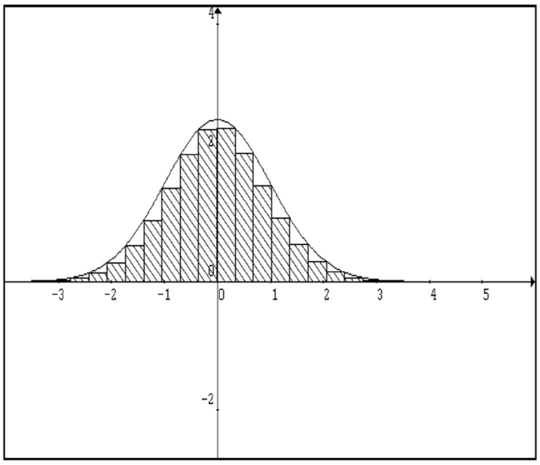
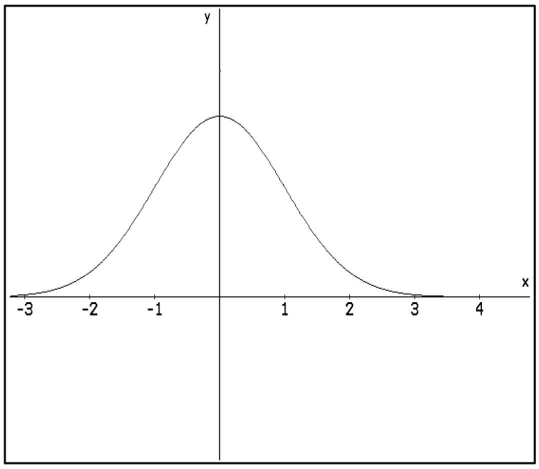
Tema 9: Inferencia Estadística.

9.1 Campana de Gauss.

- Carl Friedrich Gauss (matemático alemán 1777 – 1855) estudiando los errores que se producen al medir reiteradamente una magnitud, demostró que se distribuían según la función: $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$. Siendo μ la media y σ la desviación típica de la distribución de las medidas de la magnitud.

- La misma función resulta cuando en un histograma de frecuencias relativas de una variable continua aumenta el número de intervalos indefinidamente y su amplitud disminuye. En el siguiente gráfico se observa como el polígono de frecuencias relativas se acerca a la función f(x) anterior que se llama función de densidad de la distribución normal o campana de Gauss.



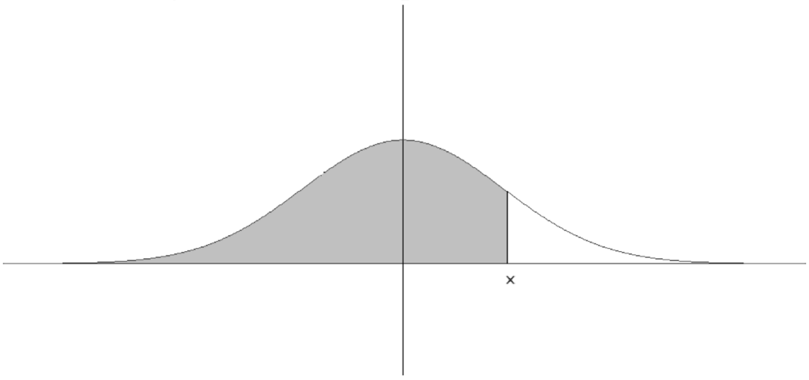


Una variable aleatoria es normal "si se rige según las leyes del azar". La mayoría de las distribuciones más importantes son normales: pocos individuos en los extremos y aumento paulatino hasta llegar a la parte central en la que están la mayoría de los datos. Ejemplo: la distribución de los pesos de los individuos de cualquier especie, la estatura de una población, la temperatura del mes de agosto a lo largo de 100 años, la longitud de los tornillos que salen de una fábrica, el consumo de cierto producto por familia, el cociente intelectual, el número de productos defectuosos etc. De hecho se ha comprobado que la mayoría de los fenómenos relacionados con psicología, pedagogía, biología, etc., siguen una distribución normal.
Sin embargo, no todas las distribuciones son normales, por ejemplo si clasificamos según el nivel de renta a los ciudadanos españoles, son muy pocos los que poseen niveles de rentas altas y en cambio son muchos los que poseen niveles de rentas bajas, por tanto la distribución no sería simétrica y en consecuencia no se adapta al modelo normal.

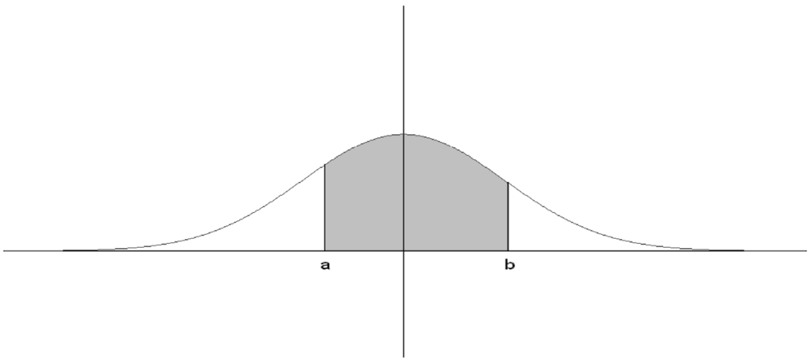
9.2 Función de densidad y función de distribución

- La función de densidad o campana de Gauss anterior $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ donde μ es la media y σ la desviación típica, cumple las siguientes propiedades:
  - $f(x)\geq 0$.
  - $\int_{-\infty}^{+\infty} f(x)dx=1$. El área encerrada bajo la curva de la función vale 1.
  - $\int_{a}^{b} f(x)dx=P(a\leq X\leq b)$. Es decir, la probabilidad coincide con el área encerrada bajo la curva en ese intervalo.

- La función de distribución o distribución normal es:
  $F(x)=P(X\leq x)=\int_{-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$. Esta probabilidad coincide con el área acumulada bajo la curva a la izquierda del número x.

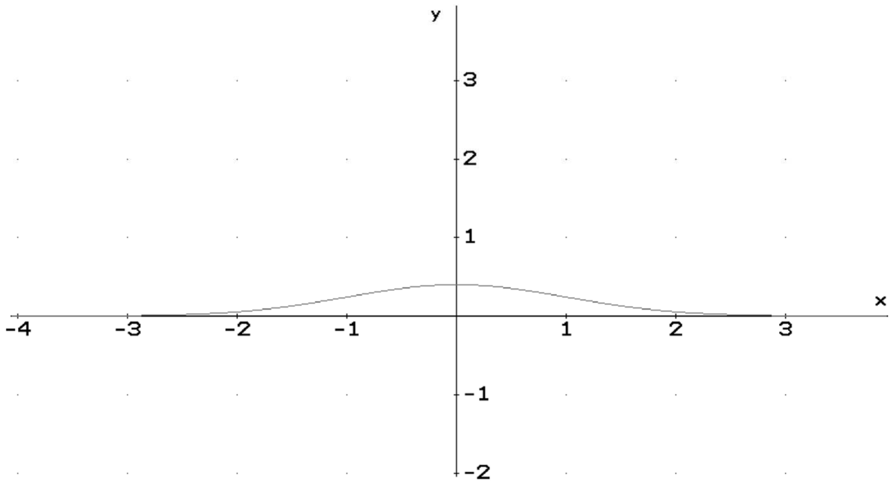


$P(a\leq X\leq b)=\int_{a}^{b} f(x)dx=F(b)-F(a)$. Que es el área de la región encerrada por la función en el intervalo comprendido entre a y b.

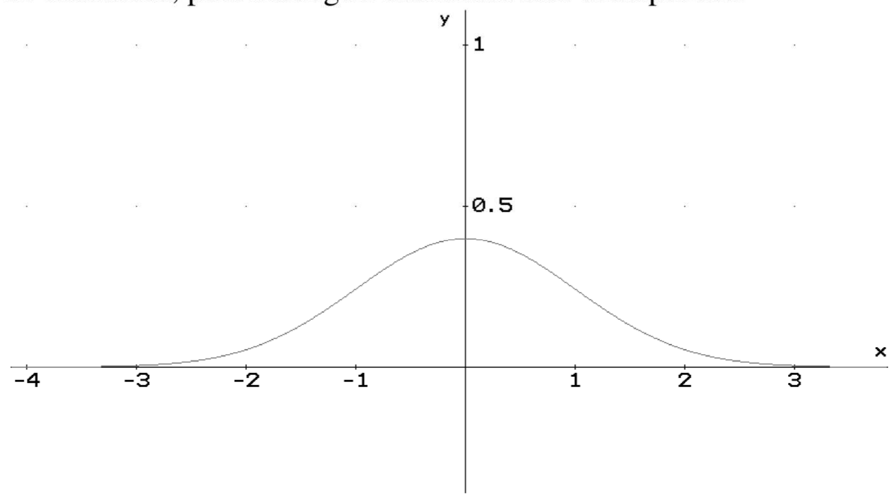


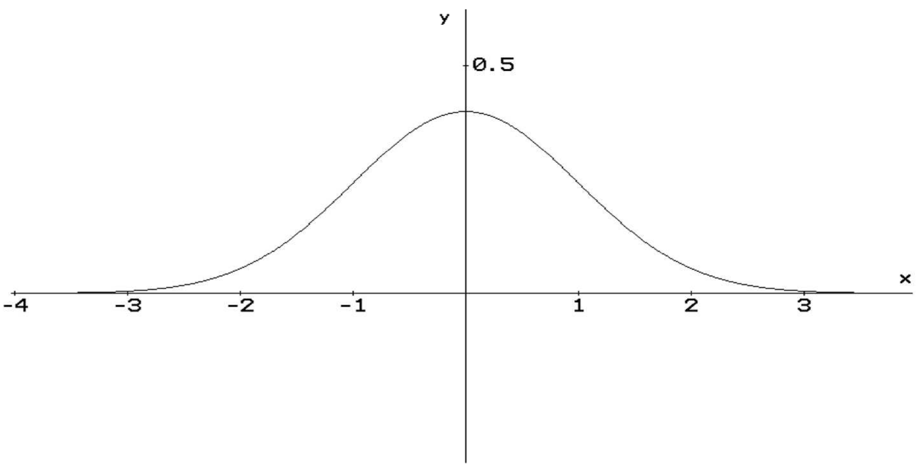
9.3 Tipificación de la variable.

- Si $\mu = 0$ y $\sigma = 1$, es decir, la distribución normal $N(0,1)$, llamada reducida, estándar, o tipificada. Su función de densidad es $f(x) = \dfrac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}x^2}$.



En la práctica se utiliza un dibujo de la función un poco deformada en el eje de ordenadas, para conseguir una forma mas acampanada.





La distribución $N(0,1)$ se encuentra tabulada, lo cual permite un cálculo rápido de las probabilidades asociadas a esta distribución. En la práctica, cada distribución normal tiene su propia media y desviación típica, lo que dificulta el cálculo directo de probabilidades, por lo que se hace una transformación que se llama **tipificación de la variable**, que consiste en hacer el siguiente cambio de variable:

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

A partir de este cambio de variable obtenemos una variable Z que sí es $N(0,1)$ y, por lo tanto, se pueden calcular sus probabilidades utilizando las tablas.

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}z^2} \sigma \cdot dz$$

Haciendo una utilización conjunta de μ y σ y buscando en las tablas, obtenemos unas relaciones muy importantes**:**

- En $(\mu \pm \sigma)$ está el 68,26% de los datos ya que:

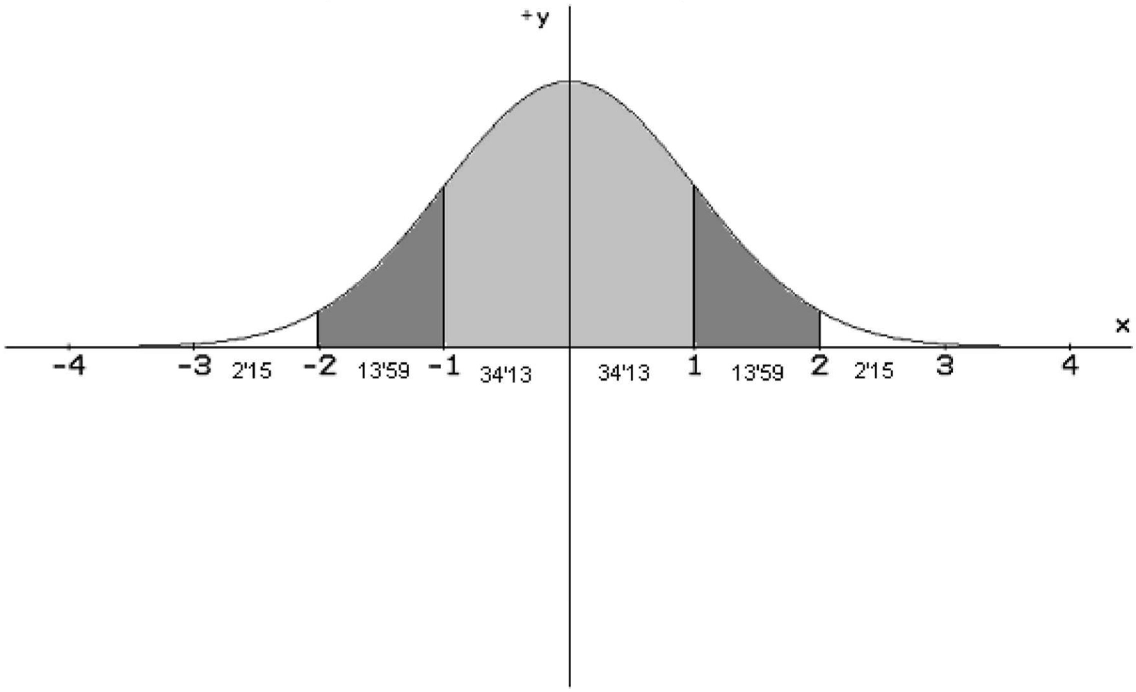
$$P(\mu-\sigma, \mu+\sigma) = P\left(\frac{\mu-\sigma-\mu}{\sigma} < Z < \frac{\mu+\sigma-\mu}{\sigma}\right) = P(-1 < Z < 1) = 0{,}6826.$$

- En $(\mu \pm 2\sigma)$ está el 95,44% de los datos ya que:

$$P(\mu-2\sigma, \mu+2\sigma) = P\left(\frac{\mu-2\sigma-\mu}{\sigma} < Z < \frac{\mu+2\sigma-\mu}{\sigma}\right) = P(-2 < Z < 2) = 0{,}9544.$$

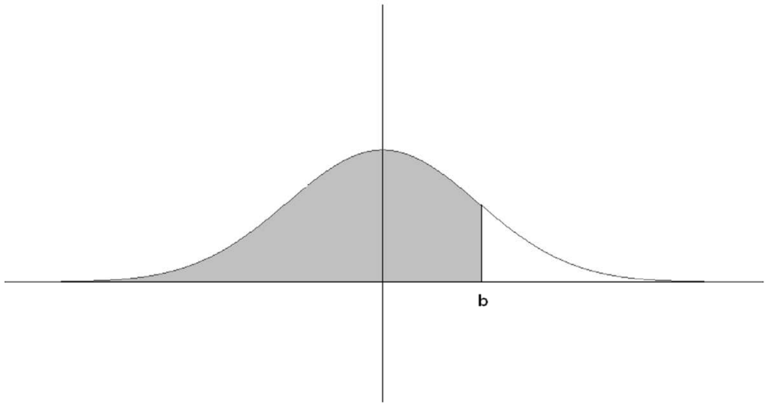
- En $(\mu \pm 3\sigma)$ está el 99,73% de los datos ya que:

$$P(\mu-3\sigma, \mu+3\sigma) = P\left(\frac{\mu-3\sigma-\mu}{\sigma} < Z < \frac{\mu+3\sigma-\mu}{\sigma}\right) = P(-3 < Z < 3) = 0{,}9973.$$



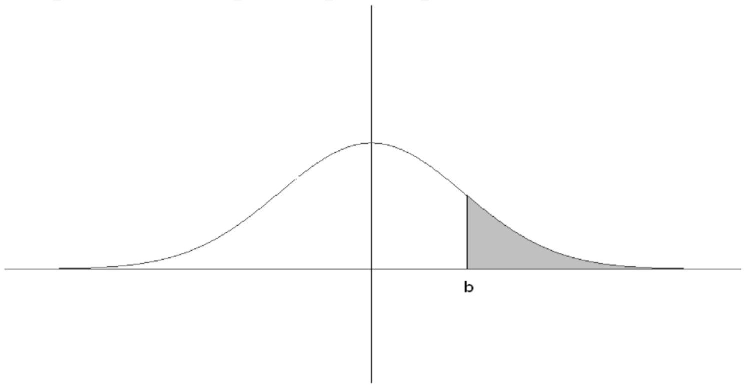
9.4 Manejo de tablas.

- $P(Z<b)$. En este caso se busca en la tabla directamente.



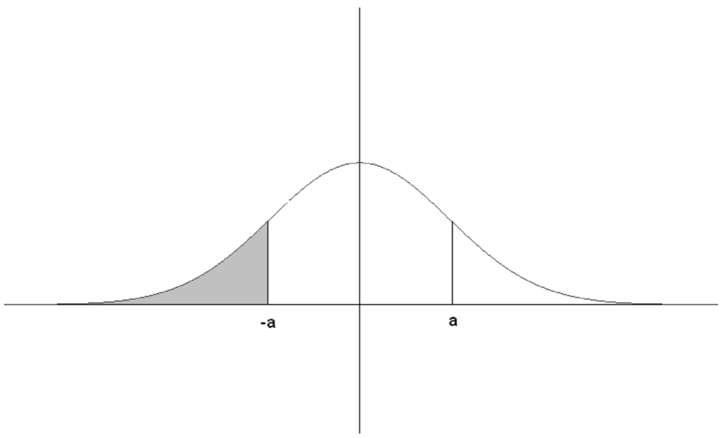
Ejemplo: $P(Z<1{,}45)=0{,}9265=92{,}65\%$.

- $P(Z>b)=1-P(Z<b)$. Aquí hay que tener en cuenta que $Z>b$ y $Z<b$ son complementarios, por lo que sus probabilidades suman 1.



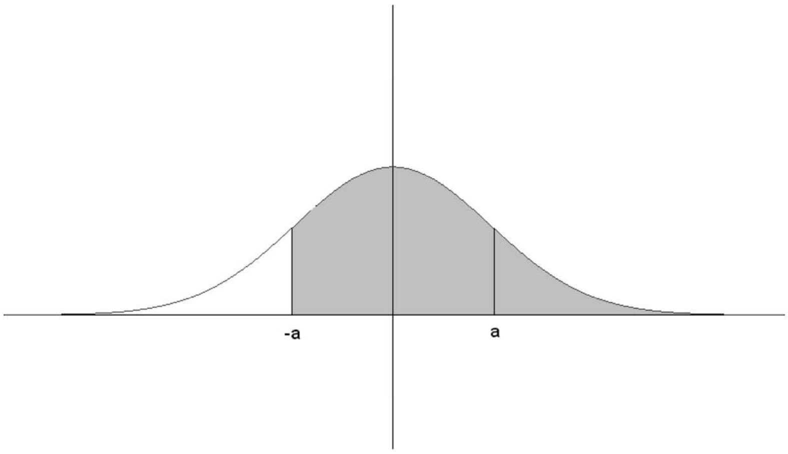
Ejemplo: $P(Z>1{,}45)=1-P(Z<1{,}45)=1-0{,}9265=0{,}0735=7{,}35\%$.

- $P(Z<-a)=P(Z>a)=1-P(Z<a)$.



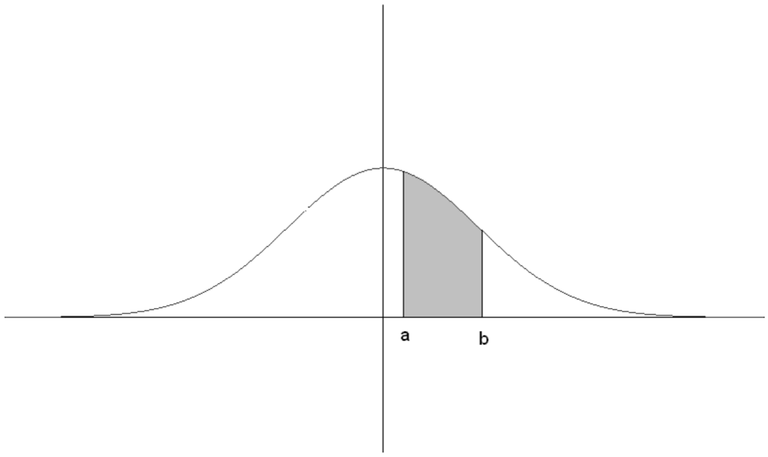
Ejemplo:
$P(Z<-0{,}75)=P(Z>0{,}75)=1-P(Z<0{,}75)=1-0{,}7734=0{,}2266=22{,}66\%$.

- $P(Z > -a) = 1 - P(Z < -a) = 1 - P(Z > a) = 1 - [1 - P(Z < a)] = 1 - 1 + P(Z < a) = P(Z < a)$



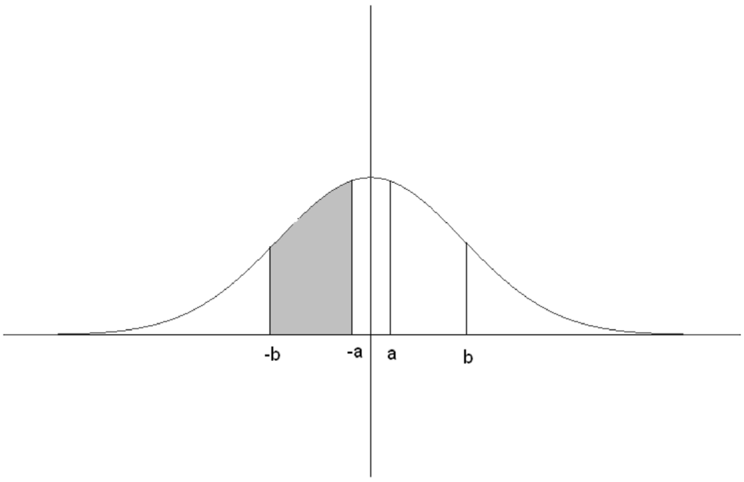
Ejemplo:
$P(Z > -0,75) = 1 - P(Z < -0,75) = 1 - P(Z > 0,75) = 1 - [1 - P(Z < 0,75)] =$
$= 1 - 1 + P(Z < 0,75) = P(Z < 0,75) = 0,7734 = 77,34\%$.

- $P(a < Z < b) = P(Z < b) - P(Z < a)$ y se busca en las tablas.



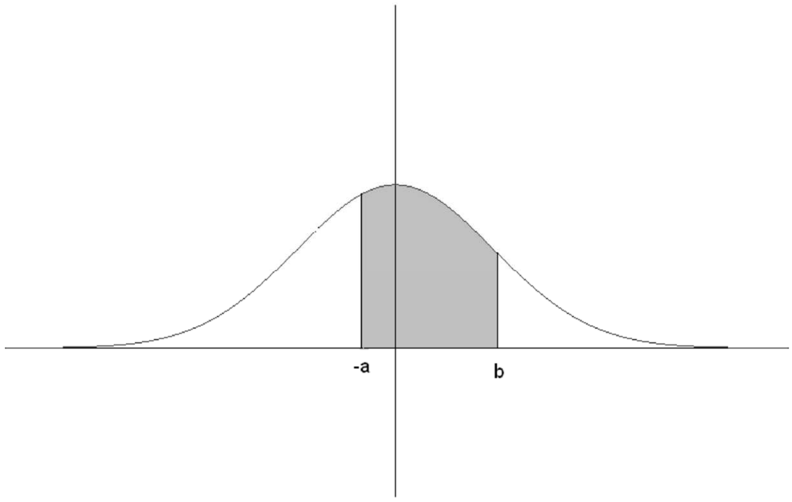
Ejemplo: $P(1,25 < Z < 2,57) = P(Z < 2,57) - P(Z < 1,25) = 0,9949 - 0,8944 =$
$= 0,1005 = 10,05\%$.

- $P(-b < Z < -a) = P(a < Z < b)$



Ejemplo: $P(-2{,}57 < Z < -1{,}25) = P(1{,}25 < Z < 2{,}57) = 0{,}1005 = 10{,}05\%$.

- $P(-a < Z < b) = P(Z < b) - P(Z < -a) = P(Z < b) - P(Z > a) = P(Z < b) - [1 - P(Z < a)] =$
$= P(Z < b) - 1 + P(Z < a)$ y se busca en las tablas.



Ejemplo:
$P(-0{,}53 < Z < 2{,}46) = P(Z < 2{,}46) - P(Z < -0{,}53) = P(Z < 2{,}46) - P(Z > 0{,}53) =$
$= P(Z < 2{,}46) - [1 - P(Z < 0{,}53)] = P(Z < 2{,}46) - 1 + P(Z < 0{,}53) =$
$= 0{,}9931 - 1 + 0{,}7019 = 0{,}6950 = 69{,}50\%.$
Ejemplo: Si tenemos una distribución normal $N(2{,}4)$. Calcular $P(X < 7)$:
$$P(X < 7) = P\left(\frac{X-2}{4} < \frac{7-2}{4}\right) = P\left(Z < -\frac{5}{4}\right) = P(Z > 1{,}25) = 1 - P(Z < 1{,}25) =$$
$= 1 - 0{,}8944 = 0{,}1056 = 10{,}56\%.$

Ejemplo: una cadena hotelera quiere ofrecer a un grupo de personas nuevos destinos turísticos. Para realizar la selección, tiene en cuenta dos factores: la edad y los ingresos mensuales. Se selecciona aleatoriamente un grupo de personas cuyas edades e ingresos siguen unas distribuciones $N(44{,}5)$ y $N(1900{,}150)$ respectivamente. Calcula el porcentaje de personas cuya edad está comprendida entre 38 y 50 años.
$$P(38 < X < 50) = P\left(\frac{38-44}{5} < \frac{X-44}{5} < \frac{50-44}{5}\right) = P(-1{,}2 < Z < 1{,}2) =$$
$= P(Z < 1{,}2) - P(Z < -1{,}2) = P(Z < 1{,}2) - P(Z > 1{,}2) = P(Z < 1{,}2) - [1 - P(Z < 1{,}2)] =$
$= P(Z < 1{,}2) - 1 + P(Z < 1{,}2) = 0{,}8849 - 1 + 0{,}8849 = 0{,}7698 = 76{,}98\%.$

¿Qué porcentaje de personas tienen ingresos mensuales entre 1675€ y 2095€?
$$P(1675 < X < 2095) = P\left(\frac{1675-1900}{150} < \frac{X-1900}{150} < \frac{2095-1900}{150}\right) = P(-1{,}5 < Z < 1{,}3) =$$
$= P(Z < 1{,}3) - P(Z < -1{,}5) = P(Z < 1{,}3) - P(Z > 1{,}5) = P(Z < 1{,}3) - [1 - P(Z < 1{,}5)] =$
$= P(Z < 1{,}3) - 1 + P(Z < 1{,}5) = 0{,}9032 - 1 + 0{,}9332 = 0{,}8364 = 83{,}64\%.$

Ejemplo: el C.I. de los 5600 alumnos de una provincia se distribuyen según una distribución normal $N(112,6)$. Calcula, aproximadamente, cuántos de ellos tienen: (practica ahora tú y comprueba los resultados)

a) más de 112 ....2800 alumnos................la mitad de los alumnos.
b) entre 106 y 118 ....3823 alumnos ....este es el caso $(\mu \pm \sigma)$.
c) entre 106 y 112 ....1911 alumnos
d) menos de 100 ....128 alumnos
e) más de 130 ....7 alumnos
f) entre 118 y 124 ....761 alumnos

9.5 Aproximación de la distribución binomial.

- Cuando los valores de una distribución binomial son elevados y superan los valores de la tabla binomial, se puede obtener un resultado aproximado mediante la distribución normal. Según demostró De Moivre en su teorema: Dada una binomial $B(n,p)$, siendo: $np \geq 5$ y $nq \geq 5$, podemos aproximar las probabilidades mediante la distribución normal $N\left(np, \sqrt{npq}\right)$. Que tipificando, tendremos: $Z = \dfrac{X - np}{\sqrt{npq}}$ que ahora es una distribución $N(0,1)$.
  Este procedimiento de aproximación es tanto más fiable cuanto mayor es el tamaño de la muestra n y cuanto más cerca está p de $0,5$.
  Ejemplo: Se ha comprobado que la probabilidad de que un individuo tenga los ojos marrones es $0,6$. Sea X la variable aleatoria que representa el número de individuos que tienen los ojos marrones de un grupo de 1100. Calcular $P(X > 680)$ y $P(X = 680)$.

$$P(X > 680) = 1 - P(X < 680) = 1 - P\left(Z < \frac{680 - 1100 \cdot 0,6}{\sqrt{1100 \cdot 0,6 \cdot 0,4}}\right) = 1 - P(Z < 1,23) =$$
$$= 1 - 0,8907 = 0,1093 = 10,93\%.$$

$$P(X = 680) = P(679,5 < X < 680,5) = P\left(\frac{679,5 - 1100 \cdot 0,6}{\sqrt{1100 \cdot 0,6 \cdot 0,4}} < Z < \frac{680,5 - 1100 \cdot 0,6}{\sqrt{1100 \cdot 0,6 \cdot 0,4}}\right) =$$
$$= P(1,20 < Z < 1,26) = P(Z < 1,26) - P(Z < 1,20) = 0,8962 - 0,8849 = 0,0113 = 1,13\%.$$

En este apartado debe operarse así, porque en una variable continua, la probabilidad de un valor puntual es nula.

9.6 Introducción a la estadística inferencial.

- La estadística es la ciencia o rama de las Matemáticas que se ocupa de recoger datos, analizarlos y organizarlos, y de realizar las predicciones que sobre esos datos puedan deducirse, tiene dos vertientes básicas:

  a) Estadística descriptiva: se ocupa de recoger datos, analizarlos y organizarlos. Es aquí donde tiene sentido calcular la media, mediana, moda, desviación media, varianza, desviación típica, etc.

b) Estadística inferencial: su nombre viene de inferir, que es extraer una consecuencia o deducir una cosa de otra. Se ocupa de predecir, extraer conclusiones para una población tomando como base una muestra (es decir, una parte de dicha población). Como todas las predicciones, siempre han de hacerse bajo un cierto grado de fiabilidad o confianza.

En la primera parte del tema estudiaremos las diferentes distribuciones muestrales, en ellas se esperan ciertos resultados sobre las características de las muestras cuando los parámetros de la población son conocidos. En la segunda parte del tema se pretende lo contrario, a través de los llamados intervalos de confianza, se buscan conclusiones generales sobre la población a partir de los datos obtenidos en las muestras; además, se intenta medir el grado de confianza que las conclusiones obtenidas merecen.

9.7 Varianza y cuasivarianza.

- Para cálculos posteriores es necesario manejar las siguientes definiciones:

Varianza: $s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} = \dfrac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2$

Desviación típica: $s = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} = \sqrt{\dfrac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2}$

Cuasivarianza: $\hat{s}^2 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$ también se escribe s_{n-1}^2

Desviación típica muestral: $\hat{s} = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$ también se escribe s_{n-1}.

Este valor sirve para aproximar la desviación típica poblacional σ, cuando ésta es desconocida y tenemos una cantidad de datos n>30.

La relación entre varianza y cuasivarianza es: $\hat{s}^2 = \left(\dfrac{n}{n-1}\right)s^2$ y por tanto la relación entre desviación típica y desviación típica muestral es $\hat{s} = \sqrt{\dfrac{n}{n-1}} \cdot s$

9.8 Métodos de muestreo.

- Para no tener que trabajar con toda la población se utiliza el muestreo. Puede ser:
  - Muestreo no probabilístico: no se usa el azar, sino el criterio del investigador, suele presentar grandes sesgos y es poco fiable.
  - Muestreo probabilístico: se utilizan las leyes del azar. Puede ser:

- **Muestreo aleatorio simple** (es el más importante): cada elemento de la población tiene la misma probabilidad de ser elegido, las observaciones se realizan con reemplazamiento, de manera que la población es idéntica en todas las extracciones, o sea, que la selección de un individuo no debe afectar a la probabilidad de que sea seleccionado otro cualquiera, aunque ello comporte que algún individuo pueda ser elegido más de una vez ("se hacen tantas papeletas numeradas como individuos hay, se coge una y se devuelve, se vuelve a coger otra y se devuelve, etc.")
- **Muestreo sistemático**: los elementos de la población están ordenados por listas. Se elige un individuo al azar y a continuación a intervalos constantes se eligen todos los demás hasta completar la muestra. Si el orden de los elementos es tal que los individuos próximos tienden a ser más semejantes que los alejados, el muestreo sistemático tiende a ser más preciso que el aleatorio simple, al cubrir más homogéneamente toda la población.
- **Muestreo estratificado**: debe utilizarse cuando nos interesa que la muestra tenga la misma composición a la de la población la cual se divide en clases o estratos. Si por ejemplo en la población el 20% son mujeres y el 80% hombres, se mantendrá la misma proporción en la muestra.

9.9 Distribuciones de muestreo.

- **Distribución de medias muestrales** $\overline{X}$: Consiste en tomar repetidamente muestras de n-elementos y calcular su media.

  El Teorema central del límite afirma que aunque la población de estudio no siga una distribución normal, la distribución de medias muestrales sí que es normal y tiene los siguientes parámetros:
  - La media de las medias muestrales es igual a la media real de la población es decir: Media $\boxed{\overline{x} = \dfrac{\overline{x}_1 + \overline{x}_2 + \dots\dots\dots + \overline{x}_n}{\text{nº de muestras posibles}} = \mu}$
  - La desviación típica de las medias muestrales es: $\boxed{\sigma_{\overline{x}} = \dfrac{\sigma}{\sqrt{n}}}$

  Esto significa que la distribución de medias muestrales $\overline{X}$ de tamaño n de una población (normal o no normal) se distribuye según una normal $\boxed{N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)}$.

  Ejemplo: Supongamos que tenemos los elementos 2, 4, 6, 8.
  En esta población vamos a tomar todas las muestras posibles de tamaño 2:

  La media y desviación típica son:

  $$\mu = \frac{2+4+6+8}{4} = 5 \qquad\qquad \sigma = \sqrt{\frac{2^2+4^2+6^2+8^2}{4} - 5^2} = 2'23$$

  Queremos comprobar experimentalmente que la media y desviación típica muestrales son respectivamente las siguientes:

  $$\overline{x} = \mu = 5 \qquad\qquad \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2'23}{\sqrt{2}} = 1'5768.$$

  ¿Es esto cierto?

Elementos e_1 e_2		Media de la muestra
2	2	2
2	4	3
2	6	4
2	8	5
4	2	3
4	4	4
4	6	5
4	8	6
6	2	4
6	4	5
6	6	6
6	8	7
8	2	5
8	4	6
8	6	7
8	8	8

La media de las medias muestrales será: $\overline{x} = \dfrac{2+3+4+5+3+....}{16} = 5$

La varianza de las medias muestrales será:

$$\sigma_{\overline{x}} = \sqrt{\frac{(2-5)^2+(3-5)^2+...}{16}} = 1'58$$

Por lo tanto, hemos comprobado que efectivamente: $\overline{x} = \mu$ y $\sigma_{\overline{x}} = \dfrac{\sigma}{\sqrt{n}}$.

Ejemplo: El peso de los recién nacidos en una maternidad se ha distribuido según una ley normal de media μ=3100gr y de desviación típica σ=150gr. Se toman 300 muestras de 100 recién nacidos. ¿Cuál será la probabilidad de que la media de una muestra de se superior a 3130gr?

La distribución muestral sigue una $N\left(3100, \dfrac{150}{\sqrt{100}}\right) = N(3100,15)$ por lo que

$$p(\overline{X} > 3130) = \left(Z > \frac{3130-3100}{15}\right) = p(Z>2) = 1-p(Z<2) =$$

$$= 1 - 0'9772 = 0'0228 = 2'28\%.$$

Por lo tanto la probabilidad pedida es sólo del 2'28%. Se espera por tanto, que de las 300 muestras tengamos $300 \cdot \dfrac{2'28}{100} = 6'84 \approx 7$ muestras con una media superior a 3130 gr.

- **Distribución de proporciones muestrales** $\hat{P}$: Dada una población en la que un determinado suceso tiene una cierta probabilidad, se trata de tomar una muestra de n-elementos, obtener la proporción con la que se da el suceso en dicha muestra y repetir este proceso sucesivamente. La media y la desviación típica de las proporciones muestrales de la variable aleatoria $\hat{P}$ son:

$\boxed{\overline{p} = p}$ $\boxed{\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}}}$

Esta distribución es aproximadamente normal para valores grandes de n (mayores de 30) y en consecuencia puede estudiarse como: $\boxed{N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)}$.

Ejemplo: En un lote de bombillas la probabilidad de que sea defectuosa es del 5%. Se toman 50 muestras de 100 bombillas. Hallar la probabilidad de que en una muestra tengamos más de 7 bombillas defectuosas.

La distribución muestral de proporciones sigue una distribución

$N\left(0'05, \sqrt{\frac{0'05 \cdot 0'95}{100}}\right) = N(0'05, 0'022)$ por lo que:

$p(\hat{P} > 0'07) = \left(Z > \frac{0'07 - 0'05}{0'022}\right) = p(Z > 0'91) = 1 - 0'8186 = 0'1814.$

La probabilidad pedida es pues del $18'14\%$, y se en las 50 muestras habrá $50 \cdot \frac{18'14}{100} = 9'07 \approx 9$ muestras con más de 7 bombillas defectuosas.

- **Distribución de sumas muestrales** T: Tomamos una muestra de n-elementos y apuntamos la suma de los resultados obtenidos, y así sucesivamente con varias muestras. Se puede afirmar que:
  - La media de las sumas muestrales es $n\mu$.
  - La desviación típica de las sumas muestrales vale: $\sigma\sqrt{n}$.

  Esto significa que la distribución de sumas muestrales de tamaño n se distribuye según una normal: $\boxed{N(n\mu, \sigma\sqrt{n})}$.

  Ejemplo: El peso de una población sigue una distribución normal N(70,5). Tomamos 80 muestras de 50 personas y sumamos los pesos de cada muestra. Hallar la probabilidad de que la suma de los pesos de una muestra sea mayor de 3550 Kg.

  La distribución de sumas muestrales sigue una distribución normal $N(50 \cdot 70, 5 \cdot \sqrt{50}) = N(3500, 35'35)$ por lo que

  $p(T > 3550) = p\left(Z > \frac{3550 - 3500}{35'35}\right) = p(Z > 1'41) = 1 - 0'9207 = 0'0893.$

  La probabilidad pedida es del $8'93\%$. Se espera que en las 80 muestras habrá $80 \cdot \frac{8'93}{100} = 7'14 \approx 7$ muestras con más de 3550 Kg. de peso total.

- **Distribución de diferencias muestrales** $X_1 - X_2$: Consiste en tomar dos muestras, una de cada población, formamos parejas y apuntamos la diferencia de los resultados, para continuar este proceso con varias muestras. Se puede afirmar que:
  - La media de las diferencias muestrales es $\mu_1 - \mu_2$.
  - La desviación típica de las diferencias muestrales vale: $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Esto significa que la distribución de diferencias muestrales se distribuye según una normal $\boxed{N\left(\mu_1-\mu_2,\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)}$.

Ejemplo: Las alturas de los chicos de 13 y 15 años siguen distribuciones normales N(165,3) y N(175,5) respectivamente. Tomamos 70 muestras de 50 parejas de chicos de cada una de las edades y calculamos la diferencia de sus alturas en cada pareja. Hallar la probabilidad de que la diferencia de alturas medias sea menor de 9cm.

La distribución de sumas muestrales sigue una distribución normal $N\left(175-165,\sqrt{\frac{3^2}{50}+\frac{5^2}{50}}\right)=N(10{,}0'825)$ por lo que la probabilidad es:

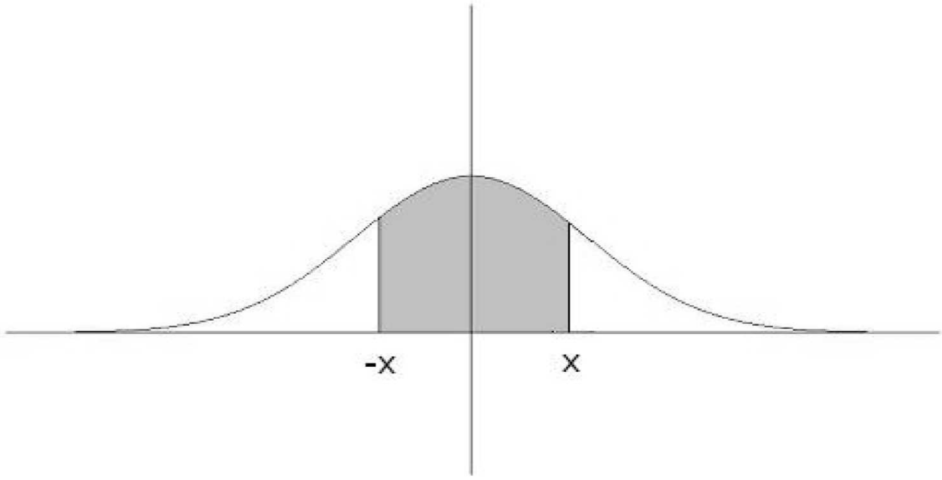
$$p(X_1-X_2<9)=p\left(Z<\frac{9-10}{0'825}\right)=p(Z<-1'21)=p(Z>1'21)=1-p(Z<1'21))=$$
$$=1-0'8869=0'1131.$$

La probabilidad pedida es del $11'31\%$.

9.10 Intervalos de confianza.

- Observemos la siguiente igualdad en la distribución normal N(0,1):

$$p(-x<Z<x)=p(Z<x)-p(Z<-x)=p(Z<x)-p(Z>x)=$$
$$=1-p(Z>x)-p(Z>x)=1-2p(Z>x)$$



Supongamos que $x=Z_{\frac{\alpha}{2}}$ es el valor en el que $p\left(Z<-Z_{\frac{\alpha}{2}}\right)=p\left(Z>Z_{\frac{\alpha}{2}}\right)=\frac{\alpha}{2}$

Sustituyendo en la fórmula anterior: $p\left(-Z_{\frac{\alpha}{2}}<Z<Z_{\frac{\alpha}{2}}\right)=1-2\cdot\frac{\alpha}{2}=1-\alpha$.

Multiplicando por σ en el intervalo, $p\left(-Z_{\frac{\alpha}{2}}\sigma<Z\sigma<Z_{\frac{\alpha}{2}}\sigma\right)=1-\alpha$.

Recordando que Z es la variable normal tipificada: $Z=\frac{X-\mu}{\sigma}$, tendremos:

$$p\left(-Z_{\frac{\alpha}{2}}\sigma < \frac{X-\mu}{\sigma}\sigma < Z_{\frac{\alpha}{2}}\sigma\right) = 1-\alpha \Rightarrow p\left(-Z_{\frac{\alpha}{2}}\sigma < X-\mu < Z_{\frac{\alpha}{2}}\sigma\right) = 1-\alpha$$

Restando ahora X: $p\left(-X-Z_{\frac{\alpha}{2}}\sigma < -\mu < -X+Z_{\frac{\alpha}{2}}\sigma\right) = 1-\alpha$.

Cambiando de signo y de símbolo: $p\left(X+Z_{\frac{\alpha}{2}}\sigma > \mu > X-Z_{\frac{\alpha}{2}}\sigma\right) = 1-\alpha$.

Colocando en la forma habitual: $p\left(X-Z_{\frac{\alpha}{2}}\sigma < \mu < X+Z_{\frac{\alpha}{2}}\sigma\right) = 1-\alpha$.

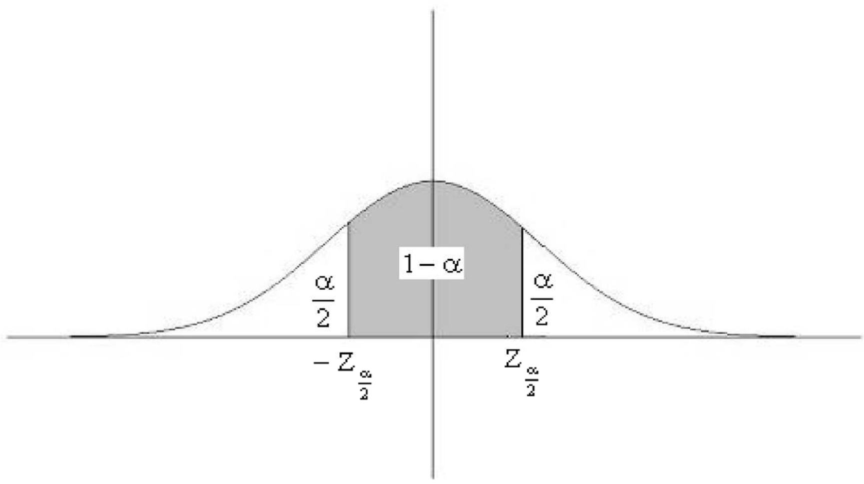
Utilizando la media en lugar de la variable:

$$\boxed{p\left(\overline{x}-Z_{\frac{\alpha}{2}}\sigma < \mu < \overline{x}+Z_{\frac{\alpha}{2}}\sigma\right) = 1-\alpha}$$

A este intervalo se le llama **intervalo de confianza**.
A la región que está fuera del intervalo se le llama **región crítica**.
Al valor $1-\alpha$ se le llama **nivel de confianza**.
Al valor α se le llama **nivel de significación o riesgo**.
El margen de error es la diferencia entre el extremo superior del intervalo y el extremo inferior.
El error máximo es: $E = Z_{\frac{\alpha}{2}}\sigma$.

Para valores $n > 30$ la desviación típica poblacional σ se puede aproximar mediante la desviación típica muestral $\hat{s}$.
Hay tres valores de α muy usados que son:

- $\alpha = 0'01 \quad 1-\alpha = 0'99$ (99 %) $\quad Z_{\frac{\alpha}{2}} = 2'575$
- $\alpha = 0'05 \quad 1-\alpha = 0'95$ (95 %) $\quad Z_{\frac{\alpha}{2}} = 1'96$
- $\alpha = 0'10 \quad 1-\alpha = 0'90$ (90 %) $\quad Z_{\frac{\alpha}{2}} = 1'645$



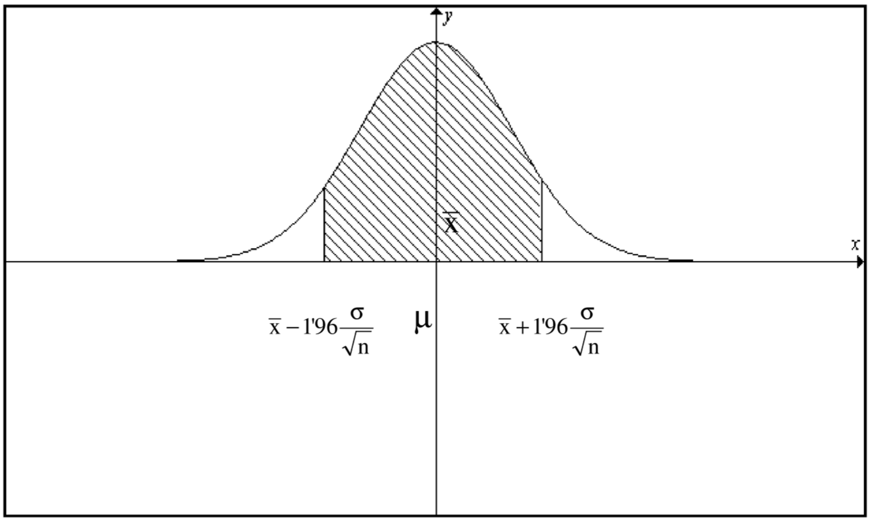
- Intervalo de confianza en la distribución de medias muestrales $\overline{X}$:

  Como la distribución de medias muestrales es $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, tendremos la siguiente expresión: $\boxed{p\left(\overline{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha}$

  Así por ejemplo para $1-\alpha = 0'95$ (95%), $\alpha = 0'05$, $Z_{\frac{\alpha}{2}} = 1'96$, podemos afirmar que la media poblacional μ se encuentra en el siguiente intervalo con un nivel de confianza del 95%.

  $$\left(\overline{x} - Z_{0'025}\frac{\sigma}{\sqrt{n}}, \overline{x} + Z_{0'025}\frac{\sigma}{\sqrt{n}}\right) = \left(\overline{x} - 1'96\frac{\sigma}{\sqrt{n}}, \overline{x} + 1'96\frac{\sigma}{\sqrt{n}}\right)$$

  Dicho de otra forma: podemos afirmar que en ese intervalo tenemos una probabilidad del 95% de que dentro esté la media poblacional μ.



- Intervalo de confianza en la distribución de proporciones muestrales $\hat{P}$:
  Esta distribución es aproximadamente normal para valores grandes de n (mayor de 30) en consecuencia puede estudiarse como una $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$ y la expresión para el intervalo de confianza será de la forma:

  $$\boxed{p\left(\overline{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \le p \le \overline{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right) = 1-\alpha}$$

- Intervalo de confianza en la distribución de diferencias muestrales $X_1 - X_2$:
  La distribución de diferencias muestrales sigue una normal $N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$, y la expresión del intervalo de confianza es:

$$p\left(\bar{x}_1-\bar{x}_2-Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\le\mu_1-\mu_2\le\bar{x}_1-\bar{x}_2+Z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)=1-\alpha$$

9.11 Error y tamaño de la muestra.

- En la expresión para el intervalo de confianza general:
  $p\left(\bar{x}-Z_{\frac{\alpha}{2}}\sigma<\mu<\bar{x}+Z_{\frac{\alpha}{2}}\sigma\right)=1-\alpha$. El error máximo cometido es $E=Z_{\frac{\alpha}{2}}\sigma$
  Si estamos trabajando con una distribución de medias muestrales, entonces $\sigma=\frac{\sigma}{\sqrt{n}}$ y el error es $\boxed{E=Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}}$, por lo que despejando $n=\left(Z_{\frac{\alpha}{2}}\frac{\sigma}{E}\right)^2$.
  Ejemplo: Se desea realizar una investigación para estimar el peso medio de los hijos de madres fumadoras. Se admite un error máximo de 50 gr., con una confianza del 95%. Si por estudios se sabe que la desviación típica es de 400 gr. ¿Qué tamaño mínimo de muestra se necesita en la investigación?
  La muestra debe tener: $n=\left(1'96\frac{400}{50}\right)^2=245'86$, es decir 246 niños.
- Si estamos trabajando con una distribución de proporciones muestrales, entonces $\sigma=\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ y el error es $\boxed{E=Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}}$, por lo que despejando $n=\left(\frac{Z_{\frac{\alpha}{2}}}{E}\right)^2\cdot\bar{p}\cdot(1-\bar{p})$

  Ejemplo: Se desea realizar una investigación para estimar el número de votantes de un candidato a las elecciones. Se ha tomado una muestra de 100 personas de las que 47 han respondido que le votarán en las próximas elecciones. Con un error máximo de $0'1$, determinar el tamaño de una nueva muestra para que la confianza en que el candidato sea elegido sea del 95%.
  El tamaño mínimo de la muestra debe ser $n=\left(\frac{1'96}{0'1}\right)^2\cdot 0'47\cdot 0'53=95'69$, es decir 96 votantes de los que 45 (el 47%) probablemente votarán al candidato.

9.12 Contraste de hipótesis.

Para **resolver un problema por un contraste de hipótesis**, hay que seguir los siguientes pasos:

1. Formular de forma clara y concisa la hipótesis nula H_0 y la hipótesis alternativa H_1.
2. Fijar el nivel de significación α, llamado también probabilidad de Error del tipo I. Nos interesa que el error a cometer sea pequeño, por lo que α será de un valor próximo a 0 ($0'05$; $0'01$; etc).

3. Elegir el estadístico del contraste. Este estadístico tiene que tener una distribución en el muestreo conocida y depender del parámetro que estamos estudiando en el contraste de hipótesis.
4. Hallar las regiones de aceptación y de rechazo. Para ello, debes observar primero si el contraste es unilateral o bilateral.
5. Tomar una muestra en la población y calcular en ella el valor muestral del estadístico del contraste.
6. Observar si el valor calculado en el apartado anterior cae dentro de la región de aceptación o de la región crítica, y decidir por tanto, si se acepta la hipótesis nula o se rechaza.

- **Contraste de hipótesis sobre la media poblacional:**
  La media muestral puede ser diferente de la media poblacional. Lo normal es que estas diferencias sean pequeñas y estén justificadas por el azar, pero podría suceder que no fuesen debidas al azar sino a que los parámetros poblacionales sean otros, y que por los motivos que sean, han cambiado.
  El contraste de hipótesis es el instrumento que permite decidir si esas diferencias pueden interpretarse como fluctuaciones del azar (hipótesis nula) o bien, son de tal importancia que requieren una explicación distinta (hipótesis alternativa). Al igual que en los intervalos de confianza las conclusiones se formularán en términos de probabilidad.

Comparando la media poblacional y la media muestral, ¿podemos asegurar que esa muestra procede de una población de media μ_0? (**hipótesis nula**). La respuesta será sí cuando μ_0 pertenezca al intervalo de confianza de $\overline{x}$, para el nivel de significación prefijado, por el contrario la respuesta será no, cuando no pertenece a dicho intervalo.
Otra forma de realizar el contraste consiste en utilizar el llamado **estadístico de contraste** y observar si pertenece **a la región de aceptación.**

— <u>Se acepta la hipótesis nula</u> $\boxed{\mu = \mu_0}$ Si $\mu_0 \in \left(\overline{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$.

Otra forma más práctica es la siguiente: se acepta la hipótesis si el estadístico de contraste $\boxed{\dfrac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}}$ pertenece a la región de aceptación $\boxed{\left(-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}\right)}$. En caso contrario se rechaza la hipótesis nula.

— <u>Se acepta la hipótesis nula</u> $\boxed{\mu \le \mu_0}$ si el estadístico de contraste $\boxed{\dfrac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}}$ pertenece a la región de aceptación $\boxed{(-\infty, Z_\alpha)}$

— <u>Se acepta la hipótesis nula</u> $\boxed{\mu \ge \mu_0}$ si el estadístico de contraste $\boxed{\dfrac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}}$ pertenece a la región de aceptación $\boxed{(-Z_\alpha, +\infty)}$

Nota: No olvidemos que en todas las expresiones anteriores si se desconoce la desviación típica de la población debemos sustituirla por la desviación típica muestral (o raíz cuadrada de la cuasivarianza de la muestra).

Ejemplo: un informe indica que el precio medio del billete de avión entre Canarias y Madrid es, como máximo, de 120 € con una desviación típica de 40 €. Se toma una muestra de 100 viajeros y se obtiene que la media de los precios de sus billetes es de 128 €. ¿Se puede aceptar, con un nivel de significación del 10%, la afirmación de partida?

1.- La hipótesis nula es $H_0: \mu \leq 120$ y la hipótesis alternativa es $H_1: \mu > 120$.

2.- El nivel de significación es $\alpha = 0'10$.

3.- El estadístico de contraste es $\dfrac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

4.- La región crítica es $(1'645, +\infty)$.

5.- El valor de μ en el muestreo es 128 y por tanto el valor del estadístico de contraste en el muestreo es $\dfrac{128 - 120}{\frac{40}{\sqrt{100}}} = 2$.

6.- Como ese valor está en la región crítica, rechazamos H_0 a ese nivel de significación.

Conclusión: No es correcto ese precio medio máximo, sino que es mayor.

- **Contraste de hipótesis sobre la proporción:**

Por analogía con el apartado anterior para responder a la pregunta: ¿Puede asegurarse que esa muestra de proporción $\overline{p}$ procede de una población con proporción p_0?

La respuesta será sí, cuando p_0 pertenezca al intervalo de confianza de $\overline{p}$, para el nivel de significación prefijado, por el contrario la respuesta será no en caso contrario.

— Se acepta la hipótesis nula $\boxed{p = p_0}$ si el estadístico de contraste $\boxed{\dfrac{\overline{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}$ pertenece a la región de acceptación $\boxed{\left(-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}\right)}$

— Análogamente, se acepta la hipótesis nula $\boxed{p \leq p_0}$ si el estadístico de contraste $\boxed{\dfrac{\overline{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}$ pertenece a la región de aceptación $\boxed{(-\infty, Z_\alpha)}$

— Por último, se acepta la hipótesis nula $\boxed{p \geq p_0}$ si el estadístico de contraste $\boxed{\dfrac{\overline{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}$ pertenece a la región de aceptación $\boxed{(-Z_\alpha, +\infty)}$

Ejemplo: en febrero de 2011 los medios de comunicación se hacían eco de un informe según el cual aproximadamente cerca del 30% de los niños y niñas con edades comprendidas entre 3 y 12 años sufría sobrepeso. ¿Será esto verdad? El área de Sanidad de un ayuntamiento, al conocer la noticia, decidió comprobar si en su pueblo esto también era así de alarmante. Así, se puso en contacto con una empresa de estudios estadísticos que realizaron 250 encuestas, o sea, analizaron el peso de 250 niños y niñas de la localidad en esa franja de edad y comprobaron que 53 sufrían sobrepeso. ¿Confirmarán a partir del estudio con un nivel de significación del 1% que efectivamente la proporción de sobrepeso infantil en la localidad es del 30%?

1.- La hipótesis nula es $H_0:\ p = 0'3$ y la hipótesis alternativa es $H_1:\ p \neq 0'3$.

2.- El nivel de significación es $\alpha = 0'01$.

3.- El estadístico de contraste es $\dfrac{\overline{p}-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$

4.- La región crítica es $(-\infty,-2'575)\cup(2'575,+\infty)$.

5.- El valor de p en el muestreo es $\dfrac{53}{250} = 0'212$ y por tanto el valor del estadístico de contraste en el muestreo es $\dfrac{0'212-0'3}{\sqrt{\dfrac{0'3 \cdot 0'7}{250}}} = -3'0362$.

6.- Como ese valor está en la región crítica, rechazamos H_0 a ese nivel de significación.

Conclusión: no es esa la proporción de obesidad infantil, sino que es inferior.

En cualquiera de los dos contrastes estudiados pueden darse las siguientes posibilidades:

	H_0 ES CIERTA	**H_0 ES FALSA**
Aceptamos H_0	Decisión correcta $p = 1-\alpha$	Error tipo II $p = \beta$
Rechazamos H_0	Error tipo I $p = \alpha$	Decisión correcta $p = 1-\beta$

Error de tipo I: es el que cometemos cuando rechazamos la hipótesis nula siendo verdadera.
Error de tipo II: es el que cometemos cuando aceptamos la hipótesis nula siendo falsa.
Nivel de significación, o riesgo α: es la probabilidad de cometer un error del tipo I. Es decir, es la probabilidad de rechazar una hipótesis H_0 correcta.
Potencia de un contraste $1-\beta$: es la probabilidad de rechazar una hipótesis H_0 falsa. Si la potencia es máxima el error tipo II es mínimo.

Resumen y fórmulas:

Distribución binomial.

- Probabilidad: $P(X=k)=\binom{n}{k}p^k\cdot q^{n-k}$
- Media: $\mu = n\cdot p$
- Desviación típica: $\sigma=\sqrt{n\cdot p\cdot q}$

Distribución normal.

- Tipificación: $Z=\dfrac{X-\mu}{\sigma}$=N(0,1)
- La Binomial B(n,p) si $np\geq 5$ y $nq\geq 5$, puede aproximarse mediante la distribución normal N(np, $\sqrt{npq}$).

Distribución de medias muestrales.

- Media: $\overline{x}=\mu$
- Desviación típica: $\sigma_{\overline{x}}=\dfrac{\sigma}{\sqrt{n}}$

Distribución de proporciones muestrales.

- Media: $\overline{p}=p$
- Desviación típica: $\sigma_{\overline{p}}=\sqrt{\dfrac{p(1-p)}{n}}$

Distribución de sumas muestrales.

- Media: $n\mu$
- Desviación típica: $\sigma\sqrt{n}$

Distribución de diferencias muestrales.

- Media: $\mu_1-\mu_2$
- Desviación típica: $\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$

Intervalo de confianza.

- $p\left(\overline{x}-Z_{\frac{\alpha}{2}}\sigma<\mu<\overline{x}+Z_{\frac{\alpha}{2}}\sigma\right)=1-\alpha$. Error máximo: $E=Z_{\frac{\alpha}{2}}\sigma$
- A este intervalo se le llama intervalo de confianza. Al valor $1-\alpha$ se le llama nivel de confianza.
- A la región que está fuera del intervalo se le llama región crítica. Al valor α se le llama nivel de significación o riesgo.

- $\alpha=0'01$ (1%) $1-\alpha=0'99$ (99 %) $Z_{\frac{\alpha}{2}}=2'575$
- $\alpha=0'05$ (5%) $1-\alpha=0'95$ (95 %) $Z_{\frac{\alpha}{2}}=1'96$
- $\alpha=0'10$ (10%) $1-\alpha=0'90$ (90 %) $Z_{\frac{\alpha}{2}}=1'645$

Expresión del intervalo de confianza y error en la distribución de medias.

- $p\left(\overline{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$
- El error es $E = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$, por lo que despejando $n = \left(Z_{\frac{\alpha}{2}} \frac{\sigma}{E}\right)^2$.

Expresión del intervalo de confianza y error en la distribución de proporciones.

- $p\left(\overline{p} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \leq p \leq \overline{p} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right) = 1 - \alpha$
- El error es $E = Z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$, luego $n = \left(\frac{Z_{\frac{\alpha}{2}}}{E}\right)^2 \cdot \overline{p} \cdot (1-\overline{p})$.

Contraste de hipótesis sobre la media. Estadístico de contraste: $\boxed{\dfrac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}}$

- Hipótesis nula H_0: $\boxed{\mu = \mu_0}$ Región de aceptación: $\left(-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}\right)$
- Hipótesis nula H_0: $\boxed{\mu \leq \mu_0}$ Región de aceptación: $(-\infty, Z_{\alpha})$
- Hipótesis nula H_0: $\boxed{\mu \geq \mu_0}$ Región de aceptación: $(-Z_{\alpha}, +\infty)$

Contraste de hipótesis sobre la proporción. Estadístico de contraste: $\boxed{\dfrac{\overline{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}$

- Hipótesis nula H_0: $\boxed{p = p_0}$ Región de aceptación: $\left(-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}\right)$
- Hipótesis nula H_0: $\boxed{p \leq p_0}$ Región de aceptación: $(-\infty, Z_{\alpha})$
- Hipótesis nula H_0: $\boxed{p \geq p_0}$ Región de aceptación: $(-Z_{\alpha}, +\infty)$

Tabla de la distribución normal N(0,1) $P(Z<z_0)$

z_0	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50400	0,50800	0,51200	0,51600	0,51990	0,52390	0,52790	0,53190	0,53590
0,1	0,53980	0,54380	0,54780	0,55170	0,55570	0,55960	0,56360	0,56750	0,57140	0,57530
0,2	0,57930	0,58320	0,58710	0,59100	0,59480	0,59870	0,60260	0,60640	0,61030	0,61410
0,3	0,61790	0,62170	0,62550	0,62930	0,63310	0,63680	0,64060	0,64430	0,64800	0,65170
0,4	0,65540	0,65910	0,66280	0,66640	0,67000	0,67360	0,67720	0,68080	0,68440	0,68790
0,5	0,69150	0,69500	0,69850	0,70190	0,70540	0,70880	0,71230	0,71570	0,71900	0,72240
0,6	0,72570	0,72910	0,73240	0,73570	0,73890	0,74220	0,74540	0,74860	0,75170	0,75490
0,7	0,75800	0,76110	0,76420	0,76730	0,77040	0,77340	0,77640	0,77940	0,78230	0,78520
0,8	0,78810	0,79100	0,79390	0,79670	0,79950	0,80230	0,80510	0,80780	0,81060	0,81330
0,9	0,81590	0,81860	0,82120	0,82380	0,82640	0,82890	0,83150	0,83400	0,83650	0,83890
1,0	0,84130	0,84380	0,84610	0,84850	0,85080	0,85310	0,85540	0,85770	0,85990	0,86210
1,1	0,86430	0,86650	0,86860	0,87080	0,87290	0,87490	0,87700	0,87900	0,88100	0,88300
1,2	0,88490	0,88690	0,88880	0,89070	0,89250	0,89440	0,89620	0,89800	0,89970	0,90150
1,3	0,90320	0,90490	0,90660	0,90820	0,90990	0,91150	0,91310	0,91470	0,91620	0,91770
1,4	0,91920	0,92070	0,92220	0,92360	0,92510	0,92650	0,92790	0,92920	0,93060	0,93190
1,5	0,93320	0,93450	0,93570	0,93700	0,93820	0,93940	0,94060	0,94180	0,94290	0,94410
1,6	0,94520	0,94630	0,94740	0,94840	0,94950	0,95050	0,95150	0,95250	0,95350	0,95450
1,7	0,95540	0,95640	0,95730	0,95820	0,95910	0,95990	0,96080	0,96160	0,96250	0,96330
1,8	0,96410	0,96490	0,96560	0,96640	0,96710	0,96780	0,96860	0,96930	0,96990	0,97060
1,9	0,97130	0,97190	0,97260	0,97320	0,97380	0,97440	0,97500	0,97560	0,97610	0,97670
2,0	0,97720	0,97780	0,97830	0,97880	0,97930	0,97980	0,98030	0,98080	0,98120	0,98170
2,1	0,98210	0,98260	0,98300	0,98340	0,98380	0,98420	0,98460	0,98500	0,98540	0,98570
2,2	0,98610	0,98640	0,98680	0,98710	0,98750	0,98780	0,98810	0,98840	0,98870	0,98900
2,3	0,98930	0,98960	0,98980	0,99010	0,99040	0,99060	0,99090	0,99110	0,99130	0,99160
2,4	0,99180	0,99200	0,99220	0,99250	0,99270	0,99290	0,99310	0,99320	0,99340	0,99360
2,5	0,99380	0,99400	0,99410	0,99430	0,99450	0,99460	0,99480	0,99490	0,99510	0,99520
2,6	0,99530	0,99550	0,99560	0,99570	0,99590	0,99600	0,99610	0,99620	0,99630	0,99640
2,7	0,99650	0,99660	0,99670	0,99680	0,99690	0,99700	0,99710	0,99720	0,99730	0,99740
2,8	0,99740	0,99750	0,99760	0,99770	0,99770	0,99780	0,99790	0,99790	0,99800	0,99810
2,9	0,99810	0,99820	0,99820	0,99830	0,99840	0,99840	0,99850	0,99850	0,99860	0,99860
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997

$1-\alpha$	75%	80%	85%	90%	91%	92%	93%	94%	95%	96%	97%	98%	99%
α	25%	20%	15%	10%	9%	8%	7%	6%	5%	4%	3%	2%	1%
$Z_{\alpha/2}$	1,150	1,282	1,440	1,645	1,695	1,751	1,812	1,881	1,960	2,054	2,170	2,326	2,576
Z_{α}	0,674	0,842	1,036	1,282	1,345	1,405	1,476	1,555	1,645	1,751	1,881	2,054	2,326

Nivel de Significación Bilateral

$Z_{\alpha/2}$	α	0.37	0.711382	0.75	0.453255	1.13	0.258476	1.51	0.131043	1.89	0.058758	2.26	0.023821	2.63	0.008538
0.00	1.000.000	0.38	0.703945	0.76	0.447255	1.14	0.254286	1.52	0.128511	1.90	0.057433	2.27	0.023208	2.64	0.008291
0.01	0.992021	0.39	0.696537	0.77	0.441300	1.15	0.250144	1.53	0.126017	1.91	0.056133	2.28	0.022608	2.65	0.008049
0.02	0.984043	0.40	0.689157	0.78	0.435391	1.16	0.246049	1.54	0.123560	1.92	0.054858	2.29	0.022021	2.66	0.007814
0.03	0.976067	0.41	0.681806	0.79	0.429528	1.17	0.242001	1.55	0.121142	1.93	0.053607	2.30	0.021448	2.67	0.007585
0.04	0.968093	0.42	0.674485	0.80	0.423711	1.18	0.238000	1.56	0.118760	1.94	0.052380	2.31	0.020888	2.68	0.007362
0.05	0.960122	0.43	0.667196	0.81	0.417940	1.19	0.234046	1.57	0.116415	1.95	0.051176	2.32	0.020341	2.69	0.007145
0.06	0.952156	0.44	0.659937	0.82	0.412216	1.20	0.230139	1.58	0.114107	**1.9599**	**0.050000**	2.33	0.019806	2.70	0.006934
0.07	0.944194	0.45	0.652710	0.83	0.406539	1.21	0.226279	1.59	0.111835	1.96	0.049996	2.34	0.019284	2.71	0.006728
0.08	0.936237	0.46	0.645516	0.84	0.400908	1.22	0.222465	1.60	0.109599	1.97	0.048838	2.35	0.018773	2.72	0.006528
0.09	0.928287	0.47	0.638355	0.85	0.395325	1.23	0.218697	1.61	0.107398	1.98	0.047704	2.36	0.018275	2.73	0.006333
0.10	0.920344	0.48	0.631227	0.86	0.389789	1.24	0.214975	1.62	0.105232	1.99	0.046591	2.37	0.017788	2.74	0.006144
0.11	0.912409	0.49	0.624134	0.87	0.384300	1.25	0.211300	1.63	0.103101	2.00	0.045500	2.38	0.017313	2.75	0.005960
0.12	0.904483	0.50	0.617075	0.88	0.378859	1.26	0.207669	**1.64**	**0.101005**	2.01	0.044431	2.39	0.016848	2.76	0.005780
0.13	0.896566	0.51	0.610051	0.89	0.373466	1.27	0.204085	**1.65**	**0.098943**	2.02	0.043383	2.40	0.016395	2.77	0.005606
0.14	0.888660	0.52	0.603064	0.90	0.368120	1.28	0.200545	1.66	0.096914	2.03	0.042357	2.41	0.015953	2.78	0.005436
0.15	0.880765	0.53	0.596112	0.91	0.362823	1.29	0.197051	1.67	0.094919	2.04	0.041350	2.42	0.015521	2.79	0.005271
0.16	0.872881	0.54	0.589197	0.92	0.357573	1.30	0.193601	1.68	0.092957	2.05	0.040364	2.43	0.015099	2.80	0.005110
0.17	0.865010	0.55	0.582319	0.93	0.352371	1.31	0.190196	1.69	0.091028	2.06	0.039399	2.44	0.014687	2.81	0.004954
0.18	0.857153	0.56	0.575479	0.94	0.347218	1.32	0.186835	1.70	0.089131	2.07	0.038452	2.45	0.014286	2.82	0.004802
0.19	0.849309	0.57	0.568678	0.95	0.342112	1.33	0.183518	1.71	0.087266	2.08	0.037526	2.46	0.013894	2.83	0.004655
0.20	0.841481	0.58	0.561915	0.96	0.337055	1.34	0.180245	1.72	0.085432	2.09	0.036618	2.47	0.013511	2.84	0.004511
0.21	0.833668	0.59	0.555191	0.97	0.332046	1.35	0.177016	1.73	0.083630	2.10	0.035729	2.48	0.013138	2.85	0.004372
0.22	0.825871	0.60	0.548506	0.98	0.327086	1.36	0.173830	1.74	0.081859	2.11	0.034858	2.49	0.012774	2.86	0.004236
0.23	0.818092	0.61	0.541862	0.99	0.322174	1.37	0.170687	1.75	0.080118	2.12	0.034006	2.50	0.012419	2.87	0.004105
0.24	0.810330	0.62	0.535258	1.00	0.317311	1.38	0.167587	1.76	0.078408	2.13	0.033172	2.51	0.012073	2.88	0.003977
0.25	0.802587	0.63	0.528695	1.01	0.312495	1.39	0.164529	1.77	0.076727	2.14	0.032355	2.52	0.011735	2.89	0.003852
0.26	0.794864	0.64	0.522173	1.02	0.307728	1.40	0.161513	1.78	0.075076	2.15	0.031555	2.53	0.011406	2.90	0.003732
0.27	0.787160	0.65	0.515692	1.03	0.303010	1.41	0.158540	1.79	0.073454	2.16	0.030773	2.54	0.011085	2.91	0.003614
0.28	0.779478	0.66	0.509254	1.04	0.298340	1.42	0.155608	1.80	0.071861	2.17	0.030007	2.55	0.010772	2.92	0.003500
0.29	0.771816	0.67	0.502858	1.05	0.293718	1.43	0.152717	1.81	0.070296	2.18	0.029257	2.56	0.010467	2.93	0.003390
0.30	0.764177	0.68	0.496504	1.06	0.289145	1.44	0.149867	1.82	0.068759	2.19	0.028524	2.57	0.010170	2.94	0.003282
0.31	0.756561	0.69	0.490194	1.07	0.284619	1.45	0.147059	1.83	0.067250	2.20	0.027807	**2.5758**	**0.010000**	2.95	0.003178
0.32	0.748968	0.70	0.483927	1.08	0.280142	1.46	0.144290	1.84	0.065768	2.21	0.027105	2.58	0.009880	2.96	0.003076
0.33	0.741400	0.71	0.477704	1.09	0.275713	1.47	0.141562	1.85	0.064314	2.22	0.026419	2.59	0.009598	2.97	0.002978
0.34	0.733857	0.72	0.471525	1.10	0.271332	1.48	0.138873	1.86	0.062886	2.23	0.025747	2.60	0.009322	2.98	0.002882
0.35	0.726339	0.73	0.465390	1.11	0.266999	1.49	0.136224	1.87	0.061484	2.24	0.025091	2.61	0.009054	2.99	0.002790
0.36	0.718847	0.74	0.459300	1.12	0.262714	1.50	0.133614	1.88	0.060108	2.25	0.024449	2.62	0.008793	3.00	0.002700