Josh Richter
Engr. 315
Probability and Statistics
Project 1 (M&M)

(I)   Problem:

The task at hand was to see how one package of M&M's, my package,
compared to the other packages in the class. While doing this, we will be further
developing our skills in statistics.

Some questions that one might ask during this analysis include:

(A)   (Generally) How does one pack of M&M's compare to the average pack?
(B)   On the average, how many brown M&M's come in a pack?
(C)   What is average total quantity of M&M's in a pack?
(D)   How large is the range between least to most number of M&M's in a
       pack?

(II)   Procedure:

As part of the population it was my duty to gather information on my
package of M&M's. The first step was to separate the M&M's by color and
determine the total amount in the pack. The next step was to see how many
M&M's were defective, again sorting by color. Each individual then submitted
his/her data so that it could be pooled and tallied with the other data from all
students in the class. The information collected was then put into a spreadsheet.
After obtaining the information from the spreadsheet, I compared my package of
M&M's to the entire population of packages. Using a stem and leaf diagram I
separated the column of "total" quantity of M&M's in a pack so I could easily
manipulate the data. After obtaining equations from our text, I calculated the
mean, average, and standard deviation for various colors of M&M's.

(III) Aggregate: Data

Entire population
Mean number in a pack => 58.69
            Median => 58.00
    Standard Deviation => 59.31

Mean => 6104/10 = 58.69

(III) Aggregate: Data (Cont.)

Blue M&M's

Mean => 978/104 = 9.4

Std. Deviation => SQRT((9970-9.40)/103)
= **9.83**

Red M&M's

Mean => 8.9

Std. Deviation => SQRT((9079-8.93)/103))
= **9.38**

Total Population:

Color	Red	Orange	Yellow	Green	Blue	Brown
Mean	8.93	8.1	8.35	6.7	9.4	17.2

Group	Std. Deviation	Mean
Total Pop.	59.31	58.69
Blue	9.83	9.4
Red	9.38	8.9

IV) Single Sample:

My package of M&M's seemed to follow the general trend of the average pack. It seemed as though 59 was the average total number of M&M's in a pack. This was pretty close to my 60. The largest number of any one color of M&M seemed to be brown. For the average pack it was a quantity of 17 which wasn't too far form my 22. Overall, I don't believe that my package of 23. M&M's was unusual.

Data Collection Process:

Number in pack	Tally
51	1
52	0
53	1
54	1
55	7
56	17
57	18
58	22
59	9
60	17
61	6
62	3
63	1
141	1

Color	Red	Orange	Yellow	Green	Blue	Brown	Total
Count	6	5	9	8	10	22	60
# Defective	0	0	1	0	0	1	2

(V)   Conclusion:

The data collected from the entire class was pretty good. The quantities of total M&M's in a pack, number of red, brown, etc. were pretty consistent with the exception of one package. There was one package that didn't really follow suit. It could have been an extraordinary pack, or it could have been miss keyed while the data was being entered. This one outlier pack did not seem to effect the data calculations too much; only the range was affected. For the total population, 59 seemed to be the average quantity of M&M's in a pack. Of the total quantity of M&M's in a pack, brown seemed to be the color which occurred most in the average pack. The number of brown M&M's in the average pack was 17. While manipulating the data for the entire population I found the range, the difference between the least and most number of M&M's in a pack, to be 90. Using a trimmed mean of 1.3% or more you can eliminate any outlying points. Utilizing this method, I found the new range to be 10.

(VI)   "What I learned:"

Throughout this project I've learned quite a few neat and extremely useful tools. I learned how to eliminate inconsistent data points by using a trimmed mean. I further enhanced my knowledge of mean, median, and mode calculations, and what it means if they are close in value or far apart. I also learned how to use stem and leaf diagrams to quickly and accurately compare and record data. Some important decisions had to be made throughout this project, and if I had chosen differently, I believe that I would have had slightly different results. One decision that I had to make was whether or not to include specific data, even though it is so far from being realistic. I decided to use the data anyway, and let the trimmed mean take care of the inconsistent points. Another choice I had to make was what kind of data gathering procedure to use. Different methods are better for certain situations.