STT200
Exam 4 Prep

1. A test for efficacy of a new medication has null hypothesis that the medication is no better than placebo versus the alternative that it is better. The P-value of the test is 0.002. It has been decided to adopt a procedure that will reject H0 in one half of one percent of cases when H0 is true.

a. What is H0? What is H1?
ans. H0 is a name we reserve for the null hypothesis. It is described above. H1 is a name we give the alternative hypothesis. It is described above.

b. What action (reject H0 or fail to reject H0) is taken by the test and why?
ans. Any test rejects H0 if and only if the P-value is less than a given threshold called alpha (it is the chance we commit type one error, rejecting H0 when it is true). Here alpha = 0.01 / 2 = 0.005. Since the P-value is given as 0.002 we reject H0.

Remember, the P-value summarizes everything about the test into a number between 0 and 1. It combines all the information, including the choice of H0 and H1, the test statistic and the data, into a single number. That number will be small to the degree that the data disagrees with H0.

2. Determine a 95% t-based CI for the mean of a **normal population** based on the two samples {4.22, 5.22}. Use the fact that, for a sample of n = 2, we have $\frac{s}{\sqrt{n}} = \frac{|x1 - x2|}{2}$. Employ the t-plot for the correct degrees of freedom.
ans. xBAR +/- t $\frac{|x1 - x2|}{2}$ where t = 12.7 is found from the t-plot with DF 2-1=1 by entering C (confidence) = 0.95 on the horizontal C axis of the t-plot. By eye you will not get t so accurately (I used a table) but it will be close enought for hw and exam work. The CI works out to

4.72 +/- 12.7 |4.22-5.22| / 2 = 4.72 +/- 6.35.

Remember, such a 95% CI has around 95% chance of covering the population mean each time it is used. For a t-interval with n not large we require the population distribution to be normal or else the coverage probability may in fact not be close to the agreed upon 0.95.

3. Determine a 70% z-based CI for a random sample of n = 900 from which we find sample mean 34.6 and sample sd s = 18.4. Use the t-plot for infinite df.
ans. Enter C = 0.7 to the t-plot with df infinity to find t = z ~1.04. It is almost the same as z = 1 for 68% CI (rule of thumb). So the z-based CI is xBAR +/- z s / root(n) which works out to
34.6 +/- 1.04 18.4 / root(900)

Remember, the z-interval tends (in many practical applications) to be approximately valid when n is large even in cases where the population distribution is not normal.

4. Determine a 95% z-based CI for the fraction p of male carp in the population of carp in an estuary if, in a random sample of 400 carp, there are 88 males.
ans. z = 1.96 of course. The CI takes form
xBAR +/- z root(pHAT qHAT) / root(n)
with n = 400, sample proportion of males pHAT = 88/400 and sample proportion of females = (400-88)/400 = 1 - 88/400.

Remember, the margin of error always refers to the +/- part of a 95% CI and so in this case is
1.96 root((88/400)(312/400))/root(400)

5. Determine a 95% z-based CI for the difference of two population means based on two independent random samples x1, ..., x100 (from x-population) and y1, ..., y400 (from y-population) with
xBAR = 25.6 yBAR = 24.2
sx = 13.2 sy = 15.4
ans. The form of such a CI is
(xBAR-yBAR) +/- z (estimated sd of (xBAR-yBAR))
(25.6-24.2) +/- 1.96 root(13.2^2/100 + 15.4^2/ 400)

Remember, the theoretical sd of xBAR is $\sigma_x/\sqrt{n}$ and is estimated by s/$\sqrt{n}$. So variance of xBAR is estimated by $s_x{}^2$/n. The reason for using the sum (not difference) 13.2^2/100 + 15.4^2/ 400 is that the variance of a difference (in this case xBAR-yBAR) is the SUM of their respective variances.

6. (multi-part) A z-test of H0: mean < 14 versus H1: mean > 14 will be based on a sample of 400.

a. Is the test one-sided or two-sided?
ans. It is one-sided since the alternative hypothesis H1 lies to one side of H0.

b. Do we reject H0 for large xBAR or small xBAR?
ans. We reject H0 if the data provides evidence strongly against H0. In this case that evidence would take the form of xBAR being much larger than 14 (relative to the estimated sd of xBAR of course).

c. Give the form of the test statistic for this test. Would you reject H0 if xBAR = 13.8? Why?
ans. Since 13.8 is less than 14 it is in the null hypothesis H0 and so does not argue for rejecting H0. The test statistic is

(xBAR - 14)/(s/root(n))

which will be negative if our sample mean is 13.8.

d. Evaluate the test statistic if instead the data yields xBAR = 14.2 and s = 3.5.
ans. Here the test statistic is the positive value

CORRECTED ******************************

(xBAR - 14)/(s/root(n)) = (14.2-14)/(3.5/root(400)) = 1.14.

e. Instead of (c) or (d) suppose the data yields a test statistic value of 2.3. Use the t-plot for infinite DF to determine the P-value of the test.
ans. Enter t = 2.3 to the t-plot for df = infinity and read off C ~ 0.98. This test being one-sided the P-value is

Corrections********

*********(1-C)/2 ~ 0.01.** ******

It is small, true, but one percent is not really so rare.

f. If you are willing to falsely reject H0 with probability 0.05 what action will you take (accept H0 or fail to accept H0) based on the data of (e)?
ans. If you are willing to wrongly reject H0 at a 0.05 rate you reject H0 if the P-value is less than 0.05.

Corrections********

********It is (the P-value is 0.01) so you reject H0.**

Departure of data from H0, as measured by P-value, is rare enough (0.01 in this example) to cause you to reject H0 because you have decided on a 5% rate for type I error.*******

g. Instead of the above, if you are willing to falsely reject H0 with probability 0.03 and your data gives a P-value of 0.031 what action (reject H0 or fail to reject H0) is taken?
ans. You fail to reject H0 since 0.031 is not less than 0.03.

7. (multi-part) A test of H0: mean = 14 versus H1: mean ≠ 14 will be based on a sample of 400.

a. Does the test reject for xBAR very much larger than 14? Does it reject for xBAR very much smaller than 14? Is it one or two-sided?
ans. The test is two-sided since H1 lies on both sides of H0.

b. Give the form of the test statistic and evaluate it for xBAR = 14.2 and s = 3.5.
ans. (xBAR - 14)/(s/root(n)) = (14.2 - 14)/(3.5/root(400)) = 1.14.

c. Based on (b) give the P-value.
ans. Enter t = 1.14 to the t-plot with DF infinity finding C ~ 0.75. This being a two-sided test the P-value is (1 - C) = 0.25 (not at all rare). there is no division by two (meaning (1-C)/2) as required by a one-sided test.

d. If we are willing to wrongly reject H0 with probability 0.01 what action (reject H0 or fail to reject H0) is taken?
ans. Since P-value = 0.25 is not less than 0.01 we fail to reject H0.

8. For data having a given xBAR and s what is the relative size comparison of the CI width for n = 100 versus n = 400? We know larger n is better, but how much so?
ans. Look at the CI term +/- z s/root(n). Since root(100) = 10 and root(400) = 20 the CI for n = 400 will (other things being equal) have half the width of one with n = 100.

9. A plaza has 300 units of equal area. We choose a sample of 50 units and count the people in each unit. From these 50 scores we develop a 95% z-CI for the mean number of persons per unit. That CI (let us suppose) works out to 15.6 ± 2.1.

a. Give the 68% CI for the mean number of persons per unit.
ans. By the rule of thumb for normal distributions, z = 1.96 goes with 95% and z = 1 goes with 68%. So the 68% CI is around

15.6 +/- 2.1 / 2

b. Give the estimated margin of error for the mean number of persons per unit.
ans. 2.1. Remember, margin of error is the half width of the 95% CI.

c. Give the 95% CI for the total number of people in the plaza.
ans. The total number in the plaza is 300 units times the mean number of people in one unit. Since both xBAR and s scale linearly when x is multiplied by 300 the CI for the total plaza count is just

300 15.6 +/- 300 2.1.

10. Exercise 7, page 507.
ans. The large n1 and n2 CI for a difference between two population proportions based on two independent samples of those populations is found on page 496

pHAT1 = 16 / 347 =0.046 (surgery)

pHAT2 = 31/348 = 0.089 (no surgery)

so the 95% CI for the difference p1-p2 of population proportions is

$$(0.046-0.089) \text{ +/- } 1.96 \sqrt{\frac{0.046\ 0.954}{347} + \frac{0.089\ 0.911}{348}}$$

It is much the same idea as in exercise 5, but for the special case of proportions.