

ESTADÍSTICA INFERENCIAL. MUESTREO

Índice

1. Introducción	2
2. Muestreo	3
3. Distribución Normal	5
3.1. TIPIFICACIÓN DE LA VARIABLE	6
3.2. LA DISTRIBUCIÓN BINOMIAL SE APROXIMA A LA NORMAL	7
4. Estimación de Parámetros	8
5. Distribuciones muestrales	8
5.1. DISTRIBUCIÓN MUESTRAL DE MEDIAS	8
5.2. DISTRIBUCIÓN MUESTRAL DE PROPORCIONES	10
6. Estimación Puntual	11
7. Intervalos de Probabilidad	12
7.1. INTERVALO DE PROBABILIDAD PARA LA MEDIA MUESTRAL \bar{X}	13
7.2. INTERVALO DE PROBABILIDAD PARA LA PROPORCIÓN MUESTRAL \hat{p}	15
8. Estimación por intervalos	15
8.1. ESTIMACIÓN DE LA MEDIA DE UNA POBLACIÓN μ	16
8.2. ESTIMACIÓN DE UNA PROPORCIÓN p	17
8.3. ERROR MÁXIMO ADMISIBLE	18
8.4. TAMAÑO DE LAS MUESTRAS	18

9. Actividades Finales	19
10. Anexo I: Tabla de la Distribución Normal Estandar	21
11. Anexo II	22
12. Anexo III: La Distribución Binomial	23

1. Introducción

La palabra **Inferir** significa: Sacar una consecuencia de una cosa. Sacar consecuencia o deducir una cosa de otra. La estadística es la ciencia o rama de las Matemáticas que se ocupa de recoger datos, analizarlos y organizarlos, y también de realizar las predicciones que sobre esos datos puedan deducirse, tiene dos vertientes básicas:

- **Estadística descriptiva:** Básicamente se ocupa de la 1ª parte, es decir, a partir de ciertos datos, analizarlos y organizarlos. Es aquí donde tiene sentido calcular la media, mediana, moda, varianza, desviación típica, etc. Es la parte que estudiamos el año pasado.
- **Estadística inferencial:** Se ocupa de predecir y sacar conclusiones para una población, tomando como base una muestra (es decir, una parte) de dicha población. Teniendo en cuenta que cualquier predicción siempre ha de hacerse bajo un cierto grado de fiabilidad o confianza. Esta vertiente de la estadística es la que estudiaremos en esta unidad didáctica y en la siguiente.

En ocasiones el tamaño de la población es muy grande y frecuentemente no es posible estudiar todos sus elementos (por razones de tiempo, economía, etc). Por eso lo que nos interesa es estudiar una muestra y deducir o inferir las características de la población a partir de las características de la muestra: **La estadística Inferencial** se ocupa de deducir o inferir las características de la población a partir de las de la muestra. Podemos dividir la estadística inferencial en:

- **Estadística Inductiva**, cuyo objetivo es estimar los parámetros de una población
 - mediante un único valor: **estimación puntual**
 - mediante un intervalo: **estimación por intervalos**
- **Estadística Deductiva**, cuyo objetivo es comprobar si la información que nos proporciona la muestra permite afirmar o no, una suposición previa (hipótesis) formulada sobre la población, mediante el **contraste de hipótesis** que trataremos en la siguiente unidad.

Recordamos algunos términos estadísticos:

- **Población:** es el conjunto de elementos objeto del estudio estadístico.
- **Individuo:** es cada elemento de la población.
- **Muestra:** es el subconjunto o parte de la muestra que tomamos para hacer el estudio.

- **Tamaño de la muestra:** es el número de individuos que la componen.

Es muy importante distinguir entre los **parámetros poblacionales** o simplemente **parámetros**, que son los índices centrales, de dispersión, etc, de **TODA** la población y que en la práctica no son calculables.

Y los **parámetros muestrales** o **estadísticos**, que son los índices centrales, de dispersión, etc de la muestra, que son los que se calculan para estimar los parámetros.

Los estadísticos que más vamos a utilizar son:

- La media muestral: \bar{x}
- La desviación típica muestral: s

2. Muestreo

Ya sabemos que una población es el conjunto de individuos sobre los que hacemos cierto estudio, y que una muestra es un subconjunto de la población. Es evidente que los resultados de una determinada encuesta tendrán un mayor grado de fiabilidad si dicha encuesta se realiza sobre la población completa. Sin embargo, en la mayoría de las ocasiones esto no es posible, debido a múltiples razones, como por ejemplo:

- **Imposibilidad material** Hacer una encuesta a los casi 41 millones de españoles es imposible, o hacer un estudio sobre la fecha de caducidad de un producto. Si lo hacemos con todos los productos ¿qué vendemos luego?
- **Imposibilidad temporal** Hacer un estudio sobre la duración de una bombilla. ¿Cuánto debemos esperar para saberlo?.

Por tanto, es habitual que tengamos que manejarnos con muestras, de modo que es importante saber elegir bien una muestra de la población, una muestra que **represente** bien a dicha población y que nos permita con un alto grado de **fiabilidad** inferir o predecir las características de la población.

Hay muchas maneras de elegir una muestra de una población, Pero antes de pasar a analizar dichas formas de extracción de muestras, lo que si hemos de dejar claro es que todas las muestras han de cumplir varias condiciones indispensables.

Es evidente que para que el estudio a realizar sea fiable, hay que cuidar mucho la elección de la muestra, para que represente en la medida de lo posible a la población de la que se extrae. Si la muestra está mal elegida, diremos que no es representativa. En este caso, se pueden producir errores imprevistos e incontrolados. Dichos errores se denominan **sesgos** y diremos que **la muestra está sesgada**.

Una de las condiciones para que una muestra sea representativa es que el sistema que se utilice para elegirla sea aleatorio, es decir, que todos los individuos de la población tengan las mismas posibilidades de ser elegidos, mientras que si la elección de la muestra es subjetiva, es probable que resulte sesgada.

Las distintas maneras de elegir una muestra de una población se denominan **muestreos** y básicamente hay dos tipos de muestreos:

1. *Muestreo no probabilístico*: El investigador no elige la muestra al azar, sino mediante determinados criterios subjetivos. Los individuos de la población no tienen la misma probabilidad de ser incluidos en la muestra.

En este tipo de muestreo suele ser muy escasa la representatividad y por tanto, poco válidas las inferencias que pueden hacerse.

2. *Muestreo probabilístico o aleatorio*: Es el que se realiza teniendo en cuenta que cada miembro de la población tiene la misma probabilidad de ser elegido en la muestra. Con este tipo de muestreo, las muestras suelen ser más representativas, es posible conocer los errores cometidos y pueden hacerse inferencias estadísticas. En este caso podemos distinguir varios tipos:

- *Muestreo aleatorio simple*: Aquel en el que cada individuo de la población tiene las mismas posibilidades de salir en la muestra.

- *Muestreo sistemático*: En el que se elige un individuo al azar y a partir de él, a intervalos constantes, se eligen los demás hasta completar la muestra.

- *Muestreo estratificado*: Se divide la población en clases o estratos y se escoge, aleatoriamente, un número de individuos de cada estrato proporcional al número de componentes de cada estrato.

- *Muestreo por conglomerados*: Si no disponemos de la relación de los elementos de la población, o de los posibles estratos, no podemos aplicar los muestreos anteriores.

Aquí entra el llamado muestreo por conglomerados, donde en lugar de elegir individuos directamente, se eligen unidades más amplias donde se clasifican los elementos de la población, llamados conglomerados. En cada etapa del muestreo en lugar de seleccionar elementos al azar seleccionamos conglomerados.

Los conglomerados deben ser tan heterogéneos como la población a estudiar, para que la represente bien. Luego se elegirían algunos de los conglomerados al azar, y dentro de éstos, analizar todos sus elementos o tomar una muestra aleatoria simple. No debemos confundir estrato y conglomerado. Un estrato es homogéneo (sus elementos tienen las mismas características), mientras que un conglomerado es heterogéneo (debe representar bien a la población).

En cualquier caso hemos de asumir que un **error en el muestreo** ocasionará que los resultados que proporcione la muestra no coincidan o estén alejados de los valores reales de la población. Pueden darse dos tipos de errores:

- **Error aleatorio muestral**. Para reducir este error hay que aumentar el tamaño de la muestra.

- **Error sistemático o Sesgo**. Va asociado al proceso de selección de la muestra y se reduce mejorando esta selección.

Veamos la diferencia de estos muestreos mediante el siguiente ejemplo: *Imaginemos que hemos de recoger una muestra de 20 alumnos de entre los de un 600 instituto.*

- *Muestreo aleatorio simple*: Elegiríamos un alumno al azar (probabilidad de elegirlo $\frac{1}{600}$). Lo devolvemos a la población y se elige otro (probabilidad de elegirlo $\frac{1}{600}$), y así hasta 20. Observa que si no devolviésemos al alumno, entonces, la probabilidad de escoger al 2º alumno sería $\frac{1}{599}$, y ya no todos tendrían la misma probabilidad de ser elegidos.¹ El problema es que entonces permitimos que se puedan repetir individuos.

¹Recuerda las extracciones con y sin reemplazamiento

- *Muestreo sistemático*: Como hemos de elegir 20 alumnos de 600, es decir, 1 de cada 30, se procede así: Se ordenan los alumnos y se numeran, se elige uno al azar, por ejemplo el alumno 27, y luego los demás se eligen a partir de este a intervalos de 30 alumnos. Escogeríamos por tanto a los alumnos:

27,57,87,117,147,177,207,237,267,297,327,357,387,417,447,477,507,537,567,597

- *Muestreo estratificado*: Como queremos que la muestra sea representativa, lo mejor será conocer cuántos alumnos de cada curso hay, es decir, si hay 200 alumnos de 3º ESO, 150 de 4º ESO, 150 de 1º Bachillerato y 100 de 2º Bachillerato, procederíamos así: Como de 600 en total hemos de elegir a 20, de 200 de 3º de ESO hemos de elegir x

$$\frac{20}{600} = \frac{x}{200} \longrightarrow x = \frac{4000}{600} = 6,6 \approx 7 \text{ alumnos de } 3^\circ \text{ ESO}$$

De igual manera, utilizando una regla de tres, podemos calcular los alumnos correspondientes a los demás cursos: 7 alumnos son de 3º, 5 alumnos de 4º, 5 alumnos de 1º y 3 alumnos de 2º. Por último, para la elección de cada alumno dentro de cada curso utilizamos el muestreo aleatorio simple.

- *Muestreo por conglomerados*: Para ver este muestreo, hemos de cambiar el ejemplo. Supongamos que queremos extraer una muestra aleatoria de los estudiantes universitarios del país. Necesitaríamos una lista con todos ellos para poder realizar algún muestreo del tipo de los 3 anteriores, lo cuál es muy difícil de conseguir. Sin embargo, los estudiantes están clasificados por Universidades, Facultades y Clases. Podemos seleccionar en una primera etapa algunas Universidades, después algunas facultades al azar, dentro de las facultades algunas clases y dentro de las clases, algunos estudiantes por muestreo aleatorio simple. Los conglomerados en cada etapa serían las diferentes Universidades, las diferentes facultades y las diferentes clases. Como vemos los conglomerados son unidades amplias y heterogéneas.

Ejercicio 1: En una población de 1500 jóvenes, 7500 adultos y 1000 ancianos, se hace una encuesta a 200 personas para conocer sus actividades de ocio preferidas. Si se utiliza un muestreo estratificado, ¿qué tamaño muestral corresponde a cada estrato?

3. Distribución Normal

En este punto vamos a recordar la **distribución de probabilidad normal** que es fundamental para comprender los puntos siguientes.

La campana de Gauss, curva de Gauss o **curva normal** es una función de probabilidad **continua** y **simétrica**, cuyo máximo coincide con la media μ . Esta curva fue descrita por el matemático alemán Carl F. Gauss, que llegó a ella estudiando los errores que se cometen al medir reiteradamente una cierta magnitud².

²Al efectuar una misma observación astronómica o geodésica repetidas veces, obtenía valores diferentes debido a errores humanos y a errores de los aparatos de medida. Para resolver este problema formuló la teoría de mínimos cuadrados, de gran utilidad práctica.

La gran importancia de esta distribución se debe a la enorme frecuencia con que aparece en las situaciones más variadas, entre las muchas variables que se distribuyen *normalmente* podemos citar:

- Caracteres morfológicos como peso, talla, etc
- Caracteres fisiológicos, como por ejemplo el efecto de una misma dosis de un fármaco.
- Caracteres sociológicos, como por ejemplo el consumo de ciertos productos por individuos de un mismo grupo humano.
- Caracteres físicos, como por ejemplo la resistencia a la rotura de piezas aparentemente idénticas.
- Y en general cualquier característica que se obtenga como suma de muchos factores

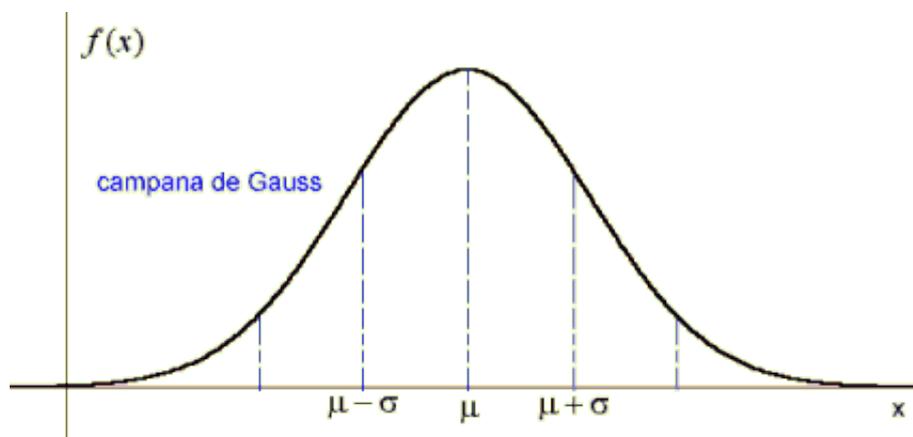


Figura 1: Distribución Normal μ es la media, σ es la desviación típica

Si X es una variable aleatoria que sigue la Distribución Normal, entonces es suficiente con conocer su media μ y su desviación típica σ para que quede completamente determinada. Se dice entonces que $X = N(\mu, \sigma)$.

La distribución $Z = N(0, 1)$ se conoce como **normal estandar** ($\mu = 0$ y $\sigma = 1$). Se encuentra tabulada y resulta sencillo calcular probabilidades que se corresponden con las áreas encerradas bajo la curva. La tabla la tienes en el epígrafe ANEXO I.

3.1. Tipificación de la variable

Cuando una variable normal X no sigue la distribución estandar $N(0, 1)$ sino una $N(\mu, \sigma)$ entonces hay que tipificar la variable, es decir, transformarla en una variable estandar. Para ello se hace el cambio de variable:

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

Con lo que el cálculo de probabilidades se reduce a:

$$P(X \leq a) = P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) = P\left(Z \leq \frac{a - \mu}{\sigma}\right)$$

y éstas se encuentran en la tabla.

Por ejemplo, en una distribución $X = N(14, 4)$ hallar $P(X \leq 20)$

$$P(X \leq 20) = P\left(\frac{X - 14}{4} \leq \frac{20 - 14}{4}\right) = P(Z \leq 1,5) = 0,9332$$

Como ya habrás observado (y recordarás del curso pasado), en la tabla sólo encontramos los valores correspondientes a $P(Z \leq a)$ con $a \geq 0$. En los demás casos hemos de usar propiedades de la curva normal, como la simetría, para llegar al resultado. En el epígrafe ANEXO II tienes ejemplos de todos los casos posibles.

3.2. La distribución Binomial se aproxima a la normal

En el curso anterior estudiamos las distribuciones binomiales $B(n, p)$ ANEXO III. Son distribuciones de probabilidad discretas³, que responden a la función de probabilidad:

$$P(X = k) = \binom{n}{k} p^k q^{n-k}$$

siendo la media $\mu = n \cdot p$ y la desviación típica $\sigma = \sqrt{n \cdot p \cdot q}$.

El cálculo de expresiones de la forma $\binom{n}{k} p^k q^{n-k}$, resulta muy laborioso y complicado sobre todo cuando n es grande.

El matemático Abraham de Moivre (1667-1754) demostró el siguiente resultado que, bajo ciertas condiciones, permite aproximar la distribución binomial por una distribución normal:

Teorema 1 *Si X es una variable discreta que sigue la distribución binomial $X = B(n, p)$, entonces la variable X se aproxima a una variable normal $Y = N(n \cdot p, \sqrt{n \cdot p \cdot q})$*

$$B(n, p) \approx N(n \cdot p, \sqrt{n \cdot p \cdot q})$$

La bondad de la aproximación⁴ es mayor cuanto mayor es n y cuanto más próximo está p a 0,5.

Ejercicio 2: Se efectúan 15 lanzamientos de una moneda. Calcula la probabilidad de que:
a) Salgan exactamente 9 caras.
b) Salgan entre 8 y 12 caras, ambas inclusive.

Nota a tener en cuenta: **Corrección de Yates**

Cuando aproximamos una distribución Binomial mediante una Normal, estamos convirtiendo una variable discreta en variable continua. Para variables continuas la probabilidad de que la variable tome un valor fijo es nula, $P(X = a) = 0$. Para evitar este problema, en la aproximación de los valores fijos, estos se corrigen sustituyéndolos por un intervalo centrado en el valor y de amplitud la unidad.

Así para $X = a$ se considera $a - 0,5 \leq X \leq a + 0,5$.

Utiliza la corrección de Yates para resolver el ejercicio propuesto.

³Son aquellas en las que la variable sólo toma un número finito o numerable de valores

⁴Esta aproximación está especialmente indicada cuando n es mayor que 10, ya que las tablas binomiales ofrecen valores hasta $n=10$

4. Estimación de Parámetros

Como el objetivo principal de la estadística inferencial es el estudio de la población y realizar predicciones acerca de ella (pero a partir de una muestra de ella, no de la población entera), en principio tendremos que estimar los índices de la población a partir de los índices correspondientes para la muestra.

En una primera aproximación, parece lógico pensar que si por ejemplo, queremos determinar la media de una cierta población, si hemos cogido una muestra representativa la media de la muestra (que es fácilmente calculable porque tenemos los datos) será muy parecida a la de la población y por tanto nos sirva para estimarla.

Por tanto debemos distinguir entre:

1. *Parámetros poblacionales*: Que son los índices centrales y de dispersión que definen a una población.

Representaremos la media poblacional μ y la desviación típica poblacional σ .

En el caso de proporciones, la proporción de población que tiene una determinada característica la denotaremos por p y la proporción que no la cumple por $q = 1 - p$ (como en la Distribución Binomial)

2. *Estadísticos poblacionales*: Son los índices centrales y de dispersión que definen a una muestra.

Representaremos la media muestral por \bar{x} y la desviación típica muestral por s .

En el caso de proporciones, la proporción de muestra que tiene una determinada característica la denotaremos por \hat{p} y la proporción que no la cumple por $\hat{q} = 1 - \hat{p}$.

¿Cuál es el problema de la estimación entonces?. Como vamos a disponer de una muestra, lo que podemos calcular es \bar{x} y s (o bien \hat{p} y \hat{q}), y a partir de estos intentar estimar quienes tienen que ser μ y σ (o bien p y q), los reales para la población.

Para ello hemos de conocer cuál es la relación entre un estadístico y el correspondiente parámetro. Es necesario conocer la distribución muestral de estos estadísticos.

5. Distribuciones muestrales

5.1. Distribución muestral de medias

Comenzamos por la situación de obtener conclusiones sobre la media de la población a partir del estudio de medias obtenidas de las muestras.

Si tenemos una población de parámetros desconocidos μ y σ , y tomamos una muestra, podemos calcular la media muestral, \bar{x}_1 , que tendrá cierta relación con μ .

Podríamos tomar otra muestra, de igual tamaño n , y calcular de nuevo su media muestral \bar{x}_2 , que también estará relacionada con μ .

Así sucesivamente, podemos considerar una variable aleatoria \bar{X}_n , que asigna a cada muestra su media y podemos estudiar entonces su distribución denominada **distribución muestral de medias**.

Teorema 2 Si la población sigue una distribución normal $N(\mu, \sigma)$, donde μ y σ son desconocidos, si elegimos todas las muestras de cierto tamaño (n), de forma que sean representativas, entonces:

- La media de las medias muestrales de todas las muestras posibles, es igual a la media poblacional, es decir:

$$\bar{x} = \frac{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_k}{k} = \mu$$

- La desviación típica de las medias muestrales posibles es:

$$s_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Conclusión: Las medias de las muestras de tamaño n extraídas de una población de parámetros μ y σ , se aproximan por una distribución:

$$\bar{X} \longrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

siempre que n sea suficientemente grande. En la práctica se considera que n es suficientemente grande si $n \geq 30$.

Notas importantes a tener en cuenta:

1. Este resultado se conoce como *Teorema central del límite*
2. Si la población es normal, el resultado se cumple para muestras de CUALQUIER tamaño (incluso menor que 30).
3. En la práctica suele ocurrir que σ es desconocida. En estos casos el resultado sigue siendo válido aproximando σ por la desviación típica muestral s , siempre que n sea suficientemente grande ($n \geq 100$).

Ejercicio 3: La altura de los estudiantes de cierta población se distribuye según una normal de media 167 y desviación típica 3,2.

- a) Calcula la probabilidad de que un estudiante mida menos de 165 cm.
- b) Se toma una muestra de 10 estudiantes. Calcula la probabilidad de que la media muestral sea menor que 165 cm.

Ejercicio 4: Los pesos de los tornillos que fabrica cierta máquina se distribuyen según una $N(142, 32, 8, 5)$ (medidas en gramos). Se toman muestras de 25 tornillos. Calcular:

- a) Distribución que siguen las medias de esas muestras.
- b) Probabilidad de que una muestra elegida al azar de 25 tornillos tenga un peso medio superior a 144,6 gramos.
- c) La misma pregunta si la muestra es de 100 tornillos.

Ejercicio 5: Una máquina ha fabricado piezas de precisión con un peso medio de 150 gr. y una desviación típica de 20 gr. Calcular la probabilidad de que una muestra de 80 piezas tenga un peso medio de más de 155 gr. (Solución: 0'0129)

Ejercicio 6: Sabemos que el gasto mensual en electricidad (por familia) se distribuye normalmente con media 142,32 € y desviación típica 8,5 €.

- Halla la probabilidad de que una muestra de 25 familias elegidas al azar, tenga un gasto medio superior a 144,6 €.
- Realiza el mismo cálculo si la muestra que se toma es de 100 familias.

5.2. Distribución muestral de proporciones

Cuando en una población estudiamos cierta característica que sólo puede tomar dos valores: **sí** (éxito) o **no** (fracaso) nos encontramos con la distribución binomial.

Nos planteamos ahora determinar qué proporción de una población posee un cierto atributo, por ejemplo si es fumador o no fumador, si tiene ordenador o no, si tiene alergia o no, etc... El estudio de este tipo de proporciones es equiparable al de una distribución binomial (donde sólo hay dos posibilidades).

Cada una de las muestras que extraigamos tendrá un porcentaje de individuos con esa misma característica.

Llamamos p al parámetro poblacional, que es la proporción de uno de los valores que presenta la variable aleatoria en la población y q al parámetro poblacional para el otro valor ($q = 1 - p$). Si extraemos muestras de tamaño n , para cada muestra tendremos un estadístico proporción muestral que llamamos \hat{p} .

La distribución que asocia a cada muestra su proporción es la **distribución muestral de proporciones**. Como para poblaciones grandes sabemos que la binomial se aproxima a una normal, aplicando razonamientos similares a los del apartado anterior se tiene el siguiente:

Teorema 3 *Las proporciones muestrales de tamaño $n \geq 30$, extraídas de una población en la que la probabilidad de éxito es p , se ajustan a una normal*

$$N\left(p, \sqrt{\frac{p \cdot q}{n}}\right)$$

- En la práctica se considera que la aproximación es buena para $n \geq 30$, $n \cdot p \geq 5$ y $n \cdot q \geq 5$.
- En la práctica habitualmente ocurre que las proporciones poblacionales p y q son desconocidas. En estos casos se aproximan por las respectivas de una muestra \hat{p} . Esto se puede hacer por ser \hat{p} un estimador insesgado, como veremos en el epígrafe siguiente.

Ejercicio 7: Una fábrica de pasteles fabrica, en su producción habitual, un 3% de pasteles defectuosos. Un cliente recibe un pedido de 500 pasteles de la fábrica.

- Probabilidad de que encuentre más del 5% de pasteles defectuosos.
- Probabilidad de que encuentre menos de un 1% de pasteles defectuosos.

Ejercicio 8: De una población de 120 alumnos, hay 48 que tienen 2 o más hermanos. Si de dicha población se toman muestras de tamaño 40.

- ¿Qué distribución siguen las proporciones muestrales?
- ¿Cuál es la probabilidad de que se encuentre en dicha muestra una proporción de más del

55% de alumnos con 2 o más hermanos?.

Ejercicio 9: En un saco mezclamos judías blancas y pintas en la relación de 14 blancas por cada pinta. Extraemos un puñado de 100 judías. Calcula la probabilidad de que la proporción de judías pintas esté comprendida entre 0,05 y 0,1.

Ejercicio 10: Una población está formada por los elementos 1, 2, 4 y 6

- Obtén todas las muestras con reemplazamiento de tamaño 2 y calcula la proporción de cifras pares de cada muestra.
- Calcula la media y la desviación típica de la distribución muestral de proporciones.
- Compara los resultados con los que conocemos teóricamente (Teorema 3).

6. Estimación Puntual

Comenzamos este tema estudiando la teoría de muestras y hemos visto cómo la **estadística inferencial** trata de inferir información sobre una población, utilizando muestras extraídas aleatoriamente de esa población.

Dentro de la estadística inferencial está la **estadística inductiva**, que trata de **estimar parámetros poblacionales** a partir de sus correspondientes **estadísticos muestrales**.

Esta estimación puede hacerse de dos formas:

- **Estimación puntual**
- **Estimación por intervalos**

Por ejemplo si decimos que la estatura media de los españoles es de 1,75 m. estamos haciendo una estimación puntual, mientras que si decimos que la estatura media está entre 1,72 y 1,78 m. estamos haciendo una estimación por intervalos.

Por tanto la **estimación puntual** consiste en estimar mediante un único valor el parámetro poblacional desconocido.

Las estimaciones puntuales son más precisas que las estimaciones por intervalos, que veremos en el epígrafe 8, sin embargo, son menos fiables.

En la estimación puntual el estadístico que usamos para la estimación se llama **estimador puntual**. Éstos pueden ser:

- **Insesgados:** Si la media de la distribución muestral del estadístico coincide con su correspondiente parámetro poblacional.
Según hemos visto en el punto anterior:
 - La media muestral \hat{x} es un estimador insesgado de la media poblacional μ
 - La proporción muestral \hat{p} es un estimador insesgado de la proporción poblacional p
- **Sesgados:** Si la media de la distribución muestral del estadístico NO coincide con su correspondiente parámetro poblacional.

En general siempre debemos escoger estimadores insesgados ⁵

⁵Si hay más de un estimador insesgado para un mismo parámetro, se escoge el más **eficiente**, en el sentido de que su distribución muestral tenga menos dispersión

7. Intervalos de Probabilidad

En una variable normal cualquiera $N(\mu, \sigma)$, se verifica que:

1. En el intervalo $(\mu - \sigma, \mu + \sigma)$ está el 68'26% de la población.
2. En el intervalo $(\mu - 2 \cdot \sigma, \mu + 2 \cdot \sigma)$ está el 95'44% de la población.
3. En el intervalo $(\mu - 3 \cdot \sigma, \mu + 3 \cdot \sigma)$ está el 99'74% de la población.

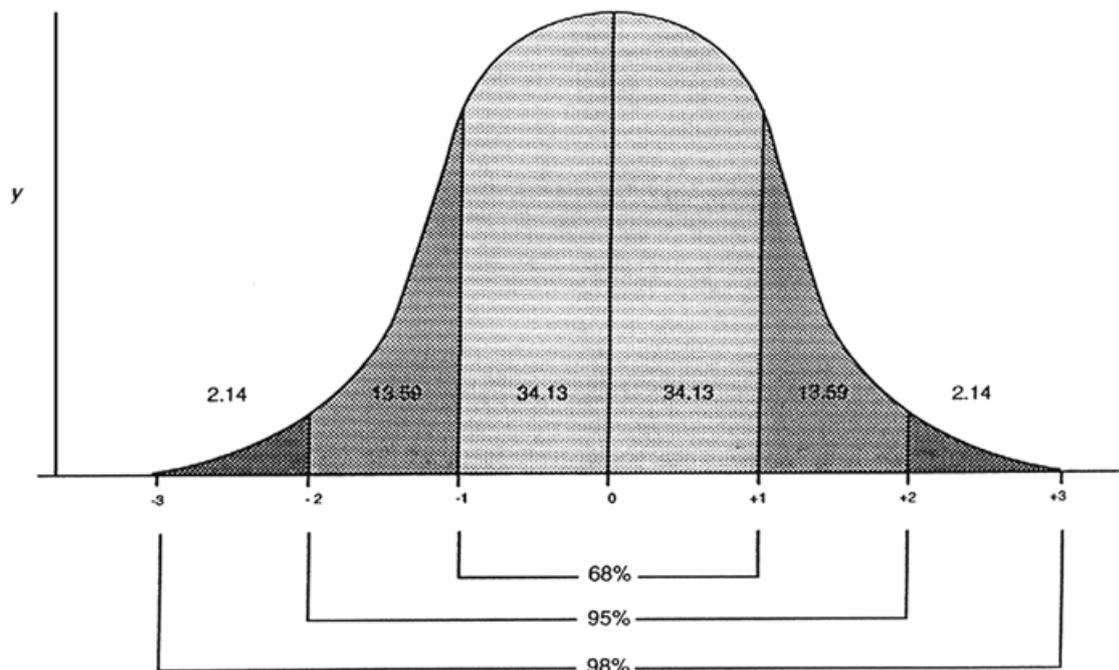


Figura 2: Porcentajes de población en los diferentes intervalos simétricos de la normal estandar $N(0, 1)$.

Es evidente que a medida que el intervalo se amplía, hay mayor porcentaje de la población en él. En general, dado un porcentaje del $N\%$, siempre es posible encontrar un intervalo simétrico respecto de la media de forma que dicho intervalo contenga a dicho porcentaje de población.

Más explícitamente, se denomina *intervalo de probabilidad* a aquel intervalo para el cuál se sabe que hay una seguridad del $N\%$ de que los parámetros muestrales (\bar{x} o \hat{p}) se encuentren en dicho intervalo. La seguridad N viene fijada previamente.

Se denomina **Nivel de confianza** al número:

$$1 - \alpha = \frac{N}{100}$$

y llamaremos **Nivel de significación** al valor α .

Nota: El nivel de confianza vendrá explicitado en las condiciones del problema, por ejemplo: Si queremos que el 85% de la población esté en el intervalo, el nivel de confianza será $1 - \alpha = \frac{85}{100} = 0,85$ mientras que el nivel de significación será $\alpha = 0,15$.

7.1. Intervalo de probabilidad para la media muestral \bar{X}

Si la población sigue una distribución de parámetros μ y σ , y las muestras son de tamaño $n \geq 30$ (o bien la población es normal y las muestras pueden ser de cualquier tamaño), sabemos que la media muestral \bar{x} sigue una distribución:

$$\bar{X} \longrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Se trata de encontrar el valor de k como en la figura:

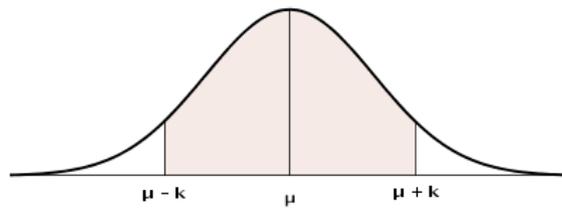


Figura 3: Buscamos el valor de k que deje en el intervalo $(\mu - k, \mu + k)$ al $(1 - \alpha) \cdot 100\%$ de la población.

Razonemos ahora sobre la normal $Z \longrightarrow N(0, 1)$ que es la que se encuentra tabulada:

Si queremos que el intervalo buscado contenga a la media muestral con una confianza de $1 - \alpha$, entonces fuera del intervalo el área tiene que ser de α , y como la curva es simétrica, en cada una de las ramas fuera de la región sombreada tenemos un área de $\frac{\alpha}{2}$.

Llamaremos $z_{\frac{\alpha}{2}}$ al punto situado en eje x que separa la región sombreada de la otra.

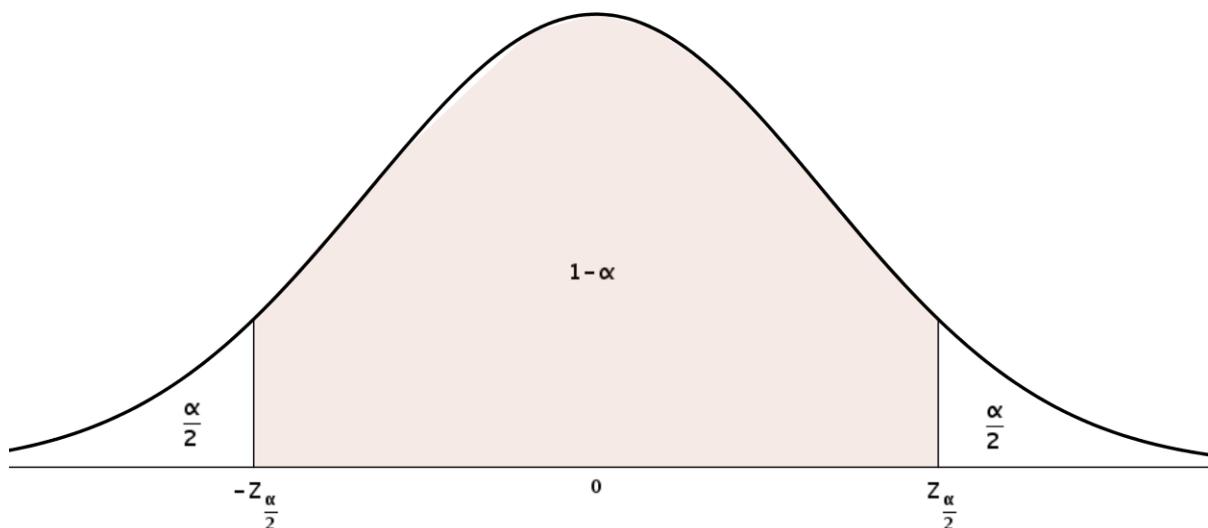


Figura 4: En la tabla de la $N(0, 1)$ buscamos el valor $z_{\frac{\alpha}{2}}$ de modo que en el intervalo $(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}})$ esté el $(1 - \alpha)\%$ de la población.

Ahora bien, este valor corresponde a una Normal $N(0, 1)$, como nosotros manejamos una $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ debemos tipificar:

$$\frac{k - \mu}{\frac{\sigma}{\sqrt{n}}} = z_{\frac{\alpha}{2}}$$

Y despejando encontramos el valor buscado:

$$k = \mu + \frac{\sigma}{\sqrt{n}} \cdot z_{\frac{\alpha}{2}}$$

De modo que, dado el nivel de significación α o el de confianza $1 - \alpha$, podemos determinar el intervalo de probabilidad para la media muestral, que será:

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

Veamos algún **ejemplo**:

Sabiendo que la población de recién nacidos sigue una normal de media $\mu=3100$ gr. y desviación típica $\sigma=150$ gr. Calcular el intervalo de probabilidad con un nivel de confianza del 95 % para la media de una muestra de 100 recién nacidos.

Solución: Para un nivel de confianza del 95 %, el nivel de significación es $\alpha = 0.05$ y en cada zona fuera de la región queda $\frac{\alpha}{2} = 0.025$. Debemos buscar en la tabla de la $N(0, 1)$ el valor $z_{0,025}$, es decir, el valor que deja a su derecha un área de 0.025

$$P(Z \geq z_{0,025}) = 0,025 \implies P(Z \leq z_{0,025}) = \underbrace{0,975}_{\text{mira en la tabla}} \implies z_{0,025} = 1,96$$

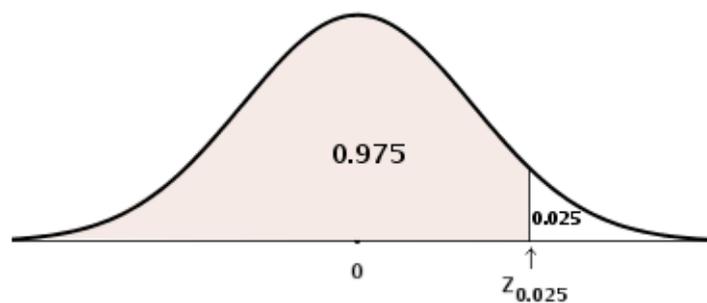


Figura 5: Obtención del valor z para un nivel de confianza del 95 %

Sabemos que la media muestral sigue una distribución:

$$N\left(3100, \frac{150}{\sqrt{100}}\right)$$

Por tanto el intervalo buscado es :

$$\left(3100 - 1,96 \cdot \frac{150}{\sqrt{100}}, 3100 + 1,96 \cdot \frac{150}{\sqrt{100}} \right) = (3070,6, 3129,4)$$

Esto significa que el 95 % de las muestras de tamaño 100 tendrá su media comprendida entre estos dos valores: (3070,6 , 3129,4).

Nota: Como n es $100 \geq 30$ el resultado también sería cierto incluso si el peso de los recién nacidos fuera una variable que no siguiera una distribución normal.

Ejercicio 11:

Las notas de una población de 150 alumnos siguen una distribución de media 5,5 y varianza $\sigma^2 = 4,1616$. Extraemos muestras de tamaño 36. Calcula el intervalo de probabilidad para un nivel de confianza del: a) 75 % b) 86'64 %, e interpreta los resultados.

7.2. Intervalo de probabilidad para la proporción muestral \hat{p}

Razonando de manera análoga podemos llegar a obtener un intervalo para la proporción muestral \hat{p} con un nivel de significación α .

$$\left(p - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot q}{n}} , p + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot q}{n}} \right)$$

donde p y q son las proporciones poblacionales y $n \geq 30$.

Ejercicio 12:

Sabiendo que la proporción de alumnos con teléfono móvil de una población de 120 alumnos es de $p = 0,7$. Halla el intervalo de probabilidad para la proporción de:

- las muestras de tamaño 30 con una confianza del 75 %.
- las muestras de tamaño 49 con una confianza del 90 %.
- las muestras de tamaño 49 con una confianza del 99 %.

8. Estimación por intervalos

Si queremos estimar la estatura de un jugador de baloncesto y decimos “1,85 metros” estamos haciendo una estimación puntual que no dice nada sobre la seguridad o duda de que esto sea cierto, sin embargo si decimos: “estoy casi seguro que mide entre 1,80 y 1,90”, entonces tenemos un cierto grado de confianza de que eso sea cierto.

Es por esto que la estimación puntual se utiliza poco, pues no tenemos datos suficientes que nos indiquen el grado de fiabilidad del dato muestral que hemos tomado. Lo que tiene más sentido plantearse es cuál es la probabilidad de que la media o proporción poblacional pertenezcan a un intervalo determinado.

Logicamente cuanto mayor sea el intervalo mayor será el grado de confianza que podamos tener, aunque también será mayor el error que cometamos al hacer la estimación.

8.1. Estimación de la media de una población μ

La media μ de una población es desconocida y deseamos conocerla. Para ello, basándonos en los intervalos de probabilidad, sabemos que si la población tiene parámetros μ y σ , la media muestral \bar{x} sigue una distribución:

$$\bar{X} \longrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

siendo n el tamaño de la muestra.

Sabemos también que el intervalo de probabilidad para \bar{x} con un nivel de confianza $(1 - \alpha)$ es:

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

De donde se tiene que:

$$\mu - \frac{\sigma}{\sqrt{n}} \cdot z_{\frac{\alpha}{2}} \leq \bar{x} \leq \mu + \frac{\sigma}{\sqrt{n}} \cdot z_{\frac{\alpha}{2}}$$

Ahora despejando μ en la primera desigualdad:

$$\mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} \cdot z_{\frac{\alpha}{2}}$$

y despejando en la segunda desigualdad:

$$\mu \geq \bar{x} - \frac{\sigma}{\sqrt{n}} \cdot z_{\frac{\alpha}{2}}$$

Por lo que:

$$\bar{x} - \frac{\sigma}{\sqrt{n}} \cdot z_{\frac{\alpha}{2}} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} \cdot z_{\frac{\alpha}{2}}$$

Conclusión: El intervalo de confianza para la media μ desconocida es:

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

Notas:

- Este resultado es cierto siempre que tengamos la certeza de que la población objeto de estudio sigue la distribución normal, o bien, que el tamaño de las muestras sea suficientemente grande ($n \geq 30$). Recuerda el TEOREMA2.
- Cuando la desviación típica poblacional σ es desconocida, en su lugar se usa la desviación típica muestral s , con lo que el intervalo queda:

$$\left(\bar{x} - \frac{s}{\sqrt{n}} \cdot z_{\frac{\alpha}{2}}, \bar{x} + \frac{s}{\sqrt{n}} \cdot z_{\frac{\alpha}{2}}\right)$$

Al valor $\frac{\sigma}{\sqrt{n}}$ se le llama **Error típico**

Ejercicio 13: Para estimar la media de los resultados que obtendrían al resolver un cierto test los alumnos de 4º de E.S.O. de la Comunidad de Andalucía, se les pasa el test a 400 alumnos escogidos al azar, con los resultados de la siguiente tabla:

Puntuación	Número de alumnos
1	24
2	80
3	132
4	101
5	163

Estima con un nivel de confianza del 95 % el valor de la media poblacional.

Ejercicio 14: De una variable estadística conocemos la desviación típica 8, pero desconocemos la media. Para estimarla, extraemos una muestra de tamaño 60 cuya media es 37. Estimar la media poblacional con una confianza del 99 %.

8.2. Estimación de una proporción p

Si para una población se desconoce la proporción p de individuos que poseen cierta propiedad, y deseamos dar un intervalo de confianza para el valor de p , como el intervalo de probabilidad para la proporción muestral, \hat{p} , para el nivel de confianza $1 - \alpha$ en una muestra de tamaño n es:

$$\left(p - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot q}{n}}, p + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot q}{n}} \right)$$

Razonando igual que en el caso anterior encontramos que el intervalo de confianza para p con un nivel de significación α es:

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

Y como además no se conocen ni p ni q se utilizan en su lugar \hat{p} y \hat{q} , quedando:

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

Notas:

- Para poder aplicar este resultado es necesario que $n \geq 30$
- Habitualmente en las encuestas se suele utilizar esta fórmula con el valor $p = q = 0,50$ porque es la situación más desfavorable.

Ejercicio 15: Determina el intervalo de confianza, con una significación del 0,05 para la proporción poblacional de fumadores entre los jóvenes menores de 18 años, a partir de una muestra de tamaño 900, cuando no se conocen valores de p anteriores. Considera los dos casos anteriores (usando \hat{p} y usando $p = q = 0,5$). La proporción de fumadores en la encuesta ha sido de $\hat{p} = 0,3$.

8.3. Error máximo admisible

Al estudiar los intervalos de confianza hemos visto que su amplitud depende del factor:

$$z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{o} \quad z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot q}{n}}$$

En el caso de la media se tiene: $|\mu - \bar{x}| = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

En el caso de la proporción: $|p - \hat{p}| = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot q}{n}}$

Se llama **Error máximo admisible** para la estimación de medias o de proporciones, respectivamente a:

$$E = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{o} \quad E = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot q}{n}}$$

Observamos que:

- El error es mayor al aumentar el nivel de confianza, ya que el valor $z_{\frac{\alpha}{2}}$ aumenta como podemos observar en la tabla:

Confianza = $1 - \alpha$	$z_{\frac{\alpha}{2}}$
0,90	1,645
0,95	1,960
0,99	2,575

- Al aumentar el tamaño muestral n disminuimos el error porque dividimos por un número mayor: \sqrt{n}

Por tanto: **Para reducir el error no hay que aumentar la confianza, sino el tamaño de la muestra elegida.**

8.4. Tamaño de las muestras

Hemos visto que el tamaño de las muestras n es inversamente proporcional al error admisible E . A partir de la expresión del error admisible podemos despejar el valor del tamaño de la muestra n :

- Para la **estimación de medias**: $n = \left(\frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{E}\right)^2$
- Para la **estimación de proporciones**: $n = \frac{Z_{\frac{\alpha}{2}}^2 \cdot p \cdot q}{E^2}$

Ejercicio 16: Al medir un tiempo de reacción, un psicólogo sabe que la desviación típica del mismo es 0,5 segundos. ¿Cuál es el número de medidas que deberá realizar para que con una confianza del 99%, el error de estimación no exceda de 0,1 segundos?.

Ejercicio 17: Se sabe que el tiempo de dedicación de los jóvenes al ocio sigue una distribución normal de media 400 minutos y desviación típica 63 minutos. Halla el tamaño mínimo de la muestra de jóvenes que garantiza con una probabilidad de 0,95 que el tiempo medio de ocio está entre 382 y 418 minutos.

Ejercicio 18: Para 96 familias españolas elegidas al azar se ha determinado que la TV permanece encendida en la casa una media de 217 minutos diarios, la desviación típica de la muestra fue de 40 minutos.

a) Para una fiabilidad del 95% ¿qué error se asume cuando se da por bueno ese dato para el total de las familias españolas?.

b) ¿Qué tamaño muestral sería necesario para reducir ese error a la mitad?.

NOTA IMPORTANTE

Diferencia entre intervalos de probabilidad y de confianza

En un intervalo de probabilidad lo que conocemos es la media y desviación típica poblacionales, y damos el intervalo donde se encontrará (para un cierto nivel de confianza) la media muestral o la proporción muestral.

Sin embargo, en un intervalo de confianza entramos ya en el terreno de la estimación, es decir:

NO conocemos la media poblacional (y en ocasiones tampoco la desviación típica poblacional) ni la proporción poblacional, sino que sólo conocemos, o podemos calcular, la media muestral o la proporción muestral, y de lo que se trata es de dar un intervalo en el que se encuentre la media poblacional (o la proporción poblacional).

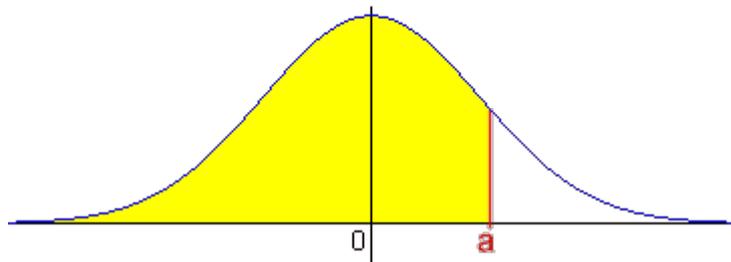
9. Actividades Finales

- Supongamos que una población se compone de niños de edades 2, 3, 6, 8 y 11 años. Considera todas las muestras posibles de 3 niños (con reemplazamiento) que pueden formarse. Halla:
 - La media y la desviación típica de la población.
 - La media y la desviación típica de la distribución muestral de medias.
 - ¿Qué relación hay entre los resultados obtenidos en a) y b) ?
- Suponiendo que las puntuaciones de un test de inteligencia se distribuyen según una Normal $N(100,15)$.
 - Calcula la probabilidad de que una muestra de tamaño 49, extraída de esa población, tenga una media inferior a 98.
 - Calcula la probabilidad de que una muestra de tamaño 81, extraída de esa población, tenga una media superior a 105.

3. Se supone que la estatura de los jóvenes de 18 años de cierta población sigue una distribución normal de media 162 cm y desviación típica 12 cm. En una muestra tomada al azar de 100 de esos jóvenes:
 - a) ¿Cuál es la probabilidad de que la media esté entre 159 y 165 cm?.
 - b) ¿Cuántos de esos jóvenes tienen su estatura entre esos valores?.
4. En una determinada población se sabe que el 20 % de las personas usan gafas graduadas y el resto no. Tomamos una muestra de 256 personas. ¿Cuál es la probabilidad de que el porcentaje de personas encuestadas que usan gafas esté entre el 15 % y el 25 %?
5. En una muestra aleatoria de 1000 personas, están a favor de que el ministerio de economía mantenga la presión fiscal el 65 %. Halla el intervalo de confianza del 99 % para la proporción.
En una encuesta realizada un año antes había resultado un 68 % favorable al mantenimiento de la presión fiscal, ¿cae este valor dentro del margen de confianza de la nueva encuesta?. ¿Qué podemos decir sobre el cambio de opinión de la población de un año a otro?
6. Se sabe que la desviación típica del peso de los individuos de cierta población es 6 kg. Calcula el tamaño de la muestra que se ha de considerar para, con un nivel de confianza del 95 %, estimar el peso medio con un error inferior a 1 kg.
7. En una encuesta de opinión, durante una campaña electoral en una ciudad, se preguntó a una muestra aleatoria de 400 personas a cuál de los dos candidatos pensaban votar. Declararon 160 que votarían a un determinado partido. Obtén un estimador puntual y un intervalo de confianza del 95 % para la proporción de ciudadanos que votará a ese partido en las elecciones.
8. La edad media de esperanza de vida de una población es 50 años, con una desviación típica de 10 años. Una compañía de seguros quiere determinar el tamaño de una muestra para que la estimación difiera del valor 50 en al menos 2 % de este valor, tomando como nivel de confianza el 95 %. Calcula el tamaño de dicha muestra.
9. Deseamos conocer el número de personas mayores de edad, que sería necesario incluir en una muestra nacional, para estimar su proporción con un error de $E=0,04$ y un nivel de confianza del 99,73 %. Se dispone de un valor $p = 0,45$ del último censo.
10. La desviación típica de los habitantes de cierto país es 10 cm. Calcula el tamaño mínimo que ha de tener una muestra de habitantes de dicho país para que el error cometido al estimar la altura media sea inferior a 1 cm con un nivel de significación $\alpha = 1 \%$.
11. La estatura media de los niños de 10 años en España es de 135 cm, con una varianza de 64 cm^2 . Calcula el tamaño de muestra necesario para que el intervalo de confianza al 95 % de la media muestral tenga una amplitud de 2 cm.
12. Según una encuesta electoral, la intención de voto a cierto partido político está entre el 42 % y el 48 %. Se trata de un intervalo de confianza, pero en la ficha técnica no aparece el nivel de confianza, sólo aparece el tamaño de la muestra $n = 1056$ individuos. Obtén el nivel de confianza.

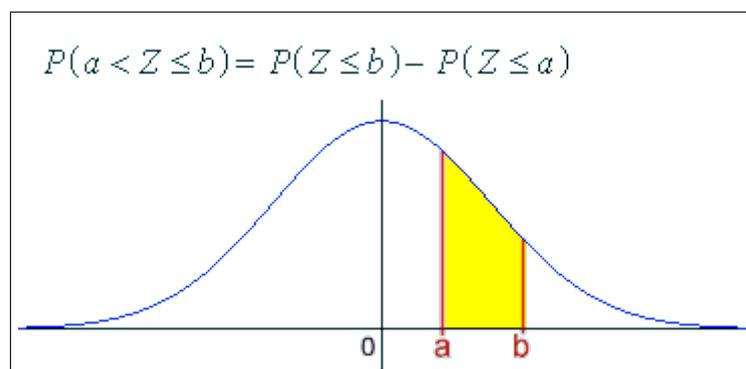
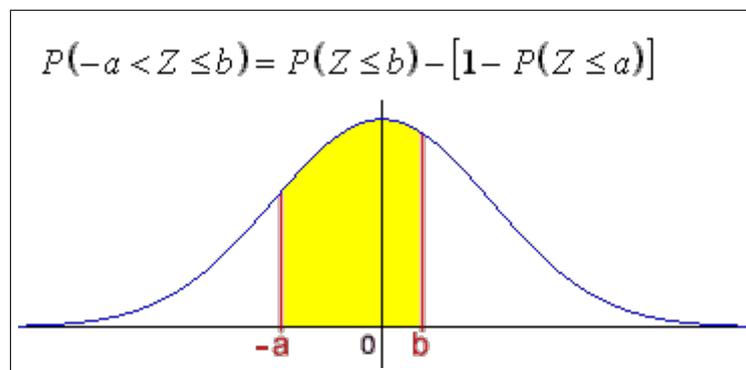
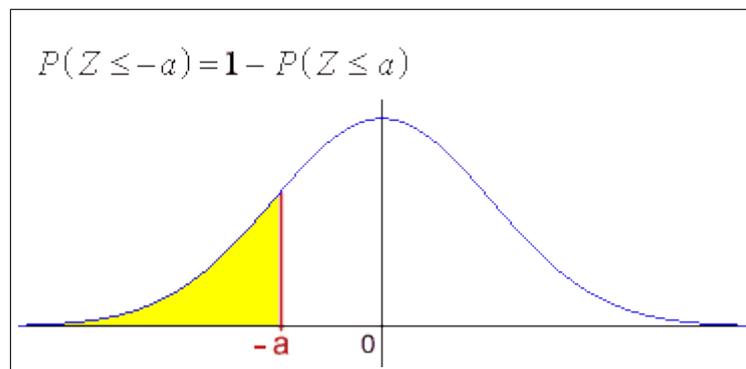
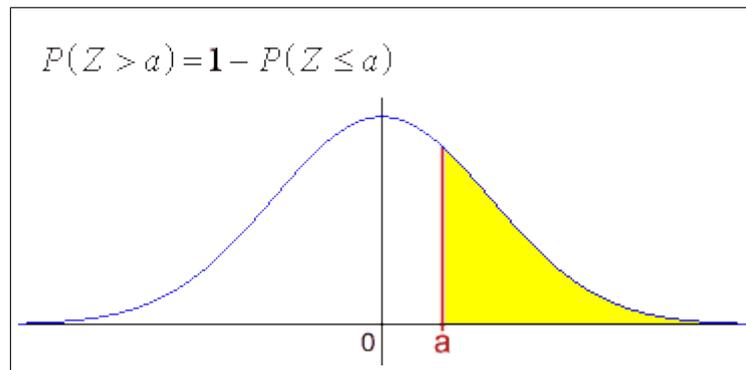
10. Anexo I: Tabla de la Distribución Normal Estandar

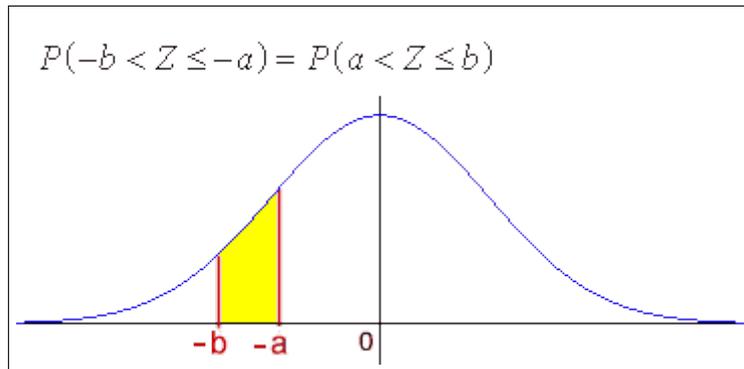
$$P(Z \leq a) \rightarrow \text{Tablas}$$



	.00	.01	.02	.03	.04	.05	.06	.07	.08	.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	0.8079	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

11. Anexo II





12. Anexo III: La Distribución Binomial

Se conoce como **experimento de Bernouilli** a un experimento aleatorio que sólo tiene dos resultados posibles, que son complementarios entre sí y se denominan: **éxito** y **fracaso**. Consideramos un experimento aleatorio con las siguientes características:

1. Cada prueba del experimento es una prueba de Bernouilli, es decir, sólo son posibles dos resultados: el suceso A (éxito) y su contrario \bar{A} (fracaso).
2. El resultado obtenido en cada prueba es **independiente** de los resultados obtenidos en las pruebas anteriores.
3. La probabilidad del suceso A es siempre **constante** y no varía de una prueba a otra. $P(A) = p$ y $P(\bar{A}) = q = 1 - p$.

Definimos la variable aleatoria X como el número de éxitos obtenidos en n pruebas. Entonces X es una variable aleatoria **discreta** que sigue la denominada como **distribución de probabilidad binomial**.

Una distribución binomial queda caracterizada por dos parámetros: El número de pruebas realizadas \mathbf{n} y la probabilidad del suceso éxito \mathbf{p} y se representa por **$\mathbf{B(n,p)}$** .