Jonothan Meyer
Linear Regression
Project 6
03/18/21

Exercise 1 (6 points): Using male subjects from the teengamb data, fit a model with gamble as the response and the other variables as predictors.

(1) Predict the mean amount gambled by all males with average (given these data) status, income and verbal score using a 95% confidence level.

```
teengamb_M <- teengamb %>% filter(sex == 0)
summary(teengamb_M) #Mean:  stat = 52, income = 4.9, verbal = 6.8, gamble = 29.775

gamb_status.lm <- lm(gamble~status+income+verbal, data=teengamb_M)
mean_predictors <- data.frame(status=52, income=4.9, verbal=6.8)
predict(gamb_status.lm, mean_predictors, interval="confidence", level=0.95)

# Confidence Interval for "Average Male Teen": (19.06, 39.7) based on the 'teengamb' data
```

(2) Predict the amount that a randomly sampled male with average (given these data) status, income and verbal score would gamble using a 95% confidence level.

```
predict(gamb_status.lm, mean_predictors, interval="predict", level=0.95)
# Prediction Interval for a randomly sampled "Average Male Teen": (-26.18, 84.94)
```

(3) Repeat the previous part for a male with maximal values (for this data) of status, income and verbal score. Which prediction interval is wider and why is this result expected?

```
min_predictors <- data.frame(status=18, income=.6, verbal=1)
max_predictors <- data.frame(status =75, income=15, verbal=10)

predict(gamb_status.lm, min_predictors, interval="prediction", level=0.95)
#Prediction Interval for teen male with minimum status, income, and verbal lifestyle: (-39.53, 91.15)

predict(gamb_status.lm, max_predictors, interval="prediction", level=0.95)
#Prediction Interval for teen male with maximum status, income, and verbal lifestyle: (10.87, 143.94)
```

(4) Predict the amount that 25 males with average (given these data) status, income and verbal score would gamble in a year using a 95% confidence level.

```
Xstar <- t(c(1,mean(teengamb_M$status), mean(teengamb_M$income), mean(teengamb_M$verbal)))

b<-coefficients(gamb_status.lm)
tval<-qt(1-(1-.95)/2,28-4)
X<-model.matrix(gamb_status.lm)
```

```
c(t(Xstar)%*%b-tval* (699.78*(1/25+t(Xstar)%*%inv(t(X)%*%X)%*%Xstar))^.5,
  t(Xstar)%*%b+tval* (699.78*(1/25+t(Xstar)%*%inv(t(X)%*%X)%*%Xstar))^.5)

#Prediction Interval for 25 average teen males: (14.75, 44.8)
```

Exercise 2 (6 points): Use the same predictors from the previous exercise (status, income and verbal score for males) to fit model of the square root of the gamble variable. Complete each part of Exercise 1 using the transformed gamble variable as your response. Take care to interpret the intervals in terms of the original units of the response variable.

```
gamble_root <- teengamb_M
gamble_root$gamble <- sqrt(teengamb_M$gamble)
#Rebuild Linear model using altered gamble response
gamble_root.lm <- lm(gamble~status+income+verbal, data=gamble_root)

predict(gamble_root.lm, mean_predictors, interval="confidence", level=0.95)
# Mean Confidence Interval: (3.43, 5.28)

predict(gamble_root.lm, mean_predictors, interval="prediction", level=0.95)
# Mean Confidence Prediction for single random average teen male: (-.612, 9.32)

predict(gamble_root.lm, min_predictors, interval="prediction", level=0.95)
# Minimum Prediction Interval: (-2.19, 9.53)

predict(gamble_root.lm, max_predictors, interval="prediction", level=0.95)
# Maximum Prediction Interval: (3.2, 15.1)

c(t(Xstar)%*%b-tval* (699.78*(1/25+t(Xstar)%*%inv(t(X)%*%X)%*%Xstar))^.5,
  t(Xstar)%*%b+tval* (699.78*(1/25+t(Xstar)%*%inv(t(X)%*%X)%*%Xstar))^.5)
# Prediction Interval for 25 average teen males :(-10.63, 19.41)
```

Exercise 3 (8 points): This is Exercise 7(a)(b)(c)(d) from Chapter 3 of the textbook Linear Models with R. The F-test required for part (c) is not covered in these notes, but is covered in the textbook readings.

(1) Fit a regression model with Distance as the response and the right and left leg strengths and flexibilities as predictors. Which predictors are significant at the 5% level?

```
punt <- punting
# Constructing our linear model using 'Distance' as response, and 'Right Leg Strength', 'Left Leg
# Strength', 'Right Leg Flexibility', and 'Left Leg Flexibility' as the predictors.
punt.lm <- lm(Hang~RStr+LStr+LFlex+RFlex, data=punt)
```

```
summary(punt.lm) #Relevant info from summary function:
```

Predictor	P-Value
RStr	0.52
LStr	0.37
RFlex	0.73
LFlex	0.42

The P-Value for all of the predictors is >.05, so using this model none of the predictors are identified to be significant at the 5% level.

(2) Use an F-test to determine whether collectively these four predictors have a relationship to the response.

The F-test results in a P-value of .00492, so at the 5% we can conclude that these four predictors collectively have a relationship to the response.

(3) Relative to the model in (1), test whether the right and left leg strengths have the same effect.

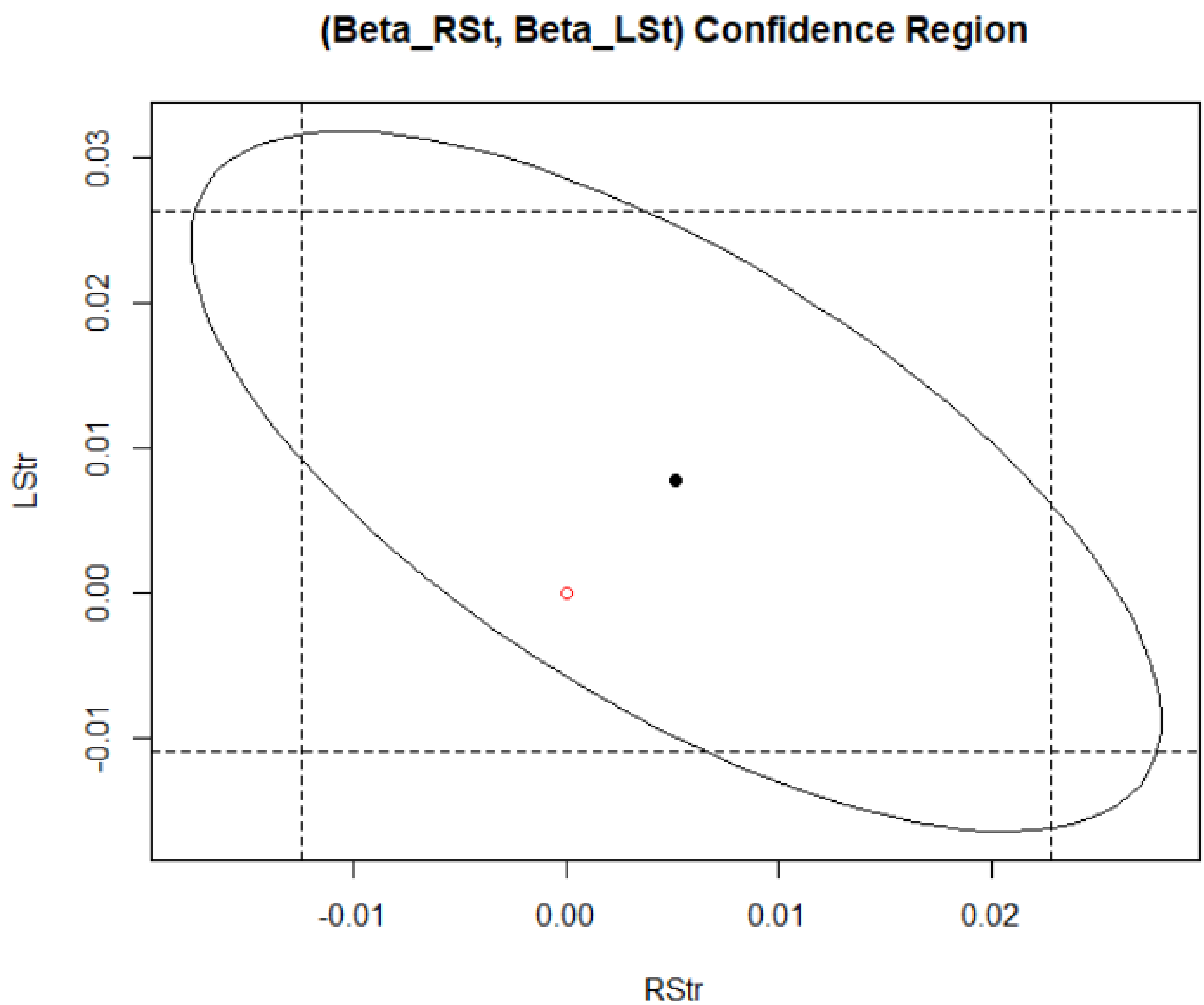
```
punt.lm <- lm(Distance~RStr+LStr, data=punt)
summary(punt.lm) #Relevant info from summary function:
```

Predictor	P-Value
RStr	0.17
LStr	0.69

At the 5% significance level neither of these two predictors are identified to be significant. However, like in (2) an F-test yeilds a P-value of .007, so we can conclude at the 5% level that these two predictors collectively have a relationship to the response.

(4) Construct a 95% confidence region for (Beta_RStr, Beta_LStr). Explain how the test in (3) relates to this region.

```
#Building our Confidence Region for Beta_RStr, Beta_LStr.
library(ellipse)
plot(ellipse(punt.lm,c(2,3)),type="l", main= "(Beta_RSt, Beta_LSt) Confidence Region")
points(coef(punt.lm)[2], coef(punt.lm)[3], pch=19)
abline(v=confint(punt.lm)[2,],lty=2)
abline(h=confint(punt.lm)[3,],lty=2)
points(0,0, col="Red")
```



This relates to part 3 by visually illustrating that when paired together for a linear model RStr and LStr are reject at 95% confidence. This is shown by the origin Red[(0,0)] inside the bounds of the oval.