

Műszaki és gazdasági adatok elemzése

Dr. Halász Gábor előadásai alapján készítette Varga Roxána

Az ábrákat Till Sára rajzolta

Frissített verzió

2018. április 19.

1. előadás

Ezen az előadáson megismerkedünk az átlag, a medián, a tapasztalati szórás és a korrigált tapasztalati szórás fogalmával. Szó esik a box-plot-ról, és az abból kiolvasható információkról, valamint megemlítődik a tapasztalati sűrűségfüggvény.

A fenti alapfogalmakat a következő egyszerű példán keresztül fogjuk bevezetni.

1. *Példa:* Tekintsük azon hallgatókat akik elégtelennél jobb jegyet kaptak matematika A2-ből. Az érdemjegyeket az alábbi táblázat tartalmazza.

	érdemjegy	gyakoriság (ν_i)	relatív gyakoriság ($\frac{\nu_i}{n}$)
η_5	5	20	$\frac{20}{500} = 0,04$
η_4	4	33	$\frac{33}{500} = 0,066$
η_3	3	150	$\frac{150}{500} = 0,3$
η_2	2	297	$\frac{297}{500} = 0,594$

A táblázat első oszlopában η_i jelöli az i . ($i = 2 \dots 5$) érdemjegyet, ezeket nevezzük a **lehetséges kimenetek**eknek. A táblázatból látható, hogy összesen 500 hallgató kapott elégtelennél jobb jegyet, ezt a számot nevezzük a **minta méret**ének, vagy nagyságának, ezt legtöbbször n vagy N jelöli. Jelölje ξ_j , ($j = 1, \dots, 500$) a j . hallgató érdemjegyét. A jegyek **átlagát** a következő képlettel szokás számítani:

$$\bar{\xi} = \frac{1}{500} \sum_{j=1}^{500} \xi_j. \quad (1)$$

Ezt átírhatjuk a következő módon:

$$\bar{\xi} = \frac{1}{500} \sum_{i=2}^5 \nu_i \eta_i = \sum_{i=2}^5 \frac{\nu_i}{500} \eta_i. \quad (2)$$

ahol ν_i a i . osztályzat **gyakorisága** és η_i az i . osztályzat értéke. Az egyenlőség jobb oldalán megjelenő $\frac{\nu_i}{500}$ értéket az i . osztályzathoz tartozó **relatív gyakoriságnak** nevezzük. Tehát a relatív

gyakoriság a kedvező és az összes esetek számának hányadosa. Megjegyezzük, hogy a relatív gyakoriságok összege mindig 1, ezt az olvasó maga is beláthatja. A (2)-es egyenlőség jobb oldalát a minta **súlyozott átlagának** nevezzük.

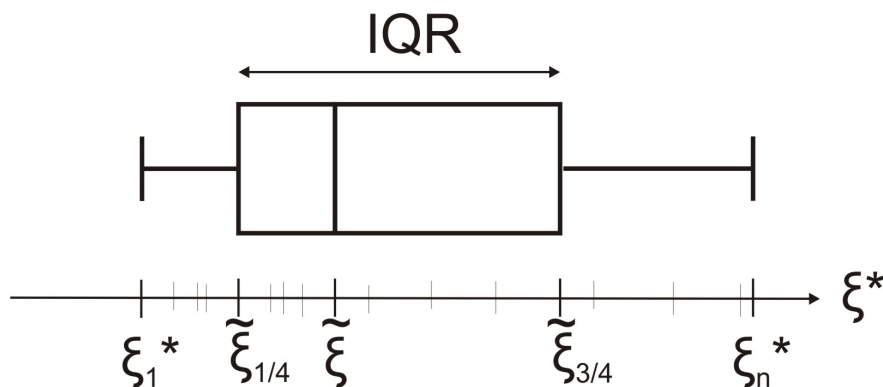
A példánkban szereplő megfigyelések **diszkrét** értékűek, hiszen az osztályzatok mindenkor csak is egész értéket vehetnek fel. **Folytonos** értékekről például akkor beszélhetnénk, ha a hallgatók testmagasságát vizsgálnánk.

1.1. A box-plot

Az átlag még nem mond el sok mindent a mért adatainkról. Ahhoz, hogy minél több információt nyerjünk a méréseinkből további fogalmakra van szükségünk. Az első ilyen fogalom a medián. A **medián** az a mintaelem amelyik a sorba rendezett mintában éppen középen van, tehát ugyanannyi tőle nagyobb elem van, mint kisebb. Ezt formailag a következőképpen fogalmazhatjuk meg. Jelölje $\xi_1^*, \xi_2^*, \dots, \xi_n^*$, a sorba rendezett megfigyeléseket (mostantól a * felső index mindig sorba rendezett mintát fog jelenteni). Ha n páratlan szám akkor a középső elem indexe $m = \frac{n+1}{2}$, és így a medián $\tilde{\xi} = \xi_{\frac{n+1}{2}}^*$. Ha pedig n páros szám, akkor nincs középső elem, ezért a két középső elem átlaga lesz a medián: $\tilde{\xi} = \frac{\xi_{\frac{n}{2}}^* + \xi_{\frac{n}{2}+1}^*}{2}$ lesz.

Vizsgálható továbbá az is, hogy a medián alatti illetve feletti mintaelemekről mi mondható el. Vethető ezen elemeknek is a mediánja, amik az eredeti adatainkat 'negyedelik'. Erre vezetjük be a **kvartilisek** fogalmát. **Alsó kvartilis** (vagy 1. kvartilis) alatt azt a mintaelemet értjük, amelyik a medián alatt elhelyezkedő mintaelemek mediánja, ezt $\tilde{\xi}_{\frac{1}{4}}$ jelöli. Hasonlóan **felső kvartilisen** (vagy 3. kvartilisen) azt a mintaelemet értjük, amelyik a medián felett elhelyezkedő mintaelemek mediánja, ezt $\tilde{\xi}_{\frac{3}{4}}$ jelöli. Értelmezhető a 0. illetve a 4. kvartilis is, ezek a rendezett mintánkban rendre a legkisebb (ξ_1^*) illetve a legnagyobb (ξ_n^*) elemek. Vagyis a k . kvartilis alatt éppen a mintában szereplő elemek $\frac{k}{4}$ -ed része található és felette a mintában szereplő elemek $1 - \frac{k}{4}$ -ed rész. Nyilvánvaló, hogy a 2. kvartilis a medián.

A kvartilisek felhasználásával tudjuk felrajzolni a mintánkhoz tartozó **box-plot**ot. Ezt az 1. ábra szemlélteti.



1. ábra. Box.plot

A box-plotot vizsgálva további fogalmak bevezetésére van szükség. A középső tartomány méretét **interkvartilis tartománynak** nevezzük, amit a továbbiakban IQR fog jelölni. Tehát

$$IQR = \tilde{\xi}_{\frac{3}{4}} - \tilde{\xi}_{\frac{1}{4}}. \quad (3)$$

Lefelé kiugró értékeken azon ξ_i mintaelemeket értjük, amelyekre fennáll:

$$\xi_i < \tilde{\xi}_{\frac{1}{4}} - 1,5 * IQR. \quad (4)$$

Lefelé extrém kiugró értékek pedig azon ξ_i mintaelemek, amelyekre

$$\xi_i < \tilde{\xi}_{\frac{1}{4}} - 3 * IQR. \quad (5)$$

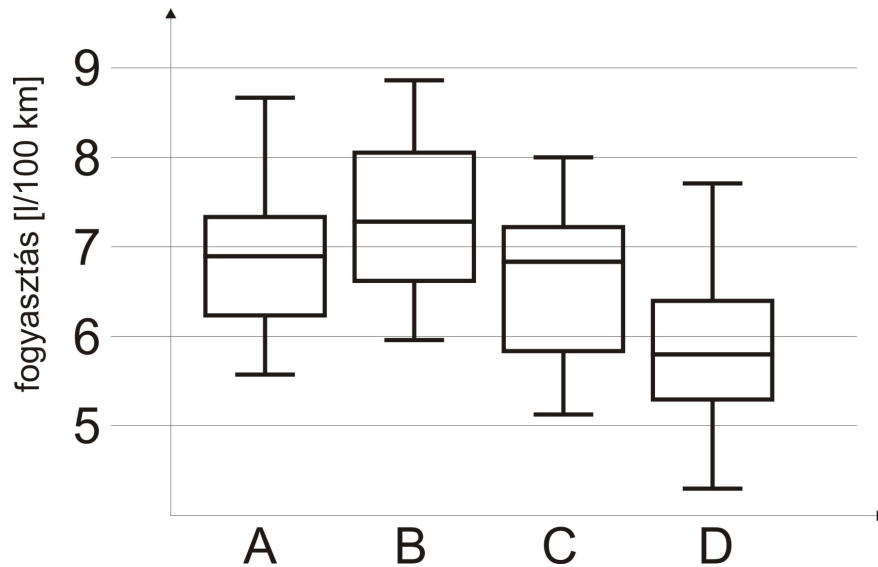
Hasonlóan a **felfelé kiugró** értékek azon ξ_i mintaelemek, amelyekre fennáll:

$$\xi_i > \tilde{\xi}_{\frac{3}{4}} + 1,5 * IQR. \quad (6)$$

Felfelé extrém kiugró értékek pedig azon ξ_i mintaelemek, amelyekre

$$\xi_i > \tilde{\xi}_{\frac{3}{4}} + 3 * IQR. \quad (7)$$

2. *Példa:* Tekintsük 4 különböző gépkocsi típus városi fogyasztását. Minden típusból több autó fogyasztását feljegyeztük és az adatokból box-plotokat készítettünk, melyeket a 2. ábrán szemléltetünk. Hasonlítsuk össze a fogyasztásokat vizuálisan! Jól látszik, hogy az A, B és C jelű kocsik fogyasztása nem tér el nagyon egymástól, míg a D kocsié lényegesen (szignifikánsan) különbözik a többitől.



2. ábra. Az autók fogyasztását szemléltető box-plot ábra.

Érdeemes lehet azt is megvizsgálni, hogy a mérési adataink az átlaghoz képest mennyire ingadoznak. Ezt a pillanatnyi megfigyelés és az átlag különbsége fogja megadni, vagyis $\xi_i - \bar{\xi}$ értéke. Könnyen meggondolható, hogy az ingadozások átlaga mindig 0. Tehát

$$\frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi}) = 0, \quad (8)$$

vagyis ez nem mond semmit a mintánkról. Ezért az átlagtól való eltérés abszolút értékét vagy pedig az átlagtól vett négyzetes eltérést szokták vizsgálni:

$$\frac{1}{n} \sum_{i=1}^n |\xi_i - \bar{\xi}| \quad \text{vagy} \quad \frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi})^2.$$

Az előbbi matematikai szempontból nem praktikus, hiszen az abszolútérték-függvény nem differenciálható folytonosan a 0-ban. Ezért az átlagos négyzetes eltérést szokás használni, melyet **tapasztalati szórásnégyzet**nek nevezünk és s^2 -el jelölünk, vagyis:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi})^2, \quad (9)$$

ebből a tapasztalati szórás ennek pozitív négyzetgyöke:

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi})^2}. \quad (10)$$

Itt az n darab tag nem független egymástól, hiszen $\bar{\xi}$ éppen a ξ_i megfigyelések átlaga. A statisztikában független adatokkal célszerű dolgozni, így bevezetjük a **korrigált tapasztalati szórásnégyzet** fogalmát, melyben az $1/n$ együttható helyett $1/(n-1)$ szerepel (mert bármely $n-1$ darab tag már lineárisan független):

$$(s^*)^2 = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2. \quad (11)$$

és a korrigált tapasztalati szórást:

$$s^* = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2}. \quad (12)$$

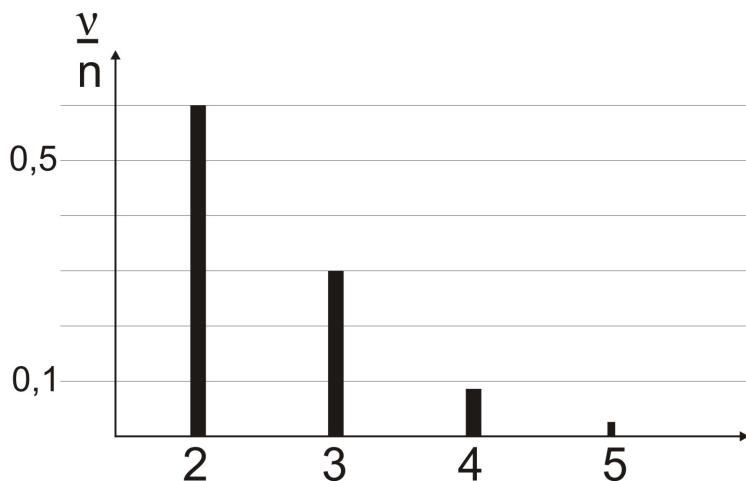
A korrigált tapasztalati szórásnak elméleti jelentősége is van, amelyet a félév során, később ismerünk meg, amikor az $n \rightarrow \infty$ esetet vizsgáljuk.

Vegyük észre, hogy az átlag, a medián és a szórás mértékegysége ugyanaz, mint a megfigyeléseké.

1.2. A tapasztalati sűrűségfüggvény

A **tapasztalati sűrűségfüggvény**, más néven **hisztogram** segítségével vizsgálhatjuk, hogy hogyan oszlanak el a megfigyelt értékeink. A sűrűségfüggvény tárgyalásakor megkülönböztetjük a diszkrét illetve folytonos eseteket.

Diszkrét esetben a koordináta-rendszer X tengelyén a megfigyelt értékeket ábrázoljuk, míg az Y tengelyen a relatív gyakoriságokat. Mennyiségi változó esetén az X tengelyen a skálázás természetes módon növekvő sorrendben történik, míg minőségi változó esetén (pl. szín) a skálázási sorrend lényegtelen, csak az oszlopok magassága számít. Az 1. példához tartozó hisztogramot a 3. ábra mutatja.



3. ábra. Az 1. példában szereplő osztályzatok alapján felrajzolt hisztogram.

2. előadás

Ezen az előadáson folytatjuk a tapasztalati sűrűségfüggvény tárgyalását folytonos esetben és bevezetjük az eloszlásfüggvény fogalmát. Megismerkedünk a tapasztalati korrelációs együtthatóval, valamint annak egy speciális esetével, a rangkorrelációval.

Folytonos változó esetén (pl. testmagasság) a tapasztalati sűrűségfüggvény felrajzolásához a sorozatból kiválasztjuk a legkisebb és legnagyobb értékeket, és a köztük levő tartományt intervallumokra osztjuk. Az egyes intervallumok fölé úgy rajzolunk téglalapokat, hogy azok területe megegyezzen az adott intervallumba eső megfigyelések relatív gyakoriságával. Ezeket precízen a következőképpen tudjuk megfogalmazni. Legyenek a megfigyeléseink $\xi_1, \xi_2, \dots, \xi_n$, a legkisebb elem legyen $x_0 = \min \xi_i$, a legnagyobb pedig $x_{IS} = \max \xi_i$, ahol IS az intervallumok száma. Az i . intervallumra bevezetjük a következő jelöléseket:

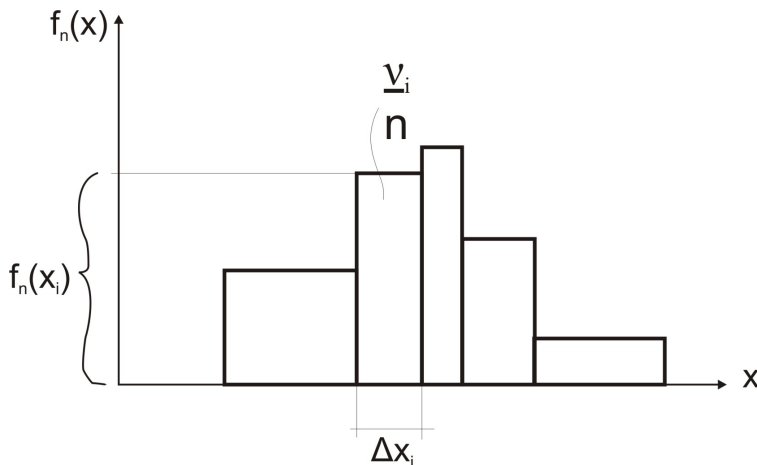
- Δx_i : az i . intervallum szélessége,
- ν_i : az i . intervallumba eső megfigyelések száma, vagyis az i . intervallum gyakorisága,
- $f_i = f_n(x_i)$ n elemű minta esetén az x_i helyen felvett téglalap magasság.

Mivel az i . intervallumba esés relatív gyakorisága megegyezik az intervallum fölé emelt téglalap területével:

$$\frac{\nu_i}{n} = \Delta x_i f_n(x_i), \quad (13)$$

ebből pedig kiszámolható a téglalap magassága:

$$f_n(x_i) = \frac{v_i}{n\Delta x_i}. \quad (14)$$



4. ábra. Folytonos változó esetén a tapasztalati sűrűségfüggvény.

Az intervallumok számának (IS) meghatározására bevett módszer, hogy:

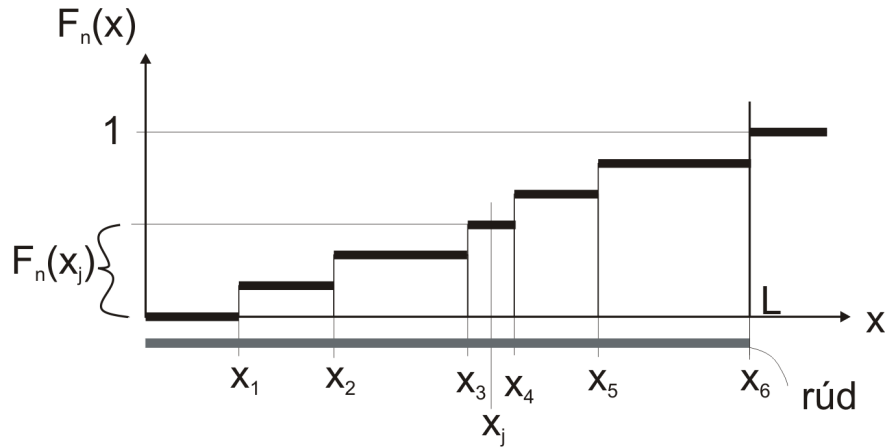
- $n < 100$ esetén $IS \approx \sqrt{n}$,
- $n \geq 100$ esetén $IS \approx \log_2 n + 1$.

Az intervallumok beosztását pedig úgy érdemes meghatározni, hogy nagyjából ugyanannyi megfigyelés essen minden intervallumba, vagyis a gyakoriságok közel azonosak legyenek.

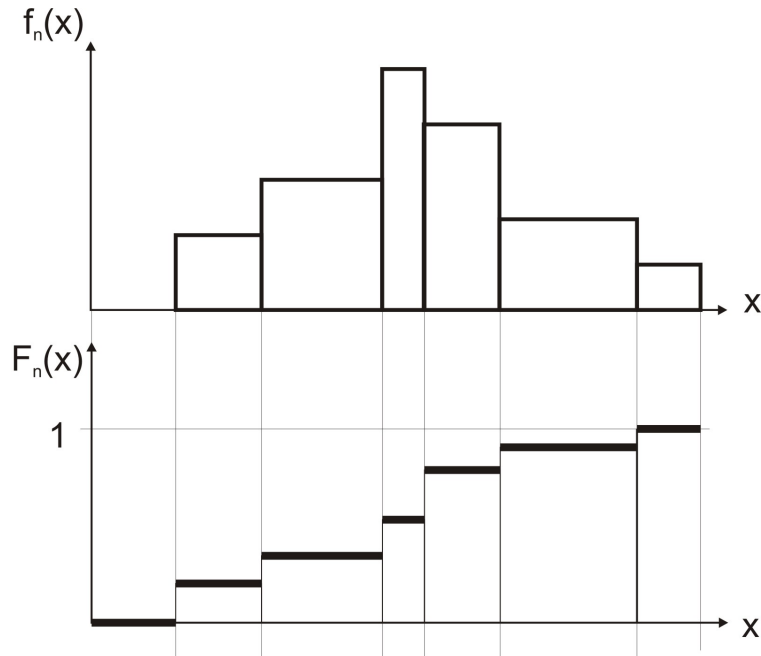
2.1. A tapasztalati eloszlásfüggvény

Egy másik módszer a megfigyelések ábrázolására a **tapasztalati eloszlásfüggvény**. Ennek ismeretése előtt végezzünk el egy kísérletet. Tekintsünk egy homogén, prizmatikus, egyenes rudat. Az egyik végét rögzítsük le a derékszögű koordináta-rendszer origójába, a rúd az X tengellyel legyen párhuzamos. Vízszintes irányban kezdjük el húzni a rudat, az előbb-utóbb egy véletlen helyen el fog szakadni. Ismételjük meg ezt a kísérletet többször is, és jegyezzük az X tengelyen a szakadási helyeket, míg az Y tengelyen az eltört rudak számának relatív gyakoriságát. A szakadási helyek növekvő sorozata legyen x_1, x_2, \dots, x_n . Ekkor az 5. ábrán látható lépcsős függvényt ábrázolhatjuk, amelyet F_n jelöl. $F_n(x_j)$ magasság éppen az olyan megfigyelések relatív gyakoriságát adja meg, amelyek x_j alatt szakadtak el.

A fenti példa alapján megfogalmazható, hogy mi is az a tapasztalati eloszlásfüggvény. A tapasztalati eloszlásfüggvény egy olyan függvény, amely megmondja annak az eseménynek a relatív gyakoriságát, amely kisebb, mint a helyettesítési érték. n elemű minta esetén a függvényt F_n -el jelöljük.



5. ábra. Homogén, prizmatikus, egyenes rúd szakadási helyeit szemléltető tapasztalati eloszlásfüggvény.



6. ábra. A tapasztalati sűrűség- és eloszlásfüggvények összehasonlítása.

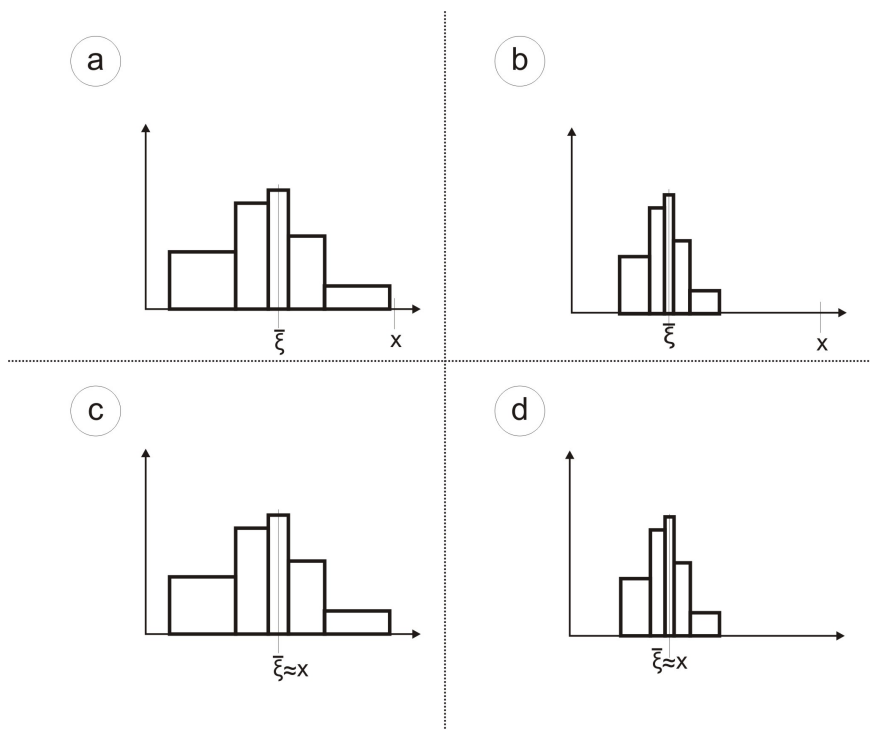
A tapasztalati sűrűségfüggvény és a tapasztalati eloszlásfüggvény között a következő kapcsolat áll fenn. Rajzoljuk fel egymás alá egy adott megfigyeléssorozathoz tartozó tapasztalati sűrűségfüggvényt és tapasztalati eloszlásfüggvényt, egyforma tengelybeosztásokkal (6. ábra). Ekkor a következő összefüggés figyelhető meg. A j . intervallum helyén az eloszlásfüggvény magassága:

$$F_n(x_j) = \frac{\nu_1 + \dots + \nu_j}{n}. \quad (15)$$

A bal oldalt másképpen írva:

$$F_n(x_j) = f_n(x_1)\Delta x_1 + \dots + f_n(x_j)\Delta x_j = \sum_{i=1}^j f_n(x_i)\Delta x_i. \quad (16)$$

Vegyük észre, hogy az utóbbi egyenlet jobb oldalán álló szumma alakja egy integrálközelítő összeghez hasonlít. Ez így is van, vagyis ha elég finom a felosztásunk, akkor $F_n(x_j)$ értéke egy integrállal megadható. Ezzel bővebben a félév során még fogunk foglalkozni.



7. ábra. A minőségellenőrzés során kapott tapasztalati sűrűségfüggvények. a.) nagy gyártási ingadozás és hibás méretbeállítás. b.) szűk gyártási ingadozás, de hibás méretbeállítás. c.) nagy gyártási ingadozás, pontos méretbeállítás. d.) szűk gyártási ingadozás és pontos méretbeállítás.

3. *Példa:* A tapasztalati sűrűségfüggvényt a valós életben minőség-ellenőrzésre is szokták alkalmazni. Ehhez tekintsünk egy terméket, mely névleges mérettel, sorozatgyártással készül. Célunk az, hogy a gyártó gépet pontosan állítsuk be és a gyártás minősége minél jobb legyen. Ezt úgy tehetjük meg, hogy a gyártás után leellenőrizzük a termékek méretét, és ezeket a mérési adatokat vizsgáljuk. Az adatokból elkészült tapasztalati sűrűségfüggvények négyféleképpen alakulhatnak, ezeket a 7. ábrán láthatjuk. A kívánt méretet x jelöli, a mért adataink átlagát pedig $\bar{\xi}$.

2.2. Két változó közötti kapcsolat

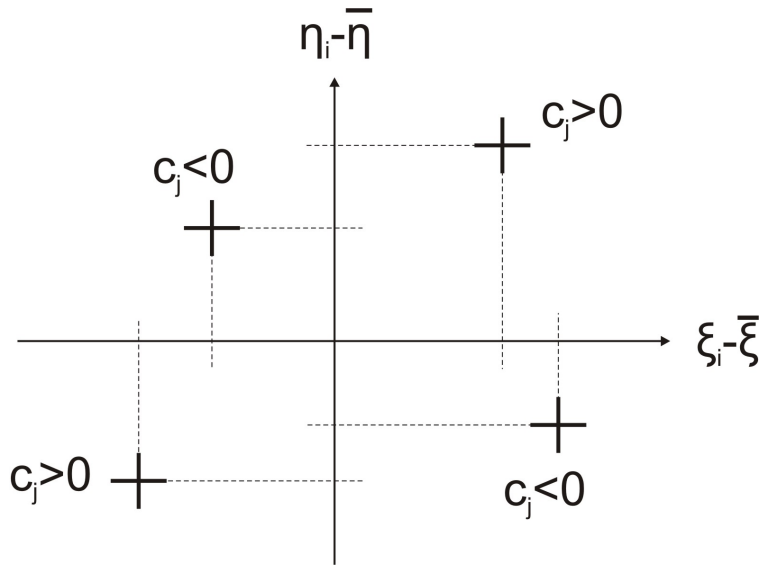
A fizikában és a mérnöki életben sokszor szükségünk lehet két változó közötti kapcsolat vizsgálatára. Gondoljunk például arra, hogy egy csap nyitásának függvényében változik a csapon átfolyó vízárám.

Persze nem mindig van ilyen kapcsolat. A következőkben arra a kérdésre fogunk választ adni, hogy van-e lineáris kapcsolat két megfigyelt változó között. Két változó esetén a megfigyeléseink pontpárok lesznek. Legyenek ezek $(\xi_1, \eta_1), (\xi_2, \eta_2), \dots, (\xi_n, \eta_n)$. Ezeket a pontpárokat ábrázolhatjuk egy derékszögű koordináta-rendszerben (8. ábra), melynek az X tengelyén $\xi_j - \bar{\xi}$ van, az Y tengelyén pedig $\eta_j - \bar{\eta}$. Vagyis az origót betöltük az átlagokba. Ekkor mind a négy síknegyedbe kerülhet pontpár. Minden megfigyeléshez definiáljuk a következő értéket:

$$c_j = (\xi_j - \bar{\xi})(\eta_j - \bar{\eta}). \quad (17)$$

c_j előjele a szorzat tényezőinek előjelétől függ, vagyis:

- $c_j < 0$ a bal felső és jobb alsó síknegyedekben,
- $c_j > 0$ a jobb felső és bal alsó síknegyedekben.



8. ábra. A pontpárok ábrázolása és a c_j együtthatók előjele.

Jelölje C a c_j értékek összegét:

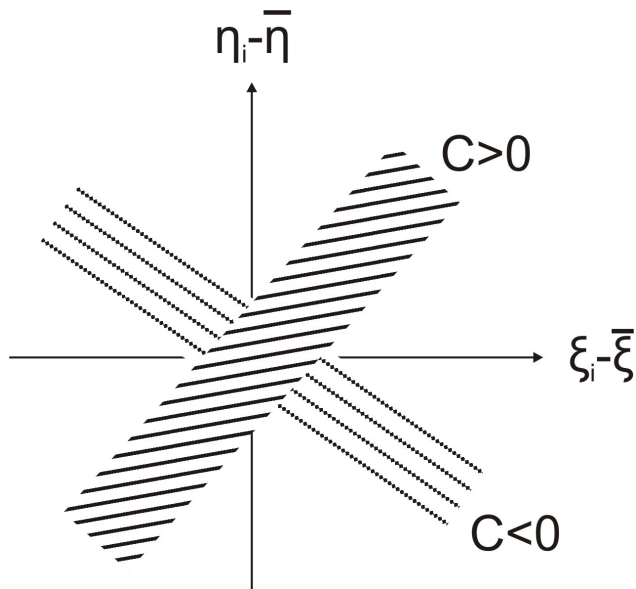
$$C = \sum_{j=1}^n c_j = \sum_{j=1}^n (\xi_j - \bar{\xi})(\eta_j - \bar{\eta}). \quad (18)$$

C előjele meghatározza, hogy a pontok hogyan helyezkednek el a koordináta-rendszerben, ezt a 9. ábán szemléltetjük. Annak érdekében, hogy a C mérőszám a pontok számától és az ingadozásoktól is független legyen, átlagolni kell, valamint leosztani a két változóhoz tartozó tapasztalati szórásokkal. Az így kapott számot nevezzük **tapasztalati korrelációs együtthatónak** és $\rho(\xi, \eta)$ -val jelöljük. Tehát

$$\rho(\xi, \eta) = \frac{\frac{1}{n} \sum_{j=1}^n (\xi_j - \bar{\xi})(\eta_j - \bar{\eta})}{s_{\xi} s_{\eta}}. \quad (19)$$

A korrelációs együtthatóra mindig fennáll, hogy $-1 \leq \rho \leq 1$. Értéke pedig választ adhat arra, hogy van-e valamiféle determinisztikus, függvényszerű kapcsolat a két változó között, de sajnos ez az érték *csak lineáris kapcsolat* szorosságát mutatja meg. Mégpedig:

- Ha $\rho \approx -1$, akkor ξ és η között ellentétesen változó kapcsolat van és a pontokra egy negatív meredekségű egyenes illeszthető.
- Ha $\rho \approx 1$, akkor ξ és η között együtt változó kapcsolat van és a pontokra egy pozitív meredekségű egyenes illeszthető.



9. ábra. Ha a pontok egy pozitív meredekségű egyenes mentén helyezkednek el, akkor C előjele is pozitív. Ha pedig a pontok egy negatív meredekségű egyenes mentén helyezkednek el, akkor C előjele negatív.

Ha például a megfigyelt változók között négyzetes kapcsolat van, vagyis $\eta = \xi^2$, akkor a pontok egy parabolán helyezkednek el, vagyis a felső félsíkon. Ekkor a c_j értékek az egyik síknegyedben negatívak, a másikban pedig pozitívak és így $\rho \approx 0$.

A korreláció egyik speciális esete a **rangkorreláció**. Ezt a fogalmat egy példán keresztül vezetjük be.

4. *Példa:* Tekintsünk egy borversenyt, ahol 4 fajta bort tesztel egy kéttagú zsűri. Jelöljük a különböző borokat A, B, C és D betűkkel, az első zsűri pontszámait ξ míg a második zsűriét η fogja megadni. A zsűri döntése az alábbi táblázatban látható. Arra vagyunk kíváncsiak, hogy összhangban van-e a két zsűritag döntése? Erre a választ a két rangsor közötti korrelációs együtthatóból kaphatjuk meg.

	A	B	C	D
1. zsűritag (ξ)	2	1	3	4
2. zsűritag (η)	1	3	2	4

3. előadás

Az előadáson folytatjuk a rangkorreláció tárgyalását. Bemutatunk két görbeillesztési módszert, a legkisebb négyzetek módszerét és a Wald-módszert. Az illesztés jóságának vizsgálatára bevezetjük a determinációs együttható fogalmát.

Rangkorreláció, más néven **Spearman-féle korreláció** esetében elsősorban olyan mennyiségeket vizsgálunk, ahol nincsen objektív mérce. Ilyen például egy olimpiai sportág zsűri általi értékelése, vagy az X-faktor.

Először a két tagból álló zsűri esetével foglalkozunk. Tegyük fel, hogy n_V számú versenyzőnk van, őket jelöljük v_1, v_2, \dots, v_{n_V} betűkel, a zsűri két tagja legyen z_1 és z_2 . Az i . ($i = 1 \dots n_V$) versenyző összesített pontszámát jelölje $c_i = r_{i,1} + r_{i,2}$, ahol $r_{i,1}$ és $r_{i,2}$ rendre az első illetve a második zsűritagtól kapott pontszám. Az a versenyző nyer akinek a legkisebb az összpontszáma. A kérdés pedig az, hogy mennyire egyhangú a zsűri döntése. Erre a két pontszámsorozat közötti korrelációs együttható ad választ:

$$\rho(r_1, r_2) = \frac{\frac{1}{n_V} \sum_{i=1}^{n_V} (r_{i,1} - \bar{r}_1)(r_{i,2} - \bar{r}_2)}{s_1 s_2}. \quad (20)$$

Ahol \bar{r}_i jelöli az i . zsűritagtól kapott átlagos pontszámot. Vezessük be a $d_i = r_{i,1} - r_{i,2}$ jelölést. Belátható, hogy ezzel a (20) egyenlet a következő alakra hozható:

$$\rho(r_1, r_2) = 1 - \frac{6 \sum_{i=1}^{n_V} d_i^2}{n_V(n_V^2 - 1)}. \quad (21)$$

Megjegyezzük, hogy a Rang-korreláció tetszőleges pontsorozat esetén megmondja, hogy a pontsorozat mennyire tekinthető monotonnak. Ehhez szükséges először a pontsorozatot mindkét koordináta szerint rangsorolni. A (21) képletet erre a transzformált sorozatpárra kell alkalmazni. Az eredmény pedig nem a lineáris kapcsolat erősségét fogja jelenteni, hanem a monotonitás erősségét.

Több zsűritag esetén a **Kendall-féle konkordancia** együtthatóval szokás összehasonlítást végezni. Ehhez legyen a zsűri létszáma n_Z , a versenyzők száma továbbra is n_V . A zsűri által adott pontszámokat írjuk a lenti táblázatba.

	Z_1	Z_2	\dots	Z_{n_Z}	Σ
v_1	$r_{1,1}$	$r_{1,2}$	\dots	r_{1,n_Z}	c_1
v_2	$r_{2,1}$	$r_{2,2}$		r_{2,n_Z}	c_2
\vdots				\vdots	\vdots
v_{n_V}	$r_{n_V,1}$	$r_{n_V,2}$		r_{n_V,n_Z}	c_{n_V}

$c_i = \sum_{j=1}^{n_Z} r_{i,j}$, az i . versenyző pontszámainak az összege, vagyis a táblázatunkban az i . sorösszeg.

Rendezzük ezeket az összegeket növekvő sorrendbe, az új jelöléssel: $c_1^* \leq c_2^* \leq \dots \leq c_{n_V}^*$. Meggondolható a következő. Ha a zsűriben teljes az összhang, vagyis mindenki ugyanazt a versenyzőt rakja az i . helyre, akkor

$$c_1^* = n_Z, \quad c_2^* = 2n_Z, \quad \dots \quad c_i^* = in_Z, \quad \dots \quad c_{n_V}^* = n_V n_Z \quad (22)$$

$$\Delta c_i^* = n_Z, \quad (i = 1, \dots, n_V - 1). \quad (23)$$

Ahol Δc_i^* jelöli a c_i^* és c_{i+1}^* közötti különbséget. Ha semmilyen összhang nincs a zsűritagok pontozása között, akkor

$$\bar{c} = \frac{1}{n_V} \sum_{i=1}^{n_V} c_i, \quad (24)$$

$$\Delta c_i^* = 0, \quad (i = 1, \dots, n_V - 1). \quad (25)$$

Tekintsük a következő, szórás-szerű összeget:

$$S^2 = \sum_{i=1}^{n_V} (\bar{c} - c_i)^2. \quad (26)$$

S^2 pontosan akkor lesz maximális, ha a zsűriben teljes az összhang, vagyis minden $\Delta c_i^* = n_Z$, ($i = 1, \dots, n_V - 1$). Valamint pontosan akkor minimális, ha a zsűri pontozása össze-vissza történt, mert ekkor $\bar{c} = c_i$, és ebből $S^2 = 0$. A **Kendall-féle konkordancia együttható** az aktuális szórást hasonlítja össze azzal az esettel, amikor a zsűri szavazata egyhangú, ezt W -vel jelöljük:

$$W = \frac{S_{akt}^2}{S_{max}^2}.$$

S_{max} kiszámolható a következő módon:

$$S_{max}^2 = \frac{1}{12} n_Z^2 n_V (n_V^2 - 1). \quad (27)$$

Tehát

$$W = \frac{12 \sum_{i=1}^{n_V} (\bar{c} - c_i)^2}{n_Z^2 n_V (n_V^2 - 1)}. \quad (28)$$

3.1. Görbeillesztés

3.1.1. Legkisebb négyzetek módszere

Felmerülhet az a kérdés, hogy ha a változók között nem lineáris a kapcsolat, akkor mennyire nem az, illetve található-e másfajta kapcsolat és mi annak a módja. Ezzel a matematikában a regresszióanalízis foglalkozik. A görbeillesztésre a legelterjedtebb módszer a **legkisebb négyzetek módszere**, amely Gauss (1795) nevéhez fűződik. Megjegyezzük, hogy az Excel beépített trendvonalillesztője is ezt a módszert használja. A görbeillesztés jóságának számszerűsítésére a determinációs együttható szolgál.

A legkisebb négyzetek módszerének alkalmazhatóságához feltétel az, hogy az abszcissa változójának ne legyen hibája, vagyis azt pontosan ismerjük, ennek fényében ezt a változót latin betűvel fogjuk

jelölni. A függő változó pontos ismerete nem feltétel, vagyis azt terhelheti hiba, így ezt görög betűvel fogjuk jelölni. Legyenek tehát a megfigyelt pontjaink (x_j, η_j) , $(j = 1, \dots, n)$. Olyan grafikont szeretnénk találni, amely a pontok között halad, és a lehető legjobban közelíti azokat. Tegyük fel először, hogy ismerjük a pontok közötti függvénykapcsolatot, legyen ez $y(x)$:

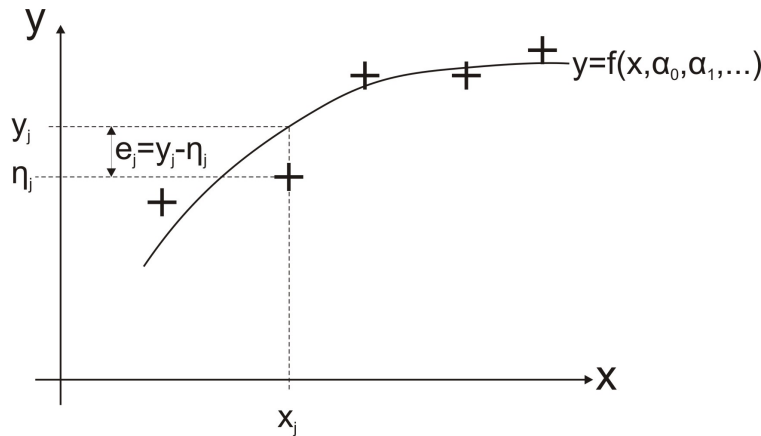
$$y(x) = y(x, \alpha_0, \alpha_1, \dots), \quad (29)$$

ahol α_i ($i = 0, 1, \dots$) paramétereket szeretnénk alkalmasan megválasztani. Az a célunk, hogy az x_j helyeken $y(x_j)$ és η_j eltérése minimális legyen. Ehhez nézzük meg minden j index esetén ezeket az eltéréseket:

$$e_j := \eta_j - y(x_j, \alpha_0, \alpha_1, \dots). \quad (30)$$

számunkra az eltérések előjele nem érdekes, ezért emeljünk négyzetre:

$$e_j^2 = (\eta_j - y(x_j, \alpha_0, \alpha_1, \dots))^2. \quad (31)$$



10. ábra. Legkisebb négyzetek módszerével illesztett y görbe és a megfigyelt pontok közötti eltérések.

Annak érdekében, hogy az összes hibát egyszerre tudjuk minimalizálni, adjuk össze a négyzetes eltéréseket, ezt jelöljük D -vel:

$$D(\alpha_0, \alpha_1, \dots) := \sum_{j=1}^n (\eta_j - y(x_j, \alpha_0, \alpha_1, \dots))^2. \quad (32)$$

Mivel x_j és η_j ismert mennyiségek minden $j = 1, \dots, n$ esetén, ezért a (32) egyenlőség jobb oldala csak $\alpha_0, \alpha_1, \dots$ paraméterektől függ. Tehát a feladatunk a $D(\alpha_0, \alpha_1, \dots)$ függvény minimumát meghatározni az α_i ($i = 0, 1, \dots$) mennyiségek függvényében. Többváltozós függvény minimumhelyét a parciális deriváltak segítségével tudjuk meghatározni. Jelen esetben azzal nem fogunk foglalkozni, hogy létezik-e minimumhely, egyértelmű-e, lokális vagy globális-e a minimum stb.

Az egyszerűség kedvéért példaként a továbbiakban $y(x)$ másodfokú függvény esetére mutatjuk be a módszert. Tehát

$$y(x, \alpha_0, \alpha_1, \alpha_2) := \alpha_0 + \alpha_1 x + \alpha_2 x^2 \quad (33)$$

Így a következő egyenletet kell megoldanunk:

$$\frac{\partial}{\partial \alpha_i} \sum_{j=1}^n (\eta_j - \alpha_0 - \alpha_1 x_j - \alpha_2 x_j^2)^2 = 0. \quad (34)$$

Minden α_i -re elvégezve a deriválást az alábbi egyenletrendszert kapjuk:

$$\begin{aligned} -2 \sum_{j=1}^n (\eta_j - \alpha_0 - \alpha_1 x_j - \alpha_2 x_j^2) &= 0, \\ -2 \sum_{j=1}^n (\eta_j - \alpha_0 - \alpha_1 x_j - \alpha_2 x_j^2) x_j &= 0, \\ -2 \sum_{j=1}^n (\eta_j - \alpha_0 - \alpha_1 x_j - \alpha_2 x_j^2) x_j^2 &= 0. \end{aligned} \quad (35)$$

Egyszerűsítve -2 -vel, a szummákat felbontva és rendezve az egyenlőségeket a következő lineáris algebrai egyenlőségrendszert kapjuk:

$$\begin{aligned} n\alpha_0 + \alpha_1 \sum_{j=1}^n x_j + \alpha_2 \sum_{j=1}^n x_j^2 &= \sum_{j=1}^n \eta_j, \\ \alpha_0 \sum_{j=1}^n x_j + \alpha_1 \sum_{j=1}^n x_j^2 + \alpha_2 \sum_{j=1}^n x_j^3 &= \sum_{j=1}^n x_j \eta_j, \\ \alpha_0 \sum_{j=1}^n x_j^2 + \alpha_1 \sum_{j=1}^n x_j^3 + \alpha_2 \sum_{j=1}^n x_j^4 &= \sum_{j=1}^n x_j^2 \eta_j. \end{aligned} \quad (36)$$

Felhívjuk a figyelmet arra, hogy a fenti egyenletrendszerben az ismeretlenek α_0 , α_1 és α_2 , az együtthatók pedig a szummás kifejezések, amelyeket a mérési eredményeinkből ki tudunk számolni. (itt kihasználjuk azt is, hogy a függvénykapcsolatot leíró egyenlet paramétereire nézve lineáris). Az egyenletrendszer megoldhatóságával az előadás keretein belül nem foglalkozunk. Viszont megjegyezzük, hogy lineáris algebrai tanulmányainkból tudjuk, hogy a fenti rendszer mátrixos alakra hozható, és így, ha az megoldható, akkor például Excel programmal könnyen meghatározhatók az ismeretlenek.

3.1.2. Determinációs együttható

Miután meghatároztuk, hogy milyen magasabb rendű polinomot illesztünk a pontjainkra, természetes módon merül fel a kérdés, hogy mennyire jó a választásunk. Erre ad választ a **determinációs együttható**. Ehhez legyenek a mérési pontjaink (x_j, η_j) , $j = 1, 2, \dots, n$, és a választott görbénk $y(x)$. Számoljuk ki az η_j értékek átlagát és az átlagtól vett tapasztalati szórásnégyzetet! Ez utóbbit S_t^2 -vel jelöljük (a t index a *teljes szórásra* vonatkozik):

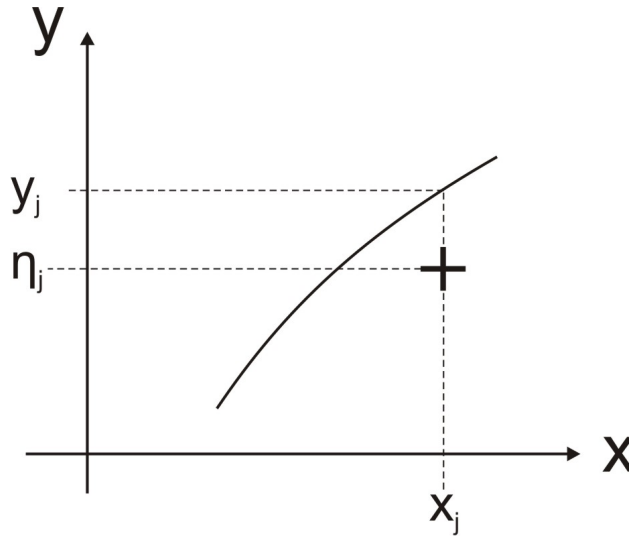
$$\bar{\eta} = \frac{1}{n} \sum_{j=1}^n \eta_j \quad (37)$$

$$S_t^2 = \frac{1}{n} \sum_{j=1}^n (\eta_j - \bar{\eta})^2 \quad (38)$$

Az η_j értékek tapasztalati szórásnégyzetét az illesztett polinomunkhoz képest is ki lehet számolni, ezt S_m^2 jelöli (az m index a *maradék szórásra* vonatkozik):

$$S_m^2 = \frac{1}{n} \sum_{j=1}^n (\eta_j - y(x_j))^2 \quad (39)$$

Tehát S_m^2 azt mondja meg, hogy az illesztett y görbe helyettesítési értékei az x_j helyeken mennyire térnek el az η_j mért értékektől (11. ábra).



11. ábra. A determinációs együttható számításakor az η_j pontok szórását az illesztett görbénkhez képest is ki kell számolni.

Ekkor a determinációs együttható:

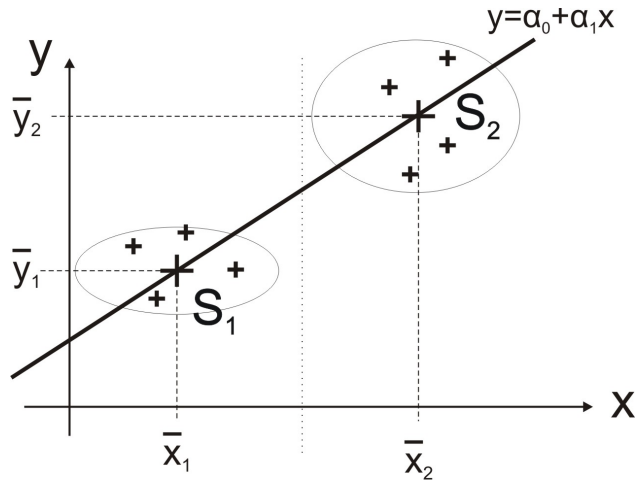
$$R^2 = \frac{S_t^2 - S_m^2}{S_t^2} \quad (40)$$

Gondoljuk meg, hogy ha $S_m^2 = 0$, akkor $\eta_j = y(x_j)$, vagyis az illesztett függvényünk átmegy minden ponton. Ekkor $R^2 = 1$. Ha pedig $y = \bar{\eta}$, azaz a közelítő függvény épp az átlag, akkor $R^2 = 0$. Tehát R^2 azt mondja meg, hogy mennyire halad közel a pontokhoz a választott függvény. Megjegyezzük, hogy $R^2 = 1$ esetén y interpolációs polinom, ami nem a legjobb választás, hiszen egy n pontra illeszkedő interpolációs polinom akár $(n - 1)$ -edfokú is lehet. Statisztikai szempontból egy n -edfokú polinom pedig nem szolgál hasznos információkkal a megfigyeléseinkre vonatkozólag. Ezért inkább arra törekedjünk, hogy az illesztett polinomunk foka jóval kisebb legyen, mint a megfigyeléseink száma. A hallgató maga is meggondolhatja, hogy egyenes illesztése esetén a tapasztalati korrelációs együttható és a determinációs együttható megegyeznek

$$\rho^2(y_j, \eta_j) = R^2. \quad (41)$$

3.1.3. Wald-módszer

Ha két változó kötött lineáris kapcsolat van, akkor egy másik egyenesillesztési módszer a **Wald-módszer**. Később látni fogjuk, hogy ez a módszer akkor alkalmazható, ha mindkét változót hiba terheli. A módszer a következő gondolaton alapul. Ábrázoljuk a megfigyelt pontpárokat és a síkon határozzunk meg egy elválasztó egyenest, ezzel két részhalmazra osztva a megfigyeléseket. Számoljuk ki a pontthalmazok súlypontjait, és illesszük a megfigyelésekre azt az egyenest amely áthalad a két súlyponton (12. ábra).



12. ábra. Wald-módszerrel illesztett egyenes.

Tehát legyenek a megfigyelt pontpárjaink $\{\xi_j, \eta_j\}_{j=1}^n$, az illesztett egyenes egyenlete pedig $y = \alpha_1 x + \alpha_0$. A pontokat úgy érdemes szétválasztani, hogy a súlypontok a lehető legtávolabb essenek egymástól. Ha a pontokat például egy pontdiagramon ábrázolva nem egyértelmű, hogy hol válasszuk el, akkor rendezzük sorba a ξ_j értékek szerint a pontpárokat. Legyen a sorba rendezett pontsorozatunk $\{\xi_j^*, \eta_j^*\}_{j=1}^n$, és legyen ξ_r^* az az érték ahol szétválasztjuk a pontokat. Ekkor a két súlypont:

$$S_1 = (\bar{x}_1, \bar{y}_1), \quad S_2 = (\bar{x}_2, \bar{y}_2), \quad (42)$$

ahol az S_1 és S_2 súlypont koordinátáit a pontthalmazok átlagaiból kaphatjuk meg:

$$\bar{x}_1 = \frac{1}{r} \sum_{j=1}^r \xi_j^*, \quad \bar{y}_1 = \frac{1}{r} \sum_{j=1}^r \eta_j^*, \quad (43)$$

$$\bar{x}_2 = \frac{1}{n-r} \sum_{j=r+1}^n \xi_j^*, \quad \bar{y}_2 = \frac{1}{n-r} \sum_{j=r+1}^n \eta_j^* \quad (44)$$

A két súlypontot összekötő egyenes iránytangense és konstans tagja:

$$\alpha_1 = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}, \quad \text{és} \quad \alpha_0 = \bar{y}_1 - \alpha_1 \bar{x}_1 = \bar{y}_2 - \alpha_1 \bar{x}_2 \quad (45)$$

4. előadás

Az elmúlt előadásokon az úgynevezett leíró statisztikával foglalkoztunk, ahol véges n számú megfigyelésekből vontunk le következtetéseket. Azonban fontos lehet azt is tudni, hogy mi történik akkor, ha a megfigyeléseink számát növeljük, vagyis ha $n \rightarrow \infty$? Vajon stabilizálódik-e az átlag? Hogyan alakulnak a már korábban tárgyalt függvények?

Az előadáson a valószínűségszámítás alapvető fogalmait ismerjük meg: definiálni fogjuk a véletlen esemény és a valószínűség fogalmát. Ismertetjük az eloszlás- és sűrűségfüggvényt, bevezetjük a várható érték fogalmát.

4.1. Véletlen esemény

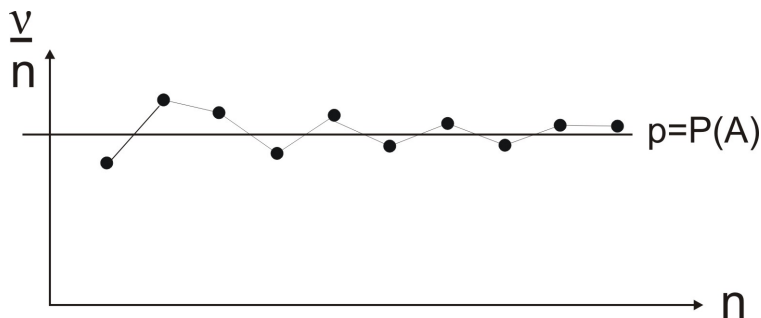
A valóságban számtalan olyan esemény van melynek bekövetkezését a figyelembe vett körülmények nem határozzák meg egyértelműen. Például tegyük fel, hogy csapágygolyók átmérőjét szeretnénk mérni, és meghatározunk egy intervallumot, amin belül kívánjuk tartani ennek értékét. A tapasztalat azt mutatja, hogy nem lesz minden csapágygolyó mérete az intervallumon belül. Ez azzal magyarázható, hogy az általunk figyelembe vett körülmények még nem határozzák meg pontosan a csapágygolyók méretét.

A fenti példa alapján a következő definíciót mondhatjuk ki. **Véletlen eseménynek** nevezzük az olyan eseményt, amely esetén a figyelembe vett körülmények nem határozzák meg egyértelműen az esemény kimenetelét.

Fontos megjegyeznünk, hogy ebben a fejezetben a szemléletességet tartjuk elsődlegesnek. A bevezetett fogalmak pontos meghatározása a valószínűségszámítás tankönyveiben megtalálhatók.

4.2. Relatív gyakoriság, valószínűség

Végezzünk egy kísérletsorozatot, és egy általunk kiválasztott esemény relatív gyakoriságát ábrázoljuk az elvégzett kísérletek számának függvényében. Ekkor azt tapasztalhatjuk, hogy a relatív gyakoriság ingadozik. Belátható, hogy ha növeljük az elvégzett kísérletek számát, akkor az ingadozás csillapodik: egy átlag körül oszcillál. Ezt az átlagot nevezzük az esemény bekövetkezésére vonatkozó **valószínűségnek** (13. ábra). Nyilvánvalóan ez a definíció szemléletes, de pontatlan.



13. ábra.

Felhívjuk a figyelmet arra, hogy a pontsorozat nem feltétlenül tart az átlaghoz, azonban egyre

jobban közelíti azt.

Legyen A egy esemény, ekkor az A bekövetkezésének valószínűségét a következő módon írjuk:

$$P(A) = p, \quad 0 \leq p < 1. \quad (46)$$

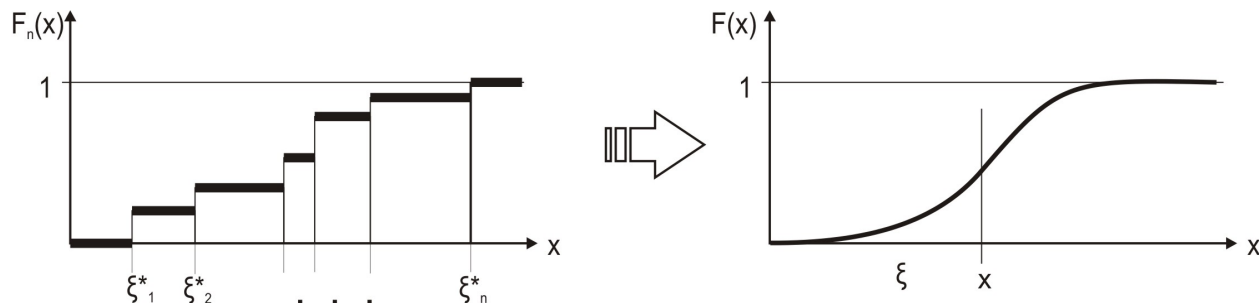
Ha B biztos esemény, akkor $P(B) = 1$. Ha pedig L lehetetlen esemény, akkor $P(L) = 0$. Fontos megjegyezni, hogy ha egy esemény bekövetkezésének valószínűsége 0, akkor nem feltétlenül lesz az esemény lehetetlen esemény. Példaképp, jelöljünk ki a síkon egy pontot, és egy körző hegyével véletlenszerűen bökjünk a síkra. Ekkor annak a valószínűsége, hogy éppen az általunk előre kijelölt pontra bökünk 0, mégis ez nem lehetetlen.

Egy esemény számszerűsítése érdekében bevezetjük a **valószínűségi változókat**. Az ilyen változó értéke nem egyértelmű, véletlenszerű ingadozást mutat, vagyis a figyelembe vett körülmények nem határozzák meg egyértelműen az értékét. Általában a valószínűségi változókat görög betűkkel jelöljük. A továbbiakban olyan módszereket mutatunk be, amelyek egy valószínűségi változó értékeinek leírására alkalmasak.

4.3. Valószínűségi változók leírása

A továbbiakban tekintsük a ξ valószínűségi változót.

Az első függvény, amit a ξ leírására használunk, az **eloszlásfüggvény**. $n \rightarrow \infty$ esetén az $F_n(x)$ tapasztalati eloszlásfüggvény ehhez a függvényhez tart (lásd a 14. ábrát). Emlékezzünk vissza,



14. ábra.

hogy egy n elemű minta esetén $F_n(x)$ azt mondta meg, hogy mennyi azon elemek relatív gyakorisága amelyek x -nél kisebbek. Ennek analógiájára a ξ valószínűségi változó F eloszlásfüggvénye az x helyen azt mondja meg, hogy mekkora annak a valószínűsége, hogy ξ az x értéknél kisebb. Matematikailag ezt a következőképpen írjuk le:

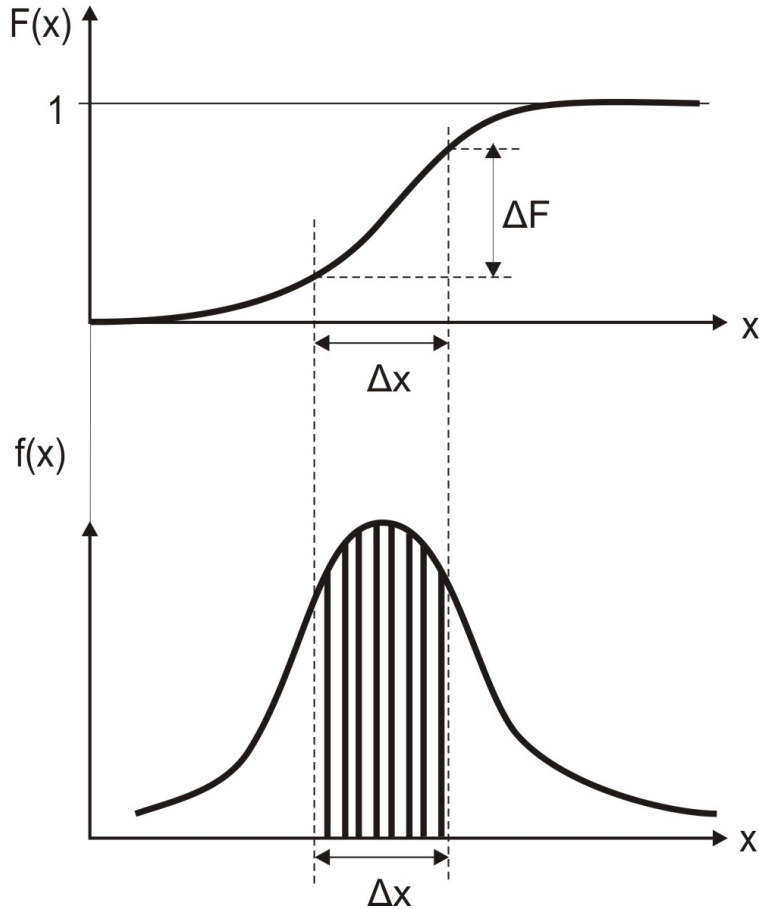
$$F(x) = P(\xi < x). \quad (47)$$

Könnyen meggondolható, hogy bármely ξ valószínűségi változó F eloszlásfüggvényére a következő tulajdonságok mindig igazak:

- $0 \leq F(x) \leq 1$ minden x -re.
- Ha $x \rightarrow -\infty$, akkor $F(x) \rightarrow 0$,
ha $x \rightarrow \infty$, akkor $F(x) \rightarrow 1$.

- $F(x)$ monoton növekvő függvény: ha $x_1 \leq x_2$ akkor $F(x_1) \leq F(x_2)$.
- $P(x_1 \leq \xi \leq x_2) = F(x_2) - F(x_1)$.

Az $f_n(x)$ tapasztalati sűrűségfüggvény $n \rightarrow \infty$ esetében beszélünk **sűrűségfüggvényről**. Ha a j . intervallumba esés gyakoriságát ν_j jelöli, akkor $f_n(x_j) = \frac{\nu_j}{n\Delta x_j}$ volt. n növelésével az intervallumokba esések gyakorisága is nő, és $n \rightarrow \infty$ esetén a tapasztalati sűrűségfüggvény folytonos függvényhez tart. Ez lesz az $f(x)$ sűrűségfüggvény. A 15. ábrán egy ξ valószínűségi változó



15. ábra. Az eloszlás és sűrűségfüggvény kapcsolata

eloszlásfüggvénye és a hozzá tartozó sűrűségfüggvény látható. A jelölések alapján a Δx intervallumba esés valószínűsége éppen ΔF , vagyis $\Delta F = P(\xi \in \Delta x)$. A sűrűségfüggvényt pontosan úgy definiáltuk, hogy egy intervallumba esés valószínűsége éppen az intervallum és a sűrűségfüggvény közötti területtel legyen egyenlő, vagyis $P(\xi \in \Delta x) \cong f(x)\Delta x$. Ezzel

$$\Delta F \cong f(x)\Delta x. \quad (48)$$

Amiből a sűrűségfüggvény:

$$\frac{\Delta F}{\Delta x} \cong f(x). \quad (49)$$

Határátmenetet véve a sűrűségfüggvény és az eloszlásfüggvény között a következő összefüggést kaptuk:

$$f(x) = \frac{dF}{dx}. \quad (50)$$

És az alábbi tulajdonságok mindig fennállnak:

- $f(x) \geq 0$ minden x -re.
- $F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx$.
- Ha $x \rightarrow \pm\infty$ akkor $f(x) \rightarrow 0$.
- $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Megjegyezzük, hogy léteznek nem folytonos eloszlású valószínűségi változók, de mi ennek a tárgynak a keretein belül ilyenekkel nem foglalkozunk.

A ξ valószínűségi változó **várható értéke** olyan állandó amely körül ξ értéke ingadozik. Ennek definiálásához először tekintsük a ξ_1, \dots, ξ_n megfigyelések átlagát:

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i. \quad (51)$$

Ezt a relatív gyakoriságok és a tapasztalati sűrűségfüggvény segítségével közelíthetjük. Ehhez legyen J a tapasztalati sűrűségfüggvény intervallumainak a száma, a j . intervallum egy pontja x_j , és a j . intervallumba esés gyakorisága ν_j ($j = 1 \dots J$). Ekkor

$$\bar{\xi} \cong \frac{1}{n} \sum_{j=1}^J \nu_j x_j = \sum_{j=1}^J \frac{\nu_j}{n\Delta x_j} x_j \Delta x_j. \quad (52)$$

Ahol Δx_j a j . intervallum szélessége. A fenti jelölésekkel a j . intervallum feletti oszlop magassága:

$$f_n(x_j) = \frac{\nu_j}{n\Delta x_j},$$

ezzel pedig

$$\sum_{j=1}^J \frac{\nu_j}{n\Delta x_j} x_j \Delta x_j = \sum_{j=1}^J f_n(x_j) x_j \Delta x_j. \quad (53)$$

$n \rightarrow \infty$ esetén a fenti összeg egy integrálközelítő összeg:

$$\sum_{j=1}^J f_n(x_j) x_j \Delta x_j \rightarrow \int_{-\infty}^{+\infty} x f(x) dx. \quad (54)$$

ezt az integrált nevezzük a ξ valószínűségi változó várható értékének, amit $M(\xi)$ -vel jelölünk:

$$M(\xi) = \int_{-\infty}^{+\infty} x f(x) dx. \quad (55)$$

Bár az átlag véletlentől függő mennyiség, az $M(\xi)$ várható érték nem valószínűségi változó, hanem állandó mennyiség. A várható érték tulajdonságai a következők:

- $M(\xi + a) = M(\xi) + a$.
- $M(a\xi) = aM(\xi)$.
- $M(\xi + \eta) = M(\xi) + M(\eta)$.

5. előadás

Ezen az előadáson definiáljuk valószínűségi változókra a szórásnégyzet fogalmát. Bemutatjuk a standardizálás módszerét. Kimondjuk a nagy számok törvényét. Megismerkedünk két nevezetes eloszlással, az egyenletes illetve a normális eloszlással. A centrális határeloszlás tételéről is szó esik.

A ξ valószínűségi változó **szórásnégyzete** annak a mértékét mondja meg, hogy ξ mennyire ingadozik a várható értéke körül. Ezt $D^2(\xi) = \sigma_\xi^2$ jelöli és

$$D^2(\xi) = M[(M(\xi) - \xi)^2]. \quad (56)$$

A szórás a szórásnégyzet pozitív gyöke: $\sigma = +\sqrt{\sigma^2}$, tulajdonságai pedig:

- $0 \leq \sigma$, és 0 , ha ξ determinisztikus változó.
- Ha $\zeta = a\xi$, a konstans, akkor $D^2(\zeta) = a^2 D^2(\xi)$.
- Ha $\zeta = a + \xi$, a konstans, akkor $D^2(\zeta) = D^2(\xi)$.
- Ha $\zeta = a\xi + b$, a, b konstans, akkor $D^2(\zeta) = a^2 D^2(\xi)$.
- Ha $\zeta = \xi_1 + \xi_2$, és ξ_1 és ξ_2 függetlenek egymástól, vagyis az egyik esemény bekövetkezése nem befolyásolja a másik esemény ismeretét, akkor $D^2(\zeta) = D^2(\xi_1) + D^2(\xi_2)$.

5.1. Alkalmazás

A következőkben megfigyelések átlagának várható értékéről és szórásáról beszélünk. Tekintsük a $\xi_1, \xi_2, \dots, \xi_n$ megfigyeléseket, melyeknek átlagát jelölje $\bar{\xi}$. Tegyük fel továbbá, hogy a ξ_i megfigyeléseket ugyanolyan körülmények között figyeltük meg és a különböző megfigyeléseink eredményét a többi nem befolyásolta. Ekkor azt szokás mondani, hogy a megfigyeléseink azonos eloszlásúak és függetlenek, vagyis a statisztikai mintánk reprezentatív.

Mivel az átlag véletlentől függő mennyiség, ezért érdekes lehet számunkra, hogy mennyi a várható értéke. Ezt az átlag definíciója és a várható érték tulajdonságai alapján így számolhatjuk:

$$M(\bar{\xi}) = M\left(\frac{1}{n} \sum_{i=1}^n \xi_i\right) = \frac{1}{n} M\left(\sum_{i=1}^n \xi_i\right) = \frac{1}{n} \sum_{i=1}^n M(\xi_i). \quad (57)$$

Mivel feltettük, hogy a megfigyeléseink ugyanolyan eloszlásúak, ezért a várható értékük is megegyezik: $M(\xi_1) = M(\xi_2) = \dots = M(\xi_n) =: M(\xi)$. Ezért:

$$M(\bar{\xi}) = \frac{1}{n} \sum_{i=1}^n M(\xi) = \frac{1}{n} n M(\xi) = M(\xi). \quad (58)$$

Így az **átlag várható értéke** megegyezik a megfigyelések várható értékével, vagyis ugyanazon érték körül ingadozik, mint a megfigyelések. Ekkor azt mondhatjuk, hogy az átlag becslése (közelítése) a várható értéknek.

Felmerül a kérdés, hogy az fentihez hasonló becsléseket hogyan kategorizáljuk. Ehhez bevezetjük a **torzítatlan becslés** fogalmát: ha a becslés várható értéke megegyezik a keresett értékkel, akkor az torzítatlan becslés. Ennek fényében az átlag torzítatlan becslése a várható értéknek.

Az **átlag szórását** hasonlóan a várható értékéhez így számolhatjuk:

$$D^2(\bar{\xi}) = D^2\left(\frac{1}{n} \sum_{i=1}^n \xi_i\right) = \frac{1}{n^2} D^2\left(\sum_{i=1}^n \xi_i\right). \quad (59)$$

Mivel a ξ megfigyelések függetlenek és azonos σ szórásúak:

$$D^2(\bar{\xi}) = \frac{1}{n^2} D^2\left(\sum_{i=1}^n \xi_i\right) = \frac{1}{n^2} \sum_{i=1}^n D^2(\xi_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}. \quad (60)$$

Ebből pedig már látszik, hogy érdemes átlagot számolni, hiszen az kevésbé ingadozik a várható érték körül, mint a megfigyelések. Valamint $n \rightarrow \infty$ esetén az átlag szórása 0-hoz tart.

5.2. Standardizálás

Legyen ξ egy valószínűségi változó, várható értéke $M(\xi) = m$ és szórásnégyzete $D^2\xi = \sigma^2$. Ekkor ξ áttranszformálható a következő η valószínűségi változóra:

$$\eta = \frac{\xi - m}{\sigma}. \quad (61)$$

Könnyen látható, hogy η várható értéke 0 és szórása 1:

$$M(\eta) = M\left(\frac{\xi - m}{\sigma}\right) = \frac{1}{\sigma}(M(\xi) - M(m)) = \frac{1}{\sigma}(m - m) = 0, \quad (62)$$

$$D^2(\eta) = D^2\left(\frac{\xi - m}{\sigma}\right) = \frac{1}{\sigma^2}(D^2(\xi) + D^2(m)) = \frac{1}{\sigma^2}(\sigma^2 + 0) = 1. \quad (63)$$

Ezt a módszert nevezzük **standardizálásnak**.

5.3. A nagy számok törvénye

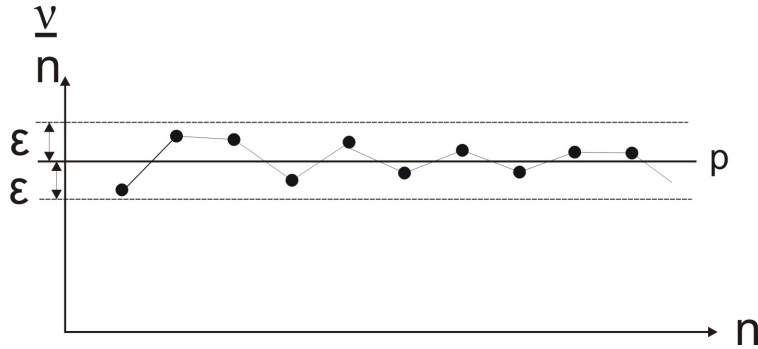
A **nagy számok törvénye** azt fogalmazza meg, hogy $n \rightarrow \infty$ esetén egy esemény relatív gyakorisága a valószínűség körül ingadozik, és annak egy ε sugarú környezetében marad. Formálisan:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{\nu}{n} - p\right| > \varepsilon\right) = 0 \quad (64)$$

A fenti konvergencia a sztochasztikus konvergencia, jelölése:

$$\frac{\nu}{n} \xrightarrow{s} p, \quad \text{ha } n \rightarrow \infty. \quad (65)$$

Ebben az értelemben az átlag sztochasztikusan tart a várható értékhez: $\bar{\xi} \xrightarrow{s} M(\xi)$.



16. ábra. Ha egy esemény elég sokszor előfordul, akkor elég nagy valószínűséggel a relatív gyakorisága az átlag egy ϵ sugarú környezetében marad

5.4. Néhány nevezetes eloszlás

A következőkben röviden ismertetünk néhány fontos valószínűségi változót és az azokat leíró függvényeket.

5.4.1. Binomiális eloszlású valószínűségi változó

Tekintsünk egy n elemű mintát, melyben egy elem vagy selejtes, vagy nem. Legyen p annak a valószínűsége, hogy egy kiválasztott elem hibás. Mekkora annak a valószínűsége, hogy az n elemű mintában éppen k darab selejt van?

A kérdés megválaszolásához vezessük be a ξ valószínűségi változót, amely azt mondja meg, hogy mekkora annak az esélye, hogy a mintában éppen k darab selejt van. ξ ekkor binomiális eloszlású lesz, melynek sűrűségfüggvénye (a fenti jelölések mellett) (17. ábrán felül):

$$P(\xi = k) = \binom{n}{k} p^k (1-p)^{(n-k)}.$$

Eloszlásfüggvénye (17. ábrán alul):

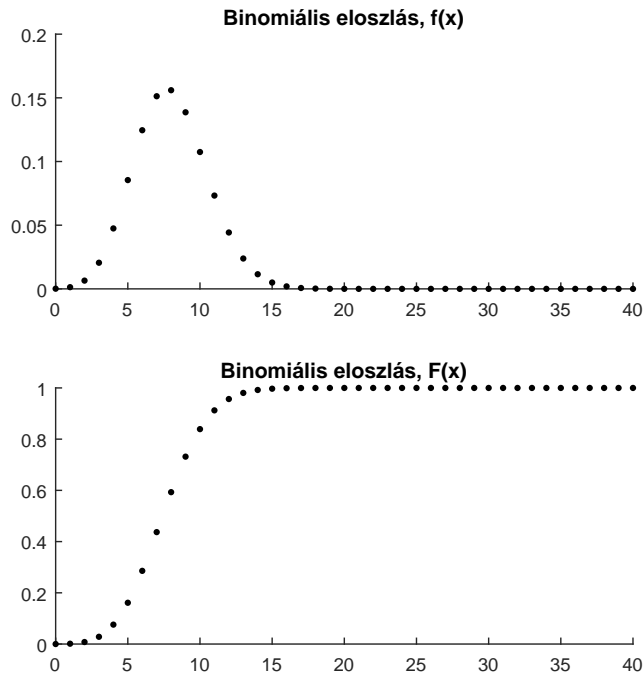
$$P(\xi < k) = \sum_{i=0}^{k-1} P(\xi = i) = \sum_{i=0}^{k-1} \binom{n}{i} p^i (1-p)^{(n-i)}.$$

Várható értéke és szórásnégyzete:

$$M(\xi) = np \quad \text{és} \quad D^2 = np(1-p).$$

ξ diszkrét eloszlású, hiszen csakis 0 és n közötti egész számokat vehet fel (ez az oka annak is, hogy az eloszlásfüggvény nem integrál alakban van megadva).

A binomiális eloszlást leíró függvényeket és konstansokat két paraméter, nevezetesen n és p egyértelműen meghatározzák, ezért szokás $\xi \in \text{Binom}(n, p)$ jelölést alkalmazni.



17. ábra. $n=40$ és $p=0,2$ paraméterű binomiális eloszlású valószínűségi változó sűrűség-, illetve eloszlásfüggvénye

5.4.2. Poisson eloszlású valószínűségi változó

Poisson-eloszlást követhetnek az alábbiak:

- adott időszakban bekövetkező véletlen események száma;
- adott tartományba jutó, véletlenül elhelyezkedő pontok száma;

Ha ξ Poisson eloszlású, akkor $k = 0, 1, 2, \dots$ értékeket vehet fel, vagyis diszkrét valószínűségi változó, és

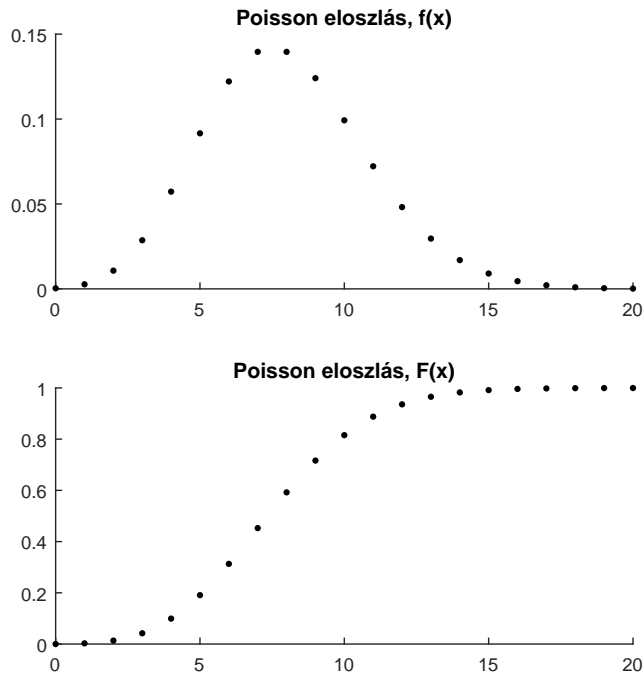
$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

λ tetszőleges rögzített valós szám a Poisson-eloszlás paramétere. Fontos, hogy a $P(\xi = k)$ valószínűség csakis az intervallum hosszától, vagy a tartomány nagyságától függ.

λ ismeretében ξ várható értéke és szórásnégyzete:

$$M(\xi) = \lambda \quad \text{és} \quad D^2 = \lambda.$$

Érdeemes megjegyezni, hogy bizonyos esetekben a Poisson-eloszlás jó közelítése a binomiális eloszlásnak. Ugyanis, ha n értéke sokkal nagyobb, mint k és p kicsi, akkor érvényes a következő közelítés:



18. ábra. $\lambda=8$ paraméterű Poisson-eloszlású valószínűségi változó sűrűség-, illetve eloszlásfüggvénye

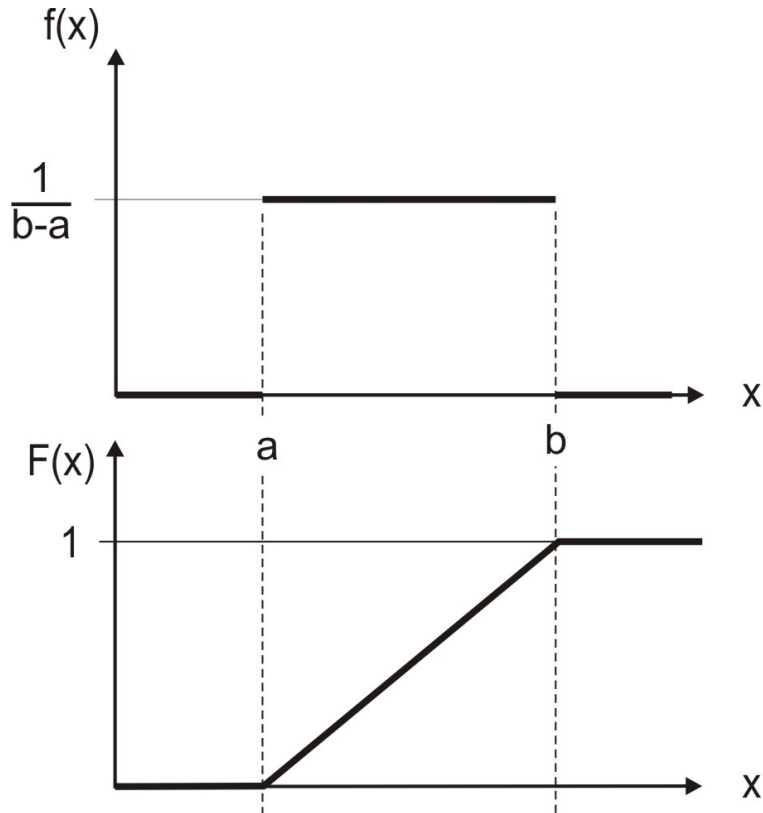
$$\binom{n}{k} p^k (1-p)^{(n-k)} \approx \frac{(np)^k}{k!} e^{-np}$$

vagyis $\lambda = np$ választással éppen a Poisson-eloszlást kapjuk.

5.4.3. Egyenletes eloszlású valószínűségi változó

A legegyszerűbb folytonos valószínűségi változó az **egyenletes eloszlású** valószínűségi változó. Ezt úgy képzelhetjük el, hogy ha adott egy $[a, b]$ intervallum, akkor ξ az intervallum bármely szakaszára ugyanakkora valószínűséggel esik. A sűrűségfüggvényét és eloszlásfüggvényét a következőképpen adhatjuk meg:

$$f(x) = \begin{cases} 0 & \text{ha } x < a \\ \frac{1}{b-a} & \text{ha } a \leq x \leq b; \\ 0 & \text{ha } b < x \end{cases} \quad F(x) = \begin{cases} 0 & \text{ha } x < a \\ \frac{x-a}{b-a} & \text{ha } a \leq x \leq b \\ 1 & \text{ha } b < x \end{cases}$$



19. ábra. Az $[a, b]$ intervallumon egyenletes eloszlású valószínűségi változó sűrűség-, illetve eloszlásfüggvénye

5.4.4. Exponenciális eloszlású valószínűségi változó

ξ exponenciális eloszlású, ha valamely esemény bekövetkezéséig eltelt időtartamot jelöli, így ez is folytonos valószínűségi változó. Sűrűség- és eloszlásfüggvénye (20. ábra):

$$f(x) = \begin{cases} 0 & \text{ha } x \leq 0 \\ \lambda e^{-\lambda x} & \text{ha } x > 0; \end{cases} \quad F(x) = \begin{cases} 0 & \text{ha } x \leq 0 \\ 1 - e^{-\lambda x} & \text{ha } x > 0 \end{cases}$$

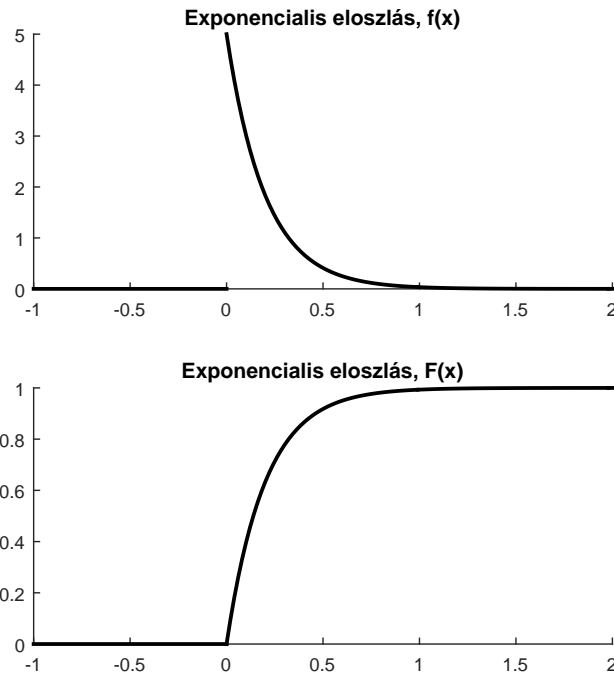
λ az exponenciális eloszlás paramétere, mellyel a várható érték és a szórásnégyzet:

$$M(\xi) = \frac{1}{\lambda} \quad \text{és} \quad D^2 = \frac{1}{\lambda^2}.$$

Az exponenciális eloszlású változó "örökifjú", ami azt jelenti, hogy az esemény bekövetkezésének esélye az idő múlásával nem nő.

5.4.5. Normális eloszlású valószínűségi változó

Egy másik nevezetes valószínűségi változó a **normális eloszlású** valószínűségi változó. Ezt a sűrűségfüggvényével szokás megadni (21. ábra):



20. ábra. $\lambda=5$ paraméterű exponenciális eloszlású valószínűségi változó sűrűség-, illetve eloszlásfüggvénye

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad (66)$$

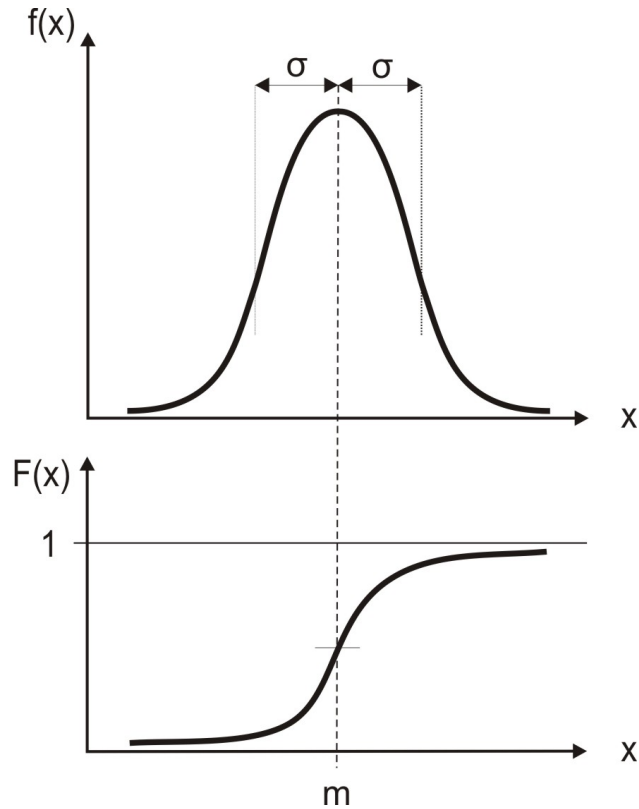
ahol m a várható érték és σ a szórás. Bármely normális eloszlású valószínűségi változó áttranszformálható egy 0 várható értékű és 1 szórású úgynevezett **standard normális eloszlású** valószínűségi változóra. Ezt az 22. ábra szemlélteti.

Emlékeztetünk arra, hogy a sűrűségfüggvény esetén egy intervallum feletti terület annak a valószínűségét adja meg, hogy a hozzá tartozó valószínűségi változó mekkora valószínűséggel esik az intervallumba. A 23. ábra szemlélteti, hogy normális eloszlású valószínűségi változó esetén mekkora annak a valószínűsége, hogy a μ_γ várható érték körül a valószínűségi változó rendre a σ_γ szórás egyszeres, kétszeres illetve háromszoros sugarába esik.

Belátható, hogy ha valamely valószínűségi változó úgy áll elő, hogy nagyon sok azonos eloszlású véletlen hatás összege, akkor ez a változó majdnem minden esetben normális eloszlású lesz. Ezt hívjuk **centrális határeloszlás tételének**.

6. előadás

Az elmúlt két előadáson a valószínűségszámítás alapvető fogalmaival ismerkedtünk meg. A mérnök hallgatóban felmerülhet a kérdés, hogy ez a tudás miért jó neki. Mérnökként nem elég a méréseket



21. ábra. Az m várható értékű és σ szórású normális eloszlású valószínűségi változó sűrűség-, illetve eloszlásfüggvényei

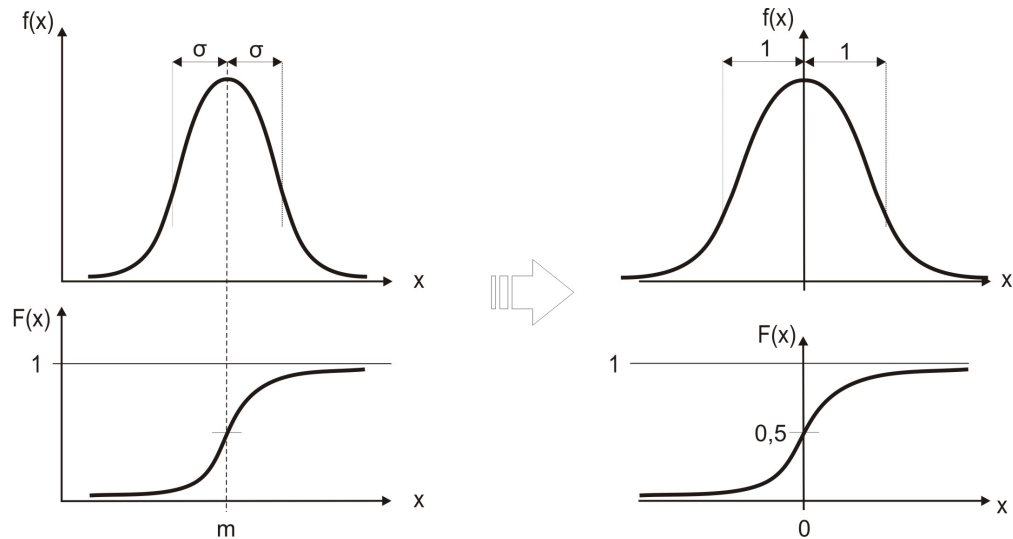
elvégezni, azokat ki is kell értékelni, és valamilyen következtetést levonni a műszer illetve a mérés hibájára vonatkozólag. Ezen az előadáson megismerkedünk a mérések kiértékelésére vonatkozó statisztikai valószínűségszámítás néhány eszközével. Szó esik a rendszeres illetve véletlen hiba fogalmáról, és a közvetlen valamint közvetett mérések hibájáról.

6.1. A mérés (megfigyelés) fogalma (Terméktervező hallgatók számára)

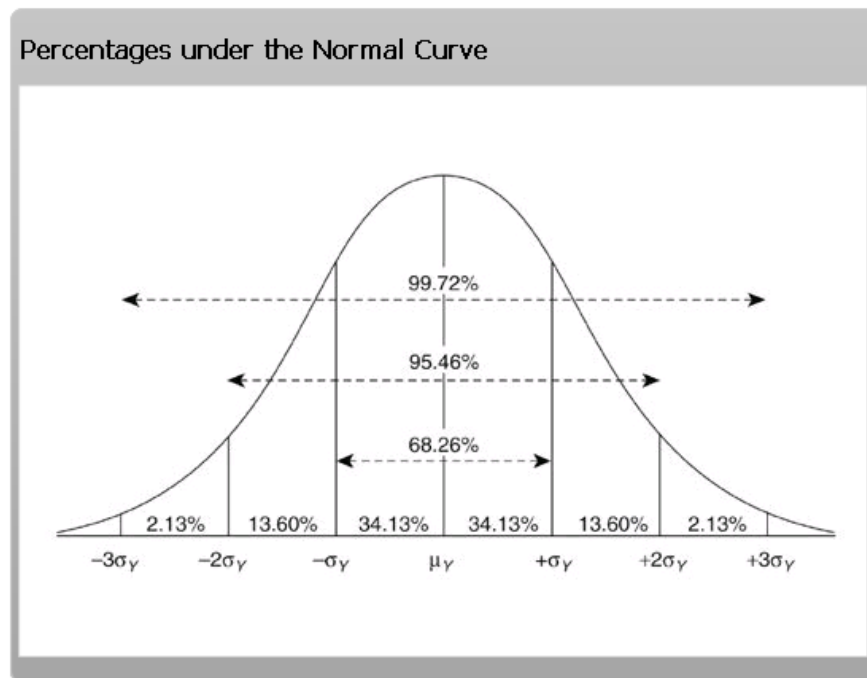
Egy természettudományos mennyiséget a számérték (mérőszám) és a hozzá kapcsolt mértékegység adja meg. Például a tengelyátmérője 20 [mm], vagy Budapest-Nyíregyháza távolság 245 [km]. A mértékegység rendszere Európa nagy részén az SI.

Mérés az a tevékenység, amikor a mennyiség mérőszámát kísérleti úton határozzuk meg, adott mértékegység mellett.

A mérés múltja egészen a Kr.e.1500-as évekig nyúlik vissza. Kairóban például kőből faragott hosszú rudat használtak hosszúság mérésére, és kőből faragott súlyokat a tömegek összehasonlítására. Az időt pedig napórával mérték. Egyiptomban az idő mérésére az úgynevezett kleszidrát is használták. Ez egy agyag tál, melynek az alján egy rés van, az agyagtál oldalába rovátkák vannak vésve. Ezt a tálat vízzel megtöltve vízóra készíthető.



22. ábra. Az ábra jobb oldalán látható a standard normális eloszlás sűrűségfüggvénye és eloszlásfüggvénye



23. ábra. A normális eloszlás sűrűségfüggvénye alatti területek a várható érték körüli egy, kettő illetve három σ sugarú környezetében. (forrás: <https://learn.bu.edu/bbcswebdav/courses/13sprgmetcj702ol/week03/metcj702W03S01T02normal.html>)

Manapság egy adott célnak megfelelő mérőberendezéseket, **modelleket** szokás készíteni, használni. Például egy csapágygolyó átmérőjének megméréséhez egy modellt állítanak fel, melyben feltételezik, hogy az a golyó gömb alakú. Persze ez a valóságban sosem igaz, hiszen egy megmunkált golyó felszínén lehetnek apró rovátkák, karcolások, sosem lesz a golyó tökéletes gömb formájú. A konyhai fiókok tervezése és tesztjei során is elhanyagolásokat tesznek, hiszen nem lehet azok során számításba venni, hogy a való életben hányszor használjuk azt a fiókot nem rendeltetésszerűen. Vagyis a méréssel egy ideális helyzetet mérünk.

A mérés végrehajtásakor először mérési modellt állítunk fel (kimondva vagy kimondatlanul). Pl. csapágygolyó mérésénél feltételezzük, hogy az gömb alakú, tengelycsapnál, hogy az henger alakú. Vagyis a "valóság" bizonyos részeit elhanyagoljuk.

A mérés elvégzéséhez mérőeszközre van szükségünk, amelynek főbb részei: érzékelő, jelfeldolgozó, kijelző (megjelenítő).

A mérési modell felállításakor elhanyagolásokat tettünk, az érzékelés, a jelfeldolgozás és a megjelenítés folyamatát zaj terheli, ezért a mérési folyamat eredménye nem pontosan a keresett érték lesz. Vagyis a mérési eredményt mérési hiba terheli.

Az érzékelés feladata a számérték meghatározása. Ehhez szükséges egy egység, amellyel a mérendő mennyiséget össze lehet hasonlítani. Ezt az egységet nevezzük **etalonnak**. Tehát például etalon az 1 m, 1 km, 1 fényév. Nem minden esetben létezik etalon. A hosszúságra, tömegre van etalonunk, de a hatásfokra már nincsen. A számérték nagyságának meghatározására különböző eljárások léteznek. Egy szabó a méréseit **összehasonlítással** végzi el a mérőszalagja segítségével. **Kitérítéssel** alapján működik egy rugós erőmérő, ahol a deformáció mértékét hozzárendeljük az erő nagyságához. A víz-, villany- és gázóra is kitérítéssel alapján működik. Ezek működéséhez persze szükséges a műszer kalibrálása is, a skála elkészítése, melynek folyamata már a szakma sajátossága. A piacon találkozhatunk **kiegyenlítéses módszerrel**, ahol a zöltséges a gyümölcs tömegét kétkarú mérleggel méri meg. A kiegyenlítéses módszer az alapja számos elektronikus eszköznek is. Amikor a pontos etalonhoz képesti eltérést vizsgáljuk, akkor **különbségmérést** végzünk.

Az etalonnal való összehasonlítást **közvetlen mérésnek** nevezzük. Nem tudunk közvetlen mérést végezni, ha nincsen a birtokunkban etalon, vagy a mért mértékegység kiesik az etalon nagyságrendjéből. Ilyenkor **közvetett mérést** alkalmazunk és ebből vonunk le valamilyen következtetést. Például egy higanyos hőmérőről nem a hőmérsékletet olvashatjuk le, hanem a higanyos oszlop magasságát, amelyhez beállítható a hőmérséklet.

6.2. Közvetlen mérés

Valamilyen termék (pl. tengelycsap) névleges mérete ismert. Ezt az értéket pontosan sohasem tudjuk megvalósítani, mert a gyártást hiba terheli (anyagtechnológiai, gyártási, stb. hibák). A megvalósított méretet nem tudjuk pontosan megmérni a modellezési és mérési hibák miatt. De a mért adatok alapján mégis következtetést kell levonnunk a gyártott termék jóságára vonatkozóan.

Jelölje x a névleges méretet vagy megkívánt értéket, melyet pontosan ismerünk, valamint jelölje ξ valószínűségi változó a mérési eredményt. Ekkor tegyük fel, hogy

$$\xi = x + \varepsilon, \quad (67)$$

ahol ε a mérési vagy gyártási hiba, amely a mérésből vagy a folyamatból adódik. Tegyük fel továbbá, hogy ε Δx várható értékű és σ szórású normális eloszlásból származik, vagyis $\varepsilon \in N(\Delta x, \sigma)$.

Megjegyezzük, hogy ez a feltevés a centrális határeloszlás tétele miatt jogos. ε várható értékét tekintve két esetet különböztethetünk meg. Ha $\Delta x \neq 0$, akkor ε **rendszeres hiba**. Ha pedig $\Delta x = 0$, akkor ε **véletlen hiba**.

6.2.1. Rendszeres hiba

Ha $M(\varepsilon) = \Delta x \neq 0$, akkor ez a hiba, minden mérést ugyanúgy fog terhelni. Így a hiba jelenlétére csak akkor jöhetünk rá, ha a mérést össze tudjuk hasonlítani egy jobb eszközzel vett méréssel. Például, ha egy kopott tolómérővel mérünk, akkor minden mérést ugyanaz a pár tized mm hiba fog terhelni. A rendszeres hiba ismert, ha tudjuk a nagyságát és az előjelét. A hiba kiküszöbölése a kalibrálás, amely után már $\Delta x = 0$. A kalibráció szakterület kérdése, ezzel ennek a tárgynak a keretein belül nem foglalkozunk.

6.2.2. Véletlen hiba

Ha ε véletlen hiba, akkor a várható érték tulajdonságai alapján kiszámítható ξ várható értéke:

$$M(\xi) = M(x + \varepsilon) = M(x) + M(\varepsilon) = x. \quad (68)$$

Összevetve a fenti egyenlőségsorozat elejét és végét, azt mondhatjuk, hogy *a közvetlen mérés kiértékelése a várható érték becslése*.

6.3. Közvetett mérés

Mérési hiba terjedéséről közvetett mérés esetén beszélünk. Ekkor ismerünk valamiféle kapcsolatot a változók között, amiből ki tudjuk számolni a számunkra érdekes értéket. Ezt tesszük például a határfok mérésénél, vagy térfogatszámításkor is. Tehát legyenek x_1, x_2, \dots, x_n a közvetlenül mérhető változók és y ezeknek egy függvényéből megkapható érték:

$$y = f(x_1, x_2, \dots, x_n). \quad (69)$$

Célunk megtudni, hogy hibával terhelt mérések esetén mennyi lesz y hibája. Legyenek a mérési eredményeink $\xi_1, \xi_2 \dots \xi_n$, és tegyük fel, hogy azokat rendre valamilyen hiba terhel. y hibája a mérések hibájától függ.

6.3.1. y rendszeres hibája

Tegyük fel, hogy az egyes komponensek hibáját ismerjük, vagyis ismerjük azok nagyságát és előjelét. Legyenek ezek $\Delta x_1, \Delta x_2, \dots, \Delta x_n$. Felhívjuk a figyelmet arra, hogy Δx_i nem hibakorlát. Jelölje y rendszeres hibáját Δy :

$$\Delta y = f(x_1 + \Delta x_1, x_2 + \Delta x_2, \dots, x_n + \Delta x_n) - f(x_1, x_2, \dots, x_n). \quad (70)$$

Feltéve, hogy f Taylor-sorba fejthető és a Δx_i hibák elég kicsik, akkor a Δy első tagját (x_1, x_2, \dots, x_n) körül sorbafejtve és a lineárisnál magasabb rendű tagokat elhagyva a következőt kapjuk:

$$f(x_1 + \Delta x_1, x_2 + \Delta x_2, \dots, x_n + \Delta x_n) \cong f(x_1, x_2, \dots, x_n) + \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Delta x_i. \quad (71)$$

Ekkor (70) így írható:

$$\Delta y = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Delta x_i. \quad (72)$$

A fenti képlet alapján könnyen meggondolható, hogy y hibája akár 0 is lehet, még akkor is, ha a mérések hibával terheltek. Δx_i együtthatója, $\frac{\partial f}{\partial x_i}$ neve **érzékenységi együttható**. Ha az érzékenységi együttható nagy, akkor a hozzá tartozó hiba nagyobb mértékben befolyásolja y hibáját, ha pedig ez az együttható 1-nél kisebb, akkor csökkenti a neki megfelelő hiba hatását y -ra nézve. Ebből következik, hogy nem mindegy, hogy melyik komponensnek mi a hibája. Speciális esetekben az érzékenységi együttható igen könnyen kiszámolható. Tegyük fel, hogy y a következő alakban kapható az x_i adatok ismeretében:

$$y = C x_1^{k_1} \dots x_n^{k_n}. \quad (73)$$

Ekkor y egy x_i változó szerinti deriváltja:

$$\frac{\partial y}{\partial x_i} = C x_1^{k_1} \dots k_i x_i^{k_i-1} \dots x_n^{k_n} = k_i \frac{y}{x_i}. \quad (74)$$

Ezzel a y hibája így írható:

$$\Delta y = \sum_{i=1}^n k_i \frac{y}{x_i} \Delta x_i, \quad (75)$$

és y relatív hibája:

$$\frac{\Delta y}{y} = \sum_{i=1}^n k_i \frac{\Delta x_i}{x_i}. \quad (76)$$

6.3.2. y véletlen hibája

Tegyük fel, hogy a ξ_i méréseket véletlen ε_i ($i = 1, 2, \dots, n$) hiba terheli:

$$\xi_i = x_i + \varepsilon_i. \quad (77)$$

Mivel ε_i ($i = 1, 2, \dots, n$) véletlen hiba, ezért $\varepsilon_i \in N(0, \sigma_i)$. Tegyük fel továbbá, hogy a ξ_i valószínűségi változók függetlenek egymástól, vagyis az egyik ismerete nem befolyásolja a többi változó ismeretét. Jelölje most y hibáját ε_y :

$$\varepsilon_y = f(\xi_1, \xi_2, \dots, \xi_n) - f(x_1, x_2, \dots, x_n). \quad (78)$$

Most is ugyan úgy járunk el, mint a rendszere hiba esetében, az (78) jobb oldalán álló első tagot (x_1, x_2, \dots, x_n) körül Taylor-sorba fejtjük, és a lineárisnál magasabb rendű tagokat elhanyagoljuk:

$$\varepsilon_y \cong \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{(\xi_1, \xi_2, \dots, \xi_n)} \varepsilon_i. \quad (79)$$

Az egyszerűség kedvéért a szumma mögött álló parciális deriváltat a következőképp fogjuk jelölni $\frac{\partial f}{\partial x_i} \Big|_{(\xi_1, \xi_2, \dots, \xi_n)} = \frac{\partial f}{\partial x_i} \Big|_{\xi_i}$, amely valójában egy valószínűségi változó, ám ez elhanyagolható, és konstansnak tekinthető. Ezzel azt mondhatjuk, hogy ε_y az ε_i véletlen hibák lineáris kombinációja.

Mivel $\varepsilon_i \in N(0, \sigma_i)$, ezért ezek lineáris kombinációja is normális eloszlást fog követni. A várható érték tulajdonságait felhasználva ki tudjuk számolni $M(\varepsilon_y)$ értékét:

$$M(\varepsilon_y) = M\left(\sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{\xi_i} \varepsilon_i\right) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{\xi_i} M(\varepsilon_i) = 0. \quad (80)$$

A második egyenlőség azért igaz, mert elhanyagoltuk az érzékenységi együttható valószínűségi változó voltát, és konstans értéknek tekintjük. Azt pedig már jól tudja a kedves olvasó is, hogy a várható érték lineáris operátor, és így a konstans kiemelhető. ε_y szórásnégyzete is kiszámolható:

$$D^2(\varepsilon_y) = D^2\left(\sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{\xi_i} \varepsilon_i\right) = \sum_{i=1}^n D^2\left(\frac{\partial f}{\partial x_i} \Big|_{\xi_i} \varepsilon_i\right) = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \Big|_{\xi_i}\right)^2 D^2(\varepsilon_i) = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \Big|_{\xi_i}\right)^2 \sigma_i^2. \quad (81)$$

A második egyenlőség most azért igaz, mert a ξ_i valószínűségi változókról feltettük, hogy függetlenek. $D^2(\varepsilon_y)$ értékét szokás **eredő szórásnégyzetnek** nevezni. Mivel a véletlen hibák szórájáról nem tettük fel, hogy 0, ezért az eredő szórásnégyzet sem lesz zérus. Ha csökkenteni szeretnénk $D^2(\varepsilon_y)$ értékét, akkor a nagyobb érzékenységi együtthatóhoz tartozó szórásnégyzetet kell csökkenteni. Ahogyan azt már a rendszeres hibaterjedés tárgyalásánál is láttuk, ha $y = Cx_1^{k_1} \dots x_n^{k_n}$ alakú, akkor a relatív szórásnégyzet:

$$D^2\left(\frac{\varepsilon_y}{y}\right) = \sum_{i=1}^n k_i^2 \left(\frac{\sigma_i}{x_i}\right)^2. \quad (82)$$

5. *Példa:* Egy egyszerű példa a fenti módszer alkalmazására egy henger térfogatának kiszámítása. Egy henger térfogatát a következő módon számíthatjuk ki:

$$V = \frac{d^2 \pi}{4} h. \quad (83)$$

Tehát V egy kétváltozós függvénnyel számítható, a paraméterek pedig a henger alapkörének sugara (d) és a henger magassága (h). Tegyük fel, hogy az átmérőt $\pm 2\%$ -os relatív pontossággal ismerjük, a magasságot pedig $\pm 3\%$ -ossal. Ekkor a henger térfogatának relatív szórásnégyzete:

$$D^2\left(\frac{\varepsilon_y}{y}\right) = 2^2(2\%)^2 + 1^2(3\%)^2 = 25\%. \quad (84)$$

A relatív szóráns ennek a négyzetgyöke, azaz

$$D\left(\frac{\varepsilon_y}{y}\right) = 5\%. \quad (85)$$

Ha ezt az értéket csökkenteni szeretnénk, akkor az átmérő hibáját érdemes csökkenteni, mert az négyzetesen befolyásolja a térfogatot, vagyis a relatív szórásnégyzetben 2^2 az együtthatója.

6.4. Közvetlen mérés kiértékelése

Amint azt már korábban is említettük, közvetlen mérés kiértékelése valójában az átlag hibájának a becslése.

Legyen ξ a mérési eredmények valószínűségi változója, és legyenek a véletlen hibákkal terhelt mérési eredményeink $\xi_1, \xi_2, \dots, \xi_n$. Tehát azt mondhatjuk, hogy $\xi = x + \varepsilon$, ahol $M(\varepsilon) = 0$ és normális eloszlást követ, valamint x a keresett érték. Ekkor $M(\xi_i) = x$, $(i = 1, 2, \dots, n)$. Tegyük fel továbbá, hogy ξ is normális eloszlásból származik. Tekintsük a ξ_i mérések átlagát:

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i. \quad (86)$$

A várható érték tárgyalásánál már láttuk, hogy az átlag várható értéke megegyezik a megfigyelések várható értékével, vagyis $M(\bar{\xi}) = x$, azonban általában $\bar{\xi} \neq x$. Érdekes megvizsgálni azt, hogy vajon az átlag mekkora sugarú környezetébe esik bele a várható érték. Tekintsük az átlag körüli a sugarú intervallumot. Azt szeretnénk meghatározni, hogy mekkora legyen a ahhoz, hogy a várható érték beleessen ebbe az intervallumba, vagyis igaz legyen

$$\bar{\xi} - a \leq M(\xi) \leq \bar{\xi} + a. \quad (87)$$

A fenti intervallumot nevezzük **konfidencia intervallumnak**, és a -t a **konfidencia intervallum sugarának**. Mivel $\bar{\xi}$ valószínűségi változó, az értéke nem fix, ezért ezt az intervallumot sosem tudjuk biztosan meghatározni, ellenben megkívánhatjuk (87) egyenlőtlenség teljesülését bizonyos p valószínűséggel:

$$P(\bar{\xi} - a \leq M(\xi) \leq \bar{\xi} + a) = p. \quad (88)$$

p -t szokás **szignifikancia szintnek** nevezni, ennek értéke adott. Tehát adott szignifikancia szint mellett keressük a konfidencia intervallum sugarát.

Először tegyük fel, hogy ismerjük a ξ valószínűségi változó szórását, legyen ez σ . Ekkor a következő ekvivalens átalakításokat tehetjük meg:

$$P\left(-\frac{a}{\sigma}\sqrt{n} \leq \frac{\bar{\xi} - M(\xi)}{\sigma}\sqrt{n} \leq \frac{a}{\sigma}\sqrt{n}\right) = p. \quad (89)$$

Legyen

$$\eta = \frac{\bar{\xi} - M(\xi)}{\sigma}\sqrt{n} \quad \text{és} \quad \lambda = \frac{a}{\sigma}\sqrt{n} \quad (90)$$

A figyelmes olvasó rögtön észreveheti, hogy η valószínűségi változót a $\bar{\xi}$ standardizálásával kaptuk meg, így $M(\eta) = 0$ és $D^2(\eta) = 1$. Mivel feltettük, hogy ξ normális eloszlású, ezért $\bar{\xi}$ is az, ebből pedig az következik, hogy η standard normális eloszlású. Ekkor (89) így írható:

$$P(-\lambda \leq \eta \leq \lambda) = p. \quad (91)$$

Jelölje a standard normális eloszlás eloszlásfüggvényét Φ (... ábra). Ekkor a $[-\lambda, \lambda]$ intervallumba esés valószínűsége:

$$P(-\lambda \leq \eta \leq \lambda) = \Phi(\lambda) - \Phi(-\lambda) = p. \quad (92)$$

Mivel Φ középpontosan szimmetrikus az $y = 0,5$ -re és $-\infty$ -ben 0-hoz, $+\infty$ -ben 1-hez tart, ezért $\Phi(\lambda) + \Phi(-\lambda) = 1$. Ezzel (92) így írható:

$$2\Phi(\lambda) - 1 = p. \quad (93)$$

Ha p adott, akkor

$$\Phi(\lambda) = \frac{p+1}{2}, \quad (94)$$

amiből az eloszlásfüggvény ismeretében λ kiszámolható. Ekkor a konfidencia intervallum sugara:

$$a = \frac{\lambda\sigma}{\sqrt{n}}. \quad (95)$$

7. előadás

Az előadáson folytatjuk a konfidencia intervallumok tárgyalását. Kimondjuk a Gauss-Markov tételt és felírjuk a Gauss-féle normálegenleteket.

Sajnos általában nem ismerjük a ξ valószínűségi változó szórását. Ilyenkor a szórást a korrigált tapasztalati szórással becsüljük, melynek négyzete:

$$(s^*)^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{\xi} - \xi_i)^2. \quad (96)$$

Ezzel η más lesz:

$$\tilde{\eta} = \frac{\bar{\xi} - M(\xi)}{s^*} \sqrt{n}. \quad (97)$$

Most s^* is a véletlentől függő változó, így $\tilde{\eta}$ nem fog normális eloszlást követni. Ebből pedig az következik, hogy a konfidencia intervallum sugara nem adható meg Φ eloszlás segítségével. Azonban belátható, hogy $\tilde{\eta}$ úgynevezett **Student-eloszlásból** származik. Ennek az eloszlásfüggvénye p szignifikancia szint, és a minta elemszáma segítségével meghatározható. Ekkor a konfidencia intervallum sugara:

$$a = \frac{\lambda_{st}s^*}{\sqrt{n}}, \quad (98)$$

ahol $\lambda_{st} = \lambda(p, n-1)$ a Student-eloszlásból kapható meg.

Ahhoz, hogy a fenti módszer alkalmazható legyen, biztosítani kell olyan feltételeket, amelyek mellett a méréseket csak véletlen hiba terheli. Ekkor a mérések átlaga ($\bar{\xi}$) és korrigált tapasztalati szórásnégyzete ($(s^*)^2$) kiszámolása után egy adott p szignifikanciaszint mellett meg tudjuk határozni $\lambda(p, n-1)$ értékét, amelyből a konfidencia intervallum sugara (a) számolható. Ekkor a mérési eredmény:

$$\bar{\xi} \pm a \quad (p). \quad (99)$$

Vagyis véletlen hibával terhelt mérés esetén a konfidencia intervallum sugara a közvetlen mérés hibakorlátja. Megjegyezzük, hogy ezt a hibát mindig csak korlátok közé tudjuk szorítani.

A mérési eredmény kiértékelésekor általában $a/\bar{\xi}$ értékét szokás tekinteni. Az, hogy mikor fogadjuk el ennek az értékét, adott műszaki területtől függ. Ha úgy döntünk, hogy ezt az értéket nem fogadjuk el, akkor valamelyik paramétert meg kell változtatni. Könnyen meggondolható, hogy ha a mérések számát, azaz n értékét növeljük, akkor a értéke \sqrt{n} szerint csökken. λ_{st} értéke két paramétertől, p -tól és n -től függ. Ha p értékét növeljük, akkor nagyobb biztonsággal szeretnénk dönteni ugyanannyi információ (mérés) birtokában. Ekkor a konfidencia intervallum sugara nagyobb lesz, így λ_{st} értéke is növekszik. Ha a mérések számát, n -et növeljük, akkor a korrigált tapasztalati szórással jobban közelítjük a valódi szórást, így λ_{st} értéke közelebb kerül λ értékéhez, amelyet a standard normális eloszlás eloszlás függvényéből kaphatunk meg.

7.1. Becslések értékelése

Az 5.1-es fejezetben már megmutattuk, hogy az átlag a várható érték becslése. Beláttuk, hogy $M(\bar{\xi}) = M(\xi)$, és $\bar{\xi}$ a keresett mennyiség (a várható érték) körül ingadozik. Ezt a tulajdonságot neveztük **torzítatlan becslésnek**. A torzítatlan becslés általános definíciója a következő. Ha egy a mennyiség becslése α és $M(\alpha) = a$, akkor azt mondjuk, hogy α torzítatlan becslése a -nak.

A továbbiakban a legkisebb négyzetek módszerének alkalmazására vonatkozó általános feltételeket mondunk ki és felírjuk az együtthatók meghatározására alkalmazható Gauss-féle normálegyenleteket.

7.1.1. Gauss-Markov tétel

Legyenek x és y determinisztikus változók. Tegyük fel, hogy teljesülnek a következő feltételek:

1. $y = \sum_{i=1}^n a_i x^i$, ahol n adott természetes szám, a_i valós szám minden $i = 0, 1, 2, \dots, n$ esetén. Vagyis x és y között polinomiális kapcsolat van.
2. x mérését (beállítást) nem terheli mérési hiba. y mérését véletlen hiba terheli (y -t sosem lehet pontosan megmérni). A j . mérési eredmény $\eta_j = y_j + \varepsilon_j$, ekkor $M(\eta_j) = y$.
3. Az η_j mérési eredmények függetlenek egymástól.

Legyen továbbá $[x_j, \eta_j]_{j=1}^N$ a méréssorozat, a legkisebb négyzetek módszerével kapott együtthatók becslései α_i , ($i = 0, 1, 2, \dots, n$) értékek. Ekkor $M(\alpha_i) = a_i$, vagyis α_i torzítatlan becslése a_i -nek.

7.1.2. Gauss-Markov tétel alkalmazása a gyakorlatban

Mérnöki szempontból a következőket mondhatjuk el a Gauss-Markov tétel feltételeiről.

1. A gyakorlatban szinte sosem teljesül, hogy a két változó között polinomiális kapcsolat van.
2. Ez a feltétel problémafüggő, néha teljesíthető. Például, ha idő vagy hely függvényében mérünk, akkor az x változót pontosan tudjuk mérni.
3. A mérési eredmények függetlensége általában teljesül.

A legkisebb négyzetek módszerét ezek függvényében kell alkalmazni. Általában a két változó közötti kapcsolat $y = f(x)$ módon írható le, ám mi $f(x)$ -et egy másik $g(x)$ függvénnyel közelítjük, amelynek a mért adatok ismeretében $g_n(x)$ közelítését határozzuk meg (... ábra). Emiatt valójában g és y között semmilyen kapcsolatot nem ismerünk. Tehát a legkisebb négyzetek módszerének alkalmazása során a mért pontok és $g(x)$ közötti távolságot minimalizáljuk, így biztosítani tudjuk, hogy $g_n(x)$ a pontok között fusson, de matematikailag semmilyen állítást nem tudunk kimondani.

7.1.3. Gauss-féle normálegyenletek

Írjuk fel a legkisebb négyzetek módszerét az $[x_j, \eta_j]_{j=1}^N$ mérésorozat pontjaira polinomiális közelítés esetén:

$$F(\alpha_0, \alpha_1, \dots, \alpha_n) = \sum_{j=1}^N \left(\eta_j - \sum_{i=0}^n \alpha_i x_j^i \right)^2, \quad (100)$$

ezt az összeget szeretnénk minimalizálni az $\alpha_0, \alpha_1, \dots, \alpha_n$ értékek függvényében. Felhívjuk a figyelmet arra, hogy j a mérési pontok indexe, i pedig a közelítő polinom futóindexe. Minimumhely ott lehet, ahol az α_k ($k = 1, 2, \dots, n$) szerinti parciális derivált értéke 0:

$$\frac{\partial}{\partial \alpha_k} \left[\sum_{j=1}^N \left(\eta_j - \sum_{i=0}^n \alpha_i x_j^i \right)^2 \right] = 0. \quad (101)$$

Elvégezve a deriválást:

$$-2 \left[\sum_{j=1}^N \left(\eta_j - \sum_{i=0}^n \alpha_i x_j^i \right) \right] x_j^k = 0. \quad (102)$$

–2-vel való osztás után bontsuk fel a zárójeleket, és az α_i együtthatóktól nem függő tagokat rendezzük a jobb oldalra:

$$\sum_{j=1}^N \sum_{i=0}^n \alpha_i x_j^{i+k} = \sum_{j=1}^N \eta_j x_j^k. \quad (103)$$

Mivel a szummák végesek, ezért a bal oldalon felcserélhetőek:

$$\sum_{i=0}^n \alpha_i \sum_{j=1}^N x_j^{i+k} = \sum_{j=1}^N \eta_j x_j^k. \quad (104)$$

A fenti egyenlőséget mátrixos alakban is fel tudjuk írni. A mátrix sorai k értékei, és oszlopa az i . indexe a szummának.

$$\begin{pmatrix} N & \sum_{j=1}^N x_j & \sum_{j=1}^N x_j^2 & \sum_{j=1}^N x_j^3 & \dots \\ \sum_{j=1}^N x_j & \sum_{j=1}^N x_j^2 & \sum_{j=1}^N x_j^3 & \dots & \\ \vdots & & & & \\ \sum_{j=1}^N x_j^n & \sum_{j=1}^N x_j^{n+1} & \dots & & \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^N \eta_j \\ \sum_{j=1}^N \eta_j x_j \\ \vdots \\ \sum_{j=1}^N \eta_j x_j^n \end{pmatrix} \quad (105)$$

Ezzel egy lineáris algebrai egyenletrendszert kaptunk az α_i együtthatókra. Ha például harmadrendű közelítést szeretnénk kapni akkor ezt az egyenletet kell megoldanunk $k = 3$ esetre. Megjegyezzük, hogy az Excel program is ezt a módszert használja.

8. előadás

Az előadáson a statisztikai próbákról, azon belül is a paraméteres próbákról lesz szó.

8.1. Paraméteres próbák

6. *Példa:* Tekintsük egy élelmiszerbolt polcán sorakozó 1 kg-os kiszerelésű cukor-készletet. A vásárlóban felmerülhet a kérdés, hogy vajon valóban minden zacskó cukor tömege 1 kg-e? Tegyük fel, hogy lehetőségünk van lemérni a polcon levő cukrok aktuális tömegét. A megfigyelésekből le szeretnénk vonni valamilyen statisztikailag megalapozott döntést egy csomag cukor tömegére vonatkozólag. A kérdést matematikailag a következő módon fogalmazhatjuk meg.

Legyenek a mérési eredményeink $\xi_1, \xi_2, \dots, \xi_n$. Ezek $\bar{\xi}$ átlagát számítva meghatározhatunk az átlag körül egy a sugarú konfidencia intervallumot, amely n -től és egy általunk választott p szignifikancia szinttől függ. Legyen $m_0 = 1$ kg az elméleti várható értéke a cukros zacskóknak. Ekkor két lehetőség van.

- Ha $m_0 \in [\bar{\xi} - a; \bar{\xi} + a]$, akkor
 1. vagy valóban $m_0 = 1$ kg a csomagolt cukor tömege,
 2. vagy $m_0 \neq 1$ kg, de a mérésekből mégis az ellenkezőjét kapjuk.
- Ha $m_0 \notin [\bar{\xi} - a; \bar{\xi} + a]$, akkor
 1. vagy $m_0 \neq 1$ kg
 2. vagy $m_0 = 1$ kg, de a mérésekből mégis azt kapjuk, hogy nem.

A fenti módszer egy matematikai eszköz arra, hogy eldöntsük, hogy egy adott kérdésre megfigyelések alapján milyen választ adjunk. Ennek a módszernek az általánosítása az **U-próba**.

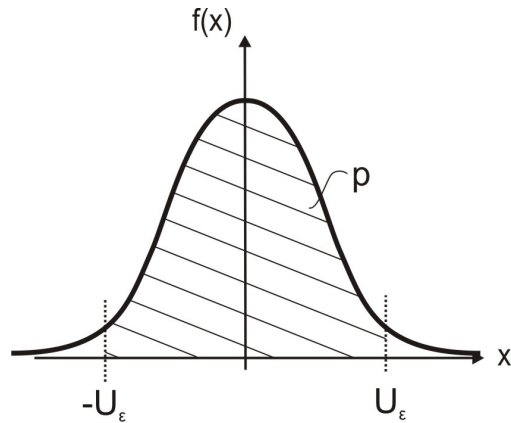
8.1.1. U-próba

Legyenek a ξ változóra vonatkozó megfigyelések $\xi_1, \xi_2, \dots, \xi_n$. Tegyük fel, hogy ξ ismert σ szórású normális eloszlású valószínűségi változó. Ekkor felmerülhet kérdésként, hogy vajon ξ várható értéke egy adott a_0 szám e. Ezt nevezzük **nullhipotézisnek**, melyet H_0 -al szokás jelölni. Vezessük be a következő U_{akt} változót, melyet a próba **aktuális értékének** nevezünk.:

$$U_{akt} = \frac{\bar{\xi} - a_0}{\sigma} \sqrt{n}. \quad (106)$$

Ekkor, ha ξ várható értéke a_0 , akkor $M(U_{akt}) = 0$ és $D^2(U_{akt}) = 1$, vagyis $U_{akt} \in N(0, 1)$, hiszen (106) éppen az átlag standardizálása. Ismerve a standard normális eloszlás sűrűségfüggvényét, egy adott p szignifikancia szinthez meg tudunk határozni egy intervallumot (24. ábra). Ha a H_0 hipotézisünk igaz, akkor p valószínűséggel U_{akt} ebben az intervallumban helyezkedik el. Fordítva a következőket mondhatjuk:

- ha $U_{akt} \in [-U_\varepsilon, U_\varepsilon]$, akkor a statisztikai adatok nem mondanak ellent H_0 -nak. Ekkor azt mondjuk, hogy **elfogadjuk** H_0 -t.
- ha $U_{akt} \notin [-U_\varepsilon, U_\varepsilon]$, akkor a statisztikai adatok ellent mondanak H_0 -nak. Ekkor azt mondjuk, hogy **elutasítjuk** H_0 -t.



24. ábra. A sűrűségfüggvény alatti terület azt adja meg, hogy a hozzá tartozó valószínűségi változó mekkora valószínűséggel esik az adott intervallumba.

8.2. A statisztikai próbák általános menete

Egy statisztikai próba célja valamely adott kérdésre statisztikai alapon választ adni, azt eldönteni. Például két különböző tolómérővel ugyanazon méréseket elvégezve, kérdés lehet, hogy a tolómérők ugyanolyanok-e? A válasz megadásához a kérdésre vonatkozólag megfigyeléseket kell tennünk, és megfogalmazzunk egy nullhipotézist. A kérdést statisztikailag át kell fogalmazzunk, és meg kell határozni a statisztika eloszlását. Ekkor választhatunk egy p szignifikanciaszintet, amelyhez az eloszlásból meghatározhatóak U_ε és $-U_\varepsilon$ határok. Ezek után a statisztika aktuális értékét meghatározva el lehet dönteni, hogy a nullhipotézist elfogadjuk-e vagy sem.

A statisztikai próbák általános menete lépésekben a következő:

1. Fogalmazzuk meg a kérdést.
2. Végezzünk megfigyeléseket.
3. Állítsuk fel a nullhipotézist.
4. Készítsünk statisztikát, amely segítségével meg lehet válaszolni a kérdést. Határozzuk meg a statisztika eloszlását.
5. p szignifikancia szinthez az eloszlásból határozzuk meg U_ε és $-U_\varepsilon$ határokat.
6. Számoljuk ki a statisztika U_{akt} aktuális értékét.
7. Döntsük el, hogy $U_{akt} \in [-U_\varepsilon, U_\varepsilon]$ igaz-e. Ha igaz, akkor H_0 -t elfogadjuk, ha pedig hamis, akkor elutasítjuk.

Megjegyezzük, hogy a 4. és 5. pontban foglaltak nem mérnöki feladatok, ezeket a mérnök csak alkalmazza. Így ennek az előadásnak a keretein belül nem törekszünk a próbák részletes, kimerítő tárgyalására, azokat inkább csak felsorolásszerűen ismertetjük az olvasóval.

Kérdésként merülhet fel, hogy milyen hibákat véthetünk amikor H_0 -t elfogadjuk vagy elutasítjuk. Valójában sosem tudjuk biztosan, hogy egy nullhipotézis igaz-e vagy sem. Ezért érdemes szót ejteni arról is, hogy mekkora annak a valószínűsége, hogy hibásan döntünk. Ha a H_0 hipotézisünk

	H_0 igaz	H_0 hamis
elfogadjuk	✓	másodfajú hiba
elutasítjuk	elsőfajú hiba	✓

igaz, de mi mégis elutasítjuk, akkor **elsőfajú hibát** vétünk. Ennek a valószínűsége $1 - p$. Ha H_0 hamis, de mi elfogadjuk, akkor **másodfajú hibát** követünk el. Ennek a valószínűségét nem tudjuk meghatározni, hiszen ekkor nem ismerjük a valódi eloszlást. A másodfajú hiba valószínűsége függ attól is, hogy mennyire rossz a nullhipotézisünk.

7. *Példa*: Tekintsük a 6. példában szereplő feladatot. Legyen a megfigyeléseink átlaga $\bar{\xi} = 0,98$ kg, és a szórás $\sigma = 0,05$ kg. Legyen $n = 25$ a megfigyelések száma. A H_0 nullhipotézisünk pedig az, hogy a megfigyelések várható értéke $m_0 = 1$ kg. Először kiszámoljuk az aktuális értéket.

$$U_{akt} = \frac{\bar{\xi} - m_0}{\sigma} \sqrt{n} = \frac{0,98 - 1}{0,05} \sqrt{25} = -2.$$

Ha $p = 95\%$, akkor $U_\varepsilon = 1,96$, és $U_{akt} \notin [-U_\varepsilon, U_\varepsilon]$, vagyis a nullhipotézisünket elvetjük. Ha $p = 99\%$, akkor $U_\varepsilon = 2,58$, és $U_{akt} \in [-U_\varepsilon, U_\varepsilon]$, és a nullhipotézist elfogadjuk. Vagyis nem célszerű a szignifikancia szintet emelni, ha biztosabb döntést szeretnénk hozni. Érdemesebb inkább több megfigyelést végezni, és újra meghatározni az aktuális értéket.

8.3. t-próba

Ha nem ismerjük a szórást, akkor nem U-próbát, hanem **t-próbát** alkalmazunk. Legyen $\xi \in N(a_0, \sigma)$, és a megfigyeléseink $\xi_1, \xi_2, \dots, \xi_n$. Mivel a korrigált tapasztalati szórás a szórás torzítatlan becslése, ezért σ -t s^* -al becsüljük. A nullhipotézisünk pedig az, hogy a megfigyelések $\bar{\xi}$ átlagának a várható értéke a_0 adott szám. Az aktuális értéket a következő módon számoljuk:

$$t_{akt} = \frac{\bar{\xi} - a_0}{s^*} \sqrt{n} \quad (107)$$

Ekkor t_{akt} Student-eloszlást követ. Ezek után egy választott p szignifikancia szinthez meghatározzuk a $-t_\varepsilon$ és t_ε határokat. t_ε értéke p -tól és n -tól függ, ezt vagy táblázatból olvashatjuk ki, vagy számítógép segítségével határozhatjuk meg. A H_0 nullhipotézisünket elfogadjuk, ha $t_{akt} \in [-t_\varepsilon, t_\varepsilon]$.

8.4. Kapcsolt t-próba

Tegyük fel, hogy ugyanazon a darabon mérünk két különböző műszerrel. Kérdés lehet, hogy vajon a két műszer ugyanúgy működik-e. Legyenek a mérési eredményeink az első műszerrel $\xi_1, \xi_2, \dots, \xi_n$, a második műszerrel pedig $\eta_1, \eta_2, \dots, \eta_n$. Ekkor a nullhipotézisünk: $M(\xi) = M(\eta)$. A két változó nem független, mert a két műszerrel ugyanazt mértük, ezért azt nem tudjuk vizsgálni, hogy a várható értékeik megegyeznek-e. Ehelyett tekintsük a mérések különbségeit, legyen $\delta_i = \xi_i - \eta_i$,

$i = 1, 2, \dots, n$. Ekkor $M(\delta) = 0$ az új nullhipotézisünk. Erre már alkalmazhatunk t-próbát. Az aktuális érték

$$t_{akt} = \frac{\bar{\delta} - 0}{s_{\delta}^*} \sqrt{n} \quad (108)$$

Megjegyezzük, hogy a gyógyszeriparban, amikor azt vizsgálják, hogy van-e szignifikáns különbség két gyógyszer hatása között, akkor is ezt a próbát alkalmazzák.

8.5. Kétmintás U-, illetve t-próba

Kétmintás U, vagy t-próbát alkalmazunk akkor, ha a két megfigyelés sorozatunk független egymástól, és azt szeretnénk megtudni, hogy azok várható értéke megegyezik-e. Ezt a próbát a jegyzetben nem tárgyaljuk, a hallgató találkozott vele gyakorlaton.

8.6. Grubbs próba

Felmerülhet a kérdés, hogy egy mérésorozat legkisebb vagy legnagyobb eleme ugyanabból az eloszlásból származik-e, mint a mérés többi eleme. Legyenek most is a mért adatok $\xi_1, \xi_2, \dots, \xi_n$ és $\xi \in N(a_0, \sigma)$. Legyen a ξ_{min} a mért adatok legkisebb, illetve ξ_{max} a legnagyobb eleme. Ekkor $\bar{\xi}$ és s^* meghatározása után számítsuk ki, hogy a legkisebb (legnagyobb) elem és az átlag eltérése hányszorosa a szórásnak. Vagyis az aktuális érték:

$$g_{akt} = \frac{\bar{\xi} - \xi_{min}}{s^*}. \quad (109)$$

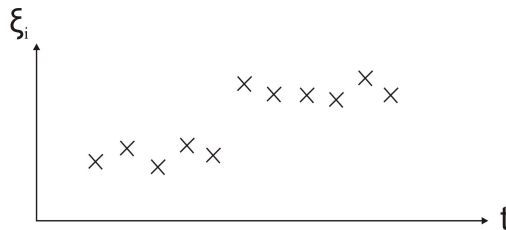
A g_{krit} kritikus értéket p szignifikancia szinttől és n darabszámtól függően táblázatból olvashatjuk ki. Ha $g_{akt} < g_{krit}$, akkor elfogadjuk a nullhipotézisünket, miszerint $\xi_{min} \in N(a_0, \sigma)$. A maximum elemre az aktuális érték:

$$g_{akt} = \frac{\xi_{max} - \bar{\xi}}{s^*}. \quad (110)$$

A döntést ugyanúgy hozzuk, mint a minimum értéknél.

8.7. Abbe próba

Egy mérés során valami történhet a műszerrel, megváltozhatnak a mérési körülmények. Kérdés lehet, hogy minden ξ_i , ($i = 1, 2, \dots, n$) mérés ugyanabból az eloszlásból származik-e.



25. ábra. t időpontban mért ξ_i mérések ábrázolva.

Tegyük fel tehát, hogy $\xi \in N(a_0, \sigma)$, és a mérési eredmények $\xi_1, \xi_2, \dots, \xi_n$. A nullhipotézisünk pedig

az, hogy minden megfigyelésnek ugyanaz a várható értéke, vagyis $M(\xi_i) = a_0$, ($i = 1, 2, \dots, n$). Azt vizsgáljuk, hogy változott-e a várható érték a mérés során. A korrigált tapasztalati szórásnégyzet helyett tekintsünk egy olyan összeget, amelyben az egymás melletti megfigyelések eltérését nézzük:

$$q^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (\xi_{i+1} - \xi_i)^2. \quad (111)$$

Felhívjuk a figyelmet arra, hogy most számít a megfigyelések sorrendje. Az aktuális érték legyen q^2 és s^{*2} hányadosa:

$$r_{akt} = \frac{q^2}{s^{*2}}. \quad (112)$$

Ha nem változik meg a várható érték a mérés során, akkor q^2 és s^{*2} nem nagyon térnek el egymástól. Ha pedig a mérés során a várható értékben egy ugrás tapasztalható, akkor q^2 -ben csak egy tag értéke változik meg, míg s^{*2} -ben minden tag megváltozik, hiszen az átlag is megváltozik. Ekkor r_{akt} értéke csökken. A próbához tartozó kritikus érték ugyancsak táblázatból olvasható ki. A nullhipotézisünket akkor fogadjuk el, ha $r_{akt} > r_{krit}$.

8.8. F-próba

F-próba, vagy más néven Fisher próba során két megfigyelés sorozat szórásnégyzetét hasonlítjuk össze. Legyen az egyik megfigyelés sorozatunk $\xi_1, \xi_2, \dots, \xi_n$ és ismeretlen szórásnégyzete σ_ξ^2 . A második méréssorozat legyen $\eta_1, \eta_2, \dots, \eta_m$ és ismeretlen szórásnégyzete σ_η^2 . Tegyük fel továbbá, hogy ξ és η normális eloszlásból származnak. A nullhipotézisünk az, hogy a két szórásnégyzet megegyezik, vagyis $\sigma_\xi^2 = \sigma_\eta^2$. Az aktuális érték meghatározásához számítsuk ki a korrigált tapasztalati szórásokat:

$$s_{\xi}^{*2} = \frac{1}{n-1} \sum_{i=1}^n (\bar{\xi} - \xi_i)^2 \quad \text{és} \quad s_{\eta}^{*2} = \frac{1}{m-1} \sum_{j=1}^m (\bar{\eta} - \eta_j)^2 \quad (113)$$

Ekkor az aktuális érték:

$$F_{akt} = \max \left\{ \frac{s_{\xi}^{*2}}{s_{\eta}^{*2}}; \frac{s_{\eta}^{*2}}{s_{\xi}^{*2}} \right\}. \quad (114)$$

A kritikus értéket p szignifikancia szinthez, $n-1$ és $m-1$ függvényében táblázatból olvashatjuk ki. A nullhipotézisünket elfogadjuk, ha az aktuális érték kisebb, mint a kritikus érték.

9. 9. előadás

Az előadáson a binomiális eloszlást tárgyaltuk, majd azt alkalmazva a minőség-ellenőrzés alapvető fogalmaival, módszereivel ismerkedtünk meg. Az előadás anyagához tartozó diasor a tárgy honlapjáról letölthető: <http://www.hds.bme.hu/letoltesek/targyak/BMEGEVGAG14/minell.pdf>

10. előadás

Az eddigiekben olyan próbákat tárgyaltunk ahol a nullhipotézisünk egy paraméterre vonatkozott. Most olyan próbákról lesz szó, ahol az eloszlásra irányul a vizsgálatunk. Ilyenkor három

fő típusa van a próbáknak, ezek a **nemparaméteres próbák**. Az első az **illeszkedésvizsgálat**, ahol a minta tapasztalati eloszlásfüggvényét hasonlítjuk össze egy elméleti eloszlásfüggvénnyel, azt vizsgáljuk, hogy mennyire illeszkedik egymásra a kettő. A második a **homogenitásvizsgálat**, ahol azt vizsgáljuk, hogy két minta vagy mérésorozat vajon ugyanolyan eloszlásból származik-e. A harmadik típus a **függetlenség-vizsgálat**, ahol arra szeretnénk választ kapni, hogy két változót tekinthetünk-e függetlennek.

Mindhárom típusra az úgynevezett χ^2 -**próbát** szokták leggyakrabban alkalmazni. Illeszkedésvizsgálatra kisebb minta elemszám esetén szokás még **Ryan-Joiner tesztet** használni.

10.1. Illeszkedésvizsgálat χ^2 -próbával

A próba lényegét és elvét egy bevezető példán keresztül mutatjuk be.

8.Példa: Tekintsünk egy hat oldalú dobókockát és döntsük el, hogy szabályos-e vagy sem. Ha szabályos a kocka, akkor azt várjuk, hogy elég sokszor feldobva nagyjából ugyanannyiszor kapjuk mind a hat számot. Tekintsük a következő dobássorozatot, ahol a kockát $N = 840$ -szer dobtuk fel, és összeszámoltuk, hogy melyik számot hányszor kaptuk. Számoljuk ki azt is, hogy ha szabályos lenne a kocka, hányszor kellett volna kapnunk az egyes értékeket. Az értékeket az alábbi táblázatban gyűjtöttük össze.

A dobás eredménye (i)	gyakoriság (ν_i)	elméleti valószínűség (p_i)	elméleti gyakoriság (Np_i)
1	124	1/6	140
2	152	1/6	140
3	130	1/6	140
4	148	1/6	140
5	152	1/6	140
6	134	1/6	140

Tekintsük a következő összeget, amelyet a χ^2 -próba aktuális értékének nevezünk:

$$\chi_{akt}^2 = \sum_{i=1}^6 \frac{(Np_i - \nu_i)^2}{Np_i}. \quad (115)$$

Ha a kocka szabályos, akkor $Np_i \approx \nu_i$, és χ_{akt}^2 kicsi. A próbához tartozó kritikus értéket a χ^2 -eloszlásból lehet meghatározni p szignifikancia szint, és r osztály esetén: $\chi_{krit}^2(1 - p; r - 1)$. A példánkban $r = 6$, mivel 6 kimenetele lehet egy hat oldalú kockával való dobásnak. A nullhipotézist akkor fogadjuk el, ha $\chi_{akt}^2 < \chi_{krit}^2$. A példánkban $p = 95\%$ -os szignifikancia szinthez az aktuális érték $\chi_{akt}^2 = 5,29$ és $\chi_{krit}^2 = 11,1$, vagyis elfogadható a nullhipotézisünk, mely szerint a dobókockánk szabályos.

10.1.1. Alkalmazás: normalitásvizsgálat

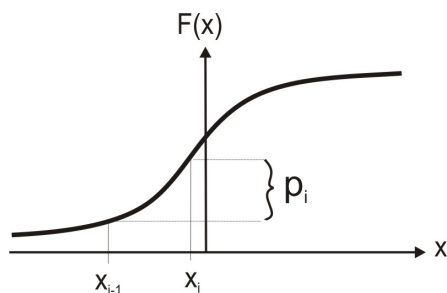
Gyakran van szükség arra, hogy egy megfigyeléssorozatról meghatározzuk, hogy normális eloszlásból származik-e vagy sem. Ehhez legyenek a megfigyelések $\xi_1, \xi_2, \dots, \xi_N$. Ezeket először soroljuk r osztályokba (intervallumokba) és határozzuk meg az osztályba (intervallumba) esés gyakoriságát. Majd a normális eloszlás eloszlásfüggvényéből határozzuk meg, hogy mekkora valószínűséggel esne

egy megfigyelés az adott intervallumokba, ha valóban normális eloszlásból származna. Ha ismert az $F(x)$ eloszlásfüggvény, akkor az egy x helyen azt mondja meg, hogy mekkora annak a valószínűsége, hogy a változónk x -nél kisebb értéket vesz fel. Így az $[x_{i-1}; x_i]$ intervallumba esés p_i valószínűsége az intervallum két végpontjában felvett függvényérték különbsége (26. ábra), vagyis:

$$p_i = F(x_i) - F(x_{i-1}). \quad (116)$$

A gyakorlat szempontjából érdemes az alábbi módon táblázatot készíteni.

Intervallum	gyakoriság	eloszlásfüggvény	p_i	$\frac{(Np_i - \nu_i)^2}{Np_i}$
$-\infty - x_1$	ν_1	$F(x_1)$	$F(x_1)$	
$x_1 - x_2$	ν_2	$F(x_2)$	$F(x_2) - F(x_1)$	
$x_2 - x_3$	ν_3	$F(x_3)$	$F(x_3) - F(x_2)$	
\vdots	\vdots	\vdots	\vdots	
$x_r - +\infty$	ν_r	$F(x_r)$	$1 - F(x_r)$	



26. ábra. Az eloszlásfüggvény ismeretében kiszámítható az $[x_{i-1}, x_i]$ intervallumba való esés valószínűsége.

Ha meghatároztuk az egyes intervallumokba esések valószínűségét, akkor meg tudjuk határozni a χ^2 -próba aktuális értékét. Számoljuk ki minden intervallumra a (115) egyenlőség jobb oldalán álló szumma megfelelő tagját, majd ezeket adjuk össze.

Érdemes úgy megválasztani az osztályokat, az intervallumok határait, hogy azokban a megfigyelések gyakorisága közel megegyezzen. Emlékeztetünk arra, hogy amikor hisztogramot készítettünk, akkor is így osztottuk fel intervallumokra a megfigyelési tartományt. Megjegyezzük továbbá, hogy $F(x_i)$ értékét pl. Excelben könnyen meg lehet határozni. A próba kritikus értékét p szignifikancia szint választása után az osztályok r számának megfelelően határozzuk meg: $\chi_{krit}^2(1 - p; r - 1)$.

10.2. Homogenitásvizsgálat χ^2 -próbával

Homogenitásvizsgálat esetén az a kérdés, hogy két kísérletsorozat tekinthető-e ugyanolyannak. Például, ha egy terméket éjszakai és nappali műszakban is előállítanak, akkor van-e különbség a két műszak teljesítménye között. Legyen ξ és η két valószínűségi változó, eloszlásfüggvényeik rendre $F(x)$ és $G(x)$. Kérdés, hogy $F(x)$ és $G(x)$ megegyeznek-e. Tehát a nullhipotézisünk az, hogy $F(x) = G(x)$. Legyen a két változóra tett megfigyelés sorozataink $\xi_1, \xi_2, \dots, \xi_n$ és $\eta_1, \eta_2, \dots, \eta_m$.

Felhívjuk a figyelmet arra, hogy nem szükséges mind a két változóra ugyanannyi megfigyelést tennünk.

χ^2 -próba esetén a változókra vonatkozó osztályokat készítünk és megnézzük mindkét változóra az osztályokba esés relatív gyakoriságát. Ha ezek a relatív gyakoriságok nagyjából azonosak, akkor elfogadjuk a nullhipotézisünket. A próba menetét egy példán keresztül mutatjuk be.

9. *Példa:* Vizsgáljuk meg egy narniai kristály törőszilárdságát éjszaka és nappal. A mért adatokat az alábbi táblázat tartalmazza: Jelen esetben $n = 120$, $m = 100$. Homogenitásvizsgálat esetén χ_{akt}^2

$\sigma_t 10^3 \frac{N}{m^2}$	$\nu_i(\xi)$	$\mu_i(\eta)$
10-15	36	28
16-18	41	36
19-21	28	22
22-25	11	8
26-29	3	4
30-	1	2

értékét a következő képlet adja:

$$\chi_{akt}^2 = nm \sum_{i=1}^r \frac{(\frac{\nu_i}{n} - \frac{\mu_i}{m})^2}{\frac{\nu_i}{n} + \frac{\mu_i}{m}}. \quad (117)$$

A példánkban $r = 6$ az intervallumok vagy osztályok száma, $\chi_{akt}^2 = 1,19$, $\chi_{krit}^2 = 11,1$. Mivel $\chi_{akt}^2 = 1,19 < \chi_{krit}^2 = 11,1$, ezért a nullhipotézist elfogadjuk.

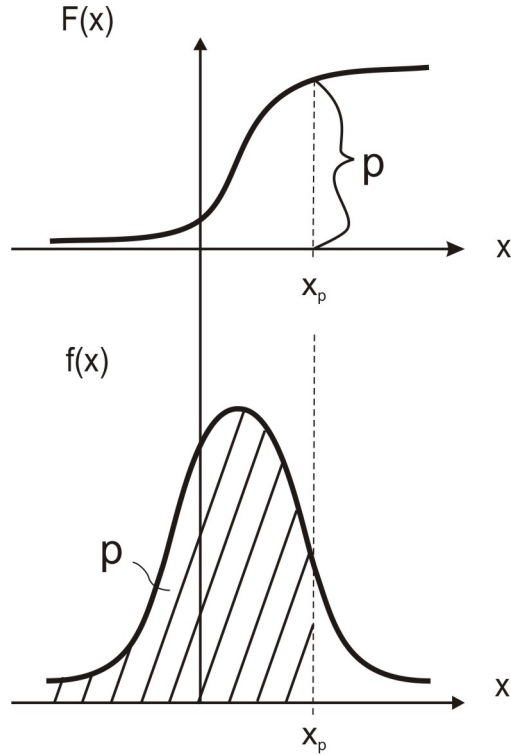
10.3. Illeszkedés vizsgálat Ryan-Joiner teszttel

Kisebb minta elemszám esetén illeszkedés vizsgálatra a Ryan-Joiner tesztet is szokták alkalmazni. A próba lényege, hogy a megfigyelésekből felrajzolt tapasztalati eloszlásfüggvényt hasonlítja össze az elméleti eloszlásfüggvénnyel. Ha elég jól illeszkedik a két függvény, akkor elfogadjuk a nullhipotézist. A próba megismertetése előtt szükségünk van egy fogalom bevezetésére. Tekintsük a ξ valószínűségi változót és annak $F(x)$ eloszlásfüggvényét.

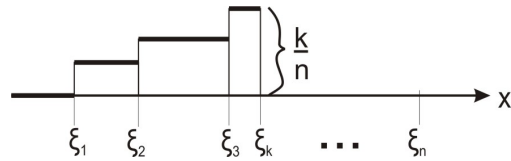
Egy $F(x)$ eloszlásfüggvény **p -percentilise** az az x_p szám, melyre $p = F(x_p)$. Ezt az 27. ábrán szemléltetjük. Érdekes kicsit jobban elmerülni a p -percentilis fogalmában. p annak a valószínűsége, hogy ξ x_p -nél kisebb értéket vesz fel. Ezt a sűrűségfüggvénnyel is meg tudjuk fogalmazni, hiszen a sűrűségfüggvény alatti terület éppen a valószínűséget adja meg:

$$p = F(x_p) = \int_{-\infty}^{x_p} f(x) dx. \quad (118)$$

A próba elvégzéséhez szükségünk van a tapasztalati eloszlásfüggvényre is. Emlékeztetünk arra, hogy ezt úgy készítettük el, hogy a megfigyeléseket sorba rendeztük, és egy olyan lépcsős függvényt rajzoltunk fel, amelynek minden lépcsője $1/n$ magas volt, és a k . megfigyelés feletti lépcső éppen k/n magas (28. ábra). Tekintsük a tapasztalati eloszlásfüggvény egy lépcsőjét, és rajzoljuk mellé az elméleti eloszlásfüggvényt is. Ez a 29. ábrán látható. Az ábrán alkalmazott jelölések mellett a k . lépcső metszéspontjához tartozó $x_{k/n}$ érték lesz a k/n -percentilis. Ilyen módon minden



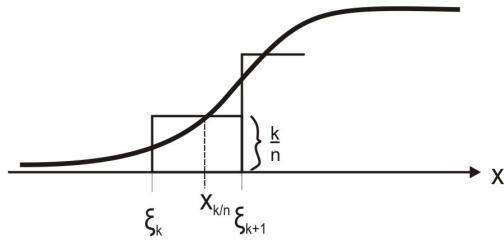
27. ábra. Az eloszlásfüggvény és a sűrűségfüggvény kapcsolata.



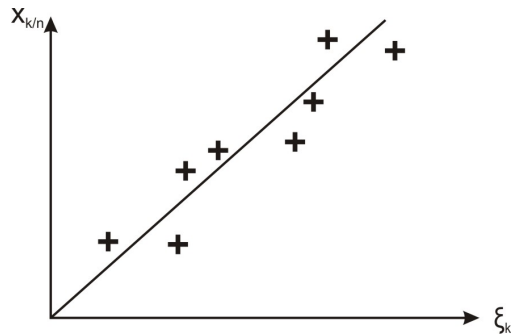
28. ábra. Tapasztalati eloszlásfüggvény.

lépcsőhöz meg tudjuk határozni az elméleti eloszlásból a hozzá tartozó percentilis értékét. Így generálható egy adatsor, amely az elméleti eloszlásból származik és ugyanannyi eleme van, mint amennyi megfigyelésünk. Azt, hogy ez a generált adatsor mennyire hasonlít a méréseinkre a korrelációs együttható mondja meg. Ha ezeket pontpároknak tekintjük és ábrázoljuk a pontokat egy koordináta-rendszerben (30. ábra), akkor ezek egy egyenes körül ingadoznak, ha a korrelációs együttható értéke elég nagy. A hozzáértő szakemberek megmutatták, hogy a p -percentilist k/n helyett jobb $\frac{k-0,375}{n+0,25}$ értékeknél számolni.

Tehát a próba aktuális értéke a $Korr(x_{p_k}, \xi_k)$ korrelációs együttható lesz. A kritikus értéket táblázatból lehet meghatározni. A próbát akkor fogadjuk el, ha az aktuális érték nagyobb, mint a kritikus érték. Megjegyezzük továbbá, hogy ez a próba nem csak normalitásvizsgálatra használható.



29. ábra. Az elméleti eloszlásfüggvény és a tapasztalati eloszlásfüggvény k . lépcsője. A k . lépcső metszéspontjához tartozó $x_{k/n}$ érték a k/n -percentilis.



30. ábra. A Ryan-Joiner módszerrel generált adatsort ábrázolása a mért adatok függvényében.

10.4. Szórásanalízis

Sok esetben szükség van arra, hogy egy termék különböző típusairól objektív eszközökkel el tudjuk dönteni, hogy azok különbözőek-e. Például megeshet, hogy különböző gépkocsi típusokról szeretnénk eldönteni, hogy azok fogyasztása minőségileg különbözik-e. Ezzel a kérdéskörrel a statisztikában a szórásanalízis foglalkozik.

Szórásanalízis esetén nem mennyiségi, hanem minőségi változókról kell döntést hozni. A módszer ismertetése előtt szükséges bevezetnünk néhány fogalmat:

- **Faktor:** a független változó, melynek a hatását kívánjuk vizsgálni.
- **Szint:** a faktor egy hatása.
- **Kísérleti eredmény:** a függő változó.

Tehát a fenti példában az autók a faktorok, a gépkocsik típusai pedig a faktor szintjei. A kérdésünk pedig az, hogy a különböző szinteknek van-e hatása a végeredményre. A válaszhoz egy statisztikai próbát fogunk elvégezni. Ezen a kurzuson feltesszük, hogy csak 1 faktorunk van. Először megfogalmazzuk a nullhipotézisünket, és utána ismertetjük a probléma matematikai megfogalmazását.

A nullhipotézis tehát:

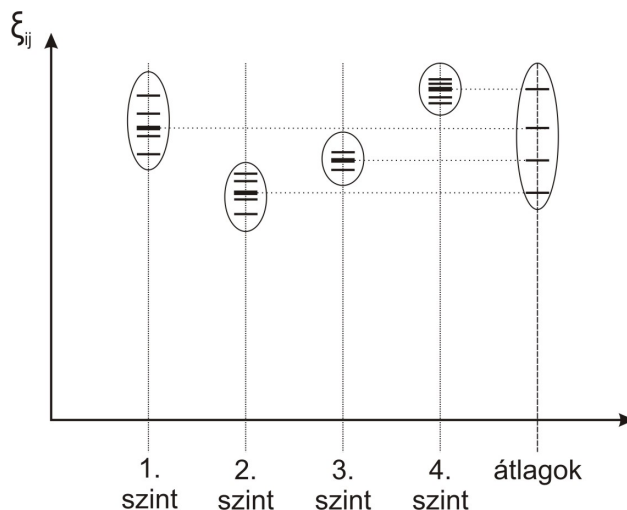
H_0 : A szintek hatása megegyezik.

Mely ekvivalens megfogalmazása:

H_0 : A szintek várható értéke megegyezik.

Vegyük észre, hogy két szint esetén épp kapcsolt T-próbát kellene alkalmazni. Ennek megfelelően több szint esetén szükségünk van egy általánosított T-próbára, melyet a következőkben mutatunk be.

Vezessük be a következő jelölést. Legyen az i . szinten mért j . mérési eredmény $\xi_{i,j}$. A szintek száma k és az i . szinten a mérések száma n_i . Vagyis például az 5. szinten mért adatok: $\xi_{5,1}, \xi_{5,2}, \dots, \xi_{5,n_5-1}, \xi_{5,n_5}$. A próba elvégzéséhez azzal a feltétellel kell élnünk, hogy az egyes szinteken tett mérések normális eloszlásból származnak, és az elméleti szórások megegyeznek.



31. ábra. Egyfaktoros kísérlet.

Az általánosított T-próbát két különböző szórás összehasonlításával fogjuk elvégezni. A 31. ábrán egy egyfaktoros kísérlet látható. Négy szinten végeztünk megfigyeléseket, ezek eredményei vékony vízszintes vonalakkal vannak ábrázolva. Számoljuk ki minden szinten az átlagokat, ezek az utolsó oszlopban láthatóak. Nézzük meg, hogy az átlagok szórása hogyan viszonyul az egyes szintek szórásához. Ha sokkal nagyobb a szórás, akkor a szinteknek van jelentősége. Ha kicsi az átlagok szórása, akkor az a faktorszintek szórásával magyarázható.

A fenti gondolatmenetet fogalmazzuk most meg matematikailag. Számoljuk ki az i . szinten ($i = 1, \dots, k$) a mérések $\bar{\xi}_i$ átlagát és s_i^* korrigált tapasztalati szórásnégyzetét:

$$\bar{\xi}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \xi_{i,j} \quad \text{és} \quad (s_i^*)^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\xi_{i,j} - \bar{\xi}_i)^2. \quad (119)$$

A teljes minta mérete $n = \sum_{i=1}^k n_i$. Ezt felhasználva az átlagok súlyozott átlaga:

$$\bar{\xi} = \sum_{i=1}^k \frac{n_i}{n} \bar{\xi}_i = \frac{1}{n} \sum_{i=1}^k n_i \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \xi_{i,j} \right) = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} \xi_{i,j}, \quad (120)$$

és az átlagok súlyozott korrigált tapasztalati szórásnégyzete:

$$(s_{\acute{a}tl}^*)^2 = \frac{1}{k-1} \sum_{i=1}^k n_i (\bar{\xi}_i - \bar{\xi})^2. \quad (121)$$

A teljes szórás pedig, ami a szinteken belüli szórás:

$$(s_{szint}^*)^2 = \frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (\xi_{i,j} - \bar{\xi})^2 = \frac{1}{n-k} \sum_{i=1}^k (n_i - 1) (s_i^*)^2. \quad (122)$$

Ez utóbbi két szórásnégyzetből kapjuk meg az aktuális értéket:

$$F_{akt} = \frac{(s_{\acute{a}tl}^*)^2}{(s_{szint}^*)^2} = \frac{(n-k) \sum_{i=1}^k n_i (\bar{\xi}_i - \bar{\xi})^2}{(k-1) \sum_{i=1}^k (n_i - 1) (s_i^*)^2}. \quad (123)$$

Mivel két szórást hasonlítunk össze, a próba kritikus értéke az F-próbánál is alkalmazott Fisher-eloszlásból számolandó:

$$F_{krit} = F^{-1}(1-p, k-1, n-k). \quad (124)$$

Itt p a szignifikancia szint. Felhívjuk a figyelmet arra, hogy az argumentumok sorrendje fontos! H_0 -t akkor fogadjuk el, ha az aktuális érték kisebb, mint a kritikus, vagyis ha $F_{akt} < F_{krit}$.

11. Műszaki és gazdasági adatok elemzése tárgy gyakorlatain használt Excel függvények

Magyar	Angol
ÁTLAG	AVERAGE
SZUM	SUM
DARAB	COUNT
SZÓRÁS	STDEV
SZÓR.M	STDEV.S
SZÓRÁSP	STDEVP
SZÓR.S	STDEV.P
GYÖK	SQRT
KVARTILIS	QUARTILE
KVARTILIS.TARTALMAZ	QUARTILE.INC
KEREKÍTÉS	ROUND
LOG	LOG
PERCENTILIS	PERCENTILE
PERCENTILIS.TARTALMAZ	PERCENTILE.INC
DARABTELI	COUNTIF
MIN	MIN
MAX	MAX
KITEVŐ	EXP
KORREL	CORREL
HA	IF
INVERZ.STNORM	NORMSINV
INVERZ.T	TINV
INVERZ.F	FINV
INVERZ.KHI	CHIINV
STNORMELOSZL	NORMSDIST
FAKT	FACT
KOMBINÁCIÓK	COMBIN
BINOM.ELOSZL (zh-n nem használható)	BINOM.DIST
VÉL()	RAND()
INVERZ.MÁTRIX	MINVERSE
MSZORZAT	MMULT