Assignment 5

1. Researchers studying the number of electric fish species living in various parts of the Amazon basin were interested in whether the presence of tributaries affected the local number of electric fish species in the main rivers (Fernandes et al. 2004). They counted the number of electric fish species above and below the entrance point of a major tributary at 12 different river locations. Here's what they found:

Tributary	**Upstream number of species**	**Downstream number of species**
Içá	14	19
Jutaí	11	18
Japurá	8	8
Coari	5	7
Purus	10	16
Manacapuru	5	6
Negro	23	24
Madeira	29	30
Trombetas	19	16
Tapajós	16	20
Xingu	25	21
Tocantins	10	12

a. What is the mean difference in the number of species between areas upstream and downstream of a tributary? What is the 95% confidence interval of this mean difference?
b. Test the hypothesis that the tributaries have no effect on the number of species of electric fish.
c. State the assumptions that you had to make to complete parts (a) and (b).

R exercise
d. Load the data *ElectricFish.csv* into R. Use the function *t.test*() and if you want to do a paired t-test set the command *paired* to TRUE; *t.test*(data.1, data.2, paired = TRUE). Note that each row has to be a paired set of data.
e. Interpret the test output and compare it with the output of your manual calculation. Are they the same? Do you understand what means what?

2. *Spot the flaw*. There are two types of males in bluegill sunfish. *Parental* males guard small territories, where they mate with females and take care of the eggs and young. *Cuckolder* males do not have territories or take care of young. Instead, they sneak in and release sperm when a parental male is spawning with a female, thereby fertilizing a portion of the eggs. A choice experiment was carried out on juvenile sunfish to test whether offspring from the two types of eggs (fertilized by parental male vs. fertilized by cuckolder male) are able to distinguish kin (siblings) from non-kin using odor cues. The researchers used a two-sample method to test the null hypothesis that fish are unable to discriminate between kin and non-kin. This null hypothesis was not rejected for offspring from parental males. However, the same null hypothesis was rejected for offspring from cuckolder males. The researchers concluded that offspring of cuckolder males are more able to discriminate kin from non-kin than are offspring of parental males. What is wrong with this conclusion? What analysis should have been conducted?

3. Alcohol consumption is influenced by price and packaging, but what about glassware? Attwood et al. (2012) measured whether the time taken to drink a beer was influenced by the shape of the glass in which it was served. Participants were given 12 oz. (about 350 ml) of chilled lager and were told that they should drink it at their own pace while watching a nature documentary. The participants were randomly assigned to receive their beer in either a straight-sided glass or a curved, fluted glass. The data below are the total time in minutes to drink the glass of beer by the 19 women participants in the study.

Straight glass: 11.63, 10.37, 17.89, 6.96, 20.4, 20.64, 9.26, 18.11, 10.33, 23.54.

Curved glass: 7.46, 9.28, 8.9, 6.73, 8.25, 6.16, 13.09, 2.1, 6.37.

a. Show the data in a graph. What trend is suggested? Comment on other differences between the frequency distributions of the two samples.
b. Test whether the mean total time to drink the beer differs depending on beer glass shape.
c. How much difference does it make? Provide a confidence interval of the difference.
d. A second test of the same hypotheses but using the data from male participants yielded the following results:

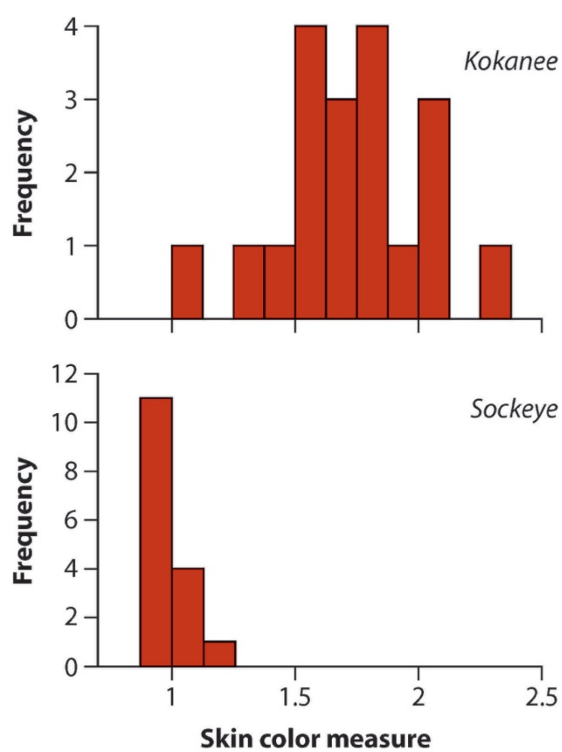
Straight: $\bar{Y}_1$ = 7.987, s_1 = 2.459,
Curved: $\bar{Y}_2$ = 6.930, s_2 = 3.748,
$\bar{Y}_1$ - $\bar{Y}_2$ = 1.057, s_p^2 = 10.048, $SE_{\bar{Y}_1 - \bar{Y}_2}$ = 1.418, t = 0.746, df = 18, P = 0.466.

Is the following conclusion from the tests valid: "There is a significantly greater effect of beer glass shape on mean time to drink in women than in men"? Explain.

R exercise
e. Load the data *BeerGlassShape.csv* into R. Make a boxplot in R with one for each category of graph. Add your own y-axis label. Getting two boxplot side by side needs a little tweak, as R like to know what the response variable (beer$drinkingMinutes) and the explanatory variable (beer$glassShape) are. These are separated by a ' ~ '; this is a general syntax in R which we will see soon for defining statistical models.
f. Perform a two-sample t-test on the data set above. t.test(response var ~ explanatory var). If you want to assume equal variances you have to set it using var.equal = TRUE. Does this test show the same results / conclusions as d.?

4. Sockeye salmon swim sometimes hundreds of miles from the Pacific Ocean, where they grow up, to rivers for spawning. Kokanee are a type of freshwater sockeye that spend their entire lives in lakes before swimming upriver to mate. In both types of fish, the males are bright red during mating. This red coloration is caused by carotenoid pigments, which the fish cannot synthesize but get from their food. The ocean environment is much richer in carotenoids than the lake environment, which raises the question: how do kokanee males become as red as the sockeye? One hypothesis is that the kokanee are much more efficient at using available carotenoids than the sockeye. This hypothesis was tested by an experiment in which both sockeye and kokanee individuals were raised in the lab with low levels of carotenoids in their diets (Craig and Foote 2001). Their skin color was measured electronically (as a* units on a L*a*b* standard). The data are as follows and are plotted in the accompanying histograms:

Kokanee: 1.11, 1.34, 1.55, 1.53, 1.50, 1.71, 1.87, 1.86, 1.82, 2.01, 1.95, 2.01, 1.66, 1.49, 1.59, 1.69, 1.80, 2.00, 2.30
Sockeye: 0.98, 0.88, 0.97, 0.99, 1.02, 1.03, 0.99, 0.97, 0.98, 1.03, 1.08, 1.15, 0.90, 0.95, 0.94, 0.99



a. List three methods that would be appropriate to test whether there was a difference in skin color between the two groups.
b. Use a transformation to test the difference in mean between these two groups. Is there a difference in the mean of kokanee and sockeye skin color?

R exercise

c. Read the data *SalmonColor.csv* into R. Make sure the x axis is the same in both histograms. See hint[4] if you need some help.

d. Check the variances of both groups. *aggregate*() is great function to do chat. Use ?*aggregate* to check how it works and calculate the standard deviations for both species.

Check the standard deviations again after a log transformation. Which test would you do for unequal variances and which for equal variances? Do both and compare the results.

5. The vuvuzela captured international attention during the 2010 World Cup in South Africa. In its modern incarnation, the vuvuzela is a plastic horn about 65 cm long, which can produce a sound loud enough to cause permanent hearing damage. Blowing a vuvuzela requires a fair amount of air pressure, and Lai *et al.* (2011) were concerned that vuvuzela use by anyone carrying a pathogen would cause air-borne contagions to be spread broadly through in a crowd. They tested this idea with an experiment that compared the concentration of aerosol droplets or particles produced by people blowing vuvuzelas to that produced by the same people shouting instead. The data, measured as thousands of particles per liter, for 8 individuals are in the table below.

	Particle concentration (1000s/l)	
Person	**Vuvuzela**	**Shouting**
1	606	6.1
2	1077	6.4
3	220	1.3
4	396	1.8
5	1197	6.0
6	178	1.5
7	645	2.9
8	944	2.9

a. Take the log transformation of each value before finding differences. Then calculate a 95% confidence interval for the mean difference in log particle concentration between vuvuzelas and shouting.

b. Carry out an appropriate test for a mean difference in log particle concentration between the two forms of cheering.

R exercise

c.. Load the data *Vuvuzela.csv* into R. Log transform the two columns and take the difference between for each person. Calculate the mean, sd and se.

d. do a one sample t-test to test if the mean difference between the two variables is significantly different from 0. Standard *t.test*() takes 0 as the reference but you can change this through e.g. *t.test*(, mu = 4) to it against 4 as your reference.

e. Use the confidence interval around the mean difference to see what range the 95% C.I. covers. Use the mu = .. option to see if your intuition of what reference mean should still be significantly different holds. Hint[5].

Hints:

[4] Provide the break command in the *hist*() with a vector of cut of points. Start from 0 to a little over the maximum value of time beer drinking the *max*(beer$drinkingMinutes). The function *seq*() is very handy to make this, it takes the first value, last and the difference between values. *seq*(0,10, 2) gives the vector *c*(0, 2, 4, .. , 10)

[5] If the reference does not fall within the 95% C.I. it should by significant. This C.I. approximation works well, but is not not ‘exact’.