

FREQUENTLY USED FORMULAS

n = sample size N = population size f = frequency

Chapter 2

Class width = $\frac{\text{high} - \text{low}}{\text{number of classes}}$ (increase to next integer)

Class midpoint = $\frac{\text{upper limit} + \text{lower limit}}{2}$

Lower boundary = lower boundary of previous class
+ class width

Chapter 3

Sample mean $\bar{x} = \frac{\sum x}{n}$

Population mean $\mu = \frac{\sum x}{N}$

Weighted average = $\frac{\sum xw}{\sum w}$

Range = largest data value – smallest data value

Sample standard deviation $s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$

Computation formula $s = \sqrt{\frac{\sum x^2 - (\sum x)^2/n}{n - 1}}$

Population standard deviation $\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$

Sample variance s^2

Population variance σ^2

Sample coefficient of variation $CV = \frac{s}{\bar{x}} \cdot 100$

Sample mean for grouped data $\bar{x} = \frac{\sum xf}{n}$

Sample standard deviation for grouped data

$$s = \sqrt{\frac{\sum (x - \bar{x})^2 f}{n - 1}} = \sqrt{\frac{\sum x^2 f - (\sum xf)^2/n}{n - 1}}$$

Chapter 4

Probability of the complement of event A
 $P(A^c) = 1 - P(A)$

Multiplication rule for independent events
 $P(A \text{ and } B) = P(A) \cdot P(B)$

General multiplication rules
 $P(A \text{ and } B) = P(A) \cdot P(B|A)$
 $P(A \text{ and } B) = P(B) \cdot P(A|B)$

Addition rule for mutually exclusive events
 $P(A \text{ or } B) = P(A) + P(B)$

General addition rule
 $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Permutation rule $P_{n,r} = \frac{n!}{(n-r)!}$

Combination rule $C_{n,r} = \frac{n!}{r!(n-r)!}$

Chapter 5

Mean of a discrete probability distribution $\mu = \sum xP(x)$

Standard deviation of a discrete probability distribution

$$\sigma = \sqrt{\sum (x - \mu)^2 P(x)}$$

Given $L = a + bx$

$$\mu_L = a + b\mu$$

$$\sigma_L = |b|\sigma$$

Given $W = ax_1 + bx_2$ (x_1 and x_2 independent)

$$\mu_W = a\mu_1 + b\mu_2$$

$$\sigma_W = \sqrt{a^2\sigma_1^2 + b^2\sigma_2^2}$$

For Binomial Distributions

r = number of successes; p = probability of success;

$$q = 1 - p$$

Binomial probability distribution $P(r) = C_{n,r} p^r q^{n-r}$

Mean $\mu = np$

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

Geometric Probability Distribution

n = number of trial on which first success occurs

$$P(n) = p(1 - p)^{n-1}$$

Poisson Probability Distribution

r = number of successes

λ = mean number of successes over given interval

$$P(r) = \frac{e^{-\lambda} \lambda^r}{r!}$$

Chapter 6

Raw score $x = z\sigma + \mu$ Standard score $z = \frac{x - \mu}{\sigma}$

Mean of \bar{x} distribution $\mu_{\bar{x}} = \mu$

Standard deviation of \bar{x} distribution $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Standard score for \bar{x} $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Mean of \hat{p} distribution $\mu_{\hat{p}} = p$

Standard deviation of \hat{p} distribution $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$; $q = 1 - p$

Chapter 7

Confidence Interval

for μ

$$\bar{x} - E < \mu < \bar{x} + E$$

where $E = z_c \frac{\sigma}{\sqrt{n}}$ when σ is known

$$E = t_c \frac{s}{\sqrt{n}} \text{ when } \sigma \text{ is unknown}$$

with $d.f. = n - 1$

for p ($np > 5$ and $n(1 - p) > 5$)

$$\hat{p} - E < p < \hat{p} + E$$

$$\text{where } E = z_c \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

$$\hat{p} = \frac{r}{n}$$

for $\mu_1 - \mu_2$ (independent samples)

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$$

where $E = z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ when σ_1 and σ_2 are known

$$E = t_c \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \text{ when } \sigma_1 \text{ or } \sigma_2 \text{ is unknown}$$

with $d.f. = \text{smaller of } n_1 - 1 \text{ and } n_2 - 1$

(Note: Software uses Satterthwaite's approximation for degrees of freedom $d.f.$)

for difference of proportions $p_1 - p_2$

$$(\hat{p}_1 - \hat{p}_2) - E < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + E$$

$$\text{where } E = z_c \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

$$\hat{p}_1 = r_1/n_1; \hat{p}_2 = r_2/n_2$$

$$\hat{q}_1 = 1 - \hat{p}_1; \hat{q}_2 = 1 - \hat{p}_2$$

Sample Size for Estimating

$$\text{means } n = \left(\frac{z_c \sigma}{E} \right)^2$$

proportions

$$n = p(1 - p) \left(\frac{z_c}{E} \right)^2 \text{ with preliminary estimate for } p$$

$$n = \frac{1}{4} \left(\frac{z_c}{E} \right)^2 \text{ without preliminary estimate for } p$$

Chapter 8

Sample Test Statistics for Tests of Hypotheses

$$\text{for } \mu \text{ (}\sigma \text{ known)} \quad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$\text{for } \mu \text{ (}\sigma \text{ unknown)} \quad t = \frac{\bar{x} - \mu}{s/\sqrt{n}}; d.f. = n - 1$$

$$\text{for } p \text{ (} np > 5 \text{ and } nq > 5) \quad z = \frac{\hat{p} - p}{\sqrt{pq/n}}$$

$$\text{where } q = 1 - p; \hat{p} = r/n$$

$$\text{for paired differences } d \quad t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}; d.f. = n - 1$$

for difference of means, σ_1 and σ_2 known

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

for difference of means, σ_1 or σ_2 unknown

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

$d.f. = \text{smaller of } n_1 - 1 \text{ and } n_2 - 1$

(Note: Software uses Satterthwaite's approximation for degrees of freedom $d.f.$)

for difference of proportions

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\bar{p} \bar{q}}{n_1} + \frac{\bar{p} \bar{q}}{n_2}}}$$

$$\text{where } \bar{p} = \frac{r_1 + r_2}{n_1 + n_2} \text{ and } \bar{q} = 1 - \bar{p}$$

$$\hat{p}_1 = r_1/n_1; \hat{p}_2 = r_2/n_2$$

Chapter 9

Regression and Correlation

Pearson product-moment correlation coefficient

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

Least-squares line $\hat{y} = a + bx$

$$\text{where } b = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$

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

Coefficient of determination = r^2

Sample test statistic for r

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \text{ with } d.f. = n - 2$$

$$\text{Standard error of estimate } S_e = \sqrt{\frac{\sum y^2 - a \sum y - b \sum xy}{n - 2}}$$

Confidence interval for y

$$\hat{y} - E < y < \hat{y} + E$$

$$\text{where } E = t_c S_e \sqrt{1 + \frac{1}{n} + \frac{n(x - \bar{x})^2}{n \sum x^2 - (\sum x)^2}}$$

with $d.f. = n - 2$

Sample test statistic for slope b

$$t = \frac{b}{S_e} \sqrt{\sum x^2 - \frac{1}{n} (\sum x)^2} \text{ with } d.f. = n - 2$$

Confidence interval for β

$$b - E < \beta < b + E$$

$$\text{where } E = \frac{t_e S_e}{\sqrt{\sum x^2 - \frac{1}{n} (\sum x)^2}} \text{ with } d.f. = n - 2$$

Chapter 10

$$\chi^2 = \sum \frac{(O - E)^2}{E} \text{ where}$$

O = observed frequency and

E = expected frequency

For tests of independence and tests of homogeneity

$$E = \frac{(\text{row total})(\text{column total})}{\text{sample size}}$$

For goodness of fit test E = (given percent)(sample size)

Tests of independence $d.f. = (R - 1)(C - 1)$

Test of homogeneity $d.f. = (R - 1)(C - 1)$

Goodness of fit $d.f. = (\text{number of categories}) - 1$

Confidence interval for σ^2 ; $d.f. = n - 1$

$$\frac{(n - 1)s^2}{\chi^2_U} < \sigma^2 < \frac{(n - 1)s^2}{\chi^2_L}$$

Sample test statistic for σ^2

$$\chi^2 = \frac{(n - 1)s^2}{\sigma^2} \text{ with } d.f. = n - 1$$

Testing Two Variances

$$\text{Sample test statistic } F = \frac{s_1^2}{s_2^2}$$

where $s_1^2 \geq s_2^2$

$$d.f._N = n_1 - 1; d.f._D = n_2 - 1$$

ANOVA

k = number of groups; N = total sample size

$$SS_{TOT} = \sum x_{TOT}^2 - \frac{(\sum x_{TOT})^2}{N}$$

$$SS_{BET} = \sum_{all\ groups} \left(\frac{(\sum x_i)^2}{n_i} \right) - \frac{(\sum x_{TOT})^2}{N}$$

$$SS_W = \sum_{all\ groups} \left(\sum x_i^2 - \frac{(\sum x_i)^2}{n_i} \right)$$

$$SS_{TOT} = SS_{BET} + SS_W$$

$$MS_{BET} = \frac{SS_{BET}}{d.f._{BET}} \text{ where } d.f._{BET} = k - 1$$

$$MS_W = \frac{SS_W}{d.f._W} \text{ where } d.f._W = N - k$$

$$F = \frac{MS_{BET}}{MS_W} \text{ where } d.f. \text{ numerator} = d.f._{BET} = k - 1;$$

$$d.f. \text{ denominator} = d.f._W = N - k$$

Two-Way ANOVA

r = number of rows; c = number of columns

$$\text{Row factor } F: \frac{MS \text{ row factor}}{MS \text{ error}}$$

$$\text{Column factor } F: \frac{MS \text{ column factor}}{MS \text{ error}}$$

$$\text{Interaction } F: \frac{MS \text{ interaction}}{MS \text{ error}}$$

with degrees of freedom for

$$\text{row factor} = r - 1$$

$$\text{interaction} = (r - 1)(c - 1)$$

$$\text{column factor} = c - 1$$

$$\text{error} = rc(n - 1)$$

Chapter 11

Sample test statistic for x = proportion of plus signs to all signs ($n \geq 12$)

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

Sample test statistic for R = sum of ranks

$$z = \frac{R - \mu_R}{\sigma_R} \text{ where } \mu_R = \frac{n_1(n_1 + n_2 + 1)}{2} \text{ and}$$

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Spearman rank correlation coefficient

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

Sample test statistic for runs test

R = number of runs in sequence