Assignment-two-final.R

shahid

2023-10-23

Budget=c(155,9,154,59,30,191,197,138,73,70,61,48,51,52,39,256,80,45,35,44,24,39,30,29,24,19,25,23,11,21,23,17,78,81,127,143,11,21,7,21,163,36,60)  
Opening=c(60.39,18.13,18.57,9.91,15.71,33.47,63.52,74.08,58.38,19.13,9.11,141.64,15.53,16.20,28.90,46.95,9.96,30.44,52.93,30.65,15.11,46.86,33.73,8.60,25.56,9.83,30.61,52.05,19.08,19.13,25.81,9.37,33.85,40.81,53.73,81.58,4.43,23.27,12.00,44.63,42.87,14.04,22.52)  
Theatres=c(3778,2858,3182,2814,2434,3620,3394,3842,3488,3566,2128,4017,3353,2741,2746,3592,3691,3191,3647,2900,2908,3280,3155,2583,1891,2426,2970,3091,3227,3236,2817,3092,3264,3182,3602,4078,2687,3027,1193,3111,4050,2754,3260)  
Ratings=c(8.2,7.3,5.0,7.3,6.6,5.5,5.4,7.7,8.2,6.3,7.5,4.3,5.4,7.4,5.6,7.4,4.1,7.1,5.5,7.5,5.2,8.3,5.7,5.1,7.3,5.6,6.3,7.5,5.2,3.6,6.0,4.7,6.3,9.0,7.2,5.8,4.5,7.3,7.7,5.9,8.0,6.5,7.3)  
USRevenue=c(289.0,59.0,134.9,30.5,42.3,84.3,163.5,256.0,254.5,52.7,38.3,294.8,62.3,36.8,57.4,198.1,40.1,98.0,172.5,93.5,37.2,280.1,149.2,17.2,82.7,28.1,80.2,119.8,57.1,37.5,48.1,21.5,101.5,124.6,105.3,174.7,12.5,73.4,31.9,66.6,216.1,50.2,91.5)  
Movies=data.frame(USRevenue, Budget, Opening, Theatres, Ratings)  
  
# (a) State the value of the F statistic used to test the hypothesis that   
# β1=β2=β3=β4=0 versus β1≠0 or β2≠0 or β3≠0 or β4≠0 .  
movies\_model <- lm(USRevenue~Budget+Opening+Theatres+Ratings, data=Movies)  
summary(movies\_model)

##   
## Call:  
## lm(formula = USRevenue ~ Budget + Opening + Theatres + Ratings,   
## data = Movies)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -76.667 -18.001 -5.485 13.606 113.050   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -183.79203 55.81658 -3.293 0.00215 \*\*   
## Budget 0.13295 0.12204 1.089 0.28283   
## Opening 1.94084 0.29377 6.607 8.42e-08 \*\*\*  
## Theatres 0.03124 0.01509 2.070 0.04531 \*   
## Ratings 18.34156 4.79275 3.827 0.00047 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 37.63 on 38 degrees of freedom  
## Multiple R-squared: 0.8016, Adjusted R-squared: 0.7807   
## F-statistic: 38.38 on 4 and 38 DF, p-value: 7.326e-13

anova(movies\_model)

## Analysis of Variance Table  
##   
## Response: USRevenue  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Budget 1 71241 71241 50.2985 1.846e-08 \*\*\*  
## Opening 1 124227 124227 87.7077 2.045e-11 \*\*\*  
## Theatres 1 1205 1205 0.8511 0.3620742   
## Ratings 1 20743 20743 14.6455 0.0004702 \*\*\*  
## Residuals 38 53822 1416   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# (b) Is there evidence that a model with Opening and Budget is better than   
# a model with just Budget? State the relevant test statistic, p -value   
# and conclusion.   
movies\_opening\_buget\_model <- lm(USRevenue~Opening+Budget, data=Movies)  
movies\_budget\_model <- lm(USRevenue~Budget, data=Movies)  
anova(movies\_budget\_model,movies\_opening\_buget\_model)

## Analysis of Variance Table  
##   
## Model 1: USRevenue ~ Budget  
## Model 2: USRevenue ~ Opening + Budget  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 41 199998   
## 2 40 75771 1 124227 65.58 5.822e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# (c) Conduct the appropriate F test to test whether a model   
# containing all the predictors is preferred over a model with Budget as the   
# predictor. State the relevant test statistic, p -value and conclusion.  
movies\_budget <- lm(USRevenue~Budget, data=Movies)  
anova(movies\_budget,movies\_model)

## Analysis of Variance Table  
##   
## Model 1: USRevenue ~ Budget  
## Model 2: USRevenue ~ Budget + Opening + Theatres + Ratings  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 41 199998   
## 2 38 53822 3 146176 34.401 6.381e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# (d) Is there evidence that Theatres is related to the response in the presence   
# of Budget, Opening and Ratings? State the relevant test statistic,   
# p -value and conclusion.  
movies\_model\_two <- lm(USRevenue~Theatres+Budget+Opening+Ratings, data=Movies)  
summary(movies\_model\_two)

##   
## Call:  
## lm(formula = USRevenue ~ Theatres + Budget + Opening + Ratings,   
## data = Movies)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -76.667 -18.001 -5.485 13.606 113.050   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -183.79203 55.81658 -3.293 0.00215 \*\*   
## Theatres 0.03124 0.01509 2.070 0.04531 \*   
## Budget 0.13295 0.12204 1.089 0.28283   
## Opening 1.94084 0.29377 6.607 8.42e-08 \*\*\*  
## Ratings 18.34156 4.79275 3.827 0.00047 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 37.63 on 38 degrees of freedom  
## Multiple R-squared: 0.8016, Adjusted R-squared: 0.7807   
## F-statistic: 38.38 on 4 and 38 DF, p-value: 7.326e-13

# (e) It had budget of $ 30 million and was shown in 3,065 theatres,   
# grossing $ 19.83 million during the first weekend. The movie's IMDb rating   
# was 7.6 out of 10. Obtain a 99% prediction interval for the USRevenue   
# based on the model with all four predictors.  
newdata <- data.frame(Budget = 30, Opening = 19.83, Theatres = 3065, Ratings= 7.6)  
pi<-predict.lm(movies\_model,newdata,interval = c("prediction"), level = 0.99); pi

## fit lwr upr  
## 1 93.83617 -12.24059 199.9129