# Objectives for Exam 2

## Introduction

This document outlines the objectives for Exam 2, covering the essential concepts from chapters 7 through 11.

## Chapter 7

- **Understanding Parameters and Statistics**: Accurately define what constitutes a parameter in the context of a population and a statistic in the context of a sample.

- **Identifying and Appreciating Sampling Distributions**:

  - Correctly identify scenarios that involve sampling distributions.
  - Explain the significance of sampling distributions in statistical inference.
  - Correctly identify the mean ($\mu$) and standard error ($\sigma/\sqrt{n}$) of the sampling distribution of the sample mean ($\bar{X}$).
  - Compute probabilities for ranges of outcomes ($\bar{x}$) and percentiles using when given a sampling distribution, assuming the population from which samples are drawn is normally distributed.

- **Application of the Central Limit Theorem (CLT)**:

  - Assess and determine the conditions under which the distribution of sample averages from an initially unknown or non-normal population distribution can be accurately described by a normal distribution. This includes recognizing the importance of sample size and the role of the CLT in justifying the normal approximation for the distribution of sample means.
  - Clearly articulate situations when the CLT would not apply.
  - Be able to calculate probabilities and percentiles using this information.

## Chapter 8

- **Understanding Observational Studies vs. Experiments**:

  - Clearly differentiate between observational studies and experiments, recognizing the unique insights each can provide.
  - Articulate the reasons behind classifying a study as observational or experimental.
  - Evaluate and identify instances of anecdotal evidence within the context of scientific research.

- **Identifying Components of Experiments**: Accurately identify the experimental units, explanatory variables, treatments or factors, levels, and response variables in various research scenarios.

- **Understanding Experimental Design Graphs**:

  - Master the ability to both draw and interpret experimental design graphs for random experiments, completely randomized, matched pairs, and block designs.
  - Justify the use of matched pair or block designs over completely randomized, understanding the specific conditions that make them preferable.

- **Evaluating Experimental Designs**: Recognize the critical factors in designing an experiment, including control, the principle of control, randomization, and replication, and assess whether an experimental design can be considered good" and the utility of single blind, and double blind experiments and when blinding is not possible.

- **Recognizing Sampling Methods**: Classify a study's sampling method and why one method may be preffered given context.

  - **Non-randomized Methods**: issues with non-random sampling and the different types.
  - **Randomized Methods**: such as simple random sampling and stratified sampling.

- **Identifying Sampling Issues**: Identify common issues in sampling such as bias, convenience sampling, self-selection, undercoverage, and nonresponse, and understand their impacts on study results.

- **Defining and Identifying Lurking Variables**:

  - Define what lurking variables are and identify potential lurking variables in given research contexts.
  - Distinguishing Between Lurking Variables: Differentiate whether a lurking variable acts as a confounding variable or is part of a common response in specific scenarios, and competently draw and interpret the corresponding diagrams.

- **Assessing Causality**: Evaluate the possibility of causality in various contexts, utilizing both statistical data and theoretical knowledge. This includes understanding when statistical inference can suggest a causal relationship and recognizing the limitations of such inferences.

## Chapter 9, 10, and 11

### Confidence Intervals/Bounds

- **Understanding Confidence Intervals and Confidence Bounds**:
  - Accurately define what constitutes a confidence interval and a confidence bound, including their purpose in statistical analysis.
  - Distinguish between point estimates and interval estimates in the context of statistical inference.
- **Interpreting Confidence Interval Results**:
  - Develop the skill to correctly interpret the results of a confidence interval, understanding what the interval range implies about the potential true value of the parameter being estimated.
  - Recognize the implications of different confidence levels (e.g., 95%, 99%) and how they affect the interpretation of confidence intervals.
- **Confidence Level and Random Sampling**:
  - Explain the relationship between the chosen confidence level and the principle of random sampling and the sampling distribution in the construction of confidence intervals.
  - Understand how the confidence level reflects the proportion of confidence intervals, from repeated samples, that are expected to contain the true parameter value.
- **Determining Factors That Control the Width of a Confidence Interval**:
  - Identify the factors that influence the width of a confidence interval, such as sample size, variability of the data, and the chosen confidence level.
  - Analyze the trade-off between the width of the confidence interval and the level of confidence.
- **Assumptions Required for Statistical Inference**:
  - List and understand the assumptions necessary for performing statistical inference, including normality, independence, and sample size.
  - Evaluate whether these assumptions are met in practical scenarios and understand the implications if they are not met.
- **Application of Confidence Intervals**:
  - Be proficient in interpreting computer output related to confidence intervals and know when and how to compute confidence intervals manually using the correct critical values.
  - Apply knowledge of confidence intervals, demonstrating an understanding of their practical importance in statistical inference, using provided graphs and output to support the analysis.
  - Correctly identify which computer output to use for constructing confidence intervals and understand how to apply the central limit theorem in calculating these intervals.
  - You may also need to compute the interval by hand using the correct critical value.
- **Precision of Confidence Intervals**:
  - Be able to calculate the sample size, $n$, needed for a particular level of precision (margin of error) for the one sample situations and determine the final answer. **ME** is the margin of error (half-width).
  - Recognize that preliminary studies or approximations may be necessary in the case where the population standard deviation is unknown.

The table below summarizes the formulas for one-sample z, one-sample t, two-sample t (independent), and two-sample t (matched pair) tests.

**Summary of Confidence Interval Formulas**

| One-Sample Z | One-Sample t | 2-Sample t (Independent) | 2-Sample t (Matched Pair) |
|---|---|---|---|
| $\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ | $\bar{x} \pm t_{\frac{\alpha}{2},n-1} \frac{s}{\sqrt{n}}$ | $\bar{x}_1 - \bar{x}_2 \pm t_{\frac{\alpha}{2},\nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ | $\bar{d} \pm t_{\frac{\alpha}{2},n-1} \frac{s_d}{\sqrt{n}}$ |

**Summary of Sample Size Calculations for Confidence Intervals**

| One-Sample Z | One-Sample t |
|---|---|
| $n = \left(\frac{\sigma z_{\alpha/2}}{ME}\right)^2$ | $n = \left(\frac{s' t_{\alpha/2,n'-1}}{ME}\right)^2$ |

**Summary of Confidence Bound Formulas**

| Test Type | Lower Bound | Upper Bound |
|---|---|---|
| One-Sample Z | $\mu > \bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}$ | $\mu < \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$ |
| One-Sample t | $\mu > \bar{x} - t_{\alpha,n-1} \frac{s}{\sqrt{n}}$ | $\mu < \bar{x} + t_{\alpha,n-1} \frac{s}{\sqrt{n}}$ |
| 2-Sample t (Independent) | $\mu > \bar{x}_1 - \bar{x}_2 - t_{\alpha,\nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ | $\mu < \bar{x}_1 - \bar{x}_2 + t_{\alpha,\nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ |
| 2-Sample t (Matched Pair) | $\mu > \bar{d} - t_{\alpha,n-1} \frac{s_d}{\sqrt{n}}$ | $\mu < \bar{d} + t_{\alpha,n-1} \frac{s_d}{\sqrt{n}}$ |

**Unknwon but equal Variance**

- **Pooled Variance Estimator**:
  - Be able to show that the pooled estimator is unbiased if the equal variance assumption is true.
  - Discuss the ramifications of assuming equal variances incorrectly.
  - Explore how bias in the pooled variance estimator is exacerbated when sample sizes are imbalanced, elucidating the mechanism behind this phenomenon.

**Pooled Variance Estimator**

$$S_p^2 = \left[\frac{n_A - 1}{n_A + n_B - 2}\right] S_A^2 + \left[\frac{n_B - 1}{n_A + n_B - 2}\right] S_B^2$$

**Degrees of Freedom when Equal Variance is Assumed:**

$$\mathbf{df} = n_A + n_B - 2$$

**Unknown and Unequal Variance and the Welch Satterthwaite Approximate Degrees of Freedom**

- **Welch–Satterthwaite Approximation**:
  - Recognize the implications of incorrectly assuming equal variances between two populations.
  - Understand how the Welch t-procedure and the Satterthwaite approximation compares to the use of the pooled variance approach used in a standard two-sample t-test with the assumption of equal variances, and know when each method is appropriate.
  - Understand that the Satterthwaite Approximation is used to approximate the degrees of freedom in the two-sample t-test when the two populations do not have equal variances and the sample sizes might be unequal.
  - Answer questions regarding what happens to the formula if the sample sizes are equal, the variances are equal, or both are equal.
  - Know the limitations of the approximation, such as precision of confidence intervals and potential loss of power compared to when population variances are equal, or when alternative nonparametric tests might be more appropriate.
  - Evaluate the implications of changes in the formula for degrees of freedom on the interpretation inference results, understanding how these changes impact hypothesis testing and statistical conclusions.

**Exact Degrees of Freedom:**

$$\mathbf{df} = \frac{\left(\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}\right)^2}{\frac{1}{n_A - 1}\left(\frac{\sigma_A^2}{n_A}\right)^2 + \frac{1}{n_B - 1}\left(\frac{\sigma_B^2}{n_B}\right)^2}$$

**Welch Satterthwaite Approximate Degrees of Freedom:**

$$\nu = \frac{\left(\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}\right)^2}{\frac{1}{n_A - 1}\left(\frac{s_A^2}{n_A}\right)^2 + \frac{1}{n_B - 1}\left(\frac{s_B^2}{n_B}\right)^2}$$

# Hypothesis Testing, Error, and Power

- **Type I and Type II Errors, and Power**:
  - Comprehend the implications of Type I error (false positive rate, denoted as $\alpha$), Type II error (false negative rate, denoted as $\beta$), and statistical power $(1 - \beta)$ in hypothesis testing.
  - Perform calculations related to Type II error and power for a specified alternative mean $\mu_a$, given sample size $n$, significance level $\alpha$, and null value $\mu_0$, when the alternative hypothesis is clearly defined.
- **Sample Size, Type II Error, and Power**: Understand the direct relationship between sample size ($n$) and its impact on Type II error and power, including how to calculate $n$ for a specified level of power $(1 - \beta)$ or Type II error ($\beta$) for a given alternative mean $\mu_a$, when $\alpha$ and $\beta$ or $1 - \beta$ are provided.

- **Test Statistic and P-value**:
  - Demonstrate clear understanding of test statistics and p-values, including how they are obtained from hypothesis testing, and correct versus incorrect interpretations.
  - Four-Step Hypothesis Testing: Be able to systematically conduct hypothesis tests in four steps, covering parameter identification, hypothesis formulation, calculation of test statistic and p-value, and conclusion writing.

**Test Statistic Formulas:**

| Test Type | Test Statistic Formula | Degrees of Freedom |
|---|---|---|
| One-Sample Z | $z_{ts} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ | $N/A$ |
| One-Sample t | $t_{ts} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ | $n - 1$ |
| 2-Sample t (Independent) | $t'_{ts} = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ | $\nu$ |
| 2-Sample t (Matched Pair) | $\frac{\bar{d} - \Delta_0}{s_D/\sqrt{n}}$ | $n - 1$ |

- **Hypothesis Test and Confidence Intervals/Bounds**:
  - Be able to compare the results of a confidence interval or bound with the associated hypothesis test. This includes how they are similar and how they are different.
  - Be able to determine which confidence interval or bound corresponds to which alternative hypothesis.
  - Given the results of the confidence interval/bound or hypothesis test, be able to predict the results of the other one and when this approach is valid ($C + \alpha = 1$).

**Hypotheses Relationships and Code:**

| Hypothesis Test | Null Hypothesis | Alternative Hypothesis | Confidence Interval or Bound | $p$-value ($z$-code) | $p$-value ($t$-code) |
|---|---|---|---|---|---|
| Upper Tailed | $H_0 : \mu \leq \mu_0$ | $H_a : \mu > \mu_0$ | Lower Bound | $P(Z > z_{ts})$ : $\mathbf{pnorm}(z_{ts}, \text{lower.tail} = \mathbf{FALSE})$ | $P(T > t_{ts})$ : $\mathbf{pt}(t_{ts}, \mathbf{df} = n-1, \text{low}$ |
| Lower Tailed | $H_0 : \mu \geq \mu_0$ | $H_a : \mu < \mu_0$ | Upper Bound | $P(Z < z_{ts})$ : $\mathbf{pnorm}(z_{ts}, \text{lower.tail} = \mathbf{TRUE})$ | $P(T < t_{ts})$ : $\mathbf{pt}(t_{ts}, \mathbf{df} = n-1, \text{lov}$ |
| Two-Tailed | $H_0 : \mu = \mu_0$ | $H_a : \mu \neq \mu_0$ | Interval | $2P(Z > z_{ts})$ : $2\mathbf{pnorm}(\text{abs}(z_{ts}), \text{lower.tail} = \mathbf{FALSE})$ | $2P(T < t_{ts})$ : $2\mathbf{pt}(\text{abs}(t_{ts}), \text{lowe}$ |

## Statistical Significance vs. Practical Significance:

- **Understanding Statistical Significance and Practical Significance**: State and differentiate between statistical significance and practical significance, including how to determine and interpret the practical significance of inference results. Understand that practical difference and effect size are critical in evaluating the real-world importance of statistical findings.

- **Determining Practical Significance**: Given sufficient information, discern and interpret the practical significance of inference results.

- **Interpretation based on Hypothesis Test Outcome**:

  - If the decision is to fail to reject $H_0$, acknowledge the absence of practical significance.
  - If the decision is to reject $H_0$, assess the appropriate difference and effect size.
- **Difference Calculation**: Compute the difference as the absolute value of the difference between the null value and the experimental value (confidence bound or the closest limit of the confidence interval).

- **Effect Size Computation**: Calculate the effect size as the difference divided by the standard deviation.Only for one-sample, two-sample paired, or two-sample independent if equal variance is reasonable.

$$\frac{\text{difference}}{\text{standard deviation}}$$

- **Evaluating Difference and Effect Size**:
  - For the difference:
    - If the difference is small, consider the values practically the same.
    - If the difference is large, deem the values practically different.
    - Stake holders deem what differences are large or important.
    - On exams, indicate which differences are practically equivalent.
- **For the effect size**:
  - If the effect size is small, regard the values as practically indistinguishable.
  - If the effect size is large, perceive the values as practically distinct.
  - Remember practical significance is primarily determined by stake holders.
  - In the absence of stakeholder guidance, adhere to rule-of-thumb thresholds: less than 0.2 (small), 0.2 – 0.8 (moderate), greater than 0.8 (large).
- **Final Interpretation**: Recognize that practical difference only exists when both the difference and effect size indicators are large. Therefore, if one indicator is large, evaluate the other to determine the final conclusion.