Chapter 19: Two-Sample Problems
Key Ideas

- The two-sample t procedures follow from the fact that

$$\frac{\text{estimate} - \text{hypothesized value}}{SE(\text{estimate})}$$

has an approximate t distribution (though finding the right degrees of freedom is more complicated).

Conditions for comparing two means:

- We have two independent SRSs from distinct populations with unknown means and standard deviations.
    - Separate samples from each treatment or population
- Both populations are *Normally distributed*.
    - In practice, the two-sample t procedures are valid if the distributions have similar shapes and no strong outliers..

t test for $H_0 : \mu_1 = \mu_2$:

1. **State** the problem in context

2. **Formulate**

   a. Define the parameters of interest

   b. State H_0 and H_a.

3. **Solve**

   a. Check that the conditions are satisfied.

   b. Compute the two-sample *t* statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

c. Compute the *p*-value using the t distribution

$$H_a: \mu_1 > \mu_2 \Rightarrow p-\text{value} = P(T \geq t)$$

$$H_a: \mu_1 < \mu_2 \Rightarrow p-\text{value} = P(T \leq t)$$

$$H_a: \mu_1 \neq \mu_2 \Rightarrow p-\text{value} = 2P(T \geq |t|)$$

with degrees of freedom equal to

- The smaller of $n_1 - 1$ and $n_2 - 1$, **if computed by hand.**
- The expression at the top of p. 474 of the text, **if computed using technology.**

4. **Conclude** in the context of the **stated** problem.

Note: This test can be carried out using the two-sample t-test procedure in SPSS. Use Analyze…Compare Means…Independent Samples t-test. Select the test and grouping variables, define the groups, and refer to the "Equal variances not assumed" portion of the output.

t **confidence interval for** $\mu_1 - \mu_2$

1. **State** the parameter estimation problem in context

2. **Formulate**

    a. Define the parameters of interest

    b. State the level of confidence to be used

3. **Solve**

    a. Check that the conditions are satisfied (see practical guidelines below).

    b. Calculate the level *C* confidence interval for $\mu_1 - \mu_2$

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Where t^* is the upper (1-*C*)/2 critical value from the *t* distribution (with degrees of freedom as described above for the significance test.)

4. **Conclude** in the context of the **stated** problem.

Notes:

- This interval has confidence level at least C no matter what the population standard deviations may be.
- These computations need not be carried out by hand. Use the Independent-Samples *t* procedure in SPSS and refer to the "Equal Variances not Assumed" portion of the output.

Practical Guidelines for use of the *t* procedures

- When sample sizes are similar and distributions are of similar shapes, *t* procedures are valid for $n_1 \geq 5$ and $n_2 \geq 5$.
- If distribution shapes are different, larger samples are needed (so CLT applies).
  - Use guidelines from one-sample *t* procedures (Chapter 18 notes) substituting $n_1 + n_2$ for *n*.