Intro to Probability & Stats

Trinity Taylor

Dr. Nicholas Jacob

January 15, 2023

Longest Running TV Shows

The data set I will be studying is data over the longest running TV shows. I found my data set from Dr. Jacob's GitHub, (https://github.com/nurfnick/Data_Sets_For_Stats). On this website, the data set can be found under (LongestRunningTVSHows.xlsx). The variables of my data set include number of seasons, network, second network, first air date, last air date, number of episodes, presently on air, and ran on multiple networks. The categorical variables include network, second network, presently on air, and ran on multiple networks. The quantitative variables include number of seasons, first air date, last air date, and number of episodes. In Table 1.1 these variables are shown.

Table 1.1:

Number of Seasons	**Series**	**Network**	**Second Network**	**First Air Date**	**Last Air Date**	**Number of Episodes**	**Presently on Air**	**Ran on Multiple Networks**
31	The Simpsons	FOX	None	12/17/89	Present	684	Yes	No
21	Law & Order: Special Victims Unit	NBC	None	9/20/99	Present	478	Yes	No
20	Gunsmoke	CBS	None	9/10/55	3/31/75	635	No	No
20	Law & Order	NBC	None	9/13/90	5/24/10	456	No	No
19	Lassie	CBS	None	9/12/54	3/21/71	591	No	No
18	Family Guy	FOX	None	1/31/99	Present	349	Yes	No

In Table 1.1 all the variables I will be studying are shown. My data set is ranked on the number of seasons the show has. I am very interested in TV show series that continue over the span of years. We live in a digital age with many different social media platforms. I feel like as the years go on TV series become less popular. In this data I want to understand what variables have contributed to these shows' success.

In conclusion I hope to find out if the network the shows appear on play a big part in how long they run. Another variable could be if they appear on multiple networks. I also noticed a trend in the first air dates. Most of the shows at the top of this ranking aired in September. Does this have anything to do with their success? I hope to find out by using statistics.

Project Part 2:

In Table 2.1 I have provided a frequency table showing the TV shows currently on air. Most of the shows I am studying are not on the air anymore. In Table 2.2 I have included another frequency table showing the amount of TV shows ran on multiple networks. Most TV shows were run on a single network. In both tables I have also included the categorical variables relative frequency. In Table 2.3 I have included a two-way table of the different networks and if their shows are still on air. Most of these longest running TV shows on these networks weren't on the air anymore.

Table 2.1

Presently on Air	Frequency	Relative Frequency
Yes	17	0.09139785
No	169	0.90860215
Total:	186	1

Table 2.2

Ran on Multiple Networks	Frequency	Relative Frequency
Yes	30	0.16129032
No	156	0.83870968
Total:	186	1

Table 2.3

Presently on Air

Network	Yes	No	Total
FOX	4	12	16
NBC	4	48	52
CBS	4	57	61
ABC	3	42	45
WB	2	6	8
CW	0	2	2
UPN	0	2	2
Total:	17	169	186

Project Part 3:

The quantitively variable I chose to display was the number of seasons. The mean of the number of seasons was 9.308108108. The standard deviation of the seasons was 3.246601767. The five number summary for the box plot (Table 3.2) is Minimum: 7, Q1: 7, Median: 8, Q3: 10, and the Maximum was 31. By using Table 3.2, I was able to identify the outliers as 15, 16, 17, 18, 19, 20, 21, and 31. The total number of outliers was 8. By looking at the histogram (Table 3.1), I can see that the data is skewed to the right. When looking at both the box plot and the histogram I see a trend of these shows only lasting 7-9 seasons. I feel like maybe the shows lose popularity after so many seasons. Another reason for less seasons could be the number of episodes in each season is a large amount.

Table 3.1

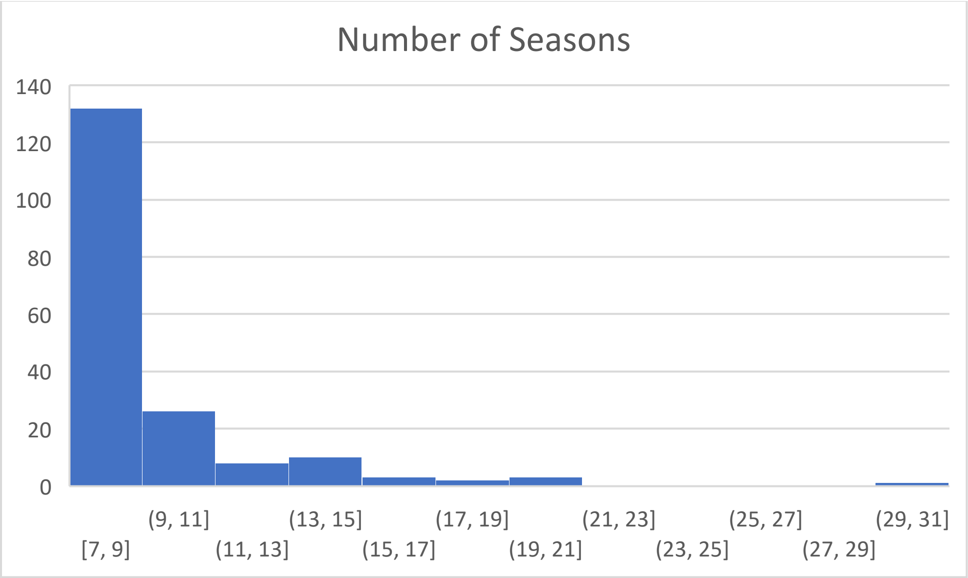
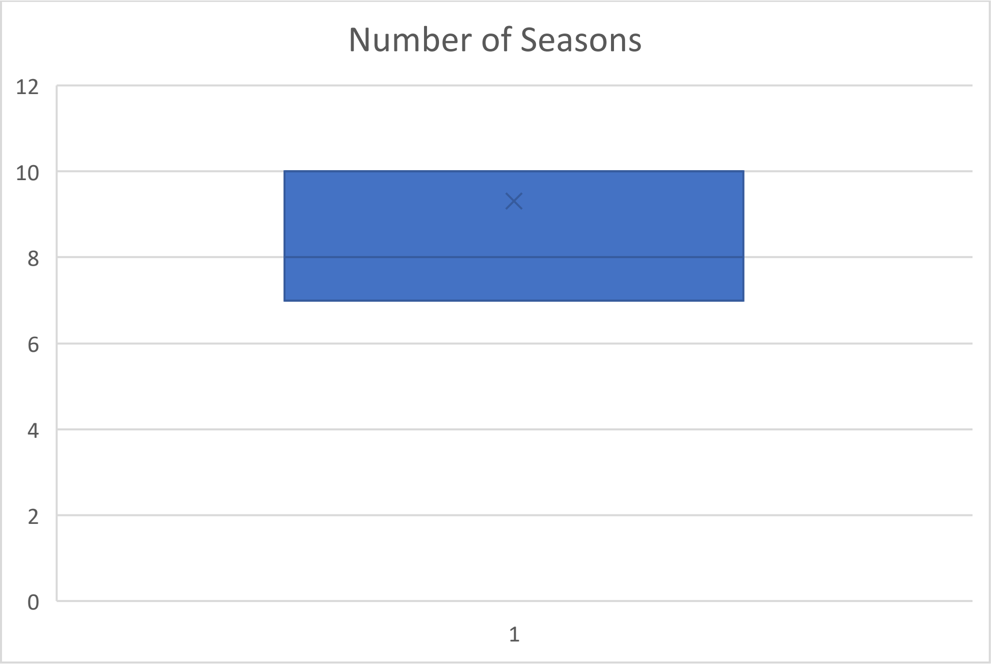


Table 3.2



Project Part 4:

For my quantitative variable I chose was the number of episodes. My null hypothesis is that the list of shows episode count averaged at about 200. I feel like the more epsiodes a show has, the more popularity the show will gain.

Quantitative Variable

Ho: $\mu = 200$

Ha: $\mu \neq 200$

For my categoical value I chose to examine if the TV show is still presently on air. Since there are so many shows on this list with success, I feel like they would still be on air to keep making profit.

Categorical Variable

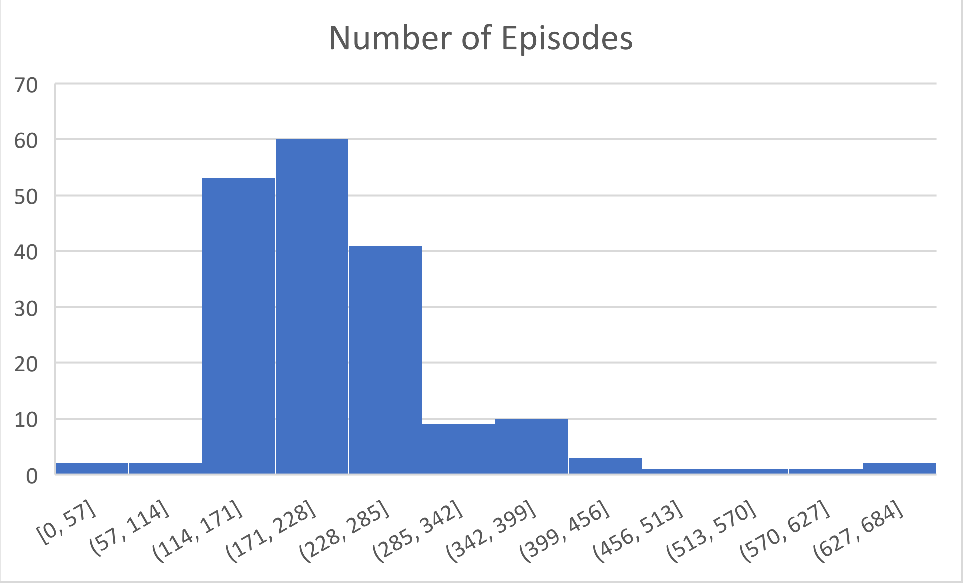
Ho: P = The amount of shows still presntly on air $p = 0.05$

Ha: P = The amount of shows presently on air $p < 0.05$

Project Part 5:

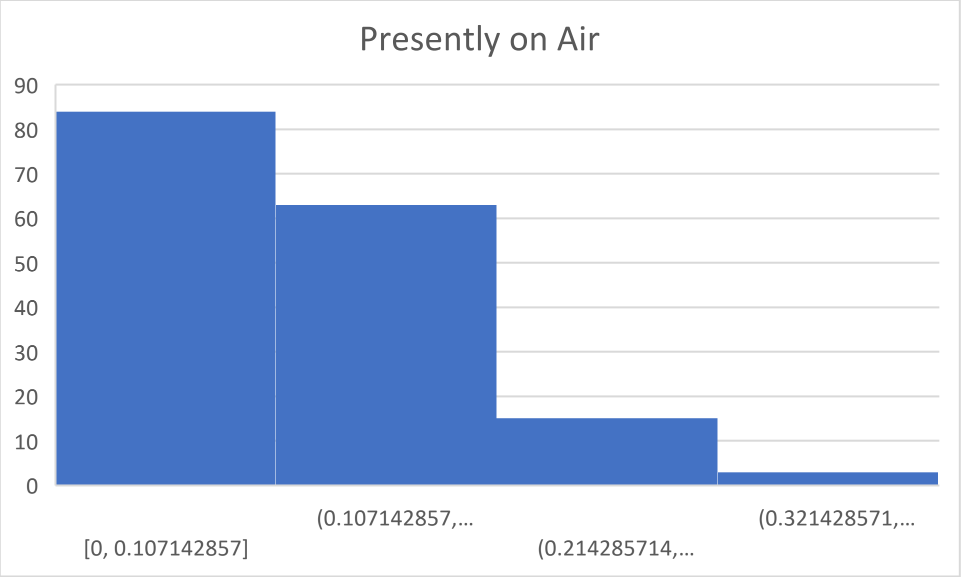
In this part of the project I performed a bootstrap so I could figure out if I reject or fail to reject the quantative and categorical hypothesis I came up with. The quantative variable I used was the number of episodes. The standard error was 6.72. The 95% confidence interval for the mean had a lower range of 212.19 and a higher range of 238.32. I found that I have rejected my null hypothesis. The histogram in Figure 5.1 that is based on my findings.

Figure 5.1



The categorical variable was p = 0.5. The standard error was 6.72. The 95% confidence interval for the mean had a lower range of 0.0502 and a higher range of 0.1335. I found that I have rejected my null hypothesis. The histogram in Figure 5.2 that is based on my findings.

Figure 5.2



Project Part 6:

I used the different formulas we learned on our homework assignments this week to configure the numbers in Table 6.1. This table is for the portion of shows that are currently on the air. These are the “yes” values. My lower confidence interval has a value of 0.06007665 and my higher confidence interval has a value of 0.12271905. My bootstrap had a lower confidence range of 0.0502 and a higher confidence range of 0.1335. I feel like the numbers for my bootstrap are close to the data I found for this part of the project. After looking at this data, I am still able to reject my null hypothesis. The formulas I used were Z= (p-hat – p), SE= SQRT(p*(1-p)/n), P=Norm.S.Dist(Z, True), and Z*= Norm.S.INV(1-(Alpha/2)).

Table 6.1

Sample	Proportion	Statistic	Significance		
n	p	P hat	alpha		
186	0.5	0.09139785	0.05		
				Cl 95%	
SE	z	z*	p	Lower	Higher
0.0159805	2.59052297	1.95996398	0.99520849	0.06007665	0.12271905
			0.00479151		

Project Part 7:

I used a t test to determine if the mean of the number of episodes does not equal 200 episodes. For this test I used 200 episodes as my hypothesis mean. To find my x bar I took the average of all the television shows episodes. My x bar is 225.26. I calculated my standard error by dividing my standard deviation by the square root of the population. My standard error is 6.667309. I found my t score by taking the x bar minus my mu. Then I divided by my standard error to find 3.788635. I found my t* by using the excel function "T.INV" to find -1.9728699. I calculated my 95% confidence interval to be between 208.086145 and 242.433855. In my bootstrap, my 95% confidence intervals were between 212.19 and 238.32. My confidence intervals were similar. Since 200 is not in these intervals I am still able to reject my null hypothesis.

mu	**x bar**	**n**	**SD**	**Alpha**
200	225.26	186	90.93	0.05
t	**SE**	**t***	**CI**	**CI**
3.788635	6.667309	-1.9728699	Lower	Upper
(x bar-mu/SE)	(SD/SQRT(n))	T.INV(alpha,n-1)	208.086145	242.433855

Project Part 8:

Using **Table 8.1** that I created for part 2 of this project, I have created two conditional probabilities. The first I will be testing if a show is randomly picked what is the probability that the show is still presently on air, which is "yes" on the table, and that the tv station is FOX. I started with the probability of a show being on FOX **(P(B)) = 16/186 = 0.08602151**. Then I had to look at the intersection of how many shows were presently on air, which there was only 4 shows. So, I calculated this by **P(A cap B) = 4/186 = 0.02150538.** Then I found P(A|B), I did

P(A|B) = A cap B/P(B) = 0.02150538/0.08602151 = 0.25000003. So, there is a **25%** chance of choosing a show that is presently on air and on the tv station FOX.

Next, I did if a show was picked randomly, what is the probability of a show to be presently on air and on the tv station NBC. I found all my answers using the same steps in the first paragraph. I found **P(B) = 52/186 = 0.27956989.** Then I did **P(A cap B) 4/186 = 0.02150538. P(A|B) = 0.7692308.** So, there is a **7.69 %** chance of picking a tv show presently on air and showing on the tv station NBC.

Table 8.1

Presently on Air

Network	**Yes**	**No**	**Total**
FOX	4	12	16
NBC	4	48	52
CBS	4	57	61
ABC	3	42	45
WB	2	6	8
CW	0	2	2
UPN	0	2	2
Total:	17	169	186