Second course in Statistics. Lecture 22.

Simple linear regression analysis

- Parameter estimation (2): maximum likelihood estimation (MLE)
- Statistical inference: testing and confidence interval
- Diagnostics: residual analysis

Maximum likelihood estimation

Difference between LSE and MLE:

Least squares estimates were determined without having to specify the probability distribution of the random errors ε_i. If we are willing to assume that the ε_i are independent normally distributed random variables with mean zero and variance σ^2 for all $i = 1, 2, \cdots, n$, it is possible to determine maximum likelihood estimates of β_0, β_1 and σ^2.

Express probability distribution function of Y_i in terms of probability distribution of ε_i

$\varepsilon_i \sim N(0, \sigma^2)$

$Y_i - \beta_0 - \beta_1 x_i \sim N(0, \sigma^2)$

Density function of response variables

Joint probability distribution of $Y_1, Y_2, \cdots, Y_n$ is

$$f_{Y_1,\cdots Y_n}\left(y_1, y_2, \cdots y_n \mid \beta_0, \beta_1, \sigma^2\right)$$

$$= f_{Y_1}\left(y_1 \mid \beta_0, \beta_1, \sigma^2\right) \cdot f_{Y_2}\left(y_2 \mid \beta_0, \beta_1, \sigma^2\right) \cdot \cdots \cdot f_{Y_n}\left(y_n \mid \beta_0, \beta_1, \sigma^2\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(y_1 - \beta_0 - \beta_1 x_1\right)^2\right] \cdots \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(y_n - \beta_0 - \beta_1 x_n\right)^2\right]$$

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

Likelihood function

Likelihood function is therefore given by

$$L\left(y_1, y_2, \ldots y_n \mid \beta_0, \beta_1, \sigma^2\right) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2\right]$$

In the joint probability distribution function, Y_i's are regarded as random variables, whereas in likelihood function, y_i's are deterministic variables. For convenience, the maximum likelihood estimates are determined by maximizing natural log of the likelihood function.

$$\ln\left[L\left(\beta_0, \beta_1, \sigma^2\right)\right] = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\Sigma\left(y_i - \beta_0 - \beta_1 x_i\right)^2$$

Maximum likelihood estimation

Taking partial derivatives with respect to β_0, β_1 and σ^2 and equating them to zero, then

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

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

Maximum likelihood estimation

LSE and MLE yield the same estimation for β_0 and β_1, but estimation of σ^2 is different. MLE of σ^2 is biased.

Why do we bother with the maximum likelihood estimators since they are the same as the LS estimators? One of the reasons is that maximum likelihood estimators possess the desirable properties of consistency, sufficiency, and minimum variance. In addition, they provide the necessary means to develop sampling distribution for β_0 and β_1.

Properties of ML estimators

General properties of estimators of β_0, β_1 and $\hat{Y}$

Theorem 1: LS estimators or MLE $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1, i.e. $E\left(\hat{\beta}_1\right) = \beta_1$ and $E\left(\hat{\beta}_0\right) = \beta_0$

Definition: Best linear unbiased estimators (BLUE): $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1. If among these unbiased estimators of β_0 and β_1 there exist estimators whose variances are smaller than those of any other unbiased estimators of β_0 and β_1, then these are the best linear unbiased estimators (BLUE) of β_0 and β_1.

Properties of ML estimators

Theorem 2: $\hat{\beta}_0$ and $\hat{\beta}_1$ shown as above are the best linear unbiased estimators of β_0 and β_1.

$Var\left(\hat{\beta}_1\right) = \dfrac{\sigma^2}{S_{xx}}$ or equivalently $s.d.\left(\hat{\beta}_1\right) = \dfrac{\sigma}{\sqrt{S_{xx}}}$ and

$Var\left(\hat{\beta}_0\right) = \sigma^2\left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{S_{xx}}\right)$ or equivalently $s.d.\left(\hat{\beta}_0\right) = \sigma\sqrt{\dfrac{1}{n} + \dfrac{\bar{x}^2}{S_{xx}}}$

Theorem 3: $\hat{Y}_i$ is unbiased estimator of $E(Y_i)$, i.e. $E\left(\hat{Y}_i\right) = E(Y_i)$ and

$$s.d.\left(\hat{Y}_i\right) = \sigma\sqrt{\frac{1}{n} + \frac{\left(x_i - \bar{x}\right)}{S_{xx}}}$$

Sampling distribution of estimators in simple linear model

We have calculated the means and variances of $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{Y}_i$, what are their sampling distributions?

Recall:

ε_i is i.i.d. $n(0,\sigma^2)$ for $i = 1,2,\ldots,n$

σ^2 is estimated by $s^2 = \dfrac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n-2} = \dfrac{SSE}{n-2}$

First define $s\left(\hat{\beta}_1\right) = \dfrac{s}{\sqrt{S_{xx}}}$, $s\left(\hat{\beta}_0\right) = s\sqrt{\dfrac{1}{n} + \dfrac{\bar{x}^2}{S_{xx}}}$ and $s\left(\hat{Y}_i\right) = s\sqrt{\dfrac{1}{n} + \dfrac{\left(x_i - \bar{x}\right)^2}{S_{xx}}}$

Sampling distribution of estimators in simple linear model

Sampling distribution of $\left(\hat{\beta}_1 - \beta_1\right) / s\left(\hat{\beta}_1\right)$ is Student's t with $n-2$ degrees of freedom.

Sampling distribution of $\left(\hat{\beta}_0 - \beta_0\right) / s\left(\hat{\beta}_0\right)$ is Student's t with $n-2$ degrees of freedom.

Sampling distribution of $\left(\hat{Y}_i - EY_i\right) / s\left(\hat{Y}_i\right)$ is Student's t with $n-2$ degrees of freedom.

Confidence interval estimation

Confidence interval for β_0 with confidence coefficient $1-\alpha$

$\hat{\beta}_0 \pm t_{\alpha/2}(n-2)\cdot s\left(\hat{\beta}_0\right)$, and t is based on degree of freedom $n-2$.

Confidence interval for β_1 with confidence coefficient $1-\alpha$

$\hat{\beta}_1 \pm t_{\alpha/2}(n-2)\cdot s\left(\hat{\beta}_1\right)$, and t is based on degree of freedom $n-2$.

Confidence interval for Y_i with confidence coefficient $1-\alpha$

$\hat{Y}_i \pm t_{\alpha/2}\cdot s\left(\hat{Y}_i\right)$, and t is based on degree of freedom $n-2$.

Example

Construct 95% confidence interval for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{Y}$ in salary and GPA example.

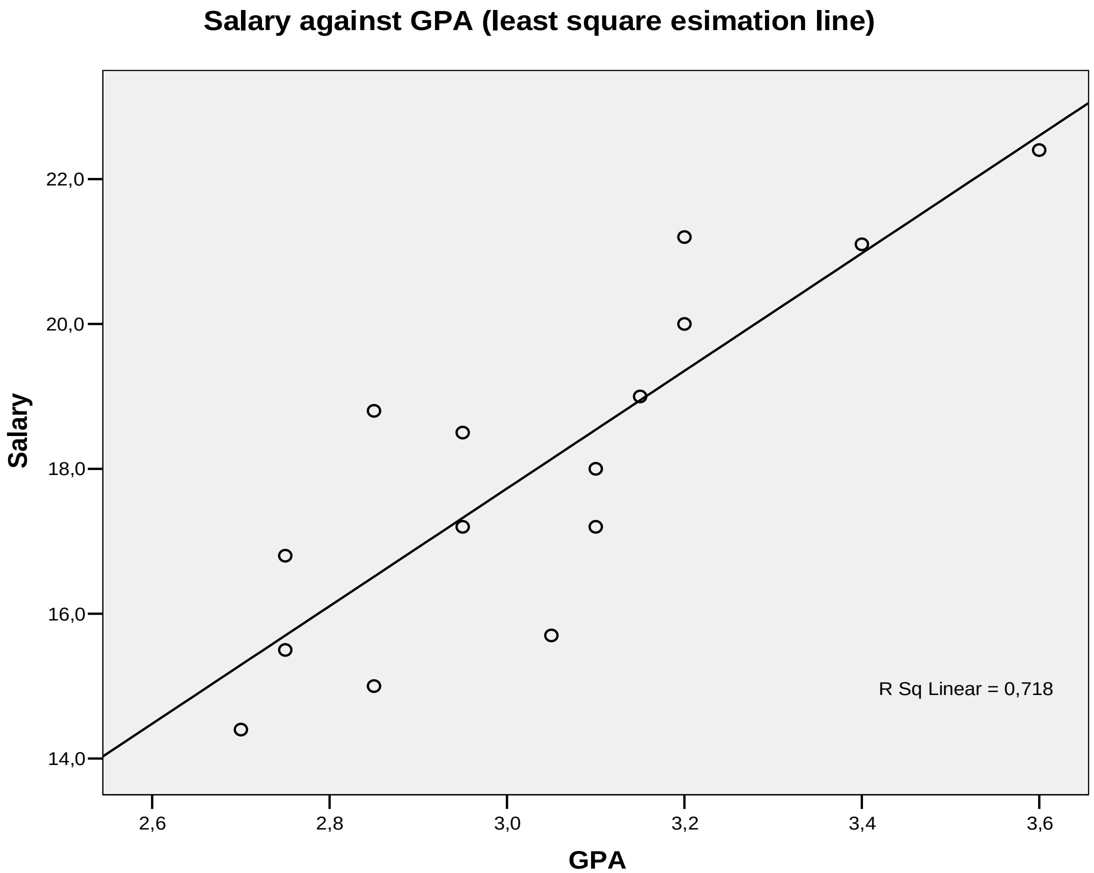
Compare the above results with the SPSS regression coefficients.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-6,627	4,298		-1,542	,147	-15,913	2,659
	GPA	8,119	1,409	,848	5,760	,000	5,074	11,163

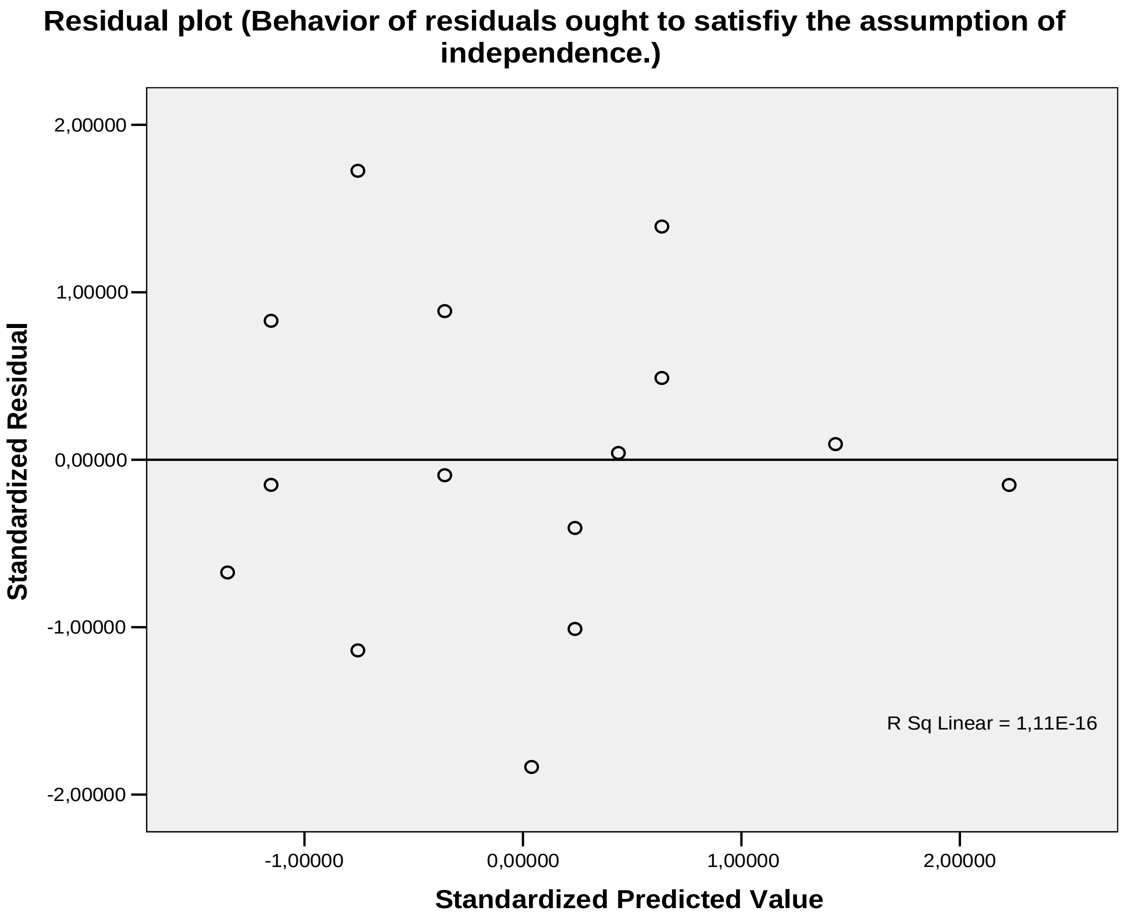
a. Dependent Variable: Salary

Example



Salary against GPA (least square esimation line)
22,0
20,0
18,0
16,0
14,0
Salary
R Sq Linear = 0,718
2,6
2,8
3,0
3,2
3,4
3,6
GPA

Residual analysis



Residual analysis