week_3_exercises

Lindsey Greenhill

2/10/2021

Question 1

Part a

```
## lm(formula = y ~ x1 + x2, data = df_train)
##             coef.est coef.se
## (Intercept) 1.32     0.39
## x1          0.51     0.05
## x2          0.81     0.02
## ---
## n = 40, k = 3
## residual sd = 0.90, R-Squared = 0.97
```

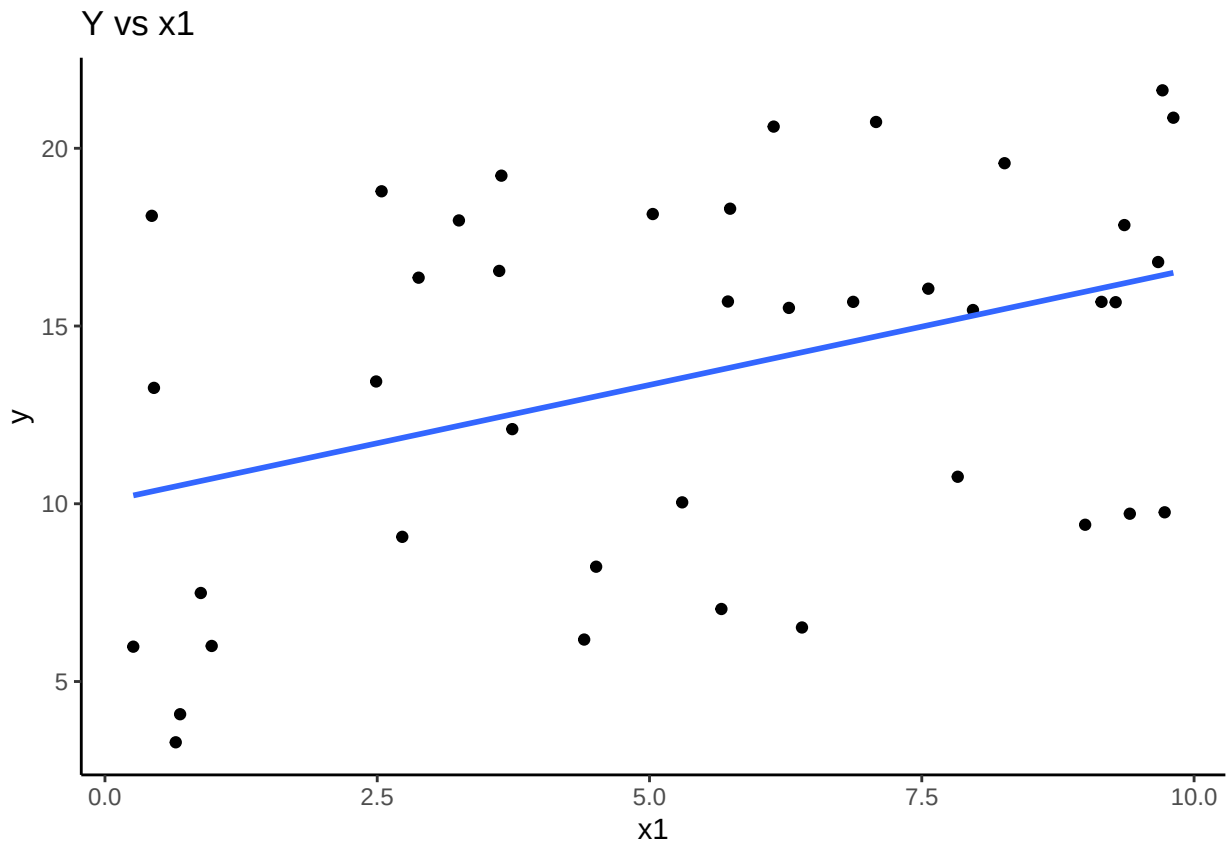
The intercept is equal to 1.31, meaning that the average value of y when x1 and x2 are both 0 is 1.13. The coefficient for x1 is .514, meaning that for every 1 unit increase in x1, y increases by .514 on average, holding x2 constant. The coefficient for x2 is .806, meaning that for every 1 unit increase in x2, y increases by .806 on average, holding x1 constant.

The R squared value is .97, meaning that about 97% of the variance is explained by the model

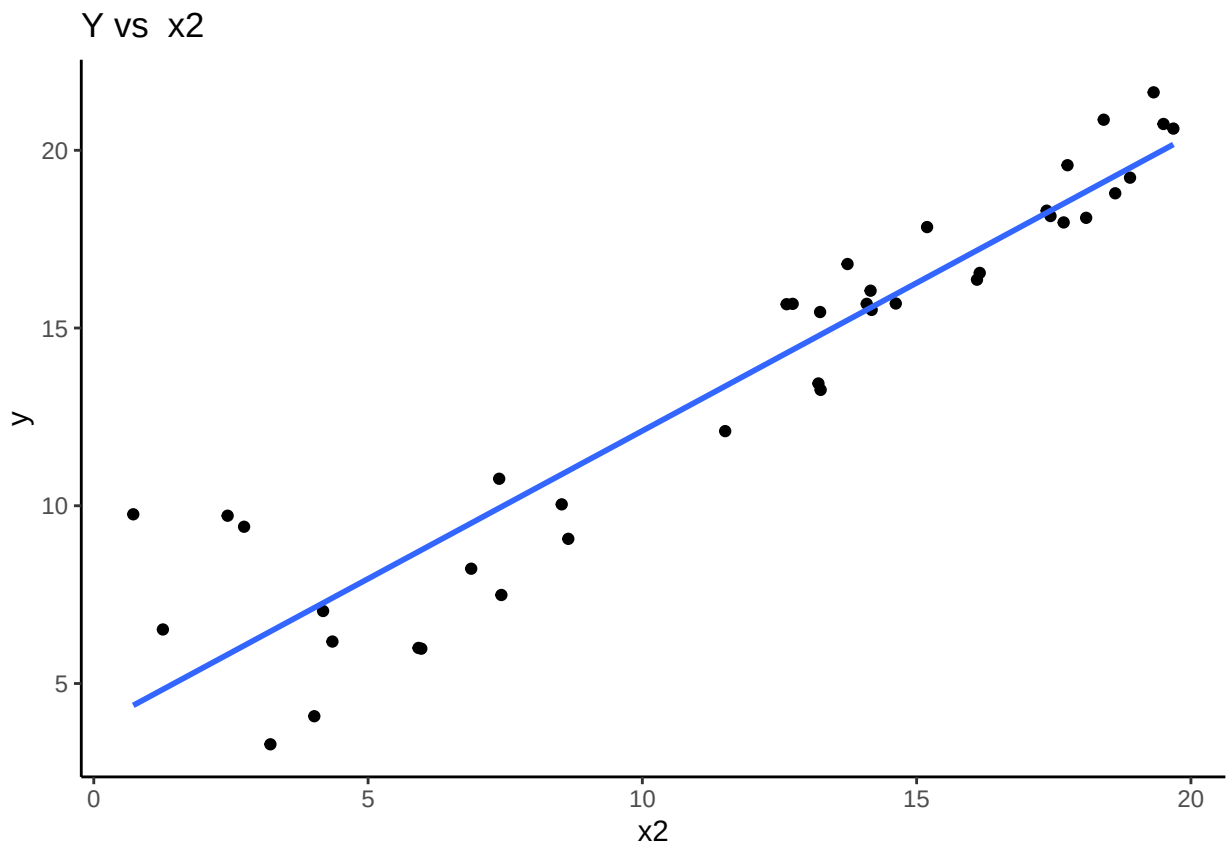
Part b

The graphs below visualize the relationship between y and x1 and x2. The blue line is the least squared regression line.

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

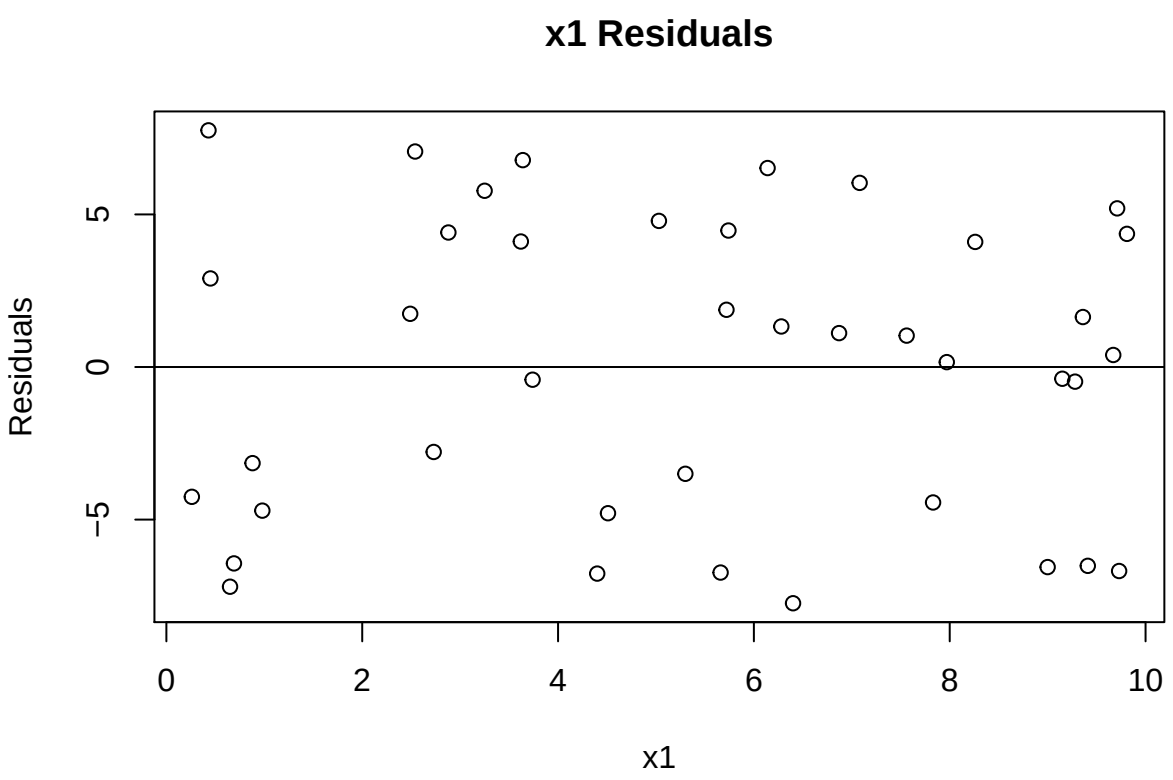


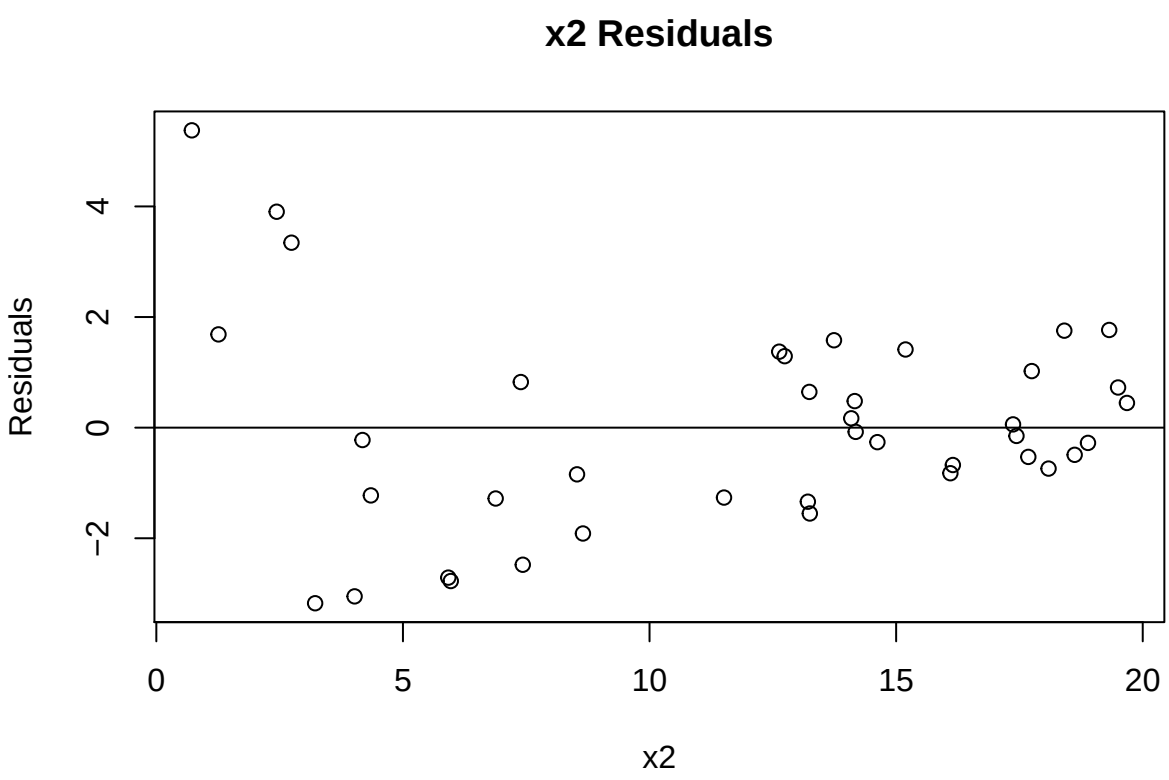

```
## `geom_smooth()` using formula 'y ~ x'
```



Part c

The plots below show the residuals of the two above models plotted against the observed values from the data frame. For the x1 plot, the residuals seem to be relatively evenly spread throughout the x1 values with no clear patterns emerging. For the x2 plot, the residuals seems to be larger for the x2 values closer to 0. However, the residuals for x2 seem to be smaller on average than the residuals for x1.


x1 Residuals
Residuals
5
0
−5
0
2
4
6
8
10
x1



Part d

```
##          fit       lwr       upr
## 41 14.812484 12.916966 16.708002
## 42 19.142865 17.241520 21.044211
## 43  5.916816  3.958626  7.875005
## 44 10.530475  8.636141 12.424809
## 45 19.012485 17.118597 20.906373
## 46 13.398863 11.551815 15.245911
## 47  4.829144  2.918323  6.739965
## 48  9.145767  7.228364 11.063170
## 49  5.892489  3.979060  7.805918
## 50 12.338639 10.426349 14.250929
## 51 18.908561 17.021818 20.795303
## 52 16.064649 14.212209 17.917088
## 53  8.963122  7.084081 10.842163
## 54 14.972786 13.094194 16.851379
## 55  5.859744  3.959679  7.759808
## 56  7.374900  5.480921  9.268879
## 57  4.535267  2.616996  6.453539
## 58 15.133280 13.282467 16.984094
## 59  9.100899  7.223395 10.978403
## 60 16.084900 14.196990 17.972810
```

Question 3

```
##
## Call:
## lm(formula = var1 ~ var2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2849 -0.7127 -0.0411  0.7346  3.3710
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.01568    0.03314   0.473    0.636
## var2        -0.02686    0.03268  -0.822    0.411
##
## Residual standard error: 1.047 on 998 degrees of freedom
## Multiple R-squared:  0.0006763,  Adjusted R-squared:  -0.000325
## F-statistic: 0.6754 on 1 and 998 DF,  p-value: 0.4114
```

Part a

The slope is not statistically significant.

Part b

Out of the 100 simulations, 2 estimates are statistically significant.

Out of the 100 simulations, only