# STATISTICS FORMULAS

$$\overline{x} = \frac{\sum x_i}{n}$$

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{J} x_j f_j \text{ and } n = \sum_{j=1}^{J} n_j$$

$$R = H - L$$

$$SS_{xx} = \sum (x_i - \overline{x})^2 = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}$$

$$SS_{xx} = \sum_{j=1}^{J} x_j^2 f_j - \frac{1}{n}\left(\sum_{j=1}^{J} x_j f_j\right)^2$$

$$s^2 = \frac{SS_{xx}}{n-1} = \frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}$$

$$s = \sqrt{s^2} = \sqrt{\frac{SS_{xx}}{n-1}} = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}}$$

$$z_x = \frac{x - \mu}{\sigma} \qquad x_z = \sigma z + \mu$$

$$z = \frac{x - \overline{x}}{s} \qquad x = s z + \overline{x}$$

## Quantiles

$$ML = \frac{n+1}{2}$$

$$PL = p(n+1), \quad 0 < p < 1$$

$$x_p = x_{(k)} + (PL - k)(x_{(k+1)} - x_{(k)});$$

*Where* $x_p$ = 100*pth percentile*, $k = [PL]$, *and*

$x_{(k)}$ = *kth value of x in order of magnitude*.

$$IQR = x_{75} - x_{25} = Q_3 - Q_1$$

## Probability

$$P(A') = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cap B) = P(A)\,P(B|A)$$

$$P(A \cap B) = P(B)\,P(A|B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ where } P(B) \neq 0$$

**Finite Population**

$$\mu = E(x) = \sum_x x\, p(x)$$

$$\sigma^2 = E[(x-\mu)^2] = \sum_x x^2\, p(x) - \mu^2$$

**Binomial Distribution**

$$p(x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \ldots, n$$

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

$\mu = np \qquad \sigma^2 = npq$

## Sampling Distributions of Several Sample Statistics

| *Statistic (Estimator)* | *Mean of Estimator* | *Standard Deviation of Estimator* | *z-score of Estimator* |
|---|---|---|---|
| $\bar{x}$ | $\mu$ | $\frac{\sigma}{\sqrt{n}}$ | $\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ |
| $\hat{p} = \frac{x}{n}$ | $p$ | $\sqrt{\frac{pq}{n}}$ | $\frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$ |
| $\bar{x}_1 - \bar{x}_2$ | $\mu_1 - \mu_2$ | $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ | $\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ |
| $\hat{p}_1 - \hat{p}_2$ | $p_1 - p_2$ | $\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$ | $\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$ |

**Test Statistics**

$H_o$: $\mu = \mu_o$ $\qquad z = \dfrac{\bar{x}-\mu_o}{\frac{\sigma}{\sqrt{n}}}$

$H_o$: $\mu = \mu_o$ $\qquad t = \dfrac{\bar{x}-\mu_o}{\frac{s}{\sqrt{n}}}, \quad df = n-1$

$H_o$: $p = p_o$ $\qquad z = \dfrac{\hat{p}-p_o}{\sqrt{\frac{p_o q_o}{n}}}, \quad \hat{p} = \dfrac{x}{n}$

$$\chi^2_{(r-1)(c-1)} = \sum \frac{(O-E)^2}{E}$$

$r$ = # of rows, $c$ = # of columns, $O$ = # observed, $E$ = # expecte

$H_o$: $p_1 - p_2 = (p_1 - p_2)_o$

Case 1: $(p_1 - p_2)_o \neq 0$ $\quad z = \dfrac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)_o}{\sqrt{\dfrac{\hat{p}_1 \hat{q}_1}{n_1} + \dfrac{\hat{p}_2 \hat{q}_2}{n_2}}}$

Case 2: $(p_1 - p_2)_o = 0$ $\quad z = \dfrac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$, $\quad \hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$, $\quad \hat{q} = 1$ -

$H_o: \sigma^2{}_1 = \sigma^2{}_2$ $\quad F = \dfrac{s^2_{\max}}{s^2_{\min}}$

$\nu_1 = df_{numerator} = n_{\max} - 1$, $\quad \nu_2 = df_{denominator} = n_{\min} - 1$

$RR$: $F \geq F_{\frac{\alpha}{2}, \nu_1, \nu_2}$

$H_o$: $\mu_1 - \mu_2 = (\mu_1 - \mu_2)_o$ $\quad t = \dfrac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)_o}{\sqrt{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$, $\quad df = n_1 + n_2 - 2$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$H_o$: $\mu_1 - \mu_2 = (\mu_1 - \mu_2)_o$ $\quad t = \dfrac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)_o}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$,

$\min(n_1 - 1, n_2 - 1) \leq df' \leq n_1 + n_2 - 2$

The number of degrees of freedom for this test statistic is determined by a computer algorithm, and is denoted by df'.

$H_o$: $\mu_D = \mu_{D_o}$ $\quad t = \dfrac{\bar{x}_D - \mu_{D_o}}{\dfrac{s_D}{\sqrt{n_D}}}$, $\quad df = n_D - 1$

## Power for Tests about $\mu$

$$H_a:\ \mu > \mu_o \quad RR:\ z > z_{\alpha} \quad or \quad \bar{x} > z_{\alpha}\frac{\sigma}{\sqrt{n}} + \mu_o = \bar{x}_{CV}$$

$$Power = P\left(\bar{x} > z_{\alpha}\frac{\sigma}{\sqrt{n}} + \mu_o \middle| \mu_a\right) = P\left(z > z_{\alpha} - (\mu_a - \mu_o)\frac{\sqrt{n}}{\sigma}\right)$$

$$H_a:\ \mu < \mu_o \quad RR:\ z < -z_{\alpha} \quad or \quad \bar{x} < -z_{\alpha}\frac{\sigma}{\sqrt{n}} + \mu_o = \bar{x}_{CV}$$

$$Power = P\left(\bar{x} < -z_{\alpha}\frac{\sigma}{\sqrt{n}} + \mu_o \middle| \mu_a\right) = P\left(z < -z_{\alpha} - (\mu_a - \mu_o)\frac{\sqrt{n}}{\sigma}\right)$$

$$H_a:\ \mu \neq \mu_o \quad RR:\ z < -z_{\frac{\alpha}{2}} \quad or \quad z > z_{\frac{\alpha}{2}}$$

$$or \quad \bar{x} < -z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} + \mu_o = \bar{x}_{LCV} \quad or \quad \bar{x} > z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} + \mu_a = \bar{x}_{UCV}$$

$$Power = P\left(\bar{x} < -z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} + \mu_o \middle| \mu_a\right) + P\left(\bar{x} > z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} + \mu_o \middle| \mu_a\right) =$$

$$P\left(z < -z_{\frac{\alpha}{2}} - (\mu_a - \mu_o)\frac{\sqrt{n}}{\sigma}\right) + P\left(z > z_{\frac{\alpha}{2}} - (\mu_a - \mu_o)\frac{\sqrt{n}}{\sigma}\right)$$

$$d = \frac{|\mu_a - \mu_o|}{\sigma} \qquad \delta = d\sqrt{n}$$