MTH211A: Theory of Statistics

Quiz1

Name: SOLUTION SET **Roll number:**

Time: 15 minutes **Total marks:** $5+5=10$

Suppose that the joint distribution of daily study time (X) and daily screen time (Y) of students at IITK is assumed to be normally distributed with parameters $(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$.
A person randomly selects 10 students from IITK, and inquires about their daily study time and screen time. The person then tabulates the responses as follows:

Serial no →	1	2	3	4	5	6	7	8	9	10
Study time (in hrs)	11.76	10.27	9.34	9.00	10.77	10.57	12.99	9.93	10.10	9.52
Screen time (in hrs)	4.00	5.60	6.50	5.50	4.68	6.41	4.68	5.11	6.04	6.46

Q.1 Find an estimate of the variability of daily screen time for a randomly chosen student of IITK based on this sample. Justify why you think this is a good estimate.

Let X be the random variable indicating study time of IITK students, and Y be the random variable indicating sreen time.

Given $\underset{\sim}{Z} = \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho\right)$

Variability of a random variable can be measured by its variance. So, we need to estimate $\text{var}(Y) = \sigma_Y^2$.

We know that $S_Y^{*2} = \frac{1}{(n-1)} \sum_{i=1}^{n} \left(Y_i - \bar{Y}_n\right)^2$ is the unbiased sample variance,

and it satisfies $E\left[S_Y^{*2}\right] = \sigma_Y^2$. So, on an average S_Y^{*2} neither over-estimates nor under-estimates σ_Y^2. So, we use S_Y^{*2} as an estimator of σ_Y^2.

Based on the 10 given realizations $y_1 = 4.00, \ldots, y_{10} = 6.46$, we calculate s_y^{*2} = realization of S_Y^{*2}, ~~based on the given~~ and use the same as an estimate of σ_Y^2.

Q.2 Suppose you know that the daily study time of a particular student is 10 hours. How will you modify the estimate of variability of screen time for that student?

When it is given that $X = 10$, the instead on considering the marginal distributn of Y, we would consider the conditional distn. of Y given $X = 10$.

We know that

$$Y \mid X=10 \sim N\left(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (10 - \mu_X) , \sigma_Y^2 (1-\rho^2) \right)$$

The variability of this distn. is measured by its variance, i.e., $\sigma_Y^2 (1-\rho^2) = \sigma_Y^2 - \sigma_Y^2 \rho^2$

As we do not have realizatns available from the conditional distributn., we would separately estimate σ_Y^2 and ρ^2.

From part (a) we have an estimate of σ_Y^2.

Further, ρ can be estimated by Sample correlation coefficient

$$R_{xy} = \frac{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{S_X \, S_Y} .$$

Based on the 10 realization, we will ~~estimate~~ calculate r_{xy} (realization of R_{xy}), and the proposed estimate of variability would be

$$s_y^{*2} \left(1 - r_{xy}^2\right).$$