Tema 10. Muestreo. Intervalos de confianza

1. Introducción al muestreo

El empleo de encuestas es uno de los métodos de investigación más empleado en la actualidad. Por medio de encuestas se puede conocer la opinión de la gente sobre cualquier asunto: preocupaciones sociales; hábitos de consumo; inclinaciones políticas; gustos deportivos…
Por procedimientos similares se puede determinar la eficacia de un medicamento, la duración en condiciones saludables de un alimento, el tiempo medio dedicado al uso del teléfono móvil o la puntualidad de una compañía aérea ...
Para conocer esas opiniones y hechos se hace necesario seleccionar una parte de la población, lo más pequeña posible, pero representativa del total, en la que sea posible medir las características deseadas. El muestreo estadístico es el instrumento que la Matemática ha generado para tal fin.

Que la muestra de estudio sea lo más pequeña posible es una exigencia de tiempo y de costes; además, el aumento de datos no siempre acarrea una certeza considerablemente mayor. Más importante que los muchos datos es que estén bien elegidos: que sean representativos de la población que se desea estudiar. Por ejemplo, parece lógico que, si se desea estimar la altura media de los hombres de 20 años, no se pregunte por su estatura, exclusivamente, a los individuos que salen de presenciar un partido de baloncesto, pues, aunque no es necesario ser alto para que guste el baloncesto, es más probable que el baloncesto guste a los altos. Esta consideración pone de manifiesto la necesidad de estudiar cómo hay que seleccionar una muestra para que sea representativa. Parece que el azar debe jugar un papel importante en la elección.

Pero el azar comporta riesgo. ¿Qué seguridad se tiene de que los resultados obtenidos por azar sean fiables? ¿Cómo se mide esa seguridad? La “estadística inferencial” mide esa seguridad en términos de probabilidad, lo que lleva implícito cierta inseguridad; así, por ejemplo, cuando se afirma que la probabilidad de acierto es 0,95, se admite que la probabilidad de fallo es 0,05.

1.1. Terminología

Antes de seguir adelante vamos a concretar el significado de algunos términos usuales.

• Población. Es el conjunto total de individuos susceptibles de poseer la información buscada. Aunque se utilicen las palabras "población" e "individuo", no se refiere exclusivamente a personas; por ejemplo, la población puede estar formada por todos los árboles de un bosque, o por todos los coches de una determinada marca fabricados el último año. En todos los casos la población debe estar claramente definida y actualizada.

Ejemplos:
a) En cuestiones demográficas la población viene determinada por el censo de habitantes o por el padrón municipal.
b) En una universidad, la población de estudiantes viene dada por la relación de alumnos matriculados, en ese momento, en los distintos grados o másteres.

• Muestra. Es la parte de la población en la que se miden las características estudiadas.
El número de individuos de la muestra se llama tamaño.
El tamaño de la muestra debe decidirse antes de comenzar la encuesta. Los criterios para determinarlo dependerán de las características de la población y de la fiabilidad que se desee dar al resultado de la encuesta. Conviene aumentar el tamaño si la población es heterogénea y si se quiere mayor precisión en la estimación.

• Muestreo. Es el proceso seguido para la extracción de una muestra. El muestreo puede ser probabilístico o no. En el muestreo probabilístico la muestra se elige por métodos aleatorios; además, permite acotar el posible error y fijar la fiabilidad del trabajo. (Obviamente es el que da más garantías de acierto).

• Encuesta. Es el proceso de obtener la información buscada entre los elementos de la muestra. Por ejemplo, las encuestas a personas pueden ser telefónicas, callejeras, domiciliarias, por internet... La fiabilidad de los resultados dependerá del método utilizado y de cómo se ha seleccionado la muestra de estudio.

1.2. Métodos de muestreo

Los métodos de muestreo probabilístico más frecuentes son los siguientes:

Muestreo aleatorio simple
Este método de muestreo debe satisfacer dos criterios:
1. Cada individuo debe tener la misma probabilidad de ser elegido para la muestra.
2. La selección de un individuo no debe afectar a la probabilidad de que sea seleccionado otro cualquiera. Esto implica que la elección debería hacerse con reemplazamiento; aunque ello comporte que algún elemento pueda ser elegido más de una vez.

Ejemplo:
Si se quiere conocer las aficiones deportivas de los estudiantes de una universidad puede seleccionarse, por sorteo, a cierto número de ellos y preguntarles.
La población de estudio está constituida por los alumnos matriculados en ese momento.
Para asegurar la fiabilidad de los resultados, el número de encuestados, que es el tamaño de la muestra, debe decidirse con antelación.

Muestreo sistemático
En el muestreo sistemático solo se elige un individuo al azar, los restantes se toman según un criterio prefijado. Este muestreo exige ordenar previamente los individuos de la población, después se elige uno de ellos al azar, y a continuación, a intervalos constante, se toman todos los demás hasta completar la muestra.
→ Observación: Este método tiene la ventaja de que es sencillo y rápido, pero vulnera los criterios de aleatoriedad pues una vez elegido el primer individuo, los demás tienen probabilidad 1 o 0 de salir. Además, si la población de partida presenta algún tipo de regularidad, la muestra puede no ser representativa de ella.

Ejemplos:
a) Si en una universidad hay 20000 estudiantes y se desea encuestar a 200 de ellos:
1) Se observa que hay que entrevistar a 1 de cada 100.
2) Se ordenan todos los estudiantes de alguna manera (por orden alfabético; por número de expediente; por el número del carnet de estudiante …). Una vez ordenados se elige al azar uno de ellos entre los 100 primeros de la lista (o entre 100 consecutivos). Si, por ejemplo, ha resultado elegido el número 73, los siguientes encuestados serán los estudiantes correspondientes a los números 173, 273, 373 …

b) En los procesos de calidad de muchas empresas suele utilizarse alguna forma de muestreo sistemático. Así, es frecuente examinar un número determinado de elementos de cada lote producido. (En la industria esos procesos están normalizados: normas ISO y AENOR).

Muestreo estratificado (proporcional)
Este tipo de muestreo divide la población total en clases homogéneas, llamadas estratos. Hecho esto, la muestra se escoge aleatoriamente en número proporcional al de los componentes de cada clase o estrato. (La mayoría de las encuestas a personas se hacen utilizando el muestreo estratificado, eligiendo la muestra con proporción de sexos, de edad, de situación laboral…)

Ejemplo:
Si en la universidad que se está poniendo de ejemplo, los estudiantes pueden distribuirse, por estudios y sexos, como se indica en la siguiente tabla.

Población	Ciencias	Biosanitarios	CC Sociales	Ingenierías	Total
Hombres	1900	1700	2700	3600	9900
Mujeres	2100	2000	2600	3400	10100
Total	4000	3700	5300	7000	20000

Si se quiere hacer un muestreo estratificado de tamaño 200, el número de individuos de cada estrato será el dado en esta otra tabla:

Muestra	Ciencias	Biosanitarios	CC Sociales	Ingenierías	Total
Hombres	19	17	27	36	99
Mujeres	21	20	26	34	101
Total	40	37	53	70	200

Nota: Si se obtuviesen restos decimales habría que redondear.

→ Observación. En el muestreo no probabilístico no se usa el azar, sino el criterio del investigador. Este tipo de muestreo se utiliza con frecuencia en el mundo periodístico para conocer la opinión de oyentes o lectores sobre una cuestión de actualidad. Suele presentar grandes sesgos y es estadísticamente poco fiable. Lo mismo sucede con "redes sociales".

2. Distribuciones de probabilidad
Una distribución de probabilidad es un modelo teórico que trata de explicar el comportamiento de un fenómeno real. Actúa como una función que asigna a cada suceso, cuantificado mediante una variable aleatoria, la probabilidad correspondiente.
La variable aleatoria asocia a cada suceso del espacio muestral un número real. Puede ser discreta o continua. Se llama discreta cuando solo puede tomar ciertos valores aislados. Es continua cuando puede tomar todos los valores de un intervalo.
→ Un caso particularmente importante de distribución discreta es la llamada binomial.
→ La distribución continua más utilizada es la normal.

Ejemplos:
a) El número de veces que hay que lanzar uno de esos dados hasta que salga el número 6 es una variable aleatoria discreta
b) El tiempo que una persona tiene que esperar al autobús es una variable continua.

Función de probabilidad
Una distribución de probabilidad queda determinada a partir de su función de probabilidad (llamada función de densidad, en el caso de las distribuciones continuas), que asigna a cada uno de los valores de la variable aleatoria su probabilidad: $f(x_i) = P\left(X = x_i\right) = p_i$

Función de distribución
A partir de la distribución de probabilidad de la variable X se define la función de distribución, $F(x)$, de dicha variable como sigue: $F(x) = P\left(X \le x\right)$ $\to$ $F(x)$ acumula probabilidades.

3. Distribución binomial

Se emplea cuando el fenómeno de estudio queda determinado por dos sucesos complementarios: si/no; a favor/en contra; … Esos dos sucesos aleatorios (y contrarios) reciben el nombre genérico de éxito y fracaso. Una distribución de estas características también se llama de pruebas de Bernoulli. (Jacob I: suizo, 1654–1705. Es interesante "entrar" en Familia Bernoulli).

La variable aleatoria X, cuenta el número r de éxitos en las n pruebas: $r = 0, 1, \ldots, n$. Por tanto, los valores que puede tomar X son: 0, 1, 2, ..., n.
La distribución binomial queda determinada por los parámetros n y p (número de veces que se realiza el experimento y probabilidad de éxito en cada prueba).
Se indica simbólicamente por $B(n, p)$.

Ejemplo:
Si en una determinada región, la tasa de paro entre su población activa es del 12 %, si se pregunta a 10 personas de esa población, elegidos al azar, por su situación laboral, el número de parados viene descrito por la binomial de parámetros $n = 10$ y $p = 0{,}12$: $B(10,\ 0{,}12)$.

3.1. Probabilidad de *r* éxitos

La función de probabilidad que mide el número r de éxitos cuando una prueba de carácter binomial, $B(n, p)$, se realiza n veces, viene dada por:

$$P(X = r) = \binom{n}{r} p^r \cdot q^{n-r}, \quad r = 0, 1, 2, \ldots, n$$

Recuerda que $\binom{n}{r} = \dfrac{n!}{r!(n-r)!}$ → Por ejemplo: $\binom{8}{3} = \dfrac{8!}{3!(8-3)!} = \dfrac{8\cdot7\cdot6\cdot5!}{3!\cdot5!} = \dfrac{8\cdot7\cdot6}{3\cdot2\cdot1} = 56$.

(El triángulo de Tartaglia facilita el cálculo de algunos números combinatorios, si n es pequeña).

Ejemplo:
Para la binomial $B(10,\ 0{,}12)$ que estudia el número de parados entre 10 personas elegidas al azar cuando la tasa de paro es del 12 %, se tendrá:

$$P(2 \text{ parados}) = P(X = 2) = \binom{10}{2} \cdot 0{,}12^2 \cdot 0{,}88^8 = \frac{10\cdot9}{2} \cdot 0{,}12^2 \cdot 0{,}88^8 = 45 \cdot 0{,}0051787\ldots \approx 0{,}2330.$$

- Para este mismo ejemplo, la probabilidad de que haya menos de 2 parados es:

$$P(X < 2) = P(X = 0) + P(X = 1) = \binom{10}{0} \cdot 0{,}12^0 \cdot 0{,}88^{10} + \binom{10}{1} \cdot 0{,}12^1 \cdot 0{,}88^9 =$$

$$= 1\cdot0{,}88^{10} + 10\cdot0{,}12\cdot0{,}88^9 = 0{,}278500976 + 0{,}379774058 = 0{,}658275034 \approx 0{,}6583.$$

3.2. Media y varianza de la binomial *B*(*n*, *p*)

La media y varianza de la distribución $B(n, p)$ se obtiene a partir de sus parámetros, siendo:
→ Media: $\mu = n \cdot p$ → Varianza: $\sigma^2 = n \cdot p \cdot q$
En consecuencia, la desviación típica vale $\sigma = \sqrt{npq}$.

Ejemplo:
Si se considera la binomial $B(50,\ 0{,}12)$, que puede servir para determinar el número de parados en muestras de tamaño $n = 50$, se tiene:

Media: $\mu = 50 \cdot 0{,}12 = 6$ parados; desviación típica: $\sigma = \sqrt{50\cdot0{,}12\cdot0{,}88} = \sqrt{5{,}28} \approx 2{,}3$.

3.3. Algunos ejercicios de tipo binomial

Ejercicio 1. Si se tira 12 veces una moneda, ¿cuál es la probabilidad de obtener exactamente 8 caras?
Solución:
El experimento es de tipo binomial, con n = 12 y p = 0,5: B(12, 0,5).

$$P(X=8)=\binom{12}{8}\cdot 0,5^8\cdot 0,5^4=\frac{12\cdot 11\cdot 10\cdot 9}{4\cdot 3\cdot 2\cdot 1}0,5^{12}=495\cdot 0,5^2=0,12085.$$

Ejercicio 2. Un examen de tipo test consta de 6 preguntas con 4 posibles respuestas cada una, de las que solo una de ellas es correcta. Si se responde al azar, ¿cuál es probabilidad de que acertar 4 o más preguntas?
Solución:
Es un experimento binomial B(6, 0,25). Por tanto:

$$P(X\geq 4)=P(X=4)+P(X=5)+P(X=6)=\binom{6}{4}\cdot 0,25^4\cdot 0,75^2+\binom{6}{5}\cdot 0,25^5\cdot 0,75+\binom{6}{6}\cdot 0,25^6$$

$$=\frac{6\cdot 5}{2}0,25^4\cdot 0,75^2+6\cdot 0,25^5\cdot 0,75+0,25^6=0,03296+0,00439+0,00024=0,03759\ .$$

Ejercicio 3. En una ciudad, el 15 % de sus ciudadanos tocan algún instrumento musical. Si se eligen 8 personas al azar, ¿cuál es la probabilidad de que al menos 2 de ellas toquen un instrumento?
Solución:
Se trata de una binomial B(8, 0,15). Por tanto:

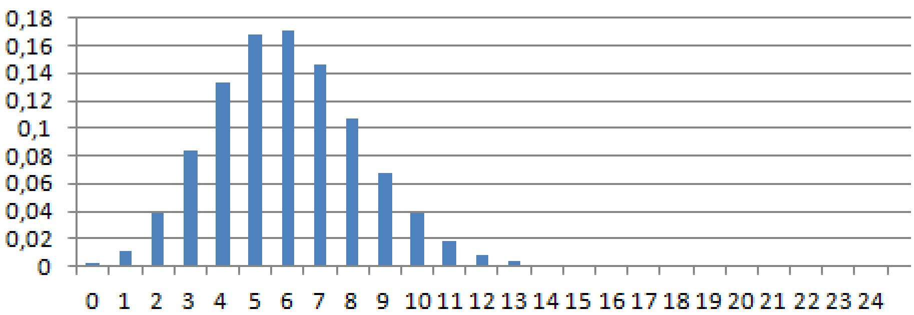
$$P(X\geq 2)=1-P(X=0)-P(X=1)=$$

$$=1-\binom{8}{0}\cdot 0,15^0\cdot 0,85^8-\binom{8}{1}\cdot 0,15^1\cdot 0,85^7=1-0,27249-0,38469=0,34282.$$

Observación. Para valores grandes de n, la probabilidad de cada uno de los posibles sucesos (de un número r de éxitos) es muy pequeña, sobre todo para valores de X alejados de la media. Así, por ejemplo, para la binomial B(50, 0,12), pueden darse las siguientes probabilidades:

$$P(X=2)=0,03816514;\ P(X=8)=0,10754701;\ P(X=12)=0,0084088.$$

En el gráfico de barras que sigue se dan las probabilidades de cada uno de los sucesos asociados a la B(50, 0,12). En el eje OX se indican los valores de X; en OY sus probabilidades.



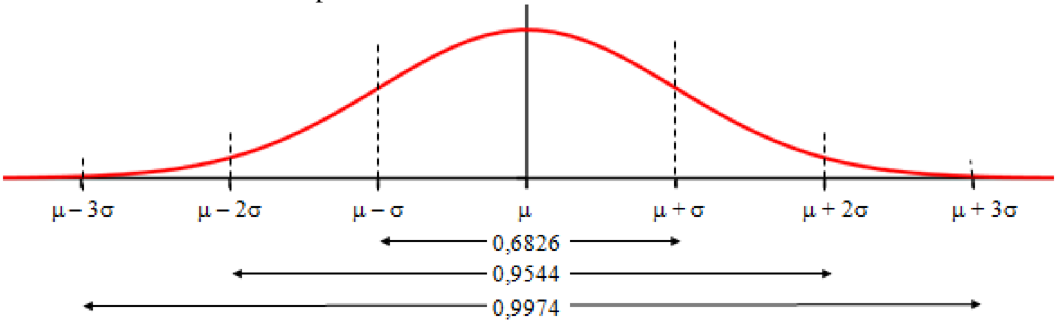
- En el apéndice de Recursos informáticos se dan las pautas para obtener los valores de probabilidad y para confeccionar el gráfico.

4. Distribución de probabilidad normal

Es una distribución de probabilidad continua, asociada (teóricamente) a multitud de fenómenos naturales y cotidianos, que se caracterizan porque la mayoría de los resultados tienden a agruparse en torno a su media. (Altura de las personas; tamaño de los frutos de una planta; …)
Una variable con distribución normal queda definida por su media μ y por su desviación típica σ. Se denota como $N(\mu, \sigma)$.

→ La expresión de la función de densidad de la distribución normal es: $f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$.

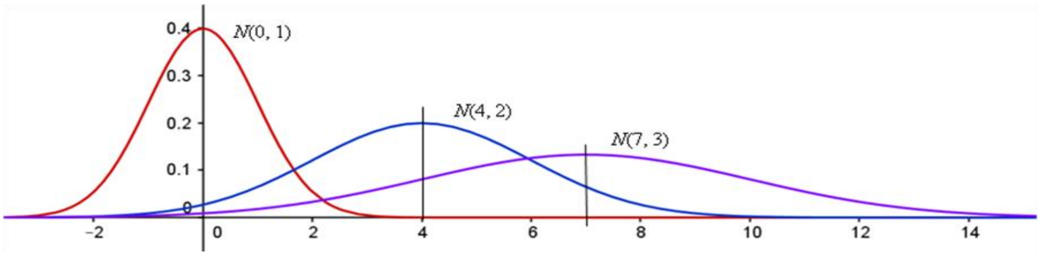
Su gráfica es la conocida "campana de Gauss".



Esta función cumple las siguientes propiedades:

- Está definida para todo número real, es decir, en el intervalo $(-\infty, +\infty)$: la variable puede tomar cualquier valor; siendo $f(x)>0$ para todo x.
- El área por debajo de la curva vale 1. Esto es: $\displaystyle\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx=1$.
- Es simétrica respecto a su media μ.
- El eje de abscisas es una asíntota de la curva.
- Aunque la variable puede tomar cualquier valor entre $-\infty$ y $+\infty$, la probabilidad de que tome valores alejados de la media es prácticamente nula, pues se cumple que: el área delimitada por la curva y el eje *OX* entre $\mu-\sigma$ y $\mu+\sigma$ es 0,6826; entre $\mu\pm 2\sigma$ es de 0,9544; y entre $\mu\pm 3\sigma$ es de 0,9974. A los valores que están a una distancia superior a $3{,}5\sigma$ de μ se les asigna probabilidad 0.

→ La variación de la media y de la desviación típica originan cambios en la curva, desplazándose a izquierda o derecha o haciéndose más esbelta o más baja, como puede verse en la figura.



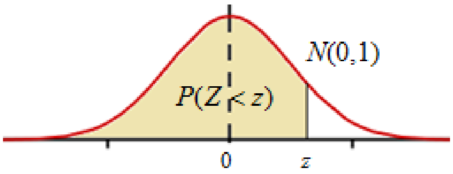
Recuérdese que la desviación típica es una medida de la dispersión de los elementos de una población. Una desviación típica más grande significa que los datos son más heterogéneos; por eso las curvas normales con mayor desviación típica son más planas, que es un indicador de que los datos pueden estar más alejados de la media. (Cuando la igualdad entre los datos es grande, la desviación típica es pequeña; y al revés).

4.1. Distribución normal de media 0 y desviación típica 1: *N*(0, 1)

El comportamiento estadístico normal hace que puedan asignarse valores de probabilidad a cualquier suceso de la variable estudiada. Esto es, se puede saber (pues está tabulado) la probabilidad de que la variable tome valores pertenecientes a cualquier intervalo dado.

→ Lo que está tabulado es la función de distribución en el caso de la curva normal de media μ = 0 y desviación típica σ = 1, la normal *N*(0, 1).

La función de distribución, $F(x)$, da la superficie del recinto limitado por la función de densidad, curva $y = f(x)$ definida más arriba, y el eje *OX*, desde –∞ hasta un valor determinado *z*, esto es



$$P(Z < z) = F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx .$$

Esta área da la probabilidad de que la variable *Z*, tome valores menores que *z*. (Cuando se trata de la *N*(0, 1), la variable aleatoria suele designarse por la letra *Z*).

4.2. Tabla normal estándar: *N*(0, 1)

En la tabla normal *N*(0, 1) se da la probabilidad de que la variable *Z* tome valores entre –∞ y +*z*: $P(Z < z)$; los demás valores se obtienen teniendo en cuenta la simetría de la curva y que el área por debajo de la curva vale 1. Así, a partir del valor $P(Z < z)$, pueden obtenerse los valores de $P(Z > z)$, $P(Z < -z)$, $P(Z > -z)$, $P(0 < Z < z)$ y $P(-Z < z < Z)$.

En esta tabla la cifra de las unidades y de las décimas se muestran en la columna de la izquierda, la de las centésimas en la fila superior.

Los valores de *Z* están dados en desviaciones típicas: *Z* = 1, indica un valor superior en una desviación típica a la media; *Z* = 1,26 indica 1,26 desviaciones típicas más que la media. *Z* = 0 indica que la variable se sitúa justamente en la media.

Z	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
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

Ejemplos:

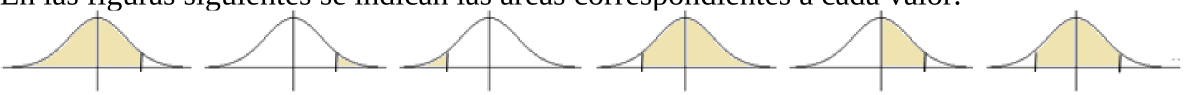
La probabilidad de que *Z* < 1,26, $P(Z < 1,26) = 0,8962$.

A partir de ese valor se deducen:

$P(Z > 1,26) = 1 - 0,8962 = 0,1038$; $P(Z < -1,26) = 0,1038$; $P(Z > -1,26) = 0,8962$;

$P(0 < Z < 1,26) = 0,8962 - 0,5 = 0,3962$; $P(-1,26 < Z < 1,26) = 2 \cdot 0,3962 = 0,7924$.

En las figuras siguientes se indican las áreas correspondientes a cada valor.

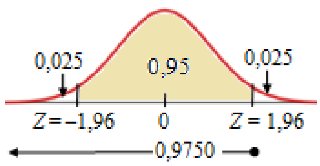


$P(Z < 1,26)$ $P(Z > 1,26)$ $P(Z < -1,26)$ $P(Z > -1,26)$ $P(0 < Z < 1,26)$ $P(-1,26 < Z < 1,26)$

4.3. Cálculo del valor de *Z* a partir de su probabilidad asociada

La tabla normal se emplea también en sentido contrario, para hallar la abscisa (el valor de *Z*) correspondiente a una probabilidad determinada.
Esto es, igual que se sabe que $P(Z < 1) = 0{,}8413$, en sentido contrario la pregunta sería: ¿cuánto debe valer z para que $P(Z < z) = 0{,}8413$? La respuesta es evidente: el valor de z debe ser 1.

Ejemplo:
Un caso que se presenta con frecuencia es encontrar el intervalo $(-z, z)$ que contiene el 95% de los datos de la variable estadística.
Esto es, hallar el valor z tal que $P(-z < Z < z) = 0{,}95$.



Por el ejemplo anterior:

$$P(-z < Z < z) = 0{,}95 \Rightarrow 2 \cdot P(Z < z) - 1 = 0{,}95 \Rightarrow P(Z < z) = 0{,}9750 \Rightarrow z = 1{,}96.$$

→ Si el valor de probabilidad no figura en la tabla se tomará el más cercano. También puede optarse por la interpolación. Así, para $P(Z < z) = 0{,}9950$ se toma $z = 2{,}575$, intermedio entre 2,57 y 2,58, cuyos valores de probabilidad respectivos son 0,9949 y 0,9951.

4.4. Tipificación

Las distribuciones normales con las que se trabaja en la práctica no son la estándar: la $N(0, 1)$.
Son distribuciones con media μ (la que sea) y desviación típica σ: $N(\mu, \sigma)$.
Para calcular valores de probabilidad de una variable *X*, normal $N(\mu, \sigma)$, se hace el cambio de variable $Z = \dfrac{X - \mu}{\sigma}$, siendo *Z* la $N(0, 1)$.
Con esto, las diferencias de *X* respecto de su media μ se "tipifican": se calculan en desviaciones típicas, pues lo significativo no es el valor que tome *X*, sino cuántas desviaciones típicas es mayor o menor que su media.
Este cambio permite hallar valores de probabilidad para cualquier la variable $X \approx N(\mu, \sigma)$, pues:

$$P(X < k) = P\left(\frac{X - \mu}{\sigma} < \frac{k - \mu}{\sigma}\right) = P\left(Z < \frac{k - \mu}{\sigma}\right),$$ donde *Z* es $N(0, 1)$.

Ejemplos:
a) Si *X* es una variable normal $N(45, 7)$, la probabilidad de que *X* tome valores menores de 52, mayores de 52; o entre 52 y 59 es:

$$P(X < 52) = P\left(Z < \frac{52 - 45}{7}\right) = P(Z < 1) = 0{,}8413;$$

$$P(X > 52) = 1 - P(Z < 52) = 1 - P\left(Z < \frac{52 - 45}{7}\right) = 1 - 0{,}8413 = 0{,}1587;$$

$$P(52 < X < 59) = P\left(\frac{52 - 45}{7} < Z < \frac{59 - 45}{7}\right) = P(1 < Z < 2) = P(Z < 2) - P(Z - 1) =$$

$$= 0{,}9772 - 0{,}8413 = 0{,}1359.$$

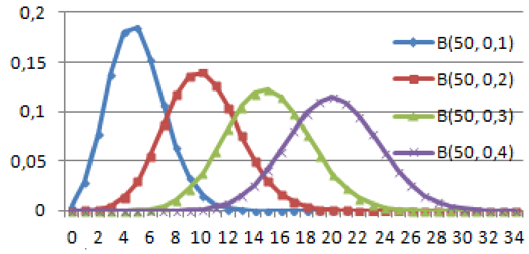
b) En sentido inverso, también se puede encontrar el valor de *X* a partir de la probabilidad asociada. Esto es, para la misma $N(45, 7)$, si se desea conocer el valor de k tal que, por ejemplo, $P(X < k) = 0{,}9772$, se resuelve

$$P(X < k) = P\left(Z < \frac{k - 45}{7}\right) = 0{,}9772 \Rightarrow \frac{k - 45}{7} = 2 \Rightarrow k = 45 + 2 \cdot 7 = 59.$$

Nota: Existen Recursos informáticos (p. 15) que facilitan estos cálculos.

5. Aproximación de la distribución binomial mediante una normal

En la figura adjunta puede observarse la evolución de las poligonales de frecuencias de la distribución binomial $B(50, p)$ cuando p se acerca a 0,5. Cada vez se aproximan más a una curva normal.



$\rightarrow$ En general, se admite que la aproximación es buena cuando $n \geq 25$ y el producto $np \geq 5$.
En ese caso, la distribución normal que mejor aproxima a la $B(n, p)$ es la que tiene por media y desviación típica la de la distribución binomial.
Como la media y desviación típica de la variable $X \approx B(n, p)$ son $\mu = np$ y $\sigma = \sqrt{npq}$, el ajuste se hace por la variable $X' \approx N\left(np,\ \sqrt{npq}\right)$; que se tipifica haciendo $Z = \dfrac{X'-\mu}{\sigma} = \dfrac{X'-np}{\sqrt{npq}}$.

$\rightarrow$ Con esto, las probabilidades asociadas a la variable $X \approx B(n, p)$ se calculan como sigue:

$$P(X < k) = P(X' < k) = P\left(Z < \frac{k-np}{\sqrt{npq}}\right).$$

Ejemplo:
La variable binomial $X = B(50, 0,3)$ se estudia mediante la variable normal $X' = N(15,\ 3{,}24)$.

Así, por ejemplo: $P(X < 18) = P(X' < 18) = P\left(Z < \dfrac{18-15}{3{,}24}\right) = P(Z < 0{,}93) = 0{,}8238$.

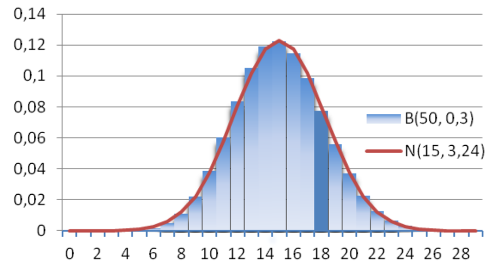
6.1. Corrección de continuidad

En la distribución normal la probabilidad de que la variable tome un valor concreto es 0, pero en una binomial no es así. Por seguir con el ejemplo anterior:

- Para la normal $X' = N(15,\ 3{,}24)$, $P(X' < 18) = P(X' \leq 18)$, pues $P(X' = 18) = 0$.
- Para la binomial $X = B(50, 0,3)$, $P(X < 18) < P(X \leq 18)$, pues $P(X = 18) = 0{,}077247062$,

que es el resultado de $P(X = 18) = \binom{50}{18} \cdot 0{,}3^{18} \cdot 0{,}7^{22}$.

Este valor coincide con el área de la barra sobre $X = 18$ (sombreada con mayor intensidad), que se muestra en el gráfico adjunto.



Esta anomalía puede paliarse mediante la llamada corrección de continuidad, por la que probabilidades puntuales (de valor 0 en las distribuciones continuas) son sustituidas por probabilidades de intervalo de la forma siguiente:

$$P(X = k) = P(k - 0{,}5 < X' < k + 0{,}5);\quad P(X \leq k) = P(X' < k + 0{,}5);$$

$$P(X < k) = P(X' < k - 0{,}5);\quad P(X \geq k) = P(X' > k - 0{,}5);\quad P(X > k) = P(X' > k + 0{,}5).$$

Ejemplo:
Para la binomial $X = B(50, 0,3) \approx X' = N(15,\ 3{,}24)$, se tiene:

$$P(X = 18) = P(17{,}5 < X' < 18{,}5) = P\left(\frac{17{,}5-15}{3{,}24} < Z < \frac{18{,}5-15}{3{,}24}\right) =$$

$$= P(Z < 1{,}08) - P(Z < 0{,}77) = 0{,}8599 - 0{,}7794 = 0{,}0805.$$

6. Distribuciones de las medias muestrales

En esta distribución los elementos no son cada uno de los individuos de la población, sino las medias de cada una de las posibles muestras. Esto es, si se consideran muestras de tamaño *n*, y los elementos de la muestra M_1 son $x_{11}, x_{12}, ..., x_{1n}$, el elemento $\bar{X}_1$ de esta distribución es

$\bar{X}_1 = \dfrac{x_{11} + x_{12} + ... + x_{1n}}{n}$; y, en general, para la muestra M_i, será $\bar{X}_i = \dfrac{x_{i1} + x_{i2} + ... + x_{in}}{n}$.

Cuando una población tiene muchos elementos el número de sus muestras puede resultar enorme.

Así, por ejemplo, en una población de 120 individuos el número de muestras distintas de tamaño 30 que pueden extraerse sin repetir elementos, es $\binom{120}{30} = 1,697 \cdot 10^{28}$ → un número de 29 cifras.

Y si 1los elementos pueden repetirse, que sería el caso del muestreo aleatorio simple, su número será $\binom{149}{30} \approx 2,58 \cdot 10^{31}$: las combinaciones con repetición de 120 elementos tomados 30 a 30.

La media de las medias muestrales es $\bar{X} = \dfrac{\bar{X}_1 + \bar{X}_2 + ... + \bar{X}_i + ...}{\text{nº total de muestras}}$.

Pues bien, si la población de partida es una distribución normal de media μ y desviación típica σ, *N*(μ, σ), la distribución de las medias muestrales extraídas de ella, se comporta como una normal de media $\bar{X} = \mu$ y desviación típica $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$.

Esto es:

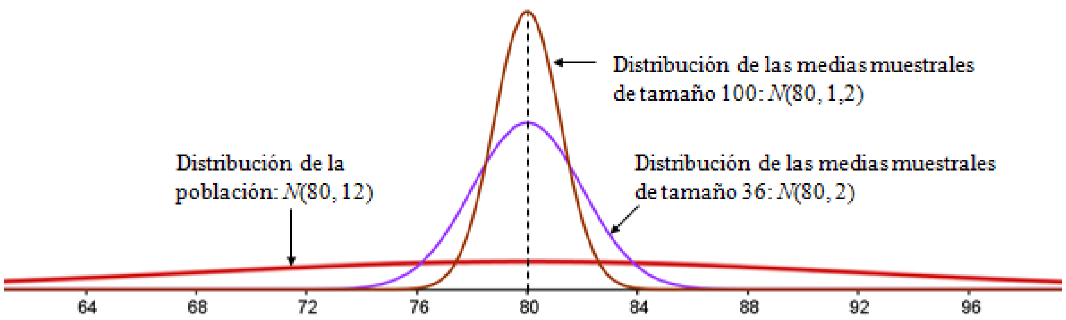
si la población es *N*(μ, σ),

la distribución de las medias muestrales de tamaño *n* es $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.

Ejemplo:

En una población *N*(80, 12), la media de las medias muestrales de tamaño 36 se distribuye según la normal $N\left(80, \dfrac{12}{\sqrt{36}}\right)$ = *N*(80, 2).

Asimismo, la media de las medias muestrales de tamaño 100 se distribuye según la normal $N\left(80, \dfrac{12}{\sqrt{100}}\right)$ = *N*(80, 1,2).



Como puede verse en la figura, si el tamaño muestral aumenta la campana se estrecha. Esto genera mayor precisión en las estimaciones.

6.1. Valores de probabilidad en las muestras

Que las medias muestrales se distribuyan normalmente permite determinar los v alores de probabilidad de cualquier media elegida por métodos aleatorios de una población normal. Basta con tipificar dichos valores y usar la $N(0, 1)$.

La distribución de medias muestrales, $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$, se tipifica haciendo el cambio $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$.

Ejemplo:

Supongamos que la altura de una determinada raza de ovejas se distribuye normalmente con media 80 cm y desviación típica 12 cm: $N(80, 12)$. Para esa raza de ovejas se pueden obtener los valores de probabilidad es los siguientes casos:

a) De que una oveja elegida al azar mida más de 83 cm de alta.

b) De que la altura media de una muestra aleatoria de 36 ovejas supere los 83 cm.

c) De que la altura media de una muestra de 100 ovejas supere los 83 cm.

→ Para dar respuesta a esas preguntas hay que proceder como sigue:

a) Los elementos individuales de la población se ajustan a la normal $N(80, 12)$, que se tipifica haciendo $Z = \dfrac{x-\mu}{\sigma} = \dfrac{x-80}{12}$. Por tanto:

$$P(x>83) = P\left(Z > \frac{83-80}{12}\right) = P(Z>0,25) = 1 - P(Z<0,25) = 1 - 0,5987 = 0,4013.$$

b) Las muestras de tamaño 36 se distribuyen según la normal $N\left(80, \dfrac{12}{\sqrt{36}}\right) = N(80, 2)$; que se tipifica haciendo $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \dfrac{\bar{X}-80}{12/\sqrt{36}} = \dfrac{\bar{X}-80}{2}$. Por tanto:

$$P(\bar{X}>83) = P\left(Z > \frac{83-80}{2}\right) = P(Z>1,5) = 1 - P(Z<1,5) = 1 - 0,9332 = 0,0668.$$

c) Las muestras de tamaño 100 se distribuyen según la normal $N\left(80, \dfrac{12}{\sqrt{100}}\right) = N(80, 1,2)$; que se tipifica haciendo $Z = \dfrac{\bar{X}-80}{1,2}$. Por tanto:

$$P(\bar{X}>83) = P\left(Z > \frac{83-80}{1,2}\right) = P(Z>2,5) = 1 - P(Z<2,5) = 1 - 0,9938 = 0,0062.$$

<u>Comentarios</u>:

1. Observa que no sería extraño encontrar una oveja con altura superior a 83 cm, pero es muy improbable que la altura media de 100 ovejas, elegidas aleatoriamente, supere los 83 cm.

2. Si una muestra está bien realizada y tiene el tamaño suficiente, su media estará muy próxima a la media real de la población de partida. Este es el fundamento de la inferencia estadística, lo que da consistencia a sus resultados; aunque una muestra solo proporciona la seguridad puntual de su media o de de su desviación típica. (En el apartado siguiente se concretará un poco más).

3. Lo importante es que las medias muestrales tienen un comportamiento que se ajusta a una normal: se distribuyen normalmente.

4. La distribución de las medias muestrales es normal incluso en el caso de que estas procedan de poblaciones no normales, siempre que el tamaño de la muestra sea grande ($n \geq 30$).

5. El fundamento matemático de lo dicho forma parte del *teorema central del límite*.

7. Intervalo de confianza para la media de la población

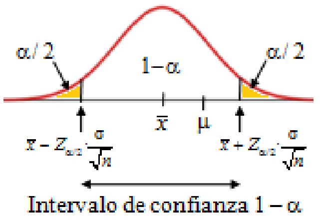
La media, $\bar{x}$, de los elementos de una muestra es una estimación puntual de la media, μ, de la población de partida. Esto significa que la media real está próxima a $\bar{x}$; pero: ¿cuánto de próxima?, y ¿qué seguridad se tiene de que sea así?
Por eso la estimación no se hace solo en términos puntuales, dando $\bar{x}$, si no definiendo un intervalo alrededor de la media muestral e indicando la probabilidad que se tiene de sea realmente así.
Ese intervalo se llama de intervalo de confianza. A la probabilidad de que tal estimación sea cierta se la llama nivel de confianza.

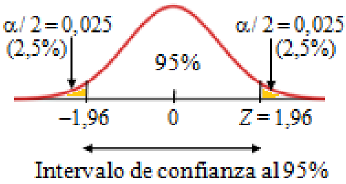
7.1. Intervalo de confianza para la media

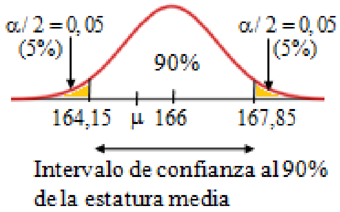
El intervalo de confianza de la media poblacional que se obtiene a partir de una muestra de media $\bar{x}$, es: $\left(\bar{x} - Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right)$, siendo σ la desviación típica poblacional, n el tamaño muestral y $Z_{\alpha/2}$ el valor correspondiente en la tabla normal, $N(0, 1)$, para un nivel de confianza de $1 - \alpha$.

Por tanto: $P\left(\bar{x} - Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$.

El nivel de confianza $1 - \alpha$, da la probabilidad que hay de que la media poblacional, μ, pertenezca a ese intervalo; es frecuente darla en porcentajes. Son equivalentes las expresiones "nivel de confianza $1 - \alpha$" y "significación α"; así, suele hablarse de obtener un intervalo con un nivel de confianza del 95% ($1 - \alpha = 0,95$) o para una significación $\alpha = 0,05$. (Este valor de α indica la probabilidad que hay de errar en la estimación).







Ejemplo:
Se sabe que la desviación típica de la estatura de las chicas de 18 años es 9 cm. Si se toma una muestra aleatoria de tamaño 64, en la que se obtiene una estatura media de $\bar{x} = 166$, el intervalo de confianza para un nivel del 90 % de la estatura media de la población será:

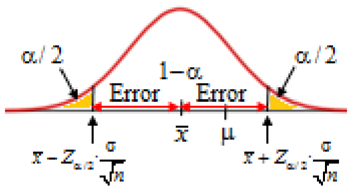
$$\left(166 - 1,645\cdot\frac{9}{\sqrt{64}},\quad 166 + 1,645\cdot\frac{9}{\sqrt{64}}\right) = (166 - 1,85,\ 166 + 1,85) = (164,15,\ 167,85).$$

Esto significa que la media de estaturas para toda la población, μ, estará entre 164,15 cm y 167,85 cm, con una probabilidad de 0,90: $P(164,15 < \mu < 167,85) = 0,90$.
Puede suceder que la verdadera media sea mayor que 167,85 cm, pero la probabilidad de eso es escasa: $P(\mu > 167,85) = 0,05$ ($\to \alpha/2 = 0,05$).
También puede suceder que la verdadera media sea menor que 164,15 cm, pero con una probabilidad igualmente pequeña: $P(\mu < 164,15) = 0,05$ ($\to \alpha/2 = 0,05$).
Observación: Para una confianza del 90 % $\Rightarrow 1 - \alpha = 0,90 \to 1 - \alpha/2 = 0,9500$. Hay que buscar en la tabla normal ese valor, que no aparece: está entre $Z = 1,64$ y $Z = 1,65$. Se toma 1,645.

7.2. Error admitido

Al dar el intervalo de confianza para la media se afirma, con una confianza de $1-\alpha$, que

$$\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$



Esto significa que, en la estimación, se admite un error máximo de $E = Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$.

→ El error depende de σ, α y *n*: aumenta al hacerlo la desviación típica y la confianza; disminuye cuando aumenta el tamaño muestral. Por tanto:

1) Para una confianza dada, si se quiere disminuir el error hay que aumentar el tamaño muestral.

2) Para un mismo tamaño muestral, si se desea mayor seguridad, mayor confianza en el resultado, hay que asumir un error mayor.

Ejemplo:

En el ejemplo anterior (σ = 9 cm, $1-\alpha = 0{,}90 \to Z_{0{,}05} = 1{,}645$ y $n = 64$), el error máximo es:

$$E = 1{,}645\cdot\frac{9}{\sqrt{64}} = 1{,}85 \text{ cm.}$$

→ Como la estimación puntual es $\bar{x} = 166$ cm, podría afirmarse que $\mu = 166 \pm 1{,}85$ cm.

- Si el tamaño muestral aumentase a $n = 100$, el error será: $E = 1{,}645\cdot\dfrac{9}{\sqrt{100}} = 1{,}48$ cm.
- Para $n = 64$ y una confianza del 95%, ($\alpha = 0{,}05$; $Z_{\alpha/2} = 1{,}96$): $E = 1{,}96\cdot\dfrac{9}{\sqrt{64}} \approx 2{,}2$ cm. En

este caso, $\mu \in (166 - 2{,}2,\ 166 + 2{,}2) = (163{,}8,\ 168{,}2)$, con una probabilidad de 0,95.

7.3. Tamaño muestral

Si inicialmente se fija el error máximo (*emax*) admisible, puede calcularse el mínimo tamaño muestral necesario para tal propósito, pues basta con despejar. Así:

$$E = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < emáx \Rightarrow Z_{\alpha/2}\frac{\sigma}{(emáx)} < \sqrt{n} \Rightarrow n > \left(Z_{\alpha/2}\frac{\sigma}{(emáx)}\right)^2$$

Ejemplo:

a) Si en el ejemplo anterior se quiere estimar la media poblacional, μ, con un error máximo de 1,5 cm y con un nivel de confianza del 95 %, ¿cuál será el tamaño muestral necesario para determinarla?

→ Se desea que $E = Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} < 1{,}5$. Como $Z_{0{,}025} = 1{,}96$ y σ = 9 $\Rightarrow$ $n > \left(1{,}96\cdot\dfrac{9}{1{,}5}\right)^2 = 138{,}3$.

Por tanto, habrá que encuestar a 139 personas de esa población, al menos.

b) Si en el mismo ejemplo se desea un error máximo de 1 cm, con un nivel de confianza del 80% (significación α = 0,20, $Z_{0{,}10} = 1{,}28$), se tendría: $1{,}28\cdot\dfrac{9}{\sqrt{n}} < 1 \Rightarrow n > (1{,}28\cdot 9)^2 = 132{,}7$.

El tamaño mínimo de la muestra debe ser $n = 133$.

Observación: Si se desconoce la desviación típica poblacional, entonces se emplea la desviación típica de la muestra. (En rigor habría que emplear la "cuasidesviación" típica de la muestra, cuyo valor se obtiene pulsando la tecla σ_{n-1} de la calculadora. Como $\sigma_{n-1} > \sigma_n$, el intervalo de confianza se amplia y da más seguridad. Aquí se utilizará siempre σ).

8. Distribución de la proporción de las muestras

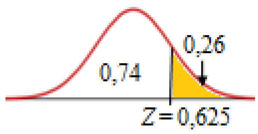
Cuando se trata de determinar la proporción de una población que posee cierto atributo (está de acuerdo/no está de acuerdo; vota al partido A/no vota al partido A; éxito/fracaso), su estudio es equiparable al de una distribución binomial. Así pues, si la probabilidad de éxito en la población es p, la de fracaso será $q = 1 - p$, y se toman muestras aleatorias de tamaño n, entonces la proporción, $\hat{p}$, y la desviación típica, $\sigma_{\hat{p}}$, de la muestra se distribuyen aproximadamente como una normal de media $\mu = p$ y $\sigma_p = \sqrt{\dfrac{pq}{n}}$: $N\left(p,\ \sqrt{\dfrac{pq}{n}}\right)$.

Recuerda que los parámetros de las proporciones muestrales de tamaño n se obtienen dividiendo los parámetros binomiales entre n:
Distribución binomial $X \approx B(n, p)$:
Su media y desviación típica son $\mu = np$ y $\sigma = \sqrt{npq}$. Se aproxima por $N\left(np,\ \sqrt{npq}\right)$.
Distribución muestral: $\hat{p} = \dfrac{np}{n}$ y $\sigma_{\hat{p}} = \dfrac{\sqrt{npq}}{n} = \sqrt{\dfrac{pq}{n}}$. Se aproxima por $N\left(p,\ \sqrt{\dfrac{pq}{n}}\right)$.

Ejemplo:
Supongamos que en una comunidad autónoma el porcentaje de familias con dos o más hijos es el 60 %. Si en esa comunidad se toma una muestra de 36 familias, se pueden obtener los valores de probabilidad en los siguientes casos:
a) De que el porcentaje de familias (en la muestra) con dos o más hijos sea superior al 65 %.
b) De que el porcentaje de familias con dos o más hijos esté entre el 57 % y el 63 %.
c) De que al menos 18 de esas familias tenga dos o más hijos.
→ Para dar respuesta a esas preguntas hay que proceder como sigue:
La proporción de la población es $p = 0{,}60$, luego $q = 0{,}40$. Por tanto, las muestras de tamaño 36, se distribuyen según la normal $N\left(0{,}6,\ \sqrt{\dfrac{0{,}6 \cdot 0{,}4}{36}}\right) \to N(0{,}6,\ 0{,}0816) \approx N(0{,}6,\ 0{,}08)$.

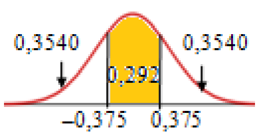
Se tipifica haciendo $Z = \dfrac{\hat{p} - p}{\sigma_p} = \dfrac{\hat{p} - 0{,}60}{0{,}08}$. Por tanto:

a) $P(\hat{p} > 0{,}65) = P\left(Z > \dfrac{0{,}65 - 0{,}60}{0{,}08}\right) = P(Z > 0{,}625) =$
$= 1 - P(Z < 0{,}625) = 1 - 0{,}740 = 0{,}26$.



Por tanto, el 26 % de las muestras de tamaño 36 dará una proporción superior a 0,65 (65 %).

b) $P(0{,}57 < \hat{p} < 0{,}63) = P\left(\dfrac{0{,}57 - 0{,}60}{0{,}08} < Z < \dfrac{0{,}63 - 0{,}60}{0{,}08}\right) =$
$= P(-0375 < Z < 0{,}375) = 2P(Z < 0{,}375) - 1 = 2 \cdot 0{,}6460 - 1 = 0{,}292$.



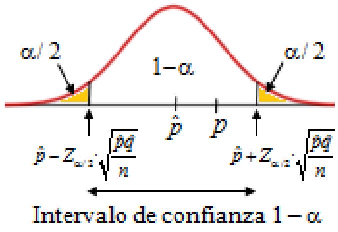
c) Si al menos 18 familias tienen más de 2 hijos, entonces $\hat{p} > \dfrac{18}{36} = 0{,}5$.

$$P(\hat{p} > 0{,}5) = P\left(Z > \dfrac{0{,}5 - 0{,}60}{0{,}08}\right) = P(Z > -0{,}125) = P(Z < 0{,}125) = 0{,}5498.$$

8.1. Intervalo de confianza para la proporción

El intervalo de confianza para la proporción de la población, que se obtiene a partir de una muestra de tamaño *n*, con un nivel de confianza de 1 – α, es:

$$\left(\hat{p}-Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}},\ \hat{p}+Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right),$$



siendo $\hat{p}$ la proporción de la muestra, $\hat{q}=1-\hat{p}$, y $Z_{\alpha/2}$ el valor correspondiente en la tabla *N*(0, 1) para una significación α.

Por tanto: $P\left(\hat{p}-Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}<p<\hat{p}+Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)=1-\alpha$.

→ La proporción de la muestra, $\hat{p}$, es una estimación puntual de la proporción real.

Ejemplo:
Para conocer, con una probabilidad del 95 %, el grado de satisfacción de los ciudadanos en relación con los servicios municipales, se pregunta a 250 vecinos de la localidad en cuestión, de los que 85 dicen estar satisfechos.

Con esto, como $\hat{p}=\frac{85}{250}=0,34$ ($\hat{q}=1-0,34=0,66$) y $Z_{\alpha/2}$ = 1,96 (1 – α = 0,95 → 95 %), el intervalo de confianza para la proporción será:

$$\left(0,34-1,96\sqrt{\frac{0,34\cdot 0,66}{250}},\ 0,34+1,96\sqrt{\frac{0,34\cdot 0,66}{250}}\right)\approx (0,34 - 0,0587,\ 0,34 + 0,0587) \approx$$

≈ (0,34 – 0,06, 0,34 + 0,06) = (0,28, 0,40) → Se ha redondeado a centésimas.

→ Como la estimación puntual es $\hat{p}$ = 0,34, podría afirmarse que *p* = 0,34 ± 0,06. Esto significa que el porcentaje "real" de ciudadanos satisfechos estará entre el 28 % y el 40 %.
(La probabilidad de que sea inferior al 28 % o superior al 40% es 0,025 en ambos casos: $P(p<0,28)=P(p>0,40)=0,025$).

Observaciones:

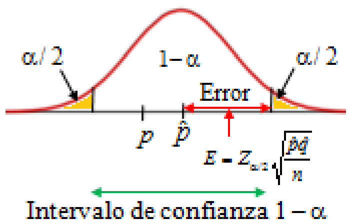
1. La estimación se considera válida si *np* y *nq* son mayores que 5, lo cual se consigue cuando *n* es suficientemente grande.

2. Si se conociese *p* por estudios anteriores, se emplea la desviación típica $\sqrt{\frac{pq}{n}}$ en vez de $\sqrt{\frac{\hat{p}\hat{q}}{n}}$, siendo el intervalo de confianza: $\left(\hat{p}-Z_{\alpha/2}\sqrt{\frac{pq}{n}},\ \hat{p}+Z_{\alpha/2}\sqrt{\frac{pq}{n}}\right)$.

3. En estudios profesionales (encuestas) suele asignarse inicialmente *p* = 0,5 y *q* = 0,5; con esto se consigue un intervalo más amplio, lo que da más seguridad a la estimación.
→ El producto *pq*, con *p* + *q* = 1, es máximo cuando *p* = *q* = 0,5.
En efecto, el máximo de la función $f(p)=pq=p(1-p)=p-p^2$ se obtiene cuando $f'(p)=0$ y $f''(p)<0$.
Derivando con respecto a *p* e igualando a 0: $f'(p)=1-2p=0\Rightarrow p=0,5$.
Como $f''(p)=-2<0$, para ese valor se obtiene el máximo de $f(p)=pq$.

Pero si pq es máximo cuando *p* = *q* = 0,5, también es máximo el valor de $\sqrt{\frac{pq}{n}}$; lo que hace que el intervalo de confianza se amplíe.

8.2. Error admitido



El error máximo admitido es: $E = Z_{\alpha/2}\sqrt{\dfrac{\hat{p}\hat{q}}{n}}$.

Puede controlarse variando la significación α y el tamaño muestral *n*.

Ejemplos:

a) En el caso del ejemplo anterior: $\hat{p}$ = 0,34, 1 − α = 0,95 ($Z_{\alpha/2} = Z_{0,025}$ = 1,96) y *n* = 250, el error máximo es: $E = 1{,}96\sqrt{\dfrac{0{,}34 \cdot 0{,}66}{250}}$ = 0,0587 ≈ 0,06.

Luego la proporción de la población será *p* = 0,34 ± 0,06, con una probabilidad de 0,95.

b) Si se varía el tamaño muestral y el nivel de confianza el error cambia. Por ejemplo, para *n* = 400 y un nivel de confianza del 92 %, 1 − α = 0,92 ⇒ $E = 1{,}75\sqrt{\dfrac{0{,}34 \cdot 0{,}66}{400}}$ ≈ 0,041.

8.3. Tamaño muestral

Si se fija el error máximo (= *emáx*) deseado, se puede deducir el mínimo tamaño muestral *n*, pues: $E = Z_{\alpha/2}\sqrt{\dfrac{\hat{p}\hat{q}}{n}} < emáx \Rightarrow n > \left(Z_{\alpha/2}\right)^2 \dfrac{\hat{p}\hat{q}}{(emáx)^2}$.

Ejemplos:

a) Continuando con el caso de $\hat{p}$ = 0,34. Si se quiere obtener una estimación de la proporción con un error máximo del 3 %, con un nivel de confianza del 98 %, el tamaño muestral mínimo será:

$$n = (2{,}33)^2 \frac{0{,}34 \cdot 0{,}66}{0{,}03^2} = 1353{,}6 \rightarrow \text{se toma } n = 1354.$$

El 3% ⇒ *E* = 0,03; 1 − α = 0,98 ⇒ α = 0,02, α/2 = 0,01 y $Z_{0,01}$ = 2,33.

b) En trabajos profesionales suele fijarse una confianza del 95,44 (se dice "sigma 2": σ = 2, aludiendo al valor de $Z_{\alpha/2} = Z_{0,0228} = 2$, pues 1 − α = 0,9544, α = 0,0456 y 1 − α/2 = 0,9772).

También es frecuente exigir un error menor a igual al 3 %.

Además, se supone que la proporción a favor y en contra es la misma: *p* = *q* = 0,5.

Con estos supuestos, el tamaño muestral mínimo será: $n = 2^2 \dfrac{0{,}50 \cdot 0{,}50}{0{,}03^2}$ = 1111,1.

En muchas encuestas aparece ese tamaño muestral.

c) ¿A cuántas personas tendríamos que estudiar para conocer la prevalencia de esquizofrenia en un área residencial de Madrid? Se pide un nivel de confianza del 95 % y un error inferior al 5 %. Se sabe, por estudios anteriores, que la proporción esperada es del 10 %.

→ Se supone que *p* = 0,10. Para una confianza del 95 %, $Z_{\alpha/2}$ = 1,96; como el error *E* < 0,05:

$$E = 1{,}96 \cdot \sqrt{\frac{0{,}10 \cdot 0{,}90}{n}} < 0{,}05 \Rightarrow n > \left(\frac{1{,}96 \cdot \sqrt{0{,}10 \cdot 0{,}90}}{0{,}05}\right)^2 = 138{,}3.$$

Debe tomarse una muestra de, al menos, 139 personas.

9. Ejercicios finales

Ejercicio 1. Para determinar el peso medio de las jóvenes de 20 años se ha tomado una muestra aleatoria de tamaño 81. Si la media de la muestra ha sido de 55 kg y se sabe, por estudios anteriores, que la desviación típica de la población es de 4 kg:
a) Halla un intervalo de confianza al 94 % para la media poblacional.
b) Si para la misma confianza se exige un error máximo del 1kg, cuál debe ser tamaño muestral mínimo para determinarlo.
Solución:
a) Para una confianza del 94 %, 1 – α/2 = 0,97, $Z_{\alpha/2}$ = 1,88. El intervalo de confianza de la media poblacional será:

$$IC=\left(\bar{x}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)=\left(55-1{,}88\cdot\frac{4}{\sqrt{81}},\ 55+1{,}88\cdot\frac{4}{\sqrt{81}}\right)=$$

= (55 – 0,84, 55 + 0,84) = (54,16, 55,84).

b) Como el error viene dado por $E=Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Si se desea que E < 1, se tendrá:

$$E=1{,}88\cdot\frac{4}{\sqrt{n}}<1\Rightarrow n>\left(1{,}88\cdot 4\right)^2\approx 56{,}6\,.$$

Debe tomarse una muestra de, al menos, 57 jóvenes.

Ejercicio 2. Con un nivel de confianza igual a 0,95, a partir de un estudio muestral, el intervalo de confianza de la proporción de jóvenes entre 20 y 30 años que realizan compras *online* es (0,36, 0,42).
a) ¿Cuál es la proporción muestral de jóvenes entre 20 y 30 años que realizan compras *online*? ¿Cuál es el tamaño de la muestra?
b) ¿Cuál debería ser el tamaño muestral para estimar la citada proporción, con una confianza del 95 %, con un error máximo de 0,04?
Solución:
a) La proporción de la muestra es el punto medio del intervalo: $\hat{p}=\frac{0{,}36+0{,}42}{2}=0{,}39$. Ese valor será también la estimación puntual de la proporción de la población estudiada.
El margen de error, que es la mitad de la amplitud del intervalo, vale E = 0,03; y, para una confianza del 95 %, $Z_{\alpha/2}$ = 1,96.

Como $E=1{,}96\cdot\sqrt{\frac{0{,}5\cdot 0{,}5}{n}}=0{,}03\ \Rightarrow\ n=\left(1{,}96\right)^2\frac{0{,}5\cdot 0{,}5}{0{,}03^2}=1067{,}1$.

El tamaño muestral mínimo es n = 1068.

b) Si se parte del valor de $\hat{p}=0{,}39\ \rightarrow\ \hat{q}=1-0{,}39=0{,}61$, se tendrá:

$$E=1{,}96\cdot\sqrt{\frac{0{,}39\cdot 0{,}61}{n}}=0{,}04\ \Rightarrow\ n=1{,}96^2\,\frac{0{,}39\cdot 0{,}61}{0{,}04^2}=571{,}2\,.$$

El tamaño muestral debería ser de 572 jóvenes entre 20 y 30 años.
Si se desconociese $\hat{p}$, habrá que suponer que p = q = 0,5, luego

$$E=1{,}96\cdot\sqrt{\frac{0{,}5\cdot 0{,}5}{n}}=0{,}04\ \Rightarrow\ n=1{,}96^2\,\frac{0{,}5\cdot 0{,}5}{0{,}04^2}=600{,}25\,.$$

El tamaño muestral debería ser de 601 jóvenes entre 20 y 30 años.

Problemas Propuestos

Muestreo

1. En una ciudad se quiere hacer una encuesta para conocer el porcentaje de ciudadanos que aprueban la gestión del ayuntamiento en cuestiones medioambientales (limpieza de calles, contaminación, cuidado de parques…). Se pretende que la muestra sea representativa por sexo y edad; para la edad se establecen tres estratos: 10 a 25 años (jóvenes), 25 a 60 años (adultos) y mayores de 60. El número de personas de cada grupo es: jóvenes, 3000; adultos, 8500; mayores de 60, 2500. Por sexo, la distribución es: 6800 hombres y 7200 mujeres, que se suponen proporcionales a cada grupo de edad.
Si el tamaño de la muestra es de 500 personas, determina, redondeando si es necesario, el tamaño muestral correspondiente a cada estrato.

2. Supongamos que en un centro escolar los alumnos y docentes se distribuyen de acuerdo con la tabla:

Alumnos	1 y 2º ESO	3º y 4º ESO	BACH	PROFS
Hombres	110	95	115	20
Mujeres	130	120	130	25

Si se quiere realizar una encuesta entre ellos de tamaño $n = 50$, por el método de muestreo estratificado por sexo y nivel de trabajo, ¿a cuántas personas de cada *clase* hay que preguntar?

3. Supongamos que en tu ciudad (o en tu comarca) hay 4 institutos o colegios (IES), cuyo alumnado se distribuye como sigue:

	IES 1	IES 2	IES 3	IES 4
Alumnos	370	460	620	520
Alumnas	340	500	650	540

Si se quiere realizar un muestreo de tamaño 100, estratificado por sexo y centro escolar, ¿cuáles serían los tamaños muestrales correspondientes a cada IES y a cada estrato?

4. (Selectividad, Galicia 2016).
a) Se desea tomar una muestra estratificada de las personas mayores de edad de un municipio, cuyos estratos son los siguientes intervalos de edades, en años: de 18 a 30, de 31 a 45, de 45 a 60 y mayores de 60. En el primer intervalo hay 7500 personas, en el segundo 8400, en el tercero 5700 y en el cuarto 3000. Calcula el tamaño de la muestra total y su composición, sabiendo que el muestreo se hace con afijación proporcional y se han elegido 375 personas del primer estrato.
b) Dada la población {2, 4, 6}, construye todas las muestras posibles de tamaño 2, que puedan formar mediante muestreo aleatorio simple, y halla la varianza de las medias muestrales de todas las muestras.

Distribución Binomial (Repaso)

5. Un dado, cuyas caras están numeradas del 1 al 6, se lanza cinco veces. Halla la probabilidad de que el número 3 salga:
a) Exactamente dos veces. b) Una vez a lo sumo. c) Más de una vez.

6. En un Centro Escolar el 25 % de los alumnos son de origen extranjero. Si se eligen 6 estudiantes al azar, ¿cuál es la probabilidad de 4 o más sean de origen extranjero?

7. Se lanza una moneda correcta 10 veces y se mide el número de caras y cruces obtenidas.
a) ¿Cuántos resultados forman el espacio muestral? ¿Cuál es la probabilidad de cada uno de los resultados posibles?
b) ¿Cuál es la probabilidad de que salgan 4 caras?

8. En una moneda trucada la probabilidad de obtener cara es 0,4. Si se lanza 5 veces, calcula la probabilidad de obtener al menos 3 caras.

9. Un examen consta de 8 preguntas con 3 posibles respuestas cada una, de las que sólo una de ellas es correcta. Si un estudiante responde al azar marcando las respuestas aleatoriamente, calcula la probabilidad de que:

a) No acierte ninguna respuesta correcta. b) Acierte 6 o más preguntas.

<u>Distribución Normal (Repaso)</u>

10. Utilizando la tabla normal *N*(0, 1) calcula:

a) $P(Z<1,2)$ b) $P(Z<1,27)$ c) $P(Z<-1,2)$ d) $P(Z<-1,27)$

11. Utilizando la tabla normal *N*(0, 1) calcula interpolando:

a) $P(Z<1,325)$ b) $P(Z<1,645)$ c) $P(Z<0,666)$ d) $P(Z<1,863)$

12. Utilizando la tabla normal *N*(0, 1), determina el valor de *k* que cumple:

a) $P(Z<k)=0,9115$ b) $P(Z<k)=0,9452$

c) $P(Z<k)=0,1587$ d) $P(Z<k)=0,95$

13. Para una distribución normal *N*(50, 5), halla:

a) $P(X<56)$ b) $P(X>58)$ c) $P(X<48)$ d) $P(48<X<56)$

14. Supongamos que la estatura media de las alumnas de bachillerato se distribuye normalmente con media μ = 166 cm y desviación típica 9 cm. Si se elige una alumna al azar halla la probabilidad de que su estatura sea:

a) Superior a 175 cm. b) Inferior a 155 cm. c) Esté entre 155 cm y 175 cm.

15. Para una distribución normal *N*(60, 5), determina el valor de *k* que cumple:

a) $P(X<k)=0,90$ b) $P(X>k)=0,95$ c) $P(60-k<X<60+k)=0,9544$

16. Supongamos que los chicos de 15 años de un determinado país tienen una estatura que se distribuye según una normal de media 168 cm y desviación típica 12 cm. Si se quieren seleccionar al 5% de los chicos más altos, ¿a partir de qué altura debe hacerse?

17. El diámetro de las ciruelas de una determina variedad se distribuye normalmente con media 4,5 cm y desviación típica 0,3 cm. Si se desea seleccionar, para su exportación, el 10% de las más grandes, ¿a partir de qué tamaño hay que cogerlas?

18. La edad de los habitantes de cierta ciudad se distribuye normalmente, con una media de 40 años. Se sabe además que el 2,28 % de los habitantes tiene más de 60 años.
a) ¿Cuál es la desviación típica?
b) ¿Cuál es el porcentaje de habitantes con menos de 35 años?

19. La duración de una determinada marca de lavadoras se ajusta a una normal de media 8,4 años y desviación típica 6 meses. El fabricante asegura que sus lavadoras duran más de 7 años, comprometiéndose a: "si una lavadora se estropea antes de 7 años le damos otra nueva". ¿Cuántas lavadoras nuevas tendrá que reponer por cada 10000 vendidas?

20. Los envases de cartón de una determinada marca de leche contienen 1 litro de media, siendo la desviación típica de 5 ml.
a) ¿Qué porcentaje de envases sobrepasan los 1005 ml?
b) Si el control de calidad rechaza los envases que contengan menos de 990 ml y más de 1010 ml, ¿qué porcentaje de envases habrá que rechazar?

21. Los ingresos anuales de los ejecutivos de una multinacional se distribuyen normalmente con media 45000 € y desviación típica de 3000 €. Si se elige un ejecutivo al azar se pide calcular las siguientes probabilidades:
a) De que sus ingresos anuales sean superiores a 50000 euros.
b) De que sus ingresos anuales estén comprendidos entre 42000 € y 46000 €.
c) De que sus ingresos anuales sean inferiores a 39000 euros.
d) Sabiendo que la probabilidad de que sus ingresos anuales sean superiores a una determinada cantidad es del 1 %, ¿cuál es esa cantidad?

Aproximación de la binomial mediante una normal (Repaso)

22. Mediante la aproximación normal de la binomial $B(50, 0{,}12)$ calcula:
a) $P(X=6)$ b) $P(X=12)$ c) $P(6<X\le 12)$

23. El 42 % de los habitantes de un pueblo pasa cada día por la calle mayor. Elegidos 60 habitantes al azar, ¿qué probabilidad hay de que más de 30 de ellos pasen ese día por la calle mayor?

24. Un examen de respuesta múltiple consta de 80 preguntas, cada una con 4 opciones, una de ellas correcta y erróneas las otras tres. Si un estudiante contesta al azar, ¿cuál es la probabilidad de que acierte 25 o más preguntas? ¿Y menos de 10?

Distribución de las medias muestrales:
Intervalo de confianza; Error; Tamaño muestral

25. (Selectividad, Baleares 2014). El cociente intelectual de unos universitarios se distribuye normalmente con una media de 100 y una desviación típica de 10.
a) Se elige una persona al azar. Busca la probabilidad de que su cociente intelectual se encuentre entre 98 y 103.
b) Se elige una muestra de veinticinco personas al azar. Busca la probabilidad de que la media de sus cocientes intelectuales se encuentre entre 98 y 103.

26. (Selectividad, Canarias 2016). En un invernadero que se dedica a la producción de tomates, se ha comprobado que el peso de los tomates sigue una distribución normal con media 100 g y desviación típica 10 g. A la hora de comercializarlos se toman para la clase *A* los comprendidos entre 80 y 120 g. Hallar la probabilidad de que:
a) Elegido un tomate al azar, corresponda a la clase *A*.
b) Elegidos una docena de tomates al azar, su peso medio sea superior a 105 g.

27. Supongamos que la estatura media de las alumnas de 2º de bachillerato es de 165 cm, con desviación típica 8 cm.
a) Halla los parámetros de las medias muestrales de tamaño $n = 36$ y $n = 64$.
b) ¿Cuál es la probabilidad de que una muestra de 36 alumnas tenga una media de 167 o más cm? ¿Y de que una muestra de 64 alumnas supere esa misma medida?
c) ¿Tiene algo de extraño que una muestra de tamaño 36 dé una media de 170 cm?

28. A pesar de acudir con "cita previa", el tiempo de espera de los pacientes de una clínica dental sigue una distribución normal de media desconocida y desviación típica 10 minutos. A partir de una muestra aleatoria de 144 pacientes, se obtuvo una media de espera es de 20 minutos.
a) Calcula los intervalos de confianza del 90 %, 95 % y 99 % para la media del tiempo de espera de la población.
b) Explica la relación entre la amplitud de los intervalos y el nivel de confianza.

29. El número de *tweet* diarios generados por jóvenes de entre 16 y 24 años sigue una distribución normal de media desconocida y desviación típica 6 tweet. Tomada una muestra de 81 usuarios de Twitter, su media diaria ha resultado ser de 12 tweet. Calcula los intervalos de confianza del 80 % y 85 % para la media de la población.

30. A lo largo de las diferentes pruebas de Selectividad se ha observado que la distribución de las calificaciones en el examen de Matemáticas siguen una ley normal de media 5,7 puntos y desviación típica 1,8.
a) ¿Cuál es la probabilidad de que la nota de un estudiante elegido al azar sea superior a 6,3?
b) ¿Cuál es la probabilidad de que una muestra de 49 alumnos tenga una media superior a 6,5 puntos?

31. En el último año de Selectividad (2016), en una universidad se tomó una muestra aleatoria de tamaño 49, obteniéndose una nota media en el examen de Matemáticas de 5,5 puntos. Admitiendo que la desviación típica sigue siendo de 1,8 puntos, se pide:
a) Un intervalo de confianza al 90 % para la nota media de la población.
b) Si para esa misma confianza se quiere un error máximo de 0,5 puntos, ¿cuál debe ser el tamaño de la muestra?

32. Se sabe que el número de horas que duermen los habitantes de una ciudad se puede aproximar mediante una distribución normal con desviación típica de 0,64 horas. Si se toma una muestra aleatoria de tamaño 16 y se pregunta por el número de horas que dedican a dormir se obtienen los siguientes datos:

6,5	8,5	6,5	8,5	7,5	7,0	5,5	7,5
7,5	6,5	7,0	7,0	6,5	8,0	7,5	6,0

a) Calcula la media muestral del número de horas que se duermen.
b) Halla el intervalo de confianza para la media de la población con una confianza del 95%
c) Si el que el intervalo de confianza para la media es (6,7, 7,48) horas, ¿con qué nivel de confianza se da?

33. (Selectividad Castilla la Mancha, 2016). Se estudió el cociente intelectual de 10 estudiantes de 2ª de Bachillerato elegidos aleatoriamente de un determinado centro escolar, siendo estos valores:

80, 96, 87, 104, 105, 99, 112, 89, 90, 100.

Sabiendo que el cociente intelectual se distribuye según una normal con desviación típica 15, se pide:
a) Halla el intervalo de confianza al nivel del 95 % para la media del cociente intelectual de los estudiantes de 2º de Bachillerato de dicho centro escolar.
b) Razona y explica qué se podría hacer para que el intervalo de confianza tuviera menor amplitud con el mismo nivel de confianza.

34. El peso de los paquetes de espagueti de una determinada marca, sigue una distribución normal con desviación típica 20 gramos. Se seleccionan al azar 50 paquetes de esos espaguetis y se observa que tienen un peso medio de 745 gramos.
Halla el intervalo de confianza para el peso medio de los paquetes de espaguetis de esa marca con un nivel de confianza del 97 %.

35. Se realiza una encuesta a 100 trabajadores menores de 25 años, sobre su sueldo mensual, obteniéndose una media de 900 € con una desviación típica de 120 €.
a) ¿Cuál es el intervalo de confianza para la media de ingresos de los trabajadores de ese sector, con un nivel de confianza del 92 %?
b) ¿Cuál debe ser el tamaño mínimo de la muestra necesario para estimar la media de ingresos mensuales con un error menor de 20 € y con una confianza del 94 %?

36. El gasto familiar medio en material escolar (por hijo) sigue una distribución normal con desviación típica $\sigma = 50$ €. A partir de una muestra de 100 personas se ha obtenido un gasto medio de 325 euros.
a) Halla un intervalo de confianza al 95 % para el gasto medio por hijo.
b) ¿Qué tamaño deberá tener la muestra para obtener un intervalo de confianza al 99 % con una amplitud igual a la anterior?

37. (Selectividad, La Rioja 2012). La valoración de las instituciones por parte de los ciudadanos se mide en unas unidades ficticias que denominaremos “u”. Se sabe que, en el caso de los españoles, dicha valoración sigue una normal con desviación típica 25 u.
a) Se elige una muestra de 100 españoles, dando una media de 180 u. Calcula un intervalo de confianza para la media poblacional de la valoración de las instituciones, con una probabilidad del 90 %.
b) Si se conoce que la media poblacional es 182 u, calcula la probabilidad de que una muestra de tamaño 100 tenga media inferior a 180 u.

38. En la lotería navideña, una muestra aleatoria de 10 personas mayores de edad, jugaron las siguientes cantidades (en euros):

74, 72, 65, 75, 80, 81, 82, 84, 87, 90

Sabiendo que el gasto por persona, en la lotería navideña, sigue una distribución normal con desviación típica $\sigma = 20$ €, halla un intervalo de confianza para el gasto medio de la población con un nivel de confianza del 95,44 %.

39. La edad de los atletas participantes en la última olimpiada seguía una distribución normal con desviación típica de 5 años. Una muestra aleatoria de 150 atletas dio como resultado una media de edad de 25,4 años.
a) Halla el intervalo de confianza del 94 % para la media de edad de todos los atletas olímpicos.
b) ¿Cuál fue el tamaño mínimo de una la muestra si se exigió estimar la media con un nivel de confianza del 92 % y con un error máximo de 0,5 años?

40. (Selectividad, Andalucía 2013).
El tiempo que los españoles dedican a ver la televisión los domingos es una variable aleatoria que sigue una distribución Normal de media desconocida y desviación típica 75 minutos. Elegida una muestra aleatoria de españoles se ha obtenido, para la media de esa distribución, el intervalo de confianza (188,18, 208,82), con un nivel del 99 %.
a) Calcula la media muestral y el tamaño de la muestra.
b) Calcula el error máximo permitido si se hubiese utilizado una muestra de tamaño 500 y un nivel de confianza del 96 %.

41. (Selectividad, Cantabria 2014).
a) El tiempo diario que los estudiantes de bachillerato de Cantabria dedican al estudio en las dos semanas previas al inicio de los exámenes de Selectividad de la convocatoria de junio, sigue una distribución normal de media desconocida y desviación típica 15 minutos. Para estimar el tiempo medio se elije una muestra de 300 alumnos. ¿Con qué nivel de confianza debe realizarse la estimación si el error cometido es de 1,88 minutos?
b) Con vistas a la convocatoria de septiembre del mismo año se realiza un análisis similar. El tiempo diario que los estudiantes destinan al estudio las dos semanas anteriores al inicio de los exámenes, sigue una distribución normal con desviación típica 11 minutos. Con una muestra aleatoria de 150 alumnos se ha obtenido un tiempo medio de 173 minutos. Obtener el intervalo de confianza del 93 % para el tiempo medio de estudio.

Distribución de la proporción:
Intervalo de confianza; Error; Tamaño muestral

42. En unas elecciones a alcalde, el 56 % de los votantes optó por el candidato *A* mientras que el 44 % lo hizo por el candidato *B*.
a) Halla la distribución de probabilidad de las muestras de tamaño 50 extraídas de da población. Haz lo mismo para $n = 100$.
b) Calcula la probabilidad de que en una muestra de 50 votantes haya, al menos, 30 favorables al candidato *A*.
c) Si la muestra es de tamaño 100, ¿cuánto es la probabilidad de que una mayoría apoye al candidato *B*?

43. De una muestra aleatoria de 120 graduados universitarios, 15 de ellos no han conseguido trabajo después de un año de su graduación.
a) Halla el intervalo de confianza, al 99 %, para estimar la proporción de graduados que sí han encontrado trabajo trascurrido un año de su graduación.
b) Para la misma confianza, ¿cuál debe ser el tamaño mínimo de la muestra para estimar la proporción de graduados con trabajo, admitiendo un error máximo del 5 %?

44. Se quiere estimar la proporción de personas que esperan que su situación económica mejore el año próximo. Para ello se ha preguntado a 500 personas de esa población, de las 175 esperan que su situación económica mejore: son optimistas.
a) Calcula un intervalo de confianza para la proporción de personas optimistas en esta población, con un nivel de confianza del 94 %.
b) Si antes de conocer el resultado se quiere determinar un intervalo de confianza con el mismo nivel (94 %) y un error máximo de 0,02, ¿cuál debe ser el tamaño mínimo de la muestra?

45. Para estimar el grado de satisfacción de sus clientes, una compañía de reparto realiza una encuesta aleatoria entre 122 de sus clientes. De ellos, 103 declararon estar satisfechos.
a) ¿Cuál es la estimación de la proporción de clientes satisfechos?
b) Halla el intervalo de confianza al 99 % para la estimación de la proporción de clientes satisfechos.

46. (Selectividad, Canarias 2014). En una zona escolar, para una muestra de 200 alumnos, 30 son repetidores.
a) Construir un intervalo de confianza con un nivel del 95 %, para estimar la proporción de alumnos repetidores.
b) Si se ignoran los datos iniciales y con un nivel de confianza del 90 %, ¿cuál es el tamaño mínimo muestral para estimar la proporción de alumnos repetidores con un error máximo del 2 %?

47. En una encuesta realizada a 70 jóvenes, 14 se declararon contrarios a la Unión Europea. Halla un intervalo de confianza del 92 % para determinar el porcentaje de la proporción de jóvenes que es contrario a la UE.

48. En una encuesta telefónica realizada a 125 personas, 35 de ellas declaran que el problema del paro juvenil es el más preocupante.
a) Halla un intervalo de confianza, con una significación α = 0,06, para obtener el porcentaje de la población que sitúan el problema del paro juvenil como el más preocupante.
b) Si la encuesta se hubiese realizado a 900 personas y se hubiese obtenido un valor de $\hat{p} = 0,28$, para la misma confianza, ¿qué margen de error se estaría asumiendo?

49. (Selectividad, Murcia 2016). Para estimar la proporción de individuos de una población que utilizan el comercio electrónico se ha realizado una encuesta a una muestra aleatoria de 200 individuos, de los cuales 90 han respondido que utilizan el comercio electrónico. Con estos datos, hallar un intervalo de confianza del 95 % para la proporción de individuos de la población que utilizan el comercio electrónico.

<u>Otros problemas</u>

50. Una variable aleatoria se distribuye normalmente con desviación típica conocida σ. Halla el valor de σ sabiendo que a partir de una muestra aleatoria simple de tamaño 100, se ha obtenido una media de 25 y se dice que (24,02, 25,98) es el intervalo de confianza para la media poblacional, con un nivel de significación 0,05.

51. (Selectividad, Madrid 2012). Se supone que el gasto que hacen los individuos de una determinada población en regalos de Navidad se puede aproximar por una variable aleatoria con distribución normal de media μ y desviación típica igual a 45 euros.
a) Se toma una muestra aleatoria simple y se obtiene el intervalo de confianza (251,6, 271,2) para μ, con un nivel de confianza del 95 %. Calcula la media muestral y el tamaño de la muestra elegida.
b) Se toma una muestra aleatoria simple de tamaño 64 para estimar μ. Calcula el error máximo cometido por esa estimación con un nivel de confianza del 90 %.

52. (Selectividad, Castilla La Mancha 2014). En un aeropuerto, el tiempo de espera de un viajero frente a la cinta transportadora hasta que sale su maleta sigue una distribución normal

de media desconocida y desviación típica $\sigma = 3$ minutos. Se tomó una muestra aleatoria de 50 viajeros, y se observó que el tiempo medio de espera era de 17 minutos.
a) Halla un intervalo de confianza para la media poblacional del tiempo de espera de la maleta en ese aeropuerto con un nivel de confianza del 95 %.
b) ¿Se puede admitir que la media poblacional sea $\mu = 16$ con un nivel de confianza del 95%? ¿Cómo podríamos disminuir la amplitud del intervalo de confianza sin variar el nivel de confianza? Razona tus respuestas.

53. El nivel medio de colesterol (en mg/dl) en individuos sanos depende de la edad y el sexo; para los hombres con menos de 21 años su distribución es normal con media $\mu = 160$ y desviación típica $\sigma = 10$. Un nivel fuera de $\mu \pm 2\sigma$ resulta extraño: indica que puede haber alguna anomalía. Lo mismo cabe decir de las muestras: un nivel muestral fuera de $\mu \pm 2\sigma_{\bar{X}}$ resulta extraño.
a) ¿Cuál es el intervalo de probabilidad admisible (no extraño) para las muestras de tamaño: 1; 9 y 100?
b) ¿Qué porcentaje de individuos o muestras caen en los intervalos hallados?

54. (De un examen MIR). Se desea conocer la media de la colesterolemia basal de una población, con una seguridad del 95 % y una precisión de ± 4 mg/dl (error), y se tiene información por un estudio piloto o revisión bibliográfica de que la varianza es de 300 mg/dl. ¿Cuál debe ser el tamaño muestral?

55. El perímetro torácico de los individuos adultos (hombres) en una población se distribuye normalmente con desviación típica $\sigma = 6$ cm. Si a partir de una muestra de tamaño *n* se afirma que el intervalo de confianza al 88 % es (87, 91), ¿cuál fue el tamaño de la muestra estudiada?

56. La ficha técnica de una encuesta indica:
Universo: Mayores de 18 años. *Ámbito*: Nacional. *Muestra*: 1000 entrevistas con un margen de error ±3,16 para datos globales, con un nivel de confianza del 95,5 % (dos sigma) y un p/q = 50/50. *Selección*: Estratificada, aleatoria. *Entrevista*; Telefónica. *Fecha de trabajo de campo*: del 16 al 18 de febrero de 2016. *Realización*: SIGMA DOS. *Dirección*: José Miguel de Elías.
Comprueba si el margen de error es el indicado

57. Si a la empresa encuestadora le exigen un margen de error de ±3, ¿cuál debe ser el tamaño muestra? (Los demás parámetros permanecen iguales).

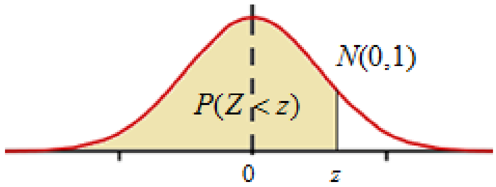
58. (Selectividad, Canarias 2016). Un estudio sobre los kilogramos de residuos no minerales que genera cada español al año, ha dado, para una muestra de 100 personas, el intervalo de confianza [1470,6; 1529,4]. Si la desviación típica es de 150 kilogramos, suponiendo que la generación de residuos sigue una distribución normal:
a) ¿Cuál es la media muestral?
b) ¿Cuál es el nivel de confianza utilizado?
c) ¿Cuál sería el correspondiente intervalo con la misma información muestral, pero con un nivel de confianza igual a 0,9?

Soluciones

1. Hombres: 52; 148; 43. Mujeres: 55; 156; 46.
2. Hombres: 7; 6; 8. Mujeres: 9; 8; 9.
3. Alumnos: 9; 11; 16; 13. Alumnas: 9; 13; 16; 13,
4. a) 1230: 375 + 420 + 285 + 150.
b) {2, 2}; {2, 4}; {2, 6}; {4, 2}; {4, 4}; {4, 6}; {6, 2}; {6, 4} y {6, 6}; $\sigma_{\bar{x}}^2 = \frac{4}{3}$.
5. a) 0,16075. b) 0,80375. c) 0,19625.
6. 0,03759.
7. a) 1024; 1/1024. b) 105/5122.
8. 0,31744.
9. a) 256/6561. b) 129/6561.
10. a) 0,8849. b) 0,8980. c) 0,1151.
d) 0,1020.
11. a) 0,9074. b) 0,95. c) 0,74732.
d) 0,96881. e) 0,0735. f) 0,9222.
12. a) 1,35. b) 1,6. c) –1. d) 1,645.
13. a) 0,8849. b) 0,0548. c) 0,3446.
d) 0,5403.
14. a) 0,1587. b) 0,1112. c) 0,7301.
15. a) 66,4. b) 51,775. c) 10.
16. 187,74 cm.
17. 4,884 cm.
18. a) 10. b) 30,85 %.
19. 26.
20. a) 15,87% b) 4,56%
21. a) 0,0475. b) 0,4706. c) 0,0228.
d) 51990 €.
22. a) 0,1742. b) 0,0061. c) 0,5787.
23. 0,0823.
24. 0,1230; 0,0034.
25. a) 0,1972. b) 0,7745.
26. a) 0,9544. b) 0,0418.
27. a) $N\left(165,\ \frac{4}{3}\right)$; $N(165,\ 1)$. b) 0,0668; 0,0228. c) Muy improbable, menos de 0,0002.
28. a) (18,63, 21,37); (18,37, 21,63); (17,85, 22,15).
29. (11,15, 12,85); (11,04, 12,96).
30. a) 0,3707. b) 0,0009.
31. a) (5,077, 5,923). b) 36 personas.
32. a) 7,09 h. b) (6,78, 7,4). c) 98,54%.
33. a) (86,9, 105,5). b) Aumentar *n*.
34. (738,86, 751,14).
35. a) (879, 921). b) n = 128.
36. a) (315,2, 334,8). b) 173.
37. a) (175,9, 184,1). b) 0,2119.
38. (66,35, 91,65)
39. a) (24,63, 26,17). b) 307 atletas.
40. a) 198,5; 350. b) ±6,88.
41. a) 97%. b) (171,37, 174,63).
42. a) N(0,56, 0,07); N(0,56, 0,05).
b) 0,2843. c) 0,1151.
43. a) (0,797, 0,953). b) 291.
44. a) (0,31, 0,39). b) 2010.
45. a) 0,844. b) (0,759, 0,929).
46. a) (0,125, 0,175). b) 260.
47. (0,152, 0,248).
48. a) (0,204, 0,356). b) 2,8%.
49. (0,381, 0,519).
50. $\sigma = 5$.
51. a) $\bar{x}$ = 261,6; n = 100. b) 10.28.
52. a) (16,17, 17,83). b) Es muy improbable. Aumentando *n*.
53. a) (140, 180); (153,4, 166,6); (158, 162). b) 95,44%.
54. 73 personas.
55. 22 hombres.
57. 1111.
58. a) $\bar{x}$ = 1500 kg. b) 95%.
c) (1475,3, 1524,7).

ÁREAS BAJO LA DISTRIBUCIÓN DE PROBABILIDAD NORMAL ESTÁNDAR

Los valores en la tabla representan el área bajo la curva normal hasta un valor positivo z.



Z	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,5000	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