STAT 5 Final Study Guide

Antonio Aguirre

Department of Statistics, University of California, Santa Cruz
Spring 2025

Purpose of this Guide

This document outlines the most important topics and formulas to review for the STAT 5 final exam. Use it as a roadmap to organize your studying and focus on essential concepts.

Study Topics and Formulas

Descriptive Statistics

- **Sample Mean:** $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- **Sample Variance:** $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$
- **Standard Deviation:** $s = \sqrt{s^2}$

*Know how to compute and **interpret** these summary statistics.*

Probability

- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cap B) = P(A|B) \cdot P(B)$

Be comfortable applying these rules in different scenarios, including when events are independent or disjoint

Normal Distribution

- **Standard Normal:** $Z = \frac{X - \mu}{\sigma}$, where X follows a $N(\mu, \sigma^2)$
- **Empirical Rule:** 68%, 95%, 99.7% chance within 1σ, 2σ, 3σ of μ.

Be able to standardize data and use the empirical rule.

Sampling and Study Design

- Distinguish between sampling methods: simple random, stratified, cluster, convenience.
- Understand controlled experiments vs. observational studies.

Recognize implications for statistical inference.

Point Estimates

- **One Mean:** $\hat{\mu} = \overline{x}$
- **One Proportion:** $\hat{p} = \frac{X}{n}$, X being the number of subjects in the samples that have a specific characteristic.
- **Two Proportions:** $\hat{p}_1 - \hat{p}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$, X_1 and X_2 being the number of subjects in the sample 1 and 2 that have a specific characteristic.

Understand what point estimators represent and how to compute them.

Confidence Intervals

- **One Mean:** $\overline{x} \pm z^* \frac{\sigma}{\sqrt{n}}, \quad \overline{x} \pm t^* \frac{s}{\sqrt{n}}$
- **One Proportion:** $\hat{p} \pm z^* SE(\hat{p})$
- **Two Proportions:** $(\hat{p}_1 - \hat{p}_2) \pm z^* SE(\hat{p}_1 - \hat{p}_2)$
- $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- $SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Be able to interpret confidence intervals and margin of error.

Hypothesis Tests

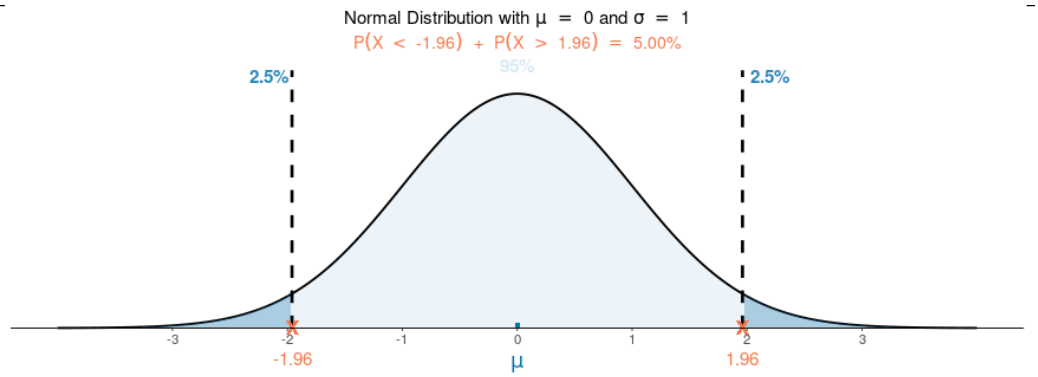
- **One Mean:** $Z = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}, \quad t = \frac{\overline{x}-\mu_0}{s/\sqrt{n}}$
- **One Proportion:** $Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
- **Two Proportions:** $Z = \frac{(\hat{p}_1-\hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, where $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$

Know how to set up hypotheses, calculate test statistics, and interpret p-values.

Significance Level and Statistical Errors

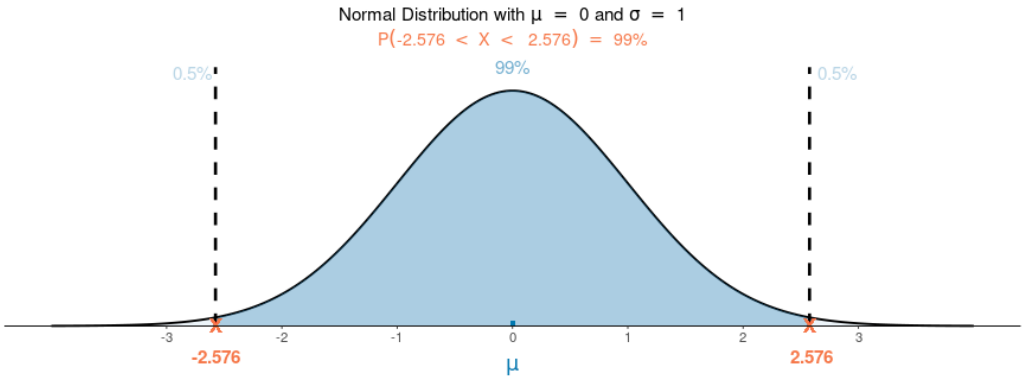
- Understand the meanings of significance level (α), Type I error, and Type II error.

Be able to define and recognize the consequences of errors in hypothesis testing.



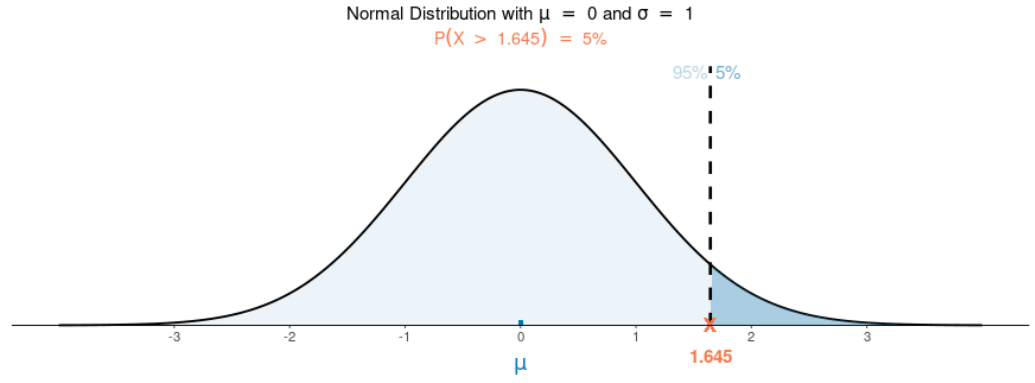
Normal Probability (Tail Probability):

Mean μ	Std. Dev. σ	Value of x1	Value of x2	Probability P(X < -1.96) + P(X > 1.96)	Lower z-Score	Upper z-Score
0	1	-1.96	1.96	5.00%	-1.96	1.96



Normal Percentile (Two-Tailed):

Mean μ	Std. Dev. σ	Central Percent	Value of x1	Value of x2
0	1	99%	-2.576	2.576



Normal Percentile (Upper Tail):

Mean μ	Std. Dev. σ	Percent in Upper Tail	Value of x
0	1	5%	1.645

Figure 1: Plots for p-value extraction

Regression

- $\hat{y} = b_0 + b_1 x$
- $b_1 = r\frac{s_y}{s_x}$, $b_0 = \overline{y} - b_1\overline{x}$
- $R^2 = \frac{\text{Var}(y) - \text{Var}(e)}{\text{Var}(y)} = r^2$ for simple linear regression.

Interpret the regression line, slope, intercept, and R^2.

Reading and Interpreting Data Tables

- Extracting relevant proportions or probabilities from contingency tables.
- Translating tabular information into meaningful statistical statements.

Assumptions for Statistical Inference

- Know when inferential methods are valid: random sampling, independence, sample size/normality.

Communicating Statistical Findings

- Explaining results in plain language, especially for non-technical audiences.
- Connecting statistical output to real-world questions.

Advice for Effective Studying

- Focus on understanding the "why" behind each concept, not just memorizing formulas.
- Practice stating hypotheses and interpreting statistical results in context.
- Be able to explain statistical reasoning in clear, everyday language.
- Review worked examples and try to solve a variety of practice problems.
- Don't hesitate to ask questions if any concept is unclear!

Antonio Aguirre
Department of Statistics, UCSC