STAT 3011 (006) Fall 2022 Final Exam Information

Yu Yang

1 General Information About the Exam

1. **Final Exam will be held on Tuesday, December 20th, 5:00 pm - 7:00 pm, at Anderson Hall 310 (West Bank).**
2. Final Exam has 17 multiple choice questions (worth 50% of the exam score, including one question asking "did you circle your multiple-choice answers on the last page of the exam?" See item 9) and 3 short answers questions with sub-questions (worth 50% of the exam score). Final Exam will be based on all chapters that we learned in class, lab, and homework.
   - 5 - 8 multiple choice questions from Chapter 1, 2, 4, 5, 6, 7
   - 5 - 7 multiple choice questions from Chapter 8, 9, 10
   - 3 - 5 multiple choice questions from Chapter 3, 11, 12, 14
   - 3 short answer problems from Chapter 3, 11, 12, and 14
3. You can bring **TWO letter-sized (8.5" by 11")** cheat sheet either typed or hand-written to Final Exam. You may use both sides of your cheat sheet. You can use any content from Canvas/homework/lecture, etc. **Please write your name on both sides of your cheat sheet. Please note that your cheat sheet will be collected with your exam and will not be returned. If you want to keep your cheat sheet, please make a photocopy of it before the exam.**
4. Bring a scientific or graphic calculator. Phone/computer calculators are NOT allowed. If you have a graphing calculator with a color screen, please plan to seat in the front. I do not want to wrongly accuse you of using a cell phone during the exam and it is hard to tell from far away. **If you use a graphic calculator, please only use the basic algebraic functions and do not use the fancy functions that a scientific calculator does not provide.**
5. Bring your U-card or photo ID. Please have your ID out while taking the exam.
6. I will provide scratch paper if needed. Do not bring your own.
7. When you finish your exam :
   - Hand in your exam & cheat sheet to the teaching staff (instructor/TAs).
   - Show your ID.
   - Find your name on the class roster and write your initial.
   - Please DO NOT discuss the exam with other students when you are in line waiting to hand in your exam.
8. Sharing cheat sheets and/or sharing a calculator during the exam is NOT allowed.
9. Please circle your multiple choice answer on the last page of the exam. Failing to do so will result 0 for your multiple choice part.
10. All necessary R command results will be provided. But not all are useful.

2 Topics Covered

1. Chapter 14 Analysis of Variance
   - Can compare between-group variability and within-group variability from the side-by-side boxplot
   - Know that F-test stat is defined as (between group variability)/(within-group variability) and the meaning of large F-test statistic
   - Can conduct five-step ANOVA hypothesis testing
     - Know the assumptions of ANOVA F-test and correctly determine whether each assumption is met or not from the problem description and/or boxplot.
     - State the null and alternative hypotheses using correct notations.
     - Test statistic
       - Know the distribution of ANOVA F-test statistic with correct degrees of freedom under H_0
       - Can fill in ANOVA table
       - Find/calculate the F-test statistic from ANOVA table
     - P-value : calculate p-value using `pf()` command, or find p-value from `summary(aov())` R output
     - Conclusion : draw a conclusion and interpret in the context of the problem.
   - Know the properties of F-distribution
   - Tukey's HSD
     - Know when to perform Tukey's HSD
     - Given Tukey's HSD R output, can identify pairs that are significantly different and interpret the confidence interval in context.
2. Chapter 11 Association Between Two Categorical Variables
   - Can conduct five-step Chi-squared test for independence
     - State and check assumptions
     - State null and alternative hypothesis
     - Test stat :
       - Calculate expected cell counts
       - Know the distribution of the test statistic with correct degrees of freedom under H_0.
       - Calculate the value of test statistic
     - P-value : Use `1-pchisq(`X^2`, df=(r-1)(c-1))` or `pchisq(`X^2`, df=(r-1)(c-1), lower.tail=FALSE)` to find p-value
     - Draw a conclusion and interpret it in the context of problem.
   - Know the properties of Chi-squared distribution
   - Risk and relative risk
     - Risk: can calculate risk for each group from a contingency table
     - Relative risk: can calculate relative risk and interpret in the context of the problem.
3. Chapter 12 Regression Analysis
   - Correlation
     - Interpret correlation and scatter plot
     - Know the properties of correlation (r)
     - Know the relationship between correlation (r) and slope (b)
     - Know the relationship between correlation (r) and r-squared (r^2), note: $r = sign(b)\sqrt{r^2}$.

- Least squares regression
  - From `summary(lm(y ~ x))` output :
    - Can find y-intercept (a) and slope (b) and write out the sample regression equation
    - Can find and interpret r^2
  - Can interpret the slope in context of the problem
  - Can calculate slope and y-intercept using when the sample means ($\overline{x}, \overline{y}$), sample standard deviations (s_x, s_y), and correlation (r) are given
  - Know the definition of residual (observed y - predicted y), and calculate residual of a particular observation
  - Know and understand properties least square regression line
  - Know what extrapolation means and how to tell from scatter plot
  - Know what lurking variable is and can tel from a given scenario
- Inference of beta (β)
  - Know assumptions of regression model and inference, can determine whether each assumption is met based on corresponding plot
  - Can calculate and interpret the confidence interval for β
    - From R `summary(lm(y ~ x))` output, can find b and the standard error of b
    - Can find t-multiplier value using R command `qt(1-`$\alpha/2$`, df=n-2)` or `qt(`$\alpha/2$`, df=n-2)`
  - Can conduct five step hypothesis test for β
    - State null and alternative hypotheses using statistical notations
    - From R `summary(lm(y ~ x))` output, can find test stat and p-value
    - Draw a conclusion and interpret in context of the problem

4. Chapter 1 - 10

(a) Chapter 1 Introduction

- Know the definitions of sample, population, statistic, parameter and can identify them from a given scenario.

(b) Chapter 2 Exploring Data

- Identify different types of data (categorical, quantitative (discrete, continuous))
- Understand properties of numerical summaries (sample mean, median, range, IQR, standard deviation, etc.) including whether it is resistant to outliers or not.

(c) Chapter 4 Gathering Data

- Know the definition of explanatory variable, response variable, and identify each from a given scenario.
- Know types of study (randomized experiment vs observational study) and whether we can establish a cause-and-effect relationship.

(d) Chapter 5 Probability

- Understand definitions/meaning of intersection and union.
- Know how to use general addition rule
- Understand meaning of two events being disjoint
- Know how/when to use addition rule for disjoint events
- Understand meaning of two events being independent
- Know how/when to use multiplication rule for two independent events

(e) Chapter 6 Probability Distribution

- Understand properties of normal distribution/standard normal distribution
- Understand and use the 68-95-99.7 rule to calculate approximate probabilities of normal distributions

- Calculate z-score of x, and interpret.
- Find probabilities standardizing x into z-distribution and then `pnorm()` (or `1-pnorm()`) command.
- Find a value of z-score/quantile that marks the p-th percentile using R command `qnorm(p)` and interpret
- Determine whether a distribution is normal or not using Q-Q plot
- Know the conditions of binomial probability distribution
- When conditions are met, can identify parameters of binomial distribution (n=number of trials, p=probability of success as in $X \sim Binom(n, p)$)
- Calculate binomial distribution probabilities

(f) Chapter 7 Sampling Distributions

- Understand meaning of sampling distribution of sample statistic and distinguish it from population distribution/probability distribution
- The sampling distribution of the sample proportion
  - Understand properties of, and know how to calculate the mean of (the sampling distribution of) the sample proportion and the standard deviation of (the sampling distribution of) the sample proportion
  - Know when the sampling distribution of sample proportion is approximately normal (sample size is large enough so that) (Central Limit Theorem).
  - Understand the effect of sample size on sampling distribution of a sample proportion (shape, center, spread)
- The sampling distribution of the sample mean
  - Understand properties of, and know how to calculate the mean of (the sampling distribution of) the sample mean and the standard deviation of (the sampling distribution of) the sample mean
  - Understand the effect of sample size on the sampling distribution of a sample mean (shape, center, spread)
  - Know when the sampling distribution of sample mean is normal / approximately normal
    i. when population distribution is normal
    ii. when population is unknown/not normal, if n is large (Central Limit Theorem)

(g) Chapter 8 Confidence Intervals

- Understand the definitions/notations//meaning of the point estimate vs interval estimate, sample statistic, parameter
- Understand the definition of the standard error of a sample statistic and how to calculate.
- Understand properties of t-distribution
- Know the assumptions of the confidence interval and determine whether they are met or not
- Obtain the correct z-multiplier (in R: `qnorm(1-`$\alpha/2$`)` or `qnorm(`$\alpha/2$`, lower.tail=FALSE))` to construct a confidence interval for p (population proportion)
- Obtain the correct t-multiplier (in R: `qt(1-`$\alpha/2$`, df=n-1)` or `qt(`$\alpha/2$`, lower.tail=FALSE)` to construct a confidence interval for μ (population mean).
- Interpret the confidence interval in the context of the problem

(h) Chapter 9 Hypothesis Tests

- Conduct a five-step hypothesis test
  - Understand the nature of null and alternative hypothesis using correct notations
  - Know the formula of the test statistic and its distribution under the null hypothesis
  - Calculate and interpret z-test statistic for proportion, t-test statistic with correct degrees of freedom for mean.
  - In testing population mean(s), use t-distribution with correct degrees of freedom

    - for one sample t-test for μ, df=n-1,
    - for independent two samples t-test for $\mu_1 - \mu_2$, df=$\min(n_1 - 1, n_2 - 1)$
    - for matched pair t-test for μ_D, df=$n_D - 1$
  - Know how P-value is defined for each alternative hypothesis ($<, >, \neq$) and find "new p-value" under a new alternative hypothesis (using the relationship between one-sided p-value and two-sided p-value).
  - Draw a conclusion and interpret in the context of the problem.
- Understand the equivalence between two-sided hypothesis tests and confidence interval for μ
- Apply Type I error and Type II error in different scenarios and know a possible error you could have made after drawing a conclusion.

(i) Chapter 10 Comparing Two Groups

- Find point estimate of $\mu_1 - \mu_2$
- Calculate and interpret the confidence interval for $\mu_1 - \mu_2$ in context (when CI contains 0 vs when CI does not contain 0).
- Determine whether two samples are matched pairs or independent and find the correct degrees of freedom in each case
- Understand the equivalence between CI for $\mu_1 - \mu_2$ and two-sided hypothesis test.