9 Glossary

Symbols:

$\Rightarrow$	implies
$\equiv$	is equivalent to
$\bar{x}$	mean of a set of values of x
ε	error
$\hat{\varepsilon}$	Greenhouse–Geisser correction (see p. 25)
$\tilde{\varepsilon}$	Huynh–Feldt correction (see p. 25)
μ	mean
ρ	population correlation
r	sample correlation
r_{xy} or $r_{x.y}$	correlation between x and y
$r_{y.a,b,c}$	multiple correlation between y and (a, b, c)
$r_{y.(x\|z)}$	semipartial correlation between y and x, having partialled out z (see p. 100)
$r_{y.x\|z}$	partial correlation between y and x, having partialled out z (see p. 100)
$\sum$	sum of (see p. 209)
σ_X	population standard deviation of X
s_X	sample standard deviation of X
σ_X^2	population variance of X
s_x^2	sample variance of X

- **Additive model.** In *within-subjects* ANOVA, a structural model that assumes the effects of within-subjects treatments are the same for all subjects.
- **ANCOVA.** Analysis of covariance: an ANOVA that uses a *covariate* as a *predictor variable*.
- **ANOVA.** Analysis of *variance*. See p. 8→ for an explanation of how it works.
- ***A priori* tests.** Tests planned in advance of obtaining the data; compare *post hoc tests*.
- **Balanced ANOVA.** An ANOVA is said to be balanced when all the cells have equal n, when there are no missing cells, and if there is a *nested design*, when the nesting is balanced so that equal numbers of levels of the nested factor appear in the levels of the factor(s) that they are nested within. This greatly simplifies the computation.
- **Between-subjects** (factor or covariate). If each subject is only tested at a single level of an independent variable, the independent variable is called a between-subjects factor. Compare *within-subjects*.
- **Carryover effects.** See *within-subjects.*
- **Categorical predictor variable.** A variable measured on a nominal scale, whose categories identify class or group membership, used to predict one or more dependent variables. Often called a *factor*.
- **Continuous predictor variable.** A continuous variable used to predict one or more dependent variables. Often called a *covariate*.
- **Covariance matrix.** If you have three variables x, y, z, the covariance matrix, denoted $\sum$, is $\Sigma = \begin{array}{c} \\ x \\ y \\ z \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} \sigma_x^2 & \mathrm{cov}_{xy} & \mathrm{cov}_{xz} \\ \mathrm{cov}_{xy} & \sigma_y^2 & \mathrm{cov}_{yz} \\ \mathrm{cov}_{xz} & \mathrm{cov}_{yz} & \sigma_z^2 \end{bmatrix} \end{array}$ where cov_{xy} is the covariance of x and y ($= \rho_{xy}\sigma_x\sigma_y$ where ρ_{xy} is the correlation between x and y and σ_x is the variance of x). Obviously, $\mathrm{cov}_{xx} = \sigma_x^2$. It is sometimes used to check for **compound symmetry** of the covariance matrix, which is a fancy way of saying

$\sigma_x^2 = \sigma_y^2 = \sigma_z^2$ (all numbers on the leading diagonal the same as each other). and $\text{cov}_{xy} = \text{cov}_{yz} = \text{cov}_{xz}$ (all numbers not on the leading diagonal the same as each other). If there is compound symmetry, there is also *sphericity*, which is what's important when you're running ANOVAs with *within-subjects factors*. On the other hand, you can have sphericity without having compound symmetry; see p. 25→.

- **Conservative.** Apt to give *p* values that are too large.
- **Contrast.** See *linear contrast.*
- **Covariate.** A continuous variable (one that can take any value) used as a *predictor variable*.
- **Degrees of freedom (*df*).** Estimates of parameters can be based upon different amounts of information. The number of independent pieces of information that go into the estimate of a parameter is called the degrees of freedom (d.f. or *df*). Or, the number of observations free to vary. For example, if you pick three numbers at random, you have 3 *df* — but once you calculate the sample mean, $\bar{x}$, you only have two *df* left, because you can only alter two numbers freely; the third is constrained by the fact that you have 'fixed' $\bar{x}$. Or, the number of measurements exceeding the amount absolutely necessary to measure the 'object' (or parameter) in question. To measure the length of a rod requires 1 measurement. If 10 measurements are taken, then the set of 10 measurements has 9 *df*. In general, the df of an estimate is the number of independent scores that go into the estimate minus the number of parameters estimated from those scores as intermediate steps. For example, if the population variance σ^2 is estimated (by the sample variance s^2) from a random sample of n independent scores, then the number of degrees of freedom is equal to the number of independent scores (n) minus the number of parameters estimated as intermediate steps (one, as μ is estimated by $\bar{x}$) and is therefore $n - 1$.
- **Dependent variable.** The variable you measure, but do not control. ANOVA is about predicting the value of a single dependent variable using one or more *predictor variables*.
- **Design matrix.** The matrix in a *general linear model* that specifies the experimental design — how different factors and covariates contribute to particular values of the dependent variable(s).
- **Doubly-nested design.** One in which there are two levels of nesting (see *nested design*). Some are described on p. 159→.
- **Error term.** To test the effect of a predictor variable of interest with an ANOVA, the variability attributable to it ($\text{MS}_{\text{variable}}$) is compared to variability attributed to an appropriate 'error term' (MS_{error}), which measures an appropriate *error variability*. The error term is valid if the *expected mean square* for the variable, $E(\text{MS}_{\text{variable}})$, differs from $E(\text{MS}_{\text{error}})$ only in a way attributable solely to the variable of interest.
- **Error variability** (or error variance, σ_e^2). Variability among observations that cannot be attributed to the effects of the independent variable(s). May include measurement error but also the effects of lots of irrelevant variables that are not measured or considered. It may be possible to reduce the error variability by accounting for some of them, and designing our experiment accordingly. For example, if we want to study the effects of two methods of teaching reading on children's reading performance, rather than randomly assigning all our students to teaching method 1 or teaching method 2, we could split our children into groups with low / medium / high intelligence, and randomly allocate students from each level of intelligence to one of our two teaching methods. If intelligence accounts for some of the variability in reading ability, accounting for it in this way will reduce our error variability. *Within-subjects* designs take this principle further (but are susceptible to *carryover effects*).
- **Expected mean square (EMS).** The value a mean square (MS) would be expected to have if the null hypothesis were true.
- ***F* ratio.** The ratio of two variances. In ANOVA, the ratio of the *mean square (MS)* for a *predictor variable* to the MS of the corresponding *error term*.

- **Factor.** A discrete variable (one that can take only certain values) used as a *predictor variable.* A categorical predictor. Factors have a certain number of *levels.*
- **Factorial ANOVA.** An ANOVA using factors as predictor variables. The term is often used to refer to ANOVAs involving more than one factor (compare *one-way ANOVA*). Factorial designs range from the completely randomized design (subjects are randomly assigned to, and serve in only one of several different treatment conditions, i.e. completely between-subjects design), via mixed designs (both between-subjects and within-subjects factors) to completely within-subjects designs, in which each subject serves in every condition.
- **Fixed factor.** A *factor* that contains all the levels we are interested in (e.g. the factor 'sex' has the levels male and female). Compare *random* factor and see p. 31.
- **Gaussian distribution.** Normal distribution.
- **General linear model.** A general way of predicting one or more *dependent variables* from one or more *predictor variables*, be they categorical or continuous. Subsumes regression, multiple regression, ANOVA, ANCOVA, MANOVA, MANCOVA, and so on.
- **Greenhouse–Geisser correction / epsilon.** If the *sphericity assumption* is violated in an ANOVA involving within-subjects factors, you can correct the *df* for any term involving the WS factor (and the *df* of the corresponding error term) by multiplying both by this correction factor. Often written $\hat{\varepsilon}$, where $0 < \hat{\varepsilon} \leq 1$. Originally from Greenhouse & Geisser (1959).
- **Heterogeneity of variance.** Opposite of *homogeneity of variance.* When variances for different treatments are *not* the same.
- **Hierarchical design.** One in which one variable is *nested* within a second, which is itself nested within a third. A doubly-nested design (such as the split-split plot design) is the simplest form of hierarchical designs. They're complex.
- **Homogeneity of variance.** When a set of variances are all equal. If you perform an ANOVA with a factor with *a* levels, the homogeneity of variance assumption is that $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_a^2 = \sigma_e^2$, where σ_e^2 is the *error variance*.
- **Huynh–Feldt correction / epsilon.** If the *sphericity assumption* is violated in an ANOVA involving within-subjects factors, you can correct the *df* for any term involving the WS factor (and the *df* of the corresponding error term) by multiplying both by this correction factor. Often written $\tilde{\varepsilon}$, where $0 < \tilde{\varepsilon} \leq 1$. Originally from Huynh & Feldt (1970).
- **Independent variable.** The variables thought to be influencing the *dependent variable(s).* In experiments, independent variables are manipulated. In correlational studies, independent variables are observed. (The advantage of the experiment is the ease of making causal inferences.)
- **Interaction.** There is an interaction between factors A and B if the effect of factor A depends on the level of factor B, or vice versa. For example, if your dependent variable is engine speed, and your factors are 'presence of spark plugs (Y / N)' (A) and 'presence of petrol (Y / N)' (B), you will find an interaction such that factor A only influences engine speed at the 'petrol present' level of B; similarly, factor B only influences engine speed at the 'spark plugs present' level of B. This is a binary example, but interactions need not be. Compare *main effect, simple effect.*
- **Intercept.** The contribution of the grand mean to the observations. See p. 65. The *F* test on the intercept term ($MS_{intercept} / MS_{error}$) tests the null hypothesis that the grand mean is zero.
- **Level** (of a factor). One of the values that a discrete predictor variable (factor) can take. For example, the factor Weekday might have five levels — Monday, Tuesday, Wednesday, Thursday, Friday. We might write the factor as $Weekday_5$ in descriptions of ANOVA models (as in 'Tedium = $Drowsiness_2 \times Weekday_5 \times S$'), or write the levels themselves as $Weekday_1$ …$Weekday_5$.
- **Levene's test** (for heterogeneity of variance). Originally from Levene (1960). Tests the assumption of *homogeneity of variance*. If Levene's test produces a 'significant' result, the assumption of homogeneity of variance cannot be made (this is generally a Bad Thing and suggests that you might need to transform your data to improve the situation; see p. 34).

- **Liberal.** Apt to give *p* values that are too small.
- **Linear contrasts.** Comparisons between linear combinations of different groups, used to test specific hypotheses. See p. 75→.
- **Linear regression.** Predicting *Y* from *X* using the equation of a straight line: $\hat{Y} = bX + a$. May be performed with *regression ANOVA*.
- **Logistic regression.** See Howell (1997, pp. 548-558). A logistic function is a sigmoid (see www.mathworld.com). If your dependent variable is dichotomous (categorial) but ordered ('flight on time' versus 'flight late', for example) and you wish to predict this (for example, by pilot experience), a logistic function is often better than a straight line. It reflects the fact that the dichotomy imposes a cutoff on some underlying continuous variable (e.g. once your flight delay in seconds — continuous variable — reaches a certain level, you classify the flight as late — dichotomous variable). Dichotomous variables can be converted into variables suitable for linear regression by converting the *probability* of falling into one category, *P*(flight late), into the *odds* of falling into that category, using $\text{odds} = \frac{P(A)}{P(\neg A)}$, and then into the *log odds*, using the natural (base *e*) logarithm $\log_e(\text{odds}) = \ln(\text{odds})$. The probability is therefore a logistic function of the log odds: $\text{probability} = \frac{e^{\ln(\text{odds})}}{1 + e^{\ln(\text{odds})}}$, so performing a linear regression on the log odds is equivalent to performing a logistic regression on probability. This is pretty much what logistic regression does, give or take some procedural wrinkles. Odds ratios (likelihood ratios), the odds for one group divided by the odds for another group, emerge from logistic regression in the way that slope estimates emerge from linear regression, but the statistical tests involved are different. Logistic regression is a computationally iterative task; there's no simple formula (the computer works out the model that best fits the data iteratively).
- **Main effect.** A main effect is an effect of a factor *regardless* of the other factor(s). Compare *simple effect; interaction.*
- **MANCOVA.** Multivariate analysis of covariance; see *MANOVA* and *ANCOVA*.
- **MANOVA.** Multivariate ANOVA — ANOVA that deals with multiple *dependent variables* simultaneously. Not covered in this document. For example, if you think that your treatment has a bigger effect on dependent variable Y_2 than on variable Y_1, how can you see if that is the case? Certainly not by making categorical decisions based on *p* values (significant effect on Y_1, not significant effect on Y_2 — this wouldn't mean that the effect on Y_1 and Y_2 were significantly different!). Instead, you should enter Y_1 and Y_2 into a MANOVA.
- **Mauchly's test** (for sphericity of the covariance matrix). Originally from Mauchly (1940). See *sphericity*, *covariance matrix*, and p. 25.
- **Mean square (MS).** A *sum of squares (SS)* divided by the corresponding number of *degrees of freedom (df)*, or number of independent observations upon which your SS was based. This gives you the mean 'squared deviation from the mean', or the 'mean square'. Effectively, a *variance*.
- **Mixed model.** An ANOVA model that includes both *between-subjects* and *within-subjects* predictor variables. Alternatively, one that includes both *fixed* and *random* factors. The two uses are often equivalent in practice, since Subjects is usually a random factor.
- **Multiple regression.** Predicting a dependent variable on the basis of two or more continuous variables. Equivalent to ANOVA with two or more *covariates*.
- **Nested design.** An ANOVA design in which variability due to one factor is 'nested' within variability due to another factor. For example, if one were to administer four different tests to four school classes (i.e. a between-groups factor with four levels), and two of those four classes are in school A, whereas the other two classes are in school B, then the levels of the first factor (four different tests) would be nested in the second factor (two different schools). A very common example is a design with one between-subjects factor and one within-subjects factor, written A × (U × S); variation due to subjects is nested within variation due to A (or, for short-hand, S is nested within A), because each subject is only tested at one level of the between-subjects factor(s). We might write this S / A ('S is nested within A'); SPSS uses the alternative notation of S(A). See also *doubly-nested design*.

- **Nonadditive model.** In *within-subjects* ANOVA, a structural model that allows that the effects of within-subjects treatments can differ across subjects.
- **Null hypothesis.** For a general discussion of null hypotheses, see handouts at www.pobox.com/~rudolf/psychology. In a one-way ANOVA, when you test the main effect of a factor A with *a* levels, your null hypothesis is that $\mu_1 = \mu_2 = \ldots = \mu_a$. If you reject this null hypothesis (if your *F* ratio is large and significant), you conclude that the effects of all *a* levels of A were not the same. But if there are >2 levels of A, you do not yet know which levels differed from each other; see *post hoc tests.*
- **One-way ANOVA.** ANOVA with a single between-subjects factor.
- **Order effects.** See *within-subjects.*
- **Overparameterized model.** A way of specifying a *general linear model* design matrix in which a separate predictor variable is created for each group identified by a factor. For example, to code Sex, one variable would be created in which males score 1 and females score 0, and another variable would be created in which males score 0 and females score 1. These two variables contain mutually redundant information: there are more predictor variables than are necessary to determine the relationship of a set of predictors to a set of dependent variables. Compare *sigma-restricted model.*
- **Planned contrasts.** *Linear contrasts* run as *a priori tests.*
- **Polynomial ANCOVA.** An *ANCOVA* in which a nonlinear term is used as a *predictor variable* (such as x^2, x^3…, rather than the usual x). See Myers & Well (1995, p. 460).
- ***Post hoc* tests.** Statistical tests you run after an ANOVA to examine the nature of any main effects or interactions you found. For example, if you had an ANOVA with a single *between-subjects factor* with three *levels*, sham/core/shell, and you found a main effect of this factor, was this due to a difference between sham and core subjects? Sham and shell? Shell and core? Are all of them different? There are many *post hoc* tests available for this sort of purpose. However, there are statistical pitfalls if you run many post-hoc tests; you may make Type I errors (see handouts at www.pobox.com/~rudolf/psychology) simply because you are running lots of tests. *Post hoc* tests may include further ANOVAs of subsets of your original data — for example, after finding a significant Group × Difficulty *interaction*, you might ask whether there was a *simple effect* of Group at the 'easy' level of the Difficulty factor, and whether there was a *simple effect* of Group at the 'difficult' level of the Difficulty factor (see pp. 20, 39→).
- **Power of an ANOVA.** Complex to work out. But things that increase the expected *F* ratio for a particular term if the null hypothesis is false increase power, and $F = \frac{\mathrm{MS}_{\mathrm{predictor}}}{\mathrm{MS}_{\mathrm{error}}} = \frac{\mathrm{SS}_{\mathrm{predictor}} \times df_{\mathrm{error}}}{\mathrm{SS}_{\mathrm{error}} \times df_{\mathrm{predictor}}}$. Bigger samples contribute to a larger *df* for your error term; this therefore decreases $\mathrm{MS}_{\mathrm{error}}$ and increases the expected *F* if the null hypothesis is false, and this therefore increases your power. The larger the ratio of $E(\mathrm{MS}_{\mathrm{treatment}})$ to $E(\mathrm{MS}_{\mathrm{error}})$, the larger your power. Sometimes two different structural models give you different EMS ratios; you can use this principle to find out which is more powerful for detecting the effects of a particular effect (see p. 73→). For references to methods of calculating power directly, see p. 102.
- **Predictor variable.** Factors and covariates: things that you use to predict your dependent variable.
- **Pseudoreplication.** What you do when you analyse correlated data without accounting for the correlation. A Bad Thing — entirely Wrong. For example, you could take 3 subjects, measure each 10 times, and pretend that you had 30 independent measurements. No, no, no, no, no. Account for the correlation in your analysis (in this case, by introducing a Subject factor and using an appropriate ANOVA design with a within-subjects factor).
- **Random factor.** A *factor* whose levels we have sampled at random from many possible alternatives. For example, Subjects is a random factor — we pick our subjects out of a large potential pool, and if we repeat the experiment, we may use different subjects. Compare *fixed factor* and see p. 31.

- **Regression ANOVA.** Performing linear regression using ANOVA. A simple linear regression is an ANOVA with a single *covariate* (i.e. ANCOVA) and no other factors.
- **Repeated measures.** Same as within-subjects. 'Repeated measures' is the more general term — within-subjects designs involve repeated measurements of the same subject, but things other than subjects can also be measured repeatedly. In general, within-subjects / repeated-measures analysis is to do with accounting for *relatedness* between sets of observations above that you'd expect by chance. Repeated measurement of a subject will tend to generate data that are more closely related (by virtue of coming from the same subject) than data from different subjects.
- **Robust.** A test that gives correct *p* values even when its assumptions are violated to some degree ('this test is fairly robust to violation of the normality assumption…').
- **Sequence effects.** See *within-subjects.*
- **Sigma-restricted model.** A way of specifying a *general linear model* in which a categorical variable with *k* possible levels is coded in a design matrix with *k* – 1 variables. The values used to code membership of particular groups sum to zero. For example, to code Sex, one variable would be created in which males score +1 and females –1. Compare *overparameterized model.*
- **Simple effect.** An effect of one factor *considered at only one level* of another factor. A simple effect of A at level 2 of factor B is written 'A at B_2' or 'A / B_2'. See *main effect, interaction,* and pp. 20, 39→.
- **Source of variance (SV).** Something contributing to variation in a dependent variable. Includes *predictor variables* and *error variability*.
- **Sphericity assumption.** An important assumption applicable to *within-subjects (repeated measures)* ANOVA. Sphericity is the assumption of *homogeneity of variance of difference scores.* Suppose we test 5 subjects at three levels of A. We can therefore calculate three sets of difference scores ($A_3 - A_2$), ($A_2 - A_1$), and ($A_3 - A_1$), for each subject. Sphericity is the assumption that the variances of these difference scores are the same. See p. 25→.
- **Standard deviation.** The square root of the *variance.*
- **Structural model.** An equation giving the value of the *dependent variable* in terms of *sources of variability* including *predictor variables* and *error variability.*
- **Sum of squares (SS).** In full, the sum of the squared deviations from the mean. See *variance*. Sums of squares are used in preference to actual variances in ANOVA, because sample sums of squares are additive (you can add them up and they still mean something) whereas sample variances are not additive unless they're based on the same number of *degrees of freedom*.
- ***t* test, one-sample.** Equivalent to testing $MS_{intercept} / MS_{error}$ with an ANOVA with no other factors (odd as that sounds). $F_{1,k} = t_k^2$ and $t_k = \sqrt{F_{1,k}}$. See *intercept.*
- ***t* test, two-sample, paired.** Equivalent to ANOVA with one within-subjects factor with two levels. $F_{1,k} = t_k^2$ and $t_k = \sqrt{F_{1,k}}$.
- ***t* test, two-sample, unpaired.** Equivalent to ANOVA with one between-subjects factor with two levels. $F_{1,k} = t_k^2$ and $t_k = \sqrt{F_{1,k}}$.
- **Variance.** To calculate the variance of a set of observations, take each observation and subtract it from the mean. This gives you a set of deviations from the mean. Square them and add them up. At this stage you have the sum of the squared deviations from the mean, also known as the *sum of squares (SS).* Divide by the number of independent observations you have (*n* for the population variance; *n*–1 for the sample variance; or, in general, the number of *degrees of freedom*) to get the variance. See the Background Knowledge handouts at www.pobox.com / ~rudolf / psychology.
- **Within-subjects** (factor or covariate). See also *repeated measures*. If a score is obtained for every subject at each level of an independent variable, the independent variable is called a within-subjects factor. See also *between-subjects.* The advantage of a within-subjects design is that the different treatment conditions are automatically matched on many irrelevant variables — all those that

are relatively unchanging characteristics of the subject (e.g. intelligence, age). However, the design requires that each subject is tested several times, under different treatment conditions. Care must be taken to avoid *order*, *sequence* or *carryover* effects — such as the subject getting better through practice, worse through fatigue, drug hangovers, and so on. If the effect of a treatment is permanent, it is not possible to use a within-subjects design. You could not, for example, use a within-subjects design to study the effects of parachutes (versus no parachute) on mortality rates after falling out of a plane.

10 Further reading

- A very good statistics textbook for psychology is Howell (1997).
- Abelson (1995) examines statistics as an technique of argument and is very clear on the logical principles and some of the philosophy of statistics.
- Keppel (1991) is a fairly hefty tome on ANOVA techniques. Winer (1991) is another monster reference book. Neither are for the faint-hearted.
- Myers & Well (1995) is another excellent one. Less fluffy than Howell (1997) but deals with the issues head on.

There is also an excellent series of Statistics Notes published by the British Medical Journal, mostly by Bland and Altman. A list is available at

www.mbland.sghms.ac.uk / pbstnote.htm

and the articles themselves are available online from

www.bmj.com

This series includes the following:

- The problem of the 'unit of analysis' (Altman & Bland, 1997). Correlation and regression when repeated measurements are taken, and the problem of pseudoreplication (Bland & Altman, 1994a). The approach one should take to measure correlation within subjects (Bland & Altman, 1995a) and correlation between subjects (Bland & Altman, 1995b).
- Why correlation is utterly inappropriate for assessing whether two ways of measuring something agree (Bland & Altman, 1986).
- Generalization and extrapolation (Altman & Bland, 1998).
- Why to randomize (Altman & Bland, 1999b), how to randomize (Altman & Bland, 1999a), and how to match subjects to different experimental groups (Bland & Altman, 1994b).
- Blinding (Day & Altman, 2000; Altman & Schulz, 2001).
- Absence of evidence is not evidence of absence — about power (Altman & Bland, 1995).
- Multiple significance tests: the problem (Bland & Altman, 1995c).
- Regression to the mean (Bland & Altman, 1994e; Bland & Altman, 1994d).
- One-tailed and two-tailed significance tests (Bland & Altman, 1994c).
- Transforming data (Bland & Altman, 1996b) and how to calculate confidence intervals with transformed data (Bland & Altman, 1996c; Bland & Altman, 1996a).
- ANOVA, briefly (Altman & Bland, 1996), and the analysis of interaction effects (Altman & Matthews, 1996; Matthews & Altman, 1996a; Matthews & Altman, 1996b).
- Comparing estimates derived from separate analyses (Altman & Bland, 2003).
- Dealing with differences in baseline by ANCOVA (Vickers & Altman, 2001).

Finally, there's an excellent on-line textbook (StatSoft, 2002):

www.statsoft.nl / textbook / stathome.html

11 Bibliography

Abelson, R. P. (1995). *Statistics As Principled Argument*, Lawrence Erlbaum, Hillsdale, New Jersey.

Altman, D. G. & Bland, J. M. (1995). Absence of evidence is not evidence of absence. *British Medical Journal* **311:** 485.

Altman, D. G. & Bland, J. M. (1996). Comparing several groups using analysis of variance. *British Medical Journal* **312:** 1472-1473.

Altman, D. G. & Bland, J. M. (1997). Statistics notes. Units of analysis. *British Medical Journal* **314:** 1874.

Altman, D. G. & Bland, J. M. (1998). Generalisation and extrapolation. *British Medical Journal* **317:** 409-410.

Altman, D. G. & Bland, J. M. (1999a). How to randomise. *British Medical Journal* **319:** 703-704.

Altman, D. G. & Bland, J. M. (1999b). Statistics notes. Treatment allocation in controlled trials: why randomise? *British Medical Journal* **318:** 1209.

Altman, D. G. & Bland, J. M. (2003). Interaction revisited: the difference between two estimates. *British Medical Journal* **326:** 219.

Altman, D. G. & Matthews, J. N. (1996). Statistics notes. Interaction 1: Heterogeneity of effects. *British Medical Journal* **313:** 486.

Altman, D. G. & Schulz, K. F. (2001). Statistics notes: Concealing treatment allocation in randomised trials. *British Medical Journal* **323:** 446-447.

Bland, J. M. & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **i:** 307-310.

Bland, J. M. & Altman, D. G. (1994a). Correlation, regression, and repeated data. *British Medical Journal* **308:** 896.

Bland, J. M. & Altman, D. G. (1994b). Matching. *British Medical Journal* **309:** 1128.

Bland, J. M. & Altman, D. G. (1994c). One and two sided tests of significance. *British Medical Journal* **309:** 248.

Bland, J. M. & Altman, D. G. (1994d). Regression towards the mean. *British Medical Journal* **308:** 1499.

Bland, J. M. & Altman, D. G. (1994e). Some examples of regression towards the mean. *British Medical Journal* **309:** 780.

Bland, J. M. & Altman, D. G. (1995a). Calculating correlation coefficients with repeated observations: Part 1--Correlation within subjects. *British Medical Journal* **310:** 446.

Bland, J. M. & Altman, D. G. (1995b). Calculating correlation coefficients with repeated observations: Part 2--Correlation between subjects. *British Medical Journal* **310:** 633.

Bland, J. M. & Altman, D. G. (1995c). Multiple significance tests: the Bonferroni method. *British Medical Journal* **310:** 170.

Bland, J. M. & Altman, D. G. (1996a). Transformations, means, and confidence intervals. *British Medical Journal* **312:** 1079.

Bland, J. M. & Altman, D. G. (1996b). Transforming data. *British Medical Journal* **312:** 770.

Bland, J. M. & Altman, D. G. (1996c). The use of transformation when comparing two means. *British Medical Journal* **312:** 1153.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effect of inequality of variance and of correlation of errors in the two-way classification. *Annals of Mathematical Statistics* **25:** 484-498.

Boyd, O., Mackay, C. J., Lamb, G., Bland, J. M., Grounds, R. M. & Bennett, E. D. (1993). Comparison of clinical information gained from routine blood-gas analysis and from gastric tonometry for intramural pH. *Lancet* **341:** 142-146.

Cardinal, R. N., Parkinson, J. A., Djafari Marbini, H., Toner, A. J., Bussey, T. J., Robbins, T. W. & Everitt, B. J. (2003). Role of the anterior cingulate cortex in the control over behaviour by Pavlovian conditioned stimuli in rats. *Behavioral Neuroscience* **117:** 566-587.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. First edition, Academic Press, New York.

Day, S. J. & Altman, D. G. (2000). Statistics notes: blinding in clinical trials and other studies. *British Medical Journal* **321:** 504.

Field, A. P. (1998). A bluffer's guide to sphericity. *Newsletter of the Mathematical, Statistical and computing section of the British Psychological Society* **6:** 13-22.

Frank, H. & Althoen, S. C. (1994). *Statistics: Concepts and Applications*, Cambridge, Cambridge University Press.

Greenhouse, S. W. & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika* **24:** 95-112.

Howell, D. C. (1997). *Statistical Methods for Psychology*. Fourth edition, Wadsworth, Belmont, California.

Huynh, H. & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measures designs have exact *F*-distributions. *Journal of the American Statistical Association* **65:** 1582-1589.

Keppel, G. (1982). *Design and analysis: a researcher's handbook*. Second edition, Englewood Cliffs: Prentice-Hall, London.

Keppel, G. (1991). *Design and analysis: a researcher's handbook*. Third edition, Prentice-Hall, London.

Levene, H. (1960). Robust tests for the equality of variance. In *Contributions to probability and statistics* (Oklin, I., ed.). Stanford University Press, Palo Alto, California.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association* **62:** 399-402.

Matthews, J. N. & Altman, D. G. (1996a). Interaction 3: How to examine heterogeneity. *British Medical Journal* **313:** 862.

Matthews, J. N. & Altman, D. G. (1996b). Statistics notes. Interaction 2: Compare effect sizes not P values. *British Medical Journal* **313:** 808.

Mauchly, J. W. (1940). Significance test for sphericity of a normal *n*-variate distribution. *Annals of Mathematical Statistics* **11:** 204-209.

Myers, J. L. & Well, A. D. (1995). *Research Design and Statistical Analysis*, Lawrence Erlbaum, Hillsdale, New Jersey.

Prescott, C. E., Kabzems, R. & Zabek, L. M. (1999). Effects of fertilization on decomposition rate of Populus tremuloides foliar litter in a boreal forest. *Canadian Journal of Forest Research* **29:** 393-397.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in C*. Second edition, Cambridge University Press, Cambridge, UK.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin* **2:** 110–114.

Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52:** 591-611.

SPSS (2001). SPSS 11.0 Syntax Reference Guide (spssbase.pdf).

StatSoft (2002). *Electronic Statistics Textbook* (http://www.statsoft.com/textbook/stathome.html), Tulsa, OK.

Tangren, J. (2002). A Field Guide To Experimental Designs (http://www.tfrec.wsu.edu/ANOVA/, 2004). Washington State University, Tree Fruit Research and Extension Center.

Vickers, A. J. & Altman, D. G. (2001). Statistics notes: Analysing controlled trials with baseline and follow up measurements. *British Medical Journal* **323:** 1123-1124.

Winer, B. J. (1971). *Statistical principles in experimental design*. Second edition, McGraw-Hill, New York.

Winer, B. J., Brown, D. R. & Michels, K. M. (1991). *Statistical Principles in Experimental Design*, McGraw-Hill, New York, NY.