Formula Sheet

Population mean formula. $\mu = \dfrac{\sum x_i}{N}$ where the summation is taken over all data points in the population, and N is the population size.

Population variance formula. $\sigma^2 = \dfrac{\sum (x_i - \mu)^2}{N}$ where the summation is taken over all data points in the population, and N is the population size.

Population standard deviation formula. $\sigma = \sqrt{\sigma^2}$.

Sample mean formula. $\overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$, where n is the sample size.

Sample variance formula. $s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$, where n is the sample size.

Sample standard deviation formula. $s = \sqrt{s^2}$.

Normal distrubution/Bell curve

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-1/2[(x-\mu)/\sigma]^2)$$

---

z_α formulas.

- $P(z \geq z_\alpha) = \alpha$.
- For any $0 \leq \alpha \leq 1$ we have $-z_\alpha = z_{1-\alpha}$.

Hypothesis test for a population mean, change of variable to t.

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}.$$

If x is approximately a normal variable, or if $n > 30$, this variable t follows Student's t-distribution with $df = n - 1$.

Hypothesis test for two population proportions, independent samples.
If $n\hat{p}_1(1-\hat{p}_1) \geq 10$ and $n\hat{p}_2(1-\hat{p}_2) \geq 10$, then the variable $\hat{p}_1 - \hat{p}_2$ has an approximately normal distribution. We may convert $\hat{p}_1 - \hat{p}_2$ to a standard normal variable via

$$z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$.

Hypothesis test for two population means, matched pair data.

$d_i = x_i - y_i$

If d is approximately normally distributed, or if $n > 30$, then the new variable

$$t = \frac{\overline{d} - \mu_d}{s_d/\sqrt{n}}$$

follows Student's t-distribution with $df = n - 1$.

Hypothesis test for two population means, independent samples.

If the two populations are both normally distributed, or if both n_1 and n_2 are > 30, then the new variable

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

approximately follows Student's t-distribution with the smaller of $n_1 - 1$ or $n_2 - 1$ degrees of freedom.