MATH 1070 Introductory Statistics
Annotated Example for Computing Correlation

Methodology

Correlation is computed by using the following steps (from Cryer and Miller, p. 147-148) [1]:

1. Compute the means and standard deviations for each variable.
2. Standardize each observation, case by case.
3. Multiply the standardized values of the observations together, case by case.
4. Add the resulting products and divide by one less than the number of cases.

Problem statement

A researcher wishes to determine whether a person's age is related to the number of hours he or she jogs per week. The data for the sample follow: [2]

Age, x	34	22	48	56	62
Hours, y	5.5	7	3.5	3	1

Step 1: Compute means and standard deviations

Means

$$\begin{aligned} \text{Age } \bar{x} &= \frac{34+22+48+56+62}{5} \\ &= 44.4 \end{aligned}$$

$$\begin{aligned} \text{Hours } \bar{y} &= \frac{5.5+7+3.5+3+1}{5} \\ &= 4 \end{aligned}$$

Standard deviations

$$\begin{aligned} \text{Age } s_x &= \sqrt{\frac{(34-44.4)^2+(22-44.4)^2+(48-44.4)^2+(56-44.4)^2+(62-44.4)^2}{5-1}} \\ &= 16.3340 \end{aligned}$$

$$\begin{aligned} \text{Hours } s_y &= \sqrt{\frac{(5.5-4)^2+(7-4)^2+(3.5-4)^2+(3-4)^2+(1-4)^2}{5-1}} \\ &= 2.3184 \end{aligned}$$

[1] *Statistics for Business: Data Analysis and Modeling, 2nd ed.*, Jonathan D. Cryer, Robert B. Miller, Duxbury Press, 1994

[2] *Elementary Statistics: A Step by Step Approach*, Allan G. Bluman, William C. Brown publishers, p. 384, 1992

Step 2: Standardize each observation, case by case

Case	Age		Hours	
i	x_i	$\frac{x_i-\bar{x}}{S_x}$	y_i	$\frac{y_i-\bar{y}}{S_y}$
1	34	$\frac{(34-44.4)}{16.3340} = -0.6367$	5.5	$\frac{(5.5-4)}{2.3184} = 0.6470$
2	22	$\frac{(22-44.4)}{16.3340} = -1.3714$	7	$\frac{(5.5-4)}{2.3184} = 1.2940$
3	48	$\frac{(48-44.4)}{16.3340} = 0.2204$	3.5	$\frac{(5.5-4)}{2.3184} = -0.2157$
4	56	$\frac{(56-44.4)}{16.3340} = 0.7102$	3	$\frac{(5.5-4)}{2.3184} = -0.4313$
5	62	$\frac{(62-44.4)}{16.3340} = 1.0775$	1	$\frac{(5.5-4)}{2.3184} = -1.2940$

			Standardized		
Case	Age	Hours	Age	Hours	Product
i	x_i	y_i	$\frac{x_i-\bar{x}}{S_x}$	$\frac{y_i-\bar{y}}{S_y}$	$\left(\frac{x_i-\bar{x}}{S_x}\right)\left(\frac{y_i-\bar{y}}{S_y}\right)$
1	34	5.5	-0.6367	0.6470	??
2	22	7	-1.3714	1.2940	??
3	48	3.5	0.2204	-0.2157	??
4	56	3	0.7102	-0.4313	??
5	62	1	1.0775	-1.2940	??
Sum					??

Step 3: Multiply the standardized values of the observations together

			Standardized		
Case	Age	Hours	Age	Hours	Product
i	x_i	y_i	$\frac{x_i-\bar{x}}{S_x}$	$\frac{y_i-\bar{y}}{S_y}$	$\left(\frac{x_i-\bar{x}}{S_x}\right)\left(\frac{y_i-\bar{y}}{S_y}\right)$
1	34	5.5	-0.6367	0.6470	-0.4119
2	22	7	-1.3714	1.2940	-1.7745
3	48	3.5	0.2204	-0.2157	-0.0475
4	56	3	0.7102	-0.4313	-0.3063
5	62	1	1.0775	-1.2940	-1.3943
Sum					??

Step 4: Add the resulting products ...

			Standardized		
Case	Age	Hours	Age	Hours	Product
i	x_i	y_i	$\frac{x_i-\bar{x}}{S_x}$	$\frac{y_i-\bar{y}}{S_y}$	$\left(\frac{x_i-\bar{x}}{S_x}\right)\left(\frac{y_i-\bar{y}}{S_y}\right)$
1	34	5.5	-0.6367	0.6470	-0.4119
2	22	7	-1.3714	1.2940	-1.7745
3	48	3.5	0.2204	-0.2157	-0.0475
4	56	3	0.7102	-0.4313	-0.3063
5	62	1	1.0775	-1.2940	-1.3943
Sum					-3.9346

Step 4: ... and divide by one less than the number of cases.

$$\begin{aligned} r &= -3.9346/(5-1) \\ &= -0.984 \end{aligned}$$

Interpretation

So what can we conclude from this value? First, the relationship is a negative relationship as indicated by the sign on the value. It is a strong relationship since the value (-0.984) is close to -1.0. As we know, the closer to one of the extreme values (-1.0, 1.0) the stronger the relationship. So we could conclude from this analysis that there is a strong negative correlation between the person's age and the number of hours he or she jogs.