CIMT Further Statistics p110 Ex 5.4

1.

bottles	0	1	2	3	4	5
f_o	41	62	49	12	5	1

$\sum f_o = 170.$

from calc, $\bar{x} = 1.3$

so if $X \sim B(n,p)$ then $np = 1.3$

$p = \frac{1.3}{5}$

$p = 0.26$

f_e 37.7 66.3 46.6 16.4 2.9 0.2. ← from 170 x binompdf(5, 0.26)

now one $f_e < 1$ and only $\frac{2}{3}$ $f_e \geq 5$ so we need to combine categories.

bottles	0	1	2	3	4 - 5
f_o	41	62	49	12	6
f_e	37.7	66.3	46.6	16.4	3.1

now 80% of $f_e \geq 5$ which is acceptable.

H_0: data fits $B(5, 0.26)$

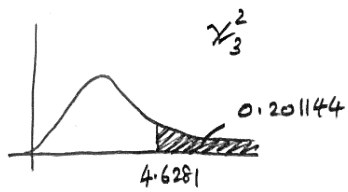
H_1: data does not fit $B(5, 0.26)$

Assume H_0 to be true.

$\alpha = 5\%$, one-tail test

$df = 5 - 2 = 3$

$X^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 4.6281$



$P(X^2 > 4.6281) = 0.201144$

> 0.05

so do not reject H_0

we do not have evidence to suggest that the data does not fit $B(5, 0.26)$

This suggests that the numbers of bottles in samples of size 5, that contain less than 1160 mL, follows a $B(5, 0.26)$ distribution.

2.

x	0	1	2	3	4	5	6+
f_o	728	447	138	48	26	13	0

$\Sigma f_o = 1400$

$\Sigma f_o x = 1036$

$\therefore \bar{x} = \frac{1036}{1400} = 0.74$

also $s_x^2 = 1.00455$, Hmm. as $\bar{x} \neq s_x$, maybe not Poisson ☹

f_e 667.9 494.3 182.9 45.1 8.34 1.23 0.17

H_0: data fits $Po(0.74)$

H_1: data does not fit $Po(0.74)$

Assume H_0 to be true

$\alpha = 1\%$, one-tail test

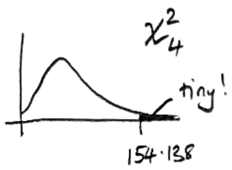
we have one $f_e < 1$ so we need to combine categories

x	0	1	2	3	4	5+
f_o	728	447	138	48	26	13
f_e	667.9	494.3	182.9	45.1	8.3	1.4

now we have $\frac{5}{6}$ of $f_e \geq 5$ which is acceptable.

so $df = 6 - 2 = 4$

$$X^2 = \Sigma \frac{(f_o - f_e)^2}{f_e} = 154.138$$



$P(X^2 > 154.138) = 2.6 \times 10^{-32}$
≈ 0

So we have evidence to reject H_0, and we conclude that the data is not $Po(0.74)$ distributed.

b) The engineer' claim that the breakdown rate is constant is effectively saying that it is Poisson distributed. In the light of part (a), where we rejected this notion, we disagree that the engineers' claim is true.

c)

	A	B	C	D
f_o	230	303	270	233
f_e	259	259	259	259

H_0: breakdowns are uniform, $U(4)$

H_1: breakdowns not uniform, $U(4)$

Assume H_0 to be true

$\alpha = 5\%$, one-tailed test

$df = 4 - 1 = 3$

$$X^2 = \Sigma \frac{(f_o - f_e)^2}{f_e} = 13.7992$$

$P(X^2 > 13.7992) = 0.003192 < 0.05$

Evidence to reject H_0, meaning that breakdowns do not occur at an equal rate on each production line.

4\. a) H_0: books borrowed uniformally, $U(5)$

H_1: books not borrowed uniformally, $U(5)$

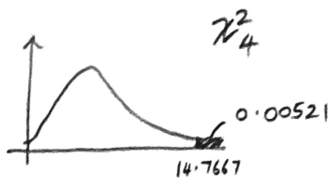
Assume H_0 to be true

$\alpha = 1\%$, one-tail test

	M	T	W	Th	F	
f_o	518	431	485	443	523	$\sum f_o = 2400$
f_e	480	480	480	480	480.	

$df = 5-1 = 4$

$$X^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 14.7667$$



$P(X^2 > 14.7667) = 0.00521 < 0.01$

Evidence to reject H_0 at 1% level and conclude that books are not borrowed evenly over the week

b) $P(X=r) = (r-1)\,p^2(1-p)^{r-2}$ $\quad r = 2, 3, \ldots$

H_0: data distributed as stated

H_1: data not distibuted as stated

Assume H_0 to be true

$\alpha = 5\%$, one tail test

r	2	3	4	5	6+
f_o	18	17	12	3	0
f_e	22.2	14.8	7.4	3.3	2.26.

$\sum f_o = 50$

we have $\bar{x} = 3$, so estimate, $p = \frac{2}{3}$.

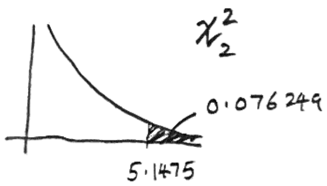
$\leftarrow$ from $50 \times P(X=r)$

now we have 60% of $f_e \geq 5$, so we must combine categories

r	2	3	4	5+
f_o	18	17	12	3.
f_e	22.2	14.8	7.4	5.55

so $df = 4 - 2 = 2$

$$X^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 5.1475$$



so $P(X^2 > 5.1475) = 0.076$
$\qquad > 0.05$

so we do not reject H_0

we do not have evidence to suggest that the data is not distributed as stated.

6. a) H_0: data fits $Po(\lambda)$

H_1: data not fit $Po(\lambda)$

Assume H_0 to be true

$\alpha = 5\%$, one-tailed test

from calc, mean strings $= 2.8$

also $s_x^2 = 2.21$

$\neq 2.8$

so Poisson doubtful!

anyway, let $\lambda = 2.8$

strings	0	1	2	3	4	5	6	7	
f_o	14	29	57	48	31	41	0	0	$\sum f_o = 220$
f_e	13.4	37.5	52.4	48.9	34.2	19.2	8.9	5.37	$\leftarrow 220 \times$ poisspdf

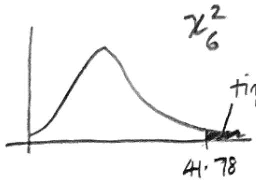
all $f_e \geq 5$, so no need to combine categories.

$df = 8 - 2 = 6$

$X^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 41.7868$

so $P(X^2 > 41.7868) = 2 \times 10^{-7}$

$\ll 0.05$



So, evidence to reject H_0, meaning Poisson is not an adequate distribution model for the number of blemishes in cloth.

b)

strings	0	1	2	3	4	≥ 5
f_o	14	29	57	48	31	41
f_e	10.96	32.85	49.29	49.29	36.98	40.63

all $f_e \geq 5$ ✓☺

H_0: data fits $Po(3)$

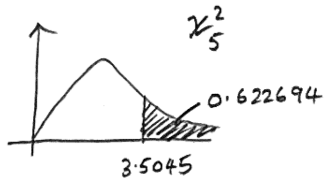
H_1: data does not fit $Po(3)$

assume H_0 to be true

$\alpha = 5\%$ one tail test

$df = 6 - 1 = 5$

$X^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 3.50459$



$P(X^2 > 3.5045) = 0.622694$

> 0.05 so do not reject H_0

We do not have evidence to suggest that the data is not $Po(3)$ distributed.

c) Given that a poisson model only fits if the data is truncated, and that if we don't truncate the data, the very large number of 5 string cloths (41, compared to expected number of 19.2) means that Poisson model is not a good fit.

This leads me to conclude that blemishes do not occur at random at a constant average rate through the cloth, as we have excessive cloths with 5 strings in them.

8 coin is biased

X = no. heads

x	0	1	2	3	4	5
f_o	5	39	70	52	25	9

$\Sigma f_o = 200.$

from calc, $\bar{x} = 2.4$

if $X \sim B(5, p)$, then $5p = 2.4$

$p = 0.48$

(so coin is biased towards tails)

f_e	7.6	35.1	64.8	59.8	27.6	5.10

all $f_e \geq 5$, so no need to combine categories.

H_0: data fits $B(5, 0.48)$

H_1: data not fit $B(5, 0.48)$

Assume H_0 to be true

here $df = 6 - 2 = 4$, if we were to do $X^2 = \Sigma \frac{(f_o - f_e)^2}{f_e}$

a) $X_1^2 = \Sigma \frac{f_o - f_e}{f_e} = 0.390348$

b) $X_2^2 = \Sigma \frac{|f_o - f_e|}{f_e} = 1.52502$

I consider that X_2^2 would be better, as it sums the 'positive deviations', in a similar way to $\frac{(f_o - f_e)^2}{f_e}$, so X_2^2 is always increasing though the calculation

The downside to X_1^2 is that 'negative deviations' will reduce the total sum, thereby giving a false impression that it is a good fit.

Also worth noting that as X_4^2 does not involve any squaring, it would not be compared against a χ^2 distribution, but rather a χ distribution, of sorts.

9.

H_0: data fits $Po(\lambda)$

H_1: data not fit $Po(\lambda)$

Assume H_0 to be true.

$\alpha = 5\%$, one-tail test

rooms required	0	1	2	3	4	5	6	7	8	
f_o	2	9	16	26	33	25	20	11	5	$\Sigma f_o = 147$
f_e	2.27	9.5	19.7	27.4	28.6	23.9	16.6	9.9	9.1	

all $f_e \geq 5$ ✓ ☺

now, from calc, $\bar{x} = 4.17007$

also $s_x^2 = 3.3065$

$\not\approx 4.17$, so poisson not looking good so far ☹

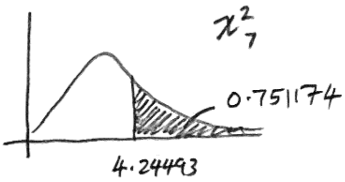
let $\lambda = 4.17$

generate f_e with $147 \times$ poisspdf(4.17, rooms)

all $f_e > 1$ and only $\frac{1}{9}$ of $f_e < 5$, so no need to combine categories.

$df = 9 - 2 = 7$

$X^2 = \Sigma \frac{(f_o - f_e)^2}{f_e} = 4.24493$



$P(X^2 > 4.24493) = 0.751174$

$\gg 0.05$ so do not reject H_0

We do not have evidence to suggest that room demand is not $Po(4.17)$ distributed.

let X = demand for rooms, so $X \sim Po(4.17)$

let Y = no. rooms occupied per night

y	0	1	2	3	4
$P(Y=y)$	0.015	0.064	0.134	0.187	0.599

so $P(Y=0) = P(X=0)$
$P(Y=1) = P(X=1)$
$P(Y=2) = P(X=2)$
$P(Y=3) = P(X=3)$
but $P(Y=4) = P(X \geq 4)$ as 4 rooms will be occupied when at least 4 people want a room.

so $E(Y) = \Sigma y P(Y=y) = 3.28944$

$E(Y^2) = \Sigma y P(Y=y) = 11.8668$

$Var(Y) = E(Y^2) - E^2(Y) = 1.04641$

Hence mean no. rooms occupied is 3.3 (1dp) with variance of 1.0 (1dp)

10. a) H_0: data fits $B(5,p)$

H_1: data not fit $B(5,p)$

we assume H_0 to be true

$\alpha = 5\%$ one-tail test

x	0	1	2	3	4	5	
f_o	0	3	12	27	26	12.	$\Sigma f_o = 80.$
f_e	0·27	2·8	12·1	25·7	27·4	11·6.	

we calculate $\bar{x} = 3{\cdot}4$

so $5p = 3{\cdot}4$

$p = 0{\cdot}68$

we have one $f_e < 1$ and $\frac{4}{6}$ of $f_e \geq 5$,

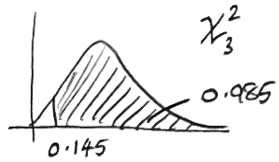
so we need to combine categories.

x	0-1	2	3	4	5
f_o	3	12	27	26	12
f_e	3·12	12·1	25·7	27·4	11·6.

now we have 80% of $f_e \geq 5$, which is acceptable.

$df = 5 - 2 = 3$

$X^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 0{\cdot}145809$



$P(X^2 > 0{\cdot}145809) = 0{\cdot}985823$

$\gg 0{\cdot}05$ so do not reject H_0

We do not have evidence to suggest that driver only cars are not $B(5, 0{\cdot}68)$ distributed.

b) for the given table, we have several $f_e < 1$, so we need to combine categories.

so

no. cars	0	1	2	3	4	≥ 5
f_o	28	40	32	19	7	4
f_e	25·85	41·75	33·72	18·16	7·33	3·19

$\bar{x} = \frac{210}{130} = 1{\cdot}61538$

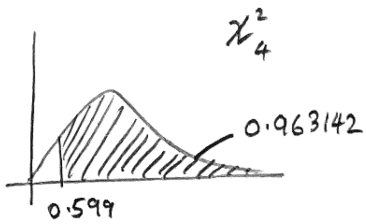
we H_0: data fits $Po(1{\cdot}615)$

H_1: data not fit $Po(1{\cdot}615)$

Assume H_0 to be true

$\alpha = 5\%$, one-tail test

$df = 6 - 2 = 4$



$X^2 = 0{\cdot}599293$

$P(X^2 > 0{\cdot}599293) = 0{\cdot}963142$

$\gg 0{\cdot}05$ so do not reject H_0

We do not have evidence to suggest that the $Po(1{\cdot}615)$ model is not adequate.

c) The p-values for both tests are exceptionally high, at 0·98 and 0·96 respectively. This consistently high match between theory and reality is very suspicious indeed.