STA 110B – FALL 1998: PRACTICE FINAL COMPREHENSIVE EXAM (3 HOURS)

In addition to solving these problems, reviewing problems on the midterm exam is a very good idea.

1. (15 pts.) Assume, for purposes of this question, that, in reality, ESP doesn't exist.

Two researchers independently decide to conduct a research study to investigate the existence of ESP. Both researchers compare ESP to a control group and perform a statistical test that rejects the null hypothesis of no ESP effect if the p-value is smaller than .05. In this way they have a probability of rejecting the null hypothesis, even though it is false, of .05.

a) (10 pt) What is the probability that at least one of the researchers will find a significant result (reject the null) and falsely "prove" that ESP does exist?

b) (5 pt) As a), except now there are 7 researchers who independently conduct studies. What is the probability that at least one of the 7 researchers will find significant results and falsely "prove" that ESP does exist?

2. (15 pts.)

This problem is based on a very interesting idea from one of the C2000 projects. The investigator took a random sample of 15 students and measured their level of agreement with the curriculum 2000 proposal, on a scale from 1 to 10. Next, each student asked one parent (randomly chosen, let's say) to learn a bit about the curriculum 2000 proposal and then measured the parent's level of agreement. Here are the results (they are fictitious, but imitate the actual findings).

Student	2	4	2	2	5	5	2	1	2	7	3	6	2	8	1
His/Her parent	6	6	10	10	5	7	8	7	2	5	9	10	2	10	5

a) (5 pts) Compute a 95% probability interval on the difference in agreement level between students and parents. Use a flat prior and assume that the two populations can be considered normal. If you use a Student-t, the computer output below could be useful.

```
MTB > invcdf .95;
SUBC> t 13.
    0.9500    1.7709
MTB > invcdf .95;
SUBC> t 14.
    0.9500    1.7613
MTB > invcdf .95;
SUBC> t 28.
    0.9500    1.7011
SUBC> t 13.
    0.9750    2.1604
MTB > invcdf .975;
SUBC> t 14.
    0.9750    2.1448
MTB > invcdf .975;
SUBC> t 28.
    0.9750    2.0484
MTB > invcdf .975;
```

b) (5 pts) Based on your results, would you say that the agreement levels differ? If so, is the difference statistically significant? Is it practically significant?

c) (5 pts) Do that data suggest that the normal distribution is appropriate for this analysis?

Note: this problem can be approached in one of two ways, depending on whether or not you assume that the two groups are independent. Only one is appropriate. However, it is a good idea to practice both approaches and also to compare the answers and to understand why they differ.

2. (15 pts.)

A random sample of 90 voters in Hillsborough, NC reveals that 54 of them are republican.

a) (10 pts.) What is the probability that the next randomly sampled voter in Hillsborough is also a republican? State any assumptions you make.

b) (5 pts.) What is the probability that the majority of the voters in Hillsborough are republican? You may use a normal approximation.

3. (20 pts.) A small hospital has a blood bank from which it wishes to satisfy its day to day needs for blood transfusions. For blood type A+ in any given month the number of transfusions requested is a random variable T. The hospital must decide how much blood to store at the beginning of the month. Purchasing and storing blood for one transfusion costs, say, \$100 a month. At the end of the month the blood deteriorates and is wasted. If there are more transfusion requested than blood stored, blood has to be obtained from a regional blood bank at significant penalty in terms of costs and delays. Hospital administrators judge that every transfusion demand that has to be satisfied with externally supplied blood costs \$600. A good indication of the number of transfusion requested is the number S of surgical operations scheduled at the hospital in that month.

The joint probability distribution of S and T is given below.

	$T = 0$	$T = 1$	$T = 2$
S = none	.1	0	0
S = low	.1	.09	.01
S = medium	.05	.15	.10
S = high	.01	.19	.20

a) (5 pts.) Are T and S independent? Check the definition and then provide intuition to your answer.

b) (5 pts.) Compute $E(T)$ and $E(T|S = \text{low})$

c) (5 pts.) Suppose $S = \text{low}$. Should the hospital store blood for zero, one or two transfusions?

d) (5 pts.) Suppose that no information is available on S. Should the hospital store blood for zero, one or two transfusions?

5. (20 pt.) It is hard to measure the volume of wood in a tree without taking it down. There are, however, predictors of the volume that are much simpler to measure, such as the diameter of the tree at the base. The sample below is for volume and diameter of 29 black cherry trees in the Allegheny National Forest, PA.

```
        -
      75+                                                     *
        -
volume  -
        -                                           *
        -                                        **
      50+                                           2
        -                                    *
        -                                   *
        -                           * *
        -                            *
      25+               **       * 2
        -              2**2**   *
        -             * *
        -    ***
        -
          +---------+---------+---------+---------+---------+---------+------diameter
        7.5      10.0      12.5      15.0      17.5      20.0
```

A linear regression of the volume on the diameter gave the following results:

```
The regression equation is:  volume = - 37.2 + 5.07 diameter

Predictor        Coef        Stdev    t-ratio
Constant     -37.210        3.331     -11.17
diameter      5.0724       0.2445      20.75

s = 4.201        R-sq = 93.9%
```

a) (5 pts.) Does the value of R^2 suggest that the diameter is a good predictor of the volume?

b) (5 pts.) Construct a 99% confidence interval on the slope.

(continued in the next page)

c) (5 pts.) Prediction intervals for the wood volume of two trees not included in the sample are listed below. Explain briefly why the P.I. at 20 is wider than the P.I. at 15.

	Fit	Stdev.Fit	95% P.I.
diameter = 20	64.238	1.818	(54.859, 73.617)
diameter = 15	38.876	0.877	(30.082, 47.670)

d) (5 pts.) Below is a plot of residuals vs fitted values for the regression. Does the plot suggest that using the diameter squared as a predictor could give better results? Give both a statistical and a physical justification to your answer.

```
          -                                                       *
  resids  -
          -
          -
       1.5+   *
          -                                         *  *
          -    *          **                         *
          -     *        *
          -             *  *          *
       0.0+            * ***            *
          -                 *
          -              *         *            *      *
          -                            *               *
          -                   *
      -1.5+                       *  *         *
          -                          *
          -
            --------+---------+---------+---------+---------+--------fits
                   12        24        36        48        60
```

6. (15 pt.) You are asked to design an experiment to compare the effects of three different drugs (A,B,C), believed to relieve pollen allergies.

a. (5 pts.) Suppose it is possible to successively administer each type of drug to the same patient, without biasing the results. Would you choose to perform:

□ *Design 1.* An experiment with 10 patients each of which receives successively each of the three drugs.

□ *Design 2.* An experiment with 30 patients, 10 receiving drug A, 10 receiving drug B and 10 drug C.

Justify your answer.

b. (5 pts.) If you choose 2, what is the best way to assign patients to the three groups? What if 12 of the patients available for the study are men and the rest are women?