10 Discrete Data Analysis

SW Chapter 10

Earlier this semester we discussed inference for a single proportion problem. In this section we will generalize those methods in two directions. First we consider single sample problems involving categorical variables with multiple categories. Second, we consider problems with two or more samples.

Goodness-of-Fit Tests

Example The following data set was used as evidence in a court case. The data represent a sample of 1336 individuals from the jury pool of a large municipal court district for the years 1975-1977. The fairness of the representation of various age groups on juries was being contested. The strategy for doing this was to challenge the representativeness of the pool of individuals from which the juries are drawn. This was done by comparing the age group distribution within the jury pool against the age distribution in the district as a whole, which was available from census figures.

Age group (yrs)	Obs. Counts	Obs. Prop.	Census Prop.
18-19	23	.017	.061
20-24	96	.072	.150
25-29	134	.100	.135
30-39	293	.219	.217
40-49	297	.222	.153
50-64	380	.284	.182
65-99	113	.085	.102

A statistical question here is whether the jury pool population proportions are equal to the census proportions across the age categories. This comparison can be formulated as a **goodness-of-fit test**, which generalizes the large sample test on a single proportion to a categorical variable (here age) with $r > 2$ levels. For $r = 2$ categories, the goodness-of-fit test and large sample test on a single proportion are identical. Although this problem compares two populations, only one sample is involved because the census data is a population summary!

In general, suppose each individual in a population is categorized into one and only one of r levels or categories. Let $p_1,\ p_2, ..., p_r$ be the population proportions in the r categories, where each $p_i \geq 0$ and $p_1 + p_2 + \cdots + p_r = 1$. The hypotheses of interest in a goodness-of-fit problem are $H_0 : p_1 = p_1^0, p_2 = p_2^0, ..., p_r = p_r^0$ and H_A : not H_0, where $p_1^0,\ p_2^0, ..., p_r^0$ are given category proportions.

The plausibility of H_0 is evaluated by comparing the hypothesized category proportions to estimated (i.e. observed) category proportions $\hat{p}_1,\ \hat{p}_2,\ ..., \hat{p}_r$ from a random or representative sample of n individuals selected from the population. The discrepancy between the hypothesized and observed proportions is measured by the Pearson chi-squared statistic:

$$\chi_s^2 = \sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i},$$

where O_i is the **observed** number in the sample that fall into the i^{th} category ($O_i = n\hat{p}_i$), and $E_i = np_i^0$ is the number of individuals **expected** to be in the i^{th} category when H_0 is true.

The Pearson statistic can also be computed as the sum of the squared residuals:

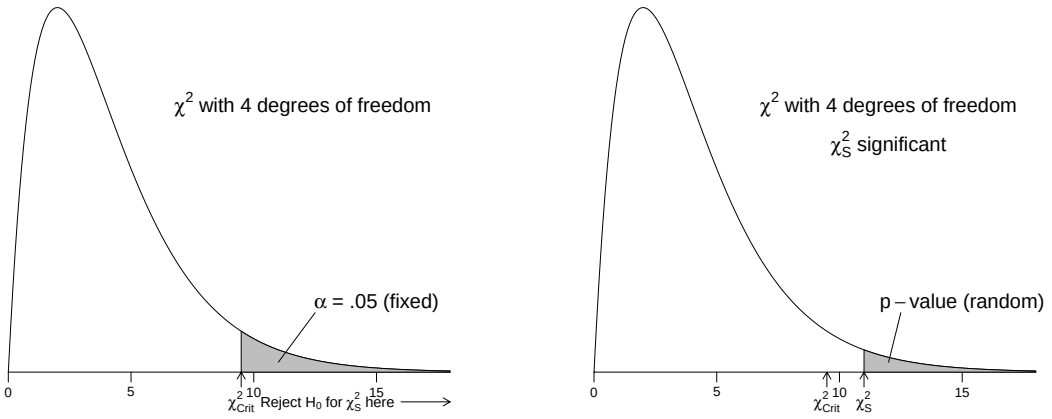
$$\chi_s^2 = \sum_{i=1}^{r} Z_i^2,$$

where $Z_i = (O_i - E_i)/\sqrt{E_i}$, or in terms of the observed and hypothesized category proportions

$$\chi_s^2 = n \sum_{i=1}^{r} \frac{(\hat{p}_i - p_i^0)^2}{p_i^0}.$$

The Pearson statistic χ_s^2 is "small" when all of the observed counts (proportions) are close to the expected counts (proportions). The Pearson χ^2 is "large" when one or more observed counts (proportions) differs noticeably from what is expected when H_0 is true. Put another way, large values of χ_s^2 suggest that H_0 is false.

The critical value χ^2_{crit} for the test is obtained from a chi-squared probability table with $r - 1$ degrees of freedom. A chi-squared table is given on page 686 of SW. The picture below shows the form of the rejection region. For example, if $r = 5$ and $\alpha = .05$, then you reject H_0 when $\chi_s^2 \geq \chi^2_{crit} = 9.49$. The p-value for the test is the area under the chi-squared curve with $df = r - 1$ to the right of the observed χ_s^2 value.



Example (Jury pool problem) Let p_{18} be the proportion in the jury pool population between ages 18 and 19. Define p_{20}, p_{25}, p_{30}, p_{40}, p_{50} and p_{65} analogously. You are interested in testing $H_0 : p_{18} = .061$, $p_{20} = .150$, $p_{25} = .135$, $p_{30} = .217$, $p_{40} = .153$, $p_{50} = .182$ and $p_{65} = .102$ against H_A : not H_0, using the sample of 1336 from the jury pool.

The observed counts, the expected counts, and the category residuals are given in the table below. For example, $E_{18} = 1336 * (.061) = 81.5$ and $Z_{18} = (23 - 81.5)/\sqrt{81.5} = -6.48$ in the 18-19 year category.

The Pearson statistic is

$$\chi_s^2 = (-6.48)^2 + (-7.38)^2 + (-3.45)^2 + .18^2 + 6.48^2 + 8.78^2 + (-1.99)^2 = 231.26$$

on $r - 1 = 7 - 1 = 6$ degrees of freedom. Here $\chi^2_{crit} = 12.59$ at $\alpha = .05$. The p-value for the goodness-of-fit test is less than .001, which suggests that H_0 is false.

Age group (yrs)	Obs. Counts	Exp. Counts	Residual
18-19	23	81.5	-6.48
20-24	96	200.4	-7.38
25-29	134	180.4	-3.45
30-39	293	289.9	0.18
40-49	297	204.4	6.48
50-64	380	243.2	8.78
65-99	113	136.3	-1.99

Adequacy of the Goodness-of-Fit Test

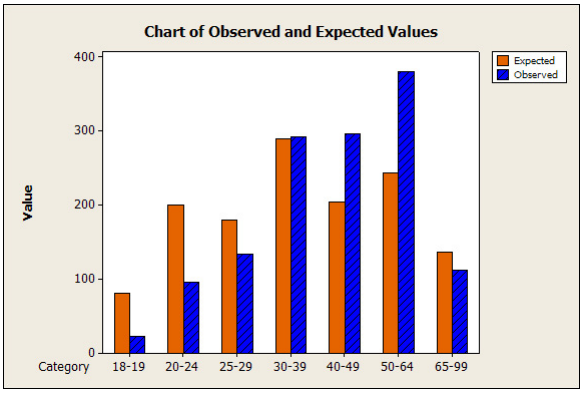
The chi-squared goodness-of-fit test is a large sample test. A conservative rule of thumb is that the test is suitable when each **expected** count is at least five. This holds in the jury pool example. There is no widely available alternative method for testing goodness-of-fit with smaller sample sizes. There is evidence, however, that the chi-squared test is **slightly conservative** (the p-values are too large, on average) when the expected counts are smaller. Some statisticians recommend that the chi-squared approximation be used when the minimum expected count is at least one, provided the expected counts are not too variable.

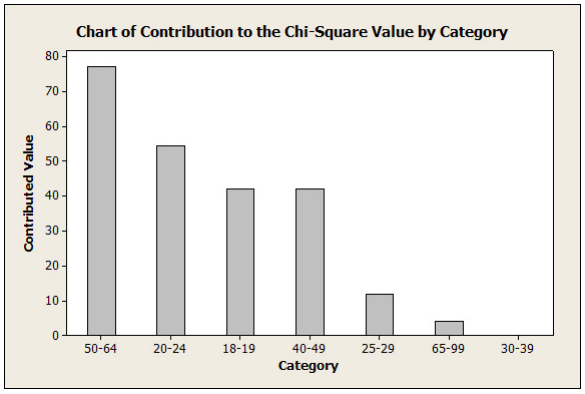
Minitab Implementation

Minitab will do a chi-squared goodness-of-fit test in the by following the menu path `Stat > Tables > Chi-Square Goodness-of-Fit Test (One Variable)`. Unlike the method we used for a single proportion of entering summarized data from a dialog box, the summarized data need to be entered into the worksheet (having counts for categories is summarized data). Following is the Minitab output for the jury pool problem:

```
Data Display

Row  Age    Count  CensusProp
  1  18-19     23       0.061
  2  20-24     96       0.150
  3  25-29    134       0.135
  4  30-39    293       0.217
  5  40-49    297       0.153
  6  50-64    380       0.182
  7  65-99    113       0.102
```

```
Chi-Square Goodness-of-Fit Test for Observed Counts in Variable: Count

Using category names in Age

                                 Test                  Contribution
Category   Observed  Proportion  Expected     to Chi-Sq
18-19            23       0.061    81.496       41.9871
20-24            96       0.150   200.400       54.3880
25-29           134       0.135   180.360       11.9164
30-39           293       0.217   289.912        0.0329
40-49           297       0.153   204.408       41.9420
50-64           380       0.182   243.152       77.0192
65-99           113       0.102   136.272        3.9743

   N  DF   Chi-Sq  P-Value
1336   6  231.260    0.000
```





The term "Contribution to Chi-Square" refers to the values of $\frac{(O-E)^2}{E}$ for each category. χ_s^2 is the sum of those contributions.

Comparing Two Proportions: Independent Samples

The New Mexico state legislature is interested in how the proportion of registered voters that support Indian gaming differs between New Mexico and Colorado. Assuming neither population proportion is known, the state's statistician might recommend that the state conduct a survey of registered voters sampled independently from the two states, followed by a comparison of the sample proportions in favor of Indian gaming.

Statistical methods for comparing two proportions using independent samples can be formulated as follows. Let p_1 and p_2 be the proportion of populations 1 and 2, respectively, with the attribute of interest. Let $\hat{p}_1$ and $\hat{p}_2$ be the corresponding sample proportions, based on independent random or representative samples of size n_1 and n_2 from the two populations.

Large Sample CI and Tests for $p_1 - p_2$

A large sample CI for $p_1 - p_2$ is $(\hat{p}_1 - \hat{p}_2) \pm z_{crit} SE_{CI}(\hat{p}_1 - \hat{p}_2)$, where z_{crit} is the standard normal critical value for the desired confidence level, and

$$SE_{CI}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

is the CI standard error.

A large sample p-value for a test of the null hypothesis $H_0 : p_1 - p_2 = 0$ against the two-sided alternative $H_A : p_1 - p_2 \neq 0$ is evaluated using tail areas of the standard normal distribution (identical to 1 sample evaluation) in conjunction with the test statistic

$$z_s = \frac{\hat{p}_1 - \hat{p}_2}{SE_{test}(\hat{p}_1 - \hat{p}_2)},$$

where

$$SE_{test}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}} = \sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

is the test standard error for $\hat{p}_1 - \hat{p}_2$. The **pooled proportion**

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

is the proportion of successes in the two samples combined. The test standard error has the same functional form as the CI standard error, with $\bar{p}$ replacing the individual sample proportions.

The pooled proportion is the best guess at the common population proportion when $H_0 : p_1 = p_2$ is true. The test standard error estimates the standard deviation of $\hat{p}_1 - \hat{p}_2$ assuming H_0 is true.

Example Two hundred and seventy nine French skiers were studied during two one-week periods in 1961. One group of 140 skiers receiving a placebo each day, and the other 139 receiving 1 gram of ascorbic acid (Vitamin C) per day. The study was double blind - neither the subjects nor the researchers knew who received what treatment. Let p_1 be the probability that a member of the ascorbic acid group contracts a cold during the study period, and p_2 be the corresponding probability for the placebo group. Linus Pauling and I are interested in testing whether $p_1 = p_2$. The data are summarized below as a two-by-two table of counts (a contingency table)

Outcome	Ascorbic Acid	Placebo
# with cold	17	31
# with no cold	122	109
Totals	139	140

The sample sizes are $n_1 = 139$ and $n_2 = 140$. The sample proportion of skiers developing colds in the placebo and treatment groups are $\hat{p}_2 = 31/140 = .221$ and $\hat{p}_1 = 17/139 = .122$, respectively. The pooled proportion is the number of skiers that developed colds divided by the number of skiers in the study: $\bar{p} = 48/279 = .172$.

The test standard error is:

$$SE_{test}(\hat{p}_1 - \hat{p}_2) = \sqrt{.172 * (1 - .172)\left(\frac{1}{139} + \frac{1}{140}\right)} = .0452.$$

The test statistic is

$$z_s = \frac{.122 - .221}{.0452} = -2.19.$$

The p-value for a two-sided test is twice the area under the standard normal curve to the right of 2.19 (or twice the area to the left of -2.19), which is $2 * (.014) = .028$ At the 5% level, we reject the

hypothesis that the probability of contracting a cold is the same whether you are given a placebo or Vitamin C.

A CI for $p_1 - p_2$ provides a measure of the size of the treatment effect. For a 95% CI

$$z_{crit}SE_{CI}(\hat{p}_1 - \hat{p}_2) = 1.96\sqrt{\frac{.221 * (1 - .221)}{140} + \frac{.122 * (1 - .122)}{139}} = 1.96 * (.04472) = .088.$$

The 95% CI for $p_1 - p_2$ is $(.122 - .221) \pm .088$, or $(-.187, -.011)$. We are 95% confident that p_2 exceeds p_1 by at least .011 but not by more than .187.

On the surface, we would conclude that a daily dose of Vitamin C decreases a French skier's chance of developing a cold by between .011 and .187 (with 95% confidence). This conclusion was somewhat controversial. Several reviews of the study felt that the experimenter's evaluations of cold symptoms were unreliable. Many other studies refute the benefit of Vitamin C as a treatment for the common cold.

Example A case-control study was designed to examine risk factors for cervical dysplasia (Becker et al. 1994). All the women in the study were patients at UNM clinics. The 175 cases were women, aged 18-40, who had cervical dysplasia. The 308 controls were women aged 18-40 who did not have cervical dysplasia. Each women was classified as positive or negative, depending on the presence of HPV (human papilloma virus).

The data are summarized below.

HPV Outcome	Cases	Controls
Positive	164	130
Negative	11	178
Sample size	175	308

Let p_1 be the probability that a case is HPV positive and let p_2 be the probability that a control is HPV positive. The sample sizes are $n_1 = 175$ and $n_2 = 308$. The sample proportions of positive cases and controls are $\hat{p}_1 = 164/175 = .937$ and $\hat{p}_2 = 130/308 = .422$.

For a 95% CI

$$z_{crit}SE_{CI}(\hat{p}_1 - \hat{p}_2) = 1.96\sqrt{\frac{.937 * (1 - .937)}{175} + \frac{.422 * (1 - .422)}{308}} = 1.96 * (.03336) = .0659.$$

A 95% CI for $p_1 - p_2$ is $(.937 - .422) \pm .066$, or $.515 \pm .066$, or $(.449, .581)$. I am 95% confident that p_1 exceeds p_2 by at least .45 but not by more than .58.

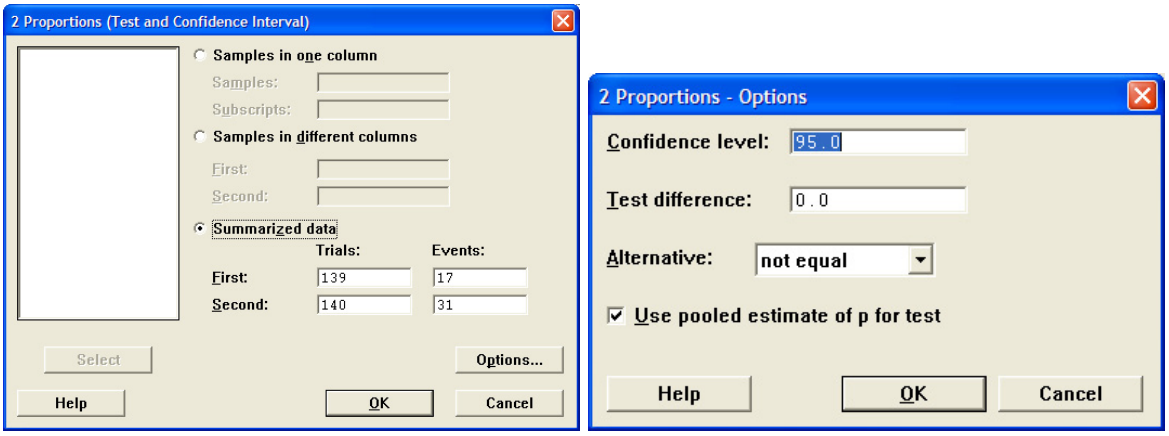
Not surprisingly, a two-sided test at the 5% level would reject $H_0 : p_1 = p_2$. In this problem one might wish to do a one-sided test, instead of a two-sided test. Let us carry out this test, as a refresher on how to conduct one-sided tests.

Appropriateness of Large Sample Test and CI

The standard two sample CI and test used above are appropriate when each sample is large. A rule of thumb suggests a minimum of at least five successes (i.e. observations with the characteristic of interest) and failures (i.e. observations without the characteristic of interest) in each sample before using these methods. This condition is satisfied in our two examples.

Minitab Implementation

For the Vitamin C example, in order to get Minitab to do all the calculations as presented, it is easiest to follow the menu path `Stat > Basic Statistics > 2 Proportions` and enter summary data as follows (you need to check the box for pooled estimate of p for test).




```
Test and CI for Two Proportions

Sample   X    N  Sample p
1       17  139  0.122302
2       31  140  0.221429


Difference = p (1) - p (2)
Estimate for difference:  -0.0991264
95% CI for difference:  (-0.186859, -0.0113937)
Test for difference = 0 (vs not = 0):  Z = -2.19  P-Value = 0.028
```

For the cervical dysplasia example, Minitab results are as follows:

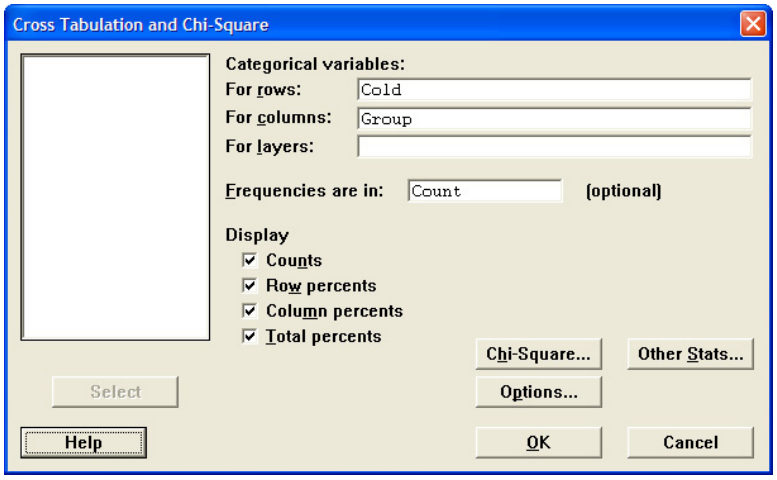
```
Test and CI for Two Proportions

Sample    X    N  Sample p
1       164  175  0.937143
2       130  308  0.422078


Difference = p (1) - p (2)
Estimate for difference:  0.515065
95% CI for difference:  (0.449221, 0.580909)
Test for difference = 0 (vs not = 0):  Z = 11.15  P-Value = 0.000
```

The above analyses are not the most common way to see data like this presented. The ability to get a confidence interval is particularly nice, and I do recommend including such an analysis. Usually, though, we present such data as a two-by-two contingency table. We need this structure in the rest of this section, so let us do that for these two examples.

The basic structure of data entry (it must be in the worksheet) is similar to our earlier use of stacked data. This is how SAS, Stata, and most other packages want it as well. For the Vitamin C example, the data are entered as follows:

```
Data Display

Row  Cold  Group      Count
  1  1Yes  1Vit C        17
  2  1Yes  2Placebo      31
  3  2No   1Vit C       122
  4  2No   2Placebo     109
```

The values for Cold could be entered as just Yes and No, but then Minitab alphabetizes in the presentation. What I have done is one way to get Minitab to present the table in the order we want it. Now we follow the menu path `Stat > Tables > Cross Tabulation and Chi-Square` and fill in the following box appropriately:



The various Display options and Other Stats are reflected in the following output. I structured this to present what I usually get out of SAS by default.

```
Tabulated statistics: Cold, Group

Using frequencies in Count


Rows: Cold   Columns: Group

         1Vit C  2Placebo     All

1Yes         17        31      48
          35.42     64.58  100.00
          12.23     22.14   17.20
           6.09     11.11   17.20
           23.9      24.1    48.0
         1.9990    1.9847       *

2No         122       109     231
          52.81     47.19  100.00
          87.77     77.86   82.80
          43.73     39.07   82.80
          115.1     115.9   231.0
         0.4154    0.4124       *

All         139       140     279
          49.82     50.18  100.00
         100.00    100.00  100.00
          49.82     50.18  100.00
          139.0     140.0   279.0
```

```
            *         *       *

 Cell Contents:      Count
                     % of Row
                     % of Column
                     % of Total
                    Expected count
                    Contribution to Chi-square

 Pearson Chi-Square = 4.811, DF = 1, P-Value = 0.028
 Likelihood Ratio Chi-Square = 4.872, DF = 1, P-Value = 0.027

 Fisher's exact test: P-Value =  0.0384925
```

The Pearson $\chi^2_s = 4.811$ is just the square of $Z_s = -2.19$, so for this case it's really an identical test (only for the two-sided hypothesis, though). The Likelihood Ratio Chi-Square is another large-sample test. Fisher's Exact test is another test that does not need large samples - I use it in practice very frequently. Minitab only performs this test for two-by-two tables — for more complicated tables, this is can be a very hard test to compute. SAS and Stata will at least try to compute it for arbitrary tables, though they do not always succeed. Let us examine the output to see what all these terms mean.

For the cervical dysplasia data, the results are:

```
Data Display

Row  HPV   Group    Count
  1  1Pos  Case       164
  2  1Pos  Control    130
  3  2Neg  Case        11
  4  2Neg  Control    178

 Tabulated statistics: HPV, Group

 Using frequencies in Count


 Rows: HPV   Columns: Group

            Case  Control     All

 1Pos        164      130     294
           55.78    44.22  100.00
           93.71    42.21   60.87
           33.95    26.92   60.87
           106.5    187.5   294.0
           31.01    17.62       *

 2Neg         11      178     189
            5.82    94.18  100.00
            6.29    57.79   39.13
            2.28    36.85   39.13
            68.5    120.5   189.0
           48.25    27.41       *

 All         175      308     483
           36.23    63.77  100.00
           100.00   100.00  100.00
           36.23    63.77  100.00
           175.0    308.0   483.0
```

```
             *        *       *

Cell Contents:       Count
                     % of Row
                     % of Column
                     % of Total
                     Expected count
                     Contribution to Chi-square


Pearson Chi-Square = 124.294, DF = 1, P-Value = 0.000
Likelihood Ratio Chi-Square = 144.938, DF = 1, P-Value = 0.000

Fisher's exact test: P-Value =  0.0000000
```

Effect Measures in Two-by-Two Tables

Consider a study of a particular disease, where each individual is either exposed or not-exposed to a risk factor. Let p_1 be the proportion diseased among the individuals in the exposed population, and p_2 be the proportion diseased among the non-exposed population. This population information can be summarized as a two-by-two table of population proportions:

Outcome	Exposed population	Non-Exposed population
Diseased	p_1	p_2
Non-Diseased	$1 - p_1$	$1 - p_2$

A standard measure of the difference between the exposed and non-exposed populations is the **absolute difference:** $p_1 - p_2$. We have discussed statistical methods for assessing this difference.

In many epidemiological and biostatistical settings, other measures of the difference between populations are considered. For example, the relative risk

$$\text{RR} = \frac{p_1}{p_2}$$

is commonly reported when the individual risks p_1 and p_2 are small. The odds ratio

$$\text{OR} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

is another standard measure. Here $p_1/(1-p_1)$ is the odds of being diseased in the exposed group, whereas $p_2/(1-p_2)$ is the odds of being diseased in the non-exposed group.

We will discuss these measures more completely next semester. At this time I will note that each of these measures can be easily estimated from data, using the sample proportions as estimates of the unknown population proportions. For example, in the vitamin C study:

Outcome	Ascorbic Acid	Placebo
# with cold	17	31
# with no cold	122	109
Totals	139	140

the proportion with colds in the placebo group is $\hat{p}_2 = 31/140 = .221$. The proportion with colds in the vitamin C group is $\hat{p}_1 = 17/139 = .122$.

The estimated absolute difference in risk is $\hat{p}_1 - \hat{p}_2 = .122 - .221 = -.099$. The estimated risk ratio and odds ratio are

$$\hat{RR} = \frac{.122}{.221} = .55$$

and

$$\hat{OR} = \frac{.122/(1-.122)}{.221/(1-.221)} = .49,$$

respectively.

Testing for Homogeneity of Proportions

Example The following two-way table of counts summarizes the location of death and age at death from a study of 2989 cancer deaths (Public Health Reports, 1983):

(Obs Counts)	Location of death			
Age	Home	Acute Care	Chronic care	Row Total
15-54	94	418	23	535
55-64	116	524	34	674
65-74	156	581	109	846
75+	138	558	238	934
Col Total	504	2081	404	2989

The researchers want to compare the age distributions across locations. A one-way ANOVA would be ideal if the actual ages were given. Because the ages are grouped, the data should be treated as categorical. Given the differences in numbers that died at the three types of facilities, a comparison of proportions or percentages in the age groups is appropriate. A comparison of counts is not.

The table below summarizes the proportion in the four age groups at each location. For example, in the acute care facility $418/2081 = .201$ and $558/2081 = .268$. The **pooled proportions** are the Row Totals divided by the total sample size of 2989. The pooled summary gives the proportions in the four age categories, ignoring location of death.

The age distributions for home and for the acute care facilities are similar, but are very different from the age distribution at chronic care facilities.

To formally compare the observed proportions, one might view the data as representative sample of ages at death from the three locations. Assuming independent samples from the three locations (populations), a chi-squared statistic is used to test whether the population proportions of ages at death are identical (homogeneous) across locations. The **chi-squared test for homogeneity** of population proportions can be defined in terms of proportions, but is traditionally defined in terms of counts.

(Proportions)	Location of death			
Age	Home	Acute Care	Chronic care	Pooled
15-54	.187	.201	.057	.179
55-64	.230	.252	.084	.226
65-74	.310	.279	.270	.283
75+	.273	.268	.589	.312
Total	1.000	1.000	1.000	1.000

In general, assume that the data are independent samples from c populations (strata, groups, sub-populations), and that each individual is placed into one of r levels of a categorical variable. The raw data will be summarized as a $r \times c$ **contingency table** of counts, where the columns correspond to the samples, and the rows are the levels of the categorical variable. In the age distribution problem, $r = 4$ and $c = 3$. (SW uses k to identify the number of columns.)

To implement the test:

1. Compute the (estimated) **expected** count for each cell in the table as follows:

$$\mathrm{E} = \frac{\text{Row Total} * \text{Column Total}}{\text{Total Sample Size}}.$$

2. Compute the Pearson test statistic

$$\chi^2_s = \sum_{\text{all cells}} \frac{(O-E)^2}{E},$$

   where O is the **observed** count.

3. For a size α test, reject the hypothesis of homogeneity if $\chi^2_s \geq \chi^2_{crit}$, where χ^2_{crit} is the upper α critical value from the chi-squared distribution with $df = (r-1)(c-1)$.

The p-value for the chi-squared test of homogeneity is equal to the area under the chi-squared curve to the right of X^2; see the picture on page 98.

For a two-by-two table of counts, the chi-squared test of homogeneity of proportions is identical to the two-sample proportion test we discussed earlier.

Minitab Analysis

Enter data as follows (just as for the two-by-two table):

```
Data Display

Row  Age    Care     Count
  1  15-54  1Home       94
  2  15-54  Acute      418
  3  15-54  Chronic     23
  4  55-64  1Home      116
  5  55-64  Acute      524
  6  55-64  Chronic     34
  7  65-74  1Home      156
  8  65-74  Acute      581
  9  65-74  Chronic    109
 10  75+    1Home      138
 11  75+    Acute      558
 12  75+    Chronic    238
```

Follow the same path `Stat > Tables > Cross Tabulation and Chi-Square` with these results:

```
Tabulated statistics: Age, Care

Using frequencies in Count

* NOTE * Fisher's exact test available only for 2 x 2 tables.

Rows: Age   Columns: Care

           1Home   Acute  Chronic     All

15-54         94     418       23     535
           17.57   78.13     4.30  100.00
           18.65   20.09     5.69   17.90
            3.14   13.98     0.77   17.90
            90.2   372.5     72.3   535.0
           0.159   5.564   33.627       *

55-64        116     524       34     674
           17.21   77.74     5.04  100.00
           23.02   25.18     8.42   22.55
            3.88   17.53     1.14   22.55
           113.6   469.3     91.1   674.0
           0.049   6.388   35.789       *

65-74        156     581      109     846
           18.44   68.68    12.88  100.00
           30.95   27.92    26.98   28.30
            5.22   19.44     3.65   28.30
           142.7   589.0    114.3   846.0
           1.249   0.109    0.250       *

75+          138     558      238     934
           14.78   59.74    25.48  100.00
           27.38   26.81    58.91   31.25
            4.62   18.67     7.96   31.25
           157.5   650.3    126.2   934.0
           2.412  13.092   98.937       *

All          504    2081      404    2989
           16.86   69.62    13.52  100.00
          100.00  100.00   100.00  100.00
           16.86   69.62    13.52  100.00
           504.0  2081.0    404.0  2989.0
               *       *        *       *

Cell Contents:      Count
                    % of Row
                    % of Column
                    % of Total
                    Expected count
                    Contribution to Chi-square


Pearson Chi-Square = 197.624, DF = 6, P-Value = 0.000
Likelihood Ratio Chi-Square = 200.972, DF = 6, P-Value = 0.000
```

The Pearson statistic and the likelihood ratio statistic, which is an alternative statistic for testing homogeneity, both report a p-value of 0 to three places. The data strongly suggest that there are differences in the age distributions among locations. The likelihood ratio statistic leads to the same conclusion. The various summaries help us to explain what is the nature of the differences.

Testing for Homogeneity in Cross-Sectional and Stratified Studies

Two-way tables of counts are often collected either by **stratified sampling** or by **cross-sectional** sampling.

In a stratified design, distinct groups, strata, or sub-populations are identified. Independent samples are selected from each group, and the sampled individuals are classified into categories. The HPV study is an illustration of a stratified design (and a case-control study). Stratified designs provide estimates for the strata (population) proportion in each of the categories. A test for **homogeneity of proportions** is used to compare the strata.

In a **cross-sectional design**, individuals are randomly selected from a population and classified by the levels of **two** categorical variables. With cross-sectional samples you can test homogeneity of proportions by comparing either the row proportions or by comparing the column proportions.

Example The following data (*The Journal of Advertising,* 1983, p. 34-42) are from a cross-sectional study that involved soliciting opinions on anti-smoking advertisements. Each subject was asked whether they smoked and their reaction (on a five-point ordinal scale) to the ad. The data are summarized as a two-way table of counts, given below:

	Str. Dislike	Dislike	Neutral	Like	Str. Like	Row Tot
Smoker	8	14	35	21	19	97
Non-smoker	31	42	78	61	69	281
Col Total	39	56	113	82	88	378

The row proportions are

(Row Prop)	Str. Dislike	Dislike	Neutral	Like	Str. Like	Row Tot
Smoker	.082	.144	.361	.216	.196	1.000
Non-smoker	.110	.149	.278	.217	.245	1.000

For example, the entry for the (Smoker, Str. Dislike) cell is: $8/97 = .082$.

Similarly, the column proportions are

(Col Prop)	Str. Dislike	Dislike	Neutral	Like	Str. Like
Smoker	.205	.250	.310	.256	.216
Non-smoker	.795	.750	.690	.744	.784
Total	1.000	1.000	1.000	1.000	1.000

Although it may be more natural to compare the smoker and non-smoker row proportions, the column proportions can be compared across ad responses. There is no advantage to comparing "rows" instead of "columns" in a formal test of homogeneity of proportions with cross-sectional data. The Pearson chi-squared test (and the likelihood ratio test) treats the rows and columns interchangeably, so you get the same result regardless of how you view the comparison. However, one of the two comparisons may be more natural to interpret.

Note that checking for homogeneity of proportions is meaningful in stratified studies only when the comparison is across strata! Further, if the strata correspond to columns of the table, then the column proportions or percentages are meaningful whereas the row proportions are not.
Question: How do these ideas apply to the age distribution problem?

Testing for Independence in a Two-Way Contingency Table

The row and column classifications for a population where each individual is cross-classified by two categorical variables are said to be independent if each **population** cell proportion in the two-way table is the product of the proportion in a given row and the proportion in a given column. One can show that independence is equivalent to homogeneity of proportions. In particular, the two-way table of population cell proportions satisfies independence if and only if the population column proportions are homogeneous. If the population column proportions are homogeneous then so are the population row proportions.

This suggests that a test for independence or **no association** between two variables based on a cross-sectional study can be implemented using the chi-squared test for homogeneity of proportions. This suggestion is correct. If independence is not plausible, I tend to interpret the dependence as a deviation from homogeneity, using the classification for which the interpretation is most natural.

Example

```
Data Display

Row  Smoker  Opinion         Count
  1  1Yes    1 Str. Dislike      8
  2  1Yes    2 Dislike          14
  3  1Yes    3 Neutral          35
  4  1Yes    4 Like             21
  5  1Yes    5 Str. Like        19
  6  No      1 Str. Dislike     31
  7  No      2 Dislike          42
  8  No      3 Neutral          78
  9  No      4 Like             61
 10  No      5 Str. Like        69

Tabulated statistics: Smoker, Opinion

Using frequencies in Count

* NOTE * Fisher's exact test available only for 2 x 2 tables.

Rows: Smoker   Columns: Opinion

         1 Str.
        Dislike  2 Dislike  3 Neutral   4 Like  5 Str. Like     All

1Yes          8         14         35       21           19      97
           8.25      14.43      36.08    21.65        19.59  100.00
          20.51      25.00      30.97    25.61        21.59   25.66
           2.12       3.70       9.26     5.56         5.03   25.66
          10.01      14.37      29.00    21.04        22.58   97.00
        0.40286    0.00955    1.24259  0.00009      0.56819       *

No           31         42         78       61           69     281
          11.03      14.95      27.76    21.71        24.56  100.00
          79.49      75.00      69.03    74.39        78.41   74.34
           8.20      11.11      20.63    16.14        18.25   74.34
```

```
          28.99      41.63      84.00      60.96      65.42  281.00
        0.13907    0.00330    0.42894    0.00003    0.19614       *

All          39         56        113         82         88     378
          10.32      14.81      29.89      21.69      23.28  100.00
         100.00     100.00     100.00     100.00     100.00  100.00
          10.32      14.81      29.89      21.69      23.28  100.00
          39.00      56.00     113.00      82.00      88.00  378.00
              *          *          *          *          *       *

Cell Contents:       Count
                     % of Row
                     % of Column
                     % of Total
                     Expected count
                     Contribution to Chi-square

Pearson Chi-Square = 2.991, DF = 4, P-Value = 0.559
Likelihood Ratio Chi-Square = 2.980, DF = 4, P-Value = 0.561
```

The Pearson chi-squared test is not significant (p-value = .561). The observed association between smoking status and the ad reaction is not significant. This suggests, for example, that the smoker's reactions to the ad were not statistically significantly different from the non-smoker's reactions, which is consistent with the smokers and non-smokers attitudes being fairly similar.