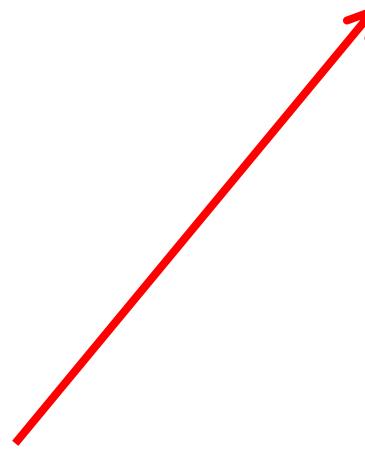


8. Dealing with variance heterogeneity & temporal autocorrelation

- Learn to recognize when you violate the assumptions of variance homogeneity
- Learn different methods that can help correct this problem

Assumptions of LMs

- $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ $\epsilon_i \sim N(0, \sigma^2)$
- Normality of errors
- Linearity and additivity
- Independence of data
- **Variance homogeneity**
 - Assumes similar variances across entire dataset (and in relation to all predictor variables)
 - Violating it (more or less) completely invalidates any test you wish to do! Your results become unreliable



How to fix it?

- Some variance heterogeneity is because the data is fitting something other than a Gaussian distribution
 - E.g. transform log-normal data
 - Clam data in R
- Some other factor is influencing the variance in your data
 - Change you the assumptions of your model to explicitly include this!

Variance heterogeneity

- You can alter the assumptions of your error term to account for variance heterogeneity
 - This is an alteration of the RANDOM portion of the model
 - But the key is that you have to know where the variance heterogeneity is coming from (in order to know what to alter)!
 - If you don't know this, then you might need to transform or use a different distribution all together

Variance heterogeneity in squid

- You are interested in what factors influence reproductive success in squid. So you measure testis weight in 768 squid. You also measure their body size and the month in which you measured it (to account for the reproductive cycle).
- What is your model?
- Go to R

Variance heterogeneity in squid

- $\text{Testisweight}_i = \text{intercept} + \text{DML}_i + \text{Month}_i + \text{DML:Month}_i + \varepsilon_i$
 $\varepsilon_i \sim N(0, \sigma^2)$
- Fixed variance structure – assumes variance increases with increasing DML
 $\varepsilon_i \sim N(0, \sigma^2 \times \text{DML}_i)$

Variance heterogeneity in Squid

- $\text{Testisweight}_i = \text{intercept} + \text{DML}_i + \text{Month}_i + \text{DML:Month}_i + \varepsilon_i$
 $\varepsilon_i \sim N(0, \sigma^2)$
- Identity variance structure – let's instead assume that the variances will be different across the different months (or any categorical variable)
 $\varepsilon_i \sim N(0, \sigma_j^2), \quad j = 1, \dots, 12$
j = different months
 - If there are no differences among months, then we get the same model as above. These models are nested!

Variance heterogeneity in squid

- $\text{Testisweight}_i = \text{intercept} + \text{DML}_i + \text{Month}_i + \text{DML:Month}_i + \varepsilon_i$
 $\varepsilon_i \sim N(0, \sigma^2)$
- Var power structure – allows the variance to increase as a power of some covariate
 $\varepsilon_i \sim N(0, \sigma^2 \times |\text{DML}_i|^{2\delta})$
 - If $\delta = 0$ then we have the same model as above, so these models are nested!

Other variance structures

Function in R	Example expression	Models the variance according to...
VarFixed	$\varepsilon_i \sim N(0, \sigma^2 \times DML_i)$	Linear increase in variance according to continuous predictor
VarIdent	$\varepsilon_i \sim N(0, \sigma_j^2), j = \text{months}$	Different variances per level of predictor
VarPower	$\varepsilon_i \sim N(0, \sigma^2 \times DML_j ^{2\delta})$	Allows variance to increase with covariate; covariate raised to power
VarExp	$\varepsilon_i \sim N(0, \sigma^2 \times e^{2\delta \times DML_j})$	Allows variance to increase as exponential power of covariate
VarConstPower	$\varepsilon_i \sim N(0, \sigma^2 \times (\delta + DML_j ^{2\delta})^2)$	Constant plus power of the variance covariate
VarComb		Combination of the variance functions

Dealing with variance heterogeneity

- Learn to recognize when you violate the assumptions of variance homogeneity
- Learn different methods that can help correct this problem

FURTHER READING on variance heterogeneity:

- Zuur Chapter 4

Temporal Autocorrelation

- Understand when violations of independence occur and what it looks like (graphically)
- Learn about the different correlation structures that can be used to account for this
- Learn how to compare the fit of different correlation structures

Violations of independence

- Random effects assume that all measurements from the same level are all equally related to each other
- But what if some measurements are more related than others?
 - E.g. time – measurements closer in time will probably be more similar than measurements farther away

Assumption of independence

- $y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \epsilon_i \sim N(0, \sigma^2)$

	Time 1	Time 2	Time 3	Time 4
Time 1	1	0	0	0
Time 2	0	1	0	0
Time 3	0	0	1	0
Time 4	0	0	0	1

This means that we should never be able to predict the next observation based on the previous observation (within a group)!

Compound symmetry

- Assumes that all measurements are related by the same amount to each other, regardless of distance

	Time 1	Time 2	Time 3	Time n
Time 1	σ^2	ρ	ρ	ρ
Time 2	ρ	σ^2	ρ	ρ
Time 3	ρ	ρ	σ^2	ρ
Time n	ρ	ρ	ρ	σ^2

- Often too simplistic for most time series, but useful for shorter series

Auto-regressive 1 (AR1)

- Assumes that measurements closer in time are more related. Correlation among measurements degrades over time

	Time 1	Time 2	Time 3	Time n
Time 1	σ^2	ρ	ρ^2	ρ^n
Time 2	ρ	σ^2	ρ	ρ^2
Time 3	ρ^2	ρ	σ^2	ρ
Time n	ρ^n	ρ^2	ρ	σ^2

e.g.

Time 1-2: $\rho = 0.5$

Time 1-3: $\rho = 0.5^2$

- The ρ parameter is unknown and needs to be estimated from the data!

Auto-regressive moving average (ARMA)

- More generalized version of AR1 where you can manipulate the parameters (ρ, q)
 - p = auto-regressive parameter
 - q = moving average parameter
- AR1 refers to the special case of ARMA(1,0)
- (p,q) values must be estimated from data, and values > 2 or 3 can generate convergence problems
- Generally viewed as “black box” to fix auto-corre problems

Temporal auto-correlation in moorhens

- Researchers measured moorhen abundance every year in Hawaii from 1958 – 2003. We want to test whether these abundances are related to Year and Rainfall
- What does our model look like?



Temporal Autocorrelation

- Understand when violations of independence occur and what it looks like (graphically)
- Learn about the different correlation structures that can be used to account for this
- Learn how to compare the fit of different correlation structures

FURTHER READING on temporal autocorrelation:

- Zuur Chapter 6
- Barnett et al. 2010. Using information criteria to select the correct variance-covariance structure for longitudinal data in ecology. *Methods in Ecology & Evolution* 1.

On spatial correlation:

- Zuur Chapter 7