Linear models

Brian Kissmer

USU Department of Biology

Nov. 5th, 2024

Learning objectives

1. Understand conceptual basis of linear models
2. Know how to implement linear models in R

Today’s outline

1. Linear models
2. Linear models in R

What is a linear model?

What do you associate with the term linear model? Spend a minute thinking about what you personally know, then write a post on the “Linear models” discussion on Canvas

Linear models

Linear models include **linear regression, analysis of variance,** and **analysis of covariance**

Linear models

Linear models have several key components:

1. Response (dependent) variable
2. One or more predictor (independent) variables, sometimes called covariates
3. Model structure with regression coefficients
4. Error terms (related to residuals)

Linear models

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \epsilon_i$$

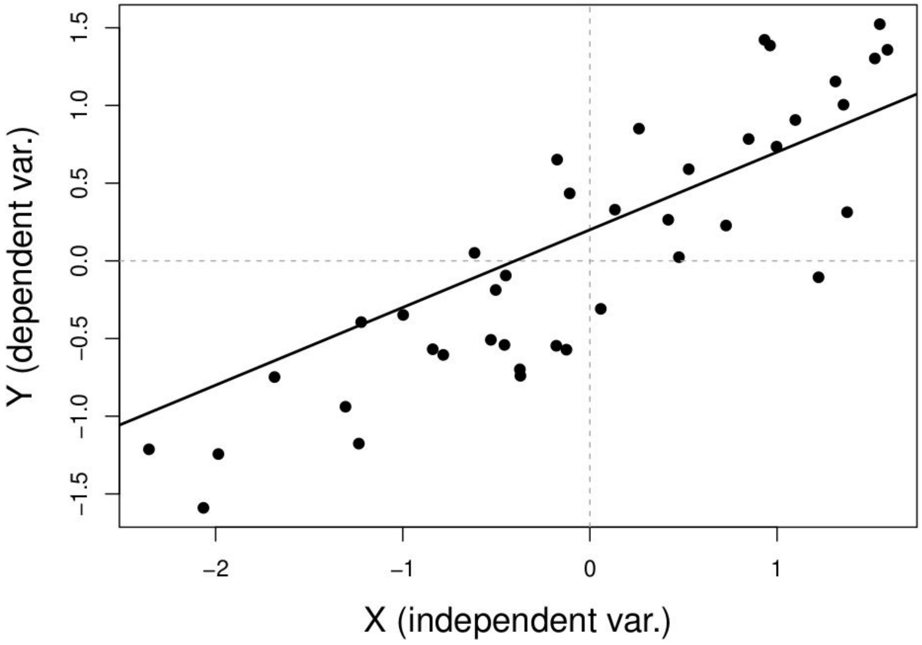
- Response (dependent) variable (Y_i)
- One or more predictor (independent) variables, sometimes called covariates (X_{ki})
- Model structure with regression coefficients (β_0, β_1 , etc.)
- Error term (related to residuals; ϵ)
- Observation = i, covariate number = k

What do we do with linear models?

GOAL: determine the relationship between independent and dependent variables

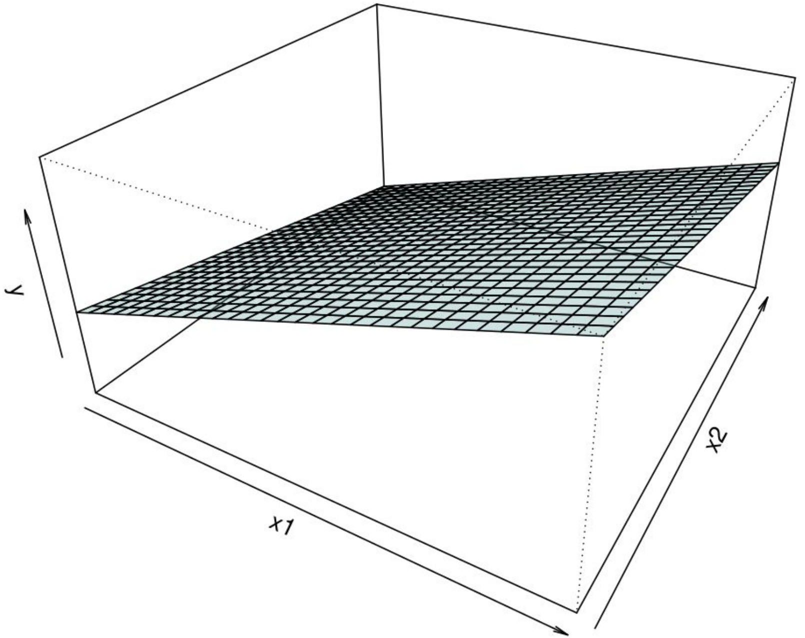
- ➢ Relationship is captured by regression coefficients (βs)
- ➢ Association does not always indicate causality

Linear model with a single, continuous covariate (linear regression)



$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i = 0.2 + 0.5X_{1i} + \epsilon_i$$

Linear model with two covariates



$$Y_i = 0.1 + 0.5X_{1i} - 0.3X_{2i} + \epsilon_i$$

How do you estimate the regression coefficients?

You need data (X and Y), and some means of estimation.

Several exist:

1. Minimize sum-of-squares function

$$S = \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \epsilon_i))^2$$

2.
3.

How do you estimate the regression coefficients?

You need data (X and Y), and some means of estimation.

Several exist:

1. Minimize sum-of-squares function

$$S = \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \epsilon_i))^2$$

2. Maximize a likelihood (or minimize a negative log-likelihood)
3. Bayesian methods derive the posterior probability of the parameters given the data and model

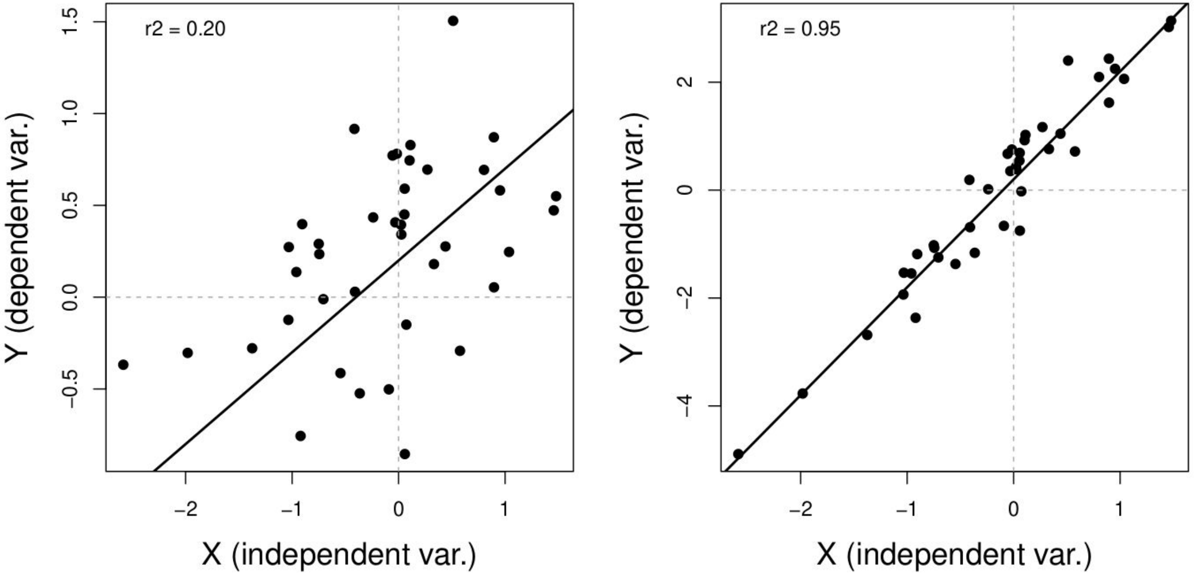
Hypothesis testing with linear regression models

You have an estimate of β_1 describing the effect of some variable on your response variable of interest.

1. How do you know whether the model is any good?
2. How you know if you should be confident in your estimate?
3. How do you know if you can rule out the point value of 0 for β_1 (i.e. no effect)?
4. Discuss in groups and provide a brief summary of your groups answer on the Linear model 2 discussion on Canvas

Hypothesis testing with linear regression models

Coefficient of determination (r^2) measures the proportion of variation in the dependent variable explained by the independent variables.



Extensions of simple linear regression

Several extensions of simple linear regression exist:

- Multiple linear regression = multiple independent variables with possible interactions
- Multivariate linear regression = multiple correlated dependent variables
- Generalized linear models = likelihood-based, allow for flexibility in terms of data distribution, error distribution, and link function
- Mixed or random effect models

Extensions of simple linear regression

Several extensions of simple linear regression exist (cont.):

- Analysis of variance (ANOVA) = categorical linear regression
- Analysis of covariance (ANCOVA) = mixture of continuous and categorical values

Encoding ANOVA as a linear model

Design matrix (dummy variable) encoding can be used to fit an ANOVA as a linear model (done for you in R):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

For three treatments, A, B, and C denoted by *X:*

	Intercept	X_{1i} = Trt. B	X_{2i} = Trt. C
A	1	0	0
B	1	1	0
C	1	0	1

R code for linear models

See the linear models R handout