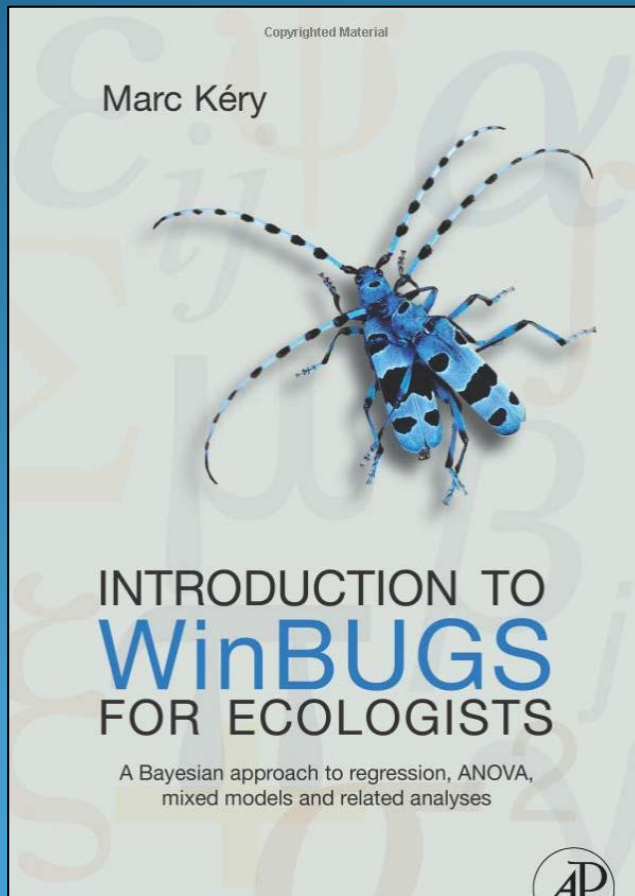


# Topic 3: Key Components of (Generalized) Linear Models



## Chapter 6

# Linear model key components

**response = deterministic part + stochastic part**

- deterministic also called systemic part
- stochastic also called random part

# Linear model key components

- For the description of the stochastic part of the model, we use a statistical distribution. In order to choose the “right” distribution, we need to know the typical sampling situation that leads to one rather than another distribution.
- Four of the most common statistical distributions: normal, uniform, binomial, Poisson
- Linear models are so called because they assume that the mean (i.e., the expected) response can be treated as the result of explanatory variables whose effects add together.



# Linear model key components

- To be able to specify how exactly we think the explanatory variables affect the response, we need to understand the so-called linear predictor of the model, the design matrix, and different parameterizations of a linear model. It is the design matrix along with the observed values of the covariates that makes up the deterministic part of a linear statistical model.

# Stochastic part of a linear model:

## Statistical distributions

- Things whose outcome is affected by chance and thus are only predictable up to a certain degree are called random variables.
- A probability distribution assigns to each possible realization (value or event) of a random variable a probability of occurrence.
- Probability distributions are themselves governed (described) by one or a few parameters, and their actual form depends on the particular values of these parameters.



# Statistical distributions

- Among the four distributions (normal, uniform, Poisson, and binomial), one big divide is whether your response (i.e., the data) is discrete or continuous. Counts are discrete and point to the two latter distributions, while measurements are continuous and point to the two former distributions.
- Under certain circumstances (e.g., large sample sizes, many observed unique values), the two discrete distributions can often be well approximated by a normal distribution. For instance, large counts in practice are often modeled as coming from a normal distribution.

# Normal distribution

- Mathematical description: It includes two parameters, the mean (location) and standard deviation (spread, average deviation from the mean) or, equivalently, the variance (squared standard deviation).
- Mean and standard deviation:

$$E(y) = \mu \quad \text{mean}$$

$$sd(y) = \sigma \quad \text{sd}$$



# Normal distribution

```
n <- 100000
```

# Sample size

```
mu <- mean <- 600
```

# Body mass of male peregrines

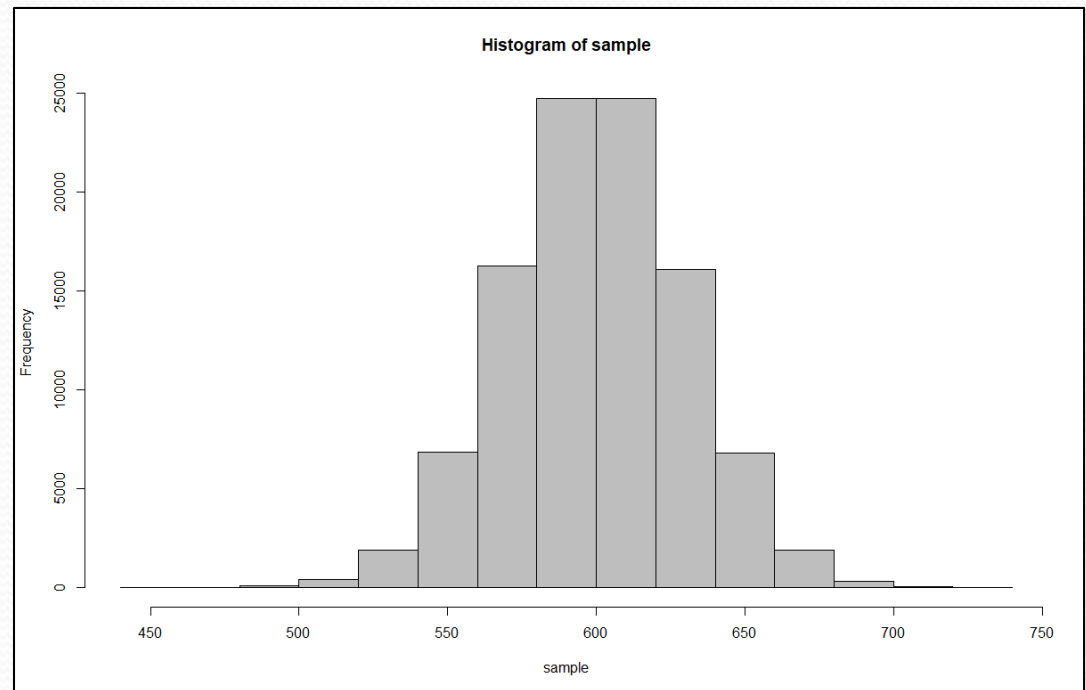
```
sd <- st.dev <- 30
```

# SD of body mass of male peregrines

```
sample <- rnorm(n = n, mean = mu, sd = sd)
```

```
print(sample, dig = 4)
```

```
hist(sample, col = "grey")
```





# Continuous uniform distribution

- Sampling situation: Measurements are taken, which are all equally likely to occur in a certain range of values.
- Mathematical description: It includes two parameters, lower ( $a$ ) and upper limits ( $b$ ).
- Mean and standard deviation:

$$E(y) = (a + b)/2 \quad \text{mean}$$

$$sd(y) = \sqrt{(b - a)^2/12} \quad \text{sd}$$

# Continuous uniform distribution

```
n <- 100000
```

# Sample size

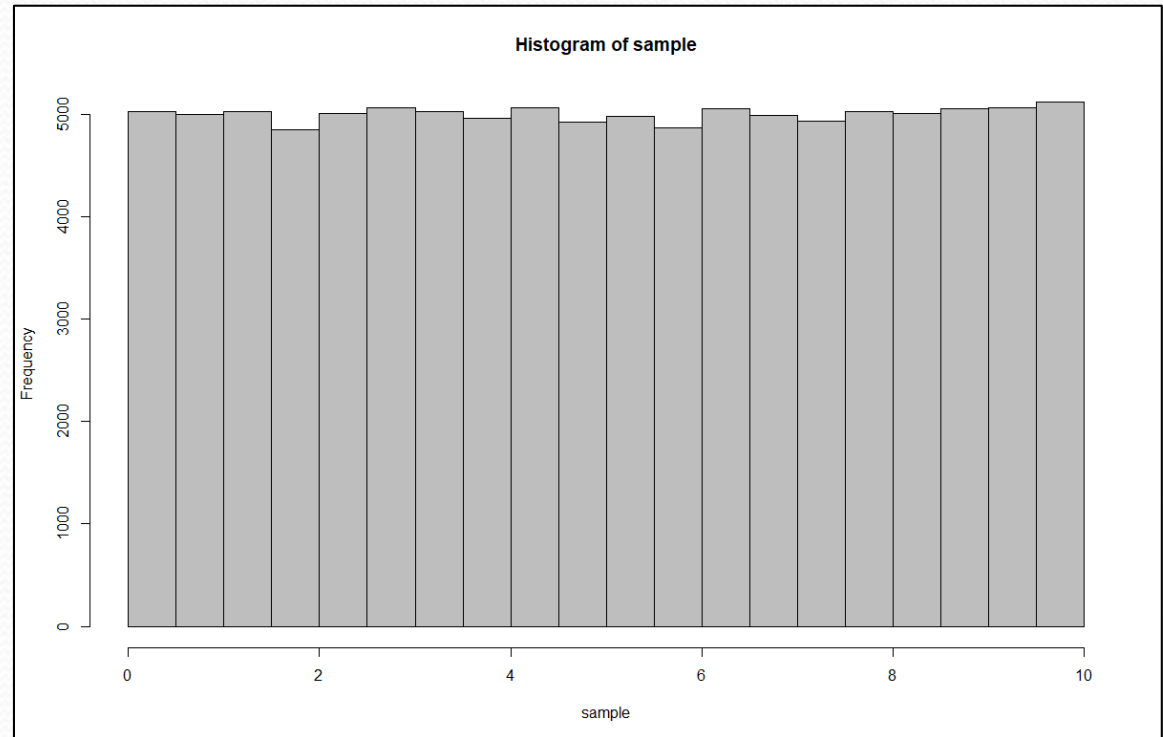
```
a <- lower.limit <- 0
```

```
b <- upper.limit <- 10
```

```
sample <- runif(n = n, min = a, max = b)
```

```
print(sample, dig = 3)
```

```
hist(sample, col = "grey")
```





# Binomial Distribution: The “Coin-Flip Distribution”

- Sampling situation: When  $N$  available things all have the same probability  $p$  of being in a certain state (e.g., being counted, or having a certain attribute, like being male or dead), then the number  $x$  that is actually counted in that sample, or has that attribute, is binomially distributed.
- Classical examples: Number of males in a clutch, school class, or herd of size  $N$ ; number of times heads shows up among  $N$  flips of a coin.
- Varieties: The Bernoulli distribution corresponds to a single coin flip and has only a single parameter,  $p$ . Actually, a binomial is the sum of  $N$  Bernoullis (or coin flips).

# Binomial Distribution: The “Coin-Flip Distribution”

- Mathematical description: It includes one or two parameters, the probability of being chosen or having a certain trait (male, dead), often called success probability  $p$ , and the “binomial total”  $N$ , which is the sample or trial “size.”  $N$  represents a ceiling to a binomial count; this is an important distinction to the similar Poisson distribution.
- Mean and standard deviation:

$$E(y) = N * p \quad \text{mean}$$

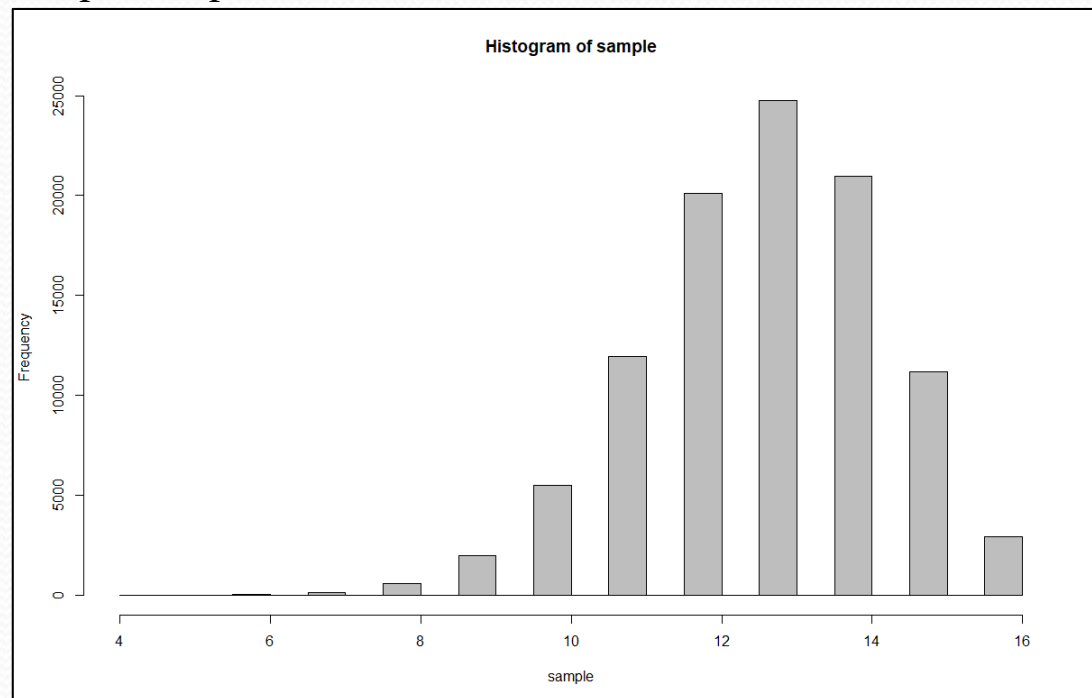
$$sd(y) = \sqrt{N * p * (1 - p)} \quad \text{sd}$$



# Binomial Distribution: The “Coin-Flip Distribution”

```
n <- 100000      # Sample size  
N <- 16          # Number of individuals that flip the coin  
p <- 0.8         # Probability of being counted (seen), dead or a male
```

```
sample <- rbinom(n = n, size = N, prob = p)  
print(sample, dig = 3)  
hist(sample, col = "grey")
```



# Poisson Distribution

- Sampling situation: When things (e.g., birds, cars, or erythrocytes) are randomly distributed in one or two (or more) dimensions and we randomly place a “counting window” along that dimension or in that space and record the number of things, then that number is Poisson distributed.
- Classical examples: Number of passing cars during 10-min counts at a street corner; number of birds that fly by you at a migration site, and number of birds or hares per sample quadrat.



# Poisson Distribution

- Mathematical description: It includes a single parameter called  $\lambda$ , which is equal to the mean (= expectation, average count, intensity), as well as the variance (i.e., variance = mean). That is, as for the binomial distribution, the variance is not a free parameter but is a function of the mean.
- Mean and standard deviation:

$$E(y) = \lambda \quad \text{mean}$$

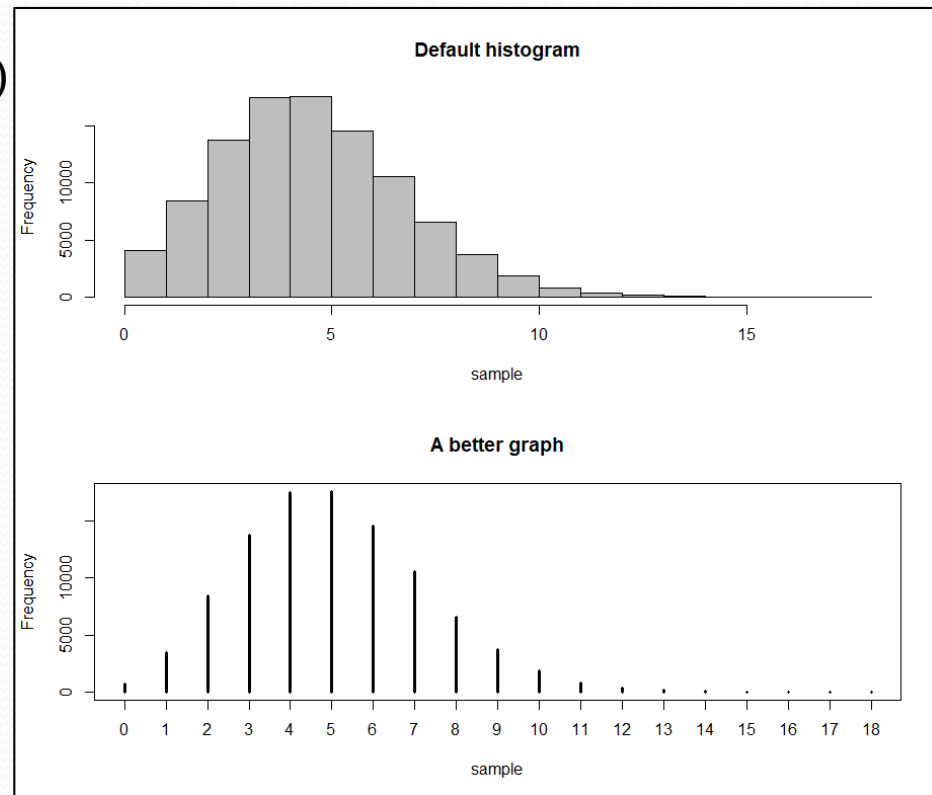
$$sd(y) = \sqrt{\lambda} \quad \text{sd}$$

# Poisson Distribution

```
n <- 100000      # Sample size  
lambda <- 5      # Average no. individuals per sample, density
```

```
sample <- rpois(n = n, lambda = lambda)  
print(sample, dig = 3)
```

```
par(mfrow = c(2,1))  
hist(sample, col = "grey",  
      main = "Default histogram")  
plot(table(sample),  
      main = "A better graph",  
      lwd = 3, ylab = "Frequency")
```





# Deterministic part of linear models: linear predictors and design matrices

- Linear models describe the expected response as a linear combination of the effects of discrete or continuous explanatory variables. That is, they directly specify the relationship between the response and one or more explanatory variables.

# Design matrix

- The material in the rest of this chapter may be seen by some as a nuisance. It may look difficult at first and indeed may not be totally necessary when fitting linear models using one of the widely known stats packages. However, understanding design matrices will greatly increase your grasp of statistical models in general.
- For each element of the response vector, the design matrix  $n$  index indicates which effect is present for categorical (= discrete) explanatory variables and what “amount” of an effect is present in the case of continuous explanatory variables.
- The design matrix contains as many columns as the fitted model has parameters, and when matrix-multiplied with the parameter vector, it yields the linear predictor, another vector.
- In the remainder of this chapter, we look at a progression of typical linear models (e.g., t-test, simple linear regression, one-way ANOVA, and analysis of covariance (ANCOVA)).



# Design matrix example

- six snakes; body mass (mass, in units of 10 g; a continuous **response variable**)
- in three populations (pop)
- two regions (region)
- three habitat types (hab)
- continuous explanatory variable snout–vent length (svl)

TABLE 6.1 Our Toy Data Set for Six Snakes

mass	pop	region	hab	svl
6	1	1	1	40
8	1	1	2	45
5	2	1	3	39
7	2	1	1	50
9	3	2	2	52
11	3	2	3	57

# Example: model of the mean

`lm(mass ~ 1)`

$$\text{mass}_i = \mu + \varepsilon_i$$

$\mu$  = overall mean

$\varepsilon_i$  = residual for snake  $i$

$\varepsilon_i \sim \text{Normal}(0, \sigma^2)$



# Example: model of the mean

```
> model.matrix(mass~1)
      (Intercept)
1                1
2                1
3                1
4                1
5                1
6                1
attr(,"assign")
[1] 0
```

# Example: t-Test

`lm(mass ~ region)`

$$\text{mass}_i = \alpha + \beta * \text{region}_i + \varepsilon_i$$

$\alpha$  = constant

$\beta$  = constant to be multiplied with region indicator

$\varepsilon_i$  = residual for snake  $i$

$$\varepsilon_i \sim \text{Normal}(0, \sigma^2)$$



# Example: t-Test

```
> model.matrix(mass ~ region)
      (Intercept) region2
1                1      0
2                1      0
3                1      0
4                1      0
5                1      1
6                1      1
```

TABLE 6.1 Our Toy Data Set for Six Snakes

mass	pop	region	hab	svl
6	1	1	1	40
8	1	1	2	45
5	2	1	3	39
7	2	1	1	50
9	3	2	2	52
11	3	2	3	57

# Example: t-Test

```
> model.matrix(mass ~ region)
      (Intercept) region2
1                1      0
2                1      0
3                1      0
4                1      0
5                1      1
6                1      1
```

$$\begin{aligned}6 &= \alpha * 1 + \beta * 0 + \varepsilon_1 \\8 &= \alpha * 1 + \beta * 0 + \varepsilon_2 \\5 &= \alpha * 1 + \beta * 0 + \varepsilon_3 \\7 &= \alpha * 1 + \beta * 0 + \varepsilon_4 \\9 &= \alpha * 1 + \beta * 1 + \varepsilon_5 \\11 &= \alpha * 1 + \beta * 1 + \varepsilon_6\end{aligned}$$

To get a solution for this system of equations, i.e., to obtain values for the unknowns  $\alpha$  and  $\beta$  that are “good” in some way, we need to define some criterion for dealing with the residuals  $\varepsilon_i$ . Usually, in this system of equations, the unknowns  $\alpha$  and  $\beta$  are chosen such that the sum of the squared residuals is minimal. This is called the **least-squares method**, and for normal GLMs, the resulting parameter estimates for  $\alpha$  and  $\beta$  are equivalent to those obtained using the more general maximum likelihood method.



# Example: t-Test

$$\begin{pmatrix} 6 \\ 8 \\ 5 \\ 7 \\ 9 \\ 11 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} * \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}$$

```
> lm(mass~region)
```

```
call:
```

```
lm(formula = mass ~ region)
```

```
coefficients:
```

(Intercept)	region2
6.5	3.5

# Example: t-Test (another way)

$$\begin{pmatrix} 6 \\ 8 \\ 5 \\ 7 \\ 9 \\ 11 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} * \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}$$

```
> lm(mass~region-1)
```

```
call:
```

```
lm(formula = mass ~ region - 1)
```

```
coefficients:
```

```
region1  region2  
    6.5     10.0
```



# Example: simple linear regression

`lm(mass ~ svl)`

$$\text{mass}_i = \alpha + \beta * \text{svl}_i + \epsilon_i$$

$\alpha$  = constant

$\beta$  = constant to be multiplied with snout–vent leng.

$\epsilon_i$  = residual for snake  $i$

$\epsilon_i \sim \text{Normal}(0, \sigma^2)$

# Example: simple linear regression

```
> model.matrix(mass ~svl)
      (Intercept)  svl
1                1   40
2                1   45
3                1   39
4                1   50
5                1   52
6                1   57
```

$$\begin{aligned}6 &= \alpha + \beta * 40 + \varepsilon_1 \\8 &= \alpha + \beta * 45 + \varepsilon_2 \\5 &= \alpha + \beta * 39 + \varepsilon_3 \\7 &= \alpha + \beta * 50 + \varepsilon_4 \\9 &= \alpha + \beta * 52 + \varepsilon_5 \\11 &= \alpha + \beta * 57 + \varepsilon_6\end{aligned}$$



# Example: simple linear regression

$$\begin{pmatrix} 6 \\ 8 \\ 5 \\ 7 \\ 9 \\ 11 \end{pmatrix} = \begin{pmatrix} 1 & 40 \\ 1 & 45 \\ 1 & 39 \\ 1 & 50 \\ 1 & 52 \\ 1 & 57 \end{pmatrix} * \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}$$

```
> lm(mass~svl)

call:
lm(formula = mass ~ svl)

coefficients:
(Intercept)          svl 
   -5.5588         0.2804
```

# Example: simple linear regression

$$\begin{pmatrix} 6 \\ 8 \\ 5 \\ 7 \\ 9 \\ 11 \end{pmatrix} = \begin{pmatrix} 1 & 40 \\ 1 & 45 \\ 1 & 39 \\ 1 & 50 \\ 1 & 52 \\ 1 & 57 \end{pmatrix} * \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}$$

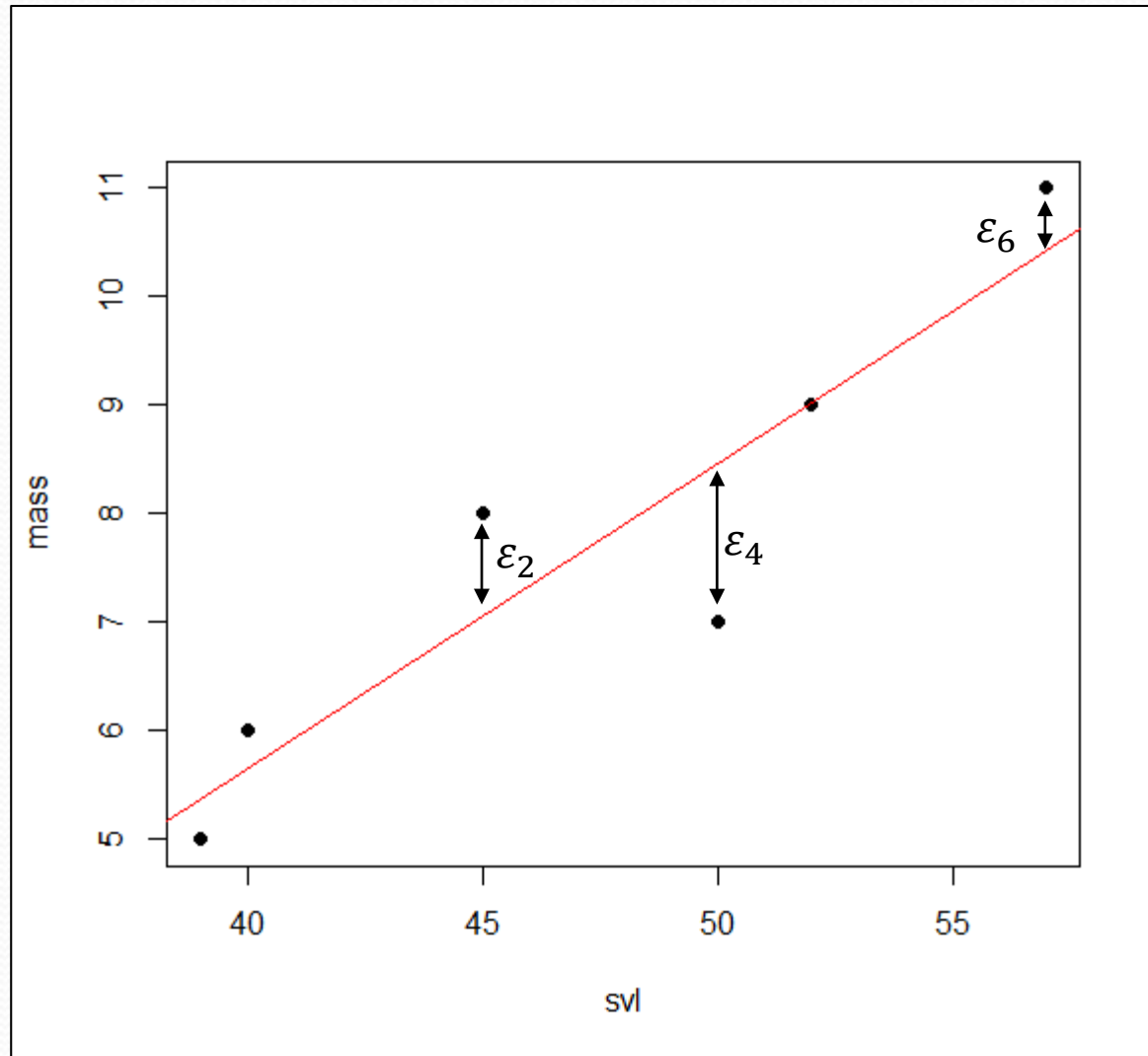
```
> lm(mass~svl)
```

```
call:
```

```
lm(formula = mass ~ svl)
```

```
Coefficients:
```

(Intercept)	svl
-5.5588	0.2804





# Example: One-way analysis of Variance

`lm(mass ~ pop)`

$$\text{mass}_i = \alpha + \beta_{j(i)} * \text{pop}_i + \varepsilon_i$$

$\alpha$  = constant

$\beta$  = one constant per population  $j$

$\varepsilon_i$  = residual for snake  $i$

$\varepsilon_i \sim \text{Normal}(0, \sigma^2)$

# Example: One-way analysis of Variance

```
> model.matrix(~pop)
      (Intercept) pop2 pop3
1                1    0    0
2                1    0    0
3                1    1    0
4                1    1    0
5                1    0    1
6                1    0    1
```

```
> model.matrix(~pop-1)
      pop1 pop2 pop3
1         1    0    0
2         1    0    0
3         0    1    0
4         0    1    0
5         0    0    1
6         0    0    1
```



# Example: One-way analysis of Variance

$$\begin{pmatrix} 6 \\ 8 \\ 5 \\ 7 \\ 9 \\ 11 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} * \begin{pmatrix} \alpha \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}$$

```
call:  
lm(formula = mass ~ pop)
```

coefficients:

(Intercept)	pop2	pop3
7	-1	3

# Example: Two-way analysis of Variance

`lm(mass ~ region + hab)`

$$\text{mass}_i = \alpha + \beta_{j(i)} * \text{region}_i + \delta_{k(i)} * \text{hab}_i + \varepsilon_i$$

$\alpha$  = constant

$\beta$  = one constant per region  $j$

$\delta$  = one constant per habitat  $k$

$\varepsilon_i$  = residual for snake  $i$

$\varepsilon_i \sim \text{Normal}(0, \sigma^2)$



# Example: Two-way analysis of Variance

$$\text{mass}_i = \alpha + \beta_{j(i)} * \text{region}_i + \delta_{k(i)} * \text{hab}_i + \epsilon_i$$

```
> model.matrix(~region + hab)
      (Intercept) region2 hab2 hab3
1                1      0  0    0
2                1      0  1    0
3                1      0  0    1
4                1      0  0    0
5                1      1  1    0
6                1      1  0    1
```

TABLE 6.1 Our Toy Data Set for Six Snakes

mass	pop	region	hab	svl
6	1	1	1	40
8	1	1	2	45
5	2	1	3	39
7	2	1	1	50
9	3	2	2	52
11	3	2	3	57

# Example: Two-way analysis of Variance

$$\begin{pmatrix} 6 \\ 8 \\ 5 \\ 7 \\ 9 \\ 11 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} * \begin{pmatrix} \alpha \\ \beta_2 \\ \delta_2 \\ \delta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}$$

```
> lm(mass ~ region + hab)
```

```
call:
```

```
lm(formula = mass ~ region + hab)
```

```
coefficients:
```

(Intercept)	region2	hab2	hab3
6.50	3.50	0.25	-0.25

# Example: Analysis of Covariance

$\text{lm}(\text{mass} \sim \text{pop} + \text{svl})$

$$\text{mass}_i = \alpha + \beta_{j(i)} * \text{pop}_i + \delta * \text{svl}_i + \varepsilon_i$$

$\alpha$  = constant

$\beta$  = one constant per population  $j$

$\delta$  = constant to be multiplied with snout–vent length

$\varepsilon_i$  = residual for snake  $i$

$\varepsilon_i \sim \text{Normal}(0, \sigma^2)$



# Example: Analysis of Covariance

$$\text{mass}_i = \alpha + \beta_{j(i)} * \text{pop}_i + \delta * \text{svl}_i + \varepsilon_i$$

```
> model.matrix(lm(mass ~ pop + svl))  
  (Intercept) pop2 pop3 svl  
1           1    0    0  40  
2           1    0    0  45  
3           1    1    0  39  
4           1    1    0  50  
5           1    0    1  52  
6           1    0    1  57
```

# Example: Analysis of Covariance

$$\begin{pmatrix} 6 \\ 8 \\ 5 \\ 7 \\ 9 \\ 11 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 40 \\ 1 & 0 & 0 & 45 \\ 1 & 1 & 0 & 39 \\ 1 & 1 & 0 & 50 \\ 1 & 0 & 1 & 52 \\ 1 & 0 & 1 & 57 \end{pmatrix} * \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \delta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}$$

```
> lm(mass ~ pop + svl) # Additive model
```

```
call:
```

```
lm(formula = mass ~ pop + svl)
```

```
Coefficients:
```

(Intercept)	pop2	pop3	svl
-3.43860	-1.49123	0.05263	0.24561

# Example: Analysis of Covariance (Interaction)

`lm(mass ~ pop * svl)`

$$\text{mass}_i = \alpha + \beta_{j(i)} * \text{pop}_i + \delta * \text{svl}_i + \gamma_{j(i)} * \text{svl}_i * \text{pop}_i + \varepsilon_i$$

$\alpha$  = constant

$\beta$  = one constant per population  $j$

$\delta$  = constant to be multiplied with snout-vent length

$\gamma$  = one constant per population  $j$

$\varepsilon_i$  = residual for snake  $i$

$$\varepsilon_i \sim \text{Normal}(0, \sigma^2)$$



# Example: Analysis of Covariance (Interaction)

$$\text{mass}_i = \alpha + \beta_{j(i)} * \text{pop}_i + \delta * \text{svl}_i + \gamma_{j(i)} * \text{svl}_i * \text{pop}_i + \varepsilon_i$$

```
> model.matrix(lm(mass ~ pop * svl))      # Interactive model
      (Intercept) pop2 pop3 svl pop2:svl pop3:svl
1              1    0    0  40         0        0
2              1    0    0  45         0        0
3              1    1    0  39        39        0
4              1    1    0  50        50        0
5              1    0    1  52         0       52
6              1    0    1  57         0       57
```

# Example: Analysis of Covariance (Interaction)

$$\begin{pmatrix} 6 \\ 8 \\ 5 \\ 7 \\ 9 \\ 11 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 40 & 0 & 0 \\ 1 & 0 & 0 & 45 & 0 & 0 \\ 1 & 1 & 0 & 39 & 39 & 0 \\ 1 & 1 & 0 & 50 & 50 & 0 \\ 1 & 0 & 1 & 52 & 0 & 52 \\ 1 & 0 & 1 & 57 & 0 & 57 \end{pmatrix} * \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \delta \\ \gamma_1 \\ \gamma_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}$$

```
> lm(mass ~ pop * svl)
```

```
# Interactive model
```

```
call:
```

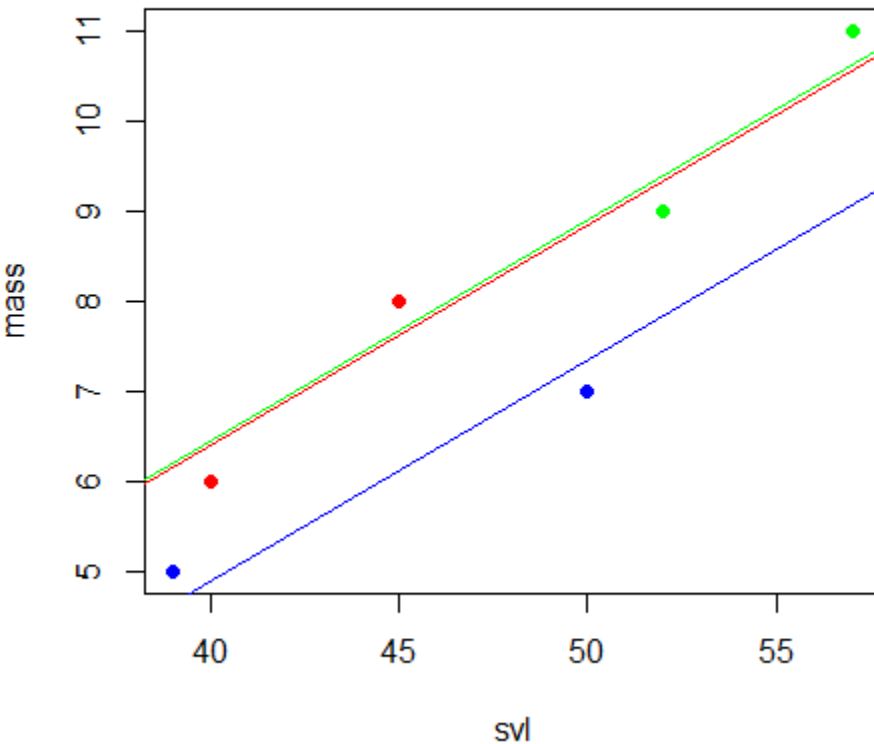
```
lm(formula = mass ~ pop * svl)
```

```
Coefficients:
```

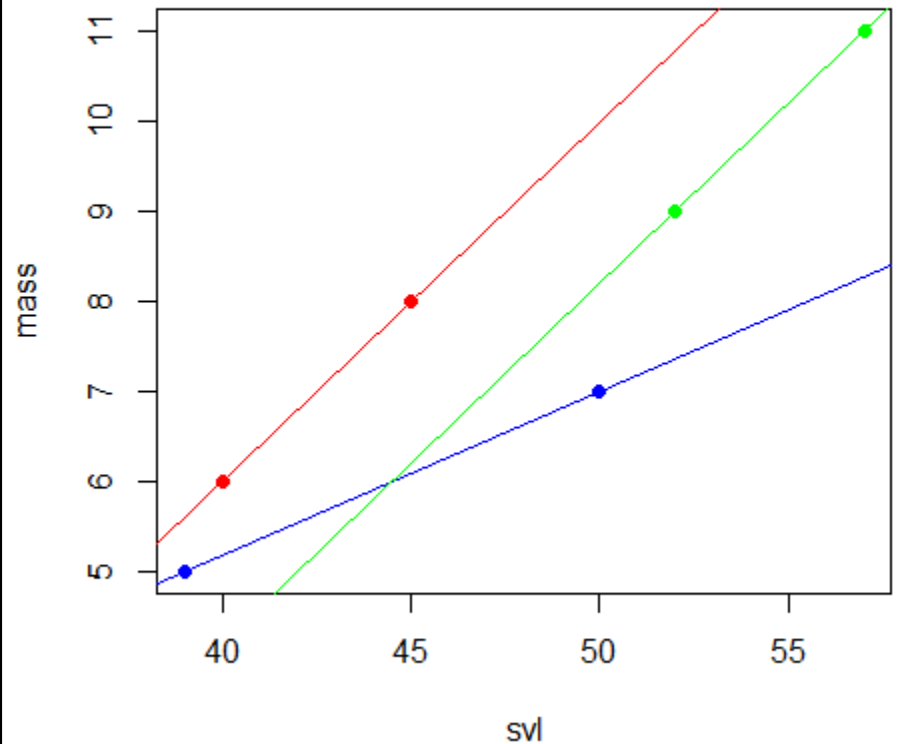
(Intercept)	pop2	pop3	svl	pop2:svl	pop3:svl
-1.000e+01	7.909e+00	-1.800e+00	4.000e-01	-2.182e-01	-1.599e-15

# Example: Analysis of Covariance

ANCOVA



ANCOVA + Interaction





# Summary

- Briefly reviewed the two key components of linear statistical models:
  1. Statistical distributions
  2. Linear predictor = design matrix \* parameter vector
- Understanding both is essential for applied statistics.