10-3
Contingency Tables

Definition

- **Contingency Table** (or two-way frequency table)

  a table in which frequencies correspond to two variables.

(One variable is used to categorize rows, and a second variable is used to categorize columns.)

Contingency tables have at least two rows and at least two columns.

Contingency Table

	Homicide	Robbery	Assault
Stranger	12	379	727
Acquaintance or Relative	39	106	642

Definition

- **Test of Independence**

  tests the null hypothesis that there is no association between the row variable and the column variable.

(The null hypothesis is the statement that the row and column variables are <u>independent</u>.)

Assumptions

1. **The sample data are randomly selected.**
2. **The null hypothesis H_0 is the statement that the row and column variables are independent; the alternative hypothesis H_1 is the statement that the row and variables are dependent.**
3. **For every cell in the contingency table, the expected frequency E is at least 5. (There is <u>no</u> requirement that every observed frequency must be at least 5.)**

Tests of Independence

H_0: The row variable is independent of the column variable

H_1: The row variable is dependent (related to) the column variable

This procedure cannot be used to establish a direct cause-and-effect link between variables in question.

Dependence means only there is a relationship between the two variables.

Test of Independence
Test Statistic

$$X^2 = \Sigma \frac{(O - E)^2}{E}$$

Critical Values

1. Found in Table A-4 using

   degrees of freedom = (r - 1)(c - 1)

r is the number of rows and c is the number of columns

2. **Tests of Independence are always right-tailed.**

$$E = \frac{\text{(row total) (column total)}}{\text{(grand total)}}$$

↑

Total number of all observed frequencies in the table

Contingency Table

	Homicide	Robbery	Assault
Stranger	12	379	727
Acquaintance or Relative	39	106	642

Expected Frequency for Contingency Tables

$$E = \text{grand total} \cdot \frac{\text{row total}}{\text{grand total}} \cdot \frac{\text{column total}}{\text{grand total}}$$

Expected Frequency for Contingency Tables

$$E = \text{grand total} \cdot \frac{\text{row total}}{\text{grand total}} \cdot \frac{\text{column total}}{\text{grand total}}$$

$$n \cdot p$$

(probability of a cell)

Expected Frequency for Contingency Tables

$$E = \cancel{\text{grand total}} \cdot \frac{\text{row total}}{\cancel{\text{grand total}}} \cdot \frac{\text{column total}}{\text{grand total}}$$

$$n \cdot p$$

(probability of a cell)

$$E = \frac{\text{(row total) (column total)}}{\text{(grand total)}}$$

Is the type of crime independent of whether the criminal is a stranger?

	Homicide	Robbery	Assault	
Stranger	12	379	727	
Acquaintance or Relative	39	106	642	

H_0: Type of crime is independent of knowing the criminal

H_1: Type of crime is dependent with knowing the criminal

Is the type of crime independent of whether the criminal is a stranger?

	Homicide	Robbery	Assault	Row Total
Stranger	12	379	727	1118
Acquaintance or Relative	39	106	642	787
Column Total	51	485	1369	1905

H_0: Type of crime is independent of knowing the criminal

H_1: Type of crime is dependent with knowing the criminal

Is the type of crime independent of whether the criminal is a stranger?

	Homicide	Robbery	Assault	Row Total
Stranger	12 (29.93)	379 (284.64)	727 (803.43)	1118
Acquaintance or Relative	39 (21.07)	106 (200.36)	642 (565.57)	787
Column Total	51	485	1369	1905

$$E = \frac{\text{(row total) (column total)}}{\text{(grand total)}}$$

$$E = \frac{(1118)(51)}{1905} = 29.93 \qquad E = \frac{(1118)(485)}{1905} = 284.64$$

etc.

Is the type of crime independent of whether the criminal is a stranger?

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

	Homicide	Robbery	Assault
Stranger	12 (29.93) [10.741]	379 (284.64) [31.281]	727 (803.43) [7.271]
Acquaintance or Relative	39 (21.07) [15.258]	106 (200.36) [44.439]	642 (565.57) [10.329]

(E)

$\left[\frac{(O - E)^2}{E}\right]$

Upper left cell: $\frac{(O - E)^2}{E} = \frac{(12 - 29.93)^2}{29.93} = 10.741$

Is the type of crime independent of whether the criminal is a stranger?

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

	Homicide	Robbery	Assault
Stranger	12 (29.93) [10.741]	379 (284.64) [31.281]	727 (803.43) [7.271]
Acquaintance or Relative	39 (21.07) [15.258]	106 (200.36) [44.439]	642 (565.57) [10.329]

(E)

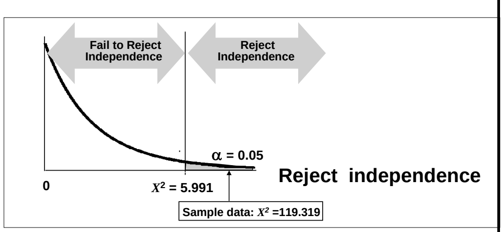
$\left[\frac{(O - E)^2}{E}\right]$

Test Statistic $X^2 = 10.741 + 31.281 + ... + 10.329 = 119.319$

Test Statistic: $X^2 = 119.319$

with $\alpha = 0.05$ and $(r - 1)(c - 1) = (2 - 1)(3 - 1) = 2$ degrees of freedom

Critical Value: $X^2 = 5.991$ (from Table A-4)



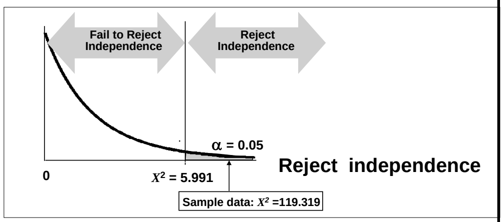
H_o : The type of crime and knowing the criminal are independent

H_1 : The type of crime and knowing the criminal are dependent

Test Statistic: $X^2 = 119.319$

with $\alpha = 0.05$ and $(r - 1)(c - 1) = (2 - 1)(3 - 1) = 2$ degrees of freedom

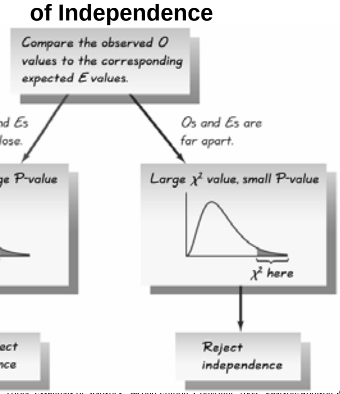
Critical Value: $X^2 = 5.991$ (from Table A-4)



It appears that the type of crime and knowing the criminal are related.

Relationships Among Components in X^2 Test of Independence

Figure 10-8



Definition

Test of Homogeneity

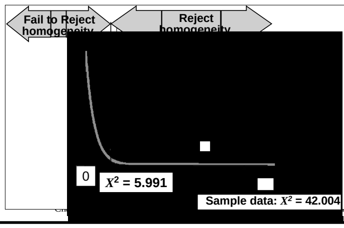
tests the claim that *different populations* have the same proportions of some characteristics

Example - Test of Homogeneity

Seat Belt Use in Taxi Cabs

		New York	Chicago	Pittsburgh
Taxi has usable seat belt?	Yes	3	42	2
	No	74	87	70

Claim: The 3 cities have the same proportion of taxis with usable seat belts
H_0: The 3 cities have the same proportion of taxis with usable seat belts
H_1: The proportion of taxis with usable seat belts is not the same in all 3 cities



There is sufficient evidence to warrant rejection of the claim that the 3 cities have the same proportion of usable seat belts in taxis; appears from Table Chicago has a much higher proportion.