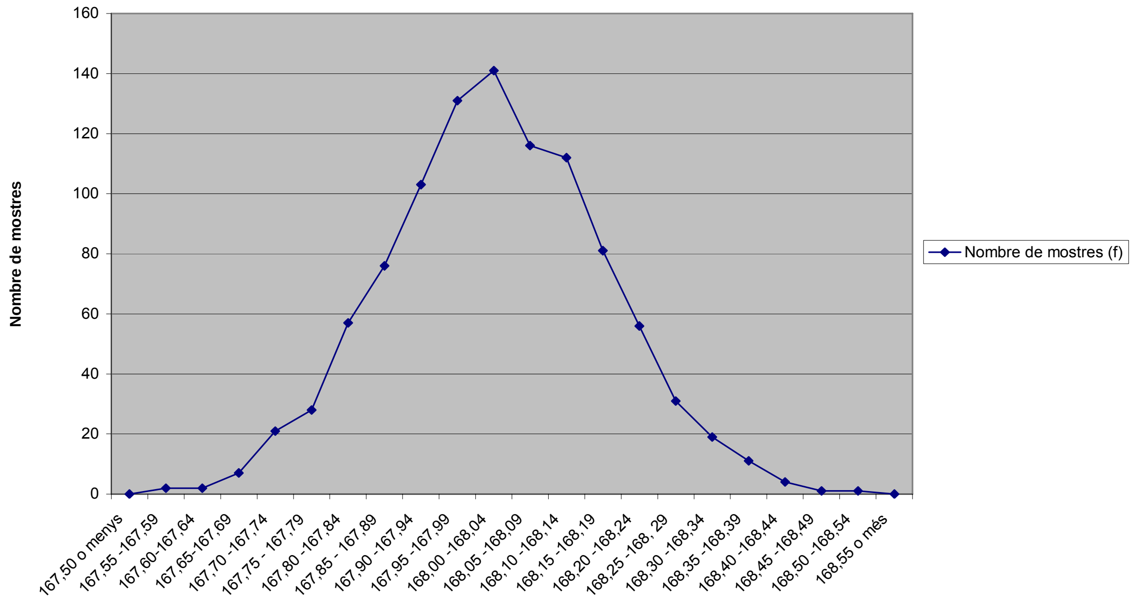
5. Inferències sobre la mitjana d'una població

Tot el que hem explicat al capítol d'introducció a l'estadística inferencial ens serveix per entendre com podem fer inferències a la població a partir de l'estudi d'una mostra. Recordem que al capítol 4 ja hem vist la distribució mostral de molts estadístics, entre els quals ens fixarem en les mitjanes. Així, hem observat que, si escollim un nombre elevat de mostres d'una població, totes de la mateixa mida, busquem la mitjana de cadascuna i representem gràficament la distribució de freqüències d'aquestes mitjanes, la gràfica resultant s'aproxima molt a la corba normal. El valor que té més freqüències d'aquesta distribució el podem considerar la mitjana de la població.

Per exemple, hem vist que representant gràficament la distribució mostral de les mitjanes d'alçada de 1.000 mostres de 100 persones cada una, ens sortia el gràfic següent:

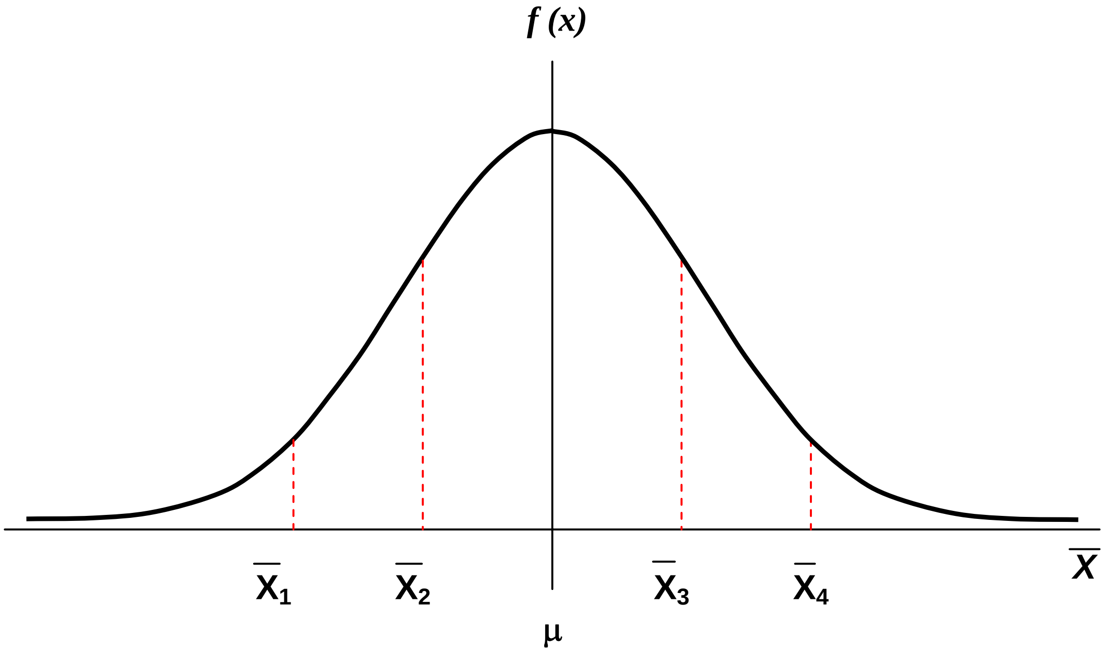
Gràfic 5.1: Distribució hipotètica de le alçades mitjanes de 1000 mostres de N = 100



Si en lloc d'extreure mil mostres n'extraiéssim moltes més —fins a un nombre infinit—, la forma de la distribució mostral de mitjanes seguiria el model de la corba normal. El nombre més elevat de freqüències correspondria a la mitjana de les mitjanes, que podríem considerar la mitjana de la població.

Però abans d'efectuar inferències sobre la mitjana de la població, cal tenir una mesura de dispersió d'aquesta distribució. Si la variable que estem analitzant és l'alçada de les persones adultes, i sabem que la mitjana és de 168 cm, podem entendre que és molt poc probable que una persona seleccionada a l'atzar tingui una alçada de 157,00 cm. Però encara és molt menys probable que una mostra de N = 100 seleccionada tingui una mitjana d'alçada de 157 cm. És a dir, *la distribució mostral de mitjanes és molt menys dispersa que la distribució de freqüències de les puntuacions directes d'una variable*. Dit d'una altra manera, un valor extrem d'una mitjana d'una mostra és menys probable que el mateix valor d'una puntuació directa. Com més gran sigui la *N* de les mostres amb què treballem per fer la distribució mostral, menys dispersió hi haurà en aquesta distribució.

Gràfic 5 2. Distribució mostral de mitjanes



Per calcular la dispersió de la distribució mostral de mitjanes podríem utilitzar la fórmula que ja coneixem de la desviació típica, tenint en compte que els valors serien les diferents mitjanes. No obstant això, normalment els investigadors només disposen d'una mostra, i és aquesta mostra la que ha de servir per fer inferències. Les teories estadístiques mostren que la dispersió d'una distribució de mitjanes mostrals a partir d'una sola mostra escollida aleatòriament, de mida *N*, es pot calcular de la manera següent:

$$S_x = \frac{S}{\sqrt{N}} \qquad (11)$$

S = desviació típica de la mostra
N = nombre de casos de la mostra

Aquesta fórmula s'anomena *error típic de la mitjana*.

L'error típic de la mitjana (S_x) és menor que la desviació típica de les puntuacions directes (*S*). Com més gran sigui la mostra més petit serà l'error típic, és a dir, menor serà la dispersió de la distribució mostral de mitjanes. Es fa servir la paraula *error* perquè expressa la idea que la diferència entre la mitjana poblacional i la mitjana d'una mostra de la població que s'escull aleatòriament és un error produït pel mostreig, és a dir, per l'atzar.

Seguint amb l'exemple de les alçades:

Imaginem-nos que tenim una mostra de N = 100 homes dels països escandinaus i que s'ha calculat que la seva alçada mitjana és de 169 cm i que la desviació típica d'aquesta mostra és de 7 cm. Llavors, l'error típic de la mitjana el podem calcular de la manera següent:

$$S_x = \frac{S}{\sqrt{N}} = \frac{7}{\sqrt{100}} = 0{,}7$$

Se sap també que la mitjana de l'alçada de la població d'homes d'Europa és de 168 cm.

Ara ens pot interessar esbrinar si l'alçada mitjana dels homes dels països escandinaus és diferent de la població completa d'homes adults d'Europa. Sense mesurar l'alçada de tots els homes dels països escandinaus no podem respondre amb certesa a aquesta pregunta, però podem utilitzar la mitjana mostral i l'error típic de la mitjana per fer inferències sobre la població. Hi ha dues possibilitats:

a) La mostra de mitjana 169 cm prové d'una població amb mitjana 168 cm, i per tant la diferència d'1 cm és deguda a l'error de mostreig.

b) La mostra no prové d'una població amb mitjana 168 cm sinó d'una amb qualsevol altra mitjana.

Per decidir entre aquestes dues alternatives ens caldria saber fins a quin punt és estrany obtenir una mostra de mitjana $\overline{X} = 169$ d'una població amb una $\mu = 168$. Quina de les dues alternatives —*a* o *b*— és la més encertada?

5.1. Comprovació d'hipòtesis

En l'exemple de la moneda vàrem veure que calia calcular la probabilitat que per atzar pogués sortir 6 vegades cara en 6 llançaments consecutius. Vàrem deduir que això era bastant improbable prenent com a regla de decisió el 0,05, és a dir, vàrem rebutjar la hipòtesi 0 (la que establia que la moneda era correcta) i vàrem acceptar la hipòtesi 1 (la que establia que la moneda estava trucada). El procediment és molt semblant en el cas de les alçades: cal determinar la probabilitat d'escollir una mostra amb una mitjana de 169 cm, sent la mitjana poblacional de 168, i comparar aquesta probabilitat amb una regla de decisió.

5.1.1. Les hipòtesis nul·la i alternativa

El primer pas consisteix a establir la hipòtesi nul·la, que simbolitzarem amb H_0, que especifica la suposició sobre el paràmetre poblacional. En l'exemple que ens ocupa la hipòtesi nul·la implica que la mostra, amb mitjana = 169 cm, és una mostra aleatòria d'una població amb μ = 168 cm.

A continuació es formula la hipòtesi alternativa, que simbolitzarem amb una H_1. Aquesta hipòtesi estableix que la mitjana poblacional de la qual s'ha extret la mostra amb mitjana 169 cm, no és 168 sinó qualsevol altre valor. És a dir, la diferència d'1 punt entre la mitjana de la mostra i la de la població no pot ser deguda a l'atzar.

Cal tenir en compte que les proves estadístiques no determinen mai si la hipòtesi nul·la és veritable o és falsa. Una prova estadística indica si la H_0 és prou probable o no, segons la regla de decisió que s'estableixi. Sense mesurar la població completa és impossible provar res, i la decisió final que es pren pot ser correcta o incorrecta. De fet, existeixen quatre possibilitats:

1. La mitjana poblacional és realment 168 (és a dir, la mostra procedeix d'una població amb μ = 168). El fet de rebutjar una H_0 veritable s'anomena *error de tipus I,* i la probabilitat (o risc) de cometre aquest error se simbolitza amb la lletra grega alfa (α).

2. La mitjana poblacional no és realment 168 i incorrectament acceptem H_0. L'error comès en acceptar una hipòtesi nul·la que en realitat és falsa s'anomena *error de tipus II*, i la probabilitat (risc) de cometre'l se simbolitza amb la lletra grega beta (β).

3. La mitjana poblacional no és realment 168 i es rebutja correctament H_0. La probabilitat d'assolir aquesta decisió correcta s'anomena *potència* de la prova estadística i és igual a 1 - β.

4. La mitjana poblacional és realment 168 i s'accepta correctament H_0. La probabilitat de prendre aquesta decisió correcta és igual a 1 - α .

Aquestes quatre possibilitats estan resumides a la taula 5.1.

Taula 5.1: Model de comprovació d'hipòtesis estadístiques

		Situació real o estat de la població	
		H_0 és realment certa	**H_0 és realment falsa**
El resultat de l'experiment aconsella:	Acceptar H_0	*Decisió correcta* La probabilitat d'acceptar la H_0 certa és 1 - α.	*Error de tipus II* La probabilitat (risc) d'acceptar una H_0 falsa és β.
	Rebutjar H_0	*Error de tipus I* La probabilitat (risc) de rebutjar una H_0 certa és α.	*Decisió correcta* La probabilitat de rebutjar una H_0 falsa (potència) és 1 - β.

Aquests dos tipus d'error no es poden eliminar ja que són complementaris, és a dir, encara que reduíssim al màxim la probabilitat de cometre un error de tipus I, el que ens passaria és que augmentaria la probabilitat de fer un error de tipus II, i a l'inrevés. Si disminuïm el risc de rebutjar una hipòtesi nul·la que és correcta, augmenta el risc d'acceptar una hipòtesi nul·la que és falsa.

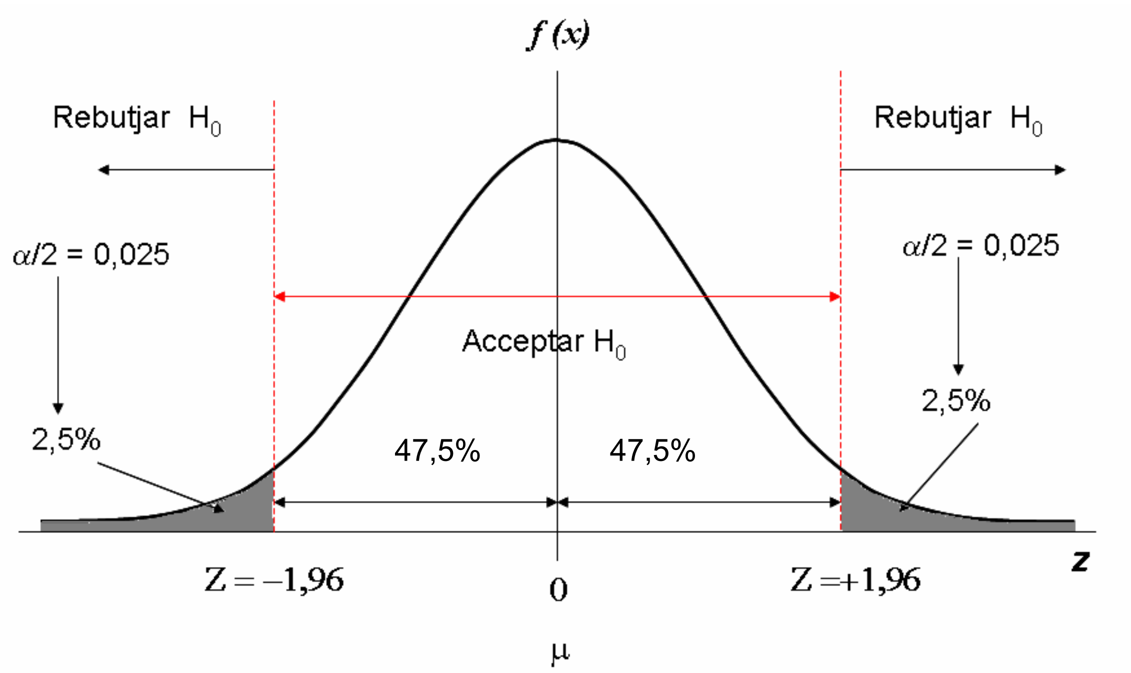
5.1.2. Nivell de significació

Una vegada establertes les hipòtesis nul·la i alternativa, es necessita una regla que indiqui com és d'improbable que una determinada mitjana mostral es produeixi, per rebutjar la H_0 i acceptar la H_1. Ja hem vist, en parlar de l'estratègia general de l'estadística inferencial, que no hi ha regles absolutes per establir valors a partir dels quals les mitjanes mostrals s'hagin de considerar prou diferents per rebutjar la H_0.

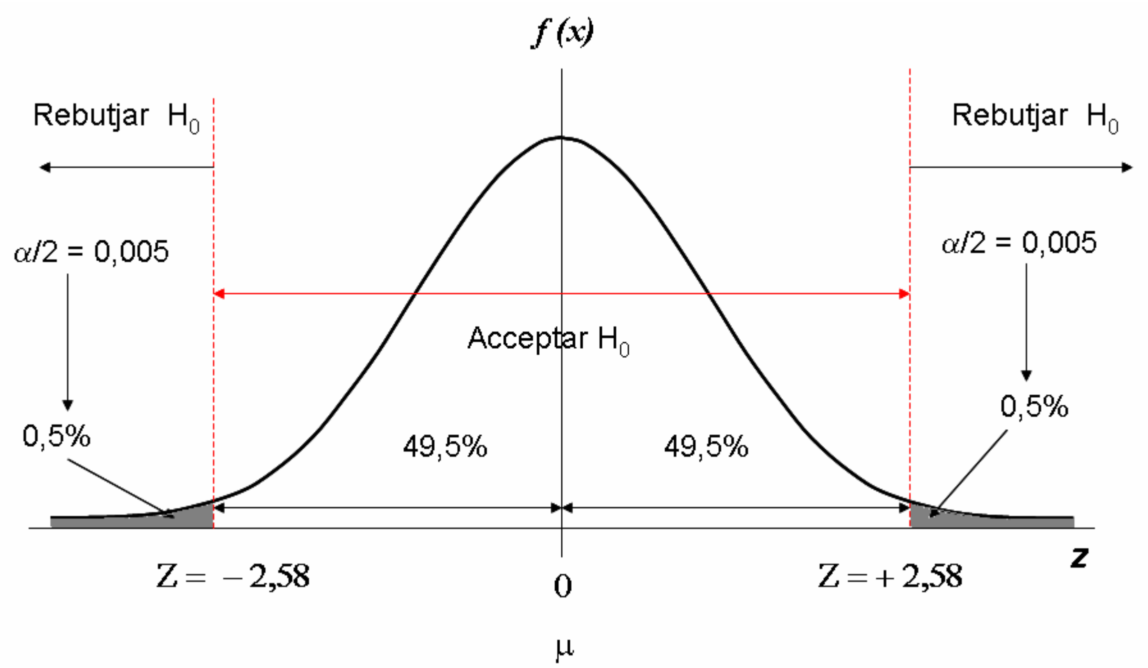
El valor numèric que s'especifica com a regla de decisió rep el nom de *nivell de significació*. Així, quan s'escull la "regla de decisió del 0,05", s'està emprant un nivell de significació del 5 %, és a dir, s'està dient que, si s'obté una probabilitat igual o menor al 0,05, es considerarà prou improbable i, per tant, es rebutjarà la H_0. El nivell de significació s'expressa amb $\alpha = 0{,}05$. Tal com hem dit, aquest valor α indica la probabilitat de cometre un error de tipus I, és a dir, indica el risc que tenim de rebutjar una H_0 que sigui veritable.

Als gràfics 3 i 4 trobem representats els errors de tipus I per $\alpha = 0{,}05$ i $\alpha = 0{,}01$. Cal tenir en compte que l'error α es reparteix a ambdós costats de la mitjana, a cada una de les cues de la corba normal.

Gràfic 5.3: Nivell de significació per α = 0,05



Gràfic 5.4: Nivell de significació per α = 0,01



El nivell de significació de l'l,1 % és més restrictiu en el rebuig de H_0 que el del 5 %, ja que només les mitjanes mostrals molt extremes es consideraran prou improbables. A més, té l'avantatge de reduir l'error de tipus I. No obstant això, com a conseqüència, el nivell de significació de l'l,1 % és menys exigent en l'acceptació de H_0, per la qual cosa augmenta la probabilitat de l'error de tipus II (és a dir, augmenta el risc d'acceptar una H_0 que és falsa).

5.2. Prova estadística per a la mitjana d'una població quan σ és coneguda

Tornem ara a l'exemple de les alçades. Suposem que la desviació típica de la població és 6 i que volem utilitzar la regla de decisió del 0,05, és a dir, un nivell de significació de 0,05.

El primer que cal fer és tipificar el valor de la mitjana mostral. Recordem que la mitjana mostral que havíem suposat era de 169 cm, per a una mostra de *N* = 100. Recordem també que tipificar significa convertir els valors observats als valors *Z*. Com que no es tracta ara de tipificar una puntuació directa d'alçada sinó de tipificar la mitjana d'una mostra, haurem d'utilitzar l'error típic de la mitjana com a mesura de dispersió. La fórmula que haurem d'emprar és la següent:

$$Z = \frac{\overline{X} - \mu}{\sigma_x} \qquad (12)$$

$\overline{X}$ = valor observat de la mitjana mostral
μ = valor de la mitjana poblacional
σ_x = error típic de la mitjana quan la mitjana de la població és coneguda ($= \sigma / \sqrt{N}$)

Com que $\sigma = 6$

$$\sigma_x = \frac{6}{\sqrt{100}} = 0,6$$

$$Z = \frac{169 - 168}{0,6} = 1,67$$

Veiem que el valor +1,67 cau entre els valors -1,96 i +1,96, que són els valors crítics quan $\alpha = 0,05$. És a dir, que una $Z = 1,67$ té una probabilitat que cau dins de l'àrea d'acceptació de la hipòtesi nul·la. Per tant, és probable obtenir, per a un nivell de significació del 0,05, una mostra de 100 casos amb una mitjana de 169 d'una població que té una mitjana de 168.

Suposem ara que la desviació típica de la població fos 4 i la mitjana, 168. És probable obtenir una mostra de 100 casos amb una mitjana de 169 d'aquesta població?

Busquem primer l'error típic de la mitjana:

$$\sigma_x = \frac{4}{\sqrt{100}} = 0,4$$

Ara tipifiquem el valor de la mitjana:

$$Z = \frac{169 - 168}{0,4} = 2,5$$

En aquest cas, el valor +2,5 queda fora de l'àrea d'acceptació compresa entre els valors -1,96 i +1,96. Per tant, en aquest cas rebutgem la H_0 i deduïm que és molt poc probable obtenir una mostra de 100 casos amb una mitjana de 169 d'una població que té una $\mu = 168$ i una $\sigma = 4$. Una altra manera de dir-ho és que la diferència entre la mitjana de la mostra i la de la població és estadísticament significativa per a un nivell de significació del 0,05.

5.3. Prova estadística per a la mitjana d'una població quan σ és desconeguda

Cal tenir en compte que quan es treballa amb mostres el més habitual és que no coneguem la desviació típica de la població. La majoria de vegades és necessari calcular una estimació de l'error típic de la mitjana a partir de la desviació típica de la mostra. En aquest cas, cal utilitzar la fórmula (11):

$$S_x = \frac{S}{\sqrt{N}}$$

I per tipificar els valors de les mitjanes mostrals s'utilitza :

$$t = \frac{\overline{X} - \mu}{S_x} \qquad (13)$$

Els valors obtinguts amb aquesta fórmula no estan distribuïts seguint el model de la corba normal, encara que la corba normal constitueix una bona aproximació si la mida de la mostra és més gran de 30 casos. El procediment de la corba normal podria portar a respostes errònies perquè en aquesta fórmula s'està utilitzant un estimador de l'error típic de la mitjana, calculat a partir de la desviació típica de la mostra.

El model que s'utilitza en aquest cas, és a dir, quan no coneixem σ de la població i tenim menys de 30 casos en la mostra, és la *distribució t*. Existeix una distribució *t* per a cada mida de les mostres. Els matemàtics ja han tabulat les diferents probabilitats per les corbes corresponents a cada possible mida de la mostra.

La distribució t

La distribució *t* la trobem en unes taules, que estan expressades en funció dels graus de llibertat (g. ll.) i no segons la mida de la mostra. Els graus de llibertat són, en cada cas, iguals a la mida de la mostra menys 1:

$$\text{g. ll.} = N - 1$$

A la taula 5.2 es mostra un fragment de la taula de valors crítics de *t*. Cada fila horitzontal representa una distribució diferent, que correspon als graus de llibertat. A l'annex 3 trobareu la taula de valors crítics de *t* completa.

Tabla C. *Valores críticos de t (*)*

	Nivel de significación para una prueba unilateral					
gl	0,10	0,05	0,025	0,01	0,005	0,0005
	Nivel de significación para una prueba bilateral					
	0,20	0,10	0,05	0,02	0,01	0,001
1	3,078	6,314	12,706	31,821	63,657	636,619
2	1,886	2,920	4,303	6,965	9,925	31,598
3	1,638	2,353	3,182	4,541	5,841	12,941
4	1,533	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,859
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,449	5,405
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	$\boxed{2,262}$	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587

Taula 5.2: Fragment de la taula de valors crítics de t

Exemple

Imaginem-nos que agafem una mostra de 10 persones adultes i els mesurem l'alçada. Trobem que la mitjana és de 166 cm i la desviació típica, de 7 cm. Volem saber si aquesta mostra pot pertànyer a la població en què la mitjana d'alçada és de 168 cm.

El primer que farem serà una tipificació, però, com que N = 10, no podem convertir-ho a un valor de Z, i per tant utilitzarem els valors t.

- En primer lloc, formularem les hipòtesis nul·la i alternativa:

H_0: la mostra pertany a la població. Podem simbolitzar-ho dient que $\overline{X} = \mu$
H_1: la mostra no pertany a la població. $\overline{X} \neq \mu$

- Tot seguit buscarem l'error típic de la mitjana, utilitzant l'expressió (11):

$$S_{\overline{x}} = \frac{S}{\sqrt{N}} = \frac{7}{\sqrt{10}} = 2{,}21$$

- A continuació tipificarem la mitjana mostral, utilitzant la fórmula (13):

$$t = \frac{\overline{X} - \mu}{S_x} = \frac{166 - 168}{2,21} = -0,90$$

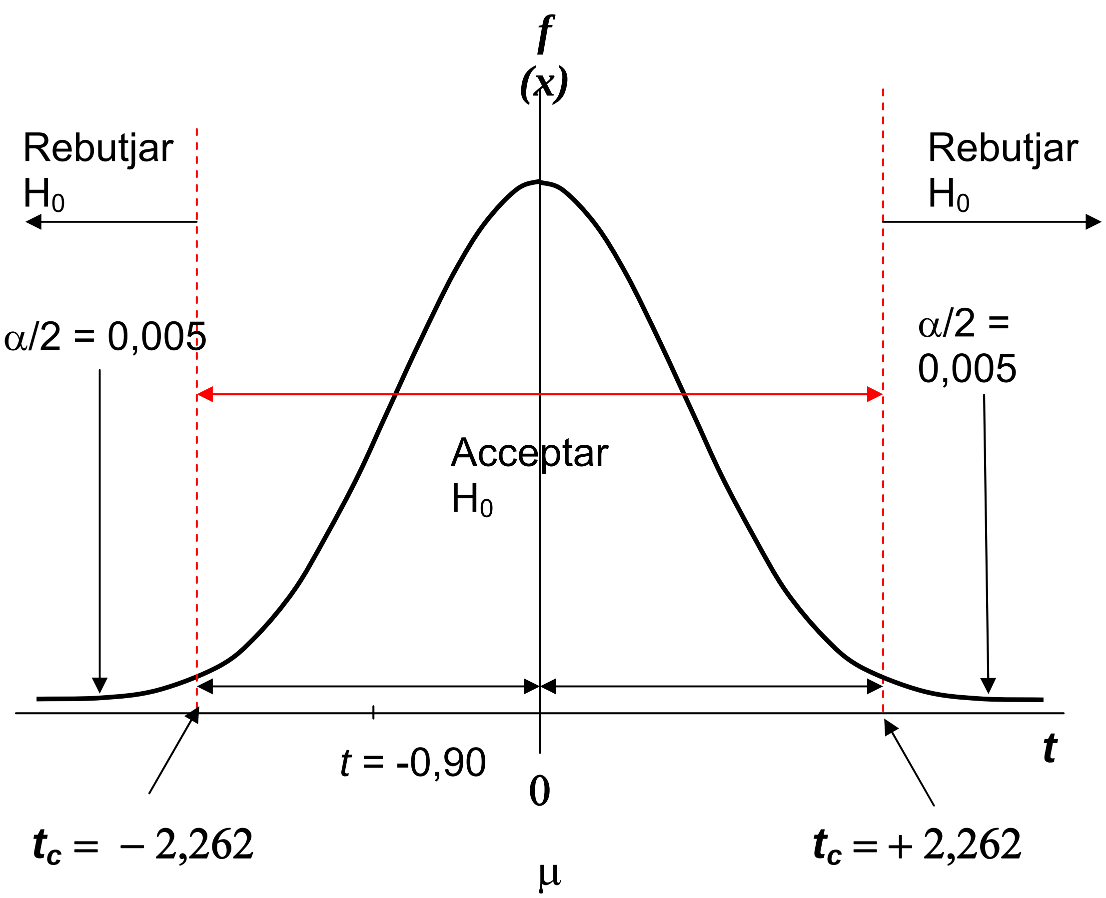
- Ara hem de buscar els valors crítics de *t*, per poder comparar el valor que ens ha sortit. Farem la interpretació per a un nivell de significació de α = 0,05.

Mirem les taules i busquem el valor crític:

t_c (g. ll. = N - 1; α = 0,05)
t_c (g. ll. = 9; α = 0,05) = 2,262

El gràfic 5.5 ens pot ajudar en la interpretació. Com que el valor de la *t* que hem trobat és dins de l'interval definit per $-t_c$ i $+t_c$, podem dir que és dins de la zona d'acceptació de la H_0.

Gràfic 5.5: Situació dels valors crítics de t *i de la* t *observada en l'exemple*



- Per tant, amb un nivell de significació de 0,05, acceptem la hipòtesi nul·la, és a dir, que és possible obtenir una mostra aleatòria de $N = 10$ amb una mitjana de 166 cm d'una població que té una mitjana de 168 cm. Dit d'una altra manera, tenim un 95 % de probabilitat que, en escollir una mostra a l'atzar de 10 persones adultes d'una població que té de mitjana d'alçada 168 cm, ens surti que la mitjana de la mostra és de 166 cm.

Si observem les taules de valors crítics de *t* veurem que ,com més petita és la mida de la mostra, més gran ha de ser *t* per rebutjar H_0, amb el mateix nivell de significació. També es veu que, quan la mida de la mostra és gran, les distribucions *Z* i *t* són equivalents. Així, per exemple, per 40 graus de llibertat, la t_c és 2,02 (per $\alpha = 0{,}05$), mentre que per a 120 graus de llibertat, $t = 1{,}98$. Aquest valor ja s'aproxima molt al de $Z = 1{,}96$.

RESUM

La tipificació de valors de les mitjanes mostrals ens permet saber si una mostra pot pertànyer a una població determinada, amb mitjana μ.

Quan μ i σ són conegudes

- Tipificarem el valor de la mitjana de la mostra amb la fórmula:

$$Z = \frac{\overline{X} - \mu}{\sigma_x} \qquad (12)$$

$\overline{X}$ = valor observat de la mitjana mostral
μ = valor de la mitjana poblacional
σ_x = error típic de la mitjana quan la mitjana de la població és coneguda ($= \sigma / \sqrt{N}$)

- El valor de *Z* que surti el compararem amb els valors corresponents de la corba normal. Per $\alpha = 0{,}05$ utilitzarem el valor $Z = 1{,}96$; i per $\alpha = 0{,}01$ utilitzarem els valors de $Z = 2{,}58$. No és necessari fer el dibuix cada vegada.
  - Per a $\alpha = 0{,}05$, si el valor de *Z* que ens surt, en valor absolut, és més petit que 1,96, acceptarem H_0, i si és més gran rebutjarem H_0 i acceptarem H_1.
  - Per a $\alpha = 0{,}01$, si el valor de *Z* que ens surt, en valor absolut, és més petit que 2,58, acceptarem H_0 i si és més gran rebutjarem H_0 i acceptarem H_1.

Quan μ és coneguda i σ és desconeguda

- En general, tendirem a utilitzar els valors de la taula *t* per decidir si acceptem la H_0 o si la rebutgem. Utilitzarem la fórmula:

$$t = \frac{\overline{X} - \mu}{S_x} \qquad (13)$$

$\overline{X}$ = valor observat de la mitjana mostral
μ = valor de la mitjana poblacional
S_x = error típic de la mitjana quan la mitjana de la població és desconeguda

$$S_x = S / \sqrt{N} \qquad (11)$$

- A l'hora de fer la interpretació, no cal que fem el dibuix cada vegada, sinó que compararem el valor de *t* que ens ha sortit amb el valor de la t_c, en valor absolut. Tenint en compte que els valors de les taules ens indiquen el nivell a partir del qual, per marge d'error determinat, la diferència entre la mitjana mostral i la poblacional comença a ser significativa, la interpretació la farem de la manera següent:
    - Si t < o = t_c, acceptarem H_0
    - Si $t > t_c$, rebutjarem H_0

5.4. Interval de probabilitat o interval de confiança

Una altra manera de resoldre el problema d'acceptar o rebutjar la H_0 consisteix a estimar un interval en el qual és probable que trobem la mitjana de la població. Si el valor de μ expressat per la hipòtesi nul·la no cau dins de l'interval, es rebutja H_0. *L'interval de probabilitat o de confiança* és l'interval de valors en què és més probable trobar la mitjana de la població de la qual s'ha extret la mostra. Així, l'exemple que hem posat a l'apartat anterior el podem resoldre també buscant la mitjana de la població de la qual s'ha extret la mostra de 10 casos, amb una mitjana de 166 cm. No podem trobar el valor exacte de la mitjana poblacional, però podem estimar un interval de valors en què és més probable —en què podem tenir més confiança de— trobar la mitjana de la població. Si la mitjana coneguda de la població μ = 168 es troba dins de l'interval, podrem acceptar la H_0, és a dir, podrem concloure que la mostra pertany a una població amb una mitjana de 168 cm.

Vegem el procediment que cal seguir, posant un altre supòsit.

Exemple

Volem saber si és possible obtenir una mostra de 20 casos amb una mitjana de 167 cm d'alçada i una desviació típica S = 5, d'una població que té de mitjana μ = 168 cm. Volem interpretar-ho amb un nivell de significació del 95 %.

Aquesta vegada optarem pel procediment de l'interval de probabilitat. Buscarem la mitjana de la població de la qual s'ha extret la mostra.

1. Com que la mostra és petita, anirem a les taules de valors crítics de t per buscar el valor corresponent a 20-1 graus de llibertat i α = 0,05:

   t_c (19 g. ll.; μ = 0,05) = 2,093

2. Busquem l'error típic de la mitjana:

$$S_x = \frac{S}{\sqrt{N}} = \frac{5}{\sqrt{20}} = 1{,}12$$

3. Substituïm la fórmula que ja coneixem pels valors que sabem. Com que estem buscant el valor de la mitjana de la població de la qual s'ha extret la mostra, la μ constitueix la incògnita, el valor que busquem.

$$t = \frac{\overline{X} - \mu}{S_x}$$

   Cal tenir en compte que, com que estem treballant sempre amb proves bilaterals, el valor de t_c pot adoptar el signe positiu o el negatiu. Per tant,

$$+2{,}093 = \frac{167 - \mu_1}{1{,}12}$$

$$-2{,}093 = \frac{167 - \mu_2}{1{,}12}$$

4. Si aïllem la mitjana de la població en cada una d'aquestes fórmules, tindrem el següent:

   $\mu_1 = 167 – 1{,}12 \cdot 2{,}093 = 164{,}66$
   $\mu_2 = 167 + 1{,}12 \cdot 2{,}093 = 169{,}34$

5. Fem la interpretació:

   164,66 < 168 < 169,34

   Com que la mitjana suposada de la població es troba dins de l'interval, amb un nivell de confiança del 95 % podem acceptar la H_0. És a dir, és probable que una mostra de 20 casos, amb una mitjana de 167 i una desviació típica de 5, l'hàgim tret d'una població amb una mitjana de 168.

En general, la fórmula que hem d'utilitzar per calcular intervals de probabilitat quan σ és desconeguda és:

$$\boxed{\overline{X} - t_c S_x < \mu < \overline{X} + t_c S_x} \qquad (14)$$

t_c = valor crític de t per N-1 graus de llibertat i α
$\overline{X}$ = valor observat de la mitjana mostral
S_x = error típic de la mitjana (= $S/\sqrt{N}$)

Qualsevol valor per sobre o per sota d'aquest interval es considera improbable. Per això, si el valor que comparem cau dins de l'interval, acceptarem la hipòtesi nul·la, i si cau fora, rebutjarem aquesta hipòtesi.

Si les mostres són grans

$$\boxed{\overline{X} - Z S_x < \mu < \overline{X} + Z S_x} \qquad (15)$$

Z = valor de Z segons el marge d'error α
$\overline{X}$ = valor observat de la mitjana mostral
S_x = error típic de la mitjana (= $S/\sqrt{N}$)

5.5. Exercicis d'inferències sobre la mitjana d'una població

1. Es vol saber quina és la mitjana d'edat dels alumnes que assisteixen a la Universitat. Per això s'ha agafat una mostra aleatòria de 400 alumnes i ens dóna una mitjana de 24,3 anys i una desviació típica de 12,8 anys. Calcula els resultats en els nivells de confiança del 95 % i del 99 %.

2. Sabem que la nota mitjana del primer quadrimestre dels estudiants de l'assignatura Bases metodològiques de la investigació educativa del curs 2005-2006 és de 6,02 i que la desviació típica és d'1,58. Aquesta mostra està composta per 63 estudiants. Quina serà la mitjana de puntuació de les notes del primer

quadrimestre que podem inferir per als estudiants d'aquesta assignatura dels cursos vinents? Utilitza un nivell de significació del 0,05.

3. Es vol conèixer la mitjana d'euros que es gasten els estudiants de la UdG per esmorzar cada setmana. S'ha seleccionat una mostra de 250 estudiants i se'ls ha demanat aquesta dada. La mitjana d'euros que es gasten a la setmana és de 5 euros, amb una desviació típica d'1,5. Respon a les qüestions següents:

   3.1. Amb una confiança del 95 %, entre quins valors es troba la mitjana d'euros que els estudiants es gasten per esmorzar cada setmana?
   3.2. Quina és la probabilitat d'obtenir una mitjana mostral de 5 euros/setmana o més quan la mitjana de la població és igual a 4,8?
   3.3. Podem admetre a un nivell de confiança del 0,05 que la nostra mostra pertany a la població amb una mitjana de 4,8?
   3.4. Calcula de nou l'apartat 1 a un nivell de confiança del 0,01. Amb aquest nivell de confiança, podem dir que la mostra pot pertànyer a una població amb una mitjana de 4,8?

4. Tenim una mostra de 200 estudiants de la FEP, amb una mitjana d'edat de 24 anys. En el nivell de significació del 0,05, podem admetre que aquesta mostra pertany a la població d'estudiants de la FEP amb una mitjana d'edat de 22 i una desviació típica de 4?

5. S'ha agafat una mostra de 300 persones que viuen en pisos de lloguer. Investigada la quantitat que paguen de lloguer, surt una mitjana de 400 € i una desviació típica de 25 €. A partir d'aquestes dades, podem afirmar amb un 95 % de confiança que la mitjana de lloguer per a la població és de 390 €?

6. Se sap que la mitjana d'ingressos anuals dels titulats universitaris de grau mitjà (diplomats) és de 16.000 €. Hem seleccionat una mostra de 30 educadores socials i els hem preguntat sobre els seus ingressos anuals. La mitjana d'ingressos d'aquesta mostra d'educadores és de 15.000 € anuals, amb una desviació típica de 1.500 €. Podem dir, basant-nos en aquest resultat, que els diplomats en Educació Social estan més mal pagats en comparació del conjunt de diplomats universitaris? Interpreta el resultat per a un marge d'error del 0,05.

5.6. Respostes als exercicis d'inferències sobre la mitjana d'una població

Exercici 1

Per a un nivell de confiança del 95 % la mitjana d'edat estaria entre els 23,05 i els 25,55 anys.
Per a un nivell de confiança del 99 % la mitjana d'edat se situaria entre els 22,64 i els 25,95 anys.

Exercici 2

Aquí se'ns demana que busquem l'interval de probabilitat agafant 63 estudiants del curs 2005-2006 com a mostra representativa de la població. A més, no coneixem la desviació típica de la població. Per tant, buscarem l'interval de probabilitat amb la fórmula:

$\overline{X}$ +/- t. $S/\sqrt{N}$

t (g. ll. N -1 = 62; α = 0,05) = 2 (Utilitzem g. ll. = 60, ja que el següent ja és 120).

$6{,}02 + 2.\ 1{,}58/\sqrt{63} = 6{,}42$

$6{,}02 - 2.\ 1{,}58/\sqrt{63} = 5{,}62$

Per tant, amb un marge d'error del 0,05, es pot esperar per als anys vinents que la mitjana del primer quadrimestre estarà entre una puntuació de 5,62 i 6,42.

Exercici 3

3.1. $\overline{X} = 5$ $S = 1{,}5$ $N = 250$
Per α = 0,05 la Z = 1,96 (fem servir Z perquè N és molt gran.)

$\overline{X}$ +/- $Z \cdot$ S / $\sqrt{N}$

$\overline{X} + Z. \cdot S/\sqrt{N} = 5 + 1{,}96 \cdot (1{,}5/\sqrt{250}) = 5{,}19$

$\overline{X} - Z \cdot S/\sqrt{N} = 5 - 1{,}96 \cdot (1{,}5/\sqrt{250}) = 4{,}81$

Resposta: a un nivell de confiança del 95 %, els estudiants de la UdG es gasten entre 4,81 i 5,19 euros de mitjana cada setmana per esmorzar.

3.2. Per saber la probabilitat exacta haurem de tipificar el valor de la mitjana mostral. Les dades que tenim són les següents:

$\overline{X} = 5$ $\mu = 4{,}8$ $S = 1{,}5$ $N = 250$

Desconeixem la desviació típica de la població, per la qual cosa haurem d'utilitzar la desviació típica de la mostra.

$$Z = \frac{\overline{X} - \mu}{S / \sqrt{N}} = \frac{5 - 4{,}8}{1{,}5 / \sqrt{250}} = \frac{0{,}2}{0{,}095} = 2{,}11$$

Ara cal buscar en les taules de la corba normal el valor 2,11. Veiem que per a una Z = 2,17 l'àrea entre aquest valor i la mitjana és del 48,26 %. Com que la pregunta és la probabilitat de tenir una mitjana de 5 euros o més a la setmana, i el que tenim és el percentatge entre la mitjana i Z, hem de restar aquest valor de 50 %:

P (5 o més) = 50 % - 48,26 % = 1,74 %

Per tant, la probabilitat de tenir una mostra de 250 casos amb mitjana de 5 euros o més a la setmana és de l'1,5 %.

3.3. Primer formulem les hipòtesis estadístiques:

H_0: la mostra de 250 casos amb una mitjana de 5 i una desviació típica d'1,5 pertany a una població amb una mitjana de 4,8.
H_1: la mostra no pertany a una població amb una mitjana de 4,8.

En aquest cas, d'entrada veiem que si l'interval de probabilitat es troba entre 4,81 i 5,19, una μ de 4,8 se situa, per poc, fora d'aquest interval. Per tant, rebutgem la hipòtesi nul·la i diem que no és probable que la mostra pertanyi a una població amb una mitjana de 4,8.

3.4.

$\overline{X} = 5$ $\quad S = 1{,}5$ $\quad N = 250$

Per $\alpha = 0{,}01$ la $Z = 2{,}58$ (fem servir Z perquè N és molt gran.)

$$\overline{X} \pm Z \cdot S / \sqrt{N}$$

$$\overline{X} + Z. \cdot S / \sqrt{N} = 5 + 2{,}58 \cdot (1{,}5/\sqrt{250}) = 5{,}24$$

$$\overline{X} - Z \cdot S / \sqrt{N} = 5 - 2{,}58 \cdot (1{,}5/\sqrt{250}) = 4{,}76$$

Amb aquest nivell de confiança sí que podem admetre que la mostra de 250 casos amb una mitjana de 5 i S = 1,5 pot ser extreta d'una població amb μ = 4,8 ja que aquest valor queda dins de l'interval.

Exercici 4

En aquest cas tenim les dades següents:

$\overline{X} = 24$ $\mu = 22$ $\sigma = 4$ $N = 200$ Per $\alpha = 0{,}05$

Formulem les hipòtesis:

H_0: la mostra de 200 estudiants amb una mitjana de 24 pertany a la població amb una mitjana de 22.
H_1: la mostra de 200 estudiants amb una mitjana de 24 no pertany a la població amb una mitjana de 22.

Tipifiquem el valor de la mitjana mostral, tenint en compte que σ és coneguda.

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{N}} = \frac{24 - 22}{4/\sqrt{200}} = \frac{2}{0{,}28} = 7{,}14$$

Fixem-nos que el valor de Z que surt està molt per sobre d'1,96, que és el valor de *Z* per al nivell de confiança del 0,05. Per tant, rebutgem la H_0 i acceptem la hipòtesi alternativa. Això vol dir que a un nivell de confiança del 0,05 és molt poc probable que es pugui extreure una mostra de 200 estudiants de la FEP que tinguin una mitjana d'edat de 24 anys.

I seria probable obtenir una mostra de 20 estudiants amb una mitjana d'edat de 24 anys?

$$t = \frac{\overline{X} - \mu}{\sigma / \sqrt{N}} = \frac{24 - 22}{4\ \sqrt{20}} = \frac{2}{0{,}89} = 2{,}25$$

En aquest cas, com que la mostra és petita, hem de fer una comparació amb els valors de *t*. Busquem la *t* crítica en les taules:

t_c (a 0,05; g. ll. 19) = 2,093

Com que 2,093 < 2,25, podem concloure que també és poc probable obtenir una mostra a l'atzar de 20 casos i que surti una mitjana de 24 anys.

Exercici 5

$\overline{X} = 400$ € $S = 25$ $N = 300$ Per $\alpha = 0{,}05$ la $Z = 1{,}96$ (fem servir *Z* perquè *N* és molt gran.)

Formulem les hipòtesis:

H_0: és molt probable que la mitjana de lloguer de la població sigui de 390 euros. És a dir, la mitjana de la població de la qual s'ha extret la mostra de 300 persones pot ser de 390 euros.
H_1: la mostra de 300 persones amb una mitjana de 400 euros no pertany a una població amb una mitjana de 390 euros.

El que fem és buscar l'interval de probabilitat de la mitjana de la població de la qual s'ha extret la mostra de 300 persones amb una mitjana de 400 € i una desviació típica de 25 €.

$\overline{X}$ +/- $Z \cdot S / \sqrt{N}$

$\overline{X} + Z. \cdot S/\sqrt{N} = 400 + 1,96 \cdot (25 /\sqrt{300}\,) = 400 + 1,96 \cdot 1,44 = 402,82$

$\overline{X} - Z \cdot S / \sqrt{N} = 400 - 1,96 \cdot (25/\sqrt{300}\,) = 400 - 1,96 \cdot 1,44 = 397,18$

Per tant, a un nivell de confiança del 95 %, rebutgem la H_0. Per tant, és molt poc probable que la mitjana de la població sigui de 390 euros, ja que aquest valor es troba per sota del límit inferior de l'interval de probabilitat.

També es pot resoldre aquest exercici tipificant el valor de la mitjana de la mostra, agafant la mitjana de 390 com a mitjana de la població.

$$Z = \frac{\overline{X} - \mu}{S / \sqrt{N}} = \frac{400 - 390}{25/\sqrt{300}} = \frac{10}{1,44} = 6,94$$

Tenint en compte que ens demanen que utilitzem un nivell de confiança del 95 %, el valor de *Z* corresponent a aquest nivell de confiança és d'1,96. El valor de *Z* que ens ha sortit és molt més gran que 1,96. Per tant, haurem de concloure que cal rebutjar la hipòtesi nul·la i que, per tant, acceptem que no és probable, tenint en compte les dades obtingudes amb l'estudi de la mostra, que la mitjana de la població pugui ser de 390 €.

Exercici 6

H_0: La mostra pertany a la població. És a dir, la diferència entre la mitjana de la mostra i la de la població és deguda a l'atzar. Això significa que aquesta diferència no és estadísticament significativa.

H_1: La mostra no pertany a la població. La diferència entre les dues mitjanes no és deguda a l'atzar. És a dir, la diferència entre la mitjana de la mostra i la de la població és estadísticament significativa.

$$t = \frac{\overline{X} - \mu}{\sigma / \sqrt{N}} = \frac{15.000 - 16.000}{1.500 \ \sqrt{30}} = \frac{-1.000}{273,8} = -3,65$$

t_c (g. ll. 29; a 0,05) = 2,045

-3,65<-2,045

Rebutgem la H_0. La mostra no pertany a la població.

Per tant, segons aquests resultats, podem dir amb un 95 % de confiança que la mitjana d'ingressos anuals de les educadores socials és significativament menor que la mitjana d'ingressos del conjunt de diplomats universitaris.