

Data Analysis In-Class Worksheet #10: Regression 2 of 2

Does the number of clicks (CLICKS) predict the number of stars (STAR) in ratings?

To answer this question, we can rephrase it as a pair of hypotheses:

H_0 : Number of clicks does not predict number of stars (no linear relationship)

H_1 : Number of clicks does predict number of stars (there is a linear relationship).

Express these hypotheses in terms of the regression β_1 slope

H_0 : $\beta_1=0$

H_1 : $\beta_1 \neq 0$

For H_0 : $\hat{Y} = \beta_0$

For H_1 : $\hat{Y} = \beta_0 + \beta_1 * \text{Clicks}$ ($\beta_1 \neq 0$)

Assume the sample size is 101, and the variance of the STAR variable is 2.

Calculate the Sum of Squares Total (SST) = $(n-1) \cdot \text{Var}_{\text{STAR}} = (101-1) \cdot 2 = 100 \cdot 2 = 200$

What is the Degrees of Freedom Total (DFT)? $N-1 = 101-1=100$

Assume that the Sum of Squares Residual (SSR) is 120.

Calculate the Sum of Squares Model (SSM) = $\text{SST}-\text{SSR} = 200-120=80$

What is the Degrees of Freedom Model (DFM)? No of predictors= 1

Calculate Mean Squares Model (MSM) = $\text{SSM}/\text{DFM}=80/1=80$

What is the Degrees of Freedom Residual (DFR) ? $n-2=101-2=99$

Calculate Mean Squares Residual (MSR) = $\text{SSR}/\text{DFR}= 120/99=1.212$

Calculate the F ratio = $\frac{\text{MSM}}{\text{MSR}} = \frac{80}{1.212}=66$

Use Google Sheets to determine the p-value of this F ratio

What Google Sheets formula did you use? $\text{F.DIST.RT}(66, 1, 99)$ $p<0.001$

Calculate R^2 using SSM & SST= $\text{SSM}/\text{SST}= 80/200=0.4$

So the correlation between CLICK and STAR is $\text{sqrt}(R)=0.632$

Calculate Adjusted $R^2 = 1 - ((1-R^2) * ((n-1)/(n-2))) = 1 - (0.6 * (100/99)) = 0.394$

APA Writeup

Regression analysis was used to test if Clicks significantly predicted Stars. Results indicated that the predictor explained 40% of the variance (Adjusted R^2 = 0.394,)

$F(1, 99) = 66, p < 0.001$.