#9 Modelling I & II: Regression Analysis

Iris E. Acquarone

POLI 102

Spring 2023, Rice University

Up-to-date

- ▶ WE'R DONE (after this anyways...)! The usual friendly reminder of **all you've accomplished so far** :) :
  - ▶ Fundamentals and R Data ✓
  - ▶ Data wrangling/cleansing ✓
  - ▶ Data visualization I, II, III & IV ✓
  - ▶ Data summarization and MCT&D ✓
  - ▶ Hypothesis testing (Measures of Association I) ✓
  - ▶ Measures of Association II ✓
  - ▶ HW #1, #2, #3, #4, #5 & #6 ✓
- ▶ H #7 ⇝ Run a regression & visualize
- ▶ **Final Project**: Draft 4/14, Final 4/21

Canvas chat for attendance

From correlation to regression

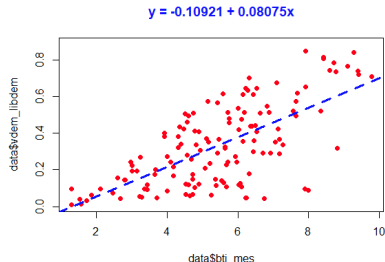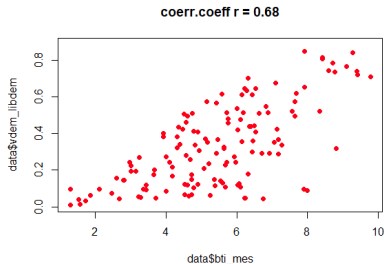coerr.coeff r = 0.68
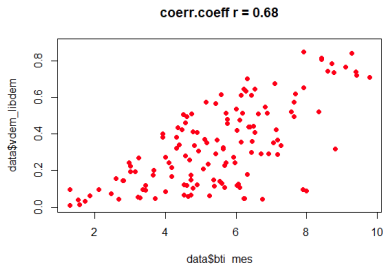
What is it?	Inter-connection/association between two variables
Variable types?	There is no independent/dependent variable, just two variables that we want to see how they 'run' together
What for?	Find out a value representing the strength of the linear association of two variables
Coefficient interpretation?	Whether and how two variables change together

From correlation to regression

coerr.coeff r = 0.68

y = -0.10921 + 0.08075x

What is it?	Inter-connection/association between two variables	How a *said* independent variable is associated with (and explains) a *said* dependent variable
Variable types?	There is no independent/dependent variable, just two variables that we want to see how they 'run' together	There is an independent and known/fixed predictor variable that we assume it helps us estimate a random dependent/outcome variable
What for?	Find out a value representing the strength of the linear association of two variables	Estimate the expected value of an outcome variable based on the values of a predictor variable
Coefficient interpretation?	Whether and how two variables change together	The effect on the outcome variable of a unit change in the independent variable

Regression analysis 101

1. Is your outcome variable continuous or categorical (most. binary?)

▶ Linear regression: predict value of **continuous** outcome

▶ Logistic regression: predict probability of an **event occurring**

2. Will you consider one or multiple independent variables when predicting your outcome?

▶ Bivariate regression: **One** predictor, $Y_i = \beta_0 + \beta_1 X_i + \epsilon$

▶ Multivariate regression: **Multiple** predictors (and/or controls), $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i + \beta_n X_i + \epsilon$

Linear Regression in R

```r
# lm() function for linear regression models
lm(y ~ x, # regression formula: y (DV) ~ x (IV)
   data=data) # dataset

lm(y ~ x + z + w,
   data=data)
```
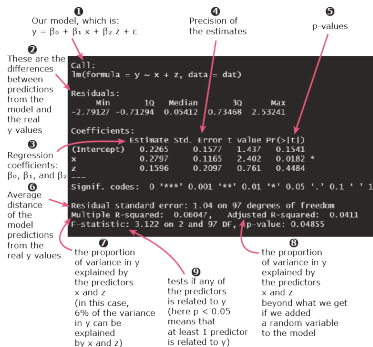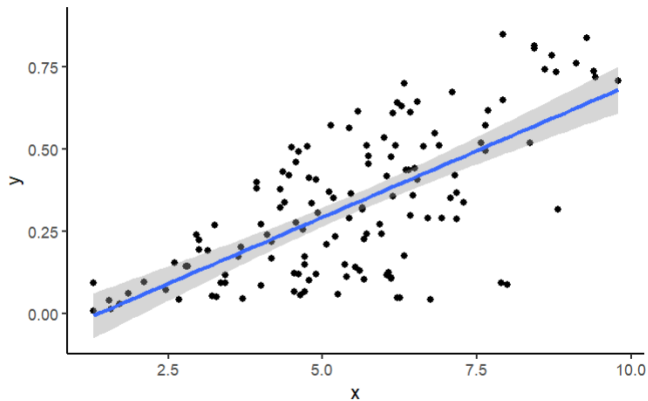

Figure 1: Link

Linear Regression in R

Logistic Regression in R

```r
# glm() function for generalize linear models
glm(y ~ x,# regression formula: y (DV) ~ x (IV)
    family = "binomial", #family = binomial for logistic regression
    data=data) # data set

glm(y ~ x + z + w,
    family = "binomial",
    data=data)
```
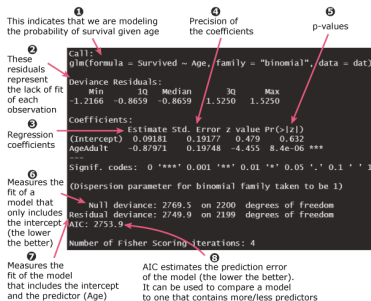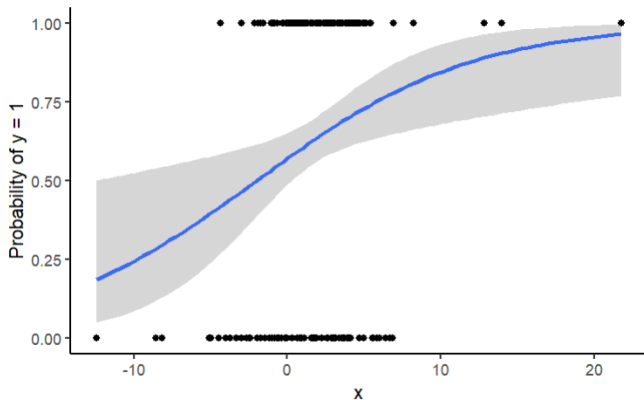

Figure 2: Link

Logistic Regression in R

> Model specification

- ► How many and which predictors?
- ► Transformations?
  - ► Polynomials, logs
  - ► Standardized
- ► Interactions?

HW #7 posted, deadline next week

Next weeks, all about your project.

iea@rice.edu