

# Psy 300 Statistics Workbook

James Van Slyke

2021-06-01



# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Independent t-test</b>	<b>7</b>
<b>3</b>	<b>Paired samples t-test</b>	<b>11</b>
<b>4</b>	<b>One Way ANOVA</b>	<b>15</b>
<b>5</b>	<b>Two-Way ANOVA</b>	<b>19</b>



# Chapter 1

## Introduction

This is a workbook of various examples of statistical analysis to explore psychology experiments. The workbook uses R for all its calculations and R Studio is assumed to be the program used for those calculations.

If you have not downloaded R and R Studio you can go to these websites

R

R Studio

On the left you'll see examples from each of the primary statistical analyses used in this class. Two other books will be important to use as references.

R for Data Science

Using R to something something Psychology



## Chapter 2

# Independent t-test

Independent t-test example

1. First Step, upload dataset from SPSS
2. Get data set named “Invisibility” from SPSS datasets
3. Use *import dataset* tool under the environment tab
4. Find file called *invisibility.sav*

```
library(haven)
Invisibility <- read_sav("Invisibility.sav")
View(Invisibility)
```

Inspect variables

```
Invisibility$Cloak
Invisibility$Mischief
```

Create bar graph to examine data First step is to create a dataset with your descriptive variables We'll use the dplyr package to do this, which is part of tidyverse

```
library(dplyr)
Invis_Descriptives <- Invisibility %>%
  group_by(Cloak) %>%
  summarize(n = n(),
            mean = mean(Mischief),
            sd = sd(Mischief),
            se = sd / sqrt(n),
            ci = qt(0.975, df = n - 1) * sd / sqrt(n))
```

First let's make a simple graph with just the basics

```
ggplot(Invis_Descriptives,
       aes(x = Cloak,
           y = mean)) +
  geom_bar(stat = "identity")
```

Next let's add error bars

```
ggplot(Invis_Descriptives,
       aes(x = Cloak,
           y = mean)) +
  geom_bar(stat = "identity") +
  geom_errorbar(aes(ymin=mean-ci,
                   ymax=mean+ci))
```

We can add in labels to improve the look of our chart. We've added labels to our factor variable and labels for the title and x and y variables.

```
ggplot(Invis_Descriptives,
       aes(x = factor(Cloak, labels=c("No Cloak", "Cloak")),
           y = mean)) +
  geom_bar(stat = "identity") +
  geom_errorbar(aes(ymin=mean-ci,
                   ymax=mean+ci)) +
  labs(title = "Mean Number of Mischievous Acts with or without Cloak",
       y="Mean Number of Mischievous Acts", x="Were They Wearing a Cloak?")
```

Next let's add some color to our chart

```
ggplot(Invis_Descriptives,
       aes(x = factor(Cloak, labels=c("No Cloak", "Cloak")),
           y = mean)) +
  theme_minimal() +
  geom_bar(stat = "identity", fill="cornflowerblue") +
  geom_errorbar(aes(ymin=mean-ci,
                   ymax=mean+ci), width=.3, size=1) +
  labs(title = "Mean Number of Mischievous Acts with or without Cloak",
       y="Mean Number of Mischievous Acts", x="Were They Wearing a Cloak?")
```

Based on the graph, it looks like there is a difference between the groups. Let's run a t test to make sure.

```
t.test(Mischief ~ Cloak, data = Invisibility)
```

List the effect size as well. Cohen's d - Subtract the means from each other and divide by the standard deviation of the control group.

```
Cohens_d <- (5.00-3.75)/1.91
```

Write out conclusion



*On average, participants given a cloak of invisibility engaged in more acts of mischief ( $M = 5$ ,  $SE = 0.48$ ), than those not given a cloak ( $M = 3.75$ ,  $SE = 0.55$ ). This difference, 1.25, 95% CI[-2.76, 0.26], was not significant  $t(21.54) = -1.71$ ,  $p = 0.101$ ; however, it did represent a medium-sized effect  $d = 0.65$ .*

Steps to writing results

1. Write out the means and include standard error
2. Write out the difference between the means (Subtract the sample means) and the confidence intervals for the difference
3.  $t(df) = t$  score,  $p$  value, Cohen's  $d$



## Chapter 3

# Paired samples t-test

Check out dataset [View Spreadsheet](#)

```
View(ch14ds1)
```

Info on dataset [Research hypothesis -](#)

[View in Console](#)

```
ch14ds1
```

[View Descriptive Statistics](#)

```
summary(ch14ds1)
```

Graph data using a boxplot to get a first glimpse

```
boxplot(ch14ds1$Pretest, ch14ds1$Posttest)
```

Run the t test for paired samples (also sometimes called the Dependent t test)

```
t.test(ch14ds1$Posttest, ch14ds1$Pretest, paired = TRUE)
```

This time let's use a ggplot to make a better looking bar graph. Check out the format of the data

```
ch14ds1
```

The data is set up for a paired t test, so it includes 2 numeric variables. This is a problem because we are missing a single factor variable to describe our two bars (i.e. pre and post test). Thus, we are going to create a second dataset that has a factor ("group") and numeric ("Test") variable. Notice I'm creating a new dataframe, calling it "TestData". I'm creating the factor variable from scratch using the "rep" repeat function and then adding together the Pre and Post Test scores into a single number variable.

```
TestData <- data.frame(
  group= rep(c("Pretest", "Posttest"), each=25),
  Test = c(ch14ds1$Pretest, ch14ds1$Posttest)
)
```

Next, I'll find my descriptive statistics for this dataset

```
library(dplyr)
Test_Descriptives <- TestData %>%
  group_by(group) %>%
  summarize(n = n(),
    mean = mean(Test),
    sd = sd(Test),
    se = sd / sqrt(n),
    ci = qt(0.975, df = n - 1) * sd / sqrt(n))
```

Check your descriptives, make sure they turned out ok.

```
Test_Descriptives
```

Finally, create a basic bar graph

```
ggplot(Test_Descriptives,
  aes(x = group,
    y = mean)) +
  geom_bar(stat = "identity")
```

Then let's add in some error bars with the `geom_errorbar` function

```
ggplot(Test_Descriptives,
  aes(x = group,
    y = mean)) +
  geom_bar(stat = "identity") +
  geom_errorbar(aes(ymin=mean-ci,
    ymax=mean+ci))
```

Finally, let's make it look really nice

```
ggplot(Test_Descriptives,
  aes(x = group,
    y = mean)) +
  theme_minimal() +
  geom_bar(stat = "identity", fill="cornflowerblue") +
  geom_errorbar(aes(ymin=mean-ci,
    ymax=mean+ci), width=.3, size=1) +
  labs(title = "Change in Performance from Pre to Post Tests",
    y="Mean Score", x="Tests")
```

You can also flip the columns to make it look cleaner

```
ggplot(Test_Descriptives,
       aes(x = group,
           y = mean)) +
  theme_minimal() +
  geom_bar(stat = "identity", fill="cornflowerblue") +
  geom_errorbar(aes(ymin=mean-ci,
                   ymax=mean+ci), width=.3, size=1) +
  labs(title = "Change in Performance from Pre to Post Tests",
       y="Mean Score", x="Tests") +
  scale_x_discrete(limits=c("Pretest", "Posttest"))
```



## Chapter 4

# One Way ANOVA

Practice dataset: How does preschool affect language development? Three groups that differ by how many hours they spent in preschool per week

Variables IV = Preschool DV = Language Development Language development was measured based on a language development test score

Null Hypothesis: Attendance at preschool has no effect on language development

View Dataset

```
View(ch15ds1)
```

View in R Console

```
ch15ds1
```

Let's look at a bar graph of the data first Step 1 - create table of descriptive statistics

```
library(dplyr)
Preschool_Descriptives <- ch15ds1 %>%
  group_by(Group) %>%
  summarize(n = n(),
            mean = mean(Language.Score),
            sd = sd(Language.Score),
            se = sd / sqrt(n),
            ci = qt(0.975, df = n - 1) * sd / sqrt(n))
```

Check it out

```
Preschool_Descriptives
```

Now graph it based on the descriptive statistics

```
ggplot(Preschool_Descriptives,
       aes(x = Group,
           y = mean)) +
  geom_bar(stat = "identity") +
  geom_errorbar(aes(ymin=mean-ci,
                   ymax=mean+ci))
```

Run your ANOVA First install Psych package

```
library(psych)
```

Check out the variables Psych package uses a different command for this “describeBy” Must specify DV and groups

```
describeBy(ch15ds1$Language.Score, group = ch15ds1$Group)
```

Run your ANOVA! Something new, must save results in an object rather than the results being there automatically

```
m1 <- aov(Language.Score~Group, data = ch15ds1)
```

To get ANOVA table must use “summary” function

```
summary(m1)
```

Figure out the effect size - Eta squared The formula is  $SS_{\text{between}}/SS_{\text{Total}}$  or  $SS_{\text{between}}/(SS_{\text{between}}+SS_{\text{Residual}})$

```
1133/(1133+1738)
```

Write out conclusion

Number of hours in Preschool had a significant effect on language development,  $F(2, 27) = 8.799$ ,  $p = 0.00114$ ,  $\eta^2 = 0.39$ .

Where is the difference? Need to use post hoc tests TukeyHSD will tell us where the differences are between the individual groups. Run TukeyHSD on saved ANOVA results

```
TukeyHSD(m1)
```

Finally, write out the whole conclusion TukeyHSD post hoc tests revealed that 20 hours a week of preschool ( $M=91.6$ ,  $SE=1.96$ ) resulted in significantly higher levels of language development in comparison to 5 hours ( $M=76.6$ ,  $SE=3.78$ ). This difference, -15 95% CI[-23.90, -6.10] was significant with an adjusted  $p = .0008$ .

Improve the graph

```
ggplot(Preschool_Descriptives,
       aes(x = Group,
           y = mean)) +
```



```
geom_bar(stat = "identity") +
geom_errorbar(aes(ymin=mean-ci,
                  ymax=mean+ci)) +
scale_x_discrete(limits=c("5 Hours", "10 Hours", "20 Hours"))
```

Make it look real nice Clark

```
ggplot(Preschool_Descriptives,
       aes(x = Group,
           y = mean)) +
theme_minimal() +
geom_bar(stat = "identity", fill="steelblue") +
geom_errorbar(aes(ymin=mean-ci,
                  ymax=mean+ci), width=.3, size=1) +
labs(title = "Does Preschool Effect Language Development?",
      y="Mean Score on Language Test", x="Number of Hours Spent in Preschool") +
scale_x_discrete(limits=c("5 Hours", "10 Hours", "20 Hours"))
```



## Chapter 5

# Two-Way ANOVA

Testing the effects of Alcohol and Gender on the “beer goggles effect” # Two Independent Variables (Main Effects) # IV 1 = Alcohol 3 Levels (None, 2 Pints, 4 Pints) IV 2 = Gender 2 Levels (Male and Female) DV = Attractiveness of the partner selected at the end of the evening

Hypotheses H1 Alcohol has an effect on the attractiveness level of the selected partner H2 Gender has an effect on the attractiveness level of the partner. H3 There is an interaction effect between Alcohol and Gender

Get the dataset and import it

```
library(haven)
goggles <- read_sav("~/iCloud Drive (Archive)/Documents/FPU/Spring 2017/Statistics Psy 300 SP17/S
View(goggles)
```

Check out Gender variable

```
goggles$Gender
```

Check out Alcohol variable

```
goggles$Alcohol
```

Check out Attractiveness variable

```
goggles$Attractiveness
```

Alcohol and Gender variables are factors When they get imported from SPSS they don't function as well because the focus is on the numbers, not the labels or words We can use tidyverse and the mutate function to fix this.

```
goggles <- goggles %>%
  mutate(Gender = factor(Gender, levels = c(0,1),
    labels = c("Male", "Female")))
```

We can do the same thing with the Alcohol variable. Make sure you know the levels or numbering of the variable.

```
goggles$Alcohol
```

Then go ahead and mutate the variable as you did with Gender.

```
goggles <- goggles %>%
  mutate(Alcohol = factor(Alcohol, levels = c(1,2,3),
    labels = c("None", "2 Pints", "4 Pints")))
```

Now we can move to the Two-Way analysis. “m3” is the new object to save results so the formula will have this structure: `m3 <- aovfunction`

```
m3 <- aov(Attractiveness ~ Gender + Alcohol +
  Gender*Alcohol, data = goggles)
```

Key Findings from our output: Main effect of Alcohol on attractiveness. Interaction effect of Alcohol and Gender. What does this mean?

Let’s use a graph to understand this better. First let’s look at each variable individually.

```
GogglesAlcohol <- goggles %>%
  group_by(Alcohol) %>%
  summarize(n = n(),
    mean = mean(Attractiveness),
    sd = sd(Attractiveness),
    se = sd/sqrt(n),
    ci = qt(0.975, df = n - 1) * sd / sqrt(n))
```

Graph the Alcohol variable individually.

```
ggplot(GogglesAlcohol, aes(x = Alcohol,
  y = mean)) +
  geom_point(size = 3) +
  geom_line(size = 1) +
  geom_errorbar(aes(ymin = mean - ci,
    ymax = mean + ci),
    width = .1)
```

Graph the Gender variable individually.

```
GogglesGender <- goggles %>%
  group_by(Gender) %>%
  summarize(n = n(),
    mean = mean(Attractiveness),
    sd = sd(Attractiveness),
```

```
se = sd/sqrt(n),
ci = qt(0.975, df = n - 1) * sd / sqrt(n))
```

Graph it

```
ggplot(GogglesGender, aes(x = Gender,
                          y = mean)) +
  geom_point(size = 3) +
  geom_line(size = 1) +
  geom_errorbar(aes(ymin = mean - ci,
                    ymax = mean + ci),
                width = .1) +
  ylim(0,70)
```

Finally, we can graph relationships for both variables First find your descriptive statistics, but this time based on two independent variables

```
GogglesDescriptives <- goggles %>%
  group_by(Alcohol, Gender) %>%
  summarize(n = n(),
            mean = mean(Attractiveness),
            sd = sd(Attractiveness),
            se = sd/sqrt(n),
            ci = qt(0.975, df = n - 1) * sd / sqrt(n))
```

Check it

```
GogglesDescriptives
```

Use a line graph to graph the relationship

```
ggplot(GogglesDescriptives, aes(x = Alcohol,
                                y = mean,
                                group=Gender,
                                color=Gender)) +
  geom_point(size = 3) +
  geom_line(size = 1) +
  geom_errorbar(aes(ymin = mean - ci,
                    ymax = mean + ci),
                width = .1)
```

Use dodge functions to make graph clearer

```
pd <- position_dodge(0.2)
ggplot(GogglesDescriptives,
      aes(x = Alcohol,
          y = mean,
          group=Gender,
          color=Gender)) +
```

```
geom_point(position = pd,
           size = 3) +
geom_line(position = pd,
          size = 1) +
geom_errorbar(aes(ymin = mean - ci,
                  ymax = mean + ci),
              width = .1,
              position = pd)
```

All the bells and whistles

```
pd <- position_dodge(0.2)
ggplot(GogglesDescriptives,
       aes(x = Alcohol,
           y = mean,
           group=Gender,
           color=Gender)) +
geom_point(position=pd,
           size = 3) +
geom_line(position = pd,
          size = 1) +
geom_errorbar(aes(ymin = mean - ci,
                  ymax = mean + ci),
              width = .1,
              position = pd,
              size = 1) +
scale_color_brewer(palette="Set1") +
theme_minimal() +
labs(title = "Beer Goggles Effect",
     x = "Amount of Alcohol Consumed",
     y = "Mean Level of Attractiveness",
     color = "Gender")
```

Use Tukey to look at specific differences in the groups you are interested in

```
TukeyHSD(m3)
```

For us, the most important difference is that males are more likely to choose a less attractive person at 4 pints of alcohol than females

How to write results section: First write out the main effects There was a significant main effect of the amount of alcohol consumed on the attractiveness of the date that was selected,  $F(2, 42) = 20.07$ ,  $p < .001$ . There was not a significant main effect of gender on the attractiveness of the date that was selected,  $F(1, 42) = 2.03$ ,  $p = .161$ . Second, write out the interaction effect There was a significant interaction effect between amount of alcohol consumed and gender on the attractiveness of the date that was selected  $F(2, 42) = 11.91$ ,  $p < .001$ .

Third, write out any relevant Tukey findings TukeyHSD post hoc tests revealed that at the largest amount of alcohol consumption (4 pints) Males were significantly more likely to choose a less attractive date ( $M=35.6$ ,  $SE=3.83$ ) in comparison to females ( $M=57.5$ ,  $SE=57.5$ ). This difference, 21.86, 95% CI[-23.90, -6.10] was significant with an adjusted  $p = .0003$ .

## Introduction

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter ???. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter ???.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

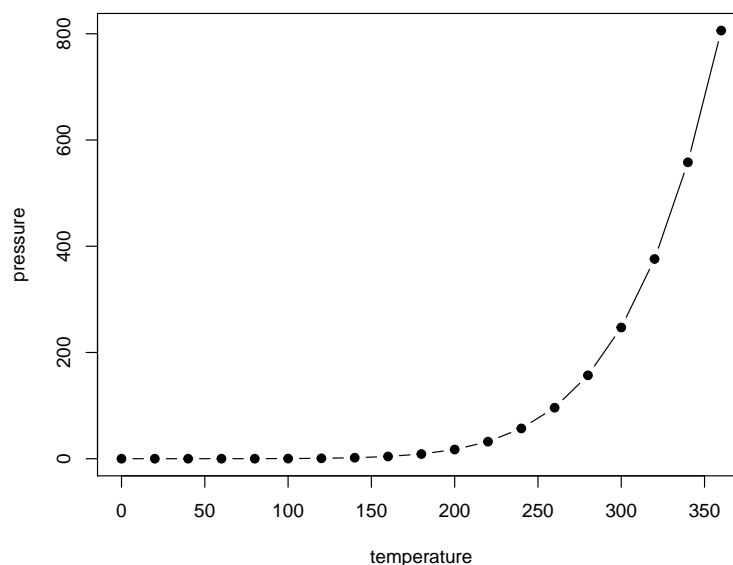


Figure 5.1: Here is a nice figure!

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 5.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 5.1.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package

Table 5.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa



(Xie, 2021) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

#Front Matter from Example

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports, e.g., a math equation  $a^2 + b^2 = c^2$ .

The **bookdown** package can be installed from CRAN or Github:

```
install.packages("bookdown")  
# or the development version  
# devtools::install_github("rstudio/bookdown")
```

Remember each Rmd file contains one and only one chapter, and a chapter is defined by the first-level heading #.

To compile this example to PDF, you need XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): <https://yihui.org/tinytex/>.



# Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2021). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.22.