Homework 1 , Stat 632

Chandnee das

1/29/2022

# Github page link

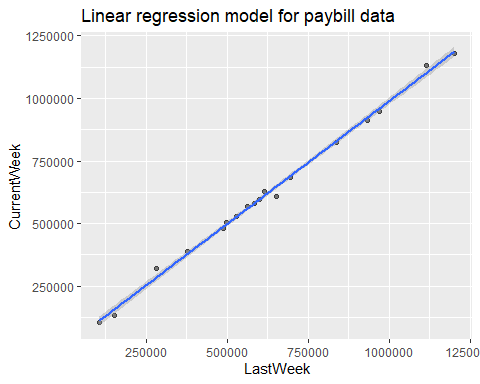
<https://github.com/chandnee-das>

# Exercise 3

### (a)

playbill <- read.csv('playbill.csv')  
library(ggplot2)  
ggplot(playbill, aes(LastWeek, CurrentWeek)) + geom\_point(size=1.5, alpha=0.5) + geom\_smooth(method='lm')+ggtitle("Linear regression model for paybill data")

## `geom\_smooth()` using formula 'y ~ x'



### (b) 95% confidence interval calculation:

lm1 <- lm(LastWeek ~ CurrentWeek,data=playbill)  
confint(lm1, level=0.95)

## 2.5 % 97.5 %  
## (Intercept) -2.632500e+04 16796.579076  
## CurrentWeek 9.831401e-01 1.046343

1 is a plausible value for since this value falls within 95% confidence interval for slope.

### (c) Prediction of current week’s box office result

df <- data.frame(CurrentWeek=400000)  
pred\_val = predict(lm1,newdata=df, interval = 'prediction')  
pred\_val

## fit lwr upr  
## 1 401132.4 360675.2 441589.6

From the above calculation, we see that it is unlikely to get $450000 box office result for current week.

### (d)

The promoter’s rule is not appropriate because each week the box office result will be lower than previous week’s result by 26325 on average.

# Exercise 4

### (a)

library(alr4)

## Loading required package: car

## Loading required package: carData

## Loading required package: effects

## lattice theme set by effectsTheme()  
## See ?effectsTheme for details.

lm3 <- lm(Interval ~ Duration,data=oldfaith)  
summary(lm3)

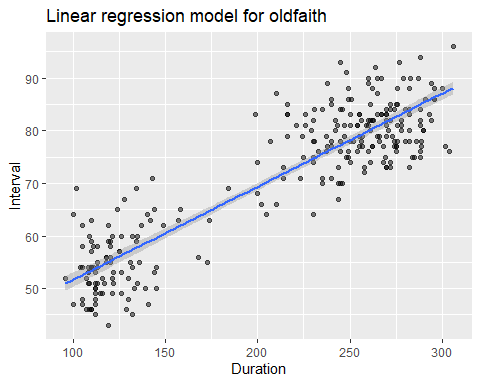
##   
## Call:  
## lm(formula = Interval ~ Duration, data = oldfaith)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.3337 -4.5250 0.0612 3.7683 16.9722   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 33.987808 1.181217 28.77 <2e-16 \*\*\*  
## Duration 0.176863 0.005352 33.05 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.004 on 268 degrees of freedom  
## Multiple R-squared: 0.8029, Adjusted R-squared: 0.8022   
## F-statistic: 1092 on 1 and 268 DF, p-value: < 2.2e-16

### (b)

##### Making Scatter Plot

library(ggplot2)  
ggplot(oldfaith, aes(Duration, Interval)) + geom\_point(size=1.5, alpha=0.5) + geom\_smooth(method='lm')+ggtitle("Linear regression model for oldfaith")

## `geom\_smooth()` using formula 'y ~ x'

 ### (c)

df <- data.frame(Duration=250)  
pred\_val = predict(lm3,newdata=df, interval = 'prediction')  
pred\_val

## fit lwr upr  
## 1 78.20354 66.35401 90.05307

pred2 <- predict(lm3)

The predicted amount is 78.20354 seconds. For 95% prediction interval, the wait time range is 66.35401 to 90.05307 seconds.

### (d)

Calculated value of 0.8029 means that about 80% of the variability of the interval is explained by duration.