# Elemi statisztika fizikusoknak

Pollner Péter

Biológiai Fizika Tanszék

pollner@elte.hu

# 4-6. fejezet
# Valószínűségek kiszámítása szimulációval

# Kulcsfogalmak

**Ebben a fejezetben egy másik módszert mutatunk be a valószínűségek kiszámítására, amivel az előző fejezetekben bevezetett formális módszerek nehézségeit ki lehet kerülni.**

# Definíció

**Egy folyamat szimulációja egy olyan másik folyamat, ami ugyanúgy viselkedik, mint az valódi (jó közelítéssel).**

# Szimulációs példa (nagyon egyszerű példa)

**Nemek (F,N) szelekciója Ha valamilyen nemi szelekciós módszert tesztelünk, akkor tudnunk kell, hogy mi a valószínűsége annak, hogy 100 újszülött közül legalább 60 lány, feltéve, hogy a fiú és lány születések egyforma gyakoriak.**

**Találjunk ki egy egyszerű szimulációt, amivel ki tudjuk számítani ezt a valószínűséget!**

# Példa: Generáljuk 100 újszülött nemét

## 1. megoldás:

- Dobjunk fel 100-szor egy érmét és

fej = nő
írás = férfi

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| F | F | Í | F | Í | Í | F | F | F | F |
| ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| nő | nő | férfi | nő | férfi | férfi | nő | nő | nő | nő |

## 2. megoldás:

- Generáljunk 0 és 1 sorozatokat egy számítógéppel, ahol

0 = férfi
1 = nő

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ | ↓ |
| férfi | férfi | nő | férfi | nő | nő | nő | férfi | férfi | férfi |

- Generáljunk nagyon sokszor (N alkalommal) 100 db véletlen 0 vagy 1 számot 50-50% valószínűséggel (vagy pl. 51,12% - 48,88 %).
- Számoljuk meg hányban van 60 vagy több 1-es ($N_{60}$ alkalommal).
- P(60 vagy több lány 100 születésből)=$N_{60}$/N

# Véletlen számok

**Sok szimulációban, véletlen számokat használunk a valóságos események szimulációjára. Különböző véletlen szám generálási módszerek:**

- **Véletlen számok táblázata**
- **Excel (VÉL() függvény, 0 és 1 között egyenletesen)**
- **C (y=random(100) ) véletlen 0 és 100 közötti egész egyenletesen**

# Szimuláció ellenőrzése

Valóban véletlenek-e a számok, vagy elértük a véletlenszám-generátor határát? Pl. vannak-e nem-várt periodicitások?

A modell megfelel-e a tapasztalatoknak, előzetes ismereteknek?

Pl: kocka dobás eredményének szimulálása azonosan valószínű számok generálásával?

1 db kocka esete?

2 db kocka esete?

# Összefoglalás

**Ebben a fejezetben megvitattuk a:**

- **Szimulációkat.**
- **Véletlen szám generálást.**

# 5. előadás
# Valószínűség eloszlások

5-1 Áttekintés

5-2 Véletlen változók

5-3 A binomiális eloszlás

5-4 A binomiális eloszlás átlaga, varianciája és szórása

5-5 A Poisson eloszlás

6-1 Áttekintés

6-2 A normális eloszlás

# 5-1. fejezet
# Áttekintés

# Áttekintés

## Alap filozófia:

**A valószínűség eloszlások azt írják le, hogy valószínűleg mi fog történni és nem azt, hogy valójában mi történt.**

**Pl: kockadobás**

# 5-2. fejezet
# Véletlen változók

# Kulcsfogalmak

**Ebben a fejezetben bevezetjük a valószínűségi eloszlás fogalmát, ami megadja egy változó véletlen által meghatározott értékeinek a valószínűségét.**

**Figyelembe veszi, hogy egy adott kimenet gyakran következik-e be, vagy pedig egy szokatlan értékkel van dolgunk, ami ritkán fordul elő véletlenül.**

# Definíciók

- **Véletlen változó**

  **egy változó (tipikusan *x* vagy *k* jelöli) aminek az egyes számértékeit a véletlen kísérlet véletlenszerű kimenetei határoznak meg**

- **Valószínűség eloszlás**

  **egy olyan leírás, ami a véletlen változó minden egyes értékéhez hozzárendeli annak valószínűségét; gyakran grafikonként vagy táblázatként vagy képlettel van kifejezve**

# Definíciók

- **Diszkrét véletlen változó**

  **vagy véges sok, vagy megszámlálhatóan sok számú értéket vehet fel**

**(nominális, ordinális változók is!)**

- **Folytonos véletlen változó**

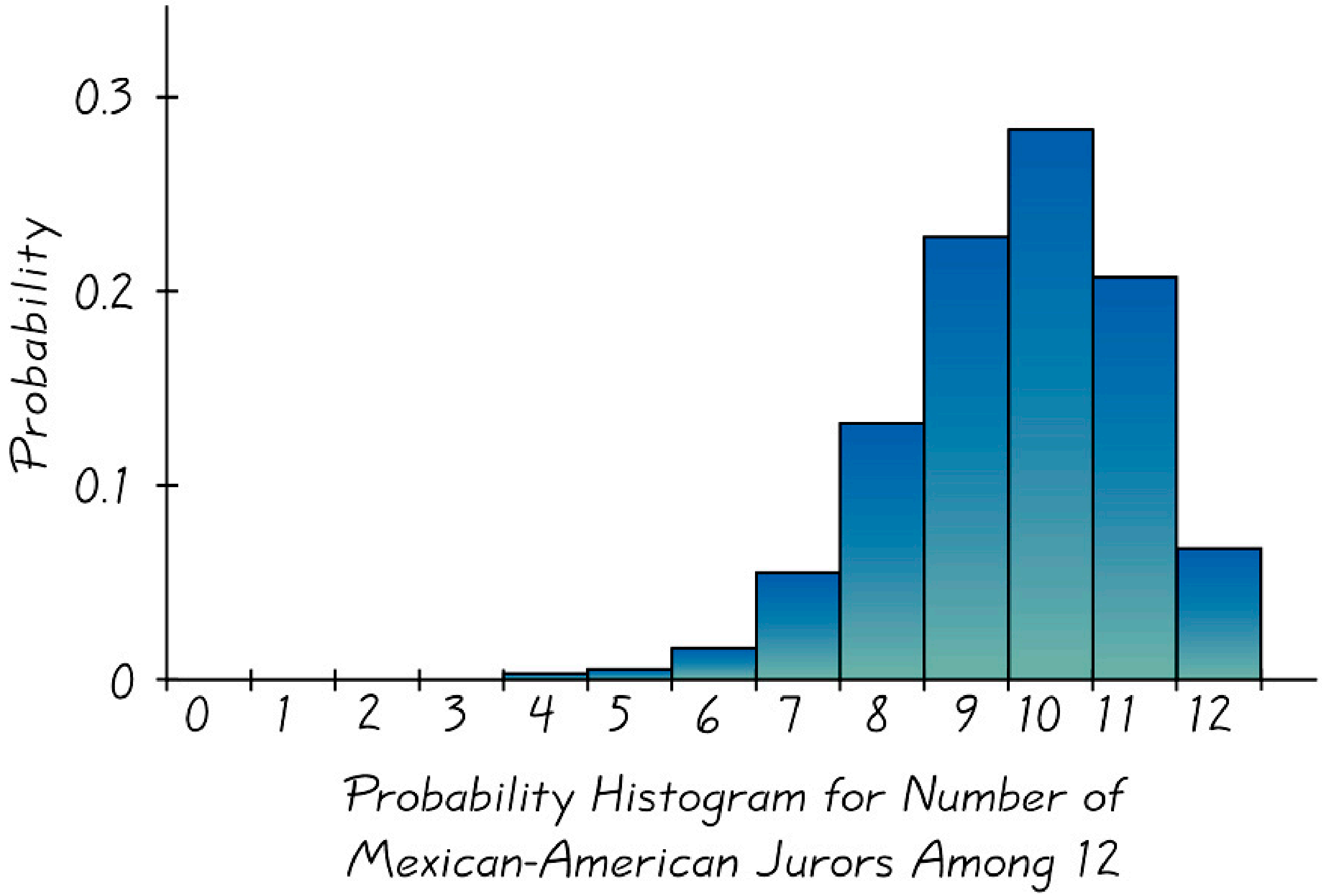
  **végtelen sok értéket vehet fel, melyek valamilyen folytonos skálán megadható mérés eredményeiként adódnak**

# Példa

- Texasban, Hidalgo járásban annak a valószínűsége, hogy az esküdtszék 12 tagja közül hány Mexikói-Amerikai. A lakosság 80%-a Mexikói-Amerikai.
- Szokatlan-e, hogy egy esküdtszék 7 tagja Mexikói-Amerikai vagy nem?

# Grafikonok

**A valószínűség hisztogram nagyon hasonló a relatív gyakoriság hisztogramhoz, de a függőleges skála most a valószínűségeket mutatja.**



# A valószínűség eloszlás fontos tulajdonságai

$$\sum P(x) = 1$$

**ahol *P* pozitív értékeket vehet fel.**

$$0 \leq P(x) \leq 1$$

**minden *x értékre.***

# A valószínűség eloszlások átlaga, varianciája és szórása

$$\mu = \Sigma\, [x \cdot P(x)]$$ Átlag

$$\sigma^2 = \Sigma\, [(x - \mu)^2 \cdot P(x)]$$ Variancia

$$\sigma^2 = [\Sigma\, x^2 \cdot P(x)] - \mu^2$$ Variancia (rövidített)

$$\sigma = \sqrt{\Sigma\, [x^2 \cdot P(x)] - \mu^2}$$ Szórás

# Definíció

**Az $x$ diszkrét véletlen változó várható értékét általában $E(x)$ jelöli, ami a kimenetek átlaga. Értékét úgy kaphatjuk meg, ha kiszámítjuk a $\Sigma\, [x \cdot P(x)]$ kifejezés értékét.**

$$E(x) = \sum [x \cdot P(x)]$$

# Ritkán előforduló értékek azonosítása

**Az értékek nagy része az átlag 2 (3) szórásnyi környezetébe esik. Ezen kívül találhatók a ritka értékek.**

**A “szokatlan” értékek az alábbi határokon kívülre esnek:**

**A szokásos értékek maximuma = $\mu + 2\sigma$**

**A szokásos értékek minimuma = $\mu - 2\sigma$**

# A ritka értékek azonosítása

## Ritka esemény szabály

**Ha bizonyos feltevés mellett (mint pl. hogy egy érme szabályos) egy bizonyos bekövetkező esemény megfigyelése (mint pl. 992 fej 1000 dobásból) nagyon kicsi, akkor arra következtetünk, hogy a feltevés nem igaz.**

- **Szokatlanul sok**: ***x*** **siker** ***n*** **próbálkozásból szokatlanul sok ha *P*(*x* vagy több siker) ≤ 0.05 (0.003).**
- **Szokatlanul kevés**: ***x*** **siker** ***n*** **próbálkozásból szokatlanul kevés ha *P*(*x* vagy kevesebb siker) ≤ 0.05 (0.003).**

# Összefoglalás

Ebben a fejezetben megvitattuk:

- A leíró statisztika és a valószínűségek kombinálását.
- Véletlen változókat és eloszlásukat.
- Valószínűség histogrammokat.
- A valószínűség eloszlások tulajdonságait.
- Átlagot, varianciát és szórást a vsz. eloszlás esetén.
- A különös esetek azonosítását.
- A várható értéket.

# 5-3. fejezet
# Binomiális Eloszlás

# Kulcsfogalmak

**A binomiális eloszlást akkor tudjuk használni, ha a kimeneteket két csoportra lehet osztani, mint elfogadható/nem elfogadható, túlélő/elpusztult stb.**

# Definíciók

**A binomiális eloszlás akkor lép fel, ha a véletlen kísérletre a következő feltételek teljesülnek:**

1. **Mindig fixen rögzített számú kísérletet végzünk .**

2. **A kísérletek függetlenek. (Bármely egyes kísérlet kimenetele nem befolyásolja a többit.)**

3. **Minden kísérlet kimeneteleit két csoportba lehet sorolni (általában sikeres és sikertelen).**

4. **A siker valószínűsége állandó a különböző kísérletekben.**

# Jelölések a binomiális eloszlással kapcsolatban

**S** és **F** (success és failure) jelöli a két lehetséges kimenet csoportot; $p$ és $q$ jelöli az **S** és **F** valószínűségeit, azaz

$P(S) = p$ ($p$ = az siker valószínűsége)

$P(F) = 1 - p = q$ ($q$ = a sikertelenség vsz.-e)

siker = esemény bekövetkezése

Sikertelenség = nem az esemény következett be

# Jelölések (folyt.)

| | |
|---|---|
| $n$ | jelöli a próbálkozások fix számát. |
| $x$ | jelöli $n$ próbálkozás közül a sikeresek számát, így $x$ bármely egész szám lehet 0 és $n$ között, beleértve a határokat is. |
| $p$ | jelöli a siker valószínűségét egy-egy kísérletben. |
| $q$ | jelöli a sikertelenség valószínűségét egy-egy kísérletben. |
| $P(x)$ | jelöli annak valószínűségét, hogy pontosan $x$ próbálkozás lesz sikeres $n$ próbálkozás közül. |

# A binomiális eloszlás képlete

$$P(x) = \frac{n!}{(n-x)!x!} \cdot p^x \cdot q^{n-x}$$

**ahol $x = 0, 1, 2, \ldots, n$**

**és**

**$n$ = a kísérletek száma**

**$x$ = a sikerek száma az $n$ próbálkozásból**

**$p$ = a siker valószínűsége egy-egy kísérletben**

**$q$ = a sikertelenség valószínűsége ($q = 1 - p$)**

# A képlet indoklása

$$P(x) = \underbrace{\frac{n!}{(n-x)!x!}} \cdot \; p^x \; \cdot \; q^{n-x}$$

**A pontosan *x* sikert tartalmazó kimenetek száma az *n* kísérlet esetén**

# Indoklás (folyt.)

$$P(x) = \underbrace{\frac{n!}{(n-x)!x!}} \cdot \underbrace{p^{x} \cdot q^{n-x}}$$

A pontosan *x* sikert tartalmazó kimenetek száma *n* kísérlet esetén

bármilyen sorrendben bekövetkező *x* siker valószínűsége az *n* kísérlet esetén

# Összefoglalás

**Ebben a fejezetben bemutattuk a:**

- **A binomiális eloszlás definícióját.**
- **Jelölések.**
- **A képlet indoklása.**

# 5-4. fejezet
# A binomiális eloszlás átlaga, varianciája, és szórása

# Kulcsfogalmak

**Ebben a fejezetben a binomiális eloszlás fontosabb tulajdonságait tekintjük át, kiszámítjuk az átlagát, a varianciáját és szórását.**

**Ugyanúgy mint eddig, a cél nem az, hogy ezeket kiszámítsuk, hanem hogy interpretáljuk és megértsük .**

# Diszkrét eloszlásokra vonatkozó képletek:

**Átlag** $\mu = \sum[x \cdot P(x)]$

**Variancia** $\sigma^2 = [\sum x^2 \cdot P(x)] - \mu^2$

**Szórás** $\sigma = \sqrt{[\sum x^2 \cdot P(x)] - \mu^2}$

# A binomiális eloszlásra vonatkozó képletek:

**Átlag** $\mu = n \cdot p$

**Variancia** $\sigma^2 = n \cdot p \cdot q$

**Szórás** $\sigma = \sqrt{n \cdot p \cdot q}$

**Ahol**

**$n$ = a kísérletek rögzített száma**

**$p$ = a siker valószínűsége**

**$q$ = a sikertelenség valószínűsége**

# Összefoglalás

**Ebben a fejezetben megvitattuk a:**

- **A binomiális eloszlás átlagát, varianciáját és szórását.**
- **Az eredmény interpretálását.**

# 5-5. fejezet
# A Poisson eloszlás

# Kulcsfogalmak

**A Poisson eloszlás azért fontos, mert nagyon gyakran használjuk ritka (kis valószínűségű) események eloszlásának leírására.**

# Definíció

**A Poisson eloszlás egy diszkrét eloszlás, ami bizonyos események előfordulásának számát adja meg egy adott intervallumban. Az *x* véletlen változó az események előfordulási száma abban az intervallumban. Az intervallum lehet idő, távolság, terület, térfogat vagy hasonló.**

## Képlete:

$$P(x) = \frac{\mu^{x} \cdot e^{-\mu}}{x!} \quad \text{ahol } e \approx 2.71828$$

# Példa

- Rutherford és Geiger (1910)
- Polonium radioaktív bomlása során az alfa-részecskék számát mérték
- 10.097 alfa részecske 52.16 óra alatt
- 0.0538 alfa részecske/másodperc

# A Poisson eloszlás feltételei

- **Az *x* véletlen változó bizonyos események előfordulásának számát adja meg egy adott intervallumban.**
- **Az előfordulásoknak véletlenszerűeknek kell lenniük.**
- **Az előfordulásoknak függetleneknek kell lenniük egymástól.**
- **Az előfordulásoknak egyenletesen kell eloszlaniuk az intervallumon belül.**

## Paraméterek

- **Az átlaga** $\mu$.
- **A szórás** $\sigma = \sqrt{\mu}$ .

# Eltérés a binomiálishoz képest

**A Poisson és a binomiális között a következő fontos különbségek vannak:**

- **A binomiális eloszlás külön-külön függ a minta $n$ méretétől és a $p$ valószínűségtől, miközben a Poisson csak a $\mu$ átlagtól.**

- **A binomiális esetén az $x$ lehetséges értékei $0, 1, \ldots n$, míg a Poisson eloszlásnál $x$ lehetséges értékei $0, 1, \ldots$, felső határ nélkül.**

# A binomiális közelítése Poissonnal

**A Poisson eloszlással jól közelíthető a binomiális, ha *n* nagy és *p* kicsi.**

## Ökölszabály

- $n \geq 100$
- $np \leq 10$

# A binomiális eloszlás közelítése Poissonnal - $\mu$

- $n \geq 100$
- $np \leq 10$

**$\mu$ kifejezése**

$$\mu = n \cdot p$$

# Összefoglalás

**Ebben a fejezetben megvitattuk a:**

- **Poisson eloszlás definícióját.**
- **A Poisson eloszlás feltételeit.**
- **A Poisson és a binomiális közötti különbséget.**
- **A binomiális Poisson közelítését.**

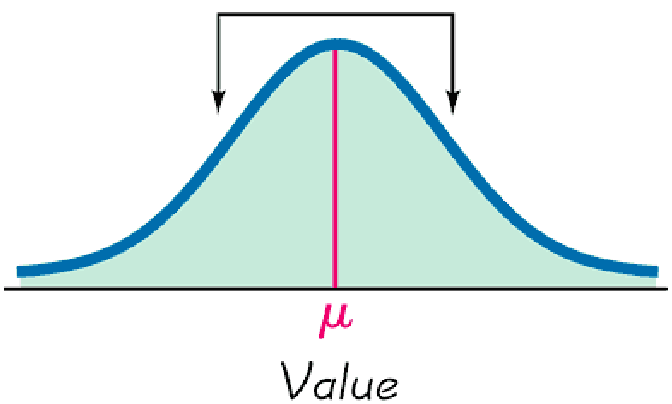
# 6-1. fejezet
# Áttekintés

# Áttekintés

**A következő fejezetek a:**

- **Folytonos változókról**
- **Normális eloszlásról szólnak**

**Harang alakú**
**és szimmetrikus**




$$f(x) = \frac{e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$$

**6-1 képlet**

**6-1 ábra**

# 6-2. fejezet
# A standard normális eloszlás

# Kulcsfogalmak

**Ebben a fejezetben a standard normális eloszlást mutatjuk be, aminek három fő tulajdonsága van:**

**1. Harang alakú.**

**2. Átlaga 0.**

**3. Szórása 1.**

**Nagyon fontos, hogy megtanuljuk, hogyan kell kiszámítani a standard normális eloszlás különböző részei alatti területeket (valószínűségeket vagy relatív gyakoriságot).**

# Definíció

- **Egy folytonos véletlen változó eloszlása egyenletes eloszlás, ha értékei egyenletesen oszlanak el valamilyen intervallumban. Az egyenletes eloszlás téglalap formájú.**

# Definíció

- **Sűrűség függvény** egy folytonos valószínűség eloszlás görbéje. A következő tulajdonságokkal rendelkezik:

1. A görbe alatti teljes terület 1.
2. A görbe minden pontja 0 vagy annál nagyobb.

(A görbe soha nem eshet az x tengely alá.)

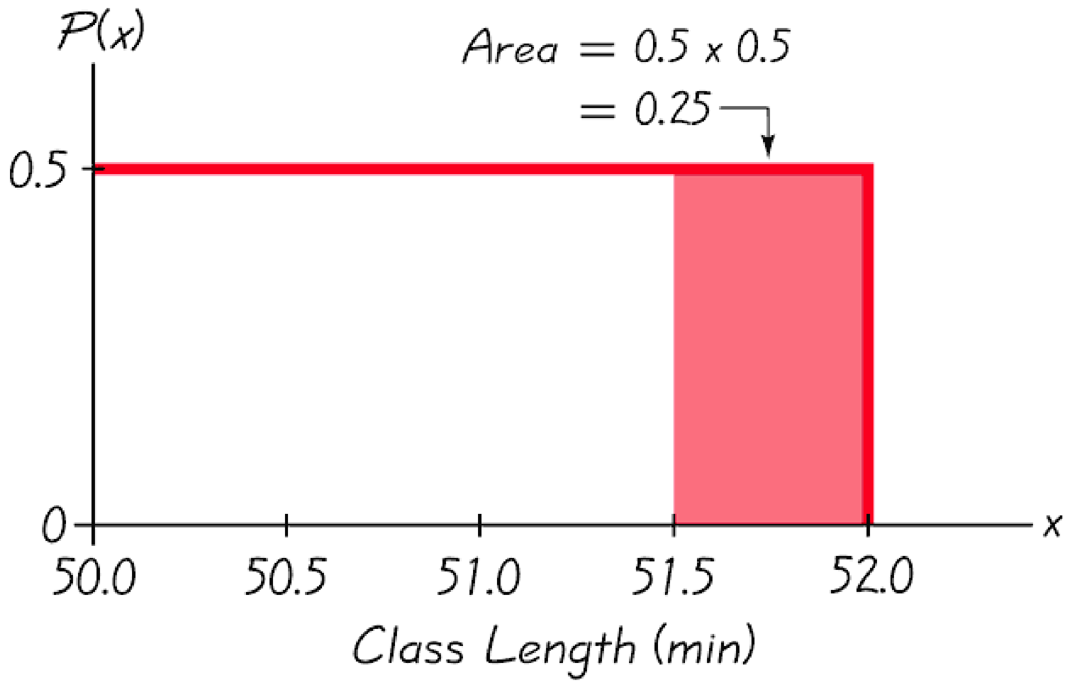
# Terület és valószínűség

**Mivel a görbe alatti terület 1, kapcsolat van a terület és a valószínűség között.**

# Példa

- Mivel az elemi statisztika előadások olyan izgalmasak, hosszuk 50 és 52 perc közötti egyenletes eloszlást mutat ☺.
- Neked 51.5 percnél el kell menned. Mi a valószínűsége annak, hogy lekésed a 6-os villamost?

# A valószínűség kiszámítása a területből



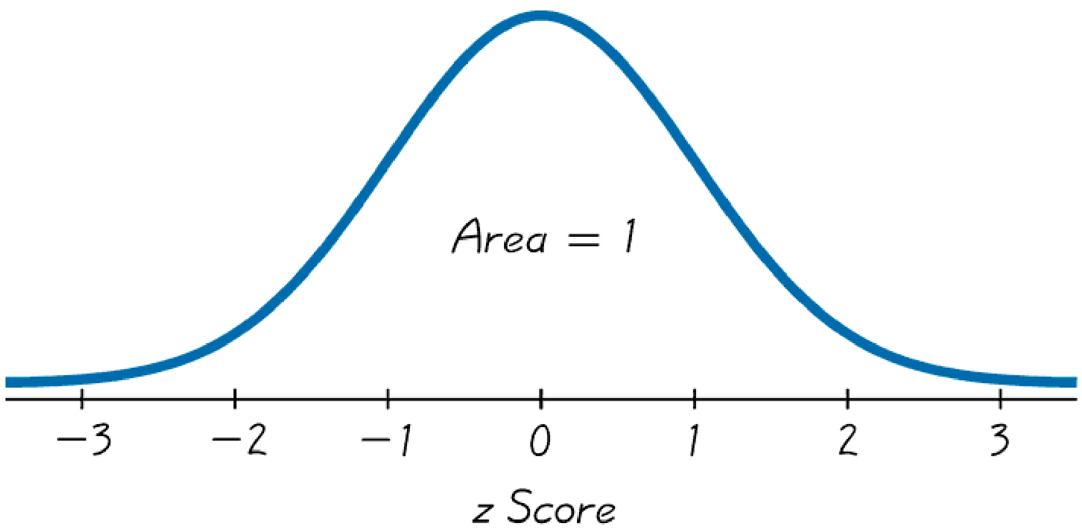

**6-3. ábra**

# Példa - Hőmérők

**Legyenek hőmérőink, amelyek átlagban 0-t mutatnak 1 fok szórással ha fagypontban lévő vízbe helyezzük őket. Számítsuk ki, mi a valószínűsége, hogy egy ilyen hőmérő kevesebb mint 1.58 fokot mutat, ha fagypontban lévő vízbe helyezzük.**

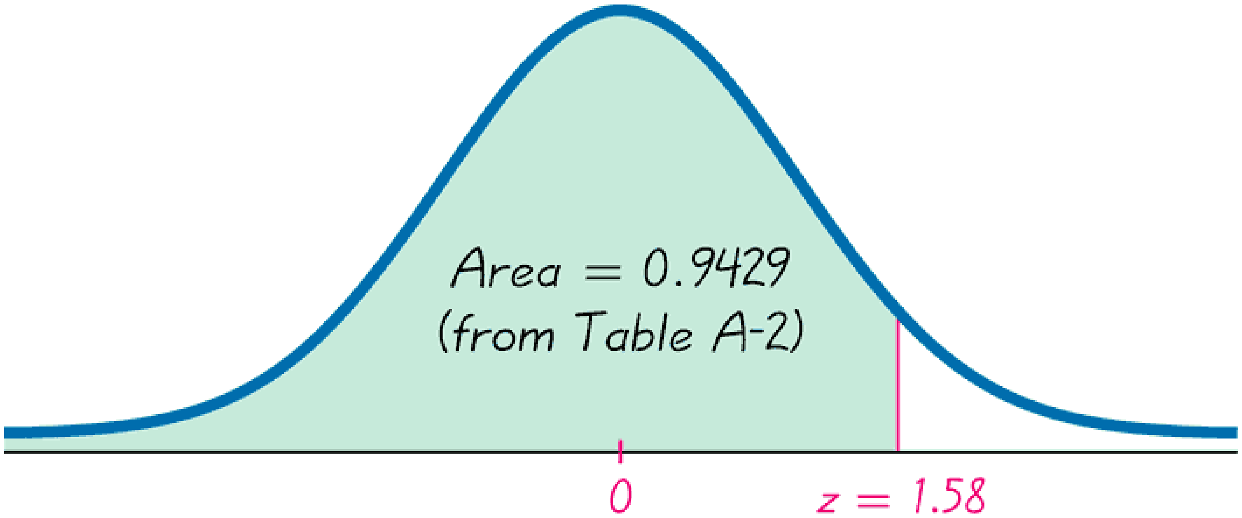
# Definíció

- A **standard normális eloszlás** **egy folytonos valószínűség eloszlás, aminek 0 az átlaga, szórása 1 és a sűrűség függvénye alatti terület is 1.**



# Példa – folyt.

**P(*z* < 1.58) =**




**6-6. ábra**

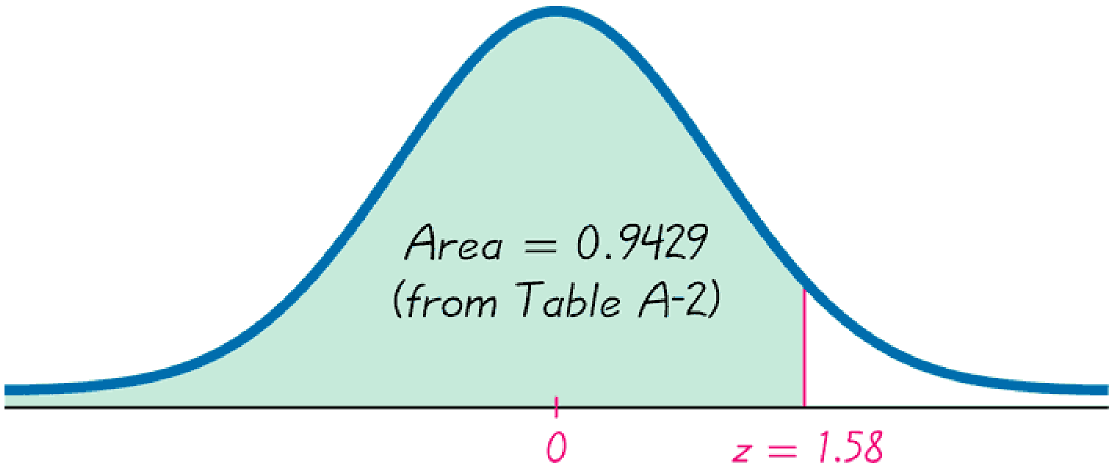
# Standard Normál Eloszlás Táblázat

**Table A-2** Standard Normal (*z*) Distribution: Cumulative Area from the LEFT

| *z* | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| −3.50 and lower | .0001 | | | | | | | | | |
| −3.4 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0002 |
| −3.3 | .0005 | .0005 | .0005 | .0004 | .0004 | .0004 | .0004 | .0004 | .0004 | .0003 |
| −3.2 | .0007 | .0007 | .0006 | .0006 | .0006 | .0006 | .0006 | .0005 | .0005 | .0005 |
| −3.1 | .0010 | .0009 | .0009 | .0009 | .0008 | .0008 | .0008 | .0008 | .0007 | .0007 |
| −3.0 | .0013 | .0013 | .0013 | .0012 | .0012 | .0011 | .0011 | .0011 | .0010 | .0010 |
| −2.9 | .0019 | .0018 | .0018 | .0017 | .0016 | .0016 | .0015 | .0015 | .0014 | .0014 |
| −2.8 | .0026 | .0025 | .0024 | .0023 | .0023 | .0022 | .0021 | .0021 | .0020 | .0019 |
| −2.7 | .0035 | .0034 | .0033 | .0032 | .0031 | .0030 | .0029 | .0028 | .0027 | .0026 |
| −2.6 | .0047 | .0045 | .0044 | .0043 | .0041 | .0040 | .0039 | .0038 | .0037 | .0036 |
| −2.5 | .0062 | .0060 | .0059 | .0057 | .0055 | .0054 | .0052 | .0051 | * .0049 | .0048 |
| −2.4 | .0082 | .0080 | .0078 | .0075 | .0073 | .0071 | .0069 | .0068 | .0066 | .0064 |
| −2.3 | .0107 | .0104 | .0102 | .0099 | .0096 | .0094 | .0091 | .0089 | .0087 | .0084 |
| −2.2 | .0139 | .0136 | .0132 | .0129 | .0125 | .0122 | .0119 | .0116 | .0113 | .0110 |
| −2.1 | .0179 | .0174 | .0170 | .0166 | .0162 | .0158 | .0154 | .0150 | .0146 | .0143 |
| −2.0 | .0228 | .0222 | .0217 | .0212 | .0207 | .0202 | .0197 | .0192 | .0188 | .0183 |
| −1.9 | .0287 | .0281 | .0274 | .0268 | .0262 | .0256 | .0250 | .0244 | .0239 | .0233 |
| −1.8 | .0359 | .0351 | .0344 | .0336 | .0329 | .0322 | .0314 | .0307 | .0301 | .0294 |
| −1.7 | .0446 | .0436 | .0427 | .0418 | .0409 | .0401 | .0392 | .0384 | .0375 | .0367 |
| −1.6 | .0548 | .0537 | .0526 | .0516 | .0505 | * .0495 | .0485 | .0475 | .0465 | .0455 |
| −1.5 | .0668 | .0655 | .0643 | .0630 | .0618 | .0606 | .0594 | .0582 | .0571 | .0559 |

# Példa – folyt.

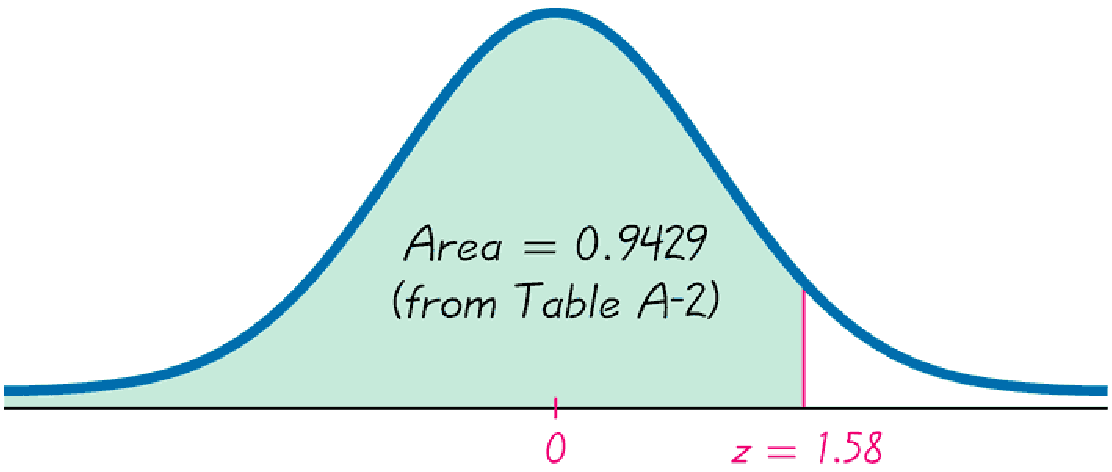
**P ($z$ < 1.58) = 0.9429**




**6-6. ábra**

# Példa – folyt.

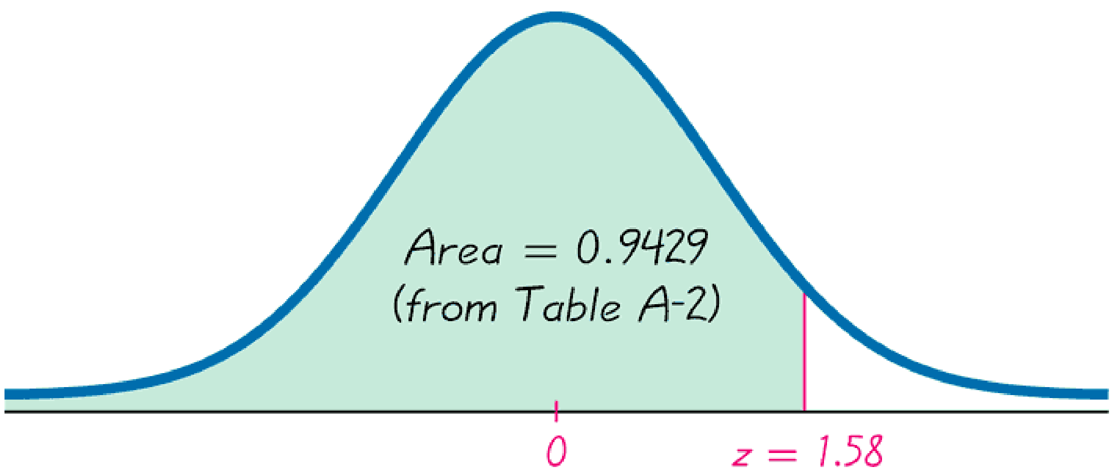
**$P(z < 1.58) = 0.9429$**




**Annak a valószínűsége, hogy az egyik hőmérő kevesebb mint 1.58 fokot mutat 0.9429.**

# Példa – folyt.

**P ($z$ < 1.58) = 0.9429**

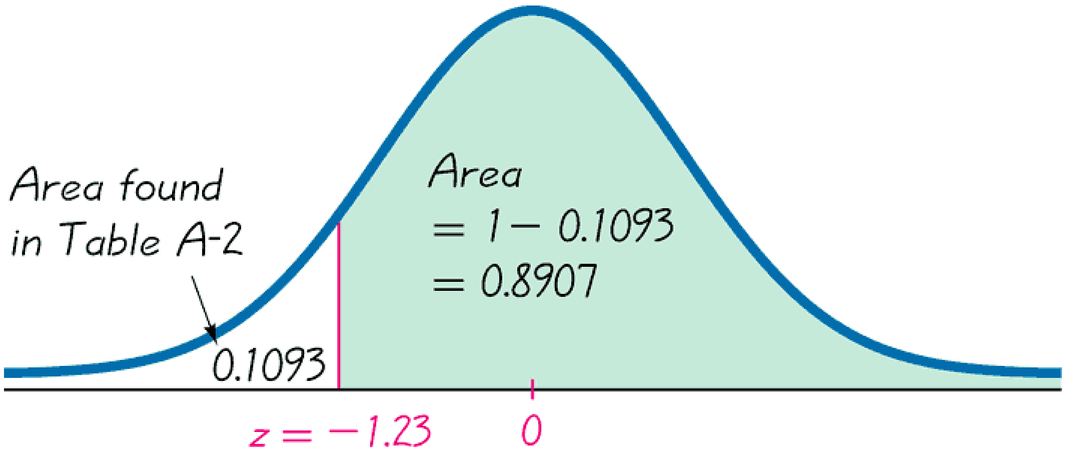



**94.29%-a a hőmérőknek kevesebbet mutat mint 1.58 fok.**

# Példa – folyt.

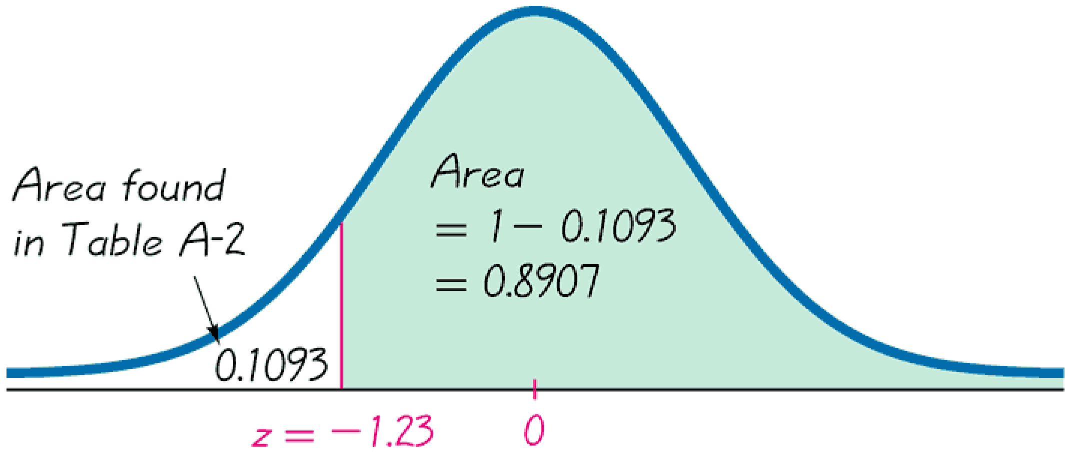
**Ugyanolyan hőmérők esetén mi a vsz.-e, hogy egy véletlenül választott hőmérő többet mutat mint –1.23 fok.**

**P ($z$ > –1.23) = 0.8907**



# Példa – folyt.

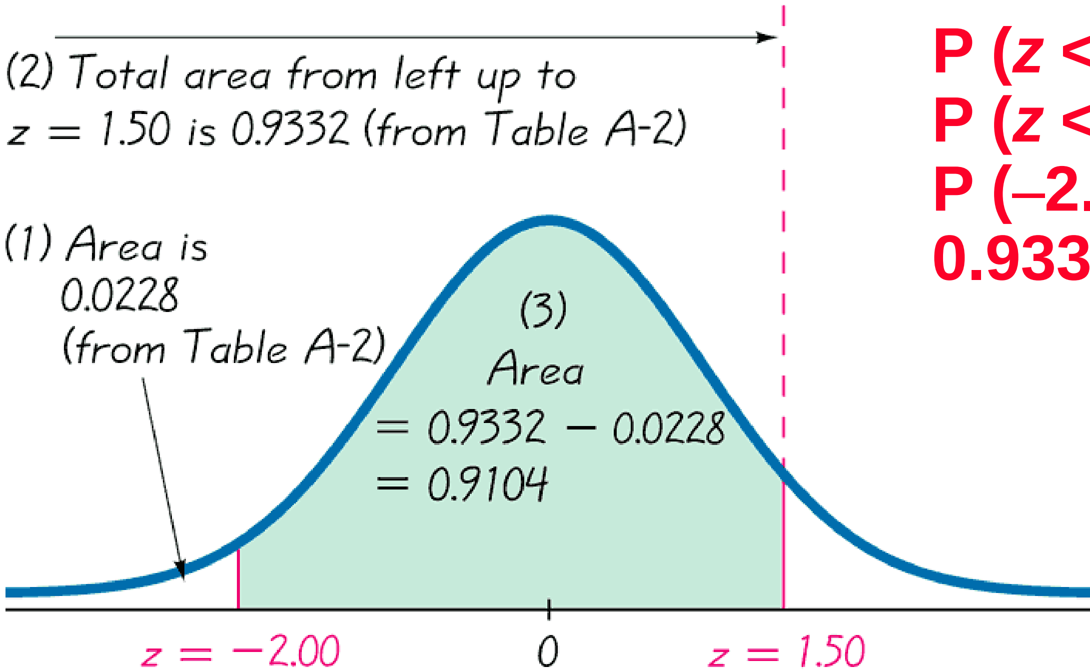
**P ($z$ > –1.23) = 0.8907**




**89.07%-a a hőmérőknek többet mutat mint –1.23 fok.**

# Példa – folyt.

**Mi a vsz.-e, hogy egy véletlenül választott hőmérő –2.00 és 1.50 fokok közötti értéket mutat.**




**$P(z < -2.00) = 0.0228$**
**$P(z < 1.50) = 0.9332$**
**$P(-2.00 < z < 1.50) =$**
**$0.9332 - 0.0228 = 0.9104$**

**Annak a vsz.-e hogy a hőmérő – 2.00 és 1.50 fokok közötti értéket mutat 0.9104.**

# Jelölés

$$P(a < z < b)$$

**jelöli annak a valószínűségét, hogy a *z* érték *a* és *b* közé esik.**

$$P(z > a)$$

**jelöli annak a valószínűségét, hogy egy *z* érték nagyobb mint *a*.**

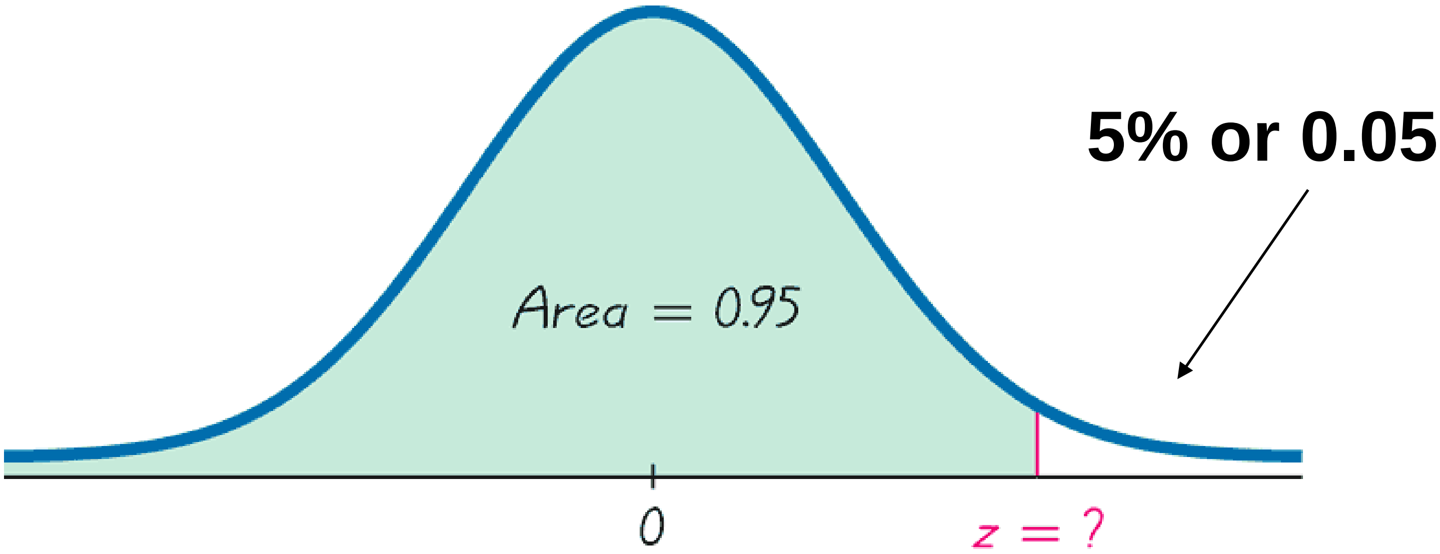
$$P(z < a)$$

**jelöli annak a valószínűségét, hogy egy *z* érték kisebb mint *a*.**

# A *z* érték meghatározása a valószínűségből

**Rajzolj egy haranggörbét és határozd meg azt a területet, ami egy adott valószínűséghez tartozik. Ha ez nem egy baloldalról kumulált terület lenne, akkor vezesd vissza valahogy a problémát ilyenre!**

# z érték meghatározása a valószínűséghez

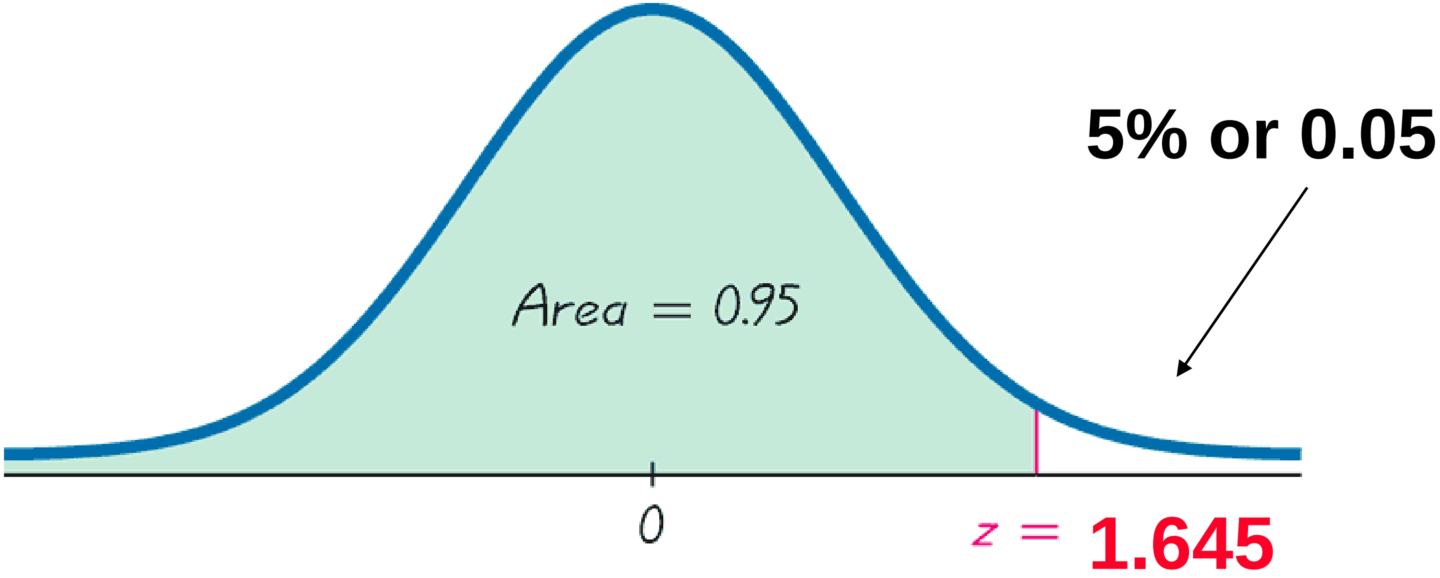



(*z* érték pozitív lesz)

**6-10. ábra**

**A 95. Percentilis meghatározása**

# z érték meghatározása a valószínűséghez

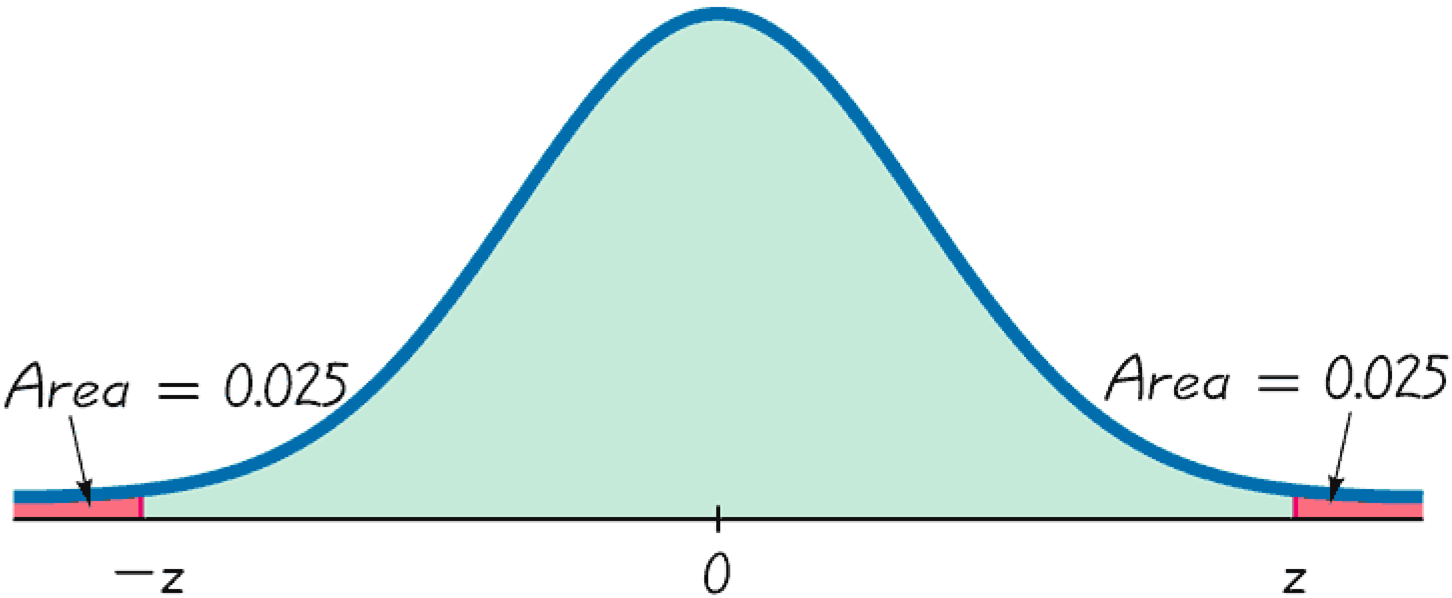



**(*z* érték pozitív lesz)**

**6-10. ábra**

**A 95. Percentilis meghatározása**

# z érték meghatározása a valószínűséghez

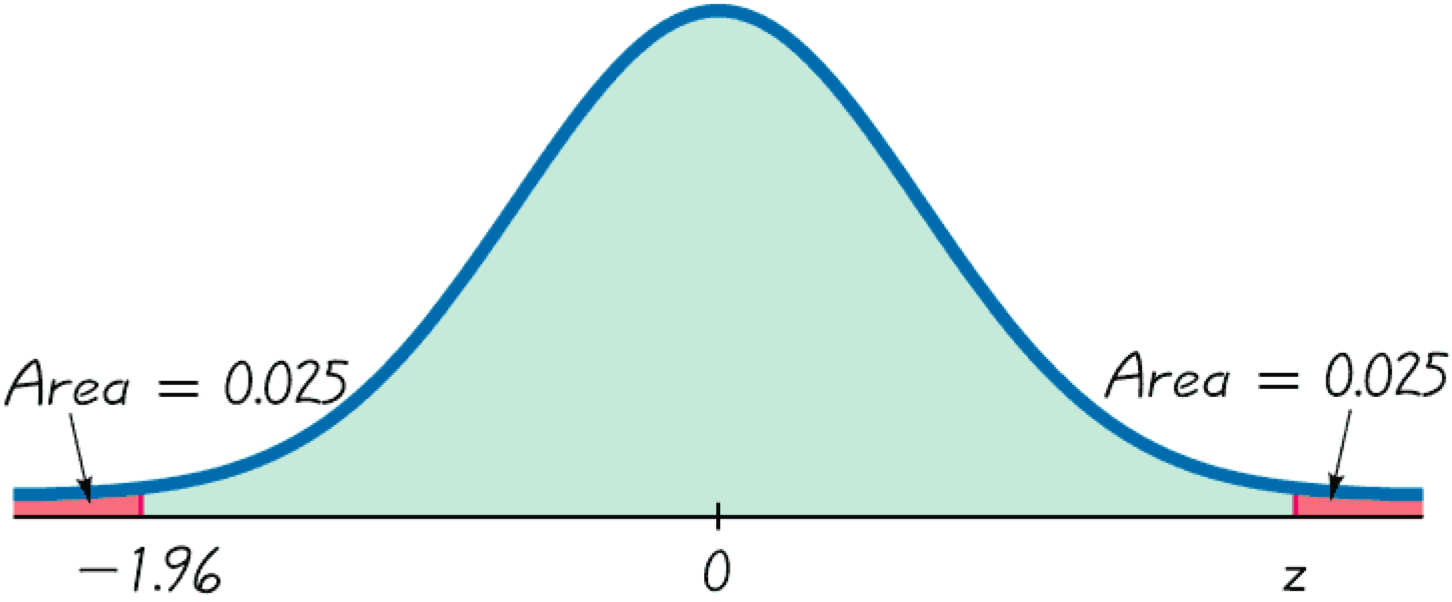



**( Az egyik *z* érték negatív, a másik pozitív lesz)**

**6-11. ábra**

**Az alsó 2.5% és a felső 2.5% meghatározása**

# z érték meghatározása a valószínűséghez

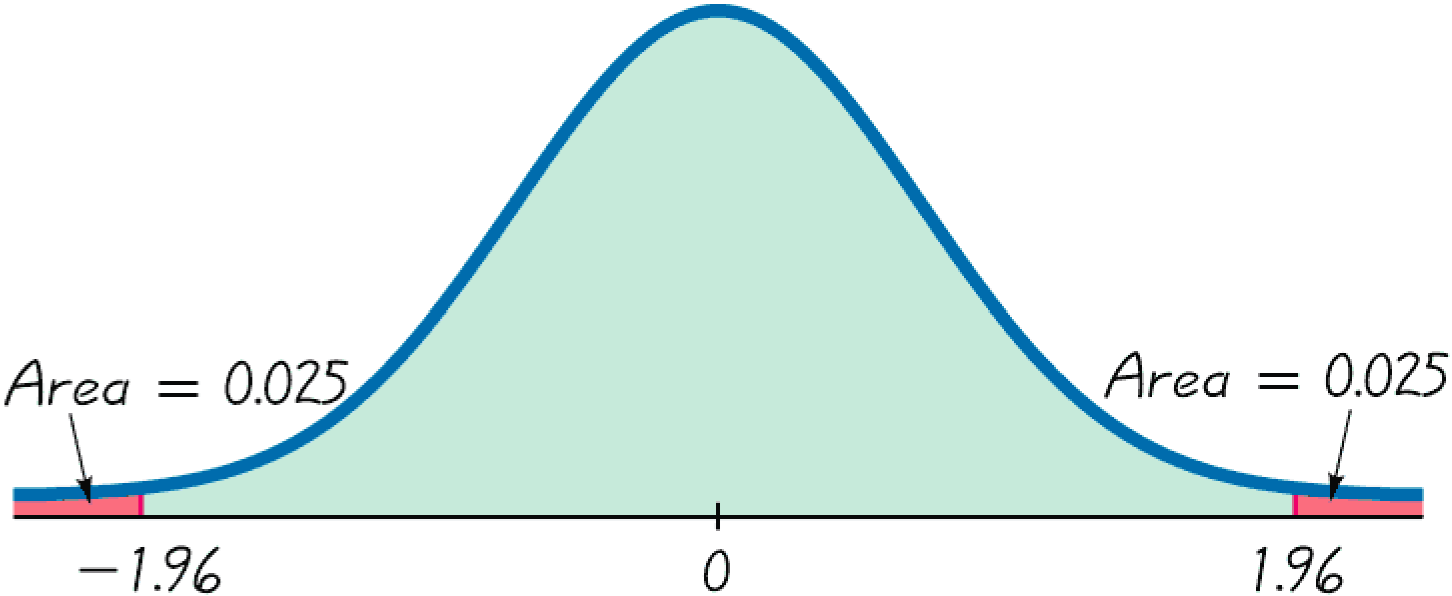



**( Az egyik *z* érték negatív, a másik pozitív lesz)**

**6-11. ábra**

**Az alsó 2.5% és a felső 2.5% meghatározása**

# z érték meghatározása a valószínűséghez




**( Az egyik *z* érték negatív, a másik pozitív lesz)**

**6-11. ábra**

**Az alsó 2.5% és a felső 2.5% meghatározása**

# Összefoglalás

**Ebben a fejezetben megvitattuk:**

- **A sűrűség függvényt.**
- **A terület és a valószínűség közti kapcsolat**
- **Standard normális eloszlás.**