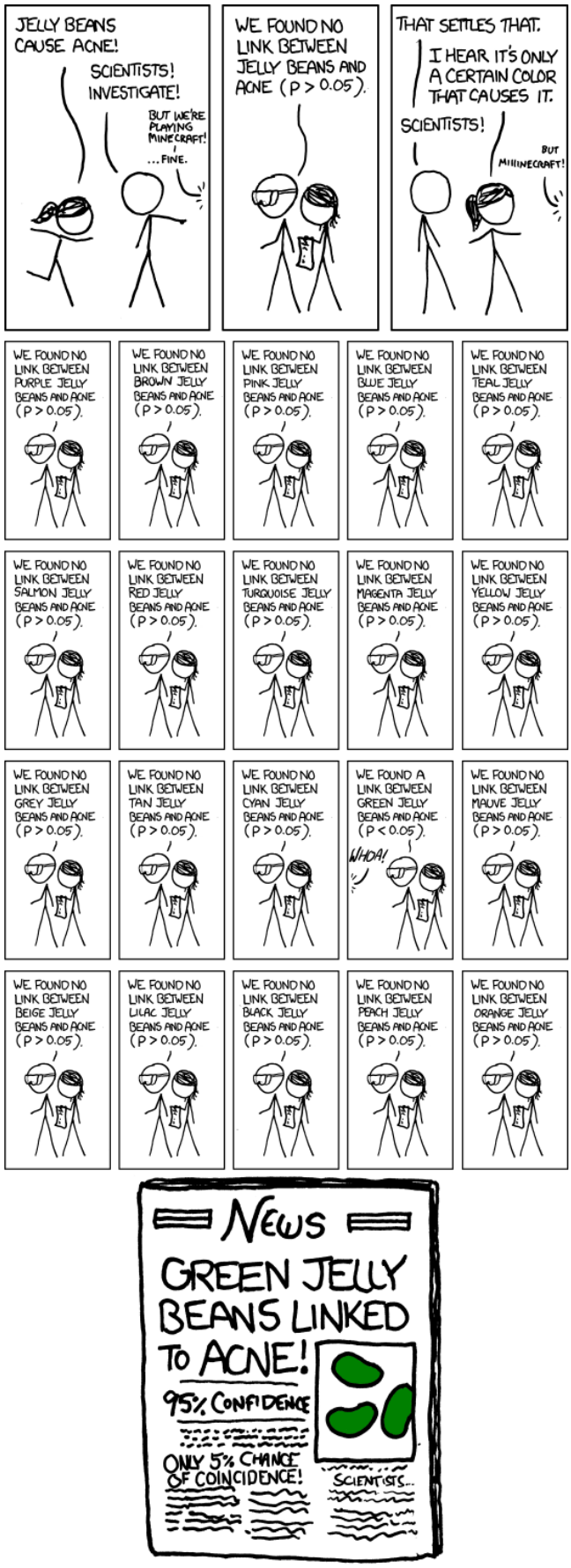
# Problem sheet 5 :: Chi square



Have you ever wondered why the package of M&Ms you bought never seems to have enough of your favorite color? Or, why is it that you always seem to get the package of mostly brown M&Ms? What's going on at the Mars Company? Is the number of the different colors of M&Ms in a package really different from one package to the next, or does the Mars Company do something to ensure that each package gets the correct number of each color of M&M?

Mars company claims to produce the M&Ms peanut chocolate candies in the following proportions: Brown (30%), Red (20%), Blue (10%), Orange (10%), Green (10%), Yellow (20%). To determine if the company's claim is true, we need to sample a package of M&Ms and do a type of statistical analysis known as "goodness of fit" test. This type of statistical test allows us to determine if any differences between our observed measurements (counts of colors from our M&M sample) and our expected (what the Mars Co. claims) are simply due to chance or some other reason (i.e. the Mars company's sorters aren't doing a very good job of putting the correct number of M&M's in each package).

You set out to investigate if this is true using a goodness of fit test, more particularly a Chi Square (X2) analysis. You buy a large bag of M&Ms and systematically count how much of each color you have: Brown (58), Red (40), Blue (34), Orange (22), Green (30), Yellow (36).

## Questions

Four questions, each worth two marks with two marks for attendance.

Q1. Fill in the 4th column (see table below) which describes the amount of M&Ms you should have expected to find in your bag of 220 M&Ms.

Q2. Fill in the 5th column (see table below) using the formula provided at the end of this document and compute the $X^2$ value.

Q3. Using the CHI square distribution table provided at the end of this document and a significance level of 0.05, calculate the P-value for this test and explain whether you should you accept or reject the null hypothesis and conclude.

Q4. Using the CHI square distribution table provided at the end of this document and a significance level of 0.01, calculate the P-value for this test and explain whether you should you accept or reject the null hypothesis and conclude.

| Color | Official | Observed by counting a bag of M&Ms | Expected | $X^2$ |
|---|---|---|---|---|
| Brown | .30 | 58 | 0.30 * 220= 66 | 0.9696 |
| Red | .20 | 40 | 44 | 0.3636 |
| Blue | .10 | 34 | 22 | 6.5454 |
| Orange | .10 | 22 | 22 | 0 |
| Green | .10 | 30 | 22 | 2.9090 |
| yellow | .20 | 36 | 44 | 1.4545 |
| | 1.00 (total) | 220 (total) | 220 (total) | $X^2$ =12.242 (total) |

**Q3.**

*Determine the degrees of freedom: number of categories – 1 = 6-1 = 5.*
*Using the table, we find that with df=5 and $X^2$ = 12.242 : p-value < 0.05*

*Because our p value of the x goodness of fit test is less than 0.05, we reject the null hypothesis. i.e. there is a significant difference in M&Ms color ratios between the actual store-bought bags and what the Mars company claims are the actual ratios.*

**Q4.**

*Determine the degrees of freedom: number of categories – 1 = 6-1 = 5.*
*Using the table, we find that with df=5 and $X^2$ = 12.242 : p-value > 0.01*

*Because our p value of the x goodness of fit test is more than 0.01, we cannot reject the null hypothesis. i.e. there is no evidences of a significant difference in M&Ms color ratios between the actual store-bought bags and what the Mars company claims are the actual ratios.*

## Useful

$$X^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

| Degrees of Freedom | Probability | | | | | |
|---|---|---|---|---|---|---|
| | **0.90** | **0.50** | **0.25** | **0.10** | **0.05** | **0.01** |
| **1** | 0.016 | 0.46 | 1.32 | 2.71 | 3.84 | 6.64 |
| **2** | .0.21 | 1.39 | 2.77 | 4.61 | 5.99 | 9.21 |
| **3** | 0.58 | 2.37 | 4.11 | 6.25 | 7.82 | 11.35 |
| **4** | 1.06 | 3.36 | 5.39 | 7.78 | 9.49 | 13.28 |
| **5** | 1.61 | 4.35 | 6.63 | 9.24 | 11.07 | 15.09 |