Overview of Lecture

- Last Week
- Per comparison and familywise error
- Post hoc comparisons
- Testing the assumptions of ANOVA
- Summary of a one-way between groups ANOVA
- Using SPSS to conduct a one-way between groups ANOVA

Last Week - Analysis of Variance

- A one-way between groups ANOVA conducted on:
  - IV - three lecturing styles (each assigned 5 students)
  - DV - exam score (0….20)
- Results
  - $F_{1,12}$=7.41, MSe=14.17, p<.01

Lectures	Worksheets	Both
6 (1.41)	9 (1.87)	15 (1.73)

Table 1: The means (and standard errors) of the exam scores for the three different teaching styles

Last Week - Planned comparisons

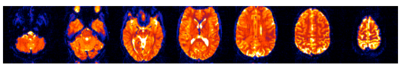
- Before we set out to collect the data, we made specific predictions about the direction of the effects
  - used a technique known as planned (a priori) comparisons.
- Tested the prediction that lectures+worksheets would produced better performance on the exam than worksheets alone
- The result was that lectures+worksheets did indeed lead to better performance on the exam ($F_{1,12}$=14.29, MSe=14.17, p<0.01)

Per Comparison and Familywise Error

- A Type I error has been defined as the probability of rejecting the null hypothesis when in fact the null hypothesis is true.
- This applies to every statistical test that we perform on a set of data.
- If we perform several statistical tests on a set of data we can effectively increase the chance of making a Type I error.

An example of familywise error

- fMRI data:
  - often 64x64x64 voxels
  - Chance of one of these voxels being active at the 0.05 level is very high.
  - By chance, we expect 13,107 voxels at 0.05!
  - How can we control for Type I errors?



Per comparison and familywise error rates

- If we perform two statistical tests on the same set of data then we have a range of opportunities of making a Type I error.
  - Type I error on the first test only
  - Type I error on the second test only
  - Type I error on both the first and the second test
- Type I errors involving single tests are known as per comparison errors
- The whole set of Type I errors above is known as the familywise error.

Per comparison and familywise error rates

- The relationship between the two error rates is very simple:

$$\alpha_{fw} = c(\alpha_{pc})$$

  - where c is the number of comparisons.
- So if we have made three comparisons, we can expect 3*(0.05) = 0.15 errors. If we make twenty comparisons, we will on average make one error [20*0.05=1.0].
- Of course, if we make twenty comparisons, it is possible that we may be making 0, 1, 2 or in rare cases even more errors.
- The chance we will make at least one error is given by the formula: $1- (1-\alpha)^C$.
- So if we make twenty comparisons at $p < .05$, we have a $1-(.95)^{20}$ = 0.64% chance that our we are reporting at least one erroneous findings.

Type I error rates and analytical comparisons

- With planned comparisons :
  - Ignore the theoretical increase in familywise type I error rates and reject the null hypothesis at the usual per comparison level.
- With post hoc or unplanned comparisons between the means we cannot afford to ignore the increase in familywise error rate.

Post hoc analytical comparisons

- A variety of different post hoc tests are commonly used - for example
  - Scheffé
  - Tukey HSD
  - t-tests
- These tests vary in their ability to protect against Type I errors.
- Increasing Type I protection reduces Type II protection.

The Scheffé test

- The Scheffé is calculated in exactly the same way as a planned comparison
- Scheffé differs in terms of the $F_{Critical}$ that is adopted.
- For the one-way between groups analysis of variance the critical F associated with an $F_{Scheffé}$ is given by:

$$F_{Scheffe} = (a-1)F_{(df_A, df_{S/A})}$$

  - where *a* is the number of treatment levels and $F_{(dfA, dfS/A)}$ is the critical value of F for the overall, omnibus analysis of variance.
- For our example
  - Omnibus ANOVA critical value F(2,12)= 3.885. There were three treatment levels so (3-1)*3.885= 7.77.
  - $F_{observed}$ = 14.29 when comparing lectures+worksheets with lectures alone

Tukey HSD

- The Tukey (Honestly Significant Difference) test establishes a value for the smallest possible significant difference between two means.
- Any mean difference greater than the critical difference is significant
- The critical difference is given by:

$$D = q_{(\alpha, df, a)}\sqrt{\frac{MS_{Error}}{n}}$$

  - where q(α,df,a) is found in tables of the studentized range.
- This particular formula only works for between groups analysis of variance with equal cell sizes
- A variety of different formulae are used for different designs

t-tests

- When comparing two means, a modified form of the t-test is available.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{2MS_{Error}}{n}}}$$

- For multiple comparisons the critical value of t is found using
  - p=0.05/c
  - where c is the number of comparisons.
- This is known as a *Bonferroni correction*

Post hoc tests

- Post-hoc tests are conservative – they reduce the chance of type I errors by greatly increasing type II errors.
- Only very robust effects will be significant.
- Null results using these tests are not easy to interpret.
- Many different post hoc tests exists and have different merits and problems
- Many post hoc tests are available on computer based statistical packages (e.g. SPSS or Experstat)

The assumptions of the F-ratio

- Independence
  - The numerator and denominator of the F-ratio are independent
- Random Sampling
  - Observations are random samples from the populations
- Homogeneity of Variance
  - The different treatment populations have the same variance.
- Normality
  - Observations are drawn from normally distributed populations

Testing Assumptions of Anova

- Each of these assumption should be met before progressing onto the analysis.
- There are two assumptions that we have to assume have been met by the experimenter
  - Independence and Random Sampling
  - If an experiment has been designed appropriately both of these assumptions will be true.
- Both the homogeneity of variance and the normality assumptions need not necessarily be true.

Testing Homogeneity of Variance

- When looking a between groups designs use
  - Hartley's F-max
  - Bartlett
  - Cochran's C
- When looking at within or mixed designs use
  - Box's M
- All these tests are sensitive to departures from normality
- All of these tests are available in SPSS (as are a number of other tests)

Testing Homogeneity of Variance - A heuristic

- For hand calculations, there is a quick and dirty measure of homogeneity of variance:

$$\frac{\text{largest variance}}{\text{smallest variance}} < 4$$

- Note: this is a heuristic. When you have the option, use one of the specific tests (e.g. Bartlett).

Testing normality

- The three most commonly used tests for normality are:
  - Skew
  - Lilliefors
  - Shapiro-Wilks
- These tests compare the distribution of the data to a theoretically derived normal distribution.
- All these tests are very sensitive to departures from normality when there are large samples.
- The Lilliefors and Shapiro-Wilks are difficult to calculate by hand, but both are available on SPSS.

Testing normality by examining skew

- Since we assume
  - that the distributions of the population from which the samples are taken are normal
  - and the skew of a normal distribution is equal to zero
- Then
  - One test of normality is to see if the skew is significantly different to zero
- In other words, test the value of skew to see if it deviates significantly from a normal distribution.

Testing skew

- The simplest test we can use is a z-score. In the case of skew the z-score is given by:

$$z = \frac{skew - 0}{SE_{skew}}$$

- The standard error of skew is given by

$$SE_{skew} = \sqrt{6/N}$$

  - where N is the number of cases in the sample.
- If a z score associated with the skew is greater than |±1.96| then the sample is significantly different from normal.
- In other words, a value of skew which is significantly different from zero, would mean that we do not have normally distributed data

Data transformations

- What can we do in order to meet the assumption of the analysis of variance?
- In order to return our data to normality and establish homogeneity of variance we can use transformations.
  - These are simply mathematical operations that are applied to the data before we conduct an analysis of variance.
- However, there are three circumstances where no transformation to the data will work:
  - Variance are heterogenous
  - Distributions are heterogenous
  - Variances are heterogeneous and distributions are heterogeneous

Data transformations

- The following table shows the kinds of transforms that we can use
- They depend on the amount of skew in the data

	Moderate $1.96 \leq z \leq 2.33$	Substantial $2.34 \leq z \leq 2.56$	Severe $z > 2.56$
Positive Skew	Square Root	Logarithm	Reciprocal
Negative Skew	Square Root (K-X)	Logarithm (K-X)	Reciprocal (K-X)

- Where K is the largest number in the data set plus 1

Transforming data

- Transforming data reduces the probability of making a type II error
  - A type II error occurs when we fail to reject the null hypothesis when it is false
  - If an assumption is broken, ANOVA fails gracefully: we will miss real effects (type II) but we will not increase our rate of making claiming effects that do not exist (type I)
- Data should be transformed when either the data is not homogenous or not normal
- Solving the homogeneity problem often solves the normality problem and vice versa

Transforming data

- What happens when transforming the data is impossible?
- In general we proceed with the analysis but advise caution to the reader when reporting the results
- This is particularly important if the observed F value has an associated probability, p, such that

$$0.1<p<0.01$$

- In these circumstances it is difficult to know whether a type I error or a type II error is being made or if no error is being made at all.