Summary/ Abstract

This experiment studied the United States one-cent coins also known as pennies. The purpose of this experiment is apply statistical test such as t-test on the data given. In different year, the material used to make the one-cent coin is different due to the source of material. Therefore, we can make a hypothesis that different composition of pennies have different masses. In 1982, the composition changed, thus the data is separated into 1982 and 1982.5. The main group of masses to be analyzed is before 1982 and after 1982.5. As a result, we noticed that …….

Introduction

United States one-cent coins, also called as pennies are widely used in the United States of America. Pennies are minted by the United Mint. The four branches that produce coins are Denver, Philadelphia, San Francisco and West Point. The coin normally minted with the location marks such as D for Denver, P for Philadelphia, S for San Francisco and W for West Point. Sometimes, the penny does not have any marks on it, and we can assume it is from Philadelphia as Philadelphia didn’t put the mark on the penny before 1980. The composition of penny changed by year due to the limited amount of resources and the cost of the resources. This causes that the penny has different in the particular year. In 1982, there is a change in composition. The mass of penny changed from approximate 3.1g to approximate 2.5g due to different composition in the penny. In this year, the pennies with approximate 3.1g are categorized in 1982 group while the 2.5g pennies are categorized in 1982.5 group. In this experiment, we are not interested in how it is made, what the composition in it is or how to take the mass. Instead, we are interested in using the statistical tests to test on the data collected from pennies.

Experimental Procedure (Materials and Methods)

In this experiment, we started by removing the outliers of the data. There are no pennies with the same weight. They might be corroded, worn or other things that cause the decrease or increase in weight. When the weights are more or less than the arbitrary limit, those are outliers and should be removed from the data. We separated the data into 2 main groups: pre-1982.0 and post-1982.5 and ranked the mass in increasing sequence. Then we used the function in excel to get the average (avg), standard deviation (s), minimum value (min) and maximum value (max) of each group. We also calculated the relative standard deviation in parts-per-thousand (ppt) by using the formula: (s/avg)\*1000. We also calculated the arbitrary limit of the mass. Any value than exceed the avg+3s and below the avg-3s were removed. The remaining data underwent the same procedure until there are no more outliers.

Next, we started the year-by-year analysis. The final data after all removal of outliers is used for the analysis. The avg and s of each year were calculated and were used to plot a graph with year on the x-axis and mass of penny on the y-axis. Error bars were included in the graph. Then, all the 2009’s masses were removed from the post-1982.5 data and a new graph was plotted. A linear trend line was added to get the value of slope (m), intercept (b) and correlation coefficient (R). The standard deviation of the slope was calculated by using the formula: Then, we used the student t-test at 95% confidence level to test the data. The data is very large, therefore we use t table = 1.960 for an infinite number of degree of freedom. Then, we compare the highest and lowest average masses using the t-test. These procedures were repeated for the pre-1982 data starting from the year 1963 to the year 1982.

After that, we started the comparison of mint location. From the above data, we resort the data according to locations. The data were pooled and the avg, s, min, max were calculated for each location. The values were reported in 4 decimal places. The t-test is used to compare the avg of each location at 95% confidence level. This procedure was again repeated for the pre-1982 data starting from 1963 to 1982. Lastly, we conducted the distribution of masses. All the data above (1963-1982 and post-1982.5 except 2009) were combined into a single column. The masses were ranked from lowest to highest. Avg, s, min, max were calculated and reported in 4 decimal places. A histogram was generated using these data. Analysis Toolpack was installed in Excel to plot the histogram. A column of bins with a spacing of 0.0100g starting from 2.4000 until 2.6000 was set up. Then, the bins column and the data were used to plot the histogram and we were done with this experiment

Results and Discussion

**Removal of outliers**

The data that exceed the arbitrary limit are removed. Before each removal, the mean (avg), standard deviation (s), minimum, maximum and relative standard deviation in ppt (RSD) of the data are obtained. The formula below is used to obtain the data needed. For finding the minimum and maximum, just used the function in Excel to compare because the data are too much. All the data calculated are shown in the “Table Question 1-6”.

Formula used:

, x is the data

, n is the amount of data

**Year-by-Year Analysis**

To plot the graph with year on the x-axis and mass on the y-axis, the average masses and the standard deviation of each year are calculated with same formula as above. The data calculated is shown as “Table Question 7” in the below. A graph of average mass versus year is plotted and shown in “Graph of Question 7”. The standard deviation of each year is used for the error bars.

**Year-by-Year Analysis 1982.5-present**

The data from 2009 are removed from the post-1982.5 data. A new graph of mass versus year is plotted and shown in “Graph of Question 8”. A trend line is added to the graph and the gradient (m), intercept (b) and correlation coefficient (R) are obtained from the equation of trend line. LINEST OUTPUT is used to obtain other information needed to find the standard deviation of m (), the data are shown in “Table of Question 8”. Besides that, we can compare the m,b and R to make sure the m,b and R value obtained from the graph are correct. To calculate , we used formula below and get . To compare the m and , the and the values are calculated. Both value are negative, -0.0006 and -0.0007 respectively. Therefore we can confidently conclude that the historical trend in the mass of pennies since 1982.5 is decreasing.

Then, Student t-Test at 95% confidence is used to see the value of the slope, whether it is positive, negative or zero. The data are very large, so 1.960 is used for the t-value. The t-value is used to calculate the confidence intervals (C.I.). Formula for C.I. is shown below.

We get -0.0006 for and -0.0008 for . Therefore, we can conclude that the mass is decreasing with only 0.05% of masses are not in the range.

Then, we conduct t-test for comparison of means to compare the highest average mass and the lowest average mass at 95% confidence level. To run the t-test, we need to conduct the F-test first. We use the formula below and found that the F calculated < F value from table. From here, we can conclude that the standard deviation are not significantly different and the F-test passes. Then, we conduct the t-test (case2a). We found that the t calculated > t table. We can conclude that the average masses are significantly difference. All the data calculated are shown in “Table of Question 10 (Calculated)” and the DATA ANALYSIS function in Excel is used to make sure the calculation of F-test and t-test are correct.

Formulas:

, is obtained from the LINEST OUTPUT

,

**Year-by-Year Analysis of 1963-1982.0**

The analysis in this section are same as the upper section but with different range of data. The data start from 1963 to 1982.0. A graph is plotted and shown as “Graph of Question 11”. The σ\_m is 0.0059. 〖m+u〗\_m is -0.0002 and the m-u\_m is -0.0005 (Table of Question 11). Both values are decreasing, therefore the historical trend in mass of pennies between 1963 and 1982.0 is decreasing. We also get -8.8181E-05 for m+(t×u\_m) and -6.2242E-04 for m-(t×u\_m ), shown in “Table of Question 12”. For F-test, the F calculated (2.3248) > F table (1.5395). From here, we know that the standard deviation are significantly different and F-test failed. We conduct case 2b for the t-test and the formula is shown below. We found that t calculated (5.6190) > t table (1.9812). Hence, the average masses are significant different. All the data calculated are shown in “Table of Question 13”.

Formulas:

Degre of freedom=〖((s\_1^2)/n\_1 +(s\_2^2)/n\_2 )〗^2/(〖〖(s〗\_1^2/n\_1)〗^2/(n\_1-1)+〖〖(s〗\_2^2/n\_2)〗^2/(n\_2-1)) t calculated= (〖ABS(avg〗\_1-〖avg〗\_2))/(√(〖(s〗\_1^2/n\_1)+) 〖(s〗\_2^2/n\_2))

Comparison of Mint Locations

All the data from the year-by-year analysis 1982.5-present are sorted into P (Philadelphia) and D (Denver). Then, avg, s, min., max are obtained using the same way as above and reported in 0.0001-gram (refer to “Table of Question 14”). Then, the t-test for comparisons of means to compare the average masses of each sets at 95% confidence level is conducted, same formulas as above. The F calculated (1.0080) < F table (1.2295). We can say that the standard deviation are not significantly different. Then we get t calculated (0.0238) < t table (1.9611). Thus, the average masses are not significantly different. Then, it is repeated again with data from 1963 to 1982.0. The F calculated (1.2230) > F table (1.1675). The standard deviation are significantly difference. Then, t calculated (2.2064) > t table (1.9677), the average masses are significantly difference. All data are included in “Table of Question 15 (Calculated)” and “Table of Question 15 (Calculated)”.

Distribution of Masses

All the data from the year-by-year analysis 1982.5-present are sorted into single column and avg, s, min, max are calculated and reported in 0.0001-gram for each year (shown in “Table of Question 17”). DATA ANALYSIS function is added to the Excel. A “bins” column is created with spacing 0.0100g, starting from 2.4000 to 2.6000 (refers to “Table of Question 19”). Then the histogram is plotted and shown in “Histogram of Question 20”. From the histogram, we notice that the mass of pennies are mostly located at the mean, around 2.5g.

Conclusion

As a conclusion, we as student learned to apply statistical test on the data to interpret the data. We also get familiarize with the Excel. We also know that when F calculated > F table, the standard deviation are significantly difference and vice versa. Then we run the t-test and t calculated > t table, the average masses are significantly difference and vice versa. We also noticed that most of the pennies are weighted around 2.5g according to the histogram.