Introduction to analysing data in Psychology – Inferential Statistics

How do we decide if our results can be trusted ?

- So, we've done a nicely designed study and collected some data. How do we decide if what the data seem to show is a real effect or just a fluke ?
- That is, do the data show something that we can generalize to the population, or just sampling error ?
- The methods to answer this are collectively called inferential statistics
- This lecture, I will :
  - Introduce the logic of inferential statistics
  - Show you one way to apply it, in the between-groups design, as used in your Lab One

Back to the flatworms

- I train one flatworm to turn right, put it in the blender, feed it to another flatworm, and test the new flatworm
- On the first trial, it turns right … Are you convinced ?
- So I give it 100 trials and it turns right every time …
- The strength of the evidence depends on how likely it is that we might have got the same result just by chance
- The criterion probability, called a significance level, that we normally use is .05, or 5%, or 1/20
- If the probability of our data occurring just by chance is less than 1 in 20, we'll believe that it's a real effect

Back to the flatworms

- How many times would the flatworm have to turn right in a row to reach the .05 significance level ?
- One right turn : probability = 1/2
- Two right turns : probability = 1/4
- Three right turns : probability = 1/8
- Four right turns : p = 1/16
- Five right turns : p = 1/32 which is less than 1/20
- So if the flatworm made five successive right turns, that would be enough evidence that we'd start to think there was a real effect

More formal logic of inferential statistics

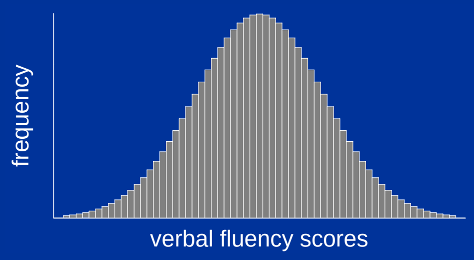
- Set up two rival hypotheses about the data :
- H0 – the null hypothesis
  - There is no real effect, results obtained just by chance or sampling error
- H1 – the experimental or alternative hypothesis
  - There is a real effect, results not obtained just by chance or sampling error
- Work out the probability of our results if H0 were true
- If that probability is less than the significance level (usually .05) reject the null hypothesis and conclude that there is a real effect

Logic of inferential statistics

- All statistical tests are applying this same logic, even if they look very different from each other
- A common situation is when we have two independent samples of scores and we want to know if they're significantly different
- We're really asking – how likely is it that these two samples were drawn from the same population, and the apparent difference between them is just sampling error?
- If that's not likely, then they come from different populations and represent a real difference
- So we need a test for the between-groups design …

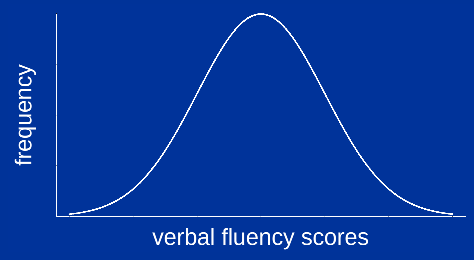
A picture of a sample

- If we draw a graph of the number of times each score occurs in the sample (frequency) we get a picture of the sample as a whole, called a frequency distribution



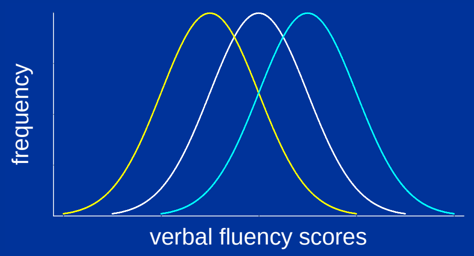
A picture of a sample

- We could draw it as a bar graph or histogram, as in the last slide, or just as a curve like this. Doesn't matter.



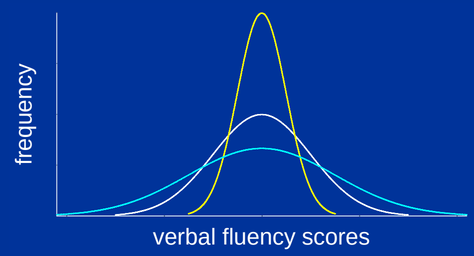
A picture of a sample

- Frequency distributions can differ in terms of their location or average value …



A picture of a sample

- And in terms of their variability, or how spread out they are …



How do we measure location or average ?

Usually by calculating the mean

Add up all the scores and divide by the number of scores

Equation for mean $\overline{X}$

$$\overline{X} = \frac{\Sigma X}{N}$$

X means "score"

Σ means "add these things up"

N means "number of scores"

Participant	Score
1	5
2	3
3	1
4	6
5	3
6	4
7	5
8	6
9	3
10	4
ΣX	40
mean	4

How do we measure variability or spread ?

Usually by calculating the standard deviation

Subtract the mean from each score, square the differences, add them up, divide by $N - 1$, and take the square root (sorry!)

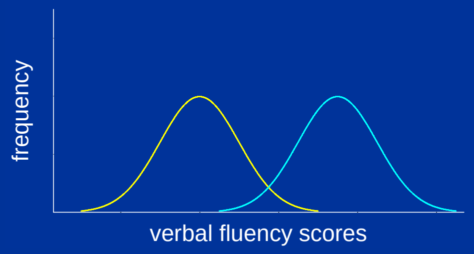
Equation for standard deviation

$$s = \sqrt{\frac{\Sigma(X - \overline{X})^2}{N - 1}}$$

Score	(score – mean)²
5	1
3	1
1	9
6	4
3	1
4	0
5	1
6	4
3	1
4	0
Σ	22
s	1.56

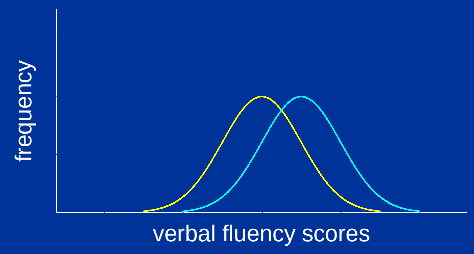
Do we think these two samples differ significantly ?

- Probably yes, they don't overlap much, they look as if they don't come from the same population



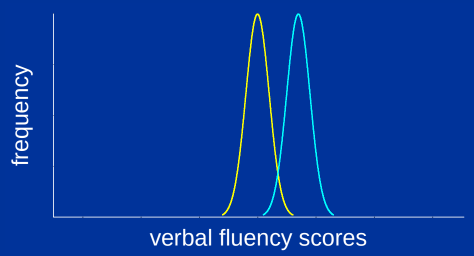
What about these two ?

- Probably no, their means are quite close together, so they overlap a lot and look as if they easily could come from the same population



And these two ?

- Probably yes, they are different. Although their means are quite close together, there is little variability, so they don't overlap much and don't seem to come from the same population.



A test for the between-groups design

- So whether two samples are significantly different or not depends on two things :
  - How far apart are their means ?
  - How variable are the samples ?
- Bigger difference between means is more likely to be significant
- But more variability implies less likely to be significant
- That is, if your samples are quite variable, you're going to need a big difference between their means to be significant
- We need a test that will reflect this …

A test for the between-groups design

- It's called Student's *t* and it compares the difference between means with the variability or standard error :

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\text{standard error}}$$

- The formula for standard error is horrible :

$$\sqrt{\frac{((N_1-1)s_1^2 + (N_2-1)s_2^2)(N_1+N_2)}{(N_1+N_2-2)N_1N_2}}$$

- But fortunately you don't have to know that, because the computer will work out *t* for you automatically … See Lab 1

A test for the between-groups design

- So to test whether two samples are significantly different, you work out *t* (or get SPSS to do it for you) and find out its probability if the null hypothesis were true
- SPSS will also work out this probability for you
- If the probability of your obtained value of *t* is less than the significance level (usually .05), it's significant and you can reject the null hypothesis
- That is, you conclude that the two samples really do come from different populations – there is a real difference
- If it's greater than the significance level, you have to conclude that the difference between the means is really only sampling error, or chance variation

A test for the between-groups design

- The probability of a particular value of *t* depends on the sample sizes
- Smaller samples mean that you need a bigger *t* to be significant
- The mathematics behind this means that it's best to express sample size as degrees of freedom, or df
- df = the number of scores that are free to vary while still producing the same two means
- In this two-sample case, df = $N_1 + N_2 - 2$
- SPSS will tell you the number of df for your data and whether your result shows a significant difference between the groups

Conclusion and a warning

- What I've tried to do is give you an intuitive understanding of how inferential statistics work, without going into the maths in any detail. I hope it made it less scary.
- But remember, all that statistics can do is tell you whether the average scores of the two samples are different
- Statistics can't tell you why they're different – because of your IV or because of a confounding variable ?
- Good statistics can't make up for bad research design !
- Enjoy the rest of the course, and I'll see you later in the year