窗体顶端

|  |  |
| --- | --- |
|  | 第4章第2题 双因素方差分析 |
|  | 某同类产品有三种型号（X、Y、Z）在两个地区（A、B）的月销售数量如下表所示。   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | A地区 | | | | | B地区 | | | | | | X | 28 | 22 | 31 | 19 | 26 | 30 | 25 | 26 | 21 | 23 | | Y | 33 | 27 | 30 | 24 | 31 | 35 | 31 | 31 | 25 | 33 | | Z | 23 | 16 | 17 | 26 | 18 | 21 | 26 | 25 | 19 | 17 |   请利用双因素方差分析，分析产品销量与控制因素“型号”和“地区”之间的作用关系。 |
| Python | import pandas as pd  import numpy as np  d = np.array([  [28, 22, 31, 19, 26, 30, 25, 26, 21, 23],  [33, 27, 30, 24, 31, 35, 31, 31, 25, 33],  [23, 16, 17, 26, 18, 21, 26, 25, 19, 17]  ])  df = pd.DataFrame(d)  df.index=pd.Index(['X','Y','Z'],name='型号')  df.columns=pd.Index(['A','A','A','A','A','B','B','B','B','B'],name='地区')  df  df1 = df.stack().reset\_index().rename(columns={0:'销量'})  df1  from statsmodels.formula.api import ols  from statsmodels.stats.anova import anova\_lm  model = ols('销量~C(型号) + C(地区)+C(型号):C(地区)', df1).fit()  anova\_lm(model)  print(anova\_lm(model)) |
| C++ |  |
| 输出 | df sum\_sq mean\_sq F PR(>F)  C(型号) 2.0 423.800000 211.900000 13.525532 0.000117  C(地区) 1.0 9.633333 9.633333 0.614894 0.440622  C(型号):C(地区) 2.0 6.866667 3.433333 0.219149 0.804792  Residual 24.0 376.000000 15.666667 NaN NaN |
| 书籍 |  |
| 软件 | Python，C++（附加orsci包）。 |

窗体底端