

Department of Epidemiology and Biostatistics  
McGill University

EPI 513-607 (Inferential Statistics)  
Midterm Examination  
May 26, 1992

**INSTRUCTIONS**

The answers are to be written in the spaces provided.

Be brief.

PLEASE WRITE LEGIBLY.

For tests of significance, always indicate the null hypothesis and the direction(s) of the alternative hypothesis.

Exact calculations to 4 decimal places are not needed.

For calculations you leave unfinished, say what Table in Colton or Moore&McCabe or Armitage&Berry or what formula would be appropriate; explain where one obtains each of the components of the formula.

Each of the 25 questions is worth 5 points. The best 20 answers will be counted.

The completed examination is to be brought to the 08:30 class on Thursday May 28.

---

your ID number or *nom-de-plume*

**Consider the 3rd and 4th sentences (“The Norwegian researchers... ..to treat tallness”) of the second paragraph of the “Too Tall?” editorial.**

- a If the relevant population base is 600,000 (six hundred thousand) girls, how many of them would meet the inclusion criteria? Use the data given and state any assumptions you make.*
- b From the data given, and any assumptions you make, calculate the 95th percentile of height in the female population.*
- c If the heights of men have a mean of 180 cm, but have the same standard deviation as those of women, what is the probability that a randomly chosen woman is taller than a randomly chosen man?*

**Consider the second and third sentences in the first paragraph of the editorial “Too Tall?”.**

- d State the hypothesis/claim, implied in the second sentence, in formal statistical terms.*
  
  
  
  
  
  
  
  
  
  
- e State the null hypothesis against which you can statistically test the claim.*
  
  
  
  
  
  
  
  
  
  
- f There are a few ways to test this; what test statistic you would use?*
  
  
  
  
  
  
  
  
  
  
- g What reference distribution will you use to describe the sampling variation of the test statistic under the null hypothesis?*

(Third sentence) The number of elections on which the data are based is not given, but say for the sake of this exercise that it is 50.

- h Lay out the steps involved in carrying out the statistical test. You do not need to carry out the detailed calculations but you should give sufficiently clear instructions that a research assistant could follow them in your absence.*



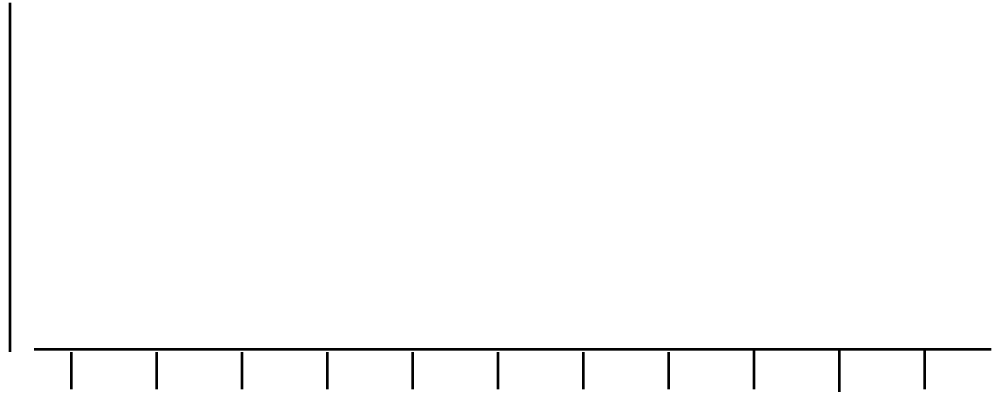
- e The 12.0% in the first line is an average, over 30 patients, of the observed within patient differences between active and placebo treatments. Does this mean that the difference was greater than zero in all 30 patients? Hint: Assume these 30 observations come from a Gaussian distribution and calculate approximately what percentage of the patients showed a greater than zero difference.*

**For parts f and g, refer to the marked sentence in the second column of page 326:** “At the end of the trial, more patients felt that their symptoms were better controlled during cyclosporin than during placebo treatment (21 vs. 8;  $p=0.02$ ; 1 patient reported no difference)”

- f State the null and alternative hypotheses being tested*
- g Verify the  $p$ -value corresponding to this  $21/29 = 72\%$  “preference” for cyclosporin over placebo.*

**Refer to the article entitled “Breast milk and subsequent intelligence quotient in children born preterm”.**

- a *(Table I and 1st sentence of Children and Methods) Add relevant axis labels and scales and draw a rough sketch of the histogram of birthweights for group I.*



- b *(Table I) Are the birthweights in group I more variable than those in Group II? Given the sample sizes involved, is this what you would have expected?*

- c *(Table I) Explain in a layperson's words what the numbers “30 (22,45)” in line 4 represent*

- d (Table I) Does it make sense to use this “3-number” presentation format for describing variation in 'days in study' and quite another format for variation in birthweight and gestation? Why? / Why not?*
- e (Table II) Reconstruct the CI in line 3 (overall IQ) from the means and SEMs given.*
- f The abstract describes the 8.3 point advantage as somewhat over “half a standard deviation”. Does this fit reasonably with the data in Table II ? (the authors may have been using a slightly smaller s.d. than you can calculate; this would reflect the remaining within-group variation in IQ after the variations associated with social class, mother's education, etc. were removed)*
- g (Table IV) From the quoted CI, calculate the SE(Mean Increase) corresponding to the 4.2 points for females in Table IV. Use this SE to calculate the test statistic and check that it agrees with the p value quoted.*

*h Explain this p-value to a general television audience (keep your explanation short enough for a 10 second “soundbite” and avoid the word “significant”).*

*i There are important imbalances in the demographic and socioeconomic composition of groups I and II (Table I). Consider for now just the sex-imbalance. Given this sex-imbalance, and given the estimated difference in average IQ scores of 4.2 between the sexes (Table IV), should the “real” IQ advantage “attributed” to mother's milk be more or less than the “crude” difference seen in Table II? Why? Approximately by how much?*

*[This is a classic case of confounding, and of the use of regression analysis to adjust for it; we haven't come to regression yet, but this is a way to introduce the concept without any fancy equations, just commonsense].*

*j Do you agree with the authors' conclusions (last sentence of Abstract)? Why/why not?*