[redacted]

Into to Prob. and Stats.

Dr. Jacob

August 20, 2020

Longest Running TV Shows

The data that I found is a list of TV shows that has or is still running for a long period of time. I found this data on Wikipedia in an article called, "List of longest-running scripted U.S. primetime television series" (https://en.wikipedia.org/wiki/List_of_longest-running_scripted_U.S._primetime_television_series). I chose to do a data set about TV shows, because I am always looking for something new to watch. I like to watch shows that are complete and had a long run time, so this list will be very useful to me. Table 1.1 shows just how diverse the long running TV shows are. My variables include number of seasons, network, second network, first air date, last air date, number of episodes, presently on air, and if they ran on multiple networks. The number of seasons, number of episodes, first air date, and last air date are the quantitative variables. The network, second network, presently on air, and ran on multiple networks are the categorical variables. Table 1.2 is an explanation of the variables and restating if they are categorical or quantitative. I am interested to learn the secret of what allows a TV show to run for such a long period of time. The Simpsons are already on their 31st season and is still going strong. The secret could be the first air date or the networks the shows appeared on. There is only one way to solve this mystery, and that is with statistics.

Number of Seasons	Series	Network	Second Netowrk	First air date	Last air date	Number of Episodes	Presently on Air	Ran On Multiple Networks
31	The Simpsons	FOX		12/17/89	Present	684	yes	no
21	Law & Order: Special Victims Unit	NBC		9/20/99	Present	478	yes	no
16	Grey's Anatomy	ABC		3/27/05	Present	363	yes	no
15	Supernatural	WB	CW	9/13/05	Present	320	yes	yes
12	The Big Bang Theory	CBS		9/24/07	5/16/19	279	no	no
11	M*A*S*H	CBS		9/17/72	2/28/83	256	no	no
9	The Beverly Hillbillies	CBS		9/26/62	3/23/71	274	no	no
7	Agents of S.H.I.E.L.D.	ABC		9/24/13	8/12/20	136	no	no

Table 1.1

Number of Seasons	How many seasons were produced	Quantitative
Network	Primary network	Categorical
Second Network	Did it appear on another network	Categorical
First air date	When it first aired	Quantitative
Last air date	When it was last aired	Quantitative
Number of Episodes	Number of episodes in the series	Quantitative
Presently on Air	Is it still being produced	Categorical
Ran on multiple networks	Did it run on other networks	Categorical

Table 1.2

Project Part 2:

I have provided two frequency tables (Table 2.1 and Table 2.2) for this part of the project. Table 2.1 shows the frequencies and relative frequencies of shows that ran on multiple networks. Table 2.2 shows the frequencies and relative frequencies of shows that are presently on air. The two-way table (Table 2.3) is describing the different networks the shows appeared on and if those shows are presently on air. The table reveals that ABC, CBS, and NBC produced more of the longer running TV shows. This is not surprising, considering that these three networks are some of the older networks. What surprised me was that WB (Warner Brothers), which is the oldest network on this list, only produced seven shows that ran for a long time.

Ran on multiple networks	Frequency	Relative Frequency
yes	30	0.162162162
no	155	0.837837838
Total	185	1

Table 2.1

Presently on air	Frequency	Relative Frequency
yes	17	0.091891892
no	168	0.908108108
Total	185	1

Table 2.2

Network	Presently on air		
	yes	no	Total
ABC	3	41	44
CBS	4	57	61
CW	0	2	2
FOX	5	12	17
NBC	4	48	52
UPN	0	2	2
WB	1	6	7
Total	17	168	185

Table 2.3

Project Part 3:

The tables below show statistical charts showing the number of TV shows of the longest running series. The mean of the seasons was 9.31. The standard deviation was 3.24660177. The five-number summary for the box and whisker chart (Table 3.1) was: Minimum = 7, Q1 = 7, Median = 8, Q3 = 10 and the Maximum = 31. The box and whisker plot revealed that there were eight outliers with values of 31, 21, 20, 19, 18, 17, 16, and 15. The histogram (Table 3.2) is skewed to the right, because only a few shows were able to go for more than 9 seasons. There are fewer shows to be aired longer due to the audience interest in the shows. It seems that after 7-9 seasons, the audience loses interest in the show which leads the show to be cancelled. Since the majority of the TV shows were cancelled after the 7^{th} -9^{th} season, it caused the histogram to be skewed right and to display outliers in the box plot.

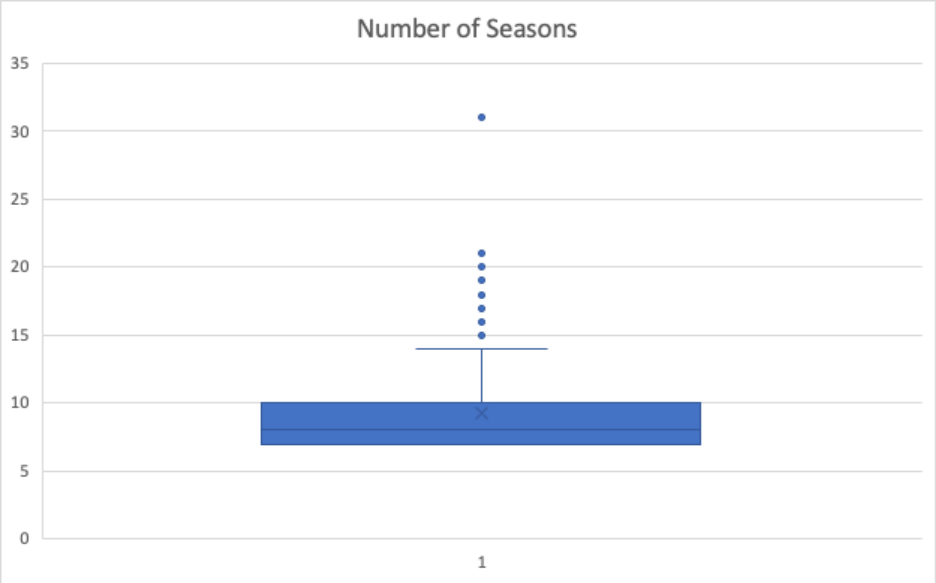

Table 3.1

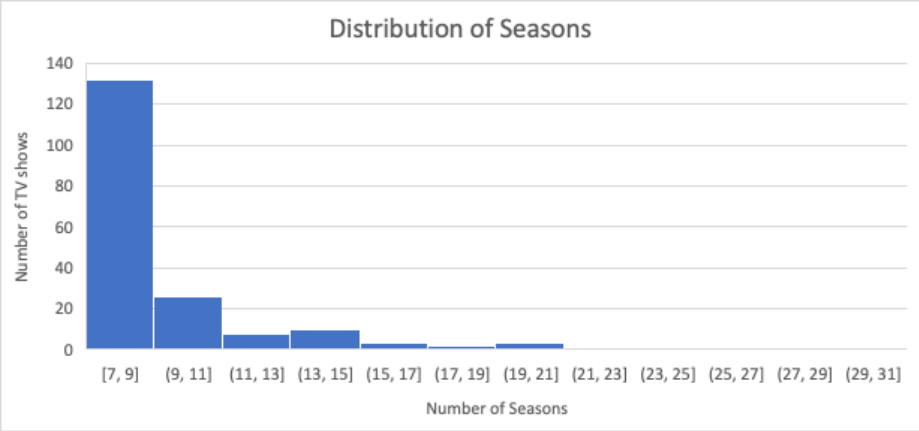

Table 3.2

Project Part 4:

I wanted to look at the number of episodes for the quantitative hypothesis.
H_o: μ=300
H_a: $\mu \neq 300$
I want to know the average number of episodes each series produced while they were running. By calculating the average number of episodes per show, I can calculate the average number of episodes per season.

For the categorical variable, I chose to look into which shows are presently on air.
H_o: The proportion of shows still on air is p= 0.05
H_a: The proportion of shows still on air is p<0.05
There are a lot of the series still running on TV, so I wanted to see if the number of episodes prolonged their production.

Project Part 5:
I performed a bootstrap to find out if I could reject or fail to reject my quantitative hypothesis in part 4. The standard error was 7.04517. The 95% confidence interval for the mean had a lower range of 210.770631 and a higher range of 238.95132. I have included the histogram of the bootstrap distribution of the means below (Figure 5.1). Based on the confidence interval, I can reject my null hypothesis, because the null estimated mean was outside of the confidence interval range.

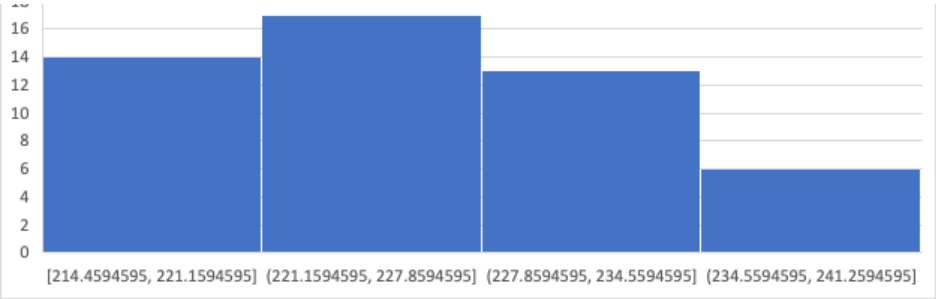

Figure 5.1

I also performed a bootstrap to test my categorical hypothesis in part 4. The standard error is 0.023032. The 95% confidence interval had a lower range of 0.0498278 and a higher range of 0.14195598. I have also included the histogram of the bootstrap distribution of the proportion of shows that are still on the air below (Figure 5.2). Based on this data, I can fail to reject my null hypothesis, because the estimated p value is included in the range.

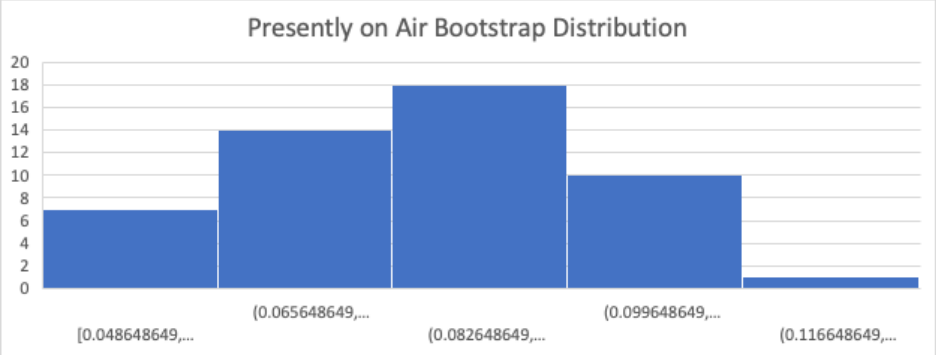

Figure 5.2

Project Part 6:

Using the formulas we practiced with for the homework, I was able to form a table below (Table 6.1) for the proportion of shows that are still on the air, the "yes" values. I will be able to use -1.644854 as the beginning of the lower rejection zone, because the z* score is symmetrical on both sides of p and since I am using less than in my alternative hypothesis. I was not able to reject my null hypothesis due to the z value I created not being below the critical z value. The 95% confidence interval I calculated had a lower range of 0.04941512 and an upper range of 0.13436866. Based on this data, I am able to fail to reject my null hypothesis. My bootstrap confidence interval had a lower range of 0.0498278 and an upper range of 0.14195598. The 95% confidence interval that I computed for this part of the project has a smaller range than the bootstrap 95% confidence interval. Using these formulas may provide more accurate confidence intervals of shows that are still on the air than the bootstrap distribution.

Sample	Proportion	Statistic	Significance		
n	p	p hat	alpha		
185	0.05	0.09189189	0.05		
				CI 95%	
SE	z	z*	p	Lower	Higher
0.01602363	2.61438213	1.644854	0.99553055	0.04941512	0.13436866
			0.00446945		

Table 6.1

Project Part 7:

For Part 7 of the project, I used a t test to determine if the mean number of episodes is not equal to 300 episodes. I used 300 episodes as my hypothesis mean and found x bar by taking the average of my samples. I calculated the standard error, t score, and t* by using the formulas listed in the table (Table 7.1). I calculated the 95% confidence interval using t*. The confidence interval was between 213.673991 and 235.709793. My Bootstrapping confidence intervals were

210.770631 and 238.95132. These confidence intervals are very similar, and by using both, I am able to reject my null hypothesis because 300 is not in the range of the confidence intervals.

mu	x bar	n	SD	Alpha
300	225	185	90.6519694	0.05
t	SE	t*	CI	
-11.299269	6.66486534	-1.653131869	Lower	Upper
(x bar-mu/SE)	(SD/SQRT(n))	T.INV(alpha,n-1)	235.709793	213.673991

Table 7.1

Project Part 8:

Using Table 2.3, I created two conditional probabilities. The first one that I tested was, if one TV station was chosen, what is the probability that the show is still on air (yes), given that the TV station is ABC? I first looked at the probability of getting a show in ABC **(P(B)) = 44/185 = 0.23784**. Next I looked at the intersection for how many shows were still on air and are broadcasted on ABC. There were 3 shows that met the criteria, so to calculate **P (A cap B) = 3/185 = 0.01622**.
To find P(A|B), I used **P(A|B) = A cap B/P(B) = 0.01622/0.23784 = 0.0681812**. There is a **6.81%** chance of choosing a show that is still on air given the TV station ABC.

The second conditional probability that I tested was if a show was randomly picked, what is the probability that the project will still be on air, given that the TV station is CBS. Following the same steps previously listed, I found **P(B) to be 61/185 = 0.32973**.
P (A cap B) to be 3/185 = 0.01622. **P(A|B) = 0.049192**. There is a **4.92%** chance of choosing a show that is still on air given the TV station is CBS.

Network	Presently on air		
	yes	no	Total
ABC	3	41	44
CBS	4	57	61
CW	0	2	2
FOX	5	12	17
NBC	4	48	52
UPN	0	2	2
WB	1	6	7
Total	17	168	185

Table 2.3