地理建模实验3 实验报告

42109232　吕文博　地信2101班

2024-05-16

R 语言拟合一元线性回归模型

加载数据

```
dt = readxl::read_xls('../data/exp3/3-SPSS.xls')
head(dt)
```

```
## # A tibble: 6 x 2
##   `平均气温x/oC` `降雨量y/mm`
##            <dbl>        <dbl>
## 1            3.8         77.7
## 2            4           51.2
## 3            5.8         60.1
## 4            8           54.1
## 5           11.3         55.4
## 6           14.4         56.8
```

拟合一元线性回归模型

```
lm.model = lm(`降雨量y/mm` ~ `平均气温x/oC`,data = dt)
summary(lm.model)
```

```
##
## Call:
## lm(formula = `降雨量y/mm` ~ `平均气温x/oC`, data = dt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.3225  -7.8229   0.0991   9.8908  11.9445
##
```

```
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      74.3266     7.2342  10.274 1.24e-06 ***
## `平均气温x/oC`  -1.2010     0.6766  -1.775    0.106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.71 on 10 degrees of freedom
## Multiple R-squared:  0.2396, Adjusted R-squared:  0.1636
## F-statistic: 3.151 on 1 and 10 DF,  p-value: 0.1063
```

R方为0.2396,说明数据有23.96%的可能被该回归方程解释,数据与模型的拟合程度较低.

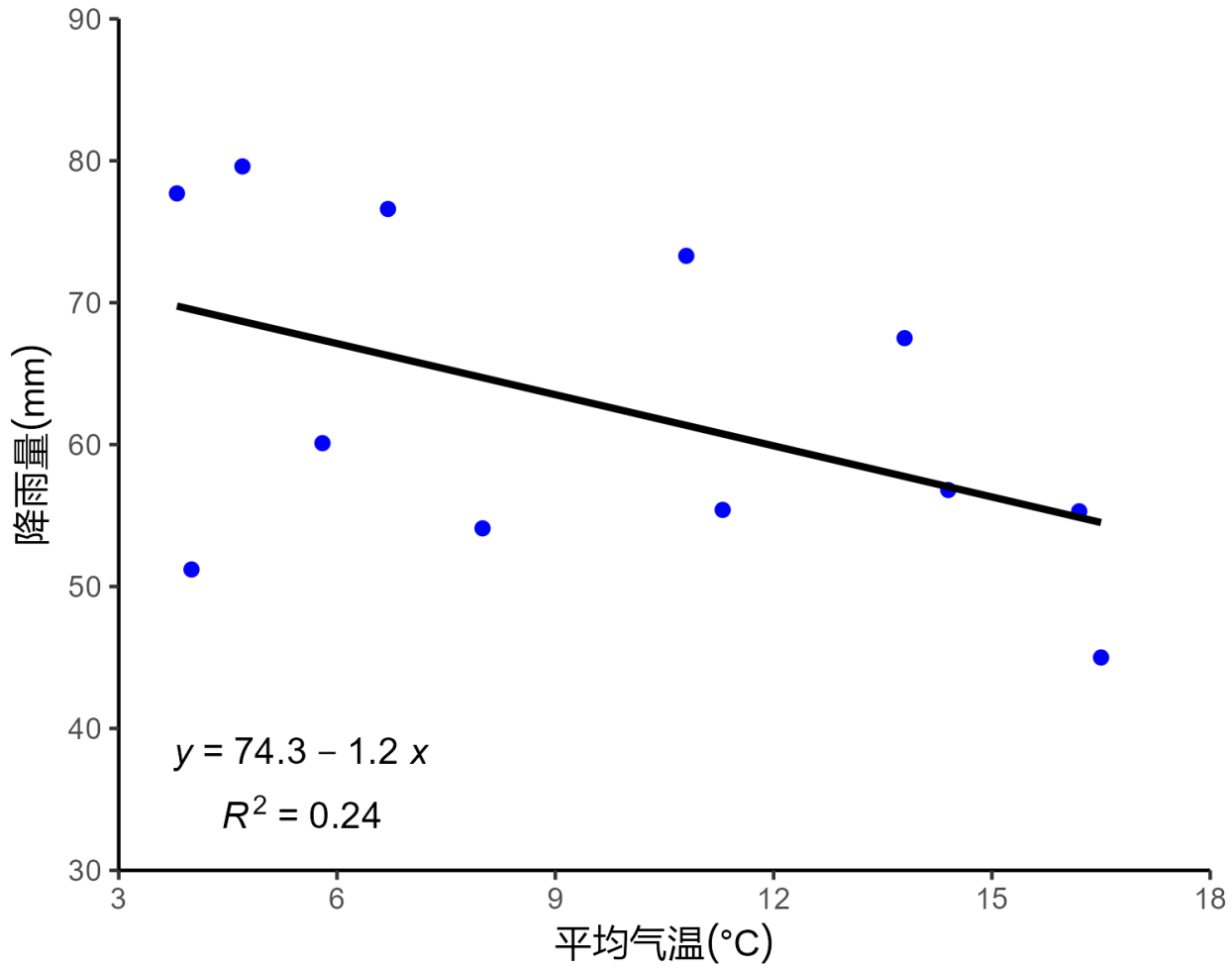
由回归方程的F检验可得,$P = 0.1063 > 0.05$, F检验无法通过,说明回归方程不显著,自变量不能显著影响因变量.

回归模型的方程为 $y = -1.201x + 74.3266$,对自变量进行t检验,$P_{(x)} = 0.106 > 0.05$,t检验无法通过,说明自变量不能显著影响因变量.

一元线性回归拟合图表绘制

```
library(ggplot2)

ggplot(data = dt,
       aes(x = `平均气温x/oC`,
           y = `降雨量y/mm`)) +
  geom_point(shape = 20,size = 2.5,color = 'blue') +
  geom_smooth(method = 'lm',se = F,color = 'black') +
  scale_x_continuous(name = latex2exp::TeX("$平均气温 (\\degree C)$"),,
                     expand = c(0,0),limits = c(3,18)) +
  scale_y_continuous(name = latex2exp::TeX("$降雨量 (mm)$"),
                     expand = c(0,0),limits = c(30,90)) +
  ggpmisc::stat_poly_eq(aes(
    label = paste0("atop(", after_stat(eq.label),
                   ",",
                   after_stat(rr.label), ")")),
    formula = y ~ x,
    label.x = "left",
    label.y = "bottom",
    parse = TRUE) +
  theme_classic()
```



Python 拟合一元线性回归模型

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

dt = pd.read_excel('../data/exp3/3-SPSS.xls')
dt.head()
```

```
##    平均气温x/oC  降雨量y/mm
## 0       3.8     77.7
## 1       4.0     51.2
## 2       5.8     60.1
## 3       8.0     54.1
```

```
## 4      11.3     55.4
```

```
X = dt.loc[:,'平均气温x/oC']
y = dt.loc[:,'降雨量y/mm']
X = sm.add_constant (X)
lm_model = sm.OLS(y,X).fit()
print(lm_model.summary())
```

```
##                             OLS Regression Results
## ==============================================================================
## Dep. Variable:                降雨量y/mm   R-squared:                       0.240
## Model:                            OLS   Adj. R-squared:                  0.164
## Method:                 Least Squares   F-statistic:                     3.151
## Date:                周四, 16 5月 2024   Prob (F-statistic):              0.106
## Time:                        16:15:15   Log-Likelihood:                -44.387
## No. Observations:                  12   AIC:                             92.77
## Df Residuals:                      10   BIC:                             93.74
## Df Model:                           1
## Covariance Type:            nonrobust
## ==============================================================================
##                  coef    std err          t      P>|t|      [0.025      0.975]
## ------------------------------------------------------------------------------
## const         74.3266      7.234     10.274      0.000      58.208      90.445
## 平均气温x/oC      -1.2010      0.677     -1.775      0.106      -2.709       0.307
## ==============================================================================
## Omnibus:                        1.571   Durbin-Watson:                   1.008
## Prob(Omnibus):                  0.456   Jarque-Bera (JB):                0.898
## Skew:                          -0.287   Prob(JB):                        0.638
## Kurtosis:                       1.789   Cond. No.                         25.2
## ==============================================================================
##
## Notes:
## [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

*R*方为*0.2396*,说明数据有*23.96%*的可能被该回归方程解释,数据与模型的拟合程度较低.

由回归方程的*F*检验可得,$P = 0.1063 > 0.05$, *F*检验无法通过,说明回归方程不显著,自变量不能显著影响因变量.

回归模型的方程为 $y = -1.201x + 74.3266$,对自变量进行*t*检验,$P_{(x)} = 0.106 > 0.05$,*t*检验无法通过,说明自变量不能显著影响因变量.