STA 207 Practice Midterm

10:00 am to 10:50 pm, Feb 26th 2020

Print name:____

Print student ID:__________________________________

(Initial) ____ I will not disclose the content of the exam to students not in this class.

(Initial) ____ I have read the instructions on the second page of the exam.

Sign name: ____

The best thing about being a statistician is that you get to play in everyone's backyard.

John Tukey

	Total	Part I	Part II
Grade			

General Instructions:

1. Do not start the exam until your instructor authorizes you to do so.
2. All models and numbers are fictional in this exam.
3. This exam is closed book/closed notes. However, you are allowed to consult your written **one-page** summary.
4. Calculators are allowed, but **no** other computing devices (including cellphones).
5. Scratch papers are provided. You may remove these papers during the exam. If you remove those pages, please, do not hand them in.
6. You need to write down the full derivations to receive full credits.
7. Partial credits can only be given if your thoughts and handwriting can be followed.
8. Answer the question that is asked – you do not need to give reasons unless you are specifically asked. Writing **too much** and touching on topics not immediately relevant to a question will count against you. For example, a correct answer buried within irrelevant, incorrect or incomprehensible information will not receive credits.
9. **Answers written with pencils or erasable pens will not be accepted**.
10. **Only answers within the given box for each question will be considered.**

Part I: ANOVA with summary statistics (80 pts)

A client of yours wants to find out the best microbial environment for *C. elegans*. In previous meetings, the client told you that *C. elegans* feed on bacteria but may also be killed by certain bacteria. Therefore, it is important to figure out what bacteria are beneficial to *C. elegans*. In particular, the client was interested in studying the association between the densities of ***Gluconobacter***, ***Enterobacter***, and ***C. elegans***. The client had collected some pilot data for this study from $n_T = 36$ agar plates. To be specific, the client considered three concentration levels of ***Gluconobacter***, where the density of ***Gluconobacter*** is low, medium, or high. In a similar manner, the client considered three levels for the density of ***Enterobacter*** (low, medium or high). There were 4 agar plates in each combination of the two factors. Each plate was initialized with a fixed density of *C. elegans* (5×10^{-4} g/ml). After **10 days** since the onset of experiments, the client measured the density of *C. elegans* in each agar plate, which is our response.

However, the records of data set was accidentally deleted. The client brought to the meeting a table (see Table 1) that summarizes the mean densities of *C. elegans* under each combination. Your task is to help the client make as much as possible out of this table.

		Gluconobacter		
		Low	Medium	High
Enterobacter	Low	8.76	3.47	1.69
	Medium	11.75	8.20	7.17
	High	17.34	13.56	10.04

Table 1: Mean densities of *C. elegans* in the experiment (unit: 10^{-4} g/ml).

For your convenience, some summary statistics of Table 1 are as follows

$$3\sum_i (\bar{Y}_{i,\cdot} - \bar{Y}_{\cdot,\cdot})^2 = 121.6, \quad 3\sum_j (\bar{Y}_{\cdot,j} - \bar{Y}_{\cdot,\cdot})^2 = 62.0, \quad \text{and} \sum_i \sum_j (Y_{i,j} - \bar{Y}_{\cdot,\cdot})^2 = 186.8,$$

where we use Y to denote the number in Table 1 and i indicates the row, and j indicates the column.

1. (15 pts) Write down a two-way ANOVA model for the data in Table 1. For consistency, choose the letters from $\{Y, \alpha, \beta, \mu, \epsilon\}$ and use the factor-effect form. (Hint: Only the means are available to you in Table 1.)

2. (0 pts) Plot the means in order to investigate if the interaction effects are present. Also, obtain the plot to investigate the presence of main effects. What can you see from these plots? (Hint: this question will not be graded, but your answers will guide you through the remaining questions.)

3. (15 pts) Set up the ANOVA table for your model, include the F-statistics.

4. (15 pts) The client managed to recover part of an ANOVA table from an analysis done on the full data (see Table 2). Complete the ANOVA table for the client, include the F-statistics.

Source	d.f.	S.S.	M.S.	F
Enterobacter		486.3		
Gluconobacter		248.0		
Interaction		12.9		
Residuals				
Total		776.3		

Table 2: ANOVA table evaluated on the lost raw data.

5. (9 pts) Write down the factor-effect model that might generate the above ANOVA table.

6. (8 pts) Is there any terms in the new model that can be dropped, at the significance level $\alpha = 0.01$?

7. (10 pts) Carry out a test to decide if the effect of the densities of ***Enterobacter*** is present on the full data set, at the significance level $\alpha = 0.01$.

8. (8 pts) The client wanted to find out if there is a combination of the densities of ***Enterobacter*** and ***Gluconobacter*** that leads to the highest density of *C. elegans*. Find this combination using an appropriate statistical method (use $\alpha = 0.01$).

Part II: ANOVA with no numbers (20 pts)

Consider a one-factor ANOVA model with fixed effects

$$Y_{i,j} = \mu + \alpha_i + \epsilon_{i,j},\ j = 1, \ldots, n_i, i = 1, \ldots, r,$$

where $\{\alpha_i\}$ satisfies that $\sum_i n_i \alpha_i = 0$ and $\{\epsilon_{i,j}\}$ are i.i.d. $N(0, \sigma^2)$.

1. (7 pts) Write down the least squares estimate of α_i. Find the mean and variance of this estimate in terms of $\{n_i\}$ and the parameters of the model.

2. (7 pts) Suppose that we are interested in a contrast $L = \sum_{i=1}^{r} c_i \alpha_i$ where $\sum_{i=1}^{r} c_i = 0$. Write down the least squares estimator $\hat{L}$ of L. Find the mean and variance of $\hat{L}$ in terms of $\{c_i\}$ and parameters of this model.

3. (6 pts) For this part, assume that the factor effects are random, where $\{\alpha_i\}$ are i.i.d. $N(0, \sigma_\alpha^2)$, $\{\alpha_i\}$ are independent of $\{\epsilon_{i,j}\}$, and the model is unbalanced. Find the mean and variance of $\hat{L}$ in Part II Q2.

Quantiles of F-distribution

$$F(1-0.1;2,4)=4.32,\quad F(1-0.05;2,4)=6.94,\quad \text{and } F(1-0.025;2,4)=10.65.$$

$$F(1-0.1;1,3)=5.54,\quad F(1-0.05;1,3)=10.13,\quad \text{and } F(1-0.025;1,3)=17.44.$$

$$F(1-0.01;4,27)=4.10,\quad \text{and}\quad F(1-0.005;4,27)=4.74.$$

$$F(1-0.01;6,27)=3.55,\quad \text{and}\quad F(1-0.005;6,27)=4.06.$$

Quentiles of studentized range distribution

$$q(1-0.01;9,27)=5.72\quad q(1-0.005;9,27)=6.11$$

(Scratch paper)

(Scratch paper)

(Scratch paper)