BASIC STATISTICS FOR BIOLOGISTS – EXAM INFORMATION

**Questions** on the exam should be **in English** (I have asked for this and handed in English-language questions but not received a guarantee that they won’t be back-translated to German) and **should be answered in the language they are posed in**!

## **Seminar Breakdown:**

Seminar 1 – Introduction to Statistics

* Be familiar with terminology of:
  + Population/Sample
  + Training Data/ Test Data
  + Randomisation (what makes data random and how to achieve that in R)

##### Seminar 2 – Introduction to R

* Assigning and Removing Objects
* Load Data on a certain name (provide the code)
* Object Types (list object types and modes)
* Object Dimensions (know how str, class, summary, dim, and view work and when to use which)
* Basic commands of R (getwd, setwd, rm,?, etc.)

##### Seminar 3 – A Primer for Statistical Tests

* Variable scales and what they imply
* Tests for normality of the data

Seminar 4 – Descriptive Statistics

* Differences between descriptive and inferential statistics and when you would use which
* Descriptive parameters (no calculation formulas):
  + What do they describe?
  + How do they react to outliers?
  + How do you obtain them in R?

Seminar 5 – Data visualization

* Not on the exam (but important for your research)

Seminar 6 – Inferential Statistics

* Be able to establish Null and Alternative Hypotheses
* How to Go About Testing Hypotheses
* Choosing The Appropriate Test (table with statistical tests)  
   On a personal note, I detest having to have this on the exam and would not ask this of you if I had the freedom to take the possibility of this being off the exam out back

##### Seminar 7 – Data Handling

* README file (what is it for and what does it contain)
* Common Issues (be able to spot these in a data set):
  + **NAs in Excelsheet?**
  + **Dots or Commas?**
  + **Redundancy and Sparsity?**

##### Seminars 8-12 – Statistical Tests

* For each test we **covered**:
  + **Know what it does**
  + **Its R notation**
  + **The assumptions that come with it**

## **Example Questions:**

1) what does setwd() do?  
 setting working directory

2) How do you identify the working directory?  
 getwd()

3) What is a binary scale and what does it imply?  
 a special case of a nominal scale that it implies that the data values fall onto 2 states

4) Your data: flight length and avian species membership. You want to run a Spearman-correlation. How do you do this?  
 it is not an option, because both the data need to be continuous, but species is categorical!

5) You want to read data from an excel file from an csv-format, how do you do that?  
 read.csv/read.table

6) Subtract 1 from every element in vector a <- c(1,2,3,4,5,6)  
 a–1

7) Select the third element of the vector a and add 5  
 a[3] + 5

8) How do I convert a numeric vector into a character vector  
 as.character()

9) You want to run an anaylsis to answer the question whether toad length influences toad eye span. A shapiro test of the two data sets (length and eye span) has concluded with p-values 0.012 and 0.004, respectively. Which test do you use and why?  
 Pearson correlation as both variable values are metric and normal distributed.