Statistics 300: Elementary Statistics

Section 9-3

Section 9-3 concerns Confidence Intervals and Hypothesis tests for the difference of two means, $(\mu_1 - \mu_2)$

What is the variance of

$(\bar{x}_1 - \bar{x}_2)$?

- **Apply the new concept here also:**

$$s^2_{x-y} = s^2_x + s^2_y$$

Application: Use sample values

$$s^2_{\bar{x}-\bar{y}} = s^2_{\bar{x}} + s^2_{\bar{y}}$$

$$s^2_{\bar{x}-\bar{y}} = s^2_{\bar{x}} + s^2_{\bar{y}}$$

$$s^2_{\bar{x}-\bar{y}} = \frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}$$

Alternative approach when two samples come from populations with equal variances

$$s_1^2 = s_2^2$$

["homogeneous variances" or "homoscedastic"]

When samples come from populations with equal variances:

$$s_1^2 = s_2^2 = s^2_{common}$$

and

s_1^2 is an estimate of s^2_{common}

s_2^2 is an estimate of s^2_{common}

When samples come from populations with equal variances:

the best estimate of s^2_{common} is:

$$s_p^2 = \frac{(df_1)s_1^2 + (df_2)s_2^2}{(df_1)+(df_2)}$$

Alternative notation for the preceding formula:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1)+(n_2 - 1)}$$

When doing confidence intervals of hypothesis tests involving (m1 – m2) one must <u>first</u> decide whether the variances are different or the same, heterogeneous or homogeneous.

If the variances are different (heterogeneous), then use the following expression as part of the confidence interval formula or the test statistic:

$$\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

If the variances are the same (homogeneous), then use this alternative expression as part of the confidence interval formula or the test statistic:

$$\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}$$

where "p" means "pooled"

When the variances are pooled to estimate the common (homogeneous) variance, then the degrees of freedom for both samples are also pooled by adding them together.

Application to CI(μ_1 - μ_2)

$$CI(m_1 - m_2) = (\bar{x}_1 - \bar{x}_2) \pm E$$

$$E = t_{a/2} \cdot s_{\bar{x}_1 - \bar{x}_2}$$

Application to CI(μ_1 - μ_2) when variances are different

$$CI(m_1 - m_2) = (\bar{x}_1 - \bar{x}_2) \pm E$$

$$E = t_{a/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

In this case, the degrees of freedom for "t" will be the smaller of the two sample degrees of freedom.

Application to CI(μ_1 - μ_2) when variances are the same

$$CI(m_1 - m_2) = (\bar{x}_1 - \bar{x}_2) \pm E$$

$$E = t_{a/2} \cdot \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

In this case, the degrees of freedom for "t" will be the sum of the two sample degrees of freedom.

Tests concerning (m_1 - m_2)
Test Statistic

$$\frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)_0}{s_{(\bar{x}_1 - \bar{x}_2)}}$$

Test statistic for (m_1 - m_2)
when variances are different

$$\frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

In this case, the degrees of freedom for "t" will be the smaller of the two sample degrees of freedom.

Test statistic for (m_1 - m_2)
when variances are the same

$$\frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

In this case, the degrees of freedom for "t" will be the ssum of the two sample degrees of freedom.

Section 9-3
Handling Claims / Hypotheses

- **Write the claim in a symbolic expression as naturally as you can**
- **Then rearrange the expression to have the difference between the two means on one side of the relational operator (< > = …)**

Section 9-3
Handling Claims / Hypotheses

- **Statement: Mean #2 is less than four units more than Mean #1**
- **So:** $m_2 < m_1 + 4$
- **Rearrange:** $(m_2 - m_1) < 4$
- **H_0:** $(m_2 - m_1) \geq 4$
- **H_1:** $(m_2 - m_1) < 4$

Section 9-3
Handling Claims / Hypotheses

- **Statement: On average, treatment A produces 18 more than treatment B**
- **So:** $m_A = m_B + 18$
- **Rearrange:** $(m_A - m_B) = 18$
- **H_0:** $(m_A - m_B) = 18$
- **H_1:** $(m_A - m_B) \neq 18$