A-Modern-Approach-to-Regression-with-R–Chapter 2,Exercises 1

Soodabeh Ramezani

September 2

1. The web site www.playbill.com provides weekly reports on the box office ticket sales for plays on Broadway in New York. We shall consider the data for the week October 11–17, 2004 (referred to below as the current week). The data are in the form of the gross box office results for the current week and the gross box office results for the previous week (i.e., October 3–10, 2004). The data, plotted in Figure 2.6 , are available on the book web site in the file playbill.csv.

Fit the following model to the data: Y=beta0+beta1X+e where Y is the gross box office results for the current week and x is the gross box office results for the previous week.

Answer:

```
setwd("C:/Users/soodr/Documents/STAT5310_Applied_Regression_Analysis/data")
playbill<- read.delim("playbill.csv", header = T, sep = ",")
y=playbill$CurrentWeek
x=playbill$LastWeek
Model_pb=lm(y~x)
summary(Model_pb)

##
## Call:
## lm(formula = y ~ x)
##
## 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
## x           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

Mod_sum=summary(Model_pb)
```

```
#library(tidyverse)
#ggplot(playbill, aes(x = LastWeek, y = CurrentWeek)) +
 # geom_point() +
 # stat_smooth(method = "lm", col = "red")
```

a) Find a 95% confidence interval for the slope of the regression model,b1. Is 1 a
   plausible value for b1? Give a reason to support your answer.

Answer_a_part1:

```
ci=confint(Model_pb,level=0.95)
cat("95% confidence interval for the slope:", ci[2,])
```

```
## 95% confidence interval for the slope: 0.9514971 1.012666
```

Answer_a_part2:

1 is a plausible value for b1 since it is between 9.514971e-01 and 1.012666.

b)Test the null hypothesis H0:beta0=10000 against a two-sided alternative. Interpret your
result.

Answer_b:

The large p-value~0.7517807 indicates weak evidence against the null hypothesis (fail to
reject). We therefore conclude that beta_zero may have a value equal to 10000.

```
beta_null=10000
beta1_Estimate=coefficients(Mod_sum)["(Intercept)","Estimate"]
se=coefficients(Mod_sum)["(Intercept)","Std. Error"]
t=(beta1_Estimate-beta_null)/se
n=dim(playbill)[1]
p_value= 2*pt(-abs(t),df=n-2)
cat('t test:',t,'p_value:', p_value,sep = "\t")
```

```
## t test:   -0.3217858   p_value:      0.7517807
```

c) Use the fitted regression model to estimate the gross box office results for the current
   week (in $) for a production with $400,000 in gross box office the previous week. Find
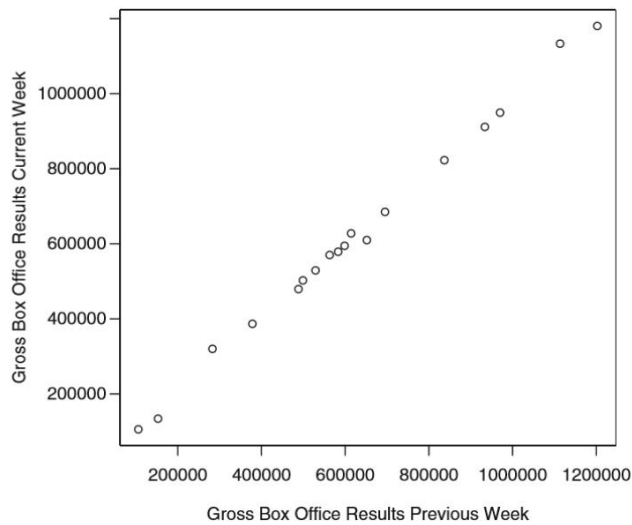   a 95% prediction interval for the gross box office

Figure 2.6 Scatter plot of gross box office results from Broadway

results for the current week (in $) for a production with $400,000 in gross box office the previous week. Is $450,000 a feasible value for the gross box office results in the current week, for a production with $400,000 in gross box office the previous week? Give a reason to support your answer.

Answer_c:

since $450,000 is not between the 95% prediction interval, it is not a feasible value.

```
#part_1
b1=coefficients(Mod_sum)[2,1]
b0=coefficients(Mod_sum)[1,1]
x=400000
y=b0+b1*x
cat("current gross box office for a production with $400,000:",y)

## current gross box office for a production with $400,000: 399637.5

cat(sep = "\n" )

cat(sep = "\n" )

#part_2
predict(Model_pb,data.frame(x = c(400000)), interval = "prediction", conf.int
= 0.95)

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

d)Some promoters of Broadway plays use the prediction rule that next week's gross box office results will be equal to this week's gross box office results. Comment on the appropriateness of this rule.

Answer_d:

logically, next week's gross box office may not be equal to this week's gross box office results due to some unexpected conditions such as inclement weather conditions which make the visitors not show up in a week compared to its previous week. Another reason could be introduction of new shows which make the gross box office deviate from its existing trend. Statistically, the fitted regression line has a slope of ~0.98 with 95% confidence interval between 0.9514971 and 1.012666 which is not exactly One, and our best estimate is that a 1 unit increase in gross box office for the previous week corresponds to an increase of 0.98 unit in gross box office for the current week.