Chi-Squared Tests

Semester 1
Goodness of Fit

Up to now, we have tested hypotheses concerning the values of population parameters such as the population mean or proportion.

We have not considered testing hypotheses about the form of a population’s distribution.

We next consider the problem of determining whether or not a population follows a particular distribution.
For example, we may be interested in determining whether the number of emails arriving per minute at a server follows a Poisson distribution or not.

Similarly, we may wish to test if the lengths of components from an automated process follow a normal distribution.

Another similar question is whether a 6-sided die is fair or not.
The general procedure for testing hypotheses on the distribution of a population is as follows.

(i) The null hypothesis $H_0$ is that some distribution describes the population.

(ii) We choose a sample of size $n$ from the population. This might involve recording the results of $n$ rolls of the die or recording the number of mails arriving in $n$ minute-long intervals.

(iii) The observations in our sample are grouped into $k$ “bins” or “classes” and we record the number of observations that fall into each bin or the frequency of each bin. $O_i$ represents the number of observations or observed frequency for the $i^{th}$ bin.

(iv) Under the null hypothesis, we can calculate the expected frequency $E_i$ for each bin.
(v) The test statistic we use is

\[ \chi_0^2 = \sum_{i=1}^{k} \frac{(E_i - O_i)^2}{E_i}. \]

If the null hypothesis is true, then \( \chi_0^2 \) has approximately a chi-squared distribution with \( k - p - 1 \) degrees of freedom. Here \( p \) is the number of parameters of the distribution that we have to estimate with our sample data.

(vi) We reject \( H_0 \) at the significance level \( \alpha \) if the value of \( \chi_0^2 \) calculated with our sample data exceeds the critical value \( \chi_{k-p-1,\alpha}^2 \) which we obtain from a table of chi-square critical values.
For example:

- For testing the fairness of a die, we would use 6 bins for the numbers 1 to 6 and $O_i$ would count the number of times the number $i$ came up in the $n$ rolls.

- For the email server example, the bins might represent 0 emails, 1 email, 2 emails, 3 emails and $\geq 4$ emails. In this case, we would have 5 bins in total, and $O_1$ would count the number of minutes in which we received 0 emails, $O_2$ the number of minutes in which we received 1 email, $O_3$ the number of minutes in which we received 2 emails and so on.
If the 6-sided die is fair, then the expected frequency is\[ E_i = \frac{n}{6} \]for each bin.

If the number of emails arriving at the server per minute follows a Poisson distribution with mean \( \lambda \), the expected number of minutes in which no emails arrive would be\[ e^{-\lambda} n. \]

The expected number of minutes in which 1 email arrives would be\[ \frac{e^{-\lambda} \lambda^1}{1!} n \]

and so on.
The key point is that we can compute expected frequencies based on the null hypothesis and then compare the expected frequencies with the actual frequencies observed in a sample.

When the deviation between the expected frequencies and the observed frequencies is too large we reject the null hypothesis concerning the population.

To determine when the difference between observed and expected frequencies is too large, we use a special distribution known as the chi-squared distribution.
The number of emails arriving at a server per minute is claimed to follow a Poisson distribution. To test this claim, the number of emails arriving in 70 randomly chosen 1-minute intervals is recorded. The table below summarises the results.

<table>
<thead>
<tr>
<th>Number of emails</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>≥ 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>13</td>
<td>22</td>
<td>23</td>
<td>12</td>
<td>0</td>
</tr>
</tbody>
</table>

Test the hypothesis that the number of emails per minute follows a Poisson distribution? Use a significance level of $\alpha = 0.05$. 

Chi-Squared Tests
Poisson Goodness of Fit

To calculate the expected frequencies, we need the Poisson parameter $\lambda$. This is simply the mean number of emails per minute. We need to estimate this from the sample data:

$$\lambda = \frac{13(0) + 22(1) + 23(2) + 12(3)}{70} = 1.49.$$ 

- Our Null Hypothesis is $H_0$: Number of emails per minute has a Poisson Distribution with $\lambda = 1.49$.
- $H_1$: Number of emails per minute does not have a Poisson Distribution with $\lambda = 1.49$.
- Significance Level: $\alpha = 0.05$.
- Test Statistic: We treat the last two bins as one (as no minutes contained 4 or more calls) so the number of bins is $k = 4$.

$$\chi^2_0 = \sum_{i=1}^{4} \frac{(E_i - O_i)^2}{E_i}.$$ 

Chi-Squared Tests
We reject $H_0$ if our sample data gives a value of $\chi^2_0 > \chi^2_{2,0.05} = 5.99$. We have lost two degrees of freedom because we have to estimate the parameter $\lambda$ from sample data.

To do the calculation, we require:

<table>
<thead>
<tr>
<th>Number of emails</th>
<th>Observed Freq.</th>
<th>Expected Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>13</td>
<td>15.78</td>
</tr>
<tr>
<td>1</td>
<td>22</td>
<td>23.51</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>17.5</td>
</tr>
<tr>
<td>$\geq 3$</td>
<td>12</td>
<td>13.2</td>
</tr>
</tbody>
</table>

Chi-Squared Tests
Poisson Goodness of Fit

The actual value of $\chi^2_0$ is then:

\[
\frac{(15.78 - 13)^2}{15.78} + \frac{(23.51 - 22)^2}{23.51} + \frac{(17.51 - 23)^2}{17.51} + \frac{(13.2 - 12)^2}{13.2}
\]

which is equal to 2.417.

We cannot reject the null hypothesis at the 5% level of significance.
It is also possible to perform a goodness of fit test for distributions other than the Poisson distribution.

The approach is essentially the same - all that changes is the distribution used to calculate the expected frequencies.

We next consider an example based on the Binomial distribution.
Example

Bits are sent over a communications channel in packets of 8. In order to characterise the performance of this channel, 80 packets are sent over the channel and the number of corrupted bits in each packet is recorded. The results of this experiment are recorded below.

<table>
<thead>
<tr>
<th>Number of Corrupt Bits</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>≥ 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Packets</td>
<td>35</td>
<td>31</td>
<td>10</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

Test the hypothesis that the number of corrupted bits in a packet sent over this channel follows a binomial distribution. Use a significance level of $\alpha = 0.025$. 
To calculate the expected frequencies, we need the binomial parameter $p$.

We need to estimate this from the sample data.

Out of the 640 bits sent over the channel, 63 were corrupt.

So our estimate of $p$ is $\frac{63}{640} = 0.098$
Binomial Goodness of Fit

- \( H_0 \): Population is binomial with \( p = 0.098 \).
- \( H_1 \): Population is not binomial.
- Significance Level: \( \alpha = 0.025 \).
- Test Statistic: We treat the last two bins as one (as no packets contain 4 or more corrupt bits) so the number of bins is \( k = 4 \).

\[
\chi_0^2 = \sum_{i=1}^{4} \frac{(E_i - O_i)^2}{E_i}.
\]

We reject \( H_0 \) if our sample data gives a value of \( \chi_0^2 > \chi_{2,0.025}^2 = 7.378 \). We have lost two degrees of freedom because we have to estimate the parameter \( p \) from sample data.
Binomial Goodness of Fit

To do the calculation, we require:

<table>
<thead>
<tr>
<th>Number of Corrupt Bits</th>
<th>Observed Freq.</th>
<th>Expected Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>35</td>
<td>35.04</td>
</tr>
<tr>
<td>1</td>
<td>31</td>
<td>30.48</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>11.6</td>
</tr>
<tr>
<td>≥ 3</td>
<td>4</td>
<td>2.88</td>
</tr>
</tbody>
</table>

The actual value of $\chi^2$ is then:

$$\frac{(35.04 - 35)^2}{35.04} + \frac{(30.48 - 31)^2}{30.48} + \frac{(11.6 - 10)^2}{11.6} + \frac{(2.88 - 4)^2}{2.88} = 0.665.$$  

We cannot reject the null hypothesis at the 1% level of significance.
The final example of goodness of fit that we shall consider is for the Normal distribution.

For this case, the situation is a little more complicated as the distribution is continuous. This means that we need to be more careful in selecting the bins.

In practice, it is usual to choose bins so that the expected frequency for each bin is the same.

We shall see how to do this in an example below.
Normal Goodness of Fit

Example

A text processing tool can be downloaded from a particular webserver. The administrator of the server wishes to test if the download times are adequately described by a normal distribution. A random sample of 80 users is selected and their download times recorded. The mean and standard deviation of the download times (in seconds) for the sample are 20.2 and 2.1 respectively.
Suppose we wish to use 8 bins.

We first find the intervals that divide the standard normal distribution into 8 equal parts.

From the table of standard normal probabilities we can see that these intervals are:

\[ (-\infty, -1.15], (-1.15, -0.675], (-0.675, -0.32], (-0.32, 0] \]

and their mirror images on the other side of 0.

This allows us to construct the bins in which to group our data.
The first bin will be $x \leq 20.2 - 1.15(2.1) = 17.785$, the second bin will be $17.785 < x \leq 20.2 - 0.675(2.1) = 18.7825$ and so on. The table below summarises the observed download times for the 80 users in the sample as well as the expected frequencies for each bin.

<table>
<thead>
<tr>
<th>Download Time ($x$)</th>
<th>Observed Freq.</th>
<th>Expected Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x \leq 17.785$</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>$17.785 &lt; x \leq 18.7825$</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>$18.7825 &lt; x \leq 19.528$</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>$19.528 &lt; x \leq 20.2$</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>$20.2 &lt; x \leq 20.872$</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>$20.872 &lt; x \leq 21.6175$</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>$21.6175 &lt; x \leq 22.615$</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>$22.615 &lt; x$</td>
<td>12</td>
<td>10</td>
</tr>
</tbody>
</table>
Normal Goodness of Fit

We can now test the data for normality at a 5% level of significance following the same procedure as before.

1. $H_0$: The form of the distribution is normal.
2. $H_1$: The form of the distribution is non-normal.
3. Significance Level: $\alpha = 0.05$.
4. Test Statistic
   \[
   \chi^2_0 = \sum_{i=1}^{8} \frac{(E_i - O_i)^2}{E_i}.
   \]

5. We reject $H_0$ if our sample data gives a value of $\chi^2_0 > \chi^2_{5,0.05} = 11.07$. The number of degrees of freedom is $8 - 2 - 1 = 5$ because we estimated two parameters from the data.

6. The actual value of $\chi^2_0$ for our sample is 3. As this is not greater than the critical value of 11.07 we cannot reject the null hypothesis at the 5% level of significance.

Chi-Squared Tests
Another use of the chi-square distribution is to assess the independence of two different ways of classifying a population.

For example, we could classify drivers according to age and according to the insurance premium they pay and test whether these classifications are independent.

Another example would be to classify computer users by the number of times their computer crashes per week and also by the operating system they use.
Contingency Tables

- If the first way of classifying the population has \( r \) levels (\( r \) different age categories for the drivers) and the second has \( c \) levels (\( c \) different categories of insurance premiums), the observed frequencies can be recorded in an \( r \times c \) contingency table with \( r \) rows and \( c \) columns.

- A sample of \( n \) observations is selected. For \( 1 \leq i \leq r \) and \( 1 \leq j \leq c \), we let \( O_{ij} \) denote the frequency observed for level \( i \) in the first classification and level \( j \) in the second classification.

- \( t_i \) denotes the total number of observations in category \( i \) in the first way of classifying the population and \( s_j \) denotes the total number of observations in category \( j \) in the second way of classifying the population.
If the two methods of classification are independent then the expected frequency of “cell” \((i,j)\), denoted \(E_{ij}\) is

\[
\frac{t_i s_j}{n}.
\]

Similar to goodness of fit tests, we use the test statistic

\[
\chi^2_0 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(E_{ij} - O_{ij})^2}{E_{ij}}
\]

which has a chi-square distribution with \((r - 1)(c - 1)\) degrees of freedom if the hypothesis of independence is true.
A car rental company wishes to test if the age of a car rented to a customer and the customer’s level of satisfaction are independent of each other. When a customer returns a car they rate their level of satisfaction as one of **dissatisfied, no opinion, satisfied, very satisfied**. The company only rents out cars that are two years old or less. A random sample of 60 customers is selected and the results of the sample are recorded in the table below.
<table>
<thead>
<tr>
<th>Age of Car</th>
<th>Dissatisfied</th>
<th>No Opinion</th>
<th>Satisfied</th>
<th>Very Satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>4</td>
<td>6</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>1 Year</td>
<td>3</td>
<td>10</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>2 Years</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

Test the hypothesis that the level of satisfaction is independent of the age of the car at a 5% level of significance.
Contingency Tables

1. **H₀**: The two methods of classification are independent.
2. **H₁**: The two methods of classification are not independent.
3. Significance Level: \( \alpha = 0.05 \).
4. Test Statistic:
   \[
   \chi^2 = \sum_{i=1}^{3} \sum_{j=1}^{4} \frac{(E_{ij} - O_{ij})^2}{E_{ij}}.
   \]
5. Reject \( H₀ \) if the value of \( \chi^2 \) is greater than \( \chi^2_{6,0.05} = 12.59 \). We have 6 = (3 – 1)(4 – 1) degrees of freedom.
6. The value of \( \chi^2 \) for our sample is 9.91.
7. As this is not greater than the critical value of 12.59, we cannot reject \( H₀ \) at the 5% level of significance.