CLASSICAL TWO-SAMPLE RESULTS

We consider two random samples

$X_1, \dots, X_{n_1} \overset{iid}{\sim} F_1$, where F_1 has mean μ_1 and variance $\sigma_1^2 < \infty$.

$Y_1, \dots, Y_{n_2} \overset{iid}{\sim} F_2$, where F_2 has mean μ_2 and variance $\sigma_2^2 < \infty$.

COMPARING MEANS

I. Assuming Normal populations:

Let $X_1, \dots, X_{n_1} \overset{iid}{\sim} \text{Normal}(\mu_1, \sigma_1^2)$

$Y_1, \dots, Y_{n_2} \overset{iid}{\sim} \text{Normal}(\mu_2, \sigma_2^2)$

and let $\bar{X} = \frac{1}{n_1}(X_1 + \dots + X_{n_1})$ and $\bar{Y} = \frac{1}{n_2}(Y_1 + \dots + Y_{n_2})$.

Construct a $(1-\alpha)100\%$ C.I. for $\mu_1 - \mu_2$:

Case (Ø): σ_1^2 and σ_2^2 known

Use the following result:

Result:
$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \text{Normal}(0,1).$$

Proof: We have $\bar{X} \sim \text{Normal}\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$ and $\bar{Y} \sim \text{Normal}\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$.

Since $\bar{X}$ and $\bar{Y}$ are independent, we have

$$\bar{X} - \bar{Y} \sim \text{Normal}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right).$$ (see pg. 10 of Lec 02)

Therefore
$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \text{Normal}(0,1).$$

So this is a pivot quantity.

Application to C.I.s

Because of the above result, we may write

$$P\left(-z_{\alpha/2} < \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < z_{\alpha/2}\right) = 1-\alpha$$

$\Leftrightarrow$

$$P\left(\bar{X}-\bar{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu_1-\mu_2 < \bar{X}-\bar{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right) = 1-\alpha,$$

so that

$$\bar{X}-\bar{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

is a $(1-\alpha)^*100\%$ C.I. for $\mu_1-\mu_2$.

Case (i): σ_1^2 and σ_2^2 unknown, and $\sigma_1^2 \neq \sigma_2^2$:

We now consider using the estimators

$$S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\bar{X})^2 \quad \text{and} \quad S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2$$

in place of σ_1^2 and σ_2^2.

Use the following result:

Result (Satterthwaite):

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \overset{\text{approx}}{\sim} t_{\hat{\nu}}, \quad \hat{\nu} = \left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2\left[\frac{\left(\frac{S_1^2}{n_1}\right)^2}{(n_1-1)}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{(n_2-1)}\right]^{-1}.$$

Partial proof: Rewrite

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \Bigg/ \sqrt{\frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}\left(\frac{\sigma_1^2}{n_1(n_1-1)}\right)+\frac{(n_2-1)S_2^2}{\sigma_2^2}\left(\frac{\sigma_2^2}{n_2(n_2-1)}\right)}{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

This can be represented as

$$\frac{Z}{\sqrt{\frac{W_1\left(\frac{\sigma_1^2}{n_1(n_1-1)}\right)+W_2\left(\frac{\sigma_2^2}{n_2(n_2-1)}\right)}{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}},$$

where $Z \sim \text{Normal}(0,1)$, $W_1 \sim \chi^2_{n_1-1}$, $W_2 \sim \chi^2_{n_2-1}$, Z, W_1, W_2 indep.

This can be rewritten as

$$\frac{Z}{\sqrt{\underbrace{\frac{W_1\left(\frac{\sigma_1^2}{n_1(n_1-1)}\right)+W_2\left(\frac{\sigma_2^2}{n_2(n_2-1)}\right)}{\left[\frac{\left(\frac{\sigma_1^2}{n_1(n_1-1)}\right)^2(n_1-1)+\left(\frac{\sigma_2^2}{n_2(n_2-1)}\right)^2(n_2-1)}{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]}}_{\overset{\text{approx}}{\sim}\ \chi^2_{\nu^*}\ \text{(Welch-Satterthwaite)}} \Bigg/ \underbrace{\left[\frac{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^2}{\left(\frac{\sigma_1^2}{n_1(n_1-1)}\right)^2(n_1-1)+\left(\frac{\sigma_2^2}{n_2(n_2-1)}\right)^2(n_2-1)}\right]}_{\nu^*}}},$$

which has the form

$$Z \Big/ \sqrt{\frac{W^*}{\nu^*}}$$

where $Z \sim \text{Normal}(0,1)$, $W^* \overset{\text{approx}}{\sim} \chi^2_{\nu^*}$, ν^* as above.

Since Z and W^* are independent, $Z \Big/ \sqrt{\frac{W^*}{\nu^*}} \overset{\text{approx}}{\sim} t_{\nu^*}$.

Plugging in $\hat{\nu}$ for ν^*, i.e. replacing σ_1^2 and σ_2^2 with S_1^2 and S_2^2 in the expression for ν^*, we get the result.

$\square$

<u>Application to C.I.s</u>

Because of the above result, we may write

$$P\left(-t_{\hat{\nu},\alpha/2} < \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} < t_{\hat{\nu},\alpha/2}\right) \approx 1-\alpha$$

$\Leftrightarrow$

$$P\left(\bar{X}-\bar{Y}-t_{\hat{\nu},\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} < \mu_1-\mu_2 < \bar{X}-\bar{Y}+t_{\hat{\nu},\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right) \approx 1-\alpha,$$

so that

$$\bar{X}-\bar{Y} \pm t_{\hat{\nu},\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

is an approximate $(1-\alpha)^* 100\%$ C.I. for $\mu_1-\mu_2$.

<u>Case (ii): σ_1^2 and σ_2^2 unknown, but $\sigma_1^2=\sigma_2^2$:</u>

If we believe $\sigma_1^2=\sigma_2^2=\sigma^2_{common}$, for some $\sigma^2_{common}<\infty$, we we combine the two samples to estimate the common variance.

Define

$$S^2_{pooled} = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}.$$

Then we have the following result:

Result:

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{S^2_{pooled}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}.$$

Proof: Rewrite

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{S^2_{pooled}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma^2_{common}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \Bigg/ \sqrt{\frac{S^2_{pooled}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}{\sigma^2_{common}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$= \underbrace{\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma^2_{common}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}}_{\sim \text{Normal}(0,1)} \Bigg/ \sqrt{\underbrace{\frac{(n_1+n_2-2)\,S^2_{pooled}}{\sigma^2_{common}}}_{\sim \chi^2_{n_1+n_2-2}} \Big/ (n_1+n_2-2)},$$

$$\frac{(n_1+n_2-2)\,S^2_{pooled}}{\sigma^2_{common}} = \underbrace{\frac{(n_1-1)\,S_1^2}{\sigma^2_{common}}}_{\chi^2_{n_1-1}} + \underbrace{\frac{(n_2-1)\,S_2^2}{\sigma^2_{common}}}_{\chi^2_{n_2-1}} \sim \chi^2_{n_1+n_2-2}$$

independent

which has the form

$$Z \Big/ \sqrt{W/(n_1+n_2-2)},$$

where $Z \sim \text{Normal}(0,1)$, $W \sim \chi^2_{n_1+n_2-2}$, Z, W independent.

Thus $Z \Big/ \sqrt{W/(n_1+n_2-2)} \sim t_{n_1+n_2-2}$. $\square$

Application to C.I.s:

Because of the above result, we may write

$$P\left(-t_{n_1+n_2-2,\,\alpha/2} < \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{S^2_{pooled}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} < t_{n_1+n_2-2,\,\alpha/2}\right) = 1-\alpha$$

$\Leftrightarrow$

$$P\left(\bar{X}-\bar{Y}-t_{n_1+n_2-2,\,\alpha/2}\sqrt{S^2_{pooled}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} < \mu_1-\mu_2 < \bar{X}-\bar{Y}+t_{n_1+n_2-2,\,\alpha/2}\sqrt{S^2_{pooled}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right) = 1-\alpha,$$

so that

$$\bar{X} - \bar{Y} \pm t_{n_1+n_2-2,\,\alpha/2}\sqrt{S^2_{pooled}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

is a $(1-\alpha)^*100\%$ C.I. for $\mu_1 - \mu_2$.

II. Assuming non-Normal populations:

Let

$X_1, \ldots, X_{n_1} \overset{iid}{\sim} F_1$, where F_1 has mean μ_1 and variance $\sigma_1^2 < \infty$.

$Y_1, \ldots, Y_{n_2} \overset{iid}{\sim} F_2$, where F_2 has mean μ_2 and variance $\sigma_2^2 < \infty$.

and let $\bar{X} = \frac{1}{n_1}(X_1 + \cdots + X_{n_1})$ and $\bar{Y} = \frac{1}{n_2}(Y_1 + \cdots + Y_{n_2})$.

We wish to construct a $(1-\alpha)^*100\%$ C.I. for $\mu_1 - \mu_2$:

Case (Ø): σ_1^2 and σ_2^2 known

Use the following result:

This is how we ensure that both samples grow to ∞ together

Result: If $\overbrace{n_1/n_2 \to c}$, $0 < c < \infty$, as $n_1, n_2 \to \infty$, then

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \xrightarrow{D} Z, \quad Z \sim \text{Normal}(0,1)$$

as $n_1, n_2 \to \infty$.

Proof: Rewrite

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\sqrt{\frac{\sigma_1^2}{n_1}}\left(\frac{\bar{X} - \mu_1}{\sigma_1/\sqrt{n_1}}\right) - \sqrt{\frac{\sigma_2^2}{n_2}}\left(\frac{\bar{Y} - \mu_2}{\sigma_2/\sqrt{n_2}}\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \sqrt{\frac{\frac{\sigma_1^2}{n_1}}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \left(\frac{\bar{X} - \mu_1}{\sigma_1 / \sqrt{n_1}} \right) - \sqrt{\frac{\frac{\sigma_2^2}{n_2}}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \left(\frac{\bar{Y} - \mu_2}{\sigma_2 / \sqrt{n_2}} \right)$$

$$= \underbrace{\sqrt{\frac{\sigma_1^2}{\sigma_1^2 + \left(\frac{n_1}{n_2}\right)\sigma_2^2}}}_{\to \sqrt{\frac{\sigma_1^2}{\sigma_1^2 + c\sigma_2^2}}} \underbrace{\left(\frac{\bar{X} - \mu_1}{\sigma_1 / \sqrt{n_1}} \right)}_{\to^D Z_1} - \underbrace{\sqrt{\frac{\frac{n_1}{n_2}\sigma_2^2}{\sigma_1^2 + \frac{n_1}{n_2}\sigma_2^2}}}_{\to \sqrt{\frac{c\sigma_2^2}{\sigma_1^2 + c\sigma_2^2}}} \underbrace{\left(\frac{\bar{Y} - \mu_2}{\sigma_2 / \sqrt{n_2}} \right)}_{\to^D Z_2}$$

$$\to^D \sqrt{\frac{\sigma_1^2}{\sigma_1^2 + c\sigma_2^2}}\, Z_1 - \sqrt{\frac{c\sigma_2^2}{\sigma_1^2 + c\sigma_2^2}}\, Z_2 ,$$

as $n_1, n_2 \to \infty$, where $Z_1, Z_2 \sim \text{Normal}(0,1)$, Z_1, Z_2 independent.

Finally, we have that

$$\sqrt{\frac{\sigma_1^2}{\sigma_1^2 + c\sigma_2^2}}\, Z_1 - \sqrt{\frac{c\sigma_2^2}{\sigma_1^2 + c\sigma_2^2}}\, Z_2 \sim \text{Normal}(0, 1),$$

since $\mathbb{E} Z_1 = \mathbb{E} Z_2 = 0$ and

$$\text{Var}\left[\sqrt{\frac{\sigma_1^2}{\sigma_1^2 + c\sigma_2^2}}\, Z_1 - \sqrt{\frac{c\sigma_2^2}{\sigma_1^2 + c\sigma_2^2}}\, Z_2 \right] = \frac{\sigma_1^2}{\sigma_1^2 + c\sigma_2^2} \text{Var}\, Z_1 + \frac{c\sigma_2^2}{\sigma_1^2 + c\sigma_2^2} \text{Var}\, Z_2$$

$$= \frac{\sigma_1^2}{\sigma_1^2 + c\sigma_2^2} + \frac{c\sigma_2^2}{\sigma_1^2 + c\sigma_2^2}$$

$$= 1,$$

since Z_1 and Z_2 are independent. $\square$

Application to C.I.s:

Because of the above result, we may write

$$\lim_{n_1,n_2\to\infty} P\left(-z_{\alpha/2} < \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < z_{\alpha/2}\right) = 1-\alpha$$

$\Leftrightarrow$

$$\lim_{n_1,n_2\to\infty} P\left(\bar{X}-\bar{Y}-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu_1-\mu_2 < \bar{X}-\bar{Y}+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right) = 1-\alpha,$$

so that for $\overbrace{\text{large } n_1, n_2}^{\text{A rule of thumb is } \min\{n_1,n_2\}\geq 30}$

$$\bar{X}-\bar{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

is an approximate $(1-\alpha)^*100\%$ C.I. for $\mu_1-\mu_2$.

Case (i): σ_1^2 and σ_2^2 unknown, and $\sigma_1^2 \neq \sigma_2^2$:

Use the following result:

Result: If $n_1/n_2 \to c$, $0<c<\infty$, as $n_1, n_2 \to \infty$, then

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \xrightarrow{D} Z, \quad Z \sim \text{Normal}(0,1)$$

as $n_1, n_2 \to \infty$.

Proof: Rewrite

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sqrt{\frac{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

$$= \underbrace{\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}}_{\xrightarrow{D} Z,\ Z\sim \text{Normal}(0,1)}\sqrt{\underbrace{\frac{\sigma_1^2+\frac{n_1}{n_2}\sigma_2^2}{S_1^2+\frac{n_1}{n_2}S_2^2}}_{\xrightarrow{P} 1}}\ .$$

, since S_1^2 and S_2^2 are consistent estimators of σ_1^2 and σ_2^2.

So the result follows by applying Slutzky's theorem. $\square$

Application to C.I.s:

Because of the above result, we may write

$$\lim_{n_1,n_2\to\infty} P\left(-z_{\alpha/2} < \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} < z_{\alpha/2}\right) = 1-\alpha$$

$\Leftrightarrow$

$$\lim_{n_1,n_2\to\infty} P\left(\bar{X}-\bar{Y}-z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} < \mu_1-\mu_2 < \bar{X}-\bar{Y}+z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right) = 1-\alpha$$

so that for $\overbrace{\text{large } n_1, n_2}^{\text{A rule of thumb is } \min\{n_1,n_2\}\geq 30}$

$$\bar{X}-\bar{Y} \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

is an approximate $(1-\alpha)^*100\%$ C.I. for $\mu_1-\mu_2$.

Case (ii): σ_1^2 and σ_2^2 unknown, but $\sigma_1^2 = \sigma_2^2$:

Use the following result:

Result: If $n_1/n_2 \to c$, $0 < c < \infty$, as $n_1, n_2 \to \infty$, then

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S^2_{pooled}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \xrightarrow{D} Z, \quad Z \sim \text{Normal}(0,1)$$

as $n_1, n_2 \to \infty$.

Proof: Let $\sigma_1^2 = \sigma_2^2 = \sigma^2_{common}$ and rewrite

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S^2_{pooled}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma^2_{common}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sqrt{\frac{\sigma^2_{common}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{S^2_{pooled}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \underbrace{\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma^2_{common}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}}_{\xrightarrow{D} Z,\ Z \sim \text{Normal}(0,1)} \underbrace{\sqrt{\frac{\sigma^2_{common}}{S^2_{pooled}}}}_{\xrightarrow{P} 1} .$$

$\xrightarrow{P} 1$, since S^2_{pooled} is a consistent estimate of σ^2_{common}.

So the result follows by applying Slutzky's theorem. $\square$

Application to C.I.s:

Because of the above result, we may write

$$\lim_{n_1,n_2\to\infty} P\left(-z_{\alpha/2} < \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{S^2_{pooled}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} < z_{\alpha/2}\right) = 1-\alpha$$

$\Leftrightarrow$

$$\lim_{n_1,n_2\to\infty} P\left(\bar{X}-\bar{Y}-z_{\alpha/2}\sqrt{S^2_{pooled}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} < \mu_1-\mu_2 < \bar{X}-\bar{Y}+z_{\alpha/2}\sqrt{S^2_{pooled}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right) = 1-\alpha,$$

so that for $\overbrace{\text{large } n_1, n_2}^{\text{A rule of thumb is } \min\{n_1,n_2\}\geq 30}$

$$\bar{X}-\bar{Y} \pm z_{\alpha/2}\sqrt{S^2_{pooled}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

is an approximate $(1-\alpha)^*100\%$ C.I. for $\mu_1-\mu_2$.

SUMMARY OF C.I.s FOR THE DIFFERENCE IN MEANS

Let $X_1,\dots,X_{n_1}$ be independent rvs with the same distribution as X.
Let $Y_1,\dots,Y_{n_2}$ be independent rvs with the same distribution as Y.

case(i) $\sigma_1^2 \neq \sigma_2^2$:

- X and Y Normal: $\boxed{\bar{X}-\bar{Y} \pm t_{\hat{\nu},\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$
- X or Y non-Normal
  - $\min\{n_1,n_2\}\geq 30$: $\boxed{\bar{X}-\bar{Y} \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$
  - $\min\{n_1,n_2\}<30$: $\boxed{\text{Beyond scope of course.}}$

These are both approximate $(1-\alpha)^*100\%$ C.I.s

Case (ii) $\sigma_1^2 = \sigma_2^2$:

- X and Y Normal $\rightarrow$ $\boxed{\bar{X} - \bar{Y} \pm t_{n_1+n_2-2,\, \alpha/2} \sqrt{S^2_{pooled}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
- X or Y non-Normal
  - $\min\{n_1, n_2\} \geq 30$ $\rightarrow$ $\boxed{\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{S^2_{pooled}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ Approximate $(1-\alpha)^* 100\%$ C.I.
  - $\min\{n_1, n_2\} < 30$ $\rightarrow$ $\boxed{\text{Beyond scope of course.}}$

COMPARING PROPORTIONS

Let $X_1, \ldots, X_{n_1} \overset{iid}{\sim} \text{Bernoulli}(p_1)$, p_1 unknown

$Y_1, \ldots, Y_{n_2} \overset{iid}{\sim} \text{Bernoulli}(p_2)$, p_2 unknown.

Define $\hat{p}_1 = \frac{1}{n_1}(X_1 + \cdots + X_{n_1})$ and $\hat{p}_2 = \frac{1}{n_2}(Y_1 + \cdots + Y_{n_2})$.

We wish to build a $(1-\alpha)^* 100\%$ C.I. for $p_1 - p_2$.

We use the following result:

Result: If $n_1/n_2 \to c$, $0 < c < \infty$, as $n_1, n_2 \to \infty$, then

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \xrightarrow{D} Z, \quad Z \sim \text{Normal}(0,1)$$

as $n_1, n_2 \to \infty$.

Proof: We can follow the proof of (◆), using the fact that

$$\hat{p}_1(1-\hat{p}_1) \xrightarrow{P} p_1(1-p_1) \quad \text{as } n_1 \to \infty$$

and $$\hat{p}_2(1-\hat{p}_2) \xrightarrow{P} p_2(1-p_2) \quad \text{as } n_2 \to \infty.$$

Application to C.I.s:

We can use the above result to write

$$\lim_{n_1,n_2\to\infty} P\left(-z_{\alpha/2} < \frac{\hat{p}_1-\hat{p}_2-(p_1-p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} < z_{\alpha/2} \right) = 1-\alpha$$

$\Leftrightarrow$

$$\lim_{n_1,n_2\to\infty} P\left(\hat{p}_1-\hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} < p_1-p_2 < \hat{p}_1-\hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \right) = 1-\alpha,$$

giving that for large n_1 and n_2, Rule of thumb: $\min\{n_1\hat{p}_1, n_1(1-\hat{p}_1)\} \geq 15$, $\min\{n_2\hat{p}_2, n_2(1-\hat{p}_2)\} \geq 15$

$$\hat{p}_1-\hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

is an approximate $(1-\alpha)*100\%$ C.I. for p_1-p_2.

COMPARING VARIANCES

Suppose we have two independent random samples

$$X_1,\dots,X_{n_1} \sim \text{Normal}(\mu_1,\sigma_1^2)$$
$$Y_1,\dots,Y_{n_2} \sim \text{Normal}(\mu_2,\sigma_2^2),$$

and let

$$S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}\left(X_i-\bar{X}_{n_1}\right)^2 \quad \text{and} \quad S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}\left(Y_i-\bar{Y}_{n_2}\right)^2.$$

Result:

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,\,n_2-1}.$$

Application to C.I.s:

$$P_{S_1^2, S_2^2}\left(F_{n_1-1, n_2-1, 1-\alpha/2} < \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < F_{n_1-1, n_2-1, \alpha/2} \right) = 1-\alpha$$

$\iff$

$$P_{S_1^2, S_2^2}\left(\frac{S_2^2}{S_1^2} F_{n_1-1, n_2-1, 1-\alpha/2} < \frac{\sigma_2^2}{\sigma_1^2} < \frac{S_2^2}{S_1^2} F_{n_1-1, n_2-1, \alpha/2} \right) = 1-\alpha$$

$\iff$

A $(1-\alpha)^*100\%$ C.I. for $\frac{\sigma_2^2}{\sigma_1^2}$ is

$$\left(\frac{S_2^2}{S_1^2} F_{n_1-1, n_2-1, 1-\alpha/2} \, , \, \frac{S_2^2}{S_1^2} F_{n_1-1, n_2-1, \alpha/2} \right).$$