Will V. Denzel

Joshua Bennett *Washington State University*

In this Article we will discuss who is the better artor between Will Smith and Denzel Washington.

Keywords: T-tests, Histogram, ScatterPlot, correlation tables

December 14, 2020

1   Introduction

2   Methods of finding issues in the Data

3   Research Question: What is my primary question

My primary question for this is who is objectively a better actor?

3.1   What is my secondary question

My secondary question would be what quantifies what a good actor is? Movie monetary returns? reviews?

3.2   What is my other secondary question

How did the success change over time? (this might be a good factor for determining acting skill)

4   Data Description

This is a large IMDB dataset that was harvested by my professor Monte Shaffer. It was taken from the IMBD website in September of 2020. The data is following Will Smith and Denzel Washington's careers. The movies that they have been in and the reviews and gross that the movie made.There was numerous other categories that the data has as well, such as the Titles of the movies that they were in the rating of the movie (R,PG-13, PG) a description of the movie how many votes the movie has and the year it was released. It was compiled and uploaded by my professor as well to be able to be access and the data analyze in Rstudio.

4.1   Summary of Sample

This is a summary of the Data Set that was used for this analysis[1]

```
##      ttid              nmid                rank              year
##   Length:111        Length:111        Min.   :  1.0    Min.   :1992
##   Class :character  Class :character  1st Qu.: 28.5    1st Qu.:2002
##   Mode  :character  Mode  :character  Median : 56.0    Median :2007
##                                       Mean   : 56.0    Mean   :2007
##                                       3rd Qu.: 83.5    3rd Qu.:2014
##                                       Max.   :111.0    Max.   :2021
##                                                        NA's   :35
##      title            genre              rated              minutes
##   Length:111        Length:111        Length:111          Min.   : 52.0
##   Class :character  Class :character  Class :character    1st Qu.: 93.5
##   Mode  :character  Mode  :character  Mode  :character    Median :105.0
##                                                           Mean   :106.3
##                                                           3rd Qu.:118.0
##                                                           Max.   :157.0
##                                                           NA's   :40
##      ratings           metacritic          votes            millions
##   Min.   :2.300     Min.   :15.00     Min.   :    34    Min.   :  0.02
##   1st Qu.:5.700     1st Qu.:41.50     1st Qu.: 11735    1st Qu.: 24.24
##   Median :6.300     Median :56.00     Median : 56408    Median : 56.48
##   Mean   :6.228     Mean   :51.98     Mean   :131488    Mean   : 93.80
##   3rd Qu.:6.875     3rd Qu.:62.50     3rd Qu.:206926    3rd Qu.:161.54
##   Max.   :8.600     Max.   :78.00     Max.   :675160    Max.   :355.56
##   NA's   :37        NA's   :56        NA's   :45        NA's   :59
```

summary of Wills data [2]

```
       ttid                    nmid                    rank            year
 Length:61              Length:61              Min.    : 1    Min.     :1981
 Class :character       Class :character       1st Qu.:16     1st Qu.:1994
 Mode  :character       Mode  :character       Median :31     Median :2001
                                               Mean    :31    Mean     :2002
                                               3rd Qu.:46     3rd Qu.:2011
                                               Max.    :61    Max.     :2021
                                                              NA's     :2
      title                   genre                  rated              minutes
 Length:61              Length:61              Length:61          Min.    : 60.0
 Class :character       Class :character       Class :character   1st Qu.:103.5
 Mode  :character       Mode  :character       Mode  :character   Median :118.0
                                                                  Mean     :117.0
                                                                  3rd Qu.:127.5
                                                                  Max.     :202.0
                                                                  NA's     :6
      ratings              metacritic            votes            millions
 Min.    :5.000        Min.    :30.00        Min.    :   330   Min.    :   0.19
 1st Qu.:6.500         1st Qu.:53.00         1st Qu.: 13118    1st Qu.:  21.43
 Median :6.850         Median :61.00         Median :  74561   Median :  49.42
 Mean     :6.815       Mean     :61.15       Mean    :113021   Mean     :  52.12
 3rd Qu.:7.300         3rd Qu.:71.00         3rd Qu.:183051    3rd Qu.:  78.82
 Max.     :8.500       Max.     :79.00       Max.    :383980   Max.     :130.16
 NA's     :7           NA's     :20          NA's    :11       NA's     :13
  paragraph
 Length:61
 Class :character
 Mode  :character
```

summary of Denzels data [3]

4.2   Summary Statistics of Data

Using the Data given to us by our professor from IMDB, I used the will dataframe and the Denzel dataframe to work through the data. There were multiple different methods that I used and they were mostly comparing the two dataset against each other and other factors of their own. For example, we took the will dataframe and took the data from the "metacritic" column and compared it to the years the movies came out and you can clear see a downward trend. I would also do the same for the Denzel dataframe to check for similar results.

   I also used multiple T-Tests to judge the data as well as some Histograms with a Shapiro Normality test. This was to show the mean and standard deviation between the variable that was being tested. To help show the trends over time, I implemented a Albine line that intersects the plot to show the general trajectory of the data. This shows if time goes by how do their movies change? Was there an initial burst of popularity that caused the actor to be popular and that wore off? or do their acting chops shine through giving the ability to keep a steady straight line. Maybe even improving. Running the welchs T-Test we see a p-value of 0.001401 which shows that there isn't a statistically significant difference between them, but if you look at the mean of the data you see that Will Smiths mean is 6.328 over his career and Denzels is 6.852 and the standard deviation is .913 to .654. So a strong indicator of quality. You can also look at the graphs for the differences between their "millions" category (movie earnings). From Will smith you can clearly see That he

has made movies that completely smash in the box office but still has a trend of only slightly rising. That might not even account for inflation. But he still averaged over his career around 50 million more. When comparing their movie gross there was a 95% confidence interval of 21.695 to 80.82291.

Correlation tables were also run, but we didn't get much out of it that we didn't find using other methods.

5 Key Findings

After going through this Will v Denzel dataset. We found many interesting things. Looking at the amount of money over time they make as well as their reviews over time I found to be the most interesting ways to look at the data. When comparing reviews against each other from the start of their careers you can see a much more consistent line for Denzel Washington and a mean of 6.8 for his movies with will smith having an average of 6.3 for his movies. While that doesn't appear at face value to be a significant difference in quality, the biggest thing to take into account for this is the trend over time. Will Smith has a significant downward trend whereas Denzel has only a slight downward trend. This could be due to movie critics being harder on movies nowadays, but the difference over time is significant in Denzels favor. The most telling of the findings is when looking at the Denzel movie gross vs the Will Smith movie gross over time. Will Smiths movie gross is almost a straight line across the graph from the moment he became a movie star he has averaged almost the same gross. But for Denzel his average is in a constant state of going up. Overtime his movies are getting more and more revenue, whereas Wills is staying almost the same. Running the correlation tables as well will show a .001 significance.
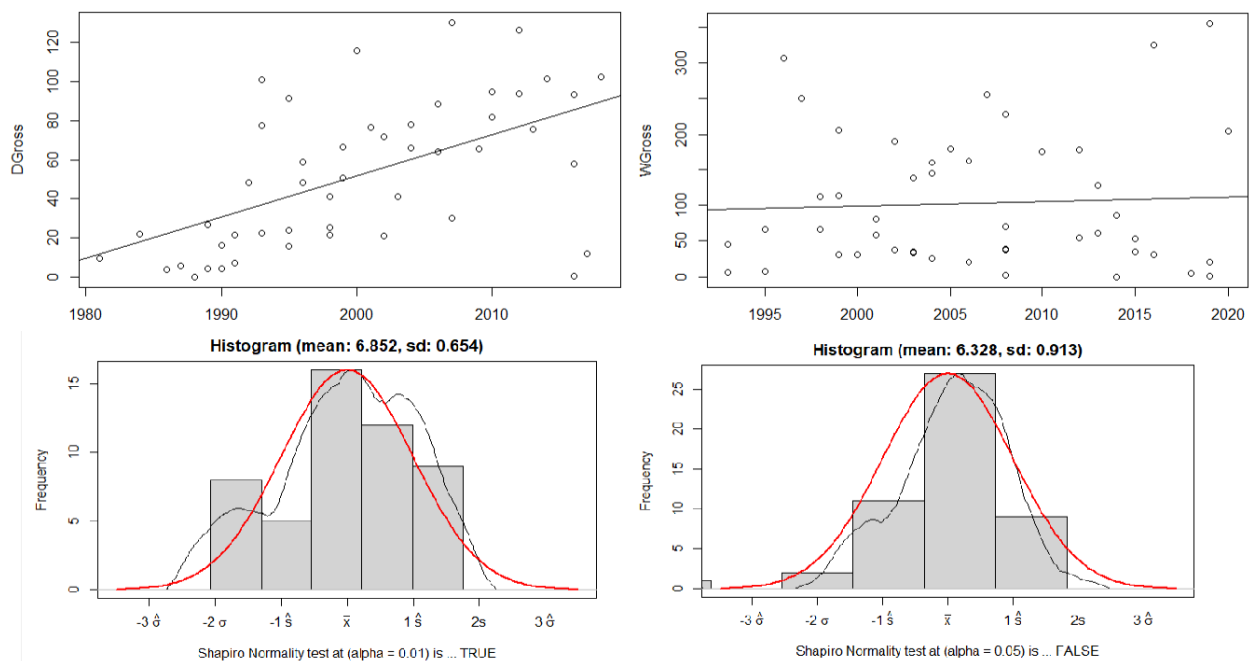
Overview of the important data. Will is on the right, Denzel on the left [4]

6 Conclusion

After going through this Will v Denzel dataset. I can see a clear winner in my mind, with one graph specifically being my main reason for my thoughts on the who is the better actor. Starting from the beginning, will Smith clearly makes more money per movie on average. I wish we had the data to go through of how much the actors were paid, but without that we can still make a solid conclusion of who is the better actor. While Will Smiths movies have made more money they haven't been consistent at all. This is similar for all the different methods statistical methods that I used to check. Denzel is consistent and Will is not. So to choose who the better actor is we need to quantify what makes an actor good. From the data I gathered, I think Denzel is clearly the better actor. His movies are constantly going up for how

much he is making, he has significantly more consistent reviews, whereas Will has some reviews that are down at around 15%, Denzels lowest is at 30% and it was an outlier. Will has 12 movies below a 40% and Denzel has only 2. The graphic that I found to be the most damning was the Shapiro Normality test, Wills failed and Denzel you can see just how stable his gross and reviews are, and how varied Will's is. Denzel is a more conistent actor which directly plays into acting ability, if he is the star of the movie and his movies constantly are rated better over 30 years. It has to be the abilities of the actor. There are arguments to be made the Will is the more successful actor, but not the BETTER actor. Going by average reviews and income overtime, Denzel Washington is the better actor.

7   Appendix

7.1   Data Provenance

7.1.1   Data Collection and organization

The beginning of the Data Provenance would be the method of collection collected. For my Data Collection personally, it was collected through our instructor. He then took that data and combined it into a much larger Data Set that we could Analyze with much more efficiency. The functions that are used to manipulate the data are in my RMD file for this that I used to look at all the data, as well as in the Functions folder on GitHub. I then organized the data by using the Will and Denzel dataframes, playing around with the different columns to see what results we could pull out. Doing this gave me results that I could compare against each other to judge the actors abilities.

ENDNOTES

[1] summary of Will and Denzels data
[2] This is an overview of the data I went through to come up with a conclusion.
[3] This is an overview of the data I went through to come up with a conclusion.
[4] Showing the differences between the two actors in the most importany metrics

TABLE OF CONTENTS