Question

An article by R.J. Payne and J.J. Pilgram (1981): "Changing Evaluations of Flood Plain Hazard, The Hunter Valley, Australia", *Environment and Behaviour*, 13, no.4, pp. 461-480, is a report on the results of a survey of the attitudes of people at risk of flood hazards. One aspect of this survey involved asking people about the **COST** (low, moderate or high) of any preparations they made in response to the threat of a flood and how the expectation of the amount of flood **DAMAGE** (major or minor) affected that response. The results of the 110 interviews are summarised in the following contingency table:

	DAMAGE		
COST	Major	Minor	Total
Low	43	16	59
Moderate	10	28	38
High	4	9	13
Total	57	53	110

(a) Find the expected cell frequencies for each of the six **COST-DAMAGE** categories in the above contingency table.

		DAMAGE		
		Major	Minor	
COST	Low	$\frac{59\times57}{110}=30.57$	$\frac{59\times53}{110}=28.43$	59
	Moderate	$\frac{38\times57}{110}=19.69$	$\frac{38\times53}{110}=18.31$	38
	High	$\frac{13\times57}{110}=6.74$	$\frac{13\times53}{110}=6.26$	13
		57	53	110

(b) Find the Pearson X^2 statistic for a test of association between **COST** and **DAMAGE**. Is there sufficient evidence at the $\alpha = 0.05$ level of significance to indicate that the level of preparation (ie. the **COST**) depends upon the extent of the perceived threat of flood damage (ie. the **DAMAGE**)?

$$X^2 = \sum_i \sum_j \frac{(O_{ij}-E_{ij})^2}{E_{ij}} = \frac{(43-30.57)^2}{30.57} + \frac{(16-28.43)^2}{28.43} + \dots + \frac{(9-6.26)^2}{6.26}$$

$$= 5.05 + 5.43 + 4.77 + 5.13 + 1.11 + 1.20 = 22.69$$

This is $\sim \chi^2_{(r-1)(c-1)} = \chi^2_{2\cdot1} = \chi^2_2$ and $\chi^2_2(1-\alpha = 0.95) = 5.99$

So, as $22.69 > 5.99$ reject H_0: no association, and conclude that there is a significant association between COST and DAMAGE.

(c) Find the likelihood ratio G^2 statistic for the same test as in part (b). Does this statistic lead you to the same conclusion as you reached in part (b)?

$$G^2 = 2\sum_i\sum_j O_{ij} \ln\left(\frac{O_{ij}}{E_{ij}}\right)$$

$$= 2\left[43 \ln\frac{43}{30.573} + 16 \ln\frac{16}{28.43} + \dots + 9 \ln\frac{9}{6.26}\right]$$

$$= 2\left[14.67 - 9.20 - 6.78 + 11.89 - 2.08 + 3.26\right]$$

$$= 23.53$$

This is also χ^2_2, so we reach the same conclusion as in part (b).

"Classical" Analysis of Contingency Tables
(aka Analysis of Cross-classified categorical data)

Think of the entries in the cross-classified table as a series of multinomial counts:

use these counts to estimate ⟹

		Factor 1: 1	Factor 1: 2	
Factor 2	1	Y_{11}	Y_{12}	$Y_{1.}$
	2	Y_{21}	Y_{22}	$Y_{2.}$
	3	Y_{31}	Y_{32}	$Y_{3.}$
		$Y_{.1}$	$Y_{.2}$	$Y_{..}$

π_{11}	π_{12}	$\pi_{1.}$
π_{21}	π_{22}	$\pi_{2.}$
π_{31}	π_{32}	$\pi_{3.}$
$\pi_{.1}$	$\pi_{.2}$	1

eg π_{11} could be estimated by $Y_{11}/Y_{..}$

$\pi_{.1}$ " " " " $Y_{.1}/Y_{..}$

$\pi_{1.}$ " " " " $Y_{1.}/Y_{..}$

Now, if Factor 1 & Factor 2 are unrelated (independent)

$\pi_{1.}$	$\pi_{1.}$	$\pi_{1.}$
$\pi_{2.}$	$\pi_{2.}$	$\pi_{2.}$
$\pi_{3.}$	$\pi_{3.}$	$\pi_{3.}$
		1

$\pi_{.1}$	$\pi_{.2}$	
$\pi_{.1}$	$\pi_{.2}$	
$\pi_{.1}$	$\pi_{.2}$	
$\pi_{.1}$	$\pi_{.2}$	1

Under this assumption $\pi_{ij} = \pi_{i.} \times \pi_{.j}$

eg $\pi_{11} = \pi_{1.} \times \pi_{.1}$

$$\Longrightarrow E[Y_{11}] = \frac{Y_{1.} \times Y_{.1}}{Y_{..}}$$

So, observed counts for each cell, eg $O_{11} = Y_{11}$ for cell 1,1

& the expected count in each cell is

$$E_{11} = \frac{Y_{1.} \times Y_{.1}}{Y_{..}} = \frac{\text{corresponding column total} \times \text{corresponding row total}}{\text{overall total}}$$

for cell 1,1 under the assumption of independence

$\longrightarrow \equiv$ Factor 1 & factor 2 are not associated (unrelated)

$\Longrightarrow$ so, the test of whether Factor 1 & factor 2 are related is a question of whether (or not)

$$O_{ij} = E_{ij} \quad \forall i, j$$

if they are, evidence of H_0: no association

if not , " " H_A: Factor 1 & Factor 2 associated

There are two "classical" tests

$$\text{Pearson } X^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\text{Likelihood Ratio } G^2 = 2 \sum_i \sum_j O_{ij} \ln\left(\frac{O_{ij}}{E_{ij}}\right)$$

both of which have an asymptotic χ^2 dist? with $(r-1)(c-1)$ df

r = # rows c = # columns

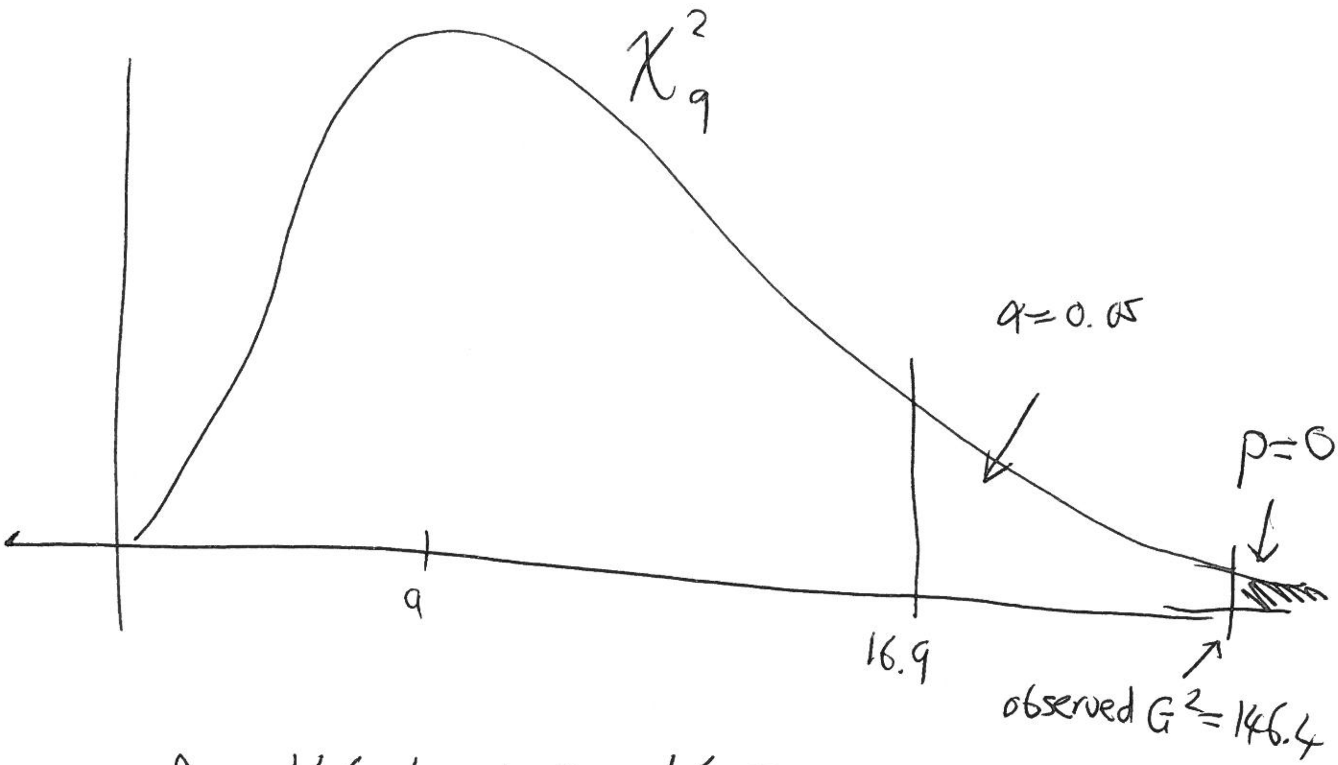
Eye Colour Example

Likelihood ratio G^2 test

H_0 : no association between hair colour & eye colour

H_A : Some " " " " " " "

$$G^2 \sim \chi^2_{(4-1)(4-1)=9}$$



As $146.4 >> 16.9$

or $p = 0 << \alpha = 0.05$

reject H_0 & conclude that there is an association between hair & eye colour