JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Example

Growth of Hudson River striped bass; years 1965–1976.

y_t = incremental growth rate in year t

x_{1t} = population density in year t

x_{2t} = rise in temperature in year t

x_{3t} = a measure of river flow in year t

Fit the model

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \epsilon_t \quad \text{where } \epsilon_t \sim \text{ iid Normal}(0, \sigma^2)$$

Plot y vs each x
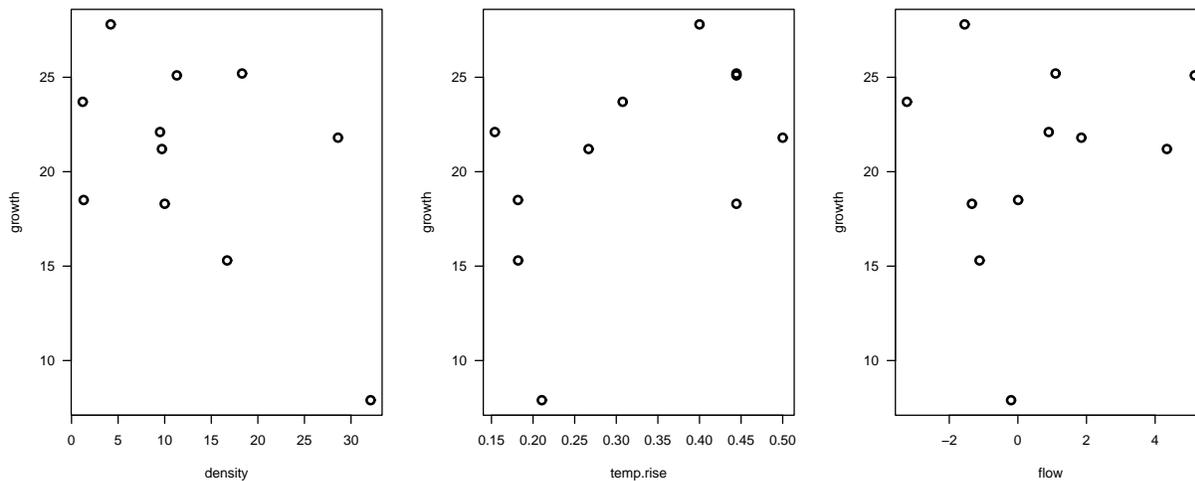
Results

	Est	SE	t-value	P value
Intercept	17.00	3.22	5.3	0.001
density	–0.37	0.11	–3.2	0.015
temp.rise	25.36	9.01	2.8	0.026
flow	0.45	0.47	1.0	0.362

Diagnostics

Assumptions	Diagnostics
ϵ's independent	Plot residuals vs time
ϵ's normally distributed	QQ plot of residuals
ϵ's have constant SD	Plot residuals vs fitted values
y's linear in each of the x's	Plot residuals vs each x
No other x's belong in the model	Plot residuals vs other x's

residuals vs time
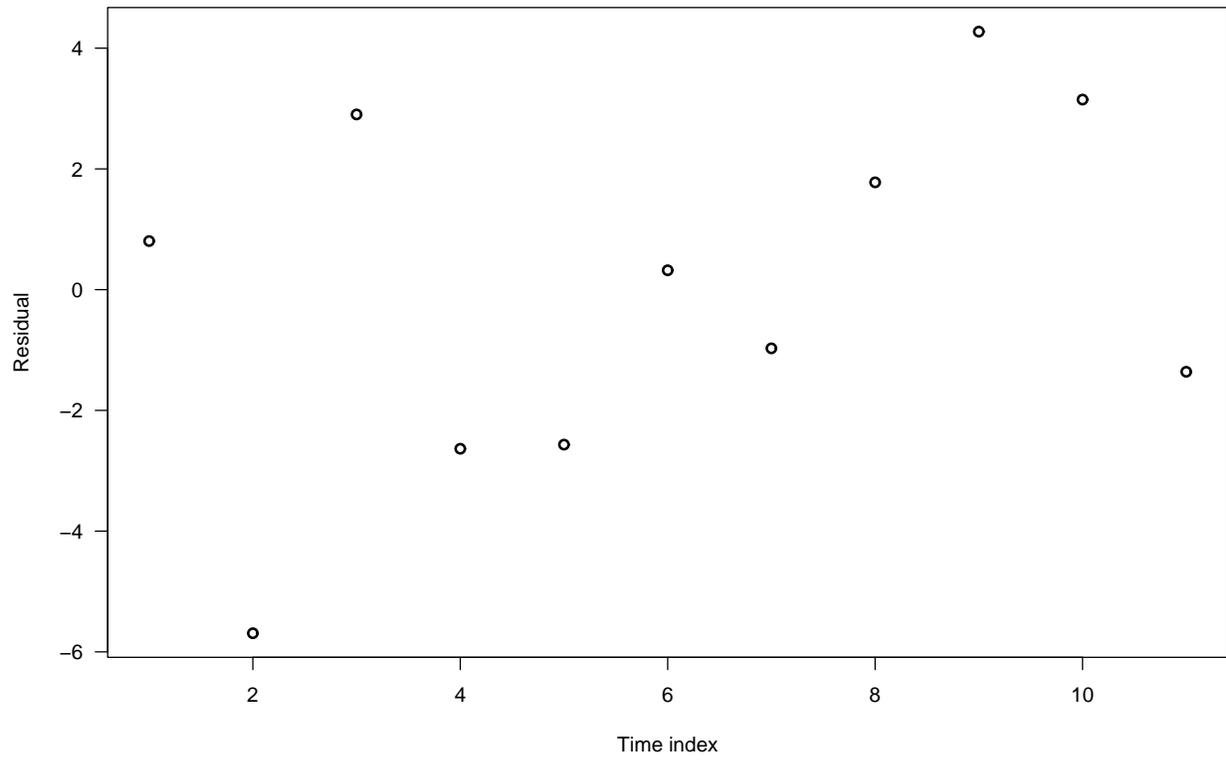
residual(t) vs residual(t+1)

corr = −0.09

Permutation test

P−value = 78%

correlation

QQ plot of residuals

Sample Quantiles
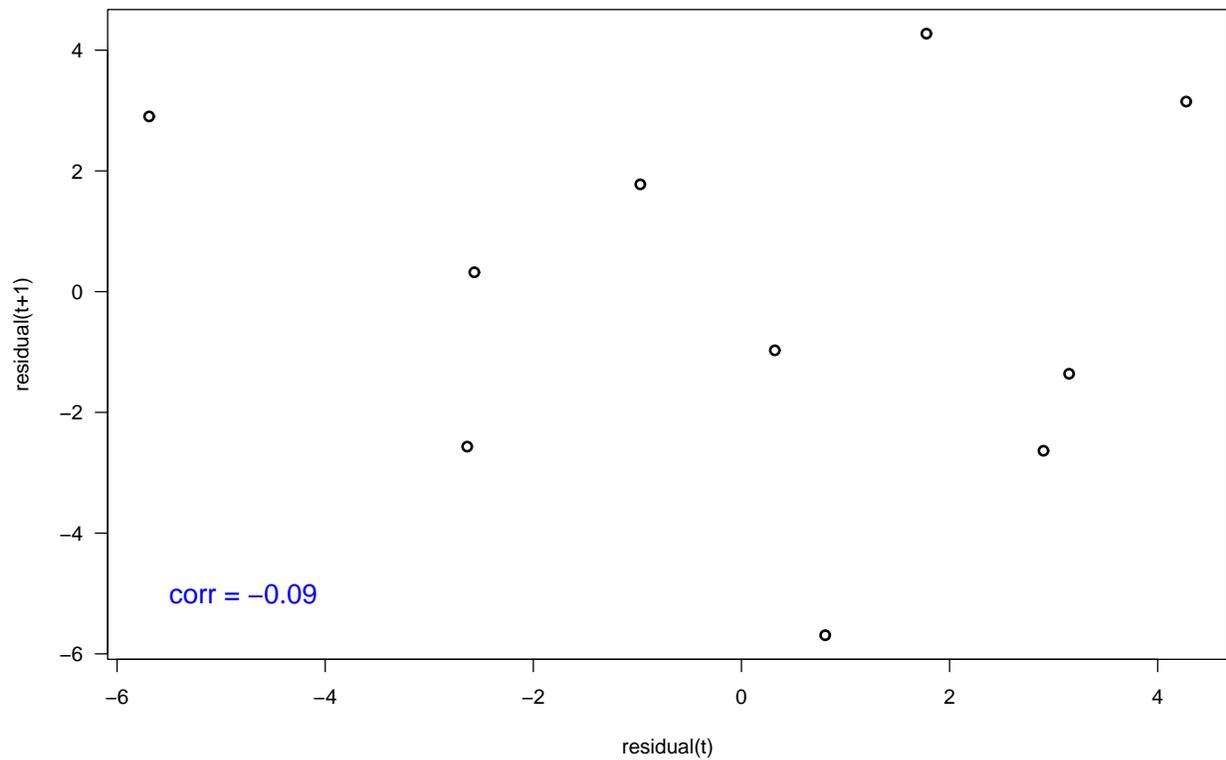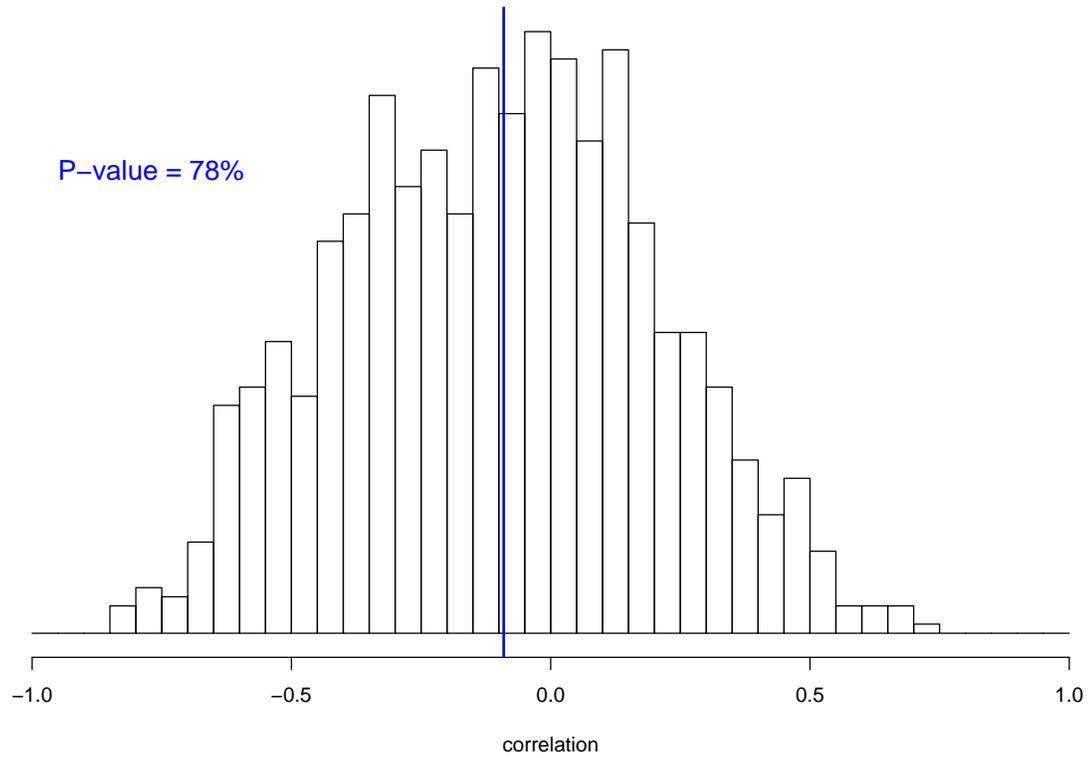
Theoretical Quantiles
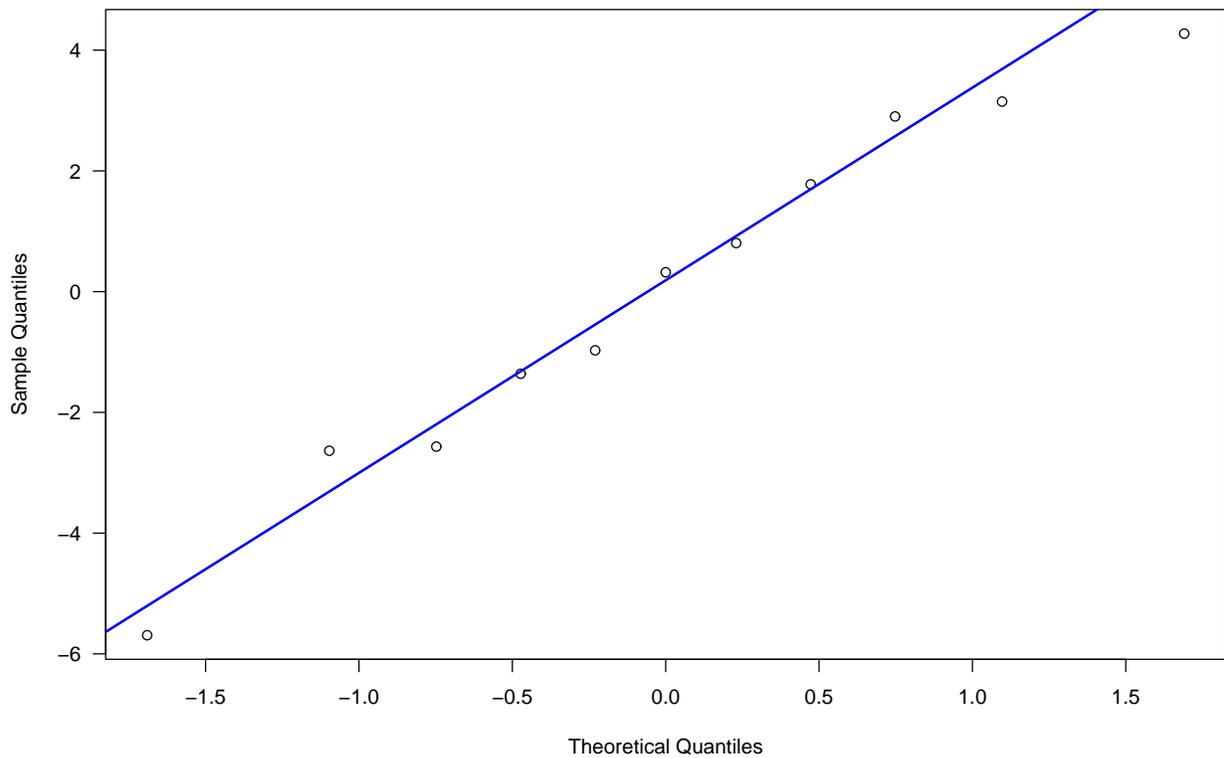
Residuals vs fitted values
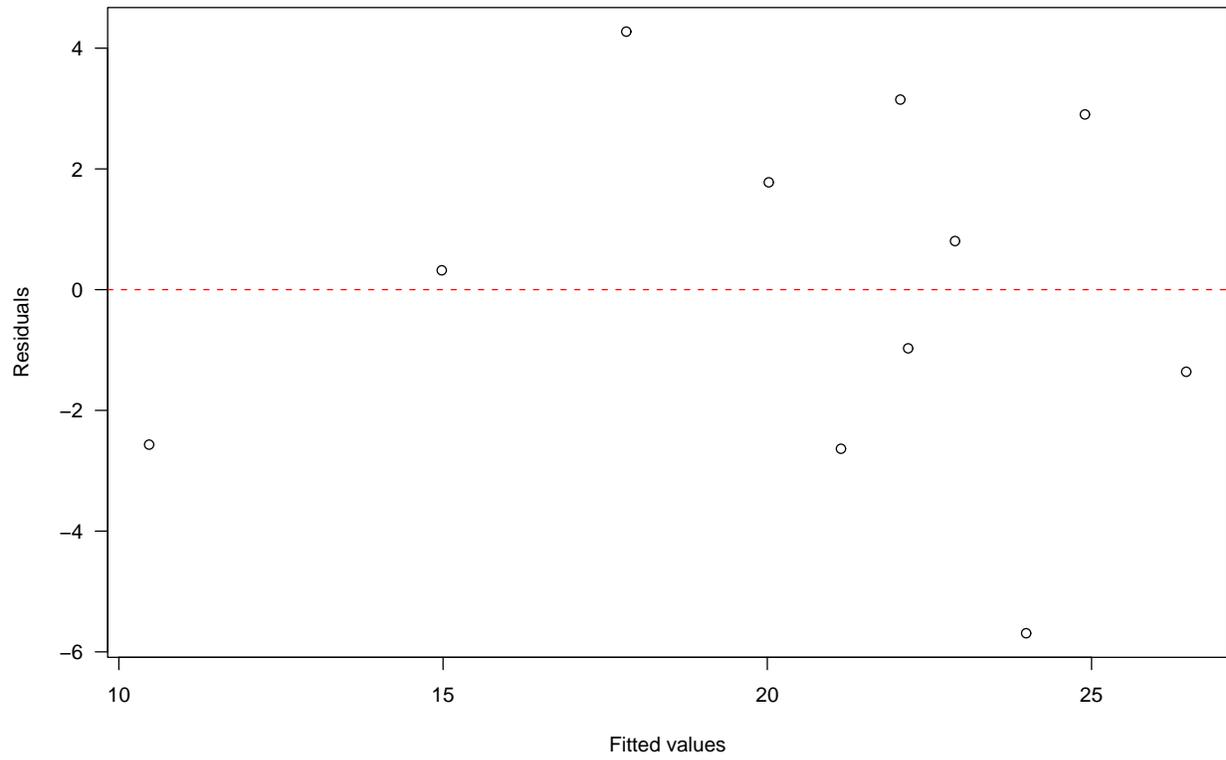
Residuals vs x's
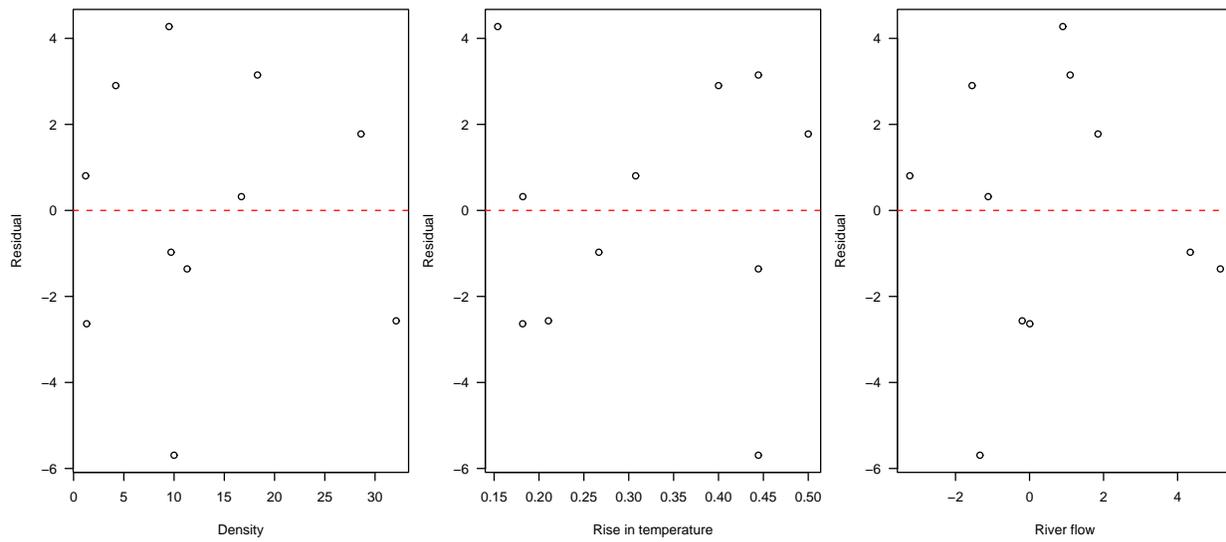
A fake example
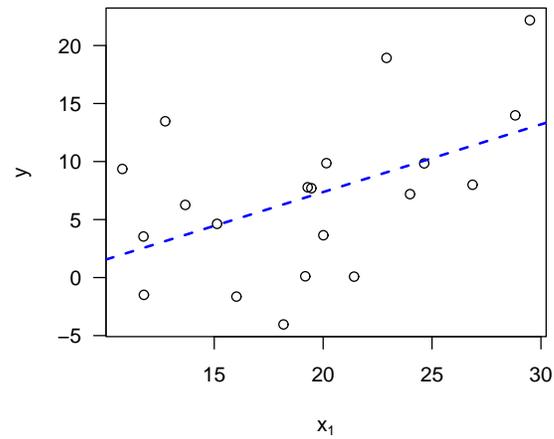
	Est	SE	t-val	P-val
Intercept	-4.29	4.97	-0.86	0.40
x1	0.58	0.25	2.36	0.03

Residuals vs another x
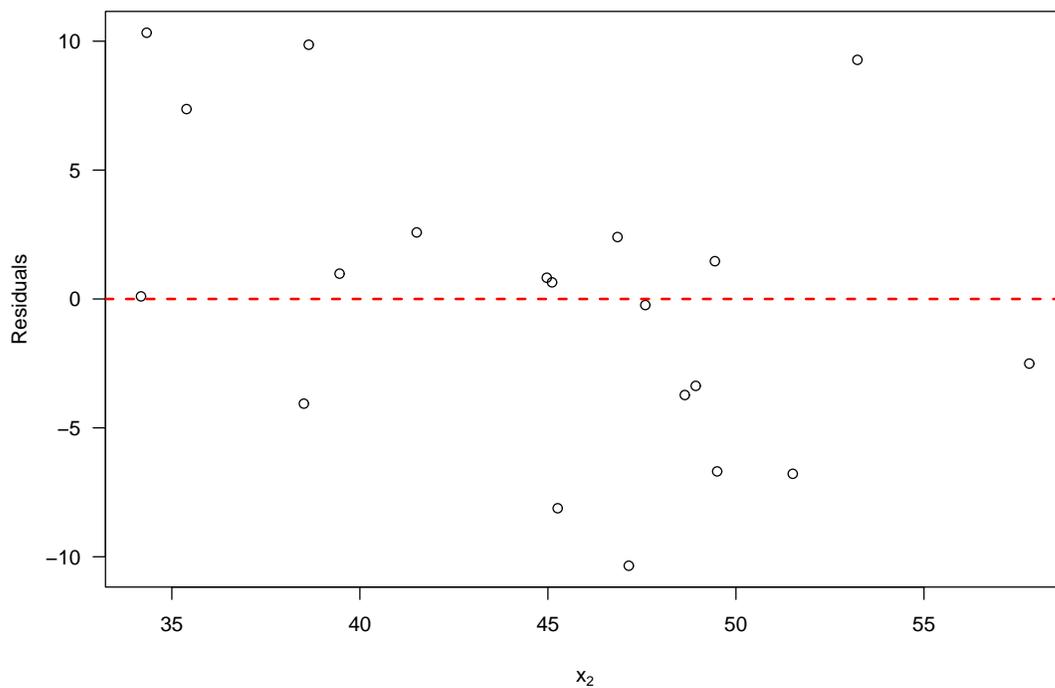
Regress y on both x's

	Est	SE	t-val	P-val
Intercept	18.02	8.50	2.12	0.0491
x1	1.21	0.29	4.13	0.0007
x2	-0.77	0.26	-3.00	0.0080

One last example

Sediment ingestion by the mud snail, *Hyrobia minuta*.

y = Amount ingested

x = Time allowed to eat
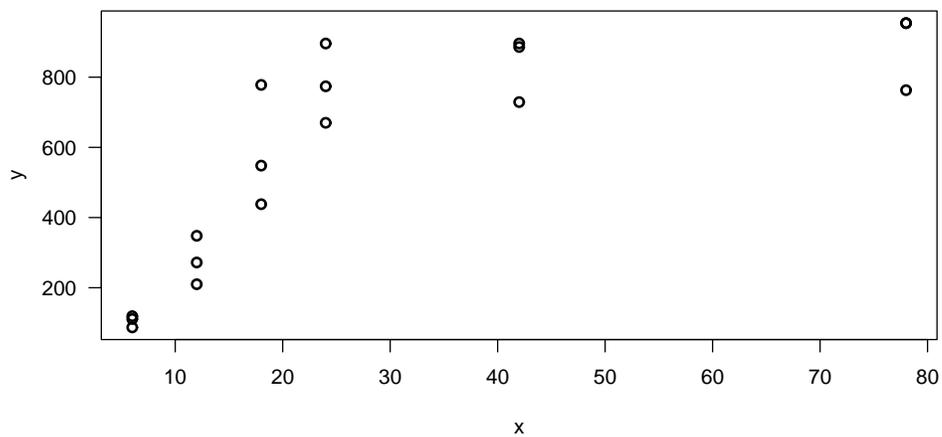
A model

Let's consider the model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_i \quad \text{where } \epsilon_t \sim \text{ iid Normal}(0, \sigma^2)$$
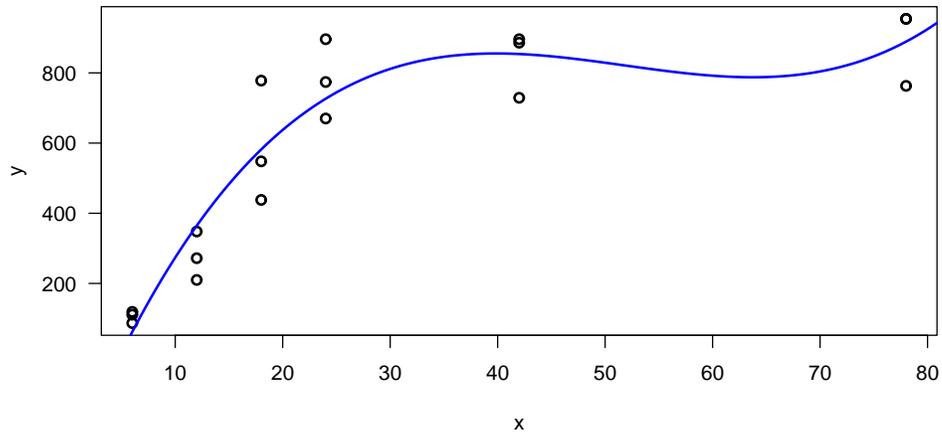
Estimated coefficients

	Est	SE	t-val	P-val
Intercept	-339	127	-2.66	0.019
time	75.7	15.4	4.91	<0.001
time^2	-1.55	0.48	-3.22	0.006
time^3	0.010	0.004	2.52	0.024

Diagnostic plots