第四章

##### 10193903446 汤博

代码可以通过[Quantitative-Geography/practice2 at practice2 · Plutoset/Quantitative-Geography (github.com)](https://github.com/Plutoset/Quantitative-Geography/tree/practice2/practice2)访问。

1. 根据提供的数据，计算四个变量所有的二级偏相关系数与所有有三个自变量的复相关系数，并进行显著性检验。
   * **计算二级偏相关系数**

* clear; clc;  
    
  % open dataset  
  table = readtable("课后作业数据.xls","VariableNamingRule","preserve");  
  Y = table.("农民人均纯收入y/(元·人-1)");  
  X1 = table.("人均粮食产量x1 (kg·人-1)");  
  X2 = table.("经济作物占农作物播面比例x2/％");  
  X3 = table.("人均耕地面积x3/hm2)");  
  arr = table2array(table);  
  % calculate the XY correlation coefficient  
  CC = corrcoef([Y, X1, X2, X3]);

| * CC | * R\_1 | * R\_2 | * R\_3 | * R\_4 |
| --- | --- | --- | --- | --- |
| * **R1\_** | * 1 | * 0.3825 | * 0.0690 | * 0.6441 |
| * **R2\_** | * 0.3825 | * 1 | * 0.7342 | * 0.4197 |
| * **R3\_** | * 0.0690 | * 0.7342 | * 1 | * 0..2549 |
| * **R4\_** | * 0.6441 | * 0.4197 | * 0.2549 | * 1 |

* % calculate First-order partial correlation coefficient  
    
  fun = @(data, x, y, z) ...  
   (data(x, y)-data(x, z)\*data(y, z)) ...  
   / sqrt( (1 - data(x, z) ^ 2) \* (1 - data(y, z) ^ 2) );  
  R12\_3 = fun(CC, 1, 2, 3);  
  R13\_2 = fun(CC, 1, 3, 2);  
  R13\_4 = fun(CC, 1, 3, 4);  
  R14\_2 = fun(CC, 1, 4, 2);  
  R14\_3 = fun(CC, 1, 4, 3);  
  R23\_1 = fun(CC, 2, 3, 1);  
  R24\_1 = fun(CC, 2, 4, 1);  
  R24\_3 = fun(CC, 2, 4, 3);  
  R34\_1 = fun(CC, 3, 4, 1);  
  R34\_2 = fun(CC, 3, 4, 2);
* **一级偏相关系数：**

| * R12\_3 | * R13\_2 | * R14\_2 | * R14\_2 | * R14\_3 |
| --- | --- | --- | --- | --- |
| * 0.4900 | * -0.3377 | * -0.1287 | * 0.5767 | * 0.6495 |

| * R23\_1 | * R24\_1 | * R24\_3 | * R34\_1 | * R34\_2 |
| --- | --- | --- | --- | --- |
| * 0.7679 | * 0.2452 | * 0.3542 | * 0.2758 | * -0.0864 |

* % calculate Second-order partial correlation coefficient   
  ffun = @(data, a, b, c, d) ...   
   ( fun(data, a, b, c) - ( fun(data, a, d, c) \* ( fun(data, b, d, c) ) ) ...   
   / sqrt( (1-fun(data, a, d, c)) \* (1-fun(data, b, d, c)) ) )   
     
  R12\_34 = ffun(CC,1,2,3,4);   
  R13\_24 = ffun(CC,1,3,2,4);   
  R14\_23 = ffun(CC,1,4,2,3);   
  R23\_14 = ffun(CC,2,3,1,4);   
  R24\_13 = ffun(CC,2,4,1,3);   
  R34\_12 = ffun(CC,3,4,1,2);
* **二级偏相关系数：**

| * R12\_34 | * R13\_24 | * R14\_23 | * R23\_14 | * R24\_13 | * R34\_12 |
| --- | --- | --- | --- | --- | --- |
| * 0.3655 | * -0.3537 | * 0.5838 | * 0.7515 | * 0.0543 | * 0.1409 |

* % t-student significance test   
  t\_test = @(data, m, n) data \* sqrt(n - m - 1) / sqrt(1 - data ^ 2);   
  m = 2;   
  n = length(arr);   
  t12\_34 = t\_test(R12\_34, m, n);   
  t13\_24 = t\_test(R13\_24, m, n);   
  t14\_23 = t\_test(R14\_23, m, n);   
  t23\_14 = t\_test(R23\_14, m, n);   
  t24\_13 = t\_test(R24\_13, m, n);   
  t34\_12 = t\_test(R34\_12, m, n);   
  % the degree of freedom is 21 - 2 - 1 = 18

| * t12\_34 | * t13\_24 | * t14\_23 | * t23\_14 | * t24\_13 | * t34\_12 |
| --- | --- | --- | --- | --- | --- |
| * 1.6659 | * -1.6044 | * 3.0509 | * 4.8327 | * 0.2307 | * 0.6038 |

* 自由度为 21 - 2 - 1 = 18 时，查 *t* 分布表如下：

| 双尾 \ α | 0.5 | 0.2 | 0.1 | 0.05 | 0.02 | 0.01 | 0.002 | 0.001 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| n = 18 | 0.688 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 | 3.922 |

* 在置信水平 α = 0.001 上，R23\_14 是显著的；
* 在置信水平 α = 0.01 上，R14\_23, R23\_14 是显著的。
  + **计算复相关系数**

因为计算复相关系数的代码较为复杂，重新定义函数(存储在mcc.m中)：

function Ry\_123 = mmc(CC, ny, n1, n2, n3)   
% calculate multiple correlation coefficient   
% ------ INPUT ------   
% CC(4\*4 double): correlation coefficient matrix   
% ny(int): the index for y   
% n1(int): the index for x1   
% n2(int): the index for x2   
% n3(int): the index for x3   
% ------ OUTPUT ------   
% Ry\_123: the multiple correlation coefficient between y and x1, x2, x3   
   
% calculate First-order partial correlation coefficient   
   
 fun = @(data, x, y, z) ...   
 (data(x, y)-data(x, z)\*data(y, z)) ...   
 / sqrt( (1 - data(x, z) ^ 2) \* (1 - data(y, z) ^ 2) );   
   
 % calculate Second-order partial correlation coefficient   
 ffun = @(data, a, b, c, d) ...   
 ( (fun(data, a, b, c) - ( fun(data, a, d, c) \* ( fun(data, b, d, c) ) )) ...   
 / sqrt( (1 - fun(data, a, d, c) ^ 2) \* (1 - fun(data, b, d, c) ^ 2) ) );   
   
 Ry1 = CC(ny, n1);   
 Ry2\_1 = fun(CC, ny, n2, n1);   
 Ry3\_12 = ffun(CC, ny, n3, n1, n2);   
 Ry\_123 = sqrt(1 - (1 - Ry1 ^ 2) \* (1 - Ry2\_1 ^ 2) \* (1 - Ry3\_12 ^ 2));   
end

应用函数，分别计算所有有三个自变量的复相关系数

% calculate multiple correlation coefficient   
Ry\_123 = mmc(CC, 1, 2, 3, 4);   
R1\_y23 = mmc(CC, 2, 1, 3, 4);   
R2\_y13 = mmc(CC, 3, 1, 2, 4);   
R3\_y12 = mmc(CC, 4, 1, 2, 3);

| R1\_y23 | R2\_y13 | R3\_y12 | Ry\_123 |
| --- | --- | --- | --- |
| 0.8067 | 0.7744 | 0.6790 | 0.7081 |

F\_test = @(Mcc, n, k) (Mcc ^ 2 \* (n - k - 1)) / ((1 - Mcc ^ 2) \* k);   
   
n = length(arr);   
k = 3;   
Fy\_123 = F\_test(Ry\_123, n, k);   
F1\_y23 = F\_test(R1\_y23, n, k);   
F2\_y13 = F\_test(R2\_y13, n, k);   
F3\_y12 = F\_test(R3\_y12, n, k);

| t1\_y23 | t2\_y13 | t3\_y12 | ty\_123 |
| --- | --- | --- | --- |
| 10.5595 | 8.4918 | 4.8471 | 5.7002 |

自由度为 21 - 3 - 1 = 17 时，查 *t* 分布表如下：

| 双尾 \ α | 0.5 | 0.2 | 0.1 | 0.05 | 0.02 | 0.01 | 0.002 | 0.001 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| n = 17 | 0.689 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 | 3.965 |

在置信水平 α = 0.001 上，R1\_y23, R2\_y13, R3\_y12, Ry\_123 都是显著的。

1. 根据提供数据，拟合y与x1, x2, x3 的多元线性回归方程，并进行显著性检验。

根据课本上的代码进行拟合

% formulas according to textbook   
n = length(arr);   
X = ones(n,1);   
X = [X,X1,X2,X3];   
k = 3;   
b = X' \* X \ (X' \* Y);   
% F-student significance test   
L\_yy = sum((Y-mean(Y)) .^ 2);   
L\_ny = (Y-mean(Y))' \* [X1,X2,X3]-mean([X1,X2,X3]);   
U = sum(b(2:end)' .\* L\_ny);   
Q = L\_yy - U;   
F = (U / 3) / (Q / (n - k - 1))

使用matlab自带模型

% use built-in models   
mdl = fitlm([X1,X2,X3],Y)   
es = mdl.Coefficients.Estimate(2:end);   
p = mdl.Coefficients.pValue(2:end);

mdl =

线性回归模型:  
 y ~ 1 + x1 + x2 + x3

| 估计系数: | Estimate | SE | tStat | pValue |
| --- | --- | --- | --- | --- |
| (Intercept) | 367.26 | 214.22 | 1.7144 | 0.10462 |
| x1 | 1.7808 | 1.0999 | 1.619 | 0.12385 |
| x2 | -17.312 | 11.103 | -1.5592 | 0.13737 |
| x3 | 524.54 | 176.91 | 2.965 | 0.0086785 |

观测值数目: 21，误差自由度: 17  
均方根误差: 311  
R 方: 0.501，调整 R 方 0.414  
F 统计量(常量模型): 5.7，p 值 = 0.00688

拟合多元线性回归的结果：

参数：

| b0 | b1 | b2 | b3 |
| --- | --- | --- | --- |
| 367.2575 | 1.7808 | -17.3120 | 524.5408 |

统计量***F***: 5.6954, p 值: 0.00688

所以该模型在置信水平α=0.01上，该线性回归模型是显著的。