回歸分析

Linear Regression

盧政良 (Arthur)

研究兩個樣本之間的關係

簡単線性回帰

Simple Linear Regression

$$y = \beta x + \alpha + \varepsilon,$$

$$\varepsilon \sim N(0, \sigma_\varepsilon)$$

$$\mathbf{E}[\, Y = y_i \mid X = x_i] = \beta\, x_i + \alpha$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \hat{\beta} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \hat{\alpha} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

最小平方誤差法

Ordinary Least Square Error (OLS)

$$\varepsilon = Y - (\beta X + \alpha)$$

OLS Regression Results

Dep. Variable:	y	R-squared:	0.903
Model:	OLS	Adj. R-squared:	0.901
Method:	Least Squares	F-statistic:	451.9
Date:	Wed, 11 Mar 2020	Prob (F-statistic):	6.95e-50
Time:	16:41:04	Log-Likelihood:	-137.70
No. Observations:	100	AIC:	281.4
Df Residuals:	97	BIC:	289.2
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.3776	0.286	4.811	0.000	0.809	1.946
x1	-1.3230	1.323	-1.000	0.320	-3.950	1.304
x2	10.9470	1.281	8.549	0.000	8.405	13.489

Omnibus:	1.873	Durbin-Watson:	2.016
Prob(Omnibus):	0.392	Jarque-Bera (JB):	1.638
Skew:	-0.313	Prob(JB):	0.441
Kurtosis:	2.989	Cond. No.	22.5

OLS Regression Results

Dep. Variable:	y	R-squared:	0.903
Model:	OLS	Adj. R-squared:	0.901
Method:	Least Squares	F-statistic:	451.9
Date:	Wed, 11 Mar 2020	Prob (F-statistic):	6.95e-50
Time:	16:41:04	Log-Likelihood:	-137.70
No. Observations:	100	AIC:	281.4
Df Residuals:	97	BIC:	289.2
Df Model:	2		
Covariance Type:	nonrobust		

p-value confidence interval

estimator

模型參數的估計

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.3776	0.286	4.811	0.000	0.809	1.946
x1	-1.3230	1.323	-1.000	0.320	-3.950	1.304
x2	10.9470	1.281	8.549	0.000	8.405	13.489

Omnibus:	1.873	Durbin-Watson:	2.016
Prob(Omnibus):	0.392	Jarque-Bera (JB):	1.638
Skew:	-0.313	Prob(JB):	0.441
Kurtosis:	2.989	Cond. No.	22.5

OLS Regression Results

Dep. Variable:	y	R-squared:	0.903
Model:	OLS	Adj. R-squared:	0.901
Method:	Least Squares	F-statistic:	451.9
Date:	Wed, 11 Mar 2020	Prob (F-statistic):	6.95e-50
Time:	16:41:04	Log-Likelihood:	-137.70
No. Observations:	100	AIC:	281.4
Df Residuals:	97	BIC:	289.2
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.3776	0.286	4.811	0.000	0.809	1.946
x1	-1.3230	1.323	-1.000	0.320	-3.950	1.304
x2	10.9470	1.281	8.549	0.000	8.405	13.489

Omnibus:	1.873	Durbin-Watson:	2.016
Prob(Omnibus):	0.392	Jarque-Bera (JB):	1.638
Skew:	-0.313	Prob(JB):	0.441
Kurtosis:	2.989	Cond. No.	22.5

檢查殘差的分佈

Durbin-Watson: 自我相關性統計量

Jarque-Bera (JB): 偏態/峰態統計量

https://en.wikipedia.org/wiki/Coefficient_of_determination

https://en.wikipedia.org/wiki/Akaike_information_criterion

https://en.wikipedia.org/wiki/Bayesian_information_criterion

OLS Regression Results

Dep. Variable:	y	R-squared:	0.903
Model:	OLS	Adj. R-squared:	0.901
Method:	Least Squares	F-statistic:	451.9
Date:	Wed, 11 Mar 2020	Prob (F-statistic):	6.95e-50
Time:	16:41:04	Log-Likelihood:	-137.70
No. Observations:	100	AIC:	281.4
Df Residuals:	97	BIC:	289.2
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.3776	0.286	4.811	0.000	0.809	1.946
x1	-1.3230	1.323	-1.000	0.320	-3.950	1.304
x2	10.9470	1.281	8.549	0.000	8.405	13.489

Omnibus:	1.873	Durbin-Watson:	2.016
Prob(Omnibus):	0.392	Jarque-Bera (JB):	1.638
Skew:	-0.313	Prob(JB):	0.441
Kurtosis:	2.989	Cond. No.	22.5

最大概似估計

Maximum Likelihood Estimation (MLE)

https://en.wikipedia.org/wiki/Maximum_likelihood_estimation