Contingency Tables

Starter

1. **(Review of last lesson)**
The accountant of a company monitors the number of items produced each month together with the total cost of production. The data collected for a random sample of 12 months is:

Number of items (x) (1000s)	21	39	48	24	72	75	15	35	62	81	12	56
Production cost (y) (£1000)	40	58	67	45	89	96	37	53	83	102	35	75

(a) Use your calculator to find an equation for the regression line of y on x.
(b) The selling price of each item is £2.20. Find the level of output at which income and total costs are equal. Interpret this value.

Working: (a) $y = 20.985 + 0.978x$ or $Y = 20985 + 978X$

(b) Need to solve the equation: $20.985 + 0.978x = 2.2x$

$$x = 17172$$

The level of output at which income and total costs are equal is 17172 — this is called the break even point

2. The following table shows the grades achieved by random students in two schools.

Observed frequencies		Grade A	Grade B	Grade C	Totals
School	X	18	12	20	50
	Y	26	12	32	70
Totals		**44**	**24**	**52**	**120**

(a) From the table, write down:
(i) the probability of choosing a student from school X,
(ii) the probability of choosing a student who got a grade A.
(b) If the two probabilities from (a) are independent (i.e. the factors school and grade are independent), calculate the probability of choosing a students from school X who achieved an A grade.
(c) Using your answer to (b), calculate how many students we would expect to get a grade A in school X if ‘school' and ‘grade’ were independent.
(d) Draw a new version of the table and calculate the expected values for the values in blue

Working: (a) (i) $\dfrac{50}{120}$

(ii) $\dfrac{44}{120}$

(b) $\dfrac{50}{120} \times \dfrac{44}{120} = \dfrac{11}{72}$

(c) $\frac{11}{72} \times 120 = \frac{55}{3} = 18\frac{1}{3}$

(d)

Expected frequencies		Grade A	Grade B	Grade C	Totals
School	X	$18\frac{1}{3}$	10	$21\frac{2}{3}$	50
School	Y	$25\frac{2}{3}$	14	$30\frac{1}{3}$	70
Totals		44	24	52	120

E.g. 1 Carry out a chi-squared test at the $5\ \%$ level to decide whether there is a connection between school and grades achieved. State your null and hypothesis clearly.

Working: H_0 : The variables school and grade are independent.
H_1 : The variables school and grade are not independent

$$\chi^2_{calc} = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{\left(18 - 18\frac{1}{3}\right)^2}{18\frac{1}{3}} + \frac{\left(12 - 10\right)^2}{10} + \frac{\left(20 - 21\frac{2}{3}\right)^2}{21\frac{2}{3}} + \frac{\left(26 - 25\frac{2}{3}\right)^2}{25\frac{2}{3}} + \frac{\left(12 - 14\right)^2}{14} + \frac{\left(32 - 30\frac{1}{3}\right)^2}{30\frac{1}{3}}$$

$\chi^2_{calc} = 0.9165$
Degrees of freedom, $\nu = 2$
The critical value at the $5\ \%$ level is $\chi^2_2(5\%) = 5.991$
Since $\chi^2_{calc} = 0.9165 < 5.991 = \chi^2_2(5\%)$, we do not reject H_0 :
There is evidence to suggest the variables are independent
i.e. the grade achieved by students is independent of school attended

E.g. 2 Seven different types of locality were studied to see if ownership, or non-ownership, of a television was or was not related to locality. $\chi^2_{calc} = \sum \frac{(O_i - E_i)^2}{E_i}$ was found to be 13.1.

(a) Carry out a hypothesis test at the $5\ \%$ level to see if locality and ownership of TV are related.

(b) Would your conclusion differ if the significance level was $2.5\ \%$? Explain your answer.

Working: (a) H_0 : The variables locality and TV ownership are independent.
H_1 : The variables locality and TV ownership are not independent
$\chi^2_{calc} = 13.1$
Degrees of freedom, $\nu = 6 \times 1 = 6$
The critical value at the 5% level is $\chi^2_6(5\%) = 12.59$
Since $\chi^2_{calc} = 13.1 > 12.59 = \chi^2_6(5\%)$, we reject H_0.
There is evidence to suggest locality and TV ownership are not independent

(b) The critical value at the 2.5% level is $\chi_4^2(2.5\%) = 14.45$
Since $\chi_{calc}^2 = 13.1 < 14.45 = \chi_6^2(2.5\%)$, we do not reject H_0.
So our conclusion would differ.

E.g. 3 A university requires all science students to study non-science subject in their first year. Here are the choices they made:

	French	Poetry	Russian	Sculpture
Male	2	8	15	10
Female	10	17	21	37

Test at the 1% level whether choice of subject is independent of sex.

Working: The expected frequency for males who choose French is $\dfrac{12 \times 35}{120} = 3.5$.
Since the expected frequency is below 5 we must combine columns.
It makes sense to combine French and Russian as they are both languages.
The new observed frequencies are:

New observed frequencies	French/Russian	Poetry	Sculpture
Male	17	8	10
Female	31	17	37

The expected frequencies are:

Expected frequencies	French/Russian	Poetry	Sculpture
Male	14.00	7.29	13.71
Female	34.00	17.71	33.29

$$\chi_{calc}^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(17-14.00)^2}{14.00} + \frac{(8-7.29)^2}{7.29} + \frac{(10-13.71)^2}{13.71} + \frac{(31-34.00)^2}{34.00} + \frac{(17-17.71)^2}{17.71} + \frac{(37-33.29)^2}{33.29}$$

$\chi_{calc}^2 = 2.422$
H_0 : sex and choice of subject are independent
H_1 : sex and choice of subject are not independent
Degrees of freedom, $\nu = 2 \times 1 = 2$
The critical value at the 1% level is $\chi_2^2(1\%) = 9.210$
Since $\chi_{calc}^2 = 2.422 < 9.210 = \chi_2^2(1\%)$, we do not reject H_0.
There is no evidence of a relationship between sex and choice of subject

Video: Contingency tables
Video: Chi-squared hypothesis tests
Video: Contingency tables example

Solutions to Starter and E.g.s

Exercise
p99 6A Qu 1i, 2-5, (6-7 red)