CIMT Further Statistics p102 Ex 5B.

1. H_0 : data fits $B(5, 0.5)$ distribution

H_1 : data does not fit $B(5, 0.5)$ distribution

Assume H_0 to be true

$\alpha = 5\%$, one-tail test (as cows only have one tail ☺)

Heifers	0	1	2	3	4	5	
f_o	4	19	41	52	26	8	$\Sigma f_o = 150$
f_e	4.7	23.4	46.9	46.9	23.4	4.7	← from 150 × binompdf (5, 0.5)

all $f_e > 1$ ✓

$33\frac{1}{3}\%$ $f_e < 5$ ✗ so we need to combine categories, as we need $\geq 80\%$ of $f_e \geq 5$

Combining 0 & 1

Heifers	0-1	2	3	4	5
f_o	23	41	52	26	8
f_e	28.1	46.9	46.9	23.4	4.7

$df = 5 - 1 = 4$

$$X^2 = \Sigma \frac{(f_o - f_e)^2}{f_e} = 4.85156$$

$P(X^2 > 4.85156) = 0.302871$

OR

combining 4 & 5

Heifers	0	1	2	3	4-5
f_o	4	19	41	52	34
f_e	4.7	23.4	46.9	46.9	28.1

$df = 5 - 1 = 4$

$$X^2 = \Sigma \frac{(f_o - f_e)^2}{f_e} = 3.46489$$

$P(X^2 > 3.46489) = 0.483237$

either way, the p-value is > 5% and we do not have evidence to reject H_0 so we do not have evidence to suggest that the data does not fit the $B(5, 0.5)$ distribution.

If p had not been specified, we would have estimated it from the observed data (to be $p = 0.534667$)
This value of p would then generate the expected frequencies.
We would again check to see that none $f_e < 1$ and no more than 20% $f_e < 5$
The degrees of freedom would be 2 less than the number of categories
X^2 would be recalculated, along with p-value, and end comparison as normal.

2. H_0 : data is distributed with Po (4)

H_1 : data is not distributed with Po(4)

Assume H_0 to be true

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

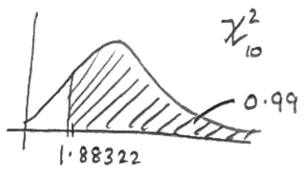
x	0	1	2	3	4	5	6	7	8	>8	
f_o	5	12	31	40	38	29	22	14	5	4	$\sum f_o = 200$.
f_e	3·7	14·6	29·3	39·1	39·1	31·3	20·8	11·9	5·9	4·27269	← from 200 × poisspdf(4, x)

this from $200 - \sum_1^8 f_e$

no $f_e < 1$ and 80% of $f_e \geq 5$, so no need to combine categories

df = 10 − 1 = 9

$$X^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 1{\cdot}88322$$



$P(X^2 > 1{\cdot}88322) = 0{\cdot}993183$

$> 0{\cdot}05$

So we do not reject H_0

So we do not have evidence to suggest that the data is not distributed with Po(4)

3. a)

x	0	1	2	3	4	5
f_o	8	19	25	22	5	1

$\Sigma f_o = 80$

from this data, $\bar{x} = 2$

we presume that $X =$ no. correct forecasts

$X \sim B(5, p)$

so $np = 2$

$5p = 2$

$p = 0.4$

$\therefore$ we try to fit a $B(5, 0.4)$ distribution to the data.

b)

x	0	1	2	3	4	5
f_o	8	19	25	22	5	1
f_e	6.2	20.7	27.6	18.4	6.1	0.8

← from $80 \times$ binompdf(5, 0.4)

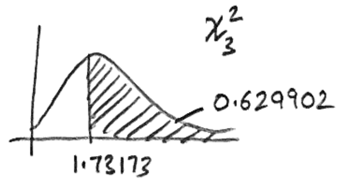
now one of f_e is < 1, so we need to combine categories:

x	0	1	2	3	4 - 5
f_o	8	19	25	22	6
f_e	6.2	20.7	27.6	18.4	6.9

we conduct a χ^2 test with $\alpha = 10\%$ and H_0: data fits $B(5, 0.4)$ and $df = 5 - 1 - 1 = 3$.

H_1: data not fit $B(5, 0.4)$

$$X^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 1.73173$$



so $P(X^2 > 1.73173) = 0.629902 > 0.10$ so we do not reject H_0.

So no evidence to suggest that the data does not fit $B(5, 0.4)$

c) If $B(5, 0.4)$ is a good model, then $P(X=5) = 0.01024$

This probability is just more than 0.01 (1 in 100), and so mathematically, she should win in the long run. But it would be a very long time, given how close 0.01024 is to 0.01.

I would advise her to not continue with the competition, as she will not likely make much money on it. However, if they raised the odds to, say, 1 in 200, it would be worth playing!

4.

x	0	1	2	3	4	5+	
f_o	180	173	69	20	6	2	$\sum f_o = 450.$
f_e	182.9	164.7	74.1	22.2	5.0	1.05	$\sum f_e = 450.$

we have $\bar{x} = 0.9$ (from 1-var stats)

we also have $s^2 = 0.914254 \approx 0.9$

so it's sensible to proceed with Poi(0.9) model

all $f_e > 1$ ✓

and $\frac{5}{6}$ of $f_e \geq 5$ ✓ so no need to combine categories.

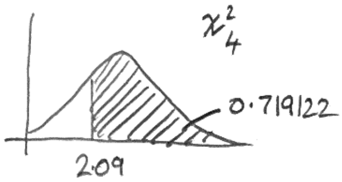
b) H_0: data fits Po(0.9)

H_1: data does not fit Po(0.9)

Assume H_0 to be true

$\alpha = 0.10$ one tail test

$df = 6 - 1 - 1 = 4$



$X^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 2.09048$

$P(X^2 > 2.09048) = 0.719122$

> 0.10

so do not reject H_0

we do not have evidence to suggest that the data does not fit Po(0.9)