

Research Design & Analysis

Lecture 5: From correlation to regression

Lecture overview

- Quick recap from last year
- Experimental designs
 - Between-subjects designs
 - Within-subjects designs
- Non-experimental designs
 - Correlational designs

Hypotheses

- Hypothesis: a testable statement about what is likely to occur between two or more factors/variables in a study
- People with higher levels of happiness will spend more money on other people than people with lower levels of happiness.

Null hypothesis:

- People with higher levels of happiness will not spend more money on other people

Variables

- Independent Variable
 - A factor either manipulated by the experimenter (e.g., condition A Versus condition B), or believed to explain differences in an observed phenomenon (e.g., gender differences in leadership).
- Dependent Variable
 - Some measure of performance (e.g., test scores), or scores on a specific variable (e.g., level of extraversion, socio-economic status)

Variables continued

- Extraneous Variable
 - A variable that interferes with or effects the outcome of your study
 - Should always try and control for any possible extraneous variable
 - E.g., noise, lighting, social influence
- Confounding Variable
 - A variable that is systematically linked to changes in the IV and may therefore explain differences in the DV
 - E.g., Performance in morning V Performance in night
 - Confounding variable of breakfast?

Correlational Research: three components:

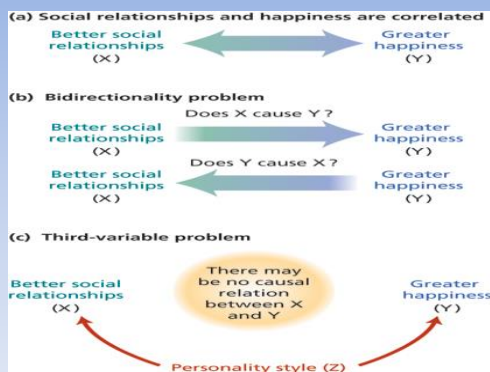
- Researcher measures one variable (X)
- Researcher measures a second variable (Y)
- Researcher statistically determines whether (X) and (Y) are related
- E.g.,
 - are *height* and *shoe-size* related i.e., is a bigger shoe size associated with being taller?
 - How strong is this association?
 - NB: no manipulation of *height* or *shoe-size* – simply measuring pre-existing variables.

- Correlational Research: Advantages

- Can establish generalization outside the laboratory
- Can generate ideas for further laboratory study
- Allows the examination of variables that may not be practical or ethical to study in the laboratory

Correlation refresher

- Nature of correlational designs:
 - Not looking for differences between experimental groups or conditions
 - Investigating if there is a consistent relationship between variables (shared variance)
 - Is there an association?
 - What is the 'direction' of this relationship?
 - What is the strength of the association?
 - Is it statistically significant?



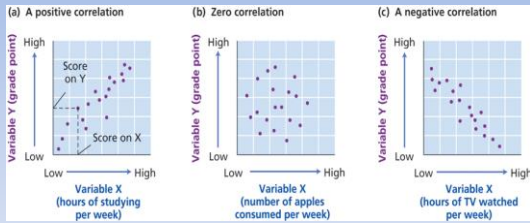
Correlational Design

- E.g.
- Are happiness and prosocial behaviour associated with each other?
- No IV, No DV, but “co-relating” variables...
- Not exploring whether one **causes** another...

- Correlation does not equal causation
 - Some interpretive possibilities:
 - X caused Y
 - Y caused X
 - Both variables influenced each other
 - A third variable influenced both X and Y (spurious)

- Positive Correlation: higher scores on one variable are associated with higher scores on a second variable
- Negative Correlation: higher scores on one variable are associated with lower scores on a second variable
- No Correlation: the two sets of scores are completely unrelated

Methods of Research



- Correlation Coefficient: a statistic that indicates the direction and strength of the relation between two variables
 - Correlation coefficient is measured from 0 to 1.
 - 0 = no correlation.
 - 1 = perfect correlation.
 - + or – sign simply denotes the direction of the correlation, positive or negative.

- Correlation Coefficient: a statistic that indicates the direction and strength of the relation between two variables
 - Positive Correlation:
 - 0 to 1
 - Negative Correlation:
 - <0 to -1
 - No Correlation = zero

- So, in reality you will find values between -1 and 1.
- Remember this value is just a descriptive value
- As a guide to describing r-values (Cohen, 1988):
 - $r = \pm .10$ to $.29$ – small
 - $r = \pm .30$ to $.49$ – medium
 - $r = \pm .50$ to 1.00 – large

Data Associations:

From correlation to regression

Correlational designs

- Features of correlation:
 - Variables that are continuous not categorical
 - Ordinal or interval data
 - Direction of correlation
 - Positive: Increase A $\rightarrow$ Increase B
 - Negative: Increase A $\rightarrow$ Decrease B
 - Strength/significance of correlation
 - r^2 value provides amount of shared variance
 - p-value (sig.) tells us whether the correlation is significant

Correlational designs

- So far have looked at correlations between two variables (bivariate)
 - E.g. Altruism score and Extroversion score
 - Temperature and Ice-cream sales
- No causal relation can be assumed without theory
- All theorise is that scores vary in related way
 - Variance from their mean score

Regression

- Research questions:
 - Which of a **range** of variables are most strongly related to a specific variable of interest?
 - Do these variables do better than chance or simple mean scores at **predicting** scores on a variable of interest?
 - How might these variables **combine** to predict a variable?
- Multiple variables:
 - Could just carry out many simple correlations
 - Provide several Bivariate associations (2 variables correlated)
 - Still only a non-causal correlation produced
 - Will not indicate our overall success with all variables used

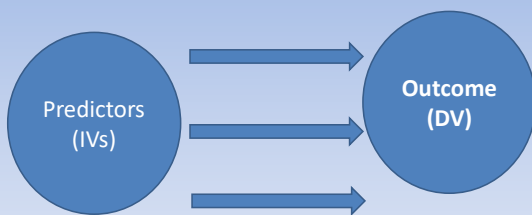
Regression

- Important extension of 'regression' over 'correlation':
 - **Combined** association effect of the predictor variables
 - **Multiple** regression
 - Treating the predictors (separate variables) as a combined variable in relation to the dependent variable
 - How 'associated' they are as a whole
 - How 'predictive' of the DV they are as a whole
 - (nb. Will also look at each in isolation)

- How do you choose which variables to use as your “predictors” or IVs in a regression design?



- What guides selection of IV's?



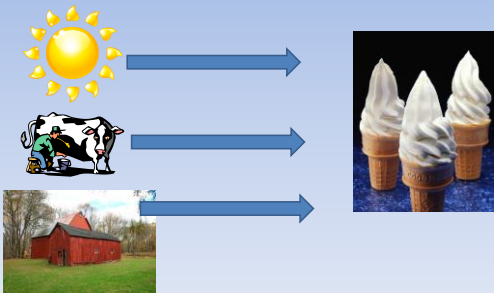
Regression

- Regression can be used to analyse a more ‘causal’ relation between variables
- Theory suggests a dependent variable and independent (or explanatory) variable(s)
 - Can also use more than one explanatory variable (IV)
- Regression analyses amount of change in the dependent variable based on change in the explanatory/independent variables
- Can be used to make **predictions** for the DV based on scores for the IV

- Ice-cream sales example
 - What guided selection of IV's?



- Ice-cream sales example
 - What guided selection of IV's?



- Ice-cream sales example
 - Was this “empirically driven good science?”
 - Better approach if funded research project?
 - Would you actually expect a “good fit” from this model?

Assumptions

- Continuous data (NB: dummy variables)
- Normal distribution of data
- Linear relationships between IVs and DVs
- Homogeneity of variance

Before next week..

- Read up on:
 - Assumptions of regression
 - Regression analysis
- Next week's session:
 - Analysing and reporting regression
