**Movie Success: Analyzing Factors that Effect the Box Office Performance of Movies**

Chetan Kumar Suppala

Lubar School of Business, University of Wisconsin-Milwaukee

BUS ADM 813: Social Media Analytics for Business

Dr. Yang Wang

May 16, 2023

|  |  |
| --- | --- |
| *DESCRIPTIVE STATISTICS* | |
|  |  |
| Mean | 5.483069977 |
| Standard Error | 0.024130186 |
| Median | 5.7 |
| Mode | 6.2 |
| Standard Deviation | 1.606062842 |
| Sample Variance | 2.579437852 |
| Kurtosis | -0.92156881 |
| Skewness | -0.17883764 |
| Range | 6.6 |
| Minimum | 2.1 |
| Maximum | 8.7 |
| Sum | 24290 |
| Count | 4430 |
| Largest(1) | 8.7 |
| Smallest(1) | 2.1 |
| Confidence Level(95.0%) | 0.047307224 |

A basic descriptive statistics test has been run to determine the normality of the data. By using data analysis add-in in excel descriptive statistics test has been run. According to the kurtosis value, we can conclude that the data is normal. (For normal data this value is between -1 to 1)



A correlation test has been run between overall rating and movie reviews. The formula used for calculating has been shown in the image above. The correlation result shows a value of 0.43 approximately which indicates that it is positively correlated i.e., increase in one variable induces an increase in another and vice versa.

A picture containing text, screenshot, diagram, plot

Description automatically generated

A picture containing text, screenshot, plot, line

Description automatically generated

Scatter plot shown above has been drawn using excel graph feature. The scatter plot line indicates the trajectory of the scatter plot. The line drawn is approximation of the overall data points and not exact. Thus, some data points may lie on the line and some may not.

|  |  |
| --- | --- |
| 0.819175 | positive correl |
|  |  |

The above correlation shows correlation between box office and screens, which is also positively correlated indicating increase in one variable increase in another.

Multiple Regression:

Multiple regression test has been run on box office records as dependent variable and overall rating and number of screens as independent variable. We wanted to predict box office values with respect to number of screens and overall rating. The results are as follows:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.727527174 |  |  |  |  |  |  |  |
| R Square | 0.529295789 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.508830388 |  |  |  |  |  |  |  |
| Standard Error | 2.13600045 |  |  |  |  |  |  |  |
| Observations | 49 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 2 | 235.999382 | 117.9996911 | 25.8629578 | 2.9731E-08 |  |  |  |
| Residual | 46 | 209.874904 | 4.562497923 |  |  |  |  |  |
| Total | 48 | 445.874287 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | -15.50835609 | 9.29390954 | -1.668657955 | 0.10197877 | -34.2160257 | 3.19931352 | -34.2160257 | 3.19931352 |
| overall\_rating | 6.729653745 | 4.01805658 | 1.674852907 | 0.10074947 | -1.35827466 | 14.8175822 | -1.35827466 | 14.8175822 |
| screens | 0.00065952 | 0.00010451 | 6.310439157 | 9.9022E-08 | 0.00044915 | 0.00086989 | 0.00044915 | 0.00086989 |

The multiple R value in summary output represents that there is good regression possible between selected variables.

A picture containing text, screenshot, font, line

Description automatically generated

A picture containing text, screenshot, font, line

Description automatically generated

**PREDICTED BOX OFFICE:**

|  |  |  |  |
| --- | --- | --- | --- |
| *Observation* | *Predicted box\_office* | *Residuals* | *Standard Residuals* |
| 1 | -1.295621732 | 1.3286926 | 0.63542585 |
| 2 | 8.516723446 | -4.578141 | -2.189422139 |
| 3 | 7.197682556 | -3.827604 | -1.830489921 |
| 4 | 5.878641667 | -3.305413 | -1.580760721 |
| 5 | 5.219121222 | -3.011247 | -1.440080511 |
| 6 | 3.409397122 | 0.2976895 | 0.142365209 |
| 7 | 6.551607041 | 7.7901252 | 3.725502017 |
| 8 | 6.551607041 | 7.448393 | 3.562074053 |
| 9 | 4.007836686 | -2.631459 | -1.258452786 |
| 10 | 1.929028244 | -1.272335 | -0.608473894 |
| 11 | 1.807016961 | -1.211741 | -0.579495808 |
| 12 | 1.328864639 | -0.853274 | -0.408064599 |
| 13 | 0.689789328 | 0.1511556 | 0.072287711 |
| 14 | 3.483517932 | 2.9180569 | 1.395513733 |
| 15 | 3.574531754 | 2.8175942 | 1.347469092 |
| 16 | 0.84081951 | -0.527434 | -0.25223667 |
| 17 | 0.141068318 | 0.0069632 | 0.003330028 |
| 18 | 0.102156612 | 0.0490245 | 0.023445174 |
| 19 | 0.01114279 | 0.1463375 | 0.069983566 |
| 20 | -0.252665388 | 0.4542402 | 0.21723306 |
| 21 | 1.156070282 | 0.8864494 | 0.42393016 |
| 22 | 1.1824511 | 0.8018009 | 0.38344836 |
| 23 | -0.252665388 | 0.3566024 | 0.170539356 |
| 24 | -0.309384146 | 0.3818251 | 0.182601704 |
| 25 | -0.325212637 | 0.394504 | 0.188665177 |
| 26 | -0.440628714 | 0.5036208 | 0.240848561 |
| 27 | 0.248824671 | -0.0441 | -0.02109024 |
| 28 | 1.081139473 | 0.1109865 | 0.053077513 |
| 29 | 1.292186015 | 0.1912786 | 0.091475889 |
| 30 | 0.264653162 | -0.08985 | -0.042969322 |
| 31 | 0.09911353 | 0.0268707 | 0.01285049 |
| 32 | 0.052947099 | 0.017919 | 0.008569494 |
| 33 | 1.466299412 | 1.3588974 | 0.649870829 |
| 34 | 0.771824384 | 0.4077032 | 0.194977482 |
| 35 | -0.022897752 | 0.0874647 | 0.041828576 |
| 36 | 0.681065083 | -0.147207 | -0.07039929 |
| 37 | 0.668534195 | -0.3394 | -0.162312748 |
| 38 | 0.665236593 | -0.60067 | -0.287260599 |
| 39 | 0.654024745 | -0.589458 | -0.281898713 |
| 40 | 0.650067622 | -0.623296 | -0.298081265 |
| 41 | 0.652046184 | -0.618975 | -0.296014983 |
| 42 | 0.657981868 | -0.561919 | -0.268728658 |
| 43 | 0.657322347 | -0.569133 | -0.272178875 |
| 44 | 0.654684266 | -0.57122 | -0.273176627 |
| 45 | 0.649408102 | -0.614762 | -0.294000239 |
| 46 | 0.648748582 | -0.587331 | -0.280881722 |
| 47 | 0.654684266 | -0.56807 | -0.271670379 |
| 48 | 0.652705704 | -0.586564 | -0.280514779 |
| 49 | 0.6461105 | -0.603591 | -0.288657593 |

Residuals can be simply defined as

RESIDUAL = ACTUAL VALUE - PREDICTED VALUE

Therefore, by using above residual formula we can calculate actual value too by rearranging the formula as

ACTUAL VALUE = RESIDUAL + PREDICTED VALUE.