705604096_stats101a_hw4

Jade Gregory

2023-04-28

Question 1

a)

a)

```
playbill <- read.csv('playbill.csv')
playbill_lm <- lm(CurrentWeek ~ LastWeek, data = playbill)
summary(playbill_lm)
```

```
##
## Call:
## lm(formula = CurrentWeek ~ LastWeek, data = playbill)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -36926  -7525  -2581   7782  35443
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.805e+03  9.929e+03   0.685    0.503
## LastWeek    9.821e-01  1.443e-02  68.071   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18010 on 16 degrees of freedom
## Multiple R-squared:  0.9966, Adjusted R-squared:  0.9963
## F-statistic:  4634 on 1 and 16 DF,  p-value: < 2.2e-16
```

```
confint(playbill_lm)
```

```
##                    2.5 %       97.5 %
## (Intercept) -1.424433e+04 27854.099443
## LastWeek     9.514971e-01     1.012666
```

The confidence interval is (9.514971e-01, 1.012666). Because the value of 1 falls in between the bounds for the 95% confidence interval, we are able to conclude that 1 is a plausible value for

$$\beta_1$$

b)

```
confint(playbill_lm)["(Intercept)", ]
```

```
##      2.5 %     97.5 %
## -14244.33  27854.10
```

Since we are 95% confident that

$$\beta_0$$

falls in the interval (-14244.33, 27854.10) we fail to reject the null hypothesis.

c)

```
playbill_new = data.frame(LastWeek = 400000)
PI <- predict(playbill_lm, playbill_new, interval = "predict", level = .95)
PI
```

```
##        fit      lwr      upr
## 1 399637.5 359832.8 439442.2
```

Our predicted interval is (359832.8, 439442.2). It has the fitted value 399637.5 with lower bound 359832.8 and upper bound 439442.2. This interval does not include the value of 450,000 and therefore this value is not feasible.

d) I do not believe this is true because our predicted intervals will contain a range of options that do not necessarily equal the exact value from the previous week. There are many factors that can affect the observed value rather than the predicted value.

b

```
res <- resid(playbill_lm)
plot(fitted(playbill_lm), res)
abline(0,0)
```
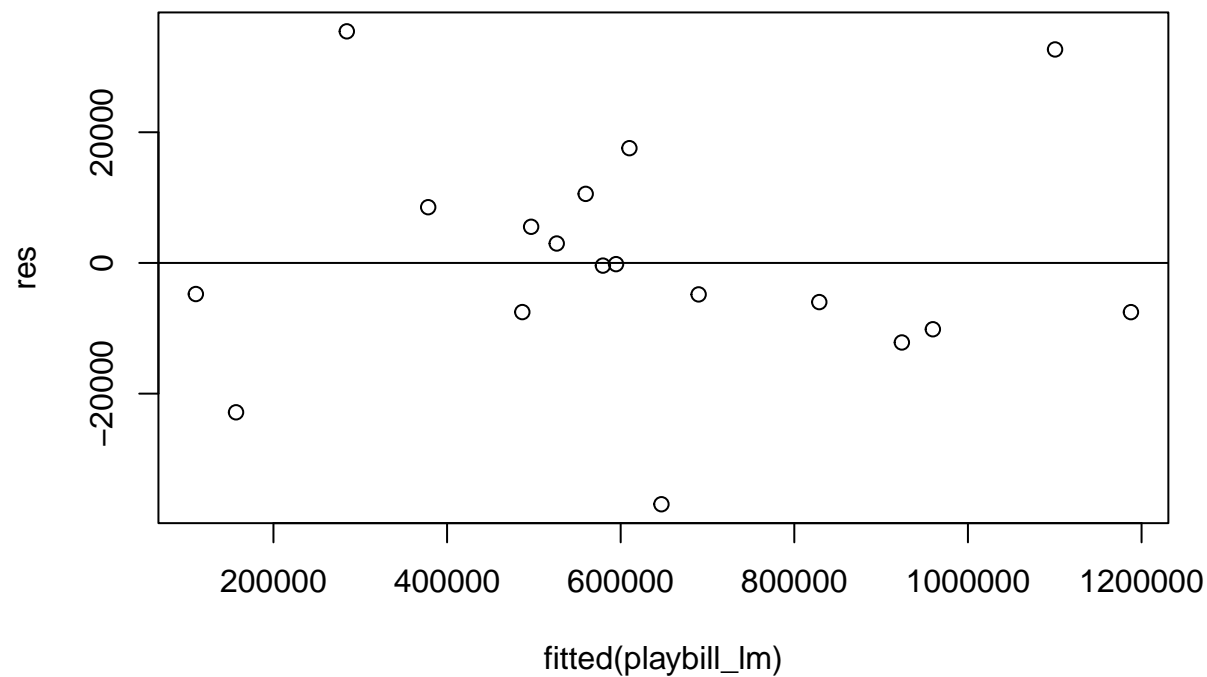
This plot seems valid as there is no observable pattern in the plotted points.

Question 2

a

a)

```
indicators <- read.delim('indicators.txt')
prices_lm <- lm(PriceChange ~ LoanPaymentsOverdue, data = indicators)
confint(prices_lm, "LoanPaymentsOverdue")
```

```
##                          2.5 %     97.5 %
## LoanPaymentsOverdue -6.759119 0.3731191
```

The data does suggest a negative linear association. The 95% confidence interval for

$$\beta_1$$

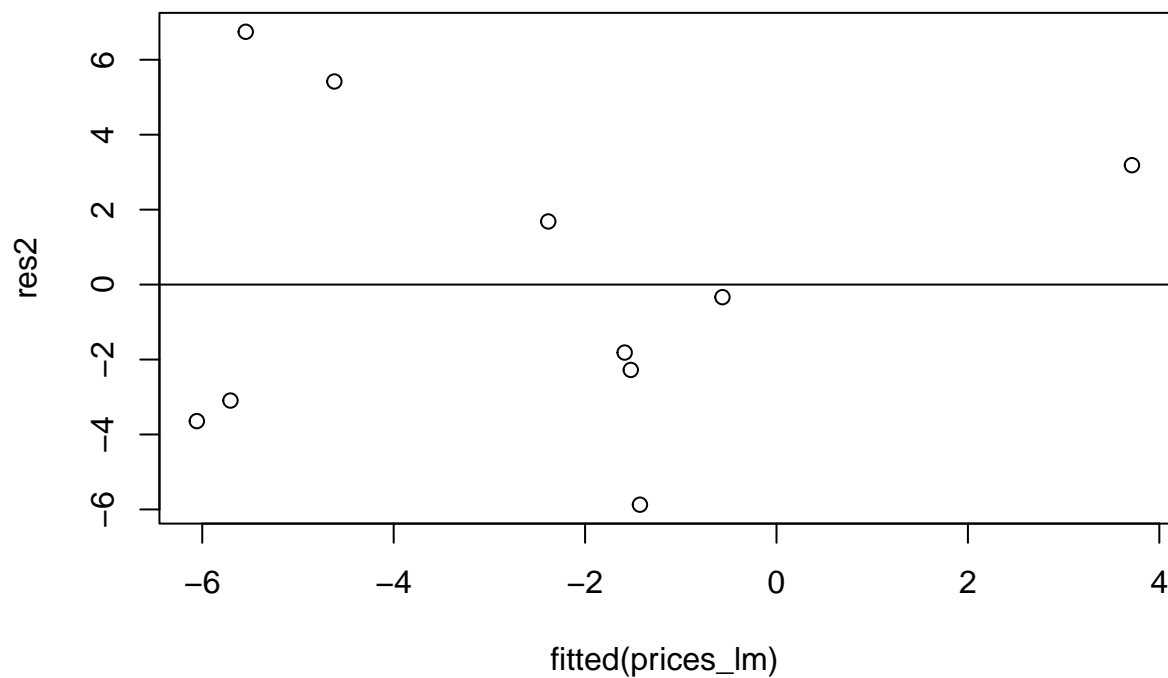is (-6.759119, 0.3731191)

b)

```
predict(prices_lm, data.frame("LoanPaymentsOverdue" = 4), interval = "predict")
```

```
##        fit      lwr      upr
## 1 -3.789726 -14.48141 6.901961
```

The expected value is -3.789726. Our 95% confidence interval is (-14.48141,6.901961). 0 is a feasible value because it falls in this interval.

b

```
res2 <- resid(prices_lm)
plot(fitted(prices_lm), res2)
abline(0,0)
```


This residual plot appears to be valid because there is no apparent pattern in the plotted points.

Question 3

a)

```
t_val <- qt(0.975, 28)
se <- 0.1222707
0.6417099 + c(-1,1) * t_val * se
```

[1] 0.3912497 0.8921701

The 95% CI for the start up time is (0.3912497, 0.89217016).

   b)

```
t_val <- qt(0.975, 28)
se2 <- 0.0008184

0.0112916 + c(-1,1) * t_val * se2
```

[1] 0.009615184 0.012968016

From our hypothesis test, we can see our 95% CI is (0.009615184, 0.012968016). This means we are 95% confident that we will observe a value between our bounds. Our value of 0.01 falls within this range, and therefore we fail to reject the null hypothesis since there is no significant evidence to suggest otherwise.

   c)

```
intercept <- 0.6417099
inv_slope <- 0.0112916
inv_num <- 130

point_estimate <- intercept + (inv_slope * inv_num)

t_val <- qt(0.975, 28)
rse <- 0.3298

point_estimate + c(-1,1) * t_val * rse
```

[1] 1.434053 2.785183

```
point_estimate
```

[1] 2.109618

Our 95% CI is (1.434053, 2.785183). Our point estimate is 2.109618.

Question 4

The correct option would be D. This is because the data points in model 1 are more close together and concentrated when compared to the data points in model 2, meaning that the RSS would be lesser in model 1 than in model 2. The SSreg is greater in model 1 because the regression line better fits the data set when compared to the regression line that is fit to the data in model 2, meaning the SSreg would be greater in model 1.