MSDS 596 Regression & Time Series

Lecture 01 Introduction to linear regression

Department of Statistics
Rutgers University

Sept 12, 2022

Welcome to MSDS 596: Regression & Time Series

- Meeting time: **Mondays 6-9pm**
- Instructor: **Koulik Khamaru** (kk1241@rutgers.edu)
  Office hour: Monday 4-5pm, 479 Hill center
- Teaching Assistant: **Suvrojit Ghosh** (sg1565@scarletmail.rutgers.edu)
  Office hour: Wednesdays 3:30-4:30pm
- Course website (piazza):
  https://piazza.com/rutgers/fall2022/1695459602
  The course website is our main means of communication for all things class-related.

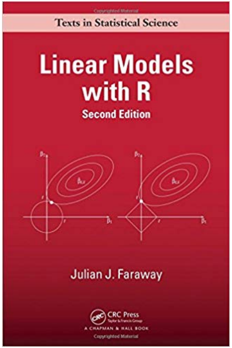
About the course

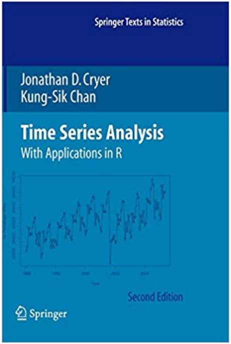
Course description: An introduction to statistical methods with linear models, time series models and forecasting. Inference and prediction of linear models and their extensions. Model checking, refinement and selection. Classical time series including autoregressive, moving average, ARIMA models. Select advanced time series topics (state space models and Kalman filter, seasonal models). Data analysis with computational software is emphasized.

Your preparations

Prerequisites: Undergraduate level calculus, linear algebra, probability and mathematical statistics. Prior experience with statistical programming software, preferably R, is highly desirable.

Textbooks





- Faraway, Julian J. *Linear models with R*. 2nd ed., CRC press, 2014. Website: `http://people.bath.ac.uk/jjf23/LMR/`
- Cryer, J. D. and Chan, K.-S. *Time Series Analysis with Applications in R* (2nd Ed). Springer, 2008.

Schedule (subject to change)

Week	Date	Topic
1	9/12	Intro to linear regression
2	9/19	Estimation
3	9/26	Inference I
4	10/3	Inference II
5	10/10	Inference and prediction
6	10/17	Explanation; model diagnostics
7	10/24	Transformation and model selection
8	10/31	Shrinkage methods
9	11/7	Time series exploratory analysis
10	11/14	Linear time series: ARIMA models
11	11/21	Model specification and estimation
12	11/28	Diagnostics and forecasting
13	12/05	Seasonal models
14	12/12	Project & final evaluation
15	12/17	Project & final evaluation

Evaluation

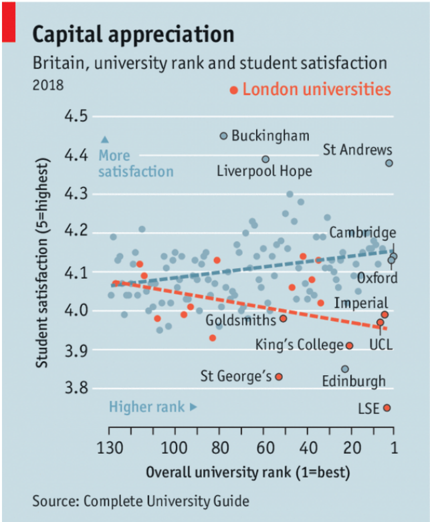
- Homework assignments: 30%;
  - About eight homeworks, weekly or bi-weekly.
  - no late submissions accepted; least homework grade will be dropped
- Midterm project: 30%;
- Final evaluation: 40%.

Questions?

Why do **you** want to study regression?

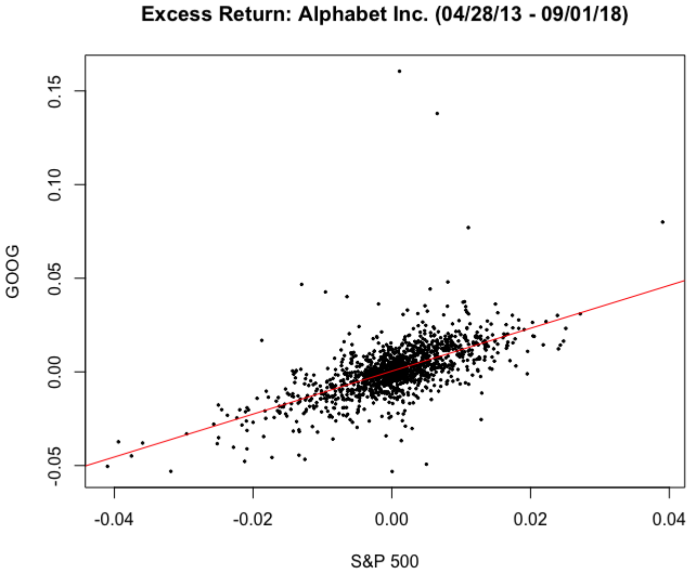
British university ranking vs. student satisfaction

Read the full story [here] *(The Economist, 07/10/18)*



Economist.com

Stock market



Child growth assessment study

Archives of Disease in Childhood, 1970, **45**, 755.

Standards for Children's Height at Ages 2-9 Years Allowing for Height of Parents

J. M. TANNER, H. GOLDSTEIN, and R. H. WHITEHOUSE

From the Department of Growth and Development, Institute of Child Health, University of London

Tanner, J. M., Goldstein, H., and Whitehouse, R. H. (1970). *Archives of Disease in Childhood*, **45**, 755. **Standards for children's height at ages 2–9 years allowing for height of parents.** Charts* are presented which give centile standards for boys' and girls' heights at ages 2 to 9 when parents' height is allowed for. Mid-parent height is used (i.e. the average of father's and mother's height).

A comparison is made with results from the existing 'parent-unknown' British standard charts. A child at the 3rd centile on the parent-unknown charts is (i) at the 20th centile on the new charts if his parents are small enough to average 3rd centile for adults, (ii) at about the 1st centile if his parents average the 97th centile. Conversely a

Linear model

- We want to model Y (e.g. child's height) as a function of X (e.g. mid-parent height).

- A model is an expression for $E[Y \mid X = x]$. Let us call it $\mu(x)$.

- What is a linear model?
    - Linear model means $\mu(x)$ is linear in some parameters.

$$\mu(x) = \beta_0^* + \beta_1^* x.$$

- Defining $\epsilon(x) := Y(x) - \mu(x)$

$$\begin{aligned} Y(x) &= \mu(x) + \epsilon(x) \\ &= \beta_0^* + \beta_1^* x + \epsilon(x) \quad \text{(Linear model)} \end{aligned}$$

Linear model

- We want to model Y (e.g. child's height) as a function of X (e.g. mid-parent height).
- A model is an expression for $E[Y \mid X = x]$. Let us call it $\mu(x)$.
- What is a linear model?
    - Linear model means $\mu(x)$ is linear in some parameters.

$$\mu(x) = \beta_0^* + \beta_1^* x.$$

- Defining $\epsilon(x) := Y(x) - \mu(x)$

$$\begin{aligned} Y(x) &= \mu(x) + \epsilon(x) \\ &= \beta_0^* + \beta_1^* x + \epsilon(x) \quad \text{(Linear model)} \end{aligned}$$

Linear model

- We want to model Y (e.g. child's height) as a function of X (e.g. mid-parent height).
- A model is an expression for $E[Y \mid X = x]$. Let us call it $\mu(x)$.
- What is a linear model?
  - Linear model means $\mu(x)$ is linear in some parameters.

$$\mu(x) = \beta_0^* + \beta_1^* x.$$

- Defining $\epsilon(x) := Y(x) - \mu(x)$

$$\begin{aligned} Y(x) &= \mu(x) + \epsilon(x) \\ &= \beta_0^* + \beta_1^* x + \epsilon(x) \quad \text{(Linear model)} \end{aligned}$$

Linear model

- We want to model Y (e.g. child's height) as a function of X (e.g. mid-parent height).
- A model is an expression for $E[Y \mid X = x]$. Let us call it $\mu(x)$.
- What is a linear model?
  - Linear model means $\mu(x)$ is linear in some parameters.

$$\mu(x) = \beta_0^* + \beta_1^* x.$$

- Defining $\epsilon(x) := Y(x) - \mu(x)$

$$\begin{aligned} Y(x) &= \mu(x) + \epsilon(x) \\ &= \beta_0^* + \beta_1^* x + \epsilon(x) \quad \text{(Linear model)} \end{aligned}$$

Linear model

- We want to model Y (e.g. child's height) as a function of X (e.g. mid-parent height).

- A model is an expression for $E[Y \mid X = x]$. Let us call it $\mu(x)$.

- What is a linear model?
    - Linear model means $\mu(x)$ is linear in some parameters.

$$\mu(x) = \beta_0^* + \beta_1^* x.$$

- Defining $\epsilon(x) := Y(x) - \mu(x)$

$$\begin{aligned} Y(x) &= \mu(x) + \epsilon(x) \\ &= \beta_0^* + \beta_1^* x + \epsilon(x) \quad \text{(Linear model)} \end{aligned}$$

Linear model

$$Y(x) = \beta_0^* + \beta_1^* x + \epsilon(x)$$

- Note that $\mathbb{E}[\epsilon(x)] = 0$. Meaning noise is zero mean.
- We will use Y and ϵ instead of $Y(x)$ and $\epsilon(x)$ for simplicity.
- We will also assume that $\text{var}(\epsilon(x)) = \sigma^2$ (independent of x)
- Shorthand:

$$Y = \beta_0^* + \beta_1^* x + \epsilon$$

Are these linear models?

$$Y = \beta_0^* + \beta_1^* \log(x) + \epsilon$$

$$Y = \beta_0^* + \beta_1^* \sqrt{x} + \epsilon$$

$$Y = \beta_0^* + \beta_1^* x^2 + \epsilon$$

But we do not know β_0^* and β_1^*?

$$y = \beta_0^* + \beta_1^* x + \epsilon$$

- We have data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ from the linear model. They satisfy

$$y_i = \beta_0^* + \beta_1^* x_i + \epsilon_i$$

- Problem: How do we find β_0^* and β_1^*?
- Idea: Good values of more accurate prediction or equivalently less noise, i.e. small ϵ_i.

But we do not know β_0^* and β_1^*?

$$y_i = \beta_0^* + \beta_1^* x_i + \epsilon_i \quad \text{i = 1, \ldots n}$$

- Linear regression finds $(\widehat{\beta}_0, \widehat{\beta}_1)$ via

$$(\widehat{\beta}_0, \widehat{\beta}_1) = \arg\min_{\beta_0, \beta_1} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \right\}$$

- Least squares method finds the minimizers to the sum of squares error.

Solving for a simple linear regression: least squares

- $$(\widehat{\beta}_0, \widehat{\beta}_1) = \arg\min_{\beta_0,\beta_1} Q(\beta_0, \beta_1) = \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \right\}$$

- Q is a continuous function, thus $\widehat{\beta}_0, \widehat{\beta}_1$ is solution to

$$\frac{\partial Q(\beta_0, \beta_1)}{\partial \beta_0} = 0, \qquad \frac{\partial Q(\beta_0, \beta_1)}{\partial \beta_1} = 0,$$

- Normal equations:

$$\bar{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \bar{x}$$
$$\overline{yx} = \widehat{\beta}_0 \bar{x} + \widehat{\beta}_1 \overline{x^2}$$

Solving for a simple linear regression: least squares

- $$(\widehat{\beta}_0, \widehat{\beta}_1) = \arg\min_{\beta_0, \beta_1} Q(\beta_0, \beta_1) = \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \right\}$$

- Q is a continuous function, thus $\widehat{\beta}_0, \widehat{\beta}_1$ is solution to

$$\frac{\partial Q(\beta_0, \beta_1)}{\partial \beta_0} = 0, \qquad \frac{\partial Q(\beta_0, \beta_1)}{\partial \beta_1} = 0,$$

- Normal equations:

$$\bar{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \bar{x}$$
$$\overline{yx} = \widehat{\beta}_0 \bar{x} + \widehat{\beta}_1 \overline{x^2}$$

Solving for a simple linear regression: least squares

- $$(\widehat{\beta}_0, \widehat{\beta}_1) = \arg\min_{\beta_0, \beta_1} Q(\beta_0, \beta_1) = \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \right\}$$

- Q is a continuous function, thus $\widehat{\beta}_0, \widehat{\beta}_1$ is solution to

$$\frac{\partial Q(\beta_0, \beta_1)}{\partial \beta_0} = 0, \qquad \frac{\partial Q(\beta_0, \beta_1)}{\partial \beta_1} = 0,$$

- Normal equations:

$$\bar{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \bar{x}$$
$$\overline{yx} = \widehat{\beta}_0 \bar{x} + \widehat{\beta}_1 \overline{x^2}$$

Solving for a simple linear regression: least squares

- $$(\widehat{\beta}_0, \widehat{\beta}_1) = \arg\min_{\beta_0,\beta_1} Q(\beta_0, \beta_1) = \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \right\}$$
- Q is a continuous function, thus $\widehat{\beta}_0, \widehat{\beta}_1$ is solution to
$$\frac{\partial Q(\beta_0, \beta_1)}{\partial \beta_0} = 0, \qquad \frac{\partial Q(\beta_0, \beta_1)}{\partial \beta_1} = 0,$$
- Solution:
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

Residual and fitted value

- Model

$$y_i = \beta_0^* + \beta_1^* x_i + \epsilon_i$$

- $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$: fitted values.
- $\hat{\epsilon}_i = y_i - \hat{y}_i$: residuals.
- $(\widehat{\beta}_0, \widehat{\beta}_1) \approx (\beta_0^*, \beta_1^*)$ when we have lots of data (we will show this mathematically at some point of time), but we can not say $(\widehat{\beta}_0, \widehat{\beta}_1) = (\beta_0^*, \beta_1^*)$.
- Similarly, $\hat{y}_i \neq y_i$ and $\epsilon_i \neq \hat{\epsilon}_i$.

More examples

- Suppose I know $\beta_1^* = \beta_0^*$.
- How do I fit the model

$$Y = \beta_0^* + \beta_0^* x + \epsilon?$$

Multiple linear model

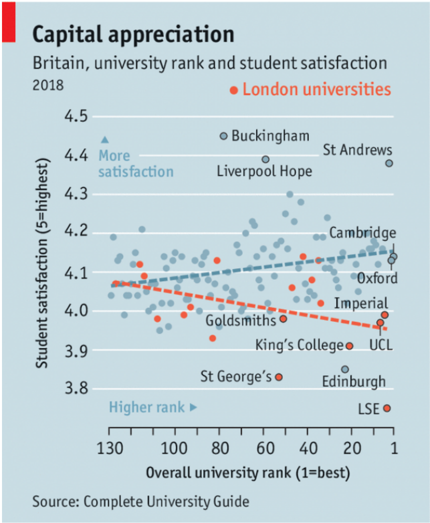
- Simple linear regression: $\mu(X) = \beta_0^* + \beta_1^* X$.
- Generally, model Y using p predictors

$$\begin{aligned} \mu(X_1, \ldots, X_p) &= \beta_0^* + \beta_1^* X_1 + \cdots + \beta_p^* X_p \\ Y &= \mu(X_1, \ldots, X_p) + \epsilon \end{aligned}$$

  - Y: response, output, outcome, dependent variable
  - X_j: predictor, input, covariate, independent variable
  - β_j: parameter, regression coefficient
- Linear means μ is a linear function of β_j^*, $0 \le j \le p$.
  - $Y = \beta_0^* + \beta_1^* \cdot \log X_1 + \beta_2^* \cdot X_2^{X_3} + \epsilon$ is a linear model;
  - $Y = \beta_0^* + \beta_1^{*2} X_1 + X_2^{\beta_2^*} + \epsilon$ is NOT a linear model!

University ranking vs. student satisfaction *(The Economist, 07/10/18)*

What is the response variable? What are the predictor variables?



Economist.com

Matrix representation

- Multiple linear model:

$$Y_i = \beta_0^* + \beta_1^* X_{i1} + \beta_2^* X_{i2} + \cdots + \beta_p^* X_{ip} + \epsilon_i, \quad 1 \le i \le n.$$

- Define

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}, \quad \boldsymbol{\beta}^* = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}.$$

- Let $\boldsymbol{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip})'$, then for each individual observation (unit)

$$y_i = \boldsymbol{x}_i' \beta^* + \epsilon_i, \quad 1 \le i \le n.$$

- The overall linear model can be represented as

$$\boldsymbol{y} = \boldsymbol{X}\beta^* + \epsilon.$$

  - $\boldsymbol{X}$ is called design matrix or model matrix.
  - Assume $n > p + 1$, and $\boldsymbol{X}$ has full rank $p + 1$.

- Next week: review of matrix algebra

Matrix representation

- Multiple linear model:

$$Y_i = \beta_0^* + \beta_1^* X_{i1} + \beta_2^* X_{i2} + \cdots + \beta_p^* X_{ip} + \epsilon_i, \quad 1 \le i \le n.$$

- Define

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}, \quad \boldsymbol{\beta}^* = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}.$$

- Let $\boldsymbol{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip})'$, then for each individual observation (unit)

$$y_i = \mathbf{x}_i' \boldsymbol{\beta}^* + \epsilon_i, \quad 1 \le i \le n.$$

- The overall linear model can be represented as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\epsilon}.$$

  - $\boldsymbol{X}$ is called design matrix or model matrix.
  - Assume $n > p + 1$, and $\boldsymbol{X}$ has full rank $p + 1$.

- Next week: review of matrix algebra

Least squares for multiple regression

- Minimize the sum of squared errors

$$\min_{\beta_0,\beta_1,\ldots,\beta_p} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} + \cdots + \beta_p x_{ip})]^2.$$

- In matrix notation

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2.$$

- Normal equation:

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}.$$

- Least squares estimate:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$

- $\hat{y}_i = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}$: fitted values.
- $\hat{\epsilon}_i = y_i - \hat{y}_i$: residuals.

Exploratory data analysis and statistical graphics

An introduction to R/Handout on Piazza