Happy Halloween

R Rocks!

Quantitative Thinking in the Life Sciences

October 31st – Simple linear regressions, multiple linear regressions, and non-linear models

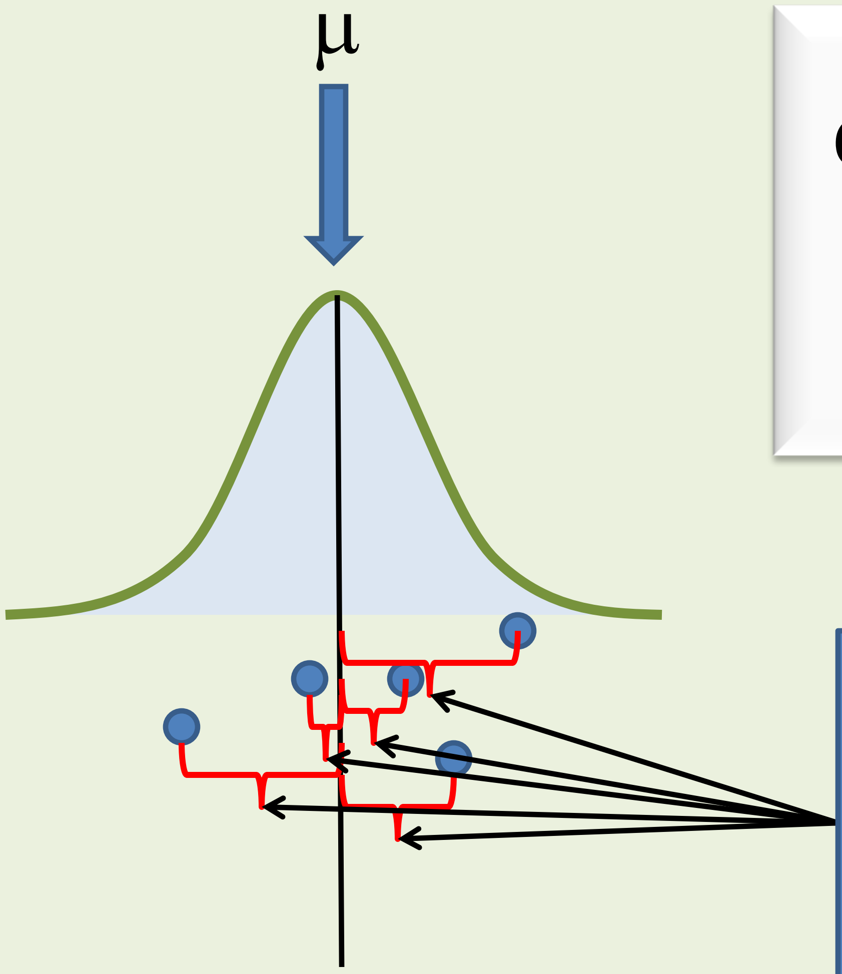
Today

- Simple Linear models
- Multiple regression
- Non-linear modeling
- Assignment C

Housekeeping

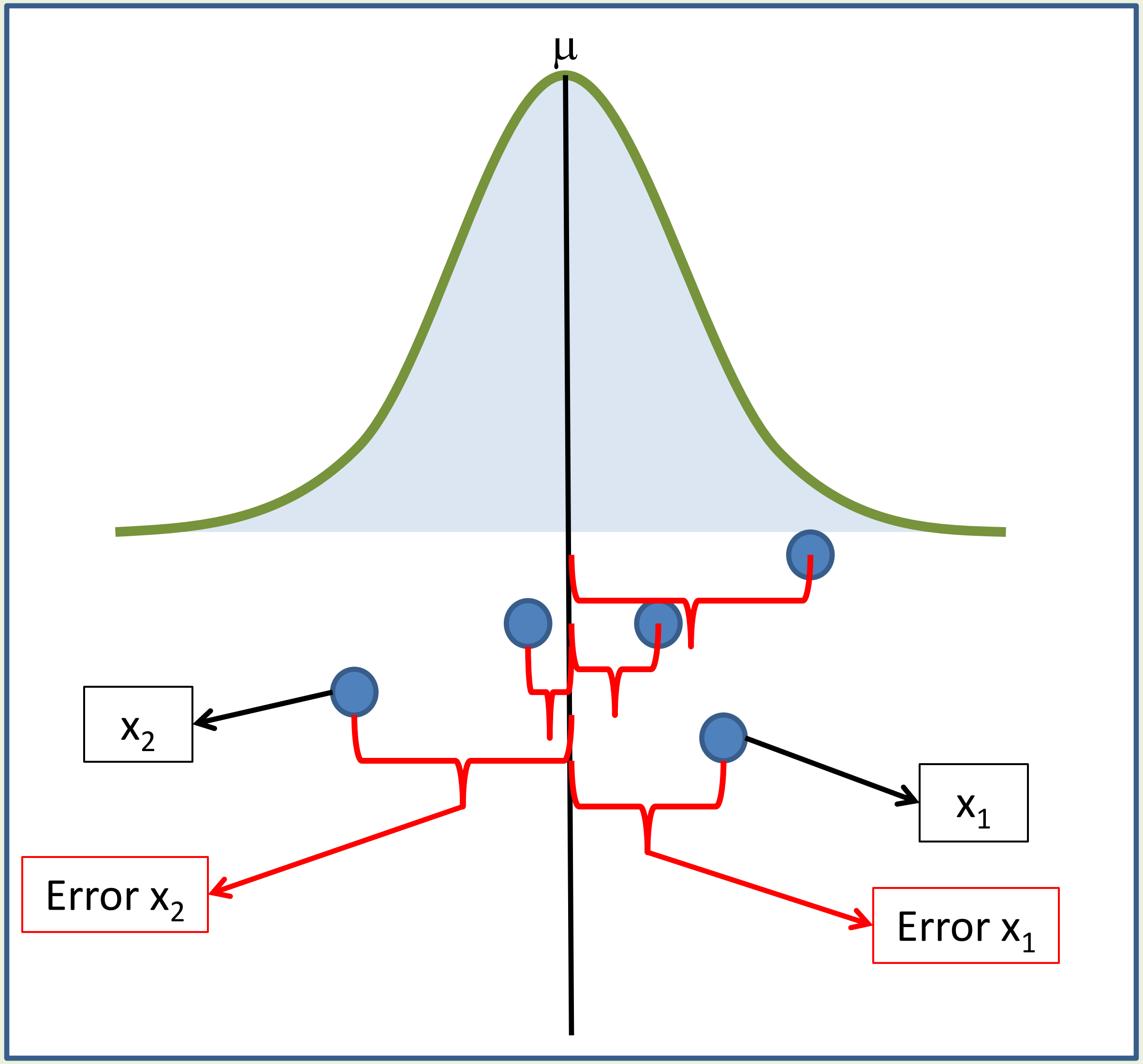
- November 14th absence
- After today, only three class sessions left
- Homework B is due today
  - First attempt at simulating your data distributions
- Homework C is due on Nov 7th
  - Chapter 8 R code: Modeling Elk populations in Rocky Mountain National Park

Developing a test statistic with a normal distribution



Calculate probability for each data point, each error

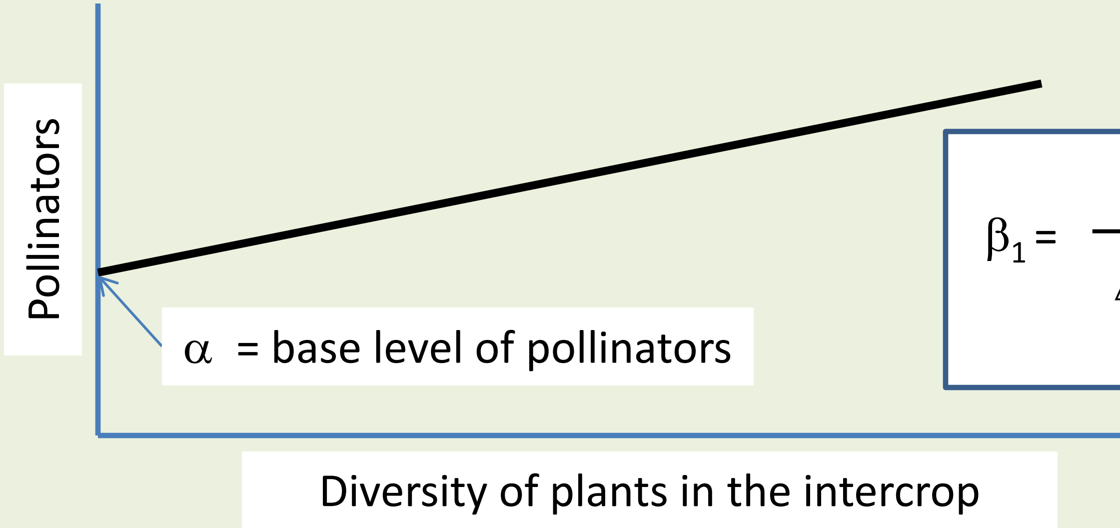
How far away are the data from their expected value(s)



$x_i - \mu$ = Distance or Error

Allows us to quantify the probability of x's occurrence

Linear model: $y = \alpha + \beta_1 * x$



$$\beta_1 = \frac{\Delta \text{Pollinators}}{\Delta \text{Intercrop.plant.diversity}}$$

$$\text{Pollinators} = \alpha + \beta_1 * \text{Intercrop.plant.diversity}$$

Does $\beta_1 = 0$?

Example in R!

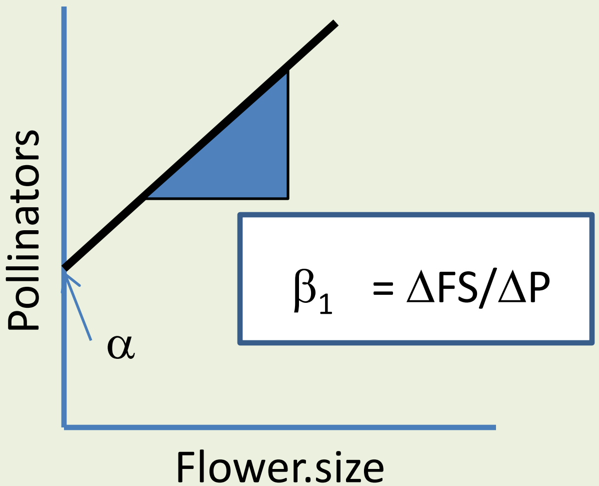
Linear regression: Assumptions about the data

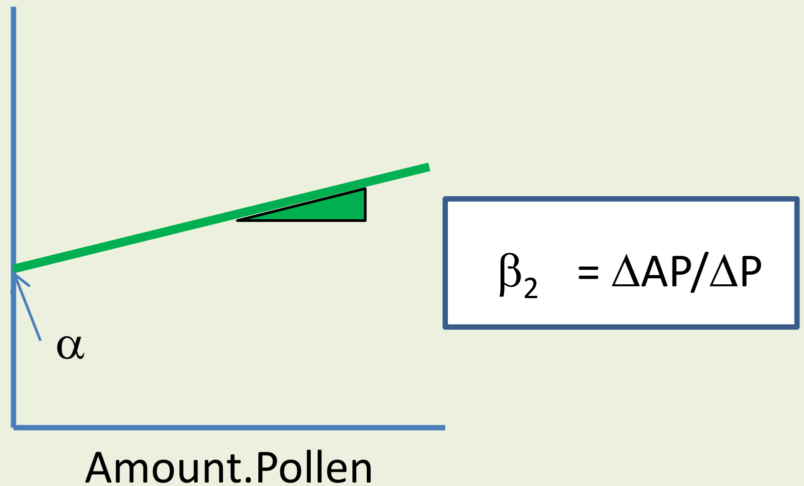
- There is no measurement error in your predictor variables (Ouch! – reinforces need for good design)
- Linearity (just witnessed in R)
- Constant variance in your errors (R example)
- Independence of errors in your response variable (y, e.g., # of pollinators)

Linear model multiple effects
multiple linear regression

$$y = \alpha + \beta_1 * x_1 + \beta_2 * x_2 \ldots$$

$$\text{Pollinators} = \alpha + \beta_1 * \text{Flower.size} + \beta_2 * \text{Amount.Pollen}$$



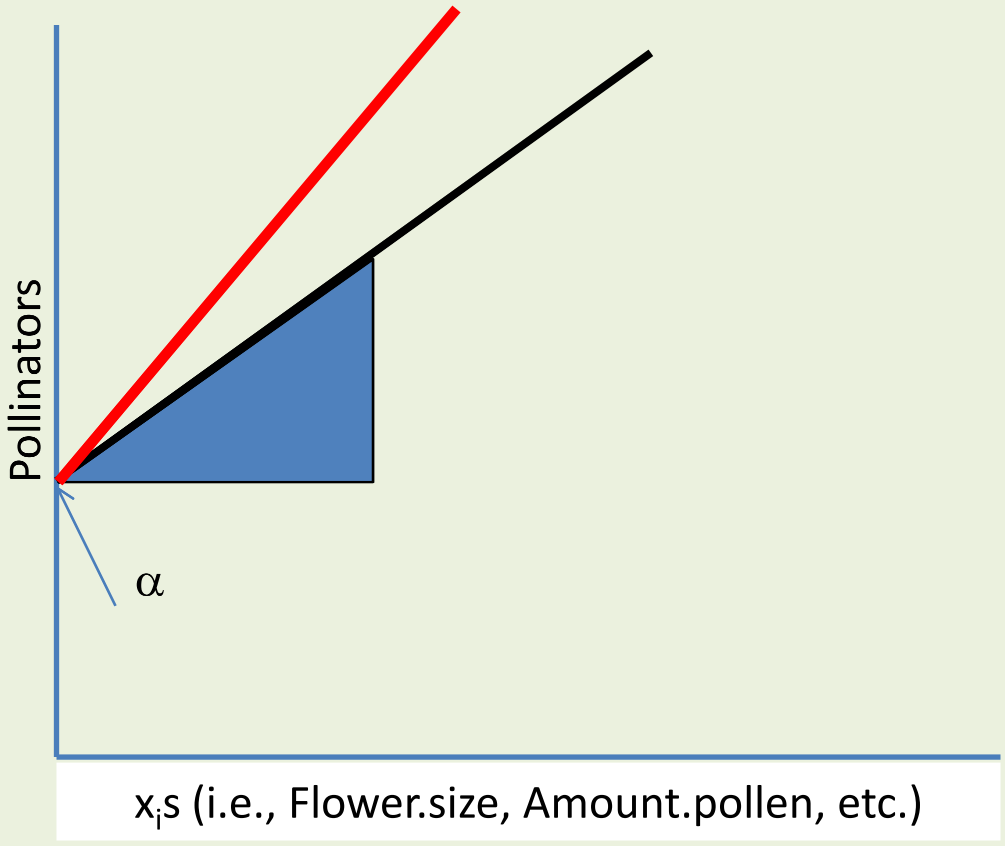


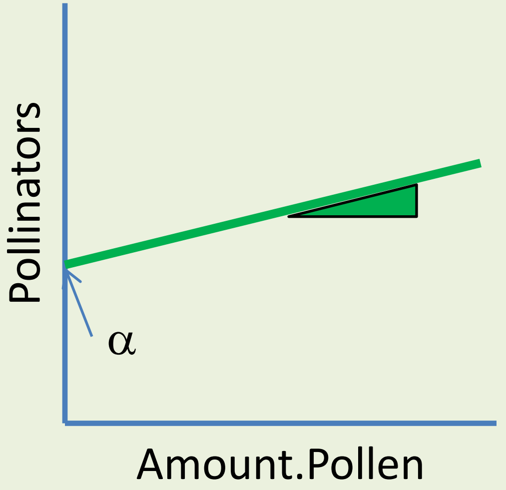
Does $\beta_1 = 0$?

Does $\beta_2 = 0$?

…and…

Pollinators = $\alpha + \beta_1$* Flower.size **+** β_2* Amount.Pollen





What is the prob that the overall model slope = 0?
Could the slope of the red line be equal to zero?

Time check!

- Load up r-squared example

Side note: R-squared

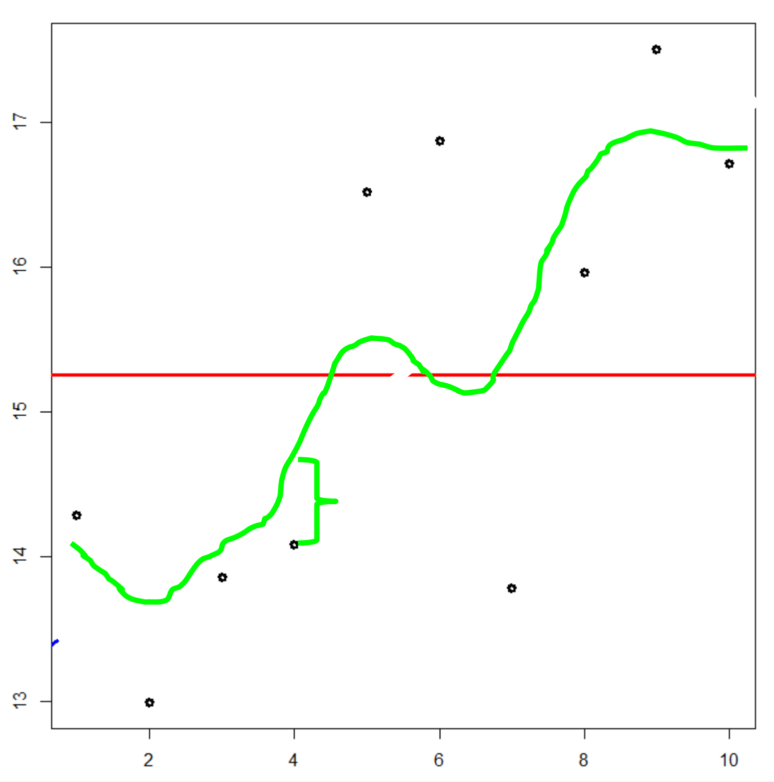
$$R^2 = 1 - \frac{SS_{error}}{SS_{total}}$$

SS_{total} = (y - mean(y))^2

SS_{total} = sum (red lines)^2

SS_{error} = error^2

SS_{error} = sum(blue lines)^2



Non-linear models

The concepts are the same as with linear models

- How much error is explained by your model
- How much variation exists in your system
- How much information is available for the model to work with
  - Really complex models that are fitting lots of parameters should fit the data better
  - Models with lots of parameters have reduce degrees of freedom (less information to work with)

Co2 model in R

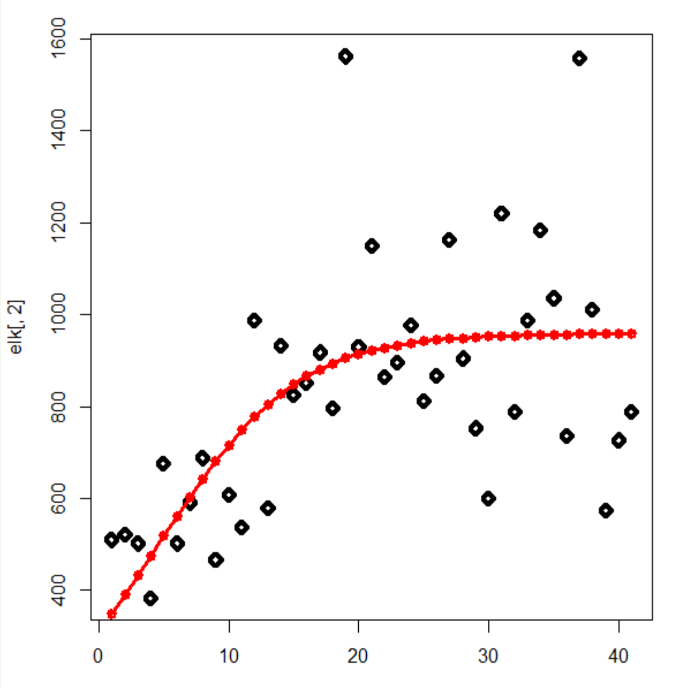
Assignment C

- Modeling the population dynamics of elk in Rocky Mountain National Park
- Data set of elk populations will be available on my website
  - Year
  - Population
  - Sample size (don’t worry about this for now)
- We will use the Logistic formula to model the population
  - Exponential growth, with a growth rate r
  - With a population limited by a carrying capacity K

Assignment C

$$N_{(time)} = N_{(time-1)} + r * N_{(time-1)} * \left(1 - \frac{N_{(time-1)}}{K}\right)$$

SSE = (Model predicted elk density – observed elk density)^2



Assignment C

- Assignment C is due on November 7th
- Worth 50 points
- R code - Fitting models
  - Using data
    - plot different curves
    - calculate mse of each
  - Write up in manuscript form for a few of the components. That is, introduce the system (you can self-plagiarize but make it clean), describe how you will sample (or already sampled) components (Methods section), describe your simulation inputs, include output plots. Discuss in brief.

Steps

- Look at the data distributions that you have created for your concept map
- Look over the R Chapter 7 distributions
- Figure out one that looks like it fits
- Adjust the values so that distribution parameters fit your data

Assignment B

- Reintroducing the system
- Describing your actual sampling methodology (in brief)
- Describe with figures what you expect your data distributions to look like using histograms of your data
- Discuss in brief (or not)