September 26th – Variability, distributions, clarification and catching up

Quantitative Thinking in the Life Sciences

Today

- Assignment 3 R code revisited
- Distributions & Variability!
- Assignment # 4
- More R fun!
  - Chapter 4 (or other) questions?
  - Chapter 5 Elements and more loops

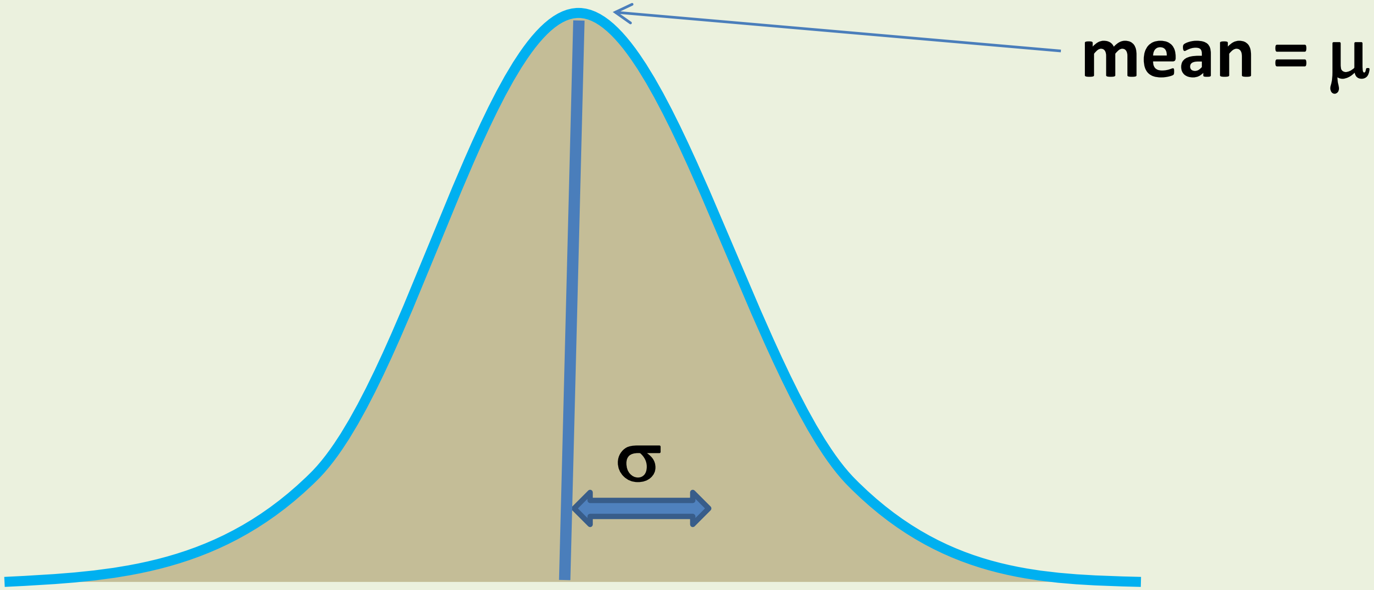
Housekeeping

Jumping right in!
Assignment 3 R code review

- To R we go!
  - Dropbox\\Quantitative Thinking\\Sept 26 notes_assignment 3 r code revisited.R

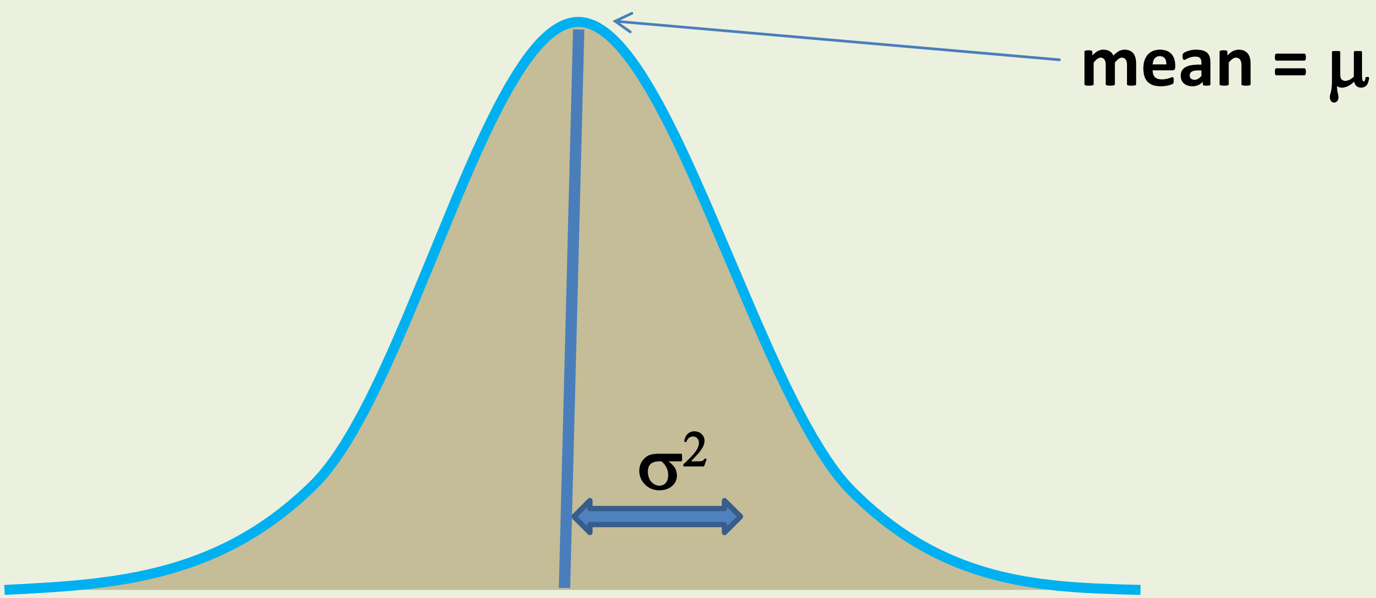
Variance (σ)

- Expected difference from the mean (the mean of the difference from the mean)
- On average (mean) how far are data expected to be away from the mean value



Standard Deviation(σ^2)

- Variance squared (Why?)



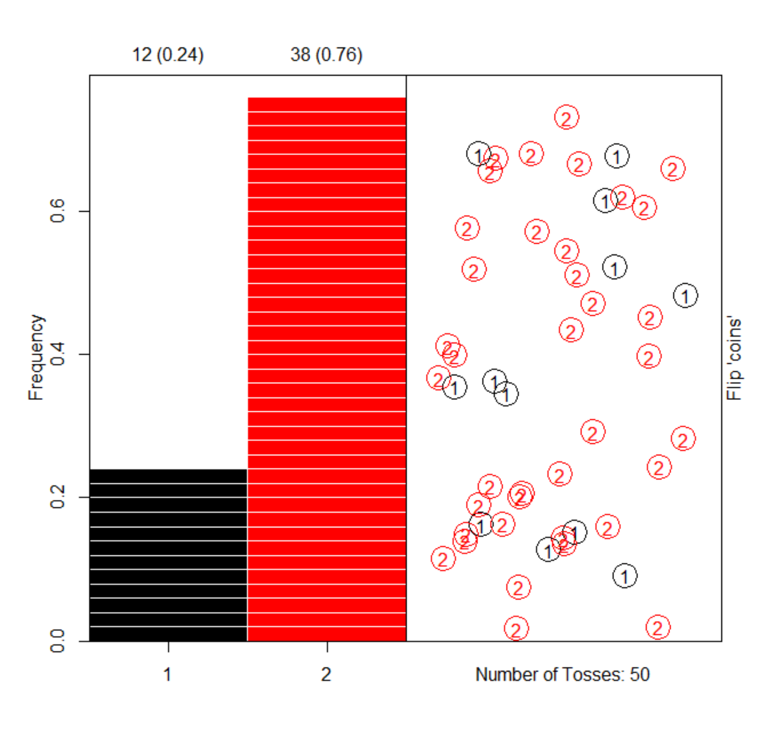
Standard Deviation vs Standard Error

- Standard deviation is a measure of the variability in your true population (frequently unknown)
- Standard error is an estimate of the variability in your measured population
  - Standard error approaches the standard deviation with increased sample size

Why do we care about the distribution type / shape?

- What should the system look like if there is no effect of our variables?

Binomial distribution

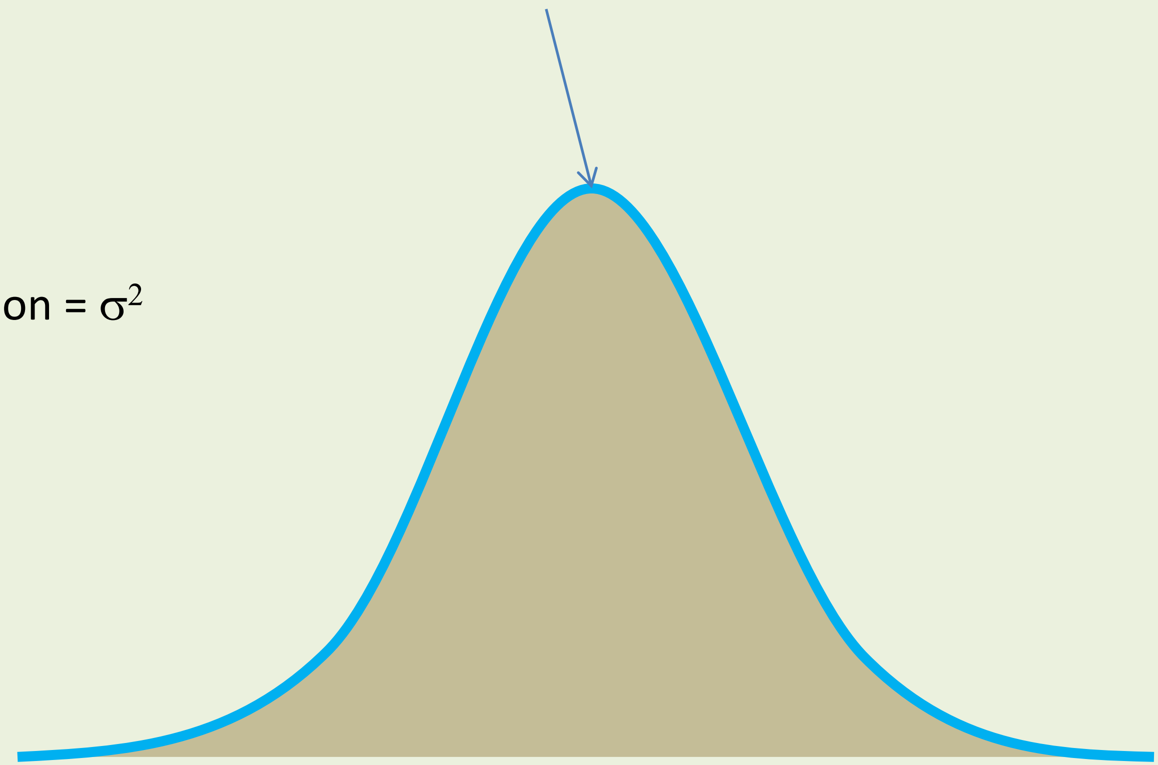


Coin flip – we assume that coin flips have a 50% chance of landing on heads (p = 0.5) and a 50% chance of landing on tails (p = 0.5), flips are independent

We can test and possibly reject the assumptions (possibly not independent, possibly not evenly weighted)

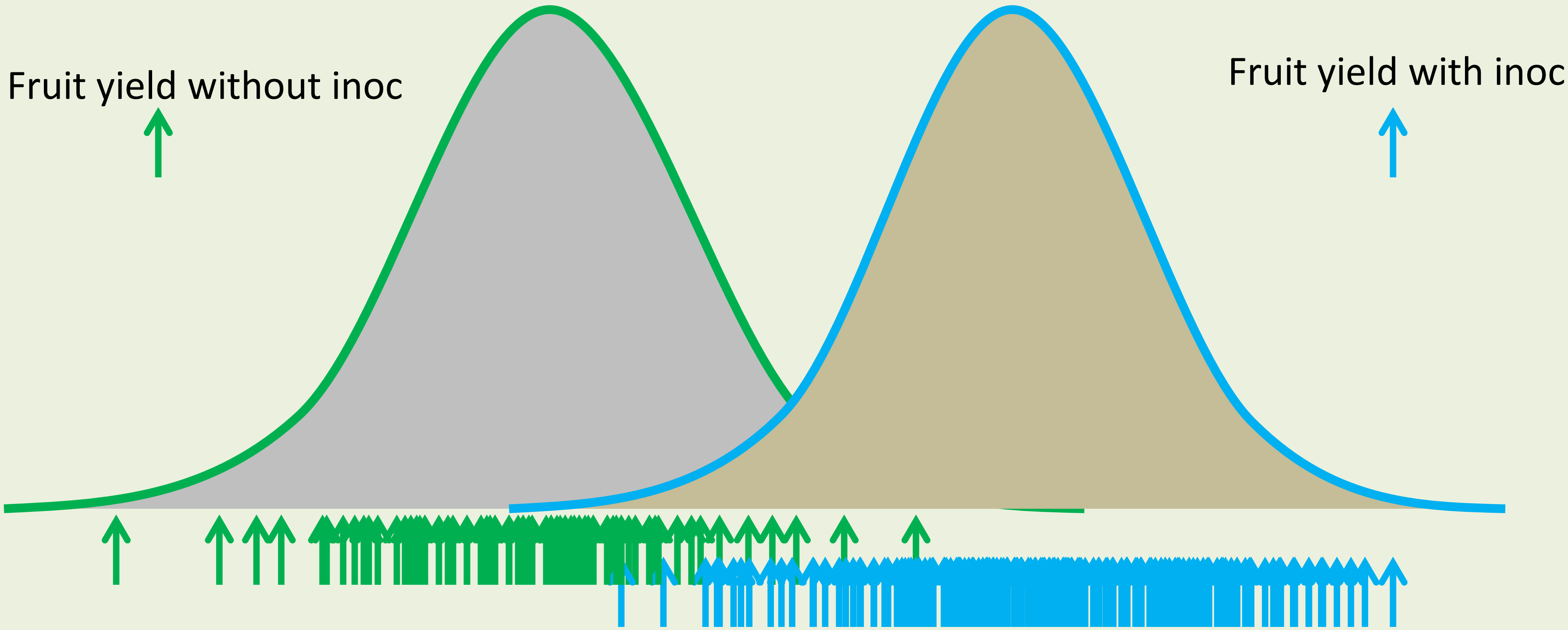
Normal probability distribution: Probability DENSITY function (PDF)

mean = μ



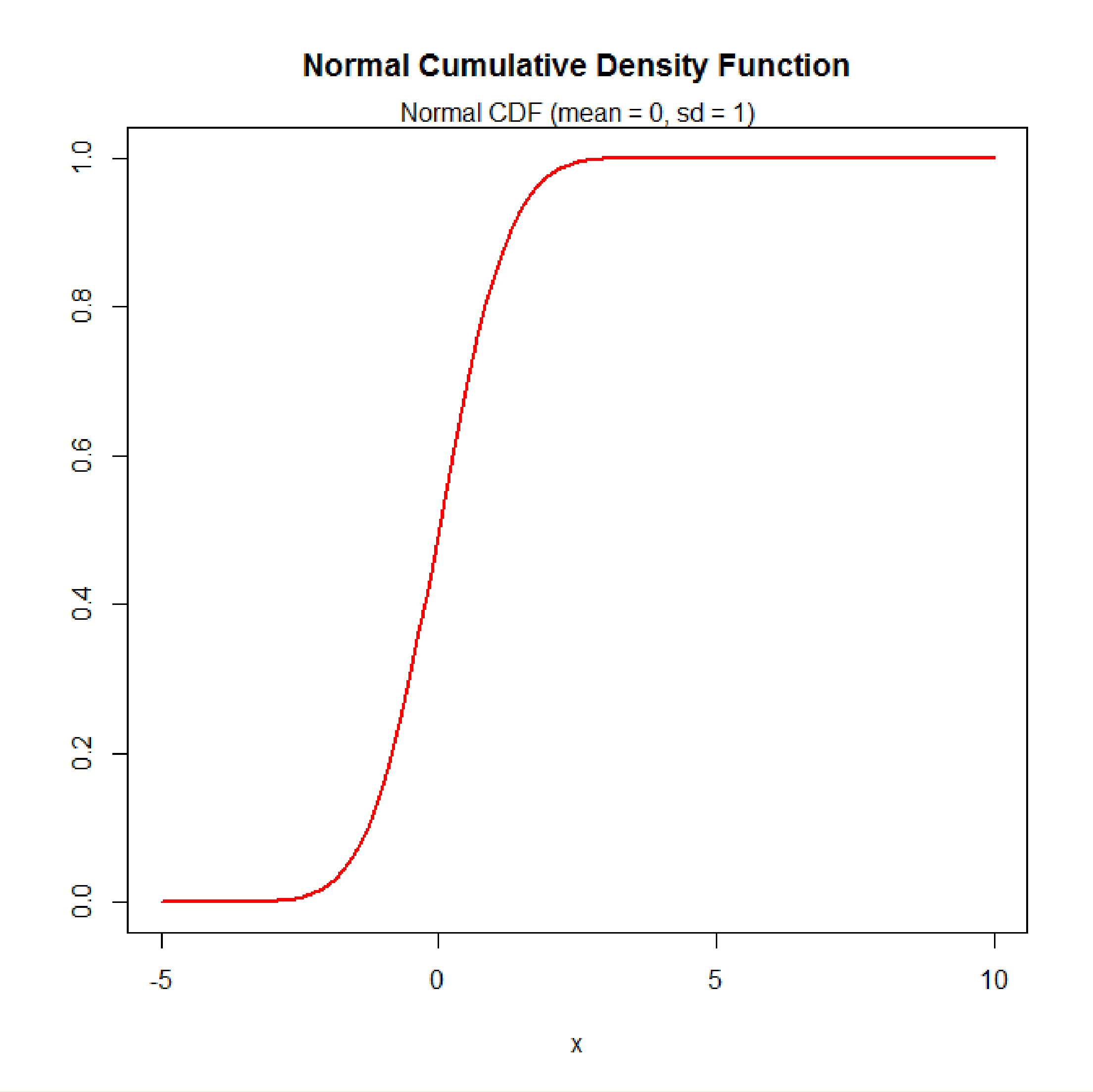
Independent
Mean = μ
Variance = σ
Standard deviation = σ^2

Testing assumptions of normality



Test to see if the assumptions have been violated

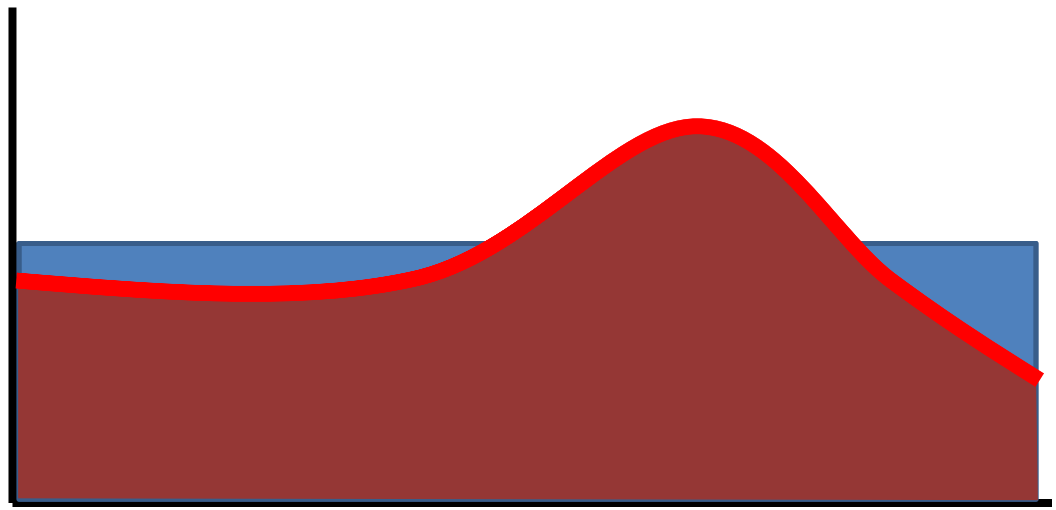
Normal Cumulative DENSITY function



Go to R

- Dropbox/Sept 26 notes_variability_distributions.R

Uniform probability DENSITY function (equal probability)

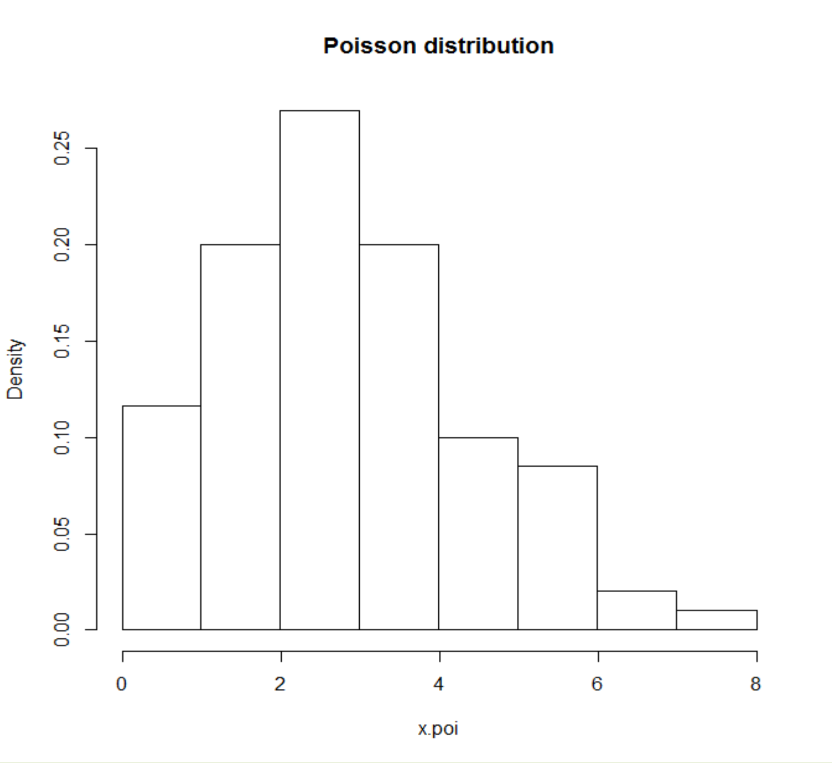


Poisson Distribution

- Mean number of events occurring is small relative to total possible occurrences
- Independent events
- Occurrences are random

Why do we care?

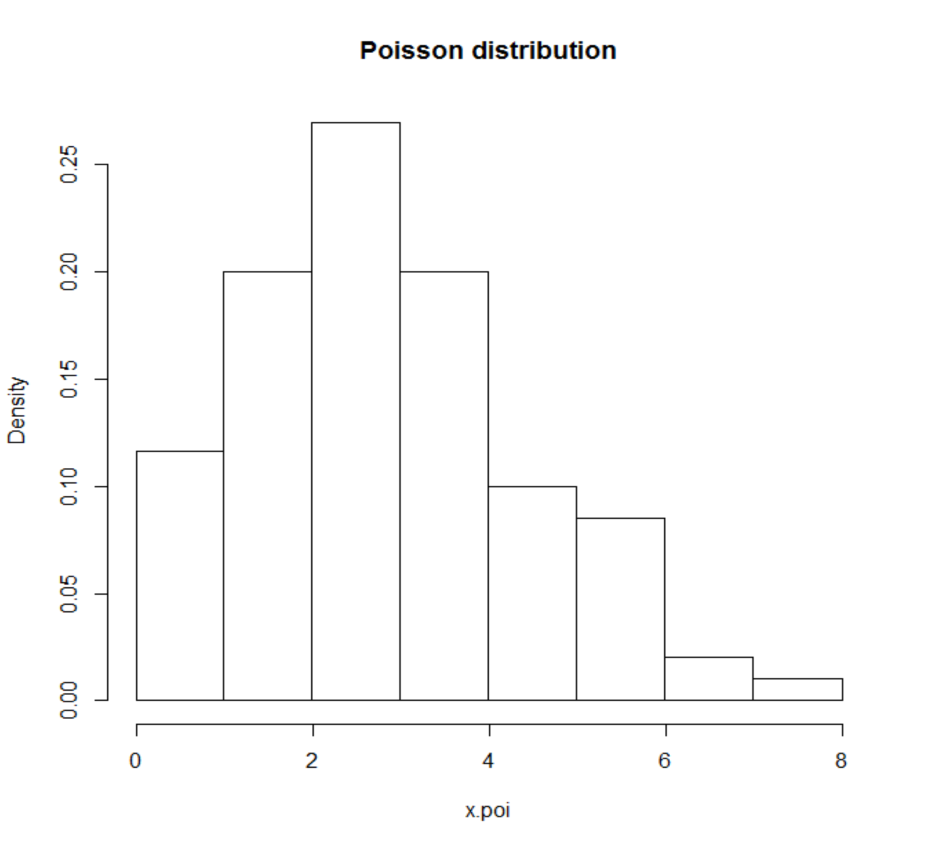
- If a goodness of fit test tells us that are distribution is NOT poisson distributed then we know that one of the assumptions has been violated!



> x.poi<-rpois(n=200,lambda=2.5)
> histogram(x.poi,nint=8,main="Poisson distribution")

Poisson Distribution

- Remember, we can test if something does not fit
- Difficult to say something definitely fits



E.g., Worms in a soil core

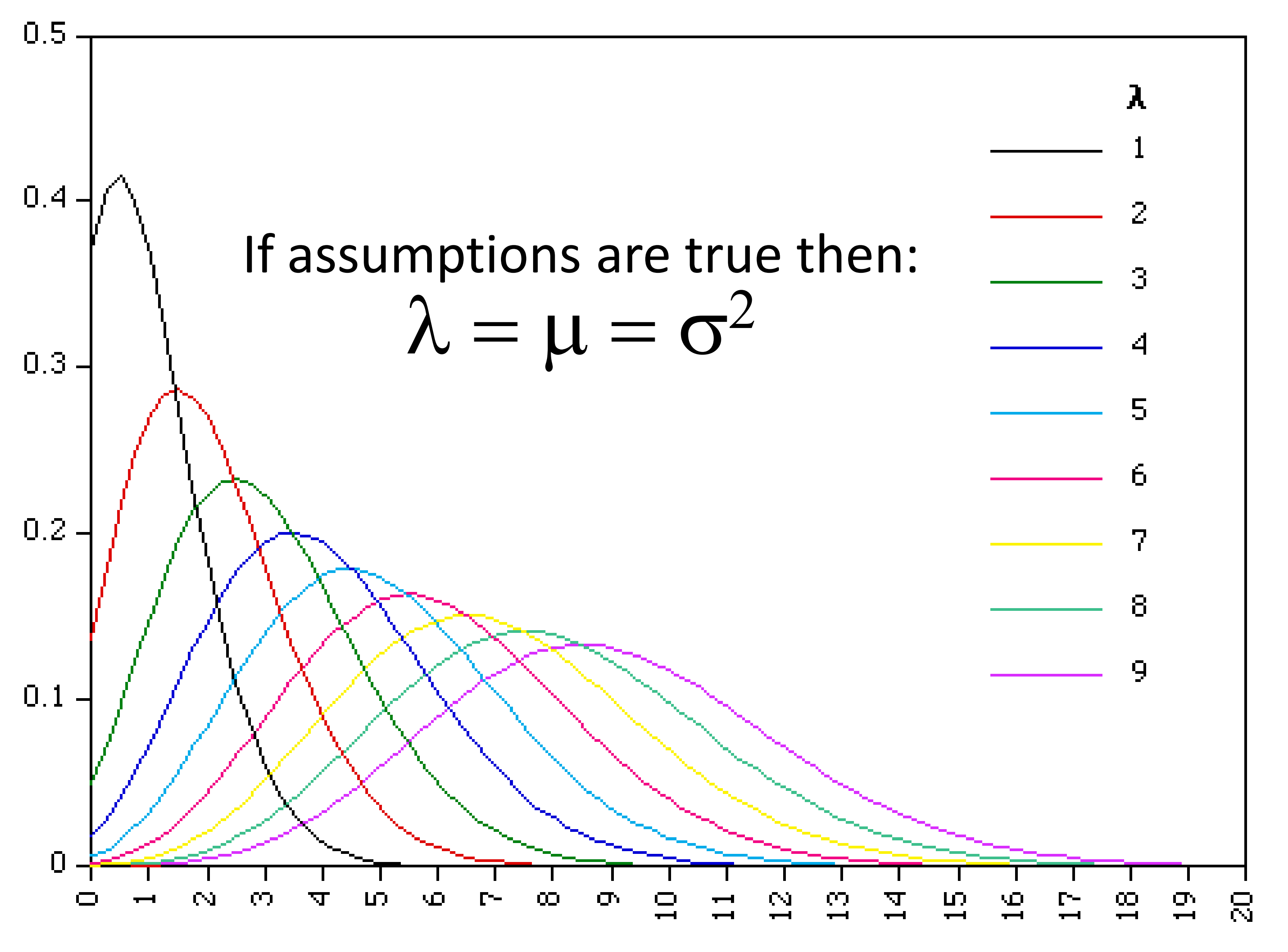
- Assume that nothing is influencing their presence (random)
- Lots could fit in a sample
- typically would only find a few
- the worms are independent

Check assumptions

- Find that variance is greater than mean, then worms might be clumping (on to next ? why?
- Where might you see the reverse (variance less than mean?)

Animals with territories, spacing of trees in a forest (light / shading issues)

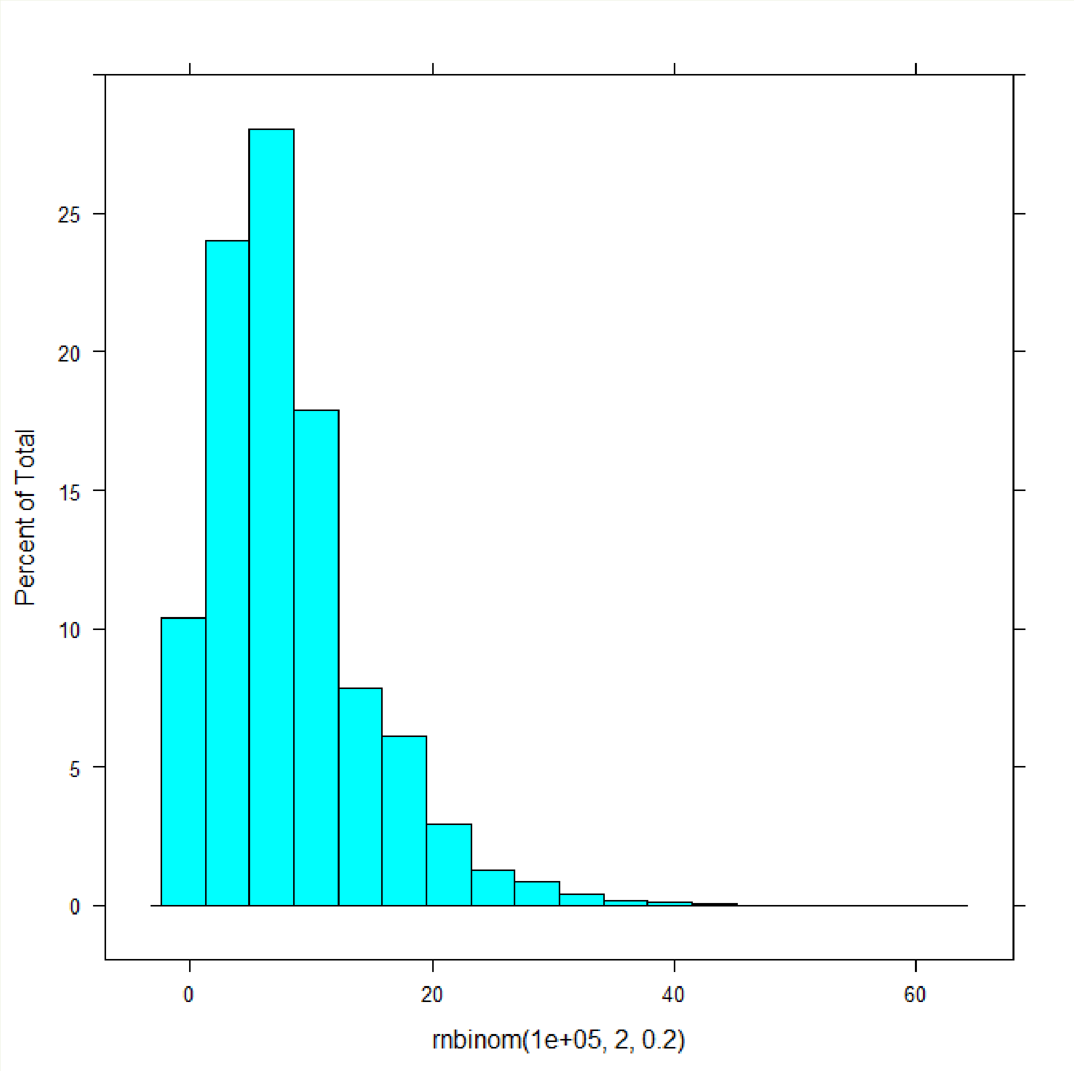
Poisson Distribution



Example: Negative binomial distribution:

Null model is that data (observations, individuals, etc) are clumped in time or space

So, if you want to REJECT that individuals are clumped, then you would need to create a null model that they are clumped.



Getting started on your assignment

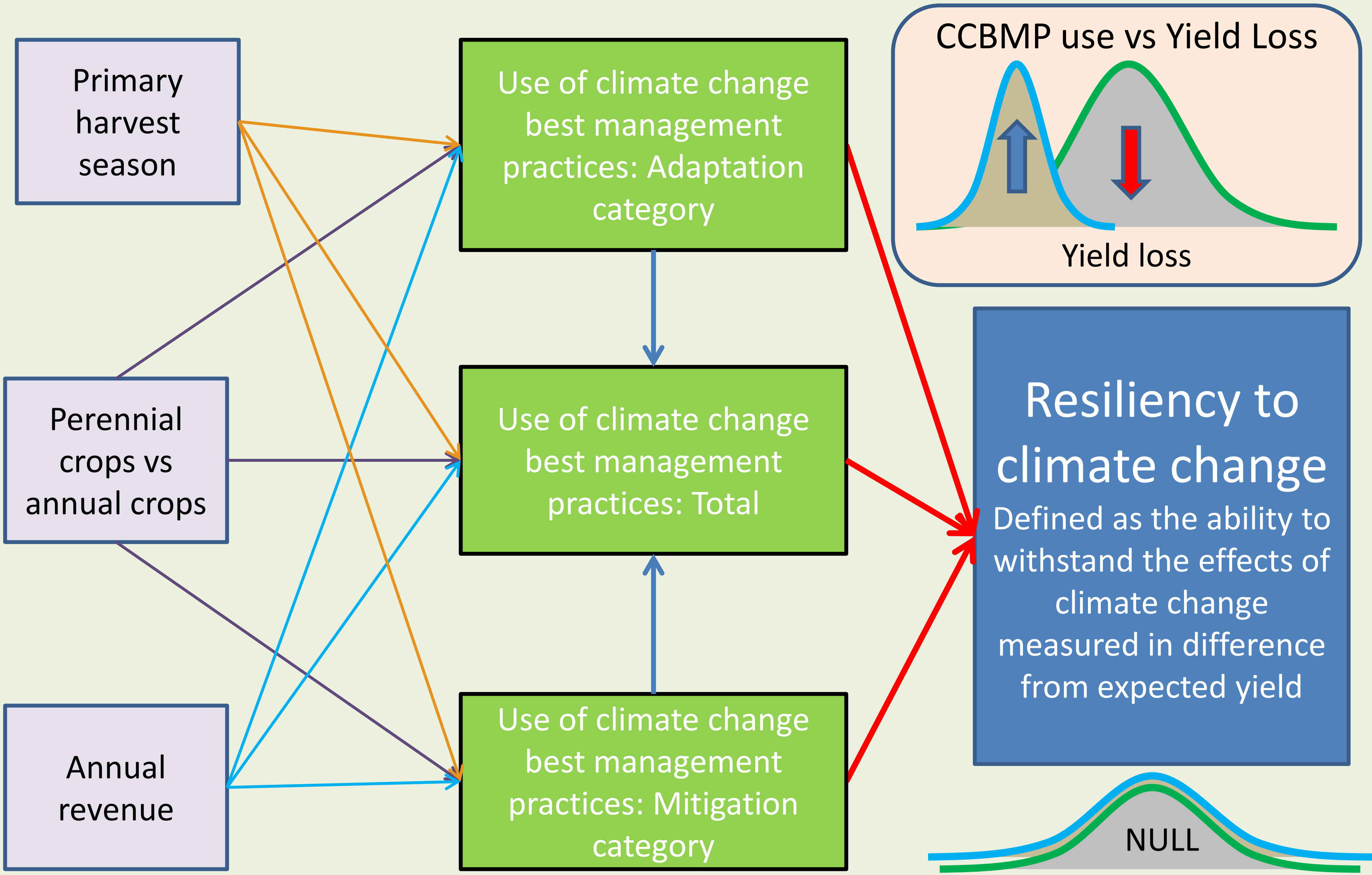
- Assignment # 4 is due on Oct 3rd
- Worth 50 points (Not 100 points – no simulation):
  - Three parts
    - Part 1: Distributions and variability for your system's factors/components/variables
      - Distributions and variability estimates
      - Relationships between it and connected factors
    - Parts 2 & 3: Chapters 4 and 5 in R
      - Chapter 4 was given to you last week
      - Chapter 5 will be given to you this week (distributions and variability)

Assignment 4: Part 1

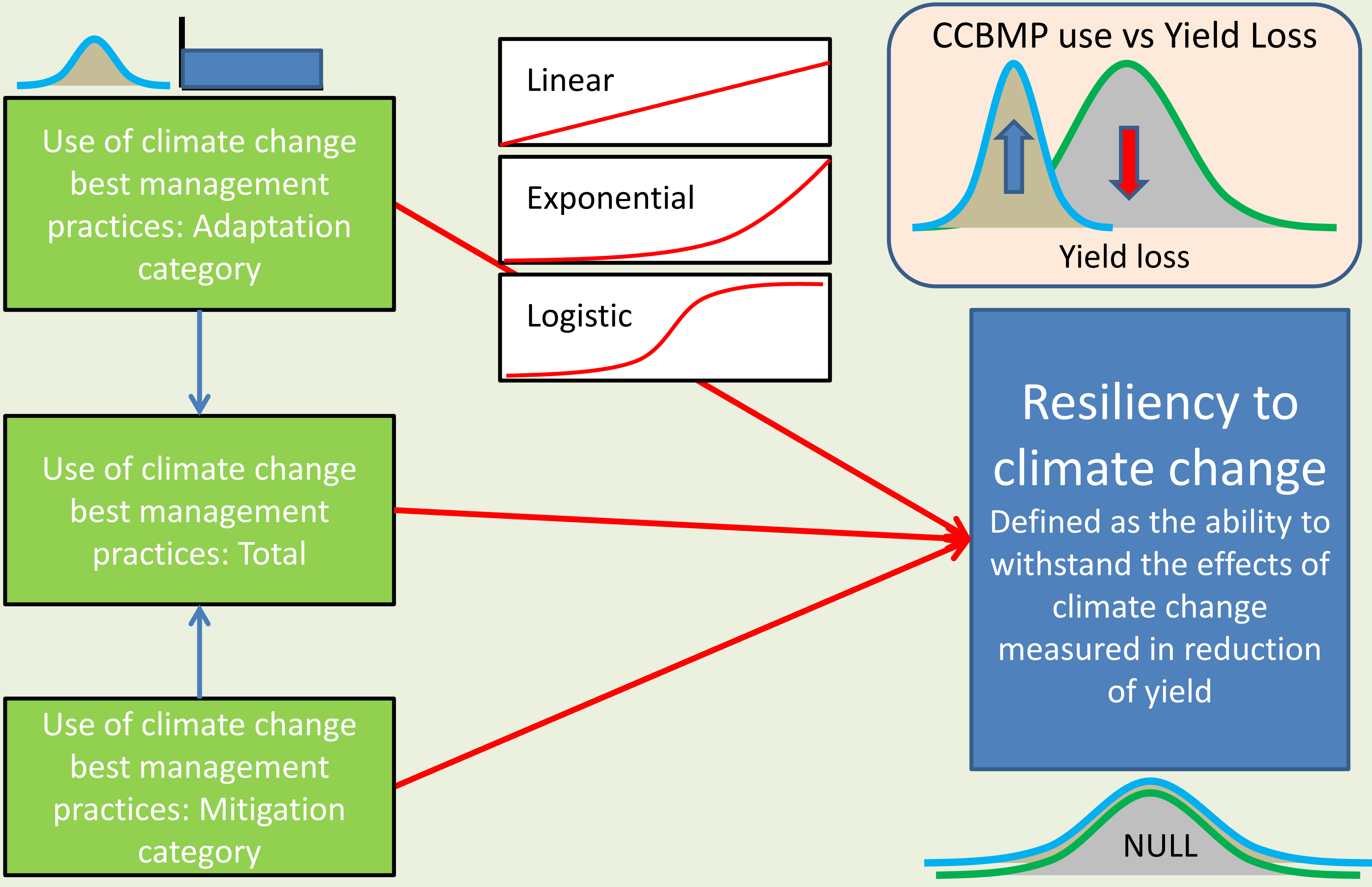
Use our concept maps to Distributions and variability for your system's factors/components/variables

Distributions and variability estimates
Relationships between it and connected factors

Cropping system resiliency to climate change



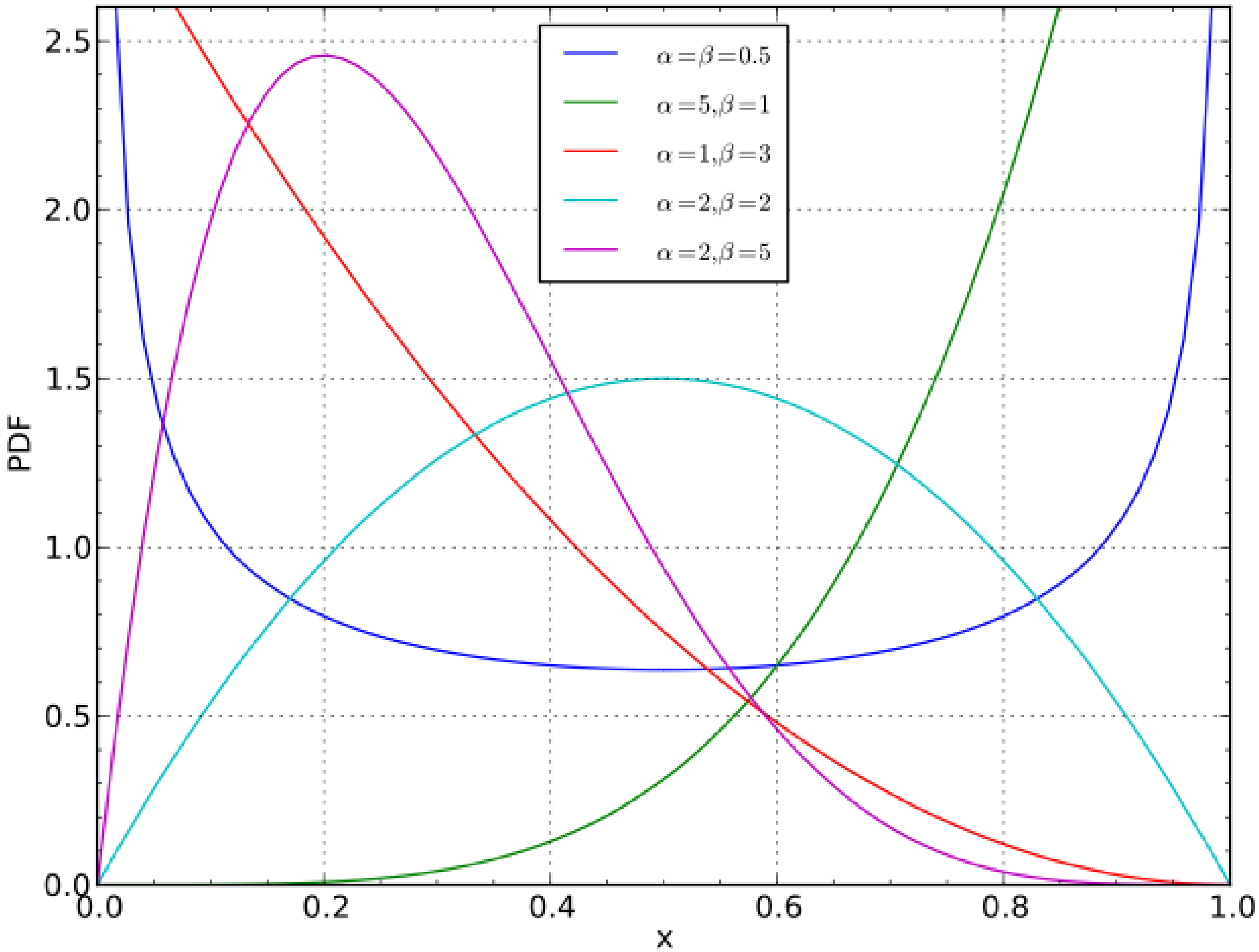
Cropping system resiliency to climate change



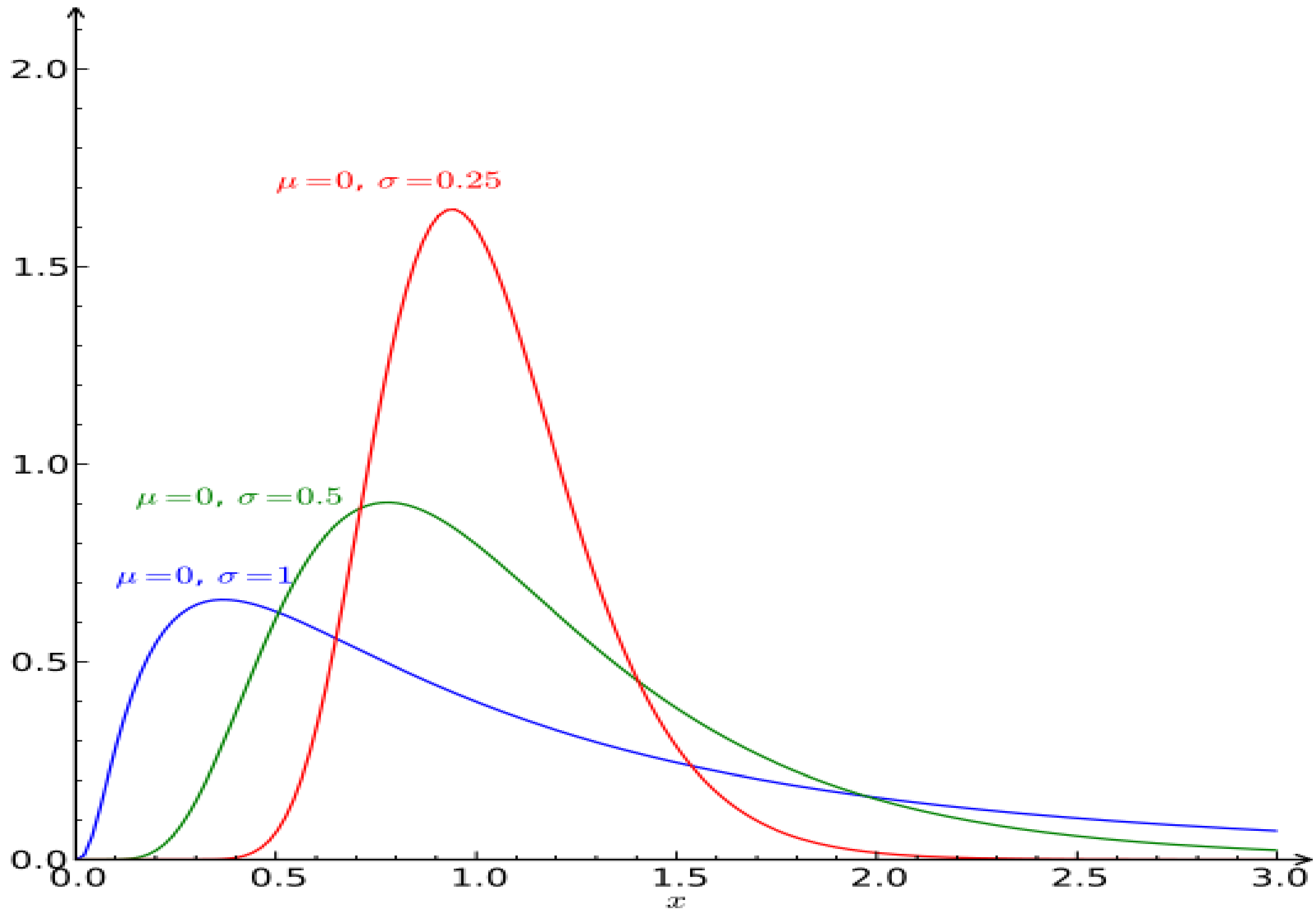
Endless fun with R!

- Questions from last week?
- This week – Chapter 5: Formulas and distributions
- jpeg('rplot_population_growth_exercise.jpg')

Beta distribution: probability distribution function
(α and β are shape parameters)



Lognormal Distribution: probability DENSITY function
(Always positive, its logarithm is normally distributed)



Gamma Distribution: probability distribution function
(k and θ are shape and scale parameters)