

# 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** =  $\frac{\sum(\text{value} \times \text{weight})}{\sum(\text{value} \times \text{weight})}$

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

**Standard deviation (not table):**

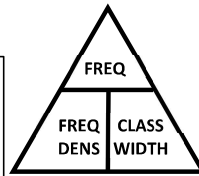
=  $\sqrt{\frac{\sum(x-\bar{x})^2}{n}}$  or =  $\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$

**Standard deviation (frequency table):**

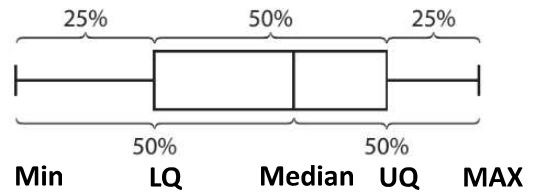
=  $\sqrt{\frac{\sum f(x-\bar{x})^2}{\sum f}}$  or =  $\sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$

**Histograms** → frequency is area of the bar

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



**BOX PLOT**



**Spearman's rank correlation coefficient:**

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

Small **outlier** is less than  $LQ - (1.5 \times IQR)$   
 Large outlier is greater than  $UQ + (1.5 \times IQR)$   
 OR  $\text{mean} + 3\sigma$  ( $\sigma$  = standard deviation)

For a set of data:

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

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)

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

**Expected frequency** of event A =  $P(A) \times$  number of trials

**Estimated probability** =

$\frac{\text{number of trials with successful outcomes}}{\text{total number of trials}}$

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

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

**Mutually exclusive**, A and B:  $P(A \cup B) = P(A) + P(B)$

**Exhaustive events**:  $P(A) + P(\text{not } A) = 1$  or  $P(\text{not } A) = 1 - P(A)$

**General addition law** (for not mutually exclusive):

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

**Independent events**, A and B:

$P(A \cap B) = P(A) \times P(B)$ ;  $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$

**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** =  $\frac{\text{price}}{\text{base year price}} \times 100$

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

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

**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**  $N(\mu, \sigma^2)$

Mean  $(\mu)$  and variance  $(\sigma^2)$

Variance =  $(\text{standard deviation} - \sigma)^2$

**Warning limits** set at  $\mu \pm 2\sigma$

**Action limits** set at  $\mu \pm 3\sigma$

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

**Binomial distribution** =  $B(n, p)$

Binomial expansion =  $(p + q)^n$

Mean of binomial expansion =  $np$