Tests of Homogeneity and Independence

Lecture 50
Sections 14.4 - 14.5

Robb T. Koether

Hampden-Sydney College

Wed, Apr 25, 2012

Outline

Outline

Homogeneous Populations

- Now we wish to compare two or more populations by observing the values of a qualitative variable for each of them.

Homogeneous Populations

- Now we wish to compare two or more populations by observing the values of a qualitative variable for each of them.
- For example, is there any difference of voters approval or disapproval of the president based on their political ideology?

Homogeneous Populations

- Now we wish to compare two or more populations by observing the values of a qualitative variable for each of them.
- For example, is there any difference of voters approval or disapproval of the president based on their political ideology?

	Approve	Disapprove	Neutral
Conservative	10	50	20
Liberal	20	10	10
Moderate	20	30	30

Homogeneous Populations

Definition (Homogeneous)

Two or more populations are called homogeneous if they exhibit the same distributions over the same categories.

- If there is no significant change in the *distribution* of opinion among the different political groups, then the groups are homogeneous.

The Test of Homogeneity

- The null hypothesis is that the populations are homogeneous.
- The alternative hypothesis is that the populations are not homogeneous.
  - H_0: The populations are homogeneous.
  - H_1: The populations are not homogeneous.

Outline

Independent Variables

- Statisticians at Christopher Newport University recently conducted a survey.

Independent Variables

- Statisticians at Christopher Newport University recently conducted a survey.
- They asked men and women whether they supported or opposed the new Virginia requirement that woman have an ultrasound procedure before having an abortion.

Independent Variables

- Statisticians at Christopher Newport University recently conducted a survey.
- They asked men and women whether they supported or opposed the new Virginia requirement that woman have an ultrasound procedure before having an abortion.
- There are two variables.

Independent Variables

- Statisticians at Christopher Newport University recently conducted a survey.
- They asked men and women whether they supported or opposed the new Virginia requirement that woman have an ultrasound procedure before having an abortion.
- There are two variables.
  - Whether the person is a man or a woman.

Independent Variables

- Statisticians at Christopher Newport University recently conducted a survey.
- They asked men and women whether they supported or opposed the new Virginia requirement that woman have an ultrasound procedure before having an abortion.
- There are two variables.
  - Whether the person is a man or a woman.
  - Whether the person supports or opposes the requirement or has no opinion.

Independent Variables

	Support	Oppose	No Opinion
Male	34%	53%	13%
Female	36%	55%	9%

Independent Variables

	Support	Oppose	No Opinion
Male	34%	53%	13%
Female	36%	55%	9%

- We need to express the data as actual counts.

Independent Variables

- The sample size was 1018, but I will change it to 1000 for simplicity.
- To make this example more interesting, we will assume that there were 600 men and 400 women.

	Support	Oppose	No Opinion
Male	204	318	78
Female	144	220	36

Independence

Definition (Independent Variables)

Two variables are independent if the value of one has no bearing (no predictive value) on the value of the other.

- If knowing the sex of the person is of no help in predicting their opinion, then the variables are independent.
- Equivalently, if knowing the opinion of the person is of no help in predicting their sex, then the variables are independent.

The Test of Independence

- The null hypothesis is that the variables are independent.
- The alternative hypothesis is that the variables are not independent.
  - H_0: The variables are independent.
  - H_1: The variables are not independent.

Outline

The Test of Homogeneity or Independence

- The two preceding examples could both be interpreted as problems concerning homogeneous populations or problems concerning independent variables.
- The procedure is the same.
- It is a matter only of how the problem is stated.

An Example

- We will work an example involving the political groups and their opinions.
- Are their opinions distributed in the same way?
- In other words, are the three populations homogeneous.

	Approve	Disapprove	Neutral
Conservative	10	50	20
Liberal	20	10	10
Moderate	20	30	30

An Example

Example (Test of Homogeneity)

(1) H_0 : The populations are homogeneous.
H_1 : The populations are not homogeneous.

(2) $\alpha = 0.05$.

(3) The test statistic is

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

Outline

Expected Counts

- The question now is, how do we find the expected counts?
- Under the assumption of homogeneity (H_0), the different rows should exhibit the same proportions.
- We can get the best estimate of those proportions by pooling the rows.
- That is, find the column totals then compute the proportions from them.

Row and Column Proportions

	Approve	Disapprove	Neutral
Conservative	10	50	20
Liberal	20	10	10
Moderate	20	30	30
Column Total			
Percent			

Row and Column Proportions

	Approve	Disapprove	Neutral
Conservative	10	50	20
Liberal	20	10	10
Moderate	20	30	30
Column Total	50	90	60
Percent			

Row and Column Proportions

	Approve	Disapprove	Neutral
Conservative	10	50	20
Liberal	20	10	10
Moderate	20	30	30
Column Total	50	90	60
Percent	25%	45%	30%

Expected Counts

- Similarly, the columns should exhibit the similar proportions, so we can get the best estimate of them by pooling the columns.
- That is, find the row totals and then compute the row proportions from them.

Row and Column Proportions

	Approve	Disapprove	Neutral	Row Total	Percent
Conservative	10	50	20		
Liberal	20	10	10		
Moderate	20	30	30		
Column Total	50	90	60		
Percent	25%	45%	30%		

Row and Column Proportions

	Approve	Disapprove	Neutral	Row Total	Percent
Conservative	10	50	20	80	
Liberal	20	10	10	40	
Moderate	20	30	30	80	
Column Total	50	90	60		
Percent	25%	45%	30%		

Row and Column Proportions

	Approve	Disapprove	Neutral	Row Total	Percent
Conservative	10	50	20	80	40%
Liberal	20	10	10	40	20%
Moderate	20	30	30	80	40%
Column Total	50	90	60		
Percent	25%	45%	30%		

Row and Column Totals

	Approve	Disapprove	Neutral	Row Total	Percent
Conservative	10	50	20	80	40%
Liberal	20	10	10	40	20%
Moderate	20	30	30	80	40%
Column Total	50	90	60	200	100%
Percent	25%	45%	30%	100%	

Expected Counts

- Now apply the appropriate row and column proportions to each cell to get the expected count.
- For example, in the upper-left cell, according to the row and column proportions, it should contain

$$40\% \text{ of } 25\% \text{ of } 200.$$

- That is, the expected count is

$$0.40 \times 0.25 \times 200 = 20.$$

Expected Counts

- This is the same as

$$\left(\frac{80}{200}\right) \times \left(\frac{50}{200}\right) \times 200 = \frac{80 \times 50}{200} = 20.$$

- Therefore, the quick formula is

$$\text{Expected Count} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}.$$

- Apply that formula to each cell to find the expected counts and add them to the table.

Expected Counts

Example (Test of Homogeneity)

(4)

	Approve	Disapprove	Neutral	Row Total
Conservative	10 (20)	50 (36)	20 (24)	80
Liberal	20 (10)	10 (18)	10 (12)	40
Moderate	20 (20)	30 (36)	30 (24)	80
Column Total	50	90	60	200

The Test Statistic

Example (Test of Homogeneity)

Now compute χ^2 in the usual way:

$$\begin{aligned}\chi^2 &= \frac{(10-20)^2}{20} + \frac{(50-36)^2}{36} + \frac{(20-24)^2}{24} \\ &\quad + \frac{(20-10)^2}{10} + \frac{(10-18)^2}{18} + \frac{(10-12)^2}{12} \\ &\quad + \frac{(20-20)^2}{20} + \frac{(30-36)^2}{36} + \frac{(30-24)^2}{24} \\ &= 5 + 5.44 + 0.67 + 10 + 3.56 + 0.33 + 0 + 1 + 1.5 \\ &= 27.5.\end{aligned}$$

Degrees of Freedom

Example (Test of Homogeneity)

(5) The number of degrees of freedom is

$$df = (\text{No. of rows} - 1) \times (\text{No. of cols} - 1).$$

In this example, $df = (3-1) \times (3-1) = 4$.
So the p-value is

$$p\text{-value} = \chi^2\texttt{cdf(27.5,E99,4)} = 1.575 \times 10^{-5}.$$

(6) Reject H_0.

(7) The opinions of the different groups are significantly different.

Outline

Degrees of Freedom

Example (Test of Homogeneity)

- Perform the test for the data concerning the ultrasound requirement.

	Support	Oppose	No Opinion
Male	204	318	78
Female	144	220	36

Outline

Assignment

Homework

- Read Section 14.4, pages 940 - 947.
- Let's Do It! 14.4, 14.5.
- Exercises 17, 18, p. 948, work "by hand."