P-set1 **For your personal use in this course only. Do not circulate or post.**

1. (a) Let R denote the outcome number in red, B the number in blue. List the joint probabilities of (R, B) in a 3×3 table, then add the margins (row/column probabilities).

   The joint probability table and margins:

Probability	B (blue)			
R (red)	B=1	B=3	B=5	R margin
R=0			$\frac{1}{3}$	$\frac{1}{3}$
R=2	$\frac{1}{3}$			$\frac{1}{3}$
R=4		$\frac{1}{3}$		$\frac{1}{3}$
B margin	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

   (b) Rule-I: A single egg is taken at random from the basket, determining both the blue (B) and the red (R) numbers. Which color should you choose under Rule-I?

   Support your choice by computing your winning probability, using joint probabilities of (R, B).

   Under Rule-I, (R, B) are sampled jointly. The winning probability of picking Red is

$$P(R > B) = P[(R, B) = (4, 3)] + P[(R, B) = (2, 1)] = \frac{2}{3} > \frac{1}{2}$$

   So we should pick Red.

   (c) Rule-II: You take an egg at random and read the number of your color. Then you put the egg back into the basket and cover all eggs. Then Mr Trick chooses an egg at random and reads the number of his color. Which color should you choose under Rule-II?

   Compute your winning probability, using marginal probabilities of R and B.

   Under Rule-II, R and B are sampled independently. The winning probability of picking Red can be calculated using marginal distributions of R and B as

$$P(R > B) = P(R = 4)[P(B = 1) + P(B = 3)] + P(R = 2)P(B = 1) = \frac{1}{3}\left(\frac{1}{3} + \frac{1}{3}\right) + \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{3}$$

   Thus $P(B > R) = P(B \geq R) = 1 - P(R > B) = \frac{2}{3}$. We should pick Blue.

   (d) Rule-III: Like Rule-II, but your egg is not put back, so Mr Trick only has two remaining eggs (properly covered) to choose from. Which color should you choose under Rule-III?

   Compute your winning probability, using conditional probabilities of $B|R$ or $R|B$, by applying the law of total probability $P(Y = y) = \sum_x P(Y = y|X = x)P(X = x)$.

   Under Rule-III, if your color were red, that is, R is sampled first, then B is sampled conditioned on the output of R. The winning probability of picking Red can be calculated using conditional distribution of $B|R$:

$$\begin{aligned} P(R > B) &= \sum_i P(R = i)P(i > B|R \neq i) \\ &= P(R = 4)P(4 > B|R \neq 4) \;=\; P(R = 4)P(B < 4|R = 0 \textit{ or } R = 2) \\ &= P(R = 4)\frac{P(B < 4, R = 0 \textit{ or } R = 2)}{P(R = 0 \textit{ or } R = 2)} \;=\; P(R = 4)\frac{P(B = 1, R = 2)}{P(R = 0) + P(R = 2)} = \frac{1}{6} \end{aligned}$$

   Analogously, if B is sampled first, $P(B > R) = \sum_j P(B = j)P(j > R|B \neq j)$ can be computed using conditional distribution of $R|B$ as given in the table here.

Conditional prob.	$B\|R$		
R (red)	$B = 1\|R$	$B = 3\|R$	$B = 5\|R$
R=0	0	0	1
R=2	1	0	0
R=4	0	1	0

   Either scenario gives $P(B > R) = \frac{5}{6}$, we should pick Blue.

   Note: The winning probability can be calculated with other methods such as using joint probabilities, here you are asked to utilize conditional probability.

2. (a) The averages of the variables are


```
   100m    200m    400m    800m   1500m   3000m Marathon
  11.31   23.07   51.82    2.02    4.19    9.07   153.31
```

   The average of the variable "Country" is not meaningful, since it is a nominal variable - its values have no numerical meaning.

```
> ladyrun = read.table("ladyrun23.dat")
> colnames(ladyrun)=c("Country","100m","200m","400m","800m","1500m","3000m","Marathon")
> round(colMeans(ladyrun[,-1]),2)      # -1 means all but the first column
```

   (b) The sample covariance and correlation matrices of variables other than "Country" can be obtained by commands

```
> round(cov(ladyrun[,-1]),2)
> round(cor(ladyrun[,-1]),2)
```

   (c) Sample correlation matrix using Kendall's τ for variables other than "Country":

```
round(cor(ladyrun[,-1],method="kendall"),2)
```

   (d) Sample correlation matrix using Spearman's ρ for variables other than "Country":

```
round(cor(ladyrun[,-1],method="spearman"),2)
```

   (e) Sample correlation matrices on the logarithm of the data can be obtained by commands

```
> round(cor(log(ladyrun[,-1]),method="pearson"),2)
> round(cor(log(ladyrun[,-1]),method="kendall"),2)
> round(cor(log(ladyrun[,-1]),method="spearman"),2)
```

   The correlation matrix using Pearson has changed, because the Pearson correlation measures linear relationship, and logarithm transformation is not linear.

   However correlation matrices using Kendall's τ and Spearman's ρ are the same as those of the original data. This is because these two correlations are based on ranked data, log function is monotone which preserves rank.

   (f) Eigenvalues $\lambda_1, \cdots, \lambda_7$ of $\boldsymbol{R}$ in descending order are $5.70, 0.74, 0.29, 0.11, 0.09, 0.05, 0.02$. The sum of the eigenvalues $\sum_{i=1}^{7} \lambda_i = 7 = p$, which is the number of variables in the data.

```
eigen(cor(ladyrun[,-1]))
round(eigen(cor(ladyrun[,-1]))$values,2); round(eigen(cor(ladyrun[,-1]))$vectors,2)
sum(eigen(cor(ladyrun[,-1]))$values)
```

3. (a) $c = 1/9$, from total probabilities $= 1$. The probability table can be extended to

$P(X = x, Y = y)$	Y=1	y=2	y=3	y=4	$f_X(x)$	$E(Y \mid X = x)$	$V(Y \mid X = x)$
x=1	1/9	1/9	0	0	2/9	1.5	1/4
x=2	1/9	1/9	1/9	0	1/3	2	2/3
x=3	1/9	1/9	1/9	1/9	4/9	2.5	5/4
$f_Y(y)$	1/3	1/3	2/9	1/9			

   $f_X(x) = 2/9, 1/3, 4/9$ for $x = 1, 2, 3$, respectively.
   $f_Y(y) = 1/3, 1/3, 2/9, 1/9$ for $y = 1, 2, 3, 4$, respectively.

   (b) $g(x) = E(Y \mid X = x) = 1.5, 2, 2.5$ for $x = 1, 2, 3$, respectively.

   (c) $V(Y|X = x) = 1/4, 2/3, 5/4$ for $x = 1, 2, 3$, respectively.

   (d) $E\big[E(Y \mid X)\big] = 1.5 * 2/9 + 2 * 1/3 + 2.5 * 4/9 = 19/9$.
   $E(Y) = 1 * 1/3 + 2 * 1/3 + 3 * 2/9 + 4 * 1/9 = 19/9$.

   (e) $V\big[E(Y \mid X)\big] = 1.5^2 * 2/9 + 2^2 * 1/3 + 2.5^2 * 4/9 - 19^2/9^2 = 25/162$
   $E(V(Y|X)) = 1/4 * 2/9 + 2/3 * 1/3 + 5/4 * 4/9 = 5/6$
   $V(Y) = V\big[E(Y \mid X)\big] + E(V(Y|X)) = 25/162 + 5/6 = 80/81$. Which is the same as if we derive $V(Y)$ directly from $V(Y) = 1/3 + 2^2/3 + 3^3 * 2/9 + 4^2 * 1/9 - (19/9)^2 = 80/81$.

4. (a) (*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).

$$|A - \lambda I| = \begin{vmatrix} 1 - \lambda & \rho \\ \rho & 1 - \lambda \end{vmatrix} = (1 - \lambda)^2 - \rho^2 = (1 - \lambda - \rho)(1 - \lambda + \rho) = 0 \quad \Rightarrow \quad \lambda_1 = 1 + \rho, \, \lambda_2 = 1 - \rho.$$

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

$$(A - \lambda_1 I)v_1 = \begin{bmatrix} -\rho & \rho \\ \rho & -\rho \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = 0 \quad \Rightarrow \quad x_1 = y_1 \quad \Rightarrow \quad v_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$(A-\lambda_2 I)v_2 = \begin{bmatrix} \rho & \rho \\ \rho & \rho \end{bmatrix}\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = 0 \quad\Rightarrow\quad x_2 = -y_2 \quad\Rightarrow\quad v_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \ \textit{or}\ v_2' = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

$v_1^T v_2 = v_2^T v_1 = 1/\sqrt{2}\cdot(-1/\sqrt{2}) + 1/\sqrt{2}\cdot 1/\sqrt{2} = 0$, so $v_1 \perp v_2$ (similar steps to get $v_1 \perp v_2'$).

iii. Write out the spectral decomposition (a.k.s 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.

$$A = V\Lambda V^T = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

If: v_2' is uased instead of v_2,

$$A = V\Lambda V^T = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}\begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

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

$$\begin{aligned} A^{-1} &= (V\Lambda V^T)^{-1} = (V^T)^{-1}\Lambda^{-1}V^{-1} = V\lambda_{-1}V^T \\ &= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}\begin{bmatrix} 1/(1+\rho) & 0 \\ 0 & 1/(1-\rho) \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \end{aligned}$$

where we used the inverse of a product $(BC)^{-1} = C^{-1}B^{-1}$, the inverse of a matrix transpose $(B^T)^{-1} = (B^{-1})^T$, and $V^{-1} = V^T$ since $VV^T = V^TV = I$. (The last step checks it agrees with the inverse matrix formula.)

v. Use the spectral decomposition to find $R = A^{1/2}$ (in terms of operations of V and Λ) such that $A = R^2$.

Denote $\Lambda^{1/2} = \begin{bmatrix} \sqrt{1+\rho} & 0 \\ 0 & \sqrt{1-\rho} \end{bmatrix}$, let $R = V\Lambda^{1/2}V^T$, and use $V^TV = I, \Lambda^{1/2}\Lambda^{1/2} = \Lambda$, then

$$A = V\Lambda V^T = V\Lambda^{1/2}I\Lambda^{1/2}V^T = V\Lambda^{1/2}V^T\,V\Lambda^{1/2}V^T = RR = R^2$$

(b) $\Sigma' = \Sigma$, all eigenvalues of symmetric matrix are real. For any $a \in \mathbb{R}^p$,

$$a'\Sigma a = a'Cov(X)a = Var(a'X) \geq 0$$

Then by definition, Σ is a positive semi-definite matrix. Again by definition, all eigenvalues of positive semi-definite matrix are ≥ 0.

(c) Since the sum and product of r.v.'s are continuous r.v.'s with point mass probability 0, $Pr[(X_1-X_2)(Y_1-Y_2) = 0] = 0$ (used in the 2nd "=" below). Since $(X_1, Y_1)$and (X_2, Y_2) are independent and follow the same distribution, $P[X_1 > X_2, Y_1 > Y_2] = P[X_2 > X_1, Y_2 > Y_1]$ (used in the 4th "=" below).

$$\begin{aligned} \tau &= Pr[(X_1-X_2)(Y_1-Y_2) > 0] - (1 - P[(X_1-X_2)(Y_1-Y_2) \geq 0]) = 2P[(X_1-X_2)(Y_1-Y_2) > 0] - 1 \\ &= 2\{P[X_1 > X_2, Y_1 > Y_2] + P[X_1 < X_2, Y_1 < Y_2]\} - 1 = 4P[X_1 < X_2, Y_1 < Y_2] - 1 \\ &= 4\int_{\mathbb{R}}\int_{\mathbb{R}} P(X_1 < x, Y_1 < y \mid X_2 = x, Y_2 = y)F_{X_2,Y_2}(dx, dy) - 1 \\ &= 4\int_{\mathbb{R}}\int_{\mathbb{R}} P(X_1 < x, Y_1 < y)F_{X_2,Y_2}(dx, dy) - 1 \\ &= 4\int_{\mathbb{R}}\int_{\mathbb{R}} F_{X_1,Y_1}(x,y)F_{X_2,Y_2}(dx, dy) - 1 = 4\iint_{\mathbb{R}^2} F(x,y)dF(x,y) - 1 \end{aligned}$$

(d) $\mathbf{C}$ is of dimensions $k \times r$. $C_{ij} = \sum_{m=1}^p \sum_{n=1}^p A_{im}X_{mn}B_{nj}$
$E(C_{ij}) = \sum_{m=1}^p \sum_{n=1}^p E(A_{im}X_{mn}B_{nj}) = \sum_{m=1}^p \sum_{n=1}^p A_{im}E(X_{mn})B_{nj}$

5. Obtain the eigenvalue and eigenvectors of Σ:

```
> sigma=matrix(c(2,-1,1,-1,4,0,1,0,3),3,3)
> eigen(sigma)
eigen() decomposition
$values
[1] 4.532089 3.347296 1.120615

$vectors
           [,1]      [,2]       [,3]
[1,]  0.4490988 0.2931284 -0.8440296
[2,] -0.8440296 0.4490988 -0.2931284
[3,]  0.2931284 0.8440296  0.4490988
```


(a) The eigenvalues are the eigenvalues $\lambda_1 = 4.53, \lambda_2 = 3.35, \lambda_3 = 1.12$

(b) Verify that, in the R output, the eigenvectors (the column vectors) $(a_{i1}, a_{i2}, a_{i3})'$ already scaled to $a_{i1}^2 + a_{i2}^2 + a_{i3}^2 = 1$, for $i = 1, 2, 3$. The PC variables $Y_i = \boldsymbol{a}_i'X = a_{i1}X_1 + a_{i2}X_2 + a_{i3}X_3$ are

$$\begin{aligned} Y_1 &= 0.45X_1 - 0.84X_2 + 0.29X_3, \\ Y_2 &= 0.29X_1 + 0.45X_2 + 0.84X_3, \\ Y_3 &= 0.84X_1 + 0.29X_2 - 0.45X_3 \end{aligned}$$

(c) We have

$$\begin{aligned} V(Y_1) = V\left(\sum_{j=1}^n a_{1j}X_j\right) &= \sum_{j=1}^n a_{1j}^2 V(X_j) + 2\sum_{1\leq i<j\leq n} a_{1j}a_{1i}Cov(X_i, X_j), \\ &= \boldsymbol{a}_1^T\Sigma\boldsymbol{a}_1 \text{ (by re-writing the sum in quadratic forms)} \\ &= \lambda_1 \cdot a_1^T a_1 \text{ (by } a_1 \text{ being an eigenvector)} \\ &= \lambda_1 \end{aligned}$$

So the variance of Y_1 is equal to λ_1. The proof also implies that in general the variance $V(Y_i) = \lambda_i$.

6. The joint density of random variables (X, Y) is $f_{XY}(x,y) = \begin{cases} \dfrac{c}{(1+x+y)^3} & \text{if } 0 \leq x, y, \\ 0, & \text{otherwise.} \end{cases}$

(a) Derive the value of c.

$$\int_0^\infty \left(\int_0^\infty \frac{1}{((1+x+y)^3}dx\right)dy = \int_0^\infty \left(\frac{-1/2}{(1+x+y)^2}\Bigg|_{x=0}^{x=\infty}\right)dy = \int_0^\infty \frac{1}{2(1+y)^2}dy = \frac{-1}{2(1+y)}\Bigg|_{y=0}^{y=\infty} = \frac{1}{2}$$

Thus $c = 2$, since $\int_0^\infty\int_0^\infty f(x,y)dxdy = 1$.

(b) Derive the marginal density of X.

$$f_X(x) = \int_0^\infty f_{XY}(x,y)dy = \frac{2(-1/2)}{(1+x+y)^2}\Bigg|_{y=0}^{y=\infty} = \frac{1}{(1+x)^2}, \quad x \in [0, \infty)$$

(c) Derive the conditional density $f_{Y|X}(y \mid x)$ for $x > 0$. $f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{2(1+x)^2}{(1+x+y)^3}$

(d) i. Analogous to the derivation for X, the density of Y is $f_Y(y) = \int_0^\infty f_{XY}(x,y)dy = \frac{1}{(1+y)^2}$ for $y > 0$.

$$\mathbb{E}(Y) = \int_0^\infty yf_Y(y)dy = \int_0^\infty \frac{y+1-1}{(1+y)^2}dy = \int_0^\infty \left(\frac{1}{(1+y)} - \frac{1}{(1+y)^2}\right)dy = \log(1+y)\Big|_0^\infty + \frac{1}{(1+y)}\Bigg|_0^\infty = \infty - 1 = \infty$$

That is, $E(Y)$ is $+\infty$, a type of non-existence of the mean.

ii. Derive the conditional expectation
$g(x) = \mathbb{E}(Y \mid X = x)$ for $x > 0$, with detailed integration steps without quoting integral formulas.
The following gives one way to do the calculation.

$$\begin{aligned} \mathbb{E}(Y|X = x) &= \int_0^\infty yf_{Y|X}(y|x)dy = 2(1+x)^2\int_0^\infty \frac{y+1+x-(1+x)}{(1+x+y)^3}dy \\ &= 2(1+x)^2\int_0^\infty \left(\frac{1}{(1+x+y)^2} - \frac{(1+x)}{(1+x+y)^3}\right)dy \\ &= 2(1+x)^2\left(\frac{-1}{(1+x+y)} - \frac{(-1/2)(1+x)}{(1+x+y)^2}\right)\Bigg|_{y=0}^{y=\infty} \\ &= 2(1+x)^2\left(\frac{1}{(1+x)} - \frac{-(1+x)/2}{(1+x)^2}\right) = x+1, \quad x > 0. \end{aligned}$$