Lecture 2 - 19 Oct 2023

GAUSSIAN SIMPLE LINEAR REGRESSION

Assume that on n statistical units we observe (y_i, x_i), $i=1,\dots,n$.
We assume that each y_i is realisation of a random variable Y_i, and that $Y_1,\dots,Y_n$ are independent.
We only consider one covariate x_i, $i=1,\dots,n$.

Consider the model
$$Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \qquad i=1,\dots,n$$

HYPOTHESES:

$$\left.\begin{array}{l} 1.\ \mathbb{E}[\varepsilon_i]=0 \quad i=1,\dots,n \\ 2.\ \mathrm{Var}(\varepsilon_i)=\sigma^2 \quad \text{for all } i=1,\dots,n \\ 3.\ \mathrm{cov}(\varepsilon_i,\varepsilon_k)=0 \quad i\neq k;\ i,k=1,\dots,n \end{array}\right\} \text{hyp. from last time}$$

4. ε_i have Gaussian distribution

Recall that for a normal r.v. corr $=0 \Rightarrow$ independence
$$\Rightarrow \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2) \qquad i=1,\dots,n$$
$$\Rightarrow Y_i \sim N(\beta_1+\beta_2 x_i, \sigma^2) \text{ independent (but not identically distributed)}$$

Since now we have distributive assumptions, we can derive the estimators for $\beta_1, \beta_2, \sigma^2$ using the maximum likelihood method.

parameter space $\Theta = \mathbb{R}^2 \times \mathbb{R}^+$ here, $\theta = (\beta_1, \beta_2, \sigma^2)$
sample space $\Omega = \mathbb{R}^n$

likelihood function $L(\theta) \propto f(y_1,\dots,y_n;\theta) \overset{ind}{=} \prod_{i=1}^n f(y_i;\theta)$
$$L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i-\beta_1-\beta_2 x_i)^2\right\}$$
$$= \cancel{(2\pi)^{-\frac{n}{2}}}\,(\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i)^2\right\}$$

loglikelihood $\ell(\theta) = \log L(\theta)$
$$= -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i)^2$$

score function $\ell_*(\theta) = \left[\frac{\partial \ell(\theta)}{\partial\theta_1}, \dots, \frac{\partial \ell(\theta)}{\partial\theta_q}\right]$ (here, $q=3$)

$$\begin{cases} \frac{\partial}{\partial\beta_1}\ell(\theta) = -\frac{1}{\cancel{2}\sigma^2}\sum_{i=1}^n (\cancel{-2})(y_i-\beta_1-\beta_2 x_i) = \frac{1}{\sigma^2}\sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i) \\ \frac{\partial}{\partial\beta_2}\ell(\theta) = -\frac{1}{\cancel{2}\sigma^2}\sum_{i=1}^n (-x_i)\cancel{2}(y_i-\beta_1-\beta_2 x_i) = \frac{1}{\sigma^2}\sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i)x_i \\ \frac{\partial}{\partial\sigma^2}\ell(\theta) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i)^2 \end{cases}$$

the MLE is found as $\hat\theta$ s.t. $\ell_*(\hat\theta)=0$

$$\begin{cases} \frac{1}{\sigma^2}\sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i) = 0 & \rightsquigarrow \sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i)=0 \quad (1) \\ \frac{1}{\sigma^2}\sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i)x_i = 0 & \rightsquigarrow \sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i)x_i = 0 \quad (2) \\ -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i)^2 = 0 \quad (3) & \end{cases}$$

(1) and (2) are exactly the same equations we already solved using OLS
they do not depend on σ^2
hence
$$\hat\beta_1 = \bar y - \hat\beta_2 \bar x \quad \text{and} \quad \hat\beta_2 = \frac{\sum_{i=1}^n (x_i-\bar x)(y_i-\bar y)}{\sum_{i=1}^n (x_i-\bar x)^2}$$
are maximum likelihood estimates.

Solving (3)
$$-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i)^2 = 0$$
$$-\frac{1}{\cancel{2}(\sigma^2)^2}\left[n\sigma^2 - \sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i)^2\right] = 0 \Rightarrow \hat\sigma^2 = \frac{\sum_{i=1}^n (y_i-\hat\beta_1-\hat\beta_2 x_i)^2}{n}$$
estimate of σ^2

The matrix of the 2nd derivatives
$$\ell_{**}(\theta) = \left\{\frac{\partial^2 \ell(\theta)}{\partial\theta_s\partial\theta_r}\right\}_{s,r=1,2,3}$$

$$\frac{\partial^2}{\partial\beta_1^2}\ell(\theta) = -\frac{n}{\sigma^2} \qquad \frac{\partial^2}{\partial\beta_1\partial\beta_2}\ell(\theta) = -\frac{n\bar x}{\sigma^2} \qquad \frac{\partial^2}{\partial\beta_1\partial\sigma^2}\ell(\theta) = -\frac{1}{(\sigma^2)^2}\sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i)$$
$$\frac{\partial^2}{\partial\beta_2^2}\ell(\theta) = -\frac{\sum_{i=1}^n x_i^2}{\sigma^2} \qquad \frac{\partial^2}{\partial\beta_2\partial\sigma^2}\ell(\theta) = -\frac{1}{(\sigma^2)^2}\sum_{i=1}^n x_i(y_i-\beta_1-\beta_2 x_i)$$
$$\frac{\partial^2}{\partial(\sigma^2)^2}\ell(\theta) = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\sum_{i=1}^n (y_i-\beta_1-\beta_2 x_i)^2$$

these are the arguments of the lik. equations (1) and (2). Hence they are $=0$ if evaluated at $(\hat\beta_1, \hat\beta_2, \hat\sigma^2)$

We need to evaluate these derivatives at $(\hat\beta_1, \hat\beta_2, \hat\sigma^2)$
Both $\frac{\partial^2}{\partial\beta_1\partial\sigma^2}\ell(\theta)$ and $\frac{\partial^2}{\partial\beta_2\partial\sigma^2}\ell(\theta)$ are $=0$ at $(\hat\beta_1, \hat\beta_2, \hat\sigma^2)$.

$$\left.\frac{\partial^2}{\partial(\sigma^2)^2}\ell(\theta)\right|_{\theta=\hat\theta} = \frac{n}{2(\hat\sigma^2)^2} - \frac{1}{(\hat\sigma^2)^3}\underbrace{\sum_{i=1}^n (y_i-\hat\beta_1-\hat\beta_2 x_i)^2}_{=n\hat\sigma^2}$$
$$= \frac{n}{2(\hat\sigma^2)^2} - \frac{n}{(\hat\sigma^2)^2}$$
$$= -\frac{n}{2(\hat\sigma^2)^2}$$

the observed information $j(\hat\theta) = -\ell_{**}(\hat\theta)$ then is

$$j(\hat\theta) = \left[\begin{array}{cc|c} \frac{n}{\hat\sigma^2} & \frac{n\bar x}{\hat\sigma^2} & 0 \\ \frac{n\bar x}{\hat\sigma^2} & \frac{\sum_{i=1}^n x_i^2}{\hat\sigma^2} & 0 \\ \hline 0 & 0 & \frac{n}{2(\hat\sigma^2)^2} \end{array}\right] = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{\underset{2\times2}{A}} & \begin{matrix}0\\0\end{matrix} \\ \hline 0 & 0 & b \end{array}\right]$$

and it is possible to show that $(\hat\beta_1, \hat\beta_2, \hat\sigma^2)$ is a maximum.

The maximum likelihood ESTIMATORS of $(\beta_1, \beta_2, \sigma^2)$ are
$$\hat\beta_1(\underline{Y}) = \bar Y - \hat\beta_2 \bar x \quad \text{and}$$
$$\hat\beta_2(\underline{Y}) = \frac{\sum_{i=1}^n (x_i-\bar x)(Y_i-\bar Y)}{\sum_{i=1}^n (x_i-\bar x)^2}$$
$$\hat\sigma^2(\underline{Y}) = \frac{\sum_{i=1}^n (Y_i-\hat\beta_1-\hat\beta_2 x_i)^2}{n}$$

From the theory of MLE, we automatically obtain an approximate distribution of $(\hat\beta_1(\underline{Y}), \hat\beta_2(\underline{Y}), \hat\sigma^2(\underline{Y}))$

$$\begin{bmatrix} \hat\beta_1(\underline{Y}) \\ \hat\beta_2(\underline{Y}) \\ \hat\sigma^2(\underline{Y}) \end{bmatrix} \dot\sim N_3\left(\begin{bmatrix} \beta_1 \\ \beta_2 \\ \sigma^2 \end{bmatrix}; j(\hat\theta)^{-1}\right)$$

the variance is $j(\hat\theta)^{-1}$
we won't compute it (but it's very simple, you can do it for exercise). However, notice that

$$j(\hat\theta) = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{\underset{2\times2}{A}} & \begin{matrix}0\\0\end{matrix} \\ \hline 0 & 0 & b \end{array}\right] \Rightarrow j(\hat\theta)^{-1} = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{A^{-1}} & \begin{matrix}0\\0\end{matrix} \\ \hline 0 & 0 & \frac{1}{b} \end{array}\right]$$

$$\Rightarrow \mathrm{cov}(\hat\beta_1, \hat\sigma^2) = \mathrm{cov}(\hat\beta_2, \hat\sigma^2) = 0$$
Moreover, they are Gaussian $\Rightarrow \hat\beta_1 \perp\!\!\!\perp \hat\sigma^2,\ \hat\beta_2 \perp\!\!\!\perp {}^\wedge 2$