A Statistical Analysis of a sample of Skittles

Abstract

- The project I chose to address is: The statistical analysis of a packet of skittles. I chose this idea as I wanted to gain a better understanding of applying statistical and analytical methods to real – world applications and to learn more about the manufacturing and distribution of a packet of skittles. I also wanted to understand critical problem solving steps through statistical mathematical equations.
- In these experiments, I aimed to develop my understanding of these methods by applying them to a simple problem of determining the percentage of skittles that are a given colour. While this can be solved to a good approximation by taking a large sample size, statistical methods had to be interpreted to analyse the results.

Hypothesis

- According to the skittles website (www.skittles.com), each flavour is equally distributed into a package; thereby 20% of each flavour should occur in each package. The given hypothesis is that 20% of each bag should contain a given flavour. Alternatively, the null hypothesis is that the distribution of colours in a packet of skittles is unequal.
- I also researched the weight of each packet of skittles and the amount of skittles per bag. The average weight of a fun size packet of skittles is 26 grams and the average number of skittles per packet is 25.
- I also analysed a regular packet of skittles to compare the results to visualise if the distribution was similar. The average weight of a regular packet of skittles is 55g and the average number of skittles per packet is 51 skittles.

Background Research

Before I started my project, I carried out a lot of research. I looked into statistical and analytical methods to interpret data and the manufacturing and distribution of a packet of skittles.

Production Of Skittles

During my research, I also looked into the manufacturing and distribution of a packet of skittles. Skittles are manufactured by using sugar, gelatine, hydrogenated vegetable oil, along with fruit juice, citric acid and artificial colourings and flavourings. The Ingredients are heated and turned into syrup which is the poured into candy shaped dents. The rounded gelatine sweets are coated with liquid syrup. Each skittle is coated several times, thus resulting in a hard shell. After the manufacturing of the skittles, they are distributed into packets. The coloured skittles are combined to create a multi – coloured batch. An automated packaging machine weighs and fills each package with the correct amount of skittles. The packages are then heat sealed and ready for shipment.

Statistical Methods

Throughout my investigation, I also looked into statistical and analytical methods to interpret data. I looked into different methodologies to calculate and graph data. I observed how to use the standard deviation which is a quantity expressing by how much the members of a group differ from the mean value for the group. I also found different techniques to represent collected data on a graph. I researched into bar charts, line plots and scatter plots. I studied many different statistical formulas such as the binomial distribution and the Chi Squared Test to obtain a better understanding of my analysis.

Theory

- In these experiments in which the same quantity was to be measured multiple times, it is desirable to have a method by which I can analyse the results of the experiments. By using the same procedure for both experiments, a simple arithmetic mean is a good choice to interpret the collected data. To obtain the mean of a sample of n, the formula is:

$$\mu = (\Sigma Xi) / N$$

- In the experiments it is necessary to consider the specific distribution which describes the results. To theoretically analyse my data, I studied the Chi Square Test. A chi square statistic is a measurement of how expectations compare to results. If the chi square value is small, we can accept our null hypothesis. If it is a large value, we can refute our null hypothesis. The formula for the Chi-square Test is:

$$x^2 = \sum \frac{(o - e)^2}{e}$$

- I calculated the standard deviation to analyse the dispersion of the data from the mean. To calculate the standard deviation, the formula is:

$$\sigma = \sqrt{\frac{\sum(x - \mu)^2}{N}}$$

Experimental Method

- I carried out a number of experiments for during my project, using regular size packets of skittles that weigh 55g and smaller packets of skittles that weigh 26g. Once I had acquired my equipment, I started the analysis:
  1. I measured the weight of each packet of skittles.
  2. I counted the amount of skittles per packet. This allowed me to quantify the mean amount of skittles per packet.
  3. Next I counted each colour by splitting the packets into subsets. These subsets were red, green, yellow, orange and purple. I then compared the sum of all the colours in each pack to the total of the pack as a form of error checking.
  4. All numbers were entered into a database to allow for easy analysis of the data.
- In order to carry out a precise analysis, I needed to include variables in my project. For the first experiment, I measured the weight of each packet of skittles. I used the same weighing scale to avoid errors. For the second experiment, I counted the number of skittles per packet. For the third experiment, I counted each colour by splitting them into subsets which were red, green, yellow, orange and purple. I used the same scientific weighing scale in the third experiment to avoid errors. I entered all the numbers into a spreadsheet to allow for easy analysis of the data. As to avoid calculational errors in my analysis, I used Microsoft Excel to calculate my results. This was then repeated for the regular packet of skittles.

Discussion

- The hypothesis for the project was that there is an even distribution of 20% between the five colours. The null hypothesis for the experiments was that there is an unequal distribution between the colours. This may be due to sampling or experimental errors. I expected to find that the mean weight of the regular packet of skittles was going to be 55g and the smaller packet of skittles to be 26g.
- From the data that I collected from my experiments, I analysed and calculated the data using many statistical methods. I applied different mathematical theories to aid my analysis. By using standard deviation, I could analyse the spread of the collected data. I also used the chi – square goodness of fit in my calculations to further statistically analyse my data. The chi – square test was used to determine whether the collected data was consistent with a hypothesized distribution of 20%.
- During the analysis of the data, I found that some of my results were unexpected. In the experiment with the regular packets of skittles, I discovered that one of the packets weighed 60.9g. This was 5g above the expected weight. This may have been due to experimental error or in the distribution process of the skittles. My collected data supported previous research to the project. Many statistical analysis's have found that there is an equal distribution among the colours, which aided my analysis.

Results

Regular Packet of Skittles (55g)

- The hypothesis of the project was that there would be a equal distribution of 20 percent among the five colours. For the regular packets of skittles, I used a sample size of n = 1856 skittles.
- Below are the results for the mean weights and standard deviations for the skittles with and without the packet.

Mean Weight of Packet	58.66
S.D of Packet Weight	0.77
Chi Square	1.00

Mean Weight of Skittles	57.44
S.D of Skittle Weight	0.76
Chi Square	1.00

- Below are the results for the mean number, mean percentage and standard deviation of the skittles.

	Total	Red	Green	Orange	Yellow	Purple
Mean Number of Skittles	51.56	9.56	9.72	11.86	8.97	11.42
Mean Percentage of Coloured Skittles		19%	19%	23%	17%	22%
S.D of Number of Skittles	0.84	2.92	2.78	2.78	2.96	3.36
Chi Square Goodness of Fit Test	1.00	1.00	1.00	1.00	1.00	1.00

Smaller Packet of Skittles (26g)

- Below are the results for the mean weights and standard deviations for the skittles with and without the packet. For the smaller packets of skittles, I used a sample size of n = 1423 skittles.

Mean Weight of Packet	27.91
S.D of Packet Weight	0.85
Chi Square	1.00

Mean Weight of Skittles	26.86
S.D of Skittle Weight	0.82
Chi Square	1.00

- Below are the results for the mean number, mean percentage and standard deviation of the skittles.

	Total	Red	Green	Orange	Yellow	Purple
Mean Number of Skittles	24.53	5.33	4.76	5.10	4.78	4.57
Mean Percentage of Coloured Skittles		22%	19%	21%	19%	19%
S.D of Number of Skittles	1.74	2.49	2.09	1.88	1.61	2.02
Chi Square	1.00	1.00	1.00	1.00	1.00	1.00

Conclusions

Packet Weight

- For the regular packet of skittles, I expected the weight of the packet to be 55 grams, but from my analysis, I saw that the average weight of the packet was 58.66 grams. I also examined the standard deviation of the regular packet of skittles, which was 0.77. I calculated the chi – square test for the weight of the packet, the weight of the skittles and the number of skittles and the result was 1.00. This meant that the hypothesis was stated truly and we fail to reject it.
- For the smaller packet of skittles, I expected the weight of the packet to be 26 grams, but from my results table, I could see that the mean weight of the packet was 27.91 grams. I also examined the standard deviation of the smaller packet of skittles and the result was 0.82. I calculated the chi – square test for the weight of the packet, the weight of the skittles and the number of skittles and the result was 1.00. Correlating with the regular packets, we can see that the hypothesis is true.

Colour Distribution

- For the colour distribution of the regular packet of skittles, I expected there would be an equal distribution of skittles among the five colours. From my data, I realised that the average distribution percentage was close to the hypothesis. By using the chi – square test, I analysed if the results were compatible with the hypothesized distribution. I calculated the chi – square test for all the colours and realised that the result was 1.00. This meant that the expected distribution was correct and therefore, we fail to reject the hypothesis.
- For the colour distribution of the smaller packet of skittles, I expected there would be an equal distribution among the colours. From my collected data, I realised that the average distribution percentage was close to the hypothesis, correlating to the regular packet. I calculated the chi – square test for the colour distribution and all the colours had a result of 1.00. Therefore, the hypothesis was true and we failed to reject the stated hypothesis.

Overview

- During the statistical analysis of a packet of skittles, I have examined the statistical properties of the distribution of colour using analysing techniques and statistical methods. With sample sizes of n = 1856 for the regular packet of skittles and n = 1423 for the smaller packet of skittles, I was able to obtain 95% confidence intervals less than ±3% for each colour. The results are consistent with an even distribution between each of the five colours found in a bag of skittles.