MSDS 596 Regression & Time Series

Lecture 02 Gauss Markov theorem and the R^2

Department of Statistics
Rutgers University

Sept 19, 2022

Welcome to MSDS 596: Regression & Time Series

- Meeting time: **Mondays 6-9pm**
- Instructor: **Koulik Khamaru** (kk1241@rutgers.edu)
  Office hour: Monday 4-5pm, 479 Hill center
- Teaching Assistant: **Suvrojit Ghosh** (sg1565@scarletmail.rutgers.edu)
  Office hour: Wednesdays 3:30-4:30pm, 260 Hill center.
- Course website (piazza):
  https://piazza.com/rutgers/fall2022/1695459602
  The course website is our main means of communication for all things class-related.

Gauss-Markov Theorem

Gauss-Markov Theorem

Suppose $\mathbb{E}(\boldsymbol{y}) = \boldsymbol{X}\beta$, where $\boldsymbol{X}$ has full rank and $\mathrm{Var}(\boldsymbol{y}) = \sigma^2\boldsymbol{I}$. The least squares estimator

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

is the **best linear unbiased estimator (BLUE)** of $\boldsymbol{\beta}$, in the sense that for any $\psi = \boldsymbol{c}'\boldsymbol{\beta}$, of all unbiased estimators of ψ that are linear in $\boldsymbol{y}$, $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ has minimum variance.

Gauss Markov theorem proof

- Let

$$\tilde{\beta} = By = \left\{ (X^\top X)^{-1} X^\top + G \right\} y$$

- Show that $GX = 0$.
- Show $\mathrm{var}(Gy) = \mathrm{var}(\hat{\beta}_{OLS}) + \sigma^2 GG^\top$.

Variance estimate

- Recall the linear model is

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i, \quad 1 \leq i \leq n,$$

  assuming $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent and identically distributed (iid) with mean zero and variance σ^2.

- The estimate of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{n - (p+1)} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

  - Unbiased estimate, i.e. $\mathbb{E}\hat{\sigma}^2 = \sigma^2$.

- The covariance matrix of $\hat{\beta}$ is

$$\mathrm{Cov}(\hat{\beta}) = \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

  - It is estimated by $\widehat{\mathrm{Cov}}(\hat{\beta}) = \hat{\sigma}^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}$.

Variance estimate

- Recall the linear model is

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i, \quad 1 \leq i \leq n,$$

  assuming $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent and identically distributed (iid) with mean zero and variance σ^2.

- The estimate of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{n-(p+1)} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

  - Unbiased estimate, i.e. $\mathbb{E}\hat{\sigma}^2 = \sigma^2$.

- The covariance matrix of $\hat{\beta}$ is

$$\mathrm{Cov}(\hat{\beta}) = \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

  - It is estimated by $\widehat{\mathrm{Cov}}(\hat{\beta}) = \hat{\sigma}^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}$.

Variance estimate

- Recall the linear model is

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i, \quad 1 \le i \le n,$$

  assuming $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent and identically distributed (iid) with mean zero and variance σ^2.
- The estimate of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{n-(p+1)} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

  - Unbiased estimate, i.e. $\mathbb{E}\hat{\sigma}^2 = \sigma^2$.
- The covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

  - It is estimated by $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

Goodness of fit

We need to measure how well the model fits the data.

- Total sum of squares (Total SS): $\sum_{i=1}^{n}(y_i - \bar{y})^2$.
- Residual sum of squares (RSS): $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.
- Coefficient of determination:

$$R^2 := 1 - \frac{\text{RSS}}{\text{Total SS}},$$

  noting that $0 \leq R^2 \leq 1$.
- hint: Look at the definition of least square estimate.

Goodness of fit

We need to measure how well the model fits the data.

- Total sum of squares (Total SS): $\sum_{i=1}^{n}(y_i - \bar{y})^2$.
- Residual sum of squares (RSS): $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.
- Coefficient of determination:

$$R^2 := 1 - \frac{\text{RSS}}{\text{Total SS}},$$

  noting that $0 \leq R^2 \leq 1$.
- hint: Look at the definition of least square estimate.

Goodness of fit

We need to measure how well the model fits the data.

- Total sum of squares (Total SS): $\sum_{i=1}^{n}(y_i - \bar{y})^2$.
- Residual sum of squares (RSS): $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.
- Coefficient of determination:

$$R^2 := 1 - \frac{\text{RSS}}{\text{Total SS}},$$

  noting that $0 \leq R^2 \leq 1$.
- hint: Look at the definition of least square estimate.

Goodness of fit

We need to measure how well the model fits the data.

- Total sum of squares (Total SS): $\sum_{i=1}^{n}(y_i - \bar{y})^2$.
- Residual sum of squares (RSS): $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.
- Coefficient of determination:

$$R^2 := 1 - \frac{\text{RSS}}{\text{Total SS}},$$

  noting that $0 \leq R^2 \leq 1$.

- hint: Look at the definition of least square estimate.

Goodness of fit

We need to measure how well the model fits the data.

- Total sum of squares (Total SS): $\sum_{i=1}^{n}(y_i - \bar{y})^2$.
- Residual sum of squares (RSS): $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.
- Coefficient of determination:

$$R^2 := 1 - \frac{\text{RSS}}{\text{Total SS}},$$

  noting that $0 \leq R^2 \leq 1$.
- hint: Look at the definition of least square estimate.

Goodness of fit

- Coefficient of determination:

$$R^2 := 1 - \frac{\text{RSS}}{\text{Total SS}}.$$

- Large value of R^2 indicates better fit.
  - In biological and social sciences, and finances, the variables are usually weakly correlated, with a lot of noise, we would expect low values of R^2.
  - In physics and engineering, most data come from controlled experiments, typically get higher R^2.
  - Background knowledge may be required to judge the R^2 value.
- Do NOT rely on R^2 as the only measure of fit! (Chapters 6,7,8,10)

Maximum Likelihood Estimation (Next class)

- Linear model $y_i = \boldsymbol{x}_i'\beta + \epsilon_i$, $1 \le i \le n$.
- If the assumption is stronger: $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are iid $N(0, \sigma^2)$, then...
- In the simple regression case, likelihood and log likelihood:

$$L(\beta) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right] \right\},$$

$$\log L(\beta) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 .$$

Solve for the MLEs: $\frac{\partial \log L}{\partial \beta_0} \Big|_{\hat\beta_0, \hat\beta_1} = 0$ and $\frac{\partial \log L}{\partial \beta_1} \Big|_{\hat\beta_0, \hat\beta_1} = 0$.

- $\hat\beta$ is also the maximum likelihood estimate (MLE) of β.

[1]

[1] Read chap 1 and 3 of the book Statisical Inference 2nd ed. by Casella and Berger

Maximum Likelihood Estimation (Next class)

- Linear model $y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \epsilon_i$, $1 \leq i \leq n$.
- If the assumption is stronger: $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are iid $N(0, \sigma^2)$, then...
- In the simple regression case, likelihood and log likelihood:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right] \right\},$$

$$\log L(\boldsymbol{\beta}) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 .$$

Solve for the MLEs: $\frac{\partial \log L}{\partial \beta_0} \mid_{\hat{\beta}_0, \hat{\beta}_1} = 0$ and $\frac{\partial \log L}{\partial \beta_1} \mid_{\hat{\beta}_0, \hat{\beta}_1} = 0$.

- $\hat{\beta}$ is also the maximum likelihood estimate (MLE) of β. [1]

[1] Read chap 1 and 3 of the book Statisical Inference 2nd ed. by Casella and Berger

Maximum Likelihood Estimation (Next class)

- Linear model $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i$, $1 \le i \le n$.
- If the assumption is stronger: $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are iid $N(0, \sigma^2)$, then...
- In the simple regression case, likelihood and log likelihood:

$$
\begin{aligned}
L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right] \right\}, \\
\log L(\boldsymbol{\beta}) &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 .
\end{aligned}
$$

  Solve for the MLEs: $\frac{\partial \log L}{\partial \beta_0} \mid_{\hat{\beta}_0, \hat{\beta}_1} = 0$ and $\frac{\partial \log L}{\partial \beta_1} \mid_{\hat{\beta}_0, \hat{\beta}_1} = 0$.
- $\hat{\boldsymbol{\beta}}$ is also the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$.

[1]

[1] Read chap 1 and 3 of the book Statisical Inference 2nd ed. by Casella and Berger

Distribution of $\hat{\beta}$ (Next class)

- σ knwon
  - Distribution of $\hat{\beta}$.
  - Distribution of $a^\top \hat{\beta}$.
  - What happends when σ^2 is unknown?

- σ^2 is uknown
  - What is the MLE for $\hat{\beta}$?
  - What is the MLE for $\hat{\sigma}^2$?