

INTRO to DATA SCIENCE

LINEAR REGRESSION

I. LINEAR REGRESSION

EXERCISES:

II. LINEAR REGRESSION IN R

I. LINEAR REGRESSION

	<i>continuous</i>	<i>categorical</i>
<i>supervised</i>	???	???
<i>unsupervised</i>	???	???

	<i>continuous</i>	<i>categorical</i>
<i>supervised</i>	regression	classification
<i>unsupervised</i>	dimension reduction	clustering

Q: What is a **regression** model?

Q: What is a **regression** model?

A: A functional relationship between input & response variables.

Q: What is a **regression** model?

A: A functional relationship between input & response variables

The **simple linear regression** model captures a linear relationship between a single input variable x and a response variable y :

Q: What is a **regression** model?

A: A functional relationship between input & response variables

The **simple linear regression** model captures a linear relationship between a single input variable x and a response variable y :

$$y = \alpha + \beta x + \varepsilon$$

Q: What do the terms in this model mean?

$$y = \alpha + \beta x + \varepsilon$$

Q: What do the terms in this model mean?

$$y = \alpha + \beta x + \varepsilon$$

A: y = **response variable** (the one we want to predict)

Q: What do the terms in this model mean?

$$y = \alpha + \beta x + \varepsilon$$

A: y = **response variable** (the one we want to predict)

x = **input variable** (the one we use to train the model)

Q: What do the terms in this model mean?

$$y = \alpha + \beta x + \varepsilon$$

A: y = **response variable** (the one we want to predict)

x = **input variable** (the one we use to train the model)

α = **intercept** (where the line crosses the y-axis)

Q: What do the terms in this model mean?

$$y = \alpha + \beta x + \varepsilon$$

A: y = **response variable** (the one we want to predict)

x = **input variable** (the one we use to train the model)

α = **intercept** (where the line crosses the y-axis)

β = **regression coefficient** (the model “parameter”)

Q: What do the terms in this model mean?

$$y = \alpha + \beta x + \varepsilon$$

A: y = **response variable** (the one we want to predict)

x = **input variable** (the one we use to train the model)

α = **intercept** (where the line crosses the y-axis)

β = **regression coefficient** (the model “parameter”)

ε = **residual** (the prediction error)

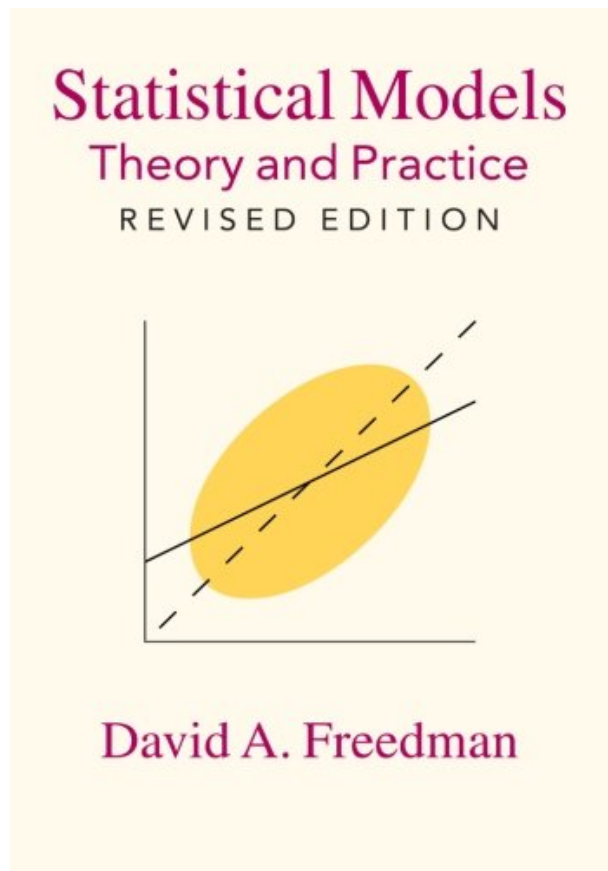
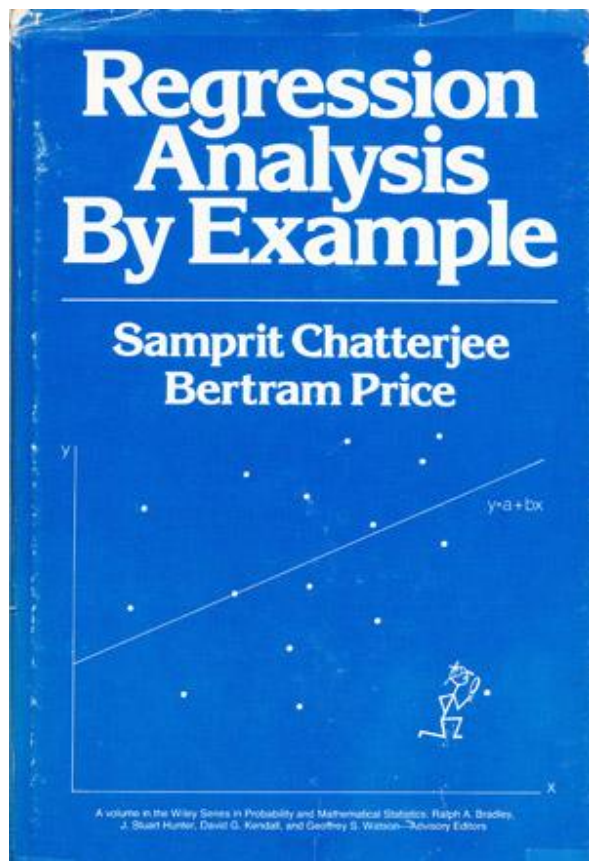
We can extend this model to several input variables, giving us the **multiple linear regression** model:

We can extend this model to several input variables, giving us the **multiple linear regression** model:

$$y = \alpha + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon$$

Linear regression involves several technical assumptions and is often presented with lots of mathematical formality.

The details are not very important for our purposes, but you can check them out if you're interested.



Q: How do we fit a regression model to a dataset?

Q: How do we fit a regression model to a dataset?

A: Minimize the sum of the squared residuals (Ordinary Least Squares).

Q: How do we fit a regression model to a dataset?

A: Minimize the sum of the squared residuals (Ordinary Least Squares).

In practice, any respectable piece of software will do this for you.

Q: How do we fit a regression model to a dataset?

A: Minimize the sum of the squared residuals (Ordinary Least Squares).

In practice, any respectable piece of software will do this for you.

And there are other ways.

Q: How do we fit a regression model to a dataset?

A: Minimize the sum of the squared residuals (Ordinary Least Squares).

In practice, any respectable piece of software will do this for you.

And there are other ways.

And software implements them as well.

Q: So, how do we select a linear regression model?

Q: So, how do we select a linear regression model?

- Techniques for producing designs

Q: So, how do we select a linear regression model?

- Techniques for producing designs
 - Theory

Q: So, how do we select a linear regression model?

- Techniques for producing designs
 - Theory
 - Forward selection

Q: So, how do we select a linear regression model?

- Techniques for producing designs
 - Theory
 - Forward selection
 - Backward elimination

Q: So, how do we select a linear regression model?

- Techniques for producing designs
 - Theory
 - Forward selection
 - Backward elimination
 - More

Q: So, how do we select a linear regression model?

- Techniques for producing designs
 - Theory
 - Forward selection
 - Backward elimination
 - More
- Techniques for evaluating designs

Q: So, how do we select a linear regression model?

- Techniques for producing designs
 - Theory
 - Forward selection
 - Backward elimination
 - More
- Techniques for evaluating designs
 - Training metrics

Q: So, how do we select a linear regression model?

- Techniques for producing designs
 - Theory
 - Forward selection
 - Backward elimination
 - More
- Techniques for evaluating designs
 - Training metrics
 - Test set performance / Cross-validation

II. EXERCISES