Goodness of fit test
Did you know that M&M's® Milk Chocolate Candies are supposed to come in the following percentages: 24% blue, 20% orange, 16% green, 14% yellow, 13% red, 13% brown?

http://www.scientificamerikeren.com/candy5.asp

“To our surprise M&Ms met our demand to review their procedures in determining candy ratios. It is, however, noted that the figures presented in their email differ from the information provided from their website (http://us.mms.com/us/about/products/milkchocolate/). An email was sent back informing them of this fact. To which M&Ms corrected themselves with one last email:

In response to your email regarding M&M'S CHOCOLATE CANDIES

Thank you for your email.
On average, our new mix of colors for M&M'S® Chocolate Candies is:

M&M'S® Milk Chocolate: 24% blue, 20% orange, 16% green, 14% yellow, 13% red, 13% brown.

M&M'S® Peanut: 23% blue, 23% orange, 15% green, 15% yellow, 12% red, 12% brown.

M&M'S® Kids MINIS®: 25% blue, 25% orange, 12% green, 13% yellow, 12% red, 13% brown.

M&M'S® Crispy: 17% blue, 16% orange, 16% green, 17% yellow, 17% red, 17% brown.

M&M'S® Peanut Butter and Almond: 20% blue, 20% orange, 20% green, 20% yellow, 10% red, 10% brown.

Have a great day!

Your Friends at Masterfoods USA
A Division of Mars, Incorporated

How to accept or reject the null hypothesis that these probabilities are correct from a finite sample?
Pearson chi\(^2\) Goodness of Fit Test

- Assume there is a sample of size \(n\) from a population
- There are \(k\) classes (e.g. 6 M&M colors) in the populations
- Null hypothesis \(H_0\): class \(i\) has frequency \(f_i\) in the population
- Alternative hypothesis \(H_1\): some population frequencies are inconsistent with \(f_i\)
- Let \(O_i\) be the observed number of sample elements in the \(i\)th class.
- Let \(E_i = n f_i\) be the expected number of sample elements in the \(i\)th class.
- Group any bin with \(E_i < 3\) with
  a) if numerical value of \(i\) is important, group it with its neighbor (\(k=i-1\) or \(k=i+1\)) which has the smallest \(E_k\) until \(E_{group} \geq 3\);
  b) If numerical value of \(i\) is irrelevant, group together all \(E_i < 3\) bins until \(E_{group} \geq 3\)

\[
X_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{9-47}
\]

P-value is calculated based on the chi-square distribution with \(k-1\) degrees of freedom:

\[
P\text{-value} = \text{Prob}(H_0 \text{ is correct}) = 1 - \text{CDF}_\text{chi-squared}(X_0^2, k-1)
\]
The chi-square Goodness of Fit Test is a **one-sided** hypothesis.

\[
X_0^2 = \text{GOF} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}
\]

Say \( X_0^2 = 10 \)

For M&M

\( k = 6 \to k-1 = 5 \)

\( X_0^2 \) is the p-value that null hypothesis is correct.
M&M group exercise

• **DO NOT EAT CANDY BEFORE COUNTING IS FINISHED! THEN, PLEASE, DO.**

• We will be testing three null hypotheses one after another:
  
  – M&M official data: **24% blue, 20% orange, 16% green, 14% yellow, 13% red, 13% brown**
  
    - **18.36% blue, 20.76% orange, 18.44% green, 14.08% yellow, 14.20% red, 14.16% brown**
  
  – Uniform distribution: 1/6~16.67% of each candy color

• You will estimate P-values for each of these null hypothesis

• Hints: \( O_i \) – is the observed # of candies of color i; calculate the expected # \( E_i=(\# \text{ candies in your sample}) \times f_i \)

Use \( 1-\text{chi2cdf}(X_0\text{ squared}, 5) \) for P-value

\[
X_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}
\]
M&M matlab exercise

- observed=mm_table(group,:); group % use when analyzing one group
- f_mm=[0.24,0.2,0.16, 0.14, 0.13,0.13];
- f_u=1./6.*ones(1,6);
- f_website=[18,21,18,14,14,14,14];
- f_website=f_website./sum(f_website);
- %p_website=[0.1836, 0.2076, 0.1844, 0.1408, 0.1420, 0.1416]
- %p_u=[0.1500, 0.2200, 0.2100, 0.1200, 0.1600, 0.1500];
- n=sum(observed)
- expected_u=n.*f_u;
- expected_mm=n.*f_mm;
- expected_website=n.*f_website;
- gf_mm=0; gf_u=0; gf_website=0;
- for m=1:6;
-   gf_mm=gf_mm+(observed(m)... -expected_mm(m)).^2./expected_mm(m);
-   gf_u=gf_u+(observed(m)-expected_u(m)).^2./expected_u(m);
-   gf_website=gf_website+(observed(m)... -expected_website(m)).^2./expected_website(m);
- end;
- disp('goodness of fit of MM ='); disp(num2str(gf_mm));
- disp('p-value of MM ='); disp(num2str(1-chi2cdf(gf_mm,5))); disp(' ');
- disp('goodness of fit of website ='); disp(num2str(gf_website));
- disp('p-value of MM ='); disp(num2str(1-chi2cdf(gf_website,5))); disp(' ');
- disp('goodness of fit of uniform ='); disp(num2str(gf_u));
- disp('p-value of uniform='); disp(num2str(1-chi2cdf(gf_u,5)));
Statistical tests of independence
• Did I mix M&M candy well?

<table>
<thead>
<tr>
<th></th>
<th>blue</th>
<th>orange</th>
<th>green</th>
<th>yellow</th>
<th>red</th>
<th>brown</th>
</tr>
</thead>
<tbody>
<tr>
<td>group 1</td>
<td>56</td>
<td>62</td>
<td>36</td>
<td>36</td>
<td>37</td>
<td>35</td>
</tr>
<tr>
<td>group 2</td>
<td>59</td>
<td>67</td>
<td>29</td>
<td>39</td>
<td>32</td>
<td>25</td>
</tr>
<tr>
<td>group 3</td>
<td>58</td>
<td>63</td>
<td>29</td>
<td>28</td>
<td>33</td>
<td>24</td>
</tr>
<tr>
<td>group 4</td>
<td>58</td>
<td>60</td>
<td>36</td>
<td>22</td>
<td>37</td>
<td>36</td>
</tr>
</tbody>
</table>
How to **test the hypothesis** if multiple samples are drawn from the same population?

- **Table**: samples (Student groups) – rows, classes (M&M colors) – columns
- Test if color fractions are independent from group
- \( P(\text{Group 1 and Color } = \text{ green}) = P(\text{Group 1}) \times P(\text{Color green}) \)
- Compute for all groups/colors 6*4=24 in our case

\[
E_{\text{green (group 1)}} = n_{\text{tot}} \times (\text{group 1}/n_{\text{tot}}) \times (\text{green}/n_{\text{tot}})
\]

\[
\chi^2 = \sum_{\text{groups & colors}}^{n_{\text{tot}}} \frac{(O_{\text{color (group)}} - E_{\text{color (group)}})^2}{E_{\text{color (group)}}}
\]

- # degrees of freedom = (colors-1)*(groups-1)
• Did I mix M&M candy well?

<table>
<thead>
<tr>
<th></th>
<th>blue</th>
<th>orange</th>
<th>green</th>
<th>yellow</th>
<th>red</th>
<th>brown</th>
</tr>
</thead>
<tbody>
<tr>
<td>group 1</td>
<td>56</td>
<td>62</td>
<td>36</td>
<td>36</td>
<td>37</td>
<td>35</td>
</tr>
<tr>
<td>group 2</td>
<td>59</td>
<td>67</td>
<td>29</td>
<td>39</td>
<td>32</td>
<td>25</td>
</tr>
<tr>
<td>group 3</td>
<td>58</td>
<td>63</td>
<td>29</td>
<td>28</td>
<td>33</td>
<td>24</td>
</tr>
<tr>
<td>group 4</td>
<td>58</td>
<td>60</td>
<td>36</td>
<td>22</td>
<td>37</td>
<td>36</td>
</tr>
</tbody>
</table>

• Using $\chi^2 = \sum_{\text{groups} \& \text{colors}} \frac{(O_{\text{color (group)}} - E_{\text{color (group)}})^2}{E_{\text{color (group)}}}$

with # degrees of freedom (colors-1)*(groups-1)

Find P-value of null hypothesis $H_0$ that samples are independent from each other
My solution

- `disp(mm_table)`
- `sumt=sum(sum(mm_table))`
- `sum_color=sum(mm_table, 1)`
- `sum_group=sum(mm_table, 2)`
- `mm_exp=kron(sum_group,sum_color)./sumt`
- `gof=sum(sum((mm_table-mm_exp).^2./mm_exp))`
- `P_value_gof=1-chi2cdf(gof, (4-1)*(6-1))`
- `%gof =10.8858; P_value_gof =0.7606`
- Null model that samples are independent is **not rejected**
- **I mixed this bag well!**
Batch effect
Does color composition vary from bag to bag?

- Last year: 231 252 130 125 139 120
- 2 years ago: 283 286 304 124 191 173
- Test if they are significantly different from each other:
- Same test expect ngroups=2; ncolors=6;
- Results:
  
  Goodness of Fit = 39.7111
  
  P_value = 1.7077e-07
- Batch effect is highly statistically significant!
Goodness of fit with a PDF defined by \( m \) parameters

- As before: \( k \) classes (e.g. M&M colors)
- Use parameter estimators to find the best parameters for the fit
  - Method of moments
  - MLE: method of maximum likelihood
- Use chi-squared distribution with \( k-1-m \) degrees of freedom
- As before: if \( E_i < 3 \), group it together with another group and reduce \( k \) by 1

\[
X_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}
\]

(9-47)
Example 9-12

EXAMPLE 9-12 Printed Circuit Board Defects
Poisson Distribution
The number of defects in printed circuit boards is hypothesized to follow a Poisson distribution. A random sample of $n = 60$ printed boards has been collected, and the following number of defects observed.

<table>
<thead>
<tr>
<th>Number of Defects</th>
<th>Observed Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Example 9-12

The mean of the assumed Poisson distribution in this example is unknown and must be estimated from the sample data. The estimate of the mean number of defects per board is the sample average, that is, \((32.0 + 15.1 + 9.2 + 4.3)/60 = 0.75\). From the Poisson distribution with parameter 0.75, we may compute \(p_i\), the theoretical, hypothesized probability associated with the \(i\)th class interval. Since each class interval corresponds to a particular number of defects, we may find the \(p_i\) as follows:

\[
P_1 = P(X = 0) = \frac{e^{-0.75}(0.75)^0}{0!} = 0.472
\]

\[
P_2 = P(X = 1) = \frac{e^{-0.75}(0.75)^1}{1!} = 0.354
\]

\[
P_3 = P(X = 2) = \frac{e^{-0.75}(0.75)^2}{2!} = 0.133
\]

\[
P_4 = P(X \geq 3) = 1 - (p_1 + p_2 + p_3) = 0.041
\]
Example 9-12

The expected frequencies are computed by multiplying the sample size $n = 60$ times the probabilities $p_i$. That is, $E_i = np_i$. The expected frequencies follow:

<table>
<thead>
<tr>
<th>Number of Defects</th>
<th>Probability</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.472</td>
<td>28.32</td>
</tr>
<tr>
<td>1</td>
<td>0.354</td>
<td>21.24</td>
</tr>
<tr>
<td>2</td>
<td>0.133</td>
<td>7.98</td>
</tr>
<tr>
<td>3 (or more)</td>
<td>0.041</td>
<td>2.46</td>
</tr>
</tbody>
</table>
Example 9-12

Since the expected frequency in the last cell is less than 3, we combine the last two cells:

<table>
<thead>
<tr>
<th>Number of Defects</th>
<th>Observed Frequency</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>32</td>
<td>28.32</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>21.24</td>
</tr>
<tr>
<td>2 (or more)</td>
<td>13</td>
<td>10.44</td>
</tr>
</tbody>
</table>

The chi-square test statistic in Equation 9-47 will have \( k - p - 1 = 3 - 1 - 1 = 1 \) degree of freedom, because the mean of the Poisson distribution was estimated from the data.
9-7 Testing for Goodness of Fit

Example 9-12

The seven-step hypothesis-testing procedure may now be applied, using $\alpha = 0.05$, as follows:

1. **Parameter of interest:** The variable of interest is the form of the distribution of defects in printed circuit boards.

2. **Null hypothesis:** $H_0$: The form of the distribution of defects is Poisson.

3. **Alternative hypothesis:** $H_1$: The form of the distribution of defects is not Poisson.

4. **Test statistic:** The test statistic is

$$
\chi^2_0 = \sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i}
$$
Example 9-12

5. **Reject $H_0$ if:** Reject $H_0$ if the $P$-value is less than 0.05.

6. **Computations:**

\[
\chi^2_0 = \frac{(32 - 28.32)^2}{28.32} + \frac{(15 - 21.24)^2}{21.24} \\
+ \frac{(13 - 10.44)^2}{10.44} = 2.94
\]

7. **Conclusions:** We find from Appendix Table III that $\chi^2_{0.10,1} = 2.71$ and $\chi^2_{0.05,1} = 3.84$. Because $\chi^2_0 = 2.94$ lies between these values, we conclude that the $P$-value is between 0.05 and 0.10. Therefore, since the $P$-value exceeds 0.05 we are unable to reject the null hypothesis that the distribution of defects in printed circuit boards is Poisson. The exact $P$-value computed from Minitab is 0.0864.