Ejercicio a mano

Nota: Para simplificar la notación, en vez de X_0 para referirme a una variable centrada en su media, usaré X_c. Si me refiero a la variable X_1 centrada en su media escribiré $X_{1,\text{c}}$. Si escribo X sin índices, me refiero a la matriz, pero si lo hago con índices, como X_1, me refiero a una *columna* (o variable), de la matriz. Para la covarianza de muestras usaré S. Por ejemplo, $\text{cov}[X_1, X_2]$, la covarianza de la primera columna con la segunda, la abreviaré como S_{12}.

Supón que en 6 personas se realizaron 3 exámenes de 15 preguntas en diferentes momentos. El promedio de respuestas correctas para cada persona se coloca en la siguiente matriz.

$$X = \begin{bmatrix} \text{Examen}_1 & \text{Examen}_2 & \text{Examen}_3 \\ \hline 9.44 & 10.46 & 10.40 \\ 9.77 & 8.73 & 10.11 \\ 11.56 & 9.31 & 9.44 \\ 10.07 & 9.55 & 11.79 \\ 10.13 & 11.22 & 10.50 \\ 11.72 & 10.36 & 8.03 \end{bmatrix}$$

Al examen 1 le llamaremos X_1, y así para los otros dos.

Supón que queremos ver la dependencia lineal de un examen con respecto a los otros (los detalles no importan)

1. Obtener la matriz de covarianza.
   1. Obtener primero la matriz X_c, para lo cual necesitamos $\bar{X}_1, \bar{X}_2, \bar{X}_3$. Luego, cada celda debe ser restada por el promedio de su columna. Por ejemplo, $9.44 - \bar{X}_1$ para la celda de la primera fila y primera columna, $10.46 - \bar{X}_2$ para la celda en la primera fila y segunda columna.
   2. Obtener las covarianzas usando las variables centradas usando el producto punto. Por ejemplo, la covarianza entre la primera columna y la segunda se obtiene así

      $$S_{12} = \frac{1}{n-1} \mathbf{x}_{1,\text{c}}^\top \mathbf{x}_{2,\text{c}}$$

      para la covarianza de la primera columna con la segunda, S_{13} para Va covarianza de la primera columna con la tercera, S_{23} para Va covarianza de la segunda con la tercera; además, obtener las covarianzas que van en la diagonal de la primera matriz de covarianzas, que es la covarianza de una columna consigo misma (o la varianza), por ejemplo $S_{11} = \text{Var}[X_1]$.

3. El arreglo quedaría

$$S = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix}$$

Nota que las celdas con el mismo color son es el mismo valor, porque $\mathrm{Cov}[x_1, x_2] = \mathrm{Cov}[x_2, x_1]$, esto te será obvio si revisas la ecuación 5 del tema 1.1 Correlaciones y covarianzas, dado que la covarianza se obtiene de un producto, su orden no altera el resultado.

2. Obtener la matriz de correlaciones

$$R = \begin{bmatrix} \frac{1}{n-1}\mathbf{z}_1^\top \mathbf{z}_1 & \frac{1}{n-1}\mathbf{z}_1^\top \mathbf{z}_2 & \frac{1}{n-1}\mathbf{z}_1^\top \mathbf{z}_3 \\ \frac{1}{n-1}\mathbf{z}_2^\top \mathbf{z}_1 & \frac{1}{n-1}\mathbf{z}_2^\top \mathbf{z}_2 & \frac{1}{n-1}\mathbf{z}_2^\top \mathbf{z}_3 \\ \frac{1}{n-1}\mathbf{z}_3^\top \mathbf{z}_1 & \frac{1}{n-1}\mathbf{z}_3^\top \mathbf{z}_2 & \frac{1}{n-1}\mathbf{z}_3^\top \mathbf{z}_3 \end{bmatrix}$$

1. Primero transforma cada valor en la matriz un valor z. Por ejemplo, para transformar 9.44 a un valor z es $(9.44 - \bar{X}_1)/SD(X_1)$, en donde $SD(X_1) = \sqrt{S_{11}}$, y S_{11} ya la obtuviste en el primer ejercicio. El valor z para 11.79 (cuarta fila, tercera columna) lo obtendrías de forma similar $(11.79 - \bar{X}_3)/SD(X_3)$, y $SD(X_3) = \sqrt{S_{33}}$.

2. Una vez transformes todas las columnas (por ejemplo, todos los valores de la columna 1 $z-$transformados les llamaremos $\mathbf{z}_1$), obtienes cada una de las entradas de la matriz de correlaciones usando el producto punto de los vectores $\mathbf{z}$. Por ejemplo, la correlación entre la variable 1 y la 2 sería

$$r_{12} = \frac{1}{n-1}\mathbf{z}_1^\top \mathbf{z}_2 = \frac{1}{n-1}\sum_{i=1}^{6} z_{i1} z_{i2}$$

por ejemplo, para la primera fila, el producto sería $z_{11}z_{12} = (-1.05)(0.58)$, para la segunda, $z_{21}z_{22} = (-0.71)(-1.33)$

3. ¿Cuál es el par de variables con la correlación más fuerte (positiva o negativa)? ¿Qué tipo de dependencia nos indica? La matriz de covarianza debería salirte:

$$S = \begin{bmatrix} 0.91 & -0.03 & -0.89 \\ -0.03 & 0.82 & -0.09 \\ -0.89 & -0.09 & 1.56 \end{bmatrix}$$

La matriz de correlaciones daría

$$R = \begin{bmatrix} 1.00 & -0.03 & -0.74 \\ -0.03 & 1.00 & -0.08 \\ -0.74 & -0.08 & 1.00 \end{bmatrix}$$

Nota de nuevo que $r_{12} = r_{21}$.