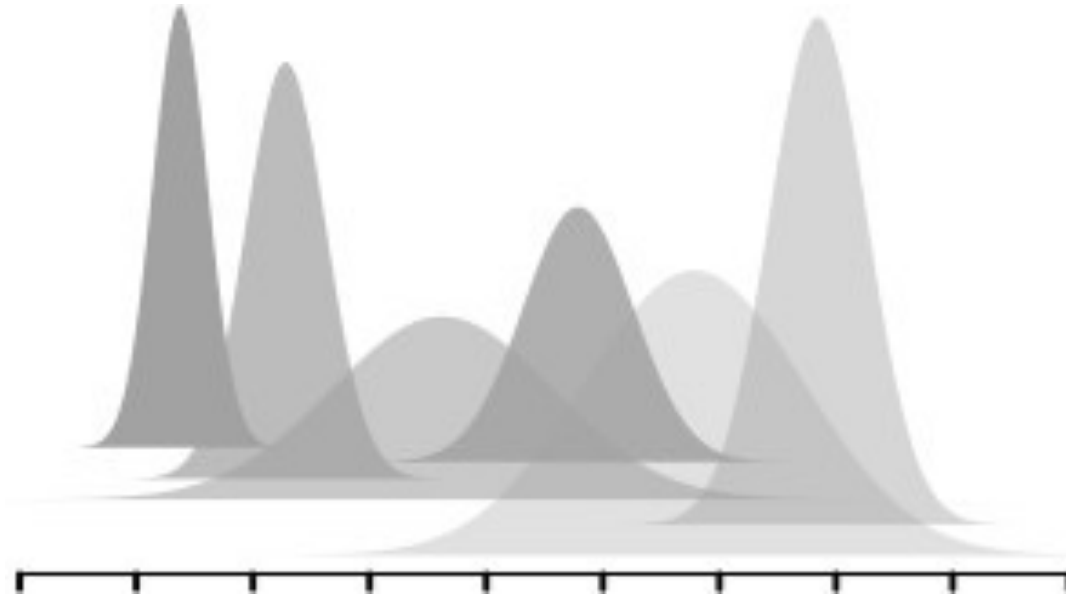


4.1b How To Use Ecological Models



Jelena H. Pantel

Faculty of Biology

University of Duisburg-Essen

jelena.pantel@uni-due.de

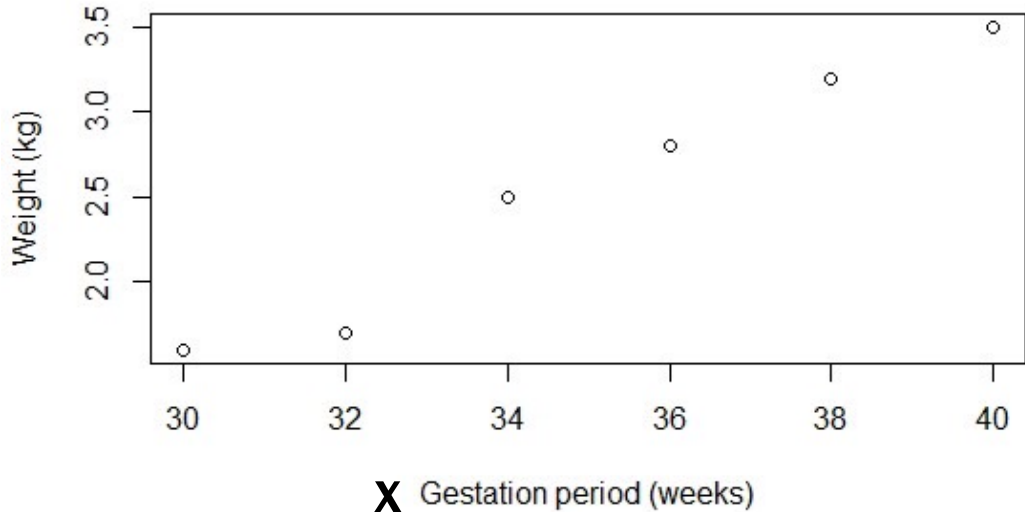
How to use a mathematical model?

Step 1. Evaluate potential model equations

$$y = mx + b$$

$$\text{weight} = m(\text{gestation period}) + b$$

Estimated baby weights during pregnancy



Step 2. State clearly the model parameters you wish to estimate

$$\text{weight} = m(\text{gestation period}) + b$$

How to use a mathematical model?

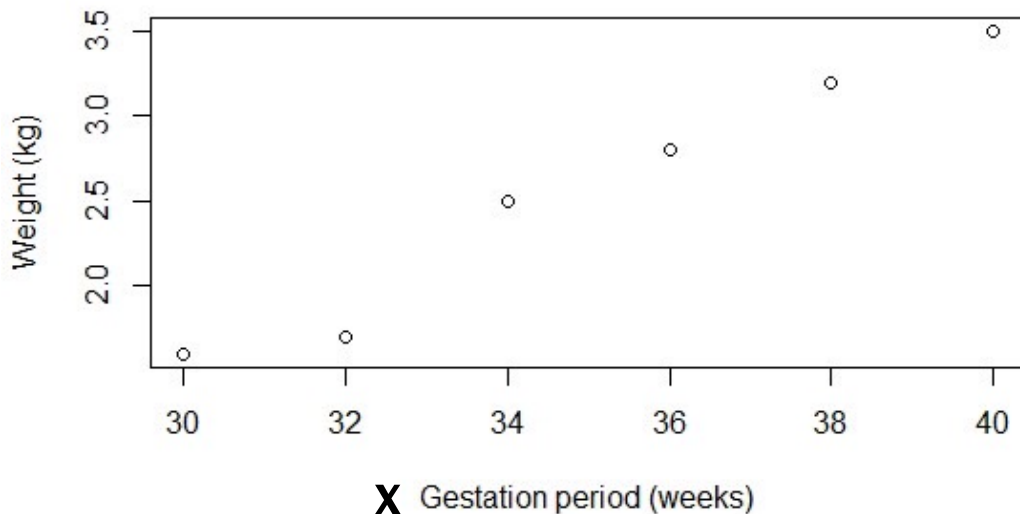
gestation	weight
30	1.6
32	1.7
34	2.5
36	2.8
38	3.2
40	3.5

Example 1. Linear causal model (linear regression)

$$y = mx + b$$

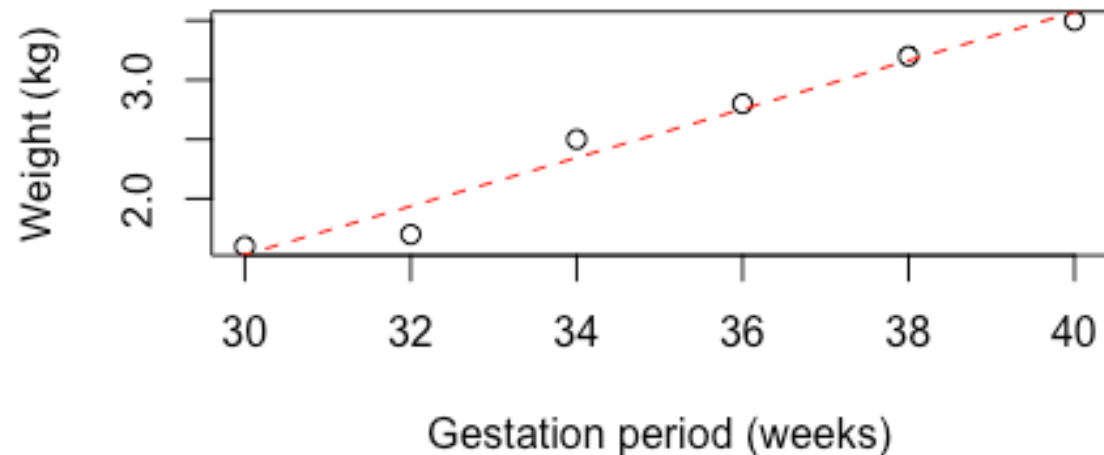
$$\text{weight} = m(\text{gestation period}) + b$$

Estimated baby weights during pregnancy



```
Coefficients:  
(Intercept)      gestation  
      -4.6000         0.2043  
  
> confint(model, level = 0.95)  
                2.5 %      97.5 %  
(Intercept) -6.3862379 -2.8137621  
gestation    0.1534916  0.2550798
```

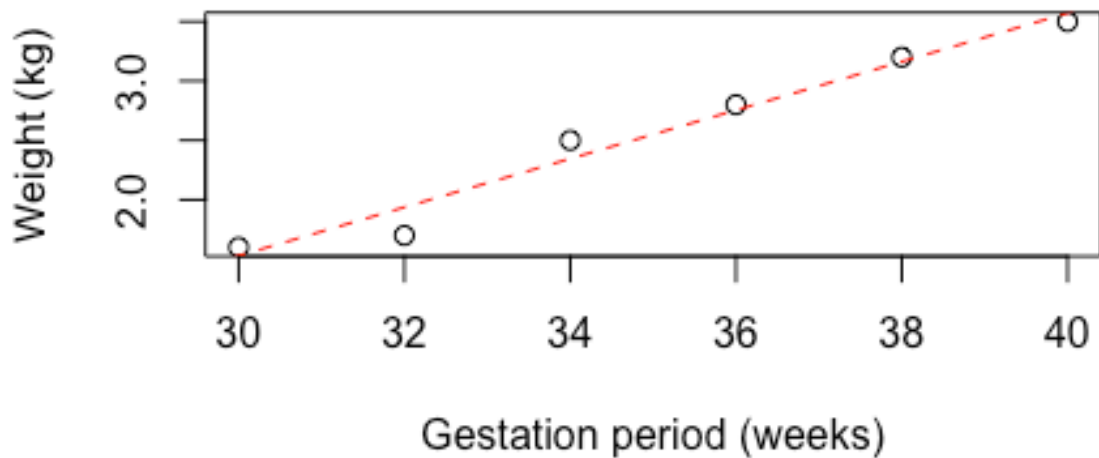
Estimated baby weights during pregnancy



How to use a mathematical model?

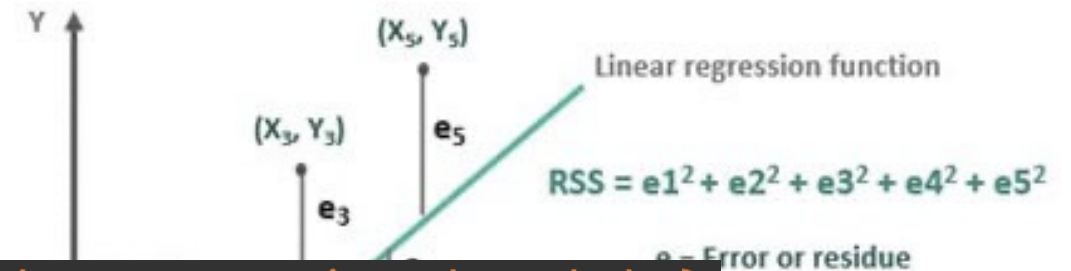
$weight = m(gestation\ period) + b$ Least squares regression (line that minimizes sum of squared distances from observed points to model-fit line)

Estimated baby weights during pregnancy



Residual Sum of Squares

Residual Sum of Squares measures the extent of variability of observed data not predicted by the regression model.



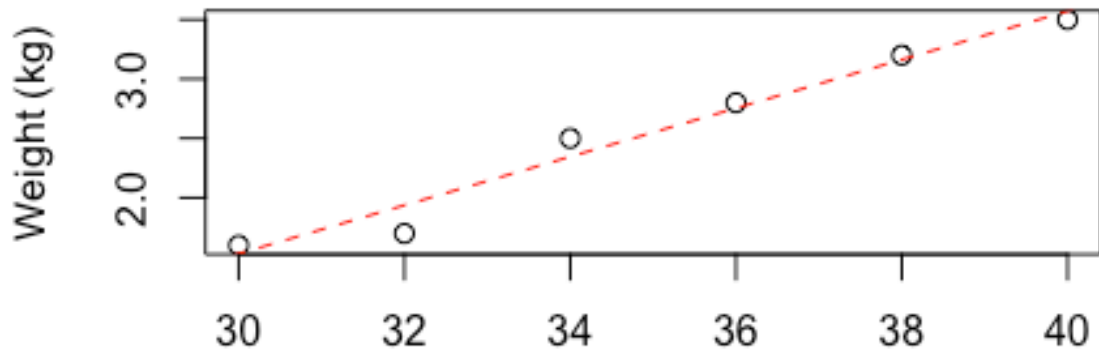
	predicted	observed	predicted_minus_observed
1	1.528571	1.6	-0.07142857
2	1.937143	1.7	0.23714286
3	2.345714	2.5	-0.15428571
4	2.754286	2.8	-0.04571429
5	3.162857	3.2	-0.03714286
6	3.571429	3.5	0.07142857

```
> model = lm(weight ~ gestation, data=baby)
> sum(dat$predicted_minus_observed^2)
[1] 0.09371429
> deviance(model)
[1] 0.09371429
```

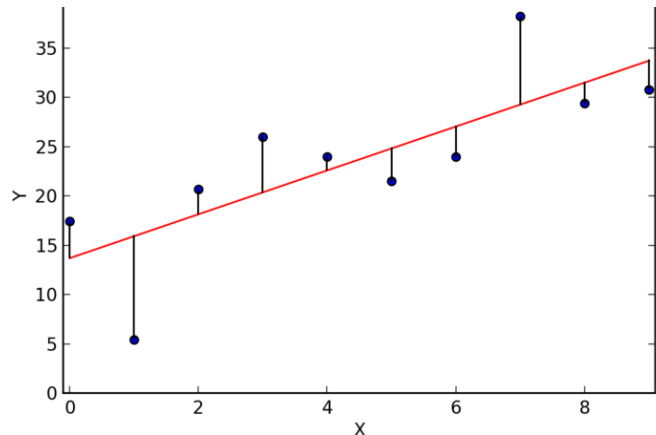
How to use a mathematical model?

$weight = m(gestation\ period) + b$ Least squares regression (line that minimizes sum of squared distances from observed points to model-fit line)

Estimated baby weights during pregnancy

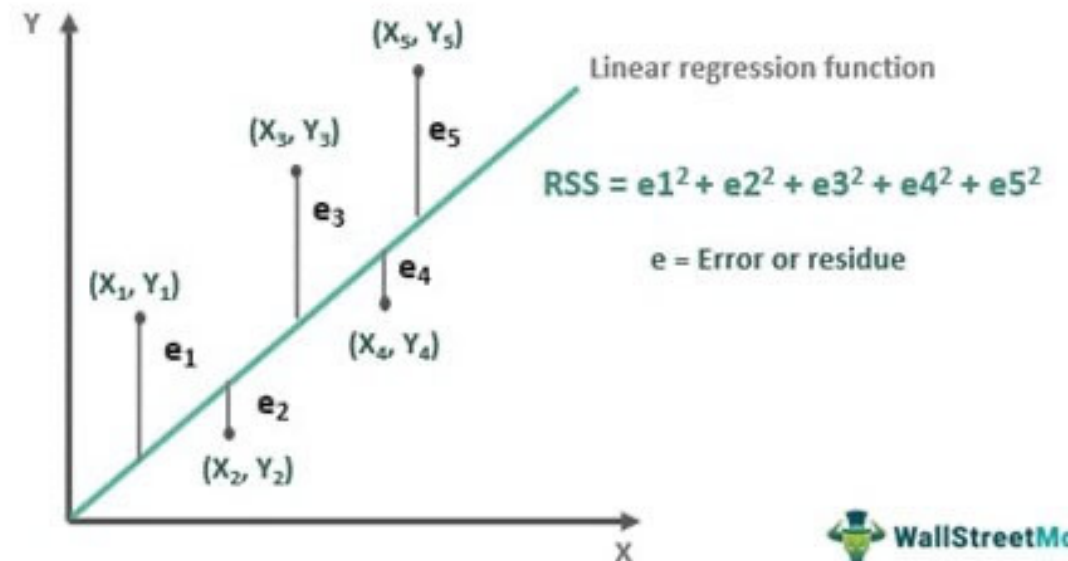


Gestation period (weeks)



Residual Sum of Squares

Residual Sum of Squares measures the extent of variability of observed data not predicted by the regression model.

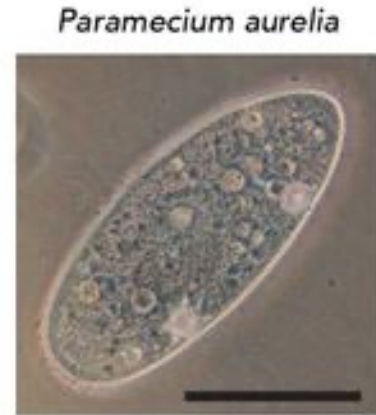
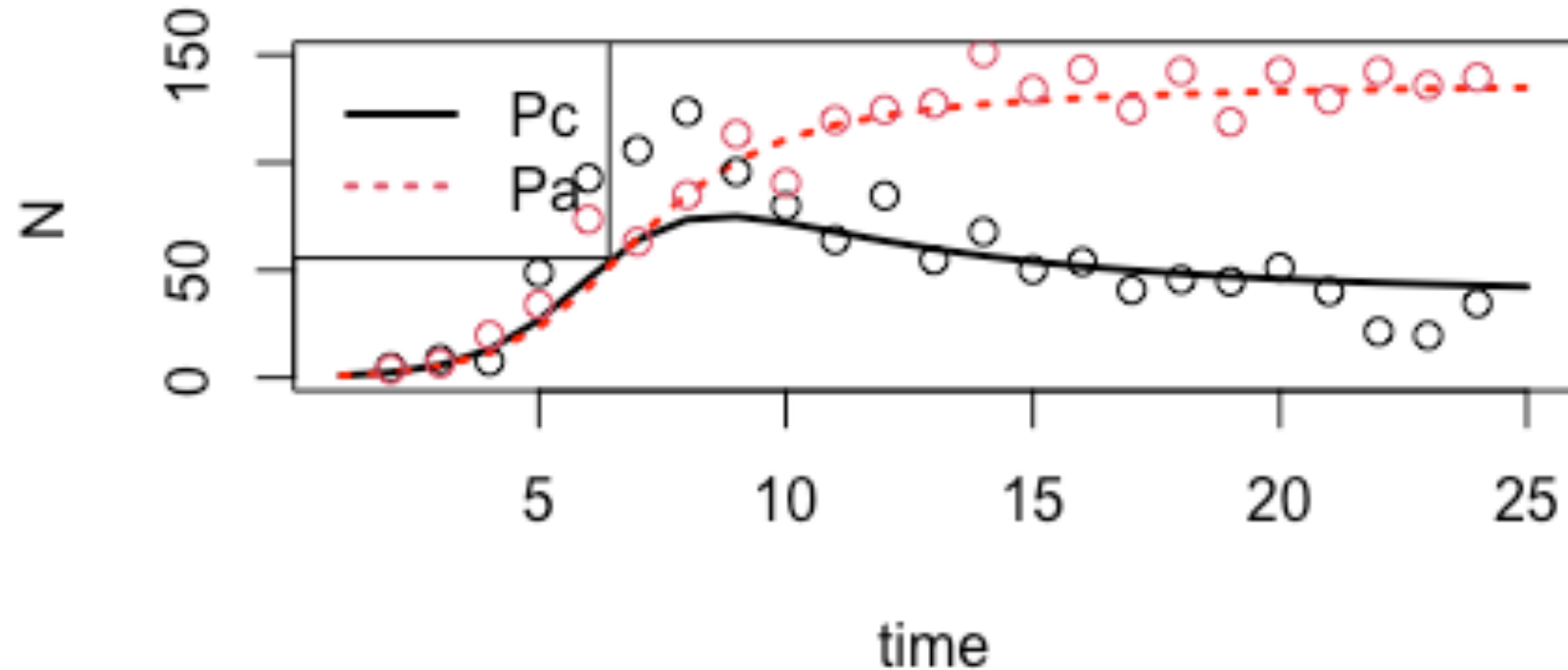


How to use a mathematical model?

Lotka Volterra model

$$\frac{dn_i}{dt} = rn_i \left(1 - \frac{n_i + \alpha_{ij}n_j}{K_i} \right) \quad \frac{dn_j}{dt} = rn_j \left(1 - \frac{n_j + \alpha_{ji}n_i}{K_j} \right)$$

Non-linear least squares regression – find parameter values that minimize sum of squared distances from observed points to model-fit line

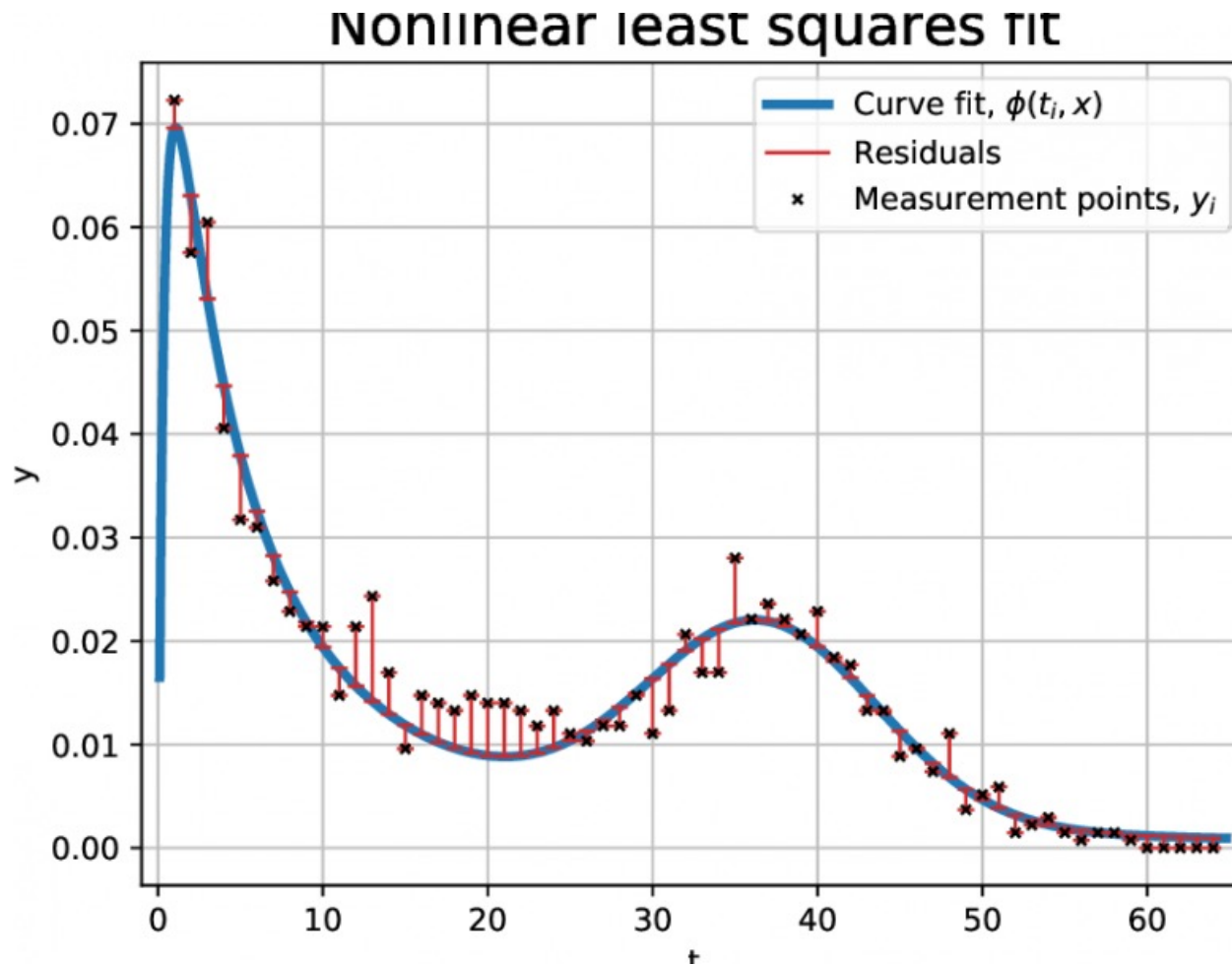


How to use a mathematical model?

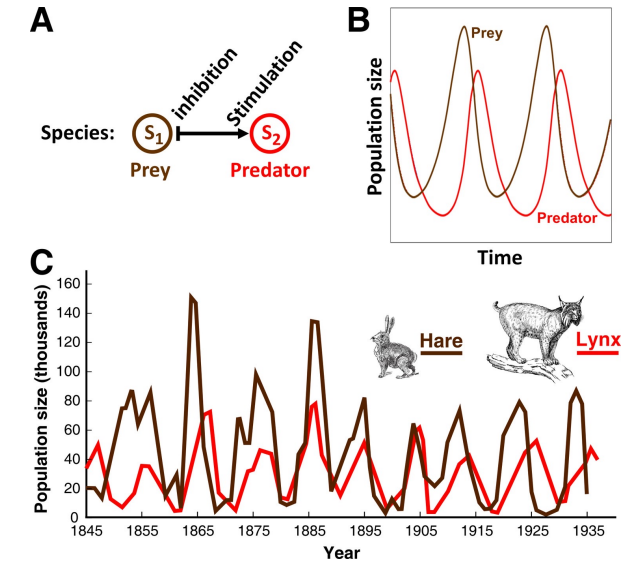
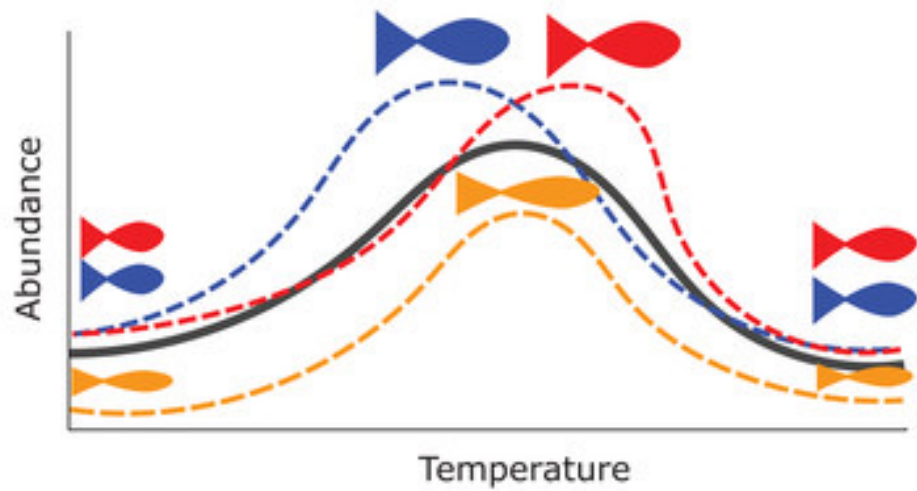
Lotka Volterra model

$$\frac{dn_i}{dt} = rn_i \left(1 - \frac{n_i + \alpha_{ij}n_j}{K_i} \right) \quad \frac{dn_j}{dt} = rn_j \left(1 - \frac{n_j + \alpha_{ji}n_i}{K_j} \right)$$

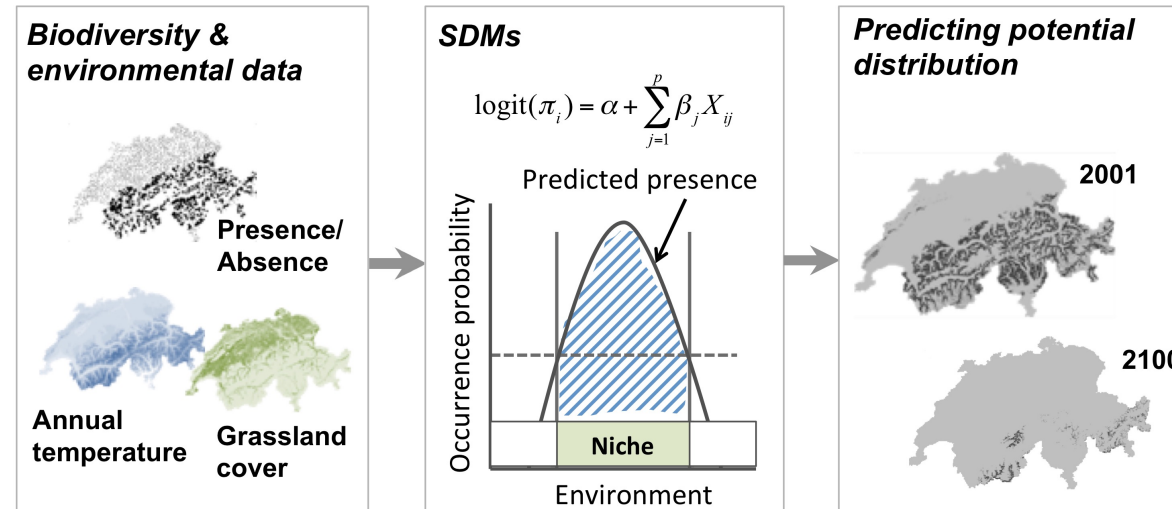
Non-linear least squares regression – find parameter values that minimize sum of squared distances from observed points to model-fit line



Fitting data to models: probability distributions and statistical models



Key link between data and models
– estimating model parameter
values given the observed data



What is a statistical model?

Definition

- a mathematical model that embodies a set of statistical assumptions concerning the generation of data sampled from a larger population
- A statistical model represents, often in considerably idealized form, the data-generating process

Réale & Festa-Bianchet, 2000



Bighorn sheep,
Ovis canadensis

trait	mean	variance (σ^2)
Longevity (L,year)	7.06	19.08
Lifetime fecundity (F,no. offspring produced)	5.33	14.88
Adult body mass (m,kg)	71.1	20

$y_i \sim N(\mu, \sigma)$

Longevity of sheep i → a normal distribution → is drawn from / is distributed as

with mean μ

and standard deviation $\sigma = \sqrt{\sigma^2}$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

Calculate variance for: 3, 6, 9

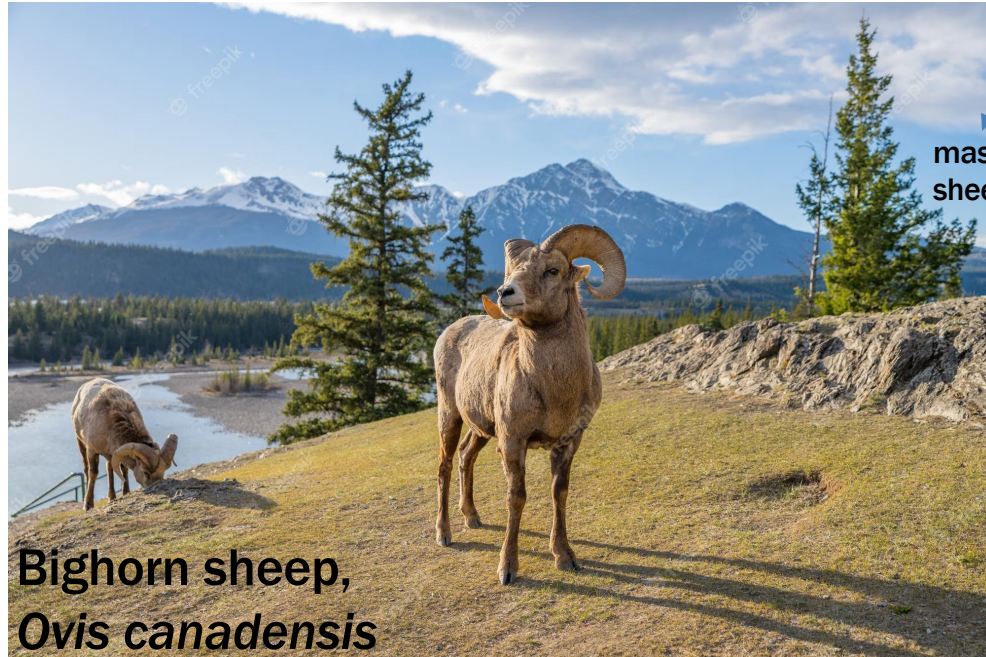
$L_i \sim N(7.06, \sqrt{19.08})$ $m_i \sim N(71.1, \sqrt{20})$

$F_i \sim N(5.33, \sqrt{14.88})$

What is a statistical model?

Definition

- a mathematical model that embodies a set of statistical assumptions concerning the generation of data sampled from a larger population
- A statistical model represents, often in considerably idealized form, the data-generating process



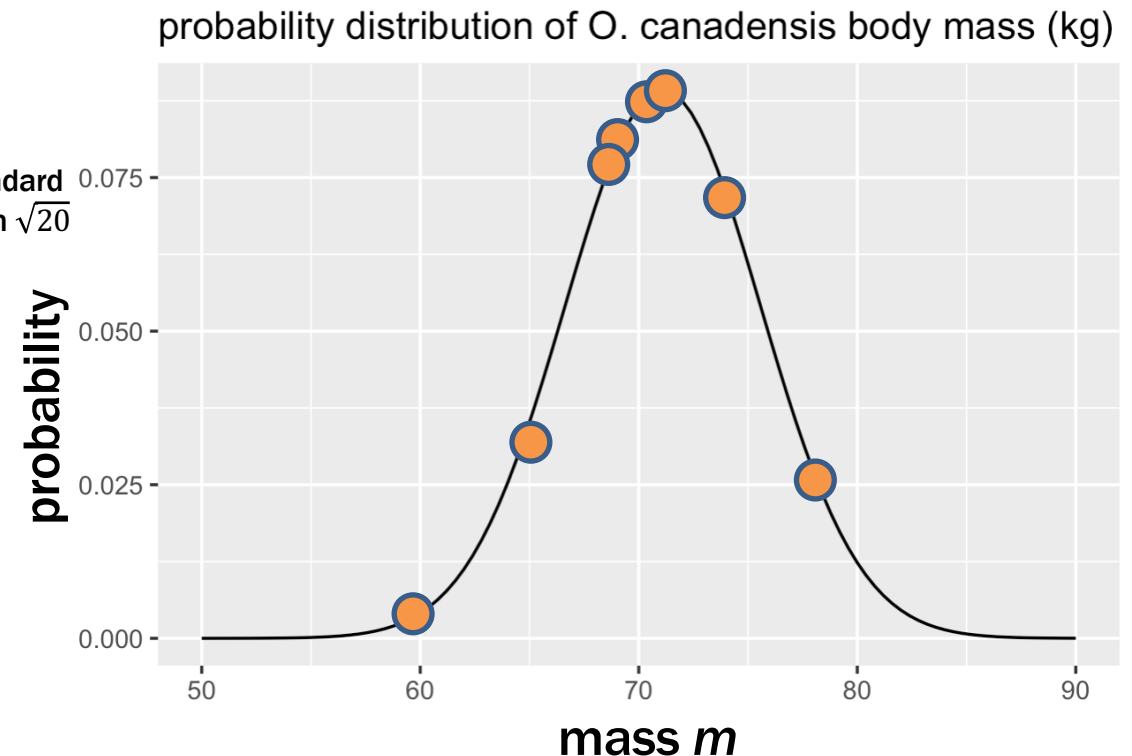
$m_i \sim N(71.1, \sqrt{20})$

mass of sheep i

with mean 71.1 and standard deviation $\sqrt{20}$

a normal distribution

is drawn from / is distributed as



Probability distribution: Normal (Gaussian)

```
> dnorm(74,mean=71.1,sd=(sqrt(20)))  
[1] 0.07229107  
> dnorm(90,mean=71.1,sd=(sqrt(20)))  
[1] 1.180421e-05  
> dnorm(71,mean=71.1,sd=(sqrt(20)))  
[1] 0.08918391  
>
```

```
> rnorm(1,mean=71.1,sd=(sqrt(20)))
[1] 64.41786
> rnorm(1,mean=71.1,sd=(sqrt(20)))
[1] 72.2535
> rnorm(1,mean=71.1,sd=(sqrt(20)))
[1] 71.86588
> rnorm(1,mean=71.1,sd=(sqrt(20)))
[1] 70.97046
> hist(rnorm(10000,mean=71.1,sd=(sqrt(20))))
```

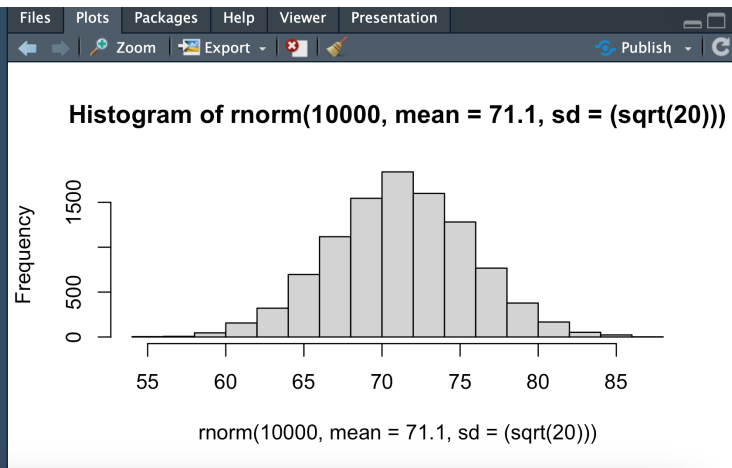
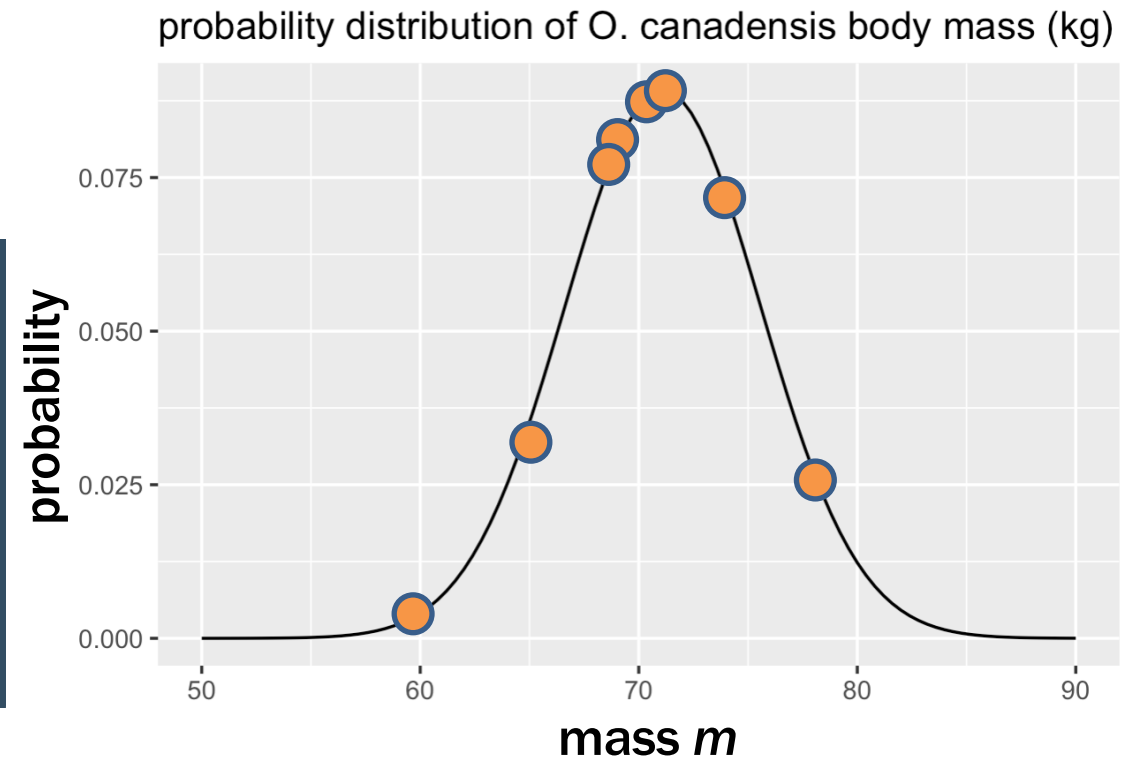


Diagram illustrating the components of the notation $m_i \sim N(71.1, \sqrt{20})$:

- m_i : mass of sheep i
- \sim : is drawn from / is distributed as
- N : a normal distribution
- 71.1 : with mean 71.1
- $\sqrt{20}$: and standard deviation $\sqrt{20}$



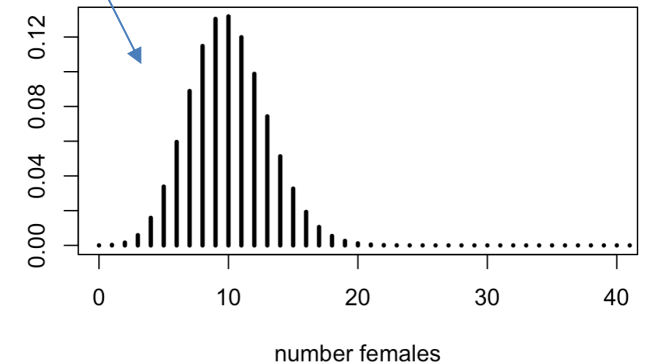
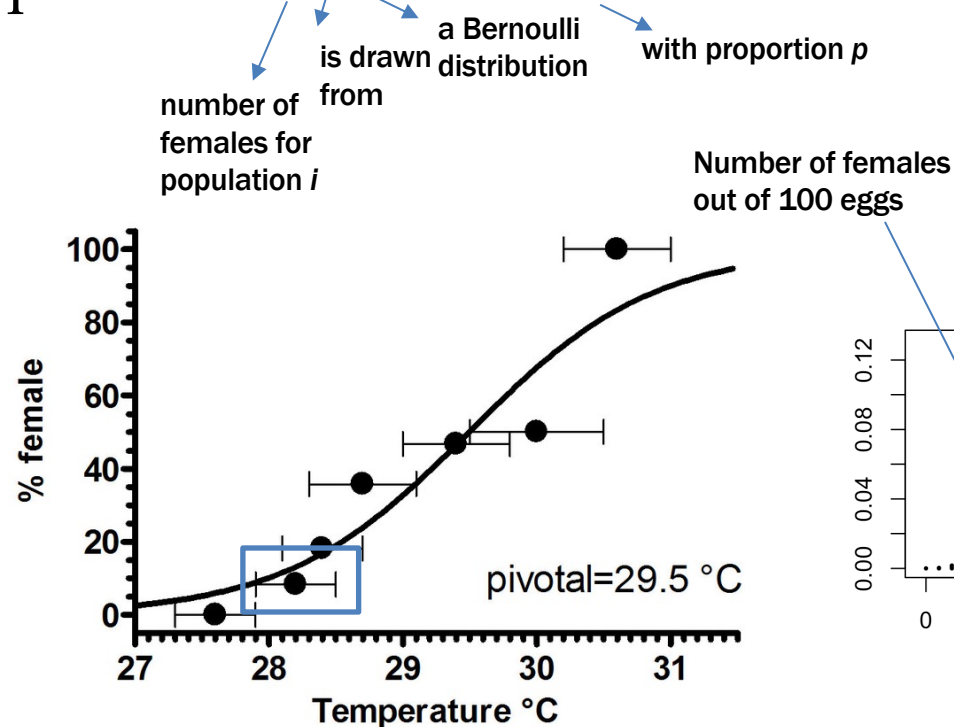
What is a statistical model?

Definition

- a mathematical model that embodies a set of statistical assumptions concerning the generation of data sampled from a larger population
- A statistical model represents, often in considerably idealized form, the data-generating process $y_i \sim \text{Bernoulli}(p)$

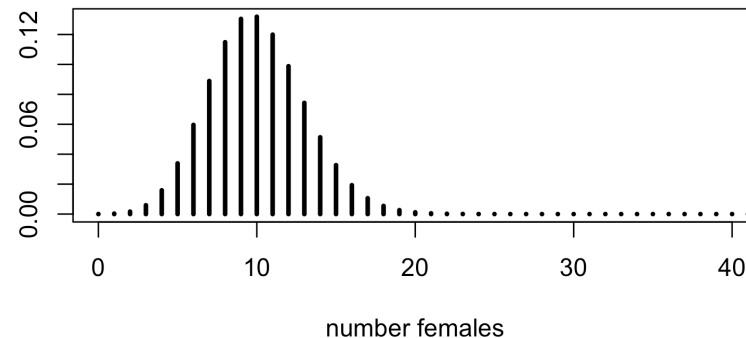
$$f_i \sim \text{Bernoulli}(0.1)$$

Green sea
turtle,
*Chelonia
mydas*



Probability distribution: Bernoulli ($n=1$), Binomial ($n > 1$)

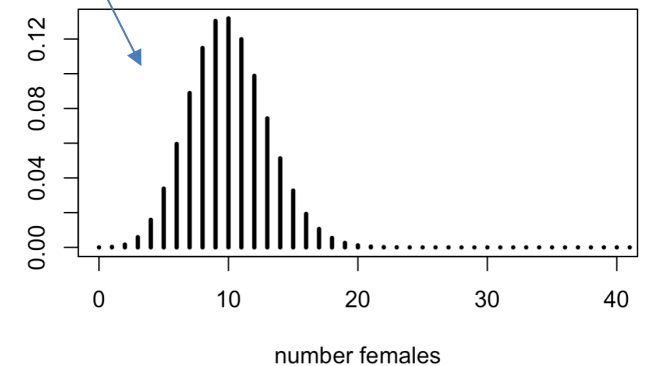
```
> rbinom(1,size=100,prob=0.1)
[1] 12
> rbinom(1,size=100,prob=0.1)
[1] 10
> rbinom(1,size=100,prob=0.1)
[1] 5
> rbinom(1,size=100,prob=0.1)
[1] 15
> rbinom(1,size=100,prob=0.1)
[1] 7
> dbinom(10,size=100,prob=0.1)
[1] 0.1318653
> dbinom(40,size=100,prob=0.1)
[1] 2.470212e-15
>
```



$$f_i \sim \text{Bernoulli}(0.1)$$



Number of females
out of 100 eggs



Probability distribution: Poisson

```
> load("~/OneDrive/Documents/UDE/teaching/2023_summer/L5_Regres  
sion1/germany.rda")  
> head(germany)  
      Date Season      home  
1 1963-08-24   1963      Werder Bremen  
2 1963-08-24   1963      1. FC Saarbrücken  
3 1963-08-24   1963      TSV 1860 München  
4 1963-08-24   1963 Frankfurter SG Eintracht  
5 1963-08-24   1963      FC Schalke 04  
6 1963-08-24   1963      Preussen Münster  
      visitor FT hgoal vgoal tier division  
1 Borussia Dortmund 3-2      3      2      1      1  
2      1. FC Köln 0-2      0      2      1      1  
3 Eintracht Braunschweig 1-1      1      1      1      1  
4 1. FC Kaiserslautern 1-1      1      1      1      1  
5 VfB Stuttgart 2-0      2      0      1      1  
6 Hamburger SV 1-1      1      1      1      1  
> hist(germany$hgoal)  
> unique(germany$Season)  
 [1] 1963 1964 1965 1968 1969 1970 1971 1972 1973 1974 1975  
[12] 1976 1977 1978 1979 1980 1981 1982 1984 1985 1986 1987  
[23] 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998  
[34] 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009  
[45] 2010 2011 2012 2013 1983 1967 1966 2014 2015 2016  
>
```

```
> fitdistrplus::fitdist(data=germany$hgoal,distr="pois")
Fitting of the distribution ' pois ' by maximum likelihood
Parameters:
      estimate   Std. Error
lambda 1.819764 0.007313536
> hist(rpois(10000,lambda=1.819764))
>
>
>
>
>
>
>
>
>
>
```

The screenshot shows the RStudio interface. The top console displays the command `int [1:101] 0 1 2 3 4 5 6 7 8 9 ...`. The top menu bar includes **Files**, **Plots**, **Packages**, **Help**, **Viewer**, and **Presentation**. The toolbar contains icons for navigation, zoom, export, and publish. The main plot area displays a histogram titled **Histogram of germany\$goal**. The y-axis is labeled **Frequency** and ranges from 0 to 15,000. The x-axis is labeled **germany\$goal** and ranges from 0 to 12. The histogram shows a right-skewed distribution with a peak frequency of approximately 16,000 for the first bin (0-1 goals).

germany\$goal (Bin)	Frequency
0-1	16000
1-2	10000
2-3	6000
3-4	3000
4-5	1500
5-6	1000
6-7	500
7-8	200
8-9	100
9-10	50
10-11	20
11-12	10

$$g_i \sim \text{Poisson}(\lambda)$$

A histogram showing the frequency distribution of 10,000 random samples drawn from a Poisson distribution with a mean (lambda) of 1.819764. The x-axis is labeled 'rpois(10000, lambda = 1.819764)' and ranges from 0 to 8. The y-axis is labeled 'Frequency' and ranges from 0 to 2500. The distribution is unimodal and slightly right-skewed, with the highest frequency occurring at x=1 (approximately 2800) and x=2 (approximately 2600). The frequency drops significantly for x=3 (approximately 1500) and continues to decrease for higher values of x.

Value (x)	Frequency (y)
0	~1500
1	~2800
2	~2600
3	~1500
4	~700
5	~300
6	~100
7	~20
8	~10

$$g_i \sim \text{Poisson}(\lambda = 1.819)$$

Probability distributions: diverse families

