Leíró statisztika

A leíró statisztika (**descriptive statistics**) csupán azokat a statisztikai jellemzőket számítja ki, amelyekhez nem kell feltételeznünk semmi sem az adatok eloszlásáról. Csak egyetlen statisztikai változó adatait tartalmazza, ezért egyváltozós (**univariate**) statisztikának nevezzük.

Kiinduló adataink táblázata (a mintajegyzőkönyv adataival)

sorszám	mérőrúd	mérőszalag	ultrahangos távolságmérő
	cm	cm	cm
1	182	177	183
2	179	179	183
3	170	179	183
4	181	184	177
5	177	176	178
6	186	183	183
7	183	178	179
8	184	183	180
9	184	184	180
10	183	182	182

Adataink elnevezése eltérő lehet, részben azért, mert különféle statisztikai elemzéshez használjuk fel példaként, de azért is, mert a számítógépes programok eltérő néven használják őket. Egyes programok vektorként, mások mátrixként kezelik. Például a SCILAB, vagy MATLAB program számára a fenti adathalmaz mátrix, amely három oszlopból és tíz sorból áll. A pontosvessző választja el egymástól a sorokat ebben a példában. Egy példa erre (hőmérséklet- és nyomásértékek). Ez valójában értékadó utasítás, és a szögletes zárójel mátrixot jelent.

```
t=[ 88.85 90.85 92.85 94.85 96.85 98.85 100.85; 151530 163067 175321 188324
202112 216720 232200]
Ha lekérdezzük a mátrix tartalmát, látjuk, hogy hét oszlopból és két sorból áll:
t  =
   88.85     90.85     92.85     94.85     96.85     98.85     100.85
  151530.    163067.    175321.    188324.    202112.    216720.    232200.
(ez a program nem kezeli a tizedesvesszőt).
```

A mi példánk adatait beírva ezt látjuk (ezúttal a mátrix neve **a1**):

```
a1  =
   182.   177.   183.
   179.   179.   183.
   170.   179.   183.
   181.   184.   177.
   177.   176.   178.
   186.   183.   183.
   183.   178.   179.
   184.   183.   180.
   184.   184.   180.
   183.   182.   182.
```

Le is kérdezhetjük, hány sorból és oszlopból áll a mátrixunk („row”, „column”):
nr = size(a1, "r")
nc = size(a1, "c")

Az egyszerűség kedvéért itt csupán a mérési eredményeink második oszlopában szereplő számokkal mutatunk be módszereket (mérőszalagos hosszúságmérések cm mértékegységben). $\sum_{i=1}^{n} x_i$

Az adatsokaság összege EXCELben, oszloponként:
=SZUM(C3:C12)=1809, =SZUM(D3:D12)=1805 és =SZUM(E3:E12)=1808
A SCILAB esetén az oszloponkénti összeg:
sum (a1, "r")
ans =
1809. 1805. 1808.
Ellenőrizzük le a második oszlopra! 177+179+179+184+176+183+178+183+184+182= 1805
Az oszlopokat azért különítettük el, mert tudjuk, hogy valamilyen okból eltérnek egymástól. Három oszlopunk van, ezeket csoportoknak, osztályoknak, vagy kezeléseknek nevezzük (**group, class, treatment**).
Az alábbiak jobb érthetősége kedvéért létrehozunk egy SCILAB változót msz („mérőszalag”) néven.
msz = a1(:,2) a kettőspont bármelyik értéket jelenti, a 2 számjegy a második oszlopot választja ki.
A statisztikai változó összege: sum(msz)=1805

Értékkészlet

Az adatsokaságnak vannak kitüntetett értékei, amelyek jellemzőek lehetnek rájuk.

A minimum és a maximum

EXCELben: =MIN(D3:D12)=176 és =MAX(D3:D12)=184

SCILAB esetén: min(msz)=176 , max(msz)=184

A medián

A medián a nagyság szerint sorba rendezett adatok *középső* értéke. Ha páros számú adatunk van, akkor a medián a két középső érték számtani középarányosa.

Lehetőségünk van az adatokat nagyság szerint sorbarendezni: msz1 = gsort(msz,'lr','i') (increasing)

Adataink: 177 179 179 184 176 183 178 183 184 182

sorba rendezve: 176 177 178 179 179 182 183 183 184 184 (rendezett minta)

=MEDIÁN(D3:D12)=180,5

SCILAB: median(msz)=180,5

a két középső adat: 179 és 183, átlaguk: (179+182)/2=180,5 *véletlenül* megegyezik az adatok átlagával.

Módusz

A módusz a legnagyobb gyakorisággal előforduló szám.

=MÓDUSZ(D3:D12)=179

Esetünkben három olyan adat is van, amelynek legnagyobb a gyakorisága, ugyanis a 179, a 183 és a 184 egyaránt kétszer fordul elő. A program ezek közül a nagyság szerint sorba rendezett adatok közül az első előfordulást választotta.

Terjedelem

A terjedelem a legnagyobb és a legkisebb szám különbsége (**range**).

EXCELben MAX(D3:D12)-MIN(D3:D12)= 8, illetve. Az excelben a *tartomány* szó jelzi.

SCILAB: strange(msz)=8

Kvartilis

Az adatokat nagyság szerint sorba rendezzük, majd részekre bontjuk. Ha négy részre bontottuk, akkor a részeket kvartilisnak nevezzük. Ehhez kiválasztjuk a mediánt. A mediánnál kisebb és nagyobb adatok gyakorisága azonos. Ha ez megtörtént, akkor a kisebb értékeket ismét két részre bontjuk, a nagyobb részeket éppúgy két részre bontjuk. Az első lépésben a következő adatsorhoz jutottunk:

176 177 178 179 179 < 180,5

182 183 183 184 184 > 180,5

Második lépésként ezt is megfelezzük:

=KVARTILIS(D3:D12;0)= 176 a legkisebb érték

=KVARTILIS(D3:D12;1)=178,25 a kisebb csoport középértéke

=KVARTILIS(D3:D12;2)= 180,5 a medián

=KVARTILIS(D3:D12;3)= 183 a nagyobb csoport középértéke

=KVARTILIS(D3:D12;4)= 184 a legnagyobb érték

Az első kvartilis [176 177] < 178,25

A második kvartilis: 178,25 < [178 179 179] < 180,5

A harmadik kvartilis: 180,5 < [182 183 183] <= 183 (egész értékek számításánál)

A negyedik kvartlis: 183 < [184 184]

A kvartilisek közti értékek (**interquartile range**) az első és a harmadik kvartilis közötti tartomány. Esetünkben a 178,25 és 183 közé eső értékek száma: IR=6 (felsorolva: 178 179 179 182 183 183), az IR terjedelem nagysága: 183-178,25 = 4,75

SCILAB: quart(msz) kilistázza mindhárom kvartilist (a definíció szerint egész számok formájában).

Átlag

Az átlag jellegű mennyiségek mértékegysége azonos a mért mennyiség mértékegységével. Példánk esetében centiméter. Jele igen gyakran M az angol **mean** kifejezés alapján.
Nézzük ezek után az egyes mérések átlagát! Tíz mérésünk van, tehát zsebkalkulátoron:
(177+179+179+184+176+183+178+183+184+182)/10= 180,5
Excelben: =ÁTLAG(D3:D12)=180,5

SCILABben (angolul a *számtani középarányos* az average, de szakszerűbb a **mean** kifejezés):
mean(a1, "r")
ans =
180.9 180.5 180.8
Amennyiben szóráselemzést végzünk, szükségünk lehet az egész adathalmaz összegére is. EXCEL: =SZUM(C3:E12)=5422 (elnevezése: **grand total**)
SCILAB változata:
sum (a1)
ans =
5422.
A metrológiában az átlag a *várható érték* (**expected value**) egyik legjobb becslése.

Mértani átlag

A mértani (geometriai) átlag az elemek szorzatának annyiadik gyöke, amennyi az elemek száma

$$M_G=\sqrt[n]{\prod_{i=1}^{n} x_i}$$

Az EXCELben: =MÉRTANI.KÖZÉP(D3:D12)=180,4771

SCILAB: geomean(msz)=180,4771

Harmonikus középarányos

Képlete (repiroka) $\frac{1}{H_i}=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}$ (a repirok értékek összegének reciprok értéke)
Az EXCELben: =HARM.KÖZÉP(D3:D12)= 185,4542
SCILAB: harmean(msz)=185,4542

Az átlagtól való eltérések

Erre csak ideiglenesen van szükségünk, a további számítások elvégzése céljából. Mértékegységük azonos a mért mennyiség mértékegységével; példánkban centiméter.
Ha a *metrológiában* a mért értékből kivonjuk az átlagot, akkor a **hiba** értékét kapjuk.
Néhány példa erre: 177-180,5= -3,5 (az adat „lefelé” tér el; a hiba negatív)
182-180,5= 1,5 (az adat „felfelé” tér el; a hiba pozitív)
Ugyanez EXCELben az első és az utolsó sorban szereplő adat példájával:
=D3-ÁTLAG(D3:D12)= -3,5
=D12-ÁTLAG(D3:D12)= 1,5
SCILABben a második oszlop első sorában álló adatból kivonjuk az átlagot:
a1(1, 2) -180.5
ans =
- 3.5
vagy az utolsó sor estén:
a1(10, 2) -180.5
ans =
1.5
tehát ugyanazt kaptuk.
SCILAB: a hiba értékek listáját kapjuk a következőképp: center(msz)

Az átlagtól való eltérések előjeles összege

Ennek az adatnak nullát kell adnia, hiszen épp ennek érdekében számítottuk az átlagot.

(177-180,5)+(179-180,5)+(179-180,5)+(184-180,5)+(176-180,5)+(183-180,5)+(178-180,5)+(183-180,5)+(184-180,5)+(182-180,5)=0
SCILAB: sum(msz-mean(msz))=0 vagy:
sum(center(msz))=0
Ha nem így van, akkor valamelyik adatot eltévesztettük. Nem a számítógép téved! Még a legegyszerűbb zsebkalkulátor is legalább tíz tizedesjegyig képes számolni. A leggyakoribb hiba ott történik, amikor kerekítve visszük be valamelyik mérési eredmény adatát – esetleg, ha az adatok törtrészének számjegyei (nagyságrendje) váltakozó értékű.

Az átlagtól való abszolút eltérések összege

$$S_{ab}=\sum_{i=1}^{n}|x_i-\overline{x}|$$

(|177-180,5|)+(|179-180,5|)+(|179-180,5|)+(|184-180,5|)+(|176-180,5|)+(|183-180,5|)+(|178-180,5|)+(|183-180,5|)+(|184-180,5|)+(|182-180,5|)=27
SCILAB: sum(abs(msz-mean(msz)))=27

Az átlagtól való eltérések négyzetének összege

Ez az adatok eloszlásának egyik legfontosabb mutatója, ez a *négyzetöszeg* **(Sum of Squares)**.

$$SS=\sum_{i=1}^{n}(x_i-\overline{x})^2$$

$(177-180{,}5)^2+(179-180{,}5)^2+(179-180{,}5)^2+(184-180{,}5)^2+(176-180{,}5)^2+(183-180{,}5)^2+(178-180{,}5)^2+(183-180{,}5)^2+(184-180{,}5)^2+(182-180{,}5)^2=82{,}5$
SCILAB: először képezzük a négyzetes eltérések vektorát: negyz=(msz-mean(msz))² (ez, mint a kiinduló adat, tíz értéket tartalmaz)
Ennek ismeretében számítjuk a négyzetes eltérések összegét: sum(negyz)=82,5

Momentum

A momentum kifejezést a középérték másfajta megfogalmazására használjuk. Nézzük egymás után az összeg, a momentum, majd a centrális momentum képletét!

$$SZUM=\sum_{i=1}^{n}(x_i)$$

$$M_1=\frac{1}{n}\sum_{i=1}^{n}x_i \qquad \text{(ez az átlag)}$$

$$CM=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x}) \qquad \text{(ez az eltérések átlaga)}$$

A momentum számításának jelentőségét az adja, hogy a függvény argumentumának hatványozásával új, fontos információra tehetünk szert.

$$M_r=\frac{1}{n}\sum_{i=1}^{n}x_i^r$$

$$CM_r=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^r$$

ahol r pozitív egész szám. Az egyszerű momentum esetében kizárólag az r=1 esetét használjuk; ez az átlag (számtani középarányos). A centrális momentum r=2 esetén azonos a variancia (szórásnégyzet) értékével. r = 3 és r = 4 használható az eloszlás ferdeségének (**skewness**) és csúcsosságának (**kurtosis**) becslésére. Hangsúlyozzuk: ehhez nem kell ismerni az eloszlás típusát.
Scilab:
moment(msz,1) = 180,5 (az átlag)
moment(msz,2= 32588,5
moment(msz,3)= 5885198,9 (ezeknek már nincs jelentősége)

cmoment(msz,1) = 0 (az előjeles értékek összege nulla)
cmoment(msz,2) = 8,25 (ez a korrigálatlan variancia)
cmoment(msz,3) = -3,6
cmoment(msz,4) = 99,2625
EXCEL: a momentum és a centrális momentum függvényként közvetlenül nem érhető el
=FERDESÉG(D3:D12) = -0,18016 (negatív érték: balra ferdült eloszlás)
=CSÚCSOSSÁG(D3:D12) = -1,76104 (negatív érték: kissé lapult eloszlás)
SCILAB:
skewness(msz)
- 0,1297135 (negatív érték: balra ferdült eloszlás)
kurtosis(msz)
- 1,8186942 (negatív érték: kissé lapult eloszlás)
A két számítási mód eltér. Magyarázatát ezúttal mellőzzük, mert a fizika laborgyakorlatokon ezeket az értékeket nem kell számolni.

A mérési bizonytalanság becslése

A mérési bizonytalanság az adatsokaság szóródási tulajdonsága alapján becsülhető. Lépésről-lépésre ez a következőképpen történik.
Számítjuk a négyzetes eltéréseket
Számítjuk a négyzetes eltérések összességét.
Számítjuk ennek középértékét.
Számítjuk a varianciát.
A variancia négyzetgyöke a szórás.
A szórás segítségével számítjuk az átlag szórását.
EXCEL:
Ha az adatok a D3:D12 cellákban vannak, akkor az átlag a D13 cellába kerül
=ÁTLAG(D3:D12= 180,5
Az első eltérés: =(D3-D13)^ 2 = 12,25 (D14 cella)
Az utolsó eltérés: =(D12-D13)^ 2 = 2,25 (D23 cella)
Összegük: =SZUM(D14:D23)= 82,5 (D24 cella)
Átlaguk: =D24/DARAB(D3:D12) = 8,25 (tíz darab mérési adatunk volt)
A szabadsági fokkal korrigált érték (D.F.=10-1= 9)
=D24/9= 9,16666 (a variancia a D27 cellába kerül))
=GYÖK(D27)= 3,02765 (a korrigált empirikus szórás)
Ugyanerre az eredményre jutunk a következőképp =GYÖK(VAR(D3:D12))= 3,02765
Az átlag szórása =GYÖK(D27/10)= 0,957427
Magyarázat: az adatösszesség szórása 3,027 valamennyi adat mérési bizonytalanságának becslésére használható, míg az átlag ennél sokkal megbízhatóbban becsülhető, hiszen tíz adatból számítottuk= 0,957427 – más szóval, az adatok (egyfajta becslés szerint) a 180,5±3,027 tartományban lehetnek; az átlaguk viszont a 180,5±0,957427 tartományban.
SCILAB megoldása ugyanennek:
(msz(1)-mean(msz))^2 = 12,25
(msz(10)-mean(msz))^2 = 2,25
negyz=(msz-mean(msz))^ 2 (tíz adatot látunk...)
sum(negyz) = 82,5
negyzatl=sum(negyz)/10 = 8,25

varia=sum(negyz)/9
varia = 9,1666667
variance(msz) = 9,1666667
sqrt(varia=sum(negyz)/9) = 3,0276504
és végül az átlag szórása:

sqrt(variance(msz)/10) = 0,9574271

Merjünk-e jóslásokba bocsátkozni?

Nem. Ez nem a jóslás tudománya. Az általunk megismert adatok tapasztalati értékek. Csak annyit mondanak, amennyit magunk is megismertünk. Ám, ha a méréseinket állandó hiba nem terheli, lehetőségünk van *előre jelezni* az eddig el nem végzett mérések eredményét. A fenti adatok példájával mondhatjuk, hogy – ha végeznénk egy *tizenegyedik* mérést – az valószínűleg (de nem biztosan) az előre jelzett tartományba fog esni, tehát 180,5-3,027=177,472 és 180,5+3,027=183,527 közé. (Tessék utánanézni! Már a mérési eredmények között is van olyan, amely kívül esik ezen az intervallumon.)
Ha végzünk egy *tizenegyedik* mérést, az természetesen meg fogja változtatni az átlagot is. De, ha a fenti feltételek teljesülnek, akkor az új átlag valószínűleg nem lesz kisebb, mint 180,5-0,957427=179,543 és nem lesz nagyobb, mint 180,5+ 0,957427=181,457.
A szóban forgó labormérés folytonos értékkészletű változókból származó, de csak egészek formájában ismert adatokat tartalmaz. Márpedig a folytonos változók legkisebb értéke a mínusz végtelen, legnagyobb értéke a plusz végtelen. Várható-e az eddig mért legkisebb értékél, 176-nál kisebb érték? Természetesen igen. Minél kisebb értékre számítunk, annak az előfordulási gyakorisága (és lehetősége) egyre kisebb, és tart a nullához. De azt, hogy ezt hogyan, milyen függvény szerint teszi, sajnos, ennyire kevés adat alapján nem tudhatjuk. Bár a szakirodalom ezt nem javasolja, megkísérelhetjük az értékeket a normál eloszlás feltételezésével vizsgálni.
Ha a szóban forgó tíz adatra megkísérelnénk többféle eloszlást illeszteni, a vizsgálati eredmények közötti hiba annyira kicsiny lenne, hogy nem tudnánk egyértelműen eldönteni, hogy jobb-e a normáleloszlás, mint a lognormál eloszlás, az exponenciális eloszlás, a Student-eloszlás eloszlás, vagy az F-eloszlás, vagy bármelyik másik.
Mindezt azért említettük, mert az EXCEL program a leíró statisztika eredményei között tartalmazza a konfidencia intervallumot is. Ennek egy lehetséges értelmezése a következő.
u kiterjesztési tényezőnél 1-ε valószínűséggel a szórás értéke számítható. $u\frac{\sigma}{\sqrt{n}}$ (angolul coverage factor, az **u** betű pedig az ucertainty kifejezés rövidítése)
E függvény értéke és a (kétoldali) valószínűség közötti kapcsolatot általában a normáleloszlás alapján határozzák meg, kétszeres kiterjesztési tényezővel. Az EXCEL magyar változatában MEGBÍZHATÓSÁG a neve, ugyanis ennek alapján lehet a metrológiai értelemben vett mérési bizonytalanság értékét becsülni.
=MEGBÍZHATÓSÁG.NORM(0,05;D28;10)
Paraméterek sorrendben: az alfa (fent epszilon jelzi), a szórás, végül a mérések darabszáma. Hivatalosan 0,05 értéket kell beírni, ami 1-0,05=0,95; tehát 95%-os valószínűségi szintet jelent. E fejezet végén bemutatjuk az EXCEL által készített eredmények táblázatát, ott azonban a 0,0236877 alfa értéket kell megadnunk ahhoz, hogy a jelzett eredményt kaphassuk.
=MEGBÍZHATÓSÁG.NORM(0,0236877;D28;10) = 2,16585059 (mértékegysége azonos a mért mennyiség mértékegységével, aktuálisan centiméter, ez a szórás mértékegységéből következik).
SCILAB megoldás (konfidencia intervallum 0,025 kétoldali szinten):
[h,p, z, ci] = hypt_ztest (msz,180.5,3.0276504,0.025)
h = F
ci =   178,35402
         182,64598
(ci(2)-ci(1))/2
 ans = 2,1459798
A **z** eloszlás a standardizált nomáleloszlás, amelynél a várható érték egy és a szórás is egy.
A **h = F** (false) elutasítja a nullhipotézist, tehát az adatok nagy része kívül esik a konfidencia intervallumon. A **ci** (confidence interval) kételemű vektor, amelynek (mint fent látható) ki is lehet

számítani a terjedelmét. A hypt_ztest paraméterei sorrendben: az adatokat tartalmazó vektor; a várható érték; a szórás, valamint az alfa (a megbízhatósági szint).
Választható egy másik függvény is ugyanerre a célra
[inter,mu]=intnormalm(msz,level=0.95)
mu =
180,5
inter =
178,33415
182,66585
(182,66585-178,33415)/2 = 2,16585
Itt a **mu** (az egyenlőségjel baloldalán) a várható érték jele. A függvény nevében a *normal* az eloszlást jelenti, az **m** pedig a várható értékre utal (görög μ). A függvény kimenete a kért valószínűségi szinten (most 0,95) megadja a konfidencia intervallum alsó és felső határát. 95%-os szinten csaknem valamennyi adat az eredményül kapott intervallum belsejében van.
Az intnormal**m** függvény párja az intnormal**s**, amely az eloszlás tulajdonságait a *szórásra* vonatkoztatva határozza meg (az **s** betű a görög szigma, σ betűre utal):
[inter,mu]=intnormals(msz,level=0.95)
mu =
3,0276504
inter =
2,0825245
5,5273093
Ez az eredmény azt jelenti, hogy a tíz adatból kapott szórásnak 5,527 és 2,0825 közé kell esnie.

A fenti példában említett adatokra az EXCEL az alábbi kimenetet képezi. Kiszámítása a következőképpen történik. Először is be kell tölteni az A*nalysis ToolPak*-et. Ekkor a menüsoron megjelenik az adatok menüpont alatt az *adatelemzés* felirat. Rákattintva válasszuk a leíró statisztika menüt. Bemeneti tartományként érdemes egy sorral többet kell kijelölni, mint ahol az adataink találhatóak, ugyanis az első sort címsorként használja az EXCEL; pl. (D1:D12) Ehhez bejelöljük a *feliratok az első sorban* melletti négyszöget.

mérőszalag
177
179
179
184
176
183
178
183
184
182

mérőszalag	
Várható érték	180,5
Standard hiba	0,957427108
Medián	180,5
Módusz	179
Szórás	3,027650354
Minta varianciája	9,166666667
Csúcsosság	-1,76103896
Ferdeség	-0,18015771
Tartomány	8
Minimum	176
Maximum	184
Összeg	1805
Darabszám	10
Konfidenciaszint(95,0%)	2,16585059

Emlékeztetőül: a standard hiba az átlag szórása; a szórás és a variancia a szabadsági fokkal korrigált értéket tartalmazza; a tartomány másik neve: terjedelem. Itt az adatok mértékegysége centiméter; a varianciáé centiméter négyzet; a csúcsosság és a ferdeség egy-mértékegységű mennyiségek.

mérések centiméterben	mérőszalag		
adatok	177		
	179		
	179		
	184		
	176		
	183		
	178		
	183		
	184		
	182		
átlag	180,5		
eltérések négyzete	12,25		
	2,25		
	2,25		
	12,25		
	20,25		
	6,25		
	6,25		
	6,25		
	12,25		
	2,25		
eltérések négyzetösszege	82,5		
eltérések négyzetösszegének átlaga	8,25		
(empirikus szórásnégyzet)			
korrigált empirikus szórásnégyzet	9,166667		
korrigált empirikus szórás	3,02765	177,4723	183,5277
átlag szórása	0,957427	179,5426	181,4574
mérési bizonytalanság	2,165851		

mérések centiméterben	mérőszalag	
adatok	177	
	179	
	179	
	184	
	176	
	183	
	178	
	183	
	184	
	182	
átlag	=ÁTLAG(D3:D12)	
eltérések négyzete	=(D3-D13)^2	
	=(D4-D13)^2	
	=(D5-D13)^2	
	=(D6-D13)^2	
	=(D7-D13)^2	
	=(D8-D13)^2	
	=(D9-D13)^2	
	=(D10-D13)^2	
	=(D11-D13)^2	
	=(D12-D13)^2	
eltérések négyzetösszege	=SZUM(D14:D23)	
eltérések négyzetösszegének átlaga	=D24/DARAB(D3:D12)	
(empirikus szórásnégyzet)		
korrigált empirikus szórásnégyzet	=D24/9	
korrigált empirikus szórás	=GYÖK(D27)	=D13-D28
átlag szórása	=GYÖK(D27/10)	=D13-D29
mérési bizonytalanság	=MEGBÍZHATÓSÁG.NORM(0,0236877;D28;10	

Baloldalt az ecxcel szokványos kimenete látható, számadatokkal. Jobboldalt számadatok helyett a kiszámításukhoz szükséges képletek. Zöld színnel (felül) az *értékekre* vonatkozó konfidencia intervallum, alul az *átlagra* vonatkoztatott konfidencia intervallum, azonos valószínűségi szintre.
Nem szabad elfelejtenünk, hogy nem vizsgáltuk, hogy az adataink normál eloszlásúak-e, ezért a konfidencia intervallumhoz valószínűségi szint nem rendelhető hozzá. Vagyis: nem tudjuk a valószínűségi szintet; csupán azt tudjuk, hogy *azonos* feltételezésekkel éltünk.