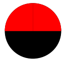
Tests de hipótesis

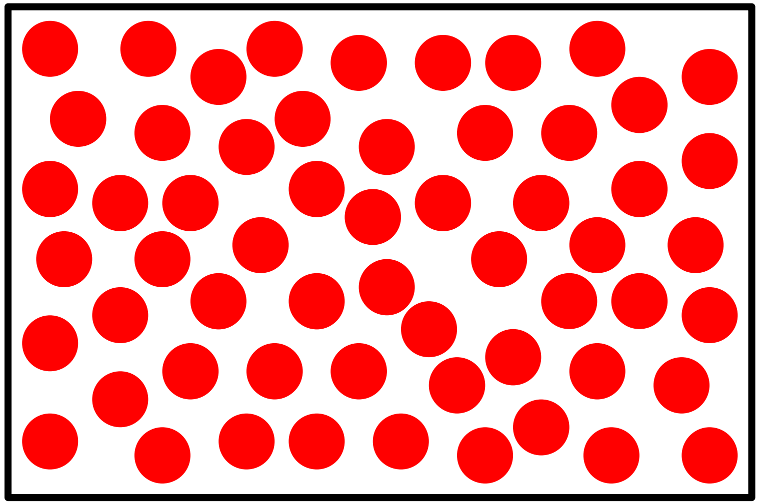
Diego Shalom

Un ejemplo sencillo:

Supongamos que tengo una caja con bolitas rojas.

Encuentro una bolita en el piso y quiero saber si puede o no pertenecer a esta caja.





Sabemos que en la caja, todas las bolitas son rojas.

● ⟶ Estoy re seguro (?) de que esta bolita no se cayó de esa caja.

● ⟶ Ummm… Podría haber salido de ahí. ¿Puedo estar seguro?

No. Tal vez no noté que las de la caja son de goma y esta de madera.
O son más pesadas. **O son iguales, pero no viene de ahí.**

Hipótesis nula (H_0): “La bolita que encontré pertenece a la caja.”

Evalúo las consecuencias de que se cumpla (o no se cumpla H_0)

● $\longrightarrow$ Rechazo la hipótesis nula (“estoy re seguro”)

● $\longrightarrow$ No puedo rechazar la hipótesis nula
(“no hay diferencias significativas”)

En cinco pasos:

1- Hipótesis nula (H_0): "La bolita que encontré pertenece a la caja."

2- Mido: color **rojo** o **negro**

3- Conozco **la distribución** a la que debería pertenecer lo que mido si se cumple **H_0**: Son todas rojas

4- Calculo la **probabilidad** de medir lo que medí, si se cumple **H_0** :

- **negra** → p=0
- **roja** → p=alta

5- Reporto el resultado:

- **negra** → rechazo **H_0** (la bolita no pertenece a la caja).
- **roja** → no hay diferencias significativas. No puedo decir más.
  (La bolita podría pertenecer a la caja)

Repaso

-Muestra vs población

-Normalidad: tests paramétricos – no-paramétricos

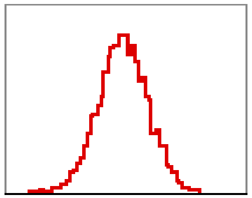
-Media, desvío y error estándar

Población: el universo de datos (desconocido) sobre el que quiero determinar algo.

Muestra: lo que mido, una parte (idealmente aleatoria) de la población.

Normalidad

Distribución normal (gaussiana), tests paramétricos:



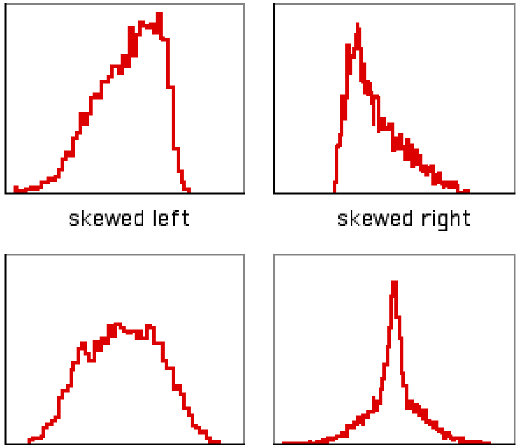
Media: el promedio de la distribución.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Desvío estándar: ancho de la distribución.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Distribución no-normal, tests no-paramétricos



Mediana: separa la mitad de los datos de un lado y la mitad del otro. $Q_2 = P_{50}$

Moda: el valor más repetido, el máximo (no tiene sentido para variables continuas).

Percentil: P_n es el valor de la variable por debajo de la cual se encuentra n% de las observaciones.

Rango intercuartil: . El rango en el que se encuentra la mitad central de los datos, desde el primer cuartil $Q_1 = P_{25}$ hasta el tercer cuartil $Q_3 = P_{75}$.

Muestras de distintos tamaños, de una **población** normal N(m=2.05,s=0.5)

Media: el promedio de la distribución
Tiende a cte al aumentar N. Ancho típico: SEM

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

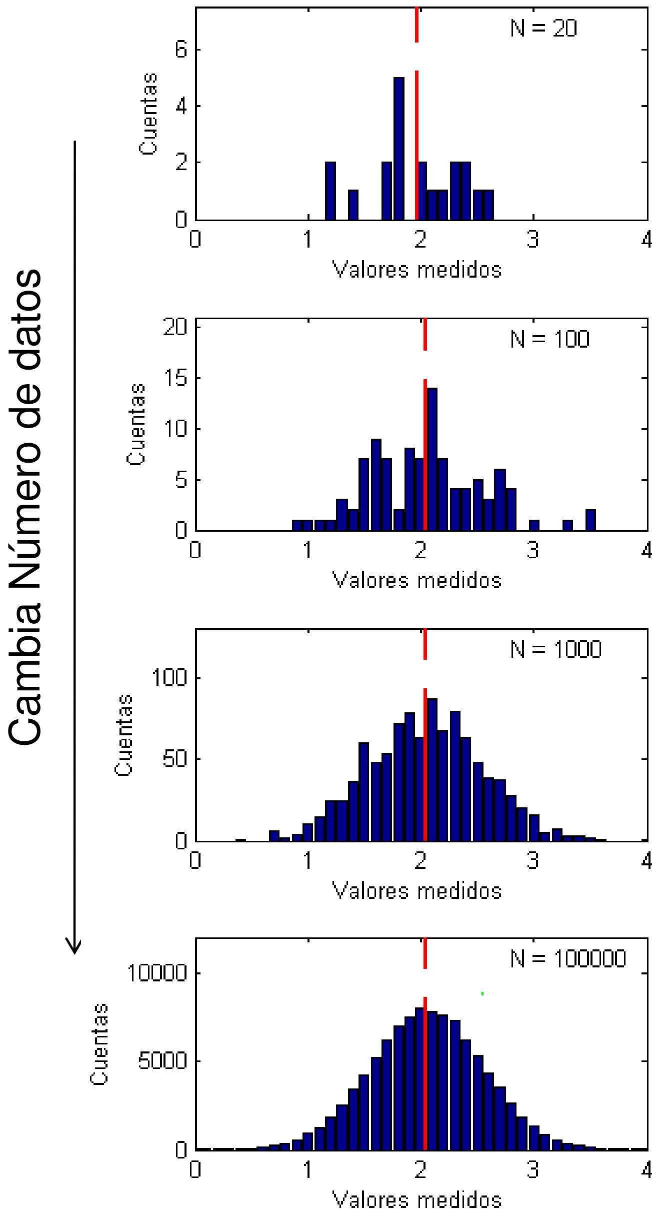
Desvío estándar: el ancho de la distribución. Describe la distribución.
Tiende a cte al aumentar N.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

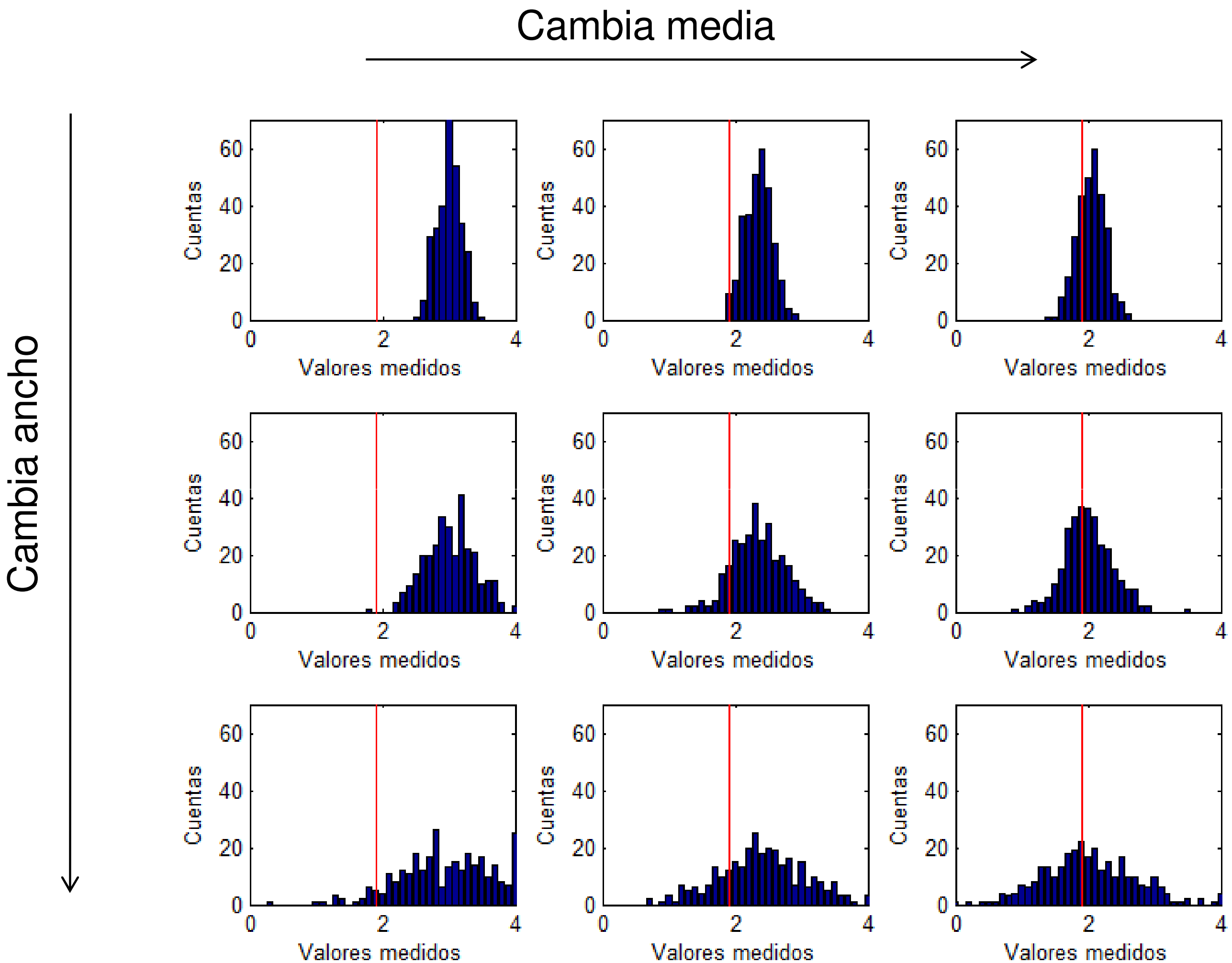
Error estándar de la media (SEM): incerteza de la media.

$$SEM = \frac{\sigma}{\sqrt{N}}$$

Se achica al aumentar N, cada vez la media tiene menos variabilidad.



Diferentes posibilidades, algunas más fáciles y otras más difíciles de decidir.



Un poco más formalmente…

Tests de hipótesis

1- Enunciar una **"Hipótesis nula" (H_0)**. Y una hipótesis alternativa $\mathbf{H_1}$.

2- Tomar una muestra **aleatoria** (medir).

3- Calcular un **estadístico** basado en la muestra.

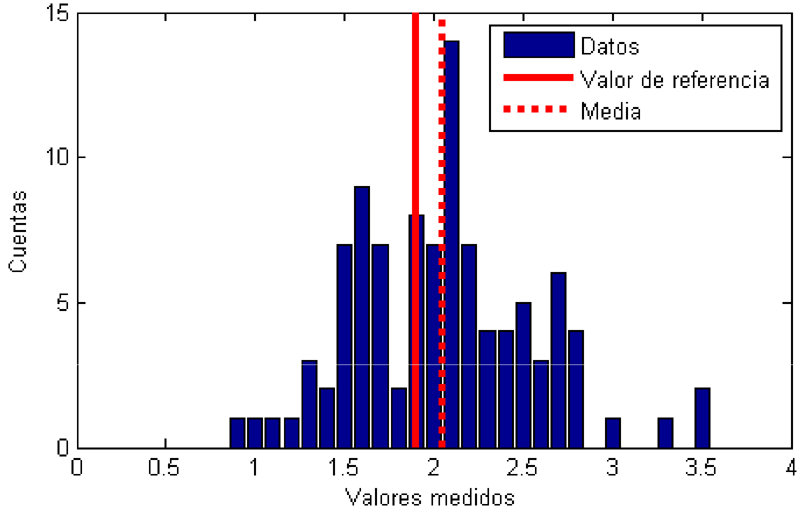
4- Usar el estadístico y sus propiedades para calcular el **p-valor**, la probabilidad de obtener un estadístico al menos tan extremo como el observado, suponiendo que se cumple la hipótesis nula.

5- Elegir un umbral (criterio) para tomar una **decisión** sobre la significancia.

Un ejemplo más específico:

En un experimento mido estos datos, y quiero determinar si su valor medio es o no es un valor determinado.

HISTOGRAMA



Ejemplo: El voltaje medido con la luz apagada es **1.9 (valor de referencia)**. Al prender la luz obtengo **estos valores experimentales (media 2.049)** de voltaje.

ó El tiempo de respuesta de ratas normales es **1.9 (valor de referencia)**. Al darles una droga obtengo **estos valores experimentales (media 2.049)** de tiempo de respuesta.

Quiero saber: ¿Afecta la luz al voltaje?

¿Afecta la droga al tiempo de respuesta?

Comparar la media (2.049) con un valor de referencia (1.9).

1- Enunciar una "Hipótesis nula" (H_0). Y una hipótesis alternativa H_1.

Planteamos la **Hipótesis nula H_0:**

"La droga **no** afecta el tiempo de reacción de las ratas" ó

"La luz **no** produce un cambio en el voltaje medido".

Asumir que esto vale implica suponer que el valor medio que medimos ($\overline{X}$) que medimos viene de una **distribución normal** con media μ_o=**1.9** (y σ=0.5). Y su diferencia (**2.049-1.9**) es debida **puramente al azar**.

La hipótesis alternativa H_1:

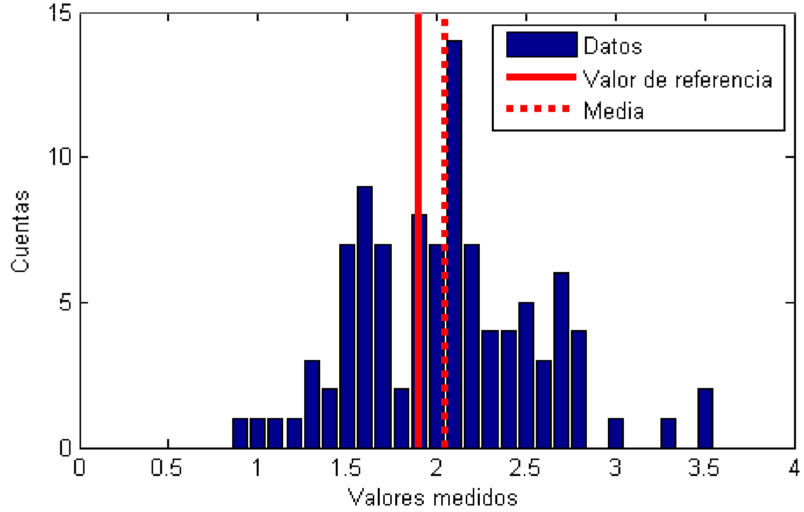
"La droga AUMENTA (o MODIFICA o DISMINUYE) el tiempo de reacción de las ratas" ó

"La luz AUMENTA (o MODIFICA o DISMINUYE) el voltaje medido".

2- Tomar una muestra aleatoria (medir).

Luego medimos:

N: **100**
Media: $\overline{X}$ = **2.049**
Valor de referencia: μ_o**= 1.9**
Desvío Estándar: **σ = 0.5**
Error Estándar de la media: SEM = **0.05**



3- Calcular un estadístico basado en la muestra.

El caso más simple sería usar la diferencia entre el valor medido y el valor de referencia:

$$\bar{X} - \mu_o = 2.049 - 1.9 = 0.149$$

En lugar de eso, lo vamos a normalizar (dividir) por el SEM:

Estadístico:

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{N}} = \frac{\bar{X} - \mu_0}{SEM}$$

Con σ la desviación estándar (conocida)

$$Z = \frac{2.049 - 1.9}{0.05} = 2.98$$

¿Y eso es mucho o poco? Allá vamos.

4- Usar el estadístico y sus propiedades para calcular el p-valor (se puede hacer teórico, acá numéricamente).

Por H_0, supongo que la población tiene una media 0 y desvío estándar 0.5 conocidos. Y que la diferencia que obtuve fue producto del azar.

Armo una muestra aleatoria:

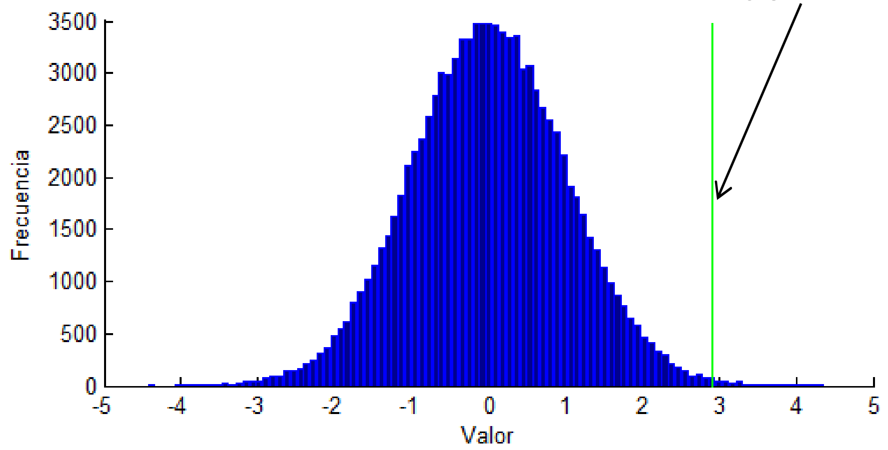
```
>>  N=100;     SD=0.5;
   PRUEBA=randn(N,1)*SD;
   [mean(PRUEBA) std(PRUEBA) sem(PRUEBA) mean(PRUEBA)/sem(PRUEBA)]
ans =      0.0796     0.4744     0.0474     1.6786
```

Notar que la media no es exactamente 0.

Armo **muchas** muestras aleatorias. Y miro su distribución.

```
>> PRUEBAS=randn(N,100000)*SD;
   m_PRUEBAS=mean(PRUEBAS);
   z_PRUEBAS=mean(PRUEBAS)./sem(PRUEBAS);
   hist(z_PRUEBAS,100)
```

El valor medido, nuestro estadístico 2.98



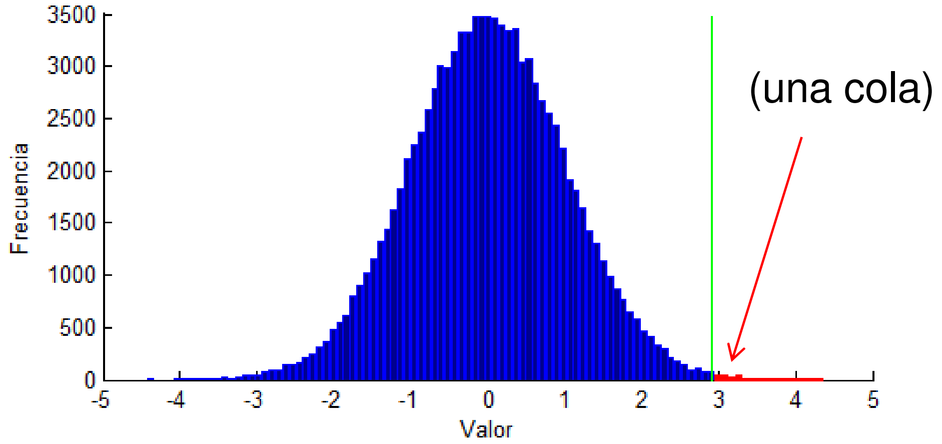
4- Usar el estadístico y sus propiedades para calcular el p-valor

La probabilidad de obtener EXACTAMENTE ese valor es cero.

Puedo cuantificar la probabilidad de obtener un valor así **o mayor**:

```
mean( z_PRUEBAS > 2.98 )
ans =
    0.0014
```

p=0.0014



4- Usar el estadístico y sus propiedades para calcular el p-valor.

La probabilidad de obtener EXACTAMENTE ese valor es cero.

Puedo cuantificar la probabilidad de obtener un valor así **o mayor**:

```
mean( z_PRUEBAS > 2.98 )
ans =
    0.0014
```

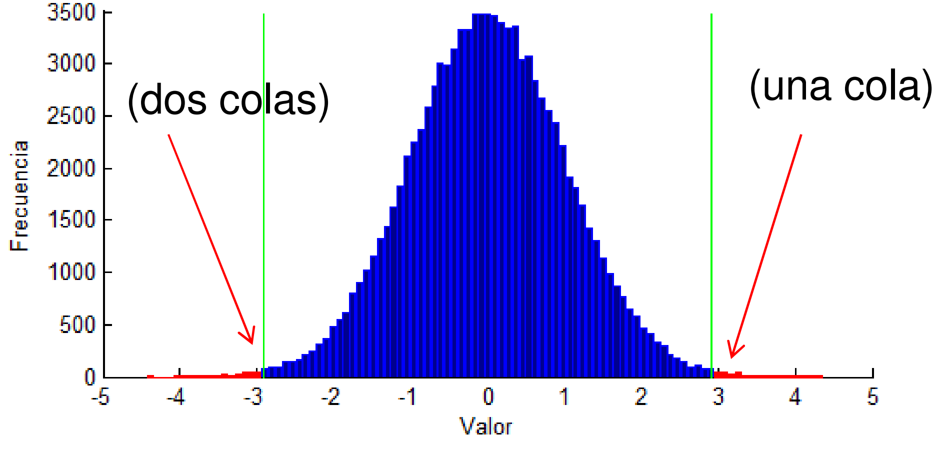
p=0.0014

O así de **extremo** para los dos lados:

```
mean( z_PRUEBAS > 2.98 )+mean( z_PRUEBAS < -2.98 )
ans =
    0.0028
```

p=0.0028

¿Usamos una o dos colas?
Depende de la hipótesis $\mathbf{H_1}$. Si queremos probar si aumenta o si cambia para cualquier lado.



¿Pero qué es p?

Es la probabilidad de obtener un valor tan extremo como el medido, asumiendo la H_0.

En otras palabras,

p es la probabilidad de equivocarnos al rechazar la H_0 (y aceptar H_1) teniendo en cuenta lo medido.

Aquí, **bajo** es “bueno”. Pero, ¿**bajo** comparado con qué?

5- Elegir un umbral (criterio) para tomar una decisión sobre la significancia.

Se elige un **criterio**, un umbral de significancia α (arbitrario).
Y si $p<\alpha$ decimos que el resultado es **"estadísticamente significativo"**.

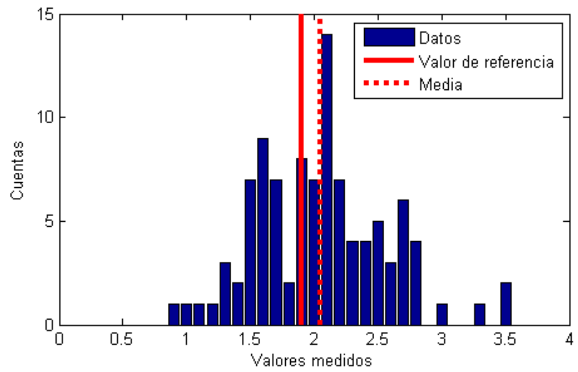
Convencionalmente se usa $\alpha = 0.05 = 5\% = 1/20$.
"Si hiciera el mismo experimento 100 veces, toleraría equivocarme en 5 por puro azar."

En nuestro experimento:

0.0014<0.05 → "**2.049** es significativamente **mayor** que **1.9**".

0.0028<0.05 → "**2.049** es significativamente **distinto** de **1.9**".

CONCLUSIÓN: Aceptaríamos ambas H_1.
"La droga AUMENTA (y MODIFICA) el tiempo de reacción de las ratas" ó
"La luz AUMENTA (y MODIFICA) el voltaje medido".



OTRO EJEMPLO: Test de Student, o t-test:
Es parecido a lo que hicimos, un poco distinto:

1- H_0: Los datos vienen de una distribución normal con media 0 y **desvío desconocido**.
   H_1: La media no es 0.
3- El estadístico es t-stat = (media-refval)/SEM (**con σ medida**)
   Sus propiedades son distintas (**tiene distribución t de Student**, de ahí el nombre del test).
   (en lugar de distribución normal)

```
>> [h p ci s]=ttest(data-1.9)

h =
     1
```

h es 0 o 1, p<α (dos colas y α=0.05 por defecto), acepto o rechazo, hay o no diferencias significativas

```
p =
    0.0044
```

p-valor, la probabilidad de obtener un valor así de extremo

```
ci =
    1.9479
    2.1520
```

El **intervalo de confianza** 95% para la media (¿agarra o no 1.9?)

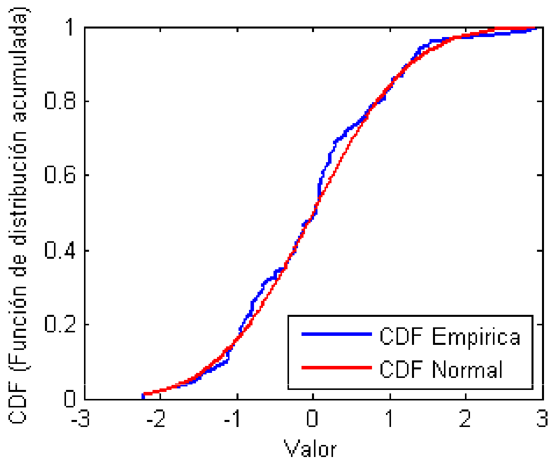
```
s =
    tstat: 2.9148
       df: 99
       sd: 0.5144
```

tstat: el estadístico del t-test
df: grados de libertad
sd: desvío estandar

EJEMPLO 2:

Normalidad: Kolmogorov-Smirnov

1- H_0: Los datos vienen de una distribución normal.
   H_1: Los datos NO vienen de una distribución normal.



Calcula la CDF y la compara con la teórica. Calcula la probabilidad de obtener una distribución así, suponiendo Normalidad.

```
>> [h p ksstat cv]=kstest(x)
h =
     0
```

No tengo evidencias para rechazar H_0

```
p =
    0.6064
```

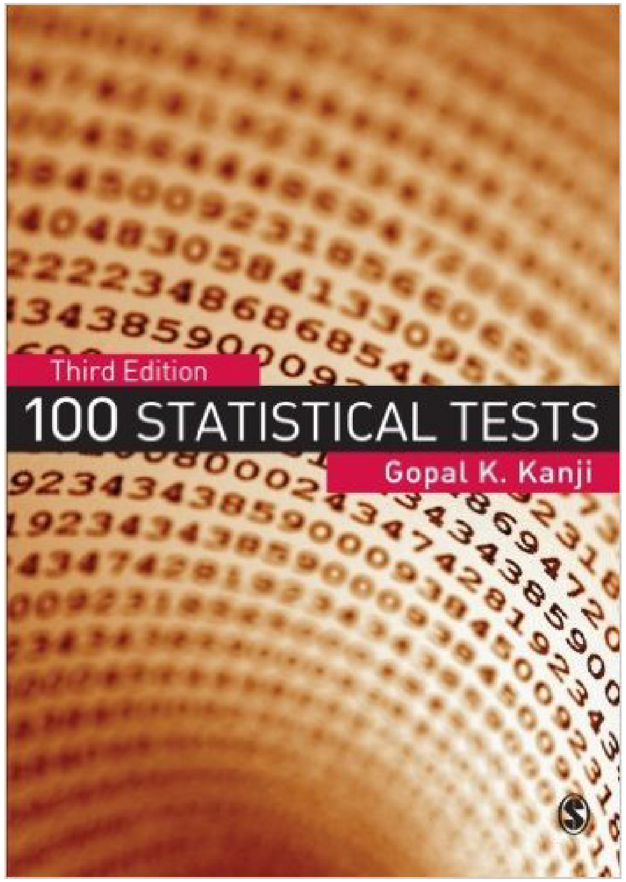
La probabilidad de equivocarme al rechazar H_0 es alta.

```
ksstat =
    0.0746
```

El estadístico.

```
cv =
    0.1340
```

El cutoff value del estadístico para p=0.05, para determinar significancia.



Test 1 Z-test for a population mean (variance known)

Object

To investigate the significance of the difference between an assumed population mean μ_0 and a sample mean $\bar{x}$.

Limitations

1. It is necessary that the population variance σ^2 is known. (If σ^2 is not known, see the t-test for a population mean (Test 7).)
2. The test is accurate if the population is normally distributed. If the population is not normal, the test will still give an approximate guide.

Method

From a population with assumed mean μ_0 and known variance σ^2, a random sample of size n is taken and the sample mean $\bar{x}$ calculated. The test statistic

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

may be compared with the standard normal distribution using either a one- or two-tailed test, with critical region of size α.

Example

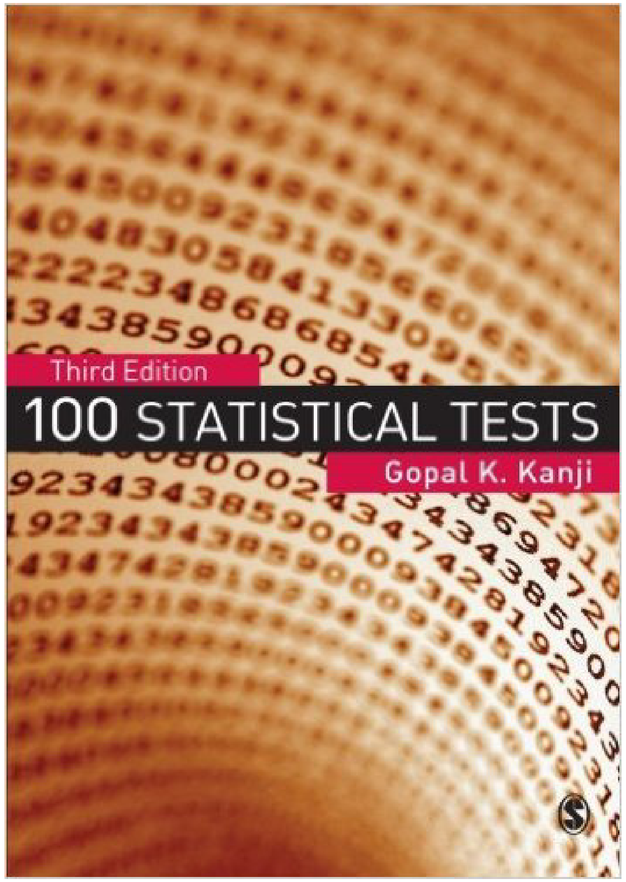
For a particular range of cosmetics a filling process is set to fill tubs of face powder with 4 gm on average and standard deviation 1 gm. A quality inspector takes a random sample of nine tubs and weighs the powder in each. The average weight of powder is 4.6 gm. What can be said about the filling process?

A two-tailed test is used if we are concerned about over- and under-filling.

In this $Z = 1.8$ and our acceptance range is $-1.96 < Z < 1.96$, so we do not reject the null hypothesis. That is, there is no reason to suggest, for this sample, that the filling process is not running on target.

On the other hand if we are only concerned about over-filling of the cosmetic then a one-tailed test is appropriate. The acceptance region is now $Z < 1.645$. Notice that we have fixed our probability, which determines our acceptance or rejection of the null hypothesis, at 0.05 (or 10 per cent) whether the test is one- or two-tailed. So now we reject the null hypothesis and can reasonably suspect that we are over-filling the tubs with cosmetic.

Quality control inspectors would normally take regular small samples to detect the departure of a process from its target, but the basis of this process is essentially that suggested above.



EXAMPLES OF TEST PROCEDURES

Test 1 ***Z*-test for a population mean (variance known)**

Hypotheses and alternatives

1. H_0: $\mu = \mu_0$
   H_1: $\mu \neq \mu_0$
2. H_0: $\mu = \mu_0$
   H_1: $\mu > \mu_0$

Test statistics

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

n is sample size
$\bar{x}$ is sample mean
σ is population standard deviation

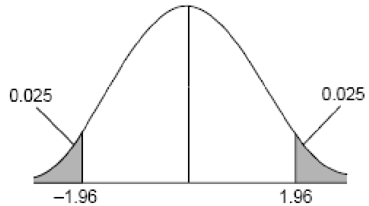
When used

When the population variance σ^2 is known and the population distribution is normal.

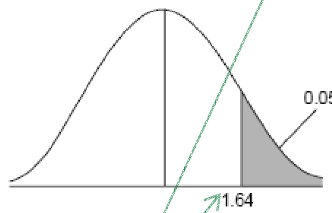
Critical region

Using $\alpha = 0.05$ [see Table 1]

1.



2.

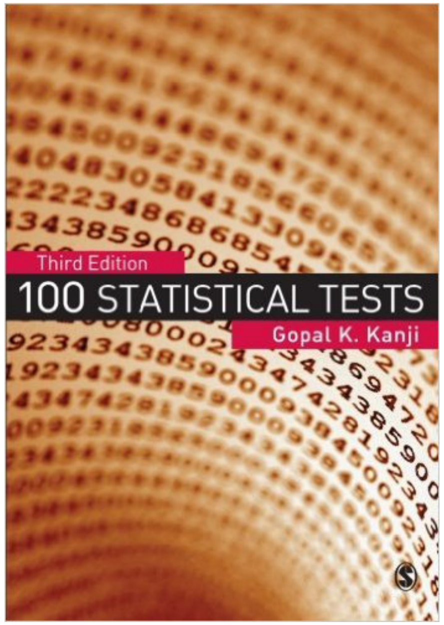


Data

H_0: $\mu_0 = 4.0$
$n = 9, \bar{x} = 4.6$
$\sigma = 1.0$
$\therefore\ Z = 1.8$

Conclusion

1. Do not reject H_0 [see Table 1].
2. Reject H_0



Test 1 *Z*-test for a population mean (variance known)

Object

To investigate the significance of the difference between an assumed population mean μ_0 and a sample mean $\bar{x}$.

Test 7 *t*-test for a population mean (variance unknown)

Object

To investigate the significance of the difference between an assumed population mean μ_0 and a sample mean $\bar{x}$.

Test 35 The Kolmogorov–Smirnov test for goodness of fit

Object

To investigate the significance of the difference between an observed distribution and a specified population distribution.

Hay muchos tests, para preguntarles distintas cosas a los datos:
-Tipos de datos
-Tipos de preguntas

	Test numbers		
For linear data	1 sample	2 samples	*K* samples
Parametric classical tests			
for central tendency	1, 7, 19	2, 3, 8, 9, 10, 18	22, 26, 27, 28, 29, 30, 77, 78, 79, 80, 87
for proportion	4	5, 6, 25	–
for variability	15, 21, 24, 34	16, 17	31, 32, 86
for distribution functions	20, 33, 75, 88, 89, 94	–	76
for association	11, 12, 13, 81, 82	14, 23, 84, 92	85
for probability	83	–	–
Parametric tests			
for distribution function	35, 37	36, 39, 40	38, 41, 42, 43, 44
Distribution-free tests			
for central tendency	45, 47	46, 48, 50, 52, 93	51, 54, 55, 56, 57
for variability	–	53	–
for distribution functions	–	49	–
for association	58, 59	–	72, 73, 74
for randomness	63, 64, 65, 66, 67, 69, 70, 71	68	–
Sequential tests			
central tendency	60, 90	–	–
variability	61	–	–
for proportion	62	–	–
for ratio	91	–	–

	Test numbers		
For circular data	1 sample	2 samples	*K* samples
Parametric tests			
for randomness	95	–	–
for distribution function	96	–	–
for central tendency	–	97, 98	–
for variability	–	99	100