Probability and Statistics Formulas

Descriptive Statistics

- Sample Variance

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

Probability

- Conditional probability:

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}.$$

- Bayes' Theorem:

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A)}.$$

- Binomial probability distribution:

$$P(X = k) = ^n C_k \times p^k \times (1-p)^{n-k} \qquad \left(\text{where } ^nC_k = \frac{n!}{k!\,(n-k)!}.\right)$$

- Poisson probability distribution:

$$P(X = k) = \frac{m^k \mathrm{e}^{-m}}{k!}.$$

- Exponential probability distribution:

$$P(X \leq k) = \begin{cases} 1 - e^{-k/\mu}, & k \geq 0, \\ 0, & k < 0. \end{cases} \qquad \left(\text{where } \mu = \frac{1}{\lambda}\right)$$

Confidence Intervals

One sample

$$S.E.(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

$$S.E.(\hat{P}) = \sqrt{\frac{\hat{p} \times (1-\hat{p})}{n}}.$$

Two samples

$$S.E.(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

$$S.E.(\hat{P}_1 - \hat{P}_2) = \sqrt{\frac{\hat{p}_1 \times (1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2 \times (100-\hat{p}_2)}{n_2}}.$$

Hypothesis tests

One sample

$$S.E.(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

$$S.E.(p) = \sqrt{\frac{p \times (1-p)}{n}}$$

Two large independent samples

$$S.E.(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

$$S.E.(\hat{P}_1 - \hat{P}_2) = \sqrt{(\bar{p} \times (1-\bar{p}))\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}.$$

Two small independent samples

$$S.E.(\bar{X}_1 - \bar{X}_2) = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}.$$

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

Paired sample

$$S.E.(\bar{d}) = \frac{s_d}{\sqrt{n}}.$$

Standard Deviation of case-wise differences (computational formula)

$$s_d = \sqrt{\frac{\sum d_i^2 - n\bar{d}^2}{n-1}}.$$

Chi Square Tests of Independence

$$\chi^2_{TS} = \sum \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Regression Estimates

$$
\begin{aligned}
S_{XY} &= \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} \\
S_{XX} &= \sum x_i^2 - \frac{(\sum x_i)^2}{n} \\
S_{YY} &= \sum y_i^2 - \frac{(\sum y_i)^2}{n}
\end{aligned}
$$

Slope Estimate

$$b_1 = \frac{S_{XY}}{S_{XX}}$$

Intercept Estimate

$$b_0 = \bar{y} - b_1 \bar{x}$$

Pearson's correlation coefficient

$$r = \frac{S_{XY}}{\sqrt{S_{XX} \times S_{YY}}}$$

Standard error of the Slope

$$S.E.(b1) = \sqrt{\frac{s^2}{S_{XX}}}$$

where $s^2 = \frac{SSE}{n-2}$ and SSE $= S_{YY} - b_1 S_{XY}$