EPI 513-607 **Mid-term Examination** May 31, 1990

Calculations need not be exact.

Q1-Q9: 8 points Q9-Q11(2 of 3):11 pts Q12-Q14(2 of 3):11 pts

1. An investigator has a computer file showing family incomes for 1,000 subjects in a study. These range from $2,800 a year to $78,600 a year. By accident, the highest income in the file gets changed to $786,000.
(a) Does this affect the average? If so, in which direction? **It will increase the mean**
(b) Does this affect the median? If so, in which direction? **No.**

2. Both of the following lists have the same average of 50. Which one has the smaller SD, and why? No computations are necessary.
(i) 50, 40, 60, 30, 70, 25, 75.
(ii) 50, 40, 60, 30, 70, 25, 75, 50, 50, 50.

(ii) has smaller SD. SD based on average (squared) deviation from mean. 2nd string has same mean as 1st but adds 3 deviations of zero... so average (squared) deviation lower than in 1st.

3. As part of a survey, a large company asked 1000 of its employees how far they commute to work each day (round trip). The average round trip commute distance was 11.3 miles, with an SD of 16.2 miles. Would a rough sketch of the histogram for the data look like (i) or (ii) or (iii)? Or is there a mistake somewhere? Explain your answer.

In a Gaussian distribution, mean +/- 2SD comprises approx 95% of all observations. Here, because the SD is much larger than the mean, and because a negative travel distance is impossible, the distribution must be skewed to the right i.e. have a long right tail.

4. A coin will be tossed either 2 times or 100 times. You will win $2.00 if the number of heads is equal to the number of tails, no more and no less. Which is correct?
(i) 2 tosses is better. (ii) 100 tosses is better.
(iii) Both offer the same chance of winning.

(i). Compare binomials in Table in material on Binomial. P(1H1T)=0.5; P(5H5T) = 0.246; P(10H10T) = 0.176....

5. In which of the following would X not have a Binomial distribution? Why?
a. X = number of women in McGill's graduating medical class (n=160 per year) in a randomly selected year from the last 30 years or so.
π(women) has been increasing over time, so wider-than binomial variation over the years.
b. X = number of women listed in different random samples of size 20 from the 1990 directory of statisticians.
yes
c. 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)
if sampling is done completely at random, then OK. If sample consecutive blocks, then there is additional variability because of clustering of activity.

d. X = number of months of the year in which it snows in Montréal.
No. Even though the range is (theoretically 0/12 to 12/12, the aggregate over the 12 is not made up of equiprobable 0/1 events from month to month. P(snow|Jan) is different form P(snow|July)! Assume that on average it snows in 5 of the 12 months so that the probability of snow on a randmly selected month is 5/12 or 0.4. The variation in the number of months with snow from year to year is less than would be predicted from var(count) = 12 x 0.4 x 0.6 = 2.88 or SD(count)=1.3 across years.

6. On the average, hotel guests weigh about 150 pounds with an SD of 25 pounds. An engineer is designing a large elevator for a convention hotel, to lift 100 people. If he designs it to lift 15,500 pounds, the chance it will be overloaded by a random group of 100 people is closest to 0.1 of 1%, 2%, 5%, 50%, 95%, 98%,99.9%
SD(sum)=√n x SD(individuals) so SD(sum of 100) = √100 x 25 = 250. 15,000 is 500 lbs or 2 x SD(sum)'s over the mean or "expectation" of 15000. Thus, the probability of overload is the probability of a z value of 2 or greater i.e. 2.3%

7. The speed of light is measured 25 times by a new procedure. The 25 measurements are recorded, and show no trend or pattern. Then the investigators work out the average and SD of the 25 numbers; the average is 299,789.2 kilometers per second and the SD is 12 kilometers per second. Find an approximate 95% confidence interval for the speed of light, showing your work.
We are interested in what c, the speed of light (if we were to be consistent in our notation, we would have used a Greek letter for the speed of light, since it is a parameter). We are given n=25; xbar=299,789.2; SD(individual measurements)=12. Thus, 95% CI for c is xbar +/- $t_{95,24}$ x 12/√25. $t_{95,24}$ is 2.064.

8. True or false: "If the data do not follow the normal curve, you can't use the curve to get confidence intervals". Explain your answer.
No so. The Central Limit theorem operates... the variability of xbar's and sample proportions is much closer to Gaussian than the variablity of the individuals, especially if n is reasonably large.

9. In a simple random sample (n=225) of all institutions of higher learning in the U.S., the average enrollment was 3,700, with an SD of 6,000. A histogram for the enrollments was plotted and did not follow the normal curve. However, the average enrollment at all institutions in the U.S. was estimated to be around 3,700 (SE = 400).
Say whether each of the following statements are true or false, and explain why.

(a) It is estimated that 95% of the institutions of higher learning in the U.S. enroll between 3,700-800 = 2,900 and 3,700 + 800 = 4,500 students.
False: The interval is for the <u>average</u> enrollment in all US colleges. Individual colleges will vary according to some curve that may be entirely skewed (this is a good example of how colleges have their own distribution, and the fact that you measure a sample of them does not change the inherent variability of all colleges). The reason the variation (of a sample mean) is symmetric (even Gaussian) is because of the Central Limit Theorem.

(b) An approximate 95%-confidence interval for the average enrollment of all institutions runs from 2,900 to 4,500.
Correct

(c) If someone takes a simple random sample of 225 institutions and goes two SE's either way from the average enrollment of the 225 sample schools, there is about a 95% chance that this interval will cover the average enrollment of all schools.
Yes, this is te reasoning behind a CI. We speak of the long-run performance of the <u>procedure</u>.

10. A colony of laboratory mice consisted of several hundred animals. Their average weight was about 30 grams, with an SD of about 5 grams. As part of an experiment, graduate students were instructed to choose 25 animals haphazardly, without any definite method. The average weight of these animals turned out to be around 33 grams. Is choosing animals haphazardly the same as drawing them at random? Discuss briefly, carefully formulating the null hypothesis, and computing Z and P. (There is no need to formulate an alternative hypothesis)
Under random sampling, the average of a group of n=25 should vary about the colony mean (30) with a SD (or more commonly used term a SE) of the mean of $5/\sqrt{25}$ or 1 gram. The particular mean observed is 33, which is (33 - 30)/1 or 3 SE's above the colony average. Under random sampling, a deviation this big or bigger (in <u>either</u> direction from 30) is quite small (0.0027 or 27 times in 10000). So... Note: by calculating prob(Z > 3) PLUS prob (z < -3), we are doing a two-sided test.

11. Which of the following questions does a test of significance deal with?
   (i) Could the difference be due to chance? **Yes; but better to say "chance ALONE"**
   (ii) Is the difference important? **If the average height of a random sample of n = one million Canadians was 0.132 cm lower than than the average height of a random sample of n = one million Americans, this would be statistically significant. However, the 0.132 cm is statistically significant not because of the big difference, but because of the very small SE of 0.007 cm for of the difference in the estimated means due to the very precise estimates obtained with such large n's. looked at another way, through CI's, we can say we are pretty sure (taking +/- 3 SE's) that the difference in POPULATION means is in the range 0.132 +/- 0.030 i.e. in the range 0.102 to 0.162 cm's. Whatever value you pick from this range, it is a trivial difference.**
   (iii) What does the difference prove? **NO.**
   (iv) Was the experiment properly designed?
   NO. Only a serious look at the design can tell you that. Remember the story of the scientist who said, after repeating it consistently in insect after insect, that when you remove all the legs from insects, they cannot hear any more.

12.TREATMENT OF OSTEOPOROSIS NEJM May 3 1990
ABSTRACT
   (i) "Vertebral bone mineral content increased significantly (P <0.01) ..." (First sentence, 3rd paragraph of abstract) What test did they use to arrive at the P value?
   Paired t-test on self-paired differences pre-post.
   If you had been given just the confidence interval of 2.0 to 8.6, what could you have said about the P-value?
   Since the CI does not include zero, we know at least that the p-value is less than 0.05 (2-sided). We could work back from the CI to the SE, and get a more accurate statement if we wanted.

   (ii) The CI of 2.0 to 8.6 for the average change in the treatment group is based on a SE of 1.6; the CI of -7.3 to 1.9 is based on SE of 2.2. Using these two SE's, explain how to construct the CI of 2.4 to 13.6 for the difference between groups (second sentence of 3rd paragraph). Do not worry if your calculation does not match exactly
   The SE of the difference has the form $\sqrt{SE^2 + SE^2}$. so we could compute it as $\sqrt{1.6^2 + 2.2^2}$.

13.PSEUDOHYPERKALEMIA NEJM May 3, 1990
Results

(i) "Alone, the application of the tourniquet had no effect, whereas the addition of clenching increased potassium levels in both the patient and the controls." (first sentence)
The authors used statistical tests on the data from the 4 controls. What test was appropriate?
paired t-test with 4 pre-post pairs so 3 df. Presumably the pre- in each control was an average of all pre measures for that ss and the post- an average of the post measures.

(ii) "Handgrip exercise also raised plasma potassium concentration in the patient (104 mmol per liter) and..." (second sentence).
How would you use the first 4 measurements on just the patient (solid circles) to formally test if the increase was real? (think of Mr. W.P.!)
One could compare the average of the 4 pre-clench measurements with the average of the 2 (or 3?) during clenching. One would use an unpaired t-test.

(iii) If statistical tests for the patient indicated that the effect is real, why study 4 additional subjects?
To see if this is a phenomenon that is generalizable over persons or peculiar to this one patient (its a little like a child who discovers that a switch turns on an electric light; (s)he likes to test it out again in different rooms to see if it is a 'general' phenomenon.

14.BRIGHT LIGHT & NIGHT WORK NEJM May 3, 1990

Results (p 1256)

Sleep-Wake Schedules:

(i) From the SEM of 0:18 for bedtimes (n = 5 individuals in control studies) estimate the interindividual variation in bedtimes.
SEM = SD/√n so SD = √n SD. i.e. SD = √5 x 0.18.

(ii) Why would it be wrong to base the SEM on the total number of nights (7 x 5) rather than the total number of subjects (5)?
The study is about persons, not person-nights. Otherwise, why not save on the costs of recruiting subjects and just study 1 person for 35 nights? The nights within the same person are more likely to be more homogeneous than nights in general.

(iii) What statistical test would you use to compare the average bedtimes of 00:22 and 00:04?
t-test for 2 independent groups

(iv) Assuming we calculate rest time as waking time minus bedtime, why is it not correct to calculate
$[SD(\text{rest time})]^2 = [SD\ (\text{waking time})]^2 + [SD(\text{bedtime})]^2$?
This formula for the variation of a difference only applies if the two copmponents are uncorrelated. Here, one would expect a correlation between the two. There is likely to be a positive covariance. This in turn reduces the variability of the differences, since we subtract a term from the sum of the variances. It is
Var(rest time) =
Var(waking time) + Var(bedtime) - 2Cov(waking time, bedtime)

Circadian Phase

(i) "In contrast, in the treatment studies, the mean final temperature nadir occurred 9.6 hours later than the mean initial temperature nadir (14.53 vs 05:19, P <0.0001)."

What test is appropriate here?
paired t-test

(ii) "The mean shift in the treatment group (-9.6 hours) was significantly greater than in the control group (1.1 hours)."
What test is appropriate here?
t test for two independent groups (although there are some irregularities... 2 subjects served as both treatment and control subjects, 6 others served in only 1 group. etc...).