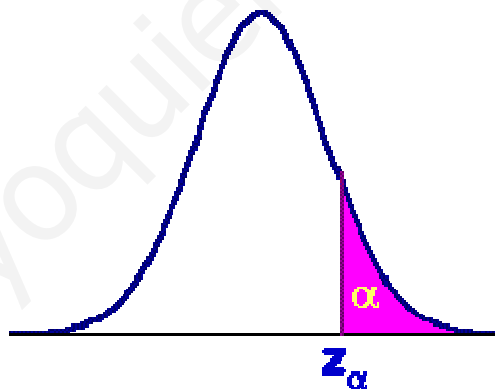


INFERENCIA ESTADÍSTICA



MATEMÁTICAS APLICADAS A LAS CC. SS. II



Alfonso González
IES Fernando de Mena
Dpto. de Matemáticas

I) REPASO ESTADÍSTICA ELEMENTAL

El concepto de variable estadística es bastante intuitivo y fácil de comprender. Existen los siguientes tipos de variables o caracteres:

1. **Caracteres cualitativos:** Se describen mediante palabras y no se pueden medir, es decir, no se les puede asignar un número; más bien, presentan modalidades. Ejemplos: el estado civil, la nacionalidad, el nivel de estudios, el color del pelo, etc.
2. **Caracteres cuantitativos:** Se describen mediante números, es decir, son medibles. A su vez, se dividen en:
 - 2.1 **Caracteres cuantitativos discretos:** Sólo toman valores puntuales aislados. Ejemplos: número de hijos, número del calzado, nota de la evaluación de Matemáticas, etc.
 - 2.2 **Caracteres cuantitativos continuos:** Pueden tomar cualquier valor intermedio entre dos cualesquiera. Ejemplos: estatura, peso, nota en un examen, etc.

Nosotros vamos a tratar en este tema generalmente variable estadística continua.

Supongamos que, en vísperas de unas elecciones, queremos tener una idea sobre las intenciones de voto del electorado. Evidentemente, no podemos preguntárselo al conjunto de todos los votantes (eso ocurrirá más bien el día de las elecciones), por lo cual tendremos que limitarnos a hacer un sondeo a una parte, naturalmente lo más representativa posible. Surgen así los siguientes conceptos:

Población: Conjunto de todos los individuos objeto del estudio.

Muestra: Subconjunto extraído de la población, cuyo estudio sirve para inferir características de toda la población.

La elección de la muestra (lo que se conoce como **muestreo**) debe realizarse con cuidado para que la muestra sea lo más representativa posible; para ello hay dos factores fundamentales a tener en cuenta: el tamaño de la muestra –es evidente que si ésta es demasiado pequeña no será representativa, pero se pueden obtener estimaciones muy certeras con sondeos muy ajustados- y el cómo se lleva a cabo el muestreo: condición indispensable para que una muestra sea representativa es que el muestreo sea aleatorio¹. Los años 30 del siglo pasado vieron el nacimiento de la Estadística Inductiva, que permite obtener conclusiones válidas para toda la población a partir del estudio de una muestra.

- Para poder manejar cómodamente datos estadísticos se suelen construir tablas de frecuencias. Veámoslo, en primer lugar, para un ejemplo sencillo de variable discreta:

Ejemplo 1: Supongamos que las edades de los 25 alumnos/as de una clase de 2º de Bachillerato son las siguientes:

17, 17, 17, 18, 17, 17, 19, 17, 17, 18, 18, 17, 17, 20, 19, 17, 18, 17, 17, 17, 18, 17, 18, 17, 17

Con estos datos, construimos la siguiente tabla:

x_i	f_i	h_i	%
17	16	0,64	64
18	6	0,24	24
19	2	0,08	8
20	1	0,04	4
	25	1	100

- La 1ª columna contiene los valores que toma la variable aleatoria x_i , en este caso la edad, por orden creciente.
- La 2ª columna muestra el recuento de cada valor; se llama **frecuencia absoluta**, y se designa como f_i
- Dividiendo las frecuencias absolutas entre el número total de individuos (25) se obtienen las **frecuencias relativas** (3ª columna), que se designan como h_i

¹ Pueden verse con más profundidad todos estos conceptos en las págs. 264 y 265 de libro de texto.

- La última columna son las frecuencias relativas expresadas en tanto por ciento.
- En el caso de variable continua, la tabla se confecciona utilizando **intervalos de clase**, es decir, agrupando los datos en intervalos:

Ejemplo 2: Las estaturas, en metros, de los 40 alumnos/as de 2º de Bachillerato de un centro son las siguientes:

1,81 1,70 1,86 1,78 1,85 1,69 1,77 1,65 1,70 1,80 1,66 1,75
 1,79 1,92 1,76 1,74 1,81 1,68 1,63 1,60 1,73 1,84 1,80 1,81
 1,73 1,77 1,71 1,72 1,68 1,89 1,86 1,80 1,82 1,79 1,76 1,76
 1,72 1,72 1,65 1,76

Construimos una tabla agrupando estos datos en intervalos de amplitud 5 cm:

Estatura	x_i	f_i	h_i
[1,60;1,65)	1,625	2	0,05
[1,65;1,70)	1,675	6	0,15
[1,70;1,75)	1,725	9	0,225
[1,75;1,80)	1,775	10	0,25
[1,80;1,85)	1,825	8	0,2
[1,85;1,90)	1,875	4	0,1
[1,90;1,95]	1,925	1	0,025
		40	1

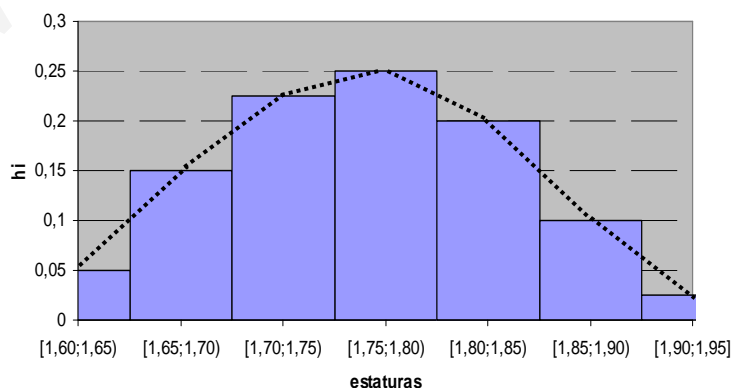
Nótese que en cada intervalo no se incluye el extremo derecho, excepto en el último. El punto medio de cada intervalo se llama **marca de clase**. ¿Para qué se utilizan las marcas de clase? Entre otras cosas, como veremos enseguida, para hallar la media de estaturas de esta distribución.

- Una práctica muy habitual es representar gráficamente la distribución estadística. La ventaja es que así podemos visualizarla fácilmente, captando así de un vistazo todos sus detalles. Para representarla disponemos de dos posibilidades:

Histograma: Se trata de levantar para cada intervalo de clase un rectángulo cuya altura² indique su frecuencia absoluta (histograma de frecuencias absolutas), o la relativa (histograma de frecuencias relativas). (NOTA: en el caso de distribuciones discretas, como la del ejemplo 1, más bien se llama diagrama de barras, puesto que, en vez de rectángulos, se dibujan líneas verticales).

Polígono de frecuencias: Se obtiene uniendo el punto medio del lado superior de cada rectángulo. Obviamente, puede tratarse de un polígono de frecuencias absolutas o de un polígono de frecuencias relativas.

Lo más usual es dibujar ambos a la vez, superpuestos. Veamos cómo serían el histograma y el polígono de frecuencias relativas de la distribución anterior:



² En el caso de que los intervalos de clase no sean de amplitud constante, más bien hay que hacer que cada rectángulo sea de área proporcional a la frecuencia de dicho intervalo.

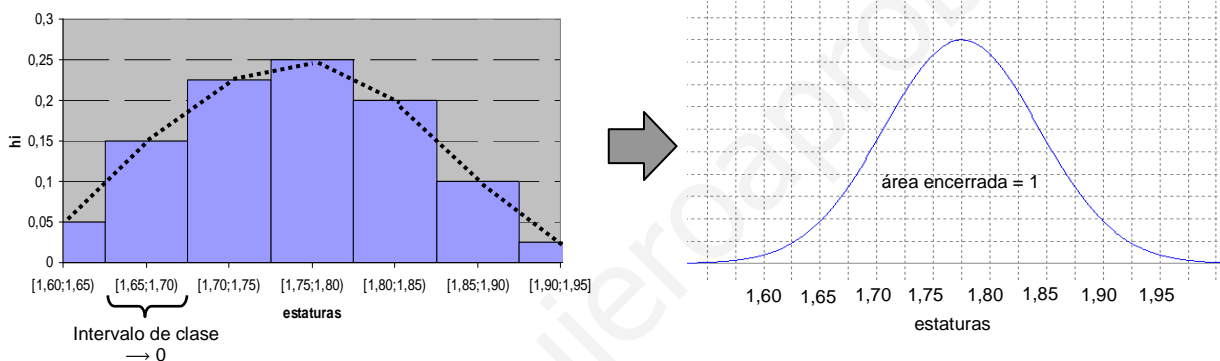
Por último, para poder definir correctamente la distribución, también se acostumbra a definir una serie de parámetros estadísticos; los más importantes son dos: la media aritmética y la desviación típica.

El concepto de media aritmética es bastante intuitivo: la **media aritmética** de una serie de valores se obtiene sumándolos, y dividiendo la suma por el número de valores. Se designa³ mediante \bar{x} o μ . En el caso de datos agrupados en intervalos, se escoge como valor de cada intervalo el punto medio o marca de clase. En el ejemplo anterior, puede comprobarse que sale 1,765 m de estatura media.

El otro parámetro aludido es la **desviación típica**, que se designa⁴ como s o σ . Su significado lo veremos más adelante (apdo. VI.4); baste decir, por el momento, que está relacionada con la forma de la distribución: cuando mayor es σ , más dispersos son los datos, y viceversa.

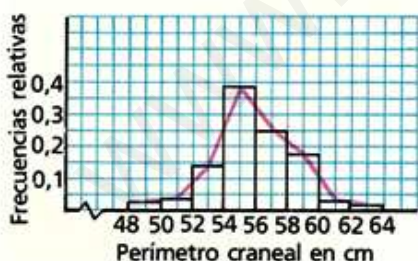
II) IDEA INTUITIVA DE DISTRIBUCIÓN NORMAL

Continuando con el ejemplo de las estaturas de los 40 alumnos/as, podemos suponer que la distribución de las tallas de los alumnos/as de 2º de Bachillerato de toda España, a medida que los intervalos de clase van siendo pequeños, tiende al diagrama representado a la derecha:



El diagrama de la derecha proporciona una idea intuitiva de la que se conoce como distribución normal.

Sabemos que la suma de todas las frecuencias relativas h_i de una distribución de frecuencias es igual a la unidad. Por lo tanto, **el área encerrada debajo de la curva normal⁵ es igual a 1.**



La distribución del perímetro craneal de una serie de individuos tiende al modelo campaniforme, es decir, a la distribución normal.

La distribución normal se llama así porque durante mucho tiempo se pensó que ese era el comportamiento de todos los fenómenos de la vida real. Conviene tener en cuenta que existe un gran número de fenómenos de la vida real cuyos histogramas de frecuencias observadas pueden considerarse como la imagen empírica de una distribución normal, es decir, simétrica; por ejemplo, en la distribución del perímetro craneal de una serie de individuos (ver figura), los que tienen poco y mucho perímetro se encuentran igualmente separados por la media y se presentan con frecuencias similares. Otros ejemplos de situaciones que pueden responder, aproximadamente, a una curva normal, pueden ser: el peso de

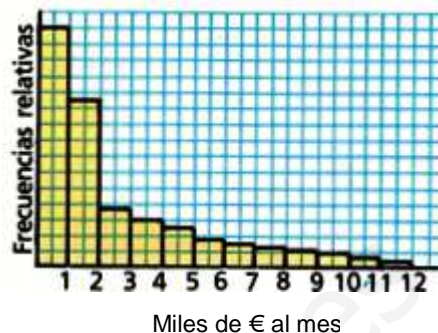
³ En realidad, se designa como \bar{x} en el caso de variable estadística, y μ en el caso de variable aleatoria; en este último caso, la media también se llama esperanza matemática o valor esperado.

⁴ En realidad, se designa como s en el caso de variable estadística, y σ en el caso de variable aleatoria.

⁵ La curva normal también se llama curva de Gauss o campana de Gauss, en honor al matemático alemán Carl Friedrich Gauss (1777-1855), que fue el primero en ver sus propiedades y utilidad.

los individuos de la población, el cociente intelectual, etc., es decir, medidas psicométricas. Pero también se obtiene una distribución que tiende a la normal en otro tipo de medidas: duración media de una bombilla, esperanza de vida de una mascota, etc... Por supuesto, todo esto no se puede demostrar, pero es algo que se comprueba estadísticamente...

¡Cuidado!: No todas las distribuciones se comportan según el modelo normal. Consideremos el siguiente contraejemplo: supongamos que clasificamos a los ciudadanos españoles según su nivel de renta (ver figura). Evidentemente, son muy pocos los que poseen niveles de rentas altas y en cambio son muchos los que poseen niveles de rentas bajas. Por tanto, la distribución ya no es simétrica y, en consecuencia, no se adapta al modelo normal.



III) DEFINICIÓN DE DISTRIBUCIÓN NORMAL: CURVA DE GAUSS

Volviendo al ejemplo 2, si el número de alumnos/as crece indefinidamente y vamos haciendo cada vez más pequeña la amplitud de los intervalos, el polígono de frecuencias tomará entonces la forma de la distribución normal, es decir, de una campana de Gauss. La variable en este caso sería continua y su recorrido sería, en principio, el intervalo $[1,60;1,95]$, aunque teóricamente sería desde 0 hasta ∞ .

Vamos a definir la distribución normal teórica:

Una variable aleatoria continua X sigue una distribución normal de media μ y desviación típica σ , y se designa por $N(\mu, \sigma)$, si cumple:

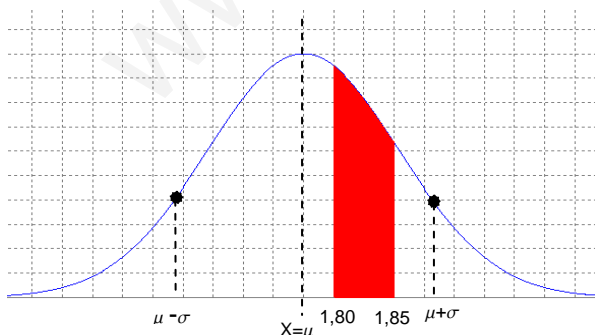
- 1º) La variable recorre toda la recta real, es decir $(-\infty, \infty)$
- 2º) La distribución tiene la forma de la curva de Gauss, cuya ecuación⁶ es:

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

Observaciones: 1º) μ y σ se llaman parámetros de la distribución normal, y cumplen: $-\infty < \mu < \infty$ y $\sigma > 0$

2º) El significado de μ es intuitivo; el de σ lo veremos más adelante, en el apartado VI.4)

La forma de la curva de Gauss es la siguiente:



Se trata de una curva simétrica respecto a $X = \mu$. Además, tiene su máximo en $X = \mu$. Por otra parte, ya podemos empezar a ver el significado de la desviación típica, la cual nos indica la forma de la curva: una desviación típica σ a ambos lados de la media μ nos marca la posición de los dos puntos de inflexión⁷ simétricos de la curva. Por último, y muy importante: recordar que el área total encerrada en el histograma de frecuencias relativas era 1 \Rightarrow **el área encerrada bajo la curva de Gauss es 1**

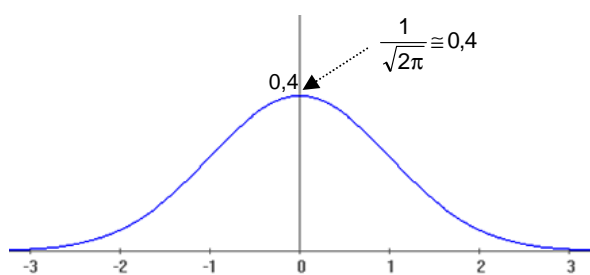
⁶ $e \cong 2,7182818...$ es un número irracional llamado constante de Euler, en honor al matemático suizo Leonhard Euler (1707-1783)

⁷ Los puntos de inflexión de una curva son los puntos donde cambia la curvatura de ésta.

¿Qué representa la curva de Gauss? Por ejemplo, **el área rayada** de la curva de la figura representa la frecuencia relativa de individuos con estatura comprendida entre 1,80 y 1,85, es decir, **representa la probabilidad de que un individuo tenga una talla comprendida entre dichos valores.**

IV) DISTRIBUCIÓN NORMAL ESTÁNDAR: N(0,1)

A la vista de lo anterior, es evidente que para cada valor de μ y σ tendremos una curva de Gauss distinta. Pero de las infinitas distribuciones normales posibles tiene interés la distribución **N(0,1)**, llamada **distribución normal estándar**, distribución normal reducida, distribución normada o tipificada. Corresponde a la ecuación:



(NOTA: Se acostumbra a modificar la escala del eje y para conseguir una forma más acampanada de la curva)

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

, cuya representación gráfica se muestra al margen. Decimos que tiene interés pues el área que deja está tabulada (ver tabla suministrada en la PAU), y además toda distribución se puede poner en función de la N(0,1), como veremos en el siguiente apartado.

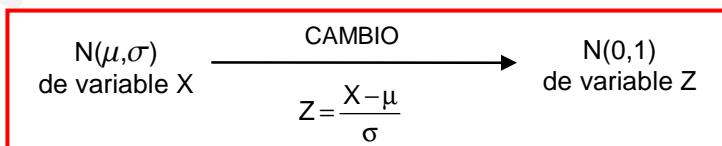
V) TIPIFICACIÓN DE LA VARIABLE (pág. 278 del libro de texto)

Consiste en transformar la variable **X** que sigue una distribución **N(μ , σ)** en otra variable **Z** que siga una distribución **N(0,1)**, la cual está tabulada. Para ello, hay que hacer dos pasos:

- 1º) **CENTRAR:** Trasladar la media de la distribución al origen de coordenadas, haciendo $x-\mu$ (esto equivale a hacer $\mu=0$)
- 2º) **CONTRAER** la gráfica de la distribución σ unidades, lo cual se consigue dividiendo por σ (esto equivale a hacer $\sigma=1$)

Estos dos pasos se consiguen simultáneamente haciendo el cambio de variable $Z = \frac{X-\mu}{\sigma}$ (1)

En resumen:



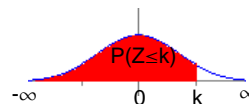
En el próximo apartado haremos uso de la expresión (1).

VI) MANEJO DE TABLAS (CÁLCULO DE ÁREAS BAJO LA CURVA DE GAUSS)

Para hallar áreas (y por lo tanto probabilidades) bajo la curva normal mediante tabla a menudo tendremos en cuenta que dicha curva es simétrica y por lo tanto deja un 50 % de observaciones a ambos lados del origen. Además, no debe olvidarse que el área encerrada total es 1, es decir el 100 % de observaciones, o lo que es lo mismo, la probabilidad del suceso seguro. Vayamos por pasos:

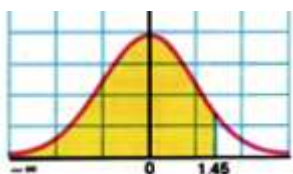
VI.1) Cálculo de probabilidades en una distribución N(0,1) (págs. 276 y 277 del libro)

Vamos a ver todos los casos que se pueden presentar, desde el más fácil hasta el último, el más "complicado". Sea Z una variable que sigue una distribución normal N(0,1). Veamos algunos ejemplos, utilizando la tabla más usual, que es una tabla de frecuencias –es decir, de probabilidades- acumuladas (precisamente la suministrada en la PAU), como la que figura al margen. Esta tabla nos proporciona la probabilidad de que la variable Z sea menor o igual que un determinado valor k. Este valor k se busca en la tabla⁸ de la siguiente forma: sus unidades y décimas en la columna izquierda y las centésimas en la superior:



Ejemplo 3: Calcular $P(Z \leq 1,45)$

La probabilidad pedida es el área sombreada:



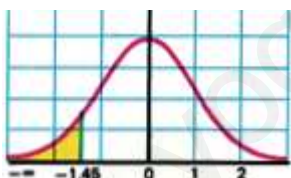
Por lo tanto, buscamos en la tabla $k=1,4$ en la columna izquierda, y 0,05 en la fila superior; su intersección nos da:

$$P(Z \leq 1,45) = \boxed{0,9265}$$

Es decir, el 92,65% de las observaciones se distribuyen entre $-\infty$ y 1,45

Ejemplo 4: Calcular $P(Z \leq -1,45)$

La probabilidad pedida es el área sombreada:

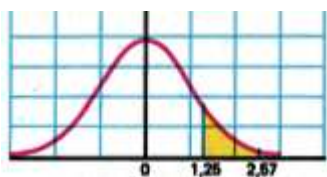


La tabla solo proporciona probabilidades para valores de k positivos. Ahora bien, teniendo en cuenta la simetría de la curva de Gauss, y que el área total encerrada por esta es 1:

$$P(Z \leq -1,45) = P(Z > 1,45) = 1 - P(Z \leq 1,45) = 1 - 0,9265 = \boxed{0,0735}$$

NOTA: En una distribución de variable continua (caso de Z) las probabilidades puntuales son nulas; por lo tanto, $P(Z \leq k) = P(Z < k)$

Ejemplo 5: Calcular $P(1,25 < Z \leq 2,57)$



La probabilidad pedida es el área sombreada; por lo tanto, restaremos al área mayor la menor:

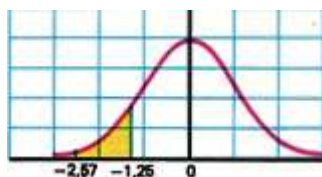
$$P(1,25 < Z \leq 2,57) = P(Z \leq 2,57) - P(Z \leq 1,25) = 0,9949 - 0,8944 = \boxed{0,1005}$$

k	0.00	0.01	0.02	0.03	0.04	0.05
0.0	.5000	.5040	.5080	.5120	.5160	.5199
0.1	.5298	.5438	.5478	.5517	.5557	.5596
0.2	.5793	.5832	.5871	.5910	.5948	.5987
0.3	.6179	.6217	.6255	.6293	.6331	.6368
0.4	.6554	.6591	.6628	.6664	.6700	.6736
0.5	.6915	.6950	.6985	.7019	.7054	.7088
0.6	.7257	.7291	.7324	.7357	.7389	.7422
0.7	.7580	.7611	.7642	.7673	.7704	.7734
0.8	.7881	.7910	.7939	.7967	.7995	.8023
0.9	.8159	.8186	.8212	.8238	.8264	.8289
1.0	.8413	.8438	.8461	.8485	.8508	.8531
1.1	.8643	.8665	.8686	.8708	.8729	.8749
1.2	.8849	.8869	.8888	.8907	.8925	.8944
1.3	.9032	.9049	.9066	.9082	.9099	.9115
1.4	.9192	.9207	.9222	.9236	.9251	.9265
1.5	.9332	.9345	.9357	.9370	.9382	.9394

Buscamos unidades y décimas de k en la columna izquierda, y centésimas en la fila superior

⁸ En el apdo. VII veremos que k habitualmente se llama valor crítico, y lo designaremos como $z_{\alpha/2}$

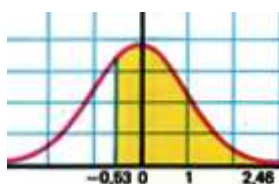
Ejemplo 6: Calcular $P(-2,57 < Z \leq -1,25)$



La probabilidad pedida es el área sombreada; por lo tanto, por simetría de la curva normal:

$$P(-2,57 < Z \leq -1,25) = P(1,25 < Z \leq 2,57) = \boxed{0,1005}$$

Ejemplo 7: Calcular $P(-0,53 < Z \leq 2,46)$



La probabilidad pedida es el área sombreada. Este es el caso más general, y se resuelve teniendo en cuenta todo lo anterior:

como en el ejemplo 5

como en el ejemplo 4

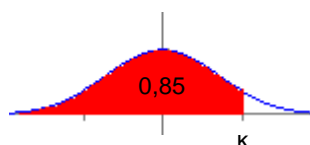
$$P(-0,53 < Z \leq 2,46) = P(Z \leq 2,46) - P(Z \leq -0,53) = P(Z \leq 2,46) - P(Z \geq 0,53) = P(Z \leq 2,46) - [1 - P(Z < 0,53)] = 0,9931 - (1 - 0,7019) = \boxed{0,695}$$

Ejercicios final tema: 1 a 3

Ejercicios libro: pág. 277: 1

- Por otra parte, tenemos el problema inverso: dada una determinada probabilidad, hallar el valor de k asociado. Veamos un ejemplo:

Ejemplo 8: Calcular el valor de k (exacto o aproximado) tal que $P(Z \leq k) = 0,85$



En las tablas encontramos que $P(Z \leq 1,03) = 0,8485$ y $P(Z \leq 1,04) = 0,8508$; por lo tanto, asignaremos a k el valor intermedio de $1,03$ y $1,04$, es decir $\boxed{k = 1,035}$

Ejercicio final tema: 4

VI.2) Cálculo de probabilidades en una distribución $N(\mu, \sigma)$

(Tipificación de la variable) (pág. 278 del libro)

Como hemos visto en el apartado V, utilizando la fórmula (1) podemos convertir una variable X que sigue una distribución $N(\mu, \sigma)$ en otra variable Z que siga una distribución $N(0, 1)$, con la ventaja de que esta última está tabulada.

$$N(\mu, \sigma) \text{ de variable } X \xrightarrow{\text{CAMBIO}} N(0, 1) \text{ de variable } Z$$

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

Esto se puede hacer porque, en esencia, todas las distribuciones normales son la misma. Por lo tanto, es más cómodo buscar probabilidades en una distribución $N(0, 1)$. Veamos ejemplos:

Ejemplo 9: En una distribución $N(66,8)$, calcular: **a)** $P(X < 70)$ **b)** $P(X > 80)$ **c)** $P(70 < X < 80)$

$$\text{a) } X=70 \rightarrow Z = \frac{X-\mu}{\sigma} = \frac{70-66}{8} = 0,5$$

$$X=80 \rightarrow Z = \frac{X-\mu}{\sigma} = \frac{80-66}{8} = 1,75$$

Por tanto, $P(X < 70) = P(Z < 0,5) = \boxed{0,6915}$

b) $P(X > 80) = P(Z > 1,75) = 1 - P(Z < 1,75) = 1 - 0,9599 = \boxed{0,0401}$

c) $P(70 < X < 80) = P(0,5 < Z < 1,75) = P(Z < 1,75) - P(Z < 0,5) = 0,9599 - 0,6915 = \boxed{0,2684}$

Ejercicios final tema: 5 y 6

VI.3) Problemas de aplicación

Ejemplo 10: Los pesos de los individuos de una población se distribuyen normalmente con media 70 kg y desviación típica 6 kg. De una población de 2000 personas, calcular cuántas personas tendrán un peso entre 64 y 76 kg.

$$\left. \begin{array}{l} \mu = 70 \text{ kg} \\ \sigma = 6 \text{ kg} \end{array} \right\} \begin{array}{l} N(\mu, \sigma) \\ \text{de variable } X \end{array} \xrightarrow[\text{la variable}]{\text{tipificamos}} N(0,1) \text{ de nueva variable } Z = \frac{X-\mu}{\sigma}$$

$$X=64 \rightarrow Z = \frac{X-\mu}{\sigma} = \frac{64-70}{6} = -1$$

$$X=76 \rightarrow Z = \frac{X-\mu}{\sigma} = \frac{76-70}{6} = 1$$

Por tanto, $P(64 < X < 76) = P(-1 < Z < 1) = P(Z < 1) - P(Z < -1) = P(Z < 1) - [1 - P(Z < 1)] = 0,8413 - (1 - 0,8413) = 0,6826$

Es decir, el 68,26% de la población tendrán un peso entre 64 y 76 kg. Por consiguiente:

Solución: $2000 \cdot 0,6826 = \boxed{1365 \text{ personas}}$

Ejercicio 1: Se ha aplicado a 300 alumnos/as un test y se ha observado que se distribuyen normalmente con media 30 y desviación típica 12

a) ¿Cuántos alumnos/as tendrán una puntuación superior a 42?

b) ¿Qué proporción de alumnos/as tendrá una puntuación entre 20 y 25?

a)

[Soluc: $\cong 47$ alumnos/as]

b)

[Soluc: $\cong 40$ alumnos/as]

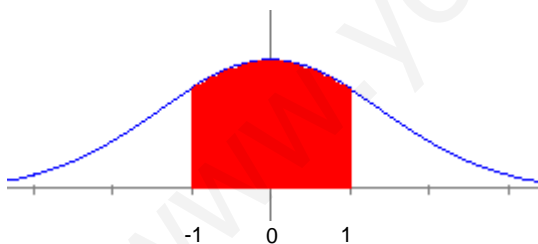
Ejercicios final tema: 7 a 19

VI.4) Casos particulares de áreas bajo la curva. Significado de σ

A veces tiene interés saber qué proporción de individuos de la población se distribuye en intervalos de la forma $(\mu-\sigma, \mu+\sigma)$, $(\mu-2\sigma, \mu+2\sigma)$, etc.

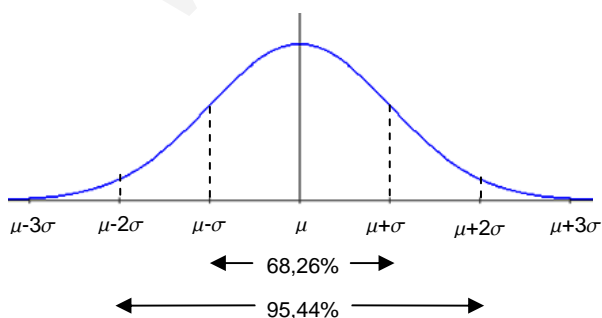
Supongamos que tenemos una variable aleatoria X que sigue una distribución $N(\mu, \sigma)$. Veamos, en primer lugar, qué proporción de individuos se encontrará en el intervalo $(\mu-\sigma, \mu+\sigma)$; para ello, tipificaremos la variable, como hemos hecho en ejemplos anteriores:

$$\left. \begin{array}{l} X=\mu-\sigma \xrightarrow[\begin{array}{c} Z = \frac{X-\mu}{\sigma} \\ \text{CAMBIO} \end{array}]{\text{CAMBIO}} Z = \frac{(\mu-\sigma)-\mu}{\sigma} = -1 \\ X=\mu+\sigma \xrightarrow{\text{CAMBIO}} Z = \frac{(\mu+\sigma)-\mu}{\sigma} = 1 \end{array} \right\} \Rightarrow \text{hay que hallar } P(-1 < Z \leq 1)$$



$$\begin{aligned} \text{Área sombreada} &= P(-1 < Z < 1) = P(Z < 1) - P(Z < -1) = \\ &= P(Z < 1) - [1 - P(Z < 1)] = 0,8413 - (1 - 0,8413) = 0,6826 \end{aligned}$$

Es decir, el 68,26% de las observaciones se encuentran alejadas una desviación típica a ambos lados de la media.



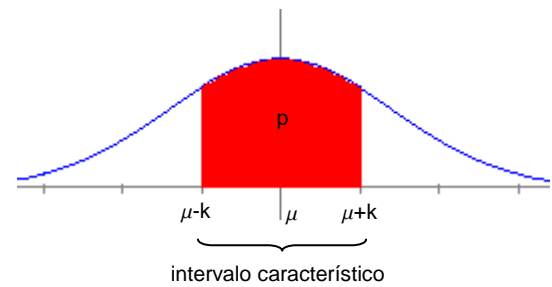
En general, puede probarse que:

- El 68,26% de las observaciones están en $(\mu-\sigma, \mu+\sigma)$
- El 95,44% de las observaciones están en $(\mu-2\sigma, \mu+2\sigma)$
- El 99,74% de las observaciones están en $(\mu-3\sigma, \mu+3\sigma)$

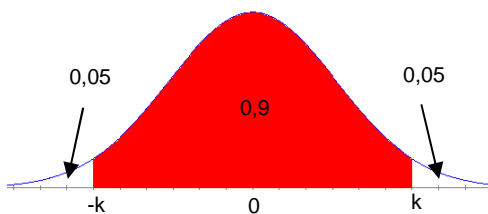
VII) INTERVALOS CARACTERÍSTICOS (pág. 279 del libro)

Definición: Dada una variable X que sigue una distribución de media μ , se llama **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 :

$$P(\mu-k < X < \mu+k) = p$$



Ejemplo 11: Hallar el intervalo característico de una distribución normal $N(0,1)$ correspondiente a la probabilidad $p=0,9$



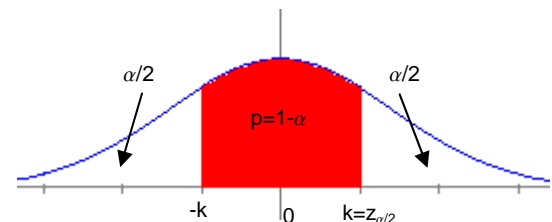
Viendo la figura adjunta, si dentro del intervalo tiene que haber un área de 0,9, fuera de él habrá por tanto 0,1; ahora bien, como el intervalo es, por definición, simétrico, el área de las dos *colas* será 0,05.

Por tanto, se trata de buscar k tal que $P(Z \leq k) = 0,95$

Este es una situación idéntica a la ya vista en el ejemplo 8; en las tablas encontramos que $P(Z \leq 1,64) = 0,9495$ y $P(Z \leq 1,65) = 0,9505$; por lo tanto, asignaremos a k el valor intermedio 1,645.

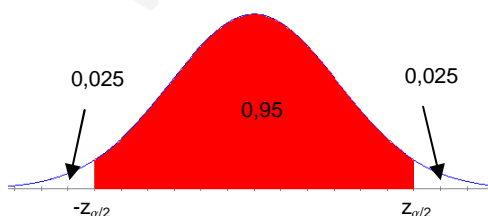
Solución: el intervalo característico pedido es $(-1,645; 1,645)$; se trata de un intervalo simétrico respecto al origen y en el que se sitúan el 90% de las observaciones.

Definición: En una distribución $N(0,1)$, si $(-k, k)$ es el intervalo característico correspondiente a una probabilidad p , diremos que k es el **valor crítico** correspondiente a p



Habitualmente, el valor crítico se suele denominar $z_{\alpha/2}$, siendo $p=1-\alpha$ (ver figura).

Ejemplo 12: Calcular razonadamente el valor crítico correspondiente a una probabilidad de 0,95



En la figura se ve que obviamente $\alpha/2=0,025$; por tanto, $z_{\alpha/2}$ será tal que la probabilidad que acumule sea $0,95+0,025=0,975$:

$$P(z \leq z_{\alpha/2}) = 0,975$$

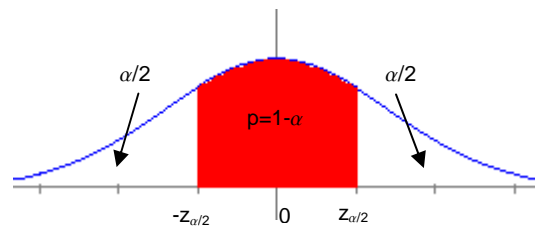
De la tabla deducimos que, entonces, $z_{\alpha/2} = 1,96$

Ejercicios libro: pág. 279: 1 y 2

Intervalos característicos en distribuciones $N(\mu, \sigma)$:

Acabamos de ver que en una variable z que siga una distribución $N(0,1)$ el intervalo característico correspondiente a la probabilidad $p=1-\alpha$ (ver figura) es $(-Z_{\alpha/2}, Z_{\alpha/2})$. Esto significa que:

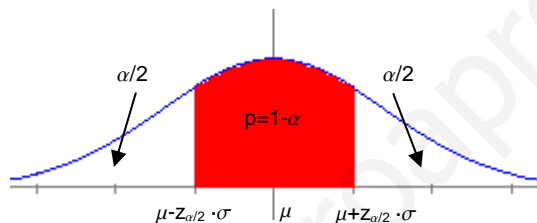
$$-Z_{\alpha/2} < z < Z_{\alpha/2} \quad (2)$$



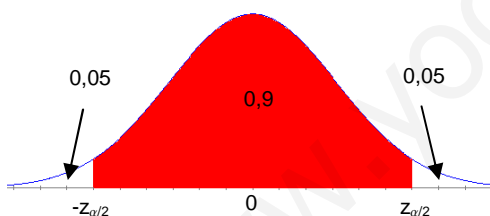
con una probabilidad $p=1-\alpha$. Ahora bien, esto es para el caso particular de una distribución $N(0,1)$; pero, ¿qué pasa en el caso general de una $N(\mu, \sigma)$? Para pasar de una a otra, sustituyamos en (2) el cambio de variable ya utilizado, $z = \frac{x-\mu}{\sigma}$:

$$-Z_{\alpha/2} < \frac{x-\mu}{\sigma} < Z_{\alpha/2} \Rightarrow -Z_{\alpha/2} \cdot \sigma < x-\mu < Z_{\alpha/2} \cdot \sigma \Rightarrow \mu - Z_{\alpha/2} \cdot \sigma < x < \mu + Z_{\alpha/2} \cdot \sigma$$

Es decir, el intervalo $(\mu - Z_{\alpha/2} \cdot \sigma, \mu + Z_{\alpha/2} \cdot \sigma)$ es el intervalo característico correspondiente a una probabilidad $p=1-\alpha$ para una variable x que sigue una distribución $N(\mu, \sigma)$; se trata de un intervalo centrado respecto a la media μ :



Ejemplo 13: En una distribución $N(66,8)$, obtener el intervalo característico para el 90%. Interpretarlo.



$$90\% \Rightarrow p=0,9=1-\alpha \Rightarrow \alpha=0,1 \Rightarrow \alpha/2=0,05$$

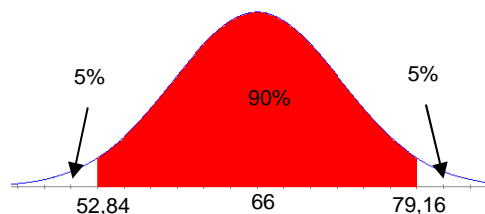
$0,9+0,05=0,95 \rightarrow$ hay que buscar el valor crítico $Z_{\alpha/2}$ tal que:

$$P(z \leq Z_{\alpha/2}) = 0,95$$

De la tabla deducimos que $Z_{\alpha/2} \cong 1,645$. Por lo tanto, el intervalo característico pedido será:

$$(\mu - Z_{\alpha/2} \cdot \sigma, \mu + Z_{\alpha/2} \cdot \sigma) = (66 - 1,645 \cdot 8, 66 + 1,645 \cdot 8) = (52,84; 79,16)$$

Es decir, el 90% de los individuos están en ese intervalo, o lo que es igual, $P(52,84 < x < 79,16) = 0,9$



Ejercicio 2: En una distribución $N(66,8)$, obtener el intervalo característico para el 95% y el 99%. Interpretarlos.

a)

[Soluc: (50,32;81,68)]

b)

[Soluc: (45,4;86,6)]

Ejercicios libro: pág. 280: 3 y 4; pág. 294: 1 a 4

VIII) DISTRIBUCIÓN DE LAS MEDIAS MUESTRALES. TEOREMA CENTRAL DEL LÍMITE (pág. 281 del libro)

Está demostrado estadísticamente que si en una distribución (no necesariamente normal, se entiende) extraemos muestras de tamaños $n=2$, $n=3$, $n=4$, etc.... la distribución de sus correspondientes medias se va pareciendo a una distribución normal a medida que aumenta n . Además, la media se mantiene constante, mientras que la desviación típica va disminuyendo (puede verse un ejemplo de esto en la pág. 274 de libro, para el caso del lanzamiento de 2, 3 y 4 dados). Este resultado se conoce como

Teorema central del límite:

Dada una población de media μ y desviación típica σ (¡no necesariamente normal!), la **distribución de las medias \bar{x} de las muestras** de tamaño n presenta las siguientes características:

- 1º) Tiene la misma media, μ , que la población.
- 2º) Su desviación típica es σ/\sqrt{n}
- 3º) Cuando $n \geq 30$, \bar{x} es prácticamente normal.

Observaciones:

1. Este teorema es válido cualquiera que sea la distribución de partida, tanto si es discreta como continua.

- Este teorema es básico para estimar la media de una población a partir de la media de una muestra extraída de ella, con ayuda de la distribución normal. Esto lo veremos en el próximo apartado.
- Si la población de partida es normal, también lo será la distribución de las medias muestrales, cualquiera que sea el valor de n (caso de los ejercicios 4B de la PAU)
- Aunque la población de partida no sea normal, para $n > 30$ es seguro que se consigue una gran aproximación a la normal.

Tabla resumen:

	población (x)	distribución de las medias de las muestras de tamaño n (\bar{x})
media	μ	μ
desviación típica	σ	σ/\sqrt{n}
		será normal si $n \geq 30$

Ejercicios libro: pág. 294: 5, 6, 7 y 12 (Teorema central del límite)

pág. 294 y ss.: 8, 9, 10, 11 y 13 (Cálculo de probabilidades en la distribución de las medias muestrales)

IX) INTERVALO DE CONFIANZA PARA LA MEDIA (págs. 285 a 287 del libro)

En este apartado veremos como, a partir de una muestra aleatoria de tamaño n de la cual conocemos su media, podemos estimar el valor de la media de toda la población dando un intervalo, llamado **intervalo de confianza**, dentro del cual confiamos que esté dicha media. Este es el objetivo final al que ha ido encaminado todo lo visto en los apartados anteriores.

Para ello, disponemos del siguiente teorema:

Se desea estimar la media, μ , de una población cuya desviación típica σ es conocida.

Para ello se recurre a una muestra de tamaño n en la cual se obtiene una media muestral, \bar{x} .

Si la población de partida es normal, o si el tamaño de la muestra es $n \geq 30$, entonces el intervalo de confianza de μ con un nivel de confianza $p=1-\alpha$ viene dado por:

$$\left(\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right) \quad (3)$$

Observaciones:

- La demostración de este hecho, bastante sencilla, puede verse en la pág. 286 del libro, y se basa en el teorema central del límite; en efecto, en el apartado VII obtuvimos que el intervalo $(\mu - Z_{\alpha/2} \cdot \sigma, \mu + Z_{\alpha/2} \cdot \sigma)$ es el intervalo característico correspondiente a una probabilidad $p=1-\alpha$ para una variable x que sigue una distribución $N(\mu, \sigma)$.

Por tanto, por el teorema central del límite, en dicho intervalo podemos reemplazar μ por \bar{x} , y σ por σ/\sqrt{n} , obteniendo así el intervalo dado por (3).

- Podemos ahora completar definitivamente la tabla resumen del apartado anterior:

	población (x)	distribución de las medias de las muestras de tamaño n (\bar{x})
media	μ	$\bar{x} = \mu$
desviación típica	σ	σ/\sqrt{n}
		será normal si $n \geq 30$
	intervalo característico: $(\mu - Z_{\alpha/2} \cdot \sigma, \mu + Z_{\alpha/2} \cdot \sigma)$	intervalo de confianza: $(\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$

3. Como ya hemos dicho al principio de este apartado, el **significado del intervalo de confianza** obtenido es que podemos afirmar que **existe** una probabilidad o, mejor dicho, **un nivel de confianza $p=1-\alpha$ (en %)** de que la media de toda la población esté en dicho intervalo.
4. Los ejercicios 4B de la PAU siempre presentan el mismo esquema:

DATOS DE PARTIDA: Tamaño de la muestra, n
Media de la muestra, \bar{x}
Desviación típica de la muestra, σ
Nivel de confianza, $p=1-\alpha$
(Además, se supone siempre que la población de partida es normal)

¿QUÉ NOS PIDEN?: El intervalo de confianza al $p=1-\alpha$ (en %) para la media de toda la población.

PASOS A SEGUIR: Se trata de obtener todos los términos del intervalo dado por (3):

$$\left(\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

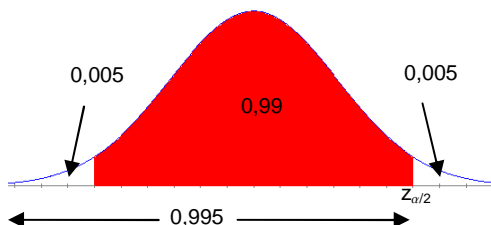
Para ello:

1. Obtenemos el valor crítico $Z_{\alpha/2}$ correspondiente a la probabilidad $p=1-\alpha$, de acuerdo con lo visto en el apdo. VII, para lo cual se recomienda un gráfico explicativo.
2. Sustituimos en el intervalo mencionado arriba el $Z_{\alpha/2}$ recién obtenido, así como los datos n , \bar{x} y σ
3. Finalmente, se interpreta el intervalo obtenido con una sencilla frase.

Ejemplo 14: (UCLM, 4B junio 2007) Para determinar cómo influye en la osteoporosis una dieta pobre en calcio, se realiza un estudio sobre 100 afectados por la enfermedad, obteniéndose que toman una media de calcio al día de 900 mg. Suponemos que la toma de calcio en la población de afectados por la enfermedad se distribuye normalmente con una desviación típica de 150 a) Encontrar un intervalo de confianza al 99 % para la media de calcio al día que toma toda la población afectada. b) Interpretar el significado del intervalo obtenido.

- a) Del enunciado deducimos que los datos son: Tamaño de la muestra: $n=100$
 Media de la muestra: $\bar{x}=900$ mg
 Desviación típica de la muestra: $\sigma=150$ mg
 Nivel de confianza, $p=1-\alpha=0,99$

Vamos a obtener el valor crítico $Z_{\alpha/2}$ correspondiente a la probabilidad $p=1-\alpha$, tal y como hacíamos en el apdo. VII:



$$99\% \Rightarrow p=0,99=1-\alpha \Rightarrow \alpha=0,01 \Rightarrow \alpha/2=0,005$$

$0,99+0,005=0,995 \rightarrow Z_{\alpha/2}$ será el valor crítico que acumule una probabilidad de 0,995, es decir:

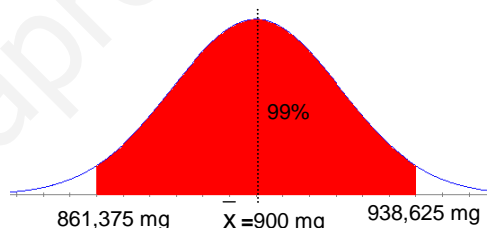
$$P(Z \leq Z_{\alpha/2})=0,995$$

De la tabla deducimos que, entonces, $Z_{\alpha/2} \cong 2,575$

Sustituimos todo lo anterior en el intervalo de confianza:

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right) = \left(900 - 2,575 \cdot \frac{150}{10}, 900 + 2,575 \cdot \frac{150}{10} \right) = (861,375; 938,625)$$

- b) Existe una probabilidad de 0,99 de que la media de calcio al día en mg que toma toda la población afectada esté en el intervalo anterior.



Ejercicios libro: pág. 287: 1; pág. 295: 14, 15 y 16

Ejercicios PAU: cualquier apartado 4B de cualquiera de las convocatorias pasadas

Ejercicios de Distribución normal

NOTA: En algunos ejercicios conviene identificar el rango de valores y el significado de la variable aleatoria continua X , y acompañar el planteamiento con una gráfica explicativa.

Cálculo de áreas:

- Utilizando la tabla de la curva normal tipificada, calcular las siguientes áreas: **a)** Área entre 0 y 0,25 **b)** Área desde $-\infty$ hasta 1,32 **c)** Área entre -2,23 y 1,15 (Soluc: 0,0987; 0,9066; 0,862)
- Sea Z una variable aleatoria $N(0,1)$. Calcular: **a)** $P(Z \geq 1,32)$ **b)** $P(Z \leq 2,17)$ **c)** $P(1,52 < Z \leq 2,03)$ (Soluc: 0,0934; 0,9850; 0,0431)
- Sea Z una variable aleatoria $N(0,1)$. Calcular: **a)** $P(Z \geq -1,32)$ **b)** $P(Z \leq -2,17)$ **c)** $P(-2,03 < Z \leq 1,52)$ (Soluc: 0,9066; 0,015; 0,9145)

Problema inverso:

- Calcular el valor de k (exacto o aproximado) en cada uno de los siguientes casos:

- | | | |
|---------------------------|-------------------------|---------------------------|
| a) $P(Z \leq k) = 0,5$ | d) $P(Z \leq k) = 0,33$ | g) $P(Z \geq k) = 0,9971$ |
| b) $P(Z \leq k) = 0,8729$ | e) $P(Z \leq k) = 0,2$ | h) $P(Z \geq k) = 0,6$ |
| c) $P(Z \leq k) = 0,9$ | f) $P(Z > k) = 0,12$ | |

(Soluc: a) $k=0$; b) $k=1,14$; c) $k \geq 1,28$; d) $k=-0,44$; e) $k \geq -0,84$; f) $k \geq 1,175$; g) $k=-2,76$; h) $k \geq -0,25$)

Tipificación de la variable:

5. En una distribución $N(18,4)$, hallar las siguientes probabilidades: a) $P(X \leq 20)$ b) $P(X \geq 16,5)$
c) $P(X \leq 11)$ d) $P(19 \leq X \leq 23)$ e) $P(11 \leq X < 25)$

(Soluc: a) 0,6915; b) 0,6480; c) 0,0401; d) 0,2957; e) 0,9198)

6. En una distribución $N(6;0,9)$, calcular k para que se den las siguientes igualdades:
a) $P(X \leq k) = 0,9772$ b) $P(X \leq k) = 0,8$ c) $P(X \leq k) = 0,3$ d) $P(X \geq k) = 0,6331$

(Soluc: a) $k=7,8$; b) $k \geq 6,756$; c) $k \geq 5,532$; d) $k=5,694$)

Problemas de aplicación:

7. Las tallas de los individuos de una población se distribuyen normalmente con media igual a 175 cm y desviación típica igual a 8 cm. Calcular la probabilidad de que un individuo tenga una talla: a) Mayor que 180 cm b) Menor que 170 cm c) Entre 170 y 180 cm

(Soluc: 0,27; ídem; 0,4648)

8. Los opositores que se presentan a unas plazas de cierto organismo se distribuyen normalmente con una puntuación media igual a 70,5 y con una desviación típica igual a 9 ¿Cuántas plazas se adjudicarán en la oposición de este año, si el tribunal ha decidido de antemano dejar sin plaza a todos aquellos que obtengan una puntuación inferior a 80?

(Soluc: consiguen plaza $\cong 14,5$ % de los opositores)

9. En un determinado examen la media de las calificaciones es 6 y la desviación típica 1,2. Calcular la probabilidad de que un alumno tenga una calificación: a) Mayor que 7 b) Menor que 5 c) Entre 5,5 y 7

(Soluc: 0,2633; ídem; 0,46)

10. El Ministerio de Educación ha hecho una encuesta sobre la distribución de las edades del profesorado de Educación Especial, y ha observado que se distribuyen normalmente con media de 38 años y desviación típica 6. De un total de 500 profesores, a) ¿Cuántos profesores hay con edades menores o iguales a 35 años? b) ¿Cuántos mayores de 55 años?

(Soluc: $\cong 154$ profesores; 1 profesor)

11. Los pesos de los individuos de una población se distribuyen normalmente con media 70 kg y desviación típica 6 kg. De una población de 2000 personas, calcular cuántas tendrán un peso comprendido entre 65 y 75 kg. (Soluc: 1187)

12. El peso teórico de una tableta de cierto medicamento es de 324 mg. Si suponemos que los pesos de las tabletas siguen una normal de desviación típica 10 mg por tableta, calcular: a) ¿Cuál será el porcentaje de tabletas con peso menor o igual a 310 mg? b) ¿Cuál será el porcentaje de tabletas con peso superior a 330 mg?

13. La duración media de un determinado modelo de televisor es de ocho años, con una desviación típica de medio año. Si la vida útil del televisor se distribuye normalmente, hallar la probabilidad de que un televisor cualquiera dure más de nueve años.

(Soluc: $\cong 2$ % de los televisores)

14. Se llama cociente intelectual (C.I.) al cociente entre la edad mental y la edad real. Supongamos que la distribución del C.I. de 2000 personas sigue una distribución normal de media 0,8 y desviación típica 0,2. Calcular: a) El número de personas con C.I. superior a 1,4 b) El número de personas con C.I. inferior a 0,9

(Soluc: 2 personas; 1383 personas)

- 15) La duración de un determinado modelo de pila se distribuye según una normal con media 70 horas y desviación típica de 2 horas. A un establecimiento le quedan del pedido anterior 20 pilas. **a)** ¿Cuántas tendrán una duración superior a 70 horas? **b)** ¿Cuántas tendrán una duración entre 75 y 82 horas?
- 16) Por estudios realizados sobre un gran número de niñas al nacer, se ha determinado que su talla se distribuye según una normal de media 50 cm y desviación típica 1,8 cm. **a)** Hallar la probabilidad de que una niña al nacer tenga una talla superior a 54 cm. **b)** Si durante un mes en una maternidad nacen 100 niñas, ¿cuántas tendrán al nacer una talla entre 48,2 y 51,8?
(Soluc: 0,0132; 68 niñas)
- 17) Un almacén de camisas ha determinado que el cuello de los varones adultos se distribuye normalmente con media 38 cm y desviación típica 1,5 cm. Con el fin de poder preparar la producción de la próxima temporada, y teniendo en cuenta que su producción está en 10 000 camisas, ¿cuántas camisas de los números 35, 36, 37, 38 y 39 tendrá que fabricar?
(Soluc: 376 camisas del 35; 1112 del 36; 2120 del 37; 2586 del 38)
18. Continuando con el problema anterior, ¿cuántas camisas habrá que fabricar del 43? ¿Y del 33?
19. En una población de 1 000 individuos se establecen dos grupos, A y B. Los cocientes intelectuales (C.I.) de ambos grupos se distribuyen según $N(100,30)$ y $N(120,35)$, respectivamente. Elegido, aleatoria e independiente, un individuo de cada grupo, se pide:
a) ¿Cuál es la probabilidad de que el individuo del grupo A tenga un C.I. superior a 90?
b) ¿Cuál es la probabilidad de que el individuo del grupo B tenga un C.I. superior a 90?
c) ¿Cuál es la probabilidad de que ambos tengan un C.I. superior a 90?
(Soluc: 0,6293; 0,8051; el producto de las dos anteriores)