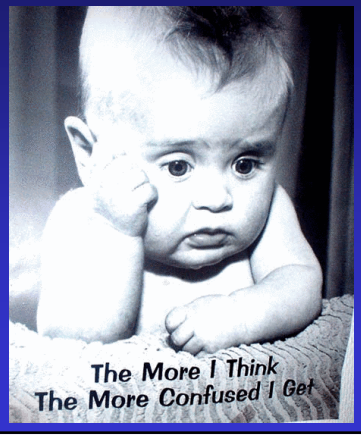
Deciding which statistical test to use:



Tests covered on this course:

(a) Nonparametric tests:

Frequency data -
Chi-Square test of association between 2 IV's (contingency tables)
Chi-Square goodness of fit test

Relationships between two IV's -
Spearman's rho (correlation test)

Differences between conditions -
Wilcoxon (repeated-measures, two conditions)
Friedman's (repeated measures, 3 or more conditions)

Mann-Whitney (independent measures, two conditions)
Kruskal-Wallis (independent measures, 3 or more conditions)

(b) Parametric tests:

z-scores (one score compared against the distribution of scores to which it belongs)

Relationship between two IV's -
Pearson's r (correlation test)

Differences between conditions -

Repeated-measures t-test (repeated measures, two conditions)
One-way repeated-measures ANOVA (repeated measures, 3 or more conditions)

Independent-measures t-test (independent measures, two conditions)
One-way independent-measures ANOVA (independent measures, 3 or more conditions)

Questions to ask yourself, in order to decide which test is appropriate for a set of data:

1. How has each participant contributed to the data?

One or more **scores** from each participant?

Or are they merely **frequency** data (i.e., the number of participants obtaining some result)?

2. If scores, then what kind of scores are they?

Are they "real" numbers (interval or ratio data) or ranks (ordinal data)?

4. Looking for similarities (relationships) or differences?

Correlations vs other tests

3. How many Independent and Dependent Variables?

One IV and one DV? (t-test, Mann-Whitney, Wilcoxon, ANOVA, Friedman's, Kruskal-Wallis)

Two IV's, with scores for each? (Correlation)

One or two IV's, and frequencies of occurrence? (Chi-Square)

5. Independent measures or repeated measures?

Does each participant do just one experimental condition, or do they perform more than one?

6. Do the data meet the requirements for a parametric test?

Interval or ratio data

Normally distributed

Similar variances

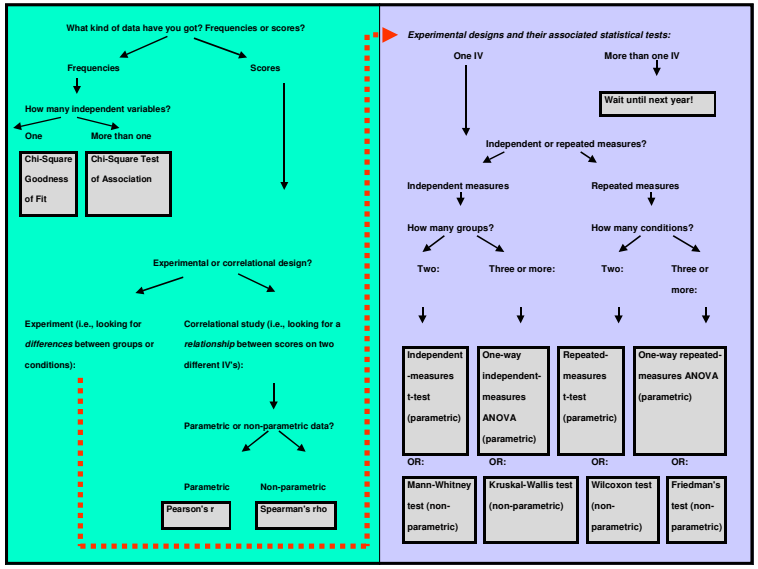
Sex differences in attachment to comfort objects:

Version 1:

20 children, 10 of each sex.

Each of them has a special toy or object.

Experimenter removes the object and measures the length of time each child spends crying.



1. Have a *score* for each participant.

2. *Ratio data* (time) - assume normal distribution and homogeneity of variance; so, a parametric test.

3. Looking for *differences* between groups.

4. We have two groups, which represent two different levels of one IV (sex).

5. *Independent measures* (each child is in only one group).

Appropriate test: independent-measures *t*-test.

Sex differences in attachment to comfort objects:

Version 2:

20 children, 10 of each sex.

Each of them has a special toy or object.

So, same problem as before; but now the experimenter removes the object and *rates* each child's level of distress on a 7-point scale.

Sex differences in attachment to comfort objects:

Version 3:

20 children, 10 of each sex.

Each of them has a special toy or object.

Same problem as before; now the experimenter removes the object and *counts* the number of children who cry for 5 minutes or longer.

1. Still have a *score* for each participant.

2. *Ordinal data* (ratings) - so, a non-parametric test.

3. Looking for *differences* between groups.

4. We have two groups, which represent two different levels of one IV (sex).

5. *Independent measures* (each child is in only one group).

***Appropriate test:* Mann-Whitney test.**

1. No longer have a *score* for each participant - now have *frequency* data.

Each child falls into one of four categories -

permutations of sex (male/female) and crying (short/long).

2. Looking for an association between sex and crying.

Appropriate test: Chi-Square test of association (2x2 contingency table)

Does sociometric status affect health?

Version 1:

Sample of 1000 people; measure SES and health (latter measured by number of visits to doctor in past 5 years).

Does sociometric status affect health?
Version 2:

Same problem, but take 30 low SES people and 30 high SES people.
For each person, measure their number of visits to the doctor.
Is there a difference between the two groups?

1. Each participant provides *two scores*: SES and health.
2. Interested in whether there is a *relationship* between two IV's - so, a correlation of some kind.
3. SES is probably not interval or ratio data. "Visits to doctor" is ratio.

Appropriate test: Spearman's rho.

1. Have a score from each participant (number of visits).
2. Ratio data.
3. Independent measures (*either* high SES *or* low SES).
4. One IV - SES.
5. Two groups, and we are looking for *differences* between them.

Appropriate test: independent-measures t-test

Does sociometric status affect health?
Version 3:
Same problem, but we find 30 high SES, 30 medium SES and 30 low SES people.
Count how many participants in each group have been to the doctor more than 5 times in the past two years.

Do people like Coronation Street better than they like Eastenders?	**Wilcoxon test**
Do European countries differ in their suicide rates?	**Chi-Square Goodness of Fit**
Do Tory and Labour voters differ in their choice of car?	**Chi-Square test of association**
Is there a relationship between cost of car and size of contribution to Tory Party funds?	**Correlation**
Do men have higher pain thresholds after watching videos of Schwarzenegger than after watching videos of Noddy?	**Repeated-measures t-test**
Do 5-, 10- and 15-year olds differ in their opinions of Sooty?	**Kruskal-Wallis test**

1. *Frequency* data (each participant merely falls into one of six categories).

2. Two IV's: SES and doctor-visiting. We are looking for an association between them.

Appropriate test: **Chi-Square test of association (2x3 contingency table).**

Effects of food additives on children's activity levels: **Group A: eat tartrazine-containing nosh.** **Group B: same nosh without tartrazine.** **DV: time spent running around.**	**Independent-measures t-test**
Effectiveness of different types of diet: **Group A: MacDonalds-a-day.** **Group B: Atkins Diet.** **Group C: banana and lettuce leaf diet.** **Group D: eat a diet book every day.** **DV: weight loss after 1 month.**	**Independent-measures ANOVA.**
Effect of lobotomy on ratings of Conservative Party: **Group A: No lobotomy.** **Group B: Lobotomy.** **DV: ratings of attractiveness of Conservative Party "policy" on immigration.**	**Mann-Whitney test.**

Effects of age on levels of grumbliness:
IV A: age.
IV B: assessment of extent to which the world is going to the dogs .

Spearman's rho

Effects of mobile phones on situational awareness:
Each person does two conditions: they walk along Western Road either using a phone or not using a phone.
DV: number of people bumped into.

Repeated-measures *t*-test

Effects of 4x4 ownership on driving ability:
Group A: 4x4 owners.
Group B: non-twatty-car owners.
DV: average speed at which each driver passes cyclists and horse-riders.

Independent-measures *t*-test.

Conclusions:

Statistics and experimental design are closely interrelated.

How you design your study affects what statistical analyses you will be able to use.

Wherever possible, avoid getting frequency data - try to get at least one score per participant.

Use a repeated-measures design if possible, as these require fewer participants and are more sensitive to effects of experimental manipulations.

Never design and run a study without thinking about how you will analyse the data obtained!