STAT 230 HW 1

Owen Forman

1. (Conceptual review) Chapter 2, exercises 3, 4, and 7

The answers to the following questions can be found on page 56. You should think about these questions before looking at the answers. You will receive full credit for answering the questions.

Chapter 2, Exercise 3 Answer: False

Chapter 2, Exercise 4 Answer: True

Chapter 2, Exercise 7 Answer: If we have a 95% confidence interval of (0.61, 1.19) this means we are 95% confident that the mean head circumference of babies whose mother did not smoke marijuana is between .61cm and 1.19 cm larger than babies whose mother did smoke marijuana. If we assume our null hypothesis is that smoking marijuana has no effect on mean baby head circumference, then this confidence interval would imply that the p-value is less than .05, since the value 0 (no difference in mean baby head circumference) does not fall within our 95% confidence interval.

2. (Grizzlies)

Load the following dataset:

```r
#load grizzly dataset
grizzly <- read.csv('https://www.math.carleton.edu/ckelling/data/Grizzly.csv', header=TRUE)
```

The data record the length and girth measurements (in inches) for a sample of grizzlies.

(a) Plot the bear's girth (response variable) against its length (predictor variable) and show the least squares line. Make sure the plot has appropriate axis labels. Comment on the apparent relationship (strength, shape, direction).

Answer: The relationship appears to have a relatively strong positive correlation. The relationship also appears to follow a linear trend.

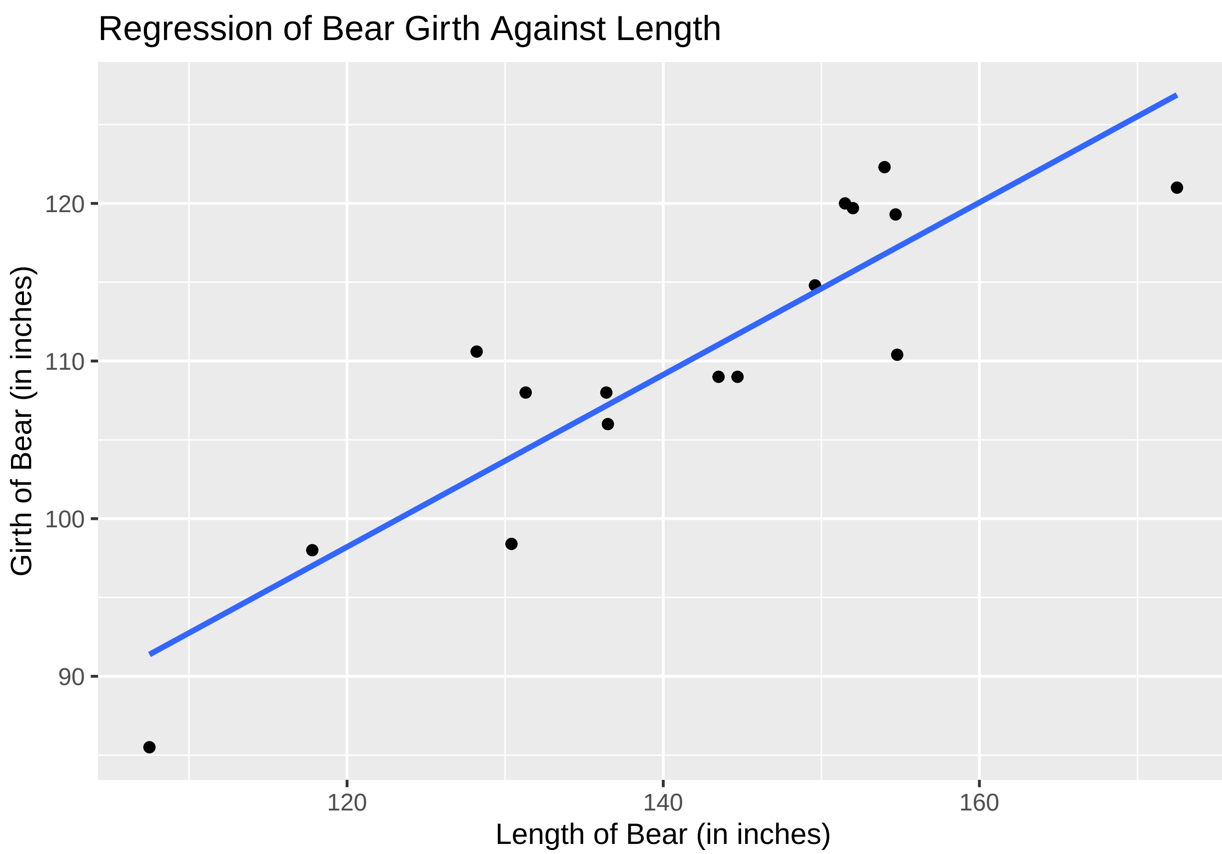
```r
#code with comments here

#load libraries
library(ggplot2)

#creates base graph (girth against length) with points shown
bear_girth_on_length <- ggplot(data=grizzly) +aes(x = Length, y = Girth) + geom_point() + labs(title =

#adds least squares line
bear_girth_on_length + geom_smooth(method = "lm", se = FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



(b) Find (and write out) the equation of the estimated linear model of the expected girth given length. (Here I am just asking for the equation of the fitted values, $\hat{Y}$; it is not necessary to state the assumptions about the errors here. Do not just print the R output; you need to write down the equation.)

Answer: $\hat{Girth} = 32.6678 + .5462(\text{Length})$

```r
#code with comments here

#creates fit model
fit <- lm(Girth ~ Length, data = grizzly)

#prints the coefficient of model
fit
```

```
##
## Call:
## lm(formula = Girth ~ Length, data = grizzly)
##
## Coefficients:
## (Intercept)       Length
##     32.6678       0.5462
```

(c) What is the interpretation of the slope coefficient?

Answer: "for each extra inch of length of a bear , the estimated average girth of the bear increases by .5462 inches"

(d) Give and interpret a 95% confidence interval for the slope coefficient. (You can use the `confint()` command.)

Answer: We are 95% confident that for every extra inch of Length of a Bear, the average girth of the bear increases between .3855978 and .7067615 inches.

```r
#code with comments here

#finds 95% confidence interval
confint(fit)
```

```
##                 2.5 %     97.5 %
## (Intercept) 9.7925051 55.5430779
## Length      0.3855978  0.7067615
```

(e) Suppose for a different species of bear, the scientists sample 25 bears and for their estimated linear model they find a slope coefficient estimate of $\hat{\beta} = 0.425$ and $\text{SE}(\hat{\beta}) = 0.068$. Give a 90% confidence interval for the true slope coefficient. (Note: this question does not rely on the dataset used for the questions above. Use the estimated slope and SE given here, and the `qt()` function to get the appropriate critical value.)

Answer: The 90% confidence interval for the true slope coefficient is (.308, .542)

Note: these values are rounded to the third decimal place.

```r
#code with comments here

#sets given variables (note df is n-2)
df <- 25 - 2
SE <- 0.068
beta1 <- 0.425

#runs qt function
crit_val <- qt(.95, df)


#lower bound calculation
lower_bound <-beta1 - crit_val * SE

#upper bound calc
upper_bound <- beta1 +crit_val * SE

#print lower and upper bound values for viewing
lower_bound
```

```
## [1] 0.3084567
```

```r
upper_bound
```

```
## [1] 0.5415433
```

Recommended conceptual exercises (do not turn in, answers in book/Moodle):

Chapter 7 exercises 3, 4, 5, 9, 10, 12