GOODNESS OF FIT

DISCRETE DATA

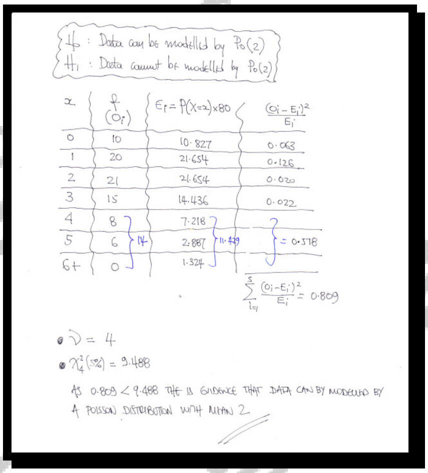
Question 1 (**)

The number of car immobilizations, carried out by a security company patrolling a car park, over a period of eighty days is summarized in the table below.

No of Immobilizations	No of Days
0	10
1	20
2	21
3	15
4	8
5	6
6+	0

Use a χ^2 test, at 5% level of significance, to investigate whether the above data can be modelled by a Poisson distribution with a mean of 2 immobilizations per day.

excellent fit, $0.809 < 9.488$



H_0 : Data can be modelled by Po(2)
H_1 : Data cannot be modelled by Po(2)

x	f (O_i)	$E_i = P(X=x) \times 80$	$\frac{(O_i - E_i)^2}{E_i}$
0	10	10.827	0.063
1	20	21.654	0.126
2	21	21.654	0.020
3	15	14.436	0.022
4	8 } 14	7.218 } 11.429	} = 0.578
5	6	2.887	
6+	0	1.324	

$\sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} = 0.809$

- $\nu = 4$
- $\chi_4^2(5\%) = 9.488$

AS $0.809 < 9.488$ THE IS EVIDENCE THAT DATA CAN BY MODELLED BY A POISSON DISTRIBUTION WITH MEAN 2

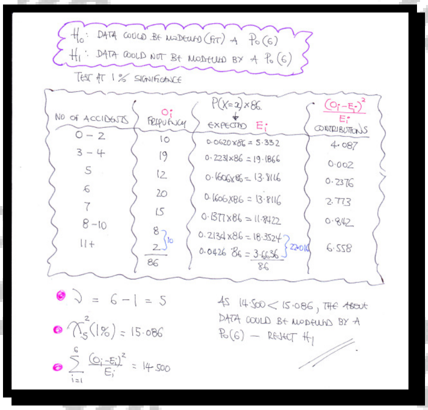
Question 2 (**)

The number of accidents, a certain police force were asked to attend, , over a period of 86 days is summarized in the table below.

No of Accidents	No of Days
0 – 2	10
3 – 4	19
5	12
6	20
7	15
8 – 10	8
11+	2

Use a χ^2 test, at 1% level of significance, to investigate whether the above data can be modelled by a Poisson distribution with mean 6 accidents per day.

good fit, $14.500 < 15.086$



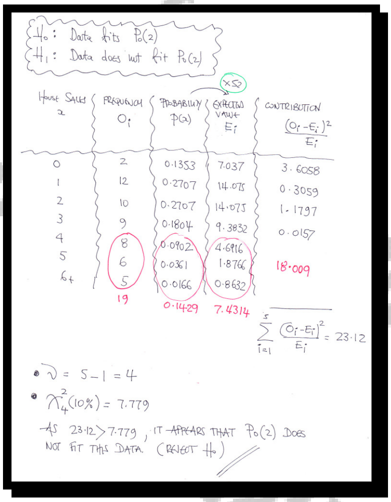
Question 3 (**)

The number of house sales achieved by an estate agent per week over a period of fifty two weeks is summarized in the table below.

No of house sales	No of weeks
0	2
1	12
2	10
3	9
4	8
5	6
6+	5

Use a χ^2 test, at 10% level of significance, to investigate whether the above data can be modelled by a Poisson distribution with a mean of 2 sales per week.

not a good fit, $23.12 > 7.779$



Question 4 (**)

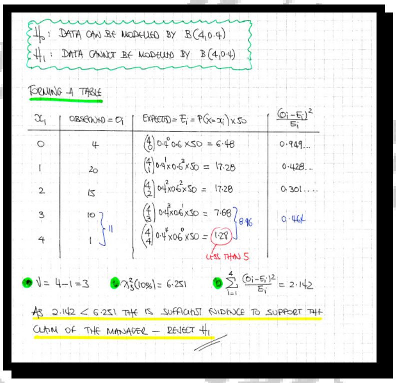
An airport manager believes the number of suitcases checked in by passengers follows a binomial distribution with $p = 0.4$.

He records the number of suitcases checked in by a group of 50 passengers, and his results are summarized in the table below.

Number of Suitcases	Frequency
0	4
1	20
2	15
3	10
4	1

Use a χ^2 test, at 5% level of significance, to investigate the validity of the airport manager's claim.

[ ] , [good fit, $2.142 < 6.251$]



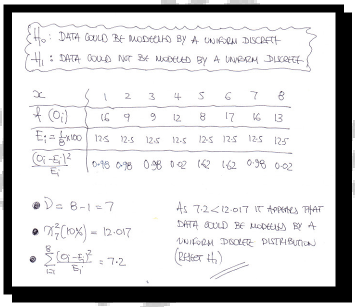
Question 5 (**)

The table summarizes the results obtained by a spinner numbered with the positive integers from 1 to 8.

Number	1	2	3	4	5	6	7	8
Frequency	16	9	9	12	8	17	16	13

Use a χ^2 test, at 10% level of significance, to investigate whether the above data can be modelled by a Discrete Uniform distribution.

good fit, $7.2 < 12.017$



H_0 : DATA COULD BE MODELLED BY A UNIFORM DISCRETE

H_1 : DATA COULD NOT BE MODELLED BY A UNIFORM DISCRETE

x	1	2	3	4	5	6	7	8
$f\ (O_i)$	16	9	9	12	8	17	16	13
$E_i = \frac{1}{8} \times 100$	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5
$\frac{(O_i - E_i)^2}{E_i}$	0.98	0.98	0.98	0.02	1.62	1.62	0.98	0.02

- $D = 8 - 1 = 7$
- $\chi^2_7(10\%) = 12.017$
- $\sum_{i=1}^{8} \frac{(O_i - E_i)^2}{E_i} = 7.2$

AS $7.2 < 12.017$ IT APPEARS THAT DATA COULD BE MODELLED BY A UNIFORM DISCRETE DISTRIBUTION (REJECT H_1)

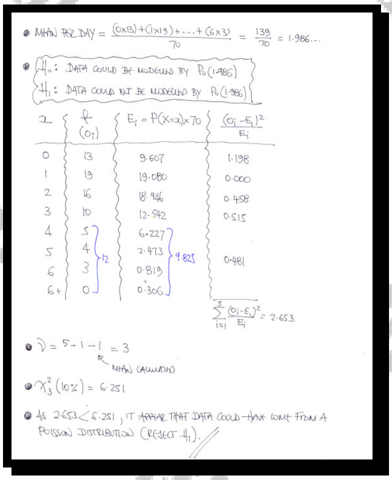
Question 6 (**+)

The number of shoplifting incidents, dealt by the security company patrolling a shopping centre, over a period of seventy days is summarized in the table below.

No of Incidents	No of Days
0	13
1	19
2	16
3	10
4	5
5	4
6	3

Use a χ^2 test, at 10% level of significance, to investigate whether the above data can be modelled by a Poisson distribution with parameter λ, where λ is the number of shoplifting incidents per day.

good fit, $2.653 < 6.251$


- MEAN PER DAY $= \frac{(0\times13)+(1\times19)+\dots+(6\times3)}{70} = \frac{139}{70} = 1.986\dots$
- H_0: DATA COULD BE MODELLED BY $Po(1.986)$
  H_1: DATA COULD NOT BE MODELLED BY $Po(1.986)$

x	f (O_i)	$E_i = P(X=x)\times70$	$\frac{(O_i-E_i)^2}{E_i}$
0	13	9.607	1.198
1	19	19.080	0.000
2	16	18.946	0.458
3	10	12.542	0.515
4	5 } 12	6.227 } 9.825	0.481
5	4	2.473	
6	3	0.819	
6+	0	0.306	

$\sum_{i=1}^{5} \frac{(O_i-E_i)^2}{E_i} = 2.653$

- $\nu = 5-1-1 = 3$ (MEAN CALCULATED)
- $\chi^2_3(10\%) = 6.251$
- AS $2.653 < 6.251$, IT APPEAR THAT DATA COULD HAVE COME FROM A POISSON DISTRIBUTION (REJECT H_1).

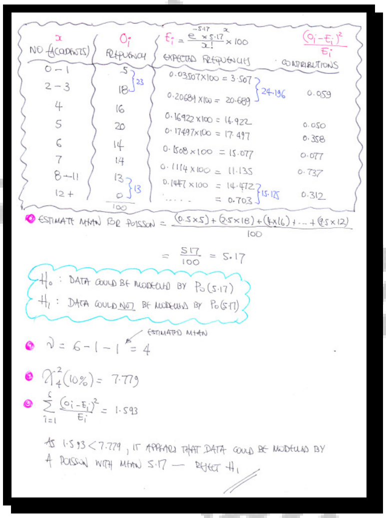
Question 7 (**+)

The admissions due to accidents, dealt by the Accident and Emergency Department of a hospital over a period of 100 hours, is summarized in the table below.

No of Accidents	No of Hourly Periods
0 – 1	5
2 – 3	18
4	16
5	20
6	14
7	14
8 – 11	13
12+	0

Use a χ^2 test, at 10% level of significance, to investigate whether the above data can be modelled by a Poisson distribution with mean λ, where λ is the mean number of accident admissions per hour.

excellent fit, $1.593 < 7.779$



Question 8 (**+)

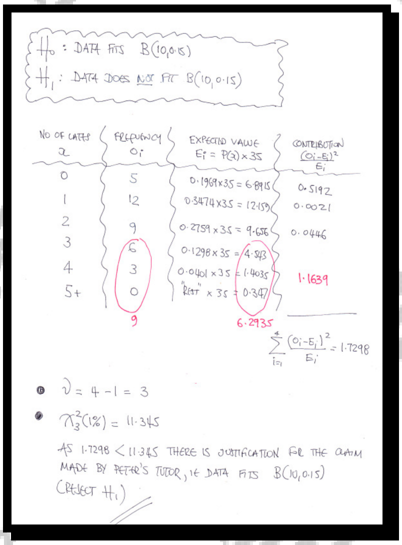
Peter's tutor feels he is late in 15% of the daily school registrations, which occur twice daily, i.e. 10 in total in a five day week.

The actual data is summarized in the table below

Late Registrations in a Week	No of Weeks
0	5
1	12
2	9
3	6
4	3

Use a χ^2 test, at 1% level of significance, to investigate whether the above data can be modelled by a Binomial distribution, as claimed by Peter's tutor.

very good fit, $1.730 < 11.345$


H_0 : DATA FITS $B(10, 0.15)$
H_1 : DATA DOES NOT FIT $B(10, 0.15)$

NO OF LATES x	FREQUENCY O_i	EXPECTED VALUE $E_i = P(x) \times 35$	CONTRIBUTION $\frac{(O_i - E_i)^2}{E_i}$
0	5	$0.1969 \times 35 = 6.8915$	0.5192
1	12	$0.3474 \times 35 = 12.159$	0.0021
2	9	$0.2759 \times 35 = 9.656$	0.0446
3	6	$0.1298 \times 35 = 4.543$	
4	3	$0.0401 \times 35 = 1.4035$	1.1639
5+	0	"REST" $\times 35 = 0.347$	
	9	6.2935	

$\sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} = 1.7298$

- $\nu = 4 - 1 = 3$
- $\chi_3^2(1\%) = 11.345$

AS $1.7298 < 11.345$ THERE IS JUSTIFICATION FOR THE CLAIM MADE BY PETER'S TUTOR, IE DATA FITS $B(10, 0.15)$ (REJECT H_1)

Question 9 **(***)**

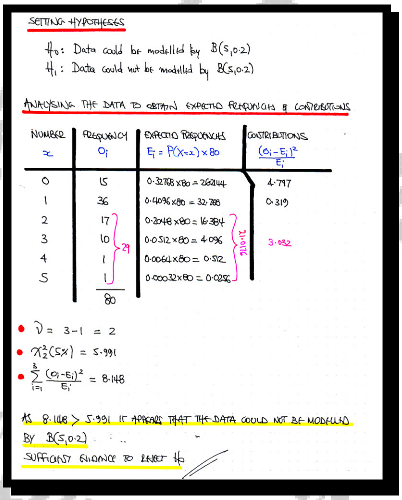
The discrete variable X is thought to have distribution $\mathrm{B}(5,0.2)$.

Some actual observations of X are summarized in the table below.

X	**Frequency**
0	15
1	36
2	17
3	10
4	1
5	1

Use a χ^2 test, at 5% level of significance, to investigate whether the above data can be modelled by $\mathrm{B}(5,0.2)$.

[ ] , not a good fit, $8.148 > 5.991$



Question 10 (***)

A bank manager investigates the number of customers served by his staff.

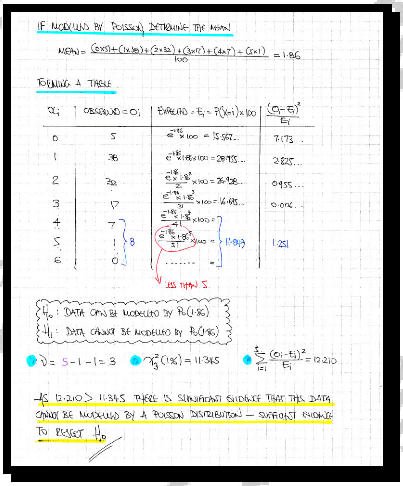
He records the number of customers being served in 100 consecutive five minute time intervals and his data is summarized in the table below.

Number of Customers	Frequency
0	5
1	38
2	32
3	17
4	7
5	1
6 or more	0

The manager further asserts that his data can be modelled by a Poisson distribution.

Use a χ^2 test, at 1% level of significance, to investigate the validity of the manager's assertion.

, not a good fit, $12.210 > 11.345$


IF MODELLED BY POISSON, DETERMINE THE MEAN

$$\text{MEAN} = \frac{(0\times5)+(1\times38)+(2\times32)+(3\times17)+(4\times7)+(5\times1)}{100} = 1.86$$

FORMING A TABLE

x_i	OBSERVED $= O_i$	EXPECTED $= E_i = P(X=i)\times100$	$\frac{(O_i-E_i)^2}{E_i}$
0	5	$e^{-1.86}\times100 = 15.567\ldots$	7.173...
1	38	$e^{-1.86}\times1.86\times100 = 28.955\ldots$	2.825...
2	32	$\frac{e^{-1.86}\times1.86^2}{2}\times100 = 26.928\ldots$	0.955...
3	17	$\frac{e^{-1.86}\times1.86^3}{3!}\times100 = 16.695\ldots$	0.006...
4	7 } 8	$\frac{e^{-1.86}\times1.86^4}{4!}\times100 =$ } 11.849	1.251
5	1	$\frac{e^{-1.86}\times1.86^5}{5!}\times100 =$	
6	0	...... =	

LESS THAN 5

H_0: DATA CAN BE MODELLED BY $Po(1.86)$
H_1: DATA CANNOT BE MODELLED BY $Po(1.86)$

$\nu = 5-1-1 = 3$ • $\chi^2_3(1\%) = 11.345$ • $\sum_{i=1}^{5}\frac{(O_i-E_i)^2}{E_i} = 12.210$

AS $12.210 > 11.345$ THERE IS SIGNIFICANT EVIDENCE THAT THIS DATA CANNOT BE MODELLED BY A POISSON DISTRIBUTION — SUFFICIENT EVIDENCE TO REJECT H_0

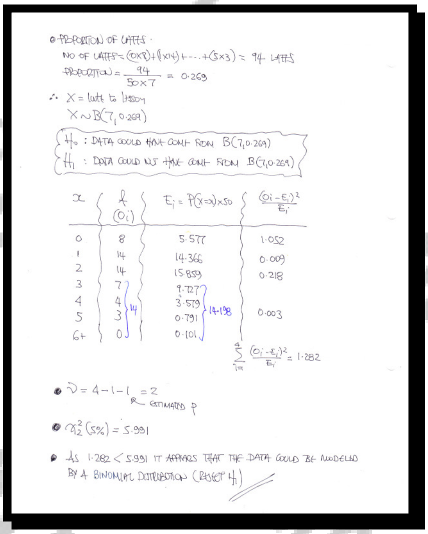
Question 11 (***)

There are 7 periods in Lilith's school. The number of late arrivals to her lessons for a random sample of 50 days is summarized in the table below.

Late Arrivals in a Day	No of Days
0	8
1	14
2	14
3	7
4	4
5	3

Use a χ^2 test, at 5% level of significance, to investigate whether the above data can be modelled by a Binomial distribution.

good fit, $1.282 < 5.991$


• PROPORTION OF LATES:
NO OF LATES $= (0\times8)+(1\times14)+\dots+(5\times3) = 94$ LATES
PROPORTION $= \frac{94}{50\times7} = 0.269$
$\therefore$ X = late to lesson
$X \sim B(7, 0.269)$
H_0 : DATA COULD HAVE COME FROM $B(7, 0.269)$
H_1 : DATA COULD NOT HAVE COME FROM $B(7, 0.269)$

x	f (O_i)	$E_i = P(X=x)\times 50$	$\frac{(O_i-E_i)^2}{E_i}$
0	8	5.577	1.052
1	14	14.366	0.009
2	14	15.859	0.218
3	7	9.727	
4	4 } 14	3.579 } 14.198	0.003
5	3	0.791	
6+	0	0.101	

$\sum_{i=1}^{4} \frac{(O_i-E_i)^2}{E_i} = 1.282$

• $\nu = 4-1-1 = 2$ (ESTIMATED p)
• $\chi^2_2(5\%) = 5.991$
• AS $1.282 < 5.991$ IT APPEARS THAT THE DATA COULD BE MODELLED BY A BINOMIAL DISTRIBUTION (REJECT H_1)

Question 12 (***)

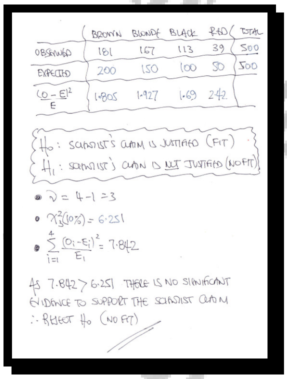
A scientist believes that the proportion of people in Britain whose hair colour is brown, blonde, black or red, is in the ratio $4:3:2:1$.

A random sample is taken and their hair colour is recorded.

- 181 had brown hair.
- 167 had blonde hair.
- 113 had black hair.
- 39 had red hair.

Test the scientist's claim at the 10% level of significance.

claim not justified, $7.842 < 6.251$



	BROWN	BLONDE	BLACK	RED	TOTAL
OBSERVED	181	167	113	39	500
EXPECTED	200	150	100	50	500
$\frac{(O-E)^2}{E}$	1.805	1.927	1.69	2.42	

H_0: SCIENTIST'S CLAIM IS JUSTIFIED (FIT)
H_1: SCIENTIST'S CLAIM IS NOT JUSTIFIED (NO FIT)

- $\nu = 4-1 = 3$
- $\chi^2_3(10\%) = 6.251$
- $\sum_{i=1}^{4} \frac{(O_i-E_i)^2}{E_i} = 7.842$

AS $7.842 > 6.251$ THERE IS NO SIGNIFICANT EVIDENCE TO SUPPORT THE SCIENTIST CLAIM
$\therefore$ REJECT H_0 (NO FIT)

Question 13 (**+)

A spinner numbered 1, 2, 3 and 4 is spun 100 times and the results are summarized in the table below

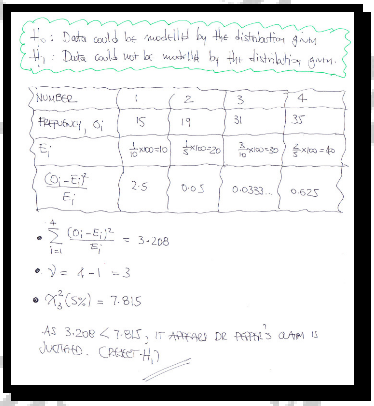
Number	Frequency
1	15
2	19
3	31
4	35

Dr Pepper claims that the probability distribution of number shown on the spinner is given by the following table.

number	1	2	3	4
probability	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{2}{5}$

Use a χ^2 test, at 5% level of significance, to investigate Dr Pepper's claim.

good fit, $3.208 < 7.815$



H_0: Data could be modelled by the distribution given

H_1: Data could not be modelled by the distribution given.

NUMBER	1	2	3	4
FREQUENCY, O_i	15	19	31	35
E_i	$\frac{1}{10}\times100=10$	$\frac{1}{5}\times100=20$	$\frac{3}{10}\times100=30$	$\frac{2}{5}\times100=40$
$\frac{(O_i-E_i)^2}{E_i}$	2.5	0.05	0.0333...	0.625

- $\sum_{i=1}^{4}\frac{(O_i-E_i)^2}{E_i} = 3.208$
- $\nu = 4-1 = 3$
- $\chi^2_3(5\%) = 7.815$

AS $3.208 < 7.815$, IT APPEARS DR PEPPER'S CLAIM IS JUSTIFIED. (REJECT H_1)

Question 14 (***+)

A clothes company has five shops.

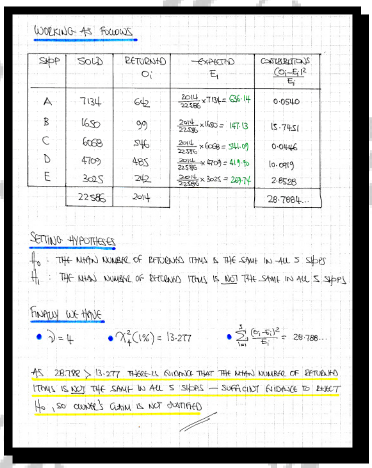
The items sold and the items returned in these shops, for the month of August, are summarised in the table below.

Shop	Sold Items	Returned Items
A	7134	642
B	1650	99
C	6068	546
D	4709	485
E	3025	242

The owner claims that the mean number of returned items per 100 items sold is the same in all five shops.

Test the owner's claim, at the 1% level of significance, stating your hypotheses clearly.

[ ] , not justified, $28.788 > 13.277$


WORKING AS FOLLOWS

SHOP	SOLD	RETURNED O_i	EXPECTED E_i	CONTRIBUTIONS $\frac{(O_i-E_i)^2}{E_i}$
A	7134	642	$\frac{2014}{22586}\times7134 = 636.14$	0.0540
B	1650	99	$\frac{2014}{22586}\times1650 = 147.13$	15.7451
C	6068	546	$\frac{2014}{22586}\times6068 = 541.09$	0.0446
D	4709	485	$\frac{2014}{22586}\times4709 = 419.90$	10.0919
E	3025	242	$\frac{2014}{22586}\times3025 = 269.74$	2.8528
	22586	2014		28.7884...

SETTING HYPOTHESES

H_0 : THE MEAN NUMBER OF RETURNED ITEMS IS THE SAME IN ALL 5 SHOPS
H_1 : THE MEAN NUMBER OF RETURNED ITEMS IS NOT THE SAME IN ALL 5 SHOPS

FINALLY WE HAVE

- $\nu = 4$ • $\chi^2_4(1\%) = 13.277$ • $\sum_{i=1}^{5}\frac{(O_i-E_i)^2}{E_i} = 28.788...$

AS $28.788 > 13.277$ THERE IS EVIDENCE THAT THE MEAN NUMBER OF RETURNED ITEMS IS NOT THE SAME IN ALL 5 SHOPS — SUFFICIENT EVIDENCE TO REJECT H_0, SO OWNER'S CLAIM IS NOT JUSTIFIED

Question 15 (***+)

An investigation was carried out to determine the effectiveness of four different blood pressure lowering medications.

Each of the 100 patients who took part in the investigation was given one of the four available medications A, B, C or D.

- 10 of the 17 patients that were given medication A had a positive response.
- 16 of the 26 patients that were given medication B had a positive response.
- 15 of the 28 patients that were given medication C had a positive response.
- 11 of the 29 patients that were given medication D had a positive response.

The following claims are made.

a) Claim 1

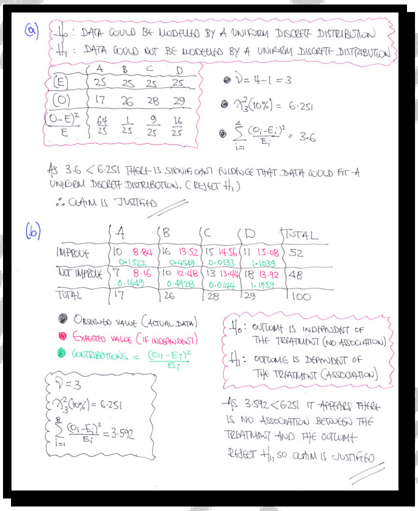
Each patient was randomly given one of the four medications.

b) Claim 2

The patient's response is independent of the medication that was given.

Test each of these claims at the 10% level of significance.

$\boxed{\text{claim 1 justified, } 3.6 < 6.251}$, $\boxed{\text{claim 2 justified, } 3.592 < 6.251}$



CONTINUOUS DATA

Question 1 (+)**

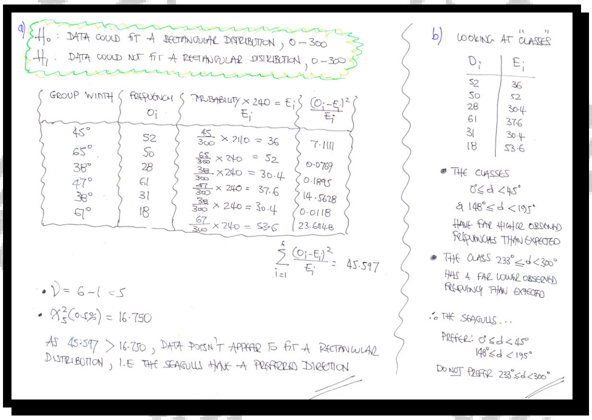
A flock of seagulls have nests on a rock face, and when they leave their nests the direction d they headed were recorded as a bearing, using certain landscape features.

Bearing d (degrees)	**No of seagulls**
$0 \le d < 45$	52
$45 \le d < 110$	50
$110 \le d < 148$	28
$148 \le d < 195$	61
$195 \le d < 233$	31
$233 \le d < 300$	18

a) Show, by using a χ^2 test at 0.5% level of significance, that seagulls have a preferred direction when flying off their nests.

b) State with justification which are, and which are not, the preferred directions of these seagulls.

preferred: $0° \le d < 45°$ & $148° \le d < 195°$
not preferred: $233° \le d < 300°$


a) H_0: DATA COULD FIT A RECTANGULAR DISTRIBUTION, 0 – 300
H_1: DATA COULD NOT FIT A RECTANGULAR DISTRIBUTION, 0 – 300

GROUP WIDTH	FREQUENCY O_i	PROBABILITY × 240 = E_i	$\frac{(O_i - E_i)^2}{E_i}$
45°	52	$\frac{45}{300} \times 240 = 36$	7.1111
65°	50	$\frac{65}{300} \times 240 = 52$	0.0769
38°	28	$\frac{38}{300} \times 240 = 30.4$	0.1895
47°	61	$\frac{47}{300} \times 240 = 37.6$	14.5628
38°	31	$\frac{38}{300} \times 240 = 30.4$	0.0118
67°	18	$\frac{67}{300} \times 240 = 53.6$	23.6448

$\sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} = 45.597$

• $\nu = 6 - 1 = 5$
• $\chi^2_5(0.5\%) = 16.750$

AS $45.597 > 16.750$, DATA DOESN'T APPEAR TO FIT A RECTANGULAR DISTRIBUTION, I.E THE SEAGULLS HAVE A PREFERRED DIRECTION

b) LOOKING AT "CLASSES"

O_i	E_i
52	36
50	52
28	30.4
61	37.6
31	30.4
18	53.6

• THE CLASSES $0° \le d < 45°$ & $148° \le d < 195°$ HAVE FAR HIGHER OBSERVED FREQUENCIES THAN EXPECTED
• THE CLASS $233° \le d < 300°$ HAS A FAR LOWER OBSERVED FREQUENCY THAN EXPECTED

∴ THE SEAGULLS...
PREFER: $0° \le d < 45°$
$148° \le d < 195°$
DO NOT PREFER $233° \le d < 300°$

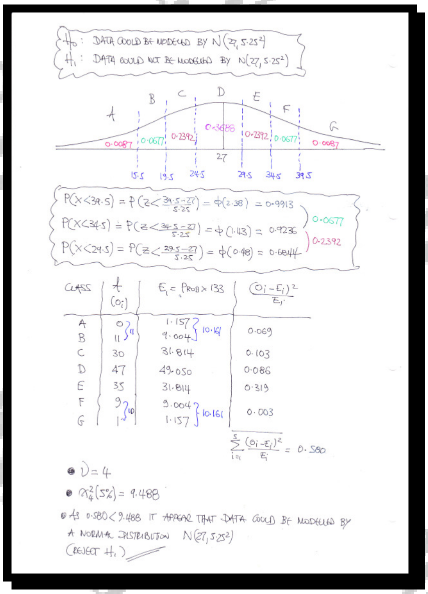
Question 2 (***)

The times taken to complete a four mile charity run for a group of people is summarized in the table below.

Time (nearest minute)	No of Runners
– 15	0
15 – 19	11
20 – 24	30
25 – 29	47
30 – 34	35
35 – 39	9
40 +	1

Use a χ^2 test, at 5% level of significance, to investigate whether the above data can be modelled by a Normal distribution with mean of 27 minutes and standard deviation $5\frac{1}{4}$ minutes.

excellent fit, $0.580 < 9.488$



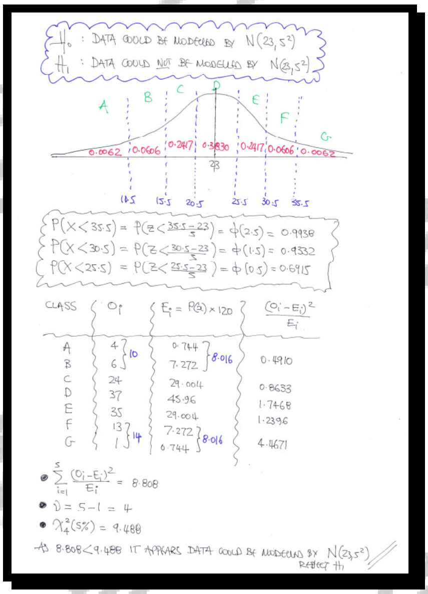
Question 3 (***)

The time that a group of people could hold the breath, rounded to the nearest second, is summarized in the table below.

Time (nearest second)	No of People
– 10	4
11 – 15	6
16 – 20	24
21 – 25	37
26 – 30	35
31 – 35	13
36 +	1

Use a χ^2 test, at 5% level of significance, to investigate whether the above data can be modelled by a Normal distribution with mean of 23 seconds and standard deviation 5 seconds.

good fit, $8.808 < 9.488$



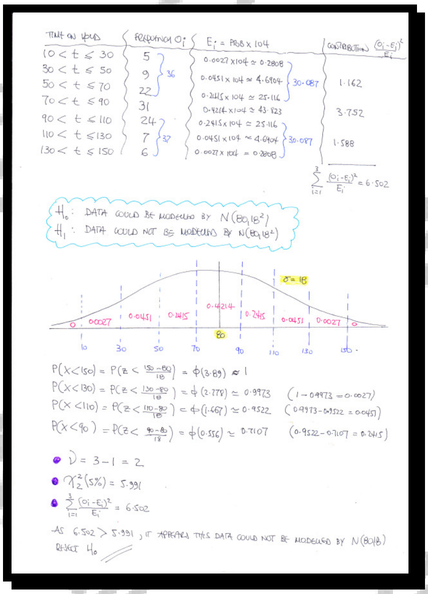
Question 4 (***)

The time spent waiting by 104 callers, before their call is answered by a local hospital, is summarized in the table below.

Waiting Time (t)	No of Callers
$10 < t \le 30$	5
$30 < t \le 50$	9
$50 < t \le 70$	22
$70 < t \le 90$	31
$90 < t \le 110$	24
$110 < t \le 130$	7
$130 < t \le 150$	6

Use a χ^2 test, at 5% level of significance, to investigate whether the above data can be modelled by a Normal distribution with mean of 80 seconds and standard deviation 18 seconds.

not a good fit, $6.502 > 5.991$



Question 5 (***+)

It is suggested that the daily takings in a shop X, in thousands of £, can be modelled by the probability density function

$$f(x) = \begin{cases} \frac{1}{30}(2x+1) & 0 \le x < 5 \\ 0 & \text{otherwise} \end{cases}$$

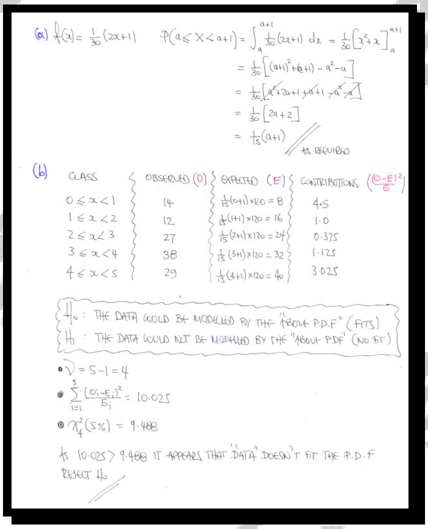
a) Show that according to this model, $\mathrm{P}(a \le X < a+1) = \frac{1}{15}(a+1)$.

The takings in 120 randomly selected days are summarized in the table below.

Takings x (£ 000)	No of days
$0 \le x < 1$	14
$1 \le x < 2$	12
$2 \le x < 3$	27
$3 \le x < 4$	38
$4 \le x < 5$	29

b) Use a χ^2 test, at 5% level of significance, to investigate whether the daily takings in this shop could be modelled by the probability density function of part (a).

not a good fit, $10.025 > 9.488$



(a) $f(x) = \frac{1}{30}(2x+1)$ $\quad \mathrm{P}(a \le X < a+1) = \int_a^{a+1} \frac{1}{30}(2x+1)\,dx = \frac{1}{30}\left[x^2+x\right]_a^{a+1}$

$= \frac{1}{30}\left[(a+1)^2+(a+1)-a^2-a\right]$

$= \frac{1}{30}\left[a^2+2a+1+a+1-a^2-a\right]$

$= \frac{1}{30}\left[2a+2\right]$

$= \frac{1}{15}(a+1)$ // AS REQUIRED

(b)

CLASS	OBSERVED (O)	EXPECTED (E)	CONTRIBUTIONS $\left(\frac{(O-E)^2}{E}\right)$
$0 \le x < 1$	14	$\frac{1}{15}(0+1)\times 120 = 8$	4.5
$1 \le x < 2$	12	$\frac{1}{15}(1+1)\times 120 = 16$	1.0
$2 \le x < 3$	27	$\frac{1}{15}(2+1)\times 120 = 24$	0.375
$3 \le x < 4$	38	$\frac{1}{15}(3+1)\times 120 = 32$	1.125
$4 \le x < 5$	29	$\frac{1}{15}(4+1)\times 120 = 40$	3.025

H_0 : THE DATA COULD BE MODELLED BY THE "ABOVE P.D.F" (FITS)
H_1 : THE DATA COULD NOT BE MODELLED BY THE "ABOVE PDF" (NO FIT)

- $\nu = 5-1 = 4$
- $\sum_{i=1}^{5} \frac{(O_i-E_i)^2}{E_i} = 10.025$
- $\chi_4^2(5\%) = 9.488$

AS $10.025 > 9.488$ IT APPEARS THAT "DATA" DOESN'T FIT THE P.D.F
REJECT H_0 //

Question 6 (****)

The length of certain type of fresh water eel is investigated by a marine biologist.

The lengths of 100 such eels are summarized in the table below.

Length l (cm)	**No of Eels**
$10 \le l < 15$	9
$15 \le l < 20$	12
$20 \le l < 22$	17
$22 \le l < 25$	44
$25 \le l < 30$	17
$30 \le l < 35$	1

Use a χ^2 test, at 1% level of significance, to investigate whether the above data can be modelled by a Normal distribution.

not a good fit, $17.573 > 6.635$