

Project 1: The Great M & M Count

Problem Statement:

This statistical experiment is designed to accumulate statistical data on the average package of M & M's (non-peanut) as well as information regarding one particular package, labeled 'my package'. Once the data is accumulated, statistical analysis is used to answer possible questions regarding a package of M & M's. For instance, was 'my package' a reasonable representative of the average package of M & M's or one could ask what the average number of M& M's per package is. The analysis provides answers to these questions, and more, with detailed information on the class sample of M & M's.

Procedure:

To start the experiment, a random package of M & M's is opened and recorded as 'my package'. Each member of the class also opens a package, which, as a whole, represents the class sample. A count of the contents of the package is made after sorting M & M's into subsets. These subsets are red, blue, green, yellow, orange and brown. Once a count is made, a further subset of each color is sorted and labeled defective. To be defective an M & M must have some fault in the candy shell, aside from misprint of the M. Recorded data from the class is compiled and made available for comparison to 'my package'.

The general analysis is made using statistical commands within Microsoft Excel. This analysis includes sample mean, median and mode calculations, as well as standard deviation and variance. Finally, a histogram is created to visually examine the sample data.

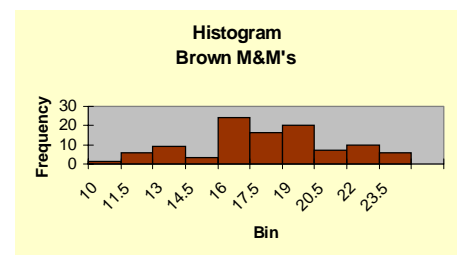
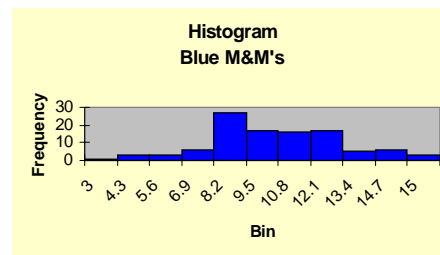
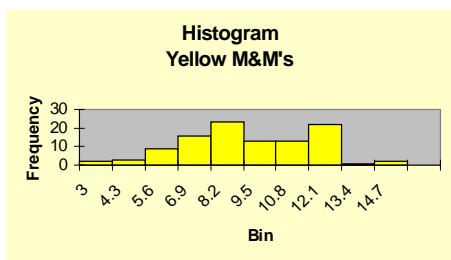
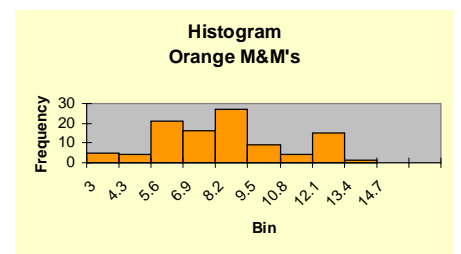
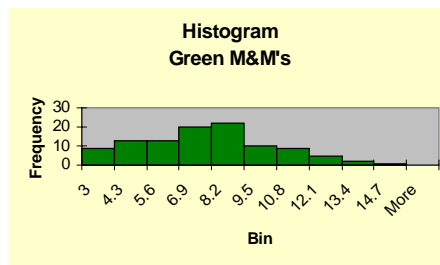
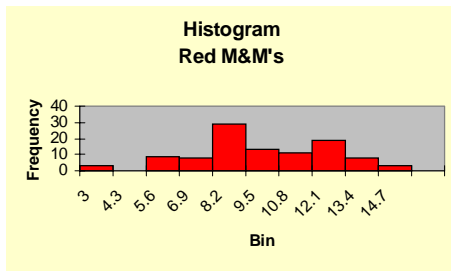
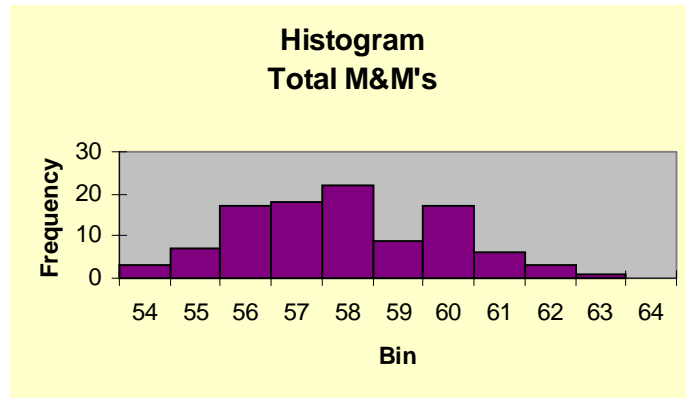
Collected Class Sample:

The results of the analysis on the class sample are in Table 1, including mean, median, mode, standard deviation, variance and range for the total and each subset (defective subset is shown as the total of all defects in all colors).

	Mean	Median	Mode	Standard Deviation	Variance	Range
Total	58.7	58	58	8.4	70.7	90
Red	8.9	9	8	2.8	7.6	13
Orange	8.1	7	5	8.4	71.1	87
Yellow	8.3	8	6	2.6	6.6	12
Green	6.7	6	6	2.5	6.4	12
Blue	9.4	9	9	2.7	7.5	15
Brown	17.2	17	17	3.4	11.8	15
Defective	1.9	1	0	2.1	4.2	10

Table 1: Statistical analysis data for total sample and sample subsets (color and defects).

The analyzed data is also shown graphically in the following histograms representing the total sample and each corresponding color.



Using both the data in Table 1 and the histograms above it can be shown that the sample chosen has a very smooth distribution. Each color does have a slightly different distribution and it is very obvious, from the mean values and histograms, that brown M&M's are much more common in a package than any other color. One data point did have very high values with a package total of 141 M&M's, yet this deviance didn't effect was removed from the histograms and only slightly effects the mean, mode and median. Variance, range and standard deviation are effected by the unusually high value, though.

Single Sample:

The single package, 'my packet', is then compared with the overall sample results. The count is shown in Table 2 and also graphically in Figure 1 as a comparison against the sample mean values.

Color	Count	# Defective
Red	12	0
Orange	6	0
Yellow	9	0
Green	6	0
Blue	10	0
Brown	17	0
Total	60	0

Table 2: My package values

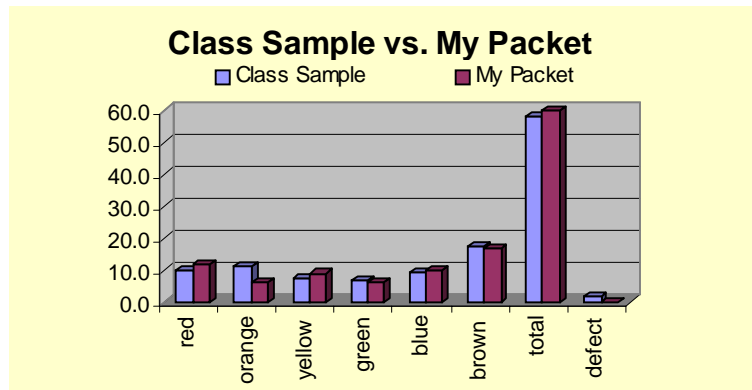


Figure 1: Comparison of single packet to sample.

Given the graphical representation of the single packet vs. the mean values of the class sample, it appears that 'my packet' is a very good representative of the sample. It is slightly unusual in that there were no defects in 'my packet'; however, the mean for the class sample was 1.9 for the total number of defects. Therefore zero defects is fairly close given a package total of M&M's at 60.

Summary and Conclusions:

The goal of this exercise was to examine a sample of M&M packages and accumulate data in a manner that provides important information regarding the sample. This information can then be used to answer questions, such as, "Was my packet a good representation of the class packages?" or "What is the average number of M&M's in a packet?"

The analysis shows that the single packet was indeed a good representation of the classes packets. The average number of M&M's in a packet turned out to be 58.7 M&M's and much more information was gathered to answer further questions regarding the sample set.

What I Learned:

I think the most valuable tool I came out of this experiment with is the ability to use statistical tools and analysis in Excel. I had no idea that Excel would calculate and plot histograms and that so much statistical analysis tools existed. The Excel package made this project much simpler than it would have been to complete by hand.