Introduction

• Many experiments result in measurements that are qualitative or categorical rather than quantitative.
  – People classified by ethnic origin
  – Cars classified by color
  – M&M®s classified by type (plain or peanut)
• These data sets have the characteristics of a multinomial experiment.

The Multinomial Experiment

1. The experiment consists of $n$ identical trials.
2. Each trial results in one of $k$ categories.
3. The probability that the outcome falls into a particular category $i$ on a single trial is $p_i$ and remains constant from trial to trial. The sum of all $k$ probabilities, $p_1 + p_2 + \ldots + p_k = 1$.
4. The trials are independent.
5. We are interested in the number of outcomes in each category, $O_1, O_2, \ldots, O_k$ with $O_1 + O_2 + \ldots + O_k = n$.

The Binomial Experiment

• A special case of the multinomial experiment with $k = 2$.
• Categories 1 and 2: success and failure
• $p_1$ and $p_2$: $p$ and $q$
• $O_1$ and $O_2$: $x$ and $n-x$
• We made inferences about $p$ (and $q = 1 - p$)

Pearson’s Chi-Square Statistic

• We have some preconceived idea about the values of the $p_i$ and want to use sample information to see if we are correct.
• The expected number of times that outcome $i$ will occur is $E_i = np_i$. If the observed cell counts, $O_i$, are too far from what we hypothesize under $H_0$, the more likely it is that $H_0$ should be rejected.
Pearson’s Chi-Square Statistic

- We use the Pearson chi-square statistic:

\[ X^2 = \sum \frac{(O - E)^2}{E} \]

- When \( H_0 \) is true, the differences \( O - E \) will be small, but large when \( H_0 \) is false.
- Look for large values of \( X^2 \) based on the chi-square distribution with a particular number of degrees of freedom.

Degrees of Freedom

- These will be different depending on the application.
1. Start with the number of categories or cells in the experiment.
2. Subtract 1 df for each linear restriction on the cell probabilities. (You always lose 1 df since \( p_1 + p_2 + \ldots + p_k = 1 \).)
3. Subtract 1 df for every population parameter you have to estimate to calculate or estimate \( E_i \).

The Goodness of Fit Test

- The simplest of the applications.
- A single categorical variable is measured, and exact numerical values are specified for each of the \( p_i \).
- Expected cell counts are \( E_i = np_i \)
- Degrees of freedom: \( df = k - 1 \)

\[ \text{Test statistic: } X^2 = \sum \frac{(O - E)^2}{E} \]

Example

- A multinomial experiment with \( k = 6 \) and \( O_1 \) to \( O_6 \) given in the table.

<table>
<thead>
<tr>
<th>Upper Face</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of times</td>
<td>50</td>
<td>39</td>
<td>45</td>
<td>62</td>
<td>61</td>
<td>43</td>
</tr>
</tbody>
</table>

- A multinomial experiment with \( k = 6 \) and \( O_1 \) to \( O_6 \) given in the table.
- We test:

\( H_0: p_1 = 1/6; p_2 = 1/6; \ldots p_6 = 1/6 \) (die is fair)
\( H_a: \) at least one \( p_i \) is different from 1/6 (die is biased)

Some Notes

- The test statistic, \( X^2 \) has only an approximate chi-square distribution.
- For the approximation to be accurate, statisticians recommend \( E_i \geq 5 \) for all cells.
- Goodness of fit tests are different from previous tests since the experimenter uses \( H_0 \) for the model he thinks is true.

\( H_a: \) model is correct (as specified)
\( H_0: \) model is not correct

- Be careful not to accept \( H_0 \) (say the model is correct) without reporting \( \beta \).
Contingency Tables: A Two-Way Classification

- The experimenter measures two qualitative variables to generate bivariate data.
  - Gender and colorblindness
  - Age and opinion
  - Professorial rank and type of university
- Summarize the data by counting the observed number of outcomes in each of the intersections of category levels in a contingency table.

Chi-Square Test of Independence

$H_0$: classifications are independent
$H_a$: classifications are dependent

- Observed cell counts are $O_{ij}$ for row $i$ and column $j$.
- Expected cell counts are $E_{ij} = np_{ij}$
  - If $H_0$ is true and the classifications are independent, $p_{ij} = p_pj = P$(falling in row $i$)$P$(falling in row $j$)

The Furniture Problem

- Furniture defects are classified according to type of defect and shift on which it was made.

### Chi-Square Test of Independence

$H_0$: type of defect is independent of shift
$H_a$: type of defect depends on the shift

<table>
<thead>
<tr>
<th>Type</th>
<th>Shift 1</th>
<th>Shift 2</th>
<th>Shift 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>15</td>
<td>26</td>
<td>33</td>
<td>74</td>
</tr>
<tr>
<td>B</td>
<td>21</td>
<td>31</td>
<td>17</td>
<td>69</td>
</tr>
<tr>
<td>C</td>
<td>45</td>
<td>34</td>
<td>49</td>
<td>128</td>
</tr>
<tr>
<td>D</td>
<td>13</td>
<td>5</td>
<td>20</td>
<td>38</td>
</tr>
</tbody>
</table>

Do $H_0$ hold?

1. Calculate the expected cell counts. For example: $E = \frac{15 \times 69}{74} = 13.92$.
2. Calculate the test statistic $X^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$.
3. Compare $X^2$ to the chi-square distribution with $df = (r-1)(c-1)$.

The test statistic has an approximate chi-square distribution with $df = (r-1)(c-1)$.

The Furniture Problem

- The experimenter measures $O_{ij}$ for row $i$ and column $j$.
- The experimenter measures $E_{ij} = np_{ij}$.
- Summarize the data by counting the observed number of outcomes in each of the intersections of category levels in a contingency table.

Chi-Square Test of Independence

- The contingency table has $r$ rows and $c$ columns—$rc$ total cells.

<table>
<thead>
<tr>
<th>Shift 1</th>
<th>Shift 2</th>
<th>Shift 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$O_{11}$</td>
<td>$O_{12}$</td>
<td>$O_{13}$</td>
</tr>
<tr>
<td>2</td>
<td>$O_{21}$</td>
<td>$O_{22}$</td>
<td>$O_{23}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$r$</td>
<td>$O_{r1}$</td>
<td>$O_{r2}$</td>
<td>$O_{r3}$</td>
</tr>
</tbody>
</table>

Does the distribution of measurements in the various categories for variable 1 depend on which category of variable 2 is being observed? If not, the variables are independent.

We study the relationship between the two variables. Is the type of defect contingent or dependent on the shift?
Comparing Multinomial Populations

- Sometimes researchers design an experiment so that the number of experimental units falling in one set of categories is fixed in advance.
- **Example:** An experimenter selects 900 patients who have been treated for flu prevention. She selects 300 from each of three types—no vaccine, one shot, and two shots. The column totals have been fixed in advance!

<table>
<thead>
<tr>
<th>Flu</th>
<th>No Vaccine</th>
<th>One Shot</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favor A</td>
<td>76</td>
<td>53</td>
<td>129</td>
</tr>
<tr>
<td>Do not favor A</td>
<td>124</td>
<td>147</td>
<td>271</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>200</td>
<td>400</td>
</tr>
</tbody>
</table>


**The Voter Problem**

Since we know that there are differences among the four wards, what are the nature of the differences? Look at the proportions in favor of candidate A in the four wards.

<table>
<thead>
<tr>
<th>Ward</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favor A</td>
<td>76</td>
<td>53</td>
<td>59</td>
<td>48</td>
<td>236</td>
</tr>
<tr>
<td>Do not favor A</td>
<td>124</td>
<td>147</td>
<td>141</td>
<td>152</td>
<td>564</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>800</td>
</tr>
</tbody>
</table>

H₀: p₁ = p₂ = p₃ = p₄
where pᵢ = fraction favoring A in each of the four wards

Equivalent Statistical Tests

A multinomial experiment with two categories is a binomial experiment.

The data from two binomial experiments can be displayed as a two-way classification.

There are statistical tests for these two situations based on the statistic of Chapter 9:

- One sample: \( \hat{p} = \frac{k}{n} \) (Successes / Total)  
- Two samples: \( \hat{p} = \frac{n₁ \hat{p}_₁ + n₂ \hat{p}_₂}{n₁ + n₂} \) (Successes + Failures)
Other Applications

✓ **Goodness-of-fit tests**: It is possible to design a goodness-of-fit test to determine whether data are consistent with data drawn from a particular probability distribution, possibly the normal, binomial, Poisson, or other distributions.

✓ **Time-dependent multinomials**: The chi square statistic can be used to investigate the rate of change of multinomial proportions over time.

Other Applications

✓ **Multidimensional contingency tables**: Instead of only two methods of classification, it is possible to investigate a dependence among three or more classifications.

✓ **Log-linear models**: Complex models can be created in which the logarithm of the cell probability \( \ln(p_{ij}) \) is some linear function of the row and column probabilities.

Assumptions

**Assumptions for Pearson’s Chi-Square:**

1. The cell counts \( O_1, O_2, \ldots, O_k \) must satisfy the conditions of a multinomial experiment, or a set of multinomial experiments created by fixing either the row or the column totals.
2. The expected cell counts \( E_1, E_2, \ldots, E_k \) should equal or exceed five.

Assumptions

✓ When you calculate the expected cell counts, if you find that one or more is less than five, these options are available to you:

1. **Choose a larger sample size** \( n \). The larger the sample size, the closer the chi-square distribution will approximate the distribution of your test statistic \( X^2 \).
2. It may be possible to **combine one or more of the cells** with small expected cell counts, thereby satisfying the assumption.

Key Concepts

I. **The Multinomial Experiment**

1. There are \( n \) identical trials, and each outcome falls into one of \( k \) categories.
2. The probability of falling into category \( i \) is \( p_i \) and remains constant from trial to trial.
3. The trials are independent, \( \Sigma p_i = 1 \), and we measure \( O_i \), the number of observations that fall into each of the \( k \) categories.

II. **Pearson’s Chi-Square Statistic**

\[ X^2 = \sum \frac{(O_i - E_i)^2}{E_i} \]

where \( E_i = np_i \)

which has an approximate chi-square distribution with degrees of freedom determined by the application.

Key Concepts

III. **The Goodness-of-Fit Test**

1. This is a one-way classification with cell probabilities specified in \( H_0 \).
2. Use the chi-square statistic with \( E_i = np_i \) calculated with the hypothesized probabilities.
3. \( df = k - 1 - \) (Number of parameters estimated in order to find \( E_i \))
4. If \( H_0 \) is rejected, investigate the nature if the differences using the sampling proportions.
Key Concepts

IV. Contingency Tables
1. A two-way classification with \( n \) observations into \( r \times c \) cells of a two-way table using two different methods of classification is called a contingency table.
2. The test for independence of classifications methods uses the chi-square statistic

\[
X^2 = \sum \frac{(o_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}
\]

with \( \hat{E}_{ij} = \frac{r_c \hat{E}_c}{n} \) and \( df = (r-1)(c-1) \)
3. If the null hypothesis of independence of classifications is rejected, investigate the nature of the dependency using conditional proportions within either the rows or columns of the contingency table.

Key Concepts

V. Fixing Row or Column Totals
1. When either the row or column totals are fixed, the test of independence of classifications becomes a test of the homogeneity of cell probabilities for several multinomial experiments.
2. Use the same chi-square statistic as for contingency tables.
3. The large-sample \( z \) tests for one and two binomial proportions are special cases of the chi-square statistic.

Key Concepts

VI. Assumptions
1. The cell counts satisfy the conditions of a multinomial experiment, or a set of multinomial experiments with fixed sample sizes.
2. All expected cell counts must equal or exceed five in order that the chi-square approximation is valid.