HW5

# Wei Ye

## Date: Jun 25, 2022

### Question1

A=[[1:10]',[1 1 0 0 0 1 1 1 0 1]',[100 95 125 140 110 85 110 100 125 105]',[4 2 4 7 4 2 4 4 6 4]']

a. Using for-loops to create two vectors, one containing the salaries of only maile(sal\_m), the other for the salaries of only females(sal\_f)

sal\_m=[];

sal\_f=[];

for i =1:size(A,1)

if A(i,2)==0

sal\_m=[sal\_m;A(i,3)]

end

end

for i =1:size(A,1)

if A(i,2)==1

sal\_f=[sal\_f;A(i,3)]

end

end

b. using for loops to creat 2 matrices that have in their first column as id of the individuals and in the second column as the corresponding salary. Do this for male(id\_sal\_m) and females(id\_sal\_f)

id\_sal\_m=[];

id\_sal\_f=[];

for i =1:size(A,1)

if A(i,2)==0

id\_sal\_m=[sal\_m;A(i,3)]

end

end

for j=1:size(A,1)

if A(j,2)==1

id\_sal\_f=[id\_sal\_f;A(j,[1,3])]

end

end

c. Separate the data into males and females and compute the descriptive statisticals for sal\_m and sal\_f. Estimate the mean, median,variance, standard deviation, max, min and number of obseravations. The matrix names should be desc\_stat\_male and desc\_stat\_female;

desc\_stats\_male=[mean(sal\_m),median(sal\_m),var(sal\_m),std(sal\_m),max(sal\_m),min(sal\_m),size(sal\_m,1)]

desc\_stats\_female=[mean(sal\_f),median(sal\_f),var(sal\_f),std(sal\_f),max(sal\_f),min(sal\_f),size(sal\_f,1)]

d. Estimate the correlation between the salaries and the years of education after high school. Name the correlation corr\_sal\_years. Also estimate the p-values of the correlations. Name it as p\_corr\_sal\_years.

[corr\_sal\_years,p\_corr\_sal\_years]=corrcoef(A(:,3),A(:,4))

e. Do the same as before, but this time for males and females independently. Create a matrix with two columns: first column the years and second column the salaries for male and females. Name the matrices sal\_year\_m and sal\_year\_f. After this, estimate the correlations and p-values. Name the correlation matrices as corr\_sal\_years\_m and corr\_sal\_years\_f; the p-values should be named as p\_corr\_sal\_years\_m and p\_corr\_sal\_years\_f.

sal\_year\_m=[];

sal\_year\_f=[];

for i =1:size(A,1)

if A(i,2)==0

sal\_year\_m=[sal\_year\_m;A(i,[4,3])]

end

end

for j =1:size(A,1)

if A(j,2)==1

sal\_year\_f=[sal\_year\_f;A(j,[4,3])]

end

end

[corr\_sal\_years\_m,p\_corr\_sal\_years\_m]=corrcoef(sal\_year\_m(:,1),sal\_year\_m(:,2))

[corr\_sal\_years\_f,p\_corr\_sal\_years\_f]=corrcoef(sal\_year\_f(:,1),sal\_year\_f(:,2))

subplot(3,1,1)

scatter(A(:,4),A(:,3))

title("years vs salary")

xlabel("years after high school")

ylabel("salary")

subplot(3,1,2)

scatter(sal\_year\_m(:,1),sal\_year\_m(:,2))

title("years vs salary (male)")

xlabel("years after high school")

ylabel("Salary for male")

subplot(3,1,3)

scatter(sal\_year\_f(:,1),sal\_year\_f(:,2))

title("years vs salary (female)")

xlabel("years after high school")

ylabel("salary for female")

### Question 2,(OLS question, same with last hw) see last homework file.