地理建模实验4 实验报告

42109232　吕文博　地信2101班

2024-05-23

R 语言多元线性回归

加载数据

```
dt = readxl::read_xlsx('../data/exp4/4.xlsx')
skimr::skim(dt)
```

表 1: Data summary

Name	dt
Number of rows	52
Number of columns	4
Column type frequency:	
character	1
numeric	3
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
台站	0	1	2	5	0	52	0

Variable type: numeric

skim_variable	mean	sd	p25	p50	p75	hist
年降水量P/mm	372.20	215.17	159.01	420.52	542.56	▆▂▃▇▂
纬度坐标Y（北纬0°）	36.75	2.38	34.93	35.79	38.80	▃▇▂▃▂
海拔Z/m	1756.98	608.32	1343.00	1560.90	2036.92	▇▇▂▁▁

拟合多元线性回归模型

```
lm.model = lm(`年降水量P/mm` ~ `纬度坐标Y（北纬0°）` + `海拔Z/m`,
              data = dt)
summary(lm.model)
```

```
##
## Call:
## lm(formula = `年降水量P/mm` ~ `纬度坐标Y（北纬0°）` +
##     `海拔Z/m`, data = dt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -188.63  -57.85  -12.86   40.47  178.58
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)           3260.83143  200.51255  16.262   <2e-16 ***
## `纬度坐标Y（北纬0°）`  -80.62259    5.24426 -15.373   <2e-16 ***
## `海拔Z/m`                0.04235    0.02055   2.061   0.0446 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 88.68 on 49 degrees of freedom
## Multiple R-squared:  0.8368, Adjusted R-squared:  0.8301
## F-statistic: 125.6 on 2 and 49 DF,  p-value: < 2.2e-16
```

*R*方为*0.8368*,说明数据有*83.68%*的可能被该回归方程解释,数据与模型的拟合程度较高.

由回归方程的*F*检验可得,$P < 2.2e-16 << 0.05$, *F*检验通过,说明回归方程显著,自变量能显著影响因变量.

回归模型的方程为 $年降水量 = -80.62259 \times 纬度坐标 + 0.04235 \times 海拔 + 3260.83143$,对自变量进行*t*检

验,$P_{\text{纬度坐标}} < 2e-16 << 0.05$且$P_{\text{海拔}} = 0.0446 < 0.05$,纬度坐标和海拔的$t$检验都通过,说明两个自变量能显著影响因变量.

Python 多元线性回归

```
import numpy as np
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt

dt = pd.read_excel('../data/exp4/4.xlsx')
dt.head()
```

```
##    台站  年降水量P/mm  纬度坐标Y（北纬0°）   海拔Z/m
## 0  安西     48.25    40.500000  1170.8
## 1  白银    193.72    36.599998  1707.2
## 2  定西    413.94    35.533000  1908.8
## 3  古浪    358.60    37.483003  2072.4
## 4  和政    615.04    35.432998  2136.4
```

```
X = dt.loc[:,['纬度坐标Y（北纬0°）','海拔Z/m']]
y = dt.loc[:,'年降水量P/mm']
X = sm.add_constant (X)
lm_model = sm.OLS(y,X).fit()
print(lm_model.summary())
```

```
##                             OLS Regression Results
## ==============================================================================
## Dep. Variable:               年降水量P/mm   R-squared:                       0.837
## Model:                            OLS   Adj. R-squared:                  0.830
## Method:                 Least Squares   F-statistic:                     125.6
## Date:                周四, 23 5月 2024   Prob (F-statistic):           5.16e-20
## Time:                        15:06:23   Log-Likelihood:                -305.46
## No. Observations:                  52   AIC:                             616.9
## Df Residuals:                      49   BIC:                             622.8
## Df Model:                           2
## Covariance Type:            nonrobust
## ==============================================================================
##                  coef    std err          t      P>|t|      [0.025      0.975]
## ------------------------------------------------------------------------------
```

```
## const         3260.8314     200.513      16.262       0.000     2857.886     3663.776
## 纬度坐标Y（北纬0°）    -80.6226       5.244    -15.373       0.000      -91.161      -70.084
## 海拔Z/m            0.0424       0.021       2.061       0.045        0.001        0.084
## ==============================================================================
## Omnibus:                        1.609   Durbin-Watson:                   1.438
## Prob(Omnibus):                  0.447   Jarque-Bera (JB):                1.588
## Skew:                           0.358   Prob(JB):                        0.452
## Kurtosis:                       2.530   Cond. No.                     3.03e+04
## ==============================================================================
##
## Notes:
## [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
## [2] The condition number is large, 3.03e+04. This might indicate that there are
## strong multicollinearity or other numerical problems.
```

*R*方为*0.837*,说明数据有*83.7%*的可能被该回归方程解释,数据与模型的拟合程度较高.

由回归方程的*F*检验可得,$P = 5.16e-20 << 0.05$, *F*检验通过,说明回归方程显著,自变量能显著影响因变量.

回归模型的方程为 年降水量 $= -80.6226 \times$ 纬度坐标 $+ 0.0424 \times$ 海拔 $+ 3260.8314$,对自变量进行*t*检验,$P_{\text{纬度坐标}} = 0 << 0.05$且$P_{\text{海拔}} = 0.045 < 0.05$,纬度坐标和海拔的*t*检验都通过,说明两个自变量能显著影响因变量.