Exercises:

Q8: The following 2x3 contingency table contains observed freq. for a sample of 200. Test for independence of the Row and column variables using the χ^2 test with $\alpha = 0.05$.

			column variable				
	f_{ij}		f_{ij}		f_{ij}		
Row variable	A	e_{ij}	B	e_{ij}	C	e_{ij}	total
P	20	28.5	44	39.9	50	45.6	114
Q	30	21.5	26	30.1	30	34.4	86
	50		70		80		200

H_0: Row variable and column variable are independent.

H_1: Row variable and column variable are not independent.

$$e_{ij} = \frac{(\text{Row } i \text{ table})(\text{column } j \text{ table})}{\text{sample size}}$$

$$e_{11} = \frac{(114)(50)}{200}, \; e_{21} = \frac{(86)(50)}{200}, \; e_{12} = \frac{114 \times 70}{200}, \; e_{22} = \frac{86 \times 70}{200}, \; e_{13} = \frac{114 \times 80}{200}, \; e_{23} = \frac{86 \times 80}{200}$$

$$\rightarrow \chi^2 = \frac{(20-28.5)^2}{28.5} + \frac{(44-39.9)^2}{39.9} + \frac{(50-45.6)^2}{45.6} + \frac{(30-21.5)^2}{21.5} + \frac{(26-30.1)^2}{30.1} + \frac{(30-34.4)^2}{34.4}$$

$$\boxed{\chi^2 = 7.86 \quad \text{with} \quad df = (2-1)(3-1) = 2}$$

- p-value approach:

df \ α	0.025	0.01
2	7.378	9.210

↑ p-value

p-value $\in (0.01, 0.025) < \alpha$

- critical value approach: χ^2_α

df \ α	0.05
2	$5.991 = \chi^2_\alpha$

$$\chi^2 \geq \chi^2_\alpha$$

So we Reject H_0, so conclude the Row and column are not independent

Q9: Test for independence of the row and column variables using χ^2 test with $\alpha = 0.05$.

	Column variable						
Row variable	A	eij	B	eij	C	eij	total
P	20	17.5	30	30.63	20	21.88	70
Q	30	28.75	60	50.31	25	35.94	115
R	10	13.75	15	24.06	30	17.19	55
total	60		105		75		240

H_0: Row variable and column variable are independent.

H_1: Row variable and column variable are independent.

$$\rightarrow \chi^2 = \frac{(20-17.5)^2}{17.5} + \frac{(30-28.75)^2}{28.75} + \frac{(10-13.75)^2}{13.75} + \frac{(30-30.63)^2}{30.63} + \frac{(60-50.31)^2}{50.31} + \frac{(15-24.06)^2}{24.06} +$$

$$\frac{(20-21.88)^2}{21.88} + \frac{(25-35.94)^2}{35.94} + \frac{(30-17.19)^2}{17.19}$$

$$\Rightarrow \chi^2 = 19.76 , \quad \text{with } df = (3-1)(3-1) = 4$$

By critical approach:

df \ α	0.05
4	9.488

$\chi^2_\alpha = 9.488$

$\chi^2 \geq \chi^2_\alpha$

so we Reject H_0, so Row and column are not indep.

Q10: use $\alpha = 0.05$ and test for the indep. of type of flight and type of ticket. What is your conclusion?

Type of flight

Type of ticket	Domestic flights	e_{ij}	$\chi^2 = \frac{(f_{ij}-e_{ij})^2}{e_{ij}}$	International flights	e_{ij}	$\chi^2 = \frac{(f_{ij}-e_{ij})^2}{e_{ij}}$	total
First class	29	35.59	1.220233212	22	15.41	2.818176509	51
Business class	95	150.73	20.60527367	121	65.27	47.58438639	216
Economy class	518	455.68	8.523047753	135	197.32	19.68265964	653
total	642			278			920

H_0: Type of ticket and Type of flight are independent.

H_1: Type of ticket and Type of fight are not independent.

$$\chi^2 = \sum_{ij} \frac{(f_{ij}-e_{ij})^2}{e_{ij}} = 100.43 , \quad \text{with } df = (3-1)(2-1) = 2 .$$

By critical value approach :

$\chi^2_\alpha = ?$

df \ α	0.05
2	5.991

$$\chi^2 \geq \chi^2_\alpha$$

So we Reject H_0

So Type of ticket and Type of flight are not indep.

Q11: use $\alpha = 0.01$ and test for indep. of degree major and industry type.

Industry

Degree major	oil	e_{ij}	$\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$	Chemical	e_{ij}	$\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$	electrical	e_{ij}	$\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$	computer	e_{ij}	$\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$	total
Business	30	30	0	15	22.5	2.5	15	17.5	0.36	40	30	$3.\overline{33}$	100
Engineering	30	30	0	30	22.5	2.5	20	17.5	0.36	20	30	$3.\overline{3}$	100
total	60			45			35		ما بقرب بجنه كله	60			200

H_0: Degree major and Industry are independent.

H_1: Degree major and Industry are not independent.

$\chi^2 = 12.381$. with $df = (2-1)(4-1) = 3$.

By p-value approach:

df \ α	0.01	0.005
3	11.345	12.838

$\downarrow$ p-value

$\Rightarrow$ p-value $\in (0.005, 0.01)$.

$(0.005, 0.01) \leq 0.01$

p-value $\leq \alpha$

So we Reject H_0 $(\alpha = 0.01)$.

So Degree major and Industry are not independent.

Industry

Q12:

order	pharm	e_{ij}	$\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$	consumer	e_{ij}	$\frac{(f-e)^2}{e}$	computers	e_{ij}	$\frac{(f-e)^2}{e}$	Teleco.	e_{ij}	$\frac{(f-e)^2}{e}$	total
Correct	207	201.6	29.16/e	136	134.4	2.56/e	151	153.6	6.76/e	178	182.4	19.36/e	672
Incorrect	3	8.4	29.16/e	4	5.6	2.56/e	9	6.4	6.76/e	12	7.6	19.36/e	28
total	210			140			160			190			

$\alpha = 0.05$

H_0: order and Industry are independent.

H_1: order and Industry are not independent.

$\chi^2 = 7.85$ with $df = (2-1)(4-1) = 3$

a. Test whether order fulfillments is independent?

By p-value approach?

df \ α	0.05	0.025
3	7.815	9.348

p-value $\in (0.025, 0.05)$.

$(0.025, 0.05) \leq 0.05$

P-value $\leq \alpha$

So Reject Ho $(\alpha = 0.05)$.

So order and Industry are not independent $(\alpha = 0.05)$

b. which industry has the highest percenting of correctly filled orders?

[illegible] ??

?

Q14: use $\alpha = 0.05$ and test the hypothesis that part quality is independent of the production shift. What is your conclusion?

shift	# Good	e_{ij}	$\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$	#defective	e_{ij}	$\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$	total
First	368	367.47	0.00076	32	32.53	0.0086	400
second	258	250.79	0.21	15	22.20	2.34	273
third	176	183.73	0.33	24	16.27	3.67	200
	802			71			873

H_0: shift and quality are independent.

H_1: shift and quality are not independent.

$\chi^2 = 6.55$ with $df = (3-1)(2-1) = 2$

By p-value approach:

df \ α	0.05	0.025
2	5.991	7.378

$\Rightarrow$ p-value $\in (0.025, 0.05)$.

p-value $\leq \alpha$

So Reject Ho, $(\alpha = 0.05)$

So shift and quality are not independent.

Age group

Q15:

payment	18-24	e_{ij}	$\frac{(f-e)^2}{e}$	25-34	e_{ij}	$\frac{(f-e)^2}{e}$	35-44	e_{ij}	$\frac{(f-e)^2}{e}$	45 and over	e_{ij}	$\frac{(f-e)^2}{e}$	total
plastic	21	15.54	1.92	27	23.31	0.58	27	25.53	0.085	36	46.62	2.42	111
cash or cheque	21	26.46	1.13	36	39.69	0.34	42	43.47	0.49	90	79.38	1.42	189
total	42			63			69			126			300

a. Test for the independence between Age group and payment. p-value ?? , $\alpha = 0.05$.

H_0: payment and Age group are independent.

H_1: payment and Age group are not independent

$\chi^2 = 7.944$ with $df = (2-1)(4-1) = 3$

By p-value approach ?

df \ α	0.05	0.025
3	7.815	9.348

p-value $\in (0.025, 0.05)$.

p-value $\leq \alpha$

so we reject H_0 $(\alpha = 0.05)$

so payment and age group are not independent.

b. If method of payment and age group are not independent, ... ?

P/E Ratio $\alpha = 0.05$

Q6:

industry	5-9	e_{ij}	$\frac{(f-e)^2}{e}$	10-14	e_{ij}	$\frac{(f-e)^2}{e}$	15-19	e_{ij}	$\frac{(f-e)^2}{e}$	20-24	e_{ij}	$\frac{(f-e)^2}{e}$	25-29	e_{ij}	$\frac{(f-e)^2}{e}$	Total
Consumer	4	9	2.78	10	12	0.33	18	15	0.6	10	8	0.5	8	6	0.67	50
banking	14	9	2.78	14	12	0.33	12	15	0.6	6	8	0.5	4	6	0.67	50
total	18			24			30			16			12			100

H_0 : P/E Ratio and industry are independent

H_1 : P/E Ratio and industry are not independent.

$\chi^2 = 9.76$ with $df = (2-1)(5-1) = 4$

By p-value approach :

df \ α	0.05	0.025
4	9.488	11.143

p-value $\in (0.025, 0.05)$.

p-value $\leq \alpha$

So we reject H_0 $(\alpha = 0.05)$.

so P/E Ratio and industry are not independent.