Stats511 Hw5

Amy Fox

10/7/2019

```
library(dplyr)
library(tidyr)
library(coin)
```

Question 1

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.

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

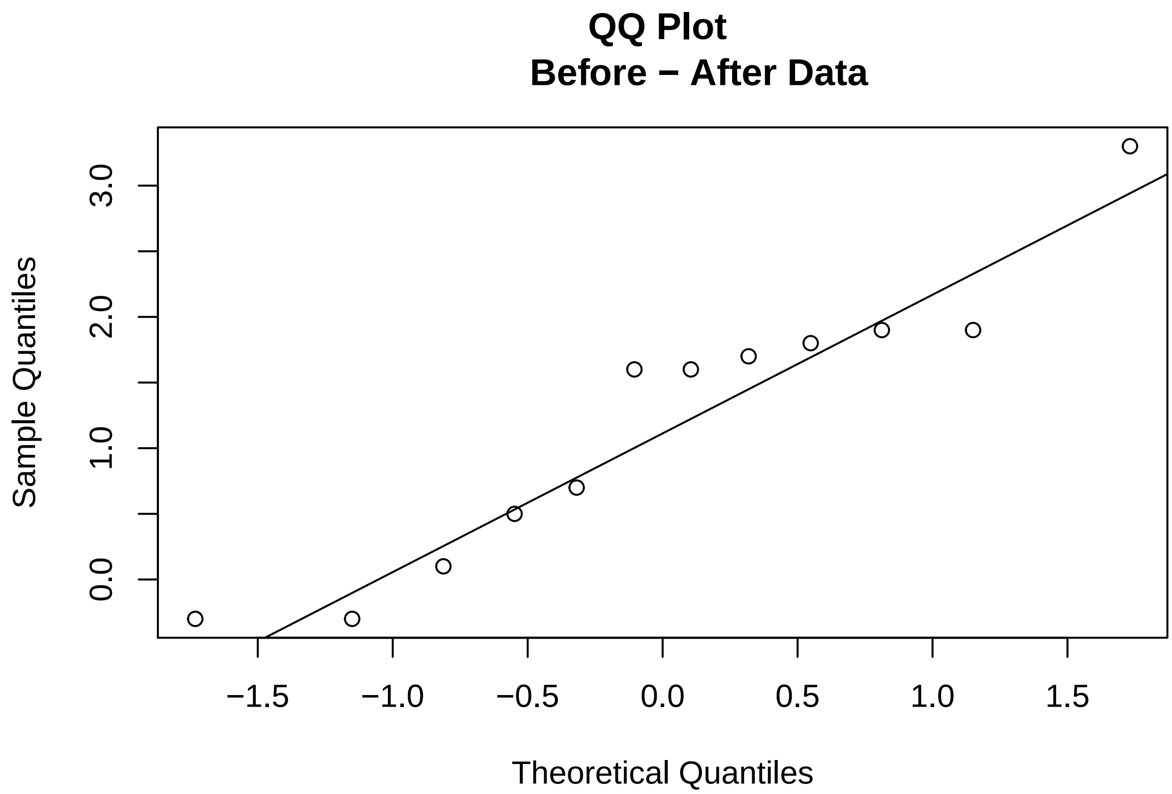
```
# read in data
lung_capacity_data <- read.csv("../Data/OTT_Final/ASCII-comma/CH06/ex6-42.txt") %>%
  rename(rat = `X.Rat.`,
         before = `X.Before.`,
         after = `X.After.`) %>%
  mutate(diff = after - before,
         sd_diff = sd(diff))

lung_capacity_data %>%
  gather(key = "time", value = "capacity", -rat) %>%
  group_by(time) %>%
  summarise(mean = mean(capacity),
            sd = sd(capacity))
```

```
## # A tibble: 4 x 3
##   time     mean    sd
##   <chr>   <dbl> <dbl>
## 1 after    9.66 0.988
## 2 before   8.45 0.516
## 3 diff     1.21 1.08
## 4 sd_diff  1.08 0
```

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

```
qqnorm(lung_capacity_data$after - lung_capacity_data$before, main = "QQ Plot
       Before - After Data")
qqline(lung_capacity_data$after - lung_capacity_data$before)
```




```r
shapiro.test(lung_capacity_data$diff)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  lung_capacity_data$diff
## W = 0.91788, p-value = 0.2689
```

It appears as if the data is mostly normally distributed based on the QQplots, however, there are a few points that could be outliers. (I also performed a Shapiro Wilks test to confirm that the data is normally distributed. Because the p-value = 0.2689 > alpha = 0.05 we assume that it is normally distributed.)

C. 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 α =0.05. Give the hypotheses, test statistic, p-value and conclusion. (4 pts)

Hypothesis

H0: μ (before) = μ (after)

HA: μ (before) $\neq$ μ (after)

Test Statistic

```r
# test statistic and p-value
t.test(lung_capacity_data$after, lung_capacity_data$before, paired = T)
```

```
##
##  Paired t-test
##
## data:  lung_capacity_data$after and lung_capacity_data$before
## t = 3.885, df = 11, p-value = 0.002541
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  0.5237735 1.8928932
## sample estimates:
## mean of the differences
##                1.208333
```

```
# another way to calculate test statistic
d_line <- mean(lung_capacity_data$before) - mean(lung_capacity_data$after)
s_d <- lung_capacity_data$sd_diff[1]
n_d <- length(lung_capacity_data$diff)

d_line/(s_d/sqrt(n_d))
```

```
## [1] -3.885013
```

Test statistic: 3.885

p-value: 0.0025

Conclusion: Reject H0 because p-value $= 0.0025 < \alpha = 0.05$. Therefore, the true difference between the means is not equal to 0.

D. 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".

```
wilcoxsign_test(lung_capacity_data$after ~ lung_capacity_data$before, paired = T, distribution = "exact
```

```
##
##  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
```

Because the p-value $= 0.004 < \alpha = 0.05$, we reject H0. Therefore, there is most likely a difference between the true mean lung capacity before and after ozone exposure.

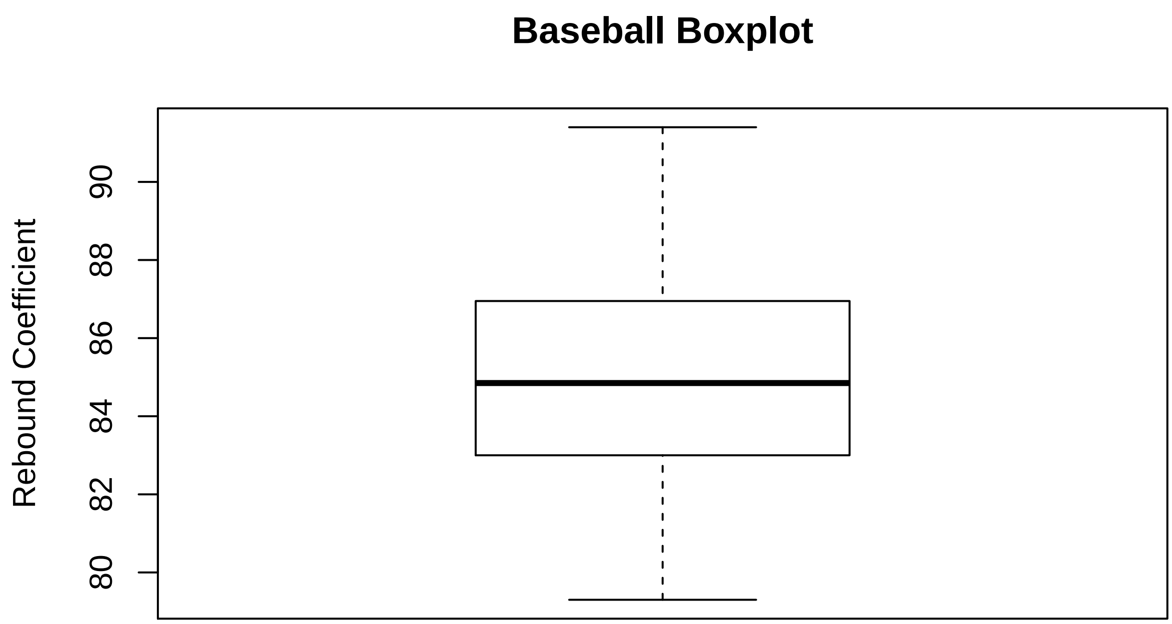
Question 2

Refer to problem 7.9 which deals with rebound coefficients of baseballs. The summary statistics are provided here for your convenience: n = 40, $\bar{y}$ = 84.798, s = 2.684. The raw data is also available from the Ott & Longnecker companion site as "exp07-9.txt". Note that Table 7 (chisquare) does not have information for df = 39, so use the qchisq() R function to calculate table values needed for parts C and D.

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

```
baseball_data <- read.csv("../Data/OTT_Final/ASCII-comma/CH07/exp07-9.txt") %>%
  rename(coefficient = `X.coefficient.`)

boxplot(baseball_data$coefficient, main = "Baseball Boxplot", ylab = "Rebound Coefficient")
```



B. Using $\alpha = 0.01$, test H0: $\mu \geq 85$ vs HA: $\mu < 85$. Give the one-sided p-value and conclusion.

```
t.test(baseball_data$coefficient, alternative = "less", mu = 85)
```

```
##
##  One Sample t-test
##
## data:  baseball_data$coefficient
## t = -0.47717, df = 39, p-value = 0.318
## alternative hypothesis: true mean is less than 85
## 95 percent confidence interval:
##      -Inf 85.51252
## sample estimates:
## mean of x
##   84.7975
```

Because the p-value $= 0.318 > \alpha = 0.01$, we fail to reject H0. Thus, the true mean is most likely greater than or equal to 85.

C. Construct a 99% CI for σ. Note: provide a standard "two-sided" CI here.

```
alpha = 0.01
n = 40
s = 2.684

chi_1 <- qchisq(alpha/2, df = 39)
chi_2 <- qchisq(1-(alpha/2), df = 39)

# one CI bound
sqrt(((n-1)*s^2)/chi_1)
```

```
## [1] 3.748389
```

```
# second CI bound
sqrt(((n-1)*s^2)/chi_2)
```

```
## [1] 2.071453
```

CI = (2.07, 3.74)

D. Using $\alpha = 0.01$, test H0: $\sigma \leq 2$ vs HA: $\sigma > 2$. Give your test statistic, rejection rule and conclusion. (4 pts) By hand

Test Statistic

```
alpha = 0.01
n = 40
s = 2.684
sigma_0 = 2

(n-1)*s^2/(sigma_0^2)
```

```
## [1] 70.2376
```

Test statistic = 70.2

Rejection rule

```
df = n-1

qchisq(1-alpha, df)
```

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

Rejection region = 62.4

Conclusion Because Chi^2 = 70.2 > 62.4 = Chi^2 alpha, n-1

Reject H0

Therefore, the true standard deviation is greater than 2.