

Stat 102

Introduction to Business Statistics

Richard P. Waterman

Wharton

Table of contents I

- 1 Today's module
- 2 Course overview
 - A quick look at the syllabus
 - The main skill: regression analysis
- 3 Concepts from Stat 101
 - Key ideas
 - M&Ms and variability
- 4 Summary
- 5 Next time

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:

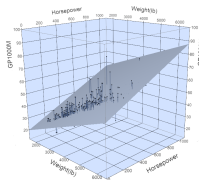
- People
- Resources
- Deliverables
- Final grade
- Collaboration

What you will be able to do by the end of the semester

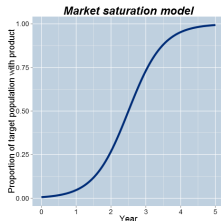
Simple regression



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

PRISM_2015 - JMP Pro														
File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help														
PRISM_2015														
Source														
Perfect model														
CaseNumber	Ant Repaid at 6 Months	Nominal Loan Amount	Total Amt to be Repaid	Repayment Percentage	Commission Upfront	Validated Monthly Batch	Historical Monthly Credit Card Receipts	Months of History	Loan Type	FICO	Years In Business	Num of Line		
1	2626	7421	20000	23800	12.5	2760	200654	21129	4 O	498	14			
2	2753	8560	14400	17856	12.5	21312	94716	9471.71429	7 O	633	10			
3	4348	7061	20000	24800	7.5	2960	2085444	20854	6 O	469	29			
4	391	10865	27700	33340	12.5	23268	2137694	2137694	4 O	487	2			
5	561	52922	60600	72417	10	3636	729385	96963.6362	11 O	689	2			
6	413	4429	14400	17856	12.5	31968	11357.56	11357.6364	11 O	404	5			
7	1526	40900	106000	131440	12.5	15688	7091534	709153	4 O	503	6			
8	4202	50653	122700	1453995	12.5	159694	10085964	100859	4 O	556	13			
9	399	194781	290000	342200	12.5	23664	2629746	234300	6 O	660	1			
10	2997	34575	45800	54731	12.5	1273.24	33718	46982	2 R	560	5			
11	3580	17107	27600	33396	12.5	1656	2668148	266815	4 O	506	4			
12	2356	21348	36000	44220	11	2520	2638044	263803333	6 O	442	2			
13	4094	4287	9900	11904	12.5	1420.8	611334	611334	3 O	602	2			
14	963	19719	46600	54988	12.5	5070.08	4543932	50738	5 O	561	12			
15	3890	5649	12300	15252	12.5	1820.4	809436	9449.33333	6 O	476	8			
16	287	3214	6000	7140	10	648.32	83729	8373.33333	3 O	526	10			
17	3552	13853	27000	32320	12.5	291596	216438	21644	3 O	702	1			
18	3317	4074	10600	12792	10	1716.26	11257.04	11256.9333	15 O	532	3			
19	2106	9000	10730	12730	12.5	1242	710582	71058	4 O	513	12			
20	153	226785	400000	470000	15	21600	4089205	438338	3 R	618	3			
21	3381	13736	26800	31860	12.5	36984	2104704	23507.3333	3 O	617	2			
22	328	22976	68000	84320	10	10064	6286202	65299	8 O	498	4			
23	454	7693	12600	15624	12.5	1491.84	10786	14142.6667	3 R	650	4			
24	4346	5941	10600	13144	12.5	1568.8	668602	6686	6 O	538	5			
25	277	20940	29400	35427	10	1764	2422062	24228.3096	7 O	546	18			
26	2680	24000	28860	34860	12.5	3666	18794466	18794574	7 O	541	5			
27	111	89628	172900	214396	12.5	255892	11372212	1137225	4 O	647	9			
28	3490	7002	11000	13475	12.5	11165	703438	7034	6 O	570	1			
29	3526	6580	12200	14518	10	1346.88	15594	20212	1 R	579	4			
30	1279	50111	156900	194556	12.5	23221.2	10321434	103214.333	6 O	536	5			
31	1493	6111	17800	22695	12.5	372464	1165546	14178.2857	7 O	573	1			
32	1257	3795	10500	12338	12.5	992.32	9450	14676	3 O	468	2			
33	1104	13052	22200	26085	12.5	1198.8	2148984	30881.3333	3 R	553	3			
34	4038	12000	14880	18060	12.5	23048	1364236	139255	4 R	528	12			
35	15	2107	3600	4284	12.5	198.70	5341.26	4304	1 R	648	11			
36	2483	10979	27000	31995	12.5	1664.56	2090164	14729.0367	4 O	510	6			
37	941	4790	10000	11800	12.5	700	898278	89825	4 O	518	2			

Example predictive model

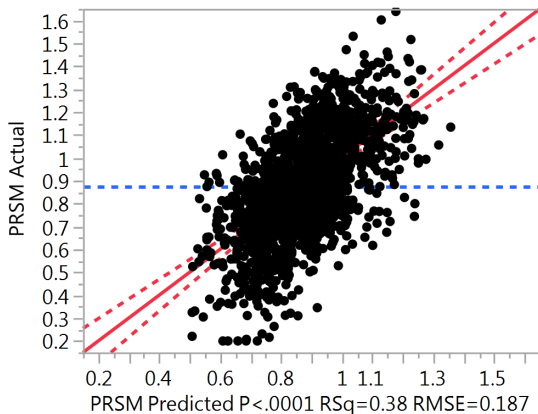
Figure 2: PRiSM data set

Prediction Expression

```
0.26228151510662
+ -0.0000007665499 * Commission Upfront
+ 0.00117891812411 * FICO
+ Match( Loan Type )  $\left\{ \begin{array}{l} \text{"O"} \Rightarrow -0.0676969162723 \\ \text{"R"} \Rightarrow 0.06769691627228 \\ \text{else} \Rightarrow . \end{array} \right.$ 
+ Match( ISO Name )  $\left\{ \begin{array}{l} \text{"A"} \Rightarrow 0.14463520535247 \\ \text{"B"} \Rightarrow -0.0666076273058 \\ \text{"C"} \Rightarrow 0.0778453122079 \\ \text{"D"} \Rightarrow -0.0263422369985 \\ \text{"E"} \Rightarrow 0.04482038500937 \\ \text{"F"} \Rightarrow 0.05172774201135 \\ \text{"H"} \Rightarrow -0.2034911415412 \\ \text{"OTHER"} \Rightarrow -0.0225876387356 \\ \text{else} \Rightarrow . \end{array} \right.$ 
```


Figure 3: PRiSM data set

Actual by Predicted Plot



Ideas you should already 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

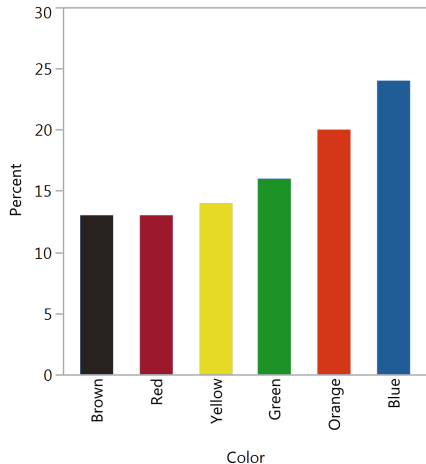
What's the deal with M&M colors?

Figure 4: M&M packet



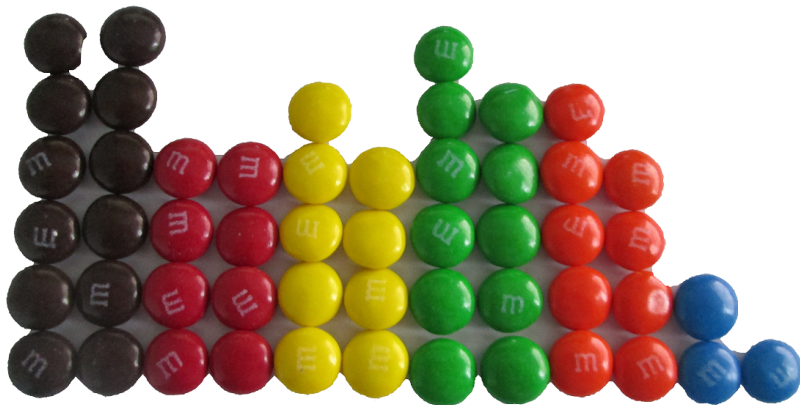
There is a historical claim that M&Ms occur in color proportions:
Brown: 13% Red: 13% Yellow: 14% Green: 16% Orange: 20% Blue: 24%

Figure 5: M&M theoretical color percentages



My observed distribution

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