R Assignment 3

Parker DeBruyne: V00837207

Note: For each of the following, carry out your calculations only using R or RStudio. Copy and paste your command(s) and the output into your document as indicated. Format your solutions in a somewhat formal way, as if you are writing a lab report (that is, use somewhat formal language and complete sentences).

Part 1 [5 marks]

In a study of factors thought to be responsible for the adverse effects of smoking on human reproduction, cadmium level determinations (nanograms per gram) were made on placenta tissue of a random sample of 22 mothers who were smokers and an independent random sample of 18 non-smoking mothers. The results were as follows:

Non-smokers: 17.4 14.1 19.7 20.2 15.5 28.2 18.7 18.0 21.4 14.6 14.3 13.9 19.6 18.7 17.4 18.8 20.1 16.0

Smokers: 25.5 24.3 20.4 18.5 29.0 31.1 24.6 19.9 18.3 24.8 30.0 13.6 25.2 27.1 20.9 22.6 26.8 20.2 25.3 22.9 20.2 32.4

(a) Store the data for non-smokers in a vector called nonsmokers. We want to test the research hypothesis that the mean cadmium level for non-smoking mothers is less than 19 nanograms per gram. Copy and paste the appropriate R command and output you would use to do this test.

```r
nonsmokers = c(17.4, 14.1, 19.7, 20.2, 15.5, 28.2, 18.7, 18.0, 21.4,
               14.6, 14.3, 13.9, 19.6, 18.7, 17.4, 18.8, 20.1, 16.0)

t.test(nonsmokers, mu=19, alternative="less")
```

```
	One Sample t-test

data:  nonsmokers
t = -1.0588, df = 17, p-value = 0.1522
alternative hypothesis: true mean is less than 19
95 percent confidence interval:
     -Inf 19.55008
sample estimates:
mean of x
```

```
 18.14444
```

(b) Clearly state the p-value you found from part (a). Using the significance level α = 0.10, should we reject the null hypothesis?

The p-value from (a) is 0.1522
This value is greater than α = 0.10, therefore we cannot reject the null hypothesis.

(c) Write a sentence stating your conclusion of the hypothesis test you just performed. (i.e. Write your conclusion in plain language so that a non-statistician could understand the conclusion of the test.)

Even though 18.14444 *is* less than 19, it is not different *enough* from 19 to prove that the mean is smaller than 19 for all non-smoking mothers everywhere; the sample mean must be significantly different for us to make that conclusion because of random sampling.

(d) Store the data for smokers in a vector called smokers. Calculate a 95% confidence interval (i.e. a two-sided confidence interval) for the mean cadmium level of mothers who smoke. Copy and paste the appropriate R command and output you would use to create your confidence interval. In an additional line, clearly state the values you get for the confidence interval from your output.

```
smokers = c(25.5, 24.3, 20.4, 18.5, 29.0, 31.1, 24.6, 19.9,
            18.3, 24.8, 30.0, 13.6, 25.2, 27.1, 20.9, 22.6,
            26.8, 20.2, 25.3, 22.9, 20.2, 32.4)

t.test(smokers, mu=25, alternative="two.sided")
```

```
	One Sample t-test

data:  smokers
t = -1.2195, df = 21, p-value = 0.2362
alternative hypothesis: true mean is not equal to 25
95 percent confidence interval:
 21.75365 25.84635
sample estimates:
mean of x
     23.8
```

95 percent confidence interval: [21.75365, 25.84635]

(e) Is it reasonable to conclude that the mean cadmium level of mothers who smoke is 25 nanograms per gram? Explain, using your confidence interval.

No, you cannot conclude the true value of a statistic from a confidence interval, but when a value falls outside of the confidence interval we can reject the null hypothesis; because 25 is within [21.75365, 25.84635], we cannot reject the null.

Part 2 [5 marks]

From a random sample of 721 items made by a particular manufacturing process, it is found that 56 are defective.

(a) Suppose we want to test the research hypothesis that to proportion of defective items is greater than 6%. Copy and paste the appropriate R command and output you would use to do this test.

```
binom.test(x=37, n=721, p=0.06, alternative="greater")
```

```
	Exact binomial test

data:  37 and 721
number of successes = 37, number of trials = 721, p-value = 0.8563
alternative hypothesis: true probability of success is greater than
0.06
95 percent confidence interval:
 0.03849977 1.00000000
sample estimates:
probability of success
            0.05131761
```

(b) Clearly state the p-value you found from part (a). Using the significance level α = 0.05, should we reject the null hypothesis?

p-value = 0.8563
We cannot reject the null because the p-value is greater than $\alpha = 0.05$

(c) Write a sentence stating your conclusion of the hypothesis test you just performed. (i.e. Write your conclusion in plain language so that a non-statistician could understand the conclusion of the test.)

From our sample, we found that the proportion of defective items is about 5%. Even though this is random, it is too far below 6% to conclude that the proportion of all defective items is greater than 6%.

(d) Suppose that in another manufacturing center in 529 items made by the same manufacturing process as above, it is found that 27 items are defective. Calculate a 95% confidence interval (i.e. a two-sided confidence interval) for the true proportion of defective items manufactured at this plant. Copy and paste the appropriate R command and output you would use to do this question. In an additional line, clearly state the values you get for the confidence interval from your output.

```
binom.test(x=27, n=529, p=0.08, alternative="two.sided")
```

```
	Exact binomial test

data:  27 and 529
number of successes = 27, number of trials = 529, p-value = 0.01269
alternative hypothesis: true probability of success is not equal to
0.08
95 percent confidence interval:
 0.03390134 0.07339402
sample estimates:
probability of success
             0.0510397
```

95 percent confidence interval: [0.03390134, 0.07339402]

(e) According to your confidence interval in part (d), is it reasonable to conclude that 8% of the items created at this plant are defective? Explain, using your confidence interval.

Since the p-value 0.0510397 falls within our confidence interval of [0.03390134, 0.07339402], we cannot reject the null hypothesis that 8% of items created at this plant are defective, but that does not prove that exactly 8% of items created at the plat are defective.