INSTRUCTIONS: Answer 8 questions; each question is worth 13 points.

1 Which of the following statements is(are) C=CORRECT and which is(are) NC=NOT CORRECT?

____ the p-value is the probability that the null hypothesis is true
____ the p-value is the probability of being wrong when we assert that a difference exists
____ the p-value is the probability of obtaining a difference as extreme or more extreme from H0 than the difference observed
____ the p-value is the probability that the observed difference is due to chance

2 Explain why it is unnecessary to estimate SE(p) from the sample when carrying out a test of a population proportion π, although it is usually necessary to estimate SE($\bar{x}$) from the sample in constructing a test of a population mean μ.

3 "The null hypothesis was H_0: π=0.20; the observed proportion was p=0.20. Therefore H_0 is true."

What is wrong with this conclusion?

4 A certain hospital emergency room (ER) admitted victims of automobile accidents at an average of 10 per week when the local highway had a speed limit of 100 Km/h. After the speed limit was reduced to 80 Km/h., 4 highway accident victims were admitted in a randomly selected week.

Give two reasons why you cannot conclude that lowering the speed limit has reduced highway accidents?

5 A representative sample was studied with regard to the occurrence of Psychological Disorders (PD) and Chronic Physical Disease (CPD). Some 8% had PD alone, 29% had CPD alone and 14% had both PD and CPD.

a Display the findings in a 2x2 table.
b Is the observed association between PD and CPD
positive negative zero ?
[statistical test not required]

6 From each of the following confidence intervals (CI's), what can you say about the corresponding p-value (assuming alternative is two-sided)?

		CI	H_0:	p-value:
a	95% CI for μ:	47 ± 3	μ = 45	
b	99% CI for $\mu_1 - \mu_2$:	10 ± 6	$\mu_1 = \mu_2$	
c	95% CI for $\mu_1 - \mu_2$:	5 ± 4	$\mu_1 = \mu_2$	
	99% CI for $\mu_1 - \mu_2$:	5 ± 6		

7 **Relation of Glucose tolerance to complications of pregnancy in nondiabetic women (see handout)**

a What is the single biggest problem with the data display in Figure 1?

In spite of this, there are several things that can be examined:

b In the results section, the authors state "We found that the 35 women who had infants weighing 4000 g or more at birth had two-hour plasma glucose levels in the third trimester significantly higher than those of the 2l4 women who had normal-sized infants mean l07.5 (SD 26) vs 96.4 (SD 23) mg /dl, p=0.0l ".

What statistical test might the authors have used to derive this p-value?

c They add in their next sentence
"We also found a significant correlation between the infant's body weight and the mother's two-hour plasma glucose level in the third trimester (Fig l), although we observed considerable scatter in the data".

What H_0 did they test?

d What is the statistical term for the equation shown at the top of Fig. 1?

What does the equation mean in words?

How would the equation read if body weight were in kilograms?

8 A relative of yours, who is pregnant but has not yet had an ultrasound exam, read the story "Ultrasound can detect Down's syndrome" in the Montreal Gazette (see handout). She is impressed by the performance data in the fourth paragraph of the story. She would like her doctor to check for the two signs when she goes for her ultrasound exam.

a What is the reported sensitivity specificity of the ultrasound screening test?

b What is the reported prevalence of Down's syndrome?

c Before seeing her doctor, what are the odds that your relative is carrying a fetus with Down's syndrome? (assume that the reported prevalence applies to persons of her age)

d She visits her doctor, who detects these two signs on ultrasound. What are the odds now?

9 **Parasympathetic nervous system in nocturnal asthma** (see handout)

Patients were studied at 4am and at 4pm on two successive days. At 4am on the first day, the patient was given either placebo or atropine and at 4pm given the opposite. Flow rates were measured 30 min after receiving the injection. The protocol was repeated on the

second day, varying the drug so that each patient received both placebo and atropine at 4am and at 4pm. The order of giving the drugs was randomized.

a Confidence intervals (CI's) can often be used as 'visual tests of significance' for comparing two means. For example, if the 95% confidence intervals around two means do not overlap substantially (ie if the two means are separated by at least 2~3 SE's of each mean), we can conclude that the means are significantly different at the 0.05 level.

The article states that the 4pm means for placebo and atropine are significantly different ($p \leq 0.04$). Yet, the two CI's [based on ±1SE(mean)] do overlap substantially and would overlap even more if the authors used mean ± $t_{9,95}$SE(mean); the two means differ by less than 1 SE(mean).

Explain the seeming discrepancy between the significance test and the confidence intervals.

b Is there enough data in the diagram to perform the appropriate statistical test comparing the rate after atropine at 4am with the rate after atropine at 4 pm (last sentence in the legend)? If not, why not?

10 **Multicenter Trial of Cryotherapy for Retinopathy** (see handout)

a Determine the α, β, z_α, z_β, π_C and π_T (Colton's notation; π_1 and π_2 in Armitage's notation) used in the sample size determination.

α β z_α z_β π_C π_T

b The majority (139) of the infants were "symmetric cases" ie they had retinopathy in both eyes; one eye (randomly chosen) was treated and the other was not. If we restrict the analysis to these 139 patients, what is the appropriate statistical test to compare the results in the treated and untreated eyes?

c The minority (33) of the infants were "asymmetric cases" ie they had retinopathy in one eye; such children were randomized to have the affected eye treated or not. If we restrict the analysis to these 33 patients, what is the appropriate statistical test to compare the results in the treated and untreated eyes?

11 **Infant mortality among subsequent siblings (brothers and sisters) of infants who died of sudden death syndrome (SIDS)**. See handout.

a From the data shown in the first analysis of Table 1, and ignoring, as they did, the fact that the controls were matched, show how they calculated an odds ratio of 1.7.

b Write a sentence that takes the information contained in the confidence interval around 1.7 and puts it in a form that is understandable to parents of SIDS victims.

c Why did they choose Fisher's exact probability rather than a p-value calculated from say a χ^2 test?

d They state "because of the small numbers of SIDS events used in the calculation of rates, statistical power is insufficient to detect less than a **four**fold difference at the 0.05 level". In other words, they had the power to detect an odds ratio of **4**. Yet, the 95% confidence interval starts at 0.4 and extends as high as 6.6, meaning that the study cannot distinguish between an odds ratio of unity and an odds ratio of 6.

Why the seeming discrepancy between the confidence interval and the power calculation?

e Later on in the article, the authors use the data in the second analysis in Table 1 and compute a 95% confidence interval of 3.5/1000 to 17.8/1000 around the observed rate of 7.4/1000 (ie 6/810).

How did they calculate this confidence interval?

4. Some 100 patients are available for a clinical trial which will compare the responses of those given a certain drug (randomly selected from the 100) to the responses of the remainder, who will receive a placebo. The response will be measured on a continuous scale.

a What statistical test is most appropriate to test whether the response is better in patients who receive the drug?

State explicitly:

H_0: H_{alt}:

Why is it statistically better to use a 50:50 division of the patients into "drug" and "placebo" groups than say a 70:30 division?

13 A recent study examined the change in knowledge of risk factors for breast cancer and breast self examination techniques following a two hour workshop for n=16 female staff in a hospital. Each participant's knowledge was tested before ('pre') and 6 months after ('post') the workshop. On both the 'pre' and 'post' test, knowledge was measured using the same 30 item quiz, which produced a score of 0 if no questions were answered correctly to 30 if all were answered correctly. The following is a summary of the data :

	'Pre' Scores	'Post' Scores
mean	14	18
sd	7	6

a What is an appropriate statistical test of whether knowledge is increased following the workshop?

b Is there enough information in the table to do the test?

c If not, what more is needed?

d If the appropriate statistical test indicated a statistically significant increase in scores, can you conclude that the workshop did in fact improve knowledge? Why?

x In an investigation to determine the influence of sampling variability on the diagnostic yield of liver biopsy, two consecutive samples were obtained from each of 100 patients by redirecting the biopsy needle through a single entry site. The results are shown below:

	Number of specimens abnormal			
	0	1	2	Total
Number of Patients with this many abnormal specimens:	40	20	40	100

a What is the (average) number of abnormal specimens per patient?

b From theoretical considerations, would you expect that the binomial distribution should fit this type of data? Why?

c What number of patients would you expect to have 0, 1, 2 abnormal specimens if the binomial distribution were appropriate?

Blindness study

a Determine the α, β, z_α, z_β, π_C and π_T (Colton's notation; π_1 and π_2 in Armitage's notation) used in the sample size determination.

α β z_α z_β π_C π_T

b The majority (139) of the infants were "symmetric cases" ie they had retinopathy in both eyes; one eye (randomly chosen) was treated and the other was not.

From the data given in the first row of Table 3, set up the appropriate 2x2 data display of the results. Note: 23.1% of 134 is 31; 47.8% of 134 is 64.
Display the calculation that the authors made to arrive at a $\chi 2$ of 27.92 (they didn't use a continuity correction)

What value of $\chi 2$ would have been needed for a $P<0.05$?

c The minority (33) of the infants were "asymmetric cases" ie they had retinopathy in one eye; such children were randomized to have the affected eye treated or not.

From the data given in the second row of Table 3, set up the appropriate 2x2 data display of the results. Note: 26.3% of 19 is 5; 23.1% of 13 is 3.

Clearly, in this group with only one affected eye, whether one uses the "blinded..no pun intended" outcome assessment in Table 2 or the (possibly unblinded) clinical outcome assessment (Table 3), the comparison between the treated and untreated eyes is not definitive and hardly calls for a statistical test.

If we are to be formal, what is the more usual test for the comparison when we have the amounts of data given here?

ALTERNATE Q1

a The SE of the mean is usually calculated as $s/\sqrt{n}$.

In this study, what does n refer to? what does s refer to?

b Why should the SE for the estimated difference (2) - (1) of 45 µg/ml of calcium found with infants' drinks not be calculated as $\sqrt{9^2 + 12^2}$? [the question does not refer to whether to use pooled or unpooled s^2's]