THE CHI-SQUARE TEST FOR INDEPENDENCE

Goodness of Fit Test — One-way tables (only one categorical variable)

Test For Independence — Two-way Tables (Contingency Tables) Data is classified according to two categorical variables

Example: Suppose we have two categorical variables:

Origin: American Latin Asian

Use of Product: Uses it Does not use it

and we want to find out if the two variables are independent or dependent

Step 1 Ho: Origin and Use of Product are independent

Ha: " " " are Dependent

(Knowing the level of one of the variables helps predict the level of the other)

Step2 Test statistic $\chi^2 = \sum \frac{(O-E)^2}{E}$

CORRECTED

	American	Latin	Asian	
Uses the Product	$O_{11} = 50$ $E_{11} = 42.75$	$O_{12} = 30$ $E_{12} = 24.75$	$O_{13} = 10$ $E_{13} = 22.50$	$r_1 = 90$
Does not use the product	$O_{21} = 45$ $E_{21} = 52.25$	$O_{22} = 25$ $E_{22} = 30.25$	$O_{23} = 40$ $E_{23} = 27.50$	$r_2 = 110$
	$c_1 = 95$	$c_2 = 55$	$c_3 = 50$	$n = 200$

$$E_{11} = \frac{r_1 c_1}{n} = \frac{90 * 95}{200} = 42.75 \qquad E_{21} = \frac{r_2 \cdot c_1}{n} = \frac{110 * 95}{200} = 52.25$$

$$E_{12} = \frac{r_1 \cdot c_2}{n} = \frac{90 * 55}{200} = 24.75 \qquad E_{22} = \frac{r_2 \cdot c_2}{n} = \frac{110 * 55}{200} = 30.25$$

$$E_{13} = \frac{r_1 \cdot c_3}{n} = \frac{90 * 50}{200} = 22.50 \qquad E_{23} = \frac{r_2 \cdot c_3}{n} = \frac{110 * 50}{200} = 27.5$$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(50-42.75)^2}{42.75} + \frac{(30-24.75)^2}{24.75} + \frac{(10-22.50)^2}{22.50} +$$

$$+ \frac{(45-52.25)^2}{52.25} + \frac{(25-30.25)^2}{30.25} + \frac{(40-27.50)^2}{27.50} =$$

$$\approx 16.89$$

Step 3 Rejection Region

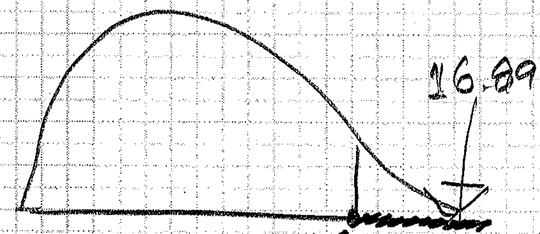
$\alpha = .05$

$df = (R-1)(C-1)$

$R = 2 \quad C = 3$

$df = (2-1)(3-1) = 1 \cdot 2 = 2$

RR: $\chi^2 > 5.99147$



C.V

$\chi^2_{.05,\,2} = 5.99147$

Step 4 Decision: Reject H_0

Step 5 conclusion: "The data provide sufficient evidence, at $\alpha = .05$, to conclude that Origin and Use of Product are dependent"