# A Level Statistics

## Effective Strategies for Exam Preparation

Dr. Ali Everett

Pearson Edexcel

# Introduction

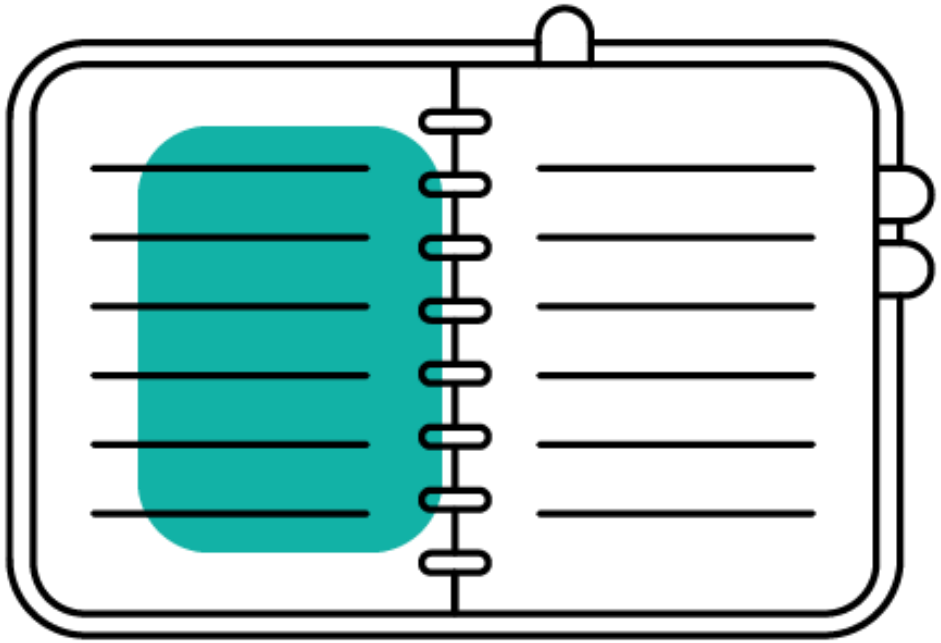
# Session Leader – Dr. Ali Everett

- Ph.D. in Pure Mathematics
- Over 10 years of experience teaching A Level Statistics
- Author of the Pearson GCE A Level Statistics (9ST0) Scheme of Work
- Pearson Credible Specialist for GCE A Level Statistics (2020-)
    - Author of practice papers, shadow papers and topic tests
- Examiner for GCE A Level Statistics (9ST0/03)
- Course leader for GCE A Level Statistics at Hereford Sixth Form College

# Aims and Objectives

In this session we are going to look at:

- Assessment objectives for A Level Statistics
- General strategies for answering questions
- Strategies for subject specific questions:
    - Technology
    - Designing Experiments / The Statistical Enquiry Cycle
    - Data Interpretation
    - Hypothesis Testing
    - Effect Size



# Assessment Objectives in A Level Statistics

# Assessment Objectives (AO)

3 Assessment objectives: AO1, AO2 and AO3

"Knowledge and Understanding"

**AO1 (55% of assessment)**
Demonstrate knowledge and understanding, using appropriate terminology and notation of standard statistical techniques used:

- To collect and represent data
- To calculate summary statistics and probabilities
- In relation to hypothesis tests and inference

"Interpretation"

**AO2 (25% of assessment)**
Interpret statistical information and results in context and reason statistically to make predictions, construct arguments, make decisions and draw conclusions.

"Critique"

**AO3 (20% of assessment)**
Critically assess the reliability and validity of statistical methodologies and the conclusions drawn through the application of the statistical enquiry cycle.

Weightings of assessment objective may be ±2%

Source: *GCE Subject Level Conditions and Requirements for Statistics (Ofqual, May 2016)*

# AO1 – “Knowledge and Understanding” (55%)

AO1 is divided into three strands:

**AO1.1**
Demonstrate knowledge and understanding, using appropriate terminology and notation, of standard statistical techniques used to collect and represent data

**AO1.2**
Demonstrate knowledge and understanding, using appropriate terminology and notation, of standard statistical techniques used to calculate summary statistics and probabilities

**AO1.3**
Demonstrate knowledge and understanding, using appropriate terminology and notation, of standard statistical techniques used in relation to hypotheses and inference

Anything relating to data collection, organisation and representation (e.g. sampling, technology, calculations from diagrams etc.)

The majority of what we would normally consider “mathematics”: calculation of probabilities, numerical measures, using probability distributions etc.

Anything relating to a hypothesis test or a confidence interval
(e.g. the calculation of a mean or a probability by itself is AO1.2 except if it is used as part of a process to find a confidence interval or a p-value which then makes it AO1.3)

Source: *GCE Subject Level Guidance for Statistics (Ofqual, May 2016)*

# AO2 – “Interpretation” (25%)

AO2 is 1 strand consisting of 2 elements

**AO2.1a**
Interpret statistical information and results in context

*Relating information and/or results to the context provided in the question/task*

Example AO2.1a questions:

- Interpreting the conclusions of a hypothesis test
- Interpretation of *a* and *b* in the regression line
- Selecting the appropriate probability model from a context

**AO2.1b**
Reason statistically to make predictions, construct arguments, make decisions and draw conclusions

*Forming a statistical judgement using appropriate statistical terminology*

Example AO2.1b questions:

- Deciding whether to reject $H_0$ or not with numerical evidence
- Determining if there are outliers in a dataset (with numerical evidence)
- Selection of blocking factors in an experiment

Source: *GCE Subject Level Guidance for Statistics (Ofqual, May 2016)*

# AO3 – “Critique” (20%)

AO3 is 1 strand consisting of 2 elements

**AO3.1a**
Critically assess the reliability and validity of statistical methodologies

**AO3.1b**
Critically assess the reliability and validity of the conclusions drawn through the application of the statistical enquiry cycle

Source: *GCE Subject Level Guidance for Statistics (Ofqual, May 2016)*

# AO3 – “Critique” (20%)

AO3 is 1 strand consisting of 2 elements

**AO3.1a**
**Critically assess** the reliability and validity of statistical methodologies

“Make a judgement relating to the strengths and weaknesses”

**AO3.1b**
**Critically assess** the reliability and validity of the conclusions drawn through the application of the statistical enquiry cycle

Source: *GCE Subject Level Guidance for Statistics (Ofqual, May 2016)*

# AO3 – “Critique” (20%)

AO3 is 1 strand consisting of 2 elements

**AO3.1a**
Critically assess the **reliability** and validity of statistical methodologies

The extent to which an experiment produces the same/compatible results on repeated trials

**AO3.1b**
Critically assess the **reliability** and validity of the conclusions drawn through the application of the statistical enquiry cycle

Source: *GCE Subject Level Guidance for Statistics (Ofqual, May 2016)*

# AO3 – “Critique” (20%)

AO3 is 1 strand consisting of 2 elements

**AO3.1a**
Critically assess the reliability and **validity** of statistical methodologies

The extent to which a procedure does what it is designed to do

**AO3.1b**
Critically assess the reliability and **validity** of the conclusions drawn through the application of the statistical enquiry cycle

Source: *GCE Subject Level Guidance for Statistics (Ofqual, May 2016)*

# AO3 – “Critique” (20%)

AO3 is 1 strand consisting of 2 elements

**AO3.1a**
Critically assess the reliability and validity of statistical methodologies

Example AO3.1a questions:

- Determining if a probability distribution is a suitable model from the context
- Identifying how the assumptions of a hypothesis test apply (or do not apply) to the given context
- Justifying the use of a particular method

**AO3.1b**
Critically assess the reliability and validity of the conclusions drawn through the application of the statistical enquiry cycle

Example AO3.1b questions

- “Ali concludes that…
  Comment on the validity of his conclusion”

Source: *GCE Subject Level Guidance for Statistics (Ofqual, May 2016)*

# General strategies for answering questions

Bullet Points

# Bullet Points

As stated in every examiners reports for the last few years:

- Use bullet points, each written in clear, specific and concise sentences

Candidates are not expected to write an essay.

It is sufficient (and highly encouraged) to use bullet points.

1 bullet point per point the candidate is trying to convey

**Example 1:** *[9ST0/01 (2022) Question 5c]*

Give **two** reasons in context why the assumptions that Tom made may **not** be justified.

**Ali's Model Answer**

- Not all tickets for each film will be sold if the film is not popular
- Family groups with multiple children may watch the film so concession tickets sold may not be independent of each other

# Bullet Points

As stated in every examiners reports for the last few years:

- Use bullet points, each written in clear, specific and concise sentences

Candidates are not expected to write an essay.

It is sufficient (and highly encouraged) to use bullet points.

1 bullet point per point the candidate is trying to convey

**Example 2:** *[9ST0/03 (2022) Question 1e]*

Highlight **one** advantage and **one** disadvantage of Giovanni's choice of box and whisker plots to display his results

**Ali's Model Answer**

- Advantage: It is very easy to compare medians and the interquartile range across the treatment groups and time periods
- Disadvantage: Information about the sample sizes are not displayed


Make *n* points…

# “Make *n* points…”

Questions often ask candidates to list a set number of points in questions:

“Make **four** distinct comments…”
“Your answer should include **three** different statements”

Candidates should make the required number of statements **and no more**.

This is to discourage the “scatter gun” approach of writing down as many points as possible to hit the target

If a candidate:

- Lists the **number of required points or fewer:**
  *The candidate will be awarded one mark per correct point given. Any incorrect points will score zero.*
- Lists **more than the required number** of points:
  - **If all points are correct OR some points are correct and some are not relevant (but not strictly incorrect)**
    The candidate will be awarded one mark per correct point given up to a maximum of the mark allocation for that question
  - **If there is at least one incorrect point** that candidate may be awarded one mark per correct point given up to a maximum of **one fewer** than the mark allocation for that question

# Example

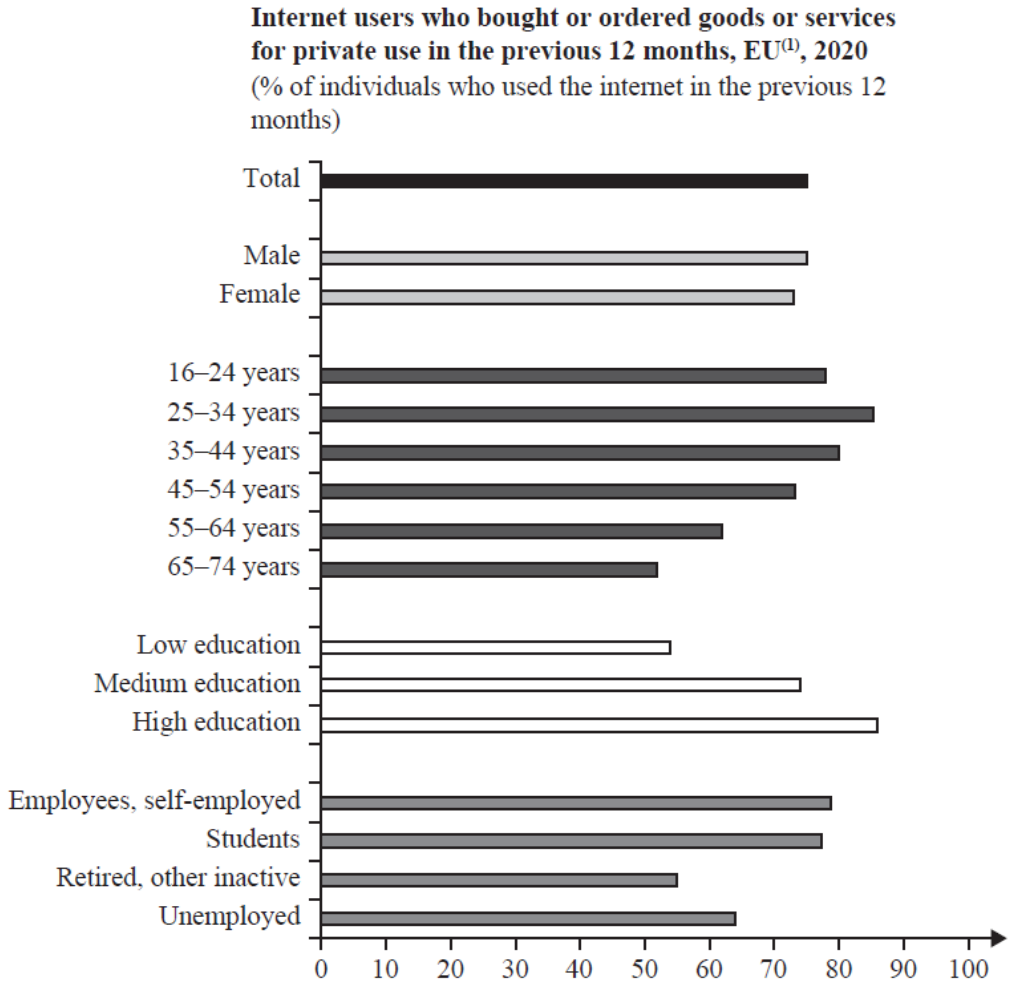



**Figure 4**

*[9ST0/03 (2023) Question 2d]*

Make **three** comments on the data in **Figure 4.** **(3)**

**Ali's Model Answer: 3 correct points**

- The proportion of males who bought or ordered goods for private use in the previous 12 months is slightly higher than that for females
- The proportion of people aged 25-34 years old who bought or ordered goods for private use in the previous 12 months was the highest out of the age groups
- People with high education are more likely to buy or order goods for private use than those with low education

E1E1E1

# Example

**Internet users who bought or ordered goods or services for private use in the previous 12 months, EU[1], 2020**
(% of individuals who used the internet in the previous 12 months)

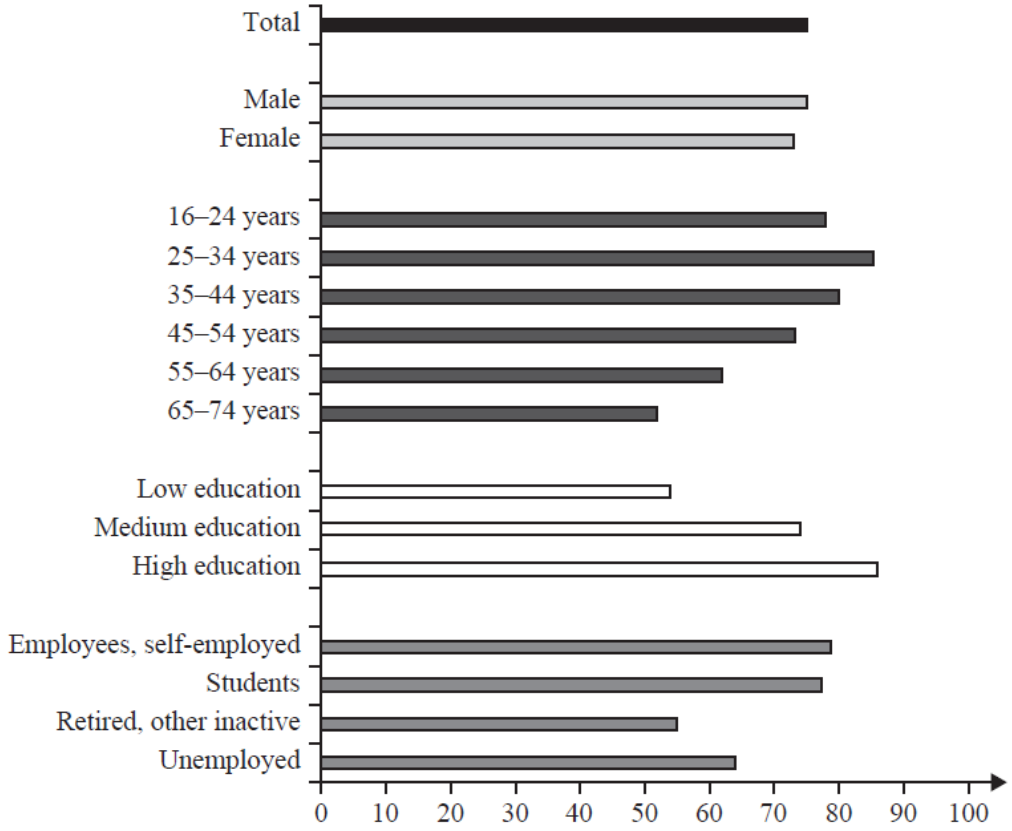



**Figure 4**

*[9ST0/03 (2023) Question 2d]*

Make **three** comments on the data in **Figure 4.** **(3)**

**Answer 2: 4 points but 1 not relevant**

- The proportion of males who bought or ordered goods for private use in the previous 12 months is slightly higher than that for females
- The proportion of people aged 25-34 years old who bought or ordered goods for private use in the previous 12 months was the highest out of the age groups
- People with high education are more likely to buy or order goods for private use than those with low education
- People with high education may have more money so they are more likely to use the internet to shop

Not relevant… but may not be incorrect

E1E1E1

# Example

**Internet users who bought or ordered goods or services for private use in the previous 12 months, EU[1], 2020**
(% of individuals who used the internet in the previous 12 months)

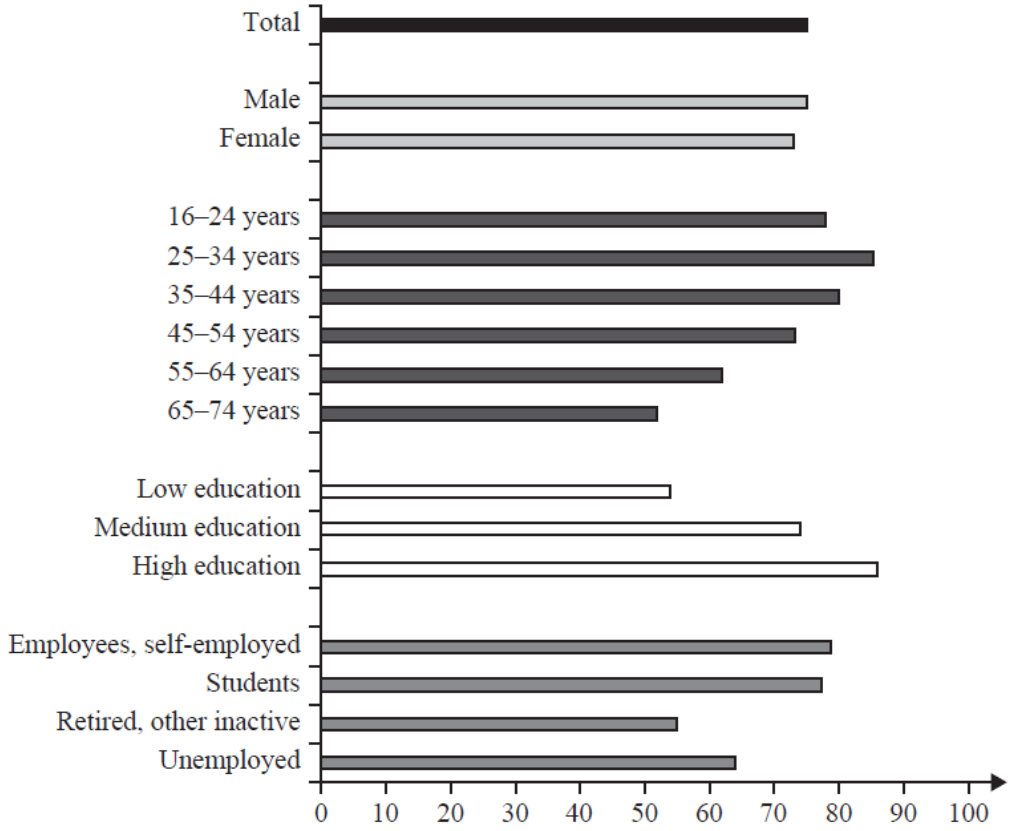



**Figure 4**

[9ST0/03 (2023) Question 2d]

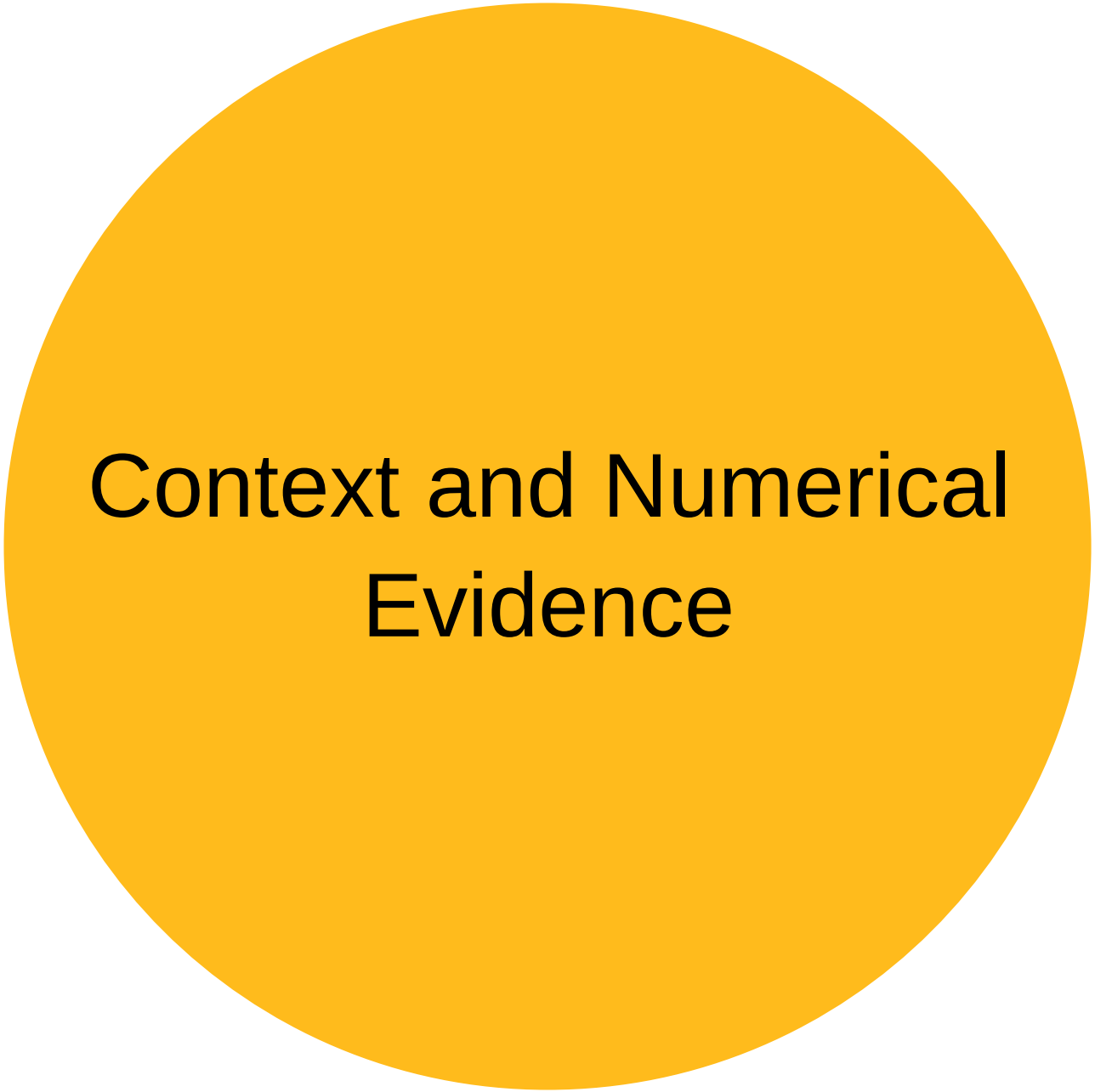
Make **three** comments on the data in **Figure 4.** **(3)**

**Answer 3: 4 points but 1 is incorrect**

- The proportion of males who bought or ordered goods for private use in the previous 12 months is slightly higher than that for females
- The proportion of people aged 25-34 years old who bought or ordered goods for private use in the previous 12 months was the highest out of the age groups
- People with high education are more likely to buy or order goods for private use than those with low education
- A larger number of students buy from the internet than unemployed people

"number" is incorrect

E1E1EO


Context and Numerical Evidence

# Context and Numerical Evidence

As stated in previous examiners reports:

- Give all explanations in context

Examples of answers that may NOT be awarded marks:

“The sample is random”

“The population is normally distributed”

“The second mean is higher than the first”

“The trials must be independent of each other”

“Events must occur at a constant average rate”

“There is significant evidence to suggest a difference in medians”

“As $x$ goes up by 1, $y$ goes up by 31.5”

**General guidance:**

Could the answer be “copy and pasted” into another question and still make sense?

If so… there is **not enough context** in the answer

e.g. if the question is about earthquakes…

…the answer should also be about earthquakes

# Context and Numerical Evidence

As stated in previous examiners reports:

- Give all explanations in context

Answers that **would** be awarded marks:

“The sample **of cows** is random”

“The population **of the heights of gymnasts** is normally distributed”

“The mean **age of Welsh people** is higher than **that for English people**”

“The **patients** must be independent of each other”

“**Sneezes** must occur at a constant average rate”

“There is significant evidence to suggest a difference in median **blood pressure between adults and children**”

“As **the time** goes up by 1 **month**, the **income** goes up by **£**31.5 **thousand**”

**General guidance:**

Could the answer be “copy and pasted” into another question and still make sense?

If so… there is **not enough context** in the answer

e.g. if the question is about earthquakes…

…the answer should also be about earthquakes

# Context and Numerical Evidence

All arguments should be supported with numerical evidence where appropriate

- If numbers are already calculated, candidates are expected to quote them again as supporting evidence
- If supporting evidence **can** be calculated from the numbers given, it is a good idea to do so!

**Tips for interpretation questions**

- Say what you see
- Say what it means
- Back it up with numerical evidence
- Keep it in context

# Example

*[9ST0/03 (2021) Question 7]*

The Northern Ireland Multiple Deprivation Measure (NIMDM) is used by the Northern Ireland Government to compare levels of social deprivation in different geographical areas of Northern Ireland.

The NIMDM combines ratings of several factors affecting deprivation including Education, Income and Employment.

Eoin is a sociologist. He is comparing rural (countryside) and urban (city) areas of Northern Ireland. He is interested in the associations between Educational deprivation and other kinds of deprivation.

He calculates the measures in **Figure 8** from the NIMDM ratings for 2017.

**A lower NIMDM value indicates greater deprivation.**

| | | Rural | Urban |
|---|---|---|---|
| **Median** measure of deprivation for Education | | 554 | 328.5 |
| **IQR** of the measure of deprivation for Education | | 252 | 469.25 |
| **Spearman's rank correlation coefficient** (SRCC) between the measures of deprivation for | Education and Income | 0.490 | 0.745 |
| | Education and Employment | 0.779 | 0.863 |

(Data source: https://www.nisra.gov.uk)

**Ali's Model Answer**

Say what you see

Back it up with numbers

- The median NIMDM for rural areas is 554 which is higher than that for urban areas at 328.5.
  This means that, on average, rural areas have greater social deprivation than urban areas

Keep it in context

Say what it means

- The IQR for rural areas is 252 which is lower than that for urban areas at 469.25.
  This means that the social deprivation in urban areas is more varied than for rural areas
- The SRCC for Education and Income for rural areas is 0.490 which is lower than that for urban areas at 0.745.
  This means the association between education and income is stronger in urban areas than in rural areas

# Defining Variables and Subscripts

# Defining variables

As stated in examiners reports

- Make sure their answers and **any symbols are clearly communicated** to the examiner

If a symbol is used to represent a variable and is not defined in the question, **it must be defined by the candidate.**

e.g.

“Let *X* be the time taken to run 100 metres”
“Let $A_1$ be the test score in group 1”
“Let $D = X - Y$ where *X* is the recovery time of a patient in the control group and *Y* is that for the experimental group”

- Using *X* in a question with one dependent variable without definition may result in a candidate not being awarded full marks
- Using *X* and *Y* in a bivariate question, or in a question with a dependent variable under two factor levels (e.g. 2 sample tests), with no indication of which letter corresponds with which may result in a candidate not being awarded full marks
  **Assuming that *X* and *Y* (or $X_1$ and $X_2$) are in the same order as the groups in question is <u>not sufficient</u>**
- This also applies to ANOVA questions: assuming e.g. $\mu_1$, $\mu_2$ etc. correspond to the means of the groups as presented in a table is **not sufficient** to warrant full marks. Subscripts and variables must be clearly defined.

# Symbols and Greek letters

This has been released as part of the “exemplar” presentations which can be found on the Maths Emporium:

- Greek letters and symbols must be **clearly identifiable**.

**Greek letters**

In some hypothesis tests we see $\mu$, $\eta$, $\rho$, or $\pi$.
It is important that candidates use the correct symbol.

If the examiner (and principal examiners) cannot clearly identify the symbol (or believe the symbol could resemble two possible symbols) then the candidate may not be awarded the mark for the hypotheses.

**Mathematical Symbols**

Some candidates use circles for decimal points. These circles have grown to huge proportions for some candidates where 1.34 looks like 1o34 (you see the problem)

This goes for:

z and 2

S and 5

> and 7

n and $\eta$

U, ∪, $\mu$ etc.

# Mark Scheme Policies

# Some final points about awarding marks

There are some differences to the general awarding of marks between A Level Statistics and A Level Mathematics.

- **M marks are, by default, “possibly implied”.**

This means that if the correct answer is obtained with no supporting working, then the method mark may be awarded as well.

This is because calculators can obtain many probabilities and values at the touch of a button.

The exceptions are if the questions asks (e.g.) “**Show that** the mean is 5.56”

In situations like this, the candidate is expected to “show that”.

- **When presented with a choice of answers:**

All attempts are marked separately, and the **average mark (rounded down)** is awarded.

# Subject Specific Questions


Technology and Software

# Technology and Software

In the real-world:

Very rarely (if ever) do people analyse data by hand.

Data are always analysed using software packages (usually statistics specific software)

In A Level Statistics, we look at only two types of software:

- Spreadsheets
- Databases

**<u>Purposes</u>**

**Spreadsheets**
This software organises data into a grid of “cells” referenced by a letter (usually column) and a number (usually row)
e.g. Cell D6 is in column D, row 6

**Advantages**
Spreadsheets are generally intuitive to use and data can be presented in a way which is easy to read

**Disadvantages**
All data is accessed at once (may not be efficient) and there is a software limit to how much data can be stored (In Microsoft Excel 365, there are 17179869184 cells available which may seem like a lot but, in the age of big data, is not actually enough!)

# Technology and Software

In the real-world:

Very rarely (if ever) do people analyse data by hand.

Data are always analysed using software packages (usually statistics specific software)

In A Level Statistics, we look at only two types of software:

- Spreadsheets
- Databases

**<u>Purposes</u>**

**Databases**
This software organises data into “tables” which consist of “fields” (usually columns) and “records” (usually rows). The tables often are themed and there are often fields of data which are duplicated across different tables to link them

**Advantages**
Databases compartmentalise the data so only data of interest can be accessed at any one time.
The limit for how much data can be stored is also raised considerably

**Disadvantages**
Databases are less user-friendly and require more knowledge to use effectively

# Spreadsheet questions

Candidates need to be aware of:

- General spreadsheet terminology
- How to carry out basic operations on a spreadsheet
- Knowledge of some basic spreadsheet formulae

# Spreadsheet questions

Candidates need to be aware of:

- **General spreadsheet terminology**
- How to carry out basic operations on a spreadsheet
- Knowledge of some basic spreadsheet formulae

**Terminology**

**Cell**
An entry of data on the spreadsheet which can be referenced by a letter and a number

**Sort**
When data is ordered either alphabetically (or reverse alphabetically), or numerically (ascending or descending)

**Filter**
When only data which satisfy a given condition are displayed

# Spreadsheet questions

Candidates need to be aware of:

- General spreadsheet terminology
- **How to carry out basic operations on a spreadsheet**
- Knowledge of some basic spreadsheet formulae

**<u>Basic Operations</u>**

**Cell arithmetic**

Cells can be added, subtracted, multiplied or divided using cell references  e.g. A3+A6

**Sort and filter**

When data need to be ordered or only data satisfying a condition need to be displayed

**Copy and Paste**

Data can be copied from one spreadsheet and pasted into another.  This is especially useful if data need to be merged across different spreadsheets

**Formulae**

Most spreadsheets use a standard set of formulae which can process the data easily

# Spreadsheet questions

Candidates need to be aware of:

- General spreadsheet terminology
- How to carry out basic operations on a spreadsheet
- **Knowledge of some basic spreadsheet formulae**

**Basic Formulae**

**=SUM([range])**
This calculates the sum of the values contained in the range provided

**=AVERAGE([range])**
This calculates the mean of the values contained in the range provided

**=COUNT([range])**

This counts the number of cells in the range which contain a number

**=COUNTIF([range],[condition])**

This counts the number of cells in the range which satisfy the given condition

**=VAR.S([range]) and =STDEV.S([range])**

This calculates the sample variance and sample standard deviation of the values contained in the range provided

# Spreadsheet Example [9ST0/02 (2024) Q7]

7 In January 2023, Michael was working at a financial advisory firm in the North West of England.

He was interested in investigating whether people living in Manchester and Liverpool have different levels of **savings**.

He decided to use his company records for his investigation. An extract of these records is shown in **Figure 7**

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | **Customer_ref** | **Surname** | **First** | **Branch** | **Type_of_account** | **Asset_value (£)** | **Savings_account (£)** |
| 2 | 1022564 | Kelly | Nigel | Preston | Flex | 312 458 | 18 210 |
| 3 | 1022565 | Littler | Mark | Manchester | Fixed | 256 874 | 34 116 |
| 4 | 1022566 | Leng | Moira | Liverpool | High risk | 303 451 | 26 415 |
| 5 | 1022567 | Tang | Peter | Chester | Fixed | 512 744 | 22 457 |
| 6 | 1022568 | Prest | Heather | Manchester | Low risk | 211 110 | 10 223 |

**Figure 7**

(a) Explain how Michael could use a spreadsheet function to help **select** appropriate data for this investigation.

**(2)**

## Ali's Model Answer

- Filter Column D to equal "Manchester" or "Liverpool"
- Use the data into Column G

| Question | Scheme | Marks | AO | Notes |
|---|---|---|---|---|
| **7(a)** | **FILTER** | B1 | 1.1 | |
| | 'Branch' to 'Manchester'<br>'Branch' to 'Liverpool' | B1 | 1.1 | oe<br>Accept 'Column D' in replace of 'Branch' |

# Spreadsheet Example [8ST0/01 (2019) Q5]

Deshandra is writing a report on the consumption of hot drinks around the world. She believes that countries where more coffee is consumed are likely to consume less tea.

She wants to test this belief by carrying out a hypothesis test.

She finds two tables of data on the internet titled

- 'The world's 50 biggest coffee drinking countries (in kg per capita per year)'
- 'Top 20 tea consuming nations (kg/capita/yr)'

She does some further research and finds some new datasets. The **top 10 rows** of each dataset are presented in **Figures 4 and 5**.

| Country | Tea consumption per capita (kg) |
|---|---|
| Turkey | 3.16 |
| Ireland | 2.20 |
| United Kingdom | 1.95 |
| Russia | 1.39 |
| Morocco | 1.22 |
| New Zealand | 1.20 |
| Egypt | 1.01 |
| Poland | 1.00 |
| Japan | 0.97 |
| Saudi Arabia | 0.90 |
| ⋮ | ⋮ |

(Source: Adapted from real data)

**Figure 4: Tea consumption by country in 2015**

| Country | Coffee consumption per capita (kg) |
|---|---|
| Finland | 9.5 |
| Norway | 7.1 |
| Netherlands | 6.6 |
| Slovenia | 6.0 |
| Austria | 5.4 |
| Serbia | 5.3 |
| Denmark | 5.2 |
| Germany | 5.1 |
| Belgium | 4.8 |
| Brazil | 4.7 |
| ⋮ | ⋮ |

(Source: Adapted from real data)

**Figure 5: Coffee consumption by country in 2015**

(*b*) Explain how you would use spreadsheet software to organise the datasets into a format suitable for investigating correlation.

**(3)**

## Ali's Model Answer

- **Sort** the data in both tables into alphabetical order by country
- Copy and paste both tables into the same spreadsheet
- Clean the data by removing countries that do not appear in each table
- Line up the tea and coffee consumption for matching countries

| | | | | |
|---|---|---|---|---|
| **5(b)** | **Sort** each table by country (alphabetically)<br>**(then)**<br>Move data into adjacent columns/rows<br>**(then)**<br>Check for differences/errors in the lists of countries, and correct or remove them. | E1<br><br>E1<br><br>E1 | 1.1<br><br>1.1<br><br>1.1 | '**Sort**' should be seen<br><br>or copy & paste etc<br>Condone 'fields' |

# Database questions

Candidates need to be aware of:

- General database terminology
- How to construct queries which select desired data
- How to understand a database structure

# Database questions

Candidates need to be aware of:

- **General database terminology**
- How to construct queries which select desired data
- How to understand a database structure

**Terminology**

**Field (columns)**
A property/trait to be recorded in a database

**Record (rows)**
A collection of field values

**Table**
A collection of records

**Query**
The mechanism to extract any information from a database

# Database questions

Candidates need to be aware of:

- General database terminology
- **How to construct queries which select desired data**
- How to understand a database structure

**<u>Query Functions</u>**

**Select**
Selects only the fields you wish to view

**Sort**
Sorts the records in an order for a field (either alphabetically/numerically in either direction)

**Filter**
Displays only records satisfying a field condition

**Group**
Returns all possible values for a field with no duplication

**Join**
Joins tables together matching the records by a common field

Databases can also combine numerical field values with standard mathematical operations in a query

# Database questions

Candidates need to be aware of:

- General database terminology
- How to construct queries which select desired data
- **How to understand a database structure**

# Database example [9ST0/01 (2022) Q4]

**Ali's Model Answer**

- Produce a query
- Filter IndidentGroup = "Fire" and DateOfCall between 1/1/18 and 31/12/18
- Sort by FirstPumpArriving_AttendanceTime from lowest to highest
- Select first ten records

| Question | Scheme | Marks | AO | Notes |
|---|---|---|---|---|
| 4 | I would use a **filter**… | | | Must see 'filter'. |
| | …on 'IncidentGroup' to find 'Fire' incidents… | | | |
| | …and on 'DateOfCall' between 1/1/18 and 31/12/18. | | | |
| | | E1, E1 | 1.1, 1.1 | E1 for any two of these.<br>E2 for all three. |
| | | | | |
| | I would **sort** by 'FirstPumpArriving _AttendanceTime" (in ascending order). | E1 | 1.1 | Must see 'sort'<br>**Condone** truncated field name. |
| | Then I would select the top/bottom 10 records (to find the fastest times) | E1 | 1.1 | |
| | **Total** | **4** | | |

**4** London Fire Brigade keeps detailed records of every incident the fire service responds to. The full dataset contains 31 fields of data for each individual incident dating from 1st January 2017 until the present day.

Some selected fields are shown in **Figure 3**.

| Field name | Data type | Description |
|---|---|---|
| IncidentNumber | Number | ID number of incident |
| DateOfCall | Date/Time | Date of call |
| TimeOfCall | Date/Time | Time of call |
| IncidentGroup | Short Text | Type of incident ('Fire', 'Special Service', or 'False Alarm') |
| PropertyCategory | Short Text | Type of property/location (e.g. 'Dwelling', 'Non-residential', 'Outdoor') |
| Postcode_full | Short Text | Full postcode of location |
| IncGeo_BoroughName | Short Text | Name of London Borough |
| Easting_m | Number | UTM geographical coordinate of location (East) |
| Northing_m | Number | UTM geographical coordinate of location (North) |
| IncidentStationGround | Short Text | Name of fire station attending incident |
| FirstPumpArriving_ AttendanceTime | Number | Time (in seconds) for first fire engine to arrive at the location |
| NumPumpsAttending | Number | Number of fire engines attending the incident. |

(Source: UK Open Government Licence v2)

**Figure 3: Table structure of 'London Fire Brigade Incident data from January 2017'**

Explain how you would use database software to find the top 10 fastest arrival times for 'Fire' incidents during **2018**.

**(4)**

# Main warning with technology questions

**If the question is about a spreadsheet:**

You must give an answer appropriate to spreadsheets.
If you give an answer for databases (e.g. produce a query…) then the candidate may not receive full (or any) marks.

**Conversely, if the question is about a database:**

You must give an answer appropriate to databases.
If you give an answer for spreadsheets (e.g. copy and paste columns) then the candidate may not receive full (or any) marks

There are only two software types being assessed and it is important that candidates know how to use each

# Designing Experiments and The Statistical Enquiry Cycle

# Designing Experiments / The SEC

These questions arise often in exam papers:

The candidate is given a scenario and is asked to detail how they would go about investigating the problem at hand.

This may include

**A: Planning**
What data should be collected, what experimental designs could be utilised, what steps can be taken to reduce bias etc.

**B: Data collection**
Details on how the data is to be collected

**C: Data organisation / processing**
How the data will be handled, what graphs could be used to represent the data

**D: Data interpretation**
What analysis or hypothesis test could be used for the data collected

**E: Evaluation and Review**
What are the advantages and disadvantages of the method they have described

This is the Statistical Enquiry Cycle which can be assessed in any of the three papers.

**Note on Stage D**: The selection of an appropriate hypothesis test comes under “SEC” and can be assessed in Paper 1, but would not have to carry out the test in Paper 1

# Designing Experiments / The SEC

**General Strategies**

- Bullet points!
- The question often give an outline of a structure the candidates are to follow. Candidates should use this as a guide when answering the questions.
- Advice for candidates:
  “Put yourself in that situation… what would YOU do? Write it down!”
- Questions like these often explicitly say what steps you would take to reduce bias.

Candidates often confuse “bias” with “experimental error”

e.g. “collect data at the same time of day will reduce bias”

This actually **increases** bias. The data are now biased to that time of day. It will, however, decrease experimental error arising from the effect of different times of day.

# Example [9ST0/01 (2024) Q3]

**3** Ned is a Physics education researcher at a university.

Ned wants to investigate whether there is any evidence of association between the number of sports activities that a student takes part in each week and their Physics grade.

(a) Design an experiment for Ned to carry out. Your answer should include details regarding

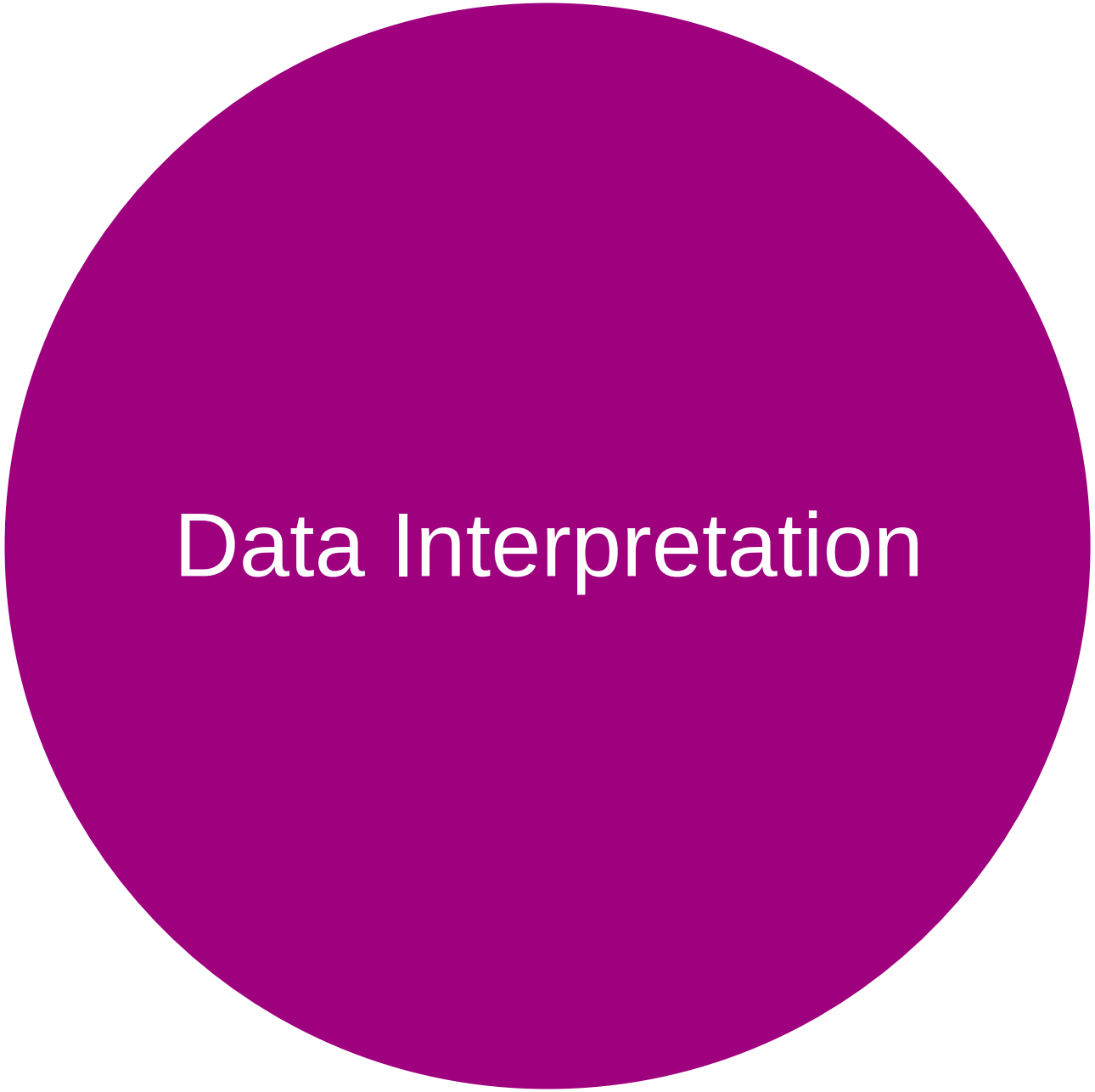
- what data he should gather, and how he can gather it
- how Ned should choose his sample
- which hypothesis test could be carried out with his data
- the hypotheses that should be used for this test.

**(6)**

**Ali's Model Answer**

- For each student, record the number of sporting activities and the Physics score
- Use a questionnaire to find out how many sports activities a student does per week and their Physics exam score
- Ned should take a large sample of students ($n \geq 30$)
- Using the university enrolment register, number the students and use a random number generator to generate (e.g.) 100 numbers without repeats and select the students assigned those numbers
- Carry out a PMCC test with $H_0$: $\rho = 0$ and $H_1$: $\rho \neq 0$

| | | | | |
|---|---|---|---|---|
| **3(a)** | For each student he needs their grade in a specific physics exam and the number of activities in categories | E1 | 1.1 | |
| | Ned could obtain this data via a survey or questionnaire | E1 | 1.1 | Or other suitable method<br><br>2 comments on gathering data |
| | Sample should be chosen at **random** | E1 | 1.1 | |
| | Sample should be reasonably **large** | E1 | 1.1 | Accept 30+ students<br><br>2 comments on sample |
| | Test on Spearman's Rank Correlation Coefficient (or PMCC) | E1 | 2.1a | Or Chi-squared test on contingency table<br><br>Accept |
| | $H_0$: No association between number of sports played per week and Physics grade ($\rho_s = 0$)<br>$H_1$: An association between number of sports played per week and Physics grade ($\rho_s \neq 0$) | E1 | 1.3 | $H_0$: Physics grade and number of sports played are independent<br>$H_1$: Physics grade and number of sports played are not independent |


Data Interpretation

# Data Interpretation

**General strategies**

- Bullet points!
- Adhere to the number of points requested in the question

**Common mistakes**

“Make 3 comments…”

Vs

“Make 3 criticisms…”

Candidates should be aware of the difference in the language

**Comments**

Talk about what the data visualisation is trying to convey
e.g.
“The Welsh income is higher than the English income”
etc.

**Criticisms**

Talk about why the data visualisation is not fit for purpose
e.g. “CPI is not defined”
or “The numbers are too difficult to read”

# Hypothesis Testing

# Hypothesis Testing

General things that should be noted about what to write in a hypothesis test

- Define the variables
- State the hypotheses clearly
- Declare the test (one- or two-tailed? Name of the test? Significance level? Important information?)
- State the test statistic clearly
- Find the *p*-value or critical values and explicitly compare the *p*-value with the significance level or the test statistic with the critical value
- Make a conclusion in context

Mentioned earlier

Any Greek symbols must be clear and any letters/subscripts must be defined

Although some of this is not required for the awarding of marks, it is generally a good idea for a candidate to mention these.

Name of the test: may be awarded marks if the question asks for it

Important information: Sample size? Degrees of freedom? Distribution being used?

These may be awarded M1 marks (which are implied by correct answers)

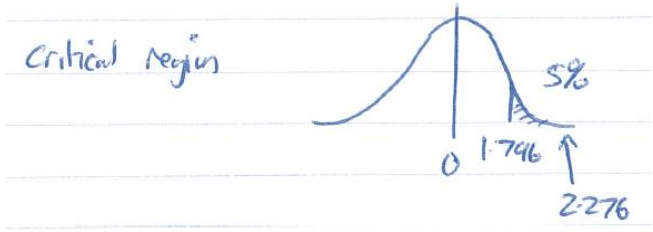
# Hypothesis Testing

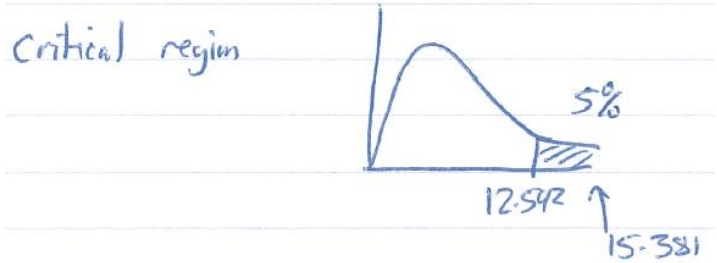
General things that should be noted about what to write in a hypothesis test

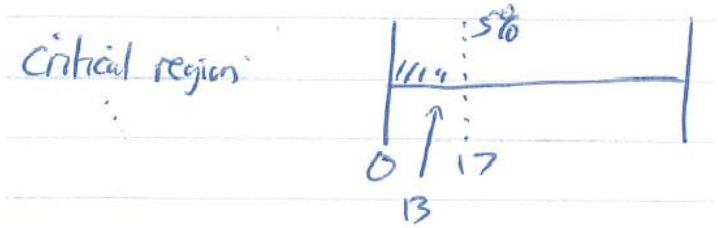
- Define the variables
- State the hypotheses clearly
- Declare the test (one- or two-tailed? Name of the test? Significance level? Important information?)
- State the test statistic clearly
- Find the *p*-value or critical values and explicitly compare the *p*-value with the significance level or the test statistic with the critical value
- Make a conclusion in context

This may be written as statements of fact:
e.g. “0.0231 < 0.05”
or “2.04 > 1.96”

OR may be displayed on a diagram (easier for CRs)



# Hypothesis Testing

General things that should be noted about what to write in a hypothesis test

- Define the variables
- State the hypotheses clearly
- Declare the test (one- or two-tailed? Name of the test? Significance level? Important information?)
- State the test statistic clearly
- Find the *p*-value or critical values and explicitly compare the *p*-value with the significance level or the test statistic with the critical value
- Make a conclusion in context

**Note about *p*-values: "The probability of observing the test statistic or more extreme if the null hypothesis is true"**

A *p*-value and the probability are **not** necessarily the same thing.

Traditionally (in Maths education pre-HE), candidates would calculate a probability and compare it with the significance level (one-tailed) or half the significance level (two-tailed).

However, the *p*-value is the probability calculated (one-tailed) or **double** the probability calculated (two-tailed) and then compared with the full significance level.

During comparisons in hypothesis tests, it does not matter which method they use
**BUT**
if the question specifically asks for the *p*-value (and the test is two-tailed), the probability must be doubled.

# Hypothesis Testing

General things that should be noted about what to write in a hypothesis test

- Define the variables
- State the hypotheses clearly
- Declare the test (one- or two-tailed? Name of the test? Significance level? Important information?)
- State the test statistic clearly
- Find the *p*-value or critical values and explicitly compare the *p*-value with the significance level or the test statistic with the critical value
- Make a conclusion in context

**What should be in the conclusion:**

- Whether to reject or not reject $H_0$
  Candidates are expected to use "Do not reject $H_0$" as opposed to the historical (and technically incorrect) "Accept $H_0$". Modern mark schemes (in both A level stats and maths) use "do not reject" and **condone** "accept"… but there will be a time in the future when "accept" may not be… accepted

- The conclusion in context, which should contain the words:
  - [significant/sufficient OR insufficient/no significant] "evidence to suggest"
  - A reference to $H_0$ being rejected **or** a suggestion that $H_1$ might be true **in words**.
  - The dependent variable (the thing being measured)
  - Any factor levels if applicable (between groups/types/trees/cars etc.)

# Example Hypothesis Test

*[9ST0/02 (2024) Question 7b]*

| City | COUNT | SUM (£) | STDEV (£) |
|---|---|---|---|
| **Liverpool** | 50 | 314250 | 425 |
| **Manchester** | 42 | 268800 | 511 |

(b) Making any necessary assumptions, carry out a hypothesis test to investigate whether people living in Manchester and Liverpool have different **average** levels of savings.

(8)

Let $X$ be the levels of savings for a person in Liverpool and let $Y$ be that for Manchester

$H_0: \mu_X - \mu_Y = 0$

$H_1: \mu_X - \mu_Y \neq 0$

Use a two-tailed two-sample $z$-test at the 5% level using $Z \sim N(0,1)$

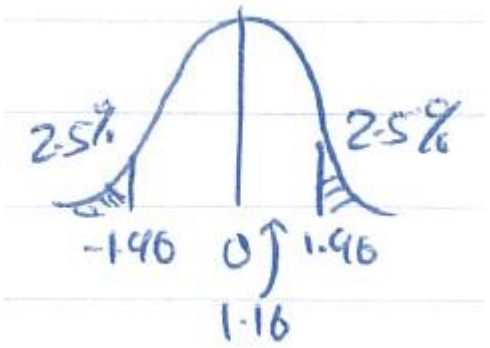
$\overline{x} = 6285 \quad \overline{y} = 6400$

$s_x^2 = 425^2 \quad s_y^2 = 511^2$

$n_x = 50, \quad n_y = 42$

Test statistic: $\frac{(6400-6285)-(0)}{\sqrt{\frac{425^2}{50}+\frac{511^2}{42}}} = 1.16$

Critical Region:




Result not significant. Do not reject $H_0$.

There is insufficient evidence to suggest a difference in mean levels of savings of a person between Liverpool and Manchester

# Example Hypothesis Test

*[9ST0/02 (2024) Question 7b]*

| City | COUNT | SUM (£) | STDEV (£) |
|---|---|---|---|
| **Liverpool** | 50 | 314250 | 425 |
| **Manchester** | 42 | 268800 | 511 |

(b) Making any necessary assumptions, carry out a hypothesis test to investigate whether people living in Manchester and Liverpool have different **average** levels of savings.

(8)

Let *X* be the levels of savings for a person in Liverpool and let *Y* be that for Manchester

$H_0: \mu_X - \mu_Y = 0$

$H_1: \mu_X - \mu_Y \neq 0$

Use a two-tailed two-sample *z*-test at the 5% level using $Z \sim N(0,1)$

$\bar{x} = 6285 \quad \bar{y} = 6400$

$s_x^2 = 425^2 \quad s_y^2 = 511^2$

$n_x = 50, \quad n_y = 42$

Test statistic: $\frac{(6400-6285)-(0)}{\sqrt{\frac{425^2}{50}+\frac{511^2}{42}}} = 1.16$

*p*-value:

$P(Z \geq 1.16) = 0.123$

*p*-value = $2 \times 0.123 = 0.246 > 0.05$

Result not significant. Do not reject $H_0$.
There is insufficient evidence to suggest a difference in mean levels of savings of a person between Liverpool and Manchester

# Hypothesis Test Glossary

**PMCC Test**

**Note:** these may be written in words using "population PMCC"

$$\mathrm{H}_0: \rho = 0$$

$$\mathrm{H}_1: \begin{cases} \rho > 0 \\ \rho < 0 \\ \rho \neq 0 \end{cases}$$

**Spearman's Rank Test**

**Note:** these may be written in words using "no association" etc.

$$\mathrm{H}_0: \rho_s = 0$$

$$\mathrm{H}_1: \begin{cases} \rho_s > 0 \\ \rho_s < 0 \\ \rho_s \neq 0 \end{cases}$$

**Binomial proportion test (with or without normal approximation)**

**Note:** $\pi$ or $p$ may be used here

$$\mathrm{H}_0: \pi = \cdots$$

$$\mathrm{H}_1: \begin{cases} \pi > 0 \\ \pi < 0 \\ \pi \neq 0 \end{cases}$$

**Two-Sample proportion test**

**Note 1:** $\pi$ should be used here since $p$ is the pooled estimate of the proportion

**Note 2**: Any equivalent rearrangement of the expressions are accepted

$$\mathrm{H}_0: \pi_X - \pi_Y = 0$$

$$\mathrm{H}_1: \begin{cases} \pi_X - \pi_Y > 0 \\ \pi_X - \pi_Y < 0 \\ \pi_X - \pi_Y \neq 0 \end{cases}$$

# Hypothesis Test Glossary

**_z_- and _t_-Tests**

$$\mathrm{H}_0: \mu = \cdots$$

$$\mathrm{H}_1: \begin{cases} \mu > \cdots \\ \mu < \cdots \\ \mu \neq \cdots \end{cases}$$

**Sign and Wilcoxon Signed-Rank Tests**

$$\mathrm{H}_0: \eta = \cdots$$

$$\mathrm{H}_1: \begin{cases} \eta > \cdots \\ \eta < \cdots \\ \eta \neq \cdots \end{cases}$$

**Note:** $\mu$ used in a Sign test is incorrect and will note be awarded marks.
$\mu$ used in a Wilcoxon Signed-Rank test is condoned but $\eta$ is preferred since the underlying method is about the median.

**Two-sample _z_- and _t_-tests**

**Note:** Any equivalent rearrangements of the expressions are accepted

$$\mathrm{H}_0: \mu_X - \mu_Y = \cdots$$

$$\mathrm{H}_1: \begin{cases} \mu_X - \mu_Y > \cdots \\ \mu_X - \mu_Y < \cdots \\ \mu_X - \mu_Y \neq \cdots \end{cases}$$

**Paired Tests**

Exactly the same as the unpaired versions but a subscript $D$ or $d$ on the Greek symbol ($\mu_D$ or $\eta_D$).
$D$ (or $d$) must be clearly defined.

# Hypothesis Test Glossary

**$\chi^2$ tests for association**

**NOTE:** Either statement is accepted

$$H_0: \begin{cases} \text{There is no association between ...} \\ \text{... and ... are independent} \end{cases}$$

$$H_1: \begin{cases} \text{There is an association between ...} \\ \text{... and ... are not independent} \end{cases}$$

**$\chi^2$ Goodness of Fit (known parameters)**

$H_0$: … is a suitable model

$H_1$: … is not a suitable model

**Note:** if parameters are known, they should be mentioned in the hypotheses and conclusion

**$\chi^2$ Goodness of Fit (estimated parameters)**

$H_0$: The … distribution is a suitable model

$H_1$: The … distribution is not a suitable model

**Note:** if parameters are estimated from the data, they should not be mentioned in the hypotheses or conclusion

**Note 2**: A general discrete random variable, or the binomial, Poisson, exponential, continuous uniform or normal distributions may be assessed for Goodness of Fit

# Hypothesis Test Glossary

**Analysis of Variance (ANOVA)**

$$\mathrm{H}_0: \mu_1 = \cdots = \mu_k$$

$\mathrm{H}_1$: At least two means are different

**Notes:**

- In $\mathrm{H}_0$, all letters/subscripts must be clearly defined.
  In situations where there are a lot of groups, and the ordinal indexing (1,2,…,*k*) are used **and defined**, then "…" may be used in between the first and last.
- The $\mathrm{H}_1$ shown here minimises the amount of writing required.
  "At least one mean is different from the others" is equivalent but is more to write down.

# Hypothesis Test Glossary

**Wilcoxon Rank-Sum Tests**

**NOTE:** Any equivalent rearrangement of the expressions are accepted

$$H_0: \eta_X - \eta_Y = 0$$

$$H_1: \begin{cases} \eta_X - \eta_Y > 0 \\ \eta_X - \eta_Y < 0 \\ \eta_X - \eta_Y \neq 0 \end{cases}$$

**Special Note:**

There are actually two versions of the Wilcoxon Rank-Sum test.

The **first version** is a test for a <u>difference in distributions</u> and relaxes the assumption of “similar shaped distributions” but the hypotheses become

$H_0$: The samples are taken from populations with identical distributions

$H_1$: The samples are taken from populations with different distributions

These hypotheses are condoned.

The **second version** is a test for a <u>difference in medians</u> and requires the assumption of “similar shaped distributions” allowing the hypotheses to be refined to those seen on the left.

Candidates are only ever tested on this second version.

# Errors and Power

A **Type I Error** is when $H_0$ is rejected but is actually true

A **Type II Error** is when $H_0$ is not rejected but is actually false

The **power** of a hypothesis test is the probability of correctly rejecting $H_0$

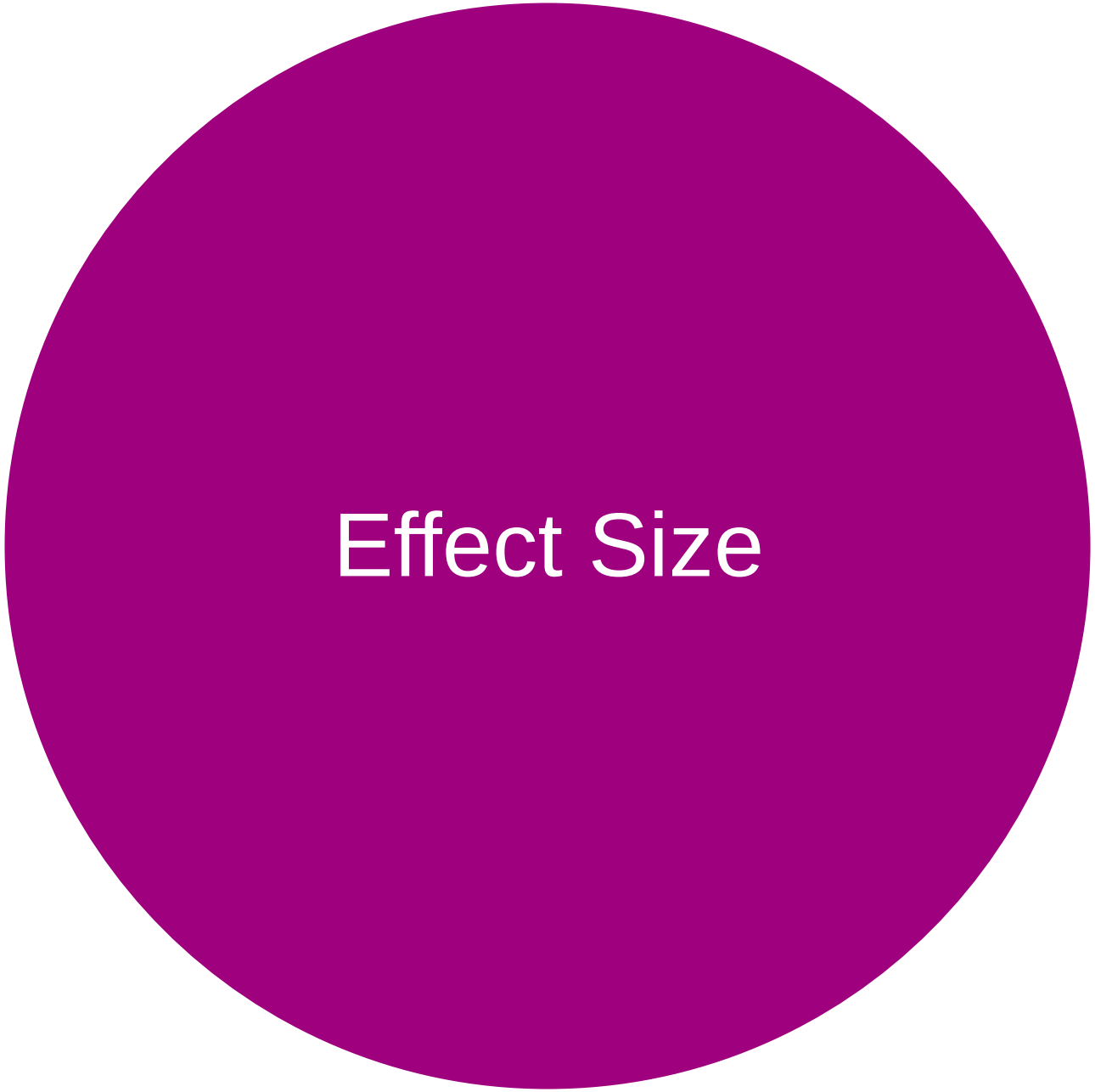
The **probability of a Type I Error** is the size of the critical region
(for continuous distributions this is just the significance level, but for discrete distributions it will likely be lower)

The **probability of a Type II Error** is $1 - \text{Power}$

**General strategies in an exam:**

- Context at all times!

e.g.
“A Type I Error is when it is concluded that there is a negative correlation between age and running speed, when in fact there is no correlation”


Effect Size

# Effect Size

This is a recent development in Statistics in general (developed in 1960s, not in widespread used until late 80s)

Effect size is a numerical way of measuring the **effect** of a factor
e.g. How **large** is the difference in means
or How **strong** is the correlation

One underlying principle of effect size is that it is **not** dependent on sample size (unlike p-values which very much are!)

Examples of effect size measures: Pearson's PMCC and Spearman's Rank

**Cohen's *d***

Measures the difference in means in terms of pooled estimates of standard deviations

i.e. (in the language of a two-sample *t*-test)
"How many $s_p$'s apart are the two sample means?"

The specification states the following guidance:

$0 \leq d < 0.3$ "Small effect"

$0.3 \leq d < 0.8$ "Medium effect"

$d \geq 0.8$ "Large effect"

**Note**: these are arbitrary guidelines based off observed data… and these were not Jacob Cohen's original guidelines but this is what is on the spec!

# Effect Size

**<u>How to interpret effect size in combination with a hypothesis test</u>**

**Case 1: If the hypothesis test provides significant evidence of a difference in means AND Cohen's *d* indicates a large effect**

"There is significant evidence to suggest a difference in means and this difference may be large."

**Case 2: If the hypothesis test does not provide significant evidence of a difference in means AND Cohen's *d* indicates a small effect**

"There is insufficient evidence to suggest a difference in means and, and if there were, this difference may be small anyway."

**Case 3: If the hypothesis test provides significant evidence of a difference in means AND Cohen's *d* indicates a small effect**

"There is significant evidence to suggest a difference in means, but this difference may be small so it may not be useful in practice."

**Case 4: If the hypothesis test does not provide significant evidence of a difference in means AND Cohen's *d* indicates a large effect**

"There is insufficient evidence to suggest a difference in means but if there were, the difference might be large. It is worth investigating further with a larger sample since the sample may be too small to detect a difference in means"

# Example [9ST0/03 (SAM) Q1]

**1** As part of an investigation into the effect of protein on weight gain, twenty female rats were to be observed from the ages of four weeks to seven weeks.

Ten of them were given a high protein diet, while ten were given a low protein diet. For reasons not connected to diet, 5 rats were lost to the study leaving 6 in the high protein group and 9 in the low protein group.

The table below gives the means and the standard deviations of the observed weight gains (in grams).

| | High protein diet | Low protein diet |
|---|---|---|
| $\bar{x}$ | 12.1 | 10.3 |
| $s$ | 2.02 | 1.95 |
| $n$ | 6 | 9 |

A t-test for comparing the mean weight gains for this data is carried out and found to have a $p$-value of 0.11

(a) Calculate the value of Cohen's $d$ for the effect of a high protein diet compared to a low protein diet on the weight gain of such rats. Assume that the variances of weight gain for the two diets are equal.

**(3)**

(b) Describe what the information provided by the Cohen's $d$ value evaluated in part (a), together with the given $p$-value, reveal about the possible effect of protein on weight gain in such rats.

You should include a comment explaining the importance of considering the effect size together with the $p$-value when carrying out this t-test.

**(3)**

**Ali's Model Answer**

(a)

$$s = \sqrt{\frac{(6-1)\times 2.02^2 + (9-1)\times 1.95^2}{6+9-2}} = 1.977$$

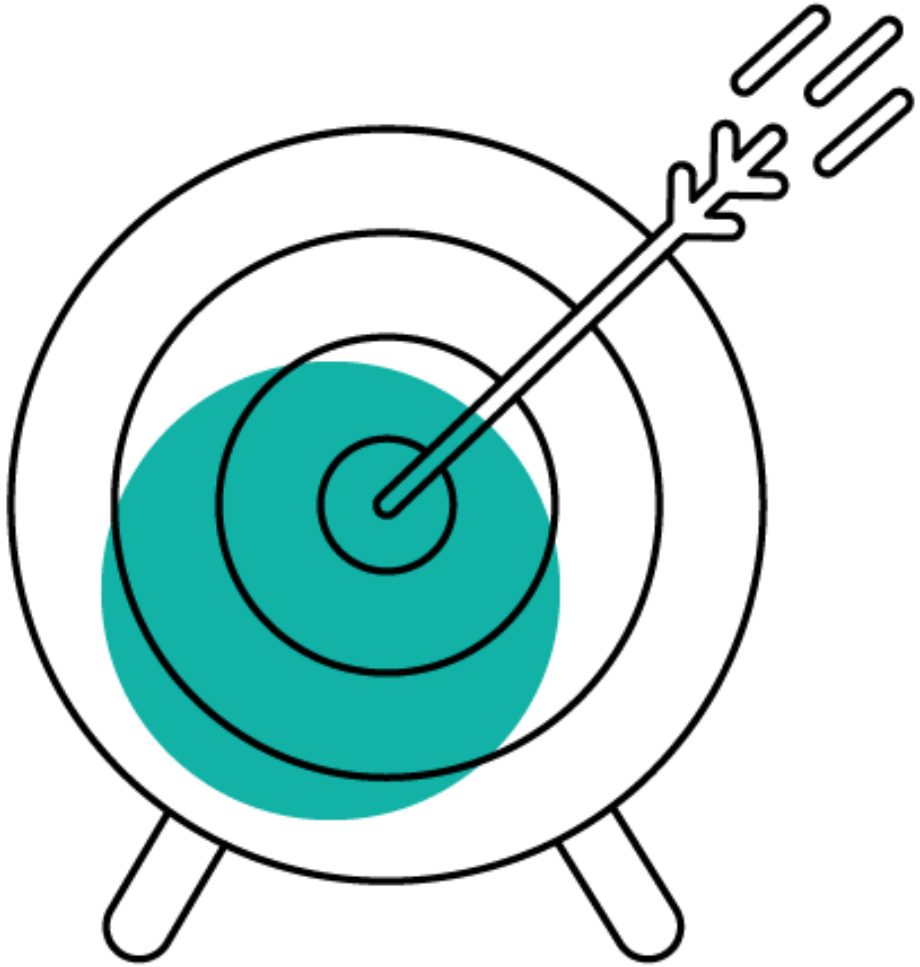
$$d = \frac{12.1 - 10.3}{1.977} = 0.910$$

(b)

- The $p$-value of 0.11 is bigger than 5% so there is insufficient evidence to suggest a difference in mean weight gains between the different diets.
- The Cohen's $d$ of 0.910 indicates a large effect on the mean weight gains between the different diets.
- This suggests that if there were a difference in mean weight gains, it may be large. Further investigation is required with larger samples since the original samples may be too small to detect a difference in mean weight gains.

# Summary: So what have we covered?

- Assessment objectives for A Level Statistics
- General strategies for answering questions
- Strategies for subject specific questions
    - Technology
    - Designing Experiments / The Statistical Enquiry Cycle
    - Data Interpretation
    - Hypothesis Testing
    - Effect Size



# Resources available on the Maths Emporium


## Exam Papers and Mark Schemes

- Sample Assessment Papers (1 set)
- Specimen Paper (Paper 3 only)
- Practice Papers (3 sets)
- Exam Papers (2019 – 2024)
- AS Exam Papers (2018-2020)
  These are still good questions for practice purposes
- Shadow Papers (2020 – 2024)

## Topic Tests

- Sampling and Data Collection
- Technology and Software
- Experimental Design
- Data Visualisation

- **Examiners Reports and exemplars**

- **"Student Friendly" Worked solutions**
  2024 and Shadow 2024 are already available.
  As a treat for you, I have created "Student Friendly" worked solutions for 2022 and 2023 (and the shadow versions)

- **Scheme of Work**
  This will be updated this year to reflect updates to assessment arising from the last 7 years.

  Includes links to Desmos activities to help visualise the graphs of probability distributions

# Subject Advisor Support

Our subject advisors are experts in their fields and are here to support you throughout the year.

**Mathematics and Statistics**

**Email**: teachingmaths@pearson.com

**Phone**: +44 (0) 344 463 2535
(Mon–Fri, 9.00–17.00)

Book an appointment with your Subject Advisor

Sign up to receive regular updates from your Subject Advisor on qualification news and support for your subject.

**Vicky Wood**
Mathematics and Statistics

Pearson

# Find out more

For more professional development courses please see Pearson's [Professional Development Academy]

# Your Feedback Matters

Following this event, you will receive an invitation to share your thoughts about the session. Your feedback is invaluable to us, as it helps us tailor our professional development materials to better meet your needs. Please don’t hesitate to let us know what you’d like to see more of and what areas you think could be improved.

Pearson