GCSE STATISTICS

KNOWLEDGE TEST
HOMEWORK REVISION BOOKLET

1 – Types of data

Definition	Terminology
1. Data as it has been collected, before it has been processed	Raw data
2. Worded data	Qualitative data
3. Data that can be split into clear, non-overlapping categories	Categorical data
4. Numerical data	Quantitative data
5. Data that can be ordered	Ordinal data
6. Numerical data that can take any value	Continuous data
7. Numerical data that can only take specific values	Discrete data
8. Involving pairs of related data	Bivariate data
9. Involving 3 or more sets of related data	Multivariate data
10. The statistics term for a number or amount	Frequency

2 – Methods of data collection

Definition	Terminology
1. Data collected by, or for, the person who is going to use it	Primary data
2. Data collected by an outside source	Secondary data
3. Everyone or everything that could be included in an investigation	Population
4. Collecting data using the entire population	Census
5. Collecting data using part of the population	Sample
6. Each individual member of the population	Sample unit
7. A list of the population	Sampling frame
8. Defines the subject area to investigate	Research question
9. An idea that can be tested	Hypothesis
10. A tendency to favour one person or thing over another	Bias

3 – Sampling methods

Definition	Terminology
1. A sample in which every member of the population is equally likely to be included	Random sample
2. A sample which is proportionally representative of the population's characteristics	Stratified sample
3. A sample selected at fixed intervals	Systematic sample
4. A sample selected by the investigator, aiming to be representative of the population's characteristics	Judgement sample
5. A sample selected by the investigator, using the quickest or easiest members of the population	Opportunity sample
6. A sample where the population is split into groups or areas	Cluster sample
7. A sample where the proportion of members with each characteristic is pre-selected	Quota sample
8. A question which gives the respondent answer boxes or options	Closed question
9. A question which allows the respondent to give their own answer	Open question
10. A question that prompts or encourages a certain response	Leading question

4 – Terminology

Definition	Terminology
1. An investigation that tests what it is supposed to test	Validity
2. An investigation that gives consistent results	Reliability
3. The variable being investigated	Explanatory variable
4. The outcome of an investigation	Response variable
5. Variables outside our control that may affect the investigation	Extraneous variable
6. A group that receives no treatment or is given a placebo to ensure results are comparable	Control group
7. To perform a scientific investigation, often in a laboratory	Experiment
8. To run a computer program or model of an investigation	Simulation
9. A natural experiment or study involving no physical intervention	Observation
10. Data collected from a series of questions or prompts	Questionnaire

5 – Presenting data 1

Definition	Terminology
1. The process of identifying and correcting errors / issues within raw data	Cleaning data
2. A graph using shapes or pictures	Pictogram
3. A graph using line markings, split into groups of five	Tally chart
4. A bar chart with data sets side by side	Dual bar chart
5. A bar chart with multiple data sets within each bar	Composite bar chart
6. A method of organising data to show its distribution	Stem and leaf diagram
7. A graphical representation of the age and sex of a population	Population pyramid
8. A map using shading or colours	Choropleth map
9. The running total of frequencies	Cumulative frequency
10. A diagram displaying the minimum value, LQ, median, UQ and maximum value	Box plot

6 – Presenting data 2

Definition	Terminology
1. A visual representation of the relationship between data sets, usually with overlapping circles	Venn diagrams
2. The statistical relationship between variables, showing how they change in relation to each other	Correlation
3. A straight line on a graph that best represents the trend of all data points	Line of best fit
4. Estimating inside the given data range	Interpolation
5. Estimating outside the given data range	Extrapolation
6. A graph using bars of varying widths, for continuous data	Histogram
7. Shown on the y-axis of a histogram	Frequency density
8. A point that lies exactly halfway between two other points	Midpoint
9. The vertical axis on a graph	Y – axis
10. The horizontal axis on a graph	X – axis

7 – Averages and range

Definition	Terminology
1. The average data value	Mean
2. The most common data value	Mode
3. The middle data value	Median
4. The difference between the highest and lowest data values	Range
5. Splitting a data set into 4 parts	Quartiles
6. Splitting a data set into 10 parts	Deciles
7. Splitting a data set into 100 parts	Percentiles
8. Shows the rate of change in prices in everyday life – such as mortgage payments, food, heating and petrol. Used to set interest rates for student loans	Retail Price Index (RPI)
9. Shows the rate of change in prices in everyday life – such as food, heating and petrol. Used to set state benefits and pensions	Consumer Price Index (CPI)
10. The value of goods and services a country produces within a given period of time	Gross Domestic Product (GDP)

8 – Probability

Definition	Terminology / Notation
1. The occurrence of one event does not affect the probability of the other	Independent events
2. The occurrence of one event does affect the probability of the other	Dependent events
3. A probability with 0% chance of occurring	Impossible
4. A probability with between 1% and 49% chance of occurring	Unlikely
5. A probability with 50% chance of occurring	Even chance
6. A probability with between 51% and 99% chance of occurring	Likely
7. A probability with 100% chance of occurring	Certain
8. Set notation to show **Not A**	A'
9. Set notation to show **A and B**	$A \cap B$
10. Set notation to show **A or B**	$A \cup B$

9 – Formulae

Topic	Formula
1. Calculating an angle for a pie chart	$\dfrac{x}{total\ frequency} \times 360°$
2. Frequency density	Frequency ÷ Class width
3. Range	Largest value – Smallest value
4. Interquartile range	Upper quartile – Lower quartile
5. Index numbers	$\dfrac{Price}{Base\ year\ price} \times 100$
6. Conditional probability	$P(A|B) = \dfrac{P(A \cap B)}{P(A)}$
7. Proportional pie charts	$\dfrac{Radius^2}{radius^2} = \dfrac{N}{n}$
8. Weighted mean	$\dfrac{\sum(value \times weight)}{\sum weights}$
9. Upper outlier	UQ + 1.5 IQR
10. Lower outlier	LQ – 1.5 IQR

10 – Year 11 topics

Definition	Terminology
1. The average used when dealing with percentages or rates of change	Geometric mean
2. A statistical method used to estimate the size of very large populations	Capture – recapture
3. The equation of the regression line – tells us the value of y when x is 0	Y – intercept
4. The equation of the regression line – tells us the rate of change	Gradient
5. A measure of comparing results for different tests and populations	Standardised scores
6. When a quality control sample crosses this line, we must immediately take another sample	Warning line
7. When a quality control sample crosses this line, we must immediately stop and reset the machine	Action line
8. The percentage of data that lies within one standard deviation of the mean	68%
9. The percentage of data that lies within two standard deviations of the mean	95%
10. The percentage of data that lies within three standard deviations of the mean	99.8%