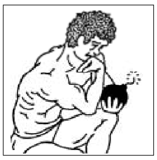
Think It Over



This section of *Resonance* presents thought-provoking questions, and discusses answers a few months later. Readers are invited to send new questions, solutions to old ones and comments, to 'Think It Over', *Resonance*, Indian Academy of Sciences, Bangalore 560 080. Items illustrating ideas and concepts will generally be chosen.

A Surprising Result on Correlation Coefficient

K S Bhanu[1] and M N Deshpande[2]
[1] Department of Statistics, Institute of Science, Nagpur, India.
2 Retd. Professor, Department of Statistics, Institute of Science, Nagpur, India.

Think-it-Over question appeared in *Resonance*, Vol.13, No.10, p.978, 2008.

Problem

Let n be a positive integer with $n \geq 3$. Let π be a random permutation of the integers $1, 2, \ldots, n$. The elements of π are denoted by $(\pi(1), \pi(2), \ldots, \pi(n))$. By a random permutation we mean a permutation selected from $n!$ possible permutations with equal probability $(1/n!)$. If $\pi(j) = i$, then $\pi^{-1}(i) = j$, i.e., $\pi^{-1}(i)$ represents the position of the integer i in the permutation π.

We associate two random variables with the selected random permutation as follows:

$$X = \text{Max}(\pi^{-1}(1), \pi^{-1}(n)),$$

$$Y = \text{Min}(\pi^{-1}(1), \pi^{-1}(n)).$$

We note that $2 \leq X \leq n, 1 \leq Y \leq (n-1)$ and $Y < X$. Compute the correlation coefficient between X and Y. It is a surprising fact that this correlation coefficient does not depend upon n.

Keywords
Correlation coefficient, random permutation.

Solution

Joint Distribution of X and Y

It is easy to see that

$$P[X = x, Y = y] = p(x, y) = \frac{2}{n(n-1)},\ 1 \leq y < x \leq n$$
$$= 0 \text{ otherwise.}$$

Marginal distributions of X and Y can be easily obtained as follows:

$$P[X = x] = h(x) = \sum_{y=1}^{x-1} p(x, y) = \frac{2(x-1)}{n(n-1)},\ 2 \leq x \leq n,$$

and

$$P[Y = y] = g(y) = \sum_{x=y+1}^{n} p(x, y) = \frac{2(n-y)}{n(n-1)},\ 1 \leq y \leq n-1.$$

Also,

$$E(X) = \sum_{x=2}^{n} xh(x) = \frac{2(n+1)}{3} ,$$
$$E(Y) = \sum_{y=1}^{n-1} yg(y) = \frac{n+1}{3} ,$$
$$E(X^2) = \sum_{x=2}^{n} x^2h(x) = \frac{(n+1)(3n+2)}{6},\ V(X) = \frac{(n+1)(n-2)}{18} ,$$
$$E(Y^2) = \sum_{y=1}^{n-1} y^2h(y) = \frac{n(n+1)}{6},\ V(Y) = \frac{(n+1)(n-2)}{18} .$$

We note that for every $n \geq 3, V(X) = V(Y)$.

Correlation Coefficient between X and Y

$$
\begin{aligned}
E(XY) &= \sum_{x=2}^{n} \sum_{y=1}^{x-1} x\, y\, p(x,y) \\
&= \sum_{x=2}^{n} \sum_{y=1}^{x-1} xy \; \frac{2}{n(n-1)} \\
&= \sum_{x=2}^{n} \frac{2x}{n(n-1)} \frac{x(x-1)}{2} \\
&= \frac{1}{n(n-1)} \left[\frac{n^2(n+1)^2}{4} - \frac{n(n+1)(2n+1)}{6} \right] \\
&= \frac{(n+1)\,(3n+2)}{12}
\end{aligned}
$$

which gives Cov $(XY) = (n+1)(n-2)/36$.

Hence the correlation coefficient between X and Y is $1/2$ which does not depend on n. We also note that the line of regression of Y on X is $Y = X/2$ and that of X on Y is $X = \frac{1}{2}Y + \frac{(n+1)}{2}$.