1. Short answer questions - 5 points each

   a. We simulate 100 data sets of 10 observations by drawing a total of 1,000 iid variates from a distribution with mean μ and variance σ^2. We take the sample mean of each of the 100 data sets and estimate the standard deviation of the resulting 100 numbers. What is that an estimate of? (explain, briefly)

   The resulting number estimates the standard deviation of the sampling distribution of the sample mean. Hence it is estimating the standard error, $\sigma/10$.

   b. Will a Student's T or Z hypothesis test for a mean with the data recorded in pounds always agree with the same test conducted on the same data recorded in kilograms? (explain)

   Yes. The T and Z statistic are scale independent as the numerator (the sample mean minus the hypothesized mean) and denominator (the standard error) have the same units. More mathematically (this was not necessary), if $X_i = Y_i * a + b$ then convince yourself that the T and Z statistic will be the same regardless of whether or not X_i or Y_i are used.

   c. What is the variance of $\bar{X} - \bar{Y}$ assuming that the means are comprised of independent groups of iid random variables? (show your work)

$$\mathrm{Var}(\bar{X} - \bar{Y}) = \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y}) = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}$$

   d. A researcher consulting you is estimating a prevalence, p, with a 95% confidence interval of $[.15, .34]$. He claims that there is a 95% probability that the true value of p is between .15 and .34. Would you have anything to say? (explain)

   In the frequentist interpretation of a confidence interval, it is the interval that is random and not the parameter. Moreover, after the data have been observed p is either in the interval or it is not. Therefore the probability is either 0 or 1. Note that Bayesians (people who adhere to a branch of statistics that we do not teach in this class), would have a different answer to this problem.

   e. Another researcher consulting you is very concerned about falsely rejecting her null hypothesis. As a result, she decides to increase the sample size of her study. Would you have anything to say? (explain)

You were given during the exam that the researcher was performing a fixed α level test. She is concerned with having a high type I error rate. The type I error rate is fixed by her choice of α. Setting $\alpha = .05$ results in a 5% error rate regardless of the sample size[1]. This was all I was asking for.

Many people confused the statement "as α goes down n goes up" (referring to sample size calculations) from the review notes with this problem. However, the reverse of this statement that people tried to use "if I choose a larger n then necessarily my α decreases", doesn't make a lot of sense. α is fixed by the researcher.

I should also mention that some people astutely noted that n has to be larger for the CLT to kick in. A poor CLT approximation would impact the actual type I error rate and hence increasing it would force better adherence to the nominal level. Kudos to those who mentioned this.

f. What does it mean that S^2 is an unbiased estimate of σ^2. Would you expect $\log S^2$ to be an unbiased estimate of $\log \sigma^2$? (explain intuitively; do not give a formal proof; be brief)

Unbiasedness means that $E[S^2] = \sigma^2$. We would not expect $\log S^2$ to be unbiased because we would not expect $E[\log S^2] = \log\{E[S^2]\} = \log \sigma^2$. That is, in general, expected values do not commute across non-linear functions like log.

2. Researchers are interested in a new blood test for diagnosing a viral disease. For this blood test a *positive* result is supposed to indicate the *presence* of the disease. A study found that of 50 patients who are known to have the disease, 45 are *positive* on the blood test. Of 200 individuals who are known not to have the disease, 51 are *positive* on the blood test.

a. (9 points) Researchers are interested in whether or not the new blood test improves on the 70% specificity of the existing standard. State relevant hypotheses, perform the relevant test and calculate and interpret the associated P-value.

Let p be the specificity of the blood test. We want to test the hypothesis that $H_0 : p = .7$ versus $H_1 : p > .7$. The estimated specificity of the new blood test is $\frac{149}{200} = .745$. The Z statistic is

$$Z = \frac{.745 - .70}{\sqrt{\frac{(.7)(.3)}{200}}} = 1.38.$$

Since Z is not larger than $Z_{.95} = 1.645$ we would fail to reject. The P-value is $P(Z > 1.38) = .08$.

In conclusion, the new test has a higher estimated sensitivity than the existing standard, and there is marginal evidence that this relationship is not due to chance alone. However, there is insufficient evidence to support this conclusion at the standard .05 level.

[1] Though keep reading for more discussion on this point.

I did not take off for using the Wald test (you should generally use the score) or doing a two sided test (they specifically wanted to know if the new test improves on the old one).

b. (9 points) A person has a positive blood test result. Interpret this positive test result without knowledge of the disease prevalence. Repeat this calculation for a negative blood test.

The instructions suggested using the estimated sensitivity and specificity to answer this and the next questions. The likelihood ratio gives a way to interpret the positive test without the disease prevalence because

$$\frac{P(D|+)}{P(D^c|+)} = \frac{P(+|D)}{P(+|D^c)}\frac{P(D)}{P(D^c)}.$$

The likelihood ratio of a positive test result is $\frac{P(+|D)}{P(+|D^c)} = \frac{45/50}{51/200} = 3.52$. Therefore the odds of disease given the positive test result is 3.52 times that of the prior odds of disease. Or, there is about 3 and one half times as much evidence supporting the hypothesis that this person has the disease than the hypothesis that they do not. Use this interpretation to the the negative result.

c. (9 points) Given that the prevalence of the disease is 10% in a subject's population with a positive test result, calculate the probability that the person has the disease.

Use Baye's rule. Sensitivity $= P(+|D) = 45/50 = .90$, Specificity $= P(-|D^c) = 149/200 = .745$, Prevalence $= P(D) = .1$.

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} = \frac{.9(.1)}{.9(.1) + (1 - .745) * .9} = 28\%.$$

3. Researchers studying brain volume found that in a random sample of 16 sixty five year old subjects with Alzheimer's disease, the average loss in grey matter volume as a person aged four years was .1 mm^3 with a standard deviation of .04 mm^3.

a. (10 points) Motivate a general formula for a $100(1-\alpha)$% confidence interval for this setting. Also calculate and interpret a confidence interval for the grey matter study given in the problem. Show your work.

The motivation for the CI is that:

$$1-\alpha = P\left(-t_{1-\alpha/2} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{1-\alpha/2}\right) = P\left(\bar{X} - t_{1-\alpha/2}S/\sqrt{n} \geq \mu \geq \bar{X} + t_{1-\alpha/2}S/\sqrt{n}\right),$$

provided that the X_i (observations that comprise $\bar{X}$) are independent identically distributed normal random variables. The CI estimate is $.1 \pm 2.13\frac{.04}{4} = [.088, .112]$.

b. (8 points) Calculate and interpret a P-value for the hypothesis that there is no loss in grey matter volumes as people age. Show your work.

The test statistic is $TS = \frac{.1}{.04/4} = 10$. We can't actually compute the P-value, because the probability that a T with 15 degrees of freedom is larger than 10, given the tables. However, 10 (was chosen because it) is quite large for a T with 15 degrees of freedom and hence we can conclude that the P-value is very small (less than the lowest percentage shown on the t table). The actual value is on the order of 10^{-8}. Because the P-value is small, there is strong evidence to suggest that there is mean loss in grey matter volume in the AD group. Furthermore, if we were conducting a hypothesis test of $H_0 : \mu = 0$ versus $H_a : \mu > 0$ (we're treating loss as positive here), our P-value indicates that we would reject the null hypothesis.

c. (10 points) The researchers would now like to plan a similar study in 100 healthy adults to detect a four year mean loss of .01 mm^3. Motivate a general formula for power calculations in this setting and calculate the power for a test with $\alpha = .05$? Assume that the variation in grey matter loss will be similar to that estimated in the Alzheimer's study.

$$
\begin{aligned}
\text{Power} &= P\left(\frac{\bar{X}}{\sigma/\sqrt{n}} \geq Z_{1-\alpha} | \mu = .01\right) \\
&= P\left(\frac{\bar{X} - .01}{\sigma/\sqrt{n}} \geq Z_{1-\alpha} - \frac{.01}{\sigma/\sqrt{n}} | \mu = .01\right) \\
&= P\left(Z \geq Z_{1-\alpha} - \frac{.01}{\sigma/\sqrt{n}}\right) \\
&= P\left(Z \geq 1.645 - \frac{.01}{.04/10}\right) \\
&= P(Z > -.855) \\
&= .80.
\end{aligned}
$$

4. A recent Daily Planet article reported on a study of a two week weight loss program. The study reported a 95% confidence interval for weight loss from baseline of [2 lbs, 6 lbs]. (There was no control group, all subjects were on the weight loss program.) The exact sample size was not given, though it was known to be over 200.

   a. (6 points) What can be said of a $\alpha = 5\%$ hypothesis test of whether or not there was any weight change from baseline? Can you determine the result of a $\alpha = 10\%$ test without any additional calculation or information? (explain your answers)

   Because 0 is not in the interval, a two sided test hypothesis test (whether or not there was any weight change from baseline) would reject the null hypothesis at the .05 level. Though we do not specifically mention what exact kind of interval is being used, for most reasonable interval procedures, a 90% interval would be narrower and contained in the 95% interval.

In particular, this would be true in the (likely) event that they were using a Z interval. Therefore the 10% test would reject also. Others also correctly mentioned that if we reject the null hypothesis with 5% percent chance of making a type I error, then we would surely also reject if we were less concerned with making a type I error. Many people actually calculated the confidence interval, this was unnecessary and (in fact) did not answer the question.

b. (9 points) Can you calculate a P-value for the hypothesis test from Part a using the information given? If so, report the P-value. If not, please explain what extra information you would need to do so. (show your work)

I think the answer "I would need to know exactly how the researchers calculated the interval." would be technically correct, though certainly not answering the spirit of the question. (No one actually tried this). Instead, pretty much everyone acknowledged that the researchers probably conducted a Z confidence interval. In which case the midpoint of the interval, 4 lbs, was the sample mean and the half width of the interval, 2 lbs, must equal the $1.96S/\sqrt{n}$ part. Therefore, $S/\sqrt{n} = 2/1.96 = 1.02$. So our test statistic is $4/1.02 = 3.92$. The probability of a Z statistic being larger than 3.92 is less than .001, so twice that (the P-value for the two sided test), remains very small.