

**Directions:** Complete the exercises below. When you are finished, turn in any required files online in Canvas, then check-in with the Lab TA for dismissal.

---

## Testing Contrasts in R

In this part of the lab, we will learn how to test contrasts in R using the data from an experiment on microwave popcorn. As a reminder, a group of students randomly assigned a total of 18 bags of popcorn to microwave at one of three times (**Time**): Short, Medium and Long. The dataset containing the volume (in cl) of popcorn produced for each of the 18 bags (**Volume**) organized by Time is given in the `popcorn.csv` file posted in Canvas.

- First, load in the data using the *Import Dataset* tool in R Studio. Be sure to change the variable type on the **Time** column to “factor” and enter “Short, Medium, Long” as the levels:

```
library(readr)
popcorn <- read_csv("popcorn.csv", col_types = cols(Time =
                                                         col_factor(levels = c("Short", "Medium", "Long"))))
```

- Then, make a scatterplot of the volume of popcorn produced for each time using the `plot()` function:

```
plot(y=popcorn$Volume, x=as.numeric(popcorn$Time), xlim=c(0.5,3.5), ylim=c(50,150),
      xlab="Time", ylab="Volume", xaxt='n', main="Microwave Popcorn Experiment")
axis(1, at=c(1,2,3), labels=c("Short", "Medium", "Long"))
```

- Next, run the code from last time to conduct the ANOVA using the `aov()` function. Then, inferences for single group means can be conducted using the `emmeans()`, `confint()`, and `test()` functions in the `emmeans` package in R. If this is your first time using this package, then you will need to install it first using `install.packages("emmeans")`:

```
my.aov <- aov(Volume~Time, data=popcorn)
summary(my.aov)
library(emmeans)
single.means <- emmeans(my.aov, "Time")
confint(single.means)
test(single.means)
```

A further description of the code and its output follows:

- The `emmeans()` function is used to save the necessary summary statistics needed to compute the single group mean inferences and it works by first inputting the ANOVA object (`my.aov`) and then naming the grouping variable (`"Time"`).
- The `confint()` function takes the `emmeans` object (`single.means`) as its only input and outputs a table containing the 95% confidence intervals for each individual group mean  $\mu_i$ .
- The `test()` function takes the `emmeans` object (`single.means`) as its only input and outputs a table containing the  $t$ -test results for  $H_0 : \mu_i = 0$ .

- Now, you can test contrasts for your data using the `contrast()` function which takes as its first input the `single.means` object and as its second input the `method=` option:

```
contrast(single.means, method="poly")
contrast(single.means, method=list(c(-1,0,1),
                                   c(1,-2,1)))
contrast(single.means, method=list("Long - Short" = c(-1,0,1),
                                   "(Long+Short)-2*Medium" = c(1,-2,1)))
contrast(single.means, method=list("Long - Short" = c(-1,0,1),
                                   "(Long+Short)/2-Medium" = c(0.5,-1,0.5)))
```

- The orthogonal polynomial contrasts can be obtained automatically using the `method="poly"` option.
- This option is the same as specifying the contrasts directly by giving the coefficients in a list, as shown in the second option.
- You can even name the contrasts with a descriptive title that will show up in the output, as shown in the third option.
- Finally, you can re-scale the contrasts by changing the values of the coefficients, as shown in the fourth and final option.

The output for each of the commands will show a table containing the (un-adjusted) results for the  $t$ -test of each contrast listed.

- Lastly, if you would like to see the 95% confidence interval for any of the contrasts, you can use the `confint()` function:

```
confint(contrast(single.means, method="poly"))
```

## Multiple Comparisons in R

In this part of the lab, we will continue using the popcorn example to illustrate how to perform pairwise and multiple comparison adjustment procedures in R.

- The pairwise and multiple comparison testing procedures can be conducted using the `pairs()` function which takes the `single.means` emmeans object as its first input and the `adjust=` option as its second input:

```
pairs(single.means, adjust=NULL)
pairs(single.means, adjust="tukey")
pairs(single.means, adjust="scheffe")
pairs(single.means, adjust="bonferroni")
```

The output table will show the tests for each one of the pairwise comparisons of group means given the following values for the `adjust=` option:

- The `NULL` option provides the results of the un-adjusted pairwise *t*-tests, which will correspond to the results of the Least Significant Difference method.
  - The `"tukey"` option provides the results for the Tukey-Kramer Honest Significant Difference (HSD) method of multiple comparison adjustment.
  - The `"scheffe"` option provides the results for the Scheffe method of multiple comparison adjustment.
  - The `"bonferroni"` option provides the results for the Bonferroni method of multiple comparison adjustment.
- Multiple comparison adjustments, more generally, can be conducted using the `contrast()` function which takes the `single.means` emmeans object as its first input, the `adjust=` option to specify the correction, and a list of contrasts for the `method` option. For example, the code below will provide tests and simultaneous confidence intervals using the Scheffe method.

```
adjC = contrast(single.means, adjust="scheffe",
               method=list("S vs. M & L" = c(-2,1,1),
                           "M vs. S & L" = c(1,-2,1),
                           "L vs. S & M" = c(1,1,-1)))
summary(adjC)
confint(adjC)
```

You can compare these to the unadjusted results by omitting the `adjust=` option.

```
regC = contrast(single.means,
               method=list("S vs. M & L" = c(-2,1,1),
                           "M vs. S & L" = c(1,-2,1),
                           "L vs. S & M" = c(1,1,-1)))
summary(regC)
confint(regC)
```

## Matched Pairs Analysis in R

In this part of the lab, we will revisit the cholesterol example to learn how to conduct an analysis of a matched-pairs study using R. As a reminder, the dataset can be found in the `cholesterol.csv` file posted in Canvas. This data frame presents data on 18 randomly sampled individuals diagnosed with high cholesterol who replaced butter in their diets with a brand (A or B) of margarine. The brand of margarine was randomized. Their doctors recorded their blood cholesterol levels at the beginning of the experiment, after four weeks of their diets, and again after eight weeks. Rather than investigating the different between the two different types of margarine, the researchers are interested in whether replacing butter with margarine can significantly lower cholesterol levels.

- First, load in the data using the *Import Dataset* tool in R Studio:

```
library(readr)
cholesterol <- read_csv("cholesterol.csv")
```

- Then, plot a scatterplot of the cholesterol levels before switching to margarine and after 4 weeks of the diet using the `plot()` function and compute the correlation using the `cor()` function:

```
plot(cholesterol$Before, cholesterol$After4weeks, main="Cholesterol Levels",
      xlab="Before (Control)", ylab="After 4 Weeks")
cor(cholesterol$Before, cholesterol$After4weeks)
```

The output should show that the groups are not independent, so the traditional  $t$ -test is not appropriate.

- Next, we can conduct the appropriate test for the matched-pairs design and compute the corresponding 95% confidence interval using the `t.test()` function with the `paired=TRUE` option:

```
t.test(cholesterol$Before, cholesterol$After4weeks, paired=TRUE)
```

Alternatively, you can compute the differences yourself and then conduct the one-sample  $t$ -test to achieve the same result:

```
d = cholesterol$Before - cholesterol$After4weeks
t.test(d)
```

- Now, we can use the calculated differences to check the normality assumption:

- You can plot the histogram and normal Q-Q plot using the code:

```
hist(d)
qqnorm(d)
qqline(d, col="red")
```

and then check to see if the histogram is approximately symmetric and bell-shaped and whether the plots in the Q-Q plot follow the red reference line.

- You can compute the summary statistics using the code:

```
mean(d)
median(d)
skewness(d)
kurtosis(d)-3
```

and then check to see if the mean and median are similar and if skew and excess kurtosis are near zero.

- You can perform the Shapiro-Wilk test for normality using the code:

```
shapiro.test(d)
```

and then check the  $p$ -value to see whether you would reject the null hypothesis of normality.

- Finally, if any of the diagnostics indicate that this assumption isn't met, you can conduct the non-parametric tests in R.

- The Wilcoxon signed rank test can be performed using the `wilcox.test()` function with the `paired=TRUE` option:

```
wilcox.test(cholesterol$Before, cholesterol$After4weeks, paired=TRUE)
```

or equivalently,

```
wilcox.test(d)
```

- The sign test can be performed using the `binom.test()` function:

```
binom.test(sum(d>0), length(d))
```

where the first input is the number of differences with a positive sign and the second input is the total number of observations. The test is for  $H_0 : Pr(+) = Pr(-) = 0.5$  against  $H_a : Pr(+) \neq Pr(-)$  and the output of the test will compute the  $p$ -value using the binomial formula described in the lecture slides.

Or equivalently, you can conduct the sign test using the `signtest()` function in the "nonpar" library (which you may need to install first using `install.packages("nonpar")`):

```
library(nonpar)
```

```
signtest(d)
```

If you want to use the normal approximation to the binomial distribution (when you have a large sample size), then counterintuitively, you can use the `exact=TRUE` option:

```
signtest(d, exact=TRUE)
```

## Assignment

1. Run the code you created in R Studio for testing contrasts in the popcorn example to complete the following exercises:
  - (a) Construct a scatterplot of the volumes popped for each time. Is there a trend in volume over time?
  - (b) For each of the single group means  $\mu_i$ , conduct the  $t$ -test of  $H_0 : \mu_i = 0$  and interpret the result in the context of the study.
  - (c) For each of the single group means  $\mu_i$ , find and interpret the 95% confidence interval in the context of the study.
  - (d) Suppose the researchers are interested in testing the orthogonal polynomial contrasts. Conduct the test and interpret the results for these contrasts.
  - (e) Using what you've learned, add code to your program to test the *sequential* contrasts and interpret the results:  $\gamma_1 = \mu_{\text{Long}} - \mu_{\text{Medium}}$  and  $\gamma_2 = \mu_{\text{Medium}} - \mu_{\text{Short}}$ .
  - (f) Are the contrasts in part (e) orthogonal? Support your choice with relevant calculations.
2. Run the code you created in R Studio for multiple comparisons in the popcorn example to complete the following exercises:
  - (a) Summarize your findings from the pairwise comparison analysis of the mean volume of popcorn popped for the 3 different microwave times.
  - (b) Summarize your findings from the Bonferroni procedure for multiple comparisons.
  - (c) Summarize your findings from the Scheffe procedure for multiple comparisons.
  - (d) Summarize your findings from the Tukey-Kramer HSD procedure for multiple comparisons.
  - (e) How do your findings compare between the methods in parts (b), (c), (d)?
3. Run the code you created in R Studio for matched pairs analysis in the cholesterol example to complete the following exercises:
  - (a) Describe the blocks being used in this experiment.
  - (b) Explain why this is a matched pairs experiment.
  - (c) Conduct the  $t$ -test to determine whether or not the mean cholesterol level is lower when after replacing butter with margarine. Be sure to include the null and alternative hypothesis, test statistic,  $p$ -value, and interpret the result in the context of the study. (Use  $\alpha = 0.05$  if you opt for the *statistical significance* method of interpretation.)
  - (d) Check the assumptions for this hypothesis test. Summarize your findings.
  - (e) Perform the Wilcoxon signed rank test. Be sure to include the null and alternative hypothesis, test statistic,  $p$ -value, and interpret the result in the context of the study.
  - (f) Perform the sign test. Be sure to include the null and alternative hypothesis, test statistic,  $p$ -value, and interpret the result in the context of the study.

**Total:** 50 points      **# correct:** \_\_\_\_\_      **%:** \_\_\_\_\_