PSYC3301 Lecture notes

Lecture 1- Unit Introduction

- Gilles Gignac

Unit structure

- Week 6, 10 and 12 will be videos
- The videos will be uploaded onto LMS

Assessment

- Mid-semester (25%)
  - MCQ
- Lab exam (15%)
  - In week 13
  - Short answers
- Final exam (60%)
  - MCQ
  - Lecture content only examinable

Pearson correlation: formula

$$\text{Pearsons } r = \frac{\sum z_x z_y}{N-1}$$

Z-score to make scores comparable on the same scale.

Z-score (standardised scores)

- Mean = 0
- SD = 1

$$z = \frac{X - \overline{X}}{SD}$$

Finding out Pearson's correlation

1- Convert to z-score
2- Multiply everyone's z-scores
3- Sum the products of z-scores
4- Divide sum by N-1

Small= .1
Medium= .3
Large= .5
Doesn't matter if it's positive or negative

p-value= probability and shows statistical significance

SPSS

- Analyse > correlate > bivariate

Coefficient of determination

- A correlation can be squared to yield the coefficient of determination r^2
- The coefficient of determination represents the proportion of variance in the DV accounted for by the IV.
- When multiplied by 100, the coefficient of determination represents the percentage of variance accounted for.

Why do we do statistics?

1- We can't access the population of interest
2- To measure the magnitude of an effect
3- To test assumptions associated with a statistical analysis

Population: the collection of units to which we want to generalise a set of findings.
Sample: A smaller collection of units from a population used to determine truths about the population.

- We have limited time/resources to analyse data from a sample and not a population.

Samples

- The smaller the sample, the less confidence we have in the accuracy of results
- We need to estimate the chances of making a mistake with statistical analysis based on a sample

What is a p-value?
Saying "no effect exists"= "null hypothesis" (no association). We want to reject the null hypothesis.

- We accept at most 5% chance of fooling ourselves. Thus, we want to see $p < .05$.

Correlation is a representation of the magnitude of an effect.

Statistic assumptions; for example:

- Random sampling (no analysis needed)
- Independence of observations (no analysis needed)
- Normally distributed data (needs analysis)
- Homogeneity of variance (needs analysis)

<u>Lecture 2: One-way ANOVA: Review</u>
Does alcohol make you dumber?

- Participants to 3 groups (N=24)
  - Low, medium and large doses
- Between-groups design (Ps measured once on a DV)

Procedure

- Mask alcohol
- Visual memory was measured
  - Between groups measured on a interval/ratio scale

Null: Memory means equal across all 3 groups
Alternative: Memory means will not be equal across 3 groups
^ AC= Lower performance

We do stats to measure the magnitude of effect

Theory of ANOVA

- We compare the amount of variability explained by the model (i.e., mean differences) to the error in the model (i.e., individual differences)
  - This ratio is called the *F*-ratio: the variability between means to the variability between individual observations.

Within group variability is the bad stuff

- The greater the amount of variability the ANOVA can explain relative to what it can't explain, the greater confidence we have that the independent variable is associated with mean differences on the dependent variable (beyond chance).
- That is, the greater the amount of "good stuff" to "bad stuff", the larger the *F*-value will be.

Thus, more likely to reject the null hypothesis.

- There are <u>three</u> sources of variance in the one-way between groups ANOVA.
  - Sums of squares between ($SS_{Between}$) (good stuff) $SS_B = N(\overline{X}_j - \overline{X}_G)^2$
  - Sums of squares within (SS_{Within}) (bad stuff) $SS_W = \sum (N_j - 1)S_j^2$

$$SS_T = SS_B + SS_W$$

- ❑ Sums of squares total (SS_{Total})

Each mean subtract the grand mean.

- In order to compare the "good stuff" versus the "bad stuff", the sums of squares (SS) need to be placed on a comparable metric.
- Accomplished by dividing the SS by their respective degrees of freedom.
- Degrees of Freedom (*df*) are the number of values that are free to vary with respect to a Sums of Squares term.
- SS divided by *df* gives Mean Square estimates.

Df = Number of groups -1

Mean Square Error
- Divide sums of squares by decrees of freedom
When between bigger than within chance of rejecting null

F-value
MS between divided by MS within

- 1 as an F-value is bad
- F-value is the analysis of variance (F-ratio)

F-value significance

- -The *p*-value needs to be p < .05 in order to declare the *F*-value of 15.25 as statistically significant.

Sig = .000 means p <.001
The null hypothesis of ____ means has been rejected

Homogeneity of variances

- Assumption: the variances are the same
- Independent sample t-test assumes homogeneity of variance

Levene's statistic you don't want to satisfy the null

Eta-Squared

- Calculates the magnitude of effect
- It is not sufficient to simply observe a *p*-value less than .05 and then call it a day.
- Must always calculate an effect size estimate.
- η^2 is an effect size measure
- It represents the percentage of variance in the dependent variable that is accounted for by the independent variable (i.e., the groups).
- It's a coefficient of determination, essentially.

Pearson's r^2 has same definition as eta-squared

$$\eta^2 = \frac{SS_{Between}}{SS_{Total}}$$

Thus, 30.7% of the variance in the memory scores was accounted for by the alcohol consumption levels.

Cohen to determine if it is big

- Small: .01
- Medium: .06
- Large: .14