**DSCI 5180 INTRODUCTION TO BUSINESS DECISION PROCESS**

**Final Project – Top 1000 Highest Grossing Movies of All Time IMBD**

**Name: Prashanth Kuchur**

**Student ID: 11626460**

**Data Source**

<https://www.kaggle.com/datasets/therealoise/top-1000-highest-grossing-movies-of-all-time>

**Applying Module 1 Learnings**

Suppose the mean Top 1000 Highest Grossing Movies rating of all movies in IMBD is 6.8 with a standard deviation of 0.9. If the movie ratings are normally distributed, what is the probability that a randomly selected movie will have a movie rating between 7.4 and 8?

Solution 1

P(7.4<X<8.0)

Conversion to Standard Normal using Z =

P(0.67<Z<1.33)

=P(Z<0.67)-P(Z<1.33)

=0.9082 – 0.7486

=0.1596

The Probability of a randomly selected movie that have a movie rating between 7.4 and 8 is 0.1596

Refer Appendix 1 for more details

**Applying Module 2 Learnings**

A movie producer wants to know the mean duration of movies. Assume that the population standard deviation is 22 minutes, mean duration per movie is 117.67 minutes for a sample of 1000 movies. Construct the 95% confidence interval for the mean duration of a movie. Round the answer to two decimal places.

Solution 2

Confidence Interval for Population Mean =  – E < µ <  + E

Population Standard Deviation Known, n 30 then E =

Confidence Interval = E

=117.671.96

116.30 and 119.03

So, we are 90% confident that the true mean duration of movies is between 116.30 and 119.03

Refer Appendix 2 for more details.

**Applying Module 3 Learnings**

IMBD claims that movies have an average gross of $156 million. Being a skeptic, you decide to test and see if the mean gross income of movies is actually more than $156 Million. for the random sample of 27 movies, Mean is $166 million with a sample standard deviation of $98 million. Assume that the population distribution is approximately normal. Is there sufficient evidence to support your statement at 0.10 level of significance?

Solution 3

:

:

t =

0.53

P=0.30022

The alternative hypothesis contains a "greater than" symbol, which indicates that this is a right-tailed test. p-value is approximately 0.3002. Since 0.3002>0.10, we have p-value>α. Thus, the conclusion is to fail to reject the null hypothesis.

There is insufficient evidence at a 0.10 level of significance that mean gross income of movies is more than $156 Million.

Refer Appendix 3 for more details.

**Applying Module 4 5 Learnings**

The following data was collected to explore how Duration and Votes effects the Movie rating. The dependent variable is Movie Rating, the first independent variable (x1) is Duration and the second independent variable (x2) is Votes.Determine if a statistically significant linear relationship exists between the independent and dependent variables at the 0.05 level of significance. If the relationship is statistically significant, identify the multiple regression equation that best fits the data.

Solution 4 & 5

R Square = 0.440366498 - Refer Appendix 2, Table 4.1

p-value = 2.1379E-126 – Refer Appendix, Table 4.2

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 5.660184588 | 0.121485859 | 46.59130416 | 1.6203E-252 | 5.42178727 | 5.898581905 | 5.42178727 | 5.898581905 |
| Duration | 0.004792171 | 0.001082093 | 4.428615132 | 1.05316E-05 | 0.002668731 | 0.006915612 | 0.002668731 | 0.006915612 |
| Votes | 1.69922E-06 | 7.2205E-08 | 23.5332275 | 9.57172E-98 | 1.55753E-06 | 1.84091E-06 | 1.55753E-06 | 1.84091E-06 |

# The equation for a multiple regression model is +…. +

**=**5.660184588 + 0.004792171+1.69922E-06