**Lecture/text homework assignment # 9**

***Note: Please circle your answers when appropriate!***

**1)** You (very carefully!) measure the length (in m) of 10 male and 13 female saltwater crocodiles (*Crocodylus porosus*) and you get the following results. Is there a difference in length? Use a Mann-Whitney U test with $\alpha = 0.05$ (***do NOT use R***):

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Males: | 4.2 | 4.4 | 4.7 | 4.9 | 5.0 | 5.1 | 5.1 | 5.3 | 5.7 | 5.9 | | | |
| Females: | 3.2 | 3.2 | 3.3 | 3.4 | 3.5 | 3.7 | 3.7 | 3.8 | 3.8 | 3.9 | 4.0 | 4.7 | 4.9 |

**2)** Now read the following data into R. ***Use R*** for all statistical procedures.

| Sample A | Sample B |
|---|---|
| 4.0353 | 6.8552 |
| 1.7120 | 51.8669 |
| 0.3804 | 3.0567 |
| 0.4975 | 6.3868 |
| 7.3198 | 6.5228 |
| 0.7533 | 6.2206 |
| 84.7224 | 6.6819 |
| 4.5062 | 5.6654 |
| 4.1913 | 9.3792 |
| 0.8560 | 5.6765 |

(a) Make *q*-*q* plots of both samples.

(b) Perform a *t*-test (Welch's *t*-test). Use $\alpha = 0.05$.

(c) Perform a MWU test. Use $\alpha = 0.05$.

(d) Which test gave you the better *p*-value now? Why?

(e) What is a "better" *p*-value? Why?

**3)** You measure the length of 5 radish seedlings at 7 days and 10 days and get the following results in mm (**do NOT use R** except for part (d)). Use $\alpha = .05$ for all tests below:

| seedling #: | 1 | 2 | 3 | 4 | 5 | $\bar{y}$ | $s$ |
|---|---|---|---|---|---|---|---|
| 5 days: | 30 | 29 | 34 | 47 | 37 | 35.4 | 7.232 |
| 7 days: | 38 | 32 | 35 | 53 | 45 | 40.6 | 8.444 |
| difference: | -8 | -3 | -1 | -6 | -8 | -5.2 | 3.114 |

(a) Is there a difference in length?

(b) Repeat (a), but now use a regular (unpaired) *t*-test.

(c) What happened in (b)?

(d) Verify the normal distribution assumption for the paired *t*-test you did in (a). (*You may use R for this part*).

**4)** Here are data on the weights (in grams) of male and female Chinstrap penguins (from the data set we used in hw2). Your job is to find out if there is a difference between male and female penguins.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| females | 3675 | 3675 | 3400 | 3700 | 3700 | 3850 | 2900 | 3650 | 3800 | 4150 | 3300 | 3525 |
| males | 3700 | 3600 | 4800 | 4550 | 3900 | 3450 | 4150 | 4300 | 3800 | 3950 | 4100 | 4050 |

It's up to you to figure out what the best procedure is, what kind of hypotheses to use, which $\alpha$ to use, what test to use, and so on. Make sure you follow all the appropriate steps. ***You may use R (and probably should)*** *as you'll get done* ***much*** *quicker. Remember to very clearly state your results in writing. Never turn in just an R printout.*

*Hint:* how do you decide which test to use? What kind of distributions do the data have?

**5)** You want to determine if there is an effect of strength training. You take 9 volunteers and measure their ability to lift a certain weight before and after strength training. You get the following data:

| Volunteer | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| before strength training | 101 | 108 | 90 | 113 | 126 | 132 | 108 | 84 | 116 |
| after strength training | 114 | 123 | 103 | 126 | 142 | 144 | 120 | 97 | 130 |

Perform a complete hypothesis test to see if there is a difference in strength before and after weight training.

*Again - notice that you are given no advice. You need to figure this out on your own, although (hint!) this one should be a bit easier than* ***(4)****.* ***Feel free to use R to solve this problem****, just make sure you don't just hand in a printout.*

**Be prepared to discuss these problems in recitation Wednesday, July 1st.**

**Remember: Copying answers from other students, web sites, previous classes or using AI is a violation of the honor code!!**

# Notes on using R.

## *1) t-tests:*

Enter your data. If you don't remember how to enter your data, check a previous homework or look at the R-notes posted on the 214 web page. Also remember that depending on how you've read your data into R, you may have to add “dataset$” in front of the examples below.

The easiest way to do a *t*-test is as follows:

```
t.test(y,x)
```

where *y* is your first variable and *x* is your second variable. R defaults to Welch's *t*-test.

If, for some reason, you want to do an equal variance (classic) *t*-test, you can do:

```
t.test(y,x, var.equal = TRUE)
```

To do a paired *t*-test (make sure your data in y and x are paired appropriately:

```
t.test(y,x, paired = TRUE)
```

*************

Next week you'll be doing some one-sided tests (you can ignore this section until next week):

To do a one sided test, you do:

t.test(y,x, alternative = “less”)     (obviously, you can also use “greater”)

This option works with all of the above examples.

*************

## *2) q-q plots*

Just give R the name of your variable you want to check. For example, if you're doing a two sample test comparing men vs. women, you would do:

```
qqnorm(men)
```

and

```
qqnorm(women)
```

(Aassuming your heights for men are in a variable named “men” and your heights for women are in a variable named “women”)

If you want to add a line to your qq plots, just repeat the above instructions exactly but change qqnorm to qqline. You need to do qqnorm first, and then qqline, for example:

```
qqnorm(men)
qqline(men)
```

Remember that for a paired *t*-test you need to do one plot of the differences. Assuming your data are correctly arranged (paired) in two columns you can do:

```
diff <- varname1 - varname2
qqnorm(diff)
qqline(diff)
```

### *3) Mann-Whitney U tests:*

This is very similar to the t.test syntax above:

```
wilcox.test(y,x)
```

Where $y$ is your response or measured variable, and $x$ is your identifying variable.

If R prints a warning message about not being able to print exact $p$-values, you can safely ignore it.

If you want to do a one sided test, the syntax is exactly the same as for the t-test:

```
wilcox.test(y,x, alternative = "less")
```
(or "greater", of course)

*A comment about your test statistic ($U^*$):*

For the Mann-Whitney U-test, R will not always give you the larger of the two values ($K_1$ or $K_2$). It'll always give you $K_1$.

If you want to get the actual $U^*$ (remember $U^* = \max(K_1,K_2)$), it's not difficult. You just need to remember that $n_1 \times n_2 = K_1 + K_2$. So here's a summary:

1) Calculate $n_1 \times n_2$
2) Subtract "$W$" (what R gives you as a test statistic)
3) pick the larger of $W$ or what you got in # 2 - that's $U^*$

Okay, so it's not as easy as having $U^*$ printed out, but other programs (e.g., Minitab) don't even give you $U^*$ (it's not at all clear what Mintab gives you!).