Statistics Training Course

for

Crop Scientists

Statistics Training Course for Crop Scientists

Velva, North Dakota
January 29 – 31, 2007

Maryann O'Donnell
Mick O'Neill
Statistical Advisory & Training Service Pty Ltd
www.stats.net.au

Statistics Training Course for Crop Scientists

Velva, North Dakota
January 29 – 31, 2007

In recent years a general algorithm, Restricted Maximum Likelihood (REML) has been developed for estimating variance parameters in linear mixed models (LMM).

This course will review classic statistical techniques (ANOVA & REGRESSION) and demonstrate how LMM (REML) can be used to analyse normally distributed data from virtually any situation. REML reproduces the statistics familiar from ANOVA for balanced data, but is *not* dependent on balance. It allows for spatial and/or temporal correlations, so can be used for time series, repeated measures or field-correlated data. It allows for changing variances, so can be used in experiments when some treatments (different spacings, crops growing over time) induce a different variance structure.

The statistical package GenStat, arguably the best statistical tool for agricultural scientists, will be used for the course. With one stroke quite complex designs can be randomly generated for immediate use in the field, and a one-click use of the mouse allows a subsequent analysis of the collected data. Participants in this course will be shown how to use GenStat to design and analyse a wide array of agricultural experiments. The GenStat license will be valid for 60 days to allow users to explore their own data back in the comfort of their office. For information on GenStat visit www.vsni.co.uk.

In general, data from familiar text books will be used as examples. The editions we used are the following.

Mead, R. and Curnow, R.N. (1983). *Statistical Methods in Agriculture and Experimental Biology*. London; New York: Chapman and Hall.

Snedecor, G.W. and Cochran, W.G. (1980). *Statistical Methods.* Seventh Edition. Ames Iowa: The Iowa State University Press.

Sokal, R.R. and Rohlf, F.J. (1995). Biometry. *The Principles and Practice of Statistics in Biological Research.* Third Edition. New York: W.H Freeman and Company.

Steel, R.G.D. and Torrie, J.H. (1980). *Principles and Procedures of Statistics: a Biometrical Approach.* Second Edition. New York: McGraw-Hill Kogakusha.

One dataset was also taken from each of the following sources.

Allen, D.M. and Cady, F.B. (1982). *Analyzing Experimental Data by Regression*. Belmont, CA: Wadsworth.

McConway, K.J., Jones, M.C. and Taylor, P.C. (1999). *Statistical modeling using GenStat*. Arnold Publishing, London.

Ratkowsky, D.R. (1990). *Handbook of nonlinear regression models.* New York: Marcel Dekker.

Reynolds, P.S. (1994). Time-series analyses of beaver body temperatures. In *Case Studies in Biometry*. N. Lange, L. Ryan, L. Billard, D. Brillinger, L. Conquest and J. Greenhouse (editors), 211–228. New York: John Wiley.

This course is sponsored by **Agro-Tech, Inc.** The training manual was prepared by the course presenters Mick O'Neill and Maryann O'Donnell from the **Statistical Advisory & Training Service Pty Ltd** (Australia). We hope you enjoy the three days. Contact details are as follows.

AGRO-TECH RESEARCH FARM

STATS

web address: www.stats.net.au

Curt Lee

Agro-Tech, Inc.
Research Farm
4489 Highway 41 North
Velva, North Dakota 58790

Voice: 701.338.2589
Fax: 701.338.2394

E-mail: agrotec@goNDTC.com

Statistical Advisory & Training Service Pty Ltd
Box 939
Tumut
NSW 2720
Australia

Mick O'Neill	mick@stats.net.au	+61 405 595 654
Maryann O'Donnell	maryann@stats.net.au	+61 419 511 161

Table of Contents

SECTION 1 CORRELATION & REGRESSION

SECTION 2 EXPERIMENTAL DESIGN

SECTION 3 LINEAR MIXED MODELS (REML)

SECTION 4 ANALYSIS OF NON-NORMAL DATA

SECTION 5 CHOICE OF SAMPLE SIZE

Correlation & Regression

Firstly, suppose we have *n* pairs of observations, (X_1, Y_1), (X_2, Y_2), …, (X_n, Y_n). Both could be *random* variates, or one (say the *X* variate) could be controlled as part of the experiment (e.g. different set temperature chambers, sowing densities) and is hence a *fixed* variate.

Both correlation and simple linear regression coefficients measure the degree of the *linear relationship* between two variables. To summarise the difference:

Regression is used when one is interested in explaining a relationship between the dependent variate Y and the fixed variate X. It may also be used to predict future observations. If X is measured with error, the regression is interpreted as conditional on the X-values observed.

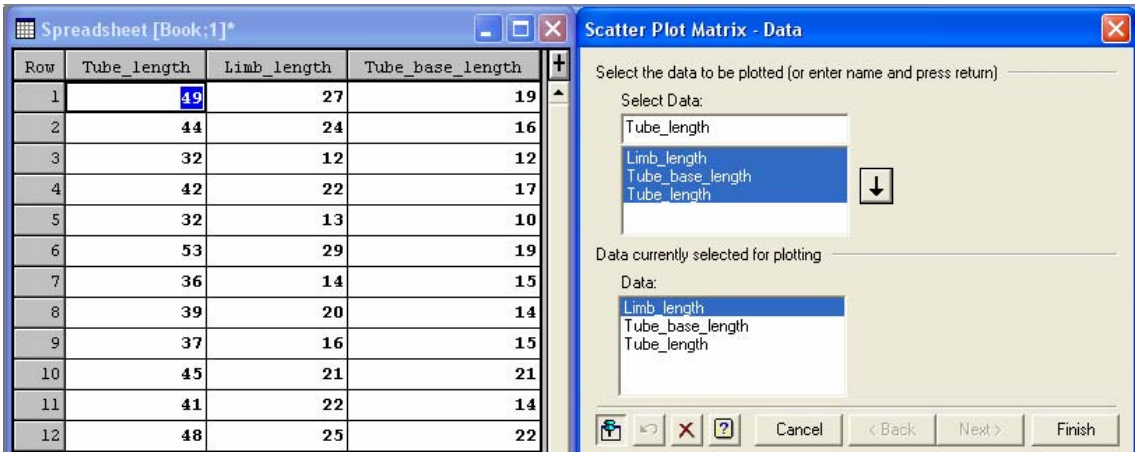
Correlation is used when one is simply interested in measuring the co-relation between two variates that appear to vary linearly with each other. Neither X nor Y is more important, they are both variates of interest.

Correlation

Example 1. Data on flowers of a Nicotiana cross (Steel and Torrie, page 276)

Tube length	49	44	32	42	32	53	36	39	37	45	41	48	45	39	40	34	37	35
Limb length	27	24	12	22	13	29	14	20	16	21	22	25	23	18	20	15	20	13
Tube base length	19	16	12	17	10	19	15	14	15	21	14	22	22	15	14	15	15	16

This is clearly when correlation is of interest. GenStat allows all three variates to be plotted against each other. Select **Graphics > Scatter Plot Matrix** and select all three variates into the Data box.



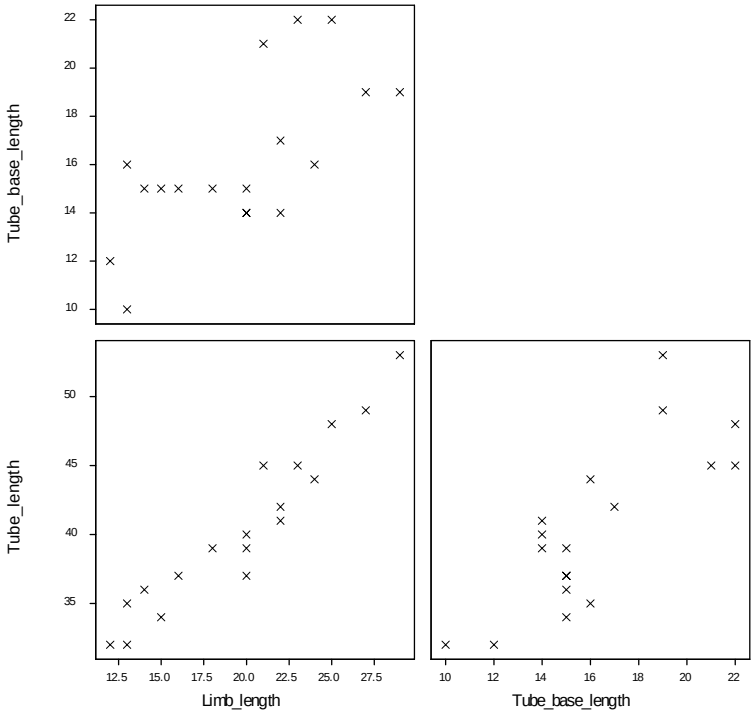
The plot on the following page shows a strong relationship between tube and limb lengths, a relatively strong relationship between tube and tube base lengths, and a slightly weaker linear relationship between tube base and limb lengths. We quantify this strength by the correlation coefficient defined as

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{x})(Y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{x})^2 \sum_{i=1}^{n}(Y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n}(X_i - \bar{x})(Y_i - \bar{y})/(n-1)}{s_x s_y} = \frac{\text{covariance}(X,Y)}{\text{sd}(X)sd(Y)} \quad (1)$$

Here we are using $\bar{x}$ and $\bar{y}$ as sample means, s as the sample standard deviation, and we introduced the concept of *covariance*. The sample standard deviation (sd) for Y is defined as

$$s_y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{y})^2}{n-1}} \quad (2)$$

The term inside the square root is the sample *variance*. The *covariance* simply replaces the squared term by an equivalent expression in the second variate thereby measuring the co-variance between Y and X. Covariances are unbounded.



For the data in (1) GenStat returns the correlation coefficients below (use **Stats > Summary Statistics > Correlations** and select the 3 variates):

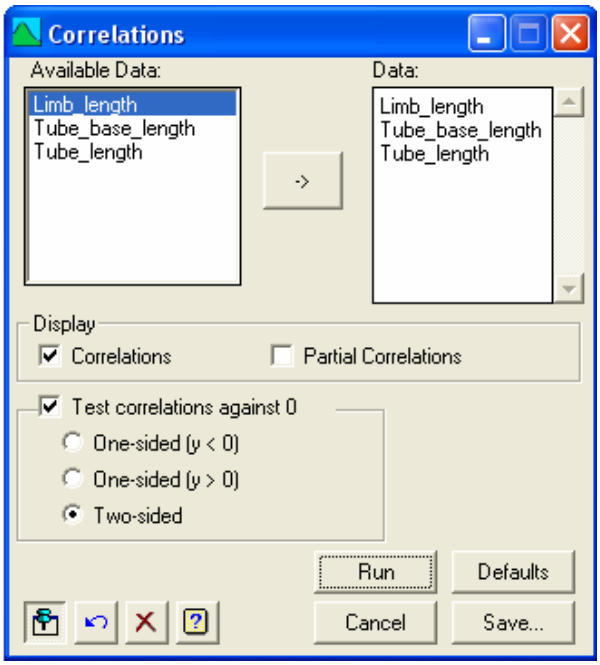
Correlation matrix

Limb_length	1.000		
Tube_base_length	0.678	1.000	
Tube_length	0.955	0.797	1.000
	Limb_length	Tube_base_length	Tube_length

Correlation coefficients are constrained to lie between -1 and +1. If one variable tends to increase as the other decreases, the correlation coefficient is negative. Conversely, if the two variables tend to increase together the correlation coefficient is positive. A correlation of 1.0

indicates a perfect linear trend with a positive slope. A correlation of 0 indicates no linear trend. If the variates are also *normally distributed*, then a correlation of 0 also indicates that *X* and *Y* are independent.

The symbol ρ (rho, Greek r) is usually used for a population correlation coefficient and *r* for a sample coefficient. A special test is available to determine whether variates are uncorrelated, that is, whether ρ = 0. The P-values from GenStat are as follows.



Two-sided test of correlations different from zero

Probabilities		
Tube_base_length	0.001981	
Tube_length	< 0.001	< 0.001
	Limb_length	Tube_base_length

Clearly, all three variates are strongly linearly related to each other.

Correlated data are extremely common in field experimentation. Sometimes the same plant or plot is measured at various times, and generally observations taken over a short time interval are more strongly correlated than those taken over a long time interval. Similarly, plants grown in a field tend to be more strongly correlated than those grown at distance. Spatial and temporal correlation models have been developed to cater for these common phenomena.

Calculation in Excel

Suppose the tube length data are named **Tube_length** in Excel and the limb length data **Limb_length**.

= CORREL(Tube_length,Limb_length) returns the value 0.9550 (to 4 decimals).

If the **Data Analysis Toolpak** has been added into Excel, the correlation macro produces:

	Tube length	*Limb length*	*Tube base length*
Tube length	1		
Limb length	0.95497792	1	
Tube base length	0.79721422	0.678111257	1

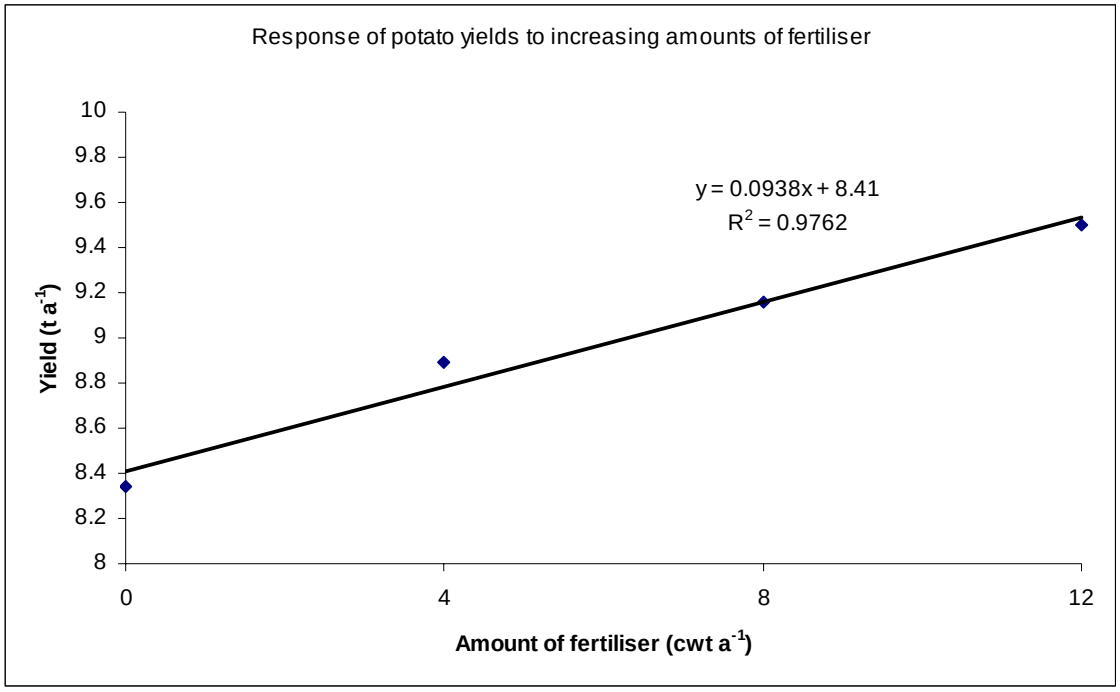
Warning on calculating covariances in Excel:
Excel has a sample variance formula **=VAR** and a sample standard deviation formula **=STDEV**. It has a "population" variance formula **=VARP** and a "population"standard deviation formula **=STDEVP**. However, the formula **=COVAR(x,y)** does ***not*** give us what we want. Instead, Excel uses *n* as a divisor instead of *n*-1!

Simple Linear Regression

Example 2. Yields of potatoes receiving amounts of fertilizer (Snedecor and Cochran, page 150).

Amount	0	4	8	12
Yield	8.34	8.89	9.16	9.50

This is not a large data set, but a scatter plot in Excel showing a linear trendline indicates a very strong predictive model for yield *over the range of fertiliser levels considered*. We could not use any model generated from these data to predict the yield for more than 12 units of fertiliser.



The line of best fit,

$$\textit{Yield} = 8.41 + 0.0938\ \textit{Fertiliser}$$

comes from a procedure known as *least squares*. Drop a perpendicular from each observation to a straight line passing through the points: these are the so-called errors, or residuals. Find the *Residual Sum of Squares*, which is simply the sum of the distances of the errors. Use a mathematical or numerical procedure to minimise the *Residual Sum of Squares*, thereby obtaining a line that goes through the points "as best as possible".

The general form of a simple linear regression line (in one predictor X_1) is

$$Y = b_0 + b_1 X_1$$

Here, b_0 is the intercept and b_1 the slope, that is, the change in Y for a unit increase in X. For the potato data, a crop with no fertiliser is predicted to produce 8.41 cwt a^{-1}, and for each additional unit of fertiliser added, an increase in yield of 0.09 cwt a^{-1} is predicted.

Furthermore, for simple linear regression, the percentage variation in yield explained by the model is 97.6%. This, in fact, is the square of the correlation coefficient, which turns out to be 0.988. (You can verify that $0.988^2 = 0.976$.)

This model is a special case of a more general linear additive model involving several predictors which we examine now in more detail.

Multiple Linear Regression

The more general multiple linear regression model applies to data taken on a dependent variable Y and a set of k predictor or explanatory variables $X_1, X_2, \ldots, X_k$. We assume we have n sets of data.

With multiple linear regression we explain the variation in the Y values by the following (usually over-simplified) relationship between Y and the set of Xs

$$Y = (\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k) + error$$

Notice that the linearity refers to the set of parameters $\beta_0, \beta_1, \ldots, \beta_k$. Polynomial equations are special cases, with X_1, $X_2 = X_1^2$, $X_3 = X_1^3$ and so on. Polynomials in X are still linear in the parameters $\beta_0, \beta_1, \ldots, \beta_k$.

The least squares procedure is again used to produce a "line of best fit".

In Example 3 we are interested in predicting burn times using a 3-predictor regression of log(leaf burn) on nitrogen (*N*), chlorine (*Cl*) and potassium (*K*) percentages in tobacco taken from farmers' fields.

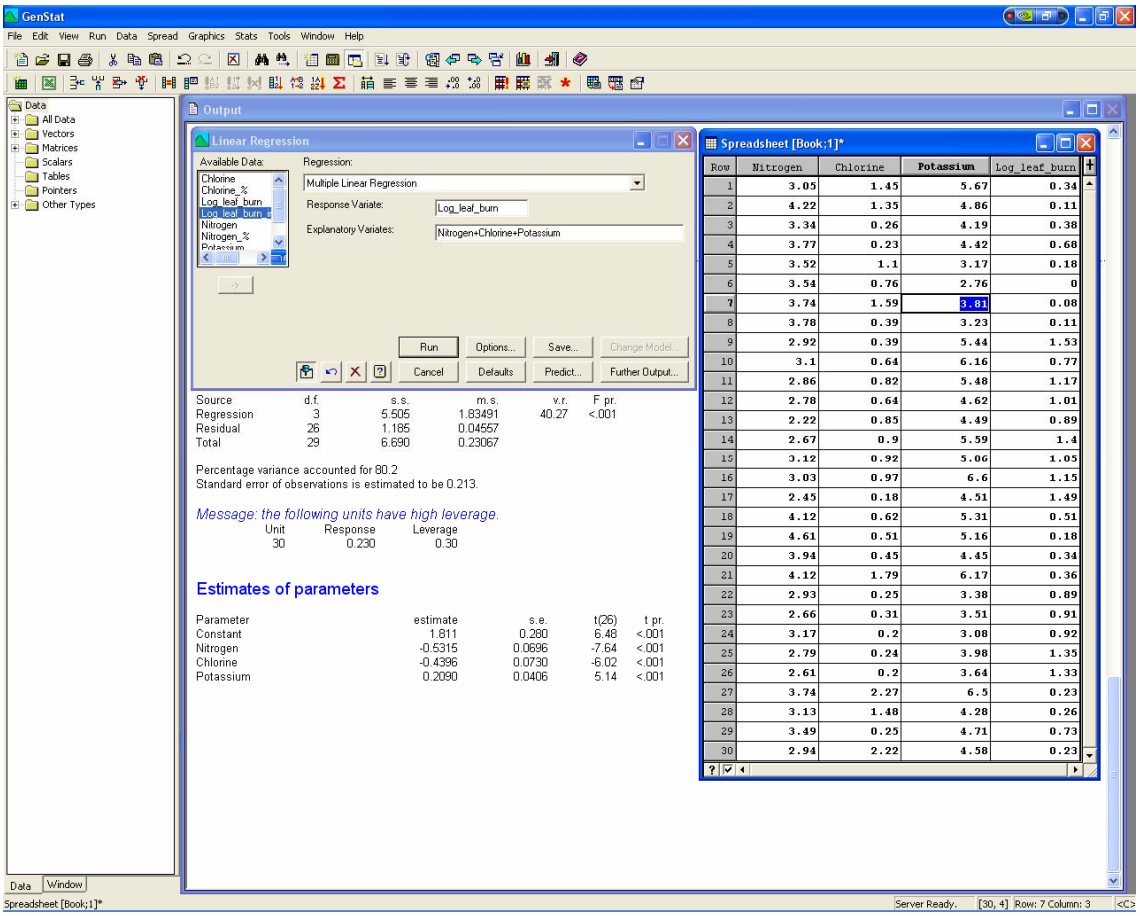
A scatter matrix of the data (see over) shows weak correlations among the dependent variates (*N*, *Cl* and *K*), as well as negative trends in log(leaf burn) on nitrogen and on chlorine, and a weak positive trend with potassium. These correlations are:

Correlation matrix

Nitrogen	1.000			
Chlorine	0.209	1.000		
Potassium	0.093	0.407	1.000	
Log_leaf_burn	-0.718	-0.500	0.179	1.000
	Nitrogen	Chlorine	Potassium	Log_leaf_burn

Example 3 (Steel and Torrie, page 319)

N	*Cl*	*K*	Log(leaf burn)
3.05	1.45	5.67	0.34
4.22	1.35	4.86	0.11
3.34	0.26	4.19	0.38
3.77	0.23	4.42	0.68
3.52	1.10	3.17	0.18
3.54	0.76	2.76	0.00
3.74	1.59	3.81	0.08
3.78	0.39	3.23	0.11
2.92	0.39	5.44	1.53
3.10	0.64	6.16	0.77
2.86	0.82	5.48	1.17
2.78	0.64	4.62	1.01
2.22	0.85	4.49	0.89
2.67	0.90	5.59	1.40
3.12	0.92	5.86	1.05
3.03	0.97	6.60	1.15
2.45	0.18	4.51	1.49
4.12	0.62	5.31	0.51
4.61	0.51	5.16	0.18
3.94	0.45	4.45	0.34
4.12	1.79	6.17	0.36
2.93	0.25	3.38	0.89
2.66	0.31	3.51	0.91
3.17	0.20	3.08	0.92
2.79	0.24	3.98	1.35
2.61	0.20	3.64	1.33
3.74	2.27	6.50	0.23
3.13	1.48	4.28	0.26
3.49	0.25	4.71	0.73
2.94	2.22	4.58	0.23



Regression analysis

Response variate: Log_leaf_burn
Fitted terms: Constant, Nitrogen, Chlorine, Potassium

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	5.505	1.83491	40.27	<.001
Residual	26	1.185	0.04557		
Total	29	6.690	0.23067		

Percentage variance accounted for 80.2
Standard error of observations is estimated to be 0.213.

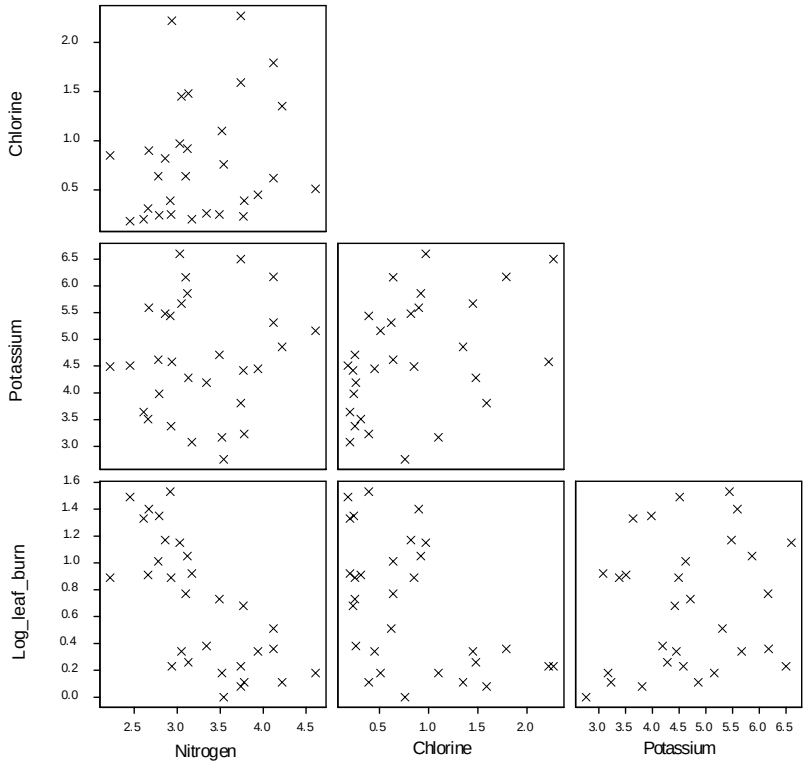
Message: the following units have high leverage.

Unit	Response	Leverage
30	0.230	0.30

Estimates of parameters

Parameter	estimate	s.e.	t(26)	t pr.
Constant	1.811	0.280	6.48	<.001
Nitrogen	-0.5315	0.0696	-7.64	<.001
Chlorine	-0.4396	0.0730	-6.02	<.001
Potassium	0.2090	0.0406	5.14	<.001

The largest R^2 is associated with nitrogen: if one were interested in a single predictor equation only, then nitrogen would be the best predictor. However, only a fraction over 50% ($-0.718^2 = 0.516$) of the variation in log(leaf burn) data is explained by this simple relationship.



Scatter matrix of nitrogen, chlorine, potassium and log(leaf burn) data of Example 3.

To perform multiple regression in GenStat, choose **Stats > Regression Analysis > Linear Models**. **Simple Linear Regression** refers to models with one predictor (including polynomials in one predictor). In this example, we select **Multiple Linear Regression** (the other choice, **with Groups**, allows regression equations to be compared across the levels of some factor). Simply select the response variate and the set of explanatory variates of interest.

The line of best fit is

log(leaf burn) = 1.811 - 0.5315 *N* - 0.4396 *Cl*+ 0.2090 *K*

Interpreting regression coefficients

A particular regression coefficient indicates the amount that *Y* will increase (or decrease) by for a unit rise in that predictor variable, *keeping the other predictor variables fixed*.

For example, for two types of tobacco with the *same* percentage of chlorine and potassium, one with 1% additional nitrogen will burn for -0.5315 fewer log-seconds compared to the other, that is, for only about 30% ($= 10^{-0.5315}$) of the time if the transformation used was base10, or about 60% ($= e^{-0.5315}$) of the time if the transformation used was the natural base.

Sometimes it is sensible to interpret the intercept, but it does not always make biological sense to do so. In this case, a value of 1.811 would indicate the log-time that tobacco would burn in the absence of any nitrogen, chlorine and potassium. However, while chlorine is as low as 0.18%, nitrogen and potassium both exceed 2% for all tobacco samples. Interpreting the intercept in this case is like predicting too far away from the experimental data, which is not valid.

In such cases, it might be better to re-write line of best fit by noting the actual solution for the intercept. For line of best fit

$$Y = b_0 + b_1 X_1 + \ldots + b_k X_k$$

using the LS solution

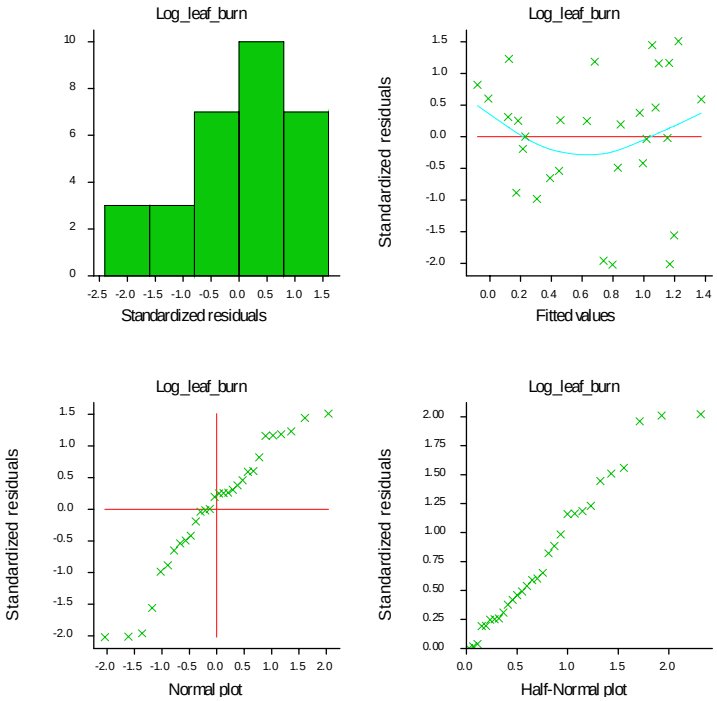
$$b_0 = \bar{y} - b_1\bar{x}_1 - \ldots - b_k\bar{x}_k$$

Allows us to write the line as

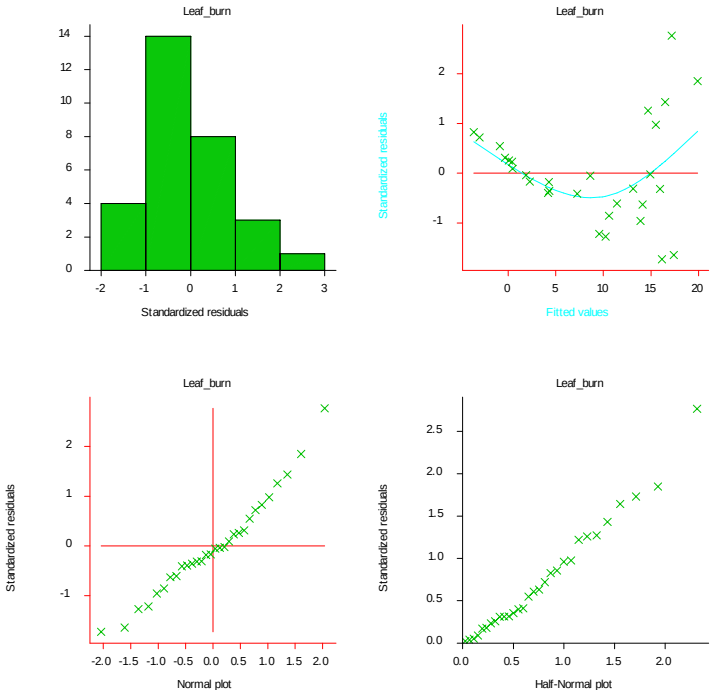
$$Y = \bar{y} - b_1\left(X_1 - \bar{x}_1\right) - \ldots - b_k\left(X_k - \bar{x}_k\right).$$

This simply emphasises predictor variates *centred to their mean*, and is an option in some GenStat procedures (eg Linear Mixed Models). For the current example, the re-arranged model is

log(leaf burn) = 0.686 - 0.5315 (N – 3.2787) - 0.4396 (Cl – 0.8077) + 0.2090 (K – 4.6537)



Standardised residual plot for the regression of log(leaf burn) on *N*, *Cl* and *K*.



Standardised residual plot for the regression of leaf burn on N, Cl and K.

Checking model assumptions

Standard practice with any analysis is to check that model assumptions appear satisfactory.

Normality. This assumption is not the most critical assumption, but can be checked in GenStat using histograms or probability plots of residuals. Histograms are not particularly useful for small data sets.

Constant variance A plot of standardised residuals against fitted values is one way to detect a problem with the variance assumption. The plot should show no trend, be randomly scattered around 0, with positive and negative values equally likely at any point. Most of the points should lie within ± 2. Fanning is indicative of data whose variance increases with the mean, and is often corrected by analyzing log-transformed data instead.

The nature of the treatments in an experiment may give rise to the suspicion that the variance may change. For example, a fanning residual plot may be due to the presence of a control treatment: plots untreated may just vary differently to treated plots. A more extreme example arises in say herbicide trials, where an increase in the amount of herbicide leads to a severe reduction in yield, with little variation. Log-transforming will not solve these problems: removing the control data is one solution, using a modern REML analysis with changing variance is preferable.

Independence Lack of independence can be detected spatially by plotting residuals in field position (an option in GenStat's ANOVA menus). If there is a time element to the design, then a plot of residuals over time is valuable.

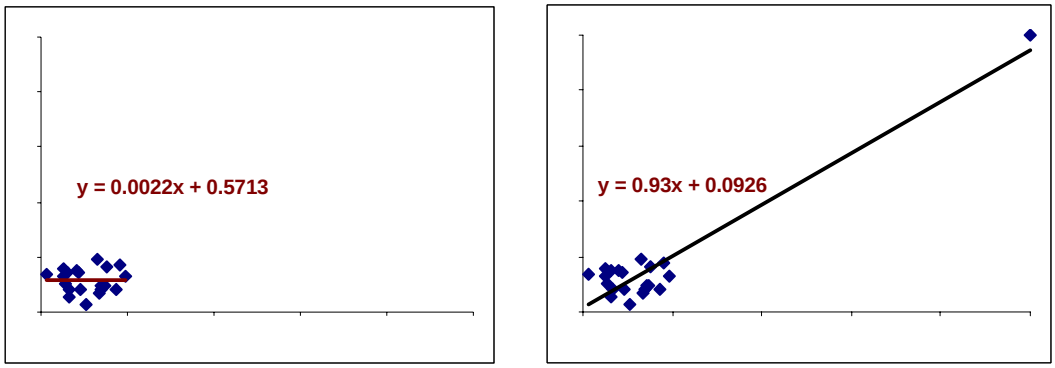
For the log(leaf burn) data, the 3-predictor model produces residuals show a very slight trend and possible fanning, but given the size of the dataset, there are no real concerns with any of the model assumptions. The plot is obtained *once the analysis is performed* by clicking on **Further Output > Model Checking**. **Deviance** residuals are standardized.

Compare the top residual plot with that obtained from an analysis of untransformed data. There is a very marked trend and fanning, which led the researchers to transform leaf burn times.

GenStat will flag potential outliers (standardised residuals outside the ± 2 region) and influential points.

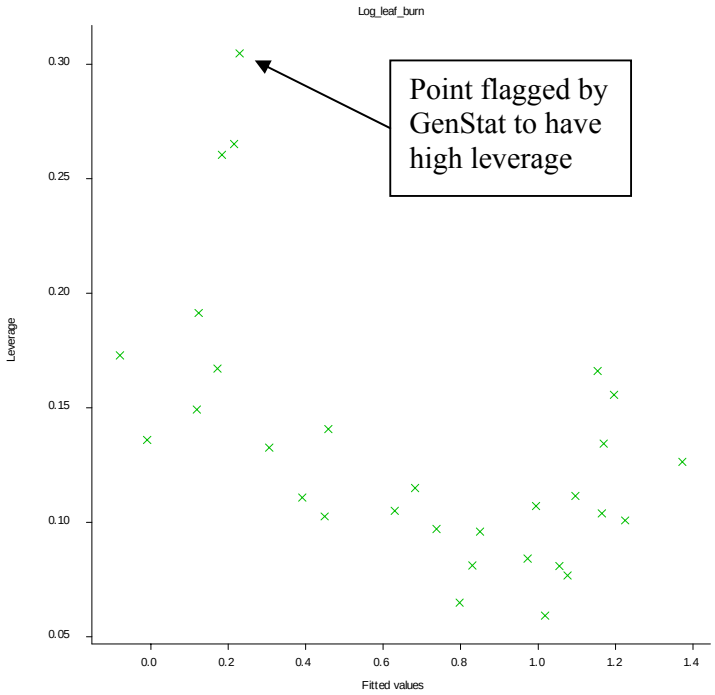
One such influential point was indicated in this analysis. What does it mean?

An **influential point** is one which has a strong influence in the fitting of the line. Consider the following hypothetical example.



The presence of just one point in the right hand diagram has dramatically affected the fitted model: the slope changes from about 0 without the point, to 1. That is not to say the point isn't important: outliers and influential points often tell you more about the system than the rest of the so-called "good data" (the discovery of the hole in the ozone layer being a dramatic example).

You can choose to plot leverages instead of standardised residuals. For the current example, two other data points appear to have high leverage, but obviously not high enough to fail GenStat's leverage test.



Plot of leverages for the log(leaf burn) analysis

The regression ANOVA

The Regression ANOVA actually tests whether *Y* is linearly dependent on the complete set of *X*s. It is not a position that in general we believe scientifically, but is often the starting point to model exploration. To drop all *X*s from the model we set up null and alternative hypotheses as follows:

H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ vs H_1: at least one β parameter ≠ 0

REGRESSION ANOVA for this set of hypotheses

Source of Variation	*df*	*SS*	*MS*	*F*	*P*
Regression	k	Regression SS	$\frac{\text{Regression SS}}{\text{Regression df}}$	$\frac{\text{Regression MS}}{\text{Residual MS}}$	✔
Residual	$n-k-1$	Residual SS	$\frac{\text{Residual SS}}{\text{Residual df}}$		
Total	$n-1$	Total SS	Sample variance of all the data		

There are mathematical formulae in any standard text book for these sums of squares (*SS*) and mean squares (*MS*). Note that GenStat uses **v.r.** (variance ratio) for the *F* statistic (for that is what it is, a ratio of two potential estimates of the same variance), and **F.pr.** for the *P* value (since the *P* value is the probability of observing a variance ratio as large as, or larger than, the one observed, assuming an F distribution).

Another feature of regression in GenStat is that it actually completes the *Total MS* cell. You can check that the sample variance of the log(leaf burn) data is 0.231. Hence the name: *analysis of variance*.

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	5.505	1.83491	40.27	<.001
Residual	26	1.185	0.04557		
Total	29	6.690	**0.23067**		

Percentage variance accounted for 80.2
Standard error of observations is estimated to be 0.213.

The *Regression SS* is the variation in log(leaf burn) data that the model explains.

The *Residual SS* is the variation in log(leaf burn) data that the model fails to explain. It is exactly what it says. Calculate the residual for each observed value

$$\begin{aligned} \textit{Residual} &= \textit{Observed} - \textit{Fitted} \\ &= Y - [b_0 + b_1 X_1 + \ldots + b_k X_k] \\ &= \text{log(leaf burn)} - [1.811 - 0.5315\, N - 0.4396\, Cl + 0.2090\, K] \end{aligned}$$

then square each residual and sum the squared residuals.

The coefficient of determination, R^2

Statistical packages generally offer two measures of the success of the model, often called R^2 and R^2(adjusted). They are fractions, but usually expressed as percentages.

$$R^2 = \frac{\text{Regression SS}}{\text{Total SS}}$$

and is therefore a measure of the proportion of the total variability (as defined by sum of squares, not variance) explained by the regression model.

An alternative definition arises as follows. Since

Total SS = Regression SS + Residual SS,

$$R^2 = \frac{\text{Total SS - Residual SS}}{\text{Total SS}} = 1 - \frac{\text{Residual SS}}{\text{Total SS}}$$

Given that the Total MS is the sample *variance*, and the Residual MS is an estimate of σ^2, the variance of a value of Y given the set of Xs, it is more natural to switch the last definition to *variances* rather than *sums of squares*. When you do this, the resulting statistic is less biased, and a better measure to use when comparing models with different numbers of parameters.

$$R^2_{adj} = 1 - \frac{\text{Residual MS}}{\text{Total MS}}$$

and is therefore a measure of the proportion of the total *variance* explained by the regression model. In fact, GenStat prefers to use the description **Percentage variance accounted for**, in this case, about 80%.

GenStat also presents $\sqrt{\text{Residual MS}}$ as Standard error of observations is estimated to be 0.213.

Testing whether a smaller model is adequate

A general rule in statistics is that, for normally distributed statistics,

$$t_{obs} = \frac{statistic}{s.e.(statistic)} \sim \text{t variable}$$

and tests that mean value of the statistic = 0.

Hence, under the regression assumptions, dividing each parameter estimate by its standard error tests whether that parameter is zero.

Estimates of parameters

Parameter	estimate	s.e.	t(26)	t pr.
Constant	1.811	0.280	6.48	<.001
Nitrogen	-0.5315	0.0696	-7.64	<.001
Chlorine	-0.4396	0.0730	-6.02	<.001
Potassium	0.2090	0.0406	5.14	<.001

Care must be taken with this table. Consider H_0: $\beta_1 = 0$ where β_1 is the coefficient of *N* in the regression model. This is tested using t_{obs} = -0.5315/0.0696 = -7.64, which is highly significant ($P<0.001$). This says that in a model involving *N*, *Cl* and *K*, *N* cannot be dropped, (the effect of allowing $\beta_1 = 0$ is effectively to drop the variate from the model, providing that *Cl* and *K* remain in the model). Proceeding to ask whether chlorine can be dropped is dangerous: this test assumes *N* and *K* remain.

A more general test is possible. What if we were to ask whether a *subset* of the predictor variables can be omitted? For example, can we drop both chlorine and potassium?

This is equivalent to the following.

We have a ***maximal model*** involving *k* conceivable predictor or explanatory variables. These are ordered for convenience only. We are interested in dropping the last *s* predictors from this model.

$$Y = (\beta_0 + \beta_1 X_1 + \ldots + \beta_{k-s} X_{k-s} + \beta_{k-s+1} X_{k-s+1} + \ldots + \beta_k X_k) + error$$

Dropping the last *s* predictors gives rise to a ***reduced model***

$$Y = (\beta_0 + \beta_1 X_1 + \ldots + \beta_{k-s} X_{k-s}) \qquad + error$$

$$Y = (\beta_0 + \beta_1 X_1 + \ldots + \beta_{k-s} X_{k-s}) + error$$

This is equivalent to testing H_0: $\beta_{k-s+1} = \ldots = \beta_k = 0$. The way we test this is as follows.

Step 1. Fit the *maximal* model and note the regression ANOVA.

Step 2. Drop the (potentially) superfluous predictors and fit the *reduced* model.

Step 3. Calculate the *difference* in Residual SS and test this against the Residual MS from the full model.

Testing H_0: $\beta_{k-s+1} = \ldots = \beta_k = 0$

Regression Analysis	Source of Variation	*SS*	*df*	*MS*	*F*	*P*
Maximal	Using all *k Xs*	*Reg* SS_{Full}	*k*		*ignore*	
Reduced	Using first (*k-s*) *Xs*	*ReS* $SS_{Reduced}$	*k-s*		*ignore*	
Calculate by differencing	***Lack of fit***	***Diff.***	***s***	$\frac{diff.}{s}$	$\frac{Lack\ of\ fit\ MS}{Res\ MS}$	✔
Maximal	Residual	*Res SS*	***n – k -1***	*Res MS*		
Maximal	Total	*Tot SS*	*n* – 1			

For non-normal data, or for testing random effects in Linar Mixed Models (REML), the equivalent technique is known as *change in deviance*. More about this later.

Firstly, let us illustrate this with the question: can we drop *K*? This is equivalent to

H_0: $\beta_3 = 0$ vs H_1: $\beta_3 \neq 0$

Step 1 Fit the *maximal* model and note the ANOVA

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	5.505	1.83491	ignore	ignore
Residual *N*, *Cl*, *K*	26	1.185	0.04557		
Total	29	6.690	**0.23067**		

Step 2 Fit the *reduced* model and note the Residual SS

Residual *N*, *Cl* only	27	2.390			

Step 3 combine into one analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression using *N*, *Cl*, *K*	3	5.505	1.83491	ignore	ignore
Residual using *N*, *Cl* only	27	2.390			
Change in Residual SS	**1**	**1.205**	**1.205**	**26.44**	**<0.001**
Residual using *N, Cl, K*	**26**	**1.185**	**0.04557**		
Total	29	6.690	**0.23067**		

Clearly, there is strong evidence ($P<0.001$) to retain *K* in a model that has *N* and *Cl*.

Note that $\sqrt{26.44} = 5.14$, the value we already have alongside the parameter estimate in the output. (This is just another example of the fact that $t_\nu^2 = F_{1,\nu}$.)

Next, let us ask the question: can we drop *Cl* and *K*? This is equivalent to

H_0: $\beta_2=\beta_3 = 0$ vs H_1: either $\beta_2 \neq 0$ and/or $\beta_3 \neq 0$.

Step 1 Fit the *maximal* model and note the ANOVA

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	5.505	1.83491	ignore	ignore
Residual *N*, *Cl*, *K*	26	1.185	0.04557		
Total	29	6.690	**0.23067**		

Step 2 Fit the *reduced* model and note the Residual SS

Residual *N* only	28	3.244

Step 3 combine into one analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression using *N*, *Cl*, *K*	3	5.505	1.83491	ignore	ignore
Residual using *N* only	28	3.244			
Change in Residual SS	**2**	**2.059**	**1.030**	**22.59**	**<0.001**
Residual using *N, Cl, K*	**26**	**1.185**	**0.04557**		
Total	29	6.690	**0.23067**		

Not surprising here, there is strong evidence that we cannot drop *Cl* and *K* from the maximal model. Both predictors of log(leaf burn) cannot be dropped, a significantly worse predictor model eventuates.

Note. In GenStat's **Options**, clicking **Accumulated** allows the effect of adding each variable into the model to be tested (in the order in which the variables are entered).

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Nitrogen	1	3.44601	3.44601	75.62	<.001
+ Chlorine	1	0.85384	0.85384	18.74	<.001
+ Potassium	1	1.20488	1.20488	26.44	<.001
Residual	26	1.18479	0.04557		
Total	29	6.68952	0.23067		

This concept will become useful when we analyse non-normal data from designed experiments.

Redundant predictor variables

An aliased (or redundant) predictor occurs when a set of variables already included in a model completely explain the values of a new predictor. A simple example is as follows.

Suppose you have in mind a 2-variable model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + error$$

This apparently will explain 2 df. However, suppose that X_2 and X_1 are linearly related:

$$X_2 = a + b X_1$$

Then the original model only *apparently* involves two independent variates. In fact there is just one:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + error \\ &= \beta_0 + \beta_1 X_1 + \beta_2 (a + b X_1) + error \\ &= (\beta_0 + a \beta_2) + (\beta_1 + b \beta_2 X_1) + error \\ &= \beta_0^* + \beta_1^* X_1 + error \end{aligned}$$

GenStat is helpful, in that it tells you the relationship between the predictor variables in the process of removing redundant predictors.

Allen and Cady (1982) have an example where water samples were taken along a river. A land survey was conducted at each sampling site, and the percentage of land allocated to agriculture, residential, industrial and forest use recorded. A fifth variate, *other*, was included. Thus at each site,

$$agriculture + residential + industrial + forest + other = 100\%.$$

The fifth variate *other* is redundant. If you did include this variate with the other four predictors, GenStat would respond:

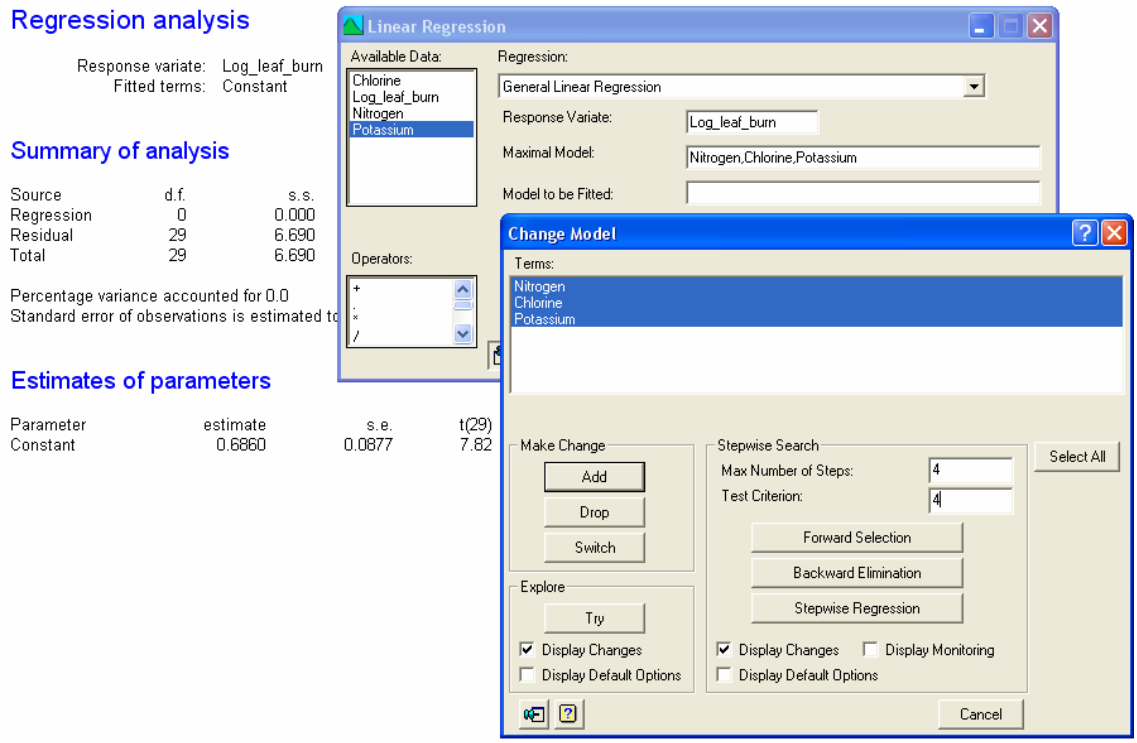
Message: term Other cannot be included in the model because it is aliased with terms already in the model.

(Other) = 100.0 - (Agriculture) - (Forest) - (Industrial) - (Residential)

The resulting model and analysis has just the first four predictors mentioned:

Estimates of parameters

Parameter	estimate	s.e.	t(15)	t pr.
Constant	1.72	1.23	1.40	0.183
Agriculture	0.0058	0.0150	0.39	0.705
Forest	-0.0130	0.0139	-0.93	0.367
Industrial	0.305	0.164	1.86	0.082
Residential	-0.0072	0.0338	-0.21	0.834



Step 1: Residual mean squares

0.1158 Adding Nitrogen
0.1793 Adding Chlorine
0.2307 No change
0.2312 Adding Potassium

Chosen action: adding Nitrogen.

Step 2: Residual mean squares

0.08851 Adding Chlorine
0.10503 Adding Potassium
0.11584 No change
0.23067 Dropping Nitrogen

Chosen action: adding Chlorine.

Step 3: Residual mean squares

0.04557 Adding Potassium
0.08851 No change
0.11584 Dropping Chlorine
0.17927 Dropping Nitrogen

Chosen action: adding Potassium.

Step 4: Residual mean squares

0.04557 No change
0.08851 Dropping Potassium
0.10503 Dropping Chlorine
0.14235 Dropping Nitrogen

Chosen action: no change.

Selection of predictor variables

Stepwise regression is a procedure for automatic selection of potentially important predictors.

Firstly, from the **Linear Regression** menu select **General Linear Regression** procedure. Enter the response variable to be analysed, and all potential predictors in the Maximal Model box, separated by **+** or **,**. There are several approaches that GenStat offers.

1. ***Forward Selection.***
   Start with *no predictor variables* in the model. Sequentially *add* variables to the model, one at a time, with the most significant predictor going in first. Variables are added until no more 'significant' terms are left to be added.

2. ***Backward Elimination***: Start with *all predictor variables* in the model. Sequentially *drop* variables from the model, one at a time, with the least significant predictor dropped first. Variables are removed until no more 'non significant' terms can be removed. What remains are significant predictors.

3. ***Stepwise***: Combines aspects of both forwards selection and backwards elimination. At each step, a new variable can be tested to determine if it significantly contributes to the model, and a variable entered earlier into the model can be tested to determine if it now can be removed from the model (because other variables entered later combine to be better predictors).

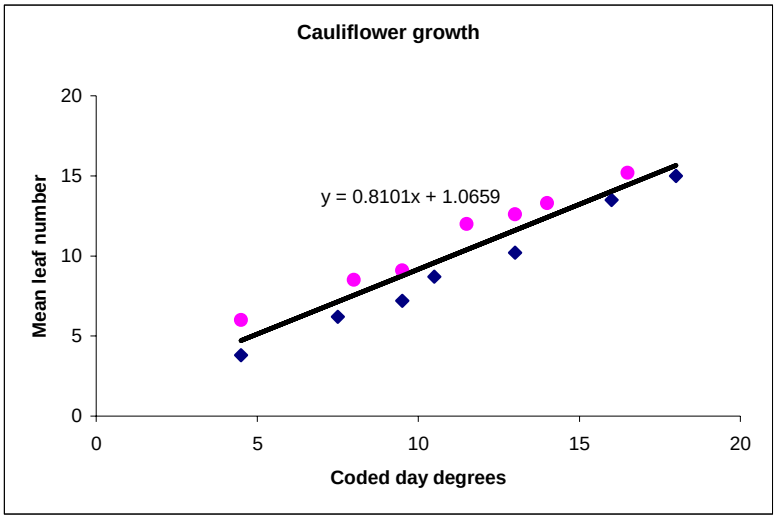
Each of these procedures is available in **Further Output**.

You need to

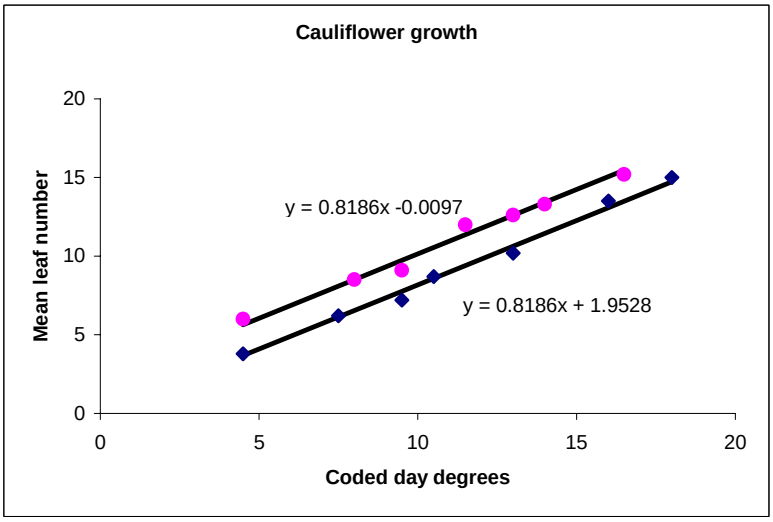
- select the predictors you wish to explore – typically by clicking **Select All**;
- set the **Max Number of Steps** you wish GenStat to use – typically the same as the number of variates and factors;
- set the **Test Criterion**. Note that the criterion of "significant" terms in the model is somewhat problematic here. Since repeated testing on the same variables is being conducted, the usual significance levels do not really apply here. Frequently, rather than using *P*-values as the testing criteria, fixed critical *F*-values (**Criterion**) are used which are not based on the actual *F*-distribution. Typically, criterion values of **4.0** are used for stepwise methods. Why? Dropping or adding a single predictor would lead to a t test with 1 numerator *df*; an $F_{1,\nu}$ critical value is the same as a (t critical value)^2 which tends to $(1.96)^2 \approx (2)^2 = 4$, so that value makes sense.

We know already that for the data in Example 3, *N*, *Cl* and *K* are all significant predictors. We should expect therefore to find the same model irrespective of which approach we use, but that is not always the case.

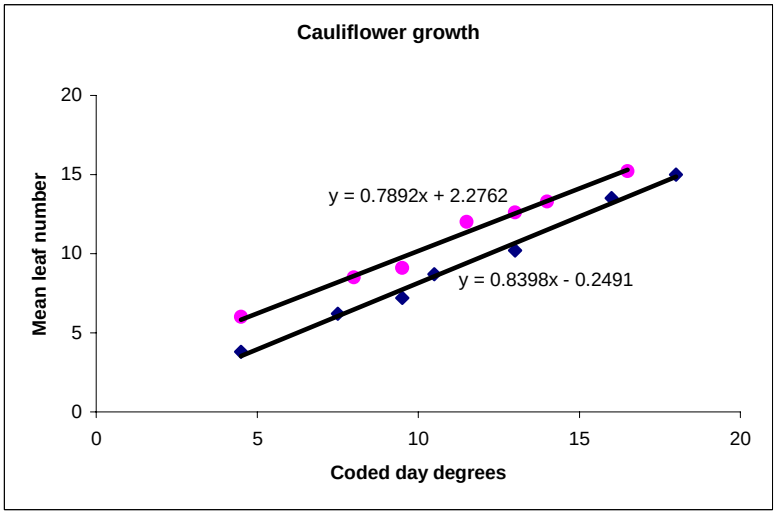
The output for stepwise regression is shown for these data. Notice that GenStat prints out the Res MS from each analysis it trials, tests the change in Res SS between the previous model and the new model, and orders the variates by the significance of the impact from the action. We are told what the chosen action is. If we are interested in the final model, we return to regression and fit the suggested model.



common line



parallel lines



separate lines

Regression with groups (factors)

One of GenStat's great strengths is its ability to allow any of the predictors to be a *factor*. Remember, a factor is a column whose entries simply identify different conditions. So Variety 1, 2, 3 is a factor; there is no relation necessarily between the 1 and 2, 2 and 3: they could have been labels A, B, C.

Mead and Curnow present the numbers of leaves (labelled Number, averaged from 10 cauliflower plants) in each of two years, and wished to relate cauliflower growth with temperature (labelled DD, measured in day degrees above 32 °F, divided by100).

Example 4. From Mead and Curnow (1990 Page 161)

1956/7 season		1957/8 season	
DD	Number	DD	Number
4.5	3.8	4.5	6.0
7.5	6.2	8.0	8.5
9.5	7.2	9.5	9.1
10.5	8.7	11.5	12.0
13.0	10.2	13.0	12.6
16.0	13.5	14.0	13.3
18.0	15.0	16.5	15.2

Interest lay in which model best describes both years (year = 1, 2) of data:

Common line: Mean leaf number = b_0 + b_1 DD 1 intercept + 1 slope = 2 parameters
Parallel lines: Mean leaf number = $b_{0,year}$ + b_1 DD 2 intercept + 1 slope = 3 parameters
Separate lines: Mean leaf number = $b_{0,year}$ + $b_{1,year}$ DD 2 intercept + 2 slope = 4 parameters

These are simple applications of testing various reduced models in a general linear model framework. The maximal model is the *separate* lines situation.

To compare the two regression lines in GenStat the data need to be stacked first, and a factor column created to identify *year*. Note that the data being analysed, average number of leaves, is related to a Poisson distribution. In fact, if the numbers of leaves on one plant is Poisson with mean μ, then the total numbers of leaves on 10 plants is Poisson with mean 10μ. The variance of a Poisson distribution is the same as the mean, so if the mean changes across day degrees or years, so must the variance. Hence, we might anticipate that the residual plots following regression will cast doubt about the constant variance assumption. To overcome this problem, we should analyse the data using log-linear modelling (to be done later).

A *common* regression line is obtained by simply analysing the stacked data ignoring the year factor; the model involves just **DD**.

DD	Number	Year
4.5	3.8	1956_7
7.5	6.2	1956_7
9.5	7.2	1956_7
10.5	8.7	1956_7
13.0	10.2	1956_7
16.0	13.5	1956_7
18.0	15.0	1956_7
4.5	6.0	1957_8
8.0	8.5	1957_8
9.5	9.1	1957_8
11.5	12.0	1957_8
13.0	12.6	1957_8
14.0	13.3	1957_8
16.5	15.2	1957 8

Parallel regression lines are obtained by adding the factor **Year** to the model to be fitted. GenStat will give output for a reference line, and the regression coefficients allow adjustments to be made for the other levels of the included factor.

Regression analysis – output for parallel lines

Response variate: Number
Fitted terms: Constant + **DD + Year**

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	2	165.532	82.7660	506.57	<.001
Residual	11	1.797	0.1634		
Total	13	167.329	12.8715		

Percentage variance accounted for 98.7
Standard error of observations is estimated to be 0.404.

Estimates of parameters

Parameter	estimate	s.e.	t(11)	t pr.
Constant	-0.010	0.337	-0.03	0.978
DD	0.8186	0.0266	30.81	<.001
Year 1957_8	1.962	0.216	9.08	<.001

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Year	1956_7

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ DD	1	152.0694	152.0694	930.74	<.001
+ Year	**1**	**13.4626**	**13.4626**	**82.40**	**<.001**
Residual	11	1.7972	0.1634		
Total	13	167.3293	12.8715		

Separate regression lines are obtained by adding the factor **DD.Year** to the model to be fitted. GenStat will give output for a reference line, and the regression coefficients allow adjustments to be made for the other levels of the included factor. Note that you now have a model

DD + Year + DD.Year

which can be shortened to **DD*Year**. More of this later in the design section.

Basically, when you have a factor with say *t* levels, GenStat uses *t* columns, each column representing a different level of the factor, with a value +1 for an observation belonging to that level of the factor, and a 0 otherwise.

The model that GenStat prints out is appropriate for the "reference" level it chooses. You can change this reference level if you wish.

In the case of the cauliflower data, the two parallel lines of best fit are given following the GenStat output.

Output from GenStat's Linear Regression with Groups

Part 1 –common model

Response variate: Number
Fitted terms: Constant + DD

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	1	152.07	152.069	119.58	<.001
Residual	12	15.26	1.272		
Total	13	167.33	12.871		

…

Estimates of parameters

Parameter	estimate	s.e.	t(12)	t pr.
Constant	1.066	0.879	1.21	0.248
DD	0.8101	0.0741	10.94	<.001

Part 2 – parallel models

Response variate: Number
Fitted terms: Constant + DD + Year

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	2	165.532	82.7660	506.57	<.001
Residual	11	1.797	0.1634		
Total	13	167.329	12.8715		
Change	**-1**	**-13.463**	**13.4626**	**82.40**	**<.001**

…

Estimates of parameters

Parameter	estimate	s.e.	t(11)	t pr.
Constant	-0.010	0.337	-0.03	0.978
DD	0.8186	0.0266	30.81	<.001
Year 1957_8	1.962	0.216	9.08	<.001

Part 3 – separate models

Response variate: Number
Fitted terms: Constant + DD + Year + DD.Year

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	165.676	55.2255	334.12	<.001
Residual	10	1.653	0.1653		
Total	13	167.329	12.8715		
Change	**-1**	**-0.144**	**0.1444**	**0.87**	**0.372**

Estimates of parameters

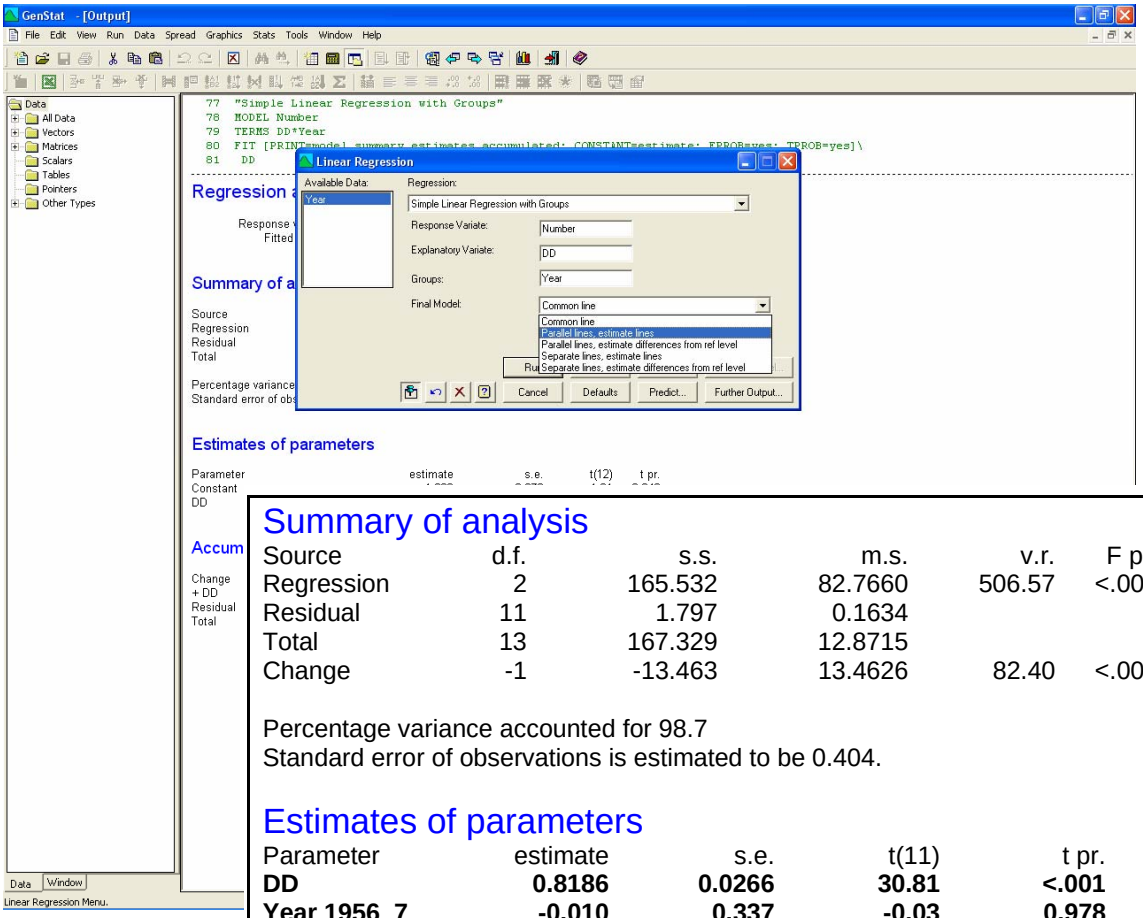
Parameter	estimate	s.e.	t(10)	t pr.
Constant	-0.249	0.425	-0.59	0.570
DD	0.8398	0.0351	23.95	<.001
Year 1957_8	2.525	0.640	3.94	0.003
DD.Year 1957_8	-0.0506	0.0542	-0.93	0.372

Mean leaf number = -0.010 + 0.8186 DD for 1956/7, the reference year,
and
Mean leaf number = (-0.010 + 1.962) + 0.8186 DD for 1956/7.
= 1.952 + 0.8186 DD

Notice in the **Accumulated** part of the analysis that the addition of this third parameter is judged to be highly significant (P<0.001).

There is actually a procedure which tests whether common, parallel or separate models are better: namely, **Linear Regression with Groups**. There are three parts to the output. There is strong evidence (P<0.001) to conclude that parallel regression lines are necessary, but no significant evidence (P=0.372) that separate lines are needed. On average, there are two extra leaves per cauliflower in the first season, however, growth over the season is similar, with about 82 leaves added for a 100 increase in (coded) day degrees.

Once the comparisons are done, you can choose which model to go with. In fact, you have the choice of re-running (and plotting, in **Further Output**) the chosen analysis so that the actual models are printed out (together with standard errors of all the intercepts and slopes) for the different factor levels, not just as they differ from the reference model.



Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	2	165.532	82.7660	506.57	<.001
Residual	11	1.797	0.1634		
Total	13	167.329	12.8715		
Change	-1	-13.463	13.4626	82.40	<.001

Percentage variance accounted for 98.7
Standard error of observations is estimated to be 0.404.

Estimates of parameters

Parameter	estimate	s.e.	t(11)	t pr.
DD	**0.8186**	**0.0266**	**30.81**	**<.001**
Year 1956_7	**-0.010**	**0.337**	**-0.03**	**0.978**
Year 1957_8	**1.953**	**0.330**	**5.92**	**<.**001

Polynomial and non-linear regression

A plot of the pasture data shows a strong linear trend with a sigmoid shape typical of plants growing over time. Again ignoring any variance problem, polynomial regression can be used, though a more biologically meaningful model may be available.

time	yield
9	8.93
14	10.80
21	18.59
28	22.33
42	39.35
57	56.11
63	61.73
70	64.62
79	67.08

Example 5. Pasture data from Ratkowsky (1990)

Polynomial regression is performed using simple or general linear regression, replacing **time** with a function **pol(time;3)**, where 3 governs the degree of the polynomial. We choose 3 with these data, anticipating the curvature at both ends.

While the model explains 99.78% of the variation in yield, it is still only a mathematical approximation for growth over the period 9 to 79 days. The fitted model is

$$\text{Yield} = 7.8838 - 0.15728\ time + 0.03336\ time^2 - 0.00028\ time^3$$

Regression analysis

Response variate: yield
Fitted terms: Constant + time
Submodels: POL(time; 3)

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	4641.734	1547.245	1222.20	<.001
Residual	5	6.330	1.266		
Total	8	4648.063	581.008		

Percentage variance accounted for 99.8
Standard error of observations is estimated to be 1.13.

Message: the following units have large standardized residuals.

Unit	Response	Residual
3	18.59	2.02

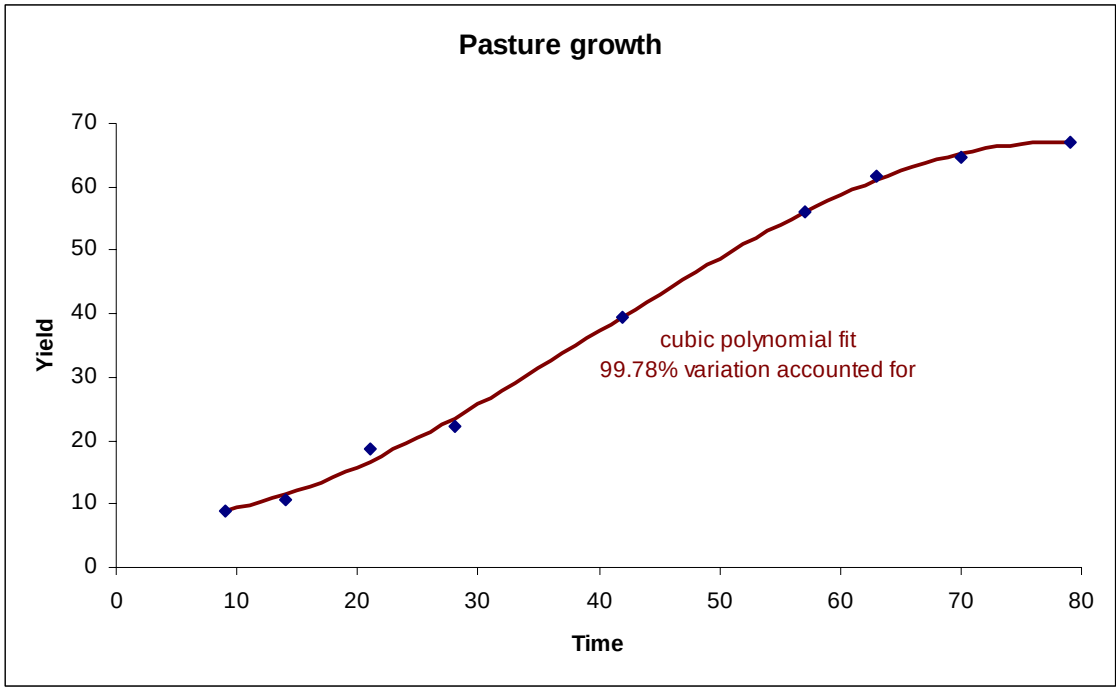
Estimates of parameters

Parameter	estimate	s.e.	t(5)	t pr.
Constant	7.88	2.43	3.25	0.023
time Lin	-0.157	0.230	-0.68	0.524
time Quad	0.03336	0.00584	5.71	0.002
time Cub	-0.0002772	0.0000436	-6.36	0.001

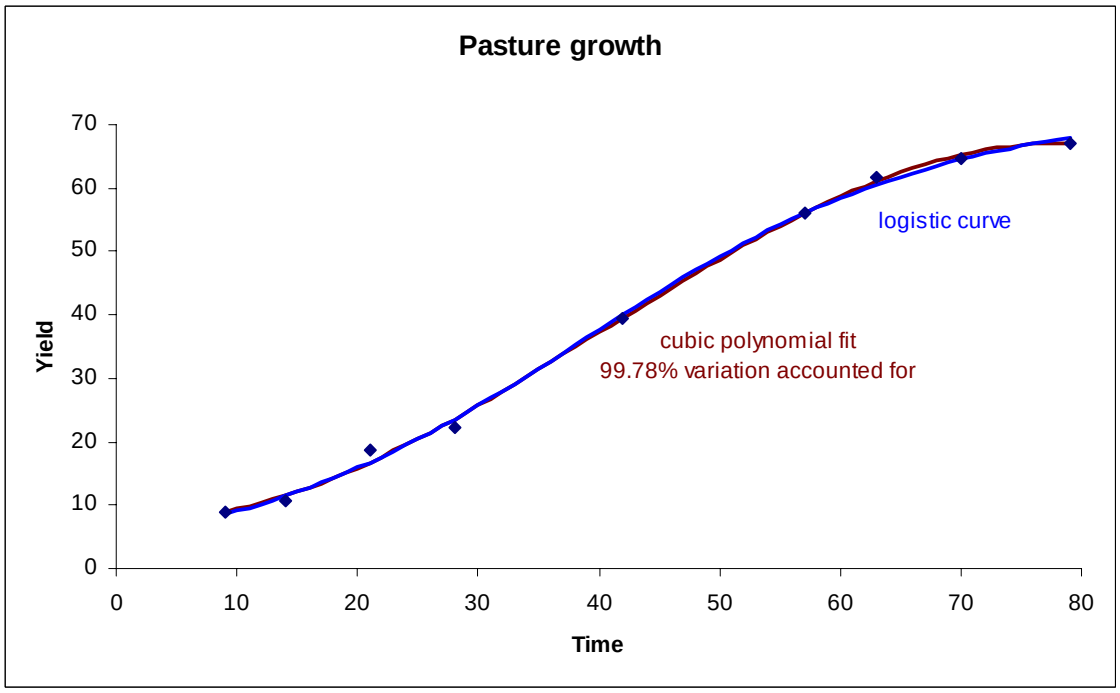
GenStat has a suite of non-linear models, including a logistic equation in the form

$$Y = A + \frac{C}{1 + e^{-B(t-M)}}.$$

This equation is commonly used for data of this type, usually with A set to 0.

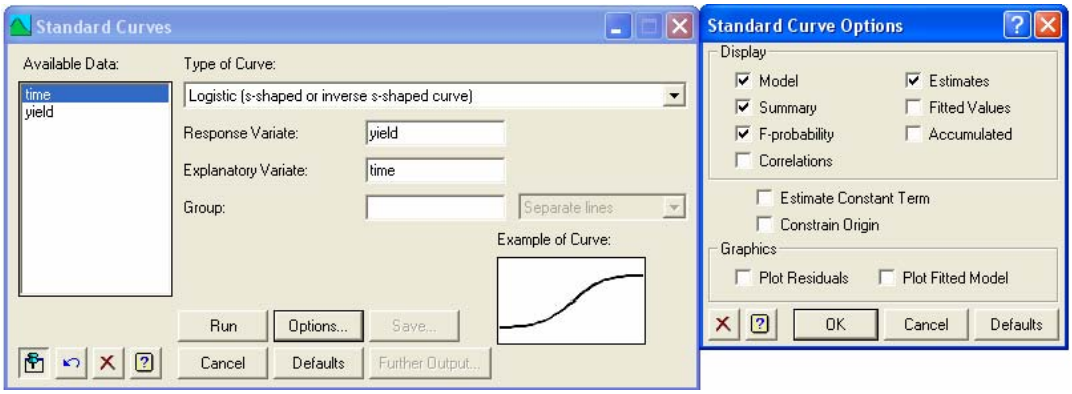


Cubic polynomial regression for pasture growth



Logistic regression (superimposed on cubic model) for pasture growth

There is very little difference visually between the two fitted curves. However, *C* is the eventual pasture yield (72.5 units) and *M* the day on which the crop is growing fastest (about day 39). On that day, the crop has grown to ½*C* (36.2 units). Relative growth rates decline from about *B* (6.7%) on day 0, to ½*B* (3.4%) on day 39 (*M*), and eventually decline to 0.



Nonlinear regression analysis

Response variate: yield
Explanatory: time
Fitted Curve: A + C/(1 + EXP(-B*(X - M)))
Constraints: A = 0.0

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	3	18215.364	6071.788	4521.89	<.001
Residual	6	8.057	1.343		
Total	9	18223.420	2024.824		

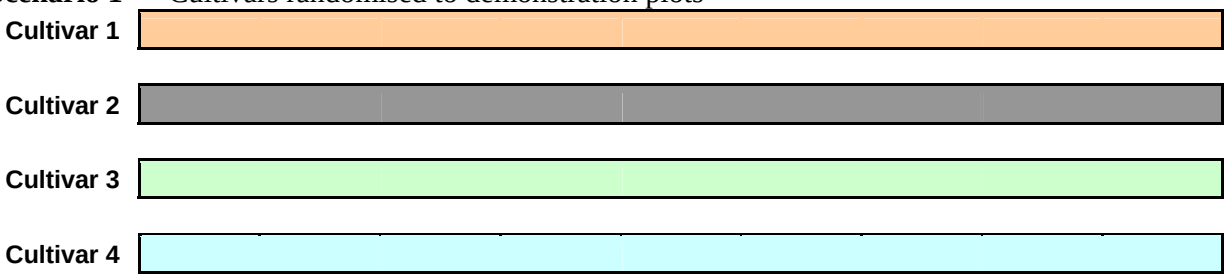
Percentage variance accounted for 99.8
Standard error of observations is estimated to be 1.16.

Estimates of parameters

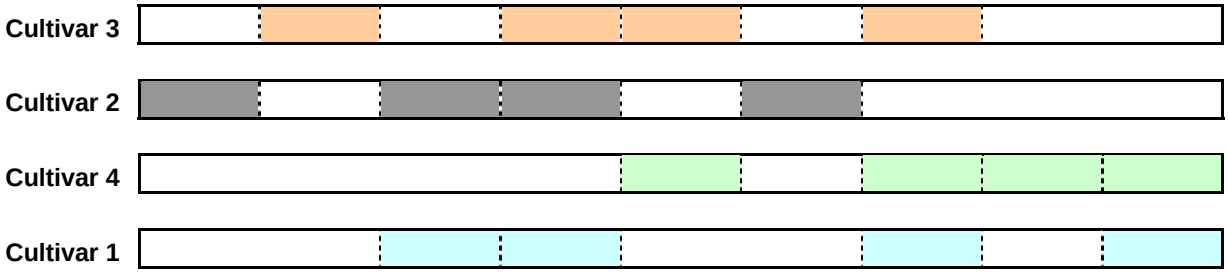
Parameter	estimate	s.e.
B	0.06736	0.00345
M	38.87	1.18
C	72.46	1.73

Various experimental scenarios

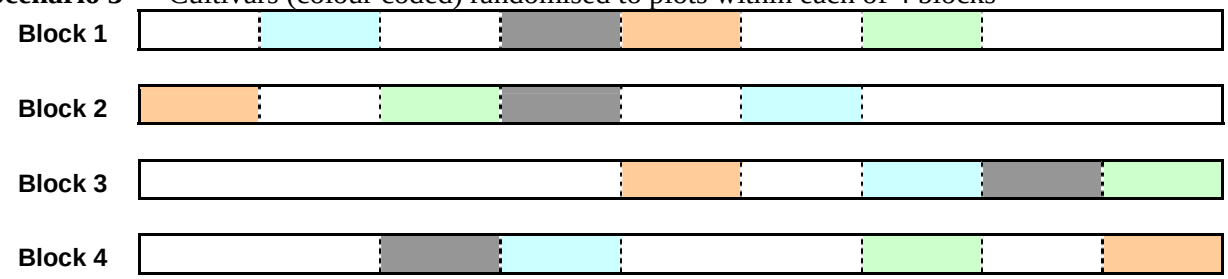
Scenario 1 Cultivars randomised to demonstration plots



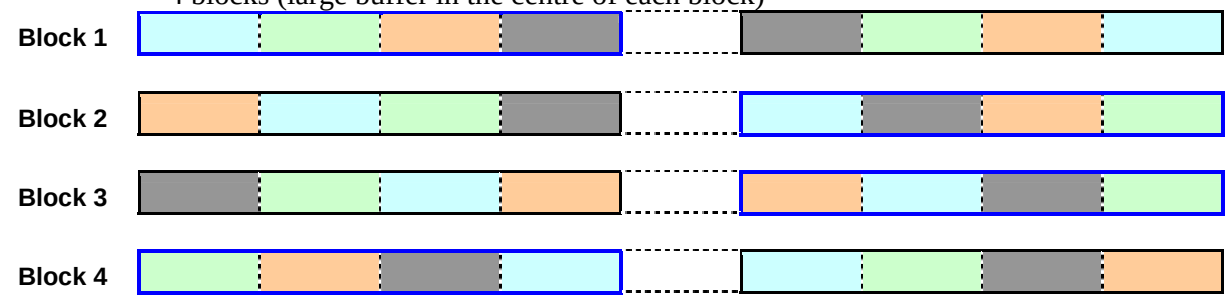
Scenario 2 Cultivars randomised to demonstration plots, 4 random grid samples taken in each



Scenario 3 Cultivars (colour coded) randomised to plots within each of 4 blocks



Scenario 4 A different method of **cultivation** (colour coded) is chosen at random to half of each block, then cultivars (colour coded) randomised to plots within each of 4 blocks (large buffer in the centre of each block)



Experimental Design

Summary of basic concepts

Before discussing experimental design and data analysis, let's review a few key concepts from standard statistics courses.

Random sampling is important to remove bias and to allow the parameters (mean, standard deviation, and so on) of the distribution from which the sample is drawn to be estimated. The more **replicates** you can provide, the more accurate will be your estimates. How many replicates to provide is often the most difficult question to answer: as we will see, we need (a) some idea of the anticipated variation in our data, as well as (b) an understanding of how large a difference we are hoping to demonstrate, before a decision can be made. When it comes to designing an experiment, GenStat will always provide a random plan for the experiment: a "blueprint" that can be used in the field. The plan is a simple spreadsheet which we augment with the data available, and analyse by simple point and click.

Treatments can only be compared if they are properly replicated. Suppose you prepare four demonstration plots and sow out four cultivars, one in each plot (Scenario 1). You cannot then compare the yields from these plots, even if you obtain several sampling areas from each plot (Scenario 2). The cultivars are not replicated. Any differences in total yield could well be accidental location differences; there is no way of separating out the cultivar effects and the location effects.

Often you perform a number of randomisations in the field, leading to differently shaped experimental units. Treatments can only be compared using replicates of the same shape. We call these different shapes *strata*.

This leads to some basic principles.

i) An **experimental unit** is the smallest amount of experimental material that one treatment is randomised to.

ii) A **sampling unit** is the smallest amount of experimental material that is actually measured.

iii) Experimental units are used in forming tests of particular treatments. Sampling units just measure how "uniform" the experimental material is, and provide no degrees of freedom for these tests.

Basically, the way you design your experiment affects the way you analyse your data.

Scenario 3 is a properly replicated trial, with each cultivar sown out in different areas. Replicates are ¼ block shapes. Blocks form one stratum (and blocks are not replicated, so strictly cannot be tested) and plots in a block form a second stratum.

Scenario 4 is also properly replicated trial. However, the blocks (stratum 1) are first divided into two large areas (stratum 2) and different cultivation techniques applied to these two areas. Cultivars are applied to smaller plots (stratum 3) within these areas, thereby affecting the way we analyse the data, as we will see.

Next, we list what needs to be known about random (uncorrelated) sampling before illustrating various *t*-tests.

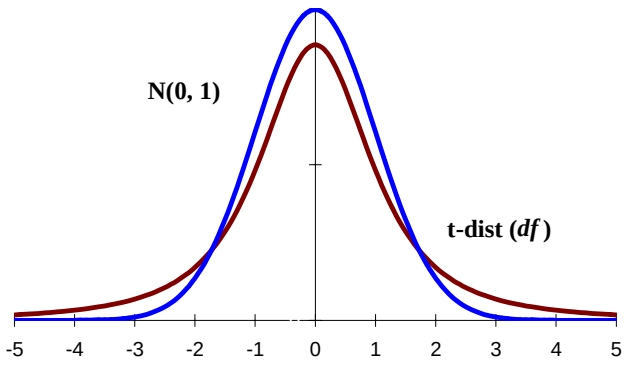
Distribution of a **sample mean** of n data values from a normal distribution with mean μ and standard deviation σ	$\bar{y}$ is normally distributed with mean μ and standard deviation $\sqrt{\sigma^2/n}$
The **standard error of a mean** (sem)	sem = $\sqrt{\sigma^2/n}$ or $\sigma/\sqrt{n}$
Distribution of the **difference between two sample means** of n_1, n_2 data values (resp.) from normal distributions with means μ_1 and μ_2 and standard deviations σ_1 and σ_2	$\bar{y}_1 - \bar{y}_2$ is normally distributed with mean μ_1-μ_2 and standard deviation $\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$
The **standard error of a difference** between two means (sed)	sed $= \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ $= \sqrt{\sigma^2(1/n_1 + 1/n_2)}$ when $\sigma_1 = \sigma_2$ $= \sqrt{2\sigma^2/n}$ when $\sigma_1 = \sigma_2$ and $n_1 = n_2$
The sample variance of Y_1, Y_2, … Y_n, defined as s^2, estimates σ^2	$s^2 = \dfrac{\sum_{i=1}^{n}(Y_i - \bar{y})^2}{n-1}$
The sample variance of $\bar{y}_1, \ldots, \bar{y}_t$ estimates σ^2/n	providing each mean comes from the same numbers of replicates from a common distribution

In experimental work, one almost never knows the true population variance σ^2, and hence it needs to be estimated. This affects the distribution used in analysing experimental data.

One-sample test statistic (we are usually interested in $\mu_1 = 0$)	$t = \dfrac{\bar{y}_1 - \mu_1}{\sqrt{s_1^2/n_1}} = \dfrac{\bar{y}_1 - \mu_1}{sem}$, $df = n$-1
Two-sample test statistics (we are usually interested in $\mu_1 - \mu_2 = 0$). When we are happy to assume $\sigma_1^2 = \sigma_2^2$ we use a *pooled* estimate of variance obtained as a weighted variance with *df* as weights: $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$	$t = \dfrac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{sed}$, where $sed = \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$ if $\sigma_1^2 \neq \sigma_2^2$, *df* complex $\sqrt{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$ if $\sigma_1^2 = \sigma_2^2$, $df = (n_1-1)+(n_2-1)$
95% confidence interval for μ	$\bar{y}_1 \pm t_{crit} sem$
95% confidence interval for $\mu_1 - \mu_2$	$(\bar{y}_1 - \bar{y}_2) \pm t_{crit} sed = (\bar{y}_1 - \bar{y}_2) \pm lsd$ where $lsd = t_{crit} sed$ is known as the "least significant difference"

For more complex analyses the estimate of variance used is based on the appropriate stratum variance (with appropriate degrees of freedom).

Background to the *t* distribution

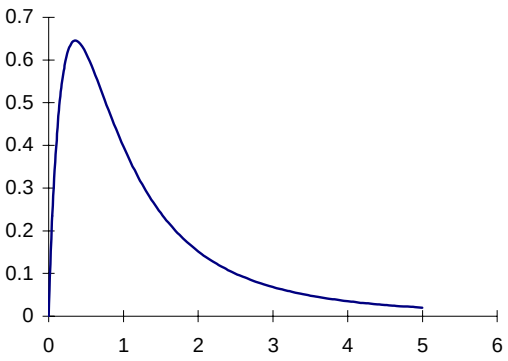


t distribution in comparison to a standard normal

The *t* distribution has "heavier" tails than the normal distribution. This distribution is used when the standard deviation has to be estimated from the sample. The larger the sample size the better the estimate of the standard deviation, thus as degrees of freedom (ν) increase, *t* tends to a normal distribution. For a *t* distribution:

mean = 0
variance = $\nu/(\nu-2)$

Background to the *F* distribution



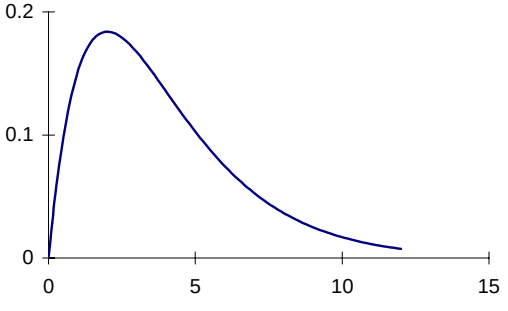
Typical shape of an *F* distribution ($\nu_2>2$)

The *F* distribution is a "heavy tailed" distribution and describes the distribution of the ratio of two competing estimates of variance, both of which are derived from normal data. There are two parameters for this distribution, namely the degrees of freedom of the numerator (ν_1) and denominator (ν_2).

For an $F_{\nu1,\nu2}$ distribution:

mean = $\nu_2/(\nu_2-2)$ (and tends to 1)
variance = function of both ν_1 and ν_2

Background to the χ^2 distribution



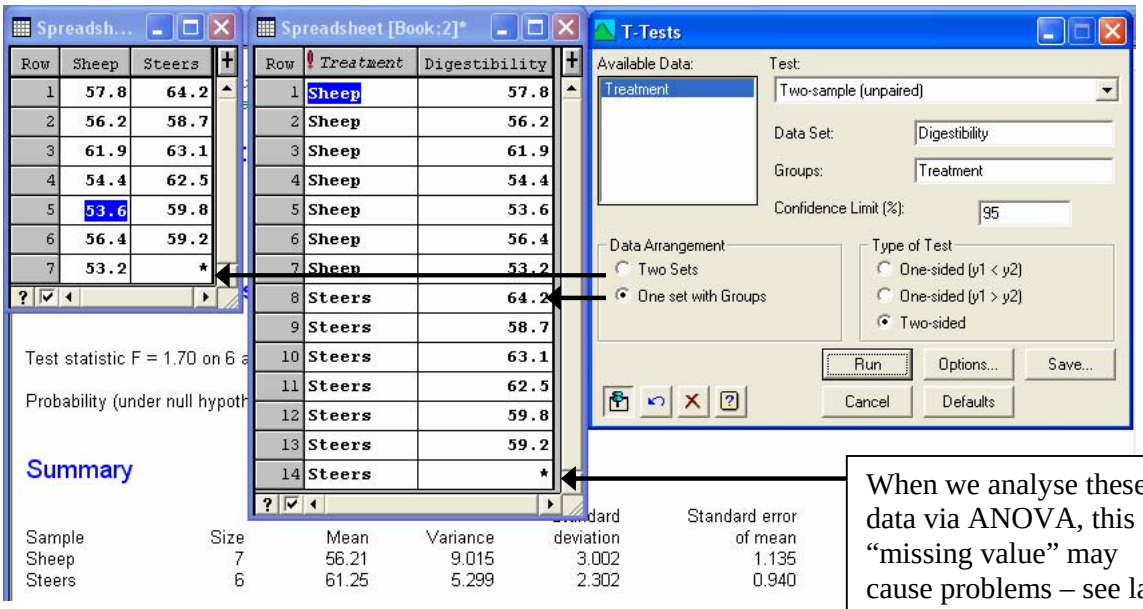
Typical shape of a χ^2 distribution (*df*>2)

The χ^2_ν distribution is a "heavy tailed" distribution and describes the distribution of the sum of squares of ν independent, standardised normal variables. The sample variance derived from *n* data values from a normal distribution is proportional to a χ^2_{n-1}.

For a χ^2_ν distribution:

mean = ν
variance = 2ν

GenStat's unpaired t test procedure



When we analyse these data via ANOVA, this "missing value" may cause problems – see later

Two-sample t-test

Variate: Digestibility
Group factor: Treatment

Test for equality of sample variances

Test statistic F = 1.70 on 6 and 5 d.f.
Probability (under null hypothesis of equal variances) = 0.58

Step 1. GenStat tests H_0: $\sigma_1^2 = \sigma_2^2$ using $F = s_1^2 / s_2^2$.
Here there is *no evidence* that the population variances are not equal (P=0.580).

Summary

Sample	Size	Mean	Variance	Standard deviation	Standard error of mean
Sheep	7	56.21	9.015	3.002	1.135
Steers	6	61.25	5.299	2.302	0.940

Difference of means: -5.036
Standard error of difference: 1.506

95% confidence interval for difference in means: (-8.350, -1.721)

Test of null hypothesis that mean of Digestibility with Treatment = Sheep is equal to mean with Treatment = Steers

Test statistic t = -3.34 on 11 d.f.
Probability = 0.007

Unpaired *t* test – special case of a one-way treatment design (no blocking)

Example 1 Coefficients of digestibility of dry matter, feed corn silage, in percent (Steel and Torrie, page 93)

	Sheep	Steers
	57.8	64.2
	56.2	58.7
	61.9	63.1
	54.4	62.5
	53.6	59.8
	56.4	59.2
	53.2	
mean	56.21	61.25
sd	3.00	2.83

To test H_0: $\sigma_1^2 = \sigma_2^2$ for normally distributed data:

$F_{obs} = \frac{s_1^2}{s_2^2}$ ~ *F* variable with (n_1-1) and (n_1-1) *df*

The first decision to make is whether you are prepared to believe that the two population variances are equal. There is a variance ratio test for this, *but this test relies very heavily on the data being normally distributed*, so use it with care.

Unless you change the default in **Options**, GenStat does the *F* test for you.

If the test does not fail, then the unpaired *t* test is used to test the means, with

$sed = \sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ and $df = (n_1-1)+(n_2-1)$

If the test does fail, then an *approximate t* test is used to test the means, with $sed = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$. The degrees of freedom are calculated from a formula; if the two sample variances are close, the *approximate df* are close to (n_1-1)+(n_2-1). When the two sample variances are different, the *approximate df* will be closer to the *df* associated with the larger variance. The formula used is shown alongside.

$$df = \left[\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1-1}+\frac{\left(s_2^2/n_2\right)^2}{n_2-1}}\right]$$

To analyse the data, use **Stats** > **Statistical Tests** > **One- and two-sample t-tests…**. GenStat allows the data to be organized either in separate columns for the separate treatments, or in one combined data column *plus* a factor column to identify which observation each treatment belongs to. Since this is a special case of a more general design, we chose to illustrate the latter approach, see the output on the left hand page.

For the coefficients of digestibility of dry matter,

- there is no evidence (*P*=0.580) that the population variances are not equal
- there is strong evidence (*P*=0.007) that the population means are different. Steers have coefficients of digestibility that are, on average, 5.0% higher than for sheep. We are 95% confident that the true difference is between 1.7% and 8.4%.

In the regression section we pointed out that a *t* test is a special case of a regression ANOVA *F* test, using Treatment as a explanatory *factor*. Here is GenStat's output for **General Linear Regression**. The model is referenced to level 1 (Sheep), hence Constant is the estimate of the Sheep mean. The coefficient Treatment Steers is what you add to the Constant to obtain the mean for the second level (Steers) and hence is the difference in means (Steers-Sheep).

Regression analysis

Response variate: Digestibility
Fitted terms: Constant, Treatment

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	1	81.93	81.927	11.18	0.007
Residual	11	80.58	7.326		
Total	12	162.51	13.543		

Percentage variance accounted for 45.9
Standard error of observations is estimated to be 2.71.

Message: the following units have large standardized residuals.

Unit	Response	Residual
3	61.90	2.27

Estimates of parameters

Parameter	estimate	s.e.	t(11)	t pr.
Constant	56.21	1.02	54.95	<.001
Treatment Steers	5.04	1.51	3.34	0.007

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Treatment	Sheep

- Same *P*-value as that for *t* test of means
- v.r. = t^2 $11.18 = 3.34^2$
- 7.326 is the pooled estimate of variance
- Constant is mean for level 1 Sheep
- *Difference in means* is 5.04
- *sem* = 1.02 for Sheep
- *sed* = 1.51

As mentioned, comparing two means is a special case of testing whether *t* means are *all* equal. Here we have randomly chosen sheep and randomly chosen steers, so this is a design with replicates taken from different groups of animals. It is known as a one-way treatment design with no blocking. Use **Stats** > **Analysis of Variance**. There is a special menu item for this design, but we prefer to select GenStat's special analysis menus from the **General** analysis of variance. We have also gone into **Options** and selected **LSD**s. Without changing the stacked spreadsheet, the output is as follows.

Analysis of variance

Variate: Digestibility

Source of variation	d.f.	(m.v.)	s.s.	m.s.	v.r.	F pr.
Treatment	1		88.754	88.754	12.12	0.005
Residual	11	(1)	80.584	7.326		
Total	12	(1)	162.511			

Message: the following units have large residuals.

units 3 5.69 s.e. 2.40

Tables of means

Grand mean 58.73

Treatment	Sheep	Steers
	56.21	61.25

Standard errors of differences of means

Table	Treatment
rep.	7
d.f.	11
s.e.d.	1.447

(Not adjusted for missing values)

Least significant differences of means (5% level)

Table	Treatment
rep.	7
d.f.	11
l.s.d.	3.184

(Not adjusted for missing values)

This is not exactly the same analysis, because with unequally replicated treatments, if you leave a row in with an asterisk (***)** to signify a missing value, GenStat assumes you want to estimate the missing value. This is rather an old fashioned approach. It *over-estimates* the Treatment SS and the resulting variance ratio is therefore too large.

If you really do have missing values, there is an **Unbalanced Treatment Structure** you can use in this case. (Basically, GenStat analyses the data via regression for you.) Using **Linear Mixed Models (REML)** is preferable, however we will defer this until the next section.

If this is a case of a deliberate choice of sample size (for example, these are the only steers you could get hold of), then a correct analysis is obtained after deleting the row with the *****.

Here are both analyses. The similarities are obvious.

Including the row with the missing value, choosing Unbalanced Treatment Structure

Analysis of an unbalanced design using GenStat regression

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Treatment	1	81.927	81.927	**11.18**	**0.007**
Residual	11	80.584	7.326		
Total	12	162.511	13.543		

Predictions from regression model

Treatment	Prediction
Sheep	56.21
Steers	61.25

Standard error of differences between predicted means	**1.506**
Least significant difference (at 5.0%) for predicted means	**3.314**

Deleting the row with the non-observed value, choosing General Analysis of Variance

Analysis of variance

Variate: Digestibility

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Treatment	1	81.927	81.927	11.18	0.007
Residual	11	80.584	7.326		
Total	12	162.511			

Message: the following units have large residuals.

units 3	5.69	approx. s.e.	2.49	

Tables of means

Variate: Digestibility

Grand mean 58.54

Treatment	Sheep	Steers
	56.21	61.25
rep.	7	6

Standard errors of differences of means

Table	Treatment
rep.	unequal
d.f.	11
s.e.d.	1.506

Least significant differences of means (5% level)

Table	Treatment
rep.	unequal
d.f.	11
l.s.d.	3.314

This analysis gives almost the same information as the *t* test did. You obtain

- the equivalent test statistic (ANOVA *F* instead of t^2)
- the same *P*-value for testing the difference between the two means (0.007)
- the same estimate of variance (7.326) and hence the same s.e.d. value (1.506).

A disadvantage is that you need to add and subtract the l.s.d. (3.314) to the mean difference (61.25-56.21) yourself to obtain the confidence interval for (μ_{steers}-μ_{sheep}), whereas the *t* test output did this calculation for you.

Advantages to the ANOVA approach are (i) it flags unusual values (standardized residuals outside the range (-2, +2)), and (ii) allows you to generate a residual plot and/or a means plot.

GenStat output for the automatic *t* test of the fine gravel data

Two-sample t-test

Variate: Fine_gravel
Group factor: Soil

Test for equality of sample variances

Test statistic F = 5.77 on 6 and 6 d.f.
Probability (under null hypothesis of equal variances) = 0.05

Step 1. Test for equality of variances

Summary

Sample	Size	Mean	Variance	Standard deviation	Standard error of mean
good	7	10.914	40.12	6.334	2.394
poor	7	3.943	6.95	2.636	0.996

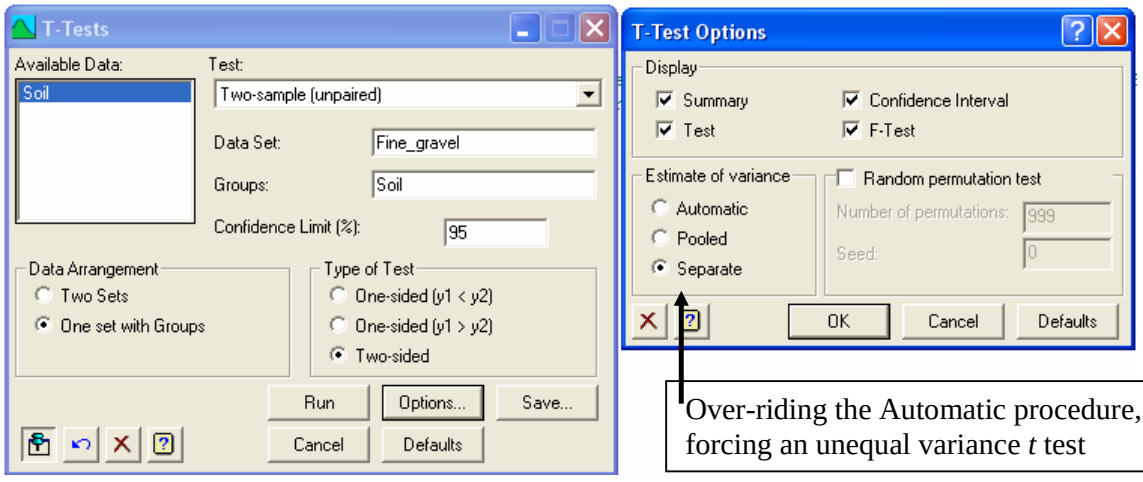
Difference of means: 6.971
Standard error of difference: 2.593

95% confidence interval for difference in means: (1.321, 12.62)

Test of null hypothesis that mean of Fine_gravel with Soil = good is equal to mean with Soil = poor

Test statistic t = 2.69 on 12 d.f.
Probability = 0.020

Step 2. Test for equality of means



Over-riding the Automatic procedure, forcing an unequal variance *t* test

Difference of means: 6.971
Standard error of difference: 2.593

95% confidence interval for difference in means: (0.9937, 12.95)

Test of null hypothesis that mean of Fine_gravel with Soil = good is equal to mean with Soil = poor

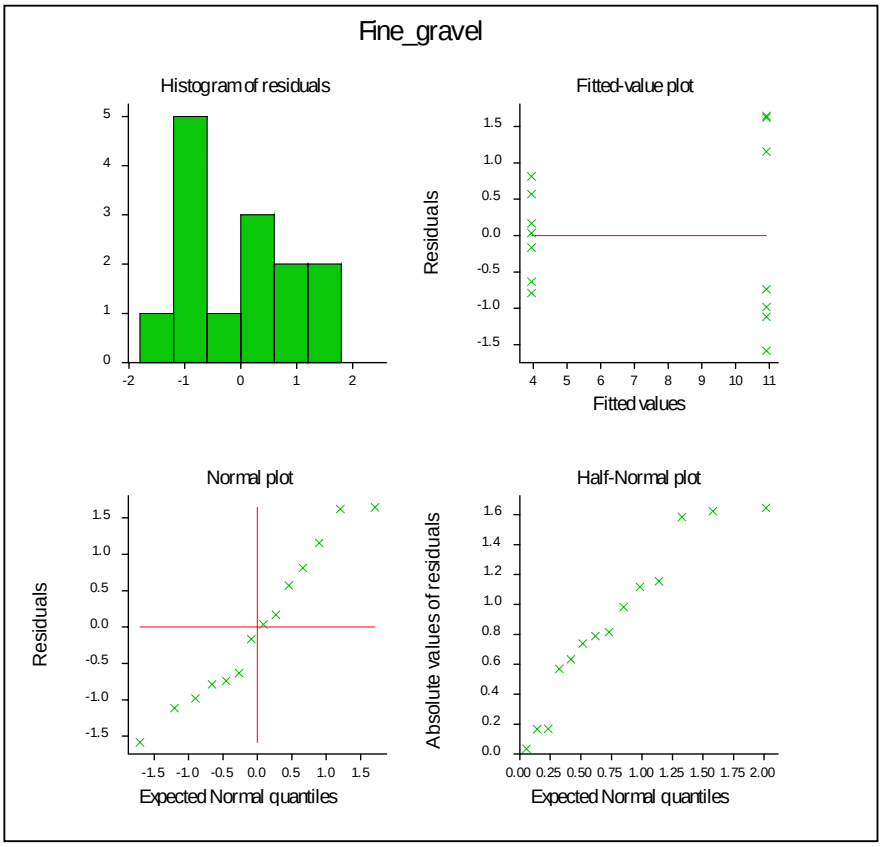
Test statistic t = 2.69 on approximately 8.02 d.f.
Probability = 0.028

Change to Step 2. Calculates approximate *df* for *t* test (8 instead of 12) and gives new *P*-value

Unpaired *t* test – example of unequal variances – Satterthwaite's approximate *t* test

Example 2 Fine gravel in soil, in percent (Steel and Torrie, page 107)

	Good soil	Poor soil
	5.9	7.6
	3.8	0.4
	6.5	1.1
	18.3	3.2
	18.2	6.5
	16.1	4.1
	7.6	4.7
mean	10.91	3.94
variance	40.12	6.95



Both means and variances in the two samples *appear* to be different. What statistical evidence is there that the mean percentage of fine gravel in the soil differs in the two soil types?

We first analysed the data via a one-way (no blocking) analysis of variance, and examined the residual plot. It is clear that the soil with the higher fitted value (obviously the good soil) has a larger visual scatter of residuals compared to that for the poor soil. This is a reflection of the different variances in the two samples.

An analysis in GenStat via a *t* test results in strong statistical evidence ($P = 0.020$) that the mean percentages of fine gravel differ. However, the test of equal variances is marginal. GenStat actually proceeds to use the standard unpaired *t* test because technically the *F* test does not fail ($P = 0.05$ to two decimals; it is actually 0.0509). We make three points.

- The *F* test depends heavily on normally distributed data, and percentages are unlikely to be normally distributed, so the *P*-value is somewhat unreliable.
- Failure to reject in this case is most likely to be caused by the low level of replication.
- We often make decisions about homogeneity of variance in more complex analyses of variance from an inspection of the standardized residual plot, rather than a formal test.

As mentioned previously, the default in GenStat for this test is to allow it to decide automatically what test to use for the means. To illustrate the approximate procedure, we over-rode GenStat by going into the **Options** menu, as shown. The change for an equally replicated experiment is only in the *df* of the *t* test (and hence in the *P*-value). Remember, it is not an exact t test. Here, the *df* used are closer to 6 than to 12, since the variances are quite different in the sample.

Paired *t* test – special case of a one-way treatment design (in randomised blocks)

Example 3 Sugar concentrations of nectar in half heads of red clover kept at different vapor pressures for eight hours (from Steel and Torrie, page 103)

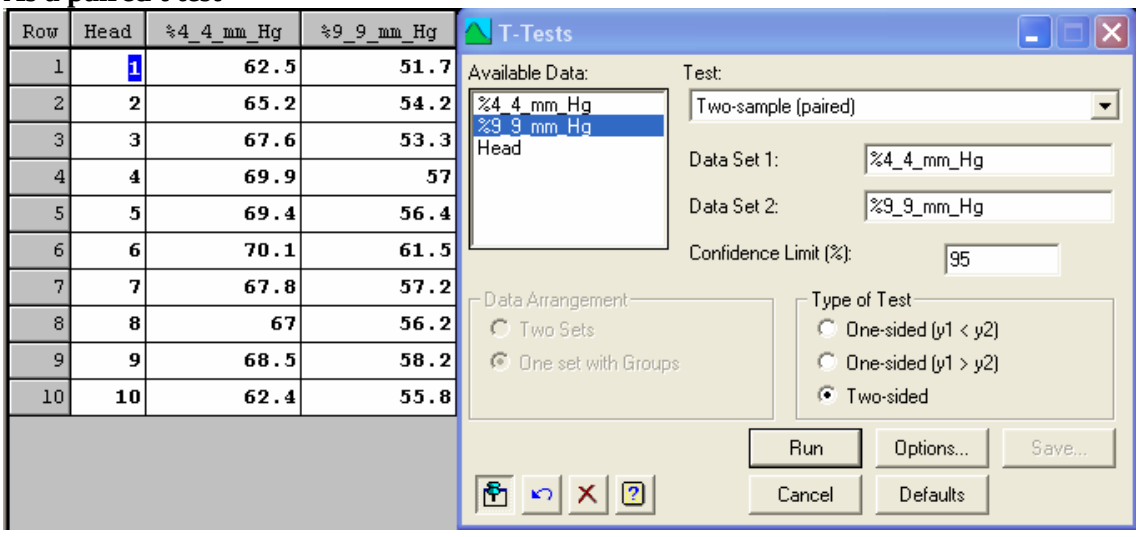
Head	4.4 mm Hg	9.9 mm Hg	difference
1	62.5	51.7	10.8
2	65.2	54.2	11.0
3	67.6	53.3	14.3
4	69.9	57.0	12.9
5	69.4	56.4	13.0
6	70.1	61.5	8.6
7	67.8	57.2	10.6
8	67.0	56.2	10.8
9	68.5	58.2	10.3
10	62.4	55.8	6.6
mean	67.04	56.15	10.89
sd	2.82	2.72	2.22

This example is quite different to the previous two examples. In this case, we cannot place the 10 concentrations in any order in each column: they are *paired*. The heads of red clover are divided into half heads; one is randomly subjected to a vapor pressure of 4.4 mm Hg, the other to a vapor pressure of 9.9 mm Hg. Each head of clover is likely to vary in its sugar concentration, and the only way to remove this variation is to take differences, and analyse these in a one sample *t* test.

When we have more than two treatments in an experiment that is blocked in some way, then we need to analyse the data using an ANOVA *F* test, setting up a “block” factor as well as a “treatment” factor.

Firstly, in GenStat, paired t test data must be set up in separate columns for separate treatments.

As a paired t test



One-sample t-test

Variate: Y[1].

Sample	Size	Mean	Variance	Standard deviation	Standard error of mean
VP_4_4-VP_9_9	10	10.89	4.914	2.217	0.7010

95% confidence interval for mean: (9.304, 12.48)

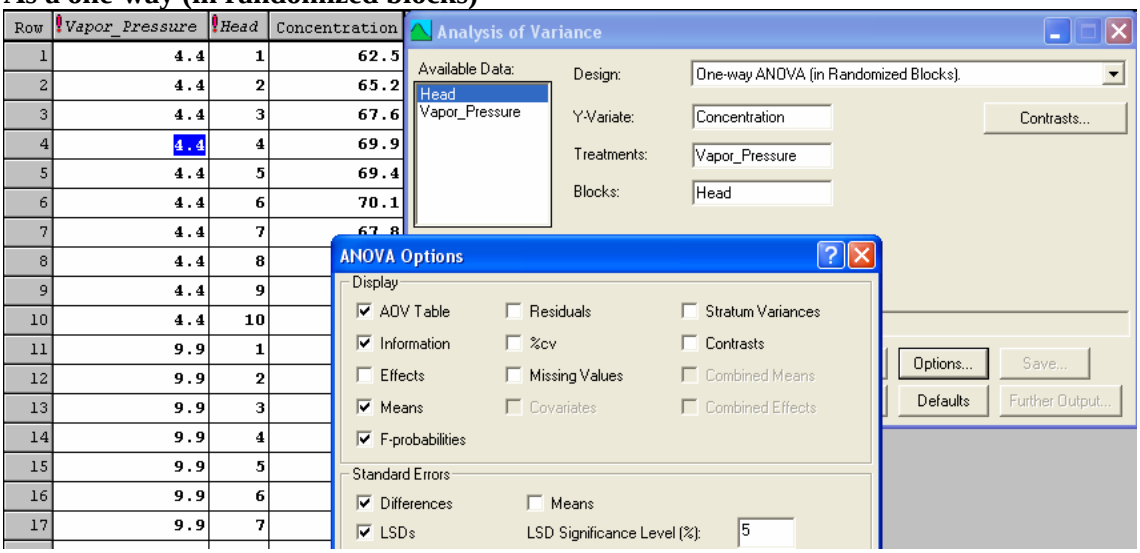
Test of null hypothesis that mean of VP_4_4-VP_9_9 is equal to 0

Test statistic t = 15.53 on 9 d.f.
Probability < 0.001

There is strong statistical evidence (P<0.001) that the mean sugar concentration of nectar differs in heads of red clover kept at different vapor pressures for eight hours. The best estimated mean difference is 10.89%, and we are 95% confident that the true difference lies between 9.30% and 12.48%.

To analyse the data via ANOVA or regression, we must stack the data, and provide a factor column to identify the various head (acting as blocks).

As a one-way (in randomized blocks)



Notice in the output that GenStat organizes the ANOVA into the two strata for this experiment. Individual heads form the top stratum, and since these are not replicated (there is no other “head 1” or “head 2” etc), there is no *P*-value for this variance ratio. The second stratum is the “Heads.Units” stratum, that is, the half head put into one of two vapor pressure treatments (at random). These are replicated in a balanced way (each treatment occurs once in each block).

Analysis of variance

Variate: Concentration

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Head stratum	9	116.114	12.902	5.25	
Head.*Units* stratum					
Vapor_Pressure	1	592.960	592.960	241.32	<.001
Residual	9	22.115	2.457		
Total	19	731.189			

Message: the following units have large residuals.

Head 10 *units* 1	-2.14	s.e.	1.05
Head 10 *units* 2	2.14	s.e.	1.05

Tables of means

Variate: Concentration

Grand mean 61.60

Vapor_Pressure	4.4	9.9
	67.04	56.15

Standard errors of differences of means

Table	Vapor_Pressure
rep.	10
d.f.	9
s.e.d.	0.701

Least significant differences of means (5% level)

Table	Vapor_Pressure
rep.	10
d.f.	9
l.s.d.	1.586

Estimated stratum variances

Stratum	variance	effective d.f.	variance component
Head	12.902	9.000	5.222
Head.*Units*	2.457	9.000	2.457

Again, notice

- the relationship between *t*-value of 15.53, and the *F*-value of 241.32 ($15.53^2 = 241.32$)
- the same *P*-value (*P*<0.001, though it is hard to see the similarity, *P* is so small)
- the mean difference is 67.04 - 56.15 = 10.89 ± 1.586, giving rise to the same confidence interval

Remember that a *t* test is just a special case of regression. There are two models to consider when testing whether the vapor pressure treatment effect is zero.

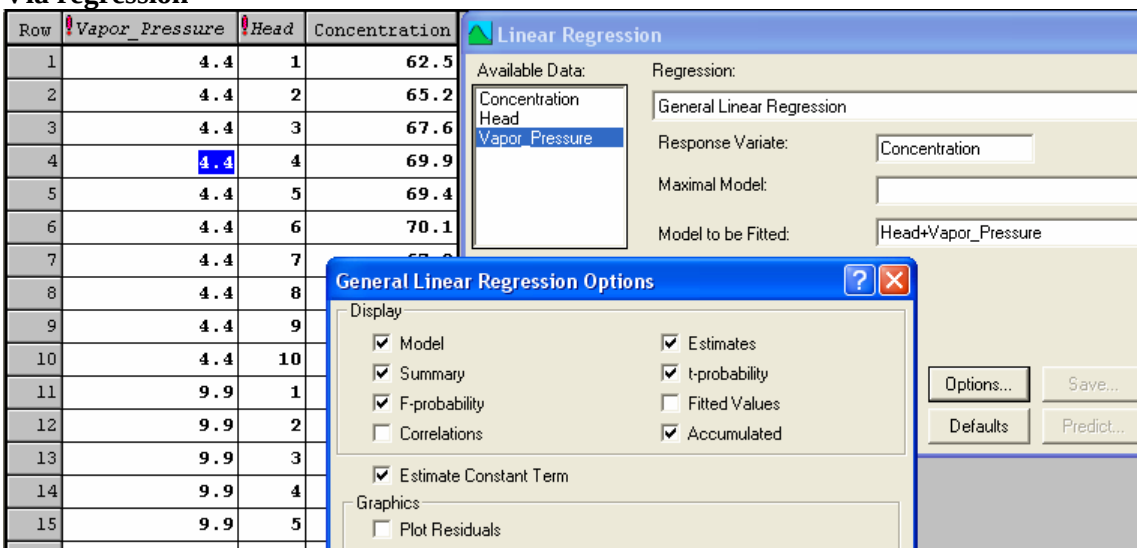
Maximal model
Sugar concentration = *overall mean* + *Head* effect + *Vapor pressure* effect + *Error*

dropped

Reduced model
Sugar concentration = *overall mean* + *Head* effect + *Error*

In GenStat's **General Linear Regression Option** menu, the effect of blocks (*Heads*) and treatments (*vapor pressure*) can be assessed by turning on **Accumulated**.

Via regression



Regression analysis

Response variate: Concentration
Fitted terms: Constant + Head + Vapor_Pressure

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	10	709.08	70.908	28.86	<.001
Residual	9	22.11	2.457		
Total	19	731.19	38.484		

Percentage variance accounted for 93.6
Standard error of observations is estimated to be 1.57.

Message: the following units have large standardized residuals.

Unit	Response	Residual
10	62.40	-2.04
20	55.80	2.04

Estimates of parameters

Parameter	estimate	s.e.	t(9)	t pr.
Constant	**62.55**	**1.16**	**53.80**	**<.001**
Head 2	2.60	1.57	1.66	0.132
Head 3	3.35	1.57	2.14	0.061
Head 4	6.35	1.57	4.05	0.003
Head 5	5.80	1.57	3.70	0.005
Head 6	8.70	1.57	5.55	<.001
Head 7	5.40	1.57	3.44	0.007
Head 8	4.50	1.57	2.87	0.018
Head 9	6.25	1.57	3.99	0.003
Head 10	2.00	1.57	1.28	0.234
Vapor_Pressure 9.900	**-10.890**	**0.701**	**-15.53**	**<.001**

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Head	1
Vapor_Pressure	4.400

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Head	9	116.115	12.902	5.25	0.011
+ Vapor_Pressure	**1**	**592.961**	**592.961**	**241.32**	**<.001**
Residual	9	22.114	2.457		
Total	19	731.190	38.484		

The default model produces a **Constant** (the mean for vapor pressure 4.4) and a mean difference of -10.890, labeled **Vapor_Pressure 9.900**. This is highly significant, with a *t*-value of -15.53, the same (apart from sign) as was produced by the paired *t* test. The **Accumulated** analysis is the RCBD ANOVA, though it is an application of the general technique for comparing a maximal and reduced model outlined on page 15.

Notice also that 1.16 will be the s.e.m. and 0.701 the s.e.d..

Overview of analysis of variance

We begin by having another look at the analysis of variance for a one-way treatment design, firstly for the unblocked analysis and then for the randomised block analysis.

One-way treatment design, (no blocking)

ANOVA for one-way (no blocking)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Treatment	1	81.927	81.927	11.18	0.007
Residual	11	80.584	7.326		
Total	12	162.511			

rep	7	6
mean	56.21	61.25
variance	9.015	5.299

Firstly, the sample variance of the 13 data values is 13.534. In the ANOVA table, this is the Total MS, and equals 162.511/12. GenStat does not complete this entry in the table (except in the regression menu). See page 12 for an earlier mention of this relationship.

We pointed out previously (page 12) that the Residual SS (80.584) is the sum of squared residuals, (defined as observed – fitted). The Residual MS turns out to be the pooled variance estimate, that is, a weighted average of the individual treatment variances, with weights equal to the individual degreed of freedom of the sample variances:

$$7.326 = (6\times 9.015+5\times 5.299)/(6+5)$$

The Treatment MS is calculated as follows. Assuming common variances, if there *are* no treatment mean differences, the data from both treatments come from the same population. In that case, the i^{th} treatment mean is an estimate of σ^2/n_i. Accordingly, a weighted variance of these sample means, under the null hypothesis that the means are equal, will estimate σ^2. It also turns out that the Treatment MS and Residual MS are independent.

Thus, under the null hypothesis that the means are equal, the ratio

F = Treatment MS / Residual MS is

is distributed as an F variable with 1, 11 degrees of freedom.

For t treatments, the situation is no different. The mean squares are interpreted as follows.

To summarise:

ANOVA for one-way (no blocking)

Source of variation	**d.f.**	**m.s.**
Treatments	t-1	Weighted variance of treatment means
Residual	N-t	Pooled estimate of variance
Total	N-1	sample variance of the data

One-way treatment design, (in randomised blocks)

This is actually easier to explain, because the design is balanced.

Analysis of variance

Variate: Concentration

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Head stratum	9	116.114	12.902	5.25	
Head.*Units* stratum					
Vapor_Pressure	1	592.960	592.960	241.32	<.001
Residual	9	22.115	2.457		
Total	19	731.189			

- The **Total MS** is still the sample variance of all the data. Thus, 731.189/19 = 38.484.
- The **Treatment MS** is a weighted variance of the treatment means, with weights the number of blocks. The two vapor pressure means are 67.04 and 56.15. Each is based on 10 replicates. Thus, the **Treatment MS** is 10×sample variance of (67.04, 56.15) = 592.96.
- The **Block MS** is a weighted variance of the block means, with weights the number of treatments. There are 10 block means, (57.1, …, 59.1) and each is based on two observations, one from each treatment. Thus, the **Block MS** is 2×sample variance of (57.1, …, 59.1) = 12.902.
- The **Block SS** is still the sum of squares of the residuals. The **Block MS** is a Treatment × Block interaction: it measures the failure of the treatments to respond alike in each block. More of interactions later.

To summarise:

ANOVA for one-way (in randomised blocks)

Source of variation	**d.f.**	**m.s.**
Blocks	b-1	Weighted variance of block means, with weights t
Treatments	t-1	Weighted variance of treatment means, with weights b
Residual	(b-1)(t-1)	Pooled estimate of variance
Total	bt-1	sample variance of the data

More complex balanced designs have similar structures.

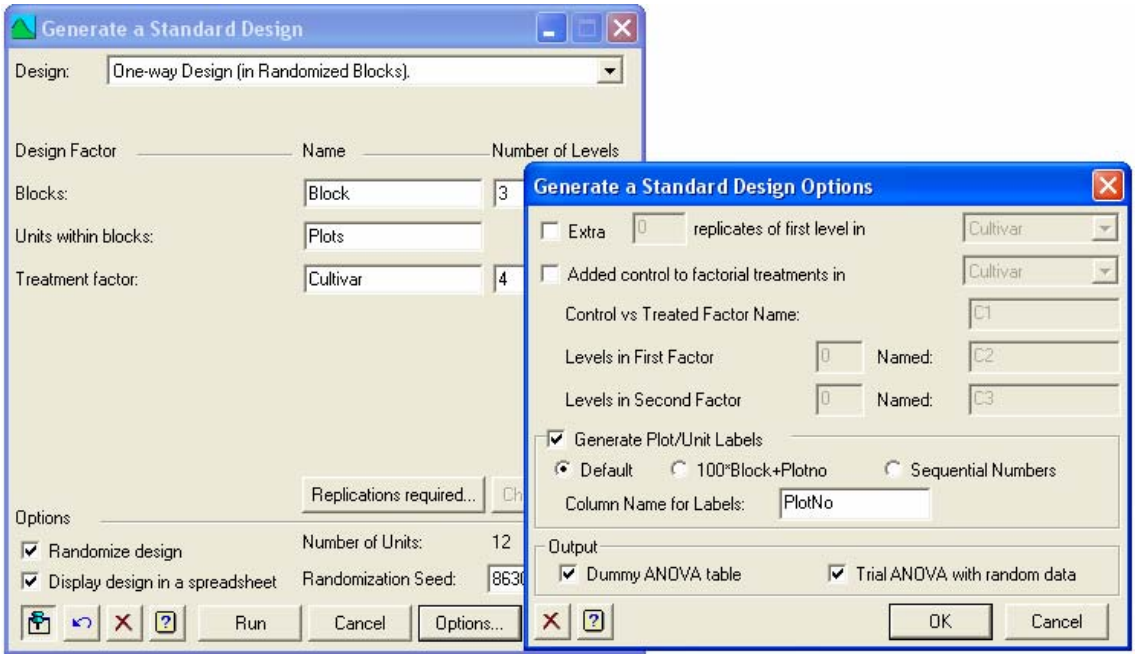
GenStat's Design menu

GenStat has the ability to generate a random design for you. Most of the common designs are available, including incomplete factorial designs, and designs with additional replication for (say) a control treatment.

The design is a blueprint for conducting the experiment. It assigns the treatments to experimental units randomly. At the end of the experiment, add your data to the spreadsheet and, at least for normally or log-normally distributed data, all you need to do is point and click to have the analysis performed.

Firstly, let's illustrate the method with a simple one-way treatment design with four cultivars of oats (Vicland (1), Vicland (2), Clinton and Branch), set out in three randomised blocks in the field.

Use **Stats** > **Design** > **Generate a Standard Design**. Choose **One-way Design (in Randomized Blocks)**. Name the treatment factor and (optionally) the units to which the treatments are to be applied. Indicate the number of blocks and levels. In **Options**, you can **Trial ANOVA with random data**: this produces an analysis of random data, scaled so that the Residual MS is always 1.



GenStat creates a spreadsheet and outputs the analysis. Notice the following:

- The first column is a key to the plots in the field. The second integer is the block number, the first integer the plot number in that block. GenStat will use as many digits as required. Thus, for a design with 12 treatments in 3 blocks, the first two columns will indicate plots and the final column the block.

Analysis of variance

Variate: _Rand_

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block stratum	2	13.973	6.986	6.99	
Block.Plots stratum					
Cultivar	3	6.448	2.149	2.15	0.195
Residual	6	6.000	**1.000**		
Total	11	26.421			

Tables of means

Variate: _Rand_

Grand mean 21.69

Cultivar	1	2	3	4
	22.94	21.05	21.36	21.41

Standard errors of means

Table	Cultivar
rep.	3
d.f.	6
e.s.e.	0.577

Least significant differences of means (5% level)

Table	Cultivar
rep.	3
d.f.	6
l.s.d.	1.998

Stratum standard errors and coefficients of variation

Variate: _Rand_

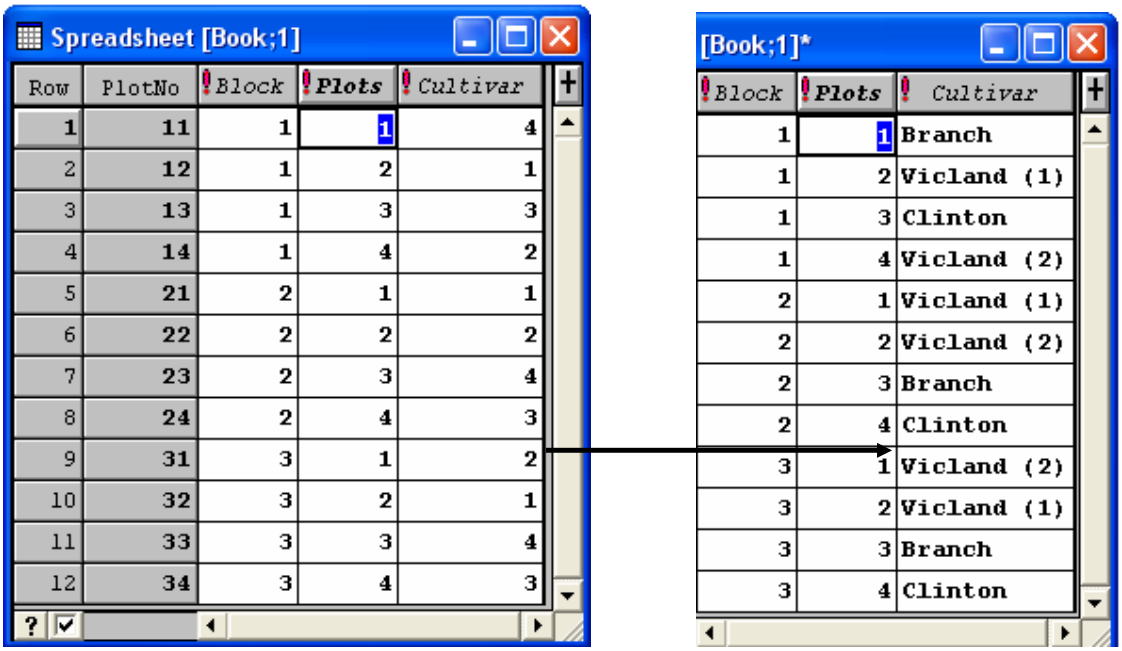
Stratum	d.f.	s.e.	cv%
Block	2	1.322	6.1
Block.Plots	6	1.000	4.6

Diagrammatic field plan

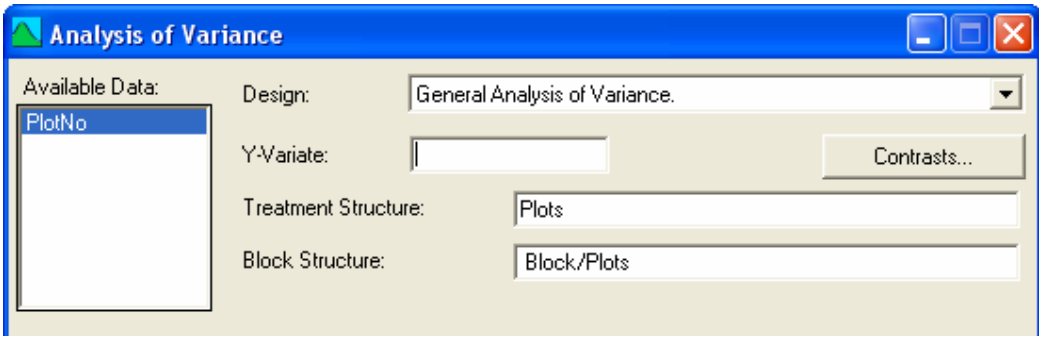
	Plot 1	**Plot 2**	**Plot 3**	**Plot 4**
Block 1	Branch	Vicland (1)	Clinton	Vicland (2)
Block 2	Vicland (1)	Vicland (2)	Branch	Clinton
Block 3	Vicland (2)	Vicland (1)	Branch	Clinton

GenStat will always generate a factor column for every stratum in the experiment. We have seen that for a block design, blocks, while unreplicated, form one stratum, and plots (which provide the replication for treatment comparisons) form the second stratum.

- The final column indicates which treatment to use in each plot in the field. This is the field plan. It is preferable at this stage to edit the column attributes (**F9** is the shortcut). In this case, change the 1, 2, 3, 4 for cultivars to their actual names. These names are then part of your statistical analysis once the data become available.



Having entered the experimental data into the spreadsheet, you can simply right click on the spreadsheet, select **Analysis** > **Analysis of Variance**. The necessary structure is completed for you: your only task is to choose which variate you want analysed this way.



The analysis will be like the one shown (which is for GenStat's random, scaled data).

Before proceeding to other designs, we need to discuss the shortcuts that GenStat uses for treatment and block structures.

Basic rules for expansion of formulae

The principle underlying a correct formulation of the blocking structure is to properly declare every type of experimental unit. For each stage of randomization a new experimental unit is created. Since the analysis exactly mimics the way the experiment is conducted in the field, a new stratum is created in the ANOVA table.

GenStat, however, allows you to omit the lowest level of randomization on the Block Structure line. If you omit the lowest level stratum in Linear Mixed Models (REML), GenStat (tells you that it) adds it to the model.

Block and treatment structures can be simplified using certain rules and operators.

Terms within parentheses are evaluated first. Otherwise, the order that GenStat uses to evaluate formulae which include operators is as follows (see GenStat Reference Manual):
1. .
2. //
3. /
4. *
5. + - -/ -*

Generally we use **. / * +** and **-**. Formulae involving a mixture of operators of rank (5) are computed left to right.

Let A, B, C … represent the names of factors and L and M a set of terms in a formula.

Rule 1	L.M	Sum of all pairwise combinations of terms in L with terms in M using the dot operator. For example: (A+B).(C+D.E) is the same as A.C + B.C + A.D.E + B.D.E
Rule 2	L*M	L+M+L.M. For example: A*C is the same as A + C + A.C (A+B)*C is the same as A + B + C + A.C + B.C
Rule 3	L/M	L+*L*.M where *L* is a term formed by combining all terms in L with the dot operator. For example: A/C is the same as A + A.C (A+B)/(C+D.E) is the same as A + B + A.B.C + A.B.D.E
Rule 4	L-M	L without any terms that appear in M. For example: (A+B)-(A+C) is the same as B A*B*C-A.B.C is the same as A+B+C+A.B+A.C+B.C

For an experiment with replication but no blocks, there should be a factor indexing the units that form replicates (plots, pots, animals, …). If there is sub-sampling within the replicate, provide an additional column to index those units. It is better to use Plot 1, 2, 3, … *p* rather than Treatment 1 (Plot 1, 2, 3), Treatment 2 (Plot 1, 2, 3) and so on. The **Block Structure** for this design can be left blank (as mentioned in paragraph 2 above), or written as **Plot** with the first method of indexing plots, or **Treatment.Plot** with the second. For the **Random Model:** in Linear Mixed Models (REML), there is an occasional advantage one way or another.

Completely randomized design (CRD), or one-way design (no blocking)

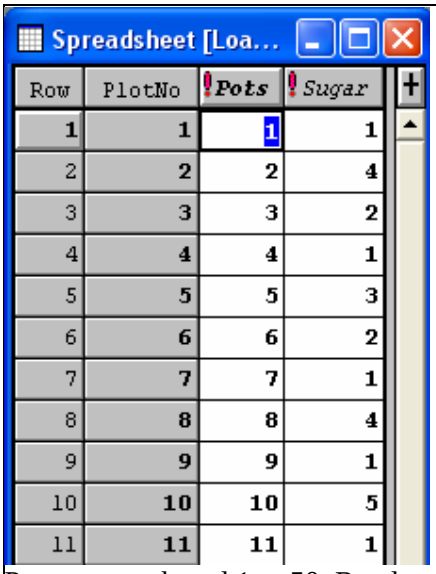
The data are from an experiment in plant physiology. Lengths of pea sections grown in tissue culture with auxin present were recorded. The purpose of the experiment was to test the effects of various sugar media on growth as measured by length.

Treatment structure: Single factor with 5 levels: sugar treatments (including a control)
Block Structure: None: 10 replicates for all treatments

Example 4. The effect of different sugars on length, in ocular units (× 0.114 = mm), of pea sections grown in tissue culture with auxin present (Sokal & Rohlf 3rd Ed. page 218)

Replicate	*Control*	*2% glucose added*	*2% fructose added*	*1% glucose + 1% fructose added*	*2% sucrose added*
1	75	57	58	58	62
2	67	58	61	59	66
3	70	60	56	58	65
4	75	59	58	61	63
5	65	62	57	57	64
6	71	60	56	56	62
7	67	60	61	58	65
8	67	57	60	57	65
9	76	59	57	57	62
10	68	61	58	59	67

In this experiment we have 50 pots (labelled 1 to 50) with no blocking required. The pots are placed in a growth chamber, and the treatments randomized to the pots (eg using GenStat's **Design** menu; notice that GenStat creates a factor column Pots, with levels 1 to 50):

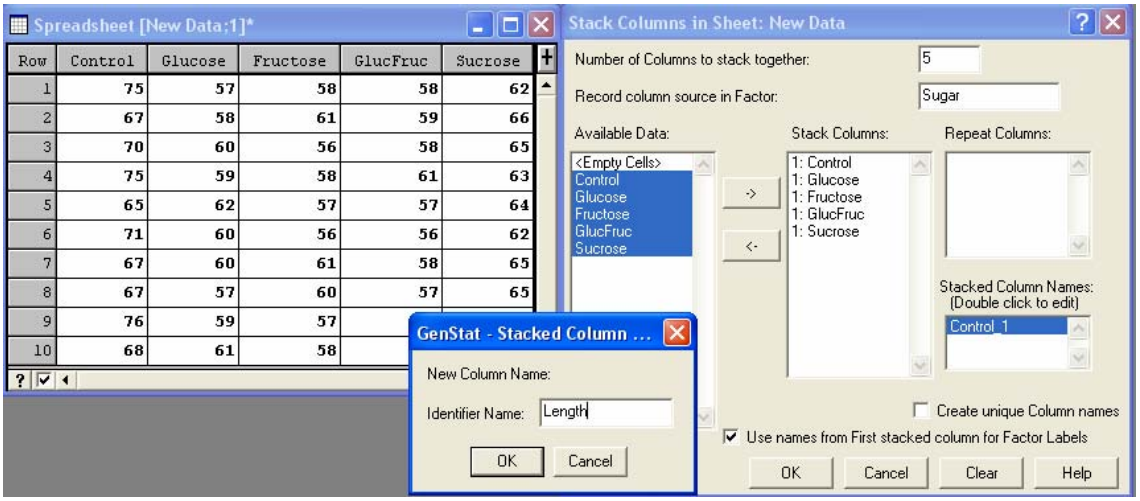


Pots are numbered 1 to 50. Random allocation of the *Control* treatment is shown

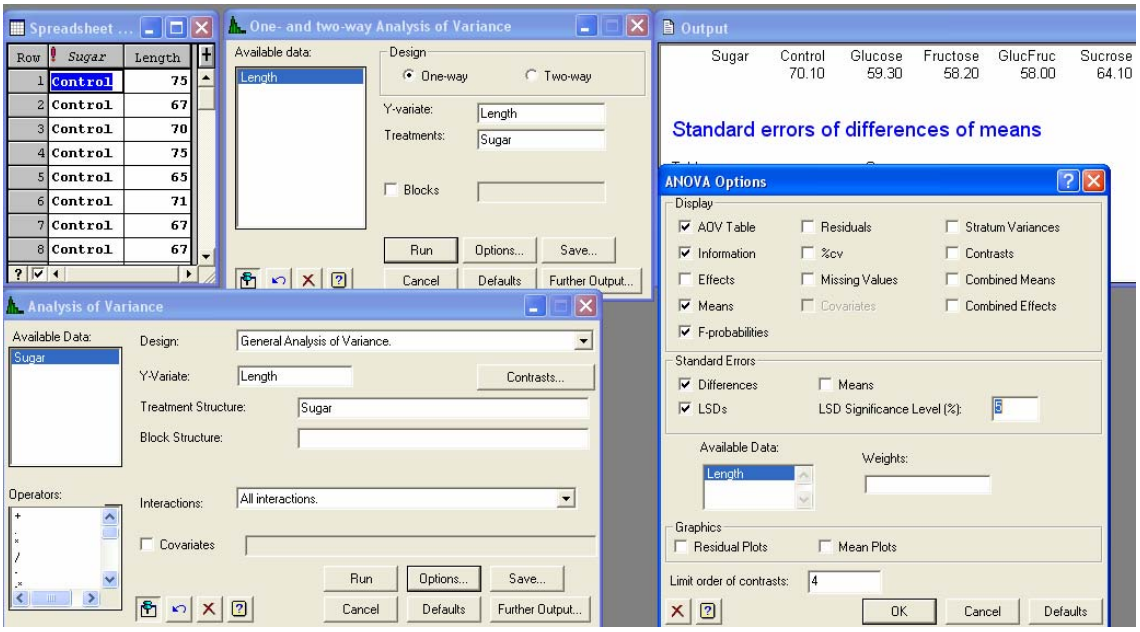
1 Control	2	3	4 Control	5
6	7 Control	8	9 Control	10
11 Control	12	13	14	15
16	17	18	19 Control	20 Control
21	22	23 Control	24	25
26 Control	27	28	29	30
31	32	33	34	35
36	37	38	39	40
41	42	43	44 Control	45
46	47	48	49	50

Data and analysis in GenStat

We firstly stack the data into a *variate* labelled `Length`, and create an identifier *factor* for the `Sugar` treatments. It is much more sensible to use treatment **labels** or treatment **levels** where possible. (Note that this can be done while stacking the data.) GenStat will always use labels or levels in its output. You can see that GenStat replaces the identifying numbers with actual labels.



Choose **One- and Two-way** to obtain the basic CRD ANOVA; alternatively, choose **General Analysis of Variance** and use Pots as the **Block Structure**. Note that GenStat allows the final stratum to be omitted, so you can, for this design, leave the **Block Structure** blank. Notice that we selected to output the 5% l.s.d. values. The s.e.(difference) is set as the default output; we could also have chosen to obtain the s.e.(mean). The (standardised) residual plot can be drawn once the analysis is obtained: return to the **Analysis of Variance** window, select **Further Output**, **Residual Plots** and **Standardized**.



Analysis of variance

Variate: Length

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Sugar	4	1077.320	269.330	49.37	<.001
Residual	45	245.500	5.456		
Total	49	1322.820			

Message: the following units have large residuals.

units 5	-5.10	s.e.	2.22
units 9	5.90	s.e.	2.22

Tables of means

Variate: Length

Grand mean 61.94

Sugar	Control	Glucose	Fructose	GlucFruc	Sucrose
	70.10	59.30	58.20	58.00	64.10

Standard errors of means

Table	Sugar
rep.	10
d.f.	45
e.s.e.	0.739

Standard errors of differences of means

Table	Sugar
rep.	10
d.f.	45
s.e.d.	1.045

Least significant differences of means (5% level)

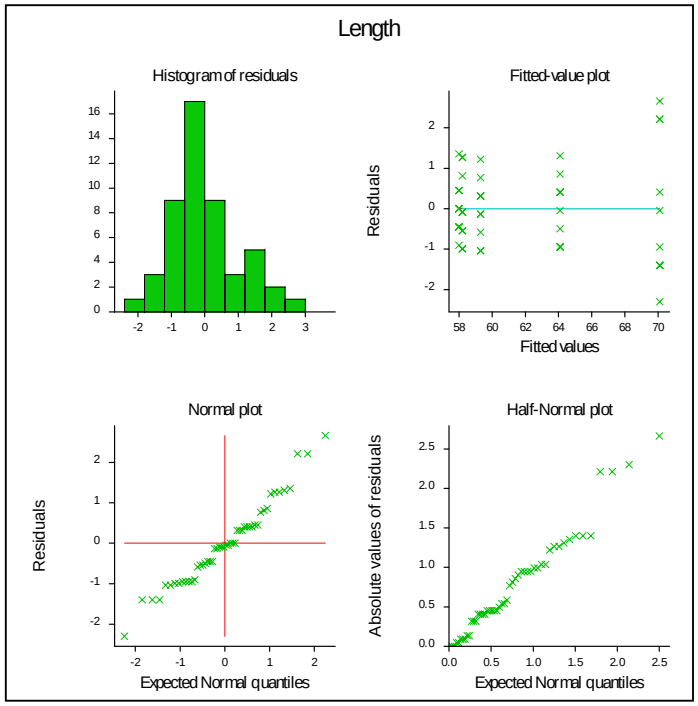
Table	Sugar
rep.	10
d.f.	45
l.s.d.	2.104

Notice:

- 5.456 is the average of the sample variances 15.878, 2.678, 3.511, 2.000, 3.211, each with (10-1) = 9 *df.*
- 269.33 is the weighted sample variance of the sugar means 70.1, 59.3, 58.2, 58.0, 64.1. Since an unweighted variance would (if the population treatment means were all equal) estimate $\sigma^2/10$, the Sugar MS is 10 × sample variance.

Before discussing the analysis in any more detail, we should inspect the (standardized) residual plot.

There are problems with this analysis. The standardised residual plot uncovers a large variance for the data in the treatment with the largest fitted value, which on inspection is the `Control` treatment. This is common in agricultural trials, and leads to special ways of analysing the data.



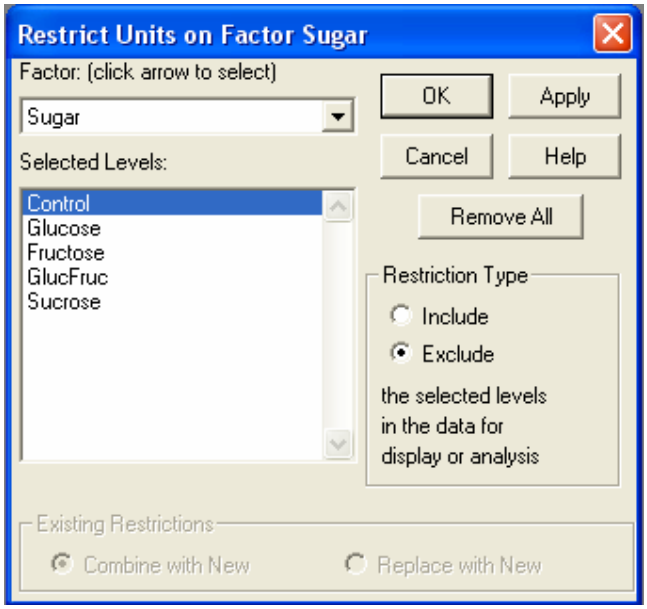
Sometimes it is possible to find a transformation that overcomes the problem, especially if the problem is one of fanning. Fanning often indicates log-normal (rather than normal) data, or data for which the variance increases as mean^2.

In this case, untreated data simply behave differently to treated data in terms of variability. One possibility is to separate out the treated and control data, and analyse these sets of data separately. The variance for the untreated data is very large (15.878 with 9 *df*) compared to the variances for the treated data (whose average is 2.850 with 4 × 9 = 36 *df*). Keeping the treated data allows fair comparisons among the four sugar treatments. If one really wanted to compare the control mean with one of the four sugar means, a variation of Satterthwaite's approximate *t* test (see page 39) can be used.

Alternatively, a Linear Mixed Model can be used that allows two variances, one for untreated data and another for treated data. Both tests (tests of equality of the four sugar treatment means, test of the mean of the untreated data versus the mean of the treated data) are done in the one analysis.

Restricting the analysis to a subset of treatments

There are several ways to do this, but the easiest is click inside the spreadsheet, then select **Spread** > **Restrict/Filter** > **To Groups (factor levels)**, select the Control treatment and **Exclude**.



Now click back into the **Analysis of Variance** box and click on **OK** to re-run the analysis. The sugar means are the same (as they must be) but the Control mean is left blank. The Residual MS is now only 2.850 instead of the earlier 5.456, representing a much fairer variance estimate for comparing the 4 sugar means (resulting in a reduced l.s.d. value of 1.531 instead of the earlier 2.104).

Analysis of variance

Variate: Length

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Sugar	3	245.000	81.667	28.65	<.001
Residual	36	102.600	2.850		
Total	39	347.600			

Tables of means

Grand mean 59.90

Sugar	Control	Glucose	Fructose	GlucFruc	Sucrose
		59.30	58.20	58.00	64.10

Standard errors of differences of means

Table	Sugar
rep.	10
d.f.	36
s.e.d.	0.755

Least significant differences of means (5% level)

Table	Sugar
rep.	10
d.f.	36
l.s.d. 1.531	

To compare the Control mean (which has an estimated standard deviation of s_1 = 3.985 with 9 *df*) with one of the 4 sugar means (which has an estimated standard deviation of $s_p = \sqrt{2.850}$ = 1.688 with 36 *df*) is achieved by an extension of Satterthwaite's test.

Approximate t test of $\mu_{untreated} = \mu_{sucrose}$

Difference in means = 70.1 – 64.1 = 6.0. $sed = \sqrt{\frac{s_1^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{1.873}$ = 1.368. Hence,

t_{obs} = 6.0/1.368 = 4.38.

The degrees of freedom are calculated from a formula modified using the formula on page 34.

$$df = \left[\frac{\left(\frac{s_1^2}{n_1} + \frac{s_p^2}{n_2} \right)^2}{\frac{\left(s_1^2 / n_1 \right)^2}{n_1 - 1} + \frac{\left(s_p^2 / n_2 \right)^2}{df \text{ of } s_p^2}} \right] = 12.42.$$

There is strong statistical evidence (P<0.001) that the control and sucrose means are different.

Two-way design (no blocking) with subsamples

Mint plants were assigned at random to pots, 4 plants per pot, 18 pots in all and grown in a nutrient solution. Three pots were randomly assigned to one of six treatment combinations, as follows. All pots were randomly located during the time spent at either 8, 12 or 16 hours of daylight. Each group of pots was completely randomised within low- or high-temperature greenhouses during the time spent in darkness. Individual plants stem lengths were measured after one week.

Example 5 One week stem lengths (cm, Steel and Torrie pages 153-9)

Temperature																		
High										Low								
Hours of Daylight										Hours of Daylight								
8			12			16				8			12			16		
pot			pot			pot				pot			Pot			pot		
1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	
3.5	2.5	3.0	5.0	3.5	4.5	5.0	5.5	5.5	8.5	6.5	7.0	6.0	6.0	6.5	7.0	6.0	11.0	
4.0	4.5	3.0	5.5	3.5	4.0	4.5	6.0	4.5	6.0	7.0	7.0	5.5	8.5	6.5	9.0	7.0	7.0	
3.0	5.5	2.5	4.0	3.0	4.0	5.0	5.0	6.5	9.0	8.0	7.0	3.5	4.5	8.5	8.5	7.0	9.0	
4.5	5.0	3.0	3.5	4.0	5.0	4.5	5.0	5.5	8.5	6.5	7.0	7.0	7.5	7.5	8.5	7.0	8.0	

This design is slightly complex, in that half the pots have a restricted randomization for the time spent in one of the two greenhouses, each set at a different temperature. Ignoring that problem, it is clear that pots form replicates for the six treatment combinations: a pot containing 4 plants is moved to a random daylight position and a random position in a greenhouse; the 4 plants form sampling units.

Treatment Structure

You need to supply two factor columns, properly labeled, to identify the six **Temperature** and **Light** treatment combinations applied to each pot. The **Treatment Structure** is then **Temperature** + **Light** + **Temperature.Light**. By the Rule 2 simplifies to **Temperature*Light**.

Block Structure

Choice 1

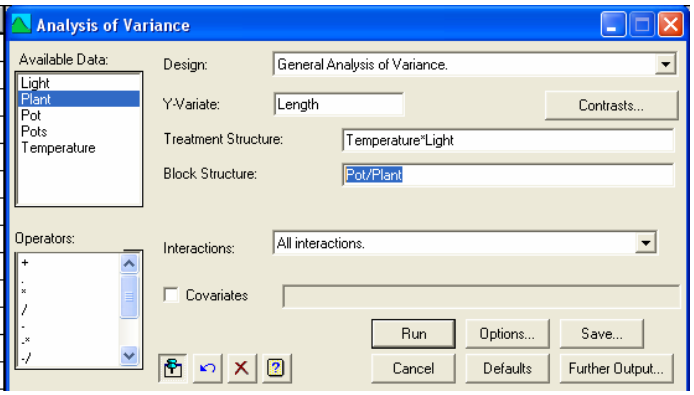
Earlier, we recommended that the replicates be numbered from 1 to the total number. There are 18 pot replicates, and in our spreadsheet we called this column **Pots**. Plants in pots are samples. There are two strata, and hence the **Block Structure** is **Pots+Pots.Plant**. By Rule 3 this simplifies to **Pots/Plant**. GenStat also allows the final stratum's factor to be omitted, so **Pots** is also permissible.

Choice 2 (not recommended)

Steel and Torrie, however, used 1, 2, 3 for each treatment combination. If you decide to use this numbering system, then the **Block Structure** *cannot* be **Pot/Plant**: as mentioned, this expands to **Pot+Pot.Plant**, and GenStat will assume that Pot #1 in every treatment is a block. Rather, you need to use **Pot.Treatment/Plant**, which expands to **Pot.Treatment + Pot.Treatment.Plant**. Here, **Treatment** is a factor that enumerates all six treatments and **Pot** has levels 1, 2, 3. We don't have such a treatment factor column, so you would need to **Insert** a new column and **Fill** this column from 1 to 6, each number repeated nine times. The analysis is identical to that obtained in *Choice 1*.

Analysis of Two-way Design (no Blocking) with subsamples

Row	Pots	Plant	Light	Temperature	Length
1	1	1	8	High	3.5
2	1	2	8	High	4
3	1	3	8	High	3
4	1	4	8	High	4.5
5	2	1	8	High	2.5
6	2	2	8	High	4.5
7	2	3	8	High	5.5
8	2	4	8	High	5
9	3	1	8	High	3
10	3	2	8	High	3
11	3	3	8	High	2.5
12	3	4	8	High	3
13	4	1	12	High	5



Analysis of variance

Variate: Length

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Pots stratum					
Temperature	1	151.6701	151.6701	70.45	<.001
Light	2	22.2986	11.1493	5.18	0.024
Temperature.Light	**2**	**5.6736**	**2.8368**	**1.32**	**0.304**
Residual	12	25.8333	2.1528	2.30	
Pots.Plant stratum	54	50.4375	0.9340		
Total	71	255.9132			

…

Tests these means

Tables of means

Variate: Length

Grand mean 5.78

Temperature	High	Low
	4.33	7.24

Light	8.	12.	16.
	5.50	5.29	6.56

Temperature	Light	8.	12.	16.
High		3.67	4.12	5.21
Low		7.33	6.46	7.92

Standard errors of differences of means

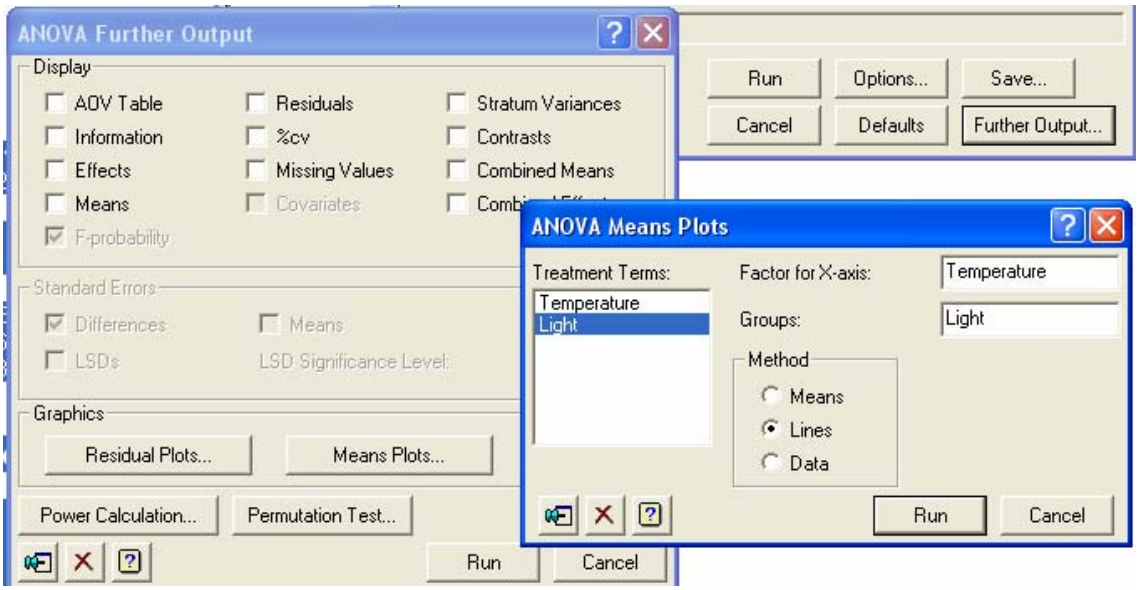
Table	Temperature	Light	Temperature Light
rep.	36	24	12
d.f.	12	12	12
s.e.d.	0.346	0.424	0.599

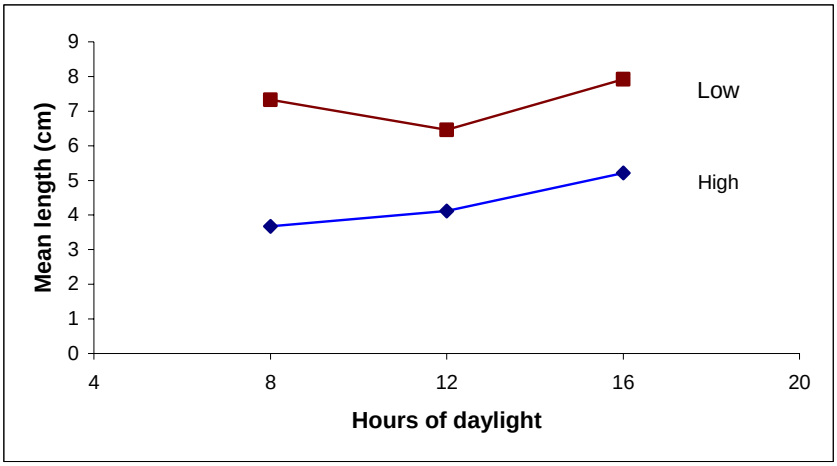
Least significant differences of means (5% level)

Table	Temperature	Light	Temperature Light
rep.	36	24	12
d.f.	12	12	12
l.s.d.	0.753	0.923	1.305

When interpreting this analysis, it is important to interpret the interaction first (for more complex designs, from highest-order interaction backwards). A two-way interaction tests whether any change in the response of the plant to temperature is consistent for both high and low temperatures. Thus, it examines the response to temperature in the following table. The response is best plotted (**Further Output** > **Means Plot**).

	Hours of light		
Temperature	8	12	16
High	3.67	4.12	5.21
Low	7.33	6.46	7.92





The responses are parallel within statistical variation (P = 0.304). Hence, attention can focus on the average effect of temperature, as well as the average effect of light. These are known as *main effects*. Both are strongly significant – see the ANOVA table.

Interest focuses on how much variation is there from plant to plant (the sampling variance) as opposed to pot to pot variation. Note that each of 6 treatments provides (3-1) = 12 residual df for estimating σ^2.

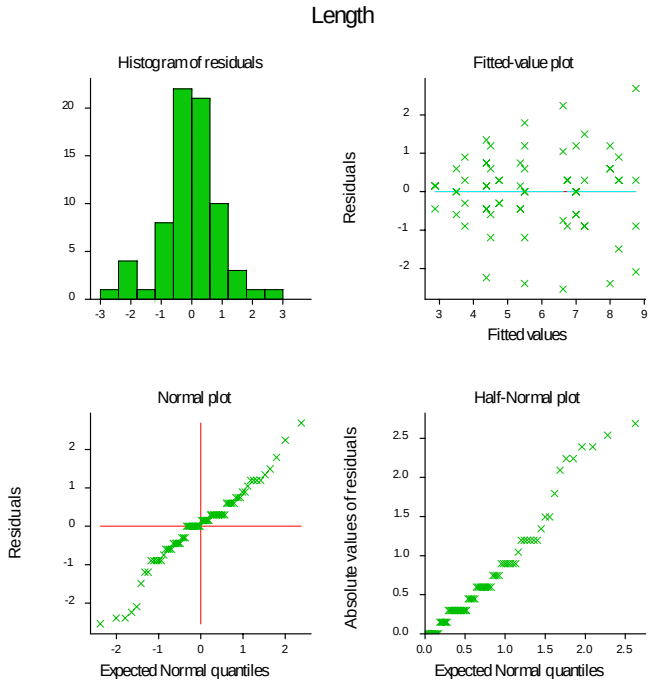
Estimates of the sampling and experimental variances are obtained by clicking on **Stratum Variances** in **Options** prior to running the analysis. The output is the following. There is three times more variation between plants in a pot than between pots.

Estimated stratum variances

Variate: Length

Stratum	variance	effective d.f.	variance component	
Pots	2.153	12.000	0.305	pot to pot variance
Pots.Plant	0.934	54.000	0.934	plant to plant variance

Finally, below is the standardised residual plot. You can make up your own mind whether the variation across all sampling units is constant.



Two-way design (in randomized blocks)

Snedecor and Cochran present the yields of cowpea hay (pounds per 1/100 Morgen plot) from 3 varieties, each grown with 3 row spacings (4", 8" and 12" apart).

Firstly, let's use GenStat's **Design** menu to generate a field plan (the monograph does not give us a field layout). One random design is the following:

Spreadsheet [Book;3]*

Row	PlotNo	Block	Plots	Variety	Spacing
1	11	1	1	1	4
2	12	1	2	1	12
3	13	1	3	3	4
4	14	1	4	1	8
5	15	1	5	2	12
6	16	1	6	3	8
7	17	1	7	3	12
8	18	1	8	2	4
9	19	1	9	2	8
10	21	2	1	3	8
11	22	2	2	3	4
12	23	2	3	1	4
13	24	2	4	2	8
14	25	2	5	1	8
15	26	2	6	3	12
16	27	2	7	2	4
17	28	2	8	2	12
18	29	2	9	1	12
19	31	3	1	1	4
20	32	3	2	3	4
21	33	3	3	2	8
22	34	3	4	3	8
23	35	3	5	2	4
24	36	3	6	1	8
25	37	3	7	2	12
26	38	3	8	3	12
27	39	3	9	1	12

	BLOCK 1	BLOCK 2	BLOCK 3
1	Variety 1 **Spaced 4"**	Variety 3 **Spaced 8"**	Variety 1 **Spaced 4"**
2	Variety 1 **Spaced 12"**	Variety 3 **Spaced 4"**	Variety 3 **Spaced 4"**
3	Variety 3 **Spaced 4"**	Variety 1 **Spaced 4"**	Variety 2 **Spaced 8"**
4	Variety 1 **Spaced 8"**	Variety 2 **Spaced 8"**	Variety 3 **Spaced 8"**
5	Variety 2 **Spaced 12"**	Variety 1 **Spaced 8"**	Variety 2 **Spaced 4"**
6	Variety 3 **Spaced 8"**	Variety 3 **Spaced 12"**	Variety 1 **Spaced 8"**
7	Variety 3 **Spaced 12"**	Variety 2 **Spaced 4"**	Variety 2 **Spaced 12"**
8	Variety 2 **Spaced 4"**	Variety 2 **Spaced 12"**	Variety 3 **Spaced 12"**
9	Variety 2 **Spaced 8"**	Variety 1 **Spaced 12"**	Variety 1 **Spaced 12"**

Note that spacing experiments, by definition, are unlikely to produce plot mean (or plot total) yields *whose variances are constant*. Why is that?

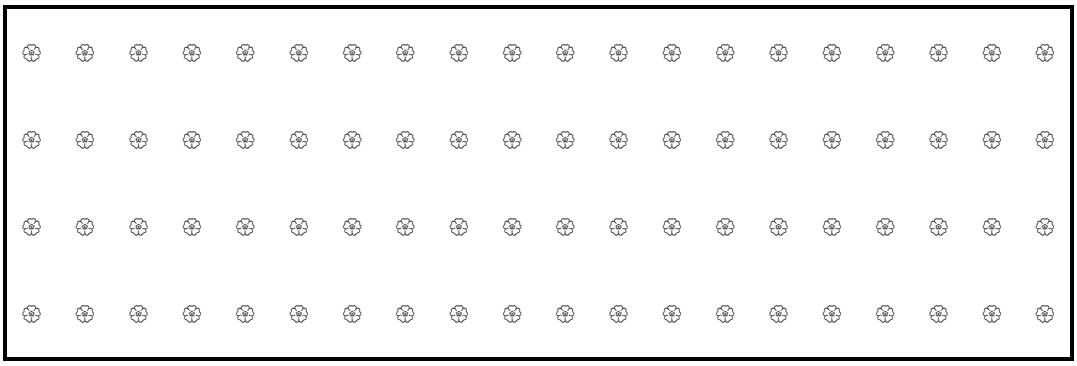
From statistical theory, if you add independent variates whose individual variances are the same, the variance of the sum is the sum of the individual variances. Let σ^2 be the variance *on a per plant basis*. Then, for independently growing plants,

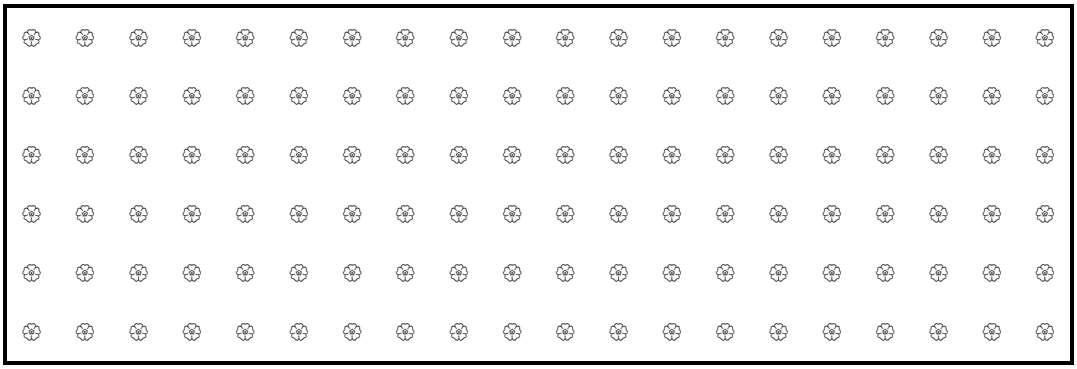
$$\text{var}(\textit{Total yield}) = \text{var}(Y_1 + \ldots + Y_n) = n\,\sigma^2$$

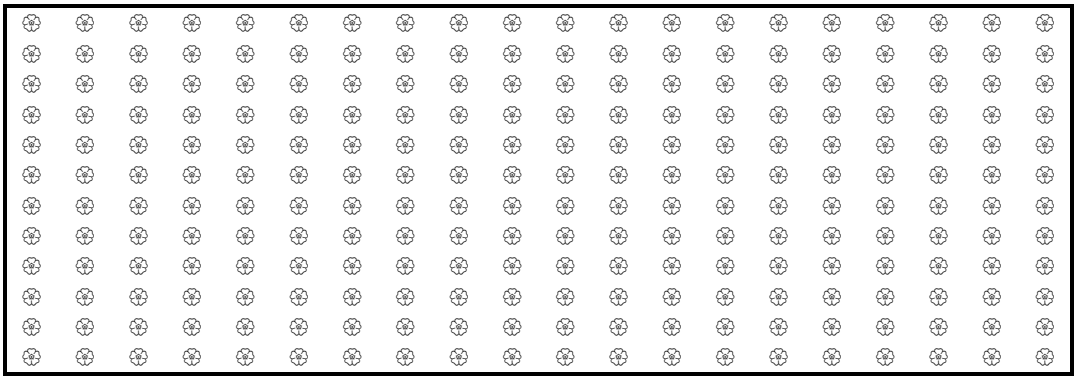
and hence

$$\text{var}(\textit{Mean yield}) = \text{var}(\bar{y}) = \sigma^2 / n$$

Now put that in the context of this spacing experiment. The plot area is 0.01 Morgen which is about 86 m^2. Spacings are about 10, 20, 30 cm. The number of rows of varying shapes depends on the shape of the plot. We'll assume for illustration that we have multiples of 1.2m areas for rows. The 12" spacing is equivalent to 30cm row spacing, so 4 rows are used at that spacing, 6 rows at 20cm spacing and 12 rows at 10cm spacing.







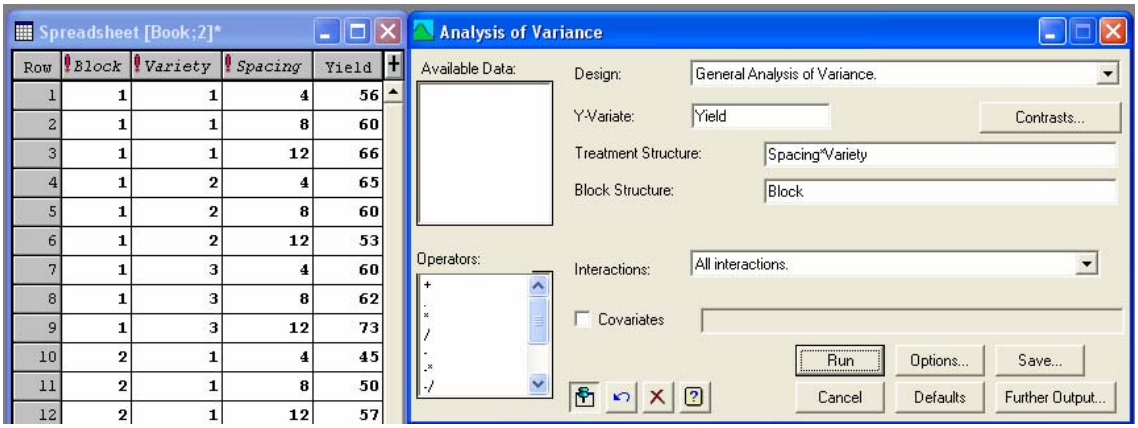
Plots like these (with row spacings 12", 8", 4") consist of varying numbers of plants (in the ratio 4:6:12). Other combinations are possible. The point is, total yield (or mean yield) obtained from plots with varying numbers of plants will have changing variance *if the plants grow independently*.

With plant competition, the variance of total yield could well even out across all shaped plots. Plant competition means that the yields become spatially correlated. We will ignore this problem for the moment. Changing variance and correlated yield models are available in **Linear Mixed Models (REML)** section.

Example 6 Yields (pounds) of cowpea hay from Snedecor and Cochran, page 309.

Variety	Spacing	Block 1	Block 2	Block 3	Block 4
I	4	56	45	43	46
	8	60	50	45	48
	12	66	57	50	50
II	4	65	61	60	63
	8	60	58	56	60
	12	53	53	48	55
III	4	60	61	50	53
	8	62	68	67	60
	12	73	77	77	65

There are two strata in this experiment, **Block** and **Block.Plot**. The **Block Structure** is therefore **Block** + **Block.Plot**, or simply **Block/Plot**. Since the smallest stratum can be omitted, **Block** is sufficient.



The full analysis of the data, including L.S.D. values and stratum variances, is as follows.

Analysis of variance

Variate: Yield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block stratum	3	255.64	85.21	4.82	
Block.*Units* stratum					
Variety	2	1027.39	513.69	29.07	<.001
Spacing	2	155.06	77.53	4.39	0.024
Variety.Spacing	4	765.44	191.36	10.83	<.001
Residual	24	424.11	17.67		
Total	35	2627.64			

Tables of means

Variate: Yield

Grand mean 57.81

Variety	1	2	3
	51.33	57.67	64.42

Spacing	4.	8.	12.
	55.25	57.83	60.33

Variety	Spacing	4.	8.	12.
1		47.50	50.75	55.75
2		62.25	58.50	52.25
3		56.00	64.25	73.00

Standard errors of differences of means

Table	Variety	Spacing	Variety Spacing
rep.	12	12	4
d.f.	24	24	24
s.e.d.	1.716	1.716	2.972

Least significant differences of means (5% level)

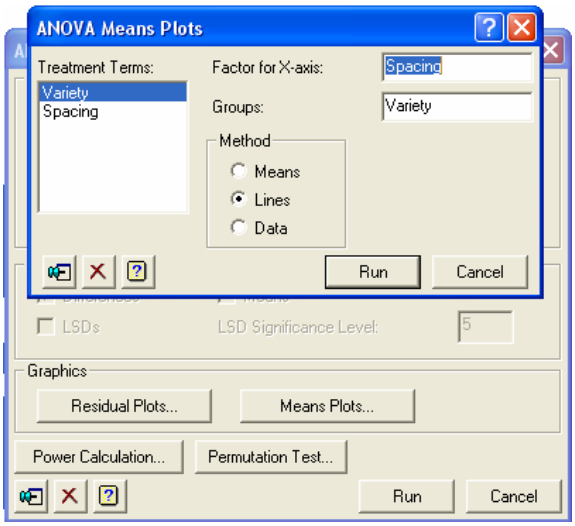
Table	Variety	Spacing	Variety Spacing
rep.	12	12	4
d.f.	24	24	24
l.s.d.	3.542	3.542	6.135

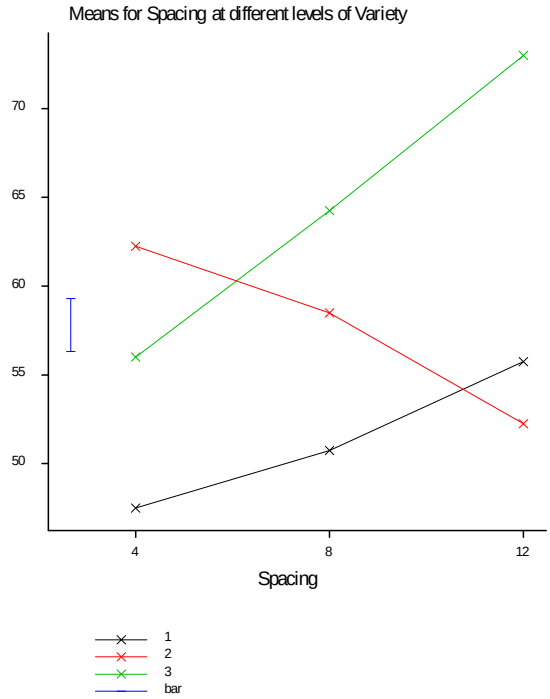
Estimated stratum variances

Variate: Yield

Stratum	variance	effective d.f.	variance component
Block	85.213	3.000	7.505
Block.*Units*	17.671	24.000	17.671

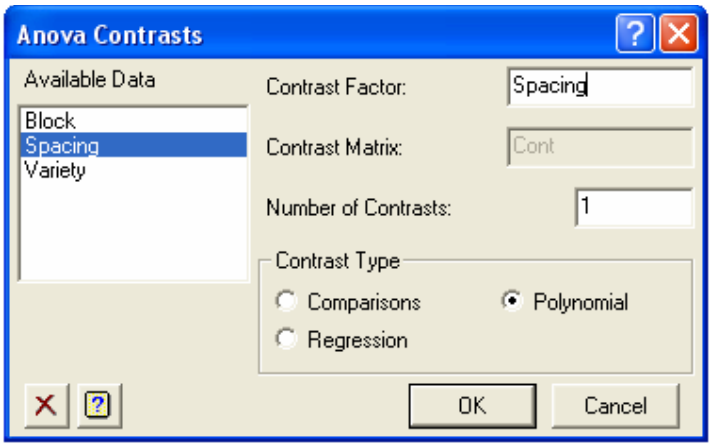
There is strong statistical evidence ($P<0.001$) that the change in mean yield at different row spacings is not the same for all three varieties. A means plot illuminates the differences:





There is a strong linear trend in mean yield, but the means for variety 2 *decrease* with increasing spacing. Varieties 1 and 3 must have heavy vegetative growth that requires at least 12" to approach optimal yield.
These linear trends can be incorporated into the ANOVA, using the **Contrast** button on the ANOVA table.

Firstly, for the factor **Spacing** we are interested in a linear trend: this is a situation where **POL** (polynomial regression/contrast) can be used.



Click on the **Contrast** button, select the **Spacing** factor and nominate **Polynomial**. The degree of the polynomial you wish to fit is the **Number of Contrasts**. In this case leave this as 1 and click **OK**. GenStat replaces **Spacing** in the treatment structure with **POL(Spacing;1)**.

We are also interested in sub-hypotheses for the **Variety** factor. In this case, two are more natural than other choices:

- H_0: Variety 1 and Variety 3 means are equal: we wish to assess $\mu_3-\mu_1$.
- H_0: Variety 2 mean is equal to the *average mean* of Variety 1 and Variety 3 means are equal: we wish to assess $(\mu_3+\mu_1)/2-\mu_2$.

Contrasts are simply the coefficients of the means in the questions asked. For any contrast, the coefficients will add to zero. GenStat allows two types of questions, labelled **Comparisons** and **Regression**.

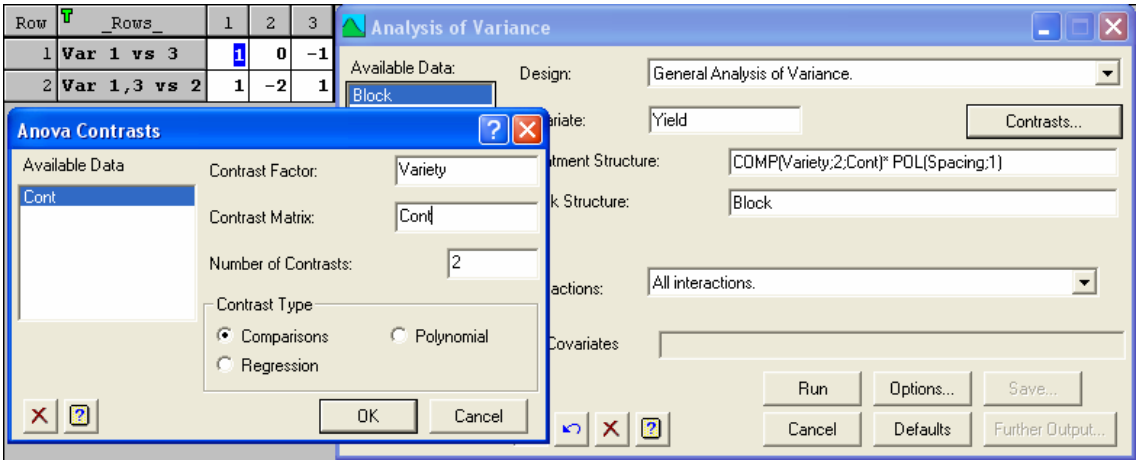
Comparisons allows any number of questions to be asked, with no restrictions on the questions asked. Their component sums of squares will not add to the Variety SS.

For *t* treatments, **Regression** allows up to (*t*-1) questions, with restrictions on the questions asked. The questions must be *orthogonal*, that is, balanced in a special way. The component sums of squares for all (*t*-1) contrasts will add to the Treatment SS. Even if the contrasts are orthogonal, the **Comparisons** choice can be used. The only difference is that GenStat does not report deviations when **Comparisons** is selected.

- Variety 1 vs 3: $\mu_3-\mu_1$ is equivalent to (-1, 0 ,1) multipliers of (μ_1, μ_2, μ_3) respectively

- Variety 1&3 vs 2: $(\mu_3+\mu_1)/2-\mu_2$ is equivalent to (½, -1, ½) multipliers of (μ_1, μ_2, μ_3). It is preferable to enter integers rather than fractions, so multiplier by a constant (in this case 2) to remove fractions. The contrast is then (1, -2, 1)

Click on the **Contrast** button, select the **Variety** factor and nominate **Regression** and enter the **Number of Contrasts** you wish to make (here 2). GenStat opens up a table (which is names, by default, **Cont**, or **Cont_1** if **Cont** exists) with (here) 2 rows (questions) and 3 columns (levels). Names of the levels are placed above the columns. Enter the contrast coefficients, and double click on the grey areas of the rows, where the names of each contrast can be set up. Then return to the **ANOVA Contrasts** menu and click **OK**. GenStat replaces **Variety** in the treatment structure with **REG(Variety;2;Cont)** or **COMP(Variety;2;Cont)** if you chose **Comparisons**.



The new ANOVA table is as follows.

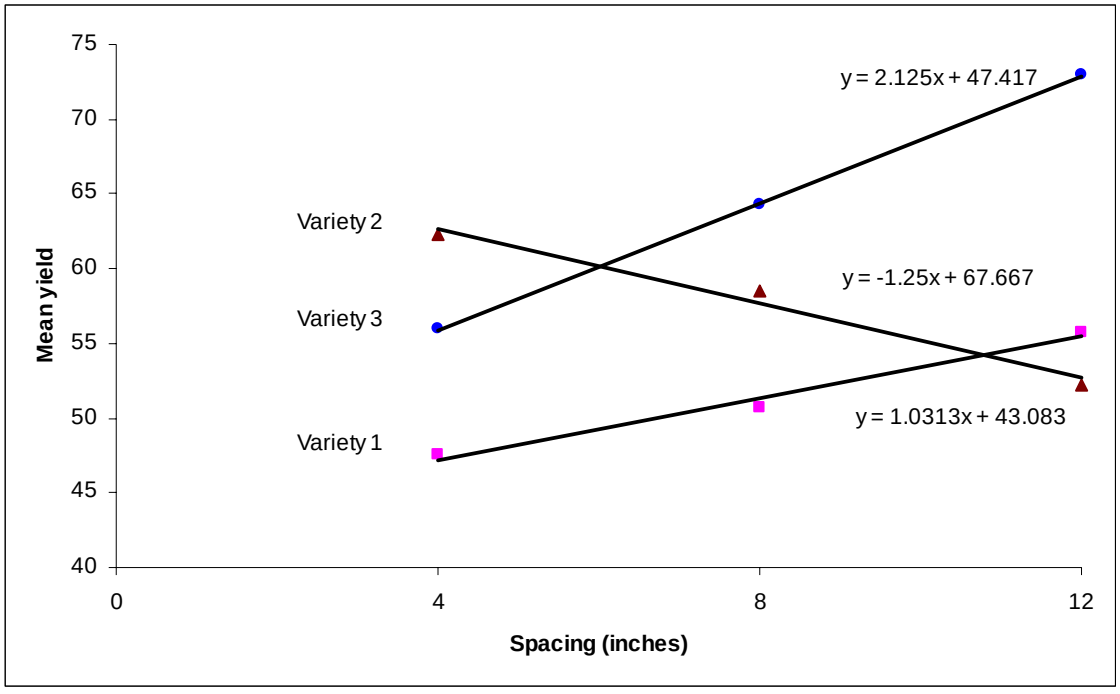
Analysis of variance

Variate: Yield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block stratum	3	255.64	85.21	4.82	
Block.*Units* stratum					
Variety	2	1027.39	513.69	29.07	<.001
Var 1 vs 3	1	1027.04	1027.04	58.12	<.001
Var 1,3 vs 2	1	0.35	0.35	0.02	0.890
Spacing	2	155.06	77.53	4.39	0.024
Lin	1	155.04	155.04	8.77	0.007
Deviations	1	0.01	0.01	0.00	0.978
Variety.Spacing	4	765.44	191.36	10.83	<.001
Var 1 vs 3.Lin	1	76.56	76.56	4.33	0.048
Var 1,3 vs 2.Lin	1	682.52	682.52	38.62	<.001
Var 1 vs 3.Dev	1	0.52	0.52	0.03	0.865
Var 1,3 vs 2. Dev	1	5.84	5.84	0.33	0.571
Residual	24	424.11	17.67		
Total	35	2627.64			

Note that with 3 spacing levels, Dev is identical to the quadratic term. With 4 spacing levels and a linear model requested, Dev will be the combined quadratic and cubic components: it's what is left after the requested polynomial is fitted.
This table adds the following to what we knew already. The slope in the regression of the means of varieties 1 and 3 are marginally different (P=0.048), whereas the slope for variety 2 in comparison is strikingly different (P<0.001) to an average slope for variety 1 and 3 means.

Here are trend lines added in Excel:



If we just wish to estimate the fitted regressions using GenStat, it is easier to use a general regression ignoring blocks (because the design is orthogonal). The factor column **Spacing** needs to be converted to a variate instead (simply point to the column, right click and select **Convert to Variate**). The **Model to be fitted** is **Variety.Spacing**.

Estimates of parameters

Parameter	estimate	s.e.	t(30)	t pr.
Constant	43.08	3.65	11.80	<.001
Spacing	1.031	0.423	2.44	0.021
Variety 2	24.58	5.17	4.76	<.001
Variety 3	4.33	5.17	0.84	0.408
Spacing.Variety 2	-2.281	0.598	-3.82	<.001
Spacing.Variety 3	1.094	0.598	1.83	0.077

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Variety	1

The model for the reference Variety 1 comes out immediately:
Mean yield = 43.08 + 1.031 *Spacing*

For (say) variety 2 we add 24.58 to the intercept and -2.281 to the slope:
Mean yield = 67.66 - 1.250 *Spacing*

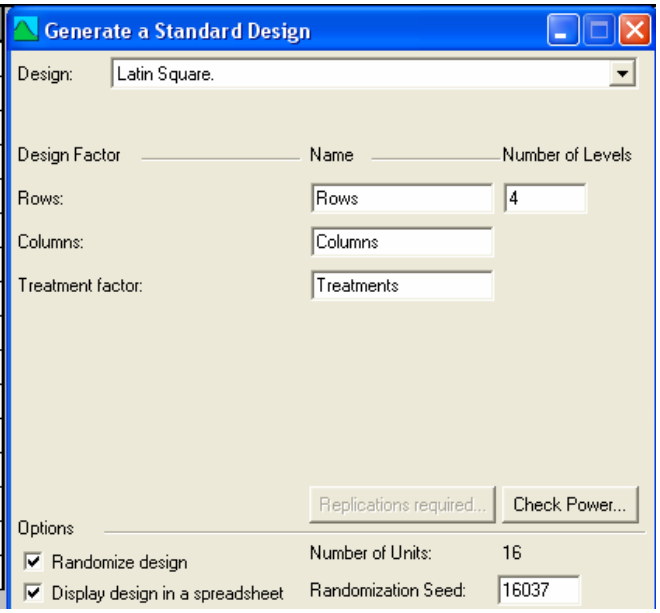
Latin Square design

Occasionally we need to block in two directions in the field (especially in animal trials, where individual animals form one block, and the experiment is repeated over time, time forming a second block).

For a Latin Square design, we need to have as many blocks in both directions as we have treatments. We then balance the allocation of treatments so that each occurs just once in each row and once in each column.

Here is GenStat's **Design** menu for generating a random 4×4 design:

Row	PlotNo	Rows	Columns	Treatments
1	11	1	1	4
2	12	1	2	1
3	13	1	3	3
4	14	1	4	2
5	21	2	1	3
6	22	2	2	2
7	23	2	3	4
8	24	2	4	1
9	31	3	1	2
10	32	3	2	3
11	33	3	3	1
12	34	3	4	4
13	41	4	1	1
14	42	4	2	4
15	43	4	3	2
16	44	4	4	3



Treatment allocation for this random design:

	Column block			
Row block	1	2	3	5
1	**4**	**1**	**3**	**2**
2	**3**	**2**	**4**	**1**
3	**2**	**3**	**1**	**4**
4	**1**	**4**	**2**	**3**

We have marked a typical row block, a typical column block, and a typical plot (the intersection of a row block and a column block). Thus, there are three strata, and hence the **Block Structure** is

Row + Column + Row.Column

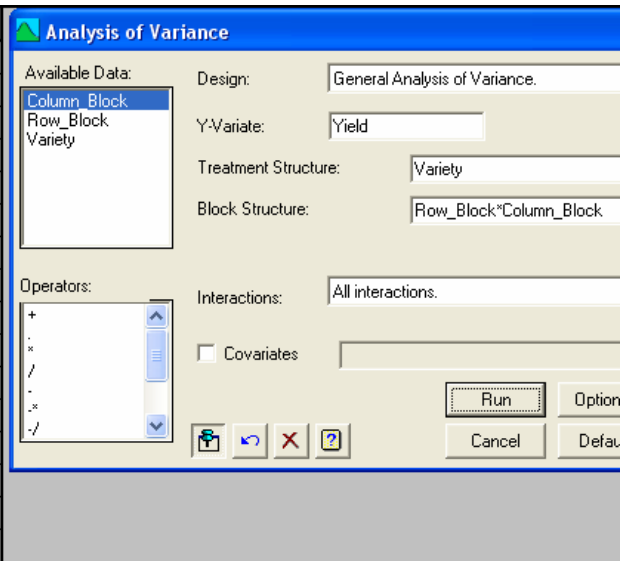
which can be shortened to Row*Column, or, since the final stratum can always be omitted, Row + Column.

Example 7 Wheat yields (kg per plot) from Steel and Torrie, page 224.

	Column block			
Row block	1	2	3	4
1	C	D	B	A
2	B	A	C	D
3	D	C	A	B
4	A	B	D	C

Column block			
1	2	3	4
10.5	7.7	12.0	13.2
11.1	12.0	10.3	7.5
5.8	12.2	11.2	13.7
11.6	12.3	5.9	10.2

Row	Row_Block	Column_Block	Variety	Yield
1	1	1	C	10.5
2	2	1	B	11.1
3	3	1	D	5.8
4	4	1	A	11.6
5	1	2	D	7.7
6	2	2	A	12
7	3	2	C	12.2
8	4	2	B	12.3
9	1	3	B	12
10	2	3	C	10.3
11	3	3	A	11.2
12	4	3	D	5.9
13	1	4	A	13.2
14	2	4	D	7.5
15	3	4	B	13.7
16	4	4	C	10.2



Analysis of variance

Variate: Yield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Row_Block stratum	3	1.9550	0.6517	1.44	
Column_Block stratum	3	6.8000	2.2667	5.00	
Row_Block.Column_Block stratum					
Variety	**3**	**78.9250**	**26.3083**	**58.03**	**<.001**
Residual	6	2.7200	0.4533		
Total	15	90.4000			

Message: the following units have large residuals.

Row_Block 4 Column_Block 4 -0.85 s.e. 0.41

Tables of means

Variate: Yield

Grand mean 10.45

Variety	A	B	C	D
	12.00	12.27	10.80	6.72

Standard errors of differences of means

Table	Variety
rep.	4
d.f.	6
s.e.d.	0.476

Least significant differences of means (5% level)

Table	Variety
rep.	4
d.f.	6
l.s.d.	1.165

Estimated stratum variances

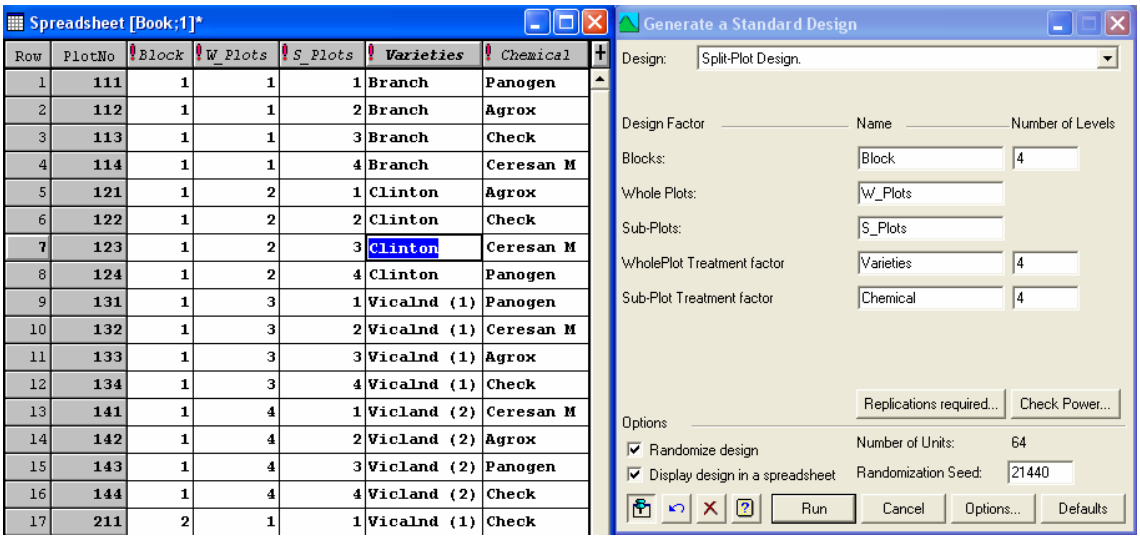
Variate: Yield

Stratum	variance	effective d.f.	variance component
Row_Block	0.652	3.000	0.050
Column_Block	2.267	3.000	0.453
Row_Block.Column_Block	0.453	6.000	0.453

From the stratum variances, columns show more variability than rows.

Split-plot design (in randomized blocks)

Firstly, we will use GenStat's **Design** menu to generate a field plan to correspond to Steel and Torrie's oats experiment (page 383) with four varieties randomised to whole plots and four chemical seed treatments (one of which is a control) to split plots. Appropriate factor labels have replaced numbers.



Spreadsheet [Book;1]*

Row	PlotNo	Block	W_Plots	S_Plots	Varieties	Chemical
1	111	1	1	1	Branch	Panogen
2	112	1	1	2	Branch	Agrox
3	113	1	1	3	Branch	Check
4	114	1	1	4	Branch	Ceresan M
5	121	1	2	1	Clinton	Agrox
6	122	1	2	2	Clinton	Check
7	123	1	2	3	Clinton	Ceresan M
8	124	1	2	4	Clinton	Panogen
9	131	1	3	1	Vicalnd (1)	Panogen
10	132	1	3	2	Vicalnd (1)	Ceresan M
11	133	1	3	3	Vicalnd (1)	Agrox
12	134	1	3	4	Vicalnd (1)	Check
13	141	1	4	1	Vicland (2)	Ceresan M
14	142	1	4	2	Vicland (2)	Agrox
15	143	1	4	3	Vicland (2)	Panogen
16	144	1	4	4	Vicland (2)	Check
17	211	2	1	1	Vicalnd (1)	Check

Notice that GenStat creates three factor columns (Block, W_Plot and S_Plot), one for each of the three strata in this experiment. The field plan is also printed in the **Output** window.

Treatment combinations on each unit of the design

Block	W_Plots \ S_Plots	1	2	3	4
1	1	4 3	4 4	4 1	4 2
	2	3 4	3 1	3 2	3 3
	3	1 3	1 2	1 4	1 1
	4	2 2	2 4	2 3	2 1
2	1	1 1	1 4	1 3	1 2
	2	2 4	2 1	2 3	2 2
	3	3 1	3 4	3 2	3 3
	4	4 2	4 4	4 1	4 3
3	1	4 4	4 1	4 2	4 3
	2	1 3	1 1	1 2	1 4
	3	3 2	3 3	3 1	3 4
	4	2 2	2 4	2 3	2 1
4	1	1 2	1 3	1 1	1 4
	2	3 3	3 4	3 1	3 2
	3	2 4	2 1	2 3	2 2
	4	4 4	4 3	4 1	4 2

Treatment factors are listed in the order: Varieties, Chemical.

This field plan is reproduced graphically with labels:

Block 1	Panogen	Agrox	Check	Ceresan M	Branch
	Agrox	Check	Ceresan M	Panogen	Clinton
	Panogen	Ceresan M	Agrox	Check	Vicland (1)
	Ceresan M	Agrox	Panogen	Check	Vicland (2)
Block 2	Check	Agrox	Panogen	Ceresan M	Vicland (1)
	Agrox	Check	Panogen	Ceresan M	Vicland (2)
	Check	Agrox	Ceresan M	Panogen	Clinton
	Ceresan M	Agrox	Check	Panogen	Branch
Block 3	Agrox	Check	Ceresan M	Panogen	Branch
	Panogen	Check	Ceresan M	Agrox	Vicland (1)
	Ceresan M	Panogen	Check	Agrox	Clinton
	Ceresan M	Agrox	Panogen	Check	Vicland (2)
Block 4	Ceresan M	Panogen	Check	Agrox	Vicland (1)
	Panogen	Agrox	Check	Ceresan M	Clinton
	Agrox	Check	Panogen	Ceresan M	Vicland (2)
	Agrox	Panogen	Check	Ceresan M	Branch

There are clearly three strata here: blocks, the ¼ block strips (the whole-plots) that the varieties are randomised to, and the ¼ whole-plot shapes (the split-plots) that the seed protectants were assigned to at random. The **Block Structure** is therefore

Block + Block.Whole_Plot + Block.Whole_Plot.Split_plot

with the shortcut

Block/Whole_Plot/Split_plot

which describes the way the units were formed in the field: whole-plots were formed as large units within blocks, and split-plots were formed as smaller units within whole-plots.

Providing you set up these three factors, this structure is what you would use irrespective of the complexity of the whole-plot treatment and the split-plot treatment structures. For example, the treatments applied to whole-plots could have a 3×4 factorial structure, while those applied to the split-plots a (2×2+1) incomplete factorial structure.

For this example, there were simple structures for both whole-plot and split-plot treatment structures. Hence the following **Block Structure** can be used instead:

Block + Block.Variety + Block.Variety.Chemical

In fact this design can be thought of in two ways.

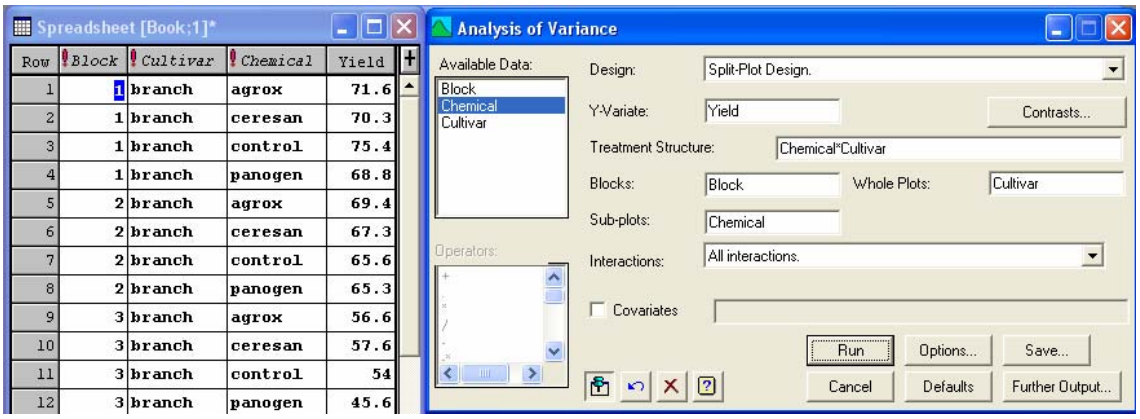
1. RCBD with varieties as treatments.

This, in fact, forms the whole-plot part of the combined split-plot ANOVA.

2. Four separate RCBDs, one per variety, with seed chemical protectants as treatments. This is one such layout, for Branch.

Block 1	Panogen	Agrox	Check	Ceresan M	Branch
Block 2	Ceresan M	Agrox	Check	Panogen	Branch
Block 3	Agrox	Check	Ceresan M	Panogen	Branch
Block 4	Agrox	Panogen	Check	Ceresan M	Branch

In fact, this is an important concept in checking the assumptions at the split-plot level. This ANOVA produces 9 *df* for the Residual MS. There are four such residuals to check for "homogeneity"; their average is, in fact, the split-plot Residual MS in the combined analysis. The combined analysis is feasible only when these individual variance components are commensurable.



Example 8 From Snedecor and Cochran page 384

		Seed chemical protectant			
Cultivar	Block	Control	Ceresan M	Panogen	Agrox
Vicland (1)	1	42.9	53.8	49.5	44.4
	2	41.6	58.5	53.8	41.8
	3	28.9	43.9	40.7	28.3
	4	30.8	46.3	39.4	34.7
Vicland (2)	1	53.3	57.6	59.8	64.1
	2	69.6	69.6	65.8	57.4
	3	45.4	42.4	41.4	44.1
	4	35.1	51.9	45.4	51.6
Clinton	1	62.3	63.4	64.5	63.6
	2	58.5	50.4	46.1	56.1
	3	44.6	45.0	62.6	52.7
	4	50.3	46.7	50.3	51.8
Branch	1	75.4	70.3	68.8	71.6
	2	65.6	67.3	65.3	69.4
	3	54.0	57.6	45.6	56.6
	4	52.7	58.5	51	47.4

First, the standard split-plot ANOVA is obtained (using the specific split-plot menu).

Analysis of variance

Variate: Yield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block stratum	3	2842.87	947.62	13.79	
Block.Cultivar stratum					
Cultivar	3	2848.02	949.34	13.82	0.001
Residual	9	618.29	68.70	3.38	
Block.Cultivar.Chemical stratum					
Chemical	3	170.54	56.85	2.80	0.054
Cultivar.Chemical	9	586.47	65.16	3.21	0.006
Residual	36	731.20	20.31		
Total	63	7797.39			

Message: the following units have large residuals.

Block 2 Cultivar clinton	-7.27	s.e. 3.11
Block 2 Cultivar vicland2	6.45	s.e. 3.11
Block 2 Cultivar clinton Chemical panogen	-8.24	s.e. 3.38
Block 2 Cultivar vicland2 Chemical agrox	-9.09	s.e. 3.38
Block 3 Cultivar clinton Chemical panogen	9.81	s.e. 3.38
Block 4 Cultivar vicland2 Chemical control	-8.34	s.e. 3.38

Tables of means

Variate: Yield

Grand mean 52.81

Cultivar	branch	clinton	vicland1	vicland2
	61.07	54.31	42.46	53.41

Chemical	agrox	ceresan	control	panogen
	52.23	55.20	50.69	53.13

Cultivar	Chemical	agrox	ceresan	control	panogen
branch		61.25	63.43	61.93	57.68
clinton		56.05	51.38	53.93	55.88
vicland1		37.30	50.63	36.05	45.85
vicland2		54.30	55.38	50.85	53.10

Standard errors of differences of means

Table	Cultivar	Chemical	Cultivar Chemical
rep.	16	16	4
s.e.d.	2.930	1.593	4.025
d.f.	9	36	26.78
Except when comparing means with the same level(s) of Cultivar			3.187
d.f.			36

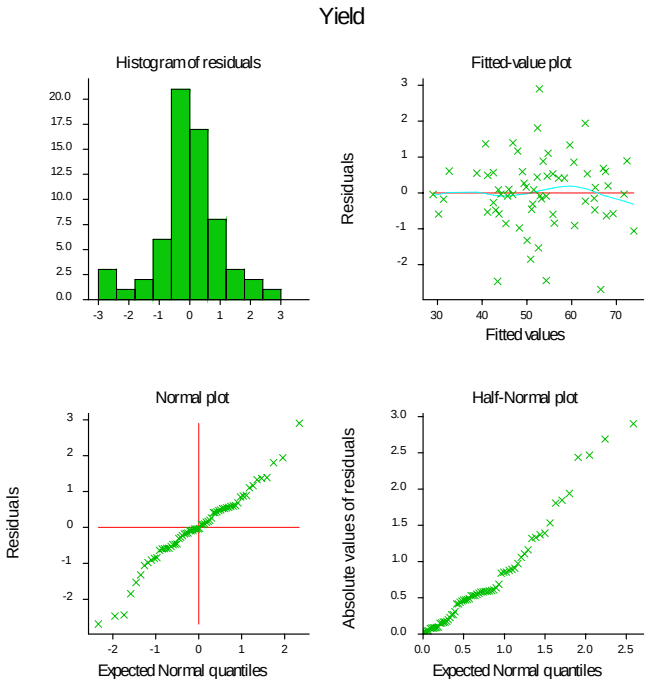
Least significant differences of means (5% level)

Table	Cultivar	Chemical	Cultivar Chemical
rep.	16	16	4
l.s.d.	6.629	3.232	8.263
d.f.	9	36	26.78
Except when comparing means with the same level(s) of Cultivar			6.463
d.f.			36

GenStat organizes the analysis into three strata corresponding to what was done in the field. Notice the following.

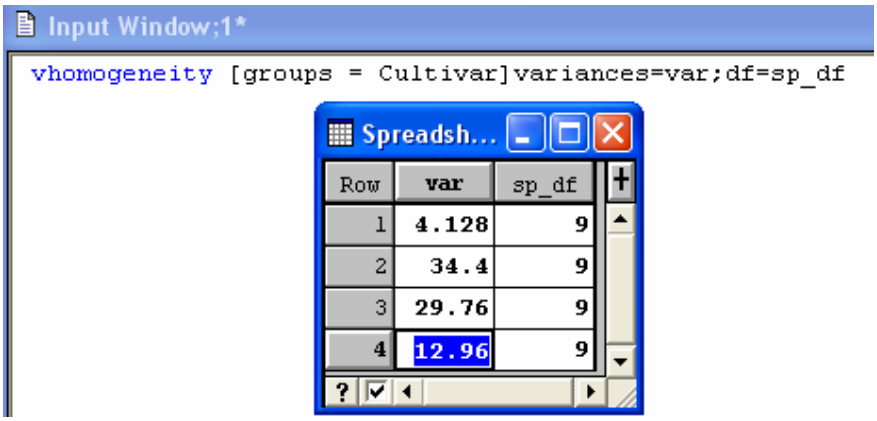
- Cultivar is tested in the whole-plot stratum, since whole-plots are the replicates for this treatment factor.
- Chemical and Cultivar.Chemical are tested in the split-plot stratum, since split-plots are the replicates for this treatment/interaction.
- There are several s.e.d. and l.s.d. values. Each is used for an appropriate treatment mean comparison. Not all comparisons lead to exact *t* tests. Performing a two stage randomization in the field has made the subsequent analysis slightly more complex than a one stage randomization.

Before interpreting the analysis, we should check the residual plot. Maybe there is some fanning, but nothing jumps out as a major problem.



Before interpreting the analysis, the components that form the split-plot error should be checked.

We do this in GenStat by clicking in the spreadsheet, then **Restrict/Filter** > **To Groups (factor levels)**. Select Cultivar and, one by one, each of the levels to perform a simple RCBD ANOVA. The Residual MS values (each with 9 *df*) are 4.128(Vicland (1)), 34.40 (Vicland (2)), 29.76 (Clinton), 12.96 (Branch). These appear quite different. Their average is 20.312, which is the split-plot Residual MS, with 4×9 = 36 *df*. In fact, performing a Bartlett test of homogeneity of variances on these indicates significance at *P*=0.021.

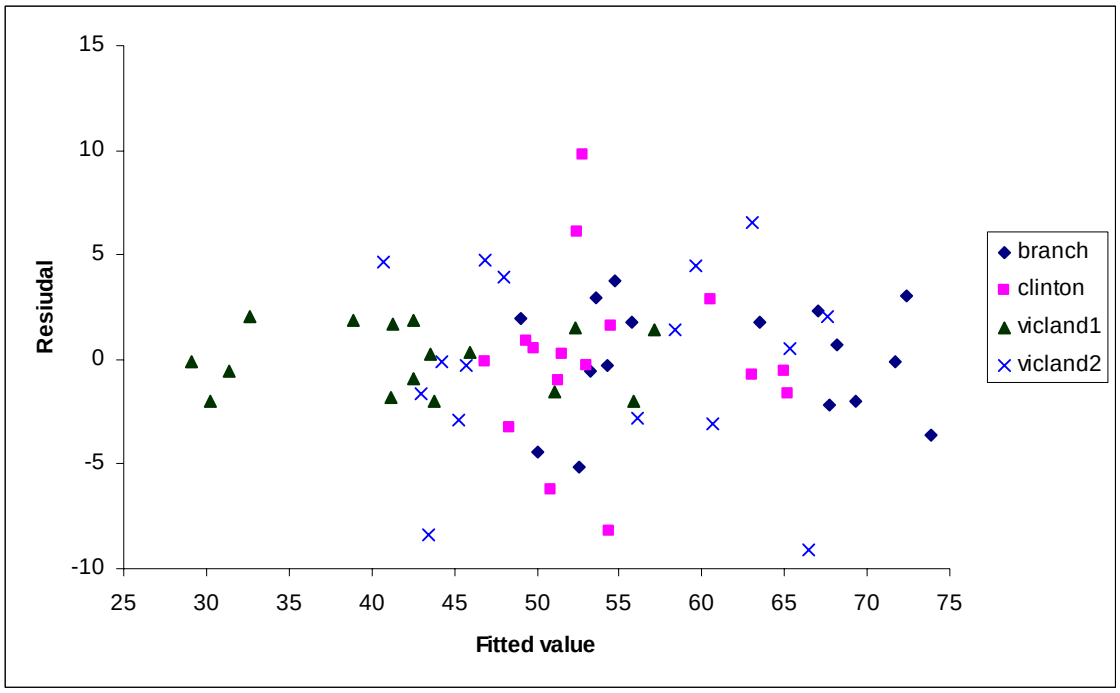


Bartlett's test for homogeneity of variances

Chi-square 9.75 on 3 degrees of freedom: probability 0.021

Steel and Torrie give further information about these varieties. Vicland (1) is a variety infected with *H. victoriae*, Vicland (2) is the same variety but is not infected. Clinton and Branch are varieties resistant to *H. victoriae*. The variation in the Vicland (1) data appears smaller than for the other varieties. It is possible that the actual levels of this factor are associated with different variances: one level is expected to have consistently smaller yields, since these seeds have been infected. Linear Mixed Models (REML) allows us to model this.

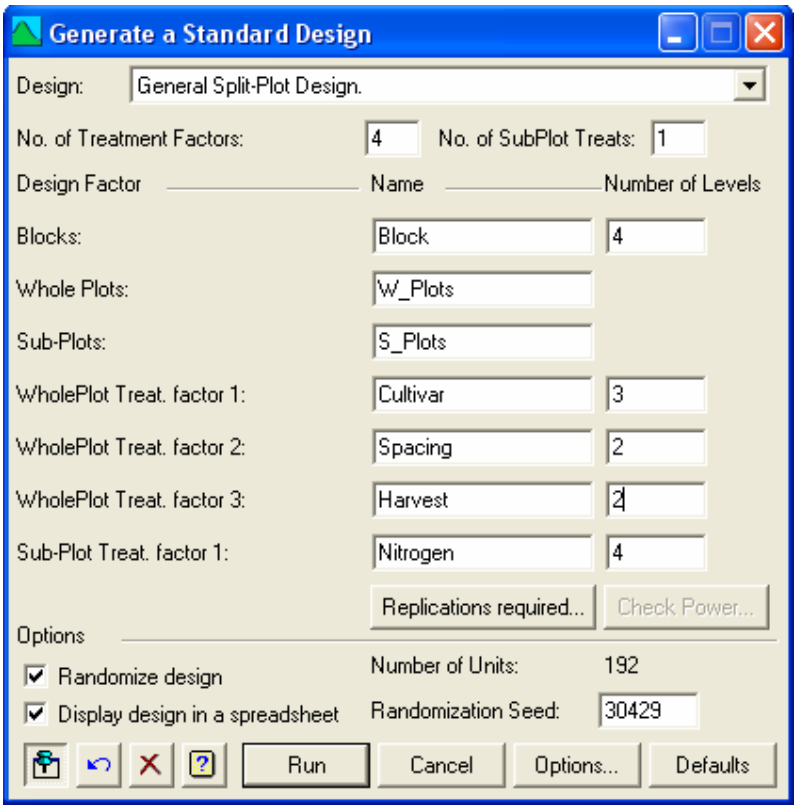
Has the combined analysis overlooked this problem? If we **Save** the fitted values and residuals, we can obtain a residual plot with different colours for the different varieties.



In this plot, the residuals from Vicland (1) appear less varied than the other varieties (corresponding to the significantly smaller variance in the yields of this variety). It would appear that the combined split-plot analysis is inappropriate for these data. We will explore alternative analyses in the LMM (REML) section of the manual.

General split-plot design

The split-plot design in the previous section had just one treatment factor applied to whole-plots and to split-plots. There is no restriction on the treatment structure in either stratum. GenStat's Design menu allows for a general split-plot design. You simply indicate how many treatment factors there are altogether, and how many of these are allocated to split-units. The following example produces a random design with cultivar × spacing × harvest treatments (3×2×4 = 24 combinations) allocated to whole-plots, and four levels of nitrogen allocated to split-plots within each whole-plot.



GenStat creates, as before, a Block stratum, a W_Plot stratum and a S_Plot stratum. This time, there are three factors required to fully define the whole-plots. Nevertheless, the Block Structure remains as Block/W_Plot/S_Plot.

PlotNo	Block!	W_Plots!	S_Plots!	Cultivar!	Spacing!	Harvest!	Nitrogen!
1101	1	1	1	2	2	2	1
1102	1	1	2	2	1	2	2
1103	1	1	3	2	2	2	4
1104	1	1	4	2	1	1	3
1105	1	1	5	2	1	2	3
1106	1	1	6	2	1	2	1
1107	1	1	7	2	1	1	2
1108	1	1	8	2	2	2	3
1109	1	1	9	2	2	1	4
1110	1	1	10	2	2	2	2
1111	1	1	11	2	1	1	1
1112	1	1	12	2	2	1	3
1113	1	1	13	2	1	1	4
1114	1	1	14	2	2	1	2
1115	1	1	15	2	2	1	1
1116	1	1	16	2	1	2	4
1201	1	2	1	3	2	1	2

etc …

Example. Curt Lee (Agro-Tech, Inc., Velva, North Dakota, USA) kindly supplied data from the following experiment on wheat.

Six blocks were set up and each divided into two whole-plots (WP). One whole-plot was randomly fertilized with a full recommended rate of nitrogen fertilizer (Standard), the other not fertilized (Reduced). (There was a small alley between the left-hand three blocks and the right-hand three blocks, but not vertically.) There was also residual nitrogen in the field, in fact the final applied N plus residual N was 100 lbs for the standard fertility and 50 lbs for the reduced fertility plots.

Each whole-plot was divided into four split-plots (SP). The four treatments allocated randomly to these plots were a fungicide treatment (or a blank treatment), and an early (at the tillering stage) or a late (at the flag leaf stage) application of the fungicide and the blank.

Block	WP	SP 1	SP 2	SP 3	SP 4	SP 1	SP 2	SP 3	SP 4	WP	Block
1	1	1.77	2.08	1.79	2.68	2.07	2.01	2.56	3.89	1	4
	2	2.79	3.69	2.77	2.97	2.03	2.16	2.03	2.50	2	
2	1	2.58	3.55	2.74	3.29	2.15	2.13	2.07	2.41	1	5
	2	2.05	2.37	2.19	2.25	2.97	2.28	2.53	3.53	2	
3	1	3.21	2.81	2.16	2.61	2.93	3.44	2.95	3.49	1	6
	2	2.69	3.65	3.15	3.10	2.30	2.76	2.49	3.54	2	

KEY

= Reduced Fertility

= Standard Fertility

= Early Timing No Fungicide

= Late Timing No Fungicide

= Early Timing Fungicide

= Late Timing Fungicide

The blank plots were actually sprayed with the treatments that contained all the carrier material (water, solvents, etc), except the actual fungicide. Thus, since a treatment was actually applied to the blank plots, the split-plot treatments can be thought of as a 2 × 2 factorial combination.

Alternatively, you can think of the split-plot treatments as a simple set of four treatments, and extract three **contrasts** to estimate the following characteristics.

a) Estimate the effect of the fungicide versus no fungicide, by comparing the mean yields from fungicide (early and late) plots to no fungicide (early and late) plots.

b) Estimate the effect of different timing by comparing the mean yields from fungicide early plots to fungicide late plots.

c) Estimate the effect of the two check treatments by comparing the mean yields from no fungicide early plots to no fungicide late plots. (They should yield the same, unless they are getting something out of the carrier materials.)

With the split-plot treatment as a 2 × 2 factorial

Analysis of variance

Variate: Yield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block stratum	5	2.1795	0.4359	4.18	
Block.W_Plot stratum					
Fert	1	4.7376	4.7376	45.38	0.001
Residual	5	0.5220	0.1044	0.97	
Block.W_Plot.S_Plot stratum					
Fung	1	2.7552	2.7552	25.51	<.001
Timing	1	0.5043	0.5043	4.67	0.039
Fert.Fung	1	0.2002	0.2002	1.85	0.183
Fert.Timing	1	0.0261	0.0261	0.24	0.626
Fung.Timing	1	0.6674	0.6674	6.18	0.019
Fert.Fung.Timing	1	0.0000	0.0000	0.00	0.993
Residual	30	3.2398	0.1080		
Total	47	14.8322			

Message: the following units have large residuals.

Block 4 W_Plot 1 S_Plot 4	**0.710**	**s.e.**	**0.260**
Block 5 W_Plot 2 S_Plot 4	**0.642**	**s.e.**	**0.260**
Block 6 W_Plot 1 S_Plot 4	**0.583**	**s.e.**	**0.260**

Tables of means

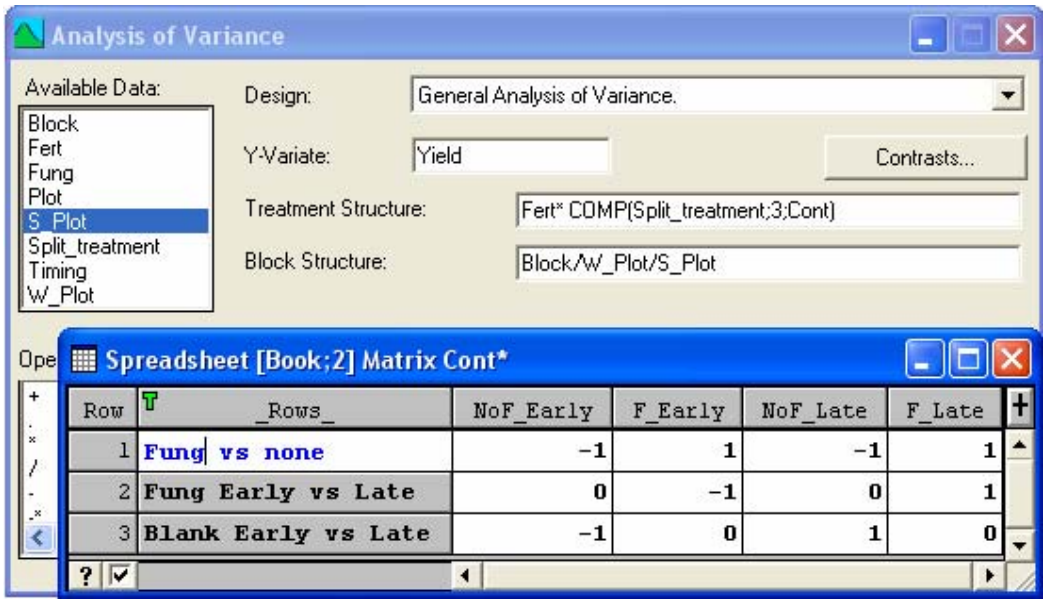
etc…

Estimated stratum variances

Stratum	variance	effective d.f.	variance component
Block	0.4359	5.000	0.0414
Block.W_Plot	0.1044	5.000	-0.0009
Block.W_Plot.S_Plot	0.1080	30.000	0.1080

The three residuals were all from edge plots in blocks 4, 5 and 6. On checking, the research company discovered that these plots had not been trimmed to equal length. For their analysis they went back, measured each plot and corrected the yield based on actual harvested plot length. We will not do that here.

With the split-plot treatment as 4 simple treatments with structure



Analysis of variance

Variate: Yield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block stratum	5	2.1795	0.4359	4.18	
Block.W_Plot stratum					
Fert	1	4.7376	4.7376	45.38	0.001
Residual	5	0.5220	0.1044	0.97	
Block.W_Plot.S_Plot stratum					
Split_treatment	3	3.9269	1.3090	12.12	<.001
Fung vs none	**1**	**2.7552**	**2.7552**	**25.51**	**<.001**
Fung Early vs Late	**1**	**1.1660**	**1.1660**	**10.80**	**0.003**
Blank Early vs Late	1	0.0057	0.0057	0.05	0.820
Fert.Split_treatment	3	0.2263	0.0754	0.70	0.560
Fert.Fung vs none	1	0.2002	0.2002	1.85	0.183
Fert.Fung Early vs Late	1	0.0126	0.0126	0.12	0.735
Fert.Blank Early vs Late	1	0.0135	0.0135	0.13	0.726
Residual	30	3.2398	0.1080		
Total	47	14.8322			

Message: the following units have large residuals.

Block 4 W_Plot 1 S_Plot 4	0.710	s.e.	0.260
Block 5 W_Plot 2 S_Plot 4	0.642	s.e.	0.260
Block 6 W_Plot 1 S_Plot 4	0.583	s.e.	0.260

Tables of effects and contrasts

Block.W_Plot.S_Plot stratum

Split_treatment contrasts

Fung vs none 0.96, s.e. 0.190, ss.div. 3.00

Fung Early vs Late 0.44, s.e. 0.134, ss.div. 6.00

Blank Early vs Late -0.03, s.e. 0.134, ss.div. 6.00

Fert.Split_treatment contrasts

Fert.Fung vs none, e.s.e. 0.268, ss.div. 1.50

Fert	Reduced	Standard
	-0.26	0.26

Fert.Fung Early vs Late, e.s.e. 0.190, ss.div. 3.00

Fert	Reduced	Standard
	-0.05	0.05

Fert.Blank Early vs Late, e.s.e. 0.190, ss.div. 3.00

Fert	Reduced	Standard
	-0.05	0.05

From the ANOVA, we see that:

- applying the fungicide late, at the flag leaf stage, gives significantly better yields ($P = 0.003$). The difference in means (for which see below) is 0.44 (± 0.134) kg/plot.
- Using fungicide has a yield advantage, on average, of ½(0.96 ± 0.190) = 0.48 ± 0.095 kg/plot ($P < 0.001$). The ½ arises because the contrast we want is ½(μ_2+μ_4)–½(μ_1+μ_3) and we currently have (μ_2+μ_4)-(μ_1+μ_3) which is essentially row 1 of the contrast matrix in the screen capture above.

Tables of means

Grand mean 2.670

Fert	Reduced	Standard
	2.356	2.984

Split_treatment	NoF_Early	F_Early	NoF_Late	F_Late
	2.446	2.689	2.415	3.130

Fert	Split_treatment	NoF_Early	F_Early	NoF_Late	F_Late
Reduced		2.220	2.333	2.142	2.728
Standard		2.672	3.045	2.688	3.532

Standard errors of differences of means

Table	Fert	Split_treatment	Fert Split_treatment
rep.	24	12	6
s.e.d.	0.0933	0.1342	0.1889
d.f.	5	30	32.32
Except when comparing means with the same level(s) of			
Fert			0.1897
d.f.			30

This design is straightforward and will not be repeated in LMM (REML).

Possible field layout for split-split-plot experiment

Block 1

1 2 3	13000
3 2 1	16000
3 1 2	10000
3 1 2	16000
2 3 1	13000
2 3 1	10000

Key to fertilizer:

1= 60 lb nitrogen
2=120 lb nitrogen
3=180 lb nitrogen

Key to irrigation:

Irrigated	
Non-irrigated	

Key to Spacing:

Spacing	13000	
Spacing	16000	
Spacing	10000	

Block 2

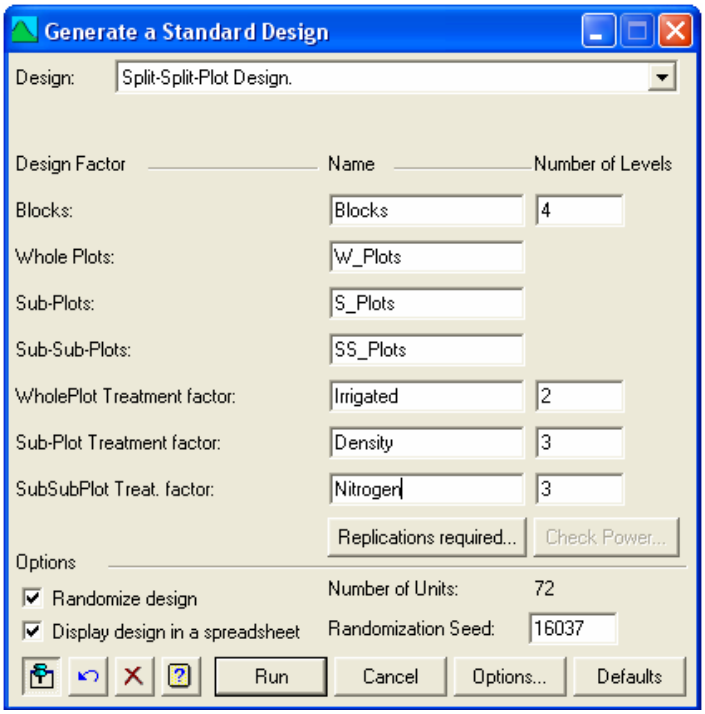
3 2 1	10000
2 3 1	16000
2 1 3	13000
3 1 2	13000
1 2 3	16000
1 3 2	10000

Block 3

2 3 1	16000
3 2 1	10000
3 1 2	13000
1 2 3	16000
1 3 2	10000
3 2 1	13000

Block 4

1 3 2	10000
1 3 2	16000
1 2 3	13000
2 1 3	10000
3 1 2	16000
3 2 1	13000



Split-split-plot design (in randomized blocks)

An experiment was conducted to determine that effects of irrigation, planting density (or stand), and fertilizer level on the yield of corn. The smallest area that could be irrigated was half a block – or one whole-plot. The two irrigation treatments were randomly allocated to the whole-plots in each of four blocks. Each whole-plot was divided into three split-plots, and with three planting densities (rates of 10,000, 13,000 and 16,000 plants acre^{-1}) randomly allocated to each. Finally, each split-plot was divided into three split-split-plots, with three fertilisers (60, 120 and 180 lb of nitrogen) randomly allocated to each.

This is quite a different layout compared to a simple RCBD in which all 18 treatment combinations could occur in any plot of each block. In this case, practical limitations dictated the layout; the penalty is a more complex analysis. The **Block Structure** comes about as follows.

- Blocks were identified in the field, so Block forms the first stratum.
- Half block areas were prepared and one of these in each block was (randomly) irrigated, forming a Block.Irrigated stratum. Irrigated and non-irrigated plot means are compared within this stratum, which is basically an RCBD with 4 blocks and 2 treatments.
- Each half-block was split into three areas and one of three spacings used (randomly) in each. Thus, we have a third stratum, Block.Irrigated.Spacing, and these units are used in constructing Spacing and Spacing.Irrigated *F*-tests.
- Each spacing strip was split into three even smaller areas and one of three fertilisers applied (randomly) in each. This gives rise to a fourth and final stratum, Block.Irrigated.Spacing.Fertiliser, and these units are used in constructing *F*-tests for the Fertiliser main effect and any interaction involving this factor.

To summarise, the **Block Structure** is

Block + Block.Irrigated + Block.Irrigated.Spacing + Block.Irrigated.Spacing.Fertiliser

which simplifies to Block/Irrigated/Spacing/Fertiliser.

Example 9 Yields of corn (bushels acre^{-1}) from Snedecor & Cochran page 328

		Non-irrigated blocks				Irrigated blocks			
Stand	Fertilizer	1	2	3	4	1	2	3	4
10,000	60	90	83	85	86	80	102	60	73
	120	95	80	88	78	87	109	104	114
	180	107	95	88	89	100	105	114	114
13,000	60	92	98	112	79	121	99	90	109
	120	89	98	104	86	110	94	118	131
	180	92	106	91	87	119	123	113	126
16,000	60	81	74	82	85	78	136	119	116
	120	92	81	78	89	98	133	122	136
	180	93	74	94	83	122	132	136	133

Analysis of variance

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block stratum	3	194.44	64.81	0.14	
Block.Irrigated stratum					
Irrigated	1	8277.56	8277.56	17.59	0.025
Residual	3	1411.78	470.59	2.03	
Block.Irrigated.Stand stratum					
Stand	2	1758.36	879.18	3.78	0.053
Irrigated.Stand	2	2747.03	1373.51	5.91	0.016
Residual	12	2787.94	232.33	2.69	
Block.Irrigated.Stand.Fertilizer stratum					
Fertilizer	2	1977.44	988.72	11.45	<.001
Irrigated.Fertilizer	2	953.44	476.72	5.52	0.008
Stand.Fertilizer	4	304.89	76.22	0.88	0.484
Irrigated.Stand.Fertilizer	4	234.72	58.68	0.68	0.611
Residual	36	3108.83	86.36		
Total	71	23756.44			

Message: the following units have large residuals.

Block 1 Irrigated Irrigated Stand 13,000	12.7	s.e.	6.2
Block 1 Irrigated Irrigated Stand 16,000	-13.6	s.e.	6.2
Block 2 Irrigated Irrigated Stand 10,000 Fertilizer 60.	14.7	s.e.	6.6
Block 3 Irrigated Irrigated Stand 10,000 Fertilizer 60.	-14.6	s.e.	6.6

Tables of means

Grand mean 99.7

Irrigated	Non-irrigated	Irrigated
	89.0	110.4

Stand	10,000	13,000	16,000
	92.8	103.6	102.8

Fertilizer	60.	120.	180.
	92.9	100.6	105.7

Irrigated	Stand	10,000	13,000	16,000
Non-irrigated		88.7	94.5	83.8
Irrigated		96.8	112.7	121.8

Irrigated	Fertilizer	60.	120.	180.
Non-irrigated		87.3	88.2	91.6
Irrigated		98.6	113.0	119.8

Stand	Fertilizer	60.	120.	180.
10,000		82.4	94.4	101.5
13,000		100.0	103.8	107.1
16,000		96.4	103.6	108.4

Irrigated	Stand	Fertilizer	60.	120.	180.
Non-irrigated	10,000		86.0	85.3	94.8
	13,000		95.2	94.2	94.0
	16,000		80.5	85.0	86.0
Irrigated	10,000		78.8	103.5	108.3
	13,000		104.7	113.2	120.2
	16,000		112.2	122.3	130.7

Standard errors of differences of means

Table	Irrigated	Stand	Fertilizer	Irrigated Stand
rep.	36	24	24	12
s.e.d.	5.11	4.40	2.68	7.21
d.f.	3	12	36	9.53
Except when comparing means with the same level(s) of				
Irrigated				6.22
d.f.				12

Table	Irrigated Fertilizer	Stand Fertilizer	Irrigated Stand Fertilizer
rep.	12	8	4
s.e.d.	5.98	5.81	8.99
d.f.	5.54	30.80	21.28
Except when comparing means with the same level(s) of			
Irrigated	3.79		8.22
d.f.	36		30.80
Stand		4.65	
d.f.		36	
Irrigated.Stand			6.57
d.f.			36
Irrigated.Fertilizer		8.22	
d.f.			30.80

Least significant differences of means (5% level)

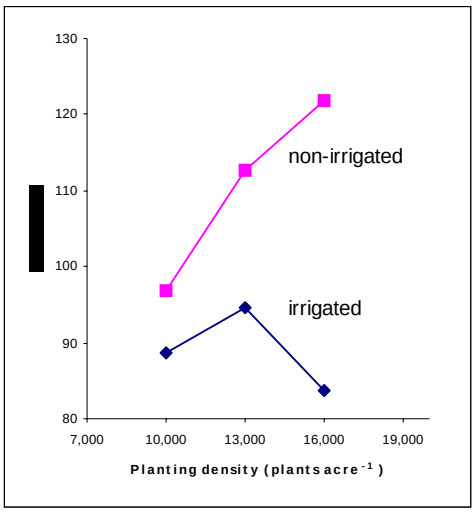
Table	Irrigated	Stand	Fertilizer	Irrigated Stand
rep.	36	24	24	12
l.s.d.	16.27	9.59	5.44	16.17
d.f.	3	12	36	9.53
Except when comparing means with the same level(s) of				
Irrigated				13.56
d.f.				12

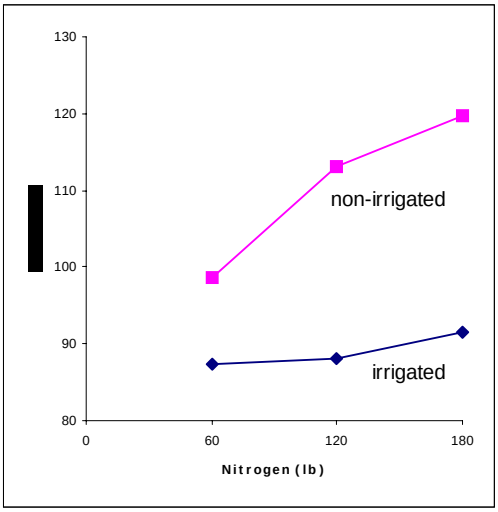
Table	Irrigated Fertilizer	Stand Fertilizer	Irrigated Stand Fertilizer
rep.	12	8	4
l.s.d.	14.92	11.85	18.67
d.f.	5.54	30.80	21.28
Except when comparing means with the same level(s) of			
Irrigated	7.69		16.76
d.f.	36		30.80
Stand		9.42	
d.f.		36	
Irrigated.Stand			13.33
d.f.			36
Irrigated.Fertilizer		16.76	
d.f.			30.80

Estimated stratum variances

Stratum	variance	effective d.f.	variance component
Block	64.81	3.000	-22.54
Block.Irrigated	470.59	3.000	26.47
Block.Irrigated.Stand	232.33	12.000	48.66
Block.Irrigated.Stand.Fertilizer	86.36	36.000	86.36

Comparing 2-way and 3-way means is now a complex procedure. Note, however, that comparing two densities (/two fertilizers) both of which were irrigated (or non-irrigated) is straightforward (the l.s.d. values are 13.56/7.69), and so on. The differences in means comes down to two significant interactions, and the following plots make these differences clear:

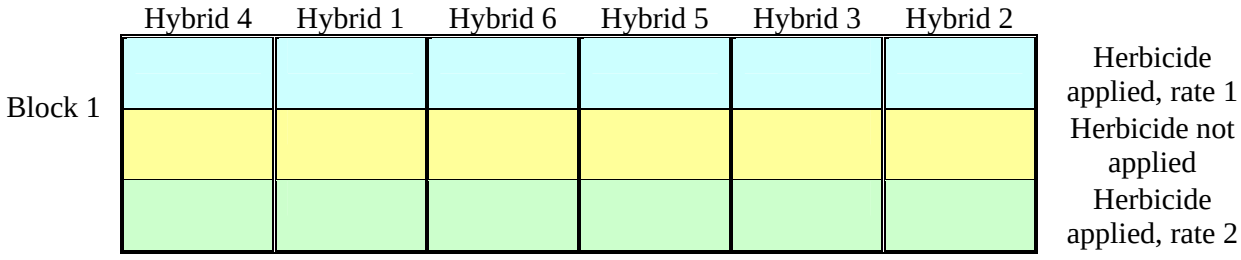




Note that the Block MS is smaller than the highest stratum Residual MS, which is unusual. When analysing via REML we would be advised to force variance components to be positive. In the analysis above, we also ignored the potential variance problem we discussed previously brought about by having varying planting densities.

Criss-cross/split-block/strip-plot design

This design has various names in the literature, but the essential difference is that a second (possibly factorially structured) treatment is randomly applied across large areas of each block, generally at right angles to the first treatment. For example, this is one block from a factorial trial in which hybrids are allocated to four plots in the block, and a herbicide treatment (absent, or one of two rates) is applied to one-three block areas stripped across the plots.



A corresponding split-plot design has the herbicide treatment applied at random to the three small plots within each whole-plot. This more complex arrangement is often the only practical way of running the experiment, but comes at the cost of greater complexity in treatment comparisons.

The levels of the herbicide treatment are also applied to large areas in each block. Thus, there are two types of whole-plots. There are now four strata: Block, Block.Hybrid, Block.Herbicide, and Block.Hybrid.Herbicide (an individual plots whose yields are measured).

Example. Curt Lee (Agro-Tech, Inc., Velva, North Dakota, USA) kindly supplied data from the following experiment on sunflower (yield in lb/acre). Hybrid number shown in each block (V1 to V7).

Block	Herbicide	V1	V2	V3	V4	V5	V6	V7
1	check	810.6	1369.7	1830.8	1335.8	1563.6	1419.5	726.8
	rate 1	776.8	1115.4	1497.0	1610.8	1637.0	1236.2	679.4
	rate 2	595.2	1175.9	1260.0	1204.3	1465.2	1172.2	669.8
		V6	V5	V4	V7	V2	V1	V3
2	rate 1	1429.4	1152.8	1150.4	744.1	1099.0	735.2	1413.9
	check	1517.5	1971.4	1737.6	643.4	916.2	608.3	1747.6
	rate 2	1696.2	1467.0	1456.1	662.8	906.7	562.5	1417.2
		V4	V2	V6	V7	V3	V1	V5
3	rate 2	1383.6	1328.2	1301.4	671.6	1805.0	709.7	1536.6
	check	1638.7	1250.8	1411.5	762.6	1827.9	601.4	1685.0
	rate 1	1727.8	1201.4	1576.8	748.2	1340.2	670.8	2193.3
		V4	V1	V7	V2	V5	V6	V3
4	rate 1	1414.4	562.3	833.6	1085.4	1480.6	1323.9	1683.9
	rate 2	1329.2	845.3	884.5	1069.9	1822.1	1277.1	1734.2
	check	1318.4	760.4	842.6	1147.4	1729.5	1212.6	1450.5

This company uses varieties and herbicides in sequential order in Block 1 as a routine, to allow growers to walk through the field and see expected differences. There is a debate as to whether the demonstration block should be used as part of the research data, but we will do so here.

Using ANOVA, the **Treatment Structure** is clearly Hybrid*Herbicide.

The **Block Structure** is slightly more complex to formulate with a shortcut. The four strata mentioned above technically is all that is needed to set up the block structure, so:

Block + Block.Hybrid + Block.Herbicide + Block.Hybrid.Herbicide

which by the rules is abbreviated to Block/(Hybrid*Herbicide).

Analysis of variance

Variate: Yield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block stratum	3	165289.	55096.		
Block.Hybrid stratum					
Hybrid	6	10890886.	1815148.	65.92	<.001
Residual	18	495649.	27536.	0.83	
Block.Herbicide stratum					
Herbicide	2	111333.	55666.	1.22	0.360
Residual	6	274730.	45788.	1.38	
Block.Hybrid.Herbicide stratum					
Hybrid.Herbicide	12	130254.	10855.	0.33	0.979
Residual	36	1192168.	33116.		
Total	83	13260309.			

Message: the following units have large residuals.

Block 2 Hybrid V6	205.	s.e. 77.
Block 2 Hybrid V5 Herbicide Check	281.	s.e. 119.
Block 2 Hybrid V5 Herbicide H1	-275.	s.e. 119.
Block 3 Hybrid V3 Herbicide H1	-277.	s.e. 119.
Block 3 Hybrid V5 Herbicide H1	355.	s.e. 119.

Tables of means

Grand mean 1231.

Hybrid	V1	V2	V3	V4	V5	V6	V7
	687.	1139.	1584.	1442.	1642.	1381.	739.

Herbicide	Check	H1	H2
	1280.	1219.	1193.

Hybrid	Herbicide	Check	H1	H2
V1		695.	686.	678.
V2		1171.	1125.	1120.
V3		1714.	1484.	1554.
V4		1508.	1476.	1343.
V5		1737.	1616.	1573.
V6		1390.	1392.	1362.
V7		744.	751.	722.

Standard errors of differences of means

Table	Hybrid	Herbicide	Hybrid Herbicide
rep.	12	28	4
s.e.d.	67.7	57.2	128.6
d.f.	18	6	54.21
Except when comparing means with the same level(s) of			
Hybrid			132.1
d.f.			41.33
Herbicide			125.0
d.f.			53.62

Least significant differences of means (5% level)

Table	Hybrid	Herbicide	Hybrid Herbicide
rep.	12	28	4
l.s.d.	142.3	139.9	257.8
d.f.	18	6	54.21
Except when comparing means with the same level(s) of			
Hybrid			266.8
d.f.			41.33
Herbicide			250.7
d.f.			53.62

Estimated stratum variances

Stratum	variance	effective d.f.	variance component
Block	55096.4	3.000	708.9
Block.Hybrid	27536.0	18.000	-1859.9
Block.Herbicide	45788.3	6.000	1810.4
Block.Hybrid.Herbicide	33115.8	36.000	33115.8

There are strongly significant differences ($P < 0.001$) among hybrids, but no interaction or herbicide effect. The interpretation is therefore straightforward. In the presence of a significant interaction, individual means will have to be compared using one of three L.S.D> values, none of which leads to a strict t test (notice the non-integer degrees of freedom).

Notice also the negative Block.Hybrid stratum variance. When using LMM (REML) we would set that to be non-negative, as will be seen with other examples.

More complex field designs: a split-strip plot experiment

This experiment was used by Schabenberger and Pierce (2001), page 599, to illustrate a REML analysis in SAS. Four soybean cultivars were used as whole-plots in each of four replicate blocks. Two row spacings (9", 18") were used, each applied at random to half of each whole-plot in a vertical direction. In addition, five target plant populations (60, 120, …, 300 thousand per acre) were used, each applied at random to one-fifth of each whole-plot in a horizontal direction. The field plan therefore appears as follows.

Block 1

AG4601		AG4701		AG3701		AG3601	
120	120	300	300	60	60	300	300
300	300	240	240	240	240	60	60
180	180	60	60	300	300	180	180
240	240	120	120	180	180	120	120
60	60	180	180	120	120	240	240
9	18	9	18	9	18	9	18

Block 2

AG4601		AG3701		AG3601		AG4701	
180	180	180	180	240	240	120	120
60	60	240	240	60	60	300	300
240	240	120	120	120	120	60	60
300	300	60	60	300	300	180	180
120	120	300	300	180	180	240	240
9	18	9	18	18	9	18	9

Block 3

AG3701		AG4701		AG3601		AG4601	
60	60	60	60	120	120	120	120
180	180	180	180	240	240	60	60
240	240	300	300	180	180	180	180
300	300	120	120	60	60	300	300
120	120	240	240	300	300	240	240
18	9	18	9	18	9	18	9

Block 4

AG3701		AG4601		AG3601		AG4701	
60	60	120	120	60	60	120	120
300	300	240	240	180	180	300	300
240	240	300	300	120	120	180	180
120	120	180	180	300	300	60	60
180	180	60	60	240	240	240	240
18	9	9	18	18	9	9	18

There are five strata in this experiment, and the block structure is the sum of these terms:

1. Block stratum
2. Block.Cultivar stratum
3. Block.Cultivar.Row stratum
4. Block.Cultivar.Plant stratum
5. Block.Cultivar.Row.Plant stratum

The yields for the corresponding treatments are as follows.

	Column							
Row	1	2	3	4	5	6	7	8
1	19.5	26.2	26.4	32.5	23.4	21.3	29.4	32.0
2	23.9	23.3	25.7	24.2	24.0	25.9	25.2	26.1
3	22.0	21.9	19.0	16.3	27.6	28.1	31.5	29.1
4	19.4	20.0	22.9	21.7	21.8	21.9	26.6	25.0
5	19.0	15.8	26.0	27.9	25.9	22.0		
6	23.4	22.4			26.0	32.9	21.9	23.9
7	20.6	19.7	26.9	25.9		27.9	31.4	26.5
8	28.2	27.9	25.6	24.8	32.1	34.2	24.5	21.4
9	25.9	28.5	23.0	23.3	26.5	40.2	28.9	30.5
10	22.0	30.3	28.8	30.4	25.1	35.9	28.0	23.3
11	17.8	22.3	16.5	19.3	22.0	28.9	23.6	21.6
12	20.9	23.3	23.3	26.6	27.9	36.9	17.2	20.8
13	26.5	26.2	28.0	30.4	27.0	32.1	24.9	24.6
14	25.9	24.2	24.2	30.1	23.2	26.9	33.0	35.3
15	22.8	19.0	22.0	26.9	26.9	34.5	30.7	25.3
16	16.2	13.0	20.4	23.6	21.4	17.6	25.2	21.1
17	26.5	25.4	21.0	24.4	23.3	26.9	26.7	26.1
18	27.5	21.9	23.2	26.2	16.0	23.2	25.5	23.5
19		17.9	24.4	21.7	21.3	27.1	14.7	15.6
20	19.8	22.2	15.6	17.7	26.2	32.4	26.0	26.4

There are six missing yields. GenStat will analyse the data via **General Analysis of Variance**. However, missing values are inserted and therefore F tests are inflated upwards. In addition, there may well be a change in variance across both row spacings and plant populations, and there may well be a better spatially correlated model to use, so it is preferable to use LMM (REML).

Treatment Structure: Cultivar*RowSpacing*PlantPop

Block Structure:

Block+Block.Cultivar+Block.Cultivar.Row+Block.Cultivar.Plant+Block.Cultivar.Row.Plant

Here is part of the ANOVA output.

Analysis of variance

Variate: Yield

Source of variation	d.f.	(m.v.)	s.s.	m.s.	v.r.	F pr.
Block stratum	3		419.420	139.807	7.24	
Block.Cultivar stratum						
Cultivar	3		531.821	177.274	9.18	0.004
Residual	9		173.723	19.303		

Block.Cultivar.PlantPop stratum						
PlantPop	4		1173.923	293.481	33.72	<.001
Cultivar.PlantPop	12		139.997	11.666	1.34	0.230
Residual	46	(2)	400.406	8.704	2.20	
Block.Cultivar.RowsSpacing stratum						
RowsSpacing	1		38.125	38.125	3.47	0.087
Cultivar.RowsSpacing	3		185.301	61.767	5.63	0.012
Residual	12		131.682	10.974	2.77	
Block.Cultivar.PlantPop.RowsSpacing stratum						
PlantPop.RowsSpacing	4		18.891	4.723	1.19	0.327
Cultivar.PlantPop.RowsSpacing	12		122.997	10.250	2.59	0.011
Residual	44	(4)	174.146	3.958		
Total	153	(6)	3388.902			

Message: the following units have large residuals.

Block 2 Cultivar AG3601 PlantPop 120.	4.51	s.e. 1.58
Block 2 Cultivar AG3601 PlantPop 240.	-3.89	s.e. 1.58
Block 3 Cultivar AG4601 PlantPop 300.	4.41	s.e. 1.58
Block 1 Cultivar AG3601 RowsSpacing 9.	-2.02	s.e. 0.91
Block 1 Cultivar AG3601 RowsSpacing 18.	2.02	s.e. 0.91
Block 2 Cultivar AG3601 PlantPop 120. RowsSpacing 9.	-2.60	s.e. 1.04
Block 2 Cultivar AG3601 PlantPop 120. RowsSpacing 18.	2.60	s.e. 1.04

May well be due to a changing variance in the field. ANOVA assumes constant variance

Estimated stratum variances

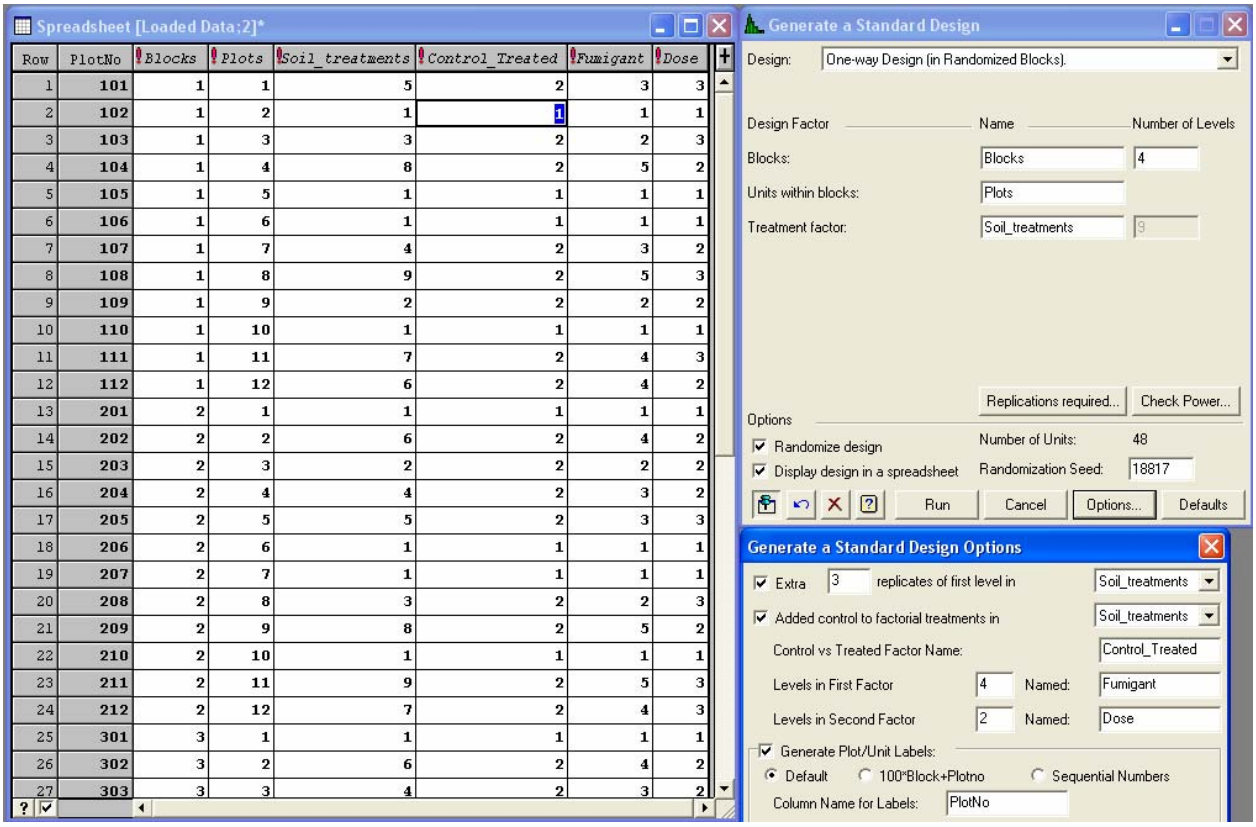
Stratum	variance	effective d.f.	variance component
Block	139.807	3.000	3.013
Block.Cultivar	19.303	9.000	0.358
Block.Cultivar.PlantPop	8.704	46.000	2.373
Block.Cultivar.RowsSpacing	10.974	12.000	1.403
Block.Cultivar.PlantPop.RowsSpacing	3.958	44.000	3.958

Two-way design (in randomized blocks) *plus* a control *plus* extra replication of the control *plus* a covariate

An experiment was laid out in four randomized blocks, designed to determine the effectiveness of four soil fumigants in keeping down the numbers of eelworms in the soil. The fumigants were chlorodinitrobenzene (CN), carbon disulphide jelly (CS) and two proprietary preparations, "Cymag" (CM) and "Seekay" (CK). Each fumigant was tested both in a single and double dose. There was a 9th treatment, viz a control (no fumigant): four plots in each block were left untreated. The purpose was to supply an accurate standard against which the performance of the fumigants was measured. The fumigants were ploughed in during spring, after which a crop of oats was sown. Before and after harvest, 400g of soil was taken from each plot and the number of eelworm cysts counted.

Generating a random design in GenStat prior to running the experiment

Although there is a 4 × 2 factorial structure (Fumigant × Dose), once the control treatment is added the treatment structure is a bit more complex. Since the control is "no fumigant", there is no way of having a single and double dose of "nothing". So initially, we need to think of this as a one-way treatment design with (4×2+1) levels. We have 9 treatments, 8 of which are factorially structured. So in the **Design** menu we select **One-way (in Randomized Blocks)**, set the number of treatments to 9, then go into **Options**. We set up a 1 *df* contrast for the treated versus untreated plots, and set up the 4 × 2 factorial structure in that menu. In addition, we can get GenStat to replicate the Control treatment 4 times (an additional 3 replicates per block):



Row	PlotNo	Blocks	Plots	Soil_treatments	Control_Treated	Fumigant	Dose
1	101	1	1	5	2	3	3
2	102	1	2	1	1	1	1
3	103	1	3	3	2	2	3
4	104	1	4	8	2	5	2
5	105	1	5	1	1	1	1
6	106	1	6	1	1	1	1
7	107	1	7	4	2	3	2
8	108	1	8	9	2	5	3
9	109	1	9	2	2	2	2
10	110	1	10	1	1	1	1
11	111	1	11	7	2	4	3
12	112	1	12	6	2	4	2
13	201	2	1	1	1	1	1
14	202	2	2	6	2	4	2
15	203	2	3	2	2	2	2
16	204	2	4	4	2	3	2
17	205	2	5	5	2	3	3
18	206	2	6	1	1	1	1
19	207	2	7	1	1	1	1
20	208	2	8	3	2	2	3
21	209	2	9	8	2	5	2
22	210	2	10	1	1	1	1
23	211	2	11	9	2	5	3
24	212	2	12	7	2	4	3
25	301	3	1	1	1	1	1
26	302	3	2	6	2	4	2
27	303	3	3	4	2	3	2

Notice that GenStat creates a factor (with 1s and 2s) to compare treated and untreated plots: a 1 represents an untreated plot (throughout the spreadsheet) and 2 a treated plot. Then, in the **Output** window, the **Treatment Structure** is shown as Control_Treated/(Fumigant*Dose). Remember that the / operator has a higher priority than the * operator, so the parentheses are important in this structure, to force the / operator on all three terms in the factorial structure. This might be clearer with the following explanation.

If you examine the other factor levels in the spreadsheet you will see that the combination of fumigant number (2, 3, 4, 5) and dose number (2 = single, say, and 3 = double) occurs only when the Control_Treated level is 2 (ie treated). Fumigant and dose treatments are "nested" inside the treated versus control contrast. The effect is that, in the ANOVA, apparent first-order interactions (like Control_Treated.Fumigant) are actually main effects and the apparent second-order interaction (Control_Treated.Fumigant.Dose) is first-order interaction

Analysis of variance

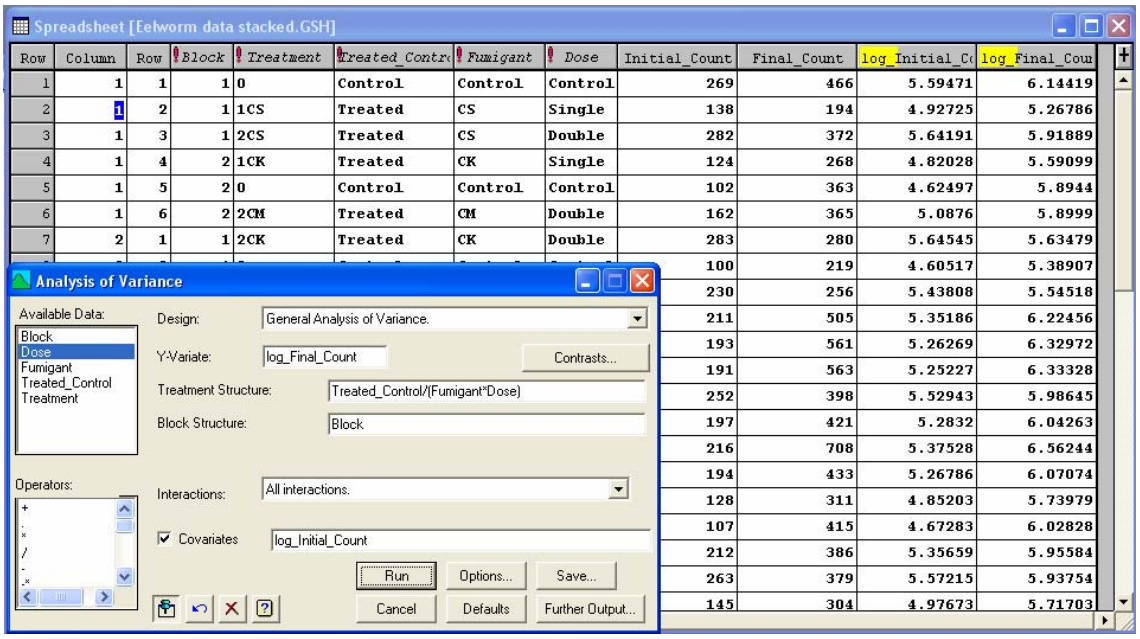
Source of variation	d.f.	**think of this component as:**
Blocks stratum	3	Blocks
Blocks.Plots stratum		
Control_Treated	1	**Control_Treated** contrast
Control_Treated.Fumigant	3	**Fumigant** main effect (for treated plots)
Control_Treated.Dose	1	**Dose** main effect (for treated plots)
Control_Treated.Fumigant.Dose	3	**Fumigant.Dose** interaction (for treated plots)
Residual	36	

Example 10. Dose (1 = single, 2 = double) and type of fumigant, and eelworm counts (initial above final) in field position, from Cochran and Cox page 46

0	2CK	1CN	1CM	2CM	2CS	2CK	0
269	283	252	212	95	127	80	134
466	280	398	386	199	166	142	590
1CS	0	0	2CM	1CK	1CN	1CM	0
138	100	197	263	107	89	41	74
194	219	421	379	236	332	176	137
2CS	1CK	0	2CN	0	0	2CN	1CS
282	230	216	145	88	25	42	62
372	256	708	304	356	212	308	221
1CK	0	1CS	2CK	2CK	0	1CK	1CM
124	211	194	222	193	209	109	153
268	505	433	408	292	352	132	454
0	2CN	2CS	1CN	0	2CN	2CS	0
102	193	128	42	29	9	17	19
363	561	311	222	254	92	28	106
2CM	0	1CM	0	1CS	1CN	0	2CM
162	191	107	67	23	19	44	48
365	563	415	338	80	114	268	298

Had we used GenStat to design the trial, we need only add the two data columns (final and initial counts) and **Run** the analysis via the **Spread** menu.

The analysis is performed in GenStat by initially setting up two factor columns: a Block factor with 4 levels and a soil Treatment factor with 9 levels. Then in Options, we set up a factor to identify treated and untreated plots, and two treatment factor columns, Dose (Single, Double) and Fumigant (CK=Seekay, CM=Cymag, CN=chlorodinitrobenzene, CS= carbon disulphide jelly). We have the added complication that the control is replicated 4 times in each block.



Row	Column	Row	Block	Treatment	Treated_Contr	Fumigant	Dose	Initial_Count	Final_Count	log_Initial_C	log_Final_Cou
1	1	1	1	0	Control	Control	Control	269	466	5.59471	6.14419
2	1	2	1	1CS	Treated	CS	Single	138	194	4.92725	5.26786
3	1	3	1	2CS	Treated	CS	Double	282	372	5.64191	5.91889
4	1	4	2	1CK	Treated	CK	Single	124	268	4.82028	5.59099
5	1	5	2	0	Control	Control	Control	102	363	4.62497	5.8944
6	1	6	2	2CM	Treated	CM	Double	162	365	5.0876	5.8999
7	2	1	1	2CK	Treated	CK	Double	283	280	5.64545	5.63479
								100	219	4.60517	5.38907
								230	256	5.43808	5.54518
								211	505	5.35186	6.22456
								193	561	5.26269	6.32972
								191	563	5.25227	6.33328
								252	398	5.52943	5.98645
								197	421	5.2832	6.04263
								216	708	5.37528	6.56244
								194	433	5.26786	6.07074
								128	311	4.85203	5.73979
								107	415	4.67283	6.02828
								212	386	5.35659	5.95584
								263	379	5.57215	5.93754
								145	304	4.97673	5.71703

There are some issues to sort out with data like these.

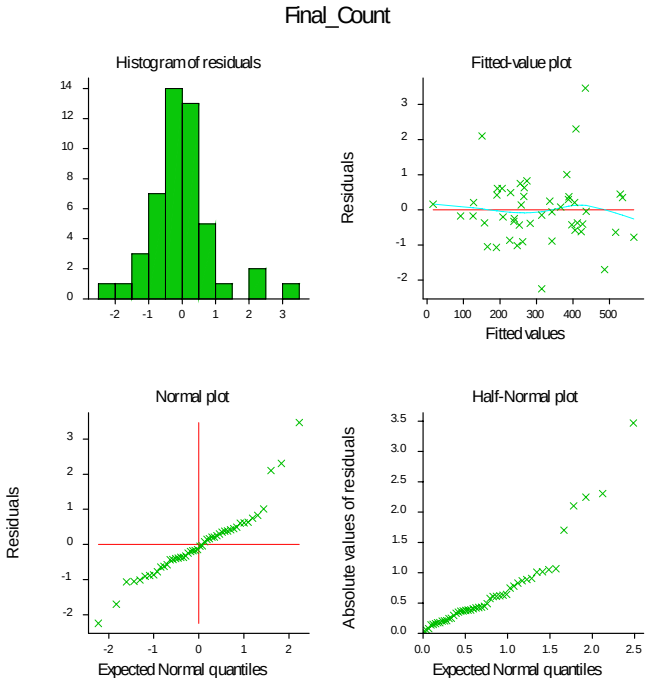
- The data are not normally distributed. It is possible that they are Poisson, in which case the variance is the same as the mean, and if the means change then so must the variances. Hence a logistic regression might be preferable to ANOVA. Alternatively, we could transform the data to achieve approximate constant variance. For Poisson data the square root transformation used to be recommended. With large counts, a log transformation may be better: differences in means are then more easily back-transformed and interpreted.

- The final counts may well depend on the initial worm counts: if the worms are not uniformly spread at the start of the experiment, then differences at the end may be misleading. We should incorporate initial counts as a covariate. If we log-transform final counts, then we should log-transform initial counts as well.

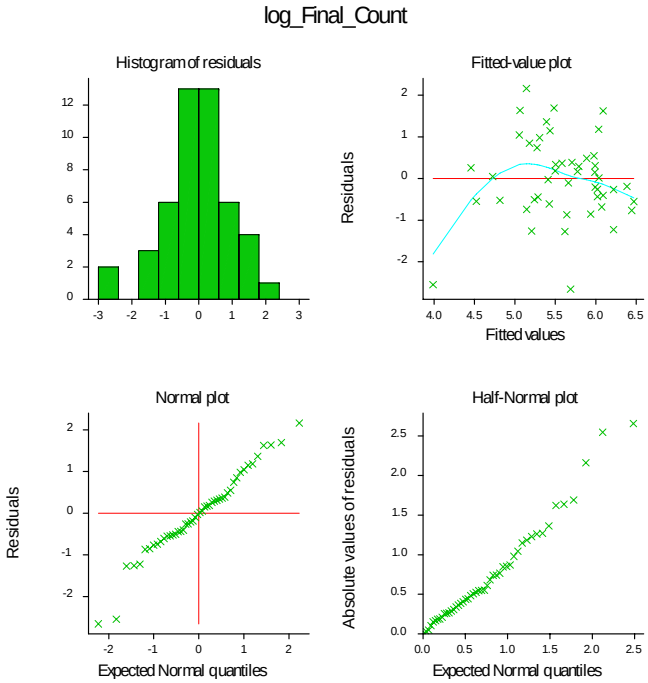
- The Poisson distribution tends to a normal distribution with increasing mean count. Thus, we could use LMM (REML) assuming an approximate normal distribution with a changing variance, and possibly a spatially correlated error structure. Notice that the four blocks are formed as a 2 × 2 layout in the field, and in each block the plots are arranged in a 3×4 grid. If there is a gradient left to right and top to bottom across blocks, we might expect a gradient left to right and/or top to bottom *within* the blocks. What has become

known as a Row-Column analysis might then remove a trend in the field more successfully than the 2 × 2 block layout.

We will look at some of these actions. Firstly, an analysis of final counts with initial counts as a covariate shows a distinct fanning in the standardised residuals:



We therefore analyse the data log-transformed:



Analysis of log(final counts), with log(initial counts) as a covariate

Analysis of variance (adjusted for covariate)

Variate: log_Final_Count
Covariate: log_Initial_Count

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Block stratum						
Covariate	1	4.76145	4.76145	11.74		0.076
Residual	2	0.81127	0.40563	4.23	4.58	
Block.*Units* stratum						
Treated_Control	1	1.16420	1.16420	12.13	1.00	0.001
Treated_Control.Fumigant	3	2.08349	0.69450	7.24	0.92	<.001
Treated_Control.Dose	1	0.04506	0.04506	0.47	0.99	0.498
Treated_Control.Fumigant.Dose	3	0.31977	0.10659	1.11	1.00	0.358
Covariate	1	5.21084	5.21084	54.31		<.001
Residual	35	3.35793	0.09594		2.48	
Total	47	16.92526				

Message: the following units have large residuals.

Block 3 *units* 11	-0.770	approx. s.e.	0.264
Block 4 *units* 8	-0.654	approx. s.e.	0.264

Tables of means (adjusted for covariate)

Variate: log_Final_Count
Covariate: log_Initial_Count

Grand mean 5.582

Treated_Control	Control	Treated
	5.805	5.470
rep.	16	32

Treated_Control	Dose	Control	Double	Single
Control		5.805		
Treated			5.432	5.508

Treated_Control	Fumigant	Control	CK	CM	CN	CS
Control		5.805				
	rep.	16				
Treated			5.195	5.667	5.798	5.220
	rep.		8	8	8	8

Treated_Control	Dose	Fumigant	Control	CK	CM	CN	CS
Control	Control		5.805				
		rep.	16				
Treated	Double			5.216	5.589	5.882	5.041
		rep.		4	4	4	4
	Single			5.174	5.745	5.713	5.399
		rep.		4	4	4	4

Standard errors of differences of means

Table	Treated_Control	Treated_Control Dose	Treated_Control Fumigant	Treated_Control Dose Fumigant	
rep.	unequal	16	unequal	unequal	
d.f.	35	35	35	35	
s.e.d.			0.1596	0.2226	min.rep
	0.0949	0.1097	0.1382	0.1760	max-min
			0.1129X	0.1113X	max.rep

(No comparisons in categories where s.e.d. marked with an X)

Least significant differences of means (5% level)

Table	Treated_Control	Treated_Control Dose	Treated_Control Fumigant	Treated_Control Dose Fumigant	
rep.	unequal	16	unequal	unequal	
d.f.	35	35	35	35	
l.s.d.			0.3241	0.4520	min.rep
	0.1927	0.2227	0.2806	0.3573	max-min
			0.2291X	0.2260X	max.rep

(No comparisons in categories where l.s.d. marked with an X)

Estimated stratum variances (adjusted for covariate)

Variate: log_Final_Count
Covariate: log_Initial_Count

Stratum	variance	effective d.f.	variance component
Block	0.3029	2.746	0.0173
Block.*Units*	0.0953	35.254	0.0953

Clearly initial counts go a long way to explaining differences in final counts. Incorporating the initial counts as a covariate:

- is strongly significant ($P<0.001$);
- reduces the Residual MS from 0.2380 to less than *half* that value, 0.0959;
- more accurately tests whether treated plots have significantly lower eelworm cysts than control plots, taking initial counts into account ($P=0.001$);
- detects that the type of fumigant is very important ($P<0.001$).

A very important feature of interpreting means of log-transformed data should be mentioned.

- The back-transformed mean of log-transformed data is the *geometric mean* of the original data. For log-normal data, the geometric mean is a much better estimate of a "typical" value than the arithmetic mean, since the importance of very large values in the calculation is greatly reduced.

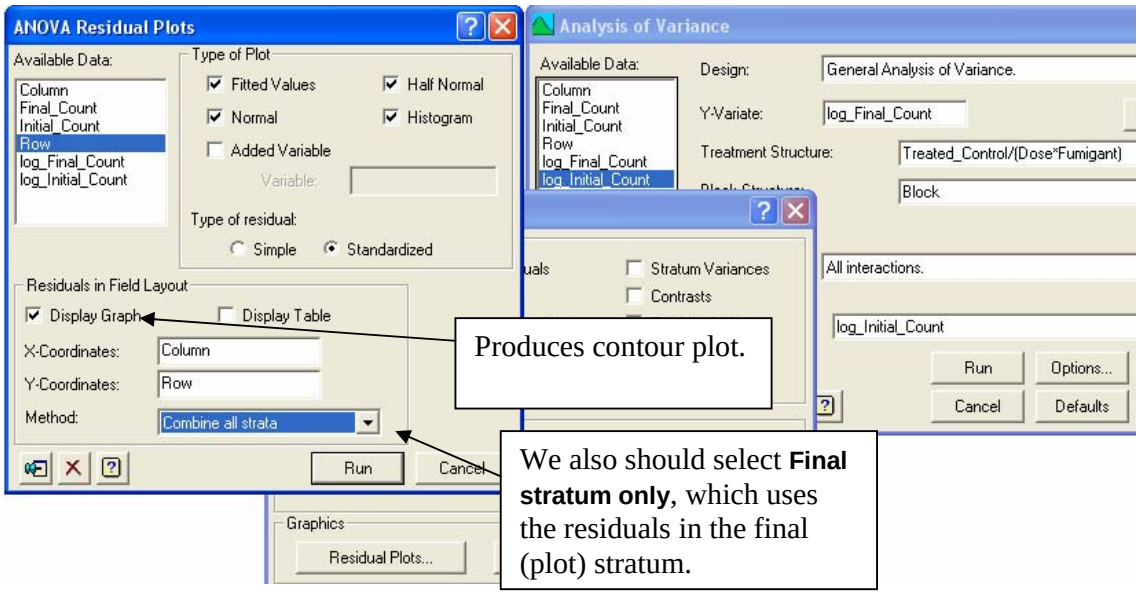
- The back-transformed difference in two means of log-transformed data is the *ratio of the two geometric means* of the original data. For example, for the carbon disulphide jelly (CS) fumigant, the effect of a single compared to a double dose is 5.399 – 5.041 = 0.358 on the log-scale. This back-transforms to 1.43. Thus, a plot with a single dose of carbon disulphide jelly applied typically has 43% more eelworms cysts than a similar plot with a double dose.

- The l.s.d. value for the comparison above is 0.4520 and this is based on 35 *df* for which *tcrit* is 2.030. The value to add and subtract to the difference in means above is 2.030×0.4520 = 0.918. The 95% confidence interval on the log-scale is (-0.560, 1.276). Back-transforming the end points gives a confidence interval for the ratio of (0.571, 3.581). Thus, while a plot with a single dose of carbon disulphide jelly applied typically has 43% more eelworms cysts than a similar plot with a double dose, we are only 95% confident that this ratio is between just over a half (0.571×), to a little more than three and a half times (3.581×). Other differences are treated similarly.

There is still one other plot to check: a plot of the residuals *in field position*, with an accompanying contour plot. To obtain this plot, we need to supply *two variates*: the X-coordinate and the Y-coordinate of each plot in field position. Imagine an X-Y coordinate system overlaying the experimental site (consisting of plots in a 6×8 layout) with the origin in the bottom left hand corner of the site.

Y=6	0 269 466	2CK 283 280	1CN 252 398	1CM 212 386	2CM 95 199	2CS 127 166	2CK 80 142	0 134 590
Y=5	1CS 138 194	0 100 219	0 197 421	2CM 263 379	1CK 107 236	1CN 89 332	1CM 41 176	0 74 137
Y=4	2CS 282 372	1CK 230 256	0 216 708	2CN 145 304	0 88 356	0 25 212	2CN 42 308	1CS 62 221
Y=3	1CK 124 268	0 211 505	1CS 194 433	2CK 222 408	2CK 193 292	0 209 352	1CK 109 132	1CM 153 454
Y=2	0 102 363	2CN 193 561	2CS 128 311	1CN 42 222	0 29 254	2CN 9 92	2CS 17 28	0 19 106
Y=1	2CM 162 365	0 191 563	1CM 107 415	0 67 338	1CS 23 80	1CN 19 114	0 44 268	2CM 48 298
(0,0)	X=1	X=2	X=3	X=4	X=5	X=6	X=7	X=8

The data are ordered down column 1 first, so we need to set up Y as (6, 5, 4, 3, 2, 1, 6, 5, …) and X as (1, 1, 1, 1, 1, 1), …, (8, 8, 8, 8, 8, 8) by right-clicking on each column and selecting **Fill** (with the **Starting Value** for Y being 6, the **Ending Value** 1 and **Increment** -1).

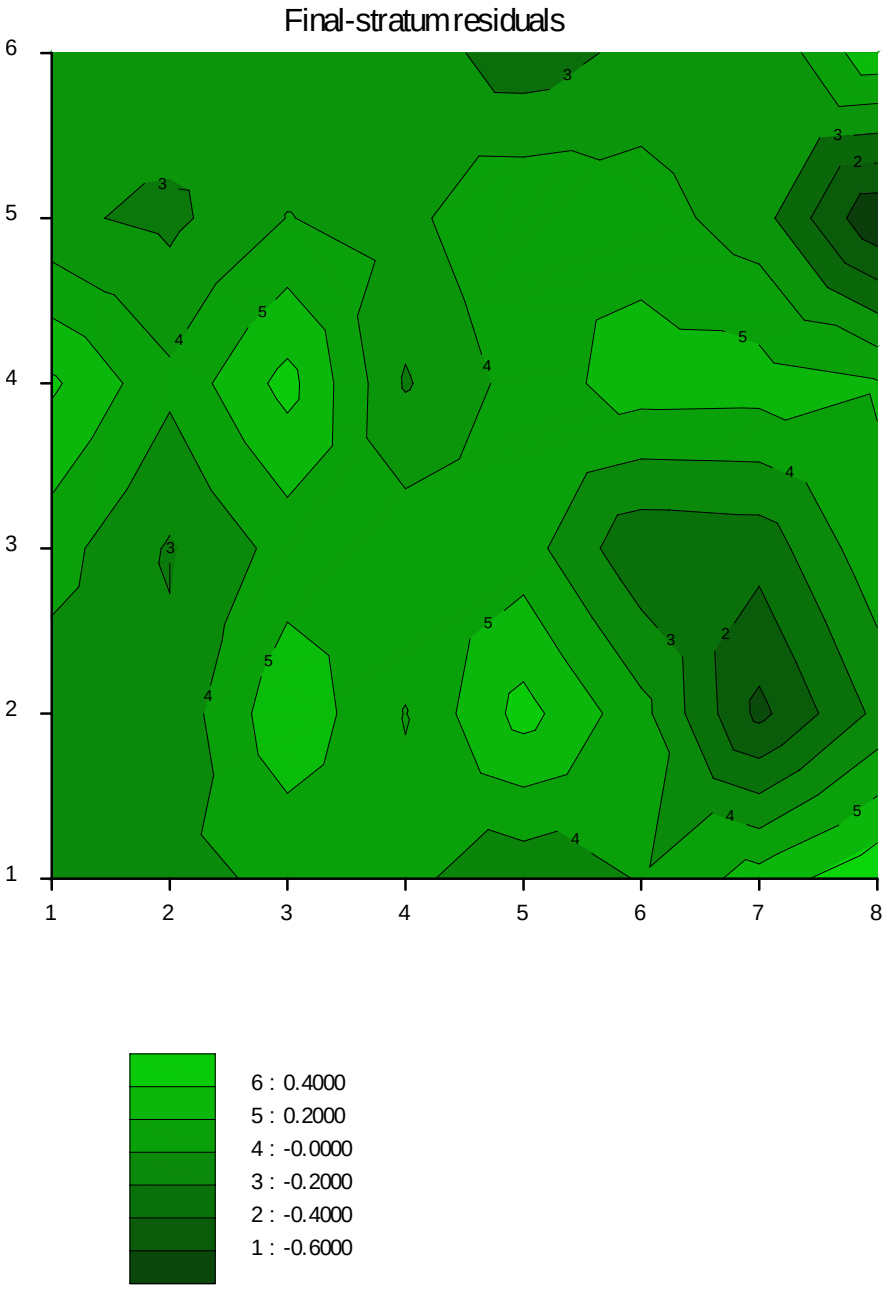


.

The residuals in field position are:

	Final_stratum_residuals							
_['Column']	1	2	3	4	5	6	7	8
_['Row']								
6	**-0.077**	**-0.027**	**-0.104**	**-0.066**	**-0.326**	**-0.130**	**-0.191**	0.343
5	**-0.157**	**-0.253**	0.004	**-0.054**	0.190	0.099	**-0.114**	**-0.770**
4	0.434	0.047	0.470	**-0.219**	0.084	0.302	0.295	0.218
3	0.087	**-0.222**	0.079	0.124	0.093	**-0.356**	**-0.324**	0.141
2	**-0.127**	**-0.142**	0.350	**-0.007**	0.474	0.066	**-0.654**	**-0.152**
1	**-0.175**	**-0.055**	0.039	0.048	**-0.140**	0.012	0.284	0.555

These residuals should be random +/- across the field, since block effects are supposed to have dealt with any gradient in the field. Within each block the residuals will add to 0. Given that, deciding if the residuals are random in the field is fairly subjective. The accompanying contour plot smoothes over the individual residuals, but again, deciding if the light areas represent plots whose fitted counts are consistently larger than the observed counts is again subjective.



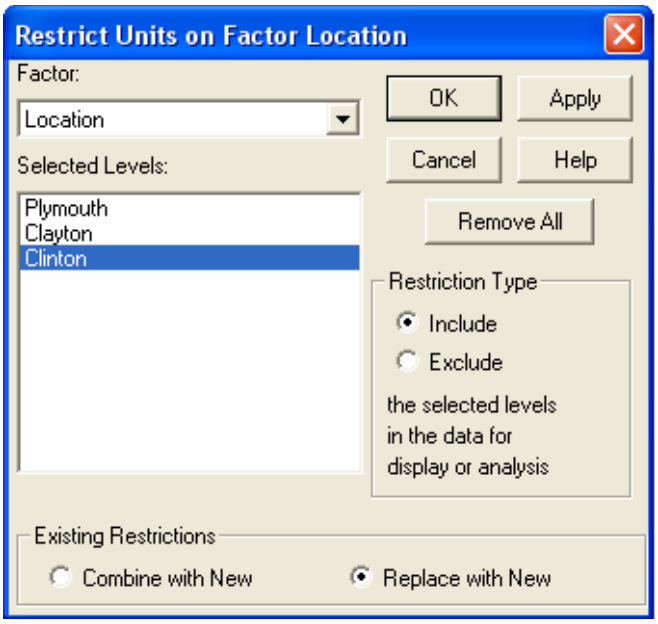
We will return to this data set in the LMM (REML) section of the manual.

Multi-site experiments

Example 11. Twelve strains of soybeans were compared in separate randomized blocks at three locations in North Carolina. Data from Steel and Torrie page 399, 400

	Plymouth			Clayton			Clinton		
Variety	BL 1	BL 2	BL 3	BL 1	BL 2	BL 3	BL 1	BL 2	BL 3
Tracy	1307	1365	1542	1178	1089	960	1583	1841	1464
Centennial	1425	1475	1276	1187	1180	1235	1713	1684	1378
N72-137	1289	1671	1420	1451	1177	1723	1369	1608	1647
N72-3058	1250	1202	1407	1318	1012	990	1547	1647	1603
N72-3148	1546	1489	1724	1345	1335	1303	1622	1801	1929
R73-81	1344	1197	1319	1175	1064	1158	1800	1787	1520
D74-7741	1280	1260	1605	1111	1111	1099	1820	1521	1851
N73-693	1583	1503	1303	1388	1214	1222	1464	1607	1642
N73-877	1656	1371	1107	1254	1249	1135	1775	1513	1570
N73-882	1398	1497	1583	1179	1247	1096	1673	1507	1390
N73-1102	1586	1423	1524	1345	1265	1178	1894	1547	1751
R75-12	911	1202	1012	1136	1161	1004	1422	1393	1342

The first thing to decide is whether the variation at each site is consistent. Three separate RCBD analyses produced the following Residual MS estimates. These are obtained by clicking in the spreadsheet, selecting **Restrict/Filter** > **To Groups (factor levels)**. Select the **Location** factor and each level with **Replace with new**.



Location	df	Residual MS
Plymouth	22	24149
Clayton	22	12124
Clinton	22	22851
Average	**66**	**19708**

Do we have any right to combine the three estimates into a pooled estimate with 66 *df*? Since we assume normal data and independent experiments across locations, these can be tested by Bartlett's variance homogeneity test, (Chi-square 2.90 on 2 degrees of freedom: probability 0.234).

Next, locations are really included to make better breeding choices, so interest lies in interpreting the Strain.Location interaction. However, locations are unreplicated, (as are blocks at each location). In the LMM (REML) section of the manual we re-analyse the data with Location a random factor.

Next, block 1 at one location is not the same as block 1 at a different location. Hence we need to combine blocks within locations, thereby obtaining (3-1)×3 = 6 *df*.

The **Block Structure** we would recommend is then Location+ Block.Location, and the **Treatment Structure** Strain+Strain.Location. This is the analysis that such a general ANOVA produces.

Analysis of variance

Variate: Yield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Location stratum	2	3113626.	1556813.	134.87	
Location.Block stratum	6	69256.	11543.	0.59	
Location.Block.*Units* stratum					
Strain	**11**	**925090.**	**84099.**	**4.27**	**<.001**
Location.Strain	22	532900.	24223.	1.23	0.256
Residual	66	1300723.	19708.		
Total	107	5941596.			

Message: the following units have large residuals.

Location Clayton Block 1	54.	s.e.	25.
Location Plymouth Block 1 *units* 9	287.	s.e.	110.
Location Plymouth Block 3 *units* 9	-282.	s.e.	110.
Location Clayton Block 3 *units* 3	299.	s.e.	110.

Tables of means

Variate: Yield

Grand mean 1403.

Strain	Centennial	D74-7741	N72-137	N72-3058	N72-3148	N73-1102
	1395.	1406.	1484.	1331.	1566.	1501.

Strain	N73-693	N73-877	N73-882	R73-81	R75-12	Tracy
	1436.	1403.	1397.	1374.	1176.	1370.

Location	Strain	Centennial	D74-7741	N72-137	N72-3058	N72-3148
Plymouth		1405.	1395.	1473.	1299.	1599.
Clayton		1402.	1308.	1652.	1308.	1529.
Clinton		1378.	1517.	1327.	1385.	1570.

Location	Strain	N73-1102	N73-693	N73-877	N73-882	R73-81
Plymouth		1524.	1476.	1391.	1506.	1300.
Clayton		1464.	1476.	1414.	1375.	1334.
Clinton		1517.	1357.	1405.	1309.	1488.

Location	Strain	R75-12	Tracy
Plymouth		1055.	1418.
Clayton		1302.	1277.
Clinton		1172.	1415.

Standard errors of differences of means

Table	Strain	Location Strain
rep.	9	3
d.f.	66	66
s.e.d.	66.2	114.6

Least significant differences of means (5% level)

Table	Strain	Location Strain
rep.	9	3
d.f.	66	66
l.s.d.	132.1	228.9

Estimated stratum variances

Variate: Yield

Stratum	variance	effective d.f.	variance component
Location	1556813.0	2.000	42924.2
Location.Block	11542.7	6.000	-680.4
Location.Block.*Units*	19707.9	66.000	19707.9

Notice that the Location.Block MS (11543) is unexpectedly smaller than the Residual MS (19708) which gives rise to the negative variance component above. When then data are analysed using LMM (REML), it is advisable to force a zero bound for this variance component.

The Location MS is much larger than the Residual MS, indicating large variation in the overall mean yields over the three locations. Differences in means between the strains, however, are consistent across these locations (P=0.256).

Experiments repeated annually

Snedecor and Cochran presented an analysis of asparagus yields taken from an experiment in which planting occurred in 1929 and cuttings commenced in 1930. Data are available for four years from the same plots. This was a randomized block, with four plots in each block. The four plots corresponded to cuttings taken on June 1 each year, but for three of the plots additional cuttings were taken (but not analysed). The intent of the analysis was to detect if repeated cutting of asparagus affected plant vigour.

Example 12. Asparagus yields from four annual cuttings, from Snedecor and Cochran, page 330-2.

		Cuttings ceased			
Block	Year	Jun 1	Jun 15	Jul 1	Jul 15
1	1930	230	212	183	148
	1931	324	415	320	246
	1932	512	584	456	304
	1933	399	386	255	144
2	1930	216	190	186	126
	1931	317	296	295	201
	1932	448	471	387	289
	1933	361	280	187	83
3	1930	219	151	177	107
	1931	357	278	298	192
	1932	496	399	427	271
	1933	344	254	239	90
4	1930	200	150	209	168
	1931	362	336	328	226
	1932	540	485	462	312
	1933	381	279	244	168

This experiment will be re-analysed using repeated measures in the next section. Clearly, the same plot is repeatedly measured, and hence yields for the same plot are almost surely correlated across years.

Snedecor and Cochran overcame that problem by analysing linear, quadratic and cubic (orthogonal) components, which was then a way of overcoming the correlated nature of the data.

We present an analysis below along the lines of the previous multi-site analysis, simply to enable a comparison of the LMM (REML) analysis that we do in the next section. We put Year in the **Block Structure** simply to force the ANOVA not to produce a *P*-value for Year.

The blocks in each year are the same: the experiment was conducted over several years and the design used in the first year was studied unchanged over four years. The **Block Structure** we used here is Year + Block + Block.Year, or simply Block*Year.

Bartlett's test for homogeneity of variances confirms no significant differences (P=0.481) between the annual variance estimates, 460.5, 929.2, 1368, 825.9 each with 9 df and an average of 895.9:

Asparagus mean yields clearly change over years

Analysis of variance

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block stratum	3	30169.6	10056.5	17.77	
Block.Year stratum					
Year	3	518721.9	172907.3	305.61	<.001
Residual	9	5092.0	565.8	0.63	
Block.Year.*Units* stratum					
CuttingTime	3	241376.6	80458.9	89.80	<.001
Year.CuttingTime	9	51177.5	5686.4	6.35	<.001
Residual	36	32255.4	896.0		
Total	63	878793.0			

Tables of means

Grand mean 290.6

Year	1930.	1931.	1932.	1933.
	179.5	299.4	427.7	255.9

CuttingTime	Jun_01	Jun_15	Jul_1	Jul_15
	356.6	322.9	290.8	192.2

Year	CuttingTime	Jun_01	Jun_15	Jul_1	Jul_15
1930.		216.2	175.8	188.8	137.2
1931.		340.0	331.2	310.2	216.2
1932.		499.0	484.8	433.0	294.0
1933.		371.2	299.8	231.2	121.2

Standard errors of differences of means

Table	Year	CuttingTime	Year CuttingTime
rep.	16	16	4
s.e.d.	8.41	10.58	20.17
d.f.	9	36	44.81
Except when comparing means with the same level(s) of			
Year			21.17
d.f.			36

Estimated stratum variances

Stratum	variance	effective d.f.	variance component
Block	10056.54	3.000	593.17
Block.Year	565.78	9.000	**-82.55**
Block.Year.*Units*	895.98	36.000	895.98

We have another look at a superior analysis in the next section.

Linear Mixed Models (REML)

What are Linear Mixed Models?

When we do an ANOVA the treatment structure generally consists of treatment effects of particular interest. In a block design, blocks are usually regarded as random, in the sense that the behaviour of the treatments in the blocks used in the experiment is assumed to be consistent across all possible blocks. In the RCBD model

yield = overall mean + block effect + variety effect + Error

the *overall mean* is a fixed effect: we are interested in the overall yield, across all varieties and all blocks; the *block effect* is random, as suggested; The *variety effect* could be fixed (this experiment is designed to answer questions on these and only these varieties), or random (from the randomly chosen varieties we want to generalize to a population of varieties); and the error term is random, assumed normal $N(0, \sigma^2)$.

This model is an example of a **Linear Mixed Model**. The general framework for a LMM is

yield = overall mean + fixed effects + random effects + Error

The *random effects* are assumed independent of the residual *Error*. They are assumed to be normally distributed with means of 0, and with variance parameters separate from σ^2 in the *Error* distribution.

Theoretical statisticians usually write a LMM as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

Here **y** is a vector of data, $\boldsymbol{\beta}$ a vector of fixed effects to be estimated, **u** a vector of random effects, $\boldsymbol{\varepsilon}$ a vector of errors. **X** is the design matrix identifying the layout for fixed effects, and **Z** is the design matrix identifying the layout for random effects

We mention this only because in GenStat's output it often mentions the following:

var(y) = Sigma2(ZGZ'+R), i.e. relative to the residual variance, Sigma2.

In other words, when setting up a LMM like the one above, the variance-covariance matrix of the data vector involves Sigma2 **G**, the variance-covariance matrix of the random vector **u**, and Sigma2 **R**, the variance-covariance matrix of the residual error vector **ε**. GenStat allows two ways to parameterize this structure: set Sigma2=1 and estimate the variances and covariances contained in **G** and in **R**; or let Sigma2=σ^2, in which case the variances and covariances contained in G are to be multiplied by σ^2. This is controlled by the PARAMETERIZATION option in the REML procedure.

The LMM clearly covers an experiment in which there are no other random effects apart from the *Error* term (so leave the **Random Model** of LMM blank), as well as no fixed effects (so leave the **Fixed Model** of LMM blank). This covers ANOVA and regression models considered to date.

What is REML?

Early monographs such as Steel and Torrie and Snedecor and Cochran introduced the idea of estimating parameters like the mean and standard deviation of a normal population without reference to the concept of **maximum likelihood** (ML). To justify the use of (n-1) as the divisor in estimating σ^2 they talk about bias or sampling with and without replacement. Some authors talk about using n as the divisor when calculating the *population* variance and (n-1) when calculating the *sample* variance. Indeed, scientific calculators have σ_n and σ_{n-1} buttons. Excel has VARP and VAR formulae for the two sorts of variances (which we label s_n^2 and s_{n-1}^2 respectively), and STDEVP and STDEV for the equivalent standard deviations.

Maximum likelihood estimates are obtained by maximizing what is known as the likelihood, or equivalently the log-likelihood. For the normal population $Y \sim N(\mu, \sigma^2)$ the likelihood of a random sample of size n is simply the product of the normal density function evaluated at each of the data points. The log-likelihood is therefore

$$\log L = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{Y_i - \mu}{\sigma}\right)^2 .$$

It is straightforward to show that the ML estimators of μ and σ^2 are

$$\hat{\mu} = \bar{y} ,$$

$$\hat{\sigma}^2_{ML} = \frac{\sum_{i=1}^{n}(Y_i - \bar{y})^2}{n} = s_n^2 .$$

Maximum likelihood estimators do not necessarily have optimal small-sample properties. It is true that the ML estimate of σ^2 is biased, in the sense that the mean over repeated sampling settles down on the value (n-1)/n × σ^2 rather than on σ^2 itself.

The idea of **Residual Maximum Likelihood** (REML) is only a couple of decades old. The idea is this:

We take the likelihood and partition into two components. The first component is a likelihood of one or more statistics and involves all *fixed* parameters like μ (and may involve variance parameters as well). The second component is a likelihood of statistics that involves variance parameters relating to *random* effects (and hence, no fixed parameters). *We then maximize each component separately.* The estimates of the variance parameters are known as REML estimates.

For the normal example above, the first component turns out to be the likelihood for the sample mean $\bar{y}$, and the second likelihood that of the variates that give rise to the sample variance. Specifically,

$$\log L = \left[-\frac{1}{2}\ln\left(2\pi\sigma^2/n\right) - \frac{1}{2}\left(\frac{\bar{y}-\mu}{\sigma/\sqrt{n}}\right)^2 \right] + \left[\frac{n-1}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2}\ln(n) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{Y_i-\bar{y}}{\sigma}\right)^2 \right]$$

The separate solutions are

$$\hat{\sigma}^2_{REML} = \frac{\sum_{i=1}^{n}(Y_i-\bar{y})^2}{n-1} = s^2_{n-1},$$

$$\hat{\mu} = \bar{y}.$$

Thus, the familiar estimate for σ^2 is actually a REML estimate, and this estimate is unbiased.

It turns out that the REML estimate of σ^2 in a CRD model is the Residual MS of the CRD ANOVA. For more complex models, the REML estimate is less biased than the ML estimate.

Using a REML algorithm in experiments involving fixed effects and random effects is not restricted to independent data, or data with the same variance in any one stratum. It is an extremely flexible estimating tool, and has become the standard way of analyzing data from agricultural trials.

REML estimates for a single sample

Let us take the coefficients of digestibility (%) for sheep from Example 1 on page 33.

Sheep	57.8	56.2	61.9	54.4	53.6	56.4	53.2

For these data, $\hat{\mu} = 56.214$, $s^2_n = 7.727$ and $s^2_{n-1} = 9.015$ (each to 3 decimals). Let us see what GenStat gives when we take the data and use LMM (REML).

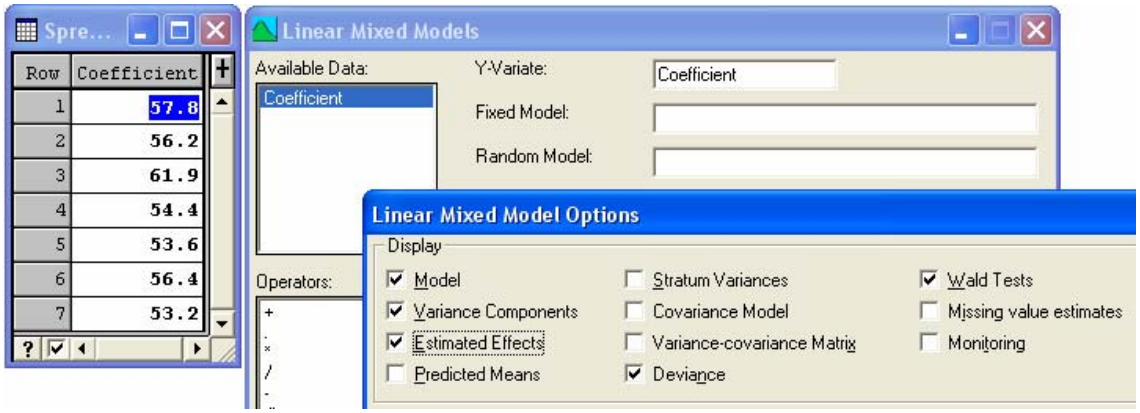
Firstly, unless otherwise specified in **Options**, GenStat will always fit a constant term (μ) and, if you do not include an error term, it will add one for you.

The model for these data is

coefficient of digestibility = μ + *Error*

where *Error* ~ N(0, σ^2). The parameter μ is a fixed parameter, and the parameter σ^2 is the only parameter in the random part of the model.

We can simply enter the *coefficient of digestibility* variate and leave the **Fixed Model** and **Random Model** blank. We need to click **Predicted Means** in **Options**, and as a general rule, click **Deviance** as well.



REML variance components analysis

Response variate: Coefficient
Fixed model: Constant
Number of units: 7

Residual term has been added to model

Non-sparse algorithm with Fisher scoring

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
Residual		**Identity**	**Sigma2**	**9.015**	**5.205**

Deviance: -2*Log-Likelihood

Deviance	d.f.
21.14	5

Note: deviance omits constants which depend on fixed model fitted.

Table of effects for Constant

56.21 Standard error: 1.135

The REML estimate of variance is 9.015 ± 5.205.

The estimate of mean is 56.21 ± 1.135.

Notice that a "Residual term has been added to model". We can deliberately put an error term if we wish (for example, if we decide to include a correlation into our model). For a sample of size *n* there are *n* error terms, each being independent with the same distribution, $N(0, \sigma^2)$. We therefore need to set up a factor that contains *n* levels corresponding to the *n* data values. Actually, GenStat has a device to do this for us: simply type **'*Units*'** in the **Random Model**. The analysis is identical apart from replacing the message

```
Residual term has been added to model
```

with

```
'*units*' used as residual term
```

Deviance

In the previous output GenStat reported a deviance of 24.14 with 5 *df*. Deviance plays the role that the Residual SS played in ANOVA. The deviance that GenStat prints out is proportional to -2×LogL, where LogL is the log-likelihood of the variance components. (The actual definition actually has the constant 2π removed): Thus, for the normal sample,

$$\text{Deviance} = (n-1)\ln\left(\hat{\sigma}^2\right) + \ln(n) + \sum_{i=1}^{n}\left(\frac{Y_i - \bar{y}}{\hat{\sigma}}\right)^2 .$$

Deviance really is only used to compare models. Asymptotically, a *change in deviance* for one (nested) model compared to a larger model is χ^2 and the degrees of freedom to use are the *change in df*. The nested model arises by replacing in the larger model the new parameters that given in the null hypothesis. We demonstrate this technique with the two *t* test examples.

REML estimates for two samples (equal variances)

We firstly use the coefficient of digestibility data for sheep and steers to illustrate using change in deviance.

Firstly, we stack the data, providing a 2-level factor to identify the animal type. We also need to anticipate how to construct a residual term. If we provide a factor called say Replicate that identifies the seven sheep and 6 steers, then in an ANOVA context the **Block Structure** is Replicate.Animal. Using Replicate.Animal as the **Random Model** in LMM (REML) allows us later to declare the Animal part of Replicate.Animal to have a **Diagonal** variance-covariance matrix. This is GenStat's way of allowing different variances for each level of Animal (hence for sheep and for steers).

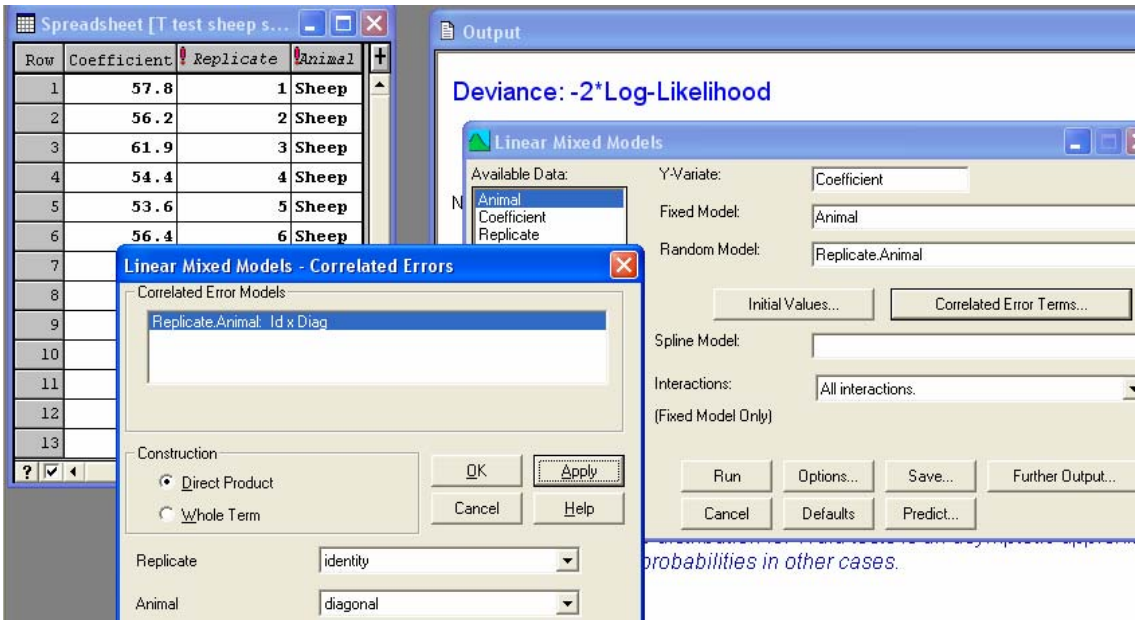
The model is

coefficient of digestibility = μ + *Animal* + *Error*

where *Animal* is a fixed effect, and *Error* is random component with the first 7 elements being normal N(0, σ^2_{sheep}) and the last 6 elements normal N(0, σ^2_{steers}). The null hypothesis that tests whether the two variances are equal is H_0: $\sigma^2_{sheep} = \sigma^2_{steers}$. Under this hypothesis, the reduced model is one with equal variances. The test of equality of means in the output for that scenario should then be equivalent to the unpaired *t* test. If the variances are judged to be different, then the test of equality of means in the output for the maximal model should then be equivalent to the approximate *t* test.

	Estimates of parameters in model	**Deviance**	**d.f.**	***P* value**
maximal model	σ^2_{sheep} = 9.0 (6 *df*), σ^2_{steers} =5.3 (5 *df*)	36.64	10	
reduced model	σ^2 = weighted average = 7.326	36.27	9	
change in deviance		0.37	1	0.543

Clearly there is no need to use an unequal variance model. (As a precursor to the *t* test, GenStat tested the variances and returned a *P* value of 0.58 based on the *F* distribution – see page 32.) The output for the final (equal variance) model is the following.



REML variance components analysis

Response variate: Coefficient
Fixed model: Constant + Animal
Random model: Replicate.Animal
Number of units: 13

Replicate.Animal used as residual term

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
	Replicate.Animal	Identity	Sigma2	7.326	3.124

Deviance: -2*Log-Likelihood

Deviance	d.f.
36.64	10

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Animal	11.18	1	11.18	<0.001

Table of predicted means for Constant

58.73 Standard error: 0.753

Table of predicted means for Animal

Animal	Sheep	Steers
	56.21	**61.25**

Standard error of differences: **1.506**

In **Options**, there are two choices of methods to use for maximizing the log-likelihood, **AI** (the default) and **Fisher scoring**. If you are unsure of what degrees of freedom to use in constructing a confidence interval for the difference in means, you can request **Stratum Variances** *but only if you use Fisher scoring*, and only if the model is identity. An **Identity** model is basically uncorrelated with constant variance, and hence the degrees of freedom should be the same as the ANOVA would give you. Here, we obtain the additional output:.

Approximate stratum variances

Stratum	variance	effective d.f.
Replicate.Animal	7.326	11.00

The df correspond to (7-1)+(6-1) = 11 as expected.

Notice that

- the Wald/df statistic, 11.18, is the same as the *F* value for the test of means on page 36. The *P* value there was 0.007, here it is <0.001. The reason is that the Wald test is an asymptotic test and is based on χ^2. If you do have a balanced design with an **Identity** model, then you are quite at liberty to replace the *P* value of Wald/df and base the new one on an *F* distribution.

- The same means and the same sed value are obtained by both analyses.

REML estimates for two samples (unequal variances)

The unequal variance *t* test data from page 37 is an example of where the test fails.

	Good soil	Poor soil
	5.9	7.6
	3.8	0.4
	6.5	1.1
	18.3	3.2
	18.2	6.5
	16.1	4.1
	7.6	4.7
count	7	7
mean	10.914	3.943
sd	6.334	2.636
variance	40.125	6.950

Test for equality of sample variances

Test statistic F = 5.77 on 6 and 6 d.f.
Probability (under null hypothesis of equal variances) = 0.05

Difference of means: 6.971
Standard error of difference: 2.593

95% confidence interval for difference in means: (0.9937, 12.95)

Test of null hypothesis that mean of Fine_gravel with Soil = good is equal to mean with Soil = poor

	Estimates of parameters in model	Deviance	d.f.	P value
maximal model	σ^2_{good} = 40.1 (6 *df*), σ^2_{poor} =6.9 (6 *df*)	49.68	10	
reduced model	σ^2 = weighted average = 7.326	53.79	11	
change in deviance		4.11	1	0.043

Again, the change in deviance is based on an asymptotic χ^2 distribution, not the F distribution. The unequal variance output is as follows.

REML variance components analysis

Response variate: Fine_gravel
Fixed model: Constant + Soil
Random model: Rep.Soil
Number of units: 14

Rep.Soil used as residual term with covariance structure as below

Covariance structures defined for random model

Covariance structures defined within terms:

Term	Factor	Model	Order	No. rows
Rep.Soil	Rep	Identity	0	7
	Soil	Diagonal	2	2

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
	Rep.Soil	Sigma2	40.12	23.17	
	Rep	Identity	-	-	-
	Soil	Diagonal	d_1	**1.000**	fixed
			d_2	**0.1732**	0.1414

Deviance: -2*Log-Likelihood

Deviance	d.f.
49.68	10

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Soil	7.23	1	7.23	0.007

Table of predicted means for Constant

7.429 Standard error: 1.2966

Table of predicted means for Soil

Soil	good	poor
	10.914	3.943

Standard error of differences: 2.593

This form of the output presents the two variance estimates as *multiples* (that GenStat calls "gammas") of an overall estimate "Sigma2 40.12". The multiples are 1.000 for level 1 (good soil) and 0.1732 for level 2 (poor soil). Hence 40.12 is the estimate of variance for good soil and 0.1732×40.12 = 6.950 for poor soil. There is an option to allow the *actual variance estimates* to be printed instead. You need to capture the REML line in the **Input** window, copy it to a new window, adjust the PARAMETERIZATION parameter and re-run the

analysis. The **AI** method, not **Fisher scoring**, needs to be the method used to obtain this parameterization.

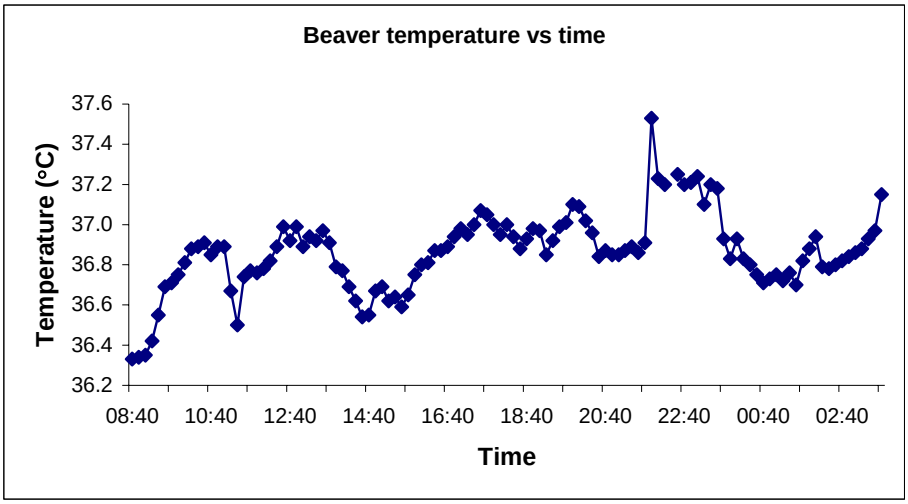
```
REML [PRINT=model,components,means,deviance,waldTests; PSE=differences;\
PARAMETERIZATION=sigmas;MVINCLUDE=*; METHOD=ai; MAXCYCLE=20000] Fine_gravel
```

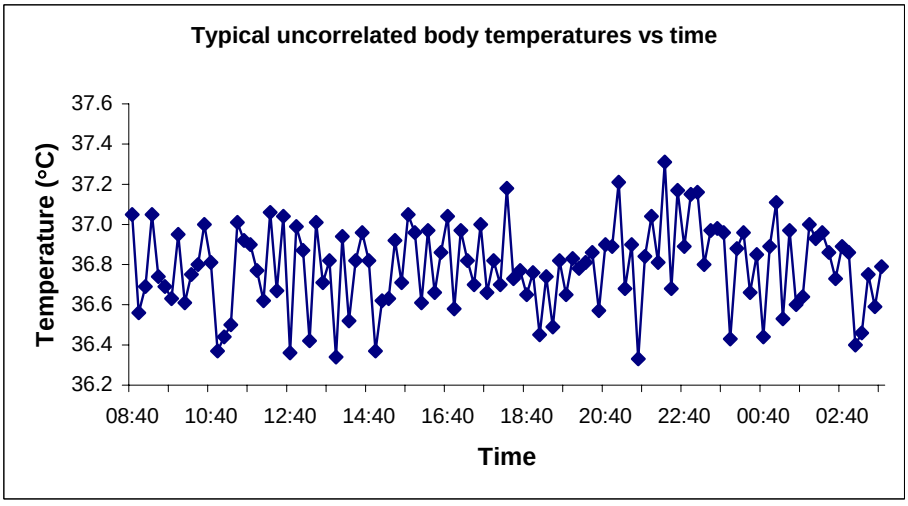
Change to previous output:

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
Rep.Soil	Sigma2	1.000	fixed		
	Rep	Identity	-	-	-
	Soil	Diagonal	d_1	**40.12**	23.17
			d_2	**6.950**	4.012

Plots for serially correlated data



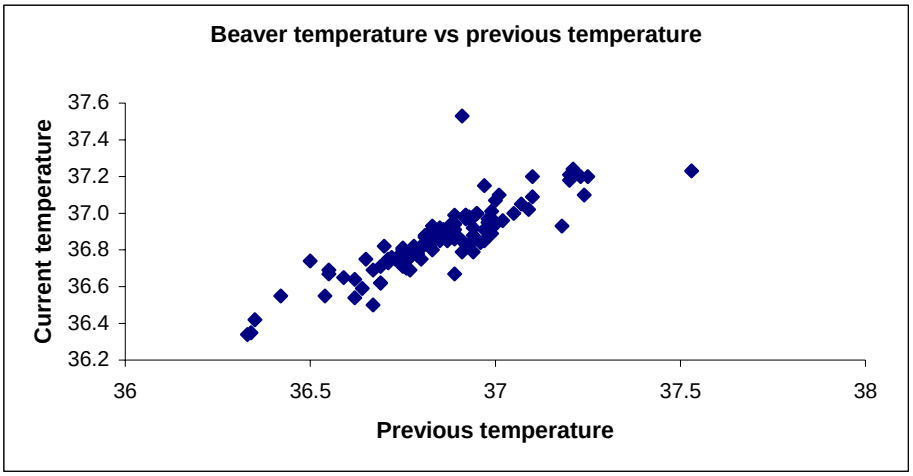


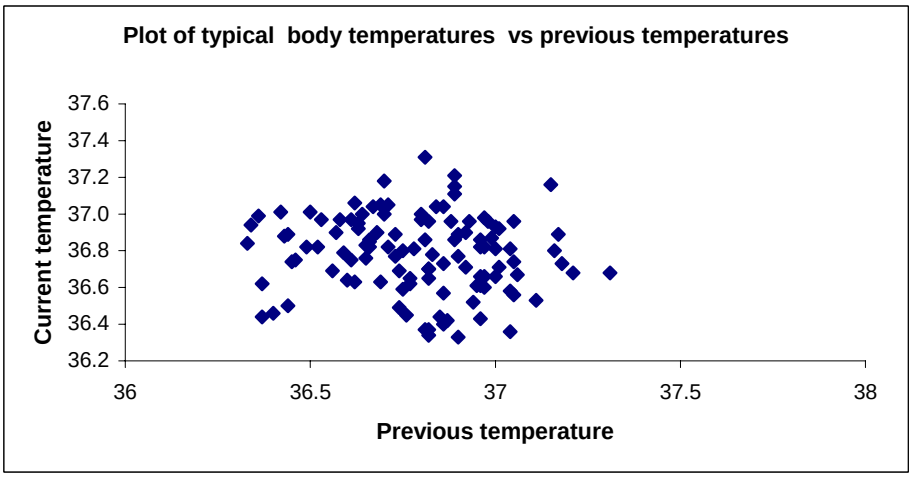
Common correlation models

This manual is not a place to describe in great detail the concepts of correlated data over time. What follows is a very brief summary of the major concepts used in modelling temporal or spatial data via LMM (REML).

A good example to illustrate serially correlated data is the famous beaver body temperatures taken every 10 minutes, taken from *Case Studies in Biometry* (Lange *et al*. 1994). A plot of these temperatures for a single animal is shown on the left hand page, and for comparison, a plot of notional temperatures randomly sampled from a normal distribution at each time with the same mean and variance as the overall beaver data had. It is clear that there is an essential difference between the two plots. The randomly generated data show random +/- values compared to the mean at any stage (a spiky plot), unlike the smooth beaver plot.

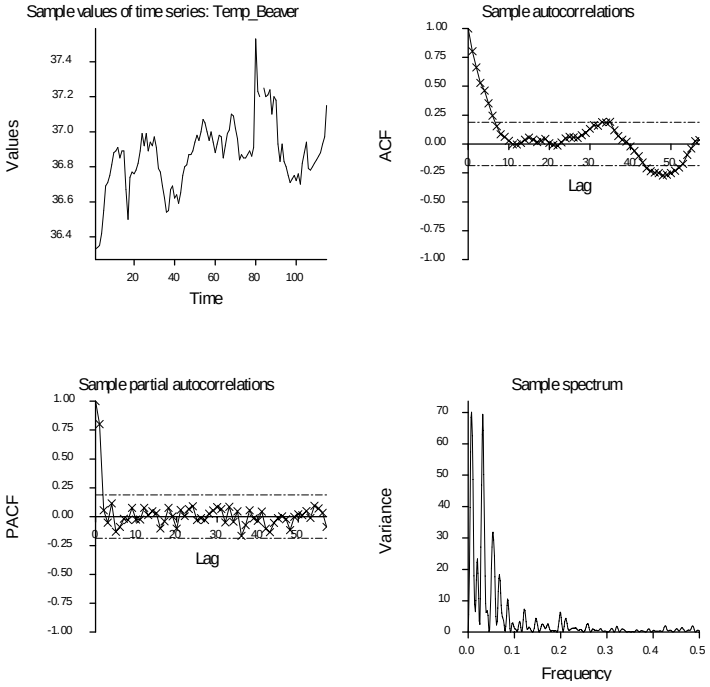
To emphasise this even more, here are plots of the temperatures at time *t* plotted against the temperatures at time *t*-1.



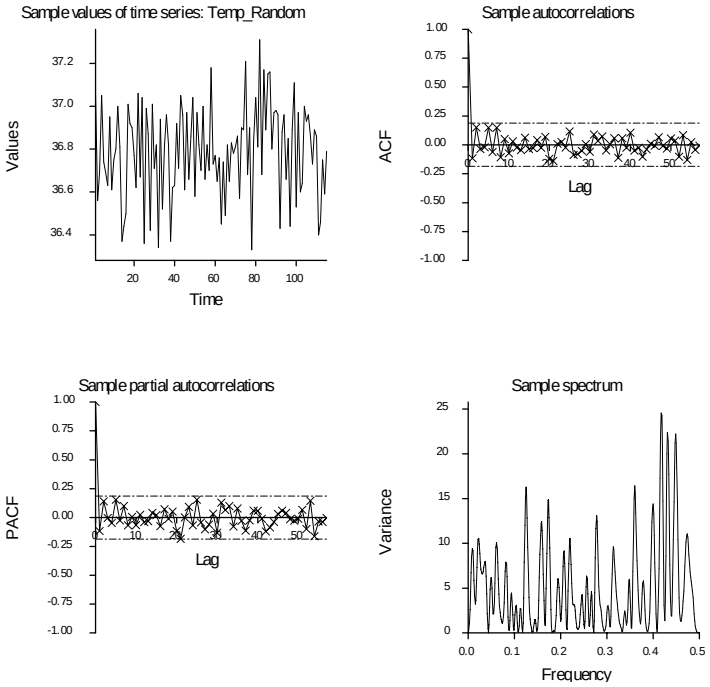


Beaver temperatures are clearly correlated in time.

Beaver data



Random data



There are various ways that we can model this correlation structure. In time series literature, they define autoregressive (AR) models, moving average (MA) models, combinations of these known as ARMA models for data, or ARIMA models for differences in data values.

It is not always easy to identify which structure to use for a given data set. Two types of correlations are helpful. The set of these is known as the *autocorrelation function* (ACF) and *partial autocorrelation function* (PACF).

The *autocorrelation* $\boldsymbol{r_1}$ is the correlation between successive pairs of data, $\{Y_t, Y_{t-1}\}$, lagged by one time period.

The *autocorrelation* $\boldsymbol{r_2}$ is the correlation between successive pairs of data, $\{Y_t, Y_{t-2}\}$, , lagged by two time periods, … and so on for other autocorrelations.

The *partial autocorrelation* $\boldsymbol{r_{2.1}}$ is the correlation between successive pairs of data, $\{Y_t, Y_{t-2}\}$, *adjusted for the effect of* Y_{t-1}. It is like performing a regression of Y_t on Y_{t-1}, saving the residuals and calculating a correlation of these with Y_{t-2}. This is extended to higher-order lags as well. As a starting point it is conventional to define $\boldsymbol{r_{1.0}}$ as $\boldsymbol{r_1}$, the first autocorrelation.

	Beaver	Random	Beaver	Random
Unit	ACF	ACF	PACF	PACF
1	1	1	1	1
2	**0.802**	-0.117	**0.802**	-0.117
3	0.663	0.151	0.055	0.139
4	0.527	-0.036	-0.053	-0.004
5	0.463	-0.021	0.115	-0.047
6	0.353	0.149	-0.130	0.153
7	0.245	-0.063	-0.089	-0.026
8	0.153	0.148	-0.017	0.099
9	0.085	-0.107	-0.030	-0.068
10	0.061	0.050	0.077	0.005
11	0.027	-0.074	-0.024	-0.066
12	-0.004	0.029	-0.026	0.024
13	-0.004	-0.023	0.075	-0.042
14	0.009	-0.046	0.013	-0.031
15	0.036	0.061	0.046	0.039
16	0.056	-0.037	0.030	0.021
17	0.039	-0.029	-0.103	-0.074
18	0.015	0.041	-0.042	0.071
19	0.029	-0.025	0.076	-0.011
20	0.044	0.068	0.002	0.051

Both AC and PAC functions have specific forms for the different types of correlation structures.

For the beaver data and the random temperature data, the ACF and PACF values are obtained as follows. Select **Time Series** > **Data Exploration** and the data to be investigated. In Options, choose **Partial Autocorrelation Functions** if these are required. The default should include ACF and PACF plots.

ACF and PACF plots for beaver temperatures and random temperatures are given on the left hand page for the first twenty lags. The horizontal lines on each plot are confidence bands around zero values.

There is clearly a difference. For the beaver data, the ACF declines steadily while the PACF values are basically zero (note that, by definition, lag-1 correlations are unity). For the random data, both ACF and PACF functions are zero.

So how can we incorporate serial correlations into our models? We start by allowing the error term in a linear model to be correlated. To do this, define

ε_t^* = random error at time *t*, assumed $N(0,\sigma^2)$ *but correlated*.

A particularly simple model is $Y_t = \mu + \varepsilon_t^*$. This is termed a ***stationary model*** as the mean does not change systematically over time. We will deal with stationery models first.

An *autoregressive* (**AR**) error structure of lag-*k* is one where *the random error at time t depends linearly on k of the previous errors*. Two very common AR models in agricultural data analysis are lag 1 and lag 2 processes, so we confine our attention to these. Thus:

AR(1), or AR1 in the LMM (REML) menu: a lag 1 process: (condition: $\phi_1^2 < 1$)

For this process the dependent error depends only on the previous dependent error. Text books use α for the slopes in these types of regression, but GenStat uses ϕ so we use this symbol below.

$\varepsilon_t^* = \phi_1 \varepsilon_{t-1}^* + \varepsilon_t$ where ε_t is an (unobserved) independent error

For a stationary model, $\varepsilon_t^* = (Y_t - \mu)$ and hence

$Y_t = \mu + \varepsilon_t^* = \mu + \phi_1 \varepsilon_{t-1}^* + \varepsilon_t = \mu + \phi_1(Y_{t-1} - \mu) + \varepsilon_t$

$Y_t = \mu(1-\phi_1) + \phi_1 Y_{t-1} + \varepsilon_t.$

Hence for an AR(1) process, the observation at time *t* depends only on the observation at time *t*-1. A regression of Y_t on Y_{t-1} will produce a LS estimate of ϕ_1 and hence of μ. For this process, the autocorrelations turn out to be:

$r_1 = \phi_1,\ r_2 = \phi_1^2,\ r_3 = \phi_1^3, ...,$ the ACF declines exponentially as $r, r^2, r^3, r^4, \ldots$

All partial autocorrelations are zero beyond

If the beaver data followed an AR1 process, we would expect the pattern of autocorrelations to be 0.8, 0.8^2=0.64, 0.8^3=0.51, 0.8^4=0.41, 0.8^5=0.33, 0.8^2=0. 26, …. The actual pattern is 0.8, 0.66, 0.53, 0.46, 0.35, 0.25, ….

The partial autocorrelations all fluctuate between ± 0.15 (and lie within the confidence band centred around zero values). We should not be surprised to find that an AR1 process fits the beaver data well.

AR(2), or AR2: a lag 2 process: (conditions: $\phi_1^2 + \phi_2^2 < 1$ and $\phi_1 + \phi_2 < 1$)

For this process the dependent error depends only on the previous *two* dependent errors:

$\varepsilon_t^* = \phi_1 \varepsilon_{t-1}^* + \phi_2 \varepsilon_{t-2}^* + \varepsilon_t$

Again for a stationary model,

$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \varepsilon_t$

$Y_t = \mu(1-\phi_1-\phi_2) + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$

Hence for an AR(2) process, the observation at time *t* depends only on the previous *two* observations, those at time *t*-1 and at time *t*-2.

You would need to perform a bivariate regression of Y_t on Y_{t-1} and Y_{t-2} to obtain LS estimates of μ, ϕ_1 and ϕ_2. The regression coefficients in this process have more complex relationships with the autocorrelations. We obtain:

$$r_1 = \frac{\phi_1}{1-\phi_2} \qquad \phi_1 = \frac{r_1(1-r_2)}{1-r_1^2}$$

$$r_2 = \phi_2 + \frac{\phi_1^2}{1-\phi_2} \qquad \phi_2 = \frac{r_2 - r_1^2}{1-r_1^2}$$

By definition, the first partial autocorrelation is $r_{2.1} = \phi_2$. All other partial autocorrelations are zero.

MA(1), or MA1 in the LMM (REML) menu: a lag 1 process

Moving average processes are not so widely used in time or space dependent models so we will not describe them in detail. An AR(*p*) process has the dependent error depending on the previous *p dependent* errors, while an MA(*q*) process has the dependent error depending on the previous *q independent* errors. Text books use b for the slopes in these types of regression, but GenStat uses θ so we use this symbol below..

AR(*p*): $\varepsilon_t^* = \varepsilon_t + \alpha_1\varepsilon_{t-1}^* + \alpha_2\varepsilon_{t-2}^* + \ldots + \alpha_p\varepsilon_{t-p}^*$

MA(*q*): $\varepsilon_t^* = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}$

An ARMA(*p*, *q*) process is a combination of these two models.

The autocorrelations for an MA(*q*) process are **zero after *q* lags**.

The partial autocorrelations for an MA(*q*) process continue to have appreciable terms after *q* lags.

This is the opposite of an AR process. AR and MA processes can usually be distinguished from ACF and PACF plots.

In the remaining pages of this section we will apply some of these models to experimental data.

Simple linear regression via LMM (REML)

Let us firstly re-visit example 2 of the regression section of the manual.

Yields of potatoes receiving amounts of fertilizer (Snedecor and Cochran, page 150).

Amount	0	4	8	12	*mean fertiliser* = 6.000
Yield	8.34	8.89	9.16	9.50	*mean yield* = 8.973

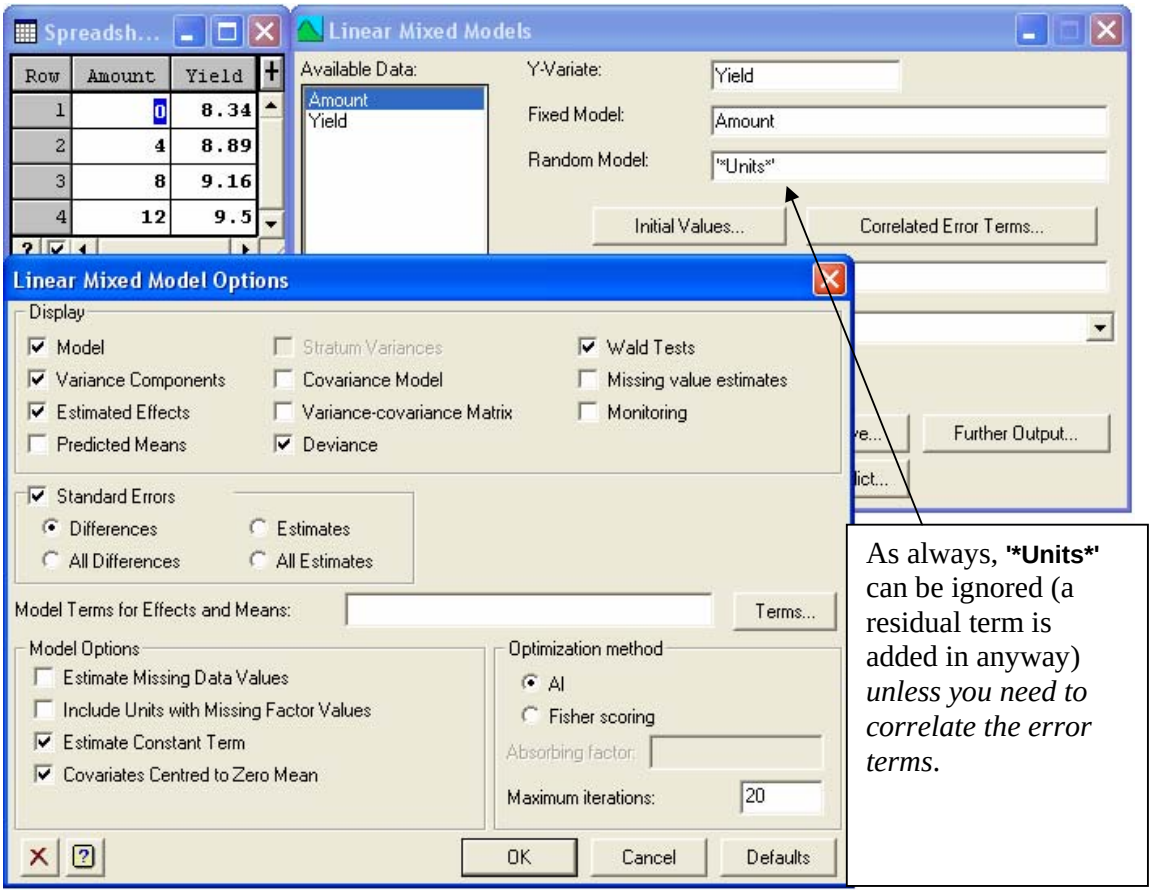
As mentioned earlier, the model can be expressed either as

Yield = intercept + slope × Fertiliser + Error

or as

Yield = mean yield + slope (*Fertiliser – mean fertiliser*) + *Error*

It is the second form of the model that GenStat has as the default in its LMM (REML) menu. To obtain the first form, go into **Options** and untick **Covariates Centred to Zero Mean**. You should also click **Deviance** and, for regression, the **Estimated Effects** (that is, mean Y and slope, or intercept and slope respectively).



Regression analysis

Response variate: Yield
Fitted terms: Constant, Amount

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	1	0.70312	0.703125	**82.00**	0.012
Residual	2	0.01715	0.008575		
Total	3	0.72028	0.240092		

Percentage variance accounted for 96.4
Standard error of observations is estimated to be 0.0926.

Estimates of parameters

Parameter	estimate	s.e.	t(2)	t pr.
Constant	8.4100	0.0775	108.55	<.001
Amount	0.0938	0.0104	9.06	0.012

REML variance components analysis

Response variate: Yield
Fixed model: Constant + Amount
Random model: '*units*'
Number of units: 4

'*units*' used as residual term

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
'	*units*'	Identity	Sigma2	0.00858	0.008575

Deviance: -2*Log-Likelihood

Deviance	d.f.
-1.75	1

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Amount	82.00	1	**82.00**	<0.001

and, for the default **Covariates Centred to Zero Mean:**

Table of effects for Constant

8.973 Standard error: 0.0463

Table of effects for Amount

0.09375 Standard error: 0.010353

otherwise:

Table of effects for Constant

8.410 Standard error: 0.0775

Table of effects for Amount

0.09375 Standard error: 0.010353

So LMM (REML):

- produces the same Wald/df (82.00) as regression produces for the ANOVA v.r.. The *P*-value is based on an asymptotic χ^2 distribution. Replace this with an $F_{1,\,2}$ *P*-value in this case.

- produces the same line of best fit
  Yield = 8.410 + 0.09375 *Fertiliser*

or equivalently

  Yield = 8.973 + 0.09375 (*Fertiliser* – 6.0)

The mean amount of fertilizer (6.0) is not part of the REML output, it needs to be calculated separately.

Simple linear regression with correlated errors

We saw that the temperatures for an individual beaver were correlated, so this dataset is ideal to illustrate a correlated-error model

Assume an AR1 stationary model for temperature. We can use change in deviance to test this model, namely

$Temperature_t = \mu + \varepsilon_t$ independent model

against the AR1-correlated model

$Temperature_t = \mu + \phi_1\varepsilon^*_{t-1} + \varepsilon_t$ AR1-correlated model

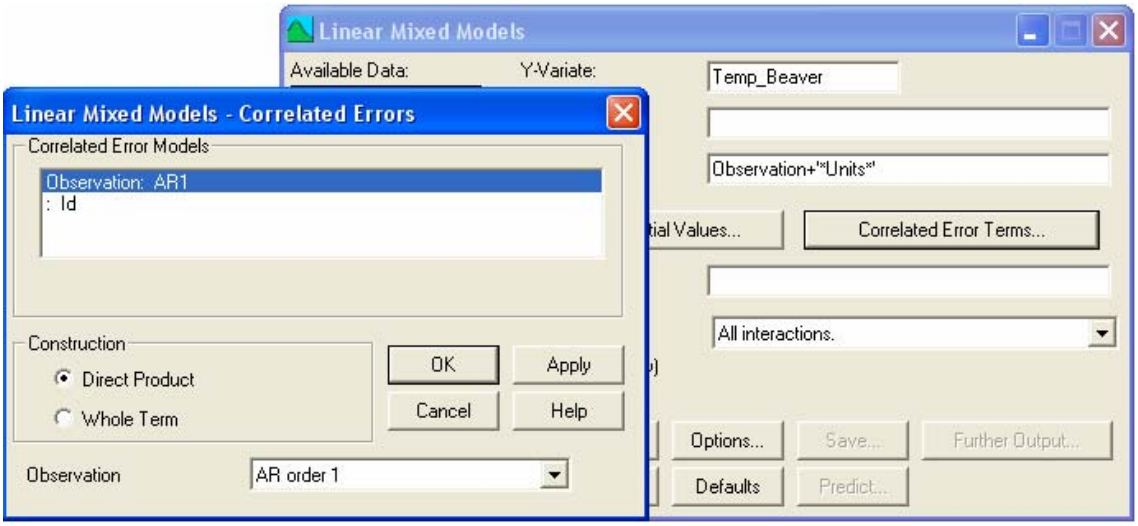
Note that the estimates will be slightly different than those obtained using GenStat's Time Series menu. LMM (REML) used REML rather than ML to estimate the variance parameters.

For the independent model, we leave the **Fixed Model** blank (there is no predictor variate). The **Random Model** consists of a factor to identify the *n* units, so we could set up our own Observation factor (with *n* = 115 levels), or just use the in-built '*Units*', or just leave it blank (since GenStat will add an independent error term for us).

For the independent model, we leave the **Fixed Model** blank (there is no predictor variate). The **Random Model** consists of a factor to identify the *dependent* units ε^*_{t-1} (we will use Observation for this), and a second factor to identify the *independent* units ε_t (we will use the in-built '*Units*' for this). We need to declare an AR1 structure for Observation in the random model. The deviances and test values are as follows. Clearly the AR1 model is superior.

Model	deviance	d.f.	change in deviance	change in d.f.	*P*-value
Identity	-253.56	112			
AR1	-411.23	110	157.67	2	<0.001

To maximize the explanation in GenStat's output we also use click **Covariance Model** in the LMM (REML) **Options**.



REML variance components analysis

Response variate: Temp_Beaver
Fixed model: Constant
Random model: Observation + '*units*'
Number of units: 114 (1 units excluded due to zero weights or missing values)

'*units*' used as residual term

Covariance structures defined for random model

Covariance structures defined within terms:

Term	Factor	Model	Order	No. rows
Observation	Observation	Auto-regressive (+ scalar)	1	115

Estimated parameters for covariance models

Random term(s)	Factor	Model(order)	Parameter	Estimate	s.e.
Observation	Observation	AR(1)	phi_1	0.9337	0.0472
			Scalar	113.4	218.2

Note: the covariance matrix for each term is calculated as G or R where
var(y) = Sigma2(ZGZ'+R), i.e. relative to the residual variance, Sigma2.

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
	'*units*'	Identity	Sigma2	0.000580	0.0010881

Estimated covariance models

Variance of data estimated in form:
V(y) = Sigma2(gZGZ' + I)

where: V(y) is variance matrix of data
Sigma2 is the residual variance
g is a gamma for the random term — G is the covariance matrix for the random term
Z is the incidence matrix for the random term — I is the residual (identity) covariance matrix

Note: a gamma is the ratio of a variance component to the residual (Sigma2)
Random Term: Observation

G is a single matrix
Scalar Sigma2*g: 0.06575

Factor: Observation
Model : Auto-regressive

Covariance matrix (first 10 rows only):

1	1.000									
2	0.934	1.000								
3	0.872	0.934	1.000							
4	0.814	0.872	0.934	1.000						
5	0.760	0.814	0.872	0.934	1.000					
6	0.710	0.760	0.814	0.872	0.934	1.000				
7	0.663	0.710	0.760	0.814	0.872	0.934	1.000			
8	0.619	0.663	0.710	0.760	0.814	0.872	0.934	1.000		
9	0.578	0.619	0.663	0.710	0.760	0.814	0.872	0.934	1.000	
10	0.539	0.578	0.619	0.663	0.710	0.760	0.814	0.872	0.934	1.000
	1	2	3	4	5	6	7	8	9	10

Residual term: '*units*'
Sigma2: 0.0005800

I is an identity matrix (114 rows)

Deviance: -2*Log-Likelihood

Deviance	d.f.
-411.23	110

Table of predicted means for Constant

36.87

Interpretation of the analysis

- The estimate of ϕ_1 is 0.9337. An AR1 model assumes that the correlations between the temperatures are $(0.9337)^2 = 0.872$ for periods two apart, $(0.9337)^3 = 0.814$ for periods three apart, $(0.9337)^4 = 0.760$ for periods four apart, $(0.9337)^5 = 0.710$ for periods five apart, and so on. These values form the covariance matrix printed above.

- The scalar 113.4 is multiplied by the variance estimate 0.000580 giving 0.066. This is confirmed in the output (Scalar Sigma2*g: 0.06575). This is the variance of the dependent error term in the model.

- The estimated model is

$$\begin{aligned} Temperature_t &= 36.87 + 0.9337\,\varepsilon^*_{t-1} + \varepsilon_t \\ &= 36.87(1\text{-}0.9337) + 0.9337\times Temperature_{t\text{-}1} + \varepsilon_{t\text{-}1} \\ &= 2.444 + 0.9337\times Temperature_{t\text{-}1} + \varepsilon_{t\text{-}1} \end{aligned}$$

  Thus, the temperature at time *t* is approximately 2.444°C + 0.9337 *times* the temperature at time *t*-1.

One-way design (in randomised blocks)

The next REML analysis is the RCBD data of example 3 of the design section.

Example 3 Sugar concentrations of nectar in half heads of red clover kept at different vapor pressures for eight hours (from Steel and Torrie, page 103)

Head	4.4 mm Hg	9.9 mm Hg	difference
1	62.5	51.7	10.8
2	65.2	54.2	11.0
3	67.6	53.3	14.3
4	69.9	57.0	12.9
5	69.4	56.4	13.0
6	70.1	61.5	8.6
7	67.8	57.2	10.6
8	67.0	56.2	10.8
9	68.5	58.2	10.3
10	62.4	55.8	6.6

Once Head is declared a block factor, GenStat really treats it as a random term. If you ask for **Stratum Variances** in ANOVA **Options**, GenStat will provide estimates of the variance of heads, and the error variance:

Analysis of variance

Variate: Concentration

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Head stratum	9	116.114	12.902	5.25	
Head.*Units* stratum					
Vapor_Pressure	1	592.960	592.960	241.32	<.001
Residual	9	22.115	2.457		
Total	19	731.189			

…

Estimated stratum variances

Stratum	variance	effective d.f.	**variance component**
Head	12.902	9.000	**5.222**
Head.*Units*	2.457	9.000	**2.457**

To do this analysis in GenStat, it makes sense to assume that heads are randomly chosen from a potentially large population of heads, and hence Head is a random component. The experimental unit is the concentration of one head at one concentration, hence the residual term is Head.Vapor_Pressure. So:

The **Fixed Model** is Vapor_Pressure

The **Random Model** is Head + Head.Vapor_Pressure or Head/Vapor_Pressure or just Head.

REML variance components analysis

Response variate: Concentration
Fixed model: Constant + Vapor_Pressure
Random model: Head + Head.Vapor_Pressure
Number of units: 20

Head.Vapor_Pressure used as residual term

Estimated variance components

Random term	component	s.e.
Head	5.222	3.096

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
	Head.Vapor_Pressure	Identity	Sigma2	2.457	1.158

Deviance: -2*Log-Likelihood

Deviance	d.f.
53.71	16

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Vapor_Pressure	241.32	1	241.32	<0.001

Table of predicted means for Constant

61.59 Standard error: 0.803

Table of predicted means for Vapor_Pressure

Vapor_Pressure	1	2
	67.04	56.15

Standard error of differences: 0.7010

Notice:

- The Wald/d.f. (241.32) is identical to the variance ratio of the ANOVA. Adjust the *P*-value (not necessary here) for one based on an *F* distribution since this is a balanced design with an uncorrelated, constant variance error structure.
- The REML estimate of variance is identical to the Residual MS of the ANOVA.
- The Head (block) variance (5.222) is identical to the Head stratum variance of the ANOVA.
- Means and s.e.d. values are identical.

One-way design (no blocking) – unequal variances

In the sugar treatment example (Example 4) of the Design section, we were introduced to a simple set of treatments in which the variance for one was different to the variance for the other treatments. We will reproduce the overall analysis using LMM (REML).

Firstly, the treatment variances (each with 9 *df*) fall into two groups. The variance for the untreated pots (15.878) appears wuite different to that for the treated pots. The average variance for treated pots is 2.850.

Treatment variances

Control	glucose 2%	fructose 2%	gluc_fruct 1%	sucrose 2%
15.878	2.678	3.511	2.000	3.211

As before, the **Fixed Model** is the Sugar factor with 5 levels.

The **Random Model** is Pot (a factor with levels 1 to 50). However, this model assumes that the variance is constant (Identity). We are interested in allowing the variance to change depending on the treatment.

The worst case is when every treatment has a different variance. What is believed is that only the Control treatment has a different variance.

Another way of extracting the tests of interest is

- to compare treated and untreated pots;
- for the treated pots, to compare among the four sugar treatments.

The spreadsheet can be set up with a contrast to measure the difference between control (Control_Rest = 0) and treated (Control_Rest = 1) pots. Among the treated pots, the four sugar treatments can be compared using GenStat's nested shortcut. In other words, the treatment structure is:

Fixed Model: Control_Rest/Sugar

The following choices set up difference variance structures among the treatments

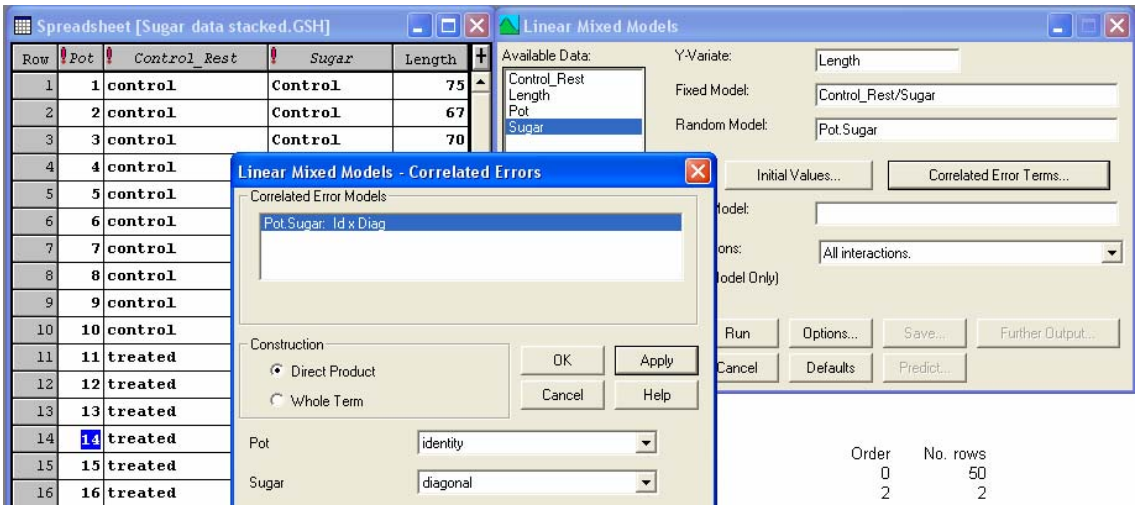
Random Model: Pot.Sugar	allows a different variance for all 5 sugar treatments by selecting **Diagonal** for Sugar in **Correlated Error Terms**
Random Model: Pot.Control_Rest	sets up one variance for the control treatment, and a separate variance for the other 4 sugar treatments, by selecting **Diagonal** for Control_Rest in **Correlated Error Terms**;
Random Model: Pot	sets up a constant variance for all 5 treatments by selecting **Identity** for Pot in **Correlated Error Terms**.

The models can be compared by change in deviance as usual.

Note that the default in GenStat is to produce multipliers rather than actual variances when selecting a **Diagonal** variance structure. To have GenStat print out the different variance estimates instead, use the

```
PARAMETERIZATION=sigmas
```

option of REML. You will need to run the default model, copy the three lines from the **Input** window, add the option and re-run the window.



The deviances for each of the three models are as follows.

Model	Random Model	Deviance	d.f.	Change in deviance	Change in d.f.	*P* value
All 5 treatment variances different	Pots.Sugar	118.3	40	0.80	3	0.849
Control variance different	Pot.Control_Rest	119.1	43	13.76	1	<0.001
Common variance	Pots	132.86	44			

Clearly allowing the control treatment to have a different variance is a better assumption than one with all variances equal ($P<0.001$); it appears unnecessary to allow all five treatments variances to be different ($P=0.8491$).

Having the **Fixed Model** as Control_Rest/Sugar allows the comparison of the control treatment with the remaining sugar treatments to be equivalent to a *t* test with unequal variances. The apparent interaction Control_Rest.Sugar is actually a main effect, testing the differences among the four sugar treatments – it is not unlike the eelworm analysis in the Design section of the manual.

The full analysis is as follows.

REML variance components analysis

Response variate: Length
Fixed model: Constant + Control_Rest + Control_Rest.Sugar
Random model: Pot.Control_Rest
Number of units: 40

Residual term has been added to model

Covariance structures defined for random model

Covariance structures defined within terms:

Term	Factor	Model	Order	No. rows
Pot.Control_Rest	Pot	Identity	0	40
	Control_Rest	Diagonal	2	2

EITHER: Sigma2 not scaled to 1:

Estimated parameters for covariance models

Random term(s)	Factor	Model(order)	Parameter	Estimate	s.e.
Pot.Control_Rest	Pot	Identity	-	-	-
	Control_Rest	Diagonal	d_1	4.571	2.626
			d_2	0.000	bound

Note: the covariance matrix for each term is calculated as G or R where
var(y) = Sigma2(ZGZ'+R), i.e. relative to the residual variance, Sigma2.

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
Residual		Identity	Sigma2	2.850	aliased

Here, Z is an identity matrix and G a diagonal matrix with a variance estimate of 4.817 in diagonal elements corresponding to control data, and 0.000 for treated data. R is an identity matrix. The estimate of the variance for the residual errors is 2.730. Thus, individual variances are estimated to be

var(yield) = 2.850 × (4.571+1.000) = 15.877 for control data, and
= 2.850 × (0.000+1.000) = 2.850 for treated data.

Notice that 15.877 is actually the sample variance of the control data, whereas 2.850 is the average of the four sugar variances, each with 9 *df*. Hence the variance estimate for he control data has 9 *df*, while the average sugar variance has 36 *df*.

OR Sigma2 scaled to 1 (via the PARAMETERIZATION option of REML):

Estimated parameters for covariance models

Random term(s)	Factor	Model(order)	Parameter	Estimate	s.e.
Pot.Control_Rest	Pot	Identity	-	-	-
	Control_Rest	Diagonal	d_1	14.88	7.48
			d_2	1.850	0.672

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
Residual		Identity	Sigma2	1.000	aliased

For this parameterization, individual variances are estimated to be

var(yield) = 1.000 × (14.88+1.000) = 15.88 for control data, and
= 1.000 × (1.850+1.000) = 2.85 for treated data.

In other words, the same estimates are obtained as with the earlier expressions. The deviance and test values are identical by both methods as well.

Deviance: -2*Log-Likelihood

Deviance	d.f.
119.10	43

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Control_Rest	62.71	1	62.71	<0.001
Control_Rest.Sugar	85.96	3	**28.65**	<0.001

Table of predicted means for Control_Rest.Sugar

Sugar:	Control	gluc_2%	fruc_2%	gluc_fruc_1%	gluc_fruc_1%
Control_Rest					
control	70.10	*	*	*	*
treated	*	59.30	58.20	58.00	64.10

Since the means have one of two estimated variances, the s.e.d. values will differ depending on whether a control mean is involved (1.37), or not (0.75). Use the **Standard Errors All Differences** option to obtain a complete set of s.e.d and l.s.d. values.

Notice the following.

- The Wald/d.f. statistic for the (nested) component Control_Rest.Sugar is the same value as the variance ratio of an ANOVA of just the treated data:

Analysis of variance

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Sugar	3	245.000	81.667	**28.65**	<.001
Residual	36	102.600	2.850		
Total	39	347.600			

- The Wald/d.f. statistic for the component Sugar is similar to the Satterthwaite approximate *t* test:

Control mean = 70.1 (based on 10 observations), var = 15.878, $df = 9$
Sugar mean = 59.9 (based on 40 observations), var = 2.850, $df = 36$

Difference in means = 10.2

$$\text{s.e.d.} = \sqrt{\frac{15.878}{10} + \frac{2.850}{40}} = 1.288$$

$t = 10.2/1.288 = 7.919$, or $F = t^2 = 7.919^2 = 62.711$.

Wald/d.f. statistic uses an asymptotic χ^2 distribution with (here) 2 *df* rather than an approximate *t* distribution with degrees of freedom closer to 9 than to 36.

Two-way design (no blocking) with subsamples

We have another look at the mint length data on page 55 of the manual, in which 18 pots were used in a 3 × 2 factorial treatment design. The **Treatment Structure** is Temperature*Light and the **Block Structure** is Pots.Plants. The estimates of the two stratum variances were

Estimated stratum variances

Stratum	variance	effective d.f.	variance component
Pot	2.153	12.000	**0.305**
Pot.Plant	0.934	54.000	**0.934**

Here is the LMM (REML) analysis. The means are as before and are suppressed in this output.

REML variance components analysis

Response variate: Length
Fixed model: Constant + Light + Temperature + Light.Temperature
Random model: Pot + Pot.Plant
Number of units: 72

Pot.Plant used as residual term

Estimated variance components

Random term	component	s.e.
Pot	0.3047	0.2243

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
	Pot.Plant	Identity	Sigma2	0.934	0.1798

Approximate stratum variances

Use Fisher scoring to obtain this

Stratum	variance	effective d.f.
Pot	**2.1528**	**12.00**
Pot.Plant	**0.9340**	**54.00**

Wald tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Light	10.36	2	5.18	0.006
Temperature	70.45	1	70.45	<0.001
Light.Temperature	2.64	2	1.32	0.268

Notice:

- The variance estimates (and *df*) are the same as obtained from ANOVA;
- The Wald/d.f. statistics are the same as the variance ratios from the ANOVA. *P* values differ slightly, but since the design is balanced, correcting to *P* values from *F* distributions is valid.

However, this analysis did not depend on balance.

Two-way design (in randomized blocks)

We have another look at the cowpea hay data on page 59 of the manual. The **Treatment Structure** is Spacing*Variety and the **Block Structure** is Block/Plot. In the earlier discussion, there was consideration about whether the variance was constant, proportional to the number of plants in a plot, or somewhere in between. We explore these issues using change in deviance.

The estimates of the stratum variances were:

Estimated stratum variances

Stratum	variance	effective d.f.	variance component
Block	85.213	3.000	7.505
Block.*Units*	17.671	24.000	17.671

In order to allow a changing variance model for different spacings, we need to ensure that Spacing appears in the **Block Structure** so we can use **Correlated Error Terms**. We can change Block/Plot for an expression in which the Plot part is replaced by a factor expression which ranges over the same set of values. Plot goes from 1 to 9 in each block. These track which combination of variety and spacing is used in each plot. Hence an equivalent expression for the **Block Structure** is Block.Spacing.Variety. The deviances for common variance (Identity) and variances changing over Spacing levels (Diagonal) are as follows:

	deviance	d.f.	Change in deviance	Change in d.f.	P value
Identity	121.74	25			
Diagonal	120.37	23	1.37	2	0.504

For this experiment, there is no evidence that a changing variance model is necessary (P=0.504). The rest of the analysis gives the same variance estimates and equivalent test values as for ANOVA.

REML variance components analysis

Response variate: Yield
Fixed model: Constant + Variety + Spacing + Variety.Spacing
Random model: Block + Block.Variety.Spacing
Number of units: 36

Block.Variety.Spacing used as residual term

Estimated variance components

Random term	component	s.e.
Block	**7.50**	7.75

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
	Block.Variety.Spacing	Identity	Sigma2	17.67	5.10

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Variety	58.14	2	29.07	<0.001
Spacing	8.77	2	4.39	0.012
Variety.Spacing	43.32	4	10.83	<0.001

Two-way design (in randomized blocks) – changing variance

McConway *et al.* (1999) reported the results of an experiment which had a randomised block design, in more or less the following words. There were 64 plots, arranged in four blocks each of size sixteen. Each block was a rectangular piece of land, measuring 3m × 32m. Each block was divided into sixteen plots by splitting the long side of the block into sixteen 2m pieces. So, each plot was a 3m× 2m rectangle of land. The River Thames runs along one edge of the field used in this experiment, and usually floods part of the field each year. The blocks were designed so that the long side of each block was parallel to the river-bank. The blocks were different distances from the river-bank.

The experiment was about growing turnips for fodder. The turnips would not normally be harvested because they are grown to provide food for farm animals in winter; the farmer simply releases animals into the field and the animals graze on the turnips. The turnips are not even the main crop in the field during the growing season; the turnips are sown after the main crop is removed.

There were sixteen treatments in this experiment. The combinations are formed from: two different varieties – Barkant or Marco; two different sowing dates – one as soon as possible after the main crop has been harvested, the other a week later; and four different sowing densities – 1, 2, 4 or 8 kg ha^{-1}. Treatment combinations were allocated to plots within blocks at random.

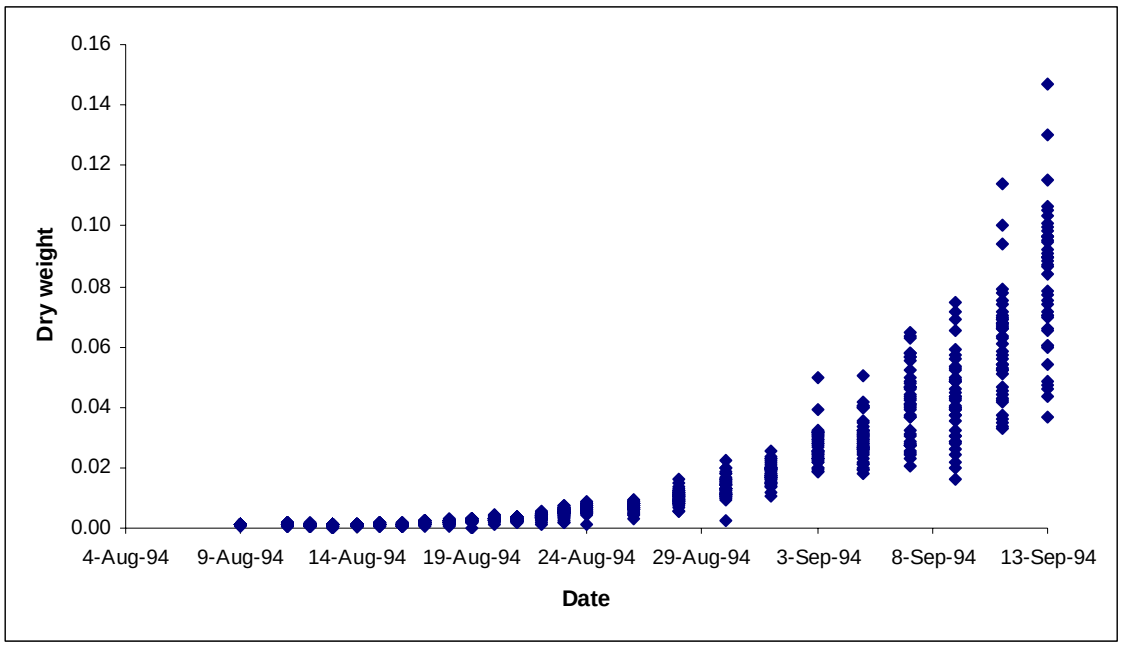
Yield of turnips (kg)

variety	sowing date	sowing density (kg ha^{-1})	Block 1	Block 2	Block 3	Block 4
Barkant	21/08/1990	1	2.7	1.4	1.2	3.8
		2	7.3	3.8	3.0	1.2
		4	6.5	4.6	4.7	0.8
		8	8.2	4.0	6.0	2.5
	28/08/1990	1	4.4	0.4	6.5	3.1
		2	2.6	7.1	7.0	3.2
		4	24.0	14.9	14.6	2.6
		8	12.2	18.9	15.6	9.9
Marco	21/08/1990	1	1.2	1.3	1.5	1.0
		2	2.2	2.0	2.1	2.5
		4	2.2	6.2	5.7	0.6
		8	4.0	2.8	10.8	3.1
	28/08/1990	1	2.5	1.6	1.3	0.3
		2	5.5	1.2	2.0	0.9
		4	4.7	13.2	9.0	2.9
		8	14.9	13.3	9.3	3.6

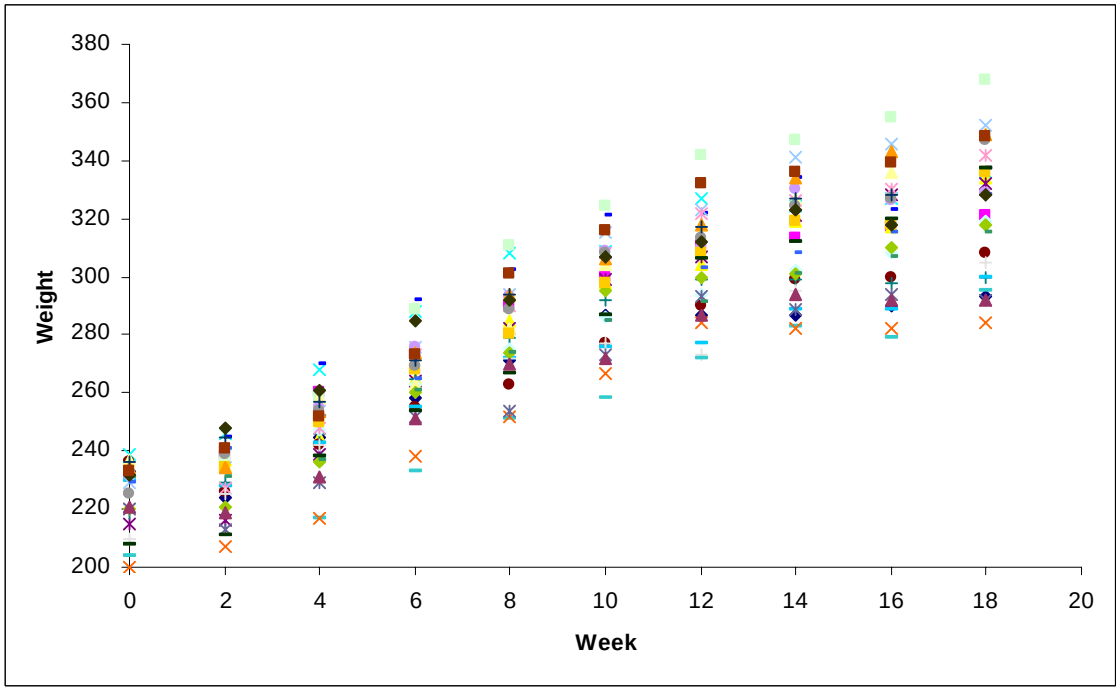
Again, this is a density trial, and hence the variance *may* change over different planting densities.

The plants are also grown for two different time periods. It is almost always the case that the variance of plant yield increases over time.

Consider the following information from an experiment conducted by a former student at The University of Sydney (Jason Moodie) on lettuce growth for the first 30 days after transplanting seedlings. Dry weights, fresh weights and leaf areas were measured every day or every second day. It is clear that the variance increases over time.



A second example is calf weight for the first nineteen weeks after birth:



Again, the variance appears to increase as the calves grow. The means and variances over time for these thirty calves are as follows.

Week	0	2	4	6	8	10	12	14	16	18	19
Mean	226.20	230.33	246.87	265.63	281.17	294.87	304.73	312.87	315.13	324.07	325.47
Variance	105.54	155.13	165.22	184.86	242.97	283.77	306.55	340.67	389.15	470.06	444.60

The point is that we should expect the variance to change when plants are grown for two different periods; we may also expect the variance to change with density, depending on the extent of plant competition.

For this experiment, the **Fixed Model** is Variety*Date*Density and the **Random Model** is Block/Plot. As before, plots are completely described by the combination of Variety*Date*Density, leading to Block+ Block.Variety*Date*Density as the **Random Model**. That allows use to investigate Diagonal structures for Date and/or Density.

Block	Variety	Date	Density	deviance	d.f.	Change in deviance	Change in d.f.	P value
Identity	Identity	Diagonal	Diagonal	162.05	42			
Identity	Identity	Identity	Diagonal	168.10	43	6.05	1	0.014
Identity	Identity	Diagonal	Identity	175.71	45	13.66	3	0.003
Identity	Identity	Identity	Identity	183.92	46			

If we start assuming that the variance changes over time as well as over densities, we can then test whether an adequate model has only a changing variance over densities ($P = 0.014$), or a changing variance over time ($P = 0.003$). We clearly should allow the variance to change over both factors.

REML variance components analysis

Response variate: weight
Fixed model: Constant + density + sowing + variety + density.sowing + density.variety + sowing.variety + density.sowing.variety
Random model: block + block.density.sowing.variety
Number of units: 64

block.density.sowing.variety used as residual term with covariance structure as below

Covariance structures defined for random model

Covariance structures defined within terms:

Term	Factor	Model	Order	No. rows
block.density.sowing.variety				
	block	Identity	0	4
	density	**Diagonal**	**4**	**4**
	sowing	**Diagonal**	**2**	**2**
	variety	Identity	0	2

Estimated variance components

Random term	component	s.e.
block	0.160	0.328

Output using `PARAMETERIZATION=sigmas`

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
block.density.sowing.variety		Sigma2	1.000	fixed	
	block	Identity	-	-	-
	density	Diagonal	d_1	1.000	fixed
			d_2	2.195	1.358
			d_3	10.48	6.35
			d_4	7.682	4.661
	sowing	Diagonal	d_1	1.030	0.507
			d_2	3.143	1.481
	variety	Identity	-	-	-

Estimated covariance models

Variance of data estimated in form:

V(y) = sZZ' + Sigma2.R

where: V(y) is variance matrix of data
s is the variance component for the random term
Z is the incidence matrix for the random term
Sigma2 is the residual variance
R is the residual covariance matrix

Random Term: block
Scalar s: 0.1604

Residual term: block.density.sowing.variety
Sigma2: 1.000

R uses direct product construction

To assist in understanding this output, we turned on the option **Covariance Model**. GenStat has scaled σ^2 to 1. The information on variance estimates is then obtained in the diagonal covariance matrices of the factors making up the residual term. To take one block and one variety, the variance of Y is obtained by evaluating the direct product of the two diagonal covariance matrices:

$$\begin{pmatrix} 1.000 & 0 & 0 & 0 \\ 0 & 2.195 & 0 & 0 \\ 0 & 0 & 10.48 & 0 \\ 0 & 0 & 0 & 7.682 \end{pmatrix} \otimes \begin{pmatrix} 1.030 & 0 \\ 0 & 3.143 \end{pmatrix} =$$

This matrix is

1.000×1.030	0	0	0	0	0	0	0
0	1.000×3.143	0	0	0	0	0	0
0	0	2.195×1.030	0	0	0	0	0
0	0	0	2.195×3.143	0	0	0	0
0	0	0	0	10.48×1.030	0	0	0
0	0	0	0	0	10.48×3.143	0	0
0	0	0	0	0	0	7.682×1.030	0
0	0	0	0	0	0	0	7.682×3.143

	Density	1		2		4		8	
Density	Date	21	28	21	28	21	28	21	28
1	21	1.030	0	0	0	0	0	0	0
	28	0	3.143	0	0	0	0	0	0
2	21	0	0	2.261	0	0	0	0	0
	28	0	0	0	6.899	0	0	0	0
4	21	0	0	0	0	10.794	0	0	0
	28	0	0	0	0	0	32.939	0	0
8	21	0	0	0	0	0	0	7.912	0
	28	0	0	0	0	0	0	0	24.145

Thus, the variance of an observation in any block and variety, density 1 kg ha^{-1} and sown on 21/08/1990 is 0.1604 (= block variance) + 1.030 = 1.190. For a similar combination but sown a week later, it is 0.1604 + 3.143 = 3.301.

The same variances are obtained using `PARAMETERIZATION=gammas`. GenStat estimates σ^2 to be 1.030 and scales the leading diagonal element of the covariance matrix for sowing date:

	sowing	Diagonal	d_1	1.000	fixed
			d_2	3.053	1.328

Deviance: -2*Log-Likelihood

Deviance	d.f.
162.05	42

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
density	38.47	3	12.82	<0.001
sowing	7.86	1	7.86	0.005
variety	9.35	1	9.35	0.002
density.sowing	**14.50**	**3**	**4.83**	**0.002**
density.variety	0.40	3	0.13	0.941
sowing.variety	2.03	1	2.03	0.154
density.sowing.variety	1.44	3	0.48	0.697

Next we present just the two-way means for density and sowing for illustration. Since there are changing variances over the levels of some factors, we should turn on the option **Standard Errors All Differences** so that individual differences can be compared or estimated with the correct precision.

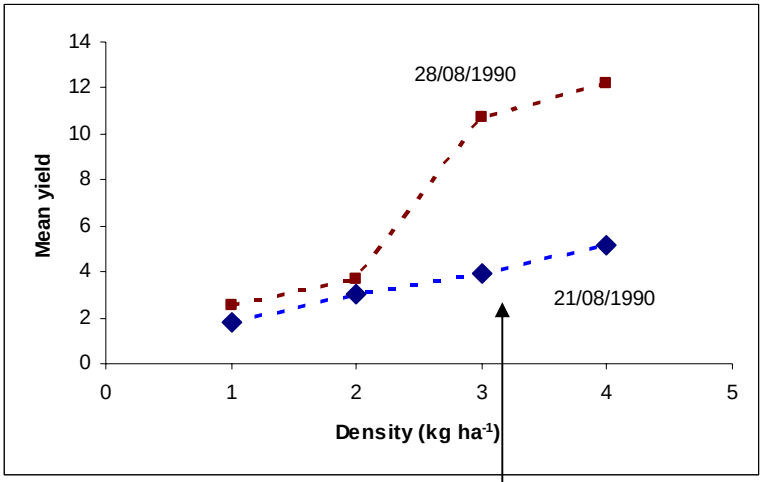


Table of predicted means for density.sowing

sowing	21/8/90	28/8/90
density		
1kg/Ha	1.763	2.512
2kg/Ha	3.013	3.688
4kg/Ha	3.912	10.737
8kg/Ha	5.175	12.212

To compare these means, use

Standard errors of differences between pairs

density 1.sowing 21/8/90	**1**	*							
density 1.sowing 28/8/90	**2**	0.722	*						
density 2.sowing 21/8/90	**3**	0.641	0.822	*					
density 2.sowing 28/8/90	**4**	0.996	1.120	1.070	*				
density 4.sowing 21/8/90	**5**	1.216	1.320	1.277	1.487	*			
density 4.sowing 28/8/90	**6**	2.061	2.124	2.098	2.232	2.338	*		
density 8.sowing 21/8/90	**7**	1.057	1.175	1.127	1.361	1.529	2.260	*	
density 8.sowing 28/8/90	**8**	1.774	1.847	1.817	1.970	2.090	2.671	2.002	*
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**

Latin Square design

The wheat yields on page 68 are from a field-based Latin Square design. For that design there were three variance estimates coming from the three strata – rows, columns and plots. As before, the **Fixed Model** contains the one factor, Variety, while the **Random Model** is Row_Block*Column_Block.

REML variance components analysis

Response variate: Yield
Fixed model: Constant + Variety
Random model: Row_Block + Column_Block + Row_Block.Column_Block
Number of units: 16

Row_Block.Column_Block used as residual term

Estimated variance components

Random term	component	s.e.
Row_Block	0.0496	0.1482
Column_Block	0.4533	0.4673

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
Row_Block.Column_Block		Identity	Sigma2	0.453	0.2617

Deviance: -2*Log-Likelihood

Deviance	d.f.
13.97	9

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Variety	174.10	3	58.03	<0.001

Table of predicted means for Constant

10.45 Standard error: 0.393

Table of predicted means for Variety

Variety	A	B	C	D
	12.00	12.27	10.80	6.73

Standard error of differences: 0.4761

Notice, as usual:

- The estimates of variance are the same as the stratum variances given in the ANOVA.
- The Wald/d.f. statistic is the same as the variance ratio of the ANOVA, however the *P* value is based on an asymptotic χ^2 distribution rather than an *F* distribution.
- The means and s.e.d. values are the same as from ANOVA.

Split-plot design (in randomized blocks)

The discussion on page 76 of the residual plot suggested that for the oats data of Example 8, the different cultivars may well have a different variance because of the treatments. We will use LMM (REML) to explore this possibility.

For this split-plot there are three strata: blocks, whole-plots and split-plots. Hence, the **Random Model** is Block/W_Plot/S_Plot. In order to allow a changing variance across cultivars, we need to mention them in the **Random Model**. Cultivars were allocated at random to the whole plots, so we can express the **Random Model** as Block/Cultivar/S_Plot, Block/Cultivar/Chemical, or simply as Block/Cultivar since the final stratum can be omitted. The stratum variances were estimated in ANOVA as follows:

Estimated stratum variances

Stratum	variance	effective d.f.	variance component
Block	947.624	3.000	54.933
Block.Cultivar	68.699	9.000	12.097
Block.Cultivar.Chemical	20.311	36.000	20.311

Standard split-plot analysis via LMM (REML)

REML variance components analysis

Response variate: Yield
Fixed model: Constant + Cultivar + Chemical + Cultivar.Chemical
Random model: Block + Block.Cultivar + Block.Cultivar.Chemical
Number of units: 64

Block.Cultivar.Chemical used as residual term

Estimated variance components

Random term	component	s.e.
Block	**54.93**	48.40
Block.Cultivar	**12.10**	8.18

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
	Block.Cultivar.Chemical	Identity	Sigma2	**20.31**	4.79

Deviance: -2*Log-Likelihood

Deviance	d.f.
237.21	45

Wald/d.f. statistics same as ANOVA *F* statistics

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Cultivar	41.46	3	13.82	<0.001
Chemical	8.40	3	2.80	0.038
Cultivar.Chemical	28.87	9	3.21	<0.001

Table of predicted means for Constant

52.81 Standard error: 3.848

Table of predicted means for Cultivar

Cultivar	Branch	Clinton	Vicland (1)	Vicland (2)
	61.07	54.31	42.46	53.41

Standard error of differences: 2.930

Table of predicted means for Chemical

Chemical	Control	Ceresan	Panogen	Agrox
	50.69	55.20	53.12	52.22

Standard error of differences: 1.593

Table of predicted means for Cultivar.Chemical

Chemical Cultivar	Control	Ceresan	Panogen	Agrox
Branch	61.92	63.42	57.67	61.25
Clinton	53.93	51.38	55.88	56.05
Vicland (1)	36.05	50.63	45.85	37.30
Vicland (2)	50.85	55.38	53.10	54.30

Standard errors of differences

Average:	3.858
Maximum:	4.025
Minimum:	3.187

Average variance of differences: 14.99

Standard error of differences for same level of factor:

	Cultivar	Chemical
Average:	3.187	4.025
Maximum:	3.187	4.025
Minimum:	3.187	4.025

LMM (REML) gives the same means and s.e.d. values as from ANOVA. Unfortunately, there is no guidance as to *df* to use. Compare the corresponding information from ANOVA:

Standard errors of differences of means

Table	Cultivar	Chemical	Cultivar Chemical
rep.	16	16	4
s.e.d.	2.930	1.593	4.025
d.f.	9	36	26.78
Except when comparing means with the same level(s) of			
Cultivar			3.187
d.f.			36

Next, we demonstrate how to check for changing variance across cultivars. Given the nature of the cultivars and seed chemical protectants, we might expect this variance to change only at the split-plot level. The following change in deviance table explores various models for

Cultivar in firstly the split-plot error term (Block.Cultivar.Chemical) and then in the whole-plot error term (Block.Cultivar).

Model for Cultivar in Block.Cultivar	Model for Cultivar in Block.Cultivar.Chemical	deviance	d.f.	change in deviance	change in d.f.	*P* value
Identity	Identity	237.21	45			
Identity	Diagonal	225.78	42	**11.43**	**3**	**0.010**
Diagonal	Diagonal	223.69	39	2.09	3	0.554

The analysis allowing for a changing variance at the split-plot level is as follows. Use **Save** if you want to take the s.e.d. values into Excel or Word most efficiently.

REML variance components analysis

Response variate: Yield
Fixed model: Constant + Cultivar + Chemical + Cultivar.Chemical
Random model: Block + Block.Cultivar + Block.Cultivar.Chemical
Number of units: 64

Block.Cultivar.Chemical used as residual term with covariance structure as below

Sparse algorithm with AI optimisation

Covariance structures defined for random model

Covariance structures defined within terms:

Term	Factor	Model	Order	No. rows
Block.Cultivar.Chemical	Block	Identity	0	4
	Cultivar	Diagonal	4	4
	Chemical	Identity	0	4

Estimated variance components

Random term	component	s.e.
Block	55.842	48.137
Block.Cultivar	7.728	6.384

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
	Block.Cultivar.Chemical	Sigma2	1.000	fixed	
	Block	Identity	-	-	-
	Cultivar	Diagonal	d_1	12.81	5.98
			d_2	33.19	16.03
			d_3	4.060	1.898
			d_4	37.03	17.41
	Chemical	Identity	-	-	-

Estimated covariance models

Variance of data estimated in form:

V(y) = SUM{ s(i)Z(i)Z(i)' } + Sigma2.R

where: V(y) is variance matrix of data
s(i) is a variance component for random term i
Z(i) is the incidence matrix for random term i
Sigma2 is the residual variance
R is the residual covariance matrix

Random Term 1: Block
Scalar s(1): 55.84

Random Term 2: Block.Cultivar
Scalar s(2): 7.728

Residual term: Block.Cultivar.Chemical
Sigma2: 1.000

R uses direct product construction

Factor: Block
Model : Identity (4 rows)

Factor: Cultivar
Model : Diagonal

Covariance matrix:
1 12.81
2 33.19
3 4.06
4 37.03

Factor: Chemical
Model : Identity (4 rows)

Deviance: -2*Log-Likelihood

Deviance	d.f.
225.78	42

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Cultivar	73.54	3	24.51	<0.001
Chemical	93.98	3	31.33	<0.001
Cultivar.Chemical	58.23	9	6.47	<0.001

Table of predicted means for Constant

52.81 Standard error: 3.845

Table of predicted means for Cultivar

Cultivar	Branch	Clinton	Vicland (1)	Vicland (2)
	61.07	54.31	42.46	53.41

Standard errors of differences between pairs

Cultivar Branch	1	*			
Cultivar Clinton	2	2.60	*		
Cultivar Vicland (1)	3	2.22	2.49	*	
Cultivar Vicland (2)	4	2.64	2.87	2.54	*
		1	2	3	4

Standard errors of differences

Average:	2.559
Maximum:	2.873
Minimum:	2.218

Average variance of differences: 6.586

Table of predicted means for Chemical

Chemical	Control	Ceresan	Panogen	Agrox
	50.69	55.20	53.12	52.22

Standard errors of differences between pairs

Chemical Control	1	*			
Chemical Ceresan	2	1.65	*		
Chemical Panogen	3	1.65	1.65	*	
Chemical Agrox	4	1.65	1.65	1.65	*
		1	2	3	4

Standard error of differences: 1.650

Table of predicted means for Cultivar.Chemical

Chemical Cultivar	Control	Ceresan	Panogen	Agrox
Branch	61.92	63.42	57.67	61.25
Clinton	53.93	51.38	55.88	56.05
Vicland (1)	36.05	50.63	45.85	37.30
Vicland (2)	50.85	55.38	53.10	54.30

Standard errors of differences between pairs

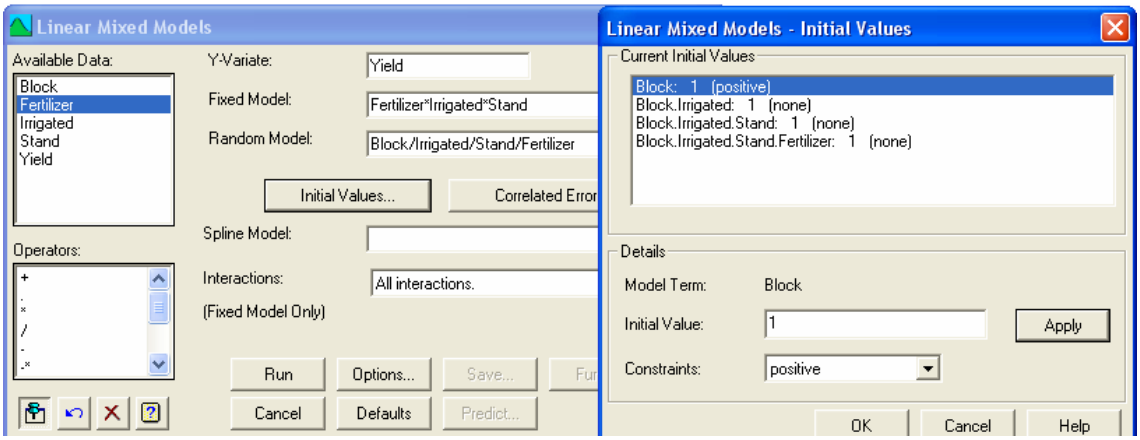
Cultivar Branch.Chemical Control	1	*			
Cultivar Branch.Chemical Ceresan	2	2.53	*		
Cultivar Branch.Chemical Panogen	3	2.53	2.53	*	
Cultivar Branch.Chemical Agrox	4	2.53	2.53	2.53	*
Cultivar Clinton.Chemical Control	5	3.92	3.92	3.92	3.92
Cultivar Clinton.Chemical Ceresan	6	3.92	3.92	3.92	3.92
Cultivar Clinton.Chemical Panogen	7	3.92	3.92	3.92	3.92
Cultivar Clinton.Chemical Agrox	8	3.92	3.92	3.92	3.92
Cultivar Vicland (1).Chemical Control	9	2.84	2.84	2.84	2.84
Cultivar Vicland (1).Chemical Ceresan	10	2.84	2.84	2.84	2.84
Cultivar Vicland (1).Chemical Panogen	11	2.84	2.84	2.84	2.84
Cultivar Vicland (1).Chemical Agrox	12	2.84	2.84	2.84	2.84
Cultivar Vicland (2).Chemical Control	13	4.04	4.04	4.04	4.04
Cultivar Vicland (2).Chemical Ceresan	14	4.04	4.04	4.04	4.04
Cultivar Vicland (2).Chemical Panogen	15	4.04	4.04	4.04	4.04
Cultivar Vicland (2).Chemical Agrox	16	4.04	4.04	4.04	4.04
		1	2	3	4

etc.

Split-split-plot design (in randomized blocks)

The irrigated corn experiment on page 83 illustrates the occasional need to restrict the variance estimates to be positive. In the split-plot ANOVA, the variance of the block stratum was estimated as -22.54 simply because the Block MS was smaller than the Residual MS in the whole-plot analysis.

For this design there are four strata, the **Fixed Model** being the same as the **Treatment Structure** of ANOVA (Fertilizer*Irrigated*Stand) and the **Random Model** being the same as the **Block Structure** (Block/Irrigated/Stand/Fertilizer). To ensure that all stratum variances are positive, you need to click **Initial Values**, choose Block and select **positive** for **Constraints**.



Because the Block variance is estimate on the boundary, the REML estimate of the Block.Irrigated variance is affected.

REML variance components analysis

Response variate: Yield
Fixed model: Constant + Irrigated + Stand + Fertilizer + Irrigated.Stand + Irrigated.Fertilizer + Stand.Fertilizer + Irrigated.Stand.Fertilizer
Random model: Block + Block.Irrigated + Block.Irrigated.Stand + Block.Irrigated.Stand.Fertilizer
Number of units: 72

Block.Irrigated.Stand.Fertilizer used as residual term

Estimated variance components

Random term	component	s.e.
Block	**0.00**	**bound**
Block.Irrigated	3.93	20.15
Block.Irrigated.Stand	48.66	32.34

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
	Block.Irrigated.Stand.Fertilizer	Identity	Sigma2	86.36	20.35

Deviance: -2*Log-Likelihood

Deviance	d.f.
338.38	50

Wald tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Irrigated	30.92	1	30.92	<0.001
Stand	7.57	2	3.78	0.023
Fertilizer	22.90	2	11.45	<0.001
Irrigated.Stand	11.82	2	5.91	0.003
Irrigated.Fertilizer	11.04	2	5.52	0.004
Stand.Fertilizer	3.53	4	0.88	0.473
Irrigated.Stand.Fertilizer	2.72	4	0.68	0.606

Notice also that the main effect in the whole plot analysis is also affected by the zero estimate of the Block variance. Essentially, when a REML estimate of 0 is obtained for Block, the Wald/d.f. statistic becomes the same as the *F* statistic from a whole-plot analysis *with no block terms*, that is, from a CRD analysis of the whole-plots:

the whole-plot analysis

Variate: Yield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block stratum	3	194.44	64.81	0.14	
Block.Irrigated stratum					
Irrigated	1	8277.56	8277.56	17.59	0.025
Residual	3	1411.78	470.59	2.03	
Block.Irrigated.Stand stratum					

…

changes to…

Variate: Yield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block.Irrigated stratum					
Irrigated	1	8277.56	8277.56	**30.92**	0.001
Residual	6	1606.22	267.70	1.15	
Block.Irrigated.Stand stratum					

…

Two-way design (in randomized blocks) – spatial correlation model

Firstly, we re-produce the analysis of the eelworm Log(Final_Count) data of Example 10 on page 94. Recall that the **Treatment Structure** is Treated_Control/(Dose*Fumigant). The **Block Structure** is Block+Block.Plot, however on page 95 we simply used Block, dropping the final stratum. In addition, we used Log(Initial_Count) as a covariate. The stratum variances from the ANOVA are:

Estimated stratum variances (adjusted for covariate)

Stratum	variance	effective d.f.	variance component
Block	0.3029	2.746	0.0173
Block.*Units*	0.0953	35.254	0.0953

To replicate this analysis using LMM (REML), we move the covariate into the **Fixed Model**, which becomes Log_Initial_Count+Treated_Control/(Dose*Fumigant).

The **Random Model** is Block+Block.Plot, or simply Block. Neither formulation allows us to use a correlation structure spatially. We will discuss this issue after the basic REML analysis is completed:

REML variance components analysis

Response variate: log_Final_Count
Fixed model: Constant + log_Initial_Count + Treated_Control + Treated_Control.Fumigant + Treated_Control.Dose + Treated_Control.Fumigant.Dose
Random model: Block
Number of units: 48

All covariates centred

Estimated variance components

Random term	component	s.e.
Block	0.01730	0.02169

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
Residual		Identity	Sigma2	0.0953	0.02271

Deviance: -2*Log-Likelihood

Deviance	d.f.
-30.98	36

Wald tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
log_Initial_Count	61.54	1	61.54	<0.001
Treated_Control	12.25	1	12.25	<0.001
Treated_Control.Fumigant	21.93	3	7.31	<0.001
Treated_Control.Dose	0.46	1	0.46	0.497
Treated_Control.Fumigant.Dose	3.33	3	1.11	0.344

REML estimates of the block and error variances are the same as the stratum variances. The Wald/d.f. statistics are similar, but not identical to, the variance ratios from ANOVA. Once a covariate is added, the main effects depend on the order the factors are entered into the model (just as they would in the ANOVA).

As always as an option, the standard errors for **All Differences** can be printed, or saved in a spreadsheet and later saved as an Excel file.

How do we incorporate a spatial correlation for this experiment? Points for discussion:

- For a simple randomised block design, the same Wald statistic is obtained for treatments irrespective of whether Block is regarded as fixed or random.

- When Block is a random term in the model, then every plot in the block is actually uniformly correlated, that is, the correlation between any two plots in any one block is the same. This is the **Uniform** correlation model for the error terms. Plots in different blocks are uncorrelated.

  We can illustrate this with two plots, *j* and *k*, in block *i*. The simple RCBD model with fixed treatments implies

  $Y_{ij} = mean + Block_i + Treatment_j + Error_{ij}$
  and
  $Y_{ik} = mean + Block_i + Treatment_k + Error_{ik}$

  Since $Block_i \sim N\left(0,\sigma^2_{Block}\right)$ independently of $Error_{ij} \sim N\left(0,\sigma^2\right)$ we obtain

  $$var\left(Y_{ij}\right) = var\left(Y_{ik}\right) = \sigma^2_{Block} + \sigma^2$$
  and
  $$covar\left(Y_{ij},Y_{ik}\right) = \sigma^2_{Block}$$

  giving the following correlation between the two plots:

  $$corr\left(Y_{ij},Y_{ik}\right) = \frac{\sigma^2_{Block}}{\sigma^2_{Block} + \sigma^2} = \theta \text{ say.}$$

Let us firstly illustrate these observations using the cowpea hay yields of Example 6.

ANOVA variance ratios:

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block stratum	3	255.64	85.21	4.82	
Block.*Units* stratum					
Variety	2	1027.39	513.69	**29.07**	<.001
Spacing	2	155.06	77.53	**4.39**	0.024
Variety.Spacing	4	765.44	191.36	**10.83**	<.001
Residual	24	424.11	17.67		

The Wald/d.f. statistics from a REML analysis with Block a *random* component are, from page 133:

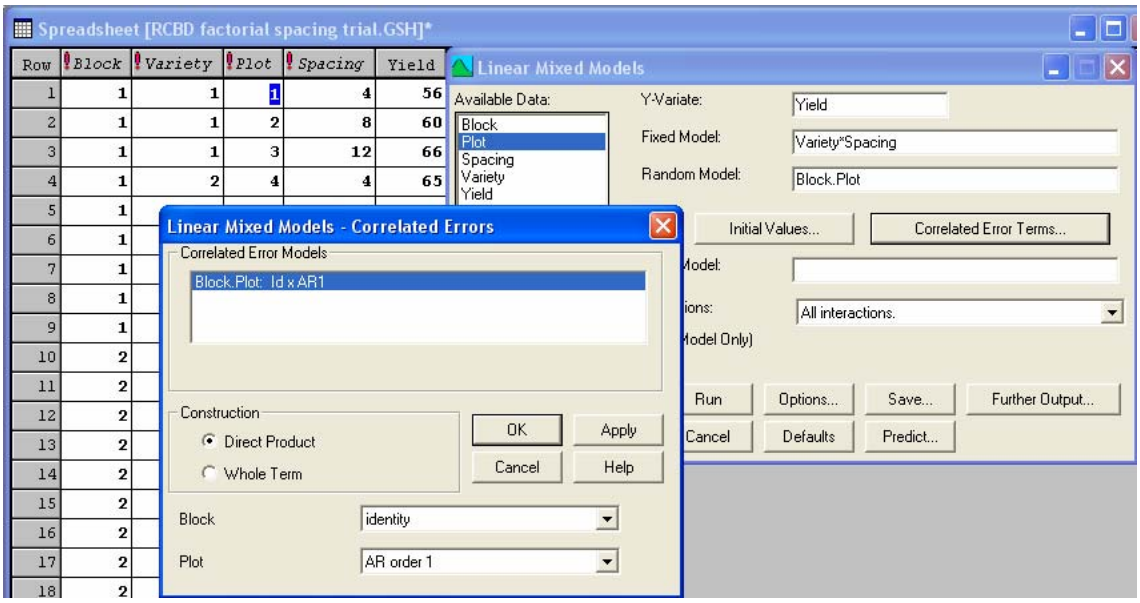
Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Variety	58.14	2	**29.07**	<0.001
Spacing	8.77	2	**4.39**	0.012
Variety.Spacing	43.32	4	**10.83**	<0.001

With Block a *fixed* component and Block.Plot the **Random Model** (Plot going from 1 to 9 in each block):

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Block	14.47	3	4.82	0.002
Variety	58.14	2	**29.07**	<0.001
Spacing	8.77	2	**4.39**	0.012
Variety.Spacing	43.32	4	**10.83**	<0.001

The estimated stratum variances from the ANOVA are $\hat{\sigma}^2_{Block}$ = 7.505 and $\hat{\sigma}^2$ = 17.671. This implies that the yields in any two plots in each block are uniformly correlated with an estimated correlation of 7.505/(7.505+17.671) 0.298. When you wish to use a correlated error structure in LMM (REML) you need to drop Block from the **Random Model**, and use just Block.Plot, since the correlation model supercedes the two random components model. (This is more fully described on page 656 in GenStat's *Statistics Guide* via the **Help** screen.)

Unfortunately, **Uniform** is not listed in the menu's available **Correlated Error Terms**, but it is an option in the actual procedure. The way around this is to run a different correlation structure, copy the appropriate lines of code to a new **Input Window**, modify the line and re-run the window of code. Here we chose AR1:




```
VCOMPONENTS [FIXED=Block+Variety*Spacing; FACTORIAL=9; CADJUST=none]
RANDOM=Block.Plot; INITIAL=1; CONSTRAINTS=none
VSTRUCTURE [TERMS=Block.Plot; FORMATION=direct] MODEL=identity,ar1;
ORDER=*,1; FACTOR=Block,Plot
```

```
REML [PRINT=model,components,deviance,waldTests; PSE=differences;
MVINCLUDE=*; METHOD=AI; MAXCYCLE=20] Yield
```

Change to **uniform**

REML variance components analysis

Response variate: Yield
Fixed model: Constant + Variety + Spacing + Variety.Spacing
Random model: Block.Plot
Number of units: 36

Block.Plot used as residual term with covariance structure as below

Covariance structures defined for random model

Covariance structures defined within terms:

Term	Factor	Model	Order	No. rows
Block.Plot	Block	Identity	1	4
	Plot	**Uniform**	1	9

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
		Block.Plot	Sigma2	**25.18**	8.96
	Block	Identity	-	-	-
	Plot	Uniform	theta1	**0.2981**	0.2286

Deviance: -2*Log-Likelihood

Deviance	d.f.
121.74	25

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Variety	58.14	2	29.07	<0.001
Spacing	8.77	2	4.39	0.012
Variety.Spacing	43.32	4	10.83	<0.001

We can work the formula for the uniform correlation backwards to calculate the block variance component. The estimate 25.18 is actually the *combined* estimate ($\hat{\sigma}^2_{Block} + \hat{\sigma}^2$). The uniform correlation is 0.2981 = $\hat{\sigma}^2_{Block} / (\hat{\sigma}^2_{Block} + \hat{\sigma}^2) = \hat{\sigma}^2_{Block} / 25.18$, so that $\hat{\sigma}^2_{Block}$ = 0.2981×25.18 = 7.506 (as was obtained earlier).

GenStat's examples go even further. To quote:

```
VCOMPONENTS [FIXED=Cv] RANDOM=Row.Column+'*units*'
VSTRUCTURE [TERM=Row.Column] MODEL=ar,ar; \
   ORDER=1; FACTOR=Row,Column
```

Here, the Row and Column terms have been removed from the random model, as they are superceded by the correlation from the composite term Row.Column. The term '*units*' has been retained to provide an estimate of independent random error in addition to that predicted by the AR(1)⊗AR(1) structure. This model might be interpreted as the correlation structure

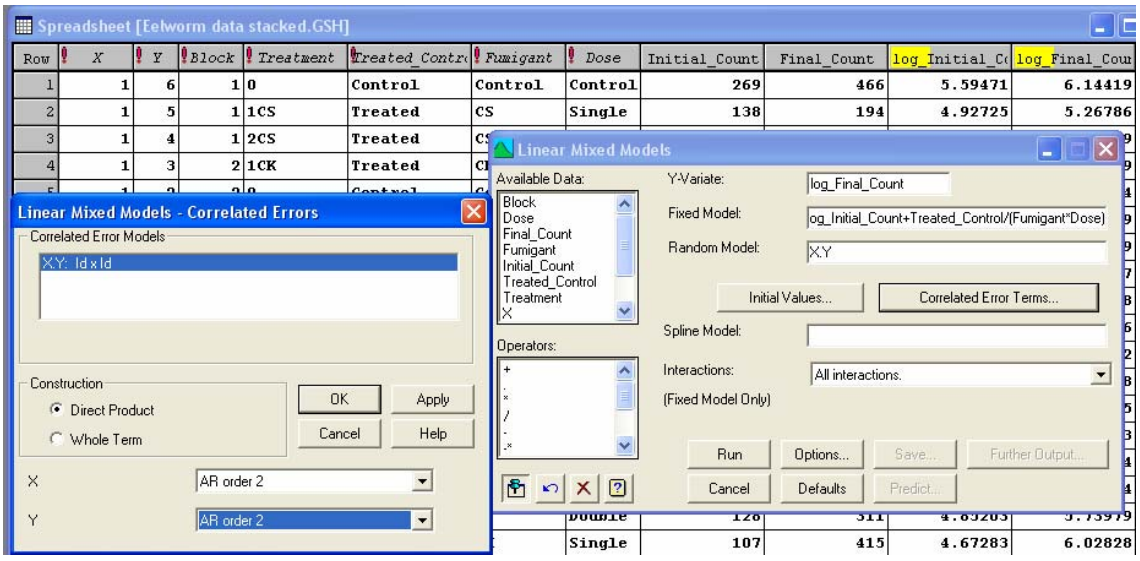
describing the underlying spatial trend in the field, with the extra residual accounting for experimental and measurement error in the data.

In field trials, it is unlikely that a uniform correlation applies spatially or temporally. It is more likely that plots closer together (in time or space) are more strongly correlated than plots further apart. Hence, AR models are very commonly used in the modern analyses of field trials. The example above does not have a known field plan, so we illustrate this with the eelworm data.

Firstly, the field really consists of plots in a row by column layout. The original layout had four blocks in a 2×2 layout. As hypothesized earlier, if there *is* a block effect, is it left to right across the field, or top to bottom, or both? If any of these, why is the gradient no reflected in the plots within a block?

To investigate these possibilities, we inserted a factor labelled Y with 6 levels, and a factor labelled X with 8 levels.

The **Random Model** is then X.Y with at most an AR2 ⊗ AR2 spatially correlated model. We do not expect exactly the same scaled Wald statistics as before, since the assumed error structure is now different.



We can use change in deviance to check whether we a less complex model is adequate.

Model for X.Y	deviance	df	Change in deviance	Change in df	*P*-value
AR2.AR2	-34.32	33			
AR2.AR1	-34.17	34	0.15	1	0.699
AR2.Identity	-34.12	35	0.05	1	0.823
AR1.Identity	-33.76	36	0.36	1	0.549
Identity.Identity	-28.50	37	5.26	1	0.022

It would appear that an AR1 correlated model left to right is what is required in this case. The analysis is as follows.

REML variance components analysis

Response variate: log_Final_Count
Fixed model: Constant + log_Initial_Count + Treated_Control + Treated_Control.Fumigant + Treated_Control.Dose + Treated_Control.Fumigant.Dose
Random model: X.Y
Number of units: 48

X.Y used as residual term with covariance structure as below

Covariance structures defined for random model

Covariance structures defined within terms:

Term	Factor	Model	Order	No. rows
X.Y	X	Auto-regressive (+ scalar)	1	8
	Y	Identity	0	6

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
X.Y	Sigma2	0.113	0.0306		
	X	AR(1)	phi_1	0.4127	0.1753
	Y	Identity	-	-	-

Deviance: -2*Log-Likelihood

Deviance	d.f.
-33.76	36

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
log_Initial_Count	61.99	1	61.99	<0.001
Treated_Control	17.34	1	17.34	<0.001
Treated_Control.Fumigant	27.00	3	9.00	<0.001
Treated_Control.Dose	0.92	1	0.92	0.337
Treated_Control.Fumigant.Dose	3.29	3	1.10	0.349

Means, and all standard errors of differences of means, are suppressed: they can be saved into an Excel file.

Lattice Design

The GenStat *Statistics Guide* is available via the **Help** menu. On page 672 is an analysis of an experiment set out as a lattice square with 6 replicates. The tables are transposed to fit on the page. In addition in the guide, the yields are divided by 100 for numerical accuracy.

Yields

	Row									
Column	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	1003	1531	1126	1261	1458	1623	1331	1211	1388	1443
2	1356	1540	1400	1423	2036	1862	1417	1411	1453	1667
3	1412	1250	1329	1110	2119	1645	1611	1183	1384	1549
4	1239	1658	1287	1735	1912	1888	1454	1550	1669	1459
5	1508	1185	1555	1617	1893	1527	1790	1660	1738	1722
6	1967	1605	1395	1820	1748	1606	1767	1526	1845	1583
7	1572	1550	1696	1351	1450	1842	1917	1681	1700	1490
8	1969	1500	1570	1297	1740	1186	1264	1545	1528	1607
9	1747	1642	1404	1412	1450	1462	1060	1290	1373	1315
10	1598	1504	1285	1506	1523	1242	951	976	1240	1174
11	1630	1680	1473	1512	1364	1082	1130	1240	1252	1443
12	1633	1526	1761	1355	1690	1304	1266	1181	1591	1649
13	1255	1452	1695	1524	1334	1267	1289	917	1428	1407
14	1277	1480	1364	1478	1239	1266	1260	1287	1509	1315
15	1572	1482	1790	1371	1557	1200	1174	975	1273	1318

Layout for varieties (1, …, 25)

	Row									
Column	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	1	6	21	11	16	3	1	5	2	4
2	2	7	22	12	17	18	16	20	17	19
3	4	9	24	14	19	8	6	10	7	9
4	3	8	23	13	18	13	11	15	12	14
5	5	10	25	15	20	23	21	25	22	24
6	19	8	11	22	5	16	12	4	25	8
7	23	12	20	1	9	24	20	7	3	11
8	2	16	24	10	13	10	1	18	14	22
9	6	25	3	14	17	13	9	21	17	5
10	15	4	7	18	21	2	23	15	6	19
11	18	5	6	24	12	10	12	19	21	3
12	25	7	13	1	19	4	6	13	20	22
13	9	16	22	15	3	17	24	1	8	15
14	11	23	4	17	10	11	18	25	2	9
15	2	14	20	8	21	23	5	7	14	16

The guide presents four analyses to demonstrate various features of spatial models.

Firstly, you can think of the layout as a randomised block design with 6 blocks.

Lattice designs were introduced to obtain more efficient estimates of the standard errors of differences of varietal means. In each replicate block, fertility trends in rows and columns meant that the varieties needed to be arranged in the rows and columns in balanced pairs in

such a way that each pairs of varieties is estimated with the same precision. These designs were very common in past years where the number of treatments was large, and often provided substantial reduction in the s.e.d. values, enabling for example better selection in breeding trials.

A modern analysis would impose a spatial correlation on the rows and columns in the two-way layout. In addition, as described earlier, it is possible to allow for extra measurement error (separate from the variance induced by the spatial process) by directly adding a term to identify the 150 individual error terms, or by using the in-built'*units*' term. In either case GenStat produces a warning about the two residual terms and tells you which one is to be used to provide the R matrix. For the two spatial models we are assuming a separable correlation model whereby the correlation between two plots in the field is a multiple of the correlation between them in the row direction and in the column direction: this is represented by Row ⊗ Column.

With the row and column factors specified in the layout above, the random models and deviances are obtained as follows:

Design	Random Model	Correlated Error Terms		
			deviance	d.f.
RCBD	Block	Identity	335.39	123
Lattice	Block/(Row*Column)	Identity for all terms	264.28	121
Row ⊗ Column	Row.Column	AR1×AR1	249.35	122
Row ⊗ Column+measurement error	Row.Column+'*Units*'	AR1×AR1+Identity	242.35	121

The s.e.d. values are as follows. For the two spatial models, the spatial process varies continuously along the rows and columns in the field, and hence the s.e.d. values will depend on the placement of the varieties in the field.

Design	s.e.d.
RCBD	1.075
Lattice	0.620
Row ⊗ Column	varies between 0.525 and 0.642, average is 0.590
Row ⊗ Column+measurement error	varies between 0.569 and 0.638, average is 0.605

The change in deviance between the last two spatial models is significant (P=0.008). The REML analysis for this model is as follows.

Response variate: Yield
Fixed model: Constant + Variety
Random model: Row.Column + '*units*'
Number of units: 150

Row.Column used as residual term with covariance structure as below

Covariance structures defined for random model

Term	Factor	Model	Order	No. rows
Row.Column	Row	Auto-regressive (+ scalar)	1	10
	Column	Auto-regressive	1	15

Estimated variance components

Random term	component	s.e.
Extra units term	**0.486**	0.179

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
		Row.Column	Sigma2	**4.580**	1.670
	Row	AR(1)	phi_1	**0.6827**	0.1023
	Column	AR(1)	phi_1	**0.8438**	0.0684

- Estimated variance of the measurement error
- Estimated variance of the spatial process
- The correlations in yields in neighbouring row plots, and in neighbouring column plots

Estimated covariance models

Variance of data estimated in form:

V(y) = Sigma2(gZZ' + R)

where: V(y) is variance matrix of data
Sigma2 is the residual variance
g is the gamma for the random term
Z is the incidence matrix for the random term
R is the residual covariance matrix

Note: a gamma is the ratio of a variance component to the residual (Sigma2)

Random Term: Extra units term
Scalar Sigma2*g: 0.4862

Residual term: Row.Column
Sigma2: 4.580

R uses direct product construction

Factor: **Row**
Model : Auto-regressive

Covariance matrix:

1	1.000									
2	0.683	1.000								
3	0.466	0.683	1.000							
4	0.318	0.466	0.683	1.000						
5	0.217	0.318	0.466	0.683	1.000					
6	0.148	0.217	0.318	0.466	0.683	1.000				
7	0.101	0.148	0.217	0.318	0.466	0.683	1.000			
8	0.069	0.101	0.148	0.217	0.318	0.466	0.683	1.000		
9	0.047	0.069	0.101	0.148	0.217	0.318	0.466	0.683	1.000	
10	0.032	0.047	0.069	0.101	0.148	0.217	0.318	0.466	0.683	1.000
	1	2	3	4	5	6	7	8	9	10

Factor: **Column**
Model : Auto-regressive

Covariance matrix (first 10 rows only):

1	1.000									
2	0.844	1.000								
3	0.712	0.844	1.000							
4	0.601	0.712	0.844	1.000						
5	0.507	0.601	0.712	0.844	1.000					
6	0.428	0.507	0.601	0.712	0.844	1.000				
7	0.361	0.428	0.507	0.601	0.712	0.844	1.000			
8	0.305	0.361	0.428	0.507	0.601	0.712	0.844	1.000		
9	0.257	0.305	0.361	0.428	0.507	0.601	0.712	0.844	1.000	
10	0.217	0.257	0.305	0.361	0.428	0.507	0.601	0.712	0.844	1.000
	1	2	3	4	5	6	7	8	9	10

Deviance: -2*Log-Likelihood

Deviance	d.f.
242.35	121

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Variety	245.39	24	10.22	<0.001

Table of predicted means for Variety

Variety	1	2	3	4	5	6	7	8
	12.46	15.16	14.04	14.05	14.72	15.22	13.73	14.53

Variety	9	10	11	12	13	14	15	16
	12.62	11.96	13.29	14.42	16.24	12.99	14.70	12.88

Variety	17	18	19	20	21	22	23	24
	14.93	15.27	16.49	16.46	15.15	16.09	13.17	15.58

Variety	25
	15.74

Standard errors of differences between pairs

Variety 1	1	*					
Variety 2	2	0.62	*				
Variety 3	3	0.62	0.63	*			
Variety 4	4	0.62	0.60	0.62	*		
Variety 5	5	0.63	0.63	0.61	0.61	*	
Variety 6	6	0.60	0.62	0.61	0.62	0.62	*

etc

More complex field designs: a split-strip plot experiment

We return to the experiment used on page 91 of this manual. There were four blocks, three fixed factors (4 cultivars × 2 row spacings × 4 target plant populations) in a five stratum layout. To obtain a better analysis than ANOVA, we use LMM (REML) with the following models:

Fixed Model: Cultivar*PlantPop*RowSpace
Random Model:
Block+Block.Cultivar+Block.Cultivar.Row+Block.Cultivar.Plant+Block.Cultivar.Row.Plant

REML variance components analysis

Response variate: Yield
Fixed model: Constant + Cultivar + PlantPop + RowsSpacing + Cultivar.PlantPop + Cultivar.RowsSpacing + PlantPop.RowsSpacing + Cultivar.PlantPop.RowsSpacing
Random model: Block + Block.Cultivar + Block.Cultivar.PlantPop + Block.Cultivar.RowsSpacing + Block.Cultivar.PlantPop.RowsSpacing
Number of units: 154 (6 units excluded due to zero weights or missing values)

Block.Cultivar.PlantPop.RowsSpacing used as residual term

Estimated variance components

Random term	component	s.e.
Block	3.037	2.896
Block.Cultivar	0.452	1.054
Block.Cultivar.PlantPop	2.421	1.017
Block.Cultivar.RowsSpacing	1.245	0.868

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
Block.Cultivar.PlantPop.RowsSpacing		Identity	Sigma2	3.927	0.835

Approximate stratum variances

Stratum	variance	effective d.f.
Block	133.551	3.00
Block.Cultivar	18.841	8.99
Block.Cultivar.PlantPop	8.688	45.83
Block.Cultivar.RowsSpacing	9.834	11.90
Block.Cultivar.PlantPop.RowsSpacing	3.927	44.29

Deviance: -2*Log-Likelihood

Deviance	d.f.
395.74	109

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Cultivar	24.38	3	8.13	<0.001
PlantPop	131.45	4	32.86	<0.001
RowsSpacing	3.52	1	3.52	0.061
Cultivar.PlantPop	14.69	12	1.22	0.259
Cultivar.RowsSpacing	18.85	3	6.28	<0.001
PlantPop.RowsSpacing	4.19	4	1.05	0.380
Cultivar.PlantPop.RowsSpacing	31.20	12	2.60	0.002

The similarities are clear, with the differences between the two analyses (apart from *P* values) due to the fact that REML uses just the data and ignores missing values.

However, we should investigate whether the variance changes with changing row spacing and changing plant population. Unfortunately, GenStat's analysis failed to converge when we tried this. To make headway, we tried the following.

The Block.Cultivar variance component is very small (0.452) and in fact can be deleted (the change in deviance is 395.96 - 395.74 = 0.22 with 1 d.f.). This is a simpler analysis which, apart from round-off error due to iteration with many parameters, produces the same variance components and close *P* values, with the exception that the individual Block and Block.Cultivar variance components of the first analysis (3.037 and 0.452) are replaced by a combined variance component of 3.465. This analysis is equivalent to treating the $b \times c$ plots (b blocks $\times$ c cultivars) as strips in the field into which the other factors are randomised (in two different ways). The analysis with changing variances for these factors did converge.

Random Model: Strip+ Strip.PlantPop+ Strip*RowSpace+ Strip.PlantPop.RowSpace, or simply Strip/(PlantPop*RowSpace)

Correlated Error Terms: use Identity⊗Diagonal⊗Diagonal for Strip.PlantPop.RowSpace

It turns out that that this more complex model is unnecessary, with a change in deviance of 401.01 – 397.48 = 3.53 with 5 d.f. (3.53 would be not significant if there was just 1 d.f.). Statistically, the first LMM (REML) analysis is the one to use for decisions; biologically, the plants within plots are competing to the point that a common variance model appears adequate.

The only point to add is that the design is unbalanced (with 6 missing values) and hence the *P* values depend on the order the factors are added to the model. As usual with unbalanced data, the *P* value to use for a factor should be the one obtained from an analysis with that factor entered last.

Multi-site experiments

The block and treatment structures used in the ANOVA of the data on page 102 were:

Treatment Structure: Strain+ Location.Strain
Block Structure: Location+ Location.Block

Placing Location in the **Block Structure** was purely a device to prevent the unreplicated factor Location from having a P-value printed in the ANOVA. The same analysis is produced when Location is placed in the **Treatment Structure**, but no stratum variance is obtained then (GenStat treats factors in the **Treatment Structure** as fixed terms).

Generally, when a factor is regarded as random then any interaction involving that factor is also random. With the Steel and Torrie data it is unclear whether the three locations, or the twelve strains, were randomly chosen or were of specific interest. It is common that Strains, and hence Strains.Location, are random, and that is what we will assume. What often occurs, moreover, is that the residual variances differ across locations. This was tested on page 102 via Bartlett's test of homogeneity of variance (and found to be not significant). Here we test it by change in deviance.

1. Reproducing the ANOVA

Treatment Structure: Location+Strain+ Location.Strain or simply Location*Strain
Block Structure: Location.Block

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block.Location stratum					
Location	2	3113626.	1556813.	134.87	<.001
Residual	6	69256.	11543.	0.59	
Block.Location.*Units* stratum					
Strain	11	925090.	84099.	4.27	<.001
Location.Strain	22	532900.	24223.	1.23	0.256
Residual	66	1300723.	19708.		
Total	107	5941596.			

Estimated stratum variances

Stratum	variance	effective d.f.	variance component
Block.Location	11542.7	6.000	-680.4
Block.Location.*Units*	19707.9	66.000	19707.9

Equivalent REML analysis:

Fixed Model: Location+Strain+ Location.Strain or simply Location*Strain
Random Model: Location.Block (+ Location.Strain.Block)

Estimated variance components

Random term	component	s.e.
Block.Location	-680.	625.

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
Residual		Identity	Sigma2	19708.	3431.

The Block.Location variance is estimated to be negative, which indicates that the Block.Location MS in the equivalent ANOVA is smaller than the combined Residual MS (11543 versus 19708, see page 102). Using a REML analysis, it is preferable to constrain Block.Location to be non-negative (via **Initial Values**). The estimate is given at the bound, namely zero:

REML variance components analysis

Response variate: Yield
Fixed model: Constant + Location + Strain + Location.Strain
Random model: Block.Location
Number of units: 108

Residual term has been added to model

Estimated variance components

Random term	component	s.e.
Block.Location	0.	bound

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
Residual		Identity	Sigma2	19027.	3171.

Wald tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Location	163.64	2	81.82	<0.001
Strain	48.62	11	4.42	<0.001
Location.Strain	28.01	22	1.27	0.175

2. Location fixed, Strain and Strain.Location random.

(a) Test whether the residual variances at each location are the same:

Fixed Model: Location
Random Model: Strain + Location.Strain + Location.Block + Location.Strain.Block
(Location.Block constrained to be positive)

Model	**Deviance**	**d.f.**	**χ^2 P-value**
Identity for Location in Location.Strain.Block	1172.04	101	
Diagonal for Location in Location.Strain.Block	1170.32	99	
Change	1.72	2	0.423

So, the simpler model with a constant residual variance at each location suffices (P = 0.423). The estimated variances at each location are slightly different to those used in Bartlett's test in the design section, because in this analysis we constrained the Location.Block term to be non-negative:

Location	Variance from individual ANOVAs	Variance from combined REML
Plymouth	24149	21778 ± 5896
Clayton	12124	13591 ± 3882
Clinton	22851	21217 ± 5818

The output from the constant variance model is as follows.

REML variance components analysis

Response variate: Yield
Fixed model: Constant + Location
Random model: Strain + Strain.Location + Location.Block + Strain.Location.Block
Number of units: 108

Strain.Location.Block used as residual term

Estimated variance components

Random term	component	s.e.
Strain	6653.	4066.
Strain.Location	1732.	2654.
Location.Block	0.	bound

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
	Strain.Location.Block	Identity	Sigma2	19027.	3171.

Deviance: -2*Log-Likelihood

Deviance	d.f.
1172.04	101

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Location	128.54	2	64.27	<0.001

Clearly there are yield differences across locations (P<0.001), but this is neither surprising nor of interest. As a breeding trial, we are more interested in strain differences. However, we need to determine firstly whether there are genotype × environment interactions.

To test whether the random Location.Strain interaction is significant is equivalent to testing whether the Location.Strain variance is 0. The estimate from the analysis above is 1732 ± 2564. However, we can only test this hypothesis using change in deviance, with the new model omitting the random term to be tested.

Model	Deviance	d.f.	χ^2 P-value
Including Location.Strain	1172.04	101	
Excluding Location.Strain	1172.55	102	
Change	0.51	1	0.475

This result indicates that strain differences are consistent across locations (P = 0.475).

Are there any differences among the strains themselves? Since Strain is also a random effect, we can only decide this by change in deviance. We take the no interaction model and drop Strain:

Model	Deviance	d.f.	χ^2 P-value
Including Strain	1172.55	102	
Excluding Strain	1186.87	103	
Change	14.32	1	<0.001

Strain differences are strongly significant (P<0.001). The final analysis is as follows.

REML variance components analysis

Response variate: Yield
Fixed model: Constant + Location
Random model: Strain + Location.Block + Strain.Location.Block
Number of units: 108

Strain.Location.Block used as residual term

Estimated variance components

Random term	component	s.e.
Strain	7095.	3998.
Location.Block	0.	bound

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
Strain.Location.Block		Identity	Sigma2	20243.	2953.

Deviance: -2*Log-Likelihood

Deviance	d.f.
1172.55	102

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Location	153.81	2	76.90	<0.001

The next question is how to estimate these effects. GenStat provides Best Linear Unbiased Predictor (BLUP) means and/or effects of random effects using the **Save** menu. Before looking at these, what are they? For the following discussion we are indebted to Keith Boldman (Global Data Analysis Methods, Monsanto Company, Iowa).

A BLUP estimate applies to random effects only. The Strain effect technically has a mean of zero, and a variance of σ_s^2 say. However, we really wish to predict the genotype mean for each strain. If we write the current model (omitting the random term Location.Block which has a zero variance and hence can be dropped from the model) as

$$Yield = \mu + stain\ effect + Error$$

At one extreme, we could use the i^{th} sample mean as an estimate of (μ + *stain effect*) for the i^{th} strain. This is appropriate when Strain is fixed, and is known as the Best Linear Unbiased Estimator (BLUE). This estimate is unbiased but may have a relatively large variance.

At the other extreme, with no genetic variance, the grand mean is the appropriate estimator for every strain. For our data, we have a genetic variance σ_s^2 which is significantly different to 0.

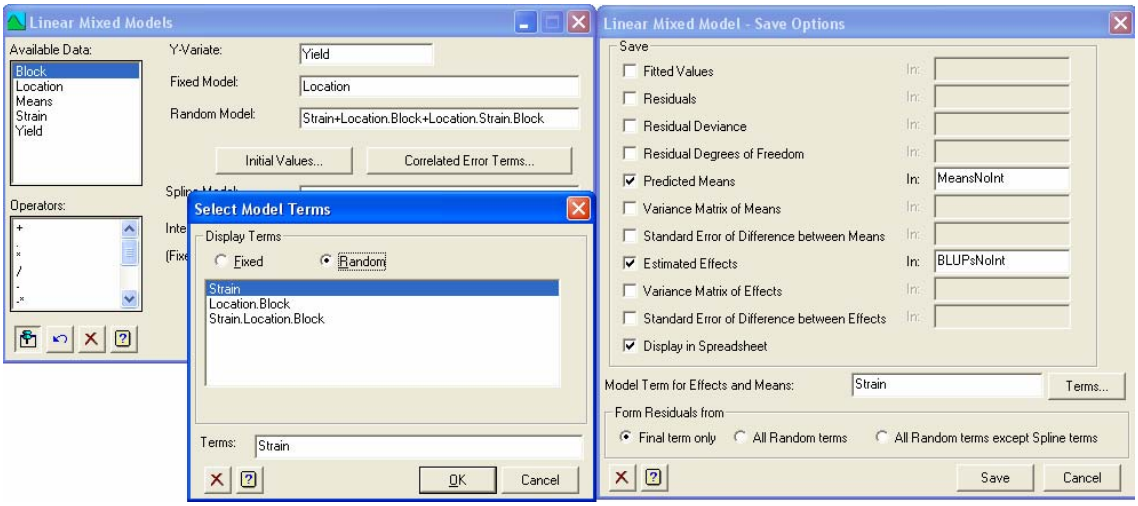
A BLUP mean is a compromise, or trade-off, between these two estimators. It is calculated by *shrinking* each sample strain mean somewhat toward the grand mean. The degree of shrinkage depends on the estimates of the genetic and environmental variance. The shrinkage ratio, h^2, is given by

$$h^2 = \frac{genetic\ var\ iance}{phenotypic\ var\ iance} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma^2 / r}$$

where r is the number of replicates of each strain and σ^2 is the residual variance. For our data, h^2 = 7095/(7095+20243/9) = 0.80. This ratio is applied to the *deviations* (differences between strain sample means and the grand mean). This reduces the various deviations, giving rise to BLUP effects and hence BLUP means. They are consequently "shrunk" toward the grand mean.

Strain	BLUP effect	BLUP Mean	Sample mean	deviation	BLUP effect /deviation
R75-12	-172.6	1231	1176	-227.4	0.8
N72-3058	-55.1	1348	1331	-72.6	0.8
Tracy	-25.3	1378	1370	-33.4	0.8
R73-81	-22.4	1381	1374	-29.5	0.8
Centennial	-6.4	1397	1395	-8.5	0.8
N73-882	-5.0	1398	1397	-6.6	0.8
N73-877	0.06	1403	1403	0.08	0.8
D74-7741	2.4	1406	1406	3.2	0.8
N73-693	25.0	1428	1436	33.0	0.8
N72-137	61.2	1464	1484	80.6	0.8
N73-1102	74.6	1478	1501	98.2	0.8
N72-3148	123.6	1527	1566	162.8	0.8
Overall	**0**	**1403**	**1403**	**0**	

The BLUP effects and BLUP means were captured using Save in GenStat. Select to display the possible random terms, as shown in the following screen capture. Double click on the random term whose BLUPS you wish to save.



Since the information on each strain is the same, the rankings of the strains are the same for the BLUPS as for the BLUES, but the BLUPs are smaller (closer to the grand mean) due to the shrinkage. In the previous table the strains were ranked in increasing order for the predicted yields.

Experiments repeated annually

The analysis of the asparagus yields of Example 12 is an example of the need for a temporal correlation model for plots measured annually.

Year in this case could be either fixed or random. We'll assume we are interested in the yields for these specific years, and take it as fixed.

Fixed Model: Year*CuttingTime
Random Model: Block+Year.Block

The ANOVA indicated that the estimated variance in the Block.Year stratum should be forced to 0 (in ANOVA the estimate was -82.55). Recall that this is achieved in the **Initial Values** menu.

REML variance components analysis

Response variate: Yield
Fixed model: Constant + Year + CuttingTime + Year.CuttingTime
Random model: Block.Year
Number of units: 64

Covariance structures defined for random model

Covariance structures defined within terms:

Term	Factor	Model	Order	No. rows
Block.Year	Block	Identity	1	4
	Year	Auto-regressive	1	4

Estimated parameters for covariance models

Random term(s)	Factor	Model(order)	Parameter	Estimate	s.e.
Block.Year	Block	Identity	Scalar	**0.7162**	0.6518
	Year	AR(1)	phi_1	0.9799	0.0914

Note: the covariance matrix for each term is calculated as G or R where
var(y) = Sigma2(ZGZ'+R), i.e. relative to the residual variance, Sigma2.

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
Residual		Identity	Sigma2	814.7	182.0

Estimated covariance models

Variance of data estimated in form:

V(y) = Sigma2(gZGZ' + I)

where: V(y) is variance matrix of data
Sigma2 is the residual variance
g is a gamma for the random term
Z is the incidence matrix for the random term
G is the covariance matrix for the random term
I is the residual (identity) covariance matrix

This is an estimate of the Error variance. In ANOVA it was estimated to be 895.98, however the Block.Year variance is negative for ANOVA and constrained to 0 in REML.

Note: a gamma is the ratio of a variance component to the residual (Sigma2)

Random Term: Block.Year

G uses direct product construction
Scalar Sigma2*g: 583.5

This is an estimate of the Block variance. In ANOVA it was estimated to be 593.17, from 0.7162×814.7. However, the Block.Year variance is negative for ANOVA and constrained to 0 in REML.

Factor: Block
Model : Identity (4 rows)

Factor: Year
Model : Auto-regressive

Covariance matrix:

1	1.000			
2	0.980	1.000		
3	0.960	0.980	1.000	
4	0.941	0.960	0.980	1.000
	1	2	3	4

The estimated correlation matrix over the four years has the AR1 structure: the lag-1 autocorrelation is r = phi_1 (0.9799) and the structure is r, r^2, r^3, r^4.

Residual term: added to model

Sigma2: 814.7

I is an identity matrix (64 rows)

Deviance: -2*Log-Likelihood

Deviance	d.f.
400.21	45

Wald tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Year	588.39	3	196.13	<0.001
CuttingTime	296.27	3	98.76	<0.001
Year.CuttingTime	62.82	9	6.98	<0.001

Table of predicted means for Constant

290.6 Standard error: 12.45

Table of predicted means for Year

Year	1930	1931	1932	1933
	179.5	299.4	427.7	255.9

Standard errors of differences between pairs

Year 1930	1	*			
Year 1931	2	10.4	*		
Year 1932	3	10.7	10.4	*	
Year 1933	4	10.9	10.7	10.4	*
		1	2	3	4

Standard errors of differences

Average:	10.56
Maximum:	10.91
Minimum:	10.38

Average variance of differences: 111.5

Table of predicted means for CuttingTime

CuttingTime	Jun_01	Jun_15	Jul_1	Jul_15
	356.6	322.9	290.8	192.2

Standard errors of differences between pairs

CuttingTime Jun_01	1	*			
CuttingTime Jun_15	2	10.1	*		
CuttingTime Jul_1	3	10.1	10.1	*	
CuttingTime Jul_15	4	10.1	10.1	10.1	*
		1	2	3	4

Standard error of differences: 10.09

Table of predicted means for Year.CuttingTime

CuttingTime	Jun_01	Jun_15	Jul_1	Jul_15
Year				
1930	216.3	175.8	188.8	137.3
1931	340.0	331.2	310.2	216.2
1932	499.0	484.7	433.0	294.0
1933	371.2	299.7	231.2	121.2

Standard errors of differences between pairs

Year 1930.CuttingTime Jun_01	1	*				
Year 1930.CuttingTime Jun_15	2	20.2	*			
Year 1930.CuttingTime Jul_1	3	20.2	20.2	*		
Year 1930.CuttingTime Jul_15	4	20.2	20.2	20.2	*	
Year 1931.CuttingTime Jun_01	5	20.3	20.3	20.3	20.3	*
Year 1931.CuttingTime Jun_15	6	20.3	20.3	20.3	20.3	20.2
Year 1931.CuttingTime Jul_1	7	20.3	20.3	20.3	20.3	20.2
Year 1931.CuttingTime Jul_15	8	20.3	20.3	20.3	20.3	20.2
Year 1932.CuttingTime Jun_01	9	20.5	20.5	20.5	20.5	20.3
Year 1932.CuttingTime Jun_15	10	20.5	20.5	20.5	20.5	20.3
Year 1932.CuttingTime Jul_1	11	20.5	20.5	20.5	20.5	20.3
Year 1932.CuttingTime Jul_15	12	20.5	20.5	20.5	20.5	20.3
Year 1933.CuttingTime Jun_01	13	20.6	20.6	20.6	20.6	20.5
Year 1933.CuttingTime Jun_15	14	20.6	20.6	20.6	20.6	20.5
Year 1933.CuttingTime Jul_1	15	20.6	20.6	20.6	20.6	20.5
Year 1933.CuttingTime Jul_15	16	20.6	20.6	20.6	20.6	20.5
		1	2	3	4	5

Etc. Better to save the s.e.d. values into an Excel spreadsheet

With this model, there are slightly smaller s.e.d. values for comparisons of cutting times within years (which was 21.17 in ANOVA).

General comments on repeated measures data

GenStat has a special menu for analyzing repeated measures data in various ways. Remember, when you have a field trial with plots that you repeatedly measure, you need to build into the analysis some sort of correlation structure.

The simplest approach is to think of time in a split-plot framework. Thus, if say varieties are replicated in a CRD in the field, varieties form the whole-plot treatment and time the split-plot treatment. We have seen that a model with two error terms induces a uniform correlation structure for units over time. This is unlikely to be a realistic assumption. Moreover, it can only be an approximation to a correct analysis. A split-plot experiment assumes that the split-plot treatments can be randomised to the split-plots; time cannot be jumbled around randomly: time 1 precedes time 2 and so on.

In the fifties Box suggested a test of this "uniform correlation structure". At about the same time two statisticians, Greenhouse and Geisler, independently arrived at a way of modifying the conventional split-plot analysis to cater for non-uniform correlation structures. These two tests are available in GenStat's **Stats** > **Repeated Measurements** > **Analysis of Variance** menu.

The Department of Agronomy & Range Science, University of California Davis used SAS to analyse an experiment in which four alfalfa cultivars were tested with 5 replications in a CRD, with repeated measurements (4 cuts of yields, Sept. 10/74; June 25/75; Aug. 5/75; Sept. 16/75 labeled Cut 1 to Cut 4). Source: www.agronomy.ucdavis.edu/agr205/Lectures/Topic12b.pdf

Replicate	Variety	Cut 1	Cut 2	Cut 3	Cut 4
1	1	2.80191	3.73092	3.09856	2.50965
1	2	2.76212	5.40530	3.82431	2.72992
1	3	2.29151	3.81140	2.92575	2.39863
1	4	2.56631	4.96070	2.81734	2.05752
2	1	2.96602	4.43545	3.10607	2.57299
2	2	3.09636	3.90683	3.26229	2.58614
2	3	2.54027	3.82716	2.86727	2.16287
2	4	2.31630	3.96629	2.91461	2.15764
3	1	2.43232	4.32311	2.81030	2.07966
3	2	3.09917	4.08859	3.13148	2.60316
3	3	2.41199	4.08317	3.03906	2.07076
3	4	2.65834	3.71856	2.92922	2.15684
4	1	2.93509	3.99711	2.77971	2.44033
4	2	2.65256	5.42879	2.70891	2.30163
4	3	2.30420	3.27852	2.72711	2.04933
4	4	2.47877	3.92048	3.06191	2.35822
5	1	2.42277	3.85657	3.24914	2.34131
5	2	2.63666	3.77458	3.09734	2.30082
5	3	2.36941	3.44835	2.50562	2.08980
5	4	2.23595	4.02985	2.85279	1.85736

To use the menu **Stats** > **Repeated Measurements** > **Analysis of Variance** you need to have this sort of layout, where each column represents a different set of measurements for one time. For a uniform correlation structure, the length of each period is irrelevant. The output is as follows.

Box's tests for symmetry of the covariance matrix

Chi-square 36.02 on 8 degrees of freedom: probability 0.000

F-test 4.49 on 8 and 3479 degrees of freedom: probability 0.000

The assumption that the correlations are uniform over time is strongly rejected ($P<0.001$).

Greenhouse-Geisser epsilon

epsilon 0.4984

This epsilon value is used to multiply the degrees of freedom in the split-plot part of the ANOVA prior to calculating *P* values. This is the form of the modification proposed by Greenhouse and Geisler, hence the name of the adjustment in GenStat. Note the comment following the ANOVA:

Analysis of variance

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Subject stratum					
Variety	3	2.84075	0.94692	7.40	0.003
Residual	16	2.04868	0.12804	1.34	
Subject.Time stratum					
d.f. correction factor 0.4984					
Time	3	37.44769	12.48256	130.46	<.001
Time.Variety	9	0.54772	0.06086	0.64	0.659
Residual	48	4.59261	0.09568		
Total	79	47.47745			

(d.f. are multiplied by the correction factors before calculating F probabilities)

Tables of means

Grand mean 2.994

Time	1	2	3	4
	2.599	4.100	2.985	2.291

Variety	1	2	3	4
	3.044	3.270	2.760	2.901

Time	Variety	1	2	3	4
1		2.712	2.849	2.383	2.451
2		4.069	4.521	3.690	4.119
3		3.009	3.205	2.813	2.915
4		2.389	2.504	2.154	2.118

Standard errors of differences of means

Table	Time	Variety	Time Variety
rep.	20	20	5
s.e.d.	0.0978	0.1132	0.2037
d.f.	23.93	16	38.56
Except when comparing means with the same level(s) of Variety			0.1956
d.f.			23.93

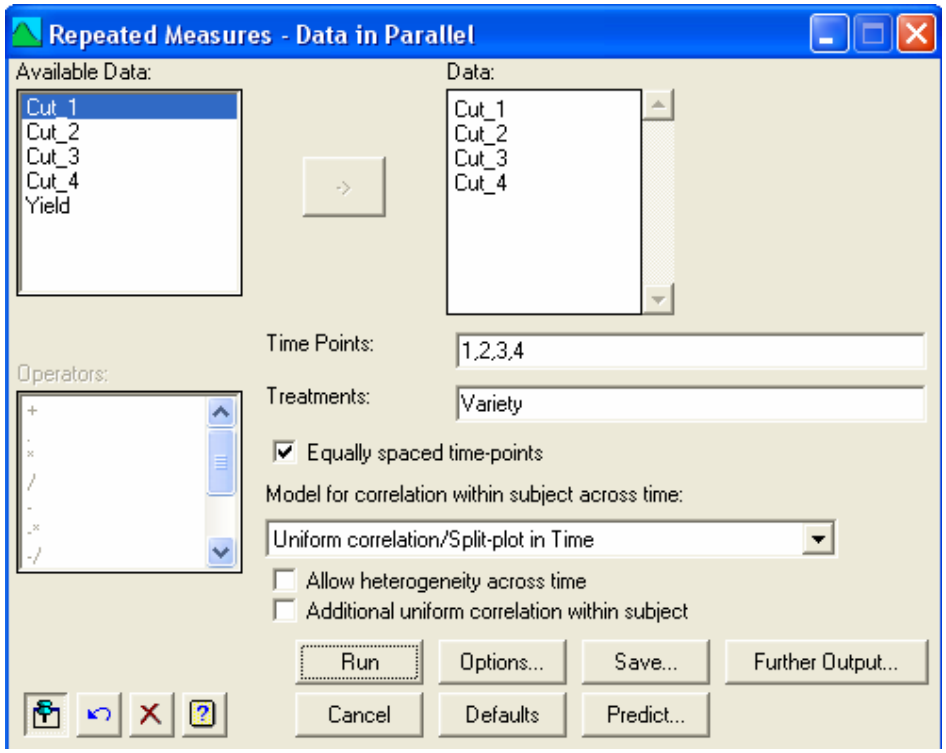
Correction factors have been applied to residual d.f.(see analysis-of-variance table for details)

The output is basically that of a split-plot ANOVA. If you stack the data and use the split-plot analysis of variance menu, you obtain this output (except that the *P* values are not modified). When you do this, the stratum variances turn out to be:

Estimated stratum variances

Stratum	variance	effective d.f.	**variance component**
Replicate.Variety	0.1280	16.000	**0.0081**
Replicate.Variety.CuttingDate	0.0957	48.000	**0.0957**

The split-plot analysis of variance can be reproduced in **Stats** > **Repeated Measurements** > **Correlation Models by REML**. You can perform this analysis using either the **Data in Multiple Variates** layout as above, or **Data in One Variate**, i.e. stacked. You can see that the default analysis assumes **Uniform Correlation** and equates this to a split-plot in time. It is irrelevant what time points you use, but the menu requires values.



REML variance components analysis

Response variate: _Data
Fixed model: Constant + %_Time + %_Variety + %_Time.%_Variety
Random model: %_subject.%_Time
Number of units: 80

%_subject.%_Time used as residual term with covariance structure as below

Covariance structures defined for random model

Covariance structures defined within terms:

Term	Factor	Model	Order	No. rows
%_subject.%_Time	%_subject	Identity	1	20
	%_Time	Uniform	1	4

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
%_subject.%_Time	Sigma2	0.104	0.0185		
	%_subject	Identity	-	-	-
	%_Time	Uniform	**theta1**	**0.07797**	0.11612

Deviance: -2*Log-Likelihood

Deviance	d.f.
-55.78	62

Wald tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
%_Time	391.39	3	130.46	<0.001
%_Variety	22.19	3	7.40	<0.001
%_Time.%_Variety	5.72	9	0.64	0.767

Table of predicted means for Constant

2.994 Standard error: 0.0400

Table of predicted means for %_Time

%_Time	1	2	3	4
	2.599	4.100	2.985	2.291

Standard error of differences: 0.09782

Table of predicted means for %_Variety

%_Variety	1	2	3	4
	3.044	3.270	2.760	2.901

Standard error of differences: 0.1132

Table of predicted means for %_Time.%_Variety

%_Variety	1	2	3	4
%_Time				
1	2.712	2.849	2.383	2.451
2	4.069	4.521	3.690	4.119
3	3.009	3.205	2.813	2.915
4	2.389	2.504	2.154	2.118

Standard errors of differences

Average:	0.2021
Maximum:	0.2037
Minimum:	0.1956

Average variance of differences: 0.04086

Standard error of differences for same level of factor:

	%_Time	%_Variety
Average:	0.2037	0.1956
Maximum:	0.2037	0.1956
Minimum:	0.2037	0.1956

This output appears similar to that from the split-plot ANOVA. Notice:

- The Wald/d.f. statistics for Variety, Time and Variety.Time are the same as the variance ratios from the ANOVA,as are s.e.d. values.
- The estimate Sigma2 (0.104) is actually the *combined* variance $\hat{\sigma}^2_{Block} + \hat{\sigma}^2$.
- Stratum variances need to be generated if you used the REML analysis but you need to understand how theta (θ) was generated. On page 148 we saw that

  $\theta = \hat{\sigma}^2_{Block} / (\hat{\sigma}^2_{Block} + \hat{\sigma}^2)$

  and thus $\hat{\sigma}^2_{Block} = \theta(\hat{\sigma}^2_{Block} + \hat{\sigma}^2) = 0.07797 \times 0.104 = 0.0081$.

A more realistic assumption is one which has the correlation reducing over time. Typically, an AR1 or AR2 model would be used. Unfortunately these assume that the cutting dates are equally spaced, which of course they are not. The alternative correlation structure for unequal times is a power model. Such a model is the same as an AR1 model when the times *are* equally spaced. When they are not, the correlation between two times k units apart is r^k.

For the cutting dates, we want to avoid using a time period which indices arithmetic overflow in calculating r^k terms. For example, if we define the Time Points as the number of days since Sep. 10 1974, we would need to use something like 0, 288, 329, 371, resulting in an overflow error with the only hint appearing in the estimate of r (phi_1):

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
%_subject.%_Time	Sigma2	0.104	0.0183		
	%_subject	Identity	-	-	-
	%_Time	Power(1)	phi_1	**0.1000**	**aliased**

.

So, if we used a 90-day time period we would use 0, 9.6000, 10.9667, 12.3667 resulting in an estimate phi_1 = 0.2699. If we used a 30-day time period we would use 0, 3.2000, 3.6556, 4.1222 resulting in an estimate phi_1 = 0.01968. These are not inconsistent. The former periods are three times the length of the latter periods, and $0.2699^3 = 0.01968$.

The output for the 30-day periods is as follows:

REML variance components analysis

Response variate: _Data
Fixed model: Constant + %_Time + %_Variety + %_Time.%_Variety
Random model: %_subject.%_Time
Number of units: 80

%_subject.%_Time used as residual term with covariance structure as below

Covariance structures defined for random model

Covariance structures defined within terms:

Term	Factor	Model	Order	No. rows
%_subject.%_Time	%_subject	Identity	1	20
	%_Time	Power - city block distance	1	4

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
%_subject.%_Time	Sigma2	0.104	0.0185		
	%_subject	Identity	-	-	-
	%_Time	Power(1)	phi_1	0.01968	0.04401

Estimated covariance models

Variance of data estimated in form:

V(y) = Sigma2.R

where: V(y) is variance matrix of data
Sigma2 is the residual variance
R is the residual covariance matrix

Residual term: %_subject.%_Time

Sigma2: **0.1036**

R uses direct product construction

Factor: %_subject
Model : Identity (20 rows)

Factor: %_Time
Model : Power - city block distance

Covariance matrix:

1	1.000			
2	0.000	1.000		
3	0.000	0.167	1.000	
4	0.000	0.027	0.160	1.000
	1	2	3	4

Deviance: -2*Log-Likelihood

Deviance	d.f.
-56.22	62

Wald tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
%_Time	370.80	3	123.60	<0.001
%_Variety	23.96	3	7.99	<0.001
%_Time.%_Variety	5.57	9	0.62	0.782

Table of predicted means for Constant

2.994 Standard error: 0.0390

Table of predicted means for %_Time

%_Time	0.000	3.200	3.656	4.122
	2.599	4.100	2.985	2.291

Standard errors of differences

Average:	0.09866
Maximum:	0.1018
Minimum:	0.09289

Average variance of differences: 0.009749

Table of predicted means for %_Variety

%_Variety	1	2	3	4
	3.044	3.270	2.760	2.901

Standard error of differences: 0.1104

Table of predicted means for %_Time.%_Variety

%_Variety	1	2	3	4
%_Time				
0.000	2.712	2.849	2.383	2.451
3.200	4.069	4.521	3.690	4.119
3.656	3.009	3.205	2.813	2.915
4.122	2.389	2.504	2.154	2.118

Standard errors of differences

Average:	0.2023
Maximum:	0.2036
Minimum:	0.1858

Average variance of differences: 0.04095

Standard error of differences for same level of factor:

	%_Time	%_Variety
Average:	0.2036	0.1973
Maximum:	0.2036	0.2036
Minimum:	0.2036	0.1858

Average variance of differences:

0.04144	0.03900

You can, instead, request **Standard Errors** of **All Differences** if you wish

Previously, we suggested that the variance very often changes with plants growing over time. In this case it may well be that repeated cutting of alfalfa gives yields with changing variances over time. To do this analysis, simply click **Allow heterogeneity across time**.

Model	**deviance**	**d.f.**	
Same variance	-56.22	62	
Changing variance	-84.74	59	
Change in deviance	28.52	3	$P<0.001$

It would seem the more complex model is required here.

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
		%_subject.%_Time	Sigma2	1.000	fixed
	%_subject	Identity	-	-	-
	%_Time	Power(1) het	phi_1	0.1012	0.1064
			Scale row 1	0.04132	0.01461
			Scale row 2	0.3061	0.1115
			Scale row 3	0.06389	0.02264
			Scale row 4	0.02939	0.01011

The overall variance of 0.104 of the former analysis changes now to one with variances 0.041 in September 1974, 0.306 in June, 0.064 in August and 0.029 in September of 1975.

The change in test statistics is:

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
%_Time	269.65	3	89.88	<0.001
%_Variety	35.93	3	11.98	<0.001
%_Time.%_Variety	4.02	9	0.45	0.910

and there are resulting changes also to s.e.d values.

Finally, the previous models forced the correlation structure to be of particular forms. It is possible to allow GenStat to estimate the complete correlation matrix over time, free of any structure. To do this, select **Unstructured** from the list. This allows both unequal variances, and correlations that are estimated separately. We assess this model against the unequal variance model:

Model	**deviance**	**d.f.**	
Changing variance	-84.74	59	
Changing variance and changing correlations	-96.18	54	
Change in deviance	11.44	5	*P*=0.043

The change in test statistics now is:

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
%_Time	515.18	3	171.73	<0.001
%_Variety	27.52	3	9.17	<0.001
%_Time.%_Variety	6.13	9	0.68	0.726

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
%_subject.%_Time	Sigma2	1.000	fixed		
	%_subject	Identity	-	-	-
	%_Time	Unstructured	v_11	0.04131	0.01461
			v_21	-0.01337	0.02699
			v_22	0.2779	0.0982
			v_31	0.003430	0.012864
			v_32	0.01267	0.03344
			v_33	0.06381	0.02256
			v_41	0.02012	0.01040
			v_42	-0.0007076	0.0236023
			v_43	0.02640	0.01309
			v_44	0.03207	0.01134

Here v_11 is the variance at time 1, v_12 the covariance between time 1 and time 2, and so on. If you request it, GenStat gives the rearranged out put for the covariance matrix:

Factor: %_Time
Model : Unstructured

Covariance matrix:

1	0.0413			
2	-0.0134	0.2779		
3	0.0034	0.0127	0.0638	
4	0.0201	-0.0007	0.0264	0.0321
	1	2	3	4

Correlations are obtained by dividing covariances by appropriate standard deviations:

Covariance matrix:

1	1			
2	-0.125	1		
3	0.066	0.095	1	
4	0.552	-0.007	0.583	1
	1	2	3	4

As before, there are resulting changes to s.e.d values.

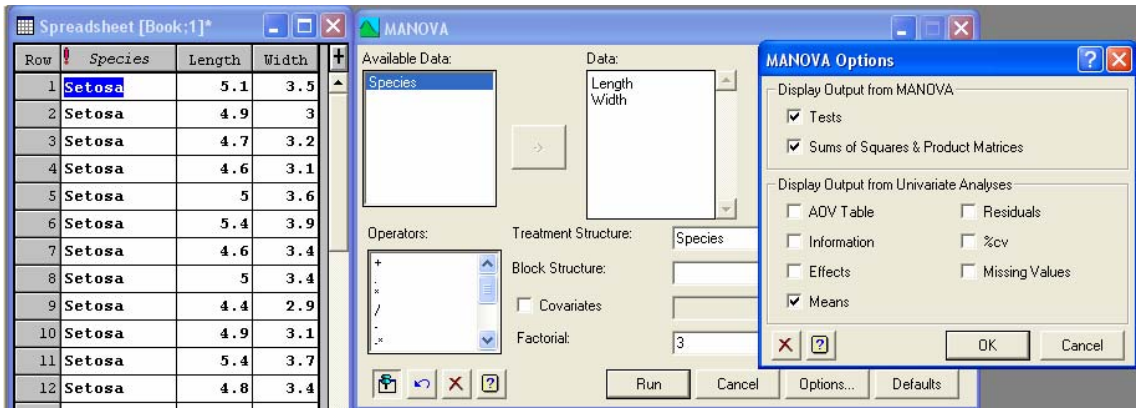
Multivariate analysis of variance via REML

REML offers an alternative to MANOVA which becomes very useful for unbalanced data.

As an example, consider Fisher's famous sepal length and width data. There were 50 randomly selected plants of two species. The example is generally used to illustrate discriminant analysis, but we will use it for MANOVA.

Length	Width	Length	Width	Length	Width	Length	Width	Length	Width
Setosa									
5.1	3.5	5.4	3.7	5.4	3.4	4.8	3.1	5.0	3.5
4.9	3.0	4.8	3.4	5.1	3.7	5.4	3.4	4.5	2.3
4.7	3.2	4.8	3.0	4.6	3.6	5.2	4.1	4.4	3.2
4.6	3.1	4.3	3.0	5.1	3.3	5.5	4.2	5.0	3.5
5.0	3.6	5.8	4.0	4.8	3.4	4.9	3.1	5.1	3.8
5.4	3.9	5.7	4.4	5.0	3.0	5.0	3.2	4.8	3.0
4.6	3.4	5.4	3.9	5.0	3.4	5.5	3.5	5.1	3.8
5.0	3.4	5.1	3.5	5.2	3.5	4.9	3.6	4.6	3.2
4.4	2.9	5.7	3.8	5.2	3.4	4.4	3.0	5.3	3.7
4.9	3.1	5.1	3.8	4.7	3.2	5.1	3.4	5.0	3.3
Versicolour									
7.0	3.2	5.0	2.0	5.9	3.2	5.5	2.4	5.5	2.6
6.4	3.2	5.9	3.0	6.1	2.8	5.5	2.4	6.1	3.0
6.9	3.1	6.0	2.2	6.3	2.5	5.8	2.7	5.8	2.6
5.5	2.3	6.1	2.9	6.1	2.8	6.0	2.7	5.0	2.3
6.5	2.8	5.6	2.9	6.4	2.9	5.4	3.0	5.6	2.7
5.7	2.8	6.7	3.1	6.6	3.0	6.0	3.4	5.7	3.0
6.3	3.3	5.6	3.0	6.8	2.8	6.7	3.1	5.7	2.9
4.9	2.4	5.8	2.7	6.7	3.0	6.3	2.3	6.2	2.9
6.6	2.9	6.2	2.2	6.0	2.9	5.6	3.0	5.1	2.5
5.2	2.7	5.6	2.5	5.7	2.6	5.5	2.5	5.7	2.8

To perform a MANOVA in GenStat, use **Stats** > **Multivariate Analysis** > **MANOVA**. We chose to include sums of squares and products matrices, as well as species means.



GenStat offers a choice of four test statistics.

Tables of means

Variate: Length

Grand mean 5.471

Species	Setosa	Versicolour
	5.006	5.936

Standard errors of differences of means

Table	Species
rep.	50
d.f.	98
s.e.d.	0.0884

Tables of means

Variate: Width

Grand mean 3.099

Species	Setosa	Versicolour
	3.428	2.770

Standard errors of differences of means

Table	Species
rep.	50
d.f.	98
s.e.d.	0.0696

Multivariate analysis of variance

Species

SSP-matrix, with 1 degree of freedom

Length	21.62		
Width	-15.30	10.82	
	Length	Width	

Tests

Wilk's Lambda: 0.1643
Approximate Chi sq: 175.20 on 2 d.f.; probability < 0.001
Approximate F test: 246.73 on 2 and 97 d.f.; probability < 0.001
Pillai-Bartlett trace: 0.8357
Roy's maximum root test: 0.8357
Lawley-Hotelling trace: 5.087

Residual SSP matrix, with 98 degrees of freedom

Length	19.14		
Width	9.04	11.87	
	Length	Width	

There is strong statistical evidence (P<0.001) that, based on length and width, the two species are different.

A sum of squares and products (SSP) matrix is a generalization of the component sum of squares of an ANOVA. In an ANOVA, the Residual MS is the appropriate estimate of σ^2.

In a MANOVA, the Residual SSP matrix needs to be divided by its *df* to produce the estimate of the population variance-covariance matrix of the variates. This matrix then provides an estimate of the correlation between length and width, since species effects are removed.

$$\textit{estimate of the population variance-covariance matrix} = \begin{pmatrix} 19.14/98 & 9.04/98 \\ 9.04/98 & 11.87/98 \end{pmatrix} = \begin{pmatrix} 0.195 & 0.092 \\ 0.092 & 0.121 \end{pmatrix}$$

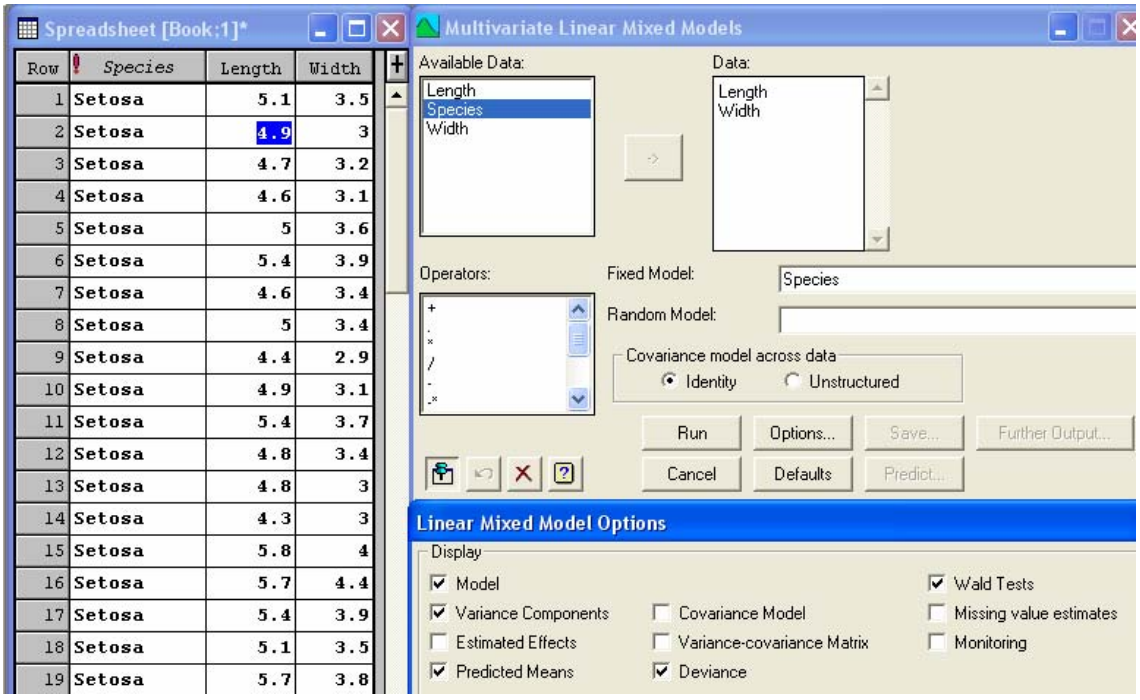
.

The correlation is estimated to be $0.092/\sqrt{(0.195\times0.121)} = 0.60$.

Analysis via LMM (REML)

Next, use **Stats** > **Mixed Models (REML)** > **Multivariate Linear Mixed Models**. Enter the data and fixed model as before, and request means and deviances.

It is possible that lengths and widths are uncorrelated. By selecting an **Identity** for the **Covariance matrix across data**, you are basically running a univariate LMM (REML) for each of the two variates. By selecting an **Unstructured**, you are setting up a correlation between the variates being analysed. Whether this approach is superior is tested by change in deviance as usual.



The analysis of deviance produces a strongly significant change in deviance (P<0.001):

Model:	Deviance	d.f.
Identity	-155.29	194
Unstructured	-198.94	193
change	**43.65**	**1**

REML variance components analysis

Response variate: _Data
Fixed model: %_variable + %_variable.%_Species
Random model: %_units.%_variable
Number of units: 200

Covariance structures defined for random model

Covariance structures defined within terms:

Term	Factor	Model	Order	No. rows
%_units.%_variable	%_units	Identity	0	100
	%_variable	Unstructured	1	2

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
	%_units.%_variable		Sigma2	1.000	fixed
	%_units	Identity	-	-	-
	%_variable	Unstructured	**v_11**	**0.1953**	0.0279
			v_21	**0.09220**	0.01811
			v_22	**0.1211**	0.0173

Deviance: -2*Log-Likelihood

Deviance	d.f.
-198.94	193

Note: deviance omits constants which depend on fixed model fitted.

Wald tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
%_variable	15667.11	2	7833.56	<0.001
%_variable.%_Species	**498.55**	**2**	**249.27**	**<0.001**

Table of predicted means for %_variable

%_variable	Length	Width
	5.471	3.099

Standard error of differences: 0.03633

Table of predicted means for %_variable.%_Species

%_Species	Setosa	Versicolour
%_variable		
Length	5.006	5.936
Width	3.428	2.770

Standard errors of differences between pairs

%_variable Length.%_Species Setosa	1	*				
%_variable Length.%_Species Versicolour	2	**0.088**	*			
%_variable Width.%_Species Setosa	3	0.051	0.080	*		
%_variable Width.%_Species Versicolour	4	0.080	0.051	**0.070**	*	
		1	2	3	4	

Notice:

- The first Wald statistic (%_variable) tests the overall mean lengths and overall mean widths are the same. This is a silly hypothesis, and should be ignored. It is needed in the output because GenStat basically stacks the different variates into one long column, and provides a factor (%_variable) to identify which variate each row is taken from.

- The hypothesis that the *vector* of variate means for Setosa is the same as that for Versicolour is tested by the apparent interaction, %_variable.%_Species. Notice that the test value, 249.27, is almost identical to the approximate F statistic of Wilk's Lambda from the MANOVA (246.73).

- The variances (0.195 for length, and 0.121 for width) and the covariance (0.092) from the MANOVA are identical to those produced by REML. This should not be a surprise, since we have seen that for a balanced design, the Residual MS from ANOVA is actually a REML estimate of variance, and this therefore extends to MANOVA. REML, however, is available for unbalanced designs as well.

- The means from the two analyses are identical.

- The s.e.d values from MANOVA for comparing species means for both length (0.0884) and width (0.0696) are identical to those produced by REML (in bold in the output above).

Analysis of non-normal data

Background

With normally distributed data, the distribution involves a mean parameter μ and a variance parameter σ^2. As we have seen in the previous sections, the model for data from one population can be expressed as

$$Y = \mu + Error$$

where *Error* is N(0, σ^2). We used maximum likelihood or residual maximum likelihood to estimate μ and σ^2.

For non-normal data, it is generally not possible to impose an additive model such as this. For example, if *Y* is binomial with known *n* and unknown *p*, we can write down a likelihood expression and maximise it to estimate *p*. If *Y* is Poisson with unknown mean μ, we can write down a likelihood expression and maximise it to estimate μ.

When we come to many treatments involving binomial or Poisson data, we need to ensure that the maximum likelihood estimates are properly defined, in particular the probability of a success in each treatment must fall in the region (0,1), while for Poisson data each mean must be positive.

Finney was among the first to suggest a method for analysing binomial data for designed experiments involving herbicides, insecticides and so on. The method became known as probit analysis. More often these days, scientists in this area will use logistic regression.

The modern method of analysing non-normal data is by maximum likelihood, in which the mean is modelled on a scale guaranteed to produce well defined estimates.

For Poisson data, we generally assume that

$$E(Y) = \mu = e^{b_0 + b_1 X_1 + \ldots}$$

where X_1, …could be covariates to explain the change in the Poisson mean, or design features (treatment effects and so on). Thus,

$$\log_e(\mu) = b_0 + b_1 X_1 + \ldots$$

We call this a *linear predictor with a log-link*. Estimation of the parameters in the linear function is by ML.

For binomial data, Finney noticed that the percentage of insects dying at low doses was small, increased rapidly as the dose increased and obviously asymptoted to 100% kill with sufficiently high dose. He noted that such a shape is typical of the cumulative distribution function of a normal variable, and proposed that method to estimate the parameters of the binomial. As mentioned these days the logistic distribution is more usual. We allow the probability of a success to depend on linear predictors via the logistic

$$p = \frac{1}{1+e^{-(b_0+b_1X_1+\cdots)}}.$$

This can be transformed as follows. Note that

$$\frac{p}{1-p} = e^{(b_0+b_1X_1+\cdots)}.$$

This is known as the *odds*. If you toss a fair coin you have a 50:50 chance of a head, or an odds of 1. If seeds have about an 80% germination rate, the odds are 0.8:0.2, or 4:1 – an odds of 4.

Taking logs now gives

$$log_e\left(\frac{p}{1-p}\right) = b_0 + b_1X_1 + \cdots$$

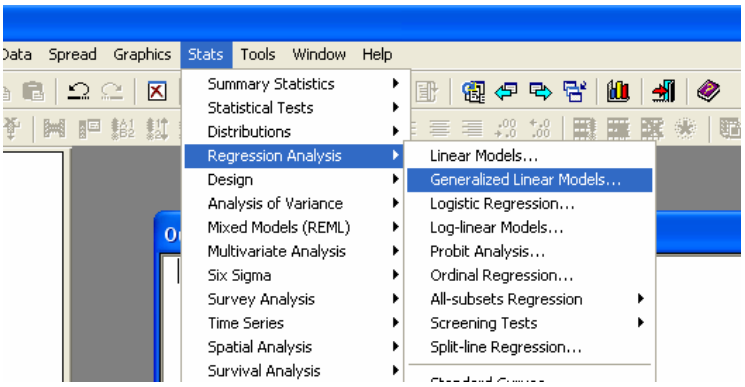
Thus, the link for the binomial is known as the logit link.

Once you estimate the parameters of this linear predictor, you calculate the odds, then the estimate of the probability:

$$probability = \frac{odds}{1+odds}.$$

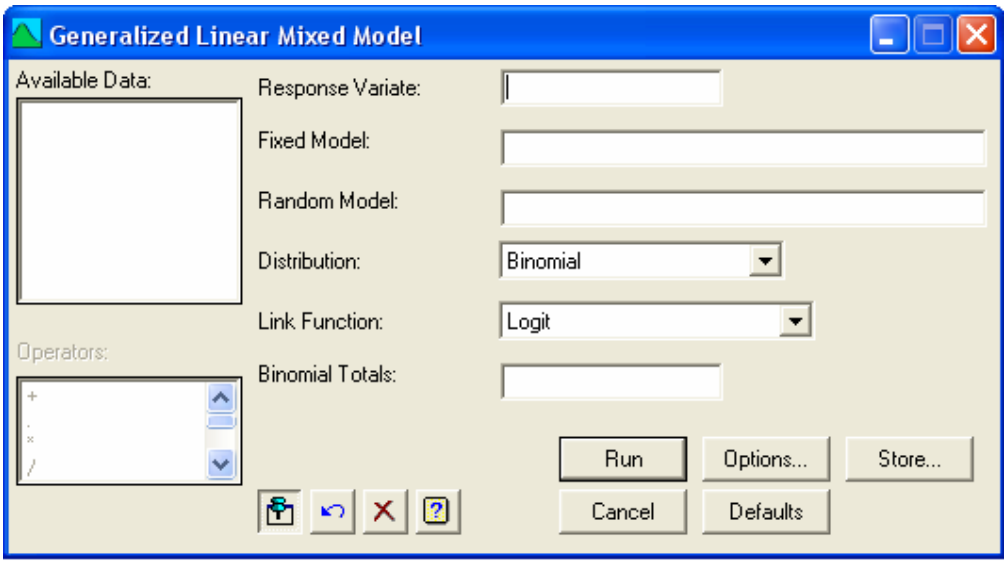
To summarise, for non-normal data, for each distribution we have a different linear predictor and link function, use maximum likelihood to estimate the parameters of the linear predictor and use change in deviance to compare models.

This type of regression model is called a *generalised linear model*. GenStat has a general menu for this, and you can select the distribution and link function from that menu. It also offers specific choices in the first **Regression Analysis** menu itself (**Logistic Regression** or **Probit Analysis** for binomial data, **Log-linear Models** for Poisson data).

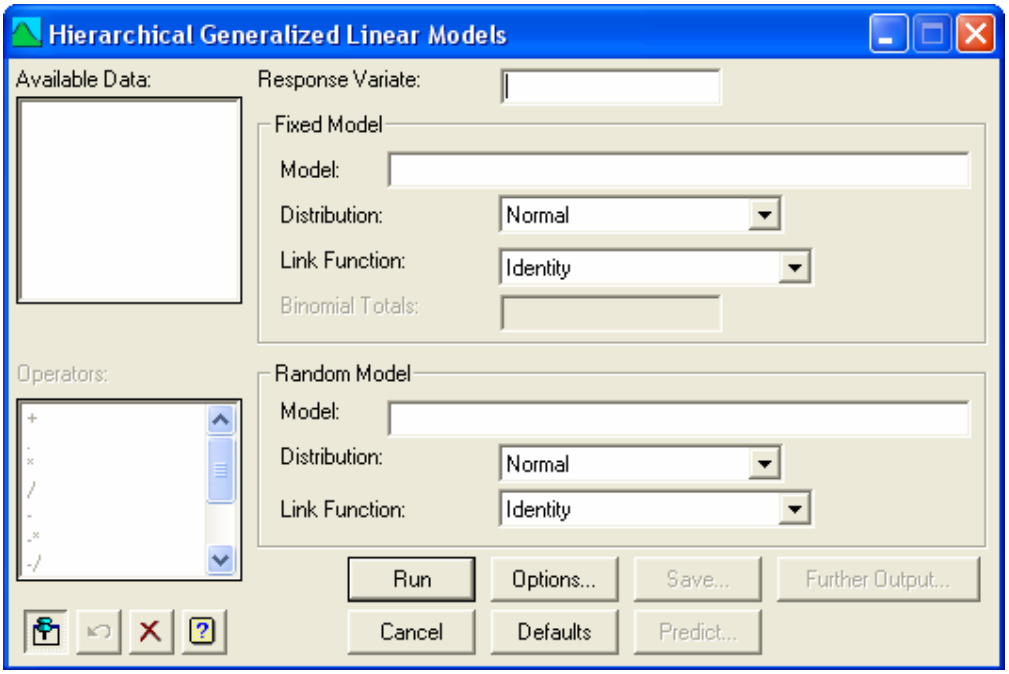


In the **Model to be Fitted** you should really only enter fixed effects. These correspond to the set of predictors X_1, X_2, … in the linear predictor. If you do have an experiment involving random effects, then a different menu is available. We are now dealing with mixed models

again (fixed and random effects) and the menu, for a **Generalized Linear Mixed Model** (GLMM), is therefore available via **Mixed Models (REML)**. The random effects are assumed to be normal for GLMMs.



If you believe that the random effects have a non-normal distribution then the analysis is very complex. There is a new menu with a selection of distributions to choose from for the random effects. Again, select **Mixed Models (REML)** and **Hierarchical Generalized Linear Mixed Models** (HGLMM).



We include one example of GLMMs in this manual, but leave the more complex HGLMMs for another occasion.

Binary logistic regression

Firstly, let us take the 2×2 contingency table, where the rows represent different treatments. (There are other types of contingency tables, some of which we consider later in this section.)

Example 1 Incidence of rust in Kentucky bluegrass pastures, from Steel and Torrie, page 504

Pasture field type	Rust	No Rust	Total
1	372	24	396
5	330	48	378

Readers may be familiar with Pearson's χ^2 goodness of fit statistic used to test whether the probability of rust is the same for the two pasture types.

$$X^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$ asymptotically χ^2 with (*rows*-1)(*columns*-1) *df.*

Under a hypothesis of equal probabilities, the best (ie ML) estimate of "Rust" is p=(372+330)/(396+378) = 0.907. This allows us to work out how many rust-affected clonal isolations are expected for each pasture type. (For example, for pasture type 1, we expect 0.907×396=359.2 to be affected.)

This test is available in GenStat. However, a more common test is now used, the maximum likelihood χ^2 test. It is, in fact, the same as the deviance in a binary logistic analysis of these data.

$$X^2 = 2\sum Observed \times ln\left(\frac{Observed}{Expected}\right)$$ asymptotically χ^2 with (*rows*-1)(*columns*-1) *df.*

Choose **Stats** > **Statistical Tests** > **Contingency Tables**. If you have not already done so, click **Create Table** and choose **Spreadsheet**. Then enter or copy the data to the table, and click back to the menu. Choose the **Method** (**Pearson** or **Maximum Likelihood**) and, in **Options**, if you wish to see expectations or not.

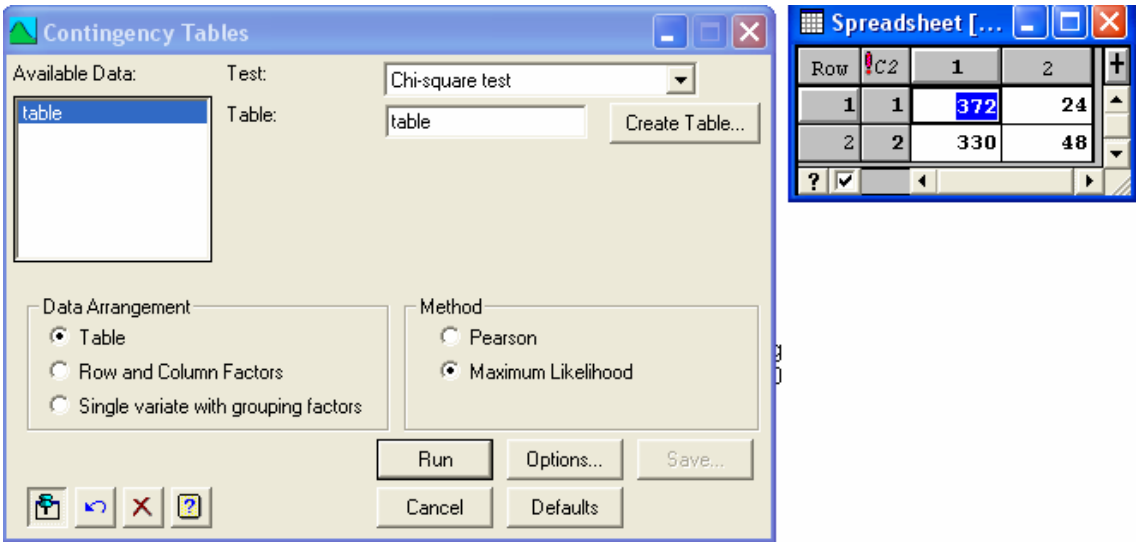
For these data the test values are virtually identical:

Chi-square test for association between C2 and C3

Pearson chi-square value is 10.10 with 1 d.f.
Probability level (under null hypothesis) p = 0.001

Chi-square test for association between C2 and C3

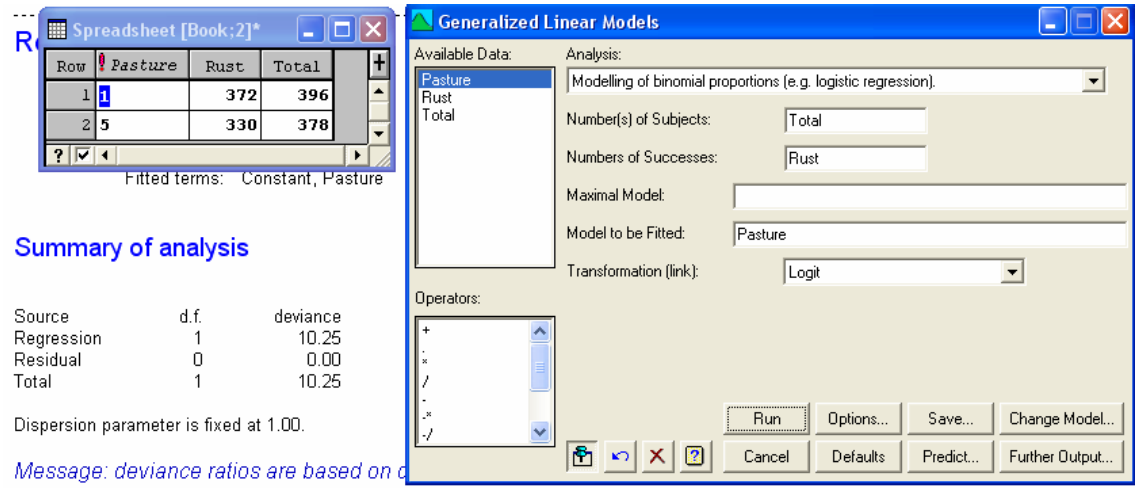
Likelihood chi-square value is 10.25 with 1 d.f.
Probability level (under null hypothesis) p = 0.001



Clearly, there is strong evidence (P=0.001) that individual probability estimates are required for the two pasture types. We would use

Pasture type 1: p = 372/396 = 0.939
Pasture type 2: p = 330/378 = 0.873

Now let us do this in GenStat's **Regression** > **Logistic Regression** menu. You need a factor column to identify the two pasture types, a column of rust numbers and a column of totals.



Regression analysis

Response variate: Rust
Binomial totals: Total
Distribution: Binomial
Link function: Logit
Fitted terms: Constant, Pasture

Summary of analysis

Source	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
Regression	**1**	**10.25**	**10.25**	**10.25**	**0.001**
Residual	0	0.00	*		
Total	1	10.25	10.25		

Dispersion parameter is fixed at 1.00.

Message: deviance ratios are based on dispersion parameter with value 1.

Estimates of parameters

Parameter	estimate	s.e.	t(*)	t pr.	antilog of estimate
Constant	2.741	0.211	13.01	<.001	15.50
Pasture 5	-0.813	0.261	-3.11	0.002	0.4435

Message: s.e.s are based on dispersion parameter with value 1.

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Pasture	1

This is very similar to a regression output with a factor. Notice:

- The Regression Deviance is 10.25, identical to the ML contingency table X^2 statistic.

- The linear predictor on the logit scale is (2.741 – 0.813 X), where X is 0 for pasture type 1, and 1 for pasture type 5. That is, the constant identifies the model for pasture type 1. The model for pasture type 5 is obtained by adding 2.741 and -0.813 = 1.928.

  The *odds* for the two pasture types are $e^{2.741}$ = 15.50 (which GenStat produces as the antilog of estimate), and $e^{2.741-0.813}$ = 6.876. This is also available by *multiplying* the two antilogs: 15.50×0.4435 = 6.874 (round-off).

  Once the odds are available, the probabilities can be calculated as odds/(1+odds). We obtain 15.50/(1+15.50) = 0.939 for pasture type 1, and 6.876/(1+6.876) = 0.873.

  These are what we calculated following the contingency table test.

- If you **Save** the fitted values, in this case you obtain the actual data.

  Thus, if a factor is significant in a model, the fitted values are identical to the actual ***totals*** *for that factor.*

Linear regression in logistic regression

Example 2 Age of parent tree, and reaction of grafts to blister dust, from Steel and Torrie, page 515

Age of parent tree (years)	Healthy	Diseased	Total
4	7	14	21
10	6	11	17
20	11	5	16
40+	15	8	23
Total	39	38	77

The individual estimated probabilities are 0.333, 0.353, 0.688, 0.652. Steel and Torrie used coded years of 1, 2, 4, 8 to partition the Pearson χ^2 (8.17) into a linear (5.58, 1 *df*) and a non-linear (2.58, 2 *df*) term. We will use the years as shown, and change in deviance.

To do this, we need to enter a factor column for age, a column of healthy counts and a column of total counts.

To test whether a linear relationship exits on the logit scale in age, we either run a second analysis with Age converted to a variate for the second analysis, or provide a new variate which we will call AgeV.

Firstly, we could treat this as a 4×2 contingency table. GenStat gives a test value of 8.33 on 3 df for this, with a *P* value of 0.040.

Treating this as a logistic regression, we obtain the same test value:

Summary of analysis

Source	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
Regression	3	**8.329**	2.776	2.78	0.040
Residual	0	0.000	*		
Total	3	8.329	2.776		

Treating Age as an explanatory variate instead of a factor gives:

Summary of analysis

Source	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
Regression	1	5.758	5.758	5.76	0.016
Residual	2	2.570	1.285		
Total	3	8.329	2.776		

The model deviance is 5.758 which tests whether a linear trend is present or not. Clearly there is evidence ($P = 0.016$) that there is a relationship between the probability of being healthy, and age of the parent tree.

The change in deviance for the two models is 2.570 on 2 df. This is not significant, which suggests that the linear model is an adequate explanation of the change in probability of being healthy.

Poisson Regression

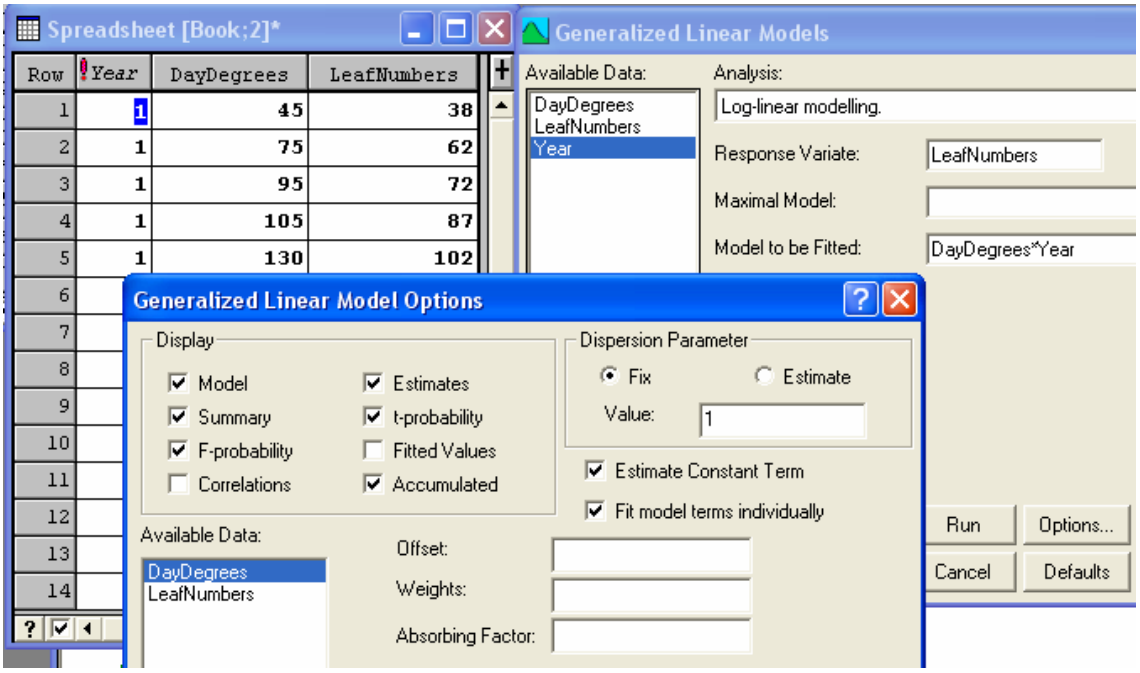
We can re-visit the analysis of mean leaf numbers from 10 cauliflower plants over two years for this section (see page 21). Multiplying by 10 to obtain total leaf numbers gives

1956/7 season		1957/8 season	
DD	Number	DD	Number
45	38	45	60
75	62	80	85
95	72	95	91
105	87	115	120
130	102	130	126
160	135	140	133
180	150	165	152

For a Poisson distribution, the sum of independent Poisson variates is Poisson. We therefore expect a Poisson distribution for the data. It is true that a Poisson distribution tends to a normal, but we can use the exact distribution for numbers like these, and use a Poisson regression analysis.

We need a factor column to identify year, a column of total leaf numbers and a column of day degrees, the predictor.

We are interested in whether a single model fits the data, or parallel or separate models over the two years. The **Model to be Fitted** is therefore **DayDegrees*Years**. To evaluate the effect of each term as they are entered into the model, click **Accumulated** > **Fit model terms individually** in **Options**.



Regression analysis

Response variate: LeafNumbers
Distribution: Poisson
Link function: Log
Fitted terms: Constant + DayDegrees + Year + DayDegrees.Year

Summary of analysis

Source	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
Regression	3	170.693	56.8976	56.90	<.001
Residual	**10**	**4.596**	**0.4596**		
Total	13	175.289	13.4838		
Change	-1	-1.463	1.4630	1.46	0.226

Dispersion parameter is fixed at 1.00.

Message: deviance ratios are based on dispersion parameter with value 1.

Message: the following units have high leverage.

Unit	Response	Leverage
7	150.00	0.58

Estimates of parameters

Parameter	estimate	s.e.	t(*)	t pr.	antilog of estimate
Constant	3.398	0.127	26.71	<.001	29.91
DayDegrees	0.009259	0.000928	9.98	<.001	1.009
Year 2	0.426	0.181	2.35	0.019	1.531
DayDegrees.Year 2	-0.00168	0.00138	-1.21	0.226	0.9983

Message: s.e.s are based on dispersion parameter with value 1.

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Year	1

Accumulated analysis of deviance

Change	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
+ DayDegrees	1	152.8308	152.8308	152.83	<.001
+ Year	1	16.3991	16.3991	16.40	<.001
+ DayDegrees.Year	1	1.4630	1.4630	1.46	0.226
Residual	10	4.5963	0.4596		
Total	13	175.2892	13.4838		

Message: ratios are based on dispersion parameter with value 1

Notice the following.

- If we have Poisson data, the "dispersion parameter", which is the mean deviance for the residual term, that is, residual deviance/residual d.f., should be 1. In this case it is under-dispersed, with a of dispersion parameter 0.4596. Is this a problem? What we do is test whether the deviance of 4.596 is likely to have come by chance from a χ^2 distribution with (in this case) 10 *df*. A lower critical probability is 0.0835. Before we do the

experiment, there is no reason why the deviance will be greater than or less than what is expected, by chance. Hence the *P* value we would quote for this is 0.167. We would not reject a hypothesis that the data are Poisson.

- The test of equal slopes is not significant ($P = 0.226$). Both Year and DayDegrees are significant predictors, so the conclusion is the same as previously: parallel regressions are required, one for each year. In the present analysis, however, linearity is on the log-scale. We therefore re-run the analysis with a model DayDegrees+Year:

Regression analysis

Response variate: Number
Distribution: Poisson
Link function: Log
Fitted terms: Constant + DayDegrees + Year

Summary of analysis

Source	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
Regression	2	169.230	84.6149	84.61	<.001
Residual	11	6.059	0.5508		
Total	13	175.289	13.4838		

Dispersion parameter is fixed at 1.00.

Estimates of parameters

Parameter	estimate	s.e.	t(*)	t pr.	antilog of estimate
Constant	3.4949	0.0972	35.95	<.001	32.95
DayDegrees	0.08513	0.00689	12.35	<.001	1.089
Year 2	0.2169	0.0537	4.04	<.001	1.242

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Year	1

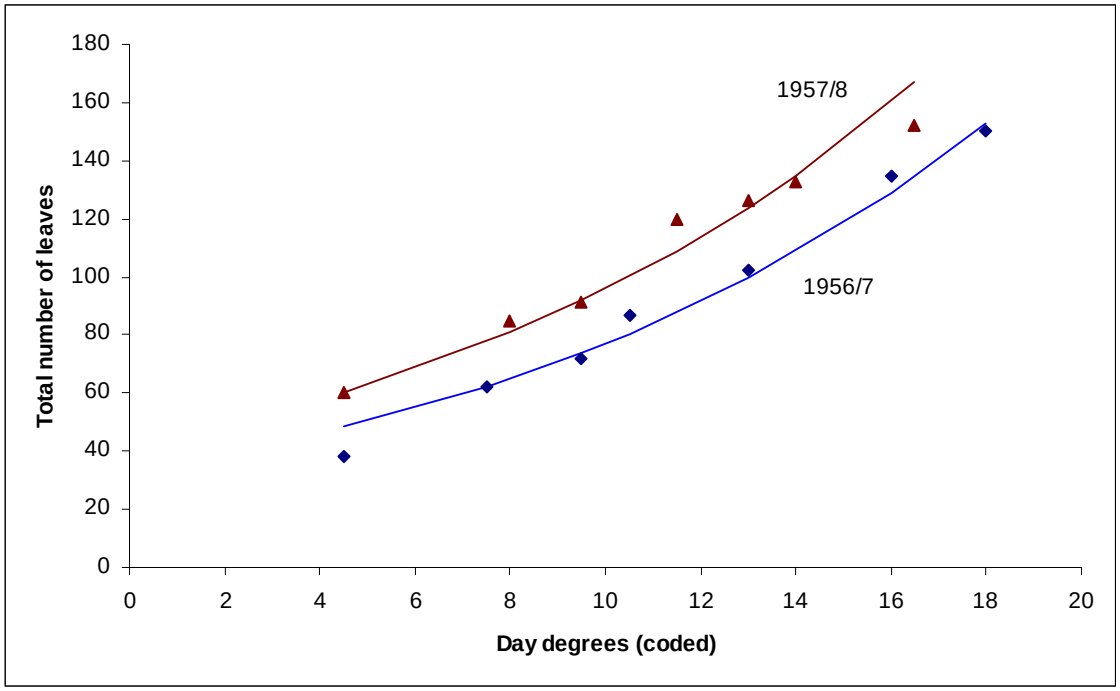
For Year = 1 (1956/7) the model is

Total number of leaves from 10 plants = $e^{3.4949 + 0.08513\ Day\ Degrees} = 32.95 \times 1.089^{Day\ Degrees}$

For Year = 2 (1957/8) the model is

Total number of leaves from 10 plants $= e^{3.4949 + 0.2169 + 0.08513\ Day\ Degrees}$
$= e^{3.7188 + 0.08513\ Day\ Degrees}\ 32.95 \times 1.242 \times 1.089^{Day\ Degrees}$
$= 40.924 \times 1.089^{Day\ Degrees}$

These models are plotted on the following page.



Note. A linear model on the log-scale makes some sense in that the predicted number of leaves can never be negative. However, with data like these, we can still use a linear predictor (as we did in the regression section) on the *count* scale assuming Poisson data. Simply choose **Regression Analysis** > **Generalized Linear Models** and select **Identity** for the **Link Function**. The Residual Deviance appears under-dispersed, so in the following output we chose **Estimate** rather than **Fix**. The final models are very similar to those from regression on page 20. We cannot compare models with different links, except to say that the smallest deviance the better (a zero deviance indicates a perfect fit). A model with a deviance of 0.184 is clearly superior to one with a deviance of 6.059 (assuming a log-link).

Regression analysis

Response variate: Number
Distribution: Poisson
Link function: Identity
Fitted terms: Constant, DayDegrees, Year

…

Estimates of parameters

Parameter	estimate	s.e.	t(11)	t pr.
Constant	0.096	0.268	0.36	0.728
DayDegrees	0.8068	0.0247	32.63	<.001
Year 2	2.015	0.206	9.80	<.001

Accumulated analysis of deviance

Change	d.f.	deviance	mean deviance	deviance ratio	approx F pr.
+ DayDegrees	1	15.72952	15.72952	939.78	<.001
+ Year	1	1.61529	1.61529	96.51	<.001
Residual	11	0.18411	0.01674		
Total	13	17.52892	1.34838		

Dispersion parameter is estimated to be 0.0167 from the residual deviance.

Log-linear modelling

This general analysis is used to analyse more complex contingency tables. It turns out that binomial data can be treated as Poisson data *conditional on the totals being fixed*. Thus, provided we use terms in the model to fix the totals, we should obtain the same analysis using log-linear modelling as we do from logistic regression. Log-linear modelling, of course, is more general - one can have any numbers of outcomes, not just two (success/failure).

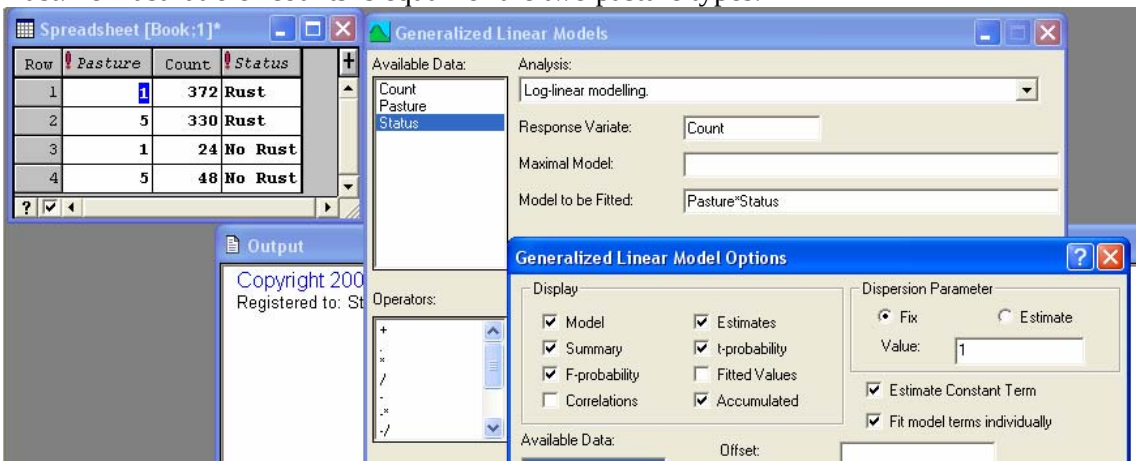
To start, consider the incidence of rust in Kentucky bluegrass pastures again. This time we stack the successes and failures, and provide a factor column to identify each. Thus, the information in the table will need to be prepared as shown:

Pasture field type	Rust	No Rust
1	372	24
5	330	48

Pasture	Count	Status
1	372	Rust
5	330	Rust
1	24	No Rust
5	48	No Rust

The **Model to be Fitted** is Pasture*Status.

We need to keep the pasture totals fixed, so Pasture must be present in the model simply to fix these. Status alone tests whether the counts are equal, and is of no interest. The only factor of interest in this experiment is the apparent interaction Pasture.Status. It assesses whether the Rust:No Rust ratio of counts is equal for the two pasture types.



Regression analysis

Response variate: Count
Distribution: Poisson
Link function: Log
Fitted terms: Constant + Pasture + Status + Pasture.Status

Summary of analysis

Source	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
Regression	3	604.6	201.5	201.53	<.001
Residual	0	0.0	*		
Total	3	604.6	201.5		
Change	-1	-10.3	10.3	10.25	0.001

Dispersion parameter is fixed at 1.00.

Message: deviance ratios are based on dispersion parameter with value 1.

Estimates of parameters

Parameter	estimate	s.e.	t(*)	t pr.	antilog of estimate
Constant	3.178	0.204	15.57	<.001	24.00
Pasture 5	0.693	0.250	2.77	0.006	2.000
Status Rust	2.741	0.211	13.02	<.001	15.50
Pasture 5 .Status Rust	-0.813	0.261	-3.11	0.002	0.4435

Message: s.e.s are based on dispersion parameter with value 1.

Parameters for factors are differences compared with the reference level:

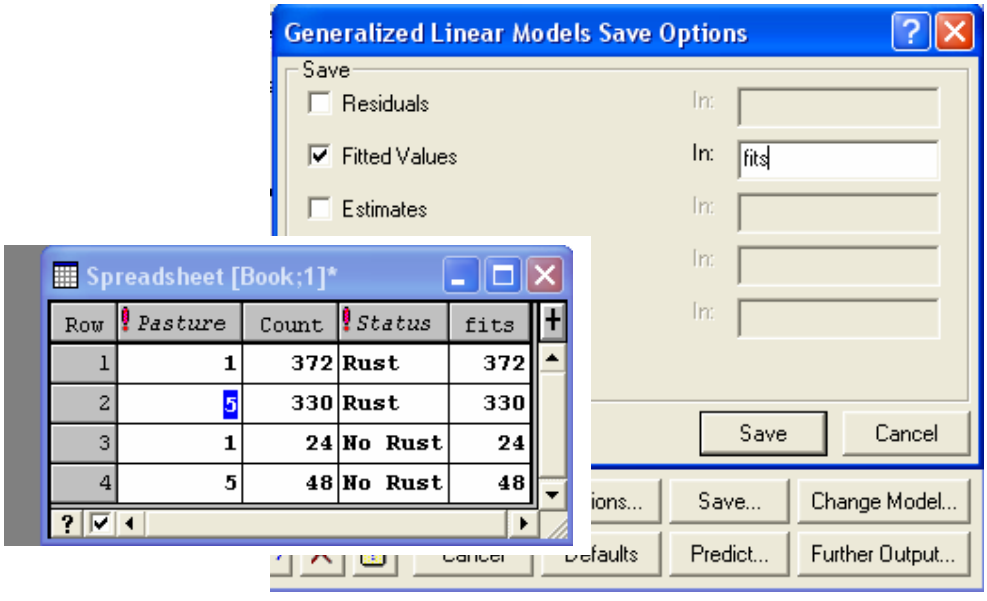
Factor	Reference level
Pasture	1
Status	No Rust

Accumulated analysis of deviance

Change	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
+ Pasture	1	0.42	0.42	0.42	0.518
+ Status	1	593.92	593.92	593.92	<.001
Residual	1	10.25	10.25		
+ Pasture.Status	**1**	**10.25**	**10.25**	**10.25**	**0.001**
Total	3	604.59	201.53		

Notice:

- The Residual deviance is 0 and has 0 *df*. This is because there are only 4 cells in the table, hence 3 *df;* and we are modelling the data with 3 one-*df* terms. This is known as a saturated model. A saturated model must reproduce the data as Fitted Values:



- The Pasture.Status component (10.25) is identical to that obtained using logistic regression.

Estimates of parameters

Parameter	estimate	s.e.	t(*)	t pr.	antilog of estimate
Constant	3.178	0.204	15.57	<.001	24.00
Pasture 5	0.693	0.250	2.77	0.006	2.000
Status Rust	2.741	0.211	13.02	<.001	15.50
Pasture 5 .Status Rust	-0.813	0.261	-3.11	0.002	0.4435

- The fitted model is referenced to pasture type 1, no rust. The saturated model gives an antilog of 24, namely the actual count for that combination. Pasture type 5 then has a fitted count of 2×24 = 48, again the actual count. The fitted count for pasture type 1, rust is 24×15.5 = 372, again the actual count.

  This is another illustration of the rule that the presence of any main effect or interaction in a generalized linear model induces the fitted counts to be identical to the observed counts for the table concerned.

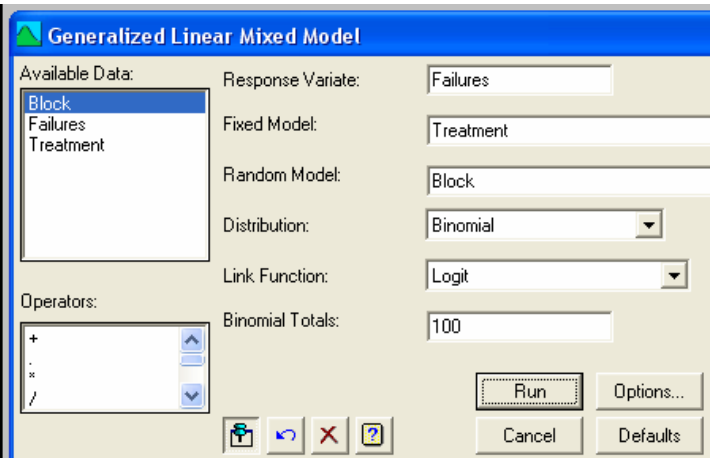
Generalized Linear Mixed Model

Example 3 The number of soybean plants that failed to emerge (each out of 100 plants) using seeds that had one of four treatments or no treatment, from Snedecor and Cochran, page 256.

Treatment	Block 1	Block 2	Block 3	Block 4	Block 5
Control	8	10	12	13	11
Arasan	2	6	7	11	5
Spergon	4	10	9	8	10
Semesan, Jr.	3	5	9	10	6
Fermate	9	7	5	5	3

If these were normal data the analysis would be a standard RCBD – as was done in the monograph. However, we have binomial counts in a block design. Blocks are usually regarded as random. Hence we need to use a GLMM.

Row	Block	Treatment	Failures
1	1	Control	8
2	1	Arasan	2
3	1	Spergon	4
4	1	Semesan, Jr.	3
5	1	Fermate	9
6	2	Control	10
7	2	Arasan	6
8	2	Spergon	10
9	2	Semesan, Jr.	5
10	2	Fermate	7
11	3	Control	12
12	3	Arasan	7



We had no reason to suspect that the distribution would be over- or under-dispersed so fixed the dispersion index at 1. (We allowed this index to be estimated and obtained 0.932 ± 0.331, which lends no weight against our supposition.)

Generalized linear mixed model analysis

Method:	c.f. Schall (1991) Biometrika
Response variate:	Failures
Binomial totals:	100
Distribution:	binomial
Link function:	logit
Random model:	Block
Fixed model:	Constant + Treatment

Dispersion parameter fixed at value 1.000

Estimated variance components

Random term	component	s.e.
Block	**0.024**	0.038

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
	Dispersn	Identity	Sigma2	**1.000**	fixed

Estimated variance matrix for variance components

Block	1	0.0014346		
Dispersn	2	0.0000000	0.0000000	
		1	2	

Wald tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	d.f.	Wald/d.f.	chi pr
Treatment	11.81	4	2.95	0.019

Tables of means

Table of predicted means for Treatment

Treatment	
Arasan	-2.720
Control	-2.115
Fermate	-2.791
Semesan, Jr.	-2.654
Spergon	-2.419

Standard errors of differences between pairs

Treatment Arasan	1	*				
Treatment Control	2	0.235	*			
Treatment Fermate	3	0.267	0.240	*		
Treatment Semesan, Jr.	4	0.259	0.231	0.263	*	
Treatment Spergon	5	0.247	0.218	0.251	0.243	*
		1	2	3	4	5

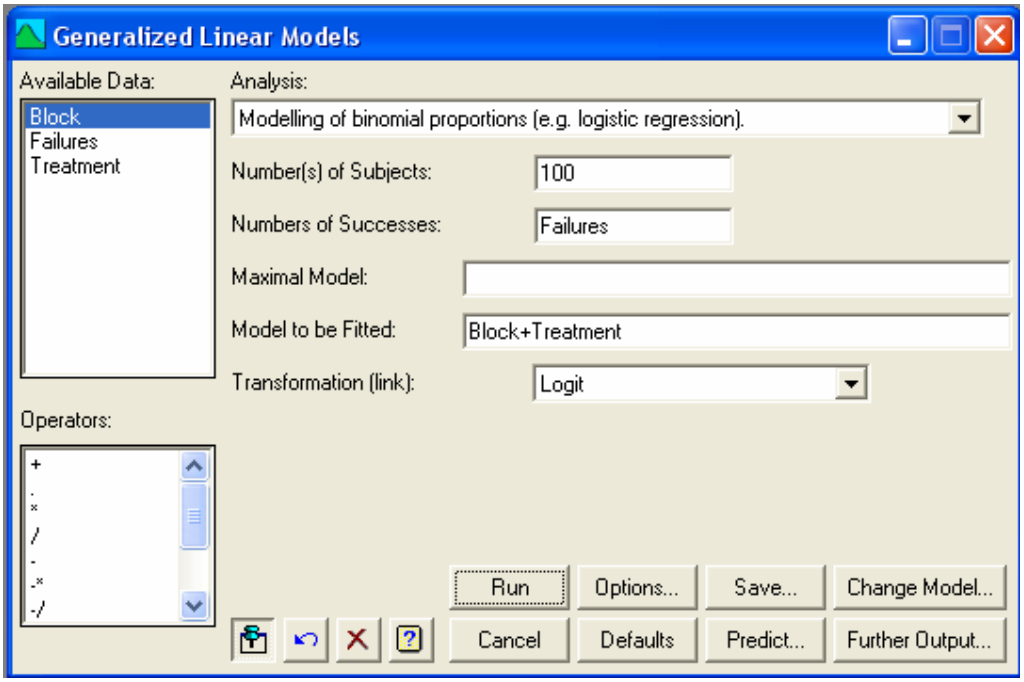
Back-transformed Means (on the original scale)

Treatment	
Arasan	6.178
Control	10.767
Fermate	5.779
Semesan, Jr.	6.577
Spergon	8.173

There is evidence that the probability of failure differs across treatments. To estimate the probabilities, divide the back-transformed means by $n = 100$. We obtain: 0.10767 (Control), 0.06178 (Arasan), 0.05779 (Fermate), 0.06577 (Semesan, Jr.) and 0.08173 (Spergon).

At the stage of writing, the model could not be printed, nor could we obtain s.e.d values on the back-transformed scale.

The data could also be analysed via a basic generalized linear model if we are prepared to assume blocks are fixed. We turned on Accumulated and Fit model terms individually to allow the contribution from blocks to be measured as well as the effect of treatments.



Regression analysis

Response variate: Failures
Binomial totals: 100
Distribution: Binomial
Link function: Logit
Fitted terms: Constant, Block, Treatment

A similar dispersion parameter as was found when estimating using a GLMM. A deviance of 15.01 can be tested using a χ^2 distribution with 16 df.

Summary of analysis

Source	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
Regression	8	18.91	2.3639	2.36	0.015
Residual	**16**	**15.01**	**0.9380**		
Total	24	33.92	1.4133		
Change	-4	-11.50	2.8741	2.87	0.022

Dispersion parameter is fixed at 1.00.

Message: deviance ratios are based on dispersion parameter with value 1.

Estimates of parameters

Parameter	estimate	s.e.	t(*)	t pr.	antilog of estimate
Constant	-3.113	0.263	-11.82	<.001	0.04448
Block 2	0.407	0.263	1.55	0.122	1.502
Block 3	0.516	0.258	2.00	0.046	1.676
Block 4	0.640	0.253	2.53	0.011	1.897
Block 5	0.318	0.267	1.19	0.234	1.374
Treatment Control	**0.607**	**0.235**	**2.58**	**0.010**	**1.835**
Treatment Fermate	-0.071	0.266	-0.27	0.790	0.9314
Treatment Semesan, Jr.	0.067	0.259	0.26	0.796	1.069
Treatment Spergon	0.302	0.247	1.22	0.222	1.353

Message: s.e.s are based on dispersion parameter with value 1.

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Block	1
Treatment	Arasan

Accumulated analysis of deviance

Change	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
+ Block	4	7.4145	1.8536	1.85	0.116
+ Treatment	**4**	**11.4964**	**2.8741**	**2.87**	**0.022**
Residual	16	15.0087	0.9380		
Total	24	33.9196	1.4133		

The GLMM gave a Wald/d.f. statistic of 2.95, consistent with the above. The model was referenced to Arasan, and we can use the P value of 0.010(from the **Treatment Control line in the** Estimates of parameters part of the analysis)to conclude that the probability of failure for untreated seeds is different to that for Arasan-treated seeds. Furthermore, no other seed treatment was significant when compared to Arasan.

Ordinal logistic regression

Occasionally scientists will only be able to score plants or plots and special analyses need to be used for score data. In this section, we introduce ordinal logistic regression for ordered scores.

We will take the data from Snedecor and Cochran page 205. Their data involves ordered scores for improvement in health of leprosy sufferers. They analysed the data as a *t* test. We will use the same data, but imagine them to come from the following plant pathology experiment. Suppose *Sclerotinia sclerotiorum* was tested as a biological control of the noxious weed bitoubush (*Chrysanthemoides monilifera* ssp. *Rotundata*). Two isolates were assessed for pathogenicity, and varying numbers of plants were assessed per isolate. We will use the following 5-point ordered scale.

1 = no reaction
2 = lesions confined to <20% of leaves
3 = lesions confined to 20% to 50% of leaves
4 = lesions confined to 50% to 70% of leaves
5 = lesions confined to >70% of leaves

The data are the same as in Snedecor and Cochran. Here we present it in two ways. Firstly, we have 144 random plants with isolate 1 with varying scores, followed by 52 random plants.

Isolate 1	2	3	2	2	3	4	2	2	5	2	2	4	2	1	3	2	2	3	2	4
	2	3	2	3	4	4	2	3	3	2	3	3	5	2	4	1	2	3	4	2
	3	3	2	2	3	2	5	2	3	3	3	1	4	2	4	2	3	1	3	4
	3	1	5	4	3	2	3	5	3	3	1	4	3	3	1	2	3	4	2	3
	4	2	2	4	3	2	2	2	4	2	2	2	4	4	2	4	2	5	2	3
	1	2	4	4	3	5	2	2	2	5	3	2	2	1	4	2	3	3	3	4
	3	4	2	2	2	3	3	5	2	3	4	5	3	2	5	4	1	2	2	3
	2	4	3	1																

Isolate 2	2	3	3	2	4	3	5	5	4	2	2	3	3	2	3	4	4	4	2	5
	3	5	2	3	2	5	2	4	3	5	4	4	4	1	3	3	2	4	2	4
	2	5	4	3	4	3	3	4	4	3	3	2								

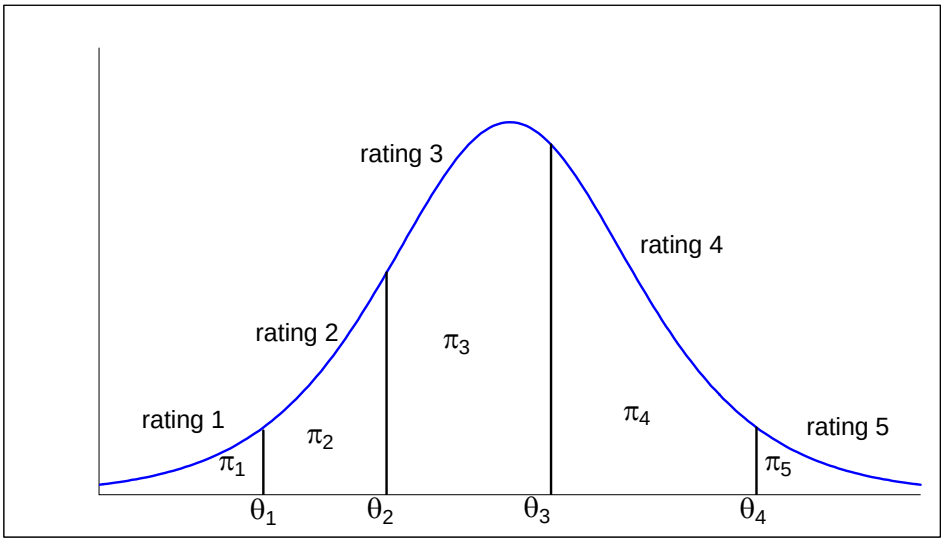
GenStat would need the scores in a single ***factor*** column as well as a factor column to identify the isolate for each plant.

Alternatively, we could supply the data in frequency form. The scores would need to be in five ***variate*** columns, and we would need a factor to identify the isolates for each row of frequencies.

Isolate	Score1	Score2	Score3	Score4	Score5
1	11	53	42	27	11
2	1	13	16	15	7

Now plants do not suddenly jump in discrete steps from one score to the next. Rather, there is a continuous change in the severity of damage of the plant. A severity score of 1 represents undamaged plants, although damage may be slowly taking place, perhaps unseen. These plants have a score while the damage is confined to a point we call the first cut-point, θ_1. A score of 2 then represents plants with damage from θ_1 to some new cut-point θ_2.

So, generally some underlying continuous distribution is assumed, such as logistic. For the 5-point rating scale under discussion, this would appear as follows, assuming an underlying logistic distribution.



We do not say that the scores necessarily represent equal spacings on this continuous scale. To quantify the discussion to date, suppose we use Y for the continuous damage variable. Then a rating of 1 represents plants whose damage value on the continuous scale is any $Y < \theta_1$. The underlying probability of obtaining a plant with this rating is $\pi_1 = P(Y < \theta_1)$. We do not know θ_1 and we don't know π_1.

Similarly, a rating 2 represents a plant whose damage value on the continuous scale is anything between θ_1 and θ_2. The underlying probability of obtaining a plant with this rating is $\pi_2 = P(\theta_1 < Y < \theta_2)$. We don't know θ_2 and we don't know π_2. And so on.

It is actually simpler to model the *cumulative* probabilities $P(Y < \theta_1) = \pi_1$, $P(Y < \theta_2) = \pi_1+\pi_2$, $P(Y < \theta_3) = \pi_1+\pi_2+\pi_3$,…. For notation we will define

$\gamma_1 = P(Y < \theta_1) = \pi_1$,
$\gamma_2 = P(Y < \theta_2) = \pi_1+\pi_2$,
$\gamma_3 = P(Y < \theta_3) = \pi_1+\pi_2+\pi_3$ and so on. (The last value must be 1.)

Thus, for a 5-point scale we need to estimate four cut-points θ_1 to θ_4 and four probabilities π_1 to π_4 (since the 5th rating is a value larger than θ_4 and π_5 is 1- π_1- π_2- π_3- π_4) and hence γ_1 to γ_4.

Now we propose a set of logistic regression equations for the cumulative probabilities:

$$\gamma_i = \frac{1}{1+e^{-(\theta_i - b_1 X_1 - \ldots)}}$$

where $\{\theta_i\}$ are the cut-points for the ordered scale and X_1, …, are covariates, or, as in the case of a designed experiment, the usual design features. On the logit scale this becomes

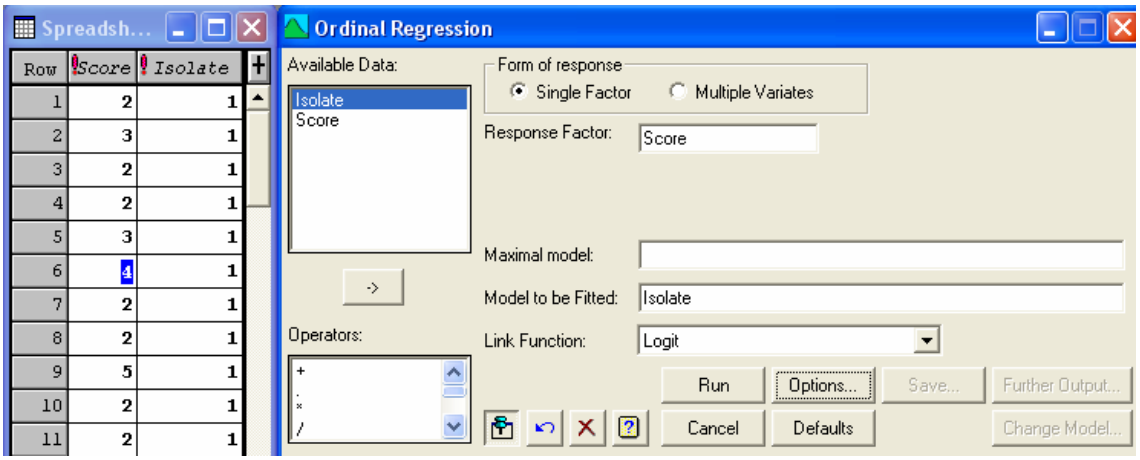
$$\log\left(\frac{\gamma_i}{1-\gamma_i}\right) = \theta_i - b_1 X_1 - \ldots \qquad \text{for ratings } i = 1, 2, \ldots$$

Note that this is sometimes referred to as the *proportional odds model*, because, for a given state (score), the ratio of the odds does not depend on the state. Notice that the *log-odds* value, and hence the odds ratio, are calculated on the *cumulative* scale (ie using the γ_i, not the π_i). By difference, once we have estimated the cumulative probabilities we can calculate the individual probabilities.

The parameters are estimated by maximum likelihood as with ordinary logistic regression.

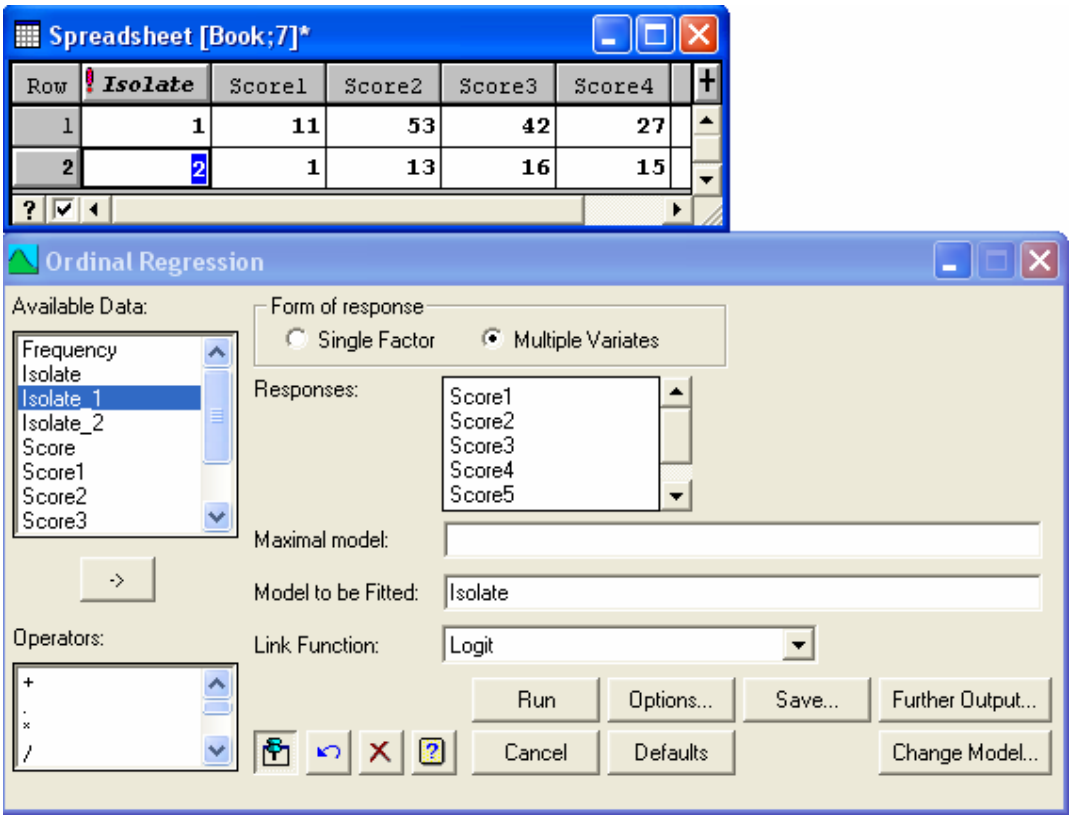
Choose **Stats** > **Regression Analysis** > **Ordinal Regression**. Then use either one of the following methods.

Method 1 Individual plant scores (as a factor) with a treatment factor column:



The same analysis is obtained by the following method.

Method 2 Individual score variates of frequencies with a treatment factor column:



Response variates: ordinal model for categories defined by Score1, Score2, Score3, Score4, Score5
Distribution: Multinomial
Link function: Logit
Fitted terms: Isolate

Summary of analysis

Source	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
Regression	1	6.7	6.679	6.68	0.010
Residual	191	560.6	2.935		
Total	192	567.3	2.955		

Dispersion parameter is fixed at 1.00.

Estimates of parameters

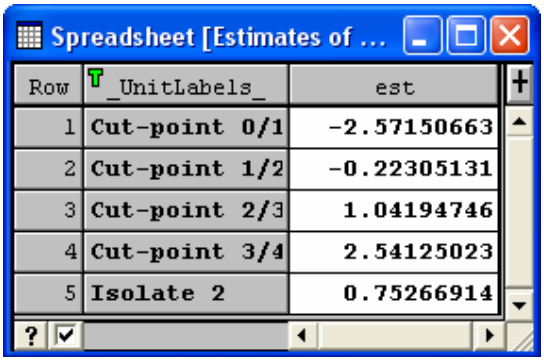
Parameter	estimate	s.e.	t(*)	t pr.	antilog of estimate
Cut-point 0/1	-2.572	0.303	-8.48	<.001	0.07642
Cut-point 1/2	-0.223	0.164	-1.36	0.173	0.8001
Cut-point 2/3	1.042	0.180	5.78	<.001	2.835
Cut-point 3/4	2.541	0.270	9.41	<.001	12.70
Isolate 2	0.753	0.295	2.55	0.011	2.123

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Isolate	1

The two isolates have significantly different (P=0.010) probability distributions of the five scores.

Before estimating the individual probability distributions for the two isolates, it is wise to save the estimates so they can be opened in Excel with full accuracy.


Spreadsheet [Estimates of ...

Row	_UnitLabels_	est
1	Cut-point 0/1	-2.57150663
2	Cut-point 1/2	-0.22305131
3	Cut-point 2/3	1.04194746
4	Cut-point 3/4	2.54125023
5	Isolate 2	0.75266914

We can interpret the model as follows.

The model is referenced to isolate 1. With a treatment factor with only two levels (isolate 1 and 2) we have only one predictor in the model. Hence X_1 = 1 for isolate 2 and 0 otherwise.

For isolate 1

$$\log\left(\frac{\gamma_i}{1-\gamma_i}\right) = \theta_i \text{ for } i = 1, 2, 3 \text{ and } 4.$$

The back-transform is given in the output as the antilog of estimate. Thus, the odds for a score of 1 are 0.076420. Hence the estimate of the probability γ_1 is 0.076420/(1+0.076420) = 0.070995. For this cut-point, γ_1 and π_1 are the same.

The odds for a score of 1 or 2 are 0.800074. Hence the estimate of the probability γ_2 is 0.800074/(1+0.800074) = 0.444467. By subtraction, the estimate for π_2 is 0.444467-0.070995 = 0.373472. And so on.

For isolate 2

$$\log\left(\frac{\gamma_i}{1-\gamma_i}\right) = \theta_i + 0.752669 \text{ for } i = 1, 2, 3 \text{ and } 4.$$

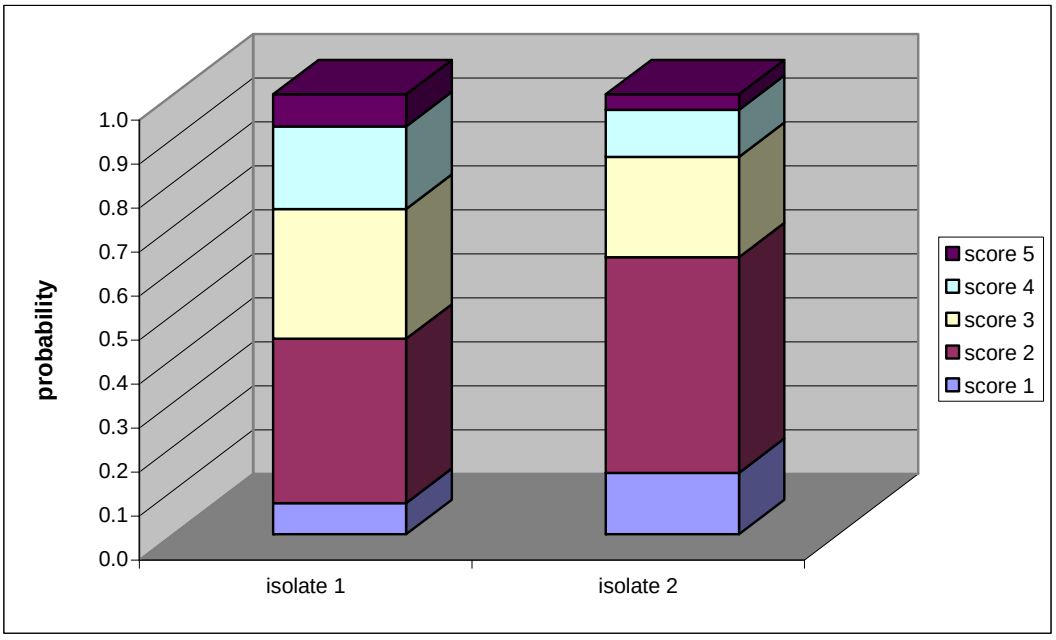
This means that the odds already worked out for the reference isolate simply need to be multiplied by $e^{0.752669}$, the antilog of b_1 being 2.122658. Thus:

The odds for a score of 1 are 0.076420×2.122658 = 0.162214. Hence the estimate of the probability γ_1 is 0.162214/(1+0.162214) = 0.139537. For this cut-point, γ_1 and π_1 are the same. And so on.

This is very easy to do in Excel. Here the cells used are marked (starting from V3), and formulae for the two calculation columns shown alongside.

	V	W	X	Y	X	Y
13			antilog	antilog	antilog	antilog
14			isolate 1	isolate 2	isolate 1	isolate 2
15			**Odds**	2.122658	**Odds**	=EXP(W20)
16	Cut-point 0/1	-2.571507	0.076420	0.162214	=EXP(W16)	=X16*Y$15
17	Cut-point 1/2	-0.223051	0.800074	1.698283	=EXP(W17)	=X17*Y$15
18	Cut-point 2/3	1.041947	2.834732	6.017167	=EXP(W18)	=X18*Y$15
19	Cut-point 3/4	2.541250	12.695533	26.948277	=EXP(W19)	=X19*Y$15
20	isolate 2	0.752669				
21						
22		**score**	**gammas**		**gammas**	
23		1	0.070995	0.139573	=X16/(1+X16)	=Y16/(1+Y16)
24		2	0.444467	0.629394	=X17/(1+X17)	=Y17/(1+Y17)
25		3	0.739226	0.857492	=X18/(1+X18)	=Y18/(1+Y18)
26		4	0.926983	0.964220	=X19/(1+X19)	=Y19/(1+Y19)
27		5	1	1	1	1
28		**score**	**probabilities**		**probabilities**	
29		1	0.070995	0.139573	=X23	=Y23
30		2	0.373472	0.489821	=X24-X23	=Y24-Y23
31		3	0.294758	0.228098	=X25-X24	=Y25-Y24
32		4	0.187758	0.106727	=X26-X25	=Y26-Y25
33		5	0.073017	0.035780	=X27-X26	=Y27-Y26

Excel has very good plotting techniques to illustrate the difference in the estimated probability distributions in cells X29:Y33:



Choice of sample size

While this topic has been left to the end of the manual, it is arguably the most important topic. In our experience, failure to think through the issues has resulted in many wasted experiments.

Why is sample size important?

Let us take a normal population by way of demonstration. This distribution involves the mean parameter μ which we try and estimate as accurately as we can. When we conduct an experiment involving several treatments, we do so in order to estimate the difference between treatment means as accurately as we can. That might entail blocking to control as much variation as we can, so that the estimate of the ancillary parameter σ^2 is as efficient as possible. If we fail to think through how many replicates we need to obtain a difference of the order of magnitude we have among our treatments, we may end up with an experiment that fails to find any differences at all.

We know that the most efficient estimate of μ for normally distributed data is the sample mean, $\bar{y}$, and that this estimate has a standard error equal to $\sigma/\sqrt{n}$. So you can see that increasing the number of replicates (*n*) is the way we obtain a small standard error, and hence a precise estimate of μ. This pre-supposes that we also have a reliable estimate of σ. Of course, practical or cost considerations may force a compromise along the way.

So we start with a two sample *t* test to consolidate basic concepts.

Two sample *t* test

Let $\bar{y}_1$ and $\bar{y}_2$ represent the sample means from two normal populations (corresponding to two treatments). The distribution of the difference between the two means is also normal. For simplicity we assume equal replication and equal variance. Then

$$\bar{y}_2 - \bar{y}_1 \sim N\left(\mu_2 - \mu_1, \frac{2\sigma^2}{n}\right)$$

There variation in $\bar{y}_2$ - $\bar{y}_1$ is measured by the standard deviation of the difference of the two means, s.e.d. given by

$$s.e.d. = \sqrt{\frac{2\sigma^2}{n}}$$

so in general this also needs to be estimated.

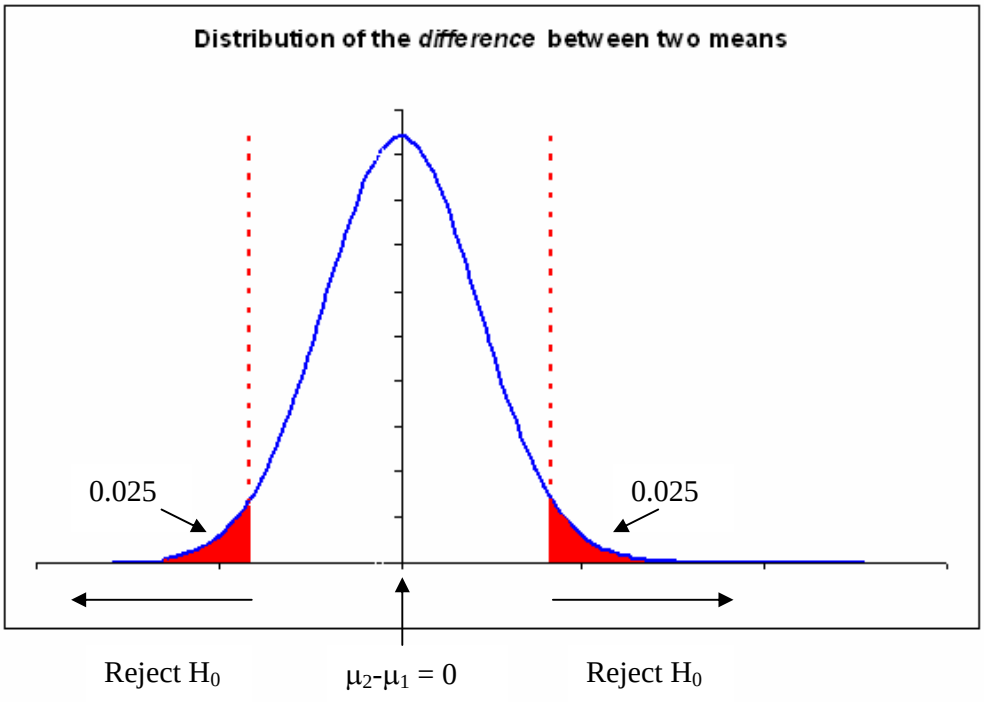
The first decision to make is at what level do you want to set α, the level of significance? There is nothing magical about α: pilot trials are often more relaxed than other experiments, and maybe α is set to 0.1 (10%). If you really need to be sure when you reject a null hypothesis H_0: $\mu_2 - \mu_1 = 0$ you might set α to be 0.001 (0.1%).

We will take a 5% level of significance, which means that the probability we reject that the mean difference is zero when it really is zero is 0.05 (or a one in twenty chance).

We will assume a two sided alternative, H_0: $\mu_2 - \mu_1 \neq 0$.

This leads to two "criteria" or "test values". These are found by allocating half of α to each tail of the distribution of the difference between two means. Here each red area in the tails corresponds to 2½%, 0.025. A test statistic outside the dotted red lines will lead us to reject that the means are equal. On average, 1 in 20 of all such life-time experimental decisions will be incorrect. That's the risk an experimenter is prepared to take in order to make a decision one way or another when testing the null hypothesis.

Diagram of the *level of significance* for a two sample *t* test (in red).



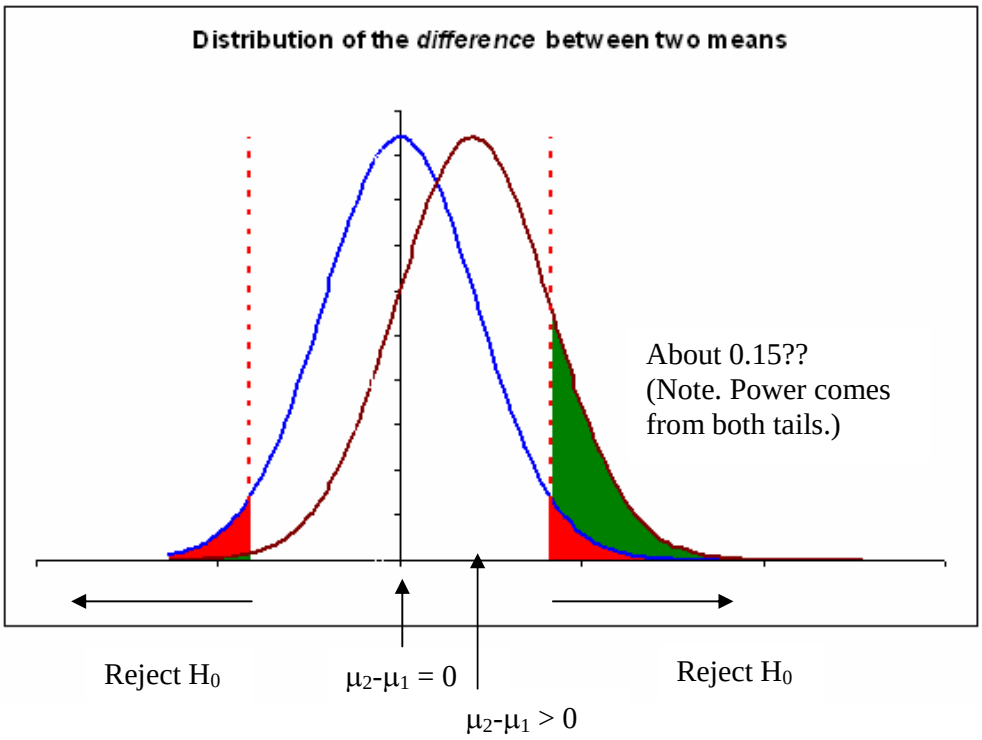
So rejecting a null hypothesis when it should *not* be rejected is one sort of error an experimenter can make. This is often called a Type 1 error. The level of significance is the probability of making such a Type 1 error.

There is another error an experimenter can make: *not* rejecting a null hypothesis when it *should* be rejected. This is often called a Type 2 error. What we would prefer is to reject H_0 as often as we can when H_1 is known to be true. The probability of rejecting H_0 when H_1 is known to be true is called the *power* of a test.

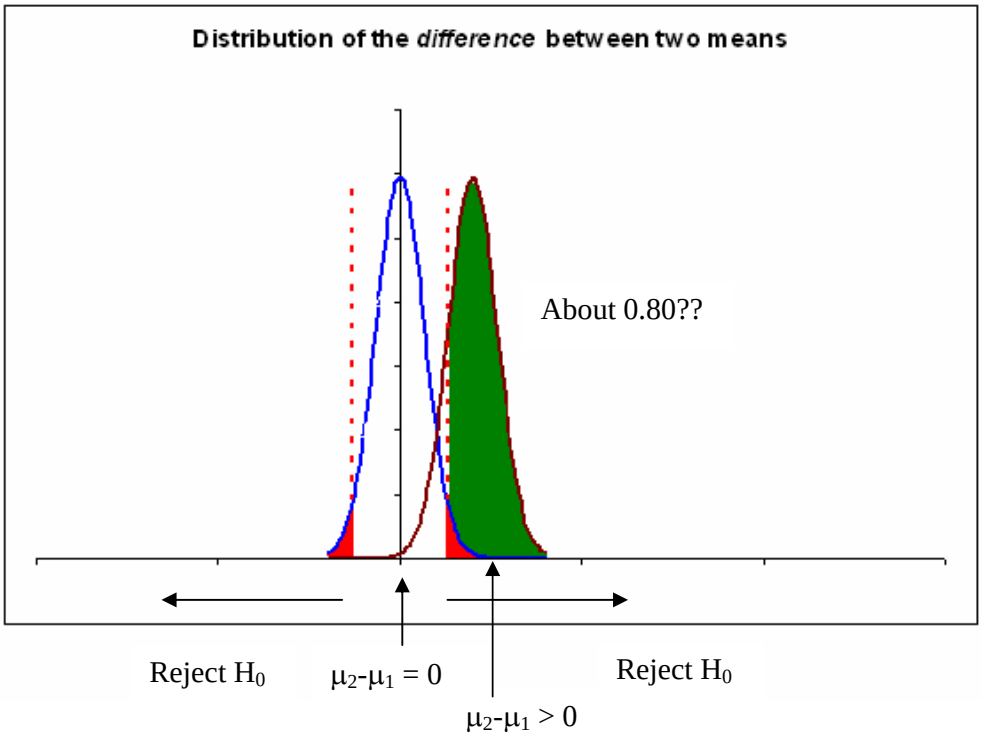
So the second decision to make is how often you wish to reject that the mean difference is zero when you *know* the mean difference is *not* zero. This really depends on how different the two means are. Power depends on the size of the real treatment mean difference. Diagrammatically, it is the area in the rejection region but under a curve centred at a non-zero value for $\mu_2 - \mu_1$.

While wanting to maximise the power, generally it is a compromise between what is desired and what is feasible in an experiment. *However, an experiment with a low power is an experiment that may as well not be conducted.*

Diagram of the *power* of the two sample *t* test (in green).



How can the power be increased? For this trial, we have used $n = 4$ replicates per treatment and have a specified value of $\mu_2-\mu_1$ in mind. Given that we have no control over σ, the only way is to increase n:



Increasing *n* reduces the s.e.d., pulls the H_0 curve (and hence the dotted red line indicating the rejection regions) in towards $\mu_2-\mu_1 = 0$, thereby *increasing* the area (that is, the power) under the H_1 curve.

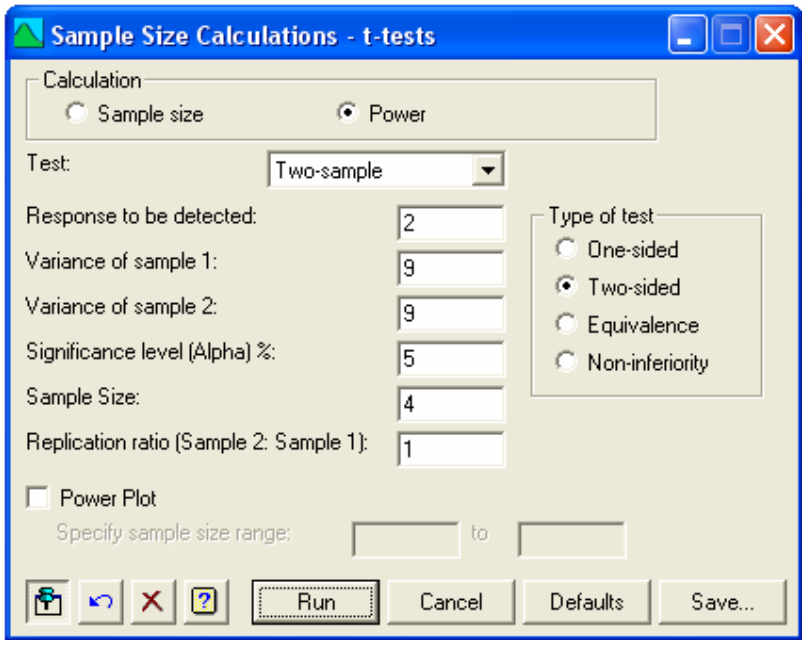
Notice what essential information was required to calculate power:

- the level of significance to be used in the test;
- prior knowledge of the variance σ^2;
- prior knowledge of the target difference in means, $\mu_2-\mu_1$.

With this information, power calculation is straightforward for this experiment.

The diagrams above related to a distribution of differences for which $\sigma = 3$ and there were 4 replicates. The mean difference of interest was 2 units.

In GenStat, select **Stats** > **Sample Size** > **t-tests** and **two-sample**. Enter the response to be detected (a difference of 2 units) in a two-tailed test (corresponding to a two-sided alternative). Enter the variances (they were equal) and indicate equal replication by a **Replication ratio** of 1.



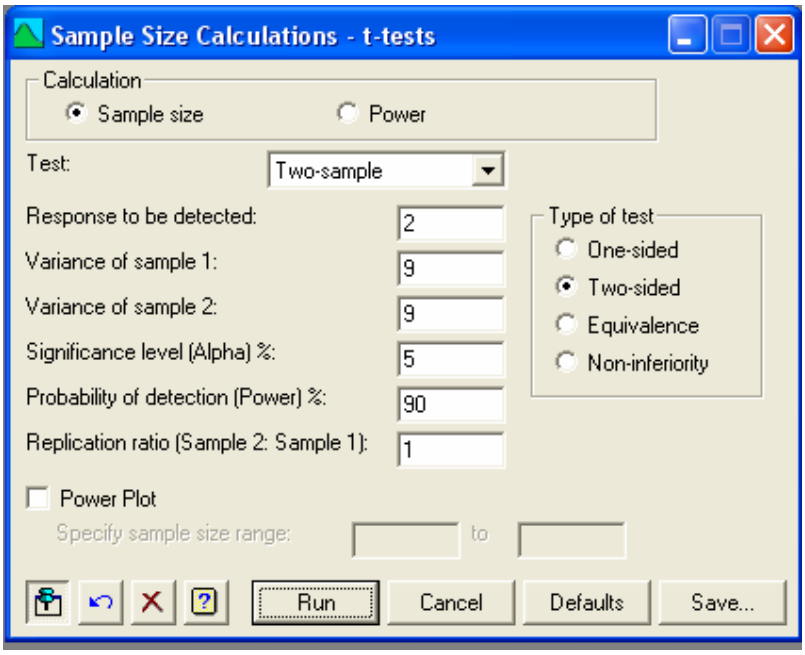
Power for a two-sample t-test

Response 2.000, sample 1 variance 9.000, sample 2 variance 9.000, two-sided significance level 0.050.

No. replicates	Power
4	0.123

We guessed 15%, the actual value is 12.3%. For 40 replicates per treatment, the power turns out to be 83.8%.

If you wish to be 90% confident of picking up a mean difference of 2 units, switch to **Sample Size** in this dialogue box and enter 90 as the **Probability of detection (Power) %**.



Sample size for t-tests

Replication

To detect a response of 2.000, at a two-sided significance level of 0.050 with a power of 0.900 using a two-sample t-test, requires a **replication of 49 for each sampl**e.

Sample size for designed experiments

Recall that an unpaired *t* test is a special case of a completely randomized design. For more than two treatments, the best estimate of σ^2 comes from the Residual MS in the stratum in which treatment means are being tested.

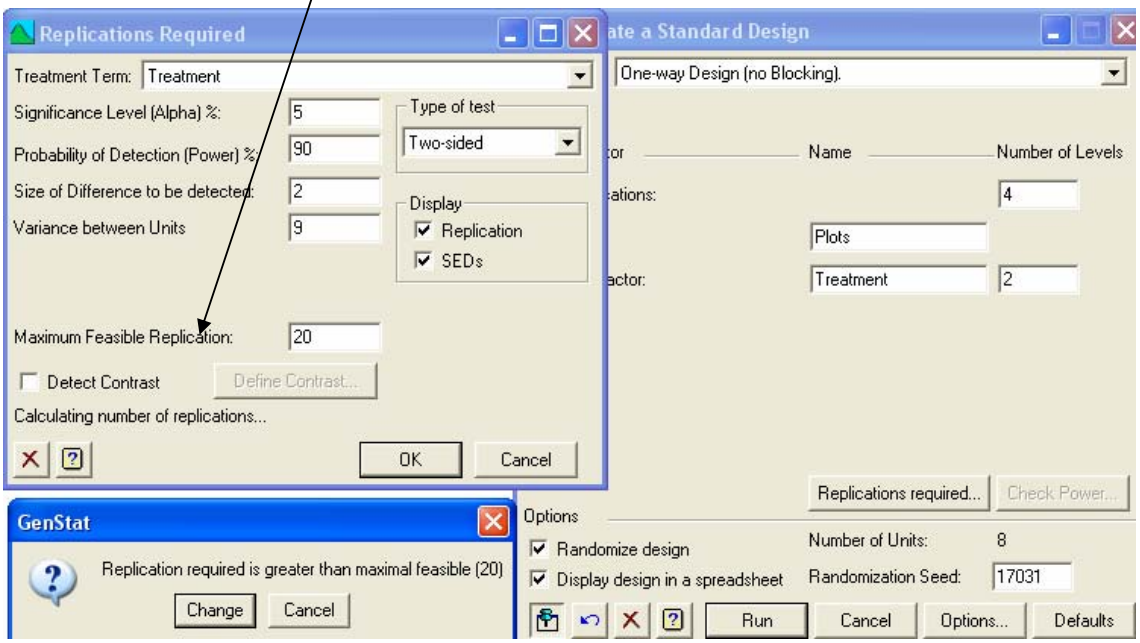
When you have more than two treatments in a design, you have at least two choices of what question to ask.

Firstly, you may have a mean difference in mind for a specified pair of treatments. This is the same question as was asked in the last section, except that we have a superior estimate of variance coming from all treatments.

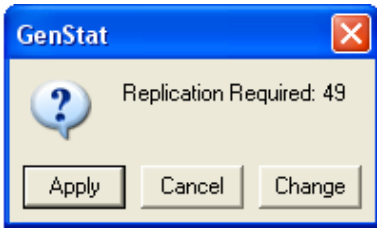
Next, you may be interested only in detecting a difference among the treatment means, but not for two specified treatments. In this case, you might say: how many replicates do I need per treatment in an experiment where the smallest detectable difference among the set of treatment means is a certain amount. What we do then is to calculate the power for the worst case scenario, knowing that if the means have a pattern that is not the extreme set, our power will be larger than the one we calculate. The power is worked out on the basis of areas under the central *F* distribution and under the non-central *F* distribution.

GenStat offers sample size calculations via its **Design** menu.

Firstly, we verify the previous sample size requirement for a two treatment design in which the standard deviations are both 3 (variances = 9), the test will be two-sided and we wish to detect a difference in means of 2 units. We entered 4 replicates per treatment, hoping that that would be enough. Initially, GenStat has a default of 20 replicates per treatment. If you leave that in this instance, a message is returned which forces you to increase that value.



When you increase this number (to say 100), the calculation returns a value of 49 replications (as was calculated in the previous menu). You can then decide if you want GenStat to change your entered number (4) to the one it calculates. If you allow this, the random design generated will have 49 replicates per treatment. Otherwise, it is back to the drawing board!



Consider the experiment reported by Steel and Torrie on page 201 on the effect of washing and removing excess moisture by wiping or by air current on the ascorbic acid content of turnip greens. The data, in mg per 100 g dry weight, are as follows.

	Block 1	Block 2	Block 3	Block 4	Block 5
Control	950	887	897	850	975
Washed and blotted dry	857	1189	918	968	909
Washed and dried in air current	917	1072	975	930	954

Were five blocks sufficient to demonstrate a difference of the order of magnitude found in this experiment? The analysis shows no significant difference between the three means:

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block stratum	4	39609.	9902.	1.50	
Block.*Units* stratum					
Treatment	2	10873.	5436.	0.82	**0.472**
Residual	8	52734.	6592.		
Total	14	103216.			

Tables of means

Variate: Dry_weight

Grand mean 950.

Treatment	1	2	3
	912.	**968.**	**970.**

Standard errors of differences of means

Table	Treatment
rep.	5
d.f.	8
s.e.d.	51.3

Least significant differences of means (5% level)

Table	Treatment
rep.	5
d.f.	8
l.s.d.	118.4

It turns out that with this type of variation and with the actual means observed, there was only a 6% chance to detect a treatment difference. So why do the experiment?

Let us plan a better experiment.

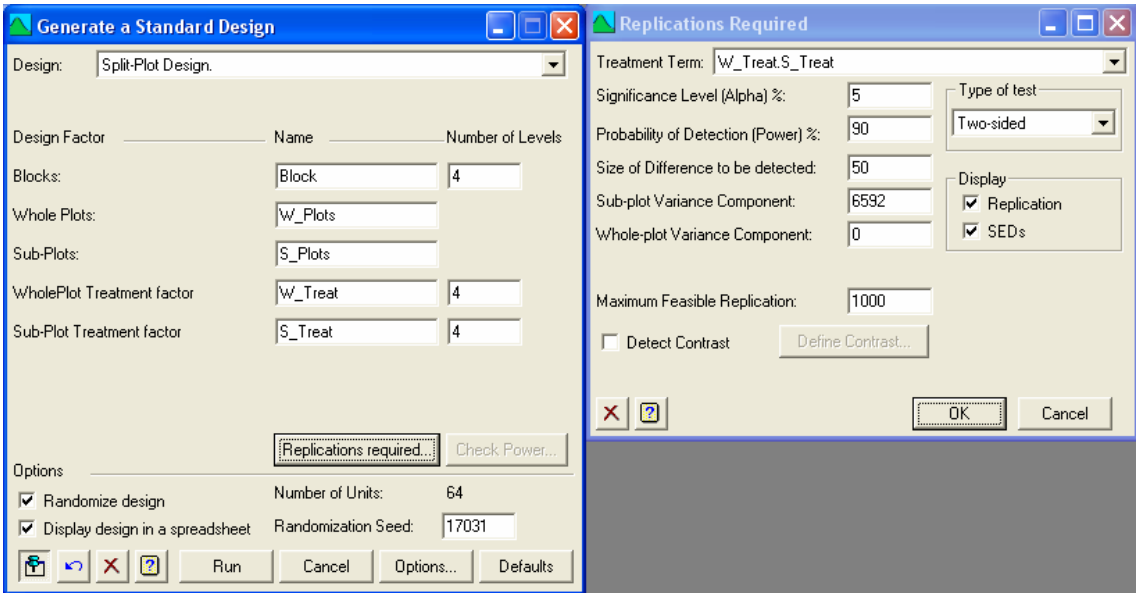
There is no way that a treatment difference of 2 mg per 100 g dry weight will ever be detected if the true variance is 6592. A replication of more than 1000 blocks is required.

So, suppose we concentrate on an experiment designed to measure a difference of at least 50 mg per 100 g dry weight among the treatment means. GenStat returns a value of 69 blocks!

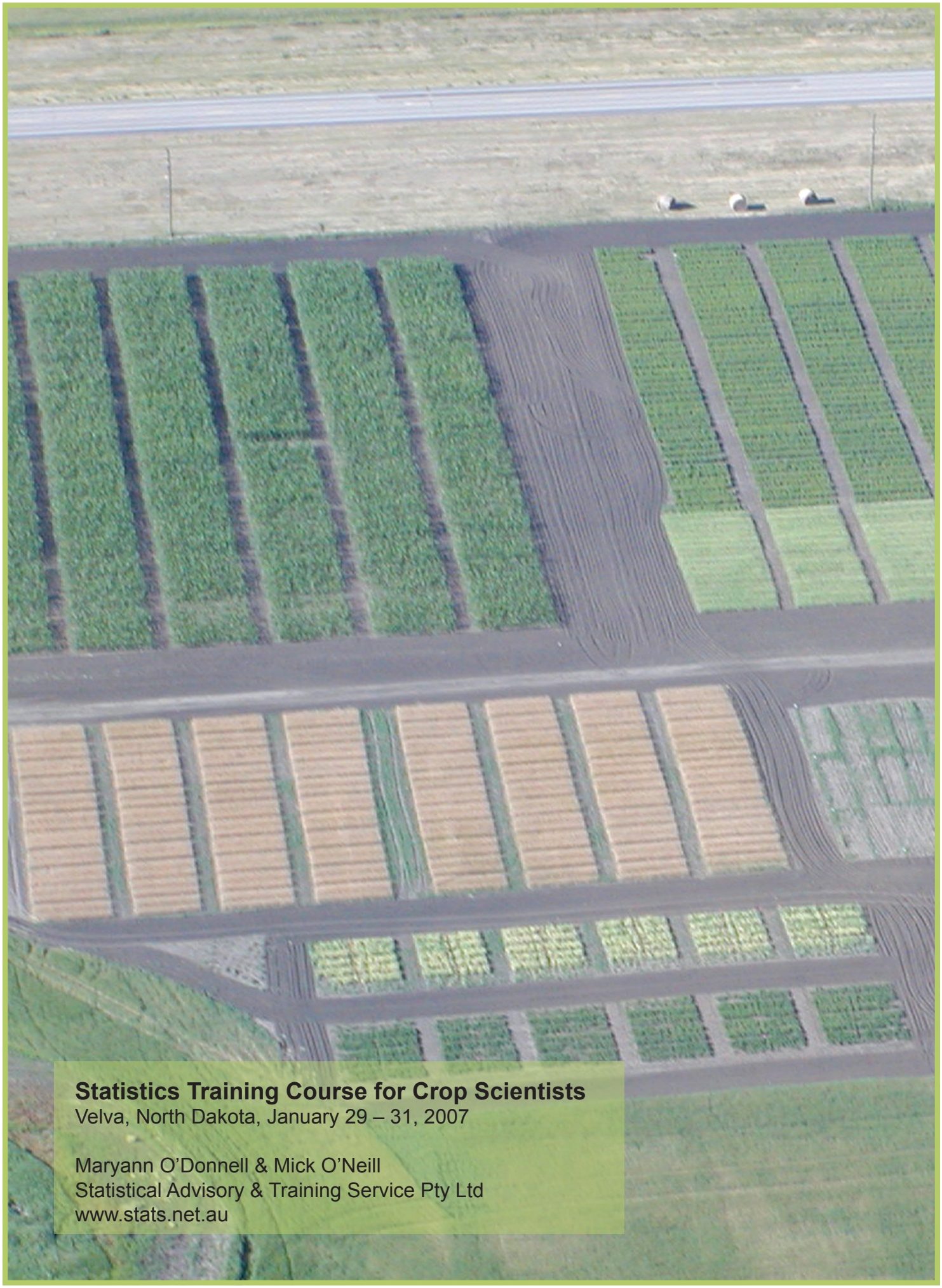
The situation is not improved if the target is defined to be a particular treatment comparison, rather than differences among all three treatments. GenStat calculates that even with this requirement, 57 blocks are required.

This is an example where the variation is much too large for any small mean difference to be detectable.

To summarise, GenStat will calculate the desired number of replicates for quite complex designs. This is a snapshot of a split-plot design. The question can be posed at the whole plot level, or between split-plot treatment means or two-way means in the split-plot stratum. The question could be a difference in specific means, or a minimum detectable difference among all treatment means.



This just touches the surface. There are many other statistics that require different calculations. The point at this stage is to think through your requirements to ensure that experiments you plan to conduct are feasible and achievable.



Statistics Training Course for Crop Scientists

Velva, North Dakota, January 29 – 31, 2007

Maryann O'Donnell & Mick O'Neill
Statistical Advisory & Training Service Pty Ltd
www.stats.net.au