Statistics for the Sciences

Generalized Linear Mixed Models

Xuemao Zhang
East Stroudsburg University

January 18, 2025

Outline

- Generalized Linear Mixed Models
- Lab

Generalized Linear Mixed Models

- Recall that a linear mixed model can be represented as

$$Y = X\beta + Zb + \varepsilon$$

  - Systematic part $X\beta$: β are the parameters for the fixed effects
  - Random part Zb: b is the random effects and Z is the design matrix for the random effects
  - ε is the experimental random error vector

Generalized Linear Mixed Models

- In Generalized linear models(GLMs),random effects are absent.
- Generalized Linear Mixed Models (GLMMs) generalize the GLMs by adding random effects in the systematic component:

- **Random effects**: $\boldsymbol{b}$ follows a normal distribution
- **Response**: Conditional probability distribution for $\mathbf{Y}|\boldsymbol{b}$ is from the exponential family
- **linear predictor**: The fixed and random effects are combined to form the conditional linear predictor:

$$\boldsymbol{\eta} = \boldsymbol{X}\beta + \boldsymbol{Z}b$$

- **Link function**: Connects $\boldsymbol{\eta} = g(\boldsymbol{\mu})$, where $\boldsymbol{\mu}$ is the conditional expectation $\boldsymbol{\mu} = E(\boldsymbol{Y}|\boldsymbol{b})$.

Generalized Linear Mixed Models

- Example (`ryeland.csv`): Ryeland et al (2017) studied the roosting behaviour of several species of shorebirds. They recorded the **proportion of time** (number of minutes as a proportion of total minutes in a video bout) individuals of various species spent in the backrest position while roosting. They used a binomial model with a logit link for proportions with four fixed predictors recorded for each video bout: ambient temperature, wind speed, size of group focal bird was in, and distance focal bird was from the observer. We will analyze the data for a single species, the sharp-tailed sandpiper (Calidris acuminata). This would be a standard binomial GLM except that more than one bird was sometimes recorded in each bout so bout was included as a random effect since birds closer together may be correlated in their behaviour.
  - ▶ Response: `timehb/filmp`
  - ▶ Fixed effects: `dist`, `groupsize`, `temp`, `wind`
  - ▶ Random effects: `vbout`

Generalized Linear Mixed Models

- First 15 rows of data

```
##    dist groupsize  temp  wind     filmp    timehb vbout olre
## 1    30        40 15.375 2.075  9.492067  0.000000    s1    1
## 2    30        40 15.375 2.075 13.090400 11.424667    s1    2
## 3    50        20 15.560 3.220 22.553333 18.456400    s6    3
## 4    50        20 15.560 3.220 22.553333 15.066133    s6    4
## 5    50        30 15.300 2.900  5.025000  1.217117    s7    5
## 6    50        30 15.300 2.900  5.025000  1.217333    s7    6
## 7    78        53 14.980 2.160 24.090867 22.866733    s8    7
## 8    78        53 14.980 2.160 24.090867  6.533400    s8    8
## 9   115        22 16.100 3.000 14.064600  6.623800    s9    9
## 10  115        22 16.100 3.000 14.064600 12.082533    s9   10
## 11   30       100 24.500 2.340 20.135200 18.994150   s10   11
## 12   30       100 24.500 2.340 20.135200 10.842400   s10   12
## 13   60         6 20.900 1.540 21.264150 15.222450   s11   13
## 14   60         6 20.900 1.540 21.264150 14.659650   s11   14
## 15   80        30 27.700 1.720 16.703550  0.000000   s12   15
```

Generalized Linear Mixed Models

- Create proportion of time facing back prophb
- Create integer varaiables success= timehb and fail=filmp-timehb (it is not required in SPSS)

```
##    dist groupsize   temp  wind      filmp     timehb vbout olre    prophb success
## 1    30        40 15.375 2.075   9.492067   0.000000    s1    1 0.0000000       0
## 2    30        40 15.375 2.075  13.090400  11.424667    s1    2 0.8727515      11
## 3    50        20 15.560 3.220  22.553333  18.456400    s6    3 0.8183447      18
## 4    50        20 15.560 3.220  22.553333  15.066133    s6    4 0.6680225      15
## 5    50        30 15.300 2.900   5.025000   1.217117    s7    5 0.2422123       1
## 6    50        30 15.300 2.900   5.025000   1.217333    s7    6 0.2422554       1
## 7    78        53 14.980 2.160  24.090867  22.866733    s8    7 0.9491868      22
## 8    78        53 14.980 2.160  24.090867   6.533400    s8    8 0.2711982       6
## 9   115        22 16.100 3.000  14.064600   6.623800    s9    9 0.4709554       6
## 10  115        22 16.100 3.000  14.064600  12.082533    s9   10 0.8590741      12
##     fail
## 1      9
## 2      1
## 3      4
## 4      7
## 5      3
## 6      3
## 7      1
## 8     17
## 9      7
## 10     1
```

Generalized Linear Mixed Models

- Scatter plots
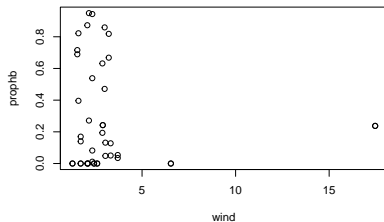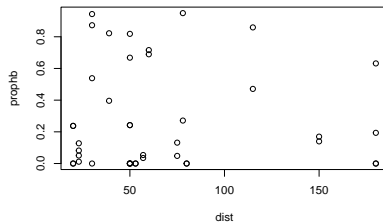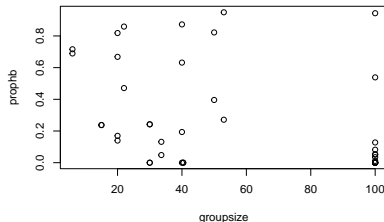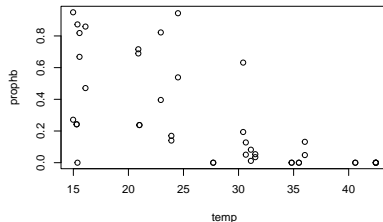
Generalized Linear Mixed Models

- Box plots
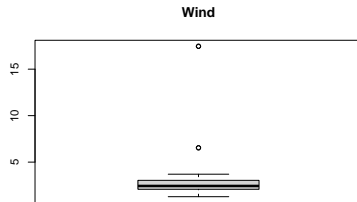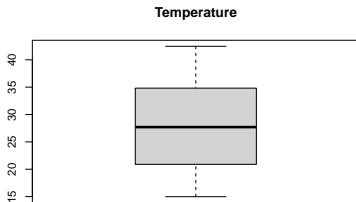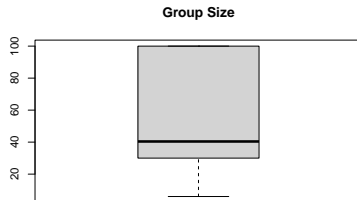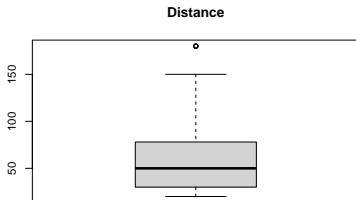
Generalized Linear Mixed Models

- Check collinearity

```
##                    dist  groupsize         temp         wind
## dist         1.00000000 -0.2030191  0.01623776 -0.27381162
## groupsize   -0.20301907  1.0000000  0.65247217 -0.15742617
## temp         0.01623776  0.6524722  1.00000000 -0.04701293
## wind        -0.27381162 -0.1574262 -0.04701293  1.00000000
```

Generalized Linear Mixed Models

- Fit GLMM with binomial distribution

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: ryeland.prop ~ dist + groupsize + temp + wind + (1 | vbout)
##    Data: ryeland
##
##      AIC      BIC   logLik deviance df.resid
##    224.0    234.4   -106.0    212.0       36
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5293 -0.6966 -0.2213  0.4084  3.3384
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  vbout  (Intercept) 1.428    1.195
## Number of obs: 42, groups:  vbout, 21
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.467829   1.231919   3.627 0.000287 ***
## dist         0.003639   0.007703   0.472 0.636656
## groupsize    0.003162   0.013106   0.241 0.809377
## temp        -0.249287   0.055305  -4.507 6.56e-06 ***
## wind        -0.031059   0.093908  -0.331 0.740841
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Generalized Linear Mixed Models

- ANOVA table

```
## Analysis of Variance Table
##             npar  Sum Sq Mean Sq F value
## dist           1  0.0831  0.0831  0.0831
## groupsize      1  8.3311  8.3311  8.3311
## temp           1 21.9971 21.9971 21.9971
## wind           1  0.1094  0.1094  0.1094
```

Lab

- After importing the data `ryeland.csv`, Convert both `filmp` and `timehb` to integers
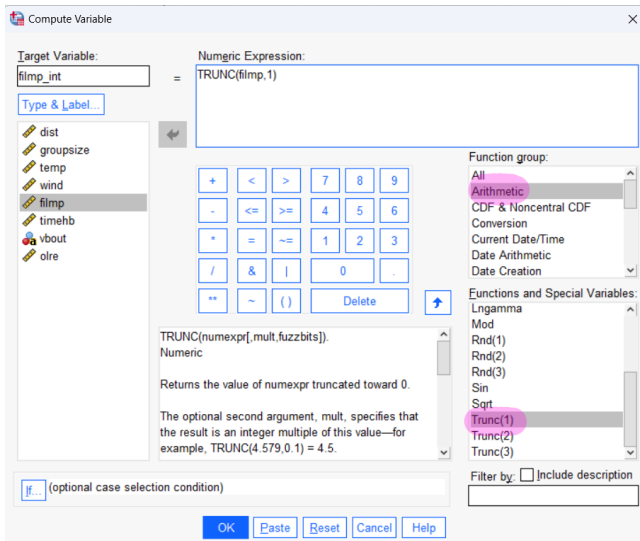
Lab

- After importing the data `ryeland.csv`, Convert both `filmp` and `timehb` to integers

Lab
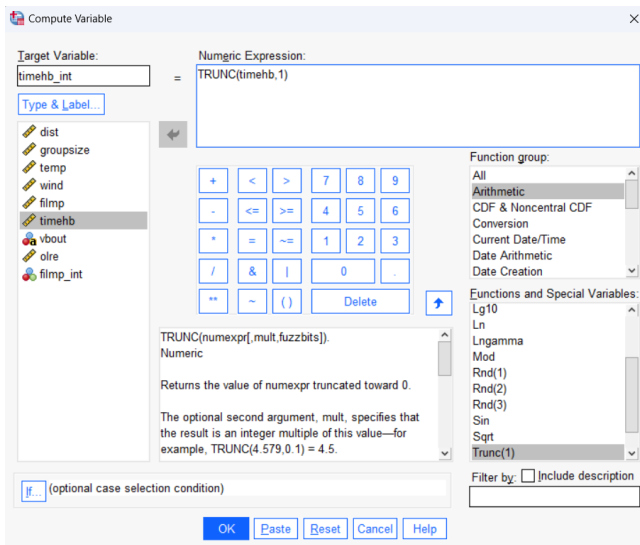
- The measures of the two new variables should be `Scale`

Lab

- Click on `Analyze → Mixed Models → Generalized Linear`
  - In the `Fields & Effects`, define our target which should be

Lab

- Define the Fixed Effects

Lab

- Define the Random Effect(s)

Lab

- Define the Random Effect(s)

Lab

- Build Options

Lab

- Build Options
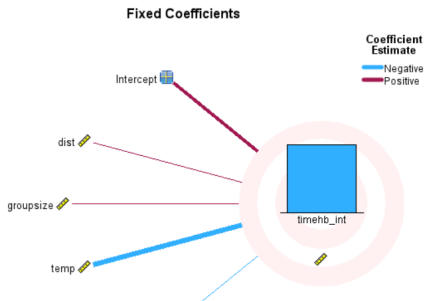 - and then run the analysis

Lab

Fixed Coefficients[a]

Model Term	Coefficient	Std. Error	t	Sig.	95% Confidence Interval		Exp(Coefficient)	95% Confidence Interval for Exp (Coefficient)	
					Lower	Upper		Lower	Upper
Intercept	4.035	1.8024	2.239	.031	.383	7.687	56.539	1.467	2179.675
dist	.003	.0086	.364	.720	-.015	.021	1.003	.985	1.022
groupsize	.002	.0144	.166	.870	-.028	.033	1.002	.972	1.034
temp	-.230	.0591	-3.885	.001	-.354	-.106	.795	.702	.900
wind	-.027	.1053	-.258	.801	-.257	.203	.973	.773	1.225

Probability distribution: Binomial
Link function: Logit

a. Target: timehb_int/filmp_int

Fixed Coefficients

License