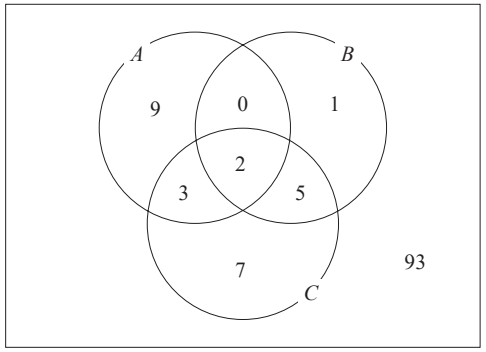
Leave blank

**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

Try this Q whilst we wait for everyone to arrive!




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

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

**(1)**

(b) Find $P(A \cap B \cap C')$

**(1)**

(c) Find $P(A \cup B \cup C')$

**(2)**

(d) Find the probability that the shoe has at most one type of defect.

**(2)**

(e) Given the selected shoe has at most one type of defect, find the probability it has defective stitching.

**(2)**

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

} Some aspect of LDS

## Possible topics (not covering in full)

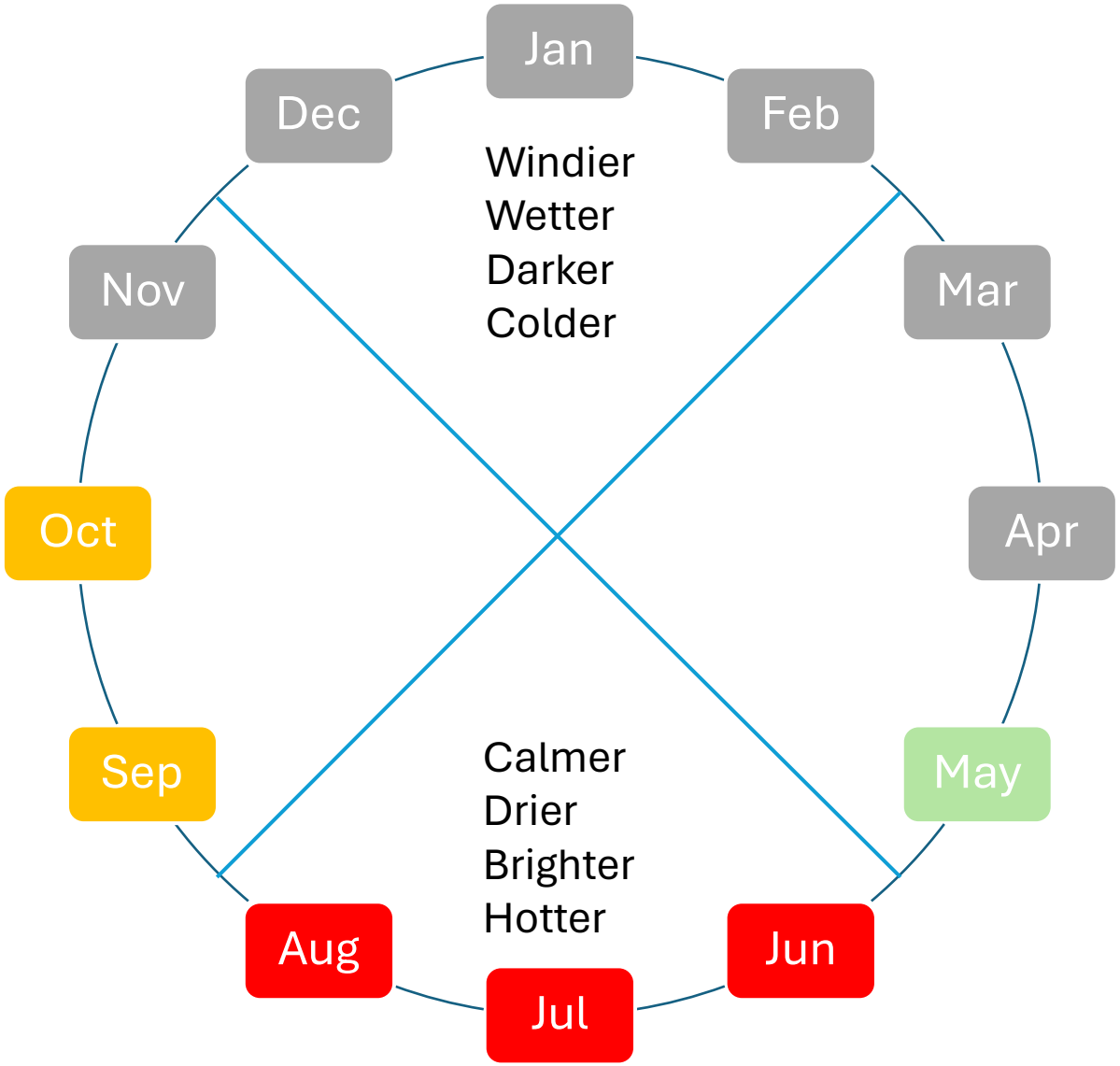
- → histograms
- → probability tables/distributions
- → cumulative freq/box plots
- → tree diagrams

# 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:**

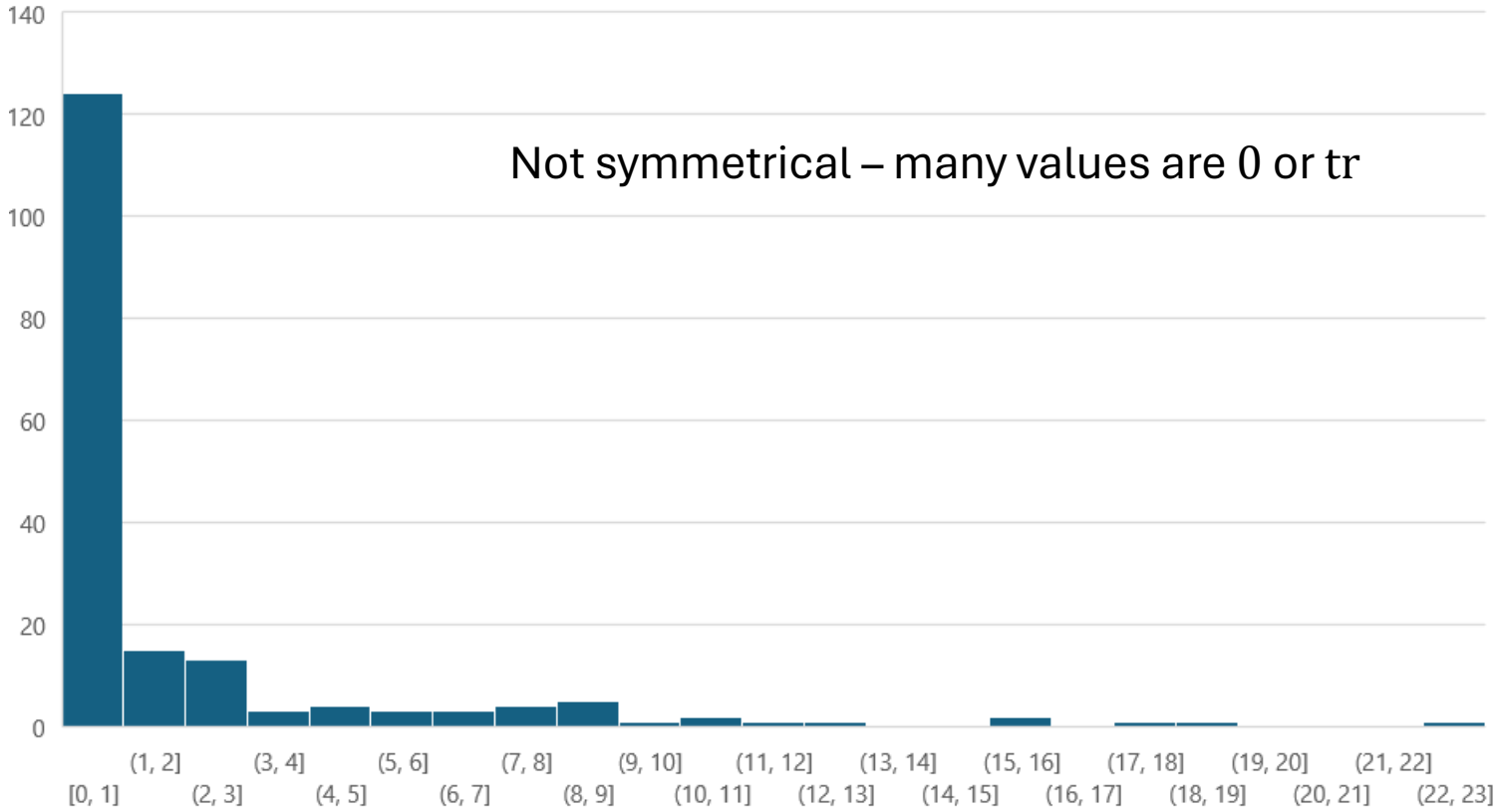
If the temperature increases…
… the amount of sunshine _____
… the amount of rainfall _____
… the windspeed _____

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** | *How hot it is* | °C | ~5°C to ~24°C | 12.0°C<br>14.8°C | Warmer in summer |
| Daily total **rainfall** | *How much it rained* | mm | 0 to ~20mm | 0 mm<br>10.4 mm<br>tr | Not symmetrical – many 0 and tr values |
| Daily total **sunshine** | *How many hours of sunshine* | hours | 0 to ~14 hours | 3.3 hrs<br>10.4 hrs | More sunshine in the summer |
| **Cloud cover** | *How much of the sky is covered in clouds* | oktas | 0 to 8 | 3<br>4<br>0 | Integers, measuring what fraction of the sky is covered |
| **Humidity** | *How much water vapour is in the air – above 95% is associated with fog* | % | ~70% to 100% | 100%<br>77%<br>95% | Integers |
| Daily mean **visibility** | *How far you can see* | Dm<br>1 Dm = 10 m<br>'decametre' | ~200 Dm to ~4000 Dm | 1300 Dm<br>2200 Dm<br>3100 Dm | Rounded to nearest 100 |
| Daily mean **pressure** | *How much the atmosphere is pushing down* | hPa<br>'hectopascals' | ~990 hPa to ~1040 hPa | 1017 hPa<br>1006 hPa<br>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** | *How windy it is* | kn (knots) | ~3 kn to ~10 kn | 4 kn<br>11 kn | Integers only |
| **Windspeed, Beaufort conversion** | | Qualitative | Light to Moderate | Light<br>Moderate<br>Fresh<br>Strong | Qualitative. Most days are 'light' |
| Daily **maximum gust** | *The strongest gust of wind that day* | kn (knots) | ~8 kn to ~50 kn | 17 kn<br>25 kn | Integers only |
| **Wind/gust direction (bearings)** | *Which direction the wind is blowing **from*** | ° | 10° to 360° | 240°<br>70° | Multiples of 10 only |
| **Wind/gust direction (cardinal)** | *Which direction the wind is blowing **from*** | Compass direction | – | N<br>SW<br>ENE | Describes where the wind is blowing **from**, not to |


**Jacksonville, Florida, USA**
Hot summers
Hurricanes in October (but mean windspeed still relatively low in our data)
**Beijing, China**
Hotter, wetter summers
Colder winters
**Perth, Australia**
Flipped seasons
Beijing
Perth
United Kingdom
Ireland
Denmark
Germany
Poland
Belarus
Ukraine
France
Austria
Romania
Italy
Spain
Greece
Türkiye
Kazakhstan
Mongolia
Kyrgyzstan
Turkmenistan
Syria
Iraq
Iran
Afghanistan
Pakistan
Tunisia
Algeria
Libya
Egypt
Western Sahara
Saudi Arabia
Oman
Yemen
Mauritania
Mali
Niger
Chad
Sudan
Burkina Faso
Guinea
Nigeria
Ghana
Ethiopia
Somalia
Kenya
Gabon
DRC
Tanzania
Angola
Zambia
Mozambique
Namibia
Zimbabwe
Botswana
Madagascar
South Africa
Cuba
Puerto Rico
Caribbean Sea
Nicaragua
Colombia
Guyana
Suriname
Ecuador
Peru
Brazil
Bolivia
Paraguay
Chile
Uruguay
Argentina
South Atlantic Ocean
Indian Ocean
Arabian Sea
Bay of Bengal
Laccadive Sea
Gulf of Aden
Gulf of Guinea
Gulf of Thailand
Sea of Japan
Sea of Okhotsk
Japan
South Korea
China Sea
Philippine Sea
Thailand
Vietnam
Philippines
Malaysia
Indonesia
Banda Sea
Arafura Sea
Papua New Guinea
Australia
Great Australian Bight

# International Measurements

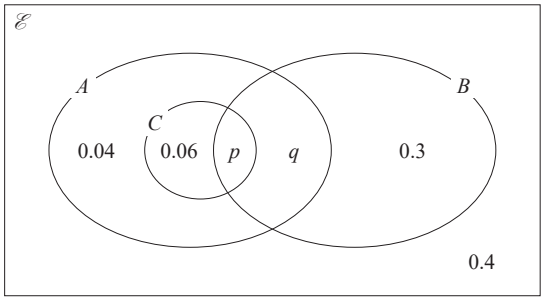
| Measurement | | Units | Details |
|---|---|---|---|
| Daily mean **temperature** | How hot it is | ℃ | Warmer in summer… but Perth is colder |
| Daily total **rainfall** | How much it rained | mm | Beijing is rainy in the summer |
| Daily mean **pressure** | How much the atmosphere is pushing down | hPa<br>'hectopascals' | |
| Daily mean **windspeed** | *How windy it is* | 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)
- Mutually exclusive – cannot happen at same time, circles won't overlap at all

Leave blank

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




(a) Find $P(B)$

**(1)**

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

**(2)**

Given that $P(C \mid B) = P(C)$

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

**(3)**

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 $D$

**(1)**

DO NOT WRITE IN THIS AREA

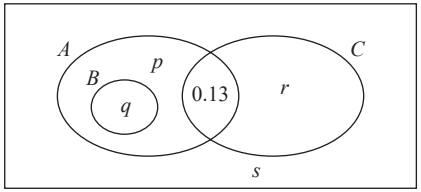
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

**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)**

# Correlation and Regression

- Describe / Interpret correlation
  - ↳ what it means in context.
  - + or −

- Hyp Testing – 3 easy marks!

$H_0: \rho = 0$ or $H_0: \rho = 0$

$H_1: \rho < 0$ or $\rho > 0$ $\quad$ $H_1: \rho \neq 0$

- Compare your observed value from the sample with the critical value
  - ... if more "extreme", there's evidence for correlation, so reject $H_0$!

## Critical Values for Correlation Coefficients

These tables concern tests of the hypothesis that a population correlation coefficient $\rho$ 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.

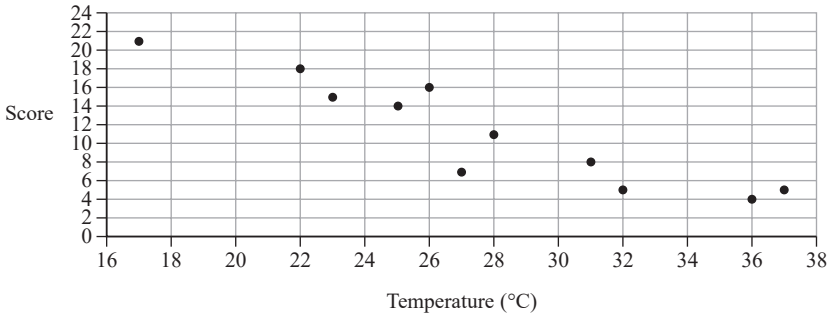
| Product Moment Coefficient | | | | | | Spearman's Coefficient | | |
|---|---|---|---|---|---|---|---|---|
| Level | | | | | Sample size, $n$ | Level | | |
| 0.10 | 0.05 | 0.025 | 0.01 | 0.005 | | 0.05 | 0.025 | 0.01 |
| 0.8000 | 0.9000 | 0.9500 | 0.9800 | 0.9900 | 4 | 1.0000 | – | – |
| 0.6870 | 0.8054 | 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, $x$ °C.

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

The results are shown in the scatter diagram below.




(a) Use the scatter diagram to find

(i) the median score

(ii) the range of the scores.

**(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)**

**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.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $t$ (mm) | 191 | 197 | 208 | 180 | 188 | 210 | 196 | 191 | 179 | 208 |
| $w$ (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.

**(1)**

(ii) Explain the dangers of extrapolation.

**(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.

**(1)**

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

**(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.

**(1)**

# 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 $y = ax^b$

- 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!

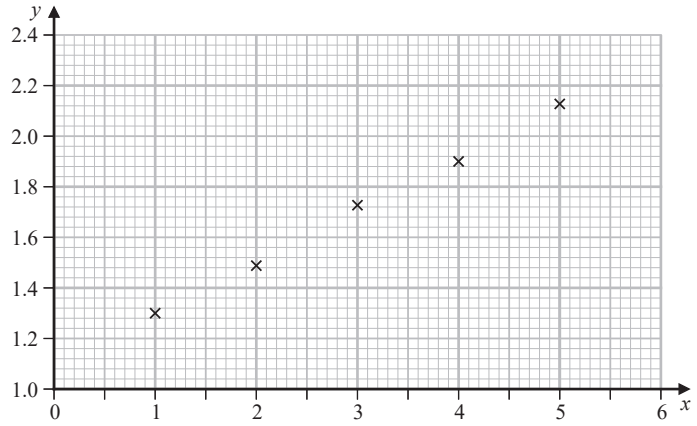
**6.** Roberta started a business five years ago.

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

$y = \log_{10} c$




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

(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)**

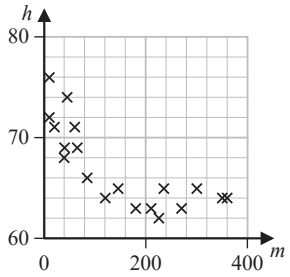
(c) Comment on the reliability of using Roberta's model in your answer to part (b). You must give a reason for your answer.

**(1)**

**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 $m$

**(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)**

# Binomial and Normal Distributions

- Why are we looking at this as one topic?

## Binomial, $B(n,p)$

(I) Binomial Calculation

eg $P(X \leq 3)$ "less than 7"

$P(X = 4)$ "at most 15"

$P(X > 5)$ "at least 6"

(II) Binomial within a Binomial

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

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

(III) Binomial Hyp Testing

↙ crit regions ↘ p-values

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

(I) Normal Calculations

eg $P(Y > 22.5)$

$P(13 < Y \leq 29)$

(II) Inverse normal calculations

eg $P(Y < a) = 0.3$

(closely linked)

(III) Missing $\mu$, $\sigma$, or both (→ sim equations)

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

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

(IV) Hyp. Testing

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

(V) Binomial → Normal

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

+ continuity corrections

(VI) 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

**(4)**

**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%

**(1)**

**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 $P(T > 445) = 0.1$

(d) find the value of $\mu$ and the value of $\sigma$, giving your answers, in Hz, to 1 decimal place.

**(6)**

FINAL CHALLENGE



**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

**(2)**

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

**(6)**

**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)**

**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

(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

(ii) more than 31°C

**(2)**

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)**

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.

**(4)**

**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,

**(1)**

(ii) the standard deviation of the times.

**(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.

**(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.

**(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)**