

一元线性回归

关键概念

$$y_i = \alpha + \beta x_i + \epsilon_i$$

- 残差 (residual) : $e_i \equiv y_i - \hat{\alpha} - \hat{\beta}x_i$
- 平方和分解公式 $TSS = ESS + RSS$: $\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y})^2 + \sum_{i=1}^n e_i^2$
- 拟合优度: $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$

Stata 代码解析

```
1 | list s lnw in 1/10
```

list: 列出指定变量的值

s 和 lnw: 选定的变量

in 1/10: 只显示数据集中的第 1 行到第 10 行的数据

```
1 | twoway scatter lnw s || lfit lnw s
```

twoway: 告诉 stata 开始绘图

scatter lnw s: 绘制 lnw 和 s 两个变量的散点图

||: 散点图上加一个图层

lfit lnw s: 拟合 lnw 和 s 之间的线性回归直线

在 Stata 中, `lfit` 命令通常需要与 `scatter` 命令一起使用, 因为 `lfit` 命令是在散点图的基础上进行回归线的拟合。在单独使用 `lfit` 命令时, Stata 会提示错误, 因为没有提供需要拟合的数据点。因此, 为了正确地使用 `lfit` 命令, 通常需要先通过 `scatter` 命令绘制散点图, 然后在此基础上使用 `lfit` 命令来拟合回归线。这种做法能够更清晰地展示数据的分布情况和回归线的拟合效果。

```
1 | reg y x, noc
```

$$y_i = \alpha + \beta x_i + \epsilon_i$$

一元线性回归，y 是解释变量，x 是被解释变量

noc 参数可选，表示无常数项： $y_i = \beta x_i + \epsilon_i$

何时可能用到无常数项回归？

例如比较个人所得税对收入的依赖性，收入为 0，个人所得税自然也为 0。

Source	SS	df	MS	Number of obs	=	758
Model	35.2039946 ESS	1	35.2039946	F(1, 756)	=	255.70
Residual	104.082155 RSS	756	.137674809	Prob > F	=	0.0000
				R-squared	=	0.2527
				Adj R-squared	=	0.2518
Total	139.28615 TSS	757	.183997556	Root MSE	=	.37105

lnw	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
s	.0966245	.0060425	15.99	0.000	.0847624	.1084866
常数项 _cons	4.391486	.0821136	53.48	0.000	4.230288	4.552684

```
1 | sum s // 单变量特征值
2 | return list // 针对上一条命令生成数据列表
3 |
4 | // e 类命令：估计命令
5 | reg lnw s // 执行了一个估计命令 reg
6 | ereturn list // 针对上一条命令生成数据列表，需要用 ereturn
```