ERRORES EN QUÍMICA ANALÍTICA

- **Introducción**
- **Estadística descriptiva**
  Medidas de centralización, dispersión y forma de la distribución
- **Medida de la calidad**
  Calidad y errores
  Errores sistemáticos y aleatorios
- **Precisión y bias de las medidas**
  Precisión, exactitud y veracidad
  Detección y determinación de bias
- **Otros tipos de error**
- **Distribución normal o gaussiana**
  Propiedades de la distribución normal
  Distribución normal tipificada o estandarizada
- **Teorema del límite central y distribución de medias muestrales**
  Intervalos de confianza de la media
  Estadística de muestras pequeñas
- **Otras distribuciones**

INTRODUCCIÓN

Poblaciones y muestras

POBLACIÓN todas las posibles determinaciones que puedan llevarse a cabo

MUESTRA determinaciones que realmente se llevan a cabo

POBLACIÓN: Número infinito de individuos, objetos o determinaciones, de la cual se toma una **MUESTRA** finita (y de tamaño mucho más reducido)

A partir de la **MUESTRA** se extraen conclusiones acerca de toda la **POBLACIÓN**

Hay un problema de **terminología**

- (IUPAC) En química ***muestra*** es una porción de un material seleccionado a partir de una cantidad más grande de material.
- Si los errores de muestreo son despreciables la IUPAC sugiere el uso de ***porción de prueba*** *(test portion), alícuota o espécimen*.

VARIABLES

Propiedades respecto de las que los elementos individuales de una muestra difieren de alguna forma mensurable

ESCALAS DE MEDIDA

Nominal — **Palabras (Cualitativa: Bueno, Aceptable, Malo...)**
Ordinal (ranked) — **Orden: de 1 a 10**
Escalas cuantitativas — **Numéricas**
Intervalo (el cero no tiene significado) — **Escala centígrada de temperatura**
de ratio (el cero tiene significado) — **Escala Kelvin de temperatura**, **cualquier variable numérica**
Continuas o discretas

INTRODUCCIÓN

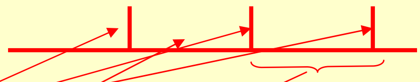
Histogramas y distribuciones

Descripción gráfica de gran número de datos.

Se ordenan (rango: intervalo entre el mayor y el menor) y se dividen en clases

$$\text{nº}_{\text{ clases}} = \sqrt{\text{nº de datos}}$$

Según la **ISO**:
Clase es cada uno de los intervalos consecutivos en los cuales se divide el intervalo total de variación.



Límites de clase son los valores que definen los límites inferior y superior de una clase.
El **punto medio** (mid-point) es la media aritmética de los límites y se denomina a veces **marca** (class mark) de la clase (la ISO no lo recomienda)
El intervalo de clase es la diferencia entre los mismos.

FRECUENCIA RELATIVA de cada clase: nº de individuos en cada clase dividido por el nº total de individuos. De ella se obtiene la **DISTRIBUCIÓN DE FRECUENCIAS RELATIVAS (Histograma habitual)**

FRECUENCIAS ACUMULADAS O ACUMULATIVAS: Suma de todas las frecuencias hasta una determinada clase, se obtiene la **DISTRIBUCIÓN DE FRECUENCIAS (RELATIVAS) ACUMULADAS O ACUMULATIVAS.**

Añadir EXCEL

HISTOGRAMA

DISTRIBUCIÓN DE FRECUENCIAS RELATIVAS

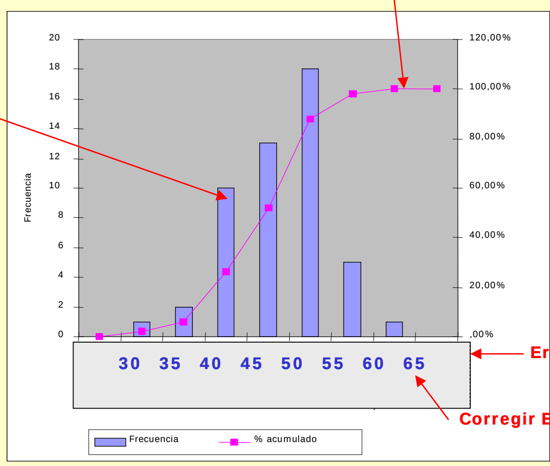
EXCEL

	Clase	Frecuencia absoluta	% Frecuencia relativa	% Frecuencia acumulada
	30	0	0,00	0,00
30	35	1	2,00	2,00
35	40	2	4,00	6,00
40	45	10	20,00	26,00
45	50	13	26,00	52,00
50	55	18	36,00	88,00
55	60	5	10,00	98,00
60	65	1	2,00	100,00
65	y mayor...	0	0,00	100,00

DISTRIBUCIÓN DE FRECUENCIAS RELATIVAS ACUMULADAS

DISTRIBUCIÓN DE FRECUENCIAS ABSOLUTAS

- Si la variable es continua se obtienen distribuciones continuas
- Si los puntos son representativos de la población, estas distribuciones son distribuciones de probabilidad o funciones de densidad de probabilidad
- La distribución de frecuencias describe la muestra.
- La distribución de probabilidad describe la población



Erróneo

Corregir Excel

ESTADÍSTICA DESCRIPTIVA

Promedio y medidas de centralización

EXCEL

Media (muestral) $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ **promedio(datos)**

Media poblacional $\mu = \frac{\sum_{i=1}^{n} x_i}{N} = E(x)$

Media de datos agrupados $\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_i x_i}{n}$

Media ponderada $\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$

	Serie 1	Serie 2
	29.8	29.8
	30.5	30.5
	30.8	30.8
	31.6	31.6
	31.9	31.9
	32.1	32.1
	32.6	32.6
	33.1	33.1
	33.6	33.6
	35.0	**40.8**
Media	**32.1**	**32.7**
Mediana	32.0	32.0

Mediana: Valor central de una serie de datos
n impar: dato que aparece en el lugar (n+1)/2 **mediana(datos)**
n par: media de los datos de los lugares n/2 y (n+2)/2
La mediana es más **robusta** o insensible que la media a las variaciones

Moda: Valor que se presenta con mayor frecuencia **moda(datos)**

ESTADÍSTICA DESCRIPTIVA

Promedio y medidas de centralización

EXCEL

Cuartiles:
Al ordenar los datos de menor a mayor, la mediana divide a un conjunto en dos partes iguales. Cada una de esas dos partes puede dividirse en otras dos y se generan cuatro partes o cuartiles divididos por los datos $\mathbf{Q_1}$, $\mathbf{Q_2}$ **y** $\mathbf{Q_3}$. La mediana coincide con Q_2

cuartil(datos;cuartil)
nº : 0 mínimo
1 primer cuartil
2 mediana
3 tercer cuartil
4 máximo

Deciles:
Valores de los datos D_1 a D_9 que dividen a toda la muestra en 10 partes iguales

Percentiles:
Valores de los datos P_1 a P_{99} que dividen a toda la muestra en 100 partes iguales

percentil(datos;k)
k : percentil (entre 0 y 1)

ESTADÍSTICA DESCRIPTIVA

Medidas de dispersión

EXCEL

Rango $R = x_{max} - x_{min}$

Varianza muestral $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$
var(datos)

Desviación típica muestral $s = \sqrt{s^2}$
desvest(datos)

Varianza poblacional $\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{N}$
varp(datos)

Desviación típica poblacional $\sigma = \sqrt{\sigma^2}$
desvestp(datos)

Desviación típica relativa $s_r = s / \bar{x}$; $s_r(\%) = s / \bar{x} * 100$

A veces se llama coeficiente de variación pero la IUPAC lo desaconseja

Recorrido intercuartil Q3 - Q1

Cuanto menor es, más agrupados están los datos

Desviación típica promediada (pooled)

Cuando se obtienen conjuntos de muestras en **diferentes momentos** o en **diferentes** (pero similares) **muestras** de **una misma población**, y se desea obtener la varianza (desviación típica) de los datos agrupados, se emplea la varianza (desviación típica) promediada (pooled), de acuerdo a:

$$s^2_{pooled} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{(n_1 - 1) + (n_2 - 1) + \ldots + (n_k - 1)} = \frac{\sum (g.d.l.)_j s_j^2}{\sum (g.d.l.)_j}$$

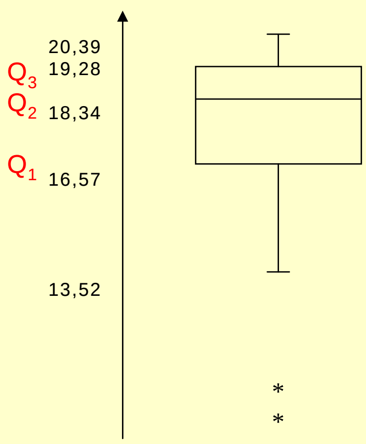
(j=1...k)

ESTADÍSTICA DESCRIPTIVA

Medidas de dispersión

Box (and whiskers) plot: Gráfico de caja y bigotes

Resultado	Orden
9,00	1
10,06	2
13,52	3
14,06	4
15,65	5
16,22	6
16,57	7
16,60	8
16,83	9
16,85	10
17,78	11
18,13	12
18,34	13
19,05	14
19,07	15
19,10	16
19,16	17
19,28	18
19,28	19
19,47	20
19,55	21
19,80	22
20,24	23
20,25	24
20,39	25



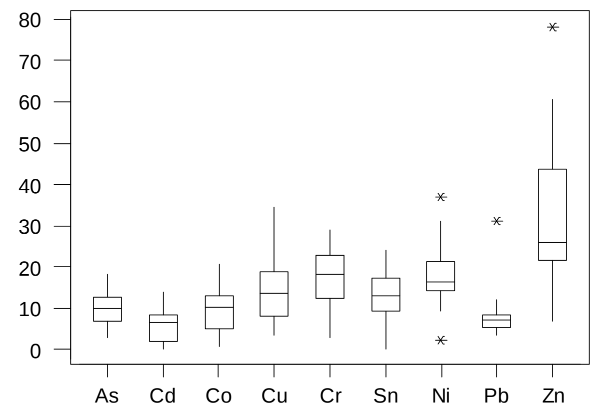
$Q_1 - 1{,}5(Q_3 - Q_1) = 12{,}51$
$Q_3 + 1{,}5(Q_3 - Q_1) = 23{,}35$

Representación gráfica de una caja (**box**) de ancho indiferente pero cuya largura corresponde a los valores Q_1, Q_2 (la mediana) y Q_3 .
También se incluye un intervalo (**whiskers**) que engloba a los puntos más extremos comprendidos dentro de $1{,}5(Q_3-Q_1)$
Lo ideal es una caja simétrica en torno a la mediana y de bigotes idénticos.

cuartil(datos;cuartil)

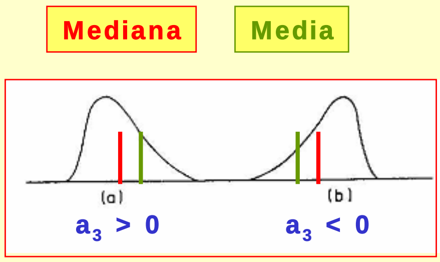
nº : 0 mínimo
1 primer cuartil
2 mediana
3 tercer cuartil
4 máximo

EJEMPLO DE BOXPLOTS MULTIPLES



ESTADÍSTICA DESCRIPTIVA

Medidas de la forma de la distribución



Coeficiente de sesgo (skewness)

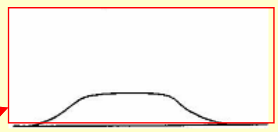
$$a_3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$$

a3 > 0 : La distribución tiene cola hacia la derecha; Media > Mediana
a3 = 0 : La distribución es simétrica; Media = Mediana
a3< 0 : La distribución tiene cola hacia la izquierda; Media < Mediana

Coeficiente de curtosis o apuntamiento

$$a_4 - 3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4 - 3$$

a4-3 > 0: La curva es apuntada (leptocúrtica)
a4-3 = 0: La curva es normal
a4-3 < 0: La curva es achatada (platicúrtica)



MEDIDA DE LA CALIDAD

Calidad y errores

Quality Assurance (Aseguramiento de la Calidad): Sistema de actividades cuyo propósito es proporcionar al productor o usuario de un producto o servicio la seguridad de que éste cumple con unas estándares o normas de calidad definidas con un nivel de confianza dado.

La calidad de una medida se obtiene si el resultado dado se aproxima al resultado correcto, es decir no está sujeto a un error mayor del que se considera aceptable.

Errores sistemáticos y aleatorios

	Determinación							
Serie	**1**	**2**	**3**	**4**	**5**	**6**	**7**	$\bar{x}$
A	10,1	9,9	10,1	10,0	9,9	10,2	9,8	10,0
B	10,3	9,7	9,6	9,8	10,1	10,5	10,2	10,0
C	11,3	10,7	11,0	10,9	11,1	11,3	10,7	11,0
D	10,0	9,8	10,1	9,9	10,2	10,3	12,7	10,4
E	9,7	9,8	9,9	10,0	10,1	10,2	10,3	10,0

Valor correcto 10,0

EL ERROR SISTEMÁTICO SIEMPRE VA ACOMPAÑADO DEL ERROR ALEATORIO

Las pruebas estadísticas se dirigen a diferenciar entre ambos tipos de error

PRECISIÓN Y BIAS DE LAS MEDIDAS

El objetivo de una medida química es encontrar el verdadero valor de un parámetro químico

VALOR VERDADERO (ISO): *"Valor que caracteriza una cantidad perfectamente definida en las condiciones que existen en el momento en que tal cantidad es observada. Es un valor ideal que solo puede obtenerse cuando todas las fuentes de error son eliminadas y la población es infinita"*.

Hay **dos causas** por las que un resultado analítico difiera del valor verdadero: existencia de **error aleatorio** o de **error sistemático**.

Un único resultado analítico, x_i, podrá diferir del valor verdadero μ_0, y esa diferencia es el **error**:

$$\boxed{e_i = x_i - \mu_0}$$

Con más medidas (muestra de una población), se obtendrá la media, $\bar{x}$, que estima μ, o la media de la población de medidas. Si la muestra es lo bastante grande, entonces $\bar{x} = \mu$

$$\boxed{e_i = (x_i - \bar{x}) + (\bar{x} - \mu_0) = (x_i - \mu) + (\mu - \mu_0)}$$

error aleatorio **error sistemático**

LOS ERRORES SISTEMÁTICOS CONDUCEN A INEXACTITUD (INACCURACY) Y BIAS, MIENTRAS QUE LOS ERRORES ALEATORIOS CONDUCEN A LA IMPRECISIÓN

Precisión

Medida del componente aleatorio

ISO *"Cercanía entre resultados de pruebas independientes obtenidos bajos condiciones estipuladas" ("closeness of agreement between independent test results obtained under stipulated conditions").*

IUPAC **IUPAC** *"Cercanía entre los resultados obtenidos aplicando el procedimiento experimental varias veces bajo las condiciones prescritas. Cuanto menor es la parte aleatoria de los errores experimentales que afectan a los resultados, más preciso es el procedimiento".*

Desviación típica "la medida cuantitativa de la precisión"

s estima σ, y si el número n de medidas replicadas es bastante grande, se puede considerar que $s=\sigma$. El valor de n varía, y según la IUPAC basta que $n \geq 10$, pero en muchos libros de estadística se dice que $n \geq 25$ o 30

Dependiendo de las condiciones experimentales, hay dos tipos de precisión:

- **REPETITIVIDAD (REPEATABILITY)** si las condiciones de medida son homogéneas: mismo método, mismo analista, mismo día, mismo laboratorio
- **REPRODUCIBILIDAD (REPRODUCIBILITY)** si las condiciones son las más adversas (o heterogéneas): generalmente diferente laboratorio
- **PRECISÓN INTERMEDIA (WITHIN-LABORATORY REPRODUCIBILITY)** las condiciones de medida son heterogéneas pero solo hay un único laboratorio

Exactitud, veracidad y bias (sesgo)

En el caso de los errores sistemáticos, la terminología es confusa y ambigua ya que aparecen los términos **exactitud (accuracy)** y **veracidad (trueness)**

ISO

Veracidad es *"la concordancia entre el valor medio de una gran serie de medidas y el valor aceptado como referencia" ("closeness of agreement between the average result obtained from a large series of test results and the acepted reference value").*

Exactitud **La veracidad es mejor concepto que la exactitud** *de una serie de medidas) y el valor aceptado como referencia" ("the closeness of agreement between test result and the acepted reference value").*

La diferencia entre ambos conceptos es que veracidad implica un gran número de series de medidas, mientras que exactitud se refiere a un único resultado (o a la media de una única serie).

IUPAC

Veracidad. No está reconocida.

Exactitud es *"la cercanía entre el valor verdadero y la media de la población obtenida aplicando el procedimiento experimental un gran número de veces.*

"accuracy was at one time used to cover only the component now named trueness"

BIAS o SESGO

"diferencia entre la esperanza de un resultado y un valor aceptado como referencia"

Bias $\Delta = \mu - \mu_0$ o $\bar{x} - \mu_0$

- **BIAS DEL MÉTODO**: Inherente al método de análisis. Diferencia entre la esperanza de los resultados de todos los laboratorios que usan ese método y el valor aceptado como referencia
- **BIAS DEL LABORATORIO**: Introducido por un laboratorio concreto. Diferencia entre la media de un gran número de resultados de un laboratorio y el valor medio global obtenido por todos los laboratorios

DETECCIÓN Y DETERMINACIÓN DE BIAS

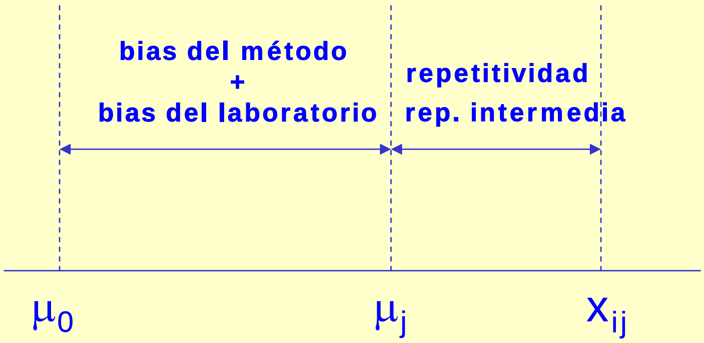
Diferenciar entre bias del laboratorio y del método, así como entre repetitividad reproducibilidad, no es tan simple como parece a primera vista y obliga a trabajar con ayuda de otros laboratorios

Un laboratorio j trabajando en **solitario únicamente puede evaluar la repetitividad (y la reproducibilidad intermedia) y no puede diferenciar entre bias del método y bias del laboratorio**, pues solo tiene acceso a la media de sus determinaciones que estima μj.y a sus resultados individuales

Repetitividad
Reprod. intermedia

$$e_{ij} = (x_{ij} - \mu_j) + (\mu_j - \mu_0)$$

bias (método + laboratorio)



DETECCIÓN Y DETERMINACIÓN DE BIAS

Si se une a otros laboratorios, podrá **diferenciar entre los bias del método y los bias de su laboratorio y podrá diferenciar entre reproducibilidad y repetitividad**. En este caso además de la media de sus determinaciones μj, tiene acceso a la media de las medias de los diferentes laboratorios, μ, que estimará mucho mejor μ0

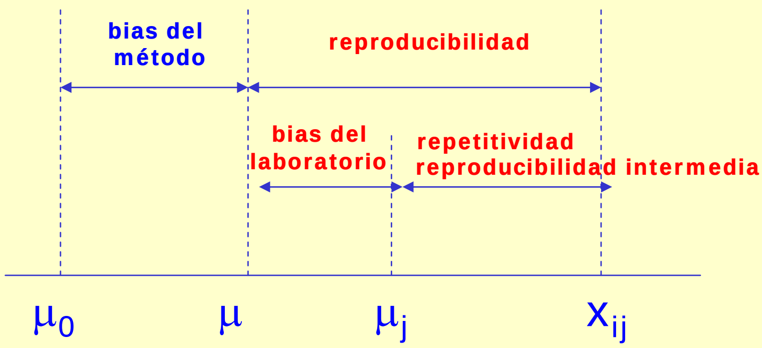
reproducibilidad

Repetitividad
Rep. interm.

bias
laboratorio

bias
método

$$e_{ij} = (x_{ij} - \mu_j) + (\mu_j - \mu) + (\mu - \mu_0)$$



Dependiendo de la situación, el **bias del laboratorio** puede ser considerado como parte del error **sistemático** (un laboratorio individual) o del **aleatorio** (muchos laboratorios tomados en conjunto).

Para evitar ambigüedades sobre la naturaleza sistemática o aleatoria de los errores, la ISO define el término **incertidumbre (uncertainty):** Un estimador añadido a un resultado que caracteriza el rango de valores dentro de los cuales se afirma que está el valor verdadero. Este rango viene afectado por diversos componentes, tanto aleatorios como sistemáticos.

Este término se calcula como una **varianza combinada** siguiendo las leyes de propagación de errores.

Otros tipos de error

	Determinación							
Serie	**1**	**2**	**3**	**4**	**5**	**6**	**7**	$\bar{x}$
A	10,1	9,9	10,1	10,0	9,9	10,2	9,8	10,0
B	10,3	9,7	9,6	9,8	10,1	10,5	10,2	10,0
C	11,3	10,7	11,0	10,9	11,1	11,3	10,7	11,0
D	10,0	9,8	10,1	9,9	10,2	10,3	$\boxed{12,7}$	10,4
E	9,7	9,8	9,9	10,0	10,1	10,2	10,3	10,0

Valor correcto 10,0

Errores espurios, groseros o inaceptables (gross errors)
Conducen a valores aberrantes (outliers). Falsean las estimaciones estadísticas y deben ser detectados y eliminados (evidentemente hay que encontrar la razón de su aparición)

Deriva (drift)
Aparecen en el caso de procesos que no están bajo control estadístico: su media y su desviación típica no son constantes.

Ruido de la línea base (baseline noise).
Las medidas en química suelen obtenerse por diferencia entre una señal obtenida cuando se mide el analito y una señal obtenida para un blanco. El ruido del blanco se superpone a la medida de la muestra y en ocasiones es indistinguible: se opera por debajo del límite de detección.

PROPAGACIÓN DE ERRORES

Si el resultado final se obtiene a partir de varias medidas **independientes** entre sí, o cuando está influido por varias fuentes de error **independientes** (p.e. muestreo y medida), los errores se propagan y pueden acumularse o compensarse.

Dada y = f(x, z, t...), los **errores aleatorios** se propagan de acuerdo a:

$$\sigma_y^2 = \left(\frac{\partial f}{\ldots}\right)^2 \sigma^2 + \left(\frac{\partial f}{\ldots}\right)^2 \sigma^2 + \left(\frac{\partial f}{\ldots}\right)^2 \sigma^2 + \ldots$$

Los errores aleatorios siempre se acumulan

Las varianzas son aditivas cuando la función es adición o sustracción, y si es multiplicación o división lo que son aditivos son los cuadrados de las desviaciones típicas relativas.

Los **errores sistemáticos** se propagan con su signo.
Para una ecuación de adición y/o sustracción $y = a + b\,x + c\,z - d\,t$

$$\Delta y = b\,\Delta x + c\,\Delta z - d\,\Delta t$$

Si la relación es multipli

Los errores sistemáticos a veces (?) se compensan

$$\Delta y/y = \Delta x/x + \Delta z/z - \Delta t/t$$

Estas ecuaciones se usan en **metrología**, porque permiten describir las fuentes individuales de error y combinarlas en lo que se denomina **incertidumbre (uncertainty).**

INCERTIDUMBRE (UNCERTAINTY): Intervalo en el cual debe estar incluido el valor verdadero. Si la incertidumbre se expresa como desviación típica, se denomina incertidumbre típica. Cuando hay varias fuentes de error debe utilizarse una incertidumbre típica combinada, o una **varianza combinada (u)** obtenida mediante las leyes de propagación de errores.

PROPAGACIÓN DE ERRORES

RESUMEN DE FÓRMULAS

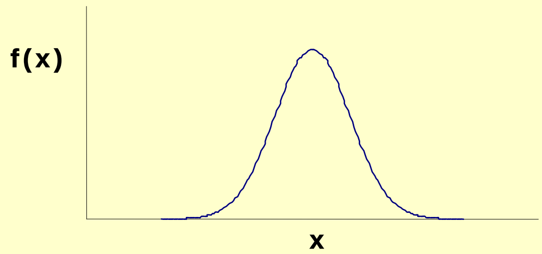
Función	Error aleatorio	Error sistemático
Combinaciones lineales $y = k + k_a a + k_b b + \ldots$	$s_y = \sqrt{(k_a s_a)^2 + (k_b s_b)^2 + \ldots}$	$\Delta y = k_a \Delta a + k_b \Delta b + \ldots$
Expresiones multiplicativas $y = k\, ab/c$	$\frac{s_y}{y} = \sqrt{\left(\frac{s_a}{a}\right)^2 + \left(\frac{s_b}{b}\right)^2 + \left(\frac{s_c}{c}\right)^2}$	$\frac{\Delta y}{y} = \frac{\Delta a}{a} + \frac{\Delta b}{b} - \frac{\Delta c}{c}$
Potencias $y = a^k$	$\frac{s_y}{y} = \left\lvert k\frac{s_a}{a}\right\rvert$	$\frac{\Delta y}{y} = k\frac{\Delta a}{a}$
Función de una variable independiente $y = f(a)$	$s_y = \left\lvert s_a \frac{dy}{da}\right\rvert$ ó $s_y^2 = s_a^2\left(\frac{dy}{da}\right)^2$	$\Delta y = \Delta a \frac{dy}{da}$
Función de varias variables independientes $y = f(a, b, \ldots)$	$s_y^2 = s_a^2\left(\frac{\partial y}{\partial a}\right)_b^2 + s_b^2\left(\frac{\partial y}{\partial b}\right)_a^2 + \ldots$	$\Delta y = \Delta a\left(\frac{\partial y}{\partial a}\right)_b + \Delta b\left(\frac{\partial y}{\partial b}\right)_a + \ldots$

DISTRIBUCIÓN NORMAL O GAUSSIANA

Con infinitos datos en vez de una diagrama de distribución de frecuencias (histograma), se obtiene una distribución continua de probabilidades (función de densidad de probabilidad)

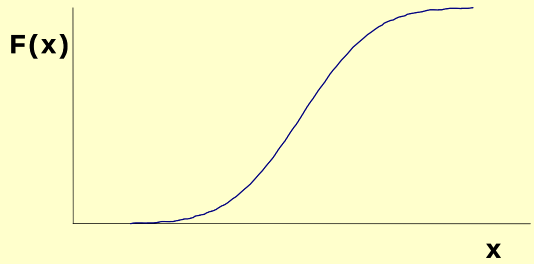
Para la distribución normal o gaussiana:

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



La **distribución de probabilidad normal acumulativa** es

$$\boxed{F(x_0) = \int_{-\infty}^{x_0} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] dx}$$

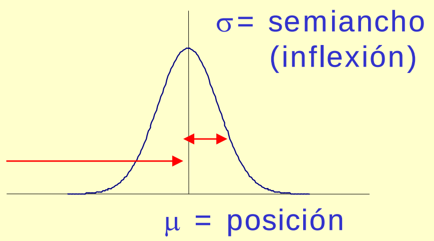


μ: **media de la población**

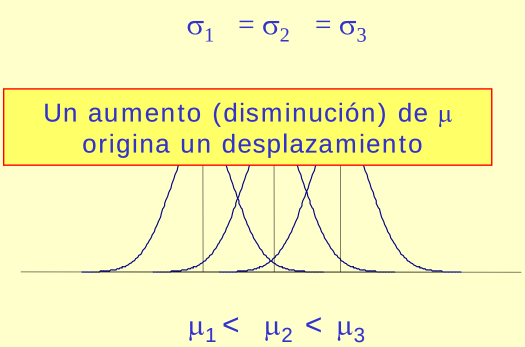
σ: **desviación típica** $\sigma = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{N}}$

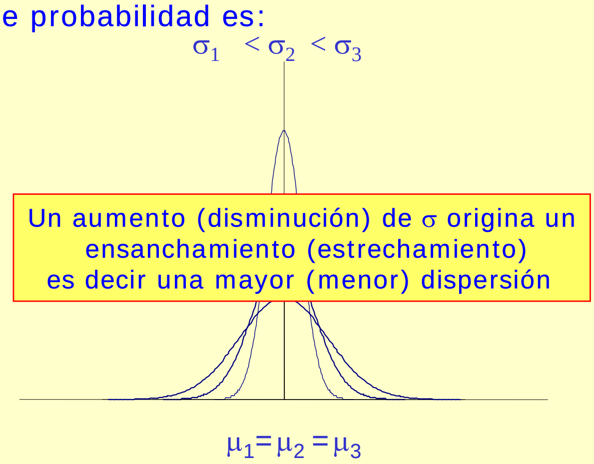
DISTRIBUCIÓN NORMAL O GAUSSIANA

Los parámetros que describen una distribución normal son μ y σ^2: **N(μ , σ^2)**

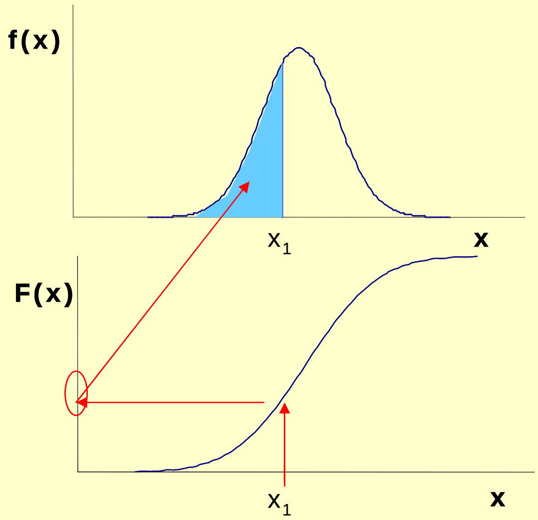


Su efecto sobre la función de densidad de probabilidad es:



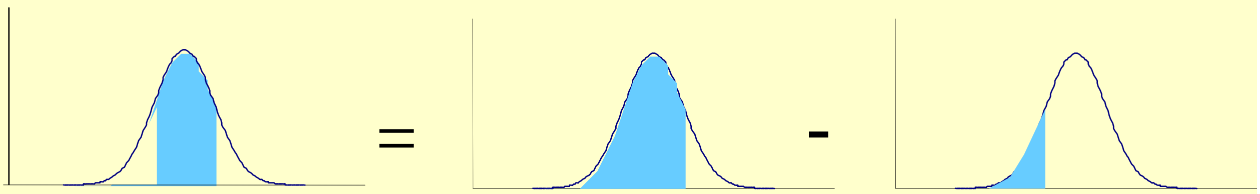


DISTRIBUCION NORMAL O GAUSSIANA

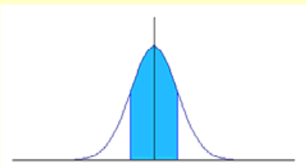


El área bajo la curva de f(x) desde $-\infty$ a x_1 es la probabilidad de encontrar un valor menor o igual que x_1

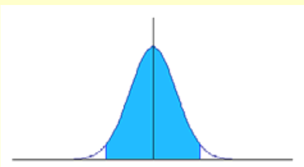
La probabilidad de encontrar un valor en el entorno de un determinado x_1 puede calcularse por diferencia de áreas



DISTRIBUCION NORMAL O GAUSSIANA

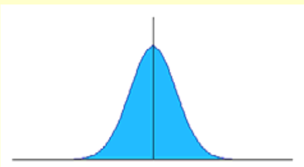


$\mu \pm \sigma$ incluye al 68,26% de los datos



Metrología

$\mu \pm 2\sigma$ incluye al 95,44% de los datos
$\mu \pm 1{,}96\ \sigma$ incluye al 95,00% de los datos



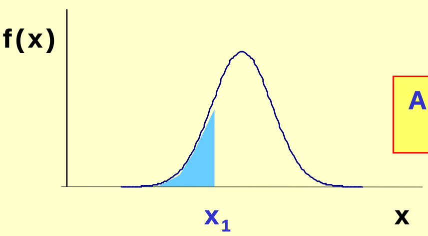
$\mu \pm 3\sigma$ incluye al 99,74% de los datos

DISTRIBUCIÓN NORMAL TIPIFICADA O ESTANDARIZADA

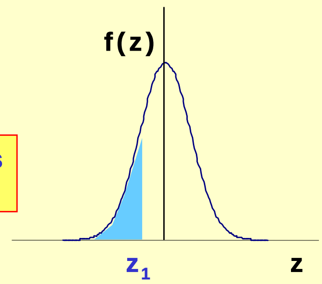
μ y σ tienen su escala y unidades, por lo que las formas y valores de f(x) y F(x) darán lugar a infinitas distribuciones N(μ,σ^2)
Para evitar este efecto se hace una transformación por escalado o autoescalado. Se calcula $\boxed{z = (x-\mu)/\sigma,}$ con lo que se obtiene una nueva distribución N(0,1)

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

$$\Phi(z_0) = \int_{-\infty}^{z_0} \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-(z)^2}{2}\right] dz$$



Ambas probabilidades son idénticas



$$P(x \le x_1) = \int_{-\infty}^{x_1} f(x)dx$$

$$P(z \le z_1) = \int_{-\infty}^{z_1} f(z)dz$$

DISTRIBUCIÓN NORMAL TIPIFICADA O ESTANDARIZADA

Tablas para la distribución tipificada o estandarizada

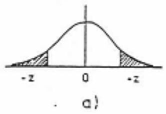
La trar … áficas para las infinita … ara la variable z, que se obti

Por tar … .

TABLE 3.1

Values of z and the two-tailed probability that its absolute value will be exceeded in a normal population (see also Fig. 3.3a)

	Second decimal in p									
p	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	∞	2.576	2.326	2.170	2.054	1.960	1.881	1.812	1.750	1.695
0.1	1.645	1.598	1.555	1.514	1.476	1.439	1.405	1.372	1.340	1.311
0.2	1.281	1.254	1.226	1.200	$\boxed{1.175}$	1.150	1.126	1.103	1.080	1.058
0.3	1.036	1.015	0.994	0.974	0.954	0.935	0.915	0.896	0.878	0.860
0.4	0.842	0.824	0.806	0.789	0.772	0.755	0.739	0.722	0.706	0.690
0.5	0.674	0.659	0.643	0.623	0.613	0.598	0.583	0.568	0.553	0.539
0.6	0.524	0.510	0.496	0.482	0.468	0.454	0.440	0.436	0.412	0.399
0.7	0.385	0.372	0.358	0.345	0.332	0.319	0.305	0.292	0.279	0.266
0.8	0.253	0.240	0.228	0.215	0.202	0.189	0.176	0.164	0.151	0.138
0.9	0.126	0.113	0.100	0.088	0.075	0.063	0.050	0.038	0.025	0.013
p	0.002		0.001	0.0001	0.00001	0.000001	0.0000001		0.00000001	
z	3.090		3.290	3.890	4.417	4.891	5.326		5.730	



a)

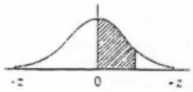
Si el área rayada (probabilidad) vale 0,24
z valdrá 1,175

DISTRIBUCIÓN NORMAL TIPIFICADA O ESTANDARIZADA

Tablas para la distribución tipificada o estandarizada

TABLE 3.2

Probability p to find a value between 0 and z (see also Fig. 3.3b)



Second decimal of z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000	0.004	0.008	0.012	0.016	0.020	0.024	0.028	0.032	0.036
0.1	0.040	0.044	0.048	0.052	0.056	0.060	0.064	0.067	0.071	0.075
0.2	0.079	0.083	0.087	0.091	0.095	0.099	0.103	0.106	0.110	0.114
0.3	0.118	0.122	0.125	0.129	0.133	0.137	0.144	0.141	0.148	0.152
0.4	0.155	0.159	0.163	0.166	0.170	0.174	0.177	0.181	0.184	0.188
0.5	0.191	0.195	0.198	0.202	0.205	0.209	0.212	0.216	0.219	0.222
0.6	0.226	0.229	0.232	0.236	0.239	0.242	0.245	0.249	[illegible]	0.255
0.7	0.258	0.261	0.264	0.267	0.270	0.273	0.276	0.279	0.282	0.285
0.8	0.288	0.291	0.294	0.297	0.299	0.302	0.305	0.308	0.311	0.313
0.9	0.316	0.319	0.321	0.324	0.326	0.329	0.331	0.334	0.336	0.339
1.0	0.341	0.344	0.346	0.348	0.351	0.353	0.355	0.358	0.360	0.362
1.1	0.364	0.366	0.369	0.371	0.373	0.375	0.377	0.379	0.381	0.383
1.2	0.385	0.387	0.389	0.391	0.392	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.3	0.403	0.405	0.407	0.408	0.410	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.4	0.419	0.421	0.422	0.424	0.425	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.5	0.433	0.434	0.436	0.437	0.438	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.6	0.445	0.446	0.447	0.448	0.449	0.450	0.451	0.452	0.453	0.454
1.7	0.455	0.456	0.457	0.458	0.459	0.460	0.461	0.462	0.462	0.463
1.8	0.464	0.465	0.466	0.466	0.467	0.468	0.469	0.469	0.470	0.471
1.9	0.471	0.472	0.473	0.473	0.474	0.474	0.475	0.476	0.476	0.477

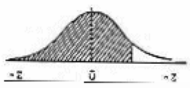
z =	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	
$F(z)$ =	0.477	0.482	0.486	0.489	0.492	0.494	0.495	0.496	0.497	0.498	
z =	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
$F(z)$	0.4987	0.4990	0.4993	0.4995	0.4997	0.4998	0.4998	0.4998	0.4999	0.49995	0.49997

Si z vale 0,78
la probabilidad es p = 0,282

DISTRIBUCIÓN NORMAL TIPIFICADA O ESTANDARIZADA

Tablas para

Probability to find a value lower than z (see also Fig. 3.3c)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500	0.504	0.508	0.512	0.516	0.520	0.524	0.528	0.532	0.536
0.1	0.540	0.544	0.548	0.552	0.556	0.560	0.564	0.567	0.571	0.575
0.2	0.579	0.583	0.587	0.591	0.595	0.599	0.603	0.606	0.610	0.614
0.3	0.618	0.622	0.625	0.629	0.633	0.637	0.641	0.644	0.648	0.652
0.4	0.655	0.659	0.663	0.666	0.670	0.674	0.677	0.681	0.684	0.688
0.5	0.691	0.695	0.698	0.702	0.705	0.709	[illegible]	[illegible]	[illegible]	[illegible]
0.6	0.726	0.729	0.732	0.736	0.739	0.742	[illegible]	[illegible]	[illegible]	[illegible]
0.7	0.758	0.761	0.764	0.767	0.770	0.773	[illegible]	[illegible]	[illegible]	[illegible]
0.8	0.788	0.791	0.794	0.797	0.799	0.802	[illegible]	[illegible]	[illegible]	[illegible]
0.9	0.816	0.819	0.821	0.824	0.826	0.829	[illegible]	[illegible]	[illegible]	[illegible]
1.0	0.841	0.844	0.846	0.848	0.851	0.853	0.855	0.858	0.860	0.862
1.1	0.864	0.866	0.869	0.871	[illegible]	0.875	0.877	0.879	0.881	0.883
1.2	0.885	0.887	0.889	0.891	0.892	0.894	0.896	0.898	0.900	0.901
1.3	0.903	0.905	0.907	0.908	[illegible]	0.911	0.913	0.915	0.916	0.918
1.4	0.919	0.921	0.922	0.924	0.925	0.926	0.928	0.929	0.931	0.932
1.5	0.933	0.934	0.936	0.937	0.938	0.939	0.941	0.942	0.943	0.944
1.6	0.945	0.946	0.947	0.948	0.949	0.950	0.951	0.952	0.953	0.954
1.7	0.955	0.956	0.957	0.958	0.960	0.961	0.962	0.962	0.962	0.963
1.8	0.964	0.965	0.966	0.966	0.967	0.968	0.969	0.969	0.970	0.971
1.9	0.971	0.972	0.973	0.973	0.974	0.974	0.975	0.976	0.976	0.977

Si z vale 1,24
la probabilidad p = 0,892

WORD

EXCEL

Funciones EXCEL

DISTR.NORMAL.ESTAND(z) devuelve la probabilidad desde -infinito a z

DISTR.NORMAL.ESTAND.INV(p) devuelve un z tal que la probabilidad desde –infinito a z valga p

TEOREMA DEL LÍMITE CENTRAL

Si se tiene una suma **y** de n variables aleatorias independientes $\mathbf{x_i}$, cuyas distribuciones no son necesariamente normales:

$$y_i = x_1 + x_2 + ... + x_n$$

con medias μ_i y varianzas σ_i^2, para grandes valores de n, la distribución de **y** es **aproximadamente normal** con una media $\Sigma\mu_i$ y varianza $\Sigma\sigma_i^2$.

METROLOGÍA

Explica por qué las distribuciones de errores aleatorios tienden a la normalidad, puesto que el error global suele ser una combinación lineal de componentes independientes.

APLICACIÓN A LAS MEDIAS MUESTRALES

Si de una población con media μ y varianza σ^2 se toman **todas las posibles muestras de tamaño n**, la distribución de las medias muestrales tendrá una media μ y una varianza σ^2/n. La distribución de las medias será normal si la población original es normal y será aproximadamente normal para cualquier tipo de distribución de la población original, y tanto más normal cuanto más grande sea n.

El valor crítico suele ser $n>30$, pero para poblaciones no normales, simétricas y uní modales, se obtienen medias distribuidas normalmente con muestras de tamaño 4-5.

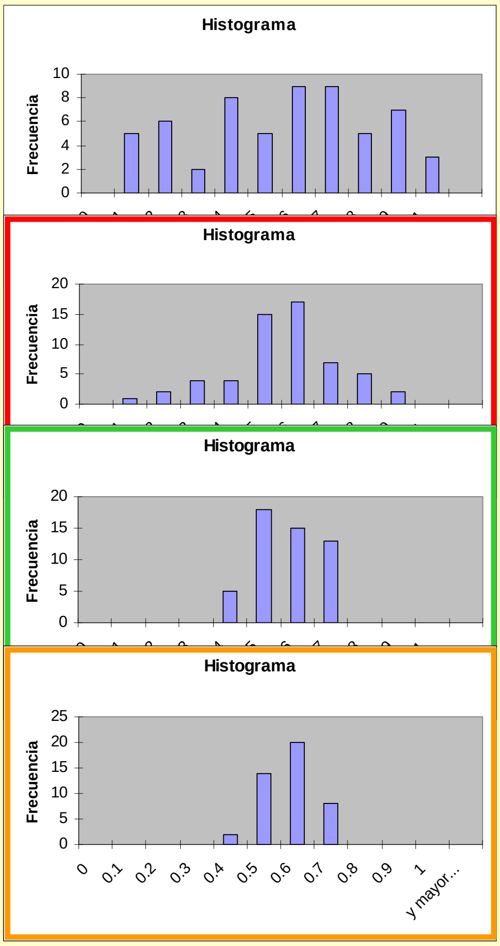
TEOREMA DEL LÍMITE CENTRAL

EXCEL

DISTRIBUCIONES DE MEDIAS			
Variable	Media (3)	Media(9)	Media(16)
0.57089422	0.47192904	0.35311918	0.42230626
0.56756603	0.5610972	0.37299614	0.40639894
0.27732686	0.4136184	0.40639443	0.39283992
0.83839872	0.35758944	0.41107005	0.41723036
0.12512961	0.09716826	0.3302782	0.39317178
0.10923998	0.16746616	0.31796386	0.42638695
0.05713519	0.22983906	0.37879498	0.4450656
0.33602332	0.46072295	0.46791218	0.45060626
0.29635867	0.63809875	0.46572926	0.49045321
0.74978687	0.64578165	0.4717585	0.51480979
0.86815071	0.43294339	0.46262371	0.47082195
0.31940739	0.14832666	0.41654713	0.46617327
0.11127207	0.26076422	0.4540099	0.49897182
0.01430052	0.5100702	0.4869906	0.5511098
0.65672008	0.61076239	0.50160008	0.59819415
0.85919001	0.50872961	0.53680612	0.57611524
0.31637708	0.44485754	0.5175699	0.55164652
....	....	....	....

aleatorio()

promedio



INTERVALOS DE CONFIANZA DE LA MEDIA

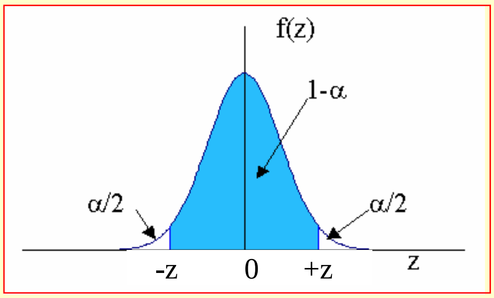
Para una distribución normal, el 95% de los datos cae dentro de los límites de z=-1,96 a z=+1,96. Esto puede ser refraseado para indicar que el 95% de los datos caen dentro del intervalo $\mu \pm 1,96\sigma$.

Si se aplica a la distribución de las medias muestrales, podemos decir que el 95% de los datos estará dentro del intervalo. $\mu \pm 1,96\sigma / \sqrt{n}$

Es decir: $\boxed{\mu = \bar{x} \pm 1,96\sigma / \sqrt{n}}$

En general $\mu = \bar{x} \pm z_{\alpha/2}\,(\sigma / \sqrt{n})$ con 100-α% de confianza, donde α se deriva de una tabla z.

de dos colas



MUESTRAS PEQUEÑAS Y LA DISTRIBUCIÓN t

Cuando no se conoce σ sino su **estimación s**, si n≥30 (ó 25 según investigadores), puede hacerse la sustitución: $\boxed{\mu = \bar{x} \pm z_{\alpha/2}\,(s / \sqrt{n}) \quad (n \geq 30)}$

Si **n<30**,s es un estimador incierto de σ, y **se sustituye z por t**

$$\boxed{\mu = \bar{x} \pm t_{\alpha/2,(n-1)}\,(s / \sqrt{n}) \quad (n < 30)}$$

siendo α el nivel de significación, y n-1 el número de grados de libertad con el que debe buscarse t en la correspondiente tabla de t de dos colas

Si se desea un nivel de confi...

- Si la tabla t es de dos cola...
- Si la tabla t es de una cola... columna p=α=0,025

En este caso la tabla es de una cola, luego hay que usar p = 0,025

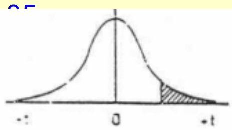


TABLE 3.4
One-sided t-table (see also Fig. 3.3d)

Area in upper tail of t-distribution

df	0.10	0.05	0.025	0.01	0.005	0.0025	0.001
1	3.078	6.314	12.706	31.821	63.657	127.32	318.310
2	1.886	2.920	4.303	6.965	9.925	14.089	22.327
3	1.638	2.353	3.182	4.541	5.841	7.453	10.215
4	1.533	2.132	2.776	3.747	4.604	5.598	7.173
5	1.476	2.015	2.571	3.365	4.032	4.773	5.893
6	1.440	1.943	2.447	3.143	3.707	4.317	5.208
7	1.415	1.895	2.365	2.998	3.499	4.029	4.785
8	1.397	1.860	2.306	2.896	3.355	3.832	4.501
9	1.383	1.833	2.262	2.821	3.250	3.690	4.297

MUESTRAS PEQUEÑAS Y LA DISTRIBUCIÓN t

Para saber si la tabla t es de una o dos colas, basta mirar la última fila que corresponde a infinitos g.d.l. y buscar la columna con t=1,96.

- Si está en una columna con p=0,05, la tabla es de dos colas
- Si está en una columna con p=0,025, la tabla es de una cola

TABLE 3.4
One-sided *t*-table (see also Fig. 3.3d)

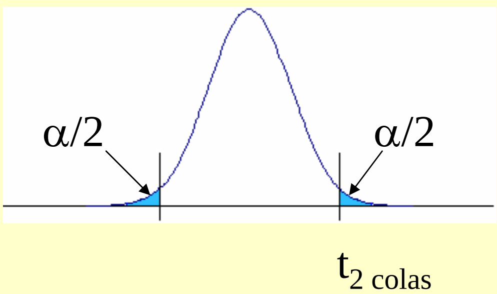
	Area in upper tail of *t*-distribution						
df	0.10	0.05	0.025	0.01	0.005	0.0025	0.001
1	3.078	6.314	12.706	31.821	63.657	127.32	318.310
2	1.886	2.920	4.303	6.965	9.925	14.089	22.327
3	1.638	2.353	3.182	4.541	5.841	7.453	10.215
60	1.296	1.671	2.000	2.390	2.660	2.915	3.232
120	1.289	1.658	1.980	2.358	2.617	2.860	3.160
∞	1.282	1.645	1.960	2.326	2.576	2.807	3.090

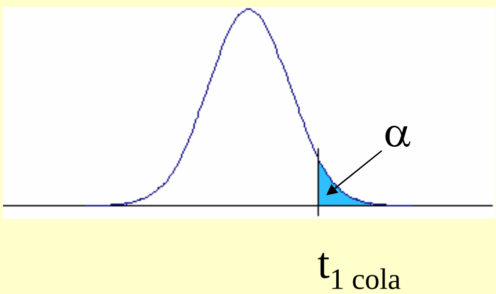
FUNCIONES EXCEL

DISTR.T.INV(p;g.d.l) devuelve el valor de t con el nº de g.d.l. elegido y el nivel de probabilidad p, en una **tabla de dos colas**

Para que sirva para **una cola** hay que ponerla como DISTR.T.INV(2p;g.d.l)

TABLAS DE UNA Y DOS COLAS





$t_{2\ colas} > t_{1\ cola}$ para igual nivel de significación

Para z es análogo

Si se dispone de una **tabla de una cola o de la función EXCEL** DISTR.NORMAL.ESTAND.INV(p)

- Si lo que se desea buscar es de una cola no hay problema: se entra por la columna encabezada por (o se sustituye en la función) p = α = 0,05
- Si lo que se desea es de dos colas, hay que entrar por la columna encabezada por (o se sustituye en la función) p = α/2 = 0,05/2= 0,025

Si se dispone de una **tabla de dos colas o de la función EXCEL** DISTR.T.INV(p;g.d.l)

- Si lo que se desea buscar es de dos colas no hay problema: se entra por la columna encabezada por (o se sustituye en la función) p = α = 0,05
- Si lo que se desea es de una cola, hay que entrar por la columna (o se sustituye en la función) p = 2α = 2*0,05 = 0,10

PRUEBAS DE NORMALIDAD

No todas las distribuciones son normales.

El **conocer si una determinada muestra o población sigue una distribución normal es importante**, ya que permite detectar un efecto que no es explicable mediante errores de medida aleatorios.

El método **RANKIT** (recomendado por ISO) es un método rápido para saber si una muestra sigue una distribución normal

Supongamos una serie de medidas:
2,286 ; 2,327 ; 2,388; 3,172 ; 3,158 ; 2,751 ; 2,222; 2,367 ; 2,247 ; 2,512 ; 2,104; 2,707

1) Se cuentan (n =12) y se ordenan los datos de menor a mayor (to rank):
2,104; 2,222; 2,247; 2,286; 2,327; 2,367; 2,388; 2,512 ; 2,707; 2,751; 3,158; 3,172

2) Estos n=12 valores pertenecen a una muestra con (n+1) = 13 intervalos de frecuencia unidad (cf. histogramas)

3) La frecuencia acumulativa se obtiene sumando uno a cada frecuencia y coincide por tanto con el número de orden de cada dato individua

4) El porcentaje de frecuencia acumulativa se obtiene mediante la fórmula:

$$\boxed{\%\ \text{frec. acum.} = (100 \times \text{fre. acum}) \ / \ (n+1)}$$

PRUEBAS DE NORMALIDAD

5) Si la **distribución fuera normal**, el **porcentaje de frecuencia relativa acumulativa de un determinado valor coincidiría con la probabilidad** de que dicho valor apareciera y se puede encontrar el valor **z** correspondiente **asociado** a dicha probabilidad.

Para ello se usa la adecuada tabla z o bien EXCEL:
Función =DISTR.NORM.ESTAND.INV(xx), siendo xx el tanto por uno de frecuencia acumulativa de un determinado dato

Se obtienen así una serie de valores de z ordenados que se denominan rankits

6) Dado que entre z y x debe haber una relación lineal, al representar z vs. x, ó x vs. z, debe obtenerse una línea recta

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

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

$$x_i = \mu + \sigma * z_i$$

PRUEBAS DE NORMALIDAD

Método rankit

Medidas (x)	Frec. Acum.	% Frec. Acum.	Z
2.104	1	7.7	-1.43
2.222	2	15.4	-1.02
2.247	3	23.1	-0.74
2.286	4	30.8	-0.50
2.327	5	38.5	-0.28
2.367	6	46.1	-0.10
2.388	7	53.8	0.10
2.512	8	61.5	0.28
2.707	9	69.2	0.50
2.751	10	76.9	0.74
3.158	11	84.6	1.02
3.172	12	92.3	1.43

Simétricos

Ordenación

n° de orden

DISTR.NORM.ESTAND.INV(xx)

% frec. Acum. = (100 x frec. Acum.) / (n+1)

PRUEBAS DE NORMALIDAD

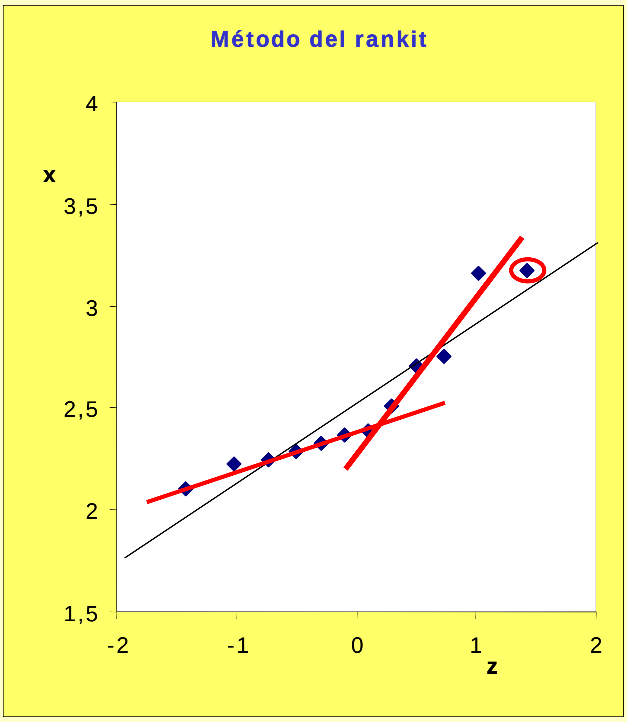
Método rankit

Relación lineal entre z y x

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

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

$$x_i = \mu + \sigma * z_i$$

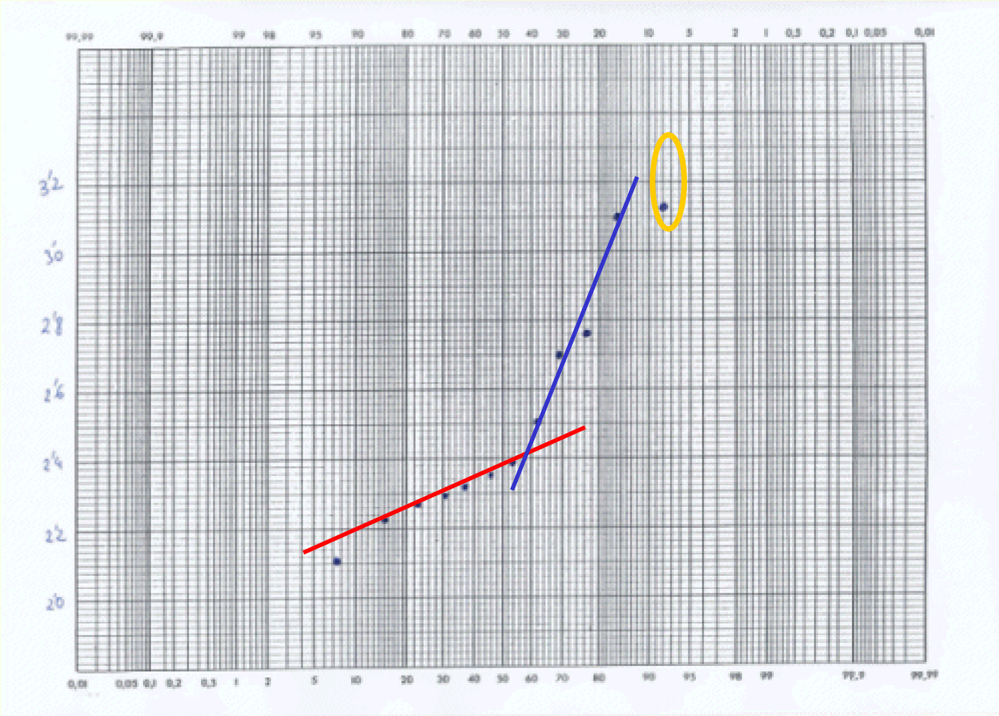


PRUEBAS DE NORMALIDAD

Papel de probabilidad normal (1)

Método rankit

Papel de probabilidad normal (2)



DISTRIBUCIÓN BINOMIAL

Una población binomial es aquella cuyos elementos pertenecen a **dos categorías mutuamente excluyentes**. Por ejemplo, un producto puede ser defectuoso o no, un paciente puede estar enfermo o sano....

Sea una serie de medidas o experimentos independientes cuyo resultado puede ser A o $\bar{A}$ (No A). La probabilidad de que ocurra A o $\bar{A}$ será:

$P(A) = \Pi$

$P(\bar{A}) = 1 - P(A) = 1 - \Pi$

Si se realizan n experimentos, la probabilidad de que A ocurra i veces será:

$$\boxed{P(i) = C_n^i \Pi^i (1-\Pi)^{n-i} = \frac{n!}{i!\,(n-i)!} \Pi^i (1-\Pi)^{n-i}}$$

Ejemplo: La probabilidad de encontrar un elemento defectuoso en un lote de productos es 0,02. Calcule la probabilidad de que aparezcan 2 elementos defectuosos en una muestra aleatoria de 10 elementos.

$$P(2) = C_{10}^2\, 0{,}02^2 (1-0{,}02)^{10-2} = \frac{10!}{2!\,(10-2)!} 0{,}02^{2i} (0{,}98)^8 = 0{,}0153 = 1{,}53\,\%$$

La media y la varianza de una distribución binomial son $\mu = n\Pi$ y $\sigma^2 = n\Pi(1-\Pi)$. (En el ejemplo $\mu =$ y $\sigma^2 =$)

La distribución binomial se aplica a:

- Muestreo sin reemplazamiento de una población muy grande comparada con el tamaño de la muestra
- Muestreo con reemplazamiento de una población finita
- Muestreo de procesos continuos de fabricación con población grande

Cuando $n\Pi$ y $n(1-\Pi)$ son grandes (>5)
la distribución binomial tiende a la normalidad

DISTRIBUCIÓN DE POISSON

Describe **variables discretas relacionadas con sucesos discretos en un intervalo continuo, tal como tiempo, espacio o volumen**. Se supone que los sucesos ocurren de forma aleatoria e independiente. Por ejemplo: el número de caramelos defectuosos de una fábrica, el número de averías mensuales en la maquinaria, el número de cuentas en la desintegración de un radioisótopo...

La probabilidad de que ocurra un número i de sucesos viene dada por

$$\boxed{P(i) = \frac{e^{-\lambda}\lambda^i}{i!}}$$

siendo λ la media, o número de sucesos que tiene lugar en un período dado de tiempo, espacio o volumen.

La media y la varianza en una distribución de Poisson coinciden: $\mu = \sigma^2 = \lambda$

La distribución de Poisson puede considerarse como una distribución binomial en la cual n tiende a infinito, Π tiende a cero y $n\Pi$ tiende a λ.

Si λ > 10, la distribución de Poisson tiende a la normalidad

DISTRIBUCIÓN χ^2 O DE PEARSON

Si un conjunto de variables independientes $z_1, z_2, \ldots, z_n$ están distribuidas según una distribución normal unidad N(0,1), la variable $\chi^2 = \sum_1^n z_i^2$ tiene una función de densidad de probabilidad con ν=n grados de libertad

$$\boxed{f(\chi^2) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)}(\chi^2)^{\nu/2-1}e^{-\chi^2/2} \qquad ; \nu \geq 0; \quad 0 \leq \chi^2 \leq \infty; \nu = n-1}$$

Γ es la función gamma

La función es asimétrica con una cola hacia la derecha cuando ν es pequeño. Según aumenta el número de grados de libertad, la distribución χ2 tiende a la normalidad.

La función χ^2 está tabulada para diversos grados de libertad, por lo que en la práctica no es preciso recordar ni calcular la función de densidad de probabilidad

La media y la varianza son: $\mu = \nu$ y $\sigma^2 = 2\nu$

Habitualmente las variables independientes iniciales no están normalizadas. Si tenemos n variables independientes $x_1, x_2, \ldots, x_n$ con distribución normal $N(\mu,\sigma^2)$ podemos obtener una nueva variable independiente

$$\sum_{i=1}^{n} \frac{\left(x_i - \bar{x}\right)^2}{\sigma^2} = (n-1)\frac{s^2}{\sigma^2}$$

que estará distribuida como χ^2 con ν = n-1 grados de libertad.

La distribución χ^2 se utiliza en el test χ^2 para comprobar si la distribución de los datos de una muestra de tamaño n se ajusta a una cierta distribución teórica, habitualmente la normal.

DISTRIBUCIÓN t DE STUDENT

Se utiliza para **describir muestreos con pocos elementos (n<30),** es decir para describir la distribución de una muestra en vez de la distribución de la población.

La distribución t con ν grados de libertad es simétrica en torno a cero, y su varianza es $\sigma^2 = \nu/(\nu-2)$. Se parece a la normal, pero es algo más apuntada y tiene la base más estrecha.
Tiende a la normalidad cuando ν tiende a infinito, y en la práctica coincide con la normal tipificada z cuando $\nu > 30$.

Sea z una variable aleatoria N(0,1) y χ^2 otra variable aleatoria independiente de la anterior, con una distribución χ^2_ν . Se puede definir una nueva variable t como

$$t = \frac{z}{\sqrt{\chi^2/\nu}}$$

La distribución t de Student con ν grados de libertad se describe mediante la función de densidad de probabilidad:

$$f(t) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma(\nu + 1/2)}{\Gamma(\nu/2)} \frac{1}{\left(1 + t^2/\nu\right)^{\nu + 1/2}}$$

Esta función está tabulada para diferentes grados de libertad.

Se utiliza cuando no se conocen las verdaderas media y varianza de una distribución para:
- **Estimar intervalos de confianza**
- **Evaluar la veracidad de un resultado (ausencia de bias)**
- **Comparar medias obtenidas por métodos diferentes**

DISTRIBUCIÓN F DE FISCHER

Describe **la distribución de dos variables aleatorias** que siguen una distribución χ^2
Su mayor utilidad es la comparación de varianzas y se basa en lo siguiente:

Sea $x_1, x_2, \ldots x_n$ una muestra aleatoria de variables aleatorias independientes con distribución $N(\mu_x, \sigma_x^2)$ e $y_1, y_2, \ldots y_n$ una muestra aleatoria de variables aleatorias independientes con distribución $N(\mu_y, \sigma_y^2)$. Las variables $F = \frac{\chi_x^2/\nu_x}{\chi_y^2/\nu_y} = \frac{s_x^2/\sigma_x^2}{s_y^2/\sigma_y^2}$ siguen una distribución χ^2 con $\nu_x = n_x - 1$ y $\nu_y = n_y - 1$ grados de libertad respectivamente.

Se puede definir una variable que seguirá una distribución F con n_x-1 y n_y-1 grados de libertad

$$\chi_x^2 = \frac{(n_x - 1)s_x^2}{\sigma_x^2} \quad y \quad \chi_y^2 = \frac{(n_y - 1)s_x^2}{\sigma_y^2}$$

Si las varianzas de ambas poblaciones (σ_x^2 y σ_y^2) son iguales, se cumplirá también :

$$\Rightarrow F = \frac{s_x^2}{s_y^2}$$

Si este cociente difiere de la unidad, σ_x^2 y σ_y^2 no serán idénticas, pues si lo fueran sus estimadores, s_x^2 y s_y^2, deberían ser muy parecidos.

El valor de la variable F está tabulado para diferentes probabilidades y diferentes grados de libertad.

Se utiliza para comparar la precisión, es decir las varianzas de dos medias o de dos procedimientos

DISTRIBUCIÓN RECTANGULAR

Aparece en los casos en los que la **densidad de probabilidad es constante en un intervalo dado**. Por ejemplo, al tirar un dado de 6 caras

En metrología se utiliza:
1) Cuando un certificado u otra especificación da unos límites sin especificar intervalos de confianza, por ejemplo el valor indicado en un matraz aforado: 25ml ± 0.05 ml.
2) Cuando se da una estimación en forma de un intervalo máximo (± a) y no se conoce la forma de la distribución

La incertidumbre (desviación típica) se estima como $\boxed{u(x) = \frac{a}{\sqrt{3}}}$

DISTRIBUCIÓN TRIANGULAR

Es una **distribución simétrica** como la que se obtiene al tirar dos dados de seis caras y sumar sus valores.

En metrología se utiliza:
1) Cuando se tiene un poco más de información que en el caso de una distribución rectangular y se sabe que los valores centrales de un intervalo tienen mayor probabilidad de aparecer que los extremos.
2) Cuando se da una estimación en forma de un intervalo máximo (± a) que sigue una distribución simétrica.

La incertidumbre (desviación típica) se estima como $\boxed{u(x) = \frac{a}{\sqrt{6}}}$