HOMEWORK 10A

Based on Problem 10.66. Researchers believe that cigarette smoking has a negative effect on lung function. One measure of lung function is the carbon monoxide diffusing capacity of the lung (DL). The average DL for a non-smoker is 100 and the associated standard deviation is 15. A random sample of 47 smokers is found to have an average DL of 91.3. Fully conduct an appropriate hypothesis test to determine whether this represents evidence that smoking negatively affects lung function. **Show all steps!** If it does, use a confidence interval to estimate the size of the effect.

We identify this as a hypothesis testing situation for means. Defining μ to be the average diffusing capacity of the lung (DL) in all smokers, we will test:

H_0: $\mu = 100$ vs. H_a: $\mu < 100$

The reason for the one-sided alternative hypothesis is that we want to show there is a negative (lowering) effect.

In order to do this test, we assume unbiased sampling. While the problem claims the sample to be random, simple logic suggests this is not reasonable. To collect a random sample one would need a complete list of all smokers from which to randomly select. The reality is that this must be some type of convenience sample. There is not much information here, but unless it was collected in some systematically weird way (e.g. from participants in an asthma convention), we would have reason to hope it wouldn't be biased. The other assumption is that 47 is a large enough sample size to apply the CLT. Since we might expect these measurements to be somewhat symmetrically distributed, any sample size greater than 30 should do.

We now consider error analysis. A type 1 error will result in incorrectly concluding that smoking is harmful, thereby potentially costing cigarette companies in terms of their profit margin. Since I do not particularly care if I make this error, I will immediately set α = P(Type I error) = 0.10. Note that a Type II error would be of greater importance as there would be a health risk but we would be unable to detect it.

The test statistic here will be $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$. There are multiple methods from here – you can either use a rejection region or the p-value method. The rejection region method will have us rejecting whenever Z is too small (since it is a onesided test). With a 10% chance of a Type I error, we will reject for any Z-value less than -1.28. Our computed Z is -3.97. Since this value of Z is in the rejection region, we conclude that smoking is associated with decreased lung function. We may make the same decision if we compute the p-value: P(Z < -3.97) = 0.0001 and note that it is less than 10%

Having concluded that there is a decrease in average lung function for smokers as compared to non-smokers, we know wish to find information about the magnitude of that decrease. This is best accomplished with a 95% upper confidence bound:

$$\bar{x} + z\frac{\sigma}{\sqrt{n}} = 91.3 + 1.28\left(\frac{15}{\sqrt{47}}\right) = 94.1$$

With 95% confidence we estimate that lung capacity is decreased by smoking by at least 6 units.