

1. Buy a packet of regular size M&M's. Records the frequency of each color and then transfer this information to a bar chart and a pie chart.
2. Represent the number of M&M's per packet on a histogram using all of the class data. Is your histogram normally distributed? Check using the empirical rule. Also choose your favorite color and draw a histogram of the number of M&M's of that color per packet based on class data.
3. There are numerous hypothesis test that can be done. Obtain the mean number of M&M's per packet for the whole class.
Consider as a null hypothesis that the actual numbers of M&M's in your packet is the actual number of M&M's per packet as intended by the manufacturer. Test this hypothesis using the sample data from the class.
Comment on the difference between a type I and type II error in this context.
(NOTE: Be careful in the choice of your sample size)
4. Another obvious test would be to test the proportion of colors present. Intuitively, we might expect each color to occur with the same probability p (Null Hypothesis). Test this hypothesis by means of the Pearson Chi-squared test, using
 - (i) the color distribution of M&M's in your first packet as sample data (Note in this case the expected number of M&M's per packet will be np , where n is the number in your packet)
 - (ii) the complete color distribution of M&M's as found in all of the class data. (Note in this case the expected number of M&M's per packet will be np , where n is the total number of M&M's for the class.)
5. This test can be further modified. Do you think that our intuitive assumption is reasonable? Maybe, every color was not meant to be in the same proportion. To find a reasonable set of proportions that would correspond to the true colors, calculate a posterior probability for **each color** based on your assumptions in part (4) by basing your model on 11 increments given by: 0, .1, .2, .3, .4, .5, .6, .7, .8, .9, 1. (Note this is Bayesian statistics question).
6. Repeat the same question but now use beta function for each color.
7. Write out a table of multinomial "prior" and "posterior" probabilities that you would obtain by adding the successive information for each packet to update your "prior" probabilities.
8. Now repeat the Pearson Chi-squared test based on the assumption that the posterior probabilities are in fact true.
9. The Mars company which makes M&M's claim that "candies contain 30% browns, 20% each of yellows and red and 10% each of oranges, greens and blues". Test this hypothesis using all the class data. Is this substantially different to the

null hypothesis in (7) above?

10. Find the confidence intervals at the 90% and 95% confidence levels for the estimate of the true mean number of M&M's in a packet using.
 - (a) all of the class data
 - (b) the date from your packet only (if it makes sense).
11. Form a contingency table of all the colors from each packet and test to see if the color distribution is independent of whatever packet there are in.
12. Formulate a null hypothesis for the value of the standard deviation in the mean number of M&M's per packet. Can you think of an appropriate test to see if the hypothesis is true? Do (Hint: Calculate the hypothesis based on the assumption that you have a multinomial distribution and use the information from question 9)

