Sets 28-29 Guide: Inference on the Difference of Two Means ($\mu_1 - \mu_2$)

General Requirements:

- ▷ Random samples
- ▷ The two populations are ***independent*** (No paired data, before/after data, etc).

For all four of the cases, the hypotheses have the same form:

Right-Tailed:	**Left-Tailed:**	**Two-Tailed:**
$H_0 : \mu_1 - \mu_2 = (\mu_1 - \mu_2)_0$	$H_0 : \mu_1 - \mu_2 = (\mu_1 - \mu_2)_0$	$H_0 : \mu_1 - \mu_2 = (\mu_1 - \mu_2)_0$
or $H_0 : \mu_1 - \mu_2 \leq (\mu_1 - \mu_2)_0$	or $H_0 : \mu_1 - \mu_2 \geq (\mu_1 - \mu_2)_0$	or $H_0 : \mu_1 - \mu_2 = (\mu_1 - \mu_2)_0$
$H_1 : \mu_1 - \mu_2 > (\mu_1 - \mu_2)_0$	$H_1 : \mu_1 - \mu_2 < (\mu_1 - \mu_2)_0$	$H_1 : \mu_1 - \mu_2 \neq (\mu_1 - \mu_2)_0$

★ Cases 1 and 2: (Using the Z-distribution)

(i) **Requirements:**

- ▷ Random samples from two independent populations.
- ▷ **Case 1**: σ_1 and σ_2 are known, and both populations are normally distributed.
- ▷ **Case 2**: n_1 and $n_2 \geq 30$.

(ii) **Standard Error and Estimated Standard Error**:

$$\text{se: } \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad \text{ese: } \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(iii) **Confidence Interval for $\mu_1 - \mu_2$:**

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad \text{or} \qquad (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(iv) **Hypothesis Testings:**

- ▷ **Test statistic** for testing $H_0 : \mu_1 - \mu_2 = (\mu_1 - \mu_2)_0$:

$$z_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad \text{or} \qquad z_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

  Test statistic is approximately normally distributed when H_0 is true.

- ▷ p-**values:**

  Right-tailed Test: $p - value = P(Z > z_{obs})$
  Left-tailed Test: $p - value = P(Z < z_{obs})$
  Two-tailed Test: $p - value = 2P(Z < -|z_{obs}|)$

★ Case 3: Small samples with equal variances ($\sigma_1^2 = \sigma_2^2$) - Pooled Two-Sample T-test

(i) **Requirements:**

- ▷ Random sample from two independent populations, that are (near) normally distributed.
- ▷ Equal Variances ($\sigma_1^2 = \sigma_2^2$):

$$\text{Rule of Thumb:} \qquad \frac{\text{larger standard deviation}}{\text{smaller standard deviation}} \leq 1.4$$

(ii) **Estimated Standard Error:**

$$\text{ese:} \quad \sqrt{\left(\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

(iii) **Confidence Interval for $\mu_1 - \mu_2$:**

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\nu,\alpha/2} \sqrt{\left(\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

with **degrees of freedom** $\nu = n_1 + n_2 - 2$.

(iv) **Hypothesis Testings:**

- ▷ **Test statistic** for testing $H_0 : \mu_1 - \mu_2 = (\mu_1 - \mu_2)_0$:

$$t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\left(\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

  Test statistic has a T_ν-distribution with $\nu = n_1 + n_2 - 2$ degrees of freedom when H_0 is true.

- ▷ p-**values**

  Right-tailed Test: p-value $= P(T_\nu > t_{obs})$
  Left-tailed Test: p-value $= P(T_\nu < t_{obs})$
  Two-tailed Test: p-value $= 2P(T_\nu > |t_{obs}|)$

★ **Case 4: Small samples with unequal variances ($\sigma_1^2 \neq \sigma_2^2$) - Unpooled Two-Sample T-test**

(i) **Requirements:**

▷ Random sample from two independent populations, that are (near) normally distributed.

▷ Unequal Variances ($\sigma_1^2 \neq \sigma_2^2$):

$$\text{Rule of Thumb:} \qquad \frac{\text{larger standard deviation}}{\text{smaller standard deviation}} > 1.4$$

(ii) **Confidence Interval for $\mu_1 - \mu_2$:**

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\nu,\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

with **degrees of freedom** ν:

$$\nu = \text{ the integer part of } \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

(iii) **Hypothesis Testings:**

- **Test statistic** for testing $H_0 : \mu_1 - \mu_2 = (\mu_1 - \mu_2)_0$:

$$t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Test statistic has a T_ν distribution with ν as in the CI, when H_0 is true.

- p-**values** : as in Case 3

Sample Size for Estimating $\mu_1 - \mu_2$

The common ($n = n_1 = n_2$) sample size needed to construct a $(1 - \alpha)100\%$ confidence interval within margin of error d (with length $2d$) for $\mu_1 - \mu_2$ is given by

$$d = z_{\alpha/2}\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}} \qquad \Rightarrow \qquad n = \left(\frac{z_{\alpha/2}}{d}\right)^2 (s_1^2 + s_2^2)$$