Lesson 6
Maximum Likelihood: Part III

(plus a quick bestiary of models)

Homework

- You want to know the density of fish in a set of experimental ponds
- You observe the following counts in ten ponds: 5,6,7,3,6,5,8,4,4,3
- What is your process model?
- What is your data model?
- Solve for the analytical MLE
- What is the estimate for this population?

- Process model: f(x) = λ (density)
- Data model: x ~ Pois(λ)

$$L = \prod Pois(x_i \mid \lambda)$$

$$L \propto \prod \lambda^{x_i} e^{-\lambda}$$

$$lnL \propto \sum x_i ln(\lambda) - \sum \lambda$$

$$\frac{dlnL}{d\lambda} = \frac{1}{\lambda} \sum x_i - n = 0$$

$$\lambda = \frac{1}{n} \sum x_i = \bar{x}$$

How the likelihood is constructed

- L = Pr(data|model parameters)
- How is the data modeled?
  - What type of data?
  - What process generated this data?
  - What distributions are an appropriate description of the data?
- How is the process modeled?
- Each analysis should be approached individually
  - Problem solving, creativity

How is the data modeled?

- What type of data is it:
  - Continuous
  - Integer / Count
  - Boolean (0/1)
  - Factor / categorical
- Are there range restrictions on the data?
  - Are negative values allowed?
  - Is there an upper bound?
  - Are the observed data near the bounds?

How is the data modeled?

- What are the dominant sources of variability in the data?
  - Observation/measurement error
  - Process variability
    - Space
    - Time
    - Individual/Site/Species
    - Random
  - Missing data

How is the data modeled?

- Are there multiple processes involved?
  - Zero-inflated data
    - Pr(abundance|present)Pr(present)
- Are there multiple types or sources of data?
  - Tree growth: tree rings + DBH
  - Tree fecundity: cone counts + seed trap
  - Tree crown: remote sensing + model + crown class
- Is the process observed directly or inferred?

How is the process modeled?

- Constant mean
- Multiple means by factor (ANOVA)
- As a function of covariates
  - Linear models
  - Generalized linear models
  - Nonlinear models
- Hierarchical models
- Mechanistic models

Note: Will shy away from “data mining” models except for EDA: regression trees, splines, neural networks, etc.

A quick beastiary of functions

- Polynomials
- Piecewise polynomials
- Rational (ratio based)
- Exponential based
- Power-based
- Sometimes chosen for mechanistic reasons, sometimes because they “fit right”

Supplemental reading: Bolker Ch 3

Polynomial

- *Linear* with respect to the model parameters
- Taylor series: Can approximate any smooth continuous function

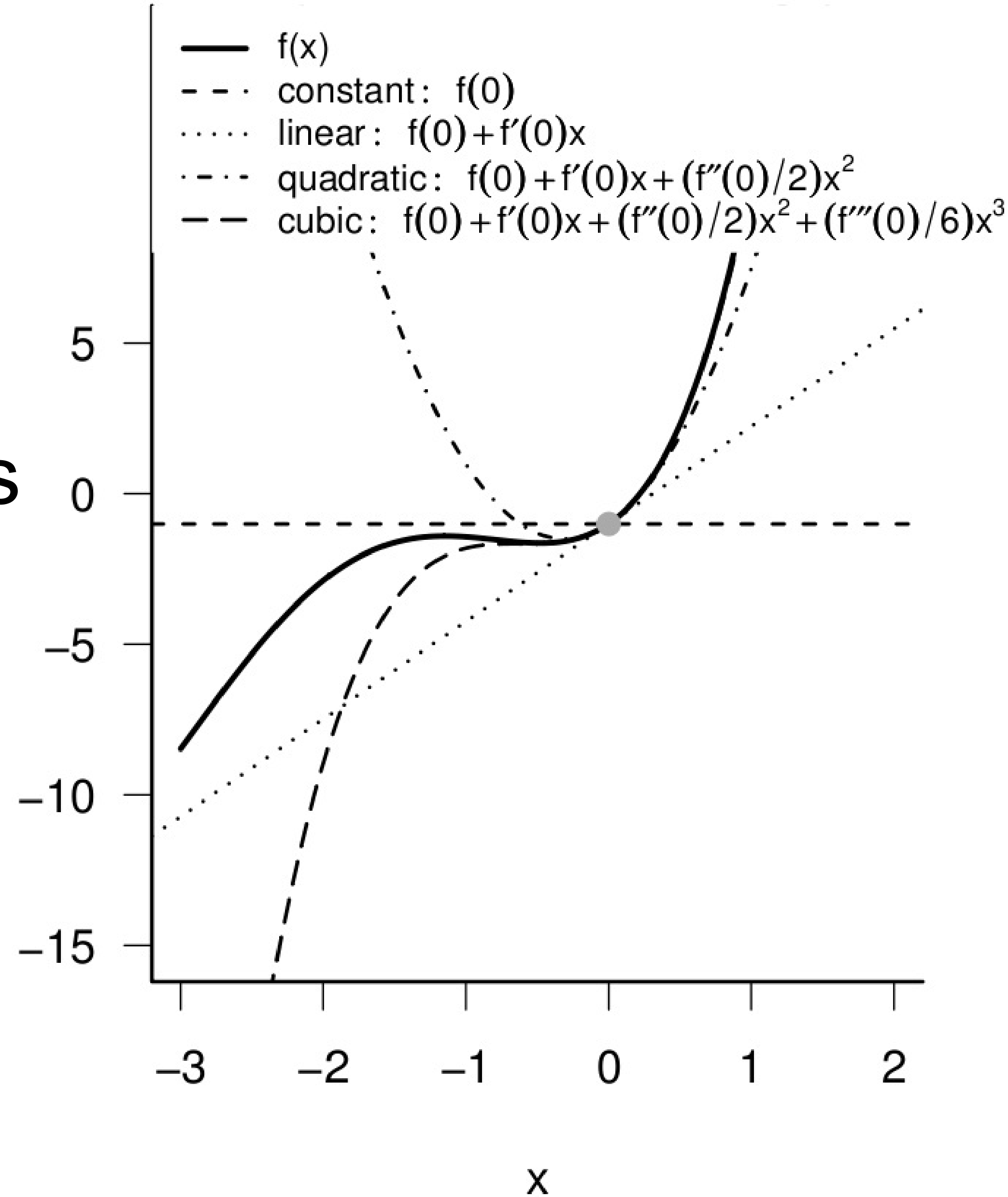


Figure 6: Taylor series expansion of a 4th-order polynomial.

Piecewise

- AKA change point analysis

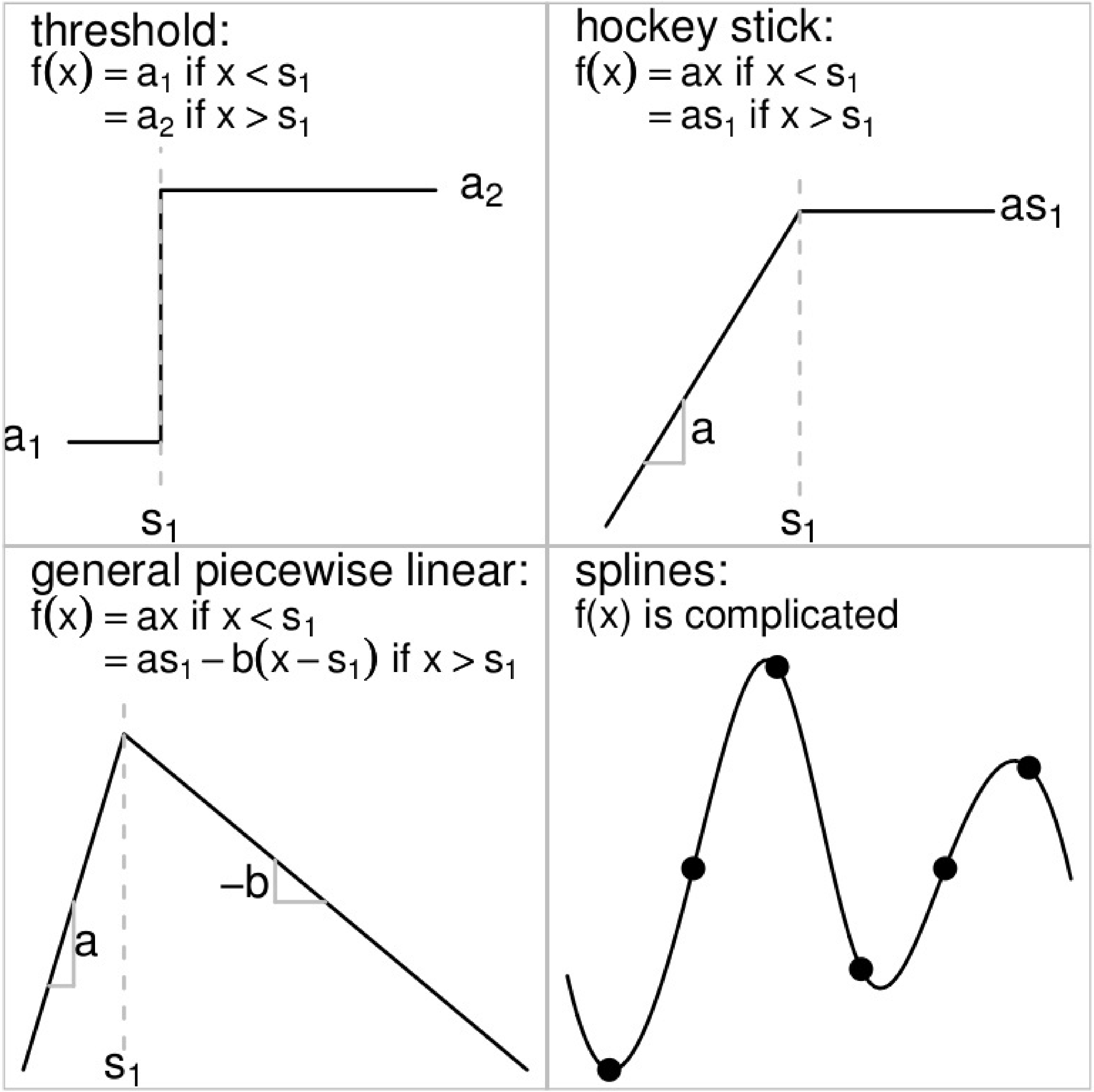


Figure 7: Piecewise polynomial functions: the first three (threshold, hockey stick, general piecewise linear) are all piecewise linear. Splines are piecewise cubic; the equations are complicated and usually handled by software (see `?spline` and `?smooth.spline`).

Rational

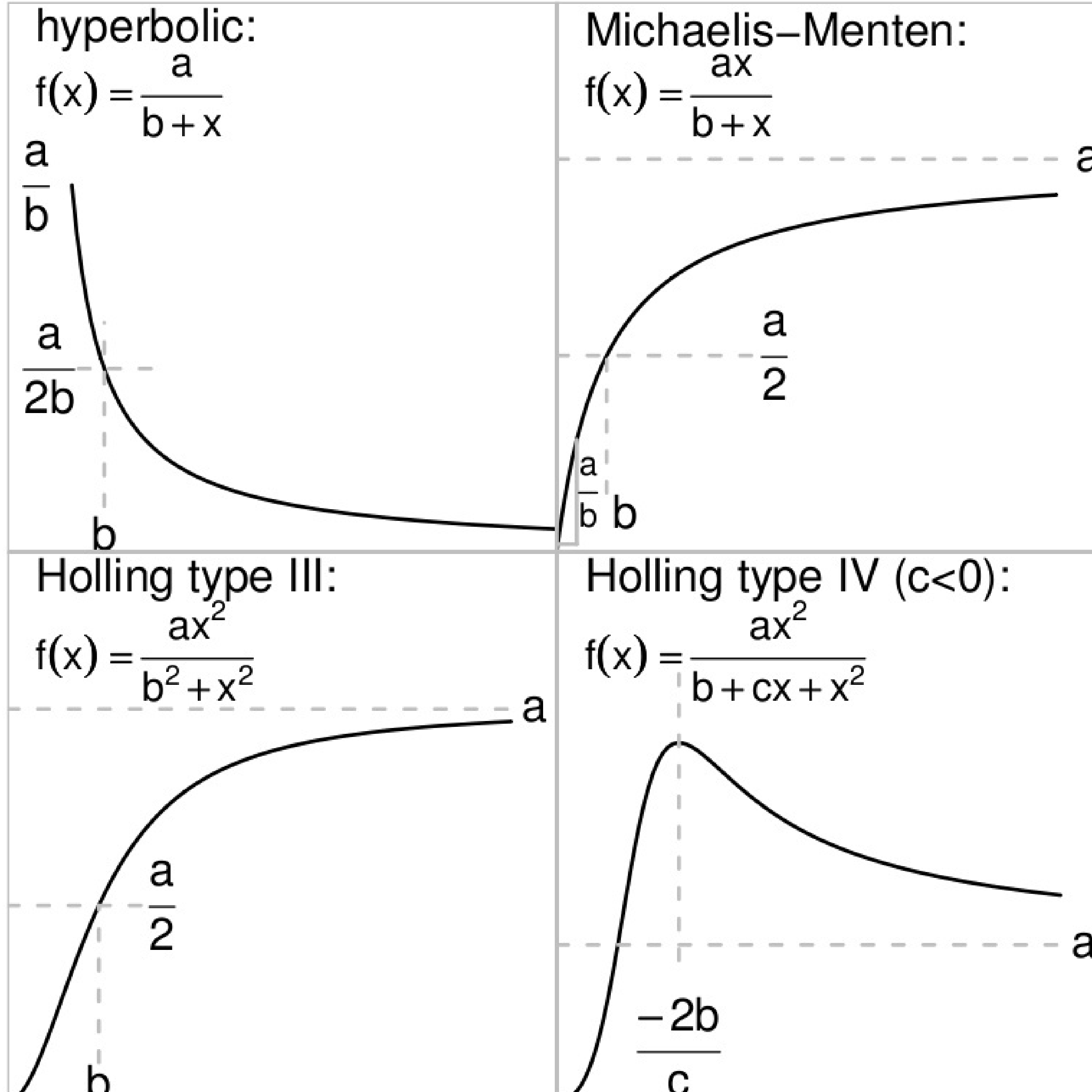
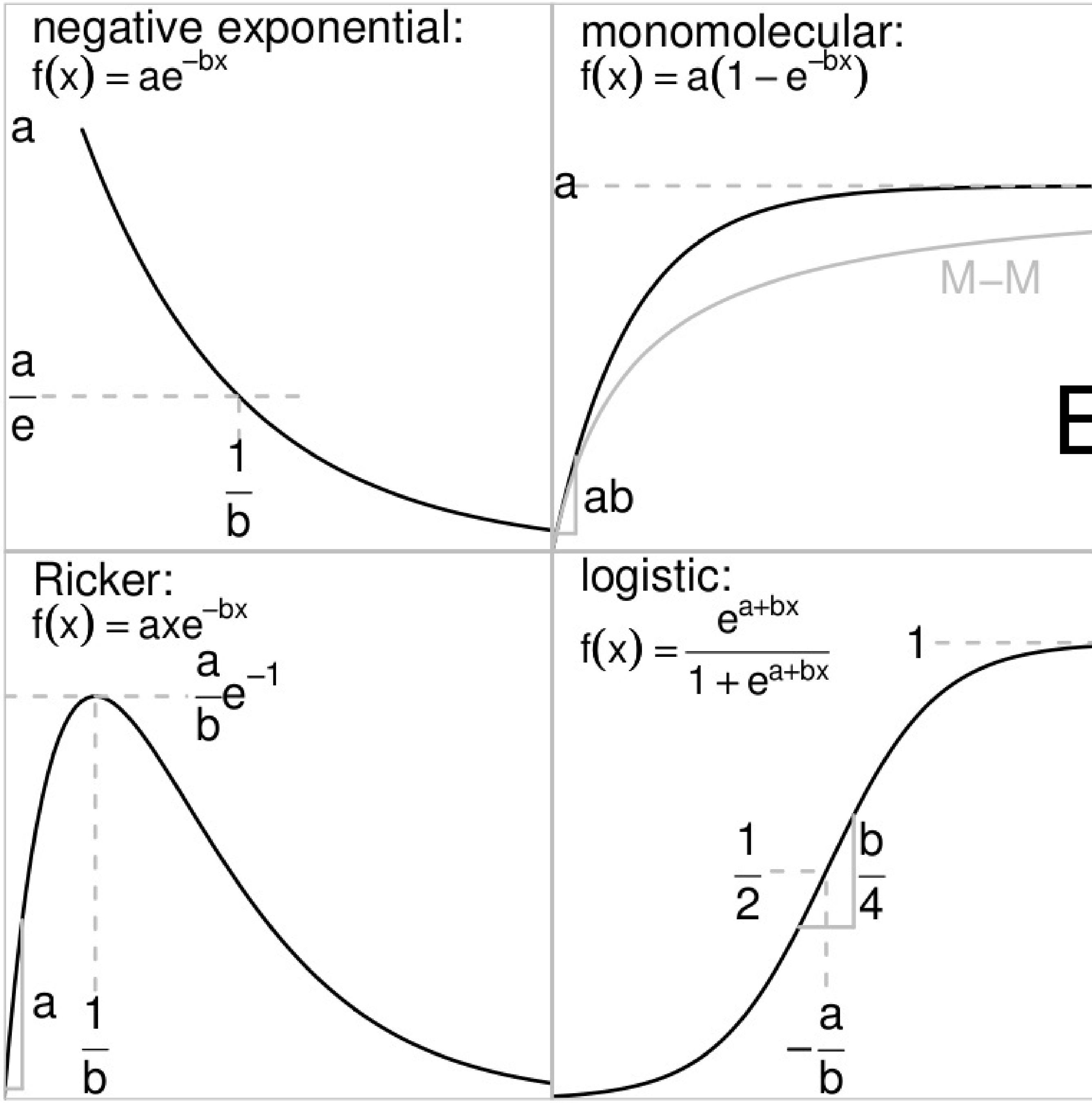


Figure 8: Rational functions.



Exponential

Figure 9: Exponential-based functions. "M-M" in the monomolecular figure is the Michaelis-Menten function with the same asymptote and initial slope.

Power

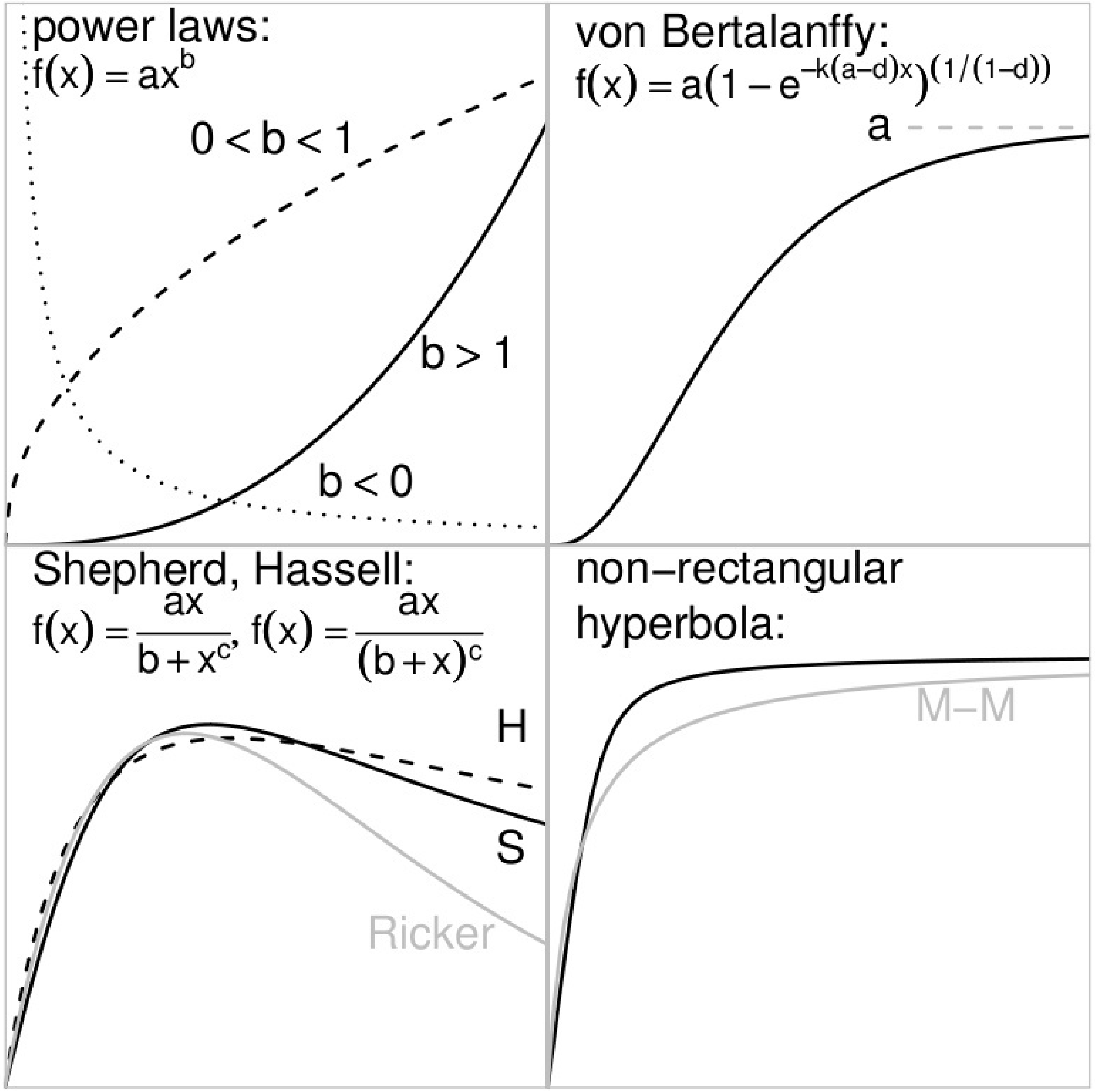


Figure 10: Power-based functions. The lower left panel shows the Ricker function for comparison with the Shepherd and Hassell functions. The lower right shows the Michaelis-Menten function for comparison with the non-rectangular hyperbola.

Function	Range	Left end	Right end	Middle
Polynomials				
Line	$\{-\infty, \infty\}$	$y \to \pm\infty$, constant slope	$y \to \pm\infty$, constant slope	monotonic
Quadratic	$\{-\infty, \infty\}$	$y \to \pm\infty$, accelerating	$y \to \pm\infty$, accelerating	single max/min
Cubic	$\{-\infty, \infty\}$	$y \to \pm\infty$, accelerating	$y \to \pm\infty$, accelerating	up to 2 max/min
Piecewise polynomials				
Threshold	$\{-\infty, \infty\}$	flat	flat	breakpoint
Hockey stick	$\{-\infty, \infty\}$	flat or linear	flat or linear	breakpoint
Piecewise linear	$\{-\infty, \infty\}$	linear	linear	breakpoint
Rational				
Hyperbolic	$\{0, \infty\}$	$y \to \infty$ or finite	$y \to 0$	decreasing
Michaelis-Menten	$\{0, \infty\}$	$y = 0$, linear	asymptote	saturating
Holling type III	$\{0, \infty\}$	$y = 0$, accelerating	asymptote	sigmoid
Holling type IV ($c < 0$)	$\{0, \infty\}$	$y = 0$, accelerating	asymptote	hump-shaped
Exponential-based				
Neg. exponential	$\{0, \infty\}$	y finite	$y \to 0$	decreasing
Monomolecular	$\{0, \infty\}$	$y = 0$, linear	$y \to 0$	saturating
Ricker	$\{0, \infty\}$	$y = 0$, linear	$y \to 0$	hump-shaped
logistic	$\{0, \infty\}$	y small, accelerating	asymptote	sigmoid
Power-based				
Power law	$\{0, \infty\}$	$y \to 0$ or $\to \infty$	$y \to 0$ or $\to \infty$	monotonic
von Bertalanffy	like logistic			
Gompertz	ditto			
Shepherd	like Ricker			
Hassell	ditto			
Non-rectangular hyperbola	like Michaelis-Menten			

Linear Regression

$$y_i = a_0 + a_1 x_i + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma^2)$$

Sections 5.4.1 & 5.4.2

Linear Regression

$$y_i = a_0 + a_1 x_i + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma^2)$$

- Step 1: Likelihood

$$L = \prod_{i=1}^{n} N(y_i | a_0 + a_1 x_i, \epsilon_i)$$
$$= \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp\left[\frac{-1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2 \right]$$

Sections 5.4.1 & 5.4.2

Intercept

$$L = \left(\frac{1}{\sqrt{2\pi}\,\sigma} \right)^n \exp\left[\frac{-1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2 \right]$$

$$\ln L = -n \ln \sigma - n \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2$$

Intercept

$$L = \left(\frac{1}{\sqrt{2\pi}\,\sigma} \right)^n \exp\left[\frac{-1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2 \right]$$

$$\ln L = -n \ln \sigma - n \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2$$

$$\frac{\partial \ln L}{\partial a_0} = \frac{1}{\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)$$

Intercept

$$L = \left(\frac{1}{\sqrt{2\pi}\,\sigma} \right)^{n} \exp\left[\frac{-1}{2\sigma^{2}} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2 \right]$$

$$\ln L = -n \ln \sigma - n \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2$$

$$\frac{\partial \ln L}{\partial a_0} = \frac{1}{\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)$$

$$0 = \sum_{t=1}^{T} y_i - n a_0 - \sum_{t=1}^{T} a_1 x_i$$

Intercept

$$L=\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{n}\exp\left[\frac{-1}{2\sigma^{2}}\sum_{t=1}^{T}(y_i-a_0-a_1x_i)^2\right]$$

$$\ln L=-n\ln\sigma-n\ln(2\pi)-\frac{1}{2\sigma^{2}}\sum_{t=1}^{T}(y_i-a_0-a_1x_i)^2$$

$$\frac{\partial\ln L}{\partial a_0} = \frac{1}{\sigma^{2}}\sum_{t=1}^{T}(y_i-a_0-a_1x_i)$$

$$0 = \sum_{t=1}^{T}y_i-n\,a_0-\sum_{t=1}^{T}a_1x_i$$

$$0 = \bar{y}-a_0-a_1\bar{x}$$

Intercept

$$L=\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{n}\exp\left[\frac{-1}{2\sigma^{2}}\sum_{t=1}^{T}(y_i-a_0-a_1x_i)^2\right]$$

$$\ln L=-n\ln\sigma-n\ln(2\pi)-\frac{1}{2\sigma^{2}}\sum_{t=1}^{T}(y_i-a_0-a_1x_i)^2$$

$$\begin{aligned}
\frac{\partial \ln L}{\partial a_0} &= \frac{1}{\sigma^{2}}\sum_{t=1}^{T}(y_i-a_0-a_1x_i)\\
0 &= \sum_{t=1}^{T}y_i-n\,a_0-\sum_{t=1}^{T}a_1x_i\\
0 &= \bar{y}-a_0-a_1\bar{x}\\
a_0 &= \bar{y}-a_1\bar{x}
\end{aligned}$$

Slope

$$\ln L = -n \ln \sigma - n \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2$$

Slope

$$\ln L = -n\ln\sigma - n\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_i - a_0 - a_1 x_i)^2$$

$$\frac{\partial \ln L}{\partial a_1} = \frac{1}{\sigma^2}\sum_{t=1}^{T} x_i(y_i - a_0 - a_1 x_i)$$

Slope

$$\ln L = -n\ln\sigma - n\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_i - a_0 - a_1 x_i)^2$$

$$\frac{\partial \ln L}{\partial a_1} = \frac{1}{\sigma^2}\sum_{t=1}^{T} x_i(y_i - a_0 - a_1 x_i)$$

$$0 = \sum_{t=1}^{T} x_i y_i - \sum_{t=1}^{T} a_0 x_i - \sum_{t=1}^{T} a_1 x_i^2$$

Slope

$$\ln L = -n\ln\sigma - n\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_i - a_0 - a_1 x_i)^2$$

$$\frac{\partial \ln L}{\partial a_1} = \frac{1}{\sigma^2}\sum_{t=1}^{T} x_i (y_i - a_0 - a_1 x_i)$$

$$0 = \sum_{t=1}^{T} x_i y_i - \sum_{t=1}^{T} a_0 x_i - \sum_{t=1}^{T} a_1 x_i^2$$

$$0 = \overline{xy} - a_0 \bar{x} - a_1 \overline{x^2}$$

Slope

$$\ln L = -n\ln\sigma - n\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_i - a_0 - a_1 x_i)^2$$

$$\frac{\partial \ln L}{\partial a_1} = \frac{1}{\sigma^2}\sum_{t=1}^{T} x_i (y_i - a_0 - a_1 x_i)$$

$$0 = \sum_{t=1}^{T} x_i y_i - \sum_{t=1}^{T} a_0 x_i - \sum_{t=1}^{T} a_1 x_i^2$$

$$0 = \overline{xy} - a_0\bar{x} - a_1\overline{x^2}$$

$$a_1 = \frac{\overline{xy} - a_0\bar{x}}{\overline{x^2}}$$

Combining slope and intercept

$$a_0 = \overline{y} - a_1 \overline{x}$$

$$a_1 = \frac{\overline{xy} - a_0 \overline{x}}{\overline{x^2}}$$

Combining slope and intercept

$$a_0 = \overline{y} - a_1 \overline{x}$$

$$a_1 = \frac{\overline{xy} - a_0 \overline{x}}{\overline{x^2}}$$

$$a_1 = \frac{\overline{xy} - \overline{x}\,\overline{y} + a_1 \overline{x}^2}{\overline{x^2}}$$

Combining slope and intercept

$$a_0 = \overline{y} - a_1 \overline{x}$$

$$a_1 = \frac{\overline{xy} - a_0 \overline{x}}{\overline{x^2}}$$

$$a_1 = \frac{\overline{xy} - \overline{x}\,\overline{y} + a_1 \overline{x}^2}{\overline{x^2}}$$

$$a_1 = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{cov[x, y]}{var[x]}$$

Combining slope and intercept

$$a_0 = \bar{y} - a_1 \bar{x}$$

$$a_1 = \frac{\overline{xy} - a_0 \bar{x}}{\overline{x^2}}$$

$$a_1 = \frac{\overline{xy} - \bar{x}\,\bar{y} + a_1 \bar{x}^2}{\overline{x^2}}$$

$$a_1 = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{cov[x, y]}{var[x]}$$

$$a_0 = \frac{\overline{x^2}\,\bar{y} - \bar{x}\,\overline{xy}}{var[x]}$$

Variance

$$\ln L = -n \ln \sigma - n \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2$$

Variance

$$\ln L = -n \ln \sigma - n \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 = 0$$

Variance

$$\ln L = -n \ln \sigma - n \ln(2\pi) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_i - a_0 - a_1 x_i)^2$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)^2 = 0$$

$$\frac{n}{\sigma} = \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)^2$$

Variance

$$\ln L = -n \ln \sigma - n \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_i - a_0 - a_1 x_i)^2$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 = 0$$

$$\frac{n}{\sigma} = \frac{1}{\sigma^3} \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

$$\sigma^2_{ML} = \frac{1}{n} \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

Matrix notation

$$y_i = \beta_1 + \beta_2 x_i$$

$$\vec{x}_i \vec{\beta} = [x_{i1} \;\; x_{i2}] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \beta_1 x_{i1} + \beta_2 x_{i2}$$

Where $x_{i1} = 1$

$$\vec{y} = \boldsymbol{X} \vec{\beta}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

Design Matrix

Design Matrices

Multiple linear regression

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1k} \\ 1 & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

$$a_1 = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - \bar{x}^2}$$

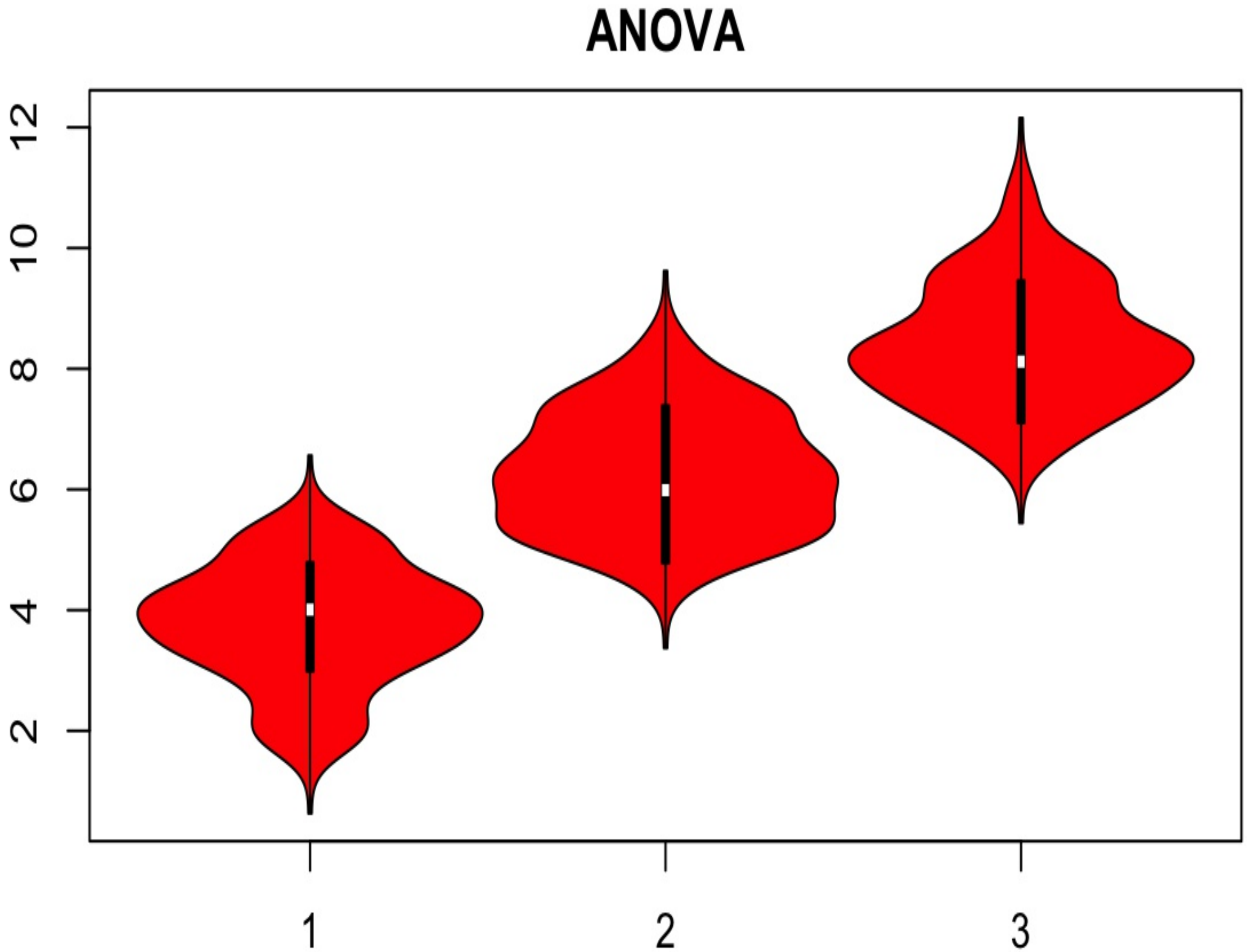
$$a_0 = \frac{\overline{x^2}\,\bar{y} - \bar{x}\,\overline{xy}}{var[x]} \longrightarrow \hat{\beta} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T y$$

ANOVA Design Matrices

One Way Anova
3 levels, n=6

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

$\beta_1 \quad \beta_2 \quad \beta_3$



$\beta_1 \qquad \beta_1+\beta_2 \qquad \beta_1+\beta_3$

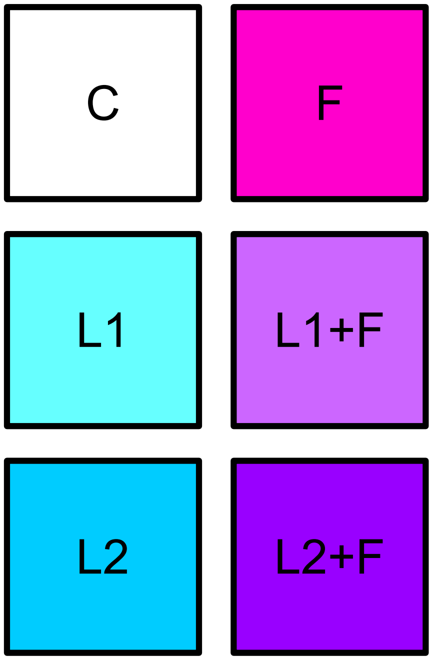
ANOVA Design Matrices

One Way Anova
3 levels, n=6

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Two Way Anova
3 levels x 2 levels
2 reps each, n=12

C	L1	L2	F	
1	0	0	1	F
1	0	0	1	
1	0	0	0	C
1	0	0	0	
1	1	0	1	L1+F
1	1	0	1	
1	1	0	0	L1
1	1	0	0	
1	0	1	1	L2+F
1	0	1	1	
1	0	1	0	L2
1	0	1	0	



ANOVA Design Matrices

One Way Anova
3 levels, n=6

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Two Way Anova
3 levels x 2 levels
2 reps each, n=12

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

ANCOVA
2 levels, 1 covariate
n=6

$$\begin{bmatrix} 1 & 0 & x_{13} \\ 1 & 0 & x_{23} \\ 1 & 0 & x_{33} \\ 1 & 1 & x_{43} \\ 1 & 1 & x_{53} \\ 1 & 1 & x_{63} \end{bmatrix}$$