Chi-Square to Test Frequency Data

Example using Nested Frequency data given in class:

Question: has spotted knapweed increased significantly on the site examined in the "West Pass" allotment?

1st What size frame should be used to compare 1995 data with 2003? (I used the plot size 2 or 1/16th m^2 for this example.

2nd Prepare contingency table

	1995	2003	Totals
Present	8	15	23
Absent	12	5	17
Totals	20	20	40

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Where: χ^2 is the chi square statistic.
Σ = summation symbol.
O = Number observed.
E = Number expected.

3. Testing the difference between two proportions (independent samples): the chi-square test

The chi-square test is used to analyze frequency data when individual quadrats are the sampling units and point cover data when individual points are the sampling units. (Even though cover is expressed as a percentage, cover data are appropriately analyzed by calculating mean values, except when individual points are the sampling units.) If the frequency data are collected on more than one species, each species is usually analyzed separately. Another alternative is to lump species into functional groups, such as annual graminoids, and analyze each of the groups.

a. 2 x 2 contingency table to compare two years

To estimate the frequency of a plant species in two separate years, we've taken two independent random samples of 400 quadrats each. In each of these quadrats the species is either present or absent. For analysis we put these data into a 2 x 2 contingency table, as follows:

	1990	1994	Totals
Present	123 (0.31)	157 (0.39)	280 (0.35)
Absent	277 (0.69)	243 (0.61)	520 (0.65)
Totals	400 (1.00)	400 (1.00)	800 (1.00)

The numbers in parentheses are frequencies of occurrence in 1990 and 1994, and, in the last column, for both years combined. The chi-square test is conducted on actual numbers

http://www.blm.gov/nstc/library/techref.htm "Measuring and Monitoring Plant Populations"

of quadrats, *not* percentages. The chi-square test is not appropriately applied to percentage data.

Just as for the *t* test and ANOVA, we must formulate a null hypothesis. Our null hypothesis states that the true proportion of the target plant species (the proportion we would get if we placed all of the quadrats of our particular size that could be placed in the sampled area) is the same in both years. This is equivalent to saying there has been no change in the proportion of the key species from 1990 to 1994.

Before we can calculate the chi-square statistic we must determine the values that would be expected in the event there was no difference between years. The total frequencies in the right hand column are used for this purpose. Thus, in both 1990 and 1994, 0.35 x 400 quadrats, or 140 quadrats, would be expected to contain the species, and in both 1990 and 1994, 0.65 x 400 quadrats, or 260 quadrats, would be expected to not contain the species. The following table shows these expected values:

	1990	1994	Totals
Present	140	140	280
Absent	260	260	520
Totals	400	400	800

Now we can compute the chi square statistic as follows:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Where: χ^2 is the chi square statistic.
Σ = summation symbol.
O = Number observed.
E = Number expected.

Applying this formula to our example we get:

$$\chi^2 = \frac{(123\text{-}140)^2}{140} + \frac{(277\text{-}260)^2}{260} + \frac{(157\text{-}140)^2}{140} + \frac{(243\text{-}260)^2}{260}$$

$$= 2.06 + 1.11 + 2.06 + 1.11 = 6.34$$

We then compare the chi-square value of 6.34 to a table of critical values of the chi-square statistic (see table in Appendix 5) to see if our chi-square value is sufficiently large to be significant.[4] The *P* value we have selected for our threshold before sampling began is 0.10. Now we need to determine the number of degrees of freedom. For a contingency table, the number of degrees of freedom, ν, is given by:

$$\nu = (r - 1)(c - 1)$$

Where: r = number of rows in the contingency table.
c = number of columns in the contingency table.

For a 2 x 2 table $\nu = (2\text{-}1)(2\text{-}1) = 1$. Therefore, we enter the table at degrees of freedom = 1, and the *P* threshold of 0.10. The critical chi-square value from the table is 2.706. Since our value of 6.34 is larger than the critical value, we reject the null hypothesis of no difference in frequency of the plant species and conclude there has been an increase in its frequency. We would also report our calculated *P* value, which we could interpolate from the chi-square table, but could obtain more easily through a statistics program. For this example, the *P* value is 0.012.