STAT120 HW 9

Victor Huang

5/30/2021

Section 9.1

9.4

β_0 intercept = 7.2777; β_0 slope = -0.3560; $y = 7.2777 - 0.3560x$

9.7

Sample Slope = -0.3560; $H_0 : \beta = 0; H_a : \beta \neq 0$; p-value = 0.087. Since the p-value is greater than the 5% significance value, we fail to reject the null hypothesis.

9.17

a.The scatter plot does not reflect anything that should be of concern.

b. $y = 2.03 + 0.00189(650) = 3.26$
c. The slope is 0.00189, this slope reflects the increase in GPA for every SAT point.
d. The test statistic if 6.99, the p-value is approximately 0. As such, we can reject the null hypothesis.
e. The r-squared value is 0.125. This means that 12.5% of the values of students GPAs in this sample is explained by their corresponding verbal SAT score.

9.26

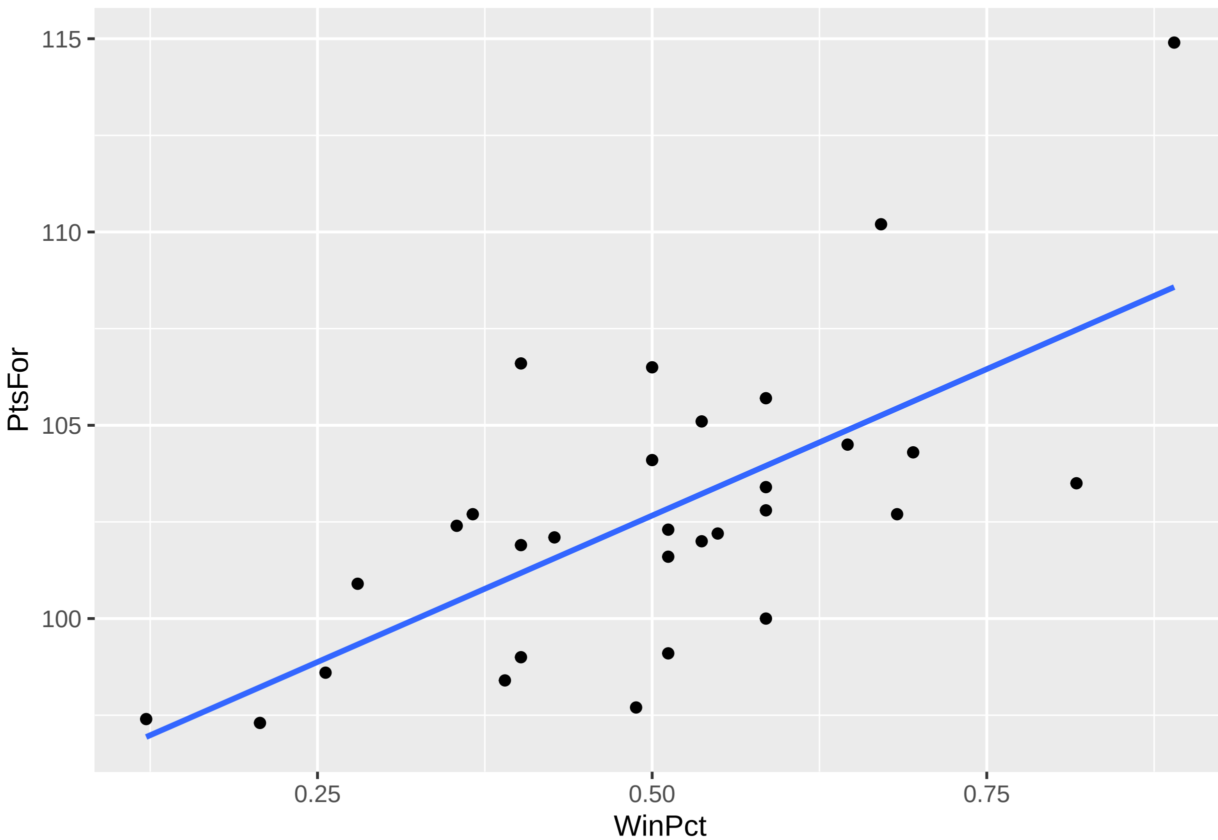
```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

##             Df Sum Sq Mean Sq F value   Pr(>F)
## Beds         1  21.05  21.046   40.77 6.53e-07 ***
## Residuals   28  14.45   0.516
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a. $H_0 : \beta = 0; H_a : \beta \neq 0$
b. Since we get a r-squared value of 0.593 (59.3%).
c. We get a f statistic of 40.77, the corresponding p-value is 6.53e-07 (approximately zero)
d. Since we get such a small p-value, we are able to reject the null hypothesis and find significant evidence that suggests there is correlation between the number of beds and baths in a house in CA. Given the

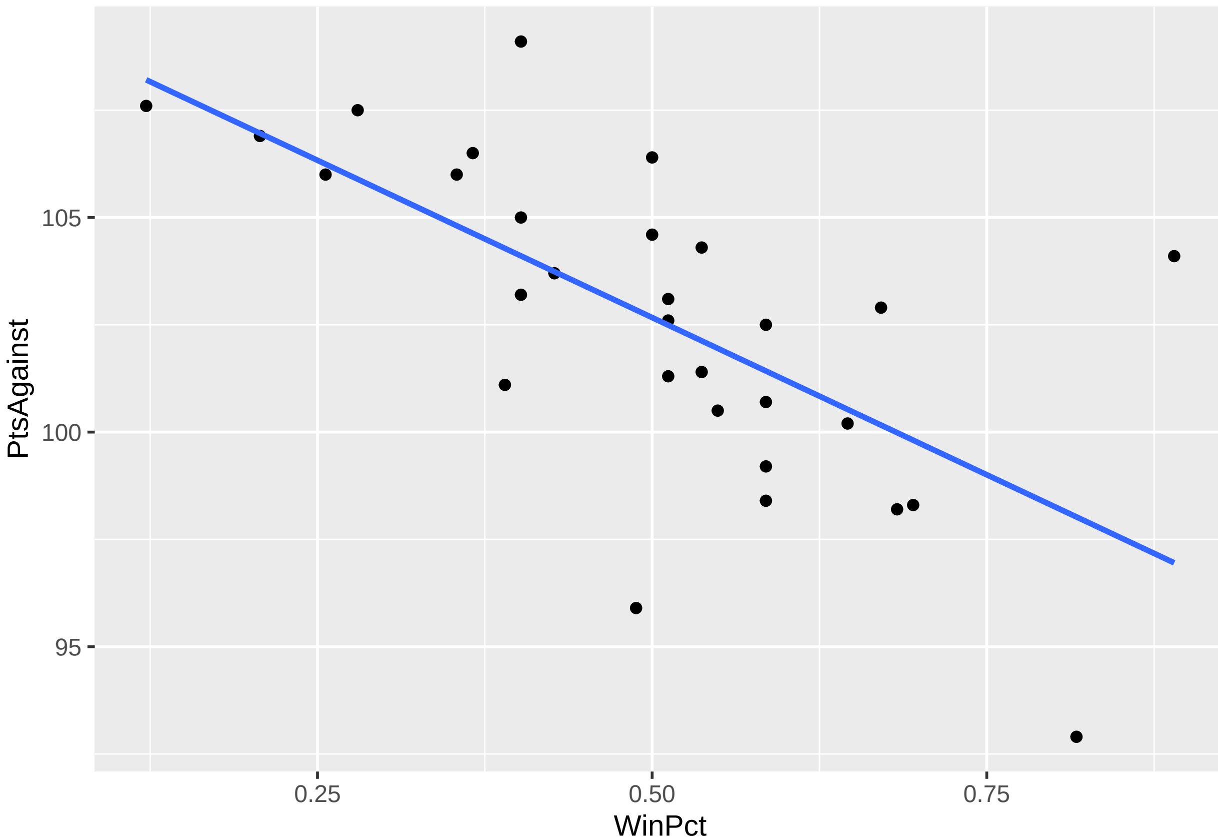
r-squared value of 59.3%, this means that 59.3% of the number of bathrooms of houses in CA in this sample is explained by the number of their bedrooms

9.28

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




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




```
##
## Call:
## lm(formula = WinPct ~ PtsFor, data = NBAStandings2016)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.21858 -0.10572  0.01607  0.07650  0.29177
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.607670   0.638332  -4.085 0.000334 ***
## PtsFor       0.030270   0.006214   4.871 3.94e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1268 on 28 degrees of freedom
## Multiple R-squared:  0.4587, Adjusted R-squared:  0.4394
## F-statistic: 23.73 on 1 and 28 DF,  p-value: 3.941e-05
```

```
##
## Call:
## lm(formula = WinPct ~ PtsAgainst, data = NBAStandings2016)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.22935 -0.04506 -0.00873  0.06213  0.43319
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.593558   0.657878   5.462 7.85e-06 ***
## PtsAgainst  -0.030132   0.006404  -4.705 6.20e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1288 on 28 degrees of freedom
## Multiple R-squared:  0.4416, Adjusted R-squared:  0.4216
## F-statistic: 22.14 on 1 and 28 DF,  p-value: 6.203e-05
```

a. As one data set shows linear increase while the other shows linear decrease, both graphs seem to fit conditions for a linear model

b. Equation: $WinPct = 3.594 - 0.030PtsFor$, since the p-value (3.941e-05) is so small, we can say that PtsFor is an effective predictor

c. Equation: $WinPct = -2.607 + 0.030PtsAgainst$, since the p-value (6.203e-05) is so small, we can say that PtsAgainst is an effective predictor

d. The r-squared value for PtsFor is 0.4587 while the r-squared value for PtsAgainst is 0.4416. The r-squared value for PtsFor against is larger than that of PtsAgainst.

e. PtsFor: $WinPct = 3.594 - 0.030(114.9) = 0.147$; PtsAgainst: $WinPct = -2.607 + 0.030(104.1) = 0.516$

f. Overall, the PtsAgainst seems to be more effective at predicting win percentage. This can be seen from the above example where PtsAginst predicted a 0.516 win percentage, much closer to the actual win percentage as compared to the 0.147 predicted by PtsFor. While the r-squared value for PtsFor is a bit higher than PtsAgainst, it is only approximately 1% and can be seen as not that significant when compared to the huge prediction difference as shown in example e.

Section 9.2

9.43

The F-stat is 7.44, the p-value is 0.011. Given this information, we can say that the linear model is appropriate for this. Since the p-value is so small, we can reject the null hypothesis.

9.46

a. Correlation: -0.366, the p-value is 0.015

b. The slope is -3.34, the t-value is -2.55, the p-value is 0.015

c. The f-stat is 6.50, the p-value is 0.015

d. They are the same

e. Since our p-value of 0.015 is less than the 5% significance level, we reject the null hypothesis and conclude that there is significant correlation between football and cognition.

9.54

```
##
## Call:
## lm(formula = LifeExpectancy ~ Health, data = SampCountries)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.034  -5.223   2.650   7.933  11.874
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  61.3202     4.7024  13.040   <2e-16 ***
## Health        0.7286     0.3637   2.003   0.0508 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.15 on 48 degrees of freedom
## Multiple R-squared:  0.07715,    Adjusted R-squared:  0.05792
## F-statistic: 4.013 on 1 and 48 DF,  p-value: 0.05083
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Health       1    414   413.7   4.013 0.0508 .
## Residuals   48   4948   103.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a. We get a correlation of 0.7286, the associated p-value is 0.05083

b. The t-stat is 2.003, the associated p-value is 0.05083.

c. We get an f-stat of 4.013 and an associated p-value of 0.0508

d. Since we have a p-value of 0.0508, it is somewhat higher than a 5% significance level. As such, we can say that this model is not that effective.

Section 9.3

9.58

A, B; 12

9.63

Confidence Interval: (-0.013, 4.783); Prediction Interval: (-2.797, 7.568)

9.68

```
##             Df  Sum Sq Mean Sq F value   Pr(>F)
## Size         1 9457748 9457748   56.88 3.25e-08 ***
## Residuals   28 4655513  166268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##        fit      lwr      upr
## 1 571.9103 445.2591 698.5614
```

```
##        fit       lwr      upr
## 1 571.9103 -133.2107 1277.031
```

a. Since we have such a small p-value (3.24e-08), we can say that size (square footage) is an effective predictor.

b. $Price = -369.63 + 470.77(2000) = 941170.37$

c. We get a 90% confidence interval of (445.2591, 698.5614). We are 90% confident that the house prices with sq footage of 2000 sqft in NY falls between the interval of (419.40, 729.42).

d. We get a 90% prediction interval of (-133.2107, 1277.031). 90% of all homes in NY with 2000 sq footage have a housing price between the interval of (-133.2107, 1277.031).

9.70

a. $GPA = 2.03 + 0.00189(500) = 2.975$; $GPA = 2.03 + 0.00189(700) = 3.353$

```
##        fit     lwr      upr
## 1 2.980022 2.91623 3.043814

##        fit     lwr      upr
## 1 2.980022 2.24318 3.716864

##       fit      lwr      upr
## 1 3.35861 3.289678 3.427542

##       fit      lwr      upr
## 1 3.35861 2.621305 4.095915
```

b.i. We get a 95% confidence interval of (2.91623, 3.043814). We are 95% confident that the students GPAs with a 500 verbal SAT score to fall between the interval of (2.91623, 3.043814).

b.ii. We get a 95% prediction interval of (2.24318, 3.716864). 95% of all students who have a 500 verbal SAT have a GPA between the interval of (2.24318, 3.716864).

b.iii. We get a 95% confidence interval of (3.289678, 3.427542). We are 95% confident that the students GPAs with a 700 verbal SAT score to fall between the interval of (3.289678, 3.427542).

b.iv. We get a 95% prediction interval of (2.621305, 4.095915). 95% of all students who have a 700 verbal SAT have a GPA between the interval of (2.621305, 4.095915).