100 Oldest Women Ever

Kimberly Miller

Statistics & Probability

Instructor: Nicholas Jacob, PhD

**Project Part 1.**

I have chosen to analyze data over the 100 oldest women ever, both living and deceased. Data for this project can be found at this link [List of the verified oldest people - Wikipedia](https://en.wikipedia.org/wiki/List_of_the_verified_oldest_people" \l ":~:text=100 verified oldest women    Rank,years, 260 days  22 more rows). Variables to be examined in this data set will include age of the women, their nationality, whether they are deceased or living, and a closer look at their birthdates, by month. In total I will be analyzing 100 of the oldest women ever, but below, in Figure 1 is a small sample portion of the data I will be using.



Figure 1.

From the sample, you may immediately notice a few trends and similarities. This data set interests me because there are other factors that may possibly contribute to how long we live, not just the obvious that we hear every day.

In conclusion, I am interested to further analyze this data. I want to take a look at whether these women’s country of origin could play a role in how long they live. I would also like to examine the possibility that maybe those born in a certain month may have a higher chance at living longer.

**Project Part 2.**

For this portion of the project, I took a closer look at my categorical variables, the nationalities of the oldest women ever, and whether they are living or deceased. First, I will analyze their nationality. I found it interesting that of the 17 countries that made the list, the top 3 with the oldest women ever were the United States, Japan, and France. Figure 2 shows the frequency and relative frequency of this for the given countries.



Figure 2.

I know several things come into play here. Obviously, more developed countries will have higher life expectancies. Things like access to adequate food, healthcare, and technology play a role. It was no surprise to me that the United States was in the top 3. I also know record keeping makes a difference as well, which relates back to technology. I was also not surprised to find Japan and France in the lead. I mentioned food above, and it has long been said that the quality of food we eat can also impact our health. I believe the people of those countries eat more beneficial foods that contribute to their health.

Next, I took a look at those same countries that made the list, and whether the women from those countries are living or deceased. Figure 3 below illustrates those findings in a two-way table.



Figure 3.

From the list of the oldest women ever, 9 are still alive. In correlation with the findings from the first variable, 6 of those 9 women alive are from the United States, Japan, and France. I found this interesting to see the relationship between the variables come in to play in this instance. To conclude, after examining these variables, I do believe the country of origin has an effect on life expectancy.

**Project Part 3.**

In project 3 here I will examine how many of the top 100 oldest women ever were born in each month. This will suggest the possible question of, do those women born in a certain month possibly live longer? Below, in figure 4, is a quick chart to show the breakdown and then I will present my summary statistics for this data.



Figure 4.

The summary statistics are presented below in figure 5.



Figure 5.

Next, in figure 6, I will look at a histogram that shows how many of the oldest women ever were born in each month. As you can see, of the 100, the top 4 months with the most born are August, May, March, and October.

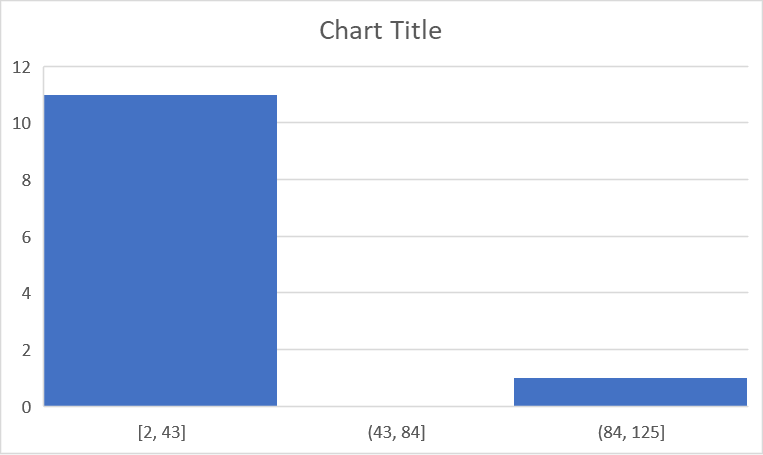


Figure 6.

Now, I will look at this same information presented on a box plot. This is presented in figure 7 below.

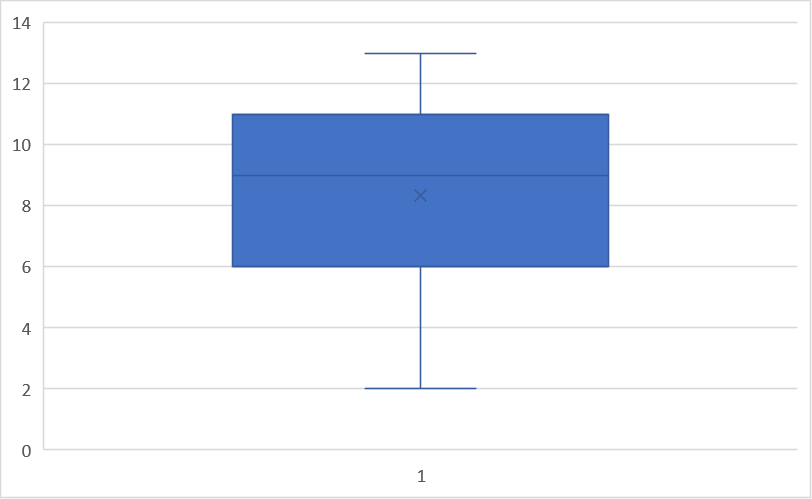


Figure 7.

From the given data, August holds the highest number born of oldest women ever. The distribution is almost symmetric with no outliers, with the mean being 8.33 and the median being 9. To me, it does raise the question of, do women born in August live longer? I know this is only a list of 100 women, and other factors and more data would have to be examined to even begin to try to answer this question without doubt, but I think this is an interesting start.

**Project Part 4.**

**Quantitative Hypothesis**

H0:  = 114

The null hypothesis is that the average age of the oldest women ever will be 114 years.

Ha:  ≠ 114

The alternative hypothesis here is that the average age of the oldest women ever will not equal 114 years.

I decided on this hypothesis after taking another look at all the ages of the women on my list. The age of 114 years appears drastically more than any other age, with not many more that are a lot higher than that.

**Categorical Hypothesis**

H0: P = 35%

The null hypothesis is that 35% of the oldest women ever will be from the United States.

Ha: P < 35%

The alternative hypothesis is that less than 35% of the oldest women ever will be from the United States.

I based my hypothesis off of previous analysis of my data. The information for this was presented earlier in the project sequence, in a frequency and two-way table.

**Project Part 5.**

To test the quantitative hypothesis, I created a bootstrap sample. I found the standard error to be 0.119492. I then figured the 95% confidence interval to be between 114.6864 and 115.1644. With this, I can reject the null hypothesis because the average age of 114 is outside of the confidence interval. Below, in figure 8, is a histogram for the bootstrap distribution.

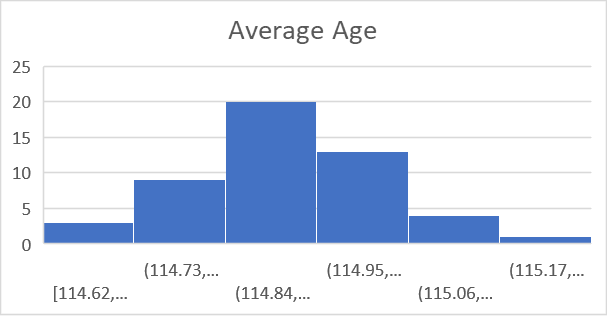


Figure 8.

Next, I tested the categorical hypothesis using bootstrapping. I found the standard error to be 0.056212. Then, I found the 95% confidence interval to be between 25% and 48%. With this, I fail to reject the null hypothesis stating that 35% of the oldest women ever will be from the United States since that percentage is within the confidence interval. Below, in figure 9, is a histogram for the bootstrap distribution.

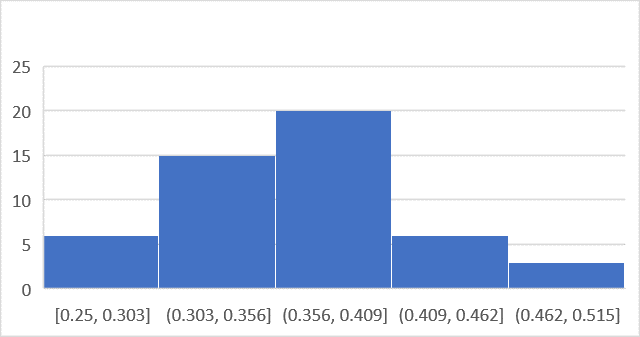


Figure 9.

**Project Part 6.**

For this part of the project, I will repeat my hypothesis test for my categorical variable using the appropriate formulas in excel. Again, my hypotheses are as follows: H0: P = 35% Ha: P < 35%.

Previously, I found my standard error to be 0.056212, using the formulas for this portion I found my standard error to be 0.047697. Also, in part 5 I found the 95% confidence interval to be between 25% - 48%. This time I found my 95% confidence interval to be between 26% - 44%. Although the confidence intervals are slightly different, I still fail to reject the null hypothesis since the percentage is within the confidence intervals. Below, in figure 10, is a snip from excel, with the formulas I used included.



Figure 10.

**Project Part 7.**

I will now repeat my hypothesis test for my quantitative variable. To refresh, my hypotheses are as follows, H0:  = 114 Ha:  ≠ 114.

Before, I had found my standard error to be 0.119492. This time, I found it to be 0.127995. In the previous portion, I found my confidence interval to be between 114.6864 and 115.1644. Now, I have found the confidence interval to be between 114.6975 and 115.1225. With this, I still reject my null hypothesis. Below, in figure 11, is a screenshot from excel with formulas used.



Figure 11.

**Project Part 8.**

For the final portion of this project, I will look at two conditional probabilities. I will be referring back to the two-way table from part 2. I am including a copy of this table below.

The first probability I will calculate is, what is the probability that of the oldest women ever, they are living given that they are from the United States. This will look like, P(living|US) so P=2/35 =0.057. From this I know that of the 35 oldest women ever who are from the United States, the probability that they are living is 5.7%.

The next conditional probability I found is, what is the probability that the person is from Japan given that they are deceased. Here, P(Japan|deceased) so P=24/91 =0.26. From this I can conclude that of the 91 deceased oldest women ever, 26% of them are from Japan.