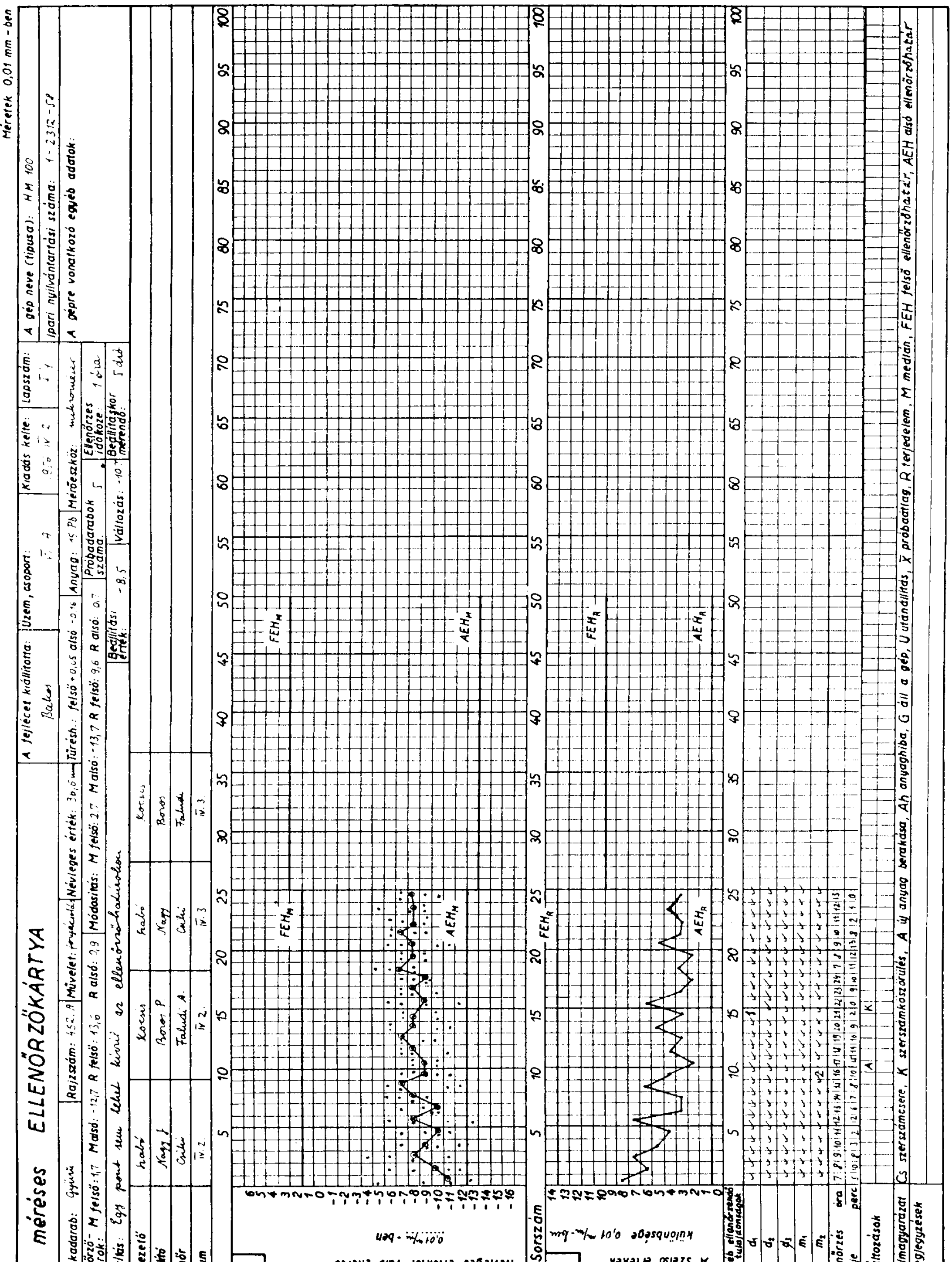


4-9. ábra. Medián kártya

$$UCL_{\tilde{x}} = \overline{\tilde{x}} + \sqrt{\pi / 2} A_2 \overline{R} \,,$$

$$LCL_{\tilde{x}} = \overline{\tilde{x}} - \sqrt{\pi / 2} A_2 \overline{R} \,.$$

A képletekben $\overline{\tilde{x}}$ a mediánok átlaga:

$$\overline{\tilde{x}} = \frac{1}{m} \sum_{i}^{m} \tilde{x}_i$$

Az átlaggal ellentétben a medián közelítőleg sem normális eloszlású, ha *x* nem volt az, mert a mediánra nem vonatkozik a centrális határeloszlási tétel. Ezért a medián-kártyára ilyenkor csak a 3.4. pont szerinti nem-paraméteres tesztek használhatók (pl. túl sok pont van a középvonal, az átlagos medián egyik oldalán). A medián előnyösebb, mint az egyedi érték volna, mert a mintaelemszám növelésével érzékenyebbé tehető.

A szóródás jellemzésére olyan esetekben, amikor egyszerűsége miatt a medián-kártyát használják, a terjedelem-kártya túlságosan bonyolult volna, egy terjedelem-ablakot szokás használni, amellyel ellenőrizhetjük, hogy a kártyán ábrázolt adatok terjedelme nem halad-e meg egy határt (Pyzdek, 1992; p. 89).

A medián-kártya, ugyanúgy, mint az átlag-kártya nemcsak terjedelem-, hanem szórás- vagy szórásnégyzet-kártyával is kombinálható.

Az MSZ 246/2-56 szabványban adott medián-terjedelem kártya-kombinációt mutat a 4-9. ábra.

4.10. Ellenőrző kártya egyedi értékekre

Sokszor a gyártás természete olyan, hogy a termékek egyenként keletkeznek (pl. a vegyipar szakaszos technológiáinál az adagok (sarzsok), az ismételt vizsgálatok csak a kémiai analízis - a meghatározás - hibájával különböznének egymástól); vagy a gyártás lassú ahhoz, hogy csoportokat lehessen formálni a termékekből, és/vagy minden egyes termék-példányt minősítenek, ellenőriznek (pl. automatikusan). Az is előfordul, hogy nincs mód többelemű minta vételére, mert maga a mérés drága (pl. roncsolással történik). Bizonyos gyártásoknál (pl. papíripar) a termék jellemzője az egyik (szélességi) irányban alig változik, a másik irányban jelentősen. Azok a mintaelemek tartoznának egy racionális csoportba, amelyek a hosszú tekercsben készülő anyag egy szakaszáról, a szélesség szerint különböző helyekről származnak, de ezek olyan kis mértékben különböznek egymástól, hogy nem tükrözik a folyamat változékonyságát, a jellemző ingadozását, ezért túlságosan kicsiny σ értéket adnának, amihez képest a hossz-menti természetes ingadozás szélsőségesen nagynak minősülne és rengeteg kiesô pontot kapnánk.

Minthogy nem átlagokat ábrázolunk, a pontok csak akkor normális eloszlásúak, ha maguk az adatok ilyenek, ezt normalitás-vizsgálattal, hisztogrammal, egyéb grafikus

módszerekkel ellenőrizhetjük. Ha a normális eloszlás nem valósul meg, a kártya statisztikai kezelése kétségessé válik.
Ilyenkor a minta egyetlen elemből áll, és természetesen nem számíthatjuk ki a szórásnégyzetét vagy terjedelmét, ezért nem magától értetődő a beavatkozási határok kijelölése.

Egyedi értékek esetén a mozgó terjedelmet (az egymás utáni két példány, adag mért jellemzője közötti különbséget) használják az ingadozás kifejezésére:

$$MR_i = |x_i - x_{i-1}|.$$

Ebből $\hat{\sigma} = \dfrac{\overline{MR}}{d_2}$, ahol $\overline{MR} = \dfrac{\sum_{i=2}^{m} MR_i}{m-1}$.

A középvonal és a beavatkozási határok:

$$CL_x = \bar{x}, \qquad UCL_x = \bar{x} + \frac{3\overline{MR}}{d_2}, \qquad LCL_x = \bar{x} - \frac{3\overline{MR}}{d_2}.$$

A d_2 konstans értéke n=2-höz veendő a függelék V. táblázatából.

Itt is először a szóródási jellemzőre vonatkozó, tehát a mozgó terjedelem-kártyát kell megvizsgálni, mert az egyediérték-kártya csak σ=konstans esetben használható megnyugtatóan.

4.11. Mozgó terjedelem-kártya (MR chart)

Egyedi adatok esetén (tehát ha az n mintaelemszám 1), az ingadozás jellemzésére a mozgó terjedelmet (MR: moving range) szokták használni.

A mozgó terjedelem-kártya paraméterei:

$$CL_{MR} = \overline{MR},$$
$$UCL_{MR} = D_4\,\overline{MR},$$
$$LCL_{MR} = D_3\,\overline{MR}.$$

A D_3 és D_4 konstansok értékei n=2-höz veendők a függelék V. táblázatából.

Az egymás után számolt terjedelem-értékek közelítően sem függetlenek egymástól, és így a szokásosan használt D_3 és D_4 értékek tulajdonképpen nem alkalmazhatók megnyugtatóan, sőt a beavatkozási határok egész számítása kérdésessé válik.

Ha az adatokban trend van (Ryan, 1991; p. 156), a mozgó terjedelemből számított becsült variancia sokkal kisebb annál, amit az adatokból a szokásos szórásnégyzet-formulával kapnánk, az utóbbi ugyanis a trendet is tartalmazza (3-8. ábra). Ennek az a következménye, hogy az elfogadási tartomány (*UCL*–*LCL*) túlságosan szűk lesz, ezért sok lesz a téves riasztás.

A kérdés az, hogy a beavatkozási határokat úgy akarjuk-e megállapítani, hogy a trenddel terhelt, de azon túl csak véletlenszerű ingadozást mutató adatok még elfogadhatók legyenek. Ekkor a σ^2 varianciát úgy kell becsülni, hogy ne csak az ingadozás varianciáját tartalmazza, hanem a trenddel megnövelt szóródásét.

Ha a trendet és a véletlenszerű ingadozást külön akarjuk választani, szokás úgy eljárni, hogy a trendre regresszióval függvényt illesztenek, és a függvénytől való eltérések szórásnégyzetével (a reziduális szórásnégyzettel) becsülik a varianciát. Ekkor ugyanis, ha a függvény alakja megfelelő, az eltérések csak az ingadozásnak tulajdoníthatók, és éppen az ingadozás varianciáját akarjuk becsülni. A regresszióhoz nemcsak analitikusan megadott függvényalakot használhatunk, hanem az idősor-elemzésnél használt trend-kiszűrési módszereket (ARIMA stb.) is. A regressziós ellenőrző kártyát és a trend esetén követhető eljárásokat az 5.8. pontban tárgyaljuk.

Amennyiben a trend nem kívánatos, és az előzetes adatfelvételnél műszaki intézkedésekkel ki is küszöbölik, az adatfelvételt a trendmentes állapotban újra el kell végezni.

Ha nincs trend az adatokban, az előzetes adatfelvételnél célszerű inkább az összes adatból kiszámított szórásnégyzetet használni a variancia becslésére.

4-9. példa

Készítsünk egyedi érték + mozgó terjedelem-kártyát a 4-4. táblázat első és második oszlopában megadott adatokból!

4-4. táblázat

	x_i	$MR_i = \lvert x_i - x_{i-1} \rvert$
1	248.49	-
2	249.84	1.35
3	250.39	0.55
4	249.96	0.43
5	250.08	0.12
6	250.04	0.04
7	250.50	0.46
8	249.95	0.55
9	249.57	0.38
10	250.09	0.52
11	251.86	1.77
12	251.32	0.54
13	250.94	0.38
14	250.63	0.31
15	252.21	1.58
16	250.83	1.38
17	250.61	0.22
18	250.64	0.03
19	250.64	0.00
20	249.88	0.76

átlag	250.4235	0.5984

Az egyediérték-kártya paraméterei:

$$CL_x = \bar{x} = 250.4235,$$

$$\hat{\sigma}_x = \frac{\overline{MR}}{d_2} = \frac{0.5984}{1.128} = 0.5305,$$

$$UCL_x = \bar{x} + \frac{3\overline{MR}}{d_2} = 250.424 + \frac{3 \cdot 0.5984}{1.128} = 252.015,$$

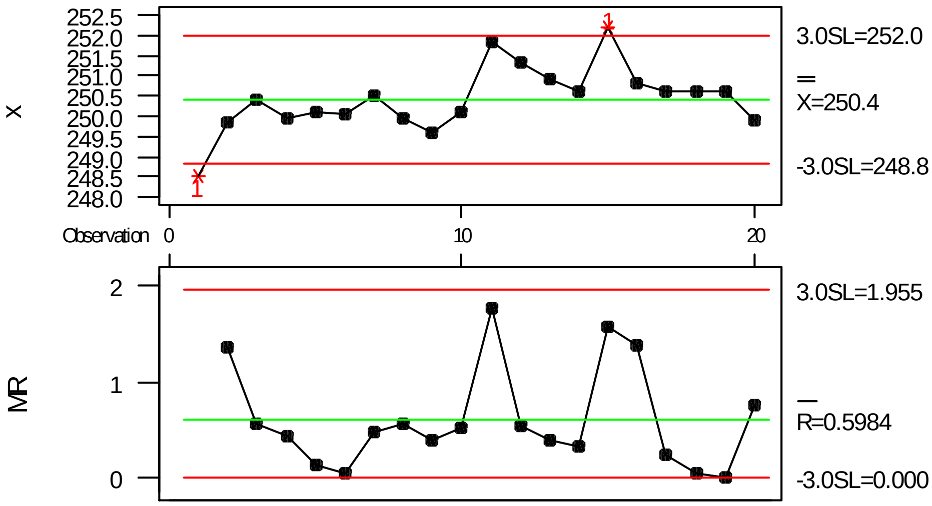
$$LCL_x = \bar{x} - \frac{3\overline{MR}}{d_2} = 250.424 - \frac{3 \cdot 0.5984}{1.128} = 248.832 .$$

A mozgó terjedelem-kártya paraméterei:

$$CL_{MR} = \overline{MR} = 0.5984,$$

$$UCL_{MR} = D_4 \overline{MR} = 3.267 \cdot 0.5984 = 1.955,$$

$$LCL_{MR} = D_3 \overline{MR} = 0 \cdot 0.5984 = 0 .$$



4-10. ábra. Egyedi érték + mozgó terjedelem-kártya, a MINITAB programmal

4.12. Az alkalmazott statisztikai feltételezések, és az eltérések következményei (normális eloszlás, függetlenség, homoszcedaszticitás)

Az ismertetett kártyák mindegyikére igaz, hogy használatuk csak akkor teljesen jogos, ha a mért *x* változó ingadozása normális eloszlású, állandó szórással (homoszcedaszticitás), és az egymás utáni mérések (minták) statisztikai értelemben függetlenek egymástól. A következőkben megvizsgáljuk a feltételek jelentőségét, nem-teljesülésük következményeit ill. ezek esetleges elhárítását.

Normális eloszlás

Nem mindig teljesül *x*-re pontosan. Sokszor a folyamat jellege miatt, pl. őrlésnél a keletkező szemcsék mérete lognormális eloszlású; hossz és tömeg nem lehet negatív, tehát bizonyos érték-tartományok nem fordulhatnak elő *x*-re. Az is előfordulhat, hogy maga a mérési módszer vagy az eredmény megadásának módja rontja el a normalitást, pl. hidrogénion-koncentráció helyett annak negatív logaritmusát (pH) használjuk.

Egyik következménye, hogy az elsőfajú hiba megengedett valószínűsége ill. a másodfajú hiba adott mintaelemszámhoz tartozó valószínűsége ténylegesen nem akkora, mint feltételezzük. Ez nem absztrakt statisztikai probléma, hanem nagyon is gyakorlati következményekkel jár. Az elsőfajú hiba ugyanis azt jelenti, hogy tévesen hisszük, hogy veszélyes hiba következett be, és fölöslegesen állítjuk le a gyártó berendezést, indokolatlanul határozunk el beállítást stb. A másodfajú hiba valószínűségének téves megítélése azt jelenti, hogy esetleg lényegesen több mintavétel után észlelünk csak egy eltolódást a beállításban, rosszul ítéltük meg a szükséges mintaelemszámot.

Minthogy nem mindig igaz, hogy *x* normális eloszlás szerint ingadozik, elfogadhatóbb a $\pm 3\sigma$ konvenció a szigorú számításhoz képest. Ha ugyanis nem tudjuk pontosan meghatározni az elsőfajú hiba valószínűségét, legalább egyszerű legyen az eljárás.

Ha ismerjük az *x* mért változó ingadozásának valószínűség-eloszlását (kellően sok pontból a hisztogram is legalább félkvantitatív képet ad), megfelelő módszerekkel számíthatunk alkalmas beavatkozási határokat.

Sok esetben az *x* változó transzformációjával elérhető, hogy normális eloszlású legyen (Box–Cox transzformáció, l. 12.1.3. alfejezetet), és akkor a transzformáltra (logaritmus, reciprok, négyzetgyök stb.) végezzük a vizsgálatokat.

Függetlenség

Sok esetben a folyamatban állandó a változás, ekkor az egymás utáni *x* értékek nemcsak véletlenszerűen különböznek egymástól. Pl. ha egy fúró kopik, a készített furatok átmérôje egyre csökken, de erre még a véletlen ingadozás is szuperponálódik. Hasonló a helyzet a szezonális (évszak, időjárás, műszak miatti) ingadozásokkal. A probléma speciális kezelést igényel, visszatérünk rá az 5. fejezetben.

Homoszcedaszticitás

Az alkalmazott statisztikai eljárások feltételezik, hogy az *x* változó varianciája konstans. Ha ez nem teljesül, hamisak (vagy gyengébbek) lesznek a következtetéseink, és nem is nagyon lehet megadni, hogy milyen mértékben. A gyakorlatban ez azt jelenti, hogy az elsőfajú és másodfajú hiba valószínűsége nem akkora, mint hisszük. Például, ha σ kétszeresére nő, a $\pm 3\sigma$ határ (amelyhez α=0.0027 tartozik) a valóságban csak $\pm 1.5\sigma$ lesz, ehhez α=0.134, tehát 1000 közül nem 3 esetben, hanem 134 esetben lesz téves riasztás. Sokszor segít itt is az adatok transzformációja (erre példát mutatunk a minősítéses ellenőrzésnél, ahol változó varianciájú Poisson-eloszlású változót alakítunk konstans varianciájú közel normális eloszlásúvá). Más esetben, ha a varianciák ismertek, az adatok standardizálhatók (skálázhatók). A módszert az 5.12. alfejezetben mutatjuk be részletesen.