Varun Khadayate A016 30-10-2020

C.S

MLR

Q1. a. $\hat{y} = -22.9898 + 1.1344 X_1 + 24.9724 X_2 - 0.4033 X_3$

b. $X_1 = 32$; $X_2 = 3$; $X_3 = 100$

$\hat{y} = -22.9898 + 1.1344 [32] + 24.9724 [3] - 0.4033 [100]$

$= 47.8982$

c. Yes, since from ANOVA table : P-Val < 0.0001. Hence the model is ~~satisfactori~~ statistically significant at 5% LOS

d. Since P-Value for Run Time is 0.1113 > 0.05 LOS, variable run time is significant. Hence, we remove variable Run Time.

e. $R^2 = 0.4739 = 47.39\%$

About 47.39% of variation in US Gross Variable is explained by variations in independent variables Budget, Stars & Run Time.

Q2. b1. No, it is not a good model. Cause, the no. of observations are really small for the model to be able to estimate anything with precision & reliability.

b2. The intercept gives the average value of ~~dep~~ demand for subscription when all the 3 independent variables take the value 0 simultaneously. There is no economic meaning in it, because subscription can't be given for free & ~~non~~ number of movies has to be strictly +ve.

b3. Monthly Price

Expected Sign: -ve, given the inverse relationship between price & demand, Actual Sign: +ve

Interpretation: If ~~no. of movies~~ monthly price is increased by 1 unit, demand for subscription will increase by 111.21 units

No. Of Movies

Expected Sign: +ve, more movies means more demand for subscription, Actual Sign: +ve

Interpretation: If no. of movies is increased by 1 unit, demand for subscription will increase by 10.08 units.

Price Of Rival

Expected Sign: +ve, given the substitutional effect of a high rival price makes own good relatively cheaper and hence more demanded Actual Sign: -ve.

Interception: If price of rival increases by 1 the demand for subscription will decrease by 605.17 units

Statistical Significance: For 5% LOS, p-Val = 0.05 & for 10% LOS p-val = 0.1
If computed p-val is less than 0.05 then the variable is significant at 5% LOS, otherwise it's not.
Same for 10%.

b4. $\hat{y} = 2905.9639 + 111.2130 X_1 + 10.0871 X_2 - 605.1733 X_3$
$X_1 = 8;\ X_2 = 4258;\ X_3 = 7$
$\hat{y} = 2905.9639 + 111.2130[8] + 10.0871[4258] - 605.1733[7]$
$\hat{y} = 42429.6298$

C. 1. Buy new data which has more observations.
2. Include more variables like income of consumer, monthly consumption etc.

Q3. There are 4 predictors: X_1, X_2, X_3 & X_4, i.e, $K = 4$.
Eqn: $\hat{y} = -55.93 + 0.0105 X_1 - 0.1072 X_2 + 0.5792 X_3 - 0.8695 X_4$

Q4. There are 2 predictors: X_1 & X_2, i.e, $K = 2$
Eqn: $\hat{y} = 203.3932 + 1.1151 X_1 - 2.2115 X_2$

Q5. $F^* = \frac{MSR}{MSE} = \frac{504.04/1}{720.27/48} = 33.59$
Associated p-val < 0.001.
We can conclude that there is statistically significant linear association between lifetime alcohol consumption & arm strength.

Q6. $F^* = \frac{0.95927/3}{0.54491/28} = 16.43$

There is sufficient evidence to conclude that at least one slope parameter is not equal to 0.

Q8. There are 3 predictors available in the model.
There are 15 observations in the model.
$\hat{y} = 657.0530 + 5.7103X_1 - 0.4169X_2 - 3.4715X_3$
Since the F-value is small the model is a very good predictor.
Since, the p-value of X_1 is quite less, it is significant predictor variable.
The overall model is very effective.

Q9. $F(3,13) = 4.61$; $p = 0.021$; $p < 0.05$
Regression model is significant.
For independent variables:
For X_1, $p = 0.005$, $p < 0.05$, X_1 is a significant variable.
For X_2, $p = 0.805$, $p > 0.05$, X_2 is not a significant variable.
For X_3, $p = 0.230$, $p > 0.05$, X_3 is not a significant variable.
Since, slopes of X_1 & X_2 are very small, so the coefficient estimate is very precise.

Q10. There are 2 predictors: X_1 & X_2, i.e, $k = 2$
Eqn: $\hat{y} =$

Q7. $F^* = \frac{MSR}{MSE} = \frac{116.65}{14.12} = 8.26$ $\quad F = 8.26$

p-val < 0.001

We can conclude that there is statistically significant linear association.