Lecture/text homework assignment # 6

Note: Please circle your answers when appropriate!

0) Problems 6 & 7 from last time are due; note that your recitation instructor may ask you to present them.

1) A male African (Cape) buffalo weighs about 747 kg, with a standard deviation of about 120 kg. Let's assume $\mu = 747$ kg, and $\sigma = 120$ kg.

a) Calculate $Pr\{Y > 850 \text{ kg}\}$

b) Now assume you have a sample of $n = 8$ buffalos. Calculate $Pr\{\bar{Y} > 850\}$.

c) Which probability is lower? ***Why?***

2) Refer to problem 1. Now assume you take a sample of $n = 5$ buffalos.

a) What is the probability of the sample average being within 70 kg of the population mean (747 kg)?

In other words, you need to figure out $Pr\{\mu - 70 < \bar{Y} < \mu + 70\}$

b) What is the probability of the sample average being within 70 kg of the population mean if the population mean is actually $\mu = 800$ kg?

Are you surprised? ***Why or why not?*** You should make sure you understand what happened here.

c) Now assume a sample of $n = 15$ buffalos and repeat (a).

What is the effect of sample size?

3) You go out and collect the following estimates of earthworms / acre:

54526 58628 51358 66770 66482 59268 57409

These data yield the following: $\bar{y} = 59{,}205.86$, and $s = 5{,}733.16$

a) Construct a 95% CI for these data.

b) Construct a 99% CI for these data.

c) Darwin once estimated that an acre of soil had about 50,000 worms in it. Is his estimate consistent with the data above? (*Historical note: His estimate was considered way too high in his day*).

4) Consider the results of 3(b). Notice that all the data fit within the 99% confidence interval. Is this usually the case (in other words, will a 99% CI contain most of the observations)?

> ***Caution: a lot of people get this wrong!*** *Here's a hint: suppose you had measured the worms in 6000 acres (instead of just 7). What happens to the confidence interval? If you're not sure, substitute 6,000 for 7 in your calculation for 3(b) to see what happens.*

5) Suppose the (sample) average height of women in the U.S. is $\bar{y} = 162.8$ cm, and the average height of men in the U.S. is $\bar{y} = 176.4$ cm. For both the sample standard deviation is about $s = 7.65$ cm.

> (a) Suppose you take a sample of 12 women and 11 men. Construct a 99% CI for both. Do the confidence intervals overlap?
>
> (b) Now repeat using a sample of 85 women and 173 men. Do the confidence intervals overlap?
>
> (c) Can you explain what happened? Why is sample size important if you're trying to find differences between groups?
>
> *(Statistical comment: comparing CI's is* ***not*** *the correct way of comparing two groups, but it can give you an idea if there's a difference).*

6) Use the data for heights that you collected in recitation last week and construct a 90% CI. *Use R to do this.*

> *While your lecture may not have learned about the t-test yet, a CI is just the reverse of a one sample t-test. To get Confidence Intervals in R see the following page.*

R notes:

To get 95% Confidence intervals in R/Rstudio:

```
t.test(height)
```

or

```
t.test(dataset$height)
```

Depending on how you've arranged your data.

R defaults to a 95% CI, ***but if you want a 90%CI***, you need to modify the above:

```
t.test(height,conf.level = .90)
```

or

```
t.test(dataset$height, conf.level = .90)
```

The confidence interval will be near the bottom of the printout from R.

Be prepared to discuss these problems in recitation the week of March 17th. As usual, problems not discussed in recitation are due at the end of class.