



NTNU – Trondheim
Norwegian University of
Science and Technology

Department of Biology

Examination paper for BI 3051 Evolutionary Analysis

Academic contact during examination: C. Pelabon

Phone: 90824159

Examination date: 30/11

Examination time (from-to): 09:00 / 13:00

Permitted examination support material: authorized calculator

Other information:

Language: English

Number of pages (front page excluded): 3

Number of pages enclosed:

Informasjon om trykking av eksamensoppgave

Originalen er:

1-sidig 2-sidig

sort/hvit farger

Checked by:

Date

Signature

Although several questions are presented as biological problems the purpose here is to test your understanding of the link between experimental design and statistical analyses. Therefore, understanding the biological question is interesting but should not be a prerequisite to answer the question and it is not necessary here to comment lengthily on the biology. Overall, try to write concise and precise. The weight of the question is reported in front of each question as the percentage of the total grade.

(20%) Question 1 – Variance, standard deviation (SD), standard error (SE), coefficient of variation (CV) and confidence interval (CI) describe different aspects of the dispersion of a variable. Define precisely these different statistics both in mathematical and “biological” terms, and explain how they are related to each other.

(15%) Question 2 – Explain what are type I and type II errors of a statistical test and what factors will affect the probability of making type II error.

(20%) Question 3 – In a recent paper, Larios & Venable (2015) present an experiment where they test whether or not plants adjust the provisioning to their seeds depending on the level of competition encountered. To do that they studied the annual plant *Dithyrea californica* from two desert populations one in Mexico and one in California. In these two populations they marked seedlings at the cotyledon stage (just after emergence from the ground) and they counted the number of neighbors less than 10 cm away from the base of each seedling. They followed individuals through the whole growing season, and collected 88 plants with seeds still attached at the time of seed set from both populations. Aboveground biomass was measured to the nearest 0.001 g using an analytical balance after oven drying at 60°C. They also estimated the average offspring seed diameter by measuring 30 to 50 seeds per plants using digital calipers and taking the mean of these measurements. The authors further present the following figure to describe the hypothetical relationships between variables.

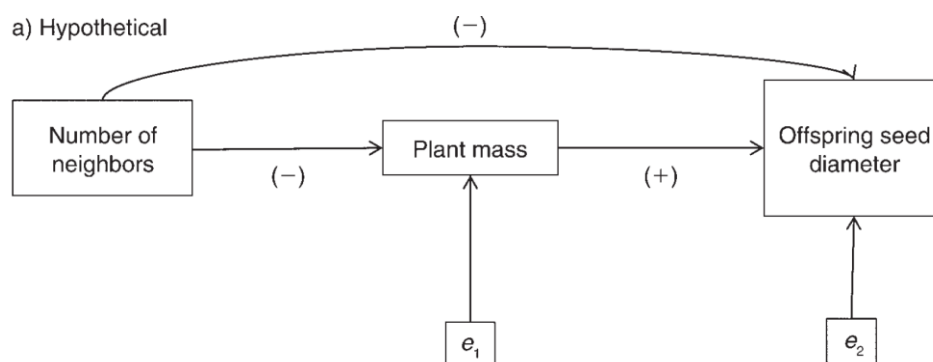


Figure 1: Path diagram showing the hypothetical relationship between the variables. The plant mass refers to the above ground biomass, and the sign “e” stands for some error.

- 1) Explain this path diagram.
- 2) Explain which type of model you would use to analyze these data. You will clearly explain which are the predictor and the response variables in your model, whether or not you include interaction terms, and what these interactions represent.

(25%) Question 4 – We conduct a study where we test whether the biodiversity estimated as the number of insect species found per tree increases with the tree height. We sample a large number of trees of various heights in two different forests (e.g. temperate and tropical). For each tree, we count the number of insect species we find anywhere on the tree. (In this data set, the number

of individual per species is not important.) The results of the statistical analysis are given below:

First part of the analysis with a GLM

```
> model1<-glm(species~height*factor(pop), poisson)
```

```
> summary(model1)
```

Call:

```
glm(formula = species ~ height * factor(pop), family = poisson)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-10.002	-4.342	-1.345	2.071	17.499

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.742288	0.102489	17.000	<2e-16	***
height	0.090784	0.004722	19.226	<2e-16	***
factor(pop)2	1.308077	0.129294	10.117	<2e-16	***
height:factor(pop)2	-0.043987	0.006253	-7.034	2e-12	***

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 6694.4 on 217 degrees of freedom

Residual deviance: 5996.2 on 214 degrees of freedom

AIC: 7123

Second part of the analysis with a GLM with negative binomial

```
> model2<-glm.nb(species~height*factor(pop))
```

```
> summary(model2)
```

Call:

```
glm.nb(formula = species ~ height * factor(pop), init.theta = 1.319332434,  
link = log)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.1086	-0.8772	-0.2470	0.3521	2.4833

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.39749	0.54319	2.573	0.0101	*
height	0.10742	0.02651	4.052	5.09e-05	***
factor(pop)2	1.70257	0.73752	2.309	0.0210	*
height:factor(pop)2	-0.06327	0.03748	-1.688	0.0914	.

(Dispersion parameter for Negative Binomial(1.3193) family taken to be 1)

Null deviance: 275.59 on 217 degrees of freedom

Residual deviance: 252.34 on 214 degrees of freedom

AIC: 2092.4

Number of Fisher Scoring iterations: 1

Theta:	1.319
Std. Err.:	0.126

- 1) Explain first the use of the different models and the differences between the two outputs as well as the reasons for these differences.
- 2) Interpret the results of this analysis. You will describe the effects in words and present a schematic figure of these effects.

(20%) Question 5 –In order to test whether the plastic response of life history traits to predator depends on the frequency of encounter with predator, Walsh and Post (2011) examined the response to predator kairomones (a molecule that indicates the presence of fish predator to prey species) in *Daphnia ambigua* (a small cladoceran) from 11 lakes: three anadromous lakes (where predator are present only part of the year), four landlocked lakes (where predator are always present) and four with no predator. These lakes do not differ significantly in size, depth, or productivity. The authors of the study established 10 clones per lake by hatching resting eggs

collected from sediment in each lake. The first laboratory generation consisted of one female per clone that was reared in a 90-mL jar under common conditions. The second laboratory generation was established by collecting neonates from the third clutch of each clone, and these individuals were reared under the same conditions as the previous generation.

They evaluated life-history responses to predator kairomones of third generation, laboratory-reared clones. The experiment was started by collecting approximately 20 neonates of similar age per clone from clutches of the second generation *Daphnia*. Four to five individuals were placed in 90-mL jars that either did ('predator' treatment) or did not ('no predator' treatment) contain predator-conditioned water. Each treatment was replicated twice per clone. The experimental conditions for the common garden experiment were similar to the rearing procedures used for the prior generations.

The traits recoded were: rates of juvenile growth, age at maturation, size at maturation and the number of embryos in the first clutch. Juvenile growth was measured by photographing random samples of two individuals per container on day 1 and 5. Length was estimated from these images and then rate of growth was calculated via: $[(\text{average length on day 5}) - (\text{average length on day 1})] / \text{no. of days}$. To estimate age at maturation, Walsh and Post monitored all *Daphnia*, twice daily (beginning on day 5), for the release of the first clutch into the brood chamber. When the release of the first clutch was confirmed, age at maturation was recorded and the first two individuals that matured in each container were photographed for estimates of size at maturation and clutch size. For each variables Walsh and Post used a single value per container by averaging the individual value recorded.

- 1 – What is the number of observational units and what is the number of experimental units considering the hypothesis tested? (You can provide a graph)
- 2 – Suggest statistical models to analyze the different variables. If you use the R syntax, you will explain clearly the structure of the model and justify why you use such model.