STAT 511A Homework 5

Kathleen Wendt

10/11/2019

Question 1

Rat data

Refer to Problem 6.42 which deals with lung capacity of rats exposed to ozone. Note: For consistency, please calculate the differences as After – Before where needed.

```r
rat_data <- readxl::read_xlsx("ex6-42.xlsx")
tibble::glimpse(rat_data)
```

```
## Observations: 12
## Variables: 3
## $ Rat    <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
## $ Before <dbl> 8.7, 7.9, 8.3, 8.4, 9.2, 9.1, 8.2, 8.1, 8.9, 8.2, 8.9, 7.5
## $ After  <dbl> 9.4, 9.8, 9.9, 10.3, 8.9, 8.8, 9.8, 8.2, 9.4, 9.9, 12.2...
```

Part 1A

Calculate the mean and standard deviation for Before and After (separately).

Before - Mean and SD

```r
mean(rat_data$Before)
```

```
## [1] 8.45
```

```r
sd(rat_data$Before)
```

```
## [1] 0.5161043
```

After - Mean and SD

```r
mean(rat_data$After)
```

```
## [1] 9.658333
```

```r
sd(rat_data$After)
```

```
## [1] 0.9876127
```

Part 1B

Add Differences Column (After - Before)

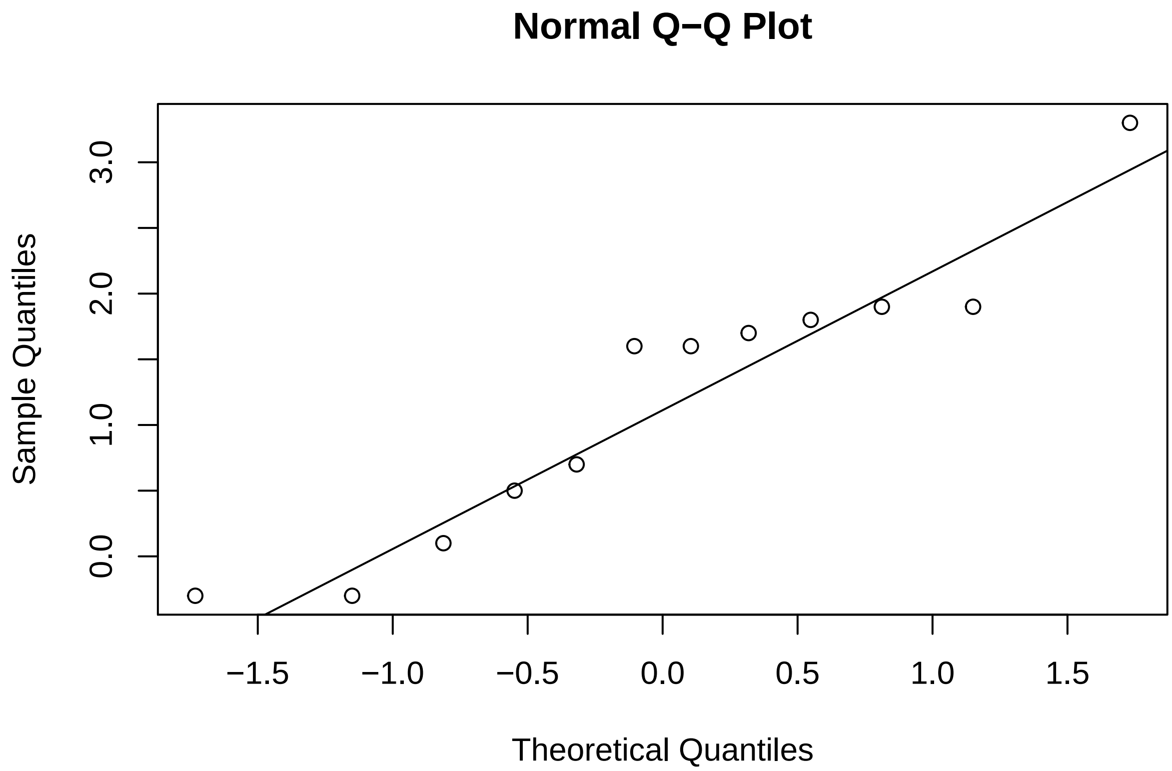
```r
library(magrittr)
rat_data <- rat_data %>%
  dplyr::mutate(diff = After - Before)
tibble::glimpse(rat_data)
```

```
## Observations: 12
## Variables: 4
## $ Rat    <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
## $ Before <dbl> 8.7, 7.9, 8.3, 8.4, 9.2, 9.1, 8.2, 8.1, 8.9, 8.2, 8.9, 7.5
## $ After  <dbl> 9.4, 9.8, 9.9, 10.3, 8.9, 8.8, 9.8, 8.2, 9.4, 9.9, 12.2...
## $ diff   <dbl> 0.7, 1.9, 1.6, 1.9, -0.3, -0.3, 1.6, 0.1, 0.5, 1.7, 3.3...
```

Q-Q Plot of Differences

Are the differences (After – Before for each rat) normally distributed? Support your answer by including a qqplot of differences in your assignment.

```r
qqnorm(rat_data$diff)
qqline(rat_data$diff)
```



The differences for each rat are approximately normally distributed; the quantiles are close (enough) to the line in the Q-Q plot, considering the small sample size.

Part 1C

Is there sufficient evidence to support the research hypothesis that there is a difference in average lung capacity after ozone exposure? Use the paired t-test with alpha=0.05. Give the hypotheses, test statistic, p-value and conclusion. (4 pts)

Hypotheses

Null (H0): $\mu 1$ - $\mu 2$ = 0 (No difference in means for Before and After) Alternative (HA): $\mu 1$ - $\mu 2$!= 0 (Sig difference in means for Before and After)

Paired Two-Sample T-Test

```
rat_t <- broom::tidy(t.test(
  x = rat_data$After,
  y = rat_data$Before,
       paired = TRUE,
       alternative = "two.sided"))
rat_t
```

```
## # A tibble: 1 x 8
##   estimate statistic p.value parameter conf.low conf.high method
##      <dbl>     <dbl>   <dbl>     <dbl>    <dbl>     <dbl> <chr>
## 1     1.21      3.89 0.00254        11    0.524      1.89 Paire~
## # ... with 1 more variable: alternative <chr>
```

Test Statistic

The test statistic from the above paired two-sample t-test is 3.8850127.

P-Value

The p-value from the above paired two-sample t-test is 0.0025413.

Conclusion

We reject the null hypothesis that there is no difference in rats' average lung capacity before and after ozone exposure. There is sufficient evidence to suggest a difference in means, such that the rats' average lung capacity is higher after ozone exposure compared to baseline levels.

Part 1D

Rerun the test from the previous question using the Wilcoxon Paired (Signed Rank) test. Give your p-value and conclusion. Use the wilcoxsign_test() function from the coin package with distribution = "exact".

```
rat_nonpara <- coin::wilcoxsign_test(After ~ Before,
                                     data = rat_data,
                                     distribution = "exact")
rat_nonpara
```

```
##
##  Exact Wilcoxon-Pratt Signed-Rank Test
##
## data:  y by x (pos, neg)
##   stratified by block
## Z = 2.6692, p-value = 0.004883
## alternative hypothesis: true mu is not equal to 0
```

We used the Wilcoxon Paired Sign Rank test to nonparametrically investigate the difference of means in rats' lung capacity before and after ozone exposure. We reject the null hypothesis at the α level of 0.05, $p = 0.005$.

Question 2

Refer to problem 7.9 which deals with rebound coefficients of baseballs. The summary statistics are provided here for your convenience: n = 40, mean = 84.798, s = 2.684. The raw data is also available from the Ott & Longnecker companion site as "exp07-9.txt".

Baseball data

```r
baseball_data <- readxl::read_xlsx("exp07-9.xlsx")
tibble::glimpse(baseball_data)
```

```
## Observations: 40
## Variables: 1
## $ coefficient <dbl> 84.8, 88.1, 85.1, 88.0, 86.6, 85.3, 85.1, 91.4, 83...
```

Summary statistics

```r
mean(baseball_data$coefficient)
```

```
## [1] 84.7975
```

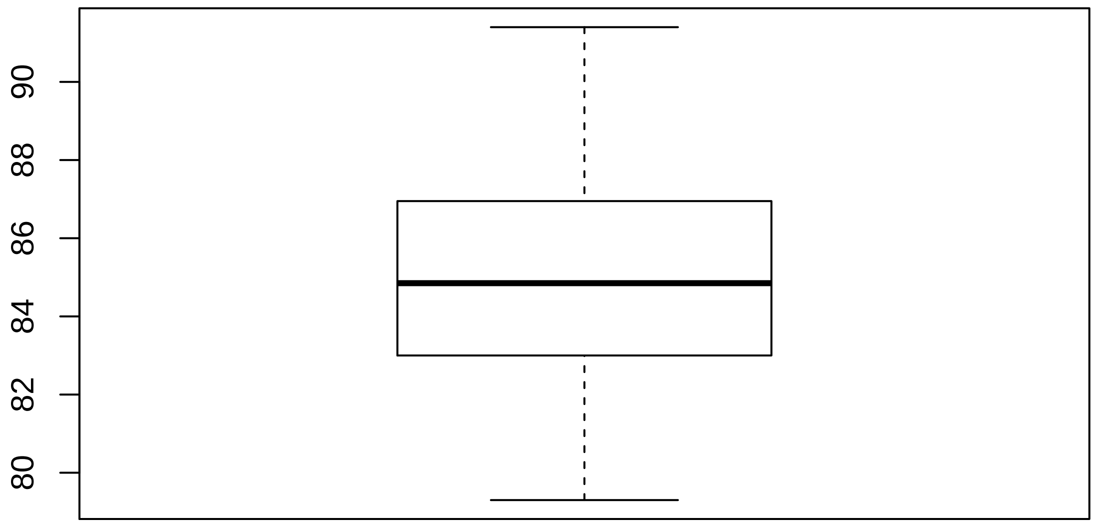
```r
sd(baseball_data$coefficient)
```

```
## [1] 2.683997
```

Part 2A

Construct a boxplot of the data and include it in your assignment.

```r
boxplot(baseball_data$coefficient)
```



Part 2B

Using $\alpha = 0.01$, test H0: mean greater than or equal to 85 vs HA: mean less than 85. Give the one-sided p-value and conclusion.

```
baseball_ttest <- (broom::tidy(t.test(baseball_data$coefficient,
                                      mu = 85,
                                      alternative = "less",
                                      conf.level = 0.99)))
baseball_ttest
```

```
## # A tibble: 1 x 8
##   estimate statistic p.value parameter conf.low conf.high method
##      <dbl>     <dbl>   <dbl>     <dbl>    <dbl>     <dbl> <chr>
## 1     84.8    -0.477   0.318        39     -Inf      85.8 One S~
## # ... with 1 more variable: alternative <chr>
```

We fail to reject the null hypothesis that μ is greater than or equal to 85. There is not sufficient evidence to suggest the μ of coefficients in baseball is less than 85, $p = 0.3179522 > \alpha = 0.01$.

Part 2C

Note that Table 7 (chi-square) does not have information for df = 39, so use the qchisq() R function to calculate table values needed for parts C and D.

Construct a 99% CI for baseball coefficient standard deviation by hand. Note: provide a standard "two-sided" CI here.

```
qchisq(0.995, df = 39)
```

```
## [1] 65.47557
```

```
qchisq(0.005, df = 39)
```

```
## [1] 19.99587
```

Construct 99% CI for SD "By Hand"

The 99% confidence interval for the standard deviation is 2.0714466, 3.7483769 (i.e., in-line R code to calculate CI using equation).

Part 2D

Using $\alpha = 0.01$, (by hand) test H0: sd less than or equal to 2 vs HA: sd greater than 2 (variance of 4). Give your test statistic, rejection rule and conclusion. (4 pts)

Hypotheses

H0: $\sigma <= 2$. HA: $\sigma > 2$.

Test Statistic

The chi-square test statistic for the right one-sided alternative is 70.237596.

Rejection Rule

Reject null if chi square test statistic is **greater than** chi square table value ($\alpha = 0.01$, df = 40 - 1 = 39).

```
qchisq(0.99, df = 39)
```

```
## [1] 62.42812
```

Conclusion

We reject the null hypothesis ($\sigma <= 2$) because the chi-square test statistic (70.237596) is greater than the right one-sided alternative table value (62.428121). The corresponding p-value can be determined by:

```
1-pchisq(70.24, df = 39)
```

```
## [1] 0.001581493
```