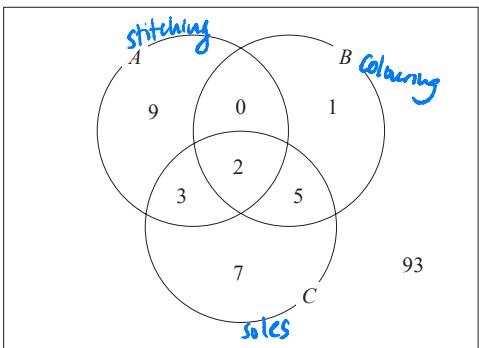


1. A factory produces shoes.

A quality control inspector at the factory checks a sample of 120 shoes for each of three types of defect. The Venn diagram represents the inspector's results.

- A represents the event that a shoe has defective stitching
- B represents the event that a shoe has defective colouring
- C represents the event that a shoe has defective soles



One of the shoes in the sample is selected at random.

- (a) Find the probability that it does **not** have defective soles.

(b) Find $P(A \cap B \cap C')$ shade only the overlap \cap

(c) Find $P(A \cup B \cup C')$ shade each part $\frac{113}{120}$

(d) Find the probability that the shoe has at most one type of defect. $\frac{110}{120}$

(e) Given the selected shoe has at most one type of defect, find the probability it has defective stitching. $\frac{9}{110}$

Leave
blank

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



General Patterns

Guaranteed Topics

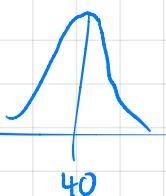
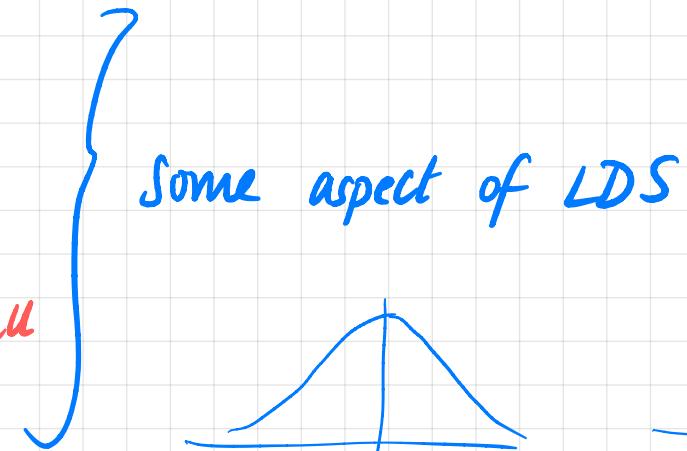
→ correlation/regression

→ normal dist

→ binomial dist

→ measures of location/spread *not covering in full*

→ conditional probability/Venn diagrams



$$\frac{x - 40}{5}$$

Possible topics (not covering in full)

→ histograms

→ probability tables/distributions

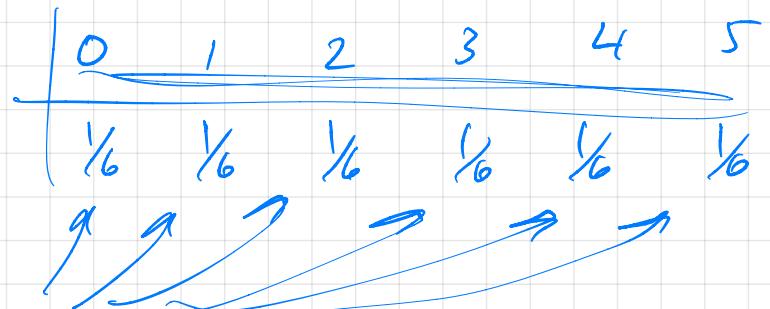
→ cumulative freq./box plots

→ tree diagrams

x	0	1	2	3
$p(x=x)$	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{4}$	$\frac{1}{25}$

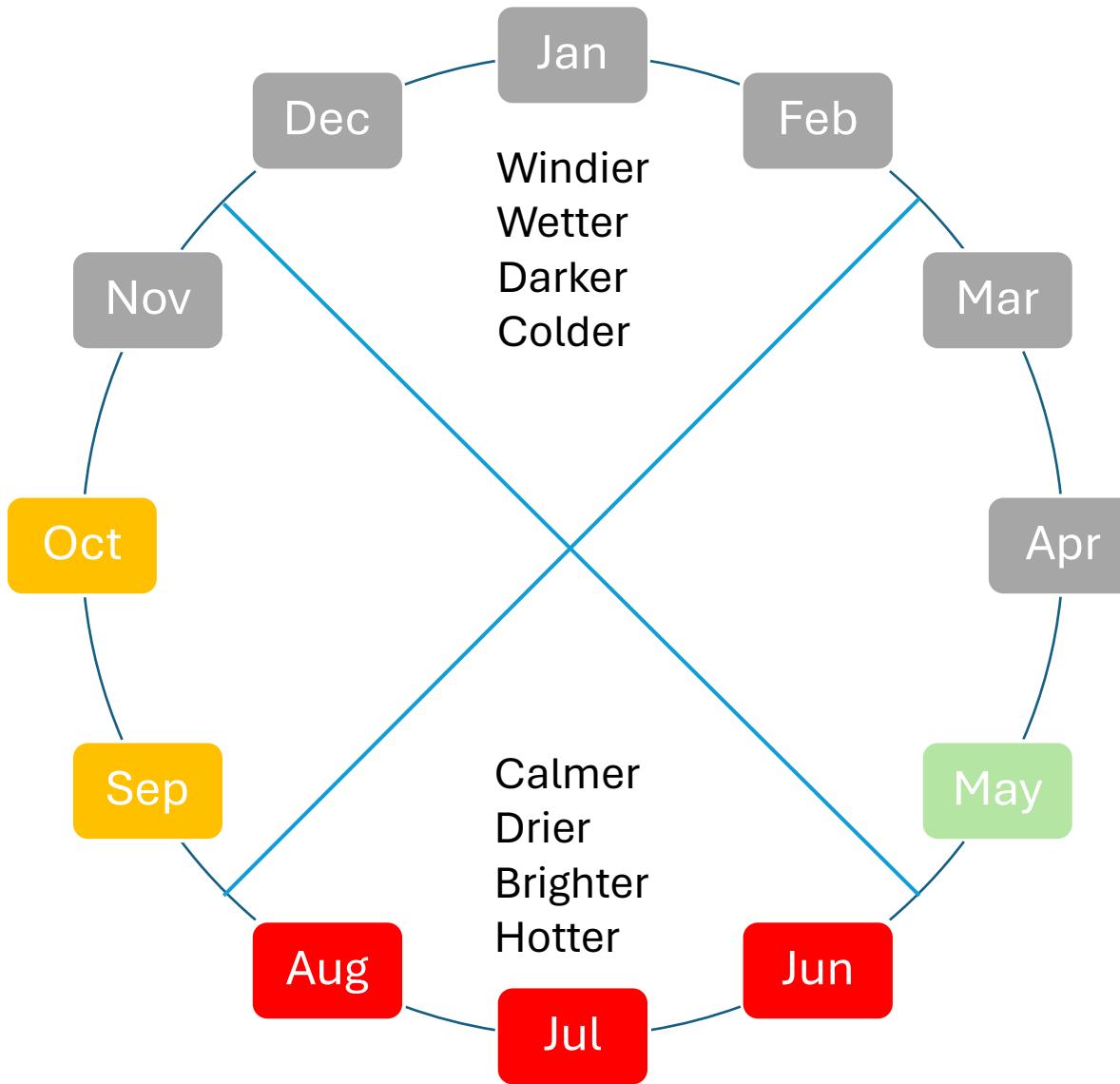
$p(x_1 + x_2) = 3$

Discrete uniform dist



The Large Data Set (Edexcel)

- In AS it has accounted for 6.7% of all the stats marks, and only 1.25% of the overall AS marks
- In A2 it has accounted for a total of 5.3% of all the stats marks, and only 0.9% of the overall A2 marks
- I'm going to give you the best chance of being successful with the questions!



We only have data for May to October.

We have data from 1987 and 2015, so can compare the same month in two different years.



From South to North:
(alphabetical order except H and H switch)

Camborne (coastal – windier)

Hurn

Heathrow

Leeming

Leuchars (coastal – windier)

Common sense box:

If the temperature increases...

... the amount of sunshine _____

... the amount of rainfall _____

... the windspeed _____

As we move further north, during May to October...

... the temperature _____

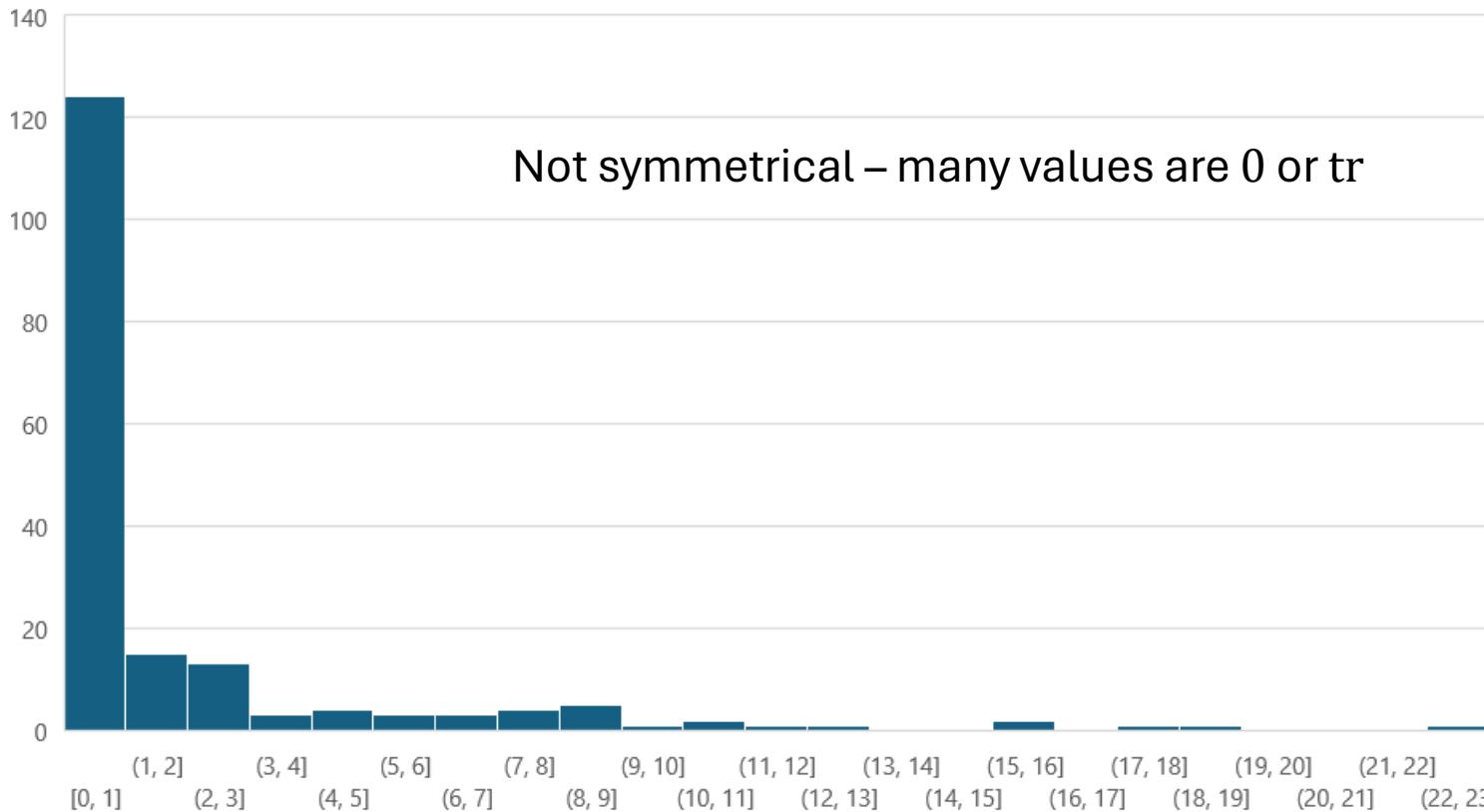
... the amount of rainfall _____

... the maximum hours of sunshine _____

UK Measurements pt. 1

Measurement		Units	Typical range	Examples	Details
Daily mean temperature	<i>How hot it is</i>	°C	~5°C to ~24°C	12.0°C 14.8°C	Warmer in summer
Daily total rainfall	<i>How much it rained</i>	mm	0 to ~20mm	0 mm 10.4 mm tr	Not symmetrical – many 0 and tr values
Daily total sunshine	<i>How many hours of sunshine</i>	hours	0 to ~14 hours	3.3 hrs 10.4 hrs	More sunshine in the summer
Cloud cover	<i>How much of the sky is covered in clouds</i>	oktas	0 to 8	3 4 0	Integers, measuring what fraction of the sky is covered
Humidity	<i>How much water vapour is in the air – above 95% is associated with fog</i>	%	~70% to 100%	100% 77% 95%	Integers
Daily mean visibility	<i>How far you can see</i>	Dm 1 Dm = 10 m 'decametre'	~200 Dm to ~4000 Dm	1300 Dm 2200 Dm 3100 Dm	Rounded to nearest 100
Daily mean pressure	<i>How much the atmosphere is pushing down</i>	hPa 'hectopascals'	~990 hPa to ~1040 hPa	1017 hPa 1006 hPa 997 hPa	Integers

Rainfall



‘Cleaning’ data

‘tr’ means there was a trace of water in the measuring instrument – it is used for values of rainfall, $0 < r \leq 0.05$. We can ‘clean’ data, which means replacing ‘tr’ with either 0 mm or 0.025 mm.

If we have n/a in any of our data, we cannot use it. Cleaning it means to remove that entry.

UK Measurements pt. 2 – wind

Measurement		Units	Typical range	Examples	Details
Daily mean windspeed	<i>How windy it is</i>	kn (knots)	~3 kn to ~10 kn	4 kn 11 kn	Integers only
Windspeed, Beaufort conversion		Qualitative	Light to Moderate	Light Moderate Fresh Strong	Qualitative. Most days are ‘light’
Daily maximum gust	<i>The strongest gust of wind that day</i>	kn (knots)	~8 kn to ~50 kn	17 kn 25 kn	Integers only
Wind/gust direction (bearings)	<i>Which direction the wind is blowing from</i>	°	10° to 360°	240° 70°	Multiples of 10 only
Wind/gust direction (cardinal)	<i>Which direction the wind is blowing from</i>	Compass direction	—	N SW ENE	Describes where the wind is blowing from, not to



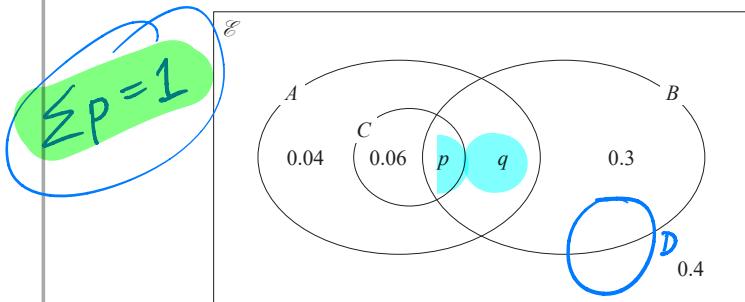
International Measurements

Measurement		Units	Details
Daily mean temperature	How hot it is	°C	Warmer in summer... but Perth is colder
Daily total rainfall	How much it rained	mm	Beijing is rainy in the summer
Daily mean pressure	How much the atmosphere is pushing down	hPa 'hectopascals'	
Daily mean windspeed	<i>How windy it is</i>	kn (knots)	Now are rounded to 1 dp
Windspeed, Beaufort conversion		Light to Moderate	Qualitative. Most days are 'light'

Conditional Probability/Venn Diagrams

- Sometimes looking at the 'complement' (not) area on a Venn can help
- When they give a fact, translate it to an equation using a formula
- $P(A|B) = \frac{P(A \cap B)}{P(B)}$ (In formula booklet it is $P(A \cap B) = P(A)P(B|A)$)
which rearranges to $\frac{P(A \cap B)}{P(A)} = P(B|A)$
- $P(A \cap B) = P(A) \times P(B)$ (If independent) $P(A|B) = P(A)$
- Mutually exclusive - cannot happen at same time, circles won't overlap at all

1. The Venn diagram shows the events A , B and C and their associated probabilities, where p and q are probabilities.



(a) Find $P(B)$

$$0.5$$

(b) Determine whether or not A and B are independent.

Given that $P(C|B) = P(C)$ indep.

(c) find the value of p and the value of q

The event D is such that

- A and D are mutually exclusive
- $P(B \cap D) > 0$

(d) On the Venn diagram show a possible position for the event D

Leave blank

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(1)
(2)
(3)

(1)

If independent then...

b) $P(A \cap B) = P(A) \times P(B)$

$$P(A \cap B) = p + q = 0.2$$

$$P(A) = 0.3$$

$$P(B) = 0.5$$

$$P(A) \times P(B) = 0.3 \times 0.5 = 0.15$$

Hence $P(A \cap B) \neq P(A) \times P(B)$
so A and B are not indep.

c) $P(C|B) = P(C)$

$$\frac{P(C \cap B)}{P(B)} = P(C)$$

$$P(C \cap B) = P(C) \times P(B)$$

$$p = (0.06 + p) \times 0.5$$

$$2p = 0.06 + p$$

$$p = 0.06$$

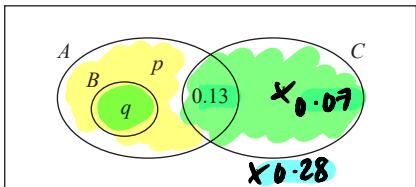
$$q = 0.2 - 0.06 = 0.14$$

YOUR TURN

IAL S1 Jun 21

2. In the Venn diagram below, A , B and C are events and p , q , r and s are probabilities.

The events A and C are independent and $P(A) = 0.65$



- (a) State which two of the events A , B and C are mutually exclusive. (1)
- (b) Find the value of r and the value of s . (5)

The events $(A \cap C')$ and $(B \cup C)$ are also independent.

- (c) Find the exact value of p and the exact value of q . Give your answers as fractions. (6)

a) B and C

$$b) P(A \cap C) = P(A) \times P(C)$$

$$0.13 = 0.65(0.13 + r)$$

$$\frac{0.13}{0.65} = 0.13 + r$$

$$0.2 = 0.13 + r$$

$$\underline{\underline{0.07 = r}}$$

$$r+s=0.35$$

$$0.07 + s = 0.35$$

$$\underline{\underline{s = 0.28}}$$

$$c) P(A \cap C') = p+q \quad P(B \cup C) = q + 0.13 + r \\ = 0.52$$

$$0.52(q + 0.2) = q$$

$$0.52q + 0.104 = q$$

$$0.104 = 0.48q$$

$$q = \frac{13}{60}$$

$$p+q = 0.52$$

$$p = 0.52 - \frac{13}{60}$$

$$\underline{\underline{p = \frac{91}{300}}}$$

Leave blank

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Correlation and Regression

- Describe / Interpret correlation

+ or -

what it means in context.

- Hyp Testing - 3 easy marks!

$$H_0: \rho = 0$$

or

$$H_1: \rho \neq 0$$

or $\rho > 0$

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

- Compare your observed value from the sample with the critical value

... if more "extreme", there's evidence for correlation, so reject H_0 !

-1

← CR

* CR |

Critical Values for Correlation Coefficients

These tables concern tests of the hypothesis that a population correlation coefficient ρ is 0. The values in the tables are the minimum values which need to be reached by a sample correlation coefficient in order to be significant at the level shown, on a one-tailed test.

Product Moment Coefficient					Sample size, n	Spearman's Coefficient			
Level						Level			
0.10	0.05	0.025	0.01	0.005		0.05	0.025	0.01	
0.8000	0.9000	0.9500	0.9800	0.9900	4	1.0000	—	—	
0.6870	0.8054	0.8783	0.9343	0.9587	5	0.9000	1.0000	1.0000	
0.6084	0.7293	0.8114	0.8822	0.9172	6	0.8286	0.8857	0.9429	
0.5509	0.6694	0.7545	0.8329	0.8745	7	0.7143	0.7857	0.8929	
0.5067	0.6215	0.7067	0.7887	0.8343	8	0.6429	0.7381	0.8333	
0.4716	0.5822	0.6664	0.7498	0.7977	9	0.6000	0.7000	0.7833	
0.4428	0.5494	0.6319	0.7155	0.7646	10	0.5636	0.6485	0.7455	
0.4187	0.5214	0.6021	0.6851	0.7348	11	0.5364	0.6182	0.7091	
0.3981	0.4973	0.5760	0.6581	0.7079	12	0.5035	0.5874	0.6783	
0.3802	0.4762	0.5529	0.6339	0.6835	13	0.4835	0.5604	0.6484	
0.3646	0.4575	0.5324	0.6120	0.6614	14	0.4637	0.5385	0.6264	
0.3507	0.4409	0.5140	0.5923	0.6411	15	0.4464	0.5214	0.6036	
0.3383	0.4259	0.4973	0.5742	0.6226	16	0.4294	0.5029	0.5824	
0.3271	0.4124	0.4821	0.5577	0.6055	17	0.4142	0.4877	0.5662	
0.3170	0.4000	0.4683	0.5425	0.5897	18	0.4014	0.4716	0.5501	
0.3077	0.3887	0.4555	0.5285	0.5751	19	0.3912	0.4596	0.5351	
0.2992	0.3783	0.4438	0.5155	0.5614	20	0.3805	0.4466	0.5218	
0.2914	0.3687	0.4329	0.5034	0.5487	21	0.3701	0.4364	0.5091	
0.2841	0.3598	0.4227	0.4921	0.5368	22	0.3608	0.4252	0.4975	
0.2774	0.3515	0.4133	0.4815	0.5256	23	0.3528	0.4160	0.4862	
0.2711	0.3438	0.4044	0.4716	0.5151	24	0.3443	0.4070	0.4757	
0.2653	0.3365	0.3961	0.4622	0.5052	25	0.3369	0.3977	0.4662	
0.2598	0.3297	0.3882	0.4534	0.4958	26	0.3306	0.3901	0.4571	
0.2546	0.3233	0.3809	0.4451	0.4869	27	0.3242	0.3828	0.4487	
0.2497	0.3172	0.3739	0.4372	0.4785	28	0.3180	0.3755	0.4401	
0.2451	0.3115	0.3673	0.4297	0.4705	29	0.3118	0.3685	0.4325	
0.2407	0.3061	0.3610	0.4226	0.4629	30	0.3063	0.3624	0.4251	
0.2070	0.2638	0.3120	0.3665	0.4026	40	0.2640	0.3128	0.3681	
0.1843	0.2353	0.2787	0.3281	0.3610	50	0.2353	0.2791	0.3293	
0.1678	0.2144	0.2542	0.2997	0.3301	60	0.2144	0.2545	0.3005	
0.1550	0.1982	0.2352	0.2776	0.3060	70	0.1982	0.2354	0.2782	
0.1448	0.1852	0.2199	0.2597	0.2864	80	0.1852	0.2201	0.2602	
0.1364	0.1745	0.2072	0.2449	0.2702	90	0.1745	0.2074	0.2453	
0.1292	0.1654	0.1966	0.2324	0.2565	100	0.1654	0.1967	0.2327	

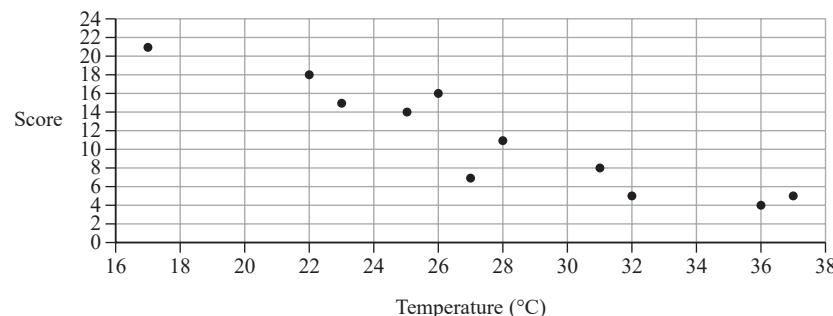
2. Xiang is investigating how room temperature affects a person's score in a task.

She gets Simon to complete the task 11 times at various controlled room temperatures, $x^{\circ}\text{C}$.

Xiang records the temperature, x , and Simon's score, y , where y is an integer.

$$\frac{11+1}{2} = 6^{\text{th}}$$

The results are shown in the scatter diagram below.



- (a) Use the scatter diagram to find

(i) the median score

11

(ii) the range of the scores. $21 - 4 = 17$

(2)

The temperature was increased each time Simon completed the task.

Xiang believes that as the room temperature increases, Simon's score will decrease.

Xiang calculates the product moment correlation coefficient from her data as -0.9286

- (b) Use this calculated value to carry out a suitable hypothesis test to investigate her belief at a 5% level of significance.

State clearly

- your hypotheses
- your critical value

(3)

Xiang is concerned that because Simon is repeating the same task his scores may improve.

- (c) Comment on how this concern may affect Xiang's conclusion to the test in part (b).

(1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

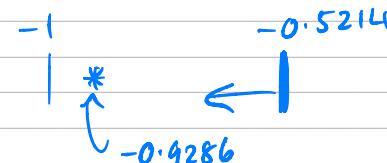
$$\alpha = 0.05$$

critical value is -0.5214

$$H_0: \rho = 0$$

$$H_1: \rho < 0$$

$$r = -0.9286$$



$$-0.9286 < -0.5214$$

reject H_0 , this supports Xiang's belief that higher temps means lower scores.

c) As temp increased his score decreased, even though he had done the task before, so conclusion is still true.



YOUR TURN

Set 1

2. An ornithologist believes that there is a relationship between the tail length, t mm, and the wing length, w mm, of female hook-billed kites. A random sample of size 10 is taken from a database of these kites and the relevant data is given in the table below.

$t(\text{mm})$	191	197	208	180	188	210	196	191	179	208
$w(\text{mm})$	284	285	288	273	280	283	288	271	257	289

The ornithologist plans to use a linear regression model based on these data and interpolate or extrapolate as necessary to estimate the wing length of other female hook-billed kites from their tail length.

- (a) (i) Explain what is meant by extrapolation.

predicting outside the range of data (1)

- (ii) Explain the dangers of extrapolation.

unreliable (1)

The ornithologist attempts to calculate the product moment correlation coefficient, r , and obtains a value of 1.3

- (b) Explain how she should be able to identify that this is incorrect without carrying out any further calculations.

$|r| \leq 1$ so 1.3 is not possible
 $-1 \leq r \leq 1$ (1)

- (c) Use your calculator to find the correct value of the product moment correlation coefficient, r .

$r = 0.7628$ (4dp) (1)

- (d) Stating your hypotheses clearly test, at the 1% significance level, whether or not there is evidence that the product moment correlation coefficient for the population is positive.

(3)

- (e) Explain what your test in part (d) suggests about female hook-billed kites.

d) $H_0: \rho = 0$ $\alpha = 1\% = 0.01$ (1)

$H_1: \rho > 0$ $n = 10$

critical value is 0.7155

$r = 0.7628 > 0.7155$

As $0.7628 > 0.7155$, we have enough evidence to reject H_0 meaning there is positive correlation between tail and wing length.



DO NOT WRITE IN THIS AREA

e) longer tails usually have longer wings

Non-Linear/Exponential Data

... same as Pure Year 1, exponential modelling.

- take logs
- use log laws
- compare coefficients

(Alt. use indices)

eg. $y = ab^x$ or

$$\log y = \log ab^x$$
$$\log y = \log a + x \log b$$

↑
 $\log y$ exponential
 ↓
 x

$$y = ax^b$$

$$\log y = \log ax^b$$
$$\log y = \log a + b \log x$$

↑
 $\log y$ non linear
 ↓
 $\log x$

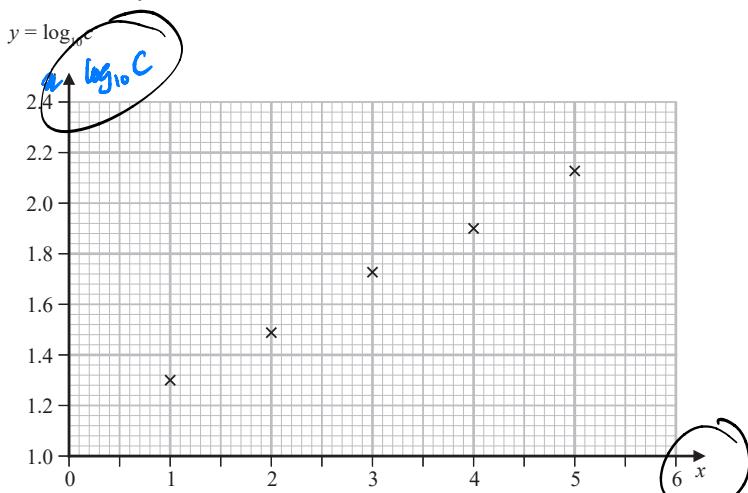
- Evidence to use a different model for points if they don't lie on line
 - ↳ if data has been coded and they appear linear, then non-linear model is suitable.

6. Roberta started a business five years ago.

She believes that the number of customers, c , is growing exponentially.

She produced the graph below by coding her data such that

x = number of years since the business started



Roberta found the regression line for this graph to be $y = 1.10 + 0.204x$

$$\log_{10} C = 1.10 + 0.204x$$

(a) (i) Explain how the graph supports Roberta's belief of exponential growth.

(ii) Find the relationship between the number of customers and number of years since the business started, in the form $c = ab^x$

(5)

Roberta claims that after 6 years she will have more than 200 customers.

(b) Show that Roberta's model supports this claim.

(1)

(c) Comment on the reliability of using Roberta's model in your answer to part (b). You must give a reason for your answer.

(1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

a) graph's points are lying very close to a straight line, and as y is a log, this corresponds to exp. growth

$$\text{ii}) \quad c = ab^x$$

$$\log_{10} C = \log a + x \log b$$

$$\log a = 1.10$$

$$a = 10^{1.10}$$

$$a = 12.6 \quad (\text{3sf})$$

$$\log b = 0.204$$

$$b = 10^{0.204}$$

$$b = 1.60 \quad (\text{3sf})$$

$$c = 12.6 \times 1.60^x$$

$$\text{b) } x = 6$$

$$c = 12.6 \times 1.60^6 = 211.39 = 211 > 200 \text{ customers}$$

c) extrapolation is unreliable,

this is making a prediction outside the given range of observed data.



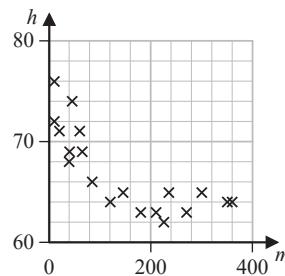
YOUR TURN

A-Level 2022

6. Anna is investigating the relationship between exercise and resting heart rate. She takes a random sample of 19 people in her year at school and records for each person

- their resting heart rate, h beats per minute
- the number of minutes, m , spent exercising each week

Her results are shown on the scatter diagram.



- (a) Interpret the nature of the relationship between h and m

The more mins spent exercising, the lower the resting heart rate ⁽¹⁾
Anna codes the data using the formulae

$$x = \log_{10} m$$

$$y = \log_{10} h$$

The product moment correlation coefficient between x and y is -0.897

- (b) Test whether or not there is significant evidence of a negative correlation between x and y

You should

- state your hypotheses clearly
- use a 5% level of significance
- state the critical value used

(3)

The equation of the line of best fit of y on x is

$$y = -0.05x + 1.92$$

- (c) Use the equation of the line of best fit of y on x to find a model for h on m in the form

$$h = am^k$$

where a and k are constants to be found.

(5)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

b) $H_0 : \rho = 0$

$$H_1 : \rho < 0$$

$n = 19$

$\alpha = 0.05$

crit value is -0.3887

$$r = -0.897 < -0.3887$$

Hence reject H_0 , there is negative correlation between x and y

c) $h = am^k$

$$\log h = \log a m^k$$

$$\log h = \log a + k \log m$$

$$k = -0.05 \quad \log a = 1.92$$

$$a = 10^{1.92}$$

$$a = 83.2 \quad (3sf)$$

$$h = 83.2 \times m^{-0.05}$$



Binomial and Normal Distributions

- Why are we looking at this as one topic?

Binomial, $B(n, p)$

~~(I)~~ Binomial Calculation

eg $P(X \leq 3)^{CD}$ "less than 7" $P(X \leq 6)$

$P(X = 4)^{PD}$ "at most 15" $P(X \leq 15)$

$\underline{\underline{P(X > 5)}} \leftarrow "at least 6"$ $P(X \geq 6)$

~~(II)~~ Binomial within a Binomial

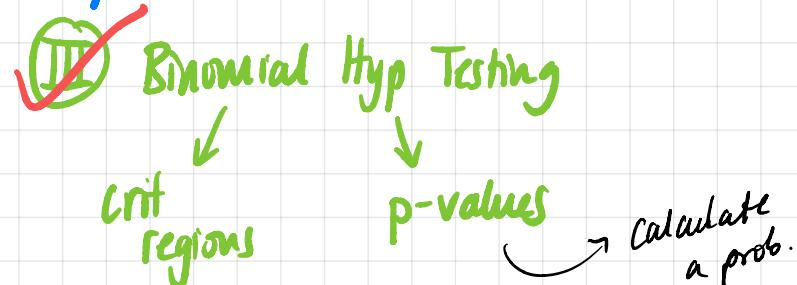
eg. There's a 20% chance I am late to school on each day.

X is being late to school $X \sim B(5, 0.2)$

What is the probability I am late on at least 3 days per week in exactly 2 weeks of a 6 week term?

$$P(X \geq 3) = 0.05792$$

Y is being late at least 3 times per week
 $Y \sim B(6, 0.05792)$ $P(Y=2) = 0.0396$ (4dp)



Normal Distribution $N(\mu, \sigma^2)$

~~(I)~~ Normal Calculations

eg $P(Y > 22.5)$
 $P(13 < Y \leq 29)$

upper = 1000000
lower = 22.5
upper = 29
lower = 13

~~(IV)~~ Hyp. Testing

$\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$

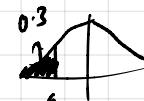
n is the size of the sample

mean of the samples

$\sigma = \frac{\sigma}{\sqrt{n}}$

Inverse normal calculations

eg $P(Y < a) = 0.3$



~~(III)~~ Missing μ, σ , or both

$$Y \sim N(\mu, \sigma^2) \rightarrow Z \sim N(0, 1^2)$$

$$\frac{X - \mu}{\sigma} = Z$$

~~(IV)~~ Binomial \rightarrow Normal

$$B(n, p) \rightarrow N(np, np(1-p))$$

+ continuity corrections

Conditional Probabilities...

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

4. In a game of *Sixes*, each player's turn involves rolling 6 identical dice. If the player gets a six on fewer than 3 dice, the player does not score points.

Assuming that the dice are fair, find the probability that in one turn a player

- (a) (i) gets a six on exactly 3 dice
(ii) gets a six on at least 3 dice

(3)

Ali and four of his friends play *Sixes* together and have one turn each. Two of the 5 players score points.

Ali claims that this suggests the dice are biased towards rolling a six.

- (b) Carry out a suitable test to investigate Ali's claim.

You should

- state your hypotheses clearly
- use a 5% level of significance
- state the p -value for the test

B(I)

(4)

X is the number of sixes rolled

$$X \sim B(6, \frac{1}{6}) \quad \checkmark$$

$$\text{i)} \quad P(X=3) = 0.0536 \quad (\text{4dp})$$

$$\text{ii)} \quad P(X \geq 3) = 1 - P(X \leq 2) \\ \underline{1, 2, 3, 4, 5, 6} \quad = 0.0623 \quad (\text{4dp})$$

B(I)

b) Y is the number of players scoring points

$$Y \sim B(5, 0.0623)$$

probability of scoring points

B(II)

$$H_0: p = 0.0623$$

$$H_1: p > 0.0623$$

Assume H_0 is true, that means $Y \sim B(5, 0.0623)$

$$P(Y \geq 2) = 1 - P(Y \leq 1) = 0.0342 < 0.05$$

B(III)

↳ this is the p -value

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Reject H_0 , Ali's claim is supported,
the dice are biased towards 6.

5. A fast food company has a scratchcard competition. It has ordered scratchcards for the competition and requested that 45% of the scratchcards be winning scratchcards.

A random sample of 20 of the scratchcards is collected from each of 8 of the fast food company's stores.

- (a) Assuming that 45% of the scratchcards are winning scratchcards, calculate the probability that in at least 2 of the 8 stores, 12 or more of the scratchcards are winning scratchcards. (5)

- (b) Write down 2 conditions under which the normal distribution may be used as an approximation to the binomial distribution. (1)

A random sample of 300 of the scratchcards is taken. Assuming that 45% of all the scratchcards are winning scratchcards,

- (c) use a normal approximation to find the probability that at most 122 of these 300 scratchcards are winning scratchcards. (4)

Given that 122 of the 300 scratchcards are winning scratchcards,

- (d) comment on whether or not there is evidence at the 5% significance level that the proportion of the company's scratchcards that are winning scratchcards is different from 45%

a) X is the num. of winning cards

$$X \sim B(20, 0.45)$$

$$P(X \geq 12) = 1 - P(X \leq 11) = 0.1308 \quad (4dp)$$

Y is the num. of stores with 12+ winning cards

$$Y \sim B(8, 0.1308)$$

$$P(Y \geq 2) = 1 - P(Y \leq 1) = 0.2820 \quad (4dp)$$

b) p close to 0.5, n is large B/N (IV)



DO NOT WRITE IN THIS AREA

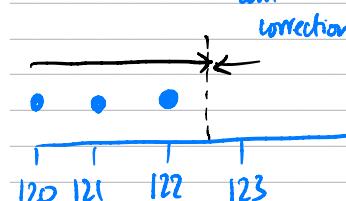
c) $X \sim B(300, 0.45)$ discrete

$$\downarrow np \quad np(1-p) \quad \sigma = \sqrt{np(1-p)}$$

$$S \sim N(135, 74.25) \quad \text{continuous}$$

$$300 \times 0.45 \quad 135 \times 0.55$$

$$\begin{array}{l} \text{binomial} \\ P(X \leq 122) \end{array} \xrightarrow{\text{cont.}} \begin{array}{l} \text{normal} \\ P(S < 122.5) = 0.6734 \quad (4dp) \end{array}$$



$$\begin{aligned} \text{lower} &= -100000 \\ \text{upper} &= 122.5 \\ \mu &= 135 \\ \sigma &= \sqrt{74.25} \end{aligned}$$

d) Two tailed, 2.5% or 0.025 \rightarrow p-value

$$P(X \leq 122) = P(S < 122.5) = 0.6734 > 0.025$$

So there is insufficient evidence to suggest it is diff. from 45%.

5. Kim and Tom are both learning to tune a violin.

Kim's teacher asks her to tune the *A-string* of her violin to the correct frequency in hertz (Hz).

When Kim tunes the *A-string*, its frequency may be modelled by a Normal distribution.

When Kim first starts learning, she tunes the *A-string* with a mean frequency of 443 Hz and a standard deviation of 6 Hz.

The correct frequency for the *A-string* is 440 Hz.

Find the probability that Kim tunes the *A-string*

(a) lower than the correct frequency, (2)

(b) more than 5 Hz away from the correct frequency. (2)

After practising for a month, Kim tunes the *A-string* with a standard deviation of 4.5 Hz.

She claims that the mean frequency when she tunes the *A-string* is now less than 443 Hz.

Kim's teacher asks Kim to tune the *A-string* 20 times and finds that the mean frequency is 442 Hz.

(c) Test at the 5% level of significance whether or not there is evidence to support Kim's claim.

You should state your hypotheses and show your working clearly. (4)

When Tom tunes the *A-string*, its frequency, T Hz, may be modelled by $T \sim N(\mu, \sigma^2)$

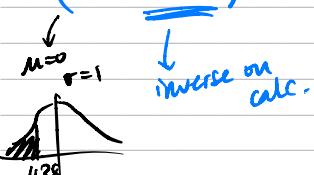
Given that $P(T < 438) = 0.2$ and that $P(T > 445) = 0.1$ unknown

(d) find the value of μ and the value of σ , giving your answers, in Hz, to 1 decimal place. (6)

$$P(T < 438) = 0.2$$

$$\begin{aligned} \text{standardise} \\ \downarrow & 438 \text{ standardise to } -0.8416 \\ & \frac{438 - \mu}{\sigma} = -0.8416 \end{aligned}$$

$$P(z < -0.8416) = 0.2$$



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

$$a) K \sim N(443, 6^2)$$

N(I)

$$a) P(K < 440) = 0.3085 \text{ (4dp)}$$

$$\mu = 443$$

$$\sigma = 6$$

$$\text{upper} = 440$$

$$\text{lower} = -\infty$$

$$\text{Alt. } 1 - P(435 < K < 445) \quad \text{N(I)}$$

$$b) 440$$

$$435$$

$$440$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445$$

$$443$$

$$445</$$

FINAL CHALLENGE

IAL S1 Jun 22

6. A manufacturer fills bottles with oil.

The volume of oil in a bottle, V ml, is normally distributed with $V \sim N(100, 2.5^2)$

(a) Find $P(V > 104.9)$

1

(b) In a pack of 150 bottles, find the expected number of bottles containing more than 104.9 ml

(c) Find the value of v , to 2 decimal places, such that $P(V > v | V < 104.9) = 0.2801$

1

**(2) conditional probabilities
(6)**

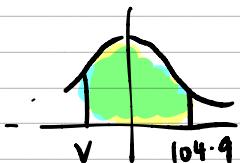
a) $P(V > 104.9) = 0.0250$ (4dp)

b) $0.0250 \times 150 = 3.75$ bottles
(4 bottles)

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

c) $P(V > v | V < 104.9) = \frac{P(V > v \cap V < 104.9)}{P(V < 104.9)}$

$$\rightarrow 1 - 0.0250$$



$$0.2801 = \frac{P(v < V < 104.9)}{0.975}$$

$$0.2801 \times 0.975 = P(v < V < 104.9)$$

$$0.2801 \times 0.975 = P(V < 104.9) - P(V < v)$$

$$0.2801 \times 0.975 = 0.975 - P(V < v)$$

$$P(V < v) = 0.975 - 0.2801 \times 0.975$$

$$P(V < v) = 0.7019025$$



$$v = 101.32 \text{ (2dp)}$$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



YOUR TURN

A-Level 2022

4. A dentist knows from past records that 10% of customers arrive late for their appointment.

A new manager believes that there has been a change in the proportion of customers who arrive late for their appointment.

A random sample of 50 of the dentist's customers is taken.

(a) Write down

- a null hypothesis corresponding to no change in the proportion of customers who arrive late
- an alternative hypothesis corresponding to the manager's belief

(1)

- (b) Using a 5% level of significance, find the critical region for a two-tailed test of the null hypothesis in (a)

You should state the probability of rejection in each tail, which should be less than 0.025

(3)

- (c) Find the actual level of significance of the test based on your critical region from part (b)

(1)

The manager observes that 15 of the 50 customers arrived late for their appointment.

- (d) With reference to part (b), comment on the manager's belief.

(1)

$p =$	0.05	0.10	0.15
$n = 50, x = 0$	0.079	0.0052	0.0003
1	0.0794	0.0338	0.0029
2	0.5405	0.1117	0.0142
3	0.7004	0.2503	0.0460
4	0.8764	0.4312	0.1221
5	0.9622	0.6161	0.2194
6	0.9882	0.7702	0.3613
7	0.9968	0.8779	0.5788
8	0.9992	0.9421	0.6681
9	0.9998	0.9755	0.7911
10	1.0000	0.9906	0.8901
11	1.0000	0.9968	0.9372
12	1.0000	0.9990	0.9699
13	1.0000	0.9997	0.9868
14	1.0000	0.9999	0.9947
15	1.0000	1.0000	0.9981
16	1.0000	1.0000	0.9993
17	1.0000	1.0000	0.9998
18	1.0000	1.0000	0.9999
19	1.0000	1.0000	1.0000
20	1.0000	1.0000	1.0000
21	1.0000	1.0000	1.0000
22	1.0000	1.0000	1.0000
23	1.0000	1.0000	1.0000
24	1.0000	1.0000	1.0000

a) $H_0 : p = 0.1$
 $H_1 : p \neq 0.1$

b) X is the number of customers late

$$X \sim B(50, 0.1)$$

$$P(X=0) = 0.0052 < 0.025$$

$$\text{Crt region } X=0 \quad >0.975$$

$$P(X \geq 10) = 1 - 0.9755 \\ = 0.0245 < 0.025$$

Critical region $X=0, X \geq 10$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

$$\begin{aligned} c) \text{act.sig} &= P(X=0) + P(X \geq 10) \\ &= 0.0052 + 0.0245 \\ &= 0.0297 \\ &= 0.030 \end{aligned}$$

d) 15 is in the critical region ($X \geq 10$).

This means the manager's belief is supported.



YOUR TURN

Set 2

2. A local sports centre has showers with two temperature settings, warm and hot.

On the warm setting, the water temperature may be modelled by a Normal distribution with mean 30°C and standard deviation 2°C

$$W \sim N(30, 2^2)$$

- (a) Using the model, find the probability that the next time the shower is used on the warm setting, the water temperature is

(i) exactly 31°C

$$P(W = 31) = 0$$

(ii) more than 31°C

$$P(W > 31) = 0.3085$$

(2)

Normal calc.

The sports centre manager thinks that a water temperature of more than 33°C is too high for the warm setting.

She tests the water temperature on the warm setting on 5 randomly selected days.

Given that the probability of the water temperature being more than 33°C is 0.0668

- (b) find the probability of the water temperature being more than 33°C

(i) on only the first of these 5 days,

(2)

(ii) on more than 1 of these 5 days.

(3)

binomial calc

On the hot setting, the water temperature may be modelled by a Normal distribution with standard deviation 1.5°C

The probability that the water temperature is more than 42°C is 0.0005

- (c) Find the mean water temperature on this setting, giving your answer to 1 decimal place.

X is the number of days with water more than 33°C (4)

$$X \sim B(5, 0.0668)$$

$$\text{bi)} \quad 0.0668 \times (1 - 0.0668)^4 = \underline{\underline{0.0507}} \quad (\text{4dp})$$

$$\text{ii)} \quad P(X > 1) = P(X \geq 2) = 1 - P(X \leq 1) \\ \text{2,3,4,5} \\ = \underline{\underline{0.0390}} \quad (\text{4dp})$$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

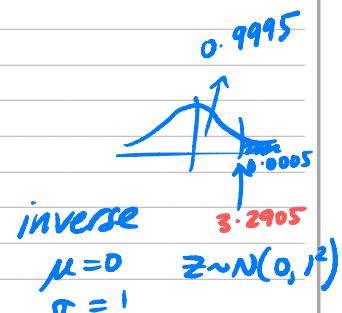
DO NOT WRITE IN THIS AREA

$$\text{c)} \quad H \sim N(\mu, 1.5^2)$$

$$P(H > 42) = 0.0005$$

$$\downarrow \quad \downarrow \quad \frac{H - \mu}{\sigma} = Z$$

$$P(Z > \underline{\underline{3.2905}}) = 0.0005$$



$$\frac{42 - \mu}{1.5} = 3.2905$$

$$42 - \mu = 3.2905 \times 1.5$$

$$42 - 3.2905 \times 1.5 = \mu$$

$$\mu = 37.06425 \\ = \underline{\underline{37.1^\circ\text{C}}}$$



YOUR TURN

Set 2

4. A large number of cyclists take part in a cycling time trial. A random sample of these cyclists are selected and their times, in minutes, are summarised in the following statistics

$$\sum x = 1680 \quad \sum x^2 = 47654.4 \quad n = 60$$

(a) Calculate, for this sample, the value of

(i) the mean time, $\bar{x} = \frac{1680}{60} = 28 \text{ mins}$

(ii) the standard deviation of the times. $\sigma = \sqrt{\frac{47654.4}{60} - 28^2} = 3.2$

Historically, the mean time for cyclists on this time trial has been 27 minutes and 30 seconds. Lucy is watching the time trial and believes that the mean time of cyclists in this time trial is greater than the mean time of cyclists in previous time trials.

The times of cyclists on this time trial are modelled by a Normal distribution with standard deviation 3 minutes.

- (b) Test, at the 5% level of significance, whether or not this sample provides evidence to support Lucy's belief. You should state your hypotheses and show your working clearly.

*normal
hyp.
test*

(5)

Speedy Wheels cycling club entered its 5 fastest riders and 5 beginners to take part in the time trial.

The fastest 20% of the cyclists in the time trial are invited to compete in a race the following week.

- (c) (i) Explain, with specific reference to the parameter p , why the distribution $B(10, 0.2)$ might not be reasonable to model the number of these *Speedy Wheels* cycling club members who are invited to compete in the race.

*binomial,
interpreting.*

(2)

- (ii) Suggest how to improve the model for the number of these *Speedy Wheels* cycling club members invited to compete in the race.

(1)

$$H_0: \mu = 27.5$$

$$n = 60$$

$$H_1: \mu > 27.5$$

Assume H_0 is true, $X \sim N(27.5, 3^2)$
then $\bar{X} \sim N(27.5, \frac{3^2}{60})$

$$P(\bar{X} > 28) = 0.0984 > 0.05$$

No evidence to support Lucy's belief, accept H_0 .

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

ci) 0.2 unlikely to be the same probability for the beginners and fastest cyclists.

The probabilities should differ for the groups

(i) Do separate binomial distribution with different p values

