METHODS QUALIFYING EXAM

JANUARY 2004

INSTRUCTIONS:

1. DO NOT put your NAME on the exam. Place the NUMBER assigned to you on the UPPER LEFT HAND CORNER of EACH PAGE of your exam.
2. Please start your answer to EACH QUESTION on a SEPARATE sheet of paper.
3. Answer all the questions.
4. Be sure to attempt all parts of every question. It may be possible to answer a later part of a question without having solved the earlier parts.
5. Be sure to hand in all of your exam. No additional material will be accepted once the exam has ended and you have left the exam room.

PROBLEM #1:

An experiment was designed to compare three different methods of assessing the knowledge obtained by students in an undergraduate statistics course:

- Method 1: Multiple choice questions
- Method 2: Student provides detailed solutions to problems
- Method 3: Individual oral examinations

To conduct the experiment, four sections of STAT 30x will be randomly selected. The four sections are taught by four different instructors. Six students will be randomly selected from each of four different sections of STAT 30x. Each student will take all three exams (a total of 72 observations). The researcher is interested in the difference in average scores on the three exams and whether the size of the differences between average scores is consistent across the various sections of STAT 30x.

(A) Suppose the order in which the three exams are taken is randomly determined for each individual student. Display an appropriate ANOVA table for this experiment with sources of variation, degrees of freedom, expected mean squares, and the F-ratio for testing all relevant hypotheses.

(B) There is concern that there may be an effect based on whether a test is taken during the first, second, or third testing period. Hence, you want to ensure that each test appears in each testing period. Will this change the design? If your answer is yes,

(i) In the following display, give an example of an appropriate assignment of the three tests (M_1, M_2, M_3) to the testing periods for the students.

	Instructor 1			Instructor 2			Instructor 3			Instructor 4		
	Period			Period			Period			Period		
Student	1	2	3	1	2	3	1	2	3	1	2	3
1												
2												
3												
4												
5												
6												

(ii) Display an appropriate ANOVA table for this experiment listing just the sources of variation and degrees of freedom. You **do not** need to determine the expected mean squares for this design.

PROBLEM #2:

A programmer claims that $U_1, \ldots, U_{50}$, is a random sample of size 50 from a uniform (0,1) distribution.

(A) Describe a graphical method to evaluate the programmer's claim. Be sure to label your axes.

(B) Describe a test of hypotheses to evaluate the programmer's claim.

(C) If $U_1, \ldots, U_{50}$ were determined to be in fact a random sample from a uniform on (0,1) distribution, show how $U_1, \ldots, U_{50}$ could be used to generate a random sample of 50 observations from a distribution having cdf given by

$$F(y) = 1 - exp(-(y - \theta)/\beta) \;\; if \;\; y \geq \theta,$$

where θ and β are **known** constants.

(D) Suppose we have $Y_1, \ldots, Y_n$ is a random sample from a population having cdf, F(y). Describe a graphical method to evaluate whether F(y) has the form:

$$F(y) = 1 - exp(-(y - \theta)/\beta) \;\; if \;\; y \geq \theta,$$

where θ and β are **unknown** constants. How can the graphical method be used to yield estimates of θ and β?

PROBLEM # 3

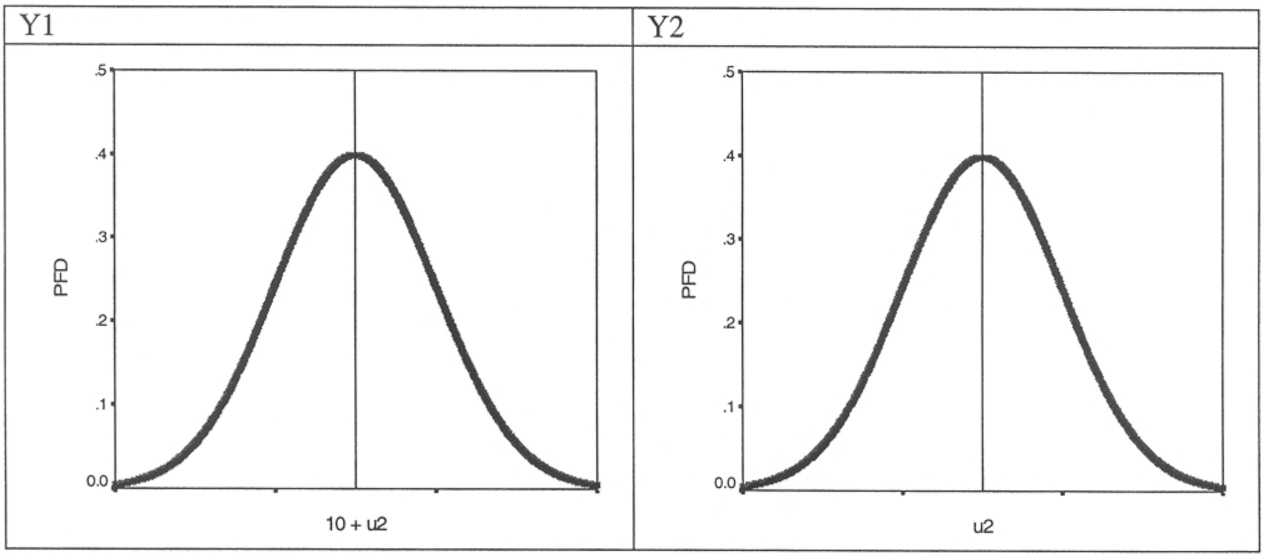
One of the purposes of this exam is to test you on your knowledge. Another purpose is to have you synthesize (bring together) a number of different concepts into a unified approach. Such are the purposes of this problem. There will be four parts to this problem. The following example will illustrate the types of questions for the other four parts.

Example. Suppose we have sample from two normal populations. $Y_1 \sim N(\mu_1, \sigma_1^2)$ and $Y_2 \sim N(\mu_2, \sigma_2^2)$. Also suppose we know that $\mu_1 = 10 + \mu_2$. Answer the following questions?

1) How many population parameters are there?
2) How many parameters do we have to estimate assuming equal variances?
3) Sketch the distribution of Y_1 and Y_2, using what you know.
4) What have we learned?

Answers:

1) there are 4 population parameters - $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$.
2) Since $\mu_1 = 10 + \mu_2$ and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ there are only 2 parameters to estimate.
3) See below.



4) By knowing that $\mu_1 = 10 + \mu_2$, we do not have to sample both populations. Instead of three unknown parameters we only have 2.

The actual problem: Note you should not do any calculations – i.e. means variance etc.

Part 1. A regression Problem

Make all of the usual assumptions.

Given the data in Table 1, do the following:

1) Write the appropriate model.

2) How many populations have been sampled? ___________
3) How many population parameters are there? ___________
4) How many population parameters do we have to estimate? ___________
5) The name of the parameter that regression analysis is most interested in is _____________(not the variance)?
6) Graph the data
7) Put in the least squares line (guess at it)
8) Sketch the distribution of Y when X = 5.

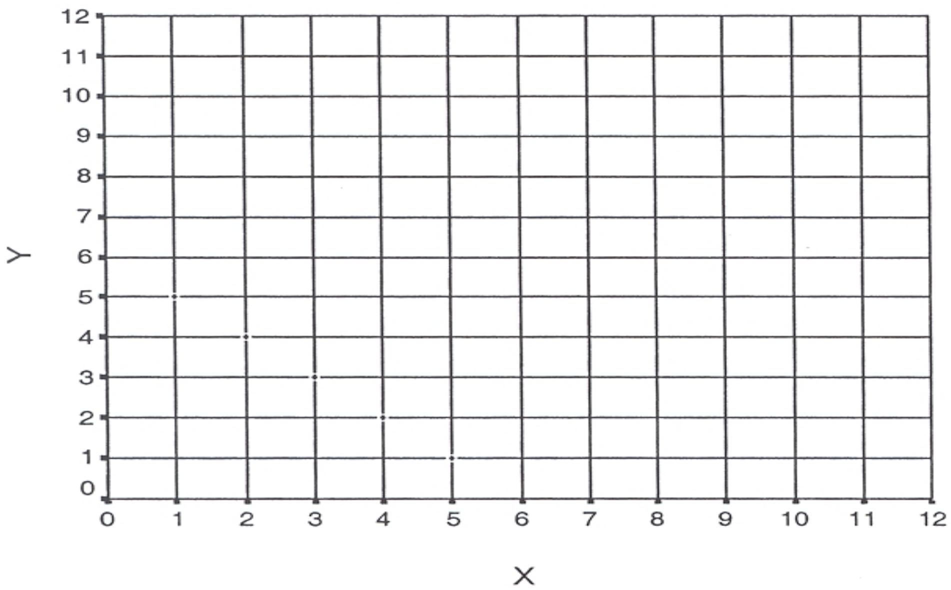


Table 1.

Y	X
1	1
2	3
4	5
5	5
5	8
4	9
7	11

Part 2. An Analysis of Covariance Problem

Make all of the usual assumptions , and assume unequal slopes.

Given the data in Table 2, do the following:

1) Write the appropriate model. Do not use $\mu + \alpha_i$ as the intercepts. Simply use α_i.

2) How many populations have been sampled? ______________
3) How many population parameters are there? ______________________
4) How many population parameters do we have to estimate? _______________
5) The names of the parameters that covariance analysis is most interested in are ________________ and __________________ (not the variance)?
6) Graph the data
7) Put in the least squares lines (guess at them)
8) Sketch the distribution of Y when X = 5.

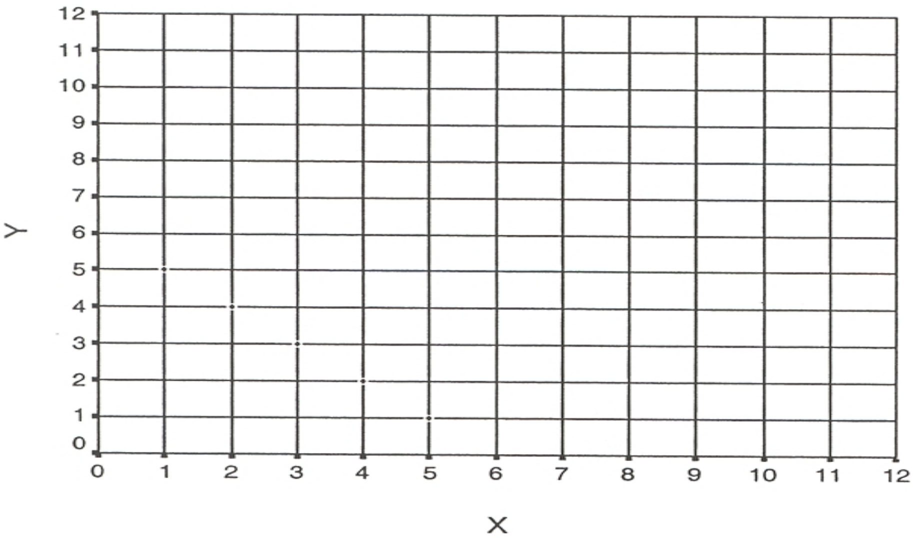


Table 2.

Y	Treatment	X
1	0	1
2	0	3
4	0	5
5	0	5
5	0	8
4	0	9
7	0	11
9	1	2
8	1	4
7	1	7
6	1	7
5	1	10
4	1	12

Part 3. An Analysis of Variance Problem

Make all of the usual assumptions.

Given the data in Table 3, do the following:

Make this look like an analysis of covariance by attaching x = 0 to each y.

1) Write the appropriate model. Do not use $\mu + \alpha_i$ as the intercepts. Simply use α_i.

2) How many populations have been sampled? _______________
3) How many population parameters are there? ______________
4) How many population parameters do we have to estimate? _________________
5) The names of the parameters that analysis of variance is most interested in are _______________ or ________________(not the variance)?
6) Graph the data
7) Put in the least squares lines (guess at them)
8) Sketch the distribution of Y when X = 0.

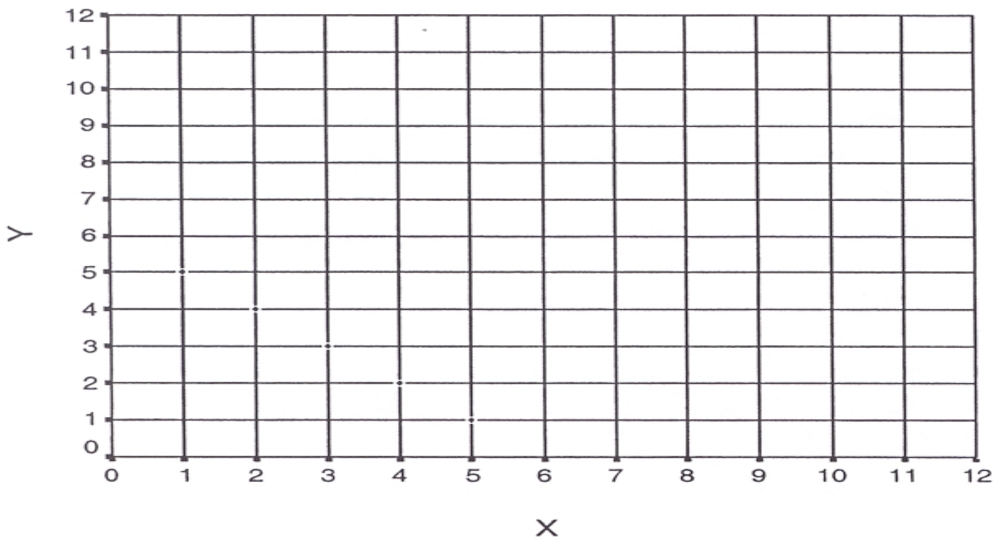


Table 3.

Y	Treatment
2	0
2	0
4	0
5	1
5	1
6	1
7	1
8	1
9	2
10	2
11	2
11	2
12	2

Part 4.

Hopefully, you have seen that you can display regression, ANCOVA and ANOVA all on a similar type Y by X graph. What have you learned by doing the above parts? What is the commonality of the three analyses? Comment on the number of samples per population needed by each approach.