Chi-Square

Big Picture

The chi-squared test is an important tool to find whether a table of data varies enough from the expected data to be significant.

Chi-Square Distribution

We use the chi-square distribution for goodness-of-fit test and test of independence. The chi-square (χ^2) test statistic is: $\chi^2 = \sum \frac{(O-E)^2}{E}$

- O is the observed frequency value for each event
- E is the expected frequency value for each event

The chi-square statistic will be small when the difference between the observed frequency and the expected frequency is small.

Contingency Table

Contingency tables are useful for using the chi-square distribution.

- Table list the variables and observational patterns that will help us run a chi-square test.
- Sometimes we calculate the expected frequency from the contingency table. To calculate the expected frequency:
  1. List the groups or subjects in rows.
  2. List the observations or responses in columns.

Example: To look at the relationship between gender and voting pattern, we might make a table like this one:

Frequency of California citizens voting for a Republican or Democratic candidate

	Democratic		Republican		Total
	Observed	Expected	Observed	Expected	Observed
Female	48	42.26	28	29.74	76
Male	36	37.74	26	24.26	62
Total	84		54		138

- Expected values calculated by using the equation $\frac{\text{(Row total)(Column total)}}{\text{Total number of observations}}$
- Degrees of freedom =(row-1)(column-1)

Goodness-of-Fit Test

- Test to see if the observed values of a categorical variable are different with the expected values of that same variable, checks how closely the sample matches expected distribution
  - For example, if all grades buy hot lunch equally, we would expect the proportion of juniors who buy hot lunch to be equal to the total proportion of students who buy hot lunch.
- Works by comparing sample frequencies with given population frequencies
- Can be used to evaluate hypotheses
  - Test proceeds based on assumption that H_0 is true
  - The higher the chi-value, the more justification to reject H_0
- Steps to create a goodness-of-fit test:
  1. Hypothesize
     - State null and alternative hypothesis
     - Choose alpha value
  2. Conditions
     - Check that the experiment was conducted with randomized subjects
     - Check that the number of each expected count ≥ 5
  3. Calculations
     - Create a contingency table that compares observed and expected frequencies (can also use proportions)
     - Expected frequency is found by multiplying the proportion of the population with the total sample number
     - Calculate the chi-squared statistic χ^2
     - Using the chi-squared statistic and the degrees of freedom (one less than the number of categories), calculate the P value
  4. Conclusion
     - Compare the P value with the alpha value to determine whether or not to reject the hypothesis
     - Write a statement expressing your conclusion

For goodness-of-fit tests, remember that we are always comparing ONE sample's proportions with a given population's proportions.

Chi-Square cont.

Test for Independence

- Test to see if there is an association between cateories
  - For example, to see if gender affects color preference.

Steps for a test for independence:

1. Hypothesize
   - State null and alternative hypothesis
   - Choose alpha level (near 0.05)
2. Check conditions
   - Check that the experiment was conducted with randomized subjects
   - Check that the number of each expected count ≥ 5
3. Calculations
   - Make a contingency table of given and expected values.
   - Calculate the chi-square value.
   - Using the chi-squared statistic and the degrees of freedom, calculate the *P* value
4. Conclusion
   - Compare the *P* value with the alpha value to determine whether or not to reject the hypothesis
   - Write a statement expressing your conclusion

For tests of independence, the question usually states explicitly to find whether two variables are independent/ related.

Test for Homogeneity

- The tests for independence and homogeneity are very similar, but the questions they try to answer are not. The test for homoegeneity is used to find if frequencies are distributed evenly throughout two different populations.
  - For example, do boys and girls watch TV equally?

Steps for a test for homogeneity:

1. Hypothesize
   - State null and alternative hypothesis
   - Choose alpha level (near 0.05)
2. Conditions
   - Check that the experiment was conducted with randomized subjects
   - Check that the number of each expected count ≥ 5
3. Calculations
   - Make a contingency table of given and expected values.
   - Calculate the chi-square value.
   - Using the chi-squared statistic and the degrees of freedom, calculate the *P* value
4. Conclusion
   - Compare the *P* value with the alpha value to determine whether or not to reject the hypothesis
   - Write a statement expressing your conclusion

For tests of homogeneity, population proportions are often given in a chart.

Notes