INTRODUCTION TO GENERAL LINEAR MODELS AND GENERAL LINEAR MIXED MODEL IN R

Ana-Johanna Voinpol-Sassu
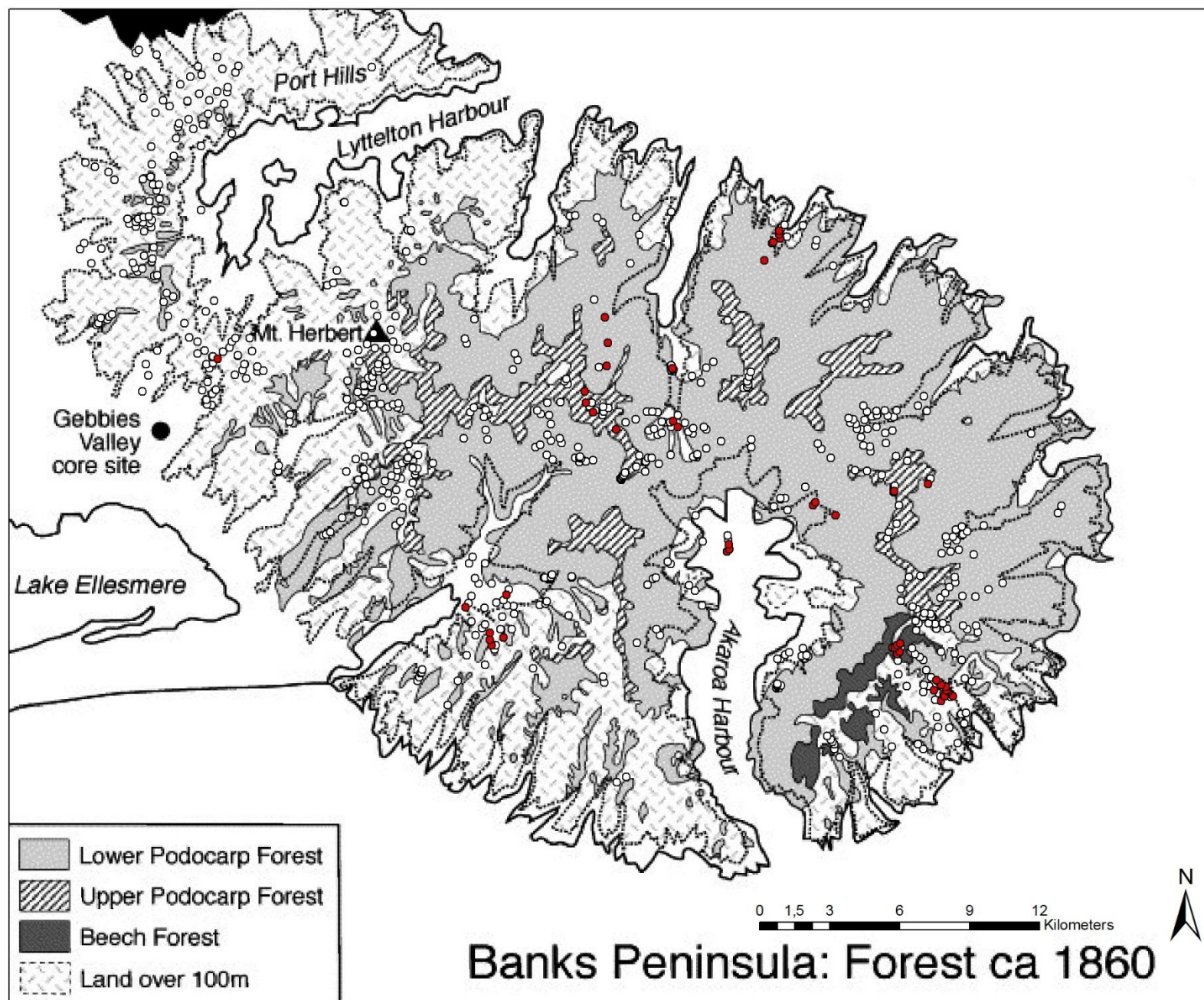
The Rifleman data and the tomtit data

Tomtit

Male rifleman,
New Zealand smallest bird (Wilson 2013)

Distribution of the Rifleman

data from Scion studies and ECan monitoring

- ● presence
- ○ abcence

Port Hills

Lyttelton Harbour

Mt. Herbert

Gebbies Valley core site

Lake Ellesmere

Akaroa Harbour

Lower Podocarp Forest

Upper Podocarp Forest

Beech Forest

Land over 100m

0 1,5 3 6 9 12 Kilometers

N

Banks Peninsula: Forest ca 1860

Autor:
Ana-Johanna Voinopol-Sassu

Tomtit abundance ~ habitat type

- Reads as Tomtit abundance predicted by habitat type
- Or also: Tomtit abundance as a function of habitat type

- Tomtit_abundance: dependent variable
- habitat_type: independent variable (predictor)

Tomtit abundance ~ habitat type + ε

fixed   random

- "ε"(read "epsilon") is a error term
  → in this model all the uncontrollable and unpredictable stuff

The model in R

- Read in the data frame (Tomtit_data.csv)

- tomtit.model <- lm (tomtit_abundance ~ habitat_type, tomtit_data)

- Note: ε not specified in lm function

- ▫ summary(tomtit.model)

Multiple R-squared: 0.1919, Adjusted R-squared: 0.1567

- **coefficient of determination**, denoted R^2 and pronounced **R squared**

- information about the **goodness of fit** of a model
- indicates how well data fit the statistical model

Multiple R-squared: 0.1919, Adjusted R-squared: 0.1567

- R^2 varies between 0 and 1
- R^2 of 1 indicates that the regression line perfectly fits the data

- Variance explained in our model is only 0.1919

F-statistic: 5.461

F = Effect Variance
    Error Variance

or

F = Between-group Variance
    Within-group Variance
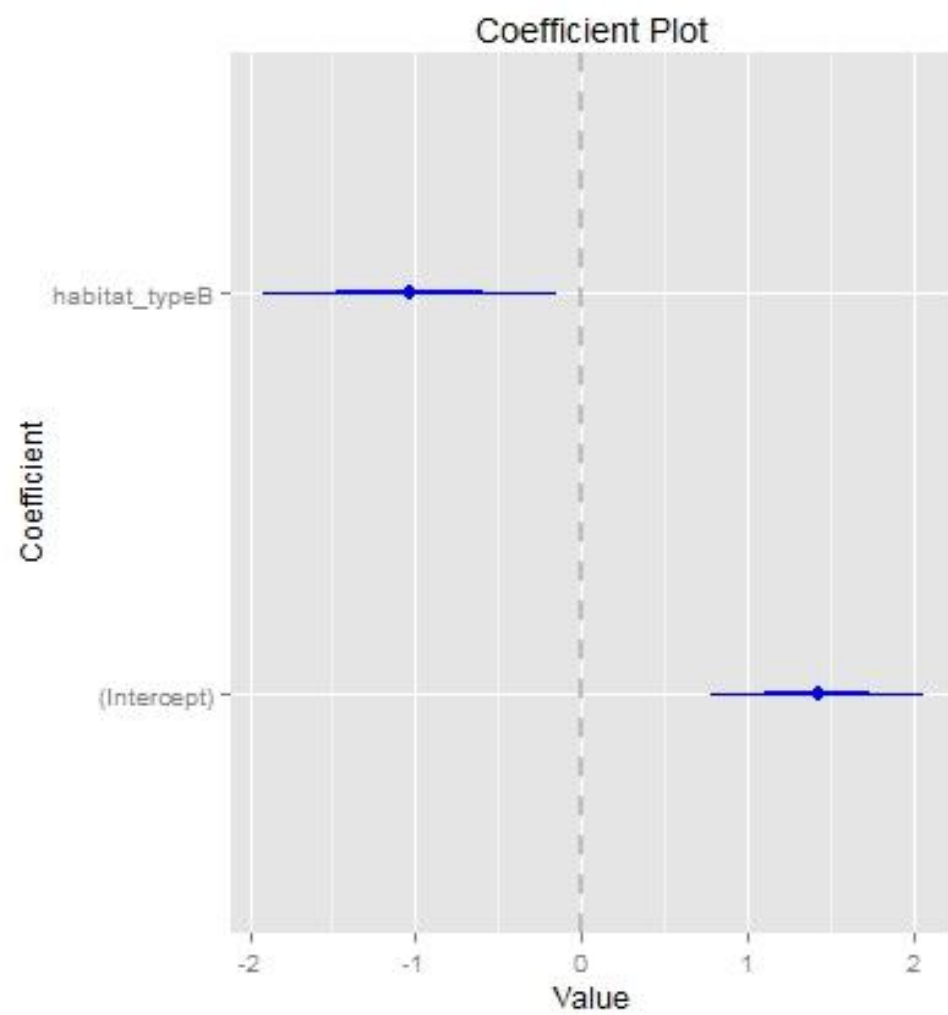
F-statistic: 5.461 on 1 and 23 DF p-value: 0.02852

Reporting results

- A linear model was build with the R version 3.0.1 (R Core Team 2013).
- This model was significant ($F_{(1, 23)}=5.461$, $p=0.028$).

▫ Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.4167    0.3185   4.448 0.000184 ***
habitat_typeB  -1.0321    0.4416  -2.337 0.028521 *
```
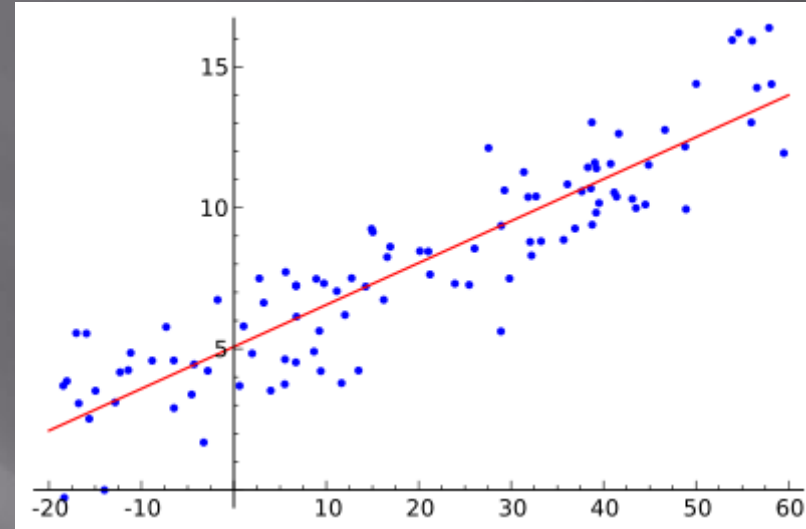
▫ Intercept is habitat type A

Coefficient Plot
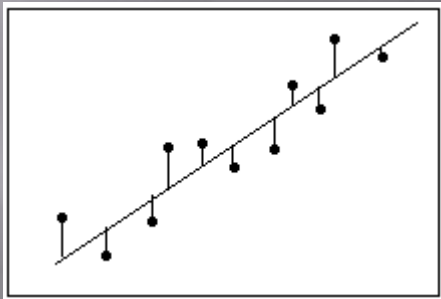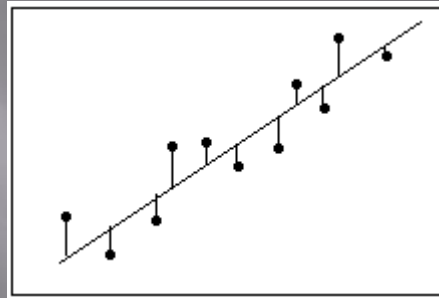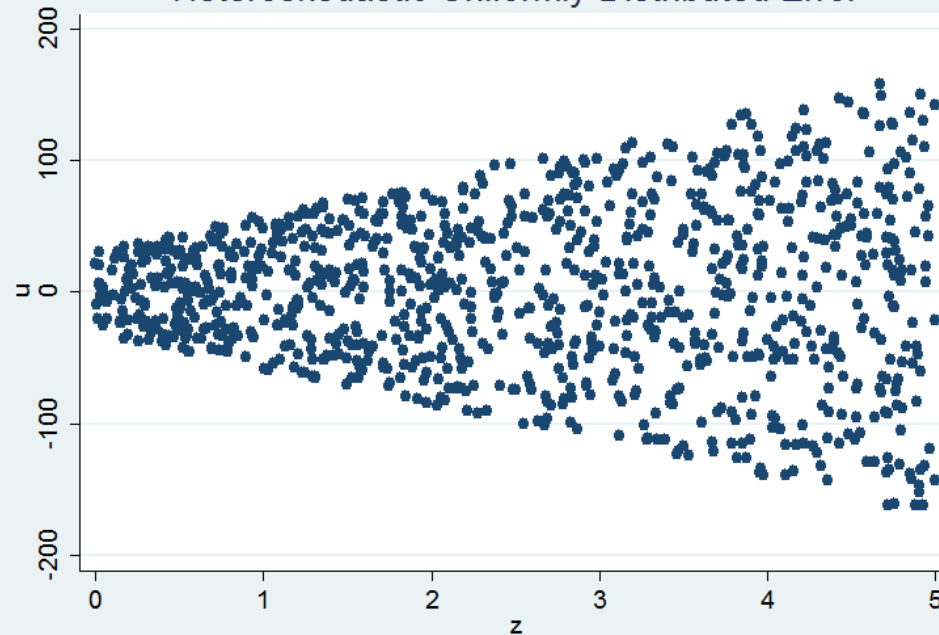
Model assumptions for GLM

Linearity

- Use residuals to check

Homoskedasticity

Heteroskedastic Uniformly Distributed Error

No multicollinearity

- multicollinearity (also collinearity) is a phenomenon in which two or more predictor variables in a multiple regression model are highly correlated,

→ one can be linearly predicted from the others with a non-trivial degree of accuracy (from Wikipedia).

What to do when there is multicollinearity

- Consider in study design
- Select the best
- Principal component Analysis (PCA)

Independence!

- Each data point comes from in independent subject

 e.g. each bird observation from a different plot

- One of most important assumption

Break!

General linear mixed model

- tomtit_abundance ~ area + connectivity + (1|patch_id)

GLMM in R

- Rif.model <- glmer(rif_presence ~ area + connectivity + (1|patch_id),family=binomial, data=Rif_data)

- summary(Rif.model)

Difference Fixed effect – Random effect

▫ Fixed effect:

→ The independent variables we have an overview of and can predict (systematic)

▫ Random effect:

→ The independent variables which seem chaotic and unpredictable

Difference Fixed effect – Random effect

- Fixed effect

→ Exhaust the population of interest

e.g. Gender, habitat type

- Random effect

→ May not exhaust the population

e.g. Plot (location)

AIC

- The **Akaike information criterion (AIC)**

→Out of collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Hence,

- AIC provides a means for model selection.

AICc

- **AICc**
- Class_Habitat + BC
- 341.6
- Class_Habitat + BC + dNL
- 343.6
- BC + dIICconnect + dNL
- 346.1
- Class_Habitat + Patch_Area + Patch_Perimeter + BC + dIICconnect + dNL
- 350.0
- Class_Habitat + Patch_Area + BC + dLC
- 371.5
- Patch_Area + BC + dIICintra
- 373.7
- Class_Habitat + Patch_Area + Patch_Perimeter + BC + dIICconnect + dLC
- 375.8
- Patch_Area + Patch_Perimeter + BC + dIICconnect + dIICintra + dNL
- 375.9
- Class_Habitat + Patch_Area + Patch_Perimeter + BC + dIICconnect + dIICintra + dLC + dNL
- 376.0

Useful package in R

- library ("MuMIn")

with function dredge(model)

Thank you!