

Engaging Students in Statistical Discovery

Amy G. Froelich
W. Robert Stephenson
William M. Duckworth

Department of Statistics
Iowa State University

Introductory Statistics Course for Mathematically Gifted Students

Developed new laboratory activities that have students:

- study and explain the variation in weight of bags of M&Ms.
- use statistical arguments to support or refute a claim of gender bias in the salary levels of men and women at a company.
- discover the effect of extrapolation on the interpretation of regression coefficients.
- develop different methods using both manipulatives and the computer for obtaining a simple random sample from a population.
- experience how statisticians solve problems by designing experiments, analyzing the resulting data, and drawing conclusions from the analysis.
- develop intuition about probability using manipulatives and the computer.
- use simulation to discover the Central Limit Theorem.
- use simulation to reveal the meaning of a confidence interval for a population parameter.
- discover the importance of randomization in experiments.
- discover the meaning of Type I error rates and power in hypothesis testing.

Descriptive Statistics with M&Ms

Problem: M&M Fun Size Packages contain smaller bags of M&Ms not intended for resale. By law, the weight of these individual bags does not have to be stated on the package.

How should you label the individual packages of M&M Fun Size Bags to accurately report the weight of the contents?

Student Activity #1: Data Collection

Each group gathers three variables on 10 bags of Fun Size M&Ms: Total Weight, Net Weight, Number of M&Ms. Based on these values, students are asked to give a label weight for these bags and explain their choice.



Activity Outcome: Students discover variation and importance of data collection.

Student Activity #2: Data Analysis

Using the combined data from Activity #1, students study the distribution of Net Weight of the bags and use this information to determine a label weight of the M&M bags.

Activity Outcome: Students discover important aspects of distributions, such as mean and median.

Student Activity #3: Using the Normal Distribution

Does the normal distribution help describe the distribution of the Net Weights of the bags? Do any of the three variables seem to follow the normal distribution?

Activity Outcome: Students discover Normal Quantile Plots and using the normal distribution as an approximation for real data.

Solution: The Fair Packaging and Labeling Act of 1967 essentially states that no bag sold should have a weight below the stated label weight. How should you label the bags given this information? How does this weight differ from the one you selected before?

Activity Outcome: Students discover the fundamental importance of variability and the role of Statistics in many applications.

Regression and Experimentation with M&Ms

Student Activity #1: Students have collected data on Total Weight and Net Weight of M&M bags and Number of M&Ms in the bag. Can we use summary statistics and regression to estimate the average weight of an empty bag and the average weight of a single M&M?

Summary Statistics

	Total Wgt	Net Wgt	Number
# of bags	110	110	110
mean	21.01g	20.24g	23.64

Regression Equations

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

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

Estimate of weight of a single M&M:

Summary Statistics: Mean Net Wgt / Mean Number = 0.867g

Regression: Slope of Net Wgt on Number = 0.761g

Estimate of weight of an empty bag:

Summary Statistics: Mean Total Wgt – Mean Net Wgt = 0.77g

Regression: Intercept of Total Wgt on Net Wgt = 2.55g

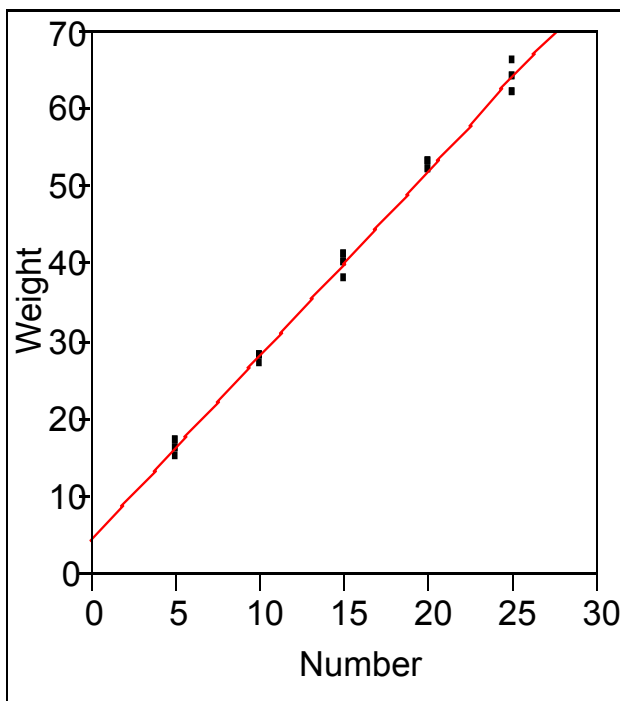
Activity Outcome: Students discover differences are due to extrapolation.

Student Activity #2: Design and conduct an experiment to obtain a reasonable estimate of the average weight of a peanut M&M and the average weight of a Ziplock bag using a regression of Weight on Number.

Usual Interpretations:

Intercept: Predicted Weight of Ziplock Bag when the number of M&Ms in the bag is zero = 4.16g.

Slope: When the number of M&Ms in the bag increases by one, the weight of the bag will increase on average by 2.392g.



$$\text{Predicted Wgt} = 4.16 + 2.392 * \text{Number}$$

Reasonable estimate of average weight of a Ziplock bag: Intercept of 4.16 grams. Reasonable estimate of average weight of a peanut M&M: Slope of 2.392 grams.

Activity Outcome: Students discover the importance of experiments.

The Importance of Randomization

Problem: Two varieties of corn are planted in 36 plots in the same field. Do Type A and Type B produce different corn yields?

Two assignments give different conclusions. Why? What kind of assignment should we use?

Convenience Assignment

A	A	A	B	B	B
130	149	139	155	137	145
A	A	A	B	B	B
149	133	152	131	147	136
A	A	A	B	B	B
141	156	137	146	132	148
A	A	A	B	B	B
150	142	155	136	152	133
A	A	A	B	B	B
139	155	139	147	137	153
A	A	A	B	B	B
155	138	150	137	145	136

Conclusion: The mean yield of Type A is slightly higher than Type B but the difference is not statistically significant.

Systematic Assignment

A	B	A	B	A	B
130	137	139	155	149	145
B	A	B	A	B	A
137	133	140	143	147	148
A	B	A	B	A	B
141	144	137	146	144	148
B	A	B	A	B	A
138	142	143	148	152	145
A	B	A	B	A	B
139	143	139	147	149	153
B	A	B	A	B	A
143	138	138	149	145	148

Conclusion: The mean yield of Type B is now slightly higher than Type A, but the difference is not statistically significant.

Solution using Manipulatives: Students make a random assignment of the two varieties to the 36 plots using a 6-sided die. Using the TRUTH, students conduct a hypothesis test to determine if Type A has a better corn yield than Type B.

THE TRUTH

A=130 B=118	A=149 B=137	A=139 B=127	A=167 B=155	A=149 B=137	A=157 B=145
A=149 B=137	A=133 B=121	A=152 B=140	A=143 B=131	A=159 B=147	A=148 B=136
A=141 B=129	A=156 B=144	A=137 B=125	A=158 B=146	A=144 B=132	A=160 B=148
A=150 B=138	A=142 B=130	A=155 B=143	A=148 B=136	A=164 B=152	A=145 B=133
A=139 B=127	A=155 B=143	A=139 B=127	A=159 B=147	A=149 B=137	A=165 B=153
A=155 B=143	A=138 B=126	A=150 B=138	A=149 B=137	A=157 B=145	A=148 B=136

Random Assignment

B	A	A	A	A	A
118	149	139	167	149	157
B	B	B	A	B	B
137	121	140	143	147	136
B	B	A	B	B	A
129	144	137	146	132	160
B	A	A	B	A	A
138	142	155	136	164	145
A	A	A	A	B	A
139	155	139	159	137	165
B	B	B	B	B	A
143	126	138	137	145	148

Under the random assignment, there is an overwhelming statistically significant difference in mean corn yields between Type A and B, with Type A having higher mean yields.

Activity Outcome: Students discover the importance of randomized experiments.

Solution using Simulation: What if another random assignment of the two varieties was made? Would we still find that Type A has a higher mean yield?

Use computer to simulate a random assignment and to conduct the hypothesis test. What do you find? Is your conclusion for this test different from the one above?

Use computer to repeat this simulation 100 times. What percentage of times do you conclude Type A has a higher mean yield? What if Type A was only 6 bushels more than Type B? Would you conclude Type A has a higher mean yield more or less often?

Activity Outcome: Students discover the Power of a Hypothesis Test.