Review of the Topics for Midterm II

I. Confidence Interval

1. Point Estimation
   a. A point estimator of a parameter is a statistic used to estimate that parameter.
   b. Properties of a “good” estimator are:

      Unbiasedness, Minimum Variance, Consistency

2. Interval Estimation
   a. Confidence Interval: A random interval which covers the true value of a parameter with a given probability.
   b. Confidence Interval for $\mu$ when $\sigma$ is known:

      A 100(1-$\alpha$)% confidence interval for $\mu$ when $\sigma$ is known is:

      $$\bar{Y} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

   c. Sample size

   d. Student’s t Distribution

      The $t$ distribution is a bell-shaped distribution looking like a normal distribution but with heavier tails. Using the t-table.

   e. Confidence Interval for $\mu$ When $\sigma$ Is Unknown:

      A 100(1-$\alpha$) % confidence interval for $\mu$ when $\sigma$ is unknown is:

      $$\bar{Y} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$
Example:

f. Sample size

g. Confidence Interval for a Population Proportion
II. Comparison of Two Independent Populations

Confidence Intervals:

a. A 100(1 - \( \alpha \))% confidence interval for (\( \mu_1 - \mu_2 \)) when the population standard deviations are known is:

b. A 100(1 - \( \alpha \))% confidence interval for (\( \mu_1 - \mu_2 \)) when the population standard deviations are unknown is:

Example: Problem #4

c. If \( \sigma_1 \) and \( \sigma_2 \) are unknown but \( \sigma_1 = \sigma_2 \), then a pooled estimate of the common variance is:

Using this estimate, a 100(1 - \( \alpha \))% c.i. for (\( \mu_1 - \mu_2 \)) is

Example: Problem #1
Hypothesis Testing:

a. Null and Alternative Hypotheses

*Null Hypothesis*: A conjecture about a parameter of a population or parameters of two population.

*Alternative Hypothesis*: Another conjecture about the same parameter(s).

b. The decision to reject or not reject the null hypothesis is based on a statistic computed from the sample. This is called the *test statistic*.

c. Use of Data and Possible Errors

d. To test \( H_0 : \mu_1 = \mu_2 \) compute \( t \) and reject \( H_0 \) in favor of \( H_A \) of

\[
H_A : \mu_1 < \mu_2 \text{ if } t < -t_{\alpha}
\]

\[
H_A : \mu_1 > \mu_2 \text{ if } t > t_{\alpha}
\]

\[
H_A : \mu_1 \neq \mu_2 \text{ if } t < -t_{\alpha/2} \text{ or } t > t_{\alpha/2}
\]

**Example**: Problem #4
e. If $\sigma_1$ and $\sigma_2$ are unknown but $\sigma_1 = \sigma_2$, then use the *pooled Procedure*.

**Example:** Problem #1

---

**Sample size Calculation and Power**

A popular formula for computing the sample size for comparing two means (one-sided alternative) is:

$$n_1 = \left(\sigma_1^2 + \sigma_2^2 / k\right) \left(z_{\alpha} + z_{\beta}\right)^2 / (\mu_1 - \mu_2)^2$$

$$n_2 = k \ n_1$$

**Example:**
The Wilcoxon-Mann-Whitney Test

Nonparametric methods are used when:

1. The populations are not normal.
2. Data are qualitative.
3. Data are ranked.

To compare two populations, suppose we have a sample of size $n_1$ from the first population and a sample of size $n_2$ from the second population ($n_1 \geq n_2$). For each observation in sample 1, count the number of observations in sample 2 that are smaller in value. Let $K_1$ be the sum of these counts. Similarly, for each observation in sample 2, count the number of observations in sample 1 that are smaller in value. Let $K_2$ be the sum of these counts. The Wilcoxon-Mann-Whitney test statistic $U_s$ is the larger of $K_1$ and $K_2$. We use Table 6 to find the critical values.

Example:

III. Design of Experiments

Observational and Experimental Designs, Confounding, Case-Control Studies, and Randomization.
IV. Comparison of Paired Samples

Paired t-test and Confidence Interval

In the matched pair designs we apply the one-sample t-procedure. *Paired designs* are used to free the comparisons from the effects of extraneous variables.

The Sign Test

In many applications we may be interested in comparing *matched pairs*, when we have *ranked* or *quantitative* data.

Let \( N_+ \) and \( N_- \) denote the number of positive and negative differences, respectively. Define \( B_\text{s} \) as the larger of \( N_+ \) and \( N_- \) and use Table 7 to find the critical values.

Example:

Wilcoxon Signed Rank Sum Test

This test for matched pairs is for quantitative data when the normality assumption is not satisfied.

Let \( W_+ \) and \( W_- \) denote the rank sum of the positive and negative differences, respectively. Define \( W_\text{s} \) as the larger of \( W_+ \) and \( W_- \) and use Table 8 to find the critical values.

**Example:** Problem #3
V. Analysis of Categorical Data

*Categorical data analysis* is used when the variable under study is classified into several categories.

**The Chi-Square Goodness-of-Fit Test**

Let \( O \) represent the *observed* and \( E \) the *expected* frequencies, respectively. Then the *chi-square test statistics* is:

\[
\chi^2 = \frac{\sum (O-E)^2}{E}
\]

The *degree of freedom* of the chi-square statistics is \( df = v - 1 \), where \( v \) is the number of categories. We use Table 9 to find the critical values.

The *2x2 contingency tables* arise in dealing with two binary categorical responses. The *chi-square test statistics* for testing *independence* or *homogeneity* in 2x2 tables is:

\[
\chi^2 = \frac{\sum (O-E)^2}{E}
\]

The degree of freedom of the chi-square statistics is \( df = 1 \). Here,

\[
E = \text{Row tot} \times \text{Column tot} / \text{Grand tot}
\]

**Example:** Problem #2

**The r×k Contingency Tables**

The *r×k contingency tables* are generalization of 2x2 tables in dealing with two categorical responses. Again if we let \( O \) represent the *observed* and \( E \) the *expected* frequencies. We can use the *chi-square test statistics*:

\[
\chi^2 = \frac{\sum (O-E)^2}{E}
\]

The *degree of freedom* of this chi-square statistics is \( df = (r-1)(k-1) \).
Fisher’s Exact Test

Example: Exercise 10.4.6, page 385

Difference Between Two Population Proportions

100(1-\(\alpha\)) % adjusted and unadjusted confidence intervals for \((p_1 - p_2)\)

To test \(H_0: p_1 = p_2\) we compute \(z\)

Example: Exercise 10.7.2, page 397.