

How much does an M&M weigh?

**Amy G. Froelich
W. Robert Stephenson
Department of Statistics
Iowa State University**

1. Data Collection – Who? What?
2. What should be the label weight of a Fun Size bag of M&Ms?
 - a. Display and numerical summaries
 - b. The Normal model
3. What is the relationship between the number of M&Ms in a bag and the weight of the bag?
 - a. Regression
 - b. The dangers of extrapolation
4. The next level
 - a. Constrained regression
 - b. Designed experiments
5. Discussion

Activity #1: Data Collection

Who do we collect data on and what data do we collect are fundamental questions for any investigator, researcher or statistician. Throughout the first part of the semester, we will be using data collected from Fun Size bags of M&Ms.

“Who” are we collecting data on?

What data should we collect?

Variable Name									
Bag 1									
Bag 2									
Bag 3									
Bag 4									
Bag 5									
Bag 6									
Bag 7									
Bag 8									
Bag 9									
Bag 10									

Data on bags of Plain M&Ms.

Group	TotalWgt	NetWgt	Number	Group	TotalWgt	NetWgt	Number	Group	TotalWgt	NetWgt	Number
1	21	18	23	5	22	21	26	9	22	22	24
1	22	21	25	5	20	20	22	9	23	22	25
1	21	20	24	5	20	19	22	9	21	21	24
1	21	20	24	5	21	21	24	9	21	20	23
1	21	21	24	5	21	20	23	9	21	20	23
1	20	19	22	5	22	21	24	9	20	20	22
1	23	22	26	5	21	21	23.5	9	21	21	24
1	21	21	24	5	22	21	25	9	23	22	25
1	22	21	23	5	22	22	25	9	22	21	24
1	21	20	23.5	5	23	22	25	9	23	21	25
2	18	17	22	6	19	18	22	10	20	19	22
2	20	19	25	6	21	20	24	10	22	21	24
2	19	19	24.5	6	20	19	23	10	20	19	23
2	18	18	23	6	21	20	24	10	21	20	23
2	19	18	24	6	21	20	23	10	22	21	24
2	20	19	25	6	20	19	22	10	21	20	23
2	18	18	23	6	21	20	24	10	21	20	23
2	18	18	22	6	22	21	25	10	19	19	22
2	19	18	23	6	20	19	23	10	20	19	22
2	18	18	23	6	21	20	24	10	22	21	24
3	21	21	23	7	23	22	24	11	19	19	24
3	21	21	23	7	21	21	23	11	21	20	24
3	23	22	25	7	21	19	22	11	20	20	24
3	23	22	25	7	22	21	24	11	20	19	23
3	22	21	24	7	21	20	23.5	11	20	19	24
3	20	19	22	7	21	20	24	11	20	19	24
3	22	22	24	7	23	22	25	11	20	19	24
3	21	20	23	7	22	21	23.25	11	20	20	24
3	23	22	25	7	22	21	23.5	11	20	19	24
3	25	25	28	7	22	20	24	11	20	19	24
4	22	21	23	8	22	21	25				
4	21	21	23	8	21	20	24				
4	22	21	24	8	21	20	23				
4	22	22	24	8	22	21	24				
4	21	20	23	8	23	21	24.5				
4	22	21	22.5	8	20	21	23				
4	21	20	23	8	21	21	23				
4	21	20	23	8	22	22	24				
4	20	19	22	8	21	20	22.75				
4	21	20	23	8	21	21	23.5				

Activity #2: The total weight of bags (bag plus contents).

How would you display and summarize these data?

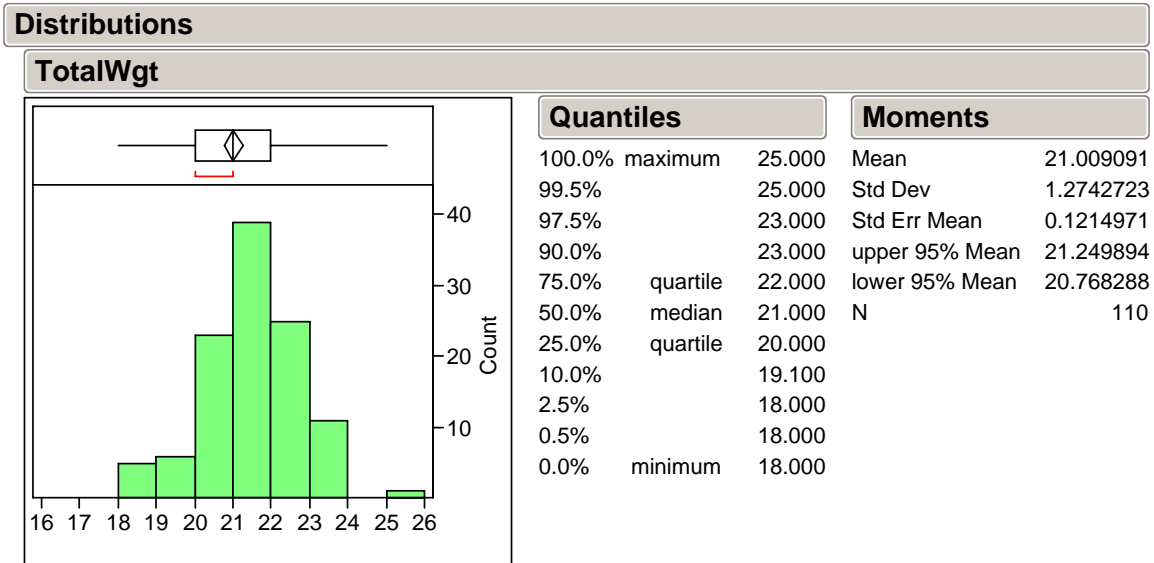
Graphical Display

Numerical Summaries

Fun Size bags of M&Ms are sold as a part of a larger package. According to the Fair Packaging and Labeling Act, the law does not require each Fun Size bag to carry a label weight; only the larger package must contain a weight of the contents in the package. If we wanted to sell Fun Size bags of M&Ms separately, we would need to determine a label weight to place on each individual bag.

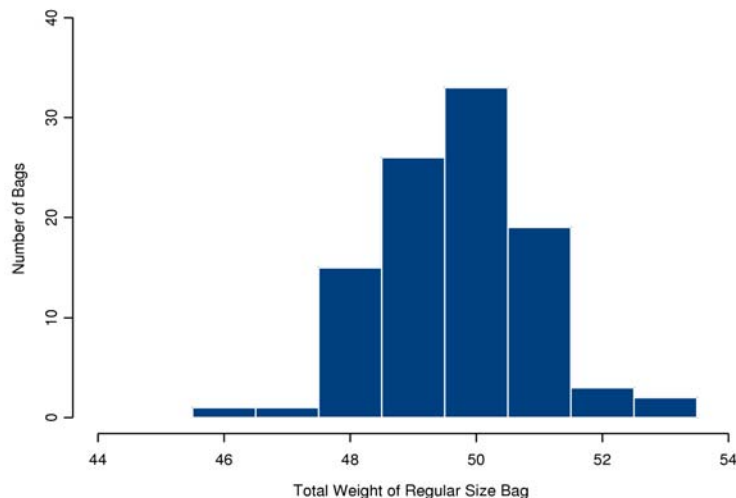
What label weight for a Fun Size bag would you choose and why?

Fun Size bags of Plain M&Ms



M&Ms are also sold in larger individual packages. These packages are sold individually and so according to the law, must carry a labeled weight for the contents in the package. Below is the distribution of the total weight of a sample of 100 regular size bags of M&Ms. The summary statistics for these data are Minimum = 46g, 1st Quartile = 49g, Median = 50g, 3rd Quartile = 50g, Maximum = 53g, Mean (\bar{y}) = 49.68g, Standard Deviation (s) = 1.2299g.

Distribution of Total Weight of 100 Regular Size Bags of M&Ms



What labeled weight do you think is listed on regular size bags of M&Ms?

The actual labeled weight of the regular size bags of M&Ms is 47.9 grams. Given this information, pick one summary value from the distribution of the total weight of the sample of 110 Fun Size Bags to use as the labeled weight of the Fun Size bags. Explain why you now chose this value.

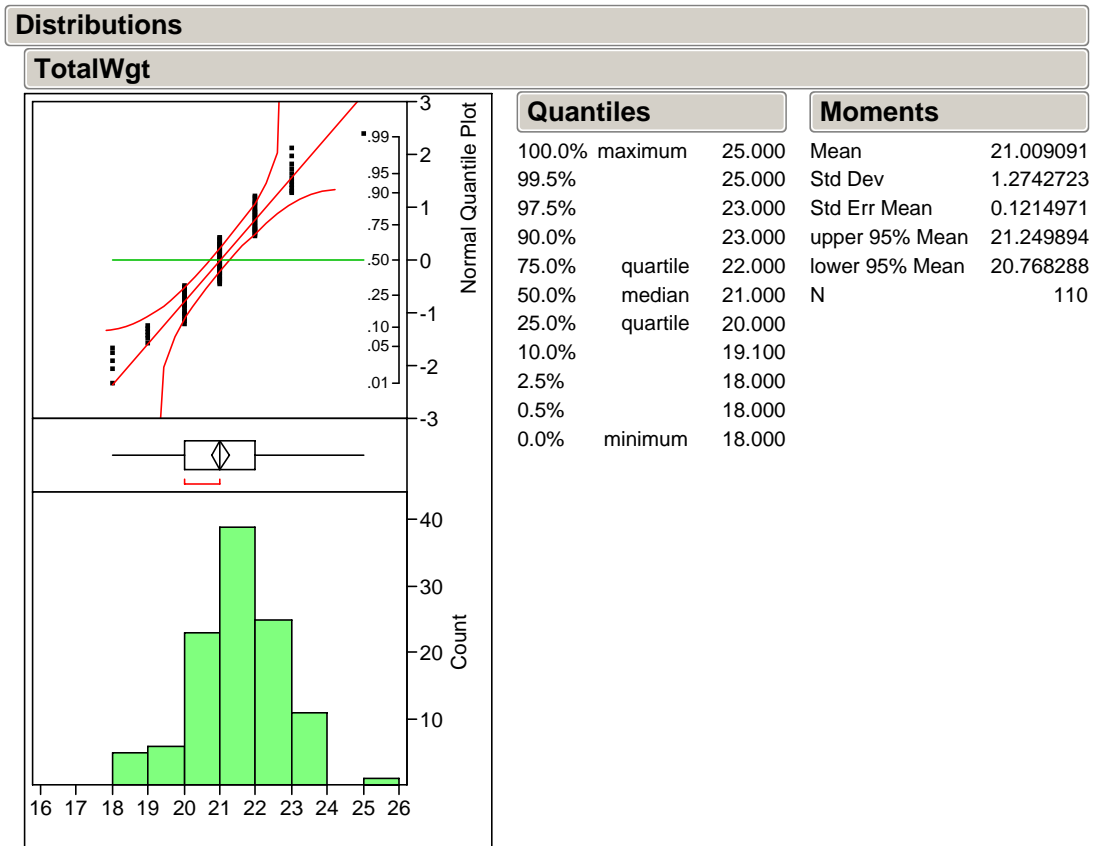
The law that regulates the requirements for packaging and labeling of food and other products is called the Fair Packaging and Labeling Act. To determine a labeled weight for a package, the following portion of the law is relevant.

Variations from the stated weight or mass, measure, or numerical count shall be permitted when caused by unavoidable deviations in weighing, measuring, or counting the contents of individual packages which occur in good packaging practice: Provided, that such variations shall not be permitted to such extent that the average of the quantities in the packages comprising a shipment or other delivery of the commodity is below the quantity stated, and no unreasonable shortage in any package will be permitted even though overages in other packages in the same shipment or delivery compensate for such shortage. Variations from stated quantity of contents shall not be unreasonably large.

Given this information, make a final determination of the labeled weight of Fun Size Bags of M&Ms. Explain the reasoning behind your answer.

The Fair Packaging and Labeling Act deals with the weight of the contents (net weight), not the weight of the contents plus the packaging (total or gross weight). The data we have been dealing with are the total weight of the Fun Size bags. How would you adjust your final labeled weight of the Fun Size bags to account for this difference?

The Normal model (Fun Size bags)

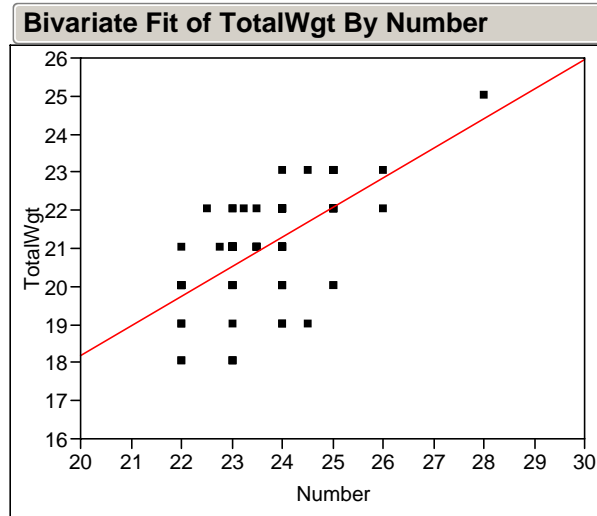


Do you think that a Normal model is appropriate for modeling the net weight of regular size bags of Plain M&Ms? Explain briefly.

Activity #3: Regression (Total weight on number of M&Ms)

Response variable: Total weight

Explanatory variable: number of M&Ms



— Linear Fit

Linear Fit

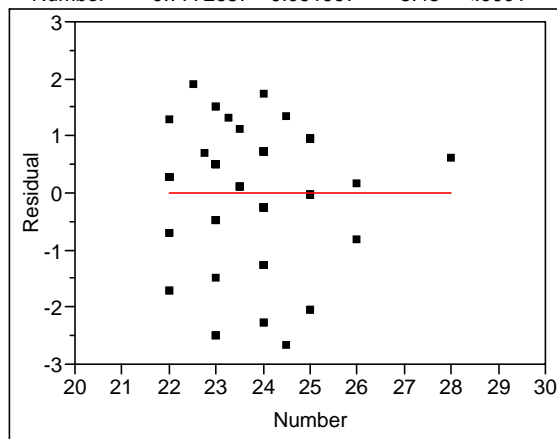
$$\text{TotalWgt} = 2.6372843 + 0.7772687 \text{ Number}$$

Summary of Fit

RSquare	0.397937
RSquare Adj	0.392362
Root Mean Square Error	0.99331
Mean of Response	21.00909
Observations (or Sum Wgts)	110

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	2.6372843	2.176536	1.21	0.2283
Number	0.7772687	0.091997	8.45	<.0001*



Give an interpretation of the slope estimate within the context of the data.

Give an interpretation of the estimated y-intercept within the context of the data.

Do you notice anything unusual about these estimates?

Activity #4: Constrained Regression Estimates

In Activity #3, we found that the slope and intercept of a regression line do not estimate the physical quantities of the average weight of an empty bag (intercept) and the average weight of a single M&M (slope) very well. In the activity below, we would like to see if we can determine better estimates of these quantities. Below are some simple summary statistics calculated from the data collected in Activity #1.

	Total Weight	Net Weight	Number
Number of bags	110	110	110
Mean	21.01g	20.24g	23.64

What is the weight of an empty bag?

- From the simple summary statistics, come up with an estimate of the weight of a single empty bag. Explain your reasoning.
- If we have Total Weight as the response variable and Net Weight as the explanatory variable, give an interpretation of the intercept of the line that relates Total Weight to Net Weight.
- The least squares regression equation relating Total Weight to Net Weight is:

$$\text{Predicted Total Weight} = 2.55 + 0.912 * \text{Net Weight}$$

- What is the value of the least squares intercept? Given your answers in (a) and (b), does this value look right? Explain your answer.
- What is the value of the least squares slope? Why does this value seem wrong? Hint: Think of the interpretation of the slope within the context of the problem. What should the value of the slope of the line relating Total Weight to Net Weight be?
 - If we force the slope to be the correct value, the least squares estimate of the intercept is 0.77. How does this compare with your estimate in (a)?

What is the weight of a single M&M?

- (a) From the simple summary statistics, come up with an estimate for the weight of a single M&M. Explain your reasoning.
- (b) If we have Net Weight as the response variable and Number as the explanatory variable, give an interpretation of the slope of the line that relates Net Weight to Number.
- (c) The least squares regression equation relating Net Weight to Number is:

$$\text{Predicted Net Weight} = 2.25 + 0.761 * \text{Number}$$

What is the value of the least squares slope? Given your answers in (a) and (b), does this value look right? Explain your answer.

- (d) What is the value of the least squares intercept? Why does this value seem wrong? Hint: Think about the interpretation of the intercept within the context of the problem. What should the value of the intercept of the line that relates Net Weight to Number be?
- (e) If we force the intercept to be the correct value, the least squares estimate of the slope is 0.856. How does this compare with your estimate in (a)?

Activity #5: Experimental Design and Regression

In Activity #1, we conducted an observational study by selecting bags of M&Ms, weighing the bags and counting the number of M&Ms in each bag. From the analysis of the data in Activity #3, we found that using least squares regression to predict the weight of an empty bag and the average weight of a single M&M was not very successful.

Your task in this activity is to design and conduct an experiment to obtain reasonable estimates of the average weight of a peanut M&M and the average weight of an empty Fun Size Bag using least squares regression. You will have a Tupperware container of M&Ms, 20 empty Fun Size Bags, and a scale to conduct the experiment.

Grading Rubric for the Design of Experiments

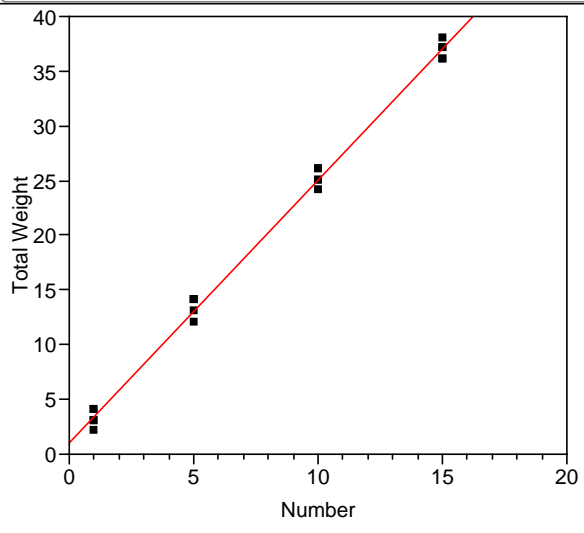
In order to have a good experiment, you must have control, randomization and replication within your experiment. For Activity #5 and the Project, these are the questions you must correctly answer when describing your experiment.

- Control - The main idea of control in experiments is to control the sources of variation other than the factors being tested. The conditions of the experiment should be as similar as possible for all experimental units. Have you described variables that could affect your experiment other than the treatment? Did you control for these variables in your experiment? How did you control for these variables?
- Randomization - The main idea of randomization is to spread the effect of unknown or uncontrollable sources of variation in your experiment evenly across the treatment groups. The experimental units should be assigned to the treatments at random. How many total trials do you have in your experiment? Did you describe how you assigned an experimental unit to a trial? Did you randomize the order of the trials? Did you describe how you randomized the order of the trials? The description of the randomization process must be complete; simply saying, “I randomized” is not sufficient. The grader should be able to exactly reproduce your experiment.
- Replication – The main idea of replication within the experiment is to apply the same treatment to several experimental units. Did you correctly identify your experimental units? Did you complete several trials at each treatment level using different experimental units? How many trials do you have at each treatment level?

Note: It is not necessary to repeat the entire experiment. The fact that other groups are doing similar experiments will give us the opportunity to see if results are repeatable.

Computer Output

Bivariate Fit of Total Weight By Number



Linear Fit

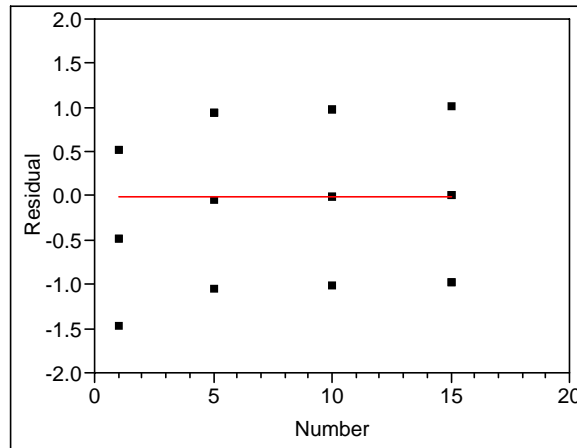
Total Weight = 1.1024831 + 2.393228 Number

Summary of Fit

RSquare	0.995941
RSquare Adj	0.995716
Root Mean Square Error	0.847369
Mean of Response	19.65
Observations (or Sum Wgts)	20

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	1.1024831	0.337318	3.27	0.0043*
Number	2.393228	0.036009	66.46	<.0001*



For more information on the activities presented in this session go to <http://duckworth.stat.iastate.edu/~amy/NSF/activities.html>