

A Few Final Exam Questions from Past Semesters

1. Does bread lose its vitamins when stored? Twenty small loaves of bread were prepared with flour that was fortified with a fixed amount of vitamin E. Immediately after baking, the vitamin E content of each of 10 loaves (randomly selected from the 20) was measured. The remaining 10 loaves were stored for three days. After storage, the vitamin E content was measured in each of the loaves that had been stored. Summary statistics (in mg/100g) are provided in the table below.

Bread	Average	Standard Deviation
Fresh	97.4	3.2
Stored	96.1	2.7

- (a) Is there significant evidence that vitamin E is lost when bread is stored? You may assume that all the standard assumptions are satisfied.

- (b) Is the investigation described here an experiment or an observational study. Explain briefly.

2. Researchers place the roots of a plant in a dish with a chemical mixture that promotes the development of shoots on the roots. They want to compare two plant genotypes (say A and B) to see if one type's roots tend to produce more shoots than the other. The researchers use 10 dishes containing the chemical mixture in the experiment. One root from a plant of each genotype is placed in each dish. At the conclusion of the experiment, the number of shoots on each root is counted. The table below contains the number of shoots that developed on each root. The researchers would like to conduct some follow-up research using the type of plant that produces the most shoots. Does it matter which genotype the researchers use in follow-up experiments? If so, which should be used? Several analyses have been conducted to answer this question. You may assume that the calculations are correct for each of the analyses. Some of the analyses, however, are inappropriate for this data set. Mark an X next to any of the analyses listed below that are inappropriate. Place a * next to the best analysis for this data set. Leave blank any analyses that could appropriately be applied to this type of data but are not as good as the best analysis in this case.

Dish	Genotype	
	A	B
1	4	12
2	5	4
3	17	18
4	4	3
5	8	6
6	5	7
7	15	21
8	6	10
9	3	9
10	5	11

sign test : Either genotype could be used in follow-up research. There is no significant evidence suggesting that one genotype has a greater chance of producing more shoots than the other (2-sided p -value=0.3438).

two-sample t-test : Either genotype could be used in follow-up research. It looks like both genotypes could have the same mean number of shoots per root ($t = -1.26$, 2-sided p -value=0.2225).

regression analysis : Either genotype A or genotype B could be used in follow-up research. There is a significant linear relationship between the number shoots produced by a root of genotype A and the number of shoots produced by a root of genotype B ($\hat{Y} = 3.51 + 0.929X$, p -value=0.0070 for test of $H_0 : \beta_1 = 0$).

paired-data t-test : Genotype B should be used in follow-up experiments. B produces significantly more shoots than genotype A on average ($t = -2.67$, 2-sided p -value=0.0258).

Wilcoxon rank-sum test : Either genotype could be used in follow-up research. It looks like both genotypes could have the same distribution for number of shoots per root ($z = -1.3663$, 2-sided p -value=0.1718).

Wilcoxon signed-rank test : Genotype B should be used in follow-up experiments. There is some mild evidence that B produces significantly more shoots than genotype A on average (2-sided p -value=0.0508).

3. Forty rats were randomly divided into 4 groups of 10 rats each. Two of the four groups were randomly chosen to receive a high-protein diet. One of these two groups was randomly chosen to receive protein through a beef-based diet. The other high-protein group received protein through a cereal-based diet. The other two groups received normal amounts of protein, one through a beef-based diet and the other through a cereal-based diet. Due to some problems in the lab, 5 rats in the high-protein beef-based diet were excluded from the study. Two rats from the high-protein cereal-based diet were excluded. The weight gain in grams over the course of the study was recorded for each of the remaining rats. Some summary statistics, SAS code, and partial SAS output are provided below.

Protein Level	Protein Source	Number of Rats	Average Weight Gain
normal	cereal	10	127
normal	beef	10	126
high	cereal	8	132
high	beef	5	138

```
proc glm;
  class level source;
  model wgain=level source level*source;
run;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	????	????	????	14.85	<.0001
Error	????	????	????		
C. Total	????	984			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
level	1	550.4761905	550.4761905	41.14	<.0001
source	1	47.6190476	47.6190476	3.56	0.0693
level*source	1	xxxxxxxxxx	xxxxxxxxxx	????	0.0132

(a) Give the missing degrees of freedom for Model, Error, and Corrected Total.

(b) Compute the Model and Error Sum of Squares.

(c) Compute the Model and Error Mean Squares.

(d) Compute lsmeans for each level of the factor called *Protein Source*.

(e) Compute the F -statistic missing from the *level*source* line of SAS output.

(f) The researchers know from past work that rats on the high-protein diets will have greater weight gains than rats on the normal-protein diets. In this experiment, they are primarily interested in getting information about how the source of the protein (beef or cereal) might affect weight gain. Conduct an appropriate statistical analysis of this data to help the researchers understand what the results of this experiment convey about the relationship between protein source and weight gain. You may assume that all the standard assumptions are satisfied. Organize your work neatly and include a final statement summarizing the results. Write on the back of this page if necessary.

4. An area in a greenhouse has 8 benches. On each bench, there are 4 pots. Each pot contains 1 plant. Researchers want to compare the effect of four treatments (A, B, D, and D) on the growth of the plants. The researchers randomly assign the four treatments to the four plants on each bench so that each treatment is represented exactly once on each bench.

- (a) Name the type of experimental design used by the researchers.
- (b) At the conclusion of the experiment, the dry weight of each of the 32 plants was obtained. Does it appear that all the treatment means could be equal? Use the code and output below to help you answer this question. Provide the appropriate test statistic, p-value, and a conclusion.

```
proc glm;
  class treatment bench;
  model dryweight=treatment bench;
run;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	73.5000000	7.3500000	2.19	0.0626
Error	21	70.5000000	3.3571429		
C. Total	31	144.0000000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
treatment	3	50.5000000	16.8333333	5.01	0.0089
bench	7	23.0000000	3.28571429	0.98	0.4721

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treatment	3	50.5000000	16.8333333	5.01	0.0089
bench	7	23.0000000	3.28571429	0.98	0.4721

- (c) Suppose the researchers had decided to ignore the factor *bench* in their analysis. In other words, suppose *bench* was deleted from the SAS code above. Provide as many entries as possible in the ANOVA table corresponding to this analysis.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	_____	_____	_____	_____	_____
Error	_____	_____	_____		
C. Total	_____	_____			