

Introduction to Multiple Linear Regression

Author: Nicholas G Reich, Jeff Goldsmith

*This material is part of the **statsTeachR** project*

Made available under the Creative Commons Attribution-ShareAlike 3.0 Unported License: http://creativecommons.org/licenses/by-sa/3.0/deed.en_US

Today's lecture

Multiple Linear Regression: basic concepts

- Motivation
- Assumptions
- Interpretation of β s
- More on confounding (omitted variable bias)
- Matrix notation for MLR

Relevant reading: Faraway Chapter 2, *ISL* Chapter 3.2-3.3

Motivation

Most applications involve more than one covariate – if more than one thing can influence an outcome, you need multiple linear regression.

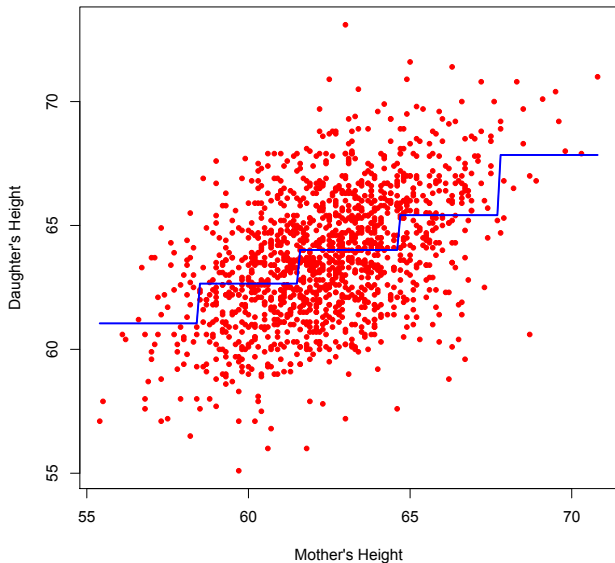
- Improved description of $y|x$
- More accurate estimates and predictions
- Allow testing of multiple effects
- Includes multiple predictor types

bo/d force
vectors

Why not bin all predictors?

- Divide x_i into k_i bins
- Stratify data based on inclusion in bins across x 's
- Find mean of the y_i in each category
- Possibly a reasonable non-parametric model

Why not bin all predictors?



Why not bin all predictors?

- More predictors = more bins
- If each x has 5 bins, you have 5^p overall categories
- May not have enough data to estimate distribution in each category
- Curse of dimensionality is a problem in a lot of non-parametric statistics

For more, see this [interactive Shiny app](#).

Multiple linear regression model

- Observe data $(y_i, x_{i1}, \dots, x_{ip})$ for subjects $1, \dots, n$. Want to estimate $\beta_0, \beta_1, \dots, \beta_p$ in the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i; \epsilon_i \stackrel{iid}{\sim} (0, \sigma^2)$$

- Assumptions (residuals have mean zero, constant variance, are independent) are as in SLR
- Impose linearity which (as in the SLR) is a big assumption
- Our primary interest will be $E(y|\mathbf{x})$
- Eventually estimate model parameters using least squares

Predictor types

- Continuous
- Categorical
- Ordinal

nominal



ordinal
age groups
grades

Interpretation of coefficients

$$\beta_0 = E(y|x_1 = 0, \dots, x_r = 0)$$

- Centering some of the x 's may make this more interpretable

Interpretation of β_1

$$E(y | x_1 = x_1, x_2 = k, x_3 = x_3) = \cancel{\beta_0} + \cancel{\beta_1} x_1 + \cancel{\beta_2} k + \cancel{\beta_3} x_3$$

$$- E(y | x_1 = x_1, x_2 = k-1, x_3 = x_3) = \cancel{\beta_0} + \cancel{\beta_1} x_1 + \cancel{\beta_2} (k-1) + \cancel{\beta_3} x_3$$

$$= \beta_2$$

Interpretation of β_1

β_1 = the expected change in y for a one unit increase in x_1 , holding all other x 's constant.

Example with two predictors

y

Suppose we want to regress weight on age and sex.

- Model is $y_i = \beta_0 + \beta_1 x_{i,age} + \beta_2 x_{i,sex} + \epsilon_i$
- Age is continuous starting with age 0; sex is binary, coded so that $x_{i,sex} = 0$ for men and $x_{i,sex} = 1$ for women

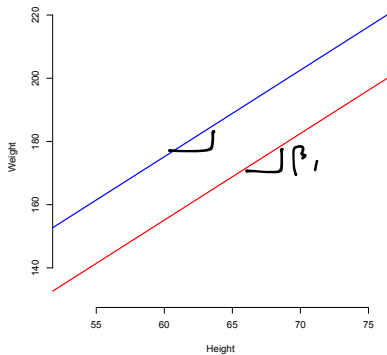
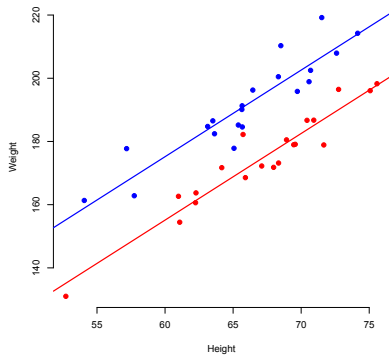
Example with two predictors

$$\text{Model: } y_i = \beta_0 + \beta_1 x_{i,\text{age}} + \beta_2 x_{i,\text{sex}} + \epsilon_i$$

β_1 = Expected change in y for a 1 unit change in age, holding sex constant

β_2 = Exp. diff in weight between a man and woman of the same age.

Example with two predictors



Omitted variable bias

What happens if the true regression model is

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i$$

but we ignore x_2 and fit the simple linear regression

$$\underbrace{y_i = \beta_0^* + \beta_1^* x_{i,1} + \epsilon_i^*}$$

Does $\beta_1^* = \beta_1$?

Omitted variable bias

When should you be concerned?

If both of the following conditions are met, then $\beta_1^* = \beta_1$:

- The omitted variable is unrelated to the outcome
- The omitted variable is uncorrelated with the retained variable

Note: A Simpson's paradox can be explained by omitted variable bias.

Matrix notation

- Observe data $(y_i, x_{i1}, \dots, x_{ip})$ for subjects $1, \dots, n$. Want to estimate $\beta_0, \beta_1, \dots, \beta_p$ in the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i; \quad \epsilon_i \stackrel{iid}{\sim} (0, \sigma^2)$$

- Notation is cumbersome. To fix this, let
 - $\mathbf{x}_i = [1, x_{i1}, \dots, x_{ip}]$
 - $\boldsymbol{\beta}^T = [\beta_0, \beta_1, \dots, \beta_p]$
 - Then $y_i = \mathbf{x}_i \boldsymbol{\beta} + \epsilon_i$

Multiple linear regression

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

■ Let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$y = X\beta + \epsilon$$

■ Then we can write the model in a more compact form:

$$\mathbf{y}_{n \times 1} = \mathbf{X}_{n \times (p+1)} \boldsymbol{\beta}_{(p+1) \times 1} + \boldsymbol{\epsilon}_{n \times 1}$$

■ \mathbf{X} is called the *design matrix*

Matrix notation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

- $\boldsymbol{\epsilon}$ is a random vector rather than a random variable
- $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}$
- Note that Cov means the “variance-covariance matrix”

$$E(\epsilon_i) = 0$$

$$\sigma^2 \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

Mean, variance and covariance of a random vector

- Let $\mathbf{y}^T = [y_1, \dots, y_n]$ be an n -component random vector. Then its mean and variance are defined as

$$E(\mathbf{y})^T = [E(y_1), \dots, E(y_n)]$$

$$\text{Var}(\mathbf{y}) = E[(\mathbf{y} - E\mathbf{y})(\mathbf{y} - E\mathbf{y})^T] = E(\mathbf{y}\mathbf{y}^T) - (E\mathbf{y})(E\mathbf{y})^T$$

- Let \mathbf{y} and \mathbf{z} be an n -component and an m -component random vector respectively. Then their covariance is an $n \times m$ matrix defined by

$$\text{Cov}(\mathbf{y}, \mathbf{z}) = E[(\mathbf{y} - E\mathbf{y})(\mathbf{z} - E\mathbf{z})^T]$$

Coming up next...

Today we covered

- Motivation
- Assumptions
- Interpretation of β s
- More on confounding (omitted variable bias)
- Matrix notation for MLR

Next time...

- ▶ estimation (more least squares)
- ▶ more detailed model diagnostics
- ▶ inference