

附录——一元线性回归演示

2023 年 9 月 18 日

一元线性回归动态演示（请自上而下运行一遍）

某公司计划研究销售人员数量对于新产品销售额的影响。

从其下属多家公司中随机抽取 10 个子公司，这 10 个子公司当年新产品销售额和销售人员数量统计数据在路径“数据/销售人员和销售量.csv”。

下面的代码演示了，通过调节斜率和截距这两个参数，可以观察 R^2 和 SER 这个评价回归效果的指标的变化。

读取数据

```
[2]: import pandas as pd
data = pd.read_csv("数据/销售人员和销售量.csv")
x = data["销售人员数量/人"]
y = data["新产品销售额/万元"]
```

回归参数估计

```
[3]: import scipy.stats as stats
result = stats.linregress(x, y)
result
```

```
[3]: LinregressResult(slope=12.230986303255557, intercept=176.2952026980522,
rvalue=0.969906207108702, pvalue=3.4603114770030177e-06,
stderr=1.0855453862397284, intercept_stderr=27.326866769281782)
```

计算 R2 和 SER 的函数

```
[4]: def SER(y_test, y_pred):
      return np.sqrt(np.mean((y_test - y_pred) ** 2))

def R2(y_test, y_pred):
    #print(y_test)
    TSS=np.sum((y_test-np.mean(y_test))**2)
    SSR=np.sum((y_test-y_pred)**2)
    ESS = np.sum((y_pred-np.mean(y_test))**2)
    r2=1-SSR/TSS
    return r2
```

使用 ipywidgets 来动态绘图

```
[6]: %matplotlib inline
from ipywidgets import interactive
import matplotlib.pyplot as plt
import numpy as np
import scipy.stats as stats

R2_record = []
SER_record = []
def regress(slope=result.slope, intercept=result.intercept):
    plt.figure(figsize=(10, 6))
    plt.scatter(x, y, marker='o',c='b', label='original data')
    plt.plot(x, intercept + slope*x, c='r', label='fitted line')
    R2_record.append(R2(y, intercept + slope*x))
    SER_record.append(SER(y, intercept + slope*x))
    print("拟合优度 R2[接近 1 最好] 目前值: %.2f" %(R2_record[-1]))
    print("SER[越小越好] 目前值: %.2f" %SER_record[-1])
    plt.legend()
    plt.show

    plt.figure(figsize=(10, 1.5))
    f = plt.plot(R2_record, label="R2")
```

```

plt.ylabel('R2')
plt.legend()

plt.figure(figsize=(10, 1.5))
plt.plot(SER_record, label="SER")
plt.ylabel('SER')
plt.legend()
plt.show()

interactive_plot = interactive(regress, slope=(0, 2*result.slope, 0.2),
    ↪intercept=(0, result.intercept*2, 5))
output = interactive_plot.children[-1]
output.layout.height = '650px'
interactive_plot

```

```

interactive(children=(FloatSlider(value=12.230986303255557, description='slope',
    ↪max=24.461972606511114, step=...

```

[]:

[]: