

Estudiantes:

- Daniel Garzón Rodríguez
- Pedro José Leal Mesa
- Luis Enrique Mantilla Sanabria

Solución Taller 1 primera parte

1. Estudie el siguiente paper: Hubert, M., & Debruyne, M. (2010). Minimum covariance determinant.

2. Explique:

a) ¿Qué es el efecto “masking”?

Ocurre cuando hay datos atípicos y la metodología no los detecta.

b) ¿Qué es el efecto “swamping”?

Ocurre cuando la metodología nombra a datos no atípicos como atípicos.

c) ¿Por qué los estimadores MCD de localización y dispersión son estimadores robustos?

Los estimadores MCD son robustos porque buscan las h observaciones dentro de los datos que generan **la menor dispersión posible** (mínimo determinante de la matriz de covarianza), lo cual está relacionado con el volumen de la hiperelipse umbral definida por la distribución de la distancia de Mahalanobis.

d) Supuestos que requieren los estimadores MCD:

- Los datos provienen de una distribución unimodal y simétrica (idealmente elíptica). Una distribución multivariante con parámetros $\mu \in \mathbb{R}^p$ y una matriz **definida positiva** $\Sigma \in \mathbb{R}^{p \times p}$ se dice que es **elípticamente simétrica y unimodal** si existe una función estrictamente decreciente g tal que su función de densidad puede expresarse como:

$$f(\mathbf{x}) = \frac{1}{\sqrt{|\Sigma|}} g((\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu))$$

- Existe un número mayoritario de observaciones “buenas” un porcentaje menor de outliers.
- El número de observaciones n debe ser considerablemente mayor que la dimensión p (se recomienda $n > 5p$).
- La matriz S debe ser definida positiva.

e) **Elipse de tolerancia**

Es una región elíptica que contiene un porcentaje determinado de las observaciones (por ejemplo, 95 % o 97.5 %) bajo el supuesto de normalidad. En multivariado, esta elipse se construye a partir de la media y la matriz de covarianza (clásica o robusta) y permite identificar observaciones atípicas que se encuentren fuera de esta región.

Note que la distancia de Mahalanobis se puede observar como una elipse de la siguiente manera:

$$MD(x, \mu) = (x - \mu)'(C\Lambda^{-1}C')(x - \mu)$$

Sea

$$z = C'(x - \mu)$$

Reemplazando:

$$MD(x, \mu) = z' \Lambda^{-1} z$$

Note que z es la proyección de x normalizado en sus autovectores.

Tomando los valores z_j , normalizando y elevando al cuadrado:

$$\sum_j \frac{z_j^2}{\lambda_j} = \chi_{p,0,975}^2$$

donde λ_j son los autovalores de la matriz de covarianza y corresponden a las varianzas en cada dirección principal.

f) **Equivarianza afín de un estimador**

Un estimador es afínmente equivariante si sus resultados se comportan adecuadamente ante transformaciones lineales de los datos. Formalmente, si se transforma el conjunto de datos como $\mathbf{X}^* = A\mathbf{X} + b$, entonces el estimador de localización y dispersión debe transformarse como:

$$\hat{\mu}^* = A\hat{\mu} + b, \quad \hat{\Sigma}^* = A\hat{\Sigma}A^\top$$

g) **Punto de ruptura de un estimador**

Es la fracción máxima de observaciones que pueden ser reemplazadas por valores arbitrarios sin que el estimador se deteriore completamente. En el MCD, el punto de ruptura depende del parámetro h y alcanza su valor máximo teórico cuando $h \approx \frac{n+p+1}{2}$, llegando a un 50 %.

Según el artículo del MCD para un estimador de localización multivariada T_n , el punto de ruptura se define como:

$$\varepsilon_n^*(T_n; X_n) = \frac{1}{n} \min \left\{ m \in \{1, \dots, n\} : \sup_{X_{n,m}} \|T_n(X_n) - T_n(X_{n,m})\| = \infty \right\}$$

Para un estimador de dispersión multivariada C_n :

$$\varepsilon_n^*(C_n; X_n) = \frac{1}{n} \min \left\{ m \in \{1, \dots, n\} : \sup_{X_{n,m}} \max_i |\log \lambda_i(C_n(X_n)) - \log \lambda_i(C_n(X_{n,m}))| = \infty \right\}$$

donde λ_i son los autovalores de la matriz de dispersión C_n . Consideramos que el estimador está roto cuando alguno de los autovalores puede hacerse arbitrariamente grande o pequeño.

Para el estimador MCD *raw*, bajo la condición de que $k(X_n) < h$, donde $k(X_n)$ es el número máximo de observaciones que yacen sobre un hiperplano en \mathbb{R}^p , se tiene:

$$\varepsilon_n^*(\hat{\mu}_0; X_n) = \varepsilon_n^*(\hat{\Sigma}_0; X_n) = \frac{\min(n - h + 1, h - k(X_n))}{n}$$

Si los datos provienen de una distribución continua (posición general), entonces casi con certeza $k(X_n) = p$, y se tiene:

$$\varepsilon_n^*(\hat{\mu}_0; X_n) = \varepsilon_n^*(\hat{\Sigma}_0; X_n) = \frac{\min(n - h + 1, h - p)}{n}$$

Esto implica que, para valores de h entre $\lfloor \frac{n+p}{2} \rfloor$ y $\lfloor \frac{n+p+1}{2} \rfloor$, el MCD alcanza el **punto de ruptura máximo**:

$$\varepsilon_n^* = \frac{n - p + 2}{2n}$$

Este es el valor más alto posible para estimadores de covarianza afínmente equivariantes.

h) **Eficiencia del estimador**

La eficiencia mide la variabilidad del estimador frente a su competencia en condiciones ideales.

i) **Efecto de los outliers en los valores propios de la matriz de covarianza clásica**

Los outliers tienden a inflar desproporcionadamente los valores propios más grandes de la matriz de covarianza clásica. Esto provoca que la forma y tamaño de las elipses de tolerancia calculadas con dicha matriz se distorsionen, generando regiones demasiado grandes que dificultan la detección de outliers.

j) **Papel de los valores propios de la matriz de covarianza clásica en la construcción de las elipses de tolerancia**

Los valores propios de la matriz de covarianza clásica determinan la longitud de los ejes principales de la elipse de tolerancia. Valores propios grandes generan elipses alargadas en ciertas direcciones, reflejando mayor dispersión de los datos en esas direcciones.

k) **Papel de los valores propios de la matriz de covarianza MCD en la construcción de las elipses de tolerancia**

Similar al caso clásico, los valores propios de la matriz de covarianza MCD determinan la forma y tamaño de las elipses de tolerancia, pero considerando solo los datos “buenos”. Esto permite obtener elipses más ajustadas a la estructura real de los datos, sin ser afectadas por outliers, mejorando así la detección de observaciones anómalas.

3. **Estudie el artículo de Zuo, Y., & Serfling, R. (2000).**

4. **Explique las siguientes profundidades:** Mahalanobis, Half Space de Tukey y espacial (Spatial Depth).

Explicación: Del artículo Zuo, Y., & Serfling, R. (2000) se explicará las cualidades que debe tener una función de profundidad y luego se adentrará en las profundidades pedidas con sus respectivas propiedades.

Cualidades que debe tener una función de profundidad:

Una función de profundidad estadística, $D(\cdot; P)$, definida para una distribución P en \mathbb{R}^d , debe proporcionar un ordenamiento de los puntos desde el centro hacia afuera. Para ser considerada ideal, debe satisfacer las siguientes cuatro propiedades matemáticamente precisas:

- a) **P1. Invarianza Afín** La profundidad de un punto $x \in \mathbb{R}^d$ no debe depender del sistema de coordenadas ni de las escalas de las variables. Formalmente, para cualquier vector aleatorio $X \in \mathbb{R}^d$, cualquier matriz no singular $A \in \mathbb{R}^{d \times d}$ y cualquier vector $b \in \mathbb{R}^d$,

$$D(Ax + b; F_{AX+b}) = D(x; F_X),$$

donde F_X es la distribución de X y F_{AX+b} la de $AX + b$. Esto asegura que la función sea robusta ante transformaciones afines.

- b) **P2. Maximalidad en el Centro** Para una distribución F con un centro único θ (por ejemplo, un punto de simetría según alguna definición), la profundidad debe alcanzar su máximo en θ :

$$D(\theta; F) = \sup_{x \in \mathbb{R}^d} D(x; F).$$

Esto implica que el punto más profundo coincide con el centro de la distribución, asumiendo que dicho centro está bien definido (e.g., en distribuciones simétricas).

- c) **P3. Monotonía Relativa al Punto Más Profundo** La profundidad debe disminuir monótonicamente a medida que un punto $x \in \mathbb{R}^d$ se aleja del punto más profundo θ a lo largo de cualquier rayo fijo que pase por θ . Matemáticamente, para cualquier $\alpha \in [0, 1]$,

$$D(x; F) \leq D(\theta + \alpha(x - \theta); F).$$

Esto garantiza que la función refleje un ordenamiento centro-afuera consistente.

- d) **P4. Desvanecimiento en el Infinito** La profundidad debe tender a cero cuando la norma del punto x crece sin límite:

$$D(x; F) \rightarrow 0 \quad \text{a medida que} \quad \|x\| \rightarrow \infty, \quad \text{para cada} \quad F \in \mathcal{F},$$

donde \mathcal{F} es la clase de distribuciones en \mathbb{R}^d . Esto asegura que los puntos extremadamente alejados tengan una profundidad mínima.

Estas propiedades hacen que una función de profundidad sea útil para análisis multivariados no paramétricos, proporcionando robustez y un ordenamiento significativo.

Existen cuatro estructuras generales para construir funciones de profundidad estadística, clasificadas como Tipo A, B, C y D. Cada tipo se define a continuación con su formulación matemática y ejemplos representativos.

- **Tipo A** Se define como el promedio de una medida de cercanía acotada de x a una muestra aleatoria de tamaño r :

$$D(x; P) = E[h(x; X_1, \dots, X_r)],$$

donde $X_1, \dots, X_r \sim P$ y $h(x; x_1, \dots, x_r)$ es una función no negativa y acotada que mide la cercanía de x a los puntos x_1, \dots, x_r .

Ejemplos:

- Profundidad simplicial ($h = \mathbf{I}\{x \in S[X_1, \dots, X_{d+1}]\}$).
- Profundidad de mayoría.

- **Tipo B** Se basa en una medida de distancia no acotada $h(x; x_1, \dots, x_r)$, transformada para ser acotada:

$$D(x; F) = (1 + E[h(x; X_1, \dots, X_r)])^{-1},$$

donde $X_1, \dots, X_r \sim F$.

Ejemplos:

- Profundidad de volumen simplicial ($h = \Delta^\alpha(S[x, X_1, \dots, X_d])$).
- Profundidad L^p ($h = \|x - X_1\|_p$).

- **Tipo C** Se construye a partir de una medida de *outlyingness* (atipicidad) $O(x; F)$, también transformada:

$$D(x; F) = (1 + O(x; F))^{-1},$$

donde $O(x; F)$ mide cuán atípico es x respecto al centro o punto más profundo de F .

Ejemplos:

- Profundidad de proyección.
- Profundidad de Mahalanobis.

- **Tipo D** Se define como el ínfimo de la masa de probabilidad de conjuntos cerrados que contienen a x :

$$D(x; P, \mathcal{E}) = \inf\{P(C) \mid x \in C \in \mathcal{E}\},$$

donde \mathcal{E} es una clase de subconjuntos cerrados de \mathbb{R}^d .

Ejemplo:

- Profundidad de halfspace (con \mathcal{E} como la clase de halfspaces cerrados).

Estas estructuras ofrecen flexibilidad para diseñar funciones de profundidad adaptadas a diferentes contextos estadísticos.

Explicación de profundidades específicas:

Profundidad de Mahalanobis La profundidad de Mahalanobis mide la cercanía de un punto x al centro de una distribución usando la distancia de Mahalanobis. Se define como:

$$MHD(x; F) = (1 + d_{\Sigma(F)}^2(x, \mu(F)))^{-1},$$

donde:

- $d_{\Sigma(F)}^2(x, \mu(F)) = (x - \mu(F))' \Sigma(F)^{-1} (x - \mu(F))$ es la distancia de Mahalanobis.
- $\mu(F)$ es una medida de localización (e.g., media o mediana).
- $\Sigma(F)$ es una matriz de covarianza asociada a F .

Propiedades:

- **P1 (Invarianza Afín):** Satisfecha si μ y Σ son equivariantes afines, i.e., $\mu(AX + b) = A\mu(X) + b$ y $\Sigma(AX + b) = A\Sigma(X)A'$.
- **P2 (Maximalidad):** Cumplida si $\mu(F)$ coincide con el punto de simetría de F en distribuciones simétricas.
- **P3 (Monotonía):** Satisfecha bajo simetría, ya que $d_{\Sigma(F)}^2$ crece monotónicamente al alejarse de $\mu(F)$.
- **P4 (Desvanecimiento):** Cumplida trivialmente, pues $d_{\Sigma(F)}^2 \rightarrow \infty$ implica $MHD(x; F) \rightarrow 0$.

Ventajas y Limitaciones:

- *Ventajas:* Computacionalmente simple y usa información de covarianza.
- *Limitaciones:* Su robustez depende de μ y Σ ; con la media y covarianza clásicas, es sensible a outliers.

Profundidad de Half Space de Tukey

Propuesta por Tukey (1975), mide la profundidad como la mínima masa de probabilidad de cualquier halfspace cerrado que contenga a x :

$$HD(x; P) = \inf\{P(H) : H \text{ es un halfspace cerrado, } x \in H\}.$$

Un halfspace cerrado es un conjunto de la forma $H = \{y \in \mathbb{R}^d : u'y \geq u'x\}$, con $u \in \mathbb{R}^d$ un vector unitario.

Propiedades:

- **P1 (Invarianza Afín):** Satisfecha, ya que los halfspaces se preservan bajo transformaciones afines.
- **P2 (Maximalidad):** Cumplida para distribuciones H-simétricas (donde $P(H) \geq 1/2$ para todo H que contenga al centro θ).
- **P3 (Monotonía):** Satisfecha arbitrariamente, pues al alejarse de θ , algún halfspace tendrá menor masa.
- **P4 (Desvanecimiento):** Cumplida, ya que $P(H) \rightarrow 0$ para halfspaces que contienen puntos lejanos.

Ventajas:

- Robustez (punto de ruptura del mediano de halfspace es 1/3 en datasets típicos).
- Proporciona un ordenamiento centro-afuera efectivo sin supuestos paramétricos.

Profundidad Espacial

La profundidad espacial mide la centralidad de un punto x en \mathbb{R}^d basándose en su grado de alejamiento medido mediante cuantiles espaciales. Se define como:

$$D_s(x; F) = 1 - \|Q_F^{-1}(x)\|,$$

donde $Q_F^{-1}(x)$ es el vector de outlyingness que resuelve:

$$-E \left\{ \frac{X - x}{\|X - x\|} \right\} = u.$$

Aquí:

- $Q_F(u)$ es el cuantil espacial asociado al vector u (dirección y magnitud).
- $Q_F^{-1}(x)$ es el valor inverso de u que genera x .
- $D_s(x; F)$ toma valores entre 0 (muy lejano) y 1 (punto más central).

Propiedades:

- **P1 (Invarianza Afín):** Satisfecha; solo satisface invariancia ante traslaciones, rotaciones y escalamiento homogéneo.
- **P2 (Maximalidad en el Centro):** Cumplida; el valor máximo se alcanza en el mediano espacial $Q_F(0)$.
- **P3 (Monotonía):** Cumplida; decrece de forma monótona desde el centro hacia el exterior.
- **P4 (Desvanecimiento):** Cumplida; tiende a 0 a medida que $\|x\| \rightarrow \infty$.

Ventajas y Limitaciones:

- *Ventajas:*
 - Intuitiva y vinculada directamente a cuantiles.
 - Computacionalmente más eficiente que otras profundidades como Halfspace Depth.

5. Utilice el dataset wine usado en Hubert, M., & Debruyne, M. (2010) y encuentre:

- a) Las observaciones outliers usando los estimadores MCD. Explique sus resultados.
 Por medio de la base de datos **wine** de **scikit-learn** compuesta por 178 observaciones y 13 variables sobre vinos italianos se concentra el análisis bivariado en las características *Proline* y *Malic Acid*. El objetivo es comparar las elipses de tolerancia en el caso clásico (calculadas a partir de la media y la distancia de Mahalanobis) con las obtenidas mediante la metodología de Máxima Densidad de Covarianza (MCD). El resultado se ilustra en la Figura 1:

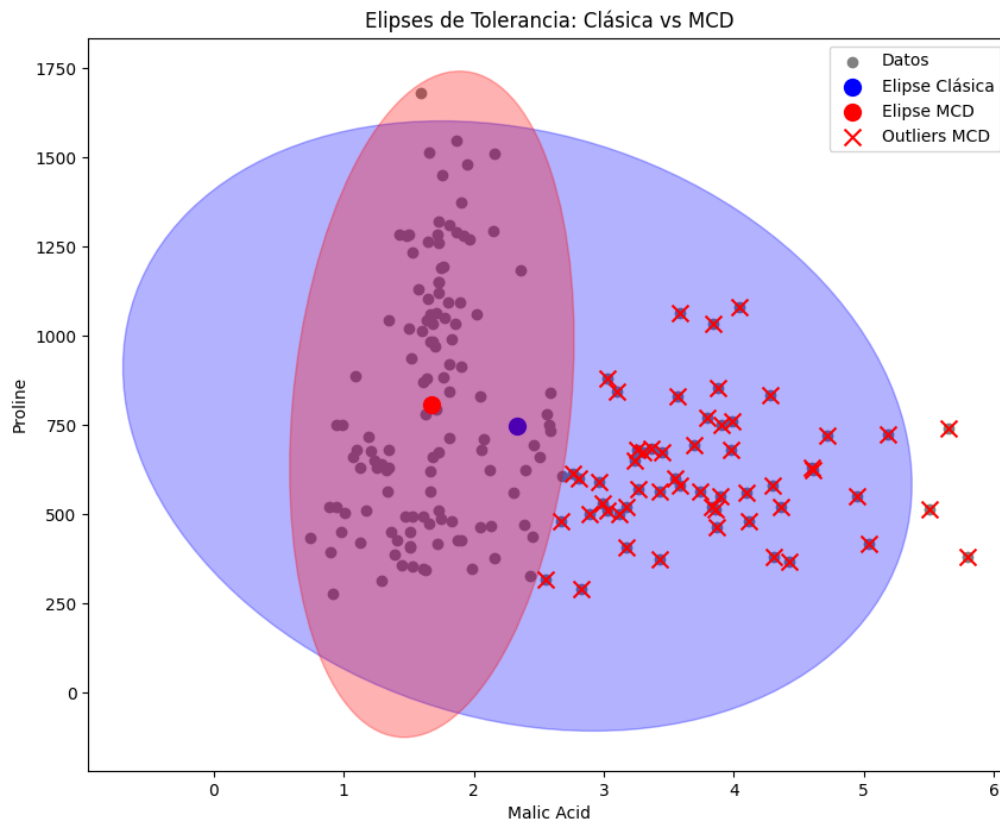


Figura 1: Comparación de las elipses de tolerancia clásica y la derivada mediante MCD para las variables Proline y Malic Acid.

Como se aprecia en la Figura 1, la elipse clásica presenta problemas de enmascaramiento (*masking*), al detectar correctamente únicamente cuatro valores atípicos. En cambio, la elipse de tolerancia obtenida mediante MCD identifica cincuenta y cinco valores atípicos y abarca un área notablemente menor. Estos resultados ponen de manifiesto la robustez del método MCD frente a la presencia de observaciones atípicas.

- b) Las observaciones outliers usando las profundidades de Mahalanobis, Half Space y Spatial. Explique sus resultados y compare con los del punto anterior.

Por medio de funciones de profundidad, también es posible detectar observaciones atípicas. Con este mismo conjunto de datos se aplican tres metodologías:

- I. **Profundidad de Mahalanobis:** Como se mencionó anteriormente, la profundidad de Mahalanobis es inversamente proporcional a la distancia de Mahalanobis y satisface las cuatro propiedades de una función de profundidad. Aplicando esta técnica con un umbral basado en el 5º percentil de las profundidades calculadas (profundidad $\approx 0,152$), se obtuvo la siguiente identificación de atípicos:

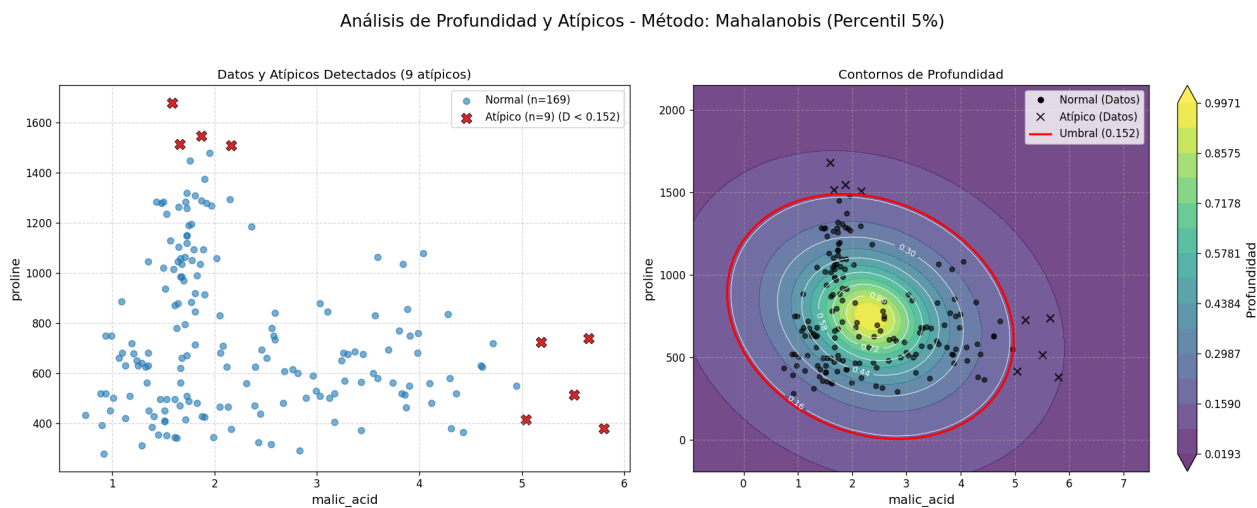


Figura 2: Valores atípicos detectados mediante la función de profundidad de Mahalanobis (Percentil 5%).

En la Figura 2 se observa que esta metodología identifica **nueve** valores atípicos (5.06 % del total). Este número supera a la detección basada en la distancia clásica de Mahalanobis (4 atípicos), pero difiere notablemente de los cincuenta y cinco atípicos hallados mediante MCD. Los datos señalados como atípicos se concentran principalmente en la región superior (valores altos de *proline*) y en la parte derecha del gráfico (valores altos de *malic_acid*). Los contornos de isoprofundidad elípticos, característicos de esta métrica basada en la media y covarianza, se visualizan en la subfigura derecha, donde la línea roja indica el umbral de profundidad mencionado.

II. **Profundidad Half Space:** La profundidad Half Space (o de Tukey), definida como la mínima masa de probabilidad de cualquier semiespacio cerrado que contenga al punto x , es una medida robusta y no paramétrica de centralidad. Su aplicación al conjunto de datos mediante una aproximación por proyección, usando un umbral del 5º percentil (profundidad $\approx 0,011$), resulta en la identificación de atípicos mostrada en la Figura 3.

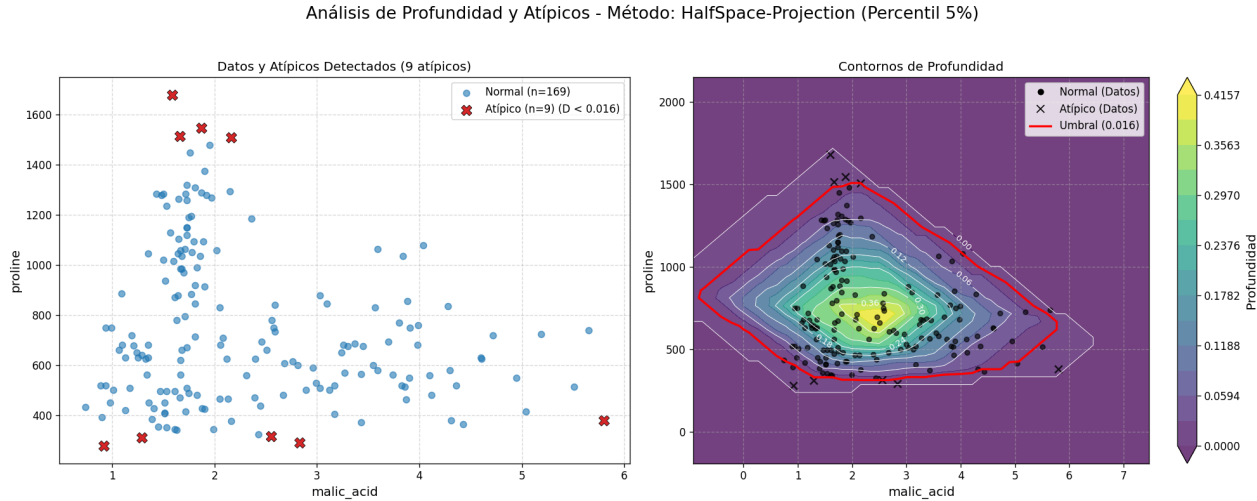


Figura 3: Valores atípicos detectados mediante la función de profundidad Half Space (aproximación por proyección, Percentil 5%).

En la Figura 3, se observa que la profundidad Half Space identifica **seis** valores atípicos (3.37 % del total). Este número es inferior al obtenido por Mahalanobis y Spatial en este caso, y también significativamente menor que los cincuenta y cinco identificados por MCD. Los puntos señalados como atípicos (marcados con 'X' rojas en la subfigura izquierda) se distribuyen en dos zonas principales: un grupo con valores elevados de *proline* (en la parte superior) y otro grupo con valores elevados de *malic_acid* (hacia la derecha).

La subfigura de la derecha visualiza los contornos de profundidad. A diferencia de los contornos elípticos de Mahalanobis, estos presentan formas poligonales irregulares, reflejando la naturaleza no paramétrica del método. La línea roja gruesa resalta el contorno umbral (profundidad $\approx 0,011$), separando la región central de las regiones exteriores consideradas atípicas según este criterio.

III. **Profundidad Espacial:** La profundidad espacial define la centralidad de un punto x como 1 menos la norma del vector promedio de direcciones unitarias desde los puntos de datos hacia x . Aplicando esta técnica con un umbral del 5º percentil (profundidad $\approx 0,100$), se obtiene la identificación de atípicos ilustrada en la Figura 4.

Análisis de Profundidad Espacial (Percentil 5%)

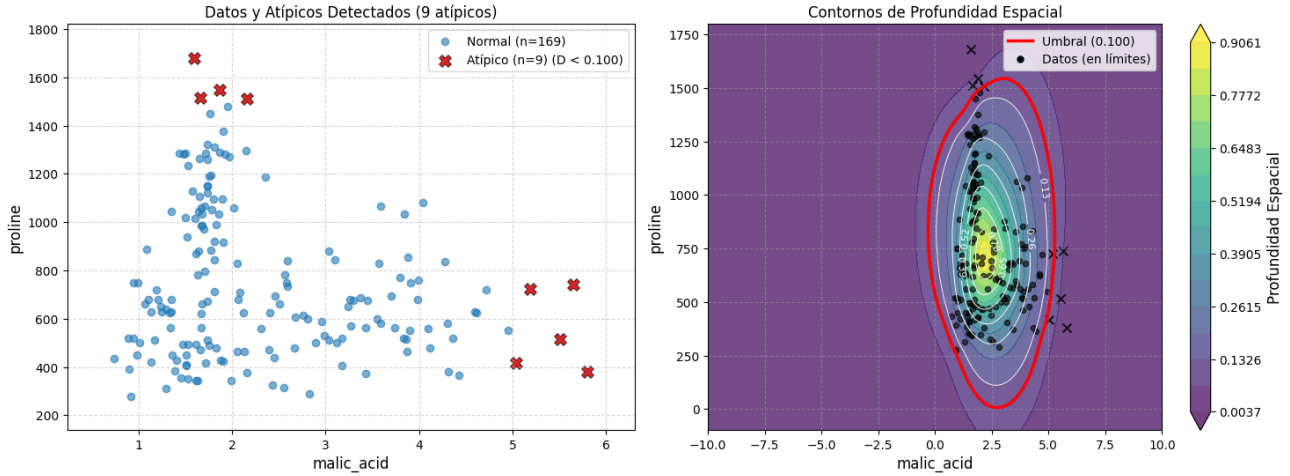


Figura 4: Valores atípicos detectados mediante la función de profundidad Espacial (Percentil 5%).

Como se muestra en la Figura 4, la profundidad espacial identifica **nueve** valores atípicos (5.06 %), coincidiendo exactamente con el número detectado por la profundidad de Mahalanobis en este análisis, pero difiriendo de los seis detectados por Half Space y los cincuenta y cinco por MCD. La localización de estos atípicos (marcados con 'X' rojas a la izquierda) es idéntica a la observada con Mahalanobis: puntos con alta *proline* y puntos con alta *malic_acid*.

La subfigura derecha presenta los contornos de isoprofundidad espacial. Estos contornos suelen ser suaves y convexos, reflejando la influencia de todos los puntos de datos y la geometría L2 subyacente en el cálculo del vector promedio de direcciones. En esta visualización específica, los contornos se adaptan a la distribución de los datos, y la línea roja gruesa indica el umbral (profundidad ≈ 0.100) que separa los puntos normales de los atípicos según este criterio.

IV. Comparación y Conclusiones Finales:

El análisis comparativo de la detección de atípicos revela diferencias entre las metodologías. Mientras que el método robusto MCD identificó un número muy alto de atípicos (55) y el método clásico solo 4, las funciones de profundidad evaluadas arrojaron resultados más moderados y diferenciados entre sí. Usando un umbral del 5^o percentil, las profundidades de Mahalanobis y Espacial identificaron cada una **nueve** observaciones (5.06 %) como atípicas, mientras que la profundidad Half Space (por proyección) identificó **seis** (3.37 %).

Un examen de las coincidencias, resumidas en la Tabla 1, muestra que los métodos Mahalanobis y Espacial identifican exactamente el **mismo conjunto de nueve puntos** en este caso. Sin embargo, la concordancia con Half Space es mucho menor: solo **tres** observaciones (1.69 %) son identificadas como atípicas simultáneamente por los tres métodos. Estos tres puntos representan la evidencia más fuerte de atipicidad según estas métricas. Half Space comparte 3 puntos con Mahalanobis y 3 puntos con Spatial, lo que indica que captura un subconjunto diferente de los puntos periféricos. Esta variación subraya que cada método captura aspectos distintos de la "atipicidad", debido a sus diferentes definiciones geométricas y los distintos valores umbrales resultantes (≈ 0.152 para Mahalanobis, ≈ 0.100 para Spatial, y ≈ 0.011 para HalfSpace). La visualización comparativa en la Figura 5 (especialmente el subgráfico inferior derecho) permite identificar qué puntos específicos son detectados por cada combinación de métodos.

Cuadro 1: Resumen de coincidencias en la identificación de atípicos (percentil 5 %) entre los diferentes métodos de profundidad (Resultados Actualizados).

Combinación de Métodos	Coincidencias	Proporción Coincidencia (%)
Mahalanobis & Spatial	9	5.06 %
Mahalanobis & HalfSpace-Projection	3	1.69 %
Spatial & HalfSpace-Projection	3	1.69 %
Todos los métodos (M & S & H)	3	1.69 %

Comparación de Outliers por Métodos de Profundidad (Percentil 5%)

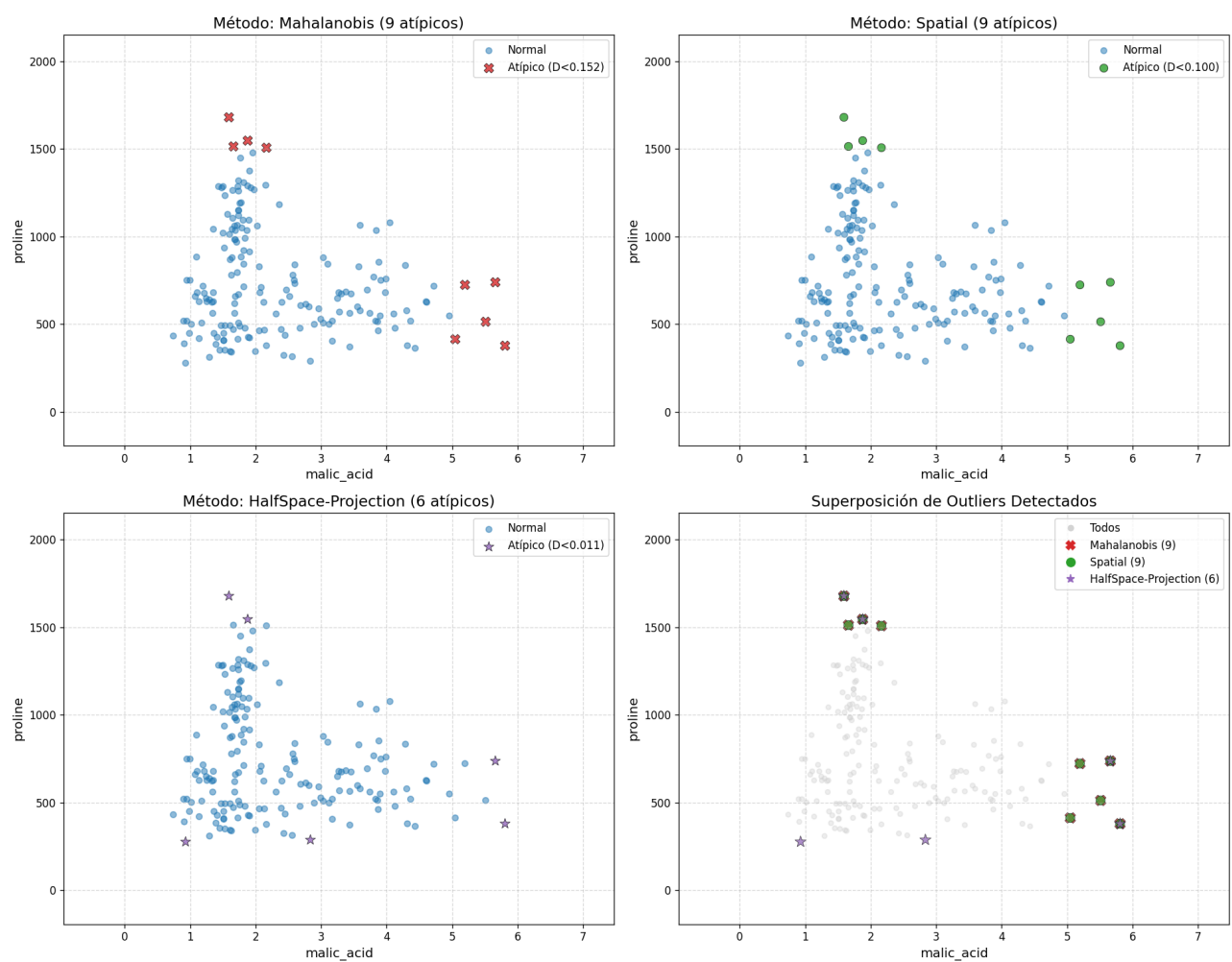


Figura 5: Comparación visual de los outliers detectados por los tres métodos de profundidad (Percentil 5 %).

En conclusión, la aplicación conjunta de varios métodos de profundidad y la evaluación de sus coincidencias, apoyada por herramientas visuales como la Figura 5, proporciona un entendimiento más completo y matizado de la estructura periférica de los datos que la aplicación de un único método. La alta concordancia entre Mahalanobis y Spatial en este caso sugiere que, para estos datos, sus sensibilidades son similares, mientras que Half Space ofrece una perspectiva diferente.

El código usado para generar estos resultados y figuras se encuentra en [código en](#)

Python.

6. Del libro de Ojeda Martínez de Castilla, I., & Gago Vargas, J. (2008), estudie la Proposición VI.1.2. (Página 161).
7. Demuestre que la matriz de covarianza muestral es singular en escenarios de alta dimensionalidad.

Demostración: Supongamos que los datos están centrados, como la matriz X tiene dimensión $n \times p$ donde $p > n$ (ya que estamos en un escenario de alta dimensionalidad), la matriz $X^T X$ tiene dimensión $p \times p$, ahora, veremos después que $r(X^T X) = r(X)$, y como $n < p$, entonces $r(X) \leq n$, de donde $r(X^T X) \leq n$, y como $n < p$, entonces, necesariamente $X^T X$ es singular.

Veamos que $r(X^T X) = r(X)$, por teorema de rango nulidad, tenemos que $r(X) + \dim(\ker(X)) = p$, y se cumple también que $r(X^T X) + \dim(\ker(X^T X)) = p$, por lo tanto, basta con mostrar que $\dim(\ker(X)) = \dim(\ker(X^T X))$, esto se tiene ya que, $\ker(X) = \ker(X^T X)$.

Veamos que $\ker(X) \subset \ker(X^T X)$:

si $v \in \ker(X)$, entonces $X^T X v = X^T (X v) = X^T 0 = 0$, luego, $v \in \ker(X^T X)$

Veamos que $\ker(X^T X) \subset \ker(X)$:

Si $v \in \ker(X^T X)$, entonces $X^T X v = 0$, así $v^T X^T X v = v^T 0 = 0$, y como $v^T X^T X v = \|X v\|^2$ se tiene que $\|X v\| = 0$, lo que implica que $X v = 0$, es decir $v \in \ker(X)$

8. Basado en el artículo de Engel, J., Buydens, L., & Blanchet, L. (2017):

- a) Explique por qué la matriz de covarianzas muestral no es un buen estimador de la dispersión en ambientes de alta dimensionalidad:

La matriz de covarianza muestral S pierde efectividad en entornos de alta dimensionalidad (cuando $p \gg n$ o incluso cuando $p \approx n$) debido a varias razones:

- **Pérdida de rango completo:** Si $p > n$, la matriz S no es de rango completo, lo que implica que varios de sus autovalores son cero. Esto hace que no sea invertible y, por tanto, inutilizable para muchas técnicas multivariadas (por ejemplo, PCA, MANOVA, LDA).
- **Sobreestimación y subestimación sistemática de autovalores:** En datos de alta dimensión, los autovalores grandes de S tienden a sobreestimarse y los pequeños a subestimarse, generando un sesgo importante. Este fenómeno se presenta incluso cuando n es 10 veces mayor que p .
- **Autovectores inconsistentes:** Las direcciones principales (autovectores) estimadas también pueden desviarse significativamente de las verdaderas, lo cual afecta directamente métodos como el PCA.
- **Alta varianza en las entradas:** Los elementos individuales s_{ij} y los de su inversa s_{ij}^{-1} presentan alta varianza, dificultando aplicaciones como el modelado de redes de correlación o de correlaciones parciales.

- b) Explique algunas propuestas para estimar la matriz de covarianza muestral en alta dimensionalidad:

Algunas propuestas para estimar la matriz de covarianza muestral en alta dimensionalidad son:

- 1) **Contracción de autovalores (Shrinkage)**

Este enfoque busca ajustar los autovalores de la matriz S hacia valores más estables, manteniendo los autovectores fijos. La idea general es:

$$\hat{\Sigma} = \sum_{i=1}^p \vartheta(\lambda_i; \delta) e_i e_i^\top,$$

donde λ_i y e_i son los autovalores y autovectores de S , y $\vartheta(\lambda_i; \delta)$ representa el valor contraído. Dado que no se modifica la dirección de los vectores propios se asume que las direcciones son las correctas.

Algunos métodos destacados:

- **Ledoit–Wolf (linear)**: contracción lineal hacia el promedio

$$\vartheta_{LW-LIN}(\lambda_i; \delta) = \delta \bar{\lambda} + (1 - \delta) \lambda_i.$$

está demostrado que usar este método garantiza que la estimación de la matriz sea positiva definida. Y es útil cuando los valores propios no son dispersos de lo contrario Ledoit y Wolf demostraron que es mejor los no lineales.

- **Contracción no lineal**:

- **CNR**: Este método busca controlar los extremos del espectro de autovalores de la matriz de covarianza mediante truncamiento. Para ello utiliza dos parámetros clave: κ_{\max} , el número de condición máximo permitido (estimado por validación cruzada), y τ^* , el autovalor mínimo permitido (estimado a partir de los datos). Los autovalores muestrales se contraen a este rango:

$$\vartheta_{CNR}(\lambda_i, \tau^*, \kappa_{\max}) = \begin{cases} \tau^*, & \lambda_i < \tau^* \\ \lambda_i, & \tau^* \leq \lambda_i \leq \kappa_{\max} \cdot \tau^* \\ \kappa_{\max} \cdot \tau^*, & \lambda_i > \kappa_{\max} \cdot \tau^* \end{cases}$$

- **CERNN**: Busca mezclar el metodo lineal y el CNR, es decir suaviza con la contracción lineal y controla los extremos

$$\vartheta_{CERNN}(\lambda_i; \delta) = \frac{-n + \sqrt{n^2 + 4\delta\alpha [n\lambda_i + \delta(1 - \alpha)]}}{2\delta\alpha}$$

donde δ se estima por validación cruzada y α por los datos.

- **LW-NONLIN**: Ledoit y Wolf, basándose en la teoría espectral de matrices aleatorias y la ecuación de Marčenko–Pastur, desarrollan la función QuEST (Quantized Eigenvalues Sampling Transform). Esta función permite estimar el espectro de la matriz de covarianza poblacional a partir de los autovalores muestrales. Aunque proporciona estimaciones de alta precisión, su implementación es computacionalmente costosa, ya que requiere resolver un problema de optimización no convexa.
- **NERCOME**: Este método es no paramétrico, computacionalmente rápido, y produce resultados asintóticamente equivalentes a **LW-NONLIN** cuando el tamaño muestral n es grande. Su estimación se construye dividiendo los datos en dos subconjuntos disjuntos de tamaños n_1 y n_2 , y se define como:

$$\hat{\Sigma} = P_1 \cdot \text{diag}(P_1^\top S_2 P_1) \cdot P_1^\top,$$

donde:

- P_1 es la matriz de autovectores de la matriz de covarianza S_1 calculada con los primeros n_1 datos.
- S_2 es la matriz de covarianza muestral calculada con los restantes n_2 datos.
- Por recomendación práctica, se toma $n_2 = \min\{30, \frac{n}{10}\}$.

Este procedimiento rompe la dependencia entre autovalores y autovectores, reduciendo el sesgo en la estimación y garantizando positividad definida.

Además, el desempeño del estimador puede mejorarse **promediando la matriz estimada sobre múltiples particiones aleatorias del conjunto de datos**. En lugar de usar una única partición (S_1, S_2), se generan varias divisiones del mismo tamaño y se calcula la estimación final como:

$$\hat{\Sigma}_{\text{NERCOME}} = \frac{1}{B} \sum_{b=1}^B \hat{\Sigma}^{(b)},$$

donde cada $\hat{\Sigma}^{(b)}$ se construye como la fórmula anterior a partir de la partición b -ésima. Este promedio reduce la varianza del estimador y mejora su estabilidad en entornos de alta dimensionalidad.

2) Estimadores tipo ridge

Una estrategia ampliamente utilizada en alta dimensión consiste en construir un estimador como una combinación convexa entre la matriz de covarianza muestral S y una matriz objetivo T :

$$\hat{\Sigma}(\delta) = \delta T + (1 - \delta)S,$$

Este enfoque busca balancear la varianza y el sesgo de la estimación mediante los siguientes componentes:

- **Matriz objetivo T** : Es un estimador de **baja varianza pero alto sesgo**, como la identidad escalada (cuando $T = \mathbb{I}$, el estimador coincide con la contracción lineal de Ledoit–Wolf), la matriz diagonal $\text{diag}(S)$, o una estructura de correlación común. La elección de T depende del conocimiento previo sobre la población.
- **Matriz muestral S** : Es insesgada, pero presenta **alta varianza** en contextos de alta dimensionalidad ($p \gg n$).
- **Parámetro de contracción δ** : Controla el peso de cada componente y puede estimarse mediante criterios de pérdida como la *norma de Frobenius*.

Una variante importante es el estimador tipo ridge propuesto por **van Wieringen et al.**, el cual introduce una forma cuadrática que permite garantizar positividad definida *sin necesidad de invertir matrices*. Su forma es:

$$\hat{\Sigma}_\delta = \delta I + \frac{1}{4}(S - \delta T)^2 + \frac{1}{2}(S - \delta T)$$

y su inversa puede calcularse explícitamente como:

$$\hat{\Sigma}_\delta^{-1} = \frac{1}{\delta} \left(\hat{\Sigma}_\delta - (S - \delta T) \right).$$

Este esquema ofrece ventajas computacionales y garantiza estabilidad espectral del estimador.

3) Thresholding (Umbralización)

Supone que muchas entradas de la matriz son cero (dispersión). Se estiman las entradas de la matriz aplicando operadores de umbral, de manera general se estima la matriz de varianzas y covarianzas de la siguiente manera:

$$\hat{\Sigma}(\delta) = \arg \min_{\Sigma} \left\{ \frac{1}{2} \|S - \Sigma\|_F^2 + p_\delta(\Sigma) \right\}$$

y entrada por entrada se puede escribir así:

$$\hat{\sigma}_{ij} = \arg \min_{\sigma_{ij}} \left\{ \frac{1}{2} (s_{ij} - \sigma_{ij})^2 + p_\delta(\sigma_{ij}) \right\}$$

Nota: si la penalización $p_\delta(\sigma_{ij})$ es muy pequeña (por ejemplo, cuando $p_\delta(\sigma_{ij}) \rightarrow 0$), entonces la estimación regularizada $\hat{\sigma}_{ij}$ se aproxima al valor muestral s_{ij} .

Existen 4 variantes como:

- **Hard thresholding**: Elimina por completo los elementos menores a cierto valor.
- **Soft thresholding**: Además de eliminar reduce en magnitud los elementos que superan el umbral.
- **SCAD**: penalización no convexa para reducir sesgo.
- **Adaptive Lasso**: Umbral adaptativo según s_{ij} .

En el cuadro 1 se puede ver la penalización de cada método y en el cuadro 2 la solución de cada método.

Tipo de thresholding	Penalización $p_\delta(\sigma_{ij})$
Hard thresholding	$\frac{\delta^2 - (\sigma_{ij} - \delta)^2}{2} \cdot \mathbf{1}_{ \sigma_{ij} < \delta}$
Soft thresholding	$\delta \sigma_{ij} $
SCAD	$\begin{cases} \delta \sigma_{ij} , & \sigma_{ij} \leq \delta \\ \frac{-\sigma_{ij}^2 + 2a\delta \sigma_{ij} - \delta^2}{2(a-1)}, & \delta < \sigma_{ij} \leq a\delta \\ \frac{(a+1)\delta^2}{2}, & \sigma_{ij} > a\delta \end{cases}$
Adaptive LASSO	$\delta \cdot w(s_{ij}) \cdot \sigma_{ij} $

Cuadro 2: Penalizaciones $p_\delta(\sigma_{ij})$ utilizadas en operadores de umbralización

Tipo de thresholding	Solución $\hat{\sigma}_{ij}(s_{ij}, \delta)$
Hard thresholding	$s_{ij} \cdot \mathbf{1}_{ s_{ij} > \delta}$
Soft thresholding	$\text{sgn}(s_{ij}) \cdot (s_{ij} - \delta)^+$
SCAD	$\begin{cases} \text{sgn}(s_{ij}) \cdot (s_{ij} - \delta)^+, & s_{ij} \leq 2\delta \\ \frac{(a-1)s_{ij} - \text{sgn}(s_{ij})a\delta}{a-2}, & 2\delta < s_{ij} \leq a\delta \\ s_{ij}, & s_{ij} > a\delta \end{cases}$
Adaptive LASSO	$\text{sgn}(s_{ij}) \cdot (s_{ij} - \delta w(s_{ij}) s_{ij})^+$

Cuadro 3: Operadores de thresholding $\hat{\sigma}(s_{ij}, \delta)$ para estimación dispersa

9. Basado en el paper de Boudt, K., Rousseeuw, P. J., Vanduffel, S., & Verdonck, T. (2020) use la metodología MRCD sobre el dataset VDP y encuentre:

- a) Las distancias T_{MRCD}^2 para 3 observaciones seleccionadas arbitrariamente. Explique sus resultados.

Resultados: En el siguiente cuadro se muestran las distancias de tres tableros

Tablero	T_{MRCD}^2
A_1	412.4262
A_4	55.67081
C_4	55.07384

Cuadro 4: Distancias T_{MRCD}^2 para tres observaciones del dataset VDP

De la anterior tabla se puede decir que los tableros A_4 y C_4 están más cerca del centro que la tabla A_1 . El código de la solución de este ejercicio está en la sección .Ejercicio 9 del taller 1A” del link: [Ejercicio 9 en Quarto](#).

- b) Las observaciones outliers de este dataset.

Resultados: Siguiendo el procedimiento recomendado en el artículo guía, se obtuvo la Figura 1, la cual me muestra el cuadrado de las distancias robustas con el umbral $\chi_{p,0,97}^2$. Claramente se observa que hay un efecto de masking en 2 tableros, que se pueden considerar datos atípicos; esto se puede deber a que $p \gg n$, pues $p = 314$ y $n = 23$. Incluso, los mismos autores del artículo mencionan este inconveniente en la página 125.

“Our experiments have shown that the usual chi-square and F-distribution results for the MCD distances (Hardin and Rocke 2005) are no longer good approximations when p is large relatively to n . A better approximation would be useful for improving the accuracy of the MRCD by reweighting.”

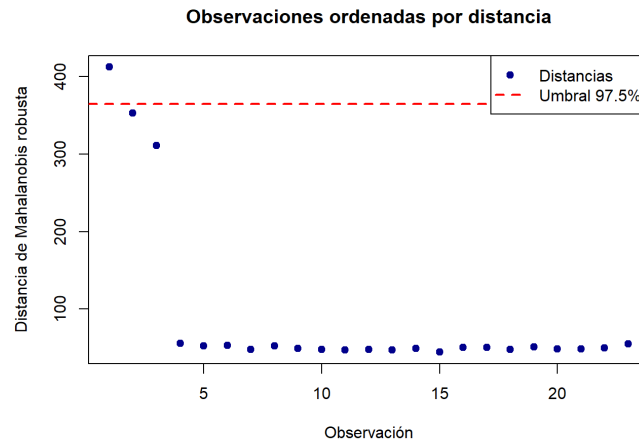


Figura 6: Distancia de mahalanobie robusta usando MRCD y umbral $\chi^2_{p,0,97}$.

Según la gráfica, el único dato atípico es el tablero A_1 ; sin embargo, por el efecto masking, está ocultando a los tableros A_2 y A_3 que pueden ser considerados datos atípicos. Puede ver todo el análisis en el siguiente link [Ejercicio 9 en Quarto](#).

10. Utilice la tabla 1.2 (Paper Quality) del libro de Johnson & Wichern (2014):

a) Verifique si en este conjunto de datos hay observaciones outliers.

Aplicando la distancia clásica de Mahalanobis identificamos un único punto outliers, dado por la observación 25, $(0,971, 126,10, 72,10)^T$, usando el umbral 0.975 para la distribución χ^2_3 .

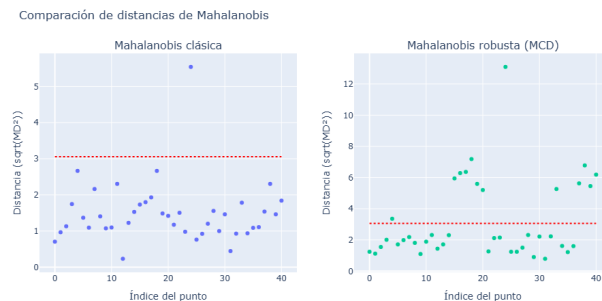


Figura 7: Comparación detección de outliers

Si usamos el los estimadores robustos basados en MCD, entonces encontramos 13 observaciones outliers, correspondientes a las observaciones outlier (basado en MCD): $[5, 16, 17, 18, 19, 20, 21, 25, 34, 38, 39, 40, 41]$, correspondiente a los puntos

Specimen	Density	Machine_direction	Cross_direction
5	0.840	135.10	71.21
16	0.796	109.81	53.10
17	0.759	109.10	50.85
18	0.770	115.10	51.68
19	0.759	118.31	50.60
20	0.772	112.60	53.51
21	0.806	116.20	56.53
25	0.971	126.10	72.10
34	0.782	107.40	54.42
38	0.788	110.60	53.52
39	0.772	103.51	48.93
40	0.776	110.71	53.67
41	0.758	113.80	52.42

Cuadro 5: Filas detectadas como outliers (basado en MCD).

Ademas de esto aplicamos la profundidad de Tukey para detectar observaciones outliers, y obtuvimos las siguientes observaciones como outliers:

Specimen	Density	Machine_direction	Cross_direction
3	0.841	129.20	78.20
5	0.840	135.10	71.21
6	0.842	131.50	78.39
16	0.796	109.81	53.10
18	0.770	115.10	51.68
21	0.806	116.20	56.53
23	0.845	131.00	74.35
25	0.971	126.10	72.10
31	0.843	123.90	78.54
35	0.795	120.70	70.41
36	0.805	121.91	73.68
39	0.772	103.51	48.93
40	0.776	110.71	53.67

Cuadro 6: Filas detectadas como outliers (profundidad de Tukey < 0.2195).

- b) Encuentre estimadores apropiados de localización y dispersión para el vector aleatorio que explica la calidad del papel.

De acuerdo a los resultados encontrados, decidimos usar los estimadores de localización y dispersión dados por el estimador MCD.

$$\text{Media robusta (MCD)} : \quad \mu_{\text{MCD}} = \begin{bmatrix} 0,8192 \\ 123,9554 \\ 73,3611 \end{bmatrix}$$

$$\text{Matriz de covarianza robusta (MCD)} : \quad \Sigma_{\text{MCD}} = \begin{bmatrix} 0,00023 & 0,04490 & 0,03264 \\ 0,04490 & 26,5837 & 11,1179 \\ 0,03264 & 11,1179 & 13,3197 \end{bmatrix}$$

También podemos ver el gráfico de la elipse umbral para la distribución χ_3^2 usando el estimador robusto dado por MCD.

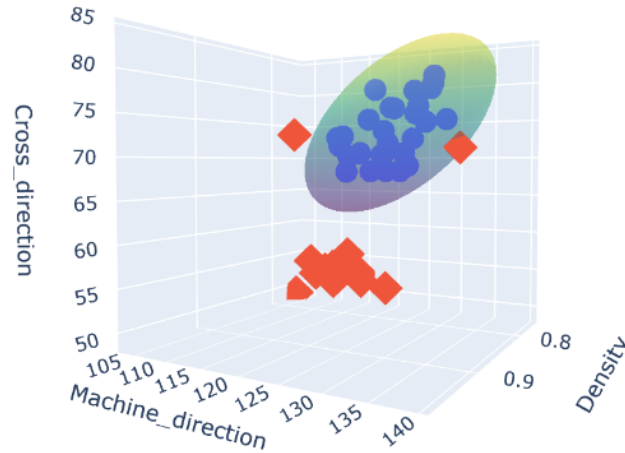


Figura 8: Elipse umbral

[Abrir notebook en Google Colab](#)

Solución: Taller 1 Segunda parte

1. Del libro de Johnson, R. A., & Wichern, D. W. (2014), resolver los siguientes ejercicios:

2.16 Sea \mathbf{V} una variable aleatoria vectorial con vector media $E(\mathbf{V}) = \boldsymbol{\mu}_v$, y matriz de covarianza

$$E[(\mathbf{V} - \boldsymbol{\mu}_v)(\mathbf{V} - \boldsymbol{\mu}_v)'] = \boldsymbol{\Sigma}_v.$$

Demuestre que

$$E(\mathbf{V}\mathbf{V}') = \boldsymbol{\Sigma}_v + \boldsymbol{\mu}_v\boldsymbol{\mu}_v'.$$

Solución:

Como $E[(\mathbf{V} - \boldsymbol{\mu}_v)(\mathbf{V} - \boldsymbol{\mu}_v)'] = \boldsymbol{\Sigma}_v$, entonces

$$\begin{aligned} E[(\mathbf{V} - \boldsymbol{\mu}_v)(\mathbf{V} - \boldsymbol{\mu}_v)'] &= \boldsymbol{\Sigma}_v \\ E[(\mathbf{V} - \boldsymbol{\mu}_v)(\mathbf{V}' - \boldsymbol{\mu}_v')] &= \\ E[\mathbf{V}\mathbf{V}' - \boldsymbol{\mu}_v\mathbf{V}' - \mathbf{V}\boldsymbol{\mu}_v' + \boldsymbol{\mu}_v\boldsymbol{\mu}_v'] &= \end{aligned}$$

por propiedades del valor esperado de un vector aleatorio, se tiene

$$\begin{aligned} E[\mathbf{V}\mathbf{V}' - \boldsymbol{\mu}_v\mathbf{V}' - \mathbf{V}\boldsymbol{\mu}_v' + \boldsymbol{\mu}_v\boldsymbol{\mu}_v'] &= \boldsymbol{\Sigma}_v \\ E[\mathbf{V}\mathbf{V}'] - E[\boldsymbol{\mu}_v\mathbf{V}'] - E[\mathbf{V}\boldsymbol{\mu}_v'] + E[\boldsymbol{\mu}_v\boldsymbol{\mu}_v'] &= \\ E[\mathbf{V}\mathbf{V}'] - \boldsymbol{\mu}_v\boldsymbol{\mu}_v' - \boldsymbol{\mu}_v\boldsymbol{\mu}_v' + \boldsymbol{\mu}_v\boldsymbol{\mu}_v' &= \\ E[\mathbf{V}\mathbf{V}'] - \boldsymbol{\mu}_v\boldsymbol{\mu}_v' &= \end{aligned}$$

Al despejar $E[\mathbf{V}\mathbf{V}']$ se tiene

$$E[\mathbf{V}\mathbf{V}'] = \boldsymbol{\Sigma}_v + \boldsymbol{\mu}_v\boldsymbol{\mu}_v'.$$

3.13 Demostrar que $|\mathbf{Q}| = +1$ o -1 si \mathbf{Q} es una matriz ortogonal de $p \times p$.

Demostración: Sea \mathbf{Q} una matriz cuadrada de tamaño $p \times p$.

Por definición, una matriz \mathbf{Q} es ortogonal si $\mathbf{Q}' = \mathbf{Q}^{-1}$. Esto implica:

$$\mathbf{Q}\mathbf{Q}' = \mathbf{I}$$

donde \mathbf{I} es la matriz identidad de tamaño $p \times p$.

Tomamos el determinante a ambos lados:

$$|\mathbf{Q}\mathbf{Q}'| = |\mathbf{I}|$$

Usamos: $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$, $|\mathbf{I}| = 1$, y $|\mathbf{Q}'| = |\mathbf{Q}|$.

$$|\mathbf{Q}||\mathbf{Q}'| = |\mathbf{I}|$$

$$|\mathbf{Q}||\mathbf{Q}| = 1$$

$$|\mathbf{Q}|^2 = 1$$

Sea $x = |\mathbf{Q}|$. La ecuación $x^2 = 1$ tiene como únicas soluciones reales $x = +1$ y $x = -1$. Por lo tanto, el determinante de una matriz ortogonal \mathbf{Q} debe ser $+1$ o -1 .

$$\boxed{|\mathbf{Q}| = \pm 1}$$

3.12 Demostrar que el determinante de una matriz simétrica cuadrada A de tamaño $p \times p$ puede expresarse como el producto de sus valores propios $\lambda_1, \lambda_2, \dots, \lambda_p$; es decir,

$$|A| = \prod_{i=1}^p \lambda_i.$$

Por el teorema espectral como A es simétrica entonces se puede expresar como $A = P\Lambda P'$, donde P es una matriz cuadrada con vectores ortonormales, de aquí tenemos que

$$|A| = |P\Lambda P'| = |P||\Lambda P'| = |P||\Lambda||P'| = |\Lambda||I|,$$

esto ya que como P es ortonormal entonces su inversa es su transpuesta y $P'P = I$

3.19 Sea

$$\mathbf{A}_{(m \times m)}^{1/2} = \sum_{i=1}^m \sqrt{\lambda_i} \mathbf{e}_i \mathbf{e}_i' = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}',$$

donde $\mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \mathbf{I}$. (Los λ_i y los \mathbf{e}_i son los valores propios y los vectores propios normalizados de la matriz \mathbf{A} .)

Demuestre las propiedades (1)–(4) de la matriz raíz cuadrada en la ecuación (3-22).

Solución:

a) $A^{1/2}$ es simétrico:

$$(A^{1/2})' = (\mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}')' = \mathbf{P}(\mathbf{\Lambda}^{1/2})'\mathbf{P}'$$

como $\mathbf{\Lambda}^{1/2}$ es diagonal, entonces es simétrica y luego

$$(A^{1/2})' = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}' = A^{1/2}.$$

b) $A^{1/2}A^{1/2} = A$: Se mostrará por medio de igualdades

$$\begin{aligned} A^{1/2}A^{1/2} &= \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}'\mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}' \\ &= \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{P}' \\ &= \mathbf{P} \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_m} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_m} \end{bmatrix} \mathbf{P}' \\ &= \mathbf{P} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix} \mathbf{P}' \\ &= \mathbf{P}\mathbf{\Lambda}\mathbf{P}' = A \end{aligned}$$

c) $(A^{1/2})^{-1} = \mathbf{P}A^{-1/2}\mathbf{P}'$: Por medio de igualdades se obtiene, de la siguiente manera :

$$(A^{1/2})^{-1} = (\mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}')^{-1}$$

dado que $\mathbf{P}\mathbf{P}' = \mathbf{I}$, la matriz inversa de $A^{1/2}$ es $A^{-1/2}$ es decir

$$\begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_m} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\lambda_m}} \end{bmatrix} = \mathbf{I}$$

entonces

$$(A^{1/2})^{-1} = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}'$$

d) $A^{1/2}A^{-1/2} = A^{-1/2}A^{1/2} = \mathbf{I}$: Por el inciso anterior se tiene que $A^{-1/2}$ es matriz inversa de $A^{1/2}$, entonces por propiedades de matriz inversa, se tiene la igualdad requerida.

3.20 Determine la matriz raíz cuadrada $\mathbf{A}^{1/2}$, usando la matriz \mathbf{A} del Ejercicio 3.3. También, determine $\mathbf{A}^{-1/2}$ y demuestre que $\mathbf{A}^{1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1/2}\mathbf{A}^{1/2} = \mathbf{I}$.

Solución:

La matriz \mathbf{A} dada en el Ejercicio 3.3 es:

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

Para encontrar $\mathbf{A}^{1/2}$ y $\mathbf{A}^{-1/2}$, utilizamos la descomposición espectral de \mathbf{A} , ya que es una matriz simétrica definida positiva (sus valores propios son positivos como se verá). La descomposición espectral es $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$, donde \mathbf{P} es una matriz ortogonal cuyas columnas son los vectores propios normalizados de \mathbf{A} , y $\mathbf{\Lambda}$ es la matriz diagonal de los correspondientes valores propios.

Calculando valores y vectores propios de \mathbf{A} : La ecuación característica es

$$|\mathbf{A} - \lambda\mathbf{I}| = (2 - \lambda)(3 - \lambda) - 1 = \lambda^2 - 5\lambda + 5 = 0$$

Los valores propios son $\lambda_1 = \frac{5+\sqrt{5}}{2}$ y $\lambda_2 = \frac{5-\sqrt{5}}{2}$.

$$\text{Para } \lambda_1: (\mathbf{A} - \lambda_1\mathbf{I})\mathbf{e}_1 = \mathbf{0} \implies \begin{bmatrix} \frac{-1-\sqrt{5}}{2} & 1 \\ 1 & \frac{1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} e_{11} \\ e_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Esto da $e_{12} = \frac{1+\sqrt{5}}{2}e_{11}$.

Un vector propio es $\begin{bmatrix} 2 \\ 1 + \sqrt{5} \end{bmatrix}$.

$$\text{Para } \lambda_2: (\mathbf{A} - \lambda_2\mathbf{I})\mathbf{e}_2 = \mathbf{0} \implies \begin{bmatrix} \frac{-1+\sqrt{5}}{2} & 1 \\ 1 & \frac{1+\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} e_{21} \\ e_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Esto da $e_{22} = \frac{1-\sqrt{5}}{2}e_{21}$.

Un vector propio es $\begin{bmatrix} 2 \\ 1 - \sqrt{5} \end{bmatrix}$.

Los vectores propios normalizados son (después de calcular las normas $\|v_1\|^2 = 10 + 2\sqrt{5}$ y $\|v_2\|^2 = 10 - 2\sqrt{5}$):

$$\mathbf{e}_1 = \frac{1}{\sqrt{10 + 2\sqrt{5}}} \begin{bmatrix} 2 \\ 1 + \sqrt{5} \end{bmatrix} \quad \text{y} \quad \mathbf{e}_2 = \frac{1}{\sqrt{10 - 2\sqrt{5}}} \begin{bmatrix} 2 \\ 1 - \sqrt{5} \end{bmatrix}$$

(Se puede verificar que $\mathbf{e}_1' \mathbf{e}_2 = 0$).

Construir las matrices \mathbf{P} , $\mathbf{\Lambda}^{1/2}$ y $\mathbf{\Lambda}^{-1/2}$:

$$\mathbf{P} = [\mathbf{e}_1 \quad \mathbf{e}_2] = \begin{bmatrix} \frac{2}{\sqrt{10+2\sqrt{5}}} & \frac{2}{\sqrt{10-2\sqrt{5}}} \\ \frac{1+\sqrt{5}}{\sqrt{10+2\sqrt{5}}} & \frac{1-\sqrt{5}}{\sqrt{10-2\sqrt{5}}} \end{bmatrix}$$

$$\mathbf{\Lambda}^{1/2} = \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{5+\sqrt{5}}{2}} & 0 \\ 0 & \sqrt{\frac{5-\sqrt{5}}{2}} \end{bmatrix}$$

$$\mathbf{\Lambda}^{-1/2} = \begin{bmatrix} 1/\sqrt{\lambda_1} & 0 \\ 0 & 1/\sqrt{\lambda_2} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{2}{5+\sqrt{5}}} & 0 \\ 0 & \sqrt{\frac{2}{5-\sqrt{5}}} \end{bmatrix}$$

Definir $\mathbf{A}^{1/2}$ y $\mathbf{A}^{-1/2}$: La matriz raíz cuadrada es $\mathbf{A}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}'$. La inversa de la raíz cuadrada es $\mathbf{A}^{-1/2} = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}'$.

Verificar $\mathbf{A}^{1/2}\mathbf{A}^{-1/2} = \mathbf{I}$ y $\mathbf{A}^{-1/2}\mathbf{A}^{1/2} = \mathbf{I}$:

Ahora calculamos el producto:

$$\begin{aligned} \mathbf{A}^{1/2}\mathbf{A}^{-1/2} &= (\mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}')(\mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}') \\ &= \mathbf{P}\mathbf{\Lambda}^{1/2}(\mathbf{P}'\mathbf{P})\mathbf{\Lambda}^{-1/2}\mathbf{P}' \quad (\text{asociatividad}) \\ &= \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{I}\mathbf{\Lambda}^{-1/2}\mathbf{P}' \quad (\text{ya que } \mathbf{P}'\mathbf{P} = \mathbf{I}) \\ &= \mathbf{P}(\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{-1/2})\mathbf{P}' \quad (\text{propiedad de } \mathbf{I} \text{ y asociatividad}) \\ &= \mathbf{P}\mathbf{I}\mathbf{P}' \quad (\text{propiedad de } \mathbf{\Lambda}^{1/2}, \mathbf{\Lambda}^{-1/2}) \\ &= \mathbf{P}\mathbf{P}' = \mathbf{I} \quad (\text{ya que } \mathbf{P}\mathbf{P}' = \mathbf{I}) \end{aligned}$$

El cálculo para $\mathbf{A}^{-1/2}\mathbf{A}^{1/2}$ es análogo:

$$\begin{aligned} \mathbf{A}^{-1/2}\mathbf{A}^{1/2} &= (\mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}')(\mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}') \\ &= \mathbf{P}\mathbf{\Lambda}^{-1/2}(\mathbf{P}'\mathbf{P})\mathbf{\Lambda}^{1/2}\mathbf{P}' \\ &= \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{I}\mathbf{\Lambda}^{1/2}\mathbf{P}' \\ &= \mathbf{P}(\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{1/2})\mathbf{P}' \\ &= \mathbf{P}\mathbf{I}\mathbf{P}' = \mathbf{P}\mathbf{P}' = \mathbf{I} \end{aligned}$$

Queda demostrado que $\mathbf{A}^{1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1/2}\mathbf{A}^{1/2} = \mathbf{I}$ basándose en las propiedades de la descomposición espectral y las matrices ortogonales.

3.21 Dada la matriz

$$A = \begin{bmatrix} 1 & 1 \\ 2 & -2 \\ 2 & 2 \end{bmatrix}$$

- Calcular $A'A$ y obtener sus valores propios y vectores propios.
- Calcular AA' y obtener sus valores propios y vectores propios. Verificar que los valores propios no nulos son los mismos que en el inciso (a).
- Obtener la descomposición en valores singulares (SVD) de A .

Solución (a)

Calculamos $A'A$:

$$A'A = \begin{bmatrix} 1 & 2 & 2 \\ 1 & -2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & -2 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 9 & 1 \\ 1 & 9 \end{bmatrix}$$

Ecuación característica:

$$\det(A'A - \lambda I) = \begin{vmatrix} 9 - \lambda & 1 \\ 1 & 9 - \lambda \end{vmatrix} = (9 - \lambda)^2 - 1 = 0$$

Valores propios: $\lambda_1 = 8$, $\lambda_2 = 10$.

Vectores propios correspondientes:

- Para $\lambda_1 = 8$: $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$
- Para $\lambda_2 = 10$: $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Solución (b)

Calculamos AA' :

$$AA' = \begin{bmatrix} 1 & 1 \\ 2 & -2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \\ 1 & -2 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 4 \\ 0 & 8 & 0 \\ 4 & 0 & 8 \end{bmatrix}$$

Ecuación característica:

$$\det(AA' - \lambda I) = -\lambda^3 + 18\lambda^2 - 80\lambda = -\lambda(\lambda^2 - 18\lambda + 80) = 0$$

Valores propios:

$$\lambda_1 = 0, \lambda_2 = 8, \lambda_3 = 10$$

Vectores propios:

- Para $\lambda_1 = 0$:

$$\begin{bmatrix} 2 & 0 & 4 \\ 0 & 8 & 0 \\ 4 & 0 & 8 \end{bmatrix} \mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v}_1 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

- Para $\lambda_2 = 8$:

$$\begin{bmatrix} -6 & 0 & 4 \\ 0 & 0 & 0 \\ 4 & 0 & 0 \end{bmatrix} \mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

- Para $\lambda_3 = 10$:

$$\begin{bmatrix} -8 & 0 & 4 \\ 0 & -2 & 0 \\ 4 & 0 & -2 \end{bmatrix} \mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

Se confirma que los valores propios no nulos (8 y 10) coinciden con los obtenidos en (a).

Solución (c)

1. Valores singulares:

$$\sigma_1 = \sqrt{10}, \sigma_2 = \sqrt{8}$$

2. Matriz V (de $A'A$):

$$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

3. Matriz U (de AA'):

$$U = \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 \\ 0 & 1 \\ \frac{2}{\sqrt{5}} & 0 \end{bmatrix}$$

4. Descomposición SVD:

$$A = U\Sigma V' = \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 \\ 0 & 1 \\ \frac{2}{\sqrt{5}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{10} & 0 \\ 0 & \sqrt{8} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

3.24 Sea \mathbf{X} una variable aleatoria con matriz de covarianza

$$\Sigma = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Hallar:

(a) Σ^{-1} : Note que

$$\begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/9 & 0 \\ 0 & 0 & 1/1 \end{bmatrix} = \mathbf{I}$$

$$\text{entonces } \Sigma^{-1} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/9 & 0 \\ 0 & 0 & 1/1 \end{bmatrix}$$

(b) Los valores propios y vectores propios de Σ : Por definición de valor y vector propio

$$\begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix} v = \lambda v$$

es fácil ver que:

$$\begin{aligned} \bullet & \begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ \bullet & \begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = 9 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\ \bullet & \begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 1 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \end{aligned}$$

(c) Los valores propios y vectores propios de Σ^{-1} .
: Por definición de valor y vector propio

$$\begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/9 & 0 \\ 0 & 0 & 1 \end{bmatrix} v = \lambda v$$

es fácil ver que:

$$\begin{aligned} \bullet & \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/9 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 1/4 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ \bullet & \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/9 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = 1/9 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\ \bullet & \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/9 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 1 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \end{aligned}$$

3.25 Sea \mathbf{X} un vector aleatorio con matriz de covarianza

$$\Sigma = \begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix}$$

(a) Determine ρ y $\mathbf{V}^{1/2}$.

(b) Multiplique sus matrices para verificar la relación $\mathbf{V}^{1/2}\rho\mathbf{V}^{1/2} = \Sigma$.

Solución:

(a) **Determinación de $\mathbf{V}^{1/2}$ y ρ**

La matriz $\mathbf{V}^{1/2}$ es una matriz diagonal que contiene las desviaciones estándar de las componentes de \mathbf{X} en su diagonal principal. Las varianzas son los elementos diagonales de Σ :

$$\sigma_{11} = 25, \quad \sigma_{22} = 4, \quad \sigma_{33} = 9$$

Las desviaciones estándar son las raíces cuadradas de las varianzas:

$$\sqrt{\sigma_{11}} = \sqrt{25} = 5$$

$$\sqrt{\sigma_{22}} = \sqrt{4} = 2$$

$$\sqrt{\sigma_{33}} = \sqrt{9} = 3$$

Por lo tanto, la matriz diagonal de desviaciones estándar es:

$$\mathbf{V}^{1/2} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

La matriz de correlación ρ se obtiene a partir de la matriz de covarianza Σ usando la relación $\rho_{ik} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}}$. Los elementos diagonales de ρ son siempre 1. Para los elementos fuera de la diagonal:

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} = \frac{-2}{5 \times 2} = \frac{-2}{10} = -\frac{1}{5} = -0,2$$

$$\rho_{13} = \frac{\sigma_{13}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{33}}} = \frac{4}{5 \times 3} = \frac{4}{15}$$

$$\rho_{23} = \frac{\sigma_{23}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{33}}} = \frac{1}{2 \times 3} = \frac{1}{6}$$

Dado que ρ es simétrica ($\rho_{ik} = \rho_{ki}$), la matriz de correlación es:

$$\rho = \begin{bmatrix} 1 & -1/5 & 4/15 \\ -1/5 & 1 & 1/6 \\ 4/15 & 1/6 & 1 \end{bmatrix}$$

(b) **Verificación de la relación $\mathbf{V}^{1/2}\rho\mathbf{V}^{1/2} = \Sigma$**

Ahora multiplicamos las matrices obtenidas en la parte (a):

$$\begin{aligned}
\mathbf{V}^{1/2} \boldsymbol{\rho} \mathbf{V}^{1/2} &= \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1/5 & 4/15 \\ -1/5 & 1 & 1/6 \\ 4/15 & 1/6 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \\
&= \begin{bmatrix} 5(1) + 0 + 0 & 5(-1/5) + 0 + 0 & 5(4/15) + 0 + 0 \\ 0 + 2(-1/5) + 0 & 0 + 2(1) + 0 & 0 + 2(1/6) + 0 \\ 0 + 0 + 3(4/15) & 0 + 0 + 3(1/6) & 0 + 0 + 3(1) \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \\
&= \begin{bmatrix} 5 & -1 & 4/3 \\ -2/5 & 2 & 1/3 \\ 4/5 & 1/2 & 3 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \\
&= \begin{bmatrix} 5(5) + (-1)(0) + (4/3)(0) & 5(0) + (-1)(2) + (4/3)(0) & 5(0) + (-1)(0) + (4/3)(3) \\ (-2/5)(5) + 2(0) + (1/3)(0) & (-2/5)(0) + 2(2) + (1/3)(0) & (-2/5)(0) + 2(0) + (1/3)(3) \\ (4/5)(5) + (1/2)(0) + 3(0) & (4/5)(0) + (1/2)(2) + 3(0) & (4/5)(0) + (1/2)(0) + 3(3) \end{bmatrix} \\
&= \begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix}
\end{aligned}$$

El resultado de la multiplicación $\mathbf{V}^{1/2} \boldsymbol{\rho} \mathbf{V}^{1/2}$ es igual a la matriz de covarianza original $\boldsymbol{\Sigma}$, verificando así la relación.

3.32 Sea el vector aleatorio $\mathbf{X}' = [X_1, X_2, \dots, X_5]$ con vector de medias

$$\boldsymbol{\mu}_X = [2, 4, -1, 3, 0]$$

y matriz de varianza-covarianza

$$\boldsymbol{\Sigma}_X = \begin{bmatrix} 4 & -1 & \frac{1}{2} & \frac{-1}{2} & 0 \\ -1 & 3 & \frac{1}{2} & -1 & 0 \\ \frac{1}{2} & 1 & 6 & 1 & -1 \\ \frac{-1}{2} & -1 & 1 & 4 & 0 \\ 0 & 0 & -1 & 0 & 2 \end{bmatrix}$$

Particionamos \mathbf{X} como:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix} \quad \text{donde} \quad \mathbf{X}^{(1)} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad \mathbf{X}^{(2)} = \begin{bmatrix} X_3 \\ X_4 \\ X_5 \end{bmatrix}$$

Sean:

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \end{bmatrix}$$

y consideremos las combinaciones lineales $\mathbf{A}\mathbf{X}^{(1)}$ y $\mathbf{B}\mathbf{X}^{(2)}$. Hallar:

- $E(\mathbf{X}^{(1)})$
- $E(\mathbf{A}\mathbf{X}^{(1)})$
- $\text{Cov}(\mathbf{X}^{(1)})$
- $\text{Cov}(\mathbf{A}\mathbf{X}^{(1)})$
- $E(\mathbf{X}^{(2)})$
- $E(\mathbf{B}\mathbf{X}^{(2)})$
- $\text{Cov}(\mathbf{X}^{(2)})$
- $\text{Cov}(\mathbf{B}\mathbf{X}^{(2)})$
- $\text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$
- $\text{Cov}(\mathbf{A}\mathbf{X}^{(1)}, \mathbf{B}\mathbf{X}^{(2)})$

Solución

(a) $E(\mathbf{X}^{(1)})$

$$E(\mathbf{X}^{(1)}) = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

(b) $E(\mathbf{A}\mathbf{X}^{(1)})$

$$E(\mathbf{A}\mathbf{X}^{(1)}) = \mathbf{A}E(\mathbf{X}^{(1)}) = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} -2 \\ 6 \end{bmatrix}$$

(c) $Cov(\mathbf{X}^{(1)})$

$$Cov(\mathbf{X}^{(1)}) = [\Sigma_{11}] = \begin{bmatrix} 4 & -1 \\ -1 & 3 \end{bmatrix}$$

(d) $Cov(\mathbf{A}\mathbf{X}^{(1)})$

$$Cov(\mathbf{A}\mathbf{X}^{(1)}) = \mathbf{A}Cov(\mathbf{X}^{(1)})\mathbf{A}' = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 9 & 1 \\ 1 & 5 \end{bmatrix}$$

(e) $E(\mathbf{X}^{(2)})$

$$E(\mathbf{X}^{(2)}) = \begin{bmatrix} \mu_3 \\ \mu_4 \\ \mu_5 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix}$$

(f) $E(\mathbf{B}\mathbf{X}^{(2)})$

$$E(\mathbf{B}\mathbf{X}^{(2)}) = \mathbf{B}E(\mathbf{X}^{(2)}) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

(g) $Cov(\mathbf{X}^{(2)})$

$$Cov(\mathbf{X}^{(2)}) = \begin{bmatrix} 6 & 1 & -1 \\ 1 & 4 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

(h) $Cov(\mathbf{B}\mathbf{X}^{(2)})$

$$Cov(\mathbf{B}\mathbf{X}^{(2)}) = \mathbf{B}Cov(\mathbf{X}^{(2)})\mathbf{B}' = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} 6 & 1 & -1 \\ 1 & 4 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 12 & 9 \\ 9 & 24 \end{bmatrix}$$

(i) $Cov(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$

$$Cov(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = [\Sigma_{12}] = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ 1 & -1 & 0 \end{bmatrix}$$

(j) $Cov(\mathbf{A}\mathbf{X}^{(1)}, \mathbf{B}\mathbf{X}^{(2)})$

$$Cov(\mathbf{A}\mathbf{X}^{(1)}, \mathbf{B}\mathbf{X}^{(2)}) = \mathbf{A}Cov(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})\mathbf{B}' = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

4.1 Considere una distribución normal bivariada con $\mu_1 = 1$, $\mu_2 = 3$, $\sigma_{11} = 2$, $\sigma_{22} = 1$ y $\rho_{12} = -0,8$.

(a) Escriba la densidad normal bivariada.

(b) Escriba la expresión de la distancia estadística cuadrada $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ como una función cuadrática de x_1 y x_2 .

Considere dos variables aleatorias normales con:

- Medias: $\mu_1 = \mathbb{E}(X_1)$, $\mu_2 = \mathbb{E}(X_2)$
- Varianzas: $\sigma_{11} = \text{Var}(X_1)$, $\sigma_{22} = \text{Var}(X_2)$
- Covarianza: $\sigma_{12} = \text{Cov}(X_1, X_2) = \rho_{12}\sqrt{\sigma_{11}\sigma_{22}}$

La matriz de covarianza es:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

Su inversa es:

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2)} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

La distancia estadística cuadrada es:

$$\begin{aligned} D^2 &= (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \\ &= \frac{1}{1 - \rho_{12}^2} \left[\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right)^2 + \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right)^2 - 2\rho_{12} \left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right) \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right) \right] \end{aligned}$$

Entonces, la densidad normal bivariada es:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2)}} \exp \left(-\frac{1}{2(1 - \rho_{12}^2)} \left[\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right)^2 + \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right)^2 - 2\rho_{12} \left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right) \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right) \right] \right)$$

Sustitución:

$$\mu_1 = 1, \quad \mu_2 = 3, \quad \sigma_{11} = 2, \quad \sigma_{22} = 1, \quad \rho_{12} = -0,8$$

La distancia estadística es:

$$\begin{aligned} D^2 &= \frac{1}{1 - (-0,8)^2} \left[\left(\frac{x_1 - 1}{\sqrt{2}} \right)^2 + (x_2 - 3)^2 - 2(-0,8) \left(\frac{x_1 - 1}{\sqrt{2}} \right) (x_2 - 3) \right] \\ &= \frac{1}{0,36} \left[\left(\frac{x_1 - 1}{\sqrt{2}} \right)^2 + (x_2 - 3)^2 + 1,6 \left(\frac{x_1 - 1}{\sqrt{2}} \right) (x_2 - 3) \right] \end{aligned}$$

Y la densidad es:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{2(1 - 0,64)}} \exp \left(-\frac{1}{2(0,36)} D^2 \right).$$

4.2 Considere una población normal bivariada con $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_{11} = 2$, $\sigma_{22} = 1$, y $\rho_{12} = ,5$. (a) Escriba la densidad normal bivariada. (b) Escriba la expresión de la distancia generalizada cuadrática $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ como función de x_1 y x_2 . (c) Determine (y dibuje) el contorno de densidad constante que contiene el 50

Solución:

(a) **Densidad Normal Bivariada**

Primero, identificamos los parámetros:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

$$\sigma_{11} = 2, \quad \sigma_{22} = 1, \quad \rho_{12} = 0,5$$

Calculamos la covarianza:

$$\sigma_{12} = \sigma_{21} = \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} = 0,5\sqrt{2}\sqrt{1} = \frac{\sqrt{2}}{2}$$

La matriz de covarianza es:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2 & \sqrt{2}/2 \\ \sqrt{2}/2 & 1 \end{bmatrix}$$

El determinante de $\boldsymbol{\Sigma}$ es:

$$|\boldsymbol{\Sigma}| = \sigma_{11}\sigma_{22} - \sigma_{12}^2 = (2)(1) - \left(\frac{\sqrt{2}}{2}\right)^2 = 2 - \frac{2}{4} = 2 - \frac{1}{2} = \frac{3}{2} = 1,5$$

También necesitamos $1 - \rho_{12}^2 = 1 - (0,5)^2 = 1 - 0,25 = 0,75 = \frac{3}{4}$.

La fórmula general para la densidad normal bivariada es:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\}$$

Sustituyendo los valores:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{1,5}} \exp \left\{ -\frac{1}{2(1 - \rho_{12}^2)} \left[\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right)^2 - 2\rho_{12} \left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right) \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right) + \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right)^2 \right] \right\}$$

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{1,5}} \exp \left\{ -\frac{1}{2(0,75)} \left[\left(\frac{x_1 - 0}{\sqrt{2}} \right)^2 - 2(0,5) \left(\frac{x_1}{\sqrt{2}} \right) \left(\frac{x_2 - 2}{1} \right) + \left(\frac{x_2 - 2}{1} \right)^2 \right] \right\}$$

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{1,5}} \exp \left\{ -\frac{1}{1,5} \left[\frac{x_1^2}{2} - \frac{x_1(x_2 - 2)}{\sqrt{2}} + (x_2 - 2)^2 \right] \right\}$$

o

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{1,5}} \exp \left\{ -\frac{2}{3} \left[\frac{x_1^2}{2} - \frac{x_1(x_2 - 2)}{\sqrt{2}} + (x_2 - 2)^2 \right] \right\}$$

(b) **Distancia Generalizada Cuadrática**

Necesitamos calcular $\boldsymbol{\Sigma}^{-1}$:

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{|\boldsymbol{\Sigma}|} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{bmatrix} = \frac{1}{1,5} \begin{bmatrix} 1 & -\sqrt{2}/2 \\ -\sqrt{2}/2 & 2 \end{bmatrix} = \frac{2}{3} \begin{bmatrix} 1 & -\sqrt{2}/2 \\ -\sqrt{2}/2 & 2 \end{bmatrix}$$

El vector $\mathbf{x} - \boldsymbol{\mu}$ es:

$$\mathbf{x} - \boldsymbol{\mu} = \begin{bmatrix} x_1 - 0 \\ x_2 - 2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - 2 \end{bmatrix}$$

Ahora calculamos la distancia generalizada cuadrática:

$$\begin{aligned}
(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) &= [x_1 \quad x_2 - 2] \left(\frac{2}{3} \begin{bmatrix} 1 & -\sqrt{2}/2 \\ -\sqrt{2}/2 & 2 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 - 2 \end{bmatrix} \\
&= \frac{2}{3} [x_1 \quad x_2 - 2] \begin{bmatrix} 1 \cdot x_1 - (\sqrt{2}/2)(x_2 - 2) \\ -(\sqrt{2}/2)x_1 + 2(x_2 - 2) \end{bmatrix} \\
&= \frac{2}{3} \left[x_1 \left(x_1 - \frac{\sqrt{2}}{2}(x_2 - 2) \right) + (x_2 - 2) \left(-\frac{\sqrt{2}}{2}x_1 + 2(x_2 - 2) \right) \right] \\
&= \frac{2}{3} \left[x_1^2 - \frac{\sqrt{2}}{2}x_1(x_2 - 2) - \frac{\sqrt{2}}{2}x_1(x_2 - 2) + 2(x_2 - 2)^2 \right] \\
&= \frac{2}{3} \left[x_1^2 - \sqrt{2}x_1(x_2 - 2) + 2(x_2 - 2)^2 \right]
\end{aligned}$$

$$\boxed{(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \frac{2}{3}x_1^2 - \frac{2\sqrt{2}}{3}x_1(x_2 - 2) + \frac{4}{3}(x_2 - 2)^2}$$

(c) **Contorno de Densidad Constante al 50 %**

Un contorno de densidad constante es una elipse definida por $(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c^2$. Para que la elipse contenga el $100(1 - \alpha) \%$ de la probabilidad, debemos tener $c^2 = \chi_p^2(\alpha)$. Aquí, $p = 2$ (bivariada) y queremos el 50%. Necesitamos el valor crítico $\chi_2^2(0,50)$. Consultando una tabla de la distribución chi-cuadrado con 2 grados de libertad, encontramos $\chi_2^2(0,50) = 1,386$ (o 1.39 como aproximación).

La ecuación del contorno es:

$$\frac{2}{3} \left[x_1^2 - \sqrt{2}x_1(x_2 - 2) + 2(x_2 - 2)^2 \right] = 1,386$$

Esta es la ecuación de una elipse. Para dibujarla, necesitamos:

- **Centro:** $\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$.
- **Orientación y longitud de los ejes:** Dadas por los vectores y valores propios de $\boldsymbol{\Sigma}$. Calculamos los valores propios de $\boldsymbol{\Sigma}$ resolviendo $|\boldsymbol{\Sigma} - \lambda \mathbf{I}| = 0$:

$$(2 - \lambda)(1 - \lambda) - (\sqrt{2}/2)^2 = \lambda^2 - 3\lambda + 2 - 0,5 = \lambda^2 - 3\lambda + 1,5 = 0$$

$$\lambda = \frac{3 \pm \sqrt{9 - 4(1,5)}}{2} = \frac{3 \pm \sqrt{3}}{2}$$

Los valores propios son $\lambda_1 = \frac{3+\sqrt{3}}{2} \approx 2,366$ y $\lambda_2 = \frac{3-\sqrt{3}}{2} \approx 0,634$. Los vectores propios correspondientes (no normalizados) son proporcionales a $\begin{bmatrix} \sqrt{2} \\ \sqrt{3} - 1 \end{bmatrix}$ para λ_1 y $\begin{bmatrix} \sqrt{2} \\ -(1 + \sqrt{3}) \end{bmatrix}$ para λ_2 . Estos vectores dan la dirección de los ejes mayor y menor de la elipse. Las semilongitudes de los ejes son $\sqrt{\lambda_i} \sqrt{\chi_2^2(0,50)}$.

◦ Semilongitud eje 1: $\sqrt{\lambda_1} \sqrt{1,386} = \sqrt{2,366} \sqrt{1,386} \approx (1,538)(1,177) \approx 1,810$

◦ Semilongitud eje 2: $\sqrt{\lambda_2} \sqrt{1,386} = \sqrt{0,634} \sqrt{1,386} \approx (0,796)(1,177) \approx 0,937$

De igual manera se realizó para 25 y 75 como se muestra a continuación:

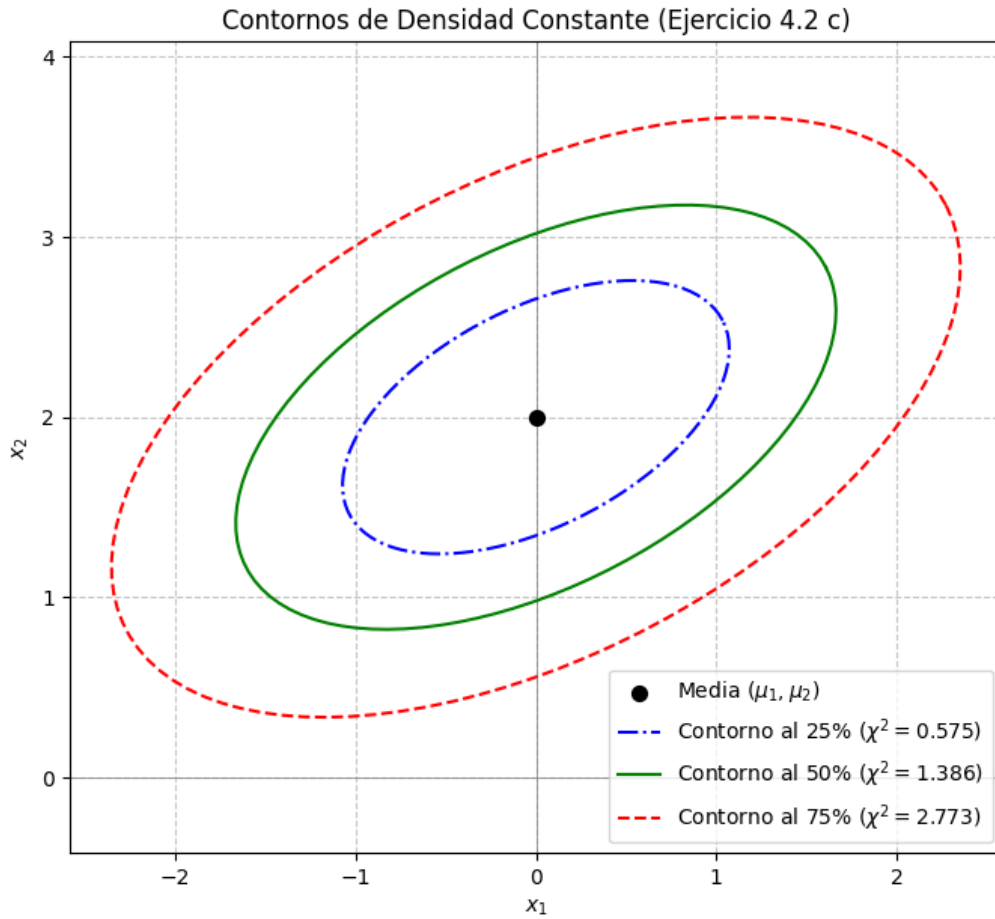


Figura 9: Contornos de una densidad al 25, 50 y 75

Como se observa en la figura 9 la elipse está centrada en el punto $(0, 2)$. Su eje mayor tiene una longitud total de $2 \times 1,810 = 3,620$ y está orientado en la dirección del vector $\begin{bmatrix} \sqrt{2} \\ \sqrt{3} - 1 \end{bmatrix} \approx \begin{bmatrix} 1,414 \\ 0,732 \end{bmatrix}$ (aproximadamente 27 grados desde el eje x_1 positivo). Su eje menor tiene una longitud total de $2 \times 0,937 = 1,874$ y es perpendicular al eje mayor.

4.3 Sea $\mathbf{X} \sim \mathcal{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con

$$\boldsymbol{\mu}' = [-3, 1, 4], \quad \text{y} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

¿Cuáles de las siguientes variables aleatorias son independientes? Justifique su respuesta.

- (a) X_1 y X_2
- (b) X_2 y X_3
- (c) (X_1, X_2) y X_3
- (d) $\frac{X_1 + X_2}{2}$ y X_3
- (e) X_2 y $X_2 - \frac{5}{2}X_1 - X_3$

Solución

Recordemos que para una distribución normal multivariada, la independencia entre componentes es equivalente a una covarianza cero.

(a) X_1 y X_2

La covarianza entre X_1 y X_2 es $\Sigma_{12} = -2 \neq 0$. Por lo tanto:

X_1 y X_2 **no son independientes**

(b) X_2 y X_3

La covarianza entre X_2 y X_3 es $\Sigma_{23} = 0$. Por lo tanto:

X_2 y X_3 **son independientes**

(c) (X_1, X_2) y X_3

La matriz de covarianza cruzada es:

$$\text{Cov} \left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, X_3 \right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Como todas las covarianzas cruzadas son cero:

(X_1, X_2) y X_3 **son independientes**

(d) $\frac{X_1 + X_2}{2}$ y X_3

Calculamos la covarianza:

$$\text{Cov} \left(\frac{X_1 + X_2}{2}, X_3 \right) = \frac{1}{2}(\Sigma_{13} + \Sigma_{23}) = \frac{1}{2}(0 + 0) = 0$$

Por lo tanto:

$\frac{X_1 + X_2}{2}$ y X_3 **son independientes**

(e) X_2 y $X_2 - \frac{5}{2}X_1 - X_3$

Calculamos la covarianza:

$$\begin{aligned} \text{Cov}(X_2, X_2 - \tfrac{5}{2}X_1 - X_3) &= \Sigma_{22} - \tfrac{5}{2}\Sigma_{21} - \Sigma_{23} \\ &= 5 - \tfrac{5}{2}(-2) - 0 \\ &= 5 + 5 - 0 = 10 \neq 0 \end{aligned}$$

Por lo tanto:

X_2 y $X_2 - \frac{5}{2}X_1 - X_3$ **no son independientes**

4.4 Sea $\mathbf{X} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\boldsymbol{\mu}' = [2, -3, 1]$ y

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}$$

(a) Encuentre la distribución de $3X_1 - 2X_2 + X_3$: Como $3X_1 - 2X_2 + X_3 = (3, -2, 1)\mathbb{X}$ entonces por el resultado 4.2 del libro guía, se tiene

$$(3, -2, 1)\mathbb{X} \sim N((3, -2, 1)\boldsymbol{\mu}, (3, -2, 1)\boldsymbol{\Sigma}(3, -2, 1)')$$

donde

$$(3, -2, 1)\mu = (3, -2, 1)(2, -3, 1)' = 6 + 6 + 1 = 13$$

y

$$(3, -2, 1) \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix} (3, -2, 1)' = 6 + 2 + 1 = 9$$

finalmente

$$3X_1 - 2X_2 + X_3 \sim N(13, 9)$$

(b) Renombre las variables si es necesario, y encuentre un vector $\mathbf{a} \in \mathbb{R}^{2 \times 1}$ tal que X_2 y

$$X_2 - \mathbf{a}' \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$$

sean independientes:

Queremos encontrar un vector $\mathbf{a} \in \mathbb{R}^{2 \times 1}$ tal que:

$$\text{Cov} \left(X_2 - \mathbf{a}' \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}, \begin{bmatrix} X_1 \\ X_3 \end{bmatrix} \right) = \mathbf{0}'$$

Expandimos esta covarianza utilizando propiedades lineales:

$$\begin{aligned} \text{Cov}(X_2 - \mathbf{a}'\mathbf{Z}, \mathbf{Z}) &= \text{Cov}(X_2, \mathbf{Z}) - \text{Cov}(\mathbf{a}'\mathbf{Z}, \mathbf{Z}) \\ &= \boldsymbol{\Sigma}_{21} - \mathbf{a}'\boldsymbol{\Sigma}_{11} \end{aligned}$$

Donde:

- $\mathbf{Z} = \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$
- $\boldsymbol{\Sigma}_{21} = \text{Cov}(X_2, \mathbf{Z}) = [\text{Cov}(X_2, X_1) \quad \text{Cov}(X_2, X_3)] = [1 \quad 2]$
- $\boldsymbol{\Sigma}_{11} = \text{Cov}(\mathbf{Z}, \mathbf{Z}) = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_3) \\ \text{Cov}(X_3, X_1) & \text{Var}(X_3) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$

Entonces la condición de independencia equivale a:

$$\boldsymbol{\Sigma}_{21} = \mathbf{a}'\boldsymbol{\Sigma}_{11} \quad \Rightarrow \quad \mathbf{a}' = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$$

Calculamos la inversa de $\boldsymbol{\Sigma}_{11}$:

$$\boldsymbol{\Sigma}_{11}^{-1} = \frac{1}{1 \cdot 2 - 1 \cdot 1} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

Entonces:

$$\begin{aligned} \mathbf{a}' &= [1 \quad 2] \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \\ &= [2 - 2 \quad -1 + 2] = [0 \quad 1] \end{aligned}$$

Por tanto:

$$\boxed{\mathbf{a} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}}$$

Y la combinación:

$$X_2 - \mathbf{a}' \begin{bmatrix} X_1 \\ X_3 \end{bmatrix} = X_2 - X_3$$

es independiente de (X_1, X_3) , como se deseaba.

4.5 Especifique cada una de las siguientes distribuciones.

Solución:

Utilizaremos el Resultado 4.6 del libro de texto. Si un vector aleatorio \mathbf{X} ($p \times 1$) se distribuye como $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y se particiona como $\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix}$ con las correspondientes particiones $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{bmatrix}$ y $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$, entonces la distribución condicional de $\mathbf{X}^{(1)}$ dado $\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$ es normal multivariada:

$$N_{p_1}(\boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$$

donde $\mathbf{X}^{(1)}$ tiene p_1 componentes.

(a) **Distribución condicional de X_1 dado $X_2 = x_2$ para la distribución conjunta del Ejercicio 4.2.**

Del Ejercicio 4.2, tenemos $p = 2$:

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 2 & \sqrt{2}/2 \\ \sqrt{2}/2 & 1 \end{bmatrix}$$

Particionamos $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$. Las partes correspondientes son:

$$\begin{aligned} \boldsymbol{\mu}^{(1)} &= \mu_1 = 0, & \boldsymbol{\mu}^{(2)} &= \mu_2 = 2 \\ \boldsymbol{\Sigma}_{11} &= \sigma_{11} = 2, & \boldsymbol{\Sigma}_{12} &= \sigma_{12} = \sqrt{2}/2 \\ \boldsymbol{\Sigma}_{21} &= \sigma_{21} = \sqrt{2}/2, & \boldsymbol{\Sigma}_{22} &= \sigma_{22} = 1 \end{aligned}$$

Necesitamos $\boldsymbol{\Sigma}_{22}^{-1} = \sigma_{22}^{-1} = 1^{-1} = 1$. La media condicional es:

$$\mu_{1|2} = \mu_1 + \sigma_{12}\sigma_{22}^{-1}(x_2 - \mu_2) = 0 + (\sqrt{2}/2)(1)(x_2 - 2) = \frac{\sqrt{2}}{2}(x_2 - 2)$$

La varianza condicional es:

$$\sigma_{11|2} = \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21} = 2 - (\sqrt{2}/2)(1)(\sqrt{2}/2) = 2 - \frac{2}{4} = 2 - \frac{1}{2} = \frac{3}{2}$$

Por lo tanto, la distribución condicional de X_1 dado $X_2 = x_2$ es:

$$\boxed{N\left(\frac{\sqrt{2}}{2}(x_2 - 2), \frac{3}{2}\right)}$$

(b) **Distribución condicional de X_2 dado $X_1 = x_1$ y $X_3 = x_3$ para la distribución conjunta del Ejercicio 4.3.**

Del Ejercicio 4.3, tenemos $p = 3$:

$$\boldsymbol{\mu} = \begin{bmatrix} -3 \\ 1 \\ 4 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Queremos la distribución de X_2 condicionada a X_1 y X_3 . Particionamos \mathbf{X} como $\mathbf{X} = \begin{bmatrix} X_2 \\ X_1 \\ X_3 \end{bmatrix}$. Las partes correspondientes son:

$$\boldsymbol{\mu}^{(1)} = \mu_2 = 1, \quad \boldsymbol{\mu}^{(2)} = \begin{bmatrix} \mu_1 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} -3 \\ 4 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_{11} = \sigma_{22} = 5, \quad \boldsymbol{\Sigma}_{12} = [\sigma_{21} \quad \sigma_{23}] = [-2 \quad 0]$$

$$\boldsymbol{\Sigma}_{21} = \begin{bmatrix} \sigma_{12} \\ \sigma_{32} \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{22} = \begin{bmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

Calculamos $\boldsymbol{\Sigma}_{22}^{-1}$: Como $\boldsymbol{\Sigma}_{22}$ es diagonal, su inversa es $\boldsymbol{\Sigma}_{22}^{-1} = \begin{bmatrix} 1/1 & 0 \\ 0 & 1/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}$. La media condicional es:

$$\begin{aligned} \mu_{2|1,3} &= \mu^{(1)} + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}) \\ &= 1 + [-2 \quad 0] \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \left(\begin{bmatrix} x_1 \\ x_3 \end{bmatrix} - \begin{bmatrix} -3 \\ 4 \end{bmatrix} \right) \\ &= 1 + [-2 \quad 0] \begin{bmatrix} x_1 + 3 \\ (x_3 - 4)/2 \end{bmatrix} \\ &= 1 + (-2)(x_1 + 3) + (0)((x_3 - 4)/2) \\ &= 1 - 2x_1 - 6 = -5 - 2x_1 \end{aligned}$$

La varianza condicional es:

$$\begin{aligned} \sigma_{22|1,3} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \\ &= 5 - [-2 \quad 0] \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \end{bmatrix} \\ &= 5 - [-2 \quad 0] \begin{bmatrix} -2 \\ 0 \end{bmatrix} \\ &= 5 - ((-2)(-2) + (0)(0)) = 5 - 4 = 1 \end{aligned}$$

Por lo tanto, la distribución condicional de X_2 dado $X_1 = x_1$ y $X_3 = x_3$ es:

$$\boxed{N(-5 - 2x_1, 1)}$$

(c) Distribución condicional de X_3 dado $X_1 = x_1$ y $X_2 = x_2$ para la distribución conjunta del Ejercicio 4.4.

Del Ejercicio 4.4, tenemos $p = 3$:

$$\boldsymbol{\mu} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}$$

Queremos la distribución de X_3 condicionada a X_1 y X_2 . Particionamos \mathbf{X} como $\mathbf{X} = \begin{bmatrix} X_3 \\ X_1 \\ X_2 \end{bmatrix}$. Las partes correspondientes son:

$$\boldsymbol{\mu}^{(1)} = \mu_3 = 1, \quad \boldsymbol{\mu}^{(2)} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_{11} = \sigma_{33} = 2, \quad \boldsymbol{\Sigma}_{12} = [\sigma_{31} \quad \sigma_{32}] = [1 \quad 2]$$

$$\boldsymbol{\Sigma}_{21} = \begin{bmatrix} \sigma_{13} \\ \sigma_{23} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{22} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}$$

Calculamos Σ_{22}^{-1} : $|\Sigma_{22}| = (1)(3) - (1)(1) = 2$.

$$\Sigma_{22}^{-1} = \frac{1}{2} \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 3/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}$$

La media condicional es:

$$\begin{aligned} \mu_{3|1,2} &= \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}) \\ &= 1 + [1 \quad 2] \begin{bmatrix} 3/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix} \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 2 \\ -3 \end{bmatrix} \right) \\ &= 1 + [(1)(3/2) + 2(-1/2) \quad (1)(-1/2) + 2(1/2)] \begin{bmatrix} x_1 - 2 \\ x_2 + 3 \end{bmatrix} \\ &= 1 + [1/2 \quad 1/2] \begin{bmatrix} x_1 - 2 \\ x_2 + 3 \end{bmatrix} \\ &= 1 + \frac{1}{2}(x_1 - 2) + \frac{1}{2}(x_2 + 3) = 1 + \frac{1}{2}x_1 - 1 + \frac{1}{2}x_2 + \frac{3}{2} \\ &= \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{3}{2} \end{aligned}$$

La varianza condicional es:

$$\begin{aligned} \sigma_{33|1,2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ &= 2 - [1 \quad 2] \begin{bmatrix} 3/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ &= 2 - [1/2 \quad 1/2] \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ &= 2 - ((1/2)(1) + (1/2)(2)) = 2 - \left(\frac{1}{2} + 1\right) = 2 - \frac{3}{2} = \frac{1}{2} \end{aligned}$$

Por lo tanto, la distribución condicional de X_3 dado $X_1 = x_1$ y $X_2 = x_2$ es:

$$\boxed{N\left(\frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{3}{2}, \frac{1}{2}\right)}$$

4.6 Sea \mathbf{X} una variable aleatoria distribuida como $\mathcal{N}_3(\boldsymbol{\mu}, \Sigma)$, donde

$$\boldsymbol{\mu}' = [1, -1, 2] \quad \text{y} \quad \Sigma = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

¿Cuáles de las siguientes variables aleatorias son independientes? Justifique su respuesta.

- (a) X_1 y X_2
- (b) X_1 y X_3
- (c) X_2 y X_3
- (d) (X_1, X_3) y X_2
- (e) X_1 y $X_1 + 3X_2 - 2X_3$

Solución

Para una distribución normal multivariada, la independencia entre componentes es equivalente a una covarianza cero.

(a) X_1 y X_2

La covarianza entre X_1 y X_2 es $\Sigma_{12} = 0$. Por lo tanto:

X_1, X_2 **son independientes**

(b) X_1 y X_3

La covarianza entre X_1 y X_3 es $\Sigma_{13} = -1 \neq 0$. Por lo tanto:

X_1 y X_3 **no son independientes**

(c) X_2 y X_3

La covarianza entre X_2 y X_3 es $\Sigma_{23} = 0$. Por lo tanto:

X_2, X_3 **son independientes**

(d) (X_1, X_3) y X_2

La matriz de covarianza cruzada es:

$$\text{Cov} \left(\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}, X_2 \right) = \begin{bmatrix} \Sigma_{12} \\ \Sigma_{23} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Como todas las covarianzas cruzadas son cero:

$(X_1, X_3), X_2$ **son independientes**

(e) X_1 y $X_1 + 3X_2 - 2X_3$

Calculamos la covarianza:

$$\begin{aligned} \text{Cov}(X_1, X_1 + 3X_2 - 2X_3) &= \Sigma_{11} + 3\Sigma_{12} - 2\Sigma_{13} \\ &= 4 + 3(0) - 2(-1) \\ &= 4 + 0 + 2 = 6 \neq 0 \end{aligned}$$

Por lo tanto:

X_1 y $X_1 + 3X_2 - 2X_3$ **no son independientes**

4.7 Refiérase al Ejercicio 4.6 y especifique cada uno de los siguientes:

(a) La distribución condicional de X_1 , dado que $X_3 = x_3$.

Sea $X_1 \mid X_3 = x_3$. Usamos la fórmula de la normal condicional univariada:

$$X_1 \mid X_3 = x_3 \sim \mathcal{N}(\mu_{1|3}, \sigma_{1|3}^2)$$

donde

$$\mu_{1|3} = \mu_1 + \frac{\sigma_{13}}{\sigma_{33}}(x_3 - \mu_3), \quad \sigma_{1|3}^2 = \sigma_{11} - \frac{\sigma_{13}^2}{\sigma_{33}}$$

Sustituyendo:

$$\begin{aligned} \mu_{1|3} &= 1 + \frac{-1}{2}(x_3 - 2) = 1 - \frac{1}{2}(x_3 - 2) \\ \sigma_{1|3}^2 &= 4 - \frac{(-1)^2}{2} = 4 - \frac{1}{2} = \frac{7}{2} \end{aligned}$$

Entonces:

$$X_1 \mid X_3 = x_3 \sim \mathcal{N}\left(1 - \frac{1}{2}(x_3 - 2), \frac{7}{2}\right)$$

(b) La distribución condicional de X_1 , dado que $X_2 = x_2$ y $X_3 = x_3$.

Aquí $X_1 | (X_2 = x_2, X_3 = x_3) \sim \mathcal{N}(\mu_{1|2,3}, \sigma_{1|2,3}^2)$

Definimos:

$$\boldsymbol{\mu}_1 = \mu_1 = 1, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{12} = \begin{bmatrix} 0 & -1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{22} = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}$$

Entonces:

$$\begin{aligned} \mu_{1|2,3} &= \mu_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \begin{bmatrix} x_2 - \mu_2^{(1)} \\ x_3 - \mu_2^{(2)} \end{bmatrix} = 1 + \begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 1/5 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} x_2 + 1 \\ x_3 - 2 \end{bmatrix} \\ &= 1 - \frac{1}{2}(x_3 - 2) \end{aligned}$$

$$\sigma_{1|2,3}^2 = \sigma_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{12}' = 4 - \begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 1/5 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = 4 - \frac{1}{2} = \frac{7}{2}$$

Por tanto:

$$X_1 | (X_2 = x_2, X_3 = x_3) \sim \mathcal{N}\left(1 - \frac{1}{2}(x_3 - 2), \frac{7}{2}\right)$$

4.8 Sea X_1 una variable aleatoria $N(0, 1)$, y sea

$$X_2 = \begin{cases} -X_1 & \text{si } -1 \leq X_1 \leq 1 \\ X_1 & \text{en otro caso (es decir, } |X_1| > 1) \end{cases}$$

Demuestre cada uno de los siguientes puntos: (a) X_2 también tiene una distribución $N(0, 1)$. (b) X_1 y X_2 *no* tienen una distribución normal bivariada.

Solución:

(a) **Demostrar que** $X_2 \sim N(0, 1)$

Para demostrar que X_2 tiene una distribución $N(0, 1)$, mostraremos que su función de distribución acumulada (CDF), $F_{X_2}(x) = P(X_2 \leq x)$, es igual a la CDF de una $N(0, 1)$, denotada por $\Phi(x)$. Usaremos la ley de probabilidad total, particionando el espacio muestral según los valores de X_1 .

Sea A el evento $X_1 \in [-1, 1]$ y A^c el evento $|X_1| > 1$.

$$P(X_2 \leq x) = P(X_2 \leq x \text{ y } A) + P(X_2 \leq x \text{ y } A^c)$$

- Si $X_1 \in [-1, 1]$ (evento A), entonces $X_2 = -X_1$.
- Si $|X_1| > 1$ (evento A^c), entonces $X_2 = X_1$.

Sustituyendo esto:

$$P(X_2 \leq x) = P(-X_1 \leq x \text{ y } X_1 \in [-1, 1]) + P(X_1 \leq x \text{ y } |X_1| > 1)$$

$$P(X_2 \leq x) = P(X_1 \geq -x \text{ y } X_1 \in [-1, 1]) + P(X_1 \leq x \text{ y } (X_1 > 1 \text{ o } X_1 < -1))$$

Ahora consideraremos diferentes rangos para x :

a) **Caso 1:** $x < -1$

- Para el primer término: $P(X_1 \geq -x \text{ y } X_1 \in [-1, 1])$. Como $x < -1$, entonces $-x > 1$. El evento $X_1 \geq -x$ (es decir, $X_1 > 1$) y $X_1 \in [-1, 1]$ es imposible. Así, el primer término es 0.
- Para el segundo término: $P(X_1 \leq x \text{ y } (X_1 > 1 \text{ o } X_1 < -1))$. Como $x < -1$, el evento $X_1 \leq x$ implica $X_1 < -1$. Por lo tanto, este término se reduce a $P(X_1 \leq x)$.

Entonces, si $x < -1$,

$$P(X_2 \leq x) = 0 + P(X_1 \leq x) = P(X_1 \leq x) = \Phi(x)$$

b) **Caso 2:** $-1 \leq x \leq 1$

- Para el primer término: $P(X_1 \geq -x \text{ y } X_1 \in [-1, 1])$. Como $-1 \leq x \leq 1$, entonces $-1 \leq -x \leq 1$. El evento es $P(-x \leq X_1 \leq 1)$.
- Para el segundo término: $P(X_1 \leq x \text{ y } (X_1 > 1 \text{ o } X_1 < -1))$. Como $-1 \leq x \leq 1$, el evento $X_1 \leq x \text{ y } X_1 > 1$ es imposible. El evento $X_1 \leq x \text{ y } X_1 < -1$ es $P(X_1 < -1)$ si $x \geq -1$, o $P(X_1 \leq x)$ si $x < -1$. Dado que estamos en el caso $-1 \leq x \leq 1$, este término se reduce a $P(X_1 < -1)$.

Entonces, si $-1 \leq x \leq 1$,

$$P(X_2 \leq x) = P(-x \leq X_1 \leq 1) + P(X_1 < -1) = P(X_1 \leq x) = \Phi(x)$$

Usando la pista descrita en el libro.

c) **Caso 3:** $x > 1$

- Para el primer término: $P(X_1 \geq -x \text{ y } X_1 \in [-1, 1])$. Como $x > 1$, entonces $-x < -1$. La condición $X_1 \geq -x$ siempre es verdadera para $X_1 \in [-1, 1]$ ya que $-x < -1 \leq X_1$. Así, este término es $P(X_1 \in [-1, 1])$.
- Para el segundo término: $P(X_1 \leq x \text{ y } (X_1 > 1 \text{ o } X_1 < -1))$. Esto se descompone en $P(X_1 \leq x \text{ y } X_1 > 1) + P(X_1 \leq x \text{ y } X_1 < -1)$. El primer sub-término es $P(1 < X_1 \leq x)$. El segundo sub-término es $P(X_1 < -1)$.

Entonces, si $x > 1$, $P(X_2 \leq x) = P(X_1 \in [-1, 1]) + P(1 < X_1 \leq x) + P(X_1 < -1)$. Combinando $P(X_1 < -1)$ y $P(X_1 \in [-1, 1])$ nos da $P(X_1 \leq 1)$. Por lo tanto,

$$P(X_2 \leq x) = P(X_1 \leq 1) + P(1 < X_1 \leq x) = P(X_1 \leq x) = \Phi(x)$$

En los tres casos, $F_{X_2}(x) = \Phi(x)$. Por lo tanto, X_2 tiene una distribución $N(0, 1)$.

(b) **Demostrar que (X_1, X_2) no es normal bivariada**

Para que (X_1, X_2) tenga una distribución normal bivariada, cualquier combinación lineal de X_1 y X_2 debe seguir una distribución normal univariada. Consideremos la combinación lineal $Y = X_1 - X_2$.

Analizamos Y según los dos casos para X_2 :

- Si $-1 \leq X_1 \leq 1$: En este caso, $X_2 = -X_1$. Entonces $Y = X_1 - (-X_1) = X_1 + X_1 = 2X_1$.
- Si $|X_1| > 1$ (es decir, $X_1 < -1$ o $X_1 > 1$): En este caso, $X_2 = X_1$. Entonces $Y = X_1 - X_1 = 0$.

La variable aleatoria Y toma el valor 0 siempre que $|X_1| > 1$. La probabilidad de que $Y = 0$ es:

$$P(Y = 0) = P(|X_1| > 1)$$

Dado que $X_1 \sim N(0, 1)$, tenemos $P(X_1 > 1) = 1 - \Phi(1)$ y $P(X_1 < -1) = \Phi(-1)$. Por la simetría de la distribución normal estándar, $\Phi(-1) = 1 - \Phi(1)$.

$$P(|X_1| > 1) = P(X_1 > 1) + P(X_1 < -1) = (1 - \Phi(1)) + (1 - \Phi(1)) = 2(1 - \Phi(1))$$

Usando una tabla normal, $\Phi(1) \approx 0,8413$.

$$P(Y = 0) \approx 2(1 - 0,8413) = 2(0,1587) = 0,3174$$

Una variable aleatoria continua con distribución normal (no degenerada) tiene probabilidad cero de tomar cualquier valor puntual específico. Como $P(Y = 0) = 0,3174 \neq 0$, la variable $Y = X_1 - X_2$ no sigue una distribución normal continua.

Dado que la combinación lineal $X_1 - X_2$ no es normal, el vector (X_1, X_2) no puede seguir una distribución normal bivariada.

4.9 Refiérase al Ejercicio 4.8, pero modifique la construcción reemplazando el punto de quiebre 1 por un valor c de modo que:

$$X_2 = \begin{cases} -X_1, & \text{si } -c \leq X_1 \leq c \\ X_1, & \text{en otro caso} \end{cases}$$

Demuestre que se puede elegir c de modo que $\text{Cov}(X_1, X_2) = 0$, pero que las dos variables aleatorias no son independientes.

Solución

Sea $X_1 \sim \mathcal{N}(0, 1)$ y definamos X_2 como:

$$X_2 = \begin{cases} -X_1, & \text{si } -c \leq X_1 \leq c \\ X_1, & \text{en otro caso} \end{cases}$$

La covarianza entre X_1 y X_2 es:

$$\text{Cov}(X_1, X_2) = E[X_1 X_2] - E[X_1]E[X_2] = E[X_1 X_2]$$

(pues $E[X_1] = 0$)

Calculamos $E[X_1 X_2]$:

$$E[X_1 X_2] = \int_{-\infty}^{\infty} x_1 x_2 f_{X_1}(x_1) dx_1 = \int_{-\infty}^{-c} x_1^2 \phi(x_1) dx_1 + \int_{-c}^c -x_1^2 \phi(x_1) dx_1 + \int_c^{\infty} x_1^2 \phi(x_1) dx_1$$

Donde $\phi(x)$ es la densidad de $\mathcal{N}(0, 1)$.

Por simetría de la normal,

$$E[X_1 X_2] = 2 \int_c^{\infty} x^2 \phi(x) dx - 2 \int_0^c x^2 \phi(x) dx$$

Usando que $E[X_1^2] = 1$:

$$1 = 2 \int_0^{\infty} x^2 \phi(x) dx = 2 \int_0^c x^2 \phi(x) dx + 2 \int_c^{\infty} x^2 \phi(x) dx$$

Por lo tanto:

$$E[X_1 X_2] = \left[1 - 2 \int_0^c x^2 \phi(x) dx \right] - 2 \int_0^c x^2 \phi(x) dx = 1 - 4 \int_0^c x^2 \phi(x) dx$$

Queremos $\text{Cov}(X_1, X_2) = 0$, es decir:

$$1 - 4 \int_0^c x^2 \phi(x) dx = 0 \quad \Rightarrow \quad \int_0^c x^2 \phi(x) dx = \frac{1}{4}$$

es claro que existe c que cumple esta condición, ya que $x^2 \phi(x)$ es una función continua, y además $\int_0^{\infty} x^2 \phi(x) dx = \frac{1}{2}$.

Aunque $\text{Cov}(X_1, X_2) = 0$, las variables no son independientes porque:

$$P(X_1 > c, X_2 > c) = P(X_1 > c) \neq P(X_1 > c)P(X_2 > c)$$

ya que $P(X_2 > c) = P(X_1 > c)$

4.17 Sean $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$ y \mathbf{X}_5 vectores aleatorios independientes e idénticamente distribuidos con vector media $\boldsymbol{\mu}$ y matriz de covarianza $\boldsymbol{\Sigma}$.

Hallar el vector media y la matriz de covarianza para cada una de las siguientes combinaciones lineales de vectores aleatorios:

$$\frac{1}{5}\mathbf{X}_1 + \frac{1}{5}\mathbf{X}_2 + \frac{1}{5}\mathbf{X}_3 + \frac{1}{5}\mathbf{X}_4 + \frac{1}{5}\mathbf{X}_5$$

y

$$\mathbf{X}_1 - \mathbf{X}_2 + \mathbf{X}_3 - \mathbf{X}_4 + \mathbf{X}_5$$

en términos de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$. Además, obtener la covarianza entre las dos combinaciones lineales anteriores.

Sean $\mathbf{X}_1, \dots, \mathbf{X}_5$ vectores aleatorios independientes e idénticamente distribuidos, con:

$$\mathbb{E}[\mathbf{X}_i] = \boldsymbol{\mu}, \quad \text{Cov}(\mathbf{X}_i) = \boldsymbol{\Sigma}, \quad \text{para } i = 1, \dots, 5$$

Definimos las siguientes combinaciones lineales:

$$\mathbf{A} = \frac{1}{5}(\mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3 + \mathbf{X}_4 + \mathbf{X}_5)$$

$$\mathbf{B} = \mathbf{X}_1 - \mathbf{X}_2 + \mathbf{X}_3 - \mathbf{X}_4 + \mathbf{X}_5$$

1. Esperanzas

Usando linealidad de la esperanza:

$$\mathbb{E}[\mathbf{A}] = \frac{1}{5} \sum_{i=1}^5 \mathbb{E}[\mathbf{X}_i] = \frac{1}{5}(5\boldsymbol{\mu}) = \boldsymbol{\mu}$$

$$\mathbb{E}[\mathbf{B}] = \mathbb{E}[\mathbf{X}_1 - \mathbf{X}_2 + \mathbf{X}_3 - \mathbf{X}_4 + \mathbf{X}_5] = \boldsymbol{\mu} - \boldsymbol{\mu} + \boldsymbol{\mu} - \boldsymbol{\mu} + \boldsymbol{\mu} = \boldsymbol{\mu}$$

2. Varianzas

Como los vectores son independientes, las varianzas de las combinaciones lineales se calculan como:

$$\text{Cov}(\mathbf{A}) = \frac{1}{5^2} \sum_{i=1}^5 \text{Cov}(\mathbf{X}_i) = \frac{5}{25}\boldsymbol{\Sigma} = \frac{1}{5}\boldsymbol{\Sigma}$$

$$\begin{aligned} \text{Cov}(\mathbf{B}) &= \text{Cov}(\mathbf{X}_1 - \mathbf{X}_2 + \mathbf{X}_3 - \mathbf{X}_4 + \mathbf{X}_5) \\ &= \boldsymbol{\Sigma} + \boldsymbol{\Sigma} + \boldsymbol{\Sigma} + \boldsymbol{\Sigma} + \boldsymbol{\Sigma} = 5\boldsymbol{\Sigma} \end{aligned}$$

Los signos cruzados no afectan porque las \mathbf{X}_i son independientes. Pero los coeficientes sí, entonces:

$$\text{Cov}(\mathbf{B}) = (1^2 + (-1)^2 + 1^2 + (-1)^2 + 1^2)\boldsymbol{\Sigma} = (1 + 1 + 1 + 1 + 1)\boldsymbol{\Sigma} = 5\boldsymbol{\Sigma}$$

3. Covarianza entre A y B

Usamos la fórmula:

$$\text{Cov}(\mathbf{A}, \mathbf{B}) = \text{Cov}\left(\frac{1}{5} \sum_{i=1}^5 \mathbf{X}_i, \mathbf{X}_1 - \mathbf{X}_2 + \mathbf{X}_3 - \mathbf{X}_4 + \mathbf{X}_5\right)$$

Como las variables son independientes, solo se conservan los términos donde los índices coinciden:

$$\begin{aligned} &= \frac{1}{5} [\text{Cov}(\mathbf{X}_1, \mathbf{X}_1) - \text{Cov}(\mathbf{X}_2, \mathbf{X}_2) + \text{Cov}(\mathbf{X}_3, \mathbf{X}_3) - \text{Cov}(\mathbf{X}_4, \mathbf{X}_4) + \text{Cov}(\mathbf{X}_5, \mathbf{X}_5)] \\ &= \frac{1}{5}(1 - 1 + 1 - 1 + 1)\boldsymbol{\Sigma} = \frac{1}{5}(1)\boldsymbol{\Sigma} = \frac{1}{5}\boldsymbol{\Sigma} \end{aligned}$$

- 4.18 Encuentre los estimadores de máxima verosimilitud del vector de medias 2×1 $\boldsymbol{\mu}$ y la matriz de covarianza 2×2 $\boldsymbol{\Sigma}$ basados en la muestra aleatoria

$$\mathbf{X} = \begin{bmatrix} 3 & 6 \\ 4 & 4 \\ 5 & 7 \\ 4 & 7 \end{bmatrix}$$

de una población normal bivariada.

Solución:

Para una población normal multivariada, los estimadores de máxima verosimilitud (EMV o MLE en inglés) para el vector de medias $\boldsymbol{\mu}$ y la matriz de covarianza $\boldsymbol{\Sigma}$ son el vector de medias muestrales $\bar{\mathbf{x}}$ y la matriz de covarianza muestral $\hat{\boldsymbol{\Sigma}}$ (usando el divisor n), respectivamente.

La muestra consiste en $n = 4$ observaciones bivariadas ($p = 2$):

$$\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 5 \\ 7 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$$

a) **Calcular el vector de medias muestrales ($\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$):**

$$\bar{x}_1 = \frac{3 + 4 + 5 + 4}{4} = \frac{16}{4} = 4$$

$$\bar{x}_2 = \frac{6 + 4 + 7 + 7}{4} = \frac{24}{4} = 6$$

Por lo tanto, el estimador de máxima verosimilitud para $\boldsymbol{\mu}$ es:

$$\boxed{\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}}$$

b) **Calcular la matriz de covarianza muestral ($\hat{\boldsymbol{\Sigma}}$):** El EMV de $\boldsymbol{\Sigma}$ es $\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'$. Primero calculamos los vectores de desviación ($\mathbf{x}_j - \bar{\mathbf{x}}$):

$$\mathbf{x}_1 - \bar{\mathbf{x}} = \begin{bmatrix} 3 - 4 \\ 6 - 6 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

$$\mathbf{x}_2 - \bar{\mathbf{x}} = \begin{bmatrix} 4 - 4 \\ 4 - 6 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

$$\mathbf{x}_3 - \bar{\mathbf{x}} = \begin{bmatrix} 5 - 4 \\ 7 - 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\mathbf{x}_4 - \bar{\mathbf{x}} = \begin{bmatrix} 4 - 4 \\ 7 - 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Ahora calculamos los productos externos y su suma:

$$(\mathbf{x}_1 - \bar{\mathbf{x}})(\mathbf{x}_1 - \bar{\mathbf{x}})' = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$(\mathbf{x}_2 - \bar{\mathbf{x}})(\mathbf{x}_2 - \bar{\mathbf{x}})' = \begin{bmatrix} 0 \\ -2 \end{bmatrix} \begin{bmatrix} 0 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}$$

$$(\mathbf{x}_3 - \bar{\mathbf{x}})(\mathbf{x}_3 - \bar{\mathbf{x}})' = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$(\mathbf{x}_4 - \bar{\mathbf{x}})(\mathbf{x}_4 - \bar{\mathbf{x}})' = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\sum_{j=1}^4 (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix}$$

Finalmente, dividimos por $n = 4$:

$$\hat{\Sigma} = \frac{1}{4} \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix} = \boxed{\begin{bmatrix} 1/2 & 1/4 \\ 1/4 & 3/2 \end{bmatrix}}$$

(Nota: La matriz de covarianza muestral insesgada \mathbf{S} sería $\frac{1}{n-1} \sum (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' = \frac{1}{3} \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix}$).

4.19 Sea $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{20}$ una muestra aleatoria de tamaño $n = 20$ proveniente de una población $N_6(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Especifique completamente la distribución de cada uno de los siguientes elementos:

- (a) La distribución de $(\mathbf{X}_1 - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_1 - \boldsymbol{\mu})$
- (b) Las distribuciones de $\bar{\mathbf{X}}$ y $\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu})$
- (c) La distribución de $(n-1)\mathbf{S}$

Solución

Sea $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{20}$ una muestra aleatoria de tamaño $n = 20$ de una población $N_6(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

(a) Distribución de $(\mathbf{X}_1 - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_1 - \boldsymbol{\mu})$

Por propiedades de la distribución normal multivariada:

- $\mathbf{X}_1 - \boldsymbol{\mu} \sim N_6(\mathbf{0}, \boldsymbol{\Sigma})$
- $(\mathbf{X}_1 - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_1 - \boldsymbol{\mu}) \sim \chi_6^2$

$$\boxed{(\mathbf{X}_1 - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_1 - \boldsymbol{\mu}) \sim \chi_6^2 \quad (\text{Chi-cuadrado con 6 grados de libertad})}$$

(b) Distribuciones de $\bar{\mathbf{X}}$ y $\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu})$

Para la media muestral:

- $\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$
- $\bar{\mathbf{X}} \sim N_6(\boldsymbol{\mu}, \frac{1}{n} \boldsymbol{\Sigma})$
- $\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \sim N_6(\mathbf{0}, \boldsymbol{\Sigma})$

$$\boxed{\begin{aligned} \bar{\mathbf{X}} &\sim N_6(\boldsymbol{\mu}, \frac{1}{20} \boldsymbol{\Sigma}) \\ \sqrt{20}(\bar{\mathbf{X}} - \boldsymbol{\mu}) &\sim N_6(\mathbf{0}, \boldsymbol{\Sigma}) \end{aligned}}$$

(c) Distribución de $(n-1)\mathbf{S}$

Para la matriz de covarianza muestral:

- $\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$
- $(n-1)\mathbf{S} = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})' \sim W_6(n-1, \boldsymbol{\Sigma})$

$$\boxed{(n-1)\mathbf{S} \sim W_6(19, \boldsymbol{\Sigma}) \quad (\text{Distribución de Wishart con 19 grados de libertad})}$$

4.21 Sea $\mathbf{X}_1, \dots, \mathbf{X}_{60}$ una muestra aleatoria de tamaño 60 tomada de una distribución normal de dimensión cuatro, con vector media $\boldsymbol{\mu}$ y matriz de covarianza $\boldsymbol{\Sigma}$.

Especifique completamente cada uno de los siguientes:

- (a) La distribución de $\bar{\mathbf{X}}$
- (b) La distribución de $(\mathbf{X}_1 - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_1 - \boldsymbol{\mu})$
- (c) La distribución de $n(\bar{\mathbf{X}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})$
- (d) La distribución aproximada de $n(\bar{\mathbf{X}} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})$

Solución:

- (a) **Distribución de $\bar{\mathbf{X}}$**

Como la media de una muestra normal multivariada también es normal:

$$\bar{\mathbf{X}} \sim N_4 \left(\boldsymbol{\mu}, \frac{1}{60} \boldsymbol{\Sigma} \right)$$

- (b) **Distribución de $(\mathbf{X}_1 - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_1 - \boldsymbol{\mu})$**

Esta es una forma cuadrática centrada y estandarizada:

$$(\mathbf{X}_1 - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_1 - \boldsymbol{\mu}) \sim \chi_4^2$$

- (c) **Distribución de $n(\bar{\mathbf{X}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})$**

Esta es una forma cuadrática de la media muestral:

$$n(\bar{\mathbf{X}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}) \sim \chi_4^2$$

- (d) **Distribución aproximada de $n(\bar{\mathbf{X}} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})$**

$$n(\bar{\mathbf{X}} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}) \approx \chi_4^2$$

Esta aproximación es válida cuando n es suficientemente grande.

- 4.23 Considere las tasas anuales de retorno (incluyendo dividendos) del promedio industrial Dow-Jones para los años 1996-2005. Estos datos, multiplicados por 100, son:

$$-0,6, 3,1, 25,3, -16,8, -7,1, -6,2, 16,1, 25,2, 22,6, 26,0$$

Use estas 10 observaciones para completar lo siguiente: (a) Construya un gráfico Q-Q. ¿Parecen los datos estar normalmente distribuidos? Explique. (b) Realice una prueba de normalidad basada en el coeficiente de correlación r_Q . [Ver (4-31).] Sea el nivel de significancia $\alpha = ,10$.

Solución:

- (a) **Construcción del Gráfico Q-Q**

Primero, ordenamos las $n = 10$ observaciones de menor a mayor:

$$x_{(1)} = -16,8, x_{(2)} = -7,1, x_{(3)} = -6,2, x_{(4)} = -0,6, x_{(5)} = 3,1,$$

$$x_{(6)} = 16,1, x_{(7)} = 22,6, x_{(8)} = 25,2, x_{(9)} = 25,3, x_{(10)} = 26,0$$

Calculamos las probabilidades $p_j = (j - 1/2)/n = (j - 0,5)/10$ para $j = 1, 2, \dots, 10$. Luego, encontramos los cuantiles normales estándar $q_{(j)} = \Phi^{-1}(p_j)$ correspondientes a estas probabilidades usando una tabla normal estándar o software.

j	$x_{(j)}$ (Ordenado)	$p_j = (j - 0,5)/10$	$q_{(j)} = \Phi^{-1}(p_j)$
1	-16.8	0.05	-1.645
2	-7.1	0.15	-1.036
3	-6.2	0.25	-0.674
4	-0.6	0.35	-0.385
5	3.1	0.45	-0.126
6	16.1	0.55	0.126
7	22.6	0.65	0.385
8	25.2	0.75	0.674
9	25.3	0.85	1.036
10	26.0	0.95	1.645

El gráfico Q-Q se construye graficando los puntos $(q_{(j)}, x_{(j)})$.

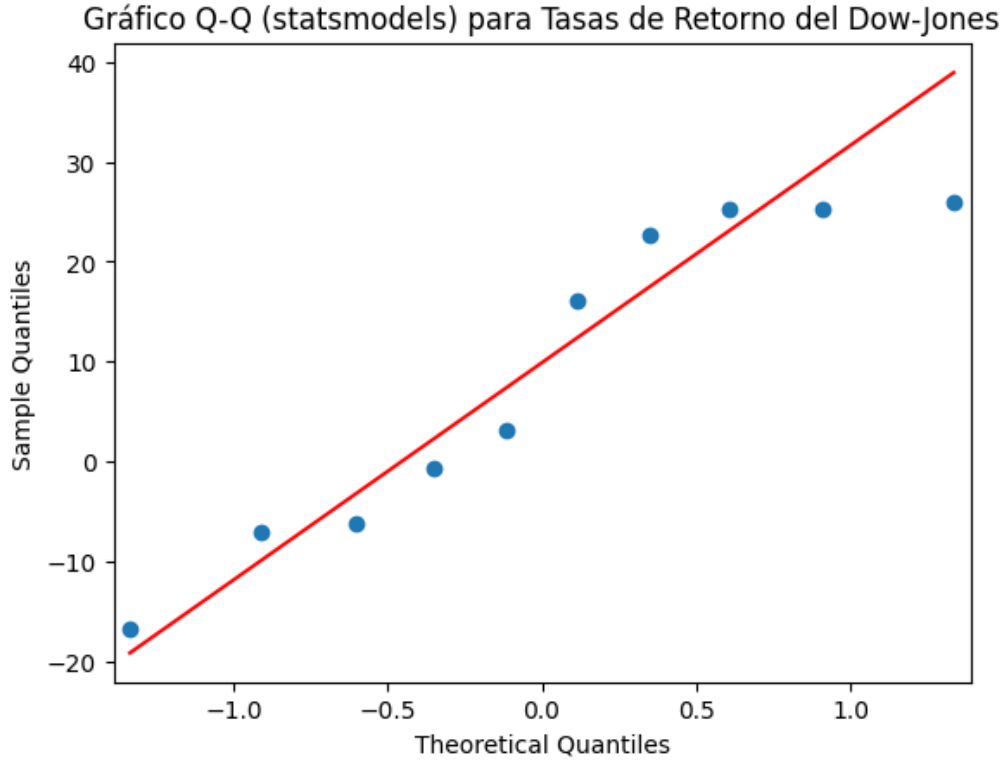


Figura 10: Q-Q plot para prueba de normalidad.

Como se observa en la figura 1 no sigue perfectamente los cuantiles esparados de una normal teorica, pero solo teniendo 10 datos no se podrian tomar conclusiones definitivas

(b) **Prueba de Normalidad usando r_Q**

El coeficiente de correlación para el gráfico Q-Q se calcula como:

$$r_Q = \frac{\sum_{j=1}^n (x_{(j)} - \bar{x})(q_{(j)} - \bar{q})}{\sqrt{\sum_{j=1}^n (x_{(j)} - \bar{x})^2} \sqrt{\sum_{j=1}^n (q_{(j)} - \bar{q})^2}}$$

Primero, calculamos las medias:

$$\bar{x} = \frac{1}{10} \sum_{j=1}^{10} x_{(j)} = \frac{-16,8 - 7,1 - \dots + 26,0}{10} = \frac{87,6}{10} = 8,76$$

$$\bar{q} = \frac{1}{10} \sum_{j=1}^{10} q_{(j)} = \frac{-1,645 - 1,036 - \dots + 1,645}{10} = \frac{0}{10} = 0 \quad (\text{por simetría})$$

Calculamos las sumas necesarias (se recomienda usar calculadora o software):

$$\sum_{j=1}^{10} (x_{(j)} - \bar{x})(q_{(j)} - \bar{q}) = \sum_{j=1}^{10} x_{(j)} q_{(j)} \approx 124,65$$

$$\sum_{j=1}^{10} (x_{(j)} - \bar{x})^2 = \sum_{j=1}^{10} x_{(j)}^2 - n\bar{x}^2 \approx 2790,46 - 10(8,76)^2 \approx 2023,4$$

$$\sum_{j=1}^{10} (q_{(j)} - \bar{q})^2 = \sum_{j=1}^{10} q_{(j)}^2 \approx 8,795$$

Ahora calculamos r_Q :

$$r_Q \approx \frac{124,65}{\sqrt{2023,4} \sqrt{8,795}} \approx \frac{124,65}{(44,98)(2,966)} \approx \frac{124,65}{133,41} \approx 0,934$$

La hipótesis nula es H_0 : Los datos provienen de una distribución normal. La hipótesis alternativa es H_1 : Los datos no provienen de una distribución normal. Usamos la Tabla 4.2 con $n = 10$ y nivel de significancia $\alpha = 0,10$. El valor crítico es 0,9351. Comparamos nuestro r_Q calculado con el valor crítico:

$$r_Q \approx 0,934 < 0,9351$$

Dado que r_Q es ligeramente menor que el valor crítico, rechazamos H_0 al nivel de significancia $\alpha = 0,10$. Hay evidencia estadística (aunque marginal, ya que los valores son muy cercanos) para sugerir que los datos no provienen de una distribución normal. Esto concuerda con la posible desviación observada en el gráfico Q-Q para el punto más bajo.

- 4.35 Examine los datos sobre las mediciones de calidad del papel en la Tabla 1.2 para evaluar la normalidad marginal y multivariante.

Solución

Para estudiar la normalidad marginal, se muestran los gráficos Q-Q por cada variable, junto con su histograma y la estimación de densidad normal, además se aplica una prueba de Shapiro-Wilk

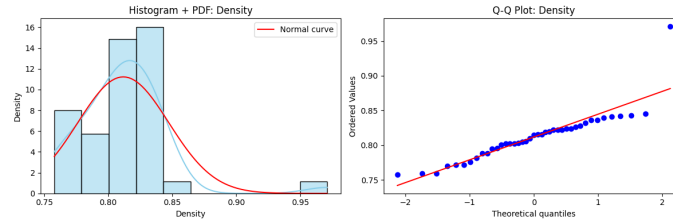


Figura 11: Estudio marginal para la variable “Density”

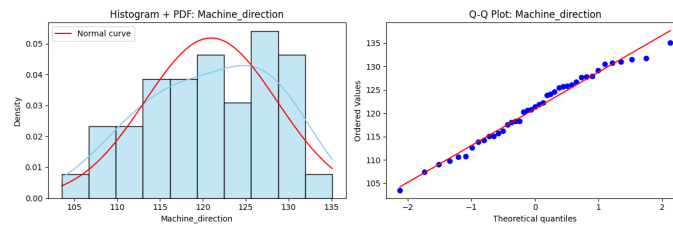


Figura 12: Estudio marginal para la variable “Machine_direction”

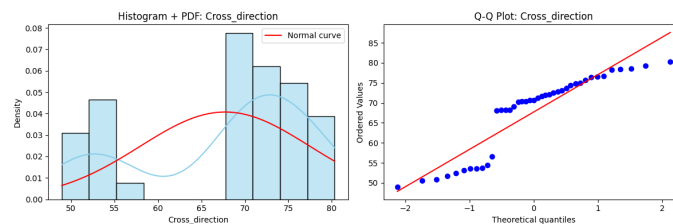


Figura 13: Estudio marginal para la variable “Cross_direction”

A partir de las gráficas y los valores de la prueba de Shapiro-Wilk en la tabla, se puede ver que la única variable para la cual que no rechazamos las hipótesis de distribución normal es para "Machine_direction"

Cuadro 7: Resultados de pruebas de normalidad marginal

Estadístico	Valor del estadístico	P-valor
Density	0.821938	0.000016
Machine_direction	0.977283	0.574077
Cross_direction	0.843518	0.000052

Para revisar la normalidad multivariada podemos revisar los cuantiles de la distribución dada por las distancias de Mahalanobis, que si los datos se distribuyen normalmente debería ser una χ^2_3

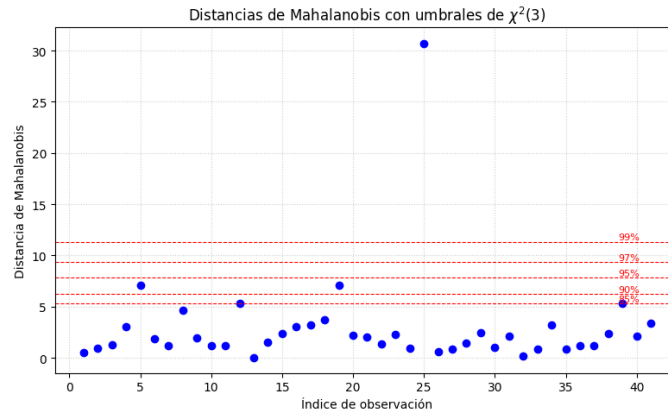


Figura 14: Revisión cuantiles χ^2_3

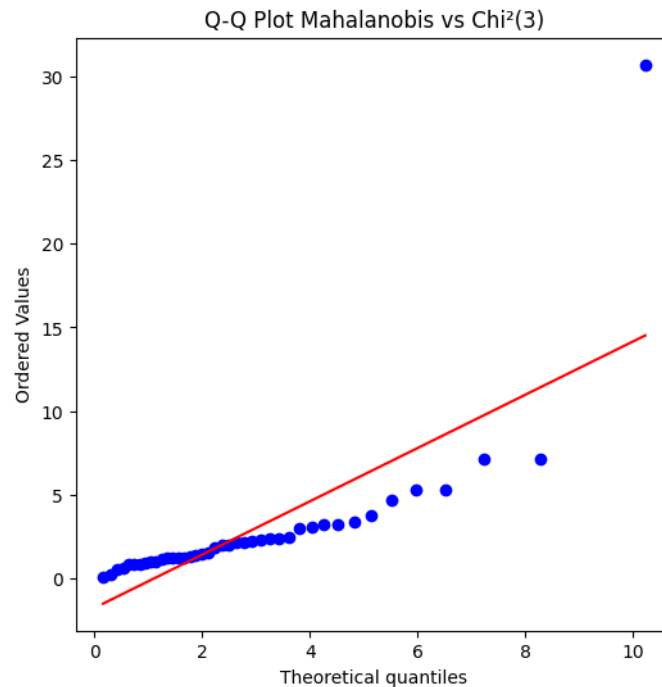


Figura 15: Gráfico Q-Q para χ^2_3

Podemos afirmar que no hay evidencia para pensar que los datos siguen una distribución normal multivariada.

2. Usando los datos de paper quality (Tabla 1.2, página 15).

- Considere las primeras 30 observaciones con el conjunto de observaciones históricas, y con base en ellas estime el vector de medias, la matriz de covarianzas y verifique si las observaciones provienen de una distribución normal p-variada.
- Con base en las últimas 11 observaciones determine si el proceso se encuentra bajo control.

Solución:

Se realizó 2 métodos para hacer el análisis de control, uno asumiendo normalidad y el otro por medio de Bootstrap. El análisis se condujo en dos fases sobre datos de especímenes con mediciones de Densidad (**Density**), Dirección de Máquina (**Machine_direction**) y Dirección Transversal (**Cross_direction**).

a) Asumiendo normalidad [Link](#):

1. Fase I: Establecimiento de Parámetros de Control y Verificación de Normalidad Se utilizaron las primeras 30 observaciones (conjunto **Y**) para esta fase, con el objetivo de estimar los parámetros del proceso (\vec{y} y **S**) y verificar la normalidad multivariante. EL vector de medias y matriz de varianzas y covarianzas estimadas son

■

$$\bar{\mathbf{x}} = \begin{bmatrix} 0,8173 \\ 122,6450 \\ 69,2000 \end{bmatrix}$$

■

$$\mathbf{S} = \begin{bmatrix} 0,001374 & 0,144762 & 0,189376 \\ 0,144762 & 52,217784 & 48,151617 \\ 0,189376 & 48,151617 & 81,925800 \end{bmatrix}$$

1.1. Detección Iterativa de Atípicos y Evaluación de Normalidad Se calculó la distancia de Mahalanobis al cuadrado (D_i^2) para cada observación respecto a la media y matriz de covarianzas del conjunto actual. Estas distancias se compararon con un Límite de Control Superior (LCS) de $\chi_{p,\alpha}^2$ (para $p = 3$ variables y $\alpha = 0,05$, $LCS \approx 7,815$). Paralelamente, se evaluó la normalidad multivariante mediante:

- **Gráficos QQ multivariantes:** Comparación visual de las D_i^2 ordenadas contra los cuantiles teóricos de una distribución χ_3^2 .

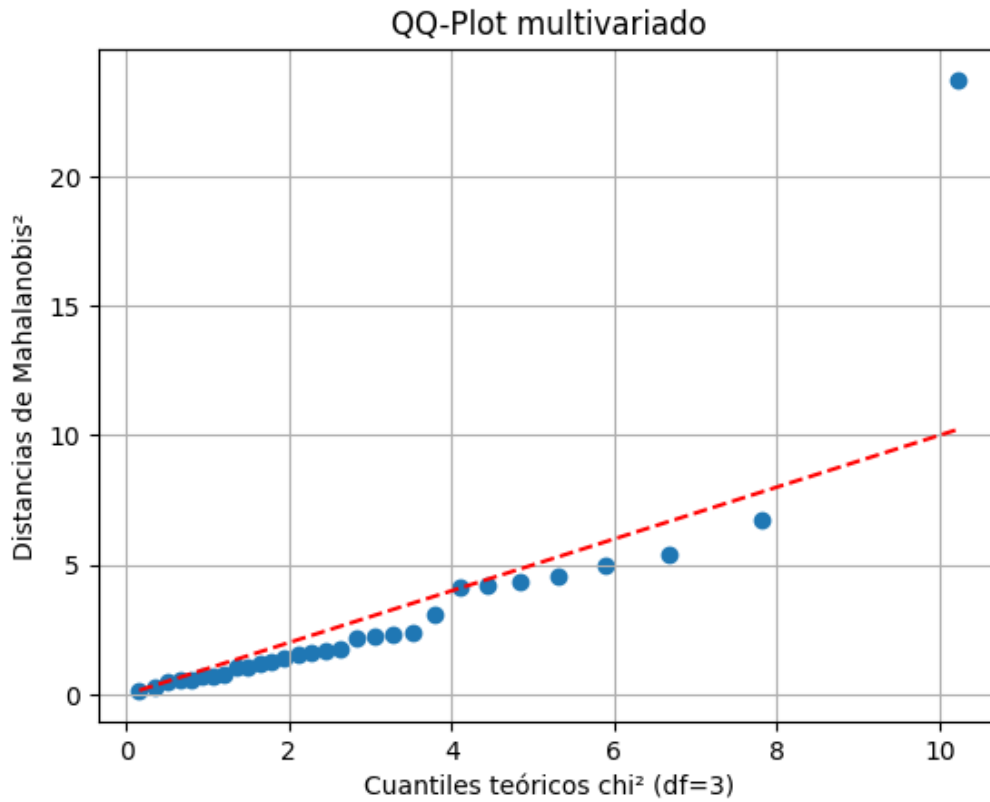


Figura 16: Q-Q Plot multivariado para los datos de la calidad del papel

La figura 16 muestra el gráfico QQ multivariado para el conjunto inicial \mathbf{Y} , donde las distancias de Mahalanobis al cuadrado (D^2) se comparan con los cuantiles teóricos de una distribución χ^2_3 . Se observa una ligera desviación hacia abajo de la línea de referencia para las distancias menores, pero más significativamente, los puntos con mayores D^2 , especialmente el último que excede un valor de 20, se desvían marcadamente hacia arriba. Esta marcada desviación en la cola superior, indicativa de observaciones con distancias al centroide multivariado mayores a las esperadas bajo normalidad, sugiere fuertemente que el conjunto original \mathbf{Y} no sigue una distribución normal multivariada y contiene datos atípicos.

- **Prueba de Normalidad Multivariante de Henze-Zirkler:** Esta prueba estadística evalúa si un conjunto de datos p -variado proviene de una distribución normal multivariante. Calcula un estadístico HZ basado en distancias entre pares de observaciones (escaladas por la covarianza y un parámetro de suavizado β). La distribución de HZ bajo la hipótesis nula de normalidad se aproxima con una distribución log-normal para obtener un p -valor. Un p -valor bajo (ej. $< 0,05$) sugiere desviación de la normalidad. Se aplicó este test de normalidad multivariada obteniendo un p -valor de $3,32 \cdot 10^{-7}$ donde demuestra que los datos no siguen una distribución normal multivariada.

El proceso iterativo fue:

- 1) **Análisis inicial de \mathbf{Y} (30 obs.):** Se identificó el Especímen 24 como atípico ($D^2 > \text{LCS}$) como se muestra en 17. La prueba de Henze-Zirkler indicó no normalidad.

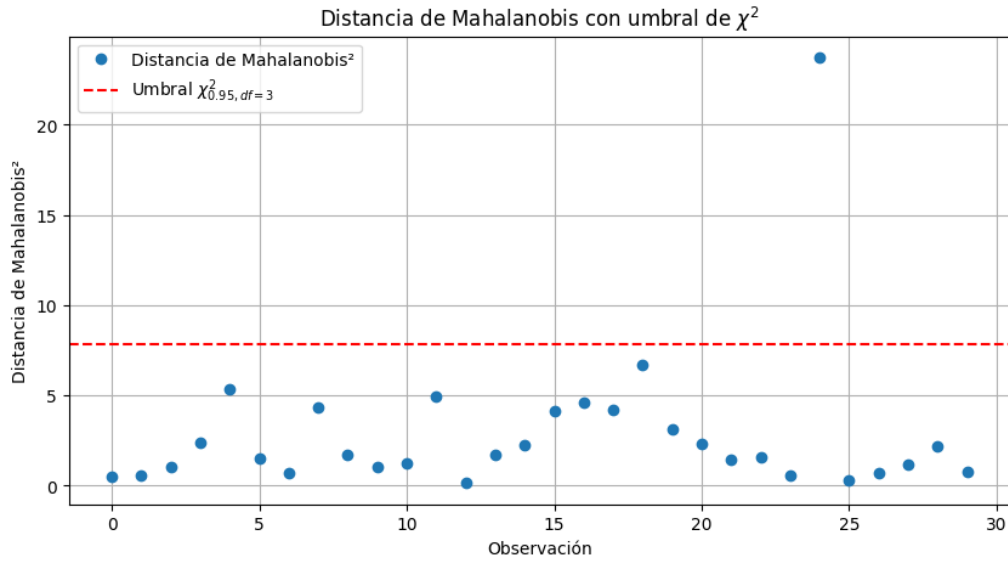


Figura 17: Identificación del umbral para las distancias de mahalanobis en el conjunto de datos \mathbf{Y}

- 2) **Conjunto \mathbf{Y}_1 (29 obs.):** Tras eliminar el Espécimen 24, se reanalizó. El Espécimen 18 resultó atípico en este nuevo contexto mostrado en la figura 18.

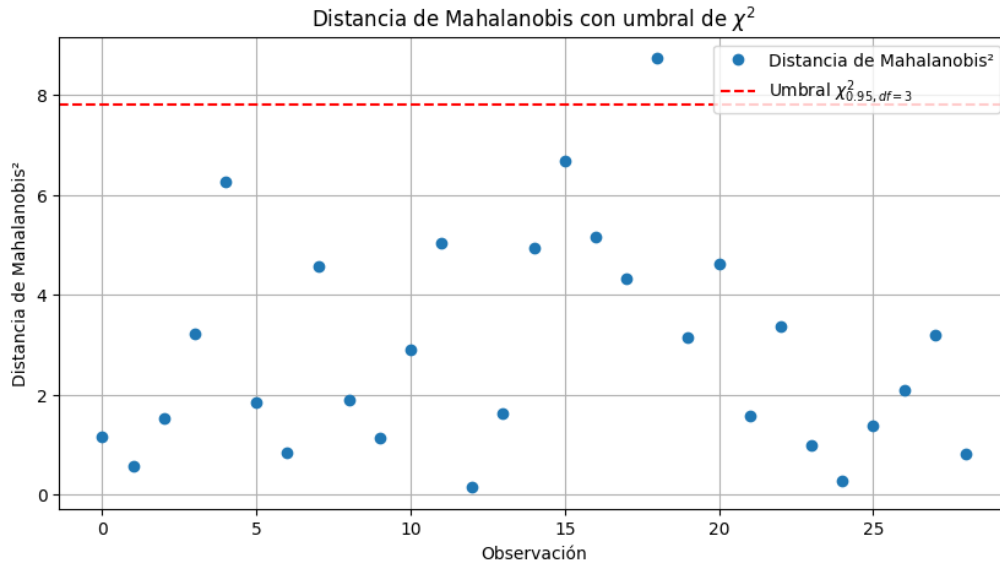


Figura 18: Identificación del umbral para las distancias de mahalanobis en el conjunto de datos \mathbf{Y}_1

- 3) **Conjunto \mathbf{Y}_2 (28 obs.):** Tras eliminar el Espécimen 18, todas las observaciones restantes en \mathbf{Y}_2 mostraron $D^2 < LCS$ como se muestra en 19. Se asumió una conformidad razonable con la normalidad multivariante para este conjunto depurado. El vector de medias $\bar{\mathbf{y}}_2$ y la matriz de covarianzas \mathbf{S}_2 de \mathbf{Y}_2 se usaron como referencia.

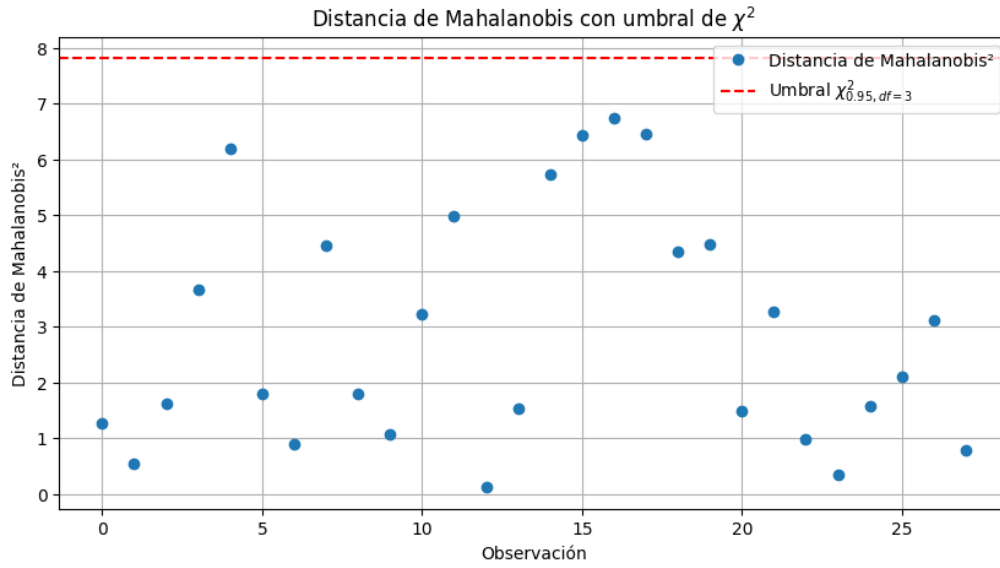


Figura 19: Identificación del umbral para las distancias de mahalanobis en el conjunto de datos \mathbf{Y}_2

2. Fase II: Determinación del Estado de Control del Proceso Las últimas 11 observaciones (Especímenes 31 al 41, conjunto \mathbf{Z}) se utilizaron para determinar si el proceso estaba bajo control.

- Para cada nueva observación $\vec{z}_j \in \mathbf{Z}$, se calculó su $D_j^2(\text{nuevo})$ usando los parámetros $\bar{\vec{y}}_2$ y \mathbf{S}_2^{-1} estimados de \mathbf{Y}_2 en la Fase I.
- Estas $D_j^2(\text{nuevo})$ se compararon con el LCS $\approx 7,815$.

Todos los datos del conjunto \mathbf{Z} son menores al limite LCS calculado anteriormente, mostrando que **el proceso esta controlado**

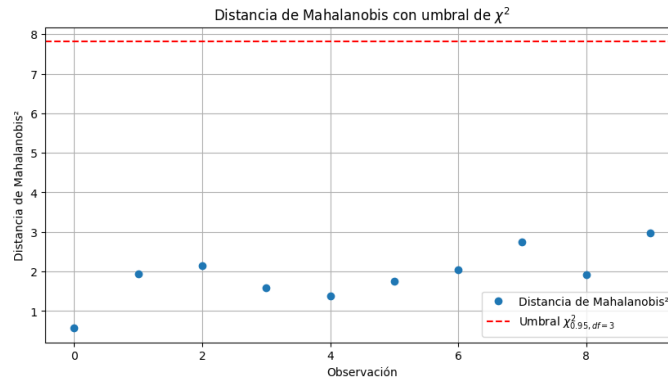


Figura 20: Identificación del umbral para los datos de la fase II.

b) Por medio de Bootstrap [Link](#):

Datos y preparación. Se parte de una matriz de datos con $p = 3$ variables: **Density**, **Machine.direction**, y **Cross.direction**. La muestra inicial consta de $n = 41$ observaciones, donde las primeras 30 se utilizan para la Fase I (estimación de parámetros) y las restantes 11 para la Fase II (monitoreo).

Cálculo de distancias T^2 . La estadística utilizada es la distancia cuadrada de Mahalanobis:

$$T_i^2 = (\mathbf{x}_i - \hat{\boldsymbol{\mu}})' \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})$$

donde:

- $\hat{\boldsymbol{\mu}}$: media muestral de las observaciones en la Fase I.

- $\hat{\Sigma}$: matriz de covarianza muestral de la Fase I.

Limpieza iterativa. Se realiza un procedimiento iterativo para eliminar posibles observaciones atípicas que podrían enmascarar el comportamiento real del proceso:

- 1) Se calcula T_i^2 para todas las observaciones actuales.
- 2) Se genera un umbral empírico mediante remuestreo bootstrap sobre el conjunto limpio actual.
- 3) Si existen observaciones con T_i^2 por encima del umbral, se elimina la observación con el mayor valor.
- 4) Se repite el procedimiento hasta que todas las observaciones estén bajo control o se alcance un número máximo de iteraciones.

Las siguientes figuras ilustran las tres iteraciones necesarias en la Fase I para eliminar los puntos fuera de control:

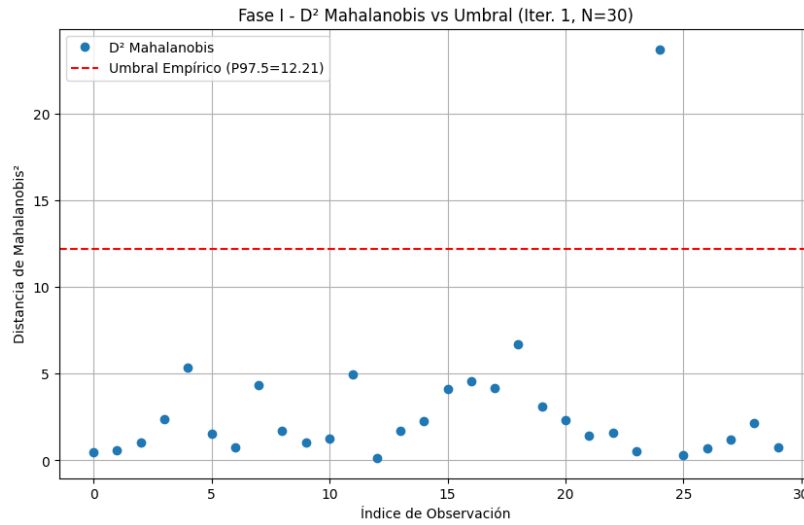


Figura 21: Primera iteración: eliminación del primer dato fuera de control.

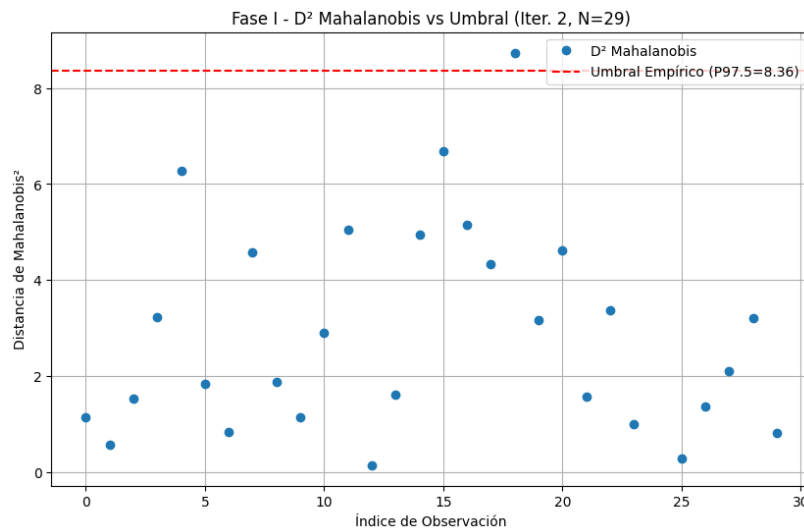


Figura 22: Segunda iteración: nuevo umbral con conjunto actualizado.

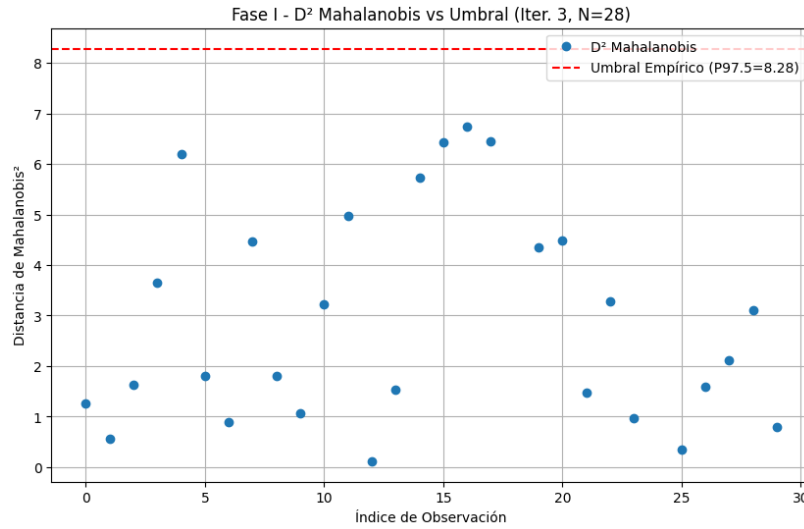


Figura 23: Tercera iteración: todas las observaciones bajo control.

Bootstrap para determinar el umbral. Se generan $B = 2000$ muestras bootstrap con reemplazo del conjunto limpio actual. Para cada remuestra se calcula la distribución empírica de distancias T^2 , de la cual se extrae el percentil deseado como umbral:

$$\text{Umbral}_{\text{Fase I}} = \text{percentil}_{97.5} \left(\{(T_b^2)^{(i)}\}_{i=1}^B \right)$$

Validación de normalidad. En cada iteración se realizan dos herramientas diagnósticas:

- Un QQ-plot de las distancias T^2 contra la distribución teórica χ_p^2
- Una prueba de normalidad multivariada de Henze-Zirkler

Fase II: Monitoreo. Con los parámetros estimados de la Fase I, se evalúan las observaciones nuevas. Se calcula T_i^2 para cada nueva observación y se compara contra el umbral empírico. Se reporta como fuera de control cualquier observación para la cual:

$$T_i^2 > \text{Umbral}_{\text{Fase I}}$$

Además, se grafican los valores de T^2 en la Fase II junto con el umbral para su interpretación visual:

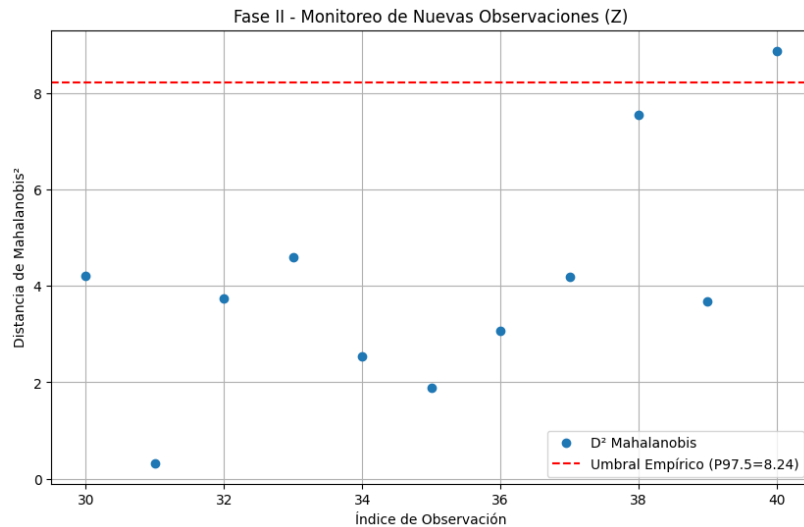


Figura 24: Carta de control en Fase II: se detecta un dato fuera de control.

Referencias

- Boudt, K., Rousseeuw, P. J., Vanduffel, S., & Verdonck, T. (2020). *The minimum regularized covariance determinant estimator*. *Statistics and Computing*, 30(1), 113–128.
- Engel, J., Buydens, L., & Blanchet, L. (2017). *An overview of large-dimensional covariance and precision matrix estimators with applications in chemometrics*. *Journal of Chemometrics*, 31(4), e2880.
- Hubert, M., & Debruyne, M. (2010). *Minimum covariance determinant*. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1), 36–43.
- Johnson, R. A., & Wichern, D. W. (2014). *Applied multivariate statistical analysis*. Pearson Educación Limited.
- Ojeda Martínez de Castilla, I., & Gago Vargas, J. (2008). *Métodos matemáticos para estadística*. Universidad de Extremadura, Servicio de Publicaciones.
- Zuo, Y., & Serfling, R. (2000). *General notions of statistical depth function*. *Annals of Statistics*, 461–482.