

# Az elméleti szórás körvizsgálati precizitási kísérletekben használandó robusztus becslése

*Wilrich, P-Th*

Berlini Műszaki Egyetem

Érkezett: 2004.október 23.

## 1. Bevezetés

Az ISO 5725-2:1994 [4] szabvány olyan laboratóriumi körvizsgálatokkal foglalkozik, melyeket azért végeznek, hogy statisztikai úton megbecsüljék a mérés szórásának két összetevőjét: az ismételhetőség körülményei között a laboratóriumon belüli, illetve a laboratóriumok közötti komponensét. Ez a módszer olyan mérési eredményeket feltételez, „melyek közelítőleg normál eloszlásúak” (ISO 5725-2, 1.4). Az elméleti szórás,  $\sigma$  becslésére a minta „ $s$ ” szórását használják.

Előfordulhat, hogy a vizsgált  $x$  jellemző normál eloszlásának feltételezése kétséges, vagy tudjuk, hogy az nem normál eloszlású. Mindkét esetben felmerül a kérdés, hogy ha az eloszlás nem normál, vajon az „ $s$ ” tapasztalati szórás a  $\sigma$  elméleti szórásnak robusztus becslése-e, azaz még mindig jó becslése  $\sigma$ -nak, vagy nem az. Ha nem elég jó a becslés, akkor felmerül a kérdés, lehet-e az adott helyzetben jobb becslést találni.

A 2. fejezetben a becslés pontosságának két mértékét vezetjük be (torzítás és precizitás), a 3. fejezetben a normál eloszlástól való eltérés két különböző helyzetét különböztetjük meg. A 4. fejezetben a normál eloszlástól való eltérés esetén  $\sigma$  becsléseivel, azok pontosságával és robusztusságával foglalkozunk. Az 5. fejezetben bemutatjuk az ISO 5725-2 megközelítését a mérési módszer precizitásának meghatározására, a 6. fejezetben ennek alternatíváját mutatjuk be az „ $s$ ” szórás robusztus alternatívájaként a Rousseeuw féle  $Q_n$  alkalmazásával.

## 2. A becslések pontossága

Vegyük azt a helyzetet, hogy egy népesség összetevőinek „ $x$ ” jellemzője érdekel minket, és ennek a jellemzőnek az eloszlását szeretnénk leírni egy vagy több paraméterrel.

Példák:

- A 2004-ben Berlinben élő (népesség) férfiak (egyedek) tömege (jellemző): érdekel minket az „átlagtömeg”, és a „tömeg szórása,  $\sigma$ ”
- 50 adag vaj (darab) víztartalma (jellemző) egy tételben vagy szállítmányban; szeretnénk tudni „ $\mu$  átlagos víztartalmat” és „a víztartalom  $\sigma$  szórását”

Hogy az általában nem ismert  $\mu$  és  $\sigma$  paraméterek becslésére statisztikai módszereket alkalmazhassunk, leggyakrabban azt feltételezzük, hogy az  $x$  jellemző eloszlása a populációban normál (Gauss eloszlás).

A minta átlagát

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

használjuk a  $\mu$  becslésére, a minta szórását

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

pedig a  $\sigma$  becslésére, és a minta szórásnégyzetét  $s^2$ -et az elméleti variancia  $\sigma^2$  becslésére.

Hogy írhatjuk le ezen a becslések pontosságát? Két egyszerű megfontolásból indulunk ki:

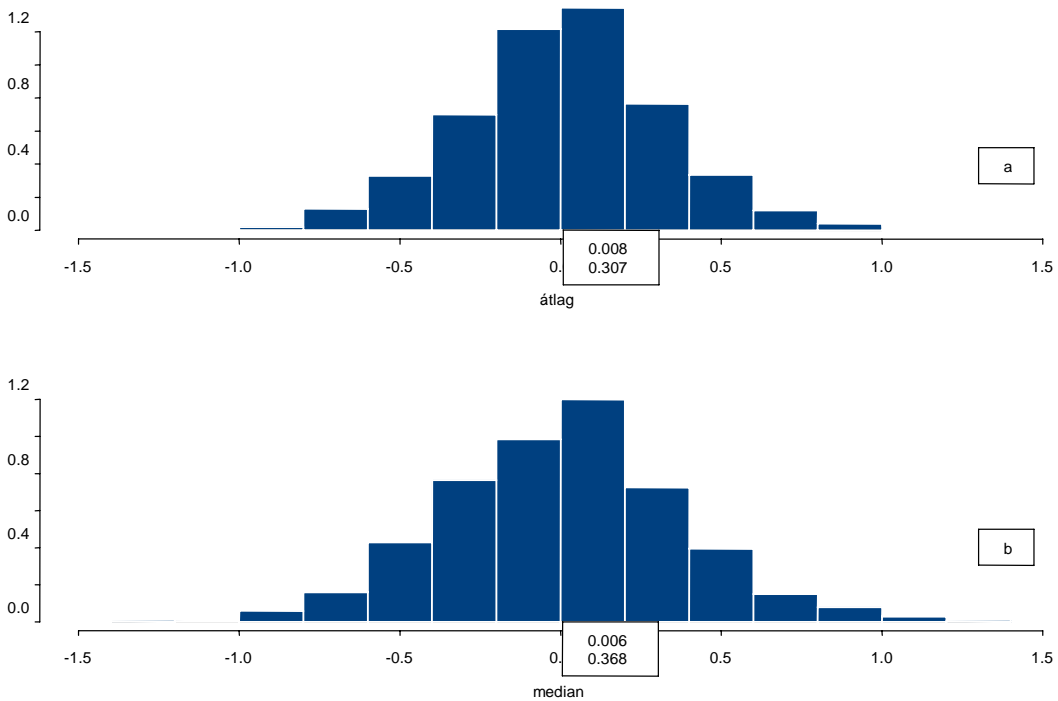
- a) Amikor populációból egy  $n$  méretű mintát kiválasztunk,  $\bar{x}$ -re (és  $s$ -re) egy partikuláris értéket kapunk. Ha egy másik ( $n$  méretű) mintát választunk ki, legnagyobb valószínűséggel más  $\bar{x}$  (és más  $s$ ) értéket fogunk kapni és így tovább: a  $\mu$   $\bar{x}$  becslései ugyanabból a populációból ismételt mintavétel mellett eloszlást mutatnak.  $\bar{x}$ -nek ez az eloszlása jellemző alakkal és paraméterekkel rendelkezik:  $\bar{x}$  átlag értéke,  $\bar{x}$  szórása és más értékek.  $\bar{x}$  eloszlásának a típusa és paraméterei nem csupán az  $n$  mintamérettől függenek, hanem az  $\bar{x}$  jellemző eloszlásának típusától is. Így ha le akarjuk írni  $\bar{x}$  mint  $\mu$  becslésének pontosságát, nem elég, ha csak  $\bar{x}$  egy értékét nézzük és hasonlítjuk össze  $\mu$ -vel, mert tudjuk, hogy  $\bar{x}$  következő értéke különböző lesz. Ehelyett meg kell vizsgálnunk  $\bar{x}$  eloszlását, különösen pedig annak paramétereit. Ugyanez a helyzet az  $s$  értékkel, mint  $\sigma$  becslésével.
- b)  $\bar{x}$  helyett használhatnánk a mediánt, azaz a legkisebbtől a legnagyobbig elrendezett minta középső értékét.

$$\tilde{x} = \begin{cases} x_{\binom{n+1}{2}} & n = 3, 5, 7, \dots \\ \frac{1}{2} \left( x_{\binom{n}{2}} + x_{\binom{n}{2}+1} \right) & n = 2, 4, 6, \dots \end{cases}$$

ahol  $x_{(i)}$  jelöli az  $i$ -edik legkisebb minta értéket, mint  $\mu$  becslését. Ugyanabból a populációból történő ismételt mintavételnél  $\tilde{x}$ -re különböző értékeket kaphatunk,  $\tilde{x}$ -nek eloszlása van, amely  $\bar{x}$  eloszlásához hasonlóan, függ az  $n$  mintaszámtól és  $x$  eloszlásától. Mivel  $\tilde{x}$  általában különbözik  $\bar{x}$ -től,  $\tilde{x}$  eloszlása is különbözik  $\bar{x}$  eloszlásától. Így ha össze akarjuk hasonlítani  $\tilde{x}$  és  $\bar{x}$  pontosságát, mint ugyanannak a  $\mu$ -nek a becsléseit, meg kell vizsgálnunk azok eloszlását, közelebbről az eloszlások paramétereit.

## 1. sz. szimulációs kísérlet:

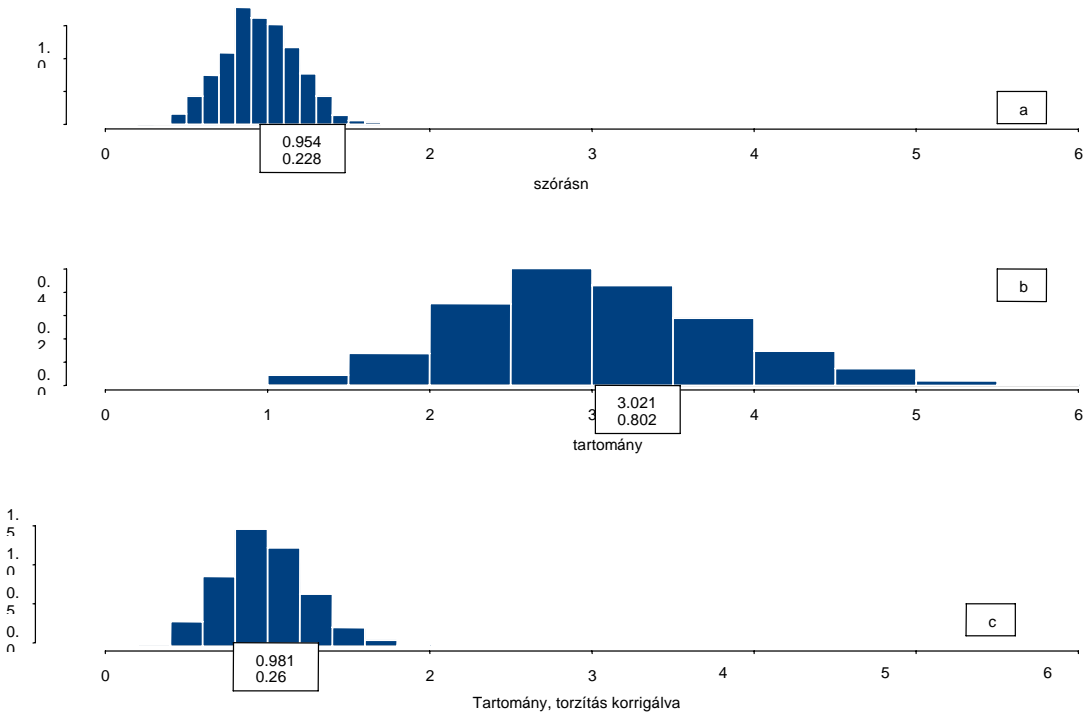
Használjuk az S-PLUS statisztikai programcsomagot a szimulációs kísérlethez.  $N=1000$  mintából  $n=10$  méretű mintát veszünk egy standardizált normál eloszlásból (az átlag  $\mu=0$ , a szórás  $\sigma=1$ ) és minden mintára kiszámítjuk  $\bar{x}$ -et és  $\tilde{x}$ -et (1. ábra). Egy-egy gyakorisági eloszlást kapunk  $x$ -re és  $\tilde{x}$ -re, ezek láthatók az alábbi hisztogrammokon. Ezen kívül kiszámítjuk  $\bar{x}$  és  $\tilde{x}$  átlagát és szórását, amit a kis négyzetekben tüntettünk fel a hisztogramokon. Azonnal láthatjuk, hogy mindkét becslésnek van egy átlagértéke, amely körülbelül egyenlő  $\mu$ -vel; azt mondjuk, hogy  $\bar{x}$  és  $\tilde{x}$  a  $\mu$  torzítatlan becslései.  $\tilde{x}$  eloszlása azonban szélesebb, mint  $\bar{x}$ -é, vagy a szórásokat tekintve,  $\tilde{x}$  szórása körülbelül 20 %-kal nagyobb, mint  $\bar{x}$ -é, azt mondjuk, hogy  $\mu$   $\tilde{x}$  becslése kevésbé pontos, mint az  $\bar{x}$  becslés.



**1. ábra: Szimulációs kísérlet standard normál eloszlásra,  $N=1000$  mintából  $n=10$  kiválasztására**

Ezen kívül az  $N=1000$  minta mindegyikére kiszámítottuk az  $s$  szórást és az  $R=x_{\max}-x_{\min}$  tartományt, azaz a legnagyobb és a legkisebb mintaérték különbségét. A 2a és 2b ábra mutatja ezek gyakorisági eloszlását és a kis négyzetekben átlagukat és szórásukat. Azonnal láthatjuk, hogy  $s$  a  $\sigma$  torzítatlan becslésese (elméletileg kimutatható, hogy ez is torzított, de a torzítás mértéke valójában nagyon kicsi), míg  $R$  erősen torzított, átlaga 3,021 azaz háromszor akkora, mint  $\sigma=1$ . Elméleti alapon belátható, hogy  $x$  normál eloszlása esetén  $R$  átlaga  $3,078\sigma$  (ha  $n=10$ ). Így ha minden  $R$ -t 3,078-cal elosztunk, megkapjuk  $\sigma$  torzítatlan becslését. A 2c ábra mutatja az  $R/3,078$

gyakorisági eloszlását. Ez az eloszlás szélesebb  $s$ -énél (2a ábra):  $R/3,078$  torzítatlan, de kevésbé pontos becslése  $\sigma$ -nak.



## 2. ábra: Szimulációs kísérlet standardizált normál eloszlás mellett minta vételére, $N=1000$ , $n=10$

Definíciók szerint: Egy  $\tau$  paraméter  $T$  becslése torzítatlan, ha eloszlásának  $E(T)=\mu_T$  átlaga (várható értéke)  $\tau$ -val egyenlő,

$$E(T) = \tau.$$

Más esetben torzult, és a torzulása

$$b = E(T) - \tau$$

A  $\tau$  paraméter  $T_1$  torzítatlan becslése *kevésbé precíz*, mint a  $T_2$  torzítatlan becslés, ha  $\sigma_{T_1}$  szórása nagyobb  $T_2$  szórásánál,  $\sigma_{T_2}$ -nél.  $T_1$  *relatív hatékonysága*  $T_2$ -höz képest  $\sigma_{T_2}^2 / \sigma_{T_1}^2$ . Ez az  $n_1$  és  $n_2$  mintamérettől függ, amin  $T_1$  és  $T_2$  alapul. Egyenlő mintaméretnél, ha  $n_1 = n_2 = n$ , a határérték

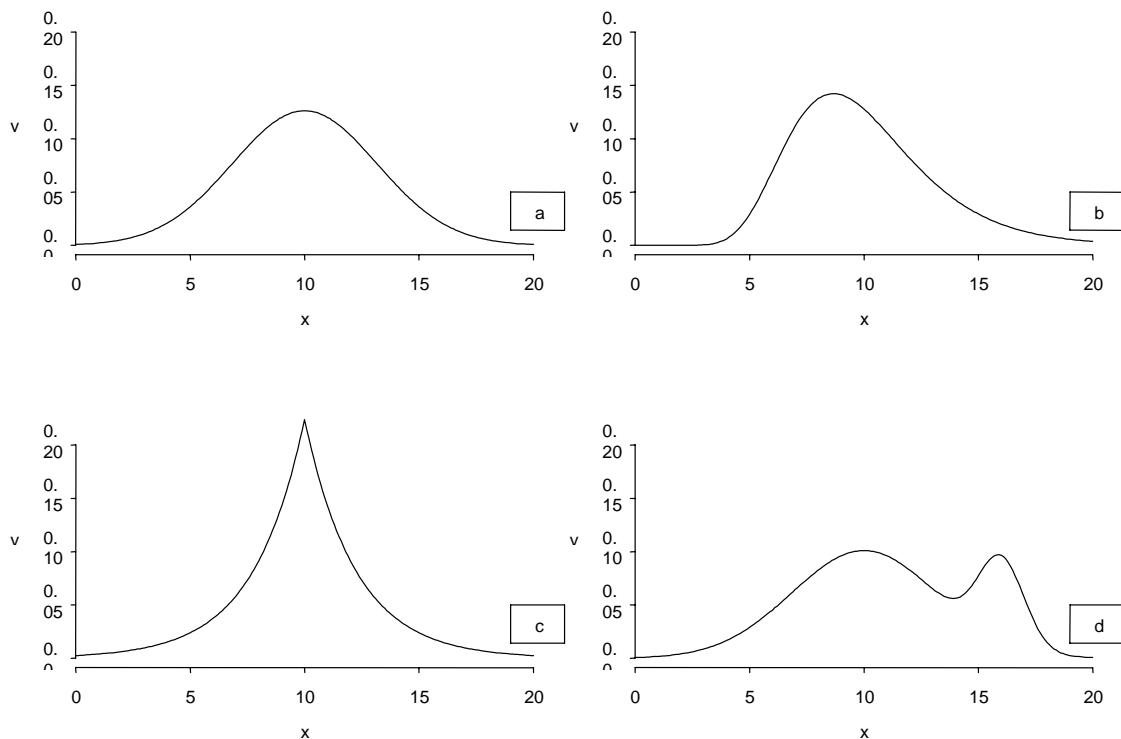
$$\lim_{n \rightarrow \infty} \sigma_{T_2}^2 / \sigma_{T_1}^2$$

amit  $T_1$ -nek  $T_2$ -re vonatkoztatott *aszimptotikus relatív hatékonyságának* nevezünk.

Ha ugyanannak a paraméternek két vagy több különböző becslése áll rendelkezésünkre, nyilvánvalóan azt választjuk, amelyik torzítatlan (vagy legkevésbé torzított) és a legpontosabb, azaz a legkisebb a szórása. A becslés ezen tulajdonságai azonban  $x$  mögöttes eloszlásától függenek, jelen esetben a normál eloszlástól.

### 3. Eltérések a normál eloszlástól

A normál eloszlásnak (Gauss-eloszlás) harang alakú a valószínűség-sűrűségi görbéje (3a ábra). Ez az eloszlási típus azonban csak egy a végtelen sokféle eloszlás között. A normalitástól tipikusan eltérnek a ferde eloszlások (3b ábra, mely lognormál eloszlást mutat), a hatványfüggvény (csúcsos) eloszlások, azaz olyan eloszlások, ahol nagyon kis és nagyon nagy értékek nagyobb valószínűséggel fordulnak elő, mint a normál eloszlásban (3c ábra amely Laplace eloszlást mutat), és a kevert eloszlások, melyek két vagy több eloszlás keverékéből állnak (pl. normál eloszlásokéből). (A 3d ábra két normál eloszlás keverékét mutatja, a bal oldaliba esik a minta értékeinek 80%-a, míg a jobbra a 20%.)



**3. ábra: Az x jellemző négy különböző eloszlása**

A 3a-3c ábrák és a 3d ábra bal normál eloszlásának ugyanaz az átlagértéke,  $\mu=10$ , és azonos a szórása is,  $\sigma=\sqrt{10}=3.16$ , azaz csak alakjukban különböznek.

A következőkben egy ilyen eloszlás paramétereinek két különböző esetét vizsgáljuk meg.

- A 3b és 3c ábra esetében a normáltól eltérő eloszlásunk van, de még mindig az egész eloszlás paramétereinek érdekelnék.
- Az eloszlások keverésének számos lehetősége közül egy speciális esetet vizsgálunk meg. A kevert eloszlás két eloszlás keveréke: az észlelések többségét magában foglaló első egy normál eloszlás (paramétereinek  $\mu$  és  $\sigma$ ), a második, amelybe a megfigyelések csak kis száma esik, egy bármilyen

típusú eloszlás, átlaga távol van  $\mu$ -tól vagy a szórása nagyobb  $\sigma$ -nál. Ezt az eloszlást *kontaminált normál* eloszlásnak nevezzük. Ha ebből az eloszlásból veszünk mintát, nagyon nagy valószínűséggel lesznek olyan megfigyeléseink, amelyek távol esnek azon észlelések többségétől, amelyeket normál eloszlásnál várnánk. Ezeket az észleléseket kiesőnek hívjuk. Ilyen kieső helyzetben nem a kevert eloszlás átlaga és szórása érdekel minket, hanem csak a nem kontaminált normál eloszlás  $\mu$  és  $\sigma$  értékei.

Mi következik ebből az  $x$  eloszlás paramétereinek becslésére, tekintettel arra, hogy típusa nem ismert?

Mivel általában azt várjuk, hogy  $x$  eloszlása normál eloszlás, a  $\mu$ -re és  $\sigma$ -ra olyan becsléseket választunk, melyek torzítatlanok és pontosak a normalitás feltételezése esetén. Ezen a becslés-csoporton kívül olyanokat választunk, melyeknek van még egy további tulajdonsága, mégpedig:

- Hogy a gyenge becsléseket elkerüljük az A) esetben, azaz olyan eloszlástípusnál, amely a normáltól különbözik, olyan becsléseket keresünk, melyek megőrzik jó tulajdonságaikat (torzítatlanság és nagy pontosság) még akkor is, ha  $x$  eloszlása kissé (nem extrém módon) különbözik a normál eloszlástól.
- Hogy a B) esetben elkerüljük a gyenge becsléseket, azaz kontaminált normál eloszlásnál, olyan becsléseket keresünk, melyek csak a nem kontaminált normál eloszlás  $\mu$  és  $\sigma$  értékét becslik (torzítatlanul és nagy pontossággal) azaz „kizárják” a kieső megfigyelések befolyását.

A) esetet könnyű kivitelezni. A statisztika elméletéből tudjuk a következőket:

- $\bar{x}$  a  $\mu$  torzítatlan becslése **minden** eloszlástípusnál.
- $x$  normál eloszlásánál az  $\bar{x}$  átlag a  $\mu$  legjobb becslése, azaz a lehető legkisebb szórású.
- $s^2$  a  $\sigma^2$  torzítatlan becslése **minden** eloszlástípusnál: ezért  $s$  nem lehet a  $\sigma$  torzítatlan becslése (még normál eloszlás esetén sem), azonban a torzítás kicsi ( $n=10$  esetén normál eloszlásnál a  $0,973\sigma$  a  $\sigma$  helyett) és még ha az  $x$  eloszlása kissé el is tér a normáltól, a torzítás még mindig elhanyagolható.
- $x$  normál eloszlása esetén az  $s^2$  variancia a  $\sigma^2$  legjobb becslése, azaz a legkisebb szórású (és gyakorlatilag ugyanez érvényes  $s$ -re mint  $\sigma$  becslésére).

#### 4. A $\sigma$ kieső értékkel szemben robusztus becslései

Az  $s$  szórás alternatíváiként  $\sigma$  három különböző becslését tárgyaljuk. A becslések működésének bemutatásához a következő példát használjuk: vegyük a már rendezett,  $n=5$  elemű mintát: 34, 42, 42, 53, 67 melyet egy olyan normál eloszlásból vettünk, melynek jellemzői:  $\mu=50$  és  $\sigma=10$ . Egy második lépésben

véletlenszerűen kiválasztottuk a 41-es minta értéket és a kieső 410-re cseréltük, egy harmadik lépésben véletlenszerűen kiválasztottuk a 67-es mintát és egy második kivülesővel, a 6700-zal helyettesítettük, egy negyedik lépésben pedig mindkét kieső értéket tízszeresével helyettesítettük.

1. minta:	34	41	42	53	67		
2. minta:	34		42	53	67	410	
3. minta:	34		42	53		410	6700
4. minta:	34		42	53		4100	67000

A minta változásának mennyiségi leírásához első megközelítésében a mintaértékek eltéréseit vizsgáljuk. Az 1. mintához a következő mátrixot számoljuk ki:

34	41	42	53	67	
34	0	-7	-8	-19	-33
41	7	0	-1	-12	-26
42	8	1	0	-11	-25
53	19	12	11	0	-14
67	33	26	25	14	0

melynek minden minta értékre egy sora és egy oszlopa van, és az  $i$ -edik sor  $j$ -edik oszlopában az érték az  $x_i - x_j$ . Az eltérések átlaga,

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j),$$

nullával egyenlő, akármik is a minta értékei, mivel minden eltérés egyszer negatív, egyszer pozitív jellel szerepel. Ezért minden eltérést csak egyszer veszünk, azaz kitöröljük az értékeket a mátrix átlójában (a minta értékei és önmaguk közötti különbségeket, ami 0) és csak a mátrix felső háromszögét használjuk és négyzetreemeléssel szabadulunk meg az előjelektől. Ennek eredményeképpen kapunk  $n(n-1)/2$  négyzetes eltérést  $(x_i - x_j)^2$  a minta értékei között, és átlagukat használjuk

$$q^2 = \frac{1}{n(n-1)/2} \sum_{i=1}^n \sum_{j>i}^n (x_i - x_j)^2 = 332.6$$

mint a változás (nem szokásos) mérőszámát. Ha azonban kiszámoljuk a minta varianciáját ismét a kiesők nélkül,  $s^2=166,3$ -at kapunk és látjuk, hogy

$$s^2 = q^2 / 2.$$

Ez elméleti úton is bizonyítható: a minta varianciája  $s^2$  az összes négyzetre emelt eltérés átlaga szorozva  $1/2$ -del. A szórás

$$s = \sqrt{s^2} = 12.9$$

annak a normál eloszlásnak a  $\sigma=10$  becslése, amiből a mintákat vettük.

Ha kiszámítjuk  $s$ -t a 2, 3 és 4. mintára, 161,9-et, 294,4-et és 29542,8-at kapunk, melyek – a kieső értékek miatt – drámaian fölébecslik a  $\sigma=10$ -et. Megfigyelhetjük, hogy már egyetlen nagy kieső érték is használhatatlan  $s$  becsléshez vezet. Ha ez a kieső érték tart a végtelenhez, az “ $s$ ” értéke is tart a végtelenhez.

Egy becslés letörési pontját úgy definiáljuk, mint a szélsőségesen nagy kieső értékek azon legnagyobb hányadát, amely még nem okozza a szórás végtelenhez tartását, azaz a becslés elromlását. Általában a letörési pont függ a mintamérettől. Az  $n \rightarrow \infty$  esetén *aszimptotikus letörési pontnak hívjuk*.

$s$  letörési pontja 0: a szórás szélsőségesen nem robusztus a kieső értékekkel szemben.

### Hogy találhatjuk meg $s$ robusztus alternatíváit?

Az első alternatíva megváltoztatja a minta szórásnégyzetének képletét:

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

az  $\bar{x}$  átlagot a  $\tilde{x}$  mediánnal helyettesítve (amely igen robusztus a kiesőkkel szemben), az eltérések négyzete helyett az eltérések abszolút értékeit használja (így a nagy eltérések nem kerülnek túlsúlyba a négyzetreemeléssel) és nem az átlagukat használja, hanem a mediánjukat. A  $\sigma$ -nak ez a becslése torzított: normál eloszlás mellett  $\sigma/1,4826$ -ad ad. Ennél fogva a torzítással korrigált *medián abszolút eltérés*

$$MAD = 1,4826 \text{ med}(|x_i - \tilde{x}|)$$

$\sigma$  torzítatlan becslése normál eloszlási körülmények között. Közelítési határfoka 36,7%, letörési pontja 50% (a lehető legjobb érték). 1, 2, 3 and 4 mintáinkra a  $MAD$  11,9; 20,8; 28,2 és 28,2. A kieső értékek a  $MAD$  –ot azért befolyásolják, mert  $\tilde{x}$  változik, ha egy  $\tilde{x}$  -nál kisebb érték annál nagyobb kiesőre változik.

Egy második alternatíva ismét a minta értékei egymástól való eltéréseinek mátrixából indul ki, és azok abszolút értékét használja. A mátrix  $i$  sora mindegyikének a mediánját veszi, majd a mediánok mediánját. Ezt Rousseeuw és Croux vezette be [1]  $S_n$ -ként,

$$S_n = \text{med}_i \left( \text{med}_j |x_i - x_j| \right).$$

Meg kell jegyezni, hogy ebben a képletben páros mintaszám mellett a mediánt nem a szokásos módon számítjuk, azaz a rendezett minta közepén levő két érték átlagaként. Ehelyett minden sor mediánját “magas” mediánként, azaz a két középső érték közül a magasabbként számoljuk, és a sor mediánok



mediánját “alacsony” mediánként, azaz a két “középső” sor medián közül az alacsonyabbnak számítjuk

$S_n$  normál eloszlás mellett  $\sigma=1,1926$  becslést ad. Így,  $1,1926S_n$  a  $\sigma$  torzítatlan becslése normál eloszlás mellett. Hatékonysága 58,2% , letörési pontja 50%. A mi 1, 2, 3 és 4. mintánkra a  $1,1926S_n$  becslések 9,5; 22,7; 22,7 és 22,7. A következőkben bemutatjuk  $S_n$  kiszámítását az 1. mintára:

	34	41	42	53	67	medián
34	0	7	8	19	33	8
41	7	0	1	12	26	7
42	8	1	0	11	25	8
53	19	12	11	0	14	12
67	33	26	25	14	0	25

Az eltérések abszolút értékeinek mátrixa, melyet az egyes sorok mediánjainak oszlopával bővítettünk. Ennek a mediánja  $S_n = 8$ , a becslés pedig  $1,1926S_n = 1,1926 \cdot 8 = 9,5$ .

Egy harmadik alternatíva ismét a minta értékeinek egymástól való eltéréseinek mátrixából indul ki, abszolút értékeit használja, de az ismétlődő értékek nélkül, azaz a mátrixnak csak a felső háromszögét használva. Ezeket növekvő sorrendbe rendezi. Válasszuk ki a mediánjukat:

$$med\left(\left|x_i - x_j\right|_{i < j}\right)$$

mint  $\sigma$  becslését. Az ezen a mediánon alapuló becslés letörési pontja azonban kisebb mint a kvantilis becslésé, ami durván 25 %, azaz az 50 % kvantilis helyett a 25 % kvantilis. Ezt a becslést Rousseeuw and Croux [1] vezette be mint  $Q_n$ ,

$$Q_n = \left( \left|x_i - x_j\right|_{i < j} \right)_{[k]} \text{ ahol } k = \binom{h}{2} \text{ és } h = [n/2] + 1$$

ahol  $[x]$  az  $x$ -et meg nem haladó legnagyobb egész szám. Normál eloszlás mellett a becslés  $\sigma/c_n$  ahol

$$c_n = 2,2219 \cdot \frac{n}{n+1,4} \quad \text{ha } n \text{ páratlan}$$

és

$$c_n = 2,2219 \cdot \frac{n}{n+3,8} \quad \text{ha } n \text{ páros}$$

Ennél fogva,  $c_n Q_n$  normál eloszlás esetén a  $\sigma$  torzítatlan becslése. Hatásfoka 82,3%, letörési pontja 50%. A mi 1, 2, 3 and 4 mintáinkra a  $c_n Q_n$

becslései 13,9; 24,3; 33,0 és 33,0. A következőkben bemutatjuk  $Q_n$  számítását az 1 mintára:

	34	41	42	53	67
34		7	8	19	33
41			1	12	26
42				11	25
53					14
67					

A minta értékeinek egymástól való abszolút eltéréseit tartalmazó mátrix felső háromszöge, értékei növekvő sorrendben 1, 7, 8, 11, 12, 14, 19, 25, 26, 33. Mivel  $n=5$ , a következőt kapjuk

$$h = [n/2] + 1 = [2.5] + 1 = 2 + 1 = 3, \quad k = \binom{h}{2} = \binom{3}{2} = 3,$$

$$c_n = 2.2219 \frac{n}{n+1.4} = \frac{2.2219 \cdot 5}{6.4} = 1.736$$

és ebből adódóan,  $Q_n = x_{[3]} = 8$ .  $c_n Q_n = 1,736 \cdot 8 = 13,9$   $\sigma$  becslése.

Az 1-4 mintára kapott minden  $\sigma$  becslést a következő táblázatban foglaltunk össze:

minta	s	MAD	$S_n$	$Q_n$
1	12,9	11,9	9,5	13,9
2	161,9	20,8	22,7	24,3
3	2940,4	28,2	22,7	33,0
4	29542,8	28,2	22,7	33,0

Az összehasonlítás azt mutatja, hogy a  $\sigma$  robusztus becslései állandók maradnak, ha a kieső értékek nőnek, de az  $s$  szórás drámaian megnő a kieső értékek növekedésével.

Összefoglalva, az  $s^2$  aszimptotikus letörési pontja 0, a másik három becslésé pedig 50%, a lehető legmagasabb érték, azaz szélsőségesen nem robusztus. A robusztus becslések viszont szélsőségesen robusztusak a kieső értékekkel szemben. A  $MAD$ ,  $S_n$  and  $Q_n$  hatékonysága  $s^2$  -tel összehasonlítva 36,7%, 58,2% és 82,3%. Világos, hogy  $s$  a legjobb választás, ha  $x$  normál eloszlású, és  $Q_n$  a legjobb választás, ha fennáll a veszélye annak, hogy kieső értékek vannak a mintákban. A négy becslést ( $s$ ,  $MAD$ ,  $1,1926S_n$ ,  $cQ_n$ ) a normál eloszlástól való eltérés körülményei között szimulációs kísérletben hasonlítjuk össze.

## 2. szimulációs kísérlet:

Mint az 1. szimulációs kísérletben is,  $N = 1000$  mintából  $n = 10$  -et választunk

- a) standard normál eloszlásból ( $\mu = 0, \sigma = 1$ ),
- b) chi négyzet eloszlásból, melynek szabadsági foka  $f = 5$   
 ( $\mu = 5, \sigma = \sqrt{10} = 3,16$ ) mint ferde eloszlási példa
- c) kontaminált normál eloszlásból, ahol 0,9 a valószínűsége annak, hogy a minta értékei egy olyan standard normál eloszlásból származnak, melynek jellemzői ( $\mu = 0, \sigma = 1$ ), és 0,1 a valószínűsége annak, hogy a minta értékei olyan standard normál eloszlásból vannak, melyre ( $\mu = 10, \sigma = 1$ ) a jellemző. Ebben az esetben a globális átlag  $\mu_G = 0,9 \cdot 0 + 0,1 \cdot 10 = 1$ , a globális szórásnégyzet  $\sigma_G^2 = 0,9 \cdot 1^2 + 0,1 \cdot 1^2 + 0,9 \cdot 0,1(10-0)^2$ , és a globális szórás  $\sigma_G = \sqrt{10} = 3,16$ . A cél azonban a nem kontaminált eloszlás ( $\mu = 0, \sigma = 1$ ) paramétereinek becslése, nem pedig a globális paramétereké.

A 4., 5. és 6. ábra a négy becslés,  $s, MAD, 1,1926S_n, cQ_n$  hisztogramjait mutatja az a), b) és c) eloszlásra. A négyzetekben tüntettük fel átlagukat és szórásukat. A 4. ábra azt mutatja, hogy normál eloszlás esetén mind a négyféle becslés torzítatlan (mivel véges számú mintán alapulnak, a minta átlag nem lehet pontosan a várt 1-es érték), pontosságuk pedig  $s, Q_n, MAD, S_n$  sorrendben csökken. Világos, hogy  $s$  a legjobb választás. Az 5. ábra azt mutatja, hogy ferde, chi négyzet eloszlás mellett, ahol a szabadsági fokok száma  $f=5$ ,  $s$  kevésbé torzított, a többi becslés negatív torzítása pedig 10% és 15% között van. A precizitásuk nem különbözik számottevően, Itt is az  $s$  a legjobb választás. A 6. ábrán látható, hogy kontaminált normál eloszlás esetén  $s$  a globális szórásra  $\sigma_G = 3,16$ -ot ad  $\sigma = 1$  helyett. A többi módszerek  $\sigma = 1$ -et adnak;

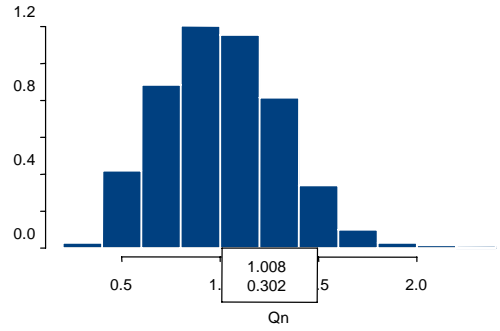
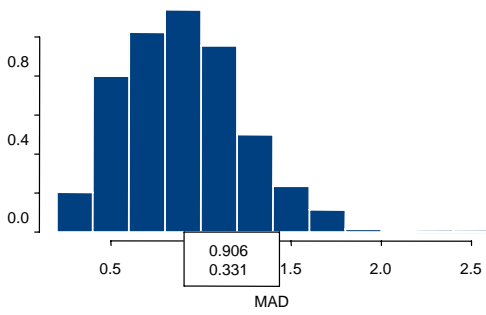
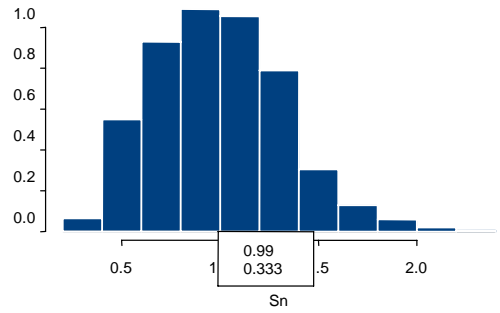
$MAD$ -nak van a legkisebb torzítása és a legjobb precizitása.  $MAD$  a legjobb választás, azonban statisztikai elméleti szempontból  $Q_n$  a kedvezőbb.

### **Javaslatok $\sigma$ robusztus becslésére:**

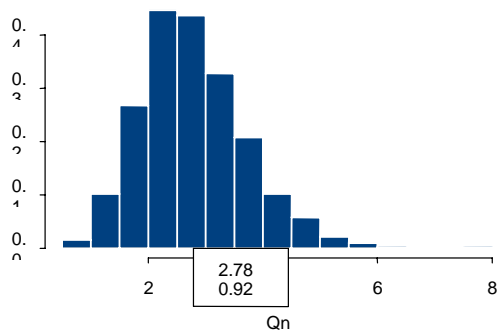
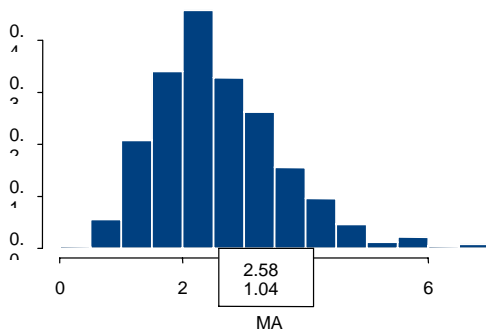
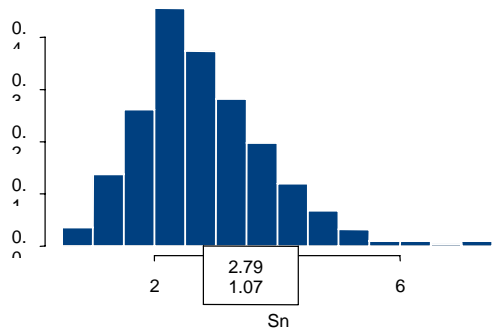
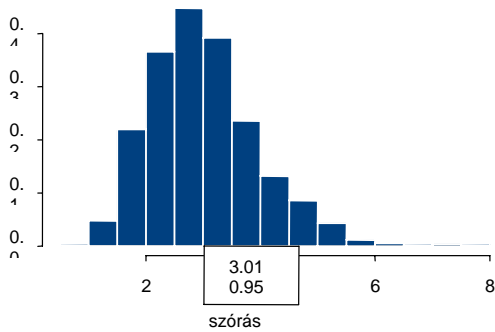
Feltételezve, hogy a vizsgált  $x$  jellemző (közelítőleg) normál eloszlású,  $\sigma$  becslésére  $s$  a legjobb választás. .

Ha  $x$  normál eloszlású, azonban előfordulhatnak kieső értékek, a legjobb választás  $Q_n$ , a robusztus becslés.

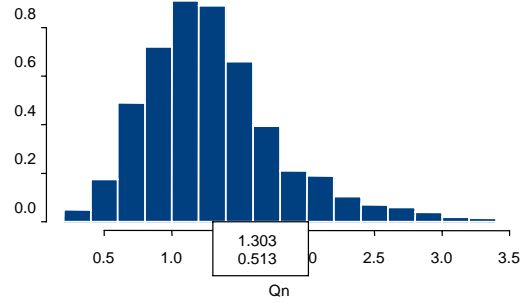
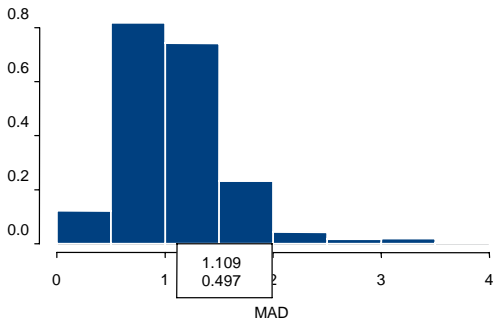
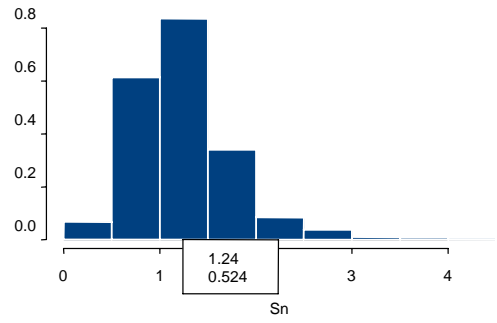
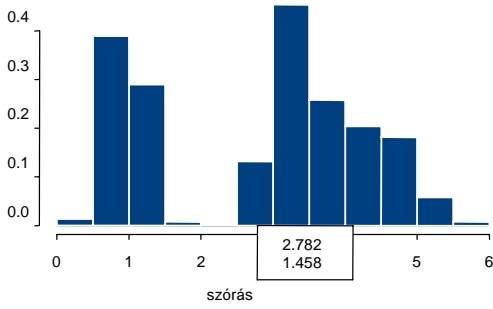
Ha ismert, hogy  $x$  nem normál eloszlású, úgy kell transzformálni, hogy  $y = f(x)$  közelítőleg normál eloszlású legyen, vagy olyan becsléseket kell alkalmazni, melyek  $x$  speciális eloszlásán alapulnak. Ebben az esetben  $\sigma$  robusztus becslése nem segít.



**4. ábra: Szimulációs kísérlet standardizált normál eloszlásból vett mintával,  $N=1000$ ,  $n=10$**



**5. ábra: Szimulációs kísérlet chi négyzet eloszlásból vett mintával,  $N=1000$ ,  $n=10$**



**6. ábra: Szimulációs kísérlet kontaminált normál eloszlásból vett mintával, N=1000 , n=10**

## 5. Kvantitatív mérési eredmények ismételhetőségének és reprodukálhatóságának megközelítése az ISO 5725-2 szabvány szerint

Az ISO 5725-2[4] szerint egy laboratóriumok közötti körvizsgálatban általában több mint egy anyag típus vagy minta szerepel, melyek például különböző összetételűek vagy különböző mennyiségű szennyezőt tartalmaznak. Ezeket a különböző típusokat hívják (vizsgálati) szinteknek. Minden szintre kapunk pontossági adatokat az ismételhetőség és reprodukálhatóság szórásának formájában. A mérési eredmények elemzésének statisztikai modellje az egyes szintekre

$$y_{ij} = m + B_i + e_{ij}; \quad i = 1, \dots, k; j = 1, \dots, n_i,$$

$k$  a résztvevő laboratóriumok száma

$n_i$  az ismételt mérések száma az  $i$ -edik laboratóriumban (egy bizonyos szinten)

$y_{ij}$   $j$  ( $j = 1, \dots, n_i$ ) mérési eredmény az  $i$  ( $i = 1, \dots, k$ ) laboratóriumban

$m$  általános átlag (várható érték)

$B_i$  a variancia laboratóriumi komponense (ismételhetőségi körülmények között minden laboratóriumban) az  $i$ -edik laboratóriumban; várható értéke 0, szórásnégyzete pedig  $\sigma_L^2$ , a laboratóriumok közötti variancia

$e_{ij}$  a  $j$  mérés véletlen hibája az  $i$  laboratóriumban (ismételhetőségi körülmények között); várható értéke 0, variáciája, a laboratóriumon belüli szórásnégyzet,  $\sigma_{ri}^2$ , feltehetően ugyanakkora mindegyik laborban és az ismételhetőség variációjaként jelölik,  $\sigma_r^2$

Minden szintre egyutas varianciaanalízist végzünk (random modell) hogy statisztikusan megbecsülhessük a variancia két komponensét,  $\sigma_r^2$  and  $\sigma_L^2$  -t, és a reprodukálhatóság variációját,  $\sigma_R^2 = \sigma_r^2 + \sigma_L^2$  [2]:

(1)	(2)	(3)	(4)
Forrás	Négyzetösszeg SQ	Szabadsági fokok f	Négyzetek átlaga SQ/f
Laboratóriumok között	$\sum_{i=1}^k n_i (\bar{y}_i - \bar{\bar{y}})^2$	$k - 1$	$s_{II}^2$
Laboratóriumon belül	$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	$N - k$	$s_I^2$

ahol

$$N = \sum_{i=1}^k n_i,$$

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}; i = 1, \dots, k,$$

$$\bar{\bar{y}} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij} = \frac{1}{N} \sum_{i=1}^k n_i \bar{y}_i.$$

A becslések

$$s_L^2 = \max(0, (s_{II}^2 - s_I^2) / \bar{\bar{n}}) \square \sigma_L^2,$$

$$s_r^2 = s_I^2 \square \sigma_r^2,$$

$$s_R^2 = s_r^2 + s_L^2 \square \sigma_R^2$$

ahol

$$\bar{\bar{n}} = \frac{1}{k-1} \left( N - \frac{\sum_{i=1}^k n_i^2}{N} \right).$$

Ha a laboratóriumokban egyenlő számú ismétlést végeztek, azaz ha  $n_i = n$  és  $i = 1, \dots, k$ ,  $\bar{\bar{n}} = n$ -et kapunk

A varianciaelemzést megelőzi a kieső értékek vizsgálata, amely magában foglalja annak a feltételezésnek a vizsgálatát, miszerint az összes laboratóriumban azonos a laboratóriumon belüli variancia.

## 6. Az ISO 5725-2 megközelítés egy robusztus alternatívája

A laboratóriumi körvizsgálatok eredményei legtöbbször a következő típusú kieső értékeket mutatják:

- a) olyan észlelt értékek, amelyek eléggé eltérnek az ugyanabban a laboratóriumban az ismételhetőségi körülmények között észlelt egyéb értékektől; ezek megnövelik a laboratóriumon belüli varianciát és ennél fogva az ismételhetőség szórásnégyzetét,  $s_r^2$ -t, azaz a laboron belüli varianciák átlagát;
- b) laboratóriumi átlagok, melyek nagymértékben eltérnek más laboratóriumok átlagától; megnövelik a laboratóriumi átlagok varianciáját és ezáltal a laboratóriumok közötti  $s_L^2$  varianciát és a reprodukálhatóság varianciáját,  $s_R^2$ -t.

A kieső értékek vizsgálatát azért végzik, hogy kimutassák ezeket a helyzeteket és hogy eldöntsék, megtartsák-e ezeket az adatokat az elemzéshez vagy hagyják el mint kieső értéket. A kieső értékeknek ezt a kezelését, melynek mindig van szubjektív eleme, elkerülhetjük, ha  $\sigma_r$  és  $\sigma_R$  robusztus becsléseket használjuk az  $s_r$  és  $s_R$  helyett.

A Royal Society of Chemistry Analitikai Módszerek Bizottsága[6] Huber[3]. javaslata alapján ajánlja  $\sigma_r$  és  $\sigma_R$  robusztus becslését. Ezek a becslések lassan konvergálnak, letörési pontjuk 25 %-nál kisebb, mint azt Reichenbach egy szimulációs vizsgálattal kimutatta [7]. Ehhez képest jobb a Svájci Élelmiszerkönyben javasolt megközelítés [8]. Az ISO 5725-5 [5] olyan módszert javasolt, amely csonkolja a távoli megfigyeléseket, ahelyett hogy kisebb súllyal venné, így tehát inkább a kiesők automatikus elvetésének módszere. Lischer [10] azt javasolja, hogy ha minden laboratóriumban  $n = 2$  ismételt mérést végeznek, használják a Rousseouw féle  $Q_n$  -t[1] a szórások becslésére. A [9] irodalom azt javasolja, hogy  $\sigma_r$  robusztus becsléseként használják a laboratóriumon belüli szórások mediánját, és a laboratóriumi átlagok Rousseouw szerinti  $S_n$  [1] értékét a  $\sigma_R$  becslésére.

Vizsgáljuk azt az esetet, ahol minden laboratóriumban  $n \geq 2$  ismételt mérést végeznek, és javasoljuk egy laboratóriumon belüli és laboratóriumok közti  $Q_n$  érték használatát  $\sigma_r$  és  $\sigma_R$  becslésére.

Az alkalmazott statisztikai modell szerint, a különböző laboratóriumokban észlelt  $y_{ij}$  értékeknek különböző az átlaga, de azonos a  $\sigma_r^2$  szórásnégyzete. Az észlelt értékek és laboratóriumi átlagaik közötti különbség

$$d_{ij} = y_{ij} - \bar{y}_i$$

azonosan oszlik meg, 0 átlaggal és  $((n-1)/n)\sigma_r^2$  varianciával.

Ezért  $s_i$  helyett a következő képletet használjuk:

$$s_{I,rob} = \sqrt{n/(n-1)}Q_n(d)$$

$\sigma_r$  becslése ar  $d = (d_{ij}; i=1, \dots, k; j=1, \dots, n)$  vektor.  $s_{II}$  helyett pedig a következőt alkalmazzuk

$$s_{II,rob} = \sqrt{n}Q_n(a)$$

$a = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_k)$  vektorral. Az  $s_r^2$  and  $s_R^2$  számítási képleteibe,  $s_{I,rob}$  és  $s_{II,rob}$  helyett  $s_i$  -et illetve  $s_{II}$  -t helyettesítjük be.

Meg kell jegyeznünk, hogy a  $d$  vektor elemei nem függetlenek, s mivel a torzítást korrigáló tényezőt  $Q_n$ -ben úgy vezettük le, hogy  $n$  független megfigyelést feltételeztünk, végül  $\sigma_r$  és  $\sigma_R$  kissé torzított becsléséhez jutunk. Ahogy azonban egyes szimulációs számítások mutatják, a torzítás legfeljebb 5%.

Abban az általános esetben, ha a laboratóriumokban az ismételt mérések  $n_i$  száma eltérő,  $n$ -et a  $s_{I,rob}$  és  $s_{II,rob}$  képletében  $\bar{n}$  -gal kell helyettesítenünk.  $s_{II,rob}$  azonban torzul, mivel nem veszi figyelembe az ismétlések különböző számát az egyes laboratóriumokban. Ennélfogva ezt az eljárást csak akkor javasoljuk, ha az  $n_i$  számok nem különböznek nagyon, azaz a kísérletet egyenlő számú  $n$  ismétléssel tervezték, de véletlenül egy ismételt mérés hiányzik.

## Egy példa

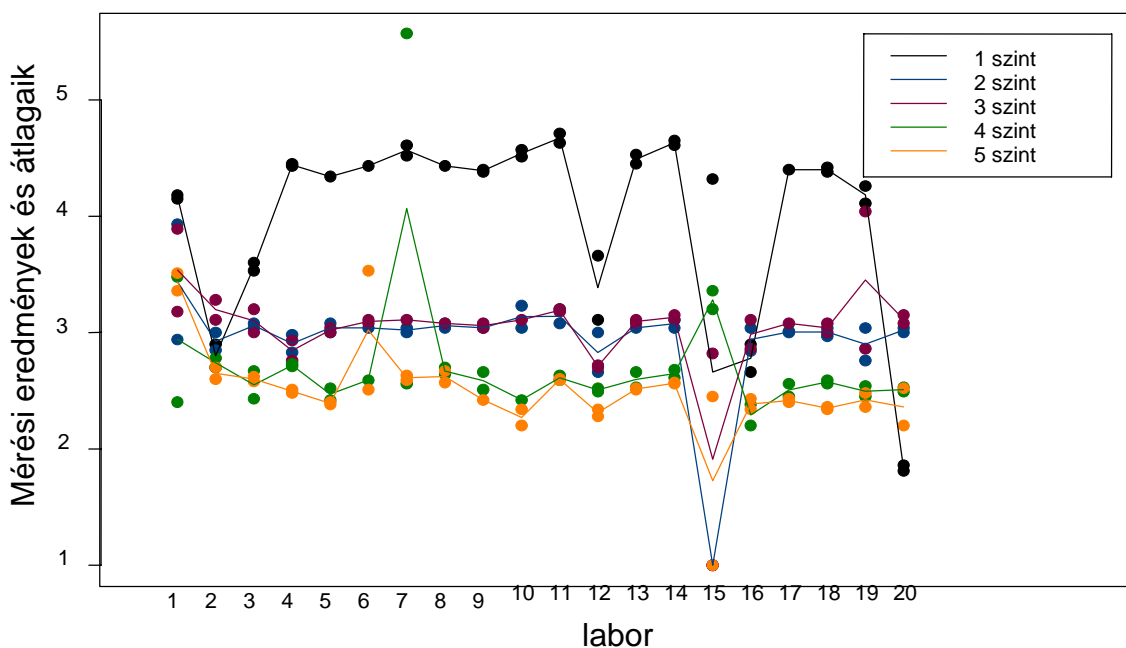
Egy aerob telepszámlálási (30 °C)-körvizsgálatban, amelyet a “Nyers és hőkezelt tej mikrobiológiai és kémiai elemzése Európai Közösségi módszereinek validálása” című project részeként az Egyesült Királyság Mezőgazdasági, Erdészeti és Halászati Minisztériuma (MAFF) [11] szervezett és kivitelezett 1991-ben, húsz résztvevő öt tétel (szint) tejből 20-20 ml-t két ismétlésben, véletlenszerűen számozva, azaz összesen 10 mintát kapott a 30 °C-on végzett telepszámláláshoz. Az eredményeket a következő táblázatban foglaltuk össze:

labor	szint									
	1	2	3	4	5	6	7	8	9	10
1	4,18	4,15	2,94	3,93	3,18	3,89	2,40	3,48	3,51	3,36
2	2,90	2,70	2,85	3,00	3,11	3,28	2,70	2,78	2,60	2,70
3	3,60	3,53	3,08	3,04	3,00	3,20	2,67	2,43	2,58	2,62
4	4,43	4,45	2,83	2,98	2,76	2,93	2,73	2,71	2,48	2,51
5	4,34	4,34	3,00	3,08	3,00	3,04	2,52	2,42	2,40	2,38
6	4,43	4,43	3,04	3,04	3,08	3,11	2,59	2,59	2,51	3,53
7	4,52	4,61	3,00	3,04	3,11	3,11	2,56	5,57	2,63	2,59
8	4,43	4,43	3,08	3,04	3,08	3,08	2,70	2,64	2,67	2,57



9	4,40	4,38	3,04	3,04	3,04	3,08	2,51	2,66	2,42	2,42
10	4,51	4,57	3,23	3,04	3,11	3,11	2,42	2,42	2,34	2,20
11	4,63	4,71	3,08	3,20	3,18	3,20	2,63	2,59	2,59	2,60
12	3,11	3,66	2,66	3,00	2,72	2,70	2,49	2,52	2,34	2,28
13	4,45	4,53	3,04	3,04	3,11	3,08	2,53	2,66	2,52	2,51
14	4,61	4,65	3,04	3,11	3,11	3,15	2,61	2,68	2,57	2,56
15	4,32	1,00	1,00	1,00	2,82	1,00	3,36	3,20	2,45	1,00
16	2,90	2,66	2,84	3,04	3,11	2,86	2,38	2,20	2,43	2,34
17	4,40	4,40	3,01	3,00	3,08	3,08	2,56	2,45	2,43	2,40
18	4,42	4,38	3,04	2,97	3,00	3,08	2,56	2,59	2,36	2,34
19	4,11	4,26	3,04	2,76	4,04	2,86	2,54	2,45	2,36	2,48
20	1,81	1,86	3,04	3,00	3,15	3,08	2,49	2,53	2,52	2,20

A 7. ábrán látható grafikon azt mutatja, hogy az egy laboratóriumon belüli mérések és a laboratóriumi átlagok is erősen eltérőek voltak.



**7. ábra: Az aerob telepszám-vizsgálat mérési értékei és átlagok**

A következő táblázatban összehasonlítjuk az ismételtetés  $s_r$  szórását és a reprodukálhatóság  $s_R$  szórását a kieső értékek elhagyása nélkül a robusztus módszer alkalmazásával, illetve az ISO 5725-2 szabályai szerint a kiesők elhagyásával. Az elhagyott kieső laboratóriumok száma 2 és 5 között változik, és ezért  $s_r$  és  $s_R$  becslései a kiesők kihagyásával, illetve anélkül meglehetősen különbözőek. A robusztus becslések jobban összhangban vannak a kiesők elhagyásával kapott eredményekkel, bár vannak figyelemreméltó különbségek, például az  $s_R$  értékek között.

A robusztus módszerrel kapott eredmények figyelemre méltóak, és a módszer szükségtelessé teszi annak megvitatását, mely értékeket kell elhagyni, mint kiesőket.

Szint	Laborok száma	A kiesők elhagyása nélkül		Robusztus		Laborok száma	A kiesők elhagyásával	
		$s_r$	$s_R$	$s_r$	$s_R$		$s_r$	$s_R$
1	20	0,536	0,905	0,072	0,331	18	0,066	0,793
2	20	0,183	0,488	0,100	0,125	18	0,099	0,107
3	20	0,367	0,411	0,057	0,119	15	0,071	0,075
4	20	0,511	0,527	0,100	0,174	17	0,073	0,122
5	20	0,289	0,383	0,072	0,193	16	0,047	0,126

## 7. Hivatkozások

- [1] Rousseeuw, P. J., and Croux, C. (1993): Alternatives to the median absolute deviation. *J. Am. Stat. Ass.* **88**, 1273-1283
- [2] Graf, U., Henning, H.-J., Stange, K., Wilrich, P.-Th. (1987): *Formeln und Tabellen der angewandten mathematischen Statistik*. Berlin: Springer
- [3] Huber, P. J. (1981): *Robust Statistics*. New York: Wiley
- [4] ISO 5725-2:1994: Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method. International Standardisation Organisation, Geneva
- [5] ISO 5725-5:1998: Accuracy (trueness and precision) of measurement methods and results – Part 5: Alternative methods for the determination of the precision of a standard measurement method. International Standardisation Organisation, Geneva
- [6] Analytical Methods Committee of the Royal Society of Chemistry (1989): Robust Statistics – How Not to Reject Outliers. Part 1: Basic concepts, *The Analyst* **114**, 1693 – 1697. Part 2: Inter-laboratory trials, *The Analyst* **114**, 1699 – 1702
- [7] Reichenbach, A. (1989): *Robuste Methoden für die Auswertung von Ringversuchen*. Diplomarbeit an der ETH Zürich
- [8] SLB 1989. *Schweizerisches Lebensmittelbuch*, Kapitel 60: Statistik und Ringversuche. Bern: EDMZ
- [9] Lombard, B. (2003): Draft ISO/IDF Standard (Version 2 - March 2003). Microbiology of food and animal feeding stuffs- Protocol for the establishment of precision characteristics of quantitative methods by interlaboratory studies
- [10] Lischer, P. (1993): Ringversuche zur Bestimmung des Qualitätsstandards von Laboratorien. Seminar der Region Oesterreich-Schweiz der Internationalen Biometrischen Gesellschaft
- [11] Scotter, S., Aldridge, M., Back, J., Wood, R. (1993): Validation of European Community methods for microbiological and chemical analysis of raw and heat-treated milk. *J. Assoc. Publ. Analysts* **29(1)**, 1-32