STATISTICS 642 - EXAM 1

February 26, 2020

Student's Name _____

Student's Email Address _____

INSTRUCTIONS FOR STUDENTS:

(1) The exam consists of 5 pages of questions including the cover page and SAS code, 6 pages of SAS output, and 21 pages of STAT 642 Tables.

(2) You have exactly **70 minutes** to complete the exam and 20 minutes for downloading exam and uploading solutions.

(3) **Exam Period:** A period of 90 minutes Starting at 11:30 a.m. CDT Feburary 26 and Concluding at 1:00 p.m. CDT February 26.

(4) You MUST upload your solutions to the exam by 1:00 p.m. CDT February 26.

(5) The exam will be proctored using Zoom. You will have two options for composing your solutions:

- You can print out the exam, put your solutions on the exam, scan pages 1-5 into a single pdf file, then upload the pdf just as you would with homework solutions. to eCampus.
- Alternatively, you can just write your solutions on paper, scan the solutions into a single pdf file, then upload the pdf file just as you would with homework solutions.

(6) Upload just your solutions, **Do Not upload the SAS Output.**

(7) Do not discuss or provide information to anyone concerning the questions on this exam or your solutions until I post the solutions to the exam.

(8) You may use the following:

- Calculator
- Summary Sheets - **6-pages,** 8.5"**x**11"**, (you may write/type/cut-paste anything on both sides of the six sheets**)
- The STAT 642 Exam Tables

(9) Do not use any other written material except for your summary sheets and STAT 642 Exam Tables. Do not communicate with anyone during the exam. The solutions must be just your own work.

I attest that I spent no more than 70 minutes to complete the exam. I used only the materials described above. I did not receive assistance from anyone during the taking of this exam.

Student's Signature_____

Problem I. (44 points) An experiment was designed to evaluate the differences in the effectiveness of three growth stimulants (S_1, S_2, S_3) on the growth of four varieties of rose plants (V_1, V_2, V_3, V_4). There were two greenhouses available for the study with 8 benches in each of the greenhouses. In both greenhouses, two benches were randomly assigned to each of the four rose varieties with 30 plants per bench. Each bench was subdivided into three regions of 10 plants each in a north to south direction with the regions randomly assigned to the types of stimulants. The initial height of each plant was recorded due a considerable difference in the plants prior to applying the stimulant. Two weeks after applying the stimulant, the height of each plant is recorded. The floriculturist's goal was to determine which stimulant produced the most robust growth in the rose plant and whether the best stimulant depended on the variety of rose plant.

	Greenhouse 1								Greenhouse 2							
Region	B1	B2	B3	B4	B5	B6	B7	B8	B1	B2	B3	B4	B5	B6	B7	B8
R1	V2S1	V1S3	V2S2	V3S1	V3S1	V4S1	V1S2	V4S1	V3S1	V4S3	V3S2	V2S1	V2S3	V1S1	V4S2	V1S1
R2	V2S2	V1S1	V2S3	V3S3	V3S2	V4S3	V1S1	V4S3	V3S2	V4S2	V3S1	V2S2	V2S1	V1S3	V4S3	V1S3
R3	V2S3	V1S2	V2S1	V3S2	V3S3	V4S2	V1S3	V4S2	V3S3	V4S1	V3S3	V2S3	V2S2	V1S2	V4S1	V1S2

Provide the details for each of the following items:

1. Type of Randomization:

2. Type of Treatment Structure:

3. Identify each of the factors as being Fixed or Random:

4. Describe the experimental units:

5. Describe the measurement units:

6. Identify any covariates:

Problem II. (36 points) Caffeine is a drug that affects the central nervous system. The main goal of this experiment is to determine whether the amount of caffeine in the nervous system alters the ability of similar compounds to move across the blood-brain barrier? The experiment involved randomly assigning 8 lab rats to each of six levels of caffeine along with an arterial injection of C^{14}-labeled adenine. Shortly after injection, the concentration of labeled adenine in the rat brains is measured. The measurements from several of the rats were not obtained.

Caffeine (mM)	Adenine								n_i	$\bar{y}_{i.}$	S_i
0	6.6	7.6	6.9	5.9	6.7	6.1	6.6	6.3	8	6.59	.525
5	5.6	5.0	5.3	5.9	6.0	4.8			6	5.43	.484
10	4.7	4.9	4.6	5.5	4.4	5.1	5.0		7	4.89	.363
15	4.2	4.2	3.4	4.5	4.8	4.4	4.6	3.7	8	4.23	.468
20	3.3	3.5	3.7	3.6	3.2	3.3	2.6		7	3.31	.363
25	1.8	1.7	1.9	1.9	1.5	1.2	1.7		7	1.67	.250
Total									43	4.38	.422

Use the attached SAS output to assist you in answering the following questions. Provide a justification for each of your answers.

1. Separate the Caffeine levels into groups of levels such that all members of the group are not significantly different from any other member of the group with respect to their average Adenine concentration. The researcher wants an experimentwise error rate of at most $\alpha = 0.05$ in reaching your conclusions?

2. The researcher wants to determine if the average Adenine concentration is a decreasing function of the level of Caffeine in the rats. Which ones, if any, of the trends in the average Adenine concentration as a function of Caffeine level are significant? The researcher wants an experimentwise error rate of at most $\alpha = 0.05$ in reaching your conclusions?

3. The researcher has decided to design a new experiment to evaluate the concentration of vitamin B_4 in the brains of the rats using the same six Caffeine concentrations. The researcher states that she wants a probability of at least 0.9 to declare a difference in the average vitamin B_4 concentrations if any pair of Caffeine levels have a difference in their vitamin B_4 concentrations of at least 1.4 units. She also wants the probability of falsely declaring a difference in the six average vitamin B_4 concentrations to be at most $\alpha = .01$. The researcher is willing to specify that $\sigma_e^2 \approx 0.4$. A statistician is consulted and states that the necessary number of replications per Caffeine concentration is 8. If the experiment uses 8 rats at each of the 6 Caffeine levels, will the researcher's goal be achieved? Justify your answer.

III. **(20 points) CIRCLE (A, B, C, D, or E) corresponding to the BEST answer for each question.**

(1) A split-plot analysis was conducted for an experiment having a treatment with t levels. Each of the EU's is measured at 5 specified locations on the EU after the treatment was applied. The treatment was designated as the whole plot treatment and the location at which the measurements were taken was designated as the split-plot factor. What is the major problem in analyzing this experiment as a split plot experiment.

A. The correlation between the 5 measurements on each EU may not be equal.

B. The 5 measurements on each EU should be considered as a covariate and the adjusted means should be analyzed.

C. The 5 measurements on each EU are subsamples and hence are a random effect.

D. In a split plot experiment, the levels of the split plot factor have to be randomized over the subplots.

E. None of the above

(2) A completely randomized design with the treatment having 5 equally spaced quantitative levels. Which of the following contrasts would be an evaluation of an increasing trend in the 5 treatments?

A. $C = -2\mu_1 - \mu_2 + \mu_4 + 2\mu_5$

B. $C = 2\mu_1 - \mu_2 - 2\mu_3 - \mu_4 + 2\mu_5$

C. $C = -\mu_1 + 2\mu_2 - 2\mu_4 + \mu_5$

D. $C = \mu_1 - 4\mu_2 + 6\mu_3 - 4\mu_4 + \mu_5$

E. All of the above contrasts would be appropriate

(3) A veterinarian wants to investigate $t = 10$ treatments for controlling heartworms in puppies. Her consulting statistician determines that she will need $r = 9$ replications per treatment. There is enormous variation in the effectiveness of the treatment so the veterinarian wants to use groups of homogeneous puppies and decides to use litters of puppies as her blocking variable. Most litters contain fewer than 10 puppies so she decides to use a BIBD, with at most 9 puppies per litter. Which of the following combinations of b litters and k puppies per litter would yield the most effective design?

A. $b = 45$ and $k = 2$

B. $b = 30$ and $k = 3$

C. $b = 18$ and $k = 5$

D. $b = 10$ and $k = 9$

E. All of the above are equally effective because they all produce a feasible BIBD.

(4) In a complete randomized design with a single factor having t levels with n_i reps for the ith treatment, the following sum of squares: $\sum_{i=1}^{t}\sum_{j=1}^{n_i}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{t} n_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$ is used to evaluate

A. the variability within the t population distributions

B. the variability between the t treatment means

C. the variability of the response variable across all t treatments

D. pooled estimate of the population variances

E. none of the above

(5) You work for a company and design a completely randomized design with $t = 5$ fixed effects treatments and $n_1 = 6$, $n_2 = 3$, $n_3 = 5$, $n_4 = 3$, $n_5 = 3$, reps/treatment. You fit the following model to the data: $Y_{ij} = \mu + \tau_i + e_{ij}$. The company has its own in house statistical program that yields the following least squares estimates of five of the six parameters in the model:

$$\hat{\mu} = 2.7; \quad \hat{\tau}_1 = -1.6; \quad \hat{\tau}_2 = 2.5; \quad \hat{\tau}_3 = 3.6; \quad \hat{\tau}_4 = -6.2$$

The least squares estimate of τ_5 is given by

A. $\hat{\tau}_5 = 0$

B. $\hat{\tau}_5 = 1.7$

C. $\hat{\tau}_5 = 0.9$

D. The program has an error, $\hat{\tau}_i$ is not estimable if sample sizes are unequal.

E. Not enough information has been provided to answer the question.

SAS Program:

```
data new; array Y Y1-Y8;
input C $ Y1-Y8; do over Y; A=Y; output; end;
      drop Y1-Y8;
      label C = 'Caffeine' A = 'Adenine Concentration';
cards;
  0        6.6  7.6  6.9  5.9 6.7 6.1 6.6 6.3
 05        5.6  5.0  5.3  5.9 6.0 4.8  .   .
 10        4.7  4.9  4.6  5.5 4.4 5.1 5.0  .
 15        4.2  4.2  3.4  4.5 4.8 4.4 4.6 3.7
 20        3.3  3.5  3.7  3.6 3.2 3.3 2.6  .
 25        1.8  1.7  1.9  1.9 1.5 1.2 1.7  .
run;
PROC GLM ORDER = DATA;
CLASS C;
MODEL A = C/ss3;
LSMEANS C/PDIFF ALPHA=.05;
MEANS C/DUNNETT('0') ALPHA=.05;
LSMEANS C/PDIFF ADJUST = TUKEY;
contrast 'LINEAR'    C -5 -3  -1  1  3 5;
contrast 'QUADRATIC' C  5 -1  -4 -4 -1 5;
contrast 'CUBIC'     C -5  7   4 -4 -7 5;
contrast 'QUARTIC'   C  1 -3   2  2 -3 1;
contrast 'QUINTIC'   C -1  5 -10 10 -5 1;
run;
```

642 EXAM I Spring 2020

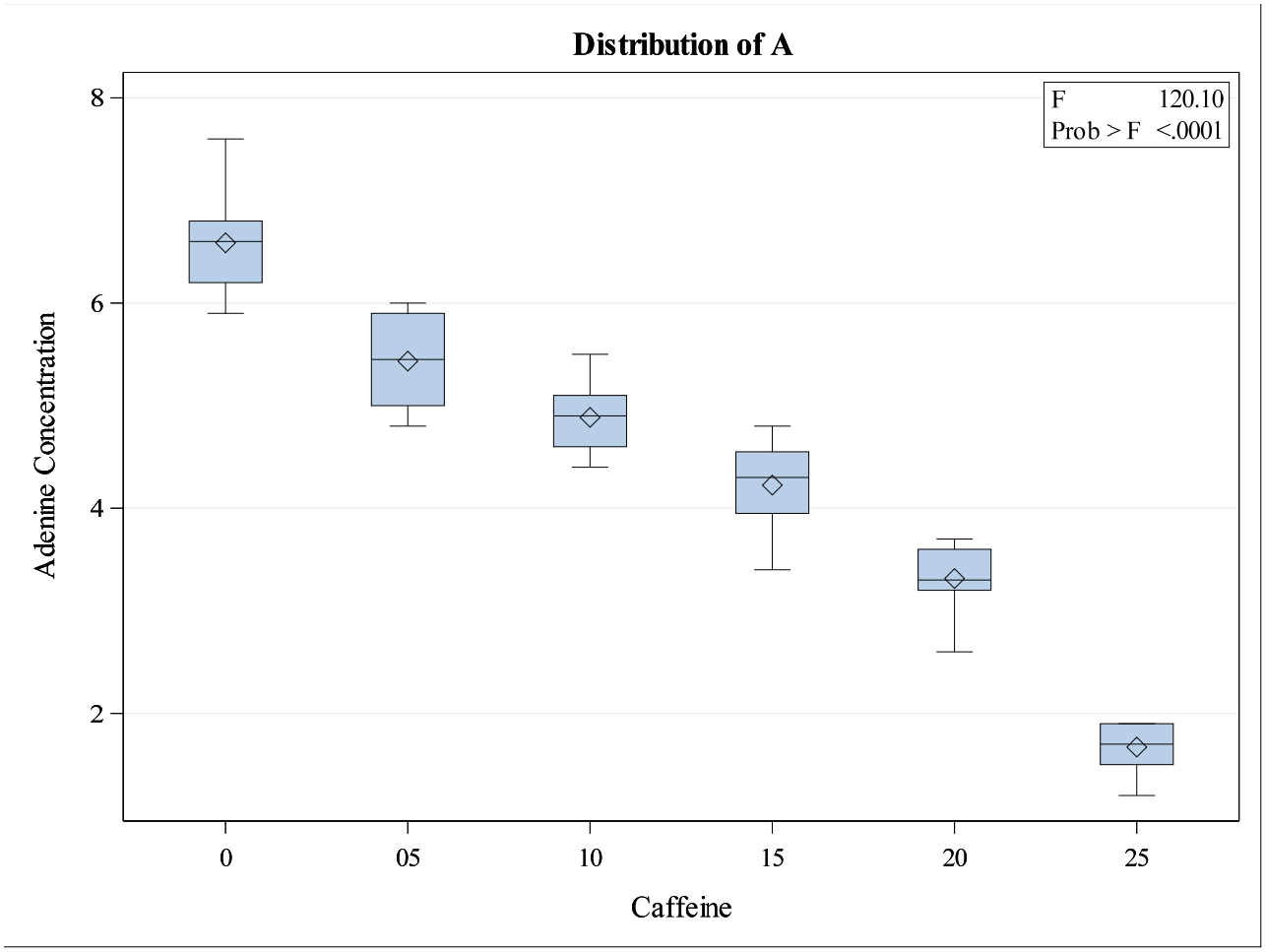
The GLM Procedure

Dependent Variable: A Adenine Concentration

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	106.9282323	21.3856465	120.10	<.0001
Error	37	6.5885119	0.1780679		
Corrected Total	42	113.5167442			

R-Square	Coeff Var	Root MSE	A Mean
0.941960	9.641434	0.421981	4.376744

Source	DF	Type III SS	Mean Square	F Value	Pr > F
C	5	106.9282323	21.3856465	120.10	<.0001



642 EXAM I Spring 2020

The GLM Procedure
Least Squares Means

C	A LSMEAN	LSMEAN Number
0	6.58750000	1
05	5.43333333	2
10	4.88571429	3
15	4.22500000	4
20	3.31428571	5
25	1.67142857	6

Least Squares Means for effect C
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: A

i/j	1	2	3	4	5	6
1		<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001		0.0252	<.0001	<.0001	<.0001
3	<.0001	0.0252		0.0045	<.0001	<.0001
4	<.0001	<.0001	0.0045		0.0002	<.0001
5	<.0001	<.0001	<.0001	0.0002		<.0001
6	<.0001	<.0001	<.0001	<.0001	<.0001	

642 EXAM I Spring 2020

The GLM Procedure

Dunnett's t Tests for A

Note: This test controls the Type I experimentwise error for comparisons of all treatments against a control.

Alpha	0.05
Error Degrees of Freedom	37
Error Mean Square	0.178068
Critical Value of Dunnett's t	2.63867

Comparisons significant at the 0.05 level are indicated by *.**				
C Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
05 - 0	-1.1542	-1.7555	-0.5528	***
10 - 0	-1.7018	-2.2781	-1.1255	***
15 - 0	-2.3625	-2.9192	-1.8058	***
20 - 0	-3.2732	-3.8495	-2.6969	***
25 - 0	-4.9161	-5.4923	-4.3398	***

642 EXAM I Spring 2020

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

C	A LSMEAN	LSMEAN Number
0	6.58750000	1
05	5.43333333	2
10	4.88571429	3
15	4.22500000	4
20	3.31428571	5
25	1.67142857	6

Least Squares Means for effect C
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: A

i/j	1	2	3	4	5	6
1		0.0002	<.0001	<.0001	<.0001	<.0001
2	0.0002		0.2071	<.0001	<.0001	<.0001
3	<.0001	0.2071		0.0476	<.0001	<.0001
4	<.0001	<.0001	0.0476		0.0023	<.0001
5	<.0001	<.0001	<.0001	0.0023		<.0001
6	<.0001	<.0001	<.0001	<.0001	<.0001	

642 EXAM I Spring 2020

The GLM Procedure

Dependent Variable: A Adenine Concentration

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
LINEAR	1	102.4048355	102.4048355	575.09	<.0001
QUADRATIC	1	1.3441344	1.3441344	7.55	0.0092
CUBIC	1	1.9300758	1.9300758	10.84	0.0022
QUARTIC	1	0.0136742	0.0136742	0.08	0.7832
QUINTIC	1	0.0247513	0.0247513	0.14	0.7114