• Stem-and-leaf diagram - vertical numbers on far left represent the 10s, numbers right of the line represent the 1s
• The mean should not be used if there are extreme scores, or for ranks and categories
• Unbiased estimator - The best guess which is just as likely to be too high as too low, if obtained repeatedly, on average it will be correct

Lecture 3

• Density - An alternative vertical scale for histograms, which ensures that the total area covered by the histogram is 1 square unit (Initial area = 10 x 10 = 100, new area = 10 x 0.1 = 1, convert the vertical scores on the scale by dividing each one by the value of the initial area)
• The area for a region of a histogram can be used to tell you the proportion of scores in that region
• The normal distribution - Bell shaped curve, it is the idealised distribution, it is symmetrical, the mean, mode and median are all the same value, scores close to the mean are much more common, and become increasingly uncommon the further away from the mean, the proportion of scores above or below a certain value can be calculated when you know the mean and standard deviation of a normally distributed variable.
• The approximate proportion of scores for a normally distributed variable lie within 1, 2, or 3 standard deviations of its mean.
• Standard normal distribution - As a result of the special properties of the normal distribution we can use standard deviations as a unit of measurement to describe how far above or below the mean a particular score is, +1 means 1 standard deviation greater than the mean (and so on...), These are called z-scores (a z-score of zero indicates the mean).

Lecture 4
\[ z = \frac{X - \mu}{\sigma} \] - Converting individual values into z-scores

\[ z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{N}}} \]

Or

\[ \frac{\bar{X} - \mu}{S.E.\text{Mean}} \] - The z-test of whether a sample mean differs significantly from a population mean

How many standard errors separate the sample mean from the population mean

**One Sample t-test**

\[ t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{N}}} \]

Degrees of freedom = N - 1

**Statistical Hypotheses** - Logical process of performing an inferential test, null/alternative hypothesis is a statement about populations that could be true, given the null hypothesis, if the observed statistics are improbable we can reject the null in favour of the alternative hypothesis

**Study/Experimental/Research Hypotheses** - Practical endeavours of empirical exploration, what would we expect to observe in a study if one particular theory is true, there are often competing hypotheses if there is more than one theory to test

**One-tailed/Two-tailed hypotheses**

One-Tailed - Directional - Mean1 is bigger/smaller than Mean2

Two-Tailed - Non-directional - Mean1 does not equal Mean2

Two-tailed should be used as often as possible, a directional hypothesis should not require a two-tailed test,

---

**Lecture 13**

**Independent Samples** - Two separate groups, no one is in both groups

**Related Samples** - One group in both conditions, each participant has a direct counterpart in both groups

**To decide between the two** - Consider two samples A and B, write A in a list, then B, if the order you put B in a list depends on the order of A, then the sample is related
Factors that affect the Power

i. The probability of a Type 1 error
ii. The true difference between the Null hypothesis and the actual state of affairs
iii. The sample size and population variance or SD
iv. One-tailed/Two-tailed
v. Parametric/Non-parametric
vi. Paired t-test is more powerful than independent t-tests
vii. Correlation of the two groups in a paired t-test

**Measures of Effect Size**

**Mean Difference** - With very large samples we can reliably detect small differences

**The Linear Relationship Between Two Variables** - With very large samples we can reliably detect small relationships

Cohen's d - $\frac{\mu_1 - \mu_2}{\sigma}$

Estimated Effect Size - $\frac{\bar{X}_1 - \bar{X}_2}{S}$

**Converting d into t and r**

\[ r = \frac{d}{\sqrt{d^2 + 4}} \]

\[ r = \frac{t}{\sqrt{df} \times \sqrt{1 + t^2 / df}} \]

**Phi** - The measure for effect size for association for a 2-by-2 contingency table

0.1 - Small

0.3 - Medium

0.5 - Large

Effect Size and Significance are related

Magnitude of a Significance Test = Effect Size X Size of Study

**Lecture 19**

**Confidence Interval** - A range estimate that is commonly used in Psychological research, to communicate the degree of uncertainty
\[
\sqrt{(a - 1) \times (Fcrit)}
\]

\[
W = q_{(r, \text{df ERROR})} \sqrt{\frac{\text{Mean Square Error}}{N_A}}
\]

\[
F = \frac{\text{Between Group Variance}}{\text{Residual Variance}}
\]

\[
SS_S = \frac{(\sum S_1)^2 + (\sum S_2)^2}{N_S} - \frac{(\Sigma Y)^2}{N}
\]

\[
df_{RESIDUAL} = (a - 1) \times (s - 1)
\]

\[
F = \frac{\text{Mean Square}_A}{\text{Mean Square}_{SXA}}
\]

**Between Group Variance:**

The variance of each mean from the grand mean:

\[
\text{Between Group Variance} = \frac{N_1(A_1 - \bar{Y})^2 + N_2(A_2 - \bar{Y})^2 (\text{And so on ...})}{\text{Between Groups Degrees of Freedom}}
\]

If each group is evenly sized, the equation can be rearranged thusly:

\[
\text{Between Group Variance} = \frac{(\Sigma A_1)^2 + (\Sigma A_2)^2}{N_A} - \frac{(\Sigma Y)^2}{N}
\]

Because:

\[
SS_{BETWEEN} = \text{Between Group Variance} \times \text{Between Groups Degrees of Freedom}
\]

We can say that:

\[
SS_{BETWEEN} = \frac{(\Sigma A_1)^2 + (\Sigma A_2)^2}{N_A} - \frac{(\Sigma Y)^2}{N}
\]

**Within-Group Variance:**
If you have two levels, and you know the grand mean one level is free to vary but once a value exists for that the remaining level becomes fixed to enable the grand mean to stay at its current value, hence we have a between group degrees of freedom of 1.

**Within group degrees of freedom:**

\[ df_{WITHIN} = a(NA - 1) \]

If we have 10 participants in a sample and we know the mean, 9 scores are free to vary but once they gain a set value the remaining score's value has to be fixed to remain consistent with the value of the mean, hence we have a degrees of freedom of 9. If we have three levels, each with 10 participants, the the degrees of freedom becomes 9 + 9 + 9, hence we have a within groups degrees of freedom of 27.

**Total degrees of Freedom:**

\[ df_{Total} = df_{BETWEEN} + df_{WITHIN} \]

This because:

\[ SS_{TOTAL} = SS_{WITHIN} + SS_{BETWEEN} \]

Mean Square and the F-ratio

The mean square is calculated by taking each Sum of Squares and dividing it by its corresponding degrees of freedom. Mean square is essentially the variance, mean square error:

\[ Mean\ Square_{A} = \frac{ss_{A}}{df_{A}} \]

\[ Mean\ Square_{S/A} = \frac{ss_{S/A}}{df_{S/A}} \]

The ratio for F is simply between subjects mean square divided by within-subjects mean square:
• ... X influences Y  
• ...Y influences X  
• ... X and Y influence each other  
• ... A third variable influences both X and Y

We can use MLR to examine the influence of variables while controlling for the others

Variable X $\rightarrow$ Variable Y

Does X cause Y, or is there a mediator value Z, that is caused by X and causes Z. If Z correlates with both, this provides support for this idea. We can use multiple linear regression to control for variable Z, to investigate the relationship between X and Y, without the possible mediating effect of Z (with those who have similar levels of Z, do higher levels of X still predict higher levels of Y).

Possible relations:

• ... Z could be the sole cause
• ... Z could be a partial cause
• ... Z could be an unrelated cause

If Z is the sole cause, when we hold constant there should be no relationship between X and Y, if Z is a partial cause when we hold constant the relationship between X and Y should be reduced.

Steps:
1. In initial regression X was found to be a predictor of Y
2. Enter Z into the regression
3. If beta weight now drops, but remains significant - There is a shared variance

$R^2$ Change gives us an indication of how much of the variance is accounted for before and after controlling for a variable

F Change gives you an indication of the change in significance when accounting for a specific variable

Hierarchical Regression - Automated way of controlling for variables in a number of stages,

Indirect Mediator - X affects Z, Z affects Y, manifests originally as X affecting Y

Partial Mediator - X affects Y a small amount, X also affects Z, Z affects Y, both predictors affect Y, but if only one is investigated it may appear direct

Baron and Kenny (1986) -

1. Significant relationship between IV and mediator (a)
Mean Square Error is the within subjects mean square, to work out the mean square divide sum of squares by degrees of freedom, to work out F ratio divide between-subjects mean square and the mean square error.

\[ df_A = \text{Number of Levels} - 1 \]

\[ df_{S/A} = \text{Number of Levels} \times (\text{Number of PPs in level A}) \]

### Single Factor Within-Subjects ANOVA

\[
\text{Between} - \frac{(\sum A_1)^2 + (\sum A_2)^2}{N_A} - \frac{\left(\sum Y\right)^2}{N} \quad \text{Subjects} - \frac{(\sum S_1)^2 + (\sum S_2)^2}{N_S} - \frac{\left(\sum Y\right)^2}{N}
\]

### Two Factor Fully Between-Subjects ANOVA

\[
\text{Between Factor A} - \frac{(\sum A_1)^2 + (\sum A_2)^2}{N_A} - \frac{\left(\sum Y\right)^2}{N} \quad \text{Between Factor B} - \frac{(\sum B_1)^2 + (\sum B_2)^2}{N_B} - \frac{\left(\sum Y\right)^2}{N}
\]
- Newman-Keuls Test - Protected, hence F value must be significant, requires 5 or fewer means, cannot be used if means are identical
- t-test - Can be used as a post hoc test, has to be used with the Bonferroni adjustment \( \frac{Old \ Stringent \ Level}{Number \ of \ Comparisons \ Performed} \) to correct for familywise errors, very stringent and can lack statistical power, F does not need to be significant
- Planned Comparison - Uses the t-test model, only crucial comparisons needed should be performed, no need for significant F, when using few comparisons (number of levels - 1) test is powerful and requires no correction, when using many comparisons (greater than number of levels - 1) the Bonferroni adjustment should
- Scheffé - Most stringent of all, uses a t-test but corrects the critical value, used after planned comparisons

Equations:

Bonferroni - \( \frac{0.05}{\text{Number of Comparisons}} \)

Tukey - \( W = q_{(\text{df ERROR})} \sqrt{\frac{\text{Mean Square Error}}{\text{Number of Scores that Make up Factor A}}} \)

Scheffé - \( \sqrt{(\text{number of levels of A} - 1) \times (\text{df ERROR})} \)

t-test - \( t = \frac{A_1 - A_2}{\text{(Mean Square Error)\times} (\frac{1}{N})} \)

Newman-Keuls - N.B. Cannot be completed by hand

Basic Notation

\([A] = \frac{(\Sigma A_1)^2 + (\Sigma A_2)^2}{N_A} = \frac{\text{Total Score of Level One A Squared} + \text{Total Score of Level Two A Squared}}{\text{Number of Scores that Make up factor A}}\)

\([T] = \frac{(\Sigma Y)^2}{N} = \frac{\text{Sum of all scores regardless of condition squared}}{\text{Total Number of Participants}}\)

\([Y] = \Sigma Y^2 = \text{Sum of each squared score}\)

\([S] = \frac{(\Sigma S_1)^2 + (\Sigma S_2)^2}{N_S} = \frac{\text{Total Score for Subject One Squared} + \text{Total Score for Subject Two Squared}}{\text{Number of Scores that Make up Subject Total}}\)
> Complete Fully Between-Subjects ANOVA
  > Effect Details
    > (IV1) * (IV2)
      > LSMeans Plot

Stacking Data
> Tables
  > Stack
    > Highlight (WSF1) and (WSF2)
      > Type file name of choice
    > Overtype Data with data name of choice
      > Overtype Label with factor name of choice
    > OK

Mixed Design ANOVA
> Stack Data
  > Analyse
    > Fit Model
      > Click (DV) into Y
      > Add (IV1) into Construct Model Effects
      > Add (IV2) into Construct Model Effects
      > Highlight (IV1) and (IV2) click Cross to add to Construct Model Effects
    > Highlight (PP), select Random Effect from Attributes
      > Change Emphasis to Minimal Report
    > Run Model

Fully Between Subjects ANOVA with Three Factors
> Fit Model
  > Click (DV) into Y
    > Add (IV1) into Construct Model Effects
    > Add (IV2) into Construct Model Effects
    > Add (IV3) into Construct Model Effect
      > Highlight both (IV1) and (IV2)
      > Cross
      > Highlight both (IV1) and (IV3)
Fully within-subjects ANOVA with three factors

> Run Model

> Fit Model
  > Click (DV) into Y
  > Add (IV1) into Construct Model Effects
  > Add (IV2) into Construct Model Effects
  > Add (IV3) into Construct Model Effects
  > Add (PP) into Construct Model Effects
    > Highlight both (IV1) and (IV2)
    > Cross
    > Highlight both (IV1) and (IV3)
    > Cross
    > Highlight both (IV2) and (IV3)
    > Cross
    > Highlight (IV1), (IV2) & (IV3)
    > Cross
    > Highlight (PP)

> Random Effects from Attributes
  > Highlight both (PP&R) and (IV1)
  > Cross
  > Highlight both (PP&R) and (IV2)
  > Cross
  > Highlight both (PP&R) and (IV3)
  > Cross