***INSTRUCTIONS**: Be brief and W R I T E C L E A R L Y . Unless specifically asked for, complete calculations [or even complete sentences] are not required. Answer in point form when possible. Write answers in space provided, or on back of sheet if necessary. Completed exam to be handed in at/before the beginning of class on Friday May 23*

Team entries welcome (maximum: 4 per team)

---

1 [5 points]

In a study of the effects of acid rain, a random sample of 100 trees from a particular forest are examined. Forty percent of these show some signs of damage. Indicate with T or F which of the following statements are true and which are not.

(a) 40% is a parameter
(b) 40% is a statistic
(c) 40% of all trees in this forest show signs of damage
(d) more than 40% of the trees in this forest show some signs of damage
(e) less than 40% of the trees in this forest show some signs of damage

2 [5 points]

An athlete suspected of having used steroids is given two tests that operate independently of each other. Test A has probability 0.9 of being positive if steroids have been used. Test B has probability 0.8 of being positive if steroids have been used. If steroids have been used, what is the probability that both tests are negative ? *A tree may help*

(a) 0.02 (b) 0.72 (c) 0.30 (d) 0.28 (e) none of the above

Suppose the two tests are indeed negative. What can now be said about the probability that the athlete has used steroids? *[one sentence]*

Use this example to explain the limitation of a p-value in the interpretation of statistical tests. If you like, make the analogy with medical tests *[one sentence]*

3 [6 points]

In which of the following would X not have a Binomial distribution? Why?

a. X = number of women in different random samples of size 20 from the 1990 directory of statisticians.

b. X = number of occasions, out of a randomly selected sample of 100 occasions during the year, in which you were indoors. (One might use this design to estimate what proportion of time you spend indoors)

c. X = number of months of the year in which it snows in Montreal.

4 [4 points]

A significance test gives a P-value of .04. From this we can... [indicate True/False for each]

(a) reject H_0 at the $\alpha = .01$ level
(b) reject H_0 at the $\alpha = .05$ level
(c) say that the probability that H_0 is false is .04.
(d) say that the probability that H_0 is true is .04.

5 [5 points]

The following is part of a table in a recent paper from the Annals of Internal Medicine on a randomized placebo-controlled trial of low-dose aspirin in patients with chronic stable angina (paper courtesy of Leslie Brailsford from a previous summer)

"Baseline Characteristics of Participants with Chronic Stable Angina in the U.S. Physicians' Health Study

Characteristic	Aspirin Group (n=119)	Placebo Group (n=102)
Mean age, years	63.6 ± 9.3	62 4 ± 8.6
Mean systolic blood pressure, mm Hg	132.5 ± 13.0	132.5 ± 14.4
Mean diastolic blood pressure, mm Hg	80.3 ± 7 8	80.2 ± 7.9
Mean cholesterol level, mmol/L	5.9 ± 1.1	5.8 ± 1.3

Plus-minus values are mean ± SE "

If you were checking this paper for typographical and other errors before it was published, would you have noticed any statistical error(s)? Explain.

6 [8 points]

25 measurements are made of the speed of light. Their average is 300,007 and their SD is 10 Km/sec.

- Fill in the blank: The speed of light is estimated to be ...

True or False? explain your answers

- The measurements differ from 300,007 by an average of 10 or so.

- The average of the 25 measurements differs from 300,007 by 2 or so.
- If a 26th measurement were made, it would differ from the speed of light by 2 or so..

- A 95% CI for the speed of light is 300,007 ± 4.

- A 95% CI for the average of the 25 measurements is 300,007 ± 4.

- Approximately 95% of measurements are within a range of 20 Km/sec.

- If another 25 measurements are made, there is a 95% chance that their average will be in the range 300,007 ± 4 Km/sec.

7 [18 points] Effects of Beer on Breast-fed Infants

JAMA 269 p 1637 April 7, 1993

To the Editor.—In response to a query in JAMA about the value of beer consumption to the breast-feeding mother[] it was concluded that there was a scientific basis for the folklore that beer is a galactagogue[] [galactagogue: "favours the production of milk" -- Dorland's Medical dictionary] Beer, unlike other alcoholic beverages, increases serum prolactin levels[]. The subjects in these studies were normal men and nonlactating women however. To our knowledge, no investigation in this area focused on the lactating women and, perhaps more importantly, determined whether milk intake by breast-fed infants is enhanced when their mothers drink beer.

Recently, we demonstrated that breast-fed infants consumed significantly less milk during the 3-hour testing session in which their mothers drank a small dose of ethanol in orange juice than when the mothers drank orange juice alone.[] Using similar methods, we now report similar effects following alcoholic beer consumption. Each of 11 nursing mothers and their infants was tested on 2 days separated by 1 week. In a counter-balanced fashion, the mother drank a 0.3-g/kg dose of alcoholic beer (Miller, 4.6% vol/vol alcohol) on one testing day and an equal volume of nonalcoholic beer (Miller Sharp's, <0.5% vol/vol alcohol) on the other day. During the next 4 hours, each infant fed on demand. Milk intake was assessed by weighing the infants immediately before and after each feed, the infants' behaviors during breast-feeding were monitored by videotape, and the mothers' perceptions of their lactational performance were determined via questionnaire.

Consistent with our previous findings[], the infants consumed significantly less milk during the testing session in which their mothers drank the alcoholic beer(l49.5± 13.1 mL) than during the session in which the mothers drank the nonalcoholic beer (193.1 ± 18.4 mL, paired t *[df,* 10]=-2.47; P=.03). The mothers were apparently unaware of this difference, however. Regardless of whether they consumed alcoholic or nonalcoholic beer, most mothers believed their infants had ingested enough milk, reported that they experienced a letdown during nursing, and felt that they had milk remaining in their breasts at the end of the feeds. Analyses of the video records revealed that infants terminated approximately 70% of the feedings on each testing day. There was no significant difference for any of these variables between the two conditions.

Because milk intake and the rate of milk synthesis varies from feed to feed,[] a 23% reduction in milk intake may be difficult for women to perceive. Additionally, unlike the bottlefeeding mother who often feeds her infant in response to the amount of formula remaining in the bottle,[] the breast-feeding mother does not have an obvious means of assessing how much milk her infant consumed. Moreover, breast-feeding imposes a more active role on the infant; the infant often determines the pace and duration of the feeding and regulates the amount of milk ingested.[] These factors may explain why the folklore that beer consumption enhances lactational performance has persisted for centuries.

These findings do not imply that occasional beer consumption would decrease overall milk intake by the infant. Nor do they directly test whether beer consumption does or does not act to stimulate the amount of milk produced by the mother. However, they do suggest that such folklore should be carefully evaluated using rigorous methods.

- How would you - a priori, obviously - have decided the sample size for this study?

- Do you have a way to reconstruct the SD of the 11 within-pair differences? If yes, explain how; if not, why not?

- What do you think the ±18.4 and ±13.1 are? What are other possibilities and why do you tend to rule them out?

- Is the p value of 0.03 1- or 2-sided?

- Are you comfortable with the statistical analysis performed? List 2 other tests that were available to the authors.

- In the last paragraph, why are the authors careful about their inferences?

8 [30 points] Paracetamol and Fever

a Entry was limited to children with temperatures between 38°C and 41°C.

Given the mean of 38.9 °C and the SD of 0.9, what can you say about the shape of the frequency distribution over the 38°- 41° interval? (give a sketch)

b *"We estimated a sample size requirement of 210 subjects completing the trial"* (Sample size — paragraph 5 of Methods)

Give the formula by which the authors estimated this (identify what numbers go with what parameters, but leave the calculations to your assistant [who has not taken a statistics course])

c *"Student's t- test and Mann-Whitney (alias Wilcoxon) test..."* (Statistical testing — paragraph 5 of Methods)

Why did the authors use the Mann-Whitney (alias Wilcoxon) test? In light of the n's and the shape of the distribution of duration of fever, was their concern about the use of the t test justified?

d *"The mean duration of fever..."* [paragraph 4 of Results]

Explain in a sentence, in non-technical words, the phrase "the differences were statistically non-significant"

e "The 95% CI for the differences between the paracetamol and placebo groups for duration of fever was -10.0 to +7.1 h"

Explain in non-technical words what this statement says.

f How does this CI add to what is shown in Figure 1?

g How was the CI calculated?

h Before the study, the authors anticipated a SD of 2 days (48 hours) for the duration of fever. The SD of the duration of fever observed in the n=225 is not reported explicitly.

How could one reconstruct this SD from the results given [assume that the SD is the same in the two treatment groups]?

i "Children..were more likely to be rated.as having at least a 1-category improvement in activity...." [2nd last paragraph of Results]

What tests could be used to compare the two groups? Do they all give the same answer?

j *"On the basis of ...completing the trial"* [sample size considerations, first sentence of paragraph 5 of Methods]

"There were no significant differences between groups in mean duration of subsequent fever" [Abstract]

If these two statements were the ONLY information you were given about the trial, what could you conclude?

9 [15 points] Melatonin and Delayed Sleep

a What sample size formula or table would you have shown the authors if they had consulted you concerning sample sizes before doing their study?

b What is it about the study design that makes the required sample size so much smaller than that in Kramer's study?

c What do <u>you</u> consider would be a clinically significant advance in sleep onset time?

d "In all 8 subjects sleep onset time was earlier during melatonin treatment than during placebo" [Abstract]

List 3 possible tests of these data, putting them in order of increasing statistical power [do not carry out the tests, but give references]

e Set up the calculation from which the $p<0.01$ for the 3.49 versus 2.12 [Table II, sleep onset time, melatonin versus placebo] was derived

10 [10 points] Statistical Power and Sample Size

Suppose that on the basis of observing a person on 10 randomly chosen occasions, you classify the person into one of two types

One who, in the 10 observations,

'+' wore the seat belt 'significantly more often than 50%' i.e. $p < 0.05$ 1 sided.
(you infer that the person wears seat belts in a MAJORITY of ALL occasions, not just the 10 observed here)

'–' did not
(because your 'test' is one sided, this category includes the person for whom you might infer that (s)he wears the seat on a 'minority' of ALL occasions and the person for whom you do not have 'sufficient evidence' that (s)he is a 'majority' user)

- On at least how many occasions out of 10 must you observe that the person used a seat belt in order to classify the person as a '+' ? Why?

- Suppose there are really only 5 groups of persons: those who wear their car seat belt on 0% of all possible occasions, those who do so on 25% of all occasions, those who do so on 50%, on 75%, and those who do so on 100%. If you use the sampling and classification scheme above, what proportion of these different groups of persons will you classify as '+' ? Draw these proportions as a type of 'power curve' below. Make sure to label the axes. If you have time, fill in the values for 55%, 60%, ... 95%.

- Repeat the calculations for a system based on samples of size 20

This question can be answered using manipulations of 'pre-calculated' probabilities found in a table in the coursepack and in the text.