Stat 102
Introduction to Business Statistics

Richard P. Waterman

Wharton

Table of contents I

Today's module

Topics to be covered in this module:

- A quick look at the syllabus
- The key skill from the course: regression analysis
- Review concepts from Stat 101
- Getting a feel for variability with M&Ms
- Summary
- Next time

Syllabus review

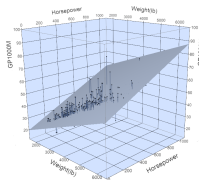The syllabus is on Canvas.
Main points:

- People
- Resources
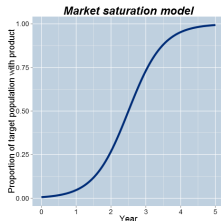- Deliverables
- Final grade
- Collaboration

Simple regression

Multiple regression

Logistic regression

An example of multiple regression

The PRiSM data set:

- Outcome variable: Loan performance (PRiSM)
- 30-40 predictor variables
- Model loan performance as a function of these variables
  1. Which variables to include (inference)?
  2. How much weight should they get (estimation)?
  3. Forecast loan performance and provide forecast intervals (prediction)
  4. Understand what's going on and critique the work of others

The data

Figure 1: PRiSM data set

Example predictive model

Figure 2: PRiSM data set

Prediction Expression

0.26228151510662

+ -0.0000007665499 * Commission Upfront

+ 0.00117891812411 * FICO

$$+ \text{Match} \left(\text{Loan Type} \right) \begin{cases} \text{"O"} \Rightarrow -0.0676969162723 \\ \text{"R"} \Rightarrow 0.06769691627228 \\ \text{else} \Rightarrow . \end{cases}$$

$$+ \text{Match} \left(\text{ISO Name} \right) \begin{cases} \text{"A"} & \Rightarrow 0.14463520535247 \\ \text{"B"} & \Rightarrow -0.0666076273058 \\ \text{"C"} & \Rightarrow 0.0778453122079 \\ \text{"D"} & \Rightarrow -0.0263422369985 \\ \text{"E"} & \Rightarrow 0.04482038500937 \\ \text{"F"} & \Rightarrow 0.05172774201135 \\ \text{"H"} & \Rightarrow -0.2034911415412 \\ \text{"OTHER"} & \Rightarrow -0.0225876387356 \\ \text{else} & \Rightarrow . \end{cases}$$

Figure 3: PRiSM data set

Actual by Predicted Plot
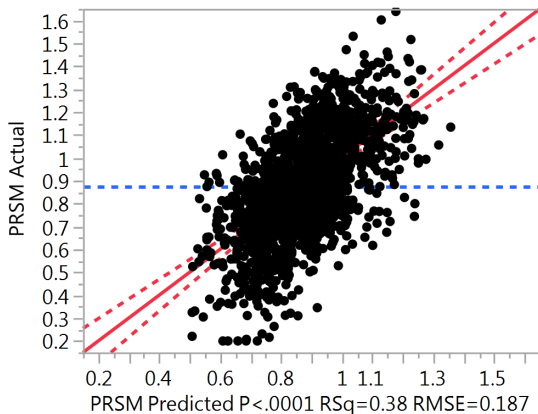
Ideas you should alreday be familiar with

- Summaries of categorical data
- Summaries of numerical data
- Association between categorical variables
- Association between quantitative variables (correlation)
- Probability models for counts: Bernoulli and Binomial
- The Normal probability model

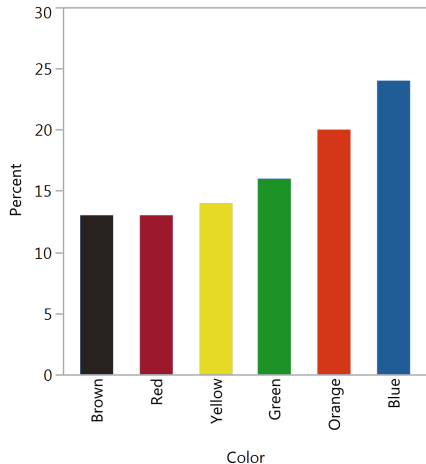# What's the deal with M&M colors?

There is a historical claim that M&Ms occur in color proportions:
Brown: 13% Red: 13% Yellow: 14% Green: 16% Orange: 20% Blue: 24%

Figure 5: M&M theoretical color percentages

Figure 6: M&M observed color frequency
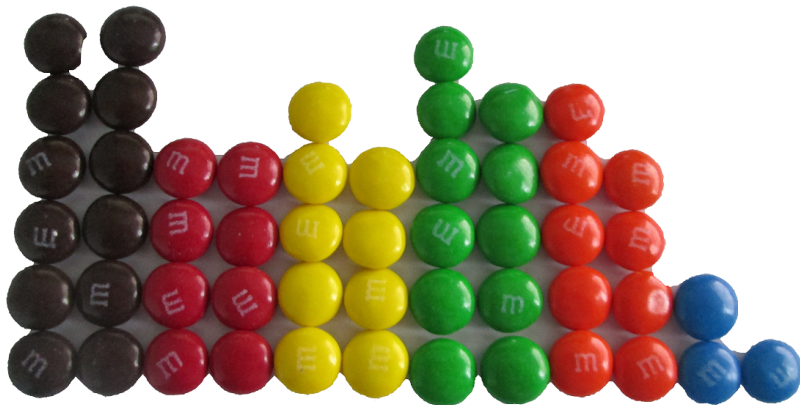
Collect your data

Team members:

_____

_____

_____

Table 1: M&M data collection

	Brown	Red	Yellow	Green	Orange	Blue	Total
Count							
Proportion							1.000

Collect your data

Team members:
Richard Waterman

Table 2: M&M data collection

	Brown	Red	Yellow	Green	Orange	Blue	Total
Count	10	8	9	11	9	3	50
Proportion	0.200	0.160	0.180	0.220	0.180	0.060	1.000

Module summary

Topics covered today include:

- Reviewed the syllabus
- Sample to sample variation
- Previewed JMP

Next time

- Standard error
- Central limit theorem