Review of the Topics for the Final

I. Comparison of Two Independent Populations

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

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:


Hypothesis Testing
Null and alternative hypotheses, type I and type II errors, test statistics

To test \(H_o: \mu_1 = \mu_2\) compute \(t\) and reject \(H_o\) in favor of \(H_\Lambda\) of

\(H_\Lambda: \mu_1 < \mu_2\) if \(t < -t_\alpha\)

\(H_\Lambda: \mu_1 > \mu_2\) if \(t > t_\alpha\)

\(H_\Lambda: \mu_1 \neq \mu_2\) if \(t < -t_{\alpha/2}\) or \(t > t_{\alpha/2}\)
Example: Problem #2 of the sample final.

If \( \sigma_1 \) and \( \sigma_2 \) are unknown but \( \sigma_1 = \sigma_2 \), then use the **pooled Procedure**.

Example: Exercise 7.5.6, page 258.

**Sample size Calculation and Power**

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

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

\[
    n_2 = k n_1
\]

**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** \( U \) is the larger of \( K_1 \) and \( K_2 \). We use Table 6 to find the critical values.

Example: Problem #2 of the sample final.
II. 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.

Example: Exercise 8.2.2, page 308.

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 $B_s$ as the larger of $N_+$ and $N_-$ and use Table 7 to find the critical values.

Example: Exercise 8.4.5, page 320.

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 $W_s$ as the larger of $W_+$ and $W_-$ and use Table 8 to find the critical values.

Example: Exercise 8.5.7, page 326.
III. 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:

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

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

**The 2x2 Contingency Tables**

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:

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

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

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

**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*:

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

The *degree of freedom* of this chi-square statistics is $df = (r - 1)(k-1)$.

**Example:** Problem #4 of the sample final.
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.
**Paired Data for 2x2 Tables**

McNemar’s test can be used to test the hypothesis that the discordant pairs are equally likely. The chi-square statistics is:

\[ \chi^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} \]

with 1 degrees of freedom.

**Example:** Exercise 10.8.2, page 400.

**Relative Risk and Odds ratio**

Ratio of two probabilities is called relative risk. The odds ratio is the ratio of two odds under two different conditions.

**Example:** Exercise 10.9.2, page 408.

**Confidence Interval for the Odds Ratio**

The estimate of the odds ratio is:

\[ \hat{\theta} = \frac{n_{11} n_{22}}{n_{12} n_{21}} \]

where \( n_{11}, n_{12}, n_{21}, \) and \( n_{22} \) are the table frequencies. For large \( n \), \( \log(\hat{\theta}) \) is approximately normal with mean \( \log(\theta) \) and standard error

\[ \text{SE}_{\log(\theta^\wedge)} = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} \]

This leads to a 100(1-\( \alpha \))% confidence interval for \( \log(\theta) \) as

\[ \log(\hat{\theta}) \pm Z_{\alpha/2} \text{SE}_{\log(\theta^\wedge)} \]

The confidence interval for \( \theta \) is exponentiation of this interval.

**Example:** Exercise 10.9.6, page 409.