

# 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

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}b + \epsilon$$

- ▶ Systematic part  $\mathbf{X}\beta$ :  $\beta$  are the parameters for the fixed effects
- ▶ Random part  $\mathbf{Z}b$ :  $b$  is the random effects and  $\mathbf{Z}$  is the design matrix for the random effects
- ▶  $\epsilon$  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:**  $\mathbf{b}$  follows a normal distribution
- **Response:** Conditional probability distribution for  $\mathbf{Y}|\mathbf{b}$  is from the exponential family
- **linear predictor:** The fixed and random effects are combined to form the conditional linear predictor:

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

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

# 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

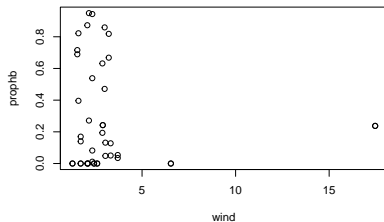
- Create proportion of time facing back prophb
- Create integer variables success= timehb and fail=filmp-timehb (it is not required in SPSS)

```
##      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
```

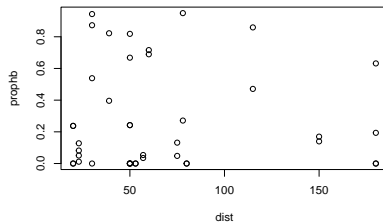
# Generalized Linear Mixed Models

- Scatter plots

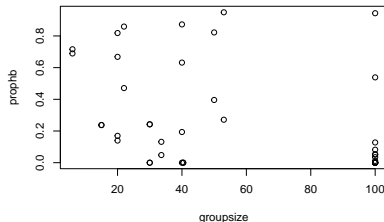
prophb vs wind



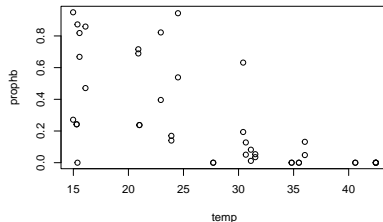
prophb vs dist



prophb vs groupsize



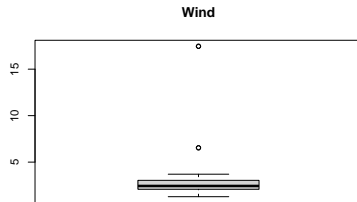
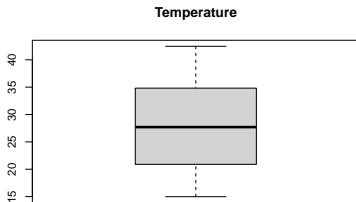
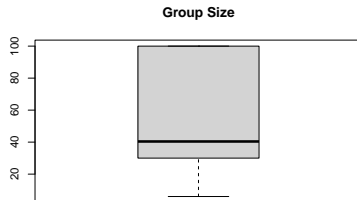
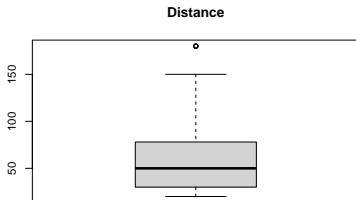
prophb vs temp





# 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
```

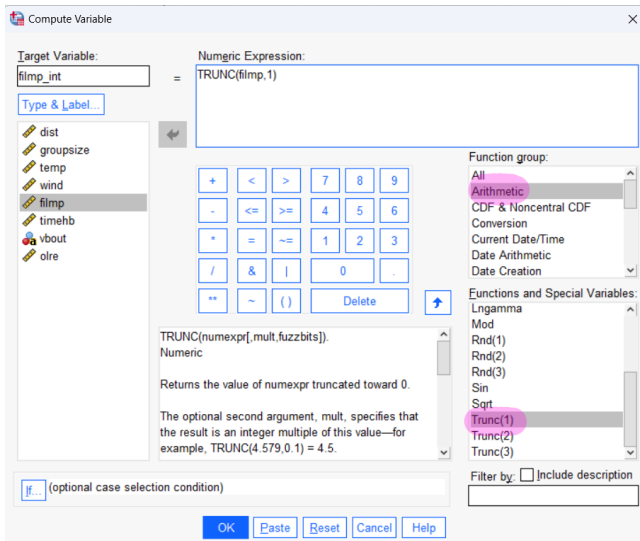
# Generalized Linear Mixed Models

- 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
```

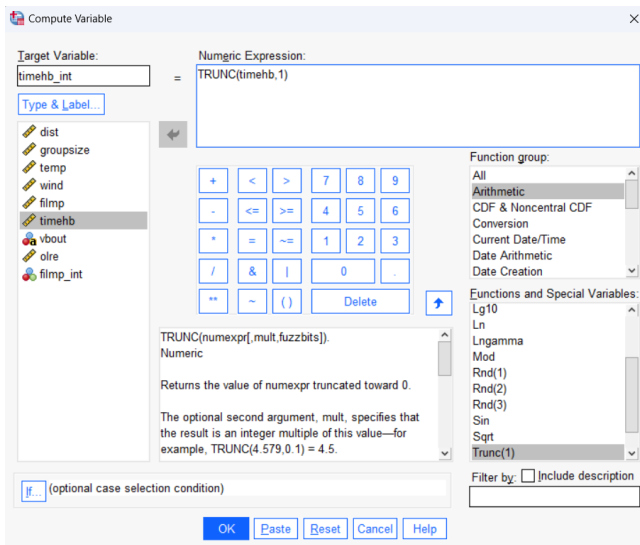
# Lab

- 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



- The measures of the two new variables should be Scale

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help											
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	dist	Numeric	3	0		None	None	8	Right	Scale	Input
2	groupsize	Numeric	4	1		None	None	8	Right	Scale	Input
3	temp	Numeric	11	8		None	None	13	Right	Scale	Input
4	wind	Numeric	11	9		None	None	13	Right	Scale	Input
5	flimp	Numeric	11	9		None	None	13	Right	Scale	Input
6	timehb	Numeric	12	9		None	None	14	Right	Scale	Input
7	vbout	String	3	0		None	None	6	Left	Nominal	Input
8	olre	Numeric	2	0		None	None	8	Right	Scale	Input
9	flimp_int	Numeric	8	2		None	None	11	Right	Scale	Input
10	timehb_int	Numeric	8	2		None	None	12	Right	Scale	Input
11											

# Lab

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

Generalized Linear Mixed Models

Data Structure Fields & Effects Build Options Model Options

Select an item:

- Target
- Fixed Effects
- Random Effects
- Weight and Offset

Target

☐ Use predefined target ☒ Use custom target

Target:

More

☒ Use number of trials as denominator ☐ Customize reference category

☒ Use field

☐ Use value

Reference value:

Target Distribution and Relationship (Link) with the Linear Model

☐ Linear model ☐ Multinomial logistic regression

☐ Gamma regression ☐ Binary logistic regression

☐ Loglinear ☐ Binary probit

☐ Negative binomial regression ☐ Interval censored survival

☒ Custom

Distribution:  Link function:

Power:

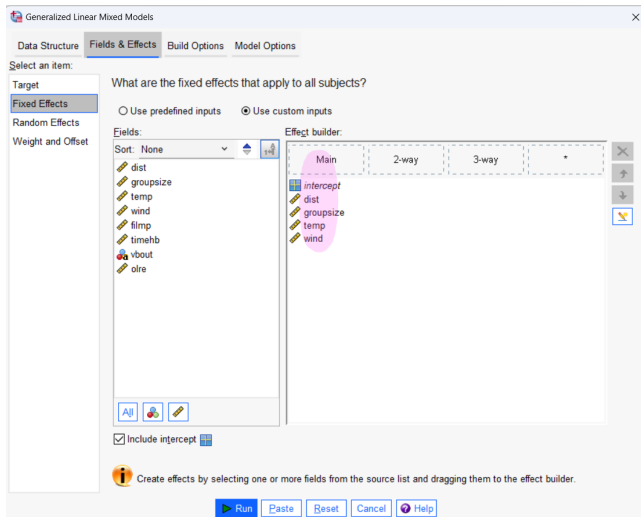
Description

Custom allows the user to choose a different target distribution and link function.

Run Paste Reset Cancel Help

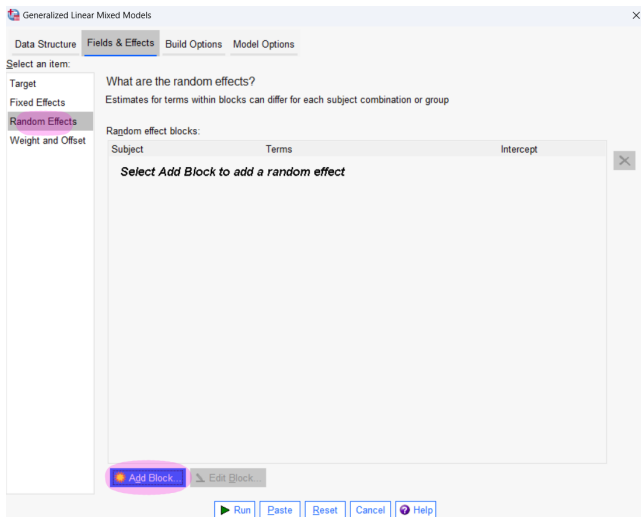


- Define the Fixed Effects

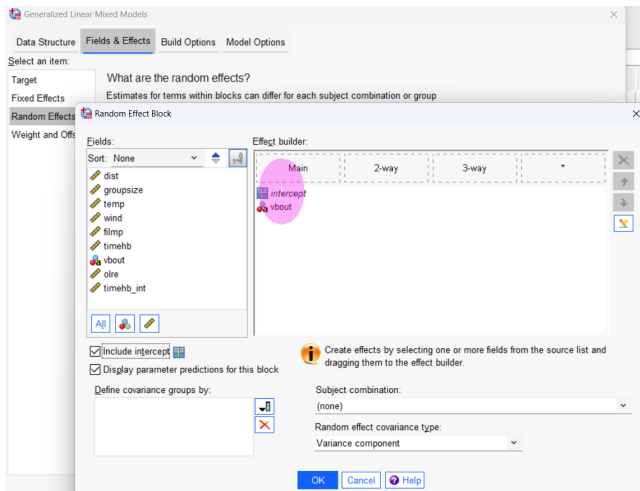


# Lab

- Define the Random Effect(s)



- Define the Random Effect(s)



## • Build Options

Generalized Linear Mixed Models

Data Structure Fields & Effects **Build Options** Model Options

Select an item:

General

Estimation

Sorting Order

Sorting order for categorical targets : Ascending

Sorting order for categorical predictors: Ascending

Stopping Rules

Maximum iterations(for both inner and outer loops): 100

Post-Estimation Settings

Confidence level(%): 95.0

Degrees of freedom

☐ Residual method

☒ Satterthwaite approximation

The Satterthwaite method is useful if the data are unbalanced or complicated covariance type is used (for example unstructured). The degrees of freedom are calculated as a function of the variance of the parameter being estimated.

☐ Kenward-Roger approximation

Tests of fixed effects and coefficients

☒ Assume model assumptions are correct (model-based covariances)

☐ Use robust estimation to handle violations of model assumptions (robust covariances)

Run Paste Reset Cancel Help

# Lab

- Build Options
  - ▶ and then run the analysis

The screenshot shows the 'Generalized Linear Mixed Models' dialog box with the 'Build Options' tab selected. The 'Estimation' sub-tab is active in the left sidebar. The 'Convergence Criteria' section contains three rows: 'Parameter Convergence' (unchecked, value 0.000001, type Absolute), 'Log-Likelihood Convergence' (checked, value 0.000001, type Absolute), and 'Hessian Convergence' (checked, value 0.00000001, type Relative). Below this, 'Maximum Fisher scoring steps' is set to 0 and 'Singularity tolerance' is set to 0.000000000001. At the bottom are buttons for Run, Paste, Reset, Cancel, and Help.

Generalized Linear Mixed Models

Data Structure Fields & Effects **Build Options** Model Options

Select an item:

- General
- Estimation**

Convergence Criteria

	Value:	Type:
<input type="checkbox"/> Parameter Convergence	0.000001	Absolute
<input checked="" type="checkbox"/> Log-Likelihood Convergence	0.000001	Absolute
<input checked="" type="checkbox"/> Hessian Convergence	0.00000001	Relative

Maximum Fisher scoring steps: 0

Singularity tolerance: 0.000000000001

Run Paste Reset Cancel Help

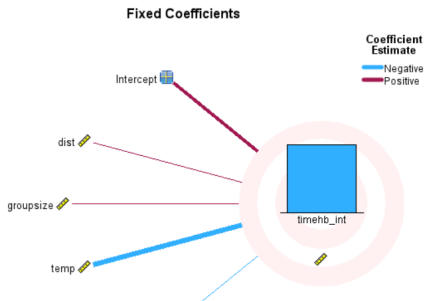
Fixed Coefficients<sup>a</sup>

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/filup\_int



# License



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).