Distribución Normal Multivariada

José A. Perusquía Cortés

Análisis Multivariado Semestre 2023-2

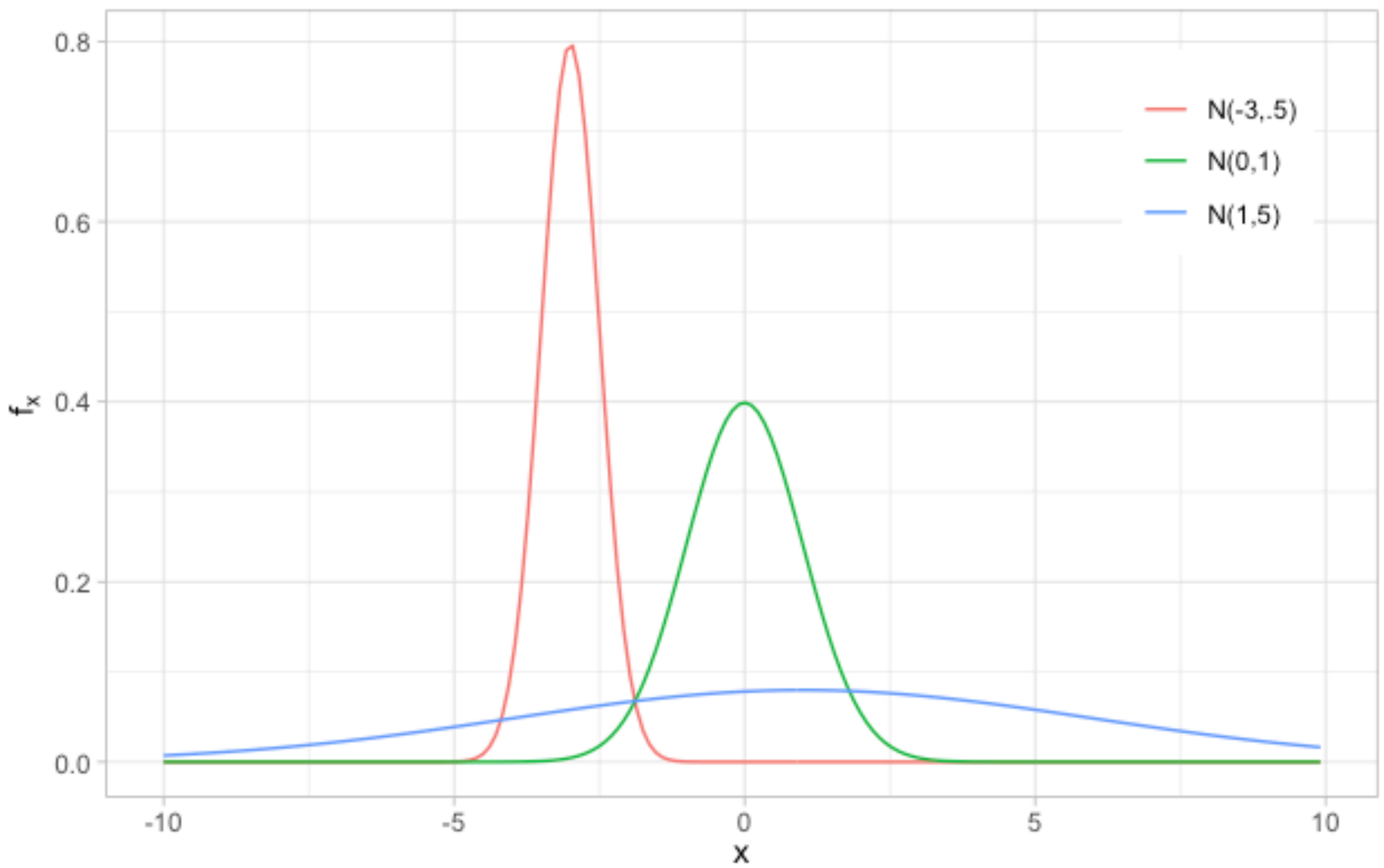
Distribución Normal Univariada

- Decimos que $X \sim N(\mu, \sigma^2)$ si tiene función de densidad

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

- Donde
  - $\mathbb{E}(X) = \mu \in \mathbb{R}$ (parámetro de localización)
  - $Var(X) = \sigma^2 \in \mathbb{R}^+$ (parámetro de escala)

Distribución Normal Univariada



- Decimos que $\mathbf{x} \sim N_p(\mu, \Sigma)$ (no singular) si tiene función de densidad

$$f(\mathbf{x}) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right]$$

- Donde
  - $\mathbb{E}(\mathbf{x}) = \mu$
  - $Var(\mathbf{x}) = \Sigma > 0$ (positiva definida)

- Para variables normales multivariadas en **R** podemos usar la librería mvtnorm

  - dmvnorm() : Evaluar la densidad.
  - pmvnorm() : Evaluar la distribución.
  - qmvnorm() : Obtener los cuantiles.
  - rmvnomr() : Obtener una muestra.

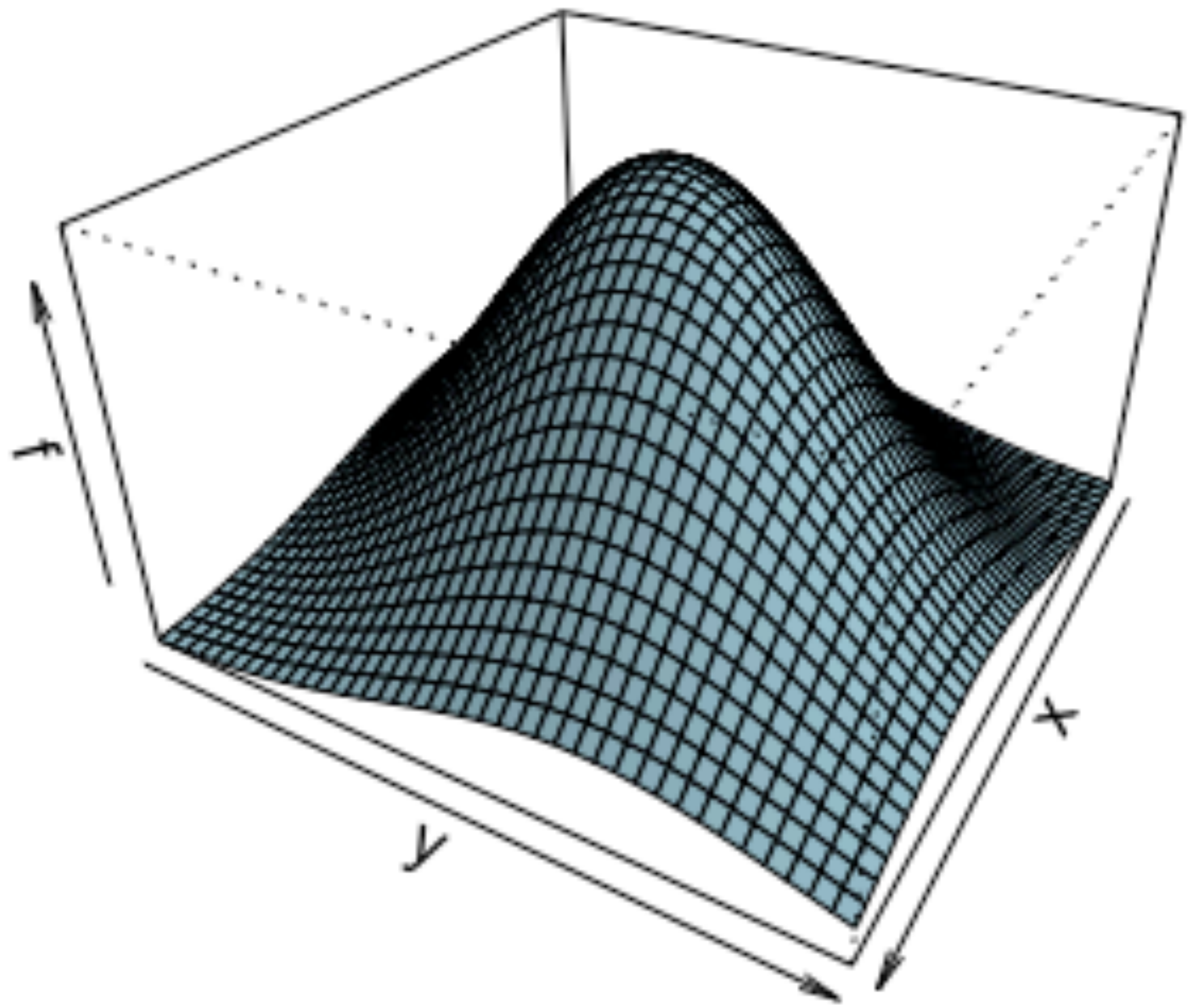
Distribución Normal Multivariada

- Por ejemplo, para dibujar la densidad

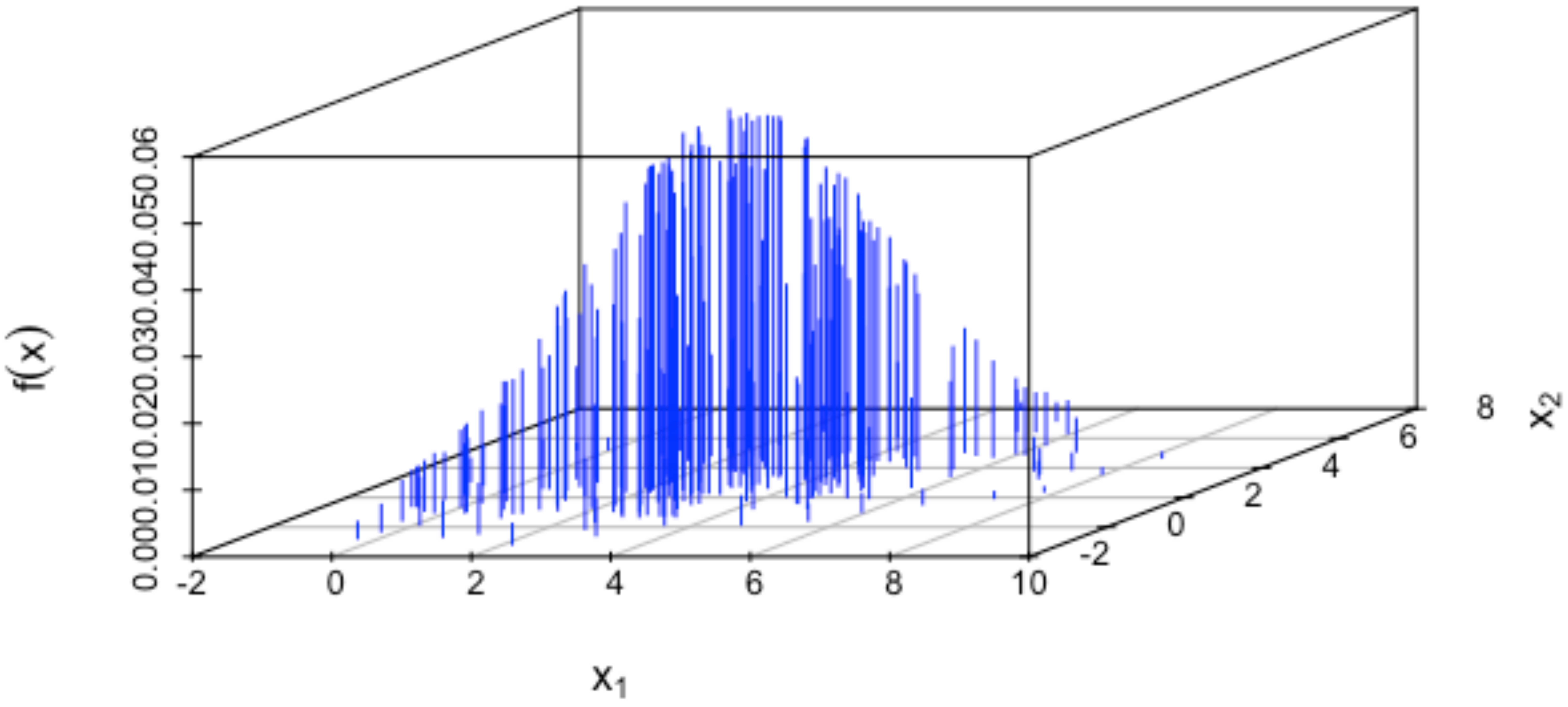
$$\mu = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

- Creamos un grid `expand.grid()`
- Evaluamos la densidad en el grid con `dmvnorm(grid, mu, sigma)`
- Convertimos a una matriz el resultado anterior
- Graficamos con `persp()`

Distribución Normal Multivariada



- Para datos bivariados también se puede crear un scatterplot en 3D con librería scatterplot3d



- Si $\text{ran}(\Sigma) = k < p$ podemos definir la densidad (singular) como

$$f(\mathbf{x}) = \frac{(2\pi)^{-\frac{k}{2}}}{(\lambda_1 \cdots \lambda_k)^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^-(\mathbf{x}-\mu)\right]$$

- Donde
  - $\mathbf{x}$ vive en el híperplano $\mathbf{N}'(\mathbf{x}-\mu)$ donde $\mathbf{N}$ es una matriz de tamaño $p \times (p-k)$ tal que $\mathbf{N}^T\Sigma = \mathbf{0}, \mathbf{N}^T\mathbf{N} = \mathbf{I_{p-k}}$
  - Σ^- es la inversa generalizada y $\lambda_1, \ldots, \lambda_k$ son los eigenvalores diferentes de cero.

- **Definición.**

  Decimos que $\mathbf{x}$ tiene una distribución normal p-variada si y solo si $\mathbf{a}^T\mathbf{x}$ tiene una distribución normal univariada para todos los vectores p-variados (no triviales) $\mathbf{a}$

- **Definición.**

  Decimos que $\mathbf{x}$ tiene una distribución normal p-variada si y solo si $\mathbf{a}^T\mathbf{x}$ tiene una distribución normal univariada para todos los vectores p-variados (no triviales) $\mathbf{a}$

- **Proposición 1**

  Sea $\mathbf{x}$ un vector normal p-variado y definamos a $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ donde $\mathbf{A}$ es una matriz de dimensión $q \times p$. Entonces $\mathbf{y}$ tiene una distribución normal q-variada tal que

$$\mathbb{E}(\mathbf{y}) = \mathbf{A}\mu + \mathbf{b} \qquad Var(\mathbf{y}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T$$

- **Corolario 1**

  Sea $\mathbf{x} \sim N_p(\mathbf{0}, \mathbf{I}_p)$ y definamos a $\mathbf{y} = \Sigma^{\frac{1}{2}}\mathbf{x} + \mu$ entonces $\mathbf{y} \sim N_p(\mu, \Sigma)$

- **Corolario 1**

  Sea $\mathbf{x} \sim N_p(\mathbf{0}, \mathbf{I}_p)$ y definamos a $\mathbf{y} = \Sigma^{\frac{1}{2}}\mathbf{x} + \mu$ entonces $\mathbf{y} \sim N_p(\mu, \Sigma)$

- **Corolario 2**

  Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$ $(\Sigma > 0)$ y definamos a $\mathbf{y} = \Sigma^{-\frac{1}{2}}(\mathbf{x} - \mu)$ donde $\Sigma^{-\frac{1}{2}}$ es la matrix raíz cuadrada de Σ^{-1}. Entonces $y_1, y_2, \ldots, y_p$ son variables aleatorias iid $N(0,1)$.

- **Corolario 1**

  Sea $\mathbf{x} \sim N_p(\mathbf{0}, \mathbf{I}_p)$ y definamos a $\mathbf{y} = \Sigma^{\frac{1}{2}}\mathbf{x} + \mu$ entonces $\mathbf{y} \sim N_p(\mu, \Sigma)$

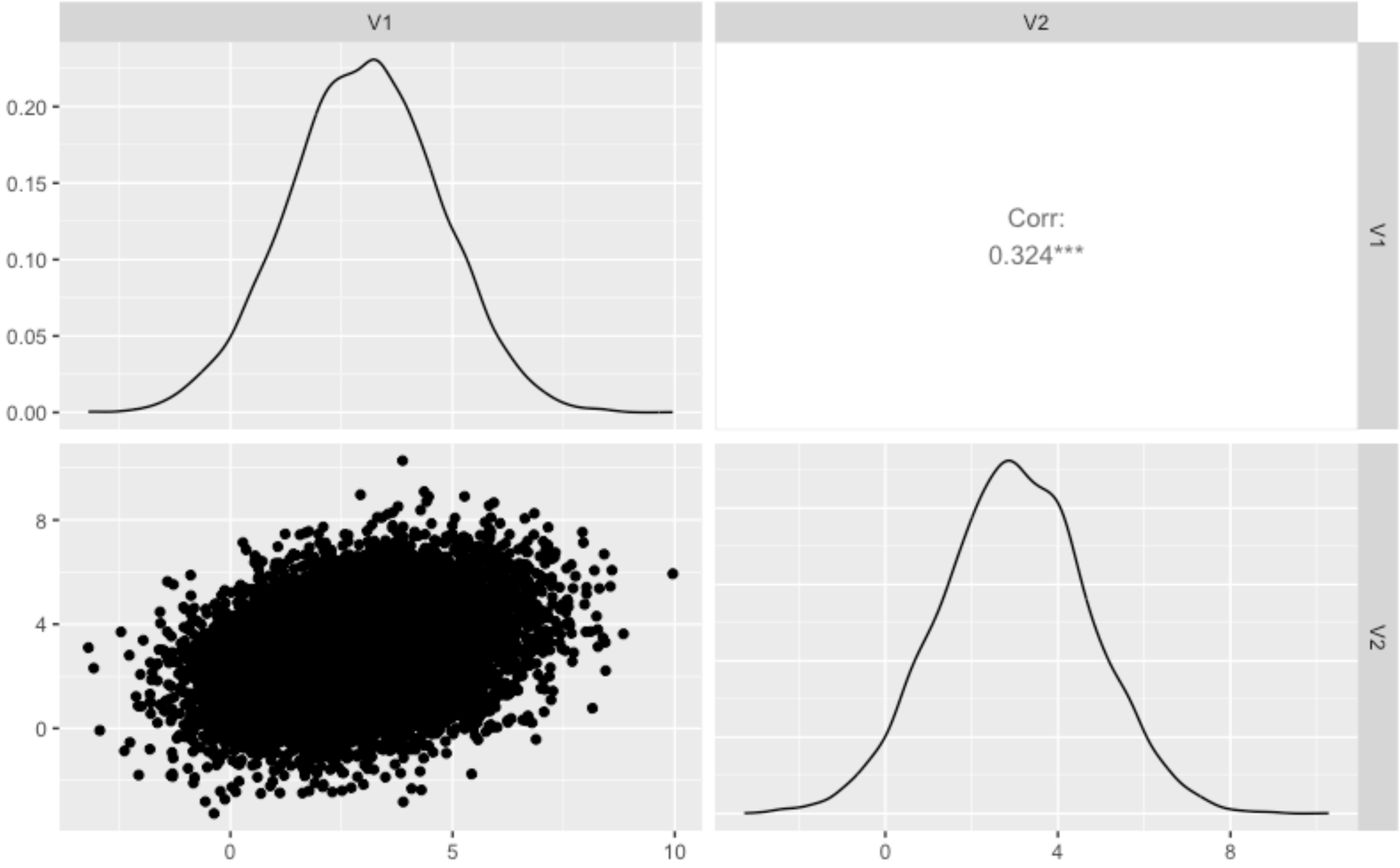
- **Corolario 2**

  Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$ $(\Sigma > 0)$ y definamos a $\mathbf{y} = \Sigma^{-\frac{1}{2}}(\mathbf{x} - \mu)$ donde $\Sigma^{-\frac{1}{2}}$ es la matrix raíz cuadrada de Σ^{-1}. Entonces $y_1, y_2, \ldots, y_p$ son variables aleatorias iid $N(0,1)$.

- En **R** la librería expm proporciona la función requerida para obtener $\Sigma^{-\frac{1}{2}}$ con sqrtm()

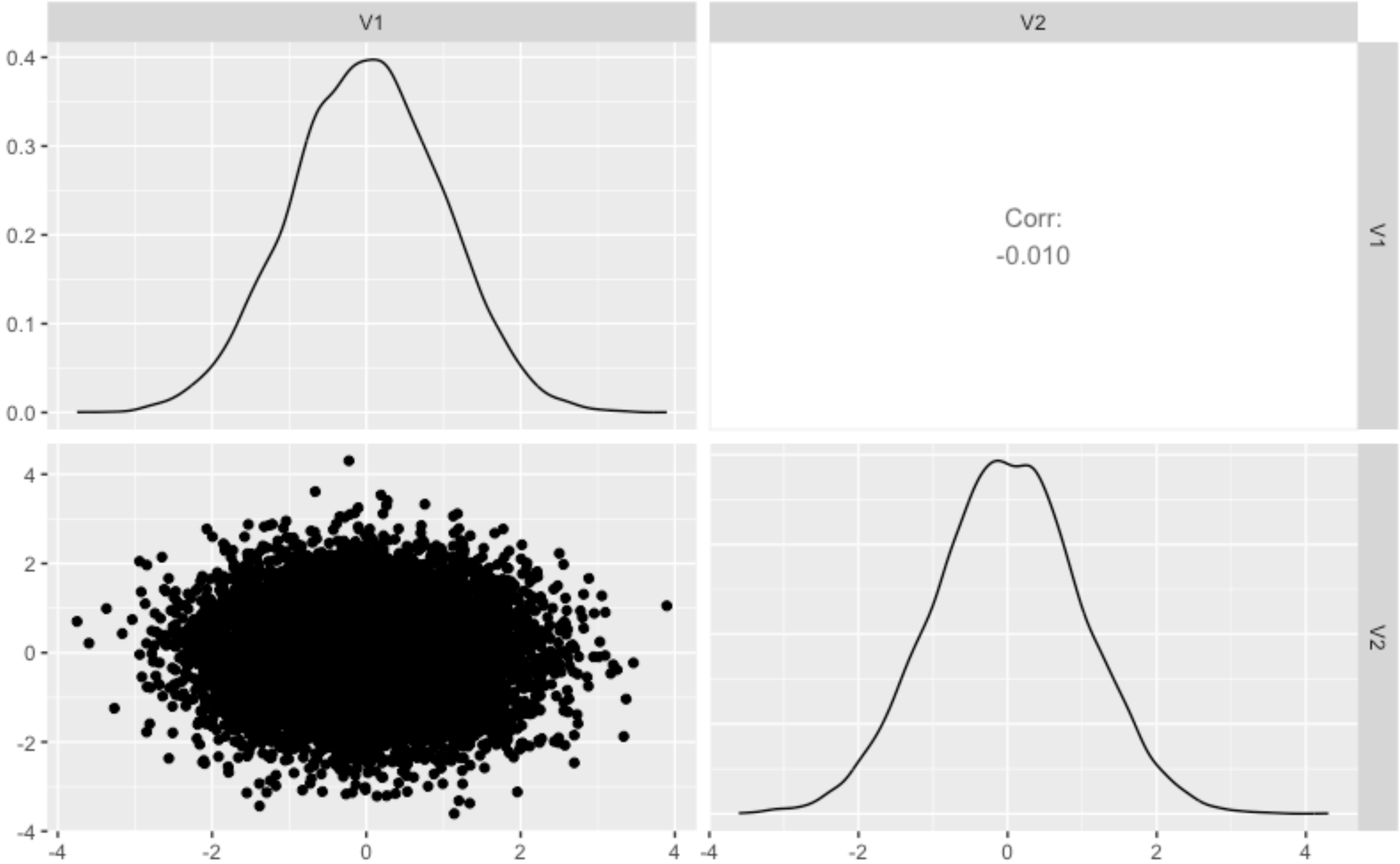
Distribución Normal Multivariada

$$\mathbf{x} \sim N_p(\mu, \Sigma) \qquad \mu = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$



Distribución Normal Multivariada

$$\mathbf{y} = \Sigma^{-\frac{1}{2}}(\mathbf{x} - \mu)$$



- **Observación**

La distribución normal multivariada tiene densidad constante en elipses (elipsoides)

$$(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)=k$$

- **Observación**

  La distribución normal multivariada tiene densidad constante en elipses (elipsoides)

$$(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)=k$$

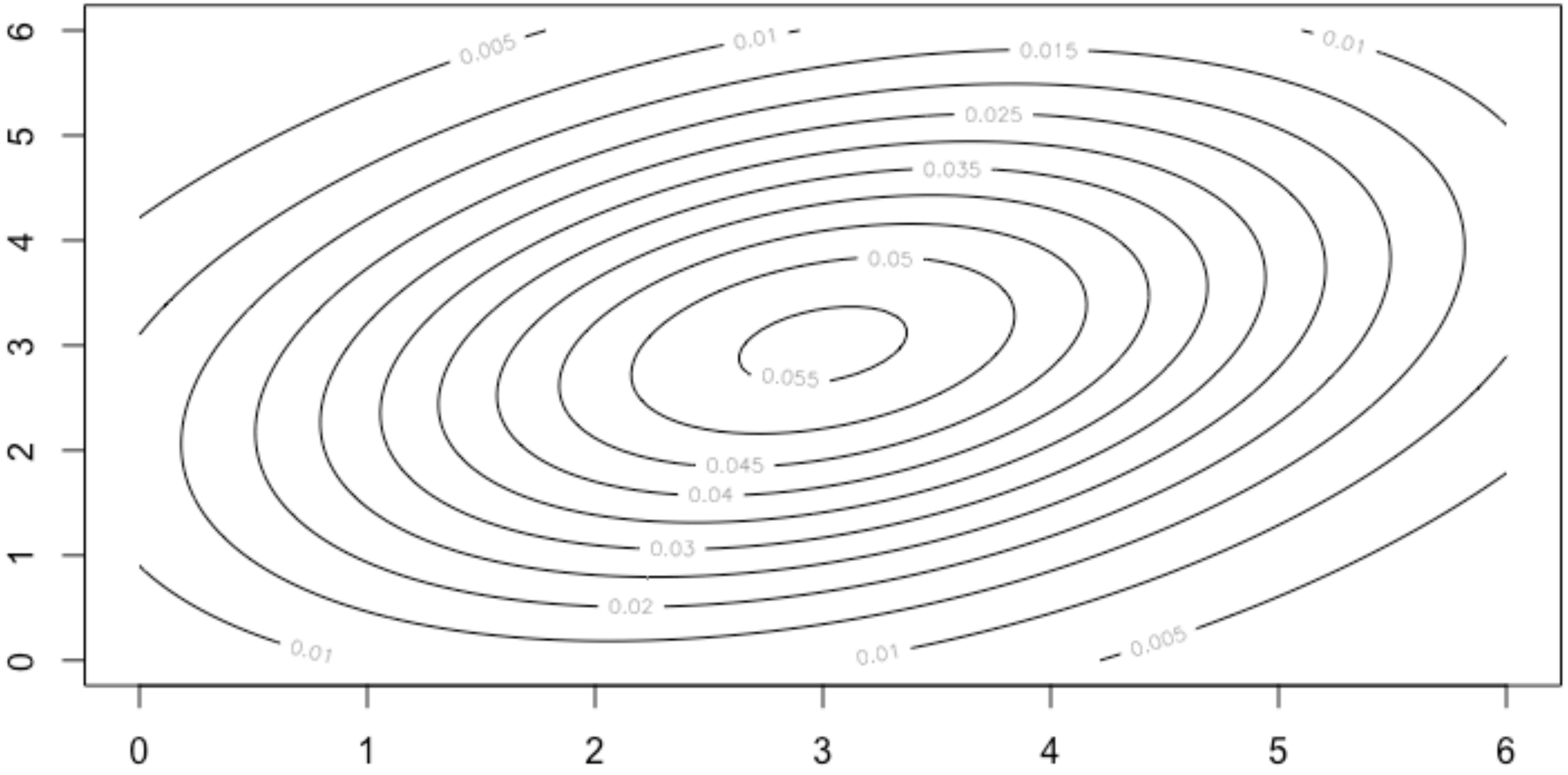
- En **R** podemos graficar las curvas mediante la función contour()

  - Generar vectores x, y donde evaluar la densidad (e.g. x=seq(0,6,length.out=40) , y =seq(0,6,length.out=40))

  - Evaluar la densidad en estos puntos con z=dmvnorm()

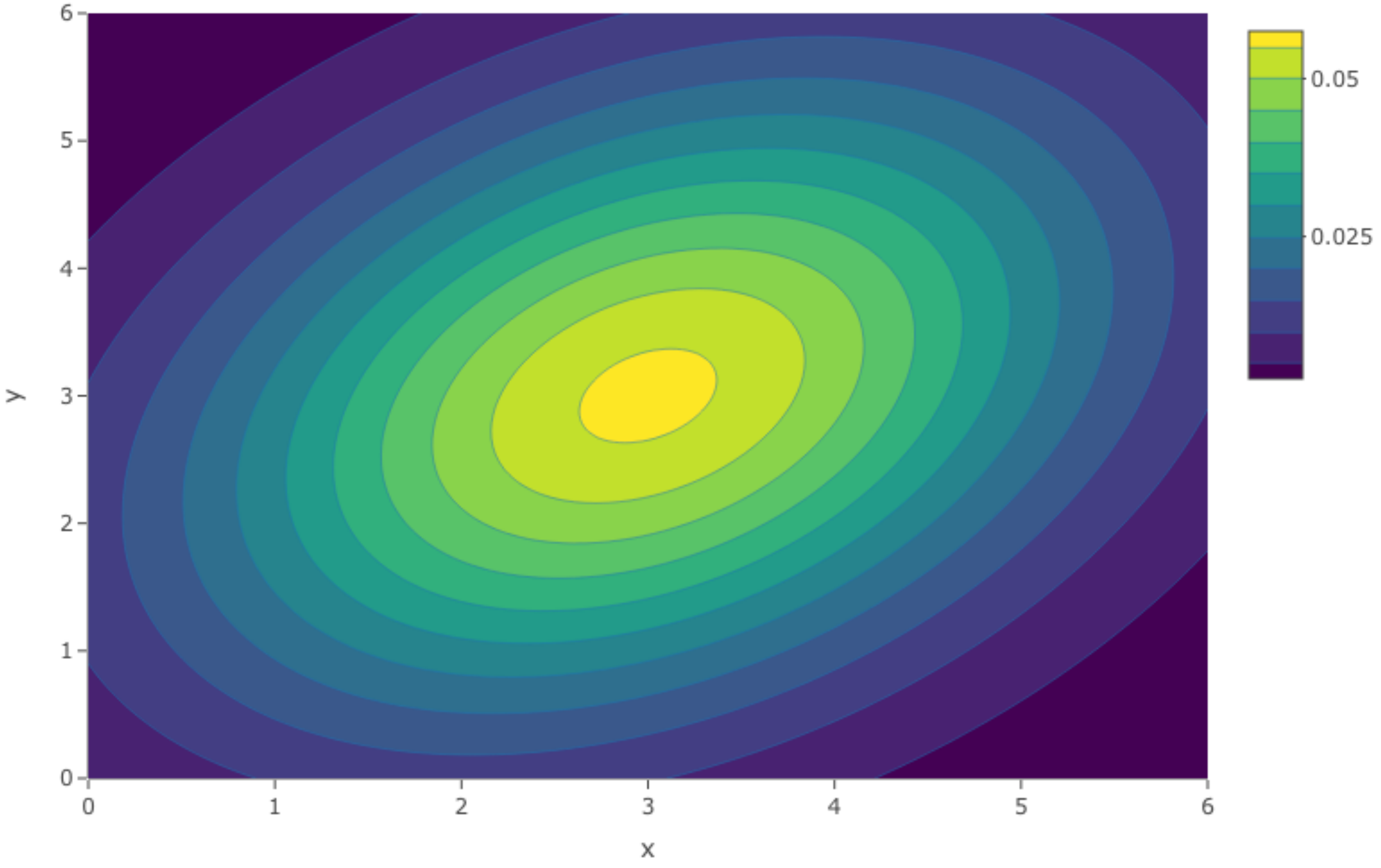
  - Usar la función contour(x,y,z)

Distribución Normal Multivariada



- Otra alternativa es usar la librería plotly para una gráfica más interactiva
  - Generar vectores x, y donde evaluar la densidad (e.g. `x=seq(0,6,length.out=40)` , `y =seq(0,6,length.out=40)`)
  - Evaluar la densidad en estos puntos con `z=dmvnorm()`
  - Usar la función `plot_ly(x,y,z,type = "contour")`

Distribución Normal Multivariada



- **Proposición 2**

  Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$ entonces $U = (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \sim \chi_p^2$.

- **Proposición 2**

  Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$ entonces $U = (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \sim \chi^2_p$.

- **Observación**

  Podemos fácilmente evaluar la probabilidad de que $\mathbf{x}$ este en un elipsoide, i.e.

$$\mathbb{P}[(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) < k]$$

- **Proposición 3 (Otras propiedades)**

  Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$ y sea

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{pmatrix} \qquad \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Entonces

1. Cualquier subconjunto de $\mathbf{x}$ se distribuye normal multivariado. En particular $\mathbf{x}^{(1)} \sim N_p(\mu^{(1)}, \Sigma_{11})$.

- **Proposición 3 (Otras propiedades)**

  Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$ y sea

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{pmatrix} \qquad \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Entonces

1. Cualquier subconjunto de $\mathbf{x}$ se distribuye normal multivariado. En particular $\mathbf{x}^{(1)} \sim N_p(\mu^{(1)}, \Sigma_{11})$.
2. $\mathbf{x}^{(1)}$ y $\mathbf{x}^{(2)}$ son independientes si y solo si $Cov(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = \mathbf{O}$.

- **Proposición 3 (Otras propiedades)**

Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$ y sea

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{pmatrix} \qquad \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Entonces

1. Cualquier subconjunto de $\mathbf{x}$ se distribuye normal multivariado. En particular $\mathbf{x}^{(\mathbf{1})} \sim N_p(\mu^{(1)}, \Sigma_{11})$.
2. $\mathbf{x}^{(1)}$ y $\mathbf{x}^{(2)}$ son independientes si y solo si $Cov(\mathbf{x}^{(\mathbf{1})}, \mathbf{x}^{(\mathbf{2})}) = \mathbf{O}$.
3. $\mathbf{x}^T\Sigma^{-1}\mathbf{x} \sim \chi^2_p(\mu^T\Sigma^{-1}\mu)$

- **Proposición 3 (Otras propiedades)**

  Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$ y sea

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{pmatrix} \qquad \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Entonces

1. Cualquier subconjunto de $\mathbf{x}$ se distribuye normal multivariado. En particular $\mathbf{x}^{(1)} \sim N_p(\mu^{(1)}, \Sigma_{11})$.
2. $\mathbf{x}^{(1)}$ y $\mathbf{x}^{(2)}$ son independientes si y solo si $Cov(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = \mathbf{O}$.
3. $\mathbf{x}^T\Sigma^{-1}\mathbf{x} \sim \chi^2_p(\mu^T\Sigma^{-1}\mu)$
4. $\mathbf{x}^{(2)} \,|\, \mathbf{x}^{(1)} \sim N_{p-k}\left(\mu^{(2)} + \Sigma_{21}\Sigma_{11}^{-1}[\mathbf{x}^{(1)} - \mu^{(1)}], \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)$

- **Teorema (Teorema Central del Límite)**

  Sean $\mathbf{x}_1, \mathbf{x}_2, \ldots \in \mathbb{R}^p$ vectores aleatorios independientes e idénticamente distribuidos con media μ y matriz (finita) de varianza Σ. Entonces se tiene que

$$\sqrt{n}(\bar{\mathbf{x}} - \mu) \to N_p(\mathbf{0}_p, \Sigma)$$

- **Definición**

Sea $\mathbf{x}$ un vector p-variado. Entonces la función característica está definida para $\mathbf{t} \in \mathbb{R}^p$ como

$$\phi(\mathbf{t}) = \mathbb{E}[\exp(i\mathbf{t}^T\mathbf{x})] = \int \exp(i\mathbf{t}^T\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

- **Definición**

Sea $\mathbf{x}$ un vector p-variado. Entonces la función característica está definida para $\mathbf{t} \in \mathbb{R}^p$ como

$$\phi(\mathbf{t}) = \mathbb{E}[\exp(i\mathbf{t}^T\mathbf{x})] = \int \exp(i\mathbf{t}^T\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

- **Propiedades (algunas)**

- $\phi(\mathbf{t})$ siempre existe, $\phi(\mathbf{0}) = 1$ y $|\phi(\mathbf{t})| \leq 1$.

- **Definición**

Sea $\mathbf{x}$ un vector p-variado. Entonces la función característica está definida para $\mathbf{t} \in \mathbb{R}^p$ como

$$\phi(\mathbf{t}) = \mathbb{E}[\exp(i\mathbf{t}^T\mathbf{x})] = \int \exp(i\mathbf{t}^T\mathbf{x})f(\mathbf{x})d\mathbf{x}$$

- **Propiedades (algunas)**

- $\phi(\mathbf{t})$ siempre existe, $\phi(\mathbf{0}) = 1$ y $|\phi(\mathbf{t})| \leq 1$.

- (Teorema de unicidad) $F(\mathbf{x}) = F(\mathbf{y}) \Leftrightarrow \phi_{\mathbf{x}}(\mathbf{t}) = \phi_{\mathbf{y}}(\mathbf{t})$

Función Característica

- **Definición**

Sea $\mathbf{x}$ un vector p-variado. Entonces la función característica está definida para $\mathbf{t} \in \mathbb{R}^p$ como

$$\phi(\mathbf{t}) = \mathbb{E}[\exp(i\mathbf{t}^T\mathbf{x})] = \int \exp(i\mathbf{t}^T\mathbf{x})f(\mathbf{x})d\mathbf{x}$$

- **Propiedades (algunas)**

- $\phi(\mathbf{t})$ siempre existe, $\phi(\mathbf{0}) = 1$ y $|\phi(\mathbf{t})| \leq 1$.
- (Teorema de unicidad) $F(\mathbf{x}) = F(\mathbf{y}) \Leftrightarrow \phi_{\mathbf{x}}(\mathbf{t}) = \phi_{\mathbf{y}}(\mathbf{t})$
- (Teorema de inversión) Si $\phi(\mathbf{t})$ es absolutamente integrable entonces $f(\mathbf{x}) = \dfrac{1}{(2\pi)^p}\displaystyle\int_{\mathbb{R}} \exp[-i\mathbf{t}^T\mathbf{x}]\phi(\mathbf{t})d\mathbf{t}$

Función Característica

- **Definición**

Sea $\mathbf{x}$ un vector p-variado. Entonces la función característica está definida para $\mathbf{t} \in \mathbb{R}^p$ como

$$\phi(\mathbf{t}) = \mathbb{E}[\exp(i\mathbf{t}^T\mathbf{x})] = \int \exp(i\mathbf{t}^T\mathbf{x})f(\mathbf{x})d\mathbf{x}$$

- **Propiedades (algunas)**

- $\phi(\mathbf{t})$ siempre existe, $\phi(\mathbf{0}) = 1$ y $|\phi(\mathbf{t})| \leq 1$.
- (Teorema de unicidad) $F(\mathbf{x}) = F(\mathbf{y}) \Leftrightarrow \phi_{\mathbf{x}}(\mathbf{t}) = \phi_{\mathbf{y}}(\mathbf{t})$
- (Teorema de inversión) Si $\phi(\mathbf{t})$ es absolutamente integrable entonces $f(\mathbf{x}) = \dfrac{1}{(2\pi)^p}\displaystyle\int_{\mathbb{R}} \exp[-i\mathbf{t}^T\mathbf{x}]\phi(\mathbf{t})d\mathbf{t}$
- $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}$ son independientes si y solo si $\phi_{\mathbf{x}}(\mathbf{t}) = \phi_{\mathbf{x}^{(1)}}(\mathbf{t}^{(1)})\phi_{\mathbf{x}^{(2)}}(\mathbf{t}^{(2)})$

- **Definición**

Sea $\mathbf{x}$ un vector p-variado. Entonces la función característica está definida para $\mathbf{t} \in \mathbb{R}^p$ como

$$\phi(\mathbf{t}) = \mathbb{E}[\exp(i\mathbf{t}^T\mathbf{x})] = \int \exp(i\mathbf{t}^T\mathbf{x})f(\mathbf{x})d\mathbf{x}$$

- **Propiedades (algunas)**

- $\phi(\mathbf{t})$ siempre existe, $\phi(\mathbf{0}) = 1$ y $|\phi(\mathbf{t})| \leq 1$.
- (Teorema de unicidad) $F(\mathbf{x}) = F(\mathbf{y}) \Leftrightarrow \phi_{\mathbf{x}}(\mathbf{t}) = \phi_{\mathbf{y}}(\mathbf{t})$
- (Teorema de inversión) Si $\phi(\mathbf{t})$ es absolutamente integrable entonces $f(\mathbf{x}) = \dfrac{1}{(2\pi)^p}\displaystyle\int_{\mathbb{R}} \exp[-i\mathbf{t}^T\mathbf{x}]\phi(\mathbf{t})d\mathbf{t}$
- $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}$ son independientes si y solo si $\phi_{\mathbf{x}}(\mathbf{t}) = \phi_{\mathbf{x}^{(1)}}(\mathbf{t}^{(1)})\phi_{\mathbf{x}^{(2)}}(\mathbf{t}^{(2)})$
- $\phi_{\mathbf{x}^{(1)}}(\mathbf{t}^{(1)}) = \phi_{\mathbf{x}}(\mathbf{t}^{(1)}, \mathbf{0})$

- **Teorema (Crámer-Wold)**

  La distribución de un vector aleatorio p-variado $\mathbf{x}$ está completamente determinado por el conjunto de todas las distribuciones de combinaciones lineales $\mathbf{t}^T\mathbf{x}$, con $\mathbf{t} \in \mathbb{R}^p$

- **Proposición 4**

  Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$ entonces

$$\phi(\mathbf{t}) = \exp\left(i\mathbf{t}^T\mu - \frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}\right)$$

- **Proposición 5 (Asimetría y Curtosis)**

  Sea $\mathbf{x} \sim N_p(\mu, \Sigma)$ entonces los coeficientes de asimetría y curtosis están dados respectivamente por

$$\beta_{1,p} = 0$$

$$\beta_{2,p} = p(p+2)$$

- **Checar normalidad**

- **Checar normalidad**
  - Todas las distribuciones univariadas son normales
    - qqplot
    - histogramas
    - Pruebas de normalidad (e.g. Anderson-Darling, Shapiro-Wilk, Lilliefors, etc.)

- **Checar normalidad**
  - Todas las distribuciones univariadas son normales
    - qqplot
    - histogramas
    - Pruebas de normalidad (e.g. Anderson-Darling, Shapiro-Wilk, Lilliefors, etc.)
  - Prueba basada en los coeficientes de asimetría y curtosis (Prueba de Mardia)

- **Checar normalidad**
  - Todas las distribuciones univariadas son normales
    - qqplot
    - histogramas
    - Pruebas de normalidad (e.g. Anderson-Darling, Shapiro-Wilk, Lilliefors, etc.)
  - Prueba basada en los coeficientes de asimetría y curtosis (Prueba de Mardia)
  - $(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu) \sim \chi^2_p$
    - qqplot

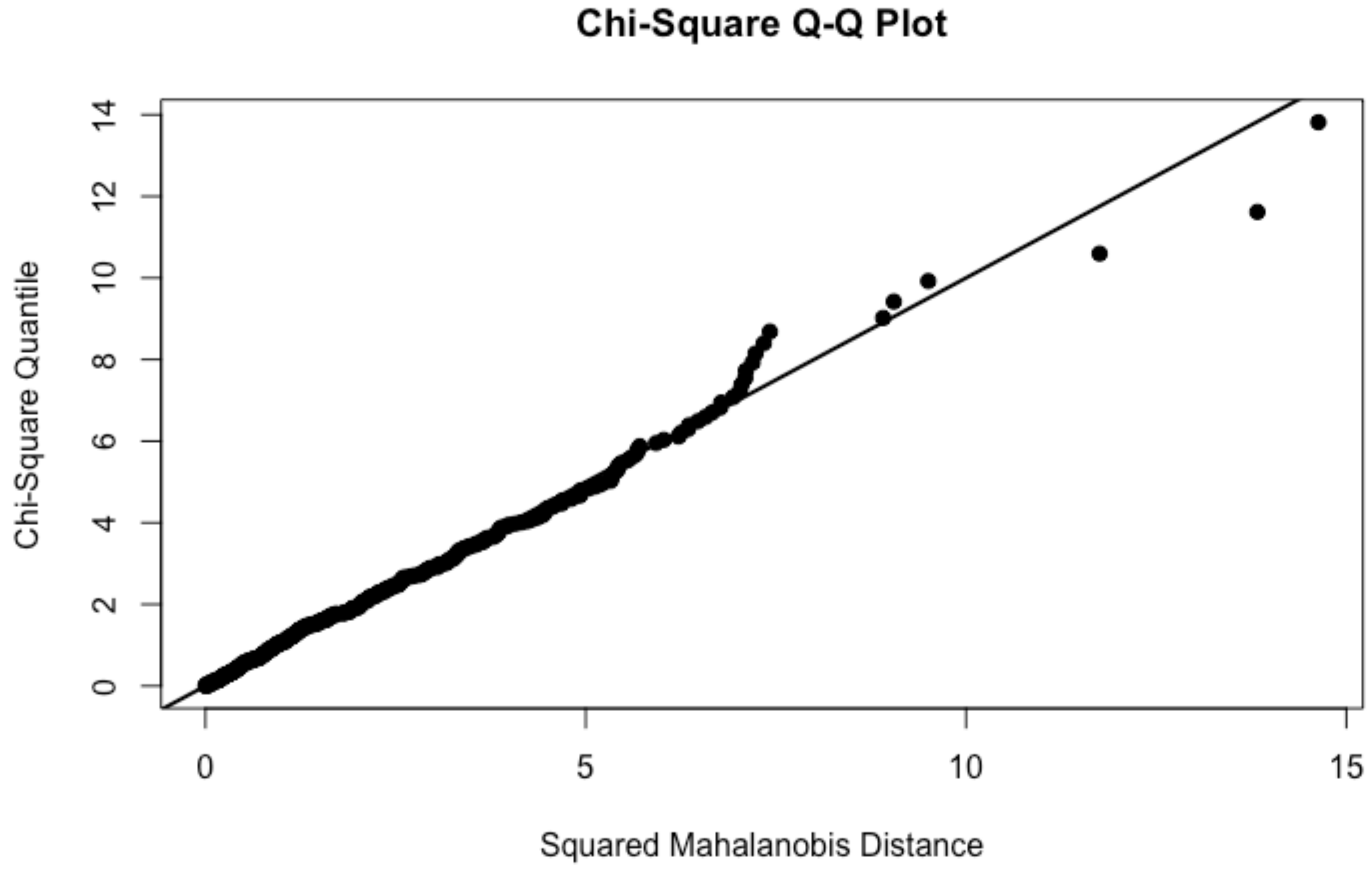
- **Checar normalidad**
  - Todas las distribuciones univariadas son normales
    - qqplot
    - histogramas
    - Pruebas de normalidad (e.g. Anderson-Darling, Shapiro-Wilk, Lilliefors, etc.)
  - Prueba basada en los coeficientes de asimetría y curtosis (Prueba de Mardia)
  - $(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu) \sim \chi^2_p$
    - qqplot
  - Otras pruebas de normalidad multivariadas

- En **R** la librería mvn provee todas las herramientas esto

```
mvn(multnorm.sample,mvnTest="mardia",multivariatePlot = "qq")
```

```
$multivariateNormality
             Test        Statistic           p value Result
1 Mardia Skewness 3.47421629178296 0.481809628750315    YES
2 Mardia Kurtosis 0.29380065853414 0.768910231846389    YES
3             MVN             <NA>              <NA>    YES

$univariateNormality
              Test  Variable Statistic   p value Normality
1 Anderson-Darling  Column1     0.3285    0.5161    YES
2 Anderson-Darling  Column2     0.2388    0.7793    YES
```



mvn(iris[,-5],mvnTest="mardia",univariatePlot = "qqplot")

```
$multivariateNormality
             Test          Statistic               p value Result
1 Mardia Skewness   67.4305087780629 4.75799820400705e-07     NO
2 Mardia Kurtosis -0.230112114480775    0.818004651478188    YES
3             MVN               <NA>                 <NA>     NO

$univariateNormality
              Test     Variable Statistic   p value Normality
1 Anderson-Darling Sepal.Length    0.8892  0.0225      NO
2 Anderson-Darling  Sepal.Width    0.9080  0.0202      NO
3 Anderson-Darling Petal.Length    7.6785  <0.001      NO
4 Anderson-Darling  Petal.Width    5.1057  <0.001      NO
```