

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: _____

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?

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

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

- 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?

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?

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

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

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?

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.)

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.)

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

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.				Letter Groupings		
newChewy LSMEAN	Sugar	Temp	LSMEAN Number			
15.822716	C	350	3			
12.494659	C	375	5			
12.206091	S	375	6			
9.204538	C	300	1			
8.358389	S	350	4			
6.597414	S	300	2			

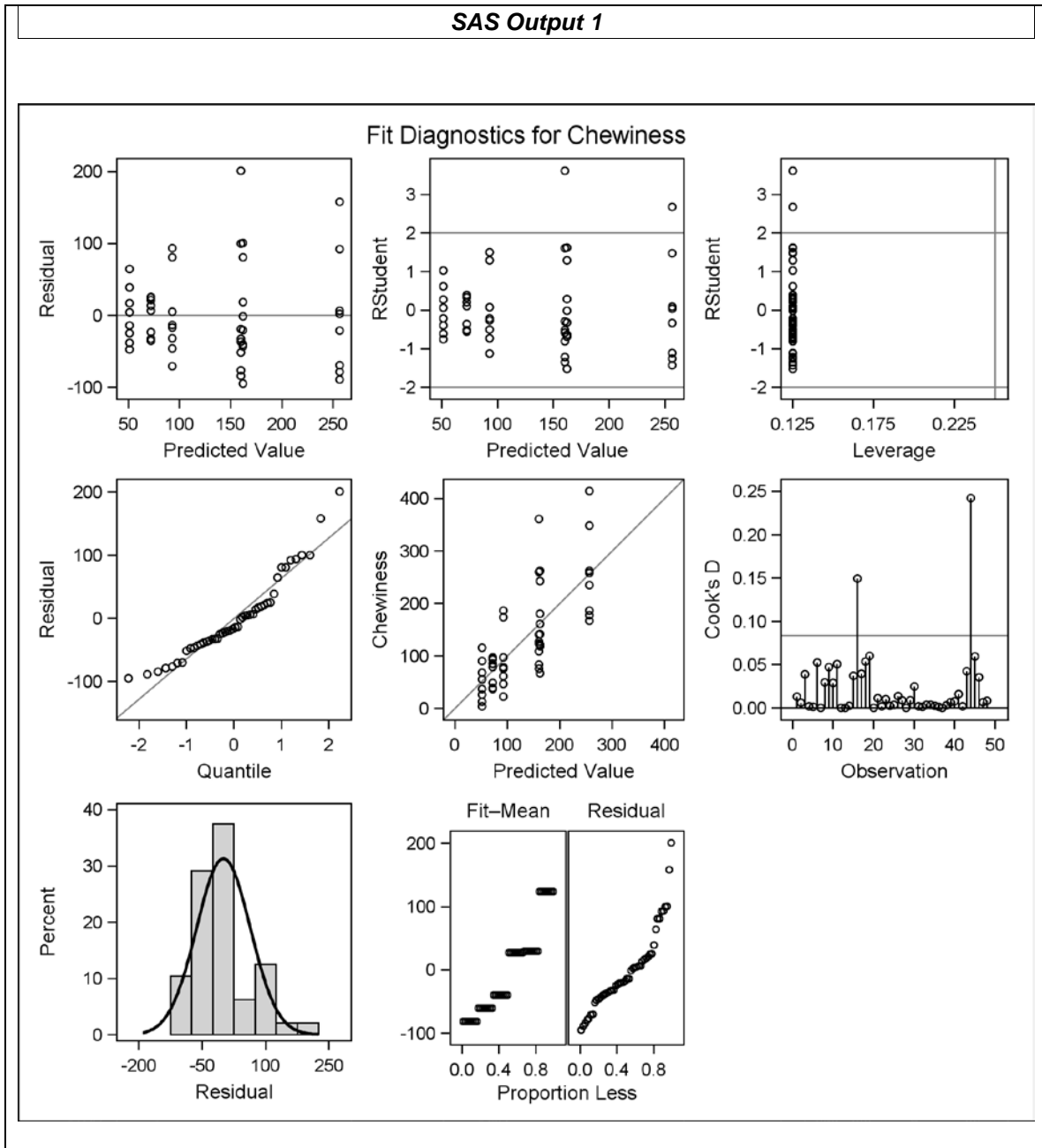
Question 7: (12 points) Looking at the “Interaction Plot” in SAS Output 3, someone comments that it looks like, averaging over the two levels of Sugar, the mean of newChewy doesn’t really change between Temp 350 and Temp 375. Let μ_i be the mean newChewy value for Combination i (see first page of exam).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

$$\psi =$$

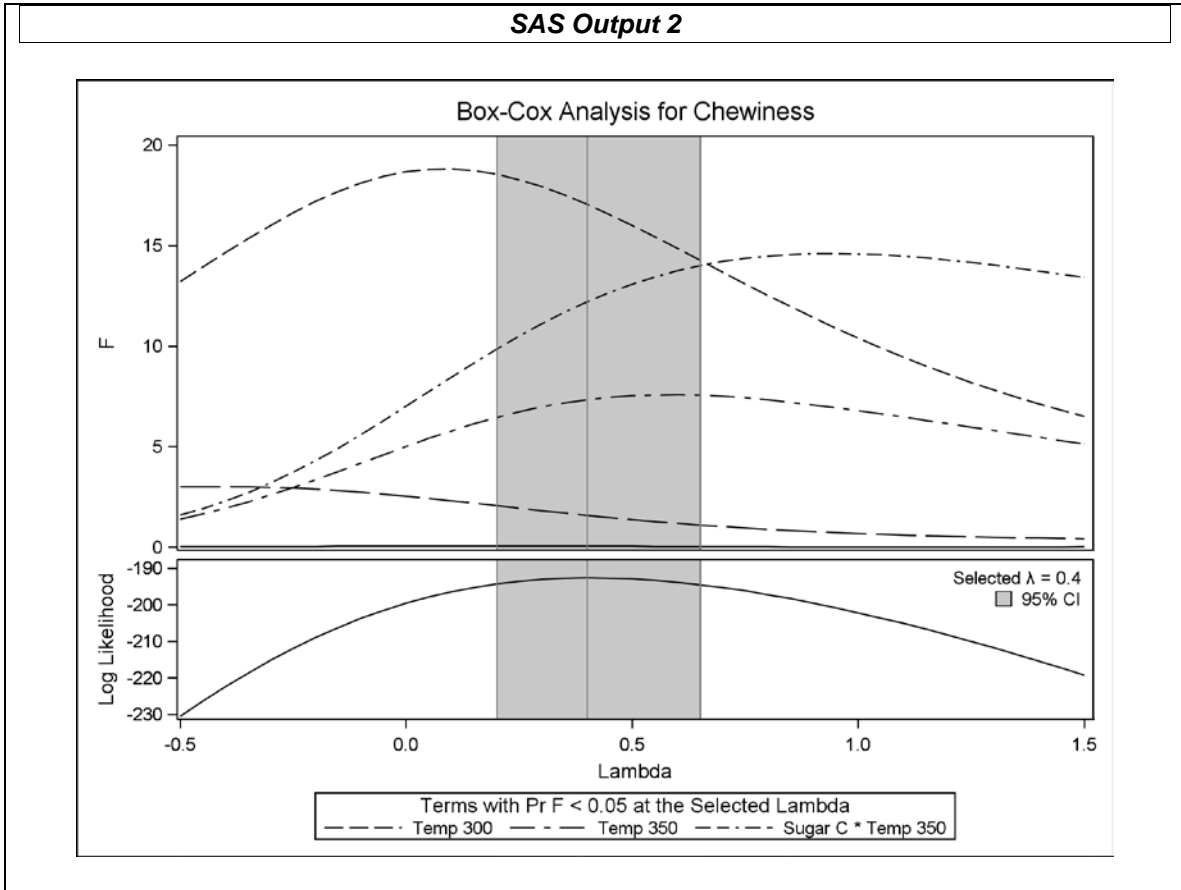
Question 8: (1 point) What topic(s) did you study most that did not appear on this exam?

SAS Code & Output for STAT 5200 Spring 2015 Midterm Exam

```
proc glm data=cookie plots=diagnostic;  
  class Temp Sugar;  
  model Chewiness = Sugar | Temp;  
  title 'SAS Output 1';  
run;
```



```
proc transreg data=cookie;  
  model boxcox(Chewiness / lambda=-0.5 to 1.5 by 0.05)  
    = class(Sugar|Temp);  
  title 'SAS Output 2';  
run;
```



```

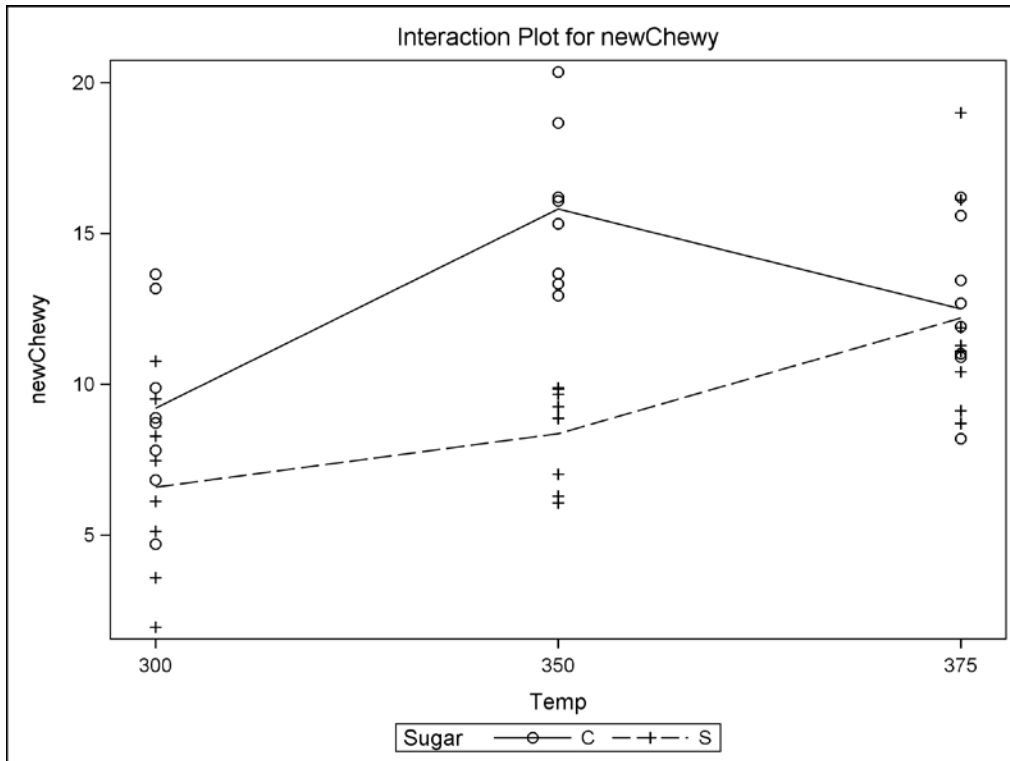
proc glm data=cookie plots=diagnostics;
  class Temp Sugar;
  model newChewy = Sugar | Temp;
  lsmeans Sugar*Temp / pdiff lines;
  title1 'SAS Output 3';
run;

```

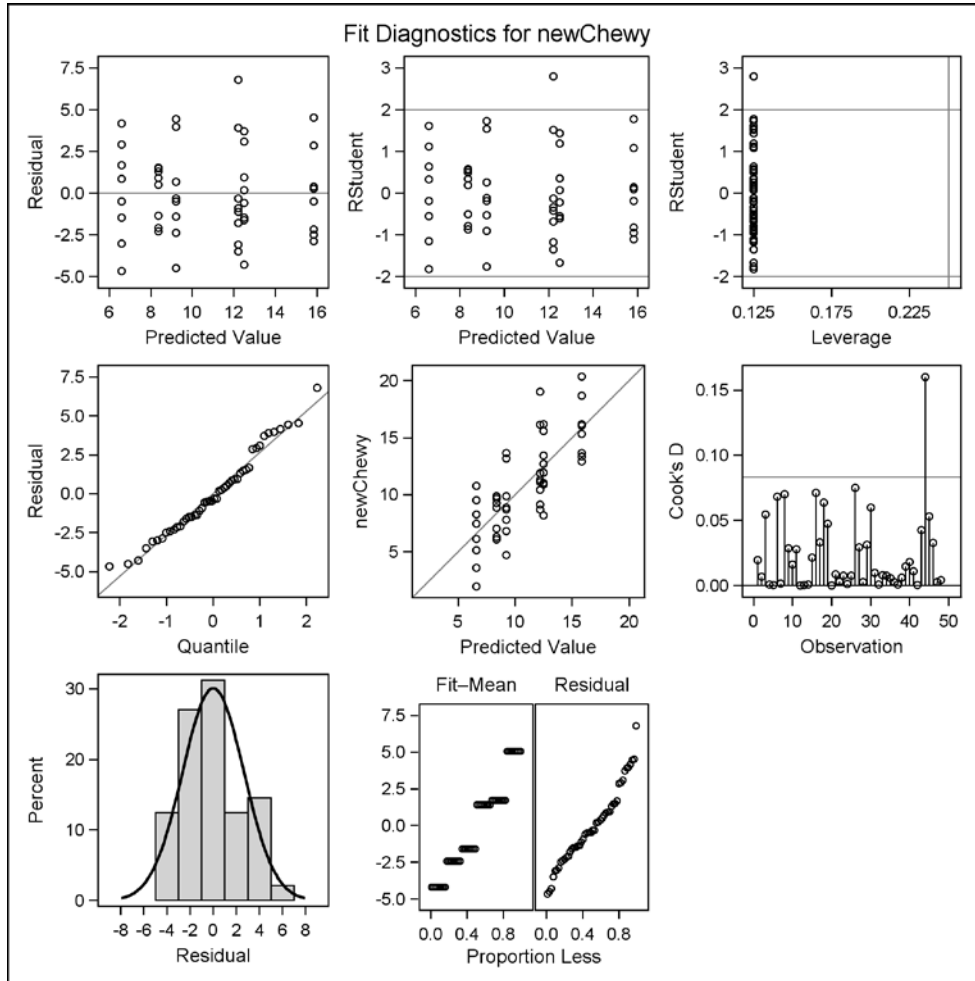
SAS Output 3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	449.9446039	89.9889208	18.20	<.0001
Error	44	330.3124199	7.8645814		
Corrected Total	47	780.2570239			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Sugar	1	143.1066602	143.1066602	18.20	0.0001
Temp	2	199.5584384	99.7792192	12.69	<.0001
Temp*Sugar	2	107.2795054	53.6397527	6.82	0.0027



(SAS Output 3, continued)



Least Squares Means

Temp	Sugar	newChewy LSMEAN	LSMEAN Number
300	C	9.2045381	1
300	S	6.5974144	2
350	C	15.8227162	3
350	S	8.3583894	4
375	C	12.4946591	5
375	S	12.2060906	6

(SAS Output 3, continued)

Least Squares Means for effect Temp*Sugar Pr > t for H0: LSMean(i)=LSMean(j) Dependent Variable: newChewy						
i/j	1	2	3	4	5	6
1		0.0700	<.0001	0.5495	0.0237	0.0382
2			<.0001	0.2161	0.0001	0.0003
3				<.0001	0.0223	0.0135
4					0.0052	0.0089
5						0.8379
6						

Note: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.

```

proc glm data=cookie plots=diagnostics;
  class Temp Sugar;
  model newChewy = Sugar|Temp;
  lsmeans Sugar*Temp / pdiff lines adjust=tukey;
  title1 'SAS Output 4';
run;

```

SAS Output 4

**Least Squares Means
Adjustment for Multiple Comparisons: Tukey**

Temp	Sugar	newChewy LSMEAN	LSMEAN Number
300	C	9.2045381	1
300	S	6.5974144	2
350	C	15.8227162	3
350	S	8.3583894	4
375	C	12.4946591	5
375	S	12.2060906	6

**Least Squares Means for effect Temp*Sugar
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: newChewy**

i/j	1	2	3	4	5	6
1		0.4406	0.0004	0.9902	0.1989	0.2870
2			<.0001	0.8066	0.0017	0.0032
3				<.0001	0.1890	0.1251
4					0.0544	0.0876
5						0.9999
6						