

## Distribución de la media o promedio muestral

a) Las variables  $X_1, \dots, X_n$  son todas variables aleatorias normales independientes con esperanza  $\mu$  y varianza  $\sigma^2$ .

En ese caso 
$$\bar{X} \sim N(\mu, \sigma^2 / n)$$

Supongamos que la capacidad vital forzada (CVF) de hombres jóvenes y sanos sigue una distribución aproximadamente normal con media  $\mu = 4.8$  litros y desviación estándar  $\sigma = 0.5$ .

Se toma una muestra de 16 hombres dentro de esta población, en este caso  $\bar{X} \sim N(4.8, 0.5^2 / 16)$  es decir que  $\bar{X} \sim N(4.8, 0.125^2)$

b) Las variables  $X_1, \dots, X_n$  son todas variables aleatorias independientes con esperanza  $\mu$  y varianza  $\sigma^2$  pero con distribución que no es necesariamente normal.

En el caso de variables normales, el promedio muestral resulta también con distribución normal, repitiendo la misma media y con una varianza igual a la de las variables que se promedian pero dividida por  $n$ , el tamaño de la muestra.

En el caso b) no es cierto que si promediamos variables con determinada distribución, dicho promedio va a tener la misma distribución que las variables consideradas.

## TEOREMA CENTRAL DEL LÍMITE

Sea  $X$  una variable con media  $\mu$  y varianza  $\sigma^2$ , con cualquier distribución. Cuando  $n$  es suficientemente grande, la distribución de la media muestral ( $\bar{X}$ ) es aproximadamente normal con media  $\mu$  y desviación estándar  $\sigma / \sqrt{n}$ .

Es decir, cuando  $n$  es suficientemente grande

$$\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right),$$

o lo que es lo mismo,

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \approx N(0, 1)$$

( $\approx$  se lee "tiene distribución aproximada").

*Comentarios.*

1. El Teorema Central de Límite (TCL) supone que las observaciones son independientes, es decir que la muestra es aleatoria.
2. ¿Cuándo consideramos  $n$  *suficientemente grande* para que valga la aproximación normal al histograma de los promedios?

Depende de la forma de la distribución de la variable. Si ésta es razonablemente simétrica, la aproximación vale incluso para valores muy pequeños de  $n$ . Si la distribución es fuertemente asimétrica o con muchos datos outliers, es necesario un  $n$  muy grande para que el histograma de los promedios pueda ser aproximado por la curva normal.

En general se considera que  $n=30$  es un tamaño suficientemente grande.

3. En la práctica uno usualmente toma una ÚNICA muestra o realiza un único experimento y no lo repite. Sin embargo, la idea de que podríamos repetirlo (con sólo tener suficiente tiempo y dinero) y que esto resultaría en muestras diferentes de la actual, constituye la base para interpretar la variabilidad del estadístico y es un paso fundamental en el razonamiento estadístico. La inferencia se basa en una pregunta simple: ¿Qué hubiera ocurrido si la muestra o el experimento hubieran sido repetidos muchas veces?

## ESTIMACIÓN

Hasta ahora hemos aprendido formas de caracterizar los datos, resumirlos y describirlos. Nuestro objetivo ahora será “trascender” los datos y obtener a partir de ellos conclusiones acerca de la población a partir de la cual éstos fueron obtenidos.

***Al analizar datos con propósitos de investigación nuestra intención es extrapolar los hallazgos de una muestra a toda la población. Aunque, desde el punto de vista clínico cada individuo interesa, el objetivo de la investigación es resumir la experiencia y extraer conclusiones generales.***

***Usaremos resultados obtenidos en la muestra como estimación de lo que ocurre en la población.***

Recordemos,

**PARÁMETRO**  $\Rightarrow$  es un número que describe a una población

**ESTADÍSTICO**  $\Rightarrow$  es cualquier variable obtenida a partir de las variables de una muestra.

Una de las armas más importantes que provee la Estadística es la posibilidad de obtener conclusiones generales a partir de un conjunto limitado, pero representativo de datos. Cuando inferimos no tenemos garantía de que la conclusión que obtenemos sea exactamente correcta. Sin embargo, la estadística permite cuantificar el error asociado a la estimación.

En lo que sigue, centraremos nuestra atención en el problema de estimar la media poblacional  $\mu$ , aunque las ideas que presentaremos son similares a las que se utilizan para realizar inferencia sobre cualquier otro parámetro poblacional.

## DISTRIBUCIÓN DE MUESTREO DE UN ESTADÍSTICO.

Distintas muestras tomadas desde la misma población (aún cuando tengan el mismo tamaño y sean obtenidas con los mismos criterios) producirán distintos valores del estadístico de interés, es decir, nuestra estimación dependerá de la muestra que hayamos seleccionado. Por lo tanto, una pregunta clave a responder es:

*¿Cuán variables son los valores del estadístico de muestra en muestra?*

Las muestras que obtenemos son aleatorias, en consecuencia, el valor del estadístico que nos interesa (en este ejemplo la media muestral) también es aleatorio. Por lo tanto, *la media muestral es una variable aleatoria* (varía de muestra en muestra) y tiene asociada una distribución de probabilidades a la que denominamos **DISTRIBUCIÓN DE MUESTREO DE LA MEDIA**.

Hemos visto que la distribución de  $\bar{X}$  es:

1.  $N(\mu, \sigma^2/n)$ , si  $X_1, \dots, X_n$  son variables aleatorias independientes con distribución  $N(\mu, \sigma^2)$ .
2. **Aproximadamente**  $N(\mu, \sigma^2/n)$ , si  $X_1, \dots, X_n$  son variables aleatorias independientes con la misma distribución y  $n$  es suficientemente grande.

En consecuencia el desvío estándar de la distribución de la media muestral, denominado **ERROR ESTÁNDAR DE LA MEDIA** es

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

*Nota:* El error estándar de la media nos indica cuán variables son las medias obtenidas en distintas muestras (tamaño  $n$ ), mientras que el desvío estándar  $\sigma$  mide la variabilidad de las observaciones poblacionales.

## INTERVALOS DE CONFIANZA

Nos concentraremos ahora en temas de inferencia estadística. Dos de las áreas sobresalientes de la inferencia son:

- Estimación.
- Test de Hipótesis.

Ambos tipos de inferencia están basados en la distribución de muestreo del estadístico involucrado. Es decir, se basan en probabilidades que establecen *qué hubiese ocurrido si hubiésemos tomado muchas muestras del mismo tamaño*.

En el resto de este curso presentaremos algunos métodos de la inferencia estadística para datos numéricos y describiremos el razonamiento usado en su construcción.

Presentaremos la noción de intervalo de confianza construyendo un intervalo de confianza para la media poblacional.

### INTERVALO DE CONFIANZA PARA LA MEDIA $\mu$ DE UNA POBLACIÓN

La media muestral  $\bar{X}$  es un estimador natural de la media poblacional  $\mu$ .

Decimos que  $\bar{X}$  es un estimador *insesgado* de  $\mu$ , esto implica que no hay tendencia sistemática a subestimar o sobreestimar el verdadero valor de  $\mu$ , sin embargo, la estimación producida en una única muestra, en general, no coincidirá con el verdadero valor.

Entonces, para ser verdaderamente informativa, la inferencia acerca de un parámetro debería indicar la probable exactitud de la estimación. Esto es, debería describir cuán cerca del verdadero parámetro esperamos que se encuentre nuestra estimación.

La exactitud de un estimador puntual depende de las características de su distribución de muestreo, puesto que es la distribución de muestreo la que nos indica la probabilidad de que el estadístico caiga a cierta distancia del parámetro.

Un *intervalo de confianza* para un parámetro es un rango de valores dentro del cual se cree que se encuentra el parámetro. La probabilidad de que un intervalo de confianza contenga al parámetro se denomina *coeficiente de confianza*, y se elige como un número cercano a 1, tal como 0.95 o 0.99.

### Intervalo de confianza para $\mu$ cuando $n$ es grande

Usamos la media muestral ( $\bar{X}$ ) y el desvío estándar muestral ( $s$ ) como estimadores de  $\mu$  y de  $\sigma$  (media y desvío estándar poblacionales).

Sabemos que si  $n$  es grande, la distribución de muestreo de  $\bar{X}$  es aproximadamente

normal, entonces, con alta probabilidad (digamos 95%) la media muestral caerá dentro de dos errores estándar de la media y casi con seguridad se encontrará a menos de tres errores estándar.

El error estándar nos ayuda a determinar la exactitud del estimador, mientras menor sea éste, más exacta será nuestra estimación.

Ahora bien, el 95% de las muestras que vienen de una población con media  $\mu$ , producirán valores de  $\bar{X}$  que se encuentran entre

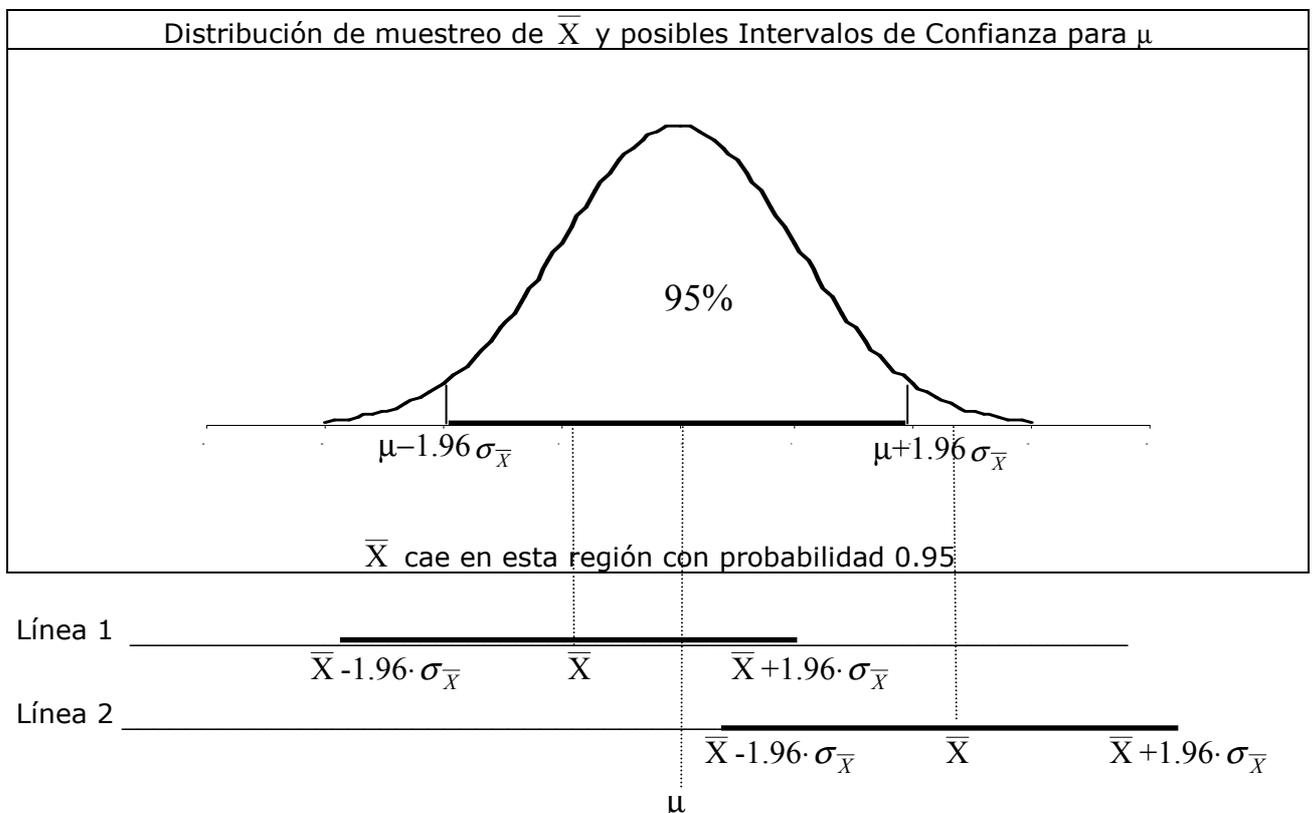
$$\mu - 1.96 \cdot \sigma_{\bar{X}} \text{ y } \mu + 1.96 \cdot \sigma_{\bar{X}}$$

Una vez que la muestra ha sido seleccionada, si  $\bar{X}$  efectivamente está a menos de  $1.96 \cdot \sigma_{\bar{X}}$  de distancia de  $\mu$ , entonces, el intervalo

$$(\bar{X} - 1.96 \cdot \sigma_{\bar{X}}, \bar{X} + 1.96 \cdot \sigma_{\bar{X}})$$

contiene a  $\mu$ . En otras palabras, con probabilidad 0.95 ocurren valores de  $\bar{X}$  tales que el intervalo  $\bar{X} \pm 1.96 \cdot \sigma_{\bar{X}}$  contiene a la media poblacional  $\mu$ .

Por otro lado, la probabilidad de que  $\bar{X}$  se encuentre más allá de  $1.96 \sigma_{\bar{X}}$  es 0.05. Si esto ocurre, el intervalo  $\bar{X} \pm 1.96 \cdot \sigma_{\bar{X}}$  no contiene a  $\mu$ .



En la línea 1 vemos un intervalo que incluye al parámetro  $\mu$ , en cambio el intervalo de la línea 2 no incluye al parámetro  $\mu$ .

El intervalo  $\bar{X} \pm 1.96 \cdot \sigma_{\bar{X}}$  es una estimación por intervalo de la media poblacional  $\mu$ , con coeficiente de confianza 0.95, y se lo llama *intervalo del 95% de confianza*.

Desgraciadamente, el valor de  $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$  en la expresión anterior es desconocido, ya que la desviación estándar poblacional  $\sigma$  es un parámetro desconocido. Si  $n$  es grande, una buena aproximación para  $\sigma_{\bar{X}}$  se obtiene sustituyendo  $\sigma$  por la desviación estándar muestral ( $s$ ).

Entonces,

$$ESM = \frac{s}{\sqrt{n}}$$

estima el verdadero error estándar de la media  $\sigma_{\bar{X}}$ .

Un intervalo del 95% confianza para  $\mu$  cuando  $n$  es grande está dado por:

$$\bar{X} \pm 1.96 \frac{s}{\sqrt{n}}$$

Veamos un ejemplo.

Se midió la presión sistólica en 100 estudiantes de una universidad. La media de la muestra, ( $\bar{X}$ ), resultó 123.4 mmHg y la desviación estándar de la muestra, ( $s$ ), resultó 14 mmHg. Consideramos esta muestra representativa de otras poblaciones (otra universidad, otro año) pero sabemos que la media de esta muestra no va a coincidir exactamente con la de la población. De todos modos como el tamaño de la muestra es grande esperaríamos no estar muy lejos del valor real de nuestro parámetro de interés  $\mu$ .

Obtengamos el intervalo de 95% de confianza:

$$123.4 \pm 1.96 * 14 / \sqrt{100}$$

De manera que el intervalo queda

$$(120.66, 126.14)$$

### *¿cómo interpretamos este intervalo?*

Notemos que no tenemos ninguna duda respecto a la media de la muestra, es decir tenemos un 100% de probabilidad de que la media de la muestra esté contenida en el intervalo, por definición el intervalo construido siempre va a estar centrado en dicha

media. Con respecto a la media poblacional  $\mu$  podemos decir que confiamos en un 95% que este intervalo contiene dicha media. Esto significa que si construyéramos muchos intervalos como este, de la misma forma, y basados en muestras del mismo tamaño, aproximadamente el 95% de dichos intervalos contendrían a  $\mu$  y sólo un 5% lo dejaría fuera del intervalo.

#### Comentarios

- El intervalo que hemos construido vale sólo cuando la muestra es grande, caso en el cual la distribución de la media muestral es aproximadamente normal.
- Debemos tener presente que el error asociado a este intervalo es sólo error de muestreo. Otros errores relevantes para un parámetro, incluyen aquellos asociados a no respuesta, datos faltantes, error de medición, etc.

En la práctica, sólo seleccionamos UNA ÚNICA muestra de tamaño  $n$  y construimos un intervalo de confianza usando estos datos. No sabemos si nuestro intervalo efectivamente contiene a  $\mu$ . Una forma de estar más tranquilos es aumentar el nivel de confianza.

#### *¿Qué efecto tiene aumentar el nivel de confianza?*

Aumentar el nivel de confianza produce un intervalo de mayor longitud. A mayor confianza menor precisión en nuestra estimación. Veremos como construir intervalos con distintos coeficientes de confianza.

Supongamos que nos interesa construir un intervalo con un nivel de confianza cercano a 1, digamos  $1 - \alpha$  (donde  $\alpha$  es un número pequeño cercano a 0, generalmente se elige  $\alpha < 0.10$ ).

La forma general del intervalo de confianza de nivel  $1 - \alpha$  para  $\mu$ , cuando la muestra es grande es:

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{cuando se conoce } \sigma, \text{ ó}$$
$$\bar{X} \pm z_{\alpha/2} \frac{S}{\sqrt{n}} \quad \text{cuando se desconoce } \sigma,$$

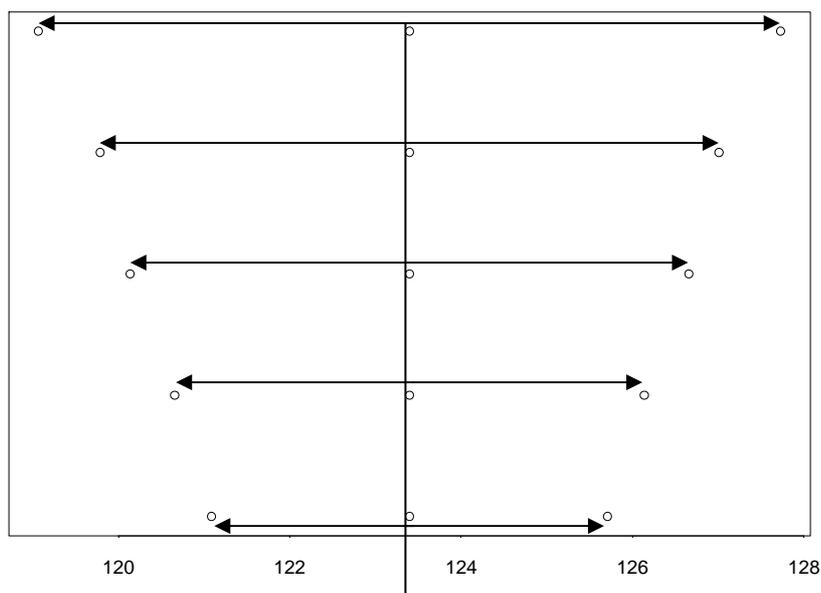
donde  $z_{\alpha/2}$  es el percentil de la distribución  $N(0, 1)$  que deja a su derecha un área  $\alpha/2$  (es decir entre  $-z_{\alpha/2}$  y  $z_{\alpha/2}$  queda encerrada un área  $1 - \alpha$ ).

Confianza	$Z_{\alpha/2}$
90%	1.65
95%	1.96
98%	2.33
99%	2.58
99.9%	3.09

A medida que aumentamos la confianza, aumenta la amplitud del intervalo. Para estar más seguros de que nuestro intervalo contiene al parámetro debemos sacrificar precisión en la estimación. El decidir con qué confianza trabajar, en general implica una solución de compromiso entre la precisión y la confianza deseadas. Esta es la razón por la cual no se usan coeficientes de confianza del 99.9999%, ya que resultarían muy seguros, pero en general el intervalo será tan amplio que no nos dirá nada acerca del valor de la media poblacional.

Veamos cual es el efecto de modificar el nivel de confianza en el ejemplo de la presión sistólica.

Nivel de confianza	Límite inferior	Límite superior
90%	121.09	125.71
95%	120.66	126.14
98%	120.14	126.66
99%	119.79	127.01
99.9%	119.07	127.73



Un modo de reducir la amplitud del intervalo para un coeficiente de confianza determinado, es aumentar el número de observaciones.

### ¿Cómo elegir el tamaño de muestra?

Recordemos la expresión para el intervalo de confianza de nivel  $1-\alpha$ :

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

La longitud del intervalo es  $2 z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$  y depende de:

- del nivel de confianza ( $\uparrow$  confianza  $\uparrow z_{\alpha/2}$   $\uparrow$  longitud).
- de la desviación estándar poblacional ( $\uparrow \sigma$   $\uparrow$  longitud).
- del tamaño de muestra ( $\uparrow n$   $\downarrow$  longitud).

La desviación estándar de la variable bajo estudio es una característica inherente a la población. Mientras que, el coeficiente de confianza y el tamaño de muestra los define el investigador. Por lo tanto, para un nivel de confianza dado, es posible controlar la precisión de la estimación (amplitud del intervalo de confianza) a través del tamaño de muestra.

En nuestro ejemplo, la longitud del intervalo es  $2 z_{\alpha/2} \frac{S}{\sqrt{n}}$ , que con un nivel de 95% de confianza es igual a  $2*1.96*14/\sqrt{10}=5.49$ . Si quisiéramos lograr un intervalo del mismo nivel de confianza pero cuya longitud fuera a lo sumo 4, ¿qué tamaño de muestra deberíamos usar?

Queremos que  $2*1.96*14/\sqrt{n} \leq 2$ , lo que es equivalente a pedir que

$$n \geq \left( \frac{2*1.96*14}{2} \right)^2 = 753$$

Por lo tanto, con una muestra de 753 estudiantes podríamos haber obtenido un intervalo de la longitud deseada.

Genéricamente,

el tamaño de muestra necesario para obtener un intervalo de nivel

$1 - \alpha$  de confianza para  $\mu$ , cuya longitud total no supere  $L_0$  es:

$$n \geq \left( \frac{2 z_{\alpha/2} \sigma}{L_0} \right)^2$$

Si no conocemos  $\sigma$ , tomamos primero una muestra piloto, obtenemos un estimador de  $\sigma$ , y a partir de allí obtenemos el tamaño de muestra adecuado para lograr la longitud deseada.

Otro ejemplo

Interesa construir un intervalo del 98% de confianza para la media del nivel de colesterol de mujeres que inmigraron recientemente del Asia. Se pretende que el intervalo tenga una semilongitud de a lo sumo 6 mg/dL.

Se sabe que la distribución del colesterol es aproximadamente normal. En una muestra previa se obtuvo una desviación estándar de 40 mg/dL.

Entonces:  $\alpha = 0.02 \Rightarrow z = 2.33$ ,  $A = 6$ ,  $\sigma^* = 40$  y por lo tanto,

$$n \geq \left( \frac{z_{\alpha/2} \cdot \sigma^*}{A} \right)^2 = \left( \frac{2.33 \cdot 40}{6} \right)^2 = 241.28,$$

es decir debemos tomar  $n \geq 242$ . Notar que utilizamos  $\sigma^*$  y no  $\sigma$  ya que este último es seguramente desconocido, 40 es una estimación obtenida de algún estudio previo.

El intervalo de confianza para  $\mu$  presentado más arriba es válido cuando la muestra es grande. Esto asegura que la distribución de muestreo de  $\bar{X}$  es aproximadamente normal. Además asegura que la desviación estándar muestral estará suficientemente cerca de la desviación estándar poblacional desconocida.

Sin embargo, en muchas investigaciones el número de observaciones es pequeño, debido por ejemplo, al costo o al tiempo necesario para obtener un dato. Consideraremos ahora una distribución propuesta por W. S. Gosset, en el año 1908, quien firmaba sus papers bajo el seudónimo de Student.

### Distribución t de Student

Gosset demostró que si las observaciones tienen distribución normal, la media muestral estandarizada usando el desvío estándar muestral, ya no tiene distribución normal, sino que tiene una distribución “parecida” a la normal. Esta distribución se conoce como DISTRIBUCIÓN T DE STUDENT.

Si  $X \sim N(\mu, \sigma^2)$  entonces, para una muestra de tamaño  $n$ , la distribución de muestreo del estadístico

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

se denomina distribución  $t$  de Student con  $n - 1$  grados de libertad.

## Notación.

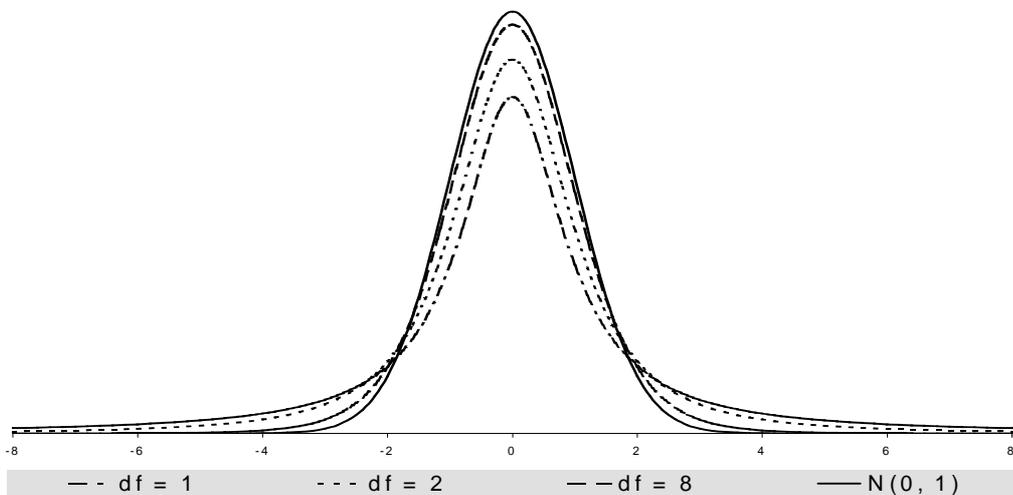
$T \sim t_{n-1}$  se lee “T tiene distribución  $t$  de Student con  $n - 1$  grados de libertad”.

## Características de la distribución $t$

- Es una distribución unimodal, simétrica y acampanada, centrada en cero.
- Tiene un único parámetro, que se denomina *grados de libertad* ( $\nu$ ).
- La dispersión de la distribución depende de los grados de libertad:

$$\sigma_T = \sqrt{\frac{\nu}{\nu - 2}}$$

- Cuando los grados de libertad aumentan la dispersión tiende a 1.
- La distribución  $t$  es más achatada que la  $N(0, 1)$  y por lo tanto acumula mayor probabilidad en las colas.
- Cuando los grados de libertad aumentan la distribución  $t$  se parece más y más a la  $N(0, 1)$ .



- En nuestro contexto (una muestra con  $n$  observaciones) los grados de libertad coinciden con  $n - 1$ , el denominador de la varianza muestral.
- La distribución  $t$  es una familia de curvas. El área debajo de una distribución particular está tabulada.

**Intervalo de confianza para  $\mu$  cuando  $X \sim N(\mu, \sigma^2)$**

El intervalo de confianza se construye en base a la distribución t de Student. Por lo tanto, un intervalo de confianza de nivel  $1 - \alpha$  para la media poblacional  $\mu$  es

$$\bar{X} \pm t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}$$

donde  $t_{n-1, \alpha/2}$  es el percentil de la distribución t con  $n - 1$  grados de libertad, que deja a su derecha un área  $\alpha/2$ .

La interpretación de este intervalo es similar a la que hemos presentado para intervalos basados en la distribución normal.

Debemos tener en cuenta que este intervalo sólo es aplicable cuando podemos suponer normalidad de la variable de interés.

### Ejemplo

Supongamos que la presión sistólica se asume siguiendo una distribución normal y queremos hallar un intervalo de confianza para  $\mu$ .

En una muestra de 5 estudiantes se obtuvieron los siguientes resultados 120 80 90 110 95

La media de la muestra es  $\bar{X} = 99$  mmHg y la desviación estándar  $s = 15.97$  Mg.

Si queremos un intervalo de 95% de confianza,  $\alpha = 0.05$  y  $\frac{\alpha}{2} = 0.025$ , necesitamos

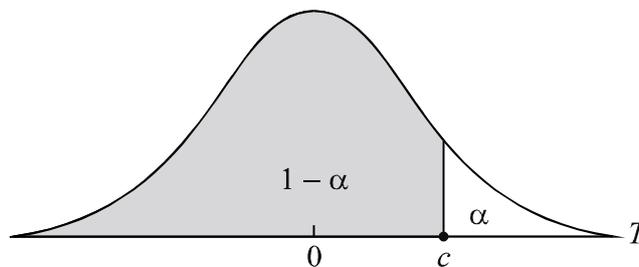
entonces el valor de  $t_{4,0.025} = 2.776$  que obtenemos de la tabla que se encuentra en la página siguiente.

$$\bar{X} \pm t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} = 99 \pm 2.776 * 15.97 / \sqrt{5} = 99 \pm 19.826$$

El intervalo queda entonces (79.174, 118.826)

### TABLA DE LA DISTRIBUCION t-Student

$$P[T \leq c] = 1 - \alpha$$



$\nu$	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576

## Ejercicios

- 1) La temperatura corporal aumenta unas décimas en los días de ovulación de la mujer. Con el objeto de estimar el día exacto de ovulación en las mujeres, se midió su temperatura corporal en días en los que con seguridad no estaban ovulando. Se tomó una muestra de 100 mujeres y se obtuvo una temperatura media de 36.22° centígrados con una desviación estándar muestral de 0.105° centígrados.
  - a) Obtenga un intervalo de nivel 99% de confianza a partir de dicha muestra para la temperatura corporal de una mujer en días en los que no está ovulando.
  - b) ¿Qué tamaño de muestra se debería usar si se quiere que la longitud del intervalo sea a lo sumo 0.04?
  
- 2) Supongamos que de estudios previos se sabe que la media de colesterol en niños entre 2 y 14 años es de 175 mg/dL. Se identificó un grupo de padres que sufrieron un ataque cardíaco y elevaron su colesterol a más de 250 mg/dL y se midió el nivel de colesterol de sus hijos. La media de un grupo de 20 niños correspondiente a hijos de esa población de padres resultó 207.3 mg/dL, con una desviación estándar de 30. Asuma que el nivel de colesterol sigue una distribución normal
  - a) Construya un intervalo de nivel 95% para la media de colesterol de los hijos de padres que sufrieron ataque cardíaco.
  - b) ¿Le parece razonable pensar que en esta población la media de colesterol es 175?

## TEST DE HIPÓTESIS

El objetivo en muchos estudios es chequear si los datos concuerdan con ciertas predicciones o hipótesis acerca de variables medidas en el estudio. En bioestadística, el interés generalmente está centrado en comparar una variable en grupos de sujetos, animales o plantas, que han sido sometidos a distintos tratamientos, procedimientos, condiciones de vida, etc. El valor numérico asociado a la comparación de interés se denomina EFECTO.

### **Hipótesis**

Una *hipótesis* es una afirmación acerca de algún parámetro que caracteriza a la distribución de una variable de interés.

En los problemas de comparación es posible establecer dos hipótesis contrapuestas

⇒ *Hipótesis Nula*

$H_0$ : No existe efecto (el efecto es cero).

⇒ *Hipótesis Alternativa*

$H_1$ : Existe efecto (el efecto es distinto de cero).

En general, la hipótesis nula ( $H_0$ ) es la negación de la hipótesis del investigador, indica la posición de que nada cambia, de que el tratamiento no tiene efecto. Mientras que la hipótesis alternativa ( $H_1$ ), representa la posición del investigador.

Si interesa probar que el tratamiento A es mejor que el B para curar el dolor de cabeza, las hipótesis serán:

$H_0$ : los dos tratamientos son igualmente efectivos

$H_1$ : el tratamiento A es mejor que el B

¿Cómo elegir con que hipótesis nos quedamos? Tenemos que tomar muestras, y tomar nuestra decisión en base a lo que observamos en ellas. Aún cuando no existan diferencias en los tratamientos, es común observar diferencias en los resultados de los dos grupos debido al hecho de haber tomado muestras aleatorias para realizar el experimento. Un TEST DE HIPÓTESIS responde a la pregunta:

**¿La diferencia observada es lo suficientemente grande como para pensar que es real o puede ser atribuida al azar?**

Las hipótesis son producidas a partir de la teoría que induce a realizar el experimento o la investigación. Cuando una hipótesis se relaciona con características poblacionales, tales como parámetros poblacionales, uno puede utilizar métodos estadísticos y un conjunto de datos para decidir cual de ambas hipótesis tiene mayor verosimilitud.

Ejemplos de hipótesis que podrían ser testeadas estadísticamente son:

- La supervivencia de pacientes con cáncer de mama tratadas con tamoxifeno es mayor que la de pacientes no tratadas.
- El nivel de colesterol sérico es mayor en varones que en mujeres de la misma edad.
- La mayoría de los argentinos afiliados a una obra social están satisfechos con el servicio de salud que la misma les presta.

A continuación presentaremos las ideas básicas de un test de hipótesis a través de un ejemplo.

Un área corriente de investigación consiste en determinar si los antecedentes familiares agregan factores de riesgo en problemas cardiovasculares. Supongamos que el nivel medio de colesterol en niños es de 175 mg/dL. Se identificó un grupo de padres que murieron el año anterior a causa de problemas cardíacos y se registró el colesterol de sus hijos. En este caso la hipótesis nula, la de “no cambio”, es:

$H_0$ : El nivel medio de colesterol de esos niños es 175

Mientras que la alternativa, la del investigador, es:

$H_1$ : El nivel medio de colesterol de esos niños es mayor a 175

Supongamos que de estudios previos se sabe que el nivel de colesterol sigue una distribución normal con un desvío de 30, es decir que el nivel de colesterol

$$X \sim N(\mu, 30^2)$$

Entonces podemos plantear las hipótesis de este modo

$$H_0 : \mu = 175$$

$$H_1 : \mu > 175$$

Si en la muestra el nivel medio de colesterol resultó 200, ¿qué valor de  $\mu$  hace más probable este resultado? ¿ $\mu = 175$  o  $\mu > 175$ ?

Asumamos que  $\mu = 175$ , es decir que los antecedentes familiares no elevan el nivel de colesterol medio de los hijos, ¿cuál es la probabilidad de que en una muestra de tamaño 10 obtengamos un promedio de 200 o más?

Calculamos la probabilidad de obtener una muestra con una media muestral como la obtenida, o más alejada aún del valor propuesto para  $\mu$  en  $H_0$ , es decir para  $\mu = 175$ .

Para calcular esta probabilidad es necesario conocer la **DISTRIBUCIÓN DEL ESTADÍSTICO**, es decir la distribución del promedio muestral.

Como asumimos que el colesterol es una variable con distribución  $N(\mu, 30^2) \Rightarrow \bar{X} \sim N(\mu, 30^2/10)$  y como además estamos suponiendo que  $\mu = 175$ ,  $\bar{X} \sim N(175, 90)$ .

$$\text{Entonces } P(\bar{X} \geq 200) = P\left(\frac{\bar{X} - 175}{\sqrt{90}} \geq \frac{200 - 175}{\sqrt{90}}\right) = 1 - \Phi(2.63) = 0.0043$$

Con este resultado podemos decir que, si la media de colesterol de los hijos de padres que murieron por ataque cardíaco es 175, hay un 0.43% de probabilidad de que en una muestra de 10 de estos niños el promedio muestral resulte ser 200 o más.

Esta probabilidad se denomina **p-valor**. Es una medida de cuan consistentes son nuestros datos con la hipótesis nula.

Mientras más pequeño sea el p-valor, mas fuerte es la evidencia para RECHAZAR la hipótesis nula.

Por lo tanto estaríamos inclinados a pensar que la media debe ser mayor a 175.

*¿Cómo interpretamos el p-valor?*

Uno comienza suponiendo que la hipótesis nula es verdadera (el colesterol medio  $\mu = 175$ ) e imagina a muchos investigadores repitiendo el experimento en idénticas condiciones. Este p-valor nos dice que sólo 43 de cada 10000 investigadores que obtuvieron muestras aleatorias tomadas con  $\mu = 175$  obtendrían un valor del estadístico mayor o igual a 200 mg/dL.

Finalmente, tomamos una DECISIÓN respecto de si nuestros datos contradicen o no la hipótesis nula.

En nuestro ejemplo, el p-valor (la probabilidad de obtener una muestra como la que hemos obtenido o más alejada aún del valor propuesto en  $H_0$ ) es muy pequeña, por lo que concluimos que la diferencia entre lo observado en la muestra y lo que esperaríamos observar si  $H_0$  es verdadera, es demasiado grande para ser atribuida al azar. Esto nos conduce a RECHAZAR  $H_0$  y concluir que el nivel medio de colesterol de estos niños es mayor a 175, es decir que nos quedamos con la hipótesis alternativa ( $H_1$ ).

Podemos concluir entonces que hemos observado una diferencia *estadísticamente significativa* entre el promedio de colesterol de estos niños y el valor poblacional 175 ( $p = 0.0043$ ).

### *¿Cómo se toma la decisión de rechazar o no la hipótesis nula?*

Cuanto más pequeño sea el p-valor, mayor es la evidencia para rechazar  $H_0$ .

Un p-valor cercano a 1 nos dice que los datos son consistentes con la hipótesis nula. Un p-valor muy pequeño es evidencia en contra de la hipótesis nula, nos dice que los datos son muy improbables cuando  $H_0$  es verdadera.

Pero, ¿qué decisión se toma frente a p-valores no tan extremos? Lo que hacemos es definir un valor de corte, que se denomina *nivel de significación del test* y que se denota  $\alpha$  y usamos el siguiente criterio:

1. si  $p\text{-valor} < \alpha$ , decimos que el resultado es *estadísticamente significativo*
2. si el  $p\text{-valor} \ll \alpha$  (es decir mucho menor que  $\alpha$ ) decimos que el resultado es *altamente significativo*.

Usualmente se considera  $\alpha = 0.05$ .

### *Comentarios.*

- En el ejemplo hemos rechazado la hipótesis nula, que dice que en la población

de niños con padres que murieron por ataque cardíaco, el colesterol medio es 175. Es decir, el aumento observado en la muestra no es atribuible a variabilidad aleatoria y esto nos lleva a concluir que el colesterol medio de esta población es mayor a 175. Sin embargo, no podemos asegurar que la CAUSA de este aumento haya sido la muerte de sus padres!!!. Podría haber muchas otras razones que expliquen un aumento en el colesterol respecto al promedio general (por ejemplo, un tipo de alimentación diferente).

- Es importante distinguir entre *significación estadística* e *importancia práctica*. Un p-valor muy pequeño, tal como  $p=0.0043$ , no necesariamente implica un hallazgo importante desde el punto de vista práctico. Simplemente significa que si la hipótesis nula fuera cierta, la muestra presente es realmente atípica.
- El test de hipótesis es un argumento POR CONTRADICCIÓN, diseñado para mostrar que la hipótesis nula lleva a conclusiones absurdas y por lo tanto debe ser rechazada. Un p-valor pequeño provee evidencia en contra de  $H_0$ , ya que la muestra observada es improbable cuando  $H_0$  es verdadera. Por otro lado, si el p-valor no es pequeño la hipótesis nula es factible. En este caso la conclusión debe reportarse como “No se rechaza  $H_0$ ”, ya que los datos no contradicen  $H_0$ .  
Esto no significa que “aceptemos  $H_0$ ” o que podamos asegurar que  $H_0$  es necesariamente verdadera. Simplemente concluimos que la evidencia en nuestros datos no ha sido suficiente para rechazarla.
- El p-valor NO ES la probabilidad de que la hipótesis nula sea verdadera. En una aplicación particular, la hipótesis nula es verdadera o es falsa, no es correcto hablar de *probabilidad* de que la hipótesis nula sea verdadera.

### DECISIONES Y TIPOS DE ERRORES EN LOS TESTS DE HIPÓTESIS

Un test de hipótesis es un procedimiento que nos permite tomar una decisión basándonos en información proveniente de una única muestra. Cualquiera sea la decisión que tomemos es posible que estemos cometiendo un error. La Tabla siguiente resume los dos tipos de errores que se pueden cometer.

		REALIDAD	
		$H_0$ verdadera	$H_0$ falsa
DECISIÓN BASADA EN LA MUESTRA	No rechazar $H_0$	O.K.	Error Tipo II
	Rechazar $H_0$	Error tipo I	O.K.

Definimos:

- *Error de Tipo I* : Rechazar la hipótesis nula cuando es verdadera.
- *Error de Tipo II* : No rechazar la hipótesis nula cuando es falsa.

$\alpha$  = P(error tipo I) = P(rechazar  $H_0$  cuando  $H_0$  es verdadera)

= NIVEL DE SIGNIFICACIÓN DEL TEST

$\beta$  = P(error tipo II) = P(no rechazar  $H_0$  cuando  $H_0$  es falsa)

$\pi$  = P(rechazar  $H_0$  cuando  $H_0$  es falsa) =  $1 - \beta$  = POTENCIA DEL TEST

El nivel de significación de un test ( $\alpha$ ) es la máxima probabilidad de cometer error de tipo I que el investigador está dispuesto a tolerar. Como ocurre con los intervalos de confianza, la elección de  $\alpha$  refleja cuanta protección pretende el investigador respecto de la posibilidad de cometer error de tipo I. Mientras menor sea  $\alpha$  mayor será la evidencia necesaria para rechazar la hipótesis nula.

Para evitar sesgos en el proceso de toma de decisión, el nivel  $\alpha$  debe seleccionarse ANTES de analizar los datos.

Una vez fijado el valor de  $\alpha$ , en función de lo observado en la muestra, si el p-valor es menor que el nivel establecido  $\alpha$ , rechazamos la hipótesis propuesta en  $H_0$

La probabilidad de cometer error tipo II ( $\beta$ ) y la potencia del TEST ( $\pi$ ) dependen de la magnitud del efecto que estamos estudiando (que es desconocida) y del tamaño de muestra. Es posible obtener un test de una potencia tan grande como se desee (o un  $\beta$  tan pequeño como se desee) eligiendo un tamaño de muestra apropiado. Más adelante veremos cómo decidir el tamaño de muestra, en función de la potencia deseada para el test, para diferentes tipos de estudio.

### *Regiones de Rechazo y de No Rechazo*

Para cada test de nivel de significación  $\alpha$ , existe un conjunto de valores del estadístico que conducen a rechazar  $H_0$  y que proveen evidencia *a favor de la hipótesis alternativa*, al que denominaremos *región de rechazo* y otro conjunto de valores del estadístico que conducen a no rechazar  $H_0$  a la que llamaremos *región de no rechazo*. Los gráficos siguientes muestran estas regiones para un test de nivel  $\alpha$ , cuyas hipótesis son:

$H_0: \mu = 175 \text{ mg/dL}$

versus

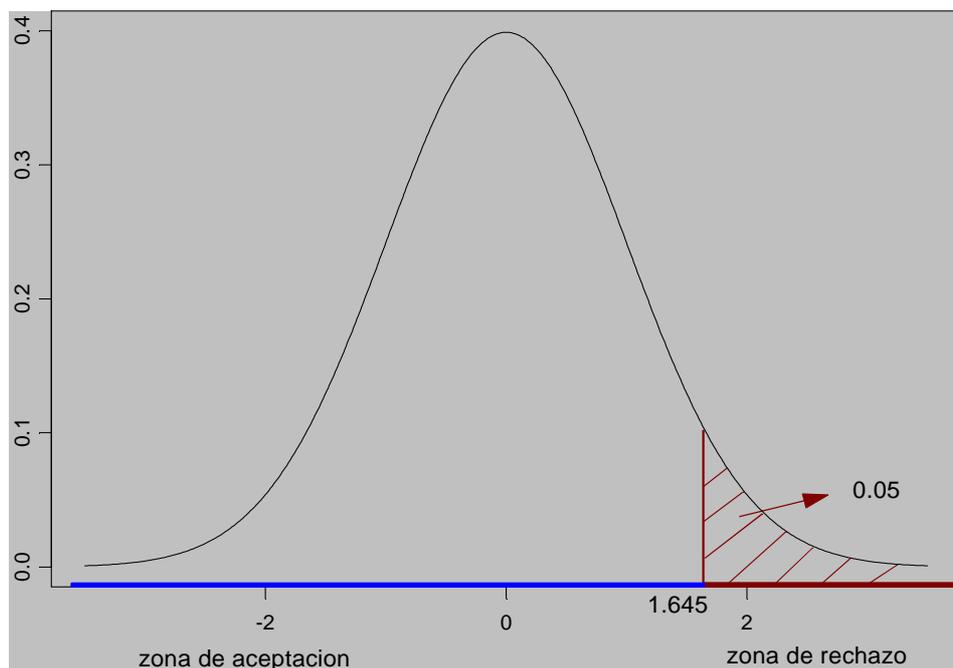
$H_1: \mu > 175 \text{ mg/dL}$

Supongamos que queremos que  $\alpha = P(\text{error tipo I}) = P(\text{rechazar } H_0 \text{ cuando } H_0 \text{ es verdadera})$  sea 0.05.

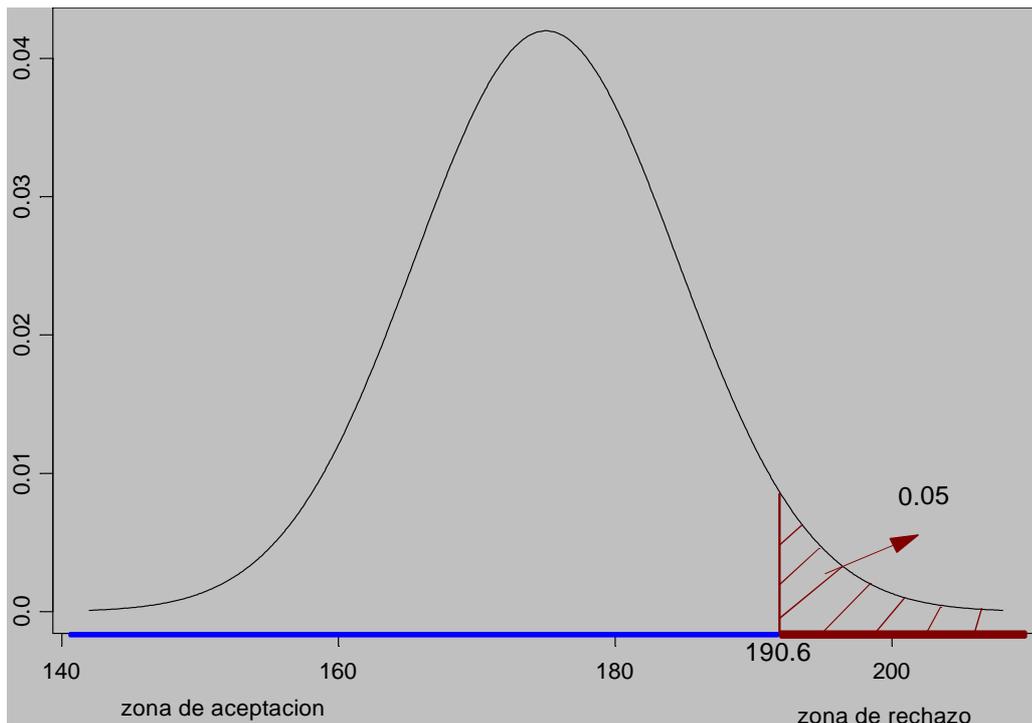
¿Cuál sería un valor de  $\bar{X}$  tal que la probabilidad de superar ese valor cuando  $\mu = 175$  sea 0.05?

Si estandarizamos la variable, obtendremos a cuantos errores estándar por encima de la media tenemos que estar, para que la probabilidad de superar ese valor sea 0.05.

Buscando en la tabla de la  $N(0,1)$  vemos que la probabilidad de estar 1.645 desviaciones por encima de la media es 0.05.



Entonces para nuestra variable  $\bar{X}$ , asumiendo que sigue una distribución  $N(175,90)$ , el valor que está 1.645 errores estándar por encima de la media es  $\bar{X} = 175 + 1.645 \sqrt{90} = 190.6$



Es decir, en este test, la región de rechazo está constituida por todos los valores de  $Z \geq 1.645$  ya que  $P(Z \geq 1.645) = 0.05$  o equivalentemente por todos los valores de  $\bar{X} \geq 190.6$ , ya que si

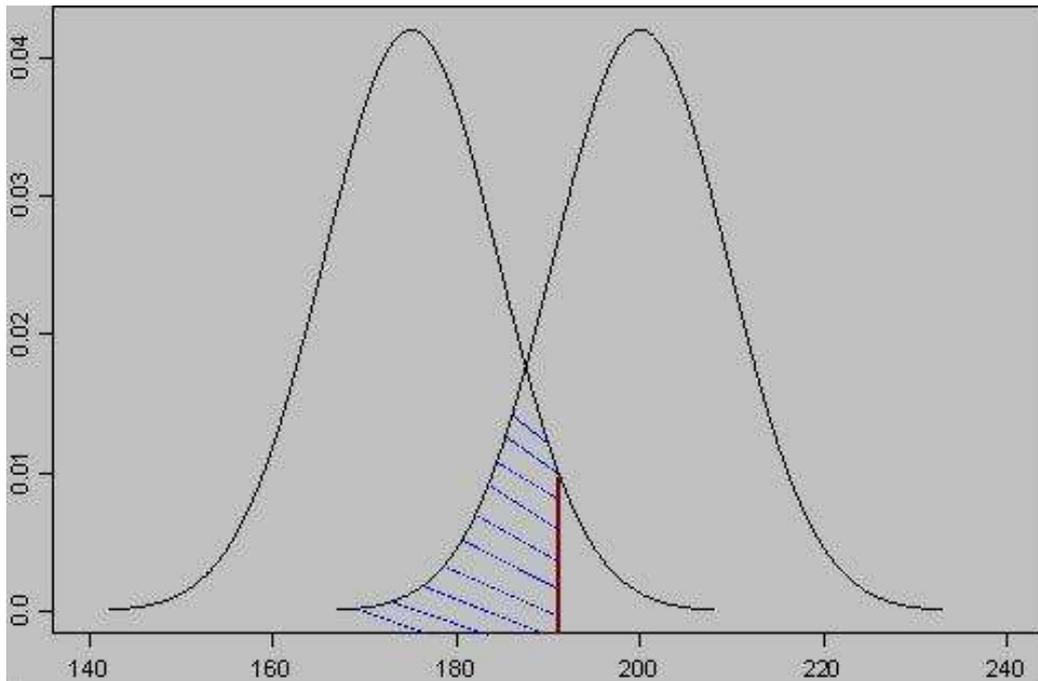
$$Z = \frac{\bar{X} - 175}{\sqrt{90}} = 1.645 \Rightarrow \bar{X} = 1.645 \cdot \sqrt{90} + 175 \Rightarrow \bar{X} = 190.6$$

La región de rechazo depende del nivel del test y de la hipótesis alternativa.

El investigador controla la probabilidad de error de tipo I al definir el nivel de significación del test.

### *Error tipo II*

Un error de tipo II ocurre cuando no rechazamos  $H_0$  siendo ésta falsa. Como puede observarse en el siguiente gráfico, esta probabilidad depende de cuán lejos cae el verdadero valor del parámetro del valor propuesto en  $H_0$ .



Cuando  $\mu = 200$  la probabilidad de error de tipo II,  $(\beta)$ , es el área sombreada. ¿Cómo se calcula?

Tenemos que calcular la probabilidad de que, siendo la verdadera media de colesterol 200, en una muestra de tamaño 10, obtengamos un promedio muestral que nos lleve a no rechazar  $H_0$ , es decir un valor menor o igual a 190.6.

Cuando  $\mu=200$ ,  $\bar{X} \sim N(200, 90)$ , entonces

$$P(\bar{X} \leq 190.6) = P\left(\frac{\bar{X} - 200}{\sqrt{90}} \leq \frac{190.6 - 200}{\sqrt{90}}\right) = P(Z \leq -0.99) = 0.1611$$

¿Qué representa la potencia  $\pi = 1 - \beta$  ?

Cuando  $\mu=200$ , tenemos una probabilidad de 16.11% de que en una muestra de tamaño 10, la media muestral resulte menor o igual que 190.6 y esto nos lleve a concluir, erróneamente, que el nivel medio de colesterol de estos niños es 175.

Entonces, tenemos una probabilidad de 83.89% de concluir en forma correcta que el colesterol de estos niños es mayor que 175.

La **potencia** de un test, es la capacidad que tiene, para cada valor del parámetro propuesto en  $H_1$ , de detectar que la hipótesis nula es falsa.

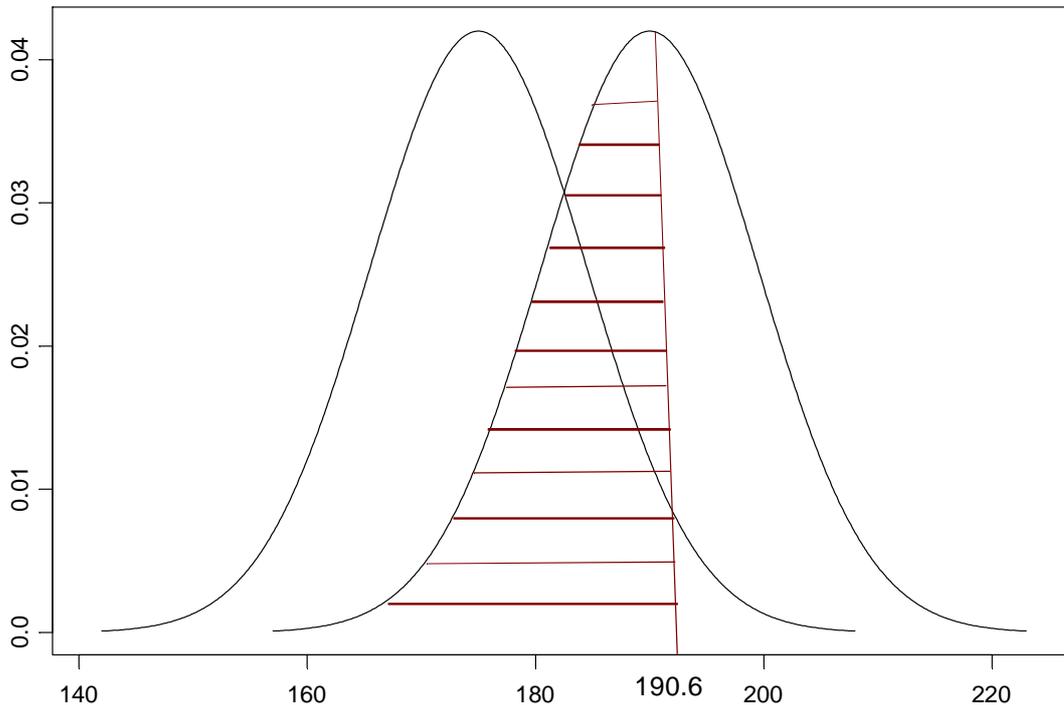
La probabilidad de error de tipo II depende de cuán lejos se encuentre el verdadero valor del parámetro del valor propuesto en  $H_0$ . Si el verdadero valor del parámetro es cercano al valor propuesto en  $H_0$ , la probabilidad de error de tipo II  $(\beta)$ , puede ser muy grande. Mientras más alejado esté el verdadero valor del parámetro del valor

especificado en  $H_0$ , menor es la probabilidad de que la muestra falle en detectar la diferencia y resultar en un error de tipo II.

En la siguiente tabla se dan las probabilidades de error II, ( $\beta$ ) y la potencia, ( $\pi$ ), para distintos valores de  $\mu$ , con  $\alpha = 0.05$

$\mu$	$\beta$	$\pi$
180	0.8681	0.1319
185	0.7225	0.2775
190	0.5252	0.4748
195	0.3214	0.6786
200	0.1609	0.8391
205	0.0645	0.9355
210	0.0204	0.9796
215	0.0051	0.9949

La probabilidad de error de tipo I y II están inversamente relacionadas. Cuando  $\alpha$  disminuye,  $\beta$  aumenta y viceversa. En otras palabras, mientras más fuerte sea la evidencia requerida para rechazar  $H_0$  (o sea mientras menor sea  $\alpha$ ), más probable es que fracasemos en detectar una diferencia real.



Para una probabilidad de error tipo I ( $\alpha$ ) fija, uno puede disminuir la probabilidad de error de tipo II ( $\beta$ ) usando una muestra de mayor tamaño. Si en nuestro ejemplo tomásemos una muestra de tamaño 64 en vez de una muestra de tamaño 10, la distribución de  $\bar{x}$  tendría un error estándar de  $30/\sqrt{64} = 3.75$  y por lo tanto, para un

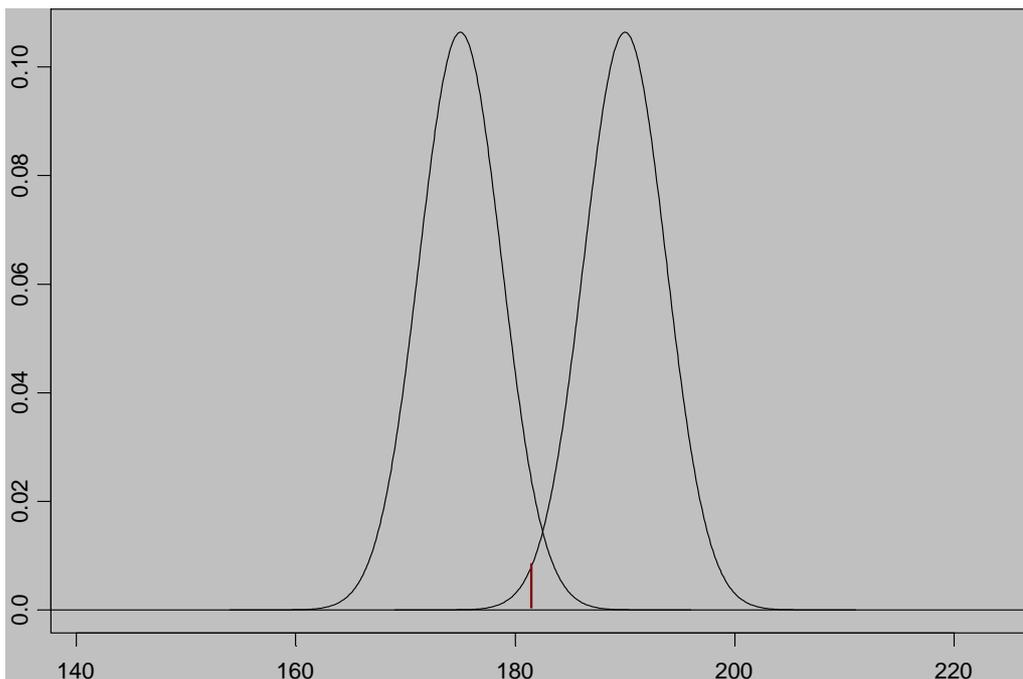
test de nivel  $\alpha=0.05$ , la probabilidad de error de tipo II, cuando  $\mu = 190$ , se modifica al aumentar el tamaño de muestra.

Debemos tener en cuenta que ahora el punto de corte no es 190.6, ya que con este error estándar, estar 1.645 desviaciones estándar por encima de la media propuesta en  $H_0$ , es estar por encima de  $175+1.645*3.75=181.16$ . Entonces la probabilidad de error II cuando  $\mu = 190$  es

$$P(\bar{X} \leq 181.16) = P\left(\frac{\bar{X}-190}{3.75} \leq \frac{181.16-190}{3.75}\right) = P(Z \leq -2.35) = 0.0094,$$

por lo tanto al aumentar el tamaño de muestra de 10 a 64 la probabilidad de error II se redujo de un 52.52% a un 0.94% y la potencia aumento de 47.48% a 99.06%.

Al aumentar el tamaño de muestra, disminuye el error estándar y es más fácil discriminar, las curvas correspondientes a las distribuciones de  $\bar{X}$ , suponiendo  $\mu = 175$  o  $\mu = 190$ , se superponen mucho menos, permitiendo distinguir mejor cual es la hipótesis con mayor verosimilitud.



Siempre que aplicamos un test, la conclusión puede o no ser correcta. Cuando rechazamos la hipótesis nula podemos estar cometiendo un error de tipo I, mientras que si la conclusión es no rechazar  $H_0$ , podemos estar cometiendo error de tipo II. Aunque en una aplicación particular no sabemos si la conclusión es correcta, el procedimiento se justifica en términos de la proporción de veces que se cometen estos dos tipos de errores en muchas repeticiones del procedimiento.

En la práctica el investigador sólo decide el nivel de significación del test. Respecto de la potencia puede:

- Calcular el tamaño de muestra necesario para asegurar una potencia razonablemente alta para un valor del parámetro lo suficientemente alejado del valor propuesto en  $H_0$  de modo que la conclusión sea técnicamente valiosa.
- Cuando los recursos son acotados, y es imposible lograr el tamaño de muestra obtenido en el cálculo anterior, es aconsejable calcular la potencia que tendrá el test para este mismo valor del parámetro y para el tamaño de muestra factible. Cuando la potencia resulta muy baja, este sencillo cálculo puede llevar a la decisión de no realizar la investigación.

No realizaremos cálculos de tamaños de muestra para distintos valores de la potencia, ya que estos cálculos son bastante complejos.

#### *Algunos comentarios acerca del tamaño muestral y la conclusión del test.*

Es frecuente escuchar la opinión de que un estudio basado en una muestra de pequeño tamaño tiene escaso valor y resulta poco concluyente. Esta afirmación, como se comenta más abajo, no necesariamente es correcta.

En el siguiente análisis, consideraremos que los estudios a los que haremos referencia han sido correctamente diseñados y están libres de sesgo.

Cuando el resultado del test es negativo (no significativo) y el tamaño de muestra es pequeño, esta afirmación es válida, ya que la probabilidad de error de tipo II puede ser muy grande. Si a pesar de tener una muestra pequeña, el resultado es positivo (test significativo), esto indica que el efecto que estamos analizando es suficientemente importante como para lograr ser detectado a pesar del tamaño de muestra.

Respecto de los casos en que el tamaño de muestra es muy grande, y el resultado es significativo, debemos ser cuidadosos en decidir sobre la importancia clínica o técnica del resultado, ya que un gran tamaño de muestra puede conducir a declarar significativos efectos que en la práctica sean poco relevantes.

Finalmente, cuando el tamaño de muestra es muy grande y el resultado es no significativo, estamos frente a un estudio concluyente ya que, si trabajando con gran

potencia no fue posible rechazar la hipótesis nula, esto implica que si el efecto existe, es prácticamente despreciable.

Como se dijo más arriba, las conclusiones anteriores sólo son válidas cuando se trata de estudios libres de sesgo, ya que de lo contrario es imposible decidir cuánto del efecto observado se debe al tratamiento aplicado y cuánto a error sistemático.

Por último, los comentarios anteriores son simplemente orientativos y no se aplican estrictamente en todas las situaciones. Resumiendo:

	n PEQUEÑO	n MUY GRANDE
RECHAZO $H_0$	¡CONCLUYENTE! El efecto es tan grande que aún con un n pequeño pudo ser detectado.	¡CUIDADO! Chequear si la magnitud del efecto es técnicamente importante.
NO RECHAZO $H_0$	NO CONCLUYENTE Es posible que el efecto exista y no pudiera ser demostrado.	¡CONCLUYENTE! Si el efecto existe es despreciable.

Veamos otro ejemplo, supongamos que se quiere testear si las madres que pertenecen a un nivel socioeconómico bajo, tienen hijos con un peso medio al nacer menor al normal. Para testear esta hipótesis se registraron los pesos de 100 bebés, cuyas madres pertenecen a este nivel socioeconómico, y se obtuvo un promedio muestral  $\bar{X} = 3.260$  kg con una desviación estándar muestral de 0.68 kg.

Supongamos que en el país la media del peso al nacer de un bebé es de 3.400 kg. ¿Podemos decir que la media de peso al nacer de estos bebés es menor a la media del país?

En este caso las hipótesis a testear serían

$$H_0 : \mu = 3.400$$

$$H_1 : \mu < 3.400$$

Como  $n=100$  es suficientemente grande, podemos pensar que  $\bar{X}$  sigue una distribución aproximadamente normal, o equivalentemente,  $\frac{\bar{X} - \mu}{s/\sqrt{100}}$  tiene una distribución aproximadamente  $N(0,1)$ .

Si fijamos la probabilidad de cometer error de tipo I,  $\alpha = 0.05$ , deberíamos rechazar  $H_0$ , si obtenemos un valor de  $\bar{X}$  que esté 1.645 errores estándar por debajo de la media propuesta en  $H_0$ ,  $\mu = 3.400$ , es decir si  $\frac{\bar{X} - 3.400}{s/10} < -1.645$ .

Reemplazando por los valores obtenidos en la muestra,

$$\frac{3.260 - 3.400}{0.68/10} = -2.06$$

Como obtuvimos un resultado menor a -1.645 concluimos que las diferencias son significativas a nivel 0.05 y el peso medio de estos bebés es menor a 3.400.

En este caso, el p-valor es la probabilidad de obtener un valor tan alejado de la media como 3.260 kg, asumiendo que la verdadera media es la propuesta en  $H_0$ , es decir, 3.400 kg.

$$\text{p-valor} = P(\bar{X} < 3.260) = P\left(\frac{\bar{X} - 3.400}{0.68/10} < \frac{3.260 - 3.400}{0.068}\right) = P(Z < -2.06) = 0.0197$$

Es decir que si muchos investigadores tomaran muestras de tamaño 100 de esta población, si asumimos que la media de la población es 3.400, sólo en el 1.97% de los casos se obtendrían valores del estadístico menores o iguales a 3.260 kg.

### HIPÓTESIS ALTERNATIVA UNILATERAL Y BILATERAL.

En nuestros dos ejemplos anteriores, propusimos las siguientes hipótesis

Ejemplo 1: nivel de colesterol de los hijos de padres que murieron por ataque cardíaco

$$H_0: \mu = 175 \text{ mg/dL} \quad \text{versus} \quad H_1: \mu > 175 \text{ mg/dL}$$

Ejemplo 2: peso al nacer de bebés cuyas madres pertenecen a un nivel socioeconómico bajo

$$H_0: \mu = 3.400 \text{ kg} \quad \text{versus} \quad H_1: \mu < 3.400 \text{ kg}$$

Estas hipótesis alternativas se denominan hipótesis unilaterales (a una cola), ya que proponen que el cambio (si se produce) es en una única dirección.

Las hipótesis BILATERALES (a dos colas), pretenden detectar apartamientos de la hipótesis nula en cualquier dirección.

Ejemplo 1:

$$H_0: \mu = 175 \text{ mg/dL} \quad \text{versus} \quad H_1: \mu \neq 175 \text{ mg/dL}$$

Ejemplo 2:

$$H_0: \mu = 3.400 \text{ kg} \quad \text{versus} \quad H_1: \mu \neq 3.400 \text{ kg}$$

En la práctica los test bilaterales (a dos colas) son más comunes que los tests unilaterales (a una cola). Aún cuando el investigador pueda predecir la dirección del cambio, los tests a dos colas permiten además detectar efectos que se producen en la dirección contraria a la esperada.

Si uno tiene evidencia previa al experimento de que el cambio se producirá en un cierto sentido, es conveniente plantear una hipótesis unilateral.

Si no se tiene información, entonces, se DEBE plantear una hipótesis bilateral. Las hipótesis deben proponerse ANTES de mirar los datos, la evidencia de la muestra no debe ser usada para decidir como plantear la hipótesis alternativa.

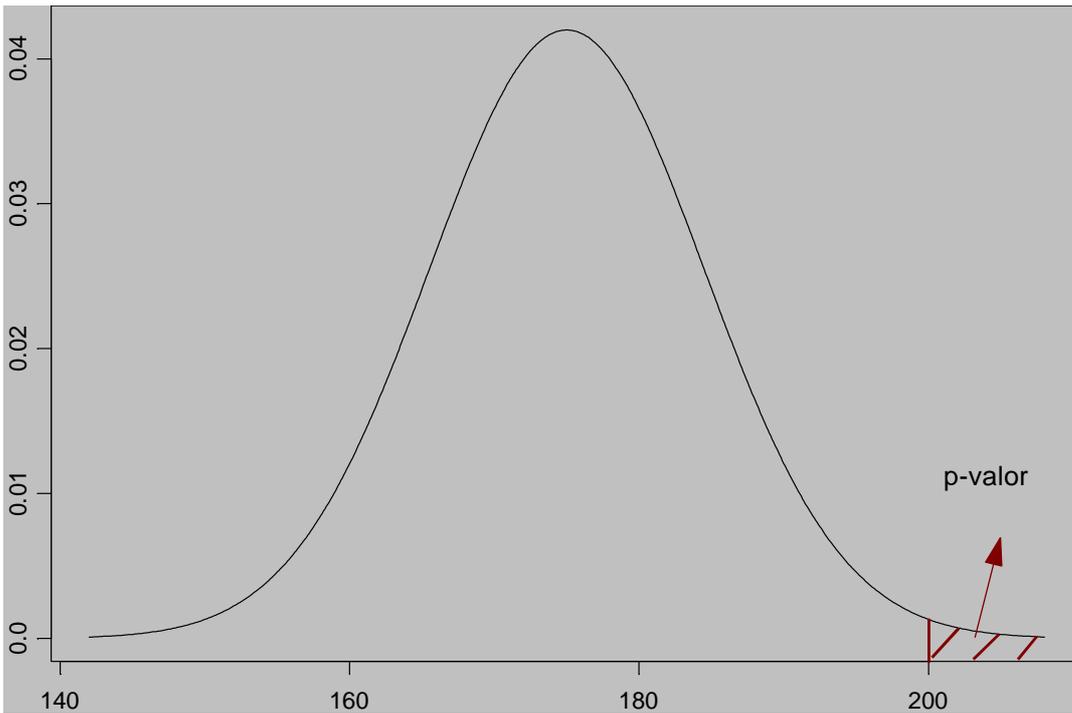
El investigador debería honestamente preguntarse si de encontrar un resultado que contradice su expectativa lo informaría. Si la respuesta es afirmativa, entonces debería plantear una alternativa bilateral.

La hipótesis alternativa de un test define la región de rechazo, así como la dirección en la que se calculará el p-valor.

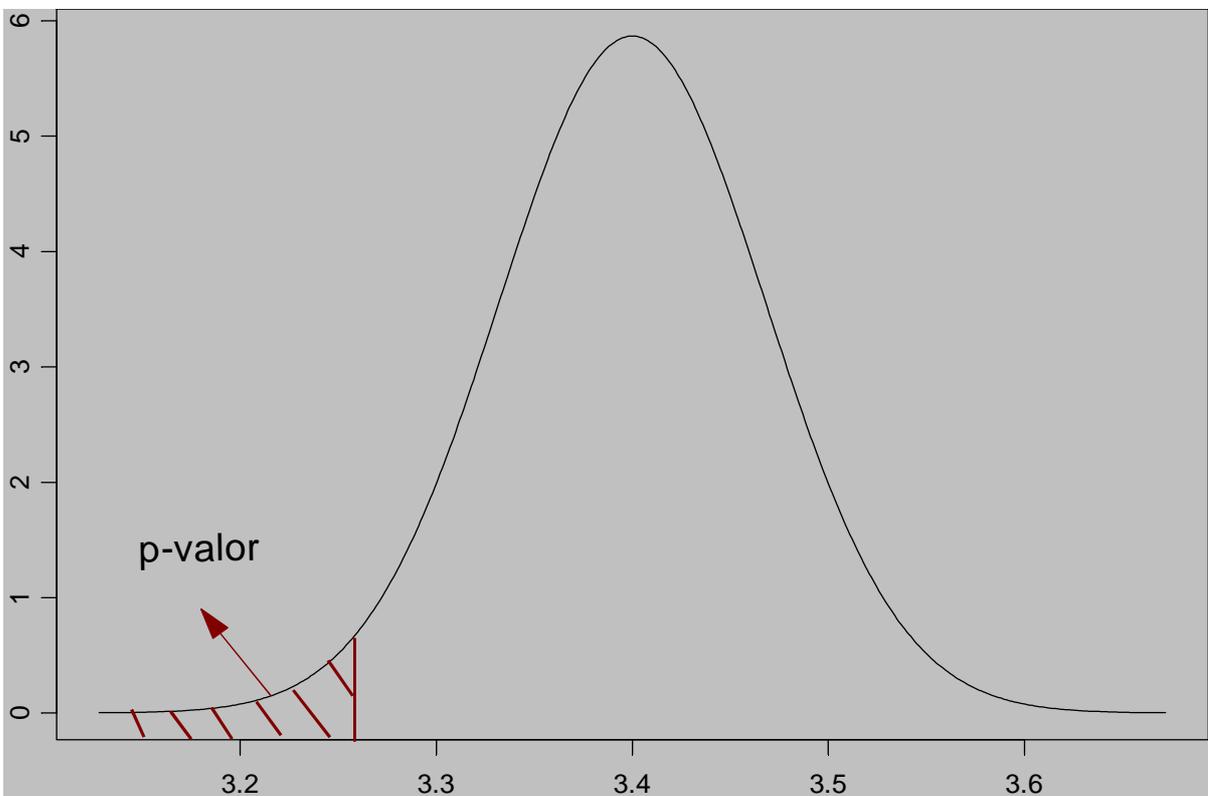
En el ejemplo 1, hemos calculado el p-valor como la probabilidad de que ocurran valores como el observado o mayores que él, es decir valores del promedio muestral mayores a 200.

Tomamos el área bajo la curva normal a la derecha del valor observado en la muestra, porque son estos los valores del estadístico que proveen evidencia en contra de  $H_0$  pero a favor de  $H_1$ .

Es la hipótesis alternativa la que determina la dirección (o las direcciones) en la que debe calcularse el p-valor.



En el ejemplo 2 obtuvimos el p-valor, como la probabilidad de obtener valores menores a 3.260 en el promedio muestral asumiendo que la media poblacional es 3.400 kg



*¿Cómo calcularíamos el p-valor en caso de hipótesis bilaterales?*

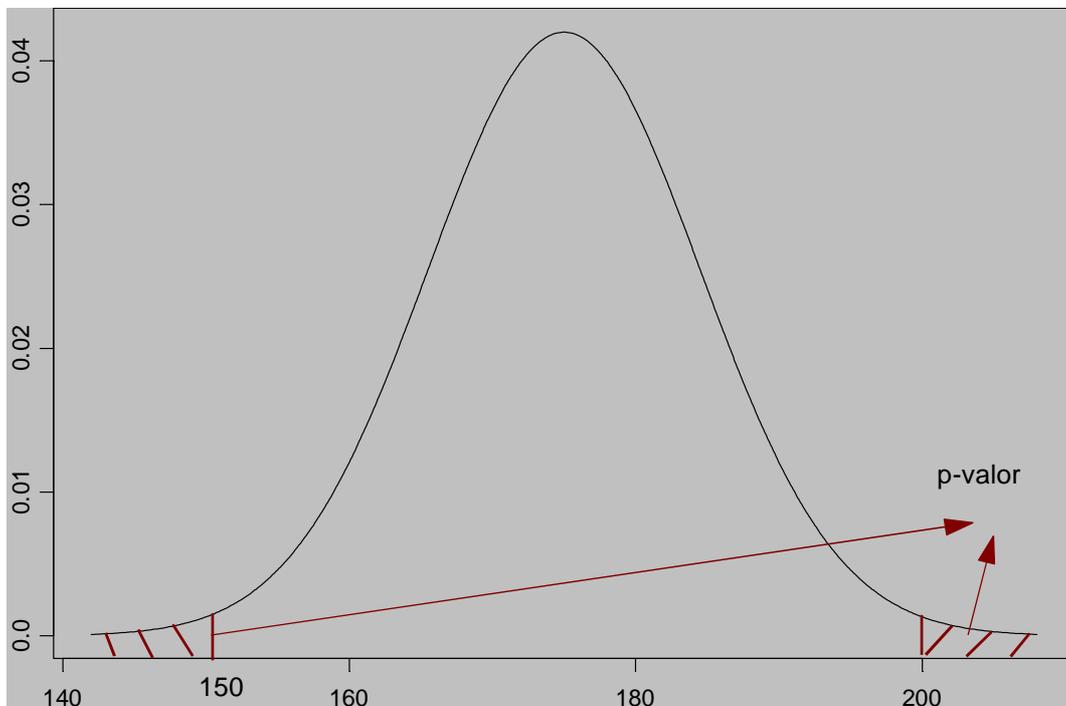
Ejemplo 1:

Tenemos que calcular la probabilidad de obtener un valor tan alejado de la media propuesta en  $H_0 : \mu = 175$ , como 200, pero en este caso este alejamiento se debe considerar en las dos direcciones, es decir si 200 estandarizado es equivalente a  $\frac{200-175}{\sqrt{90}} = 2.63$ , estamos en un valor que está a 2.63 errores estándar por encima de 175.

El p-valor es entonces la probabilidad de estar 2.63 o más errores estándar por encima de 175 o 2.63 errores estándar por debajo de 175, es decir, un promedio muestral menor a 150.

$$\text{p-valor} = P(Z > 2.63) + P(Z < -2.63) = 0.0043 + 0.0043 = 0.0086$$

Si utilizamos un nivel de significación  $\alpha = 0.05$ , como  $\text{p-valor} < 0.05$ , concluimos que las diferencias son altamente significativas y no podemos pensar que se deben al azar.



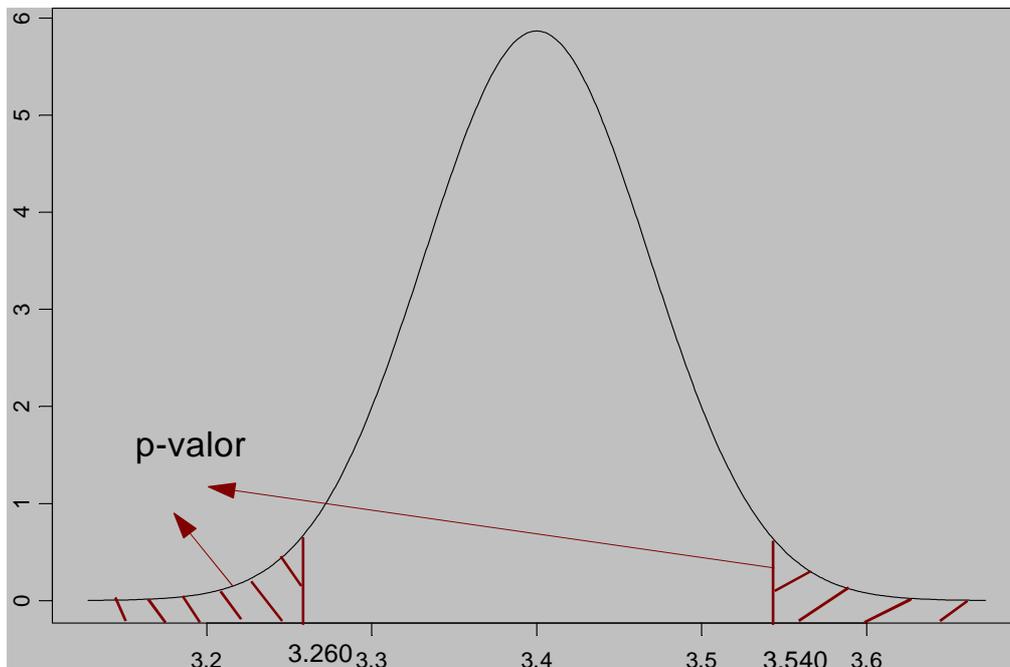
Ejemplo 2:

Tenemos que calcular la probabilidad de obtener un valor tan alejado de la media propuesta en  $H_0 : \mu = 3.400$ , como 3.260, pero en este caso este alejamiento se debe considerar en las dos direcciones, es decir si 3.260 estandarizado es equivalente a  $\frac{3.260 - 3.400}{0.68/10} = -2.06$ , estamos en un valor que está a 2.06 errores estándar por debajo de 3.400.

El p-valor es entonces la probabilidad de estar 2.06 o más errores estándar por encima de 3.400 o 2.06 errores estándar por debajo de 3.400.

$$\text{p-valor} = P(Z > 2.06) + P(Z < -2.06) = 0.0197 + 0.0197 = 0.0394$$

Si utilizamos un nivel de significación  $\alpha = 0.05$ , como  $\text{p-valor} < 0.05$ , concluimos que las diferencias son significativas y no podemos pensar que se deben al azar.



*Nota.*

Recordar que las hipótesis se plantean sobre parámetros poblacionales desconocidos (por ejemplo la media poblacional  $\mu$ ), NUNCA sobre un estadístico muestral (tal como  $\bar{X}$ ). No tenemos incerteza acerca del valor de un estadístico muestral, ya que a partir de los datos podemos calcular su valor y por lo tanto, no tenemos necesidad de hacer inferencia.

## Problemas para resolver con el statistix

### Ejercicio 1

En el archivo **DIETA.SX** encontrará los datos correspondientes a 173 personas que están siguiendo una dieta, en las que se ha registrado el consumo de grasas saturadas, el consumo de alcohol y el total de calorías.

- a) Halle intervalos de 95% de confianza para la media del consumo de grasas saturadas, del consumo de alcohol y del total de calorías de la población de personas que siguen este tipo de dieta. ¿Es necesario suponer normalidad de las variables para construir estos intervalos?

En el menú del sx

- Statistics
- One,two, Multi-sample tests
- One-sample T-test
- Elegir las variables y arrastrarlas a sample variables
- ok

- b) La dieta propone que el consumo medio de grasas saturadas debe ser en promedio 22, el consumo medio de alcohol 8 y la media del total de calorías 1400, sin embargo se cree que las personas no cumplen con los requisitos de esta dieta.

- 1) Plantee un test de nivel 5% para confirmar si el consumo medio de grasas saturadas de esta población es mayor a 22. ¿Cuál es su conclusión? ¿Cuál es el p-valor?

En el menú del sx

- Statistics
- One,two, Multi-sample tests
- One-sample T-test
- Elegir la variable sfat y arrastrarla a sample variables
- Enviar las variables alco y cal que estaban en sample variables a variables
- Null hypothesis: 22
- Althernative hypothesis: Greater Than

- 2) Plantee un test de nivel 5% para confirmar si el consumo medio de alcohol de esta población es mayor a 8. ¿Cuál es su conclusión? ¿Cuál es el p-valor?

- Statistics
- One,two, Multi-sample tests
- One-sample T-test
- Elegir la variable alco y arrastrarla a sample variables

- Enviar las variables sfat y cal que estaban en sample variables a variables
  - Null hypothesis: 8
  - Althernative hypothesis: Greater Than
- 3) Plantee un test de nivel 5% para ver si la media del total de calorías consumidas por esta población es mayor a 1400. ¿Cuál es su conclusión? ¿Cuál es el p-valor?
- Statistics
  - One,two, Multi-sample tests
  - One-sample T-test
  - Elegir la variable alco y arrastrarla a sample variables
  - Enviar las variables sfat y cal que estaban en sample variables a variables
  - Null hypothesis: 1400
  - Althernative hypothesis: Greater Than

## Ejercicio 2

Se quiere confirmar la eficacia de gotas oculares para aumentar el “tiempo de ruptura del lagrimal” (TBUT), que normalmente se ve reducido en pacientes con síndrome de ojo seco. Los datos corresponden a la diferencia en TBUT entre el momento inmediato posterior a poner las gotas y la medición hecha antes de colocarlas (difinm) y a la diferencia del TBUT a los 10 minutos de colocadas las gotas con respecto a la medición anterior a la colocación (difpost), en 14 pacientes con ese síndrome. Los datos se encuentran en el archivo **TEARS.SX**

- a) Plantee un test para confirmar si la diferencia de TBUT, inmediatamente después de haber colocado las gotas es mayor que cero, es decir si el tiempo de ruptura se incrementa con las gotas, inmediatamente después de colocarse.
- b) Plantee lo mismo pero con respecto al comportamiento a los 10 minutos de colocadas las gotas.
- c) Encuentre intervalos de confianza de nivel 0.95 para las variables difinm y difpost. ¿le parece razonable suponer que no hay diferencias entre el TBUT antes de las gotas y después de las gotas?

### Ejercicio 3

8 personas experimentaron un episodio de intoxicación por vitamina D. Se sugirió que en estos casos puede haber un exceso de calcio. Se midieron los niveles de calcio y albúmina en sangre de estos 8 individuos. Para individuos sanos, el nivel medio de calcio debe ser 2.43 y el de albúmina 43.5.

- a) ¿cree que las personas que sufren intoxicación por vitamina D tienen niveles normales de calcio y albúmina? Utilice los datos del archivo **VITAD.SX**
- b) Encuentre intervalos de confianza de 95% para las variables calcio y albúmina

### Ejercicio 4

En el archivo **BEBES.SX** se presentan los datos correspondientes a la variable MSB que mide el vigor con que los bebés succionan distintas soluciones: una solución de agua, agua con 10% de sal, agua con 30% de sal, agua con 5% de azúcar y agua con 15% de azúcar. Todas las mediciones fueron hechas en 100 bebés que se encontraban en una nursery.

- a) Encuentre intervalos de confianza para la media de la variable MSB en una solución de agua, una de agua con 10% de sal, agua con 30% de sal, agua con 5% de azúcar y agua con 15% de azúcar.
- b) Defina una nueva variable que mida la diferencia entre la variable MSB en la solución de agua y en la solución de agua con 30% de sal. Plantee un test para ver si existe diferencia en la succión entre la solución de agua y la de agua con 30% de sal. ¿Cuál es el p-valor de su conclusión?
- c) Repita lo mismo para comparar la variable MSB en la solución de agua con la solución de azúcar al 15%.