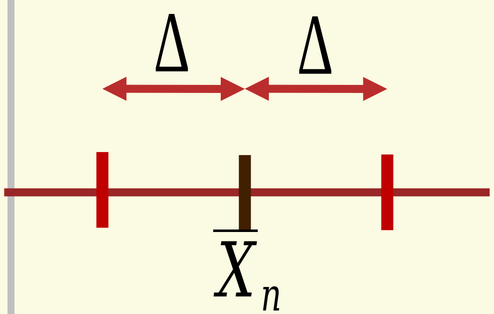
Intervalo de confianza para la media

i.i.d. y se desea estimar



$$\left(\overline{X}_n - \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}, \overline{X}_n + \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}\right)$$

Está claro que no se conoce , pero las fórmulas usan : ¿de dónde sale?

Si no es muy grande, pero las variables son normales, hay una forma de hacerlo …

Distribución t-Student

i.i.d.

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2}$$

$$T = \frac{\overline{X}_n - \mu}{\frac{S}{\sqrt{n}}} \quad t_{n-1}$$

Distribución t-Student con grados de libertad

Distribución t-Student

i.i.d.

$$T = \frac{\overline{X}_n - \mu}{\frac{S}{\sqrt{n}}} \quad t_{n-1}$$

$$E[T] = 0$$

para

para

Simétrica

para

para

Distribución t-Student

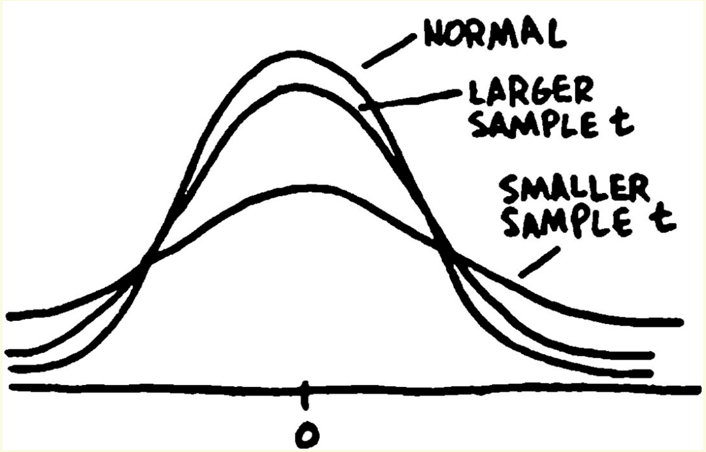
i.i.d.

$$T = \frac{\overline{X}_n - \mu}{\frac{S}{\sqrt{n}}} \quad t_{n-1}$$

$$E[T] = 0$$

$$T \underset{apr\acute{o}x.}{} \mathrm{N}(0,1)$$

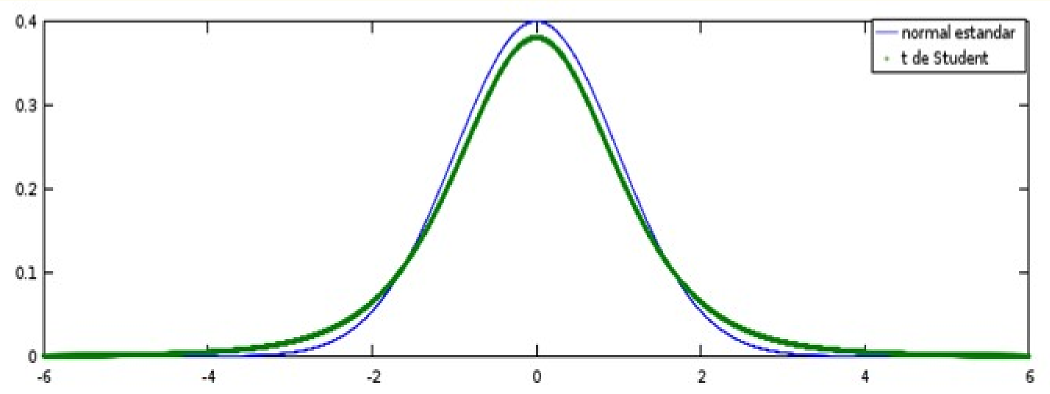
T.C.L.



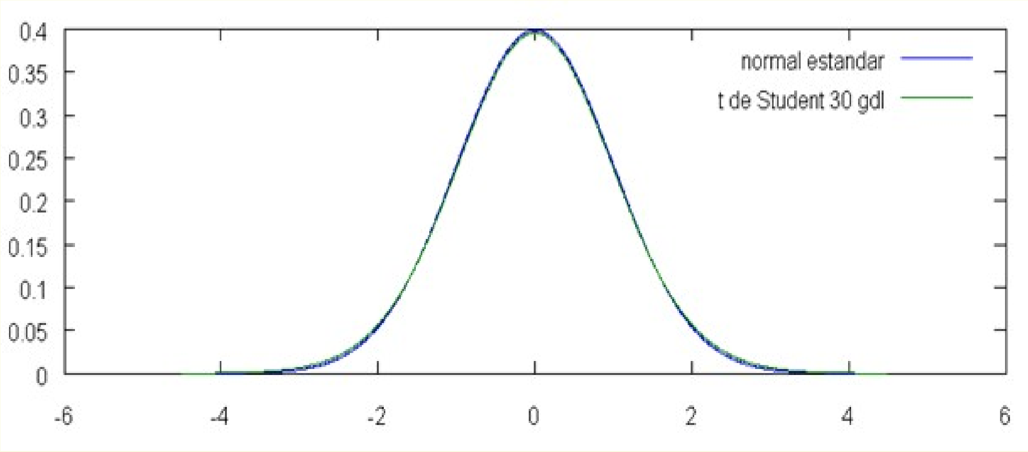
Gonick & Smith, The Cartoon Guide to Statistics

Distribución t-Student

$$T = \frac{\overline{X}_n - \mu}{\frac{S}{\sqrt{n}}} \quad t_{n-1}$$



Comparación de las funciones de densidad de probabilidad de las variables aleatorias normal estándar y t de Student con 5 grados de libertad



Comparación de las funciones de densidad de las variables aleatorias normal estándar y t de Student con 30 grados de libertad

$$T=\frac{\overline{X}_n-\mu}{\frac{S}{\sqrt{n}}} \quad t_{n-1}$$

Distribución t-Student

William Sealy Gosset

William Sealy Gosset (11 de junio de 1876 – 16 de octubre de 1937) fue un estadístico, conocido por su sobrenombre literario ***Student***.

Semblanza [editar]

Nacido en Canterbury, era hijo de Agnes Sealy Vidal y del coronel Frederic Gosset. Asistió a la famosa escuela privada Winchester College antes de estudiar química y matemática en el New College de Oxford. Tras graduarse en 1899, se incorporó a las destilerías Guinness en Dublín.

Guinness era un negocio agroquímico progresista y Gosset podría aplicar sus conocimientos estadísticos tanto a la destilería como a la granja para seleccionar las mejores variedades de cebada. Gosset adquirió ese conocimiento mediante estudio, prueba y error así como pasando dos temporadas durante 1906/7 en el laboratorio bioquímico de Karl Pearson. Gosset y Pearson tenían una buena relación y este último ayudó a Gosset con la matemática de sus artículos. Pearson contribuyó a los artículos de 1908, pero no apreció lo suficiente su importancia. Los artículos se referían a la importancia de las pequeñas muestras para la destilería, mientras que el biólogo disponía normalmente de cientos de observaciones y no veía la urgencia en el desarrollo de métodos basados en unas pocas muestras.

Otro investigador de Guinness había publicado anteriormente un artículo que contenía secretos industriales de la destilería. Para evitar futuras exposiciones de información confidencial, Guinness prohibió a sus empleados la publicación de artículos independientemente de la información que contuviesen. Esto significaba que Gosset no podía publicar su trabajo usando su propio nombre. De ahí el uso de su pseudónimo *Student* en sus publicaciones, para evitar que su empleador lo detectara. Por tanto, su logro más famoso se conoce ahora como la distribución t de Student, que de otra manera hubiera sido la distribución t de Gosset.

Gosset publicó *El error probable de una media* y casi todos sus artículos usando el pseudónimo *Student* en la publicación *Biometrika* creada por Pearson. Sin embargo, fue R.A. Fisher quien apreció la importancia de los trabajos de Gosset sobre muestras pequeñas, tras recibir correspondencia de Gosset en la que le decía *le envío una copia de las Tablas de Student, ¡ya que es la única persona que probablemente las use jamás!*. Fisher creyó que Gosset había efectuado una "revolución lógica".

William Sealy Gosset

Información personal

Nacimiento	13 de junio de 1876 Canterbury (Reino Unido de Gran Bretaña e Irlanda)
Fallecimiento	16 de octubre de 1937 (61 años) Beaconsfield (Reino Unido)

$$T = \frac{\overline{X}_n - \mu}{\frac{S}{\sqrt{n}}} \quad t_{n-1}$$

Distribución t-Student

THE PROBABLE ERROR OF A MEAN

By STUDENT

Introduction

Any experiment may he regarded as forming an individual of a "population" of experiments which might he performed under the same conditions. A series of experiments is a sample drawn from this population.

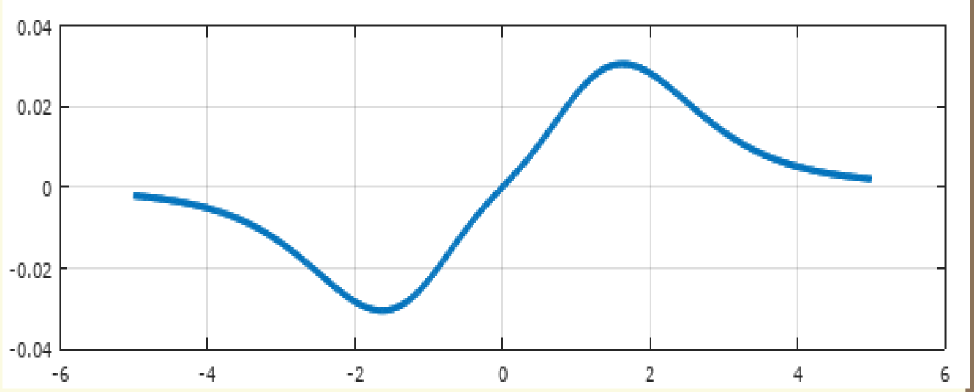
Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a greater number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

If the number of experiments be very large, we may have precise information as to the value of the mean, but if our sample be small, we have two sources of uncertainty: (1) owing to the "error of random sampling" the mean of our series of experiments deviates more or less widely from the mean of the population, and (2) the sample is not sufficiently large to determine what is the law of distribution of individuals. It is usual, however, to assume a normal distribution, because, in a very large number of cases, this gives an approximation so close that a small sample will give no real information as to the manner in which the population deviates from normality: since some law of distribution must he assumed it is better to work with a curve whose area and ordinates are tabled, and whose properties are well known. This assumption is accordingly made in the present paper, so that its conclusions are not strictly applicable to populations known not to be normally distributed; yet it appears probable that

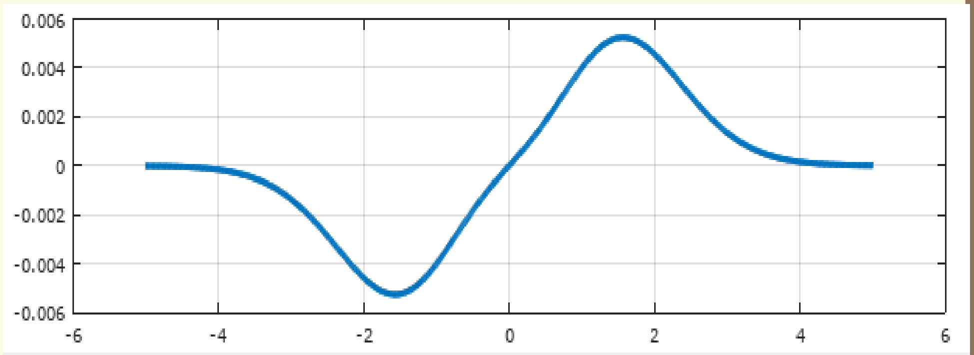
Distribución t-Student

$$T=\frac{\overline{X}_n-\mu}{\frac{S}{\sqrt{n}}} \quad t_{n-1}$$

Diferencia entre las funciones de distribución de las variables aleatorias normal estándar y t de Student con 5 grados de libertad

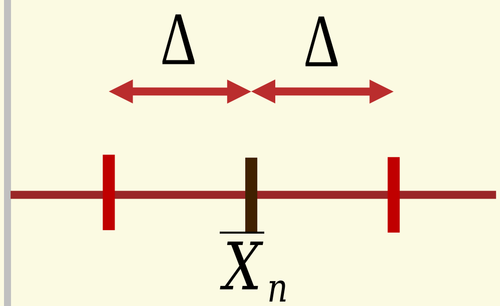


Diferencia entre las funciones de distribución de las variables aleatorias normal estándar y t de Student con 30 grados de libertad



Intervalo de confianza para la media

i.i.d., se desea estimar y se desconoce



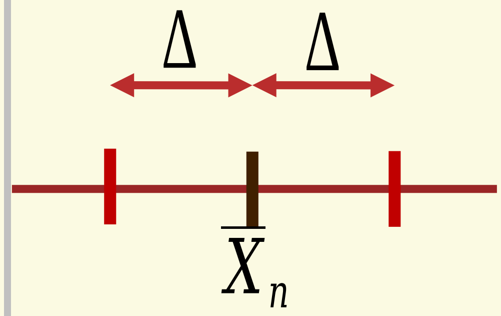
$$\left(\overline{X}_n - \frac{S}{\sqrt{n}} t_{n-1, \frac{1+\gamma}{2}}, \overline{X}_n + \frac{S}{\sqrt{n}} t_{n-1, \frac{1+\gamma}{2}} \right)$$

está definido como el valor tal que si (t-Student con grados de libertad)

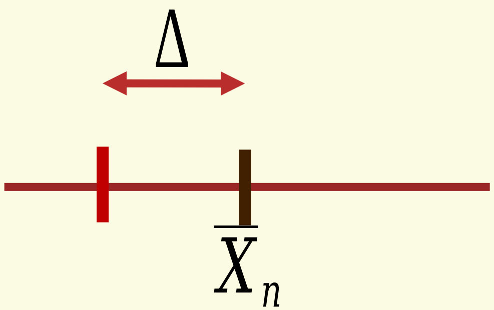
está en tablas o lo devuelven programas de computadora

Un hecho útil: (por ej.,)

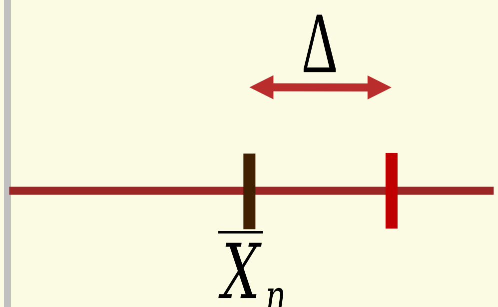
Intervalo de confianza para la media



$$\left(\overline{X}_n - \frac{S}{\sqrt{n}} t_{n-1,\frac{1+\gamma}{2}}, \overline{X}_n + \frac{S}{\sqrt{n}} t_{n-1,\frac{1+\gamma}{2}}\right)$$



$$\left(\overline{X}_n - \frac{S}{\sqrt{n}} t_{n-1,\gamma}, +\infty\right)$$



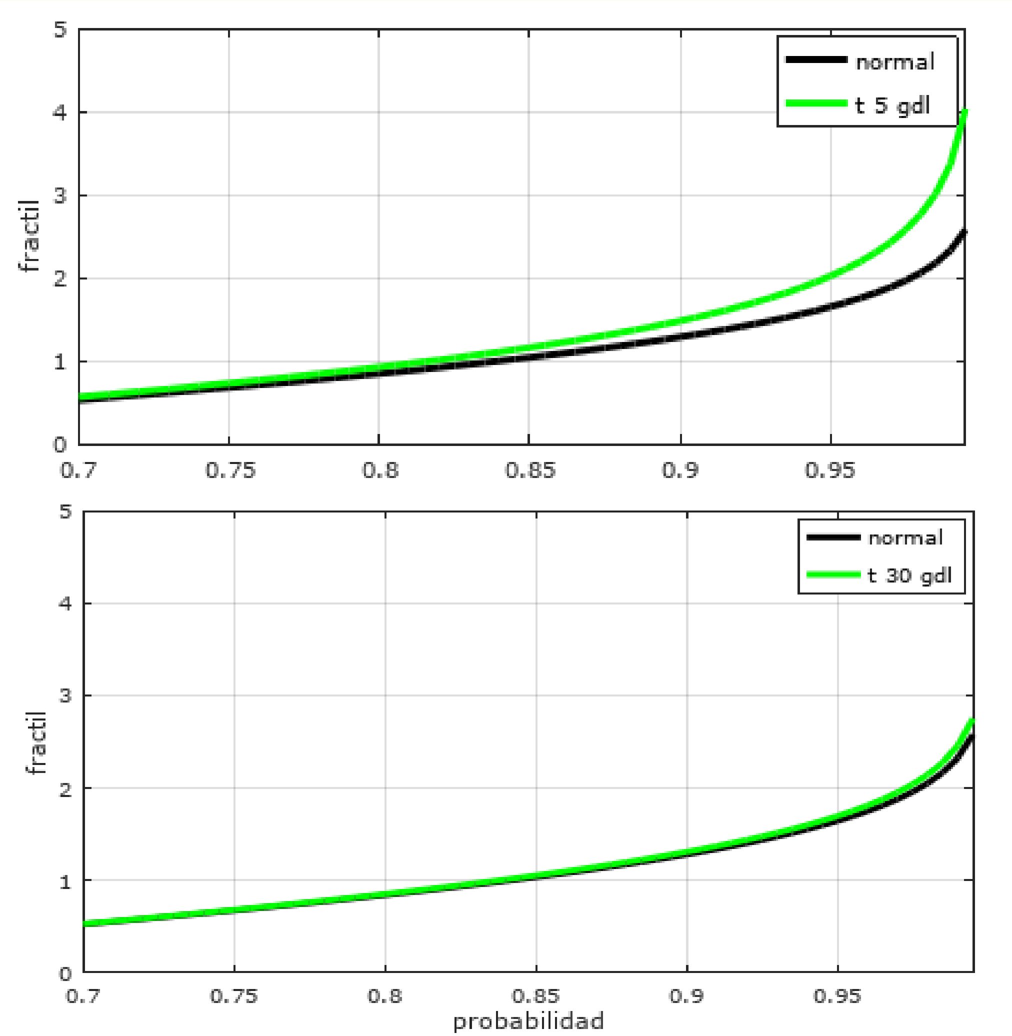
$$\left(-\infty, \overline{X}_n + \frac{S}{\sqrt{n}} t_{n-1,\gamma}\right)$$

Intervalo de confianza para la media

	1	2	3	4	5	6	7	8	9	10
0.900	3.078	1.886	1.638	1.533	1.476	1.440	1.415	1.397	1.383	1.372
0.925	4.165	2.282	1.924	1.778	1.700	1.650	1.617	1.592	1.573	1.559
0.950	6.314	2.920	2.353	2.132	2.015	1.943	1.895	1.860	1.833	1.813
0.975	12.706	4.303	3.182	2.776	2.571	2.447	2.365	2.306	2.262	2.228
0.990	31.821	6.965	4.541	3.747	3.365	3.143	2.998	2.897	2.821	2.764
0.995	63.657	9.925	5.841	4.604	4.032	3.707	3.500	3.355	3.250	3.169

	20	30	40	50	60	70	80	90	100	
0.900	1.325	1.310	1.303	1.299	1.296	1.294	1.292	1.291	1.290	1.282
0.925	1.497	1.477	1.468	1.462	1.458	1.456	1.454	1.452	1.451	1.440
0.950	1.725	1.697	1.684	1.676	1.671	1.667	1.664	1.662	1.660	1.645
0.975	2.086	2.042	2.021	2.009	2.000	1.994	1.990	1.987	1.984	1.960
0.990	2.528	2.457	2.423	2.403	2.390	2.381	2.374	2.369	2.364	2.326
0.995	2.845	2.750	2.705	2.678	2.660	2.648	2.639	2.632	2.626	2.576

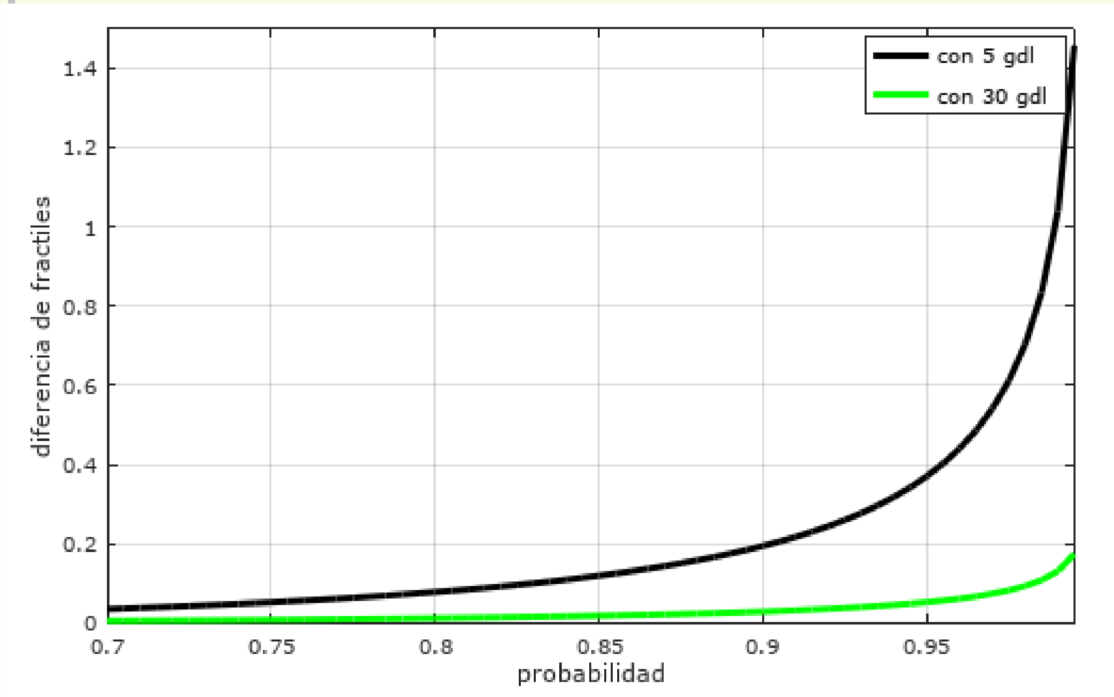
Intervalo de confianza para la media



Comparación de los fractiles de las distribuciones normal estándar y t de Student con 5 y 30 grados de libertad. Las diferencias mas importantes ocurren para valores de p cercanos a 1 y si el número de grados de libertad es pequeño (aún con 30 se nota la diferencia).

Intervalo de confianza para la media

$$t_{v,w} > z_w$$



Diferencia entre los fractiles de las distribuciones normal estándar y t de Student con 5 y 30 grados de libertad. Las diferencias mas importantes ocurren para valores de p cercanos a 1 y si el número de grados de libertad es pequeño (aún con 30 se nota la diferencia).

Ejemplo

Jorge compró 10 paquetes de 1 kg (nominal) de azúcar de la marca MuyDulce. Los pesos de esos paquetes fueron (en gramos):

986   983   981   989   991   986   986   979   997   988

Determine un IC con un nivel de confianza del 90% para el peso medio de un paquete. ¿Qué debe asumir?

Asumimos:

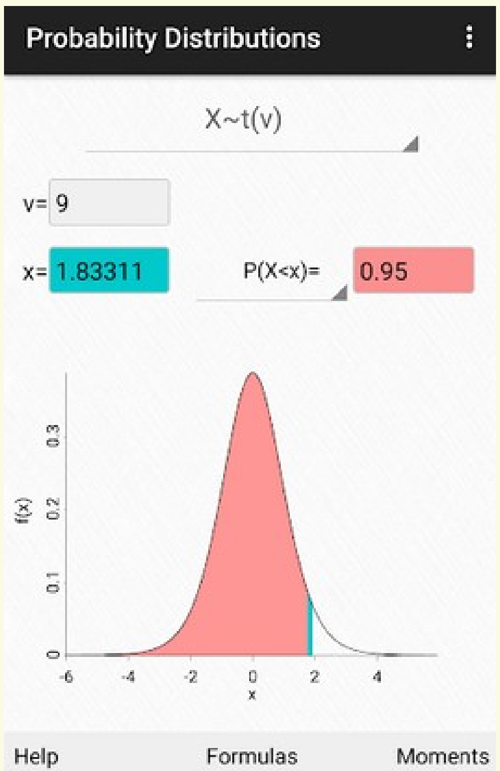
- Muestra aleatoria
- Distribución normal del peso de un paquete tomado al azar

Ejemplo

$$\overline{x}_{10} = 986.60$$

$$s = 5.15$$

$$t_{9,0.95} = 1.833$$



$$\Delta = 5.15\,1.833/\sqrt{10} \approx 3.2$$

$$IC = (983.4, 989.8)$$

Una conclusión interesante: El valor nominal indicado en el paquete (1 Kg) no pertenece al intervalo. Puede sospecharse del fabricante.

Intervalo de confianza para la media

Un error común: No usar el fractil t y reemplazarlo por el fractil normal en el cálculo de la semiamplitud del intervalo de confianza para la media poblacional de una población normal cuando se desconoce la varianza poblacional.

Octave

File Edit Debug Window Help News

Current Directory: C:\Users\paco

Command Window

```
>> mu=12;sigma=1.5;
>> M=mu+sigma*randn(10,1000); %1000 muestras de tamanio 10
>> confianza=0.95;z=1.96;t=2.262;
>> XR=mean(M);S=std(M);
>> ICnormal=[XR'-z*S'/sqrt(10)  XR'+z*S'/sqrt(10)];
>> ICt=[XR'-t*S'/sqrt(10)  XR'+t*S'/sqrt(10)];
>> MU=mu*ones(1000,1);
>> SIz=(MU>ICnormal(:,1)).*(MU<ICnormal(:,2));
>> SIt=(MU>ICt(:,1)).*(MU<ICt(:,2));
>> [mean(SIz) mean(SIt)]
ans =

   0.90200   0.94100
```

Se generan 1000 intervalos de confianza para la media poblacional suponiendo desconocida la varianza poblacional y asumiendo que las muestras provienen de una población normal. Se usan los fractiles t y z del 95%. Se calcula luego la fracción de los intervalos que contienen el verdadero valor de la media

Poco mas del 90% de los intervalos obtenidos con el fractil normal y un poco menos del 95 % de los obtenidos con el fractil t contienen el verdadero valor de la media.