

ST 411/511 Homework 4

E. Alex Soderquist

Summer 2022

Instructions

This assignment is due by 11:59 PM, August 5th on Canvas via Gradescope. **You should submit your assignment as a PDF which you can compile using the provide .Rmd (R Markdown) template.**

Note: Create a PDF by either compiling a PDF directly (via LaTeX) or from a Word Doc. Do not submit a PDF of the HTML output as they're cumbersome and difficult to read.

Include your code in your solutions and indicate where the solutions for individual problems are located when uploading into Gradescope (Failure to do so will result in point deductions). You should also write complete, grammatically correct sentences for your solutions.

Once you've completed the assignment, and before you submit it to Gradescope, you should read the document to ensure that (1) The computed values show up in the document, (2) the document "looks nice" (i.e. doesn't have extraneous code/outputs and includes *just* the essentials), and (3) ensure the document is "easy" to read.

Goals:

1. Practice using, what we often think of as, alternatives to the t-tests.
2. Familiarize ourselves with how the underlying hypotheses are different from the t-test, and amongst the alternatives, to better understand what our conclusions mean.
3. Practice hypothesis testing for equality of "scales" using Levene's test.

Question 1 (14 points): One Sample Tests

Consider the following data ($n = 18$ observations) (Assume $\alpha = 0.05$ unless otherwise specified)

```
data1 <- c(5.4, -13.2, -4.3, 1.3, -14.3, -12.3, -11.4, 9.1, 2.5,  
         -6.8, 2.6, 6.2, 10.4, -17.0, -17.1, -8.0, 16.5, -15.3)
```

(a) (5 points) Use the sign test to test the null hypothesis that the population median is 0 ($H_0 : m = 0$) using the following steps:

- State the hypotheses you are testing.
- Compute the sign test statistic K by hand.
- Compute the z -statistic corresponding to the value of K that you obtained.
- Find the two-sided p -value using the normal approximation to the null distribution of K .
- Write a conclusion for your test.

$$H_0 : m = 0 \quad H_A : m \neq 0$$

```
K <- 8  
K
```

```
## [1] 8
```

```
Z <- (K-(18/2))/sqrt(18/4)  
Z
```

```
## [1] -0.4714045
```

```
2*(1-pnorm(abs(Z)))
```

```
## [1] 0.6373519
```

With a p-value of 0.6374, we fail to reject the null hypothesis that the population median is 0.

(b) (6 points) Use the Wilcoxon Signed-Rank test to test that the “center” of this distribution is 0 using the following steps:

- State the hypotheses you are testing.
- Compute the signed-rank test statistic S by hand.
- Compute the z -statistic corresponding to the value of S that you obtained.
- Find the two-sided p -value using the normal approximation to the null distribution of S .
- Verify that you get the same result by using the `wilcox.test()` function in R with `exact=FALSE` and `correct=FALSE`.
- Write a conclusion for your test.

$$H_0 : m = 0 \quad H_A : m \neq 0$$

```
rankdata1 <- rank(abs(data1))

S <- 5+1+16+10+3+6+9+2
S

## [1] 52

Z <- (S - 18*(18+1)/4)/sqrt(18*(18+1)*(2*18+1)/24)
Z

## [1] -1.458937

2*(1-pnorm(abs(Z)))

## [1] 0.1445825

wilcox.test(data1, alternative=c("two.sided"), mu=0, exact=FALSE, correct=FALSE)

##
##  Wilcoxon signed rank test
##
##  data:  data1
##  V = 52, p-value = 0.1446
##  alternative hypothesis: true location is not equal to 0
```

Based on our p -value of 0.1446, we fail to reject the null hypothesis that our data is centered at 0.

(c) (3 points) Discuss the differences between the results in parts (a) and (b). Specifically, address:

- Do both tests come to similar conclusions (i.e. both reject or both fail to reject)?
- What are the differences between the set of hypotheses, (or are they testing the same thing)?
- In general, describe why you might choose to use one test versus the other (You might consider how you would explain the results of the test to someone who is not well versed in statistics).

Both tests fail to reject the null hypotheses, but each null hypothesis is also different for each test. The sign test is testing the value of the population median and the signed-rank test is testing the ‘pseudomedian’, or the median of all values (observations) in our dataset.

If you wanted to gauge specifically for the population median, I would use the sign test. This may be more helpful with skewed data or outliers. If you wanted to look more broadly at the ‘center’ of your data, I would consider the signed-rank test as a better analysis.

Question 2 (16 points): Two-Sample Tests

Consider the guinea pig lifetime data we looked at in Homework 3 (`ex0211` in the `Sleuth3` package). Previously, we considered an equal variance two-sample t -test. Now we will consider some alternatives.

(a) (4 points) Perform a Wilcoxon rank-sum test to test whether the population distributions of lifetime in the Control and Bacilli groups are the same using the `wilcox.test()` function. Consider a one-sided lesser (Bacilli minus Control) alternative hypothesis. You do not need to use the continuity correction. What do you conclude at significance level $\alpha = 0.01$?

$$H_0 : m_B = m_C \quad H_A : m_B < m_C \iff m_B - m_C < 0$$

```
pigcontrol <- ex0211$Lifetime[1:64]
pigbas <- ex0211$Lifetime[65:122]

wilcox.test(pigbas, pigcontrol, alternative=c("less"), mu=0, exact=FALSE, correct=FALSE, conf.level=0.99)

## 
##  Wilcoxon rank sum test
##
## data:  pigbas and pigcontrol
## W = 1478.5, p-value = 0.02647
## alternative hypothesis: true location shift is less than 0
```

Based on our p-value of 0.02647, we fail to reject the null hypothesis at the $\alpha = .01$ significance level. That is to say, we cannot dismiss the possibility that the population distributions of measured lifetimes is the same (with regards to the Control group of pigs and the Bacilli group of pigs).

(b) (4 points) Perform a Welch's two-sample t -test to test the null hypothesis that the population means of lifetime in the Control and Bacilli groups are the same using the `t.test()` function. Consider a one-sided lesser (Bacilli minus Control) alternative hypothesis. What do you conclude at significance level $\alpha = 0.01$?

$$H_0 : \mu_B - \mu_C = 0 \quad H_A : \mu_B - \mu_C < 0$$

```
t.test(pigbas, pigcontrol, alternative = "less")
```

```
##  
##  Welch Two Sample t-test  
##  
##  data:  pigbas and pigcontrol  
##  t = -3.2296, df = 97.807, p-value = 0.0008443  
##  alternative hypothesis: true difference in means is less than 0  
##  95 percent confidence interval:  
##        -Inf -49.89389  
##  sample estimates:  
##  mean of x mean of y  
##  242.5345 345.2344
```

With a p-value of 0.0008443, at the significance level $\alpha = 0.01$ we reject the null hypothesis. In other words, we have the statistical significance to say the population mean of lifetime for the Bacilli group is likely less than the population mean of lifetime for the Control group.

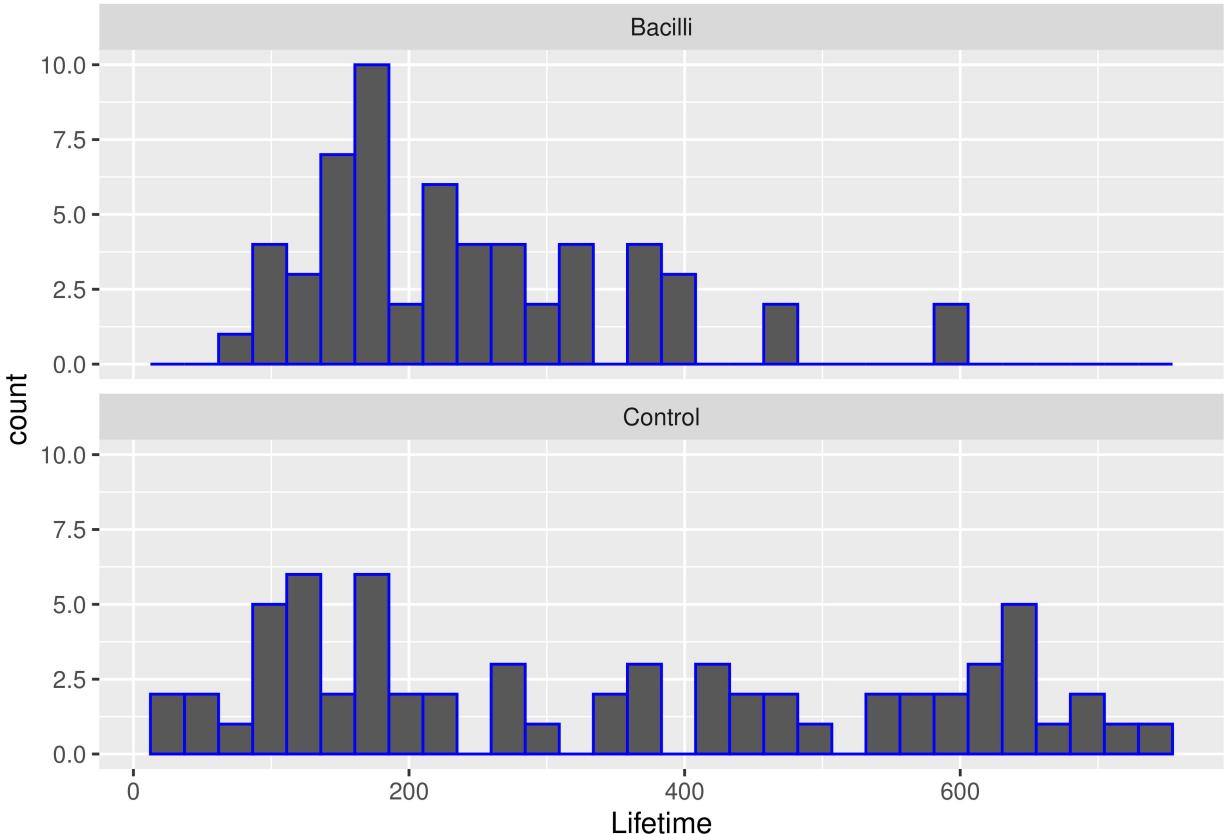
(c) (3 points) Discuss the differences between the results in parts (a) and (b). Specifically, address:

- Do both tests come to similar conclusions (i.e. both reject or both fail to reject)?
- What are the differences between the set of hypotheses, (or are they testing the same thing)?
- For this particular experiment, describe why you might choose to use one test versus the other.

The tests come to different conclusions entirely. The rank-sum test fails to reject the null while the two-sample t-test rejects the null at the same significance level with a very small p-value. Philosophically, the hypotheses are the same, it just depends on how you phrase them, either as a difference of the two not equaling zero, or one lesser than the other in the case of the alternative hypothesis. For this particular experiment, we know that the rank-sum test is more resistant to extreme outliers... but it's difficult to say whether or not we have any without performing more tests. Looking at the raw data, I would speculate that the lifetimes of the pigs form a "nice enough" distribution that we have no EXTREME outliers. Therefore, I would recommend using the normal two-sample t-test.

(d) (5 points) Previously, we analyzed the `ex0211` data using the equal variance *t*-test. Create a faceted histogram of the `Lifetime` variable where the facets are based on the `Group` variable (Here, you should consider arranging the facets so that one histogram is on top of the other to help you better compare the difference in the amount of spread. To do this in `ggplot2`, include the `ncol = 1` argument within the `facet_wrap()` function). Based on the sample distributions of each group, argue why we should use Welch's *t*-test instead of the equal variance *t*-test. Then, compare the R output for Welch's *t*-test and the equal variance *t*-test and comment about how the two tests differ in terms of their computed test statistics, degrees of freedom, and *p*-values.

```
ggplot(data = ex0211, aes(x=Lifetime))+geom_histogram(color="blue")+facet_wrap(~Group, ncol=1)
```



```
# There should be an example of how to do this in the lecture 11 slides.
```

Looking at the data this way persuades me that we should definitely use the Welch's t-test because the variances are certainly not equal. We have a slight right skew in the bacilli group and a more symmetric and “wider/thinner” spread in the control group.

```
t.test(pigbas, pigcontrol, alternative="less", var.equal=TRUE)
```

```
##  
## Two Sample t-test  
##  
## data: pigbas and pigcontrol  
## t = -3.1411, df = 120, p-value = 0.00106  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
##       -Inf -48.50152  
## sample estimates:  
## mean of x mean of y  
## 242.5345 345.2344
```

```
t.test(pigbas, pigcontrol, alternative="less", var.equal=FALSE)
```

```
##  
## Welch Two Sample t-test
```

```
##  
## data: pigbas and pigcontrol  
## t = -3.2296, df = 97.807, p-value = 0.0008443  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
##       -Inf -49.89389  
## sample estimates:  
## mean of x mean of y  
## 242.5345 345.2344
```

While both successfully reject the null hypothesis at the $\alpha = 0.01$ level, the Welch two-sample t-test has a smaller p-value and less degrees of freedom than the equal variance two-sample t-test. Because the p-value is smaller with Welch's t-test in context of a lesser-value H_A , it makes sense that the test statistic is smaller than the equal variance t-test as well. Or, vice versa.

Question 3 (5 points)

(a) (2 points) Generate two samples using the `rnorm()` function. Combine the two samples into one vector, and create another vector that indicates which group the observations belong to (See Lab 6 for help with creating the data for this question). The two samples should be drawn as follows:

- Sample A: $m = 10$ observations from a $\text{Normal}(\mu = 0, \sigma^2 = 1)$ distribution.
- Sample B: $n = 20$ observations from a $\text{Normal}(\mu = 0, \sigma^2 = 4)$ distribution. Note: You might also consider using the `set.seed()` function.

```
set.seed(441559)
sampA <- rnorm(10,0,1)
sampB <- rnorm(20,0,2)
sampComb <- c(sampA,sampB)
sampGrp <- as.factor(rep(c(1,2),c(10,20)))
df <- data.frame(sampComb,sampGrp)
```

(b) (3 points) Perform Levene's test in R using the `leveneTest()` function in the `car` library. Note: you will need to install/load the `car` package using `install.packages("car")` [Run the `install.packages()` function in the console and not the R Markdown] `library(car)` [Include this command in the R Markdown]. Report the resulting p -value and summarize your findings (state the hypothesis tested, the results of your analysis, and your conclusions).

```
library(car)
leveneTest(sampComb,group=sampGrp)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value Pr(>F)
## group    1  1.4362 0.2408
##          28
```

The hypothesis for Levene's test compares the 'spreads' of the two populations based on the samples we created, i.e., $H_0 : \sigma_A^2 = \sigma_B^2$ and $H_A : \sigma_A^2 \neq \sigma_B^2$. Based on our F-value of 1.4362 and p-value of 0.2408, we fail to reject the null hypothesis that the spreads are different at the $\alpha = 0.05$ significance level: We cannot discount the fact that the variability of the population distributions may be equal.