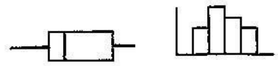
Chapter 24: Comparing Means

Plotting the data – When making a picture of the data, it is best to use a boxplot or histogram 

Comparing Two Means – The parameter of interest is μ_1-μ_2. The statistic of interest is the two sample means, y_1-y_2. As with comparing two proportions, the standard deviation is found by adding the variances and taking the square root of that sum. However, because we do not know the true population's standard deviation, σ, we substitute it with s, our estimated standard deviation.

$$SD(\bar{y}_1-\bar{y}_2)=\sqrt{Var(\bar{y}_1)+Var(\bar{y}_2)} \qquad SE(\bar{y}_1-\bar{y}_2)=\sqrt{\frac{(s_1)^2}{n_1}+\frac{(s_2)^2}{n_2}}$$

$$=\sqrt{\frac{(\sigma_1)^2}{n_1}+\frac{(\sigma_2)^2}{n_2}}$$

Confidence Interval – **two-sample *t*-interval**
Hypothesis Test – **two-sample *t*-test**
The calculated interval is the same – the statistic plus or minus the margin of error.

$$(\bar{y}_1-\bar{y}_2)\pm ME \quad \text{where } ME = t^* \times SE(\bar{y}_1-\bar{y}_2)$$

Degrees of Freedom: Let the calculator figure out the degrees of freedom by going to STAT:TEST and conducting a *t*-test or interval, or by using a *t*-table, always rounding down to be safe.

Assumptions and Conditions:

- Randomization Condition – Data should be collected with some form of randomization.
- 10% Condition (only necessary if population is small or sample is very large)
- Nearly Normal Condition – BOTH groups of data must have histograms that are nearly Normal; important for small samples ($n \leq 15$), but matters less as sample size grows.
- Independent Groups Assumption – groups must be independent of **each other**.

Pooled *t*-Tests* – We can pool the data from two groups to estimate the common variance ONLY if we are willing to assume that the groups' variances are **equal**.

$$s^2_{pooled}=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1-1)+(n_2-1)} \qquad SE_{pooled}(\bar{y}_1-\bar{y}_2)=s_{pooled}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Assumptions and conditions are the same as the two-sample *t*-test, plus:

- Equal Variances Assumption – the variances of the two populations that the samples have been drawn from are equal.
- Nearly Equal Spreads Condition – the boxplots or histograms of the data must have reasonably equal spreads.

Degrees of Freedom: $df = n_1 + n_2 - 2$

*Pooled *t*-tests often have fewer degrees of freedom and more power than a two-sample *t*-test, but the advantage is not usually enough to balance out the hassle of checking the extra conditions accurately. In other words, **it's not worth it**. Use the two-sample *t*-test instead.

Tukey's Quick Test* – a simpler alternative to the two-sample *t*-test that is surprisingly accurate, and is good for checking your results.

Assumptions and Conditions:

- Independence Assumption
- One group must have the **highest value**; the other group must have the **lowest value**.

How It Works:

1. Count how many values in the high group are larger than ALL of the values in the low group. Then count how many values in the low group are smaller than ALL of the values in the high group (count ties as ½).
2. Total up your two counts. If the total is equal to…
   - 7 or more, we can reject H_0 at $\alpha = 0.05$
   - 10 or more, we can reject H_0 at $\alpha = 0.01$
   - 13 or more, we can reject H_0 at $\alpha = 0.001$

*Tukey's quick test is not widely known or accepted; do NOT use it as a replacement for the two-sample *t*-test.