

# 线性回归

王媛媛经济系硕士 15320171151909

2019 年 3 月 27 日

## 对数线性回归（R）---真实数据的一个例子

为探究教育水平对城市生产总值的影响，对二者进行计量模型回归分析。本文的数据来源于国家统计局年鉴 2016 年的城市年度截面数据，选择国内主要的 37 个城市的国内生产总值 (亿元)、普通高等学校在校学生数 (万人)、年末总人口 (万人) 等变量，以普通高等学校在校学生数变量来表示城市的教育水平，并根据城市所在省手动处理将数据分为东部、西部和中部。在整理后的数据中，city 表示城市 ID，GDP 表示该城市的国内生产总值，edu 表示教育水平，population 表示人口规模，loc 表示东西中部的划分。<sup>1</sup>

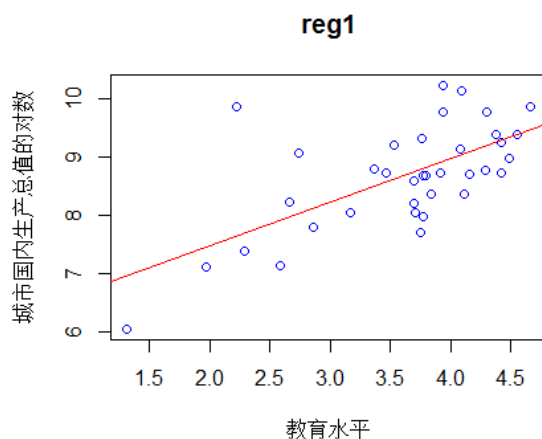
### 0.0.1 基本模型回归

$$\log(\text{GDP}) = \alpha + \beta * \log(\text{edu}) + e$$

```
> library(readxl)
> hw2 <- read_excel("D:/研二/微观计量/数据/2.0.xlsx")
> View(hw2)
> require(AER)
> reg1=lm(log(GDP)~log(edu),data=hw2)
> coeftest(reg1)

t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.99922    0.55254  10.8575 1.361e-12 ***
log(edu)      0.74547    0.14991   4.9728 1.863e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



可以从回归结果中看到教育水平总体上是对城市的国内生产总值有着非常显著的正向影响。

<sup>1</sup>数据来源：国家统计局《中国统计年鉴 2016 年主要城市数据》

### 0.0.2 分地区的模型回归

$$\log(\text{GDP}) = \alpha_{loc} + \beta_{loc} * \log(\text{edu}) + e$$

```
> reg2=lm(log(GDP)~log(edu)*loc,data=hw2)
> coeftest(reg2)

t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.54659    3.58167   0.9902  0.3300
log(edu)        1.25918    0.86071   1.4630  0.1539
loceast         3.94679    3.66803   1.0760  0.2905
locwest         1.54728    3.63482   0.4257  0.6734
log(edu):loceast -0.82352    0.88724  -0.9282  0.3607
log(edu):locwest -0.36130    0.87930  -0.4109  0.6841
```

将城市划分为东部、西部、中部后，可以从回归结果中发现，东部（east）的国内生产总值整体要高于西部地区（west），这也与经济形势的现状相符合。但是当划分了地区后，教育水平对城市国内生产总值的影响是不显著的，没有足够证据说明教育会影响城市国内生产总值。

### 0.0.3 控制人口规模后的回归分析

$$\log(\text{GDP}) = \alpha + \beta_1 * \log(\text{edu}) + \beta_2 * \text{population} + e$$

```
> reg3=lm(log(GDP)~log(edu)+population,data=hw2)
> coeftest(reg3)

t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.41921737  0.52271742  12.2805 7.466e-14 ***
log(edu)      0.49825819  0.16112910   3.0923  0.004022 **
population    0.00064438  0.00022467   2.8681  0.007145 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> reg4=lm(log(GDP)~log(edu)*loc+population,data=hw2)
> coeftest(reg4)

t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.33468842  3.11484569   1.3916 0.174620
log(edu)      0.96776173  0.75152402   1.2877 0.208021
loceast       3.43055725  3.18440741   1.0773 0.290227
locwest       1.42235000  3.15199342   0.4513 0.655163
population     0.00063581  0.00019258   3.3016 0.002556 **
log(edu):loceast -0.74104778  0.76973883  -0.9627 0.343644
log(edu):locwest -0.40844023  0.76257950  -0.5356 0.596317
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

从 reg3 回归中可以看到，加入了人口变量控制了城市规模后，教育水平对城市国内生产总值的影响没有基本回归中那么显著了，而分地区回归的结果 reg4 则表明不论是东部、西部还是中部，在划分了地区后，教育水平对城市生产总值的影响都不明显，只有人口规模会对其产生显著影响。