UNIVERSIDAD NACIONAL DE SAN LUIS

FACULTAD DE PSICOLOGÍA

METODOLOGÍA DE LA INVESTIGACIÓN I

Licenciatura en Psicología

Licenciatura en Psicomotricidad

Prof. Responsable: Dr. HORACIO DANIEL GARCIA

Trabajos prácticos: Lic. DANIEL PITONI, Lic. MAXIMILIANO SAPINO y Mg. ELIANA ZÁRATE

Unidad 5: Probabilidades y distribuciones continuas de probabilidad
Definición clásica de la probabilidad: Propiedades. Distribuciones continuas de probabilidad: Normal, Normal estándar, "t" de Student, Chi-cuadrado. Características, cálculo de probabilidades. Teorema del Límite Central: Error estándar de la media. Tipos de estimadores: Propiedades. Estimación puntual e intervalar de parámetros.

Autor: HORACIO DANIEL GARCIA

Año 2023

Introducción a las probabilidades

Lo primero que uno piensa cuando se introduce en el concepto de probabilidad es en lo engorroso y dificultoso que puede ser el tema, desconociendo que esta noción está frente a nosotros durante la mayor parte del día en nuestra vida cotidiana.

Como te darás cuenta, interactuamos permanentemente con fenómenos cuyos resultados no se pueden determinar con absoluta certeza; ¿Vendrá a visitarme mi amiga hoy? ¿Ganaré el premio si apuesto a un número en la quiniela? ¿Me encontraré algo valioso que alguien haya olvidado camino a la facultad? ¿Me tocará dar lección hoy día? etc… si bien no tenemos seguridad de la ocurrencia de estas cosas, intuimos un grado de probabilidad de que aquello pueda ocurrir.

Pensemos en el siguiente ejemplo en forma de pregunta: ¿Qué es más probable de acertar?: ¿Apostar a un número en la quiniela a la cabeza, o a los diez primeros? ¿Apostar a cuatro cifras o a dos?

Sin pensarlo demasiado, darás seguramente una respuesta bastante coherente, ¿cierto?

Definición clásica de la probabilidad: Propiedades

La teoría clásica de probabilidades surgió estudiando los denominados experimentos aleatorios, para dar respuesta a la necesidad de obtener resultados, datos, en aquellos sucesos donde está involucrado el azar. Un experimento es aleatorio si puede dar lugar a varios resultados sin que se pueda predecir con certeza un resultado concreto. Es decir, al repetir el experimento bajo condiciones similares se obtendrán resultados que, en general, serán diferentes. Por ejemplo, tirar una moneda al aire y verificar de qué lado cae.

De hecho, en 1900, un estadístico muy prestigioso llamado Pearson realizó el experimento; hizo un total de 24000 lanzamientos de una moneda y registró que unas 12012 veces cayó cara y unas 11988 veces cruces. Claro, por la Teoría clásica de la probabilidad, sabemos hoy que cuando un experimento aleatorio se repite un gran número de veces, los posibles resultados (cara o cruz) tienden a presentarse proporcionalmente; o, dicho de otro modo, luego de realizados muchos experimentos la frecuencia de aparición de cada resultado tiende a estabilizarse.

Sin embargo, lo que le interesa a esta rama de la estadística es saber la probabilidad de ocurrencia de un fenómeno sin tener que repetir tantas veces el experimento.

La probabilidad mide qué tan factible es que suceda un evento o no, y lo hace por medio de una escala que va de 0 a 1. Esto quiere decir que, si la probabilidad de que un evento suceda es 0, nos está dando la idea de que ese evento seguramente no sucederá; si el valor de probabilidad es 1, entonces estaremos 100 % seguros que el evento sucederá. Así, podemos darnos cuenta que: un valor de probabilidad cercano a 0 significa que es poco probable que suceda; en cambio, más cercano a 1 nos indica mayores probabilidades de que se produzca el suceso.

La probabilidad clásica se interesó en las situaciones en donde todos los casos posibles de un evento tienen la misma probabilidad de ocurrir (por ejemplo: tirar una moneda, los dados, etc.). Con ese principio determinó que:

La probabilidad de un evento "P(A)" es igual al número de casos favorables (CF), dividido entre todos los casos posibles (CP). Es decir: $P(A) = \frac{CF}{CP}$

Tomemos como ejemplo el experimento de la moneda que realizó Pearson, primero calcularemos su probabilidad y luego compararemos ese resultado con lo que ocurrió realmente.

Nos vamos a proponer averiguar cuál es la probabilidad de que una moneda tirada al aire, cuando caiga, muestre uno de sus lados que llamaremos: "cara".

Ok, ahora veremos las anotaciones necesarias:

P(A) = Probabilidad del evento (A) (hace referencia a que la moneda caiga del lado “cara”, es lo que nos interesa averiguar, es lo que nos preguntamos en el enunciado)

CF = 1 Esto hace referencia a la cantidad de casos FAVORABLES, ¿Cuantos lados de la moneda son “cara”?... solo uno, ¿verdad?

CP = 2 Esto hace referencia a la cantidad de casos POSIBLES, ¿Cuantos lados tiene la moneda?... dos, verdad (cara y cruz)

Entonces tenemos que: $P(A) = \frac{1}{2}$ $P(A) = 0.50$ (este resultado nos informa hay un 50% de probabilidad de que la moneda caiga del lado cara).

Ahora haremos el cálculo con el registro de los lanzamientos que realizó Pearson con su experimento….

A) Cantidad de veces que salió “cara” = 12012

B) Cantidad de veces que hizo el experimento = 24000

Si calculamos la división de A/B (la cantidad de veces que salió “cara” y la cantidad de veces que hizo el experimento), tendremos un valor de 0,5005. Como puedes observar, un valor muy similar al cálculo que hicimos averiguando la probabilidad mediante la fórmula.

Propiedades más importantes de la Teoría Clásica de la Probabilidad: (presta especial atención, necesitarás entenderlo y recordarlo más adelante)

1. La suma de las probabilidades de TODOS LOS SUCESOS POSIBLES es igual a 1.
2. Probabilidad de un SUCESO IMPOSIBLE es igual a 0.
3. Si la probabilidad de un suceso (A) y su contrario (B) vale 1, la probabilidad de B es igual a la resta de 1 – P(A) (la probabilidad de A).
4. Si un suceso está incluido en otro, su probabilidad es menor o igual a la de éste. $A \subset B \Rightarrow P(A) \leq P(B)$. La anotación dice que si A está incluido en B entonces la probabilidad de A es menor o igual que la probabilidad de B.

Si bien existen otras propiedades que han sido enunciadas, es conveniente que entiendas éstas, ya que serán clave para la comprensión del resto de la materia.

Distribuciones continuas de probabilidad: Normal, Normal estándar, "t" de Student, Chi-cuadrado. Características, cálculo de probabilidades

Cuando investigamos, muchas veces es necesario inferir y decidir sobre situaciones en las que hay diferentes probabilidades de ocurrencia en los resultados; por ejemplo, cuando pretendemos saber qué representa en la población el dato que hemos obtenido en la muestra.

En todos los procedimientos estadísticos donde infiero una característica de la población con los datos de una muestra, o cuando necesito contrastar la hipótesis que planteé para determinar si hay evidencia suficiente que la sustente, necesito apoyarme en nociones de probabilidad.

Sin embargo, no podremos utilizar la Teoría clásica de probabilidad en todos los casos. Sucede que la Teoría clásica es de mucha utilidad en fenómenos que se manifiestan como una constante probabilística, pero los fenómenos que usualmente

investigamos en nuestro campo lejos de ser constantes, son variables desde el punto de vista probabilístico, por lo que debemos pensar en términos de distribución de probabilidad de la variable.
Así, la Teoría moderna de la probabilidad introduce la idea de distribución de probabilidad, abrigando la idea de que la probabilidad puede ser variable en función de las características que tenga el fenómeno a estudiar. La idea sería que, conociendo como se distribuye una variable, se puede calcular una función matemática que describa el comportamiento de la misma y así determinar la probabilidad de que ocurra un evento. Por ejemplo, con una probabilidad del 95%, ¿entre qué valores se encuentra la media poblacional de la variable asertividad?

Uno de los aportes más importantes, derivados de la Teoría moderna de las probabilidades, ha sido el concepto de variable aleatoria. Con ello se pudo superar las restricciones que presentaba la Teoría clásica de probabilidad, en la cual sólo se podía predecir la probabilidad en casos muy concretos, no pudiendo aplicarse, por ejemplo, para determinar la probabilidad de encontrar un sujeto que mida más de 1,84m de altura en una muestra determinada.

Pero, ¿qué es una variable aleatoria?
Una variable aleatoria puede definirse como una función matemática, que tiene la capacidad de brindar valores numéricos a todos los posibles resultados de un evento que aún no ha ocurrido. En otras palabras, una variable aleatoria es una variable teórica (abstracta) a la que recurrimos para calcular el comportamiento de un dato de la variable que estoy estudiando (datos reales). Este paralelismo entre la variable aleatoria y la variable real sólo es posible hacerlo si tienen una distribución similar (de lo contrario no se podría).
¿Por qué nos apoyamos en una variable teórica para calcular algo acerca de la variable real con la que estamos trabajando? Sucede que las variables aleatorias ya están definidas por funciones matemáticas y, como ya se les ha calculado distribución de probabilidad, se las toma como distribuciones genéricas. Así, las personas que se dedican a su estudio han construido diversos modelos de distribuciones de probabilidad, que representan el comportamiento teórico de diferentes fenómenos aleatorios que aparecen en el mundo real; por lo que, en lugar de que realicemos cálculos repetidos sobre cada variable real que estudiemos, en cada situación en particular, (lo que implicaría la necesidad de contar muestras enormes, muchísimo esfuerzo y tiempo) simplemente nos apoyamos en las distribuciones de probabilidad teóricas. Logrando con ello realizar estimaciones de medias poblacionales, comparaciones entre grupos, estudiar relaciones entre variables, etc.

Un modelo de probabilidad es una simplificación de la realidad que pretende representar o predecir y, si éste tiene la capacidad, de reflejar las propiedades más importantes del mundo que nos rodea, resulta una herramienta extremadamente útil.

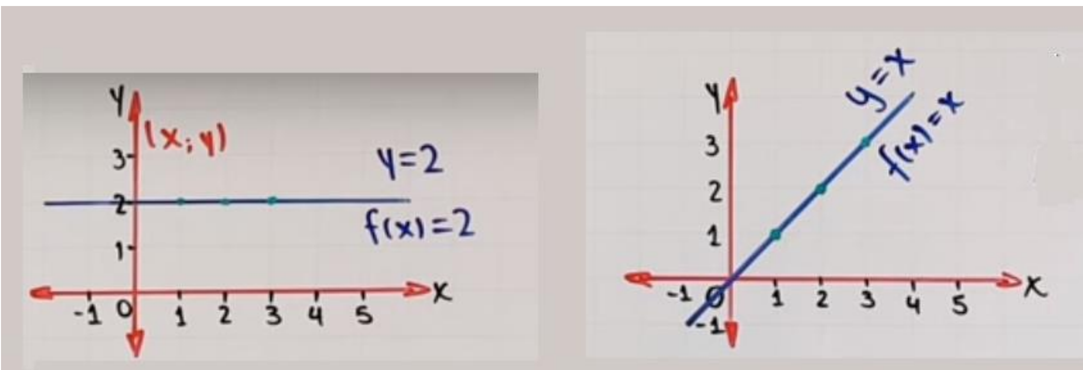


Figura 1. Algunas funciones de distribución simples

Arriba verás dos modelos (funciones) simples, que entenderás seguramente; veamos.

A la izquierda, si observas, verás que hay una variable (x) que siempre toma el valor 2 (en el eje y). Es decir, que a medida que el valor de (x) aumenta, el valor que denota (y) es constante (2), por lo tanto, si tenemos que escribir su función lo debemos hacer así **F(x) = 2** (lo que se expresa de esta manera: función de (x) igual a dos).
A la derecha, en cambio, vemos que la variable aumenta en (y) conforme nos desplazamos por (x); por cada valor de (x) aumenta uno en (y). En ese caso decimos que: la función de (x) igual a (x) por lo que se escribe así **F(x) = x**

Como podrás apreciar, estos modelos describen funciones en las variables que son útiles, pero que no siempre se adaptan a los intereses de los investigadores en ciencias sociales.
Pensemos, ¿cómo se distribuye la variable inteligencia en la realidad? … bueno por lo pronto ¿que sabemos?

a) Que hay pocas personas que tienen escasa inteligencia
b) Que la mayoría tiene una inteligencia media
c) Que hay pocas personas que son muy inteligentes

Entonces, si tuviéramos que graficar la distribución de la inteligencia en la población tendríamos que:

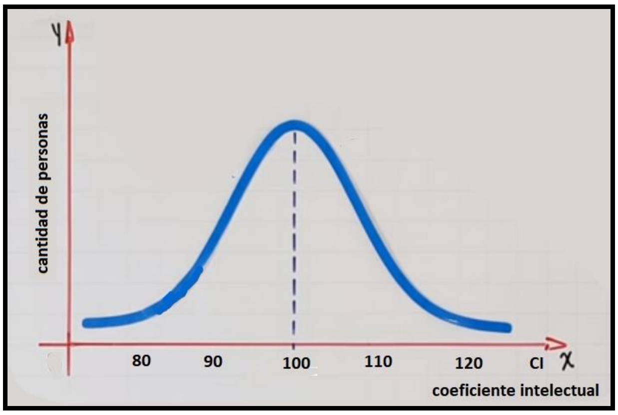


Figura 2. Distribución de la variable coeficiente intelectual

La gráfica representa que la mayoría posee un coeficiente intelectual en torno a los 100 puntos y que decrece la cantidad de personas en los extremos de los valores, lo que dibuja una curva con forma de campana.

Muy bien, si estás entendiendo hasta aquí, nos encontramos en condiciones de pasar a describir las distribuciones más usadas en ciencias humanas y de la salud, ya que la mayoría de las variables con las que trabajamos los psicólogos y psicomotricistas tienen distribuciones similares. Veremos que uno de los modelos para representar las distribuciones empíricas más frecuentes, de mayor utilidad para variables continuas, es la distribución normal.

Distribución Normal

Dada las características de las variables con las que trabajamos los investigadores, la distribución normal es una de las de mayor interés en el área de la estadística. Su desarrollo es atribuido a Carl Friedrich Gauss, quién fue el primero que aplicó esta idea al estudio de fenómenos astronómicos. Él encontró una función (una fórmula matemática) que permitía describir una gran cantidad de variables que son muy utilizadas en el campo de:

- Morfología de los seres vivos: tallas, pesos, diámetros, perímetros.
- Fisiología: efecto de los fármacos.
- Sociología: consumo, distribución de recursos, etc.
- Psicología: cociente intelectual, asertividad, tolerancia a la frustración, personalidad, etc.

- Estadística: errores cometidos al medir ciertas magnitudes, probabilidad, tipificación de variables, etc.

La función que propuso Gauss se expresa en una gráfica, en la cual es posible identificar tres partes bien diferenciadas: la zona media, en cuyo centro se encuentra el valor de la media (mayor cantidad de datos) y es cóncava; y los dos extremos (menor cantidad de datos), curvas convexas que tienden a aproximarse al "eje x".

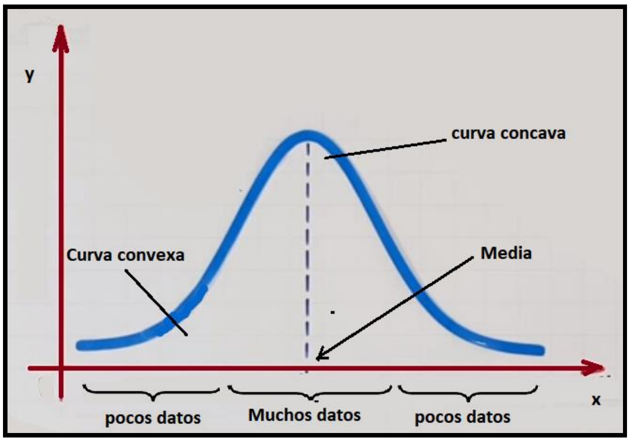


Figura 3. Elementos básicos de una distribución normal

Como podrás ver, la forma de campana describe que los elementos más comunes son los que están más centrados (ubicados en el medio), mientras que los más raros, menos frecuentes, se sitúan en los extremos.

La distribución normal, con desviaciones estándar sucesivas con respecto a la media (que le dan forma de campana), tienen la particular característica de establecer valores de referencia para estimar el porcentaje de observaciones de los datos, ya que es tomada como una variable aleatoria. Estos valores de referencia son la base de muchas pruebas de hipótesis, como las pruebas z y t que veremos más adelante.

Esta distribución considera dos parámetros: la media (μ) y la desviación estándar (σ)

Gauss la definió por esta función: $\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}$

Mediante la integral de arriba le fue posible determinar áreas debajo de una curva (campana) que representan la probabilidad de encontrar determinados rangos de valores en la distribución. No es necesario que recuerdes, ni que sepas calcular la integral de Gauss, ya que requiere un proceso complejo, además existe una tabla donde ya están calculados los valores de interés (tabla Z de distribución normal), que se encuentra al final de este documento, y que usaremos más adelante.

Veamos cómo se representan las probabilidades en una distribución Normal. Para comprender más convenientemente esto, te recomendamos que prestes atención al siguiente punto.

Propiedades de la distribución normal

La distribución normal posee ciertas propiedades importantes que conviene destacar:

1. Tiene una única moda, que coincide con su media y su mediana.
2. La curva normal toma valores entre $-\infty$ y $+\infty$, por lo tanto el área total bajo la curva es igual a 1 (esto quiere decir que incluye todos los fenómenos posibles, es decir que contiene el 100% de probabilidad).

3. Es simétrica con respecto a su media (μ). Es decir que hay la misma cantidad de casos a la derecha y a la izquierda de la media, por lo tanto, existe una probabilidad de un 50% de observar un dato mayor o igual que la media, y un 50% de observar un dato menor o igual.
4. La distancia entre la línea trazada en la media y el punto de inflexión de la curva es igual a una desviación típica (σ = 1). Cuanto mayor sea (σ), más aplanada y alargada será la curva de la densidad.
5. El área bajo la curva, comprendido entre los valores situados aproximadamente a dos desviaciones estándar de la media, es igual a 0.95. En otras palabras y con mayor precisión, hay un 95% de posibilidades de observar un valor comprendido en el intervalo (μ – 1,96 σ; μ + 1,96 σ).
6. La forma de la campana de Gauss depende de los parámetros μ y σ. La media (μ) indica la posición de la campana y la desviación estándar (σ), por ser una medida de dispersión, determina el grado de apuntamiento o de aplanamiento de la curva (un valor pequeño de este parámetro indica mayores posibilidades de obtener datos cercanos al valor medio de la distribución).

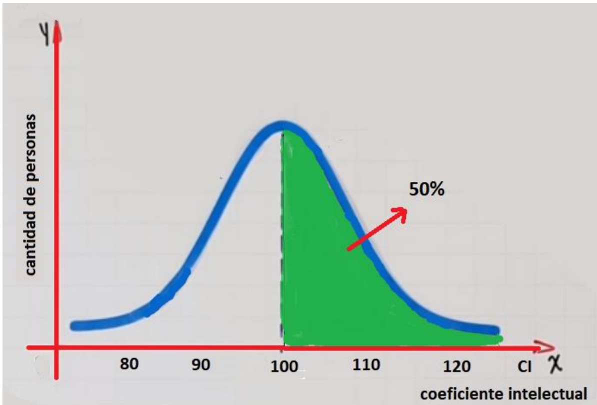


Figura 4. Distribución de la variable coeficiente intelectual

Como vemos arriba, y teniendo en cuenta la propiedad número 3, podemos saber que la probabilidad de encontrar en nuestra distribución a una persona con una inteligencia superior al valor de la media es del 50%.

Abajo te dejamos un resumen de cómo se ha determinado, en función de la integral propuesta por Gauss, las probabilidades en una variable con distribución normal.

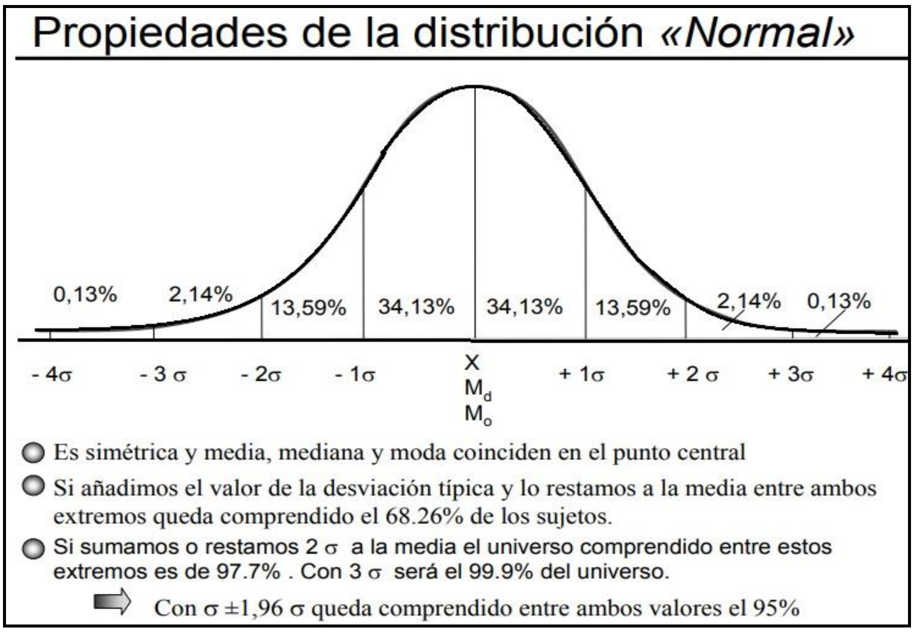


Figura 5. Determinación de probabilidades en una variable continua con distribución normal. Tomado de https://www.um.es/docencia/pguardio/documentos/Tec_resi.pdf

Cálculo de probabilidades

Arriba decíamos que una de las bondades de ésta distribución es que se ajusta para determinar con mucha eficacia la probabilidad de ocurrencia de algún valor en una variable aleatoria. <u>*Antes de ver cómo se procede para calcular la probabilidad de obtener un valor en nuestra muestra (datos reales) revisaremos el cálculo de probabilidad en una variable teórica estandarizada.*</u>

<u>*Distribución normal estándar*</u>

La distribución normal estándar, es aquella distribución normal que tiene una **media igual a cero** (μ = 0)**,** y una **desviación estándar igual a uno** (σ =1)**.** Bien, ahora lo que le vamos a prestar atención es a la función densidad normal estandarizada, que trabaja teniendo en cuenta los valores Z (en el eje horizontal).

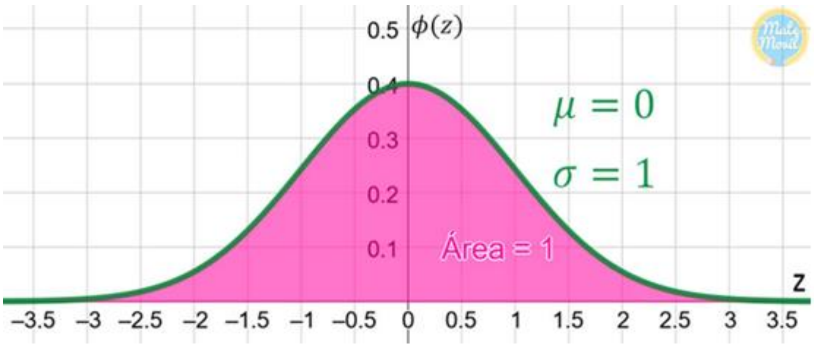


Figura 6. Distribución normal estandarizada. Tomado de https://matemovil.com/distribucion-normal-ejercicios-resueltos/

Como vemos arriba, y como lo habíamos señalado antes, bajo la curva de la normal se encuentra el área que engloba a la totalidad de los casos que toma la variable y que tiene valor 1 o, expresado en porcentaje, el 100%. Ahora, lo que nos interesa es calcular la probabilidad de que en la variable estandarizada Z, encontremos valores comprendidos entre 0 y +1,50. Entonces, debemos saber cuál es el área bajo la curva entre (Z = 0) y (Z = +1,50). Para hacerlo buscamos en la la tabla Z, que se encuentra en el anexo de este documento, y encontraremos que el valor de 1,50 nos da un área de **0,9332. Esto quiere decir que el área que va desde el - ∞ (menos infinito) hasta 1,50 representa el 93,32% de la totalidad de los casos. Como nos interesa saber el área entre Z = 0 y Z = 1,50, restaremos la probabilidad del área que está a la izquierda de Z= 0. Sabemos que es 0,50, por lo que tenemos: 0,9332 – 0,50= 0,4332.**

Ahora sí, estamos en condiciones de decir que tenemos el 43,32% de posibilidad de encontrar valores comprendidos entre 0 y 1,50 (ver figura 7).

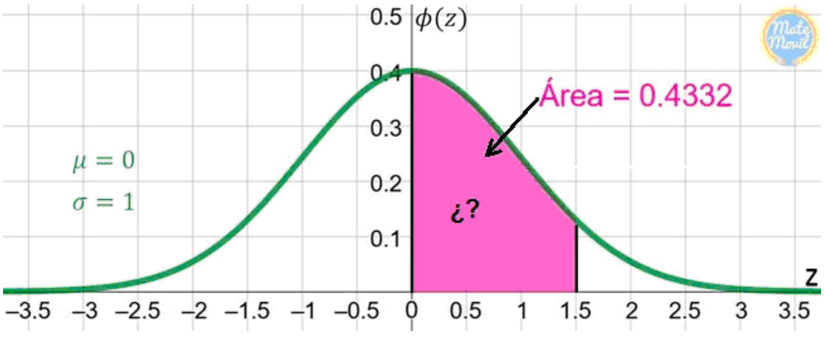


Figura 7. Determinación de área y probabilidad en distribución normal estandarizada. Tomado de https://matemovil.com/distribucion-normal-ejercicios-resueltos/

Bien, ¿cómo calculamos la probabilidad de encontrar un dato mayor a 1,5?

Para ello recordemos que a la derecha de la media encontramos el 50% de probabilidad (p = 0,50), por lo que a ese valor le tenemos que restar **0,4332,** lo que nos da un valor de 0,0668, que sería el valor del área a la derecha del valor 1,5

Entonces, podemos decir que la probabilidad de encontrar un valor por arriba de 1,5 es de 6,68% (ver figura 8).

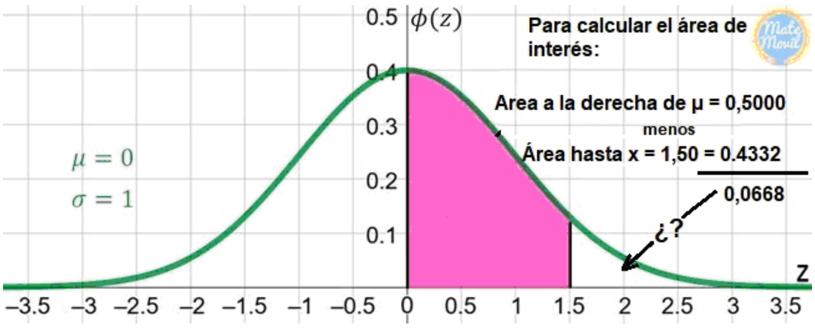


Figura 8. Determinación de área y probabilidad en distribución normal estandarizada. Tomado de https://matemovil.com/distribucion-normal-ejercicios-resueltos/

Recuerda dos cosas importantes:

- ***La tabla que usamos nosotros siempre te dará valores de área que van desde el*** $-\infty$ ***hasta el valor que escojas***.
- ***Si deseas calcular la probabilidad de un área a la derecha de un valor en particular puedes hacerlo restando el área total.***

Ahora bien, pasemos a cuestiones prácticas…

Por ejemplo, si deseamos comparar las puntuaciones de un sujeto en dos variables o, de dos sujetos en una variable, necesitamos transformar los valores obtenidos en una escala que permita dicha comparación. No podemos hacerlo si nos aferramos a la distribución original porque pueden ser de distinta naturaleza.

Para entenderlo más fácilmente, supongamos que tenemos dos muestras: la primera que es de ciudadanos de Kenia y la segunda que está compuesta por personas de Suiza; y deseamos comparar la posición económica que tiene una persona de cada población. Como podrás imaginarte, los ingresos de uno y otro ciudadano no son los mismos porque las condiciones económicas de ambos países son bien distintas, PERO COMO LO QUE DESEAMOS ES COMPARAR LA POSICIÓN ECONÓMICA QUE TIENE CADA UNO DE ELLOS EN SU PAÍS, necesitamos ver qué lugar ocupa cada uno de los sujetos en sus respectivas distribuciones para así poder compararlos. Para ello necesitamos transformar los valores que han obtenido en sus distribuciones, en una distribución teórica estandarizada. A este proceso se le denomina Tipificación de variables.

Tipificación de variables

¿Cómo hacemos?

Bueno, antes que nada, recordemos que tipificar es encontrar el valor que le corresponde en una distribución teórica, al valor que tengo en mi muestra. Por eso decimos que tipificar es transformar el valor real (el de mi estudio) en un valor teórico (un puntaje z). Este procedimiento permite medir la desviación de mis datos respecto a la media teórica, lo cual facilita la comparación relativa de los datos que provengan de distintas muestras.

Tipificar una variable X = N (μ, σ) consiste en transformarla en otra variable Z = N (0, 1). Esta afirmación quiere decir que tipificar un valor de una variable, con determinada media y desviación estándar, se logra transformando ese valor en un puntaje Z, en base a una distribución teórica que tiene como media el valor 0 y una desviación estándar de 1.

El procedimiento es simple; tomamos el valor que nos interesa, le restamos la media y luego a ese resultado lo dividimos por la desviación típica.

Para ello se utiliza la expresión: $Z = \frac{X - \mu}{\sigma}$

Luego, el puntaje de Z obtenido podemos buscarlo en la tabla, y obtendremos así la probabilidad de encontrar un valor menor al tipificado (comprendido desde –infinito hasta el valor en cuestión).

Para calcular la probabilidad de la variable X es siempre recomendable realizar la gráfica que nos determinará el área de interés, luego transformaremos los valores usando la tabla Z que se encuentra en el anexo de este documento.

Ahora veamos; supongamos que estamos evaluando Asertividad (es una variable aleatoria continua con una distribución normal no estandarizada). En nuestra investigación obtenemos una media de 10 y una desviación estándar 2, pero nos interesa calcular la probabilidad de que encontremos personas que tengan valores de asertividad comprendidos entre 10 y 11,50.

Muy bien, lo que tenemos que hacer es estandarizar los valores de la variable X aplicando la fórmula de z:

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{11{,}50 - 10}{2} = \frac{1{,}50}{2} = 0{,}75$$

Ahora, teniendo el valor **z=0,75** buscaremos en la tabla de distribución normal Z el área de probabilidad que le corresponde. Encontramos que es de 0,7734, pero como este valor es lo que le corresponde al área que está desde el - ∞ (menos infinito) hasta el puntaje z, lo que haremos será restarle 0,50 (que es el área que está a la izquierda de la media). Entonces, 0,7734 - 0,50 = 0,2734

Abajo verás las gráficas comparativas ;)

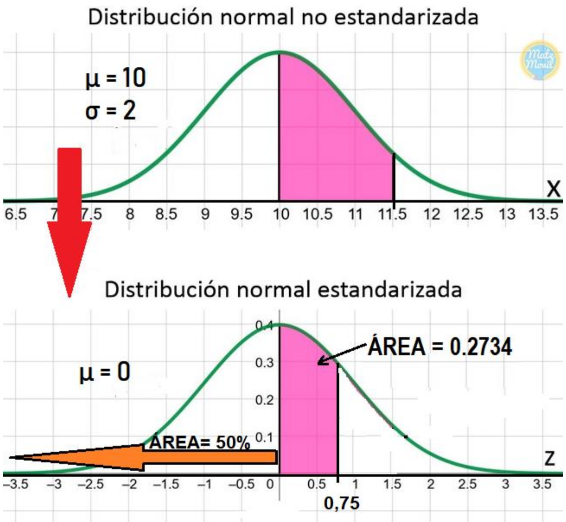


Figura 9. Determinación de probabilidad para la variable Asertividad. Proceso de tipificación

Como conclusión diríamos entonces que: como se puede observar, luego de la tipificación, podemos afirmar que hay un 27,34% de probabilidad de encontrar en nuestra muestra personas con niveles de Asertividad comprendidos entre 10 y 11,50.

La distribución tipificada se aplica en estadística inferencial para determinar intervalos de confianza para la media de una población, usualmente se utiliza un nivel de confianza del 95% para el cual z = 1,96 (esto lo veremos en las próximas unidades).

Distribuciones t de Student

William S. Gosset (1876-1937) fue quien propuso las denominadas distribuciones t de Student. Pero ¿por qué no llevan su nombre? Sucede que Gosset en 1908, trabajando para una compañía de cervezas en Dublín (Irlanda), se focalizó en encontrar una distribución que se adaptara a muestras pequeñas y con varianza desconocida; pero la empresa para la cual trabajaba había prohibido a sus empleados que publicaran información confidencial, por lo que Gosset firmó sus publicaciones usando el nombre de "Student".

Las distribuciones t de Student son parecidas a la distribución normal y se utilizan para hacer estimaciones de la media cuando se desconoce la varianza (es lo habitual) y cuando se usan muestras con un N (tamaño) menor a 30 unidades.

Una diferencia fundamental que tienen con la distribución normal es que *hay una distribución t diferente para cada tamaño de la muestra*, por eso se las denomina en plural. Estas distribuciones son una familia de distribuciones de probabilidad continuas. Las curvas de densidad son simétricas y con forma de campana (como la distribución normal estándar). Sus valores de media son 0 y sus varianzas son mayores que 1.

La distribución t difiere de la de Z (distribución normal) en que la varianza de t depende del tamaño de la muestra y siempre es mayor a uno. Si el tamaño de la muestra es n entonces decimos que la distribución t tiene **n-1** grados de libertad. Únicamente cuando el tamaño de la muestra tiende a infinito las dos distribuciones tienden a ser iguales.

Las colas de las distribuciones t disminuyen más lentamente que las colas de la distribución normal. Si los grados de libertad son mayores (es decir, si se trabaja con una muestra muy grande), más próxima a 1 es la varianza y la función de densidad es más parecida a la densidad normal (ver figura 10).

La función de densidad de la Distribución t de Student es:

$$t_n(x) = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

Pero ya sabes, afortunadamente estas integrales ya han sido calculadas y sólo tenemos que usar la *Tabla de valores críticos de la distribución t de Student* que encontrarás en el anexo de este documento para calcular la densidad y/o probabilidad.

Los Grados de libertad hacen referencia a la cantidad de datos que va a usar el investigador y que en el caso de la t de Student se calcula de la siguiente forma: ***Grados de libertad = la cantidad de datos de la muestra – 1***

gl = n – 1

Supongamos que nuestra muestra tiene 23 observaciones o individuos, entonces

gl = 23 – 1

gl = 22

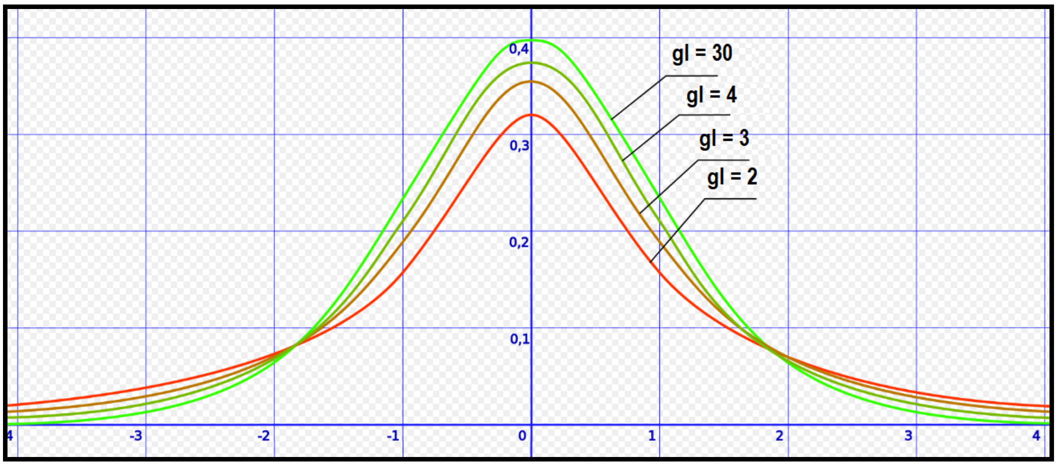


Figura 10. Distribuciones t para varios valores de grado de libertad (gl = 2 – 30).

Cálculo de probabilidad

Como la distribución t es simétrica alrededor de una media de cero, tenemos una situación parecida a la normal estandarizada. Es decir, el valor t deja un área a la derecha igual a la izquierda, y ambas suman 1.

Para encontrar los valores de t se utilizará la tabla de valores críticos de la distribución t que figura en el anexo.

Veamos un ejemplo:

Necesito determinar cuál es el valor t que, con 10 grados de libertad, deja un área de 0,05 a la derecha, y por tanto un área de 0,95 a la izquierda…. Veamos….

Lo que haremos será ir a la tabla y buscaremos el valor crítico para 0,05, con 10 grados de libertad, que es 1,812. Ese valor es el que debemos representar en la gráfica.

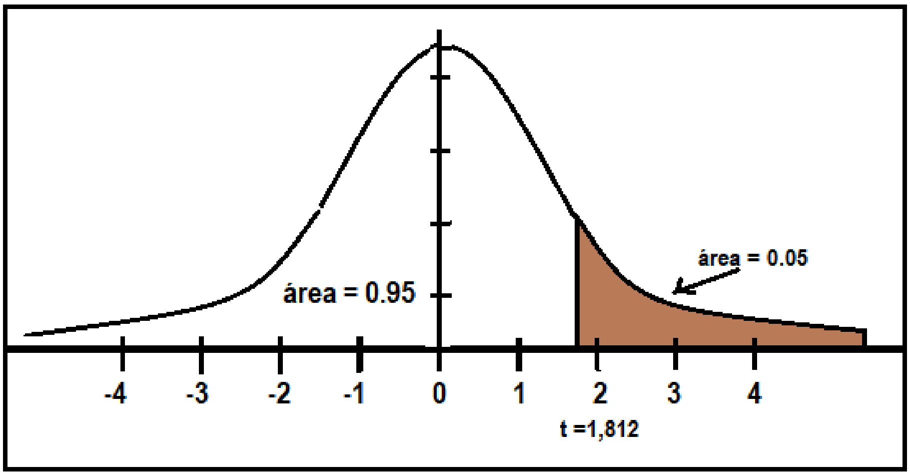


Figura 11. Determinación de probabilidad distribución t

Veamos unos ejemplos más….

Si necesitas encontrar el valor crítico de t con una probabilidad de 0,025, en una distribución de 12 grados de libertad, debes buscar en la tabla, en la intersección, el valor correspondiente, que es 2,179.

Si deseas determinar la probabilidad de que un valor que sea mayor a 1,833 en una distribución t con 9 grados de libertad; debes buscar en la tabla el valor más próximo de t en los grados de libertad señalados. Para el ejemplo que hemos propuesto es de 0,050, por lo que diríamos que en la distribución con 9 grados de libertad hay una probabilidad de 0,05 de encontrar un valor t mayor a 1,833

Distribuciones Chi-cuadrado

La distribución ji-cuadrado, también denominada Chi-cuadrado de Pearson, Chi^2 , o χ2, es una distribución de probabilidad continua cuyo parámetro es k (representa los grados de libertad de la variable aleatoria).

La función de densidad Chi-cuadrado es

$$f_k(x) = \frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

En estadística tiene mucha relevancia por su papel en el test Chi-cuadrado, en la estimación de varianzas, en la estimación de la media de una población normalmente distribuida y en el problema de estimar la pendiente de una recta de regresión lineal.

Algunas de sus características:

- La distribución es asimétrica positiva.
- A medida que aumenta el tamaño de la muestra, la curva es menos asimétrica, aproximándose a una curva normal.
- Para cada tamaño muestral, se tendrá una distribución χ2 diferente.
- El parámetro que caracteriza a una distribución χ2 son sus grados de libertad (n-1), originado una distribución para cada grado de libertad.

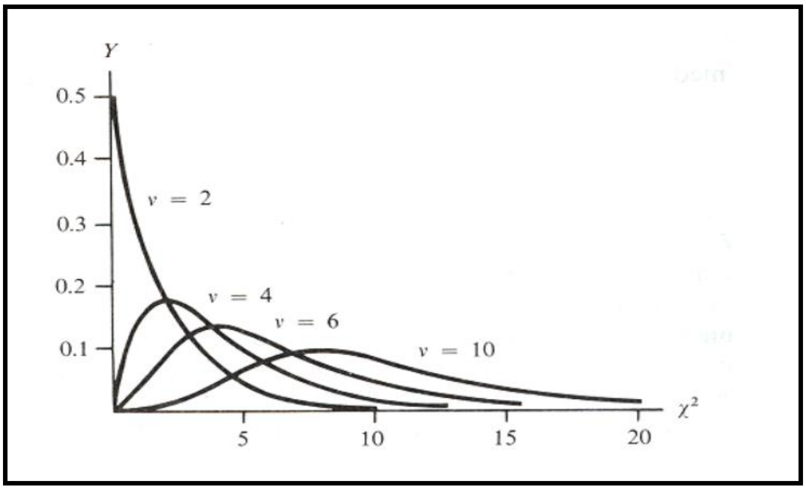


Figura 12. Distribuciones chi cuadrado para varios valores de grado de libertad (v = 2 – 10).

¿Cómo se calcula su media y su varianza?

El parámetro de la distribución X² es n y su media y su varianza son: $\mu = \nu$ y $\sigma^2 = 2\nu$

¿Cómo se calcula su función de densidad?

La Distribución chi-cuadrada, tiene por función de densidad

$$\chi_k^2(x) = \frac{x^{k/2-1}\, e^{-x/2}}{2^{k/2}\, \Gamma(k/2)}$$

Donde el parámetro k de X²k, se denomina grados de libertad de la distribución.

La Distribución chi-cuadrado no tiene sentido para valores negativos de x, como se puede ver en la figura 12.

Si deseas ampliar con más información puedes visitar:

https://es.wikibooks.org/wiki/Tablas_estad%C3%ADsticas/Distribuci%C3%B3n_chi-cuadrado

TEOREMA DEL LÍMITE CENTRAL

El Teorema del Límite Central hace referencia a una idea surgida de considerar un conjunto de resultados acerca de la distribución promedio de muchas variables aleatorias. Este teorema nos dice que *la suma de un número muy grande de variables aleatorias, con varianza finita, se aproxima a una distribución normal* y además remarca que, aunque las variables sigan distintos modelos de distribución (no importa si son exponenciales, binomiales, etc.), el promedio de ellas se distribuye con forma de una **normal.**

Explicado de otro modo sería así: **si para una variable determinada, tomamos muchas muestras de tamaño n de una población y le calculamos a cada una de ellas el valor promedio, la distribución de los valores de los promedios tendrá forma de una curva normal, independientemente de la distribución original que tengan.**

Por ejemplo: *El tiempo de realización de un proyecto complejo (como construir una casa, un submarino, un avión, una red de carreteras, un oleoducto…) es la suma de los tiempos de las distintas tareas que componen el proyecto. A pesar de que habrá tareas que tendrán un tiempo fijo, la mayoría serán variables con diferente tiempo medio y diferente variación. Pero la suma de los tiempos seguirá una distribución normal, y se podrán calcular probabilidades de finalización en un tiempo determinado (y a su vez el coste de este tiempo)*

Ejemplo 2: Supongamos que entrevistamos a 50 personas y que de esa investigación sacamos la conclusión que, con un 0,57 de probabilidad, encontraremos a uno de ellos que piensa que el hombre no llegó a la luna.

Ahora bien, si yo utilizo esa investigación para generalizarla al resto de la población cometería un error grave, la muestra es demasiado pequeña (solo 50 personas). Por lo que la probabilidad (0,57) puede haber sido por efecto del azar o por una característica puntual de la muestra. ¿Entonces qué hago?... bueno, como me resultaría imposible entrevistar a toda la población para determinar la verdadera probabilidad, tomaré muchas muestras pequeñas (unas 1000) y a cada una de ella le calcularé la probabilidad. Luego haré un histograma con los resultados obtenidos y descubriré que las probabilidades encontradas siguen una curva normal, por lo que tomaré como referencia la media de esos datos.

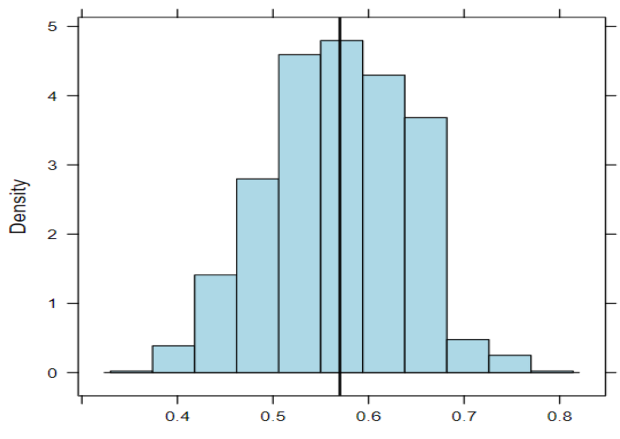


Figura 13. Histograma de probabilidad

Vemos, por lo tanto, que la distribución de los valores de probabilidad obtenidos, toman forma (aproximadamente) de una curva normal.

Es, precisamente, en este tipo de evidencia que los científicos se apoyan para demostrar que el uso de muestras puede ser adecuado en los análisis de la estadística inferencial: "Mientras que la muestra sea representativa de la población, independientemente de que los datos de la variable en estudio se ajusten a la perfección a la curva normal, se pueden usar las variables aleatorias para realizar los cálculos de probabilidad"

Ahora intentaremos explicar este teorema con el lanzamiento de dados. Sabemos que el lanzamiento de dados tiene una distribución uniforme. ¿Por qué? Porque cada uno de los números del dado tiene la misma probabilidad de ocurrencia en un lanzamiento ¿cierto?... Por lo que la distribución se asemejaría a esta:

Número	1	2	3	4	5	6
Probabilidad	1/6	1/6	1/6	1/6	1/6	1/6

Muy bien, pero ¿qué pasa cuando tiramos los dados unas 2000 veces?

Lo que vemos es que la probabilidad parece mantenerse constante (distribución uniforme), más allá de alguna pequeña diferencia.

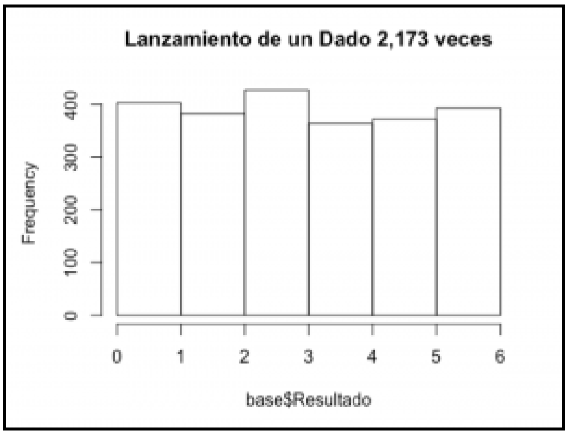


Figura 14. Histograma

Bueno, pero ahora vamos a implementar una estrategia distinta. Lanzaremos el dado 10 veces y luego calcularemos el promedio de esos 10 lanzamientos.

Lanzamiento	Resultado
1	1
2	3
3	6
4	1
5	6
6	4
7	5
8	5
9	4
10	5
Promedio	4
Suma	40

Ahora, repetiremos unas 1500 veces este experimento: lanzaremos el dado 10 veces y calcularemos el promedio de esos 10 lanzamientos

Lanzamiento	Dado 1	Dado 2	Dado 3	Dado 4	Dado...	Dado 1,500
1	1	4	5	4	...	1
2	3	6	2	2	...	5
3	6	5	2	4	...	4
4	1	6	6	6	...	5
5	6	6	5	5	...	2
6	4	5	5	6	...	6
7	5	4	4	2	...	2
8	5	4	4	2	...	6
9	4	1	3	3	...	1
10	5	3	5	4	...	2
Promedio	4	4.4	4.1	3.8	...	3.4
Suma	40	44	41	38	...	34

El histograma que mostraría la distribución de todos los promedios de los experimentos tendría esta forma:

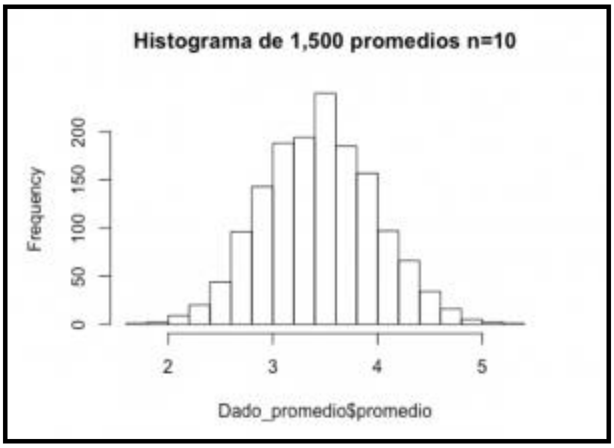


Figura 15. Histograma

Lo mismo sucede si, en lugar de tomar los promedios de cada muestra, tomamos las sumas, tal como lo muestra el gráfico de abajo.

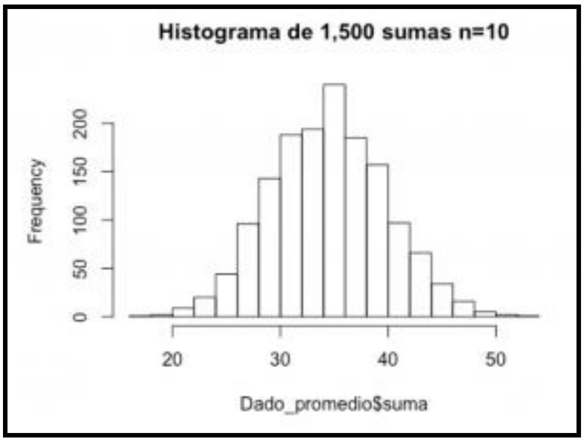


Figura 16. Histograma

Bueno, parece que nos ha quedado claro que la distribución de los promedios de muestras de tamaño n se van a comportar indefectiblemente como una distribución normal; pero ¿qué otras conclusiones podemos esperar?

Hay dos cuestiones más y que son fundamentales; para la inferencia estadística y el control estadístico de proceso. A partir del teorema del límite central podemos saber qué comportamiento tendrá la media y la varianza de los promedios. Lo explicaremos a continuación:

El error estándar de la media

La desviación estándar (σ) representa la variación en los valores de una variable, mientras que el Error estándar de la media (EEM) representa la dispersión que tendría la media de una muestra de promedios. Por lo tanto, el EEM proporciona una idea de la precisión de la media y la (σ) nos da una idea de la variabilidad de las observaciones individuales. Estos dos parámetros están relacionados:

$$\mathbf{EEM} = \frac{\sigma}{\sqrt{n}}$$

Donde:

EEM = Error estándar de la media

σ = Desviación estándar

n = tamaño de la muestra

Es decir, que el error estándar de la media se obtiene dividiendo la desviación estándar de la población entre la raíz cuadrada del tamaño de la muestra.

Los estadísticos sabemos que siempre el promedio muestral difiere en alguna medida del promedio real, es justamente ésta desviación lo que se denomina error estándar de la media. Entre más pequeño sea el error estándar de la muestra, mejor representará la muestra a la población total.

¿Cuál es la utilidad del Teorema del Límite Central?

Partiendo de la premisa de que gracias a este teorema conocemos la forma de la distribución de las variables para cualquier población, podemos comparar cualquier promedio individual con la distribución muestral para saber si proviene de la misma población. Esto tiene mucho que ver con un tema que verás más adelante, que es la prueba de hipótesis. Las pruebas de hipótesis de medias, de proporciones, de relaciones, están inspiradas en el Teorema del límite central; con ellas, por ejemplo, podemos determinar, con una probabilidad de error suficientemente bajo, si el promedio en la muestra que estoy trabajando es mayor o menor respecto a la población. Es importante que recuerdes que la mayoría de las pruebas estadísticas más comunes se apoyan en el Teorema del límite central para poder justificar los resultados y determinar el nivel de error que estamos cometiendo cada vez que afirmamos algo.

Tipos de estimadores: Propiedades

Adaptado de http://www.fca.unl.edu.ar/InferEst/EstimParam.htm

El deseo último de todo investigador es lograr describir características de la población, pero, como sabemos, este proceso no es fácil porque es muy costoso en términos de tiempo y esfuerzo, o bien porque la población es inaccesible. Es en este sentido que la estadística juega un rol preponderante, ya que mediante ella podemos inferir uno o más parámetros de la población utilizando la información que disponemos de una muestra.

Nos podemos imaginar que, lógicamente, el estadístico de la muestra puede llegar a ser diferente del parámetro de la población (que sea igual sería pura coincidencia). Las diferencias pueden obedecer a distintos motivos, uno de ellos tiene que ver con el tamaño del muestreo (claramente, una muestra, por más grande que ésta sea, siempre va a tener menos elementos que la población, pero a medida que esta sea más grande podrá representar mejor a la población). A esta diferencia (entre el valor del estadístico muestral y el correspondiente parámetro de la población) se le llama *error de estimación*.

Generalmente el error de estimación es desconocido, porque la única forma de tener una idea clara al respecto es evaluar a toda la población, lo cual, como hemos dicho, es prácticamente imposible en la en la mayoría de los casos. Ahora bien, esto no significa que el investigador no pueda continuar, ya que las inferencias estadísticas se hacen por medio de probabilidades.

Por ejemplo, tomando la media de una muestra podemos inferir la media de la población (la media de la población en la mayoría de los casos es desconocida); solo que en lugar de un valor exacto podemos pensar en la probabilidad de encontrar la media de la población entre dos valores determinados. Es importante que recuerdes que cuando inferimos no tenemos forma de garantizar que la conclusión a la que arribamos sea estrictamente (100%) correcta, pero si podemos mediante la estadística calcular el margen de error asociado a dicha estimación.

En la inferencia estadística hay dos temas importantes: los problemas de estimación y los de prueba de hipótesis (a este último lo veremos en la próxima unidad).

El procedimiento mediante el cual se llega a la obtención y se analizan los estimadores se llama *Estimación estadística*, que a su vez se divide en estimación puntual y estimación por intervalos.

Los métodos de inferencia estadística emplean el razonamiento inductivo, razonamiento de lo particular a lo general y de lo observado a lo no observado. Así, el estimador brinda información respecto al parámetro. Por ejemplo, la media de la muestra, $\bar{X}$, es un estimador de la media μ en la población.

Estimación puntual e intervalar de parámetros

Como hemos dicho la inferencia estadística busca:

- Estimar el valor de un parámetro desconocido (estimación puntual o estimación intervalar).
- Verificar si un estimador es o no igual a cierto parámetro (**prueba de hipótesis).**

Ahora nos concentraremos en el primer punto y dejaremos para la próxima unidad el siguiente.

Estimación puntual

Este tipo de estrategia busca estimar un parámetro puntual, es decir, un solo punto, un solo valor.

La estimación puntual busca, mediante una muestra, obtener el valor que mejor represente al parámetro que nos interese de la población. Por ejemplo, tengo una muestra de adultos mayores que tienen una media ($\bar{x}$) de 36,08 en Depresión. Si hago una estimación puntual, diré que la media de la población de adultos mayores tiene una media (μ) de 36,08.

La preocupación del investigador debe ser encontrar el mejor estimador, un estimador eficiente. Se entiende que un estimador es eficiente en la medida que sea insesgado (ausencia de sesgos) y estable en el muestreo (varianza mínima)

Existen dos métodos para obtener la estimación puntual de un parámetro: método de los momentos y método de máxima verosimilitud. Método de los momentos: consiste en igualar momentos poblacionales a momentos muestrales, en tanto que el Método de máxima verosimilitud consiste en tomar como valor del parámetro aquel que maximice la probabilidad de que ocurra la muestra observada. Por la complejidad de estas estimaciones y por motivos prácticos no desarrollaremos este punto y nos focalizaremos en la estimación por intervalos.

Estimación por intervalos o estimación intervalar

Un estimador por intervalo se construye sobre el concepto del estimador puntual, pero proporciona además una información adicional que tiene que ver con el grado de exactitud del estimador (valor de probabilidad asociado); por otro lado, difiere de la estimación puntual porque el estimador por intervalo es un rango, o banda, dentro de la cual se supone va a estar el parámetro.

Revisaremos a continuación los conceptos necesarios para comprender el procedimiento.

Intervalo de confianza

El intervalo de confianza contiene al parámetro estimado con un determinado nivel de confianza. Hace referencia a un par o varios pares de valores (intervalo) entre los cuales se espera que esté cierto valor desconocido con una determinada probabilidad de acierto (nivel de confianza). La probabilidad de éxito en la estimación se representa con $1 - \alpha$ y se denomina *nivel de confianza*, y nos brinda una idea de cuáles son las posibilidades de acertar en la estimación. Existe una asociación entre el nivel de confianza y la amplitud del intervalo: un intervalo más amplio cubrirá mayores probabilidades de acierto y por lo tanto será menos preciso, mientras que un intervalo más pequeño, que ofrece una estimación más precisa, aumenta su probabilidad de error.

Es un requisito fundamental conocer la distribución del parámetro a estimar para poder construir un intervalo de confianza, aunque habitualmente el parámetro suele presentar una distribución normal.

Error de la estimación

Como hemos visto arriba, es un valor que está ligado con la amplitud del intervalo de confianza. Un intervalo de confianza más estrecho nos dará mayor precisión en la estimación del parámetro; pero, si se quiere disminuir aún más el error de la estimación se debe aumentar la cantidad de observaciones (aumentar el tamaño de la muestra). Se suele llamar *E*, según la fórmula $E = (\theta_2 - \theta_1)/2$.

Nivel de confianza

Hace referencia a la probabilidad de que se encuentre el verdadero valor del parámetro estimado en el intervalo de confianza obtenido. El nivel de confianza suele expresarse con un porcentaje; en ciencias humanas es habitual tomar como nivel de confianza un 95% o un 99% (que según veras inmediatamente abajo, se corresponden con *valores* ***α*** de 0,05 y 0,01 respectivamente).

Valor α (alfa) o Nivel de significación

El Valor alfa (*α*) o nivel de significación es el complemento del nivel de confianza. Verás, es la probabilidad que existe de fallar en nuestra estimación. Es la diferencia entre la certeza (1) y el nivel de confianza (1-α) dividido por 100. Por ejemplo, en una estimación con un nivel de confianza del **95%**, el valor α (o Nivel de significación) es (100-95) /100 = **0,05**

Valor crítico

Se representa por $Z_{\alpha/2}$ y es el valor de Z necesario para construir un intervalo de confianza para la distribución.

Muy bien, ahora estamos en condiciones como para proceder a calcular un intervalo de confianza ….

***Estimación del intervalo de confianza para la media* (con varianza conocida)**

¿Cómo construimos un intervalo de confianza?, a medida que vayamos desarrollando el procedimiento, intentá incorporar los conceptos que hemos señalado arriba.

La primera idea que tenemos es que el intervalo es el rango en el cual decimos que se encuentra la media con un determinado nivel de confianza. En forma matemática sería: $\bar{X} - Z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z\frac{\sigma}{\sqrt{n}}$. Lo que nos dice esta anotación no es otra cosa que advertirnos que la media μ va a estar comprendida entre dos valores: $\underbrace{\bar{X} - Z\frac{\sigma}{\sqrt{n}}}_{\text{valor crítico 1}} \leq \mu \leq \underbrace{\bar{X} + Z\frac{\sigma}{\sqrt{n}}}_{\text{valor crítico 2}}$

Bien, ahora vamos a ver cómo se representan estos conceptos gráficamente:

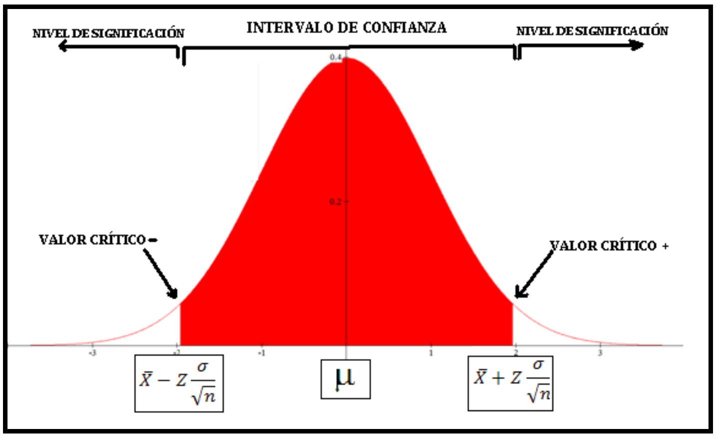


Figura 17. Esquema representativo de Intervalo de confianza

A partir de ahora estamos en condiciones de proceder a calcular los valores e interpretarlos.
Supongamos que tenemos una muestra de 36 personas a las que hemos evaluado en la variable Extroversión, y los resultados que hemos obtenido nos han dado una media de 24 y una desviación estándar de 3, ($\bar{X} = 24 ; \sigma = 3\ y\ n = 36$) y ahora deseamos construir, para la media poblacional (valor que no conocemos), un intervalo de confianza al 95%.

Como te habíamos comentado anteriormente, el nivel de significación es 100% menos el intervalo de confianza, por lo que en nuestro caso sería 100 – 95 = 5%, lo que expresado en decimales sería equivalente a 0,05. Pero si observas el gráfico de arriba, ese 0,05 se tiene que repartir en las dos colas de la distribución, por lo que cada una de las colas tendría un valor de: 0,05 / 2 = **0,025 (a esto se le llama significación bilateral).**

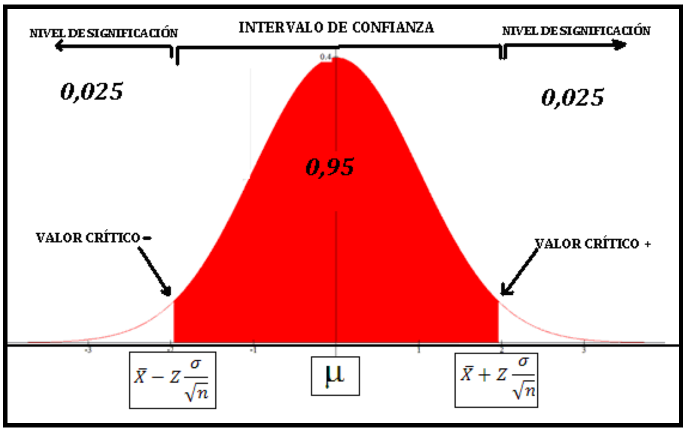


Figura 18. Determinación del Nivel de confianza / Nivel de significación

Muy bien, ahora, con el nivel de significación ya determinado, buscaremos cuál es el puntaje Z correspondiente para luego proceder a la tipificación de los valores críticos. Recordemos que al trabajar con un nivel de significación bilateral tendremos un valor de 0,025 en cada extremo. Muy bien, ahora sí…. Para ver los puntajes Z correspondientes nos dirigimos a la tabla de distribución normal Z (ver en el anexo de este documento) y buscaremos el valor de Z que tiene un área de probabilidad de 0,025…. Eso nos da: **Z = - 1,96,** y como tenemos un nivel de significación bilateral (en las dos colas de la distribución) diremos que por simetría el otro valor que corresponde es **Z = +1,96**
Ahora debemos calcular los valores críticos,

Recordamos la fórmula: $$\bar{X} - Z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z\frac{\sigma}{\sqrt{n}}$$

Remplazamos por los valores correspondientes y tendremos $24 - 1{,}96\frac{3}{\sqrt{36}} \leq \mu \leq 24 + 1{,}96\frac{3}{\sqrt{36}}$

Lo que nos da como resultado: $23{,}02 \leq \mu \leq 24{,}98$

Ahora graficamos:

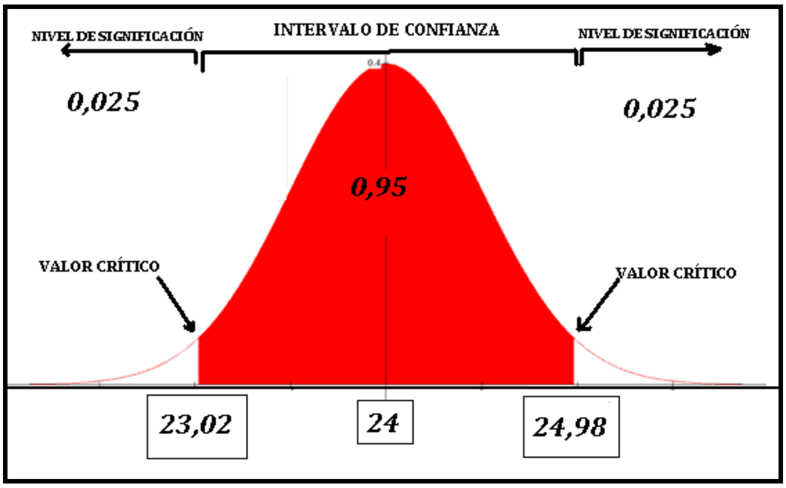


Figura 19. Determinación del intervalo de confianza

¿Cuáles son las conclusiones a las que podemos arribar con estos datos?

Estos resultados nos permiten concluir que, con una muy baja probabilidad de equivocarnos, menos de 5 veces cada 100 observaciones que hagamos con esta muestra (*Nivel de significación de 5%),* la media de Extroversión en la población estará comprendida entre los 23,02 y 24,98 puntos.

Muy bien, ahora mostraremos el procedimiento para realizar una estimación intervalar mediante el programa Jamovi.

Partamos de la idea que poseemos datos de una investigación en la cual hemos evaluado la variable Extroversión en una muestra representativa de 281 argentinos y, desconociendo la media poblacional para esta variable, deseamos realizar una estimación intervalar al 95% de confianza (es decir, con un 5% de error).

El procedimiento sería: Análisis>Pruebas T> Pruebas T en una muestra

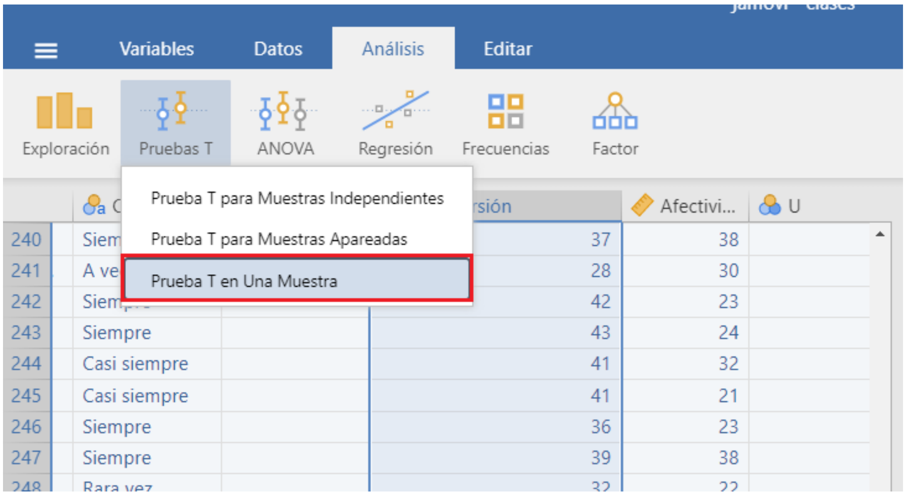


Figura 20: *Captura de pantalla del procedimiento de estimación intervalar realizado mediante el programa estadístico Jamovi*

Como se observa en la figura 20, el procedimiento a aplicar recibe el nombre de Prueba t en una muestra. A continuación, en la figura 21 mostramos las opciones seleccionadas.

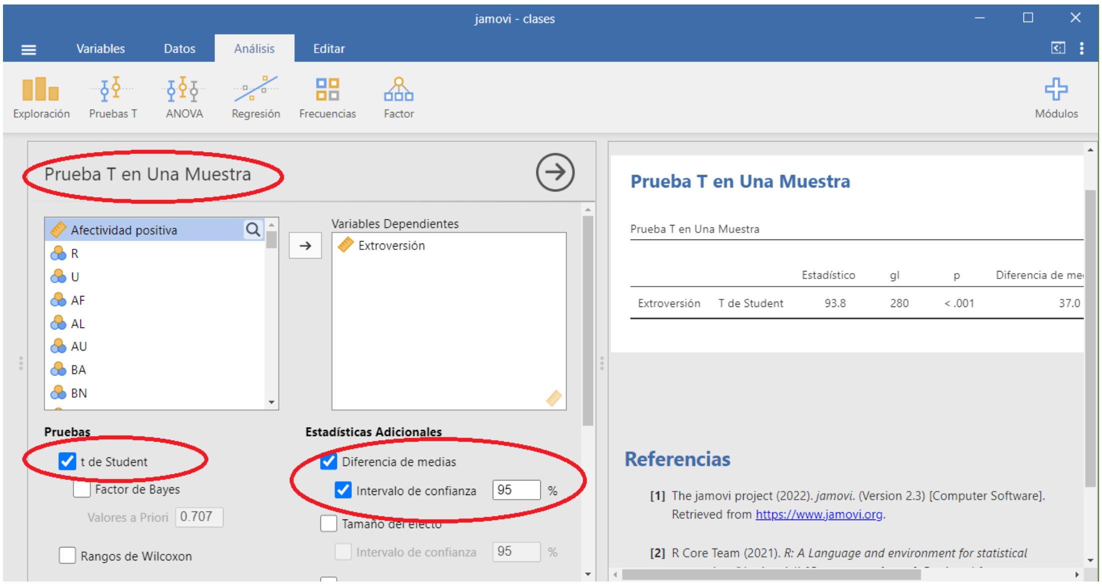


Figura 21: *Opciones necesarias para efectuar la estimación intervalar mediante la Prueba t en una muestra.*

Tabla 1.

Estimación intervalar para la variable Extraversión

Prueba T en Una Muestra

						Intervalo de Confianza al 95%	
		Estadístico	gl	p	Diferencia de medias	Inferior	Superior
Extroversión	T de Student	93.8	280	< .001	37.0	36.2	37.8

En tabla 1 se puede ver la información que ofrece el análisis. Se observa que el intervalo de confianza determinado al 95%, está comprendido entre los valores 36,2 y 37,8; por lo que podríamos decir que con un error del 5% la media de la población estaría comprendida entre dichos valores.

BIBLIOGRAFÍA

Bergagna, A. D. (2019). *Inferencia Estadística Básica para Ingenieros Agrónomos. Estimación de parámetros.* Recuperado el 02/04/19 de http://www.fca.unl.edu.ar/InferEst/EstimParam.htm

Caballero-Romero, A. E. (2009). *Metodología de la Investigación Científica, Diseños con Hipótesis Explicativas*. Lima – Perú: Editorial UDEGRAF.

Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2010). *Metodología de la Investigación* (5ta edición). México D.F.: McGraw Hill.

Hospital Universitario Ramon y Cajal (2019). *Material docente de la Unidad de Bioestadística Clínica Estimación de parámetros: Distribución muestral de medias.* Recuperado el 02/04/19 de http://www.hrc.es/bioest/esti_medias.html

Quintela del Rio, A. (2019). *Estadística Básica Edulcorada. El teorema central del límite*. Recuperado el 02/04/19 de https://bookdown.org/aquintela/EBE/el-teorema-central-del-limite.html

The jamovi project (2022). *Jamovi (Version 2.3.13) [Computer Software].* https://www.jamovi.org.

Wikilibros (2019). *Tablas estadísticas.* Recup. de https://es.wikibooks.org/wiki/Estad%C3%ADstica/Todas_las_tablas el 02/04/19

Calculadoras online

Calculadoras del Intervalo de Confianza:

http://www.learningaboutelectronics.com/Articulos/Calculadora-de-intervalo-de-confianza.php

https://www.easycalculation.com/es/statistics/confidence-limits-mean.php

Simulación de distribuciones

Estimación de parámetros: https://proyectodescartes.org/uudd/materiales_didacticos/inferencia_estadistica_JS/estimac.htm

Tabla de distribución normal z Tomado de Jose Joaquin Muñoz Gonzalez. https://www.slideshare.net/josejoaquinmunoz/tablas-de-distribucion-normal-con-la-probabilidad-definitiva-con-todos-los-valores-de-z?from_action=save

El valor de la tabla para z es el área bajo la curva de la normal estándar a la izquierda de z

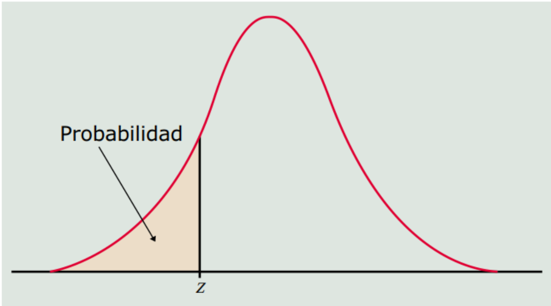


TABLA A: Probabilidades de la normal estándar

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

El valor de la tabla para z es el área bajo la curva de la normal estándar a la izquierda de z

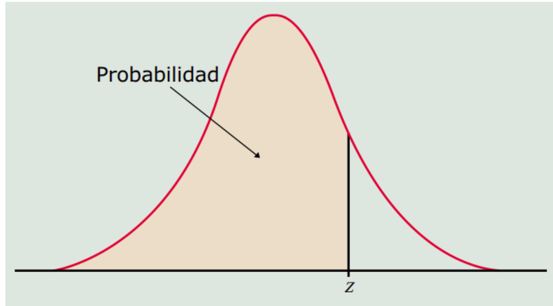
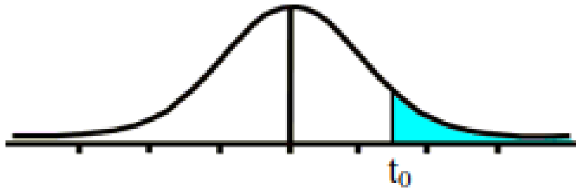


TABLA A: Probabilidades de la normal estándar (cont.)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Tabla t-Student



Valores Críticos de la Distribución t de Student

G. L.	a						
	0.2	**0.1**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**
1	1.376	3.078	6.314	12.706	31.821	63.657	318.294
2	1.061	1.886	2.920	4.303	6.965	9.925	22.327
3	0.978	1.638	2.353	3.182	4.541	5.841	10.214
4	0.941	1.533	2.132	2.776	3.747	4.604	7.173
5	0.920	1.476	2.015	2.571	3.365	4.032	5.893
6	0.906	1.440	1.943	2.447	3.143	3.707	5.208
7	0.896	1.415	1.895	2.365	2.998	3.499	4.785
8	0.889	1.397	1.860	2.306	2.896	3.355	4.501
9	0.883	1.383	1.833	2.262	2.821	3.250	4.297
10	0.879	1.372	1.812	2.228	2.764	3.169	4.144
11	0.876	1.363	1.796	2.201	2.718	3.106	4.025
12	0.873	1.356	1.782	2.179	2.681	3.055	3.930
13	0.870	1.350	1.771	2.160	2.650	3.012	3.852
14	0.868	1.345	1.761	2.145	2.624	2.977	3.787
15	0.866	1.341	1.753	2.131	2.602	2.947	3.733
16	0.865	1.337	1.746	2.120	2.583	2.921	3.686
17	0.863	1.333	1.740	2.110	2.567	2.898	3.646
18	0.862	1.330	1.734	2.101	2.552	2.878	3.610
19	0.861	1.328	1.729	2.093	2.539	2.861	3.579
20	0.860	1.325	1.725	2.086	2.528	2.845	3.552
21	0.859	1.323	1.721	2.080	2.518	2.831	3.527
22	0.858	1.321	1.717	2.074	2.508	2.819	3.505
23	0.858	1.319	1.714	2.069	2.500	2.807	3.485
24	0.857	1.318	1.711	2.064	2.492	2.797	3.467
25	0.856	1.316	1.708	2.060	2.485	2.787	3.450
26	0.856	1.315	1.706	2.056	2.479	2.779	3.435
27	0.855	1.314	1.703	2.052	2.473	2.771	3.421
28	0.855	1.313	1.701	2.048	2.467	2.763	3.408
29	0.854	1.311	1.699	2.045	2.462	2.756	3.396
30	0.854	1.310	1.697	2.042	2.457	2.750	3.385
35	0.852	1.306	1.690	2.030	2.438	2.724	3.340
40	0.851	1.303	1.684	2.021	2.423	2.704	3.307
45	0.850	1.301	1.679	2.014	2.412	2.690	3.281
50	0.849	1.299	1.676	2.009	2.403	2.678	3.261
60	0.848	1.296	1.671	2.000	2.390	2.660	3.232
70	0.847	1.294	1.667	1.994	2.381	2.648	3.211
80	0.846	1.292	1.664	1.990	2.374	2.639	3.195
90	0.846	1.291	1.662	1.987	2.368	2.632	3.183
100	0.845	1.290	1.660	1.984	2.364	2.626	3.174
200	0.843	1.286	1.652	1.972	2.345	2.601	3.131
500	0.842	1.283	1.648	1.965	2.334	2.586	3.107
800	0.842	1.283	1.647	1.963	2.331	2.582	3.100
1000	0.842	1.282	1.646	1.962	2.330	2.581	3.098
Inf.	0.842	1.282	1.645	1.960	2.326	2.576	3.090