

1. A researcher investigated the relationship between weight gains of chickens and the amount of the amino acid lysine included in their diets. A random sample of 12 two-week-old chickens was selected for the study. Each chicken was kept in its own cage and allowed to eat as much lysine-supplemented feed as desired. The mean amount of lysine consumed by the 12 chickens during the study was 0.14 grams. The standard deviation of the amounts of lysine consumed was 0.05. The mean weight gain of the 12 chickens was 18.5 grams. The standard deviation of the 12 weight gains was 2.0 grams. The correlation between the amount of lysine ingested and the amount of weight gained was 0.7.

(a) Compute the equation of the least-squares regression line for predicting weight gain from amount of lysine consumed.

(b) The sample linear correlation coefficient was $r = 0.7$ for this data. Is this correlation significantly different from zero? Provide appropriate statistical analyses to support your answer.

2. Among adults in the United States, there is a correlation between height and hair length. State whether the correlation is positive or negative, and explain why the correlation exists.

3. Researchers wished to find out if two preparations of a virus would produce different effects on tobacco leaves. Eleven tobacco leaves were used in the experiment. Half of each leaf was randomly assigned to preparation 1 and the other half to preparation 2. The number of lesions caused by the virus were counted on each half of each leaf. The data are provided below. Use an appropriate test to determine if the two preparations of the virus produce different effects on tobacco leaves. Give a test statistic, p -value, and a clear conclusion.

Leaf	Count for Preparation 1	Count for Preparation 2
1	7	6
2	19	17
3	18	14
4	17	11
5	10	5
6	41	18
7	10	9
8	13	6
9	21	12
10	19	36
11	20	17

4. Three methods of growing wheat are to be compared on the basis of yield per acre. Five farms are to be used for the experiment. Three plots are used on each farm. At each farm, the three methods are randomly assigned to the three available plots. Thus five measurements are obtained on yield per acre for each of the three growing methods. A SAS program written for the analysis of the data is provided below along with partial output.

- (a) Fill in the missing entries in the ANOVA table at the bottom of the page.
- (b) Based on the F statistic in the ANOVA table, the researchers conclude that there are no significant differences among the mean yields for the three varieties. After examining the SAS code, output, and data, a statistician claims that the researchers' conclusion is wrong. In fact, the statistician claims that there is very strong evidence that the methods differ with regard to mean yield. Explain.

```
data one;
  input method farm yield;
  cards;
1 1 24
1 2 37
1 3 22
1 4 25
1 5 11
2 1 22
2 2 36
2 3 21
2 4 24
2 5 10
3 1 17
3 2 31
3 3 16
3 4 18
3 5 5
;
```

```
proc glm;
  class method;
  model yield=method;
run;
```

Source	DF	Sum of Squares	Mean Square	F Value
Model	_____	_____	_____	0.676
Error	_____	_____	85.6	
Corrected Total	_____	_____		

5. Brain waves, heart rate, and eye movements can be used to determine if a sleeping person is about to enter a dream period. Twenty-eight volunteers for a dream-deprivation study were randomly assigned to 4 groups of 7 people each. The 7 people in group I were allowed to sleep without interruption. The 7 people in group II were awakened four times but never immediately before or during a dream. The 7 people in group III were awakened four times, twice during a dream period and twice during a non-dream period. The 7 people in group IV were awakened four times, each time during a dream period. The procedure was repeated 10 nights. Each day the subjects were given a test to determine their level of anxiety. The test results were used to determine an anxiety score for each subject. Higher scores indicate greater anxiety. Summary statistics and an ANOVA table are provided below.

Group (i)	\bar{Y}_i	s_i	n_i	Source	DF	SS	MS
I	47	11	7	Between	3	3654	1218
II	59	12	7	Within	24	3366	140.25
III	68	10	7	Total	27	7020	
IV	78	14	7				

- (a) Find the minimum significant difference (like the quantity SAS computes) between treatment means ignoring concerns about multiple testing. Use “two-sided p -value ≤ 0.05 ” as your definition of “significant difference.”

- (b) Make a line plot illustrating significant and non-significant differences between group means ignoring concerns about multiple testing. Show your work. Use “two-sided p -value ≤ 0.05 ” as your definition of “significant difference.”

Problem 5 (continued)

- (c) Suppose you computed a p -value for each possible test of equality between two group means in this problem. (You don't have to do it, but suppose you did.) Say precisely how you would convert each of these p -values into a Bonferroni-adjusted p -value.
6. The last page of this exam contains residual plots resulting from the fit of a simple linear regression model to two different sets of data.
- (a) Examine Residual Plot 1. Name all the basic assumptions of simple linear regression that appear to be violated by data set 1.
- (b) Examine Residual Plot 2. Name all the basic assumptions of simple linear regression that appear to be violated by data set 2.