Stat/For/Hort 571 **Zhu and Yandell** **18 November 2003**

Midterm II

Name:____

For the section that you *attend* please indicate:

Instructor:(circle one) Zhu Yandell

TA: (circle one) Guang Cheng Mike Kozloski Ping Wang

Instructions:

1. This exam is open book. You may use textbooks, notebooks, class notes, and a calculator.
2. You do not need to check the assumptions of the procedures that you use unless you are specifically directed to do so. In checking for normality it is sufficient to construct a stem and leaf display. It is *not* necessary to make a normal scores plot.
3. Do all your work in the spaces provided. If you need additional space, use the back of the preceding page, indicating *clearly* that you have done so.
4. To get full credit, you must show your work. Partial credit will be awarded.
5. Some partial computations have been provided on some questions. You may find some *but not necessarily all* of these computations useful. You may assume that these computations are correct.
6. Do not dwell too long on any one question. Answer as many questions as you can.
7. Note that some questions have multiple parts. For some questions, these parts are independent, and so you can work on part (b) or (c) separately from part (a).

For use by graders only:

Question	Possible Points	Score
1	20	
2	20	
3	15	
4	25	
5	20	
Total	100	

1. A mouse study of type II diabetes claims to find an association between glucose level and the genotype (A or H) at a particular genome region on chromosome 2. You have been asked to substantiate the claim based on the following data:

$$\bar{y}_A = 240.0,\ s_A = 50.0,\ n_A = 15$$
$$\bar{y}_B = 190.0,\ s_B = 40.0,\ n_B = 20$$

   (a) Find a 99% confidence interval for the difference in glucose attributable to genotype.

   (b) Interpret the confidence interval as a test for $H_0 : \mu_A = \mu_B$ versus $H_A : \mu_A \neq \mu_B$. What is your decision?

2. Certain forms of vitamin D are believed to alleviate symptoms of multiple sclerosis (MS), a debilitating disease affecting the nervous system. A clinical trial on humans of one form of vitamin D (trt 1) against a placebo (trt 2) showed the following results in terms of symptom reduction: 20 out of 30 patients on vitamin D showed reduced symptoms, as compared to 12 out of 30 patients on the placebo.

   (a) The researchers want to know: Is vitamin D better than the placebo? Perform a significance test. What do you conclude based on the data?

   (b) What assumptions did you need to make for this problem? Briefly justify them for the scientists.

3. A study is examining the effect of different cleaning agents on bacterial growth. This study includes a control (C), being the standard practice, and two new treatments (A, B). While the main interest is comparing A and B, it is also important to verify whether or not they differ from C. Thus an analysis of variance with all three seems an appropriate starting point. You are given the following information

Agent	Sample Size	Mean Growth	Sample Std. Dev.
A	10	34.0	5.0
B	10	30.0	5.0
C	18	36.0	6.0

   While you can build an ANOVA table from just these values, here are some others that may be useful: $y_{\cdot\cdot} = \sum_{\text{all obs}} y_{ij} = 1288$, $\sum_{\text{all obs}} y_{ij}^2 = 44950$, $\sum_{i=1}^{3} y_{i\cdot}^2/n_i = 43888$.

   Verify the entry in the ANOVA table below. Complete the rest of the values for this ANOVA table. Conduct a test at the 5% level of the null hypothesis, that the cleaning agents have equal effect on bacterial growth, versus the alternative that they are not all equal.

Source	df	SS	MS
Agent		231.58	
Error			
Total			

4. An agronomist wants to assess the yield of a variety of barley. He has a number of fields available to plant. Suppose the variance in barley yield is 100 g^2 per unit harvested.

   (a) What sample size would be required to have a 90% confidence interval for the yield that is at most 10 units wide? (Remember that the width of the confidence interval is from the lower end to the upper end!)

   (b) Would the sample size be smaller or larger for a 95% confidence interval? Justify your answer.

   (c) (This problem can be done independently from the above parts. All you need is the variance and the basic problem description.) Consider a two-sided test of $H_0 : \mu = 0$ in which we want the power at $H_A : \mu = 10$ to be 0.90. The agronomist can only afford ten fields. Find a level α so that the alternative $\mu = 10$ has power 0.90 against the null with a sample size of 10 pairs.

5. Consider 20 draws with replacement from an opaque bag with the following numbered blocks:

$$\boxed{1}\quad \boxed{2}\quad \boxed{2}\quad \boxed{3}\quad \boxed{3}\quad \boxed{4}$$

   (a) Find the population mean and variance for a single draw, and the approximate distribution of the sum of 20 draws. (You may assume that 20 is 'large'. State how this might help you.)

   (b) Someone draws ten 1s and ten 2s. What is the chance that they could get such an extreme set of draws? Set up appropriate hypotheses and perform a test. What assumptions are you making?