ProblemSet2

Aaron Graybill

2021-02-28

Contents

Problem 2.1

```r
df <-
  read.csv(here('Data',"playbill.csv"))
movie_lm <-
  lm(CurrentWeek~LastWeek,data=df)
summary(movie_lm)
```

```
##
## Call:
## lm(formula = CurrentWeek ~ LastWeek, data = df)
##
## 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
```

2.1.a

We can find the 95% confidence interval on the $\hat{\beta}_1$ term with the following code:

```r
confint_2.1 <-
  confint(movie_lm)
```

The confidence interval around $\hat{\beta}_1$ is $(0.9514971, 1.0126658)$ which actually contains 1, so we do not have significant evidence that the true $\beta_1 \neq 1$. We can compute this test a different way. We can run a hypothesis test where $H_0\colon \beta_1 = 1$ and $H_a\colon \beta_1 \neq 1$. We showed in class that:

$$\frac{\hat{\beta}_1 - 1}{se(\hat{\beta}_1)} \sim t_{n-2}$$

We have all of these values from our regression output, so we can compute the p-value that the null hypothesis is true in the following way:

```
test_stat=(movie_lm[["coefficients"]][["LastWeek"]]-1)/coef(summary(movie_lm))[, "Std. Error"]["LastWee

critical_region <-
  qt(c(.025,.975),movie_lm$df.residual)

p_val <- (pt(test_stat,movie_lm$df.residual)*2)
```

The 95% critical region, as computed is $(-\infty, -2.12) \cup (2.12, \infty)$, and our test statistic is equal to -1.2419935 is not in that region, so we have insufficient evidence to reject the null hypothesis that $\beta_1 = 1$ at the 95% level. In fact, the p-value from this test is 0.2321368 which is well above the required .05 at the 95% level.

2.1.b

As proven in class:

$$\frac{\hat{\beta}_0 - \beta_0^0}{se(\hat{\beta}_0)} \sim t_{n-2}$$

In this case our H_0 is $\beta_0 = 1000$ and H_a is $\beta_0 \neq 1000$. Implementing similar code to above gives:

```
test_stat=(movie_lm[["coefficients"]][["(Intercept)"]]-1000)/coef(summary(movie_lm))[, "Std. Error"]["(

critical_region <-
  qt(c(.025,.975),movie_lm$df.residual)

p_val <- (pt(-abs(test_stat),movie_lm$df.residual)*2)
```

I have to do some business with `-abs(test_stat)` to ensure that when I compute the cumulative density it's in the left tail so the p-value is just two times the computed density. Anyway, computing the p-value gives: 0.5669558 which is quite high and provides very little evidence that the true $\beta_0 \neq 1000$. In fact since $p > .5$, there is more evidence that H_0 is true than the alternative.

2.1.c

We showed in class that the prediction interval is given by:

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\frac{\alpha}{2},n-2} S \sqrt{\frac{1}{n} + \frac{x^* - \overline{x}}{SXX}}$$

But thankfully `R`will take care of that for us with the following code:

```
prediction_points <-
  data.frame(LastWeek=400000)
prediction <-
  predict(movie_lm,prediction_points,interval = 'prediction',level=.95,se.fit=T)
prediction
```

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

```
## $se.fit
## [1] 5318.889
##
## $df
## [1] 16
##
## $residual.scale
## [1] 18007.56
```

Summarizing those results, the point estimate for `CurrentWeek` is \$399637.5 with a 95% prediction interval of: $(359832.8, 439442.2)$. So we are 95% percent certain the true value of y^* would lie in the aforementioned range.

Problem 2.2