# Ecological Data Management and Analysis

*Written Exam*

*2 February 2015*

**Name: _____ Surname: _____ Matricola N° _____**

*Instructions*

The written part of the exam consists of multiple choice and open questions as well as an exercise to be done within the statistical environment "R". The written exam will represent 40 % of the final mark (calculated in 30/30). You can therefore get a maximum score of 40 points in this exam, in which each correct multiple choice reply gives you 2 points, Open questions can get a maximum of 5 points, while the correct completition of the R exercise has a value of 10 points.

1. **You have a variable, represented by 5 data points: 23.4, 46.7, 56.4, 45.0, 35.9. What definition best describes this data?**

◯ the data is continous

◯ the data is discrete

◯ the data is binomial

◯ the data is count data

2. **If predation by fox (F), puma (P) or stray dogs (D) are mutually exclusive events, what is the probability to be predated by a fox or by a dog?**

◯$P(F) * P(D)$

◯$P(F) + P(D)$

◯$P(F) + P(D) - P(F \cap D)$

◯$\frac{P(F \cap D)}{P(P)}$

3. **Which of the following is the formula of the Variance?**

◯$\sqrt{\sigma^2}$

◯$\frac{\sum (x-\overline{x})^2}{n-1}$

◯$\frac{\sigma}{\sqrt{n}}$

◯$\frac{\sum x}{n}$

4. **What is the Mode?**

◯ The mode is the number which appear less often in a sample

◯ The mode is the mean of the numbers in a sample

◯ The mode is the second quantile of a sample

◯ The mode is the number which appears most often in a sample

5. **What is the p-value?**

◯ The *p-value* is the probability of obtaining the sample results if the Null hyphotesis is false

◯ The *p-value* is the probability of obtaining the sample results if the Null hyphotesis is true

◯ The *p-value* is the probability of obtaining the sample results if the Alternative hyphotesis is false

◯ The *p-value* is the probability of obtaining the sample results if the Alternative hyphotesis is true

6. **What is the r-squared coefficient?**

◯ The R-squared coefficient represents the percentage of the variation in the response variable explained by the linear model

◯ The R-squared coefficient represents the strength of the linear association between two variables

◯ The R-squared coefficient represents the value of the slope in a linear regression

◯ The R-squared coefficient represents the value of the intercept in a linear regression

7. **What kind of regression model would be appropriate to analyse a response variable with values which can be or 0 or 1 (e.g. presence/absence or survival data)**

◯ An ordinary least squares regression model

◯ A generalized linear model with a Poisson distribution of the errors

◯ A logistic regression model

◯ A Chi-square test

8. **Which of the following sentences is false?**

◯ The Akaike Information Criterion (AIC) is a method to select the relatively best fitting model in a set of models.

◯ The model with the highest value of AIC is the best fitting model

◯ Models within $\delta$ AIC $< 2$ are considered equivalent

◯ AICc corrects for small sample sizes. But with large sample size, AICc converges to AIC

9. **In which of the following situations is it appropriate to use a linear mixed model?**

◯ In the case of data following a Poisson distribution

◯ In the case of data following a mixed distribution

◯ In the case of data in which there are repeated measurements of the same individuals

◯ In none of the above situations

10. **Which of the following definitions is true?**

◯ The 95% *confidence interval* is the interval of values between which has a 95% probability to include the true estimate

◯ The 95% *confidence interval* is the interval of values between which 95% of repeated samples from a population include the true estimate

◯ The 95% *confidence interval* is a measure of dispersion representing the interval which includes 95% of all values of the sample

◯ None of the above