[redacted]

Intro to Prob. And Stats.

Dr. Nick Jacob

January 17, 2021

Longest Running TV Shows

The collected data for the longest running TV shows was found on Dr. Jacob's GitHub, Longest Running TV Show Data. The data is formatted in an excel sheet with 186 different TV shows with variables including: number of seasons, networks, first airdate, last airdate, number of episodes, presently airing, and whether or not it ran on multiple networks. Table 1.1 displays a small sample taken from the 186 TV shows to demonstrate their diversity. In table 1.2, variables are further defined by categorical or quantitative. My interest in this data set derives from my love of television entertainment. It is a way for me to shut off my brain and decompress after a long shift at the hospital and allows for me to disconnect from the day. I am interested to see if there is any correlation to be made on success of the show (defined by number of seasons and air time) and which network it airs on. I know personally I tend to watch shows from specific networks and streaming providers. A possible lurking variable could be the show producers which is not demonstrated in this data collection.

Show	# of seasons	Network	2nd Network	1stAir Date	Last Air Date	# of Episodes	Presently Running	Ran on Multiple Networks
The Simpsons	31	FOX	-	12/17/1989	Present	684	Yes	No
Grey's Anatomy	16	ABC	-	03/27/2005	Present	363	Yes	No
ER	15	NBC	-	09/19/1994	04/02/2009	331	No	No
The Big Bang Theory	12	CBS	-	09/24/2007	05/16/2019	279	No	No
7th Heaven	11	WB	CW	08/26/1996	05/13/2007	243	No	Yes
How I Met Your Mother	9	CBS	-	09/19/2005	03/31/2014	208	No	No
One Tree Hill	9	WB	CW	09/23/2003	04/04/2012	187	No	Yes

Table 1.1 Sample Taken From Longest Running TV Shows Data

Variable	Type of Data
Number of Seasons	Quantitative
Network	Categorical
2nd Network	Categorical
1st Airdate	Quantitative
Last Airdate	Quantitative
Number of Episodes	Quantitative
Presently Running	Categorical
Ran on Multiple Networks	Categorical

Table 1.2 Description of Variables by Type of Data

Part 2:

In order to correlate TV show success to their network and provide statistical evidence, I took the networks and divided them into currently running and no longer running shows in Table 2.1 with each networks total in the data collection of 185 shows with 7 or more seasons. This demonstrates that only 10% are presently airing with the top 4 networks represented are ABC,

NBC, CBS, and FOX. Table 2.2 looks more closely at the number of shows that appear on multiple networks, frequency, and the relative frequency. This revealed a higher frequency of shows that remain on the network is first airs on. Table 2.3 establishes the cumulative relative frequency of the 185 longest running TV shows and if they are still presently running. The results are 90% of the 185 longest running television shows are no longer airing. Could this indicate that the best of TV is now behind us or is the best yet to come?

TV Network	Presently Running Yes	Presently Running No	Total
ABC	3	41	44
NBC	4	48	52
CBS	4	57	61
FOX	5	12	17
CW	1	2	3
WB	0	6	6
UPN	0	2	2
Total	17	168	185

Table 2.1

Multiple Networks	Frequency	Relative Frequency
Yes	30	0.1621621622
No	155	0.8378378378
Total	185	1

Table 2.2

Presently on Air	Frequency	Relative Frequency
Yes	17	0.0918918919
No	168	0.9081081081
Total	185	1

Table 2.3

Part 3:

The next part of statistical data to observe is the number of seasons per TV show represented in the data set. There are 185 TV shows with the minimum seasons being 7 and the maximum being 31. The mean number of seasons is 9.3 and the median is 8 with a standard deviation of 3.23. A mean greater than the median produces a right skewed distribution as shown in the histogram in Table 3.2. The 5 number summary represented by the box plot in Table 3.1 also includes the Q1=7 and Q3=10. Summarizing the data provided it appears that the vast majority of the longest running television shows run 7-9 seasons and then go off the air. This could be due to loss of interest in the audience, difficulty retaining the shows actors/actresses, or cost of show after so many seasons. Further research would need to be done to look into the reasons provided for the show cancelation to form correlations to number of seasons.

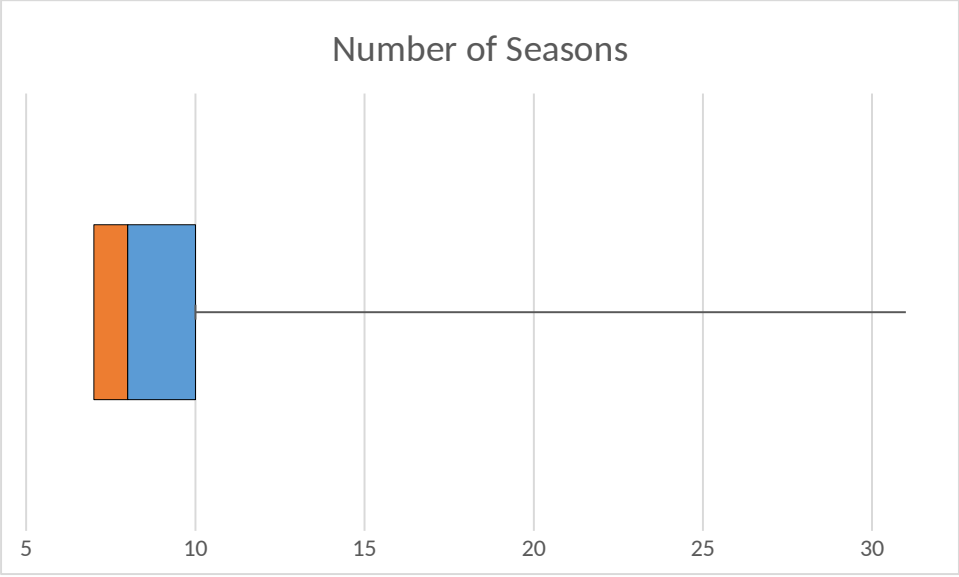

Table 3.1

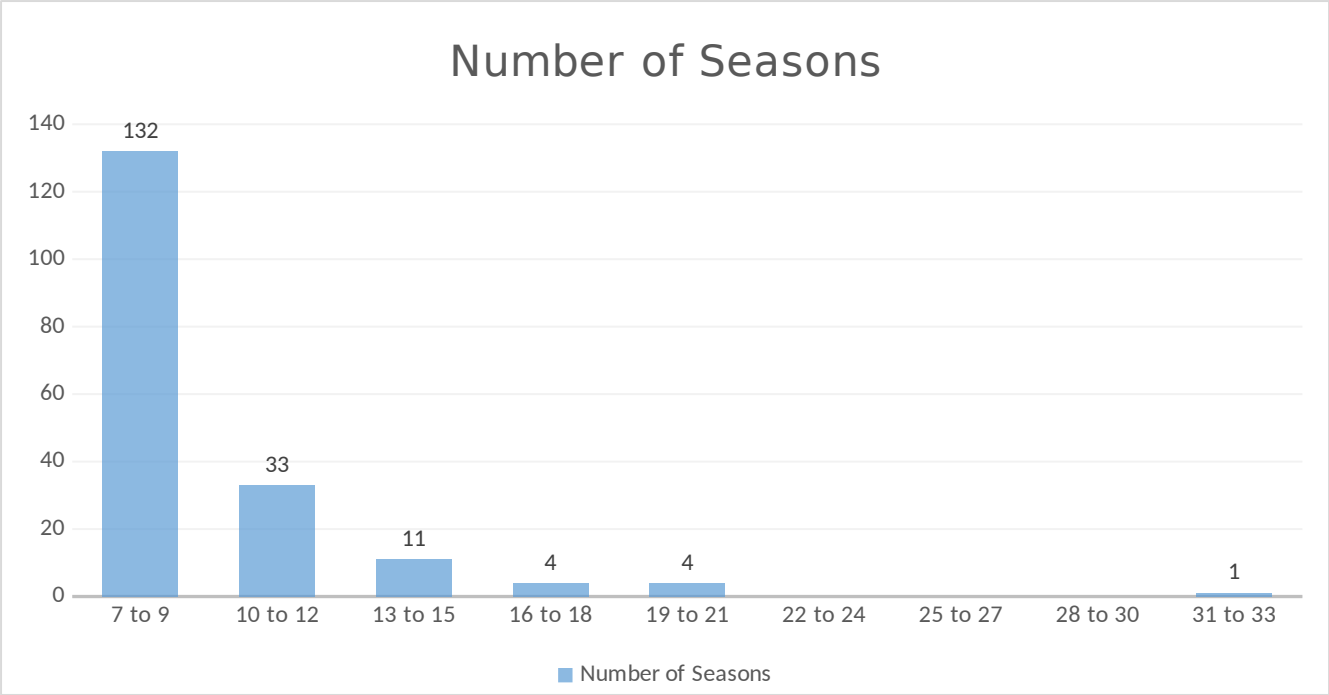

Table 3.2

Part 4:

In order to produce a quantitative hypothesis, the next variable I wanted to consider is the average number of episodes produced in total by each series listed on the longest running TV shows data set. The average number of episodes can be calculated by adding the total number of episodes divided by the number of series in the data set. I can also calculate for the average number of episodes per season if indicated. My hypothesis is that the total average number of episodes per series will be 250. I estimated this number from considering my mean number of seasons to be 9.3 and predict that most series have around 25-28 episodes each season.

<u>My null and alternative hypothesis for the quantitative data (total average number of episodes) will be:</u>

H_o: $\mu=250$

H_a: $\mu \neq 250$

A categorical hypothesis can be formed from considering which TV shows are still currently airing. Proportions of shows still on air running:

H_o: $p=0.05$

H_a: $p<0.05$

A question I draw from gathering my information to test is, does the number of episodes provide value to prolonging production and staying on air? My hypothesis of 0.05 is equivalent to 5% still currently airing. I based this off of glancing at the excel data sheet and reading the shows listed.

Part 5:

For my quantitative hypothesis, the average number of episodes for each series resulted in 224 which supports my alternate hypothesis and rejects my null of 250. The average of the bootstrap sample is 224.66 with a standard error (SE) of 6.167056. The bootstrap distribution provides a bell curved histogram with the center representing the median, which is shown in Table 5.1. The calculated 95% confidence interval is 212.6822 – 236.6265.

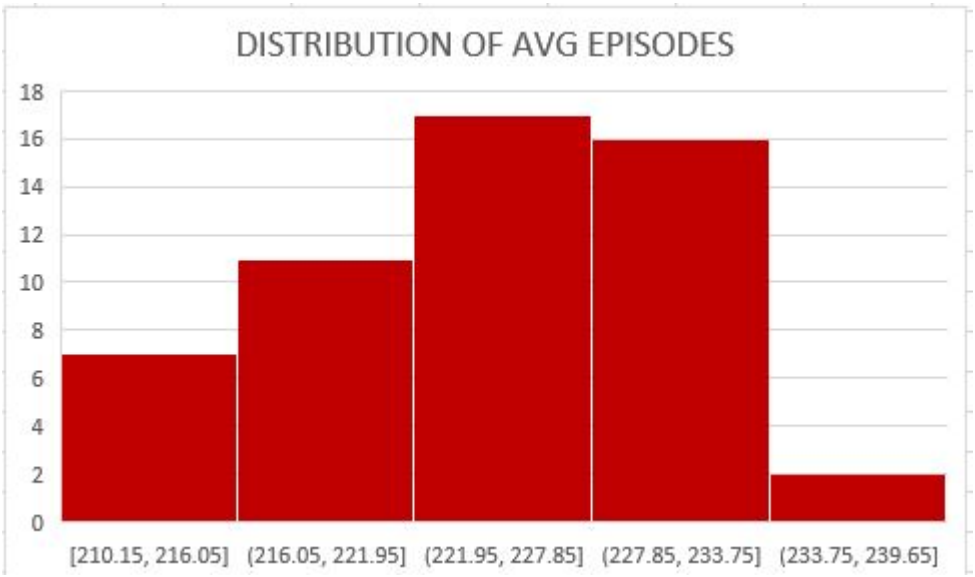

Table 5.1

For my categorical hypothesis of the proportion =0.05 (5%) is rejected by my calculated mean of 0.091892 (9.2%) of the longest running TV shows still airing. The mean of the bootstrap is 0.09106 (9.1%) which is represented as the center of the bell curve in the histogram in Table 5.2. The standard error of the distribution sample is 0.025272. The 95% confidence interval was found to be 0.040517-0.141604 (4.1%- 14.1%). The confidence interval however, supports the null hypothesis of 0.05 (5%).

Table 5.2

Part 6:

In attempt to further evaluate my statistical findings, I will now use the formulas that include the use of z and critical z values in relation to proportion and statistic value versus the bootstrap distribution for confidence intervals in part 5. I will be retesting my categorical hypothesis of the proportion of TV shows listed on the longest running data collection that are

still presently on air and then comparing my findings. My findings are presented in Table 6.1 and my comparison is listed below:

Bootstrap 95% CI= 0.040517-0.141604 (4.1%- 14.1%). SE= 0.025272
Categorical Inference 95% CI= 0.065535472-0.1182458528 (6% - 11%) SE= 0.016023631

The confidence intervals are slightly different but my new calculated confidence interval may provide a higher accuracy due to small standard error. My 0.05 (5%) null hypothesis would be rejected in this calculation whereas I failed to reject my null hypothesis in my bootstrap confidence interval. The difference in intervals would seem small but it is significant due to rejection.

Sample (n)	Poportion (p)	Statistic (p hat)	Alpha α
185	0.05	0.091892	0.05
SE	**Z**	**Z***	**p value**
0.016023631	2.614388679	1.644853627	0.995530636
			0.004469364
CI 95%			
Lower	0.065535472		
Upper	0.118248528		

Table 6.1

Part 7:

In order to further test my quantitative hypothesis of an episode mean of 250, I recalculated my values utilizing the formulas for a t test. My xbar was found by averaging the number of episodes by the number of shows (224.75). Table 7.1 shows the values I found using the quantitative inference formulas of a t test. My 95% confidence interval for this test came to (237.88-211.60). In table 7.2, the comparison of confidence intervals between the bootstrap sample and the t test. Even though the upper limit of m confidence intervals are slightly different, I am still able to reject my null hypothesis of 250 episode mean.

X BAR	N	MU	SD STDEV.S(A2:A186)	SE SD/SQRT(n)
224.75	185	250	90.61	6.66
t TEST STAT (XBAR-MU)/ SE	**t * CRITICAL** T.INV(ALPHA, n-1)	**ALPHA**	**CONFIDENCE INTERVAL**	**95%** XBAR ± t* (SE)
-3.79	1.97	0.05	**UPPER**	237.8886556
			LOWER	211.6032363

Table 7.1

Formulas for 95% Confidence Intervals	Bootstrap	T test
UPPER	239.4586213	237.8886556
LOWER	211.0584265	211.6032363

Table 7.2

Part 8:

In order to find two conditional probabilities from my data on the longest running TV series, I made Table 8.1 to show how many shows are still currently running and separated them by the network they originated from. The first probability I tested was what is the probability of choosing a show is still running given that it is on ABC network? The probability of choosing a TV show that was/is on ABC is found by **P(A) 44/185= 0.2378378378.** Next you need to know how many series are still currently running on ABC network. The formula to find this is **P(B) 3/185=0.162162162.** The probability of choosing a series still presently on air given the network being ABC is found by [redacted]**= 0.162162162/0.2378378378 = 0.681818 or 6.8%** chance of selecting an ABC network show that is currently still on air.

The second conditional probability I wanted to test is what is the probability of selecting a TV series that is still currently on air given that it is on the FOX network? [redacted] There is a 9.6% probability of selecting a show that is still currently on air given that the Network is CBS.

NETWORK	Presently on Air		TOTAL
	YES	NO	
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 8.1