Day6

Olivia Wu

2024-02-27

Problem 2.44

a) We are 95% confident that the true mean of predicted prices of a 450-page textbook lies in ($51.73, $74.02).

```
data <- data.frame(Pages=450)
predict.lm(model, data, interval="confidence")
```

```
##        fit      lwr      upr
## 1 62.87549 51.73074 74.02024
```

b) We are 95% confident that the predicted price of a 450-page textbook from the true regression line lies in ($0.90, $124.85).

```
data <- data.frame(Pages=450)
predict.lm(model, data, interval="prediction")
```

```
##        fit       lwr      upr
## 1 62.87549 0.9035981 124.8474
```

c) The midpoints of both intervals are the same, which makes sense because it should be the fitted value for a 450-page textbook from the same regression line.

d) The width of the prediction interval is much larger than the width of the confidence interval. This makes sense because the prediction interval is about individual observations, while the confidence interval talks about means.

e) We want to minimize $SE_{\hat{\mu}} = \hat{\sigma}_\epsilon \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x-\bar{x})^2}}$, so we can minimize $(x^* - \bar{x})^2$ by setting the number of pages to be about 465.

f) The prediction interval would be: (143.359, 291.782). Since the value of the explanatory variable is 1500, it is way out of the range of values we were given to start with. This is extrapolating from our model, which makes me not as confident in this interval.

```
data <- data.frame(Pages=1500)
predict.lm(model, data, interval="prediction")
```

```
##        fit      lwr     upr
## 1 217.5704 143.3587 291.782
```

Problem 2.47

```
##
## Call:
## lm(formula = adj2007 ~ distance, data = RT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -190.55  -58.19  -17.48   25.22  444.41
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  388.204     14.052  27.626  < 2e-16 ***
## distance     -54.427      9.659  -5.635 1.56e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 92.13 on 102 degrees of freedom
## Multiple R-squared:  0.2374, Adjusted R-squared:  0.2299
## F-statistic: 31.75 on 1 and 102 DF,  p-value: 1.562e-07
```

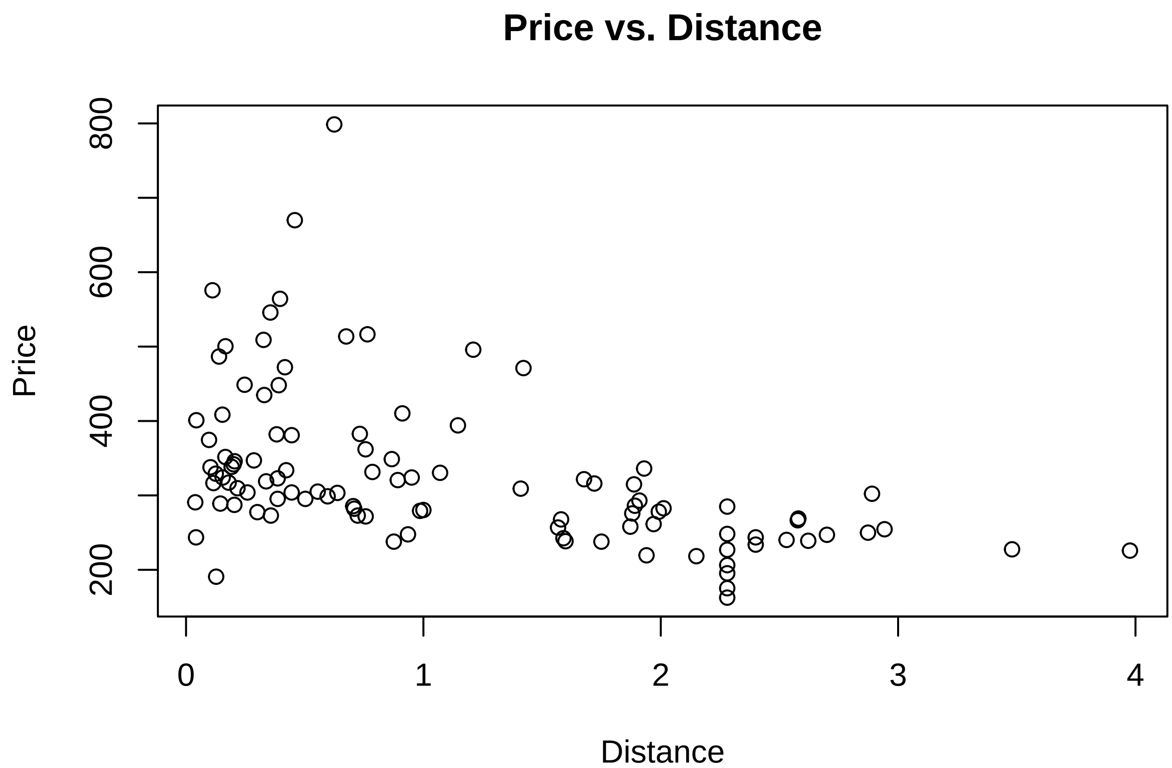
a) Price $= 388.204 - 54.427 \cdot 0.5 = \boxed{\$360.99}$

b) (177.003, 544.9774)

```r
data <- data.frame(distance=0.5)
predict.lm(model, data, interval="prediction")
```

```
##        fit     lwr      upr
## 1 360.9902 177.003 544.9774
```

c) The scatterplot shows a strong curvature, so we are not confident that our predicted price in part (a) is a good measure and we are not confident that our interval in part (b) is good.

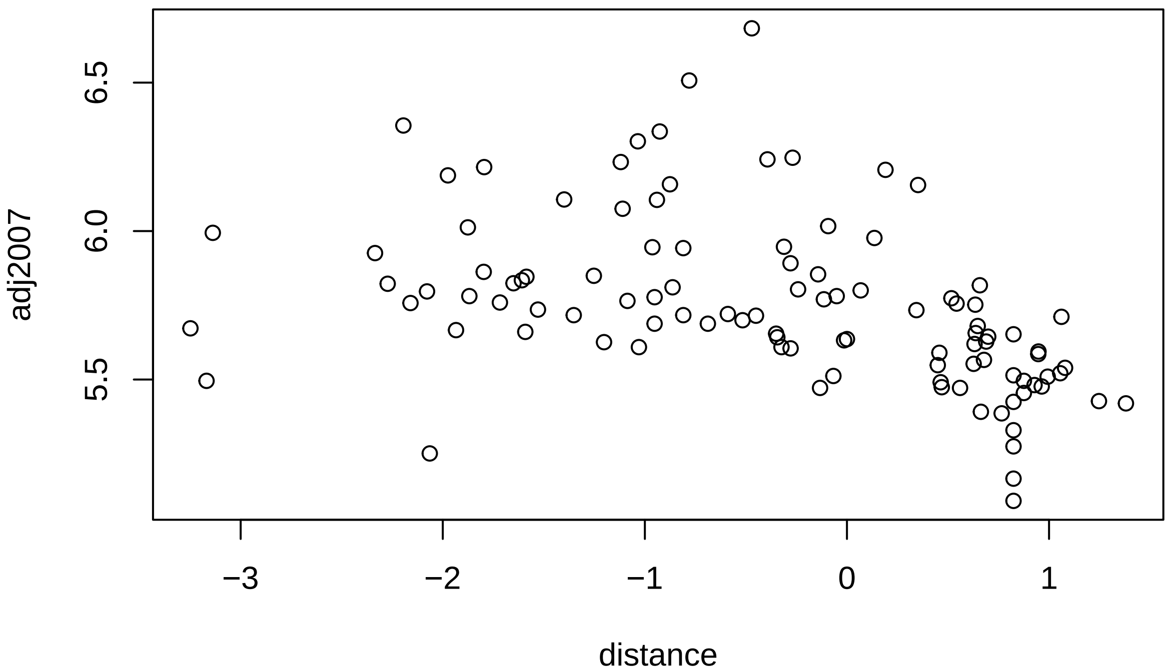


d) We are 90% confident the log of the price of a house that is 0.5 miles away from a bike trail will be between 5.272 and 6.296.

```
RT2 <- data.frame(distance=log(RT$distance),adj2007=log(RT$adj2007))
model <- lm(adj2007~distance, data=RT2)
data <- data.frame(distance= log(0.5))
predict.lm(model, data, interval="prediction")
```

```
##        fit      lwr      upr
## 1 5.784329 5.272453 6.296206
```

```
plot(adj2007~distance,data=RT2)
```



e) Converting the interval back to dollars, we get (\$194.89, \$542.51). Compared to our interval from before (\$177.003, \$544.9774), our new interval has a smaller width.