

Generalized Least squares (GLS)

BIO 599: Ecological Data Analysis

Reminders



Thursday: Paper discussion and GLS workshop



Office hours Monday 10:30-11:30 & Tuesday 4:00-5:00



Anonymous Feedback anytime!

Goal for today

- Learn how to use generalized least squares (GLS) to model data where the relationship between predictors and response is normally distributed, but the variance of the residuals is not constant and may depend on one or more predictor variables.

Outline

- Overview of heterogeneity
- Generalized Least Squares
- Variance structures
- Examples:
 - Unequal variances among groups
 - Variance increasing with a predictor
- General model selection protocol

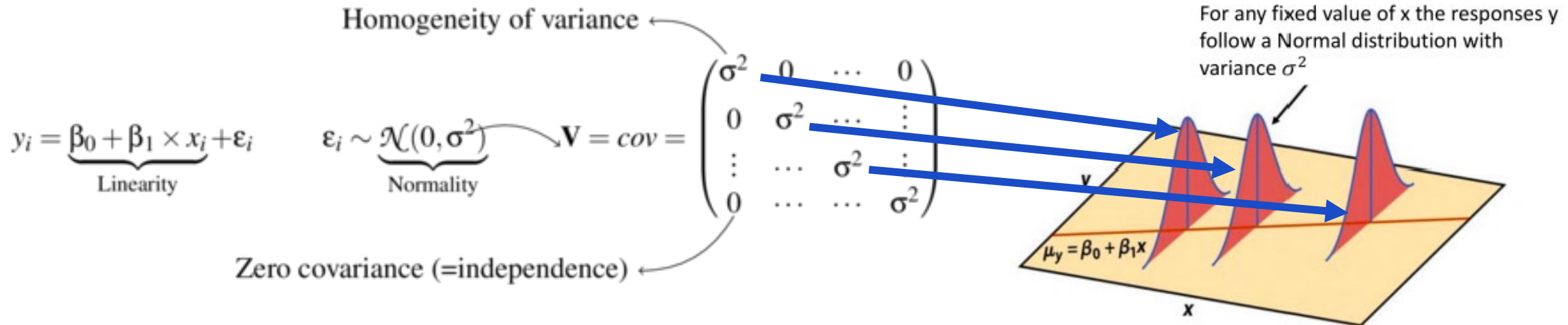
Alain F. Zuur • Elena N. Ieno
Neil J. Walker • Anatoly A. Saveliev
Graham M. Smith

**Mixed Effects
Models and
Extensions in
Ecology with R**

 Springer

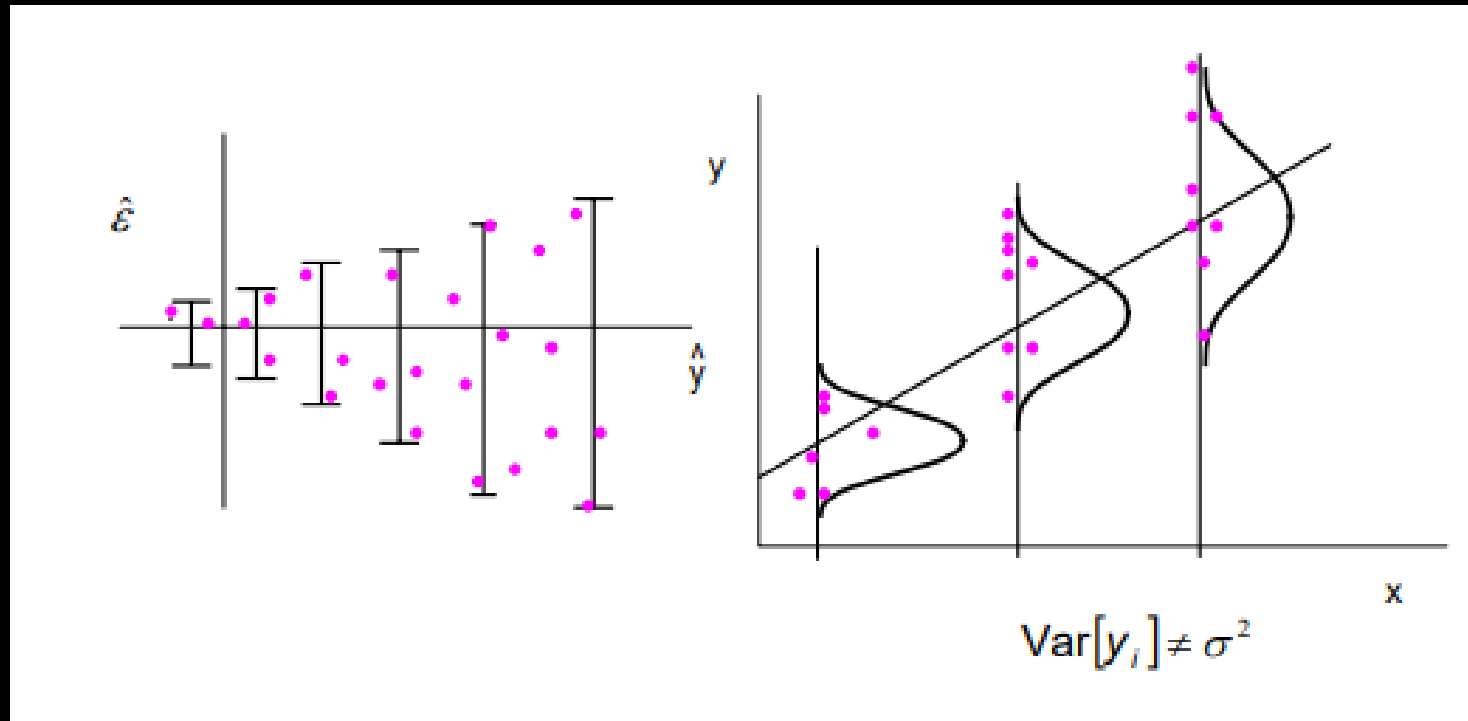
Problems with linear regression

1. Heterogeneity (today)
2. Nested data
3. Temporal and Spatial Correlation



What happens when we violate the homogeneity of variance assumption?

- Regression parameters with incorrect standard errors
- F and t statistic no longer follows an F and t distribution
- Invalidates statistics for assessing statistical significance



Solution to heterogeneity

1. Data transformation

- BUT heterogeneity is interesting ecological information – we don't want to throw it away!

2. Use Generalized least squares (GLS) to incorporate heterogeneity into models using the 'random term': using a variance structure

Outline

- Overview of heterogeneity
- Generalized Least Squares
- Variance structures
- Examples:
 - Unequal variances among groups
 - Variance increasing with a predictor
- General Model selection protocol
- Likelihood

Generalized Least Squares

OLS	GLS
Weights are the identity matrix; constant variance (σ^2) for all observations	Uses OLS to obtain estimates of fixed effects and generates maximum likelihood (ML) to estimate the variance-covariance parameters and update the fixed effects parameters
Minimizes the sum of squared residuals with equal weights for all observations.	Minimizes the weighted sum of squared residuals, giving more weight to observations with smaller variances and less weight to those with larger variances
	Weights are incorporated through the inverse of the variance-covariance matrix (Σ^{-1}), allowing for different variances and covariances.

The key difference lies in the explicit consideration of a variance-covariance matrix in GLS, allowing for more flexibility in handling heteroskedasticity and providing more efficient estimates when the assumption of constant variance is violated.

Key steps in Generalized Least Squares

Step 1: OLS Estimation of Fixed Effects

- We start with the familiar OLS estimation of fixed effects, obtaining initial estimates for our regression coefficients.

Step 2: MLE Estimation of Variances

- Next, we turn to MLE to estimate the variance-covariance matrix of the residuals. This step involves finding the parameters that maximize the likelihood of our observed data given the proposed covariance structure.

Step 3: Iterative Updating

- We iteratively update our fixed effects using the MLE estimates of variances. This process continues until convergence is achieved, refining our model with each iteration.

Generalized Least Squares

Linear Model with Constant Variance

$$y_i = \beta_0 + \beta_1 \times x_i + \epsilon_i \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Linear Model with Weighted Variance

$$y_i = \beta_0 + \beta_1 \times x_i + \epsilon_i \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2 \times \omega)$$

GLS allows observations to be given different weights based on their variances.

Outline

- Overview of heterogeneity
- Generalized Least Squares
- Variance structures
- Examples:
 - Unequal variances among groups
 - Variance increasing with a predictor
- General Model selection protocol
- Likelihood

Variance structures

- An explanatory variable that is used in the variance of the residuals is called a variance covariate.
- There are several possible types available in the nlme package:

R function	Description
VarIdent	Different variances per stratum
VarFixed	Fixed variance
VarPower	Power of the variance covariate
VarExp	Exponential of the variance covariate
VarConstPower	Constant plus power of the variance covariate
VarComb	A combination of variance functions

How to choose a variance structure

- `varIdent` if the variance covariate is nominal
- `varFixed`, `varPower`, `varExp`, and `varConstPower` – all allow for an increase (or decrease) in residual variation along a continuous variance covariate
- `varFixed` is limited – assumes that the variance of the residuals is linearly related to a variance covariate.
- `varPower` should not be used if the variance covariate is zero (or close to)
- `varExp` if variance covariate can take the value of zero
- `varComb` - combine different models for different variables

How to choose a variance structure

- Use a number of different weight functions
- Look at distribution of residuals
- Compare the fit of models using AIC or log-likelihood ratio tests
- A priori biological knowledge



Discussion

Based on BOTH biological knowledge and data exploration:

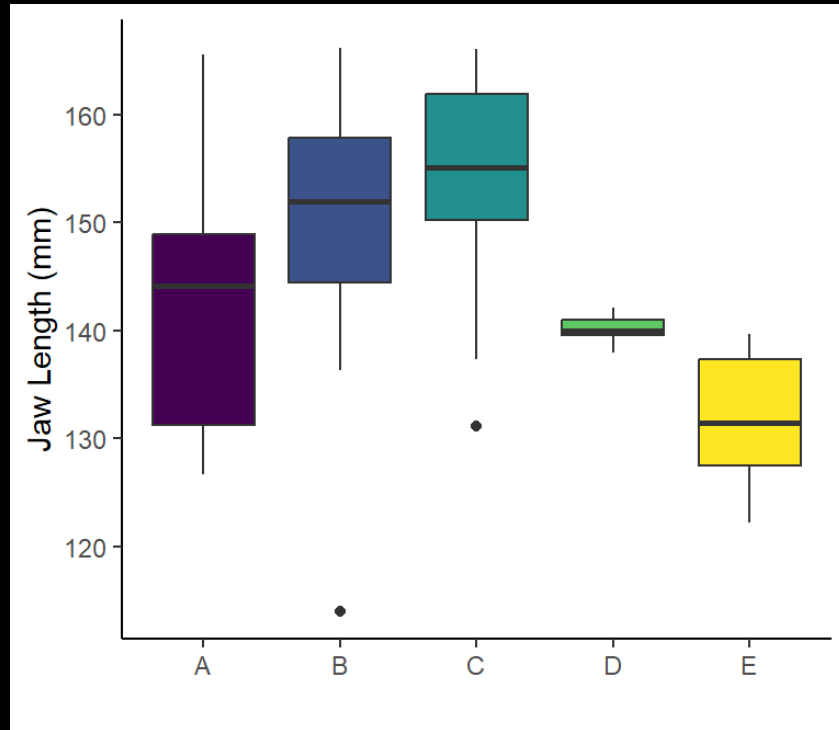
- What are potential sources of heterogeneity in your data?
- What type of variance structure might you use?

Outline

- Overview of heterogeneity
- Generalized Least Squares
- Variance structures
- Examples:
 - Unequal variances among groups
 - Variance increasing with a predictor
- General Model selection protocol
- Likelihood

Variance heterogeneity among groups

- Difference in mandible lengths (mm) of different populations of golden jackals – specimens from British Museum



Variance heterogeneity among groups

- Assume mandible length is normally distributed and that variance for each population is the same (nlme::gls)

```
> jaw.lm <- lm(jaw ~ pop, data = data.het1)
> summary(jaw.lm)

Call:
lm(formula = jaw ~ pop, data = data.het1)

Residuals:
    Min       1Q   Median       3Q      Max
-34.490  -4.012   0.877   6.512  23.409

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  142.115      3.335   42.614 <2e-16 ***
popB         6.369       4.716    1.350  0.184
popC        11.281       4.716    2.392  0.021 *
popD        -1.934       4.716   -0.410  0.684
popE       -10.640       4.716   -2.256  0.029 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.55 on 45 degrees of freedom
Multiple R-squared:  0.3584,    Adjusted R-squared:  0.3014
F-statistic: 6.285 on 4 and 45 DF,  p-value: 0.0004173
```

```
> jaw.gls <- gls(jaw ~ pop, data = data.het1, method = "REML")
> summary(jaw.gls)
Generalized least squares fit by REML
Model: jaw ~ pop
Data: data.het1
      AIC      BIC    logLik
363.2352 374.0752 -175.6176

Coefficients:
              Value Std. Error  t-value p-value
(Intercept) 142.11524  3.334962  42.61375  0.0000
popB         6.36859  4.716349   1.35032  0.1837
popC        11.28068  4.716349   2.39182  0.0210
popD        -1.93415  4.716349  -0.41009  0.6837
popE       -10.63974  4.716349  -2.25593  0.0290

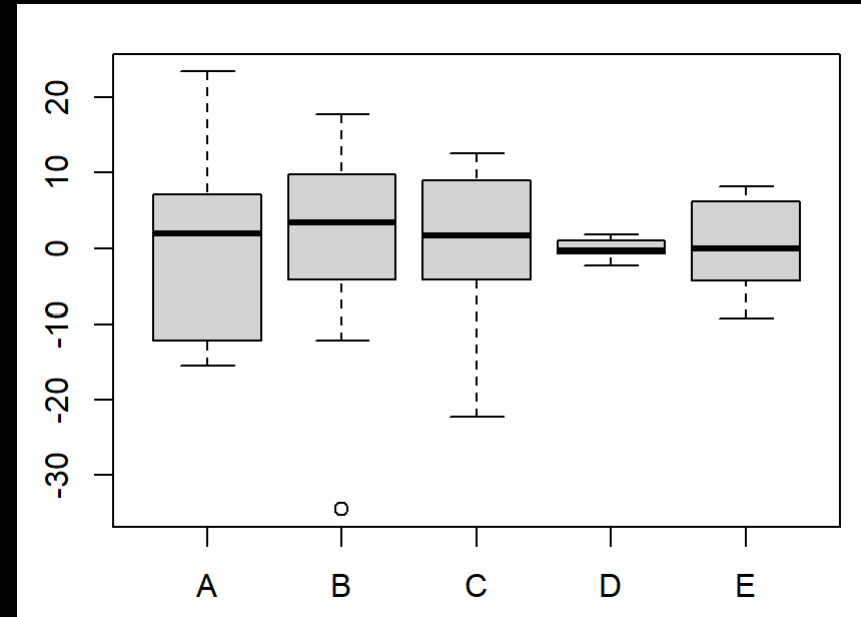
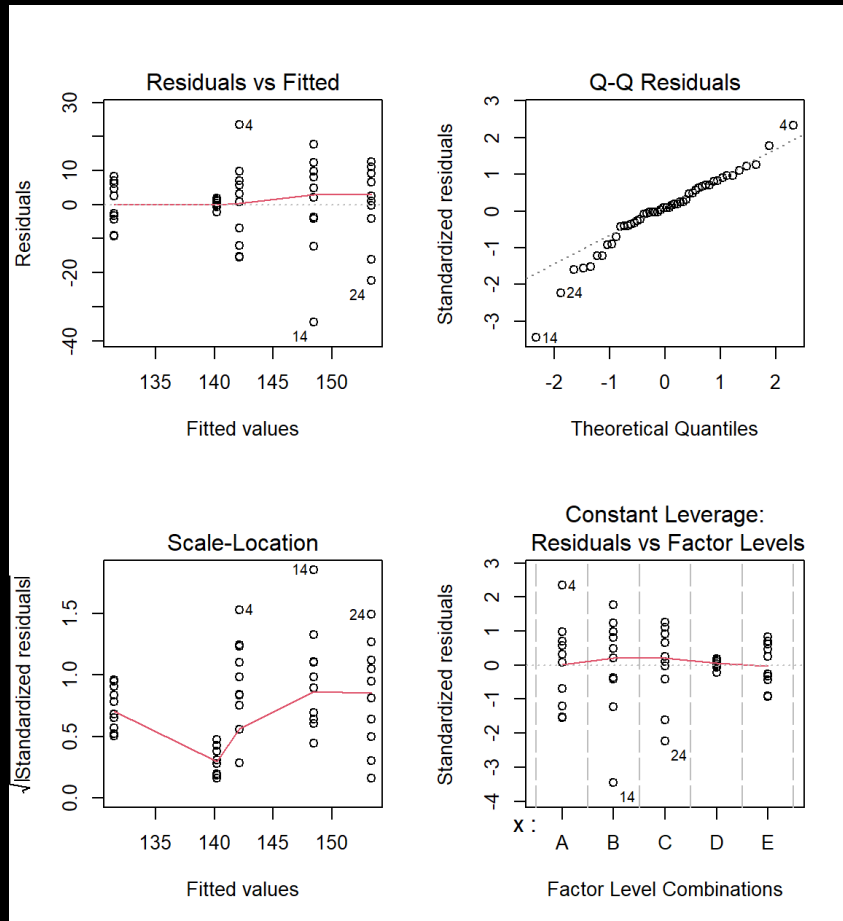
Correlation:
      (Intr) popB   popC   popD
popB -0.707
popC -0.707  0.500
popD -0.707  0.500  0.500
popE -0.707  0.500  0.500  0.500

Standardized residuals:
      Min       Q1       Med       Q3      Max
-3.27037539 -0.38045415  0.08313911  0.61750992  2.21971167

Residual standard error: 10.54608
Degrees of freedom: 50 total; 45 residual
```

Variance heterogeneity among groups

- Are the assumptions met?



Variance heterogeneity among groups

- fit this model using the gls function in the nlme package
- specify the variance model using the weights argument

```
library(nlme)
jaw.gls1 <- gls(jaw ~ pop, data = data.het1, weights = varIdent(form = ~1 | pop), method="REML")
summary(jaw.gls1)
```

Variance heterogeneity among groups

```
> summary(jaw.gls1)
Generalized least squares fit by REML
Model: y ~ x
Data: data.het1
      AIC      BIC    logLik
334.6022 352.6688 -157.3011

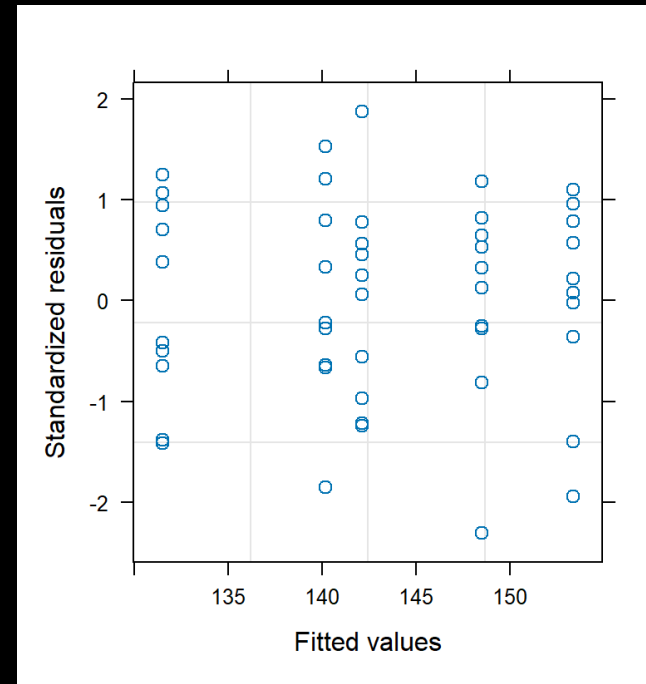
Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | x
Parameter estimates:
      A      B      C      D      E
1.00000000 1.19887818 0.91816537 0.09711061 0.52472103

Coefficients:
              Value Std. Error t-value p-value
(Intercept) 142.11524  3.949480 35.98328  0.0000
xB           6.36859  6.165882  1.03287  0.3072
xC          11.28068  5.361742  2.10392  0.0410
xD          -1.93415  3.968059 -0.48743  0.6283
xE          -10.63974  4.460172 -2.38550  0.0213

Correlation:
(Intr) xB      xC      xD
xB -0.641
xC -0.737  0.472
xD -0.995  0.638  0.733
xE -0.885  0.567  0.652  0.881

Standardized residuals:
      Min      Q1      Med      Q3      Max
-2.3034220 -0.6178649  0.1064903  0.7596770  1.8743363

Residual standard error: 12.48935
Degrees of freedom: 50 total; 45 residual
```



*Estimate the mean jaw length for each population.
Are jaw lengths more variable for population A, B,
C, D, or E?*

Variance heterogeneity among groups

- Assess if the variance structure is “worth it”
- Null hypothesis = variances of each group are equal

```
> AIC(jaw.gls, jaw.gls1)
      df      AIC
jaw.gls   6 363.2352
jaw.gls1 10 334.6022
> anova(jaw.gls, jaw.gls1)
      Model df      AIC      BIC    logLik  Test  L.Ratio p-value
jaw.gls     1   6 363.2352 374.0752 -175.6176
jaw.gls1    2  10 334.6022 352.6688 -157.3011 1 vs 2 36.63308 <.0001
```


Variance heterogeneity among groups

- Compare model summaries and residuals

```
> summary(jaw.gls)
Generalized least squares fit by REML
Model: jaw ~ pop
Data: data.het1

      AIC      BIC    logLik
363.2352 374.0752 -175.6176

Coefficients:
              Value Std.Error  t-value p-value
(Intercept) 142.11524  3.334962 42.61375  0.0000
popB         6.36859  4.716349  1.35032  0.1837
popC        11.28068  4.716349  2.39182  0.0210
popD        -1.93415  4.716349 -0.41009  0.6837
popE       -10.63974  4.716349 -2.25593  0.0290

Correlation:
      (Intr) popB   popC   popD
popB -0.707
popC -0.707  0.500
popD -0.707  0.500  0.500
popE -0.707  0.500  0.500  0.500

Standardized residuals:
      Min      Q1      Med      Q3      Max
-3.27037539 -0.38045415  0.08313911  0.61750992  2.21971167

Residual standard error: 10.54608
Degrees of freedom: 50 total; 45 residual
```

```
> summary(jaw.gls1)
Generalized least squares fit by REML
Model: jaw ~ pop
Data: data.het1

      AIC      BIC    logLik
334.6022 352.6688 -157.3011

Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | pop
Parameter estimates:
      A      B      C      D      E
1.00000000 1.19887818 0.91816537 0.09711061 0.52472103

Coefficients:
              Value Std.Error  t-value p-value
(Intercept) 142.11524  3.949480 35.98328  0.0000
popB         6.36859  6.165882  1.03287  0.3072
popC        11.28068  5.361742  2.10392  0.0410
popD        -1.93415  3.968059 -0.48743  0.6283
popE       -10.63974  4.460172 -2.38550  0.0213

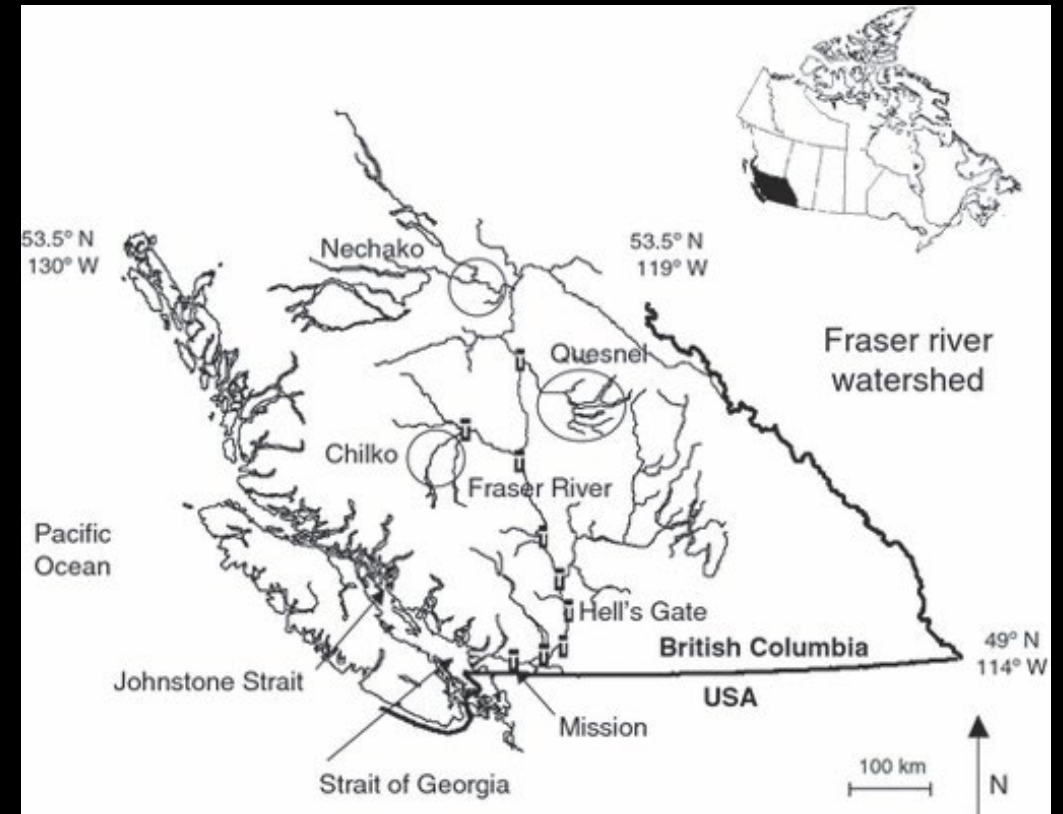
Correlation:
      (Intr) popB   popC   popD
popB -0.641
popC -0.737  0.472
popD -0.995  0.638  0.733
popE -0.885  0.567  0.652  0.881

Standardized residuals:
      Min      Q1      Med      Q3      Max
-2.3034220 -0.6178649  0.1064903  0.7596770  1.8743363

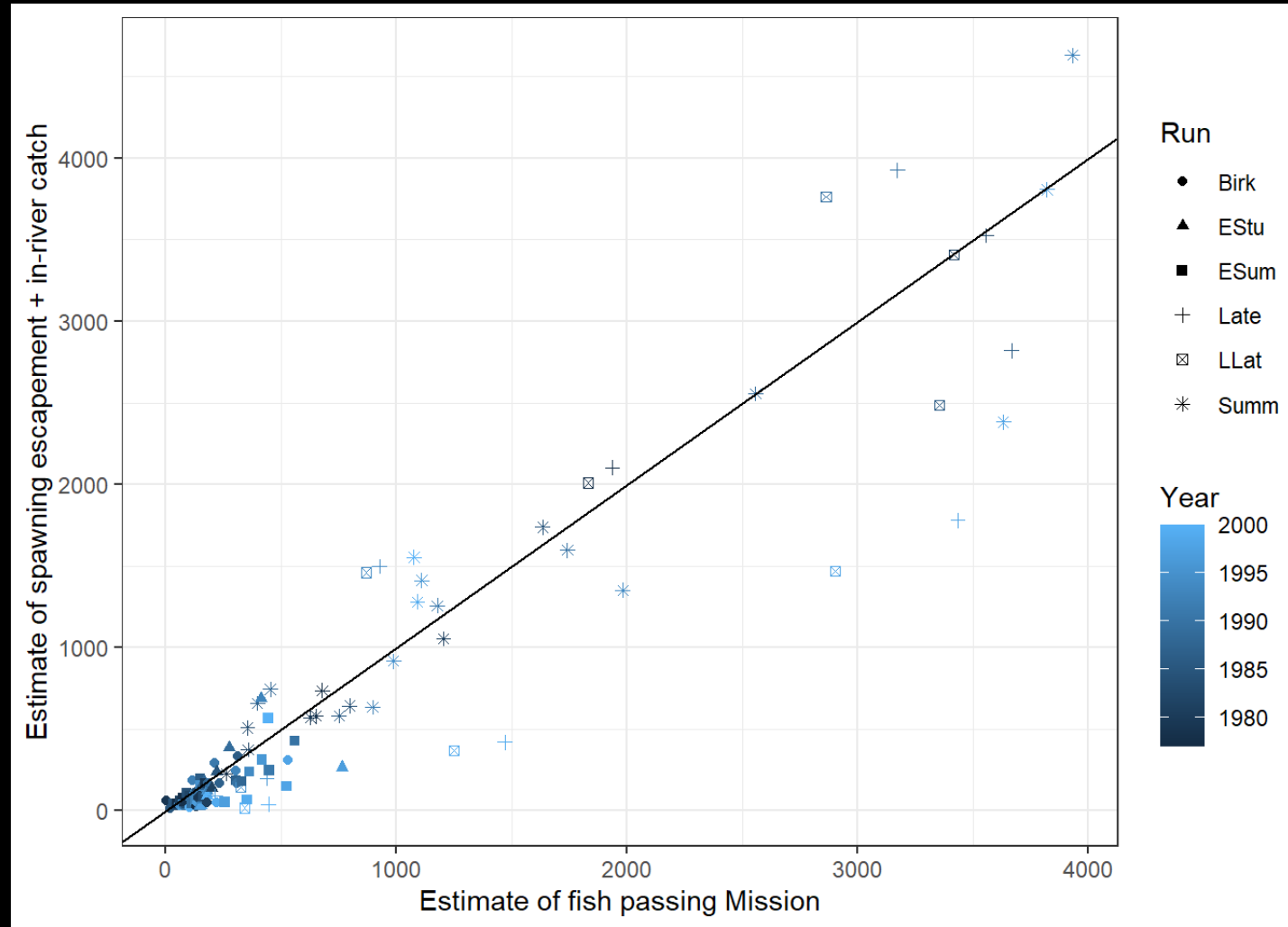
Residual standard error: 12.48935
Degrees of freedom: 50 total; 45 residual
```

Variance increasing with a predictor

- Data used to manage sockeye salmon (*Oncorhynchus nerka*) that spawn in the Fraser River system of Canada
- Can estimates of fish passing by Mission (X) predict future harvest plus spawning escapement (Y)?

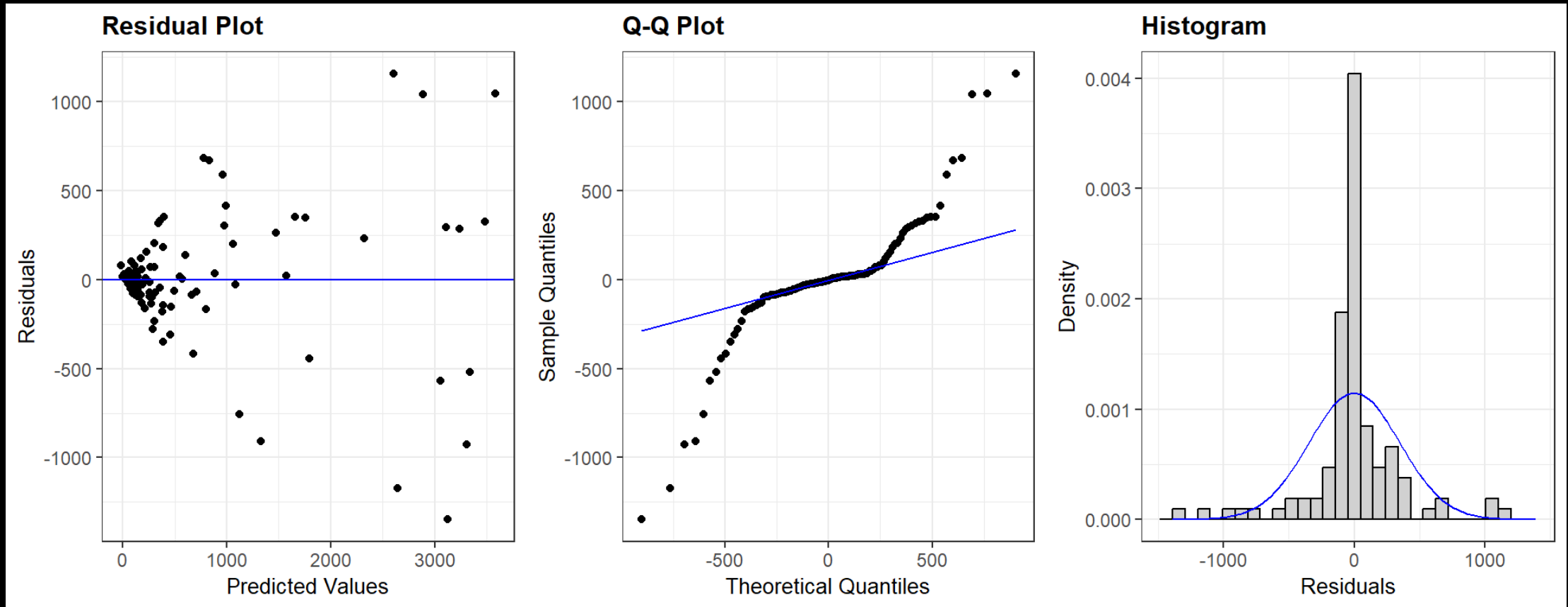


Variance increasing with a predictor



Summed catch and spawning escapement of Fraser Rivers sockeye salmon plotted against an estimate of the number of fish passing Mission. The black line is a 1-1 line which might be expected if there were no in-river mortalities above Mission.

Variance increasing with a predictor



```
lmsockeye <- lm(SpnEsc ~ MisEsc, data = sdata)
ggResidpanel::resid_panel(lmsockeye, plots = c("resid", "qq", "hist"), nrow = 1)
```

Variance increasing with a predictor

Model the variance as a function of the count at Mission:

1. Fixed variance model – covariate has to be positive
2. Power variance model – covariate $\neq 0$
3. Exponential variance model
4. Constant + power variance model – better than power when covariate is close to 0

```
fixedvar <- gls(SpnEsc ~ MisEsc, weights = varFixed(~ MisEsc), data = sdata)
varpow <- gls(SpnEsc ~ MisEsc, weights = varPower(form = ~ MisEsc), data = sdata)
varexp <- gls(SpnEsc ~ MisEsc, weights = varExp(form = ~ MisEsc), data = sdata)
varconstp <- gls(SpnEsc ~ MisEsc, weights = varConstPower(form = ~ MisEsc), data = sdata)
```

Variance increasing with a predictor

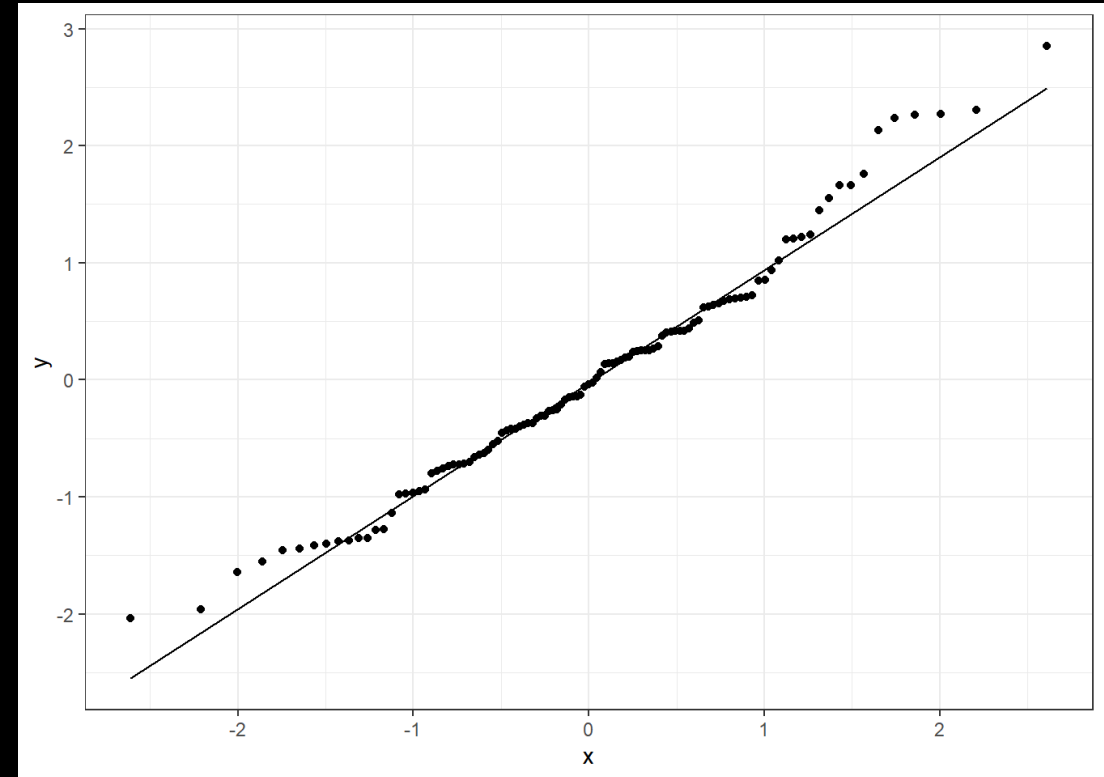
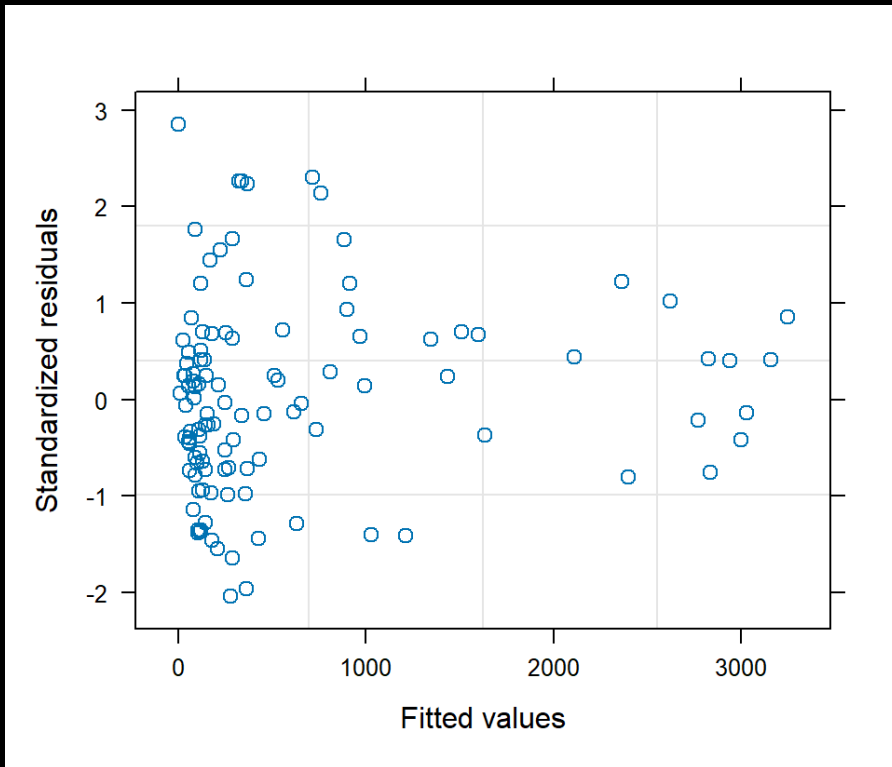
- Compare models using anova and AIC

```
> anova(lmsockeye, fixedvar, varpow, varexp, varconstp)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
lmsockeye	1	3	1614.925	1622.999	-804.4624			
fixedvar	2	3	1459.525	1467.599	-726.7625			
varpow	3	4	1446.861	1457.627	-719.4307	2 vs 3	14.66348	1e-04
varexp	4	4	1482.668	1493.434	-737.3341			
varconstp	5	5	1421.523	1434.980	-705.7614	4 vs 5	63.14534	<.0001

```
> |
```

Variance increasing with a predictor



Variance increasing with a predictor

```
> summary(varconstp)
Generalized least squares fit by REML
Model: SpnEsc ~ MisEsc
Data: sdata
      AIC      BIC    logLik
1421.523 1434.98 -705.7614

Variance function:
Structure: Constant plus power of variance covariate
Formula: ~MisEsc
Parameter estimates:
      const      power
119.49037    1.10369

Coefficients:
              Value Std.Error   t-value p-value
(Intercept) -6.122849  8.644421 -0.708301  0.4803
MisEsc       0.828173  0.052627 15.736618  0.0000

Correlation:
(Intr)
MisEsc -0.666

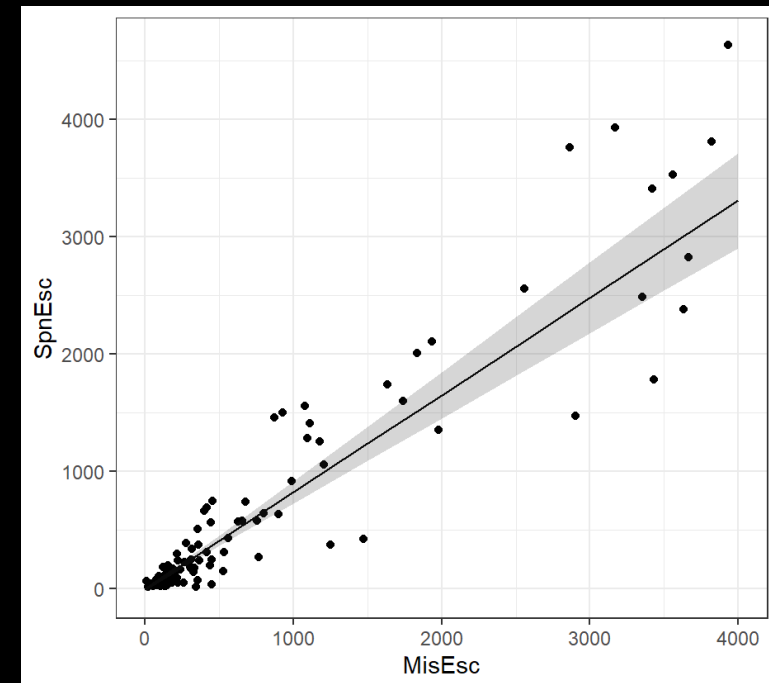
Standardized residuals:
      Min      Q1      Med      Q3      Max
-2.03554341 -0.68214833 -0.03890184  0.61977181  2.85039770

Residual standard error: 0.172081
Degrees of freedom: 111 total; 109 residual
```

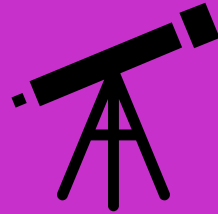
```
library(ggeffects)
pred <- ggpredict(varconstp, terms = "MisEsc")

head(pred)
pred=as.data.frame(pred)

g=ggplot()+geom_point(data=sdata, aes(y=SpnEsc, x=MisEsc))
g=g+geom_line(data=pred, aes(y=predicted, x=x))
g=g+ geom_ribbon(data=pred, aes(ymax=conf.high, ymin=conf.low, x=x), alpha=0.2)
g = g+theme_bw()
g
```



Quick reflection in pairs



WHAT'S ONE THING YOU FEEL CLEAR
ABOUT SO FAR REGARDING VARIANCE
STRUCTURES?



WHAT'S ONE THING YOU'RE STILL
CONFUSED ABOUT?

Outline

- Overview of heterogeneity
- Generalized Least Squares
- Variance structures
- Examples:
 - Unequal variances among groups
 - Variance increasing with a predictor
- General Model selection protocol
- Likelihood

General protocol for model selection

1. Start with a linear regression model – a FULL model – and test assumptions:
 - a) Residuals are normally distributed
 - b) Homogeneity – plot standardized residuals versus fitted and each explanatory variable
2. Repeat step 1 using gls from nlme (because anova cannot compare objects from lm and gls)
 - a) Ensure REML estimation is used
3. Depending on graphs in #1, choose an appropriate variance structure

General protocol for model selection

4. Fit a new gls model with the selected variance-covariance structure
 - a) use REML estimation and same explanatory variables
 - b) find the optimal random structure using as many explanatory variables in the fixed part as possible
5. Compare the new GLS model with the earlier results using the AIC, BIC, or likelihood ratio test
 - a) If new model is better, extract and inspect residuals (step 1); if OK go to #7, if not, go to #6
6. Go to step #4 and use a different variance structure. If still no good, try:
 - a) improving the fixed component (additive modeling)
 - b) try a different distribution
 - c) transformation (last resort!)

General protocol for model selection

7. Find the optimal fixed structure – significance of explanatory variables
 - a) t-statistics – summary
 - b) F-statistic - anova (remember this is sequential – only use for interaction terms)
 - c) likelihood ratio test – full and nested model (use ML)
8. Remove terms until all terms are significant
9. Reapply the model with REML estimation; check assumptions
10. Present the results in a table and interpret

Outline

- Overview of heterogeneity
- Generalized Least Squares
- Variance structures
- Examples:
 - Unequal variances among groups
 - Variance increasing with a predictor
- General Model selection protocol
- Likelihood

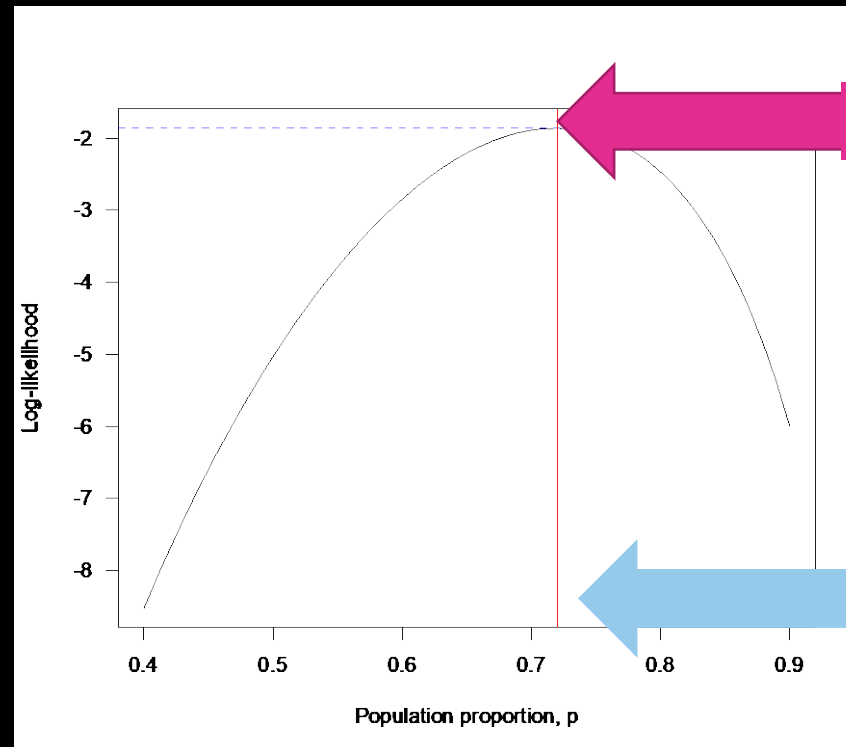
What is Likelihood?

- Data are fixed; parameters vary
- Likelihood = support for parameter values
- Higher likelihood → stronger support

Determines if we can trust the parameters in a model based on the sample data that we've observed.

Maximum likelihood estimation (ML)

- Find parameter values that maximize the likelihood
- Estimates fixed effects + variance parameters together
- Used for comparing models with different predictors



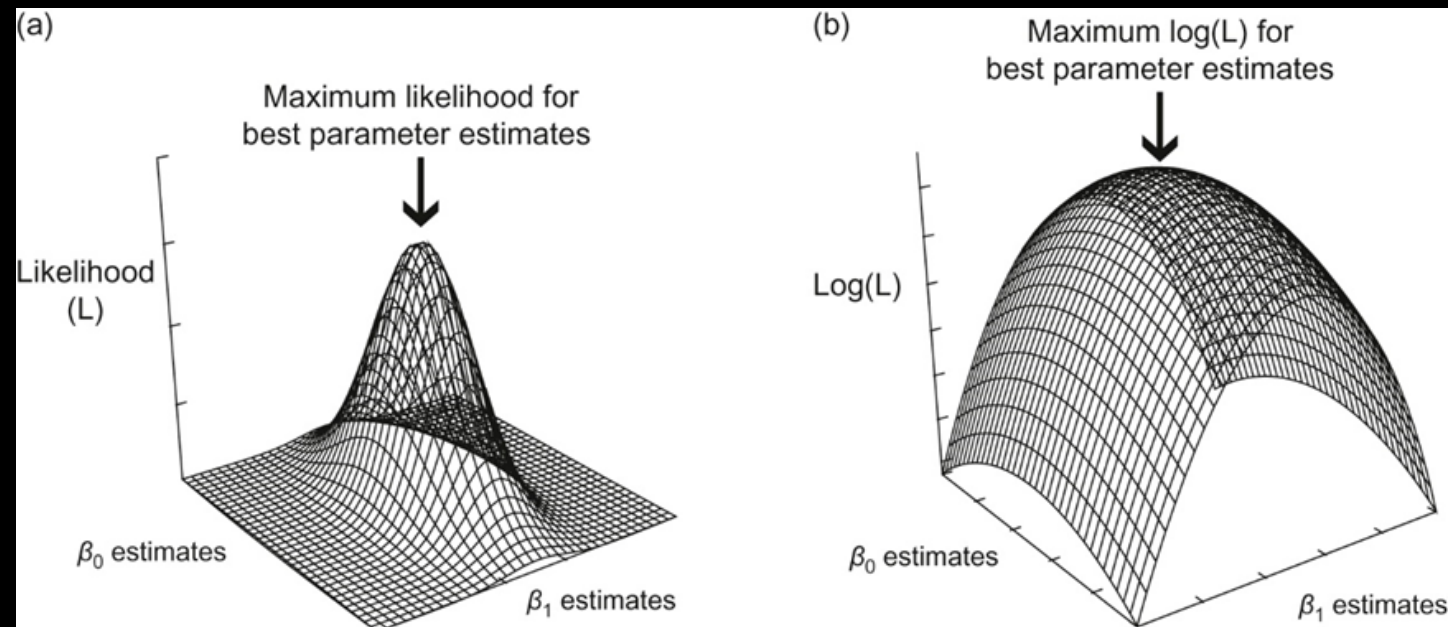
Highest log-likelihood is peak

- Maximum likelihood estimate is the parameter value that is most strongly supported
- **Parameter value that intersects with a vertical line to the peak of the curve**

Why use log-likelihood instead of likelihood?

Short answer:

Both do the same job, but log-likelihood is computationally simpler and easier to interpret and graph



Log-likelihood ratio test

- **Law of likelihood:** extent to which data supports one parameter value or hypothesis against another is equal to the ratio of their likelihoods (difference in their log-likelihoods)
- Each model has a log-likelihood.
- The likelihood ratio test looks to see if the difference between the log-likelihood from the two models is statistically significant.
- If the difference is statistically significant, then one model has a better fit.

Log-likelihood ratio test

```
> AIC(jaw.gls, jaw.gls1)
      df      AIC
jaw.gls   6 363.2352
jaw.gls1 10 334.6022
> anova(jaw.gls, jaw.gls1)
      Model df      AIC      BIC    logLik  Test  L.Ratio p-value
jaw.gls     1   6 363.2352 374.0752 -175.6176
jaw.gls1    2  10 334.6022 352.6688 -157.3011 1 vs 2 36.63308 <.0001
```

Test hypotheses:

- * Null hypothesis: the log-likelihoods from one model is equal to the other
- * Alt hypothesis: the log-likelihoods are not equal to one another

Interpretation:

If the p-value for the test is $< .05$, then we reject the null hypothesis and one model improves the fit of the model.

Maximum Likelihood (MLE)

- estimates everything together → compare fixed effects
- USE to compare fixed effects
- DO NOT use to compare random effects

Restricted ML (REML)

- ‘corrects for the degrees of freedom’
- adjusts for fixed effects → better variance estimates
- DO NOT use to compare fixed effects
- USE to compare random effects

Report parameter estimates from your final “best” REML model

GLS – summary



GLS can be used to model data where there is not constant variance



GLS assigns different weights to different observations based on the 'weighting matrix'



Choose a variance structure that reflects the structure of heteroskedasticity in the data and confirm using AIC or log-likelihood ratio tests (method=REML)