Distribution of the Sample Mean

If a random sample from a population Y with mean μ_Y and standard deviation σ_Y is taken and either

(i): The population is normally distributed, or

(ii): $n \geq 30$

then the sample mean $\bar{Y}$ is normally distributed with mean μ_Y, standard deviation $\frac{\sigma_Y}{\sqrt{n}}$ (I.e., $\bar{Y} \sim N(\mu_Y, \sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}})$).

Confidence Interval for μ

- A (1-α)100% CI for μ is:
  $\bar{y} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$
  at d.f. $= n-1$
- To calculate what sample size you should take for a margin of error within e:
  $n = \frac{t_{\alpha/2}^2 s^2}{e^2}$
  where we use d.f. $= \infty$ for $t_{\alpha/2}$.

Confidence Interval for $\mu_1 - \mu_2$

- A (1- α)100% CI for $\mu_1 - \mu_2$ is:
  $(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
  at d.f. $= \nu$ (this will be given).

Hypothesis Test for $\mu_1 - \mu_2$

- Step 1: State the null, and alternative.
- Step 2: The test-statistic is: $t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
  at d.f. $= \nu$ (this will be given)
- Step 3: The possible p-values are:
  p-value $= 2Pr\{t > |t_s|\}$
  p-value $= Pr\{t < t_s\}$
  p-value $= Pr\{t > t_s\}$

General Definitions

- p-value: The probability of observing our sample data or more extreme, if the null hypothesis is true.
- Type I error: When we reject the null, if in reality the null is true.
- Type II error: When we fail to reject the null, if in reality the null is false.
- Step 4: Decision Rule (for any hypothesis test):
  If p-value $< \alpha$, reject H_0.
  If p-value $\geq \alpha$, fail to reject H_0.

Definitions for ANOVA

- n_i = sample size for group i
  $n_\bullet$ = overall sample size $= \sum_{i=1}^{I} n_i$
- $\bar{y}_i$ = sample mean for group i
  $\bar{\bar{y}} = \frac{\sum_{i=1}^{I} n_i \bar{y}_i}{n_\bullet}$ = overall group mean
  s_i = sample standard deviation for group i
  I = total number of groups
- $SSB = \sum_{i=1}^{I} n_i(\bar{y}_i - \bar{\bar{y}})^2$
  $d.f.\{B\} = I - 1$
  $MSB = \frac{SSB}{d.f.\{B\}}$
- $SSW = \sum_{i=1}^{I}(n_i - 1)s_i^2$
  $d.f.\{W\} = n_\bullet - I$
  $MSW = \frac{SSW}{d.f.\{W\}}$
- $SSTO = SSB + SSW$
  $d.f.\{TO\} = n_\bullet - 1$
  $MSTO = \frac{SSTO}{d.f.\{TO\}}$

Hypothesis Test ANOVA

- Step 1: State the null, and alternative.
- Step 2: The test-statistic is: $F_S = \frac{MSB}{MSW}$
  with $d.f.\{numerator\} = I - 1$
  $d.f.\{denomenator\} = n_\bullet - I$
- Step 3: The p-value is:
  $Pr\{F > F_S\}$

Confidence Intervals for ANOVA

- A (1-α)100% simultaneous/overall/family-wise CI for k pairs of means is:
  $\bar{y}_a - \bar{y}_b \pm t_{\alpha/(2k)}\sqrt{MSW(\frac{1}{n_a} + \frac{1}{n_b})}$
  at d.f $= n_\bullet - I$

Confidence Intervals for a Proportion

- A $(1-\alpha)100\%$ confidence interval for p (the true proportion) is:
  $\tilde{p} \pm Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$
  where $\tilde{p} = \frac{y+2}{n+4}$ and you may find
  $Z_{\alpha/2}$ with $t_{\alpha/2}$ at d.f. $= \infty$