Parámetros de una población.

Para poder realizar un estudio estadístico, es necesaria previamente, la observación de sus individuos.

La observación de un individuo la describimos mediante uno o más caracteres. El carácter es, por tanto, una cualidad o propiedad inherente al individuo. Por ejemplo si el individuo estudiado es un libro, podemos describirlo mediante los caracteres peso, tamaño, número de hojas, color de las pastas, etc.

Dependiendo de que el carácter se pueda cuantificar (edad, peso, nº de hojas...) o no (color, opción política,...) se llama cuantitativo o cualitativo respectivamente.

El conjunto de valores que puede tomar un determinado carácter estadístico cuantitativo se llama **variable estadística**. Por ejemplo si queremos estudiar la edad de los alumnos del instituto, la variable estadística sería "edad de los alumnos del I.E.S. de Fene" y tomaría los valores: 12, 13, 14, 15, 16, 17, 18, 19 y 20.

Las variables estadísticas cuantitativas pueden ser de dos tipos:

- **Continuas:** si al menos en teoría pueden tomar infinitos valores dentro de un intervalo. Por ejemplo la talla de un individuo podría tomar cualquier valor entre 160 y 161 centímetros.
- **Discretas:** si el número posible de valores es finito o numerable (se puede contar). Por ejemplo el número de hijos de una pareja o el número de goles que marca un equipo en una temporada.

Una vez recogidos los datos, hay que organizarlos, en una tabla en la que aparezcan ordenados los distintos valores que toma la variable y el número de individuos que toma cada valor, es decir, la frecuencia absoluta de cada valor. Tenemos así una distribución de frecuencias o **distribución estadística**.

Para poder resumir la información se utilizan los parámetros de la población.

Un **parámetro** es un valor que se obtiene a partir todos los datos de una distribución estadística y que nos da una idea suficientemente clara de ella aunque desconozcamos sus datos. Un parámetro es una medida descriptiva de una población.

Sea X una variable estadística que toma los valores $x_1, x_2, \ldots\ldots x_n$ y sean ahora $f_1, f_2, \ldots\ldots f_n$ el número de veces que toma cada uno de ellos respectivamente.

Número de Individuos: $N = f_1 + f_2 + \ldots\ldots\ldots + f_n = \sum_{i=1}^{n} f_i$

Media Aritmética: es la suma de todos los valores obtenidos dividida entre el número total de datos N y se representa por $\overline{x}$

$$\overline{x} = \frac{x_1 . f_1 + x_2 . f_2 + \ldots\ldots\ldots + x_n . f_n}{f_1 + f_2 + \ldots\ldots\ldots + f_n} = \frac{\sum_{i=1}^{n} x_i . f_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} x_i . f_i}{N}$$

La media da un **valor numérico central** del conjunto de resultados obtenidos.

Varianza: es la media aritmética de los cuadrados de las desviaciones respecto de la media, siendo la desviación de un determinado valor respecto de la media, la diferencia entre este valor y la media aritmética obtenida. Se representa por Var(X), o por s^2. Se calcula mediante la fórmula:

$$s^2 = \frac{(x_1-\bar{x})^2 f_1 + \ldots\ldots\ldots + (x_n-\bar{x})^2 . f_n}{f_1 + \ldots\ldots\ldots + f_n} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 . f_i}{N} = \frac{\sum_{i=1}^{n} x_i^{\,2} . f_i}{N} - (\bar{x})^2$$

Desviación Típica: es la raíz cuadrada de la varianza, se representa por s.

$$s = \sqrt{VARIANZA} = \sqrt{s^2}\ .$$

La varianza y la desviación típica, complementan la información que nos da la media en el sentido de que si la media nos da el valor "normal" o "central", estas otras medidas nos dicen, de alguna forma, cuánto podemos desviarnos de aquella sin perder la característica de ser normales.

Ejemplo.-

Calcula la media y la desviación típica en esta distribución estadística: tiempo que emplean en ir de su casa al colegio un grupo de alumnos, dada por la siguiente tabla:

Tiempo (m)	$(0,5]$	$(5,10]$	$(10,15]$	$(15,20]$	$(20,25]$	$(25,30]$
Nº alumnos	2	11	13	6	3	1

Sol

Hallamos la marca de clase de cada intervalo y hacemos la tabla:

x_i	f_i	$x_i . f_i$	$x_i^{\,2} . f_i$
2,5	2	5	12,5
7,5	11	82,5	618,75
12,5	13	162,5	2031,25
17,5	6	105	1837,5
22,5	3	67,5	1518,75
27,5	1	27,5	756,25
	36	450	6775

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i . f_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} x_i . f_i}{N} = \frac{450}{36} = 12,5$$

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 . f_i}{N} = \frac{\sum_{i=1}^{n} x_i^{\,2} . f_i}{N} - (\bar{x})^2 = \frac{6775}{36} - 12,5^2 \approx 31,94$$

$$s = \sqrt{31,94} = 5,65$$

Distribuciones de probabilidad

Cuando se realiza un experimento aleatorio y se hace un recuento de frecuencias de cada suceso tenemos una distribución de frecuencias (o distribución estadística), si las frecuencias se sustituyen por sus probabilidades teóricas tenemos una distribución de probabilidad.

Una **distribución de probabilidad** es por tanto, un modelo matemático teórico que trata de explicar los resultados de un experimento aleatorio real. Este modelo permite asignar probabilidades a los distintos sucesos o realizar conjeturas sin necesidad de llevar a cabo el experimento. Dependiendo del tipo de fenómenos que explique, la distribución puede ser discreta o continua.

Para contar o medir estos resultados deben estar expresados en términos numéricos. Definimos para ello la **variable aleatoria** que puede definirse como una función que asigna a cada suceso elemental del espacio un número real.

Una **variable aleatoria** puede ser **discreta o continua.**

<u>Parámetros de una Variable Aleatoria Discreta</u> (nombrados con letras griegas)

La *Media o esperanza matemática* (nombrada con la letra: μ)

$$\mu = x_1.p_1 + x_2.p_2 + ... + x_n.p_n = \sum_{i=1}^{n} x_i.p_i$$

La *Varianza* (nombrada con la letra: σ^2)

$$\sigma^2 = (x_1 - \mu)^2.p_1 + ... + (x_n - \mu)^2.p_n = \sum_{i=1}^{n} (x_i - \mu)^2.p_i = \sum_{i=1}^{n} x_i^{\,2}.p_i - \mu^2$$

La *Desviación típica* (nombrada con la letra: σ)

$$\sigma = \sqrt{VARIANZA} = \sqrt{\sigma^2}$$

Ejemplo.- Consideremos la variable aleatoria X que cuenta el nº de caras al lanzar dos monedas.

Tiene por función de probabilidad la dada por la tabla siguiente:

$X = x_i$	$x_1 = 0$	$x_2 = 1$	$x_3 = 2$
$p(X = x_i) = p_i$	$p_1 = 1/4$	$p_2 = 1/2$	$p_3 = 1/4$

Calculamos los parámetros: media muestral y desviación típica:

x_i	p_i	$x_i.p_i$	$x_i^{\,2}.p_i$
0	$1/4$	0	0
1	$1/2$	$1/2$	$1/2$
2	$1/4$	$1/2$	$4/4 = 1$
TOTAL	1	1	$3/2$

$$\mu = 1$$

$$\sigma^2 = \frac{3}{2} - (1)^2 = \frac{3}{2} - \frac{2}{2} = \frac{1}{2}$$

$$\sigma = \sqrt{\frac{1}{2}} = \frac{\sqrt{2}}{2} \approx 0.7$$

DISTRIBUCIÓN BINOMIAL

Llamamos experiencia dicotómica aquella en la que sólo destacamos dos sucesos: A **(al que se suele llamar éxito) y su complementario** A' **.Si el suceso A tiene probabilidad** $p(A) = p$ **entonces el suceso complementario tendrá** $p(A') = q = 1 - p$ **.**

Son experiencias dicotómicas:

1.- Lanzar una moneda y ver la cara que ha salido:

$$A = \{C\} \quad A' = \{+\} \quad p(A) = \frac{1}{2} \quad p(A') = \frac{1}{2}$$

2.- Lanzar un dado y ver si sale o no el 3:

$$A = \{3\} \quad A' = \{1,2,4,5,6\} \quad p(A) = \frac{1}{6} \quad p(A') = \frac{5}{6}$$

3.- Extraer un naipe de una baraja española y ver si es figura:

$$A = \{FIGURA\} = \{AS, REY, CABALLO, SOTA\} \quad A' = \{2,3,4,5,6,7\} \quad P(A) = \frac{16}{40} \; P(A') = \frac{24}{40}$$

Si se repite una misma experiencia dicotómica "n" veces y si la variable X cuenta el número de veces que obtenemos éxito de entre las "n", la distribución de probabilidad que toma la variable X se le denomina BINOMIAL.

Se escribe: $X \in B(n, p)$ **de modo que la probabilidad de éxito** $p = p(A)$ **es siempre la misma en las "n" número de veces que se repite la experiencia dicotómica.**

Así, son experiencias binomiales:

1.- Lanzar 10 veces una moneda y contar las veces que ha salido cara.

Si X: "nº de veces que ha salido cara" $X \in B(10, \frac{1}{2})$

2.-Si extraemos un naipe de una baraja española observamos si es o no figura y la devolvemos al mazo. Barajamos y repetimos la experiencia 7 veces

Si X: "nº de veces que sale figura" $X \in B(7, \frac{16}{40})$

OBSERVACIÓN ***Si la extracción del naipe hubiera sido sin remplazamiento la distribución de X no sería binomial pues la probabilidad de éxito no sería siempre la misma en las 7 extracciones.***

FUNCIÓN DE PROBABILIDAD. PARÁMETROS

Si $X \in B(n, p)$ **y si "k" representa el número de éxitos de entre las "n" veces que se repite la experiencia dicotómica tendremos:**

$X = n^o \, éxitos$	0	…..	k	…..	n
$p(X = k)$	$\binom{n}{0} p^0 q^n = q^n$		$\binom{n}{k} p^k q^{n-k}$		$\binom{n}{n} p^n q^0 = p^n$

Los parámetros de una $X \in B(n, p)$ **son:**

$$\mu = n.p$$

$$\sigma = \sqrt{n.p.q}$$

Ejemplo.- En el experimento del lanzamiento de dos monedas y consideremos X: "número de caras".Pues bien en este caso $X \in B\left(2, \frac{1}{2}\right)$

$X = n^o\,éxitos$	0	1	2
$p(X=k)$	$\binom{2}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^2 = \frac{1}{4}$	$\binom{2}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^1 = \frac{1}{2}$	$\binom{2}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^0 = \frac{1}{4}$

$$\mu = n.p = 2.\frac{1}{2} = 1$$

$$\sigma = \sqrt{n.p.q} = \sqrt{2.\frac{1}{2}.\frac{1}{2}} = \sqrt{\frac{1}{2}} = \frac{\sqrt{2}}{2} \approx 0{,}7$$

AJUSTE DE UN CONJUNTO DE DATOS A UNA DISTRIBUCIÓN BINOMIAL

En ocasiones, conviene averiguar si una serie de datos obtenidos experimentalmente (distribución empírica) provienen de una distribución binomial (distribución teórica).

Veamos un ejemplo.-
Para probar la eficacia de una vacuna, se administró a 150 grupos de 4 personas con riesgo de contagio y los resultados obtenidos son los de la tabla:

Contagiados	Nº de grupos
0	30
1	62
2	46
3	10
4	2
	150

Averigua si los datos se ajustan a una Binomial y a continuación calcula la probabilidad de que en un grupo de 4 personas se contagien exactamente 3, y determina la probabilidad de que en un grupo de 4 personas se contagien menos de dos.
Sol
Nos interesa averiguar si la tabla de datos podría provenir de una variable

$$X = n^o \text{ de contagiados en un grupo de cuatro personas}$$

que se distribuye de modo binomial, es decir: B(4,p).
Para ello calculamos la media muestral (empírica) de los datos de la tabla:

x_i	f_i	$x_i.f_i$
0	30	0
1	62	62
2	46	92
3	10	30
4	2	8
Total	150	192

Como vemos la media muestral es:

$$\overline{x} = \frac{192}{150} = 1{,}28$$

En una binomial es: $\mu = n.p = 4.p$

$$\mu = \overline{x} \Rightarrow 4.p = 1{,}28 \Rightarrow p = 0{,}32$$

$$q = 1 - 0{,}32 = 0{,}68$$

Comparemos a continuación la distribución empírica con una distribución Binomial teórica $X \in B(4, p = 0{,}32)$ **.En esta distribución la variable X toma los valores:** $\{0,1,2,3,4\}$ **con probabilidades** $p_{i=}\binom{4}{i}p^i q^{4-i} \quad i \in \{0,1,2,3,4\}$.

El nº teórico de grupos de 4 personas en los que hay x_i **contagiados es:** $150.p_i$ **ya que al aplicar la Ley de los Grandes Números la probabilidad es la frecuencia relativa cuando el número de pruebas es grande;** $p_i = \frac{n^o\, teórico \quad hay \quad x_i}{150}$**).**

x_i	p_i	$150.p_i$	Nº teóricos	Nº observados	$\lvert diferencia \rvert$
0	$(0{,}68)^4 = 0{,}214$	32,1	32	30	2
1	$4.(0{,}32)(0{,}68)^3 = 0{,}4$	60	60	62	2
2	$6(0{,}32)^2(0{,}68)^2 = 0{,}28$	42	42	46	4
3	$4(0{,}32)^3(0{,}68) = 0{,}09$	13,5	14	10	4
4	$(0{,}32)^4 = 0{,}01$	1,5	2	2	0

Como las diferencias observadas son suficientemente pequeñas para suponer que el ajuste es bueno.

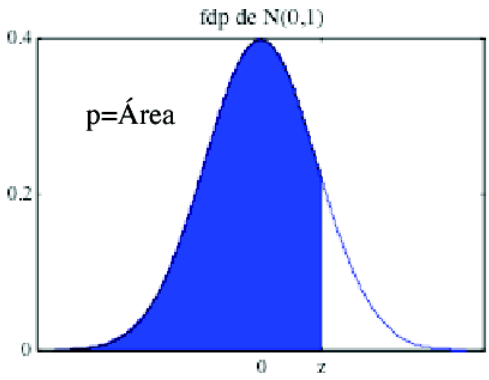
Observación.-existen métodos estadísticos para decidir de manera objetiva cuando las diferencias son suficientemente pequeñas (ahora lo decidimos a iojo!).

En consecuencia los datos provienen de una Binomial $X \in B(4, p = 0{,}32)$.
Ahora las respuestas pedidas son respectivamente:

$$p(X = 3) = \binom{4}{3}(0{,}32)^3(0{,}68) = 0{,}09 = 9\%$$

$$p(X = 0) + p(X = 1) = 0{,}214 + 0{,}4 = 0{,}614 = 61\%$$

Definición.-Se llama función de densidad a una función *f(x)* que nos permite hallar las probabilidades en las distribuciones continuas. Para que una función *f(x)* sea función de densidad debe ser positiva en su dominio de definición y el área limitada por su gráfica y el eje OX debe valer 1.



La probabilidad de que la variable aleatoria tome valores en un intervalo, es el área limitada por la gráfica y el eje OX en ese intervalo y su valor debe estar, por tanto, entre 0 y 1.

Parámetros de una Variable Aleatoria Continua

Los parámetros media, μ y desviación típica, σ tienen los mismos significados que en las distribuciones estadísticas, o sea:

a) *Media:* μ *centro de gravedad de la distribución.*

b) *Desviación típica:* σ *medida de la dispersión*.

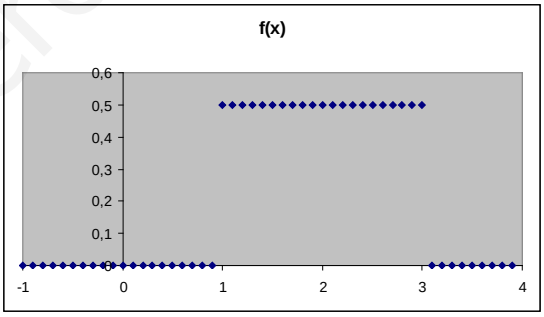
El cálculo exacto de estos parámetros requiere el dominio del cálculo integral (que está fuera de los objetivos de las matemáticas aplicadas) por ello, en este curso, y teniendo en cuenta a) y b) podremos estimarlos de forma aproximada.

$$(***)\left(\mu = \int_{-\infty}^{+\infty} x.f(x)\,dx \qquad \sigma^2 = \int_{-\infty}^{+\infty} x^2.f(x)\,dx - (\mu)^2 \right)$$

Ejemplo.- Consideremos la variable aleatoria continua con función de densidad

$$f(x) = \begin{cases} 0 & si \quad x < 1 \\ 1/2 & si \quad 1 \le x \le 3 \\ 0 & si \quad x > 3 \end{cases}$$

la función cuya gráfica es:



La función f(x) definida anteriormente en el intervalo [1,3] es una función de densidad pues verifica que

$$1.-\ f(x) \ge 0 \quad \forall x \quad y \quad 2.-\int_1^3 f(t)dt = área\ del\ rectángulo(\ base: 1\ a\ 3) = 2 * \frac{1}{2} = 1$$

Calculemos las siguientes probabilidades:

$$a)\ p(1 < X < 1{,}5) \quad b)\ p(1 < X < 2{,}5) \quad c)\ p(0 < X < 1)$$

$$a)\ p(1 < X < 1{,}5) = (1{,}5 - 1).0{,}5 = 0{,}25 \quad b)\ p(1 < X < 2{,}5) = 1{,}5.0{,}5 = 0{,}75 \quad c)\ p(0 < X < 1) = 1.0 = 0$$

Fácilmente se comprueba que la media es: μ =2(centro de gravedad de la distribución).

No es tan "fácil" determinar que la desviación típica es aproximadamente: σ =0,57

(Se necesita del cálculo integral (***), fuera del alcance del curso de CCSS).

Distribución normal.

Esta es, con seguridad, la más importante de las distribuciones de probabilidad.
El primero que la describe es Moivre (en 1733), pero no fue hasta cincuenta años más tarde cuando el matemático alemán Gauss, la redescubrió al estudiar los errores en las medidas. De ahí que también se le llama "**curva de errores**".

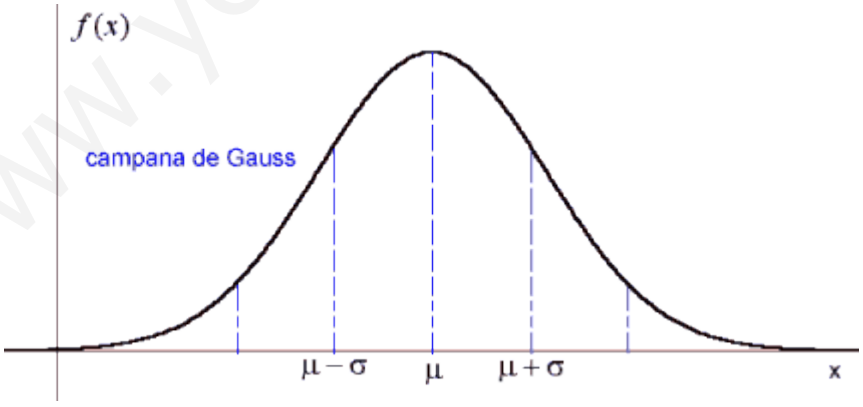
Su importancia no solo radica en el hecho de que ciertas características de los elementos de muchas poblaciones muestren una distribución de frecuencias prácticamente de tipo Normal así, por ejemplo:
1.-caracteres morfológicos (tallas, pesos…)
2.-caracteres fisiológicos (efectos de una misma dosis de un fármaco)
3.-caracteres sociológicos (consumo de ciertos productos por un mismo grupo humano)
En general:
(Cualquier característica que se obtenga como suma de muchos factores)

Sino que además,

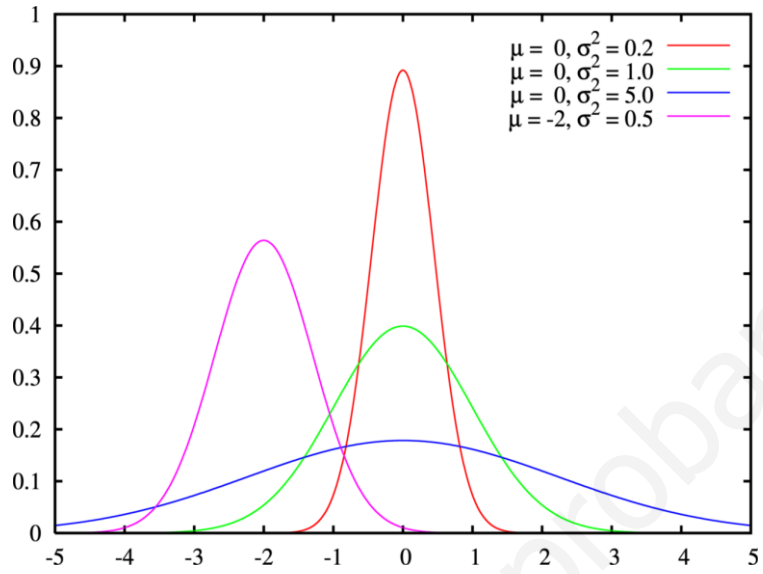
En virtud del llamado **Teorema Central del Límite,**
Si consideramos una muestra representativa de una población cualquiera, su media muestral tiene una distribución de frecuencias que se aproxima a la de una Normal a medida que aumenta el tamaño de la muestra.

Por último, presenta una propiedad muy importante desde el punto de vista práctico, y es que bajo ciertas condiciones, *para muchas distribuciones continuas y discretas* el cálculo de probabilidades puede llevarse a cabo a través de la distribución Normal.

A continuación puedes ver la gráfica de la función de densidad de una distribución Normal, que por su forma acampanada se le ha llamado: "Campana de Gauss".



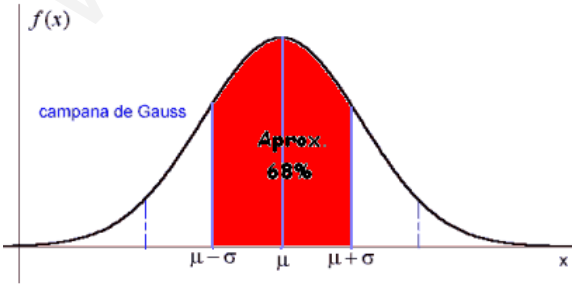
Las funciones de densidad de las variables normales constituyen una familia de curvas, en la cual cada una de ellas viene determinada por su media y su desviación típica. En las gráficas que tienes a continuación puedes ver la influencia de ambos parámetros: la desviación típica estrecha o ensancha la curva; la media la desplaza a la izquierda o a la derecha.



Las características más importantes de la curva normal son las siguientes:

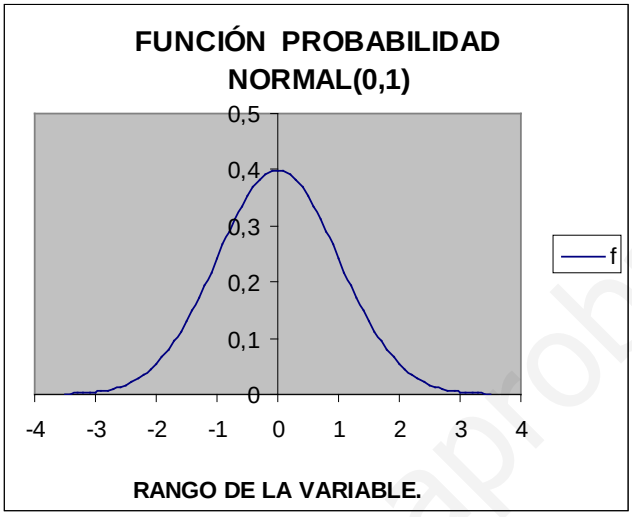
1. El dominio de la variable normal es todo $\Re$
2. *f(x)* es simétrica respecto a la media de la distribución μ
3. El máximo de *f(x)* se alcanza en $x = \mu$
4. Tiene dos puntos de inflexión con abscisas: $\mu - \sigma$ y $\mu + \sigma$.
5. El eje OX es una asíntota de *f(x)*

Para cada valor de la media μ y cada valor de la desviación típica σ hay una curva Normal (μ , σ) como puedes ver en la gráficas anteriores.
Sin embargo, el reparto de probabilidades es prácticamente el mismo en todas las distribuciones Normales, únicamente depende del valor de μ y de σ .
Así, el área comprendida bajo la curva entre los límites $\mu \pm \sigma$ es 0,6826; entre $\mu \pm 2\sigma$ es de 0,9544; y entre $\mu \pm 3\sigma$ es de 0,9974. Es decir:

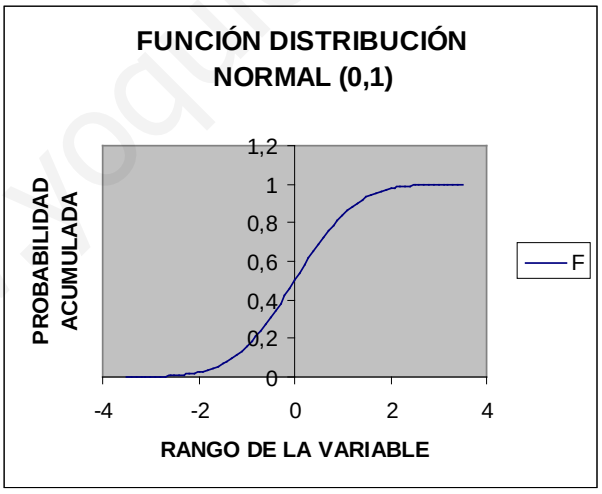


$p(\mu - \sigma \leq X \leq \mu + \sigma) = 0{,}6826$
$p(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0{,}9544$
$p(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0{,}9974$

La curva Normal usada con más frecuencia (por estar tabulada) es la llamada **Normal Estándar o N (0,1)**, se le llama así precisamente por tener como parámetros: media cero ($\mu = 0$), y desviación típica uno ($\sigma = 1$).
Su función de Densidad tiene por gráfica una curva de forma acampanada, continua, **simétrica respecto al eje Y**, con un máximo en la media $x = 0$ y dos puntos de inflexión con abscisas $x = -1, \quad x = 1$.



La función de Distribución de la N (0,1) es una función continua, creciente, y cuya gráfica puedes ver a continuación :



La distribución normal de media 0 y desviación típica 1, N (0, 1), se llama **distribución estándar o normal tipificada** y suele designarse por la **letra Z.**
En esta distribución los valores de las probabilidades para los distintos valores de Z, están tabulados, por lo que para conocerlos debemos aprender a manejar las tablas.
La tabla que viene a continuación nos da las probabilidades $p(z \leq k)$ para valores de k de 0 hasta 4 de centésima en centésima.

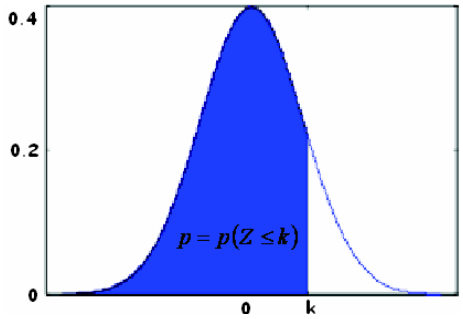


Tabla de áreas bajo la curva Normal estándar: N (0,1)

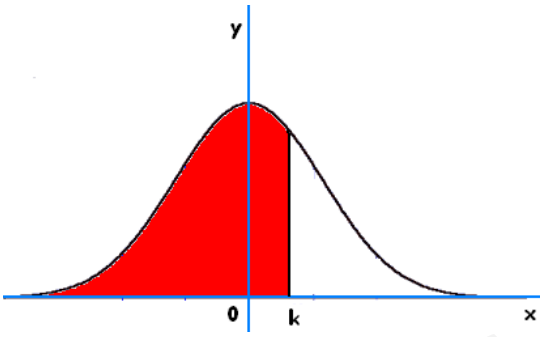
El valor de k se busca así:
unidades y décimas: columna de la izquierda,
las centésimas: en la fila superior.

$p(Z \leq k) = p = Área\ coloreada$

N(0,1)										
k	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Manejo de tablas

Vamos a ver como se calculan probabilidades utilizando la tabla de la Normal Estándar (tipificada) $Z \in N(0,1)$ **con algunos casos particulares.**
La tabla anterior da los valores de la probabilidad acumulada hasta el valor "k", es decir: $p(Z \le k)$



Por la simetría de la función de probabilidad de una Normal estándar tenemos:

a) $p(Z \le k) + p(Z \ge k) = 1$	**b)** $p(Z \ge -k) = p(Z \le k) \quad k > 0$
c) $p(Z \le -k) = p(Z \ge k) \quad k > 0$	**d)** $p(t \le Z \le k) = p(Z \le k) - p(Z \le t)$

Los ejemplos que tienes a continuación están resueltos manejando esa tabla.

1. **Probabilidad de que Z tome valores menores o iguales que 1,45**

$$p(Z \le 1'45) = 0'9265$$

2. **Probabilidad de que Z tome valores menores o iguales que -1,45**

$$p(Z \le -1'45) = p(Z > 1'45) = 1 - p(Z \le 1{,}45) = 0{,}0735$$

3. **Probabilidad de que Z tome valores entre 1´25 y 2'57**

$$p(1'25 \le Z \le 2'57) = p(Z \le 2'57) - p(Z \le 1'25) = 0'9949 - 0'8944 = 0'1005$$

4. **Probabilidad de que Z tome valores entre –2'57 y –1'25**

$$p(-2'57 \le Z \le -1'25) = (1'25 \le Z \le 2'57) = 0'1005$$

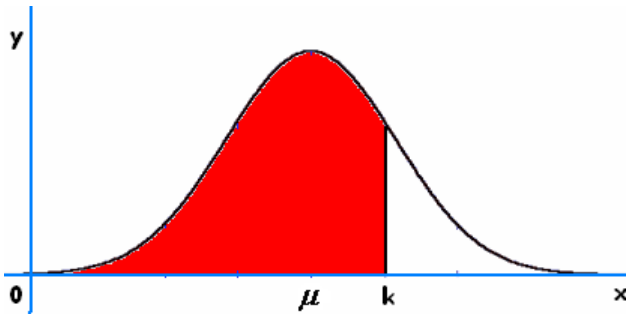
5. **Probabilidad de que Z tome valores entre –0'53 y 2'46**

$$p(-0{,}53 \le Z \le 2'46) = p(Z \le 2'46) - p(Z \le -0'53) =$$
$$= p(Z \le 2'46) - p(Z > 0'53) = p(Z \le 2'46) - (1 - p(Z \le 0'53))$$
$$= 0'9931 - (1 - 0'7019) = 0'695$$

Para calcular la probabilidad de una variable normal X: N (μ,σ) no tipificada, es decir, que no toma los valores $\mu = 0 \quad y \quad \sigma = 1$, **se transforma en una variable normal tipificada mediante el cambio:**

$$Z = \frac{X - \mu}{\sigma}$$

que sigue una distribución de media 0 y desviación típica 1 (tipificada): $Z \in N(0,1)$



En una Normal de media 6 y desviación típica 4, X: N (6,4) calcula:

6. Probabilidad de que X tome valores menores o iguales que 3

$$p(x \le 3) = p\left(z \le \frac{3-6}{4}\right) = p(z \le -0{,}75) = 1 - p(z \le 0{,}75) = 1 - 0{,}7734 = 0{,}2266$$

7. Probabilidad de que X tome valores entre 5 y 8.

$$p(5 \le x \le 8) = p\left(\frac{5-6}{4} \le z \le \frac{8-6}{4}\right) = p(-0{,}25 \le z \le 0{,}5) = 0{,}6915 - 1 + 0{,}5987 = 0{,}2902$$

Intervalos característicos

Para una variable X con distribución de media (μ) llamamos:

Intervalo Característico correspondiente a una probabilidad p, a un intervalo centrado en la media, $(\mu - k, \mu + k)$ tal que la probabilidad de que X pertenezca a dicho intervalo es $p(\mu - k < X < \mu + k) = p$

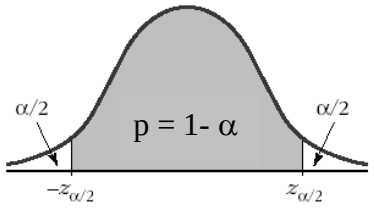
En una distribución normal estándar como $(\mu = 0)$ el intervalo característico para una probabilidad p tiene la forma: $(-k, k)$ En este caso decimos que "k" es el valor crítico correspondiente a una probabilidad "p".

Habitualmente se designa $p = 1 - \alpha$. El valor crítico correspondiente se escribe $z_{\alpha/2}$

Sin más que observar el gráfico que tienes a la derecha entenderás por qué y verás que se cumple:

a) $p\left(Z > z_{\alpha/2}\right) = \frac{\alpha}{2}$

b) $p\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1 - \alpha$



$1-\alpha$	α	$z_{\alpha/2}$
0.90	0.10	1.645
0.95	0.05	1.96
0.99	0.01	2.575

Los valores críticos para las probabilidades: 0.9; 0.95; 0.99 se utilizan frecuentemente y son: 1.645; 1.96; 2.575 respectivamente.

En una distribución $N(\mu, \sigma)$ cualquiera el intervalo característico correspondiente a una probabilidad $p = 1 - \alpha$ será pues: $\left(\mu - z_{\alpha/2}.\sigma, \mu + z_{\alpha/2}.\sigma\right)$

APROXIMACIÓN NORMAL DE LA BINOMIAL

Vamos a ver que bajo ciertas condiciones, una distribución Binomial $B(n, p)$ tiene un parecido extraordinario con la correspondiente distribución Normal.

$$Si \quad \begin{cases} n.p > 3 \\ n.q > 3 \end{cases} \Rightarrow \begin{cases} aproximación \\ aceptablemente \\ buena \end{cases} \qquad Si \quad \begin{cases} n.p > 5 \\ n.q > 5 \end{cases} \Rightarrow \begin{cases} aproximación \\ casi \\ perfecta \end{cases}$$

En estas condiciones, la binomial $B(n, p)$ se aproxima a una normal con su misma media y desviación típica, o sea a una $N(n.p, \sqrt{n.p.q})$ (teorema de De Moivre).

En esta similitud hay que tener en cuenta que la Binomial es discreta y la Normal es Continua, por ello para calcular las probabilidades de la Binomial por aproximación de la Normal se procede con la siguiente regla práctica (corrección de continuidad):

$$Si \quad X \in B(n, p) \quad y \quad X' \quad su \quad aproximación \quad normal, \quad X' \in N(n.p, \sqrt{n.p.q}), \quad tenemos:$$
$$p(X = k) = p(k - 0{,}5 < X' < k + 0{,}5)$$

Es decir, para cada valor puntual de X (0, 1, 2,3,……., n) se le asocia a la variable X' un intervalo de centro k y longitud 1.

Ejemplo 1.-

Una moneda se lanza 400 veces. Calcula la probabilidad de que el número de caras:
a) sea mayor que 200. b) esté entre 180 y 220.

Sol.

Si X es la variable aleatoria que cuenta el número de caras, tendremos que X es una Binomial de parámetros B(400, 0,5) y en consecuencia X' es N(200, 10).

a) $p(X > 200) = p(X' \geq 200 + 0{,}5) = p(z \geq (\frac{200{,}5 - 200}{10}) = p(z \geq 0{,}05) = 0{,}4801.$

b) $p(180 < X < 220) = p(180 + 0{,}5 \leq X' \leq 220 - 0{,}5) = p(\frac{180{,}5 - 200}{10} \leq z \leq \frac{219{,}5 - 200}{10}) =$

$= p(-1{,}95 \leq z \leq 1{,}95) = 0{,}9488$

Ejemplo 2.-

En un bombo de lotería tenemos 10 bolas idénticas numeradas de 0 a 9, y cada vez que hacemos la extracción de una bola la devolvemos al bombo.
a) Si sacamos 3 bolas, calcula la probabilidad de que el cero salga una sola vez.
b) Si hacemos 100 extracciones, ¿probabilidad de que el cero salga más de 12 veces?

Sol.

a) X: "Número de la bola extraída", X es una binomial de parámetros: B (3, 0.1)

$p(X = 1) = 3.(0.1).(0.9)^2 = 0{,}243$

b) X: Número de la bola extraída, X es binomial de parámetros: B (100,0.1), se puede aproximar por una Normal de parámetros: N (10, 3) (teor. de De Moivre)

$p(X > 12) = p(X' \geq 12{,}5) = p(z \geq 0{,}83) = 0{,}2033$

AJUSTE DE UN CONJUNTO DE DATOS A UNA NORMAL

En ocasiones, conviene averiguar si una serie de datos obtenidos experimentalmente (distribución empírica) provienen de una distribución normal (distribución teórica).
Veamos un ejemplo.-
La tabla dada a continuación corresponde a las estaturas de 1400 chicas. Estudia si es aceptable considerar que provienen de una distribución Normal.

x_i	141	146	151	156	161	166	171	176	181
f_i	2	25	146	327	428	314	124	29	5

x_i	f_i	$x_i.f_i$	$x_i^2.f_i$
141	2	282	39762
146	25	3650	532900
151	146	22046	3328946
156	327	51012	7957872
161	428	68908	11094188
166	314	52124	8652584
171	124	21204	3625884
176	29	5104	898304
181	5	905	163805
T	1400	225235	36294245

Calculamos la media muestral $\bar{x} = \frac{225235}{1400} \approx 160{,}9$

y calculamos también la desviación típica muestral:

$$s = \sqrt{\frac{36294245}{1400} - (160{,}882)^2} = \sqrt{41{,}39} = 6{,}43$$

Hemos de comparar, pues, la distribución empírica con una Normal : N(160.9;6.43).Para ello:

*partimos en intervalos:[138.5,143.5]……,

y vemos como se distribuirán 1400 individuos de una teórica N(160.5;6.43) en esos intervalos.

*comparamos esa distribución con la empírica y evaluamos las diferencias.

Formamos, para hacer todo lo dicho anteriormente, la siguiente tabla:

extr. interv x_k	*extr. tipif.* z_k	$p(z \le z_k)$	p_k $p(z_k < z < z_{k+1})$	$1400.p_k$	*números teóricos*	*números obtenidos*	$\lvert dif. \rvert$
138,5	-3,48	0.0003					
143,5	-2,71	0.0034	0.0031	4.34	4	2	2
148,5	-1,93	0.0268	0.0234	32.76	33	25	8
153,5	-1,15	0.1251	0.0983	137.62	138	146	8
158,5	-0,37	0.3557	0.2306	322.84	323	327	4
163,5	0,41	0.6591	0.3034	424.76	425	428	3
168,5	1,18	0.8810	0.2219	310.66	311	314	3
173,5	1,96	0.9750	0.0940	131.60	132	124	8
178,5	2,74	0.9969	0.0219	30.66	31	29	2
183,5	3,51	0.9998	0.0029	4.06	4	5	1

La mayor de las diferencias es 8, es suficientemente pequeña como para aceptar que los datos empíricos provienen de una distribución Normal, por lo que las discrepancias son atribuibles al azar.

Observación.-existen métodos estadísticos para decidir de manera objetiva cuando las diferencias son suficientemente pequeñas (ahora lo decidimos ia ojo!)

LAS MUESTRAS ESTADÍSTICAS

Población y Muestra

Se llama población o universo, al conjunto de todos los individuos objeto de nuestro estudio.

Se llama muestra de una población a un subconjunto de la población. Su estudio sirve para inferir características de toda la población.

¿Por qué se recurre a las muestras? Los motivos por los que se recurre a las muestras pueden ser:

- **La población es excesivamente numerosa: los gallegos que están viendo un determinado programa de televisión.**
- **La población es muy difícil o imposible de controlar: La totalidad de personas que entran en unos grandes almacenes en una semana.**
- **El proceso de medición es destructivo: si deseamos conocer la duración media de las bombillas que fabrica una empresa.**
- **Se desean conocer rápidamente ciertos datos de una población y se tardaría demasiado en consultar a todos: los sondeos electorales o de opinión.**

Hay dos aspectos en las muestras al que les debemos prestar mucha atención:

- **el tamaño: con muestras relativamente pequeñas se consiguen "aproximaciones" sorprendentemente buenas de la realidad poblacional. En el tema: Inferencia estadística, aprenderemos a calcular con exactitud el tamaño "n" que debe tener una muestra para conseguir lo que nos propongamos.**
- **cómo se realiza la selección de los individuos que forman la muestra. Es importante tener en cuenta que no todas las muestras son buenas, de ahí la importancia de este aspecto.**

Al sustituir el estudio de la población por el de las muestras se producen errores (pero con ellos contamos de antemano y pueden controlarse) sin embargo, si la muestra está mal elegida (no es representativa), se producen errores adicionales imprevistos e incontrolables (sesgos).

La elección de la muestra se llama muestreo. Para que un muestreo nos proporcione una muestra representativa debe ser aleatorio, es decir, <u>todos los individuos de la muestra deben elegirse al azar</u>, de modo que todos los individuos de la población tengan, *a priori*, la misma probabilidad de ser elegidos.

Veamos a continuación algunos de los tipos o métodos de muestreo más frecuentes:

Muestreo aleatorio simple

Es el tipo de muestreo más sencillo y en él se basan todos los demás. Para obtener una muestra de tamaño n, se numeran los elementos de la población y se seleccionan al azar los n elementos que debe contener la muestra.

Muestreo aleatorio sistemático

Se numeran los individuos y, a partir de uno de ellos elegido al azar, se toman los siguientes mediante saltos numéricos iguales. Al salto se llama coeficiente de elevación.

El proceso es el que sigue:

1. **Se calcula el coeficiente de elevación, h, dividiendo el número de individuos de la población N entre el tamaño de la muestra n.**
2. **Se averigua el primer elemento de la muestra, a_1, obteniéndolo aleatoriamente (sorteándolo) de entre los h primeros.**
3. **Se obtienen los restantes elementos de la muestra: $a_2 = a_1 + h$, $a_3 = a_2 + h$, $a_4 = a_3 + h$,....**

Muestreo aleatorio estratificado

Si la población puede dividirse en estratos (por ejemplo, por edades), a veces conviene elegir la muestra fijando de antemano el número de individuos de cada estrato. Este tipo de muestreo se utiliza cuando se supone que la pertenencia a uno u otro estrato influyen en la variable que estamos analizando.

Técnicas para obtener una muestra aleatoria de una población finita.

Hemos dicho que para obtener una muestra aleatoria "se sortean" los individuos de la población para decidir al azar cuales de ellos forman parte de la muestra. El "sorteo" puede hacerse de varias formas. Supongamos una población de N individuos y queremos escoger una muestra de n, podemos hacerlo:

- **Mediante extracción(o insaculación): Introduciendo en una caja bolas o papeletas previamente numeradas de 1 a N y escogiendo al azar n de ellas. Se puede hacer:**
  - **<u>Sin devolución</u>: cogiendo simultáneamente o bien de una en una las n papeletas.**
  - **<u>Con devolución</u>: Después de cada extracción se devuelve la papeleta a la caja. En este caso se podría obtener un individuo repetido con lo cual hay que devolver la papeleta a la caja y volver a extraer de nuevo.**
- **Mediante la obtención de números aleatorios: La calculadora tiene una tecla $RAN^{\#}$ que da de forma aleatoria números con tres cifras decimales comprendidos entre 0,000 y 0,999. Si la población es de menos de 1000 individuos, multiplicando N por el número que nos salga obtenemos un número entre 0 y (N-1).Por tanto si tomamos la parte entera del número obtenido por la secuencia: $N \times RAN^{\#} + 1 =$ obtenemos un número al azar entre 1 y N. Si la muestra es de más de 1000 habrá que obtener los números aleatorios con el ordenador.**

Realizado el estudio estadístico sobre los individuos de una muestra de tamaño n, debemos dar las conclusiones con un nivel de confianza y con un límite máximo de error.

El nivel de confianza es la probabilidad que tenemos de que cada vez que hagamos una estimación (en el ejemplo de más abajo sobre el grado de aceptación del alcalde), en las mismas condiciones en que hayamos hecho el estudio, obtendríamos los mismos resultados.

El límite máximo de error nos marca el margen de error (superior e inferior) sobre el cual se hace dicha estimación.

Cuando nos dan los resultados de una encuesta, siempre deben darnos la ficha técnica del estudio, que consiste en dar las características de la población, la muestra y el grado de fiabilidad.

Ejemplo: **Leemos en un periódico que el 40% de la población de un municipio tiene una opinión favorable sobre el alcalde, con la siguiente ficha técnica:**

Población: Personas del municipio mayores de 18 años

Muestra: 1.200 individuos

Tipo de muestreo: Aleatorio mediante entrevistas personales siguiendo un método estratificado por parroquias.

Límite máximo de error: Se estima en ± 2%

Nivel de Confianza: 90%

Interpretación: ***La población son los ciudadanos con derecho a voto. El tipo de muestreo es lógico ya que la opinión puede ser muy diferente de una parroquia a otra. El límite máximo de error nos dice que entre un 38% (40-2) y un 42% (40+2) de ciudadanos mayores de 18 años tiene una opinión favorable sobre el alcalde y que esta afirmación la hacemos con una probabilidad del 90%***

Ejercicio 1.- El colectivo estudiado, ¿es población o muestra?

a) En unas elecciones municipales se escrutan las papeletas de las votaciones. (P)

b) En unos grandes almacenes, para indagar sobre la eficacia de un empleado, se le pregunta a los clientes que salen por una de las puertas durante el día. (M)

c) En unos grandes almacenes, para indagar sobre la eficacia de un empleado, se le pregunta a todos los clientes atendidos por éste durante su primer día de trabajo. (P)

Ejercicio 2.-Explica por qué, en cada uno de los siguientes casos es imprescindible recurrir a una muestra.

a) En un almacén hay 4200 vasos de vidrio. Se quiere estudiar su resistencia a la rotura. Para hacerlo, se someten a presiones crecientes hasta que rompen.

Como el proceso es destructivo, es imprescindible recurrir a una muestra, y además ha de ser tan pequeña como sea posible (pero procurando que se puedan extraer de su estudio conclusiones fiables).

b) Para estudiar el tiempo de reacción de ciertas sustancias, se hacen reaccionar en 25 ocasiones, tomando medidas en cada una de ellas.

En cualquier experiencia suponemos controladas todas las variables que intervienen (presión, temperatura, cantidades,…). Sin embargo, el control de las variables no es perfecto y cada experimento puede dar lugar a un resultado distinto (aunque serán muy parecidos). La población es pues infinita, es natural pues recurrir a una muestra.

c) Un profesor, para comprobar si sus explicaciones han sido comprendidas, realiza a sus alumnos varias preguntas.

Las preguntas que realiza son una muestra que sirve para tantear lo que saben sus alumnos. Incluso de los exámenes se extrae una muestra de los conocimientos de los alumnos pues resulta imposible preguntárselo todo.

Ejercicio 3.-Disponemos de un censo electoral de 27800 electores. Deseamos extraer una muestra de 200 individuos.

a) ¿Cómo se debe realizar mediante muestreo aleatorio sistemático?

b) ¿Cómo se debe realizar mediante muestreo aleatorio simple?

Si de la población anterior sabemos que el 20% tienen entre 18 y 25 años; el 35% entre 26 y 40 años; y el 45% más de 40 años.

c) ¿Cómo se extraería una muestra de 200 individuos con estratos proporcionales a esos porcentajes?

Sol.-

a) Coeficiente de elevación: $h = \frac{27800}{200} = 139$ **lo que significa que tenemos que seleccionar un individuo de cada 139. Para saber por cuál empezamos, elegimos al azar un número de 1 a 139. Se puede realizar mediante la función** $RAN^{\#}$ **de la calculadora. Así si obtuviésemos** $RAN^{\#}$ **0.534x139+1=75.782 el primer elemento sería: 75 (parte entera) y los siguientes: 75+139=214; 214+139=353;……**

b) La secuencia $RAN^{\#}$ **x27800+1= (su parte entera) nos da un individuo al azar entre los del colectivo inicial. Esta secuencia se repetiría 200 veces para seleccionar los 200 individuos de la muestra. Si aparece algún elemento repetido se suprime y se obtiene otro en su lugar.**

Como el número inicial (27800) es mayor de 1000 individuos, los números aleatorios los obtendríamos con ordenador para que tengan, al menos, cinco cifras decimales.

c) 20% de 400 = 40; 35% de 400 =70 y 45% de 400 = 90.

Se eligen al azar 40 individuos de entre los que tienen entre 18 y 25 años, 70 individuos de entre los que tienen de 26 años a 40 y 90 individuos de más de 40años.

Inferencia estadística: Estimación de la media.

Estimación de una proporción. Test de hipótesis.

La Estadística Inferencial tiene por objeto el desarrollo de técnicas que permiten conocer o comprobar el valor de los parámetros de una población a partir de los datos obtenidos de una muestra.
La estadística Inferencial tiene dos grandes ramas:
A) Estadística Inductiva B) Estadística Hipotético-deductiva (o Teoría de la Decisión)
La primera tiene por objeto "estimar" los parámetros de la población.
La segunda tiene por objeto "comprobar" (mediante métodos matemáticos) hipótesis realizadas sobre el valor de un parámetro de la población a partir de una muestra extraída de ella.

Uno de los problemas más sencillos de la Estadística Inductiva es el de estimar el valor de la media de una población a partir de una muestra.
La estimación se realiza de forma aproximada (mediante un intervalo) y con una cierta seguridad (asignando un "nivel de confianza" al resultado). El tamaño de la muestra influye, como es obvio, en la "finura" de la estimación.
Para realizar esto se recurre a los llamados estadísticos muestrales y a una herramienta básica de la inferencia estadística: "la distribución Normal"

Estadísticos muestrales: media muestral y cuasivarianza muestral.

Como los parámetros poblacionales resultan difíciles de obtener directamente, se recurre a los estadísticos para estimarlos.

En una población sobre la que hay establecido un mecanismo para obtener muestras aleatorias se llama estadístico a cualquier variable aleatoria sobre el conjunto de posibles muestras. Los estadísticos más importantes son la media muestral y la cuasivarianza muestral.

Dada una muestra aleatoria de tamaño n procedente de una población normal con media μ y desviación típica σ, el valor medio de las n observaciones se llama media muestral y viene dada por:

$$\overline{X} = \frac{x_1 + x_2 + ... + x_n}{n}$$

Más adelante veremos como se distribuye el estadístico media muestral (T.C.L.).

Se llama cuasivarianza muestral y se representa por S^2_{n-1} a:

$$s^2_{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

La cuasivarianza muestral se utiliza para estimar la desviación típica de la población a partir de la de la muestra, aunque para muestras grandes ($n \geq 30$) se puede tomar como desviación típica de la población, la de la propia muestra.

Estimación de la media.-

Distribución de probabilidad de la media muestral.

Teorema Central del Límite

De una población formada por N elementos con media μ y desviación típica σ se pueden obtener diferentes muestras de tamaño n.

Cada muestra tendrá una media y una desviación típica.

Sea $\overline{X}$ la variable aleatoria que toma como valores las medias de las muestras.

Se puede demostrar que:

1. La media de las medias de las muestras (de tamaño n) es igual a la media real de la población, es decir: si $\overline{x}_1$ es la media de la muestra X_1, $\overline{x}_2$ es la media de la muestra X_2, ... $\overline{x}_i$ es la media de X_i, entonces:

   $\overline{x} = \dfrac{\overline{x}_1 + \overline{x}_2 + ... + \overline{x}_i + ...}{n^o \text{ de posibles muestras}} = \mu$ siendo μ la media de la población.

2. La distribución de las medias muestrales, $\overline{X}$, tiene una desviación típica que es igual a $\dfrac{\sigma}{\sqrt{n}}$ siendo σ la desviación típica de la población

3. Si n ≥ 30, la distribución de las medias muestrales es normal incluso en el caso de que estas procedan de poblaciones no normales.

Este resultado se conoce con el nombre de **teorema central del límite** que establece que:

Si una muestra aleatoria de tamaño n procede de una población con media μ y desviación típica σ, entonces en caso de que el tamaño de la muestra sea lo suficientemente grande (n ≥ 30), la media muestral $\overline{X}$ tiene aproximadamente una distribución normal de media μ y desviación típica $\dfrac{\sigma}{\sqrt{n}}$. Esto es, la distribución de las medias muestrales $\overline{X}$, es una distribución normal $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$

Es importante señalar que este teorema es válido cualquiera que sea la distribución de la población de partida, tanto si es discreta como continua.

Si la población de partida es normal, también lo será la distribución de las medias muestrales, cualquiera que sea el valor de *n*. Si la población de partida no es normal, la distribución de las medias muestrales puede ser muy parecida a la normal, incluso para valores pequeños de *n*, pero para *n* ≥ 30 es seguro que se consigue una gran aproximación a la normal cualquiera que sea la distribución de partida.

Consecuencias del Teorema Central del Límite.-

I) Control de las medias muestrales: podemos tratar de conocer la probabilidad de que la media de una muestra concreta, de tamaño n por ejemplo, esté en un cierto intervalo.

II) Control de la suma de todos los individuos de la muestra: por el teorema central del límite:

$\overline{X} \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ **.Como:** $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} \Rightarrow \sum_{i=1}^{n} x_i = n.\overline{X}$ **en consecuencia la distribución de probabilidad para la suma de todos los individuos de la muestra de tamaño n es:** $\sum_{i=1}^{n} x_i \in N\left(n\mu, \sigma\sqrt{n}\right)$ **y en consecuencia podemos tratar de conocer la probabilidad de que la suma de los elementos de una muestra esté en un cierto intervalo.**

III) Inferir la media de la población a partir de la de una muestra: Esta es la aplicación más importante del Teorema Central del Límite y lo veremos en el siguiente apartado.

Ejemplo 1.-
Los pesos en kilogramos de los soldados de una quinta siguen una distribución normal: N (69,8). Las guardias en un regimiento están formadas por 12 soldados.

a) Determina la probabilidad de que la media de los pesos de los soldados de una guardia sea superior a 71 kilos.
b) Obtén el intervalo característico para $\overline{x}$ correspondiente a una probabilidad de 0,9.
c) ¿Cuál es la probabilidad de que la suma de los pesos de los soldados de una guardia sea mayor que 800 kilos?
d) ¿Cuál es la probabilidad de que un miembro de la guardia, elegido al azar, pese más de 93 k?

Sol.
Suponemos que las guardias en el regimiento se forman tomando 12 soldados al azar (O por un proceso similar como por ejemplo: "orden alfabética").

Sabemos que si la población de los pesos de soldados es N (69,8) entonces el peso medio de las muestras de tamaño 12 es: $N(69, \frac{8}{\sqrt{12}}) = N(69; 2{,}31)$
(Aunque n<30 pero la población de partida es normal).

a) $p(\overline{x} > 71) = p(z > \frac{71-69}{2{,}31}) = p(z > 0{,}87) = 1 - p(z < 0{,}87) = 1 - 0{,}8078 = 0{,}1922$

b) sabemos que el valor crítico correspondiente a una probabilidad p =0,9 en una normal: N (0,1) es 1,645 (ya está tabulado) entonces, como también sabemos, el intervalo característico para una normal N (μ,σ) es: $\left(\mu - \sigma * z_{\alpha/2}, \mu + \sigma * z_{\alpha/2}\right)$

y por tanto en este caso para $\overline{x}$ **el intervalo es:**

$$\left(69 - 2.31*1.645,\ 69 + 2.31*1.645\right) = \left(65.20,\ 72.79\right)$$

Esto significa que el 90% de las guardias tienen un peso medio entre: 65,20 y 72,79

c) Sabemos del punto II) de las consecuencias del teorema central del Límite que:

$$\sum_{i=1}^{n} x_i \in N\left(n\mu, \sigma\sqrt{n}\right)$$

En este caso $n.\mu = 12.69 = 828 \qquad \sigma\sqrt{n} = 8.\sqrt{12} = 27{,}72$ **por tanto:**

$$\sum_{i=1}^{n} x_i \in N\left(828;27{,}2\right)$$

$$p(\sum_{i=1}^{n} x_i < 800) = p(z < \frac{800-828}{27{,}2}) = p(z < -1{,}01) = p(z > 1{,}01) = 1 - p(z < 1{,}01) = 0{,}1562$$

d) Un individuo tomado al azar de un grupo, que también fue elegido al azar, es una extracción al azar de un miembro de la población. Es decir, para calcular la probabilidad pedida, usamos la distribución de la población de pesos de los soldados que es N (69; 8).

$$p(x > 93) = p(z > \frac{93-69}{8}) = p(z > 3) = 1 - p(z < 3) = 0{,}0013$$

Esto significa que "aproximadamente" un uno por mil de los soldados de una guardia pesa más de 93 k.

Veamos cómo se aplica el Teorema Central del Límite para el caso III).

III) Inferir la media de la población a partir de la de una muestra.-

Intervalo de confianza de la media de la población, con σ conocida

En el ejemplo anterior (de los pesos de los soldados) hemos visto como de la aplicación directa del Teorema Central del Límite podemos deducir el comportamiento de las muestras a partir del conocimiento de la población.

Ahora pretendemos deducir aspectos de la población a partir del conocimiento de una muestra. En concreto, pretendemos inferir el valor de la media de la población a partir del conocimiento de la media de una muestra.

Parece razonable estimar que la media de la población: μ **será aproximadamente igual que la media de la muestra:** $\overline{x}$ **pero, ¿cómo de aproximadamente? Para determinar el grado de aproximación de** $\overline{x}$ **a** μ **procedemos a una estimación mediante intervalos.**

Queremos estimar el valor de la media, μ de una población de la que conocemos su desviación típica poblacional σ

Para ello se recurre a una muestra de tamaño n de la que podemos calcular su media muestral $\overline{x}$

Si la población de partida es normal o el tamaño de la muestra es mayor o igual de 30, el teorema central del límite nos asegura que la distribución de las medias muestrales $\overline{X}$ tiene aproximadamente una distribución normal de media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$.

Es decir, la distribución de probabilidad de la media muestral $\overline{X}$ es $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

El intervalo característico para $\overline{x}$ correspondiente a una probabilidad $p = 1-\alpha$ según hemos visto es:

$$\left(\mu - z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}, \mu + z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}\right)$$

$$Es \quad decir \quad p\left(\mu - z_{\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}} < \overline{x} < \mu + z_{\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \Rightarrow p\left(|\overline{x} - \mu| < z_{\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \quad o\ sea:$$

$$p\left(\overline{x} - z_{\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Por tanto, una vez extraída la muestra de tamaño n y calculada su media $\overline{x}$ el intervalo:

$$\left(\overline{x} - z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}\right)$$

contendrá o no a la media poblacional μ con un nivel de confianza: $1-\alpha$

Este intervalo se llama en este caso intervalo de confianza de μ.

La probabilidad de que el intervalo de confianza contenga a la media es $1-\alpha$ y habitualmente se da en porcentaje ($1-\alpha$).100% y se llama nivel de confianza.

El valor de α es el nivel de significación y determina el riesgo de que la media de la población no esté en dicho intervalo.

Si, como suele ocurrir, la desviación típica σ de la población cuya media se quiere estimar es desconocida, se puede calcular σ a partir de de la muestra.

La forma más correcta de hacerlo es mediante el estadístico cuasivarianza muestral $s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ (tomaríamos $\sigma = \sqrt{s_{n-1}^2}$)

Sin embargo para valores relativamente grandes de n ($n \geq 30$) podemos utilizar la desviación típica de la muestra s como la de la población, es decir, $\sigma = s$

Ejemplo 2.-
Se desea hacer un estudio de mercado para conocer el precio medio de los libros científicos. Para ello, se elige una muestra aleatoria formada por 34 libros, y se determina que la media muestral es de 21€ con desviación típica de 3€. Halla el intervalo de confianza para el precio medio de los libros científicos al nivel de confianza de un 99%.
Sol.-
No se conoce la desviación típica poblacional σ, pero como la muestra seleccionada contiene más de 30 individuos, se puede sustituir por la muestral $s = 3$
Por tanto el intervalo de confianza para el precio medio de los libros científicos, correspondiente al nivel de confianza del 99%, y nivel de significación (o riesgo) $\alpha = 1 - 0.99 = 0.01$ es: $\left(21 - z_{\alpha/2} * \frac{s}{\sqrt{34}}, 21 + z_{\alpha/2} * \frac{s}{\sqrt{34}}\right)$

Sustituyendo los valores conocidos es: $\left(21 - z_{0.005} * \frac{3}{\sqrt{34}}, 21 + z_{0.005} * \frac{3}{\sqrt{34}}\right)$

Puesto que $z_{0.005} \approx 2.58$ el intervalo de confianza es: $(19.67\ ,\ 22.33)$

Ejemplo 3.-
a) Un ganadero de reses bravas quiere estimar el peso medio de los toros de su ganadería con un nivel de confianza del 95%. Para eso, toma una muestra de 30 toros y los pesa. Obtiene una media de $\bar{x} = 507\ k$ y una desviación típica de $\sigma = 32\ k$
¿Cuál es el intervalo de confianza para la media μ de la población?
b) ¿Cuál será el intervalo si queremos que el nivel de confianza sea del 99%?
Sol.-
a) No se conoce la desviación típica poblacional σ, pero como la muestra seleccionada contiene 30 individuos, se puede sustituir por la muestral $s = 32$
Por tanto el intervalo de confianza para el peso medio de los toros, correspondiente al nivel de confianza del 95%, y nivel de riesgo $\alpha = 1 - 0.95 = 0.05$ es:

$$\left(507 - z_{\alpha/2} * \frac{s}{\sqrt{34}}, 507 + z_{\alpha/2} * \frac{s}{\sqrt{34}}\right) = \left(507 - z_{0.025} * \frac{3}{\sqrt{34}}, 507 + z_{0.025} * \frac{3}{\sqrt{34}}\right)$$

Puesto que $z_{0.025} = 1.96$ el intervalo de confianza es: $(495.55\ ,\ 518.45)$

b) Puesto que $z_{0.005} = 2.575$ el intervalo de confianza es: $(492\ ,\ 522)$

Ejemplo 4.-

Deseamos valorar el grado de conocimientos en historia de una población de varios miles de alumnos. Sabemos que la desviación típica es σ =2,3. Nos proponemos estimar μ pasando una prueba a 100 alumnos.

a) Calcula el intervalo característico para $\overline{x}$ correspondiente a una probabilidad de 0,95.

Una vez realizada la prueba a 100 alumnos concretos, se obtuvo una media $\overline{x}$ = 6,32.

b) Obtén el intervalo de confianza para μ con un nivel de confianza del 95%.

Sol.

La población en estudio tiene una distribución con desviación típica conocida: $\sigma = 2.3$ y media desconocida: μ Por tanto, como las muestras son de tamaño 100 (mayor que 30), la distribución de la media muestral $\overline{x}$ es $N(\mu;\frac{\sigma}{\sqrt{n}}) = N(\mu,\frac{2,3}{\sqrt{100}}) = N(\mu;0,23)$

a) el intervalo característico para $\overline{x}$ con una probabilidad de 0,95 cumple:

$$p\left(\overline{x} \in (\mu - \sigma.z_{\frac{\alpha}{2}};\eta + \sigma.z_{\frac{\alpha}{2}})\right) = 0,95 \Rightarrow p\left(\overline{x} \in (\mu - 0,23.z_{\frac{\alpha}{2}};\mu + 0,23z_{\frac{\alpha}{2}})\right) = 0,95$$

Como para una probabilidad de 0,95 sabemos que

$$1-\alpha = 0,95 \Rightarrow \alpha = 1-0,95 = 0,05 \Rightarrow \frac{\alpha}{2} = 0,025$$ **y por tanto:**

$$z_{\frac{\alpha}{2}} = 1.96$$ **(Según las tablas)**

$$p\left(\overline{x} \in (\mu - 0,23.1,96;\mu + 0,23.1,96)\right) = 0,95 \Rightarrow p\left(\overline{x} \in (\mu - 0,45;\mu + 0,45)\right) = 0,95$$

El intervalo característico para $\overline{x}$ correspondiente a una probabilidad de 0,95 es

$$(\mu - 0,45;\eta + 0,45)$$

Por tanto en el 95 % de las muestras su media $\overline{x}$ dista de μ menos de 0,45.

b) el intervalo aleatorio de confianza para μ (con un nivel de confianza del 95%) es:

$$\left(\overline{x} - 0,45;\overline{x} + 0,45\right),$$

lo que significa que en el 95% de las posibles muestras el intervalo correspondiente contiene a μ.

En este caso concreto, con $\overline{x} = 6.32$ el intervalo de confianza (no aleatorio) es:

$$(5,87;6,77)$$

ya que: {$\left(\overline{x} - 0,45;\overline{x} + 0,45\right) = (6,32 - 0,45;6,32 + 0,45) = (5,87;6,77)$**}**

¿Está μ en este intervalo: $(5,87;6,77)$? No lo podemos saber con seguridad pero tenemos una confianza del 95% de que es así.

Error en la estimación.

Determinación del tamaño de la muestra. **Determinación del nivel de confianza**.

Se llama error máximo admisible al valor $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ **(que es el radio del intervalo de confianza). Este valor depende de** α **y de *n* de la siguiente forma:**

- **Cuanto mayor sea el tamaño de la muestra, menor es *E*, con lo cual será también menor la amplitud del intervalo de confianza y afinaremos más en la estimación.**
- **Cuanto mayor sea** $1-\alpha$ **es decir, cuanto más seguros queramos estar de nuestra estimación, mayor será *E* (como se observa a continuación)**

En la tabla

$1-\alpha$	α	$z_{\alpha/2}$
0.90	0.10	1.645
0.95	0.05	1.96
0.99	0.01	2.575

están los niveles de confianza que se usan con más frecuencia. Como puedes ver cuanto mayor es $1-\alpha$ **, mayor es** $z_{\alpha/2}$ **y, por lo tanto, mayor es *E*.**

Para un error máximo admisible determinado, *E*, y un nivel de confianza, $1-\alpha$ **el mínimo tamaño que debe tener la muestra se obtiene despejando *n* en la expresión**

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow \sqrt{n} = \frac{z_{\alpha/2} \cdot \sigma}{E} \Rightarrow n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2$$

Podemos ver que:

- **El tamaño de la muestra es mayor cuanto mayor sea** $z_{\alpha/2}$ **o lo que es lo mismo, cuanto menor sea** α **y mayor** $1-\alpha$ **Para aumentar el nivel de confianza debemos aumentar el tamaño de la muestra.**
- **El tamaño de la muestra es mayor cuanto menor se *E*, por lo que para ser más precisos tenemos que aumentar el tamaño de la muestra.**

Para un error máximo admisible, *E*, y un tamaño de la muestra, *n*, el nivel de confianza con el que se realiza la estimación se obtiene despejando $z_{\alpha/2}$ **de**

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow z_{\alpha/2} = \frac{E\sqrt{n}}{\sigma}$$

Conociendo $z_{\alpha/2}$ **podemos buscar en las tablas de la normal estándar el valor de** $\alpha/2$ **y a partir de él el nivel de confianza** $1-\alpha$

Ejemplo 5.-

Sabemos de la duración de un proceso que $\sigma = 0{,}5\ s$ ¿Cuál es el número de medidas que hay que realizar para que, con un 99% de confianza, el error de estimación no exceda de 0,1 s?

Sol.

Como el error de estimación viene dado por: $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ **y si el nivel de confianza es: 99%** $\Rightarrow 1-\alpha = 0{,}99 \Rightarrow \alpha = 1-0{,}99 = 0{,}01 \Rightarrow z_{\frac{\alpha}{2}} = 2{,}575$

Ahora tendremos:

$$0{,}1 = 2{,}575 \cdot \frac{0{,}5}{\sqrt{n}} \Rightarrow \sqrt{n} = \frac{2{,}575.0{,}5}{0{,}1} = \frac{1{,}2875}{0{,}1} = 12{,}875$$

$$n = (12{,}875)^2 = 165{,}76$$

Deben realizarse, por tanto, 166 medidas (el menor entero mayor que 165,76)

Ejemplo 6.-

Al medir el tiempo de reacción, un psicólogo sabe que la desviación típica del mismo es de 0,5 s. Desea estimar el tiempo medio de reacción con un error máximo de 0,1 s, para lo cual realiza 100 experiencias.

¿Con qué nivel de confianza podrá dar el intervalo: $(\overline{x} - 0{,}1; \overline{x} + 0{,}1)$**?**

Sol

Como el error de estimación viene dado por: $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$$0{,}1 = z_{\alpha/2} \cdot \frac{0{,}5}{\sqrt{100}} \Rightarrow z_{\frac{\alpha}{2}} = \frac{0{,}1.10}{0{,}5} = 2$$

Sabemos que

$$p\left(z < z_{\frac{\alpha}{2}}\right) = p(z < 2) = 0{,}9772$$

$$\frac{\alpha}{2} = p(z > 2) = 1 - p(z < 2) = 1 - 0{,}9772 = 0{,}0228$$

$$\alpha = 2\frac{\alpha}{2} = 2.0{,}0228 = 0{,}0456 \Rightarrow 1-\alpha = 1-0{,}0456 = 0{,}9544$$

Si $\overline{x}$ es el tiempo medio obtenido con las 100 experiencias, se puede asegurar con un nivel de confianza del 95,44% que el tiempo medio de reacción está comprendido entre: $\overline{x}$ -0,1 y $\overline{x}$ +0,1 s.

Ejercicio 7.-

Lanzamos 36 dados correctos y calculamos la media de sus resultados, $\bar{x}$ si repetimos de forma reiterada esta experiencia

a) ¿Cuál será la distribución de las medias?

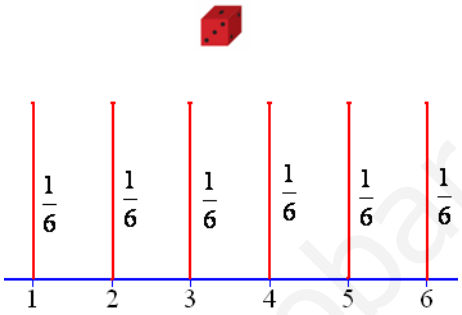
b) ¿Cuál es la probabilidad de que la media de una de las tiradas sea mayor que 4?

c) Calcula un intervalo centrado en la media en el que se encuentre el 99% de las medias de los lanzamientos.

Sol.-

a)

x_i	p_i	$x_i.p_i$	$x_i^2.p_i$
1	1/6	1/6	1/6
2	1/6	2/6	4/6
3	1/6	3/6	9/6
4	1/6	4/6	16/6
5	1/6	5/6	25/6
6	1/6	6/6	36/6
	1	3.5	91/6



Al lanzar un dado correcto la media de las puntuaciones es: $\mu = \sum_{i=1}^{6} x_i.p_i = 3.5$ **y**

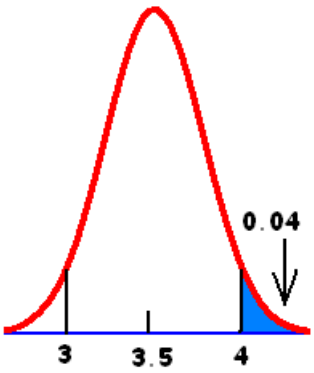
La desviación típica: $\sigma = \sqrt{\sum_{i=1}^{6} x_i^2.p_i - \mu^2} = 1.71$

Si tomamos una muestra de esta población de tamaño n=36 (n>30) sabemos que:

$$\bar{x} \in N\left(3.5\,,\,\frac{1.71}{\sqrt{36}}\right) = N(3.5\,,\,0.285)$$

b) $p(\bar{x} > 4) = p\left(z > \frac{4-3.5}{0.285}\right) = 0.0401$

Solamente en el 4% de los casos la media de las puntuaciones de los 36 dados es mayor que 4.



c) El intervalo característico de las medias muestrales para $1-\alpha = 0.99$ **es:**

$$\left(\mu - z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}, \mu + z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}\right) = (3.5 - 2.575 * 0.285\,,\,3.5 + 2.575 * 0.285) = (2.77\,,\,4.23)$$

El 99% de las puntuaciones medias de los 36 dados están dentro de ese intervalo.

Ejercicio 8.-

Para estimar la media de los resultados que obtendrían al resolver un cierto test los alumnos de 2º Bach de la comunidad autónoma de Galicia, se les pasa dicho test a 400 de esos alumnos escogidos al azar.

Los resultados obtenidos vienen dados en la tabla:

x_i	1	2	3	4	5
f_i	24	80	132	101	63

A partir de estos datos, estima con un nivel de confianza del 95% el valor de la media de la población.

Sol

x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
1	24	24	24
2	80	160	320
3	132	396	1188
4	101	404	1616
5	63	315	1575
	400	1299	4723

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{N} = \frac{1299}{400} \approx 3.25$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 \cdot f_i}{N} - (\bar{x})^2 = \frac{4723}{400} - (3.25)^2 \approx 1.245$$

$$s = \sqrt{1.245} \approx 1.12$$

Con un nivel de confianza al 95% $1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow z_{\alpha/2} = 1.96$

El intervalo de confianza al 95% es: $\left(\bar{x} - z_{\alpha/2} * \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} * \frac{s}{\sqrt{n}} \right) = (3.14\ ,\ 3.36)$

Ejercicio 9.-

Para estimar el peso medio de las niñas de 16 años de una ciudad, se toma una muestra aleatoria de 100 de ellas. Se obtienen los parámetros $\bar{x} = 52.5\ k$; $s = 5.3\ k$

Se afirma que: el peso medio de las niñas de 16 años de esa ciudad está entre 51 y 54 kilos. ¿Con qué nivel de confianza se hace tal afirmación?

Sol.-

En la fórmula del Error máximo admisible: $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ **sabemos que:**

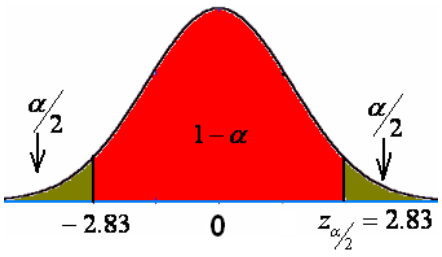
E **es la mitad de la longitud del intervalo:** $E = \frac{54-51}{2} = 1.5$; $\quad n = 100$

$$\sigma = s = 5.3 \Rightarrow z_{\alpha/2} = \frac{E\sqrt{n}}{\sigma} = \frac{1.5 * \sqrt{100}}{5.3} = 2.83$$

$$\frac{\alpha}{2} = p\left(z > z_{\frac{\alpha}{2}} \right) = p(z > 2.83) =$$

$$= 1 - p(z < 2.83) = 1 - 0.9977 = 0.0023$$

$$\alpha = 2\frac{\alpha}{2} = 2 * 0.0023 = 0.0046 \Rightarrow 1 - \alpha = 0.9954$$



El peso medio está entre 51 y 54 kilos a un nivel de confianza del 99.54%.

Estimación de una proporción.-

En los apartados anteriores aprendimos a estimar la media de una población a partir de la media muestral, con ayuda de la distribución Normal.
Cabe ahora señalar que esto sólo es posible cuando la muestra de la que se dispone es suficientemente grande. Para muestras pequeñas la curva normal falla y hay que recurrir a otra distribución (que no estudiaremos este curso) llamada t de Student (seudónimo de Gosset, químico de la fábrica de cervezas Guinness, que la inventó)

Pues bien, el segundo problema de inferencia que se estudia en este curso es el de la estimación del parámetro: proporción de individuos de una población que posee una cierta cualidad o lo que es lo mismo la probabilidad de que ocurra un cierto suceso.
En efecto, si la probabilidad de un suceso es "p" es lo mismo que decir que en la población que resulta de repetir esa experiencia aleatoria una infinidad de veces, la proporción de ocasiones en las que se da el suceso es "p". De ahí que una probabilidad pueda ser considerada como una proporción en una población infinita.

La estimación del parámetro proporción poblacional la realizaremos a través de la correspondiente proporción muestral.
Las probabilidades de los distintos valores de una proporción (variable discreta) se calculan con la ayuda de una distribución Binomial (número de individuos con una cierta cualidad), y ésta a su vez puede ser sustituida, en ciertos casos (Tª de De Moivre), por una Normal (Var. Continua) haciendo las necesarias correcciones de continuidad.
Pero, para no complicar el proceso, trataremos a la variable proporción como continua.

Distribución de probabilidad de las proporciones muestrales.

A continuación vamos a dar respuesta a la siguiente situación:
En una cierta población, la proporción de individuos que posee una cierta cualidad es "p".Consideramos todas las posibles muestras de tamaño "n" que se pueden extraer de esa población. En cada una de las muestras habrá una proporción "pr" de individuos con esa cualidad. ¿Cómo se distribuyen todos los posibles valores de "pr"?
Vamos a formalizarlo con la ayuda de un ejemplo concreto:

Una máquina produce tornillos. Se sabe que el 5% de ellos son defectuosos.
Los tornillos se empaquetan en cajas de 400.
¿Cómo se distribuye la proporción de tornillos defectuosos en las cajas de 400?

En la muestra hay 400 individuos, si X es la variable que cuenta el número de defectuosos en las cajas de 400, sabemos que X es binomial $B(n = 400\ ,\ p = 0.05)$
Sabemos que X la podemos aproximar (con corrección de continuidad) por una Normal Y: $N(np\ ,\ \sqrt{npq}) = N(20\ ,\ 4.36)$ **para simplificar el cálculo identificamos X con Y.**
Por tanto X: el número de defectuosos en una muestra de 400, es X: $N(20\ ,\ 4.36)$

La proporción de defectuosos "pr" en una muestra de 400 se obtiene:

$$pr = \frac{nº\ defectuosos\ muestra}{400} = \frac{X}{400} \quad \Rightarrow pr : N\left(\frac{20}{400}\ ,\ \frac{4.36}{400}\right) = N(0.05\ \ 0.011)$$

Si en una población la proporción de individuos que posee una cierta cualidad es p la proporción pr de individuos con dicha cualidad en una muestra de tamaño n sigue una distribución normal de media p y desviación típica $\sqrt{\frac{pq}{n}}$ es decir,

$$pr : N\left(p, \sqrt{\frac{pq}{n}} \right)$$

Siguiendo con el problema con el que iniciamos el planteamiento, podemos calcular el intervalo característico en el que se encuentre el 90% de las proporciones de los tornillos defectuosos: $1-\alpha = 0.9 \Rightarrow \frac{\alpha}{2} = 0.05 \Rightarrow z_{\alpha/2} = 1.645$ **y** $pr : N(0.05, 0.011)$

Intervalo característico: $(0.05 - 1.645 * 0.011 \, , \, 0.05 + 1.645 * 0.011) = (0.032 \, , \, 0.068)$

Esto quiere decir que el 90% de las cajas de 400 tornillos tiene una proporción de tornillos defectuosos comprendida entre 0.032 y 0.068.

Conociendo la distribución de la proporción muestral $pr : N\left(p, \sqrt{\frac{pq}{n}} \right)$ **pretendemos**

Inferir la proporción de la población a partir de la de una muestra.-

O sea, pretendemos estimar el valor de la proporción p, de individuos con una cierta cualidad que hay en una población. Para esto recurrimos a una muestra de tamaño n, en la que se obtiene una proporción muestral pr.

El intervalo de confianza de p con un nivel de confianza de $(1-\alpha).100\%$ es:

$$\left(pr - z_{\alpha/2} * \sqrt{\frac{pr(1-pr)}{n}} \, , \, pr + z_{\alpha/2} * \sqrt{\frac{pr(1-pr)}{n}} \right)$$

Como $pr : N\left(p, \sqrt{\frac{pq}{n}} \right)$ **el intervalo característico para una probabilidad** $(1-\alpha)$ **es:**

$\left(p - z_{\alpha/2} * \sqrt{\frac{pq}{n}} \, , \, p + z_{\alpha/2} * \sqrt{\frac{pq}{n}} \right)$ **O sea** $p\left(p - z_{\alpha/2}\sqrt{\frac{pq}{n}} < pr < p + z_{\alpha/2}\sqrt{\frac{pq}{n}} \right) = 1-\alpha$

Por tanto: $p\left(pr - z_{\alpha/2}\sqrt{\frac{pq}{n}} < p < \; pr + z_{\alpha/2}\sqrt{\frac{pq}{n}} \right) = 1-\alpha$ **El error máximo admisible:**

$E = z_{\alpha/2}\sqrt{\frac{pq}{n}}$ **tiene el grave inconveniente de que está dado en función de p pero para muestras grandes** $(n \geq 30)$ **estimamos** $p = pr \, ; q = 1 - pr \Rightarrow \; E = z_{\alpha/2}\sqrt{\frac{pr(1-pr)}{n}}$ **y el intervalo de confianza de** p **:** $\left(pr - z_{\alpha/2} * \sqrt{\frac{pr(1-pr)}{n}} \, , \, pr + z_{\alpha/2} * \sqrt{\frac{pr(1-pr)}{n}} \right)$

Ejemplo 10.-
Tomada una muestra de 300 personas mayores de 15 años en una gran ciudad, se encontró que 104 de ellas leían algún periódico regularmente. Hallar, con un nivel de confianza del 90%, un intervalo de confianza para estimar la proporción de lectores de periódicos entre los mayores de 15 años.
Sol.-

Proporción muestral: $pr = \frac{104}{300} = 0.347 \qquad 1-\alpha = 0.9 \Rightarrow \frac{\alpha}{2} = 0.05 \Rightarrow z_{\alpha/2} = 1.645$

Error máximo admisible (cota de error): $E = 1.645 \sqrt{\frac{0.347\ 0.653}{300}} = 0.045$

Intervalo de confianza (al 90%): $(0.347 - 0.045\ ,\ 0.347 + 0.045) = (0.302\ ,\ 0.392)$

<u>Conclusión</u>: Con un nivel de confianza al 90%, la proporción de lectores, en el colectivo de los mayores de 15 años, está entre 0.302 y 0.392.

Error en la estimación.

<u>Determinación del tamaño de la muestra</u>. <u>Determinación del nivel de confianza</u>.

El error máximo admisible o cota de error: $E = z_{\alpha/2} \sqrt{\frac{pr(1-pr)}{n}}$

(Su valor numérico es el radio del intervalo de confianza)

En el ejemplo que viene a continuación se enseña a determinar el tamaño de una muestra para conseguir una estimación de una proporción con un error máximo admisible y un nivel de confianza dados (<u>teniendo un valor estimado para</u> pr **)**

Ejemplo 11.-
A la vista del resultado del ejemplo 10, se pretende repetir la experiencia para conseguir una cota de error de 0.01 con el mismo nivel de confianza del 90%. ¿Cuántos individuos deben tener la muestra?
Sol.-

En la expresión de la cota de error $E = z_{\alpha/2} \sqrt{\frac{pr(1-pr)}{n}}$ **hemos de determinar el valor de** n**. Para ello conocemos los valores de** $z_{\alpha/2}$ **y** E **ya que según el enunciado:**

$$E = 0.01 \ ; \qquad 1-\alpha = 0.9 \Rightarrow \frac{\alpha}{2} = 0.05 \Rightarrow z_{\alpha/2} = 1.645$$

El valor de pr **no podemos obtenerlo a partir de la muestra, pues aún no se ha extraído (de echo estamos calculando el tamaño que debe tener la muestra). Tomaremos como valor estimado de** $pr = 0.347$**; mismo valor que en el ejemplo 10.**

$$0.01 = 1.645 \sqrt{\frac{0.347 * 0.653}{n}} \Rightarrow \qquad n = \left(\frac{1.645}{0.01}\right)^2 * 0.347 * 0.653 = 6131.6$$

<u>Conclusión</u>.- Habrá que tomar una muestra de n=6132 personas.
Si ahora en esa muestra de 6132 individuos calculamos la proporción y resulta (pr_0**) podremos formar el intervalo** $(pr_0 - 0.01\ ,\ pr_0 + 0.01)$ **dentro del cual se estima, con una confianza del 90%, que estará la proporción poblacional** p

En el ejemplo que viene a continuación se enseña a determinar el nivel de confianza para conseguir una estimación de una proporción con un error máximo admisible y un tamaño de muestra dados (teniendo un valor estimado para pr)

Ejemplo 12.-
A partir de una muestra de 100 individuos, se ha estimado una proporción mediante el intervalo de confianza: $(0.17\ ,\ 0.25)$
¿Cuál es el nivel de confianza con el que se ha hecho la estimación?
Sol.-

pr es el punto medio del intervalo de confianza $\Rightarrow pr = \dfrac{0.17+0.25}{2} = 0.21$

E es la mitad de la longitud de dicho intervalo $\Rightarrow E = \dfrac{0.25-0.17}{2} = 0.04$

Como: $E = z_{\alpha/2}\sqrt{\dfrac{pr(1-pr)}{n}} \Rightarrow 0.04 = z_{\alpha/2}\sqrt{\dfrac{0.21*0.79}{100}} \Rightarrow z_{\alpha/2} = 0.98$

$$p\left(z < z_{\frac{\alpha}{2}}\right) = p(z<0.98) = 0.8365 \Rightarrow \frac{\alpha}{2} = p(z>0.98) = 1-p(z<0.98) = 0.1635$$

$$\alpha = 2\frac{\alpha}{2} = 2*0.1635 = 0.3270 \Rightarrow 1-\alpha = 1-0.3270 = 0.6730$$

Conclusión.-La estimación de p mediante el intervalo $(0.17\ ,\ 0.25)$ se ha hecho a un nivel de confianza del 67.30%

En el ejemplo que viene a continuación se enseña a determinar el tamaño de una muestra para conseguir una estimación de una proporción con un error máximo admisible y un nivel de confianza dados (sin tener un valor estimado para pr)

Ejemplo 13.-
¿Cuál es el tamaño muestral que deberíamos tomar para que la proporción estimada no difiera de la verdadera en más de un 5% (E=0.05) a un nivel de confianza del 95%?
Sol.-

En la expresión del error máximo admisible: $E = z_{\alpha/2}\sqrt{\dfrac{pr(1-pr)}{n}}$ **conocemos:**

$$E = 0.05\ ;\ 1-\alpha = 0.95 \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow z_{\alpha/2} = 1.96 \Rightarrow 0.05 = 1.96\sqrt{\frac{pr(1-pr)}{n}}$$

No tenemos una estimación de pr ; pero podemos demostrar con cálculo derivadas que $pr(1-pr)$ es menor o igual que $\dfrac{1}{4}$ ($pr(1-pr)$ toma su valor máximo en $pr = 0.5$)

$$0.05 = 1.96\sqrt{\frac{pr(1-pr)}{n}} \le 1.96\sqrt{\frac{0.5*0.5}{n}} = 1.96\sqrt{\frac{1}{4n}} \Rightarrow n \ge \left(\frac{1.96}{2*0.05}\right)^2 = 384.16$$

En general, el tamaño de la muestra n sin tener estimado pr es: $n \ge \left(\dfrac{z_{\alpha/2}}{2.E}\right)^2$

Ejemplo 14.-

Para estimar el número de peces que hay en un pantano, se procede de este modo: Se pescan una cierta cantidad de ellos, 349, se marcan (con tinta indeleble) y se devuelven al pantano. Al cabo de varios días, se vuelven a pescar otro montón de ellos y se averigua que proporción están marcados. Si en esta segunda pesca se han capturado 514 peces, de los cuales hay 37 marcados.

a) Halla un intervalo de confianza, al 90% para la proporción de peces marcados.

b) Halla un intervalo de confianza al 90% para el total de peces en el pantano.

Sol.- a) $pr = \frac{37}{514} = 0.072$; $1-\alpha = 0.9 \Rightarrow z_{\alpha/2} = 1.645$; $E = z_{\alpha/2}\sqrt{\frac{pr(1-pr)}{n}}$

$E = 1.645\sqrt{\frac{0.072*0.928}{514}} = 0.019$ **El intervalo de confianza para la proporción** p **en la población de peces marcados es:** $(0.072 \pm 0.019) = (0.053\ ,\ 0.091)$

b) Para hallar el intervalo de confianza para el número total (N) de peces en el pantano, tenemos en cuenta que la proporción de peces marcados es: $p = \frac{349}{N}$

$$\left.\begin{array}{l} 0.053 = \frac{349}{N_1} \\ 0.091 = \frac{349}{N_2} \end{array}\right\} \Rightarrow \left\{\begin{array}{l} N_1 = \frac{349}{0.053} \approx 6585 \\ N_2 = \frac{349}{0.091} \approx 3835 \end{array}\right\}$$ **(Observa que se invierten los extremos)**

A nivel de confianza del 90% el <u>intervalo para el total de peces</u> $(3835\ ,\ 6585)$

Ejemplo 15.-

a) Se ha lanzado una moneda 100 veces obteniéndose 62 caras. Estima la probabilidad de "cara" mediante un intervalo, con nivel de confianza del 90%, 95% y 99%.

b) ¿Basándonos en el apartado anterior cuántas veces habremos de lanzar la moneda para estimar la probabilidad de "cara" con un error menor que 0.002 y un nivel de confianza del 95%?

Sol

a) Población infinita del lanzamiento de una moneda para la que hemos de estimar la proporción p **del suceso "cara". Muestra de 100 para la que:** $pr = \frac{62}{100} = 0.62$

$E = z_{\alpha/2}\sqrt{\frac{0.62*0.38}{100}} = z_{\alpha/2} * 0.0485$; **Intervalo confianza es:** $\left(0.62 \pm z_{\alpha/2} * 0.0485\right)$

Al 90% es: $(0.54\ ,\ 0.70)$ **al 95% es:** $(0.525\ ,\ 0.715)$ **y al 99% es:** $(0.495\ ,\ 0.745)$

b) Basándonos en el apartado anterior estimamos $pr = 0.62$ **y** n **se obtiene de:**

$$n = \frac{z^2_{\alpha/2}.pr.(1-pr)}{E^2} = \frac{(1.96)^2 * 0.62 * 0.38}{(0.002)^2} = 226270.24 \Rightarrow\ n = 226271$$

Si en esa muestra de 226271 lanzamientos calculamos la proporción de caras y resulta (pr_0**) podremos formar el intervalo** $(pr_0 - 0.002\ ,\ pr_0 + 0.002)$ **dentro del cual se estima, con una confianza del 95%, que estará la probabilidad de cara** p

Test de hipótesis.-

En 1710 el médico inglés John Arbuthnot estudió el sexo de las criaturas nacidas en una cierta localidad durante los 82 años anteriores y advirtió que la proporción de hombres fue siempre superior a la de mujeres. Con ello rebatió la creencia de que es igualmente probable que nazca un hombre o una mujer, argumentando del modo:
"El resultado no puede ser casual, ya que, haciendo corresponder Hombre y Mujer a Cara y Cruz de una moneda, es absurdo pensar que exista tal exceso de hombres".
Aunque planteado de forma matemáticamente insuficiente, puede considerarse este el primer test de hipótesis de la Historia.

En un **Test de hipótesis** se emite una afirmación estadística (relativa al valor de un parámetro de una población) y mediante una muestra se estudia si dicha afirmación (**hipótesis**) es compatible con el resultado de la experiencia (**contraste**).

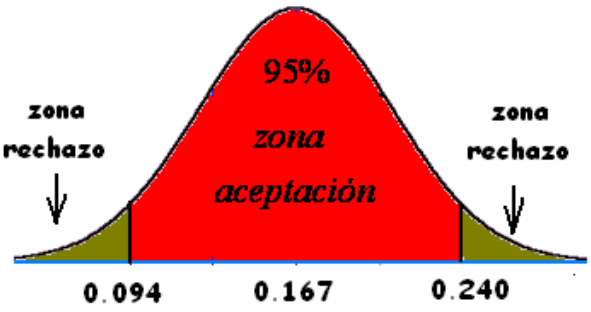
Vamos a formalizar el estudio de los test de hipótesis mediante un caso concreto:
Tenemos un dado que suponemos correcto. Lo lanzamos 100 veces y obtenemos 25 "cincos". ¿Podemos rechazar, o no, la afirmación de que el dado es correcto?
En este ejemplo se duda sobre si el parámetro $p = p(5)$ toma el valor de $1/6$
La hipótesis emitida (el dado es correcto) de este test se formaliza así:

$$p = p(5) = 1/6 = 0.167$$

Si la hipótesis fuera cierta, entonces las proporciones, pr de "cincos" en las muestras de tamaño 100 seguirían (según hemos visto anteriormente) una distribución Normal:

$$N\left(p, \sqrt{\frac{p*q}{n}}\right) = N\left(0.167\,,\, \sqrt{\frac{0.167*0.833}{100}}\right) = N(0.167\,,\,0.037)$$

En tal caso, el 95% de las proporciones muestrales de "cincos" estarían en el intervalo característico correspondiente a $1-\alpha = 0.95$: $(0.167 \pm 1.96*0.037) = (0.094\,,\,0.240)$
A este intervalo $(0.094\,,\,0.240)$ se le llama **Zona de aceptación.**
Como la proporción obtenida en la muestra de 100 tiradas para el suceso "cinco" es $pr = \frac{25}{100} = 0.25$ queda fuera de la zona de aceptación, consideramos que sería muy improbable que con un dado correcto se obtengan 25 "cincos" en 100 tiradas.
Por tanto se **rechaza la hipótesis** con un **nivel de significación** del 5%.



El **nivel de significación** (α) de una hipótesis es el valor complementario del **nivel de confianza** ($1-\alpha$) de una estimación.

Pasos para efectuar un test de hipótesis.-

1º) Enunciar la hipótesis nula H_0: Consiste en atribuirle un valor a un parámetro de cierta población. (En este curso lo haremos sobre la media y sobre la proporción).

2º) Si la hipótesis nula fuera cierta, tal parámetro de una cierta muestra se distribuirá de forma conocida:

a) se elige un nivel de significación α. Es habitual: 0.10; 0.05; 0.01.

b) se construye la zona de no rechazo: es el intervalo fuera del cual solo se encuentran el α.100% de los casos "más raros".

3º) Se extrae una muestra cuyo tamaño ya se ha decidido en el paso anterior y de ella se obtiene el valor del parámetro.

4º) Si el valor del parámetro muestral cae dentro de la zona de aceptación, no se rechaza la hipótesis con un nivel de significación α En caso contrario se rechaza.

Test de Hipótesis para la media.-

Vamos a poner en práctica los pasos enunciados anteriormente para trabajar un test de hipótesis sobre la media de una población, distinguiendo los casos en que la hipótesis nula: $H_0 : \mu = \mu_0$ (Bilateral) del caso $H_0 : \mu \leq \mu_0$ o $H_0 : \mu \geq \mu_0$ (Unilateral)

<u>Ejemplo 16</u>.- (Bilateral)

El estudio de una muestra aleatoria de 100 jóvenes que se presentan a una prueba, revela que la media de edad es de 20.2 años. Sabiendo que la variable estudiada se distribuye normalmente en la población con desviación típica $\sigma = 10$

¿Se puede aceptar con un 95% de confianza el valor de 22 años como media de edad de todos los que concurren a la prueba? Sol.-

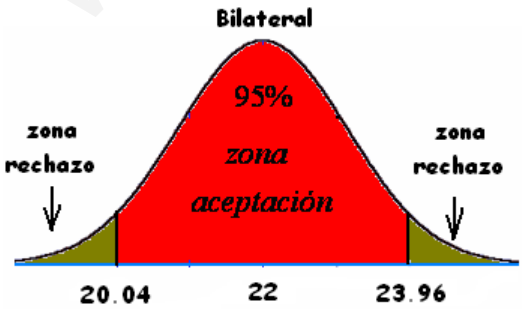
1º) $H_0 : \mu = 22$ frente a la hipótesis alternativa (la contraria) $H_1 : \mu \neq 22$

2º) Si $H_0 : \mu = 22$ es cierta; entonces las medias de edad $\overline{X}$ es $N(22, 1)$

Lo que significa que el 95% de las medias muestrales están en el intervalo característico: $(22 \pm 1.96 * 1) = (20.04, 23.96)$ zona de no rechazo

3º) Según el enunciado $\overline{X} = 20.2$

4º) Como $\overline{X} = 20.2$ cae dentro de la zona de no rechazo $(20.04, 23.96)$ concluimos que no rechazamos la hipótesis nula con un nivel de significación del 5%.

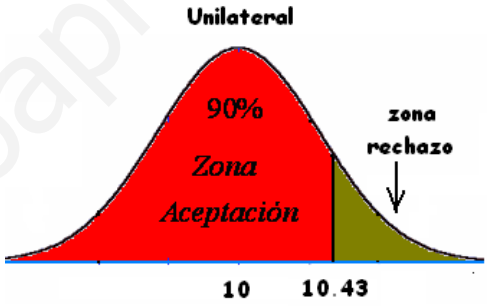


Observación: el resultado del contraste es que no hay suficiente evidencia para rechazar la hipótesis nula, pero ello no equivale exactamente a que haya evidencia de que la hipótesis sea verdadera, ya que la media 22 no está más apoyada en nuestros datos que otras medias, como 21 o 20, que tampoco serían rechazadas como hipótesis nulas. De ahí que es mejor decir "no se rechaza la hipótesis nula" a decir que "se acepta".

Ejemplo 17.- (Unilateral)
Pablo y Virginia quieren contrastar si el consumo medio en teléfono móvil entre los estudiantes es, como máximo, de 10€ frente a si es mayor.
Pablo, en una muestra de 36 estudiantes, obtuvo una media de 10.41€ con una desviación típica de 2€
Virginia obtuvo, en una muestra de 49 estudiantes, una media de 10,39€ con una desviación típica de 2€
¿Qué decisión toma Pablo con un nivel de significación del 10%?
¿Qué decisión toma Virginia con un nivel de significación del 10%?
Sol.-
1º) Ambos quieren contrastar la hipótesis nula: $H_0 : \mu \leq 10$ **frente a** $H_1 : \mu > 10$
2º) Si $H_0 : \mu \leq 10$ **es cierta, para determinar la zona de rechazo supondremos que la media de gasto es el mayor valor compatible con** $H_0 : \mu \leq 10$**, que es** $\mu = 10$
Caso de Pablo: $\mu = 10 \quad n = 36$ **entonces las medias muestrales del consumo medio en teléfono móvil** $\overline{X}$ **se distribuyen:** $N\left(10 , \frac{1}{3}\right)$

En el test unilateral, la zona de rechazo (toda una cola de la distribución) está en uno de los extremos (el derecho en este caso) por tanto, a nivel de rechazo del 10% el punto crítico es $z_\alpha = 1.28$ **y así tenemos que:**
El intervalo de no rechazo para Pablo es:

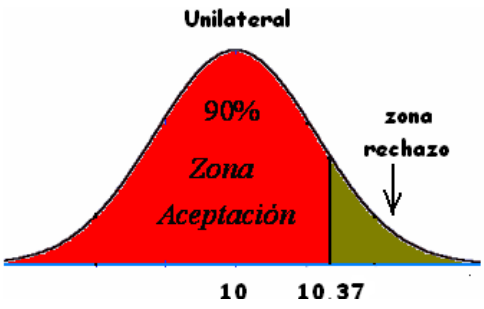
$$\left(-\infty , 10 + 1.28 * \frac{1}{3}\right) = \left(-\infty , 10.43\right)$$



Caso de Virginia: $\mu = 10 \quad n = 49$ **entonces las medias muestrales de consumo medio en teléfono móvil** $\overline{X}$ **se distribuyen:** $N\left(10 , \frac{2}{\sqrt{49}}\right)$

El intervalo de no rechazo para Virginia:

$$\left(-\infty , 10 + 1.28 * \frac{2}{\sqrt{49}}\right) = \left(-\infty , 10.37\right)$$



3º) Según el enunciado las medias muestrales para Pablo y Virginia son: 10.41 y 10.37 respectivamente.

4ª) Según los resultados obtenidos en el punto 2º) y comprobando si los obtenidos en el 3º) caen dentro de la zona de no rechazo, concluimos que:
Pablo no rechaza la hipótesis nula y Virginia sí la rechaza.

Test de Hipótesis para la proporción.-

En el caso bilateral la hipótesis nula consiste en atribuirle un valor a la proporción de individuos que tienen una cierta cualidad o la probabilidad de un suceso en una experiencia aleatoria.

Ejemplo 18.- (Bilateral)

a) Lanzamos una moneda 10 veces y obtenemos 6 caras.

b) Lazamos una moneda 100 veces y obtenemos 60 caras.

c) Lanzamos una moneda 1000 veces y obtenemos 600 caras.

¿Podemos deducir de alguna de las experiencias que la moneda es incorrecta?

Sol.-

1º) $H_0 : p = \frac{1}{2}$ (la moneda es correcta) frente $H_1 : p \neq \frac{1}{2}$

2º) Si $H_0 : p = \frac{1}{2}$ es cierta; como en los tres casos a) b) y c) $n.p = n.\frac{1}{2} \geq 5$

entonces la distribución de las proporciones muestrales es $\overline{X}$ es $N\left(0.5\,,\sqrt{\frac{0.5*0.5}{n}}\right)$

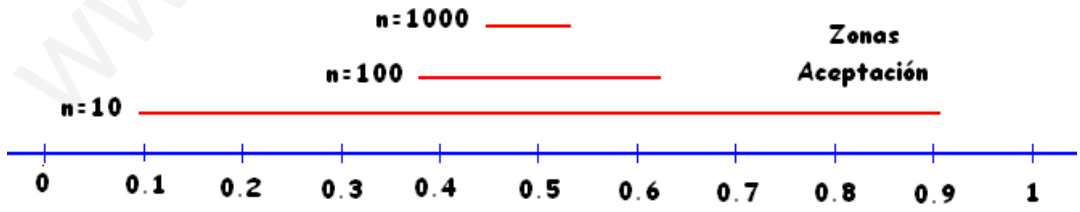
Para un nivel de significación $\alpha = 0.01$ significa que el 99% de las proporciones muestrales están en el intervalo: $\left(0.5 \pm 2.575 * \frac{0.5}{\sqrt{n}}\right)$ **zona de no rechazo.**

n	10	100	1000
Zona Aceptación	$(0.09\,,\,0.91)$	$(0.37\,,\,0.63)$	$(0.46\,,\,0.54)$

3º) Se extrae la muestra y se calcula la proporción de caras y según el enunciado en los tres casos tenemos que la proporción muestral es $pr = 0.6$

4º) Como solamente en el caso c) el valor de $pr = 0.6$ cae fuera de la zona de aceptación concluimos que **rechazamos** la hipótesis nula en dicho caso c) y en los otros dos casos a) y b) **no rechazamos** la hipótesis nula.

Es decir, sólo rechazamos que la moneda sea correcta en el caso c).



Observemos que el resultado de la prueba es acorde con la intuición: 6 caras en 10 tiradas parece de lo más razonable, en 100 lanzamientos cuesta admitir "ciertas" desviaciones sobre la proporción 50% pero en 1000 lanzamientos, aún siendo tan permisivos como es tomar un nivel de significación $\alpha = 0.01$, no resulta admisible considerar por azar que el 60% de los lanzamientos de una moneda correcta sean caras, en ese caso la moneda es incorrecta.

Ejemplo 19.- (Unilateral)
Según la ley electoral de cierto país, para obtener representación parlamentaria, un partido político ha de conseguir al menos el 5% de los votos. Poco antes de celebrarse las elecciones, una encuesta realizada sobre 1000 ciudadanos elegidos al azar revela que 36 de ellos votarán al partido X. ¿Puede estimarse, con un nivel de significación del 5% que X tendrá representación parlamentaria? ¿Y con un nivel del 1%?
Sol.
1º) Se establece la hipótesis nula: $H_0 : p \geq 0.05$ **frente a** $H_1 : p < 0.05$

2º) Si $H_0 : p \geq 0.05$ **es cierta, para determinar la zona de rechazo supondremos que la probabilidad es el menor valor compatible con** $H_0 : p \geq 0.05$**, que es** $p = 0.05$
(Que es el caso más favorable para la hipótesis nula $H_0 : p \geq 0.05$**)**
Las proporciones muestrales pr **, en muestras de tamaño 100, se distribuyen:**

$$N\left(0.05\,,\sqrt{\frac{0.05*0.95}{100}}\right) = N(0.05\,,0.007)$$

Para un nivel de significación del 5% la región de aceptación o de no rechazo:

$$(0.05 - 1.65*0.007\,,+\infty) = (0.038\,,+\infty)$$

Para un nivel de significación del 1% la región de aceptación o de no rechazo:

$$(0.05 - 2.33*0.007\,,+\infty) = (0.034\,,+\infty)$$

3º) En la muestra de tamaño 1000 se calcula la proporción de votantes del partido X y según el enunciado $pr = \frac{36}{1000} = 0.036$

4º) Como para un nivel de significación del 5% el valor de $pr = 0.036$ **cae fuera de la zona de aceptación concluimos que rechazamos la hipótesis nula.**
Como para un nivel de significación del 1% el valor de $pr = 0.036$ **cae dentro de la zona de aceptación concluimos que no rechazamos la hipótesis nula.**

Tablas de Valores Críticos más Usuales

Test Bilateral			
α	0.10	0.05	0.01
$z_{\alpha/2}$	1.645	1.96	2.575

Test Unilateral			
α	0.10	0.05	0.01
z_{α}	1.28	1.645	2.33