

Alone TV Show Data Regression

Elliot Hodge and Vella Karman
Central Arkansas Homeschool Academy
2021-2022 Data Science Class

December 31, 2021

Abstract — In order to determine the correlation between the number of days contestants remained participating on the wilderness survival TV show ‘Alone’ and various other factors, we analyzed which tools from a set list of tools were brought and the average temperatures of the contest locations for each existing season of the show as of December 2021, excluding Season 4, which was treated as an outlier. The tools that were common to all participants were excluded from the analysis.

to survival on Alone. The analysis included the number of days survived on Alone and the tools a competitor selected. The correlation between average hometown weather and average weather in the location of each season of Alone in September was also analyzed.

This paper is structured as follows: Section 2 covers the experimental set-up and methodology, including why Season 4 was excluded. Section 3 presents the results of the analysis. Section 4 concludes the paper and shows the fitted line plot for the data. Section 5 includes the dataset before translation and the sources.

1 Introduction

The Alone TV show films competitors surviving the wilderness with a basic survival pack and ten items they choose from a list of approved gear. The regression analysis in the field of Data Science determined the correlation of average temperatures and the gear they selected to the number of days contestants survived the show. Data regression in Minitab showed the calculations for significance, correlation, and randomness, as well as a formula for the line of best fit. The researchers also used a set of random data to test the formula. The results are important for showing student comprehension of data regression and foundational concepts of data science and for potential Alone competitors.

The objective of this paper is to prove or disprove the hypothesis that a particular combination of tools will provide a measurable benefit

2 Experimental Setup and Methodology

2.1 Spreadsheet Creation

1. Using the list of gear each contestant took [1], a spreadsheet was created listing contestant first name, contestant last name.
2. Using the aforementioned list [1], Season 4 contestants’ teammates’ first and last names were added to the spreadsheet. Other Seasons’ contestants, who did not work in teams, had ‘null’ listed for teammates’ first and last names.
3. Using the aforementioned list [1], the season each contestant participated in was added to the spreadsheet.

4. Using the Wikipedia article on the Alone TV show [2], the days each contestant participated in the competition were added to the spreadsheet. Participants left because they opted out, the staff of Alone chose to medically evacuate them, or they won. In most seasons, the last person participating won, but in Season 7 any contestant who lasted 100 or more days won.
5. Using the Wikipedia article on the Alone TV show [2], whether the contestant was medically evacuated or not was added to the spreadsheet as a string value of Yes or No.
6. Using the Wikipedia article on the Alone TV show [2], the hometown for each contestant was taken, and a google search for average temperature was performed. The average temperature was added to the spreadsheet. If the average temperature was listed in an array of months, the month of September was used.
7. Using the Wikipedia article on the Alone TV show [2], the location for each season was taken, and a google search for the average temperature was performed. The average temperature was added to the spreadsheet. If the average temperature was listed in an array of months, the month of September was used.
8. For each allowed item on the list of allowed items [1], a column labeled with the item's name was added to the spreadsheet. For each contestant or contestant pair, a string value of Yes was listed under the columns of the items each contestant selected, a string value of No was listed under the columns of the items the contestants did not select. If

the contestant chose to take an item twice, Double was listed instead of Yes.

The completed data set was uploaded to github [3].

2.2 Minitab Setup

Because Season 4 had two contestants working as a team, all rows from season four were removed as outliers. Because Season 4 was the only season in which multiple contestants were on a team together, the second contestant's first and last name columns were removed as they all contained the value 'null'.

Every item which no one brought, everyone brought, or only contestants from Season 4 brought was removed from the data set. Because of this, the Hatchet, Beef Jerky, Dried Snack Mix, Biltong, Hardtack Biscuits, Chocolate, Pemmican, Trail Mix, Flour, Salt, Rice/Sugar/Salt, Bear Canister, Bowl, Spoon, Lighter, Sleeping Pad, Sleeping Bag, Net Foraging Bag, Bar of Soap, Comb, Towel, Toothpaste, Floss, Shower Soap, and Razor columns were removed.

The data was copied and pasted into MiniTab. Using the Fit Regression Model found in 'Stat > Regression > Regression > Fit Regression Model,' all items brought were added to the independent factors list. The average temperature in each contestant's hometown and the average temperature in the contest location were added as independent variables as well. The days each contestant participated were added to the dependent variables list.

Using the equation generated, a prediction test was run, listing a number of items, an average home temperature, and an average contest location temperature.

Using the same Fit Regression Model in a second test, this time only the average temperature data being added as independent variables, the days each contestant participated continued to be used as the dependent variable.

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	212.7	67.9	3.13	0.003	
Average Temperature of Hometown	-1.393	0.874	-1.59	0.118	1.44
Average Temperature of Show (ba	-0.688	0.791	-0.87	0.389	1.84
Knife					
No	51.2	61.2	0.84	0.408	72.32
Yes	20.6	46.4	0.44	0.659	42.54
Multi-Tool					
Yes	-28.5	25.7	-1.11	0.274	14.48
Sharpening Stone					
Yes	-34.1	26.6	-1.28	0.206	2.60
Shovel					
Yes	-25.1	40.3	-0.62	0.537	4.03
Axe					
Yes	-43.7	24.6	-1.78	0.083	7.18
Saw					
Yes	-29.2	19.8	-1.47	0.148	5.00
Auger, Scotch Eyed					
Yes	-49.3	39.3	-1.26	0.216	1.95
Emergency Rations					
No	55.2	45.1	1.23	0.227	44.57
Yes	24.8	25.9	0.96	0.343	14.83
Pot					
Yes	-14.0	17.4	-0.80	0.426	1.46
Flint and Steel/Ferro Rod					
Yes	-38.6	32.9	-1.18	0.246	3.98
Canteen					
Yes	-33.2	28.5	-1.16	0.251	7.38
Bivvy Bag					
Yes	-15.4	25.9	-0.60	0.554	3.98
Hammock					
Yes	-0.3	37.5	-0.01	0.993	3.49
Paracord					
Yes	-18.3	22.3	-0.82	0.415	10.38
Tarp					
No	25.9	51.2	0.51	0.616	28.77
Yes	-10.2	36.5	-0.28	0.781	13.35
Rope					
Yes	-70.7	54.2	-1.30	0.199	3.71
Fishing Line + Hooks					
Yes	-58.3	30.7	-1.90	0.064	3.47
Wood Bow and Arrows					
Yes	-23.7	22.1	-1.08	0.288	10.19
Trapping Wire					
Yes	-18.3	23.5	-0.78	0.441	12.37
Gauge Gill Net					
Yes	-11.8	22.2	-0.53	0.598	9.50

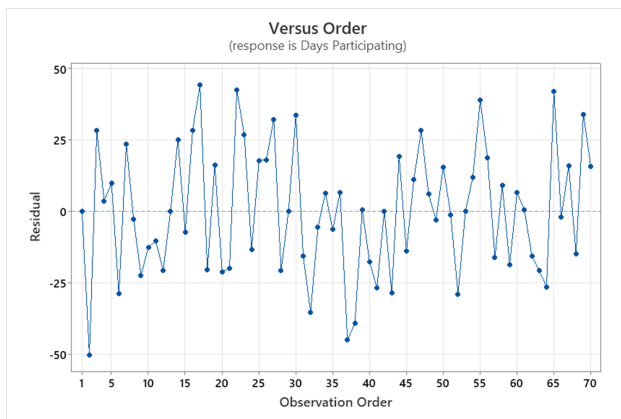
3 Results

With a 95% confidence level, the analysis gave the results below for the correlation of tools chosen and tools not chosen with the days each contestant participated:

The p-value, showing statistical significance, ranged from 0.064 to 0.993 of the coefficient calculated for each item used as a variable. P-values above 0.5 are considered strong. The

items which met that requirement and whether it was for or against bringing it were: Knife (Yes), Shovel (Yes), Bivvy Bag (Yes), Hammock (Yes), Tarp (Yes and No), and Gauge Gill Net (Yes). Only the Hammock shows a very strong (above 0.81) significance.

The coefficient of determination (r^2) was 35.42% showing a low percentage of the data fits the line of best fit. This indicated a weak relationship. The residuals plot (below) had a random scatter of data, indicating random residuals.



Regression Equation

The coefficient of determination demonstrated a weak relationship between the line of best fit and the data. While the p-values and coefficient of determination for the variables of the regression equation prove it to be inaccurate, a regression equation was still given. Due to the significance calculations and the relationship of the data to the line of best fit, we know this equation (below) cannot accurately predict results.

Regression Equation

```
Days Participating = 212.7 - 1.393 Average Temperature of Hometown
- 0.688 Average Temperature of Show (ba + 0.0 Knife_Double
+ 51.2 Knife_No + 20.6 Knife_Yes + 0.0 Multi-Tool_No
- 28.5 Multi-Tool_Yes + 0.0 Sharpening Stone_No
- 34.1 Sharpening Stone_Yes + 0.0 Shovel_No - 25.1 Shovel_Yes
+ 0.0 Axe_No - 43.7 Axe_Yes + 0.0 Saw_No - 29.2 Saw_Yes
+ 0.0 Auger, Scotch Eyed_No - 49.3 Auger, Scotch Eyed_Yes
+ 0.0 Emergency Rations_Double + 55.2 Emergency Rations_No
+ 24.8 Emergency Rations_Yes + 0.0 Pot_No - 14.0 Pot_Yes
+ 0.0 Flint and Steel/Ferro Rod_No - 38.6 Flint and Steel/Ferro Rod_Yes
+ 0.0 Canteen_No - 33.2 Canteen_Yes + 0.0 Bivvy Bag_No
- 15.4 Bivvy Bag_Yes + 0.0 Hammock_No - 0.3 Hammock_Yes
+ 0.0 Paracord_No - 18.3 Paracord_Yes + 0.0 Tarp_Double + 25.9 Tarp_No
- 10.2 Tarp_Yes + 0.0 Rope_No - 70.7 Rope_Yes
+ 0.0 Fishing Line + Hooks_No - 58.3 Fishing Line + Hooks_Yes
+ 0.0 Wood Bow and Arrows_No - 23.7 Wood Bow and Arrows_Yes
+ 0.0 Trapping Wire_No - 18.3 Trapping Wire_Yes + 0.0 Gauge Gill Net_No
- 11.8 Gauge Gill Net_Yes
```

Example Prediction

The researchers entered random items into MiniTab to predict results using the regression equation. The settings are shown below.

Settings

Variable	Setting
Average Temperature of Hometown	5
Average Temperature of Show (ba	9
Knife	Yes
Multi-Tool	No
Sharpening Stone	No
Shovel	Yes
Axe	Yes
Saw	Yes
Auger, Scotch Eyed	No
Emergency Rations	No
Pot	No
Flint and Steel/Ferro Rod	No
Canteen	No
Bivvy Bag	No
Hammock	Yes
Paracord	Yes
Tarp	Yes
Rope	Yes
Fishing Line + Hooks	Yes
Wood Bow and Arrows	No
Trapping Wire	No
Gauge Gill Net	No

Minitab displayed a warning message for the PI point predicted, “XX denotes an extremely unusual point relative to predictor levels used to fit the model.” The predictions are below.

Prediction

Fit	SE Fit	95% CI	95% PI
19.6682	60.7365	(-102.738, 142.075)	(-115.079, 154.415) XX

XX denotes an extremely unusual point relative to predictor levels used to fit the model.

This prediction shows an unbelievably large range for the CI (confidence interval); from a negative number in the hundreds to a positive number in the hundreds. This indicated the margin of error is too high to be trusted. The PI value predicted demonstrates the inputs could be an outlier. Based on these two numbers, the fit (19.6682) cannot be trusted as an accurate prediction.

Multiple Response Prediction

The researchers set the computer to predict the inputs (temperature and gear) for the given output of the longest possible days based on the regression equation and the data available. The analysis is below.

Multiple Response Prediction

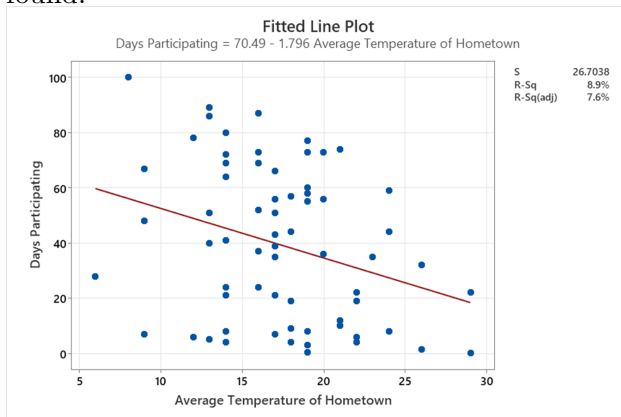
Variable	Setting
Average Temperature of Hometown	6
Average Temperature of Show (ba	3
Knife	No
Multi-Tool	No
Sharpening Stone	No
Shovel	No
Axe	No
Saw	No
Auger, Scotch Eyed	No
Emergency Rations	No
Pot	No
Flint and Steel/Ferro Rod	No
Canteen	No
Bivvy Bag	No
Hammock	No
Paracord	No
Tarp	No
Rope	No
Fishing Line + Hooks	No
Wood Bow and Arrows	No
Trapping Wire	No
Gauge Gill Net	No

Response	Fit	SE Fit	95% CI	95% PI
Days Participating	335	174	(-16, 685)	(-21, 690)

According to this prediction, the best way to succeed would be to bring no items at all.

Temperature

The other factor analyzed was the average hometown temperature in the same month as the season was filmed compared to the average temperature at the location of filming. The fitted line plot for this analysis (below) has an r^2 value of 8.9% which indicates the line of best fit explains that percentage of the data. No correlation was found.



The equation given was Days Participating + 70.49 - 1.796 Average Temperature of Hometown. The data indicates this equation can not accurately predict results either.

4 Conclusion

Based on the data analysis, the hypothesis that a particular combination of tools will provide a measurable benefit to survival on Alone must be rejected.

The analysis showed no correlation between tools a competitor decided to bring and the number of days they lasted. It is possible the skill set of the competitor is the defining factor, but more research would have to be done to reach a

conclusion.

The analysis showed no correlation between the average hometown temperature in September and the average temperature where the season was shot.

There is no relevant further research to do on this topic.

5 References

- [1] Bushcraft Info. Alone Prohibited Items & Gear List. <https://bushcraftinfo.com/alone-prohibited-items-gear-list/>.
- [2] Wikipedia. Alone (TV series). Online encyclopedia: [https://en.wikipedia.org/wiki/Alone_\(TV_series\)](https://en.wikipedia.org/wiki/Alone_(TV_series)), 2015. revision 16:48, December 17th 2021.
- [3] Github. Alone Dataset. <https://github.com/Craftidore/Assignments/blob/main/2021-22/Data%20Science/Alone%20Dataset.csv>.