MAT 1372 Stat w/ Prob classwk 27 Spring 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*.

Suppose that you do a survey of political sympathy of 300 upstate New York adults and find

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

The first you want to do is find the marginal totals.

 With Excel use autosum and fill down and right.

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

This table is called a *contingency table.*