FREQUENTLY USED FORMULAS

n = sample size N = population size f = frequency

Class Width $= \dfrac{\text{high} - \text{low}}{\text{number classes}}$ (increase to next integer)

Class Midpoint $= \dfrac{\text{upper limit} + \text{lower limit}}{2}$

Lower boundary = lower boundary of previous class + class width

Sample mean $\bar{x} = \dfrac{\Sigma x}{n}$

Population mean $\mu = \dfrac{\Sigma x}{N}$

Weighted average $= \dfrac{\Sigma xw}{\Sigma w}$

Range = largest data value − smallest data value

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

Computation formula $s = \sqrt{\dfrac{SS_x}{n - 1}}$ where $SS_x = \Sigma x^2 - \dfrac{(\Sigma x)^2}{n}$

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

Sample variance s^2

Population variance σ^2

Sample Coefficient of Variation $CV = \dfrac{s}{\bar{x}} \cdot 100$

Sample mean for grouped data $\bar{x} = \dfrac{\Sigma xf}{n}$

Sample standard deviation for grouped data

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

Regression and Correlation

In all these formulas

$$SS_x = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

$$SS_y = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$$

$$SS_{xy} = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}$$

Least squares line $y = a + bx$ where $b = \dfrac{SS_{xy}}{SS_x}$ and $a = \bar{y} - b\bar{x}$

Pearson product-moment correlation coefficient

$$r = \frac{SS_{xy}}{\sqrt{SS_x SS_y}}$$

Coefficient of determination $= r^2$

Probability of the complement of event A
$P(not\ A) = 1 - P(A)$

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

General multiplication rules
$P(A\ and\ B) = P(A) \cdot P(B, given\ A)$
$P(A\ and\ B) = P(B) \cdot P(A, given\ B)$

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

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

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

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

Mean of a discrete probability distribution $\mu = \Sigma x P(x)$

Standard deviation of a discrete probability distribution

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

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 $\sigma = \sqrt{npq}$

Raw score $x = z\sigma + \mu$

Standard score $z = \dfrac{x - \mu}{\sigma}$

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

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

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

Confidence Interval

for $\mu (n \geq 30)$

$$\bar{x} - z_c \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_c \frac{\sigma}{\sqrt{n}}$$

for μ $(n < 30)$

$d.f. = n - 1$

$$\bar{x} - t_c \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_c \frac{s}{\sqrt{n}}$$

for p $(np > 5$ and $nq > 5)$

$$\hat{p} - z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $\hat{p} = r/n$

Sample Size n for Estimating

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

proportions

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

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

Sample Test Statistics for Tests of Hypotheses

for μ $(n \geq 30)$ $\quad z = \dfrac{\bar{x} - \mu}{\sigma\sqrt{n}}$

for μ $(n < 30)$; $d.f. = n - 1$ $\quad t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

for p $\quad z = \dfrac{\hat{p} - p}{\sqrt{pq/n}}$ where $q = 1 - p$

Sample Test Statistics for Tests of Hypotheses

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

difference of means large sample

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

difference of means small sample with $\sigma_1 \approx \sigma_2$; $d.f. = n_1 + n_2 - 2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

where $s = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$

difference of proportions

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\hat{p}\hat{q}}{n_1} + \dfrac{\hat{p}\hat{q}}{n_2}}} \text{ where } \hat{p} = \frac{r_1 + r_2}{n_1 + n_2};\ \hat{q} = 1 - \hat{p};$$

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

Confidence Intervals

for difference of means (when $n_1 \geq 30$ and $n_2 \geq 30$)

$$(\bar{x}_1 - \bar{x}_2) - z_c\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_c\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

for difference of means ($n_1 < 30$ and/or $n_2 < 30$ and $\sigma_1 \approx \sigma_2$) $d.f. = n_1 + n_2 - 2$

$$(\bar{x}_1 - \bar{x}_2) - t_c s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_c s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where $s = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$

for difference of proportions

where $\hat{p}_1 = r_1/n_1;\ \hat{p}_2 = r_2/n_2;\ \hat{q}_1 = 1 - \hat{p}_1;\ \hat{q}_2 = 1 - \hat{p}_2$

$$(\hat{p}_1 - \hat{p}_2) - z_c\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_c\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$\chi^2 = \sum\frac{(O - E)^2}{E} \text{ where } E = \frac{(\text{row total})(\text{column total})}{\text{sample size}}$$

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

Goodness of fit $d.f. =$ (number of entries) $- 1$

Sample test statistic for H_0: $\sigma^2 = k$; $d.f. = n - 1$

$$\chi^2 = \frac{(n - 1)s^2}{\sigma^2}$$

Linear Regression

Standard error of estimate $S_e = \sqrt{\dfrac{SS_y - bSS_{xy}}{n - 2}}$

where $b = \dfrac{SS_{xy}}{SS_x}$

Confidence interval for y

$y_p - E < y < y_p + E$ where y_p is the predicted y value for x and

$$E = t_c S_e\sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{SS_x}} \text{ with } d.f. = n - 2$$