Terminology Review

Psy 420

Andrew Ainsworth

Concept review

Research Terminology

Variables

- IVs and DVs
  - Independent variables
    - are controlled by the experimenter
    - **and/or** are hypothesized to influence other variables (e.g. DV)
    - **and/or** represent different groups or classifications participants belong to (either assigned or ascribed)
  - Dependent variables are what the participants are being measured on; the response or outcome variable
  - Think of them as “input/output”, “stimulus/response”, etc.
  - Usually represent sides of an equation

Variables

- Qualitative vs. Quantitative
  - Qualitative variables are those that change in quality or kind
    - (e.g. male/female, ethnicity, etc.)
  - Quantitative variables are those that change in amount

Variables

- Continuous, discrete and dichotomous
  - Continuous data
    - smooth transition from one to the other rather than in steps,
    - can take on any value in a given range
    - the number of given values in the range are only limited by the precision of the measuring instrument (can be infinite)

Variables

- Continuous, discrete and dichotomous
  - Discrete
    - Categorical
    - Limited amount of values
    - And always whole values
  - Dichotomous
    - discrete variable with only two categories

Variables

- Continuous, discrete and dichotomous
  - Continuous to discrete
    - often for the sake of simplicity continuous data is “dichotomized”, “trichotomized”.
    - Often because people are obsessed with anovas or some other stat they are accustomed to (chi-square, etc.)
    - Doing this will reduce your power and cloud your interpretation
    - Reinforce use of the appropriate stat at the right time

Variables

- Continuous, discrete and dichotomous
  - Which type of data you have will decide what type of analysis you should or at least can use
  - Much of the differences in the chapters in this book have to do with what kind of data your dealing with (plus how it's collected and other things)

Levels of Measurement

- Nominal – Categorical
- Ordinal – rank order
- Interval – ordered and evenly spaced; changes in the construct represent equal changes in what you are intended to measure
- Ratio – has absolute 0; a true absence of the trait.
  - y(I, R) – one sample t-test
  - y(O, N) – one-way chi-square
  - y(I, R) and x(O, N) – two sample inde. t-test, one-way ANOVA
  - 2 xs (O, N) – two-way chi square
  - The last two are usually grouped together and treated as "continuous".

Types of input or treatment

- Qualitative input – sex (male/female), ethnicity, treatment groups, etc.
- Quantitative input – age groups, weight classes, years of education, etc. These can be quantitative categories (e.g. ANOVA) or continuous predictors (e.g. regression).

Types of output or outcome measure

- Output variables can also be discrete, ordinal or continuous.
- Research using continuous outcome measures will be the focus of this class.
  - These outcomes measure the amount of something and also track the degree the amount changes between groups or time periods.
- Analyses of discrete or ordinal data is usually limited to analyses like a chi-square test or other non-parametric tests.
- Ordinal data can be treated as continuous as long as there are enough categories (7 or more) and it is believed that there is an underlying continuum.

Number of outcomes

- Number of outcome measures changes the type of analysis you would use.
- Univariate, Bivariate, Multivariate
  - Uni - only one DV, can have multiple IVs; this is what we'll cover in this class
  - Bivariate – two variables no specification as to IV or DV (r or χ^2)
  - Multivariate – multiple DVs, regardless of number of IVs; covered in psy 524

Experimental vs. Non-Experimental

- Experimental – high level of researcher control, direct manipulation of IV, true IV to DV causal flow
- Non-experimental – low or no level of researcher control, pre-existing groups (gender, etc.), IV and DV ambiguous
- Experiments equal higher levels of internal validity (freedom from confounds), non-experiments typically will have higher generalizability (external validity)
- All of the stats we'll discuss can be applied to data collected in both experimental or non-experimental settings
- Causality in research is decided by the research design, you can apply sophisticated data analysis to crappy data and you still get crappy results

Types of research designs

- Continuous outcomes (what we'll cover in this course)
  - Randomized (between) groups
    - One-way between groups fixed effects ANOVA
    - Factorial between groups fixed effects ANOVA
  - Repeated measures (within groups)
    - One-way within groups design
    - Factorial within groups design
  - Mixed between and within groups
    - Mixed ANOVA

Types of research designs

- Continuous outcomes (what we'll cover in this course)
  - Adjusting for other variables
    - Analysis of Covariance
  - Pilot testing and incomplete designs
    - Latin squares designs
    - Screening and incomplete designs
  - Analyses of non-fixed effects
    - Random effects ANOVA and generalizability

Types of research designs

- Ordinal outcomes – non-parametric tests (Wilcoxon rank sum test, Sign test, etc.)
- Discrete outcomes
  - Chi-Square
  - Log-linear Models
  - Logistic Regression
- Time as an outcome
  - Survival Analysis

Statistics Review

- Statistic vs. Parameter
  - Statistics describe samples
  - Parameters describe populations
  - Statistical inference
    - Often statistics are used to estimate parameters (this is statistical inference)
    - The process of making decisions (inferences) about populations based on a sample of participants.
    - Researcher sets up two hypothetical states of reality

Measures of central tendency and dispersion

- Central Tendency
  - Mode – value with highest frequency
  - Median – value in the center of the distribution
  - Mean – Average value
    - For continuous variables

$$\bar{X} = \frac{\sum X}{N}$$

    - For dichotomous variables
      - 1 positive response (Success) → P
      - 0 negative response (failure) → Q = (1-P)
      - MEAN(Y) = P, observed proportion of successes
      - VAR(Y) = PQ, max when P = .50, variance depends on mean (P)

Measures of central tendency and dispersion

- Dispersion – spread of a distribution
  - Range – Max minus min
  - Deviation

$$deviation = \sum_{i=1}^{n} (X_i - \bar{X}), \text{ problem is this equals 0}$$

So often each deviation from the mean is squared,

$$\sum_{i=1}^{n} (X_i - \bar{X})^2$$

Measures of central tendency and dispersion

- Dispersion – spread of a distribution
  - Variance

$$\textit{sample}\ \text{Variance} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n};\ \text{comp formula}\ \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n}$$

$$\textit{estimated population Var} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1};\ \text{comp formula}\ \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1}$$

Measures of central tendency and dispersion

- Dispersion – spread of a distribution
  - Standard Deviation – dispersion of a single sample

$$\textit{sample}\ \mathrm{SD} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}};\ \text{comp formula}\ \sqrt{\frac{\sum_{i=1}^{n}X_i^2 - \frac{\left(\sum_{i=1}^{n}X_i\right)^2}{n}}{n}}$$

$$\textit{estimated population SD} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}};\ \text{comp formula}\ \sqrt{\frac{\sum_{i=1}^{n}X_i^2 - \frac{\left(\sum_{i=1}^{n}X_i\right)^2}{n}}{n-1}}$$

Measures of central tendency and dispersion

- Dispersion – spread of a distribution
  - Standard Error – dispersion of a sampling distribution of means

$$S_{\bar{X}} = \frac{SD_{n-1}}{\sqrt{n}}$$

Relationships between variables

- Both variables discrete
  - Chi Square
    - "Goodness of fit" test – one-way test

$$\chi^2 = \sum \frac{(o-e)}{e}$$

    - Contingency tables
    - Expected values can be given
      Or estimated

$$e = \frac{R*C}{T}$$

Both Variables Continuous

- Correlation – non-directional relationship
  - Degree of co-relation
  - Range from -1 to positive 1
  - Positive vs. Negative Correlation
  - Computational formula

$$r = \frac{\sum XY - \left[(\sum X)(\sum Y) \right] / N}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{N} \right] \left[\sum Y^2 - \frac{(\sum Y)^2}{N} \right]}}$$

Both Variables Continuous

- Regression – directional relationship

$$Y' = bx + a$$

$$b = \frac{\sum XY - \left[(\sum X)(\sum Y)\right]/N}{\sum X^2 - \frac{(\sum X)^2}{N}}$$

$$a = \overline{Y} - b\overline{X}$$

Discrete predictor, continuous outcome

- z-test
  - Z-scores

$$z = {x - \bar{x}} \Big/ {SD}$$

  - Z-test, when sigma is known

$$z = {x - \mu} \Big/ {\sigma_{\bar{X}}}$$

Discrete predictor, continuous outcome

- Z-test
  - Assumes that the population mean and standard deviation are known (therefore not realistic for application purposes)
  - Used as a theoretical exercise to establish tests that follow
  - Samples can come from any part of a distribution with a given probability, so taking one sample and comparing to the population distribution can be misleading
  - Sampling distributions are established; either by rote or by estimation (hypotheses deal with means so distributions of means are what we use)

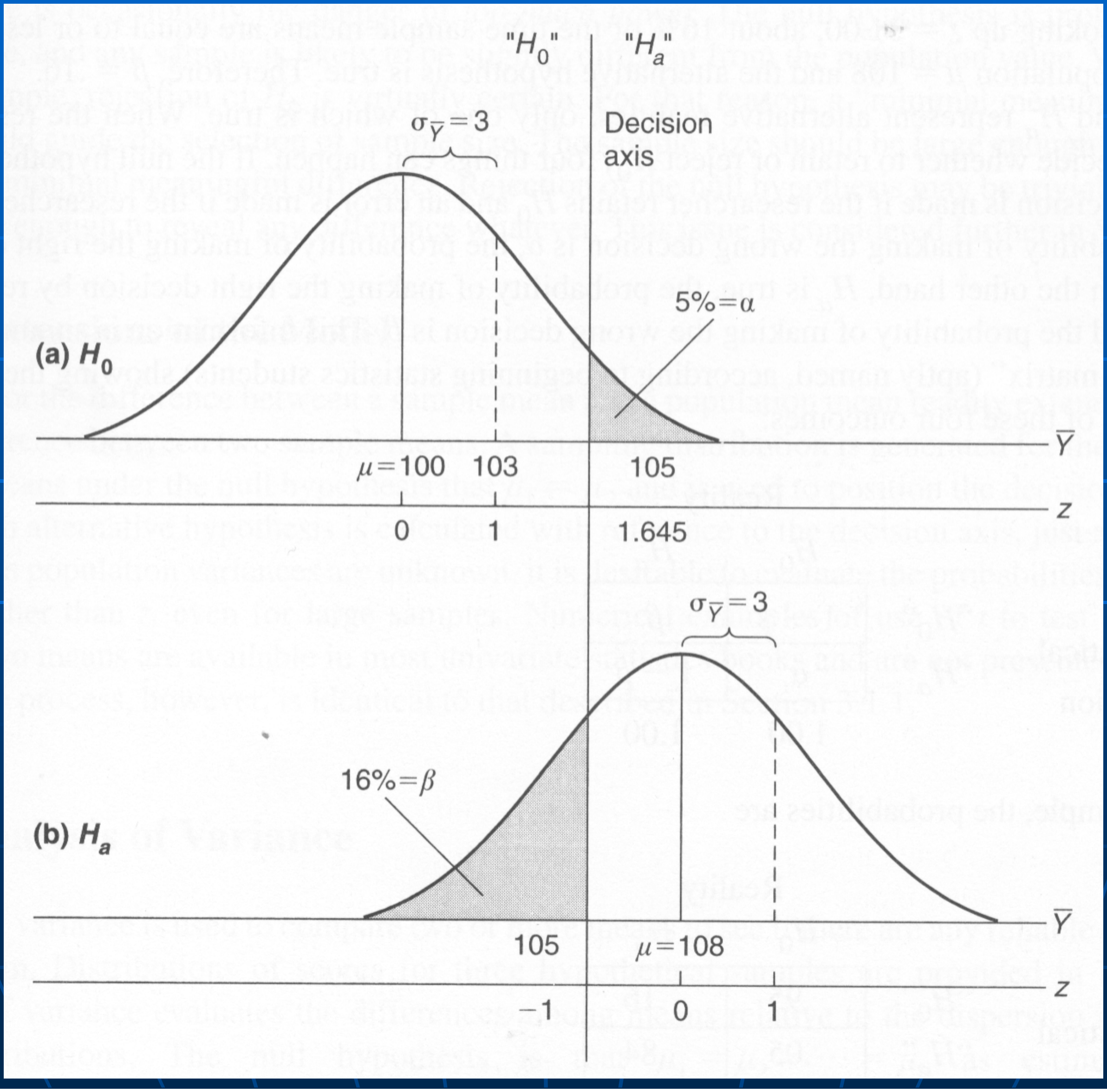
Hypothesis Testing and Z

- Decision axes established so we leave little chance for error
  - Type 1 error – rejecting null hypothesis by mistake (Alpha)
  - Type 2 error – keeping the null hypothesis by mistake (Beta)

Your Decision	Reality: H_0	Reality: H_A
"H_0"	1 - α	β
"H_A"	α	1 - β
	1.00	1.00

Your Decision	Reality: H_0	Reality: H_A
"H_0"	.95	.16
"H_A"	.05	.84
	1.00	1.00

Hypothesis Testing and Z



Power and Z

- Power is established by the probability of rejecting the null given that the alternative is true.

- Three ways to increase it
  - Increase the effect size
  - Use less stringent alpha level
  - Reduce your variability in scores (narrow the width of the distributions) more control or more subjects

- "You can never have too much power!!" – this is not true
- t-tests are realistic application of z-tests because the population standard deviation is not known (need multiple distributions instead of just one)

Discrete predictor, continuous outcome

- one sample t-test – when sigma is unknown and has to be estimated

$$t = \frac{\bar{X} - \mu}{S_{\bar{X}}}$$

Discrete predictor, continuous outcome

- independent samples t-test

$$t = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}_A - \bar{X}_B}}$$

$$s^2_{pooled} = \frac{(n_A - 1)s^2_A + (n_B - 1)s^2_B}{n_A + n_B - 2}$$

$$s_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{s^2_{pooled}}{n_A} + \frac{s^2_{pooled}}{n_B}}$$

Discrete predictor, continuous outcome

- dependent samples t-test

$$t = \frac{\bar{X}_d - 0}{s_{\bar{d}}}$$

$$s_{\bar{d}} = \frac{SD_d}{\sqrt{n}}$$