0.1 Final Exam General Info

The exam will have 5 main questions, each having multiple parts, as well as a few multiple choice or true/false questions. There will be a mix of conceptual questions, theoretical derivations, and data analysis. You should review the homeworks, midterm, lecture notes, and relevant sections in the text book. Questions will be similar in style to homework and midterm questions.

You are allowed two pages of notes, front and back. You are not allowed a calculator or any other electronic devices.

Question topics:

1. One-way ANOVA.
2. Two-way ANOVA.
3. Simple linear regression.
4. Multiple linear regression.
5. Categorical data analysis.

0.2 Hypothesis Testing

1. Definition and interpretation of P-values.
2. Definition of Type I error, Type II error, and Power.
3. How does power depend on the sample size, the effect size, and the type of test used (e.g. one-tailed vs two-tailed).
4. One- and two-sample t-tests.
5. Definition of expectation, variance, and covariance.
6. Definition of an unbiased estimator.
7. Likelihood ratio test.
8. Confidence interval interpretation.
9. Bonferroni correction for multiple comparisons.

0.3 One- and Two-way ANOVA

1. Model assumptions and constraints. Possible effects
2. Decomposition of variance $SSTOT = SSR + SSE$.
3. ANOVA Table for balanced and unbalanced designs for one-way ANOVA.
4. ANOVA table for balanced designs for two-way ANOVA.
5. Hypothesis testing equality of means, F-test, relationship to likelihood ratio test.
6. Interpretation of parameters, interpretation of main effects and interaction effects.
7. Confidence interval for group means.
8. Estimating the variance: MLE, and unbiased estimator.

0.4 Simple Linear Regression

1. Model statement, assumptions, parameter interpretations.
2. Likelihood, computing MLEs, relationship to least squares estimates.
3. Hypothesis testing $H_0 : \beta_1 = 0$.
4. Distribution of $\hat{\beta}_1$ and $\hat{Y}_h$.
5. Confidence intervals and prediction intervals.
6. ANOVA-style decomposition of variance $SSTOT = SSR + SSE$.
7. Global F-test, hypothesis statement, reference distribution, interpretation.
8. Why is the constant variance assumption important? What happens if it is violated?
9. Why is the normality assumption **not** that important, assuming we have a large sample size?
10. Diagnostics, how can we check the model assumptions?
11. What can we do if constant variance is violated?

0.5 Multiple Linear Regression

1. Observation form and matrix form of the multiple linear regression model.
2. Model assumptions and parameter interpretations.
3. Definition of a projection matrix. Role of projection matrices in estimating $\hat{\beta}$ and the residuals.
4. Variance stabilizing transformations.
5. Types of covariates: confounders, precision, and unrelated.
6. Gauss-Markov theorem: the MLE $\hat{\beta}$ from linear regression is the best linear unbiased estimate. What do "best", "linear", and "unbiased" mean in this context?

0.6 Categorical Data Analysis

1. Pearson's chi-squared tests for independence and homogeneity. Null hypotheses, differences in the two tests, formula for test statistics, interpretation of hypothesis test conclusions.
2. Theoretical justification of the chi-squared test from the Central Limit Theorem. Need for large sample size.
3. Multinomial likelihood.
4. Relationship of Pearson's chi-squared tests to the likelihood ratio test.
5. Fisher's exact test: hypergeometric distribution, null hypothesis, test statistic, interpretation, differences with Pearson's chi-squared tests.
6. Definition of the odds and odds ratio, interpretation and differences of these.
7. Advantages and disadvantages of the three study designs discussed: (i) simple random sample, (ii) prospective study, (iii) retrospective study.
8. Odds ratio confidence interval.