Models & Estimation

Merlise Clyde

STA721 Linear Models

Duke University

August 28, 2014

Outline

Readings: Christensen Chapter 1-2, Appendix A

Models

Take an random vector $\mathbf{Y} \in \mathbb{R}^n$ which is observable and decompose

Models

Take an random vector $\mathbf{Y} \in \mathbb{R}^n$ which is observable and decompose

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

into $\boldsymbol{\mu} \in \mathbb{R}^n$ (unknown, fixed)

Models

Take an random vector $\mathbf{Y} \in \mathbb{R}^n$ which is observable and decompose

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

into $\boldsymbol{\mu} \in \mathbb{R}^n$ (unknown, fixed) and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ unobservable error vector (random)

Models

Take an random vector $\mathbf{Y} \in \mathbb{R}^n$ which is observable and decompose

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

into $\boldsymbol{\mu} \in \mathbb{R}^n$ (unknown, fixed) and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ unobservable error vector (random)

Usual assumptions?

Models

Take an random vector $\mathbf{Y} \in \mathbb{R}^n$ which is observable and decompose

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

into $\boldsymbol{\mu} \in \mathbb{R}^n$ (unknown, fixed) and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ unobservable error vector (random)

Usual assumptions?

- $\mathsf{E}[\boldsymbol{\epsilon}] = \mathbf{0}$

Models

Take an random vector $\mathbf{Y} \in \mathbb{R}^n$ which is observable and decompose

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

into $\boldsymbol{\mu} \in \mathbb{R}^n$ (unknown, fixed) and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ unobservable error vector (random)

Usual assumptions?

- $\mathsf{E}[\boldsymbol{\epsilon}] = \mathbf{0} \quad \Rightarrow \mathsf{E}[\mathbf{Y}] = \boldsymbol{\mu}$ (mean vector)

Models

Take an random vector $\mathbf{Y} \in \mathbb{R}^n$ which is observable and decompose

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

into $\boldsymbol{\mu} \in \mathbb{R}^n$ (unknown, fixed) and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ unobservable error vector (random)

Usual assumptions?

- $\mathsf{E}[\boldsymbol{\epsilon}] = \mathbf{0} \quad \Rightarrow \mathsf{E}[\mathbf{Y}] = \boldsymbol{\mu}$ (mean vector)
- $\mathsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}_n$

Models

Take an random vector $\mathbf{Y} \in \mathbb{R}^n$ which is observable and decompose

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

into $\boldsymbol{\mu} \in \mathbb{R}^n$ (unknown, fixed) and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ unobservable error vector (random)

Usual assumptions?

- $\mathsf{E}[\boldsymbol{\epsilon}] = \mathbf{0} \quad \Rightarrow \mathsf{E}[\mathbf{Y}] = \boldsymbol{\mu}$ (mean vector)
- $\mathsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}_n \quad \Rightarrow \mathsf{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$ (errors are uncorrelated)

Models

Take an random vector $\mathbf{Y} \in \mathbb{R}^n$ which is observable and decompose

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

into $\boldsymbol{\mu} \in \mathbb{R}^n$ (unknown, fixed) and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ unobservable error vector (random)

Usual assumptions?

- $\mathsf{E}[\boldsymbol{\epsilon}] = \mathbf{0} \quad \Rightarrow \mathsf{E}[\mathbf{Y}] = \boldsymbol{\mu}$ (mean vector)
- $\mathsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}_n \quad \Rightarrow \mathsf{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$ (errors are uncorrelated)
- $\boldsymbol{\epsilon} \sim \mathsf{N}(\mathbf{0}, \sigma^2 \mathbf{I})$

Models

Take an random vector $\mathbf{Y} \in \mathbb{R}^n$ which is observable and decompose

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

into $\boldsymbol{\mu} \in \mathbb{R}^n$ (unknown, fixed) and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ unobservable error vector (random)

Usual assumptions?

- $\mathsf{E}[\boldsymbol{\epsilon}] = \mathbf{0} \quad \Rightarrow \mathsf{E}[\mathbf{Y}] = \boldsymbol{\mu}$ (mean vector)
- $\mathsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}_n \quad \Rightarrow \mathsf{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$ (errors are uncorrelated)
- $\boldsymbol{\epsilon} \sim \mathsf{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \quad \Rightarrow \mathbf{Y} \sim \mathsf{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$

Models

Take an random vector $\mathbf{Y} \in \mathbb{R}^n$ which is observable and decompose

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

into $\boldsymbol{\mu} \in \mathbb{R}^n$ (unknown, fixed) and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ unobservable error vector (random)

Usual assumptions?

- $\mathsf{E}[\boldsymbol{\epsilon}] = \mathbf{0} \quad \Rightarrow \mathsf{E}[\mathbf{Y}] = \boldsymbol{\mu}$ (mean vector)
- $\mathsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}_n \quad \Rightarrow \mathsf{Cov}[\mathbf{Y}] = \sigma^2 \mathbf{I}_n$ (errors are uncorrelated)
- $\boldsymbol{\epsilon} \sim \mathsf{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \quad \Rightarrow \mathbf{Y} \sim \mathsf{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$

The distribution assumption allows us to right down a likelihood function

Likelihood Functions

The likelihood function for $\boldsymbol{\mu}, \sigma^2$ is proportional to the sampling distribution of the data

Likelihood Functions

The likelihood function for $\boldsymbol{\mu}, \sigma^2$ is proportional to the sampling distribution of the data

$$\mathcal{L}(\boldsymbol{\mu}, \sigma^2) \quad \propto \quad (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2}\frac{(\mathbf{Y}-\boldsymbol{\mu})^T(\mathbf{Y}-\boldsymbol{\mu})}{\sigma^2}\right\}$$

Likelihood Functions

The likelihood function for $\boldsymbol{\mu}, \sigma^2$ is proportional to the sampling distribution of the data

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\mu}, \sigma^2) &\propto (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2} \frac{(\mathbf{Y} - \boldsymbol{\mu})^T(\mathbf{Y} - \boldsymbol{\mu})}{\sigma^2} \right\} \\
&\propto (\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2} \frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2} \right\}
\end{aligned}
$$

Likelihood Functions

The likelihood function for $\boldsymbol{\mu}, \sigma^2$ is proportional to the sampling distribution of the data

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\mu}, \sigma^2) &\propto (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2} \frac{(\mathbf{Y} - \boldsymbol{\mu})^T (\mathbf{Y} - \boldsymbol{\mu})}{\sigma^2} \right\} \\
&\propto (\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2} \frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2} \right\}
\end{aligned}
$$

Find values of $\hat{\boldsymbol{\mu}}$ and $\hat{\sigma}^2$ that maximize the likelihood $\mathcal{L}(\boldsymbol{\mu}, \sigma^2)$ for $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\sigma^2 \in \mathbb{R}^+$

Likelihood Functions

The likelihood function for $\boldsymbol{\mu}, \sigma^2$ is proportional to the sampling distribution of the data

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\mu}, \sigma^2) &\propto (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2}\frac{(\mathbf{Y}-\boldsymbol{\mu})^T(\mathbf{Y}-\boldsymbol{\mu})}{\sigma^2}\right\} \\
&\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2}\frac{\|\mathbf{Y}-\boldsymbol{\mu}\|^2}{\sigma^2}\right\}
\end{aligned}$$

Find values of $\hat{\boldsymbol{\mu}}$ and $\hat{\sigma}^2$ that maximize the likelihood $\mathcal{L}(\boldsymbol{\mu}, \sigma^2)$ for $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\sigma^2 \in \mathbb{R}^+$

Clearly, $\hat{\boldsymbol{\mu}} = \mathbf{Y}$ but $\hat{\sigma}^2 = 0$ is outside the parameter space

Restrictions on $\boldsymbol{\mu}$

Linear model $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ for fixed $\mathbf{X}$ $(n \times p)$ and some $\boldsymbol{\beta} \in \mathbb{R}^p$ restrict the possible values for $\boldsymbol{\mu}$

Restrictions on $\boldsymbol{\mu}$

Linear model $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ for fixed $\mathbf{X}$ $(n \times p)$ and some $\boldsymbol{\beta} \in \mathbb{R}^p$ restrict the possible values for $\boldsymbol{\mu}$

- Oneway Anova Model $j = 1, \ldots, J$ groups with n_j replicates in each group j

Restrictions on $\boldsymbol{\mu}$

Linear model $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ for fixed $\mathbf{X}$ $(n \times p)$ and some $\boldsymbol{\beta} \in \mathbb{R}^p$ restrict the possible values for $\boldsymbol{\mu}$

- Oneway Anova Model $j = 1, \ldots, J$ groups with n_j replicates in each group j
- "cell means" $Y_{ij} = \mu_j + \epsilon_{ij}$

$$\boldsymbol{\mu} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \ldots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \ldots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_J} & \ldots & \mathbf{0}_{n_J} & \mathbf{1}_{n_J} \end{bmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_J \end{pmatrix}$$

- $\mathbf{1}_{n_j}$ is a vector of length n_j of ones.
- $\mathbf{0}_{n_j}$ is a vector of length $n_j h$ of zeros.

Equivalent Model

Linear model $\boldsymbol{\mu} = \mathbf{X}\beta$ for fixed $\mathbf{X}$ $(n \times p)$ and some $\boldsymbol{\beta} \in \mathbb{R}^p$

Equivalent Model

Linear model $\boldsymbol{\mu} = \mathbf{X}\beta$ for fixed $\mathbf{X}$ $(n \times p)$ and some $\beta \in \mathbb{R}^p$

- Oneway Anova Model $j = 1, \ldots, J$ groups with n_j replicates in each group j

Equivalent Model

Linear model $\boldsymbol{\mu} = \mathbf{X}\beta$ for fixed $\mathbf{X}$ $(n \times p)$ and some $\beta \in \mathbb{R}^p$

- Oneway Anova Model $j = 1, \ldots, J$ groups with n_j replicates in each group j
- "Treatment effects" $Y_{ij} = \mu + \tau_j + \epsilon_{ij}$

$$\boldsymbol{\mu} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \ldots & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \ldots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_{n_J} & \mathbf{0}_{n_J} & \ldots & \mathbf{0}_{n_J} & \mathbf{1}_{n_J} \end{bmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_J \end{pmatrix}$$

Equivalent Model

Linear model $\boldsymbol{\mu} = \mathbf{X}\beta$ for fixed $\mathbf{X}$ $(n \times p)$ and some $\beta \in \mathbb{R}^p$

- Oneway Anova Model $j = 1, \ldots, J$ groups with n_j replicates in each group j
- "Treatment effects" $Y_{ij} = \mu + \tau_j + \epsilon_{ij}$

$$\boldsymbol{\mu} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \ldots & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \ldots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_{n_J} & \mathbf{0}_{n_J} & \ldots & \mathbf{0}_{n_J} & \mathbf{1}_{n_J} \end{bmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_J \end{pmatrix}$$

- Equivalent means $\mu_j = \mu + \tau_j$
- Should our inference for $\boldsymbol{\mu}$ depend on how we represent or parameterize $\boldsymbol{\mu}$?

Column Space

Many equivalent ways to represent the same mean vector – inference should be independent of the coordinate system used

Many equivalent ways to represent the same mean vector – inference should be independent of the coordinate system used

- Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p \in \mathbb{R}^n$

Column Space

Many equivalent ways to represent the same mean vector – inference should be independent of the coordinate system used

- Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p \in \mathbb{R}^n$
- The set of all linear combinations of $\mathbf{X}_1, \ldots, \mathbf{X}_p$ is the space spanned by $\mathbf{X}_1, \ldots, \mathbf{X}_p \equiv S(\mathbf{X}_1, \ldots, \mathbf{X}_p)$

Column Space

Many equivalent ways to represent the same mean vector – inference should be independent of the coordinate system used

- Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p \in \mathbb{R}^n$
- The set of all linear combinations of $\mathbf{X}_1, \ldots, \mathbf{X}_p$ is the space spanned by $\mathbf{X}_1, \ldots, \mathbf{X}_p \equiv S(\mathbf{X}_1, \ldots, \mathbf{X}_p)$
- Let $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2 \ldots \mathbf{X}_p]$ be a $n \times p$ matrix with columns $\mathbf{X}_j$ then the column space of $\mathbf{X}$, $C(\mathbf{X}) = S(\mathbf{X}_1, \ldots, \mathbf{X}_p)$ space spanned by the (column) vectors of $\mathbf{X}$

Column Space

Many equivalent ways to represent the same mean vector – inference should be independent of the coordinate system used

- Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p \in \mathbb{R}^n$
- The set of all linear combinations of $\mathbf{X}_1, \ldots, \mathbf{X}_p$ is the space spanned by $\mathbf{X}_1, \ldots, \mathbf{X}_p \equiv S(\mathbf{X}_1, \ldots, \mathbf{X}_p)$
- Let $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2 \ldots \mathbf{X}_p]$ be a $n \times p$ matrix with columns $\mathbf{X}_j$ then the column space of $\mathbf{X}$, $C(\mathbf{X}) = S(\mathbf{X}_1, \ldots, \mathbf{X}_p)$ space spanned by the (column) vectors of $\mathbf{X}$
- $\boldsymbol{\mu} \in C(\mathbf{X}) : C(\mathbf{X}) = \{\boldsymbol{\mu} \mid \boldsymbol{\mu} \in \mathbb{R}^n$ such that $\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\mu}$ for some $\boldsymbol{\beta} \in \mathbb{R}^p\}$ (also called the Range of $\mathbf{X}$, $R(\mathbf{X})$)

Column Space

Many equivalent ways to represent the same mean vector – inference should be independent of the coordinate system used

- Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p \in \mathbb{R}^n$
- The set of all linear combinations of $\mathbf{X}_1, \ldots, \mathbf{X}_p$ is the space spanned by $\mathbf{X}_1, \ldots, \mathbf{X}_p \equiv S(\mathbf{X}_1, \ldots, \mathbf{X}_p)$
- Let $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2 \ldots \mathbf{X}_p]$ be a $n \times p$ matrix with columns $\mathbf{X}_j$ then the column space of $\mathbf{X}$, $C(\mathbf{X}) = S(\mathbf{X}_1, \ldots, \mathbf{X}_p)$ space spanned by the (column) vectors of $\mathbf{X}$
- $\boldsymbol{\mu} \in C(\mathbf{X}) : C(\mathbf{X}) = \{\boldsymbol{\mu} \mid \boldsymbol{\mu} \in \mathbb{R}^n$ such that $\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\mu}$ for some $\boldsymbol{\beta} \in \mathbb{R}^p\}$ (also called the Range of $\mathbf{X}$, $R(\mathbf{X})$)
- $\boldsymbol{\beta}$ are the "coordinates" of $\boldsymbol{\mu}$ in this space

Column Space

Many equivalent ways to represent the same mean vector – inference should be independent of the coordinate system used

- Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p \in \mathbb{R}^n$
- The set of all linear combinations of $\mathbf{X}_1, \ldots, \mathbf{X}_p$ is the space spanned by $\mathbf{X}_1, \ldots, \mathbf{X}_p \equiv S(\mathbf{X}_1, \ldots, \mathbf{X}_p)$
- Let $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2 \ldots \mathbf{X}_p]$ be a $n \times p$ matrix with columns $\mathbf{X}_j$ then the column space of $\mathbf{X}$, $C(\mathbf{X}) = S(\mathbf{X}_1, \ldots, \mathbf{X}_p)$ space spanned by the (column) vectors of $\mathbf{X}$
- $\boldsymbol{\mu} \in C(\mathbf{X})$: $C(\mathbf{X}) = \{\boldsymbol{\mu} \mid \boldsymbol{\mu} \in \mathbb{R}^n$ such that $\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\mu}$ for some $\boldsymbol{\beta} \in \mathbb{R}^p\}$ (also called the Range of $\mathbf{X}$, $R(\mathbf{X})$)
- $\boldsymbol{\beta}$ are the "coordinates" of $\boldsymbol{\mu}$ in this space
- $C(\mathbf{X})$ is a subspace of $\mathbb{R}^n$

Vector spaces

A collection of vectors V is a real **vector space** if the following conditions hold: for any pair $\mathbf{x}$ and $\mathbf{y}$ of vectors in V there corresponds a vector $\mathbf{x} + \mathbf{y}$ and scalars $\alpha, \beta \in \mathbb{R}$ such that:

Vector spaces

A collection of vectors V is a real **vector space** if the following conditions hold: for any pair $\mathbf{x}$ and $\mathbf{y}$ of vectors in V there corresponds a vector $\mathbf{x} + \mathbf{y}$ and scalars $\alpha, \beta \in \mathbb{R}$ such that:

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (vector addition is commutative)

Vector spaces

A collection of vectors V is a real **vector space** if the following conditions hold: for any pair $\mathbf{x}$ and $\mathbf{y}$ of vectors in V there corresponds a vector $\mathbf{x} + \mathbf{y}$ and scalars $\alpha, \beta \in \mathbb{R}$ such that:

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (vector addition is commutative)
2. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (vector addition is associative)

Vector spaces

A collection of vectors V is a real **vector space** if the following conditions hold: for any pair $\mathbf{x}$ and $\mathbf{y}$ of vectors in V there corresponds a vector $\mathbf{x} + \mathbf{y}$ and scalars $\alpha, \beta \in \mathbb{R}$ such that:

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (vector addition is commutative)
2. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (vector addition is associative)
3. there exists a unique vector $\mathbf{0} \in V$ (the origin) such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for every vector $\mathbf{x}$

Vector spaces

A collection of vectors V is a real **vector space** if the following conditions hold: for any pair $\mathbf{x}$ and $\mathbf{y}$ of vectors in V there corresponds a vector $\mathbf{x} + \mathbf{y}$ and scalars $\alpha, \beta \in \mathbb{R}$ such that:

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (vector addition is commutative)
2. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (vector addition is associative)
3. there exists a unique vector $\mathbf{0} \in V$ (the origin) such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for every vector $\mathbf{x}$
4. for every $\mathbf{x} \in V$, there exists a unique vector $-\mathbf{x}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$

Vector spaces

A collection of vectors V is a real **vector space** if the following conditions hold: for any pair $\mathbf{x}$ and $\mathbf{y}$ of vectors in V there corresponds a vector $\mathbf{x} + \mathbf{y}$ and scalars $\alpha, \beta \in \mathbb{R}$ such that:

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (vector addition is commutative)
2. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (vector addition is associative)
3. there exists a unique vector $\mathbf{0} \in V$ (the origin) such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for every vector $\mathbf{x}$
4. for every $\mathbf{x} \in V$, there exists a unique vector $-\mathbf{x}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$
5. $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$ (multiplication by scalars is associative)

Vector spaces

A collection of vectors V is a real **vector space** if the following conditions hold: for any pair $\mathbf{x}$ and $\mathbf{y}$ of vectors in V there corresponds a vector $\mathbf{x} + \mathbf{y}$ and scalars $\alpha, \beta \in \mathbb{R}$ such that:

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (vector addition is commutative)
2. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (vector addition is associative)
3. there exists a unique vector $\mathbf{0} \in V$ (the origin) such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for every vector $\mathbf{x}$
4. for every $\mathbf{x} \in V$, there exists a unique vector $-\mathbf{x}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$
5. $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$ (multiplication by scalars is associative)
6. $1\mathbf{x} = \mathbf{x}$ for every $\mathbf{x}$

Vector spaces

A collection of vectors V is a real **vector space** if the following conditions hold: for any pair $\mathbf{x}$ and $\mathbf{y}$ of vectors in V there corresponds a vector $\mathbf{x} + \mathbf{y}$ and scalars $\alpha, \beta \in \mathbb{R}$ such that:

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (vector addition is commutative)
2. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (vector addition is associative)
3. there exists a unique vector $\mathbf{0} \in V$ (the origin) such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for every vector $\mathbf{x}$
4. for every $\mathbf{x} \in V$, there exists a unique vector $-\mathbf{x}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$
5. $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$ (multiplication by scalars is associative)
6. $1\mathbf{x} = \mathbf{x}$ for every $\mathbf{x}$
7. $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ (multiplication by scalars is distributive with respect vector addition)

Vector spaces

A collection of vectors V is a real **vector space** if the following conditions hold: for any pair $\mathbf{x}$ and $\mathbf{y}$ of vectors in V there corresponds a vector $\mathbf{x} + \mathbf{y}$ and scalars $\alpha, \beta \in \mathbb{R}$ such that:

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (vector addition is commutative)
2. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (vector addition is associative)
3. there exists a unique vector $\mathbf{0} \in V$ (the origin) such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for every vector $\mathbf{x}$
4. for every $\mathbf{x} \in V$, there exists a unique vector $-\mathbf{x}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$
5. $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$ (multiplication by scalars is associative)
6. $1\mathbf{x} = \mathbf{x}$ for every $\mathbf{x}$
7. $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ (multiplication by scalars is distributive with respect vector addition)
8. $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$ (multiplication by scalars is distributive with respect to vector addition)

Subspaces

Definition

Let V be a vector space and let V_0 be a set with $V_0 \subseteq V$. V_0 is a *subspace* of V if and only if V_0 is a vector space.

Subspaces

Definition

Let V be a vector space and let V_0 be a set with $V_0 \subseteq V$. V_0 is a *subspace* of V if and only if V_0 is a vector space.

Theorem

Let V be a vector space, and let V_0 be a non-empty subset of V. If V_0 is closed on vector addition and scalar multiplication, then V_0 is a subspace of V

Subspaces

Definition

Let V be a vector space and let V_0 be a set with $V_0 \subseteq V$. V_0 is a *subspace* of V if and only if V_0 is a vector space.

Theorem

Let V be a vector space, and let V_0 be a non-empty subset of V. If V_0 is closed on vector addition and scalar multiplication, then V_0 is a subspace of V

Theorem

Let V be a vector space, and let $\mathbf{x}_1, \ldots, \mathbf{x}_r$ be vectors in V. The set of all linear combinations of $\mathbf{x}_1, \ldots, \mathbf{x}_r$ is a subspace of V.

Subspaces

Definition

Let V be a vector space and let V_0 be a set with $V_0 \subseteq V$. V_0 is a *subspace* of V if and only if V_0 is a vector space.

Theorem

Let V be a vector space, and let V_0 be a non-empty subset of V. If V_0 is closed on vector addition and scalar multiplication, then V_0 is a subspace of V

Theorem

Let V be a vector space, and let $\mathbf{x}_1, \ldots, \mathbf{x}_r$ be vectors in V. The set of all linear combinations of $\mathbf{x}_1, \ldots, \mathbf{x}_r$ is a subspace of V.

The Column space of $\mathbf{X}$, $C(\mathbf{X})$, is a subspace of $\mathbb{R}^n$

Basis

Definition

A finite set of vectors $\{\mathbf{x}_i\}$ is *linearly dependent* if there exist scalars $\{a_i\}$ (not all zero) such that $\sum_i a_i \mathbf{x}_i = \mathbf{0}$.

Basis

Definition

A finite set of vectors $\{\mathbf{x}_i\}$ is *linearly dependent* if there exist scalars $\{a_i\}$ (not all zero) such that $\sum_i a_i \mathbf{x}_i = \mathbf{0}$. Conversely $\sum_i a_i \mathbf{x}_i = \mathbf{0}$ implies that $a_i = 0 \, \forall \, i$ the set is *linearly independent*.

Definition

A finite set of vectors $\{\mathbf{x}_i\}$ is *linearly dependent* if there exist scalars $\{a_i\}$ (not all zero) such that $\sum_i a_i \mathbf{x}_i = \mathbf{0}$. Conversely $\sum_i a_i \mathbf{x}_i = \mathbf{0}$ implies that $a_i = 0 \, \forall \, i$ the set is *linearly independent*.

Definition

If V_0 is a subspace of V and if $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is a linearly independent spanning set for V_0, then $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is a basis for V_0

Basis

Definition

A finite set of vectors $\{\mathbf{x}_i\}$ is *linearly dependent* if there exist scalars $\{a_i\}$ (not all zero) such that $\sum_i a_i \mathbf{x}_i = \mathbf{0}$. Conversely $\sum_i a_i \mathbf{x}_i = \mathbf{0}$ implies that $a_i = 0 \, \forall \, i$ the set is *linearly independent*.

Definition

If V_0 is a subspace of V and if $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is a linearly independent spanning set for V_0, then $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is a basis for V_0

Theorem

If V_0 is a subspace of V, then all bases for V_0 have the same number of vectors

Basis

Definition

A finite set of vectors $\{\mathbf{x}_i\}$ is *linearly dependent* if there exist scalars $\{a_i\}$ (not all zero) such that $\sum_i a_i \mathbf{x}_i = \mathbf{0}$. Conversely $\sum_i a_i \mathbf{x}_i = \mathbf{0}$ implies that $a_i = 0 \, \forall \, i$ the set is *linearly independent*.

Definition

If V_0 is a subspace of V and if $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is a linearly independent spanning set for V_0, then $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is a basis for V_0

Theorem

If V_0 is a subspace of V, then all bases for V_0 have the same number of vectors

Can both the collection of vectors in the cell means and the treatment effects parameterizations be a basis?

Rank

Definition

The rank of a subspace V_0 is the number of elements in a basis for V_0 and is written as $r(V_0)$. Similarly if $\mathbf{A}$ is a matrix, the rank of $C(\mathbf{A})$ is called the rank of $\mathbf{A}$ and is written $r(\mathbf{A})$.

Rank

Definition

The rank of a subspace V_0 is the number of elements in a basis for V_0 and is written as $r(V_0)$. Similarly if $\mathbf{A}$ is a matrix, the rank of $C(\mathbf{A})$ is called the rank of $\mathbf{A}$ and is written $r(\mathbf{A})$.

What is the rank of the subspace in the Oneway ANOVA model?

Orthogonality

Definition

An inner product space is a vector space V equipped with an inner product: $\langle \cdot, \cdot \rangle$ is a mapping $V \times V \to \mathbb{R}$. Two vectors are orthogonal if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, written $\mathbf{x} \perp \mathbf{y}$

Orthogonality

Definition

An inner product space is a vector space V equipped with an inner product: $\langle \cdot, \cdot \rangle$ is a mapping $V \times V \to \mathbb{R}$. Two vectors are orthogonal if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, written $\mathbf{x} \perp \mathbf{y}$ (usual $\mathbf{x}^T \mathbf{y} = 0$)

Definition

Two subspaces are orthogonal if $\mathbf{x} \in M$ and $y \in N$ imply that $\mathbf{x} \perp \mathbf{y}$

Orthogonality

Definition

An inner product space is a vector space V equipped with an inner product: $\langle \cdot, \cdot \rangle$ is a mapping $V \times V \to \mathbb{R}$. Two vectors are orthogonal if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, written $\mathbf{x} \perp \mathbf{y}$ (usual $\mathbf{x}^T\mathbf{y} = 0$)

Definition

Two subspaces are orthogonal if $\mathbf{x} \in M$ and $y \in N$ imply that $\mathbf{x} \perp \mathbf{y}$

Definition

Orthonormal basis: $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is an orthonormal basis (ONB) for M if $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0$ for $i \neq j$ and $\langle \mathbf{x}_i, \mathbf{x}_i \rangle = 1$ for all i. Length $\mathbf{x} = \|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Distance of two vectors is $\|\mathbf{x} - \mathbf{y}\|$

Find an orthonormal basis for the oneway ANOVA model. Is it unique?

Find an orthonormal basis for the oneway ANOVA model. Is it unique?

Can also use Gram-Schmidt sequential orthogonalization

Decomposition

Definition

If M and N are subspaces of V that satisfy

- $M \cap N = \{\mathbf{0}\}$

Decomposition

Definition

If M and N are subspaces of V that satisfy

- $M \cap N = \{\mathbf{0}\}$
- $M + N = V$ where $M + N = \{\mathbf{z} \mid \mathbf{z} = \mathbf{x} + \mathbf{y},\ \mathbf{x} \in M, \mathbf{y} \in N\}$ $(M \oplus N = V)$ Direct Sum

Decomposition

Definition

If M and N are subspaces of V that satisfy

- $M \cap N = \{\mathbf{0}\}$
- $M + N = V$ where $M + N = \{\mathbf{z} \mid \mathbf{z} = \mathbf{x} + \mathbf{y},\ \mathbf{x} \in M, \mathbf{y} \in N\}$
  $(M \oplus N = V)$ Direct Sum

then M and N are complementary spaces.

Decomposition

Definition

If M and N are subspaces of V that satisfy

- $M \cap N = \{\mathbf{0}\}$
- $M + N = V$ where $M + N = \{\mathbf{z} \mid \mathbf{z} = \mathbf{x} + \mathbf{y},\ \mathbf{x} \in M, \mathbf{y} \in N\}$
  $(M \oplus N = V)$ Direct Sum

then M and N are complementary spaces. N is the complement of M

Decomposition

Definition

If M and N are subspaces of V that satisfy

- $M \cap N = \{\mathbf{0}\}$
- $M + N = V$ where $M + N = \{\mathbf{z} \mid \mathbf{z} = \mathbf{x} + \mathbf{y},\ \mathbf{x} \in M, \mathbf{y} \in N\}$ $(M \oplus N = V)$ Direct Sum

then M and N are complementary spaces. N is the complement of M

Definition

If M and N are complementary spaces $M + N = \mathbb{R}^n$ and M and N are orthogonal subspaces, then M is the orthogonal complement of N, $N^{\perp}$.

Decomposition

Definition

If M and N are subspaces of V that satisfy

- $M \cap N = \{\mathbf{0}\}$
- $M + N = V$ where $M + N = \{\mathbf{z} \mid \mathbf{z} = \mathbf{x} + \mathbf{y},\ \mathbf{x} \in M, \mathbf{y} \in N\}$ $(M \oplus N = V)$ Direct Sum

then M and N are complementary spaces. N is the complement of M

Definition

If M and N are complementary spaces $M + N = \mathbb{R}^n$ and M and N are orthogonal subspaces, then M is the orthogonal complement of N, $N^{\perp}$.

If $\mathbf{z} \in \mathbb{R}^n = N + M$, then we can uniquely decompose it into a part $\mathbf{x} \in M$ and $\mathbf{y} \in N$ and $r(M) + r(N) = n$

Definition

P is an orthogonal projection operator onto $C(\mathbf{X})$ if and only if

Projections

Definition

P is an orthogonal projection operator onto $C(\mathbf{X})$ if and only if

- $\mathbf{u} \in C(\mathbf{X})$ implies $P\mathbf{u} = \mathbf{u}$ (projection)

Definition

P is an orthogonal projection operator onto $C(\mathbf{X})$ if and only if

- $\mathbf{u} \in C(\mathbf{X})$ implies $P\mathbf{u} = \mathbf{u}$ (projection)
- $\mathbf{v} \perp C(\mathbf{X})$ implies $P\mathbf{v} = \mathbf{0}$ (perpendicular)

Projections

Definition

P is an orthogonal projection operator onto $C(\mathbf{X})$ if and only if

- $\mathbf{u} \in C(\mathbf{X})$ implies $P\mathbf{u} = \mathbf{u}$ (projection)
- $\mathbf{v} \perp C(\mathbf{X})$ implies $P\mathbf{v} = \mathbf{0}$ (perpendicular)

- $C(\mathbf{X})$ and $C(\mathbf{X})^{\perp}$ are complementary spaces

Projections

Definition

P is an orthogonal projection operator onto $C(\mathbf{X})$ if and only if

- $\mathbf{u} \in C(\mathbf{X})$ implies $\mathrm{P}\mathbf{u} = \mathbf{u}$ (projection)
- $\mathbf{v} \perp C(\mathbf{X})$ implies $\mathrm{P}\mathbf{v} = \mathbf{0}$ (perpendicular)

- $C(\mathbf{X})$ and $C(\mathbf{X})^{\perp}$ are complementary spaces
- $\mathbf{y} = \mathbf{u} + \mathbf{v}$ for $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{u} \in C(\mathbf{X})$ and $\mathbf{v} \perp C(\mathbf{X})$ then $\mathrm{P}\mathbf{y} = \mathrm{P}\mathbf{u} + \mathrm{P}\mathbf{v} = \mathbf{u}$

Definition

P is an orthogonal projection operator onto $C(\mathbf{X})$ if and only if

- $\mathbf{u} \in C(\mathbf{X})$ implies $P\mathbf{u} = \mathbf{u}$ (projection)
- $\mathbf{v} \perp C(\mathbf{X})$ implies $P\mathbf{v} = \mathbf{0}$ (perpendicular)

- $C(\mathbf{X})$ and $C(\mathbf{X})^{\perp}$ are complementary spaces
- $\mathbf{y} = \mathbf{u} + \mathbf{v}$ for $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{u} \in C(\mathbf{X})$ and $\mathbf{v} \perp C(\mathbf{X})$ then $P\mathbf{y} = P\mathbf{u} + P\mathbf{v} = \mathbf{u}$
- $\mathbf{u}$ is the orthogonal projection of $\mathbf{y}$ onto $C(\mathbf{X})$

More on Projections

Prop

P is an orthogonal projection operator on $C(\mathbf{X})$ then $C(\mathbf{X}) = C(\mathrm{P})$

More on Projections

Prop

P is an orthogonal projection operator on $C(\mathbf{X})$ then $C(\mathbf{X}) = C(\mathrm{P})$

Theorem

P is an orthogonal projection operator on $C(P)$ if and only if

Prop

P is an orthogonal projection operator on $C(\mathbf{X})$ then $C(\mathbf{X}) = C(\mathrm{P})$

Theorem

P is an orthogonal projection operator on $C(P)$ if and only if

- $P = P^2$ *(idempotent)*

More on Projections

Prop

P is an orthogonal projection operator on $C(\mathbf{X})$ then $C(\mathbf{X}) = C(\mathrm{P})$

Theorem

P is an orthogonal projection operator on $C(P)$ if and only if

- $P = P^2$ *(idempotent)*
- $P = P^T$ *(symmetry)*

More on Projections

Prop

P is an orthogonal projection operator on $C(\mathbf{X})$ then $C(\mathbf{X}) = C(\mathrm{P})$

Theorem

P is an orthogonal projection operator on $C(P)$ if and only if

- $P = P^2$ *(idempotent)*
- $P = P^T$ *(symmetry)*

Claim: If $\mathbf{X}$ is $n \times p$ and $r(\mathbf{X}) = p$ then, $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ operator onto $C(\mathbf{X})$

Projections

Claim: If $\mathbf{X}$ is $n \times p$ and $r(\mathbf{X}) = p$ then, $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is an orthogonal projection operator onto $C(\mathbf{X})$

Projections

Claim: If $\mathbf{X}$ is $n \times p$ and $r(\mathbf{X}) = p$ then, $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is an orthogonal projection operator onto $C(\mathbf{X})$

- $\mathrm{P} = \mathrm{P}^2$ (idempotent)

Projections

Claim: If $\mathbf{X}$ is $n \times p$ and $r(\mathbf{X}) = p$ then, $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is an orthogonal projection operator onto $C(\mathbf{X})$

- $\mathrm{P} = \mathrm{P}^2$ (idempotent)

$$\mathrm{P}_X^2 = \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$

Claim: If $\mathbf{X}$ is $n \times p$ and $r(\mathbf{X}) = p$ then, $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is an orthogonal projection operator onto $C(\mathbf{X})$

- $\mathrm{P} = \mathrm{P}^2$ (idempotent)

$$\begin{aligned} \mathrm{P}_X^2 = \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\ &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \end{aligned}$$

Projections

Claim: If $\mathbf{X}$ is $n \times p$ and $r(\mathbf{X}) = p$ then, $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is an orthogonal projection operator onto $C(\mathbf{X})$

- $\mathrm{P} = \mathrm{P}^2$ (idempotent)

$$\begin{aligned} \mathrm{P}_X^2 = \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\ &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\ &= \mathrm{P}_{\mathbf{X}} \end{aligned}$$

Projections

Claim: If $\mathbf{X}$ is $n \times p$ and $r(\mathbf{X}) = p$ then, $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is an orthogonal projection operator onto $C(\mathbf{X})$

- $\mathrm{P} = \mathrm{P}^2$ (idempotent)

$$\begin{aligned}
\mathrm{P}_X^2 = \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\
&= \mathrm{P}_{\mathbf{X}}
\end{aligned}$$

- $\mathrm{P} = \mathrm{P}^T$ (symmetry)

Projections

Claim: If $\mathbf{X}$ is $n \times p$ and $r(\mathbf{X}) = p$ then, $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is an orthogonal projection operator onto $C(\mathbf{X})$

- $\mathrm{P} = \mathrm{P}^2$ (idempotent)

$$
\begin{aligned}
\mathrm{P}_X^2 = \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\
&= \mathrm{P}_{\mathbf{X}}
\end{aligned}
$$

- $\mathrm{P} = \mathrm{P}^T$ (symmetry)

$$
\mathrm{P}_{\mathbf{X}}^T = (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T
$$

Claim: If $\mathbf{X}$ is $n \times p$ and $r(\mathbf{X}) = p$ then, $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is an orthogonal projection operator onto $C(\mathbf{X})$

- $\mathrm{P} = \mathrm{P}^2$ (idempotent)

$$\begin{aligned}
\mathrm{P}_X^2 = \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\
&= \mathrm{P}_{\mathbf{X}}
\end{aligned}$$

- $\mathrm{P} = \mathrm{P}^T$ (symmetry)

$$\begin{aligned}
\mathrm{P}_{\mathbf{X}}^T &= (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T \\
&= (\mathbf{X}^T)^T((\mathbf{X}^T\mathbf{X})^{-1})^T(\mathbf{X})^T
\end{aligned}$$

Projections

Claim: If $\mathbf{X}$ is $n \times p$ and $r(\mathbf{X}) = p$ then, $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is an orthogonal projection operator onto $C(\mathbf{X})$

- $\mathrm{P} = \mathrm{P}^2$ (idempotent)

$$\begin{aligned} \mathrm{P}_X^2 = \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\ &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\ &= \mathrm{P}_{\mathbf{X}} \end{aligned}$$

- $\mathrm{P} = \mathrm{P}^T$ (symmetry)

$$\begin{aligned} \mathrm{P}_{\mathbf{X}}^T &= (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T \\ &= (\mathbf{X}^T)^T((\mathbf{X}^T\mathbf{X})^{-1})^T(\mathbf{X})^T \\ &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \end{aligned}$$

Projections

Claim: If $\mathbf{X}$ is $n \times p$ and $r(\mathbf{X}) = p$ then, $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is an orthogonal projection operator onto $C(\mathbf{X})$

- $\mathrm{P} = \mathrm{P}^2$ (idempotent)

$$\begin{aligned} \mathrm{P}_X^2 = \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\ &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\ &= \mathrm{P}_{\mathbf{X}} \end{aligned}$$

- $\mathrm{P} = \mathrm{P}^T$ (symmetry)

$$\begin{aligned} \mathrm{P}_{\mathbf{X}}^T &= (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T \\ &= (\mathbf{X}^T)^T((\mathbf{X}^T\mathbf{X})^{-1})^T(\mathbf{X})^T \\ &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\ &= \mathrm{P}_{\mathbf{X}} \end{aligned}$$

Projections

Claim: If $\mathbf{X}$ is $n \times p$ and $r(\mathbf{X}) = p$ then, $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is an orthogonal projection operator onto $C(\mathbf{X})$

- $\mathrm{P} = \mathrm{P}^2$ (idempotent)

$$
\begin{aligned}
\mathrm{P}_X^2 = \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\
&= \mathrm{P}_{\mathbf{X}}
\end{aligned}
$$

- $\mathrm{P} = \mathrm{P}^T$ (symmetry)

$$
\begin{aligned}
\mathrm{P}_{\mathbf{X}}^T &= (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T \\
&= (\mathbf{X}^T)^T((\mathbf{X}^T\mathbf{X})^{-1})^T(\mathbf{X})^T \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\
&= \mathrm{P}_{\mathbf{X}}
\end{aligned}
$$

- $C(X) = C(\mathrm{P}_X)$

Projections

Claim: $\mathbf{I} - \mathrm{P}_{\mathbf{X}}$ is an orthogonal projection onto $C(\mathbf{X})^{\perp}$

Claim: $\mathbf{I} - \mathrm{P}_{\mathbf{X}}$ is an orthogonal projection onto $C(\mathbf{X})^{\perp}$

- idempotent

Projections

Claim: $\mathbf{I} - \mathrm{P}_{\mathbf{X}}$ is an orthogonal projection onto $C(\mathbf{X})^{\perp}$

- idempotent

$$(\mathbf{I} - \mathrm{P}_{\mathbf{X}})^2 \quad = \quad (\mathbf{I} - \mathrm{P}_{\mathbf{X}})(\mathbf{I} - \mathrm{P}_{\mathbf{X}})$$

Projections

Claim: $\mathbf{I} - \mathrm{P}_{\mathbf{X}}$ is an orthogonal projection onto $C(\mathbf{X})^{\perp}$

- idempotent

$$
\begin{aligned}
(\mathbf{I} - \mathrm{P}_{\mathbf{X}})^2 &= (\mathbf{I} - \mathrm{P}_{\mathbf{X}})(\mathbf{I} - \mathrm{P}_{\mathbf{X}}) \\
&= \mathbf{I} - \mathrm{P}_{\mathbf{X}} - \mathrm{P}_{\mathbf{X}} + \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}}
\end{aligned}
$$

Projections

Claim: $\mathbf{I} - \mathrm{P}_{\mathbf{X}}$ is an orthogonal projection onto $C(\mathbf{X})^{\perp}$

- idempotent

$$\begin{aligned}
(\mathbf{I} - \mathrm{P}_{\mathbf{X}})^2 &= (\mathbf{I} - \mathrm{P}_{\mathbf{X}})(\mathbf{I} - \mathrm{P}_{\mathbf{X}}) \\
&= \mathbf{I} - \mathrm{P}_{\mathbf{X}} - \mathrm{P}_{\mathbf{X}} + \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} \\
&= \mathbf{I} - \mathrm{P}_{\mathbf{X}} - \mathrm{P}_{\mathbf{X}} + \mathrm{P}_{\mathbf{X}}
\end{aligned}$$

Projections

Claim: $\mathbf{I} - P_{\mathbf{X}}$ is an orthogonal projection onto $C(\mathbf{X})^{\perp}$

- idempotent

$$
\begin{aligned}
(\mathbf{I} - P_{\mathbf{X}})^2 &= (\mathbf{I} - P_{\mathbf{X}})(\mathbf{I} - P_{\mathbf{X}}) \\
&= \mathbf{I} - P_{\mathbf{X}} - P_{\mathbf{X}} + P_{\mathbf{X}}P_{\mathbf{X}} \\
&= \mathbf{I} - P_{\mathbf{X}} - P_{\mathbf{X}} + P_{\mathbf{X}} \\
&= \mathbf{I} - P_{\mathbf{X}}
\end{aligned}
$$

Projections

Claim: $\mathbf{I} - \mathrm{P}_{\mathbf{X}}$ is an orthogonal projection onto $C(\mathbf{X})^{\perp}$

- idempotent

$$
\begin{aligned}
(\mathbf{I} - \mathrm{P}_{\mathbf{X}})^2 &= (\mathbf{I} - \mathrm{P}_{\mathbf{X}})(\mathbf{I} - \mathrm{P}_{\mathbf{X}}) \\
&= \mathbf{I} - \mathrm{P}_{\mathbf{X}} - \mathrm{P}_{\mathbf{X}} + \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} \\
&= \mathbf{I} - \mathrm{P}_{\mathbf{X}} - \mathrm{P}_{\mathbf{X}} + \mathrm{P}_{\mathbf{X}} \\
&= \mathbf{I} - \mathrm{P}_{\mathbf{X}}
\end{aligned}
$$

- Symmetry $\mathbf{I} - \mathrm{P}_{\mathbf{X}} = (\mathbf{I} - \mathrm{P}_{\mathbf{X}})^T$

Projections

Claim: $\mathbf{I} - \mathrm{P}_{\mathbf{X}}$ is an orthogonal projection onto $C(\mathbf{X})^{\perp}$

- idempotent

$$
\begin{aligned}
(\mathbf{I} - \mathrm{P}_{\mathbf{X}})^2 &= (\mathbf{I} - \mathrm{P}_{\mathbf{X}})(\mathbf{I} - \mathrm{P}_{\mathbf{X}}) \\
&= \mathbf{I} - \mathrm{P}_{\mathbf{X}} - \mathrm{P}_{\mathbf{X}} + \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} \\
&= \mathbf{I} - \mathrm{P}_{\mathbf{X}} - \mathrm{P}_{\mathbf{X}} + \mathrm{P}_{\mathbf{X}} \\
&= \mathbf{I} - \mathrm{P}_{\mathbf{X}}
\end{aligned}
$$

- Symmetry $\mathbf{I} - \mathrm{P}_{\mathbf{X}} = (\mathbf{I} - \mathrm{P}_{\mathbf{X}})^T$
- $\mathbf{u} \in C(\mathbf{X})^{\perp} \Rightarrow \mathbf{u} \perp C(\mathbf{X})$ and $(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{u} = \mathbf{u}$ (projection)

Projections

Claim: $\mathbf{I} - \mathrm{P}_{\mathbf{X}}$ is an orthogonal projection onto $C(\mathbf{X})^{\perp}$

- idempotent

$$\begin{aligned}
(\mathbf{I} - \mathrm{P}_{\mathbf{X}})^2 &= (\mathbf{I} - \mathrm{P}_{\mathbf{X}})(\mathbf{I} - \mathrm{P}_{\mathbf{X}}) \\
&= \mathbf{I} - \mathrm{P}_{\mathbf{X}} - \mathrm{P}_{\mathbf{X}} + \mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} \\
&= \mathbf{I} - \mathrm{P}_{\mathbf{X}} - \mathrm{P}_{\mathbf{X}} + \mathrm{P}_{\mathbf{X}} \\
&= \mathbf{I} - \mathrm{P}_{\mathbf{X}}
\end{aligned}$$

- Symmetry $\mathbf{I} - \mathrm{P}_{\mathbf{X}} = (\mathbf{I} - \mathrm{P}_{\mathbf{X}})^T$
- $\mathbf{u} \in C(\mathbf{X})^{\perp} \Rightarrow \mathbf{u} \perp C(\mathbf{X})$ and $(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{u} = \mathbf{u}$ (projection)
- if $\mathbf{v} \in C(\mathbf{X})$, $(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{v} = \mathbf{v} - \mathbf{v} = 0$

Null Space

- Column space $C(\mathbf{X})$

Null Space

- Column space $C(\mathbf{X})$
- $N(\mathbf{A})$: Null space of $\mathbf{A}$ is $\{\mathbf{u} \mid \mathbf{A}\mathbf{u} = 0\}$

Null Space

- Column space $C(\mathbf{X})$
- $N(\mathbf{A})$: Null space of $\mathbf{A}$ is $\{\mathbf{u} \mid \mathbf{A}\mathbf{u} = 0\}$
- Null space of $\mathbf{X}^T$ is $(\mathbf{u} \mid \mathbf{X}^T\mathbf{u} = 0\}$

Null Space

- Column space $C(\mathbf{X})$
- $N(\mathbf{A})$: Null space of $\mathbf{A}$ is $\{\mathbf{u} \mid \mathbf{A}\mathbf{u} = 0\}$
- Null space of $\mathbf{X}^T$ is $(\mathbf{u} \mid \mathbf{X}^T\mathbf{u} = 0\}$
- $N(\mathbf{X}^T) = C(\mathbf{X})^\perp$ the orthogonal complement of $C(\mathbf{X})$

Null Space

- Column space $C(\mathbf{X})$
- $N(\mathbf{A})$: Null space of $\mathbf{A}$ is $\{\mathbf{u} \mid \mathbf{A}\mathbf{u} = 0\}$
- Null space of $\mathbf{X}^T$ is $(\mathbf{u} \mid \mathbf{X}^T\mathbf{u} = 0\}$
- $N(\mathbf{X}^T) = C(\mathbf{X})^\perp$ the orthogonal complement of $C(\mathbf{X})$
- $N(\mathrm{P}_{\mathbf{X}}) = C(\mathbf{X})^\perp$

Null Space

- Column space $C(\mathbf{X})$
- $N(\mathbf{A})$: Null space of $\mathbf{A}$ is $\{\mathbf{u} \mid \mathbf{A}\mathbf{u} = 0\}$
- Null space of $\mathbf{X}^T$ is $(\mathbf{u} \mid \mathbf{X}^T\mathbf{u} = 0\}$
- $N(\mathbf{X}^T) = C(\mathbf{X})^{\perp}$ the orthogonal complement of $C(\mathbf{X})$
- $N(\mathrm{P}_{\mathbf{X}}) = C(\mathbf{X})^{\perp}$

$$\mathbf{u} \in N(\mathrm{P}) \quad \Rightarrow \quad \mathrm{P}_{\mathbf{X}}\mathbf{u} = \mathbf{0}$$

Null Space

- Column space $C(\mathbf{X})$
- $N(\mathbf{A})$: Null space of $\mathbf{A}$ is $\{\mathbf{u} \mid \mathbf{A}\mathbf{u} = 0\}$
- Null space of $\mathbf{X}^T$ is $(\mathbf{u} \mid \mathbf{X}^T\mathbf{u} = 0\}$
- $N(\mathbf{X}^T) = C(\mathbf{X})^\perp$ the orthogonal complement of $C(\mathbf{X})$
- $N(\mathrm{P}_\mathbf{X}) = C(\mathbf{X})^\perp$

$$
\begin{aligned}
\mathbf{u} \in N(\mathrm{P}) \quad &\Rightarrow \quad \mathrm{P}_\mathbf{X}\mathbf{u} = \mathbf{0} \\
&\Leftrightarrow \quad \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u} = \mathbf{0}
\end{aligned}
$$

Null Space

- Column space $C(\mathbf{X})$
- $N(\mathbf{A})$: Null space of $\mathbf{A}$ is $\{\mathbf{u} \mid \mathbf{A}\mathbf{u} = 0\}$
- Null space of $\mathbf{X}^T$ is $(\mathbf{u} \mid \mathbf{X}^T\mathbf{u} = 0\}$
- $N(\mathbf{X}^T) = C(\mathbf{X})^\perp$ the orthogonal complement of $C(\mathbf{X})$
- $N(\mathrm{P}_\mathbf{X}) = C(\mathbf{X})^\perp$

$$
\begin{aligned}
\mathbf{u} \in N(\mathrm{P}) \quad & \Rightarrow \quad \mathrm{P}_\mathbf{X}\mathbf{u} = \mathbf{0} \\
& \Leftrightarrow \quad \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u} = \mathbf{0} \\
& \Leftrightarrow \quad \mathbf{u}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{0}^T
\end{aligned}
$$

Null Space

- Column space $C(\mathbf{X})$
- $N(\mathbf{A})$: Null space of $\mathbf{A}$ is $\{\mathbf{u} \mid \mathbf{A}\mathbf{u} = 0\}$
- Null space of $\mathbf{X}^T$ is $(\mathbf{u} \mid \mathbf{X}^T\mathbf{u} = 0\}$
- $N(\mathbf{X}^T) = C(\mathbf{X})^\perp$ the orthogonal complement of $C(\mathbf{X})$
- $N(\mathrm{P}_\mathbf{X}) = C(\mathbf{X})^\perp$

$$
\begin{aligned}
\mathbf{u} \in N(\mathrm{P}) \quad & \Rightarrow \quad \mathrm{P}_\mathbf{X}\mathbf{u} = \mathbf{0} \\
& \Leftrightarrow \quad \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u} = \mathbf{0} \\
& \Leftrightarrow \quad \mathbf{u}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{0}^T \\
& \Leftrightarrow \quad \mathbf{u}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\beta = 0
\end{aligned}
$$

- Column space $C(\mathbf{X})$
- $N(\mathbf{A})$: Null space of $\mathbf{A}$ is $\{\mathbf{u} \mid \mathbf{A}\mathbf{u} = 0\}$
- Null space of $\mathbf{X}^T$ is $(\mathbf{u} \mid \mathbf{X}^T\mathbf{u} = 0\}$
- $N(\mathbf{X}^T) = C(\mathbf{X})^{\perp}$ the orthogonal complement of $C(\mathbf{X})$
- $N(\mathrm{P}_{\mathbf{X}}) = C(\mathbf{X})^{\perp}$

$$
\begin{aligned}
\mathbf{u} \in N(\mathrm{P}) \quad & \Rightarrow \quad \mathrm{P}_{\mathbf{X}}\mathbf{u} = \mathbf{0} \\
& \Leftrightarrow \quad \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u} = \mathbf{0} \\
& \Leftrightarrow \quad \mathbf{u}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{0}^T \\
& \Leftrightarrow \quad \mathbf{u}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\beta = 0 \\
& \Leftrightarrow \quad \mathbf{u}^T\mathbf{v} = 0 \quad \mathbf{v} \in C(\mathbf{X})
\end{aligned}
$$

Models Again

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X})$

Models Again

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X})$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathrm{P}_{\mathbf{X}}\mathbf{Y}$

Models Again

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X})$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathrm{P}_{\mathbf{X}}\mathbf{Y}$
- Log Likelihood:

Models Again

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X})$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathrm{P}_{\mathbf{X}}\mathbf{Y}$
- Log Likelihood:

$$\log \mathcal{L}(\boldsymbol{\mu}, \sigma^2) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}$$

Models Again

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X})$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathrm{P}_{\mathbf{X}}\mathbf{Y}$
- Log Likelihood:

$$\log \mathcal{L}(\boldsymbol{\mu}, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}$$

- Decompose $\mathbf{Y} = \mathrm{P}_{\mathbf{X}}\mathbf{Y} + (\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}$

Models Again

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X})$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathrm{P}_{\mathbf{X}}\mathbf{Y}$
- Log Likelihood:

$$\log \mathcal{L}(\boldsymbol{\mu}, \sigma^2) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}$$

- Decompose $\mathbf{Y} = \mathrm{P}_{\mathbf{X}}\mathbf{Y} + (\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}$
- Use $\mathrm{P}_{\mathbf{X}}\boldsymbol{\mu} = \boldsymbol{\mu}$

Models Again

- $\mathbf{Y} \sim \mathsf{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X})$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathsf{P}_{\mathbf{X}}\mathbf{Y}$
- Log Likelihood:

$$\log \mathcal{L}(\boldsymbol{\mu}, \sigma^2) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}$$

- Decompose $\mathbf{Y} = \mathsf{P}_{\mathbf{X}}\mathbf{Y} + (\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}$
- Use $\mathsf{P}_{\mathbf{X}}\boldsymbol{\mu} = \boldsymbol{\mu}$
- and Simplify

$$\|\mathbf{Y} - \boldsymbol{\mu}\|^2$$