Population Expected Value

$$E(Y) = y_1 p_1 + y_2 p_2 + \cdots + y_k p_k = \sum_{i=1}^{k} y_i p_i,$$

Population Variance

$$\sigma_Y^2 = \text{var}(Y) = E[(Y - \mu_Y)^2] = \sum_{i=1}^{k} (y_i - \mu_Y)^2 p_i.$$

Marginal probability distribution

$$\Pr(Y = y) = \sum_{i=1}^{l} \Pr(X = x_i,\ Y = y).$$

Conditional probability distribution

$$\Pr(Y = y | X = x) = \frac{\Pr(X = x,\ Y = y)}{\Pr(X = x)}.$$

Conditional Expected Value

$$E(Y | X = x) = \sum_{i=1}^{k} y_i \Pr(Y = y_i | X = x).$$

Law of iterated expectations

$$\mathrm{E(Y)} = \sum_{\mathrm{i}=1}^{l} \mathrm{E(Y \setminus X = x_i) Pr(X = x_i)}.$$

Population covariance

$$\begin{aligned} \text{cov}(X, Y) &= \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] \\ &= \sum_{i=1}^{k} \sum_{j=1}^{l} (x_j - \mu_X)(y_i - \mu_Y) \Pr(X = x_j, Y = y_i). \end{aligned}$$

Population correlation

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\ \text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Sample mean

$$\overline{Y} = \frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n) = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Expected value, population variance and std. deviation of the sample mean

$$E(\overline{Y}) = \mu_Y.$$

$$\operatorname{var}(\overline{Y}) = \sigma^2_{\overline{Y}} = \frac{\sigma^2_Y}{n}, \text{ and}$$

$$\text{std.dev}(\overline{Y}) = \sigma_{\overline{Y}} = \frac{\sigma_Y}{\sqrt{n}}.$$

p-value when $\sigma_{\bar{Y}}$ is known:

$$p\text{-value} = \Pr_{H_0}\left(\left| \frac{\overline{Y} - \mu_{Y,0}}{\sigma_{\overline{Y}}} \right| > \left| \frac{\overline{Y}^{act} - \mu_{Y,0}}{\sigma_{\overline{Y}}} \right| \right) = 2\Phi\left(-\left| \frac{\overline{Y}^{act} - \mu_{Y,0}}{\sigma_{\overline{Y}}} \right| \right)$$

Sample variance of Y

$$s^2_Y = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

Standard error of $\bar{Y}$

$$SE(\overline{Y}) = \hat{\sigma}_{\overline{Y}} = s_Y / \sqrt{n}.$$

p-value when $\sigma_{\bar{Y}}$ is unknown:

$$p\text{-value} = 2\Phi\left(-\left| \frac{\overline{Y}^{act} - \mu_{Y,0}}{SE(\overline{Y})} \right| \right).$$

t-statistic and p-value to test a hypothesis ($\mu_{Y,0}$ is the value on the null hypothesis that the researcher is testing against, act simply means the values are calculated from the actual sample).

$$t^{act} = \frac{\overline{Y}^{act} - \mu_{Y,0}}{SE(\overline{Y})}. \quad p\text{-value} = 2\Phi(-|t^{act}|).$$

Confidence intervals

95% confidence interval for $\mu_Y = \{\overline{Y} \pm 1.96SE(\overline{Y})\}$.

90% confidence interval for $\mu_Y = \{\overline{Y} \pm 1.64SE(\overline{Y})\}$.

99% confidence interval for $\mu_Y = \{\overline{Y} \pm 2.58SE(\overline{Y})\}$.

Standard error of the differences in means:

$$SE(\overline{Y}_m - \overline{Y}_w) = \sqrt{\frac{s_m^2}{n_m} + \frac{s_w^2}{n_w}}.$$

t-statistic when comparing two means:

$$t = \frac{(\overline{Y}_m - \overline{Y}_w) - d_0}{SE(\overline{Y}_m - \overline{Y}_w)}$$

95% confidence interval for the difference in means

$$(\overline{Y}_m - \overline{Y}_w) \pm 1.96SE(\overline{Y}_m - \overline{Y}_w).$$

Sample covariance

$$s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

Sample correlation

$$r_{XY} = \frac{s_{XY}}{s_X s_Y}.$$

Estimated OLS coefficients

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{s_{XY}}{s_X^2}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}.$$

Explained and total sum of squares:

$$ESS = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2$$

$$TSS = \sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

R-squared and the standard error of the regression

$$R^2 = \frac{ESS}{TSS}. \quad R^2 = 1 - \frac{SSR}{TSS}.$$

$$SER = s_{\hat{u}}, \text{ where } s_{\hat{u}}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2 = \frac{SSR}{n-2},$$