Stat 3301: Homework 4

Due by date and time specified on Carmen

FirstName LastName (name.n)

Setup:

```
knitr::opts_chunk$set(echo = TRUE)
library(alr4)
library(tidyverse)
```

Instructions

- Replace "FirstName LastName (name.n)" above with your information.
- Provide your solutions below in the spaces marked "Solution:".
- Include any R code that you use to answer the questions; if a numeric answer is required, show how you calculated it in R. I have set the global option `echo = TRUE` to make sure the R code is displayed.
- Knit this document to HTML and upload both the HTML file and your completed Rmd file to Carmen
- Make sure your solutions are clean and easy-to-read by
  - formatting all plots to be appropriately sized, with appropriate axis labels.
  - only including R code that is necessary to answer the questions below.
  - only including R output that is necessary to answer the questions below (avoiding lengthy output).
  - providing short written answers explaining your work, and writing in complete sentences.
- Data files mentioned below are from the `alr4` package unless specified otherwise.

Concepts & Application In this assignment, you will gain experience with the following methods of statistical inference:

- confidence intervals for mean functions
- hypothesis tests about mean functions
- predictions and prediction intervals
- confidence intervals for functions of parameters

Questions This is a continuation of the question on the last homework that relates to the **NBA player** data we looked at in class (it is available on Carmen). Here we will focus on players whose position is **forward** (labeled `F` in the data set).

Use the statistical model you fit on the last homework assignment to answer the following questions.

1. Calculate and interpret a 95% confidence interval for the average **weight** of forwards who are 81.5" tall.
2. Remake the plot of weight versus height for forwards with the estimated regression line from the last homework assignment. Add lines to the plot representing point-wise 90% confidence intervals for the mean function. Also provide a formula explaining how these intervals are computed.
3. A reporter makes the claim that the average weight for NBA forwards who are 6' 7" tall is 210 pounds. Use an appropriate hypothesis test to assess whether the there is sufficient evidence to refute this claim.
4. An NBA team hopes to add a new 6' 7" tall forward to their roster. Predict the weight (reporting both point and interval predictions) for the weight of this new player.
5. In this part we are interested in the quantity $E(wt \mid ht = x + 2) - E(wt \mid ht = x)$ for some baseline height x. We will think of this quantity as a parameter and will call the parameter δ:

   $$\delta \equiv E(wt \mid ht = x + 2) - E(wt \mid ht = x).$$

   In words, this quantity is the difference in average weight for players that are two inches apart in height.

   a. Show that, in terms of the model parameters, $\delta = 2\beta_1$.
   b. Calculate an estimate $\widehat{\delta} = 2\widehat{\beta}_1$ of δ using the OLS estimate $\widehat{\beta}_1$.
   c. Find a formula (not a numeric value) for the variance of the estimator: $\text{Var}(\widehat{\delta} \mid X) = \text{Var}(2\widehat{\beta}_1 \mid X) = \ldots$
   d. Find a formula (not a numeric value) for the standard error of the estimator, $se(\widehat{\delta} \mid X)$, by taking the square root of your answer from part (c) and substituting the estimate $\widehat{\sigma}$ for the unknown parameter σ.
   e. Use the data to calculate the numeric value of $se(\widehat{\delta} \mid X)$.
   f. Use all of the above information to calculate a 95% confidence interval for δ and write a sentence interpreting this interval in terms of NBA forwards.

```
library(readr)
nba = read.csv('nbahtwt.csv')
```

Solutions Part 1:

```
forwards = nba %>% filter(pos=="F")
forwards.lm = lm(wt ~ ht, data = forwards)

predict(forwards.lm, newdata = data.frame(ht = 81.5), se.fit = TRUE, interval = "confidence")
```

```
## $fit
##        fit      lwr      upr
## 1 233.9597 231.4157 236.5036
##
## $se.fit
## [1] 1.290425
##
## $df
## [1] 209
##
## $residual.scale
## [1] 15.91395
```

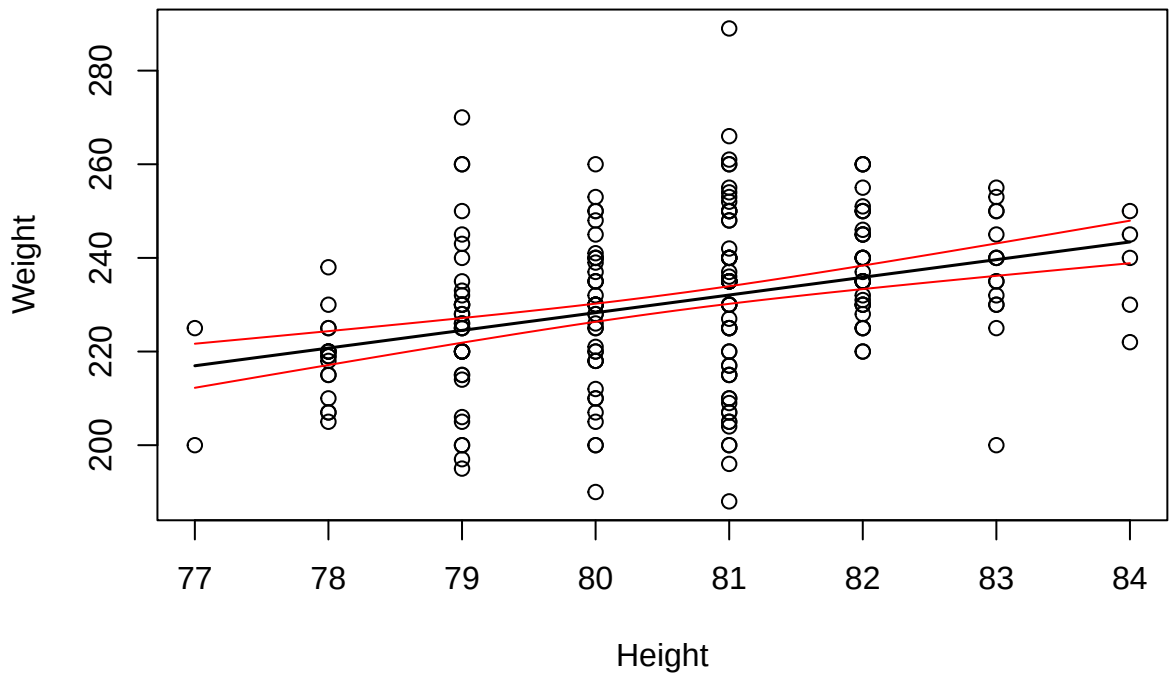
For forwards who are 81.5 inches tall, their average weight is between 231.4157 inches and 236.5036 inches with 95% confidence.

Part 2:

```
grid = seq(from = min(forwards$ht), to = max(forwards$ht), length.out = 100)

forwards.cis = predict(forwards.lm, newdata = data.frame(ht = grid), level = 0.9, interval = "confidenc

plot(forwards$ht, forwards$wt, xlab = "Height", ylab = "Weight")
lines(grid, forwards.cis[, "fit"], lwd = 1.5)
lines(grid, forwards.cis[, "lwr"], col = "red")
lines(grid, forwards.cis[, "upr"], col = "red")
```



The plot above displays the confidence interval for the mean function at any given point.

The plot uses a joint confidence region. The mean weight of each given height is calculated in the above plot. The function used is:

$$\hat{\mu}_{Y|X=x} \pm 2 * F(\alpha; df_1, df_2 - 2)^{\frac{1}{2}} sefit(\hat{y} \mid x)$$

The grid variable and forwards.cis variable has functions that capture the weight of a forward at any given height with 90% confidence which is then put into the plot.

Part 3: $H_0 : \mu_{X=79} = 210$ vs $H_1 : \mu_{X=79} \neq 210$

```
predict(forwards.lm, newdata = data.frame(ht = 79), se.fit = TRUE, interval = "confidence")
```

```
## $fit
##        fit      lwr      upr
## 1 224.5161 221.3735 227.6586
##
## $se.fit
## [1] 1.594086
##
## $df
## [1] 209
##
## $residual.scale
## [1] 15.91395
```

The mean weight for a player who is 79 inches in height is roughly 221.3735 inches to 227.6586 inches with 95% confidence. Since the reporter's hypothesis of an average of 210 pounds is not within this range, we reject the null hypothesis in favor of the alternative hypothesis. The mean weight for a player who is 79 inches in height is not 210 pounds.

Part 4: Point prediction:

```
forwards.p = predict(forwards.lm, newdata = data.frame(ht = 79))
forwards.pis = predict(forwards.lm, newdata = data.frame(ht = 79), se.fit = TRUE, interval = "predictio
forwards.p
```

```
##        1
## 224.5161
```

```
forwards.pis
```

```
## $fit
##        fit      lwr      upr
## 1 224.5161 192.9866 256.0455
##
## $se.fit
## [1] 1.594086
##
## $df
## [1] 209
##
## $residual.scale
## [1] 15.91395
```

The point prediction predicts the new 79 inch tall forward will be roughly 224.5161 pounds. The interval prediction says the new 79 inch tall forward will be between 192.9866 and 256.0455 pounds.

Part 5a: Done in other pdf.

Part 5b:

```
B_1_hat = sum((forwards$ht - mean(forwards$ht))*(forwards$wt - mean(forwards$wt)))/sum((forwards$ht-mean

delta_hat = 2 * B_1_hat
delta_hat
```

```
## [1] 7.554882
```

For an increase of 2 inches in height, there is roughly 7.554882 pounds gained on average.

Part 5c: $Var(\hat{\delta} \mid X) = Var(2\hat{\beta}_1 \mid X) = \frac{2^2\sigma^2}{S_{XX}}$

Part 5d: $se(\hat{\delta} \mid X) = \frac{2\hat{\sigma}}{\sqrt{S_{XX}}}$

Part 5e:

```
forward_data = nba %>% filter(pos=="F") %>% summarize(xbar = mean(ht), ybar = mean(wt), SXX = sum((ht -
xbar = forward_data$xbar
ybar = forward_data$ybar
SXY = forward_data$SXY
SXX = forward_data$SXX

B_1_hat = SXY/SXX
B_0_hat = ybar - (B_1_hat*xbar)
RSS = sum((forwards$wt - (B_0_hat + B_1_hat*forwards$ht))^2)
n = sum(nba$pos == "F")

sigma = sqrt((RSS)/(n-2))
se = (2*sigma)/(sqrt(SXX))

se
```

```
## [1] 1.471856
```

The variable se shows there is roughly 1.4718 pounds for the standard error for an increase of 2 inches.

Part 5f: We are looking for

$$\hat{\delta} \pm t(\frac{\alpha}{2}, n-2) * se(\hat{\sigma} \mid x)$$

```
n = sum(nba$pos == "F")
lwr = delta_hat - qt(p=0.025, df = n - 2, lower.tail = FALSE) * se
upr = delta_hat + qt(p=0.025, df = n - 2, lower.tail = FALSE) * se

lwr
```

```
## [1] 4.653295
```

```
upr
```

```
## [1] 10.45647
```

For every 2 inches increased in height of nba forwards, the average increase in weight of nba forwards is between 4.6533 and 10.4565 pounds with 95% confidence.