Biometry, Fall 2024 - Sacha Medjo-Akono

Midterm 1
Practical Exam (50 pts)

Answer the following questions in complete sentences, supporting your answers with appropriate statistics as requested.

Use R to conduct your analyses and make graphs. All data sets are provided in the Midterm 1 Practical Exam module on Canvas. Be sure to check that your data meet the assumptions of any test you use, and transform data as necessary. However, you do not need to present evidence that assumptions are met unless a question specifically asks you to do so.

When you're finished, use the Midterm 1 Practical Exam "quiz" in the module to upload a pdf copy of this document with your written answers and any figures you've created, as well as a text file containing R code used to answer each question.

1. (5 pts) Sasha is testing the effect of a new drug on cancer cell growth and metabolism. She grows cancer cells in petri dishes in LB media and measures the growth rate of cells in each dish to get a baseline measure. Then she takes a sample of cells from each dish and initiates a new culture, but this time in petri dishes with LB media infused with the new drug. She again measures the growth rate. The data are in the file *cancer.csv*. Does the drug have an effect on cancer cell growth? Provide statistics that support your answer.

Paired t-test stats:
t=1.5584, df= 19, p=0.1356, mean diff = 1.95

Since the p>0.05 and the mean difference is very low we can conclude that the new drug does not have a significant effect on cancer cell growth.

Note:
I think she should have grown the cancer cells straight onto the medicine-infused dishes. There is a chance that the 1st growth may have been affected by the first LB plates, influencing an error of independence in the final data collected.

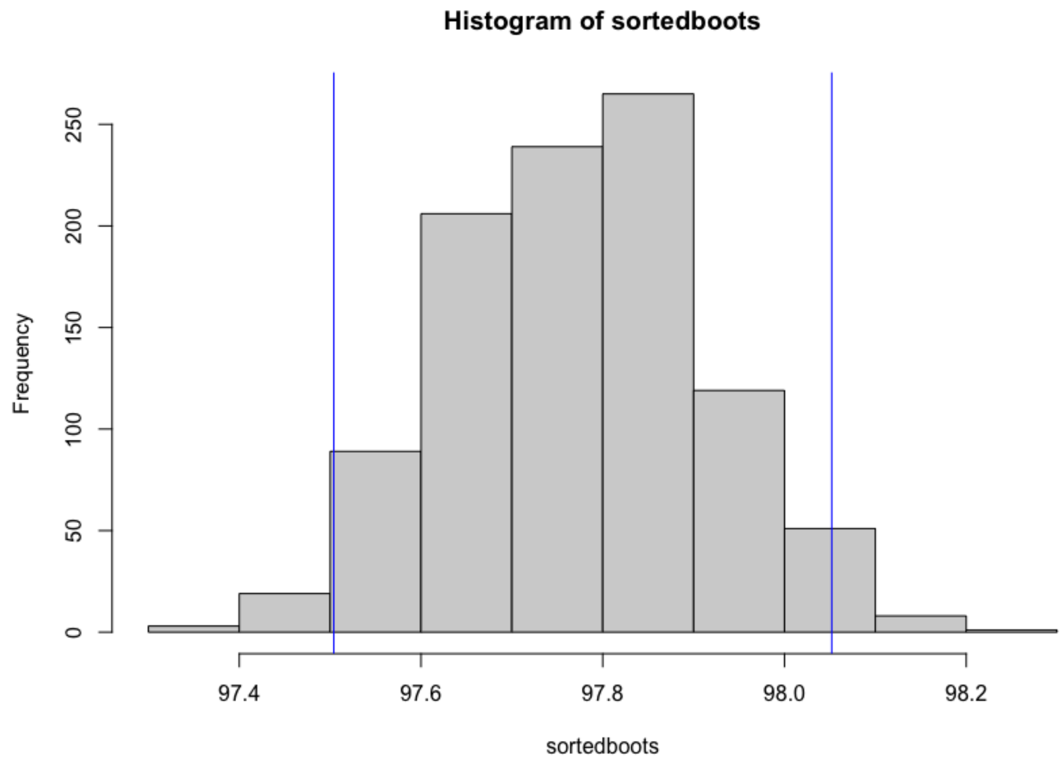
2. Usually we think of the average human body temperature as 98.6°F. This information comes from a doctor in Germany who measured the temperature of local villagers in 1851. Myra reads about this and wonders if the same is true for people outside of Germany. There is also some evidence that as humans have become more sedentary after the Industrial Revolution, the average human temperature has decreased. Myra sets out to test whether 98.6°F represents the average temperature of students at CSUN. She takes the temperature of 50 students haphazardly chosen on the CSUN campus. Her data are in the file *temps*.csv. Note: haphazard is not really random.

(a) (5 pts) Calculate the following descriptive statistics:

a. Mean = 97.774
b. Median = 97.8
c. Standard Deviation = 0.95443
d. Standard Error of the Mean = 0.13498
e. Coefficient of Variation = 0.009761626 or 0.98 %

(b) (5 pts) Use bootstrapping (generate 1000 bootstrapped samples) to generate 95% confidence intervals for the data sample above. Make a histogram of your bootstrapped samples and plot the confidence limits on the graph.

CI 95% +/-= 97.504 and 98.052



(c) (5 pts) Conduct a t-test to determine if the temperature of the CSUN population is significantly different than 98.6°F? Support your answer with appropriate statistics.

The t-value= -6.1195 and the p-value= 1.531e-07 suggest that there is a statistically significant and negative difference between the mean of today's body temp (97.77) and previous (98.6). Temperature has indeed decreased according to this experiment.

3. Kathryn, a student in Dr. Steele's lab, wants to test the hypothesis that **shading of the invasive alga *Sargassum horneri* by the native giant kelp (*Macrocystis pyrifera*) affects the biomass of *S. horneri*.** She compared the biomass (g/m^2) of *S. horneri* in 1-m^2 plots that are either under giant kelp canopy (shaded) or not shaded by giant kelp. Each individual was located at least 5-m away from other samples, **so each individual is an independent replicate.** Her data are in the file "*sargassum.csv*".

(a) (5 pts) Test the H_0 that there is no difference in biomass of the invasive alga between shaded and un-shaded plots using an appropriate test. First, write a short statement for the Methods section of this paper explaining how these data were analyzed. Second, write a short statement for the Results section of this paper (supported by statistics) about whether shading affects *Sargassum* size.

To test H_0 (null hypothesis) that there is no difference in biomass between shaded and unshaded Sargassum, I ran a two-sample t-test. This two-sample t-test is also known as Welch's test because we have two separate sample groups with differing variances and means. I interpreted whether we accept or reject the null hypothesis by noting the t-value, p-value, degrees of freedom, and 95% confidence intervals.

My results indicate that we reject the null hypothesis and conclude that shading significantly affects Sargassum size. The mean of Sargassum biomass is much higher when it is unshaded. The t-value of -5.1308, being far from 0, shows that the difference between the means of each group is large. The confidence intervals are both negative and do not include 0 supporting this difference. Finally, the p-value is less that 0.05 which confirms that the test results are statistically significant

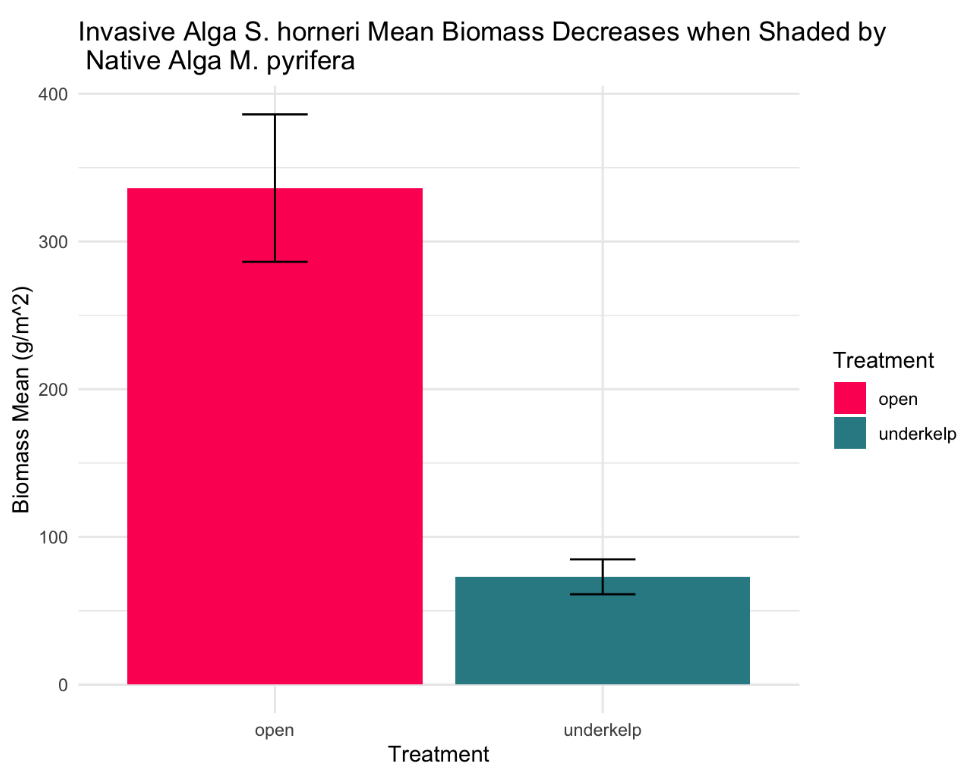
t-value = -5.1308

the p-value = 1.332e-05

df = 32.24

+/- 95% CI = -367.5339 and -158.6928

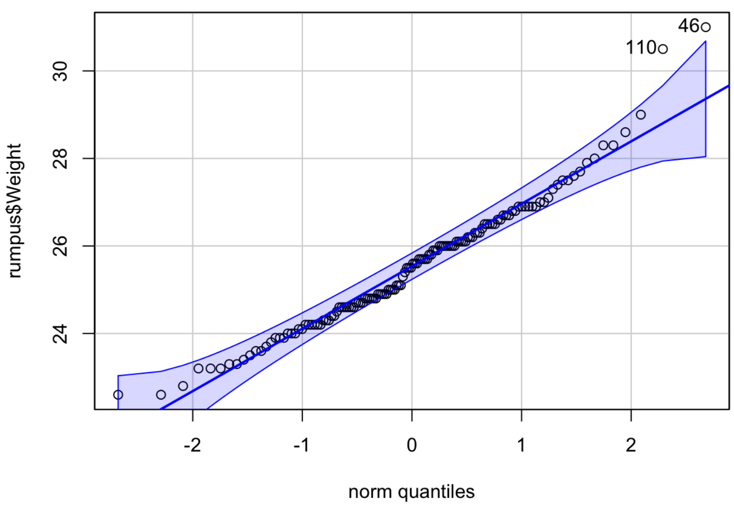
(b) (5 pts) Make a bar graph, with means +/- SEM, to illustrate these results. Include a figure caption.



4. One cold February day in 1898, an "an uncommonly severe" storm passed over New England. After the storm, the zoologist Hermon Bumpus collected 136 house sparrows that had been brought down by the storm in the vicinity of his laboratory at Brown University in. More than half of the birds recovered, but the rest died from exposure. Bumpus took this as an opportunity to study natural selection in action, and measured a number of skeletal features on all the birds, as well as recording whether they survived the storm, their sex, and (in the males) whether they were adults or yearlings. Bumpus's data are in the file "bumpus.csv."

(a) (5 pts) Confirm that the sparrows' weight in grams (column "Weight") and length in millimeters (column "Length") are more or less normally distributed, and provide your conclusion with supporting statistics and/or figures.

Figure 1. Data for Weight is normal enough except outliers 110 (30.5g) and 46 (31 g)



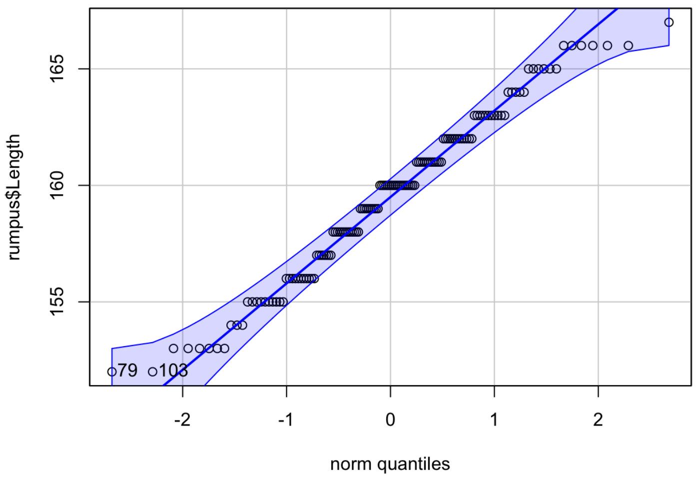


Figure 2: Data for Length is normally distributed

(b) (10 pts) Use an appropriate test, transforming the data if necessary, to determine whether male and female birds (column “Sex”) differ in average length. Answer with your conclusion, provide test statistics to support it, and illustrate it with a publication-quality figure of your choice.

The average male Sparrow has a larger body length than female Sparrow. I chose to assume equal variances after running a Bartlett test on the Length in relation to sex.

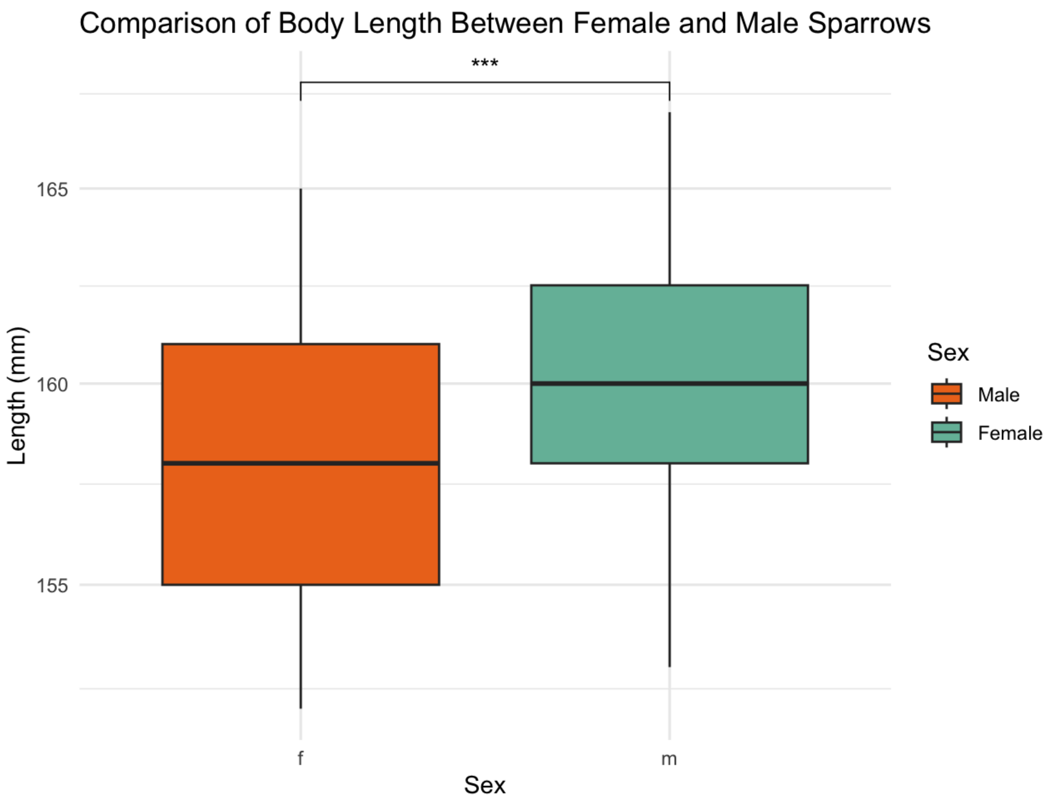
By running equal variance, two sample t-tests on both the raw data and log10 data collected by Bumpus, I was able to observe large negative t-values, p-values<0.5 and 95% confidence intervals excluding 0, confirming that the difference between their average lengths was indeed significant. Asterixes on the boxplot represent this significance.

Raw data t-test values:

t=-4.0599, p-value=8.304e-05, Female Length mean = 157.9796, Male Length mean= 160.4253, +/- 95% CI= -3.637142 and -1.254249.

Logged data t-test values:

t=-4.0825, p-value=7.617ee-05, Female Length mean = 5.062204, Male Length mean= 5.077631, +/- 95% CI= -0.022900360 and -0.007953025.



(c) (5 pts) Use an appropriate test to determine whether sparrows' weight (across both males and females) is correlated with their length. Briefly explain your choice of test, and give your conclusion with test statistics to support it.

I chose Spearman's rho because we have a large sample size of non-parametric data (i.e. weight). While I got a ranks warning for this test, I was able to verify it's functionality by using a Pearson test on ranked data.

I concluded that there is a statistically significant correlation between weight and length in sparrows since we have a positive rho close to +1 and a p-value much smaller than 0.05.

The rho= 0.5575 and p-value=1.791e-12.

(d) (10 pts) Assuming larger sparrows weigh more (that is, body size has a causal relationship with weight), use a linear regression to estimate the proportion of variation in weight that is explained by length, transforming the data if necessary, and whether this is greater than expected due to chance. Answer with your conclusions, giving statistics to support them, and illustrate them with a publication-quality scatterplot and a regression line.

Conclusion:

I concluded that while sparrow weight moderately increases with body length (or that length (x) is a predictor of weight (y)), it is not a constant relationship.

Reasoning:

A Model II linear regression predicted a significant and positive relationship between weight (y variable) and length (x independent variable).

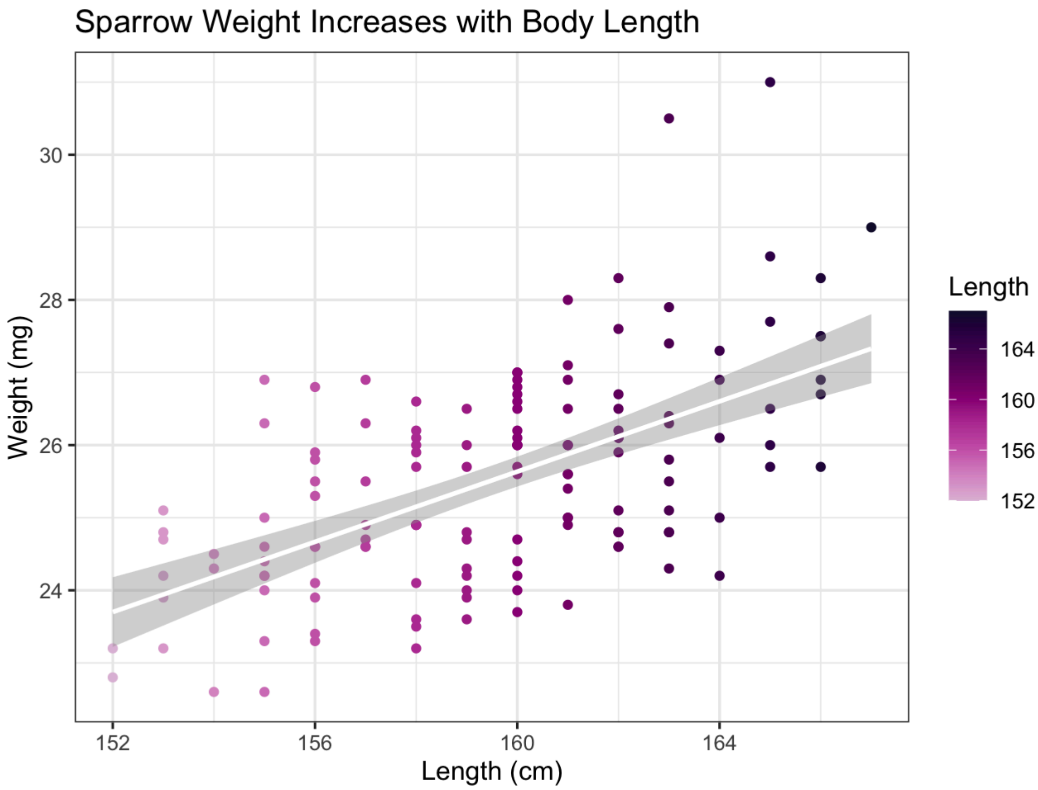
The correlation coefficient r shows us a positive correlation between x and y, while r^2 tells us that ~34% of the weights are predicted by body length.

Parametric p-values for the 2-tailed and 1-tailed tests are both <0.5 showing a statistically significant relationship between weight and length, thus confirming that correlation.

Regression and CI p-values for OLS show negative intercepts with small positive slopes showing that Weight increases with Length but it's not a 1:1 ratio.

Regression and CI p-values for RMA show a similar relationship but the slops are much close to 1, indicating a stronger sensitivity of y to changes in x.

```
Model II regression

Call: lmodel2(formula = Weight ~ Length, data = rumpus, range.y = "relative", range.x =
"relative", nperm = 99)

n = 136   r = 0.5838648   r-square = 0.3408981
Parametric P-values:   2-tailed = 8.612224e-14    1-tailed = 4.306112e-14
Angle between the two OLS regression lines = 21.76011 degrees

Permutation tests of OLS, MA, RMA slopes: 1-tailed, tail corresponding to sign
A permutation test of r is equivalent to a permutation test of the OLS slope
P-perm for SMA = NA because the SMA slope cannot be tested

Regression results
  Method Intercept     Slope Angle (degrees) P-perm (1-tailed)
1    OLS -13.06697 0.2418890*       13.59803              0.01
2     MA -17.65121 0.2706224        15.14281              0.01
3    SMA -40.57245 0.4142895        22.50371                NA
4    RMA -73.78601 0.6224674*       31.90092              0.01

Confidence intervals
  Method 2.5%-Intercept 97.5%-Intercept 2.5%-Slope 97.5%-Slope
1    OLS      -22.23770       -3.896242  0.1844224   0.2993557*
2     MA      -28.09688       -7.539422  0.2072431   0.3360944
3    SMA      -50.37377      -32.036839  0.3607895   0.4757228
4    RMA     -104.04568      -54.384097  0.5008589   0.8121307*

Eigenvalues: 13.50953 1.34625

H statistic used for computing C.I. of MA: 0.00358869
```


Sparrow Weight Increases with Body Length
Weight (mg)
Length (cm)
Length
30
28
26
24
152
156
160
164
164
160
156
152