GCSE STATISTICS FORMULAE (9-1 EDEXCEL)

Formula in green boxes are given in exam.

Refer to your Revision Guide for all topics.

Petersen capture-recapture $\frac{M}{N} = \frac{m}{n}$

M = original marked

m = new marked

n = new population

$N = total\ population$

Mean $(\bar{x}) = \frac{\sum x}{n}$

Weighted mean = $\sum(value \times weight)$

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

Standard deviation (not table):

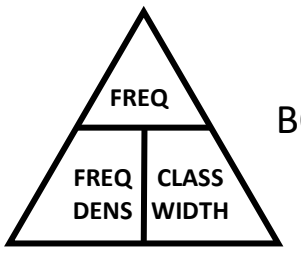
$= \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$ or $= \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$

Standard deviation (frequency table):

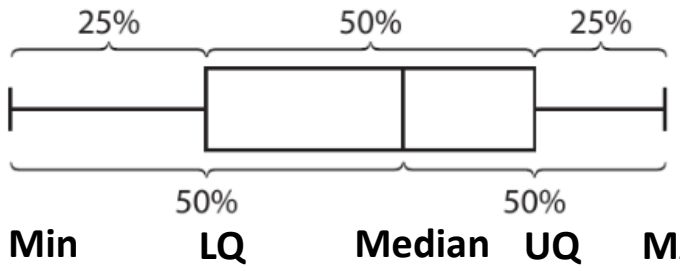
$= \sqrt{\frac{\sum f(x-\bar{x})^2}{\sum f}}$ or $= \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$

Histograms →frequency is area of the bar

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



BOX PLOT



For a set of data:

- mean > median > mode = positive skew
- mode > median > mean = negative skew

Spearman's rank correlation coefficient:

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

Small **outlier** is less than LQ – (1.5 x IQR)
Large outlier is greater than UQ + (1.5 x IQR)
OR mean $+3\sigma$ (σ = standard deviation)

Line of best fit should go through the mean point $(\bar{x}, \bar{y})$

The equation of the line $y = ax + b$

- has gradient a
- intercept on the y axis is (0, b)

Seasonal variation at a point = actual value – trend value
Estimated mean seasonal variation for any season = mean of the seasonal variations for that season
Predicted value = trend line value + estimated mean seasonal variation

Probability of an event (if possible outcomes are equally likely)

= $\frac{number\ of\ successful\ outcomes}{total\ number\ of\ possible\ outcomes}$

Expected frequency of event A = P(A) x number of trials

Estimated probability = $\frac{number\ of\ trials\ with\ successful\ outcomes}{total\ number\ of\ trials}$

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

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

Mutually exclusive, A and B: P(A∪B) = P(A) + P(B)
Exhaustive events: P(A) + P(not A) = 1 or P(not A) = 1 – P(A)
General addition law (for not mutually exclusive):
P(A∪B) = P(A) + P(B) – P(A∩B)
Independent events, A and B:
P(A∩B) = P(A) x P(B); P(A∩B∩C) = P(A) x P(B) x P(C)
Conditional probability, probability of B given A:

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

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

Two independent events A and B: P(A) = P(A|B)

Index numbers = $\frac{price}{base\ year\ price} \times 100$

Weighted index numbers = $\frac{current\ weighted\ mean\ price}{base\ year\ weighted\ mean\ price} \times 100$

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

Normal distribution:
68% ± one standard deviation of the mean
95% ± two standard deviation of the mean
99.8% ± three standard deviation of the mean

Normal distribution N(μ, σ^2)
Mean (μ) and variance (σ^2)
Variance = $(standard\ deviation - \sigma)^2$

Warning limits set at $\mu \pm 2\sigma$
Action limits set at $\mu \pm 3\sigma$

Standardised score = $\frac{score - mean}{standard\ deviation}$

Binomial distribution = B(n, p)
Binomial expansion = $(p+q)^n$
Mean of binomial expansion = np