- kiszámítjuk az átlagra vonatkozó beavatkozási határokat, ezek az átlag normális eloszlásának feltételezésére, a várható érték és a variancia ismeretére épülnek, és adott mintaelemszámhoz tartoznak
- becslést adunk a szóródási mérték (terjedelem, szórás, szórásnégyzet) várható értékére, és ezzel a becsléssel helyettesítjük is a várható értéket (ez lesz a másik kártya középvonala)
- kiszámítjuk a szóródási mérték varianciáját az x valószínűségi változó varianciájából, amelyet tudni vélünk
- kiszámítjuk a szóródási mértékre vonatkozó beavatkozási határokat, ezek a szóródási mérték varianciájának ismeretére, és egyes esetekben normális eloszlásának feltételezésére is épülnek, és adott mintaelemszámhoz tartoznak

Vegyük sorra az egyes lépéseket!

Átlag-kártya

Az *átlag várható értékének becslése* elég nagy (tipikusan 20-szor 5 elemű, azaz összesen 100 elemű) minta esetén is eltér a tényleges várható értéktől: $\bar{\bar{x}} = \mu \pm 3\sigma / \sqrt{100} = \mu \pm 0.3\sigma$. Például ha a várható érték 250, a variancia pedig 1.0, $\bar{\bar{x}} = 250 \pm 0.3$. Amikor az átlagot a várható érték helyébe írjuk a kártya vonalainak képletében, ekkora hibát véthetünk. Amikor azt kérdezzük (statisztikai próbával), hogy a vizsgált folyamat várható értéke megegyezik-e a korábbi állapotbeli várható értékkel, akkor a korábbi állapotbeli várható érték helyére annak becslését, egy véletlen hibával terhelt értéket írunk.

A föntebbi képletekben σ^2 az x valószínűségi változó (a mért jellemző) varianciája. A σ^2 becslésének jósága nagyon függhet a becslés módjától. A becslésre három lehetőséget ismertünk meg: a terjedelemből, a szórásból és a szórásnégyzetből. A valamelyik módon becsült varianciát használjuk az átlag-kártya beavatkozási határainak, a szóródási mértéket (pl. terjedelmet) ábrázoló kártya középvonalának és beavatkozási határainak számítására, a képletekben szereplő σ^2 helyén.

Egészen pontosan, az *átlag-kártyánál*, amikor az újonnan vett minta $\bar{x}_i$ átlagát a beavatkozási határokhoz mérjük, *u*-próbát végzünk, melynek próbastatisztikája:

$$u = \frac{\bar{x}_i - \mu}{\sigma / \sqrt{n}}.$$

Az ellenőrző kártya középvonala μ helyett annak becslése, a nevezőben szereplő σ helyett is annak becslése szerepel. Ez utóbbinak a hibáját is becsülhetjük. Például az előzetes adatfelvétel során kapott 20 ötelemű mintából egyesített szórásnégyzet szabadsági fokszáma 20·(5-1)=80. A szórásnégyzet $\chi^2\sigma^2 / \nu$ eloszlású, ez azt jelenti, hogy $\alpha = 0.0027$ ($\alpha / 2 = 0.00135$) esetén 1-0.0027=0.9973 valószínűséggel érvényes, hogy

$$\frac{\chi^2_{0.99865}(80)}{80} < \frac{s^2}{\sigma^2} < \frac{\chi^2_{0.00135}(80)}{80}, \quad \text{vagyis} \quad \frac{47.3}{80} < \frac{s^2}{\sigma^2} < \frac{123.4}{80},$$

azaz az arány 0.59 és 1.54 közé esik, ezek négyzetgyöke 0.77 ill. 1.24, vagyis a szórás eltérése σ-tól mindkét irányban kb. 24%. Hasonló hibakorlátokat kapunk az ötelemű mintákból végzett becslésre az átlagos terjedelemből és az átlagos szórásból is.

Helyesebb (statisztikai szempontból korrektebb) lenne figyelembe venni, hogy az aktuális mintából számított $\bar{x}_i$ átlagot egy másik átlaggal (az előzetes adatfelvételnél kapott $\bar{\bar{x}}$ -sal) hasonlítjuk össze, miközben a variancia sem ismert, csak a szórásnégyzet, vagyis tulajdonképpen kétmintás *t*-próbát kellene végezni. A 2.3.5.2. pontban megismert kétmintás *t*-próbánál két átlagértéket ($\bar{x}_1$, $\bar{x}_2$) hasonlítunk össze. A nullhipotézis általában az, hogy a két minta mögött álló sokaság várható értékei megegyeznek. A próbastatisztika:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}.$$

Az átlag-kártyánál $\bar{x}_1$ az előzetes adatfelvétel szerinti $\bar{\bar{x}}$ átlag, $\bar{x}_2$ a gyártásközi ellenőrzésnél egy *n* elemű minta átlaga. A nevezőben szereplő szórás négyzete:

$$s^2_{\bar{x}_1 - \bar{x}_2} = s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

A kétmintás *t*-próba klasszikus alkalmazásánál s^2 a két minta szórásnégyzetének egyesítésével jön létre. Eszerint itt az előzetes adatfelvételnél kapott szórásnégyzetet és a gyártásközi ellenőrzés aktuális mintájának szórásnégyzetét kellene egyesíteni. Az ellenőrző kártya filozófiája szerint azonban egyrészt a gyártásközi ellenőrzésnél használandó ellenőrző kártya paramétereit (középvonal, beavatkozási határok) az előzetes adatfelvételből kapjuk. Másrészt a 80 és 4 szabadsági fokú szórásnégyzetek egyesítésekor a nagy szabadsági fokszámú s^2 súlya 20-szorosa lenne az egy mintából kiszámolt szórásnégyzetének. Tehát nem indokolt az aktuális minta szórásnégyzetét a becslésbe bevonni.

Legyen ezért s^2 az előzetes adatfelvétel szerinti szórásnégyzet, célszerűen az ottani (a példa szerint 5 elemű) minták 4 szabadsági fokú szórásnégyzeteiből egyesítve, vagyis ν=20·4=80.

A *t*-eloszlás α elsőfajú hiba-valószínűséghez tartozó kritikus értékei:

$$P\left(-t_{\alpha/2} < t < t_{\alpha/2}\right) = 1 - \alpha.$$

Ebből a második minta (a gyártásközi ellenőrzés aktuális mintája) átlagértékére az elfogadási tartomány

$$P\left(\bar{x}_1 - t_{\alpha/2}s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \bar{x}_2 < \bar{x}_1 + t_{\alpha/2}s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = 1 - \alpha,$$

vagyis az így előálló *t*-kártya középvonala

$$CL_t = \bar{x}_1 = \bar{\bar{x}},$$

az előzetes adatfelvétel szerinti átlag. Beavatkozási határai:

$$UCL_t = \bar{x}_1 + t_{\alpha/2} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\,,$$

$$LCL_t = \bar{x}_1 - t_{\alpha/2} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\,.$$

4-5. példa

Legyen az elsőfajú hiba megengedett α valószínűsége 0.0027. Számítsuk ki a t-kártya (módosított átlag-kártya) paramétereit!

Az α=0.0027 valószínűséghez tartozó t-táblázat nem áll rendelkezésre, de számítógéppel $t_{\alpha/2}$ kiszámítható, pl. a Statistica programmal ν=80 szabadsági fokhoz $t_{\alpha/2}$=3.096; melynek helyettesítésével:

$$CL_t = \bar{\bar{x}} = 249.955\,,$$

$$UCL_t = 249.955 + 3.096 \cdot \sqrt{0.9643}\sqrt{\frac{1}{100} + \frac{1}{5}} = 251.348\,,$$

$$LCL_t = 249.955 - 3.096 \cdot \sqrt{0.9643}\sqrt{\frac{1}{100} + \frac{1}{5}} = 248.562\,.$$

Ezek a beavatkozási határok alig térnek el a hagyományos átlag-kártyára a 4-2. példában kapott 251.301 és 249.609 értékektől. Az utóbbi beavatkozási határok úgy adódtak, hogy a variancia becslésére a terjedelmet használtuk.

4-6. példa

Adjunk pontosabb becslést a hagyományos átlag-kártya elsőfajú hibájának valószínűségére, méghozzá külön a fölső és külön az alsó határ meghaladására, ha a varianciát a szórásnégyzetből becsüljük (vagyis az átlag-szórásnégyzet kártyakombinációval dolgozunk)!
Az átlag-kártya beavatkozási határai, az első egyenlőség a kétmintás t-próba, a második a kártya-technika jelöléseivel (utóbbiban $\bar{\bar{x}}$ és $\overline{s^2}$ az előzetes adatfelvételhez, n az aktuális mintához tartozik):

$$UCL_{\bar{x}} = \bar{x}_1 + 3s/\sqrt{n_2} = \bar{\bar{x}} + 3\frac{\sqrt{\overline{s^2}}}{\sqrt{n}}\,,$$

$$LCL_{\bar{x}} = \bar{x}_1 - 3s/\sqrt{n_2} = \bar{\bar{x}} - 3\frac{\sqrt{\overline{s^2}}}{\sqrt{n}}\,.$$

Fölhasználjuk, hogy t-eloszlása van a következő kifejezésnek:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \mu_1 + \mu_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

$$\alpha_{fölsô} = P\left(\bar{x}_2 > UCL \middle| \mu_1 = \mu_2\right) = P\left(\bar{x}_1 + ts\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} > \bar{x}_1 + 3s\sqrt{\frac{1}{n_2}}\right) =$$

$$= P\left(t > \frac{3\sqrt{\frac{1}{n_2}}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right).$$

Behelyettesítve n_1-re 100-at és n_2-re 5-öt:

$$\alpha_{fölsô} = P(t > 2.9277) = 1 - F(2.9277) = 0.002197.$$

Hasonlóan

$$\alpha_{alsó} = P(t < -2.9277) = F(-2.9277) = 0.002197.$$

Ez a külön a fölső, külön az alsó határra deklarált 0.00135-tôl nem nagyon tér el.

A meglepően jó egyezés oka az, hogy az előzetes adatfelvételből (a meglehetősen nagy mintából) meglehetősen pontosan becsülhető μ és σ^2. Másképpen az átlag-kártyánál a μ helyére írt $\bar{\bar{x}}$ és a σ helyén szereplő $\bar{R}/d_2$, $\bar{s}/c_4$ ill. $\sqrt{\bar{s^2}}$ elég jó közelítések. Pontosabban fogalmazva, egyrészt μ helyett $\bar{\bar{x}}$ értékét használjuk a próbában, és ez jó közelítés; másrészt az *u*-próba helyett *t*-próbát használunk egy 80 szabadsági fokú szórásnégyzettel, de ilyen nagy szabadsági foknál a *t*-eloszlás és az *u*-eloszlás között alig van különbség.

A kétmintás *t*-próba, ugyanúgy, mint a ténylegesen használt *u*-próba, feltételezi az átlagok normális eloszlását. Ez a centrális határeloszlás tétele értelmében jogos, már 4-elemű mintánál is, akkor is, ha a mért *x* változó nem követ normális eloszlást. Box, Hunter és Hunter (1978) szerint ebben az esetben azzal sem követünk el jelentős hibát, hogy a próbában használt szórásnégyzet esetleg nem normális eloszlású *x* valószínűségi változó szórásnégyzete.

A szóródási mérték kártyája

A kétmintás *t*-próba elvégzésének feltétele a két csoport (az előzetes adatfelvételnél nyert adatok és a gyártásközi ellenőrzésnél vett minta) varianciájának azonossága. Ezt a szóródási mérték kártyájával (*R*, *s*, s^2) ellenőrizzük (ezeknél éppen a varianciák azonossága a nullhipotézis).

Itt is az első fajta hibát azzal követjük el, hogy a középvonal képletében σ helyett annak becslését használjuk. Amikor azt kérdezzük (statisztikai próbával), hogy a vizsgált folyamat varianciája megegyezik-e a korábbi állapotbeli varianciával, akkor a korábbi állapotbeli variancia helyére annak becslését, egy véletlen hibával terhelt értéket írunk. Ezzel a szórásnégyzet-kártyán a középvonal hibája az elôbbi számítás szerint 41% lehet lefelé és 54% fölfelé. A szórás-kártyánál a középvonal hibája közelítőleg 24% lehet mindkét irányban. A terjedelem-kártya középvonala, mivel R arányos σ-val, ugyancsak 24% hibát tartalmazhat.

A másik fajta hibát a beavatkozási határok számításánál követhetjük el, ez a háromféle kártyánál erősen különböző. Mindhárom kártyánál közös elhanyagolás, hogy a σ helyett annak becslését helyettesítjük a beavatkozási határok képleteibe is.

A *szórásnégyzet-kártyánál* azzal a feltételezéssel élünk, hogy x normális eloszlású, mert csak ebben az esetben lesz a szórásnégyzet eloszlása $\chi^2\sigma^2/\nu$. A határok számítása ekkor a χ^2-eloszlás táblázata segítségével történik, vagyis χ^2-próbát végzünk, az i-edik új minta s_i^2 szórásnégyzetét vetjük össze a $\sigma^2 \approx \overline{s^2}$ helyettesítéssel közelített varianciával. Helyesebb (statisztikai szempontból korrektebb) lenne figyelembe venni, hogy az s_i^2 szórásnégyzetet egy másik szórásnégyzettel ($\overline{s^2}$ -gal) hasonlítjuk össze, vagyis F-próbát kellene végezni.
Így egy F-kártyát kapnánk, melyen az s_2^2/s_1^2 értékeket kellene ábrázolni, paraméterei:

$CL = 1$,

$UCL_F = F_{fölsô} = F_{0.001}$,

$LCL_F = F_{alsó} = F_{0.999}$.

Ezt a kártyát kényelmetlen lenne használni, minden mintához ki kellene számítani az s_2^2/s_1^2 arányt. Kellemesebb a módosított változat, amelyen az aktuális mintához számított s_2^2 értékeket ábrázoljuk. A vonalak helyét úgy kapjuk, hogy a megfelelő F értéket az előzetes adatfelvételnél kapott s_1^2 -tel szorozzuk.

A kapott Fs^2-kártya paraméterei:

$CL = s_1^2$,

$UCL_{Fs^2} = s_1^2 F_{0.001}$,

$LCL_{Fs^2} = s_1^2 F_{0.999}$.

4-7. példa

Számítsuk ki a módosított s^2-kártya (Fs^2-kártya) paramétereit a 4-4. példában szereplő 20·5 elemű mintából álló előzetes adatfelvételre és 5 elemű mintákkal végzett gyártásközi ellenőrzésre:

$F_{0.999}(4{,}80) = 5.123$.

$F_{0.001}(4{,}80) = 0.0224$.

A 4-4. példában s_1^2 értéke 0.9643, vagyis

$$UCL_{Fs^2} = s_1^2 F_{0.001} = 0.9643 \cdot 5.123 = 4.94,$$

$$LCL_{Fs^2} = s_1^2 F_{0.999} = 0.9643 \cdot 0.0224 = 0.0216.$$

Ez vetendő össze a 4-4. példa szerinti *UCL*=4.452, *LCL*=0.0219 értékekkel, az eltérés elhanyagolható.

4-8. példa

Számítsuk ki az s^2-kártyára az elsőfajú hiba valószínűségét pontos számítással, vagyis nem hanyagolva el a σ^2 és s^2 közötti különbséget!
Az s^2-kártya beavatkozási határai:

$$UCL_{s^2} = \frac{\chi^2_{0.001} s_1^2}{\nu} = s_1^2 \cdot \frac{18.47}{4},$$

$$LCL_{s^2} = \frac{\chi^2_{0.999} s_1^2}{\nu} = s_1^2 \cdot \frac{0.0908}{4}.$$

Fölhasználva, hogy

$$F = \frac{s_2^2}{s_1^2},$$

$$\alpha_{fölsô} = P\left(s_2^2 > UCL_{s^2} \middle| \sigma_1^2 = \sigma_2^2\right) = P\left(s_2^2 > s_1^2 \cdot \frac{18.47}{4}\right) = P\left(Fs_1^2 > s_1^2 \cdot \frac{18.47}{4}\right) =$$

$$= P\left(F > \frac{18.47}{4}\right).$$

Ilyen táblázatunk nincs, de a Statistica programmal kiszámítva $\alpha_{fölsõ} = 0.0021$, a deklarált 0.001 érték kétszerese

$\alpha_{alsó} = P\left(s_2^2 < LCL_{s^2} \middle| \sigma_1^2 = \sigma_2^2\right) = P\left(F < \frac{0.0904}{4}\right) = 0.001015$, a deklarált értékkel megegyezik.

Az *F*-próba, ugyanúgy, mint a ténylegesen használt χ^2-próba, feltételezi az *x* mért jellemző normális eloszlását, és érzékeny az attól való eltérésre.

A *terjedelem*- és a *szórás-kártyánál* lényegesen rosszabb a helyzet. A $\pm 3\sigma$ konvenció alkalmazása azt jelenti, hogy a terjedelmet és a szórást is normális eloszlásúnak tételezzük föl. Ez még akkor sem jogos, ha a mért *x* változó normális eloszlás szerint ingadozik, mert mind a terjedelem, mind a szórás sűrűségfüggvénye aszimmetrikus. Emiatt az első- és másodfajú hibáról kialakuló képünk nem teljesen megfelelő.

Ryan (1989), (p. 435, Table F; az adatok származása: Harter, 1960) szerint ötelemű minta *terjedelmének* elfogadási határai, ha az elsőfajú hiba megengedett valószínűségét 0.002-re választjuk, 0.367σ és 5.484σ. A $\pm 3\sigma$ számításmód szerint a határokat a $(d_2 \pm 3d_3)\sigma = (2.326 \pm 3 \cdot 0.864)\sigma$ összefüggéssel kell számolni, és így a jócskán különböző -0.266σ ill. 4.92σ adódik. (Az α=0.002 nem pontosan egyezik meg a $\pm 3\sigma$-nak megfelelő 0.0027-del, de ilyen α értékhez táblázat nem áll rendelkezésre.)

A *szórás*-kártyánál sem kell feltétlenül a $\pm 3\sigma$ konvenció alapján számolnunk a beavatkozási határokat, hanem közelítésként használhatjuk a χ^2-eloszlást is:

$$UCL_s = \sigma\sqrt{\frac{\chi^2_{fölsô}}{\nu}} = \frac{\bar{s}}{c_4}\sqrt{\frac{\chi^2_{fölsô}}{\nu}}\,,$$

$$LCL_s = \sigma\sqrt{\frac{\chi^2_{alsó}}{\nu}} = \frac{\bar{s}}{c_4}\sqrt{\frac{\chi^2_{alsó}}{\nu}}\,.$$

Ekkor megköthetjük az elsőfajú hiba megengedett valószínűségét, és ahhoz vesszük a táblázatból a χ^2 határokat.

A 4-1. példa adataival, ha az elsőfajú hiba valószínűségére 0.002-t választunk (ez közel áll a $\pm 3\sigma$ konvenció alkalmazásakor föltételezett 0.0027-hez), a függelék II. táblázatából $\chi^2_{0.999}(4) = 0.0908$; $\chi^2_{0.001}(4) = 18.47$; a V. táblázatból c_4=0.94, így a kártya pontosan számolt paraméterei:

$$UCL_s = \frac{\bar{s}}{c_4}\sqrt{\frac{\chi^2_{0.999}}{\nu}} = \frac{0.9181}{0.94}\sqrt{\frac{18.47}{4}} = 2.099\,,$$

$$LCL_s = \frac{\bar{s}}{c_4}\sqrt{\frac{\chi^2_{0..001}}{\nu}} = \frac{0.9181}{0.94}\sqrt{\frac{0.0908}{4}} = 0.147\,.$$

Vessük ezeket össze a $\pm 3\sigma$ konvencióval kapott paraméterekkel: ott az alsó beavatkozási határra 0, a fölsôre 1.918 adódott.

A következtetéseket összefoglalva:

- Az *átlag-kártyának* lehetséges egy olyan módosított változatát használni, amelynél a beavatkozási határokat nem az *u*-próba, hanem a kétmintás *t*-próba összefüggései szerint adják meg. Mivel azonban az előzetes adatfelvétel nagy mintát eredményez, a közelítésképpen alkalmazott *u*-próbánál használt becsült variancia igen jó közelítés. A variancia becslését *R*-kártyára is alapozhatjuk, mert egyenként kis elemszámú mintákról lévén szó, a terjedelem is jó hatásfokú statisztika., A statisztikai szempontból mégoly megalapozott eljárás (a kétmintás *t*-próba) sem ad más eredményt, mint a hagyományos átlag-terjedelem-kártya.

- Az *ingadozás mértékére* vonatkozó hipotézis vizsgálatát a gyártásközi ellenőrzés szakaszában célszerű s^2-kártyára, és így a χ^2- vagy az F-eloszlásra alapozni. Az így kapható beavatkozási határok sokkal pontosabbak, mert nem tételezik föl a szóródási mérték (terjedelem vagy szórás) normális eloszlását: Az s^2-kártya előkészítése (a kártya paramétereinek kiszámítása) az előzetes adatfelvétel stádiumában a gyártásközi ellenőrzést megelőző feladat, amelyet célszerű mindenképpen számítógéppel segíteni, így a kicsivel bonyolultabb számolás nem jelent akadályt.

4.5. A mintaelemszám változása vagy változtatása

Különböző mintaelemszám az előzetes adatfelvételnél és a gyártásközi ellenőrzésnél

Ha az előzetes adatfelvételnél a mintaelemszám nem ugyanakkora, mint az aktuális ellenőrzésnél, óvatosnak kell lennünk. Ha például az átlag-terjedelem kártya-kombinációval dolgozunk a gyártásközi ellenőrzésnél és az előzetes adatfelvételnél egyaránt, és így az átlag-kártya beavatkozási határaira az $\bar{\bar{x}} \pm \frac{3\bar{R}}{d_2\sqrt{n}}$ képletet használjuk, d_2-nél az előzetes adatfelvétel mintaelemszámát, n-re az ellenőrzés mintaelemszámát kell venni, az $\bar{\bar{x}} \pm A_2\bar{R}$ képlet nem alkalmazható. A terjedelem-kártyán a gyártásközi ellenőrzéskor a középvonal nem ugyanott lesz, mint az előzetes adatfelvételkor. Ugyanis $\bar{R}$ a kártyán a szóródás paramétere, csak figyelembe veendő, hogy az eloszlás valódi paramétere σ^2, a terjedelem várható értéke a mintaelemszámtól függ. Eszerint azzal a mintaelemszámmal, amely a paraméterek becslésére szolgál, a terjedelemből először ki kell számítani a becsült varianciát ($\sigma \approx \bar{R}/d_2$, ahol d_2 n függvénye), majd abból a másik mintaelemszámmal a terjedelem várható értékét ($R \approx d_2/\sigma$). Mindezek az összefüggések a 4-3. táblázatban találhatók. A táblázatból azt is kiolvashatjuk, hogy mi a teendő, ha a variancia becslésére az előzetes adatfelvétel eredményeiből más módszert alkalmazunk, mint a gyártásközi ellenőrzésnél a szóródás vizsgálatára.

Változó mintaelemszám

Előfordul, hogy a mintaelemszám az eddigiekben feltételezettekkel ellentétben nem állandó. Lehetséges, hogy a mintavételi eljárás okozza, hogy különböző elemszámú minták állnak elő. Ilyenkor nem $\bar{x}$ - R, hanem $\bar{x}$ - s kártya-párosítást szokás használni, mert mint a 4-3. táblázatból látható, a terjedelem-kártya középvonalának helyzete megváltozik (ugrál), ha a mintaelemszám változik, és ez kényelmetlen; a szórás-kártya középvonala jóval kevésbé változik.

Ha a mintaelemszám nem állandó, az átlagot az egyes minták n_i elemszáma szerint súlyozva kell képezni:

$$\bar{\bar{x}} = \frac{\sum_i n_i x_i}{\sum_i n_i}.$$