Before we get started:

http://arievaluation.org/projects-3/
OUTCOME EVALUATION USING R.
STATISTICAL ANALYSIS THAT ARE
MORE COMMONLY USED IN PROGRAM
EVALUATION AND HOW TO CONDUCT
THEM USING THE FREE STATISTICAL
SOFTWARE R

Antonio Olmos
Kai Schramm
Priyalathta Govindasamy

Antonio.Olmos@du.edu
AntonioOlmos@aumhc.org
Plan for the presentation

- Rationale for using R
- Outcomes analysis using Statistics
  - Introduction to R, Specifically, R Commander
  - Descriptive statistics
  - Graphics
  - Association
  - Difference
- Final thoughts
Installing R and R-commander

- You can find R and R-commander in the web.
- Follow instructions in video
- Contact us if you have any problems

http://r-project.org
R-commander

- R is command-based (scripts; very much like Syntax in SPSS)
- R-studio is much more user friendly and it has a very robust interface
  - But still, command-based (i.e., like syntax in SPSS)
- R-commander: Point-and-click
Now, what?
Statistics

- **Descriptive statistics**: describe, summarize, or make sense of a particular set of data.
- **Inferential statistics**: go beyond the sample data and infer the characteristics of populations.
Descriptive Statistics/
Graphic Representations of Data/Reliability
Descriptive Statistics

- We use statistics to summarize and make sense of a particular data set
- **Data set**: set of data which includes the cases (i.e. participants) and variables (IV & DV)

```r
load("D:/Rcmdr_AEA/Exercises&Examples/FactorialANOVA.RData")
summary(Dataset)
```

<table>
<thead>
<tr>
<th></th>
<th>attitude</th>
<th>fact_a</th>
<th>fact_b</th>
<th>coo_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>10</td>
<td>inner city:15</td>
<td>10 hours:15</td>
<td>Min. :0.000000</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>27</td>
<td>middle class:15</td>
<td>15 hours:15</td>
<td>1st Qu.:0.0005556</td>
</tr>
<tr>
<td>Median</td>
<td>35</td>
<td>upper class:15</td>
<td>5 hours :15</td>
<td>Median :0.0355556</td>
</tr>
<tr>
<td>Mean</td>
<td>35</td>
<td></td>
<td></td>
<td>Mean :0.0277778</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>42</td>
<td></td>
<td></td>
<td>3rd Qu.:0.0450000</td>
</tr>
<tr>
<td>Max.</td>
<td>64</td>
<td></td>
<td></td>
<td>Max. :0.0800000</td>
</tr>
</tbody>
</table>
Frequency Distributions

- Frequency of each unique data value are shown
- **Count (Frequency)** is shown in the second column; the third column shows the “percent distribution”

25% of the individuals are in the 120-140 range score

<table>
<thead>
<tr>
<th></th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, 20]</td>
<td>2</td>
<td>1.02</td>
</tr>
<tr>
<td>(20, 40]</td>
<td>1</td>
<td>0.51</td>
</tr>
<tr>
<td>(40, 60]</td>
<td>1</td>
<td>0.51</td>
</tr>
<tr>
<td>(60, 80]</td>
<td>1</td>
<td>0.51</td>
</tr>
<tr>
<td>(80, 100]</td>
<td>6</td>
<td>3.06</td>
</tr>
<tr>
<td>(100, 120]</td>
<td>8</td>
<td>4.08</td>
</tr>
<tr>
<td>(120, 140]</td>
<td>49</td>
<td>25.00</td>
</tr>
<tr>
<td>(140, 160]</td>
<td>81</td>
<td>41.33</td>
</tr>
<tr>
<td>(160, 180]</td>
<td>34</td>
<td>17.35</td>
</tr>
<tr>
<td>(180, 200]</td>
<td>11</td>
<td>5.61</td>
</tr>
<tr>
<td>(200, 220]</td>
<td>2</td>
<td>1.02</td>
</tr>
<tr>
<td>Total</td>
<td>196</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Bar graph

- Uses vertical bars to represent data
  - Height of bars shows frequencies of categories
  - Used for categorical variables

<table>
<thead>
<tr>
<th>Age Ranges</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-0.071, 10.1]</td>
<td>10</td>
</tr>
<tr>
<td>(20.3, 30.4]</td>
<td>40</td>
</tr>
<tr>
<td>(40.6, 50.7]</td>
<td>70</td>
</tr>
<tr>
<td>(60.9, 71.1]</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>mean</th>
<th>sd</th>
<th>se(mean)</th>
<th>cv</th>
<th>skewness</th>
<th>kurtosis</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>37.44898</td>
<td>11.27704</td>
<td>0.8055032</td>
<td>0.3011309</td>
<td>-0.1528243</td>
<td>0.07827869</td>
<td>196</td>
</tr>
</tbody>
</table>
Histogram

- Distribution of quantitative variable
  - Looks like bar graph except there is no space between the bars, and it is used only with quantitative variables
Boxplots

- Plot shows distribution of quantitative variable(s)
  - The bar inside the “box” represents the median, and the “whiskers” represent the width of the distribution of scores quantitative variables
Line graph

- Use line(s) to depict information about variable(s)
  - Simple line graph can show trend
  - Can compare groups
Scatter plot

- Depict relationship between two *quantitative* variables
  - IV or predictor variable placed on X axis (horizontal axis) and DV on Y axis (vertical axis)
Reliability
Reliability Definition

Unreliable, not valid
Results are inconsistent

Reliable, valid
Results are consistently on target

Reliable, not valid
Results are consistent, but off target
Cronbach’s alpha

- There are other measures for various internal reliability measures, but Cronbach’s Alpha is the most used and reported.
  - Describes how highly a set of items hang together
  - Values of 0.70 or better are considered ideal

- NOTE: Remember to recode reverse-worded items

- Only measures one construct at a time
Notes about Reliability

- Make sure past reliability estimates for your instrument are high:
  - > 0.80 for cognitive tests
  - > 0.70 at a minimum
  - > 0.90 minimum for tests used for placement (0.95 desirable standard)

- Sample in original study similar to your study
APA-AERA Guidelines

- The reporting of reliability coefficients alone, with little detail regarding the method used to estimate the coefficient, the nature of the group, and the conditions under which data were obtained, constitutes inadequate documentation.

- APA Task Force: reliability coefficients for data analyzed should be reported even when the focus of the research is not psychometric.
Inferential Statistics

- Inferences about characteristics of populations based on sample data
  - **Goal:** go beyond sample data; make inferences about population parameters
  - We use the laws of probability to make inferences about populations based on sample data
What Statistics Do I Choose?

Research Question

- Relationship between variables?
- Difference between groups or levels?

- Associational (crosstabulation, Correlation & Regression)
- Comparison (Hypothesis Testing)

Other very important questions you need to answer before we start:

- What kinds (nominal, ordinal, interval/ratio) of variables are being analyzed?
- Is there more than one independent/dependent variable?
- Is the data distributed normally? Can tests assumptions be met?
Association Methods
Association Methods

- Intended to determine the association between multiple variables
  - Contingency tables
  - Correlation
  - Partial correlation
  - Regression
Contingency Tables

- A.k.a. crosstabulations
  - Displays information in cells formed by the intersection of two or more categorical variables

Example of a Research Question: Is observed relationship between college major and gender statistically significant?

- Null hypothesis: the frequencies across all cells in the table are equal
- Alternative hypothesis: the frequencies across all cells in the table are not equal
Crosstabulation (Chi-Square)

- Determine whether relationship in contingency table is statistically significant
Correlation

- Association between variables
  - Variables can be ordinal, interval
  - Type of correlation coefficient is different dependent on the type of variable (e.g., ordinal, interval)

```
Pearson correlations:
     beckdepi mocscglt mohstglt
beckdepi  1.0000  0.0088  0.2005
mocscglt  0.0088  1.0000  0.7034
mohstglt  0.2005  0.7034  1.0000
```

Number of observations: 195
Partial correlations

- A correlation of two variables after controlling (holding constant) for the influence of a third variable

Partial correlations:

<table>
<thead>
<tr>
<th></th>
<th>beckdepi</th>
<th>mocsclgt</th>
<th>mohstgtlt</th>
</tr>
</thead>
<tbody>
<tr>
<td>beckdepi</td>
<td>0.00000</td>
<td>-0.18984</td>
<td>0.27332</td>
</tr>
<tr>
<td>mocsclgt</td>
<td>-0.18984</td>
<td>0.00000</td>
<td>0.71618</td>
</tr>
<tr>
<td>mohstgtlt</td>
<td>0.27332</td>
<td>0.71618</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

Number of observations: 195

Pairwise two-sided p-values:

<table>
<thead>
<tr>
<th></th>
<th>beckdepi</th>
<th>mocsclgt</th>
<th>mohstgtlt</th>
</tr>
</thead>
<tbody>
<tr>
<td>beckdepi</td>
<td>0.0080</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>mocsclgt</td>
<td>0.0080</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>mohstgtlt</td>
<td>0.0001</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>

Adjusted p-values (Holm's method):

<table>
<thead>
<tr>
<th></th>
<th>beckdepi</th>
<th>mocsclgt</th>
<th>mohstgtlt</th>
</tr>
</thead>
<tbody>
<tr>
<td>beckdepi</td>
<td>0.0080</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>mocsclgt</td>
<td>0.0080</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>mohstgtlt</td>
<td>0.0002</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>
Regression analysis

- Used to explain or predict values of quantitative dependent variable based on values of one or more independent or predictor variables.
Multiple regression

- More than one IV trying to predict a DV
- Partial Regression Coefficients: Regression coefficients in MR
  - Partial Regression Coefficient: Predict change in DV given one unit change in IV, controlling for the other IVs in equation
Regression

- Model is very powerful, but sometimes it can be misinterpreted
  - Association does not mean causation
- Also, its power depends on meeting assumptions
Differences between groups

Independent Groups
Repeated measures
Independent Groups (Between-subject designs)
Between-subject designs

- Assumptions
  - Independence (errors are independent)
  - Normality (errors are normally distributed)
  - Homogeneity of variance (distribution of errors has variance $\sigma_e^2$ in EACH population)
T-test for independent samples

- DV’s are interval or ratio scales (normally distributed)
- Make a regular practice of reporting effect size (d)

\[ d = \frac{\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}}}{s_{\text{pooled}}} \]

---

**Example:**

<table>
<thead>
<tr>
<th></th>
<th>12 hours</th>
<th>18 hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

**R-Commander Code:**

```R
# Levene's Test for Homogeneity of Variance (center = "mean")
leveneTest(steadin2 ~ group, data=dataset, center="mean")
# Df F value Pr(F)
# group 1 2.2203 0.1584
# 14

t.test(steadin2~group, alternative='two.sided', conf.level=.95,
       var.equal=FALSE, data=dataset)

# Welch Two Sample t-test
# data: steadin2 by group
# t = -2.3966, df = 10.155, p-value = 0.03718
# alternative hypothesis: true difference in means is not equal to 0
# 95 percent confidence interval:
# -3.856052 -0.1443948
# sample estimates:
# mean in group 12 hours mean in group 18 hours
# 4.25 6.25
```
Oneway ANOVA
One IV and 2 or more levels

- We could perform multiple t–tests among all the 2-group comparisons
  - This practice increases the probability of a type-I error (we reject the null hypothesis incorrectly)
    - In order to prevent type I error, we could use a Bonferroni correction
    - This correction will have an impact on power (it makes it more difficult to reject the null hypothesis)
- Instead, we use Oneway ANOVA
Post Hoc Tests in ANOVA

- **ANOVA is an omnibus test**
  - We know there is a difference, but we don’t know where

- **Post hoc testing question:** "Which pairs of means are significantly different?"

- **Example:**
  - Three comparisons made with these means
Factorial Designs

- Designs with more than one IV (factor)
- Have more information (we can examine interactions)
- More efficient and can be more powerful
- Control error variance by adding second variable
- More theory-driven
- Add complexity
- Increase size of study compared to a 2-group study
Repeated Measures designs
Repeated Measures

- Each participant tested more than once, under more than one condition
- Repeated Measures designs are extensions of paired-sample t-tests (matched-t-tests)
  - Subjects respond multiple times to the same DV
  - Repetition could be over time or over treatments
Repeated Measures: Pros

- Useful when one IV is \textit{change over time}
- Useful when population limited in size
- Reduces error variance

Repeated Measures: Cons

- More restrictive assumptions
  - And difficult to meet
- Attrition is a big problem
Matched t-test

- T-test for dependent or paired/matched samples
  - One IV, 2 levels (pre-test, posttest)
  - Computation of Effect size is recommended
Single-factor ANOVA with repeated measures

- One IV, 2 or more levels, all participants undergoing all conditions/levels
- DV is normally distributed, with equal variances
- **Critical assumption:** Sphericity (the correlations among all the different levels are equal)
  - Corrections when violated
Closing thoughts
R commander vs. SPSS

- R-commander represents a very good alternative to SPSS
  - User-friendly
  - Support
  - Frequent updates and plugins
  - FREE

- But R is much more powerful
  - Endless packages
  - Awesome support
  - Robust
  - FREE

- Steep learning curve, ONLY for programming
Questions?

Antonio Olmos
Kai Schramm
Priyalatha Govindasamy

Antonio.Olmos@du.edu
AntonioOlmos@aumhc.org
R commander vs. SPSS

- R-commander represents a very good alternative to SPSS
- But R is much more powerful
  - Much more options
  - Much more robust
- Steep learning curve, ONLY if you intend to program in R