STAT 5200 Spring 2015 Midterm Exam

Do not open the exam until you are instructed to do so.

Directions: You have 60 minutes (12:00-1:00) to complete the exam. You may use your calculator and a single page (both sides) of notes, but no laptops or wireless-capable devices are allowed. Be concise with all your responses (no more than 1-2 sentences are needed for each question). You may use 2 or 3 decimal places in all calculations. The point-worth of each question is given, and the total points sum to 100.

Student Name: Solutions

SAS Output: Partial SAS code, output, and graphics (all clearly identified by title or output number) from certain models are provided in a separate handout, and are necessary for some of the questions on the exam.

Statistical Significance: For all significance tests on this exam, use significance level $\alpha = 0.05$. Where multiple hypotheses are concerned, control the strong family-wise error rate at $\alpha = 0.05$.

Experiment: A baker specializing in cookies has developed a sugar substitute that is cheaper than regular cane sugar, but is suspected of affecting chewiness. The baker designs a study to assess the performance of the sugar substitute compared to regular cane sugar at a few baking temperatures. Six combinations of Sugar and Temperature are considered, as follows:

Sugar:	C	C	C	S	S	S
Temp:	300	350	375	300	350	375
Combination:	1	2	3	4	5	6

Sugar is coded 'C' for cane sugar and 'S' for sugar substitute. Temp is the temperature in degrees Fahrenheit. At each of the 6 combinations, 8 batches of cookie dough are created and baked separately (for fifteen minutes), so there are 48 batches total. Each batch produces one tray of 12 cookies. After cooling each tray on a cooling rack for thirty minutes, the chewiness of each cookie (on a scale where higher values means better chewiness) is measured using a device constructed by the baker.

Question 1: Refer to the description of the experiment on the first page of the exam.

a) (2 points) What is the response variable in this study?

Chewiness

b) (3 points) What are the measurement units in this study? How many are there in total?

Cookies - 576 (= 48 × 12)

+2 +1

c) (3 points) What are the experimental units in this study? How many are there in total?

Batches (or trays) - 48 (= 6 × 8)

+2 +1

d) (4 points) What should be done with the response measurements on the measurement units to proceed with a meaningful analysis based on the experimental units?

take average within each exp. unit (tray)

Question 2: (8 points) The baker's research team has only one long day and a single-tray oven to perform this study. Assume that the oven's temperature can be adjusted immediately between trays. Someone on the team proposes that the 48 batches be conducted in order by Combination number (see table on first page of exam), so all 8 trays with cane sugar at temperature 300 (Combination 1) would be done first, followed by the 8 trays with cane sugar at temperature 350 (Combination 2), and so on. Does this proposal make this a good experiment? If so, why? If not, what would be a better proposal that would make this a good experiment, and why is it better?

+2 { No

+3 { Need to randomize exp. units (trays) to treatments (or combinations)

+3 { This would avoid confounding factors [may specify] -or- would allow for valid statistical inference

okay to use other symbols (like α, β, etc.)

Question 3: Refer to the code and output for SAS Output 1.

a) (10 points) Write out the appropriate effects parameterization model corresponding to this SAS code. Define any symbols (parameters or letters) you use, and specify the range of any subscripts you use.

+5 +4 +1

$$Y_{ijk} = \mu + T_i + S_j + TS_{ij} + \varepsilon_{ijk}$$

Y_{ijk}: Chewiness

μ: overall mean

T_i: Temp effect

S_j: Sugar effect

TS_{ij}: interaction (-2 if $T_i \times S_j$)

ε_{ijk}: random error

$i = 1, 2, 3$ $\quad j = 1, 2$ $\quad k = 1, \ldots, 8$

b) (8 points) Identify two assumptions made by this model, referring to a specific part of the model. For each assumption, comment briefly on what the evidence in the output suggests regarding the appropriateness of the assumption.

+2 i) Normality of ε_{ijk}'s

+2 Normal prob. plot (or histogram) of residuals shows non-normality (may specify right skewness)

+2 ii) Constant variance of ε_{ijk}'s

+2 Plot of residuals vs. predicted values shows non-constant variance (may specify megaphone)

c) (6 points) Why must the issues you mention in part (b) be addressed before proceeding with statistical inference?

+6 If model assumptions are not met, then any inference is worthless (or invalid)

[May point to incorrect sampling distribution, or focus on how inference depends on assumptions.]

-2 if only "transform Y"

Question 4: (4 Points) Refer to the code and output for SAS Output 2. What should be done based on this result? (Be specific.)

Transform Y (or response or Chewiness), like new$Y = Y^{\lambda}$ for some λ between about 0.2 & 0.65

[Could just say "take square root of Chewiness"]

Question 5: After acting on the correct suggestion of Question 4, the variable newChewy is defined, and the model is fit again in SAS Output 3.

a) Looking at this output, and despite noting the significance of the interaction term, a member of the baker's research team said, "Since the lines in the interaction plot do not intersect, there is really not an interaction between Temp and Sugar." Correct this team member by answering the following:

i. (7 points) What does the interaction term mean in this model?

The effect of Temp (on newChewy) depends on Sugar

-or-

The effect of Sugar (on newChewy) depends on Temp

[-5 if focus on significance instead of interpretation]

ii. (7 points) How does the specific non-parallel pattern in this interaction plot communicate the interpretation in part (i) above? (That is, characterize the interaction using this interaction plot.)

Many possible, but only need one statement like these; must be specific to this plot.

- Increasing Temp steadily improves Chewiness for Sugar = 'S', but for Sugar = 'C', increasing Temp first improves but then worsens Chewiness.
- Sugar = 'C' results in higher Chewiness than Sugar = 'S', but it is much higher at Temp = 350.

[-5 if focus on significance instead of interpretation;
-4 if only refer to Temp effect for one level of Sugar]

-3 if focus on what "drives" interaction (like "Temp 350 drives the interaction") without clearly referring to HOW things (like the effect of Sugar) are different at Temp 350 (as in second example bullet point)

b) (6 points) The ANOVA table at the beginning of SAS Output 3 has some portions removed. Report the numeric value of the F test statistic and its sampling distribution. (It is equivalent to the test statistic evaluating the equality of all 6 Combination means.)

$F = \frac{89.989}{7.865} \approx 11.442 \sim F_{5,42}$

+4 +2

[okay to round differently in F calculation]

c) (4 points) What should be done differently in the SAS code to obtain the other interaction plot?

Switch the order of the variables in the CLASS statement

(class Sugar Temp;)

[-2 if MODEL statement]

→ Exact letters don't matter, but should have three clear groupings.

Question 6: (15 points) Refer to the code and output for SAS Output 4. (The variable newChewy is the same as in Question 5.) Using the appropriate output from either 'SAS Output 3' or 'SAS Output 4', fill in the columns of letter groupings in the table below (using three letters) to make the statement true that "LS-means with the same letter are not significantly different" while controlling a meaningful family-wise error rate at 0.05. (Note: You may find it useful to start at the bottom of the table, and work your way up.)

LS-means with the same letter are not significantly different.				[-1.5 for each missing or extraneous letter] **Letter Groupings** Correct groupings from Output 4		
newChewy LSMEAN	**Sugar**	**Temp**	**LSMEAN Number**			
15.822716	**C**	**350**	3		B	
12.494659	**C**	**375**	5		B	C
12.206091	**S**	**375**	6		B	C
9.204538	**C**	**300**	1	A		C
8.358389	**S**	**350**	4	A		C
6.597414	**S**	**300**	2	A		

Incorrect groupings from Output 3

C
B
B
A
A
A

4 missing letters (-6) here, so -6

Question 7: (12 points) Looking at the "Interaction Plot" in SAS Output 3, someone comments that it looks like, averaging over the two levels of Sugar, the mean of newChewy doesn't really change between Temp 350 and Temp 375. Let μ_i be the mean newChewy value for Combination i (see first page of exam). Define a contrast ψ in terms of the μ_i's, such that "H_0: $\psi = 0$" tests this.

+5 correct understanding: H_0: mean at Temp 350 = mean at Temp 375

+4 express as H_0 in terms of μ_i's: $H_0: \frac{\mu_2 + \mu_5}{2} = \frac{\mu_3 + \mu_6}{2}$

+3 write as contrast: $H_0: \mu_2 - \mu_3 + \mu_5 - \mu_6 = 0$

$\psi = \mu_2 - \mu_3 + \mu_5 - \mu_6$ (or $k \cdot \psi$ for any $k \neq 0$)

(+5 for any $\sum w_i \mu_i$ such that $\sum w_i = 0$)

Question 8: (1 point) What topic(s) did you study most that did not appear on this exam?

(Anything)