

## Formulae

### Unit 1: Collection of Data

#### Petersen Capture-Recapture (Higher only)



$$N = \frac{Mn}{m}$$

$N$  = number in the population

$M$  = size of first sample

$n$  = size of second sample

$m$  = number marked in second sample

#### Stratified Sampling



$$\text{Number selected in stratum} = \frac{\text{Stratum size}}{\text{Population size}} \times \text{sample size}$$

### Unit 2: Processing, Representing and Analysing Data

#### Pie Chart: angle of sector



$$\text{Sector angle} = \frac{x}{\text{total}} \times 360^\circ$$

$x$  = frequency of sector

total = total frequency

#### Comparative Pie Charts (Higher Only)



$$r_2 = r_1 \times \sqrt{\frac{F_2}{F_1}}$$

$r$  = radius

$F$  = frequency

#### Histograms



$$\text{Frequency density} = \frac{\text{frequency}}{\text{class width}}$$

Rearrange



$$\text{Frequency} = \text{frequency density} \times \text{class width}$$



Formula given in the exam



Must remember as not provided in the exam

## Unit 3: Summarising Data

**Median** 

**Discrete Data:** Position of Median =  $\frac{1}{2}(n + 1)th$

$n$  = number of data values

**Grouped Data:** Position of Median =  $\frac{1}{2}n th$

**Mean: Discrete Data** 

$$\bar{x} = \frac{\sum x}{n}$$

$\bar{x}$  = mean

$\sum$  = sum of

$x$  = data values

$n$  = number of data values

**Mean: Frequency Table (not grouped)** 

$$\bar{x} = \frac{\sum fx}{\sum f}$$

$\bar{x}$  = mean

$\sum$  = sum of

$x$  = data values

$f$  = frequency

**Mean: Grouped Data** 

Estimated mean =  $\frac{\sum(f \times \text{midpoint})}{\sum f}$

$\sum$  = sum of

$f$  = frequency

**Median: Interpolation** 

$$\text{Estimated median} = L + \frac{\frac{n}{2} - F}{f} \times w$$

$L$  = the lower boundary of class containing the median

$n$  = total number of data values

$F$  = cumulative frequency of the intervals *before* class containing the median

$f$  = frequency of the class containing the median

$w$  = width of the class containing the median



Formula given in the exam



Must remember as not provided in the exam

## Unit 3: Summarising Data

**Geometric Mean (Higher Only)** 

$$\text{Geometric mean} = \sqrt[n]{\text{value}_1 \times \text{value}_2 \times \dots \times \text{value}_n}$$

where  $n$  is the total number of values

**Weighted Mean (Higher Only)** 

$$\text{Weighted mean} = \frac{\sum(\text{value} \times \text{weight})}{\sum \text{weights}}$$

**Range** 

$$\text{Range} = \text{highest value} - \text{lowest value}$$

**Quartiles** **Interquartile range:**

$$\text{Upper quartile} - \text{Lower quartile}$$

**Discrete data:**

$$\text{Lower quartile position} = \frac{1}{4}(n + 1)th$$

$$\text{Upper quartile position} = \frac{3}{4}(n + 1)th$$

**Grouped data:**

$$\text{Lower quartile position} = \frac{1}{4}n th$$

$$\text{Upper quartile position} = \frac{3}{4}n th$$

**Interpercentile Range (Higher Only)** 

$$\text{Interpercentile range} = \text{larger percentile} - \text{smaller percentile}$$

**Interdecile Range (Higher Only)** 

$$\text{Interdecile range} = \text{larger decile} - \text{smaller decile}$$



Formula given in the exam



Must remember as not provided in the exam

## Unit 3: Summarising Data

**Standard Deviation (Higher Only)**PROVIDED  
IN EXAM

$$\text{Standard deviation} = \sqrt{\frac{1}{n} \sum (x - \bar{x})^2} \quad \text{OR} \quad \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$n$  = the number of data values       $\sum$  = sum of       $x$  = data values       $\bar{x}$  = mean

**Standard Deviation: data in a frequency table (Higher Only)**

$$\text{Standard deviation} = \sqrt{\frac{\sum f(x - \bar{x})^2}{\sum f}} \quad \text{OR} \quad \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

$f$  = frequency       $\sum$  = sum of       $x$  = data values       $\bar{x}$  = mean

**Standard Deviation: grouped data (Higher Only)**

$$\text{Standard deviation} = \sqrt{\frac{\sum ft^2}{\sum f} - \left(\frac{\sum ft}{\sum f}\right)^2}$$

$f$  = frequency       $\sum$  = sum of  
 $t$  is equivalent to  $x$  for data values but could be any letter e.g.  $t$  for time or  $h$  for height

**Calculating Outliers (Higher Only)**

Small outlier is  $< LQ - 1.5 \times IQR$

Large outlier is  $> UQ + 1.5 \times IQR$

Outlier is more than 3 standard deviations from the mean

**Skewness (Higher Only)**PROVIDED  
IN EXAM

$$\text{Skew} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$



Formula given in the exam



Must remember as not provided in the exam

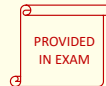
## Unit 4: Scatter Diagrams &amp; Correlation

**Equation of line of best fit**

$$y = ax + b$$

$a$  = gradient

$b$  = y-intercept

**Spearman's Rank Correlation Coefficient (Higher Only)**

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

$d$  is the difference in ranks

$n$  is the number of values in each set

## Unit 5: Time Series Analysis

**Seasonal Variation**

Seasonal variation = actual value – trend value

**Predicted Value**

Predicted value = trend line value – estimated mean seasonal variation



Formula given in the exam



Must remember as not provided in the exam

## Unit 6: Probability

**Theoretical or Expected Probability** 

$$P(\text{event}) = \frac{\text{number of favourable outcomes}}{\text{total number of outcomes}}$$

**Expected Frequency** 

$$\text{Expected frequency of event A} = P(A) \times \text{number of trials}$$

**Experimental or Estimated Probability** 

$$\text{Estimated probability} = \frac{\text{number of trials with favourable outcome}}{\text{total number of trials}}$$

**Absolute Risk** 

$$\text{Risk of event} = \frac{\text{number of trials in which event happens}}{\text{total number of trials}}$$

**Relative Risk** 

$$\text{Relative risk for a group} = \frac{\text{risk for those in the group}}{\text{risk for those not in the group}}$$

**Addition (Or) Rule for Mutually Exclusive Events** 

$$P(A \text{ or } B) = P(A) + P(B)$$



Formula given in the exam



Must remember as not provided in the exam

## Unit 6: Probability

**General Addition Law for Non-Mutually Exclusive Events** 

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

**Multiplication (And) Law for Independent Events** 

$$P(A \text{ and } B) = P(A) \times P(B)$$

Formula can be extended  
to 3, 4 or more events

$$P(A \text{ and } B \text{ and } C) = P(A) \times P(B) \times P(C)$$

**Conditional Probability** 

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

Rearrange



$$P(B|A) \times P(A) = P(A \text{ and } B)$$

**For Independent events:** 

$$P(B|A) = P(B)$$

$$P(A|B) = P(A)$$



Formula given in the exam



Must remember as not provided in the exam

## Unit 7: Index Numbers

**Index Numbers** 

$$\text{Index number} = \frac{\text{Price}}{\text{Base year price}} \times 100$$

**Weighted Index Numbers (Higher Only)** 

$$\text{Weighted index number} = \frac{\text{current weighted mean price}}{\text{base year weighted mean price}}$$

**Chain Base Index Numbers (Higher Only)** 

$$\text{Chain base index number} = \frac{\text{price}}{\text{last year's price}} \times 100$$

**Crude Rates of Change** 

$$\text{Crude rate} = \frac{\text{Number of births/deaths/other}}{\text{Total population}} \times 1\,000$$

**Standard Population (Higher Only)** 

$$\text{Standard Population} = \frac{\text{Number in age group}}{\text{Total population}} \times 1\,000$$

**Standardised Rate (Higher Only)** 

$$\text{Standardised Rate} = \frac{\text{Crude rate}}{1\,000} \times \text{standard population}$$



Formula given in the exam



Must remember as not provided in the exam

## Unit 8: Probability Distributions

**Binomial Distribution Mean (Higher Only)**Mean of binomial distribution  $B(n, p) = np$  $n$  is the number of trials $p$  is the probability of success**Normal Distribution Standardised Score (Higher Only)**Number of standard deviations from mean =  $\frac{\text{value} - \text{mean}}{\text{standard deviation}}$ 

Formula given in the exam



Must remember as not provided in the exam