p212 Ex11.3 no. 1

a) H_0 : Patients' Social class is distributed as the population

H_1 : Patients' social class is not distributed as the population.

We do a goodness-of-fit test.
Assume H_0 to be true. $\alpha = 5\%$ 1-tail test

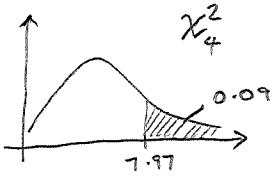
Class	I	II	III	IV	V	Total
Patient, f_o	28	63	188	173	48	500
Popn.	200	500	1600	1200	500	4000
Expected, f_e	25	62.5	200	150	62.5	

$[f_e := 500 \times \frac{\text{popn}}{4000}]$

$df = 5 - 1 = 4$

$X^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 7.97467...$

$P(X^2 > 7.97467..) = 0.092511...$



So, at the 5% level, we do not reject H_0

Based on this sample, there is evidence that the Patients are a representative sample of the whole area, with respect to social class

b) We do a 2-way contingency table test

H_0 : there is no association between reason for visit & class
H_1 : there is an association " " " " "

Observed

	I	II	III	IV	V
Minor	10	21	98	91	27
major	7	17	49	40	15
mental	11	25	41	42	6

Expected

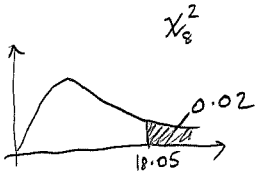
	I	II	III	IV	V
minor	13.8	31.1	92.9	85.5	23.7
major	7.2	16.1	48.1	44.3	12.3
mental	7	15.8	47	43.2	12

all $f_e \geq 5$ ✓☺

$df = (3-1) \times (5-1) = 8$

$X^2 = 18.0525$

$P(X^2 > 18.0525) = 0.020836$



So at the 5% level, we reject H_0

Based on this sample, there is evidence that there is an association between the reason for visits and their social class.

From scrutiny of the data, it would appear that 'Mental & Other' issues are not being treated at the expected rate. For social classes I & II they are above expected rates and classes III, IV and V they are under expected rates. Social stigma attached to Mental Health issues here may be a cause for this phenomenon.

Ex11·3 no. 2

a)

observed

	A	B	C	
Road	78	46	24	148
Transport	22	34	36	92
	100	80	60	

H_0: area and preferred improvement are independent

H_1: area and preferred improvement are associated in some way

We assume H_0 to be true. $\alpha = 1\%$. 1-tail test

expected

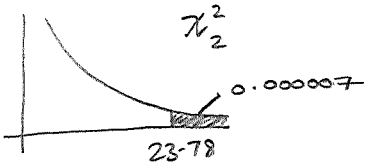
	A	B	C
Road	61·7	49·3	37
Transport	38·3	30·7	23

all $f_e \geq 5$ ✓ ☺

$df = 2$

$X^2 = 23·7885$

$P(X^2 > 23·785) = 0·000007$



So as $0·000007 < 0·01$ we have evidence that this sample does not support H_0 ⇒ reject H_0

Therefore there is an association between the preferred improvement and the areas of the city, based upon this sample.

b)

with private cars:

	A	B	C	
Road	61	30	10	101
Transport	9	10	5	24
	70	40	15	

$\chi^2 = 4·21$

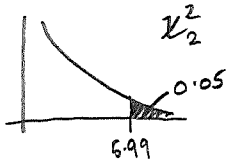
without private cars:

	A	B	C	
Road	17	16	14	47
Transport	13	24	31	68
	30	40	45	

← elements from original table minus table on left

↑ totals from original table, minus those totals from table on left.

$X^2 = 4·88$



For each value of X^2, they are both < 5.99 and outside of the critical region.

Hence, at the 5% level, we do not have evidence to reject H_0, which would have stated that area and preference are independent.

c) The results are different because the differing sample sizes are influencing the X^2 calculation's value. It would appear that combining two samples of 'small effect' gave rise to a larger sample with a very significant association.

Ex 11·3 no. 3

	Surgeon A	B	C
transfer	6	4	14
no transfer	41	68	27
	47	72	41

H_0: no association between surgeon and complications ⇒ transfer needed

H_1: there is an association.

Assume H_0 to be true $\alpha = 5\%$, 1 tail test

expected freq

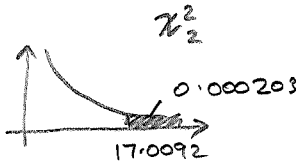
	A	B	C
transfer	7.05	10.8	6.15
no transfer	39.95	61.2	34.85

$f_e \geq 5$ ✓ ☺

$df = 2$

$X^2 = 17.0092$

$P(X^2 > 17.0092) = 0.000203$



As $0.000203 < 0.05$, we have evidence to reject H_0.

Hence, based upon this sample, there would appear to be an association between the surgeon and the need to transfer due to complications.

Scrutiny of the observed & expected frequencies suggest that surgeon C has many more transfers than expected.

Surgeon B has much fewer complications for his patients than expected (4 instead of 10.8) This may stem from Surgeon B only doing 'minor' operations, with low risk of complications attached to them. Simply offering him more (and he already does more than the other two surgeons) may have a reverse effect and diminish his quality, or have him perform operations which he is not trained for, but surgeons A or C are.

The administrator should establish which operations that A, B and C do are comparable else his analysis is not based upon comparing like-with-like.

Ex 11·3 no 4.

4 a)

observed

	CT	Bin	Nom	Poi
Good	25	12	12	11
Bad	4	11	3	12

expected

	CT	Bin	Nom	Poi
Good	19·3	15·3	10	15·3
Bad	9·7	7·7	5	7·7

fe≥5 ✓ ☺

H_0: there is no association between type of question and performance at question

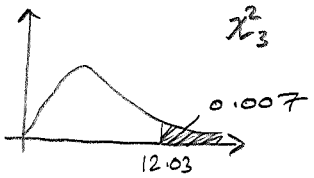
H_1: this is an association.

Assume H_0 to be true. $\alpha = 5\%$ 1-tail test

$X^2 = 12.0306$

$df = 3$

$P(X^2 > 12.0306) = 0.007279$



So as $0.007 < 0.05$, we have evidence to reject H_0

∴ based on this sample, there appears to be a connection between the type of question and how well it is answered.

It would appear, from scrutiny of the frequencies, that students are worse at Binomial and Poisson questions than we'd expect (11 vs 7·7 and 12 vs 7·7)

b) You could combine Bin & Poisson as they are the 'weak' topics, and both discrete distributions

Or, you could combine Bin & Normal as this increases the expected frequency of 5 to something larger.

I would not combine with Contingency Tables as GoF and CT questions are very different.

c)

observed

	CT	Bin	Nom	Poi
Attempted	29	23	15	23
Not attempt	1	7	15	7

expected

	CT	Bin	Nom	Poi
Attempt	22·5	22·5	22·5	22·5
Not attempt	7·5	7·5	7·5	7·5

H_0: no association between type of question and attempts made at it

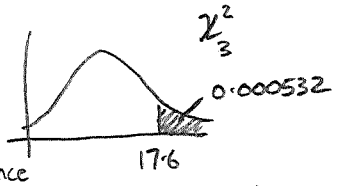
H_1: there is an association

Assume H_0 to be true $\alpha = 5\%$ 1-tail test

$X^2 = 17.6$

$df = 3$

$P(X^2 > 17.6) = 0.000532$



have evidence

Hence, as $0.0005 < 0.05$, we conclude that there is an association, based upon

Ex11·3 no. 4 cont.

this sample. From scrutiny of the frequencies, students are not choosing Normal test questions as often as they ought to, under H_0.

d) So it would appear that students like tackling Bin & Po. questions, but they are not good at them. Conversely, fewer choose Normal questions, but those few who do them are good at them.

Ex 11·3 no. 5

a) H_0: books borrowed uniformly throughout the week

H_1: books not borrowed uniformly through the week.

We assume H_0 to be true.

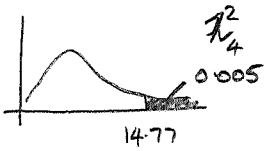
$\alpha = 1\%$ 1-tailed test

	M	T	W	Th	F	
f_o, observed	518	431	485	443	523	2400
f_e, expected	480	480	480	480	480	

$df = 5-1 = 4$

$X^2 = 14.7667$

$P(X^2 > 14.7667) = 0.00521$



as $0.00521 < 0.01$, we have evidence to reject H_0

Hence, based on this sample, it would appear that there is pattern to the distribution of books borrowed throughout the week

From scrutiny of the frequencies, much more happens on Monday & Friday than it should – maybe it is people borrowing a book for over the weekend, and when they return it on Monday they borrow another to tide them over for the next 4 days?!

b) H_0: there is no association between Grade and Length of Employment

H_1: there is an association.

Assume H_0 to be true $\alpha = 1\%$ 1 tail test

observed

	0-2	2-5	>5
Managerial	4	11	6
Skilled	32	28	21
Unskilled	25	23	50

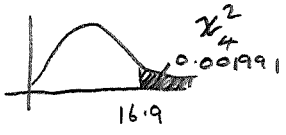
expected

	0-2	2-5	>5
M	6.4	6.5	8.1
S	24.7	25.1	31.2
US	29.9	30.4	37.7

$df = 4$

$X^2 = 16.9337$

$P(X^2 > 16.9337) = 0.001991$



As $0.001991 < 0.01$, we have evidence to reject H_0

We conclude that there is an association between Grade and Years of employment, based on this data

From scrutiny of the frequencies of these 200 leavers, we could conjecture that many more 2-5 year managers leave than we'd expect – maybe their promotion path is blocked by those above?

We also see many more long serving unskilled workers (50 vs 37.7). The company therefore ought to see if their training program could help these employees who are leaving in greater numbers than expected.