UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
ENES MÉRIDA
LICENCIATURA EN ECOLOGÍA

ESTADÍSTICA APLICADA
Tema III. Prueba de Hipótesis
(Parte 3 – Condiciones Generales)

Prof. Edlin J. Guerra Castro

Condiciones para aplicar las pruebas con buen dominio del error estadístico

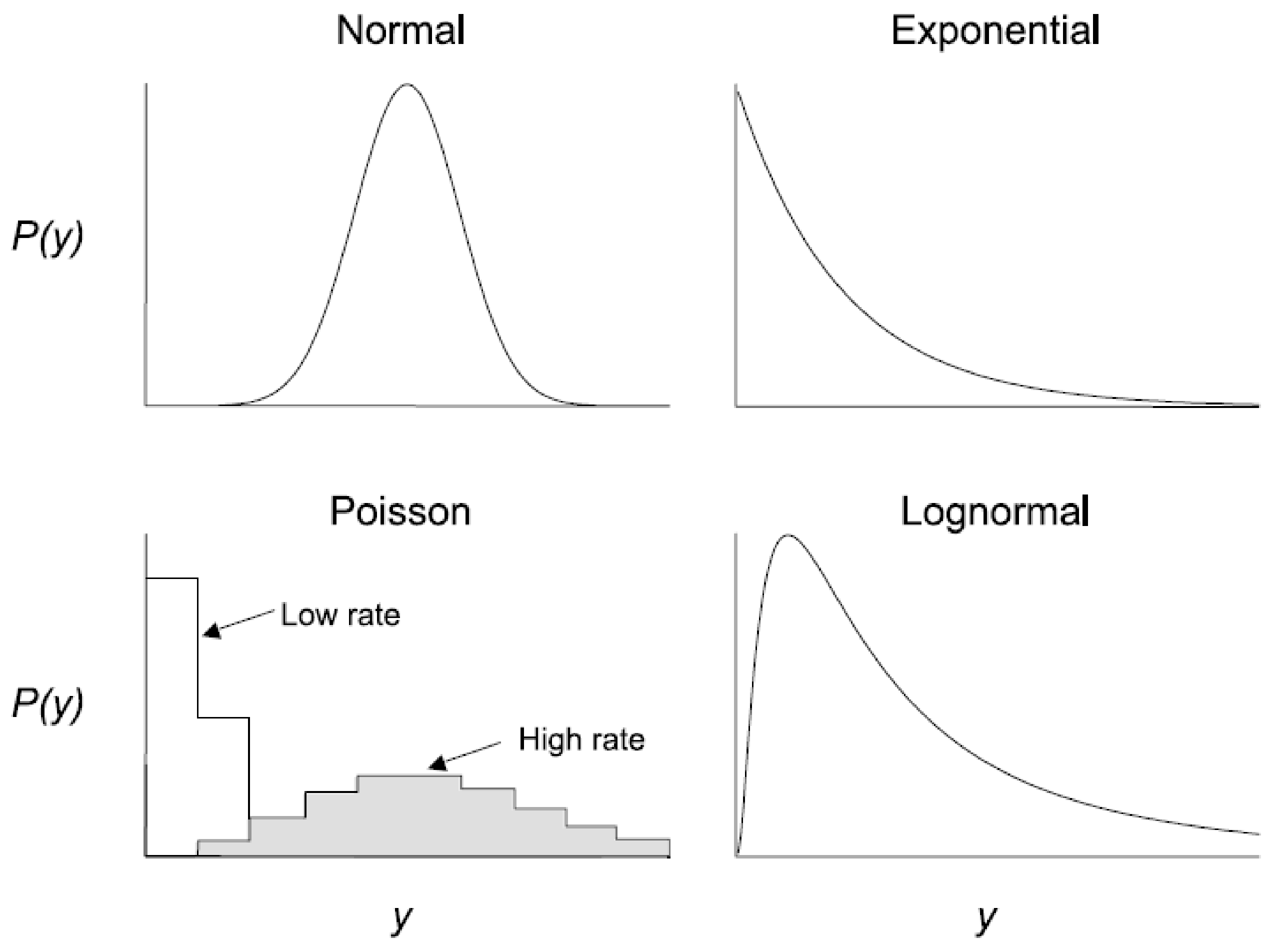
En la mayoría de los textos indican:

1. Los valores deben ser independientes e idénticamente distribuidos. Se resuelve con muestreos aleatorios.

2. Los residuales se deben ajustar a una distribución normal. Se debe evaluar, no necesariamente se debe esperar ajuste a una dist. normal.

3. Las varianzas de los grupos deben ser homogéneas. Se debe evaluar, no necesariamente se debe esperar varianzas iguales.

Distribuciones teóricas de probabilidad

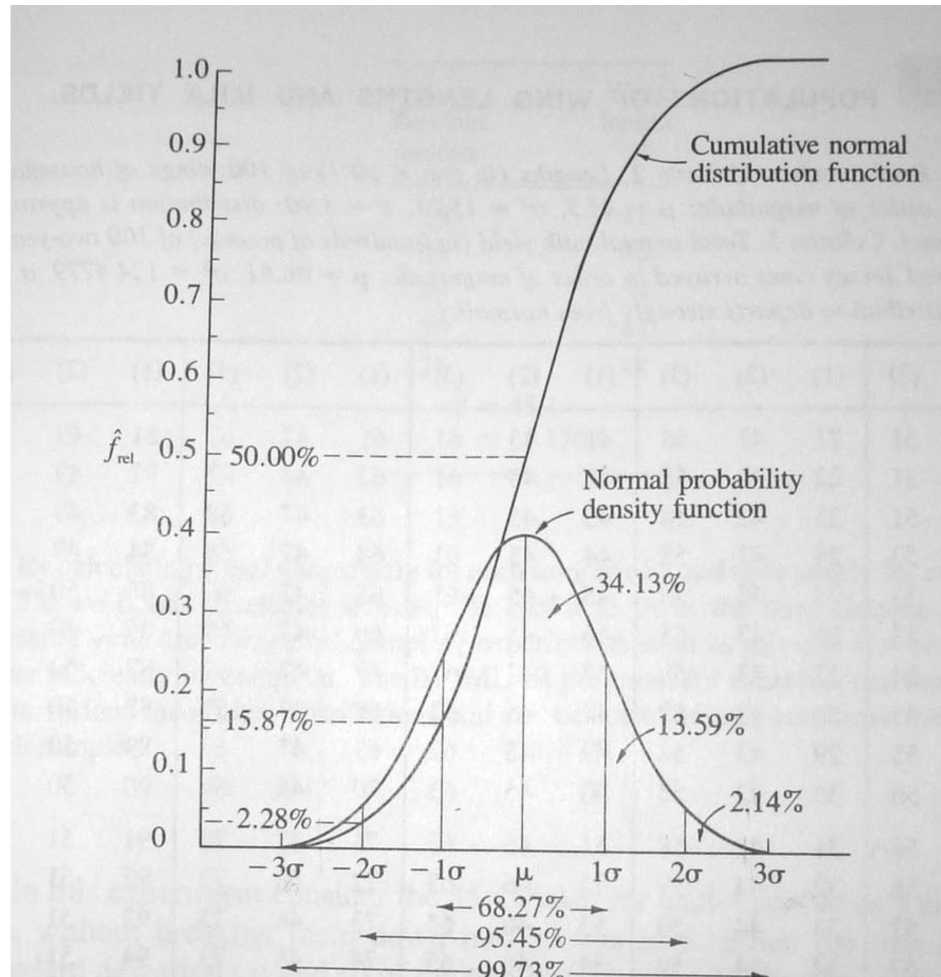


DISTRIBUCIÓN NORMAL

- Propiedades de la Distribución normal

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2}$$

$$x \approx N(\mu, \sigma^2)$$



DISTRIBUCIÓN NORMAL

- Propiedades de la distribución normal

La distribución es simétrica alrededor de la media

$\mu \pm \sigma$ contiene 68.27 % de los elementos
$\mu \pm 2\sigma$ contiene 95.45 % de los elementos
$\mu \pm 3\sigma$ contiene 99.73 % de los elementos

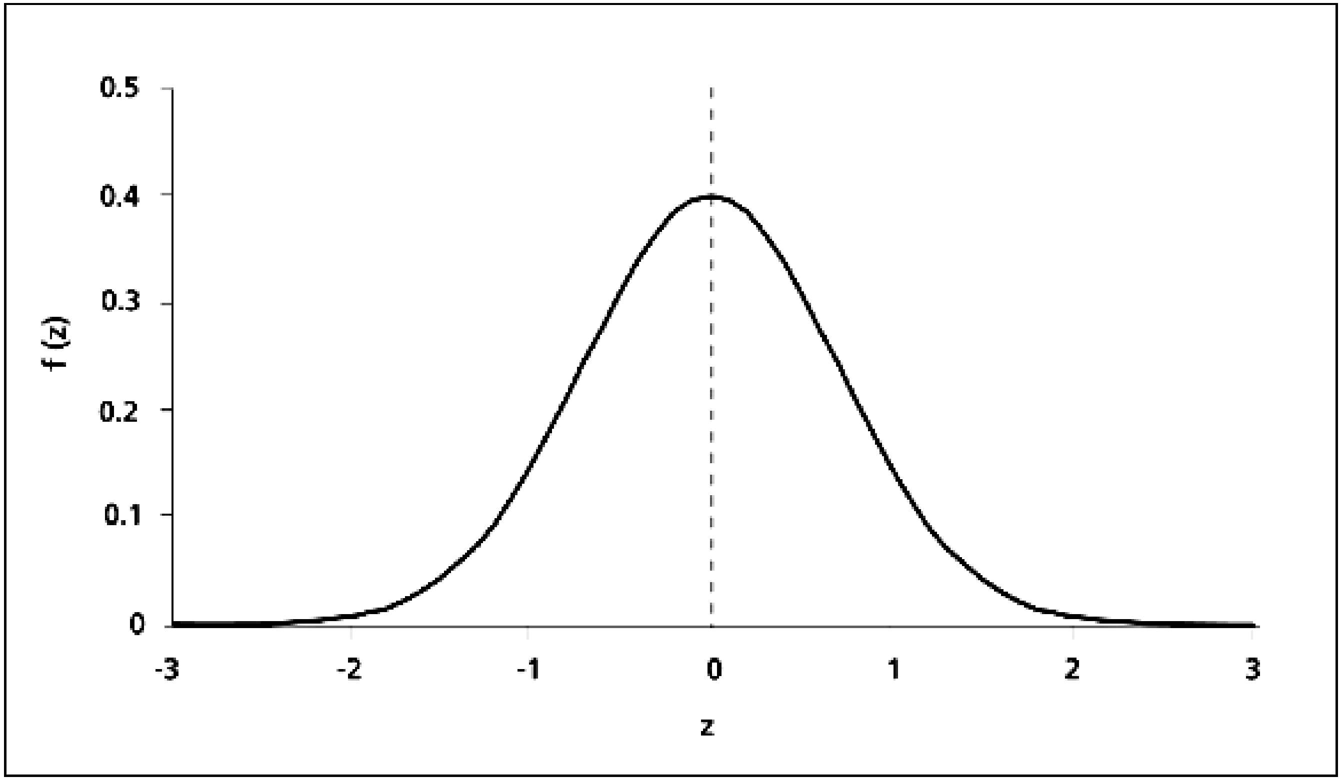
DISTRIBUCIÓN NORMAL

- Distribución normal estandarizada (Distribución Z)

$$x \approx N(\mu, \sigma^2)$$

$$x \approx N(0,1)$$

$$z_i = \frac{x_i - \mu}{\sigma}$$



DISTRIBUCIÓN NORMAL

- Métodos de detección: Gráficos

1) Gráfico de probabilidad
2) Gráfico de residuales estandarizados

- Métodos de detección: Pruebas

1) Pruebas de Bondad de ajuste (Ji-cuadrada, Kolmogorov-Smirnov)
2) Prueba de Shapiro-Wilk (1965) (n<50)
3) D'Angostino y Pearson (1973) (cualquier n)

DISTRIBUCIÓN NORMAL

Consideremos una variable Y_i donde i = 1, 2, 3,…n muestras y queremos evaluar si se distribuye normalmente

- Métodos de detección: Gráficos

1) Gráfico de Probabilidad Normal (q-q plot)

1.-Los datos se ordenan de menor a mayor: $Y_1 \leq Y_2 \leq Y_3 \leq \ldots \leq Y_n$

2.-Este ORDEN DE POSICIÓN se identifica con i = 1, 2, 3,…, n

3.-Para cada i es posible ESTIMAR cuál sería su probabilidad p(i) si perteneciera a una distribución normal con la fórmula propuesta por Bloom (1958)

4-Se grafica Y_i vs Z_i en el orden creciente

$$p(i) = \frac{i - 0.375}{n + 0.25}$$

Tabla de Bloom (1958)

DISTRIBUCIÓN NORMAL

Consideremos una variable Y_i donde i = 1, 2, 3,…n muestras y queremos evaluar si se distribuye normalmente

- Métodos de detección: Gráficos

1) Gráfico de Probabilidad Normal de Bloom (1958): :

Valor tomado	Valor Y ordenado	i	bloom	zi
297	227	1	0.041	-1.7392
340	250	2	0.107	-1.2426
325	277	3	0.172	-0.9463
227	289	4	0.238	-0.7128
277	290	5	0.303	-0.5158
337	291	6	0.369	-0.3345
250	293	7	0.434	-0.1662
290	297	8	0.500	0
293	318	9	0.566	0.1662
291	325	10	0.631	0.3345
289	337	11	0.697	0.5158
430	340	12	0.762	0.7128
510	353	13	0.828	0.9463
353	430	14	0.893	1.2426
318	510	15	0.959	1.7392

n = 15

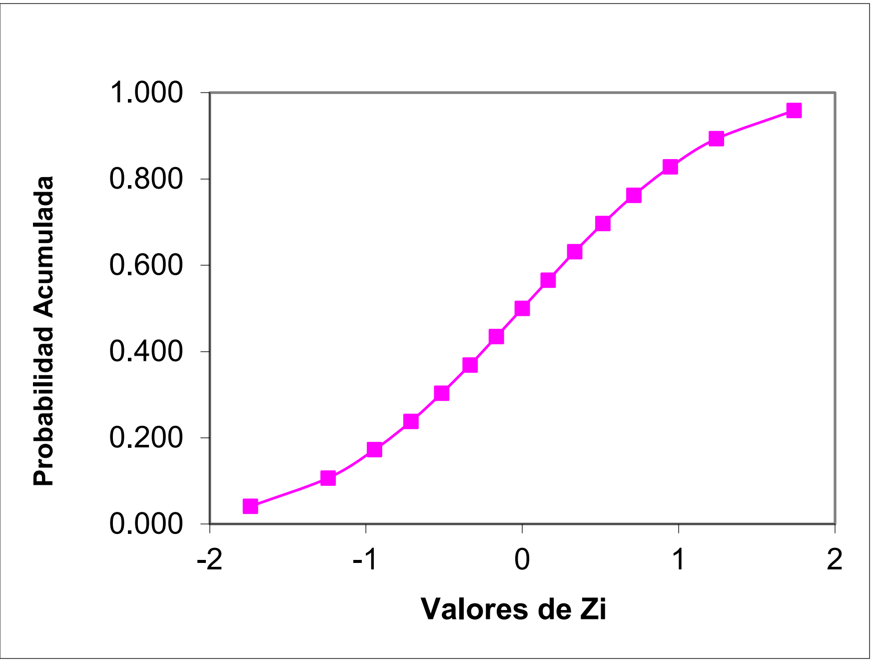
$$p(i) = \frac{i - 0.375}{n + 0.25}$$

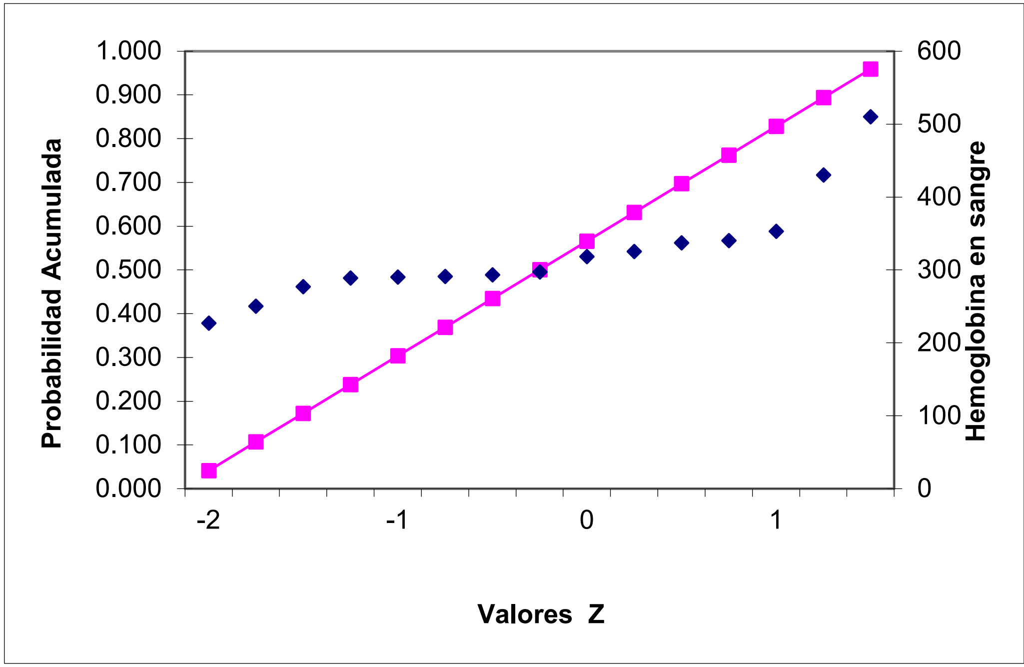
Tabla de Blom (1968)

DISTRIBUCIÓN NORMAL

Consideremos una variable Y_i donde i = 1, 2, 3,…n muestras y queremos evaluar si se distribuye normalmente

- Métodos de detección: Gráficos





DISTRIBUCIÓN NORMAL

- Métodos de detección: Gráficos

2) Gráfico de Residuales estandarizados:

1.- Se calculan los residuales de la forma

$$e_{ij} = Y_{ij} - \overline{Y} \longrightarrow es_{ij} = \frac{Y_{ij} - \overline{Y}}{s}$$

2.-Los datos se ordenan de menor a mayor: $Y_1 \leq Y_2 \leq Y_3 \leq \ldots \leq Y_n$ y se identifican con i = 1, 2, 3,…, n

3.-Para cada i es posible ESTIMAR cuál sería su probabilidad p(i) si perteneciera a una distribución normal con la fórmula propuesta por Blom (1958)

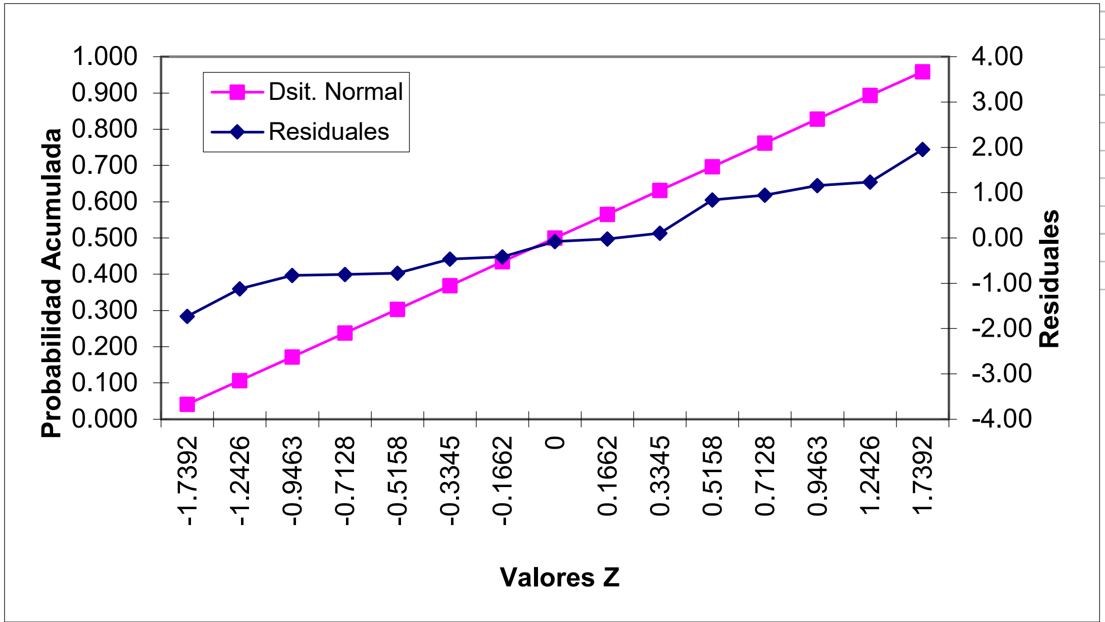
DISTRIBUCIÓN NORMAL

Consideremos una variable Y_i donde i = 1, 2, 3,…n muestras y queremos evaluar si se distribuye normalmente

- Métodos de detección: Gráficos

1) Gráfico de Residuales estandarizados:

valor	error STAN		e ordenado	i	Blom	z
297	0.10812015		-1.73	1	0.041	-1.7392
340	1.23519082		-1.12	2	0.107	-1.2426
325	0.84202663		-0.83	3	0.172	-0.9463
227	-1.72664605		-0.80	4	0.238	-0.7128
277	-0.41609876		-0.78	5	0.303	-0.5158
337	1.15655798		-0.46	6	0.369	-0.3345
250	-1.1237943		-0.42	7	0.434	-0.1662
290	-0.07535647		-0.08	8	0.500	0
293	-0.77860434		-0.02	9	0.566	0.1662
291	-0.80377867		0.11	10	0.631	0.3345
289	-0.828953		0.84	11	0.697	0.5158
430	0.94583742		0.95	12	0.762	0.7128
510	1.95281071		1.16	13	0.828	0.9463
353	-0.02337437		1.24	14	0.893	1.2426
318	-0.46392518		1.95	15	0.959	1.7392



DISTRIBUCIÓN NORMAL

Test de Shapiro-Wilk (1965)

Algoritmo:

1.-Se ordenan los datos de menor a mayor

2.-Se calcula la Sumatoria Cuadrática de los datos

$$SC = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

3.-Se calcula b, definido como…

Si n es par m = n/2

Si n es impar m = n-1/2

$$b = \sum_{i=1}^{m} a_i (Y_{n-i+1} - Y_i)$$

a_i : estadístico de orden de una dist. Normal

4.-Se calcula el estadístico

$$W = \frac{b^2}{SC}$$

5.-Si W > Wc la distribución es Normal, si es < no se aproxima a una Normal

Statistics and Computing (1992) **2**, 117–119

Approximating the Shapiro–Wilk W-test for non-normality

PATRICK ROYSTON

Department of Medical Physics, Royal Postgraduate Medical School, Ducane Road, London W12 0NN

Received October 1991 and accepted October 1991

A new approximation for the coefficients required to calculate the Shapiro–Wilk W-test is derived. It is easy to calculate and applies for any sample size greater than 3. A normalizing transformation for the W statistic is given, enabling its P-value to be computed simply. The distribution of the new approximation to W agrees well with published critical points which use exact coefficients.

Keywords: Non-normality, Shapiro–Wilk W-test

Table 2. *Some critical points for W*

n	Royston (1982a) 0.05	0.01
20	0.904	0.868
30	0.928	0.902
40	0.941	0.922
60	0.9538	0.9411
80	0.9602	0.9505
100	0.9642	0.9561
150	0.9698	0.9640
250	0.9748	0.9710
500	0.97945	0.97707
1000	0.98305	0.98159
2000	0.98602	0.98514
4000	–	–
5000	–	–

*Asymptotic points

DISTRIBUCIÓN NORMAL

- Simetría y Kurtosis

Propiedades de una distribución de frecuencia que nos permite medir la simetría y altura de la distribución. Se pueden someter a prueba hipótesis en función de una distribución esperada.

1) Prueba de D'Agostino y Pearson (1973): Se basa en la simetría y en la curtosis

La H_0 se probada con el estadístico K^2

$$k^2 > \chi^2_{v,\alpha}$$

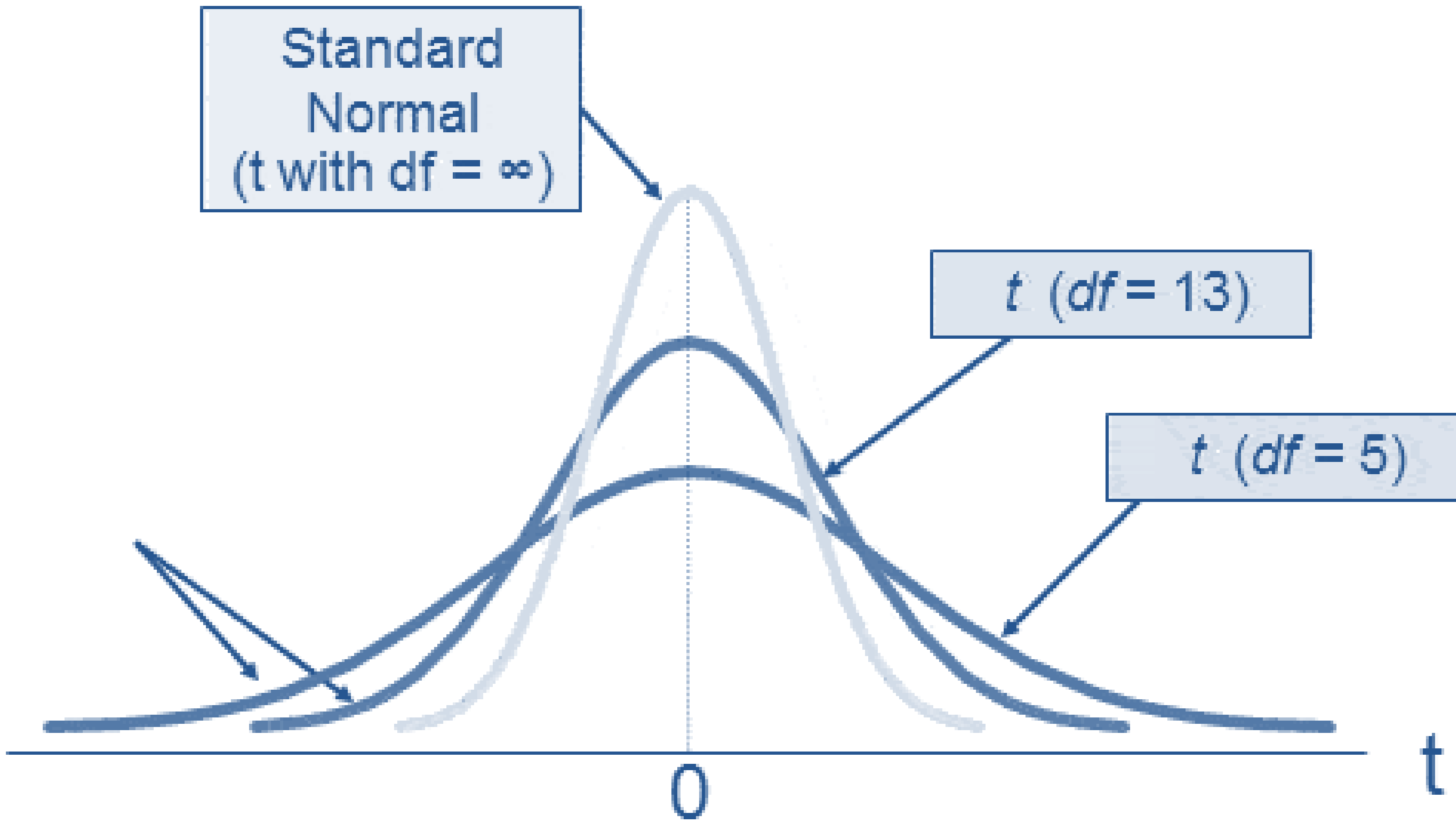
$$k^2 = Z^2_{g1} + Z^2_{g2}$$

Ver guía para procedimiento…

DISTRIBUCIÓN NORMAL

- Otras pruebas:
  - Prueba de χ^2
  - Prueba de Filliben
  - Prueba de Galton
  - Kolmogorov-Smirnov

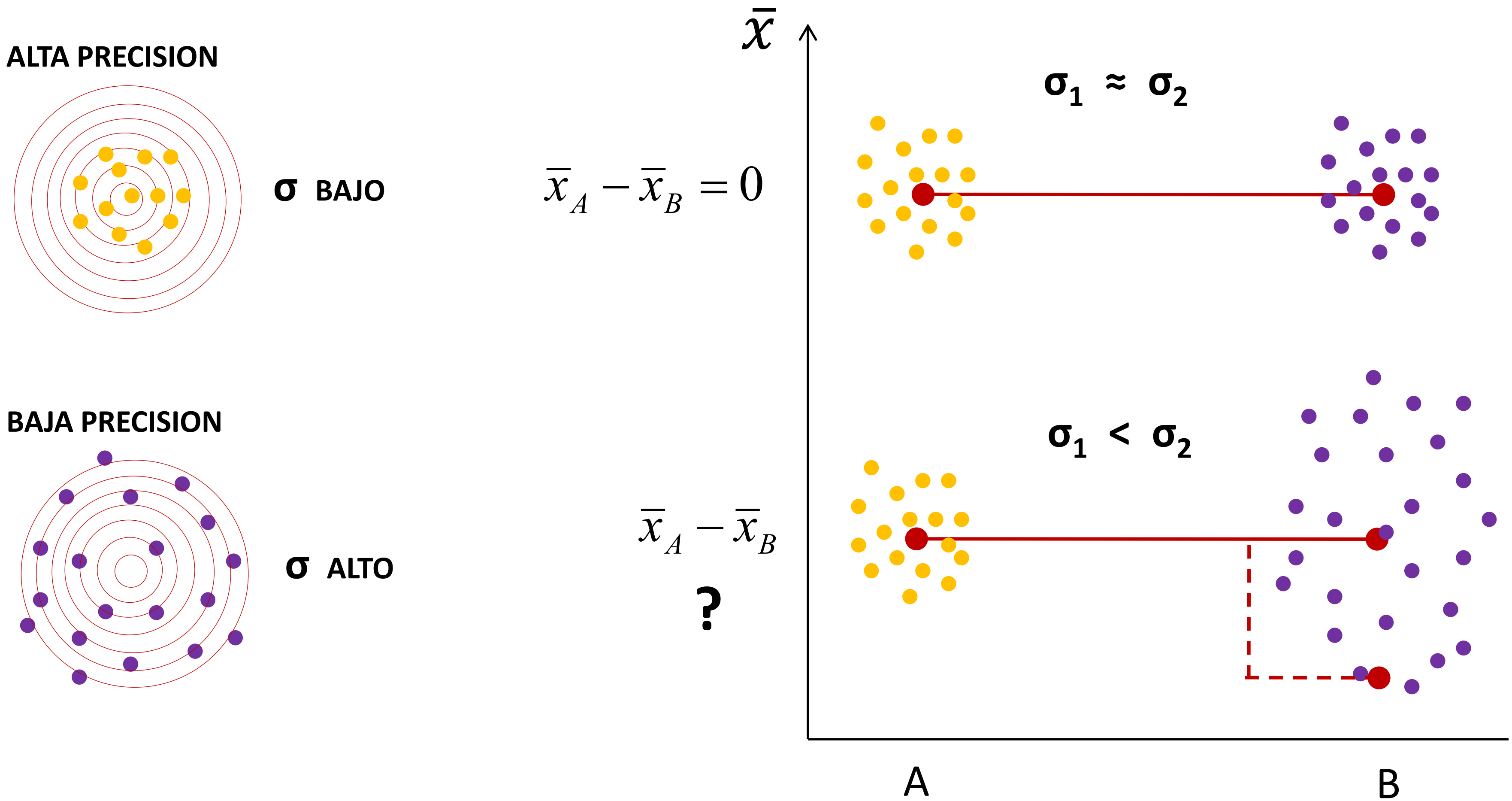
Ajuste a la distribución Normal y prueba *t*



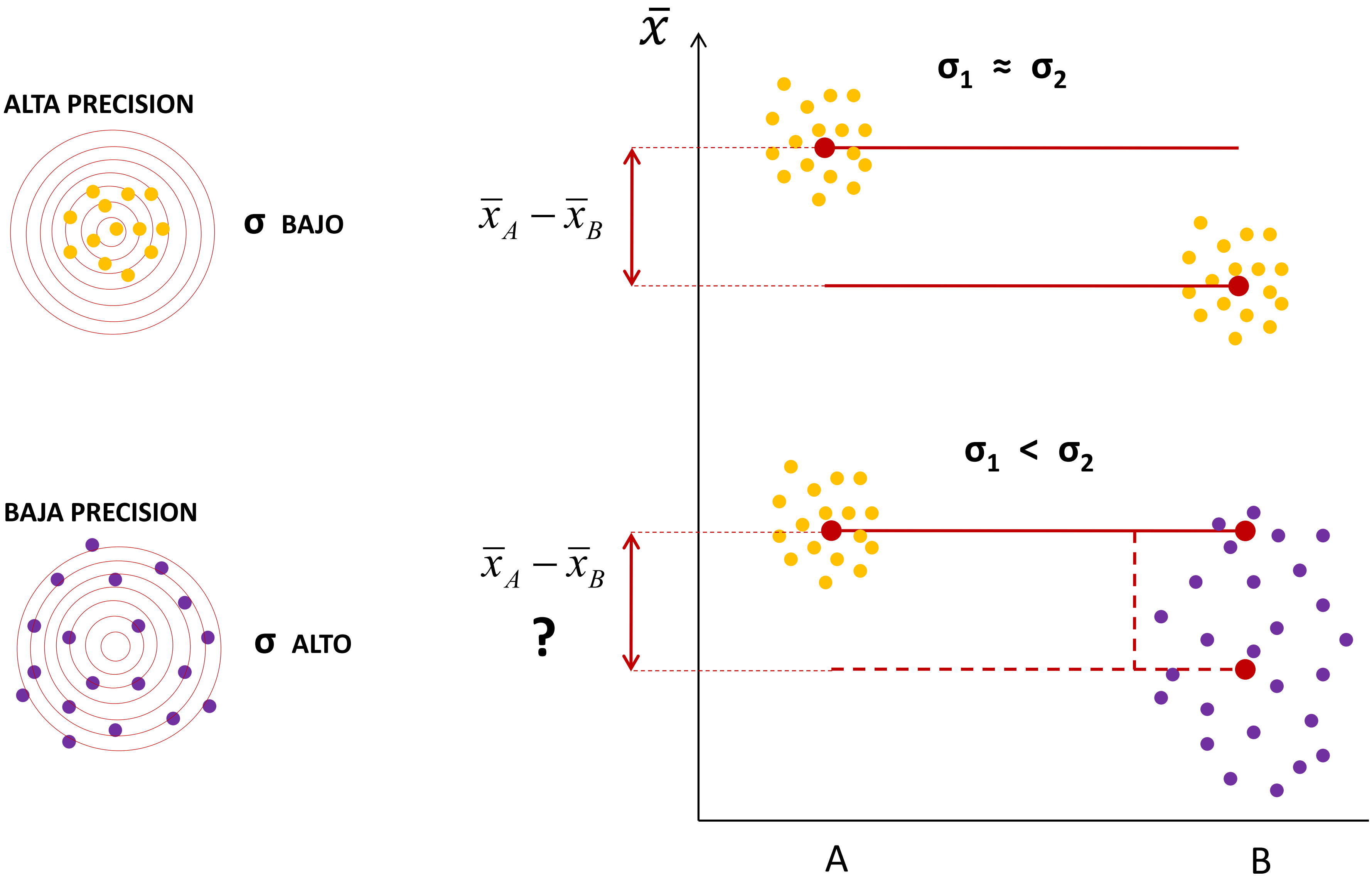
$$t = \frac{\overline{X}_A - \overline{X}_B}{S_p}$$

El problema de la heterogeneidad de varianzas

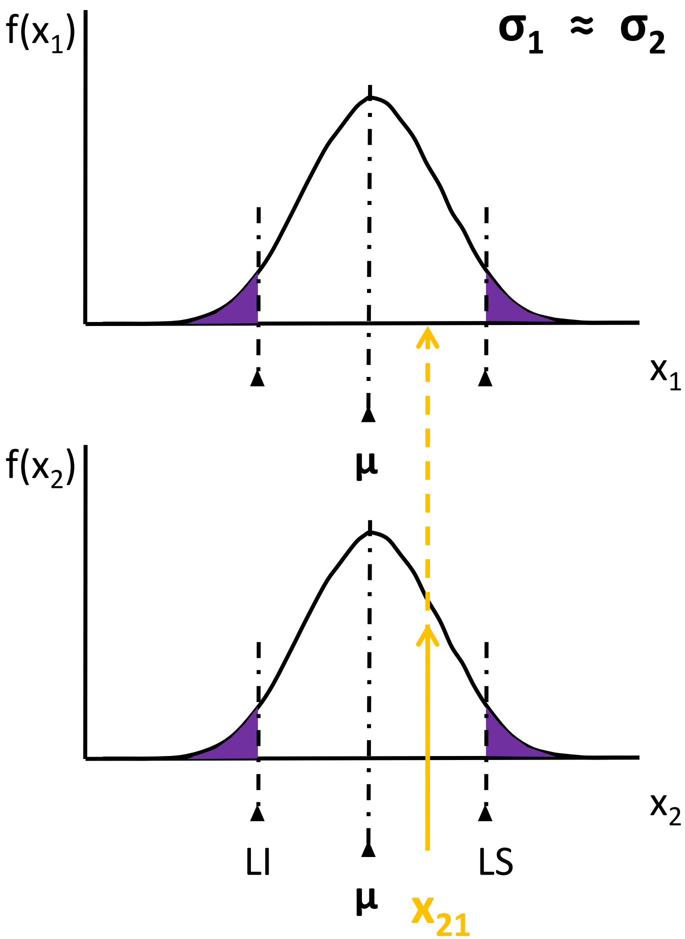
Reside en la dificultad para estimar la diferencia entre las medias poblacionales a partir de las medias de las muestras, como resultado de una diferencia en la precisión para estimar dichas medias.



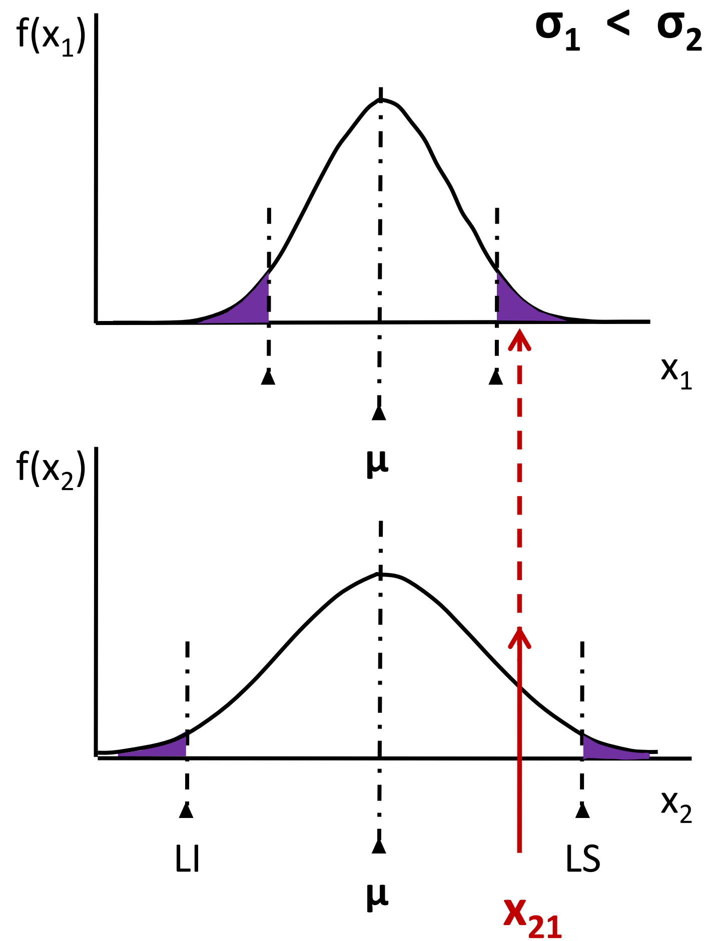
El problema de la heterogeneidad de varianzas



La H_o se cumple

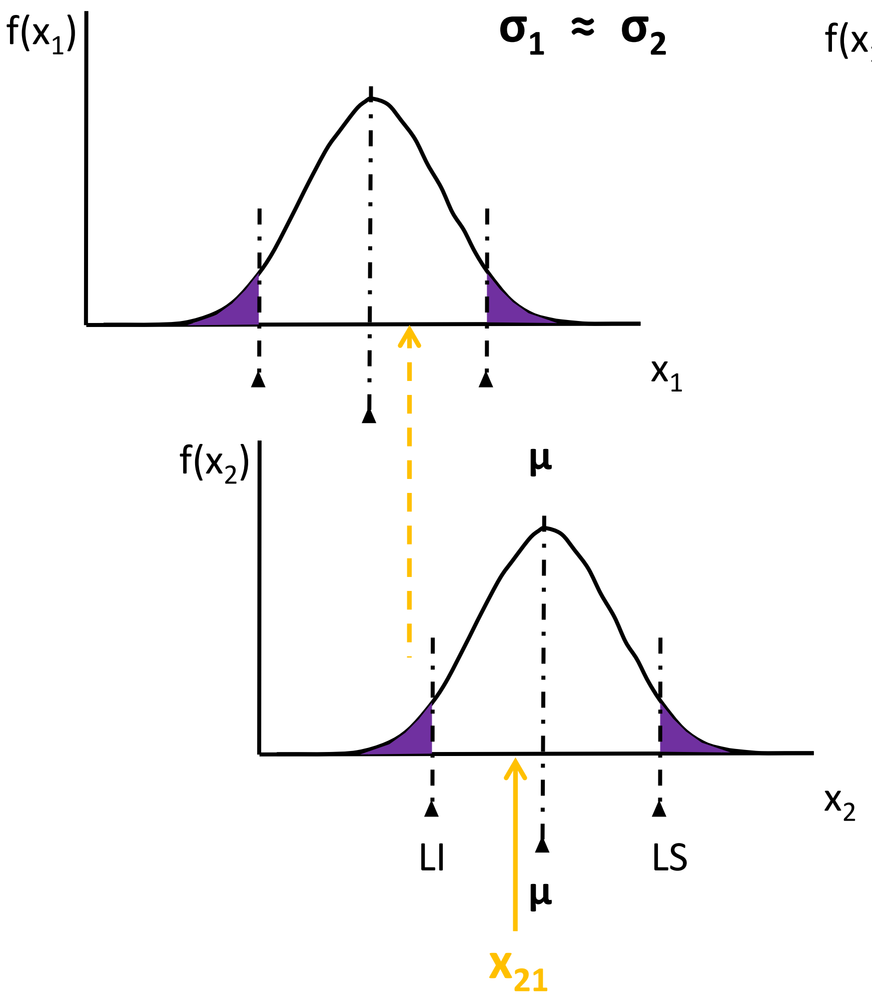


La H_o se cumple

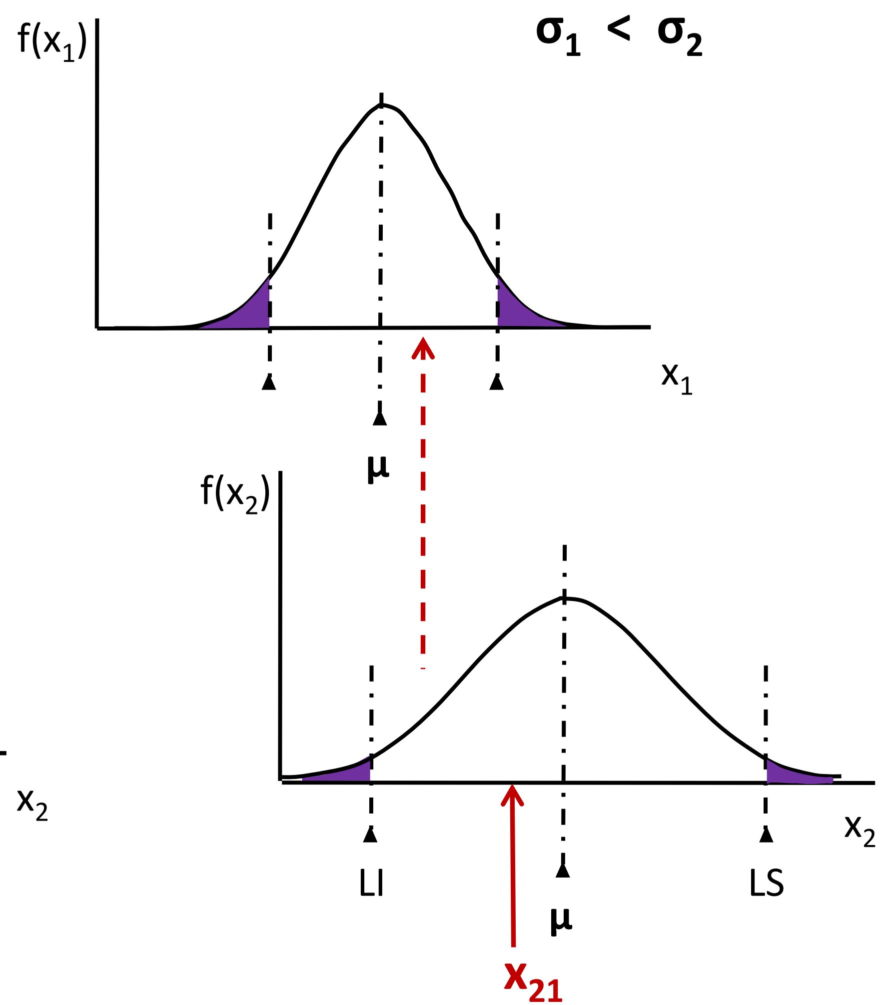


Rechazar H_o equivocadamente

La H_o NO se cumple



La H_o NO se cumple



Retener H_o equivocadamente

Para cerrar…

- Se requiere de una distribución probabilística como referencia del azar. Típicamente se usa la Dist. Normal, pero hay otras alternativas

- La variación puede influir en la inferencia estadística. Por ello es requerido entender bien las pruebas y posibles consecuencias de las diferencias en variación.