

Assignment 4: Statistical Analysis

DESCRIPTION

In this assignment, you will practice methods you learned in class for descriptive and inferential statistical analyses. You will use the data you collected and the scales you constructed in the previous assignment (Assignment 3: Scale Construction) as a basis for your analysis. The methods I would like you to practice in this assignment are calculating descriptive statistics for the populations you are studying, correlations among your scales, and inferential analysis using appropriate t-tests or analyses of variance. Finally, you will report the findings from your analyses following APA guidelines. The lecture slides on statistical analysis provide examples of how to report results of a correlation or inferential test. The handout provides the list of **R** functions that you will need to use to complete the assignment. Alternatively, you can choose to use **JMP** to complete the assignment. Please refer to the last two lectures on statistics for descriptions and example uses of the JMP functions that correspond to those provided here.

EXPECTED OUTCOME

The expected outcome of this assignment is a one-page write-up, following APA guidelines, that details the process you followed and the findings from your analysis. Please attach to your report the *.csv* data file and R functions you used and their outputs in a *.txt* file. If you used JMP, please provide the *.jmp* data file and the model output as an *.rtf* file or as *.png* image files as attachments. To save the model output in JMP, select Edit > Journal and File > Export into an *.rtf* file. Upload your submission following the naming convention "Assignment4-StudentLastName.zip." Further directions are provided on the last page of this handout.

RESOURCES

See the Moodle page for resources including the statistical analysis lecture notes and the assignment submission template.

STEP 1: DESCRIPTIVE STATISTICS

The first set of analyses you will do with your data is what is called descriptive statistics. The goal of this form of analysis is to create quantitative summaries of your dataset. Descriptive statistics inform us of average measurements of a scale or item, how much these measurements vary across observations, the range of measurements, and other aspects you might want to know about your data. This form of analysis also includes creating visual representations of your data as box plots, bar graphs, pie charts, histograms, and scatterplots. Visual tools help the researcher identify anomalies, outliers, and trends in data.

R provides a host of methods to conduct descriptive statistics and create visual representations of your data. The following method from the “psych” package provides a full list of descriptive statistics for a given dataset.

```
library(psych)
describe(mydata_scales)
describe.by(mydata_scales, mydata_conditions)
```

R also provides a host of visualization options. Below are methods for creating some of the most common forms of descriptive visualizations. You can copy the graphs you create in R simply by copying them in R and pasting them in your Word Processor.

```
boxplot(mydata_scales)
hist(mydata_scales$scale)
plot(density(mydata_scales$scale))
barplot(table(mydata_scales$scale))
barplot(table(mydata_conditions, mydata_scales$scale))
barplot(table(mydata_conditions, mydata_scales$scale),
        beside=TRUE)
```

STEP 1 INSTRUCTIONS

Study your dataset using descriptive statistics and visualization methods in R or JMP. Note down any anomalies, outliers, or extreme skews or multimodalities in your distributions that you can identify using these methods.

STEP 2. CORRELATIONS

Next you will study how the scales and items in your data might correlate with each other. The goal of looking at correlations is to identify relationships between your measurements and suggest possible causal relationships. For instance, you might have measured how much people trust a particular kind of technology and how often they shop online. You might find that these two measures are highly correlated, suggesting that shopping online and trusting technology might have some relationship or might be affected by a shared latent variable. These suggestions inform further studies or analyses of your data. For instance, you might split your population into two groups as frequent online shoppers and infrequent online shoppers and conduct inferential statistics to test whether how much people shop has an effect on how much they trust technology.

The most common measure of correlation between variables is “Pearson’s correlation” calculated by dividing the covariance between the two measures by the product of their standard deviations. This operation creates an r-value (Pearson product-moment correlation coefficient), ranging from -1 to 1, where -1 represents a strong negative relationship, 0 represents no relationship, and 1 represents strong positive relationship between the two variables. However, we cannot make a conclusive statement on the relationship between these variables simply by looking at the r-value, because the r-value is extremely sensitive to the data distribution and population size. We can, however, conduct a significance test to whether a correlation is significant given its data distribution and population size. Significance tests typically produce reliable results with sample sizes above 50.

R provides methods to calculate correlations and significance tests. Below are commands that will help you calculate r-values and p-values for significance-testing.

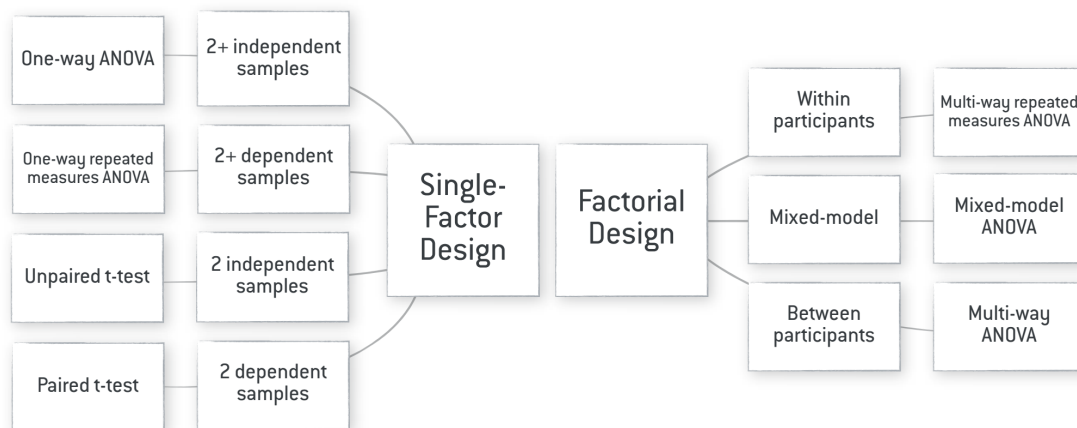
```
cor(mydata_scales, use="complete.obs", method="pearson")
      library(Hmisc)
      rcorr(as.matrix(mydata_scales), type="pearson")
```

STEP 2 INSTRUCTIONS

Identify correlations among your scales using R or JMP. Report high correlations with r-values and p-values.

STEP 3. INFERENCEAL STATISTICS

The final step in this assignment is to test hypotheses about your data using inferential tests. You will be using these tests to identify how your categorical variables affect your scales or response variables. As we discussed in class, what inferential test you should use will depend on your experimental design. The decision map below will help you choose the appropriate test for your purposes.



The first step in conducting your inferential analysis is to determine your null and alternative hypotheses. A null hypothesis suggests that your independent variables do not have any effect on your dependent variables or response variables. An alternative hypothesis suggests an effect. In your analysis, you will have to identify the “direction” of the effect suggested by your alternative hypothesis. For instance, if you have a univariate design with 2 independent samples, your alternative hypothesis has to identify whether the treatment condition in your independent variable will cause an increase, decrease, or simply a change in your dependent variable. If the alternative hypothesis suggests an increase, this analysis will use a one-tailed unpaired t-test.

R provides all the tools you will need to conduct inferential tests. Below is a list of functions that you can use to conduct these tests. Alternatively, you can use **JMP**. Examples of the use of these functions and descriptions and examples of corresponding JMP tests are provided in the related lectures. If your experiment has a multivariate design (i.e., more than one factors), you also need to conduct post-hoc analyses to make pairwise comparisons across conditions.

UNPAIRED, ONE-TAILED T-TEST

Use to compare two between-participants conditions where the alternative hypothesis suggests that the treatment condition causes an increase in the dependent variable:

```
t.test(mydata_scales~mydata_conditions, alternative="less",
```

```
paired = FALSE, var.equal=TRUE, conf.level=.95)
```

PAIRED, ONE-TAILED T-TEST

Use to compare two within-participants measurements (e.g., before stimuli and after stimuli) where the alternative hypothesis suggests that the stimuli causes an increase in the dependent variable:

```
t.test(mydata_scales$measurement1, mydata_scales$measurement2,
       alternative="less", paired=TRUE, conf.level=0.95)
```

ONE-WAY, BETWEEN-PARTICIPANTS ANOVA

Use to compare more than two between-participants samples. Tukey's HSD test can be used to conduct pairwise comparisons between conditions.

```
model = aov(mydata_scales~mydata_conditions, data=dataFile)
summary(model)
TukeyHSD(model)
```

ONE-WAY, WITHIN-PARTICIPANTS ANOVA

Use to compare more than two withinparticipants samples. Tukey's HSD test can be used to conduct pairwise comparisons between conditions; in this case, using "generalized linear models."

```
model =
aov(mydata_scales~mydata_conditions+Error(Subject/mydata_conditions),
    data=dataFile)
summary(model)
model2 = lme(mydata_scales~mydata_conditions, random =
~1|Subject/mydata_conditions, data=dataFile)
model3 <- glht(model2, linfct=mcp(mydata_conditions="Tukey"))
summary(model3)
```

TWO-WAY, BETWEEN-PARTICIPANTS ANOVA

Use to test how two or more factors affect between-participants samples. The test will look for main effects as well as interactions. You might need to simplify your model following the guidelines provided in the lecture on ANOVA.

```

model = aov(mydata_scales~
mydata_conditions1*mydata_conditions1,
            data=dataFile)
summary(model)

```

TWO-WAY, WITHIN-PARTICIPANTS ANOVA

Use to test how two or more factors affect within-participants samples. The test will look for main effects as well as interactions. You might need to simplify your model following the guidelines provided in the lecture on ANOVA.

```

model = aov(mydata_scales~
(mydata_conditions1*mydata_conditions1) +
Error(Subject/(mydata_conditions1*mydata_conditions1)),
            data=dataFile)
summary(model)

```

TWO-WAY, MIXED-MODEL ANOVA¹

Use to test how a between-participants treatment and a within-participants treatment affect the response variable. The test will calculate main effects and interactions. You might need to simplify your model following the guidelines provided in the lecture on ANOVA.

```

model = aov(mydata_scales~
(mydata_conditions1*mydata_conditions1) +
Error(Subject/mydata_conditions1) + mydata_conditions1),
            data=dataFile)
summary(model)

```

VISUALIZATION

As a part of your analysis, you should create graphics that visually support the results of your analysis. The most common form of visualization used in behavior research is bar graphs. The R package “`sciplot`” includes a method that you can use to create bar graphs. The method will allow you to determine the range of the graph, create groupings, and manipulate a number of visual elements of the graph.

```

library(sciplot)
bargraph.CI(mydata_conditions$condition1, mydata_scales$scale1,
            ylim=c(0,25), col="Red")

```

¹Alternative methods for conducting mixed-model tests and post-hoc comparisons for multi-way ANOVA are available in the lecture slides.

```
bargraph.CI(mydata_conditions$condition1, mydata_scales$scale1,  
  ylim=c(0,25), col="Red", group=mydata_conditions$condition2)
```

STEP 3 INSTRUCTIONS

Choose the appropriate inferential test based on your experimental design using the decision map. If necessary, conduct post-hoc comparisons across conditions. Create bar graphs to visually support your results. You can use the R functions provided above or JMP functions described in the lectures.

STEP 4. REPORTING USING APA GUIDELINES

The APA (American Psychological Association) provides a style guide for academic writing in behavioral sciences and HRI research follows these guidelines. In this assignment, you will write a report of your findings that resembles the “Results” section of an empirical paper and follow these guidelines. The “Resources” section of this assignment on Moodle provides links to APA style guides. The lecture slides on statistical analysis also provides examples of how to report results of a correlation or inferential test.

STEP 4 INSTRUCTIONS

Write up your results from all three sets of analyses you applied to your data. Follow APA guidelines for reporting statistical results and creating graphics.

WRITE-UP

Write a one-page report of your results. This report should resemble the “Results” section of an empirical paper and follow the assignment template. For examples on reporting statistical tests and results, you can refer to the links provided on the assignment description on Moodle, the slides for lectures on statistics, and the papers we have discussed in class as examples. Be sure to follow in your submission the naming convention “Assignment4-StudentLastName.zip” and include the following:

In your write-up as a .docx file:

- Results from descriptive analyses
- Results from correlational tests
- Results from inferential tests
- Tables or figures supporting descriptive and inferential tests

As appendices included in the .zip file:

- If you use R, the .csv data file and a .txt file with a list of the R commands you ran and the outputs
- If you use JMP, the .jmp data file and a single .rtf file with the outputs from your tests

Submit your .zip file to the Assignment 4 dropbox on Moodle.

GRADING (100 POINTS TOTAL)

- Calculating descriptive statistics (20/100)
- Calculating correlations (20/100)
- Conducting inferential tests (20/100)
 - Determining the appropriate test for the required analysis (8/20)
 - Conducting appropriate tests, pairwise comparisons, and model simplification (if necessary) (6/20)
 - Creating graphics to visually support the results (6/20)
- Following APA guidelines in reporting all results (20/100)
- Quality of the written report (20/100)