Generalized Linear Models

"The trouble with normal is that it always gets worse"

Samuel Robinson, Ph.D.

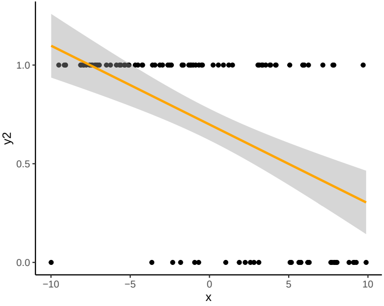
November 19, 2020

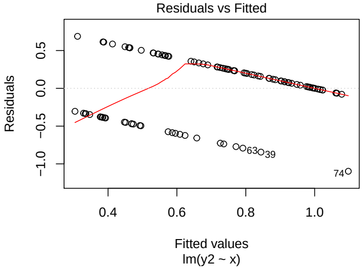
Motivation

What are Generalized Linear Models? (GLMs)

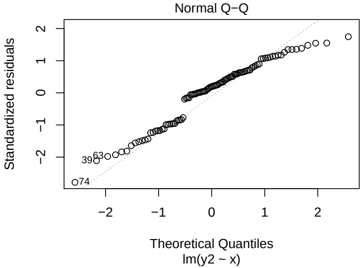
- Meet the exponential family
  - Normal, Binomial, Poisson
  - Negative Binomial, Beta, Gamma
- Tricksy hobbitses!
  - Zero-inflated models, occupancy models

Problem: not everything is normal

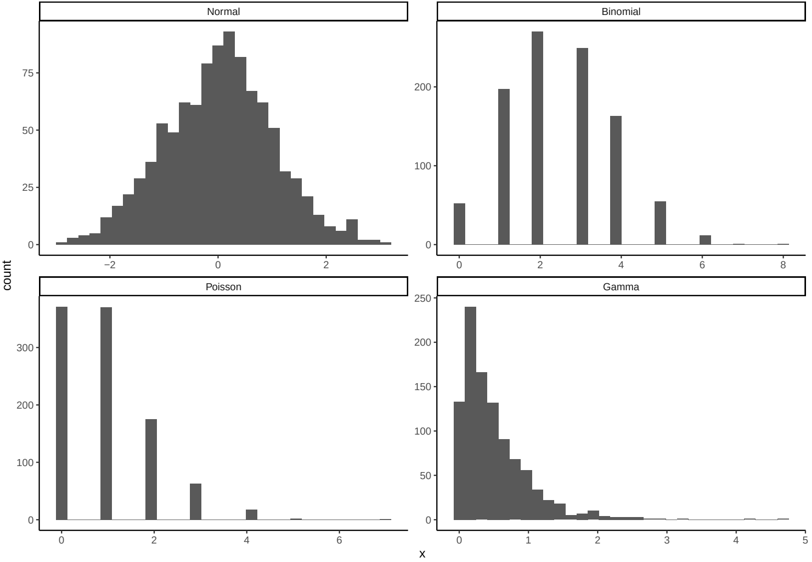




- Some types of data can never be transformed to make the residuals normal
- Solution: **use the distribution that generates the data!**



But how do I know which distribution to use?



Time to meet the family!

The Normal Distribution (aka *Gaussian*)

- Imagine many random + and - numbers added together
- If you do this *many* times:
  - Most cancel out (somewhere around 0)
  - Few are far away from 0 (tails of distribution)
- Common in nature, because of many small + and - factors adding together
  - e.g. Height is driven by many sets of genes

A Galton Board in action:

The Normal Distribution - scary math!

- 2 parameters: mean (μ) and standard deviation (σ)

$$p(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

- Probability distribution function (PDF) for the Normal distribution
- Tells you about the probability of getting some number *given* μ and σ

Example: what is the probability of getting a 4, if the mean is 5 and SD is 1?

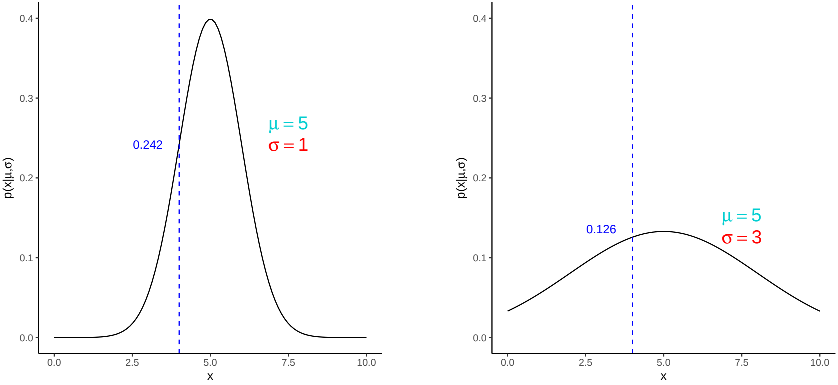
$$p(4|5,1) = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{4-5}{1})^2}$$
$$= \sim 0.24$$

In R, this is easy:

```
#d stands for "density"
dnorm(x=4,mean=5,sd=1)
```

```
## [1] 0.2419707
```

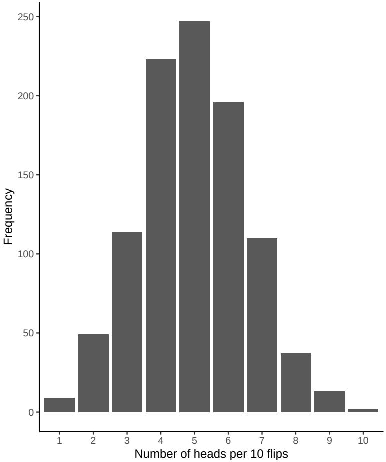
The Normal Distribution



- Probability of x changes with μ and σ
- Left: $\sigma = 1$, Right: $\sigma = 3$

The Binomial Distribution

- Imagine you have 10 coins, and you flip them all
- If you do this *many* times:
  - Most will be about 5 heads/tails
  - Few will be 1 head, 9 tails (or reverse)
- Common in nature where outcomes are binary
  - e.g. 10 seeds from a plant, how many will germinate?
- If $N = 1$, this is called a *Bernoulli trial*



The Binomial Distribution - scary math!

- 1 parameter: probability of success (ϕ), plus...
- Number of "coin flips" (N)

$$p(x|\phi, N) = \binom{N}{x}\phi^x(1-\phi)^{N-x}$$

- Probability mass function (PMF); density = continuous
- Tells you about the probability of getting x "successes" *given* ϕ and N

Example: what is the probability of getting 4 successes, if ϕ is 0.25 and N is 15?

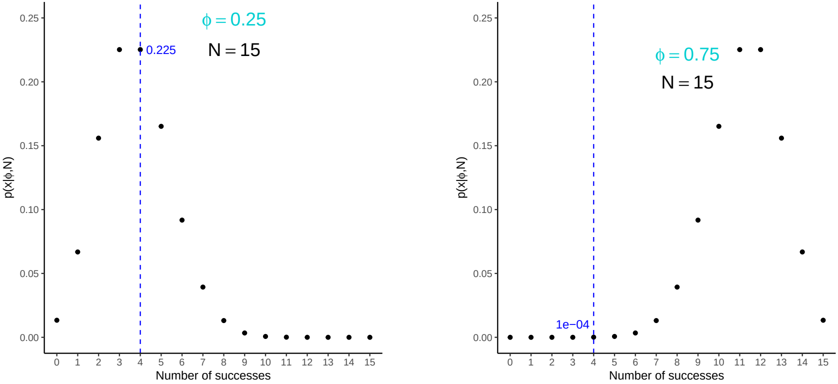
$$p(4|0.25, 15) = \binom{15}{4}0.25^4(1-0.25)^{15-4}$$
$$= \sim 0.23$$

In R, this is easy:

```
dbinom(x=4,size=15,prob=0.25)
```

```
## [1] 0.2251991
```

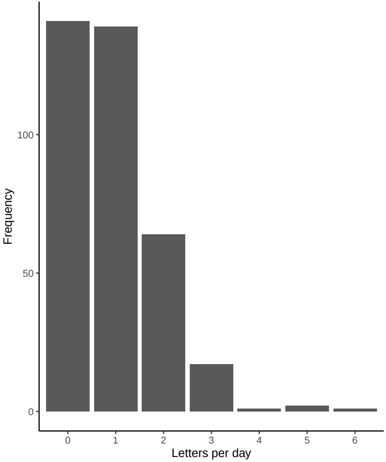
The Binomial Distribution



- Probability of x "successes" changes with ϕ and N

The Poisson Distribution

- Imagine a rare event (e.g. getting a non-junk mail letter)
- If you record the number of events every day:
  - Most days, you'll get 0 or maybe 1 letter
  - On some rare days, you'll get 3 or 4 letters
- Common in nature where rare events are measured over time/space:
  - e.g. Number of bats caught in a net (per night)



- Equivalent to Binomial distribution, where N is unknown

The Poisson Distribution - scary math!

- 1 parameter: rate parameter (λ)

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

- Probability mass function (PMF)
- Tells you about the probability of getting x counts *given* λ

Example: what is the probability of getting 2 counts, if λ is 1?

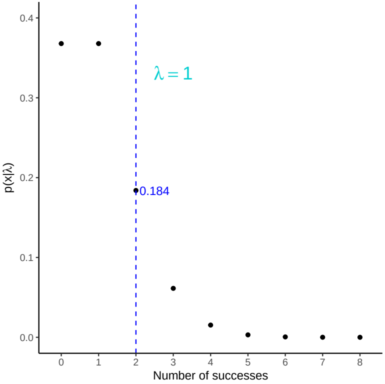
$$p(2|1) = \frac{1^2 e^{-1}}{2!}$$
$$= \sim 0.18$$

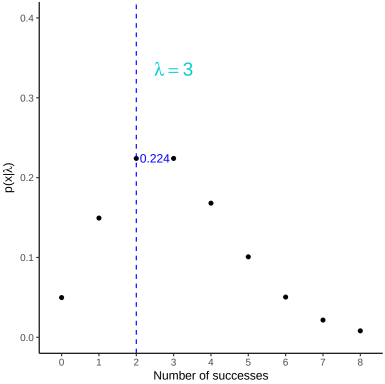
In R, this is easy:

```
dpois(x=2,lambda=1)
```

```
## [1] 0.1839397
```

The Poisson Distribution





- Probability of x counts changes with λ