Methods of Applied Statistics I

STA2101H F LEC9101

Week 1

September 14 2022

← **Tweet**

The Guy Medal in Gold is awarded to Nancy Reid @reid_nancy for her pioneering work on higher-order approximate inference

5:54 PM · Sep 12, 2022 · Twitter Web App

Royal Statistical Society 2022 International Conference

SEPTEMBER
BST 13

08:30

09:00

09:00-10:00
Contributed: Infectious disease
Topic streams Applications of Statistics

09:00-10:00
Contributed: Novel applications and data sets
Topic streams Data Science

09:00-10:00
Contributed: Spatial point process models
Topic streams Environmental & Spatial Statistics

09:00-10:00
Contributed: Causal inference
Topic streams Medical Statistics

09:00-10:00
Contributed: Clinical Trials
Topic streams Medical Statistics

09:00-10:00
Contributed: Design of experiments and studies
Topic streams Methods & Theory

09:00-10:00
Contributed: Reporting uncertainty
Topic streams Methods & Theory

10:00

10:10

10:10-11:10
Keynote 2: Campion (President's Invited) Lecture - Adrian Raftery
Topic streams Plenary

11:10

11:10-11:40
Refreshments
Topic streams Break

11:40

11:40-13:00
The role of statistics in supporting professional ballet dancers
Topic streams Applications of Statistics

11:40-13:00
Democratisation of statistics in GSK
Topic streams Business, Industry & Finance

11:40-13:00
RSS Statistical Ambassadors' Showcase
Topic streams Communicating & Teaching Statistics

11:40-13:00
Data Science in Industry - an introduction to MLOps
Topic streams Data Science

11:40-13:00
Model selection and discrimination for environmental and spatial applications
Topic streams Environmental & Spatial Statistics

11:40-13:00
What is your estimand?
Topic streams Medical Statistics

11:40-13:00
Papers from the RSS Journals: 50 years of Cox regression: developments and perspectives
Topic streams Methods & Theory

Today

1. Course introduction: course details, evaluation, syllabus, people
2. Upcoming events of interest
3. Review of linear regression
4. In the news: $\longrightarrow$ at the conference

academic integrity

STA 2101F: Methods of Applied Statistics I
Wednesday, 10am – 1 pm Eastern September 14 – December 7 2022
SF 3201

From the calendar:

This course will focus on principles and methods of applied statistical science. It is designed for MSc and PhD students in Statistics, and is required for the Applied Paper of the PhD comprehensive exams. The topics covered include: planning of studies, review of linear models, analysis of random and mixed effects models, model building and model selection, theory and methods for generalized linear models, and an introduction to nonparametric regression. Additional topics will be introduced as needed in the context of case studies in data analysis.

Prerequisites: ECO374H1/ECO375H1/STA302H1 (regression); STA305H1 (design of studies)

Course Delivery:

STA 2101F: Methods of Applied Statistics I
Wednesday, 10am – 1 pm Eastern September 14 – December 7 2022
SF 3201

From the calendar:

This course will focus on principles and methods of applied statistical science. It is designed for MSc and PhD students in Statistics, and is required for the Applied Paper of the PhD comprehensive exams. The topics covered include: planning of studies, review of linear models, analysis of random and mixed effects models, model building and model selection, theory and methods for generalized linear models, and an introduction to nonparametric regression. Additional topics will be introduced as needed in the context of case studies in data analysis.
Prerequisites: ECO374H1/ECO375H1/STA302H1 (regression); STA305H1 (design of studies)

Course Delivery:

On September 14, the class will be delivered online at the scheduled

Course Description

- Grading 60 / 40
- Academic Integrity check course info
- Computing R
- References Modules
- Contact

Use **Piazza** for course questions; email for personal questions

Linear Models = R (Faraway)
1st 2nd

Extended LM = R (F)
1st 2nd

Today

1. Course introduction: technical issues, course details, evaluation, syllabus
2. Upcoming events of interest
3. Review of linear regression
4. In the news: $\longrightarrow$ at the conference

Upcoming events

- Thursdays 3.30 — Departmental Seminar link
- Mondays 3.30 — Data Science and Applied Research Seminar link
- Fridays 12.00 — Toronto Data Workshop
- Special:
  - September 29: CANSSI Ontario Research Day link
  - September 29, 30 3.30: Distinguished Lecture Series in Statistical Sciences link

2022 DLSS: Xihong Lin

Professor, Department of BiostatisticsCoordinating Director, Program in Quantitative Genomics; Harvard T.H. Chan School of Public Health; Professor of Statistics, Department of Statistics, Harvard University

Sep 29 (3:30-4:30 pm): Lessons Learned from the COVID-19 Pandemic: A Statistician’s Reflection
Sep 30 (3:30-4:30 pm): Ensemble Methods for Testing a Global Null Hypothesis

Today

Start Recording

1. Course introduction: course details, evaluation, syllabus
2. Upcoming events of interest
3. Review of linear regression
4. Steps in analysis
5. In the news: $\longrightarrow$ at the conference

- Model:

$$\underset{n\times 1}{Y} = \underset{n\times p}{X}\ \underset{p\times 1}{\beta} + \underset{n\times 1}{\epsilon}$$

- Model:

$$Y_{n\times 1} = \quad X_{n\times p}\beta_{p\times 1} + \quad \epsilon_{n\times 1}$$

- Model:

noise

vectors = col = v.

$$Y_{n\times 1} = \boxed{X_{n\times p}\beta_{p\times 1}} + \epsilon_{n\times 1}$$

$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

- Equivalently:

$$y_i = \underset{1\times p}{\underline{x}_i^T} \underset{p\times 1}{\underline{\beta}} + \varepsilon_i$$

$$X = \begin{bmatrix} \underline{x}_1^T \\ \vdots \\ \underline{x}_n^T \end{bmatrix}$$

$$= \begin{matrix} \boxed{x_{11} \; \cdots \; x_{1p}} \\ \vdots \\ x_{n1} \; \cdots \; x_{np} \end{matrix}$$

$\underline{x}_1^T$

- Model:

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \epsilon_{n\times 1}$$

- Equivalently:

$$y_i = x_i^T\beta + \varepsilon_i$$

random

- Standard Assumptions
  - y_i independent   equivalently ϵ_i independent
  - $\mathbb{E}(\epsilon_i) = 0$
  - $\text{var}(\epsilon_i) = \sigma^2$
  - x_i known, β to be estimated

$E(y_i) = x_i^T\beta$

y is often called response

why?

constant

x_i often called explanatory variables

($y_i | x_i$ is ind't of $y_j | x_j$)

doesn't change in i

features

covariates

(independent)

- Model:

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \epsilon_{n\times 1}$$

- Equivalently:

$$y_i = x_i^T\beta + \varepsilon_i$$

- Standard Assumptions
  - y_i independent   equivalently ϵ_i independent y is often called response
  - $\mathbb{E}(\epsilon_i) = 0$ why?
  - $\text{var}(\epsilon_i) = \sigma^2$ constant
  - x_i known, β to be estimated x_i often called explanatory variables
- More concisely:

$$\mathbb{E}(Y \mid X) = X\beta \quad , \quad \text{var}(Y \mid X) = \sigma^2 I_{n\times n}$$

I ??

dist of Y, given X

$$\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

Nice big equation:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$Y = \underbrace{\qquad}_{X} \quad \beta \quad + \varepsilon$$

Nice big equation:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \\ \vdots \\ \\ \end{pmatrix} + \begin{pmatrix} \\ \vdots \\ \\ \end{pmatrix}$$

Or, if you prefer:

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ip}\beta_p + \epsilon_i, \quad \epsilon_i \sim (0, \sigma^2) \qquad i = 1, \dots, n$$

$\underline{x}_i^T \underline{\beta}$ ind't

Nice big equation:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \quad \\ \vdots \\ \quad \end{pmatrix} + \begin{pmatrix} \quad \\ \vdots \\ \quad \end{pmatrix}$$

Or, if you prefer:

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + \epsilon_i, \quad \epsilon_i \sim \qquad i = 1, \ldots, n$$

Or, if you prefer:

$$\mathbb{E}(y_i \mid x_i) = x_i^{\mathrm{T}}\beta, \qquad \mathrm{var}(y_i \mid x_i) = \sigma^2, \qquad i = 1, \ldots, n$$

$(*)$ $E(y \mid X) = X\beta$ $\quad$ var $(y \mid x) = \sigma^2 I$ $\quad$ y_i independent

- often not completely clear: X might be fixed by design, or measured on each individual

e.g.?

- often not completely clear: X might be fixed by design, or measured on each individual

  e.g.?

- If measured, then should we consider its distribution? E.g. should our model be $(y_i, x_i^{\mathrm{T}}) \sim$?? multivariate analysis

  some $(p+1)$-dimensional distribution

- often not completely clear: X might be fixed by design, or measured on each individual

  e.g.?

- If measured, then should we consider its distribution? E.g. should our model be $(y_i, x_i^{\mathrm{T}}) \sim$??

  fixed-X // model-X

  some $(p+1)$-dimensional distribution

- Almost always in regression settings we condition on X, as on previous slide

  supervised unsupervised

  ancillary statistic

- often not completely clear: X might be fixed by design, or measured on each individual

  e.g.?

- If measured, then should we consider its distribution? E.g. should our model be $(y_i, x_i^{\mathrm{T}}) \sim$??

  some $(p+1)$-dimensional distribution

- Almost always in regression settings we condition on X, as on previous slide

  ancillary statistic

- often not emphasized: interpretation of β_j

- often not completely clear: X might be fixed by design, or measured on each individual

  e.g.?

- If measured, then should we consider its distribution? E.g. should our model be $(y_i, x_i^{\mathrm{T}}) \sim$??

  some $(p+1)$-dimensional distribution

- Almost always in regression settings we condition on X, as on previous slide

  ancillary statistic

- often not emphasized: interpretation of β_j $j = 1, \dots, p$
  - version 1: effect on the expected response of a unit change in jth explanatory variable, all other variables held fixed

- often not completely clear: X might be fixed by design, or measured on each individual

  e.g.?

- If measured, then should we consider its distribution? E.g. should our model be $(y_i, x_i^{\mathrm{T}}) \sim$??

  some $(p+1)$-dimensional distribution

- Almost always in regression settings we condition on X, as on previous slide

  ancillary statistic

- often not emphasized: interpretation of β_j
  - version 1: effect on the expected response of a unit change in jth explanatory variable, all other variables held fixed
  - version 2:

$$\beta_j = \frac{\partial \mathbb{E}(y_i \mid x_{ij})}{\partial x_{ij}} \qquad \frac{\partial \mathbb{E}(y \mid x_j)}{\partial x_j}$$

notation ambiguous, see CD §6.5.2

$y_i = x_i^T \beta + \varepsilon_i$

$E(y_i) = x_i^T \beta$

Least squares estimation

- Definition

$$\hat{\beta}_{LS} := \min_{\beta} \sum_{i=1}^{n} (y_i - x_i^{\mathrm{T}} \beta)^2$$

Least squares estimation

- Definition

$$\hat{\beta}_{LS} := \min_{\beta} \sum_{i=1}^{n} (y_i - x_i^{\mathrm{T}} \beta)^2$$

- Equivalently,

$$= \min_{\beta} \|Y - X\beta\|_2^2$$

sq'd Euclidean norm

$$SS(\beta) = \min_{\beta} (\underset{n\times 1}{Y} - X\beta)^{T}(Y - X\beta)$$

$$\frac{\partial}{\partial \beta}\left\{(Y - X\beta)^{T}(Y - X\beta)\right\}\Big|_{\hat{\beta}_{LS}} = 0$$

Least squares estimation

- Definition

$$\hat{\beta}_{LS} := \min_{\beta} \sum_{i=1}^{n} (y_i - x_i^{\mathrm{T}} \beta)^2$$

- Equivalently,

- Equivalently,

$$\hat{\beta}_{LS} := (X^T X)^{-1} X^T y$$

*L*2 distance

Least squares estimation

- Definition

$$\hat{\beta}_{LS} := \min_{\beta} \sum_{i=1}^{n} (y_i - x_i^{\mathrm{T}} \beta)^2$$

- Equivalently,

$$\frac{\partial}{\partial \beta} (y - X\beta)^T (y - X\beta) \Big|_{\beta = \hat{\beta}_{LS}} = 0$$

- Equivalently,

$$\underset{p \times 1}{\hat{\beta}_{LS}} := \underbrace{(\underset{p\times n}{X^T} \underset{n\times p}{X})^{-1}}_{p\times p} \underbrace{(\underset{p\times n}{X^T} \underset{n\times 1}{y})}_{p\times 1}$$

L2 distance

- Equivalently, $\hat{\beta}_{LS}$ is the solution of the score equation

$$X^{\mathrm{T}}(y - X\hat{\beta}_{LS}) = 0$$

?how?

$X^T X$ is invertible

$$\underset{p\times n}{X^T} (\underset{n\times 1}{y} - X \hat{\beta}_{LS}) = \underset{p\times 1}{\underline{0}}$$

$p \times 1$

Least squares estimation

- Definition

$$\hat{\beta}_{LS} := \min_{\beta} \sum_{i=1}^{n} (y_i - x_i^{\mathrm{T}} \beta)^2$$

- Equivalently,

- Equivalently,

$$\hat{\beta}_{LS} :=$$

*L*2 distance

- Equivalently, $\hat{\beta}_{LS}$ is the solution of the score equation

$$X^{\mathrm{T}}(y - X\beta) = 0$$

?how?

- Solution

$$\hat{\beta}_{LS} =$$

check dimensions

... least squares estimation

- Solution

$$\hat{\beta}_{LS} = (X^{\mathrm{T}}X)^{-1}(X^{\mathrm{T}}y)$$

check dimensions

properties of $\hat{\beta}_{LS}$:

$$E(\hat{\beta}_{LS}) = E\{\quad\}$$

$$= (X^T X)^{-1} X^T E(y)$$

$$= \cancel{(X^T X)^{-1}} \cancel{X^T X}\beta = \beta$$

unbiased

ASIDE: here and following all assume X is fixed

… least squares estimation

- Solution

$$\hat{\beta}_{LS} = (X^\top X)^{-1}(X^\top y)$$

check dimensions

- Expected value

$$\mathbb{E}(\hat{\beta}_{LS}) = \quad \beta$$

why?

ASIDE: here and following all assume X is fixed

... least squares estimation

- Solution

$$\hat{\beta}_{LS} = (X^{\mathrm{T}}X)^{-1}(X^{\mathrm{T}}y)$$

check dimensions

- Expected value

$$\mathbb{E}(\hat{\beta}_{LS}) =$$

why?

- Least squares estimates are unbiased

ASIDE: here and following all assume X is fixed

… least squares estimation

$\mathrm{var}(Ay) = A^T \mathrm{var}\, y \, A$

- Solution

$$\hat{\beta}_{LS} = (X^{\mathrm{T}}X)^{-1}(X^{\mathrm{T}}y)$$

check dimensions

$$\begin{pmatrix} x_{11} & \cdots & x_{1p} \\ & & \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

- Expected value

$$\mathbb{E}(\hat{\beta}_{LS}) =$$

why?

- Least squares estimates are unbiased
- Variance

really variance-covariance matrix

$$\mathrm{var}(\hat{\beta}_{LS}) = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\underbrace{\mathrm{var}(y)}_{\sigma^2 I}X(X^{\mathrm{T}}X)^{-1} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\sigma^2 IX(X^{\mathrm{T}}X)^{-1} = \sigma^2(X^{\mathrm{T}}X)^{-1}$$

$\sigma^2 = \mathrm{var}\, \varepsilon_i$

ASIDE: here and following all assume X is fixed

- If we further assume $\underbrace{\epsilon_i \sim N(0, \sigma^2)}$ (and independent across i), then

$$y_i \sim N(x_i^T \beta, \sigma^2)$$ ← density of y_i is $f(y_i)$

$$f(y_1, \dots, y_n \mid x_1, \dots, x_n; \beta, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i - x_i^T\beta)^2}$$

Likelihood fn: $L(\beta, \sigma^2; y) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)}$

What about the normal distribution?

- If we further assume $\epsilon_i \sim N(0, \sigma^2)$ (and independent across i), then
- $y \mid X \sim N(X\beta, \sigma^2 I)$, and

What about the normal distribution?

- If we further assume $\epsilon_i \sim N(0, \sigma^2)$ (and independent across i), then
- $y \mid X \sim N(X\beta, \sigma^2 I)$, and
- the likelihood function is

$$L(\beta, \sigma^2; y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right\},$$

What about the normal distribution?

- If we further assume $\epsilon_i \sim N(0, \sigma^2)$ (and independent across i), then
- $y \mid X \sim N(X\beta, \sigma^2 I)$, and
- the likelihood function is

find β, σ^2 s.t. $L(\beta, \sigma^2)$ is max'd

$$L(\beta, \sigma^2; y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right\},$$

- the log-likelihood function is

$$\ell(\beta, \sigma^2; y) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)^{\mathrm{T}}(y - X\beta),$$

constants in params don't matter

$$\frac{\partial \ell}{\partial \beta} = -\frac{1}{2\sigma^2} X^T(y - X\beta) \Big|_{\hat\beta_{mle}} = 0$$

$$\hat\beta_{mle} = (X^TX)^{-1}X^Ty = \hat\beta_{LS}$$

What about the normal distribution?

- If we further assume $\epsilon_i \sim N(0, \sigma^2)$ (and independent across i), then
- $y \mid X \sim N(X\beta, \sigma^2 I)$, and
- the likelihood function is

$$L(\beta, \sigma^2; y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right\},$$

- the log-likelihood function is

$$\ell(\beta, \sigma^2; y) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)^{\mathrm{T}}(y - X\beta),$$

constants in params don't matter

- the maximum likelihood estimate of β is

$$\hat{\beta}_{ML} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y = \hat{\beta}_{LS}$$

... what about the normal distribution?

- maximum likelihood estimate of β is

$$\hat{\beta}_{ML} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y = \hat{\beta}_{LS}$$

… what about the normal distribution?

- maximum likelihood estimate of β is

$$\hat{\beta}_{ML} = \underbrace{(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}}_{p \times n}\underbrace{y}_{n \times 1} = \hat{\beta}_{LS}$$

$E(\hat{\beta}_{LS}) = \beta$

$var(\hat{\beta}_{LS}) = \sigma^2 I$

- distribution of $\hat{\beta}$ is normal

l.c. of y's

$$\hat{\beta} \sim N_p(\beta, \sigma^2(X^{\mathrm{T}}X)^{-1})$$

why?

… what about the normal distribution?

- maximum likelihood estimate of β is

$$\underline{\hat{\beta}_{ML}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y = \underline{\hat{\beta}_{LS}}$$

- distribution of $\hat{\beta}$ is normal

why?

$$\underline{\hat{\beta}} \sim N_{p}(\beta, \sigma^2(X^{\mathrm{T}}X)^{-1}) \quad \checkmark$$

- distribution of $\hat{\beta}_j$ is

$$N(\beta_j, \sigma^2(X^{\mathrm{T}}X)^{-1}_{jj}), \quad j = 1, \dots, p$$

$(X^{T}X)^{-1}$

$$\cancel{\mathrm{corr}}\ \mathrm{cov}(\hat{\beta}_j, \hat{\beta}_k) = \sigma^2 (X^{T}X)^{-1}_{jk}$$

$E(\hat{\sigma}^2) = \frac{n-p}{n} \sigma^2$

- maximum likelihood estimate of β is

$$\hat{\beta}_{ML} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y = \hat{\beta}_{LS}$$

$E(\tilde{\sigma}^2) = \sigma^2$

- distribution of $\hat{\beta}$ is normal

why?

$$\hat{\beta} \sim N_p(\beta, \sigma^2(X^{\mathrm{T}}X)^{-1})$$

- distribution of $\hat{\beta}_j$ is

$$N(\beta_j, \sigma^2(X^{\mathrm{T}}X)^{-1}_{jj}), \quad j = 1, \ldots, p$$

- maximum likelihood estimate of σ^2 is $\underline{\frac{1}{n}(y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta})}$ $= \frac{1}{n}\sum (y_i - x_i^{\mathrm{T}}\hat{\beta})^2$

$\left.\frac{\partial l}{\partial \sigma^2}\right|_{\hat{\sigma}^2, \hat{\beta}} = 0$

$\hat{\sigma}^2$

$\tilde{\sigma}^2 = \frac{1}{n-p}(y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta})$

… what about the normal distribution?

- maximum likelihood estimate of β is

$$\hat{\beta}_{ML} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y = \hat{\beta}_{LS}$$

- distribution of $\hat{\beta}$ is normal

$$\hat{\beta} \sim N_{p}(\beta, \sigma^2(X^{\mathrm{T}}X)^{-1})$$

- distribution of $\hat{\beta}_j$ is

$$N(\beta_j, \sigma^2(X^{\mathrm{T}}X)^{-1}_{jj}), \quad j = 1, \dots, p$$

- maximum likelihood estimate of σ^2 is $\frac{1}{n}(y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta})$
- but we use

$$\tilde{\sigma}^2 = \frac{1}{n-p}(y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta})$$

$$\frac{\partial}{\partial \beta}(y - X\beta)^{\mathrm{T}}(y - X\beta) = 0$$

$$\Downarrow$$

why?

$$X^{\mathrm{T}}(y - X\beta) = 0$$

$$\frac{\partial}{\partial \beta_j}\sum (y_i - x_i^{\mathrm{T}}\beta)^2$$

$$= \sum_{i=1}^{n} 2(y_i \dots) x_{ij}$$

Pause

(1) I'm lost

(2) I'm good

(3) I'm bored

HW Question Week 1

STA2101F 2022

Due September 21 2022 11.59 pm

Homework to be submitted through Quercus

You can submit this HW in Word, Latex, or R Markdown, but in future please use R Markdown. If you are using Word or Latex with a R script for the computational work, then this R script should be provided as an Appendix. In the document itself you would just include properly formatted output.

You are welcome to discuss questions with others, but the solutions and code must be written independently. Any R output that is included in a solution should be formatted as part of the discussion (i.e. not cut and pasted from the Console).

The dataset `wafer` concerns a study on semiconductors. You can get more information about the data with `?wafer`; you will first need `library(faraway);data(wafer)`, and possibly `install.packages("faraway")`. The questions below are adapted from LM Ch.3.

(a) Fit the linear model `resist ~ x1 + x2 + x3 + x4`. Extract the X matrix using the

Inference

- If you really like likelihood theory, the expected Fisher information is SM §8.2.3

$$\mathcal{I}(\beta, \sigma^2) = \begin{pmatrix} \sigma^{-2}X^{\mathrm{T}}X & 0 \\ 0 & \frac{1}{2}n\sigma^{-4} \end{pmatrix}$$

optional

$\mathcal{I}^{-1}$ gives (asymptotic) variance of MLE

- If you really like likelihood theory, the expected Fisher information is   SM §8.2.3

$$\mathcal{I}(\beta, \sigma^2) = \begin{pmatrix} \sigma^{-2}X^{\mathrm{T}}X & 0 \\ 0 & \frac{1}{2}n\sigma^{-4} \end{pmatrix}$$

$\mathcal{I}^{-1}$ gives (asymptotic) variance of MLE

- but just using previous slide we have

$\hat{\beta} \sim N_p(\beta, \sigma^2(X^{T}X)^{-1})$

$$\frac{\hat{\beta}_j - \beta_j}{\sigma[\{(X^{\mathrm{T}}X)^{-1}\}_{jj}]^{1/2}} \sim N(0, 1)$$

- If you really like likelihood theory, the expected Fisher information is SM §8.2.3

$$\mathcal{I}(\beta, \sigma^2) = \begin{pmatrix} \sigma^{-2} X^{\mathrm{T}} X & 0 \\ 0 & \frac{1}{2} n \sigma^{-4} \end{pmatrix}$$

$\mathcal{I}^{-1}$ gives (asymptotic) variance of MLE

- but just using previous slide we have

$$\frac{\hat{\beta}_j - \beta_j}{\sigma [\{(X^{\mathrm{T}} X)^{-1}\}_{jj}]^{1/2}} \sim N(0, 1)$$

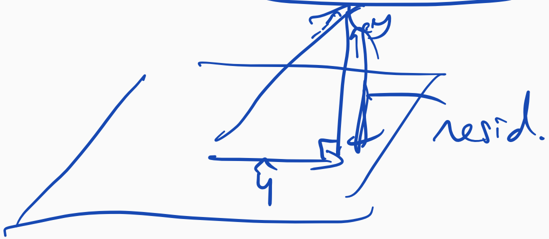
- and

$$\frac{\hat{\beta}_j - \beta_j}{\tilde{\sigma} [\{(X^{\mathrm{T}} X)^{-1}\}_{jj}]^{1/2}} \sim \underline{\underline{t_{n-p}}}$$ — inference

$$\tilde{\sigma}^2 = \frac{1}{n-p} (y - X\hat{\beta})^{\mathrm{T}} (y - X\hat{\beta})$$

Example

```
install.packages("faraway")
library(faraway)
data(prostate)
head(prostate)
```



$$\hat{y} = X\hat{\beta}$$

Example


```
install.packages("faraway")
library(faraway)
data(prostate)
head(prostate)
```

lpsa

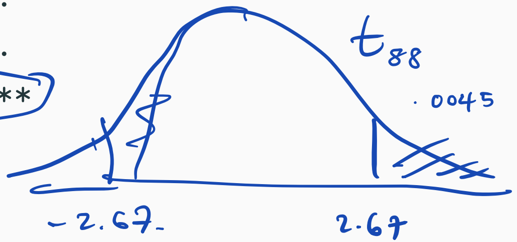
```
model1 <- lm(lpsa ~ ., data = prostate)

summary(model1)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.669337   1.296387   0.516  0.60693
lcavol        0.587022   0.087920   6.677 2.11e-09 ***
lweight       0.454467   0.170012   2.673  0.00896 **
age          -0.019637   0.011173  -1.758  0.08229 .
lbph          0.107054   0.058449   1.832  0.07040
```

```
summary(model1)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.669337   1.296387   0.516  0.60693
lcavol        0.587022   0.087920   6.677  2.11e-09 ***
lweight       0.454467   0.170012   2.673  0.00896 **
age          -0.019637   0.011173  -1.758  0.08229 .
lbph          0.107054   0.058449   1.832  0.07040 .
svi           0.766157   0.244309   3.136  0.00233 **
lcp          -0.105474   0.091013  -1.159  0.24964
gleason       0.045142   0.157465   0.287  0.77503
pgg45         0.004525   0.004421   1.024  0.30886
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\,(X^T X)^{-1}_{jj}} \sim t_{n-p}$$

test of $H_0: \beta_j = 0$ vs $H_A: \beta_j \neq 0$



$97 - 8 - 1 = 88$

322 views

Women in Statistics and Data Science

Follow @WomenInStat 18.7K followers

Jul 23rd 2020, 14 tweets, 4 min read

Bookmark Save as PDF + My Authors

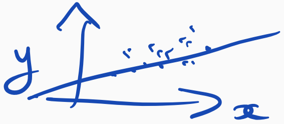
Today, we’re going to play a game I’m calling “IT’S JUST A LINEAR MODEL” (IJALM).

It works like this: I name a model for a quantitative response Y, and then you guess whether or not IJALM.

$$y = X\beta + \varepsilon$$

$$E(y) = X\beta$$

- $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n$



$mx + b$

1st column of X?

$E(y_i \mid x_i) = \beta_0 + \beta_1 x_i$

int

slope

$$E(Y \mid X) = X\beta \quad \underbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{pmatrix}}_{n\times 2} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}}_{2\times 1}$$

$$E(Y)_{n\times 1}$$

- $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n$

1st column of X?

- $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5 + \varepsilon_i$

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^5 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^5 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_5 \end{pmatrix}$$

1st column of X?

- $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n$
- $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5 \epsilon_i$
- $y_i = \beta_0 \pm \beta_1 + \epsilon_i$

$x_i = \underline{+}\, \underline{1}$

$\beta_0 + \beta_1$

$\beta_0 - \beta_1$

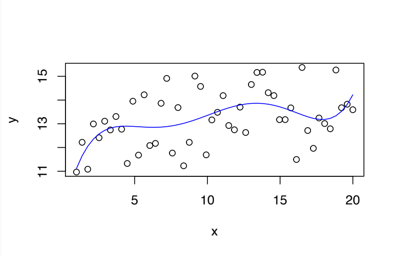
$$\begin{pmatrix} 1 & +1 \\ & \vdots \\ 1 & \\ 1 & +1 \\ 1 & -1 \\ & \vdots \\ 1 & -1 \end{pmatrix}$$

Many special cases $\mathbb{E}(Y \mid X) = X\beta$

1st column of X?

- $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n$
- $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5 \epsilon_i$
- $y_i = \beta_0 \pm \beta_1 + \epsilon_i$
- $y_i = \beta_0 + \underline{\beta_1 \sin(x_i) + \beta_2 \cos(x_i)} + \epsilon_i$

$+ \beta_3 \sin(2x_i) + \beta_4 \cos(2x_i)$



$1 \quad \sin x_1$

$\vdots \quad \vdots \quad \cdots$

$1 \quad \sin x_n$

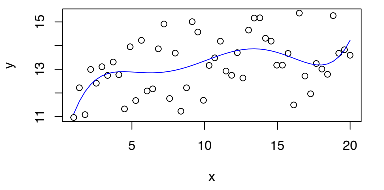
- $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \dots, n$

1st column of X?

- $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5 \epsilon_i$

- $y_i = \beta_0 \pm \beta_1 + \epsilon_i$

- $y_i = \beta_0 + \beta_1 \sin(x_i) + \beta_2 \cos(x_i) + \epsilon_i$



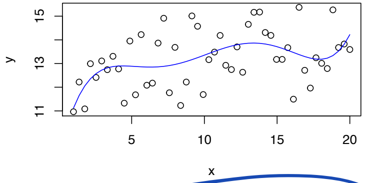
- $y_i = \gamma_0 x_{1i}^{\gamma_1} x_{2i}^{\gamma_2} \eta_i, \quad \eta_i \sim$ positive r.v.

$$\log y_i = \log \gamma_0 + \gamma_1 \underbrace{\log x_{1i}} + \gamma_2 \underbrace{\log x_{2i}} + \underbrace{\log \eta_i}_{\varepsilon_i}$$

Many special cases $\mathbb{E}(Y \mid X) = X\beta$

- $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n$ 1st column of X?
- $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \beta_5 x_i^5 \epsilon_i$
- $y_i = \beta_0 \pm \beta_1 + \epsilon_i$
- $y_i = \beta_0 + \beta_1 \sin(x_i) + \beta_2 \cos(x_i) + \epsilon_i$
- $y_i = \gamma_0 x_{1i}^{\gamma_1} x_{2i}^{\gamma_2} \eta_i, \quad \eta_i \sim$ positive r.v.
- $y_i = \varphi_0 + \sum_{k=1}^{K} \varphi_k \underbrace{s_k(x_i)}_{\text{spline basis}} + \epsilon_i$

$s(x_1) \cdots s(x_K)$



e.g smoothing splines

The linear model

- expected value $\mathbb{E}(y) =$ linear in β

$\longrightarrow$ `Sep152021.Rmd`

The linear model

- expected value $\mathbb{E}(y) =$ linear in β
- measured with additive error $y = \mathbb{E}(y) + \epsilon, \quad \epsilon \sim$

$\longrightarrow$ `Sep142022.Rmd`

The linear model

response

some matrix of data inent-dep

- expected value $\mathbb{E}(y) = X\beta$ linear in β
- measured with additive error $y = \mathbb{E}(y) + \epsilon,$ $\epsilon \sim (0, \sigma^2 I)$
- generalizations

1 → $\epsilon \sim (0, \Sigma)$

matrix of var/cov.

3 $\log(EY) = X\beta$

$g(EY) = X\beta$

gen'd linear models

2 → non-additive error

$f(y \mid x; \beta, \varphi)$

linear "variance"

$\longrightarrow$ Sep152021.Rmd

$E(y) = X\beta$

Today

1. Course introduction: technical issues, course details, evaluation, syllabus, people
2. Upcoming events of interest
3. Review of linear regression
4. In the news: $\longrightarrow$ at the conference