GOODNESS-OF-FIT TESTS

We will ask: “how are my data distributed?” Towards this aim, we consider goodness-of-fit tests. Goodness-of-fit tests compare the observed frequency distribution of a sample with a hypothesized frequency distribution. We wish to discern whether or not, apart from sampling error, the observed sample conforms to the hypothesized distribution.

Goodness-of-fit tests are based on count data, i.e., the number of individuals falling in a given category. The attribute observed on each unit sampled determines which category that individual is counted in.

Pearson’s $\chi^2$ Goodness-of-Fit Test

**Example** -- Yule (1923)

We wish to determine whether crossing a yellow-round (YR) pea with a green-wrinkled (gw) pea follows classical Mendelian genetics. Yellow is dominant to green; round is dominant to wrinkled. Theory predicts a relative frequency of $9 \ (YR) : 3 \ (Yw) : 3 \ (gR) : 1 \ (gw)$ in the second generation cross $F_2$.

<table>
<thead>
<tr>
<th>Gametes</th>
<th>YR</th>
<th>Yw</th>
<th>gR</th>
<th>gw</th>
</tr>
</thead>
<tbody>
<tr>
<td>YR</td>
<td>YYRR</td>
<td>YYRw</td>
<td>YgRw</td>
<td>YgRw</td>
</tr>
<tr>
<td>Yw</td>
<td>YYRw</td>
<td>Yyww</td>
<td>YgRw</td>
<td>Ygww</td>
</tr>
<tr>
<td>gR</td>
<td>YgRR</td>
<td>YgRw</td>
<td>ggRR</td>
<td>ggRw</td>
</tr>
<tr>
<td>gw</td>
<td>YgRw</td>
<td>Ygww</td>
<td>ggRw</td>
<td>ggww</td>
</tr>
</tbody>
</table>

Statistical Hypotheses

$H_0$: The distribution of phenotypes in the $F_2$ generation follows a 9:3:3:1 distribution, or specify as multinomial probabilities

$p_{YR} = 9/16$ and $p_{Yw} = 3/16$ and $p_{gR} = 3/16$ and $p_{gw} = 3/16$

$H_1$: The distribution of phenotypes in the $F_2$ generation does not follow a 9:3:3:1 distribution.

$p_{YR} \neq 9/16$ or $p_{Yw} \neq 3/16$ or $p_{gR} \neq 3/16$ or $p_{gw} \neq 3/16$

Multinomial Distribution

$$P(X_1 = x_1 \text{ and } X_2 = x_2 \text{ and } \ldots \text{ and } X_k = x_k) =$$

$$\begin{cases} 
\frac{n!}{x_1!x_2!\ldots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} \text{ when } \sum_{i=1}^{k} x_i = n \\
0 \text{ otherwise.} 
\end{cases}$$

Each of the $k$ components separately has a binomial distribution with parameters $n$ and $p_i$.

Test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(obs_i - exp_i)^2}{exp_i} \sim \chi^2_{df=k-1}$$

where $obs_i$ = observed count in the $i$th category

$exp_i$ = expected count in the $i$th category under $H_0$

$k$ = number of categories

and $\chi^2$ is distributed as a chi-square random variable with $k - 1$ degrees of freedom (df).

**Degrees of freedom** = number of categories ($k$) minus the number of sample constants used to calculate the expected frequencies.

In this example, only one sample constant is used

$\rightarrow n$ (total sample size $n = \sum_{i=1}^{k} obs_i$)

When $n$ is a known, only $k - 1$ frequencies can be specified; the last is determined by subtraction and therefore “fills itself in automatically”.

---

**THEORY**

Parents: $YYRR \times ggww$

First Generation ($F_1$): $YgRw$

Second Gen. ($F_2$): $9 \ YR : 3 \ Yw : 3 \ gR : 1 \ gw$ phenotypes

**Gametes**

<table>
<thead>
<tr>
<th>YR</th>
<th>Yw</th>
<th>gR</th>
<th>gw</th>
</tr>
</thead>
<tbody>
<tr>
<td>YYRR</td>
<td>YYRw</td>
<td>YgRw</td>
<td>YgRw</td>
</tr>
<tr>
<td>YYRw</td>
<td>Yyww</td>
<td>YgRw</td>
<td>Ygww</td>
</tr>
<tr>
<td>YgRR</td>
<td>YgRw</td>
<td>ggRR</td>
<td>ggRw</td>
</tr>
<tr>
<td>YgRw</td>
<td>Ygww</td>
<td>ggRw</td>
<td>ggww</td>
</tr>
</tbody>
</table>
Assumptions
- Sample observations are a random sample of the population.
- Sample observations are independent.

Constraints
- Data are categorical.
- For chi-square approximation:
  (i) none of the expected frequencies < 1.0
  (ii) no more than 20% of the expected frequencies < 5.0
  (i) is more important than (ii).

Set a significance level such as $\alpha = 0.05$
$H_1$ is always two-tailed -- we consider only general alternatives.

Critical value $\chi^2_{0.05,3} = 7.815$ (Table A-4, page 775)

Decision rule: $H_0$ rejected if $\chi^2_{obs} > \chi^2_{crit}$. If test statistic is 7.815 or greater, then reject $H_0$.

Testing Categorical Probabilities: Two-Way tables

Now consider multinomial experiments where data are classified according to two criteria---
*classification wrt two qualitative factors*

Study, based on a survey of 300 TV viewers, looking at relationship between gender of a viewer and the viewer’s brand awareness

<table>
<thead>
<tr>
<th>Category</th>
<th>Obs.</th>
<th>Expected</th>
<th>Obs - Exp</th>
<th>$\frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>YR</td>
<td>2504</td>
<td>4530 * 9/16 = 2548.1</td>
<td>-44.1</td>
<td>.763</td>
</tr>
<tr>
<td>Yw</td>
<td>853</td>
<td>4530 * 3/16 = 849.4</td>
<td>3.6</td>
<td>.015</td>
</tr>
<tr>
<td>gR</td>
<td>881</td>
<td>4530 * 3/16 = 849.4</td>
<td>31.6</td>
<td>1.176</td>
</tr>
<tr>
<td>gw</td>
<td>292</td>
<td>4530 * 1/16 = 283.1</td>
<td>8.9</td>
<td>.280</td>
</tr>
<tr>
<td>$n = \frac{4530}{4530}$</td>
<td>4530</td>
<td>$\chi^2_{obs} = 2.234$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$0.10 < \text{P-value} < 0.90$
Therefore, do not reject $H_0$.

No significant difference between observed frequency and the frequency predicted by Mendelian genetics.

Category | Obs. | Expected | Obs - Exp | $\frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}}$ |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>YR</td>
<td>2504</td>
<td>4530 * 9/16 = 2548.1</td>
<td>-44.1</td>
<td>.763</td>
</tr>
<tr>
<td>Yw</td>
<td>853</td>
<td>4530 * 3/16 = 849.4</td>
<td>3.6</td>
<td>.015</td>
</tr>
<tr>
<td>gR</td>
<td>881</td>
<td>4530 * 3/16 = 849.4</td>
<td>31.6</td>
<td>1.176</td>
</tr>
<tr>
<td>gw</td>
<td>292</td>
<td>4530 * 1/16 = 283.1</td>
<td>8.9</td>
<td>.280</td>
</tr>
<tr>
<td>$n = \frac{4530}{4530}$</td>
<td>4530</td>
<td>$\chi^2_{obs} = 2.234$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\chi^2_{obs} = 2.234 < 7.815$

This is an example of a **two-way contingency table**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand Awareness</td>
<td>Could identify product</td>
<td>$n_{11}$</td>
<td>$n_{12}$</td>
</tr>
<tr>
<td>Could not identify product</td>
<td>$n_{21}$</td>
<td>$n_{22}$</td>
<td>$r_2$</td>
</tr>
<tr>
<td>Totals</td>
<td>$c_1$</td>
<td>$c_2$</td>
<td>$n$</td>
</tr>
</tbody>
</table>

$p_{11}$, $p_{12}$, $p_{r1}$, and $p_{c2}$ are the called **marginal probabilities** for each row or column respectively-- $p_{11} = p_{11} + p_{12}$.
Suppose we want to know if the two classifications are independent?

Does knowing the gender of the TV viewer provide information about the viewer’s brand awareness?

\(H_0\): Brand awareness and gender are statistically independent.

\(H_1\): Brand awareness and gender are statistically dependent.

Remember, if two events \(A\) and \(B\) are independent, \(P(AB) = P(A)P(B)\)

In the analysis of contingency tables, if the two classifications are independent,

\[
\begin{align*}
 p_{11} &= p_{r1}p_{c1} \\
p_{12} &= p_{r1}p_{c2} \\
p_{21} &= p_{r2}p_{c1} \\
p_{22} &= p_{r2}p_{c2}
\end{align*}
\]

To test the hypothesis of independence, first need the expected or mean count in each cell

\[
E(n_{ij}) = np_{ij}
\]

For the TV viewer example:

\[
\begin{align*}
\hat{E}(n_{11}) &= \frac{r_1c_1}{n} = \frac{(136)(150)}{300} = 68 \\
\hat{E}(n_{12}) &= \frac{r_1c_2}{n} = \frac{(136)(150)}{300} = 68 \\
\hat{E}(n_{21}) &= \frac{r_2c_1}{n} = \frac{(164)(150)}{300} = 82 \\
\hat{E}(n_{22}) &= \frac{r_2c_2}{n} = \frac{(164)(150)}{300} = 82
\end{align*}
\]

Use the \(\chi^2\) statistic to test the null hypothesis of independence—compare observed and expected counts in each cell

\[
\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - \hat{E}(n_{ij}))^2}{\hat{E}(n_{ij})}
\]

\[
\chi^2 = \frac{(95 - 68)^2}{68} + \frac{(41 - 68)^2}{68} + \frac{(55 - 82)^2}{82} + \frac{(109 - 82)^2}{82} = 39.22
\]

What do large values of \(\chi^2\) imply?

To determine the cutoff value, we are making use of the fact that the sampling distribution of the \(\chi^2\) test statistic is approximately a \(\chi^2\) probability distribution when the classifications are independent (under the null hypothesis).
The appropriate degrees of freedom in a 2-way table are:

\[(r-1)(c-1)\]

where \(r\) is the number of rows and \(c\) is the number of columns.

The methods we have learned for a 2x2 two-way table can be generalized to problems with more categories. Could have a 3x2, 4x4, etc.

For the TV viewer example,

\[\text{df} = (2-1)(2-1) = 1\]

For \(\alpha = .05\), reject the hypothesis of independence when \(\chi^2 > \chi^2_{.05, 1} = 3.841\)

Since \(\chi^2 = 39.22\) exceeds 3.841 conclude that viewer gender and brand awareness are statistically dependent events at \(\alpha = .05\) significance level.

Using Table A-4 we can also conclude that \(p < 0.005\)

**Good software can be found at:**
http://faculty.vassar.edu/lowry/VassarStats.html

Choose either Chi-Square “Goodness of Fit” Test, or Chi-Square, Cramer’s V, and Lambda

---

**A Word of Caution About Chi-square Tests**

- Should be avoided when expected cell *counts* are too small (< 5)
- If \(\chi^2 < \chi^2_{\alpha, df}\) *do not accept the hypothesis of independence*—risking a Type II error and the probability, \(\beta\) of committing such an error is unknown
- If \(\chi^2 > \chi^2_{\alpha, df}\) avoid inferring that a *causal* relationship exists between the classifications
  - the alternative hypothesis states that the two classifications are statistically dependent
  - statistical dependence does not imply causality