Act 2: Correlation Analysis

Fall 2022

09/22/2022

Last Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_First Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Special note: Complete all homework in R. Please upload the completed hw on Canvas SubmitHW link.**

# 1.

Given a sample of size from a population , let .

Given the sample below

x=scan(tex="32.3 30.1 28.9 27.4 29.9 27.5 30.2 28.8 30.3 33.2")

verify the following identities via computations.

1. , where is sample variance of rank of .

x=scan(tex="32.3 30.1 28.9 27.4 29.9 27.5 30.2 28.8 30.3 33.2")  
n=length(x)  
r=rank(x)  
sr=sum(r)  
mr=mean(r)  
ssr=sum(r^2)  
ssdmr=sum((r-mr)^2)  
#(a)  
c(sr,n\*(n+1)/2)

[1] 55 55

#(b)  
c(mr,(n+1)/2)

[1] 5.5 5.5

#(c)  
c(ssr, n\*(n+1)\*(2\*n+1)/6)

[1] 385 385

#(d)  
c(ssdmr,ssr-n\*mr^2,n\*(n^2-1)/12)

[1] 82.5 82.5 82.5

#(e)  
v1=var(r)  
c(v1,ssdmr/(n-1), n\*(n+1)/12)

[1] 9.166667 9.166667 9.166667

# 2.

Given the bi-variate sample (x,y)

x=scan(tex="48.50 49.48 51.76 47.59 46.11 45.95 53.44 48.88 52.98 47.12 44.00 52.22 48.67 49.65 50.44")  
  
y=scan(tex="83.03 131.50 86.93 52.55 86.35 87.86 114.88 78.40 146.48 106.95 115.99 138.19 97.84 119.47 103.19")

complete following questions below:

1. Given a sample of ordered pairs from a bivariate population , an estimate of the population correlation is the **Pearson’s correlation** given by

Compute the value of for the given sample data.

1. It also appears that

Use this formula to compute the value of . Verify that this result matches with the result in (a).

1. An alternative formula often useful for Pearson’s correlation is given by

Use this definition to compute .

1. The built-in function **cor(x,y, method=“pearson”)** computes the Pearson correlation. Use this function to find the value of . Verify that this result matches with the result in (a), (b) or (c).
2. If the distribution of and in the bi-variate population is not normal, then the Spearman rank correlation is used and is given by

where and . This definition suggests that the Spearman’s rank correlation is the Pearson’s correlation between the ranks of and .

Compute the value of for the given sample using the definition of the Spearman’s rank correlation.

#cor(x, y, method = c("pearson", "kendall", "spearman"))  
#cor.test(x, y, method=c("pearson", "kendall", "spearman"))  
x=scan(tex="48.50 49.48 51.76 47.59 46.11 45.95 53.44 48.88 52.98 47.12 44.00 52.22 48.67 49.65 50.44")  
  
y=scan(tex="83.03 131.50 86.93 52.55 86.35 87.86 114.88 78.40 146.48 106.95 115.99 138.19 97.84 119.47 103.19")  
  
n=length(x)  
u=rank(x)  
v=rank(y)  
d=u-v  
mx=mean(x)  
my=mean(y)  
mu=mean(u)  
mv=mean(v)  
ssqx=sum((x-mx)^2)  
ssqy=sum((y-my)^2)  
ssqu=sum((u-mu)^2)  
ssqv=sum((v-mv)^2)  
spxy=sum((x-mx)\*(y-my))  
spuv=sum((u-mu)\*(v-mv))  
sx=sd(x)  
sy=sd(y)  
#(a)  
ra=spxy/sqrt(ssqx\*ssqy)  
ra

[1] 0.4426648

#(b)  
rb=1/(n-1)\*sum(((x-mx)/sx)\*((y-my)/sy))  
rb

[1] 0.4426648

#(c)  
sxy=sum(x\*y)  
sx=sum(x)  
sx2=sum(x^2)  
sy=sum(y)  
sy2=sum(y^2)  
num=sxy-sx\*sy/n  
den=sqrt((sx2-sx^2/n)\*(sy2-sy^2/n))  
rc=num/den  
rc

[1] 0.4426648

#(d)  
rd=cor(x,y,method="pearson")  
rd

[1] 0.4426648

c(ra,rb,rc,rd)

[1] 0.4426648 0.4426648 0.4426648 0.4426648

#(e)  
re=spuv/sqrt(ssqu\*ssqv)  
re

[1] 0.45