EDUR 8132
Educational Statistics 2

Introductory Notes

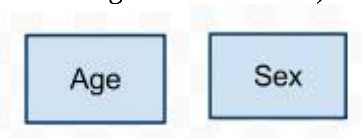
1. The Big Picture – Statistical Models of Behavior/Outcomes/Dependent Variables

Purpose of Course: – To learn how to model outcomes of interest using statistical models of regression and analysis of variance (ANOVA).

Variables – Identification of general measurement scale (qualitative/categorical vs. quantitative/continuous), and distinction between independent variables (IV) and dependent variables (DV)

Path Diagrams (or Causal Model Diagrams) – used to pictorially display theoretical models of outcomes of interest.

- Square Box indicates observed variable (observed means it is measured directly usually with one measuring item/instrument). Name of variable will appear in box.

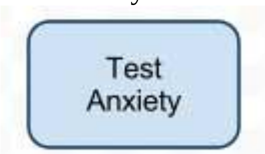


- Oval or Circle or Rounded Rectangle indicates latent variable (one that is not measured directly, but instead indirectly from several indictors. For example:

  IQ is a composite from several sub-scales;

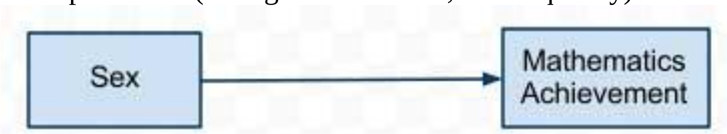
  Socio-economic Status (SES) is formally a composite of three indicators [what are they?];

  Test Anxiety is often measured from responses to 4 to 20 questionnaire items [before or during a test does your heart beat fast, do you sweat, do you think about failure, do you panic, etc.].



- Lines are used to indicate how variables are related. A line with a single arrow indicates the independent variable (IV) predictors or influences the dependent variable (DV) – the arrow points to the DV.

  Example 1: Sex (biological difference, not frequency) and mathematics achievement



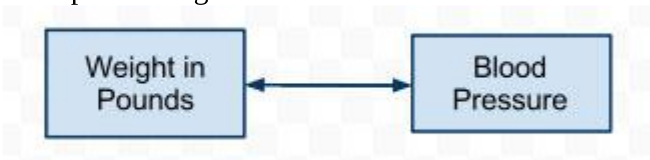
  [What type of statistical test for this model?]

Example 2: Exercise (yes vs no) and daily weight change

[What type of statistical test for this model?]

- A line with double arrows indicates two variables are related, but no clear IV or DV.

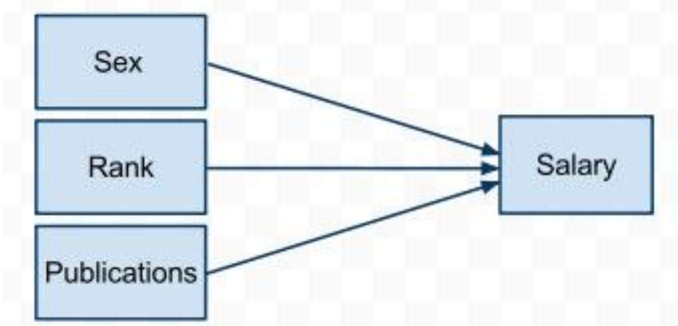
Example 3: Weight and Blood Pressure



[What type of statistical test for this model?]

- A model with multiple IV and one DV

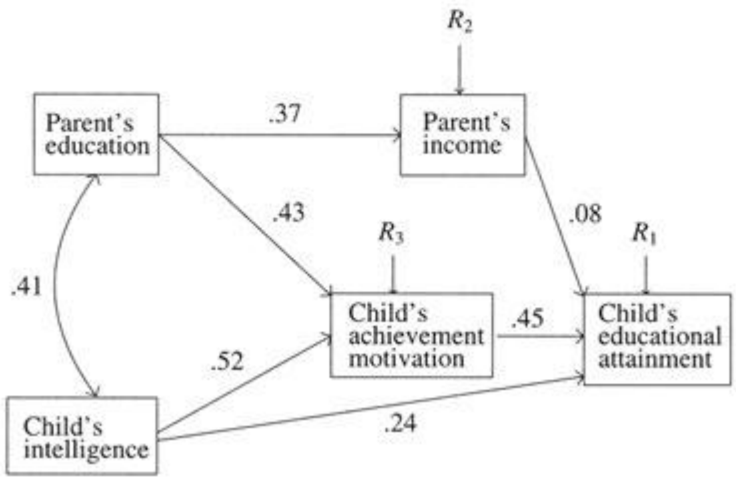
Example 4: Faculty salary, sex, rank, and number of publications



[What type of statistical analysis for Example 4?]

- Examples of path diagrams from actual research:

Example 5: [What are IV and DVs below? What does this model show?]



Numbers above indicate regression coefficients (measure of the predicted change in the DV for a unit change in the IV.
Example 6: Modeling student achievement in statistics!

Nasser, F. M. (2004). Structural Model of the Effects of Cognitive and Affective Factors on the Achievement of Arabic-Speaking Pre-service Teachers in Introductory Statistics, Journal of Statistics Education Volume 12, Number 1 (2004), www.amstat.org/publications/jse/v12n1/nasser.html
[Find Figure 2 (near bottom of site)]

2. Central Tendency and Variability Review

Central Tendency – Mean, Median, Mode (calculations in Excel and SPSS)
Variability – Variance and Standard Deviation (calculations in Excel and SPSS)

3. Review of Correlation with Pearson's r

Purpose of Pearson's r – association between two variables (zero-order Pearson's r)
Examples in Research:

Michinov, N. (2005). Social Comparison, Perceived Control, and Occupational Burnout. APPLIED PSYCHOLOGY: AN INTERNATIONAL REVIEW, 2005, 54 (1), 99118
http://www.bwgriffin.com/gsu/courses/edur7130/readings/Correlational_2.pdf

Holen, M., et al. (1970s). Student self-prediction of academic achievement. Journal of Educational Research.
http://www.bwgriffin.com/gsu/courses/edur7130/readings/Correlational_4_s.pdf

Example Calculations in SPSS

a. Is blood pressure between left and right arm correlated?
b. Is systolic or diastolic blood pressure related to weight?

4. Review Two Independent Samples t-test (two group t-test)

Purpose of Two Group t-test – difference between two groups on a quantitative DV
Examples in Research:

Eden et al. (1996). Abnormal processing of visual motion in dyslexia revealed by functional brain imaging. Nature 66-69.
http://csl.georgetown.edu/publications/Eden_et_al_Nature_1996.pdf (no t values)

Akpinar, E. (2007). The effect of dual situated learning model on students' understanding of photosynthesis and respiration concepts. Journal of Baltic Science Education, 6, 16-26.
http://www.bwgriffin.com/gsu/courses/edur7130/readings/QuasiExperimental_4.pdf (t-values)

[Show better way to present results above – discuss reporting statistical outcomes document.]

Example Calculations in SPSS

a. Is there a difference in daily weight change based upon exercise status the prior day?
[Delete weight change for dates without recordings.]
b. Is there a difference in statistics performance between males and females?
[http://tinyurl.com/26xu8hr -- Link to EDUR 8131 Grades by Sex (Sum10 Spr10 F09 Sum09) Google document]

5. Coin Toss and Hypothesis Testing Logic – Is this result real; what is the probability of such a result?

(a) Hypothesis Testing and Probabilities – All Starts with H_0

Hypothesis testing is based upon probabilities and judging those probabilities against a known or theoretical standard. The standard against which probabilities are calculated is stated in the null hypothesis. Consider, for instance this hypothesis:

H_0: a fair coin has a 50:50 chance of heads and tails ($\mu = .5$)
H_1: coin is not fair; it does not have a 50:50 chance of heads and tails ($\mu \neq .5$)

The null hypothesis above (H_0) states that a coin, if fair, should land on heads 50% of the time and tails 50% of the time.

(b) Compare Empirical Results Against What Was Expected in H_0

How can we test whether a coin appears to be fair (50:50 chance of heads/tails)?

We can empirically test that stated in the null hypothesis (H_0) by flipping a coin (taking a sample of coin tosses) and then compare our sample coin flip results to what is expected assuming the coin is fair (i.e., comparing our results to what was expected in the null hypothesis).

(c) Reject or Fail to Reject H_0 Based Upon Empirical Results

If the results we obtain in a sample are consistent with the null hypothesis (e.g., coin appears fair, the probabilities of heads/tails from our experimental coin tosses are similar to what is expected in the null hypothesis), then we will ***fail to reject*** the null and state that the sample data appear to be consistent with the null, thus the coin appears to be fair.

If, however, our sample results are odd, rare, or unexpected – that is, the results obtained are those that occur only with low probability – then we will ***reject*** the null hypothesis of fairness and conclude instead that the sample results from the coin are not consistent with the null hypothesis, therefore the null hypothesis is untenable and we reject it in favor of the alternative hypothesis.

(d) Empirical Probabilities Compared Against What Standard?

How does one judge whether empirical results obtain in a sample are consistent or inconsistent with the null hypothesis? What standard is used to judge whether results are rare if the null hypothesis is true?

Researchers set, at the outset, a predetermined standard for determining whether results appear to be change (a fluke) or due to something that may be real. In most statistical hypothesis testing, this is probability value that is set low, and this value is called ***alpha*** (the probability of committing a Type 1 error).

[An empirical example with a coin toss.]

6. Errors in Hypothesis Testing

In hypothesis testing, two decisions can be made, either reject H_0 or fail to reject H_0. Two errors can also be made in deciding whether to reject or fail to reject H_0. The table below specifies each of these errors.

		Population Situation Regarding H_0	
		H_0 True	H_0 False
One's Decision	Reject H_0	Mistake (α, alpha) Type I error	Correct ($1 - \beta$) Power
	Fail to Reject H_0	Correct ($1 - \alpha$)	Mistake (β, beta) Type II error

Case 1: Reject H_0 when H_0 is true.

This is an error because the null was rejected and it should not have been rejected. This is a Type I error (rejecting H_0 when H_0 is true). The probability of making this type of error is equal to α, the alpha level that researchers set, which is traditionally set at .05 or .01 (and sometimes .10).

Case 2: Fail to reject H_0 when H_0 is false (H_1 is true).

This is also an error, and is known as a Type II error (failing to reject H_0 when H_0 is false). This error occurs when one does not reject the null hypothesis when it should have been rejected because there really are differences (thus the alternative hypothesis is actually true). The probability of making this type of error is equal to β; unlike α, the researcher cannot directly set the level of α, but must manipulate other factors which influence α like sample size and/or the alpha level.

Case 3: Reject H_0 when H_0 is false (H_1 is true).

This is a correct decision because H_0 is not true so we adopt the alternative hypothesis, H_1. The probability of this occurring is 1 - α, and this probability is called ***power***.

The ***power*** of a test is the probability of rejecting a false H_0, p(rejecting false H_0); the probability of detecting differences if they actually exist. Power is influenced by (a) effect size, (b) n, (c) control of the variability in studies, (d) choice of hypotheses, and (e) α-level.

Case 4: Fail to reject H_0 when H_0 is true.

This is also a correct decision because H_0 was not rejected, and no differences actually exist. The probability of this occurring is 1 - α.

The researcher only has direct control of the α error level. The researcher cannot directly manipulate the α error level; however, several factors can increase or decrease α. These factors include sample size, the alpha level, type of hypothesis, and the amount of variability in the study.