

▼ 15. Konfidencia intervallum becslése átlagra és szórásra normál eloszlás esetén

Valószínűségszámítás és statisztika számítógép algebrai támogatással	Készült a PTE PMMIK-n a TÁMOP -4.1.2.A/1-11/1-2011-0098 támogatásával
Klincsik Mihály	

▼ 15.1. Bevezetés

Statisztikában a pontszerű becslés mellett fontos szerepe van az **intervallum becslés**nek. A kiszámolt értékhez megadunk egy intervallumot, amely nagy valószínűséggel tartalmazza a becsült paramétert. A mérésből számolt érték valamilyen **megbízhatósági** vagy **valószínűségi szinten** bele esik ebbe az intervallumba. Különböző megbízhatósági szinthez különböző konfidencia intervallum tartozik. Egy 95%-os megbízhatósági intervallum azt jelenti, hogy 100 mérési eredményből legalább 95 esetben a számolt érték például az átlag az intervallumba esik. Legfeljebb 5 esetben lehet az intervallumon kívül a paraméter érték. Minden mérés egy valószínűségi változó egy lehetséges értéke. Különböző $x_1, x_2, \dots, x_n$ méréshez különböző $X_1, X_2, \dots, X_n$ valószínűségi változót veszünk fel. Viszont ezek a valószínűségi változók független és azonos eloszlásúak. Jelen feladatokban feltesszük, hogy a közös eloszlás $X_i \sim N(\mu, \sigma)$ ($i = 1, 2, \dots, n$) a normál eloszlás azonos várható értékkel és szórással. Az intervallum becsléséhez **használt eloszlás** attól függ, hogy milyen statisztikával - azaz $h(X_1, X_2, \dots, X_n)$ függvényvel- becsüljük a paramétert és melyik paramétert ismerjük és melyiket kell becsülni. Így az alkalmazások során találkozhatunk **Normál** -eloszlással, a **StudentT**-eloszlással, a **chi-négyzet** eloszlással és **F-eloszlással**.

Célunk az alábbi fogalmak, összefüggések és eljárások megismerése, megértése

Konfidencia intervallum és a szint fogalma	Szórás illetve variancia konfidencia intervallumának megbízhatósági becslése normál eloszlás esetén
Átlag konfidencia intervallumának számítása, ha az adatok normál eloszlásúak és a szórás ismert	A Maple <i>OneSampleZTest</i> , <i>OneSampleZTest[SampleSize]</i> eljárásainak alkalmazása
Átlag konfidencia intervallumának számítása, ha az adatok normál eloszlásúak és a szórást az adatokból számoljuk	A Maple <i>OneSampleTTest</i> eljárásának alkalmazása
Minta méret számítása	A Maple <i>OneSampleChiSquareTest</i> eljárásának alkalmazása
Monte Carlo-módszer alkalmazása konfidencia intervallum szélességének szemléltetésére	

▼ 15.2. A várható érték konfidencia intervalluma normál eloszlás esetén

▼ 15.2.1. Az átlag konfidencia intervalluma ismert szórás esetén

15.2.1.1. DEFINÍCIÓ. Konfidencia intervallum és konfidencia szint

Statisztikában az $[a, b]$ **konfidencia intervallum** a populáció valamely **paraméterére** ad γ **megbízhatósággal** becslést. A statisztikai adatokból számolt $[a, b]$ intervallumba kell esni a vizsgált paraméternek $\gamma \in [0, 1]$ valószínűséggel, ha a kísérletet sokszor megismételjük egymástól függetlenül.

Ekkor az $[a, b]$ intervallumot nevezzük a **konfidencia intervallumnak** és γ az intervallum megbízhatósága vagy **konfidencia szintje**.

A leggyakrabban használt **megbízhatósági** vagy **konfidencia szintek** a $\gamma = 90\%, 95\%, 99\%$ és a **99.9%**. A megbízhatósági szintek a 100% közelében vannak, amelyek értéke valószínűségre visszavetítve 0.9 és 1 közé esnek. A szokásos konfidencia szintek értékeit beállítjuk egy *megbizhatóság* nevű vektorban!

> *restart* :

> *megbizhatóság := Vector([0.9, 0.95, 0.99, 0.999])*

$$\text{megbizhatóság} := \begin{bmatrix} 0.9 \\ 0.95 \\ 0.99 \\ 0.999 \end{bmatrix} \quad (1.2.1.1)$$

A γ megbízhatósági (vagy konfidencia) szint kiegészítője az $\alpha = 1 - \gamma$ (elsőfajú) **hiba maximális elkövetési szintje**, amelyek szokásos értékei $\alpha = 10\%, 5\%, 1\%$ és 0.1% .

> $\alpha := \text{map}(x \rightarrow 1 - x, \text{megbizhatóság})$

$$\alpha := \begin{bmatrix} 0.1 \\ 0.05 \\ 0.01 \\ 0.001 \end{bmatrix} \quad (1.2.1.2)$$

A számítás a standard **N(0,1)** normál eloszlásból kapott olyan $u_{\frac{\alpha}{2}}$ érték meghatározásán

alapszik, amelyre teljesül a

$$P\left(-u_{\frac{\alpha}{2}} < X < u_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

feltétel, ahol X egy 0 várható értékű és 1 szórású normál eloszlású valószínűségi változó! Ha

$\Phi(x) = \int_{-\infty}^x \phi(t) dt$ jelöli X eloszlás függvényét, akkor a szimmetria tulajdonságok

kihasználásával

$$\Phi\left(u_{\frac{\alpha}{2}}\right) - \Phi\left(-u_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

vagy

$$\Phi\left(u_{\frac{\alpha}{2}}\right) - \left(1 - \Phi\left(u_{\frac{\alpha}{2}}\right)\right) = 1 - \alpha, \text{ azaz}$$

$$\Phi\left(u_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$$

Ha a $\gamma = 1 - \alpha$ megbízhatósági szinttel számolunk, akkor

$$1 - \frac{\alpha}{2} = 1 - \frac{1 - \gamma}{2} = \frac{1 + \gamma}{2},$$

ezért a fenti képlet helyett a

$$\Phi\left(u_{\frac{\alpha}{2}}\right) = \frac{1 + \gamma}{2}$$

képlettel dolgozhatunk!

Felírjuk rendre ezeket az egyenleteket különböző α értékekre és az **fsolve** paranccsal megkeressük a közelítő megoldásait. Az összes értéket betesszük egy **u** vektorba!

> *with(Statistics) : X := RandomVariable(Normal(0, 1)) :*

$$u := \text{Vector}\left(\left[\text{seq}\left(\text{fsolve}\left(\text{CDF}(X, t) = 1 - \frac{\alpha_i}{2}\right), i = 1..4\right)\right]\right)$$

$$u := \begin{bmatrix} 1.644853627 \\ 1.959963985 \\ 2.575829304 \\ 3.290526731 \end{bmatrix}$$

(1.2.1.3)

Szemléltessük az $X \sim N(0,1)$ sűrűség függvény görbéje alatti terület 10%, 5%, 1% és 0.1%-ának beszínezésével a **megbízhatósági és hiba színeteket**!

> *suruseg := plot(PDF(X, x), x = -3..3) :*

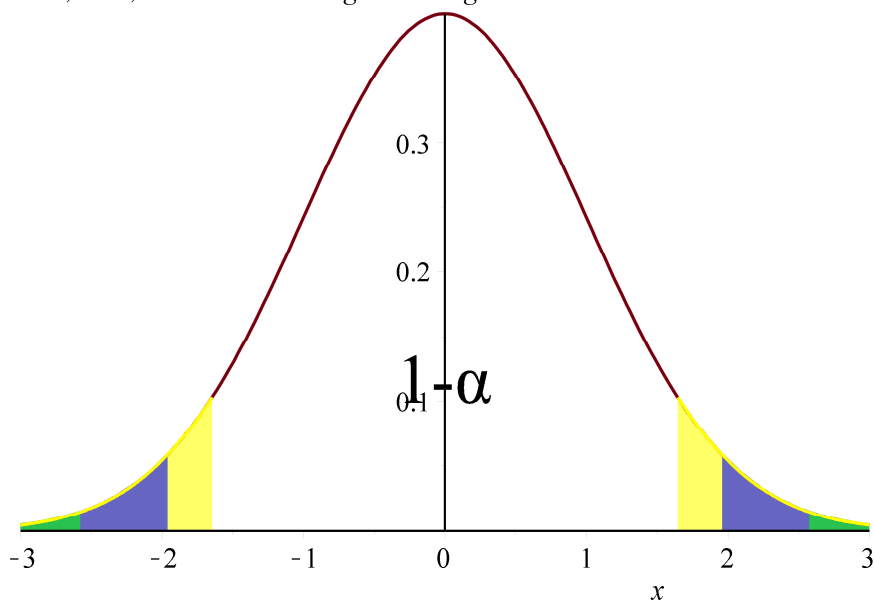
> *bal1 := plot(PDF(X, x), x = -3..-u₁, filled = true, color = yellow) :*

jobb1 := plot(PDF(X, x), x = u₁..3, filled = true, color = yellow) : bal2 := plot(PDF(X, x), x = -3..-u₂, filled = true, color = blue) : jobb2 := plot(PDF(X, x), x = u₂..3, filled = true, color = blue) : bal3 := plot(PDF(X, x), x = -3..-u₃, filled = true, color = green) : jobb3 := plot(PDF(X, x), x = u₃..3, filled = true, color = green) :

sz := plots[textplot]([0, 0.12, "1- α "], font = ["times", "roman", 26]) :

> *plots[display]([suruseg, bal3, jobb3, bal2, jobb2, bal1, jobb1, sz], title = 90%, 95%, 99% és 99.9% megbízhatósági szintekhez tartozó kritikus értékek)*

90%, 95%, 99% és 99.9% megbízhatósági szintekhez tartozó kritikus értékek



A nagy terület, amelyben az $(1 - \alpha)$ szerepel rendre a teljes terület 90%-a, 95%-a, 99%-a illetve 99.9%-a! A színes területek két részből állnak és szimmetrikusan helyezkednek el a 0 várható értékre! A baloldali terület mértéke is $\frac{\alpha}{2}$ és a jobb oldali terület mértéke is $\frac{\alpha}{2}$. Érdemes az ábrát megtekinteni 1:1 arányban, ahol az x és y tengelyen a hosszak egyformák.

A átlag konfidencia intervallumának számítását az alábbi tételre alapozzuk!

15.2.1.2. TÉTEL. Standardizált átlag normál eloszlású

Ha $X_1, X_2, \dots, X_n$ független normál eloszlású valószínűségi változók, amelyek μ várható értéke és σ^2 varianciája közös, akkor az

$$X_{\text{átlag}} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

statisztika normál eloszlású μ várható értékkel és $\frac{\sigma^2}{n}$ varianciával és ezért a

$$Z = \frac{X_{\text{átlag}} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

valószínűségi változó 0 várható értékű és 1 szórású standard normál eloszlású, melynek sűrűség függvénye

$$f(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

A tétel alapján a **várható érték konfidencia intervalluma** az alábbi képlet szerint számítható

$$\left[X_{\text{átlag}} - \frac{u_{\frac{\alpha}{2}} \sigma}{\sqrt{n}}, X_{\text{átlag}} + \frac{u_{\frac{\alpha}{2}} \sigma}{\sqrt{n}} \right]$$

ahol $X_{\text{átlag}} = \frac{\sum_{i=1}^n X_i}{n}$ a mintából számolt átlag, n a minta mérete, a minta σ szórása adott és $u_{\frac{\alpha}{2}}$ az $N(0,1)$ -ből számolt kritikus érték.

15.2.1.3. példa. Ismertek a mérési adatszám, az átlag és a szórás. Az adatok hiányoznak!

Határozzuk meg **95%-os** és **99%-os megbízhatósággal** annak a normál eloszlásnak a várható értékére **konfidencia intervallumot**, melynek **varianciája** $\sigma = 3$ a minta **$n=100$ elemű** és a minta **átlaga** $\mu=5$.

Megoldás

A feladatban nem kell számolni az átlagot, mert adott.

> *átlag := 5; szórás := 3; n := 100*

átlag := 5

szórás := 3

n := 100

(1.2.1.4)

A 95%-os megbízhatósági intervallumhoz az u_2 értéket kell használni. Kiszámoljuk az intervallum hosszát, majd ezt hozzáadjuk és levonjuk az átlagból!

> *k95 := $\frac{u_2 \text{ szórás}}{\sqrt{n}}$*

> *'95%-os konfidencia intervallum' := [átlag - k95, átlag + k95]*
k95 := 0.5879891955

95%-os konfidencia intervallum := [4.412010804, 5.587989196]

(1.2.1.5)

A 99%-os megbízhatósági intervallumhoz az u_3 értéket kell használni.

> *k99 := $\frac{u_3 \text{ szórás}}{\sqrt{n}}$*

> *'99%-os konfidencia intervallum' := [átlag - k99, átlag + k99]*
k99 := 0.7727487912

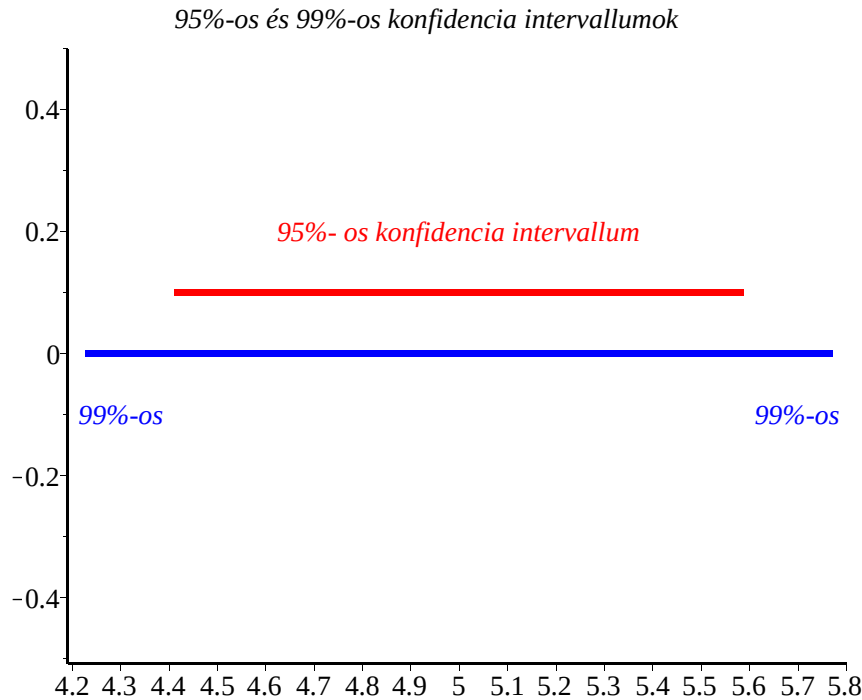
99%-os konfidencia intervallum := [4.227251209, 5.772748791]

(1.2.1.6)

Hasonlítsuk össze a két intervallumot és magyarázzuk meg a különbséget!

> *i95 := plot([['95%-os konfidencia intervallum', 0.1],
 ['95%-os konfidencia intervallum', 0.1]], thickness = 3, color = red) :*
s95 := plots[textplot]([5, 0.2, '95%-os konfidencia intervallum'], color = red) :
*i99 := plot([['99%-os konfidencia intervallum', 0],
 ['99%-os konfidencia intervallum', 0]], thickness = 3, color = blue) :*
s99 := plots[textplot]([4.3, -0.1, '99%-os'], [5.7, -0.1, '99%-os'], color = blue) :

```
plots[display]([i95, s95, i99, s99], xtickmarks = [4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5,
5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8], axes = framed, title
= 95%-os és 99%-os konfidencia intervallumok, view = [4.2 ..5.8, -0.5 ..0.5])
```



15.2.1.4. Tény.

A 99%-os megbízhatósági szinthez tartozó konfidencia intervallum hosszabb, mint a 95%-os megbízhatósági szinthez tartozó intervallum.

Miért? A kapott 99%-os szintű **K99** = [4.2272512095.772748791] konfidencia intervallum statisztikai jelentése az, hogy 100 mintavételből 99 esetben ebbe az intervallumba bele kell esni a minta számított átlagának.

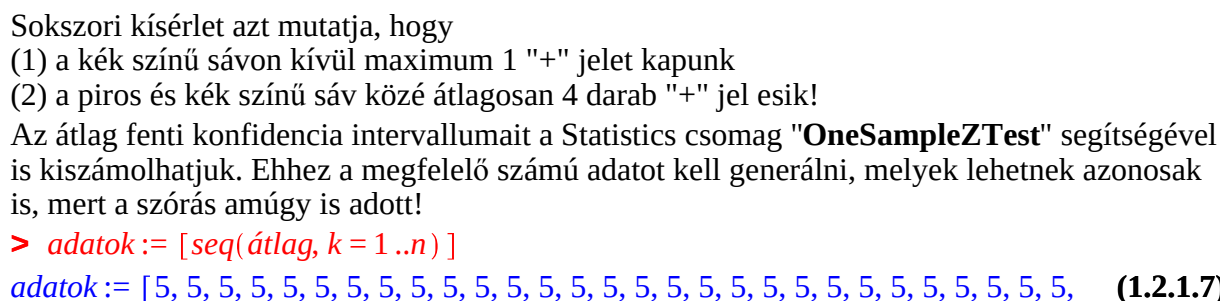
Míg a rövidebb 95%-os szintű **K95** = [4.412010804, 5.587989196] konfidencia intervallum esetén 100 mintavételből csak 95 esetben kell ebbe az intervallumba esni a minta számított átlagának.

A 95%-os eset K95 intervalluma a 100 -ból 4- gyel kevesebb minta átlag helyét **jelöli ki**, ezért valójában a 95 darab minta átlaga kisebb intervallumba esik, mint a több 99 minta átlaga. A 99 -95 = 4 többlet mérési eredmények átlaga a fenti kék színű tartományba eshet. Nem kötelező ugyan a plusz 4 mérési minta átlagának a két intervallum különbségébe esni, de a véletlen folytán megengedett!

Próbáljuk ki számítógépes **szimulációval** vagy **Monte Carlo-módszerrel** a konfidencia intervallumba esést!

Gyártunk 100 darab 100 elemű véletlen mintát és rögtön kiszámítjuk a 100 elemű minta átlagát. A kapott 100 minta átlagát és a konfidencia intervallumok alsó és felső szintjeit felrajzoljuk közös koordináta-rendszerben! A kék színű a 99%-os , a piros színű 95%-os

```
> X := RandomVariable( 'Normal'( átlag, szórás ) ) :
> próba := [ seq( [ i, Mean( Sample( X, 100 ) ) ], i = 1 ..100 ) ] :
> pontok := plot( próba, style = point, symbol = cross ) :
  szint95 := plot( `95%-os konfidencia intervallum`, x = 0 ..100, color = red, thickness
    = 2 ) :
  szint99 := plot( `99%-os konfidencia intervallum`, x = 0 ..100, color = blue, thickness
    = 2 ) :
plots[display]( [ pontok, szint95, szint99 ], title
  = Kék =99%-os szint, piros=95%-os szint)
```



$$L := 1 \quad (1.2.2.1)$$

Adott az átlag és a szórás!

$\text{átlag} := 5; \text{szórás} := 3$

$\text{átlag} := 5$

$\text{szórás} := 3 \quad (1.2.2.2)$

Adott a konfidencia - szint!

$\text{szint} := 0.95$

$\text{szint} := 0.95 \quad (1.2.2.3)$

A $\gamma=0.95$ megbízhatósági szinthez kiszámoljuk az u-értéket!

$\text{with(Statistics) : } X := \text{RandomVariable('Normal'(0, 1))} : e := \text{CDF}(X, t) = \frac{1 + \text{szint}}{2}$

$$\begin{aligned} &> u := \text{fsolve}\left(\text{CDF}(X, t) = \frac{1 + \text{szint}}{2}\right) \\ &e := \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{1}{2} t \sqrt{2}\right) = 0.9750000000 \end{aligned}$$

$u := 1.959963985 \quad (1.2.2.4)$

(i) Felírjuk az intervallum hosszára vonatkozó feltételt!

$$\text{feltetel} := \frac{2 \cdot u \cdot \text{szórás}}{\sqrt{n}} \leq L$$

$$\text{feltetel} := \frac{11.75978391}{\sqrt{n}} \leq 1 \quad (1.2.2.5)$$

Hány elemű mintát kell venni?

$> \text{solve}(\text{feltetel}, \{n\})$

$\{138.2925176000 \leq n\} \quad (1.2.2.6)$

Tehát legalább 139 elemű kell, hogy legyen a minta!

Ugyanezt kapjuk a beépített $\text{OneSampleZTest}[\text{SampleSize}](L, \text{szórás})$ hívás után is

$> \text{OneSampleZTest}[\text{SampleSize}](L, \text{szórás})$

$139 \quad (1.2.2.7)$

(ii) **Monte Carlo-módszer** alapján generáljunk $n = 139$ méretű véletlen mintát és számoljuk ezek átlagát! Megismételve a kísérletet 100-szor rajzoljuk fel a kapott átlagokat az $L = 1$ széles konfidencia intervallum határaival együtt.

$> X := \text{RandomVariable('Normal'(\text{átlag}, \text{szórás}))} : \text{randomize}() :$

$> \text{próba} := [\text{seq}([i, \text{Mean}(\text{Sample}(X, 139))], i = 1..100)] :$

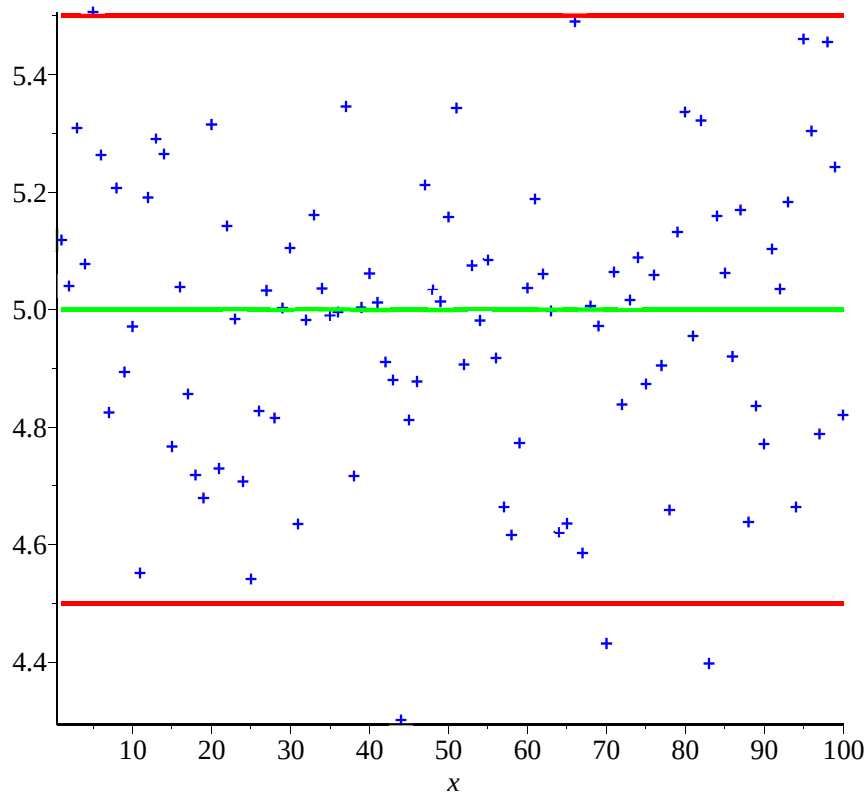
$> \text{pontok} := \text{plot}(\text{próba}, \text{style} = \text{point}, \text{symbol} = \text{cross}, \text{color} = \text{blue}) :$

$\text{szint95} := \text{plot}([4.5, 5, 5.5], x = 1..100, \text{color} = [\text{red}, \text{green}, \text{red}], \text{thickness} = 2) :$

$\text{plots[display]}([\text{pontok}, \text{szint95}], \text{title}$

$= \text{Az } L=1 \text{ hosszú } 95 \% \text{ -os konfidencia szintek közül } 5 \text{ átlag maradhat ki})$

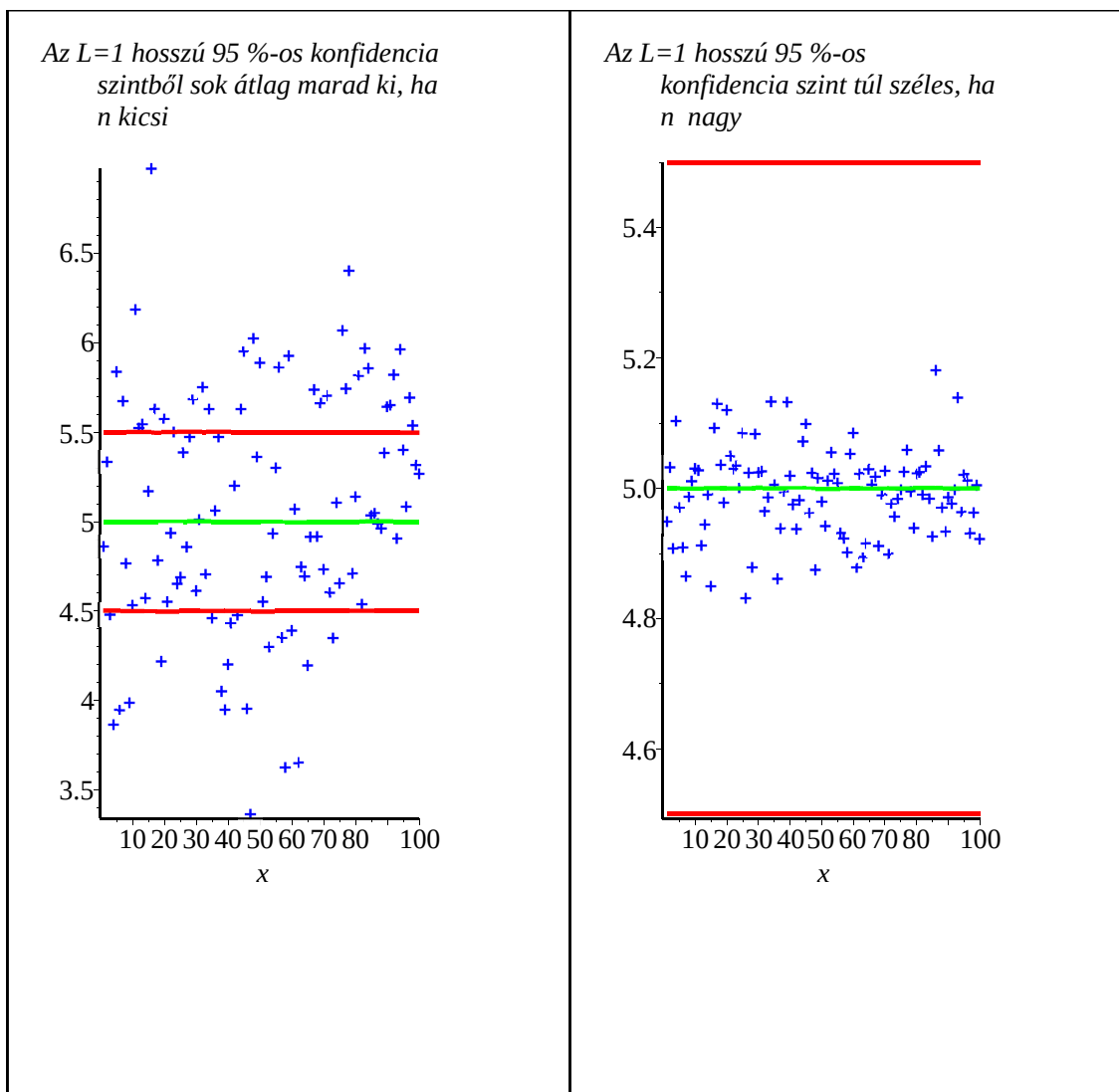
Az $L=1$ hosszú 95 %-os konfidencia szintek közül 5 átlag maradhat ki



Látható, hogy az alsó és felső piros vonalak között van a 100 pontból átlagosan 95 pont és csak 5 marad kívül. Ez a szemléletes jelentése a 95%-os konfidencia intervallumnak!

(iii) Kísérletezzünk a minta n méretével! Vegyünk az (i) pontban kapott n értéknél kisebbet ill. nagyobbat! Mi történik az átlag értékekkel?

```
> randomize( ) :
> próba1 := [seq( [i, Mean(Sample(X, 20)) ], i = 1..100) ] :
> pontok1 := plot(próba1, style = point, symbol = cross, color = blue) :
  szint95 := plot( [4.5, 5, 5.5], x = 1..100, color = [red, green, red], thickness = 2 ) :
  ábra1 := plots[display]( [pontok1, szint95], title
    = Az L=1 hosszú 95 %-os konfidencia szintből sok átlag marad ki, ha n kicsi ) :
> próba2 := [seq( [i, Mean(Sample(X, 2000)) ], i = 1..100) ] :
> pontok2 := plot(próba2, style = point, symbol = cross, color = blue) :
  ábra2 := plots[display]( [pontok2, szint95], title
    = Az L=1 hosszú 95 %-os konfidencia szint túl széles, ha n nagy ) :
plots:-display( Matrix(1, 2, [ábra1, ábra2]) )
```



A fenti kísérlet rávilágít arra, hogy n csökkentésével az átlagok szórása növekszik. Ezért, ha $n = 139$ elemű minta helyett $n = 20$ elemű minta átlagát vesszük, akkor az $L=1$ hosszú konfidencia intervallumból sokkal több pont marad ki, mint 5. Ez látható a bal oldali ábrán. Ha viszont $n = 139$ elemű minta helyett $n = 2000$ elemű minta átlagát vesszük, akkor az $L=1$ hosszú konfidencia intervallum középvonala körül kis szórással tömörülnek a pontok. Tehát az $L=1$ hossz túl nagy lesz. Ez látható a jobb oldali ábrán.

▼ 15.2.3. Az átlag konfidencia intervalluma ismeretlen szórás esetén

Ha a σ szórás ismeretlen, akkor ennek értékét az adatokból kell becsülni

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

képlet alapján! Láttuk, hogy a μ várható érték konfidencia intervallumának becsléséhez fel kellett használnunk az X_i ($i = 1, 2, \dots, n$) változók közös σ szórás értékét. Így, ha a μ várható értéket az

$$X_{\text{átlag}} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

képlet alapján becsüljük, akkor a 15.2.1.2. tételben az $T = X_{\text{átlag}} - \frac{\mu}{\frac{S}{\sqrt{n}}}$ változóra nem lesz

igaz, hogy az normál eloszlású. Pontosan azért, mert S nem egy konstans érték, hanem az $X_1, X_2, \dots, X_n$ valószínűségi változók függvénye. Normál eloszlás helyett ekkor az alábbi

15.2.3.1. tétel alapján t -eloszlással kell dolgoznunk!

15.2.3.1. TÉTEL. Standardizált átlag t -eloszlású

Ha $X_1, X_2, \dots, X_n$ független normál eloszlású valószínűségi változó, melyek μ várható értéke és σ^2 varianciája azonos, akkor a

$$T = \frac{\sqrt{n} (X_{\text{átlag}} - \mu)}{S}$$

valószínűségi változó **t -eloszlású** $(n-1)$ szabadsági fokkal, ahol

$$X_{\text{átlag}} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

az átlag és

$$S^2 = \frac{\sum_{i=1}^n (X_i - X_{\text{átlag}})^2}{n - 1}$$

a korrigált tapasztalati szórás.

Az n - szabadság fokú StudentT -eloszlás sűrűség függvénye

$$f(x) = \frac{\Gamma\left(\frac{1}{2} + \frac{n}{2}\right)}{(\pi n)^{\frac{1}{2}} \Gamma\left(\frac{n}{2}\right) \left(1 + \frac{x^2}{n}\right)^{\frac{1}{2} + \frac{n}{2}}}$$

A tétel alapján a **várható érték konfidencia intervalluma** az alábbi képlet szerint számítható

$$\left[X_{\text{átlag}} - \frac{t_{\frac{\alpha}{2}} S}{\sqrt{n}}, X_{\text{átlag}} + \frac{t_{\frac{\alpha}{2}} S}{\sqrt{n}} \right]$$

ahol $X_{\text{átlag}} = \frac{\sum_{i=1}^n X_i}{n}$ a mintából számolt átlag, **n** a minta mérete, a minta S korrigált szórását az adatokból számoljuk és $t_{\frac{\alpha}{2}}$ az $T[n-1]$ eloszlásból számolt kritikus érték, azaz

$$F\left(t_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} = \frac{1 + \gamma}{2}, \text{ ahol } F(x) \text{ az } (n-1) \text{ szabadságfokú StudentT-eloszlás}$$

eloszlásfüggvénye és $\gamma = 1 - \alpha$ a megbízhatósági szint!

15.2.3.2. PÉLDA. Konfidencia intervallum adatok várható értékére

A következő táblázat 100 darab azonos minőségű betonrúd húzó próbájának szakítási eredményeit mutatja.

[320, 380, 340, 410, 380, 340, 360, 350, 320, 370, 350, 340, 350, 360, 370, 350, 380, 370, 300, 420, 370, 390, 390, 440, 330, 390, 330, 360, 400, 370, 320, 350, 360, 340, 340, 350, 350, 390, 380, 340, 400, 360, 350, 390, 400, 350, 360, 340, 370, 420, 420, 400, 350, 370, 330, 320, 390, 380, 400, 370, 390, 330, 360, 380, 350, 330, 360, 300, 360, 360, 360, 390, 350, 370, 370, 350, 390, 370, 370, 340, 370, 400, 360, 350, 380, 380, 360, 340, 330, 370, 340, 360, 390, 400, 370, 410, 360, 400, 340, 360]

Az adatok alapján az **átlagra** adjunk 95%-os és 99%-os megbízhatósággal konfidencia intervallumot!

A szórást a minta alapján számoljuk, tehát a szórást nem ismerjük.

Megoldás

Kiszámoljuk az adatok átlagát és korrigált szórását!

> *restart* :

> *beton* := [320, 380, 340, 410, 380, 340, 360, 350, 320, 370, 350, 340, 350, 360, 370, 350, 380, 370, 300, 420, 370, 390, 390, 440, 330, 390, 330, 360, 400, 370, 320, 350, 360, 340, 340, 350, 350, 390, 380, 340, 400, 360, 350, 390, 400, 350, 360, 340, 370, 420, 420, 400, 350, 370, 330, 320, 390, 380, 400, 370, 390, 330, 360, 380, 350, 330, 360, 300, 360, 360, 360, 390, 350, 370, 370, 350, 390, 370, 370, 340, 370, 400, 360, 350, 380, 380, 360, 340, 330, 370, 340, 360, 390, 400, 370, 410, 360, 400, 340, 360]

beton := [320, 380, 340, 410, 380, 340, 360, 350, 320, 370, 350, 340, 350, 360, 370, (1.2.3.1)

350, 380, 370, 300, 420, 370, 390, 390, 440, 330, 390, 330, 360, 400, 370, 320, 350, 360, 340, 340, 350, 350, 390, 380, 340, 400, 360, 350, 390, 400, 350, 360, 340, 370, 420, 420, 400, 350, 370, 330, 320, 390, 380, 400, 370, 390, 330, 360, 380, 350, 330, 360, 300, 360, 360, 360, 390, 350, 370, 370, 350, 390, 370, 370, 340, 370, 400, 360, 350, 380, 380, 360, 340, 330, 370, 340, 360, 390, 400, 370, 410, 360, 400, 340, 360]

> *n* := *nops*(*beton*)

> *with*(*Statistics*) : *átlag* := *Mean*(*beton*); *szórás* := *StandardDeviation*(*beton*)

n := 100

átlag := 364.7000000000000

szórás := 26.8348860834383

(1.2.3.2)

A megbízhatósági szintet beállítjuk 95%-ra!

megbizhatosag1 := 0.95

megbizhatosag1 := 0.95

(1.2.3.3)

Kiszámítjuk a *t* - értéket az $(n - 1) = 99$ szabadságfokú *t*-eloszlásból!

X := *RandomVariable*('StudentT'(n - 1)) :

> *t1* := *fsolve*($CDF(X, x) = \frac{1 + \text{megbizhatosag1}}{2}$, *x* = 0..3)

t1 := 1.984216952

(1.2.3.4)

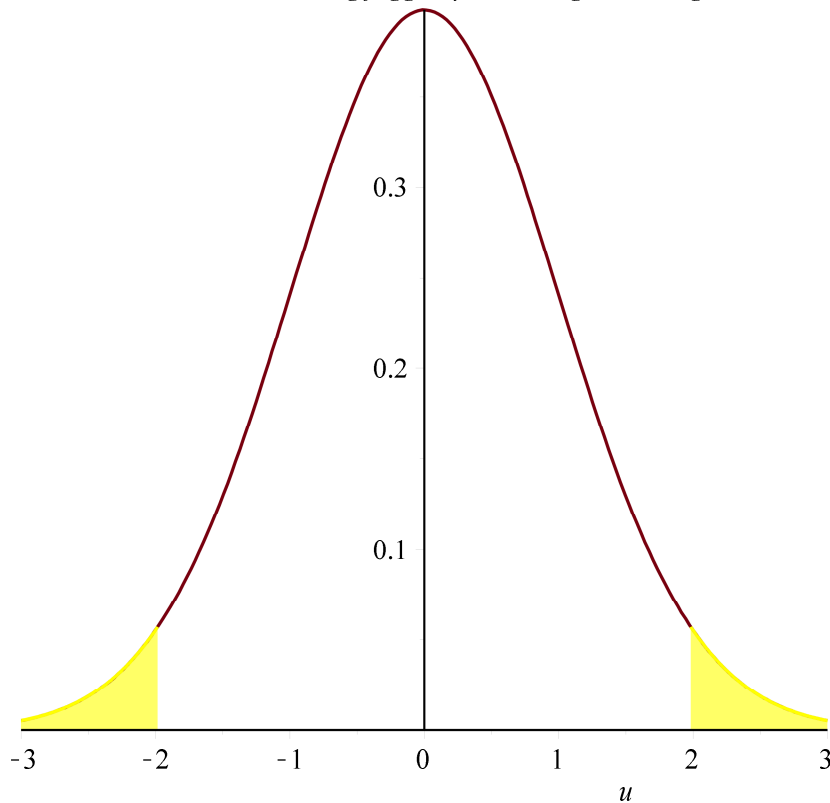
Szemléltessük a fenti számítást!

T-eloszlás sűrűség függvénye_{*n* - 1} = *PDF*(*X*, *u*)

```
> suruseg := plot(PDF(X, u), u = -3 ..3) :
  k1 := plot(PDF(X, u), u = -3 ..-t1, color = yellow, filled = true) : k2 := plot(PDF(X,
    u), u = t1 ..3, color = yellow, filled = true) : abra1 := plots[display]([suruseg, k1,
    k2], title = StudentT-eloszlás sűrűség függvénye és a megbízhatósági szint) : abra1
```

$$= \frac{39614081257132168796771975168}{315285451704888104171289053925} \frac{\sqrt{99}}{\pi \left(1 + \frac{1}{99} u^2\right)^{50}}$$

StudentT-eloszlás sűrűség függvénye és a megbízhatósági szint



A konfidencia intervallum hosszát a $\frac{t_{n-1} \text{ szórás}}{\sqrt{n}}$ képlettel számoljuk!

$$L1 := \frac{t1 \cdot \text{szórás}}{\sqrt{n}}$$

$$L1 := 5.32462358717472$$

(1.2.3.5)

A 95%-os konfidencia intervallum az átlag körüli intervallum, melynek teljes hossza $2 \cdot L$.

$$\text{konfl} := [\text{átlag} - L1, \text{átlag} + L1]$$

$$\text{konfl} := [359.375376412825, 370.024623587175]$$

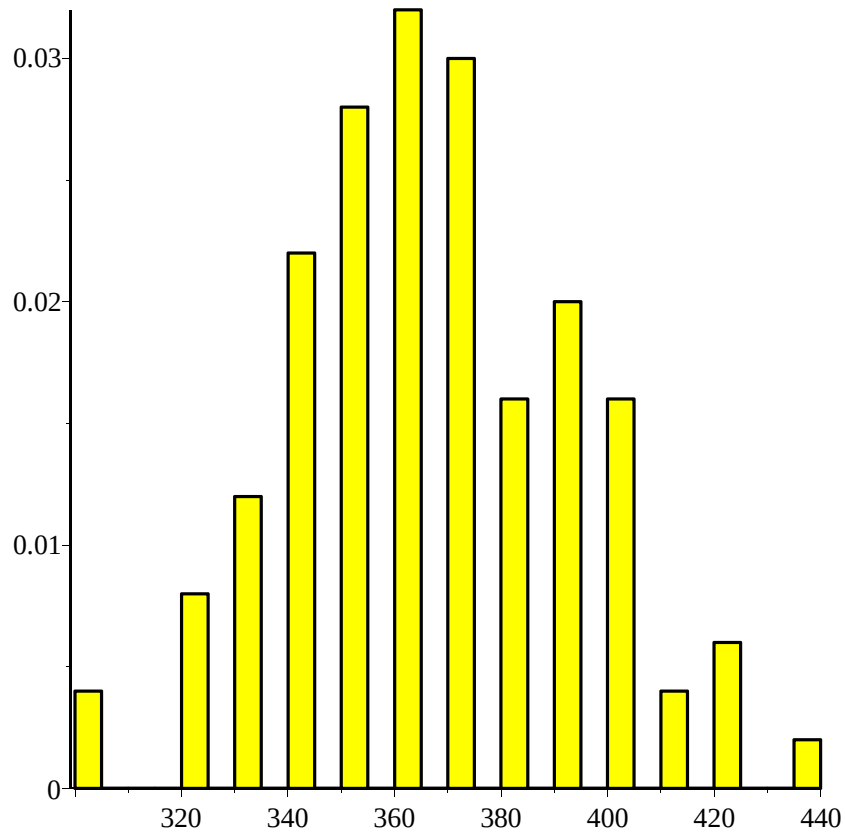
(1.2.3.6)

Szemléltessük az adatokat, majd az átlag konfidencia intervallumát!

```

A, B := min(op(beton)), max(op(beton));
HP := Histogram(beton, binwidth = 5, range = A..B, color = yellow) : HP
A, B := 300, 440

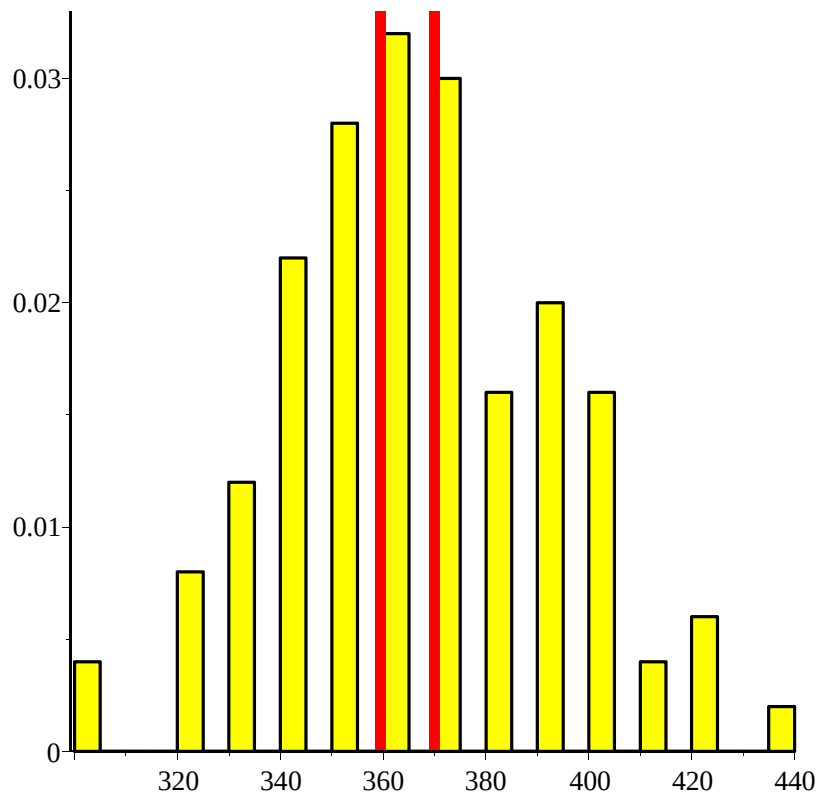
```



```

> vonal1 := plot( [[ [konf1_1, 0], [konf1_1, 0.033]], [[konf1_2, 0], [konf1_2, 0.033]]], thickness
= 5, color = red) :
> plots[display]([HP, vonal1])

```



Növeljük a megbízhatósági szintet 0.95-ről 0.99-re! Hagyan változik a konfidencia intervallum?

megbizhatosag2 := 0.99

megbizhatosag2 := 0.99

(1.2.3.7)

Számoljuk t értékét a $\gamma = 0.99$ megbízhatósághoz!

t2 := fsolve($CDF(X, x) = \frac{1 + megbizhatosag2}{2}$, x = 0 ..3)

t2 := 2.626405457

(1.2.3.8)

Szemléltessük a kapott t - értéket!

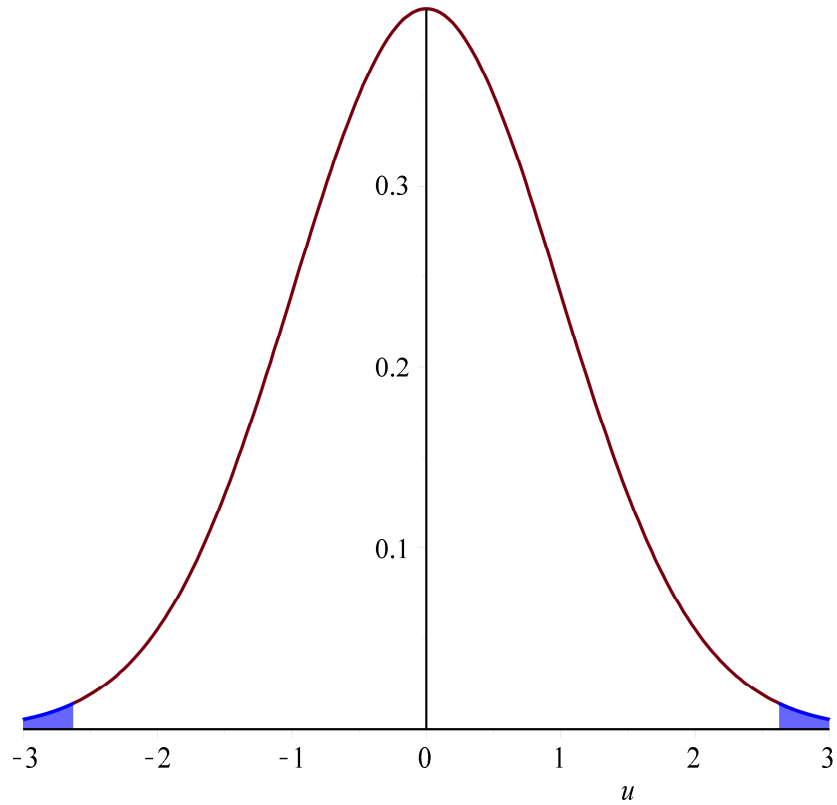
k3 := plot(PDF(X, u), u = -3 .. -t2, color = blue, filled = true) :

k4 := plot(PDF(X, u), u = t2 ..3, color = blue, filled = true) : abra2 := plots[display

]([suruseg, k3, k4], title

= StudentT-eloszlás sűrűség függvénye és a megbízhatósági szint) : abra2

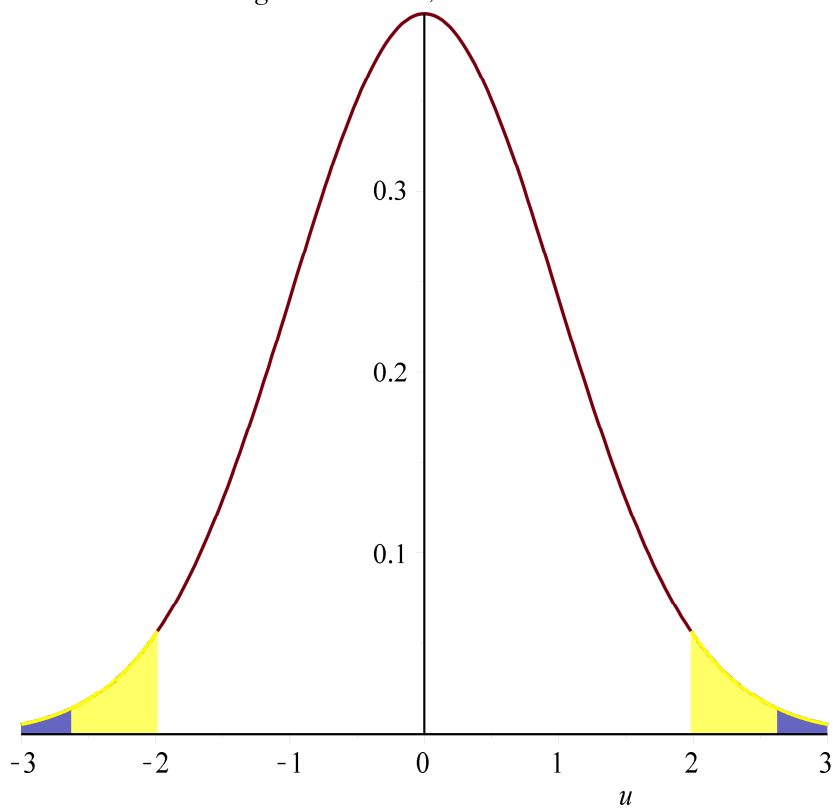
StudentT-eloszlás sűrűség függvénye és a megbízhatósági szint



A 95%-os szinthez tartozó t_2 - érték kisebb, mint a 99%-os szinthez tartozó t-érték, mert a 99%-hoz tartozó kék terület kisebb!

`plots[display]([abra2, abra1], title = Sárga terület=0.05, kék terület=0.01)`

Sárga terület=0.05, kék terület=0.01



A 99%-os konfidencia intervallum hossza és a hisztogram együtt!

$$L2 := \frac{t2 \text{ szórás}}{\sqrt{n}};$$

> $konf2 := [\text{átlag} - L2, \text{átlag} + L2]$

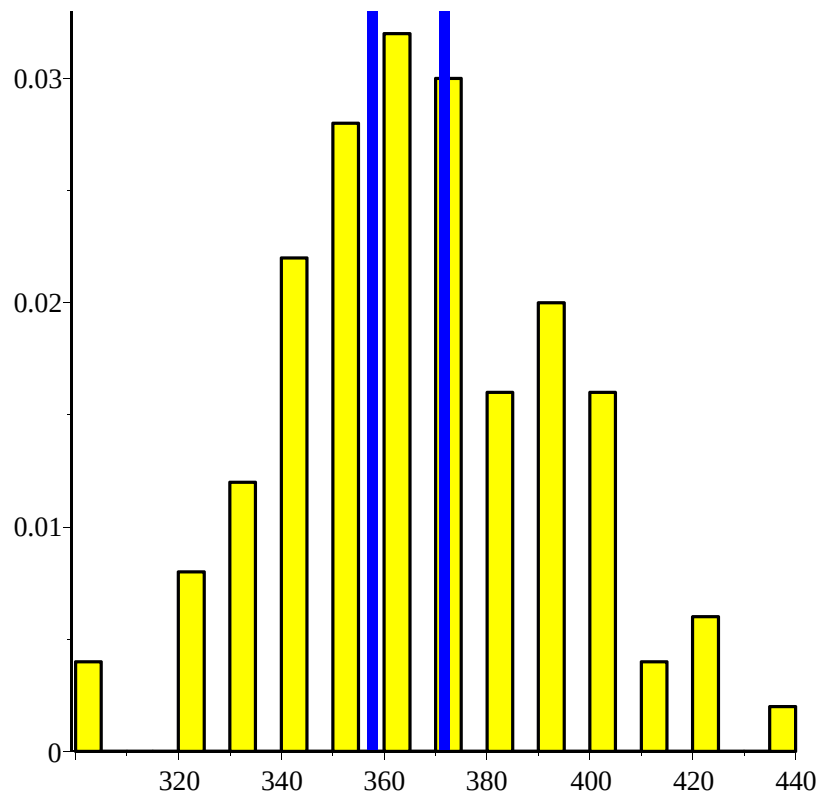
$$L2 := 7.04792912475158$$

$$konf2 := [357.652070875248, 371.747929124752]$$

(1.2.3.9)

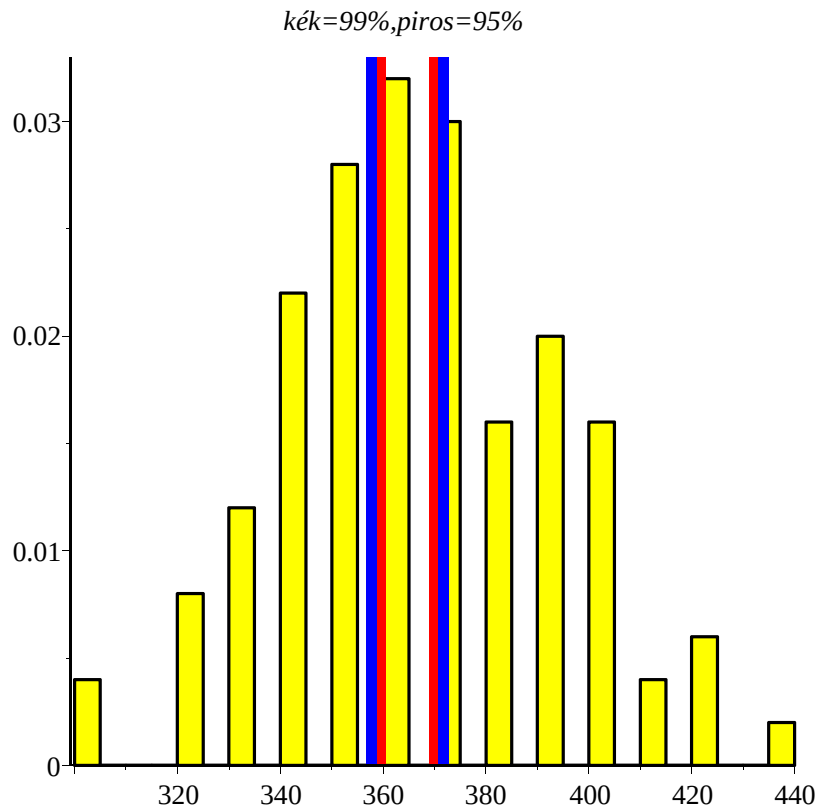
> $vonal2 := \text{plot}([[konf2_1, 0], [konf2_1, 0.033]], [[konf2_2, 0], [konf2_2, 0.033]], \text{thickness} = 5, \text{color} = \text{blue}) :$

> $\text{plots}[\text{display}]([HP, vonal2])$



A két megbízhatósági becslés együtt!

plots[display]([HP, vonal1, vonal2], title = kék=99%,piros=95%)



A **Statistics** csomag **OneSampleTTest** eljárása kiszámolja ugyanezeket a konfidencia intervallumokat, a megfelelő megbízhatósági szinthez!

inforevel[Statistics] := 1 : OneSampleTTest(beton, 364, confidence = 0.95);

Általunk számolt 95%-os konfidencia intervallum = konf1

Standard T-Test on One Sample

Null Hypothesis:

Sample drawn from population with mean 364

Alt. Hypothesis:

Sample drawn from population with mean not equal to 364

Sample size:	100
Sample mean:	364.7
Sample standard dev.:	26.8349
Distribution:	StudentT(99)
Computed statistic:	0.260854
Computed pvalue:	0.794747
Confidence interval:	359.375376413939 .. 370.024623586061
	(population mean)

Result: [Accepted]

There is no statistical evidence against the null hypothesis
hypothesis = true, confidenceinterval = 359.375376413939 ..370.024623586061,

distribution = StudentT(99), pvalue = 0.794746579162011, statistic
= 0.260854470491681

Átalunk számolt 95%-os konfidencia intervallum = [359.375376412825, (1.2.3.10)
370.024623587175]

> *OneSampleTTest(beton, 364, confidence = 0.99);*
Átalunk számolt 99%-os konfidencia intervallum = konf2

Standard T-Test on One Sample

Null Hypothesis:

Sample drawn from population with mean 364

Alt. Hypothesis:

Sample drawn from population with mean not equal to 364

Sample size:	100
Sample mean:	364.7
Sample standard dev.:	26.8349
Distribution:	StudentT(99)
Computed statistic:	0.260854
Computed pvalue:	0.794747
Confidence interval:	357.652070874505 .. 371.747929125495 (population mean)

Result: [Accepted]

There is no statistical evidence against the null hypothesis
hypothesis = true, confidenceinterval = 357.652070874505 ..371.747929125495,

distribution = StudentT(99), pvalue = 0.794746579162011, statistic
= 0.260854470491681

Átalunk számolt 99%-os konfidencia intervallum = [357.652070875248, (1.2.3.11)
371.747929124752]

Látható, hogy az általunk számolt konfidencia intervallumok megegyeznek a *OneSampleTTest* nevű egy-mintás T-teszt által kapott konfidencia intervallummal.

▼ 15.3. A szórás konfidencia intervalluma normál eloszlás esetén

A 15.2.3.2 példa esetén számoljuk a **szórás konfidencia intervallumát** a $\gamma = 0.95$ megbízhatósági szinthez!

A számítást az alábbi tételre alapozzuk!

15.3.1. Tétel. Statisztika a variancia becslésére

Ha $X_1, X_2, \dots, X_n$ független normál eloszlású valószínűségi változó, melyek μ várható értéke és a σ szórása azonos, akkor a

$$Y = \frac{(n-1) S^2}{\sigma^2}$$

valószínűségi változó **chi - négyzet eloszlású** (n-1) szabadsági fokkal, ahol

$$S^2 = \frac{\sum_{i=1}^n (X_i - X_{\text{átlag}})^2}{n-1}$$

a **korrigált tapasztalati szórásnégyzet**.

Az n szabadsági fokú chi-négyzet eloszlás sűrűség függvénye

$$f(x) = \begin{cases} 0 & \text{ha } x < 0 \\ \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} & \text{egyébként} \end{cases}$$

Mivel a chi-négyzet eloszlás nem szimmetrikus, ezért a tétel alapján a γ megbízhatósági szinthez meg kell határozunk olyan c_1 és c_2 értékeket, melyekre

$$P(Y \leq c_1) = F(c_1) = \frac{1-\gamma}{2} = \frac{\alpha}{2} \quad \text{illetve} \quad P(Y \leq c_2) = F(c_2) = \frac{1+\gamma}{2} = 1 - \frac{\alpha}{2}$$

teljesül!

Így

$$P(c_1 < Y < c_2) = F(c_2) - F(c_1) = \frac{1+\gamma}{2} - \frac{1-\gamma}{2} = \gamma = 1 - \alpha.$$

A $c_1 < Y < c_2$ egyenlőtlenséget visszatranszformáljuk a σ^2 varianciára vonatkozó egyenlőtlenségre és így az

$$\frac{n-1}{c_2} \cdot S^2 < \sigma^2 < \frac{n-1}{c_1} \cdot S^2$$

konfidencia intervallumot kapjuk a **szórásnégyzetre**!

15.3.2. PÉLDA. Konfidencia intervallum szórássra adatok alapján

A következő táblázat 100 darab azonos minőségű betonrúd húzó próbájának szakítási eredményeit mutatja.

[320, 380, 340, 410, 380, 340, 360, 350, 320, 370, 350, 340, 350, 360, 370, 350, 380, 370, 300, 420, 370, 390, 390, 440, 330, 390, 330, 360, 400, 370, 320, 350, 360, 340, 340, 350, 350, 390, 380, 340, 400, 360, 350, 390, 400, 350, 360, 340, 370, 420, 420, 400, 350, 370, 330, 320, 390, 380, 400, 370, 390, 330, 360, 380, 350, 330, 360, 300, 360, 360, 360, 390, 350, 370, 370, 350, 390, 370, 370, 340, 370, 400, 360, 350, 380, 380, 360, 340, 330, 370, 340, 360, 390, 400, 370, 410, 360, 400, 340, 360]

Az adatok alapján a **szórássra** adjunk 95%-os megbízhatósággal konfidencia intervallumot!

Megoldás

Kezdjük a számításokat az adatok megadásával!

restart :

> beton := [320, 380, 340, 410, 380, 340, 360, 350, 320, 370, 350, 340, 350, 360, 370, 350, 380, 370, 300, 420, 370, 390, 390, 440, 330, 390, 330, 360, 400, 370, 320, 350, 360, 340, 340, 350, 350, 390, 380, 340, 400, 360, 350, 390, 400, 350, 360, 340, 370, 420, 420, 400, 350, 370, 330, 320, 390, 380, 400, 370, 390, 330, 360, 380, 350, 330, 360, 300, 360, 360, 360, 390, 350, 370, 370, 350, 390, 370, 370, 340, 370, 400, 360, 350, 380, 380, 360, 340, 330, 370, 340, 360, 390, 400, 370, 410, 360, 400, 340, 360]

```

300, 360, 360, 360, 390, 350, 370, 370, 350, 390, 370, 370, 340, 370, 400, 360, 350,
380, 380, 360, 340, 330, 370, 340, 360, 390, 400, 370, 410, 360, 400, 340, 360 ] :
> n := nops(beton)
> with(Statistics) : átlag := Mean(beton); variancia := Variance(beton);
szórás :=  $\sqrt{\text{variancia}}$ 
n := 100
átlag := 364.7000000000000
variancia := 720.1111111111111
szórás := 26.8348860834383

```

(1.3.1)

Legyen a megbízhatósági szint $\gamma=0.95$.

```

megbizhatóság := 0.95
megbizhatóság := 0.95

```

(1.3.2)

A c1 és c2 számítása

```

X := RandomVariable('ChiSquare'(n - 1)) :
> c1 := fsolve(CDF(X, t) =  $\frac{1 - \text{megbizhatóság}}{2}$ , t = 70 ..100);
c2 := fsolve(CDF(X, t) =  $\frac{1 + \text{megbizhatóság}}{2}$ , t = 100 ..130)
c1 := 73.36108019
c2 := 128.4219886

```

(1.3.3)

Szemléltessük a számítást!

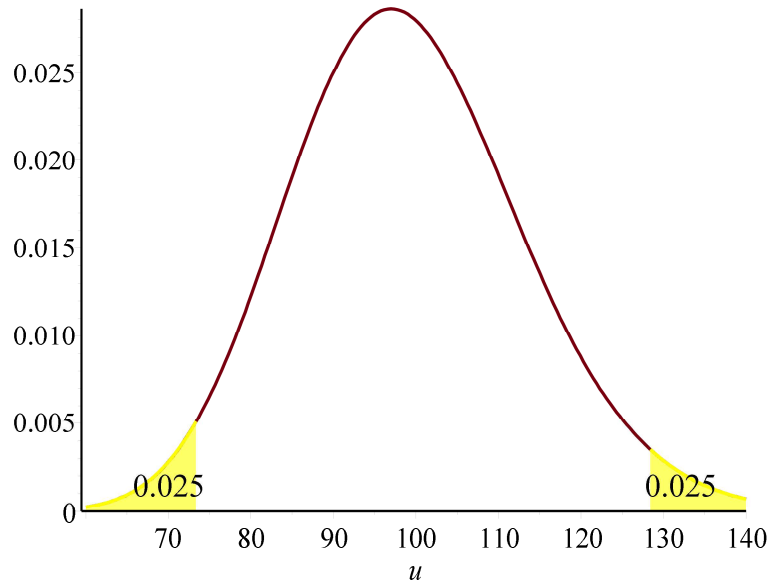
```

Chi-négyzet eloszlás sűrűség függvényen-1 = PDF(X, u)
> suruseg := plot(PDF(X, u), u = 60 ..140) :
> k1 := plot(PDF(X, u), u = 60 ..c1, color = yellow, filled = true) :
k2 := plot(PDF(X, u), u = c2 ..140, color = yellow, filled = true) :
sz := plots[textplot]( [ [ 70, 0.0015, "0.025"], [ 132, 0.0015, "0.025"] ], font = [ "times",
"roman", 12 ] ) :
abra1 := plots[display]( [ suruseg, k1, k2, sz ], title
= Khi-négyzet eloszlás sűrűség függvénye és a megbízhatósági szint ) :
abra1

```

$$\text{Chi-négyzet eloszlás sűrűség függvénye}_{99} = \begin{cases} 0 & \text{if } u < 60 \text{ or } u > 140 \\ \frac{1}{5505842706567130309051945950304886127860194607163} & \text{if } 60 \leq u \leq 140 \end{cases}$$

Khi-négyzet eloszlás sűrűség függvénye és a megbízhatósági szint



A szórás négyzet konfidencia intervallumának végpontjai!

$$> k1 := \frac{(n-1) \text{ variancia}}{c1}; k2 := \frac{(n-1) \text{ variancia}}{c2}$$

$$k1 := 971.782310393486$$

$$k2 := 555.130790117667$$

(1.3.4)

A **szórásnégyzet** konfidencia intervalluma!

$$\text{konf_variancia} := [k2, k1]$$

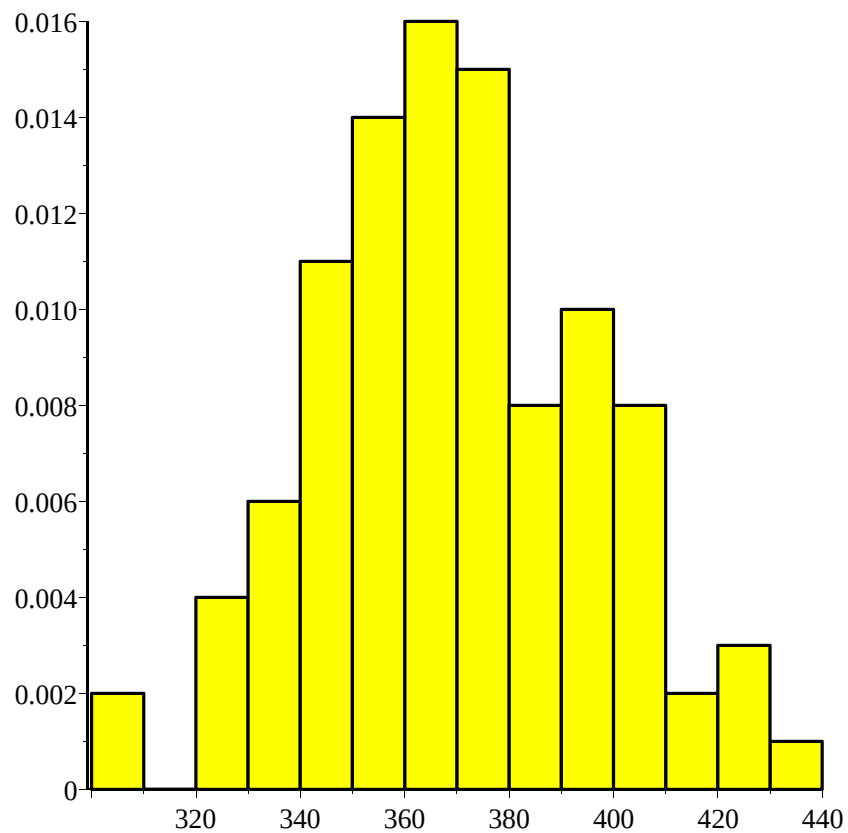
$$\text{konf_variancia} := [555.130790117667, 971.782310393486]$$

(1.3.5)

Rajzoljuk fel a hisztogramhoz a két normál eloszlást, amelyek csak a szórásban különböznek!

$$A, B := \min(\text{beton}), \max(\text{beton}) :$$

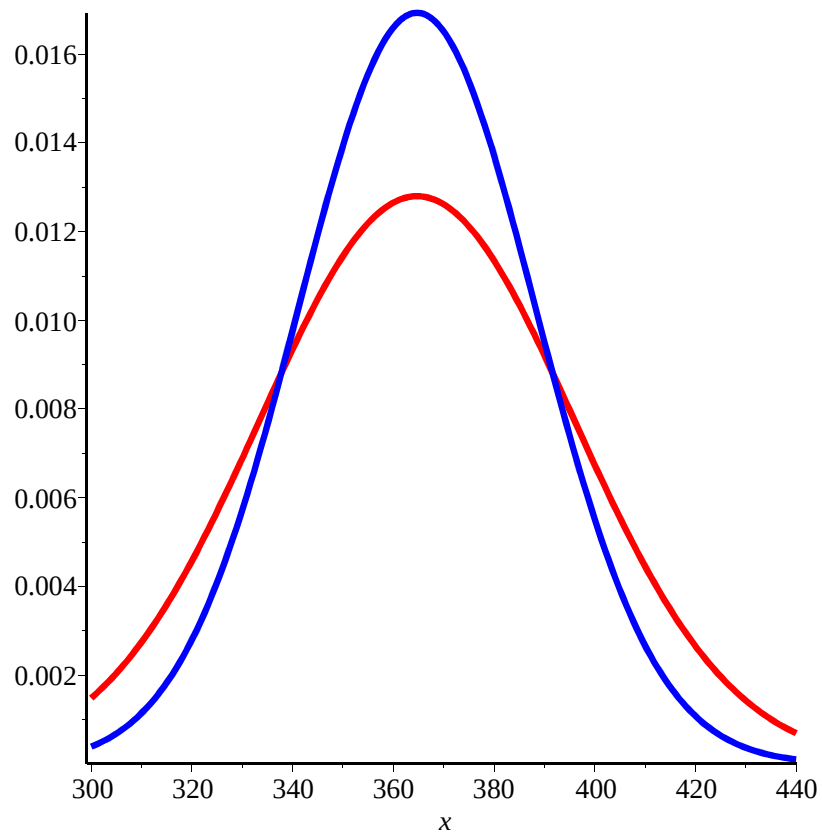
$$HP := \text{Histogram}(\text{beton}, \text{binwidth} = 10, \text{range} = A..B, \text{color} = \text{yellow}) : HP$$



```

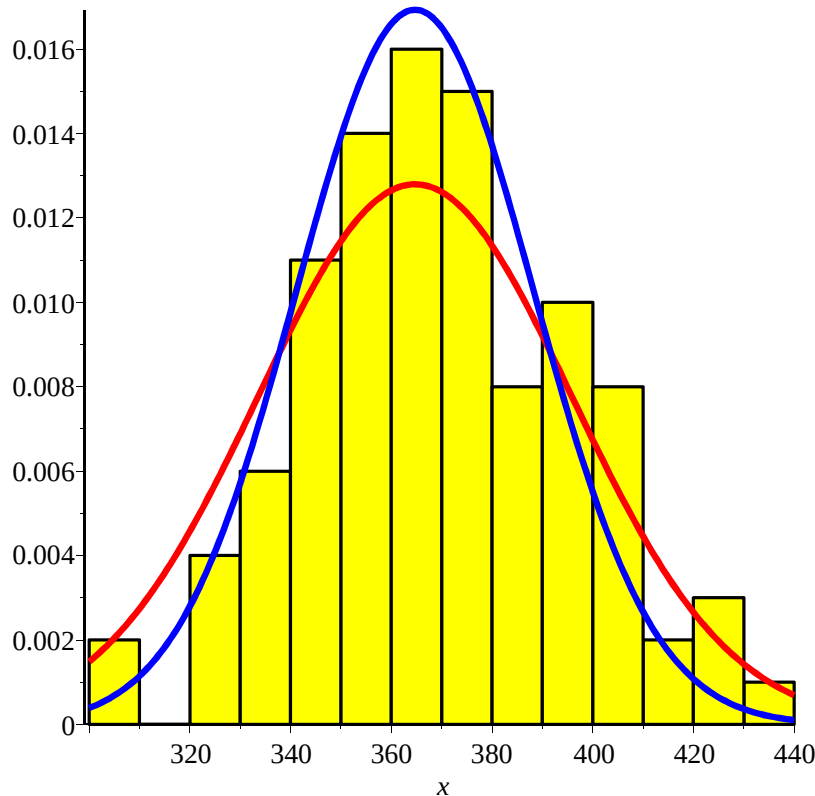
> Y1 := RandomVariable('Normal' (átlag,  $\sqrt{k1}$ )) :
Y2 := RandomVariable('Normal' (átlag,  $\sqrt{k2}$ )) :
> suruseg := plot( [PDF(Y1, x), PDF(Y2, x)], x = A..B, color = [red, blue], thickness = 3) :
suruseg

```



> plots[display]([HP, suruseg], title = Az adatok és a szórás konfidencia t  r  se)

Az adatok és a szórás konfidencia tűrése



Az ábra azt mutatja, hogy különböző 100 elemű minták vételezése során a hisztogramok
A Statistics csomag "**OneSampleChiSquareTest**" nevű eljárása az egymintás khi-négyzet tesztet futtatja a minta szórására.

```
> infolevel[Statistics] := 1 :
    teszt := OneSampleChiSquareTest(beton, szórás, confidence = 0.95)
```

Chi-Square Test on One Sample

Null Hypothesis:

Sample drawn from population with standard deviation equal to 26.8349

Alt. Hypothesis:

Sample drawn from population with standard deviation not equal to 26.8349

```
Sample size:          100
Sample standard dev.: 26.8349
Distribution:         ChiSquare(99)
Computed statistic:   99
Computed pvalue:      0.962194
Confidence interval:  23.5612136805739 .. 31.1734231484769
                     (population standard deviation)
```

Result: [Accepted]

There is no statistical evidence against the null hypothesis

teszt := hypothesis = true, confidenceinterval = 23.5612136805739 ..31.1734231484769, (1.3.6)

distribution = ChiSquare(99), pvalue = 0.962193824816528, statistic = 99.

A szórásra kapott konfidencia intervallum megegyezik a fenti intervallummal, mert mi a variancára számoltuk az intervallumot. Így, ha négyzetgyököt vonunk a fenti értékekből, akkor megkapjuk a OneSampleChiSquareTest által visszaadott konfidencia intervallumot.

> konfidencia_intervallum_szórásra = $\sqrt{k2} \dots \sqrt{k1}$

konfidencia_intervallum_szórásra = 23.5612136809135 ..31.1734231420530 (1.3.7)

▼ 15.4. Elméleti ellenőrző kérdések

15.4.1. Mit értünk egy paraméter γ megbízhatóságú becslésén?

15.4.2. Hogyan számoljuk ki egy n elemű normál eloszlású minta átlagának konfidencia intervallumát adott szinten?

15.4.3. Hogyan változik az átlag konfidencia intervallumának szélessége, ha növeljük a minta elemszámát?

15.4.4. Hogyan változik az átlag konfidencia intervallumának szélessége, ha növeljük a megbízhatósági szintet?

15.4.5. Hogyan változik az átlag konfidencia intervallumának szélessége, ha növekszik a populáció szórása?

15.4.6. Milyen eloszlást használunk az átlag becslésénél, ha ismert a szórás?

15.4.7. Milyen eloszlást használunk az átlag becslésénél, ha nem ismert a szórás?

15.4.8. Hogyan számolhatjuk ki a minta méretét, ha adott az átlag konfidencia intervallumának hossza és a szórás?

15.4.9. Milyen eloszlást használunk a szórás becsléséhez normál eloszlású populáció esetén?

▼ 15.5. Gyakorló feladatok

15.5.1. feladat.

Ha a normál eloszlás $\sigma=2$ szórása ismert és a minta mérete $n = 100$, akkor milyen széles a várható érték 99%-os konfidencia intervalluma? Mekkora legyen a minta n mérete, hogy a 99%-os konfidencia intervallum szélessége kisebb vagy egyenlő legyen, mint $L = 0.25$?

15.5.2. feladat.

Legyenek az $X_1, X_2, \dots$ azonos eloszlású független véletlen változók ismeretlen m várható értékkel és ismeretlen S szórással. Mérések alapján az $n = 100$ elemű minta átlaga

$$m_{100} = \frac{X_1 + X_2 + \dots + X_{100}}{100} = 10.083 \text{ és szórása } S_{100} = 0.568. \text{ Határozzuk meg a várható érték}$$

95%-os konfidencia intervallumát!

15.5.3. feladat.

Legyenek a $W_1, W_2, \dots$ azonos eloszlású független változók $N(0, \sigma)$ normál eloszlásúak ismeretlen σ szórással és legyen $X_i = m + W_i$, ahol m ismeretlen konstans. Mérések alapján az $n = 10$ elemű minta átlaga $m_{10} = 14.832$ és szórása $S_{10} = 1.904$. Határozzuk meg az m érték 95%-os konfidencia intervallumát!

15.5.4. feladat.

Legyenek $X_1, X_2, \dots$ azonos eloszlású független véletlen változók $N(0, \sigma)$ normál eloszlásúak ismeretlen σ szórással. Határozzuk meg a σ szórás 95%-os konfidencia intervallumát, ha $Var_{100} = 4.413$.

15.5.5. feladat.

Meg kell becsülnünk egy vegyipari üzemben naponta átlagosan készített kémiai anyag mennyiségét. Ezért $n = 100$ alkalommal rögzítjük a napi anyag mennyiségét, melynek átlagára $\bar{x} = 870$ (tonna) és standard szórására $s = 20$ (tonna) adódott.

- (i) Adjuk meg a napi mennyiség μ átlagára 95%-os konfidencia intervallumot!
- (ii) Adjuk meg a napi mennyiség μ átlagára 99%-os konfidencia intervallumot!
- (iii) Rajzoljuk fel az (i) és (ii) pontokban számolt intervallumokat és hasonlítsuk

össze!

(iv) Készítsünk Monte Carlo-szimulációt az adatokra! Interpretáljuk a konfidencia intervallumok jelentését!

(v) Számítsuk ki a minta n méretét ahhoz, hogy a konfidencia intervallum az (i) pontban kapott hossz fele legyen!

(vi) Számítsuk ki a minta n méretét ahhoz, hogy a konfidencia intervallum az (ii) pontban kapott hossz fele legyen!

15.5.6. feladat.

Egy szoftver cég tanulmányt készített a használt fájlok méretéről. Véletlenszerűen kiválasztott $n = 23$ fájl alapján az átlagos fájl méret $\bar{x} = 4822$ (kbyte) és a standard szórás $s = 127$ (kbyte). Határozzuk meg a fájl méretek átlagára 95%-os konfidencia intervallumot a mérések alapján!

15.5.6. feladat.

Melyik igaz melyik hamis az alábbi állítások közül? A választ indokoljuk!

- (i) Ha a populáció szórása növekszik és más jellemzők változatlanok maradnak, akkor az átlag konfidencia intervallumának szélessége is növekszik.
- (ii) Ha a minta mérete növekszik, akkor az átlag konfidencia intervallumának szélessége is növekszik.

15.5.7. feladat

Legyen egy púderes dobozba töltendő X mennyiség normál eloszlású $\mu_X = 100$ [gramm] várható értékkel, melynek standard szórása $\sigma_X = 1$ [gramm].

Az üres doboz Y tömege ugyancsak normál eloszlású véletlen változó $\mu_Y = 5$ [gramm] várható értékkel, melynek szórása $\sigma_Y = 0.5$ [gramm].

A tele dobozok $Z = X + Y$ tömegének ellenőrzése érdekében 20 elemű mintát vettek és mérés alapján a tömegek átlagára $Z_{\text{átlag}} = 106.5$ [gramm] adódott.

(a) Adjuk meg a 95%-os megbízhatósági szinthez tartozó konfidencia intervallumot Z átlagára, ha a szórás a fentiek szerint ismert!

(b) Vajon az átlagra kapott 95%-os megbízhatósági szintű konfidencia intervallumba beleesik-e a

$\mu_Z = 105$ [gramm]?

15.5.8. feladat

Az alábbi adatok fénycsővek működési élettartamát mutatja az elromlásig években.

Működési élettartamok

$M = [1.39, 1.07, 3.22, 3.67, 0.55, 0.81, 1.22, 1.26, 0.05, 1.54, 0.97, 1.01, 0.44, 1.97, 1.9, 0.89, 3.25, 0.85, 1.04, 0.43, 1.33, 0.82, 2.04, 1.02, 0.53, 0.13, 2.06, 2.96, 1.96, 1.5, 3.05, 0.42, 1.17, 1.72, 2.68, 0.56, 2.13, 1.56, 2.09, 1.26, 3.21, 0.74, 3.04, 2.74, 0.83, 0.79, 1.56, 1.55, 0.96, 1.23]$ (években)

(a) Számolja ki a működési élettartamok átlagát, szórását és terjedelmét?

(b) Rajzolja fel a működési élettartamok gyakoriság **hisztogramját** a terjedelem 15 egyenlő részre osztásával, és annak a **normál eloszlásnak** a sűrűség függvényét, melynek várható értéke és szórása megegyezik az adatok átlagával és standard szórásával!

(c) Vizsgálja meg, hogy a működési élettartamok **normál eloszlásúak-e** $\gamma=95\%$ -os megbízhatósági szinten a terjedelem $n=15$ részre osztása mellett!

(d) Vajon az $m=1.5$ éves élettartam benne van-e az átlag 95%-os **konfidencia intervallumában**?

(e) Az átlag értékre szimmetrikusan összesen $L=0.5$ év hosszúságú konfidencia intervallumot szeretnénk kapni 95%-os szinten. Minimálisan mennyi fénycső minta élettartamát szükséges ehhez megmérni? Ellenőrizzük a számítást véletlen adatok generálásával!

15.5.9. feladat

Egy üzem napi kén-oxid kibocsátását mérték tonnában. A mérések az alábbiak.

Napi kén-oxid kibocsátás tonnában:

Kén := [15.8, 26.4, 17.3, 11.2, 23.9, 24.8, 18.7, 13.9, 9.0, 13.2, 22.7, 9.8, 6.2, 14.7, 17.5, 26.1, 12.8, 28.6, 17.6, 23.7, 26.8, 22.7, 18.0, 20.5, 11.0, 20.9, 15.5, 19.4, 16.7, 10.7, 19.1, 15.2, 22.9, 26.6, 20.4, 21.4, 19.2, 21.6, 16.9, 19.0, 18.5, 23.0, 24.6, 20.1, 16.2, 18.0, 7.7, 13.5, 23.5, 14.5, 14.4, 29.6, 19.4, 17.0, 20.8, 24.3, 22.5, 24.6, 18.4, 18.1, 8.3, 21.9, 12.3, 22.3, 13.3, 11.8, 19.3, 20.0, 25.7, 31.8, 25.9, 10.5, 15.9, 27.5, 18.1, 17.9, 9.4, 24.1, 20.1, 28.5]
(tonna egységben)

(a) Számolja ki a kén-oxid kibocsátások **átlagát, szórását és terjedelmét**?

(b) Rajzolja fel a kén-oxid kibocsátások gyakoriság **hisztogramját** a terjedelem 20 egyenlő részre osztásával, és annak a **normál eloszlásnak** a sűrűség függvényét, melynek várható értéke és szórása megegyezik az adatok átlagával és standard szórásával!

(c) Vizsgáljuk meg, hogy a mért kén-oxid kibocsátások **normál eloszlásúak-e** $\gamma=95\%$ -os megbízhatósági szinten a terjedelem $n=20$ részre osztása mellett!

(d) Vajon a tervezett $m=15$ tonna napi kén-oxid kibocsátási érték benne van-e az átlag 95%-os **konfidencia intervallumában**?

(e) Az átlag értékre szimmetrikusan összesen $L=2$ tonna hosszú konfidencia intervallumot szeretnénk kapni 95%-os szinten. Ehhez minimálisan hány napon keresztül kell mérni a kén-oxid kibocsátást? Ellenőrizzük a számítást véletlen adatok **generálásával**!

15.5.10. feladat

Különböző fajta sajtok kalória mennyiségét mérték 1 uncia (kb. 28.3495 g) tömegre vonatkoztatva. A mérések az alábbiak.

Kalória mennyiségek

E:= [110, 45, 100, 95, 110, 110, 100, 110, 95, 120, 130, 100, 80, 105, 105, 90, 110, 70, 125, 108]
(cal egységben)

(a) Számolja ki a kalória mennyiségek **átlagát, szórását és terjedelmét**?

(b) Rajzolja fel a kalória mennyiségek gyakoriság **hisztogramját** a terjedelem 10 egyenlő részre osztásával, és annak a **normál eloszlásnak** a sűrűség függvényét, melynek várható értéke és szórása megegyezik az adatok átlagával és standard szórásával!

(c) Mutassuk meg, hogy a mért energia tartalmak **normál eloszlásúak-e** $\gamma=95\%$ -os megbízhatósági szinten a terjedelem $n=10$ részre osztása mellett!

(d) Vajon az $m=100$ cal energia érték benne van-e az átlag 95%-os **konfidencia intervallumában**?

(e) Az átlag értékre szimmetrikusan összesen $L=14$ cal hosszú konfidencia intervallumot szeretnénk kapni 95%-os szinten. Minimálisan mennyi sajt mintát szükséges ehhez megmérni? Ellenőrizzük a számítást adatok generálásával!

15.5.11. feladat

Különböző fajta üdítő italok szénhidrát mennyiségét grammokban mérték 12 uncia (1 uncia kb. 28.3495 g) tömegre vonatkoztatva. A mérések az alábbiak.

Szénhidrát mennyiségek

SH:= [48, 37, 52, 40, 43, 46, 41, 38, 41, 45, 45, 33, 35, 52, 45, 41] (gramm egységben)

(a) Számolja ki a szénhidrát mennyiségek **átlagát, szórását és terjedelmét**?

(b) Rajzolja fel a szénhidrát mennyiségek gyakoriság **hisztogramját** a terjedelem 8 egyenlő részre osztásával, és annak a **normál eloszlásnak** a sűrűség függvényét, melynek várható értéke és szórása megegyezik az adatok átlagával és standard szórásával!

(c) Mutassuk meg, hogy a mért szénhidrát tartalmak **normál eloszlásúak-e** $\gamma=95\%$ -os megbízhatósági szinten a terjedelem $n=8$ részre osztása mellett!

(d) Vajon az $m=40$ gramm energia érték benne van-e az átlag 95%-os **konfidencia intervallumában**?

(e) Az átlag értékre szimmetrikusan összesen $L=4$ gramm hosszú konfidencia intervallumot szeretnénk kapni 95%-os szinten. Minimálisan mennyi üdítőital **mintát** szükséges ehhez megmérni? Ellenőrizzük a számítást adatok **generálásával**!

15.5.12. feladat

Egy gép fém rudakat készít személyautók felfüggesztési rendszeréhez. Véletlenszerűen kiválasztottak 15 elkészített rudat és megmérték az átmérőket mm-ben. A mérések az alábbiak.

Átmérő értékek

$R := [8.24, 8.25, 8.20, 8.23, 8.24, 8.21, 8.26, 8.26, 8.20, 8.25, 8.23, 8.23, 8.19, 8.28, 8.24]$ (mm egységben)

(a) Számolja ki az átmérők **átlagát, szórását és terjedelmét**?

(b) Rajzolja fel az átmérők gyakoriság **hisztogramját** a terjedelem 8 egyenlő részre osztásával, és annak a **normál eloszlásnak** a sűrűség függvényét, melynek várható értéke és szórása megegyezik az adatok átlagával és standard szórásával!

(c) Mutassuk meg, hogy a mért átmérők **normál eloszlásúak-e** $\gamma=95\%$ -os megbízhatósági szinten a terjedelem $n=8$ részre osztása mellett!

(d) Vajon az $r=8.23$ mm átmérő benne van-e az átlag 95%-os **konfidencia intervallumában**?

(e) Az átlag értékre szimmetrikusan összesen $L=0.02$ mm hosszú konfidencia intervallumot szeretnénk kapni 95%-os szinten. Minimálisan mennyi rudat szükséges ehhez megmérni? Ellenőrizzük a számítást adatok generálásával!

(f) Az adatok hány százaléka felel meg a $8.2 \text{ mm} < R < 8.25 \text{ mm}$ minőségi követelménynek?

15.5.13. feladat

Egy tantárgy vizsgája során a tanulók által elért pontszámokat mutatja az alábbi adatsor: pontszámok: = [23,60,79,32,57,74,52,70,82,36, 80,77,81,95,41,65,92,85,55,76,

52,10,64,75,78,25,80, 98,81,67, 41,71,83,54,64,72,88,62,74,43,

60, 78,89,76,84, 48,84,90,15,79,34,67,17,82,69,74,63,80,85,61]

(a) Számolja ki a pontszámok átlagát, szórását, legkisebb- és legnagyobb értéket, a terjedelmet, az első és harmadik kvartilist!

(b) Rajzolja fel az adatsor váltakozását az átlaggal és az első és harmadik kvartilissel együtt!

(c) Rajzolja fel a pontszámok gyakoriság **hisztogramját** a terjedelem $n=15$ egyenlő részre osztásával!

(d) A leggyakrabban elért pontszám nagyobb vagy kisebb, mint az átlagos pontszám?

(e) Rajzolja be a gyakoriság hisztogramba annak a normál eloszlásnak a sűrűség függvényét, melynek várható értéke és szórása egyezik az adatok átlagával és szórásával!

(f) Vajon a pontszámok normál eloszlást követnek 95%-os megbízhatósággal, ha $n=15$ vagy $n=16$ részre osztjuk a terjedelmet?

(g) Határozzunk meg a pontszámok átlag értékére 95%-os megbízhatósággal konfidencia intervallumot! A tanár által feltételezett 70 pontos (közepes) teljesítmény benne van-e ebben a megbízhatósági intervallumban?

(h) Az átlag értékre szimmetrikusan $L=4$ sugarú intervallumot szeretnénk megkapni 95%-os konfidencia intervallumnak. Hány tanuló pontszáma szükséges ehhez? Ellenőrizzük a számítást adatok generálásával!

15.5.14. feladat

A Quality Engineering folyóirat (Vol. 4. 1992. 487-495 pp.) számában adatokat közöltek egy kémiai folyamatban keletkező anyag viszkozitási értékeiről. A minta a következő:

viszkozitas:= [13.30,14.3,14.9,15.2,15.8,14.2,16,14,

14.50,16.1,13.7,15.2,13.7,16.9,14.9,14.4,
 15.30,13.1,15.2,15.9,15.1,14.9,13.6,13.7,
 15.30,15.5,14.5,16.5,13.4,15.2,15.3,13.8,
 14.30,12.6,15.3,14.8,14.1,14.4,14.3,15.6,
 14.80,14.6,15.6,15.1,14.8,15.2,15.6,14.5,
 15.20,14.3,15.8,17,14.3,14.6,16.1,12.8,
 14.50,15.4,13.3,14.9,14.3,16.4,13.9,16.1,
 14.60,15.2,14.1,14.8,16.4,14.2,15.2,16.6,
 14.10,16.8,15.4,14,16.9,15.7,14.4,15.6]

(a) Számolja ki a viszkozitási értékek átlagát, szórását, legkisebb- és legnagyobb értéket, a terjedelmet, az első és harmadik kvartilist!

(b) Rajzolja fel az adatsor váltakozását az átlaggal és az első és harmadik kvartilissel együtt!

(c) Rajzolja fel a viszkozitási értékek gyakoriság hisztogramját a terjedelem $n = 15$ egyenlő részre osztásával!

(d) A leggyakrabban mért viszkozitás érték nagyobb vagy kisebb, mint az átlagos viszkozitás?

(e) Rajzolja be a gyakoriság hisztogramba annak a normál eloszlásnak a sűrűség függvényét, melynek várható értéke és szórása egyezik az adatok átlagával és szórásával!

(f) Vajon a viszkozitások normál eloszlást követnek 95%-os megbízhatósággal, ha $n = 15$ részre osztjuk a terjedelmet?

(g) Határozzunk meg a viszkozitások átlag értékére 95%-os megbízhatósággal konfidencia intervallumot! A feltételezett 15 viszkozitási érték benne van -e ebben a megbízhatósági intervallumban?

(h) Az átlag értékre szimmetrikusan $L = 0.2$ sugarú intervallumot szeretnénk kapni 95%-os konfidencia intervallumnak. Hány mérési adat szükséges ehhez? Ellenőrizzük a számítást adatok generálásával!

15.5.15. feladat

Az autógyártásban a felfüggesztéshez használt rúd átmérőjét automata gépsor gyártja.

Véletlenszerűen $n=15$ elemű mintát vettek a gyártásból és megmérték az átmérőket század milliméter pontosan.

Az alábbi adatok a mérési eredmények milliméterben:

átmérők:=[8.24 , 8.25 , 8.20 , 8.23, 8.24, 8.21, 8.26, 8.26, 8.20, 8.25, 8.23, 8.23, 8.19, 8.28, 8.24]

(a) Számoljuk ki az átmérők átlagát és szórását!

(b) Rajzoljuk fel az adatsor váltakozását az átlaggal együtt!

(c) Rajzoljuk fel az átmérők gyakoriság hisztogramját!

(d) Rajzoljuk be a gyakoriság hisztogramba annak a normál eloszlásnak a sűrűség függvényét, melynek várható értéke és szórása egyezik az adatok átlagával és szórásával!

(e) Vajon a gyártás normál eloszlást követ-e 95%-os megbízhatósággal? Teszteljük a hipotézist !

(f) Adjunk meg az átlag 95%-os konfidencia intervallumát!

(g) Megfelel-e az átmérőre a $\mu = 8.21$ [mm] előírás a mért átlagra?

15.5.16. feladat

30 darab 300-gramos kukoricapehelyben mérték a nátrium tartalmat . Az adatok milligrammban a következők:

natium:=[131.15, 130.69, 130.91, 129.54, 129.64, 128.77, 130.72,
 128.33, 128.24, 129.65, 130.14, 129.29, 128.71, 129.00, 129.39,
 130.42, 129.53, 130.12, 129.78, 130.92, 131.15, 130.69, 130.91,
 129.54, 129.64, 128.77, 130.72, 128.33, 128.24, 129.65]

(a) Számoljuk ki a gabonapehely nátrium tartalmának átlagát és szórását!

(b) Rajzoljuk fel az adatsor váltakozását az átlaggal együtt!

(c) Rajzoljuk fel a gabonapehely nátrium tartalmának gyakoriság hisztogramját!

(d) Rajzoljuk be a gyakoriság hisztogramba annak a normál eloszlásnak a sűrűség függvényét, melynek várható értéke és szórása egyezik az adatok átlagával és szórásával!

(e) Vajon a nátrium tartalom normál eloszlást követ-e 95%-os megbízhatósággal? Teszteljük a

hipotézist !

(f) Adjunk meg az átlag 95%-os konfidencia intervallumát!

(g) Megfelel-e a nátrium tartalomra a $\mu = 130$ [mgram] előírás a mért átlagra?

15.5.17. feladat

A "Nuclear Engineering" folyóirat 1988. februári szám 33. oldalán szereplő tudományos cikkben közzétettek egy Norvégiában működő atom reaktor néhány fűtő rúdjának jellemzőjét. Az alábbi mérési értékek 12 rúd dúsulásának százalékos értékét mutatják

dus:=[2.94, 3.00, 2.90, 2.75, 3.00, 2.95, 2.90, 2.75, 2.95, 2.82, 2.81, 3.05]

(a) Rajzoljuk fel a dúsulási százalékok hisztogramját! Feltételezhetjük-e, hogy a dúsulási értékek normál eloszlásúak 95%-os konfidencia szinten?

(b) Határozzunk meg a dúsulás átlag értékére 95%-os megbízhatósággal konfidencia intervallumot! A feltételezett 2.95%-os dúsulási érték benne van e ebben a megbízhatósági intervallumban?

(c) Az átlag érték $L=0.05$ [%] sugarú környezetét szeretnénk megkapni konfidencia intervallumnak. Hány elemű legyen a minta? Ellenőrizzük a számítást adatok generálásával!

15.5.18. feladat

Egy bizonyos típusú étkezési margarint vizsgáltak telítetlen zsírsavak százalékos szintjének megállapítása érdekében. A 6 elemű minta mérési adatai a következők:

zsírsav:= 16.8, 17.2, 17.4, 16.9, 16.5, 17.1

(a) Rajzoljuk fel a telítetlen zsírsavak százalékainak hisztogramját! Feltételezhetjük-e, hogy a kivett margarin minta telítetlen zsírsav értékei normál eloszlásúak 95%-os konfidencia szinten?

(b) Határozzunk meg a telítetlen zsírsav értékek átlagára 95%-os megbízhatósággal konfidencia intervallumot! Ráírhatják-e 95%-os megbízhatósággal a margarindra, hogy telítetlen zsírsav értéke 16.8%-os?

(c) Az átlag érték $L=0.2$ [%] sugarú környezetét szeretnénk megkapni konfidencia intervallumnak. Hány elemű legyen a minta? Ellenőrizzük a számítást adatok generálásával!