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.

P L E A S E W R I T E L E G I B L Y .

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.*

- 2.5 s.d.'s leaves a proportion 0.0062 or 0.62% or 6.2 per 1000 in the upper tail so the estimated number > 2.5 s.d.'s above the mean is 600 x 6.2 or 3720.

- Assume Gaussian distribution of heights.

- Cannot really assume predicted height = final height; If there are errors in predictions, and if they are random, then SD(predictions) > SD(Final heights)

b *From the data given, and any assumptions you make, calculate the 95th percentile of height in the female population.*

- If 181 is 2.5 s.d.'s or 14 cm above 167, then 1 s.d. must equal 14/2.5 = 5.6 cm,
  95th %-ile of Gaussian(0,1) distrn is Z = 1.645, so
  95th %-ile of height is 167 + 1.655 x 5.6 = 176.2 cm.

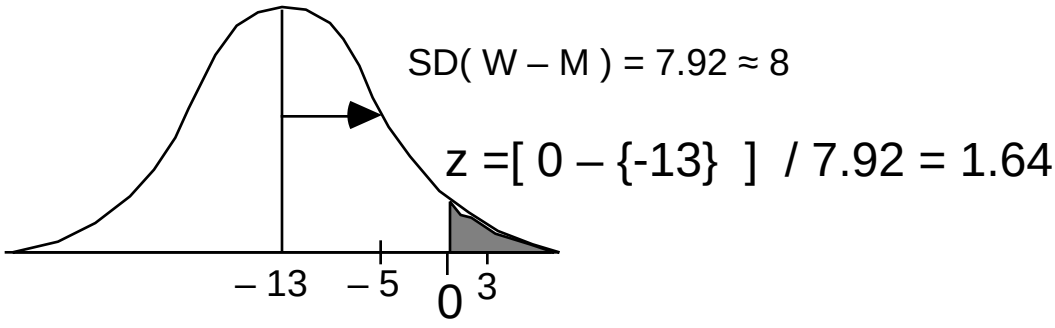
- Again, are using assumption of "Gaussian–ness" here (reasonable if Norwegians are ethnically homogeneous).

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?*

- Let W = Height of woman, M = Height of Man;
  Prob (W > M) = Prob (W - M) > 0
  E [W - M] = 167 - 180 = -13
  Var [W - M] = 5.62^2 + 5.62^2 = 62.72

  SD [W - M] = $\sqrt{62.72}$ = 7.92 [=(2) x 5.6]

  $z = \frac{0 - \{-13\}}{7.92} = 1.64$; Prob(Z > 1.64) = 0.0505 ≈ 5%



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.*

- Taller men are winners; shorter men are losers
  or
  P(Winning | Taller) > P(Winning | Shorter) {results | determinants}
  or
  P(Taller | Win) > P(Taller | Lose) {determinants | results}
  or
  (NOT SO SHARP): Mean Ht. of Winners > Mean Ht. of Losers

e *State the null hypothesis against which you can statistically test the claim.*

- P(Winning | Taller) = P(Winning | Shorter) = 0.5
  or
  P(Taller | Win) > P(Taller | Lose)
  or
  Mean(Height | Winner) = Mean(Height | Loser) i.e. $\Delta = 0$

f *There are a few ways to test this; what test statistic you would use?*

- Compare proportion of Winners in Taller men vs. $\pi = 0.5$
  or
  Compare propn of Taller men among Winners vs. $\pi = 0.5$
  or
  Compare average heights of Winners and Losers

g *What reference distribution will your use to describe the sampling variation of the test statistic under the null hypothesis?*

- Binomial with $\pi = 0.5$, n = # of elections
  or
  t distribution with df = # of elections - 1; pairing avoids the extra variation in heights due to changes in height over 2 centuries.

(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.*

- Binomial n = 50 $\pi = 0.5$, observe 49/50
  Calculate BinProb (49 | $\pi = 0.5$) + BinProb(50 | $\pi = 0.5$)
  or
  $\text{Prob}(Z \geq \frac{|49 - 25| - 0.5}{\sqrt{50 \times 0.5 \times 0.5}})$ or $\text{Prob}((\chi^2_1 \geq \frac{\{|49 - 1| - 1\}^2}{49+1})$
  (the χ^2_1 has easy form when π=0.5 in a "1 x 2" table)

- Paired t–test
  $\text{Prob } t_{49} \geq \frac{\text{average difference in height - 0}}{\text{SD[differences in height]}/\sqrt{50}}$

- Compare the probability with agreed upon alpha.
  If alt. hypothesis were 2-sided, double the P-value.

For parts a - e, refer to Table 1 from an article entitled "Trial of cyclosporin in corticosteroid-dependent chronic severe asthma". [In case you are wondering, it is permissible to do tests and CI's directly on percentage rather than absolute differences]

a *From what formula were the 95% CI's in the table calculated?*

- Average difference $\bar{d} \pm t_{29,95} \times SE[\bar{d}]$

b *What procedure was used to calculate the p-values ?*

- $t = \dfrac{\text{average difference - 0}}{\text{SD[individual differences]} / \sqrt{30}} = \dfrac{\bar{d}}{SE[\bar{d}]}$

  compare with two tails of t_{29} table

c *Calculations: For the second line of data (Morning PEFR, after bronchodilator), determine or back-calculate the values of the components used in the CI calculation; then use these components to calculate the test statistic and check the p-value. (note that the sample size is 30; see last sentence of 1st paragraph of Results)*

- $t_{29,95} = 2.045$ so $17.63 - 10.3 = 2.045\ SE[\bar{d}]$
  so $SE[\bar{d}] = 3.58$

  so t statistic $= \dfrac{10.3 - 0}{3.58} = 2.87$

  which is between the **0.005** (2.756)
  and
  0.0025 (3.038)
  cutoffs (1 tail)

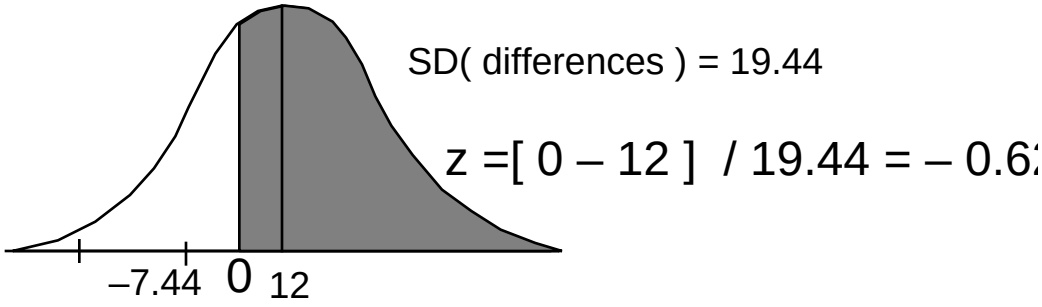
  So P 2-sided is between 0.010 and 0.005;
  (this fits with the 0.009 in the table)

d *Do the same for the fourth line of data and comment on what you find.*

- SE = (11.28 - 5.5)/2.045 = 2.83 so $t = \dfrac{5.5 - 0}{2.83} = 1.95$,
  1.95 is between the 0.05(1.699) & 0.025(2.045) cutoffs (1 tail) of t_{29},
  2 tail P-value is between 0.10 & 0.05(closer to 0.05).
  Wonder if authors used a 1-tailed test here ??
  Also, the 95% CI includes 0 => $P_{2\ tail} > 0.05$!!

e *The 12.0% in the first line is an <u>average</u> , 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.

- SE(average difference) = SD(individual differences)/√30
  So if SE(average) = (19.26 - 12.0)/2.045 = 3.55,
  then SD(individual differences) = √30 x 3.55 = 19.44
  Prob (diff ≥ 0) = Prob (Z ≥ $\frac{0 - 12}{19.44}$) = Prob (Z ≥ -0.62) = 0.7324 or 73%



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*

- Null: Patients are just as likely to feel that their symptoms are better controlled during cyclosporin as they are to feel that they are under placebo, i.e. π(better under Rx) = 0.5

  Alt π(better under Rx) > 0.5 (1-sided)
  or
  π(better under Rx) ≠ 0.5 (2-sided)

g *Verify the p-value corresponding to this 21/29 = 72% "preference" for cyclosporin over placebo.*

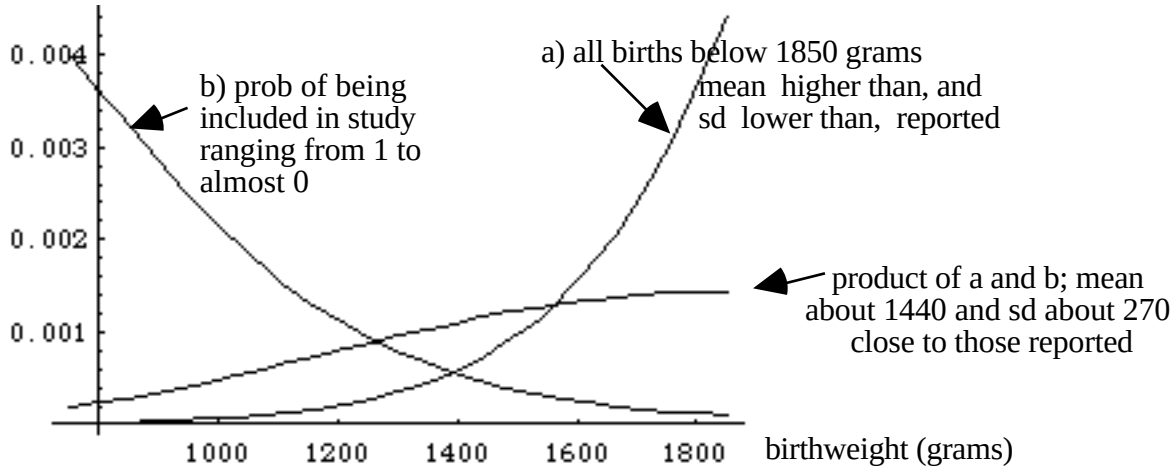
- P(21 or more | n = 29 and π = 0.5) can be evaluated from 1st principles from P(21)+...+P(29) of the binomial,
  or
  by Gaussian Approximation
  P(≥ 21) = Prob(Z ≥ $\frac{|21 - 14.5| - 0.5}{\sqrt{29 \times 0.5 \times 0.5}}$ = Prob (Z ≥ 2.22) = 0.013 [1 sided] or 0.026 [2 sided]
  or
  by chi-square test on a "1 x 2" table {see page 3}

  Doing it exactly by the binomial gives 0.024 [2-sided].

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.*

- 2 clues
  1) SE(mean) = 30 => SD(individuals) = √90 x 30 ≈ 285; distribution of individual birthweights must be skewed because 1420 + 2(285) = 1990 grams, but the data do not go beyond 1850 grams 2) preterms are the lower tail of a ≈ Gaussian curve



y scale represents the proportion of infants per gram of birthweight

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?*

- SD(group I) = √90 x 30 ≈ 285 CV = 285/1420 = 20.5%
  SD(group II) = √210 x 20 ≈ 290 CV = 290/1440 = 20.1%

  These are just about the same.

  SEM's *per se* do not help judge variation of individuals;

  one needs to translate SEM's back to SD's or report SD's in the 1st place! (but see d below).

  sample size has little to do with magnitude of σ (individuals are however variable they are, whether you measure them or not); estimates of σ will be less variable with larger n. So will estimates of μ !

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

- 25% of the individuals spent less than 22 days
  50% of the individuals spent less than 30 days

25% of the individuals spent more than 45 days
so 50% of the individuals spent between 22 and 45 days

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?*

- Quartiles show variation of individuals. If these values have skewed distrn., quartiles more helpful than SD's.

  In any case, we have a mixed format here:
  SEM refers to variation of a mean, whereas
  SD or quartile refers to variation of individuals.
  It should be all one way or all the other

  birthweights were skewed; quartiles better than SD's!

e *(Table II) Reconstruct the CI in line 3 (overall IQ) from the means and SEMs given.*

- 1st way:

  $SE(10.2) = SE(\bar{x}_2 - \bar{x}_1) = \sqrt{SE_2^2 + SE_1^2}$

  $= \sqrt{1.2^2 + 1.6^2} = \sqrt{4} = 2$

  $10.2 \pm 1.96 \times 2 = 6.28$ to 14.12

  2nd way:

  $\text{pooled } s^2 = \frac{\{n_1 - 1\}s_1^2 + \{n_2 - 1\}s_2^2}{n_1 + n_2 - 1} = 280.9;\ s = 16.8$

  $SE(\text{diff}) = \sqrt{s^2 \left\{ \frac{1}{90} + \frac{1}{210} \right\}} = 2.11$

  resulting in CI of 10.2 ± 4.14.

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)*

- $\sqrt{90} \times 1.6 = 15.2$; $\sqrt{210} \times 1.2 = 17.4$
  Estimate of common SD = $\sqrt{280.9} = 16.8$
  So 8.3 is about 1/2 of the "gross" SD.

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.*

- 7.4 = 4.2 + 1.96SE(mean increase) SE = 3.2/1.96 = 1.63
  $t = \frac{4.2 - 0}{1.63} = 2.57 \Rightarrow p = 0.0051$(1-sided) 0.0102(2 sided)

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").*

- Suppose the average IQ at 8 years isn't different between boys and girls who were born preterm; then, given the variation among children, and given the sample sizes involved, a disparity in the average IQ's of the 2 samples as big as or bigger than the one we saw here would be quite unlikely (less than 1%). We should consider the other explanations/interpretations beyond just the play of chance.

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].

- The group II (mother's milk) children had many factors going for it (i.e., advantages) but it had one going against it -- it had more boys. To put II on an equal footing with I, we would have had to add some IQ points to II, thereby eliminating the inbuilt sex advantage of I. Making this correction would increase the gap between II and I beyond the 10.2 points in Table II. If we replaced 13 boys per 100 in group II, with 13 girls, we would have 42% males in each group; this would increase the IQ in group II by 13% of 4.2 i.e., by 0.55 points.

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

- The gender imbalance creates a 0.55 point bias; factors such as education and social class, are much more imbalanced (both favour group II) and much stronger than the effect of gender. Adjusting for them should decrease real difference between the mother's milk and no mother's milk groups to < 10.2 points. The overall downward adjustment by 1.9 points to 8.3 reflects this.
  I'm not at all convinced that the causal interpretation is warranted. Those in Group II are likely to be different in many unmeasured ways that influence IQ; the measurements that were made may be an inadequate reflection of the real influences, and so it is not possible to entirely remove the confounding.
  See the very relevant article by Gray-Donald and Kramer (AJE 127, 1988 pp885-892) and their discussion of how they were unable to "control for" such factors. It is a sobering reminder that Mantel-Haenszel techniques and multiple regression are not a panacea and that the best non-experimental study can sometimes fall far short of the experimental one.