

Métodos Estadísticos Aplicados

M. González

Departamento de Matemáticas. Universidad de Extremadura

Métodos Estadísticos Aplicados

- 1 **Análisis Exploratorio de Datos**
- 2 Análisis de Datos
- 3 Diseño de Experimentos

Métodos Estadísticos Aplicados

1 Análisis Exploratorio de Datos

2 Análisis de Datos

3 Diseño de Experimentos

Métodos Estadísticos Aplicados

- 1 Análisis Exploratorio de Datos
- 2 Análisis de Datos
- 3 Diseño de Experimentos

Medidas Descriptivas

Estadístico	¿Qué mide?	Robustez
Media (\bar{x})	Centralización de la distribución	No
Media recortada (\bar{x}_{trimm})	Centralización de la distribución	Sí
Mediana (\tilde{x})	Centralización de la distribución	Sí
Media Geométrica (\bar{x}_g)	Centralización de la distribución	Sí
Varianza (s^2)	Dispersión de la distribución	No
Desviación Típica (s)	Dispersión de la distribución	No
Rango (R)	Dispersión de la distribución	No
Rango Intercuartílico (R_I)	Dispersión de la distribución	Sí
Desviación Absoluta Mediana (MAD)	Dispersión de la distribución	Sí
Coficiente de variación (C.V.)	Dispersión/centralización de la distribución	No
Coficiente de Asimetría ($\hat{\gamma}_1$)	Forma de la distribución	No
Coficiente de Curtosis ($\hat{\gamma}_2$)	Forma de la distribución	No



Medidas Descriptivas

Media Recortada (trimmed mean):

$$\bar{x}_{\text{trimm}} = \frac{1}{n - 2[\alpha n]} \sum_{i=[\alpha n]+1}^{n-[\alpha n]} x_{(i)}$$

siendo $\alpha \in [0, 0.5]$ y $[x]$ el mayor entero $\leq x$.

- Robusta a veces (dependiendo del valor de α).

Media Geométrica:

$$\bar{x}_g = \exp \left\{ \frac{1}{n} \sum_{i=1}^n \log(x_i) \right\}$$

- Se utiliza habitualmente para describir datos positivos.
- Estima la verdadera mediana de la distribución log-normal ($Y \sim LN(\mu, \sigma^2)$ si y sólo si $\log(Y) \sim N(\mu, \sigma^2)$).
- $\bar{x}_g \leq \bar{x}$ (se da la igualdad si y sólo si todas las observaciones son iguales).
- Robusta (del mismo modo que la mediana).



Medidas Descriptivas

Desviación Absoluta mediana:

$$\text{MAD} = \text{mediana}(|x_i - \tilde{x}|, \quad i = 1, \dots, n)$$

- Robusta.

Coefficiente de Asimetría:

$$\hat{\gamma}_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3}$$

- $\hat{\gamma}_1 = 0$ —distribución simétrica. $\hat{\gamma}_1 > 0$ —distribución asimétrica hacia valores grandes de la variable. $\hat{\gamma}_1 < 0$ —distribución asimétrica hacia valores pequeños de la variable.
- No robusta.

Coefficiente de Curtosis:

$$\hat{\gamma}_2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{s^4} - 3$$

- $\hat{\gamma}_2 = 0$ —distribución normal. $\hat{\gamma}_2 > 0$ —más apuntada que la distribución normal. $\hat{\gamma}_2 < 0$ —más aplastada que la distribución normal.
- No robusta.

Estimación de la densidad

Estimadores núcleo de la densidad

- Muestra: x_1, \dots, x_n .
- Estimación de la función de densidad:

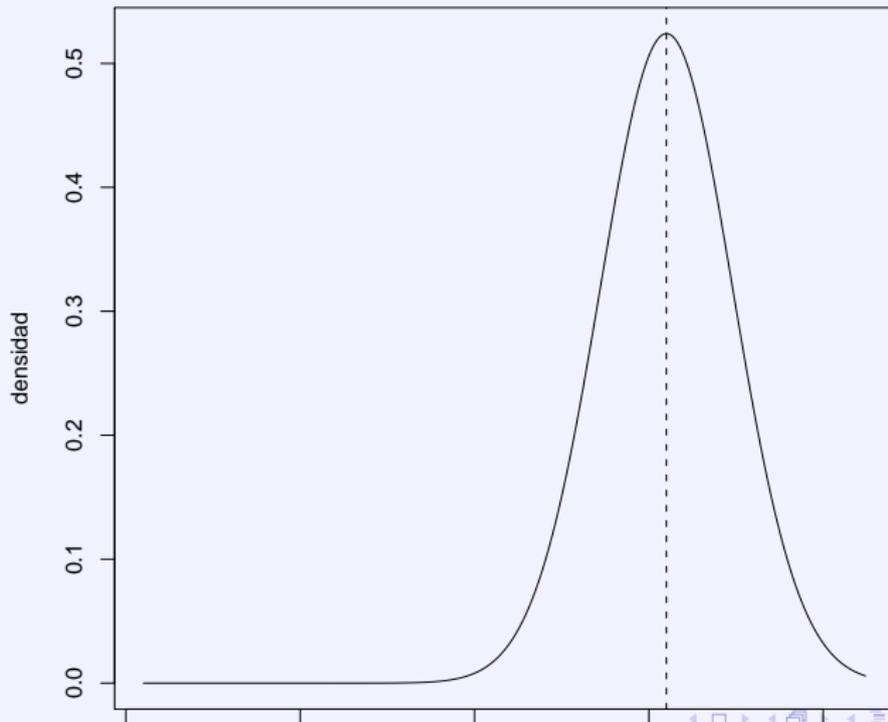
$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{b} K\left(\frac{x - x_i}{b}\right)$$

- $K(\cdot)$ es la función núcleo (habitualmente una función de densidad). Consideraremos la función de densidad de la $N(0, 1)$.
- b es el ancho de banda. En nuestro caso representa la desviación típica de la función núcleo, pues $\frac{1}{b}K\left(\frac{x-x_i}{b}\right)$ es la densidad de la distribución $N(x_i, b^2)$



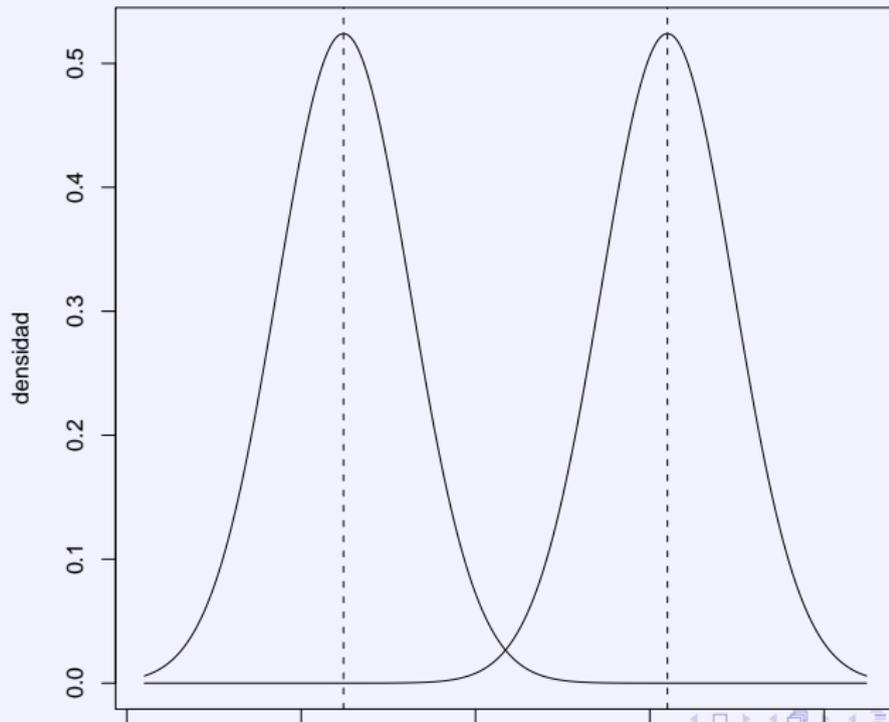
Estimación de la densidad

Estimación de la densidad



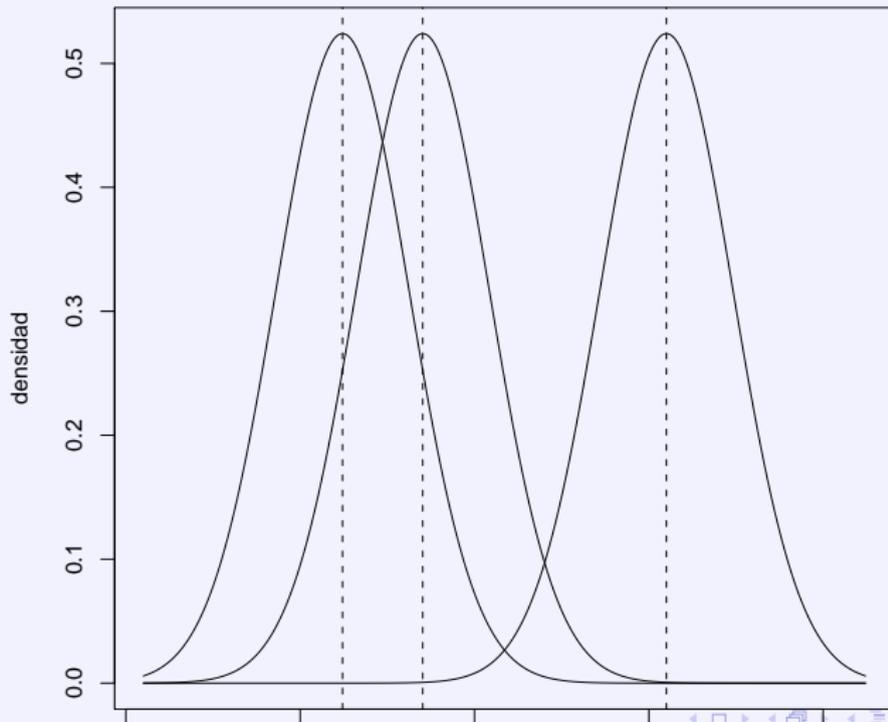
Estimación de la densidad

Estimación de la densidad



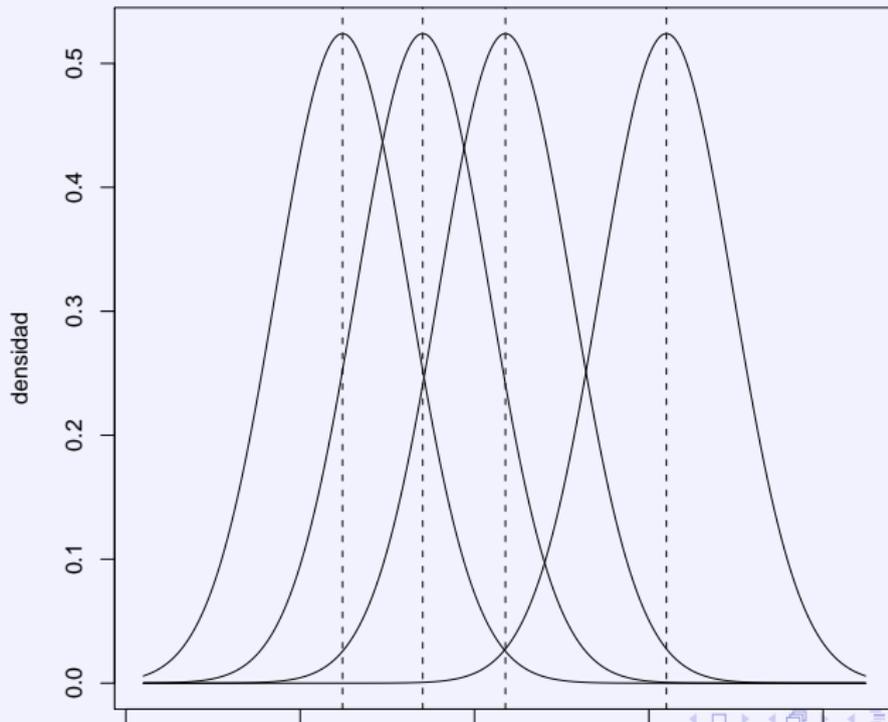
Estimación de la densidad

Estimación de la densidad



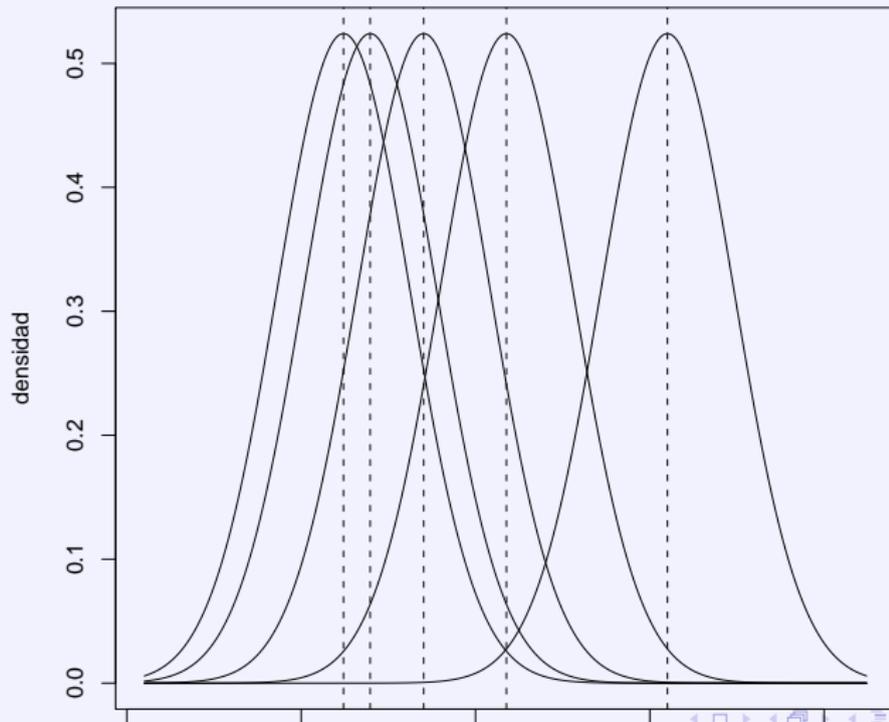
Estimación de la densidad

Estimación de la densidad



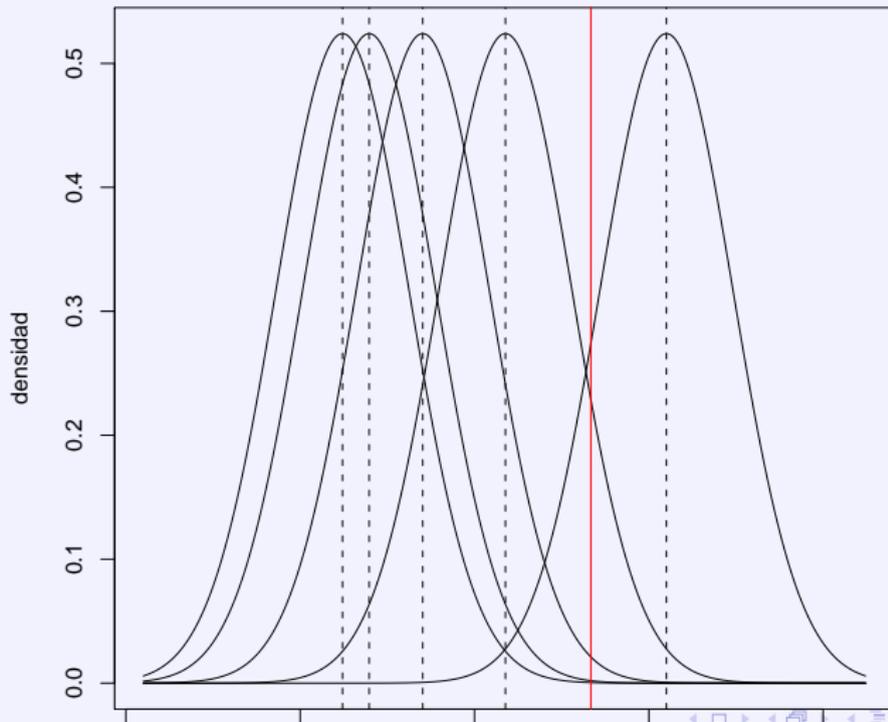
Estimación de la densidad

Estimación de la densidad



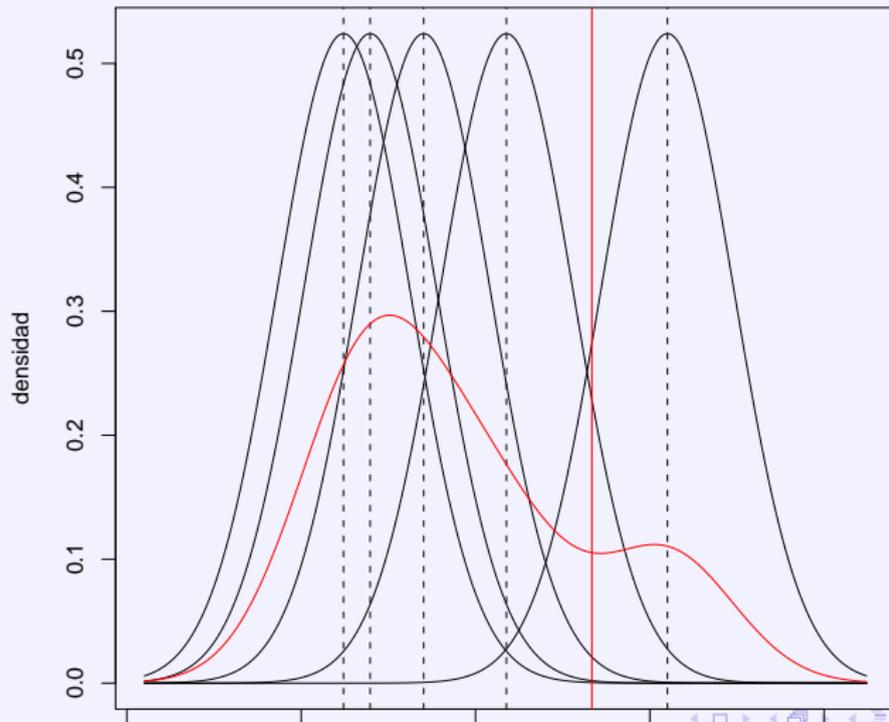
Estimación de la densidad

Estimación de la densidad



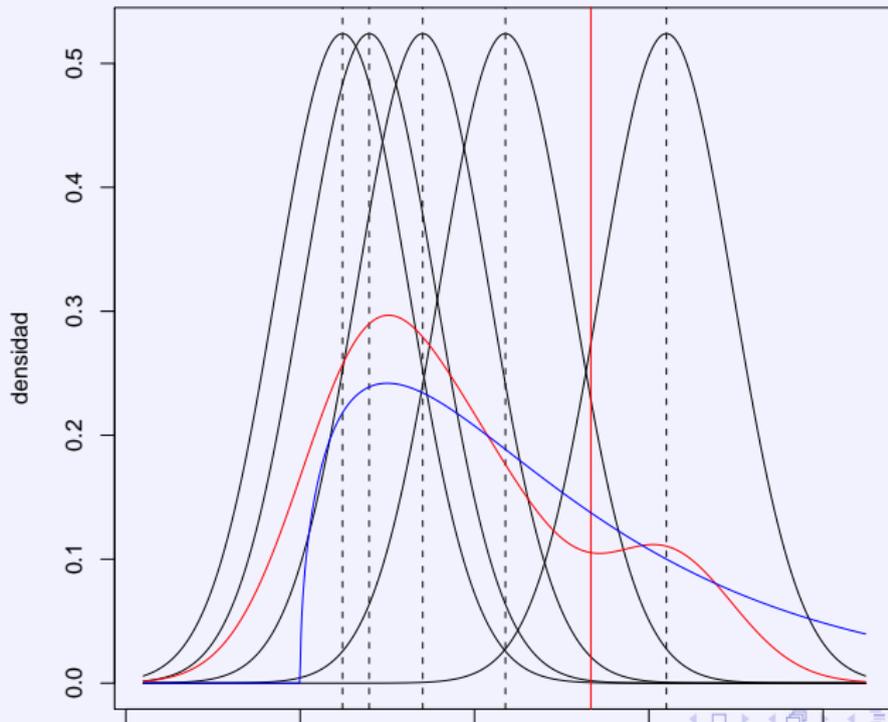
Estimación de la densidad

Estimación de la densidad



Estimación de la densidad

Estimación de la densidad



Función de distribución empírica

Gráficos de Cuantiles o de la función de distribución empírica

- Muestra ordenada de menor a mayor: $x_{(1)}, \dots, x_{(n)}$
- Estimación de $p_i = P(X \leq x_{(i)})$: $\hat{p}_i, i = 1, \dots, n$.

$$\bullet \hat{p}_i = \frac{\#\{j \in \{1, \dots, n\} : x_j \leq x_{(i)}\}}{n}.$$

$$\bullet \hat{p}_i = \frac{i - a}{n - 2a + 1}, a \in [0, 1].$$

Nombre	a	Distribución habitual
Weibull	0	Weibull, Uniforme
Mediana	0.3175	Varias
Bloom	0.375	Normal y otras
Cunnane	0.4	Varias
Gringorten	0.44	Gumbel

- Función de distribución empírica:
 - Distribución discreta: $\hat{F}(x) = \hat{p}_i$ si $x_{(i)} \leq x < x_{(i+1)}, x \in \mathbb{R}$.
 - Distribución continua: $\hat{F}(x) = (1 - r)\hat{p}_i + r\hat{p}_{i+1}$, con $r = \frac{x - x_{(i)}}{x_{(i+1)} - x_{(i)}}, x \in \mathbb{R}$.

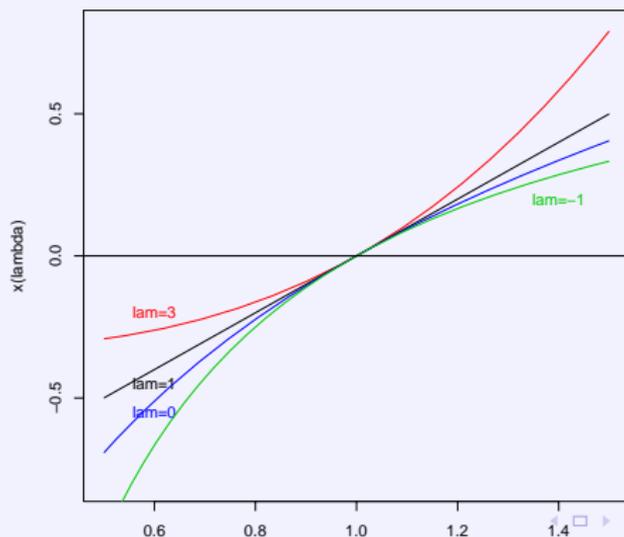


Transformaciones de Box-Cox

Dado $x > 0$, definimos

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \log(x) & \text{si } \lambda = 0 \end{cases}$$

Transformaciones de Box-Cox



Gráficos Q-Q

- **Eje X:** q_1, \dots, q_n , siendo q_i tal que $P(Z \leq q_i) = \hat{p}_i$, $i = 1, \dots, n$, y $Z \sim N(0, 1)$.
- **Eje Y:** $x_{(1)}, \dots, x_{(n)}$ muestra ordenada de menor a mayor de la población X .
- $P(X \leq x_{(i)}) \simeq \hat{p}_i$, $i = 1, \dots, n$.

Si $X \sim N(\mu, \sigma^2)$, entonces

$$\hat{p}_i = P(Z \leq q_i) = P\left(\frac{X - \mu}{\sigma} \leq q_i\right) = P(X \leq \sigma q_i + \mu), \quad i = 1, \dots, n$$

y, por tanto

$$x_{(i)} \simeq \sigma q_i + \mu, \quad i = 1, \dots, n$$

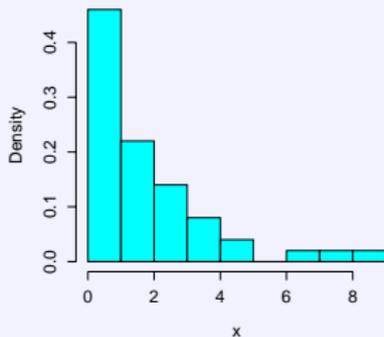


Gráficos Q-Q

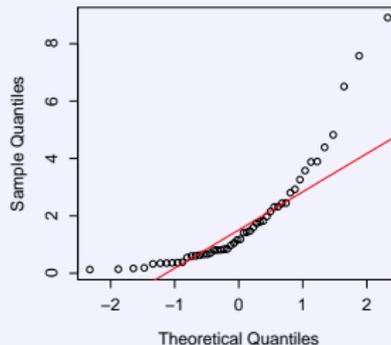
- Si $X \sim N(\mu, \sigma^2)$:
 - Nube de puntos en forma de \cup : distribución asimétrica a la derecha respecto a la campana normal.
 - Nube de puntos en forma de \cap : distribución asimétrica a la izquierda respecto a la campana normal.
 - Nube de puntos en forma de S : distribución más apuntada que la campana normal.
 - Nube de puntos en forma de S invertida: distribución más aplastada que la campana normal.
 - Nube de puntos en dos líneas separadas: mezcla de distribuciones.
 - Nube de puntos con puntos alineados salvo uno aislado: valor extremo.

Gráficos Q-Q

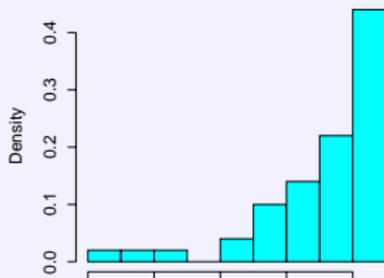
Distribución asimétrica a la derecha respecto a la Normal



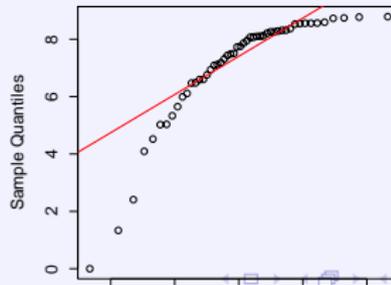
Normal Q-Q Plot



Distribución asimétrica a la izquierda respecto a la Normal

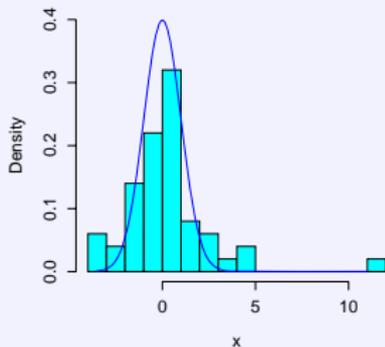


Normal Q-Q Plot

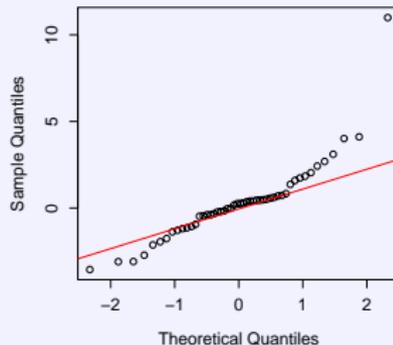


Gráficos Q-Q

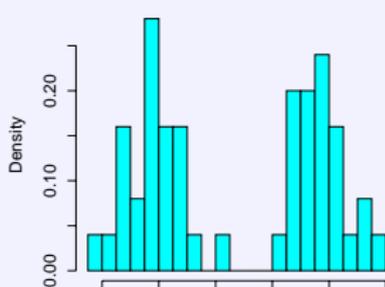
Distribución más aplastada que la Normal



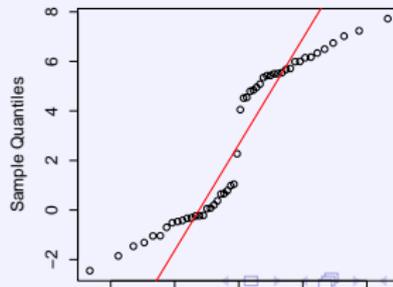
Normal Q-Q Plot



Mezcla de distribuciones



Normal Q-Q Plot



Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

- **TEST PARAMÉTRICO:**

- **VARIANZAS POBLACIONALES IGUALES:** Test de t-Student
- **VARIANZAS POBLACIONALES DIFERENTES:** Test de Welch

Para contrastar si las varianzas poblacionales son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

- **TEST PARAMÉTRICO:**

- VARIANZAS POBLACIONALES IGUALES: Test de t-Student
- VARIANZAS POBLACIONALES DIFERENTES: Test de Welch

Para contrastar si las varianzas poblacionales son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

- **TEST PARAMÉTRICO:**

- **VARIANZAS POBLACIONALES IGUALES:** Test de t-Student
- **VARIANZAS POBLACIONALES DIFERENTES:** Test de Welch

Para contrastar si las varianzas poblacionales son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

- **TEST PARAMÉTRICO:**

- **VARIANZAS POBLACIONALES IGUALES:** Test de t-Student
- **VARIANZAS POBLACIONALES DIFERENTES:** Test de Welch

Para contrastar si las varianzas poblacionales son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

● TEST PARAMÉTRICO:

- **VARIANZAS POBLACIONALES IGUALES:** Test de t-Student
- **VARIANZAS POBLACIONALES DIFERENTES:** Test de Welch

Para contrastar si las varianzas poblacionales son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

- **TEST PARAMÉTRICO:**

- **VARIANZAS POBLACIONALES IGUALES:** Test de t-Student
- **VARIANZAS POBLACIONALES DIFERENTES:** Test de Welch

Para contrastar si las varianzas poblacionales son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

- **TEST PARAMÉTRICO:**

- **VARIANZAS POBLACIONALES IGUALES:** Test de t-Student
- **VARIANZAS POBLACIONALES DIFERENTES:** Test de Welch

Para contrastar si las **varianzas poblacionales** son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

● TEST PARAMÉTRICO:

- **VARIANZAS POBLACIONALES IGUALES:** Test de t-Student
- **VARIANZAS POBLACIONALES DIFERENTES:** Test de Welch

Para contrastar si las **varianzas poblacionales** son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

● TEST PARAMÉTRICO:

- **VARIANZAS POBLACIONALES IGUALES:** Test de t-Student
- **VARIANZAS POBLACIONALES DIFERENTES:** Test de Welch

Para contrastar si las **varianzas poblacionales** son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

- **TEST PARAMÉTRICO:**

- **VARIANZAS POBLACIONALES IGUALES:** Test de t-Student
- **VARIANZAS POBLACIONALES DIFERENTES:** Test de Welch

Para contrastar si las **varianzas poblacionales** son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

- **TEST PARAMÉTRICO:**

- **VARIANZAS POBLACIONALES IGUALES:** Test de t-Student
- **VARIANZAS POBLACIONALES DIFERENTES:** Test de Welch

Para contrastar si las **varianzas poblacionales** son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

- **TEST PARAMÉTRICO:**

- **VARIANZAS POBLACIONALES IGUALES:** Test de t-Student
- **VARIANZAS POBLACIONALES DIFERENTES:** Test de Welch

Para contrastar si las **varianzas poblacionales** son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

- **TEST PARAMÉTRICO:**

- **VARIANZAS POBLACIONALES IGUALES:** Test de t-Student
- **VARIANZAS POBLACIONALES DIFERENTES:** Test de Welch

Para contrastar si las **varianzas poblacionales** son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

Comparación de dos Poblaciones: Medias o Centralización

MUESTRAS INDEPENDIENTES:

- **TEST PARAMÉTRICO:**

- **VARIANZAS POBLACIONALES IGUALES:** Test de t-Student
- **VARIANZAS POBLACIONALES DIFERENTES:** Test de Welch

Para contrastar si las **varianzas poblacionales** son iguales utilizamos el test de F-Snedecor.

- **TEST NO PARAMÉTRICO:** Test de Mann-Whitney-Wilcoxon de suma de rangos.

MUESTRAS APAREADAS O RELACIONADAS:

- **TEST PARAMÉTRICO:** Test de t-Student
- **TEST NO PARAMÉTRICO:** Test de Wilcoxon de rangos con signo.

