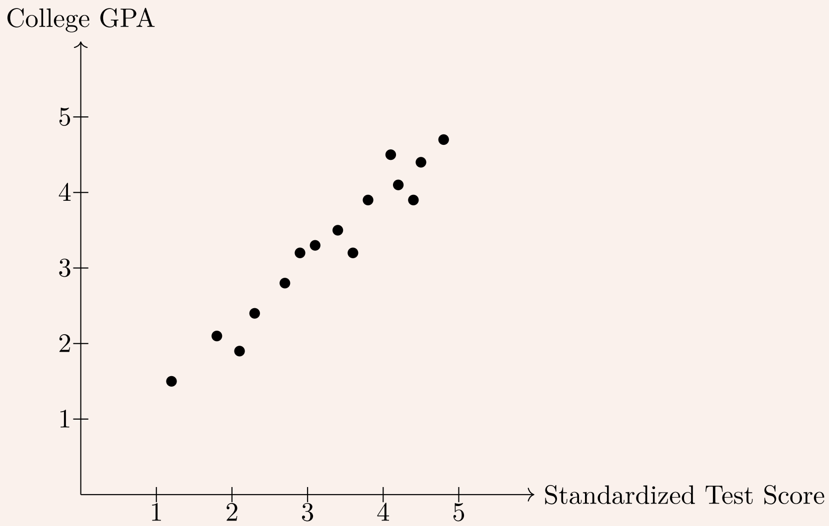
STAT 135 Lecture 17

Henry Liev

31 July 2025

Example 0.1



Test	GPA
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$

$y = \beta_0 + \beta_1 x$

$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x)^2$

Example 0.2

Generalize

Student	College Gpa	Test 1	Test 2	Test 3	$\cdots$	Test (p-1)
1	$\vdots$					
2	$\vdots$					
3	$\vdots$					
$\vdots$	$\vdots$					
n	$\vdots$					

Let y_i = college GPA for student i

Let x_{ij} = test score for student i on test j

Try to find the best linear prediction

Method 0.3 (Fitting Higher Dimensional "Line")

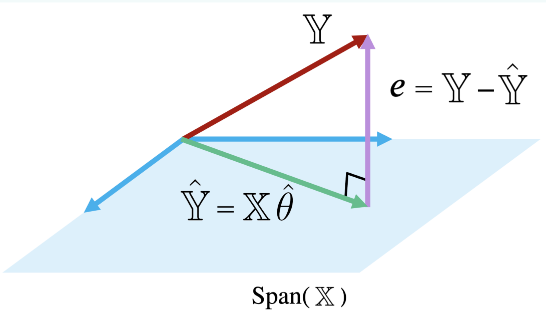
$y_i \approx \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{p-1} x_{i(p-1)}$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & \cdots & x_{1(p-1)} \\ 1 & x_{21} & x_{22} & \cdots & & \\ 1 & & & & & \\ \vdots & & & & & \\ 1 & x_{n1} & & & & x_{n(p-1)} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$$

$$\mathbb{Y} \approx \mathbb{X}\beta$$

$\mathbb{Y}$ $n \times 1$ matrix, $\mathbb{X}$ $n \times p$ matrix, β $p \times 1$ matrix

Think of $\mathbb{Y}$ as a point in $\mathbb{R}^n$



$\hat{\mathbb{Y}} - \mathbb{Y}$ is orthogonal to span$(\mathbb{X})$

$\mathbb{X}(\hat{\mathbb{Y}} - \mathbb{Y}) = 0$

$\mathbb{X}(\hat{\mathbb{Y}}) - \mathbb{X}^T\mathbb{Y}$

$\mathbb{X}^T\mathbb{X}\hat{\beta} = \mathbb{X}^T\mathbb{Y}$

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

Method 0.4

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i(p-1)} + \varepsilon_{ij}$, ε_{ij} independent

$\mathbb{E}(\varepsilon_{ij}) = 0$ OLS is BLUE (Best Linear Unbiased Estimator) of β

$\mathrm{Var}(\varepsilon_{ij}) = \sigma^2$ Gauss-Markov Theorem

OLS is "good" even under weak assumptions

Method 0.5

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i(p-1)} + \varepsilon_{ij}$, $\varepsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$, σ^2 known

Choose β's which maximize likelihood

$\varepsilon_i = Y_i - X_i^T \beta$

$$f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n | \beta) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp\left(\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 \right)$$

Minimize $\sum_{i=1}^n (y_i - x_i^T \beta)^2$ to maximize the likelihood

MLE = OLS assuming normal error $\hat{\beta}$

Remark 0.6 (Properties of $\hat{\beta}$)

$\hat{\beta}$ is unbiased: $E(\hat{\beta}) = \beta$

Expectation of a vector: $\mathbb{E}\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \mathbb{E}[Y_1] \\ \mathbb{E}[Y_2] \\ \vdots \\ \mathbb{E}[Y_n] \end{bmatrix}$

$$\begin{aligned} \mathbb{E}(\hat{\beta}) &= \mathbb{E}\left[(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T(\mathbb{X}\beta + \varepsilon)\right] \\ &= \mathbb{E}\left[(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{X}\beta\right] + \left[(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}\right]\mathbb{E}(\varepsilon) \\ &= \beta \end{aligned}$$

$\hat{\beta}$ is unbiased and MLE
$\hat{\beta}$ is consistent
$\hat{\beta}$ is asymptotically (multivariate) normal
$\hat{\beta}$ is "efficient"
Y is a random vector

$$\operatorname{Cov}(\mathbb{Y}) = \begin{bmatrix} \operatorname{Var}(Y_1) & \operatorname{Cov}(Y_1, Y_2) & \operatorname{Cov}(Y_1, Y_3) & \cdots & \operatorname{Cov}(Y_1, Y_n) \\ \operatorname{Cov}(Y_2, Y_1) & \operatorname{Var}(Y_2) & \operatorname{Cov}(Y_2, Y_3) & \cdots & \operatorname{Cov}(Y_2, Y_n) \\ \operatorname{Cov}(Y_3, Y_1) & \operatorname{Cov}(Y_3, Y_2) & \operatorname{Var}(Y_3) & \cdots & \operatorname{Cov}(Y_3, Y_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \operatorname{Cov}(Y_n, Y_1) & \operatorname{Cov}(Y_n, Y_2) & \operatorname{Cov}(Y_n, Y_3) & \cdots & \operatorname{Var}(Y_n) \end{bmatrix}$$

$M_{ij} = \operatorname{Cov}(Y_i, Y_j)$
Intersted in $\operatorname{Cov}(\hat{\beta})$
Need one fact for a matrix $\mathbb{A}$ (constant) $\operatorname{Cov}(\mathbb{A}\mathbb{Y}) = \mathbb{A}\operatorname{Cov}(\mathbb{Y})\mathbb{A}^T$

Example 0.7

$\operatorname{Cov}(\hat{\beta})$

Solution. $\mathbb{Y} = \mathbb{X}\beta + \varepsilon \to \operatorname{Cov}(\mathbb{Y}) = \sigma^2\mathbb{I}$

$$\operatorname{Cov}((\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{Y}) = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\operatorname{Cov}(\mathbb{Y})\mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1} = \sigma^2(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{I} = \sigma^2(\mathbb{X}^T\mathbb{X})^{-1}$$

■

Remark 0.8 (Residuals)

Define the i^{th} residual
$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$
The i^{th} residual is our best guess of ε_i
Claim $\sum_{i=1}^n \hat{\varepsilon}_i = 0$
Remember $Y_i - \hat{Y}_i$ is orthogonal to columns of $\mathbb{X}$
Since we have a column of $\mathbb{1}$, and $Y_i - \hat{Y}_i$ is orthogonal to $\mathbb{X}$, then the dot product between the two is 0 $\hat{\varepsilon}_i$ estimates ε_i

Remark 0.9

"Sample Variance" of $\hat{\varepsilon}_i$ could estimate σ^2

$$\hat{\sigma}^2_{reg} = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \cdots - \hat{\beta}_{p-1} x_i)^2$$

Unbiased estimate of σ^2

Remark 0.10

$M = \mathrm{Cov}(\hat{\beta}) \approx \hat{\sigma}^2_{reg}(\mathbb{X}^T\mathbb{X})^{-1}$
$\mathrm{SE}(\hat{\beta}_i) = \sqrt{M_{ii}}$
$\frac{\hat{\beta}_i - \beta_i}{\mathrm{SE}(\hat{\beta}_i)} \sim t_{n-p}$