

Analysis of variance

Rudolf N. Cardinal, MA PhD MB BChir MD MRCP

Academic Clinical Fellow

Department of Psychiatry, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, UK.

Telephone: +44-(0)7092-340641. Fax: +44-(0)7092-340645.

E-mail: rudolf@pobox.com.

[NOTE: home address for pre-publication correspondence is as previously supplied to Wiley & Sons, but that address is not for publication; address for publication is as above.]

*For Weiner IB & Craighead WE (editors),
The Corsini Encyclopedia of Psychology, John Wiley & Sons*

FILES

- Manuscript: (this file; Cardinal_ANOVA_CorsiniEncyclopedia.doc; Word 97 format)
- Figure 1: Cardinal_Figure_1_ANOVA_illustrations.ai (Adobe Illustrator 11 format)
- Figure 2: Cardinal_Figure_2_SS_and_correlated_predictors.ai (Adobe Illustrator 11 format)

KEY WORDS

analysis of variance (ANOVA); general linear model; *F* test; interaction; main effect; simple effect; between-subjects design; within-subjects design; repeated-measures design

OVERVIEW

Analysis of variance (ANOVA) was initially developed by R. A. Fisher, beginning around 1918, and had early applications in agriculture. It is now a dominant and powerful statistical technique used extensively in psychology. In ANOVA, a dependent variable is predicted by a mathematical model comprising one or more predictor variables, which may be categorical (factors) or quantitative and continuous (covariates; regressors). The model's best prediction is calculated by minimizing the sum of the squared residuals (errors, or deviations from the model's prediction). Having done this, the proportion of variance in the dependent variable accounted for by each predictor is assessed statistically, testing the null hypotheses that the mean of the dependent variable does not vary with the predictor(s); good predictors account for a large proportion of the variance, compared to unpredicted (error) variance, and poor predictors account for a small proportion. ANOVA allows the effects of predictors to be assessed in isolation, but also allows the assessment of interactions between predictors (effects of one predictor that depend on the values of other predictors).

ASSUMPTIONS

ANOVA assumes (1) that for each condition represented by a combination of predictor values, the dependent variable is normally distributed about the mean for that condition, or equivalently that error is normally distributed in each condition; (2) that the variance is the same in all such conditions, and the same as the residual variance (known as the homogeneity of variance assumption); (3) that the model accounts for all systematic influences on the data, such that the residual variability represents random error and all errors or observations in each condition are independent of each other. Checking these assumptions involves consideration of the experimental design, exploratory data analysis prior to ANOVA, and analysis of the residuals afterwards. Sometimes, the suitability of data for ANOVA may be improved by appropriate *transformations* prior to analysis. Additional homogeneity-of-variance assumptions apply to ANOVA involving *within-subjects factors*, discussed below.

LOGIC OF ONE-WAY ANOVA

The basic logic of ANOVA is simply illustrated using a single categorical predictor (single-factor or one-way ANOVA). Suppose A is a factor that can take one of k values (representing, for example, k different experimental treatments), and n values of the dependent variable Y have been sampled for each of the conditions $A_1 \dots A_k$ (giving nk observations in total). The null hypothesis is that the dependent variable Y has the same mean for each of $A_1 \dots A_k$, i.e. that A has no effect upon the mean of Y. ANOVA calculates a mean square for the predictor (MS_A) and an error mean square (MS_{error}) and then compares them.

Mean Square For The Predictor

If the null hypothesis is true, then the k samples have been drawn from the same population, and by the Central Limit Theorem, the variance of the k sample means is an estimator of σ_e^2 / n , where σ_e^2 is the population (and error) variance, so n times the variance of the sample means estimates σ_e^2 . However, if the null hypothesis is false, the sample means have come from populations with different means, and n times the variance of the k sample means will exceed this value. The sum of squared deviations (abbreviated to *sum of squares*; SS) of each condition's mean from the grand mean (\bar{Y}) is calculated for the predictor, summing across all observations (in this example,

$$SS_A = \sum_{i=1 \dots k} \sum_{j=1 \dots n} [\bar{Y}_{A_i} - \bar{Y}]^2 = \sum_{i=1 \dots k} n [\bar{Y}_{A_i} - \bar{Y}]^2$$

), and divided by the *degrees of freedom* (df) for the predictor (in this example, $df_A = k - 1$) to give the *mean square* (MS) for the predictor ($MS_A = SS_A / df_A$). If the null hypothesis is true, then the expected value of this number, $E(MS_A)$, is the error variance σ_e^2 . If the null hypothesis is false, then the expected mean square will exceed σ_e^2 , as it will contain contributions from the non-zero effect that A is having on Y.

Mean Square For Error

Whether or not the null hypothesis is true, the sample variances $s_{A_1}^2 \dots s_{A_k}^2$ estimate the corresponding population variances $\sigma_{A_1}^2 \dots \sigma_{A_k}^2$, and by the homogeneity of variance assumption, also estimate the error variance σ_e^2 . An estimate of σ_e^2 is therefore obtainable from the sample variances $s_{A_1}^2 \dots s_{A_k}^2$. The sum of squared deviations of each observation from its group mean—that is, the summed squared residual

deviations after the prediction has been made—is calculated (in this example,

$SS_{\text{error}} = \sum_{i=1 \dots k} \sum_{j=1 \dots n} [Y_{i,j} - \bar{Y}_{A_i}]^2$) and divided by the error degrees of freedom (in this

example, $df_{\text{error}} = df_{\text{total}} - df_A = [nk - 1] - [k - 1] = k[n - 1]$) to give

$MS_{\text{error}} = SS_{\text{error}} / df_{\text{error}}$. Whether the null hypothesis is true or not, $E(MS_{\text{error}}) = \sigma_e^2$.

The graphical meaning of these SS terms is shown in Figure 1a.

F Test

Comparison of MS_A to MS_{error} thus allows assessment of the null hypothesis. The ratio $F = MS_A / MS_{\text{error}}$ is assessed using an F test with $(df_A, df_{\text{error}})$ degrees of freedom. From the observed value of F , a p value may be calculated (the probability of obtaining an F this large or greater, given the null hypothesis). In conventional approaches, a sufficiently large F and small p leads to the rejection of the null hypothesis.

INTERACTIONS, MAIN EFFECTS, AND SIMPLE EFFECTS

When multiple factors are used in analysis, a key feature of ANOVA is its ability to test for *interactions* between factors, meaning effects of one factor that depend on the value (*level*) of other factor(s). The terminology will be illustrated in the abstract, temporarily ignoring important statistical caveats such as homogeneity of variance. Suppose the maximum speeds of many scrap cars are analysed using two factors: E (levels: E_0 engine broken, E_1 engine intact) and F (levels: F_0 no fuel, F_1 fuel present). There will be a *main effect* of Engine: on average, ignoring everything else, E_1 cars go faster than E_0 cars. Similarly, there will be a main effect of Fuel: ignoring everything else, cars go faster with fuel than without. Since speeds will be high in the E_1F_1 condition and very low otherwise, there will also be an *interaction*, meaning that the effect of Engine depends on the level of the Fuel factor, and vice versa (the effect of a working engine depends on whether there is fuel; the effect of fuel depends on whether there is a working engine). One can also speak of *simple effects*: for example, the simple effect of Fuel at the E_1 level is large (fuel makes intact cars go) whereas the simple effect of Fuel factor at the E_0 level is small (fuel makes no difference to broken cars). Likewise, the Engine factor will have a large simple effect at F_1 but a small simple effect at F_0 . Main effects may be irrelevant in the presence of an interaction, since they gloss over known interrelationships between the factors.

A two-factor ANOVA is illustrated in Figure 1b. ANOVA results are conventionally reported in the form shown in Table 1.

POST-HOC ANALYSIS OF INTERACTIONS AND MULTILEVEL FACTORS

Effects of interest in an ANOVA that are found to be significant are frequently analysed further. Interactions may be followed up by analysing their component parts, for example by restricting the analysis to a subset of the data (such as by examining simple effects, or simpler component interactions within a complex interaction). Likewise, significant main effects of interest may require further analysis. For example, if a three-level factor A is found to have an effect, then the null hypothesis $\mu_{A_1} = \mu_{A_2} = \mu_{A_3}$ is rejected, but the experimenter may still have an interest in which of $\mu_{A_1} \neq \mu_{A_2} = \mu_{A_3}$, $\mu_{A_1} = \mu_{A_2} \neq \mu_{A_3}$, and $\mu_{A_1} \neq \mu_{A_2} \neq \mu_{A_3}$ is the case. A range of *post hoc* tests is available for this purpose. The most important feature of such tests is that they involve the potential for multiple comparisons, and thus have the potential to inflate the Type I error rate, particularly as the number of levels of the factor increases. Appropriate tests control the maximum Type I error rate for a whole ‘family’ of comparisons.

ANOVA AS A GENERAL LINEAR MODEL (GLM)

More generally, each observed value of Y can be modelled as a sum of predictors each multiplied by a regression coefficient (*b*), plus error not accounted for by the prediction (*e*):

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{1,1} & X_{1,2} & X_{1,\dots} & X_{1,p} \\ 1 & X_{2,1} & X_{2,2} & X_{2,\dots} & X_{2,p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{n,1} & X_{n,2} & X_{n,\dots} & X_{n,p} \end{bmatrix} \times \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \dots \\ b_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$$

In matrix notation, this may be written:

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{e}$$

Here, **Y** is a matrix containing values of the dependent variable. **X** is the *design matrix*, containing columns for *p* predictors plus a column of ones to represent the additional predictor of the ‘overall mean’ of Y. Each row of **X** encodes the predictors and which levels of those predictors (for categorical predictors), or values of the predictors (for continuous predictors), apply to a given value of Y. Appropriate choice

of a design matrix allows arbitrary designs including factors with multiple levels, combinations of categorical and continuous predictors, and interactions to be encoded. The matrix \mathbf{b} contains coefficients (in this example, b_0 is the value of the overall Y mean, \bar{Y}), and \mathbf{e} contains errors. This equation may be solved for \mathbf{b} so as to minimize the summed squared residuals ($\sum e^2$). Having done so, $SS_{\text{error}} = \sum_{i=1..n} e_i^2$, $SS_{\text{total}} = SS_Y = \sum_{i=1..n} (y_i - \bar{Y})^2$, and $SS_{\text{model}} = SS_{\text{total}} - SS_{\text{error}}$. The proportion of the variance in Y accounted for by the overall model is given by $R^2 = SS_{\text{model}} / SS_{\text{total}}$. The contribution of any given predictor may be assessed by comparing the predictive value of a ‘full’ model, containing all predictors, to a ‘reduced’ model containing all predictors except the one of interest:

$$F(df_{\text{model[full]}} - df_{\text{model[reduced]}}, df_{\text{error[full]}}) = \frac{(SS_{\text{model[full]}} - SS_{\text{model[reduced]}}) \div (df_{\text{model[full]}} - df_{\text{model[reduced]}})}{SS_{\text{error[full]}} \div df_{\text{error[full]}}}$$

or, to make clear the equivalence to the logic discussed above,

$$F(df_{\text{predictor}}, df_{\text{error[full]}}) = \frac{SS_{\text{predictor}} \div df_{\text{predictor}}}{SS_{\text{error[full]}} \div df_{\text{error[full]}}} = \frac{MS_{\text{predictor}}}{MS_{\text{error[full]}}}$$

Effect sizes for individual predictors may be calculated in terms of R^2 (the proportion of variance in Y explained) or in terms of b (the change in Y for a given change in the predictor).

Viewing ANOVA in terms of a GLM makes its relationship to other well-known analytical techniques clear. For example, ANOVA with a single two-level factor is equivalent to a two-group t test; ANOVA with a single continuous predictor is equivalent to linear regression (Figure 1c); and so on. GLMs also subsume techniques such as analysis of covariance (ANCOVA) and multiple and polynomial regression, and can be extended to multiple dependent variables (multivariate ANOVA or MANOVA). General linear models may also be extended to dependent variables with non-normal (e.g. binomial) distributions via the *generalized linear model*.

CONTRASTS AND TREND ANALYSIS

GLMs may also be used to perform *contrasts* to ask specific questions of the data. In this technique, a linear combination of weighted means is created that will have a

value of zero under a given null hypothesis. For example, if a factor has 7 levels, one for each day of the week, then the linear contrast

$$L = 0.2\mu_{\text{Mon}} + 0.2\mu_{\text{Tue}} + 0.2\mu_{\text{Wed}} + 0.2\mu_{\text{Thu}} + 0.2\mu_{\text{Fri}} - 0.5\mu_{\text{Sat}} - 0.5\mu_{\text{Sun}}$$

(where μ_{day} indicates the day mean) may be used as a test of the null hypothesis that the dependent variable is equal on weekdays and weekends. In a GLM, these weights are encoded in a contrast matrix \mathbf{L} . After solving $\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, the contrast is calculated as $L = \mathbf{L}\mathbf{b}$ and assessed statistically to test the null hypothesis $L = 0$.

Trend analysis involves the use of contrasts to ask questions about categorical predictors (factors) that may be treated quantitatively. For example, if subjects' reaction times are tested with visual stimuli of length 9 cm, 11 cm, 13 cm, and 15 cm, then it may be valid to treat the lengths as categories (do reaction times to the lengths differ?) and quantitatively (is there a linear or quadratic component to the relationship between reaction time and stimulus length?).

CORRELATIONS BETWEEN PREDICTORS AND UNBALANCED DESIGNS

When an ANOVA design contains multiple predictors that are uncorrelated, assessment of their effects is relatively easy. However, it may be that predictors are themselves correlated. This may be because the predictors are correlated in the real world (for example, if age and blood pressure are used to predict some dependent variable, and blood pressure tends to rise with age). However, it may also occur if there are different sample sizes for different combinations of predictors. For example, if there are two factors, A (levels A_1 and A_2) and B (levels B_1 and B_2), then a *balanced* design would have the same number of observations of the dependent variable for each of the combinations A_1B_1 , A_1B_2 , A_2B_1 , A_2B_2 . If these numbers are unequal, the design is *unbalanced*, and this causes correlation between A and B. This problem also occurs in *incomplete factorial* designs, in which the dependent variable is not measured for all combinations of factors.

Whatever the reason for correlation between predictors, this causes a problem of interpreting their effects upon the dependent variable (Figure 2). There are various approaches to this problem; the best approach depends upon the class of hypothesis preferred on theoretical grounds (do some predictors take precedence over others?) and on whether the design is unbalanced or incomplete.

FIXED AND RANDOM EFFECTS

Up to this point, it has been assumed that factors have been *fixed effects*, meaning that the levels of the factor exhaust the population of interest and represent all possible values the experimenters would want to generalize their results to analytically; the *sampling fraction* (number of levels used \div number of levels in the population of interest) is 1. A simple example is sex: having studied male and female humans, all possible sexes have been studied. Experimental factors are usually fixed effects. It is also possible to consider *random effects*, in which the levels in the analysis are only a small, randomly selected sample from an infinite population of possible levels (sampling fraction = 0). For example, if an agronomist wanted to study the effect of fertilizers on wheat growth, it might be impractical to study all known varieties of wheat, so four might be selected at random to be representative of wheat in general. Wheat variety would then represent a random factor.

The most common random effect in psychology is that of subjects. When subjects are selected for an experiment, they are typically selected at random in the expectation that they are representative of a wider population. Thus, in psychology, discussion of fixed and random effects overlaps with consideration of between-subjects and within-subjects designs, discussed below. An ANOVA model incorporating both fixed and random effects is called a *mixed-effects model*.

With random effects, not only is the dependent variable a random variable as usual, but so is a predictor, and this modifies the analysis. It does not affect the partitioning of SS, but it affects the $E(\text{MS})$ values, and thus the choice of error terms on which F ratios are based. Regardless of the model, testing an effect in an ANOVA requires comparison of the MS for the effect with the MS for an error term where $E(\text{MS}_{\text{effect}}) = E(\text{MS}_{\text{error}})$ if the effect size is zero, and $E(\text{MS}_{\text{effect}}) > E(\text{MS}_{\text{error}})$ if the effect size is non-zero. In fixed-effects models, the error term is the ‘overall’ residual unaccounted for by the full model; in random-effects models, this is not always the case, and sometimes the calculation of an error term is computationally complex.

WITHIN-SUBJECT AND BETWEEN-SUBJECT PREDICTORS, AND MORE COMPLEX DESIGNS

Psychological experiments may be conducted with *between-subjects* factors, in which each subject is measured at just one level of the factor, or *within-subjects* factors, in

which each subject is measured at every level of the factor, or a mixture. Within-subjects designs (also known as *repeated measures* designs) allow considerable power, since an individual at one time is likely to be similar to the same individual at other times, reducing variability. The main disadvantage is the sensitivity of within-subjects designs to *order effects*.

In a simple within-subjects design, a group of subjects might be measured on three doses of a drug each (within-subjects factor: D). It would be necessary to counterbalance the testing order of the dose to avoid order effects such as improvement due to practice, or decline due to fatigue, or lingering or learned effects of a previous dose. Having done so, then variability between observations may be due to differences between subjects (S), or to differences between observations within one subject; this latter variability may be due to D, or intra-subject error variability (which might also include variability due to subjects' responding differently to the different doses, written $D \times S$).

For a design involving both between- and within-subjects factors, suppose old subjects and young subjects are tested on three doses of a drug each. Age (A) is a between-subjects factor and dose (D) is a within-subjects factor. In this example, variability between subjects may be due to A, or to differences between subjects within age groups, often written S/A ('subjects within A') and thought of as the between-subjects error. Variability of observations *within* individual subjects may be due to D, or to a $D \times A$ interaction, or within-subject error that includes the possibility of subjects responding differently to the different doses (which, since subjects can only be measured within an age group, is written $D \times S/A$). This partitioning of total variability may be accomplished for SS and *df* and analysed accordingly, with the caveat that as subject (S) represents a random factor, calculation of an appropriate error term may sometimes be complex, as described above.

Within-subjects designs also carry an additional assumption: that of *sphericity*, or homogeneity of variance of difference scores. For example, if subjects are measured at three levels of a within-subjects factor U, then three sets of difference scores can be calculated: $(U_1 - U_2)$, $(U_1 - U_3)$, and $(U_2 - U_3)$. Sphericity is the assumption that the variances of these three difference scores are the same. Violations of this assumption can inflate the Type I error rate. Violations may be tested for directly using Mauchly's test, but this may be over-conservative. Approaches to this problem include (1) using

an F test as usual but correcting the df to allow for the extent of violation, such as with the Huynh–Feldt or Greenhouse–Geisser corrections; (2) transformation of the data to improve the fit to the assumptions, if this is possible and meaningful; (3) using MANOVA, which does not require sphericity; (4) testing planned contrasts of interest, which have 1 df each and therefore cannot violate the assumption.

ANOVA designs may have arbitrary numbers of between- and within-subjects factors. In these designs, ‘Subject’ defines a type of relatedness between observations. More complex designs, involving multiple levels of ‘relatedness’, are possible and indeed commonly reflect real-world experiments. For example, suppose schoolchildren (S) each perform four tests of four levels of difficulty (factor D). These children are taught in class groups (G), and sets of groups are taught according to different teaching methods (T). In this example, D is said to be *crossed* with S (since all subjects perform in all difficulty conditions), but S is *nested* within G (any given subject is only a member of one group), and groups are nested within T. This model may be written DS/G/T, and is an example of a *hierarchical* design involving two types of ‘relatedness’. Such designs allow the analytical model to reflect correlations in the real world—such as observations that come from the same subject being likely to be more similar to each other than to randomly selected observations, or observations that come from different subjects who are in the same group being likely to be more similar to each other than to randomly selected observations. Ensuring a correct design is part of satisfying the assumption that the ANOVA model accounts for all systematic influences on the data.

REFERENCES

Cardinal, R. N., & Aitken, M. R. F. (2006). *ANOVA for the behavioural sciences researcher*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, *52*, 399–433.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd.

Howell, D. C. (2007). *Statistical Methods for Psychology*, sixth edition. Belmont, California: Thomson/Wadsworth.

Langsrud, Ø. (2003). ANOVA for unbalanced data: use type II instead of type III sums of squares. *Statistics and Computing*, *13*, 163–167.

Maxwell, S. E., & Delaney, H. D. (2003). *Designing Experiments and Analyzing Data: A Model Comparison Perspective*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Myers, J. L., & Well, A. D. (2003). *Research Design and Statistical Analysis*, second edition. Mahwah, New Jersey: Lawrence Erlbaum Associates.

StatSoft (2008). *Electronic Statistics Textbook*. Tulsa, Oklahoma: StatSoft. URL: <http://www.statsoft.com/textbook/stathome.html>.

Verrill, S. (2001). Rolling your own: linear model hypothesis testing and power calculations via the singular value decomposition. *Statistical Computing and Statistical Graphics Newsletter*, *12(1)*, 15–18.

SUGGESTED READING

Myers, J. L., & Well, A. D. (2003). *Research Design and Statistical Analysis*, second edition. Mahwah, New Jersey: Lawrence Erlbaum Associates.

StatSoft (2008). *Electronic Statistics Textbook*. Tulsa, Oklahoma: StatSoft. URL: <http://www.statsoft.com/textbook/stathome.html>.

In addition to commercial statistical software, open-source software suitable for ANOVA is available with *R* (R Development Core Team, 2008; *R: A language and environment for statistical computing*; Vienna, Austria: R Foundation for Statistical Computing; <http://www.R-project.org>) and on-line guides are available (see e.g. <http://www.psychol.cam.ac.uk/statistics/R/>).

Table 1. ANOVA of fictional data (shown with scale removed in Figure 1b), showing conventional table style as might be produced by statistical software. In general, a factor with k levels has $(k - 1)$ degrees of freedom (df). A linear predictor has 1 df , as does a linear contrast. An interaction $A \times B$, where A has a df and B has b df , has ab df . In this example, there are two factors, each with two levels each. If there are N observations of the dependent variable in total, the total df is $(N - 1)$. In this example, $N = 24$. The total line is not always shown; MS_{total} is the variance of the dependent variable. SS, sum of squares; MS, mean square.

Term	df	SS	MS	F	p
A	1	2,795.04	2,795.04	17.738	0.0004289
B	1	1,650.04	1,650.04	10.472	0.0041415
$A \times B$	1	2,147.04	2,147.04	13.626	0.0014463
Residual (error)	20	3,151.50	157.57		
Total	23	9,743.62	423.64		

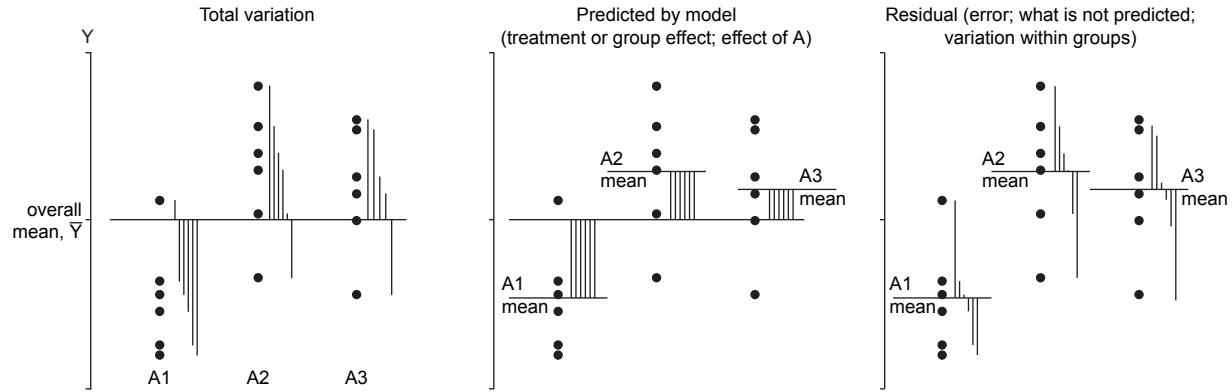
Figure 1: Illustration of simple ANOVA types. A: One-way ANOVA, illustrated with a dependent variable Y and a single factor A having three levels (A_1, A_2, A_3). In all figures, vertical lines indicate deviations that are squared and added to give a sum of squares (SS). The total sum of squares of Y (the summed squared deviation of each point from the grand mean; left panel) is divided into a component predicted by the factor A (the summed squared deviations of the A subgroup means from the grand mean, for each point; middle panel) plus residual or error variation (the summed squared deviation of each point from the prediction made using A ; right panel). The SS are additive: $SS_{\text{total}} = SS_A + SS_{\text{error}}$. When SS_A and SS_{error} have been divided by their corresponding degrees of freedom (the number of independent pieces of information associated with the estimate), they may be compared statistically: if SS_A / df_A is large compared to $SS_{\text{error}} / df_{\text{error}}$, then A is a good predictor. **B:** Two-way ANOVA. Here, two factors A and B are used, each with two levels. From left to right, panels illustrate the calculation of SS_{total} , SS_A , SS_B , the interaction term $SS_{A \times B}$, and SS_{error} . As before, the SS are additive ($SS_{\text{total}} = SS_A + SS_B + SS_{A \times B} + SS_{\text{error}}$) if the predictors are not correlated (see text). **C:** Regression ANOVA, in which Y is predicted by a single continuous variable X . As before, the total SS (left panel) may be divided into a component predicted by the model (middle panel) and residual or error variation (right panel).

Figure 2: Assessing the effects of correlated predictors. Suppose a dependent variable Y is analysed by two-way ANOVA, with predictors A , B , and their interaction ($A \times B$ or AB). The total variability in Y is represented by the sum of squares (SS) of Y , also written SS_{total} or SS_Y . This may be partitioned into SS attributable to A , to B , to the interaction AB , and to variability not predicted by the model (SS_{error}).

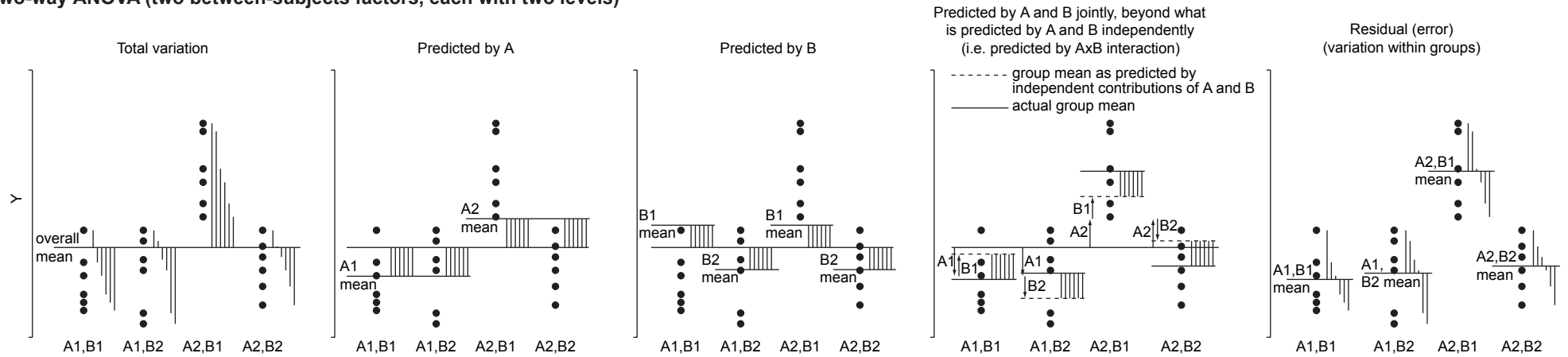
A: If the predictors are uncorrelated, then the SS are orthogonal, and are additive: $SS_{\text{total}} = SS_A + SS_B + SS_{AB} + SS_{\text{error}}$. ANOVA calculations are straightforward.

B: If the predictors are correlated, then the SS are not additive. There are various options for assessing the contribution of the predictors. For example, one possibility is to assess the contribution of each predictor over and above the contribution of all others; thus, the contribution used to assess the effect of A would be t , that for B would be x , and that for AB would be z (often termed ‘marginal’, ‘orthogonal’, or ‘Type III’ SS). A second possibility is to adjust the main effects for each other (i.e. not to include any portion of the variance that overlaps with other main effects), but not for the interaction (so SS_A is $t + w$, SS_B is $x + y$, and SS_{AB} is z , often known as ‘hierarchical’ or ‘Type II’ SS). Stated more generally, this approach adjusts terms for all other terms except higher-order terms that involve the same predictors. Alternatively, if there are reasons to treat the predictors as a hierarchy with A taking precedence, then one could assess the contribution of A ($t + u + v + w$), to assess the contribution of B above this ($x + y$), and then to assess the contribution of the interaction above these (z) (often termed ‘sequential’ or ‘Type I’ SS). These and other approaches to this problem have different advantages and disadvantages according to the type of design being analysed. In practice, unless the sequential approach is desired, Type III SS are often used but the best approach is debated. All represent ways of comparing pairs of models of the data, as described in the text, to assess whether the difference between the models is significant; the experimental hypotheses determine the comparisons to be made.

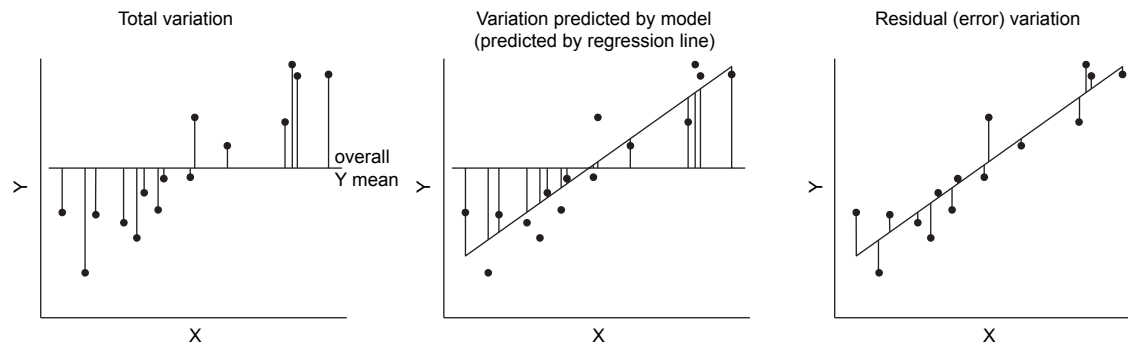
A. One-way ANOVA (one between-subjects factor with three levels)



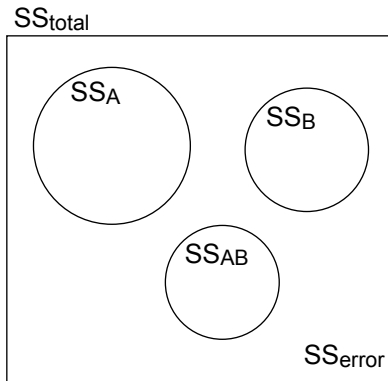
B. Two-way ANOVA (two between-subjects factors, each with two levels)



C. Regression ANOVA (linear regression; one continuous predictor)



**A. Uncorrelated predictors
(orthogonal sums of squares)**



**B. Correlated predictors
(nonorthogonal sums of squares)**

