# Maximum Likelihood Estimation

Merlise Clyde

STA721 Linear Models

Duke University

September 2, 2014

# Outline

Topics

- Projections
- Maximum Likelihood Estimates
- Spectral Decomposition

Readings: Continue reading Wakefield 5.6.1 or for more details Christensen Chapter 1-2, Appendix A, and Appendix B

# Models & MLEs

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

# Models & MLEs

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

# Models & MLEs

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X}) \Leftrightarrow \boldsymbol{\mu} = \mathbf{X}\beta$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathrm{P}_{\mathbf{X}}\mathbf{Y}$
- $\mathrm{P}_{\mathbf{X}}$ is the orthogonal projection operator on the column space of $\mathbf{X}$; e.g. $\mathbf{X}$ full rank $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$
- $\mathrm{P} = \mathrm{P}^2$ (idempotent)

# Models & MLEs

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X}) \Leftrightarrow \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathrm{P}_{\mathbf{X}}\mathbf{Y}$
- $\mathrm{P}_{\mathbf{X}}$ is the orthogonal projection operator on the column space of $\mathbf{X}$; e.g. $\mathbf{X}$ full rank $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$
- $\mathrm{P} = \mathrm{P}^2$ (idempotent)

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

# Models & MLEs

- $\mathbf{Y} \sim \mathsf{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X}) \Leftrightarrow \boldsymbol{\mu} = \mathbf{X}\beta$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathsf{P}_{\mathbf{X}}\mathbf{Y}$
- $\mathsf{P}_{\mathbf{X}}$ is the orthogonal projection operator on the column space of $\mathbf{X}$; e.g. $\mathbf{X}$ full rank $\mathsf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$
- $\mathsf{P} = \mathsf{P}^2$ (idempotent)

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

# Models & MLEs

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X}) \Leftrightarrow \boldsymbol{\mu} = \mathbf{X}\beta$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathrm{P}_{\mathbf{X}}\mathbf{Y}$
- $\mathrm{P}_{\mathbf{X}}$ is the orthogonal projection operator on the column space of $\mathbf{X}$; e.g. $\mathbf{X}$ full rank $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$
- $\mathrm{P} = \mathrm{P}^2$ (idempotent)

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

# Models & MLEs

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X}) \Leftrightarrow \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathrm{P}_{\mathbf{X}}\mathbf{Y}$
- $\mathrm{P}_{\mathbf{X}}$ is the orthogonal projection operator on the column space of $\mathbf{X}$; e.g. $\mathbf{X}$ full rank $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$
- $\mathrm{P} = \mathrm{P}^2$ (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)

# Models & MLEs

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X}) \Leftrightarrow \boldsymbol{\mu} = \mathbf{X}\beta$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathrm{P}_{\mathbf{X}}\mathbf{Y}$
- $\mathrm{P}_{\mathbf{X}}$ is the orthogonal projection operator on the column space of $\mathbf{X}$; e.g. $\mathbf{X}$ full rank $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$
- $\mathrm{P} = \mathrm{P}^2$ (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)

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

# Models & MLEs

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X}) \Leftrightarrow \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathrm{P}_{\mathbf{X}}\mathbf{Y}$
- $\mathrm{P}_{\mathbf{X}}$ is the orthogonal projection operator on the column space of $\mathbf{X}$; e.g. $\mathbf{X}$ full rank $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$
- $\mathrm{P} = \mathrm{P}^2$ (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)

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

# Models & MLEs

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X}) \Leftrightarrow \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathrm{P}_{\mathbf{X}}\mathbf{Y}$
- $\mathrm{P}_{\mathbf{X}}$ is the orthogonal projection operator on the column space of $\mathbf{X}$; e.g. $\mathbf{X}$ full rank $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$
- $\mathrm{P} = \mathrm{P}^2$ (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)

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

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X}) \Leftrightarrow \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathrm{P}_{\mathbf{X}}\mathbf{Y}$
- $\mathrm{P}_{\mathbf{X}}$ is the orthogonal projection operator on the column space of $\mathbf{X}$; e.g. $\mathbf{X}$ full rank $\mathrm{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$
- $\mathrm{P} = \mathrm{P}^2$ (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)

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

# Models & MLEs

- $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} \in C(\mathbf{X}) \Leftrightarrow \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$
- Claim: Maximum Likelihood Estimator (MLE) of $\boldsymbol{\mu}$ is $\mathsf{P}_{\mathbf{X}}\mathbf{Y}$
- $\mathsf{P}_{\mathbf{X}}$ is the orthogonal projection operator on the column space of $\mathbf{X}$; e.g. $\mathbf{X}$ full rank $\mathsf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$
- $\mathsf{P} = \mathsf{P}^2$ (idempotent)

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

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

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

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

# Projections

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

# Projections

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

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

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

- idempotent

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

# Projections

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

- idempotent

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

# 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 $\mathbf{v} \in C(\mathbf{X})$, $(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{v} = \mathbf{v} - \mathbf{v} = 0$

# Log Likelihood

- Log Likelihood:

# Log Likelihood

- Log Likelihood:

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

# Log Likelihood

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

# Log Likelihood

- 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 $\mathsf{P}_{\mathbf{X}}\boldsymbol{\mu} = \boldsymbol{\mu}$

# Log Likelihood

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

# Expand

$$\|\mathbf{Y} - \boldsymbol{\mu}\|^2 = \|\mathbf{Y} - \mathsf{P}_{\mathbf{X}}\mathbf{Y} + \mathsf{P}_x\mathbf{Y} - \boldsymbol{\mu}\|^2$$

# Expand

$$
\begin{aligned}
\|\mathbf{Y} - \boldsymbol{\mu}\|^2 &= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_x\mathbf{Y} - \boldsymbol{\mu}\|^2 \\
&= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_x\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\boldsymbol{\mu}\|^2
\end{aligned}
$$

# Expand

$$
\begin{aligned}
\|\mathbf{Y} - \boldsymbol{\mu}\|^2 &= \|\mathbf{Y} - \mathsf{P}_{\mathbf{X}}\mathbf{Y} + \mathsf{P}_{x}\mathbf{Y} - \boldsymbol{\mu}\|^2 \\
&= \|\mathbf{Y} - \mathsf{P}_{\mathbf{X}}\mathbf{Y} + \mathsf{P}_{x}\mathbf{Y} - \mathsf{P}_{\mathbf{X}}\boldsymbol{\mu}\|^2 \\
&= \|(\mathbf{I} - \mathsf{P}_{\mathbf{x}})\mathbf{Y} + \mathsf{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2
\end{aligned}
$$

# Expand

$$
\begin{aligned}
\|\mathbf{Y} - \boldsymbol{\mu}\|^2 &= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_x\mathbf{Y} - \boldsymbol{\mu}\|^2 \\
&= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_x\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\boldsymbol{\mu}\|^2 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y} + \mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2
\end{aligned}
$$

# Expand

$$
\begin{aligned}
\|\mathbf{Y} - \boldsymbol{\mu}\|^2 &= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_x\mathbf{Y} - \boldsymbol{\mu}\|^2 \\
&= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_x\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\boldsymbol{\mu}\|^2 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y} + \mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2 + {\color{red}\|\mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2}
\end{aligned}
$$

# Expand

$$
\begin{aligned}
\|\mathbf{Y} - \boldsymbol{\mu}\|^2 &= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_{x}\mathbf{Y} - \boldsymbol{\mu}\|^2 \\
&= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_{x}\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\boldsymbol{\mu}\|^2 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y} + \mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2 + \|\mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 + 2(\mathbf{Y} - \boldsymbol{\mu})^T \mathrm{P}_{\mathbf{X}}^T (\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}
\end{aligned}
$$

# Expand

$$
\begin{aligned}
\|\mathbf{Y} - \boldsymbol{\mu}\|^2 &= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_{x}\mathbf{Y} - \boldsymbol{\mu}\|^2 \\
&= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_{x}\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\boldsymbol{\mu}\|^2 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y} + \mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2 + \|\mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 + 2(\mathbf{Y} - \boldsymbol{\mu})^T \mathrm{P}_{\mathbf{X}}^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y} \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2 + \|\mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 + 0
\end{aligned}
$$

# Expand

$$\begin{aligned}
\|\mathbf{Y}-\boldsymbol{\mu}\|^2 &= \|\mathbf{Y}-\mathrm{P}_{\mathbf{X}}\mathbf{Y}+\mathrm{P}_x\mathbf{Y}-\boldsymbol{\mu}\|^2 \\
&= \|\mathbf{Y}-\mathrm{P}_{\mathbf{X}}\mathbf{Y}+\mathrm{P}_x\mathbf{Y}-\mathrm{P}_{\mathbf{X}}\boldsymbol{\mu}\|^2 \\
&= \|(\mathbf{I}-\mathrm{P}_{\mathrm{x}})\mathbf{Y}+\mathrm{P}_{\mathbf{X}}(\mathbf{Y}-\boldsymbol{\mu})\|^2 \\
&= \|(\mathbf{I}-\mathrm{P}_{\mathrm{x}})\mathbf{Y}\|^2+\|\mathrm{P}_{\mathbf{X}}(\mathbf{Y}-\boldsymbol{\mu})\|^2+2(\mathbf{Y}-\boldsymbol{\mu})^T\mathrm{P}_{\mathbf{X}}^T(\mathbf{I}-\mathrm{P}_{\mathbf{X}})\mathbf{Y} \\
&= \|(\mathbf{I}-\mathrm{P}_{\mathrm{x}})\mathbf{Y}\|^2+\|\mathrm{P}_{\mathbf{X}}(\mathbf{Y}-\boldsymbol{\mu})\|^2+0 \\
&= \|(\mathbf{I}-\mathrm{P}_{\mathrm{x}})\mathbf{Y}\|^2+\|\mathrm{P}_{\mathbf{X}}\mathbf{Y}-\boldsymbol{\mu}\|^2
\end{aligned}$$

# Expand

$$
\begin{aligned}
\|\mathbf{Y} - \boldsymbol{\mu}\|^2 &= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_{x}\mathbf{Y} - \boldsymbol{\mu}\|^2 \\
&= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_{x}\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\boldsymbol{\mu}\|^2 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y} + \mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2 + \|\mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 + 2(\mathbf{Y} - \boldsymbol{\mu})^T \mathrm{P}_{\mathbf{X}}^T (\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y} \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2 + \|\mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 + 0 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2 + \|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2
\end{aligned}
$$

Crossproduct term is zero

# Expand

$$
\begin{aligned}
\|\mathbf{Y} - \boldsymbol{\mu}\|^2 &= \|\mathbf{Y} - P_{\mathbf{X}}\mathbf{Y} + P_{x}\mathbf{Y} - \boldsymbol{\mu}\|^2 \\
&= \|\mathbf{Y} - P_{\mathbf{X}}\mathbf{Y} + P_{x}\mathbf{Y} - P_{\mathbf{X}}\boldsymbol{\mu}\|^2 \\
&= \|(\mathbf{I} - P_{\mathrm{x}})\mathbf{Y} + P_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 \\
&= \|(\mathbf{I} - P_{\mathrm{x}})\mathbf{Y}\|^2 + \|P_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 + 2(\mathbf{Y} - \boldsymbol{\mu})^T P_{\mathbf{X}}^T(\mathbf{I} - P_{\mathbf{X}})\mathbf{Y} \\
&= \|(\mathbf{I} - P_{\mathrm{x}})\mathbf{Y}\|^2 + \|P_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 + 0 \\
&= \|(\mathbf{I} - P_{\mathrm{x}})\mathbf{Y}\|^2 + \|P_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2
\end{aligned}
$$

Crossproduct term is zero

$$
P_{\mathbf{X}}^T(\mathbf{I} - P_{\mathbf{X}}) = P_{\mathbf{X}}(\mathbf{I} - P_{\mathbf{X}})
$$

# Expand

$$
\begin{aligned}
\|\mathbf{Y} - \boldsymbol{\mu}\|^2 &= \|\mathbf{Y} - \mathsf{P}_{\mathbf{X}}\mathbf{Y} + \mathsf{P}_{x}\mathbf{Y} - \boldsymbol{\mu}\|^2 \\
&= \|\mathbf{Y} - \mathsf{P}_{\mathbf{X}}\mathbf{Y} + \mathsf{P}_{x}\mathbf{Y} - \mathsf{P}_{\mathbf{X}}\boldsymbol{\mu}\|^2 \\
&= \|(\mathbf{I} - \mathsf{P}_{\mathbf{x}})\mathbf{Y} + \mathsf{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 \\
&= \|(\mathbf{I} - \mathsf{P}_{\mathbf{x}})\mathbf{Y}\|^2 + \|\mathsf{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 + 2(\mathbf{Y} - \boldsymbol{\mu})^T \mathsf{P}_{\mathbf{X}}^T(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y} \\
&= \|(\mathbf{I} - \mathsf{P}_{\mathbf{x}})\mathbf{Y}\|^2 + \|\mathsf{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 + 0 \\
&= \|(\mathbf{I} - \mathsf{P}_{\mathbf{x}})\mathbf{Y}\|^2 + \|\mathsf{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2
\end{aligned}
$$

Crossproduct term is zero

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

# Expand

$$
\begin{aligned}
\|\mathbf{Y} - \boldsymbol{\mu}\|^2 &= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_{x}\mathbf{Y} - \boldsymbol{\mu}\|^2 \\
&= \|\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\mathbf{Y} + \mathrm{P}_{x}\mathbf{Y} - \mathrm{P}_{\mathbf{X}}\boldsymbol{\mu}\|^2 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y} + \mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2 + \|\mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 + 2(\mathbf{Y} - \boldsymbol{\mu})^T \mathrm{P}_{\mathbf{X}}^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y} \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2 + \|\mathrm{P}_{\mathbf{X}}(\mathbf{Y} - \boldsymbol{\mu})\|^2 + 0 \\
&= \|(\mathbf{I} - \mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2 + \|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2
\end{aligned}
$$

Crossproduct term is zero

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

# Expand

$$
\begin{aligned}
\|\mathbf{Y}-\boldsymbol{\mu}\|^2 &= \|\mathbf{Y}-\mathrm{P}_{\mathbf{X}}\mathbf{Y}+\mathrm{P}_{x}\mathbf{Y}-\boldsymbol{\mu}\|^2 \\
&= \|\mathbf{Y}-\mathrm{P}_{\mathbf{X}}\mathbf{Y}+\mathrm{P}_{x}\mathbf{Y}-\mathrm{P}_{\mathbf{X}}\boldsymbol{\mu}\|^2 \\
&= \|(\mathbf{I}-\mathrm{P}_{\mathbf{x}})\mathbf{Y}+\mathrm{P}_{\mathbf{X}}(\mathbf{Y}-\boldsymbol{\mu})\|^2 \\
&= \|(\mathbf{I}-\mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2+\|\mathrm{P}_{\mathbf{X}}(\mathbf{Y}-\boldsymbol{\mu})\|^2+2(\mathbf{Y}-\boldsymbol{\mu})^T\mathrm{P}_{\mathbf{X}}^T(\mathbf{I}-\mathrm{P}_{\mathbf{X}})\mathbf{Y} \\
&= \|(\mathbf{I}-\mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2+\|\mathrm{P}_{\mathbf{X}}(\mathbf{Y}-\boldsymbol{\mu})\|^2+0 \\
&= \|(\mathbf{I}-\mathrm{P}_{\mathbf{x}})\mathbf{Y}\|^2+\|\mathrm{P}_{\mathbf{X}}\mathbf{Y}-\boldsymbol{\mu}\|^2
\end{aligned}
$$

Crossproduct term is zero

$$
\begin{aligned}
\mathrm{P}_{\mathbf{X}}^T(\mathbf{I}-\mathrm{P}_{\mathbf{X}}) &= \mathrm{P}_{\mathbf{X}}(\mathbf{I}-\mathrm{P}_{\mathbf{X}}) \\
&= \mathrm{P}_{\mathbf{X}}-\mathrm{P}_{\mathbf{X}}\mathrm{P}_{\mathbf{X}} \\
&= \mathrm{P}_{\mathbf{X}}-\mathrm{P}_{\mathbf{X}} \\
&= 0
\end{aligned}
$$

# Likelihood

Substitute decomposition into log likelihood

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

# Likelihood

Substitute decomposition into log likelihood

$$\begin{aligned}\log \mathcal{L}(\boldsymbol{\mu}, \sigma^2) &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2} \\ &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathsf{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}\right)\end{aligned}$$

# Likelihood

Substitute decomposition into log likelihood

$$
\begin{aligned}
\log \mathcal{L}(\boldsymbol{\mu}, \sigma^2) &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2} \\
&= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathsf{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}\right) \\
&= \underbrace{-\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2}} + \underbrace{-\frac{1}{2}\frac{\|\mathsf{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}}
\end{aligned}
$$

# Likelihood

Substitute decomposition into log likelihood

$$
\begin{aligned}
\log \mathcal{L}(\boldsymbol{\mu}, \sigma^2) &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2} \\
&= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathsf{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}\right) \\
&= \underbrace{-\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2}} + \underbrace{-\frac{1}{2}\frac{\|\mathsf{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}} \\
&= \text{constant with respect to } \boldsymbol{\mu}
\end{aligned}
$$

# Likelihood

Substitute decomposition into log likelihood

$$
\begin{aligned}
\log \mathcal{L}(\boldsymbol{\mu}, \sigma^2) &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2} \\
&= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathsf{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}\right) \\
&= \underbrace{-\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2}} + \underbrace{-\frac{1}{2}\frac{\|\mathsf{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}} \\
&= \text{constant with respect to } \boldsymbol{\mu} \qquad \leq 0
\end{aligned}
$$

# Likelihood

Substitute decomposition into log likelihood

$$\begin{aligned}
\log \mathcal{L}(\boldsymbol{\mu}, \sigma^2) &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2} \\
&= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}\right) \\
&= \underbrace{-\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2}} + \underbrace{-\frac{1}{2}\frac{\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}} \\
&= \text{constant with respect to } \boldsymbol{\mu} \qquad\qquad \leq 0
\end{aligned}$$

Maximize with respect to $\boldsymbol{\mu}$ for each $\sigma^2$

# Likelihood

Substitute decomposition into log likelihood

$$\begin{aligned}
\log \mathcal{L}(\boldsymbol{\mu}, \sigma^2) &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2} \\
&= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}\right) \\
&= \underbrace{-\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2}}_{\text{constant with respect to } \boldsymbol{\mu}} + \underbrace{-\frac{1}{2}\frac{\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}}_{\leq 0}
\end{aligned}$$

Maximize with respect to $\boldsymbol{\mu}$ for each $\sigma^2$
RHS is largest when $\boldsymbol{\mu} = \mathrm{P}_{\mathbf{X}}\mathbf{Y}$ for any choice of $\sigma^2$

# Likelihood

Substitute decomposition into log likelihood

$$
\begin{aligned}
\log \mathcal{L}(\boldsymbol{\mu}, \sigma^2) &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2} \\
&= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathsf{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}\right) \\
&= \underbrace{-\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2}}_{\text{constant with respect to } \boldsymbol{\mu}} + \underbrace{-\frac{1}{2}\frac{\|\mathsf{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}}_{\leq 0}
\end{aligned}
$$

Maximize with respect to $\boldsymbol{\mu}$ for each $\sigma^2$

RHS is largest when $\boldsymbol{\mu} = \mathsf{P}_{\mathbf{X}}\mathbf{Y}$ for any choice of $\sigma^2$

$$\therefore \quad \hat{\boldsymbol{\mu}} = \mathsf{P}_{\mathbf{X}}\mathbf{Y}$$

is the MLE of $\mu$

# Likelihood

Substitute decomposition into log likelihood

$$
\begin{aligned}
\log \mathcal{L}(\boldsymbol{\mu}, \sigma^2) &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2} \\
&= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}\right) \\
&= \underbrace{-\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2}}_{} + \underbrace{-\frac{1}{2}\frac{\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}}_{} \\
&= \text{constant with respect to } \boldsymbol{\mu} \qquad \leq 0
\end{aligned}
$$

Maximize with respect to $\boldsymbol{\mu}$ for each $\sigma^2$
RHS is largest when $\boldsymbol{\mu} = \mathrm{P}_{\mathbf{X}}\mathbf{Y}$ for any choice of $\sigma^2$

$$
\therefore \quad \hat{\boldsymbol{\mu}} = \mathrm{P}_{\mathbf{X}}\mathbf{Y}
$$

is the MLE of $\mu$ (yields fitted values $\hat{\mathbf{Y}} = \mathrm{P}_{\mathbf{X}}\mathbf{Y}$)

# MLE of $\beta$

$$\mathcal{L}(\boldsymbol{\mu}, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathsf{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}\right)$$

# MLE of $\beta$

$$\mathcal{L}(\boldsymbol{\mu}, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathsf{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}\right)$$

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathsf{P}_{\mathbf{X}}\mathbf{Y} - \mathbf{X}\beta\|^2}{\sigma^2}\right)$$

# MLE of $\beta$

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\mu}, \sigma^2) &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}\right) \\
\mathcal{L}(\boldsymbol{\beta}, \sigma^2) &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \mathbf{X}\beta\|^2}{\sigma^2}\right)
\end{aligned}$$

Similar argument to show that RHS is maximized by minimizing

$$\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \mathbf{X}\beta\|^2$$

# MLE of $\beta$

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\mu}, \sigma^2) &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}\right) \\
\mathcal{L}(\boldsymbol{\beta}, \sigma^2) &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2}{\sigma^2}\right)
\end{aligned}$$

Similar argument to show that RHS is maximized by minimizing

$$\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

Therefore $\hat{\boldsymbol{\beta}}$ is a MLE of $\boldsymbol{\beta}$ if and only if satisfies

$$\mathrm{P}_{\mathbf{X}}\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

# MLE of $\beta$

$$\mathcal{L}(\boldsymbol{\mu}, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \boldsymbol{\mu}\|^2}{\sigma^2}\right)$$

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} + \frac{\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \mathbf{X}\beta\|^2}{\sigma^2}\right)$$

Similar argument to show that RHS is maximized by minimizing

$$\|\mathrm{P}_{\mathbf{X}}\mathbf{Y} - \mathbf{X}\beta\|^2$$

Therefore $\hat{\beta}$ is a MLE of $\beta$ if and only if satisfies

$$\mathrm{P}_{\mathbf{X}}\mathbf{Y} = \mathbf{X}\hat{\beta}$$

If $\mathbf{X}^T\mathbf{X}$ is full rank, the MLE of $\beta$ is

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \hat{\beta}$$

# MLE of $\sigma^2$

- Plug-in MLE of $\hat{\boldsymbol{\mu}}$ for $\boldsymbol{\mu}$ and differentiate with respect to $\sigma^2$

# MLE of $\sigma^2$

- Plug-in MLE of $\hat{\boldsymbol{\mu}}$ for $\boldsymbol{\mu}$ and differentiate with respect to $\sigma^2$

$$\log \mathcal{L}(\hat{\boldsymbol{\mu}}, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2}$$

# MLE of $\sigma^2$

- Plug-in MLE of $\hat{\boldsymbol{\mu}}$ for $\boldsymbol{\mu}$ and differentiate with respect to $\sigma^2$

$$\begin{aligned}
\log \mathcal{L}(\hat{\boldsymbol{\mu}}, \sigma^2) &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} \\
\frac{\partial \log \mathcal{L}(\hat{\boldsymbol{\mu}}, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2} \|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2 \left(\frac{1}{\sigma^2}\right)^2
\end{aligned}$$

# MLE of $\sigma^2$

- Plug-in MLE of $\hat{\boldsymbol{\mu}}$ for $\boldsymbol{\mu}$ and differentiate with respect to $\sigma^2$

$$\begin{aligned}
\log \mathcal{L}(\hat{\boldsymbol{\mu}}, \sigma^2) &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} \\
\frac{\partial \log \mathcal{L}(\hat{\boldsymbol{\mu}}, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2} \|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2 \left(\frac{1}{\sigma^2}\right)^2
\end{aligned}$$

- Set derivative to zero and solve for MLE

$$0 = -\frac{n}{2} \frac{1}{\hat{\sigma}^2} + \frac{1}{2} \|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2 \left(\frac{1}{\hat{\sigma}^2}\right)^2$$

# MLE of $\sigma^2$

- Plug-in MLE of $\hat{\boldsymbol{\mu}}$ for $\boldsymbol{\mu}$ and differentiate with respect to $\sigma^2$

$$
\begin{aligned}
\log \mathcal{L}(\hat{\boldsymbol{\mu}}, \sigma^2) &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} \\
\frac{\partial \log \mathcal{L}(\hat{\boldsymbol{\mu}}, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2} \|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2 \left( \frac{1}{\sigma^2} \right)^2
\end{aligned}
$$

- Set derivative to zero and solve for MLE

$$
\begin{aligned}
0 &= -\frac{n}{2} \frac{1}{\hat{\sigma}^2} + \frac{1}{2} \|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2 \left( \frac{1}{\hat{\sigma}^2} \right)^2 \\
\frac{n}{2} \hat{\sigma}^2 &= \frac{1}{2} \|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2
\end{aligned}
$$

# MLE of $\sigma^2$

- Plug-in MLE of $\hat{\boldsymbol{\mu}}$ for $\boldsymbol{\mu}$ and differentiate with respect to $\sigma^2$

$$
\begin{aligned}
\log \mathcal{L}(\hat{\boldsymbol{\mu}}, \sigma^2) &= -\frac{n}{2}\log \sigma^2 - \frac{1}{2}\frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{\sigma^2} \\
\frac{\partial \log \mathcal{L}(\hat{\boldsymbol{\mu}}, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2 \left(\frac{1}{\sigma^2}\right)^2
\end{aligned}
$$

- Set derivative to zero and solve for MLE

$$
\begin{aligned}
0 &= -\frac{n}{2}\frac{1}{\hat{\sigma}^2} + \frac{1}{2}\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2 \left(\frac{1}{\hat{\sigma}^2}\right)^2 \\
\frac{n}{2}\hat{\sigma}^2 &= \frac{1}{2}\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2 \\
\hat{\sigma}^2 &= \frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{n}
\end{aligned}
$$

# Estimate of $\sigma^2$

Maximum Likelihood Estimate of $\sigma^2$

$$\hat{\sigma}^2 = \frac{\|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2}{n}$$

# Estimate of $\sigma^2$

Maximum Likelihood Estimate of $\sigma^2$

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{n} \\
&= \frac{\mathbf{Y}^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}}{n}
\end{aligned}
$$

## Estimate of $\sigma^2$

Maximum Likelihood Estimate of $\sigma^2$

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{n} \\
&= \frac{\mathbf{Y}^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}}{n} \\
&= \frac{\mathbf{Y}^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}}{n}
\end{aligned}
$$

# Estimate of $\sigma^2$

Maximum Likelihood Estimate of $\sigma^2$

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{n} \\
&= \frac{\mathbf{Y}^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}}{n} \\
&= \frac{\mathbf{Y}^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}}{n} \\
&= \frac{\mathbf{e}^T\mathbf{e}}{n}
\end{aligned}
$$

Maximum Likelihood Estimate of $\sigma^2$

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{\|(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}\|^2}{n} \\
&= \frac{\mathbf{Y}^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}}{n} \\
&= \frac{\mathbf{Y}^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}}{n} \\
&= \frac{\mathbf{e}^T\mathbf{e}}{n}
\end{aligned}
$$

where $\mathbf{e} = (\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}$ residuals from the regression of $\mathbf{Y}$ on $\mathbf{X}$

# Geometric View

- Fitted Values $\hat{\mathbf{Y}} = P_{\mathbf{X}}\mathbf{Y} = \mathbf{X}\hat{\beta}$

# Geometric View

- Fitted Values $\hat{\mathbf{Y}} = \mathsf{P}_{\mathbf{X}}\mathbf{Y} = \mathbf{X}\hat{\beta}$
- Residuals $\mathbf{e} = (\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}$

# Geometric View

- Fitted Values $\hat{\mathbf{Y}} = \mathrm{P}_{\mathbf{X}}\mathbf{Y} = \mathbf{X}\hat{\beta}$
- Residuals $\mathbf{e} = (\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}$
- $\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{e}$

# Geometric View

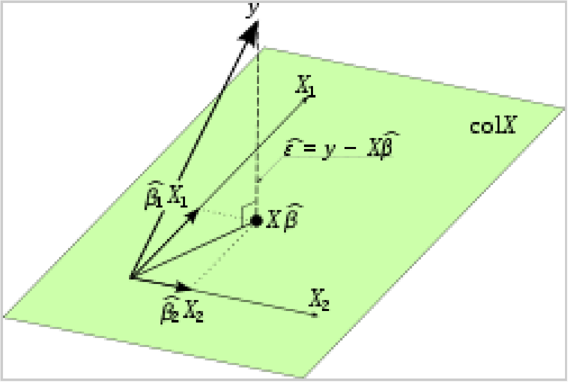
- Fitted Values $\hat{\mathbf{Y}} = \mathrm{P}_{\mathbf{X}}\mathbf{Y} = \mathbf{X}\hat{\beta}$
- Residuals $\mathbf{e} = (\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}$
- $\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{e}$

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

# Geometric View

- Fitted Values $\hat{\mathbf{Y}} = \mathsf{P}_{\mathbf{X}}\mathbf{Y} = \mathbf{X}\hat{\beta}$
- Residuals $\mathbf{e} = (\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}$
- $\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{e}$

$$\|\mathbf{Y}\|^2 = \|(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}\|^2 + \|\mathsf{P}_{\mathbf{X}}\mathbf{Y}\|^2$$



# Properties

$\hat{\mathbf{Y}} = \hat{\boldsymbol{\mu}}$ is an unbiased estimate of $\boldsymbol{\mu} = \mathbf{X}\beta$

# Properties

$\hat{\mathbf{Y}} = \hat{\boldsymbol{\mu}}$ is an unbiased estimate of $\boldsymbol{\mu} = \mathbf{X}\beta$

$$\mathsf{E}[\hat{\mathbf{Y}}] = \mathsf{E}[\mathsf{P}_{\mathbf{X}}\mathbf{Y}]$$

# Properties

$\hat{\mathbf{Y}} = \hat{\boldsymbol{\mu}}$ is an unbiased estimate of $\boldsymbol{\mu} = \mathbf{X}\beta$

$$\begin{aligned} \mathsf{E}[\hat{\mathbf{Y}}] &= \mathsf{E}[\mathsf{P}_{\mathbf{X}}\mathbf{Y}] \\ &= \mathsf{P}_{\mathbf{X}}\mathsf{E}[\mathbf{Y}] \end{aligned}$$

# Properties

$\hat{\mathbf{Y}} = \hat{\boldsymbol{\mu}}$ is an unbiased estimate of $\boldsymbol{\mu} = \mathbf{X}\beta$

$$\begin{aligned} \mathrm{E}[\hat{\mathbf{Y}}] &= \mathrm{E}[\mathrm{P}_{\mathbf{X}}\mathbf{Y}] \\ &= \mathrm{P}_{\mathbf{X}}\mathrm{E}[\mathbf{Y}] \\ &= \mathrm{P}_{\mathbf{X}}\boldsymbol{\mu} \end{aligned}$$

## Properties

$\hat{\mathbf{Y}} = \hat{\boldsymbol{\mu}}$ is an unbiased estimate of $\boldsymbol{\mu} = \mathbf{X}\beta$

$$
\begin{aligned}
\mathsf{E}[\hat{\mathbf{Y}}] &= \mathsf{E}[\mathsf{P}_{\mathbf{X}}\mathbf{Y}] \\
&= \mathsf{P}_{\mathbf{X}}\mathsf{E}[\mathbf{Y}] \\
&= \mathsf{P}_{\mathbf{X}}\boldsymbol{\mu} \\
&= \boldsymbol{\mu}
\end{aligned}
$$

# Properties

$\hat{\mathbf{Y}} = \hat{\boldsymbol{\mu}}$ is an unbiased estimate of $\boldsymbol{\mu} = \mathbf{X}\beta$

$$
\begin{aligned}
\mathsf{E}[\hat{\mathbf{Y}}] &= \mathsf{E}[\mathsf{P}_{\mathbf{X}}\mathbf{Y}] \\
&= \mathsf{P}_{\mathbf{X}}\mathsf{E}[\mathbf{Y}] \\
&= \mathsf{P}_{\mathbf{X}}\boldsymbol{\mu} \\
&= \boldsymbol{\mu}
\end{aligned}
$$

$\mathsf{E}[\mathbf{e}] = \mathbf{0}$ if $\boldsymbol{\mu} \in C(\mathbf{X})$

# Properties

$\hat{\mathbf{Y}} = \hat{\boldsymbol{\mu}}$ is an unbiased estimate of $\boldsymbol{\mu} = \mathbf{X}\beta$

$$\begin{aligned} \mathsf{E}[\hat{\mathbf{Y}}] &= \mathsf{E}[\mathsf{P}_{\mathbf{X}}\mathbf{Y}] \\ &= \mathsf{P}_{\mathbf{X}}\mathsf{E}[\mathbf{Y}] \\ &= \mathsf{P}_{\mathbf{X}}\boldsymbol{\mu} \\ &= \boldsymbol{\mu} \end{aligned}$$

$\mathsf{E}[\mathbf{e}] = \mathbf{0}$ if $\boldsymbol{\mu} \in C(\mathbf{X})$

$$\mathsf{E}[\mathbf{e}] = \mathsf{E}[(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}]$$

# Properties

$\hat{\mathbf{Y}} = \hat{\boldsymbol{\mu}}$ is an unbiased estimate of $\boldsymbol{\mu} = \mathbf{X}\beta$

$$\begin{aligned} \mathsf{E}[\hat{\mathbf{Y}}] &= \mathsf{E}[\mathsf{P}_{\mathbf{X}}\mathbf{Y}] \\ &= \mathsf{P}_{\mathbf{X}}\mathsf{E}[\mathbf{Y}] \\ &= \mathsf{P}_{\mathbf{X}}\boldsymbol{\mu} \\ &= \boldsymbol{\mu} \end{aligned}$$

$\mathsf{E}[\mathbf{e}] = \mathbf{0}$ if $\boldsymbol{\mu} \in C(\mathbf{X})$

$$\begin{aligned} \mathsf{E}[\mathbf{e}] &= \mathsf{E}[(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}] \\ &= (\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathsf{E}[\mathbf{Y}] \end{aligned}$$

# Properties

$\hat{\mathbf{Y}} = \hat{\boldsymbol{\mu}}$ is an unbiased estimate of $\boldsymbol{\mu} = \mathbf{X}\beta$

$$\begin{aligned} \mathsf{E}[\hat{\mathbf{Y}}] &= \mathsf{E}[\mathsf{P}_{\mathbf{X}}\mathbf{Y}] \\ &= \mathsf{P}_{\mathbf{X}}\mathsf{E}[\mathbf{Y}] \\ &= \mathsf{P}_{\mathbf{X}}\boldsymbol{\mu} \\ &= \boldsymbol{\mu} \end{aligned}$$

$\mathsf{E}[\mathbf{e}] = \mathbf{0}$ if $\boldsymbol{\mu} \in C(\mathbf{X})$

$$\begin{aligned} \mathsf{E}[\mathbf{e}] &= \mathsf{E}[(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}] \\ &= (\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathsf{E}[\mathbf{Y}] \\ &= (\mathbf{I} - \mathsf{P}_{\mathbf{X}})\boldsymbol{\mu} \end{aligned}$$

# Properties

$\hat{\mathbf{Y}} = \hat{\boldsymbol{\mu}}$ is an unbiased estimate of $\boldsymbol{\mu} = \mathbf{X}\beta$

$$\begin{aligned} \mathsf{E}[\hat{\mathbf{Y}}] &= \mathsf{E}[\mathsf{P}_{\mathbf{X}}\mathbf{Y}] \\ &= \mathsf{P}_{\mathbf{X}}\mathsf{E}[\mathbf{Y}] \\ &= \mathsf{P}_{\mathbf{X}}\boldsymbol{\mu} \\ &= \boldsymbol{\mu} \end{aligned}$$

$\mathsf{E}[\mathbf{e}] = \mathbf{0}$ if $\boldsymbol{\mu} \in C(\mathbf{X})$

$$\begin{aligned} \mathsf{E}[\mathbf{e}] &= \mathsf{E}[(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}] \\ &= (\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathsf{E}[\mathbf{Y}] \\ &= (\mathbf{I} - \mathsf{P}_{\mathbf{X}})\boldsymbol{\mu} \\ &= \mathbf{0} \end{aligned}$$

$\hat{\mathbf{Y}} = \hat{\boldsymbol{\mu}}$ is an unbiased estimate of $\boldsymbol{\mu} = \mathbf{X}\beta$

$$\begin{aligned} \mathsf{E}[\hat{\mathbf{Y}}] &= \mathsf{E}[\mathsf{P}_{\mathbf{X}}\mathbf{Y}] \\ &= \mathsf{P}_{\mathbf{X}}\mathsf{E}[\mathbf{Y}] \\ &= \mathsf{P}_{\mathbf{X}}\boldsymbol{\mu} \\ &= \boldsymbol{\mu} \end{aligned}$$

$\mathsf{E}[\mathbf{e}] = \mathbf{0}$ if $\boldsymbol{\mu} \in C(\mathbf{X})$

$$\begin{aligned} \mathsf{E}[\mathbf{e}] &= \mathsf{E}[(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}] \\ &= (\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathsf{E}[\mathbf{Y}] \\ &= (\mathbf{I} - \mathsf{P}_{\mathbf{X}})\boldsymbol{\mu} \\ &= \mathbf{0} \end{aligned}$$

Will not be $\mathbf{0}$ if $\boldsymbol{\mu} \notin C(\mathbf{X})$

# Estimate of $\sigma^2$

MLE of $\sigma^2$:

$$\hat{\sigma}^2 = \frac{\mathbf{e}^T\mathbf{e}}{n} = \frac{\mathbf{Y}^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}}{n}$$

# Estimate of $\sigma^2$

MLE of $\sigma^2$:

$$\hat{\sigma}^2 = \frac{\mathbf{e}^T \mathbf{e}}{n} = \frac{\mathbf{Y}^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}}{n}$$

Is this an unbiased estimate of $\sigma^2$?

# Estimate of $\sigma^2$

MLE of $\sigma^2$:

$$\hat{\sigma}^2 = \frac{\mathbf{e}^T\mathbf{e}}{n} = \frac{\mathbf{Y}^T(\mathbf{I} - \mathrm{P}_{\mathbf{X}})\mathbf{Y}}{n}$$

Is this an unbiased estimate of $\sigma^2$?

Need expectations of quadratic forms $\mathbf{Y}^T\mathbf{A}\mathbf{Y}$ for $\mathbf{A}$ an $n \times n$ matrix $\mathbf{Y}$ a random vector in $\mathbb{R}^n$

# Quadratic Forms

Without loss of generality we can assume that $\mathbf{A} = \mathbf{A}^T$

## Quadratic Forms

Without loss of generality we can assume that $\mathbf{A} = \mathbf{A}^T$

- $\mathbf{Y}^T\mathbf{A}\mathbf{Y}$ is a scalar

# Quadratic Forms

Without loss of generality we can assume that $\mathbf{A} = \mathbf{A}^T$

- $\mathbf{Y}^T\mathbf{A}\mathbf{Y}$ is a scalar
- $\mathbf{Y}^T\mathbf{A}\mathbf{Y} = (\mathbf{Y}^T\mathbf{A}\mathbf{Y})^T = \mathbf{Y}^T\mathbf{A}^T\mathbf{Y}$

# Quadratic Forms

Without loss of generality we can assume that $\mathbf{A} = \mathbf{A}^T$

- $\mathbf{Y}^T\mathbf{A}\mathbf{Y}$ is a scalar
- $\mathbf{Y}^T\mathbf{A}\mathbf{Y} = (\mathbf{Y}^T\mathbf{A}\mathbf{Y})^T = \mathbf{Y}^T\mathbf{A}^T\mathbf{Y}$

$$\frac{\mathbf{Y}^T\mathbf{A}\mathbf{Y} + \mathbf{Y}^T\mathbf{A}^T\mathbf{Y}}{2} = \mathbf{Y}^T\mathbf{A}\mathbf{Y}$$

# Quadratic Forms

Without loss of generality we can assume that $\mathbf{A} = \mathbf{A}^T$

- $\mathbf{Y}^T\mathbf{A}\mathbf{Y}$ is a scalar
- $\mathbf{Y}^T\mathbf{A}\mathbf{Y} = (\mathbf{Y}^T\mathbf{A}\mathbf{Y})^T = \mathbf{Y}^T\mathbf{A}^T\mathbf{Y}$

$$\begin{aligned}
\frac{\mathbf{Y}^T\mathbf{A}\mathbf{Y} + \mathbf{Y}^T\mathbf{A}^T\mathbf{Y}}{2} &= \mathbf{Y}^T\mathbf{A}\mathbf{Y} \\
\mathbf{Y}^T\frac{(\mathbf{A} + \mathbf{A}^T)}{2}\mathbf{Y} &= \mathbf{Y}^T\mathbf{A}\mathbf{Y}
\end{aligned}$$

# Quadratic Forms

Without loss of generality we can assume that $\mathbf{A} = \mathbf{A}^T$

- $\mathbf{Y}^T\mathbf{A}\mathbf{Y}$ is a scalar
- $\mathbf{Y}^T\mathbf{A}\mathbf{Y} = (\mathbf{Y}^T\mathbf{A}\mathbf{Y})^T = \mathbf{Y}^T\mathbf{A}^T\mathbf{Y}$

$$\begin{aligned} \frac{\mathbf{Y}^T\mathbf{A}\mathbf{Y} + \mathbf{Y}^T\mathbf{A}^T\mathbf{Y}}{2} &= \mathbf{Y}^T\mathbf{A}\mathbf{Y} \\ \mathbf{Y}^T\frac{(\mathbf{A} + \mathbf{A}^T)}{2}\mathbf{Y} &= \mathbf{Y}^T\mathbf{A}\mathbf{Y} \end{aligned}$$

- may take $\mathbf{A} = \mathbf{A}^T$

**Theorem**

*Let* $\mathbf{Y}$ *be a random vector in* $\mathbb{R}^n$ *with* $E[\mathbf{Y}] = \boldsymbol{\mu}$ *and* $Cov(\mathbf{Y}) = \boldsymbol{\Sigma}$.

**Theorem**

*Let $\mathbf{Y}$ be a random vector in $\mathbb{R}^n$ with $E[\mathbf{Y}] = \boldsymbol{\mu}$ and $Cov(\mathbf{Y}) = \boldsymbol{\Sigma}$. Then $E[\mathbf{Y}^T\mathbf{A}\mathbf{Y}] = tr\mathbf{A}\Sigma + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}$.*

# Expectations of Quadratic Forms

**Theorem**

*Let* $\mathbf{Y}$ *be a random vector in* $\mathbb{R}^n$ *with* $E[\mathbf{Y}] = \boldsymbol{\mu}$ *and* $Cov(\mathbf{Y}) = \boldsymbol{\Sigma}$. *Then* $E[\mathbf{Y}^T\mathbf{A}\mathbf{Y}] = tr\mathbf{A}\Sigma + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}$.

Result useful for finding expected values of Mean Squares; no normality required!

# Proof

Start with $(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{A} (\mathbf{Y} - \boldsymbol{\mu})$, expand and take expectations

# Proof

Start with $(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})$, expand and take expectations

$$\mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})] \;=\; \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\mu^T\mathbf{A}\mathbf{Y}-\mathbf{Y}^T\mathbf{A}\mu]$$

# Proof

Start with $(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{A} (\mathbf{Y} - \boldsymbol{\mu})$, expand and take expectations

$$
\begin{aligned}
\mathsf{E}[(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{A} (\mathbf{Y} - \boldsymbol{\mu})] &= \mathsf{E}[\mathbf{Y}^T \mathbf{A} \mathbf{Y} + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} - \mu^T \mathbf{A} \mathbf{Y} - \mathbf{Y}^T \mathbf{A} \mu] \\
&= \mathsf{E}[\mathbf{Y}^T \mathbf{A} \mathbf{Y}] + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} - \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} - \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}
\end{aligned}
$$

# Proof

Start with $(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})$, expand and take expectations

$$\begin{aligned}
\mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})] &= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\mu^T\mathbf{A}\mathbf{Y}-\mathbf{Y}^T\mathbf{A}\mu] \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}$$

# Proof

Start with $(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{A} (\mathbf{Y} - \boldsymbol{\mu})$, expand and take expectations

$$
\begin{aligned}
\mathsf{E}[(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{A} (\mathbf{Y} - \boldsymbol{\mu})] &= \mathsf{E}[\mathbf{Y}^T \mathbf{A} \mathbf{Y} + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} - \mu^T \mathbf{A} \mathbf{Y} - \mathbf{Y}^T \mathbf{A} \mu] \\
&= \mathsf{E}[\mathbf{Y}^T \mathbf{A} \mathbf{Y}] + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} - \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} - \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} \\
&= \mathsf{E}[\mathbf{Y}^T \mathbf{A} \mathbf{Y}] - \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}
\end{aligned}
$$

Rearrange

# Proof

Start with $(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})$, expand and take expectations

$$\begin{aligned}
\mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})] &= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\mu^T\mathbf{A}\mathbf{Y}-\mathbf{Y}^T\mathbf{A}\mu] \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}$$

Rearrange

$$\mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}] = \mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}$$

# Proof

Start with $(\mathbf{Y}-\mu)^T\mathbf{A}(\mathbf{Y}-\mu)$, expand and take expectations

$$
\begin{aligned}
\mathrm{E}[(\mathbf{Y}-\mu)^T\mathbf{A}(\mathbf{Y}-\mu)] &= \mathrm{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y} + \mu^T\mathbf{A}\mu - \mu^T\mathbf{A}\mathbf{Y} - \mathbf{Y}^T\mathbf{A}\mu] \\
&= \mathrm{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}] + \mu^T\mathbf{A}\mu - \mu^T\mathbf{A}\mu - \mu^T\mathbf{A}\mu \\
&= \mathrm{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}] - \mu^T\mathbf{A}\mu
\end{aligned}
$$

Rearrange

$$
\begin{aligned}
\mathrm{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}] &= \mathrm{E}[(\mathbf{Y}-\mu)^T\mathbf{A}(\mathbf{Y}-\mu)] + \mu^T\mathbf{A}\mu \\
&= \mathrm{E}[\mathrm{tr}(\mathbf{Y}-\mu)^T\mathbf{A}(\mathbf{Y}-\mu)] + \mu^T\mathbf{A}\mu
\end{aligned}
$$

# Proof

Start with $(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})$, expand and take expectations

$$
\begin{aligned}
\mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})] &= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\mu^T\mathbf{A}\mathbf{Y}-\mathbf{Y}^T\mathbf{A}\mu] \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}
$$

Rearrange

$$
\begin{aligned}
\mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}] &= \mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathrm{tr}(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathrm{tr}\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}
$$

# Proof

Start with $(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})$, expand and take expectations

$$\begin{aligned}
\mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})] &= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\mu^T\mathbf{A}\mathbf{Y}-\mathbf{Y}^T\mathbf{A}\mu] \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}$$

Rearrange

$$\begin{aligned}
\mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}] &= \mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathrm{tr}(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathrm{tr}\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}\mathsf{E}[\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}$$

# Proof

Start with $(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})$, expand and take expectations

$$
\begin{aligned}
\mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})] &= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\mu^T\mathbf{A}\mathbf{Y}-\mathbf{Y}^T\mathbf{A}\mu] \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}
$$

Rearrange

$$
\begin{aligned}
\mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}] &= \mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathrm{tr}(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathrm{tr}\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}\mathsf{E}[\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}\mathbf{A}\mathsf{E}([(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}
$$

# Proof

Start with $(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})$, expand and take expectations

$$
\begin{aligned}
\mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})] &= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\mu^T\mathbf{A}\mathbf{Y}-\mathbf{Y}^T\mathbf{A}\mu] \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}
$$

Rearrange

$$
\begin{aligned}
\mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}] &= \mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathrm{tr}(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathrm{tr}\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}\mathsf{E}[\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}\mathbf{A}\mathsf{E}([(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}\mathbf{A}\boldsymbol{\Sigma}+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}
$$

# Proof

Start with $(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})$, expand and take expectations

$$
\begin{aligned}
\mathrm{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})] &= \mathrm{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\mu^T\mathbf{A}\mathbf{Y}-\mathbf{Y}^T\mathbf{A}\mu] \\
&= \mathrm{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}
$$

Rearrange

$$
\begin{aligned}
\mathrm{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}] &= \mathrm{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{E}[\mathrm{tr}(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{E}[\mathrm{tr}\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}\mathrm{E}[\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}\mathbf{A}\mathrm{E}([(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}\mathbf{A}\boldsymbol{\Sigma}+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}
$$

$\mathrm{tr}\mathbf{A} \equiv \sum_{i=1}^{n} a_{ii}$

# Expectation of $\hat{\sigma}^2$

Use the theorem:

# Expectation of $\hat{\sigma}^2$

Use the theorem:

$$\mathsf{E}[\mathbf{Y}^T(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}] \quad = \quad \mathrm{tr}(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\sigma^2\mathbf{I} + \boldsymbol{\mu}^T(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\boldsymbol{\mu}$$

# Expectation of $\hat{\sigma}^2$

Use the theorem:

$$\begin{aligned} \mathsf{E}[\mathbf{Y}^T(\mathbf{I}-\mathsf{P}_{\mathbf{X}})\mathbf{Y}] &= \mathrm{tr}(\mathbf{I}-\mathsf{P}_{\mathbf{X}})\sigma^2\mathbf{I} + \boldsymbol{\mu}^T(\mathbf{I}-\mathsf{P}_{\mathbf{X}})\boldsymbol{\mu} \\ &= \sigma^2\mathrm{tr}(\mathbf{I}-\mathsf{P}_{\mathbf{X}}) \end{aligned}$$

# Expectation of $\hat{\sigma}^2$

Use the theorem:

$$\begin{aligned}
\mathsf{E}[\mathbf{Y}^T(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}] &= \operatorname{tr}(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\sigma^2\mathbf{I} + \boldsymbol{\mu}^T(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\boldsymbol{\mu} \\
&= \sigma^2 \operatorname{tr}(\mathbf{I} - \mathsf{P}_{\mathbf{X}}) \\
&= \sigma^2 r(\mathbf{I} - \mathsf{P}_{\mathbf{X}})
\end{aligned}$$

# Expectation of $\hat{\sigma}^2$

Use the theorem:

$$\begin{aligned}
\mathrm{E}[\mathbf{Y}^T(\mathbf{I}-\mathrm{P}_{\mathbf{X}})\mathbf{Y}] &= \mathrm{tr}(\mathbf{I}-\mathrm{P}_{\mathbf{X}})\sigma^2\mathbf{I} + \boldsymbol{\mu}^T(\mathbf{I}-\mathrm{P}_{\mathbf{X}})\boldsymbol{\mu} \\
&= \sigma^2\mathrm{tr}(\mathbf{I}-\mathrm{P}_{\mathbf{X}}) \\
&= \sigma^2 r(\mathbf{I}-\mathrm{P}_{\mathbf{X}}) \\
&= \sigma^2(n - r(\mathbf{X}))
\end{aligned}$$

# Expectation of $\hat{\sigma}^2$

Use the theorem:

$$\begin{aligned}
\mathsf{E}[\mathbf{Y}^T(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}] &= \operatorname{tr}(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\sigma^2\mathbf{I} + \boldsymbol{\mu}^T(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\boldsymbol{\mu} \\
&= \sigma^2 \operatorname{tr}(\mathbf{I} - \mathsf{P}_{\mathbf{X}}) \\
&= \sigma^2 r(\mathbf{I} - \mathsf{P}_{\mathbf{X}}) \\
&= \sigma^2(n - r(\mathbf{X}))
\end{aligned}$$

Therefore an unbiased estimate of $\sigma^2$ is

$$\frac{\mathbf{e}^T\mathbf{e}}{n - r(\mathbf{X})}$$

# Expectation of $\hat{\sigma}^2$

Use the theorem:

$$\begin{aligned} \mathsf{E}[\mathbf{Y}^T(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}] &= \operatorname{tr}(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\sigma^2\mathbf{I} + \boldsymbol{\mu}^T(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\boldsymbol{\mu} \\ &= \sigma^2\operatorname{tr}(\mathbf{I} - \mathsf{P}_{\mathbf{X}}) \\ &= \sigma^2 r(\mathbf{I} - \mathsf{P}_{\mathbf{X}}) \\ &= \sigma^2(n - r(\mathbf{X})) \end{aligned}$$

Therefore an unbiased estimate of $\sigma^2$ is

$$\frac{\mathbf{e}^T\mathbf{e}}{n - r(\mathbf{X})}$$

**Trace of a Projection Matrix**

If $\mathsf{P}$ is an orthogonal projection matrix, then its eigenvalues $\lambda_i$ are either zero or one with $\operatorname{tr}(\mathsf{P}) = \sum_i(\lambda_i) = r(\mathsf{P})$

# Spectral Theorem

**Theorem**

*If* $\mathbf{A}$ $(n \times n)$ *is a symmetric real matrix then there exists a* $\mathbf{U}$ $(n \times n)$ *such that* $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}_n$ *and a diagonal matrix* $\mathbf{\Lambda}$ *with elements* $\lambda_i$ *such that* $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$

# Spectral Theorem

**Theorem**

*If* $\mathbf{A}$ $(n \times n)$ *is a symmetric real matrix then there exists a* $\mathbf{U}$ $(n \times n)$ *such that* $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}_n$ *and a diagonal matrix* $\boldsymbol{\Lambda}$ *with elements* $\lambda_i$ *such that* $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$

- $\mathbf{U}$ is an orthogonal matrix; $\mathbf{U}^{-1} = \mathbf{U}^T$

# Spectral Theorem

**Theorem**

*If* $\mathbf{A}$ $(n \times n)$ *is a symmetric real matrix then there exists a* $\mathbf{U}$ $(n \times n)$ *such that* $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}_n$ *and a diagonal matrix* $\boldsymbol{\Lambda}$ *with elements* $\lambda_i$ *such that* $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$

- $\mathbf{U}$ is an orthogonal matrix; $\mathbf{U}^{-1} = \mathbf{U}^T$
- The columns of $\mathbf{U}$ from an Orthonormal Basis for $\mathbb{R}^n$

# Spectral Theorem

**Theorem**

*If* $\mathbf{A}$ $(n \times n)$ *is a symmetric real matrix then there exists a* $\mathbf{U}$ $(n \times n)$ *such that* $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}_n$ *and a diagonal matrix* $\mathbf{\Lambda}$ *with elements* $\lambda_i$ *such that* $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$

- $\mathbf{U}$ is an orthogonal matrix; $\mathbf{U}^{-1} = \mathbf{U}^T$
- The columns of $\mathbf{U}$ from an Orthonormal Basis for $\mathbb{R}^n$
- rank of $\mathbf{A}$ equals the number of non-zero eigenvalues $\lambda_i$

# Spectral Theorem

## Theorem

*If* $\mathbf{A}$ $(n \times n)$ *is a symmetric real matrix then there exists a* $\mathbf{U}$ $(n \times n)$ *such that* $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}_n$ *and a diagonal matrix* $\mathbf{\Lambda}$ *with elements* $\lambda_i$ *such that* $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$

- $\mathbf{U}$ is an orthogonal matrix; $\mathbf{U}^{-1} = \mathbf{U}^T$
- The columns of $\mathbf{U}$ from an Orthonormal Basis for $\mathbb{R}^n$
- rank of $\mathbf{A}$ equals the number of non-zero eigenvalues $\lambda_i$
- Columns of $\mathbf{U}$ associated with non-zero eigenvalues form an ONB for $C(\mathbf{A})$ (eigenvectors of $\mathbf{A}$)

# Spectral Theorem

## Theorem

*If* $\mathbf{A}$ $(n \times n)$ *is a symmetric real matrix then there exists a* $\mathbf{U}$ $(n \times n)$ *such that* $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}_n$ *and a diagonal matrix* $\boldsymbol{\Lambda}$ *with elements* $\lambda_i$ *such that* $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$

- $\mathbf{U}$ is an orthogonal matrix; $\mathbf{U}^{-1} = \mathbf{U}^T$
- The columns of $\mathbf{U}$ from an Orthonormal Basis for $\mathbb{R}^n$
- rank of $\mathbf{A}$ equals the number of non-zero eigenvalues $\lambda_i$
- Columns of $\mathbf{U}$ associated with non-zero eigenvalues form an ONB for $C(\mathbf{A})$ (eigenvectors of $\mathbf{A}$)
- $\mathbf{A}^p = \mathbf{U}\boldsymbol{\Lambda}^p\mathbf{U}^T$ (matrix powers)

# Spectral Theorem

## Theorem

*If $\mathbf{A}$ $(n \times n)$ is a symmetric real matrix then there exists a $\mathbf{U}$ $(n \times n)$ such that $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}_n$ and a diagonal matrix $\mathbf{\Lambda}$ with elements $\lambda_i$ such that $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$*

- $\mathbf{U}$ is an orthogonal matrix; $\mathbf{U}^{-1} = \mathbf{U}^T$
- The columns of $\mathbf{U}$ from an Orthonormal Basis for $\mathbb{R}^n$
- rank of $\mathbf{A}$ equals the number of non-zero eigenvalues $\lambda_i$
- Columns of $\mathbf{U}$ associated with non-zero eigenvalues form an ONB for $C(\mathbf{A})$ (eigenvectors of $\mathbf{A}$)
- $\mathbf{A}^p = \mathbf{U}\mathbf{\Lambda}^p\mathbf{U}^T$ (matrix powers)
- a square root of $\mathbf{A} > 0$ is $\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}^T$

# Spectral Theorem

## Theorem

*If* $\mathbf{A}$ $(n \times n)$ *is a symmetric real matrix then there exists a* $\mathbf{U}$ $(n \times n)$ *such that* $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}_n$ *and a diagonal matrix* $\mathbf{\Lambda}$ *with elements* $\lambda_i$ *such that* $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$

- $\mathbf{U}$ is an orthogonal matrix; $\mathbf{U}^{-1} = \mathbf{U}^T$
- The columns of $\mathbf{U}$ from an Orthonormal Basis for $\mathbb{R}^n$
- rank of $\mathbf{A}$ equals the number of non-zero eigenvalues $\lambda_i$
- Columns of $\mathbf{U}$ associated with non-zero eigenvalues form an ONB for $C(\mathbf{A})$ (eigenvectors of $\mathbf{A}$)
- $\mathbf{A}^p = \mathbf{U}\mathbf{\Lambda}^p\mathbf{U}^T$ (matrix powers)
- a square root of $\mathbf{A} > 0$ is $\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}^T$

# Projections

## Projection Matrix

If P is an orthogonal projection matrix, then its eigenvalues $\lambda_i$ are either zero or one with $\text{tr}(\text{P}) = \sum_i(\lambda_i) = r(\text{P})$

# Projections

## Projection Matrix

If P is an orthogonal projection matrix, then its eigenvalues $\lambda_i$ are either zero or one with $\mathrm{tr}(\mathrm{P}) = \sum_i(\lambda_i) = r(\mathrm{P})$

- $\mathrm{P} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$

# Projections

## Projection Matrix

If P is an orthogonal projection matrix, then its eigenvalues $\lambda_i$ are either zero or one with $\text{tr}(\mathrm{P}) = \sum_i(\lambda_i) = r(\mathrm{P})$

- $\mathrm{P} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$
- $\mathrm{P} = \mathrm{P}^2 \Rightarrow \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T = \mathbf{U}\boldsymbol{\Lambda}^2\mathbf{U}^T$

# Projections

**Projection Matrix**

If P is an orthogonal projection matrix, then its eigenvalues $\lambda_i$ are either zero or one with $\mathrm{tr}(\mathrm{P}) = \sum_i(\lambda_i) = r(\mathrm{P})$

- $\mathrm{P} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$
- $\mathrm{P} = \mathrm{P}^2 \Rightarrow \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T = \mathbf{U}\boldsymbol{\Lambda}^2\mathbf{U}^T$
- $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^2$ is true only for $\lambda_i = 1$ or $\lambda_i = 0$

## Projection Matrix

If P is an orthogonal projection matrix, then its eigenvalues $\lambda_i$ are either zero or one with $\mathrm{tr}(\mathrm{P}) = \sum_i(\lambda_i) = r(\mathrm{P})$

- $\mathrm{P} = \mathbf{U\Lambda U}^T$
- $\mathrm{P} = \mathrm{P}^2 \Rightarrow \mathbf{U\Lambda U}^T\mathbf{U\Lambda U}^T = \mathbf{U\Lambda}^2\mathbf{U}^T$
- $\mathbf{\Lambda} = \mathbf{\Lambda}^2$ is true only for $\lambda_i = 1$ or $\lambda_i = 0$
- Since $r(\mathrm{P})$ is the number of non-zero eigenvalues, $r(\mathrm{P}) = \sum \lambda_i = \mathrm{tr}(\mathrm{P})$

## Projection Matrix

If P is an orthogonal projection matrix, then its eigenvalues $\lambda_i$ are either zero or one with $\mathrm{tr}(\mathrm{P}) = \sum_i(\lambda_i) = r(\mathrm{P})$

- $\mathrm{P} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$
- $\mathrm{P} = \mathrm{P}^2 \Rightarrow \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T = \mathbf{U}\boldsymbol{\Lambda}^2\mathbf{U}^T$
- $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^2$ is true only for $\lambda_i = 1$ or $\lambda_i = 0$
- Since $r(\mathrm{P})$ is the number of non-zero eigenvalues, $r(\mathrm{P}) = \sum \lambda_i = \mathrm{tr}(\mathrm{P})$

$$\mathrm{P} = [\mathbf{U}_P \mathbf{U}_{P^\perp}] \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{n-r} \end{bmatrix} \begin{bmatrix} \mathbf{U}_P^T \\ \mathbf{U}_{P^\perp}^T \end{bmatrix} = \mathbf{U}_P \mathbf{U}_P^T$$

$$\mathrm{P} = \sum_{i=1}^{r} \mathbf{u}_i \mathbf{u}_i^T$$

sum of $r$ rank 1 projections.

# Prostate Example

```
> library(lasso2)
> summary(lm(lcavol ~ ., data=Prostate))

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.260101   1.259683  -1.794   0.0762 .
lweight     -0.073166   0.174450  -0.419   0.6759
age          0.022736   0.010964   2.074   0.0410 *
lbph        -0.087449   0.058084  -1.506   0.1358
svi         -0.153591   0.253932  -0.605   0.5468
lcp          0.367300   0.081689   4.496 2.10e-05 ***
gleason      0.190759   0.154283   1.236   0.2196
pgg45       -0.007158   0.004326  -1.654   0.1016
lpsa         0.572797   0.085790   6.677 2.11e-09 ***
---

Residual standard error: 0.6998 on 88 degrees of freedom
Multiple R-squared:  0.6769, Adjusted R-squared:  0.6475
F-statistic: 23.04 on 8 and 88 DF,  p-value: < 2.2e-16
```