

Directions: You have 70 minutes to complete the exam. You may use your calculator and a single page (both sides) of notes, but no laptops, cell phones, or other 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 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 output and graphics from certain models using SAS procedures are provided in a separate handout, and are necessary for some of the questions on the exam. The SAS output and graphics are clearly identified by title or output number.

Data: An experiment was carried out to evaluate the effects of multiple factors on the growth of shrimp. Two **Temperatures** were used in the experiment: 25° C and 35° C. Water **Salinity** is expressed as a percentage, and three levels were used (10, 25, and 40). Six 40-liter containers were randomly assigned to each combination of factor levels. For each container, 120 shrimp are randomly selected from a large batch of shrimp, weighed as a group, and then put in the container. After four weeks, the shrimp were harvested from the containers, weighed again as a group, and then the *average* of the 120 shrimps' weight **Gain** (in mg) in each container was recorded. The six combinations of factor levels were referred to as the six **Conditions**, summarized as follows:

Temp:	25	35	25	35	25	35
Salinity:	10	10	25	25	40	40
Condition:	1	2	3	4	5	6

Question 1: (9 points) The researchers were not lazy – they could have weighed individual [and tagged] shrimp (before and after the four weeks in their respective containers) and recorded individual weight gain. Referring to principles of experimental design, why was it appropriate to use the average of the 120 shrimps' weight gain in each container instead of the individual shrimp's weight gain?

Question 2: (7 Points) The SAS code below produces 'SAS Output 1'. What inference, if any, can be made about the effect of Condition factor level on mean Gain, based on this output? If no valid inference can be made, specifically explain why.

```
proc glm data=shrimp plots=diagnostics;  
  class Condition;  
  model Gain = Condition;  
run;
```

Question 3: (6 Points) The SAS code below produces 'SAS Output 2'. What should be done based on this result? (Be specific.)

```
proc transreg data=shrimp;  
  model boxcox(Gain / lambda=-1 to 1 by 0.05) = class(Condition);  
run;
```


Question 6: (14 points) Refer to the code (previous page) and output (attached) for ‘SAS Output 3’. The LSMEANS statement (which does not depend on the MEANS statement) produces the table below of raw, unadjusted p-values. We are interested in all pairwise comparisons of Condition level means, and we want to control the strong family-wise error rate at $\alpha=.05$ while maintaining the highest statistical power. Based on this interest and using the appropriate table in ‘SAS Output 3’, cross out the raw, unadjusted p-values below for the pairwise comparisons that are **not** statistically significant after appropriately accounting for multiple comparisons. (Here, LSMEAN number is the same as Condition number. Calling only p-values in this table greater than .05 ‘not significant’ would just control the per-comparison error rate at $\alpha=.05$, which is not our objective.)

Least Squares Means for effect Condition Pr > t for H0: LSMean(i)=LSMean(j) Dependent Variable: newGain						
i/j	1	2	3	4	5	6
1		<.0001	<.0001	<.0001	<.0001	<.0001
2			0.3085	0.0310	0.0490	0.0003
3				0.0035	0.0071	<.0001
4					0.7773	0.0380
5						0.0312
6						

Question 7: (14 points) Let μ_i be the population mean newGain value of Condition i. Define a contrast ψ in terms of the μ_i 's, such that “ $H_0: \psi = 0$ ” corresponds to a test of whether the mean newGain value is any different in the lowest Salinity level than in the highest two Salinity levels, averaging over both Temperatures.

$$\psi =$$

Question 8: The SAS code below produces ‘SAS Output 4’, which also includes one additional interaction plot.

```
proc glm data=shrimp plots=diagnostics;  
  class Temp Salinity;  
  model newGain = Temp|Salinity;  
  lsmeans Temp*Salinity;  
run;
```

- a) (14 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. Also specify model assumptions, referring to a specific part of the model.

b) (8 points) What does the interaction term mean in this model?

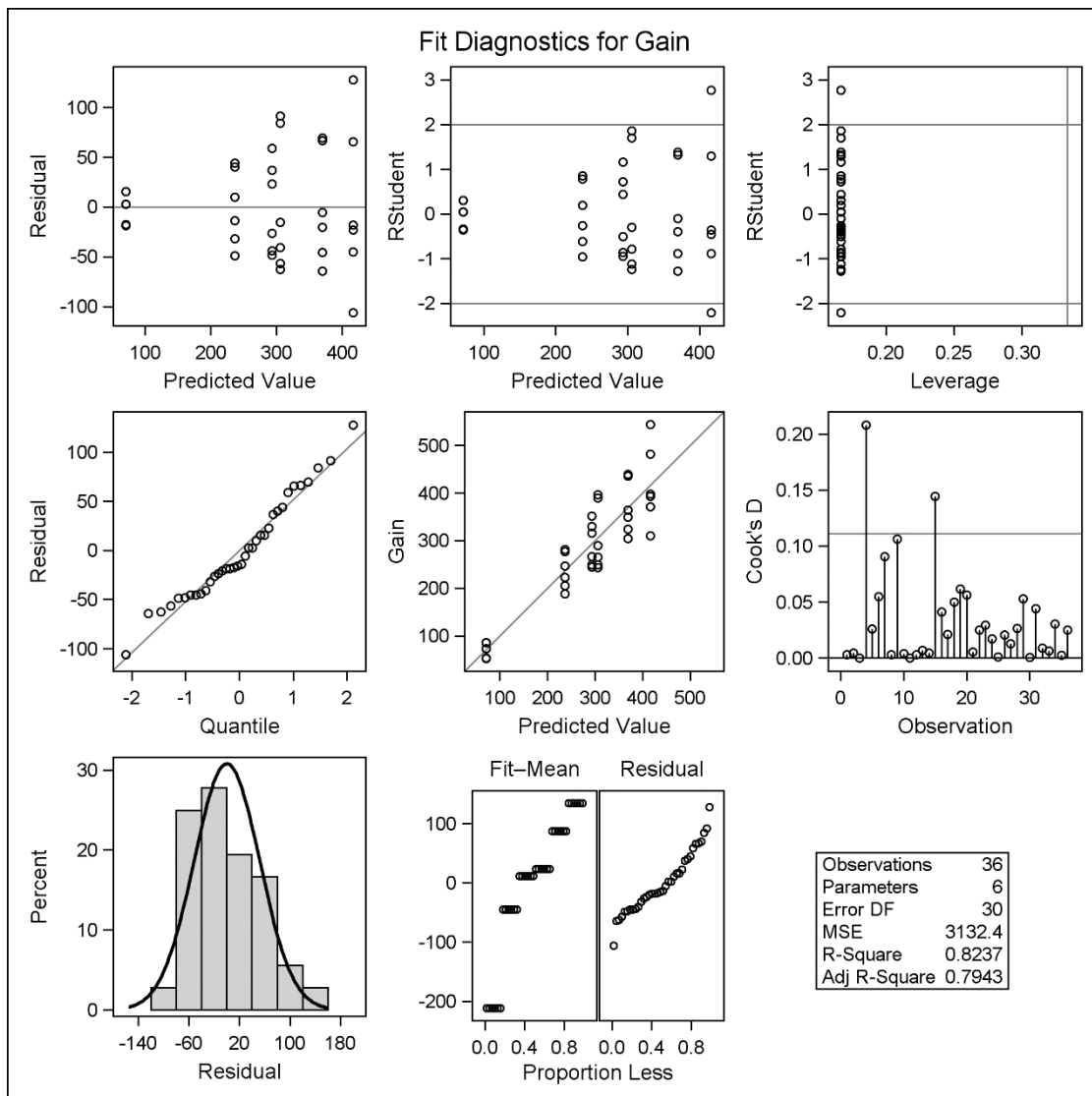
c) (6 points) Would you expect the interaction term to be statistically significant here? Explain briefly, referring to specific elements of ‘SAS Output 4’.

Question 9: (1 point) What topic(s) did you study most that did not appear on this exam?

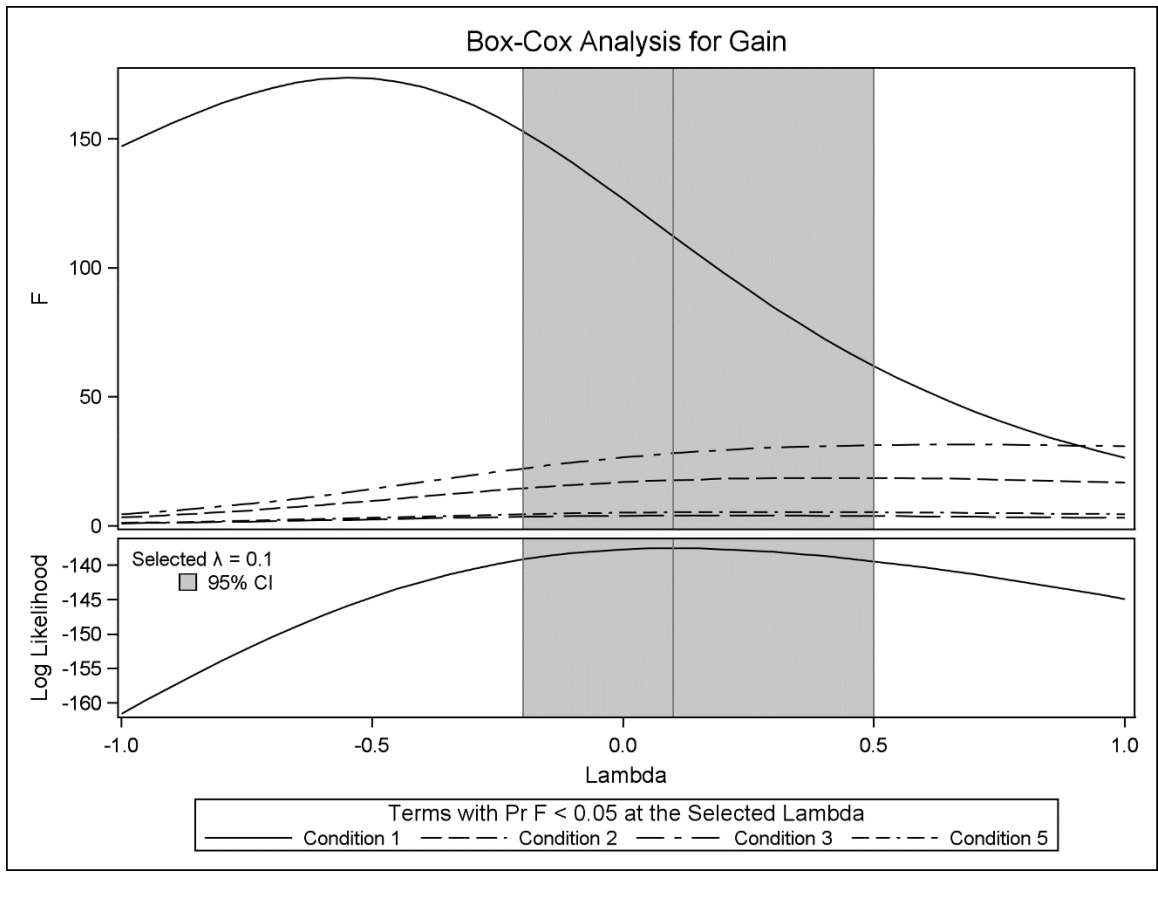
SAS Output for STAT 5200 Fall 2017 Midterm Exam

SAS Output 1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	438952.6667	87790.5333	28.03	<.0001
Error	30	93971.3333	3132.3778		
Corrected Total	35	532924.0000			

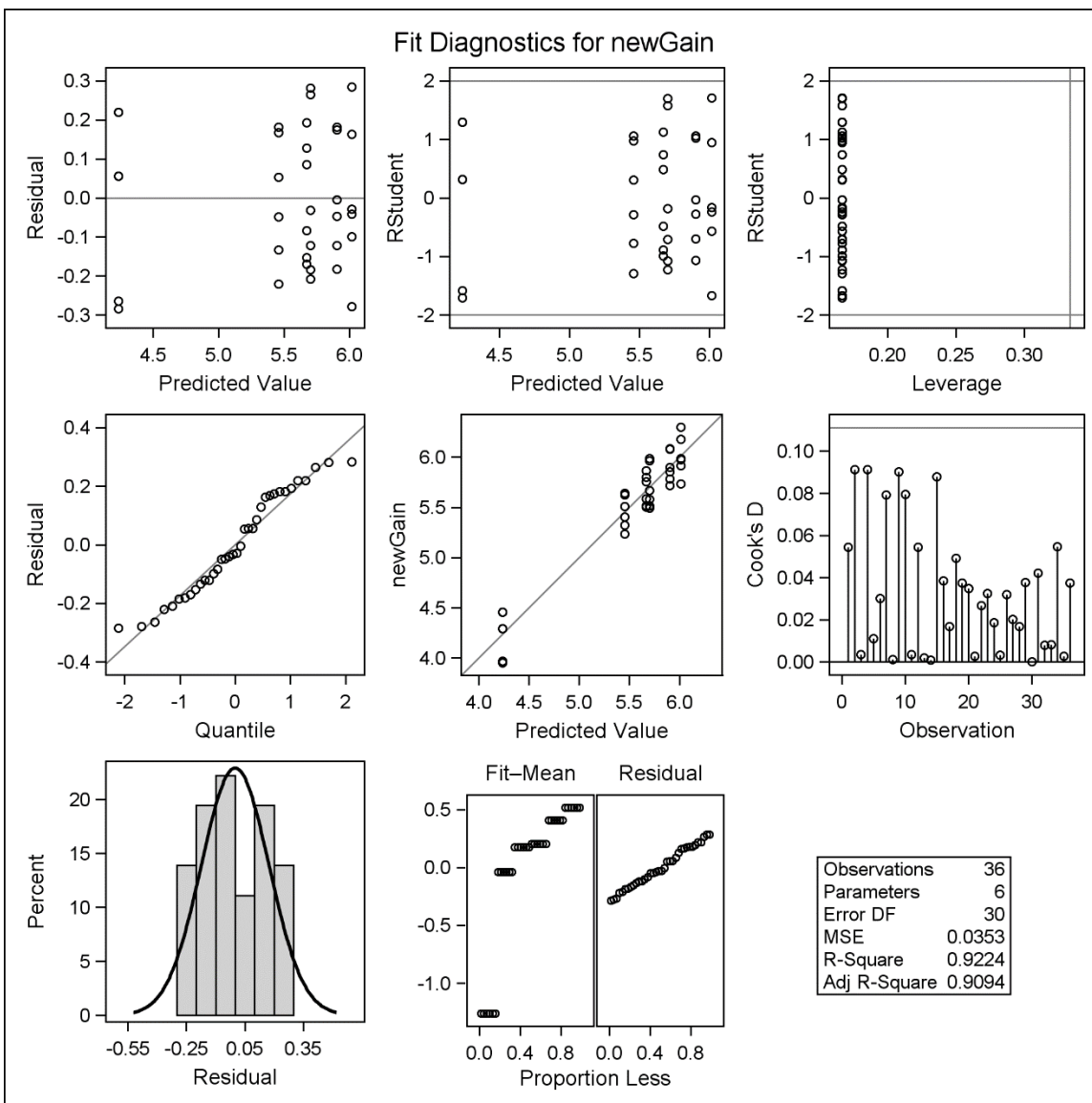


SAS Output 2



SAS Output 3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	12.59217476	2.51843495	?????	<.0001
Error	30	1.05969508	0.03532317		
Corrected Total	35	?????			



(continued)

SAS Output 3, Continued

t Tests (LSD) for newGain

Means with the same letter are not significantly different.				
t Grouping		Mean	N	Condition
	A	6.0150	6	3
	A	5.9026	6	2
B		5.7017	6	5
B		5.6707	6	4
	C	5.4564	6	6
D		4.2352	6	1

Ryan-Einot-Gabriel-Welsch Multiple Range Test for newGain

Means with the same letter are not significantly different.				
REGWQ Grouping		Mean	N	Condition
	A	6.0150	6	3
B	A	5.9026	6	2
B	C	5.7017	6	5
B	C	5.6707	6	4
	C	5.4564	6	6
	D	4.2352	6	1

(continued)

SAS Output 3, Continued

Tukey's Studentized Range (HSD) Test for newGain

Means with the same letter are not significantly different.					
Tukey Grouping			Mean	N	Condition
	A		6.0150	6	3
B	A		5.9026	6	2
B	A	C	5.7017	6	5
B		C	5.6707	6	4
		C	5.4564	6	6
	D		4.2352	6	1

Bonferroni (Dunn) t Tests for newGain

Means with the same letter are not significantly different.					
Bon Grouping			Mean	N	Condition
	A		6.0150	6	3
	A		5.9026	6	2
B	A		5.7017	6	5
B	A		5.6707	6	4
B			5.4564	6	6
	C		4.2352	6	1

SAS Output 4

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	12.59217476	2.51843495	?????	<.0001
Error	30	1.05969508	0.03532317		
Corrected Total	35	?????			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Temp	1	1.16170537	1.16170537	32.89	<.0001
Salinity	2	3.71566966	1.85783483	52.60	<.0001
Temp*Salinity	2	7.71479973	3.85739987	?????	?????

Temp	Salinity	newGain LSMEAN	LSMEAN Number
25	10	4.23519152	1
25	25	6.01498822	2
25	40	5.70170135	3
35	10	5.90257046	4
35	25	5.67073763	5
35	40	5.45639737	6

(continued)

SAS Output 4, continued

