

FORMULAE/REVISION HINTS FOR SECTION L

STATISTICS AND PROBABILITY

Mean, median, mode and standard deviation

If x = variate and f = frequency then: **mean** $\bar{x} = \frac{\sum fx}{\sum f}$

The **median** is the middle term of a ranked set of data.

The **mode** is the most commonly occurring value in a set of data

Standard deviation $\sigma = \sqrt{\left[\frac{\sum \{f(x - \bar{x})^2\}}{\sum f} \right]}$ for a population

Probability

The addition law

The probability of events A **or** B **or** C **or** ... occurring is given by:

$$p_A + p_B + p_C + \dots$$

The multiplication law

The probability of events A **and** B **and** C **and** ... occurring is given by:

$$p_A \times p_B \times p_C \times \dots$$

Binomial probability distribution

If n = number in sample, p = probability of the occurrence of an event and

$q = 1 - p$, then the probability of 0, 1, 2, 3, ... occurrences is given by:

$$q^n, nq^{n-1}p, \frac{n(n-1)}{2!}q^{n-2}p^2, \frac{n(n-1)(n-2)}{3!}q^{n-3}p^3, \dots$$

(i.e. successive terms of the $(q + p)^n$ expansion)

Normal approximation to a binomial distribution:

Mean = np Standard deviation $\sigma = \sqrt{npq}$

Poisson distribution

If λ is the expectation of the occurrence of an event then the probability of 0, 1, 2, 3, ... occurrences is given by:

$$e^{-\lambda}, \lambda e^{-\lambda}, \lambda^2 \frac{e^{-\lambda}}{2!}, \lambda^3 \frac{e^{-\lambda}}{3!}, \dots$$

Product–moment formula for the linear correlation coefficient

$$\text{Coefficient of correlation } r = \frac{\sum xy}{\sqrt{[(\sum x^2)(\sum y^2)]}}$$

where $x = X - \bar{X}$ and $y = Y - \bar{Y}$ and $(X_1, Y_1), (X_2, Y_2), \dots$ denote a random sample from a bivariate normal distribution and \bar{X} and \bar{Y} are the means of the X and Y values, respectively.

Normal probability distribution

Partial areas under the standardized normal curve – see Table 89.1 on page 944.

Student's t distribution

Percentile values (t_p) for Student's t distribution with ν degrees of freedom – see Table 92.1 on page 966.

Chi-square distribution

Percentile values (χ^2_p) for the chi-square distribution with ν degrees of freedom – see Table 94.1 on page 994.

$$\chi^2 = \sum \left\{ \frac{(o-e)^2}{e} \right\} \text{ where } o \text{ and } e \text{ are the observed and expected frequencies.}$$

Symbols

Population: number of members N_p , mean μ , standard deviation σ .

Sample: number of members N , mean \bar{x} , standard deviation s .

Sampling distributions: mean of sampling distribution of means $\mu_{\bar{x}}$, standard error of means $\sigma_{\bar{x}}$, standard error of the standard deviations σ_s .

Standard error of the means

Standard error of the means of a sample distribution, i.e. the standard deviation of the means of samples, is:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \sqrt{\left(\frac{N_p - N}{N_p - 1} \right)} \text{ for a finite population and/or for sampling without replacement, and}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \text{ for an infinite population and/or for sampling with replacement.}$$

The relationship between sample mean and population mean

$\mu_{\bar{x}} = \mu$ for all possible samples of size N which are drawn from a population of size N_p .

Estimating the mean of a population (σ known)

The confidence coefficient for a large sample size ($N \geq 30$) is z_c where:

Confidence level %	Confidence coefficient z_c
99	2.58
98	2.33
96	2.05

95	1.96
90	1.645
80	1.28
50	0.6745

The confidence limits of a population mean based on sample data are given by:

$$\bar{x} \pm \frac{z_c \sigma}{\sqrt{N}} \sqrt{\left(\frac{N_p - N}{N_p - 1} \right)} \text{ for a finite population of size } N_p, \text{ and by}$$

$$\bar{x} \pm \frac{z_c \sigma}{\sqrt{N}} \text{ for an infinite population.}$$

Estimating the mean of a population (σ unknown)

The confidence limits of a population mean based on sample data are given by

$$\mu_x \pm z_c \sigma_x$$

Estimating the standard deviation of a population

The confidence limits of the standard deviation of a population based on sample data are given by:

$$s \pm z_c \sigma_s$$

Estimating the mean of a population based on a small sample size

The confidence coefficient for a small sample size ($N < 30$) is t_c which can be determined using

Table 92.1, page 966. The confidence limits of a population mean based on sample data are given

by:

$$\bar{x} \pm \frac{t_c s}{\sqrt{(N-1)}}$$