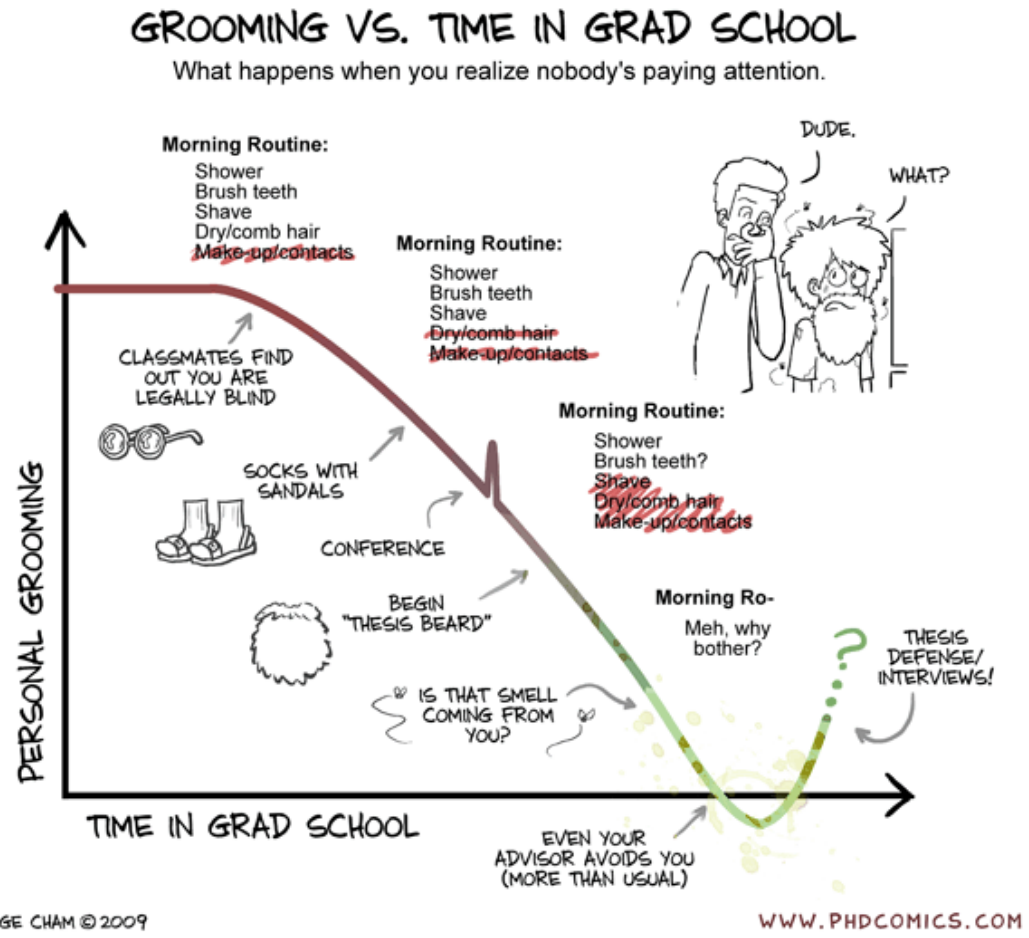
Linear Models

ENTMLGY 6707 Entomological Techniques and Data Analysis



Learning objectives

1) Understand when to conduct a simple linear regression

2) Interpret the outcome of a simple linear regression

3) Evaluate assumptions of simple linear regression

Linear regression

Table 4.1 Examples of the Generalized Linear Model as a Function of Independent Variable and Dependant Variable Type

		Responses				
		Continuous DV	*Binary DV*	*Unordered Multicategory DV*	*Ordered Categorical DV*	*Count DV*
Predictors	*Continuous IV*	OLS regression	Binary logistic regression	Multinomial logistic regression	Ordinal logistic regression	OLS, Poisson regression
	Mixed continuous and categorical IV					
	Binary/ categorical IV only	ANOVA and *t*-test	Log-linear models	Log-linear models		Log-linear models

ANOVA, analysis of variance; DV, dependent variable; IV, independent variable; OLS, ordinary least squares.

Chapter 4: Simple Linear Models With Continuous Dependent Variables: Simple ANOVA Analyses
In: Regression & Linear Modeling: Best Practices and Modern Methods

Simple linear regression

Table 4.1 Examples of the Generalized Linear Model as a Function of Independent Variable and Dependant Variable Type

		Responses				
		Continuous DV	*Binary DV*	*Unordered Multicategory DV*	*Ordered Categorical DV*	*Count DV*
Predictors	*Continuous IV*	OLS regression	Binary logistic regression	Multinomial logistic regression	Ordinal logistic regression	OLS, Poisson regression
	Mixed continuous and categorical IV					
	Binary/ categorical IV only	ANOVA and *t*-test	Log-linear models	Log-linear models		Log-linear models

ANOVA, analysis of variance; DV, dependent variable; IV, independent variable; OLS, ordinary least squares.

Chapter 4: Simple Linear Models With Continuous Dependent Variables: Simple ANOVA Analyses
In: Regression & Linear Modeling: Best Practices and Modern Methods

Linear regression

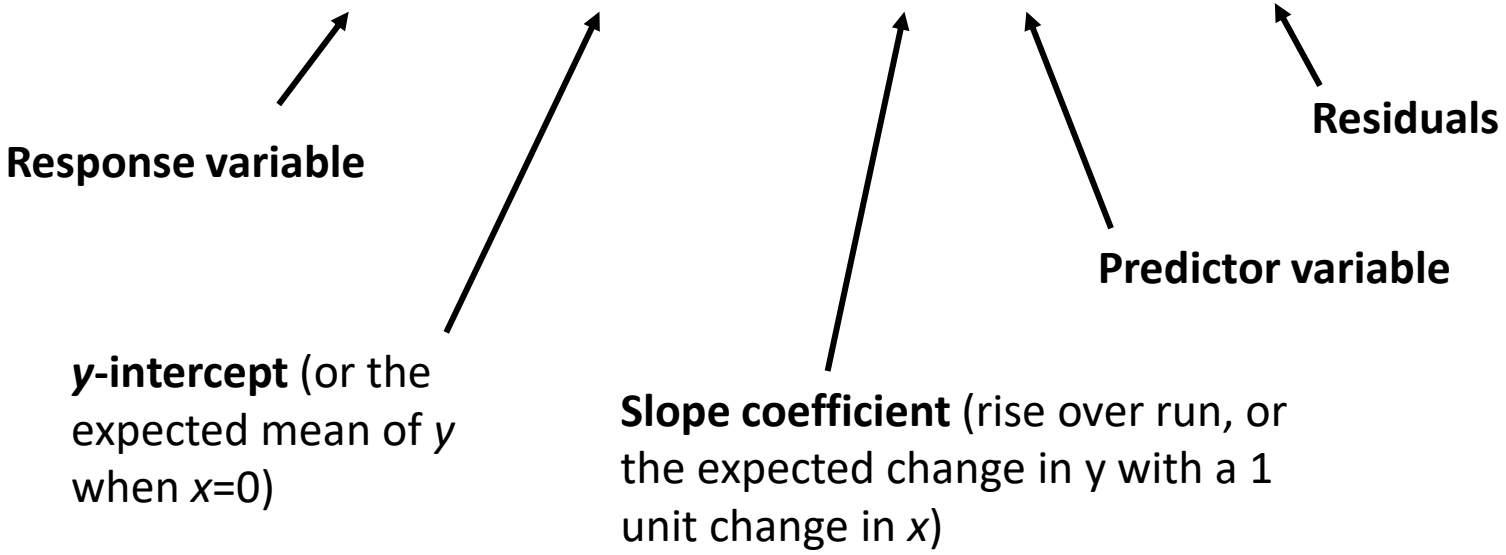
$$y = b + mx$$

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

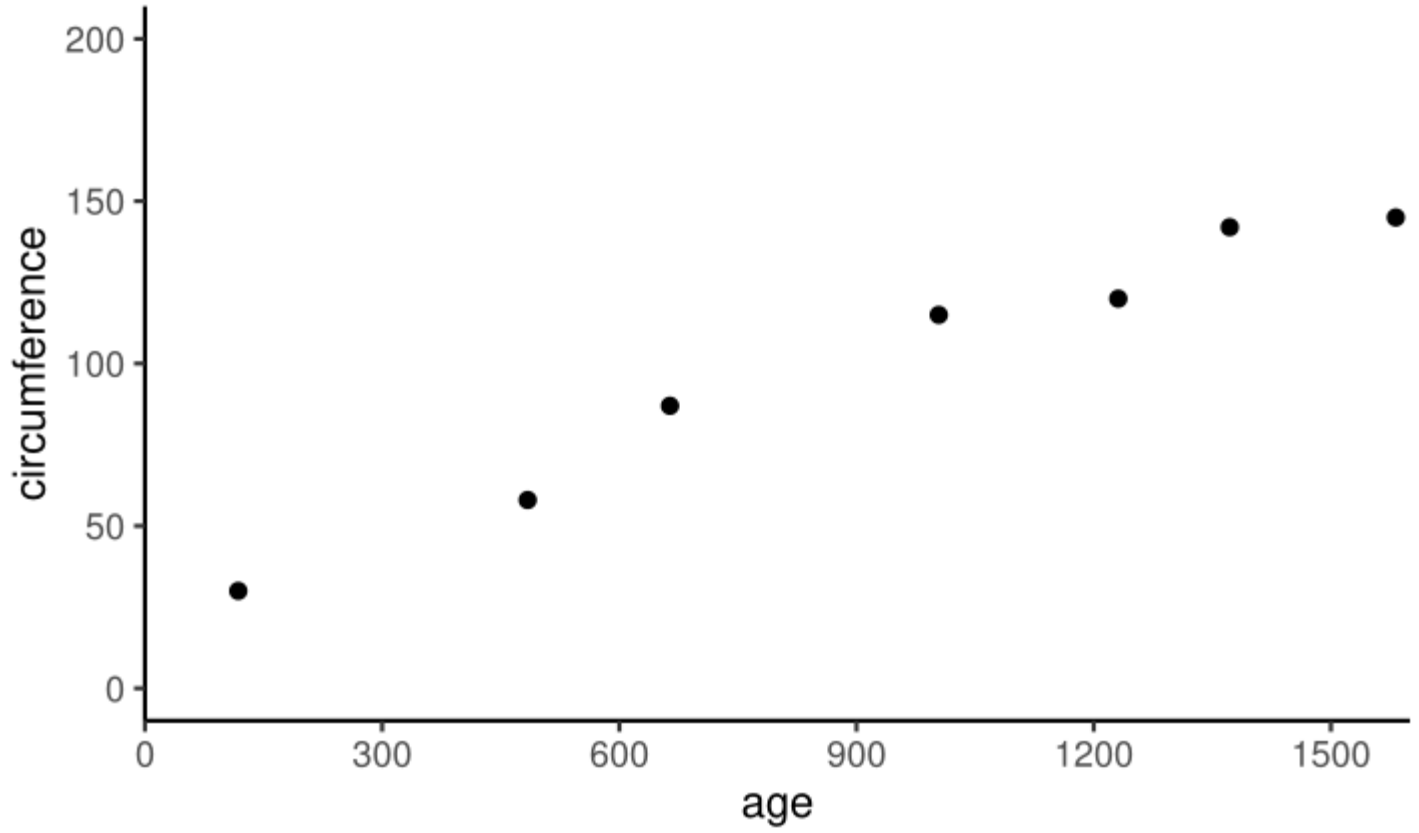
Linear regression

$$y = b + mx$$

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

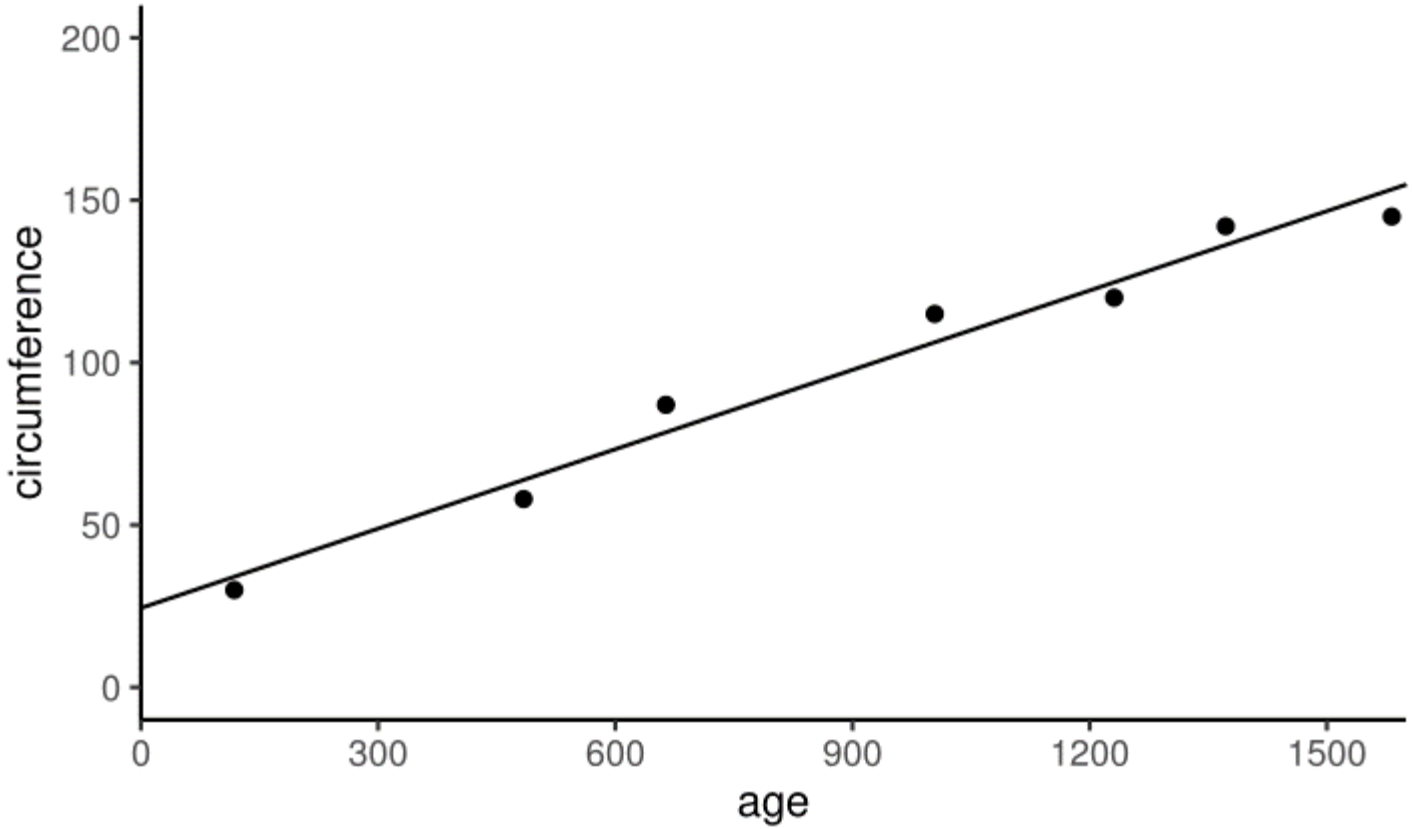


Example: size of orange trees



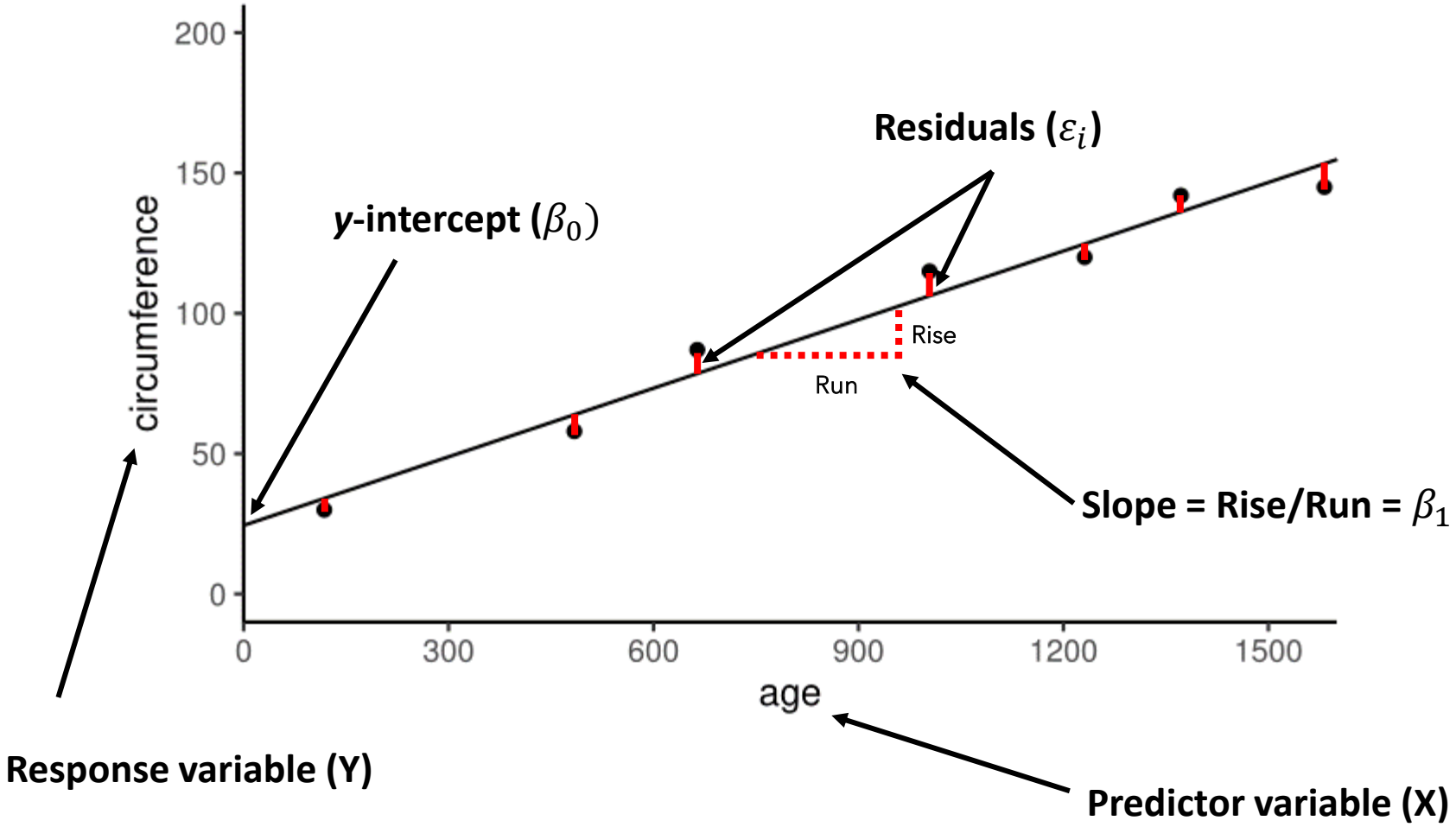
Example: size of orange trees

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$



Example: size of orange trees

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$



Linear regression in R

```
fit_oranges_1 <- lm(circumference~age, data=OT)
summary(fit_oranges_1)
```

```
##
## Call:
## lm(formula = circumference ~ age, data = OT)
##
## Residuals:
##      1      2      3      4      5      6      7
## -4.052 -5.873  8.461  8.759 -4.736  5.775 -8.335
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.437847   6.543311   3.735   0.0135 *
## age          0.081477   0.006281  12.973 4.85e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.056 on 5 degrees of freedom
## Multiple R-squared:  0.9711, Adjusted R-squared:  0.9654
## F-statistic: 168.3 on 1 and 5 DF,  p-value: 4.852e-05
```

e_i → Residuals

β_0 → (Intercept)

β_1 → age

Linear regression in R

```
fit_oranges_1 <- lm(circumference~age, data=OT)
summary(fit_oranges_1)
```

```
##
## Call:
## lm(formula = circumference ~ age, data = OT)
##
## Residuals:
##      1      2      3      4      5      6      7
## -4.052 -5.873  8.461  8.759 -4.736  5.775 -8.335
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 24.437847   6.543311   3.735   0.0135 *
## age          0.081477   0.006281  12.973 4.85e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.056 on 5 degrees of freedom
## Multiple R-squared:  0.9711, Adjusted R-squared:  0.9654
## F-statistic: 168.3 on 1 and 5 DF,  p-value: 4.852e-05
```

e_i → Residuals

β_0 → (Intercept)

β_1 → age

Assumptions

1. The relationship between response and predictor is linear

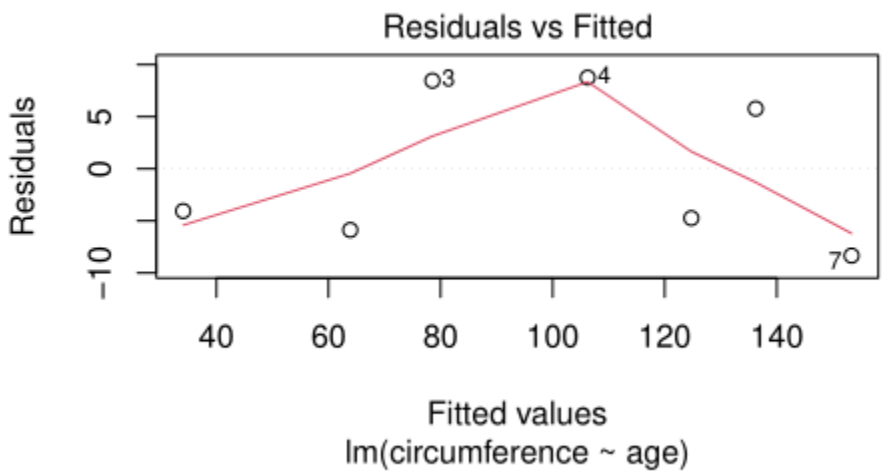
2. Residuals are independent

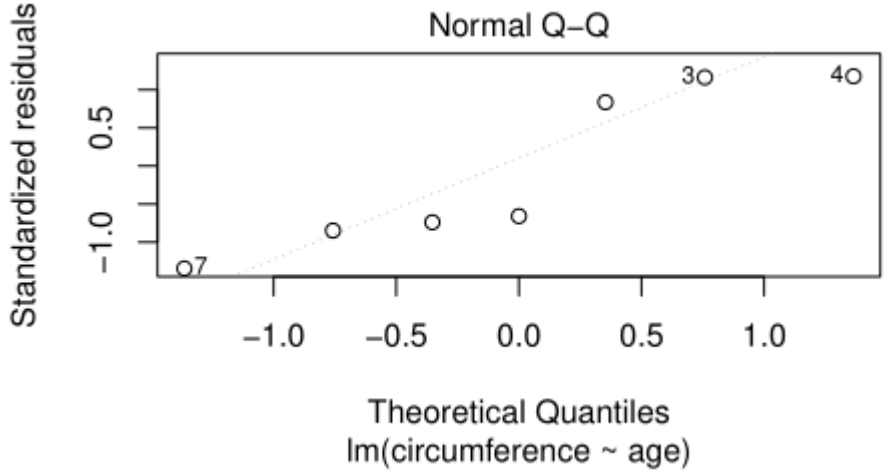
3. Residuals are normally distributed

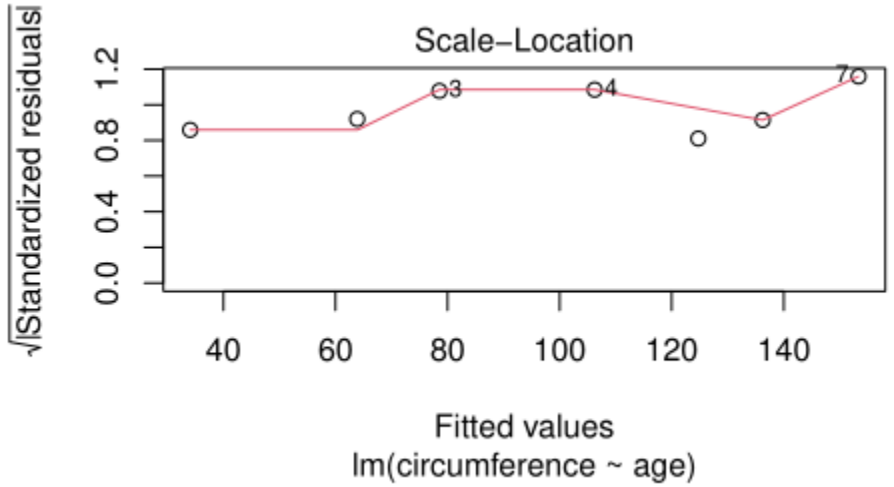
4. Residuals are homoscedastic (i.e., the variance in Y does not increase or decrease as X increases or decreases)

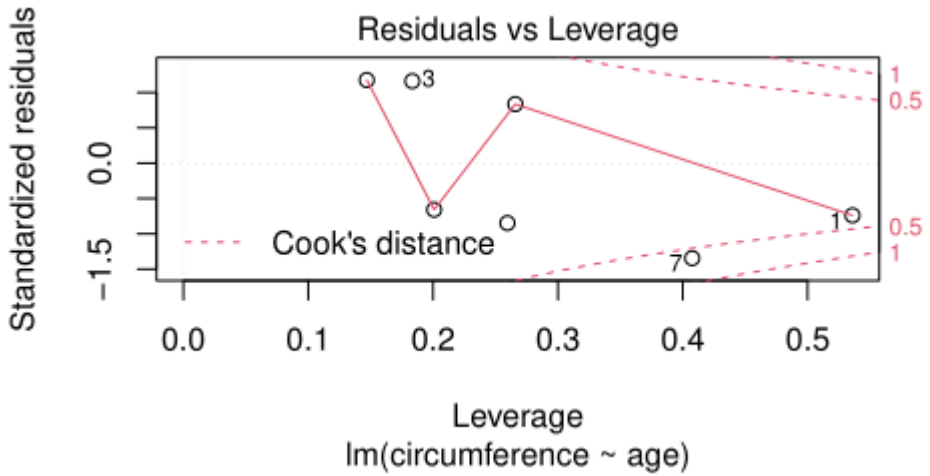
Linear regression in R

```
plot(fit_oranges_1)
```

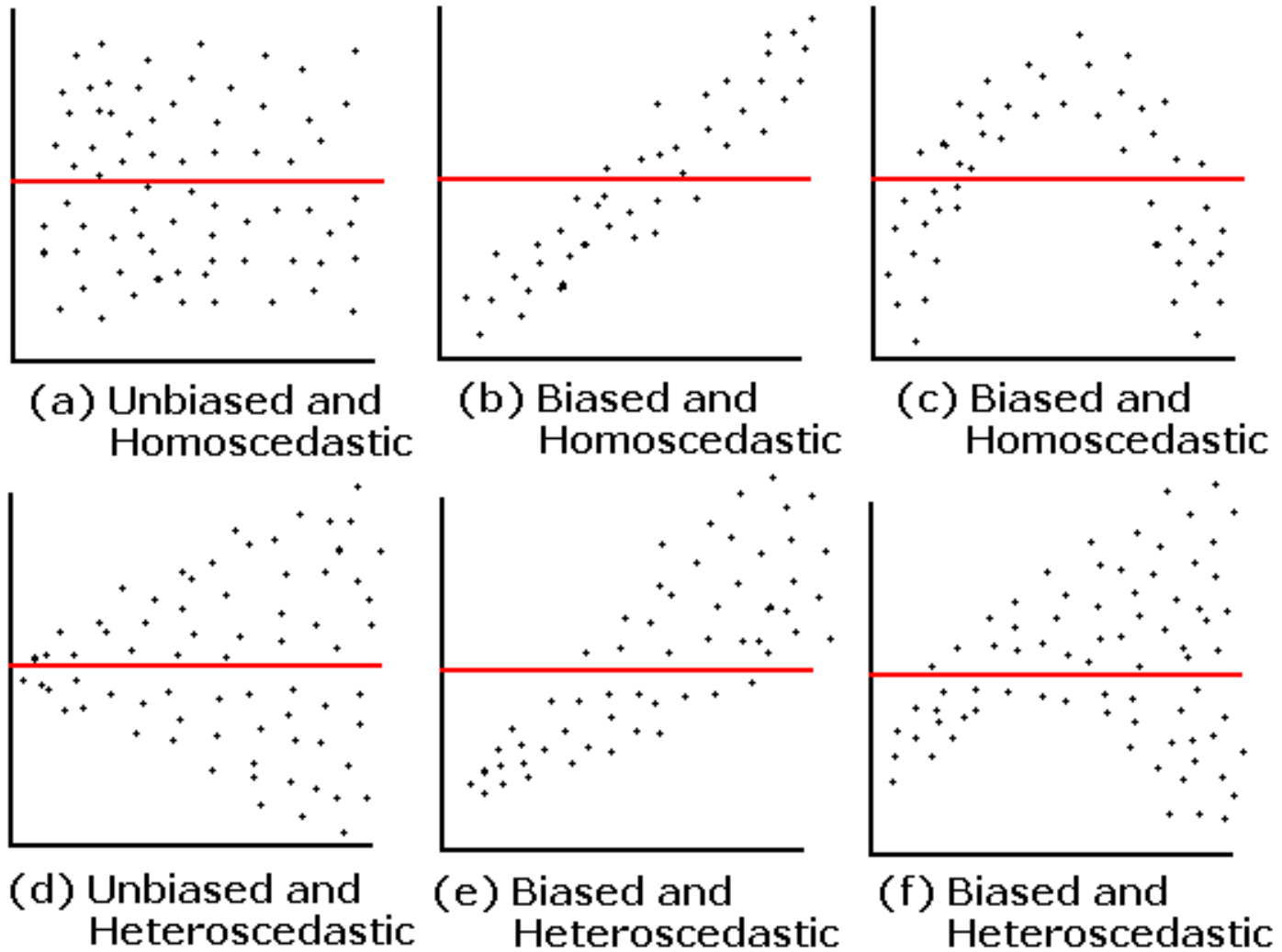








Residual plots



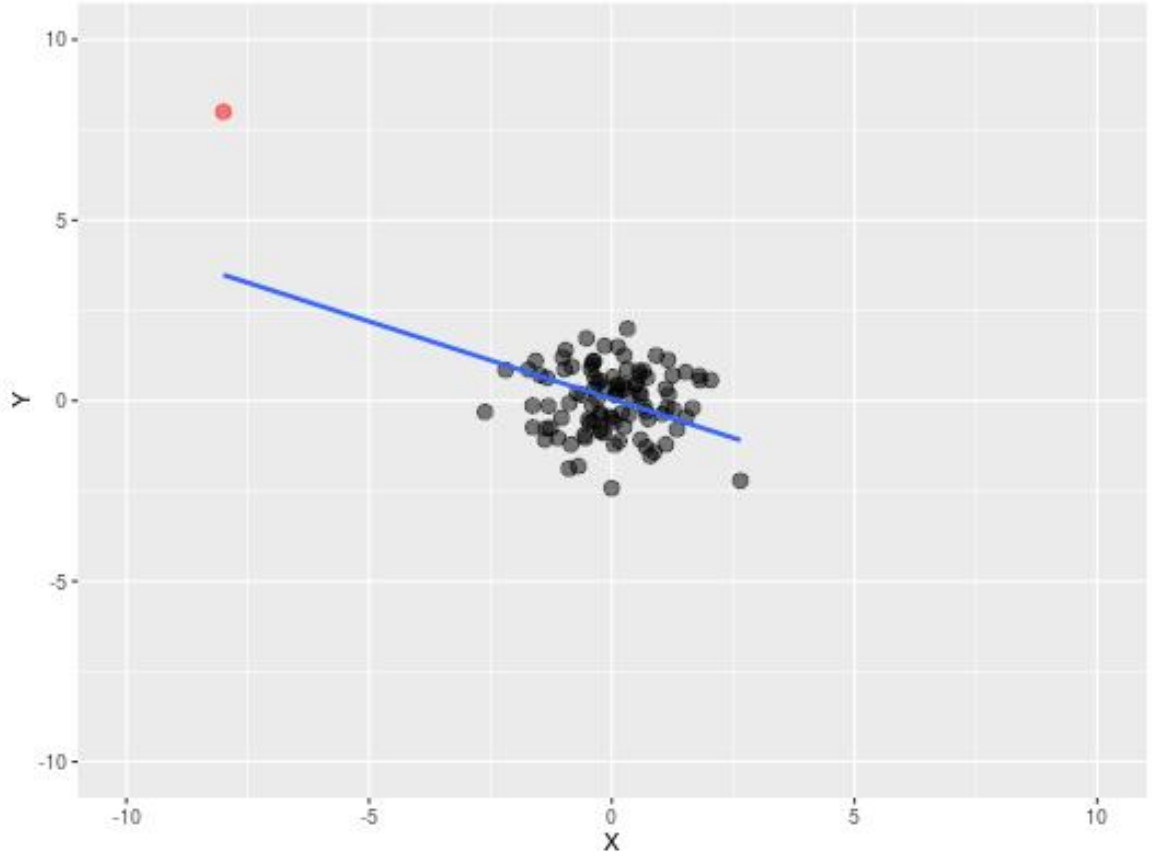
Leverage

Linear regression model coefficients

	Coefficient	Value
1	Intercept	0.04516
2	Slope	-0.42981

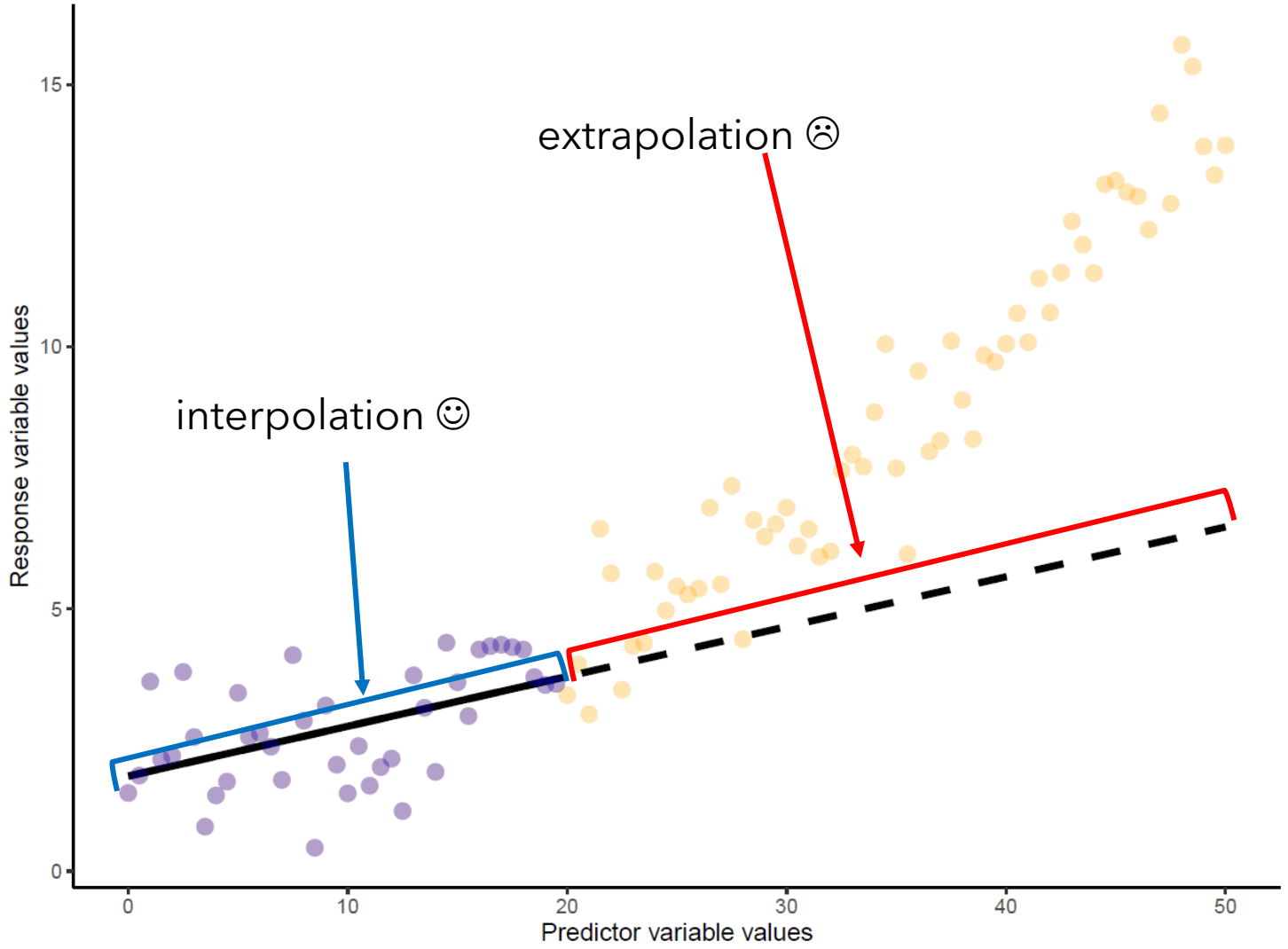
Influence measures for the adjustable point

	Measure	Value
1	hatvalue	0.41424
2	residual	4.51638
3	dfbeta.1	0.50038
4	dfbeta.x	-5.41604



https://www.r-bloggers.com/2016/06/leverage-and-influence-in-a-nutshell/

Interpolation vs. extrapolation



RESEARCH ARTICLE

Ecology and Evolution Open Access WILEY

Evolutionary and plastic variation in larval growth and digestion reveal the complex underpinnings of size and age at maturation in dung beetles

Patrick T. Rohner | Armin P. Moczek

Evolutionary and plastic variation in larval growth and digestion reveal the complex underpinnings of size and age at maturation in dung beetles

Rohner, Patrick T., Indiana University Bloomington, https://orcid.org/0000-0002-9840-1050

Moczek, Armin, Indiana University Bloomington

patrick.t.rohner@gmail.com

Published Feb 16, 2023 on Dryad. https://doi.org/10.5061/dryad.j9kd51cdc

Cite this dataset

Rohner, Patrick T.; Moczek, Armin (2023). Evolutionary and plastic variation in larval growth and digestion reveal the complex underpinnings of size and age at maturation in dung beetles [Dataset]. Dryad. https://doi.org/10.5061/dryad.j9kd51cdc

Abstract

Age and size at maturity are key life history components, yet the proximate underpinnings that mediate intra- and interspecific variation in life history remain poorly understood. We studied the proximate underpinnings of species differences and nutritionally plastic variation in adult size and development time in four species of

	ID	species	nutritional treatment	weight.L1	weight.L2	weight.L3	weight.pupa	weight.adult	weight.peak
2	2	*L. militaris*	high-quality	0.0092	0.0558	0.1443	0.1414	0.0987	0.2645
3	4	*L. militaris*	high-quality	0.0104	0.0409	0.1527	0.1535	0.0992	0.2936
4	5	*L. militaris*	high-quality	0.0083	0.0311	0.1333	0.1172	0.0791	0.2561
5	6	*L. militaris*	high-quality	0.0142	0.0531	0.1529	0.1317	0.0872	0.2986
6	7	*L. militaris*	high-quality	0.0118	0.0318	0.1411	0.1252	0.0875	0.2547
7	8	*L. militaris*	high-quality	0.0089	0.0367	0.1567	0.1365	0.0958	0.296
8	9	*L. militaris*	high-quality	0.0078	0.0312	0.1213	0.1332	0.093	0.2499
9	10	*L. militaris*	high-quality	0.0089	0.0326	0.0835	0.1402	0.0945	0.2421
10	11	*L. militaris*	high-quality	0.0094	0.0381	0.1606	0.1613	0.1053	0.2789
11	12	*L. militaris*	high-quality	0.0076	0.0293	0.1062	0.1147	0.0763	0.2255
12	13	*L. militaris*	high-quality	0.0137	0.0287	0.1198	0.1307	0.0884	0.2796
13	14	*L. militaris*	high-quality	0.0139	0.0423	0.1509	0.1347	0.0925	0.2817
14	15	*L. militaris*	high-quality	0.006	0.0276	0.1212	0.1218	0.0808	0.2445
15	16	*L. militaris*	high-quality	0.0068	0.0851	0.1127	0.1063	0.0728	0.2387
16	17	*L. militaris*	high-quality	0.0091	0.0313	0.1185	0.125	0.0837	0.2538
17	19	*L. militaris*	high-quality	0.0064	0.0315	0.14	0.1356	0.0926	0.2611
18	20	*L. militaris*	high-quality	0.0143	0.0528	0.1119	0.1211	0.0836	0.2512
19	21	*L. militaris*	high-quality	0.0104	0.0502	0.1331	0.1167	0.0797	0.2599
20	22	*L. militaris*	high-quality	0.0091	0.0365	0.1078	0.1355	0.0939	0.2656
21	23	*L. militaris*	high-quality	0.0093	0.0369	0.0886	0.1168	0.0801	0.2234
22	24	*L. militaris*	high-quality	0.0096	0.0601	0.1268	0.1353	0.0932	0.2546
23	25	*L. militaris*	high-quality	0.0068	0.0535	0.1101	0.1258	0.0841	0.2455
24	26	*L. militaris*	high-quality	0.0067	0.0256	0.1499	0.1223	0.0847	0.2128
25	27	*L. militaris*	high-quality	0.0114	0.031	0.1764	0.1619	0.0996	0.2974
26	28	*L. militaris*	high-quality	0.0101	0.0257	0.1113	0.1348	0.0931	0.2726
27	29	*L. militaris*	high-quality	0.0094	0.0364	0.1443	0.1345	0.093	0.2786
28	31	*L. militaris*	high-quality	0.0071	0.0308	0.1393	0.1322	0.0909	0.2701

Metadata provided with the dataset!

1	**sheet: weight by age and stage**	
2	ID	unique identifier for each individual
3	nutritional treatment	high-quality = cow poop from grass-fed cows; low-quality = cow poop from hay-fed cows
4	species	species identity
5	weight.L1	larval weight at the onset of the first larval instar (L1) [g]
6	weight.L2	larval weight at the onset of the second larval instar (L2) [g]
7	weight.L3	larval weight at the onset of the third larval instar (L3) [g]
8	weight.pupa	pupal weight [g]
9	weight.adult	adult weight [g]
10	weight.peak	maximum weight in L3 [g]
11	age.L1	age at the onset of the first larval instar (L1) [days]
12	age.L2	age at the onset of the second larval instar (L2) [days]
13	age.L3	age at the onset of the third larval instar (L3) [days]
14	age.pupa	age at pupation [days]
15	age.peak	age when larvae reached their peak weight [days]
16	age.adult	age at adult emergence [days]
17	pronotum width	adult pronotum width [mm]
18	FA.tibia.shape	fluctuation asymmetry score in tibia shape
19	FA.tibia.CS	fluctuation asymmetry score in tibia centroid size
20	FA.tibia.length	fluctuation asymmetry score in tibia length
21	instantaneous.growth.rate	instantaneous growth rates
22	missing data indicated wit "n/a"	
23		
24	**sheet: weight by age**	
25	ID	unique identifier for each individual
26	nutritional treatment	high-quality = cow poop from grass-fed cows; low-quality = cow poop from hay-fed cows
27	species	species identity
28	mass [g]	weight measurement at the given age

weight by age | weight and age by stage | **variable description** ⊕

```
> summary(growth)
       ID                 species    nutritional.treatment
 Min.   :  2.0   D. gazella  :48   high-quality:88
 1st Qu.: 59.0   L. militaris:50   low-quality :87
 Median :122.0   O. binodis  :36
 Mean   :138.3   O. taurus   :41
 3rd Qu.:219.5
 Max.   :289.0
   weight.L3       weight.pupa        weight.adult       we
 Min.   :0.0305   Length:175         Length:175         Le
 1st Qu.:0.0899   Class :character   Class :character   Cl
 Median :0.1270   Mode  :character   Mode  :character   Mo
 Mean   :0.1240
 3rd Qu.:0.1526
 Max.   :0.2560
     age.L2          age.L3         age.pupa           age
 Min.   :1.000   Min.   : 3.000   Length:175         Lengt
 1st Qu.:2.000   1st Qu.: 5.000   Class :character   Class
 Median :3.000   Median : 7.000   Mode  :character   Mode
```

Liatongus militaris

Genus *Liatongus*

Length 8–10 mm

Colour

brown to dark brown; distinc- tive black broken stripes on wing co- vers. Yellow ‘shoulder patches’ and a dark oval patch on the top and bot- tom of each femur.

Horns none

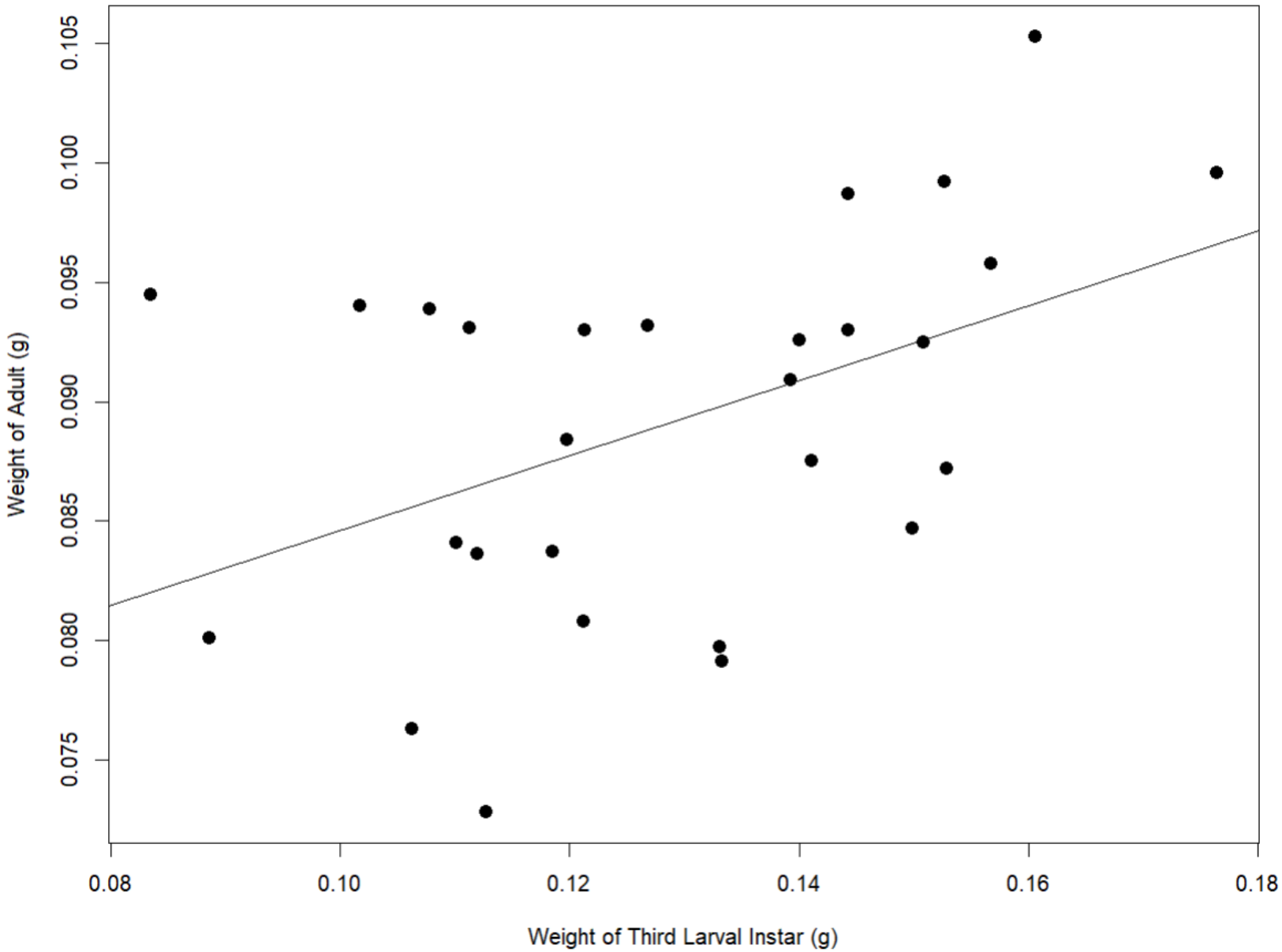
Flight time day

Yearly Activity spring to autumn

Distribution QLD, NT, northeast NSW

Similar Species Unlikely to be confused with other species.

	ID	species	nutritional treatment	weight.L1	weight.L2	weight.L3	weight.pupa	weight.adult	weight.peak
2	2	*L. militaris*	high-quality	0.0092	0.0558	0.1443	0.1414	0.0987	0.2645
3	4	*L. militaris*	high-quality	0.0104	0.0409	0.1527	0.1535	0.0992	0.2936
4	5	*L. militaris*	high-quality	0.0083	0.0311	0.1333	0.1172	0.0791	0.2561
5	6	*L. militaris*	high-quality	0.0142	0.0531	0.1529	0.1317	0.0872	0.2986
6	7	*L. militaris*	high-quality	0.0118	0.0318	0.1411	0.1252	0.0875	0.2547
7	8	*L. militaris*	high-quality	0.0089	0.0367	0.1567	0.1365	0.0958	0.296
8	9	*L. militaris*	high-quality	0.0078	0.0312	0.1213	0.1332	0.093	0.2499
9	10	*L. militaris*	high-quality	0.0089	0.0326	0.0835	0.1402	0.0945	0.2421
10	11	*L. militaris*	high-quality	0.0094	0.0381	0.1606	0.1613	0.1053	0.2789
11	12	*L. militaris*	high-quality	0.0076	0.0293	0.1062	0.1147	0.0763	0.2255
12	13	*L. militaris*	high-quality	0.0137	0.0287	0.1198	0.1307	0.0884	0.2796
13	14	*L. militaris*	high-quality	0.0139	0.0423	0.1509	0.1347	0.0925	0.2817
14	15	*L. militaris*	high-quality	0.006	0.0276	0.1212	0.1218	0.0808	0.2445
15	16	*L. militaris*	high-quality	0.0068	0.0851	0.1127	0.1063	0.0728	0.2387
16	17	*L. militaris*	high-quality	0.0091	0.0313	0.1185	0.125	0.0837	0.2538
17	19	*L. militaris*	high-quality	0.0064	0.0315	0.14	0.1356	0.0926	0.2611
18	20	*L. militaris*	high-quality	0.0143	0.0528	0.1119	0.1211	0.0836	0.2512
19	21	*L. militaris*	high-quality	0.0104	0.0502	0.1331	0.1167	0.0797	0.2599
20	22	*L. militaris*	high-quality	0.0091	0.0365	0.1078	0.1355	0.0939	0.2656
21	23	*L. militaris*	high-quality	0.0093	0.0369	0.0886	0.1168	0.0801	0.2234
22	24	*L. militaris*	high-quality	0.0096	0.0601	0.1268	0.1353	0.0932	0.2546
23	25	*L. militaris*	high-quality	0.0068	0.0535	0.1101	0.1258	0.0841	0.2455
24	26	*L. militaris*	high-quality	0.0067	0.0256	0.1499	0.1223	0.0847	0.2128
25	27	*L. militaris*	high-quality	0.0114	0.031	0.1764	0.1619	0.0996	0.2974
26	28	*L. militaris*	high-quality	0.0101	0.0257	0.1113	0.1348	0.0931	0.2726
27	29	*L. militaris*	high-quality	0.0094	0.0364	0.1443	0.1345	0.093	0.2786
28	31	*L. militaris*	high-quality	0.0071	0.0308	0.1393	0.1322	0.0909	0.2701

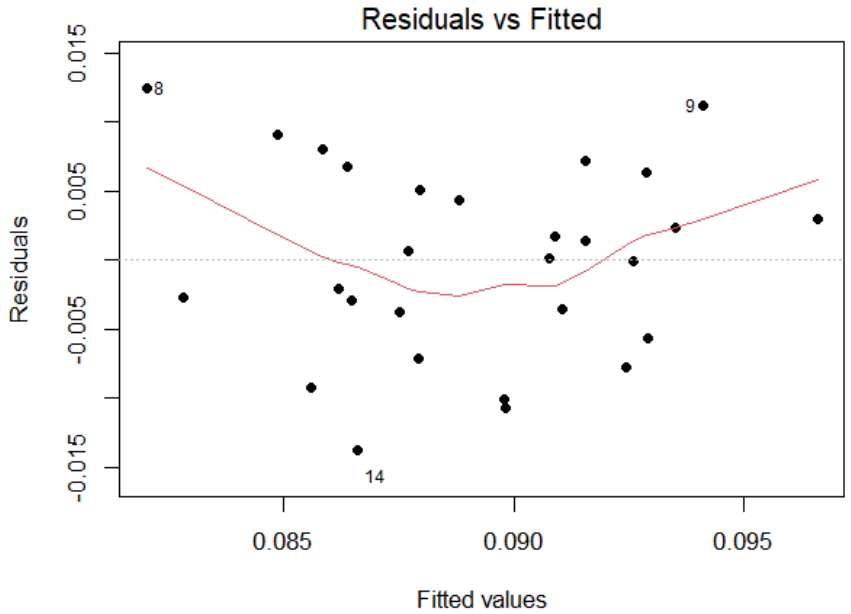

Weight of Adult (g)
0.075
0.080
0.085
0.090
0.095
0.100
0.105
0.08
0.10
0.12
0.14
0.16
0.18
Weight of Third Larval Instar (g)

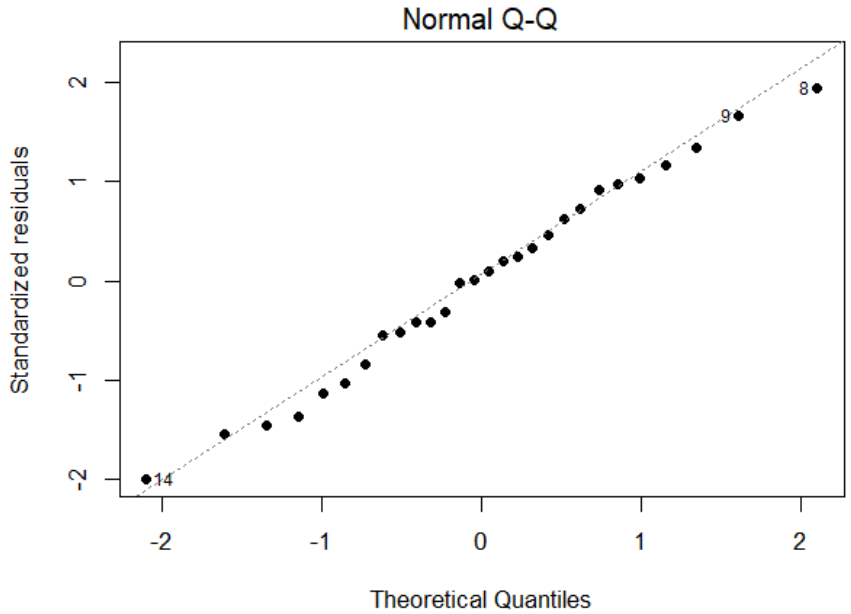
```
> summary(beetle_mod)

Call:
lm(formula = weight.adult ~ weight.L3, data = beetle_high)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0138056 -0.0042889  0.0004019  0.0053646  0.0124737

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.068932   0.007904   8.721 3.38e-09 ***
weight.L3   0.156824   0.060297   2.601   0.0151 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.00711 on 26 degrees of freedom
Multiple R-squared:  0.2065,    Adjusted R-squared:  0.1759
F-statistic: 6.764 on 1 and 26 DF,  p-value: 0.01514

> anova(beetle_mod)
Analysis of Variance Table

Response: weight.adult
          Df     Sum Sq    Mean Sq F value  Pr(>F)
weight.L3  1 0.00034194 0.00034194  6.7645 0.01514 *
Residuals 26 0.00131429 0.00005055
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

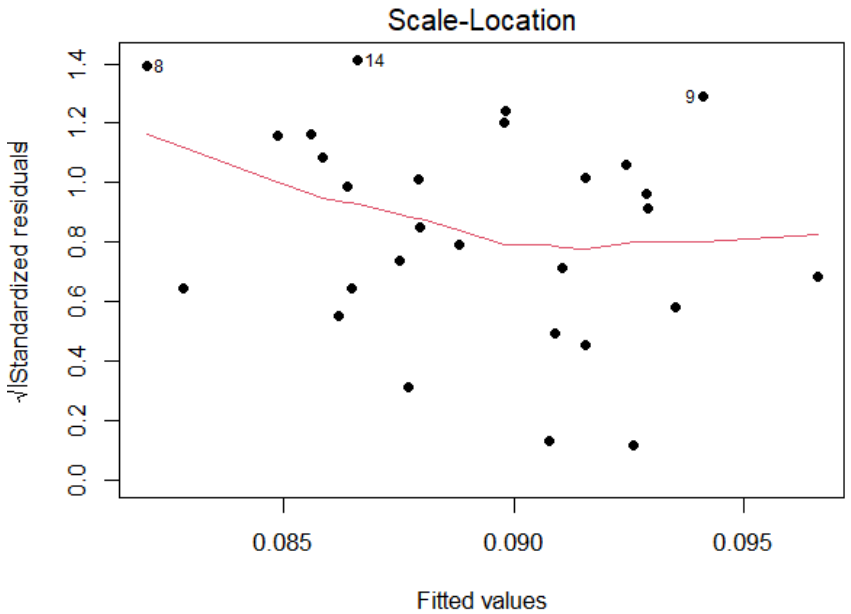
e_i → Residuals

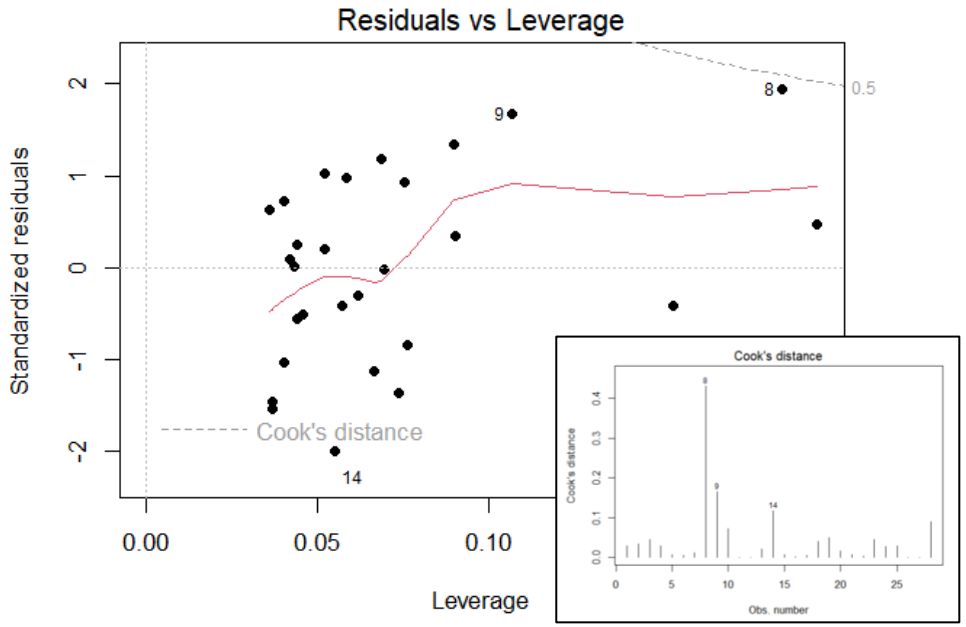
β_0 → (Intercept)

β_1 → weight.L3


Residuals vs Fitted
Residuals
Fitted values
0.085
0.090
0.095
0.015
0.005
-0.005
-0.015
8
9
14


Normal Q-Q
Standardized residuals
Theoretical Quantiles
-2
-1
0
1
2
8
9
14


Scale-Location
√|Standardized residuals|
Fitted values
0.085
0.090
0.095
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
8
14
9


Residuals vs Leverage
Standardized residuals
Leverage
0.00
0.05
0.10
-2
-1
0
1
2
8
9
14
0.5
Cook's distance
Cook's distance
Obs. number