The "Chi-Square Test" arises in the setting of "count" or "categorical" data. For each of several possible categories or "cells", the data consists of the number of observations falling into that cell. The Chi-Square Test compares these observed counts with those that are "expected" under a given $H_0$ to be tested.

**First Examples**

(A) Data: observed counts of outcomes from 60 rolls of a die

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>6</td>
<td>17</td>
<td>16</td>
<td>8</td>
<td>9</td>
<td>60</td>
</tr>
</tbody>
</table>

$H_0$: The die is "fair": each outcome has probability $\frac{1}{6}$

Expected counts if $H_0$ true:

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>10</th>
<th>10</th>
<th>10</th>
<th>10</th>
<th>10</th>
<th>Total</th>
</tr>
</thead>
</table>

This is a "goodness-of-fit" problem: How well does the observed data "fit" the probability model stated in $H_0$?

Q. What is a good test statistic for this kind of problem?
(B) Data: observed counts in the cells of a 2-way classification by categories of 2 variables

Variable 1: Age Group
Child, Young Adult, Mature Adult, Senior
(4 groups)

Variable 2: Preference for McDonald’s
Like, Dislike, Indifferent
(3 groups)

2-way classification (3x4=12 cells)

<table>
<thead>
<tr>
<th>Variable 2</th>
<th>Variable 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Child</td>
</tr>
<tr>
<td>Like</td>
<td>20</td>
</tr>
<tr>
<td>Dislike</td>
<td>0</td>
</tr>
<tr>
<td>Indifferent</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
</tr>
</tbody>
</table>

H₀: These 2 variables are independent.
Knowledge about Variable 1 for a subject is not informative about Variable 2 for that subject, and vice versa.

This is a "Test of Independence" problem:
How well does the observed data agree with the counts we expect if H₀ true?

Q₁. What are the "expected counts" in these cells if H₀ true?
Q₂. What is a good test statistic?
(C) **Data**: observed counts in the cells of a 2-way classification by categories of 1 variable by several different populations

**Variable**: Preference for McDonald's

- Like, Dislike, Indifferent
- (3 groups)

**Populations**: Beijing, London, Paris, Richardson
- (4 famous cities)

**2-way classification** (3x4=12 cells)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beijing</th>
<th>London</th>
<th>Paris</th>
<th>Richardson</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Like</td>
<td>20</td>
<td>70</td>
<td>50</td>
<td>39</td>
<td>179</td>
</tr>
<tr>
<td>Dislike</td>
<td>70</td>
<td>20</td>
<td>50</td>
<td>1</td>
<td>141</td>
</tr>
<tr>
<td>Indifferent</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>40</td>
<td>340</td>
</tr>
</tbody>
</table>

- Sample sizes from the populations
- 4 different samples

**H₀**: Preference for McD's is the same for these four populations.

This is a "Test of Homogeneity of Populations" problem.

How well does the observed data agree with the counts we expect if H₀ true?

1. What are the "expected counts" in these cells if H₀ true?
2. What is a good test statistic?
The Chi-Square Test Approach

Note. Examples (A), (B), and (C) illustrate 3 different problems:
- Testing Goodness-of-Fit of a Model (A)
- Testing Independence of 2 Variables (B)
- Testing Homogeneity of Several Populations with Respect to 1 Variable (C)

They are being treated in the following setting:

The data consists of "cell counts"

For problems having this type of data, we develop a Chi-Square Test as follows:

1. For each data cell, get an expected count if the null hypothesis $H_0$ is true.

2. For each cell, compute the quantity

\[
\frac{(\text{observed count} - \text{expected count})^2}{(\text{expected count})}
\]

3. Add these up to form the test statistic

\[
\chi^2 = \sum_{\text{all cells}} \left( \frac{(\text{observed count} - \text{expected count})^2}{(\text{expected count})} \right)
\]

4. **Fact** As total sample sizes increase, this test statistic has approximate $H_0$-distribution $\chi^2_{\nu}$, where the degrees of freedom $\nu$ is given by

- $(\text{number of cells} - 1)$ for Problem (A)
- $(\text{number of rows} - 1) \times (\text{number of columns} - 1)$ for Problems (B), (C)

Rule of Thumb

The sample sizes are big enough if each cell has expected count $\geq 5$.
Expected counts under $H_0$

(A) **Goodness-of-Fit Test**

Assume $G$ possible groups or categories and $n$ independent observations (trials).

Let $C_1, C_2, \ldots, C_G$ be the observed counts:

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$</td>
<td>$C_2$</td>
<td>$C_3$</td>
<td>$C_4$</td>
<td>$C_5$</td>
<td>$n$</td>
</tr>
</tbody>
</table>

$C_1 + C_2 + \cdots + C_G = n$

$H_0$: Group 1 has probability $\pi_{10}$

Group 2 -------- $\pi_{20}$

$\vdots$

Group $G$ -------- $\pi_{G0}$

($\pi_{10} + \pi_{20} + \cdots + \pi_{G0} = 1$)

$H_0$: Expected count for $i$-th cell

$$\text{(No. of trials)} \times \left( \text{probability of i-th cell} \right) = n\pi_{i0}$$

Test statistic:

$$\sum_{i=1}^{G} \frac{(C_i - n\pi_{i0})^2}{n\pi_{i0}}$$

Reject $H_0$ for large values of $\chi^2$

$H_0$: distribution of $\chi^2$:

$$\chi^2 \sim G-1$$

\( \ Use \ this \ to \ get \ a \ P-value \)
(B) Test of Independence

& (C) Test of Homogeneity

Assume a 2-way table of counts with \( R \) rows and \( C \) columns

Let \( O_{ij} \) be the observed count in cell \((ij)\):

For either \( H_0 \) (Independence or Homogeneity), the \( H_0 \)-Expected Count for the \((ij)\) cell is

\[
E_{ij} = \frac{O_{i\cdot} \times O_{\cdot j}}{O_{\cdot \cdot}}
\]

Test statistic: \( \chi^2 = \sum \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \)

Reject \( H_0 \) for large \( \chi^2 \) values

Relevant \( H_0 \)-distribution: \( \chi^2 \sim (R-1)(C-1) \)

(use to get \( P \)-value)
Setting

- A population grouped into $G$ groups or categories
- group probabilities: $\pi_1, \pi_2, \ldots, \pi_G$ ($\pi_1 + \pi_2 + \cdots + \pi_G = 1$)
- A random sample $X_1, \ldots, X_n$ from the parent distribution, grouped and counted:

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>$\ldots$</th>
<th>Group $G$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_1$</td>
<td>$O_2$</td>
<td>$\ldots$</td>
<td>$O_n$</td>
<td>$n$</td>
</tr>
</tbody>
</table>

(Here $O_1 + O_2 + \cdots O_G = n$.)

NOTES. (i) The probability model for the vector of counts $(O_1, O_2, \ldots, O_n)$ is $\text{Multinomial}(n; \pi_1, \pi_2, \ldots, \pi_G)$.

(ii) The model for any single count $O_g$ is $\text{Binomial}(n; \pi_g)$.

(iii) In the case of two categories ($G = 2$), this multinomial model is just a binomial model for two groups ("success" and "failure"). In this case, we just list the probability $\pi_1$ for "success", leaving understood the probability $\pi_1 = 1 - \pi_1$ for "failure".

Goal

Test a specified set of values for the unknown $\pi_1, \pi_2, \ldots, \pi_G$:

$H_0 : (\pi_1, \pi_2, \ldots, \pi_G) = (\pi_{10}, \pi_{20}, \ldots, \pi_{G0})$ (specified).

Method

(a) The Chi-square goodness of fit test statistic is developed as follows.

(i) Calculate the expected counts for the $G$ groups (cells), when $H_0$ is true: $E_g = n\pi_{g0}$, for $1 \leq g \leq G$. Here we use the fact that under $H_0$ the count $O_g$ is $\text{Binomial}(n, \pi_{g0})$.

(ii) Compare with observed counts, via the test statistic

$$\chi^2 = \sum_{g=1}^{G} \frac{(O_g - E_g)^2}{E_g}.$$  

(b) The $H_0$-distribution of $\chi^2$. Under $H_0$, this test statistic is distributed approximately as $\chi^2_{G-1}$, as the sample size $n$ increases.

(c) Get the p-value using (b).

(d) Interpret lower p-value as stronger evidence against $H_0$.

(e) For a test of $H_0$ at significance level $\alpha$, reject $H_0$ if p-value $\leq \alpha$ and otherwise accept $H_0$.

- RJS, 4/11/2011
Illustration for Example (A), p. 1

$$G = 6 \quad n = 60$$

$H_0: \pi_{10} = \frac{1}{6}, \pi_{20} = \frac{1}{6}, \ldots, \pi_{60} = \frac{1}{6}$

$H_0$-expected counts: $n \times \pi_{10} = 60 \times \frac{1}{6} = 10$, each $\pi_{10}$

(Since all $\pi_{10}$ are the same)

Test statistic: $\chi^2 = \frac{(4-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(17-10)^2}{10} + \frac{(16-10)^2}{10} + \frac{(18-10)^2}{10} + \frac{(9-10)^2}{10}$

$$= \frac{36}{10} + \frac{16}{10} + \frac{49}{10} + \frac{36}{10} + \frac{8}{10} + \frac{1}{10}$$

$$= 14.32$$

Relevant $H_0$-distribution: $\chi^2_5$

$p$-value for this data:

From the text's table for $\chi^2_5$

- the upper $1\%$ cutoff point is 15.09
- the upper $5\%$ cutoff point is 11.07

Our value 14.32 lies between these.

Conclusion: Can reject $H_0$ at the $5\%$ significance level but not at the $1\%$ significance level. The $p$-value is between 0.05 and 0.01

Exact $p$-value: 0.015

I used the "Chi Square Calculator" at

http://www.stat.tamu.edu/~west/applets/chisqdemo.html
Illustration for Example (B), p. 2

\[ R = 3, \ C = 4 \]

**H₀-expected counts**

\[
\begin{array}{ccc}
11 & 19.25 & 13.75 \\
6 & 10.50 & 7.50 \\
3 & 5.25 & 3.75 \\
\end{array}
\]

Test statistic:

\[
\chi^2 = \frac{(20-11)^2}{11} + \frac{(10-19.25)^2}{19.25} + \frac{(5-13.75)^2}{13.75} + \frac{(20-11)^2}{11} \\
+ \frac{(0-6)^2}{6} + \frac{(10-10.50)^2}{10.50} + \frac{(20-7.50)^2}{7.50} + \frac{(0-6)^2}{6} \\
+ \cdots + \frac{(0-3)^2}{3}
\]

\[
= 1316
\]

Relevant H₀-distribution: \[
\chi^2 \sim \chi^2 \text{ with } (3-1)(4-1) = 6
\]

p-value: very small

Illustration for Example (C), p. 3

\[ R = 3, \ C = 4 \]

**H₀-expected counts**

\[
\begin{array}{cccc}
52.6 & 52.6 & 52.6 & 21.1 \\
41.5 & 41.5 & 41.5 & 16.6 \\
5.9 & 5.9 & 5.9 & 2.35 \\
\end{array}
\]

Exercise: complete the details 😊
Illustration of Goodness-of-Fit Test:
Sweet pea experiment of Bateson and Punnett, 1906

- This landmark experiment in genetics investigated whether, for a certain kind of sweet pea plant, the traits flower color and pollen grain type are inherited independently ("independent segregation") or not.

In other words, this was a test of Mendelian theory of genetic inheritance, which said that traits are inherited independently of each other.

- Setting:

1. Flower color: Purple (P) or Red (Re) genes

   The trait "flower color" as seen by viewing the plant is determined by a gene pair in which "P" is dominant over "Re". The possibilities are:

   \[
   \begin{align*}
   PP & \quad \text{seen as purple} \\
   PRe & \\
   ReP & \\
   ReRe & \quad \text{seen as red}
   \end{align*}
   \]

   \[\text{The 2nd component from parent 2 through pollination}\]

   \[\text{The 1st component from parent 1}\]

2. Grain type: Long (L) dominant over Round (Ro):

   \[
   \begin{align*}
   LL & \quad \text{seen as long} \\
   LRo & \\
   RoL & \\
   RoRo & \quad \text{seen as round}
   \end{align*}
   \]
Method: Bateson and Punnett generated \( n = 256 \)

Second generation hybrid sweet pea plants, starting with two pure parent types:

- (i) pure purple (PP), pure long (LL)
- (ii) pure red (ReRe), pure round (RoRo)

The 2nd generation hybrid plants will fall into the categories:
- purple + long
- purple + round
- red + long
- red + round

with the Ho probabilities

\[ \begin{align*}
\Pi_{10} &= \frac{9}{16}, & \Pi_{20} &= \frac{3}{16}, & \Pi_{30} &= \frac{3}{16}, & \Pi_{40} &= \frac{1}{16}
\end{align*} \]

respectively. That is, if Ho: "independent segregation" is true, then these are the probabilities attached to these 4 categories.

Then the expected counts \( (E_i = n \cdot \Pi_{10}) \) are

\[
\begin{array}{c|c|c|c|c|c}
 & 1 & 2 & 3 & 4 & \text{Total} \\
\hline
144 & 48 & 48 & 16 & \sqrt{256 \times 9} \\
\end{array}
\]

The observed counts in the Bateson and Punnett experiment were

\[
\begin{array}{c|c|c|c|c|c}
 & 177 & 15 & 15 & 49 & 256 \\
\end{array}
\]

Test statistic: \( \chi^2 = \frac{(177-144)^2}{144} + \frac{(15-48)^2}{48} + \frac{(48-48)^2}{48} + \frac{(49-16)^2}{16} \)

\[ = 12.1 \] \( \approx 0 \) \( \Rightarrow \) evaluate using \( \chi^2 \) distribution

\( p\text{-value} \approx 0 \) \( \Rightarrow \) reject Ho. These genes are talking to each other!
- derivation of Ho-probability model

Initial Plants: \( PP + LL \) \( \times \) \( RR + RR \)

These are crossed to produce

1st Hybrid Generation: \( PRe + LRo \)

cross two of these to produce 2nd Hybrid Generation

\( PRe + LRo \) \( \times \) \( PRe + LRo \)

2nd Hybrid Generation:

\[
\begin{align*}
\text{purple} & \quad \begin{cases} 
PP & \frac{1}{4} \\
PRe & \frac{1}{4} \\
ReP & \frac{1}{4}
\end{cases} + \begin{cases} 
LL & \frac{1}{4} \\
LRo & \frac{1}{4} \\
RRo & \frac{1}{4}
\end{cases} \\
\text{red} & \quad \begin{cases} 
ReRe & \frac{1}{4}
\end{cases}
\end{align*}
\]

These probabilities are true whether or not Ho is true

The "independent segregation" hypothesis says that to get the probabilities for

<table>
<thead>
<tr>
<th>Event</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>purple + long</td>
<td>( \frac{3}{4} \times \frac{3}{4} = \frac{9}{16} )</td>
</tr>
<tr>
<td>purple + round</td>
<td>( \frac{3}{4} \times \frac{1}{4} = \frac{3}{16} )</td>
</tr>
<tr>
<td>red + long</td>
<td>( \frac{1}{4} \times \frac{3}{4} = \frac{3}{16} )</td>
</tr>
<tr>
<td>red + round</td>
<td>( \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} )</td>
</tr>
</tbody>
</table>

we multiply the associated probabilities

The Ho-probabilities test by the Chi-Square Test (previous page)
Carrying out the preceding procedure in Minitab

MTB > TChiSquare;
SUBC> Observed 177 15 15 49;
SUBC> GBar;
SUBC> GChiSQ;
SUBC> Pareto;
SUBC> RTable;
SUBC> Proportions .5625 .1875 .1875 .0625.

I did not need to type these commands. Minitab did. I simply used a drop-down menu.

Chi-Square Goodness-of-Fit Test for Observed Counts

<table>
<thead>
<tr>
<th>Category</th>
<th>Observed</th>
<th>Proportion</th>
<th>Expected to Chi-Sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>177</td>
<td>0.5625</td>
<td>144</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>0.1875</td>
<td>48</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>0.1875</td>
<td>48</td>
</tr>
<tr>
<td>4</td>
<td>49</td>
<td>0.0625</td>
<td>16</td>
</tr>
</tbody>
</table>

N DF Chi-Sq P-Value
256 3 121 0.000

Minitab calculated the 
N df.

The values of the terms
(obs-exp)^2/exp in \chi^2

Chart of Observed and Expected Values

Chart of Contribution to the Chi-Square Value by Category

The 4th cell is contributing the largest
\frac{(obs-exp)^2}{exp}
Illustration of Test of Independence: Are handedness and sex independent?

This follows the example treated in §28.4 of FPP

- Data (from the HANES study):

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right-handed</td>
<td>934</td>
<td>1070</td>
</tr>
<tr>
<td>Left-handed</td>
<td>113</td>
<td>92</td>
</tr>
<tr>
<td>Ambidextrous</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>1067</td>
<td>1170</td>
</tr>
</tbody>
</table>

Here we have a $3 \times 2$ classification of $N = 2237$ subjects by 2 variables, Handedness and Sex.

- Are these variables independent?
  - $H_0$: They are independent

- Expected counts under $H_0$:
  
  $E_{ij} = \frac{O_{..} \times O_{.j}}{O_{..}}$

<table>
<thead>
<tr>
<th></th>
<th>956</th>
<th>1048</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>98</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>15</td>
</tr>
</tbody>
</table>

- Test Statistic:

  $\chi^2 = \frac{(934-956)^2}{956} + \frac{(1070-1048)^2}{1048} + \frac{(113-98)^2}{98} + \frac{(92-107)^2}{107} + \frac{(20-13)^2}{13} + \frac{(8-15)^2}{15}$

  $\chi^2 = 11.806$

- Relevant $H_0$-distribution: $\chi^2_{(3)(2)-1} = \chi^2_2$

- $p$-value: 0.0027 $\Rightarrow$ Reject $H_0$

Women are right-handed at rate 91% versus 88% for men (estimated values)
Create in the Minitab worksheet 3 columns entitled “sex”, “handed”, “counts”, with entries as follows:

<table>
<thead>
<tr>
<th></th>
<th>C9</th>
<th>C10</th>
<th>C11</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sex</td>
<td>handed</td>
<td>counts</td>
</tr>
<tr>
<td>1</td>
<td>men</td>
<td>right</td>
<td>934</td>
</tr>
<tr>
<td>2</td>
<td>men</td>
<td>left</td>
<td>113</td>
</tr>
<tr>
<td>3</td>
<td>men</td>
<td>ambid</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>women</td>
<td>right</td>
<td>1070</td>
</tr>
<tr>
<td>5</td>
<td>women</td>
<td>left</td>
<td>92</td>
</tr>
<tr>
<td>6</td>
<td>women</td>
<td>ambid</td>
<td>8</td>
</tr>
</tbody>
</table>

MTB > XTABS 'handed' 'sex';
SUBC> Layout 1 1;
SUBC> Frequencies 'counts';
SUBC> Counts;
SUBC> DMissing 'handed' 'sex';
SUBC> ChiSquare;
SUBC> Expected;
SUBC> XResiduals.

**Tabulated statistics: handed, sex**

Using frequencies in counts

<table>
<thead>
<tr>
<th>Rows: handed</th>
<th>Columns: sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>men</td>
<td>women</td>
</tr>
<tr>
<td>ambid</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>13.4</td>
</tr>
<tr>
<td></td>
<td>3.306</td>
</tr>
<tr>
<td>left</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>97.8</td>
</tr>
<tr>
<td></td>
<td>2.369</td>
</tr>
<tr>
<td>right</td>
<td>934</td>
</tr>
<tr>
<td></td>
<td>955.9</td>
</tr>
<tr>
<td></td>
<td>0.500</td>
</tr>
<tr>
<td>All</td>
<td>1067</td>
</tr>
<tr>
<td></td>
<td>1067.0</td>
</tr>
<tr>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

Cell Contents: Count Expected count Contribution to Chi-square

Pearson Chi-Square = 11.866, DF = 2, P-Value = 0.003
Likelihood Ratio Chi-Square = 11.961, DF = 2, P-Value = 0.003
Illustration of Test of Homogeneity: Are 4 Ping-Pong Ball Sets Used in the California State Lottery Performing Equivalently?

This follows the example in Exercise A9 of Chap. 28 of FPP.

- **Setting**
  In the California State Lottery, the daily lottery is carried out as follows. A set of 10 Ping-Pong balls numbered 0 through 9 is placed in a glass bowl. They are mixed by an air jet, and one is forced out at random. It is replaced, and the process is repeated to generate a series of “random” digits from 0 to 9.

- **The Goal**: Test whether 4 different sets of 10 ping-pong balls behave the same.
  
  **H₀**: Sets A, B, C, D perform equivalently.

- **The Data**: For each set of balls, 120 draws are obtained from the mixing bowl (with replacement). For each set, the counts are recorded for the outcomes 0, ..., 9.

<table>
<thead>
<tr>
<th>Ball No.</th>
<th>Set A</th>
<th>Set B</th>
<th>Set C</th>
<th>Set D</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>13</td>
<td>22</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>8</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>7</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>8</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>19</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>20</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>10</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>19</td>
<td>11</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>6</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>16</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

| Total    | 120   | 120   | 120   | 120   |

**H₀**: The 4 Ball Sets are equivalent.

This is a test of homogeneity of the distributions on 0, ..., 9 associated with Ball Sets A, B, C, D.

Now let us let Minitab do the labor 😊.
Using Minitab, we get:

```
MTB > XTABS 'ball' 'set';
SUBC> Layout 1 1;
SUBC> Frequencies 'counts1';
SUBC> Counts;
SUBC> DMissing 'ball' 'set';
SUBC> ChiSquare;
SUBC> Expected;
SUBC> XResiduals.
```

### Tabulated statistics: ball, set
Using frequencies in counts1

<table>
<thead>
<tr>
<th>Rows: ball</th>
<th>Columns: set</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>13</td>
<td>22</td>
<td>12</td>
<td>16</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.75</td>
<td>15.75</td>
<td>15.75</td>
<td>15.75</td>
<td>63.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4802</td>
<td>2.4802</td>
<td>0.8929</td>
<td>0.0040</td>
<td>*</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>11</td>
<td>8</td>
<td>10</td>
<td>7</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.00</td>
<td>9.00</td>
<td>9.00</td>
<td>9.00</td>
<td>36.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4444</td>
<td>0.1111</td>
<td>0.1111</td>
<td>0.4444</td>
<td>*</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>16</td>
<td>7</td>
<td>14</td>
<td>12</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12.25</td>
<td>12.25</td>
<td>12.25</td>
<td>12.25</td>
<td>49.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.1480</td>
<td>2.2500</td>
<td>0.2500</td>
<td>0.0051</td>
<td>*</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>11</td>
<td>8</td>
<td>10</td>
<td>14</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.75</td>
<td>10.75</td>
<td>10.75</td>
<td>10.75</td>
<td>43.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0058</td>
<td>0.7035</td>
<td>0.0523</td>
<td>0.9826</td>
<td>*</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>5</td>
<td>19</td>
<td>11</td>
<td>15</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12.50</td>
<td>12.50</td>
<td>12.50</td>
<td>12.50</td>
<td>50.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.5000</td>
<td>3.3800</td>
<td>0.1800</td>
<td>0.5000</td>
<td>*</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>12</td>
<td>20</td>
<td>10</td>
<td>5</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.75</td>
<td>11.75</td>
<td>11.75</td>
<td>11.75</td>
<td>47.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0053</td>
<td>5.7926</td>
<td>0.2606</td>
<td>3.8777</td>
<td>*</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>12</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13.00</td>
<td>13.00</td>
<td>13.00</td>
<td>13.00</td>
<td>52.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0769</td>
<td>0.6923</td>
<td>3.7692</td>
<td>0.6923</td>
<td>*</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>19</td>
<td>11</td>
<td>12</td>
<td>21</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.75</td>
<td>15.75</td>
<td>15.75</td>
<td>15.75</td>
<td>63.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.6705</td>
<td>1.4325</td>
<td>0.8929</td>
<td>1.7500</td>
<td>*</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>5</td>
<td>6</td>
<td>12</td>
<td>11</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.50</td>
<td>8.50</td>
<td>8.50</td>
<td>8.50</td>
<td>34.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.4412</td>
<td>0.7353</td>
<td>1.4412</td>
<td>0.7353</td>
<td>*</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>16</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.75</td>
<td>10.75</td>
<td>10.75</td>
<td>10.75</td>
<td>43.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.5640</td>
<td>0.2849</td>
<td>0.2849</td>
<td>0.2849</td>
<td>*</td>
</tr>
<tr>
<td>All</td>
<td></td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>480</td>
</tr>
<tr>
<td></td>
<td></td>
<td>120.00</td>
<td>120.00</td>
<td>120.00</td>
<td>120.00</td>
<td>480.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Cell Contents: Count  
Expected count  
Contribution to Chi-square  

Pearson Chi-Square = 46.610, DF = 27, P-Value = 0.011

Reject Ho.

Conclusions: The ball sets are not the same.

Q. Which one is very suspicious?
- Having concluded that the 4 ball sets are not equivalent, how can we describe their differences?

Keeping in mind that a ball set performs "well" if it produces the digits 0, 1, ..., 9 with equal probability 1/10, let us test for each ball set the goodness-of-fit of this model.

1. Ball Set A

   \[ H_0: \pi_{10} = \pi_{20} = \ldots = \pi_{90} = \frac{1}{10} \]  
   \[ n = 120 \]
   \[ g = 10 \]

   Data:
   
   \[
   \begin{array}{cccccccc}
   13 & 11 & 16 & 11 & 5 & 12 & 12 & 19 & 5 & 16 & 120 \\
   \end{array}
   \]

   Expected counts:
   
   \[
   \begin{array}{cccccccc}
   12 & 12 & 12 & 12 & 12 & 12 & 12 & 12 & 12 & 120 \\
   \end{array}
   \]

   Test statistic:
   
   \[
   \chi^2 = \frac{(13-12)^2}{12} + \frac{(11-12)^2}{12} + \ldots + \frac{(16-12)^2}{12}
   \]
   
   \[
   = \frac{1}{12} \left( 1 + 1 + 16 + 1 + 49 + 0 + 0 + 49 + 49 + 49 + 16 \right)
   \]
   
   \[
   = \frac{182}{12} = \frac{15.17}{12}
   \]

   Relevant \( H_0 \)-distribution: \( \chi^2_{(10-1)} = \chi^2_9 \)

   \[
   \Rightarrow P-value = 0.0864 \Rightarrow \text{Accept} \; H_0
   \]

   Conclusion: This Ball Set performs well.
2. Ball Set B

Chi-Square Goodness-of-Fit Test for Observed Counts

<table>
<thead>
<tr>
<th>Category</th>
<th>Observed</th>
<th>Test Proportion</th>
<th>Expected</th>
<th>Contribution to Chi-Sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22</td>
<td>0.1</td>
<td>12</td>
<td>8.33333</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>0.1</td>
<td>12</td>
<td>1.33333</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>0.1</td>
<td>12</td>
<td>2.08333</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>0.1</td>
<td>12</td>
<td>1.33333</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>0.1</td>
<td>12</td>
<td>4.08333</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>0.1</td>
<td>12</td>
<td>5.33333</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>0.1</td>
<td>12</td>
<td>0.33333</td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td>0.1</td>
<td>12</td>
<td>0.08333</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>0.1</td>
<td>12</td>
<td>3.00000</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>0.1</td>
<td>12</td>
<td>0.75000</td>
</tr>
</tbody>
</table>

N DF Chi-Sq P-Value 0.002
120 9 26.6667

Reject H0 and Conclude Ball Set B does not perform well.
Chi-Square Goodness-of-Fit Test for Observed Counts

<table>
<thead>
<tr>
<th>Category</th>
<th>Observed</th>
<th>Test Proportion</th>
<th>Expected</th>
<th>Contribution to Chi-Sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>0.1</td>
<td>12</td>
<td>0.00000</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>0.1</td>
<td>12</td>
<td>0.33333</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>0.1</td>
<td>12</td>
<td>0.33333</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>0.1</td>
<td>12</td>
<td>0.33333</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>0.1</td>
<td>12</td>
<td>0.08333</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>0.1</td>
<td>12</td>
<td>0.33333</td>
</tr>
<tr>
<td>7</td>
<td>20</td>
<td>0.1</td>
<td>12</td>
<td>5.33333</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>0.1</td>
<td>12</td>
<td>0.00000</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>0.1</td>
<td>12</td>
<td>0.00000</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>0.1</td>
<td>12</td>
<td>0.75000</td>
</tr>
</tbody>
</table>

\[ N = 120, \text{ DF} = 9, \text{ Chi-Sq} = 7.5, F-Value = 0.585 \]

Accept \( H_0 \), and conclude Ball Set C does perform well.
Chi-Square Goodness-of-Fit Test for Observed Counts

<table>
<thead>
<tr>
<th>Category</th>
<th>Observed</th>
<th>Proportion</th>
<th>Expected</th>
<th>Contribution to Chi-Sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>0.1</td>
<td>12</td>
<td>1.33333</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>0.1</td>
<td>12</td>
<td>2.08333</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>0.1</td>
<td>12</td>
<td>0.00000</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>0.1</td>
<td>12</td>
<td>0.33333</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>0.1</td>
<td>12</td>
<td>0.75000</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>0.1</td>
<td>12</td>
<td>4.08333</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>0.1</td>
<td>12</td>
<td>0.33333</td>
</tr>
<tr>
<td>8</td>
<td>21</td>
<td>0.1</td>
<td>12</td>
<td>6.75000</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>0.1</td>
<td>12</td>
<td>0.08333</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>0.1</td>
<td>12</td>
<td>0.75000</td>
</tr>
</tbody>
</table>

N    DF  Chi-Sq  P-Value
120  9  16.5  0.057

Accept H₀ at 5% level, but just barely. Conclude Ball Set D OK but just barely.
Summary of Goodness-of-Fit Findings

Ball Sets A and C perform consistently with the "equally likely" model.

Ball Set B performs extremely inconsistently with this model.

Ball Set D performs marginally consistently with this model.

These differences among Ball Sets A, B, C, D help explain why the test of homogeneity resulted in "Reject Ho".

Ball Sets A and C are similar, and Ball Set D is marginally similar to these. However, Ball Set B is very different.

When Ho is rejected, we look for the explanation.