ANOVA & ANCOVA.

1. Introduction.

The term **ANOVA** comes from ANalysis Of Variance, and refers to a well established technique for comparing a number of means, based on a sample of observations. This is done by essentially comparing *between group* variances with *within group* variances.

1.1. Assumptions.

- Expected values of the errors are zero.
- Equality of error variances.
- Independent errors.
- Normally distributed errors.

2. One-Way ANOVA

The one-way ANOVA measures the effect of one independent variable *(with more than two levels say)* on a dependent variable.

- Total variability comes from:
  - Differences between groups.
  - Differences within groups.

The *traditional* layout of the ANOVA table is given below.

Source	SS	DF	MS	F-ratio
Between groups	$m\Sigma_{i=1}^{k}(Y_{i.} - Y_{..})^2$	k-1	$MS_{Groups}=\frac{SS_{groups}}{k-1}$	$\frac{MS_{Groups}}{MS_{Error}}$
Within groups *or* Error *or* Residual	$\Sigma_{i=1}^{k}\Sigma_{j=1}^{m}(Y_{ij} - Y_{i.})^2$	k(m-1)	MS$_{Error} = \frac{SS_{error}}{k(m-1)}$	
Total (corrected)	$\Sigma_{i=1}^{k}\Sigma_{j=1}^{m}(Y_{ij} - Y_{..})^2$	mk-1		

TABLE 1. Analysis of Variance Table

Where Y_{ij} denotes the jth observation in the ith group $\{j = 1 : m; i = 1 : k$, $Y_{..}$ is the overall mean, and $Y_{i.}$ is the average of the ith group.

To test the hypothesis $H_0 : \mu_1 = \mu_2 = ... = \mu_k$ the F-ratio is calculated as shown in the table and H_0 is rejected if this value is greater than the appropriate upper percentage point of the F-distribution having $(k-1)$ and $k(m-1)$ degrees of freedom, i.e. reject H_0 if the p-value is less then the significance level $\{$ usually 0.05$\}$. Hence if the F-statistic falls to the right of the *critical value* denoted by the red line in figure 1 we reject the null hypothesis. However in practice one would use a statistical computing software package to calculate the quantities required, the layout of the output is very similar.

3. Two-Way ANOVA

A two-way ANOVA measures the effect of two independent variables on the dependent variable, and is just a simple extension of the one-way ANOVA.

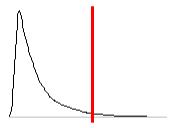

FIGURE 1. F-distribution

4. **Factorial ANOVA.**

Factorial ANOVA measures whether a combination of independent variables predict the value of a dependent variable, where the model variability comes from;

- Main effects
  - Mean differences among the levels of a particular factor.
- Interaction
  - Where the differences among cell means are not a consequence of the main effects.
  - When the effect of one factor on the dependent variable is influenced by the levels of another. We are attempting to answer if the effect of one factor depends on the level of the other factor?

5. **Repeated measures ANOVA.**

Repeated measures ANOVA also tests the equality of means, however it is used when all members of a random sample are measured under a number of different conditions. The measurement of the dependent variable is repeated, as the sample is exposed to each condition in turn.
A standard ANOVA is not appropriate as it fails to account for the correlation between the repeated measures hence the data violate the ANOVA assumption of independence.
Repeated measures ANOVA can also be used when sample members have been matched according to some important characteristic. For example, if we select a group of subjects suffering from some condition and measure some characteristic relating to the condition , we then match subjects into pairs having similar levels of the measured characteristic. One subject from each matching pair is then given a treatment , and afterwards the level of the measured characteristic of the entire sample is measured again.

6. **General Strategy**

- Set up a model.
- Find (unbiased) estimators for the parameters.
- Partition the total sum of squares and construct an appropriate ANOVA table.
- For a given data set, examine the residuals and fitted values (*if desired.*)
- Make inference as desired.

7. **Post-Hoc analysis.**

If the F-test suggests that all means are not the same then we need to find which group(s) are different. To do this one might consider either the **Bonferroni** method or **Tukey's** method, which are both multiple comparison methods.

8. **ANCOVA.**

The Analysis of Covariance (ANCOVA) essentially sits between analysis of variance (ANOVA) and regression analysis, and allows us to compare one variable in 2 or more groups taking into account the variability of other variables. It is used when:

- There are variables that might relate to the dependent variable and we want to partition out their variance from the residual variance, this leads to more statistical power.

8.1. **Assumptions.**

- Normality
- Linearity
- Homoscedasticity
- Homogeneity of regression, i.e. (Slopes between covariate and the dependent variable are similar across groups):