**Please enter your answers to this exam into ecampus by Monday May 4th at 5PM. This test is open note/internet. Do not get assistance from other students or PIs. If you have questions come to office hours (9-10 Friday or Monday)**

Questions 1-10 are worth 2.5 points each
Questions 11-15 are worth 15 points each

I would recommend that you first work out each problem before you go onto ecampus to take the test. You will be allowed only one attempt at taking the test. If any of the questions are unclear to you please come to office hours and allow me to clarify it for you.

1. What is a p-value?

2. Why do we need to correct for multiple comparisons?

3. Choose one approach to dealing with the problem of using data that come from different species.

4. An advantage of Bayesian approaches is that they _____.

5. PCA is a method that is helpful in cases _____.

6. The starting portion of an MCMC will have low likelihood _____.

7. Why don't we sample from the first portion of an MCMC run?

8. The Beta coefficient in a linear model with a continuous predictor variable describes _____.

9. The histogram below is of residuals for a linear model what potential problem do you see in them?
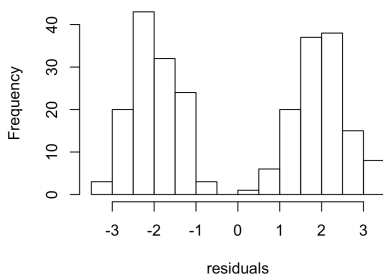
10. Often in an experiment we will have a group that receives a sham treatment. For instance, if infecting an organism with a bacteria via injection we might have a second group that receives an injection of sterile saline. What is this called and what is its purpose?

These questions do not use the data from the review ensure that you are using the files specified for the final exam.

11. Download the test.frogs.csv morphology dataset from the course website this data includes measures for 3 species in the genus Dendrobates for 10 different traits as well as 10 unidentified specimens. To what species do these 10 unidentified samples belong (there could be unidentified samples that belong to different species).

12. Download the the mcmc log file test.log.csv file from the course website. Select the most appropriate estimate of the mean for the parameter "asc1".

13. Download the test.dispersal.csv file from the course website this contains the proportion of individuals that attempt to disperse from 3 populations of beetles. Looking at the recorded variables what do you believe the most important predictor of dispersal is.

14. Use a full model with all possible predictor variables and select the answer that most closely matches the expected proportion of dispersing beetles for Marksville strain males at 27 degrees in old media.

15. Download the test.contact.maps.csv file from the course website this data is a simplified version of what we get when we do Hi-C sequencing of genomes. It tells us about the spatial arrangement of chromosomes in the nucleus. The numbers in the table tell us how often two chromosomes were in close contact in the cells that were sequenced. Our expectation (null hypothesis) is that each chromosome should have lots of contact with itself and relatively few with all others. The size of a chromosome should predict the number of contacts it has with itself. In contrast the number of contacts between any two chromosomes should be a function of their sizes. However, some preliminary data suggests that micro chromosomes (very small chromosomes present in birds and some other reptiles) may be tightly associated in the nucleus because they contain genes that are in the same genetic pathways and can be co-regulated. In the table normal chromosomes (macro-chromosomes) are labeled with the prefix "Ma" while micro-chromosomes are label with the prefix "Mi". Using a Monte Carlo approach does this data support a close association between any of the micro chromosomes in this dataset?