



Skittles and M&M's

A statistics-based approach

Racquel Bozzelli Ian Burgess Antonio Jurlina
ECO 485 INTRODUCTORY ECONOMETRICS (DR. KEITH EVANS)

Part 1

This empirical project is centered on the comparison of color proportions between two different candies: M&M's® and Skittles®. Both are manufactured under the Mars, Inc. umbrella of products and each candy brand features a multitude of variations of the original product. For the purposes of this project, smallest available packages (publicly marketed as “Fun Size”) were used, comparing specifically milk chocolate variety of M&M's® with the “original” variety of Skittles®. Data obtained details the total number of button-shaped candies per bag as well as the specific number of each color of candy present per bag. With a fairly representative sample size for each candy distribution (over 90 bags sampled), subsequent statistical comparisons were performed utilizing data analysis and statistical software, Stata®.

For basic descriptive purposes, a quick summary was rendered employing the aforementioned software, providing mean color representations across two candies (blue, green, brown, orange, red, and yellow for M&M's® and green, orange, red, purple, and yellow for Skittles®) and the mean value for bag totals across the candies. Furthermore, new variables were generated across all observations for the purpose of subsequent statistical inference, specifically with regard to colors shared among the two candies (green, orange, red, and yellow). These variables are the proportions of each color shared by the two candies, for each individual packaging (greenprop, orangeprop, redprop and yellowprop, as named in Stata). These proportions allow for the focal point of the project, the comparison between proportion means across the two candies, to gauge the probability (if any) of observing differing color distributions, between M&M's® and Skittles®, on average.

Comparing means across two different groups implies a two-sample t-test, specifically between means of the newly generated color proportion variables, to determine if the difference between the two groups is actually significant or due to random influence. Since a generic Stata® two-

sample t-test will automatically yield a 95% confidence level that is the one that shall be reported in this analysis. The basic premise, as outlined in the previous paragraph, searches for any significant difference between color distributions, implying the following hypothesis rationale:

H_0 (null): greenprop (M&M's) = greenprop (Skittles)

H_1 (alternate): greenprop (M&M's) \neq greenprop (Skittles)

(Same hypothesis rationale used for remaining proportion-related variables – orangeprop, yellowprop and redprop)

All related P-values shall, therefore, be tested against a 5% probability of making a type I error (observing equal proportions and rejecting the null hypothesis). Furthermore, the t-test shall be two-tailed since both positive and negative discrepancies between the means serve as acceptable evidence for rejecting the null hypothesis.

Part 2

In studying the results, several key comparisons in color distributions between M&M's[®] and Skittles[®] can be made. Basic summary, showing average color proportions, indicates higher percentage of yellow in Skittles[®] (19%) versus M&M's[®] (14%), higher percentage of red in Skittles[®] (22%) versus M&M's[®] (18%), higher percentage of orange in M&M's[®] (20%) versus Skittles[®] (18%), and higher percentage of green in Skittles[®] (21%) versus M&M's[®] (18%). Simple logic dictates that with 6 different colors represented in M&M's[®] and 5 represented in Skittles[®], most bags of M&M's will contain lower frequency of shared colors and this should be reflected in overall sample color distribution (see Figures 1 and 2).

More in-depth analysis, as described in Part 1, utilizes a two-sample, two-tailed t-test in comparing mean color proportion values across shared colors. Now all we need to do is look at the results of

our t-tests in Stata. Specifically, at the 95% confidence interval of the difference in proportions. If 0 is not included in the confidence interval, we can reject the null. The result summary is as follows (for more detailed summary, see Figure 3):

- For red, the interval is [-0.076, -0.012]. Zero is not in this interval, so we reject the null and conclude there is a difference in the proportions at a 95% confidence level.
- For orange, the interval is [-0.015, 0.046]. Zero is in this interval, so we fail to reject the null and conclude there is no difference in the proportions at a 95% confidence level.
- For yellow, the interval is [-0.092, -0.032]. Zero is not in this interval, so we reject the null and conclude there is a difference in the proportions at a 95% confidence level.
- For green, the interval is [-0.055, 0.008]. Zero is in this interval, so we fail to reject the null and conclude there is no difference in the proportions at a 95% confidence level.

Simply put, at a 95% confidence level, we conclude that distribution of red and yellow between the two candies (assuming same Fun Size packaging) is different at a statistically significant level. This can also be confirmed with Figure 4 (Stata log), by analyzing data distribution probabilities calculated in Stata ($\Pr(|T| > |t|)$ for $H_a: \text{diff} \neq 0$), which show that the probabilities that yellow and red are equally represented among candies when sampling at random are less than 5%. Same tests did not indicate a statistically significant difference among distribution of green and orange between the two candies (i.e. they were greater than 5%) (see Figure 3).

When contemplating a cause for distribution discrepancies, various hypotheses come to mind. Assuming that factories produce preset percentages (as noted and inquired by Madison^[1] and Wicklin^[2]), cost optimizing, bulk-production logic implies an aggregated production run satisfying all the percentages that are subsequently sequestered according to different package specifications yielding individual bags varying greatly from produced percentages that converge

on aggregate. Factory specifications further imply varying costs between two candies as they are subject to cost minimization calculations - M&M's® use more colors but one flavor and Skittles® vary flavors according to color, each attribute having a specific cost. These intrinsic differences are potentially subject to a myriad of factors. Factories located in different regions utilize different proportions (as noted by Wicklin^[2]), and if Mars, Inc. corporate set of instructions allows for slight freedom of choice for factory managers determining proportions, the list of potential conscious and subconscious influences grows significantly.

A reasonable approach would be to assume consumer preferences in Skittles somewhat dictate flavor distribution. This allows for survey-based testing to gauge for preference proportions among a representative audience. We could then test the results of the survey to see if there is a true difference in the preferences. This could be done with another two-tailed t-test gathering proportion preference means that are directly comparable to factory-set production proportions. A strong similarity between consumer flavor preferences and packaging proportions would indicate the existence of a causal link that is, by design, absent from the consumer-M&M's relationship and it is, therefore, a unique source of variation on one half of the project spectrum.

(Word Count: 1105)

References:

1. *December 2, 2007 · Posted To Projects. (n.d.). M&Ms Color Distribution Analysis. Retrieved March 01, 2018, from <https://joshmadison.com/2007/12/02/mms-color-distribution-analysis/>*
2. *Purtill, C. (2017, March 15). A statistician got curious about M&M colors and went on an endearingly geeky quest for answers. Retrieved March 01, 2018, from <https://qz.com/918008/the-color-distribution-of-mms-as-determined-by-a-phd-in-statistics/>*

Cover Image:

3. *Taste The Rainbow [Vapor Junkies take on a world famous multicolored sweet candy.]. (n.d.). Retrieved March 01, 2018, from <https://www.vaporjunkies.com/shop/consumables/taste-the-rainbow/>*

Figure 1.

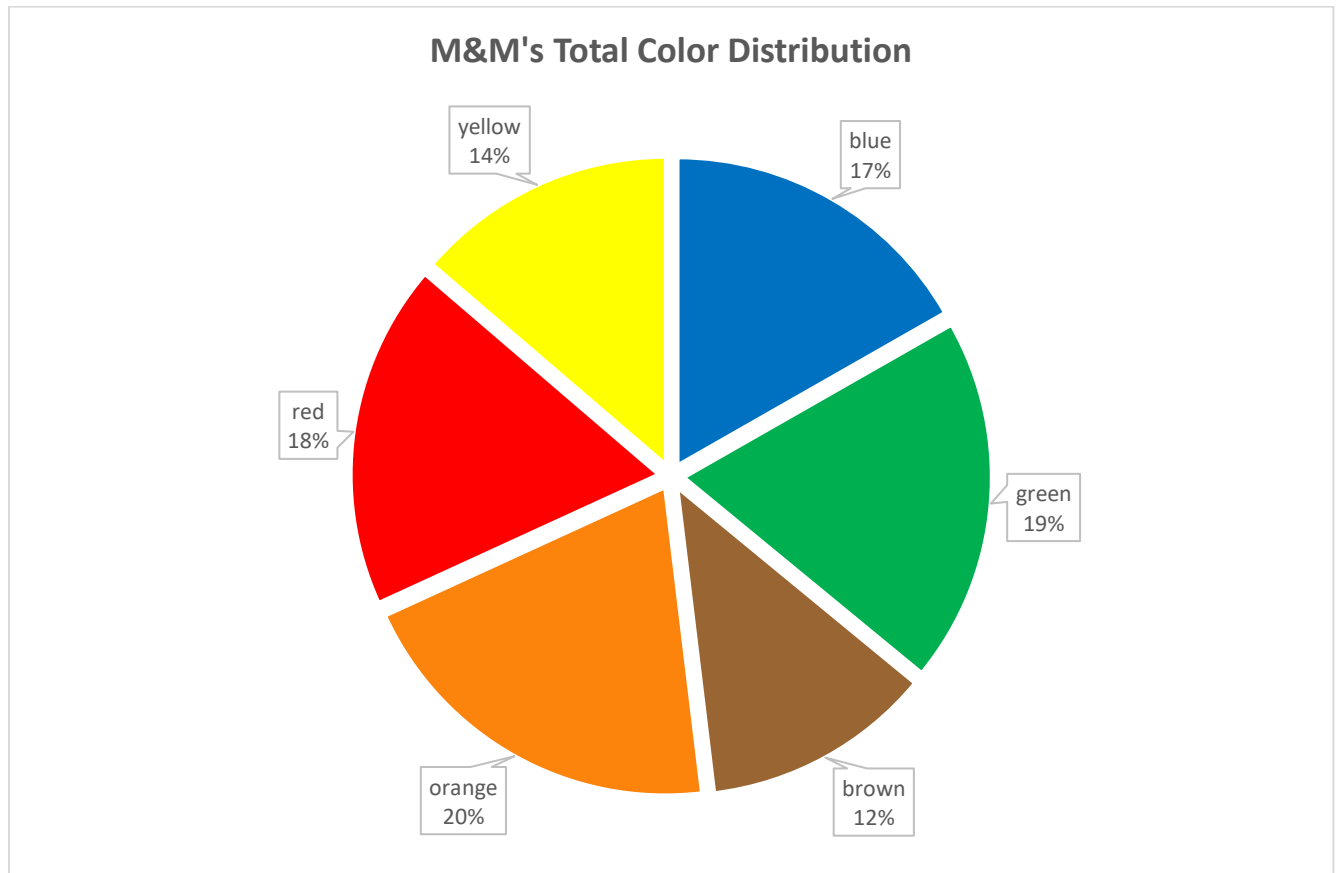


Figure 2.

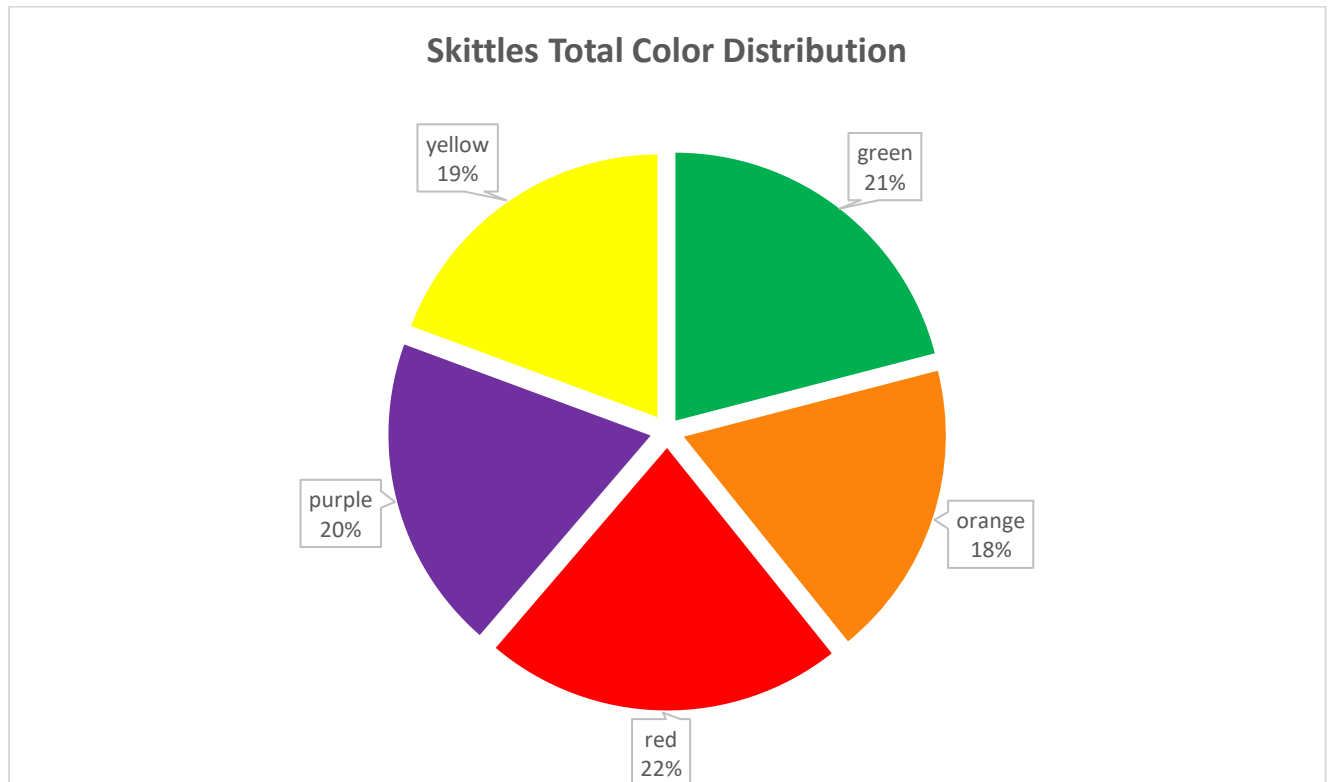
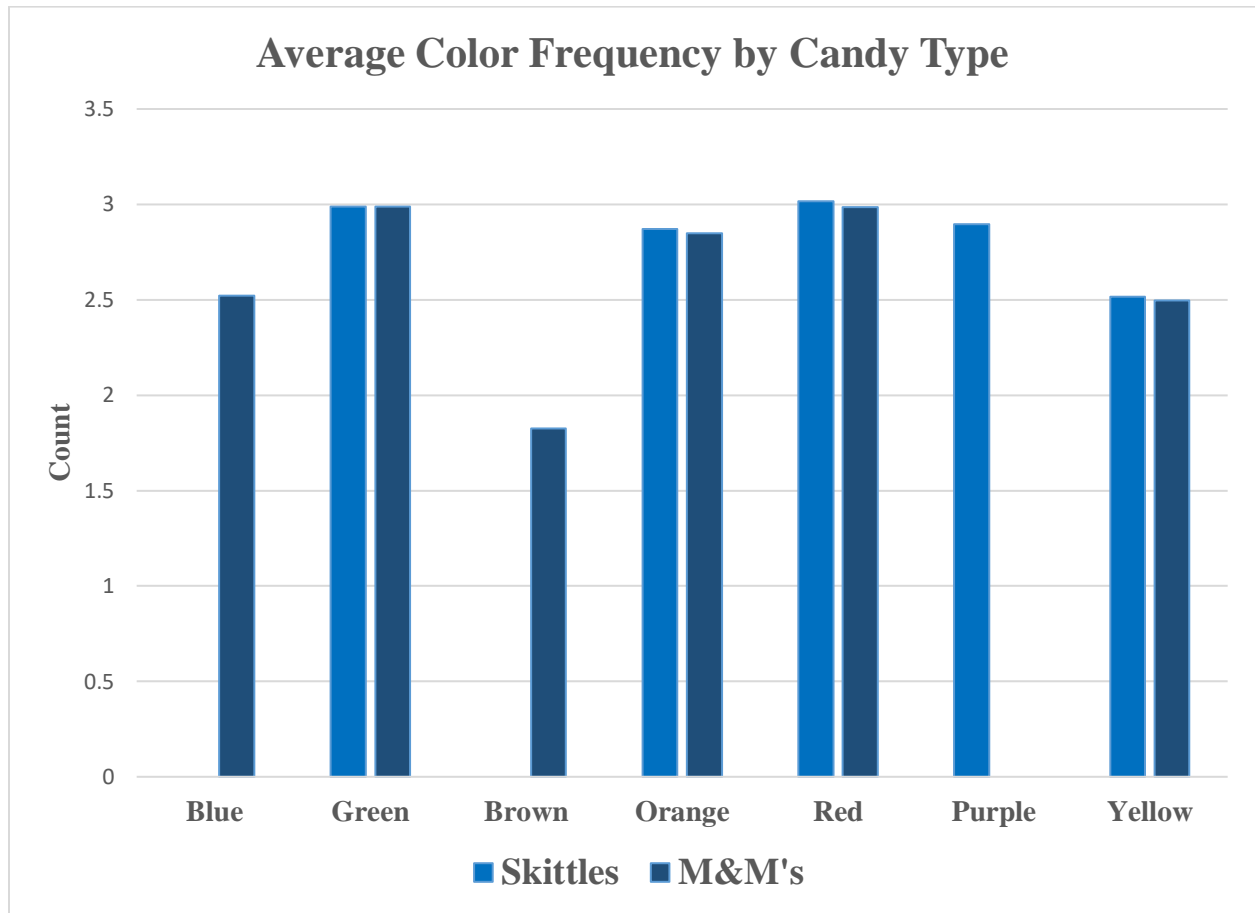


Figure 3.



Stata Log File (Figure 4).

```

-----
name: <unnamed>
log: D:\UMaine\Spring (2018)\ECO 485\Stata\Empirical_Project_1.log
log type: text
opened on: 1 Mar 2018, 20:47:33

.
. use candy_data.dta //choose my data

.
. summarize // basic data summary

```

Variable	Obs	Mean	Std. Dev.	Min	Max
id	198	99.5	57.30183	1	198
blue	92	2.521739	1.544065	0	6
green	198	3.005051	1.565807	0	7
brown	92	1.826087	1.434341	0	7
orange	198	2.853535	1.622733	0	8
red	198	3.015152	1.726113	0	8
purple	105	2.895238	1.549962	0	6
yellow	198	2.494949	1.556866	0	8
total	198	14.81818	1.438158	8	19
candy	198	1.535354	.5000128	1	2
flavor	0				

```

.
. sort candy //sort data by Skittles and M&M's

.
. //compare colors between Skittles and M&M's
. by candy: summarize blue green brown orange red purple yellow total

```

```

-----
-> candy = M&M's

```

Variable	Obs	Mean	Std. Dev.	Min	Max
blue	92	2.521739	1.544065	0	6
green	92	2.891304	1.633822	0	7
brown	92	1.826087	1.434341	0	7
orange	92	3.021739	1.690167	0	7
red	92	2.728261	1.664782	0	7
purple	0				
yellow	92	2.065217	1.532106	0	7
total	92	15.05435	.9985058	13	17

```

-----
-> candy = Skittles

```

Variable	Obs	Mean	Std. Dev.	Min	Max
blue	0				
green	106	3.103774	1.505098	1	7
brown	0				
orange	106	2.707547	1.555127	0	8
red	106	3.264151	1.747389	0	8
purple	105	2.895238	1.549962	0	6
yellow	106	2.867925	1.48693	0	8
total	106	14.61321	1.710309	8	19

```

.
. //creating variables for testing
. generate greenprop1 = green/total

. generate orangeprop = orange/total

. generate redprop = red/total

. generate yellowprop = yellow/total

.
. //ttests comparing color distributons between Skittles and M&M's
. ttest greenprop, by(candy)

Two-sample t test with equal variances
-----
      Group |      Obs      Mean   Std. Err.   Std. Dev.   [95% Conf. Interval]
-----+-----
      M&M's |       92   .1923044   .0114939   .1102454   .1694733   .2151356
    Skittles |      106   .2160217   .0111061   .1143444   .1940004   .2380431
-----+-----
    combined |      198   .2050016   .0080163   .1127992   .1891928   .2208103
-----+-----
      diff |           -.0237173   .0160245           -.0553197   .0078852
-----+-----
      diff = mean(M&M's) - mean(Skittles)                                t =  -1.4801
Ho: diff = 0                                                              degrees of freedom =    196

      Ha: diff < 0                      Ha: diff != 0                      Ha: diff > 0
Pr(T < t) = 0.0702                  Pr(|T| > |t|) = 0.1405                  Pr(T > t) = 0.9298

. ttest orangeprop, by(candy)

Two-sample t test with equal variances
-----
      Group |      Obs      Mean   Std. Err.   Std. Dev.   [95% Conf. Interval]
-----+-----
      M&M's |       92   .2006325   .0116419   .1116653   .1775072   .2237577
    Skittles |      106   .1845953   .0105643   .1087658   .1636483   .2055423
-----+-----
    combined |      198   .1920469   .0078269   .110134   .1766117   .2074821
-----+-----
      diff |           .0160372   .0156913           -.0149082   .0469826
-----+-----
      diff = mean(M&M's) - mean(Skittles)                                t =    1.0220
Ho: diff = 0                                                              degrees of freedom =    196

      Ha: diff < 0                      Ha: diff != 0                      Ha: diff > 0
Pr(T < t) = 0.8460                  Pr(|T| > |t|) = 0.3080                  Pr(T > t) = 0.1540

. ttest redprop, by(candy)

Two-sample t test with equal variances
-----
      Group |      Obs      Mean   Std. Err.   Std. Dev.   [95% Conf. Interval]
-----+-----
      M&M's |       92   .1806438   .0112816   .1082094   .1582342   .2030533
    Skittles |      106   .2247367   .0116937   .1203937   .2015503   .2479231
-----+-----
    combined |      198   .2042491   .0082294   .1167069   .1878927   .2206055
-----+-----
      diff |           -.0440929   .0163718           -.0763804   -.0118054
-----+-----
      diff = mean(M&M's) - mean(Skittles)                                t =   -2.6932
Ho: diff = 0                                                              degrees of freedom =    196

      Ha: diff < 0                      Ha: diff != 0                      Ha: diff > 0
Pr(T < t) = 0.0038                  Pr(|T| > |t|) = 0.0077                  Pr(T > t) = 0.9962

. ttest yellowprop, by(candy)

```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
M&M's	92	.1371195	.010687	.1025057	.1158912	.1583478
Skittles	106	.1991361	.0106112	.1092491	.1780961	.2201762
combined	198	.1703203	.0078421	.1103482	.1548551	.1857856
diff		-.0620166	.0151284		-.091852	-.0321812
diff = mean(M&M's) - mean(Skittles)				t = -4.0993		
Ho: diff = 0				degrees of freedom = 196		
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.0000		Pr(T > t) = 0.0001		Pr(T > t) = 1.0000		

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
M&M's	92	15.05435	.1041014	.9985058	14.84756	15.26113
Skittles	106	14.61321	.1661199	1.710309	14.28382	14.94259
combined	198	14.81818	.1022054	1.438158	14.61662	15.01974
diff		.4411403	.203015		.040766	.8415146
diff = mean(M&M's) - mean(Skittles)						t = 2.1729
Ho: diff = 0				degrees of freedom = 196		
Ha: diff < 0		Ha: diff != 0			Ha: diff > 0	
Pr(T < t) = 0.9845		Pr(T > t) = 0.0310			Pr(T > t) = 0.0155	

```
. log close //close the log
      name: <unnamed>
      log: D:\UMaine\Spring (2018)\ECO 485\Stata\Empirical_Project_1.log
      log type: text
closed on: 1 Mar 2018, 20:47:33
```

Stata Do-File.

```
1  . /******  
2  > @ Authors: Antonio J, Ian B. Racquel B. @  
3  > @ Date: 2/20/2018 @  
4  > @ Filename: EMPIRICAL_PROJECT_1.do @  
5  > *****/  
6  
7  clear //clear previous data  
8  
9  cd "D:\UMaine\Spring (2018)\ECO 485\Stata" //setting my directory  
10  
11 // Store results in a log file (diary)  
12 log using "D:\UMaine\Spring (2018)\ECO 485\Stata\Empirical_Project_1", replace text  
13  
14 use candy_data.dta //choose my data  
15  
16 summarize // basic data summary  
17  
18 sort candy //sort data by Skittles and M&M's  
19  
20 //compare colors between Skittles and M&M's  
21 by candy: summarize blue green brown orange red purple yellow total  
22  
23 //creating variables for testing  
24 generate greenprop1 = green/total  
25 generate orangeprop = orange/total  
26 generate redprop = red/total  
27 generate yellowprop = yellow/total  
28  
29 //ttests comparing color distributons between Skittles and M&M's  
30 ttest greenprop, by(candy)  
31 ttest orangeprop, by(candy)  
32 ttest redprop, by(candy)  
33 ttest yellowprop, by(candy)  
34 ttest total, by(candy)  
35  
36 log close //close the log  
37
```

Sample Data.

id	blue	green	brown	orange	red	purple	yellow	total	candy
62	2	4	1	3	3		2	15	M&M's
136	6	1	0	2	4		2	15	M&M's
121	0	6	0	4	1		3	14	M&M's
9	1	4	2	6	3		1	17	M&M's
110	1	2	3	6	2		1	15	M&M's
67	0	2	3	3	4		3	15	M&M's
25	0	5	3	1	4		2	15	M&M's
155	1	3	2	6	1		2	15	M&M's
69	3	5	1	2	5		1	17	M&M's
164	2	3	1	3	5		1	15	M&M's
169	1	2	1	6	3		2	15	M&M's
181	0	2	4	2	4		2	14	M&M's
102	3	3	1	3	1		4	15	M&M's
8	3	2	0	2	2		4	13	M&M's
1	4	3	2	0	5		1	15	M&M's
50	2	6	2	3	1		1	15	M&M's
70	2	1	3	4	3		2	15	M&M's
4	3	1	1	2	7		2	16	M&M's
124	3	1	2	3	3		4	16	M&M's
113	5	2	0	3	3		2	15	M&M's
109	2	4	3	1	3		2	15	M&M's
86	0	3	3	3	5		0	14	M&M's
59	3	5	3	3	1		2	17	M&M's
81	4	0	2	1	4		2	13	M&M's
76	2	2	2	3	1		4	14	M&M's
128	2	1	4	3	4		1	15	M&M's
51	1	3	4	5	0		3	16	M&M's
37	2	2	3	3	2		3	15	M&M's
2	3	6	2	2	1		0	14	M&M's
139	3	2	0	7	1		3	16	M&M's
34	4	3	0	4	3		1	15	M&M's
133	2	3	4	2	4		1	16	M&M's
95	6	2	1	3	2		0	14	M&M's
152	2	5	2	3	3		2	17	M&M's
27	2	3	1	2	4		3	15	M&M's
166	5	2	2	0	1		5	15	M&M's
46	1	1	0	4	4		6	16	M&M's
107	1	2	4	3	3		1	14	M&M's
159	1	2	2	3	7		1	16	M&M's
49	5	1	0	6	0		2	14	M&M's

115	6	2	3	5	1	0	17	M&M's
130	0	3	2	4	3	3	15	M&M's
105	2	4	0	6	2	1	15	M&M's
90	1	3	7	2	2	1	16	M&M's
187	3	3	1	3	1	2	13	M&M's
148	2	4	2	2	2	4	16	M&M's
163	3	2	1	3	3	3	15	M&M's
186	2	1	3	4	2	3	15	M&M's
161	6	2	2	1	0	4	15	M&M's
60	4	0	1	6	0	2	13	M&M's
72	3	2	1	2	3	3	14	M&M's
150	3	5	2	0	4	1	15	M&M's
30	2	6	0	4	1	0	13	M&M's
191	4	1	0	2	2	5	14	M&M's
97	1	4	1	5	2	2	15	M&M's
188	3	2	0	3	0	7	15	M&M's
165	4	5	2	1	3	1	16	M&M's
68	4	2	1	5	2	1	15	M&M's
84	3	4	3	4	0	2	16	M&M's
75	2	6	1	4	2	0	15	M&M's
158	3	7	0	5	0	1	16	M&M's
127	1	4	1	4	4	2	16	M&M's
39	3	2	2	4	3	1	15	M&M's
73	3	1	7	3	0	0	14	M&M's
151	2	0	0	6	7	2	17	M&M's
12	5	2	2	1	3	1	14	M&M's
174	2	2	1	3	4	3	15	M&M's
32	3	2	2	0	4	4	15	M&M's
80	1	1	4	5	2	3	16	M&M's
52	0	4	2	1	4	4	15	M&M's
114	5	1	1	2	2	4	15	M&M's
41	1	2	0	4	6	2	15	M&M's
180	4	4	5	2	1	1	17	M&M's
14	3	3	2	3	3	1	15	M&M's
82	2	3	2	1	3	5	16	M&M's
16	2	2	2	2	2	5	15	M&M's
162	1	6	1	1	5	0	14	M&M's
79	2	1	2	6	2	2	15	M&M's
143	0	3	2	5	2	3	15	M&M's
98	3	2	2	3	1	3	14	M&M's
108	1	4	1	2	2	5	15	M&M's
19	6	2	0	2	4	0	14	M&M's
94	2	4	2	2	5	1	16	M&M's

106	3	4	3	1	3		3	17	M&M's
126	3	3	1	6	2		0	15	M&M's
147	0	6	2	1	4		1	14	M&M's
92	4	0	4	2	4		2	16	M&M's
57	3	5	2	3	1		0	14	M&M's
185	4	5	1	1	4		1	16	M&M's
104	3	3	0	2	6		1	15	M&M's
120	3	5	2	0	3		0	13	M&M's
26	3	2	3	4	3		0	15	M&M's
93		5		2	2	1	3	13	Skittles
40		1		2	3	6	3	15	Skittles
96		2		4	0	6	2	14	Skittles
167		4		3	2	4	3	16	Skittles
154		4		1	3	5	3	16	Skittles
36		2		3	8	1	3	17	Skittles
20		5		1	5	4	0	15	Skittles
137		4		3	2		2	11	Skittles
138		7		2	5	5	0	19	Skittles
170		1		2	3	4	4	14	Skittles
177		3		2	2	5	3	15	Skittles
101		3		2	2	5	3	15	Skittles
22		3		4	7	3	1	18	Skittles
17		5		0	5	3	2	15	Skittles
160		3		3	2	3	4	15	Skittles
173		1		3	2	3	6	15	Skittles
142		1		6	6	2	1	16	Skittles
119		3		4	4	4	1	16	Skittles
56		3		1	7	2	2	15	Skittles
45		4		2	3	4	2	15	Skittles
100		2		4	2	1	6	15	Skittles
194		3		2	6	2	3	16	Skittles
197		4		4	2	6	2	18	Skittles
195		3		1	6	2	2	14	Skittles
179		2		0	5	1	5	11	Skittles
141		2		4	4	1	4	15	Skittles
175		1		1	2	5	6	15	Skittles
53		2		3	2	5	3	15	Skittles
15		3		6	0	3	3	15	Skittles
61		7		1	2	1	3	14	Skittles
135		6		3	2	1	2	14	Skittles
123		3		4	4	2	4	17	Skittles
63		1		3	5	4	1	14	Skittles
71		2		4	2	3	3	14	Skittles

85	3	1	2	3	5	14	Skittles
157	2	2	2	4	4	14	Skittles
48	1	6	4	3	2	16	Skittles
146	2	2	3	5	3	15	Skittles
77	1	4	5	3	1	14	Skittles
111	5	3	1	3	2	14	Skittles
129	2	2	2	3	5	14	Skittles
65	2	2	7	3	2	16	Skittles
91	2	4	6	3	1	16	Skittles
13	2	2	5	2	4	15	Skittles
192	5	3	5	1	2	16	Skittles
28	5	1	4	2	4	16	Skittles
11	3	2	3	2	3	13	Skittles
112	2	3	3	4	2	14	Skittles
18	2	4	4	1	3	14	Skittles
176	3	3	2	5	3	16	Skittles
125	3	3	3	3	3	15	Skittles
116	2	1	2	5	0	10	Skittles
87	3	3	0	5	4	15	Skittles
183	4	3	1	4	3	15	Skittles
42	4	2	5	1	3	11	Skittles
153	3	3	3	4	2	15	Skittles
168	3	7	2	2	0	11	Skittles
132	2	3	5	3	2	15	Skittles
54	1	4	5	0	4	14	Skittles
196	2	6	1	5	2	16	Skittles
44	2	2	0	6	1	11	Skittles
144	1	8	3	0	2	14	Skittles
88	2	3	4	4	4	17	Skittles
66	3	1	4	4	2	14	Skittles
89	2	3	4	2	5	16	Skittles
29	2	4	5	1	2	12	Skittles
43	1	1	4	1	6	13	Skittles
190	6	1	1	5	2	15	Skittles
55	7	3	1	1	2	14	Skittles
131	3	3	4	2	3	15	Skittles
171	3	0	5	5	2	15	Skittles
47	5	1	1	4	3	14	Skittles
189	1	1	2	2	6	12	Skittles
10	2	2	5	3	2	14	Skittles
117	4	3	2	3	4	16	Skittles
99	5	0	3	5	3	16	Skittles
156	4	3	3	1	3	14	Skittles

6	3	6	2	3	2	16	Skittles
7	4	4	3	1	2	14	Skittles
58	5	5	3	0	1	14	Skittles
182	7	3	1	3	2	16	Skittles
118	3	3	6	1	1	14	Skittles
134	4	4	1	3	2	14	Skittles
64	4	1	2	3	5	15	Skittles
78	3	5	2	3	2	15	Skittles
172	5	1	2	1	5	14	Skittles
5	3	4	3	4	3	17	Skittles
74	6	0	4	2	2	8	Skittles
103	3	3	2	4	4	16	Skittles
145	4	1	3	1	5	10	Skittles
31	4	2	2	2	4	14	Skittles
140	3	2	4	4	4	17	Skittles
83	3	4	2	4	3	16	Skittles
122	3	4	4	3	1	15	Skittles
24	2	3	5	2	4	16	Skittles
23	3	1	2	4	4	14	Skittles
35	2	1	8	2	1	14	Skittles
3	1	1	3	4	4	13	Skittles
193	4	2	4	1	3	14	Skittles
149	3	3	5	3	1	15	Skittles
184	4	4	2	1	3	14	Skittles
38	2	2	4	5	2	15	Skittles
33	6	3	5	1	2	17	Skittles
198	3	3	3	2	3	14	Skittles
21	4	3	2	1	5	15	Skittles
178	1	1	4	1	8	15	Skittles