Assignment 1 (three pages)

Statistics 32950-24620 (Spring 2024)

Due 9 am, Tuesday, March 26.

Requirements

- Your answers should be typed (allowing clear handwritten between typed texts for complex math formulas).

  Started with your name, Assignment 1, STAT 32950 or 24620; saved as `LastnameFirstnamePset1.pdf (or ...hw1.pdf)`, and uploaded to Gradescope under either 329Pset1 or 246Pset1.

  Make sure to **submit to the correct course number** you registered, and tag the pages for each question.

- When you use R (or others) to solve problems such as Question 1 in this assignment, select only relevant parts of the output, edit, then insert in your writing.

- You may discuss approaches with others. However the assignment should be devised and written by yourself.

Problem assignments

(Corresponding to Johnson and Wichern's chapters 1, 2, 3, and related background for chapters 4 and 8)

1. (*Basic description of multivariate data*)

   Download the data ladyrun24.dat (automatic download when clicked, also available next to the link of this p-set in Canvas).

   Save the dataset in your working directory.

   The data are on national track records for women, based on Table 1-9 in Johnson and Wichern.

   Measurements for 100m, 200m, and 400m are in seconds, longer distance records are in minutes.

   Variable names are not included.

   The following R command can be used to input the data (after saving the data in your working directory):

```
ladyrun = read.table("ladyrun24.dat")
colnames(ladyrun)=c("Country","100m","200m","400m","800m","1500m","3000m","Marathon")
```

   Compute the following (rounded to 2 decimal places) for the dataset.

   (a) Sample means of the variables.
   Is there any variable for which the mean is not meaningful (same judgement for the following questions)?

   (b) Sample covariance matrix and correlation matrix. Just the R command, no need to print the output.

   (c) Sample correlation matrix using Kendall's τ. Just the R command, no need to print the output.

   (d) Sample correlation matrix using Spearman's ρ. Just the R command, no need to print the output.

   (e) All three types of correlation matrix (Pearson, Kendall, Spearman) on the logarithm of the data. Again, just the R command, no need to print the output.
   Are the results using log-transformed data the same as in (b), (c), and (d)? Why?

   (f) Now for the sample correlation matrix $\boldsymbol{R}$ (of all meaningful variables), obtain the eigenvalues (show only 2 decimal places) and the eigenvectors (command only for eigenvectors, no need of output).

      i. What is the sum of all eigenvalues? Compare it to the dimensions of the variables.

      ii. Given the dimension(s) of each eigenvector.

   (Useful R commands for Question 1: `cov(..., method="...")`, `cor`, `eigen`, `sum`, `mean`, `rowMeans`, `colMeans`, `round`)

2. (*Joint distribution and conditional expectation, continuous case*)

The joint density of random variables (X, Y) is

$$f_{XY}(x, y) = \begin{cases} c(x^2 - y^2)e^{-x}, & \text{if } \ x > 0, -x \le y \le x, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Derive the value of c. (You may use the property of gamma function $\Gamma(k) = \int_0^\infty t^{k-1}e^{-t}dt = (k-1)!$ for integer k.)

(b) Derive the conditional density of Y given $X = x$.

(c) Derive the conditional expectation $g(x) = \mathbb{E}(Y \mid X = x)$ for $x > 0$.

(d) Derive the conditional variance $Var(Y \mid X = x)$ for $x > 0$.

3. (*Conditional expectation and conditional variance, discrete case*)

The following table lists the join probabilities of random variables X and Y.

	Y=1	Y=2	Y=3	Y=4
X=1	c	c	0	0
X=2	c	c	c	0
X=3	c	c	c	c

(a) Find the value of c. Derive the marginal probability mass functions $f_X(x) = \mathbb{P}(X = x)$ and $f_Y(y)$.

(b) Find the conditional expectation $g(x) = \mathbb{E}(Y \mid X = x)$ for $x = 1, 2, 3$.

(c) Find the conditional variance $Var(Y \mid X = x)$ for $x = 1, 2, 3$.

(d) Evaluate $\mathbb{E}\big[\mathbb{E}(Y \mid X)\big] = \mathbb{E}[g(X)]$. Verify that it equals $\mathbb{E}(Y) = \sum_y y f_Y(y)$ using results in (b).

(e) Evaluate $Var\big[\mathbb{E}(Y \mid X)\big]$. Derive the variance of Y using $Var(Y) = Var\big[\mathbb{E}(Y \mid X)\big] + \mathbb{E}\big[Var(Y \mid X)\big]$.

4. (*Derivations*)

(a) (*Expectation of random matrix*)

Let $\boldsymbol{C} = A\boldsymbol{X}B$, where $\boldsymbol{X}$ is a $p \times p$ random matrix, A, B are scalar (non-random) matrices of dimensions $k \times p$ and $p \times r$ respectively, and p, k, r are positive integers.

i. What are the dimensions of matrix $\boldsymbol{C}$?

ii. Write down c_{ij}, the (i, j)th entry of $\boldsymbol{C}$, in terms of elements of A, B and $\boldsymbol{X}$.
(Note: The expression has to be general, not for specific values or particular numerical dimensions.)

iii. Show that $\mathbb{E}(\boldsymbol{C}) = A\,\mathbb{E}(\boldsymbol{X})B$.

(b) (*Positive semi-definiteness of covariance matrix*)

Show that all p eigenvalues of covariance matrix $\Sigma = Cov(\boldsymbol{Y}) \in \mathbb{R}^{p \times p}$ must be nonnegative, where $\boldsymbol{Y} = [Y_1, ..., Y_p]^T$ is a random vector in $\mathbb{R}^p$.

(c) (*Spectral decomposition*) Let $A = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ for $\rho \in (0, 1)$.

i. Derive the eigenvalues (λ_i's) of A (by hand, show work).

ii. Derive(select) unit-length eigenvectors (v_i's) of A and show that they are orthogonal (by hand, show work).

iii. Write out the spectral decomposition (*a.k.a.* eigen-decomposition) $A = V\Lambda V^T$, where the columns of V are orthonormal eigenvectors, and Λ is the diagonal matrix of eigenvalues of A.

iv. Use the spectral decomposition to write A^{-1} in terms of (matrix operations of) V and Λ.

v. Use the spectral decomposition to derive R (in terms of operations of V and Λ) such that $R^2 = A$.

5. (*Joint, marginal and conditional distributions of continuous random variables*)

The trivariate random vector (W, X, Y) has joint probability density function

$$f_{W,X,Y}(w, x, y) = \frac{2}{\pi} e^{x(y+w-x-4)-\frac{1}{2}(y^2+w^2)}, \qquad x \geq 0;\ w, y \in \mathbb{R}.$$

In answering the following questions, show the steps of your derivations.

(a) Find the joint density $f_{X,Y}(x, y)$ of X and Y. (Hint: Do an appropriate completing-the-square in the component.)

(b) Find the marginal density $f_X(x)$ of X.

(c) Find $\mathbb{E}(Y \mid X = x)$.

(d) Find $\mathbb{E}(WY \mid X = x)$.

(e) (**Required for 32950 students only**. Optional for 24620.)
Show that W and Y are not independent, but they are conditionally independent given $X = x$.