Formulae

Basic Statistics: For quantitative data values, denoted y, and sample size n, the sample mean $\overline{y}$ and the sample standard deviation, s, can be found as follows.

$$\begin{aligned}\overline{y} &= \frac{\Sigma y}{n} \\ s &= \sqrt{\frac{\Sigma(y-\overline{y})^2}{n-1}}\end{aligned}$$

Z-scores: For a distribution with mean μ and standard deviation, σ,

$$z = \frac{y-\mu}{\sigma}.$$

Correlation: For a scatterplot of y vs x that meets the correlation conditions,

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

Linear Model: The Linear Model for a scatterplot is $\hat{y} = b_0 + b_1 x$.

$$\begin{aligned}b_1 &= \frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2} \\ b_0 &= \overline{y} - b_1\overline{x}\end{aligned}$$

Probabilities: Let A and B be events.

$$\begin{aligned}P(A \text{ or } B) &= P(A) + P(B) - P(A \text{ and } B) \\ P(A \text{ and } B) &= P(A \mid B)P(B)\end{aligned}$$

Random Variables: Let X be a random variable.

$$\begin{aligned} E(X) &= \Sigma(xP(X = x)) \\ Var(X) &= \Sigma((x - E(X))^2 P(X = x)) \\ SD(X) &= \sqrt{Var(X)} \end{aligned}$$

Let a be a real number.

$$\begin{aligned} E(X + a) &= E(X) + a \\ Var(X + a) &= Var(X) \\ SD(X + a) &= SD(X) \\ E(aX) &= aE(X) \\ Var(aX) &= a^2 Var(X) \\ SD(aX) &= |a| SD(X) \end{aligned}$$

Let X and Y be independent random variables.

$$\begin{aligned} E(X \pm Y) &= E(X) \pm E(Y) \\ Var(X \pm Y) &= Var(X) + Var(Y) \\ SD(X \pm Y) &= \sqrt{(SD(X))^2 + (SD(Y))^2} \end{aligned}$$

The Binomial Model: Let p be the probability of success for a set of n Bernoulli trials and let X be the number successes during the n trials.

$$\begin{aligned} P(X = x) &= \binom{n}{x} p^x (1 - p)^{n-x} \\ E(X) &= np \\ SD(X) &= \sqrt{np(1 - p)} \end{aligned}$$

The Poisson Model: Let $\lambda = np$. $Poisson(\lambda)$ is a good approximation of $Binom(n, p)$ if $n \geq 20$ and $p \leq 0.05$ or if $n \geq 100$ and $p \leq 0.10$.

$$\begin{aligned} P(X = x) &= \frac{e^{-\lambda} \lambda^x}{x!} \\ E(X) &= \lambda \\ SD(X) &= \sqrt{\lambda} \end{aligned}$$

The Sampling Distribution Model for a Proportion: Provided that the sampled values are independent and the sample size is large enough, the sampling distribution of $\hat{p}$ is modelled by a Normal model with mean p and standard deviation $SD(\hat{p}) = \sqrt{\frac{pq}{n}}$.

The Sampling Distribution Model for a Mean: When a simple random sample is drawn from any population with mean μ and standard deviation σ, its sample mean, $\bar{y}$ has a sampling distribution with the same mean μ but whose standard deviation is $\frac{\sigma}{\sqrt{n}}$.

One-Proportion Z-Interval: When the conditions are met, we are ready to find a level C confidence interval for the population proportion, p. The confidence interval is $\hat{p} \pm z^* \times SE(\hat{p})$, where the standard deviation of the proportion is estimated by $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$ and the critical value, z^*, specifies the number of SEs needed for $C\%$ of random samples to yield confidence intervals that capture the true parameter value.

One-Proportion Z-Test: When the necessary assumptions and conditions have been met we can perform the one-proportion z-test. We test the hypothesis $H_0 : p = p_0$ using the statistic $z = \frac{\hat{p}-p_0}{SD(\hat{p})}$. We use the hypothesised proportion to find the standard deviation, $SD(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}}$.

Two-Proportion Z-Interval: When the necessary assumptions and conditions have been met we can construct the confidence interval for the difference of two proportions, $p_1 - p_2$. The confidence interval is

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \times SE(\hat{p}_1 - \hat{p}_2$$

where the standard error of the difference,

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

from the observed proportions. The critical value z^* depends on the particular confidence level, C, that we specify.

Two-Proportion Z-Interval: When the necessary assumptions and conditions have been met we can perform the two-proportion z-test. We test the hypothesis $H_0 : p_1 - p_2 = 0$ Because we hypothesize that the proportions are equal, we pool the groups to find

$$\hat{p}_{pooled} = \frac{\text{Success}_1 + \text{Success}_2}{n_1 + n_2}$$

and use that pooled value to estimate the standard error:

$$SE_{pooled}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_1} + \frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_2}}$$

Now we find the test statistic,

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{SE_{pooled}(\hat{p}_1 - \hat{p}_2)}$$

One-Sample t-Interval for the Mean: When the necessary assumptions and conditions have been met, we are ready to find the confidence interval for the population mean, μ. The confidence interval is

$$\bar{y} \pm t^*_{n-1} \times SE(\bar{y})$$

where the standard error of the mean

$$SE(\bar{y}) = \frac{s}{\sqrt{n}}$$

The critical value t^*_{n-1} depends on the particular confidence level, C, that you specify and on the number of degrees of freedom, $n - 1$, which we get from the sample size.

One-Sample t-Test for the Mean: When the necessary assumptions and conditions have been met, we are ready to perfom that one-sample t-test for the mean. We test the hypothesis $H_0 : \mu = \mu_0$ using the statistic

$$t = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$$

The standard error of $\bar{y}$ is

$$SE(\bar{y}) = \frac{s}{\sqrt{n}}$$