Department of Epidemiology and Biostatistics
McGill University

EPI 513-607 (Inferential Statistics)
Final Examination
December 8-10, 1993

INSTRUCTIONS

The answers are to be written in the spaces provided.

When writing, be brief and W R I T E C L E A R L Y .

Unless specifically asked for, complete calculations are not needed. To avoid writing out formulae, just indicate which table or formula would be appropriate and give a reference; explain where one obtains each of the components of the formula.

The points (indicated in **bold** beside each question) add up to 230. The exam will be marked out of 180, so one point deserves one minute of effort.

_____
your ID number or *nom-de-plume*

4 If the government cuts the salaries of all employees by a flat amount of $70 per month, what does this do to the average monthly salary? to the SD?

If, instead, it cuts all salaries by 3%, what happens to the average and the SD?

6 A list has 100 numbers. Each number is either 1, 2 or 3.

(a) the average is 2 and the SD is 0. What is the list?

(b) the SD is 1. What is the list?

(c) can the SD be bigger than 1?

3 Is the SD of the age of 1st year university students closer to 1 month, 1 year or 5 years? Explain your reasoning

4 In a random sample of 400 workers from a working population of 20,000 persons, the average distance they travel to work was 19 Km, with a SD of 20.0 Km. Find an approximate 95% CI for the average distance that all workers in the town travel to work. If this isn't possible, explain why not.

4 4 histograms are shown below for the following variables (in a study of a small town):

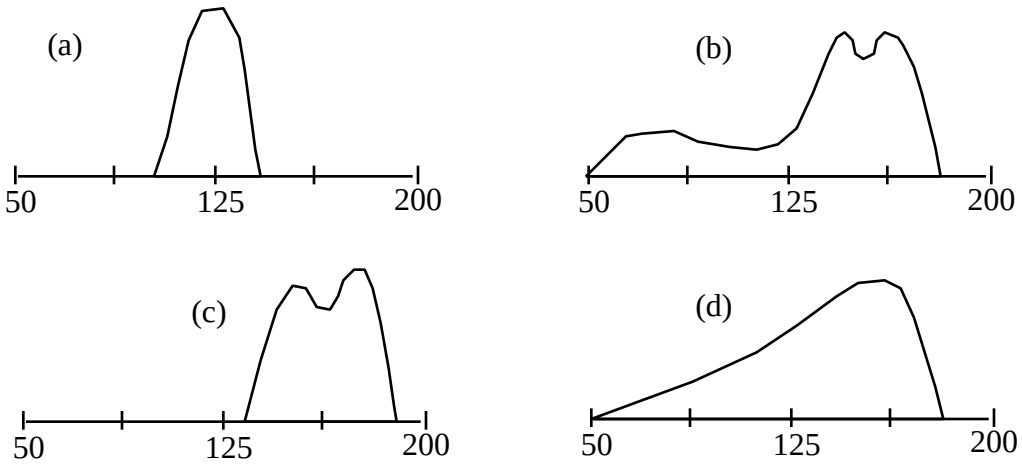
(i) heights of all members of households where both parents are less than 25 years old

(ii) heights of married couples

(iii) heights of all people

(iv) heights of automobiles (cars only)

Match the variables to the diagrams (all heights are in cm's). Explain your reasoning.



5 As genetic theory shows, there is very close to a 50:50 chance that both children in a 2-child family will be of the same sex. Here are two possibilities:
(i) 10 couples have 2 children each. In 8 or more of these families, it will turn out that both children are of the same sex.
(ii) 20 couples have 2 children each. In 16 or more of these families, it will turn out that both children are of the same sex.

Which possibility is more likely, and why?

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

2 Fill in the blanks: The speed of light is estimated to be *300,007;*

a 95% CI is approximately

True or False? explain your answers

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

1 The average of the 25 measurements differs from 300,007 by 2 or so.

1 If a 26th measurement were made, it would differ from the speed of light by 2 or so..

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

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

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

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

8 Fill in the blanks in the following letter to the editor arising from an article on how to estimate the frequency of dizygotic twinning:

> I am puzzled by a calculation in X's article. In it he states that a single ovulation with intercourse will result in a live birth with probability 1/4 and therefore a double ovulation will result in dizygotic twins with probability 1/16. He concludes that the double ovulation frequency must be 16 times the observed dizygotic twinning rate among births.
> This conclusion is incorrect since a large fraction of the double ovulations will not lead to a birth at all. A double ovulation will yield zero births with probability 3/4 x 3/4 = **9/16**, one birth with probability 1/4 x 3/4 + 3/4 x 1/4 = **6/14**, and two births with probability 1/4 x 1/4 = **1/16**. Hence the probability that a double ovulation will yield twins, conditional on there being any birth at all, is **1/16 / { 1/16** + **6/14}** i.e. **1/7.** Therefore X should conclude that the double ovulation frequency is **7** times the observed twinning rate. X should use a tree diagram to see this more clearly.

- A snail starts out to climb a wall. During the day it moves upwards an average of 22 cm (SD 4 cm); during the night, independently of how well it does during the day, it slips back down an average of 12 cm (SD 3 cm). The forward and backward movements on one day/night are also independent of those on another day/night.

5 After 16 days and 16 nights, how much vertical progress will it have made?

4 What is the chance that, after 16 days and 16 nights, it will have progressed at least 150 cm?

2 Did you have to make strong distributional assumptions in order to answer the previous part?

3 The inner planets (Mercury, Venus) are the ones closer to the sun than the Earth is. The other planets are farther away. The masses of the planets (data from 1962) are shown below, with the mass of the Earth taken as 1.

Mercury	Venus	[Earth]	Mars	Jupiter	Saturn	Uranus	Neptune	Pluto
0.05	0.81	[1.00]	0.11	318	95	15	17	0.8

The masses of the inner planets average 0.43 while the masses of the outer planets average 74. Is this difference statistically significant? Or does the question make sense?

5 One large course has 800 students, broken down into section meetings with 25 students each. The section meetings are led by teaching assistants. On the final exam, the class average is 65, with an SD of 15. However, in one section, the average is only 61. The assistant is accused of bad teaching. He argues this way

> If you took 25 students at random from the class, there is a pretty good chance that they would average below 61 on the final. That's what happened to me – chance variation.

Outline carefully the steps and calculations involved in this argument (do not complete the calculations)

- Refer to the attached material on "Math Survey" from NY Times. You need only consider the diagram entitled "How the States Fared on Math" and the 1st paragraph of the 2nd column of text (approximately 2,500 students...)

  4 Approximately what was the SD for the national data (1st row)? What assumptions are you making in order to estimate it?

  4 (BEFORE LOOKING AT THE DATA FOR INDIVIDUAL STATES) For state i, imagine the average score μ_i for ALL CHILDREN (sampled or not) in the state. Suppose there were no, or only trivial, differences in the 37 μ's. Nevertheless, even if this were true, there would still be (sampling) variation among the averages found in the 37 samples. How much would this variation be?

  3 Given your answer to the previous part, and given the observed variation in the 37 estimates of state averages (from a high of approximately 290 to a low of approximately 255), what can you say about the "supposition" that there are no differences in the 37 μ's?

  4 How would you calculate a 95% CI for the difference between the μ for New York and the μ for its neighbour, New Jersey? Do not complete the calculations.

  3 How does this fact affect the estimates of the μ's or of differences between them ?

3 An investigator at a large university is interested in the effect of exercise in maintaining mental ability. He decides to study the faculty members aged 40 to 50, looking separately at two groups: the ones who exercise regularly and the ones who don't. There are large numbers in each group, so he takes a simple random sample of 50 from each group, for detailed study. One of the things he does is to administer an IQ test to the sample people, with the following results:

	regular exercise	no regular exercise
sample size	50	50
average score	132	121
SD of scores	15	15

The difference between the averages is "highly statistically significant". The investigator concludes that exercise does indeed help to maintain mental ability among the faculty members aged 40 to 50 at his university?

Is this conclusion justified? Yes ___

Check one, and say why.

- An investigator wants to show that first-born children score higher on vocabulary tests than second-borns. She will use the WISC vocabulary test (after standardizing for age, children in general have a mean of 30 and a SD of 10 on this test). She considers two study designs:
  a In a school district find a number of 2-child families with both a 1st-born and a 2nd-born enrolled in elementary school.

b From schools in the district, take a sample of 1st-born and a sample of 2nd born children enrolled in elementary school.

4 List 1 statistical/practical advantage of each approach.

3 For the design you prefer, how would you guide her on sample size considerations?

3 For the design you prefer, what would you recommend as a statistical analysis?

4 If instead of a quantitative test, one were interested in whether first-born children are more likely to be left-handed than second-borns, what form would your recommended analysis take?

- A table in a publication went something like the following (denominators imputed, but information not otherwise altered).

  Rates and Rate ratio for hip fracture for men 65 years of age and older in communities with fluoridated vs nonfluoridated water. [MY = man-years of observation]

Fluoridation status	Hip fractures	Rate/1000MY	Rate ratio	95%CI(Rate ratio)
Fluoridated (MY= 5475*]	19	3.47	1.41	1.00 to 1.81
Nonfluoridated [MY=14159*]	32	2.26	1.00	…

  * not reported; back-calculated by JH from the numbers of cases and the rates. [Incidentally, JH is unable to figure out how 3.47 divided by 2.26 gives 1.41, so in fact it is difficult to know where the miscalculation or transcription error is. However, <u>for the purposes of the questions in the exam, assume the ratio and the CI are correct</u>; JH also finds it strange that, with only 51 fractures in all, the CI for a ratio is so "symmetric" but then the authors give an unusual reference, which he has not had time to check, for how they calculated the CI of the ratio from the SE's of the rates).

  It drew the following letter:

  [...]. I am however puzzled by the claim of a significantly increased risk of hip fracture due to artificial fluoridation in [both] men [and women]. Although the male subjects exposed to artificial fluoridation did demonstrate a higher rate than those not exposed, the 95% CI for the ratio includes 1.00. This would seem to indicate that the increased rate is not statistically significant.

  and reply:

  [...] We did have a lower confidence limit that was equal to 1.00. Our P value was equal to *0.05 (2-sided).* There is still a one in *20* chance that the data may be wrong. We chose to report our result as statistically significant: perhaps reporting the P value in this case would have been less confusing.

  3 Just from their CI, fill in their P value directly

  4 Rather than getting the P value via the CI, <u>how would</u> one calculate it from the raw data (assume, to make it easier, that they followed 5475 and 14159 persons, all of them 70 years old, for one

year each, with no losses to follow-up and no deaths, and that they found 19 and 32 hip-fractures respectively) Calculations not necessary

4 Is the authors' interpretation of a P value correct? Can you say it better?.

- In a comparison of the degree of pain experienced when 74 diabetic patients took blood samples from their fingertip versus their abdominal wall, the report states

  "Degrees of pain on finger versus abdominal sampling were, respectively, severe 20 vs 0, normal 31 vs 0, slight 22 vs 2, hardly any 1 vs 7 and none 0 vs 65. Some 88% found abdominal sampling completely painless."

  This is all that the authors said, but here is a 'translation':

	Severe	normal	slight	hardly at all	none
finger	20	31	22	1	0
abdomen	0	0	2	7	65

5 Suggest possible statistical tests of these data, and comment on each one

5 What proportion of all diabetic patients might expect to find abdominal sampling completely painless?

- Refer to attached excerpt from " Suicides by asphyxiation in NY city after publication of *Final Exit. "*

4 Fill in the blanks in the authors' section on statistical analysis:

"the method of suicides by methods recommended in Final Exit during the year after its publication was compared with the number that would have been expected if the publication of the book had no effect. The expected number was defined as

half the total of the two years

3 "The binomial test was used to determine whether there was a statistically significant change". Why does the binomial come into it? (it looks like chi-square test with observed and expected numbers)

4 For comparing the overall numbers of suicides, the authors say in a footnote that they used the normal approximation to the binomial distribution. How does one do this? (don't worry about the continuity correction, which makes only a small difference with the large numbers involved)

2 If someone used a chi-square test instead of the binomial (again without correction), would one get the same p-value? Why? why not?

2 Why is there a 'continuity correction' with these kinds of data?

- This letter appeared in the Lancet in 1991.

  "Dr X and her colleagues' report on transverse limb-reduction defects after chorion villus sampling (CVS) at 56-66 days' gestation prompted us to examine data in the Italian Multicentre Birth Defects Registry. The possibility of an association between CVS and birth defects is a matter of such concern that a report of our preliminary findings is warranted.
  From 6604 cases of malformation reported in 1988-90 we selected the 118 with a transverse limb-reduction defect, this being the one in all of Dr X's cases. The controls were the other 6486 malformed cases. 4 cases of limb-reduction and 15 controls were delivered after CVS (odds ratio 15.14, 95% CI 4.18 to 49.65; p[Fisher's exact] = 0.000305. This approach could underestimate the association, if CVS were to be associated with more than one type of defect; it is a conservative approach, appropriate to the testing of a specific hypothesis, as here...."

  6 From what JH can tell, the alternative hypothesis was one-sided. For their data, list the data points (tables) that make up the 'one tail' in Fisher's exact test.

Observrved table			Others more extreme (make CVS look worse)						
	Cases	Controls							
CVS +	4	15	5	14	6	13	.....	19	0
CVS –	114	6471	113	6472	112	6473	.....	99	6486

  2 Give a reference for how to calculate the 2-tail area if the alternative hypothesis is 2-sided.

  2 Why do you think the authors used this 'exact' technique?

- "Our survey a few years ago found that many of the 41 milk samples we tested contained amounts of Vitamin D quite different from those stated on the label. In order to assess if the increased awareness of the problem has improved matters, we recently purchased 94 milk samples in several locations in the U.S.A. and Canada. Unfortunately, the overall distribution of vitamin D concentrations in these new samples is similar to what we found earlier, i.e. there has been no improvement in the fortification process. In view of the consequences of under- and over-fortification, a unified national program is needed to ensure that proper amounts of vitamin D are included in milk". [adapted from a recent report].

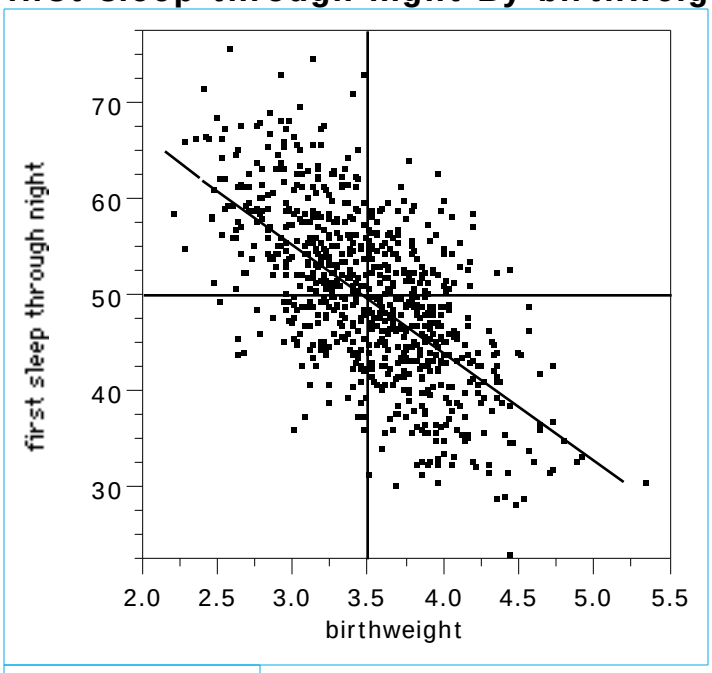
  6 How would you summarize the raw data from the more recent surveys?

  4 How would you compare the results with those of the earlier surveys?

- A study of 800 babies looked at the relationship between their weight at birth and the age at which they first slept through the night. The birth weights averaged 3.5 Kg with an SD of 0.5 Kg. The ages at which they first slept all night averaged 50 days with an SD of 10 days [JH's consultants on these data believe that both the mean and the SD for this latter variable should be much larger!]. The correlation between the two variables was –0.60.

  10 Draw a rough scatterplot of what the raw data look like (you don't have to plot all 800 points).

first sleep through night By birthweight



5 Draw in the regression line. Hint: use the "centered" form: $\mu_{y|x} = \mu_y + ß\ (x - \mu_x)$

5 If the birthweights were in grams rather than Kg, what would ß be? What would the correlation be? Likewise, if the age was measured in weeks, what would change?

5 If we consider a baby that weighed 2.5 Kg at birth, what is the probability that it will sleep through the night before it is 10 weeks (70 days) old? You don't need to DO the calculation, just indicate HOW to. What distributional assumptions do you have to make?

i.e. get Prob(Z > $\dfrac{70 - 62}{\sqrt{1 + \frac{1}{800} + \frac{[2.5-3.5]^2}{\sum\{x - 3.5\}^2}}}$ *)). With n so large here, the second and third terms under the square root sign are negligible.*

5 If we consider all babies that weigh 2.5 Kg at birth, what is the probability that the <u>average</u> age at which they will first sleep through the night is less than 70 days?

- Refer to attached article "Effects of beer on breast fed infants"

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

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

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

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

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

- Refer to attached letter entitled "Thornton Wilder's original design"

5 Suppose, before funding him, the funding agency had asked Brother Juniper to document how reproducible his ratings (and overall index) were. Suppose he had consulted you. What data would you have advised him to assemble to answer this question and what data summaries/presentations would you have recommended?

5 Suppose the pestilence left 9 survivors for every 1 it carried off, and that his budget and other constraints only allowed him to study a total sample size (cases and controls) of only 30 "souls" ("subjects"). Why would Brother Juniper use samples of 15 and 15 rather than say 27 and 3?

5 Translate the last sentence of the passage from the book ("He added up ... ") into the way one might report the results today (you do not have to use a ratio).

5 If he submitted his report to the journal Religio-Epidemiology, the reviewers might had concerns about study design, data analysis, and interpretation. What issues of a more statistical nature might they have raised?

5 Not all statistical tests/procedures available today were available in Brother Juniper's time. What relevant statistical tests would have been available to Brother Juniper if he were analyzing his data today?