MAT 1372 Stat w/ Prob classwk 27 Fall 2012

**13.3 TESTING FOR INDEPENDENCE IN POPULATIONS CLASSIFIED ACCORDING TO TWO CHARACTERISTICS**

Consider a large population in which each member is classified according to two

distinct characteristics, which we shall designate as the *X* characteristic and the

*Y* characteristic. Suppose that the possible values for the *X* characteristic are

denoted as 1 or 2 or . . . or *r*; similarly, the possible values of the *Y* characteristic

are denoted as 1 or 2 or . . . or *s*. Thus, there are *r* possible values for the *X*

characteristic and *s* possible values for the *Y* characteristic.

Let *Pij* denote the proportion of the population that has both *X* characterization

*i* and *Y* characterization *j*, for *i* being any of the values 1, 2, . . . , *r* and *j* being any

of the values 1, 2, . . . , *s*. Also, let *Pi* denote the proportion of the population who

have *X* characteristic *i*, and let *Qj* be the proportion who have *Y* characteristic *j*.

Thus if *X* and *Y* denote the values of the *X* characteristic and *Y* characteristic of a

randomly chosen member of the population, then

*P*{*X* = *i*, *Y* = *j*} = *Pij*

*P*{*X* = *i*} = *Pi*

*P*{*Y* = *j*} = *Qj*

We will be interested in developing a test of the hypothesis that the *X* characteristic

and *Y* characteristic of a randomly chosen member of the population are

independent. Recalling that *X* and *Y* are independent if

*P*{*X* = *i*, *Y* = *j*} = *P*{*X* = *i*}*P*{*Y* = *j*}

it follows that we want to test the null hypothesis

H0: *Pij* = *PiQj* for all *i* = 1, . . . , *r*, *j* = 1, . . . , *s*

against the alternative

H1: *Pij* = *PiQj* for some values of *i* and *j*

To test this hypothesis of independence, we start by choosing a random sample

of size *n* of members of the population. Let *Nij* denote the number of elements of

the sample that have both *X* characteristic *i* and *Y* characteristic *j*.

The results of a survey of gender and political party sympathy of 300 upstate New York adults given as a *contingency table* is

|  |  |  |  |
| --- | --- | --- | --- |
|  | Democrat | Republican | Independent |
| Women | 68 | 56 | 32 |
| Men | 52 | 72 | 20 |

Do we have enough evidence to show that gender does affect political sympathy?

The given table will serve as our observed values.

1. Our H0 is that gender does not affect sympathy. We need to find the expected values based on this assumption:
   1. Find the marginal totals for the observed. You can do so by selecting the data as well as the target cells and clicking on autosum. The result is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Democrat | Republican | Independent | Total |
| Women | 68 | 56 | 32 | 156 |
| Men | 52 | 72 | 20 | 144 |
| Total | 120 | 128 | 52 | 300 |

* 1. Copy the table by making a reference to a cell and filling down and right. Clear references to the observed data leaving just the marginal totals.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  | 156 |
|  |  |  |  | 144 |
|  | 120 | 128 | 52 | 300 |

* 1. Fill the inside of this 2nd table, by taking product of each marginal row and column values and dividing by the total. Write the formula for one cell with carefully placed $’s and fill down and right.

1. In a 3rd table, take the square of the difference between each observed and expected values and divide by the expected.
2. Sum these to get the test statistic TS.
3. To find the pvalue, use chidist with inputs TS and d.o.f. (r-1)(s-1), r,s the # of rows, columns respectively.