RESM2001 – FORMULAE SHEET

Notation

s	sample standard deviation
n	sample size
p	sample proportion
π	population proportion
$\bar{x}$	sample mean
μ	population mean
SE	standard error
a	sample based estimate of population slope α
b	sample based estimate of population slope β

a) **95% confidence interval for a population mean**

$$\bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$$

b) **95% confidence interval for a population proportion**

$$p \pm 1.96 \sqrt{\frac{p(1-p)}{n}}$$

c) **One sample test about a population mean**

The z-test (t-test) statistic is:

$$z = \frac{\bar{x} - \mu_0}{SE(\bar{x})},$$

where μ_0 is the hypothesised value of the population mean under H_0, and $SE(\bar{x}) = \frac{s}{\sqrt{n}}$.

d) **One sample test about a population proportion**

The test statistic is:

$$z = \frac{p - \pi_0}{SE(p)},$$

where π_0 is the hypothesised value of the population proportion under H_0, and $SE(p) = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$.

e) **Two independent samples: difference between two population means**

i) When population variances are assumed not equal, the z-test (t-test) statistic is:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

ii) When the two population variances are assumed to be equal, the z-test (t-test) statistic is:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}},$$

where the pooled variance is $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$.

f) **Two paired samples**

The z-test (t-test) statistic is:

$$z = \frac{\bar{d} - \mu_D}{SE(\bar{d})},$$

where $\bar{d}$ is the mean of the difference, μ_D is the hypothesised mean of the differences under H_0, $SE(\bar{d}\,) = s_D/\sqrt{n}$ is the standard error of the differences, and s_D is the standard deviation of the differences.

g) **Analysis of Variance (ANOVA)**

The test statistics is:

$$F = s_B^2\ /s_W^2,$$

where

$$s_B^2 = \sum_{i=1}^{G} n_i(\bar{y}_i - \bar{y})^2/(G-1)$$

is the between-group variance,

$$s_W^2 = \sum_{i=1}^{G} (n_i - 1)s_i^2\ /(n-G)$$

is the within-group variance and there are G groups.

h) **Two independent samples: difference between two population proportions**

The test statistic is:

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}},$$

where the pooled sample proportion $p = (n_1 p_1 + n_2 p_2) / (n_1 + n_2)$.

i) **Chi-squared test**

The test statistic is:

$$\chi^2 = \sum_{\text{cells}} \frac{(O-E)^2}{E},$$

where O is the observed cell count and E is the expected cell count under H_0.

j) **Regression line $\boldsymbol{y = a + b\,x}$**

$$b = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$a = \bar{y} - b\,\bar{x}$$

k) **95% confidence interval for a population slope**

$$b \pm 1.96\,SE(b),$$

where $SE(b)$ is the standard error of b.

l) **Test of $\boldsymbol{H_0: \beta = 0}$**

The z-test (t-test) statistic is:

$$z = \frac{b}{SE(b)},$$

where $SE(b)$ is the standard error of b.

m) **One-sample/paired two-sample sign test**

The test statistic is:

$$z = \frac{p - 0.5}{\sqrt{0.25/n}},$$

where p is the proportion of positive signs.

n) **Wilcoxon rank-sum (Mann-Whitney) test**

The test statistic is:

$$z = \frac{T}{\sqrt{\frac{n_A n_B (n+1)}{12}}},$$

where $T = W_A - \frac{n_A(n+1)}{2}$ and W_A is the sum of the ranks of the combined sample for Sample A.

o) **Wilcoxon signed-rank test**

The test statistic is:

$$z = \frac{T - \text{mean}}{\text{SE}},$$

where T is the smaller of T_+ and T_-, the sum of the ranks corresponding to the positive and negative differences, respectively, $\text{mean} = \frac{n(n+1)}{4}$ and $\text{SE} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$.

p) **Principal component analysis**

The j^{th} principal component is defined as:

$$Z_j = a_{j1}X_1 + a_{j2}X_2 + \cdots + a_{jp}X_{1p},$$

where $(a_{j1}, a_{j2}, \ldots, a_{jp})$ is the j^{th} (normalised) eigenvector of the correlation or covariance matrix, and the variance of Z_j is given by the j^{th} (largest) eigenvalue, λ_j.