Formula Sheet – Applying to 510

For study purposes only. Not to be taken into exam.

Arithmetic mean is $\bar{x} = \frac{\sum x}{n}$ where n is the number of observations.

Arithmetic mean for a frequency distribution is $\bar{x} = \frac{\sum fx}{\sum f}$ where f is the number of times a given value of x occurs.

For grouped data take the mid-point of each class (x) and proceed as above for the frequency distribution.

The geometric mean is the nth root of the product of the numbers $\bar{x} = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n}$

The median of a set of numbers arranged in order of magnitude is the middle value or the arithmetic mean of the two middle numbers.

The median for grouped data = $L_m + C_m \left(\frac{\frac{n}{2} - F_{m-1}}{f_m} \right)$

L_m = lower boundary of the median class

C_m = class width of the median class

n = total number of observations

F_{m-1} = cumulative frequency of all classes lower than the median class

f_m = frequency of the median class

Mode = mean – 3 (mean-median)

The standard deviation is s = $\sqrt{\frac{\sum (x - \bar{x})^2}{n}}$

For grouped data s = $\sqrt{\frac{\sum f (x - \bar{x})^2}{\sum f}}$ or s = $\sqrt{\frac{\sum fx^2}{\sum f} - \left[\frac{\sum fx}{\sum f}\right]^2}$

The coefficient of variation is $\frac{s}{\bar{x}} \times 100\%$

Pearson's coefficient of skew = $\frac{\text{3 (mean-median)}}{\text{standard deviation}}$

For mutually exclusive events P(A or B) = P(A) + P(B)

For events which are not mutually exclusive P(A or B) = P(A) + P(B) – P(A and B)

Where two independent events occur at the same time P(A and B) = P(A).P(B)

For dependent events

P(A and B) = P(A). P(B/A) where P(B/A) means the probability that B occurs given that A has occurred.

Expectation of x is $\sum P(x).\, x$

Linear regression $y = a + bx$

Using least squares regression, the line of best fit is given by:

$$b = \frac{\sum xy - n\bar{x}\,\bar{y}}{\sum x^2 - n\bar{x}^2}$$

$$a = \frac{\sum y - b\sum x}{n}$$

Correlation is defined as the degree of relationship between variables.

The coefficient of determination, r^2, is the ratio of the explained variation to the total variation and is given by the formula:

$$r^2 = \frac{a\sum y + b\sum xy - n\bar{y}^2}{\sum y^2 - n\bar{y}^2}$$

The coefficient of correlation, r, for linear relationships, is given by the formula:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[n\sum x^2 - (\sum x)^2\right]\left[n\sum y^2 - (\sum y)^2\right]}}$$

or can be taken as the square root of r^2 above.

The standard normal deviate, z, is given by the formula $z = \frac{x - \mu}{\sigma}$

where x is the observed value, μ is the mean and σ is the standard deviation.

Normal Distribution Table

Value of Z	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Area under one tail	0.500	0.460	0.421	0.382	0.345	0.308	0.274	0.242	0.212	0.189	0.159

Value of Z	1.1	1.2	1.3	1.4	1.5	1.6	1.65	1.7	1.8	1.9	1.96	2.0
Area under one tail	0.136	0.115	0.097	0.081	0.067	0.055	0.050	0.045	0.036	0.029	0.025	0.023

Value of Z	2.1	2.2	2.3	2.33	2.4	2.5	2.58	2.7	2.8	2.9	3.0
Area under one tail	0.018	0.014	0.011	0.010	0.008	0.006	0.005	0.0035	0.003	0.002	0.0013

Ref: CII – OLD4 (08/07)