Szent István Egyetem Állatorvos-tudományi Kar
Biomatematikai és Számítástechnikai Tanszék

12. előadás

Becslések. Hipotézisvizsgálatok.

Fodor János
Biomatematika 2001

Copyright © 2001 jfodor@univet.hu
Last Revision Date: November 21, 2001 Version 1.10

Table of Contents

1 Becslések **5**

1.1 A populációátlag becslése a mintaátlaggal 6

1.2 A mintaátlag eloszlása 10

1.3 A szórás becslése 18

1.4 Konfidenciaintervallumok 19

- Konfidenciaintervallum a populációátlagra ismert szórás esetén 21
- A minta elemszámának becslése . 29

- Konfidenciaintervallum a populációátlagra ismeretlen szórás esetén 30

2 Hipotézisvizsgálatok 40

2.1 Bevezetés 41

2.2 A null- és az alternatív hipotézis . . . 55

2.3 Az első- és másodfajú hiba 58

2.4 Kétoldali alternatív hipotézis 62

1. Becslések

Ebben a fejezetben arról lesz szó, hogy a mintánkból kiszámított különböző értékek hogyan használhatók a populáció tulajdonságainak becslésére, és mennyire megbízhatóak ezek a becslések.

Az ún. **pontbecslés** esetén a populáció paramétereit a mintából kiszámítható egyetlen értékkel becsüljük. Az **intervallumbecslés** esetén pedig egy olyan intervallumot adunk meg, amelyről tudjuk, hogy az általunk előírt valószínűséggel tartalmazza a populáció becsülni kívánt paraméterét.

1.1. A populációátlag becslése a mintaátlaggal

Példa.

Egy bizonyos sört fél literes üvegekben árulnak. A töltési folyamat nem tökéletes, így apró eltérések adódnak az egyes palackok között. Mivel ezek az apró eltérések nagyon sokféle hatás következtében jönnek létre, feltételezhetjük (matematikai úton bizonyítható), hogy a tényleges űrtartalmak **normális eloszlást** követnek 0.5 l átlagértékkel és - az adatok szerint - 0.01 l-es szórással.

Így, ha valaki vásárol egy üveg sört, akkor 50% a

valószínűsége annak, hogy kevesebb, mint 0.5 l sört vesz, és 50%, hogy több mint 0.5 l-t. Azt is tudjuk, hogy 95% a valószínűsége annak, hogy az üveg tartalma 0.48 és 0.52 l között lesz.

Ezek a számítások a feltételezett valószínűségi eloszláson alapulnak. Valójában senki nem tudja, hogy milyen eloszlást követnek a térfogatok addig, amíg minden egyes üveg sört ki nem bontottunk, és meg nem mértünk. Ehelyett statisztikai megfontolásokat alkalmazhatunk.

Vegyünk egy 10 elemű véletlen mintát a söröspalackokból, és mérjük le őket. A következő mérési

eredményeket kaptuk:

0.512 l, 0.495 l, 0.513 l, 0.504 l, 0.498 l,

0.519 l, 0.515 l, 0.497 l, 0.506 l, 0.491 l.

Nyilvánvalóan a minta átlagát használjuk a populációátlag becslésére:

$$\bar{x} = \frac{0.512 + 0.495 + \ldots + 0.506 + 0.491}{10} = 0.5051.$$

Vegyünk egy másik mintát:

0.502 l, 0.492 l, 0.489 l, 0.502 l, 0.509 l,

0.491 l, 0.511 l, 0.501 l, 0.498 l, 0.503 l

Ekkor az átlag 0.4998 l.

Az átlagok mintáról mintára változnak. A mintavételi eljárásból következik, hogy maga **a mintaátlag is egy valószínűségi változó**. Így meg kell mondanunk, hogy a mintaátlagnak milyen az eloszlása.

Egy sörösüvegben lévő sör mennyisége normális eloszlású valószínűségi változó, $\mu = 0.5$ l-es átlaggal, és $\sigma = 0.01$ l-es szórással. Ezek szerint a 10 mintaelem mind egy-egy ugyanilyen eloszlású valószínűségi változó. A minta átlaga ezek szerint 10 valószínűségi változó összege osztva tízzel.

Nézzük meg kicsit közelebbről a mintaátlag eloszlását általában.

1.2. A mintaátlag eloszlása

Emlékezzünk, hogy egy **statisztika** (statisztikai függvény) a megfigyeléseink függvénye. Mint ilyen, minden egyes mintához egy számot rendel. A mintánk pedig a véletlentől függ, tehát **bármely statisztika is egy valószínűségi változó**. Így viselkedése az eloszlása segítségével írható le.

Legyenek $x_1, x_2, \ldots, x_n$ a mért vagy megfigyelt konk-

rét értékek, átlaguk pedig $\bar{x}$. Ez is egy szám. A megfelelő valószínűségi változót jelölje $\overline{X}$ (a mintanagyság mindig ugyanaz: n).

A populáció átlagát jelölje μ, szórását σ.

Az $\overline{X}$ várható értékét jelölje $\mu_{\overline{X}}$, szórását pedig $\sigma_{\overline{X}}$. Az átlag szórását **standard hibának** (standard error, **SE**) is nevezik. Ekkor az alábbi szabályok érvényesek.

1. $$\mu_{\overline{X}} = \mu.$$

2. $$\sigma_{\overline{X}} = \sigma/\sqrt{n}.$$

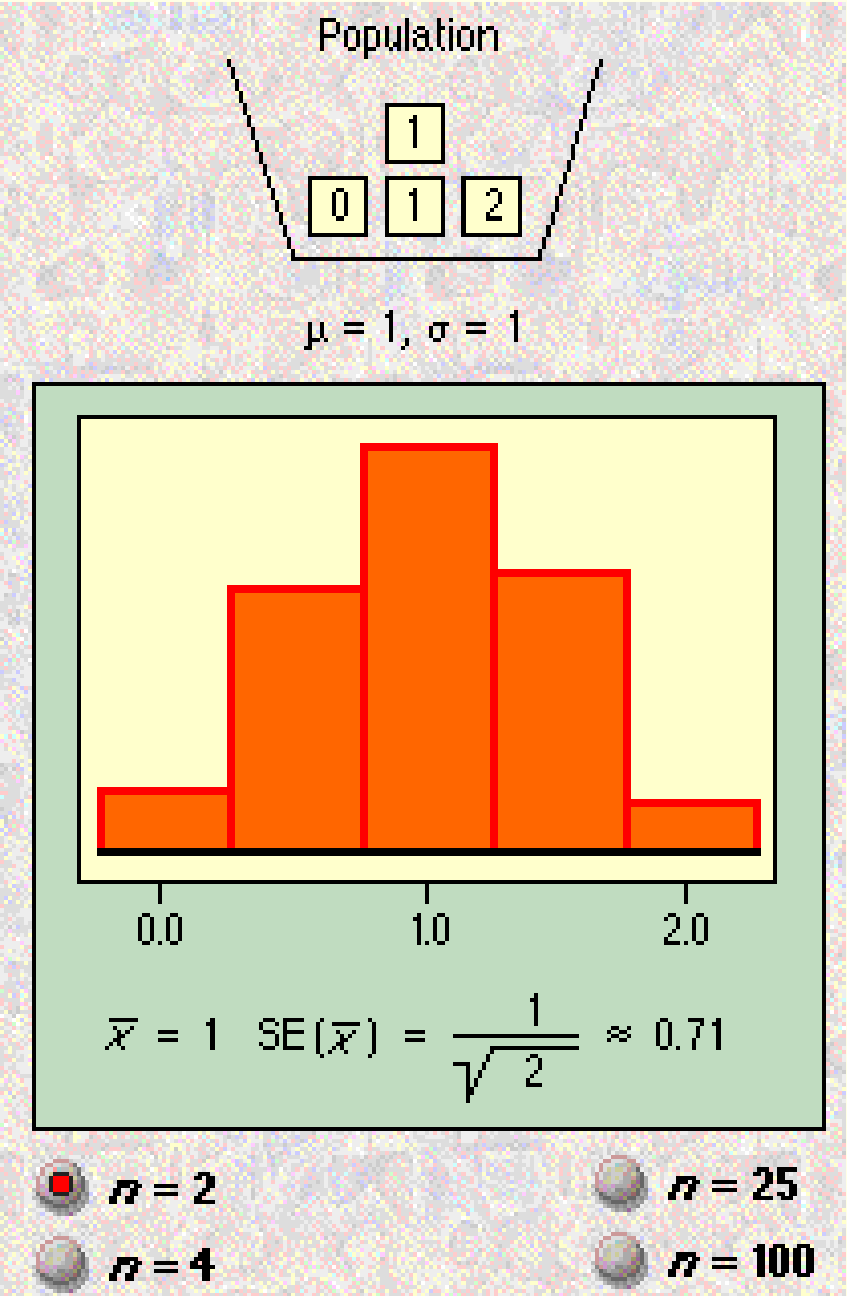
3. Amennyiben a populáció eloszlása normális, akkor $\overline{X}$ eloszlása is normális bármely n mintanagyság esetén.

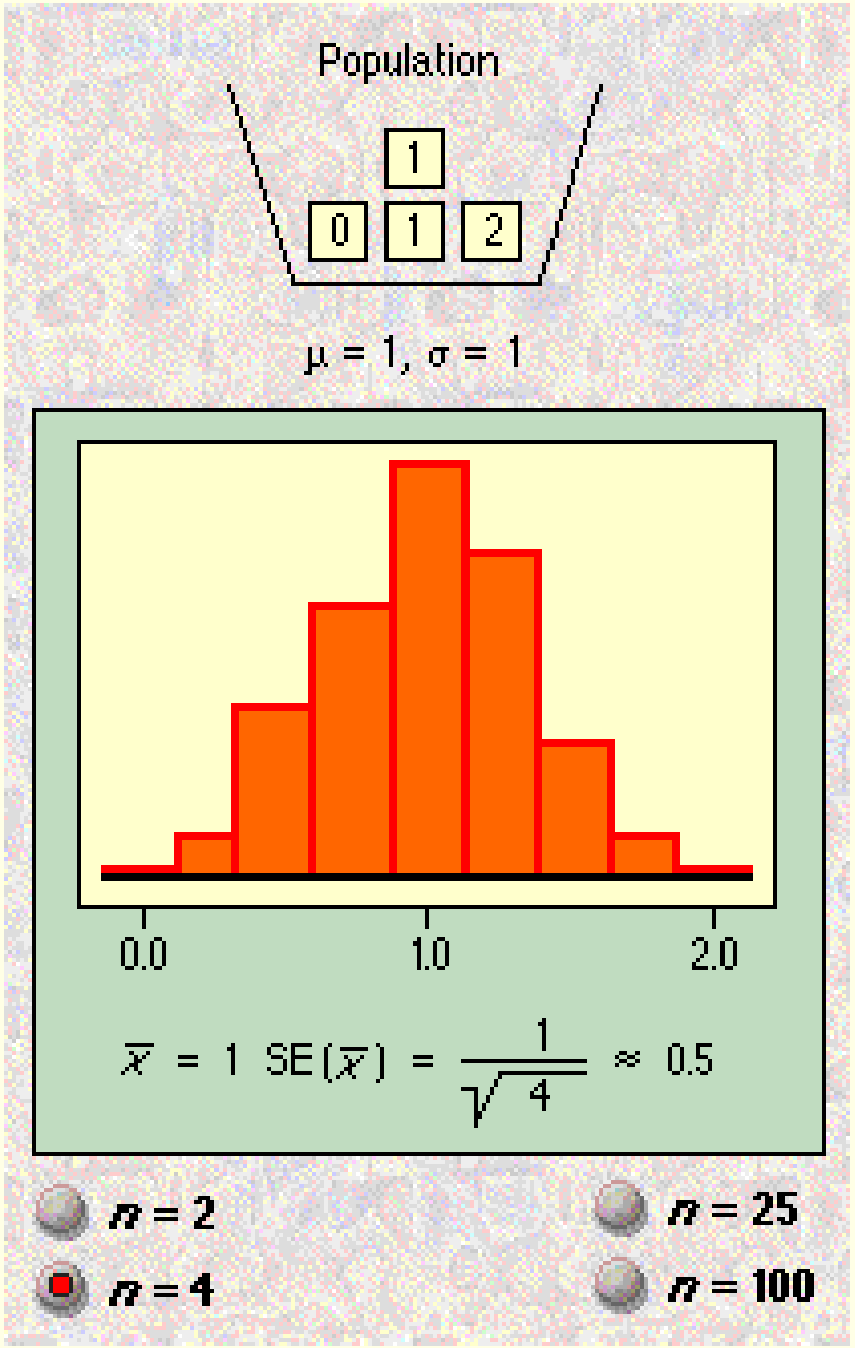
4. **(Központi határeloszlás-tétel)** Ha n elég nagy ($n \geq 30$), akkor $\overline{X}$ eloszlása jól közelíthető az $N(\mu, \sigma^2/n)$ normális eloszlással, még ha a populáció eloszlása esetleg nem is normális.

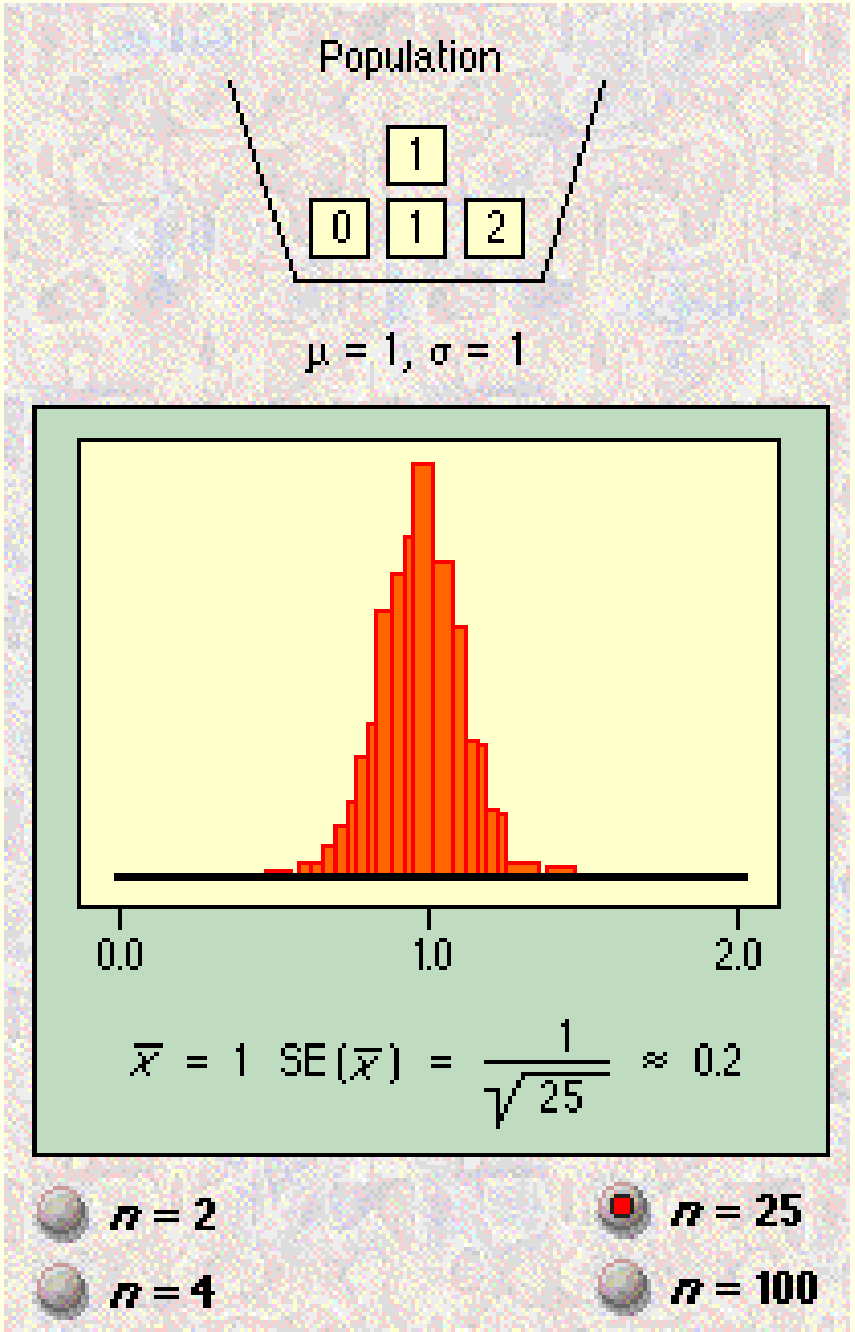
Ezeket az eredményeket a későbbiekben sokszor fel-

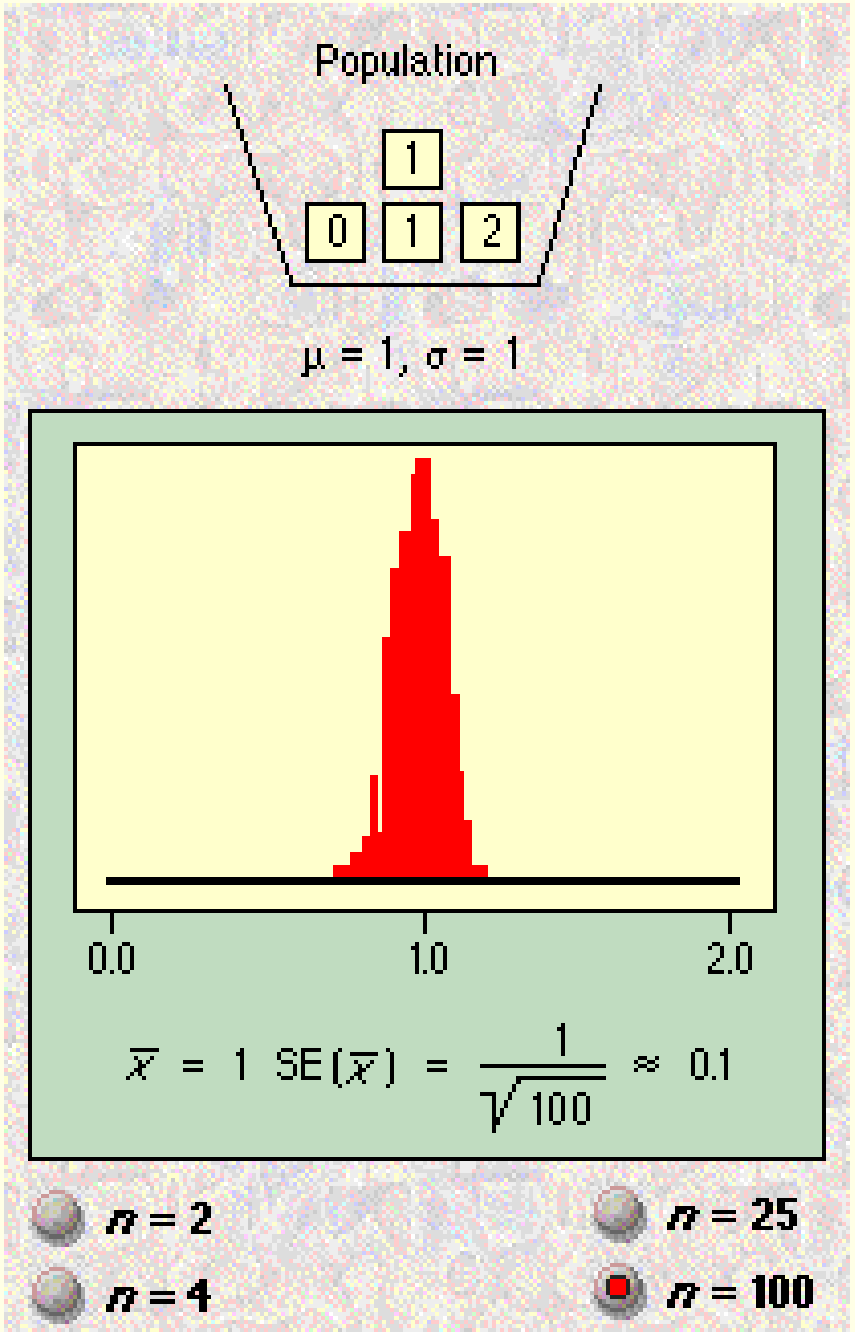
használjuk majd.

A mintaátlag szórásának képletéből láthatjuk, hogy minél nagyobb a minta, azaz minél nagyobb az n, a minta átlagának szórása annál kisebb. Ez érthető is, hiszen ekkor egyre több egyedről van adatunk a populációból. Illusztrációként lássuk az alábbi ábrát.


Population
1
0
1
2
μ = 1, σ = 1
0.0
1.0
2.0
$\bar{x} = 1 \quad SE(\bar{x}) = \frac{1}{\sqrt{2}} \approx 0.71$
n = 2
n = 25
n = 4
n = 100


Population
1
0
1
2
μ = 1, σ = 1
0.0
1.0
2.0
x̄ = 1 SE(x̄) = 1/√4 ≈ 0.5
n = 2
n = 25
n = 4
n = 100


Population
1
0
1
2
μ = 1, σ = 1
0.0
1.0
2.0
$\bar{x} = 1 \quad SE(\bar{x}) = \frac{1}{\sqrt{25}} \approx 0.2$
n = 2
n = 25
n = 4
n = 100


Population
1
0
1
2
μ = 1, σ = 1
0.0
1.0
2.0
$\bar{x} = 1 \quad SE(\bar{x}) = \frac{1}{\sqrt{100}} \approx 0.1$
n = 2
n = 4
n = 25
n = 100

1.3. A szórás becslése

Ha egy vizsgált populáció szórása nem ismert, akkor azt a minta **korrigált empirikus szórásával** becsüljük:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\mathbf{n-1}}}.$$

A populáció szórása és a minta szórása nagy minták esetén lényegében egyenlő.

1.4. Konfidenciaintervallumok

Az előzőekben tárgyalt becslések során mindig fellép a paraméter elméleti értékétől való véletlen eltérés. Sokszor meg tudunk azonban adni egy olyan **intervallumot**, amely az ismeretlen paramétert előre adott nagy valószínűséggel tartalmazza.

Mindig egy a mintából számolt **pontbecsléssel** kezdjük, majd ebből egy intervallumot csinálunk:

BECSLÉS $\pm$ tolerancia.

Itt a tolerancia (hibahatár) a becslés standard hibája szorozva a kritikus értékkel.

A kritikus érték egy eloszlástáblázatból jön, és úgy választjuk meg, hogy felhasználásával egy adott megbízhatóságot garantálhassunk.

A megbízhatósági szint azon minták relatív gyakorisága, amelyekre az intervallum tartalmazza a paraméter valódi értékét.

A tolerancia (hibahatár) a konfidenciaintervallum hosszának fele.

- **Konfidenciaintervallum a populációátlagra ismert szórás esetén**

Példa.

Térjünk vissza a sörös példánkra. Feltételeztük, hogy a sör mennyisége normális eloszlású, $\mu = 0.5$ l és $\sigma = 0.01$ l paraméterekkel. Ha kiveszünk egy 10 elemű véletlen mintát a sörös palackokból, akkor az eddigiek alapján tudjuk azt, hogy a 10 palackban lévő sör mennyiségek átlaga normális eloszlású 0.5 l várható értékkel és $\sigma/\sqrt{10} = 0.01/\sqrt{10} = 0.0032$ szórással.

Alkalmazva ,,kétszórásos" szabályt, azt mondhatjuk,

hogy annak a valószínűsége, hogy a 10 sörösüveg átlaga a

$$0.5 \pm 2 \cdot 0.0032$$

intervallumba essen, kb. **95%**. (Az intervallum pontosabban $0.5 \pm 1.96 \cdot 0.0032$.) Ezek a számítások azt mutatják, hogy összesen **5%** az esélye annak, hogy a minta átlaga ezen az **intervallumon kívülre** esik.

Ha a mintaátlagot használjuk a populációátlag becslésére, akkor összesen 5% az esély arra, hogy a kettő közötti különbség 0.0064-nél nagyobb legyen.

Ha nagyobb mintát használunk, akkor még biztosab-

bak lehetünk abban, hogy a becslésünk elég közel esik a populációátlaghoz (a szórás csökken). Ha 25 elemű mintát használunk, akkor 5% annak az esélye, hogy 0.004 l-rel eltérjünk a populációátlagtól ($2 \cdot 0.01/\sqrt{25} = 0.004$).

Annak ellenére, hogy nem ismerjük a populáció átlagát, abban 95%-os valószínűséggel bízhatunk, hogy a becslésünk - tíz elemű minta esetén - nem tér el 0.0064 liternél jobban a valóságtól. Így, ha csak a mintaátlag áll rendelkezésünkre (0.505 l), akkor azt mondhatjuk, hogy (a 10 elemű mintát alapul véve) a populációátlag 95% valószínűséggel a

$$]0.4986, 1.5114[$$

intervallumba esik. Ezt az intervallumot hívjuk **95%-os konfidenciaintervallumnak** (confidence interval). Ilyenkor nem egyetlen értékkel becsüljük a populációátlagot (pontbecslés), hanem egy intervallummal (intervallumbecslés).

Természetesen nem kötelező 95%-os valószínűséggel (megbízhatósági vagy konfidencia szinttel) dolgozni. Más megbízhatósági szintekhez ki kell keresnünk a normális eloszlás táblázatából az ún. **kritikus értéket**, és azzal kell megszorozni az átlag szórását.

Általánosan felírva:

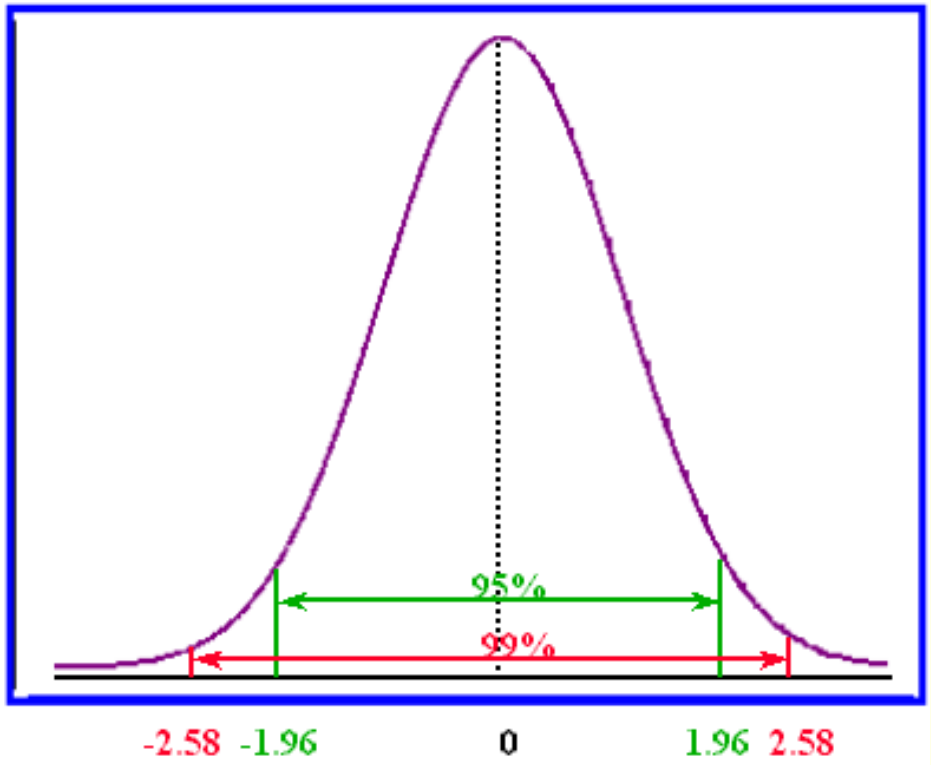
Ha n elemű mintánk van, a populáció szórását ismerjük, és a megbízhatósági szint $p\%$, akkor az ehhez tartozó, a **populációátlagra** vonatkozó konfidenciaintervallum:

$$\left]\bar{x} - z_p \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_p \cdot \frac{\sigma}{\sqrt{n}}\right[.$$

A z_p kritikus érték olyan, hogy $]-z_p, z_p[$ intervallumban a standard normális eloszlás görbéje alatti terület éppen $p\%$. A leggyakoribb két megbízhatósági

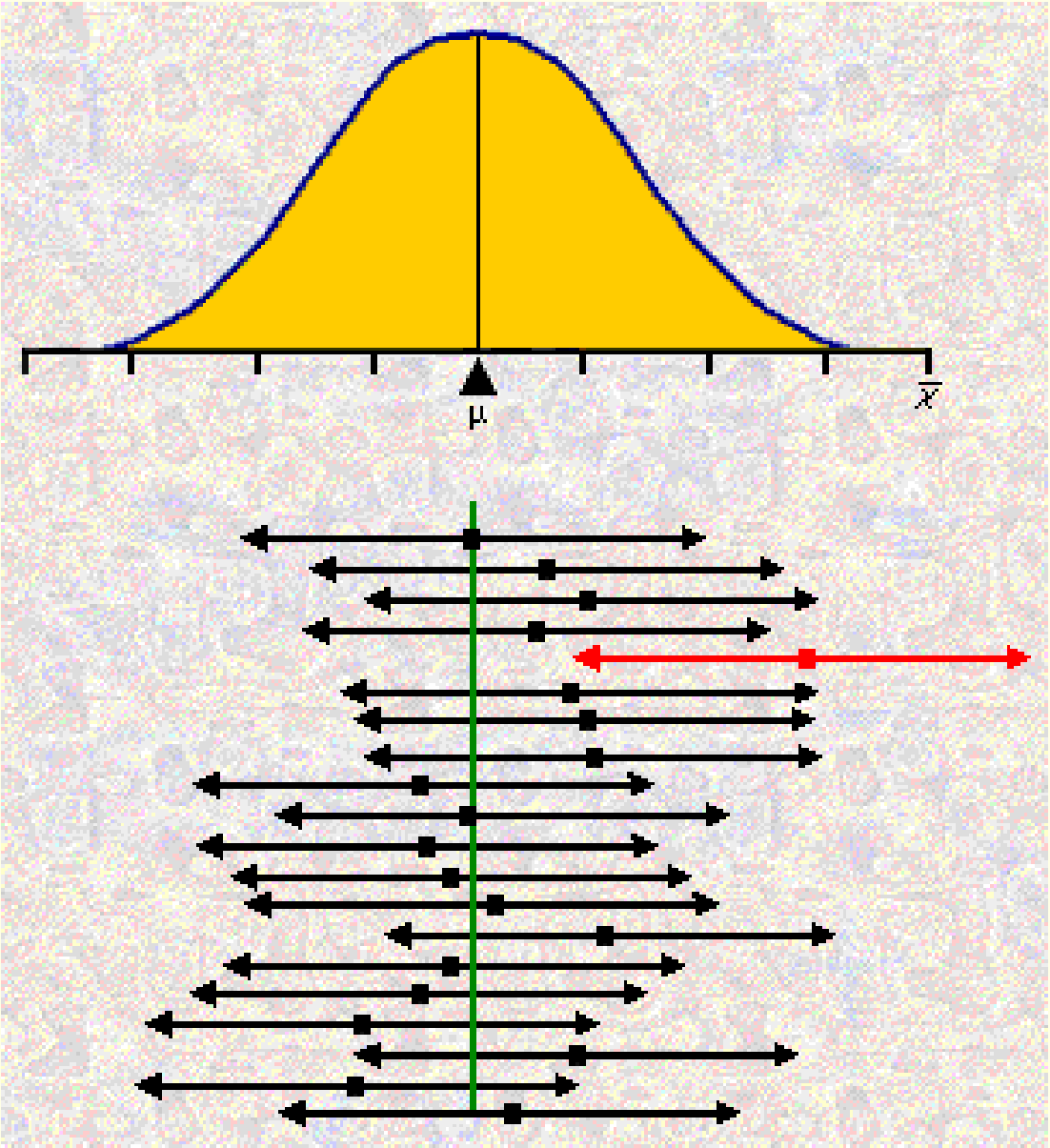
szinthez tartozó kritikus érték:

Megbízhatóság ($p\%$)	Kritikus érték (z_p)
95%	**1.96**
99%	**2.58**



Más valószínűségi eloszlás esetén a konfidenciaintervallumot más eloszlás táblázatból határozzuk meg.

A következő ábra segítségével újra hangsúlyozzuk, hogy 95%-os megbízhatósági szint azt jelenti, hogy a mintavételt 100-szor megismételve, és ezek alapján a konfidenciaintervallumokat kiszámítva, ezek közül várhatóan 95 tartalmazza az ismeretlen populáció átlagot.


μ
$\bar{x}$

- **A minta elemszámának becslése**

A konfidenciaintervallum meghatározásának módja segítséget nyújt ahhoz, hogy a **mintavétel előtt** meg tudjuk mondani, hogy bizonyos pontosság eléréséhez hány elemű mintát kell választanunk. A konfidenciaintervallum képletéből kitűnik, hogy a megengedett hiba (tolerancia):

$$h = z_p \cdot \frac{\sigma}{\sqrt{n}}.$$

Ebből kifejezve a szükséges n elemszámot:

$$n = \frac{z_p^2 \cdot \sigma^2}{h^2},$$

ahol z_p a standard normális eloszlás p%-hoz tartozó kritikus értéke.

Ez a képlet természetesen csak abban az esetben használható, ha ismerjük a populáció σ szórását.

- **Konfidenciaintervallum a populációátlagra ismeretlen szórás esetén**

Amennyiben a populáció szórása nem ismert, annak becslését tudjuk csak használni az átlagra vonatkozó

konfidenciaintervallum meghatározásánál. Ez azt is jelenti, hogy az átlagérték becslésének megbízhatósága csökken. Ezért az intervallumot egy kicsit szélesebbre kell választanunk.

Tudjuk, hogy ha standardizáljuk a mintaátlagot, azaz kivonjuk belőle a várható értéket, és osztjuk a standard hibával, akkor egy standard normális eloszlású valószínűségi változót kapunk:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}.$$

Ha a becsült s szórásértékkel osztunk a standardizálásnál, akkor egy ún. **t-eloszlású** (Student)

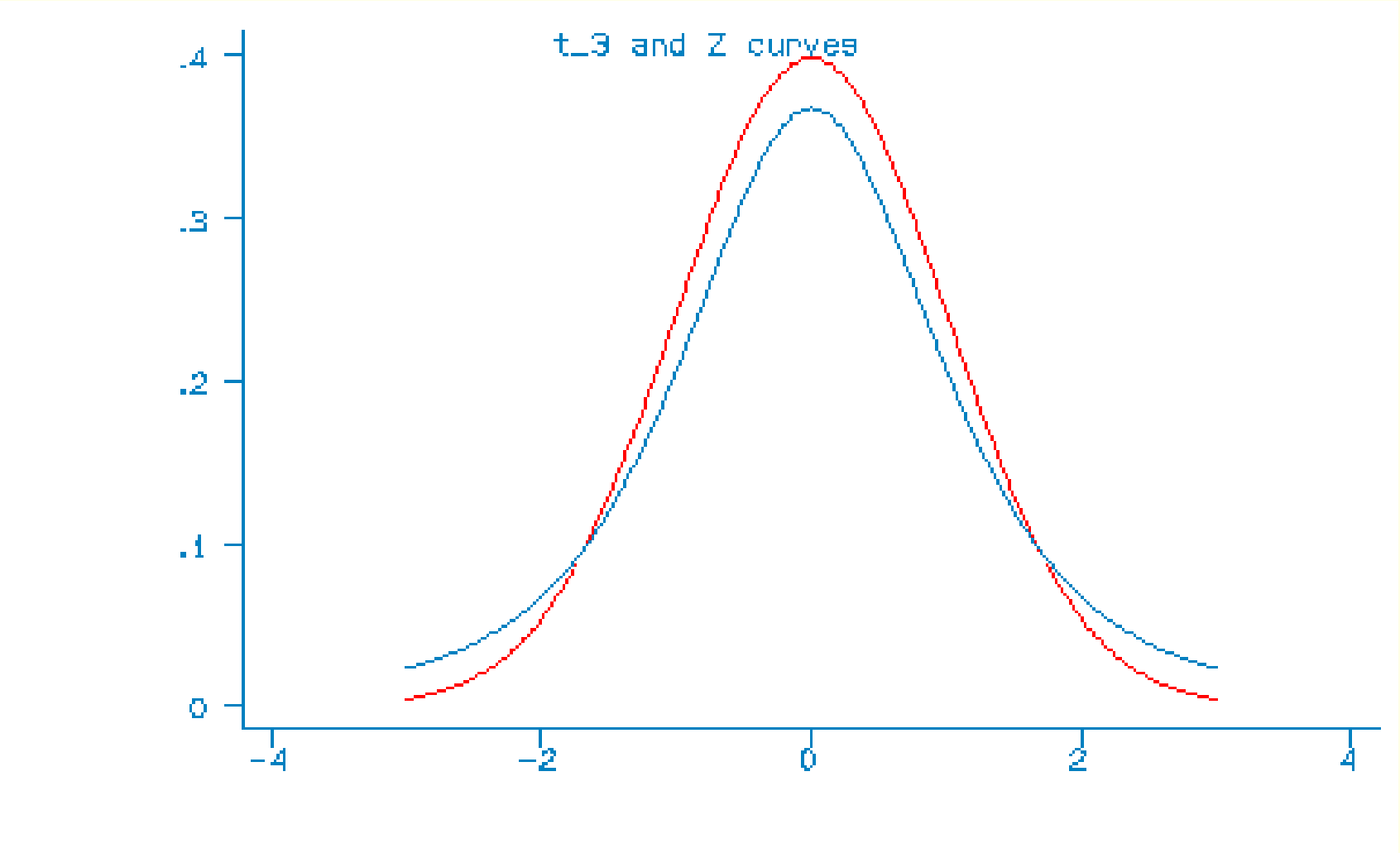
valószínűségi változót kapunk:

$$t = \frac{\bar{x} - \mu}{\mathbf{s}/\sqrt{n}}.$$

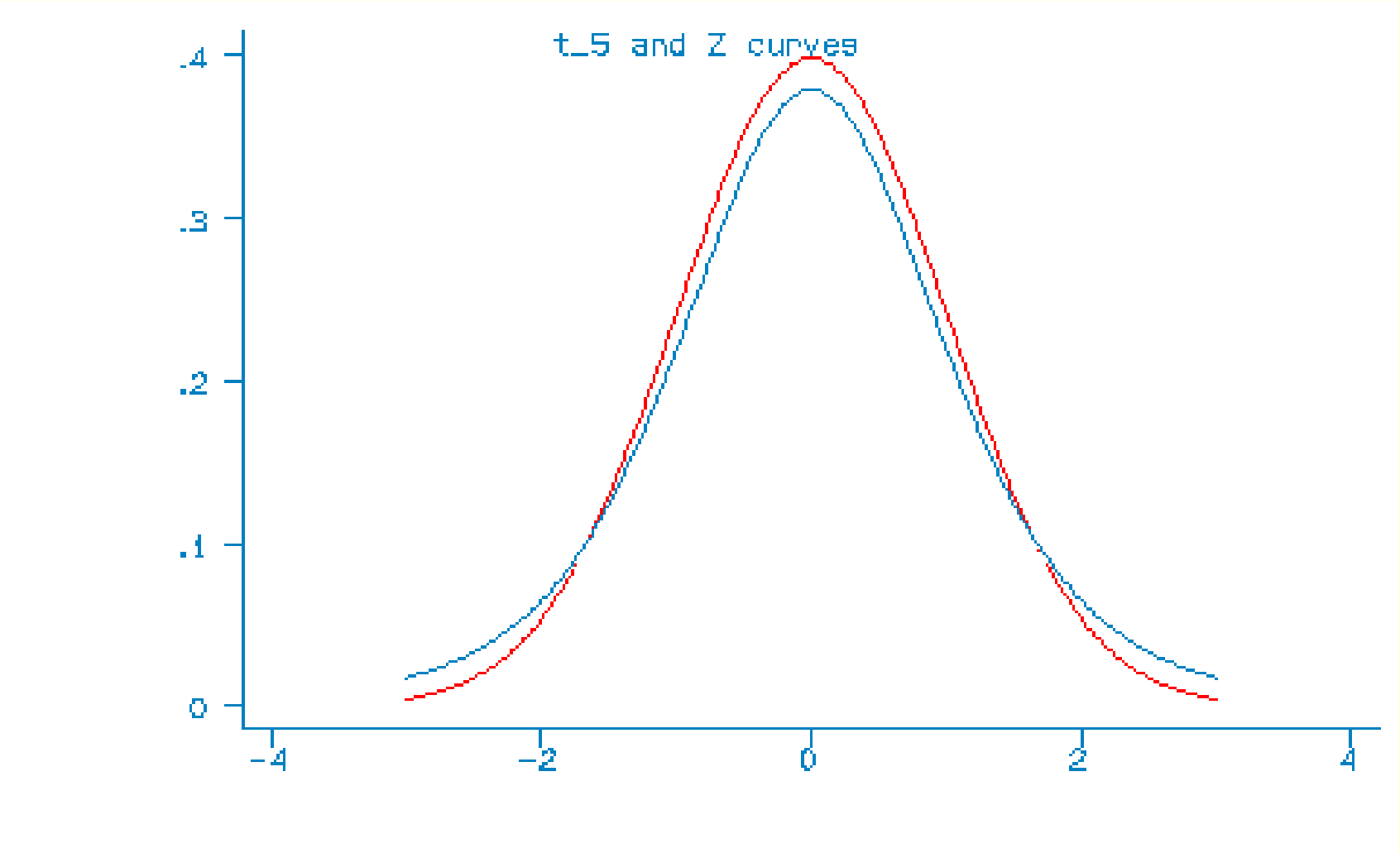
A t-eloszlás szimmetrikus és függ a minta elemszámától. Minden n értékhez egy-egy görbe tartozik, tehát egy görbecsaládról van szó. Ez természetes is, hiszen minél nagyobb a minta elemszáma, annál pontosabban tudjuk becsülni a populációátlagot. A t-eloszlás közvetlen módon a **szabadsági foktól** függ, amely **a minta elemszáma mínusz a becsült paraméterek száma**. A konfidencia-intervallum meg-

adása esetén az átlag szórását kell becsülni, amihez egy már becsült paramétert (az átlagot) használtunk fel, így a szabadsági fok:
d.f.$=n-1$.

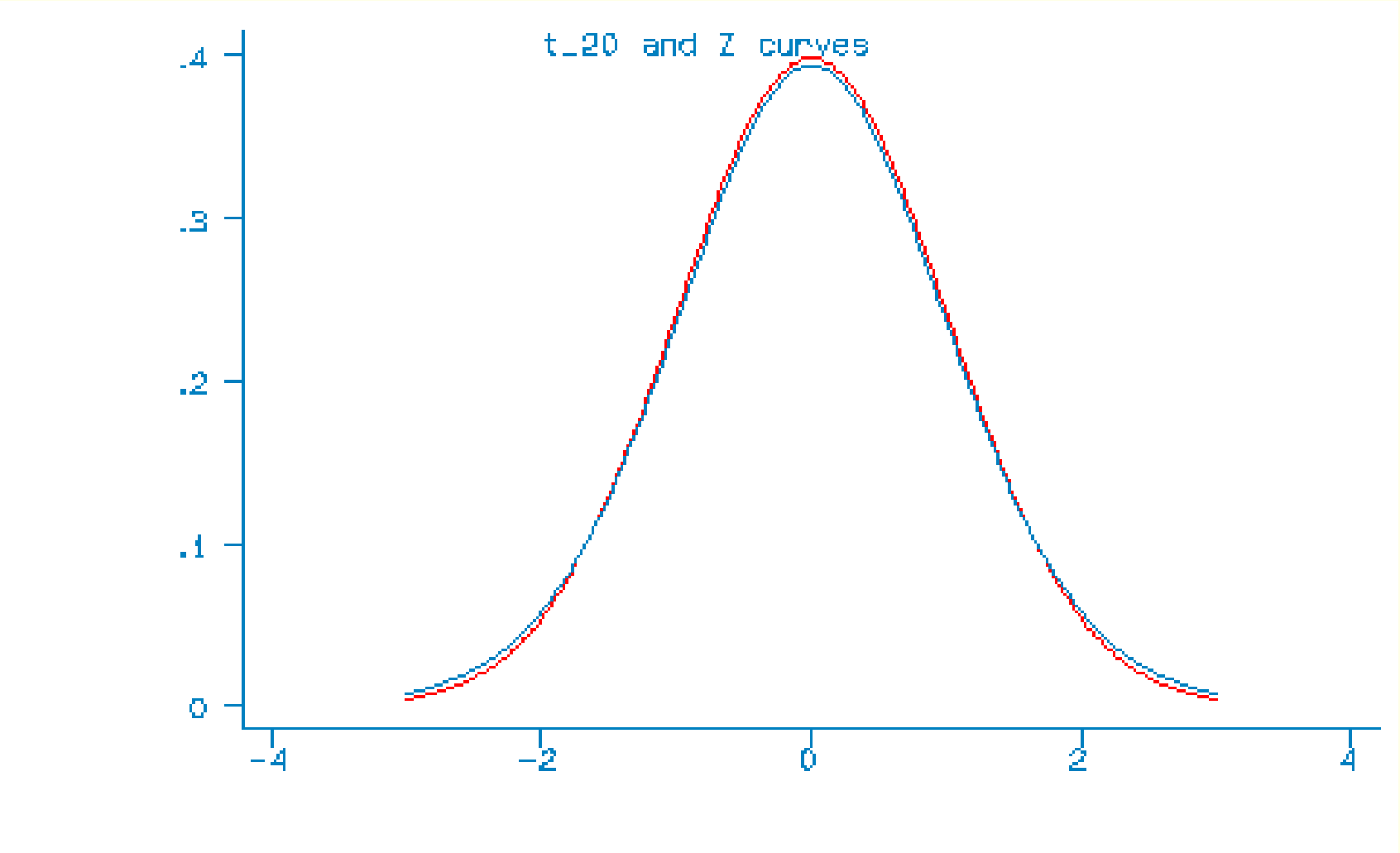
t-eloszlás (d.f. $= 3$) és a standard normális



t-eloszlás (d.f. $= 5$) és a standard normális



t-eloszlás (d.f. = 20) és a standard normális



Az utolsó ábrán is látszik, hogy nagy elemszámú ($n > 20$) minta esetén a t-eloszlás már jól közelíti a standard normális eloszlást, így nagy minták esetén az előzőekben már megismert konfidenciaintervallumokat használhatjuk, még ha a szórás ismeretlen is.

Kis minták ($n \leq 20$) esetén a konstrukció elve pontosan ugyanaz, mint korábban. A különbség mindössze az, hogy a megfelelő t-eloszlást kell használnunk a normális helyett a kritikus érték meghatározására.

Ha n elemű mintánk van, a populáció szórását **nem** ismerjük, és a megbízhatósági szint $p\%$, akkor az ehhez tartozó, a **populációátlagra** vonatkozó konfidenciaintervallum:

$$\left]\bar{x} - t_p \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_p \cdot \frac{s}{\sqrt{n}}\right[,$$

ahol a t_p kritikus érték olyan, hogy a $]-t_p, t_p[$ intervallumban az $n-1$ szabadsági fokú t-eloszlás görbéje alatti terület éppen $p\%$.

Például, a sörös példában használt mintanagyság

$n = 10$ volt, a megbízhatósági szint pedig 95%. A mefelelő kritikus érték: 2.26 (a 9 szabadsági fokú t-eloszlás táblázatából).

Előre adott pontosságú közelítéshez szükséges mintaelemek számát a korábbihoz hasonlóan, a

$$h = t_p \cdot s/\sqrt{n}$$

képletből tudjuk kiszámítani. Így most

$$n = \frac{s^2 \cdot t_p^2}{h^2},$$

ahol t_p értékét a szórás szabadsági foka alapján ke-

ressük ki a táblázatból.

2. Hipotézisvizsgálatok

Az adataink segítségével egy **hipotézis** (elképzelés) helyességéről szeretnénk meggyőződni.

Különböző minták esetén erre vonatkozóan különböző eredményeket kaphatunk. Például, egy választás eredményének előzetes becslésekor egyszer kaphatjuk azt az eredményt, hogy az "A" jelöltre 53% voksol, másszor pedig azt, hogy csupán 49%.

A hipotézisvizsgálatok során azt számítjuk ki, hogy

ha feltételezésünk igaz, akkor mekkora a valószínűsége annak, hogy olyan mintánk legyen, amilyen van.

Ha nagy a valószínűsége, akkor elfogadjuk a hipotézisünket, ha kicsi akkor pedig elutasítjuk.

2.1. Bevezetés

Példa.

Egy gazdaságban a malacok súlyát egy újfajta táp segítségével szeretnék növelni. Az eddig használt táppal az elmúlt években a malacok adott idős ko-

rukra 21.5 kilogrammosak voltak, súlyuk szórása 2.5 kg volt. Ez nagyjából minden eddigi évre igaz volt, így feltételezhetjük, hogy

a malacok súlya normális eloszlású $\mu = 21.5$ kg átlaggal és $\sigma = 2.5$ kg szórással.

Az új táp tesztelésére kiválasztunk egy véletlen mintát (pl. 10 malacot), és ezekkel az új tápot etetjük, majd megnézzük, hogy mekkora lesz az adott korban a súlyuk.

Ha a malacok átlagsúlya lényegesen nagyobb lesz, mint 21.5 kg, akkor a gazdaság elégedett lehet az

eredménnyel, ha továbbra is 21.5 kg körül lesz a malacok átlagsúlya, akkor nyilván elutasítják az új táp használatát.

Kérdés az, hogy hol húzzuk meg azt a határt, amely fölött hatásosnak tekintjük a tápot.

Előfordulhat, hogy olyan malacokat választottunk ki, amelyek mindenképpen elérték volna a kívánt súlytöbbletet. Ebben az esetben az új tápról persze alaptalanul állítanánk, hogy hatásos.

Az is előfordulhat, hogy olyan malacokat vontunk be a kísérletbe, amelyek átlagsúlya a régi táppal csak

18.5 kg lett volna, az új táppal viszont 21.5 kilogrammosak lettek. Ebben az esetben hatástalannak nyilvánítanánk a tápot, pedig az nagyon is megfelelt a kívánt célnak.

Felmerül a kérdés, hogy a mintavételezési hibát figyelembe véve a 10 malac súlyának legalább mekkorának kell lennie ahhoz, hogy az új tápot hatásosnak lehessen nyilvánítani.

22 kg nyilván nem meggyőző, de mi van abban az esetben, ha az átlag 22.5 vagy éppen 23 kg.?

Először nézzük meg, hogy mekkora a valószínűsége

annak, hogy egy valójában **nem jobb tápot jobbnak nyilvánítsunk.** Ha a táp hatástalan, akkor a 10 elemű mintánkat egy olyan populációból választottuk, amelyik normális eloszlású $\mu = 21.5$ kg várható értékkel és $\sigma = 2.5$ kg szórással:

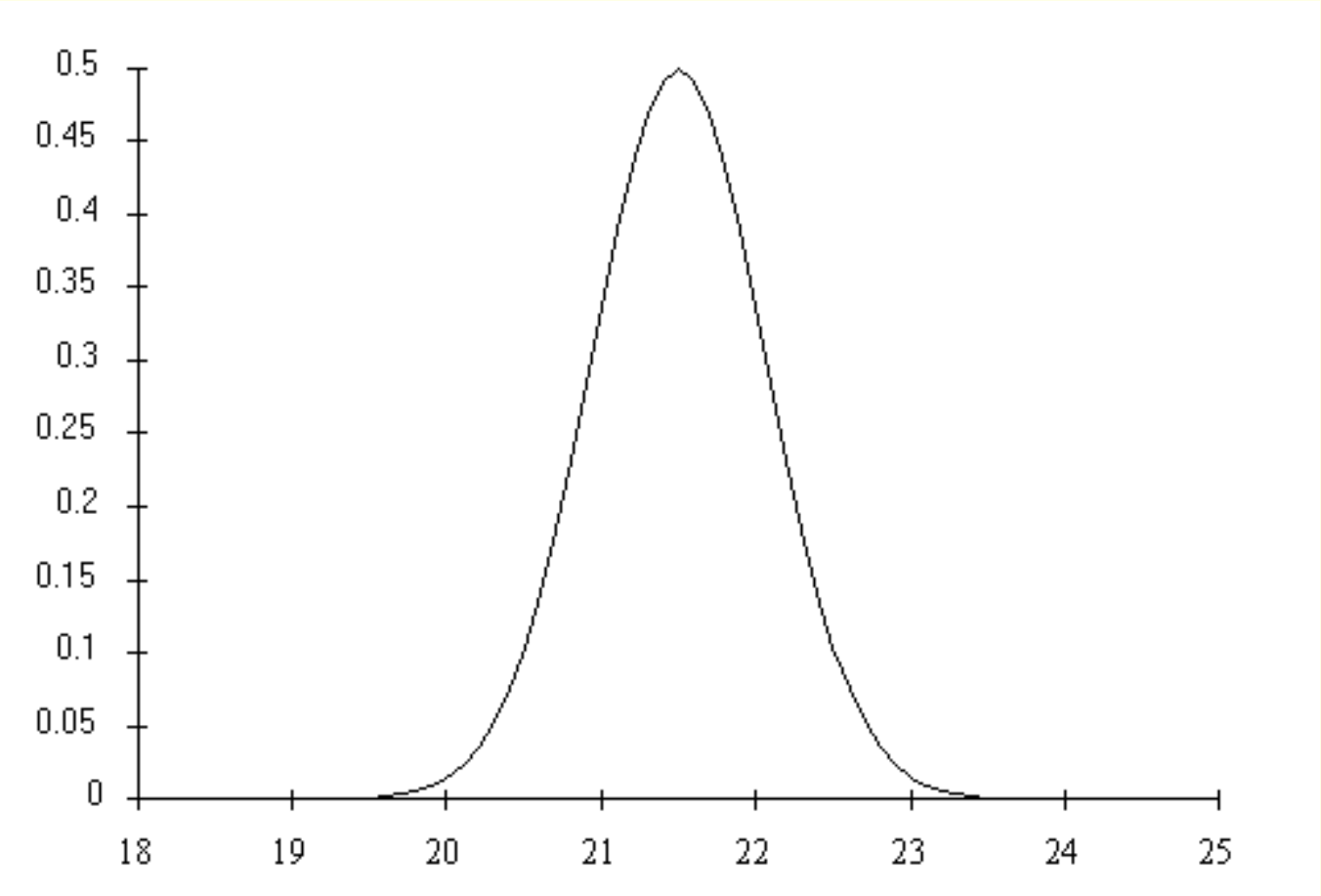
$$X \sim N(21.5, 2.5^2).$$

Az előző előadás alapján tudjuk, hogy az $\overline{X}$ mintaátlag szintén normális eloszlású

$\mu_{\overline{X}} = 21.5$ és $\sigma_{\overline{X}} = 2.5/\sqrt{10} = 0.8$ paraméterekkel:

$$\overline{X} \sim N(21.5, 0.8^2).$$

A mintaátlag eloszlását a következő ábrán láthatjuk.



Ha az átlag csak 22 kg, akkor az nem meggyőző, hiszen annak a valószínűsége, hogy 10 olyan mala-

cot választottunk, amelyek átlagsúlya 22 kg felett van 26.6%. Ezt a valószínűséget a normális eloszlás táblázatának segítségével számoltuk ki a változók standardizálása után $\left(Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)$:

$$\begin{aligned} P(\overline{X} > 22) &= P\left(Z > \frac{22 - 21.5}{2.5/\sqrt{10}}\right) \\ &= P(Z > 0.625) = 26.6\%. \end{aligned}$$

A következő gondolatmenetet alkalmazzuk. Szükségünk van egy olyan kritikus értékre, amelynél ha

nagyobb a malacok átlagsúlya, akkor már igen kicsi a valószínűsége annak, hogy egy ilyen eloszlású populációból pontosan így válasszuk a malacokat. A kritikus értéket úgy szokták meghatározni, hogy 5% legyen a valószínűsége annak, hogy akkora, vagy annál nagyobb legyen az átlag, mint az általunk meghatározott kritikus érték.

Számítsuk ki, hogy mekkora az 5%-os valószínűséghez tartozó kritikus érték:

$$
\begin{aligned}
0.05 &= P(Z > 1.65) = P\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} > 1.65\right) \\
&= P(\overline{X} > \mu + 1.65 \cdot \sigma/\sqrt{n}) \\
&= P(\overline{X} > 21.5 + 1.65 \cdot 0.8) \\
&= P(\overline{X} > 22.85).
\end{aligned}
$$

Tehát 5% a valószínűsége annak, hogy Z (az $\overline{X}$ standardizáltja) meghaladja az 1.65-ot, azaz $\overline{X}$ meghaladja a 22.85 kg-ot.

Az okoskodásunk logikája röviden:

1. Feltételezzük, hogy a populációnk olyan malacokból áll, amelyek súlya $\mu = 21.5$ kg átlagú és $\sigma = 2.5$ kg szórású normális eloszlású valószínűségi változó.

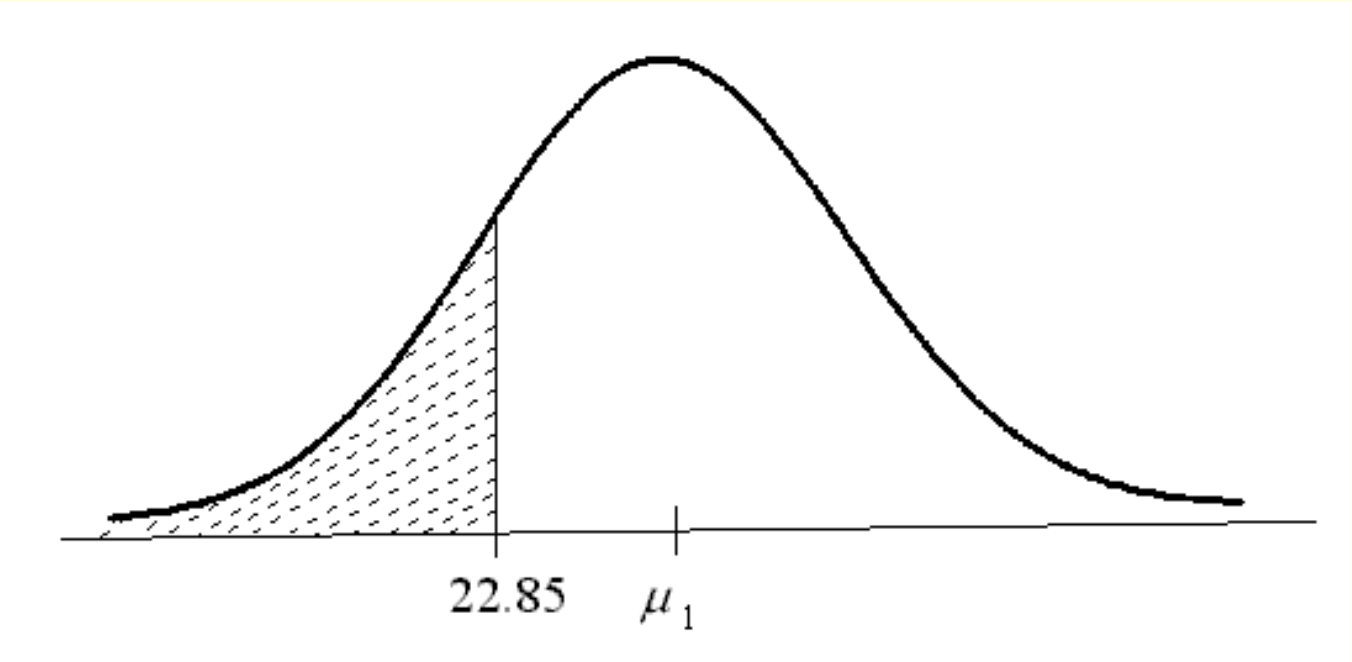
2. Kiválasztunk közülük tízet, és megnézzük, hogy mekkora az átlagsúlyuk.

3. Ha ez az átlagsúly meghaladja a 22.85 kg-ot, akkor meg lehetünk győződve arról, hogy hatásos volt a táp. (Ha hatástalan lenne, akkor csupán 5% lenne az esélye annak, hogy a kiválasztott mintánk átlaga 22.85 kg felett legyen.)

Most nézzük meg az ellenkező esetet, azaz hogy mekkora a valószínűsége annak, hogy egy **hatásos tápot hatástalannak minősítünk.**

A **mintavételi hiba** most is fennáll, hiszen lehet, hogy pont tíz olyan malacot választottunk, amelyek a normál tápon átlagosan csak 19.5 kg-osak lettek volna, de az új táppal elérték a 21.5 kg-os súlyt. Ezt az előbbi gondolatmenet alapján még nem fogadjuk el. Az is lehet, hogy éppen 22.85 kg lesz az átlag.

A következő ábra mutatja, hogy mekkora a valószínűsége annak, hogy ekkora hibát kapjunk. Itt μ_1 jelöli az új táppal etetett malacok átlagsúlyát.



Ha a táp éppen 2 kg-mal növeli a súlyt, akkor a mintánkat egy olyan populációból vesszük, amely normális eloszlású, átlaga 23.5 kg és szórása 2.5 kg.

Annak a valószínűsége, hogy a mintaátlag ebben az esetben kevesebb legyen, mint 22.85 kg az alábbi:

$$P(\overline{X} < 22.85) = P\left(Z < \frac{22.85 - 23.5}{2.5/\sqrt{10}}\right)$$
$$= P(Z < -0.82) = 20.6\%.$$

Ha a táp még jobban növelné a malacsúlyt, akkor ez a valószínűség még kisebbnek adódna.

A mintavételi hiba létezéséből következik, hogy két fajta téves statisztikai következtetést vonhatunk le:

- egyrészt azt mondhatjuk egy hatástalan tápról, hogy hatásos,

- másrészt egy hatásosról azt állíthatjuk, hogy hatástalan.

Ha 22.85 kg-nál húzzuk meg a határt, akkor 5% a valószínűsége az első esetnek, és 20.6% a másodiknak.

Tehát nem azt próbáltuk bizonyítani, hogy a táp jó. Nem a hipotézisünkről látjuk be, hogy az helyes-e vagy sem, hanem csak azt mondjuk, hogy **az adatok cáfolják hipotézist**, vagyis hogy ha az elméletünk igaz, akkor ezek az adatok nagyon valószínűtlenek.

Az a hipotézis, amelyet az adatok „elfogadnak”, konzisztens az adatokkal abban az értelemben, hogy

az elmélet és az adatok közti különbség magyarázható a mintavételi hibával.

2.2. A null- és az alternatív hipotézis

A statisztikai hipotézisvizsgálatok esetén mindig két hipotézis közül kell választanunk. Az egyik az ún. **nullhipotézis** (jele: H_0), a másik pedig az **alternatív hipotézis** (jele: H_1).

Előző példánkban:

nullhipotézis: a táp nem jobb, mint az előző.

alternatív hipotézis: a táp jobb, mint az előző

Formálisan:

$$H_0: \qquad \mu = 21.5,$$
$$H_1: \qquad \mu > 21.5.$$

Általában:

$$H_0: \qquad \mu = \mu_0,$$
$$H_1: \qquad \mu > \mu_0.$$

Az ilyen típusú H_1 esetén **egyoldali alternatív hipotézisről** beszélünk.

Az adatokat arra használjuk, hogy eldöntsük:

elfogadjuk-e H_0-t vagy sem.

A döntési szabályunk az volt, hogy H_0-t

- **elfogadjuk**, ha a mintaátlag Z standardizáltja kisebb, mint az 5%-hoz tartozó kritikus érték (azaz 1.65),
- **elutasítjuk**, ha Z meghaladja ezt az értéket.

A hipotézis elfogadásáról vagy elvetéséről egy ismert eloszlású ún. **próbástatisztika** segítségével döntünk (lásd később).

2.3. Az első- és másodfajú hiba

Mivel a minta a véletlentől függ, ezért soha nem lehetünk biztosak abban, hogy a hipotézis igaz vagy sem. A döntés során kétféle hibát követhetünk el.

Elsőfajú hibát követünk el, ha elvetjük a nullhipotézist, holott az igaz.

Ennek valószínűségét jelölje α.

Másodfajú hibát követünk el, ha elfogadjuk a nullhipotézist, holott az nem igaz.

Ennek valószínűsége pedig legyen β.

Az α és β értéke attól függ, hogy hol húztuk meg azt a **kritikus határt**, amely alatt elfogadhatjuk H_0-t, illetve amely felett elutasítjuk. Ha növeljük a kritikus értéket, akkor az esetek többségében csökkentjük α-t, és egyúttal növeljük β-t. Ha csökkentjük a kritikus értéket, akkor α csökken, de β nő.

Az α-t általában 5%-nak szokás megadni. A β pedig függvénye α-nak, H_1-nek, és az n-nek, amint azt már a példánkban láthattuk.

Ha az adataink alapján elutasítjuk H_0-t, akkor az

eredményünket statisztikusan **szignifikánsnak**, vagy egyszerűen szignifikánsnak mondjuk.

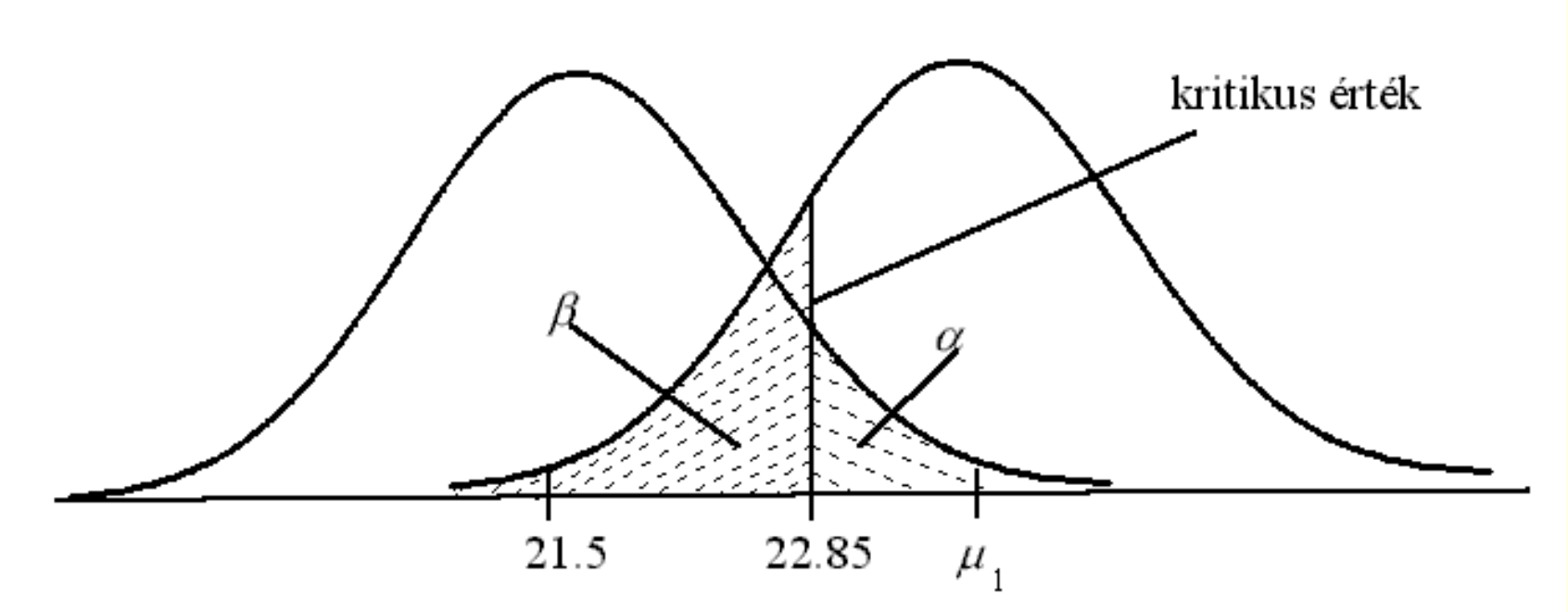
Ha $\alpha < 5\%$, például 1 vagy 0.1%, akkor a kritikus határ kitolódik, és ezzel az elsőfajú hiba valószínűsége lecsökken.

Példánk esetén az első- és másodfajú hibát az alábbi ábrán szemléltetjük.

Feltételeztük, hogy a valószínűségi változónk normális eloszlású, és a szórása 2.5 kg. Ekkor a H_0 és H_1 hipotézisünknek az ábrán látható két sűrűségfüggvény (a próbastatisztika sűrűségfüggvénye) felel meg.

Annak a valószínűsége, hogy H_0 igaz, de mi elutasítjuk: α, annak a valószínűsége, hogy H_0 -t elfogadjuk, pedig hamis: β.

Jól látható, hogy ha α-t csökkentjük, akkor β nő, és fordítva. Az α-t mi határozzuk meg, a β pedig az α-tól függ. Természetesen β értékét nem tudjuk pontosan megmondani, hiszen ez függ attól, hogy mennyi ténylegesen az új táppal etetett malacok populációjának átlagsúlya.



2.4. Kétoldali alternatív hipotézis

Eddig csak olyan eseteket vizsgáltunk, amikor az alternatív hipotézis a nullhipotézis ,,egyik oldalán" van. Nagyon gyakran előfordul azonban az, hogy

csak arra vagyunk kíváncsiak, hogy a kísérletnek **volt-e egyáltalán hatása**, mindegy hogy az pozitív vagy negatív irányú. Formálisan:

$$\begin{aligned} H_0 : &\qquad \mu = \mu_0, \\ H_1 : &\qquad \mu \neq \mu_0, \end{aligned}$$

ami a példánkban a következő lenne:

$$\begin{aligned} H_0 : &\qquad \mu = 21.5, \\ H_1 : &\qquad \mu \neq 21.5. \end{aligned}$$

A nullhipotézis az, hogy a tápnak nincs hatása, az

alternatív pedig az, hogy igenis okoz változást a malacsúlyban az új fajta táp.

Hogyan dönthetünk ekkor a táp hatásosságáról?

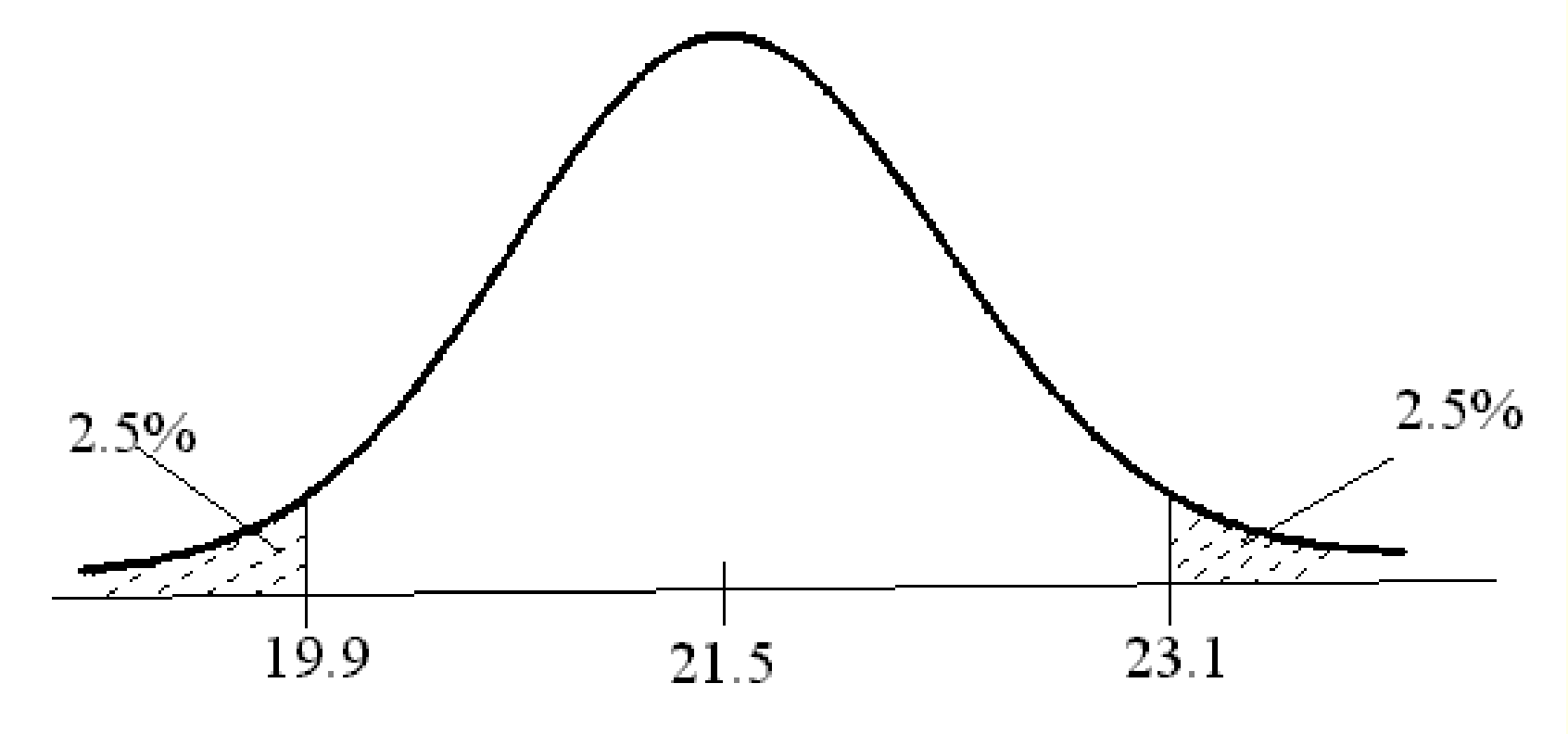
Ha a malacok súlya **sokkal nagyobb**, vagy **sokkal kisebb**, mint 21.5 kg, akkor meg lehetünk győződve arról, hogy valamilyen változást okozott az új táp bevezetése.

A kétoldali alternatív hipotézisek kétoldali vizsgálatokra vezetnek. A kritikus értéket megint az elsőfajú hiba határozza meg. A malacok véletlenszerű kiválasztása azt okozhatja, hogy a mintába 10 átlag

feletti, vagy átlag alatti malac kerül. Az első esetben az **átlag feletti** malacok súlya lehet akkora, hogy azt hisszük, hogy **pozitív hatása** volt a tápnak (pedig a régi táp esetén is körülbelül ugyanennyi lett volna a súlyuk). Az utóbbi esetben pedig az **átlag alatti** malacoknak lehetett olyan alacsony az átlagsúlya, hogy azt hisszük, hogy a táp **rossz hatással** volt a malacokra.

A kritikus súlyokat úgy szeretnénk megválasztani, hogy az elegendően távol legyen a 21.5 kg-tól, és így az elsőfajú hiba kicsi legyen. Az általános gyakorlat az, hogy 2.5% esélyt adunk annak, hogy helytelenül

döntsünk a pozitív effektusról, és 2.5%-ot annak, hogy helytelenül döntsünk a negatív effektusról. Ez összesen 5%.



Megjegyzések:

1. Mi történik akkor, ha az adatok nem normális eloszlásúak? Már láttuk korábban, hogy nagy minta esetén az átlagérték ekkor is normális eloszlású lesz. Ilyen esetben tehát használhatjuk az előbbiekben leírtakat.

2. Mi van akkor, ha nem ismerjük a szórást? Nyilván becslést kell adnunk rá. Ha nagy mintából becsüljük, akkor feltételezhetjük, hogy a becslés elegendően pontos, és alkalmazhatjuk az eddig leírtakat. Ha a populáció eloszlása normális,

akkor kis minta esetén a t-eloszlás használatával korrigálhatjuk a módszert. Ennek az lesz a hatása, hogy a kritikus értékek távolabb fognak esni a H_0-ban feltételezett μ_0 átlagértéktől.

Például, ha a malacsúlyok esetén nem ismerjük a szórást, csak becsültük a 10 elemű mintából, és az 2.5-nek adódott, akkor egyoldali próba esetén a kritikus érték:

$$
\begin{aligned}
0.05 &= P\left(\frac{\overline{X}-\mu_0}{s/\sqrt{n}} > t_{\text{krit.}}\right) \\
&= P(\overline{X} > \mu_0 + 1.83 \cdot s/\sqrt{n}) \\
&= P(\overline{X} > 21.5 + 1.83 \cdot 0.8) \\
&= P(\overline{X} > 22.96).
\end{aligned}
$$

A $t_{\text{krit.}}$ értéket a t-táblázat alapján határozzuk meg. A szabadsági fok $n-1 = 9$, a valószínűség 0.05. Így $t_{\text{krit.}} = 1.83$.

Kétoldali hipotézis esetén:

$$
\begin{aligned}
0.025 &= P\left(\frac{\overline{X}-\mu_0}{s/\sqrt{n}} > t_{\text{krit.}}\right) \\
&= P(\overline{X} > 21.5 + 2.26 \cdot 0.8) \\
&= P(\overline{X} > 23.3).
\end{aligned}
$$

Ebben az esetben a szabadsági fok szintén $n - 1 = 9$, és a valószínűség 0.025, $t_{\text{krit.}} = 2.26$.

3. A hipotézisvizsgálat elnevezés helyett nagyon gyakran a **statisztikai próba** kifejezést szokták használni. Eddig mindig csak egy mintát vizsgál-

tunk, ezért ilyenkor az **egymintás próba** kifejezés használatos.

4. Mivel ismert szórás esetén a normális eloszlás táblázatát használjuk a kritikus érték meghatározására, ezért ilyenkor **egymintás z-próbáról** (vagy u-próbáról) beszélünk. Becsült szórás esetén az **egymintás t-próba** használata a megfelelő.

5. Azokat az intervallumokat, amelyekbe ha beleesik a minta átlaga, akkor elutasítjuk a nullhipotézist, **kritikus tartományoknak** nevezzük.

A hipotáézisvizsgálat lépéseinek összefoglalása

1. lépés	A null- és az alternatív hipotézis megfogalmazása
2. lépés	A szignifikancia-szint (α) megválasztása
3. lépés	A próbastatisztika kiválasztása
4. lépés	A döntési szabály megfogalmazása
5. lépés	Döntés a minta alapján