Math 6040/7260 Linear Models

Mon/Wed/Fri 10:55am - 11:40am

Instructor: Dr. Xiang Ji, xji4@tulane.edu

1 Lecture 1:Jan 20

Today

- Introduction
- Course logistics
- Read JF chapter 1, JM Appendix A

What is this course about?

The term "linear models" describes a wide class of methods for the statistical analysis of multivariate data. The underlying theory is grounded in linear algebra and multivariate statistics, but applications range from biological research to public policy. The objective of this course is to provide a solid introduction to both the theory and practice of linear models, combining mathematical concepts with realistic examples.

A hierarchy of linear models

- The linear mean model:
$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times p}{\mathbf{X}} \underset{p\times 1}{\beta} + \underset{n\times 1}{\epsilon}$$
where $\mathbf{E}(\epsilon) = \mathbf{0}$. Only assumption is that errors have mean 0.

- Gauss-Markov model:
$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\epsilon}$$
where $\mathbf{E}(\epsilon) = \mathbf{0}$ and $\mathbf{Var}(\epsilon) = \sigma^2\mathbf{I}$. Uncorrelated errors with constant variance.

- Aitken model or general linear model:
$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\epsilon}$$
where $\mathbf{E}(\epsilon) = \mathbf{0}$ and $\mathbf{Var}(\epsilon) = \sigma^2\mathbf{V}$. $\mathbf{V}$ is fixed and known.

- Variance components models: $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma_1^2\mathbf{V}_1 + \sigma_2^2\mathbf{V}_2 + \cdots + \sigma_r^2\mathbf{V}_r)$ with $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_r$ known.

- General mixed linear Model:
$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\epsilon}$$
where $\mathbf{E}(\epsilon) = \mathbf{0}$ and $\mathbf{Var}(\epsilon) = \boldsymbol{\Sigma}(\theta)$.

- Generalized linear models (GLMs). Logistic regression, probit regression, log-linear model (Poisson regression), ... Note the difference from the general linear model. GLMs are generalization of the *concept* of linear models. They are covered in Math 7360 - Data Analysis class (https://tulane-math7360.github.io/lectures/).

Syllabus

Check course website frequently for updates and announcements.
https://tulane-math-7260-2021.github.io/

HW submission

Through Github with demo on Friday class.

2 Lecture 2:Jan 22

Last time

- Introduction
- Course logistics

Today

- Introduce yourself (remind remote students to record a short video)
  - basic info (name, department, year, ...)
  - why taking this course
- Git
- Linear algebra: vector and vector space, rank of a matrix

What is git?

Git is currently the most popular system for version control according to Google Trend. Git was initially designed and developed by Linus Torvalds in 2005 for Linux kernel development. Git is the British English slang for unpleasant person.

Why using git?

- GitHub is becoming a de facto central repository for open source development.
- **Advertise** yourself through GitHub (e.g., host a free personal webpage on GitHub).
- a skill that employers look for (according to this AmStat article).

Git workflow

Figure 2.1 shows its basic workflow.

What do I need to use Git?

- A **Git server** enabling multi-person collaboration through a centralized repository.
- A **Git client** on your own machine.
  - Linux: Git client program is shipped with many Linux distributions, e.g., Ubuntu and CentOS. If not, install using a package manager, e.g., `yum install git` on CentOS.
  - Mac: follow instructions at https://www.atlassian.com/git/tutorials/install-git.

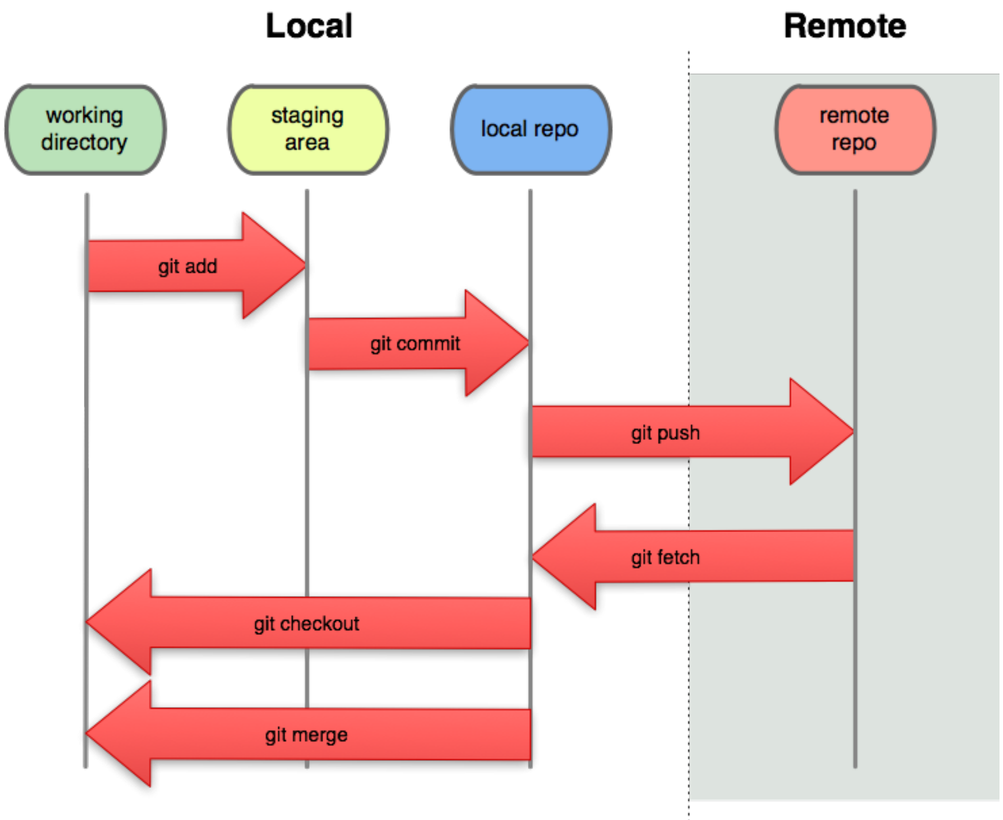

Figure 2.1

  - Windows: Git for Windows at https://gitforwindows.org (GUI) aka Git Bash.

- Do **not** totally rely on GUI or IDE. Learn to use Git on command line, which is needed for cluster and cloud computing.

Git survival commands

- `git pull` synchronize local Git directory with remote repository.
- Modify files in local working directory.
- `git add FILES` add snapshots to staging area
- `git commit -m "message"` store snapshots permanently to (**local**) Git repository
- `git push` push commits to remote repository.

Git basic usage

Working with your local copy.

- `git pull` : update local Git repository with remote repository (fetch + merge).
- `git log FILENAME` : display the current status of working directory.
- `git diff` : show differences (by default difference from the most recent commit).
- `git add file1 file2 ...` : add file(s) to the staging area.
- `git commit` : commit changes in staging area to Git directory.
- `git push` : publish commits in local Git repository to remote repository.
- `git reset –soft HEAD 1` : undo the last commit.
- `git checkout FILENAME` : go back to the last commit, discarding all changes made.
- `git rm FILENAME` : remove files from git control.

Vector and vector space

(from JM Appendix A)

- A set of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are *linearly dependent* if there exist coefficients c_j for $j = 1, 2, \ldots, n$ such that $\sum_{j=1}^n c_j \mathbf{x}_j = \mathbf{0}$ and $||\mathbf{c}||_2 = \sum_{j=1}^n c_j^2 > 0$. They are *linearly independent* if $\sum_{j=1}^n c_j \mathbf{x}_j = \mathbf{0}$ implies $c_j = 0$ for all j.

- Two vectors are *orthogonal* to each other, written $\mathbf{x} \perp \mathbf{y}$, if their inner product is 0, that is $\mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x} = \sum_j x_j y_j = 0$.

- A set of vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}$ are mutually orthogonal iff $\mathbf{x}^{(i)T} \mathbf{x}^{(j)} = 0$ for $\forall i \neq j$.

- The most common set of vectors that are mutually orthogonal are the *elementary* vectors $\mathbf{e}^{(1)}, \mathbf{e}^{(2)}, \ldots, \mathbf{e}^{(n)}$, which are all zero, except for one element equal to 1, so that $\mathbf{e}_i^{(i)} = 1$ and $\mathbf{e}_j^{(i)} = 0, \forall j \neq i$.

- A *vector space* $\mathcal{S}$ is a set of vectors that are closed under addition and scalar multiplication, that is

  - if $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are in $\mathcal{S}$, then $c_1 \mathbf{x}^{(1)} + c_2 \mathbf{x}^{(2)}$ is in $\mathcal{S}$.

- A vector space $\mathcal{S}$ is *generated* or *spanned* by a set of vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}$, written as $\mathcal{S} = \text{span}\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}\}$, if any vector $\mathbf{x}$ in the vector space is a linear combination of $\mathbf{x}_i, i = 1, 2, \ldots, n$.

- A set of linearly independent vectors that generate or span a space $\mathcal{S}$ is called a *basis* of $\mathcal{S}$.

Example A.1

Let

$$\mathbf{x}^{(1)} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{x}^{(2)} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \text{ and } \mathbf{x}^{(3)} = \begin{bmatrix} -3 \\ -1 \\ 1 \\ 3 \end{bmatrix}.$$

Then $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are linearly independent, but $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, and $\mathbf{x}^{(3)}$ are linearly dependent since $5\mathbf{x}^{(1)} - 2\mathbf{x}^{(2)} + \mathbf{x}^{(3)} = 0$

Rank

Some matrix concepts arise from viewing columns or rows of the matrix as vectors. Assume $\mathbf{A} \in \mathbb{R}^{m \times n}$.

- $\text{rank}(\mathbf{A})$ is the maximum number of linearly independent rows or columns of a matrix.

- $\text{rank}(\mathbf{A}) \leqslant \min\{m, n\}$.

- A matrix is *full* rank if $\text{rank}(\mathbf{A}) = \min\{m, n\}$. It is *full row rank* if $\text{rank}(\mathbf{A}) = m$. It is *full column rank* if $\text{rank}(\mathbf{A}) = n$.
- a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is *singular* if $\text{rank}(\mathbf{A}) < n$ and *non-singular* if $\text{rank}(\mathbf{A}) = n$.
- $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T) = \text{rank}(\mathbf{A}^T\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{A}^T)$. (Show this in HW.)
- $\text{rank}(\mathbf{AB}) \leqslant \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}$. (Hint: Columns of $\mathbf{AB}$ are spanned by columns of A and rows of of $\mathbf{AB}$ are spanned by rows of B.)
- if $\mathbf{Ax} = \mathbf{0}_m$ for some $\mathbf{x} \neq \mathbf{0}_n$, then $\text{rank}(\mathbf{A}) \leqslant n - 1$.

3 Lecture 3:Jan 25

Last time

- Git
- Linear algebra: vector and vector space, rank of a matrix

Today

- Column space and Nullspace (JM Appendix A)
- Simple Linear Regression JF Chapter 5

Column space

Definition: The column space of a matrix, denoted by $C(\mathbf{A})$ is the vector space spanned by the columns of the matrix, that is,

$$C(\mathbf{A}) = \{\mathbf{x} : \text{ there exists a vector } \mathbf{c} \text{ such that } \mathbf{x} = \mathbf{Ac}\}.$$

This means that if $\mathbf{x} \in C(\mathbf{A})$, we can find coefficients c_j such that

$$\mathbf{x} = \sum_j c_j \mathbf{a}^{(j)}$$

where $\mathbf{a}^{(j)} = \mathbf{A}_{\cdot j}$ denotes the jth column of matrix $\mathbf{A}$.

- The column space of a matrix consists of all vectors formed by multiplying that matrix by any vector.
- The number of basis vectors for $C(\mathbf{A})$ is then the number of linearly independent columns of the matrix $\mathbf{A}$, and so, $\dim(C(\mathbf{A})) = \text{rank}(\mathbf{A})$.
- The dimension of a space is the number of vectors in its basis.

Example A.2

Let $\mathbf{A} = \begin{bmatrix} 1 & 1 & -3 \\ 1 & 2 & -1 \\ 1 & 3 & 1 \\ 1 & 4 & 3 \end{bmatrix}$ and $\mathbf{c} = \begin{bmatrix} 5 \\ 4 \\ 3 \end{bmatrix}$. Show that $\mathbf{Ac}$ is a linear combination of columns in $\mathbf{A}$.

solution:

$$\mathbf{Ac} = \begin{bmatrix} 1 \times 5 + 1 \times 4 + (-3) \times 3 \\ 1 \times 5 + 2 \times 4 + (-1) \times 3 \\ 1 \times 5 + 3 \times 4 + 1 \times 3 \\ 1 \times 5 + 4 \times 4 + 3 \times 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 10 \\ 20 \\ 30 \end{bmatrix}.$$

You could recognize that

$$\mathbf{A}\mathbf{c} = 5 \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + 4 \times \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} + 3 \times \begin{bmatrix} -3 \\ -1 \\ 1 \\ 3 \end{bmatrix} = 5\mathbf{a}^{(1)} + 4\mathbf{a}^{(2)} + 3\mathbf{a}^{(3)} = \begin{bmatrix} 0 \\ 10 \\ 20 \\ 30 \end{bmatrix}.$$

Result A.1

$\text{rank}(\mathbf{AB}) \leqslant \min(\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B}))$.

proof: Each column of $\mathbf{AB}$ is a linear combination of columns of $\mathbf{A}$ (i.e. $(\mathbf{AB})_{\cdot j} = \mathbf{A}\mathbf{b}^{(j)}$), so the number of linearly independent columns of $\mathbf{AB}$ cannot be greater than that of A. Similarly, $\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{B}^T\mathbf{A}^T)$, the same argument gives $\text{rank}(\mathbf{B}^T)$ as an upper bound.

Result A.2

- (a) If $\mathbf{A} = \mathbf{BC}$, then $C(\mathbf{A}) \subseteq C(\mathbf{B})$.
- (b) If $C(\mathbf{A}) \subseteq C(\mathbf{B})$, then there exists a matrix $\mathbf{C}$ such that $\mathbf{A} = \mathbf{BC}$.

proof: For (a), any vector $\mathbf{x} \in C(\mathbf{A})$ can be written as $\mathbf{x} = \mathbf{A}\mathbf{d} = \mathbf{B}(\mathbf{C}\mathbf{d})$. For (b), $\mathbf{A}_{\cdot j} \in C(B)$, so that there exists a vector $\mathbf{c}^{(j)}$ such that $\mathbf{A}_{\cdot j} = \mathbf{B}\mathbf{c}^{(j)}$. The matrix $\mathbf{C} = (\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \ldots, \mathbf{c}^{(n)})$ satisfies that $\mathbf{A} = \mathbf{BC}$.

Null space

Definition: The null space of a matrix, denoted by $N(\mathbf{A})$, is $N(\mathbf{A}) = \{\mathbf{y} : \mathbf{A}\mathbf{y} = \mathbf{0}\}$.

Result A.3

If $\mathbf{A}$ has full-column rank, then $N(\mathbf{A}) = \{\mathbf{0}\}$.

proof: Matrix $\mathbf{A}$ has full-column rank means its columns are linearly independent, which means that $\mathbf{A}\mathbf{c} = \mathbf{0}$ implies $\mathbf{c} = \mathbf{0}$.

Theorem A.1

Assume $\mathbf{A} \in \mathbb{R}^{m \times n}$, then $\dim(C(\mathbf{A})) = r$ and $\dim(N(\mathbf{A})) = n - r$, where $r = \text{rank}(\mathbf{A})$.

See JM Appendix Theorem A.1 for the proof.

Interpretation: "dimension of column space + dimension of null space = # columns"

*Mis*Interpretation: Columns space and null space are orthogonal complement to each other. They are of different orders in general! Next result gives the correct statement.

Simple linear regression

Figure 3.1 shows Davis's data on the measured and reported weight in kilograms of 101 women who were engaged in regular exercise.

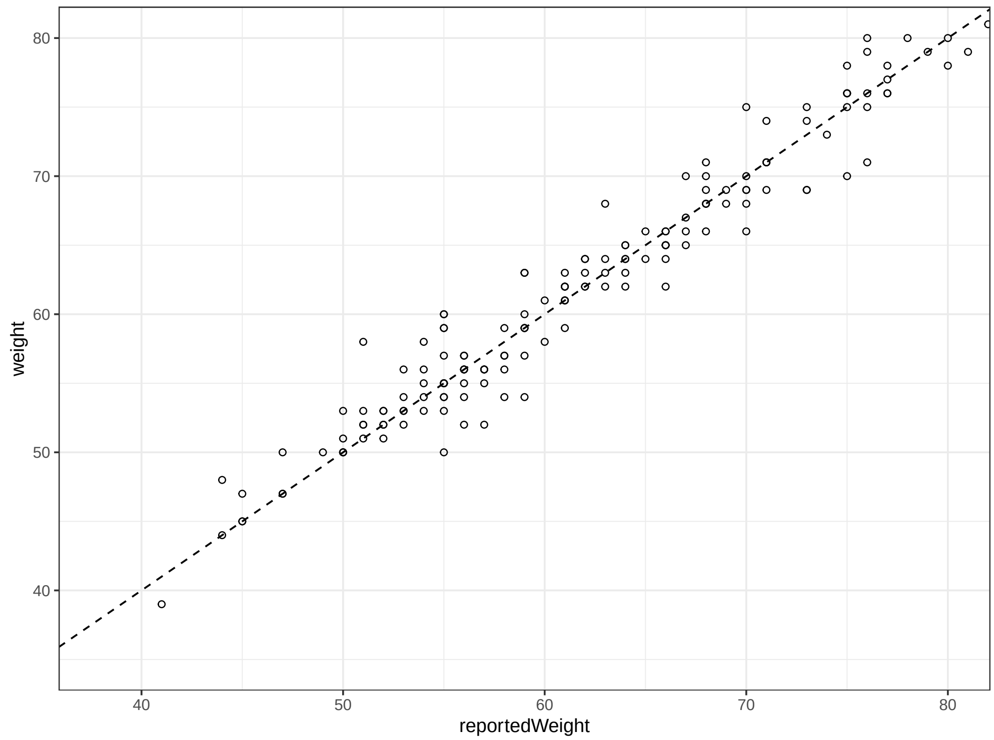

Figure 3.1: Scatterplot of Davis's data on the measured and reported weight of 101 women. The dashed line gives $y = x$.

It's reasonable to assume that the relationship between measured and reported weight appears to be linear. Denote:

- measured weight by y_i: **response variable** or **dependent variable**
- reported weight by x_i: **predictor variable** or **independent variable**
- intercept: β_0
- slope: β_1
- residual/error term ϵ_i.

Then the simple linear regression model writes:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i.$$

For given $(\hat{\beta}_0, \hat{\beta}_1)$ values, the *fitted value* or *predicted value* for observation i is:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i.$$

Therefore, the residual is

$$\epsilon_i = y_i - \hat{y}_i$$

Fitting a linear model

Choose the "best" values for β_0, β_1 such that

$$SS[E] = \sum_1^n \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)^2 = \sum_1^n (y_i - \hat{y}_i)^2 = \sum_1^n \epsilon_i^2$$

is minimized. These are **least squares** (LS) estimates:

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 &= \tfrac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}. \end{aligned}$$

Definition: The line satisfying the equation

$$y = \hat{\beta}_0 + \hat{\beta}_1 x$$

is called the linear regression of y on x which is also called the least squares line.

For Davis's data, we have

$$\begin{aligned} n &= 101 \\ \bar{y} &= \frac{5780}{101} = 57.228 \\ \bar{x} &= \frac{5731}{101} = 56.743 \\ \sum(x_i - \bar{x})(y_i - \bar{y}) &= 4435.9 \\ \sum(x_i - \bar{x})^2 &= 4539.3, \end{aligned}$$

so that

$$\begin{aligned} \hat{\beta}_1 &= \frac{4435.9}{4539.3} = 0.97722 \\ \hat{\beta}_0 &= 57.228 - 0.97722 \times 56.743 = 1.7776 \end{aligned}$$