

STATISZTIKA

Szerző:

Dr. Szalka Éva
egyetemi docens

2020

Tartalomjegyzék

I. modul: Csoportosítás, összehasonlítás	1
1. lecke: Alapfogalmak	1
2. lecke: A statisztikai adatok rendezése	6
3. lecke: Viszonyszámok	12
II. modul: Az empirikus eloszlások elemzése	16
4. lecke: Számított középértékek	16
5. lecke: Helyzeti középértékek	21
6. lecke: A szóródás és mérése	27
7. lecke: Aszimmetria, koncentráció	35
8. lecke: A sztochasztikus kapcsolatok	40
III. modul: Indexszámítás	48
9. lecke: Aggregát típusú indexszámítás	48
10. lecke: Standardizálásos indexszámítás	56
IV. modul: Mintavételi eljárások, becslés, hipotézisvizsgálat	61
11. lecke: Mintavételi eljárások	61
12. lecke: Statisztikai becslés	65
13. lecke Statisztikai próbák - Hipotézisvizsgálat	72
13.1. Alapfogalmak	72
13.2. Egymintás statisztikai próbák	76
13.2.1. Várható értékre irányuló próbák	76
13.2.2. Sokasági szórásra vonatkozó próba	78
13.2.3. Függetlenségvizsgálat	79
13.2.4. Illeszkedésvizsgálat	81
13.3. Két- és több mintás statisztikai próbák	83
13.3.1. Várható értékek különbözőségére irányuló próbák	83
13.3.2. Két sokasági szórás egyezőségére irányuló próba	86
13.3.3. Két eloszlás egyezőségének a vizsgálata	86
13.3.4. Varianciaanalízis	87
V. modul: Korreláció- és Regresszió analízis	90
14. lecke Kétváltozós korreláció- és regresszió-számítás: Lineáris	90
15. lecke Kétváltozós korreláció- és regresszió-számítás: exponenciális és hatványkitevős függvények	95
VI. modul: Idősorok elemzése	99
16. lecke: Egyszerű elemzési módszerek	99
17. lecke: Analitikus trendszámítás	102
18. lecke: Szezonális ingadozások elemzése	108
Képletek	115
Táblázatok	127

I. modul: Csoportosítás, összehasonlítás

1. lecke: Alapfogalmak

Követelmények:

Ön akkor sajátította el megfelelően a tananyagot, ha

- Felsorolásból ki tudja választani a statisztika szó jelentéseit,
- Az adott példákhoz társítani tudja a sokaság típusát.
- Egy adott tulajdonságról el tudja dönteni, hogy az mennyiségi, vagy minőségi ismerv
- Egy adott adattípusról el tudja dönteni, hogy az milyen skálán mérhető
- Adott statisztikai adatok esetén ki tudja számolni az abszolút és relatív hiba nagyságát.

1.1. A statisztika fogalma

A statisztika az információ gyűjtésének, leírásának, elemzésének, értékelésének és közlésének tudományos módszertana. A statisztika módszertani és alkalmazási alapismeretek megszerzése két okból elengedhetetlen. A gazdasági élet különböző területein szükség van a statisztikai módszerek és összefüggések feltárásának és elemzésének. Másrészt az élet egyéb területén is találkozunk számokkal, illetve adatokkal, amelyeket meg kell figyelni, fel kell dolgozni, és valamilyen szempontok szerint elemezni.

A statisztika

1. modellezni törekvő tudományos módszertan,
2. gyakorlati tevékenység, amely az adatok összegyűjtését, rendszerezését, feldolgozását és elemzését jelenti.
3. Információhalmaz, ami az összegyűjtött és rendszerezett adatok összességét jelenti.

1.2. Statisztikai sokaság

A statisztikai vizsgálat tárgyát a statisztikai sokaság képezi. A statisztikai sokaság a megfigyelés tárgyát képező egyedek összessége, halmaza.

A statisztikai sokaság típusai:

1. *Állósokaság*: állapotot fejez ki, megfigyelése mindig adott időpontban végezhető el. Ezek a sokaságok állandóan változnak, így megragadásuk csak egy időpillanatban lehetséges. Pl.: a hallgatók létszáma 2019. szeptember 1.-én.
2. *Mozgósokaság*: időben változó sokaság, egy folyamatot érzékel, időintervallum alatt figyelhető meg. Pl.: a 2019/2020. I. félévében vizsgát tett hallgatók száma.
3. *Diszkrét sokaság*: elemei jól elkülöníthetők egymástól, pl.: hallgatók
4. *Folytonos sokaság*: elemei nem különülnek el, értékeit általában intervallumban adják meg, a statisztikai sokaság egyedei csak méréssel különíthetők el, pl.: a hallgatók magassága.
5. *Véges sokaság*: véges sok elemből áll.
6. *Végtelen sokaság*: végtelen sok elemből áll, vagy olyan sok elemből, hogy végtelennek célszerű tekinteni.
7. *Egynemű sokaság*: azonos elemekből áll, nem bontható részekre, pl.: a homokbánya homokja

8. *Összetett sokaság*: az elemeknek nemcsak közös, hanem megkülönböztető tulajdonságai is vannak, pl.: hallgatók.
9. *Valós sokaság*: ténylegesen előforduló elemekből áll.
10. *Elméleti sokaság*: az adott eseményre vonatkozó bekövetkezések összes lehetséges kimenetele alkotja.

1.2.1. Statisztikai egység

A sokaságot alkotó egyedeket a sokaság egységeinek nevezzük. A statisztikai egység a vizsgált információ hordozója.

Fajtái:

- Megfigyelési: amire az adatgyűjtés vonatkozik,
- Számbavételi: akitől az adatot összegyűjtjük.

Minden statisztikai egységnek tárgyi, területi és időbeli szempontból egyaránt egyértelműen körülhatárolhatónak kell lenni.

A statisztikai sokaság olyan egyedekből áll, amelyek:

- Egyes meghatározott tulajdonságok szempontjából egyformák, - ezek alapján sorolhatjuk egy csoportba, sokaságba őket;
- Más tulajdonságok alapján különböznek – ezek biztosítják az egyedek elkülöníthetőségét.

Bemutató példa

Sokaság	Egység vagy egyed
Magyarország traktor állománya	Adott időpontban Magyarországon magánszemélyek tulajdonában lévő traktorok száma
Az Egyetem dolgozói	Az Egyetem állományába tartozó emberek

1.2.2. Statisztikai ismérv

A sokasággal szorosan összefüggő fogalom az ismérv. Mivel egy sokaság az egyedein keresztül figyelhető meg, ezért fontosak az egyedek tulajdonságai, jellemzői, melyeket *statisztikai ismérvnek* nevezzük. Az ismérvek lehetséges kimeneteleit ismérvváltozatoknak nevezzük, ezek az ismérvek által felvehető lehetséges értékeket jelentik. Pl.:

Statisztikai egyed: az Egyetem dolgozója,

Statisztikai ismérv: nem, kor, név, vagy hajszín

Statisztikai ismérvváltozó: férfi vagy nő; fiatal vagy idős;

Az olyan ismérvet, melynek két ismérvváltozata van, alternatív ismérvnek is nevezzük. Ilyen ismérv például a nem, melynek a lehetséges változata: férfi, nő.

Az ismérvek lehetnek:

- *Közös ismérv*: amelyek a statisztikai sokaságot meghatározzák, ezek alapján a sokaság egységei egyformák (pl.: évfolyam).
- *Megkülönböztető ismérv*: amely szerint a sokaság egyedei különböznek egymástól (pl.: nem, kor, lakcím, tanulmányi átlag).

Az ismérvek különböző információkat nyújtanak, így az ismérveket csoportosíthatjuk:

- *Területi ismérvek:* az egységek térbeli elhelyezésére szolgáló rendező elvek. Ismérvváltozatok általában földrajzi egységek (pl.: a hallgatók születési helye).
- *Időbeli ismérvek:* az egységek időbeli elhelyezésére szolgáló rendező elvek. Ismérvváltozatai: időpontok és időszakok (pl. a hallgatók születési ideje). Lehet állapotot (időpontra vonatkozó) és tartamot (idő intervallumra vonatkozó) kifejező.
- *Tárgyi ismérvek:*
 - *Mennyiségi ismérvek:* az egyedek számszerűen mérhető tulajdonságai. Ez lehet folytonos, amit méréssel kapunk meg, és lehet diszkrét, amihez számlálással jutunk el.
 - *Minőségi ismérvek:* az egységeket valamilyen tulajdonság alapján különböztetjük meg. Pl.: szín.

1.2.3. Az ismérvek mérhetőségi tulajdonságai

Az ismérvértékeket skálán mérjük, a különböző skálák rendszerezési kritériumaikban és rendszerezési sajátosságaiban térnek el egymástól.

- **Nominális skála:** az ismérvértékek csak azonosságuk vagy különbözőségük alapján mérhetőek. Sorrend nem állapítható meg. Pl.: foglalkozás, családi állapot
- **Ordinális skála:** Az ismérvértékeket nemcsak az azonos vagy különböző kritérium szerint rendezzük, hanem egy természetes sorrend szerint Pl.: osztályzatok (kiváló, jó, stb.), helyezési számok egy versenyen.
- **Különbség skála:** mérést jelent, ugyanis a skálaértékek különbségei is valós információt adnak a sokaság egységeiről. Pl.: hőmérséklet
- **Arány skála:** a legtöbb információt nyújtó mérést jeleníti meg. Kezdőpontja kötött, ez a „0” pont. pl.: beváltott valuta összege.

1.3. Statisztikai adat

A *statisztikai adat* valamely statisztikai sokaság elemeinek a száma, vagy a sokaság valamilyen másféle számszerű jellemzője, mérési eredménye. Két fajtája ismert:

- *alapadat:* mérés vagy számlálás útján nyerjük,
- *származtatott adat:* műveletek eredménye.

A statisztikai adatok egy része számszerű adat, amelyek lehetnek:

- *abszolút adatok:* adatgyűjtésből származnak, számítással kapjuk meg (összegezés, különbségképzés).
- *Relatívadat:* két statisztikai adat hányadosaként számítható ki.

A statisztikai adatok másik része szöveges adat, ezeket kódolni kell, hogy elemzésre tudjuk használni.

A statisztikai adatokkal szemben három követelményt támasztunk.

1. *Pontosság:* az adatok megfelelően pontosak legyenek.
2. *Gyorsaság:* gyorsan hozzájussunk az adatokhoz.
3. *Gazdaságosság:* alacsony költségek.

1.3.1. A statisztikai adatok hibája

Az összegyűjtött adatok általában korlátozott pontosságúak, azaz hibásak. A keletkezés szerint a hiba lehet:

- Adatgyűjtési hiba
- Véletlen hiba.

Jellege szerint a hiba:

- Abszolút hiba (a): a valóságos adat (A) és a mért adat (A') különbsége.
 $a = A - A'$
- Relatív hiba (α): az abszolút hiba és a valóságos adat hányadosa, amelyet százalékban szoktak megadni

$$\alpha = a/A$$

Az adatok megadásánál beszélünk szignifikáns számjegyről, ez azt jelenti, hogy nagyságrendben csak azokat a számokat írják le, amelyeket még megbízhatónak tartanak. Az adatok jelentős része kerekített szám.

A becült abszolút hiba kiszámítható az alábbi képlettel, azonban ha kiszámoljuk, akkor becült abszolút hibáról beszélünk:

$$\hat{a} \leq \frac{10^k}{2},$$

ahol $\hat{a}$: a becült abszolút hiba
k: pedig az utolsó szignifikáns számjegy helyértéke

A relatív hiba is kiszámítható, ebben az esetben becült relatív hibáról beszélünk:

$$\hat{\alpha} = \frac{\hat{a}}{A'}$$

Pl. Magyarország népessége 10.058 ezer fő. Az utolsó szignifikáns számjegy a 8, ami az ezresek helyén áll, ami tíznek harmadik hatványa, azaz $k=3$.

$$\hat{a} \leq \frac{10^3}{2} = 500 \quad \hat{\alpha} = \frac{500}{10058000} = 0,0000497 = 0,005\%$$

1.4. A statisztikai munka fázisai

A statisztikai munka négy fázisát különíthetjük el. Az egyes fázisokról később részletesen beszélünk.

Tervezés Először az Magyarországon az adatvédelmi törvény tartalmazza a célhoz kötöttség elvét. Ez azt jelenti, hogy személyes adatot gyűjteni, feldolgozni csak pontosan meghatározott jogszerű célra szabad. Az adatkezelés minden fázisában e célnak fenn kell állnia. Tilos az adatok készletezése. Éppen ezért ebben a szakaszban meg kell határozni a gyűjtendő adatok körét. Továbbá meg kell tervezni az adatgyűjtés gyakoriságát, idejét, helyét, módját.

Adatfelvétel Az adatfelvétel a statisztikai adatok beszerzését jelenti. Ez vagy más adatforrásokból való átvétel, vagy adatgyűjtést jelent. Az adatgyűjtés történhet kikérdezéssel (pl. kérdőívek alkalmazása), megfigyeléssel (pl. mérés), kísérlettel. Az adatgyűjtés (körét tekintve) teljes vagy részleges lehet. A teljes megfigyelés esetén a sokaság egészét, minden egyedét megfigyeljük, míg részleges megfigyelés esetén csak egy részét. A részleges megfigyelés egy fontos típusa a reprezentatív megfigyelés (mintavétel).

Feldolgozás Itt kell elvégezni az adatok ellenőrzését és helyesbítését; azok osztályozását, az eredmények táblákba foglalását. Ez történhet kézi vagy gépi eszközökkel. adatgyűjtés, illetve a statisztikai elemzés célját kell meghatároznunk.

Elemzés Matematikai és logikai műveletek végzését jelenti. Ekkor különböző mutatók kiszámítása, értelmezése, szöveges elemzések készítése, grafikus reprezentáció készítése történik.

2. lecke: A statisztikai adatok rendezése

Követelmények:

Ön akkor sajátította el megfelelően a tananyagot, ha

- Felsorolásból ki tudja választani az összegezzhető statisztikai sorokat,
- A megadott adatokról el kell tudni dönteni, hogy azok milyen statisztikai sorba rendezhetőek.
- Mennyiségi adatsorból ki tudja számítani a kumulált és a relatív gyakoriságot,
- A keletkezés módja és az ismerv fajtája szerint csoportosítani tudja a statisztikai sorokat,
- Felsorolásból ki tudja választani a statisztikai táblák formai követelményeit,
- adott statisztikai tábláról el tudja dönteni, hogy az milyen típusú (egyszerű, kombinációs, csoportosító),

2.1. Statisztikai sorok

A statisztikai megfigyelések során begyűjtött nagy mennyiségű adatot, a könnyebb kezelhetőség és áttekinthetőség végett valamilyen szempont szerint rendszerezniük, osztályozniuk kell. Ez a megkülönböztető ismérvek alapján megtehető. A statisztikai sor a statisztikai adatok meghatározott összefüggésben történő felsorolása, egyetlen ismerv szerinti csoportosítása eredményeként jön létre. A csoportosítás lehetőséget nyújt a sokaság szerkezetének, struktúrájának tanulmányozására.

A statisztikai sorok felosztása:

Az adatok fajtája szerint:

- *valódi sorok:* egy sokaságra vonatkoznak, azonos fajtájú adatokból keletkeznek
- *nem valódi sorok:* valamilyen jelenség leírását adják többféle szempontból, több sokaságra vonatkoznak:
 - Leíró sor: A megfigyelés tárgyának különböző tulajdonságait írja le, az adatok között logikai kapcsolat van. Az intenzitási viszonyszámok forrásai.

A keletkezés módja szerint a valódi sorokat tovább bonthatjuk:

- Összehasonlító sorokra,
- Csoportosító sorokra.

Az ismerv fajtája szerint a valódi sorok az alábbiak lehetnek:

- Idősor,
- Területi sor,
- Minőségi sor,
- Mennyiségi sor.

2.1.1. Idősorok

Bizonyos időpontban vagy időpontban meglévő vagy bizonyos időtartam alatt bekövetkező jelenségek adatait sorolja fel.

- Állapot idősor: az álló sokaságok időbeli változását mutatják be. Az adatok összegzésének nincs tárgyi értelme. Pl.: a személygépkocsi állomány éves bontásban.

- Tartam idősor: egy időintervallum alatti mozgó sokaság időbeli alakulását mutatja. Dinamikus helyzetet tükröz. Az adatok összegzésének van tárgyi értelme, az összegzéssel egy hosszabb időtartamra jellemző adatot képezhetünk. Pl. turisták száma évente.

2.1.2. Területi sorok

Ha a csoportosítás a rendező elv, akkor egy nagyobb területi egység adatait az összetartozó kisebb területi egységeknek megfelelően osztályozzuk. Azonos időpontban rögzített statisztikai sokaság térbelileg bontott részsokaságait tartalmazza. Az ismérvváltozatok sorrendje kötetlen. A gyakoriságok összeadásának van tárgyi értelme. Pl.: népesség megyék szerinti megoszlása.

Ha azonban az összehasonlítás a rendező elv, akkor az adatok összegzésének nincs tárgyi értelme.

2.1.3. Minőségi sorok

A sokaságelemek minőségi tulajdonságainak formailag előre rögzített csoportosítása révén jön létre. Betekintést nyújt a sokaság összetételébe, szerkezetébe. Az ismérvváltozatok sorrendje kötetlen, a gyakoriságok (az előfordulások száma) összeadásának van tárgyi értelme. Pl. A külföldiek megoszlása az utazás jellege szerint 2019-ben: turista, kiránduló, átutazó.

2.1.4. Mennyiségi sorok

Mennyiségi ismérvek szerinti csoportosítással jön létre. Ha az ismérvértékek konkrét pontossággal megadott számok, akkor *diszkrét mennyiségi sorról* beszélünk. Általában számlálással kapjuk az ilyen sorokat. Ha az ismérvértékek adott intervallumban bármilyen értéket felvehetnek, akkor *folytonos mennyiségi sorról* beszélünk.

A mennyiségi sorok fajtái:

- 1) **Gyakorisági sor:** az ismerv előfordulásának gyakoriságát tüntetjük fel. A gyakoriság (f_i) megmutatja, hogy az egyes ismérvváltozatok hányszor fordulnak elő a megfigyelt sokaságban. Ha az egyes gyakoriságokat azok összegéhez viszonyítjuk, akkor az adott ismérvérték relatív gyakoriságát (g_i) kapjuk meg:

$$g_i = \frac{f_i}{\sum f_i} \quad \text{ahol: } g_i: \text{ az } i\text{-edik ismérvérték relatív gyakorisága}$$

f_i : az i -edik ismérvérték gyakorisága

$\sum f_i$: a sokaság elemeinek száma, azaz egyenlő n -nel.

Ha az ismérvváltozatok száma nagy, akkor az adatokat rangsoroljuk, és ez megkönnyíti a változó osztályozását. Az osztályozás sűríti az információt. A legnagyobb és legkisebb ismérvek által adott intervallumot úgy osztjuk osztályokba, hogy az egyes osztályközökön belül a gyakoriságok közel egyenlő eloszlásúak legyenek, így az osztályközép alkalmas lesz az osztály jellemzésére.

Az osztályok olyan adatcsoportok, ahol az egyes osztályok közötti mennyiségi változás minőségi változást takar.

Az osztályközök száma: $2^k > N$

Osztályköz hossza:

$$h = \frac{x_{\max} - x_{\min}}{k}$$

Ha a csoportosító ismérv folytonos jellegű, a mennyiségi sort osztályközös gyakorisági sornak nevezzük.

Bemutató feladat:

Egy vajat gyártó üzemben az átlagosan csomagolt vaj mennyiségének meghatározásához 24 elemű mintát vettek. A mérések eredménye (g):

104,2	100,2	93,4	98,6	117,4	100,8	88,2	100,1
101,9	116,8	112,9	97,8	99,3	111,5	109,3	100,7
118,0	96,4	89,4	98,9	105,3	112,4	99,5	102,7

Feladat:

- Készítsen az adatokból rangsort!
- Írja fel az optimálisnak tartott gyakorisági sort!

Megoldás:

88,2	89,4	93,4	96,4	97,8	98,6	98,9	99,3
99,5	100,1	100,2	100,7	100,8	101,9	102,7	104,2
105,3	109,3	111,5	112,4	112,9	116,8	117,4	118,0

$n=24 \quad 2^5=32 > 24$, ezért $k=5$

Az osztályköz hossza: $h = (118,0 - 88,2)/5 = 5,96$. Az osztályköz hossza kerekítve 6 lesz

Csomagolt vaj súlya (g)	A lemért vajkockák száma (db)	Osztályközép
88,1-94,0	3	91
94,1-100,0	6	97
100,1-106,0	8	103
106,1-112,0	2	109
112,1-118,0	5	115
Összesen:	24	-

- 2) **Értékösszeg-sor:** Ha a gyakorisági soroknál az ismérv előfordulási értékeinek összegét tüntetjük fel, akkor értékösszeg-sort kapunk.

Az értékösszeg jele : s_i .

$$s_i = f_i \cdot x_i$$

Bemutató feladat:

Egy kereskedelmi egységben feljegyezték egy adott napon a tejet vásárlók megoszlását mennyiség szerint:

Vásárolt tej mennyisége (l)	Vásárlók száma (db)
0-5,0	32
5,1-10,0	48
10,1-15,0	92
15,1-20,0	63
20,1-25,0	15
Összesen	250

Feladat:

- Állapítsa meg az egyes intervallumok osztályközepét!
- Készítsen az adatokból értékösszeg-sort!

Megoldás:

<i>Vásárolt tej mennyisége (l)</i>	<i>Vásárlók száma (db)</i>	<i>Osztályközép (l)</i>	<i>Eladott tej mennyisége (l) s_i</i>
0-5,0	32	2,5	80
5,1-10,0	48	7,5	360
10,1-15,0	92	12,5	1150
15,1-20,0	63	17,5	1102,5
20,1-25,0	15	22,5	337,5
Összesen	250	-	3030

- 3) **Kumulált gyakorisági sor:** A gyakorisági soroknál sajátos információkat nyerhetünk a gyakoriságuk kumulált képzésével. A kumulálás halmozott összeadást jelent, s arról kapunk képet, hogy egy adott értékhatárnál kisebb (alulról kumulált) vagy nagyobb (felülről kumulált) érték összesen hányszor fordul elő. A kumulált gyakoriság jele: f_i' .

Bemutató feladat:

Az alábbi táblázat a gyermekvédelmi gondozásban részesülők számát mutatja be. Feladat a kumulált gyakoriság és a relatív gyakoriság kiszámítása.

Korcsoport (év)	gyermekek száma (fő) f_i	f_i'	g_i
0-3	1848	1848	0,11
4-5	1221	1848+1221=3069	0,07
6-9	3075	3069+3075=6144	0,18
10-11	2164	6144+2164=8308	0,13
12-14	3827	8308+3827=12135	0,22
15-17	5010	12135+5010=17145	0,29
Összesen	17145	-	1,00

Pl.: $g_3=(3075/17145)=0,18$

2.2. Statisztikai táblák

Több tulajdonság szerinti rendezéskor kapjuk. A statisztikai sorok összefüggő rendszere. Rendeltetés szerint lehet:

- alap vagy gyűjtő tábla,
- feldolgozási vagy munkatábla,
- közlési vagy eredménytábla.

Az összefoglalt sorok típusa szerinti felosztás alapján:

- egyszerű tábla: csoportosítást nem tartalmazó adatsorok összefüggő rendszere. Általában leíró és összehasonlító sorokat tartalmaz.

Megnevezés	2017	2018	2019
Vállalkozások száma			
Saját tőke			
Idegen tőke			

- Csoportosító tábla: egy ismerv szerinti csoportosítást tartalmazó statisztikai sorok összefüggő rendszere.

Életkor	2017	2018	2019
0-6			
7-14			
15-22			
.			
.			
Összesen			

- Kombinációs tábla: A sokaság több ismerv szerinti kombinatív osztályozásának eredményeként kapott adatokat tartalmazza.

Megnevezés	Budapest	Város	Község	Összesen
Komfortos				
Félkomfortos				
Komfort nélküli				
Összesen				

Statisztikai táblák készítése

Formai és tartalmi követelményeknek kell érvényesülni, hogy megfeleljen a vizsgálat céljára.

Formai követelmény:

- táblázat címe,
- megnevezések a fej- és oldalrovatban,
- mértékegység feltüntetése,
- forrás megjelölése.

Például:

A gazdálkodó szervezetek gyümölcsös területének méret szerinti megoszlása (2017)

Terület (ha)	Szervezetek száma (db)
-10	756
10,1-50,0	270
50,1-100,0	63
100,1-200,0	27
200,1-300,0	10
300,1-	2
Összesen	1087

Forrás: Agrárgazdasági Statisztikai Zsebkönyv, 2017.

Tartalmi követelmény.

A tábla minden celláját információközlésre kell felhasználni, ezért a következő szokásos jelöléseket alkalmazzák:

- konkrét adat: számszerű információ,
- kihúzott rovat (-): nincs információ
- ...: létezik adat, de nem áll rendelkezésre
- +: az adat előzetes becslés eredménye
- *: a táblázat alján lévő megjegyzésre utal
- 0,0: van adat, de a többi adathoz képest túl kicsi.

3. lecke: Viszonyszámok

Követelmények:

Ön akkor sajátította el megfelelően a tananyagot, ha

- Adott számról el tudja dönteni, hogy az viszonzyszám-e.
- Megadott adatról el tudja dönteni, hogy az milyen viszonzyszámmal jellemezhető,
- Ismert adatokból ki tudja számítani a viszonzyszámokat.
- A számított adatok alapján ki tudja egészíteni a számításra vonatkozó állítást,
- Ki tud egészíteni viszonzyszámokat tartalmazó táblázatot.

Két egymással összefüggő statisztikai adat hányadosát **viszonyszámnak** nevezzük. A viszonzyszámok képzésénél alapszabály, hogy csak olyan adatokból képezzük, amelyek tartalmilag összehasonlíthatóak. Az így nyert leszármaztatott szám újszerű információkat tartalmaz.

$$V = \frac{A}{B}$$

V: viszonzyszám értéke

A: a viszonyítás tárgya (amit)

B: a viszonyítás alapja (amihez)

A viszonzyszámokat számíthatjuk azonos fajta (azonos mértékegység) és különbözőfajta adatokból. Az azonos mértékegységű adatokból számított viszonzyszámok azt fejezik ki, hogy az egyik adat hányszorosa a másinak.

A viszonzyszámoknak több típusa létezik, attól függően, hogy milyen statisztika sorból számítjuk ki őket.

3.1. Megoszlási viszonzyszám

A sokaság egyes részeinek a sokaság egészéhez viszonyított arányát fejezi ki. A megoszlási viszonzyszám tehát a részsokaságnak az egészhez való viszonyát (relatív gyakoriságát) fejezi ki. A csoportosító sorok (minőségi és mennyiségi) elemzési eszköze, ahol a sor egy elemét hasonlítjuk a sor egészéhez. Értéke 0 és 1 között van, gyakran százalékban fejezzük ki, akkor 0 és 100%. A megoszlási viszonzyszám segítségével tömören jellemezhető egy sokaság ismérvváltozatok szerinti szerkezete. Állapot idősor adataiból nem számolható.

$$V_m = \frac{x_i}{\sum x_i}$$

Bemutató feladat:***A beruházások teljesítményértéke 2017-ben***

Ágazat	Összes beruházás (milliárd HUF)	Kiszámítás módja	Megoszlási viszonyszám (%)
Nem évelő növények termesztése	65	$100 \cdot 65 / 173$	37,57
Évelő növények termesztése	3	$100 \cdot 3 / 173$	1,73
Növényi szaporítóanyag termesztése	2	$100 \cdot 2 / 173$	1,16
Állattenyésztés	91	$100 \cdot 91 / 173$	52,60
Vegyes gazdálkodás	1	$100 \cdot 1 / 173$	0,58
Mezőgazdasági betakarítást követő szolgáltatás	11	$100 \cdot 11 / 173$	6,36
Összesen	173		100,00

Forrás: Agrárgazdasági Statisztikai Zsebkönyv, 2017.

A beruházások 52,60%-a az állattenyésztésben valósult meg.

3.2. Koordinációs viszonzszám

A koordinációs viszonzszám ugyanazon statisztikai sokaság két részsokaságának egymáshoz viszonyított arányát fejezi ki.

3.3. Összehasonlító viszonzszámok

Az összehasonlító viszonzszámok arról tájékoztatnak, hogy hogyan aránylik egymáshoz két azonos típusú – egymástól térben vagy időben különböző – ismervérték. Akkor alkalmazzák, ha különböző színvonalú statisztikai sorokat kívánnak összehasonlítani. Az összehasonlító viszonzszámokat főleg idősorok és területi sorok összehasonlításánál alkalmazzák.

$$V_{\bar{o}} = \frac{x_i}{x_k}$$

A sor minden elemét elosztjuk a bázisként kiválasztott elemmel. Az összehasonlító viszonzszám a fejlődés irányát mutatja meg.

$V_{\bar{o}} > 1$ (100%) növekedés $V_{\bar{o}} < 1$ (100%) csökkenés $V_{\bar{o}} = 1$ (100%) stagnálás

3.3.1. Dinamikus viszonzszámok

Az összehasonlító viszonzszámok leggyakoribb formája a **dinamikus viszonzszám**, amely általában 1-1 ismerv idősorainak értékeiből képzett hányados. A dinamikus viszonzszám jól érzékelteti a változás azonos időintervallumokra vonatkozó ütemét. Fajtái:

- **bázisviszonzszám**: az idősor értékeit egy rögzített időpont (időszak) értékéhez viszonyítjuk

$$V_b = \frac{x_i}{x_0}$$

- **lácviszonyszám**: az egyes idősor értékeket mindig közvetlenül a megelőző időpont (időszak) értékéhez viszonyítjuk:

$$V_l = \frac{x_i}{x_{i-1}}$$

Bemutató feladat:

Növénytermesztési és kertészeti termékek kibocsátás 2017-ben:

Év	Kibocsátás (milliárd HUF)	V_b	V_b (%)	V_l	V_l (%)
		kiszámítása		kiszámítása	
2013	1364	$100 \cdot 1364 / 1364$	100,00		-
2014	1455	$100 \cdot 1455 / 1364$	106,67	$100 \cdot 1455 / 1364$	106,67
2015	1451	$100 \cdot 1451 / 1364$	106,38	$100 \cdot 1451 / 1455$	99,73
2016	1562	$100 \cdot 1562 / 1364$	114,52	$100 \cdot 1562 / 1451$	107,65
2017	1492	$100 \cdot 1492 / 1364$	109,38	$100 \cdot 1492 / 1562$	95,52

Forrás: Agrárgazdasági Statisztikai Zsebkönyv, 2017.

Látható, hogy pl. 2017-ben a termelés 9,38%-kal haladta meg a 2013 évit, de 4,48%-kal csökkent 2016-hoz képest.

Az idősorokból számított bázis- és lácviszonyszámok összefüggnek egymással.

$\frac{x_1}{x_0} \cdot \frac{x_2}{x_1} \cdot \dots \cdot \frac{x_n}{x_{n-1}} = \frac{x_n}{x_0}$, azaz a lácviszonyszámok szorzata bázisviszonyszámot ad.

$\frac{x_n}{x_0} / \frac{x_{n-1}}{x_0} = \frac{x_n}{x_{n-1}}$, azaz a bázisviszonyszámok hányadosa lácviszonyszámot ad.

A fejlődés átlagos ütemét a lácviszonyszámokból lehet kiszámolni:

$$\bar{V}_l = \sqrt[n-1]{V_{l1} \cdot V_{l2} \cdot \dots \cdot V_{ln}} = \sqrt[n-1]{\prod V_{ln}} = \sqrt[n-1]{\frac{x_n}{x_0}}$$

példánkban

$$\bar{V}_l = \sqrt[4]{\frac{1492}{1364}} = 102,27\%$$

A kibocsátás 2013 és 2017. között átlagosan évi 2,27%-kal nőtt.

Ez az átlag csak akkor ad valós képet, ha a változás a teljes időszak alatt hasonlóan alakult (azaz egyenletesen nőtt vagy csökkent).

Bemutató feladat:

Egy pénzintézetnél az elmúlt években a hitel kamatából, számított lácviszonyszámok a következők:

-, 0,9998; 0,9996; 0,9992; ,09997; 0,9999; 1,1823; 0,9991.

Látható, hogy a kamatváltozást mutató lácviszonyszámok egy kivételével csökkenést mutatnak az előző évhez képest. Számítsuk ki a vizsgált időszakra az átlagos kamatváltozást:

$$\bar{V}_1 = \sqrt[7]{0,9998 * 0,9996 * 0,9992 * 0,9997 * 0,9999 * 1,1823 * 0,9991} = 1,0238 = 102,38\%$$

Látható, hogy ebben az esetben nem számolható ki az átlag, mert az azt mutatja, hogy évente átlagosan 2,38%-kal emelkedett a kamat. A valóságban azonban csak egyetlen évben volt kamatemelkedés.

3.3. Teljesítmény-viszonyszámok (terv-viszonyszámok)

Valamely számokkal meghatározott kitűzött cél végrehajtását mutatja meg. Főleg munkateljesítmény kiszámolására használják.

$$V_i = \frac{\text{elért.eredmény}}{\text{tervezett.eredmény}}$$

pl. egy munkás napi normája 150 db 8 óra alatt. Egyik munkavállaló 159 db-ot készített, tehát a teljesítménye $159/150=106\%$.

3.4. Intenzitási viszonszámok

Különböző típusú, de egymással összefüggő teljes- és részsokaság egybevetésével keletkeznek. Az intenzitási viszonszámok két különböző statisztikai sor megfelelő elemének hányadosa. A leírások elemzésének jellegzetes eszköze, dimenzióval ellátott mutatószám.

Pl. sebesség = út/idő (km/h).

Fajtái:

- reverzibilis: megfordítható t/ha $\longrightarrow$ ha/t
- irreverzibilis: nem fordítható meg pl. termelékenység, 12 db/fő. Ez azt jelenti, hogy egy ember 12 munkadarabot gyárt. Ennek fordítottja lenne a 0,0833 fő/db, ami azt jelenti, hogy egy darab terméket 0,0833 fő gyárt, és ennek nincs értelme.
- nyers: a viszonyítási alap és a viszonyítás tárgya között lazább a kapcsolat: $\text{születések száma/népesség}$. A kapcsolat azért laza, mert a népességnek csak egy része képes szülésre
- tiszta: viszonyítási alap és a viszonyítás tárgya között szoros a kapcsolat: $\text{születések száma/szülőképes korú nők}$. A kapcsolat azért szoros, mert a viszonyítás alapja a szülésre képes nők száma, ellentétben az előző esettel.

II. modul: Az empirikus eloszlások elemzése

4. lecke: Számított középértékek

Követelmény:

Ön akkor sajátította el megfelelően a tananyagot, ha

- Szöveges megfogalmazás alapján el tudja dönteni, hogy az milyen átlagot jelent.
- A felsorolásból ki tudja választani, hogy milyen átlagot lehet kiszámolni
- Felsorolásból ki tudja választani, hogy egy adott számításához milyen adatokra van szükség
- Egy sokaság adatait megismerve el tudja dönteni, hogy azokból milyen típusú átlagot lehet kiszámolni.
- Egy eloszlás számtani átlagát meg tudja határozni
- Mértani-, négyzetes- és harmonikus átlagot számolni.

A viszonyszámok mellett leggyakrabban használt elemzési eszközök az átlagok, melyeket középértéknek is szoktak nevezni. Az átlagokat azonos fajta adatok halmazának tömör, számszerű jellemzésére használjuk. Az átlagokkal szemben támasztott követelmények az alábbiak:

- a legnagyobb és a legkisebb érték közé essen,
- egy határozott szám legyen,
- az adatok többségéhez közel álljon,
- érzékenyen reagáljon a sokaság tagjainak változására,
- könnyen kiszámítható legyen.

Természetesen egyszerre minden követelménynek nem lehet eleget tenni, de szakmai megfontolásból el kell tudni dönteni, hogy melyik követelményt tartjuk be

4.1. Számtani átlag

A leggyakrabban alkalmazott középérték, jele: $\bar{x}$. A számtani átlag az a szám, amellyel az egyes átlagolandó értékeket helyettesítve azok összege változatlan marad.

Fajtai az átlagolandó értékek előfordulása alapján:

- **Egyszerű számtani átlag:** minden átlagolandó érték egyszer fordul elő az adatsorban

$$\bar{x} = \frac{\sum x_i}{n}$$

pl.: Egy hallgató az egyik héten a következő sörmennyiséget (l) fogyasztotta el naponta:

0,7; 1,6; 2,5; 3,2; 1,6; 2,4; 2,8;

$$\bar{x} = \frac{0,7 + 1,6 + 2,5 + 3,2 + 1,6 + 2,4 + 2,8}{7} = 2,11 \text{ l volt az átlagos napi sörfogyasztása.}$$

- **Súlyozott számtani átlag:** az átlagolandó értékek többször fordulnak elő az adatsorban:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

Mivel $g_i = \frac{f_i}{\sum f_i}$, ezért a relatív gyakoriságokkal is kiszámolható az átlag:

$$\bar{x} = \sum g_i x_i$$

pl. A sikeres vizsgát tett hallgatók félévi osztályzata statisztikából az alábbi volt:

Osztályzat	f_i
2	14
3	38
4	22
5	9
Összesen	83

$$\bar{x} = \frac{14 \cdot 2 + 38 \cdot 3 + 22 \cdot 4 + 9 \cdot 5}{83} = 3,31 \text{ az évfolyam átlaga}$$

Osztályközös gyakorisági sor esetén az osztályközépek töltik be az átlagolandó értékek szerepét. Az osztályközép számítása:

$$x_i = \frac{x_{i(\text{alsóh.})} + x_{i(\text{felsőh.})}}{2} \text{ ahol: } x_{i(\text{alsóh.})}: \text{ az } i\text{-edik osztályköz alsó határa,}$$

$$x_{i(\text{felsőh.})}: \text{ az } i\text{-edik osztályköz felső határa.}$$

A számításnál nem vesszük figyelembe az elkülönítést szolgáló utolsó számjegyet.

Pl.: 21-50 esetében:

$$x_i = \frac{20 + 50}{2} = 35$$

Ha az osztályközök alsó vagy felső határa nincs megadva, akkor nyitott osztályközről beszélünk. Az alsó határt úgy adjuk meg, hogy az osztályköz szélessége azonos legyen az őt követő osztályköz szélességével. A felső határ megadásakor, pedig az előző osztályköz szélességét vesszük figyelembe.

Bemutató feladat:

Pl.: A gazdálkodó szervezetek tulajdonában lévő traktorok teljesítmény szerinti megoszlása:

	f_i	<i>Oszt. Közép</i>
Traktor (kW)		
0-8	142	4
8,0-20	586	14
20,1-40	1649	30
40,1-60	6427	50
60,1-100	7176	80
100,1-	5946	120

Az utolsó osztályköz felső határa nincs megadva, mivel az előtte lévő osztályköz hossza 40, ezért az utolsó osztályköz 100,1-140 lesz. A számtani átlagot a következőképpen számoljuk ki:

$$\bar{x} = \frac{142 * 4 + 576 * 14 + 1649 * 30 + \dots + 5946 * 120}{21926} = 76$$

A traktorok átlagos teljesítménye 76 kW.

A számtani átlag típusai összetett sokaság esetén:

- **Részátlag:** az egyes részsokaságok átlaga. (A részsokaságokat itt az egyes szakok hallgatói adják.). Az átlagokat szakonként kell kiszámítani:

$$\bar{x}_{\text{közgazdász}} = \frac{20 * 45 + 40 * 55 + \dots + 10 * 95}{170} = 65$$

$$\bar{x}_{\text{jogász}} = \frac{30 * 45 + 40 * 55 + \dots + 5 * 95}{155} = 62,10$$

$$\bar{x}_{\text{mérnök}} = \frac{20 * 45 + 50 * 55 + \dots + 10 * 95}{220} = 65,90$$

Pl.: Néhány egyetemi szak hallgatóinak teljesítménye statisztika írásbeli vizsgán

Teljesítmény (pont)	Közgazdász	Jogász	Mérnök
-50	20	30	20
51-60	40	40	50
61-70	60	50	80
71-80	30	20	40
81-90	10	10	20
91-	10	5	10
Összesen	170	155	220
Részátlagok	65,0	62,10	65,9

- **Főátlag:** a teljes sokaság átlaga. Kiszámítható a részátlagokból is, a részátlagok elemszámának súlyozásával.

•

Pl.: a fenti példában

$$\bar{x} = \frac{170 * 65,0 + 155 * 62,1 + 220 * 65,9}{170 + 155 + 220} = 64,54 \text{ pont a hallgatók átlagteljesítménye.}$$

4.2 Négyzetes átlag

Az a szám, amellyel az átlagolandó értékeket helyettesítve azok négyzetösszege nem változik. Jelentősége, hogy az egyes értékek előjelei közötti különbséget eltüntet. Főleg a szóródás számításánál használjuk. Hasonlóan a számtani átlaghoz ebben az esetben is van egyszerű- és súlyozott forma.

$$\bar{x}_q = \sqrt{\frac{\sum x_i^2}{n}} \quad \bar{x}_q = \sqrt{\frac{\sum f_i x_i^2}{\sum f_i}}$$

Bemutató feladat:

A kenyér szabvány szerinti tömege 1000g. Lemértünk 7 db kenyeret, hogy kiszámoljuk a szabványtól való eltérés átlagát.

Sorszám	A kenyér tömege (g)	A szabványtól való eltérés (g)
1	988	-12
2	1023	+23
3	1010	+10
4	972	-28
5	957	-43
6	1038	+38
7	1012	+12

Az átlagos eltérés számtani átlaggal számolva:

$$\bar{x} = \frac{-12 + 23 + 10 - 28 - 43 + 38 + 12}{7} = 0, \text{ azaz a minta alapján a kenyér tömege szabványos.}$$

Az átlagos eltérés négyzetes átlaggal számolva:

$$\bar{x}_q = \sqrt{\frac{\sum x_i^2}{n}} = \sqrt{\frac{(-12)^2 + (23)^2 + (10)^2 + (-28)^2 + (-43)^2 + (38)^2 + (12)^2}{7}} = 26,71$$

A kenyerek átlagosan 26,71g-mal térnek el a szabványtól.

4.3. Mértani átlag

Az a szám, amelyet az átlagolandó adatok helyébe behelyettesítve, azok szorzata nem változik. Általában egy folyamat relatív változásának vizsgálatakor használjuk, átlagos fejlődési ütemet mutat.

$$\bar{x}_g = \sqrt[n]{\prod x_i} \quad \bar{x}_g = \sqrt[n]{\sum f_i \prod x_i^{f_i}}$$

A „ Π ”: több tényező folyamatos összeszorzásának matematikai jele.

Pl.: Egy vállalkozó 1000 eurót helyez el a bankba kamatos kamatra. Az első 2 évben 4%, majd 3 évig 5,5%, az utolsó 6. évben pedig 6% a kamat. Mennyi az átlagos kamatnövekedés?

$$k_a = \sqrt[6]{1,04 * 1,04 * 1,055 * 1,055 * 1,055 * 1,06} - 1 = 0,0508$$

Azaz a hat év alatt átlagosan 5,08% kamatra számíthat

4.4. Harmonikus átlag

Az a szám, amellyel az egyes átlagolandó értékeket helyettesítve azok reciprokának összege nem változik. Általában akkor használjuk, ha a viszonyszámok fordított arányt tükröznek.

$$\bar{x}_h = \frac{n}{\sum \frac{1}{x_i}} \quad \bar{x}_h = \frac{\sum f_i}{\sum \frac{f_i}{x_i}}$$

pl.: Egy személygépkocsi a teljes útszakasz 1/6-át 100 km/h, 1/3-át 80 km/h és a felét 50 km/h sebességgel teszi meg. Mekkora átlagsebességgel kellene mennie, hogy ezt az utat ugyanannyi idő alatt tegye meg?

$$t_1 = \frac{s_1}{v_1} \quad t = t_1 + t_2 + t_3$$

$$v_{\dot{a}} = \frac{s}{t} = \frac{1}{\frac{1}{6v_1} + \frac{1}{3v_2} + \frac{1}{2v_3}} = \frac{1}{\frac{1}{6 \cdot 100} + \frac{1}{3 \cdot 80} + \frac{1}{2 \cdot 50}} = \frac{1}{\frac{2}{1200} + \frac{5}{1200} + \frac{12}{1200}} \cdot \frac{1200}{9} = 63,158 \text{ km/h}$$

4.5 Összefüggés a különböző átlagok között

Ha ugyanazon adatsorból mindegyik átlagot kiszámítjuk, akkor az alábbi összefüggést figyelhetjük meg:

$$\bar{x}_h \leq \bar{x}_g \leq \bar{x} \leq \bar{x}_q$$

5. lecke: Helyzeti középértékek

Követelmények

Ön akkor sajátította el megfelelően a tananyagot, ha

- felsorolásból ki tudja választani a mediánt, móduszt meghatározó helyes definíciót.
- számított értékről el tudja dönteni, hogy az módusz, medián, vagy egyik sem,
- Önállóan ki tudja számítani a móduszt diszkrét és folytonos ismérvváltozó esetén
- Önállóan kiszámítani a mediánt diszkrét és folytonos ismérvváltozó esetén
- A megadott adatok alapján kiszámolni a keresett kvartilisértékeket.

A helyzeti középértékek a sokaságban elfoglalt helyzetüknél fogva jellemzik a vizsgált jelenséget vagy folyamatot. Ahhoz, hogy az egyedek sokaságon belüli elhelyezkedése jellemezhető legyen, az egyedeket valamilyen előre rögzített szabály szerint sorba kell rendezni. Általában növekvő sorrendbe rendezzük az ismérvértékeket, azaz rangsort képezünk.

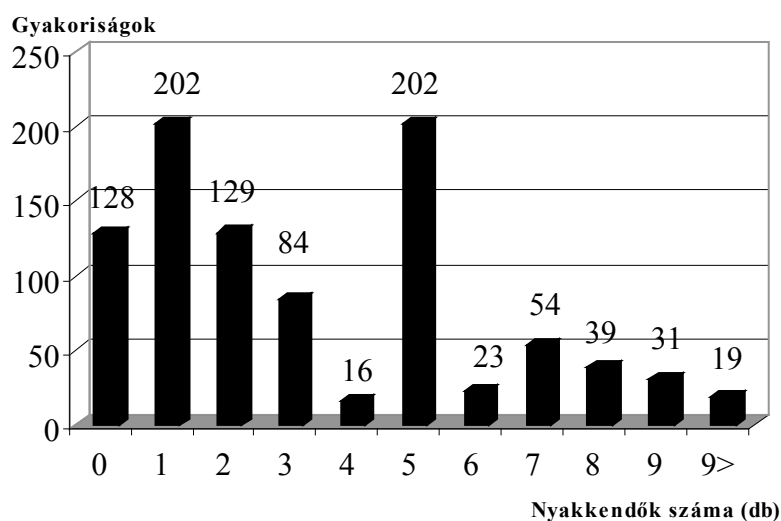
5.1. Módusz

A legegyszerűbb helyzeti mutató a leggyakoribb érték, amely egyben a „legsűrűbb” érték, és **módusznak** nevezzük. A módusz egy gyakorisági eloszlásnak az az ismérvváltozata, amely a leggyakrabban fordul elő, azaz a legnagyobb gyakorisággal. Egy eloszlásnak több módusza is lehet.

Pl. 927 férfit megkérdeztek a nyakkendőjük számáról:

Ny. száma	0	1	2	3	4	5	6	7	8	9	9>	Össz.
Gyak.	128	202	129	84	16	202	23	54	39	31	19	927

Diagramon ábrázolva:



$Mo=1$ és $Mo=5$

Azaz ennek a gyakorisági sornak két módusza van.

Folytonos gyakorisági sor esetén már nehezebb megállapítani a módusz konkrét értékét. Az osztályközös gyakorisággal megadott adatok esetében a statisztikai gyakorlat nem egységes. Kevés ismértértéket felvehető diszkrét jelenségnél adott osztályköz felső határát meghaladja a következő osztályköz alsó határa, és ekkor az osztályközepeknél az osztályközök tényleges és megadott határai megegyeznek egymással, és az osztályba sorolás egyértelmű. Ekkor az osztályközök megadott alsó és felső határaival kell számolni. Folytonos ismértértékeknel vagy sok ismértértékkal rendelkező diszkrét ismértértékeknel az osztályköz-határok kétféle megadása lehetséges. Első esetben az adott osztályköz felső határa és a következő osztályköz alsó határa megegyezik egymással, és az osztályköz-határral megegyező értékű elemeket az alacsonyabb osztályközbe soroljuk be. Második esetben közölt osztályköz-határokat adunk meg, amelyek csak arról tájékoztatnak, hogy az osztályköz-határokkal megegyező értékű elemeket az alacsonyabb osztályközbe soroljuk. Ekkor az osztályközepek kiszámításához és a módusz és medián becsléséhez az előző osztályköz felső határával számolunk.

$$M_o = m_o + \frac{f_{m_o} - f_{m_{o-1}}}{(f_{m_o} - f_{m_{o-1}}) + (f_{m_o} - f_{m_{o+1}})} * h$$

m_o : a modális osztályköz alsó határa,

f_{m_o} : a modális osztályköz gyakorisága

$f_{m_{o-1}}$: a modális osztályközt megelőző osztályköz gyakorisága,

$f_{m_{o+1}}$: a modális osztályközt követő osztályköz gyakorisága

h : az osztályköz hossza

pl.: Egy kiskereskedőnél az egyik hétvégén megvizsgálták az egyes vásárlások összegét.

Vásárláskor fizetett összeg (Ft)	Vásárlók száma (fő)
1-250	29
251-500	34
501-750	54
751-1000	39
1001-1250	20
1251-1500	18
1501-1750	12
Összesen	206

A táblázatban jól látható, hogy a legtöbb adat 501-750 között van, azaz a leggyakoribb érték itt található.

$$M_o = 500 + \frac{54 - 34}{(54 - 34) + (54 - 39)} * 250 \approx 643$$

A vásárlások 643 Ft körül sűrűsödtek.

5.2. Medián

A másik leggyakrabban alkalmazott helyzeti középérték a medián. A nagyság szerint sorba rendezett értékek közül a középső. A medián az az érték, amelynél kisebb értékek gyakorisága azonos a nálánál nagyobb értékek gyakoriságával, azaz a medián a megfigyelt értékek rangsorát két egyenlő részre osztja.

Páratlan számú adat esetén:

$$Me = \frac{n+1}{2} \text{ -edik elem a medián.}$$

Pl.: Egy vizsgán 17 hallgató az alábbi pontszámokat érte el:
8,8,9,10,10,10,11,11,12,12,13,13,13,14,14,14,14,

A medián a 9. elem, így $Me=12$

Páros számú adat esetén: a medián sorszámát (s_{Me}) az alábbiak alapján számítjuk ki

$s_{Me} = \frac{n+1}{2} - 0,5$ és $\frac{n+1}{2} + 0,5$ -edik elem lesz a medián, a hozzátartozó értékeket átlagolni kell.

pl.: 8 rúdugró az alábbi magasságokat ugrotta át:

4,6 4,8 4,8 4,9 5,0 5,0 5,1 5,3 (m)

A medián a 4. és 5. elem átlaga lesz:

$$Me = \frac{1}{2} * (4,9 + 5,0) = 4,95 \text{ m}$$

Folytonos gyakorisági sor esetén hasonlóan a móduszhoz, nehezebb megállapítani a középső értéket.

$$Me = m_e + \frac{\frac{n}{2} - f'_{m_e-1}}{f_{m_e}} * h$$

$n/2$: a medián sorszáma (középső érték)

m_e : a mediánt tartalmazó osztályköz alsó határa

f'_{m_e-1} : a mediánt tartalmazó osztályköz előző osztályköz kumulált gyakorisága

f_{m_e} : a mediánt tartalmazó osztályköz gyakorisága

h : az osztályköz hossza

pl. 200 diák testmagasságát mérték meg:

Testmagasság (cm)	f_i	f'_i
151-160	30	30
161-170	80	110
171-180	70	180
181-190	10	190
190-200	10	200
Összesen	200	-

A kumulált gyakoriságból lehet megállapítani, hogy melyik osztályközben található a medián. 200 elemünk van, így a medián elvileg a 100. és a 101. elem átlaga lenne, ha nem osztályközös gyakorisági sorból állapítanánk meg, így azonban csak azt kell meghatározni, hogy melyik osztályközben találhatóak ezek az elemek. A táblázatból leolvasható, hogy a mediánt a második osztályközben kell keresni, ugyanis itt található a nagyság szerint sorba rendezett adatok közül a 100. és 101. elem.

$$Me = 160 + \frac{\frac{200}{2} - 30}{80} * 10 = 168,75 \text{ cm}$$

a diákok fele 168,75 cm-nél alacsonyabb, fele pedig magasabb.

5.3. Kvantilisek

A középértékek mellett fontos helyzeti mutatók a kvantilisek. Ha a rangsorba rendezett sokaságot egy X -ismérvérték $q:(1-q)$ arányban osztja ketté, akkor ezt az ismérvértéket q -ad rendű vagy q -adik kvantikisnek nevezzük. A kvantilisek meghatározása egyúttal a sokaság egy osztályozását jelenti. Ezen osztályozás során egyenlő gyakoriságú osztályközöket kapunk.

5.3.1. Kvartilisek

A sokaságot négy egyenlő elemszámú részsokaságra bontjuk.

- Az alsó kvartilis: Q_1 , amely megmutatja, hogy a sokaság $1/4$ -része mely értéknél kisebb.
- A középső kvartilis: Q_2 , amely megegyezik a mediánnal.
- A felső kvartilis: Q_3 , amely megmutatja, hogy a sokaság $3/4$ -része mely értéknél kisebb, illetve $1/4$ része mely értéknél nagyobb.

Az alsó kvartilis sorszáma:

$$s = \frac{n+1}{4}$$

A felső kvartilis sorszáma pedig:

$$s = 3 * \frac{n+1}{4}$$

A képletek megegyeznek a medián képletével, természetesen a megfelelő fogalmak cseréje esetén.

Pl. a felső kvartilis kiszámítása:

$$Q_3 = q_3 + \frac{3 * \frac{n}{4} - f'_{q_3-1}}{f_{q_3}} * h$$

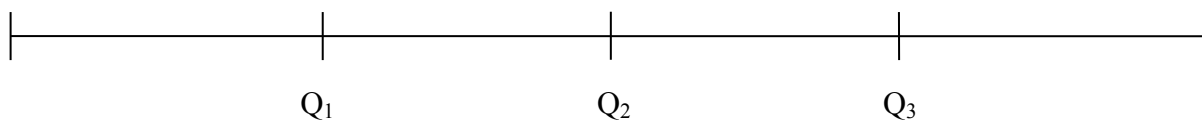
$3*n/4$: a felső kvartilis sorszáma

q_3 a felső kvartilist tartalmazó osztályköz alsó határa

f'_{q_3-1} : a felső kvartilist tartalmazó osztályközöt megelőző osztályköz kumulált gyakorisága

f_{q_3} : a felső kvartilist tartalmazó osztályköz gyakorisága

h : az osztályköz hossza



5.3.2. Tercilisek

Ha két osztópont segítségével három egyenlő részre osztjuk a sokaságot, akkor terciliseket kapunk T_1 és T_2). Az alsó tercilis (T_1) a sokaság alsó harmadolópontja, amely megmutatja, hogy mely értéknél kisebb a sokaság 1/3-ad része, illetve mely értéknél nagyobb a sokaság 2/3-ad része. A felső tercilis (T_2) a sokaság felső harmadolópontja, amely megmutatja, hogy mely értéknél kisebb a sokaság 2/3-ad része, illetve mely értéknél nagyobb a sokaság 1/3-ad része.

Kiszámításuk a mediánhoz hasonló.

Az alsó tercilis képletet:

$$T_1 = t_1 + \frac{\frac{n}{3} - f'_{t_{1-1}}}{f_{t_1}} * h$$

A felső tercilis képlete:

$$T_2 = t_2 + \frac{2 * \frac{n}{3} - f'_{t_{3-1}}}{f_{t_3}} * h$$

5.3.3. Decilisek

A decilisek számítása során 9 osztópont segítségével tíz egyenlő részre bontjuk a sokaságot, azaz tíz egyenlő gyakoriságú osztályközt hozunk létre. Az első decilis (D_1) megmutatja, hogy a sokaság 1/10-d része mely értéknél kisebb.

Kiszámítása a kvartilisekhez hasonló, természetesen a megfelelő jelölések változtatásával.

Például a 7. decilis értéke az alábbi képlettel számítható ki:

$$D_7 = d_7 + \frac{7 * \frac{n}{10} - f'_{d_{7-1}}}{f_{d_7}} * h$$

Bemutató feladat:

pl. 200 diák testmagasságát mérték meg:

Testmagasság (cm)	f_i	f'_i
151-160	30	30
161-170	80	110
171-180	70	180
181-190	10	190
190-200	10	200
Összesen	200	-

A felső kvartilis az adatok $\frac{3}{4}$ -t jelenti, azaz a 150. és 151. elem átlaga lenne ($3 * 200 / 4$), de osztályközös gyakorisági sornál a kumulált gyakoriság alapján az osztályközt kell először

meghatározni. A táblázatból kiolvasható, hogy a felső kvartilis a 3. osztályközben található, értéke:

$$Q_3 = q_3 + \frac{3 * \frac{n}{4} - f'_{q_{3-1}}}{f_{q_3}} * h = 170 + \frac{3 * \frac{200}{4} - 110}{70} * 10 = 175,71 \text{cm}, \text{ azaz a diákok } \frac{3}{4}\text{-ed része}$$

alacsonyabb 175,71cm-nél 1/4-ed része pedig ennél magasabb.

Az alsó tercilis (hasonlóan kell meghatározni, mint a mediánt vagy a kvartilist):

$$T_1 = t_1 + \frac{\frac{n}{3} - f'_{t_{1-1}}}{f_{t_1}} * h = 160 + \frac{\frac{200}{3} - 30}{80} * 10 = 164,58 \text{cm}, \text{ azaz a diákok } \frac{1}{3}\text{-ad része}$$

alacsonyabb 164,58cm-nél 2/3-ad része pedig ennél magasabb.

A 8. decilis (hasonlóan kell meghatározni, mint a mediánt vagy a kvartilist):

$$D_8 = d_8 + \frac{8 * \frac{n}{10} - f'_{d_{8-1}}}{f_{d_8}} * h = 170 + \frac{8 * \frac{200}{10} - 110}{70} * 10 = 177,14 \text{cm} \text{ azaz a diákok } \frac{8}{10}\text{-ed (80\%)}$$

része alacsonyabb 177,14 cm-nél 2/10-ed (20%) része pedig ennél magasabb.

6. lecke: A szóródás és mérése

Követelmények:

Ön akkor sajátította el megfelelően a tananyagot, ha

- Felsorolásból ki tudja választani a szóródásra igaz állítást.
- Felsorolásból ki tudja választani a szórásra igaz állítást.
- Felsorolásból ki tudja választani a szórás definícióját
- Felsorolásból ki tudja választani a belső szórásra igaz állítást.
- Önállóan meg tudja határozni az adatsor terjedelmét, szórását, relatív szórását, az abszolút és átlagos eltérést.
- Ki tudja számolni és értelmezni tudja a belső szórást, külső szórást és a teljes szórást.
- Meg tudja határozni az átlagos különbség mutatószámát, és értelmezni annak jelentését.

A középértékek csak egyetlen tulajdonságát rögzítik az eloszlásnak, ezért a statisztikai sokaság jellemzésére általában nem elegendőek. Egy sokaság eloszlása nagyon sokféle lehet, így az eloszlás formája is különböző lehet.

Az ismértértekek szóródásának mérésére és mérőszámok képzésére különböző lehetőségek vannak.

Szóródáson azonos fajta számszerű adatok (általában egy mennyiségi ismértértekek) különbözőségét értjük. A szóródás mérése az ismértértekek valamely középértéktől vett eltérései vagy egymás közötti különbségei alapján történik.

Ezen eltérések, különbségek alapján számított mérőszámok a **szóródás abszolút mutatói**, amelyek mértékegysége megegyezik a megfigyelt ismértértekek mértékegységével.

A szóródás **relatív mutatói** elvonatkoztatnak az ismértértekek mértékegységétől, nagyságrendjétől, a szóródás térbeli vagy időbeli összehasonlítására szolgálnak.

A szóródás legfontosabb mutatószámai:

- terjedelem (R),
- átlagos eltérés (δ),
- szórás (σ , s),
- relatív szórás (V, CV)
- átlagos különbség (G).

Az egyes mutatószámokra az elmélet végén talál példát!

6.1. A szóródás terjedelme (R)

Azt mutatja meg, hogy a sokaság elemei milyen értékintervallumban helyezkednek el.

A szóródás terjedelme alatt az előforduló legkisebb és legnagyobb érték különbségét értjük:

$$R = x_{\max} - x_{\min}$$

A szóródás terjedelme egy nagyon egyszerű és csak közelítő mérőszáma a szóródásnak. Az eloszlásnak csak a legkisebb és a legnagyobb értékét veszi figyelembe, és a két szélső érték közötti többi értéket nem. Arról, hogy a többi érték hol helyezkedik el, nem mond semmit.

6.2. Átlagos eltérés (δ)

Az átlagos eltérés az ismértékek számtani átlagtól vett eltérésein alapul, azaz az egyes értékek számtani átlagtól vett eltérései abszolút értékének számtani átlaga:

$$\delta = \frac{\sum |x_i - \bar{x}|}{n} \qquad \delta = \frac{\sum f_i |x_i - \bar{x}|}{\sum f_i}$$

Mivel az átlag nagyságát nem a súlyok abszolút nagysága, hanem a súlyok aránya befolyásolja ezért a relatív gyakoriságból is kiszámítható:

$$\delta = \sum g_i |x_i - \bar{x}|$$

Bemutató feladat:

Pl. 50 izzó élettartamát vizsgálva.

Élettartam (óra)	f_i	Oszt. Közép	$ x_i - \bar{x} $	g_i	$g_i * x_i - \bar{x} $
600-700	5	650	210	0,1	21,0
700-800	11	750	110	0,22	24,2
800-900	16	850	10	0,32	3,2
900-1000	10	950	90	0,2	18,0
1000-1100	8	1050	190	0,16	30,4
Összesen	50				96,8

$\delta = 96,8$

Ugyanezekkel az adatokkal dolgozva, csak a gyakoriságokat változtatjuk meg:

Élettartam (óra)	f_i	Oszt. Közép	$ x_i - \bar{x} $	g_i	$g_i * x_i - \bar{x} $
600-700	11	650	210	0,22	46,2
700-800	5	750	110	0,1	11,0
800-900	10	850	10	0,2	2,0
900-1000	16	950	90	0,32	28,8
1000-1100	8	1050	190	0,16	30,4
Összesen	50				118,4

$\delta = 118,4$

6.3. A szórás

A legismertebb és a leggyakrabban használt mutató a szórás. A szórás az egyes értékek számtani átlagtól való eltéréseinek négyzetes átlaga, azaz megmutatja, hogy az egyes ismértékek átlagosan mennyivel térnek el az átlagtól.

Tapasztalati szórás: az egész sokaságra vonatkozik.

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad \sigma = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}} \quad \sigma = \sqrt{\sum g_i (x_i - \bar{x})^2}$$

elméleti szórás: a mintából számoljuk ki.

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad s = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i - 1}}$$

Az eredeti táblázat (50 villanyégő) adataiból számítsuk ki a szórást

$$\sigma = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}} = \sqrt{\frac{5 * (650 - 860)^2 + 11 * (750 - 860)^2 + \dots + 8 * (1050 - 860)^2}{5}} = 120,42$$

Tulajdonságai:

- Ha az ismértértékhez hozzáadunk egy állandót (A), a szórás értéke nem változik, mivel ilyenkor a számtani átlag is pont ezzel az állandóval lesz nagyobb.
- Ha az ismértértéket megszorozzuk egy állandóval (B), akkor a szórás $|B|$ -szeresére változik, mivel ebben az esetben a számtani átlag értéke „B”-szer nagyobb lesz.

A szórás mindig nagyobb értékű, mint az átlagos eltérés!

6.3.1. Variancia

A variancia, vagy szórásnégyzet önálló mutatóként is használatos. Nagy jelentősége például a varianciaanalízisben van, amely több középérték összehasonlítására szolgál. Többféle jelölést találunk a szakirodalomban: s^2 , SS, MQ.

A szórásnégyzet, azaz a variancia a számtani és a négyzetes átlag ismeretében is kiszámolható:

$$\sigma^2 = \overline{x_q^2} - \bar{x}^2$$

6.3.2. A szórásnégyzet és a szórás felbontása

Ha a sokaság részekre, azaz részsokaságokra bontható, akkor a szórásnégyzetet is fel lehet bontani két részre:

- Belső szórásnégyzetre: σ_B^2
- Külső szórásnégyzetre σ_K^2

A belső- és külső szórásnégyzet összege a teljes szórásnégyzet:

$$\sigma_T^2 = \sigma_B^2 + \sigma_K^2$$

A **belső szórásnégyzet** az egyes részsokaságok (csoporton belüli elemek és a csoport főátlaga között értelmezett különbségekből meghatározott) szórásnégyzeteinek az átlaga. Ebben a mérőszámokban a csoportosító ismérven kívüli, egyéb tényezőknek a szóródó ismérve gyakorolt hatása jelenik meg.

$$\sigma_B^2 = \frac{\sum n_j * \sigma_j^2}{n}, \text{ ahol } j: \text{ a csoportosító ismért változatainak a száma}$$

σ_j^2 : a j-edik részsokaság szórásnégyzete

n_j : a j-edik részsokaság elemszáma
 n : a sokaság elemszáma

A **külső szórásnégyzet** az egyes részsokaságok átlagai (részátlag) és a teljes sokasági átlag (főátlag) eltérés négyzetösszegének az átlaga (a részsokaságok nagyságával súlyozva). Ebben a mutatószámban fejeződik ki a csoportképző ismérvnek a vizsgált mennyiségi ismérvre gyakorolt hatása.

$$\sigma_K^2 = \frac{\sum n_j * (\bar{x}_j - \bar{x})^2}{n}, \text{ ahol } \bar{x}_j : \text{ a részsokaságok átlaga}$$

$$\bar{x} : \text{ a teljes sokaság átlaga (főátlag)}$$

A **teljes szórás** (σ_T) azt fejezi ki, hogy a vizsgált sokaságban az egyes értékek átlagosan mennyivel térnek el a főátlagtól.

A **belső szórás** (σ_B) azt fejezi ki, hogy a részsokaságok egyes értékei átlagosan mennyivel térnek el a saját részátlaguktól.

A **külső szórás** (σ_K) azt fejezi ki, hogy a vizsgált sokaságban az egyes részsokaságok átlagai átlagosan mennyivel térnek el a sokasági főátlagtól.

Bemutató feladat:

A hallgatók szakok szerinti csoportosításban az alábbi eredményt érték el egy vizsgán. A maximálisan elérhető pontszám 100 pont volt.

Szak	Hallgatók száma (fő)	Az elért eredményük átlagpontszáma	Az eredmények szórás
Közgazdász	20	63	5,6
Jogász	30	60	5,9
Mérnök	25	50	6,4

Számítsa ki az adatok alapján a belső-, a külső és a teljes szórást, majd értelmezze azokat!

A vizsga átlagpontszáma, azaz a főátlag:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{20 * 63 + 30 * 60 + 25 * 50}{20 + 30 + 25} = 57,47$$

A belső szórásnégyzet:

$$\sigma_B^2 = \frac{\sum n_j * \sigma_j^2}{n} = \frac{20 * 5,6^2 + 30 * 5,9^2 + 25 * 6,4^2}{20 + 30 + 25} = 35,94$$

A belső szórás:

$\sigma_B = \sqrt{\sigma_B^2} = \sqrt{35,94} = 5,99$, Az egyes szakokon belül a hallgatók eredménye átlagosan 5,99 ponttal tér el az adott szakon elért átlagos pontszámtól.

A külső szórásnégyzet:

$$\sigma_K^2 = \frac{\sum n_j * (\bar{x}_j - \bar{x})^2}{n} = \frac{20 * (63 - 57,47)^2 + 30 * (60 - 57,47)^2 + 25 * (50 - 57,47)^2}{20 + 30 + 25} = 29,32$$

A külső szórás:

$\sigma_K = \sqrt{\sigma_K^2} = \sqrt{29,32} = 5,41$, az egyes szakok átlagos vizsgaeredménye 5,41 ponttal tér el az összes hallgató által elért átlageredménytől.

A teljes szórásnégyzet:

$$\sigma_T^2 = \sigma_B^2 + \sigma_K^2 = 35,94 + 29,32 = 65,26$$

A teljes szórás:

$\sigma = \sqrt{\sigma^2} = \sqrt{65,26} = 8,08$, A vizsgán megjelent hallgatók eredménye átlagosan 8,08 ponttal tér el az összes hallgató által elért átlageredménytől.

6.3.3. A szóráshányados

A szórásnégyzet felbontásának segítségével meghatározhatjuk a szóráshányadost, amely a csoportképző ismerv (területi, minőségi, esetleg mennyiségi) és a vizsgált mennyiségi ismerv kapcsolatát jellemzi, kapcsolatának szorosságát fejezi ki:

$$H = \sqrt{\frac{\sigma_K^2}{\sigma^2}} = \sqrt{1 - \frac{\sigma_B^2}{\sigma^2}} = \frac{\sigma_K}{\sigma}$$

Értéke: $0 \leq H \leq 1$ lehet.

6.4. Relatív szórás (V , CV)

Sok esetben a szóródás vizsgálatára a szórás, mint átlagos szóródási paraméter nem elegendő. Továbbá bizonyos esetekben szükség lehet arra, hogy az ismervértékek nagyságrendjétől illetve mértékegységétől elvonatkoztatott „tisztá szám” jellegű mérőszámmal mérjünk, és összehasonlíthatóvá tegyük a szórásokat. E célból bármelyik eddig bemutatott mérőszámot viszonyíthatjuk az átlaghoz. A legismertebb ilyen szóródási mérőszám a relatív szórás vagy variációs koefficiens, amely kifejezi, hogy a szórás az átlag hányad része. Általában százalékban kifejezve adjuk meg.

$$V = \frac{\sigma}{x}$$

Kifejezi, hogy a sokaság egyes egyedeinek értéke átlagosan hány százalékkal tér el az átlagtól.

A relatív szórás százalékosan kifejezve a gazdasági gyakorlatban a változékonyságot az alábbiak szerint minősíthetjük:

0-10%: állandóságot (homogenitást)

10-20%: közepes változékonyságot,

20-30%: erős változékonyságot

30% felett szélsőséges ingadozást fejez ki.

6.5. Átlagos különbség (G)

Az átlagos eltérés (δ) és a szórás (σ) a számtani átlag alapján méri az ismervértékek különbözőségét. A szóródás az ismervértékek egymás közötti különbségei alapján is vizsgálható, illetve mérhető. E mutató bevezetését Corrado Gini, olasz statisztikus javasolta, ezért **Gini-féle** mérőszámnak is nevezik.

Az átlagos különbség (G) az ismervértékek egymástól számított különbségei abszolút értékének számtani átlaga.

$$G = \frac{\sum \sum |x_i - x_j|}{n(n-1)} \quad G = \frac{\sum \sum f_i \cdot f_j |x_i - x_j|}{n(n-1)}$$

Az átlagos különbség azt fejezi ki, hogy az egyes ismértékek átlagosan mennyivel térnek el egymástól.

Számításához három munkatábla szükséges: $|x_i - x_j|$

	x_{j1}	x_{j2}		x_{jn}
x_{i1}				
x_{i2}				
x_{in}				

$f_i \cdot f_j$

	f_{j1}	f_{j2}		f_{jn}
f_{i1}				
f_{i2}				
x_{in}				

$f_i \cdot f_j \cdot |x_i - x_j|$

				Összesen
	A két tábla megfelelő celláinak szorzata.			

Bemutató feladat

Pl. 50 izzó élettartamát vizsgálva.

Élettartam (óra)	f_i	Oszt. Közép	$ x_i - x_j $	$f_i \cdot$ $ x_i - x_j $
601-700	5	650	210	1050
701-800	11	750	110	1210

801-900	16	850	10	160
901-1000	10	950	90	900
1001-1100	8	1050	190	1520
Összesen	50			4840

R (Terjedelem): 1100-600= 500 óra

Átlag= 860 óra: egy villanyégő átlagos élettartama.

Átl. Eltérés: 96,8 óra: a villanyégők élettartama átlagosan 96,8 órával tér el a sokaság átlagától

Szórás 120,42 óra: a villanyégők élettartama átlagosan 120,42 órával tér el a sokaság átlagától

CV= 14%: villanyégők élettartama átlagosan 14%-kal tér el a sokaság átlagától

$$|x_i - x_j|$$

Mivel a főátlóra szimmetrikus mátrixot kapunk, ezért a táblázatnak csak egy részét töltjük ki.

	650	750	850	950	1050
650	0	100	200	300	400
750		0	100	200	300
850			0	100	200
950				0	100
1050					0

$$f_i * f_j$$

	5	11	16	10	8
5	25	55	80	50	40
11		121	176	110	88
16			256	160	128
10				100	80
8					64

$$f_i * f_j * |x_i - x_j|$$

	0	5500	16000	15000	16000	52500
		0	17600	22000	26400	66000
			0	16000	25600	41600

				0	8000	8000
					0	0
						168100

Mivel a táblázatot nem teljesen töltöttük ki, ezért a kapott eredményt meg kell szorozni kettővel.

$G = 2 \cdot 168100 / 50 \cdot 49 = 137,22$ óra Az egyes égők élettartama átlagosan 137,22 órával különböznek egymástól.

7. lecke: Aszimmetria, koncentráció

Követelmények:

Ön akkor sajátította el megfelelően a tananyagot, ha

- El tudja dönteni, hogy aszimmetria esetén milyen összefüggés van a módusz és a medián közötti!
- Felsorolásból ki tudja választani ferdeségi mutató értékére vonatkozó igaz állítást.
- Ki tudja számolni az aszimmetria mutatószámait.
- Felsorolásból ki tudja választani a kiszámolt mutatószámok helyes értelmezését.
- Az eredmények alapján ki tudja választani az aszimmetria mérőszámainak jelentését.

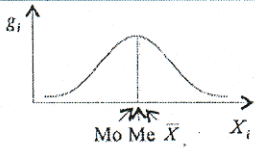
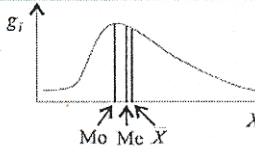
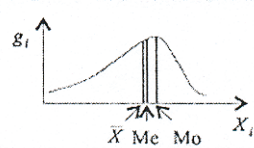
7.1. Aszimmetria

A statisztikai sorokat ábrázolva nagyon változatos görbéket kapunk, de nagy többségük szabályszerűséget mutat. Az eloszlások típusától függ a görbe lefutása. Az eloszlások az alábbiak lehetnek:

- egy móduszú,
 - Szimmetrikus: az adatok átlaga, módusza és mediánja azonos.
 - Aszimmetrikus: az adatok átlaga, módusza és mediánja nem azonos.
- több móduszú:

Az eloszlások ábrázolásánál a vízszintes tengelyen az ismérvváltozatok értékeit tüntetjük fel, a függőleges tengelyen az ismérvváltozatokhoz tartozó megfigyelt adatokat.

Az **egy móduszú** gyakorisági soroknak egy helyi maximuma van. 3 típusát különböztetjük meg.

Szimmetrikus eloszlás	Aszimmetrikus eloszlás	
	bal oldali	jobb oldali
		
$Mo = Me = \bar{X}$	$Mo < Me < \bar{X}$	$Mo > Me > \bar{X}$
$(Q_3 - Me) = (Me - Q_1)$	$(Q_3 - Me) > (Me - Q_1)$	$(Q_3 - Me) < (Me - Q_1)$

Egy móduszú eloszlásoknál az aszimmetria (vagy ferdeség) fokának jellemzésére több mutatószám is használható. A mutatószámokkal szemben támasztott főbb követelmények:

- A mérőszám dimenzió nélküli legyen,
- Szimmetrikus eloszlás esetén a nulla értéket vegye fel.

7.1.1. Pearson féle mutatószám

Az aszimmetria egyik általános úgynevezett „A mérőszámának” kidolgozása K. Person nevéhez fűződik. A mérőszám képzése abból indul ki, hogy szimmetrikus eloszlásnál a számtani átlag és a módusz megegyezik, míg aszimmetrikus eloszlásnál eltérnek egymástól.

$$A = \frac{\bar{x} - M_0}{\sigma}$$

Ha $0,5 \leq |A|$, akkor már erős aszimmetriáról beszélünk. Szélsőséges esetben 1-nél nagyobb érték is előfordulhat. Ha $A > 0$, akkor baloldali, ha $A < 0$, akkor jobboldali aszimmetriáról beszélünk. Mivel a módusz osztályközös gyakorisági sorból lett meghatározva, nem eléggé megbízható, ezért a mediánból is kiszámolható az aszimmetria egy másik Pearson-féle mutatószámmal.

Bemutató feladat:

Egy kiskereskedőnél az egyik hétvégén megvizsgálták az egyes vásárlások összegét.

Vásárláskor fizetett összeg (Ft)	Vásárlók száma (fő)	Osztályközép	Kumulált gyakoriság
1-250	29	125	29
251-500	34	375	63
501-750	54	625	117
751-1000	39	875	156
1001-1250	20	1125	176
1251-1500	18	1375	194
1501-1750	12	1625	206
Összesen	206		

$$Mo = 500 + \frac{54 - 34}{(54 - 34) + (54 - 39)} * 250 \approx 643$$

$$\bar{x} = 733$$

$$\sigma = 421$$

$$A = \frac{733 - 643}{421} = 0,214$$

Az adatsorra gyenge baloldali aszimmetria jellemző.

7.1.2. . Pearson féle mutatószám2

Ez azon gyakorlati megfigyelésen alapul, hogy a mérsékelten aszimmetrikus eloszlások esetében a medián az átlagtól az átlag és a módusz közötti teljes távolság mintegy egyharmadával balra vagy jobbra esik

$$P = \frac{3(\bar{x} - M_e)}{\sigma}$$

A mutató értéke -3 és $+3$ közé esik. A 0 -hoz közeli érték gyenge aszimmetriát jelent. Normális eloszlás esetén $P=0$. Ha $P>0$, akkor baloldali-, ha $P<0$, akkor pedig jobboldali aszimmetriáról beszélünk. Ha $P< -0,5$, illetve $P>0,5$ esetében erős ferdeség jellemzi az eloszlást.

Bemutató feladat:

Az előző feladatból:

$$Me = 685$$

$$P = \frac{3(733 - 685)}{421} = 0,342$$

Hasonló eredményt kaptunk, azaz gyenge baloldali asszimetria.

7.2. Ferdeségi mutató

A ferdeségi mutató az alsó és a felső kvartilis mediántól való eltérésének egymáshoz viszonyított nagyságán alapul.

$$F = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)}$$

A mutató megoszlási viszonzyszám jellegű, tehát értéke -1 és +1 között mozog.

Ha $F=0$, akkor normális az eloszlás. Ha $F>0$, akkor baloldali-, ha $F<0$, akkor pedig jobboldali aszimmetriáról beszélünk.

Baloldali aszimmetria esetén a medián az alsó kvartilishez (Q_1), jobboldali aszimmetria esetén pedig a felső (Q_3) kvartilishez esik közelebb. A 0-hoz közeli érték az asszimmetria hiányát mutatja, de 0,2-nél nagyobb érték esetén már nagyfokú aszimmetriáról beszélünk.

Bemutató feladat:

Az előző feladatból:

$$Q_1 = 415$$

$$Q_3 = 990$$

$$F = \frac{(990 - 685) - (685 - 415)}{(990 - 685) + (685 - 415)} = 0,061$$

7.3. Koncentráció

Koncentrációnak nevezzük azt a jelenséget, hogy a sokasághoz tartozó teljes értékösszeg (az ismerv előfordulási értékeinek összege) jelentős része a sokaság kevés egységére összpontosul. Az értékösszeg jele: **Si**

Ha a sokaság elemszáma (n) kicsi, akkor már önmagában is koncentrációt jelent, hiszen a teljes értékösszeg kevés egységre összpontosít. **A koncentráció kiszámításához:**

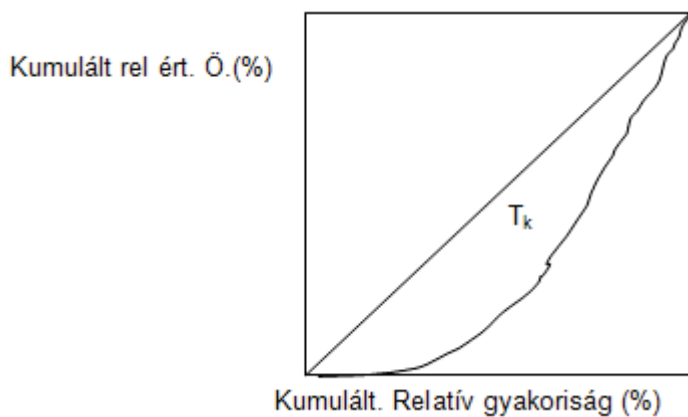
- a relatív gyakoriság, és

$$g_i = \frac{f_i}{n}$$

- a relatív értékösszeg meghatározása szükséges:

$$z_i = \frac{s_i}{\Sigma s_i}$$

A koncentráció vizsgálatának egyik legfontosabb és legelterjedtebb eszköze a **Lorenz-görbe**. Ez egy egységnyi oldalú négyzetben elhelyezett vonaldiagram, amely a kumulált relatív gyakoriság (g_i) függvényében ábrázolja a kumulált relatív értékösszegeket (z_i).



Ha az ábrázolt görbe az átlón fut, akkor a sokaságban nem figyelhető meg koncentráció, ha átlóhoz közel, akkor pedig gyenge koncentráció jellemzi a sokaságot. Minél nagyobb az átló és a görbe közötti terület, annál nagyobb a koncentráció mértéke.

A koncentrációs együttható egy megoszlási viszonzszám is egyben. A területekhez integrálszámításra van szükség, de a mutató a Gini-féle mutató segítségével is kiszámítható.

Koncentrációs együttható:
$$K = \frac{T_k}{T_b} = \frac{G}{2x}$$

ahol

T_k : a koncentrációs terület nagysága (a két görbe által bezárt terület).

T_b : az átló alatti terület

A $K=0$, a koncentráció teljes hiányát mutatja. A $K=1$ pedig a teljes koncentrációságot.

Minél kisebb a koncentrációs, annál kevésbé szóródnak az adatok.

Bemutató feladat:

A Gini-féle mutatónál található feladat folytatása.

Pl. 50 izzó élettartamát vizsgálva.

Élettartam (óra)	f_i	Öszt. Közép
601-700	5	650
701-800	11	750
801-900	16	850
901-1000	10	950
1001-1100	8	1050
Összesen	50	

Átlag= 860 óra

$G = 2 \cdot 160900 / (50 \cdot 49) = 137,22$ óra

$K = \frac{G}{2x} = \frac{137,22}{2 \cdot 860} = 0,080$, azaz gyenge koncentráció jellemzi az adatsort.

8. lecke: A sztochasztikus kapcsolatok

Követelmények:

Ön akkor sajátította el megfelelően a tananyagot, ha

- Felismeri a sztochasztikus kapcsolatokat leíró helyes definíciót.
- A fogalmakhoz párosítani tudja azok jelentését (asszociáció, vegyes kapcsolat, korreláció)
- Felsorolásból ki tudja választani, mikor használható a Yule-féle együttható.
- Felsorolásból ki tudja választani, mikor használható a Csuprov-féle együttható. (Ez a modulzáróban jön elő)
- Felsorolásból ki tudja választani, mikor használható és milyen formában fejezhető ki a H-mutató. (Ez a modulzáróban jön elő)
- Felsorolásból ki tudja választani az asszociáció jelentését.
- Önállóan meg tudja határozni a Yule-féle együttható értékét.
- A kiszámított adat alapján felsorolásból ki tudja választani két ismerv közötti kapcsolat szorosságát.
- Megadott adatok alapján ki tudja tölteni a kontingencia-táblázatot.
- Felsorolásból ki tudja választani a belső szórás fogalmát.
- Felsorolásból ki tudja választani a H-mutató jelentését.
- Felsorolásból ki tudja választani a H-mutató lehetséges értékeit.
- Ki tudja számolni a belső szórást, a külső szórást, a H-mutató és a H^2 -mutató értékét.
- Felsorolásból ki tudja választani a H-mutató és a H^2 -mutató adott értékének jelentését.
- Felsorolásból ki tudja választani, minek a jellemzésére alkalmas a korrelációs együttható,
- Listából ki tudja választani a korrelációs együttható lehetséges értékeit,
- Adott ismérvekről el tudja dönteni, milyen korrelációs együtthatóval jellemezhetők,
- Ki tudja választani a korrelációs számításra alkalmas feladatokat, ehhez nem találtam kérdést
- Adott adatsor alapján ki tudja számítani a korrelációs együtthatót.
- Ki tudja számolni és értelmezni az előjelkorrelációs együtthatót. Ehhez nem találtam kérdést
- Ki tudja számolni a rangkorrelációs együtthatót.

Ha egy sokaságot statisztikai módszerekkel vizsgálunk, akkor többféle szempont (ismerv) alapján is csoportosíthatjuk (pl. a hallgatókat: nem, magasság, tanulmányi átlag, stb.). A különféle csoportosításokkal kapcsolatban feltehető a kérdés, hogy mutatnak-e ezek hasonlóságot, azonosságot, vagy teljesen más jellegűek. Ha a vizsgált sokaság különböző ismérvek alapján feltáruló szerkezete hasonlóságot, azonosságot mutat, akkor ismérvek közötti kapcsolatról beszélünk. Ez a kapcsolat lehet:

- sztochasztikus jellegű: statisztikai valószínűségen alapuló,
- teljes meghatározottságú: pl. a születési év és az életkor között.

Sztochasztikus (valószínűségi) kapcsolatnak nevezzük az olyan összefüggést, amikor az egyik jelenség egy másik jelenségre hatással van, de nem határozza meg egyértelműen annak alakulását, mert arra más tényezők is hatnak.

Két minőségi ismerv (pl. iskolai végzettség és beosztás), vagy két területi ismerv (születési hely és lakhely), vagy egy minőségi és egy területi ismerv (választott szak és a gyakorlati hely) közötti kapcsolat elemzésére szolgál. Az említett esetekben nem számadatokkal megfogalmazható ismérvek alapján történt a csoportosítás.

8.1 Asszociáció

Az asszociáció (ismérvék közötti kapcsolat) kimutatása érdekében a sokaság elemeit úgynevezett kontingencia táblába rendezzük. Két ismérv esetén a tábla a következő:

Ismérvék	1	2	Σ
1	f_{11}	f_{12}	$f_{1\cdot}$
2	f_{21}	f_{22}	$f_{2\cdot}$
Σ	$f_{\cdot 1}$	$f_{\cdot 2}$	n

Bemutató példa: Egy vállalatnál 50 diplomás dolgozik, és közülük 15 a nő. A középfokú végzettségűek létszáma 150, és ebből 60 a nő. Rendezzük az adatokat táblázatba:

Ismérvék	Diplomás	Középfokú végzettségű	Összesen
Férfi	35	90	125
Nő	15	60	75
Összesen	50	150	200

Kettőnél több ismérvváltozat esetén értelemszerűen több sor vagy oszlop lesz a kontingencia táblázatban.

Az asszociációs mutatók a következők lehetnek:

- Yule-féle
- Csuprov-féle
- Cramer-féle.

8.1.1. Yule-féle asszociációs együttható

Csak alternatív (azaz két ismérv) ismérvek esetén használható. Alapgondolata a függetlenséghez kapcsolódik. Két egymástól független esemény együttes bekövetkezésének valószínűsége egyenlő a két esemény bekövetkezési valószínűségének szorzatával:

$$\frac{f_{\cdot 1}}{n} * \frac{f_{1\cdot}}{n} = \frac{f_{11}}{n}$$

Mivel alternatívák az ismérvek, ezért a függetlenség az alábbi:

$$\frac{f_{11}}{f_{21}} = \frac{f_{12}}{f_{22}}$$

ha $f_{11} * f_{22} - f_{21} * f_{12} = 0$, akkor nincs kapcsolat.

Minden nullától való eltérés esetén kiszámítható a kapcsolat szorossága.

$$a = \frac{f_{11} * f_{22} - f_{21} * f_{12}}{f_{11} * f_{22} + f_{21} * f_{12}} \text{ vagy } a = \frac{g_{11} * g_{22} - g_{21} * g_{12}}{g_{11} * g_{22} + g_{21} * g_{12}}$$

$$-1 \leq a \leq 1$$

A '0' a kapcsolat hiányát, azaz a függetlenséget jelenti. Az 1 pedig a teljes meghatározottságot. Ha $|a| > 0,7$, akkor a kapcsolat szorossága már erősnek mondható.

Bemutató feladat

A hallgatókat a szakterület és a karriercéllal való rendelkezés alapján csoportosítottuk:

<i>Karrier cél</i>	<i>Van</i>	<i>nincs</i>	<i>Összesen</i>
<i>szak</i>			
<i>mérnök</i>	32	68	100
<i>közgazdász</i>	22	78	100
<i>Összesen</i>	54	146	200

1. *Vizsgálja meg, hogy van-e kapcsolat a hallgató szakterülete és a karrier céllal való rendelkezés között?*
2. *Mit jelent a kiszámított mutató?*

$$a = \frac{32 * 78 - 22 * 68}{32 * 78 + 22 * 68} = 0,25$$

A vizsgált hallgatói körben a szakterület és a karrier céllal rendelkezés között gyenge kapcsolat van.

8.1.2. Csuprov-féle asszociációs együttható

Ha az ismervváltozatok száma kettőnél több, akkor használjuk. Ez is a függetlenségvizsgálathoz kapcsolódik.

$$T = \sqrt{\frac{\chi^2}{n * \sqrt{(s-1) * (t-1)}}}$$

T: Csuprov-féle együttható,

s: az egyik ismerv változatainak száma

t: a másik ismerv változatainak száma

n: a sokaság elemszáma

$$\chi^2 = \sum_{i=1}^s \sum_{j=1}^t \frac{(f_{ij} - f_{ij}^*)^2}{f_{ij}^*} \quad f_{ij}^* = \frac{f_{i.} * f_{.j}}{n}$$

f_{ij} : tapasztalati gyakoriság

f_{ij}^* : elméleti gyakoriság

$$\text{ha } s < t \quad 0 \leq T \leq \sqrt{\frac{s-1}{t-1}}$$

Ha $s=t$, akkor $T=1$, azaz az ismérvek között összefüggés nagyon intenzív. Általában úgy válasszuk meg az ismérvek jelölését, hogy $s < t$ legyen.

Bemutató feladat

A hallgatókat a szakterület és az oktatással való elégedettség rendelkezés alapján csoportosítottuk:

<i>Szak</i>	<i>Mérnök</i>	<i>Közgazdász főisk.</i>	<i>Közgazdász egyetem</i>	<i>jogász</i>	<i>Összesen</i>
<i>elégedettség</i>					
<i>Nem tudja</i>	69	83	99	53	304
<i>Nem elégedett</i>	12	56	26	5	99
<i>elégedett</i>	19	91	35	42	187

Összesen	100	230	160	100	590
----------	-----	-----	-----	-----	-----

Elméleti előfordulások számítása: $f_{ij}^* = \frac{f_{i.} * f_{.j}}{n}$

$$f_{11}^* = \frac{100 * 304}{590} = 51,5 \quad f_{12}^* = \frac{230 * 304}{590} = 118,5 \quad f_{31}^* = \frac{100 * 187}{590} = 31,7, \text{ stb.}$$

Szak elégedettség	Mérnök	Közgazdász főisk.	Közgazdász egyetem	jogász	Összesen
Nem tudja	51,5	118,5	82,5	51,5	304
Nem elégedett	16,8	38,6	26,8	16,8	99
elégedett	31,7	72,9	50,7	31,7	187
Összesen	100	230	160	100	590

$$\chi^2 = \sum_{i=1}^s \sum_{j=1}^t \frac{(f_{ij} - f_{ij}^*)^2}{f_{ij}^*} = \frac{(69 - 51,5)^2}{51,5} = 5,95$$

A χ^2 – összetevői:

Szak elégedettség	Mérnök	Közgazdász főisk.	Közgazdász egyetem	jogász	Összesen
Nem tudja	5,95	10,64	3,30	0,04	19,93
Nem elégedett	1,37	7,840	0,02	8,29	17,52
elégedett	5,09	4,49	4,86	3,35	17,79
Összesen	12,41	22,97	8,18	11,68	55,24

$$T = \sqrt{\frac{\chi^2}{n * \sqrt{(s-1) * (t-1)}}} = \sqrt{\frac{55,24}{590 * \sqrt{(3-1) * (4-1)}}} = 0,1955$$

Ki kell számolni a T-intervallumát:

$$T = \sqrt[4]{\frac{3-1}{4-1}} = 0,9036$$

azaz, $0 \leq T \leq 0,9036$

A szakterület és az oktatással való elégedettség között gyenge a kapcsolat ($T=0,1955$).

8.1.3. Cramer-féle asszociációs együttható

Az asszociációs összefüggések térbeli vagy időbeli összehasonlításakor gondot okozhat, hogy ugyanazon ismerv változatainak száma nem azonos a két helyen vagy a két időpontban (az előző bemutató példában, ha a jogászok nem minden kérdésre válaszolnak). Ilyen esetben a

Csuprov-féle asszociációs együttható nem használható az elemzés során, helyette a Cramer-féle asszociációs együtthatót használjuk.

$$C = \sqrt{\frac{\chi^2}{n^*(s-1)}}$$

$$0 \leq C \leq 1$$

8.2. Vegyes kapcsolat

Ha egy vizsgált sokaságot egy mennyiségi ismerv és egy minőségi ismerv alapján is felosztunk, és arra keressük a választ, hogy a két felosztás között, azaz a két ismerv között van-e összefüggés, akkor egy vegyes kapcsolatot elemzünk. A kapcsolat meglétét, vagy annak hiányát a H-mutató segítségével határozhatjuk meg.

Az összetett sokaságot felosztjuk részsokaságokra, és megvizsgáljuk:

1. a teljes szórast σ_T (az egyedi adatok és a főátlag eltérése).
2. a külső szórast σ_K (a részátlag és a főátlag eltérése).
3. a belső szórast σ_B (az egyedi értékek és a saját részátlaguk eltérése).

$$\sigma_T^2 = \sigma_K^2 + \sigma_B^2$$

Ez az összefüggés lehetővé teszi, hogy egy vegyes kapcsolat jellemzésére alkalmas mutatót képezzünk:

$$H = \frac{\sigma_K}{\sigma_T} = \sqrt{1 - \frac{\sigma_B}{\sigma_T}}$$

$$0 \leq H \leq 1$$

A kapcsolat jellemzésére a mutató négyzetét is használhatjuk (H^2), amelyet százalékos formában adunk meg. A mutató arra ad választ, hogy hány százalékban magyarázza a csoportosító ismerv (minőségi vagy területi) a sokaság mennyiségi ismerv szerinti szóródását.

$$H^2 = \frac{\sigma_K^2}{\sigma_T^2}$$

Bemutató feladat:

A hallgatók szakok szerinti csoportosításban az alábbi eredményt érték el egy vizsgán. A maximálisan elérhető pontszám 100 pont volt.

Szak	Hallgatók száma (fő)	Az elért eredményük átlagpontszáma	Az eredmények szórás
Közgazdász	20	63	5,6
Jogász	30	60	5,9
Mérnök	25	50	6,4

Számítsa ki az adatok alapján a belső-, a külső és a teljes szórast, majd értelmezze azokat!

A vizsga átlagpontszáma:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{20 * 63 + 30 * 60 + 25 * 50}{20 + 30 + 25} = 57,47$$

A belső szórásnégyzet:

$$\sigma_B^2 = \frac{\sum n_j * \sigma_j^2}{n} = \frac{20 * 5,6^2 + 30 * 5,9^2 + 25 * 6,4^2}{20 + 30 + 25} = 35,94$$

A belső szórás:

$$\sigma_B = \sqrt{\sigma_B^2} = \sqrt{35,94} = 5,995,.$$

A külső szórásnégyzet:

$$\sigma_K^2 = \frac{\sum n_j * (\bar{x}_j - \bar{x})^2}{n} = \frac{20 * (63 - 57,47)^2 + 30 * (60 - 57,47)^2 + 25 * (50 - 57,47)^2}{20 + 30 + 25} = 29,32$$

A külső szórás:

$$\sigma_K = \sqrt{\sigma_K^2} = \sqrt{29,32} = 5,415,$$

A teljes szórásnégyzet:

$$\sigma_T^2 = \sigma_B^2 + \sigma_K^2 = 35,94 + 29,32 = 65,26$$

A teljes szórás:

$$\sigma = \sqrt{\sigma^2} = \sqrt{65,26} = 8,078,$$

$$H = \frac{\sigma_K}{\sigma_T} = \frac{5,415}{8,078} = 0,6703 \quad H^2 = 0,4493 = 44,93\%$$

A szak és a vizsgán elért pontszám között közepes erősségű a kapcsolat. A szakterülethez való tartozás 44,93%-ban befolyásolja az elért pontszámok szóródását.

8.3. Korreláció

Mennyiségi ismérvek közötti sztochasztikus kapcsolat számszerű jellemzésével és elemzésével a korrelációs számítás foglalkozik.

Feladata:

- Megállapítani, hogy a két ismerv között valóban fennáll-e a kapcsolat, és ha van kapcsolat, akkor az milyen jellegű.
- A két ismerv között lévő kapcsolat szorosságának a megállapítása.

Két mennyiségi ismerv esetén (x;y), a változók közötti kapcsolat erősségét a korrelációs együttható (r) fejezi ki. Értéke -1 és +1 között lehet. Ha r negatív, akkor a két változó ellentétes irányban változik.

0,9 felett igen szoros

0,7-0,9 szoros

0,6-0,7 közepes

0,5-0,6 gyenge

0,4 alatt laza

8.3.1. Előjel korreláció

Az összetartozó értékpárok átlagtól való eltérései előjele alapján méri a kapcsolat szorosságát.

$$r_e = \frac{p - q}{p + q}$$

- ahol p jelenti azoknak az értékpároknak a számát, amelyeknek az átlagtól való eltéréseik előjele megegyezik.

- ahol q jelenti azoknak az értékpároknak a számát, amelyeknek az átlagtól való eltéréseik előjele eltérő.

Előnye, hogy kevés számítással, gyorsan tájékozódhatunk két mennyiségi ismerv közötti kapcsolat meglétéről és annak mértékéről.

Hátránya azonban, hogy csak „durva” mérőszám, ezért pontosabb elemzéshez nem használható.

Bemutató feladat

Vizsgáljuk meg, hogy van-e összefüggés az egy főre eső GDP (euro) és az 1000 lakosra jutó személygépkocsik számának alakulása között néhány európai ország adatai alapján.

S.szám.	1 főre jutó GDP Az Eu országaiban	Átlagtól való eltérés előjele	1000 lakosra jutó személygk. száma	Átlagtól való eltérés előjele
1.	22380	+	412	+
2.	20880	+	403	+
3.	25930	+	310	-
4.	7290	-	170	-
5.	2970	-	200	-
6.	23030	+	485	+
7.	20460	+	521	+
8.	7450	-	205	-
9.	13970	-	335	-
10.	36080	+	449	+
átlag	18044		349	

$p=9$ $q=1$, tehát szoros a kapcsolat

8.3.2. Rangkorreláció

Ha lehetőségünk van minőségi ismérvek kvantitatív jellemzésére oly módon, hogy az ismervváltozatokat rangsoroljuk, akkor a rangsor már alkalmas a két minőségi ismerv sztochasztikus kapcsolatának elemzésére.

Ezt a korrelációs együtthatót rangsorkorrelációs vagy Spearman-féle korrelációs együtthatónak nevezzük.

$$r_r = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)} \quad D_i^2 = (x_i - y_i)^2$$

ahol $D_i = x_i - y_i$

n : pedig az elemek száma

Bemutató feladat

Arra keressük a választ, hogy a tehenek fejéskori és etetéskori viselkedése között (nyugtalanság) van-e valamilyen összefüggés.

Sorszám	Fejés alatti nyugtalanság	Etetéskori nyugtalanság	D	D^2
---------	---------------------------	-------------------------	-----	-------

1	5	4	1	1
2	6	8	-2	4
3	10	7	3	9
4	9	9	0	0
5	1	2	-1	1
6	2	1	1	1
7	3	5	-2	4
8	7	6	1	1
9	8	10	-2	4
10	4	3	1	1
Össz.				26

$$r_r = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$$

$r_r = 1 - \frac{6 * 26}{10(10^2 - 1)} = 0,84$, a tehenek fejéskori és etetéskori viselkedése között szoros az összefüggés.

III. modul: Indexszámítás

9. lecke: Aggregát típusú indexszámítás

Követelmények:

Ön akkor sajátította el megfelelően a tananyagot, ha

- A meghatározások közül ki tudja választani az ár-, érték- és volumenindex helyes jelentését.
- Felsorolásból ki tudja választani, mikor használjuk a bázisidőszaki súlyozást.
- Konkrét adatok alapján ki tudja számolni az egyes indexeket és különbségeiket.
- A kiszámolt adatok alapján ki tudja egészíteni az adatok jelentésére vonatkozó mondatokat.

Az index-számítás olyan összehasonlító viszonyzámnak, melyek közvetlenül nem összesíthető mennyiségek, együttes, átlagos változását fejezi ki. A termelés, a szolgáltatás, a kereskedelem, a fogyasztás, a felhasználás, mint társadalom-gazdasági kategóriák elemzése, különböző mértékegységben rendelkezésre álló, különbözőminőséget képviselő sokaságok együttes vizsgálatát igényli.

A dinamikus viszonyszámokkal mindig csak 1-1 termék árának illetve mennyiségének alakulását vizsgáltuk. A gazdasági életben azonban szükség van olyan számításokra, amelyek egyszerre fejezik ki több termék ár- és mennyiségváltozásának hatását, pl. az árbevételre. Erre szolgálnak az indexszámítások. A változást kifejezhetjük mind relatív (I = index), mind abszolút (K =különbség) módon.

Az összehasonlítás elvégezhető:

- térben: 2 ország vagy két vállalat adatait hasonlítom,
- időben: 2 év adatait hasonlítom össze.

Az értékben történő összehasonlítást aggregálásnak, az összesített értékadatokat aggregátumoknak nevezzük.

Jelölések:

- q_i = mennyiség (volumen)
- p_i = ár
- v_i : érték $v_i = q_i * p_i$
- i = egyedi indexek
- k = egyedi eltérések

Napjainkban nem egyféle árut termelnek és forgalmaznak a vállalkozások, ezért érthető, hogy összevont mutatókkal vizsgálják tevékenységük időbeli alakulását, illetve hasonlítják azt más vállalkozások, területek adataihoz.

Természetesen az egyedi indexeket is használják, mert az egyes termékek ár, mennyiség és értékváltozását ezek fejezik ki. A termékekre számított egyedi indexek dinamikus viszonyszámok.

A termékek mennyiségi és árváltozását együtt vizsgálva legegyszerűbben az érték összehasonlításával oldható meg. A különböző termékekből származó bevételek, termelési értékek azonos mértékegységűek, így összesíthetők, és összehasonlíthatók mind hányados, mind különbség formájában is. A leggyakrabban az időbeli összehasonlítást végezzük.

9.1. Értékindeks, és értékindeks differencia.

A termelés, a forgalom, a fogyasztás értékének együttes, átlagos változását mutatja, vagyis két olyan aggregátum hányadosa, melyek a mennyiségi és az ár adatokban is eltérnek egymástól.

$$I_v = \frac{\sum q_1 * p_1}{\sum q_0 * p_0}$$

$$K_v = \sum q_1 * p_1 - \sum q_0 * p_0$$

Azaz az értékindeks megmutatja, hogy a mennyiség és az ár együttes változása esetén, hogyan változott az érték az összes terméket figyelembe véve (hányszorosára). Az érték különbség (értékindeks differencia) pedig azt, hogy mennyivel változott az érték.

Bemutató feladat

Egy piaci árus forgalma két időpontban

Termékek	2018. december		2019. december	
	Eladott mennyiség	egységár	Eladott mennyiség	egységár
A-termék (db)	1200	20	1250	25
B-termék (kg)	250	160	280	210
C-termék(kg)	700	150	500	280

1. Hogyan változott az árbevétel?

Értékindeks és értékindeks differencia

$$I_v = \frac{\sum q_1 * p_1}{\sum q_0 * p_0} = \frac{31250 + 58800 + 140000}{24000 + 40000 + 105000} = \frac{230050}{169000} = 1,3612 = 136,12\%$$

$$K_v = \sum q_1 * p_1 - \sum q_0 * p_0 = 230050 - 169000 = 61050 Ft$$

A termelési érték változását természetesen termék fajtánként, dinamikus viszonszámként, is meghatározhatjuk, amelyeket egyedi értékindeksnek nevezünk.

Az egyedi termékek értékének eltérését mutatja, azaz hogyan változott az adott termékre vonatkozó termelési érték, forgalom a bázisidőszakról a tárgyidőszakra.

$$i_v = \frac{v_{i1}}{v_{i0}} = \frac{q_{i1} * p_{i1}}{q_{i0} * p_{i0}} = i_q * i_p$$

$$k_v = v_{i1} - v_{i0}$$

A vizsgált termékek értékének változása részben az árak változásának, részben a mennyiségek változásának köszönhető.

9.2. Árindex és árindex differencia

Az árváltozás hatásának vizsgálatakor a mennyiséget állandónak tételezzük fel. Különböző statisztikusok eltérő súlyozást használtak, így a következő módon számolhatunk.

1. Tárgyévi súlyozás: Paashe-féle árindex:

$$I_p^1 = I_p^P = \frac{\sum q_1 * p_1}{\sum q_1 * p_0}$$

2. Bázisévi súlyozás: Laspeyres-féle árindex

$$I_p^0 = I_p^L = \frac{\sum q_0 * p_1}{\sum q_0 * p_0}$$

3. A két árindex mértani átlaga: Fisher-féle árindex:

$$I_p^F = \sqrt{I_p^1 * I_p^0}$$

4. Eltérések:

$$K_p^1 = \sum q_1 p_1 - \sum q_1 p_0$$

$$K_p^0 = \sum q_0 p_1 - \sum q_0 p_0$$

Az árindex kifejezi, hogy hányszorosára változott az érték, csak az árváltozás hatására, az árindex differencia pedig azt, hogy mennyivel változott az érték.

A képletekben szereplő $q_0 p_1$ és $q_1 p_0$ szorzatok összegzéseként kapott értékadatokat fiktív aggregátumoknak nevezzük, mivel ezek a valóságban nem léteznek.

Bemutató feladat

Folytassuk az előző feladatot!

Árindex és árindex differencia:

$$I_p^1 = I_p^P = \frac{\sum q_1 * p_1}{\sum q_1 * p_0} = \frac{230050}{1250 * 20 + 280 * 160 + 500 * 150} = \frac{230050}{144800} = 1,5887$$

$$I_p^0 = I_p^L = \frac{\sum q_0 * p_1}{\sum q_0 * p_0} = \frac{1200 * 25 + 250 * 210 + 700 * 280}{169000} = \frac{278500}{169000} = 1,6479$$

$$I_p^F = \sqrt{I_p^1 * I_p^0} = \sqrt{1,5887 * 1,6479} = 1,6180 = 161,80\%$$

$$K_p^1 = \sum q_1 p_1 - \sum q_1 p_0 = 230050 - 144800 = 85250 Ft$$

$$K_p^0 = \sum q_0 p_1 - \sum q_0 p_0 = 278500 - 169000 = 109500 Ft$$

Természetesen az árindexet illetve az árváltozásból adódó különbséget is meghatározhatjuk termékenként, ebben az esetben egyedi árindexről illetve egyedi árindex különbségről beszélünk.

Adott termék árváltozását fejezi ki.

$$i_p = \frac{p_{i1}}{p_{i0}}$$

$$k_p = p_{i1} - p_{i0}$$

9.3. Volumenindex és volumenindex differencia

Ebben az esetben az árat tekintjük állandónak, így itt is kétféle súlyozás lehetséges.

1. Tárgyévi súlyozás: Paashe-féle mennyiségindex:

$$I_q^1 = I_q^P = \frac{\sum q_1 * p_1}{\sum q_0 * p_1}$$

2. Bázisévi súlyozás: Laspeyres-féle mennyiségindex

$$I_q^0 = I_q^L = \frac{\sum q_1 * p_0}{\sum q_0 * p_0}$$

3. A két árindex mértani átlaga: Fisher-féle mennyiségindex:

$$I_q^F = \sqrt{I_q^1 * I_q^0}$$

4. Eltérések:

$$K_q^1 = \sum q_1 p_1 - \sum q_0 p_1$$

$$K_q^0 = \sum q_1 p_0 - \sum q_0 p_0$$

A volumenindex kifejezi, hogy hányszorosára változik az érték, csak a mennyiségi változás hatására, a volumenindex differencia pedig azt, hogy mennyivel változik az érték. Az egyedi volumindex és egyedi volumenindex különbség is meghatározható termékenként:

$$i_q = \frac{q_{i1}}{q_{i0}}$$

$$k_q = q_{i1} - q_{i0}$$

Ha érték formájában jelenítjük meg mind az árak, mind a mennyiségek változásának hatását, akkor az alábbi összefüggések írhatók fel:

- az ár hatása:

$$q_{i1} p_{i1} - q_{i1} p_{i0}$$

$$q_{i0} p_{i1} - q_{i0} p_{i0}$$

- a mennyiség hatása:

$$q_{i1} p_{i1} - q_{i0} p_{i1}$$

$$q_{i1} p_{i0} - q_{i0} p_{i0}$$

A mennyiségi változása termékenként:

az i_q : a termékek mennyiségének relatív változása a k_q a termékek eladott mennyiségének abszolút változása

Bemutató feladat

Folytassuk az előző feladatot!

Volumenindex és volumen differencia

$$I_q^1 = I_q^P = \frac{\sum q_1 * p_1}{\sum q_0 * p_1} = \frac{230050}{267500} = 0,8260$$

$$I_q^0 = I_q^L = \frac{\sum q_1 * p_0}{\sum q_0 * p_0} = \frac{144800}{169000} = 0,8568$$

$$I_q^F = \sqrt{I_q^1 * I_q^0} = \sqrt{0,8260 * 0,8568} = 0,8413 = 84,13\%$$

$$K_q^1 = \sum q_1 p_1 - \sum q_0 p_1 = 230050 - 267500 = -48450 Ft$$

$$K_q^0 = \sum q_1 p_0 - \sum q_0 p_0 = 144800 - 169000 = -24200 Ft$$

Bemutató feladat

Az eddig kiszámolt mutatók szöveges elemzése:

Egy piaci árus árbevételének alakulását vizsgáltuk 2018. és 2019. decemberi adatainak összehasonlításával. Az árak mindhárom termék esetében nőttek, a legnagyobb mértékben a C-termék egységára emelkedett 86,67%-kal, a legkisebb mértékben az A-termék eladási ára, 25%-kal. A B-termék egységára 31255-kal nőtt. Az A-termékből az eladott mennyiség csak 4,15%-kal volt több 2019. decemberében, a B-termékből azonban 12%-al adtak el többet. A C-termék eladott mennyisége csökkent 2018-hoz képest 28,57%-kal. A fentiek hatására az egyes termékekből származó árbevétel növekedése a B-termék esetében a legnagyobb, 47%. A C-termék eladásából származó árbevétel 2019-ben 32,33%-kal volt több, mint 2018-ban, az A-termék pedig 30,21%-kal.

Az egységárak és az eladott mennyiségek változása miatt a kereskedő árbevétele 2019-ben 36,12%-kal volt több mint 2018-ban, ez 61.050 Ft-nak felel meg. Az átlagos árváltozás 61,80% volt, míg az eladott mennyiségek átlagosan 15,78%-kal csökkentek.

9.4. Indexek közötti összefüggések

Mivel az ár- és a volumenindex csak az egyik tényező változását fejezi ki, ezért szoroztuk az értékindexet adja, mivel az az ár és a mennyiség együttes változásának a hatását fejezi ki.

$$I_v = I_p^0 * I_q^1 = I_p^1 * I_q^0 = I_p^F * I_q^F$$

$$K_v = K_p^1 + K_q^0 = K_p^0 + K_q^1$$

9.5. Az indexek átlagformái

Az indexek nemcsak aggregát formában számíthatók, hanem az egyedi indexek átlagával is. Ilyenkor a súly szerepét az indexben szereplő valamelyik – éppen rendelkezésre álló –

aggregátum, illetve annak mennyiségi viszonyszámai töltik be. A súlyozás módja pedig attól függ, hogy az illető aggregátum az index számlálójában vagy nevezőjében foglal helyet. Ha csak a bázisidőszaki vagy csak a tárgyidőszaki értékösszeg áll rendelkezésünkre, akkor az indexek átlagformában számíthatóak.

9.5.1. Az értékindex átlagformái

Abban az esetben, ha csak a bázisidőszaki ár és mennyiségi adatok ismertek, illetve az egyes termékcsoportok egyedi értékindexe, akkor az értékindexet számtani átlag formában tudjuk kiszámolni. A súly ebben az esetben a bázisidőszak aggregátuma.

$$I_v = \frac{\sum q_0 p_0 * i_v}{\sum q_0 * p_0}$$

Ha a tárgyidőszaki ár és mennyiségi adatok ismertek, illetve az egyes termékcsoportok egyedi értékindexe, akkor az értékindexet harmonikus átlag formában tudjuk kiszámolni. A súly ebben az esetben a tárgyidőszak aggregátuma.

$$I_v = \frac{\sum q_1 * p_1}{\sum \frac{q_1 * p_1}{i_v}}$$

9.5.2. Az árindex átlagformái

Hasonlóan az értékindexhez, ebben az esetben is kétféle számítás lehetséges:

- Számtani átlag:

$$I_p^0 = \frac{\sum q_0 p_0 * i_p}{\sum q_0 * p_0} \quad I_p^1 = \frac{\sum q_1 p_0 * i_p}{\sum q_1 * p_0}$$

- harmonikus átlag formában:

$$I_p^0 = \frac{\sum q_0 * p_1}{\sum \frac{q_0 * p_1}{i_p}} \quad I_p^1 = \frac{\sum q_1 * p_1}{\sum \frac{q_1 * p_1}{i_p}}$$

9.5.3. A volumenindex átlagformái

A volumenindex is meghatározható mindkét átlagolással.

- Számtani átlag:

$$I_q^0 = \frac{\sum q_0 p_0 * i_q}{\sum q_0 * p_0} \quad I_q^1 = \frac{\sum q_0 p_1 * i_q}{\sum q_0 * p_1}$$

- harmonikus átlag formában:

$$I_q^0 = \frac{\sum q_1 * p_0}{\sum \frac{q_1 * p_0}{i_q}} \quad I_q^1 = \frac{\sum q_1 * p_1}{\sum \frac{q_1 * p_1}{i_q}}$$

Bemutató feladat:

Egy vállalkozás kereskedelmi tevékenységére vonatkozó adatok:

Árucsoport	Árbevétel 2017-ben (eFt)	Értékesített mennyiség	Árbevétel
		2018-ban a 2017. év %-ában	
A	4.000	115,00	145,00
B	9.000	110,00	125,00
C	3.000	125,00	140,00
D	12.000	98,00	120,00

Feladat:

- Számítsa ki a négy árucsoportra vonatkozóan az érték-, ár- és volumenindexeket!

A 2017. évi árbevétel a tárgyévi adat, azaz ismert minden árucsoportra a $q_0 p_0$.

Az értékesített mennyiségi változás az i_q , az árbevétel változása az i_v .

$$I_v = \frac{\sum q_0 p_0 * i_v}{\sum q_0 * p_0} = \frac{4000 * 1,45 + 9000 * 1,25 + 3000 * 1,40 + 12000 * 1,20}{4000 + 9000 + 3000 + 12000} = 1,2732 = 127,32\%$$

$$I_q^0 = \frac{\sum q_0 p_0 * i_q}{\sum q_0 * p_0} = \frac{4000 * 1,15 + 9000 * 1,10 + 3000 * 1,25 + 12000 * 0,98}{4000 + 9000 + 3000 + 12000} = 1,0718 = 107,18\%$$

Az egyedi árindexek kiszámíthatók:

$$i_{pA} = 1,45 / 1,15 = 1,26$$

$$i_{pB} = 1,25 / 1,1 = 1,14$$

$$i_{pC} = 1,4 / 1,25 = 1,12$$

$$i_{pD} = 1,2 / 0,98 = 1,02$$

$$I_p^0 = \frac{\sum q_0 p_0 * i_p}{\sum q_0 * p_0} = \frac{4000 * 1,26 + 9000 * 1,14 + 3000 * 1,12 + 12000 * 1,02}{4000 + 9000 + 3000 + 12000} = 1,1036 = 110,36\%$$

$$I_q^1 = \frac{1,2732}{1,1036} = 1,1537$$

$$I_p^1 = \frac{1,2732}{1,0718} = 1,1879$$

Ezután a Fisher-féle ár-és volumenindex már könnyen meghatározható.

9.6. Az aggregát indexek használata területi összehasonlítás esetén

Az aggregát típusú indexeket területi összehasonlításra is használjuk. Azonban csak abban az esetben használhatók, ha:

- Azonos időszakra vagy időpontra vonatkozó adatokat hasonlítunk össze,
- Az összehasonlítás alapja (bázisa) a vizsgálttól függ, vagy az elemző dönti el.
- A számszerű eredmények megfogalmazásakor nem használhatók a növekedés vagy csökkenés kifejezések. Helyettük a nagyobb, kisebb, magasabb, eltér szavak használatosak.

A területi összehasonlítás speciális esete két ország (eltérő valutájú) adatainak összevetése, elemzése.

Az értékindexnek nincs jelentése, hiszen különböző pénznem szerepel a számlálóban és a nevezőben. Az ár- és volumenindexnél csak a Fisher-féle képleteket értelmezzük, mivel nagy lehet az eltérés mind az árakban, mind a mennyiségekben. Az árindex jelentése és kifejezési formája megváltozik. Árszínvonalat hasonlítunk összes, így nem lehet százalékos formában felírni, hanem a vizsgált valuták hányadosaként jelenik meg. A valutákat a nemzetközileg megszokott jelekkel használjuk, pl.: a forintot HUF- jelöléssel. A mennyiségi index megőrzi eredeti jelentését, a két ország lakóinak a vizsgált termékekből való fogyasztásának hányadosaként.

Bemutató feladat

Két ország élelmiszerfogyasztását az alábbi adatok jellemzik

Termék	D ország		G ország	
	Egy főre jutó fogyasztás	Eladási ár (delta pénznemben)	Egy főre jutó fogyasztás	Eladási ár (gamma pénznemben)
1. termék	20	3	40	8
2. termék	30	4	10	9

Hasonlítsa össze a két ország 1 főre jutó fogyasztásának mennyiségét és a valuták vásárlóerejét. A viszonyítási alap a D ország legyen.

$$I_q^G(G/D) = \frac{40 \cdot 8 + 10 \cdot 9}{20 \cdot 8 + 30 \cdot 9} = \frac{410}{430} = 0,9535$$

$$I_q^D(G/D) = \frac{40 \cdot 3 + 10 \cdot 4}{20 \cdot 3 + 30 \cdot 4} = \frac{160}{180} = 0,8889$$

$$I_q^F(G/D) = \sqrt{0,9535 \cdot 0,8889} = 0,9206 = 92,60$$

A termékek körét vizsgálva az egy főre jutó fogyasztás G-országban 7,4%-kal kevesebb volt, mint D-országban.

$$I_p^G(G/D) = \frac{40 \cdot 8 + 10 \cdot 9}{40 \cdot 3 + 10 \cdot 4} = \frac{410}{160} = 2,5625$$

$$I_p^D(G/D) = \frac{20 \cdot 8 + 30 \cdot 9}{20 \cdot 3 + 30 \cdot 4} = \frac{430}{180} = 2,3889$$

$$I_p^F(G/D) = \sqrt{2,5625 \cdot 2,3889} = 2,4742 G/D$$

D ország egy valutájának (delta) vásárlóereje egyenlő 2,4742 gamma valuta vásárlóerejével (G-ország valutájával).

9.7. Indexsorok

Gyakran nem kettő, hanem több időszak (időpont) adatait kell összehasonlítani. Ekkor minden időszakhoz (időponthoz) ki kell számítani az indexeket. Az így kapott értékek indexsort alkotnak.

Az összehasonlítástól függően a dinamikus viszonzszámokhoz hasonlóan lehet bázis- és lánc-indexsor. A súlyozás módja szerint az ár- és volumen-indexsor lehet

- Állandó súlyozású indexsor
- Változó súlyozású indexsor.

10. lecke: Standardizációs indexszámítás

Követelmények:

Ön akkor sajátította el megfelelően a tananyagot, ha

- Adott adatok alapján felsorolásból ki tudja választani, mit fejez ki a főátlag-index, a részátlag-index és az összetétel-index.
- Meg tudja határozni a főátlagokat,
- Adott adatok alapján ki tudja számolni a főátlag-indexet, a részátlag indexet és az összetételindexet
- A kiszámított statisztikai mutatókkal ki tudja egészíteni a mutatók értelmezésére vonatkozó mondatokat.

Ha valamilyen jelenség színvonalát (termékek önköltségét, átlagbéréket) akarjuk jellemezni, akkor heterogén sokaság esetén nem használhatjuk a már megismert (I: Modul) intenzitási viszonszámokat. A sokaságot a heterogenitást előidéző ismerv alapján homogén részsokaságokra kell bontani, és így már lehetséges az intenzitási viszonszámok számítása. A sokaság egészére kiszámított intenzitási viszonszámokat összetett intenzitási viszonszámoknak nevezzük. Az egyes részsokaságokra kiszámítottakat pedig intenzitási részviszonszámoknak.

Az átlagos színvonalat kifejező mutatókat két tényező befolyásolja:

- Milyen az egyes csoportokban a vizsgált színvonal nagysága,
- Milyen a sokaság összetétele, szerkezete.

A két tényező hatását külön kell választani és ezt a módszert standardizálásnak nevezzük.

Ha intenzitási viszonszámokat (főátlagok) hasonlítunk össze, akkor standardizációs indexszámítást alkalmazunk. Az összehasonlítás történhet úgy, hogy a relatív nagyságot (I), vagy az abszolút nagyságot (K) tudjuk megállapítani, és lehetőségünk van az eltérést alakító tényezők hatásának kimutatására is. Az összehasonlítás történhet térben és időben. A számítás során az alábbi jelöléseket használjuk:

- $\bar{V}_i$ = főátlag: összetett intenzitási viszonszám
- B_i = összetételre vonatkozó adat (pl. létszám). Ez az egyedi viszonszámok nevezőjében található adat. Az összetett intenzitási viszonszámok számításánál (a számtani átlag formájában) pedig a részviszonszámok súlyaként szerepel (f_i) (számtani átlag formában).
- $A_i = B \cdot v$: (pl. létszám * 1 főre jutó termelés = össztermelés). Az egyedi viszonszámok számlálójában szereplő adat. Az összetett intenzitási viszonszámok számításánál (a harmonikus átlag formájában) pedig a részviszonszámok súlyaként szerepel.
- V_i : részátlag, részviszonszám = A_i/B_i .
- i_i : egyedi indexek
- k_i : egyedi eltérések

Ha valamilyen ismerv szempontjából heterogén sokaságból átlagot számolunk, elemzésünk csak akkor lesz teljes, ha a sokaság összetételét nem hagyjuk figyelmen kívül. A sokaságok főátlagának térbeli vagy időbeli összehasonlításakor (azaz a sokaság átlagának időbeli változását, vagy két sokaság átlagát) figyelembe kell venni a részsokaságok átlagának és összetételének a változását.

10.1. Főátlag-index

A főátlag-index a főátlag változását fejezi ki, azaz megmutatja, hogy hogyan változik a heterogén sokaság főátlaga a részátlagok színvonalának és arányának együttes változása esetén.

$$I = \frac{\bar{V}_1}{\bar{V}_0}, \text{ általában százalékban fejezzük ki.}$$

$$\bar{V}_i = \frac{\sum A_{ij}}{\sum B_{ij}} = \frac{\sum B_{ij} * V_{ij}}{\sum B_{ij}} = \frac{\sum f_{ij} * x_{ij}}{\sum f_{ij}}, \text{ egyszerűbben kifejezve: } \bar{V}_i = \frac{\sum A_i}{\sum B_i} = \frac{\sum B_i * V_i}{\sum B_i} = \frac{\sum f_i * x_i}{\sum f_i}$$

Ez a főátlagok relatív változását mutatja meg. A változást abszolút mértékben is kifejezhetjük különbségfelbontással. Ez a mennyiségi változást mutatja meg.

$$K = \bar{V}_1 - \bar{V}_0 \text{ vagy } K = \frac{\sum B_1 * V_1}{\sum B_1} - \frac{\sum B_0 * V_0}{\sum B_0}$$

A főátlag különbség (vagy differenciál) megmutatja, hogy mennyivel változott a főátlag a részátlagok színvonalának és arányának együttes változása esetén.

Bemutató feladat

A szellemi foglalkozású munkavállalók létszám és béradatai egy vállalkozásnál

Nem	2018.		2019.	
	Létszám (fő)	Átlagbér (Ft/fő/hó)	Létszám (fő)	Átlagbér (Ft/fő/hó)
Férfiak	50	200.000	60	220.000
Nők	80	190.000	70	205.000
Összesen.	130	-	130	-

1. Számítsa ki, hogyan változott a vállalati átlagbér.
2. Mutassa ki a vállalati átlagbér változását befolyásoló tényezők hatását.
3. Írjon szöveges elemzést.

Az előző lecke során már kiszámítottuk külön a nők és a férfiak átlagbérének a változását. Férfiak átlagbérének változása:

$$i_1 = \frac{v_{11}}{v_{10}} = \frac{220000}{200000} = 1,1 = 110\%$$

$$k_1 = v_{11} - v_{10} = 220000 - 200000 = 20000$$

A nők átlagkeresetének változása:

$$i_2 = \frac{v_{21}}{v_{20}} = \frac{205000}{190000} = 1,0789 = 107,89\%$$

$$k_2 = v_{21} - v_{20} = 205000 - 190000 = 15000$$

Most megvizsgáljuk az átlagbér alakulását az egész vállalkozásra.

$$2018. \text{ év: } \bar{V}_0 = \frac{\Sigma B_0 * V_0}{\Sigma B_0} = \frac{50 * 200000 + 80 * 190000}{50 + 80} = 193846 \text{ Ft/fő/hó}$$

$$2019. \text{ év } \bar{V}_1 = \frac{\Sigma B_1 * V_1}{\Sigma B_1} = \frac{60 * 220000 + 70 * 205000}{60 + 70} = 211923 \text{ Ft/fő/hó}$$

1. A vállalati átlagbér változása:

2.

$$I = \frac{\bar{V}_1}{\bar{V}_0} = \frac{211923}{193846} = 1,0933 = 109,33\%$$

$$K = \bar{V}_1 - \bar{V}_0 = 211923 - 193846 = 18077 \text{ Ft / Fő / hó}$$

10.2. Részátlag-index

A részátlagok színvonalváltozásának hatását fejezi ki a főátlag változásában úgy, hogy állandónak tekinti az összetételt. Ez kétféleképpen történhet:

- Állandónak tekintjük a bázisidőszak adatait (amihez hasonlítunk). Ilyenkor standard átlagot számolunk:

$$\bar{V}_{s:B_0} = \frac{\Sigma B_0 * v_1}{\Sigma B_0}$$

Ebben az esetben a részátlag-index:

$$I' = \frac{\bar{V}_{s:B_0}}{\bar{V}_0}$$

- Állandónak tekintjük a tárgyidőszak adatait (amit hasonlítunk).

$$\bar{V}_{s:B_1} = \frac{\Sigma B_1 * V_0}{\Sigma B_1}$$

Ebben az esetben a részátlag-index:

$$\mathbf{I}' = \frac{\bar{\mathbf{V}}_1}{\bar{\mathbf{V}}_s}$$

A részátlag-indexet is százalékos formában adjuk meg. Ebben az esetben is kiszámoljuk az abszolút változást is.

$$\mathbf{K}' = \bar{\mathbf{V}}_1 - \bar{\mathbf{V}}_{s:B_1} \text{ vagy } \mathbf{K}' = \bar{\mathbf{V}}_{s:B_0} - \bar{\mathbf{V}}_0$$

A részátlag-index különbség megmutatja, hogy a részviszonyszámok közötti eltérések önmagukban mekkora eltérést indokolnak a két főátlag között.

Bemutató feladat:

Folytassuk az előző feladatot!

A standard főátlag kiszámítása: állandó a bázis év létszáma

$$\bar{V}_s = \frac{\Sigma B_0 * V_1}{\Sigma B_0} = \frac{50 * 220000 + 80 * 205000}{50 + 80} = 210769 \text{ Ft/fő/hó}$$

A részviszonyszámok változásának hatása:

$$I' = \frac{\bar{V}_s}{\bar{V}_0} = \frac{210769}{193846} = 1,0873 = 108,73\%$$

$$K' = \bar{V}_s - \bar{V}_0 = 210769 - 193846 = 16923 \text{ Ft / fő / hó}$$

10.3. Összetételindex

Az összetételindex a fősokaság összetételében bekövetkezett változásoknak a főátlagra gyakorolt hatását fejezi ki. Azaz az összetételindex megmutatja, hogy mennyire változott a főátlag csak az összetétel-változás hatására. Hasonlóan a részátlag-indexhez, ebben az esetben is kétféle számítás lehetséges.

- Tárgyidőszaki súlyozás:

$$\bar{V}_{s:V_1} = \frac{\Sigma B_0 * V_1}{\Sigma B_0}$$

Az összetételindex pedig:

$$I'' = \frac{\bar{V}_1}{\bar{V}_{s:V_1}}$$

- Bázis időszaki súlyozással:

$$\bar{V}_{s:V_0} = \frac{\Sigma B_1 * V_0}{\Sigma B_1}$$

Az összetételindex pedig:

$$I'' = \frac{\bar{V}_{s:V_0}}{\bar{V}_0}$$

Hasonlóan a részátlag indexhez az összetételhatást is ki lehet fejezni a különbségek felbontásával, azaz kiszámolhatjuk, hogy csak az összetétel-változás hatására mennyivel változott a főátlag.

$$K'' = \bar{V}_1 - \bar{V}_0 \text{ vagy } K'' = \bar{V}_1 - \bar{V}_s$$

Bemutató feladat:

Folytassuk az előző feladatot!

Az összetétel változásának hatása

$$I'' = \frac{\bar{V}_1}{\bar{V}_s} = \frac{211923}{210769} = 1,0055 = 100,55\%$$

$$K'' = \bar{V}_1 - \bar{V}_s = 211923 - 210769 = 1154 \text{ Ft / fő / hó}$$

10.4. Az indexek és a különbségek közötti összefüggések

Mivel a részátlag- és az összetétel-index a főátlagban bekövetkező változásokat fejezi ki külön-külön, ezért főátlag-index (amely az együttes változást fejezi ki), ezek szorzata.

$$I = I' * I'' = \frac{\overline{V_1}}{\overline{V_s}} * \frac{\overline{V_s}}{\overline{V_0}} = \frac{\overline{V_s}}{\overline{V_0}} * \frac{\overline{V_1}}{\overline{V_s}}$$

Hasonló összefüggés van a különbségek esetében is:

$$K = K' + K''$$

Természetesen ezek az összefüggések csak akkor igazak, ha „keresztbe standardizálunk”, azaz ha a részátlag-indexnél tárgyidőszaki súlyozást használunk (B_1), akkor az összetétel-index esetében bázisidőszaki (V_0) súlyozást kell használni, illetve fordítva.

Bemutató feladat:

Folytassuk az előző feladatot!

Összefüggések

$$I = I' * I'' = 1,0873 * 1,0055 = 1,0933 = 109,33\%$$

$$K = K' + K'' = 16923 + 1154 = 18077 \text{ Ft/fő/hó}$$

10.5. Az indexek és a különbségek szöveges értelmezése

A kapott számszerű eredmények értelmezésénél fontos a megfogalmazás. Időbeli összehasonlítás esetén növekedésről vagy csökkenésről beszélünk. Térbeli összehasonlításkor az eltérést a kisebb, kevesebb, nagyobb, több jelzőkkel értelmezzük.

Az indexeket százalékban fejezzük ki. A 100% feletti rész jelenti a növekedés mértékét, a 100% alatti rész pedig a csökkenést fejezi ki úgy, hogy a kapott értéket levonjuk 100-ból.

Bemutató feladat

Szöveges értékelés

A vizsgált vállalkozásnál a vállalati átlagbér a nők és férfiak átlagkeresetének és létszámának változása miatt 2018-ról 2019-re 9,33%-kal emelkedett, ami 18077 Ft-os átlagbér-növekedést jelent. A férfiak és nők átlagbérének növekedése 8,73%-os vállalati átlagbér növekedést eredményezett, azaz 16923 Ft-tal nőtt átlagosan a dolgozók bére. Csak a létszám összetételében bekövetkezett változás miatt a vállalati átlagbérek 0,55%-kal emelkedtek, ami 1154 Ft-nak felel meg.

IV. modul: Mintavételi eljárások, becslés, hipotézisvizsgálat

11. lecke: Mintavételi eljárások

Követelmények:

Ön akkor sajátította el megfelelően a tananyagot, ha

- Ki tudja választani a részleges adatgyűjtés jellemzőit,
- Ki tudja választani a reprezentatív adatgyűjtés jellemzőit,
- A felsorolásból ki tudja választani a véletlen mintavételi eljárásokat,
- Ki tudja választani a rétegezett mintavétel típusait és azok jellemzőit.

A statisztikai elemzés során, ha a megfigyelés és az adatgyűjtés a sokaság minden egyes tagjára vonatkozik, akkor teljes körű megfigyelésről beszélünk.

A társadalmi-gazdasági statisztikában általában nem tudunk minden adatot összegyűjteni, illetve elemezni, ezért általában csak részleges adatgyűjtést végzünk, amelynek egyik módja a reprezentatív mintavétel, pl. a lakosság jövedelmi helyzetének, fogyasztási szokásainak elemzéséhez gyűjtött adatok.

A reprezentatív megfigyelés (mintavétel) célja, hogy valamely sokaság egy részének megfigyelése révén következtetéseket tudjunk levonni a sokaság egészére, annak jellemzőire, összetételére vonatkozóan. A reprezentatív felvételezés során a teljes sokaságból szigorú szabályok szerint kiválasztunk egy részsokaságot, és csak ennek elemeire vonatkozóan gyűjtünk adatokat.

Azt a sokaságot, amelyre a mintavétel segítségével következtetni szeretnénk, alapsokaságnak, az alapsokaság azon részét, amely alapján a következtetéseket levonjuk, mintasokaságnak nevezzük.

A statisztikai adatfelvételek és az annak eredményeit felhasználó elemzések mindig tartalmaznak hibát. A statisztikai hiba egy része a módszertan sajátosságaiból is adódik (tömörítés, közelítés, becslés), ez velejárója a statisztikai elemzéseknek. A statisztikus célja, hogy a hibát minimálisra csökkentse (mintavételi és nem mintavételi hibát együtt). A mintavételi hiba matematikai-statisztikai eszközökkel becsülhető. A nem mintavételi hiba korábbi tapasztalatok alapján becsülhető meg. A mintavétel tervezésénél a mintavételi hibával és annak vizsgálatával foglalkozunk.

Egy sokaság elemszáma lehet véges és végtelen. A véges (N) elemszámú sokaság megadásának legegyszerűbb módja egyetlen ismerv szerint az alábbi:

$Y_1, Y_2, \dots, Y_N$

Ha a sokaság végtelen számú, akkor nem adható meg ebben a formában. Ekkor két esetet különböztetünk meg:

- diszkrét ismerv esetén, ami azt jelenti, hogy az ismervértékek véges vagy megszámlálhatatlan végtelent alkotnak, akkor a valószínűségi eloszlás használható:

$$P(Y=k)=P_k$$

- folytonos ismerv esetén pedig a:

$$P(Y<y)=F(y)$$

eloszlásfüggvénnyel és ennek deriváltjával, az

$$F'(Y)=f(y)$$

sűrűségfüggvénnyel adható meg a sokaság.

11.1. Véletlen mintavételi eljárások

A véletlen mintavétel lényege, hogy a mintát alkotó elemek a kiválasztás során egyenlő valószínűséggel kerüljenek bele a mintába. Az ember a véletlen kiválasztás végrehajtására

nem megfelelő, ugyanis a szubjektív kiválasztás általában nem felel meg az egyenlő valószínűség elvének, főleg nagy sokaság és minta esetén. A véletlen mintavétel történhet.

- sorshúzással,
- Véletlen számok segítségével:
 - véletlen számok táblázat segítségével,
 - számítógépes véletlen szám generálással.

A véletlen szám segítségével történő mintavétel csak olyan esetekben használható, ha a sokaság egyenletes eloszlású.

Folytonos sokaság esetében a felezési módszert alkalmazzák, addig felezik a mintát, amíg vizsgálható méretű mintanagyságot kapnak.

11.1.1. Egyszerű véletlen mintavétel (EV)

Homogén, véges elemszámú sokaság esetén visszatevés nélkül választjuk ki a mintát, elemenként egyenlő valószínűséggel. Véletlen szám segítségével történő mintavételkor az ismételten előforduló sorszámot átugorjuk. A mintavétel során az N elemű sokaságból $\binom{N}{n}$ -

féle (N alatt az n) különböző összetételű mintát kapunk.

Ez a módszer főleg a természettudományi kísérleteknél, főleg a biológiai eredmények értékelésekor alkalmazható. Társadalmi-gazdasági jelenségek vizsgálatára nem használható.

11.1.2. Független, azonos eloszlású minta (FAE)

FAE mintát akkor kapunk, ha homogén és végtelen sokaságból veszünk mintát. Az egyes mintaelemek kiválasztása azonos valószínűséggel történik. Tipikus alkalmazási területe a tömegtermelés minőségi ellenőrzésének.

11.1.3. Rétegzett mintavétel (R)

A rétegzett mintavétel során a vizsgált ismerv szempontjából heterogén sokaságot több homogén (minél kisebb szórású) részsokaságra bontjuk úgy, hogy a csoportok kiadják a teljes sokaságot, továbbá egyetlen sokasági elem se tartozzon két vagy több csoportba. Az egyes rétegeken belül a minta elemének a kiválasztása egyszerű véletlen mintavétellel történik. Ez a módszer a társadalomtudományok területén nagyon gyakori.

- **Egyenletes elosztás:** Lényege, hogy minden rétegben azonos számú mintaelem kerül, azaz $n_j = n/M$. Egyszerű végrehajtani, de hátránya, hogy nem veszi figyelembe a teljes sokaságot adó részsokaságok nagyságát és szórását, így nagyfokú torzítást okozhat
- **Arányos elosztás:** Lényege, hogy a mintába a sokasági arányoknak megfelelően választjuk meg az arányszámot: $n_j = n \cdot N_j / N$. A mintában ugyanazok a súlyarányok szerepelnek, mint a sokaságban. Végrehajtása egyszerű.
- **Nem arányos elosztás:** a mintában a rétegarányok nem egyeznek meg az alapsokasági rétegarányokkal: $n_j \neq n \cdot N_j / N$
- **Neyman-féle optimális elosztás:** A nem arányos elosztás egyik fajtája. Alkalmazásának feltétele, hogy előre ismerjük vagy becsülni tudjuk az egyes rétegekbe lévő adatok szórását (σ_j). A nagyobb szórású rétegekből nagyobb számú mintát veszünk. Előnye, hogy minimális hibával számolható ki az ilyen mintából a főátlag, de nehéz végrehajtani, mivel a rétegenkénti szórás szükséges hozzá.

- **Költségoptimalis elosztás:** A rétegzett mintavétel egyik módja, és feltételezi, hogy ismerjük az egyes rétegek megfigyelési egységköltségeit is. Egy elem átlagos megfigyelési költsége Π_j forint. Azonos rétegnagyság és szórás esetén minél nagyobb a mintavétel költsége, annál kisebb mintát kell venni a rétegből

11.1.4. Csoportos mintavétel (Cs)

A homogén sokaságot csoportokra bontjuk, a mintát egyszerű véletlen mintavétellel kiválasztott csoport egyedei alkotják. A csoportok meghatározása lehet természetes, azaz eleve adott, de mesterségesen is történhet. Az N elemű sokaságot M részre bontjuk, ahol az egyes csoportok n_i eleműek, és $\sum n_i = N$.

Főleg közvélemény-kutatáskor alkalmazzák ezt a mintavételt.

11.1.5. Többlépcsős mintavétel (TL)

Homogén sokaság vizsgálata esetén alkalmazható. Először csoportos mintavétellel kiválasztjuk az elsődleges mintavételi egységeket. 1-1 homogén nagyobb csoportból egyszerű véletlen kiválasztással egyedeket jelölünk ki, amelyek kiscsoportokat alkotnak. Ha megfelelő a minta, akkor kétlépcsős mintavételről beszélünk. Ha nem, akkor a kiscsoportokkal tovább ismételjük az eljárást.

11.2. Nem véletlen mintavételi eljárások

A véletlen mintavétel esetén elkövetett hibák valószínűség-számítási ismeretek segítségével meghatározhatóak. A nem véletlen mintavétel esetén kapott minta és az eredeti sokaság között azonban nehéz a mintavétel során elkövetett hibákat számszerűsíteni, a torzításokat kiszűrni.

A torzítások csak csökkenthetők, de nem szűrhetők ki teljesen. A torzítások minimalizálása érdekében célszerű, ha

- a vizsgálat alanya nem ismeri az adatfelvétel célját,
- egyértelmű kérdések vannak megfogalmazva,
- kontrollkérdéseket is beiktatnak a kérdések közé.

A nem véletlen mintavételi eljárások tipikus esetei a társadalmi vizsgálatok, sok esetben személyes megkérdezés során alakul ki az információ.

11.2.1. Szubjektív kiválasztás

Önkényesnek is nevezik, mivel a mintavevő a szakmai ismeretére támaszkodva az általa jellemzőnek tartott egyedeket választja ki a sokaságból

11.2.2. Kvóta szerinti kiválasztás:

Előre megadjuk a minta összetételét, azaz előre rögzített megoszlási viszonyzámnak megfelelő lesz a minta. Ehhez megfelelő információ szükséges a sokaságról a vizsgált ismérv szerint. A véletlennel kombinált kvóta kiválasztás azonban jobb, mint a csak kvóta szerinti. A kvótás eljárás a rétegzett mintavételhez hasonló eredményt ad. A lakosság körében végzett felmérések, az adatvédelem miatt egyre inkább kvótás eljárással készülnek.

11.2.3. Koncentrált kiválasztás:

Feltételezi, hogy a sokaság vizsgált jellemzőjét döntően kevés számú egyed határozza meg. A mintavétel során ezeket, a meghatározó elemeket választjuk ki. Pl.: a nemzetgazdasági fogyasztással kapcsolatos elemzések során a fogyasztói árindex meghatározásánál a legnagyobb mértékben fogyasztott termékeket választják ki.

11.3. Kombinált kiválasztás

A véletlen és a tudatos kiválasztás kombinációja. Az N elemű véges sokaságot a vizsgálandó ismérv alapján sorba rendezzük. Az n mintaelem-szám megadása után a sokaság minden k -adik eleme bekerül a mintába olyan módon, hogy

$$k = \left[\frac{N}{n} \right] \text{ (a hányados egész része)}$$

Összefoglalásul elmondható, hogy a mintavétel alapvető célja, hogy létrehozzon az alapsokaság helyett egy kevesebb költséggel és idővel vizsgálható részsokaságot, amely minta elegendő információt nyújt arra, hogy belőle az eredeti sokaságra levont következtetéseink valószínűségi értelemben kellően pontosak legyenek.

12. lecke: Statisztikai becslés

Követelmények:

Ön akkor sajátította el megfelelően a tananyagot, ha

- Tudja a becslőfüggvény jelentését,
- Ki tudja választani a felsorolásból az egy és kétoldali intervallum meghatározásának módját
- Használni tudja a „z”-, a „t”- és a χ^2 -táblákat.
- Meg tudja határozni a valószínűségi szintet
- Önállóan ki tudja számolni az átlagra és a szórásra a konfidencia intervallumot
- Értelmezni tudja a kiszámított intervallumok jelentését
- Meg tudja határozni a minta szükséges elemszámát

A statisztikai becslés az alapsokaságot alkotó valószínűségi változók eloszlásának, jellemzőinek és paramétereinek becslését jelenti az alapsokaságból vett mintából számított mutatók alapján. A statisztikai becsléseket úgynevezett becslőfüggvények segítségével végezzük el. A becslőfüggvény olyan valószínűségi-változó függvény, ami valamely sokasági jellemző mintából történő közelítő meghatározására szolgál.

Egy sokasági jellemzőre több becslőfüggvény is készíthető. A hogy a véletlen minta elemei valószínűségi változók, ugyanúgy a becslőfüggvény értéke is az. Egy adott n-elemű minta csak egyetlen becslt értékkel rendelkezik. A minta alapján az alapsokaságnak többféle jellemzője is becsülhető, pl.:

- számtani átlag,
- értékösszeg,
- arány és megoszlás,
- hányados.

A becslőfüggvény akkor lesz jó, ha a különféle (véletlen) minták esetén értéke a becslni kívánt jellemző körül ingadozik, és az ingadozás lehetőleg kicsi. A becslőfüggvénnyel szemben támasztott követelmények:

- torzítatlanság,
- hatásosság,
- koncisztencia.

Két becslőfüggvény közül, ha más kritériumot nem veszünk figyelembe, a torzítatlant részesítjük előnybe.

Két torzított becslőfüggvény esetén azt tekintjük a jobbnak, amelyiknél a torzítás abszolút értéke kisebb.

12.1. Pontbecslés

A becslőfüggvény értékéről közismert, hogy valószínűségi változó, de egy n elemű mintához csak egyetlen konkrét értéket rendelhetünk. Az eddig tanult sokasági jellemzők (átlag, szórás) becslését is elvégezhetjük pontbecsléssel. A számítás megegyezik például az átlag esetén a korábban tanultakkal (Statisztika I.). Az átlagról azonban tudjuk, hogy nem a sokaság minden tagjára vonatkozó adatok ismeretében számoltuk ki, hanem a mintából így bizonytalanságot rejt magába, ezért a pontbecslés helyett a leggyakrabban az ún. intervallumbecslést alkalmazzuk.

12.2. Intervallumbecslés

Azt keressük, hogy hol vannak a becslésnek azok az alsó és felső határai, amelyek nagy valószínűséggel tartalmazzák a becsülni kívánt jellemzőket. Ezt az intervallumot konfidencia intervallumnak nevezzük. A konfidencia intervallum számításának módja az, hogy a mintát és megfelelő külső információkat felhasználva előre megadott α függvényében meghatározzuk a az alsó (C_a) és a felső (C_f) határokat.

$$P(\hat{\Theta}_{a(\alpha)} < \Theta < \hat{\Theta}_{f(\alpha)}) = 1 - \alpha$$

Az $1-\alpha$ az előre megadott megbízhatósági vagy konfidencia paraméter.

Általában szimmetrikus intervallumokat keresünk. Az intervallumbecslésnél is több esetet különböztetünk meg.

12.2.1. Normális eloszlású sokaság esetén, ha a sokaság szórása (σ) ismert

A becsléshez a z-próbafüggvényt alkalmazzuk. A valószínűségi változó.

$$z = \frac{\hat{\mu} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Keressük a z_1 és z_2 konfidencia intervallumot, amely $1-\alpha$ valószínűséggel tartalmazza a z változót.

Mivel μ -re, (várható értékre) keressük az intervallumot, ezért a konfidencia intervallum az alábbi lesz:

$$\hat{\mu} - z_p * \frac{\sigma}{\sqrt{n}}; \hat{\mu} + z_p * \frac{\sigma}{\sqrt{n}};$$

Konkrét minta esetén:

$$\bar{x} - z_p * \frac{\sigma}{\sqrt{n}}; \bar{x} + z_p * \frac{\sigma}{\sqrt{n}};$$

Először ki kell számolni a standard hibát:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

A z_p értékét „A standard normális eloszlás eloszlásfüggvényeinek értékei” című táblázatból keressük ki a megfelelő valószínűségi szinten (p).

A $z_p * \frac{\sigma}{\sqrt{n}}$ mennyiséget hibahatárnak vagy maximális hibának nevezzük, és Δ -val jelöljük.

$$\Delta = z_p * \frac{\sigma}{\sqrt{n}}$$

Az intervallumbecslés általában kétoldalú, mivel a becslt érték köré szimmetrikus intervallumot szerkesztünk, azaz az intervallum alsó és felső határát pontosan meghatározzuk. Ebben az esetben a táblázatból nem az $(1-\alpha)$ valószínűségi szinthez tartozó értéket keressük, hanem az $(1-\alpha/2)$ -hez tartozót. Pl. ha $(1-\alpha)=95\%=0,95$, akkor $\alpha=0,05$, tehát $\alpha/2=0,025$, így $(1-\alpha/2)=0,975$. A „z”-értékek táblázatból a 0,975-höz tartozó értéket keressük ki, amely 1,96, azaz $z_{0,975}=1,96$.

Egyoldalú intervallum esetén, az $(1-\alpha)$ valószínűségi szinthez tartozó értéket keressük, az intervallumnak csak a felső határát tudjuk megállapítani, az alsó határ a negatív végtelen lesz. A „z”-értékek táblázatból a 0,95-höz tartozó értéket keressük ki, amely 1,65, azaz $z_{0,95}=1,65$.

A becslési hibahatárt többféleképpen csökkenthetjük:

- Csökkentjük a standard hibát: a standard hiba, a mintaátlag szórása csak a minta elemszámától függ, mégpedig annak gyökével fordítottan arányos. Így az elemszám növelésével csökken a standard hiba, és ezáltal a konfidencia intervallum is.
- Csökkentjük a meg bízhatósági szintet, így azaz z_p -értéke is kisebb lesz és szintén kisebb lesz a hibahatár.

Bemutató példa:

Egy gépkocsi típus átlagos fogyasztásának szórása a gyári adatok alapján 0,9 l/100 km. A vizsgálat céljából vett 50 elemű minta átlaga 6,2 l/100 km. 95%-os megbízhatósági szinten állapítsuk meg a fogyasztás konfidencia intervallumát!

a/ Először ki kell számolni a standard hibát:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0,9}{\sqrt{50}} = 0,13$$

A „z”-értékek a „A standard normális eloszlás eloszlásfüggvényének értékei” című táblázatból keressük ki. Megkeressük a táblázat belsejében a 0,975-ös értéket, és leolvassuk az oldal- és fejléctet. Az oldallécben 1,96, a fejlécben 6 an tehát:

$$z_{0,975}=1,96$$

$$\Delta = z_p * \sigma_{\bar{x}} = 1,96 * 0,13 = 0,25$$

(6,2 – 0,25; 6,2 + 0,25) A gépkocsik átlagos fogyasztása 5,95 és 6,45 l között van 100 km-ként.

b/ Nézzük meg, ha a minta elemszáma 100 lett volna, hogyan alakul a konfidencia intervallum.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0,9}{\sqrt{100}} = 0,09 \quad \Delta = z_p * \sigma_{\bar{x}} = 1,96 * 0,09 = 0,1764 \approx 0,18$$

(6,2 – 0,18; 6,2 + 0,18) A gépkocsik átlagos fogyasztása 6,02 és 6,38 l között van 100 km-ként.

c/ Legyen a megbízhatósági szint 90%., az elemszám az eredeti, azaz n=50

$$z_{0,95}=1,65$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0,9}{\sqrt{50}} = 0,13 \quad \Delta = z_p * \sigma_{\bar{x}} = 1,65 * 0,13 = 0,2145$$

(6,2 – 0,2145; 6,2 + 0,2145) A gépkocsik átlagos fogyasztása 5,9855 és 6,4145 között van 100 km-ként.

12.2.2. Normális eloszlású sokaság esetén, ha a sokasági szórás nem ismer, illetve nem normális, de ismert eloszlású sokaság esetén, h nagy mintát vettünk

Mivel az alapsokaság szórása nem ismert, így azt is becsülni kell, azaz a mintából számítjuk ki. Ilyenkor z-valószínűségi változó helyett t-valószínűségi változót (Student-féle eloszlás) használjuk.

$$t = \frac{\hat{\mu} - \mu}{\frac{s}{\sqrt{n}}}$$

A t-eloszlású valószínűségi változó szabadságfoka : szf=n-1.

Először ki kell számolni a standard hibát:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

A hibahatár:

$$\Delta = t_p^{(szf)} * \frac{s}{\sqrt{n}},$$

az intervallum pedig:

$$(\bar{x} - t_p^{(szf)} * \frac{s}{\sqrt{n}}; \bar{x} + t_p^{(szf)} * \frac{s}{\sqrt{n}})$$

A t-eloszlás is szimmetrikus eloszlás, azaz a z-eloszláshoz hasonlóan alakul az egy kétoldali intervallum. A szabadságfok növelésével a t-eloszlás egyre inkább közelít a normális eloszláshoz, száznál nagyobb mintaszám esetén a két eloszlás eltérése minimális.

Bemutató példa

Egy gépkocsi típus átlagos fogyasztásának vizsgálat céljából vett 50 elemű minta átlaga 6,2 l/100 km, szórása pedig 1,0 l/100km. 95%-os megbízhatósági szinten állapítsuk meg a fogyasztás konfidencia intervallumát!

Először ki kell számolni a standard hibát:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{1,0}{\sqrt{50}} = 0,1414$$

A „t”-értékek „ A t (Student) eloszlású változó eloszlásának kvantilis értékei” című táblázatban találhatóak, az szf=49-es szabadságfoknál, és a 0,975-höz tartozó értéket keressük ki:

$$t_{0,975}^{((49))} = 2,01$$

$$\Delta = t_p * s_{\bar{x}} = 2,01 * 0,1414 = 0,2842$$

(6,2 – 0,2842; 6,2 + 0,2842) A gépkocsik átlagos fogyasztása 5,9158 és 6,4842 között van 100 km-ként.

12.2.3. Értékösszegsor becslése

Nagyon nagy sokaság esetén általában nem az átlagra, hanem a sokasági értékösszegre keresünk konfidencia intervallumot. A feladat visszavezethető az átlagbecslésre.

Ha a várható érték: $\mu = \bar{x}$, akkor a sokasági értékösszeg

$$S_i = N * \bar{x}$$

A konfidencia intervallum alsó és felső határa a várhatóértékre meghatározzak N-szerese lesz.

$$N * (\bar{x} - z_p * \frac{\sigma}{\sqrt{n}}); N * (\bar{x} + z_p * \frac{\sigma}{\sqrt{n}});$$

Illetve:

$$N * (\bar{x} - t_p^{(szf)} * \frac{s}{\sqrt{n}}); N * (\bar{x} + t_p^{(szf)} * \frac{s}{\sqrt{n}})$$

Bemutató feladat:

50.000 üvegből vett 500 minta alapján az átlagos töltőtömeg 497 g, a töltőtömeg szórása 19,49g. 95%-os valószínűséggel mennyi lesz az 50000 üveg töltősúlya?

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{19,49}{\sqrt{500}} = 0,8716$$

$$t_{0,975}^{((499))} = 1,96$$

$$\Delta = t_p * s_{\bar{x}} = 1,96 * 0,8716 = 1,7083$$

Az értékösszege sor konfidencia intervalluma $50000 * (497 - 1,7083; 497 + 1,7083)$ g, azaz 24.765kg és 24935 kg között várható az 50.000 db üveg össztöltősúlya.

12.2.4. Sokasági aránybecslés

A nem homogén sokaságot valamilyen minőségi vagy mennyiségi ismerv alapján két csoportba soroljuk és az egyes csoportokba esés valószínűségét kívánjuk meghatározni, akkor aránybecslést végzünk.

A sokasági arányt p -vel jelöljük. Ez azt jelenti, hogy egy egyedet kiválasztva p a valószínűsége annak, hogy az egyed rendelkezik az adott tulajdonsággal. Továbbra is feltételezzük, hogy független, azonos eloszlású minta áll rendelkezésünkre.

A standard hibát az arányokból számoljuk ki.

$$\sigma_p = \sqrt{\frac{p * (1-p)}{n}} \text{ vagy } s_p = \sqrt{\frac{p * (1-p)}{n}}$$

A becslés a továbbiakban a z -próbafüggvénnyel történő becsléssel megegyező.

Bemutató feladat:

A fővárosban 2000 megkérdezett közül 700 ember az egyik pártra szavazna. Becsüljük meg 99%-os valószínűséggel, hogy milyen határok között lesz az erre a pártra szavazók aránya az összeszámlálás után Budapesten.

$$p = 700/2000 = 0,35$$

$$\sigma_p = \sqrt{\frac{p * (1-p)}{n}} = \sqrt{\frac{0,35 * (1-0,35)}{2000}} = 0,011$$

$$1-\alpha/2 = 1-0,005 = 0,995 \quad z_{0,995} = 2,58; \quad \Delta = 2,58 * 0,011 = 0,028 = 2,8\%$$

$$(35-2,8; 35+2,8) = (32,2; 37,8)\%$$

99%-os valószínűséggel arra számíthatunk, hogy a kiválasztott pártra szavazók aránya 32,20% és 37,8% között lesz.

12.2.5. Sokasági szórásnégyzet vagy szórási becslése

A sokasági szórásnégyzet (σ^2) becslésére a torzítatlan becslést eredményező korrigált tapasztalati szórásnégyzetet (s^2) használjuk.

$$s^2 = \frac{\sum (y - \bar{y})^2}{n-1}$$

Ha Y normális eloszlású, akkor bizonyítható, hogy az $\chi^2 = \frac{(n-1) * s^2}{\sigma^2}$ változó $(n-1)$ szabadságfokú χ^2 -eloszlást követ. A χ^2 -eloszlás aszimmetrikus, így a konfidencia intervallum az alábbi:

$$P(\chi_{\alpha/2}^2(szf) < \frac{(n-1) * s^2}{\sigma^2} < \chi_{1-\alpha/2}^2(szf)) = 1 - \alpha$$

Átrendezve:

$$P\left(\frac{(n-1) * s^2}{\chi_{1-\alpha/2}^2(szf)} < \sigma^2 < \frac{(n-1) * s^2}{\chi_{\alpha/2}^2(szf)}\right) = 1 - \alpha$$

$\chi_{1-\alpha/2}^2(szf)$ és a $\chi_{\alpha/2}^2(szf)$ értékét táblázatból kell kikeresni a megfelelő szabadságfoknál és valószínűségi szintnél.

Konkrét minta esetén a szórás intervalluma a következő:

$$\sqrt{\frac{(n-1) * s^2}{\chi_{1-\alpha/2}^2(szf)}} < \sigma < \sqrt{\frac{(n-1) * s^2}{\chi_{\alpha/2}^2(szf)}}$$

Bemutató feladat:

250 g-os kávé csomagoló gép működését vizsgálva 100 elemű mintát veszünk. A töltési tömeg normális eloszlású. Határozzuk meg, hogy milyen határok között lesz a kávécsomagok töltési tömegének szórása 95%-os valószínűségen. Az átlag töltési súly 248g, a minta szórása 5,53g.

$$Szf=100-1=99; \quad \alpha=0,05, \quad \alpha/2=0,025, \quad 1-\alpha/2=0,975,$$

$$\chi_{0,025(99)}^2=74,2; \quad \chi_{0,975(99)}^2=129,6$$

$$\Delta_1 = \sqrt{\frac{(n-1) * s^2}{\chi_{0,975}^2}} = \sqrt{\frac{99 * 5,53^2}{129,6}} = 4,83g$$

$$\Delta_2 = \sqrt{\frac{(n-1) * s^2}{\chi_{0,025}^2}} = \sqrt{\frac{99 * 5,53^2}{74,2}} = 6,39g$$

A nettó töltési tömeg szóródása (ingadozása) 95%-os valószínűségi szinten 4,83g és 6,39g között van.

12.2.6. Adott intervallumszélességhez tartozó elemszám illetve valószínűségi szint meghatározása

Eddig adott elemszám és valószínűségi szint mellett határoztuk meg a konfidencia intervallumot. A becsléskor azonban előre rögzíthetjük a hibahatárt és ehhez kell meghatározni a szükséges minta elemszámát adott valószínűségi szinten. Ha az elemszám és a hibahatár is adott, akkor meg tudjuk mondani, hogy hány százalékos valószínűséggel kerül a sokasági jellemző az adott elemszám esetén az előre meghatározott intervallumba.

Elemzés meghatározása: adott az intervallum és a valószínűség.

Mivel: $\pm \Delta = z_p * \frac{\sigma}{\sqrt{n}}$ illetve $\pm \Delta = t_p^{(szf)} * \frac{s}{\sqrt{n}}$, átalakítás után az alábbiakat kapjuk:

$$n = \left(\frac{z_p * \sigma}{\Delta} \right)^2 \quad \text{illetve} \quad n = \left(\frac{t_p^{(szf)} * s}{\Delta} \right)^2$$

Valószínűségi szint meghatározása:

$\pm \Delta = z_p * \frac{\sigma}{\sqrt{n}}$ és a $\pm \Delta = t_p^{(szf)} * \frac{s}{\sqrt{n}}$ képletek az átalakítás után az alábbiak:

$$z_p = \frac{\Delta * \sqrt{n}}{\sigma} \quad \text{illetve} \quad t_p^{(szf)} = \frac{\Delta * \sqrt{n}}{s}.$$

A kiszámított z_p -értékhez tartozó táblázatbeli $\Phi(z)$ -értékből tudjuk a valószínűséget kiszámítani:

$$P = \Psi(z) = \Phi(z) - (1 - \Phi(z)),$$

ahol $P = \Psi(z)$: a keresett valószínűség,

$\Phi(z)$: a kiszámított z -értékhez tartozó táblázatbeli érték.

A t -eloszlás esetében a kiszámított értéket megkeressük a táblázatban az adott szabadságfoknál, és leolvassuk a hozzá tartozó valószínűséget.

Bemutató feladat:

Egy vizsgálat során, megállapították, hogy a hallgatók átlagos testmagassága 169 cm és 173 cm között van 95%-os valószínűséggel.

A sokaság szórása 9,99 cm.

Hány hallgatót vontak be a vizsgálatba?

$$n = \left(\frac{z_p * \sigma}{\Delta} \right)^2 = \left(\frac{1,96 * 9,99}{2} \right)^2 = 95,84, \text{ azaz } 96 \text{ hallgató testmagasságát mérték le.}$$

Mekkora valószínűségi szinten végezték a vizsgálatot, ha 200 hallgató esetében is ugyanezt a konfidencia intervallumot kapták?

$$z_p = \frac{\Delta * \sqrt{n}}{\sigma} = \frac{2 * \sqrt{200}}{9,99} = 2,83$$

A táblázatbeli $\Phi(z) = 0,9977$

$$P = 0,9977 - (1 - 0,9977) = 0,9954 = 99,54\%.$$

Tehát a vizsgálatot 99,54%-os megbízhatósági szinten végezték.

13. lecke Statisztikai próbák - Hipotézisvizsgálat

Követelmények

Ön akkor sajátította el megfelelően a tananyagot, ha

- tudja mit jelent a hipotézisvizsgálat,
- tudja a próbafüggvény lehetséges értékeinek tartományát, és annak felbontását,
- ki tudja választani felsorolásból a szignifikancia szint jelentését,
- tudja a hipotézisvizsgálat során elkövetett hibákat.
- ha ki tudja választani, hogy mely esetekben alkalmazhatóak az egyes próbák,
- önállóan el tudja dönteni a feladat alapján, hogy a hipotézisvizsgálathoz melyik próbafüggvény szükséges,
- fel tudja írni a feladatok alapján az alternatív hipotézist,
- végre tudja hajtani a hipotézisvizsgálatokat a megfelelő próbákkal (z-próba, t-próba, χ^2 -próba),
- meg tudja határozni a próbák (z-próba, t-próba, χ^2 -próba) elfogadási tartományát.
- ki tudja választani az eredmények alapján az eredmények helyes megfogalmazását.

13.1. Alapfogalmak

Az eddigi becslési eljárások során a sokasági paramétert ismeretlennek tekintettük, és a mintából származó adatok segítségével közelítőleg meghatároztuk az ismeretlen sokasági értékét. A hipotézisvizsgálatnál a sokaságról állítunk valamit, majd a rendelkezésünkre álló minta alapján ellenőrizzük az állítás helyességét.

Az egy vagy több sokaságra vonatkozó állítást, feltevést hipotézisnek nevezzünk. A hipotézis vonatkozhat az egy vagy több sokaság eloszlására, vagy az adott eloszlások egy vagy több paraméterére is. A különféle hipotézisek vizsgálatára szolgáló eljárásokat statisztikai próbáknak nevezzük. A próba egy olyan eljárás, amelynek során a mintából származó információk alapján döntünk a hipotézis elfogadásáról, vagy elutasításáról.

A hipotézisvizsgálat első lépése a vizsgálni kívánt hipotézis megfogalmazása. Pontosabban mindig két hipotézist fogalmazunk meg, egy úgynevezett nullhipotézist (H_0), és egy ezzel szemben álló alternatív hipotézist (H_1). A vizsgálat során a két hipotézist „versenyeztetjük”, és azt fogadjuk el igaznak, amelyik a mintavétel eredménye alapján hihetőbbnek tűnik a másiknál. A két hipotézist úgy kell megfogalmazni, hogy:

- akármelyiket is tekintjük majd a másiknál hihetőbbnek, megválaszolható legyen a bennünket érdeklő kérdés;
- a formális logikai szabályai szerint kizárják egymást, azaz egyszerre ne lehessenek igazak, de együtt minden lehetőséget kimerítsenek.

A hipotézis lehet egyszerű, ha fennállásának feltételezése a sokaság eloszlását egyértelműen meghatározottá teszi. Ellenkező esetben összetett hipotézisről beszélünk, azaz az egyszerű hipotézisek halmazáról.

A hipotézisek megfogalmazása után a feladatunk a mintaelemek egy olyan függvényének a keresése, amelynek valószínűség-eloszlása a nullhipotézis helyességének feltételezése, a

sokaságra tett bizonyos kikötések és a mintavétel adott módja mellett egyértelműen meghatározható. Az e követelményeknek eleget tevő függvényt próbafüggvénynek nevezzük. A próbafüggvény hasonló szerepet tölt be a hipotézisvizsgálat során, mint a becslőfüggvény a becsléskor. A próbafüggvény konstruálása matematikai feladat.

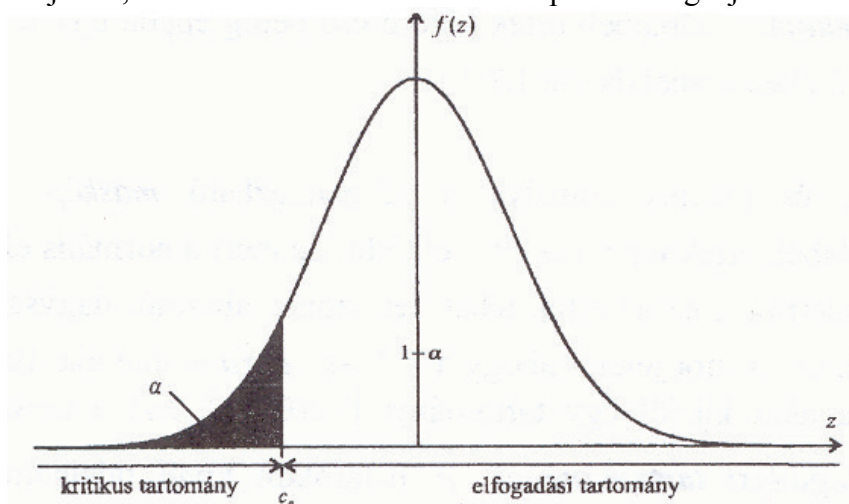
A hipotézis helyességének ellenőrzése a próbafüggvény lehetséges értékeinek teljes tartományát osztópontok segítségével (C_a ; C_f) két egymást át nem fedő tartományra bontjuk. Az egyik az elfogadási tartomány (E), a másik egy elutasítási vagy kritikus tartomány (K). Az egyes tartományok határait úgy választjuk meg, hogy a próbafüggvény értéke a nullhipotézis elfogadása esetén előre megadott valószínűséggel ($p=1-\alpha$) az elfogadási tartományba essen, és a kritikus tartományba esés csak α -valószínűséggel következzen be. Az $(1-\alpha)$ a konfidencia szint, ennek komplementere a szignifikancia szint (α). A próbafüggvény kritikus tartományba esésének valószínűségét szignifikancia szintnek nevezzük, és α -val jelöljük. Pl.: $\alpha=0,05$ szignifikancia szint azt jelenti, hogy ha a mintavételt végtelen sokszor végrehajtjuk, akkor 100 esetből összesen 5-ször fordul elő az, hogy a próbafüggvényünk minta alapján kiszámított értéke a kritikus tartományba esik.

Ha ezek után a rendelkezésre álló minta adataiból kiszámítjuk a próbafüggvény úgynevezett aktuális értékeit, és ez beleesik az elfogadási tartományba, akkor a nullhipotézist elfogadjuk, ellenkező esetben a nullhipotézist elutasítjuk, és az alternatív hipotézist fogadjuk el.

Az elfogadási és a kritikus tartomány egymáshoz viszonyított elhelyezése háromféle lehet: Egyoldali kritikus tartományhoz abban az esetben jutunk, ha az ellenhipotézisben a nullhipotézishez képest egy meghatározott irányú eltérést írunk fel.

Baloldali kritikus tartomány:

Ha a sokaság várható értékre $H_1: \mu < \mu_0$ alternatív hipotézist fogalmazzuk meg, akkor baloldali, kritikus tartományról beszélünk. A nullhipotézist abban az esetben fogadjuk el, ha a próbafüggvény számított értéke nagyobb az elfogadási tartomány alsó határánál. Az elfogadási tartomány felső határa ebben az esetben pozitív végtelen. Ellenkező esetben a nullhipotézist vetjük el, és természetesen az alternatív hipotézist fogadjuk el.



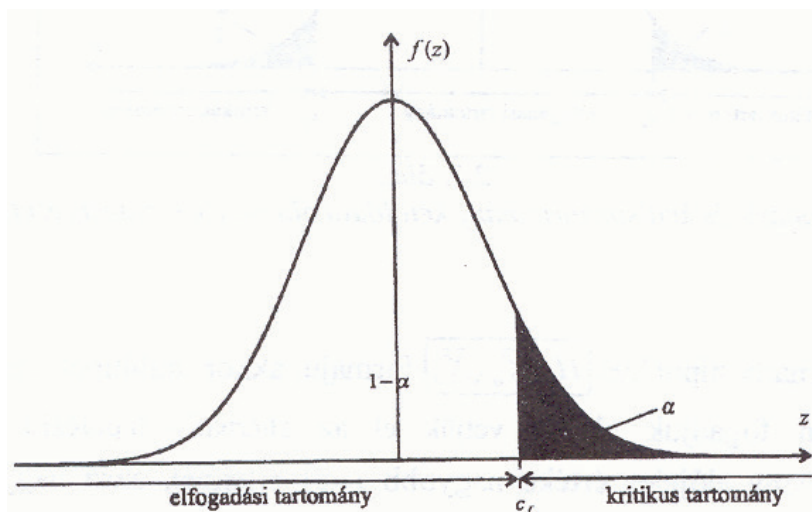
Elfogadási- és kritikus (baloldali) tartomány egyoldalú alternatív hipotézis esetén

Jobboldali kritikus tartomány:

Ha a sokaság várható értékre $H_1: \mu > \mu_0$ alternatív hipotézist fogalmazzuk meg akkor jobboldali kritikus tartományról beszélünk.

A nullhipotézist abban az esetben fogadjuk el, ha a próbafüggvény számított értéke kisebb az elfogadási tartomány felső határánál. Az elfogadási tartomány alsó határa ebben az esetben

negatív végtelen. Ellenkező esetben a nullhipotézist vetjük el, és természetesen az alternatív hipotézist fogadjuk el.

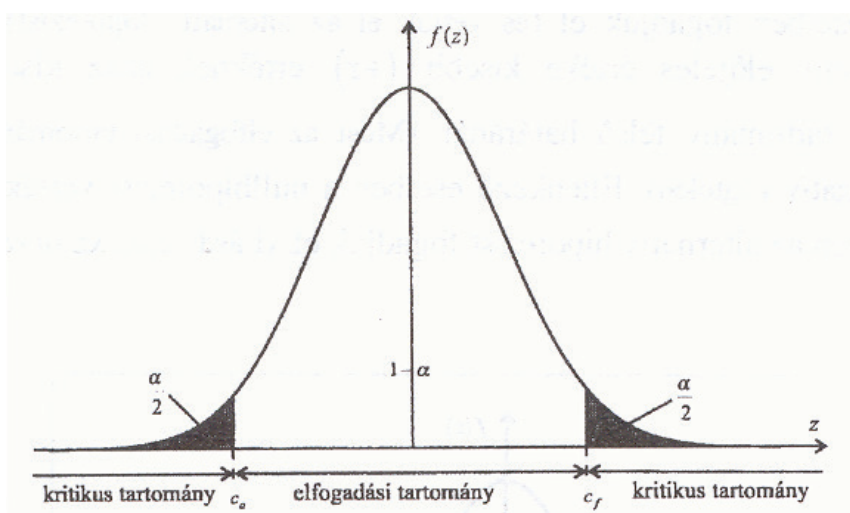


Elfogadási- és kritikus (jobboldali) tartomány egyoldalú alternatív hipotézis esetén

Kétoldali kritikus tartomány:

Kétoldali kritikus tartomány kijelölésére olyan esetben kerül sor, amikor a nullhipotézisben megfogalmazott állítástól való bármilyen irányú eltérés érdekel bennünket ($H_1: \mu \neq \mu_0$).

Ha a próbafüggvény számított értéke az elfogadási tartományba kerül, akkor a nullhipotézist fogadjuk el, ha a próbafüggvény értéke a kritikus tartományba kerül, akkor a nullhipotézist elvetjük.



Elfogadási- és kritikus tartomány kétoldali alternatív hipotézis esetén

A hipotézisvizsgálat során elkövetett hibák

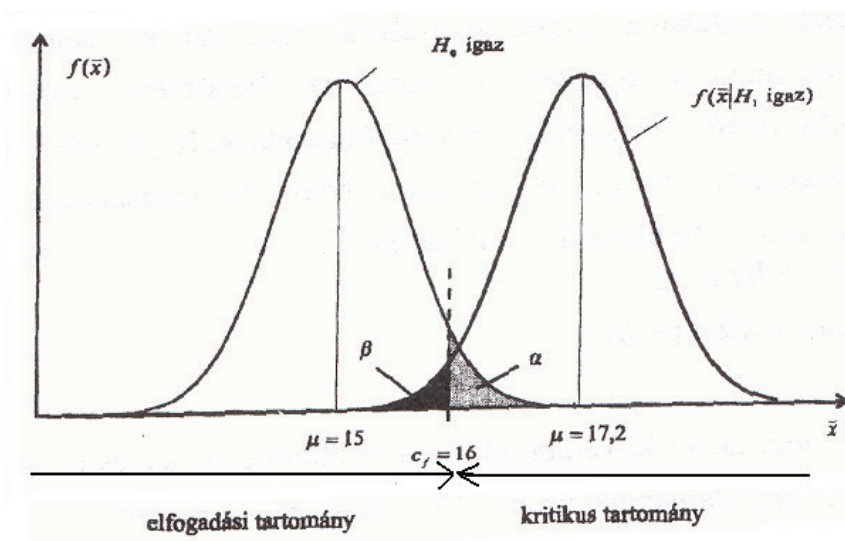
A mintából a sokaságra vonatkozóan csak valószínűségi következtetés lehetséges, így a hipotézisvizsgálat során hozott döntésünk bizonyos kockázattal jár.

Előfordulhat, hogy a nullhipotézis helyes, és a próbafüggvény adott mintából számított értéke mégis a kritikus tartományba esik. Ilyenkor a nullhipotézist annak ellenére, hogy fennáll, elutasítjuk. Ezt a hibás döntést elsőfajú hibának nevezzük. Az ilyen hiba elkövetésének valószínűsége az elfogadási és a kritikus tartomány konstrukciója alapján α , amelyet szignifikanciaszintnek nevezünk.

Előfordul, hogy a nullhipotézis nem áll fenn (nem igaz), és a próbafüggvény mintából számított értéke mégis az elfogadási tartományba esik. Ez szintén hibás döntés, és ilyenkor másodfokú hibát követünk el. Ezen esemény bekövetkezésének valószínűségét β -val jelöljük.

A hipotézisvizsgálat során elkövetett hibák

Valóságos helyzet	H ₀ -ra vonatkozó döntés	
	Elfogadjuk	elutasítjuk
H ₀ igaz	Helyes döntés 1- α	Elsőfajú hiba α
H ₁ igaz	Másodajú hiba β	Helyes döntés 1- β



A hipotézisellenőrzés során elkövethető hibák

Az $(1-\alpha)$ valószínűséget a próba megbízhatósági szintjének, az $(1-\beta)$ -t pedig a próba erejének nevezzük.

A minta elemszámának növelésével - adott szignifikancia szint és alternatív hipotézis esetén – csökkenthető a másodfajú hiba elkövetésének valószínűsége, illetve minél távolabb van μ paraméter valóságos értéke a nullhipotézisben szereplő feltételezett értéktől, annál kisebb lesz β -értéke.

A hipotézisvizsgálat menet:

1. Megfogalmazni a nullhipotézist és az alternatív hipotézist.
2. Próbafüggvény keresése a nullhipotézisben megfogalmazott állításnak megfelelően. A próbafüggvény a mintaelemeknek egy olyan függvénye, amelynek eloszlása a nullhipotézis igazságát feltételezve pontosan ismert, a mintavétel előtt azonban ennek értéke is valószínűségi változó. A próbafüggvénynek több szempont szerint kell megfelelőnek lennie, egyrészt a nullhipotézisben megfogalmazott állításnak, azaz nem

minden nullhipotézis ellenőrizhető azonos próbafüggvénnyel; másrészt a minta eloszlás-típusának, valamint a mintavétel módjának is.

3. Kiválasztani a szignifikancia szintet: az az elsőfajú hiba elkövetésének valószínűségét választjuk meg.
4. Az elfogadási és kritikus tartomány megállapítása a szignifikancia szintnek és a szabadságfoknak (ahol van) megfelelően.
5. Mintavétel, és a próbafüggvény értékének kiszámítása.
6. Döntés a nullhipotézis helyességének elfogadásáról, vagy a nullhipotézis elutasítása.

13.2. Egymintás statisztikai próbák

Az egymintás statisztikai próbák a sokaság valamely paraméterének tesztelésére szolgálnak.

13.2.1. Várható értékre irányuló próbák

Azt teszteljük, hogy egy sokaság ismeretlen várható értéke (μ), megegyezik-e az általunk feltételezett várható értékkel (μ_0). A nullhipotézis a következő:

$$H_0: \mu = \mu_0$$

Konkrét minta esetén:

$$H_0: \bar{x} = \bar{x}_0$$

Az alternatív hipotézisünk háromféle lehet:

$$H_1: \mu < \mu_0$$

$$H_1: \mu \neq \mu_0$$

$$H_1: \mu > \mu_0$$

Konkrét minta esetén:

$$H_0: \bar{x} < \bar{x}_0$$

$$H_0: \bar{x} \neq \bar{x}_0$$

$$H_0: \bar{x} > \bar{x}_0$$

13.2.1.1. Egymintás z-próba

A sokaság normális eloszlású és a sokasági szórás (σ) ismert, akkor hasonlóan a becsléshez, a z-próbafüggvényt alkalmazzunk.

$$z_0 = \frac{\hat{\mu} - \mu_0}{\frac{\sigma}{\sqrt{n}}},$$

illetve ismert minta esetén:

$$z_0 = \frac{\bar{x} - \bar{x}_0}{\frac{\sigma}{\sqrt{n}}}$$

Ez a próbafüggvény standard normális eloszlású valószínűségi változó.

A z-próba elfogadási tartományának határai az alábbiak:

Alternatív hipotézis	elfogadási tartomány
$H_1: \mu < \mu_0$	$[z_\alpha; \infty[$
$H_1: \mu \neq \mu_0$	$[z_{\alpha/2}; z_{1-\alpha/2}]$
$H_1: \mu > \mu_0$	$] -\infty; z_{1-\alpha}]$

A táblázatban csak a $z_{1-\alpha}$ értékeit találjuk meg, azonban a z_α -értékeit az alábbi összefüggés alapján meghatározhatjuk:

$$Z_\alpha = -Z_{1-\alpha}$$

Bemutató feladat:

Egy automata gépsor lisztet csomagol, a szabvány szerint 100 dkg-os tömeggel, és a megengedett szórás 3 dkg. Ellenőrzés céljából 30 db-os mintát veszünk. A lemért lisztes zacskók átlagos tömege 98 dkg. Ellenőrizzük 5%-os szignifikancia szinten, hogy a gép megfelelően csomagol-e.

$H_0: \mu = 100$ dkg.

$H_1: \mu \neq 100$ dkg

$$z_0 = \frac{98 - 100}{\frac{3}{\sqrt{30}}} = -3,65 \quad z_{0,975} = 1,96; z_{0,025} = -1,96$$

Az elfogadási tartomány: (-1,96; 1,96)

Az elfogadási tartomány nem tartalmazza a próbafüggvény aktuális értékét (-3,65), ezért a nullhipotézist elutasítjuk, azaz 5%-os szignifikancia szinten a töltési tömeg nem felel meg a szabványnak.

13.2.1.2. Egymintás t-próba

A normális eloszlású sokaság vizsgálatánál végezzük, ha nem ismerjük az eloszlás szórását. Ebben az esetben a

$$t_0 = \frac{\hat{\mu} - \mu_0}{\frac{s}{\sqrt{n}}}$$

próbafüggvényt használjuk, illetve konkrét minta esetén:

$$t_0 = \frac{\bar{x} - x_0}{\frac{s}{\sqrt{n}}},$$

a nullhipotézis ellenőrzésére.

Ha a nullhipotézis igaz, és a sokaság eloszlása valóban normális, akkor a t-próbafüggvény szf=n-1 szabadságfokú Student-féle t-eloszlást követ. A t-próba elfogadási tartományának határai az alábbiak:

Alternatív hipotézis	elfogadási tartomány
$H_1: \mu < \mu_0$	$\left[t_{\alpha}^{szf} ; \infty \right[$
$H_1: \mu \neq \mu_0$	$\left[t_{\alpha/2}^{szf} ; t_{1-\alpha/2}^{szf} \right]$
$H_1: \mu > \mu_0$	$\left] -\infty ; t_{1-\alpha}^{szf} \right]$
$t_{\alpha}^{szf} = -t_{1-\alpha}^{szf}$	

Bemutató feladat:

Az előző sokaság eloszlása normális, az átlag 98 dkg, a szórást a mintából (n=30) becsültük meg, ami 5,5 dkg. Ellenőrizzük 5%-os szignifikancia szinten, hogy a gép megfelelően csomagol-e.

$$H_0: \mu = 100 \text{ dkg.}$$

$$H_1: \mu \neq 100 \text{ dkg}$$

$$t_0 = \frac{98 - 100}{\frac{5,5}{\sqrt{30}}} = -1,99 \quad t_{0,975}^{29} = 2,05; \quad t_{0,025}^{29} = -2,05$$

Az elfogadási tartomány: (-2,05; 2,05)

Az elfogadási tartomány tartalmazza a próbafüggvény aktuális értékét (-1,99), ezért a nullhipotézist elfogadjuk, azaz 5%-os szignifikancia szinten a töltési tömeg megfelel a szabványnak.

13.2.2. Sokasági szórásra vonatkozó próba

A sokasági szórás becslésére a korrigált tapasztalati szórást használjuk. A konfidencia intervallum meghatározását a χ^2 -eloszlásra (khi) alapozzuk.

$$\chi_0^2 = \frac{(n-1) * s^2}{\sigma_0^2}$$

próbafüggvényt használjuk, amely szf=n-1 szabadságfokú χ^2 -eloszlást követ.

A nullhipotézisünk:

$$H_0: \sigma = \sigma_0$$

A χ^2 -próba elfogadási tartományának határai az alábbiak:

Alternatív hipotézis

elfogadási tartomány

$$H_1: \sigma < \sigma_0 \quad \left[\chi^2_{\alpha(szf)} ; \infty \right]$$

$$H_1: \sigma \neq \sigma_0 \quad \left[\chi^2_{\alpha/2(szf)} ; \chi^2_{1-\alpha/2(szf)} \right]$$

$$H_1: \sigma > \sigma_0 \quad \left[0 ; \chi^2_{1-\alpha(szf)} \right]$$

Bemutató feladat:

Az előző példában feltételezzük, hogy a gép 3 dkg-os szórással tölt. A 30 elemű mintából számított szórás 5,5 dkg volt. Ellenőrizzük, hogy helyes volt-e a feltevés, hogy a gép maximum 3 dkg szórással tölt, 5%-os szignifikancia szinten.

$$H_0: \sigma = 3 \text{ dkg.}$$

$$H_1: \sigma > 3 \text{ dkg}$$

$$\chi_0^2 = \frac{(30-1) * 5,5^2}{3^2} = 97,5$$

A $\chi_{0,95(29)}^2 = 42,6$, tehát az elfogadási tartomány $(0; 42,6)$, a próbafüggvény értéke nem esik bele ebbe a tartományba, ezért a nullhipotézist elutasítjuk, azaz a töltés során a szórás meghaladja az előírást.

13.2.3. Függetlenségvizsgálat

A függetlenségvizsgálat azon nullhipotézis ellenőrzésére szolgál, hogy két ismerv független egymástól. Az alternatív hipotézisben pedig azt fogalmazzuk meg, hogy nem függetlenek.

A két ismerv akkor független egymástól, ha a peremmegoszlási viszonzyszámok (relatív gyakoriságok) szorzata egyenlő s megfelelő együttes viszonzyszámokkal:

$$\frac{f_{1\bullet}}{N} * \frac{f_{\bullet 1}}{N} = \frac{f_{11}}{N}$$

Ha nem ismerjük a véges sokaságot, akkor a mintából származó adatokkal kell eldönteni a függetlenséget. Ilyenkor is egy kontingenciátáblából indulunk ki, de a táblázat ekkor a mintában észlelt gyakoriságokat tartalmazza.

$$H_0: P_{ij} = P_{i\bullet} * P_{\bullet j} \quad (i=1,2,\dots,s; j=1,2,\dots,t)$$

$$H_1: P_{ij} \neq P_{i\bullet} * P_{\bullet j}$$

P_{ij} : az első ismerv i -edik és a második ismerv j -edik változata együttes előfordulásának valószínűsége a sokaságban.

A valószínűségeket a mintából becsüljük:

$$p_{i\bullet} = n_{i\bullet} / n \quad p_{\bullet j} = n_{\bullet j} / n$$

$$\chi^2 = \sum_{i=1}^s \sum_{j=1}^t \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*} = n * \left(\sum_{i=1}^s \sum_{j=1}^t \frac{n_{ij}^2}{n_{i\bullet} * n_{\bullet j}} - 1 \right)$$

$$n_{ij}^* = n * p_{i\bullet} * p_{\bullet j} = \frac{n_{i\bullet} * n_{\bullet j}}{n}$$

Vagy a Csuprov-féle együttható szerint, ahol

$$f_i = \frac{n_i}{n}$$

$$\chi^2 = \sum_{i=1}^s \sum_{j=1}^t \frac{(f_{ij} - f_{ij}^*)^2}{f_{ij}^*}$$

$$f_{ij}^* = n * P_{i\bullet} * P_{\bullet j}$$

A szabadságfok: szf=(s-1)*(t-1)

Ez a próba jobb oldali módon hajtható végre. A minta akkor tekinthető elég nagynak, ha még a legkisebb n_{ij}^* is legalább 5, de még jobb, ha legalább 10.

Bemutató feladat:

Egy szociológiai vizsgálat során azt kívánjuk ellenőrizni, hogy az egyetemet végzett férfiak és nők előrejutási lehetőségei azonosnak tekinthetők-e. Ehhez a 15 éve végzett hallgatók közül 200 főt kiválasztva véletlenszerűen, az alábbi mintát kaptuk.

Megn.	Férfi	nő	$\Sigma n_{i\bullet}$
Beosztott	20	40	60
középvezető	60	40	100
Felső vezető	30	10	40
$\Sigma n_{\bullet j}$	110	90	200

A vizsgálatot 5%-os szignifikancia szinten végezzük el.

$H_0: P_{ij} = P_{i\bullet} * P_{\bullet j}$ $H_1: P_{ij} \neq P_{i\bullet} * P_{\bullet j}$

Megnevezés		n_{ij}	$n_{ij}^* = \frac{n_{i\bullet} * n_{\bullet j}}{n}$	$\frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}$
Férfi	Beosztott	20	$60 * 110 / 200 = 33$	5,121
	középvezető	60	55	0,455
	Felső vezető	30	22	2,909
Nő	Beosztott	40	27	6,259
	középvezető	40	45	0,556
	Felső vezető	10	18	3,556
Σ		200	200	18,856

$$\chi^2_0 = 18,856$$

$$\chi^2_{0,95(2)} = 5,99$$

$$\text{szf} = (3-1) * (2-1) = 2$$

Mivel a kritikus érték kisebb, mint a számított érték, a H_0 -t elutasítjuk, tehát az adatok alapján 5%-os szignifikancia szinten elmondható, hogy a nemhez való tartozás és a beosztás függenek egymástól, azaz elutasítjuk a függetlenséget.

13.2.4. Illeszkedésvizsgálat

Egy valószínűségi változó eloszlására vonatkozó állítás vagy feltételezés ellenőrzését illeszkedésvizsgálatnak nevezzük. Attól függően, hogy a hipotézisünket mennyire konkretizáljuk, kétféle illeszkedésvizsgálatot különböztetünk meg:

Ha a feltételezett eloszlás egyértelműen meghatározott - a típusát és a paramétereit előre rögzítjük -, akkor tiszta illeszkedésvizsgálatról beszélünk.

Ha a feltételezett eloszlásnak csak a típusát adjuk meg - a paramétereit pedig a mintából becsüljük -, akkor becsléses illeszkedésvizsgálatot végzünk.

A sokaságot egy ismerv (többnyire mennyiségi, néha minőségi) alapján k -számú részre bontjuk, azaz az adott ismerv alapján osztályozzuk a sokaság egységeit. Ugyanezt az osztályozást a mintán belül is elvégezzük.

Osztály	A kategória előfordulásának		
	valószínűsége	gyakorisága	relatív gyakorisága
		a mintában	
c_1	P_1	f_1	g_1
c_2	P_2	f_2	g_2
.	.	.	.
.	.	.	.
c_k	P_k	f_k	g_k
Összesen	1,00	n	1,00

Az általunk feltételezett sokaság eloszlása minden ismervváltozathoz egy meghatározott P_i valószínűséget rendel. A nullhipotézis tehát:

$H_0: P(c_i) = P_i \quad i=1,2,\dots,k$, az alternatív hipotézisünk pedig:

$H_1: P(c_i) \neq P_i$

A H_0 helyességét a χ^2 -próbafüggvénnyel vizsgálhatjuk meg:

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - n * P_i)^2}{n * P_i} = \sum \frac{(f_i - f_i^*)^2}{f_i^*} = n * \left(\sum \frac{g_i^2}{P_i} - 1 \right)$$

$$f_i^* = n * P_i$$

Ez a statisztikai vizsgálat a nullhipotézis helyessége esetén jó közelítéssel $szf=(k-b-1)$ szabadságfokú χ^2 -eloszlású, ahol a b a P_i valószínűségek meghatározásához szükséges olyan paraméterek száma, amelyeket a mintából becsülünk. Tiszta illeszkedésvizsgálat esetén a $b=0$. Mivel χ^2 -próbafüggvény a nullhipotézistől való jelentős eltérést nagy pozitív értékkel jelzi, ezért az illeszkedésvizsgálatot a jobb oldali kritikus tartományra kell végrehajtani, azaz a felső kritikus értéket kell keresni, tehát az elfogadási tartomány pedig:

$$[0; \chi_{1-\alpha(szf)}^2]$$

Bemutató feladat:

Egy gyorsbüfé hálózatban a vevőket 45 másodperc alatt kell kiszolgálni. A kiszolgálási idő megengedett szórása 7 másodperc. 400 véletlenül kiválasztott vendég kiszolgálási idő szerinti megoszlása a következő:

kiszolgálási idő (másodperc)	vendégek száma (fő)	Ellenőrizzük azt a feltevést, hogy a minta az előírt paraméterű (átlag=45, szórás=7 másodperc) normális eloszlásból származott, P=5%-os szignifikancia szinten.
0-35	20	
35-40	80	
40-45	100	
45-50	100	
50-55	60	
55-	40	
Összesen	400	

kiszolgálási idő (másodperc)	vendégek száma (fő)	x_{if}	$z_{if}=(x_{if}-\mu)/\sigma$	$\Phi Z_{(if)}=P_i$	P_i	$f_i^*=n \cdot P_i$	$(f_i-f_i^*)^2/f_i^*$
0-35	20	35	-1,43	0,0764	0,0764	30,56	3,65
35-40	80	40	-0,71	0,2389	0,1625	65	3,46
40-45	100	45	0	0,5	0,2611	104,44	0,19
45-50	100	50	0,71	0,7611	0,2611	104,44	0,19
50-55	60	55	1,43	0,9236	0,1625	65	0,38
55-	40	∞	∞	1	1,00	30,56	2,42
Összesen	400	-	-	-	-	400	10,79

A táblázatban vastagított számokat az alábbiak szerint kapjuk:

$$z_{if}=(x_{if}-\mu)/\sigma=35-45/7=-1,43$$

$$\Phi(-Z_{(if)})=1-P(z_{i1}) \text{ táblázatból}$$

$$P_i=P'_{ik}-P'_{ik-1}=0,2389-0,0764=0,1625$$

$$f_i^*=n \cdot P_i=400 \cdot 0,0764=30,56$$

$$(f_i-f_i^*)^2/f_i^*=(20-30,56)^2/30,56=3,65$$

Ugyan így kell a többi értéket is kiszámítani

$$\chi^2_0=10,79$$

$$Sz_f=6-1=5, \quad \chi^2_{1-0,05(5)}=11,1,$$

az elfogadási tartomány (0;11,1).

A számított érték az elfogadási tartományba esik, így elfogadjuk a nullhipotézist. A kiszolgálási időt 5%-os szignifikancia szinten 45 másodperc várható értékű és 7 másodperc szórású normális eloszlású valószínűségi változónak lehet tekinteni.

Bemutató feladat:

Egy széleskörű vizsgálat során, Magyarországon a 15 éves és idősebb népesség 15%-a sovány, 25%-a normál súlyú, és 60%-a túlsúlyos volt 1996-ban. 2015-ben 500 véletlenszerűen kiválasztott minta alapján 72 fő sovány, 176 fő normál súlyú és 252 fő pedig

túlsúlyos volt. 1%-os szignifikancia szinten állíthatjuk-e, hogy a két eloszlás egyforma. A kérdés az illeszkedésvizsgálattal válaszolható meg.

Osztály	A kategória előfordulásának		
	Valószínűsége P_i	gyakorisága	relatív gyakorisága
		a mintában	
Sovány	0,15	72	72/500=0,1444
Normál	0,25	176	176/500=0,352
túlsúlyos	0,60	252	252/500=0,504
összesen	1,00	500	1,00

$$\chi^2 = n * (\sum \frac{g_i^2}{P_i} - 1) = 500 * (\frac{0,1444^2}{0,15} + \frac{0,352^2}{0,25} + \frac{0,504^2}{0,60} - 1) = 28,608$$

$$szf=3-1=2$$

$$\chi^2_{0,99(2)}=9,21$$

az elfogadási tartomány (0; 9,21).

A két eloszlás nem egyezik, mivel a próbafüggvény számított értéke az elfogadási tartományba esik

13.3. Két- és több mintás statisztikai próbák

Gyakran előfordul, hogy két sokaságot akarunk vizsgálni, és a hipotézis két paraméter értékének egymáshoz való viszonyára vonatkozik. Ilyenkor kétmintás próbát hajtunk végre, azaz a sokaságokból 1-1 független, véletlen mintát veszünk a hipotézis ellenőrzése céljából. Az egymással összehasonlításra kerülő sokaságok időben, térben vagy bármilyen más tekintetben különbözhetnek egymástól.

13.3.1. Várható értékek különbözőségére irányuló próbák

Két sokaságból külön-külön és egymástól függetlenül vett minta alapján ellenőrizni kívánjuk a $H_0 : \mu_1 = \mu_2$ vagy konkrét minta esetén a $H_0 : \bar{x}_1 = \bar{x}_2$ hipotézis helyességét.

13.3.1.1. Kétmintás z-próba

Ha a két sokaság normális eloszlású, és ismert mindkét sokaság szórása, akkor a z-próbafüggvényt alkalmazzuk:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

A próba elfogadási tartománya megegyezik az egymintás z-próba elfogadási tartományával:

Alternatív hipotézis	elfogadási tartomány
$H_1: \mu_1 < \mu_2 \quad (x_1 < x_2)$	$[z_\alpha; \infty[$
$H_1: \mu_1 \neq \mu_2 \quad (x_1 \neq x_2)$	$[z_{\alpha/2}; z_{1-\alpha/2}]$
$H_1: \mu_1 > \mu_2 \quad (x_1 > x_2)$	$] -\infty; z_{1-\alpha}]$

Bemutató feladat:

A levegőszennyeződés vizsgálatokor az ülepedő por ($\text{g/m}^2/\text{hó}$) mennyiségét mérték meg téli és nyári időszakban. A mérés eredménye:

Télen: $n_1=60$ $\bar{x}_1 = 5,1$ $\sigma_1=3$

Nyáron: $n_2=60$ $\bar{x}_2 = 5,9$ $\sigma_2=3,9$

Megoldás:

$$H_0: \bar{x}_1 = \bar{x}_2$$

$$H_1: \bar{x}_1 \neq \bar{x}_2$$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{5,1 - 5,9}{\sqrt{\frac{3^2}{60} + \frac{3,9^2}{60}}} = -1,26$$

$$z_{0,975}=1,96$$

Az elfogadási tartomány: $(-1,96, 1,96)$

A számított érték az elfogadási tartományba esik, így a levegőszennyeződés a mért adatok alapján, azonos télen és nyáron.

13.3.1.2. Kétmintás t-próba

Ha a két normális eloszlású sokaság szórását nem ismerjük, és feltételezzük, hogy szórásuk lényegesen nem különbözik, ilyenkor t-próbát alkalmazunk:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{s_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_d = \sqrt{\frac{(n_1 - 1) * s_1^2 + (n_2 - 1) * s_2^2}{n_1 + n_2 - 2}} \quad (\text{közös szórás})$$

A szabadságfok $szf=n_1+n_2-2$

A próba elfogadási tartománya megegyezik az egymintás t-próba elfogadási tartományával.

Bemutató feladat:

Egy üzemben a szerelési műveleteket két eltérő módon tanították be. A két csoportból mintát vettek, és feljegyezték a dolgozók teljesítményét. A kérdés, hogy 5%-os valószínűségi szinten van-e különbség a két szerelési mód között?

$$n_1=16, \quad \bar{x}_1=128 \quad s_1=18$$

$$n_2=11, \quad \bar{x}_2=112 \quad s_2=19$$

Megoldás:

$$H_0: x_1 \geq x_2,$$

$$H_1: x_1 < x_2,$$

$$s_d = \sqrt{\frac{(16-1) \cdot 18^2 + (11-1) \cdot 19^2}{16+11-2}} = 18,41$$

$$t_0 = \frac{128-112}{18,41 \sqrt{\frac{1}{16} + \frac{1}{11}}} = 2,22$$

Szf=25 $t_{0,025}^{(25)} = -1,71$, az elfogadási tartomány:

$$\left[t_{\alpha}^{szf}; \infty \right], \text{ azaz } \left[-1,71; \infty \right]$$

Mivel a számított t-érték beleesik az elfogadási tartományba, ezért a nullhipotézist elfogadjuk, azaz az első betanítási módszer nem jobb, mint a második.

Ha a két minta szórása nagymértékben különbözik, akkor a kétmintás t-próba nem alkalmazható a két várható érték egyezésének eldöntésére. Ilyen esetekben az úgynevezett Welch-eljárást alkalmazzuk:

$$t_f = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Bemutató feladat:

Egy üzemben a szerelési műveleteket két eltérő módon tanították be. A két csoportból mintát vettek, és feljegyezték a dolgozók teljesítményét. A kérdés, hogy 5%-os valószínűségi szinten van-e különbség a két szerelési mód között?

$$n_1=16, \quad \bar{x}_1=128 \quad s_1=18$$

$$n_2=11, \quad \bar{x}_2=112 \quad s_2=29$$

Megoldás:

$$H_0: x_1 \geq x_2,$$

$$H_1: x_1 < x_2,$$

$$t_f = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{|128-112|}{\sqrt{\frac{18^2}{16} + \frac{29^2}{11}}} = 1,63$$

Szf=25 $t_{0,025}^{(25)} = -1,71$, az elfogadási tartomány:

$$\left[t_{\alpha}^{szf}; \infty \right], \text{ azaz } \left[-1,71; \infty \right]$$

Mivel a számított t-érték beleesik az elfogadási tartományba, ezért a nullhipotézist elfogadjuk, azaz az első betanítási módszer nem jobb, mint a második.

13.3.2. Két sokasági szórás egyezőségére irányuló próba

Ha a két sokaság normális eloszlású, a szórások egyezőségének vizsgálatára az F-próbafüggvény alkalmazható, ezért F-próbának nevezzük.

$$H_0: \sigma_1 = \sigma_2,$$

a szabadságfokok az alábbiak: $szf_1 = n_1 - 1$; $szf_2 = n_2 - 1$.

$$F = \frac{\sigma_1^2}{\sigma_2^2} \text{ vagy } F = \frac{s_1^2}{s_2^2}$$

Az F-eloszlás nem szimmetrikus, elfogadási tartománya a következő:

Alternatív hipotézis

elfogadási tartomány

$$H_1: \sigma_1 < \sigma_2$$

$$\left[F_{szf_2(\alpha)}^{szf_1}; \infty \right]$$

$$H_1: \sigma_1 \neq \sigma_2$$

$$\left[F_{szf_2(\alpha/2)}^{szf_1}; F_{szf_2(1-\alpha/2)}^{szf_1} \right]$$

$$H_1: \sigma_1 > \sigma_2$$

$$\left[0; F_{szf_2(1-\alpha)}^{szf_1} \right]$$

$$F_{szf_1(\alpha)}^{szf_2} = \frac{1}{F_{szf_1(1-\alpha)}^{szf_2}}$$

Bemutató feladat:

Az előző példa folytatása. Ellenőrizzük 10%-os szignifikancia szinten azt a feltevést, hogy a munkások teljesítményének szórása megegyezik.

$$H_0: \sigma_1 = \sigma_2, H_1: \sigma_1 \neq \sigma_2$$

$$F = \frac{18^2}{19^2} = 0,8976$$

$$szf_1 = 15; szf_2 = 10$$

$$\left[F_{10(0,05)}^{15} = 0,351; F_{10(0,95)}^{15} = 2,85 \right]$$

A számított érték belesik az elfogadási tartományba, így a nullhipotézist elfogadjuk, azaz a teljesítmények szórása között nincs szignifikáns különbség.

13.3.3. Két eloszlás egyezőségének a vizsgálata

Két eloszlás egyezőségének a vizsgálatát homogenitás vizsgálatnak is nevezzük.

Feltételezzük, hogy valamely változó két sokaságn belüli eloszlása azonos. Erre χ^2 -eloszlású próbafüggvényt alkalmazunk. A két minta elemszáma n_1 és n_2 , akkor χ^2 értéke:

$$\chi^2 = n_1 * n_2 * \sum \frac{1}{n_{1i} + n_{2i}} * \left(\frac{n_{1i}}{n_1} - \frac{n_{2i}}{n_2} \right)^2$$

$$\frac{n_{1i}}{n_1} = g_i \dots \dots \dots \text{relatívgyakoriság}$$

Bemutató feladat:

A virágárákat vizsgálva a pesti és budai virágüzletekben arra keresték a választ, hogy a virágárák eloszlása azonos-e a két helyen 99%-os konfidenciaszinten.

Ár	Buda	Pest	számítások			
	árusok száma (db)		$n_{1i}+n_{2i}$	g_{1i}	g_{2i}	$\frac{1}{n_{1i} + n_{2i}} * \left(\frac{n_{1i}}{n_1} - \frac{n_{2i}}{n_2} \right)^2$
0-55	3	5	8	0,0400	0,0595	0,000048
56-65	22	6	28	0,2933	0,0714	0,001759
66-75	18	18	36	0,2400	0,2143	0,000018
76-85	18	21	39	0,2400	0,2500	0,000003
86-95	6	18	24	0,0800	0,2143	0,000752
96-105	6	10	16	0,0800	0,1190	0,000095
105-	2	6	8	0,0267	0,0714	0,000250
összesen.	75	84	159	1	1	0,002925

$$\chi^2 = n_1 * n_2 * \sum \frac{1}{n_{1i} + n_{2i}} * \left(\frac{n_{1i}}{n_1} - \frac{n_{2i}}{n_2} \right)^2 = 75 * 84 * 0,002925 = 18,4275$$

szf=7-1=6 $\alpha/2=0,005$ $1-\alpha/2=0,995$

$$c_a=0,68 \quad c_f=18,5$$

Az elfogadási tartomány: (0,68; 18,5)

A számított érték beleesik az elfogadási tartományba, azaz a virágárák eloszlása azonos a két helyen.

13.3.4. Varianciaanalízis

Varianciaanalízissel (szórásnégyzet elemzés) kettőnél több sokaság várható értékének egyezősége tesztelhető.

A varianciaanalízis annak a nullhipotézisnek az ellenőrzésére szolgál, hogy kettőnél több, azonos szórású normális eloszlású valószínűségi változónak azonos-e a várható értéke.

$$H_0: \mu_1 = \mu_2 = \dots = \mu_n = \mu$$

A varianciaanalízis abból indul ki, hogy minden megfigyelés 3 komponens összege:

- a várható érték: μ ,
- egy a j-edik sokaságra jellemző β_j konstans,
- és egy véletlen összetevő (hibaváltozó): ε_{ij} .

A H_0 helyességét próbafüggvénnyel vizsgáljuk, és ez az F-próbafüggvény.

$$F = \frac{s_k^2}{s_b^2} = \frac{SSK / (M - 1)}{SSB / (n - M)} = \frac{MS_k}{MS_H} = \frac{MQ_k}{MQ_H}$$

SSK: a csoportok közötti eltérés négyzetösszege (külső szórás négyzete)

$$SSK = \sum_{i=1}^m n_i (\bar{x}_i - \bar{x})^2$$

M: a csoportok száma

SSB: a csoportokon belüli eltérés négyzetösszege. (belső szórás négyzete)

$$SSB = \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

Ezen kívül ki kell számolni az összes adat szórásnégyzetét is.

$$SST = \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$$

Vagy:

$$SST = SSK + SSB \text{ (teljes szórás négyzete)}$$

A varianciatáblázat a következő lesz:

A szóródás oka	SS (SQ)	DF(FG)	MS(MQ)	F
Külső (kezelés)	SSK	M-1	s_k^2	s_k^2 / s_b^2
Belső (hiba)	SSB	n-M	s_b^2	
Teljes	SST	n-1		

Az F-próba jobboldali próba. Ha a tapasztalati F-érték (számított) nagyobb az elméleti F-értéknél (táblázatbeli), akkor a várható értékek egyezőségére vonatkozó nullhipotézist az adott szignifikanciaszint mellett elvetjük, és az alternatív hipotézist fogadjuk el.

Bemutató feladat:

Egy kis élelmiszerbolt tulajdonosa feltételezte, hogy a hétfői és szombati napokon nem ugyanannyi a sajt forgalma, mint a hét többi napján. Azért, hogy a sajtrendelést jobban le tudja adni, feljegyezte a forgalmat az adott napokon:

Napok	Megfigyelt napok száma	Eladott mennyiség (kg)	Átlag	Variancia (szórásnégyzet)
Hétfő	6	30,40,54,34,44,50	42	84,8
egyéb	10	49,43,30,59,35,46,42,35,36,43	41,8	70,4
szombat	6	52,58,57,70,54,54	57,50	42,3
Össz.	22		46,136	111,17

Ellenőrizzük 5%-os szignifikancia szinten, hogy a sajtforgalom azonos a megfigyelt napokon.

$$SST = 21 \cdot 111,17 = 2334,57$$

$$SSB = 5 \cdot 84,8 + 9 \cdot 70,4 + 5 \cdot 42,3 = 1269,1$$

$$SSK = 2334,57 - 1269,1 = 1065,47$$

$$s_k^2 = SSK / (M-1) = 1065,47 / 2 = 532,735$$

$$s_b^2 = SSB / (n-M) = 1269,1 / 19 = 66,795$$

$$F = 532,735 / 66,795 = 7,976$$

A sajtforgalom varianciánálisis táblázata

A szóródás oka	SS (SQ)	DF(FG)	MS(MQ)	F
Milyen nap	1065,47	2	532,735	7,976
Hiba	1269,1	19	66,795	
Teljes	2334,57	21		

Mivel $F_{0,95}(2;19)=3,52$, azaz a táblázatbeli érték kisebb, mint a számított F-érték, így a várható értékek egyezését állító nullhipotézis elvethető. A hét vizsgált napjain 5%-os szignifikancia szinten nem egyforma a sajtforgalom átlagos nagysága.

V. modul: Korreláció- és Regresszió analízis

14. lecke Kétfváltozós korreláció- és regresszió-számítás: Lineáris

Követelmények:

Ön akkor sajátította el megfelelően a tananyagot, ha

- Tudja a korreláció és a regresszió jelentését.
- Tudja a korrelációs együttható által felvehető értékeket.
- A felsorolásból ki tudja választani a regressziós paraméterek, a relatív hiba és a rugalmassági együttható jelentését.
- A megadott adatokból ki tudja számolni a regressziós függvény paramétereit, a korrelációs együtthatót, a determinációs együtthatót, a relatív hibát, a rugalmassági együtthatót.
- Értelmezni tudja a kiszámolt értékeket.
- Becslés tud végezni a regressziós paraméterek ismeretében.
- Ki tudja számolni a regressziós paraméterek konfidencia intervallumát.
- Tudja a hatványkitevős és az exponenciális regresszió jelentését.
- Tudja a nem lineáris regresszió szorosságának mérésére szolgáló mutatót.
- Az adatok alapján ki tudja számolni a hatványkitevős és az exponenciális regressziót.
- Az adatok alapján el tudja dönteni hogy milyen típusú regressziós függvény illeszkedik a legjobban.

14.1. Kétfváltozós korreláció-számítás

A mennyiségi ismérvek között meglévő kapcsolat szorosságát és irányát a korreláció-számítással állapíthatjuk meg.

Ha a mennyiségi ismérvek között nincs kapcsolat, azaz függetlenek egymástól, akkor a szorosság mérőszáma a korrelációs együttható (r) nullával egyenlő.

Ha egyértelmű kapcsolat van a két mennyiségi ismerv között, akkor a korrelációs együttható (r) értéke $(-1;1)$ közé esik. Ha a korrelációs együttható pozitív, akkor a két ismerv közötti kapcsolat azt jelenti, hogy az egyik ismerv növekedése maga után vonja a másik ismerv növekedését. Negatív kapcsolat esetén az egyik ismerv növekedése a másik ismerv csökkenését okozza.

A mennyiségi ismérvek eloszlásainak speciális paramétere a kovariancia, amely az átlagtól való eltérések szorzatának számtani átlaga. Az együttes szóródás nagyságrendjét jellemzi, az előjele pedig kifejezi a kapcsolat irányát.

$$C = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{n}$$

A kovariancia felhasználásával kiszámítható a lineáris korrelációs együttható:

$$r = \frac{C}{\sigma_x * \sigma_y}$$

ahol: C : a kovariancia,

σ_x : az egyik változó szórása

σ_y : a másik változó szórása

Ha a változók szórására nincs külön szükségünk, akkor a korrelációs együttható másképp is kiszámolható:

$$r = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 * \sum (y_i - \bar{y})^2}}$$

Vagy:

$$r = \frac{\sum x_i * y_i - n * \bar{x} * \bar{y}}{\sqrt{(\sum x_i^2 - n * \bar{x}^2) * (\sum y_i^2 - n * \bar{y}^2)}} = \beta_1 * \frac{\sigma_x}{\sigma_y}$$

A korrelációs együttható négyzete a determinációs együttható, amely százalékban adunk meg, és azt fejezi ki, hogy az egyik ismerv hány százalékban befolyásolja a másik ismerv változását.

$$R = r^2$$

14.2. Kétváltozós lineáris regresszió-analízis

Ha két mennyiségi változó közötti függőségi viszonyt valamilyen matematikai képlettel írunk le, akkor regresszió-analízisről beszélünk.

Az ismérvek közötti függőségi viszonyok feltárásával, az összefüggésekben rejlő tendenciák matematikai függvényekkel történő leírásával a **regresszióanalízis** foglalkozik.

A gyakorlati elemző munkában a korreláció- és regresszió-számítást általában együtt, egymást kiegészítve alkalmazzák.

A regresszió-számítás a statisztikai modellezés egyik egyszerű eszköze, ám egyszerűsége ellenére szinte minden, korábban megismert módszertani elemet felhasznál.

A regresszió-számításkor általában meg szoktuk különböztetni a két- a többváltozós eseteket.

A kétváltozós regresszióanalíziskor két változó kapcsolatát vizsgáljuk, az x változó az egyik, ez a magyarázó változó, és az y-lal jelölt (eredményváltozó) változó a másik, amelynek alakulását x változó befolyásolja. A regresszió-számítás során feltételezzük, hogy az eredményváltozónk (y) sztochasztikus kapcsolatban áll a magyarázó változóval $Y=f(x)$

A regresszió típusának kiválasztásához először ábrázolni kell az adatokat, mivel az ismérvek közötti kapcsolat lényegének megismerésében fontos szerepet játszik a grafikus ábrázolás. Általában pontdiagramot készítünk.

Ezután meg kell határozni a regresszió típusát, ehhez azonban szükséges az adott terület szakmai ismerete is. Lineáris esetben az alábbi függvényt használjuk:

$$y = \beta_0 + \beta_1 * x$$

A függvénytípus kiválasztásával azonban a regressziós függvény meghatározásának problémája még nincs megoldva. A végtelen sok egyenes közül azt az egyet keressük, amely az összefüggést a lehető legjobban leírja. A függvény paramétereit a legkisebb négyzetek módszere segítségével határozzuk meg, vagyis:

$$S = \sum (y_i - \hat{y}_i)^2 \Rightarrow \text{minimum.}$$

Azt a becslőfüggvényt keressük tehát, amelyik a mintabeli és a számított értékek közötti különbségek négyzetösszege minimális. Lineáris összefüggés esetén a függvényünk:

$$y = \beta_0 + \beta_1 * x \quad \text{vagy} \quad y = a + b * x$$

Ezt behelyettesítve S-egyenletébe a következőt kapjuk:

$$S = \sum (y_i - \beta_0 - \beta_1 * x)^2$$

A függvénynek ott van minimuma, ahol a két együttható szerinti parciális differenciáhányadosa egyenlő nullával. Az egyenlet levezetéséből azt kapjuk, hogy:

$$\beta_1 = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i * y_i - n * \bar{x} * \bar{y}}{\sum x_i^2 - n * \bar{x}^2}$$

$$\beta_0 = \bar{y} - \beta_1 * \bar{x}$$

A β_0 paraméter azt fejezi ki, hogy az $x=0$ helyen a függvény éppen ezt az értéket veszi fel, ha a nulla szerepel x lehetséges értékei között.

A β_1 paraméter geometriai értelemben az egyenes meredekségét meghatározó iránytangens, regressziós együtthatóként választ ad arra, hogy az x változó egységnyi változása átlagosan mekkora változást okoz az y változóban.

A kétváltozós kapcsolatok esetében megvizsgáljuk, hogy a becslőfüggvény mennyire közelítette meg a mintabeli tapasztalati értéket. Ezt fejezi ki az ε - érték, azaz a reziduális szórás, amelyet a regressziós becslés abszolút hibája is egyben:

$$\sigma_e = \sqrt{\frac{\sum (y_i - y'_i)^2}{n - 2}} = \sqrt{\frac{\sum e_i^2}{n - 2}}$$

Ami tulajdonképpen az x_i pontban vett mintabeli y_i értékek és az adott x_i ponthoz tartozó becsült y'_i értékek közötti eltérések négyzetösszegének a gyöke.

A gyakorlatban kiszámítjuk ennek relatív nagyságát is:

$$CV_{\sigma_e} = \frac{\sigma_e}{\bar{y}} * 100$$

A relatív hiba kifejezi, hogy a regressziós becslések értékei átlagosan hány százalékkal térnek el az eredményváltozó megfigyelt értékeitől. Általában jónak mondjuk a becslőfüggvény illeszkedését, ha: $CV_{\sigma_e} < 10-15\%$.

Az eredményváltozó relatív változásának fontos szerepe van a közgazdasági elemzésekben. A relatív változást fejezi ki a rugalmassági együttható:

$$E = \frac{dy}{dx} * \frac{x_0}{y_0}$$

Az x -magyarázóváltozó adott értékének 1%-os növekedése átlagosan milyen változást eredményez az y -változó értékében. Ez az érték természetesen minden x -értékre kiszámítható:

$$E = \beta_1 * \frac{x_i}{y_i}$$

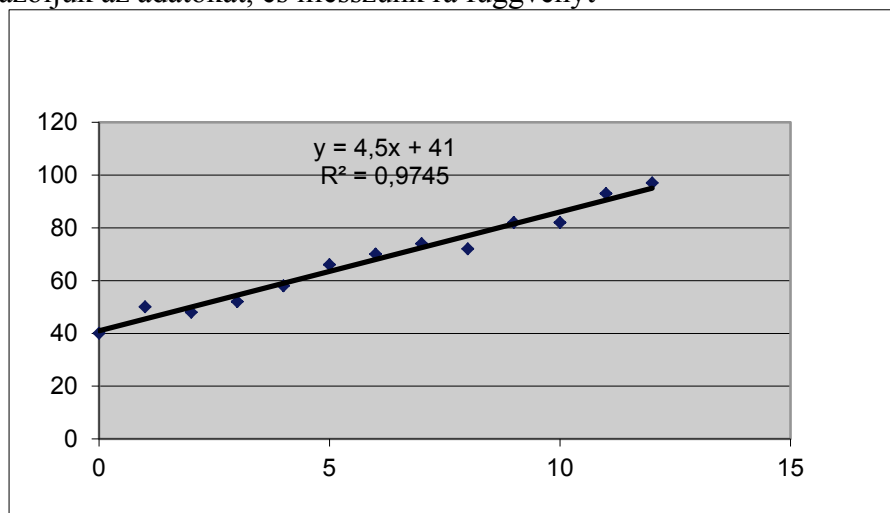
Bemutató feladat:

Egy 60 fős sokaságra a statisztika írásbeli dolgozatok eredményeit és a felkészülésre fordított idő nagyságát vizsgálva 13 elemű mintát vettek, az alábbi eredményekkel:

	Idő (óra) x változó	Dolgozat eredménye (pont) y váltó	x^2	$x - \bar{x}$	$(x - \bar{x})^2$	$y - \bar{y}$	$(y - \bar{y})^2$	$(x - \bar{x}) * (y - \bar{y})$
	0	40	0	-6	36	-28	784	168
	1	50	1	-5	25	-18	324	90
	2	48	4	-4	16	-20	400	80
	3	52	9	-3	9	-16	256	48
	4	58	16	-2	4	-10	100	20
	5	66	25	-1	1	-2	4	2
	6	70	36	0	0	2	4	0
	7	74	49	1	1	6	36	6
	8	72	64	2	4	4	16	8
	9	82	81	3	9	14	196	42
	10	82	100	4	16	14	196	56
	11	93	121	5	25	25	625	125
	12	97	144	6	36	29	841	174
össz.	78	884	650	0	182	0	3782	819
átlag	6	68						

Feladat: számítsuk ki a regressziós függvény paramétereit!

Először ábrázoljuk az adatokat, és illesszünk rá függvényt



$$\beta_1 = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{819}{182} = 4,5$$

$$\beta_0 = \bar{y} - \beta_1 * \bar{x} = 68 - 4,5 * 6 = 41$$

$$y' = 41 + 4,5x$$

β_1 : azok a hallgatók, akik 1 órával többet tanultak, átlagosan 4,5 ponttal jobb eredményt értek el.

β_0 : akik nem készültek a dolgozatra, azok átlagosan 41 pontos eredményre számíthatnak.

Ezután összehasonlítjuk a tényleges és a regresszióval becsült adatokat.

Az y' -t úgy kapom meg, hogy az egyenletbe ($y'=41+4,5x$) rendre behelyettesítem az x -értékeket. Ezután minden eredeti y -ból kivonom a kiszámított y' -értéket, majd az eredményt négyzetre emelem.

Idő (óra) x változó	Dolgozat eredménye (pont) y változó	y'	$y-y'$	$(y-y')^2$
0	40	41	-1	1
1	50	45,5	4,5	20,25
2	48	50	-2	4
3	52	54,5	-2,5	6,25
4	58	59	-1	1
5	66	63,5	2,5	6,25
6	70	68	2	4
7	74	72,5	1,5	2,25
8	72	77	-5	25
9	82	81,5	0,5	0,25
10	82	86	-4	16
11	93	90,5	2,5	6,25
12	97	95	2	4
Összesen				96,5

$$\sigma_e = \sqrt{\frac{\sum (y_i - y'_i)^2}{n - 2}} = \sqrt{\frac{96,5}{13 - 2}} = 2,96$$

$$CV_{\sigma_e} = \frac{\sigma_e}{\bar{y}} * 100 = \frac{2,96}{68} * 100 = 4,35\%$$

A becslőfüggvény illeszkedése jónak mondható.

$$r = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 * \sum (y_i - \bar{y})^2}} = \frac{819}{\sqrt{182 * 3782}} = 0,987$$

$$R^2 = 97,45\%$$

A tanulásra fordított idő 97,45%-ban meghatározza az elért pontszámot.

15. lecke Kétváltozós korreláció- és regresszió-számítás: exponenciális és hatványkitevős függvények

Követelmények

- Ön akkor sajátította el megfelelően a tananyagot, ha
- tudja a hatványkitevős és az exponenciális regresszió jelentését,
- tudja a nem lineáris regresszió szorosságának mérésére szolgáló mutatót,
- az adatok alapján ki tudja számolni a hatványkitevős és az exponenciális regressziót,
- az adatok alapján el tudja dönteni hogy milyen típusú regressziós függvény illeszkedik a legjobban.

Tananyag

Két mennyiségi ismerv közötti kapcsolat leírására sok esetben nem alkalmas a lineáris függvény. Ha az x-változó y-változóra gyakorolt hatásának mértéke függ az x-változó nagyságától, akkor a lineáris regresszió nem alkalmas az adatok közötti kapcsolat elemzésére.

15.1. Hatványkitevős regressziós függvény

Két mennyiségi ismerv közötti kapcsolat leírására sok esetben nem alkalmas a lineáris függvény. Ha az x-változó y-változóra gyakorolt hatásának mértéke függ az x-változó nagyságától, akkor a lineáris regresszió nem alkalmas az adatok közötti kapcsolat elemzésére.

Ha a független változó szorzatos növekedésével a függő változó is szorzatosan változik, akkor regressziós függvény az alábbi:

$$y = \beta_0 * x^{\beta_1}$$

Olyan esetekben alkalmazzuk, amikor az x és y változók logaritmusai között van lineáris összefüggés.

β_1 regressziós együttható azt fejezi ki, hogy az x magyarázó változó egységnyi relatív (1%-os) változása mekkora relatív (hány százalékos) változást idéz elő az eseményváltozóban.

Megoldásához linearizálni kell a regressziós függvényt:

$$\lg y = \lg \beta_0 + \beta_1 * \lg x$$

Látható, hogy az x és az y változók logaritmusai között lineáris a kapcsolat.

$$\beta_1 = \frac{\sum (\lg x_i - \overline{\lg x}) * (\lg y_i - \overline{\lg y})}{\sum (\lg x_i - \overline{\lg x})^2}$$

$$\lg \beta_0 = \overline{\lg y} - \beta_1 * \overline{\lg x}$$

Vezessünk be új ismeretleneket:

$$\lg y = Y; \lg x = X; \lg \beta_0 = B$$

Így a függvényünk az alábbi:

$$Y = B + \beta_1 * X$$

A regressziós együtthatók így már a tanultak szerint számíthatóak:

$$\beta_1 = \frac{\sum (X_i - \bar{X}) * (Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$B = \bar{Y} - \beta_1 * \bar{X}$$

$$\beta_0 = 10^{B_0}$$

A paraméterek jelentése:

- A β_0 : ha az $x=1$ része a függvény értelmezési tartományának, akkor van jelentése, azaz az $x=1$ helyen felvett regressziós érték.
- A β_1 : a magyarázóváltozó 1%-os változása az eredményváltozásban éppen A β_1 %-os változást okoz. A β_1 -együttható elaszticitási (rugalmassági) együttható is egyben, mivel megmutatja hogy az 1%-kal nagyobb x -értékhez hány százalékkal nagyobb vagy kisebb y -érték tartozik.

A kapcsolat szorosságát a korrelációs index fejezi ki:

$$I = \sqrt{1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}}$$

$$e_i^2 = (y_i - \hat{y}_i)^2$$

A korrelációs index értéke:

$$0 \leq I \leq 1$$

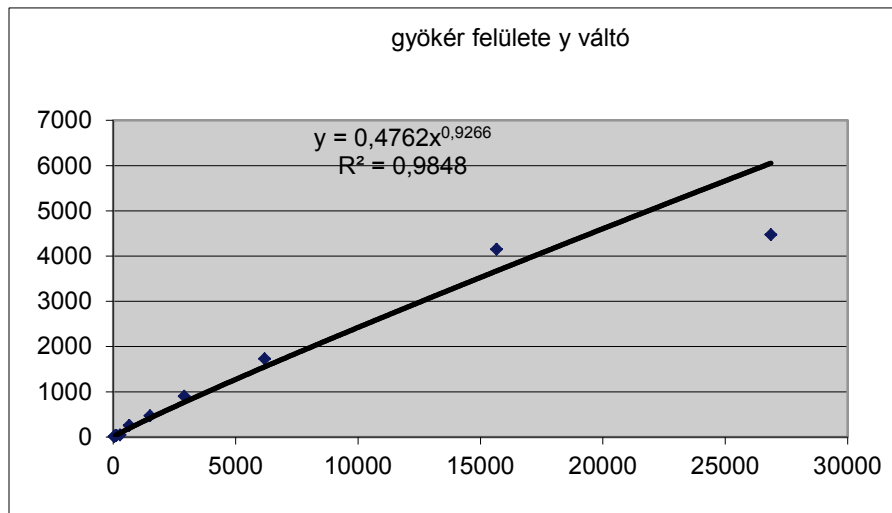
A korrelációs index négyzetét százalékban fejezzük ki.

Bemutató feladat:

Nézzük meg, hogy mennyire van hatással a gyökér súlya a gyökér felületére cukorrépa esetében.

Először az eredeti adatok logaritmusát kell meghatározni. A továbbiakban ezekkel az értékekkel dolgozunk, azaz X és Y -értékei kerülnek a képletekbe.

	gyökér súlya x változó	gyökér felülete y váltó	$\lg x = X$	$\lg y = Y$	X^2	$X - X_{\text{átlag}}$	$(X - X_{\text{átlag}})^2$	$(Y - Y_{\text{átlag}})$	$(Y - Y_{\text{átlag}})^2$	$(X - X_{\text{átlag}}) * (Y - Y_{\text{átlag}})$
	26870	4472	4,429268	3,650502	19,61841	1,461268	2,135303	1,222502	1,494511	1,786402
	15660	4152	4,194792	3,618257	17,59628	1,226792	1,505018	1,190257	1,416713	1,460198
	6180	1728	3,790988	3,237544	14,37159	0,822988	0,67731	0,809544	0,655361	0,666245
	2900	904	3,462398	2,956168	11,9882	0,494398	0,244429	0,528168	0,278962	0,261125
	1500	472	3,176091	2,673942	10,08756	0,208091	0,043302	0,245942	0,060487	0,051178
	650	260	2,812913	2,414973	7,912482	-0,15509	0,024052	-0,01303	0,00017	0,00202
	280	48	2,447158	1,681241	5,988582	-0,52084	0,271276	-0,74676	0,557649	0,388943
	130	39	2,113943	1,591065	4,468756	-0,85406	0,729413	-0,83694	0,700461	0,71479
	60	24	1,778151	1,380211	3,161822	-1,18985	1,41574	-1,04779	1,097861	1,24671
	30	12	1,477121	1,079181	2,181887	-1,49088	2,222719	-1,34882	1,819312	2,010925
össz.	54260	12111	29,68282	24,28308	97,37557	0,002824	9,268563	0,003085	8,081486	8,588538
átlag	5426	1211,1	2,968282	2,428308						



$$\beta_1 = \frac{\sum (X_i - \bar{X}) * (Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{8,5885}{9,2686} = 0,9266$$

$$B_0 = \bar{Y} - \beta_1 * \bar{X} = 2,4283 - 0,9266 * 2,9683 = -0,3222$$

$$\beta_0 = 10^{-0,3222} = 0,4762$$

$$\hat{y} = 0,4762 * x^{0,9266}$$

15.2. Exponenciális regressziós függvény

Ha az adatok közötti összefüggést a:

$$\hat{y} = \beta_0 * \beta_1^x$$

Függvénnyel írható le, akkor exponenciális regresszióról beszélünk. Olyan esetekben alkalmazzuk, ha az y-változó növekedése arányos az adott helyen felvett x-változó értékével. Az exponenciális függvények esetében is igaz, hogy lineáris összefüggés van az eredményváltozó logaritmusa és a magyarázóváltozó között. Hasonlóan a hatványkitevős regresszióhoz, ebben az esetben is visszavezetjük lineáris regresszióra:

$$\lg \hat{y} = \lg \beta_0 + x * \lg \beta_1$$

$$Y = B_0 + B_1 * x$$

$$B_1 = \frac{\sum (x_i - \bar{x}) * (Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2}$$

$$\beta_1 = 10^{B_1}$$

$$B_0 = \bar{Y} - B_1 * \bar{x}$$

$$\beta_0 = 10^{B_0}$$

A paraméterek jelentése:

A β_1 regressziós paraméter arra ad választ, hogy az x-változó egységnyi növekedése hányszorosára változtatja az y-változó értékét.

A kapcsolat szorosságát a korrelációs index fejezi ki:

$$I = \sqrt{1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}}$$

$$e_i^2 = (y_i - \hat{y}_i)^2$$

A korrelációs index értéke:

$$0 \leq I \leq 1$$

15.3. Választás a különböző regressziós egyenlet-típusok közül

Ugyanarra az adatsorra kiszámolva mindhárom regressziós függvényt, felvetődik a kérdés, hogy melyik jellemzi legjobban a változók kapcsolatát. A függvények kiválasztáshoz az egyenletek illeszkedési módszerét, azaz a legkisebb eltérések-négyzetét használjuk. Az az egyenlet illeszkedik legjobban az adatokra, ahol az $(y_i - \hat{y}_i)^2$ és az $(x_i - \hat{x}_i)^2$ is a legkisebb, illetve ahol a kapcsolat szorosságát kifejező mutató a legnagyobb.

VI. modul: Idősorok elemzése

16. lecke: Egyszerű elemzési módszerek

Követelmények:

Ön akkor sajátította el megfelelően a tananyagot, ha

- Tudja mit jelent az additív és a multiplikatív kapcsolat.
- Tud páratlan számú mozgóátlagolással trendet számolni.
- Tud páros számú mozgóátlagolással trendet számolni.

A jelenségek időbeli alakulásának, lefolyásának vizsgálata minden tudományágnak, így a gazdasági tudományoknak is egyik központi kérdése. A statisztikai soroknál és a viszonyszámoknál már találkoztunk idősorokkal, hisz a dinamikus viszonyszámokat ezekből határoztuk meg. Az idősor elemzése abból indul ki, hogy az idősort tartósan érvényesülő hosszú távú tendencia (trend), tartósan ható szabályok, jól modellezhető hullámmozgás (szezonalitás) határozza meg. A statisztikai elemzés szempontjából az idősornak három összetevője van.

- **alapirányzat (trend):** ez a legfontosabb összetevő. A trend az idősorban hosszabb időszakon át, tartósan érvényesülő tendencia. Ez is több tényező együttes hatásának a következménye.

- **Periodikus ingadozás:** az idősorokban rendszeresen ismétlődő hullámmozgást jelent. Két típusát különböztetjük meg, a szezonális vagy idényszerű hullámmozgást, és a konjunkturális ingadozást. A szezonális hullámmozgás a trendtől való abszolút vagy relatív mértékű eltérés-periodicitást mutat.

- **véletlenszerű ingadozás:** ezt az összetevőt valószínűségi változónak tekintjük.

Az idősorok elemzésének alapvető feladata a komponensek elkülönítése. Az egyes komponensek között lehet additív (összegszerű) és multiplikatív (szorzatszerű), vagy ennél bonyolultabb kapcsolat.

Additív kapcsolat: a komponensek összege adja az idősort.

$$\hat{Y}_{ij} = \hat{Y}_{ij} + \hat{S}_j + \hat{V}_{ij} \quad s_j: \text{ szezonális ingadozás}$$

A szezonális eltérés a periódusok különböző szakaszában más és más mértékű és irányú lehet. A véletlen komponens (v_{ij}) valószínűségi változó, additív kapcsolat esetén feltételezzük, hogy a várható értéke nullával egyenlő.

Multiplikatív kapcsolat: a komponensek szorzata adja az idősort.

$$\hat{Y}_{ij} = \hat{Y}_{ij}^* \cdot \hat{S}_j^* \cdot \hat{V}_{ij}^*$$

A szezonális komponens relatív módon fejezi ki a hatását, ez a szezonális ingadozás (s_j^*).

A trendszámítás során az idősor kiegyenlítése, kisimítása a célunk úgy, hogy a periodikus ingadozás és a véletlen ingadozás hatását kiszűrjük. Ez többféle módon történhet, pl: átlagolással, mozgóátlagolással, vagy analitikus trendszámítással.

16.1. Egyszerű elemzési módszerek

Átlagolás

A flow (tartam) jellegű idősor adatait egyszerű számtani átlaggal számoljuk ki.

$$\bar{y} = \frac{\sum y_i}{n}$$

A stock (állapot) jellegű idősor adatainál a számtani átlag nem használható, ekkor a kronológikus átlagot használjuk:

$$\bar{y} = \frac{\frac{y_1}{n} + y_2 + \dots + \frac{y_n}{2}}{n-1}$$

Átlagos változás mutatói:

- a változás átlagos mértéke: Az időegységre jutó átlagos változást adja meg.

$$\bar{d} = \frac{y_n - y_1}{n-1}$$

- a változás átlagos üteme: Ez viszonyyszámként adja meg a változás ütemét, fejlődést tükröz:

$$\bar{l} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

16.2. Mozgóátlagolásos trendszámítás

Általában akkor alkalmazzuk, ha nincs kellő ismeretünk a vizsgálandó folyamatról. Ez egy egyszerű idősor-elemző módszer. A mozgóátlagolásos trendszámítás során a trendet az idősor dinamikus átlagolásával határozzuk meg úgy, hogy az idősor minden eleméhez kiszámítjuk annak környezetében lévő elemek átlagát. Az átlagolással kiküszöböljük a szezonális komponenst és csökkentjük a véletlen tényező hatását.

A mozgóátlagolásnál először az átlagolás tagszámát kell meghatározni. Az időszak mindig k-nyi elemét adjuk összes és ebből a k-val való osztás során kapunk egy átlagot. Fontos, hogy 'n' egy egész periódust fogjon át. Egy periódus azt az időszakot jelenti, amely alatt ugyanaz a szezonhatás nem ismétlődik.

Ha a szezonális komponens periódusának hossza páratlan, akkor a 'n' is páratlan lesz ($n=2k+1$), így a páratlan tagszámú mozgóátlagolással kiszámított trendet az alábbi képlettel kapjuk meg:

$$\hat{y}_i = \frac{y_{i-k} + y_{i-k+1} + \dots + y_i + \dots + y_{i+k}}{2k+1}$$

Például egy 3-tagú mozgóátlagolás a következő:

Időszak	idősor adatai	mozgó összeg ($\ddot{O}_i$)	Mozgóátlag $\hat{y}_i$
1	y_1	-	-
2	y_2	$\ddot{O}_2 = y_1 + y_2 + y_3$	$\ddot{O}_2/3$
3	y_3	$\ddot{O}_3 = y_2 + y_3 + y_4$	$\ddot{O}_3/3$
4	y_4	$\ddot{O}_4 = y_3 + y_4 + y_5$	$\ddot{O}_4/3$
.	.	.	.
.	.	.	.
n-2	y_{n-2}	$\ddot{O}_{n-2} = y_{n-3} + y_{n-2} + y_{n-1}$	$\ddot{O}_{n-2}/3$
n-1	y_{n-1}	$\ddot{O}_{n-1} = y_{n-2} + y_{n-1} + y_n$	$\ddot{O}_{n-1}/3$
n	y_n	-	-

Ha a szezonális komponens periódusának hossza páros, akkor a mozgóátlagolással nem kapunk azonnal trendértéket, mivel a kiszámított átlag nem esik időponthoz. Ekkor az úgynevezett centrírozásra van szükség, amely kéttagú mozgóátlagolást jelent, és így már időponthoz tudjuk rendelni az értékeket.

Pl.: negyedéves adatok, akkor páros számú mozgóátlagolást végzünk.

Időszak	Idősor adatai	mozgó összeg ($\ddot{O}_i$)	Mozgóátlag	Centrírozás $\hat{y}_i$
1	y_1	-		-
2	y_2	$\ddot{O}_{2,5} = y_1 + y_2 + y_3 + y_4$	$\ddot{O}_{2,5}/4$	-
3	y_3	$\ddot{O}_{3,5} = y_2 + y_3 + y_4 + y_5$	$\ddot{O}_{3,5}/4$	$\ddot{O}_{2,5} + \ddot{O}_{3,5}/2$
4	y_4	$\ddot{O}_{4,5} = y_3 + y_4 + y_5 + y_6$	$\ddot{O}_{4,5}/4$	$\ddot{O}_{3,5} + \ddot{O}_{4,5}/2$
.	.	.	.	
.	.	.	.	
n-2	y_{n-2}	$\ddot{O}_{n-2,5} = y_{n-4} + y_{n-3} + y_{n-2} + y_{n-1}$	$\ddot{O}_{n-2,5}/4$	$\ddot{O}_{n-2,5} + \ddot{O}_{n-1,5}/2$
n-1	y_{n-1}	$\ddot{O}_{n-1,5} = y_{n-3} + y_{n-2} + y_{n-1} + y_n$	$\ddot{O}_{n-1,5}/4$	-
n	y_n	-	-	-

A mozgóátlagolással kiegyenlített idősor rövidebb, mint az eredeti idősor.

A módszer előnye, hogy egyszerű és gyors. Hátránya, hogy rövidebb lesz a kiegyenlített idősor, és nem kapunk analitikusan ismert trendfüggvényt, illetve nem tudunk előrejelzést végezni.

17. lecke: Analitikus trendszámítás**Követelmények:****Ön akkor sajátította el megfelelően a tananyagot, ha**

- Tudja, hogy mit jelent a trendszámítás.
- Ki tudja választani a lineáris és az exponenciális trend jellemzőjét.
- Tudja a lineáris és az exponenciális trend paramétereinek jelentését.
- Tudja milyen esetekben végezhetünk extrapolációt.
- Az adatok alapján meg tudja határozni a lineáris és az exponenciális trendfüggvény paramétereit
- El tudja dönteni, hogy milyen függvény írja le jobban az adatokat.

Ha a vizsgált jelenség tartós irányzatát az idő függvényében valamilyen regressziós függvénnyel határozzuk meg, akkor analitikus trendszámításról beszélünk. Az analitikus trendszámítás tehát a regresszió-számítás egy speciális esete, amennyiben az idősorban bekövetkezett változásokat az időtényező (t) függvényében vizsgáljuk.

Először meg kell határozni, hogy a tartós alapirányzat milyen típusú függvénnyel jellemezhető a legjobban, azaz ábrázolni kell az adatokat, és kiválasztani a függvény típusát.

Az idősor trendjét az a függvény fejezi ki a legjobban, amelynél a számított függvényértékek és az eredeti adatok közötti eltérés négyzetösszeg a legkisebb.

17.1. Lineáris trend

A lineáris trendet úgy értelmezzük, hogy a vizsgált jelenség egységnyi idő alatt azonos mértékben növekszik vagy csökken, azaz a tartós alapirányzat egyenletes abban az értelemben, hogy az abszolút változás állandó.

Ha az adatokra az egyenletes változás jellemző, akkor lineáris trenddel határozzuk meg az alapirányzat értékeit.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 * t$$

A függvény illesztésekor a legkisebb négyzetek módszerét alkalmazzuk, a normálegyenletek a következők:

$$\sum y_i = n * \hat{\beta}_0 + \hat{\beta}_1 * \sum t_i$$

$$\sum t_i * y_i = \hat{\beta}_0 * \sum t_i + \hat{\beta}_1 * \sum t_i^2$$

A megoldáshoz kódolnunk kell az idősor adatait. Ez többféleképpen lehetségesféleképpen történhet:

- Ha az idősort $t=1,2,3,\dots,n$ kódoljuk:

év	t_i
2009	1
2010	2
2011	3
2012	4
2013	5
2014	6
2015	7
2016	8

A normálegyenleteket átalakítva megkapjuk a trendfüggvény együtthatóit:

$$\hat{\beta}_1 = \frac{\sum t_i * y_i - \frac{(\sum t_i) * (\sum y_i)}{n}}{\sum t_i^2 - \frac{(\sum t_i)^2}{n}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 * \frac{\sum t_i}{n}$$

- Ha az idősort $\Sigma t=0$ módon kódoljuk, akkor különbség van a páros és páratlan számú idősor esetében.

- *Páratlan tagszámú időssor:*

év	t
2009	-3
2010	-2
2011	-1
2012	0
2013	1
2014	2
2015	3

- *Páros tagszámú idősor esetén:* kétféleképpen kódolhatunk:

év	t	t
2009	-3,5	-7
2010	-2,5	-5
2011	-1,5	-3
2012	-0,5	-1
2013	0,5	1
2014	1,5	3
2015	2,5	5
2016	3,5	7

Ebben az esetben a normálegyenletek egyszerűbbek, és a trendfüggvény együtthatóit az alábbi képletekkel kapjuk meg:

$$\sum y_i = n * \hat{\beta}_0$$

$$\sum t_i * y_i = \hat{\beta}_1 * \sum t_i^2$$

$$\hat{\beta}_0 = \frac{\sum y_i}{n} = \bar{y}$$

$$\hat{\beta}_1 = \frac{\sum t_i * y_i}{\sum t_i^2}$$

A $\hat{\beta}_1$ paraméter az időegységenkénti átlagos abszolút változás mértéke, előjelétől függően növekedést vagy csökkenést jelez.

Ha $\Sigma t=0$, és az időpontok száma páros, akkor $2 \cdot \hat{\beta}_1$ az időegységenkénti átlagos abszolút változás mértéke.

A $\hat{\beta}_0$ paraméter az alapirányzat értéke a $t=0$ időpontban:

- ha $t=1, \dots, n$, akkor a vizsgálatba bevont időpontot megelőző időpont trend szerinti értéke.
- Ha $\Sigma t=0$, és páratlan az idősornál választott tagszám, akkor a középső időpont alapirányzata és egyben a vizsgált idősor adatainak számtani átlaga.
- Ha $\Sigma t=0$, és páros az idősornál választott tagszám, és nincs $t=0$, akkor az idősor adatainak számtani átlaga.

Ezután a regresszió számításához hasonlóan, kiszámoljuk a reziduális szórásnégyzetet.

$$s_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n},$$

és a reziduális szórás mutatószámát:

$$V_e = \frac{s_e}{\bar{y}} * 100$$

Ez a valós értékektől való átlagos eltérését mutatja a trendértékeknek.

Bemutató feladat:

A villamosenergia-termelés Magyarországon.

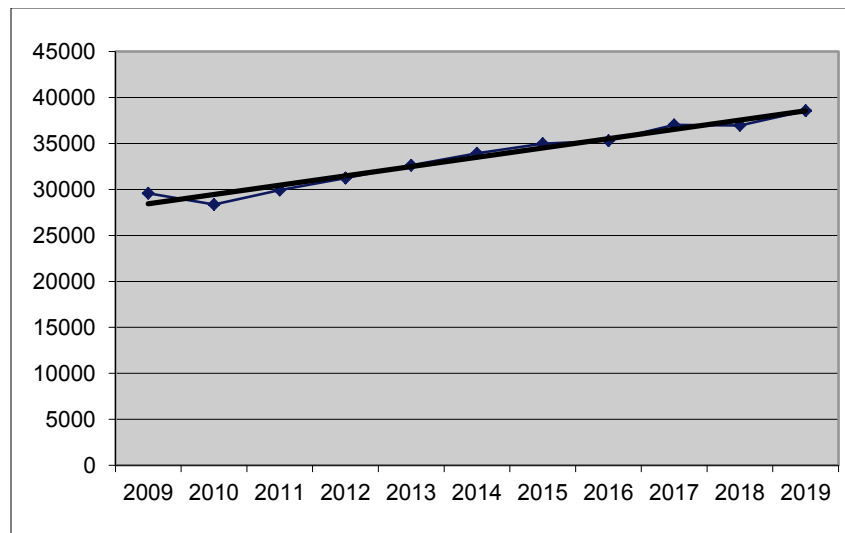
	millió kWh	t_i	t^2	$t \cdot y$	y_t	$(y - y_t)^2$
2009	29580	-5	25	-147900	28451,18	1274234,59
2010	28365	-4	16	-113460	29461,20	1201650,06
2011	29932	-3	9	-89796	30471,22	290753,89
2012	31238	-2	4	-62476	31481,23	59162,78
2013	32630	-1	1	-32630	32491,25	19251,01
2014	33928	0	0	0	33501,27	182098,49
2015	34978	1	1	34978	34511,29	217820,09
2016	35305	2	4	70610	35521,31	46788,29
2017	37023	3	9	111069	36531,32	241745,29
2018	36968	4	16	147872	37541,34	328721,05
2019	38567	5	25	192835	38551,36	244,61
Összesen	368514	0	110	111102	368513,97	3862470,15
átlag	33501,27					
$\hat{\beta}_1$	1010,018	$\hat{\beta}_0$	33501,27			

Az y_t -értékeket úgy kapjuk meg, hogy a trendegyenletbe rendre behelyettesítjük a t -értékeket.

$$y_t = 33501,27 + 1010,018 \cdot (-5) = 28451,18$$

$$y_t = 33501,27 + 1010,018 \cdot (-4) = 29461,20$$

stb.



$$s_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n} = \frac{3862470,15}{11} = 351133,65$$

$$s_e = 592,565$$

$$V_e = \frac{s_e}{\bar{y}} * 100 = \frac{592,565}{33501,27} * 100 = 1,77\%$$

A függvény szerinti villamosenergia-temelés 1,77%-os hibával illeszkednek a tényleges adatokhoz.

Nézzük meg, ha másmódon kódoljuk az időt, mi lesz a trendegyenlet!

	millió kWh	t	t ²	t*y	yt	y-yt	(y-yt) ²
2009	29580	1	1	29580	28451,17	1128,83	1274257,17
2010	28365	2	4	56730	29461,19	-1096,19	1201632,52
2011	29932	3	9	89796	30471,21	-539,21	290747,424
2012	31238	4	16	124952	31481,23	-243,23	59160,8329
2013	32630	5	25	163150	32491,25	138,75	19251,5625
2014	33928	6	36	203568	33501,27	426,73	182098,493
2015	34978	7	49	244846	34511,29	466,71	217818,224
2016	35305	8	64	282440	35521,31	-216,31	46790,0161
2017	37023	9	81	333207	36531,33	491,67	241739,389
2018	36968	10	100	369680	37541,35	-573,35	328730,223
2019	38567	11	121	424237	38551,37	15,63	244,2969
Összesen	368514	66	506	2322186	368513,97		3862470,15

átlag 33501,27

$$\sum y_t = n * b_0 + b_1 * \sum t$$

$$\sum t * y_t = b_0 * \sum t + b_1 * \sum t^2$$

$$368514 = 11 * b_0 + b_1 * 66$$

$$2322186 = b_0 * 66 + b_1 * 506$$

$$b_1 = 1010,02 \quad b_0 = 27441,15$$

$$y_t = 27441,15 + 1010,02 * t$$

17.2. Exponenciális trend

A társadalmi-gazdasági folyamatok változó környezetben nem mindig mutatnak lineáris tendenciát. Ha az időegységenkénti relatív változás ingadozik egy állandó körül, akkor a tartós irányzatot exponenciális trenddel fejezzük ki. Az exponenciális trendegyenletet logaritmus segítségével visszavezetjük lineárisra:

$$\hat{y} = \hat{\beta}_0 * \hat{\beta}_1^t$$

$$\lg \hat{y} = \lg \hat{\beta}_0 + t * \hat{\beta}_1$$

$$Y = B_0 + t * B_1$$

A normálegyenlet:

$$\sum Y_i = n * B_0 + B_1 * \sum t_i$$

$$\sum t_i * Y_i = B_0 * \sum t_i + B_1 * \sum t_i^2$$

A $\sum t=0$ esetében a normálegyenletek leegyszerűsödnek:

$$\sum Y_i = n * B_0$$

$$\sum t_i * Y_i = B_1 * \sum t_i^2$$

$$B_0 = \frac{\sum Y_i}{n}$$

$$B_1 = \frac{\sum t_i * Y_i}{\sum t_i^2}$$

$$\hat{\beta}_0 = 10^{B_0}$$

$$\hat{\beta}_1 = 10^{B_1}$$

A β_1 paraméter a trendfüggvényre nézve az időszakonkénti állandó relatív változást mutatja, az idősorra nézve pedig az időszakonkénti relatív változások átlagát fejezi ki.

Bemutató feladat:

Exponenciális trend esetében először logaritmizálni kell az y_i értékeket. A továbbiakban ezekkel az értékekkel számolunk.

Egy üdülőövezet vendégeinek száma:

	efő	t	lgy	t*lgy	t ²	y _t	y-y _t	(y-y _t) ²
2009	136	-5	2,133539	-10,6677	25	119,7789	-16,2211	263,1228
2010	152	-4	2,181844	-8,72737	16	134,9909	-17,0091	289,3107
2011	157	-3	2,1959	-6,5877	9	152,1347	-4,86529	23,67109
2012	166	-2	2,220108	-4,44022	4	171,4558	5,455813	29,76589
2013	183	-1	2,262451	-2,26245	1	193,2307	10,2307	104,6672
2014	188	0	2,274158	0	0	217,771	29,771	886,3124
2015	185	1	2,267172	2,267172	1	245,4279	60,42792	3651,533
2016	225	2	2,352183	4,704365	4	276,5973	51,59726	2662,277
2017	341	3	2,532754	7,598263	9	311,7251	-29,2749	857,0189
2018	440	4	2,643453	10,57381	16	351,3142	-88,6858	7865,17

2019	450	5	2,653213	13,26606	25	395,9311	-54,0689	2923,445
	2623	0	25,71677	5,724238	110			19556,3
	238,4545	0	2,337888					

$$B_0 = \frac{\sum Y_i}{n} = \frac{25,71677}{11} = 2,338$$

$$B_1 = \frac{\sum t_i * Y_i}{\sum t_i^2} = \frac{5,724238}{110} = 0,052039$$

$$\hat{\beta}_0 = 10^{B_0} = 10^{2,338} = 217,771$$

$$\hat{\beta}_1 = 10^{B_1} = 10^{0,052039} = 1,127$$

$$s_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n} = \frac{19556,3}{11} = 1777,8455$$

$$s_e = 42,16$$

$$V_e = \frac{s_e}{\bar{y}} * 100 = \frac{42,16}{238,4545} * 100 = 17,68\%$$

A függvény szerinti létszámok 17,68%-os hibával illeszkednek a tényleges adatokhoz.

17.3. Választás a trendegyenletek közül

A trendszámításnál sok esetben nehéz eldönteni, hogyan melyik függvényt használjuk az elemzéshez. A kiválasztáshoz a legkisebb négyzetek módszerét használjuk. Az a függvény

illeszkedik a legjobban az idősor adataihoz, amelyiknél $\sum (y_i - \hat{y}_i)^2$ a legkisebb.

17.4. Extrapoláció

A trendegyenlet meghatározásával előrejelzést (extrapoláció) végezhetünk. Fontos, hogy előrejelzést csak abban az esetben végezhetünk, ha:

- Kellően hosszú idősorból határoztuk meg a trendegyenletet.
- Illetve, ha várhatóan nem következik be olyan változás, amely a vizsgált jelenséget nagymértékben megváltoztatja.

18. lecke: Szezonális ingadozások elemzése

Követelmények:

Ön akkor sajátította el megfelelően a tananyagot, ha

- Tudja mit jelent a szezonális.
- Mit jelent a szezonális eltérés és a szezonindex.
- Ki tudja számolni az adatokból a szezonális eltérést és a szezonindexet.
- A meglévő adatok alapján előrejelzést tud végezni.

A szezonhatás vizsgálatánál arra keresünk választ, hogy a szezonális milyen mértékben vagy arányban téríti el az idősor értékét az alapirányzattól. Vizsgálatánál az idősor adataiból ki kell szűrni a trendhatást és a véletlen hatást. A szezonális komponens eltérő módon viselkedik additív és multiplikatív modell esetén. Additív összefüggéskor a szezonhatás a trendtől való abszolút eltérés, multiplikatív kapcsolat esetén pedig a relatív eltérés formájában jelentkezik. A szezonális additív modell esetén szezonális eltérésekkel, multiplikatív modell esetén pedig szezonindexekkel jellemezzük.

18.1. Szezonális eltérés

Additív összefüggés és lineáris trend esetén az idősor megfigyelt értékeit a komponensek összegeként írhatjuk fel.

$$y_{ij} = \hat{y}_{ij} + \hat{s}_j + v_{ij}$$

A trendhatást úgy szűrjük ki, hogy az idősor megfigyelt értékeiből rendre kivonjuk a trendértékeket:

$$y_{ij} - \hat{y}_{ij} = \hat{s}_j + v_{ij}$$

Az így nyert különbségeket egyedi szezonális eltéréseknek nevezzük. Minden periódusból vesszük a j-edik eltérést és ezek számtani átlagát képezzük, ezzel a véletlen hatást szűrjük ki, illetve tompítjuk:

$$\hat{s}_j = \frac{\sum (y_{ij} - \hat{y}_{ij})}{p}$$

ahol p: az adott periódusok száma az idősorban.

A szezonális eltérés megmutatja, hogy a megfigyelt idősor a j-edik szezonban átlagosan mennyivel tér el a trendértéktől a szabályosan ismétlődő szezonhatás következtében.

Lineáris trend esetében a kapott szezonális eltérések összege nullával egyenlő. Más trendfüggvények esetében:

$$\sum s_j \neq 0$$

Ilyen esetben a szezonális eltérések korrekciójára van szükség, ekkor a kiszámolt szezonális eltéréseket nyers szezonális eltéréseknek nevezzük. A korrekció azt jelenti, hogy a kiszámolt nyers szezonális eltéréseket átlagoljuk, és az így kapott ún. korrekciós tényezőt levonjuk a nyers szezonális indexekből, és megkapjuk a korrigált szezonális eltéréseket.

$$\hat{s}_j' = \hat{s}_j - \frac{\sum \hat{s}_j}{m}$$

Ahol m : a periódusok száma az idősorban (pl. 4, ha negyedéves bontásban vannak az adataim).

Bemutató feladat:

Egy építőipari vállalkozás tevékenységének árbevétele millió Ft-ban az alábbi volt.

Év	negyedév	árbevétel	t_i
2017.	I.	20	-5,5
	II.	30	-4,5
	III.	40	-3,5
	IV.	10	-2,5
2018.	I.	22	-1,5
	II.	33	-0,5
	III.	50	0,5
	IV.	13	1,5
2019.	I.	25	2,5
	II.	36	3,5
	III.	60	4,5
	IV.	15	5,5
Összesen		354	0

A kiszámított trendegyenlet: $\hat{y} = 29,5 + 0,881 \cdot t$

Helyettesítsük be a trendegyenletbe a t_i -értékeket, pl.:

$$\hat{y} = 29,5 + 0,881 \cdot (-5,5) = 24,655$$

$$\hat{y} = 29,5 + 0,881 \cdot (-4,5) = 25,536$$

stb.

Az eredeti adatokból vonjuk ki a kiszámított trendértéket.

Pl: $20 - 24,655 = -4,655$

Év	negyedév	árbevétel	y_{trend}	$(y_i - y_{\text{trend}})$ Szezonális eltérés
2017.	I.	20	24,655	-4,655
	II.	30	25,536	4,465
	III.	40	26,417	13,584
	IV.	10	27,298	-17,298
2018.	I.	22	28,179	-6,179
	II.	33	29,060	3,941
	III.	50	29,941	20,060
	IV.	13	30,822	-17,822
2019.	I.	25	31,703	-6,703
	II.	36	32,584	3,417
	III.	60	33,465	26,536
	IV.	15	34,346	-19,346
Összesen		354		

Írjuk át egy másik táblázatba az idősor adatainak és a trendértékeknek a különbségét, azaz az egyedi szezonális eltéréseket. Ezután az egyes negyedévekhez tartozó értékeket átlagolni kell. Például az I. negyedév esetében $(-4,655-6,179-6,703)/3=-5,845$

Év	$(y_i - y_{trend})$			
	I.	II.	III.	IV.
	negyedév			
2017	-4,655	4,465	13,584	-17,298
2018	-6,179	3,941	20,060	-17,822
2019	-6,703	3,417	26,536	-19,346
s_j	-5,845	3,941	20,060	-18,155

Az első negyedévben átlagosan 5,845 millió Ft-tal kevesebb az árbevétel annál, mint ami a hosszú távú tendenciából az egyes évek első negyedéveiből következne.

Korrekcióra nincs szükség, ugyanis az egyes negyedévek szezonális eltéréseinek összege nulla.

18.2. Szezonindex

Multiplikatív összefüggés és exponenciális trend esetén az idősor megfigyelt értékeit a komponensek szorzataként írhatjuk fel:

$$y_{ij} = \hat{y}_{ij} * \hat{s}_j * \hat{v}_{ij}$$

A trendhatást úgy szűrjük ki, hogy az idősor megfigyelt értékeit rendre elosztjuk a trendértékekkel:

$$\frac{y_{ij}}{\hat{y}_{ij}} = \hat{s}_j * \hat{v}_{ij}$$

Az így nyert hányadosokat egyedi szezonindexnek nevezzük. Ezt követően minden periódusból vesszük a j-edik szezonindexet és ezeket átlagoljuk vagy számtani vagy mértani átlagot számolunk:

$$\hat{s}_j^* = \frac{\sum \left(\frac{y_{ij}}{\hat{y}_{ij}} \right)}{\frac{n}{p}}$$

Ahol n: az adatok száma

P: az adott éven belül a periódusok száma

Mértani átlaggal:

$$\hat{s}_j^* = \sqrt[p]{\prod \frac{y_{ij}}{\hat{y}_{ij}}}$$

A becsült szezonindex azt fejezi ki, hogy a j-edik szezonban a megfigyelt idősor átlagosan hányszorosa, illetve hány százaléka a trendértéknek a szezonhatás következtében.

A szezonindexnél is célszerű korrekciót végezni, ha a trendet nem exponenciális függvénnyel írtuk le. A korrekciós tényező az előzőekben kiszámolt nyers szezonindexek átlaga. A

korrigált szezonindexet pedig úgy kapjuk, hogy a nyers szezonindexeket elosztjuk a korrekciós tényezővel.

Ha a nyers szezonindexet számtani átlaggal számítottuk ki, akkor a korrigált szezonindex az alábbi:

$$\hat{s}_j^{*'} = \frac{\hat{s}_j^*}{\sum_{j=1}^m \hat{s}_j^*}$$

Ahol m : a periódusok száma az idősorban (pl. 4, ha negyedéves bontásban vannak az adataim).

Ha a nyers szezonindexet mértani átlaggal számítottuk ki, akkor a korrigált szezonindex az alábbi:

$$\hat{s}_j^{*'} = \frac{\hat{s}_j^*}{\sqrt[m]{\prod_{j=1}^m \hat{s}_j^*}}$$

Bemutató feladat

Az előző feladat alapján nézzük meg a szezonális eltéréseket.

A kiszámított trendegyenlet: $\hat{y} = 25,915 \cdot 1,020^t$

Helyettesítsük be a trendegyenletbe a t_i -értékeket:

pl.

$$\hat{y} = 25,915 \cdot 1,020^{-5,5} = 23,241$$

Év	negyedév	árbevétel	y_{trend}	(y_i/y_{trend}) Szezonindex
2017.	I.	20	23,241	0,861
	II.	30	23,706	1,266
	III.	40	24,180	1,654
	IV.	10	24,663	0,405
2018.	I.	22	25,157	0,875
	II.	33	25,660	1,286
	III.	50	26,173	1,910
	IV.	13	26,696	0,487
2019.	I.	25	27,230	0,918
	II.	36	27,775	1,296
	III.	60	28,330	2,118
	IV.	15	23,241	0,861
Összesen		354		0

Írjuk át az egyedi szezonindexeket egy másik táblázatba:

Év	(y_i/y_{trend})			
	I.	II.	III.	IV.
	negyedév			
2017	0,861	1,266	1,654	0,405
2018	0,875	1,286	1,91	0,487
2019	0,918	1,296	2,118	0,519
s_j^* (számtani átlaggal)	0,885	1,283	1,894	0,470
s_j^*	0,781	1,132	1,672	0,415
s_j^* (mértani átlaggal)	0,8843	1,2826	1,8844	0,4678
s_j^*	0,8843	1,2827	1,8845	0,4678

Számtani átlaggal számolva a nyers szezonindexeket ismét átlagolni kell, azaz ki kell számolni a korrekciós tényezőt: $(0,885+1,283+1,894+0,470)/4=1,133$. Ezzel a korrekciós tényezővel kell elosztani mindegyik nyers szezonindexet.

Pl.: $0,885/1,133=0,781$

Mértani átlaggal számolva a korrekciós tényező:

$\sqrt[4]{0,8843 \cdot 1,2826 \cdot 1,8844 \cdot 0,4678} = 0,99996$, ezzel kell mindegyik nyers szezonindexet elosztani.

Pl.: $0,8843/0,99996=0,8843$

18.3. Véletlenhatás

Az alapadatok és a trendadatok közötti eltérést a szezonhatás és a véletlenhatás okozza. Az azonos szezonokra vonatkozó eltérések átlagolásával kiszámítottuk a periodikus ingadozás abszolút (szezonális eltérés) vagy relatív (szezonindex) nagyságát. Az idősor komponenseiből már csak a véletlen tagot kell meghatározni.

Additív összefüggés esetén:

$$\hat{v}_{ij} = y_{ij} - \hat{y}_{ij} - \hat{s}_j$$

A véletlenhatás számításakor nem az egyes időpontok és a trendadatok közti különbségeket, hanem az egyes szezonokra jellemző korrigált szezonhatásokkal számolunk. A véletlenhatások várható értéke általában nulla, és nem jelezhetők előre.

Bemutató feladat:

Egy építőipari vállalkozás tevékenységének árbevétele millió Ft-ban az alábbi volt.

Év	negyedév	árbevétel	y_{trend}	$(y_i - y_{trend})$ Szezonális eltérés	$\hat{v}_{ij}$ (additív esetben)
2017.	I.	20	24,655	-4,655	1,1905
	II.	30	25,536	4,465	0,5235
	III.	40	26,417	13,584	-6,4765
	IV.	10	27,298	-17,298	0,8575
2018.	I.	22	28,179	-6,179	-0,3335
	II.	33	29,060	3,941	-0,0005
	III.	50	29,941	20,060	-0,0005
	IV.	13	30,822	-17,822	0,3335
2019.	I.	25	31,703	-6,703	-0,8575
	II.	36	32,584	3,417	-0,5245
	III.	60	33,465	26,536	6,4755
	IV.	15	34,346	-19,346	-1,1905
Összesen		354			

Multiplikatív összefüggéskor pedig:

$$\hat{v}_{ij}^* = \frac{y_{ij}}{\hat{y}_{ij}^* \cdot \hat{s}_j}$$

Év	negyedév	árbevétel	y_{trend}	(y_i/y_{trend}) Szezonindex	$\hat{v}_{ij}$ (multiplikatív esetben)
2017.	I.	20	23,241	0,861	0,7611
	II.	30	23,706	1,266	1,6232
	III.	40	24,180	1,654	3,1174
	IV.	10	24,663	0,405	0,1897
2018.	I.	22	25,157	0,875	0,7734
	II.	33	25,660	1,286	1,6496
	III.	50	26,173	1,910	3,6000
	IV.	13	26,696	0,487	0,2278
2019.	I.	25	27,230	0,918	0,8119
	II.	36	27,775	1,296	1,6625
	III.	60	28,330	2,118	3,9910
	IV.	15	23,241	0,861	0,2428

A véletlenhatások szorzatának várható értéke általában egy, és nem jelezhetők előre.

18.4. Extrapoláció

A trendegyenlet meghatározásával előrejelzést (extrapoláció) végezhetünk. A meghatározott trendegyenletbe behelyettesítjük a becsülni kívánt évhez tartozó „ t_i ”-értéket, és kiszámoljuk a trendértékét. Ezután ha van szezonhatás, akkor azzal korrigálunk. Additív összefüggés esetén a kiszámított trendértéket hozzáadjuk a szezonális eltérést, multiplikatív összefüggéskor a trendértéket megszorozzuk a szezonindexszel.

Bemutató feladat:

Becsüljük meg 2020. III. negyedévére a várható árbevételt!

$t_i=8,5$

Additív esetben:

A trendegyenlet: $\hat{y} = 29,5 + 0,881 * 8,5 = 36,9885$

A III. negyedév szezonális eltérése: 20,06

A várható árbevétel $36,9885 + 20,060 = 57,0485$ millió Ft.

Multiplikatív esetben:

$\hat{y} = 25,915 * 1,020^t = 25,915 * 1,020^{8,5} = 30,666$

A III. negyedév szezonindexe: 1,8844

A várható árbevétel $30,666 * 1,8844 = 57,787$ millió Ft.

Képletek**1. ALAPFOGALMAK, VISZONYSZÁMOK****1.1. Adatközlési abszolút és relatív hiba**

Abszolút hiba:

$$\hat{a} \leq \frac{10^k}{2},$$

ahol $\hat{a}$: a becsült abszolút hiba

k: pedig az utolsó szignifikáns számjegy helyi értéke

A relatív hiba:

$$\hat{\alpha} = \frac{\hat{a}}{A'}$$

1.2. Viszonyszámok

Viszonyszámok			
Megoszlási viszonzyszám	Bázis-viszonyszám	Lánc-viszonyszám	A fejlődés átlagos üteme
$V_m = \frac{x_i}{\sum x_i}$ $w_j = \frac{n_j}{\sum n_j}$	$l_i(V_b) = \frac{y_i}{y_0}$	$l_b(V_l) = \frac{y_i}{y_{i-1}}$	$\bar{V}_l = \sqrt[n]{V_{l1} * V_{l2} * \dots * V_{ln}} =$ $= \sqrt[n]{\prod V_{ln}} = \sqrt[n]{\frac{y_n}{y_0}}$

Összefüggések $l_i = \frac{b_i}{b_{i-1}}$; $b_k = l_2 * l_3 * \dots * l_k = \prod_{i=2}^k l_i$, ahol k az idősor feladat szerint megjelölt tagja

2. EMPIRIKUS ELOSZLÁSOK ELEMZÉSE**A mennyiségi sorok jellemző értékeinek kiszámítása****2.1. Középtértékek, és kvantilsiek**

Jelölések: x_i az i -edik adat értéke vagy az i -edik osztály közepe, n az adatok száma, r az osztályok száma, f_i az i -edik osztályba tartozó adatok száma.

	Egyedi adatokból	Gyakorisági adatokból
Számtani közép	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$	$\bar{x} = \frac{\sum_{i=1}^r f_i * x_i}{\sum_{i=1}^r f_i} = \frac{\sum_{i=1}^r s_i}{\sum_{i=1}^r f_i} = \sum_{i=1}^n g_i * x_i$
Harmonikus átlag	$\bar{x}_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$	$\bar{x}_h = \frac{\sum_{i=1}^r f_i}{\sum_{i=1}^r \frac{f_i}{x_i}} = \frac{1}{\sum_{i=1}^r \frac{g_i}{x_i}}$
Geometria átlag	$\bar{x}_g = \sqrt[n]{\prod_{i=1}^n x_i}$	$\bar{x}_g = \sqrt[n]{\prod_{i=1}^r x_i^{f_i}}$

Képletek

Négyzetes átlag	$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$	$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^r f_i x_i^2}{n}}$
Módusz becslése	$Mo = x_{mo,a} + \frac{d_a}{d_a + d_f} \cdot h_{mo}$ $d_a = f_{mo} - f_{mo-1}, \quad d_f = f_{mo} - f_{mo+1}$ $M_o = m_0 + \frac{f_{m_o} - f_{m_{o-1}}}{(f_{m_o} - f_{m_{o-1}}) + (f_{m_o} - f_{m_{o+1}})} \cdot h$	
Medián becslése:	$Me = x_{me,a} + \frac{\frac{n}{2} - f'_{me-1}}{f_{me}} \cdot h_{me}$	
Kvantilisek becslése	$K_V = x_{kv,a} + \frac{\frac{j}{k} \cdot n - f'_{kv-1}}{f_{kv}} \cdot h_{kv}$	

2.2 Szóródási mutatók

Terjedelem: $R(T) = x_{\max} - x_{\min}$

Interkvartilis terjedeleml: $IQT = Q_3 - Q_1$

Szórás:

Egyedi adatokból:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}, \text{ vagy } \sigma = \sqrt{\frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n}}$$

Gyakorisági adatokból:

$$\sigma = \sqrt{\frac{\sum_{i=1}^r f_i (x_i - \bar{x})^2}{n}}, \text{ vagy } \sigma = \sqrt{\frac{\sum_{i=1}^r f_i x_i^2 - n\bar{x}^2}{n}}$$

Relatív szórás: $V(CV) = \frac{\sigma}{x}$

Teljes szórásnégyzet összetevői: $\sigma^2 = \sigma_B^2 + \sigma_K^2$,

Belső szórásnégyzet:

$$\sigma_B^2 = \frac{\sum n_j \sigma_j^2}{n}$$

Ahol: j : a csoportosító ismérv változatainak száma

Külső szórásnégyzet:

$$\sigma_K^2 = \frac{\sum n_j d_j^2}{n}$$

ahol: $d_j^2 = (\bar{x}_j - \bar{x})^2$

Átlagos eltérés:

$$\delta = \frac{\sum_{i=1}^r f_i |x_i - \bar{x}|}{\sum_{i=1}^r f_i} \text{ vagy } \delta = \sum_{i=1}^r g_i |x_i - \bar{x}|$$

Átlagos különbség:

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{n(n-1)} \quad G = \frac{\sum_{i=1}^r \sum_{j=1}^r f_i f_j |x_i - x_j|}{n(n-1)}$$

Koncentráció (Lorenz-féle területarány):

$$K = \frac{T_k}{T_b} = \frac{G}{2\bar{x}}$$

2.3. Eloszlások, gyakorisági sorok jellemzői

Aszimmetria mértéke:

$$A = \frac{\bar{x} - Mo}{\sigma}$$

$$P = \frac{3(\bar{x}_a - Me)}{\sigma}$$

Ferdeségi mutató:

$$F = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)}$$

3. ÖSSZEFÜGGÉS-VIZSGÁLAT EGYSZERŰBB ESZKÖZEI

3.1. Asszociáció:

Yule-féle együttható:

$$A(Y) = \frac{f_{11} f_{22} - f_{21} f_{12}}{f_{11} f_{22} + f_{21} f_{12}}$$

Csuprov- féle asszociációs együttható:

$$T = \sqrt{\frac{\chi^2}{n\sqrt{(s-1)(t-1)}}},$$

$$\text{ahol } \chi^2 = n \sum_{i=1}^s \sum_{j=1}^t \left(\frac{f_{ij}^2}{f_{i.} f_{.j}} - 1 \right),$$

ahol n az adatok száma, s a sorok száma, t az oszlopok száma,

vagy

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - f_{ij}^*)^2}{f_{ij}^*},$$

$$\text{ahol: } f_{ij}^* = \frac{f_{i.} \cdot f_{.j}}{n} \quad i = 1, 2 \dots s, \quad j = 1, 2 \dots t$$

Cramer-féle együttható:

$$C = \sqrt{\frac{\chi^2}{n(s-1)}}, \text{ ahol: } s \leq t$$

3.2. Korrelációs együtthatók:

Spearman-féle rangkorrelációs együttható:

$$r_r = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n * (n^2 - 1)}, \text{ ahol: } D_i = x_i - y_i$$

Előjel-korrelációs együttható:

$$r_e = \frac{p - q}{p + q}$$

3.3. Vegyes kapcsolat

Szórásnégyzet-hányados:

$$H^2 = \frac{\sigma_K^2}{\sigma^2} = \frac{SSK}{SST}$$

vagy

$$H = \frac{\sigma_K}{\sigma_T} = \sqrt{1 - \frac{\sigma_B^2}{\sigma_T^2}}$$

4. INDEXSZÁMÍTÁS

4.1. Standardizáláson alapuló indexszámítás

4.1.1. Különbségek

Főátlagok különbsége:

$$K = \bar{V}_1 - \bar{V}_0 = \frac{\sum B_1 V_1}{\sum B_1} - \frac{\sum B_0 V_0}{\sum B_0}$$

Részátlagok különbsége:

$$K' = \bar{V}_1 - \bar{V}_{st:B_1} \text{ vagy } K' = \bar{V}_{st:B_0} - \bar{V}_0,$$

ahol

$$\bar{V}_{st:B_1} = \bar{V}_{st:v_0} = \frac{\sum B_1 v_0}{\sum B_1}$$

$$\bar{V}_{st:B_0} = \bar{V}_{st:v_1} = \frac{\sum B_0 v_1}{\sum B_0}$$

Összetételből eredő különbség:

$$K'' = \bar{V}_{st:v_0} - \bar{V}_0 \text{ vagy } K'' = \bar{V}_1 - \bar{V}_{st:v_1}$$

Összefüggés: $K = K' + K''$

4.1.2. Indexek

Főátlagindex:

$$I = \frac{\bar{V}_1}{\bar{V}_0}$$

Részátlagindex:

$$I' = \frac{\bar{V}_1}{\bar{V}_{st:B_1}} \quad \text{vagy} \quad I' = \frac{\bar{V}_{st:B_0}}{\bar{V}_0}$$

$$I' = \frac{\frac{\sum B_1 V_1}{\sum B_1}}{\frac{\sum B_1 V_0}{\sum B_1}} \quad \text{illetve} \quad I' = \frac{\frac{\sum B_0 V_1}{\sum B_0}}{\frac{\sum B_0 V_0}{\sum B_0}}$$

Összetételhatás indexe:

$$I'' = \frac{\bar{V}_{st:V_0}}{\bar{V}_0} \quad \text{vagy} \quad I'' = \frac{\bar{V}_1}{\bar{V}_{st:V_1}}$$

$$I'' = \frac{\frac{\sum B_1 V_0}{\sum B_1}}{\frac{\sum B_0 V_0}{\sum B_0}} \quad \text{illetve} \quad I'' = \frac{\frac{\sum B_1 V_1}{\sum B_1}}{\frac{\sum B_0 V_1}{\sum B_0}}$$

Összefüggés: $I = I' \cdot I''$

Megjegyzés:

A 0 és 1 index-megjelölés időbeli és térbeli összehasonlításra egyaránt utalhat.

4.2. Érték-, ár- és volumenindexek

4.2.1. Egyedi értékindex:

$$i_v = \frac{v_{i1}}{v_{i0}} = \frac{q_{i1} p_{i1}}{q_{i0} p_{i0}}$$

4.2.2. Egyedi volumenindex:

$$i_q = \frac{q_{i1}}{q_{i0}}$$

4.2.3. Egyedi árindex

$$i_p = \frac{p_{i1}}{p_{i0}}$$

4.2.4. Különbségek:

$$K_v = \sum q_1 p_1 - \sum q_0 p_0$$

$$K_p^0 = \sum q_0 p_1 - \sum q_0 p_0$$

$$K_p^1 = \sum q_1 p_1 - \sum q_1 p_0$$

$$K_q^1 = \sum q_1 p_1 - \sum q_0 p_1$$

$$K_q^0 = \sum q_1 p_0 - \sum q_0 p_0$$

$$K_v = K_p^1 + K_q^0 = K_p^0 + K_q^1$$

4.2.5. Értékindex:

$$I_v = \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{\sum q_0 p_0 \cdot i_v}{\sum q_0 p_0} = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_v}}$$

4.2.6. Volumenindex:

- Laspeyres-féle (bázis súlyozás):

$$I_q^L = I_q^0 = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum q_0 p_0 \cdot i_q}{\sum q_0 p_0} = \frac{\sum q_1 p_0}{\sum \frac{q_1 p_0}{i_q}}$$

- Paasche-féle (tárgy időszaki súlyozás):

$$I_q^P = I_q^1 = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{\sum q_0 p_1 \cdot i_q}{\sum q_0 p_1} = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_q}}$$

- Fisher-féle:

$$I_q^F = \sqrt{I_q^L \cdot I_q^P}$$

- Edgeworth-Marshall féle:

$$I_q^{EM} = \frac{\sum q_1 (p_0 + p_1)}{\sum q_0 (p_0 + p_1)}$$

4.2.7. Árindex

- Laspeyres-féle (bázis súlyozás):

$$I_p^L = I_p^0 = \frac{\sum q_0 p_1}{\sum q_0 p_0} = \frac{\sum q_0 p_0 \cdot i_p}{\sum q_0 p_0} = \frac{\sum q_0 p_1}{\sum \frac{q_0 p_1}{i_p}}$$

- Paasche-féle (tárgy időszaki súlyozás):

$$I_p^P = I_p^1 = \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{\sum q_1 p_0 \cdot i_p}{\sum q_1 p_0} = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_p}}$$

$$I_p^P = \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{\sum q_1 p_0 \cdot i_p}{\sum q_1 p_0} = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_p}}$$

- Fisher-féle:

$$I_p^F = \sqrt{I_p^L \cdot I_p^P}$$

4.2.8. Indexek közötti összefüggés:

$$I_v = I_q^P \cdot I_p^L = I_q^L \cdot I_p^P = I_q^F \cdot I_p^F$$

4.2.9. Területi indexek

$$I_{q(A/B)}^A = \frac{\sum q_A p_A}{\sum q_B p_A}; \quad I_{q(A/B)}^B = \frac{\sum q_A p_B}{\sum q_B p_B}; \quad I_{q(A/B)}^F = \sqrt{I_{q(A/B)}^A \cdot I_{q(A/B)}^B}$$

$$I_{p(A/B)}^A = \frac{\sum q_A p_A}{\sum q_A p_B}; \quad I_{p(A/B)}^B = \frac{\sum q_B p_A}{\sum q_B p_B}; \quad I_p^F = \sqrt{I_{p(A/B)}^A \cdot I_{p(A/B)}^B}$$

5. Mintavétel, statisztikai becslés

Átlag becslése:

Standard hiba:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Hibahatár:

$$\Delta = z_p \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Delta = t_p^{(szf)} \cdot \frac{s}{\sqrt{n}};$$

Intervallumbecslés:

$$\left(\bar{x} - z_p \cdot \frac{\sigma}{\sqrt{n}}; \bar{x} + z_p \cdot \frac{\sigma}{\sqrt{n}} \right)$$

$$\left(\bar{x} - t_p^{(szf)} \cdot \frac{s}{\sqrt{n}}; \bar{x} + t_p^{(szf)} \cdot \frac{s}{\sqrt{n}} \right)$$

Szórás intervallumbecslése

$$\left(\sqrt{\frac{(n-1) \cdot s^2}{\chi^2_{1-\frac{\alpha}{2}}(szf)}}; \sqrt{\frac{(n-1) \cdot s^2}{\chi^2_{\frac{\alpha}{2}}(szf)}} \right)$$

Arány pontbecslése

$$p = k/n$$

Arány standard hibája

Arány intervallumbecslése

$$s_p = \sqrt{\frac{p \cdot (1-p)}{n}}$$

$$(p - z_p \cdot s_p; p + z_p \cdot s_p)$$

Elemsszám meghatározása: adott az intervallum és a valószínűség.

$$n = \left(\frac{z_p \cdot \sigma}{\Delta} \right)^2 \text{ illetve } n = \left(\frac{t_p^{(szf)} \cdot s}{\Delta} \right)^2$$

Valószínűségi szint meghatározása:

$$z_p = \frac{\Delta \cdot \sqrt{n}}{\sigma} \text{ illetve } t_p^{(szf)} = \frac{\Delta \cdot \sqrt{n}}{s}.$$

$$n_j = n \cdot \frac{w_j \cdot \sigma_j}{\sum w_j \sigma_j}$$

Értékösszegek becslése:

$$N \cdot (\bar{x} \pm z_p \cdot s_{\bar{x}}) \dots \text{vagy} \dots N \cdot (\bar{x} \pm z_p \cdot \sigma_{\bar{x}})$$

6. HIPOTÉZISVIZSGÁLATOK

6.1. Alapsokasági átlagra, szórásra, arányra tett hipotézis ellenőrzése

Null hipotézis: $H_0 : \bar{X} = \bar{X}_0$ vagy $H_0 : \bar{X} - \bar{X}_0 = 0$

Alternatív hipotézisek:

$$H_1 : \begin{array}{l} \bar{X} \neq \bar{X}_0 \\ \bar{X} > \bar{X}_0 \\ \bar{X} < \bar{X}_0 \end{array}$$

Próbafüggvény:

Ha a sokaság szórása ismert: $z_0 = \frac{\bar{x} - \bar{X}_0}{\frac{\sigma}{\sqrt{n}}}$,

Ha a sokaság szórását mintából becsültük:

$$t_0 = \frac{\bar{x} - \bar{X}_0}{\frac{s}{\sqrt{n}}}$$

szabadságfok: $n-1$

Ha a minta elemszáma nagy: $z_0 = \frac{\bar{x} - \bar{X}_0}{\frac{s}{\sqrt{n}}}$,

Arányra tett hipotézis ellenőrzése: $H_0 : P = P_0$

Próbafüggvény: $z = \frac{p - P_0}{\sqrt{\frac{P_0 \cdot (1 - P_0)}{n}}}$

Sokasági szórásra vonatkozó hipotézisvizsgálat próbafüggvénye:

$$\chi_0^2 = \frac{(n-1) \cdot s^2}{\sigma_0^2}$$

6.2. Két alapsokasági átlag különbségére tett hipotézis ellenőrzése

Null hipotézis: $H_0 : \bar{X}_1 - \bar{X}_2 = 0$

Próbafüggvény:

Ha a két sokaság szórása ismert vagy mindkét minta nagy:

$$z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Ha a két sokaság szórása nem ismert és a két minta szórása nem különbözik lényegesen:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{s_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \text{ ahol } s_d = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$$

ahol s_d a közös szórás

Szabadságfok: n_1+n_2-2

6.3. Két szórás egyezésére tett hipotézis ellenőrzése

Null hipotézis: $H_0 : \sigma_1 = \sigma_2$

Próbafüggvény: $F = \frac{s_1^2}{s_2^2}$, ahol $s_1 > s_2$,

Szabadságfok: számláló n_1-1 , nevező: n_2-1

6.4. Arányra (megoszlásra) tett hipotézis ellenőrzése

Két sokasági arányra vonatkozó:

Null hipotézis: $H_0 : P_1 = P_2$

Próbafüggvény:
$$z_0 = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

6.5. Varianciaanalízis

- Null hipotézis: $H_0 : \bar{X}_1 = \bar{X}_2 = \dots \bar{X}_m = \bar{X}$
- Alternatív hipotézis: H_1 : legalább két átlag szignifikánsan különbözik.
- Próbafüggvény:

$$F = \frac{s_k^2}{s_b^2} = \frac{SSK/(M-1)}{SSB/(n-M)} = \frac{MS_k}{MS_H} = \frac{MQ_k}{MQ_H}$$

- szabadság fok: $\nu_1 = m - 1, \nu_2 = n - m$

ahol $SSK = \sum (\bar{x}_j - \bar{x})^2 \cdot n_j$ a csoportok közötti eltérés négyzetösszege.

$SSB = \sum_j \sum_i (x_{ij} - \bar{x}_j)^2$ a csoportokon belüli eltérés négyzetösszege.

m: a csoportok száma

6.6. Illeszkedésvizsgálat próbafüggvénye

$$\chi^2 = \sum_{i=1}^K \frac{(f_i - n \cdot P_i)^2}{n \cdot P_i} = \sum \frac{(f_i - f_i^*)^2}{f_i^*} = n \cdot \left(\sum \frac{g_i^2}{P_i} - 1 \right)$$

$$f_i^* = n \cdot P_i$$

Szabadságfok $k - v - 1$, ahol v a mintából becsült paraméterek száma, k a csoportok száma.

6.7. Két eloszlás egyezőségének a vizsgálata

$$\chi^2 = n_1 \cdot n_2 \cdot \sum \frac{1}{n_{1i} + n_{2i}} \cdot \left(\frac{n_{1i}}{n_1} - \frac{n_{2i}}{n_2} \right)^2$$

$$\frac{n_{1i}}{n_1} = g_i$$

g_i : relatív gyakoriság

6.8. Függetlenségvizsgálat

$H_0: P_{ij} = P_{i\bullet} \cdot P_{\bullet j} \quad (i=1,2,\dots,s; j=1,2,\dots,t)$

$H_1: P_{ij} \neq P_{i\bullet} \cdot P_{\bullet j}$

$$\chi^2 = \sum_{i=1}^s \sum_{j=1}^t \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*} = n^* \left(\sum_{i=1}^s \sum_{j=1}^t \frac{n_{ij}^2}{n_{i\bullet}^* \cdot n_{\bullet j}} - 1 \right)$$

$$n_{ij}^* = n^* p_{i\bullet}^* \cdot p_{\bullet j} = \frac{n_{i\bullet}^* \cdot n_{\bullet j}}{n}$$

Szabadságfok: $(s-1) \cdot (t-1)$

6.9. Normalitásvizsgálat

Jarque-Bera-teszt: $JB = n \left[\frac{\alpha_3^2}{6} + \frac{\alpha_4^2}{24} \right] \rightarrow \chi^2(2)$

7. Korreláció- és regressziószámítás

Kovariancia:

$$C = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$$

korrelációs együttható:

$$r_{xy} = \frac{\sum_{i=1}^n d_{x_i} \cdot d_{y_i}}{\sqrt{\sum_{i=1}^n d_{x_i}^2 \cdot \sum_{i=1}^n d_{y_i}^2}}; \quad r = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}};$$

$$r = \frac{\sum x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{(\sum x_i^2 - n \cdot \bar{x}^2) \cdot (\sum y_i^2 - n \cdot \bar{y}^2)}} = \beta_1 \cdot \frac{\sigma_x}{\sigma_y}$$

determinációs együttható:

$$r^2 = \frac{SSR}{SST} = \frac{SST - SSR}{SST} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (y_i - \bar{y})^2 - \sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Regressziós együtthatók:

$$\beta_1 = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum x_i^2 - n \cdot \bar{x}^2}$$

$$\beta_0 = \bar{y} - \beta_1 \cdot \bar{x}$$

Rugalmassági együttható:

$$E = \frac{\beta_1 \cdot x_0}{\beta_0 + \beta_1 x_0}$$

Reziduális szórás	Relatív változás
-------------------	------------------

$\sigma_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$	$CV_{\sigma_e} = \frac{\sigma_e}{\bar{y}} * 100$
--	--

korrelációs index:

$$I = \sqrt{1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}}$$

$$e_i^2 = (y_i - \hat{y}_i)^2$$

8. Idősorok elemzése

$$\hat{\beta}_1 = \frac{\sum t_i * y_i - \frac{(\sum t_i) * (\sum y_i)}{n}}{\sum t_i^2 - \frac{(\sum t_i)^2}{n}} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 * \frac{\sum t_i}{n}$$

$$\sum y_i = n * \hat{\beta}_0$$

$$\sum t_i * y_i = \hat{\beta}_1 * \sum t_i^2$$

$$\hat{\beta}_0 = \frac{\sum y_i}{n} = \bar{y}$$

$$\hat{\beta}_1 = \frac{\sum t_i * y_i}{\sum t_i^2}$$

Reziduális szórásnégyzet:

$$s_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

Szezonális eltérés:

$$\hat{s}_j = \frac{\sum (y_{ij} - \hat{y}_{ij})}{p}$$

Korrigált szezonális eltérés:

$$\hat{s}_j' = \hat{s}_j - \frac{\sum \hat{s}_j}{m}$$

Szezonindex:

$$\hat{s}_j^* = \frac{\sum \left(\frac{y_{ij}}{\hat{y}_{ij}} \right)}{\frac{n}{p}} \quad \text{vagy} \quad \hat{s}_j^* = \sqrt[p]{\prod \frac{y_{ij}}{\hat{y}_{ij}}}$$

Korrigált szezonindex:

$$\hat{s}_j^{*'} = \frac{\hat{s}_j^*}{\sum \hat{s}_j^*} \quad \text{vagy} \quad \hat{s}_j^{*'} = \frac{\hat{s}_j^*}{\sqrt[m]{\prod \hat{s}_j^*}}$$

Véletlenhatás:

$$\hat{v}_{ij}^* = \frac{\hat{y}_{ij}}{\hat{y}_{ij}^* \hat{s}_j} \text{ vagy } \hat{v}_{ij} = \hat{y}_{ij} - \hat{y}_{ij} - \hat{s}_j$$

Táblázatok**A standard normális eloszlás eloszlásfüggvényének értékei**

z	p									
	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998

A t (Student) eloszlású változó eloszlásának kvantilis értékei

v	p						
	0,6	0,75	0,9	0,95	0,975	0,99	0,995
1	0,32	1,00	3,08	6,31	12,71	31,82	63,66
2	0,29	0,82	1,89	2,92	4,30	6,96	9,92
3	0,28	0,76	1,64	2,35	3,18	4,54	5,84
4	0,27	0,74	1,53	2,13	2,78	3,75	4,60
5	0,27	0,73	1,48	2,02	2,57	3,36	4,03
6	0,26	0,72	1,44	1,94	2,45	3,14	3,71
7	0,26	0,71	1,41	1,89	2,36	3,00	3,50
8	0,26	0,71	1,40	1,86	2,31	2,90	3,36
9	0,26	0,70	1,38	1,83	2,26	2,82	3,25
10	0,26	0,70	1,37	1,81	2,23	2,76	3,17
11	0,26	0,70	1,36	1,80	2,20	2,72	3,11
12	0,26	0,70	1,36	1,78	2,18	2,68	3,05
13	0,26	0,69	1,35	1,77	2,16	2,65	3,01
14	0,26	0,69	1,35	1,76	2,14	2,62	2,98
15	0,26	0,69	1,34	1,75	2,13	2,60	2,95
16	0,26	0,69	1,34	1,75	2,12	2,58	2,92
17	0,26	0,69	1,33	1,74	2,11	2,57	2,90
18	0,26	0,69	1,33	1,73	2,10	2,55	2,88
19	0,26	0,69	1,33	1,73	2,09	2,54	2,86
20	0,26	0,69	1,33	1,72	2,09	2,53	2,85
21	0,26	0,69	1,32	1,72	2,08	2,52	2,83
22	0,26	0,69	1,32	1,72	2,07	2,51	2,82
23	0,26	0,69	1,32	1,71	2,07	2,50	2,81
24	0,26	0,68	1,32	1,71	2,06	2,49	2,80
25	0,26	0,68	1,32	1,71	2,06	2,49	2,79
26	0,26	0,68	1,31	1,71	2,06	2,48	2,78
27	0,26	0,68	1,31	1,70	2,05	2,47	2,77
28	0,26	0,68	1,31	1,70	2,05	2,47	2,76
29	0,26	0,68	1,31	1,70	2,05	2,46	2,76
30	0,26	0,68	1,31	1,70	2,04	2,46	2,75
40	0,26	0,68	1,30	1,68	2,02	2,42	2,70
50	2,60	0,68	1,30	1,68	2,01	2,40	2,68
60	0,25	0,68	1,30	1,67	2,00	2,39	2,66
100	2,50	0,68	1,29	1,66	1,98	2,36	2,63
120	0,25	0,68	1,29	1,66	1,98	2,36	2,62
végtelen	0,25	0,67	1,28	1,65	1,96	2,33	2,58

A χ^2 eloszlású változó eloszlásának kvantilis értékei

v	p												
	0,005	0,01	0,025	0,05	0,1	0,25	0,5	0,75	0,9	0,95	0,975	0,99	0,995
1	0,00	0,00	0,00	0,00	0,02	0,10	0,45	1,32	2,71	3,84	5,02	6,63	7,88
2	0,01	0,02	0,05	0,10	0,21	0,58	1,39	2,77	4,61	5,99	7,38	9,21	10,6
3	0,07	0,11	0,22	0,35	0,58	1,21	2,37	4,11	6,25	7,81	9,35	11,3	12,8
4	0,21	0,30	0,48	0,71	1,06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9
5	0,41	0,55	0,83	1,15	1,61	2,67	4,35	6,63	9,24	11,1	12,8	15,1	16,7
6	0,68	0,87	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5
7	0,99	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,4	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	7,58	10,3	13,7	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	8,44	11,3	14,8	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	9,30	12,3	16,0	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	10,2	13,3	17,1	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	8,55	11,0	14,3	18,2	22,3	25,0	27,5	30,6	32,8
16	5,14	5,81	6,91	7,96	9,31	11,9	15,3	19,4	23,5	26,3	28,8	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	12,8	16,3	20,5	24,8	27,6	30,2	33,4	35,7
18	6,26	7,01	8,23	9,39	10,9	13,7	17,3	21,6	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,1	11,7	14,6	18,3	22,7	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,59	10,9	12,4	15,5	19,3	23,8	28,4	31,4	34,2	37,6	40,0
21	8,03	8,90	10,3	11,6	13,2	16,3	20,3	24,9	29,6	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	14,0	17,2	21,3	26,0	30,8	33,9	36,8	40,3	42,8
23	9,26	10,2	11,7	13,1	14,8	18,1	22,3	27,1	32,0	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	15,7	19,0	23,3	28,2	33,2	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	16,5	19,9	24,3	29,3	34,4	37,7	40,6	44,3	46,9
26	11,2	12,2	13,8	15,4	17,3	20,8	25,3	30,4	35,6	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	18,1	21,7	26,3	31,5	36,7	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	18,9	22,7	27,3	32,6	37,9	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	19,8	23,6	28,3	33,7	39,1	42,6	45,7	49,6	52,3
30	13,8	15,0	16,8	18,5	20,6	24,5	29,3	34,8	40,3	43,8	47,0	50,9	53,7
40	20,7	22,2	24,4	26,5	29,1	33,7	39,3	45,6	51,8	55,8	59,3	63,7	66,8
50	28,0	29,7	32,4	34,8	37,7	42,9	49,3	56,3	63,2	67,5	71,4	76,2	79,5
60	35,5	37,5	40,5	43,2	46,5	52,3	59,3	67,0	74,4	79,1	83,3	88,4	92,0
70	43,3	45,4	48,8	51,7	55,3	61,7	69,3	77,6	85,5	90,5	95,0	100,4	104,2
80	51,2	53,5	57,2	60,4	64,3	71,1	79,3	88,1	96,6	101,9	106,6	112,3	116,3
90	59,2	61,8	65,6	69,1	73,3	80,6	89,3	98,6	107,0	113,1	118,1	124,1	128,3
100	67,3	70,1	74,2	77,9	82,4	90,1	99,3	109,1	118,0	124,3	129,6	135,8	140,2
200	152,2	156,1	162,7	168,3	174,8	186,2	199,3	213,1	226,0	234,0	241,1	248,7	255,3

Az F-eloszlás kvartilis értékei

$$1 - \alpha = 0,9$$

v_2	v_1								
	1	2	3	4	5	6	7	8	9
1	39,86	49,50	53,59	55,83	57,24	58,20	58,90	59,44	59,90
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44
10	3,28	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68
∞	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63

a) v_1 a számláló, v_2 a nevező szabadságfokát jelöli.

Az F-eloszlás kvartilis értékei (Folytatás)

$$1 - \alpha = 0,9$$

v_2	v_1								
	10	12	15	20	30	40	60	120	∞
1	60,20	60,70	61,20	61,70	62,30	62,50	62,80	63,10	63,30
2	9,39	9,41	9,42	9,44	9,46	9,47	9,47	9,48	9,49
3	5,23	5,22	5,20	5,18	5,17	5,16	5,15	5,14	5,13
4	3,92	3,90	3,87	3,84	3,82	3,80	3,79	3,78	3,76
5	3,30	3,27	3,24	3,21	3,17	3,16	3,14	3,12	3,10
6	2,94	2,90	2,87	2,84	2,80	2,78	2,76	2,74	2,72
7	2,70	2,67	2,63	2,59	2,56	2,54	2,51	2,49	2,47
8	2,54	2,50	2,46	2,42	2,38	2,36	2,34	2,32	2,29
9	2,42	2,38	2,34	2,30	2,25	2,23	2,21	2,18	2,16
10	2,32	2,28	2,24	2,20	2,16	2,13	2,11	2,08	2,06
11	2,25	2,21	2,17	2,12	2,08	2,05	2,03	2,00	1,97
12	2,19	2,15	2,10	2,06	2,01	1,99	1,96	1,93	1,90
13	2,14	2,10	2,05	2,01	1,96	1,93	1,90	1,88	1,85
14	2,10	2,05	2,01	1,96	1,91	1,89	1,86	1,83	1,80
15	2,06	2,02	1,97	1,92	1,87	1,85	1,82	1,79	1,76
16	2,03	1,99	1,94	1,89	1,84	1,81	1,78	1,75	1,72
17	2,00	1,96	1,91	1,86	1,81	1,78	1,75	1,72	1,69
18	1,98	1,93	1,89	1,84	1,78	1,75	1,72	1,69	1,66
19	1,96	1,91	1,86	1,81	1,76	1,73	1,70	1,67	1,63
20	1,94	1,89	1,84	1,79	1,74	1,71	1,68	1,64	1,61
21	1,92	1,88	1,83	1,78	1,72	1,69	1,66	1,62	1,59
22	1,90	1,86	1,81	1,76	1,70	1,67	1,64	1,60	1,57
23	1,89	1,84	1,80	1,74	1,69	1,66	1,62	1,59	1,55
24	1,88	1,83	1,78	1,73	1,67	1,64	1,61	1,57	1,53
25	1,87	1,82	1,77	1,72	1,66	1,63	1,59	1,56	1,52
26	1,86	1,81	1,76	1,71	1,65	1,61	1,58	1,54	1,50
27	1,85	1,80	1,75	1,70	1,64	1,60	1,57	1,53	1,49
28	1,84	1,79	1,74	1,69	1,63	1,59	1,56	1,52	1,48
29	1,83	1,78	1,73	1,68	1,62	1,58	1,55	1,51	1,47
30	1,82	1,77	1,72	1,67	1,61	1,57	1,54	1,50	1,46
40	1,76	1,71	1,66	1,61	1,54	1,51	1,47	1,42	1,38
60	1,71	1,66	1,60	1,54	1,48	1,44	1,40	1,35	1,29
120	1,65	1,60	1,55	1,48	1,41	1,37	1,32	1,26	1,19
∞	1,60	1,55	1,49	1,42	1,34	1,30	1,24	1,17	1,00

Az F-eloszlás kvartilis értékei

$$1 - \alpha = 0,95$$

v_2	v_1											
	1	2	3	4	5	6	7	8	9	10	12	15
1	161,00	200,00	216,00	225,00	230,00	234,00	237,00	239,00	241,00	242,0	244,0	246,0
2	18,50	19,00	19,20	19,20	19,30	19,30	19,40	19,40	19,40	19,40	19,40	19,40
3	10,10	9,55	9,28	9,12	9,01	8,84	8,89	8,85	8,81	8,79	8,74	8,70
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,85	1,77
200	3,89	3,04	2,65	2,41	2,26	2,14	2,05	1,98	1,92	1,87	1,80	1,72
400	3,86	3,02	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,78	1,70
1000	3,85	3,00	2,61	2,38	2,22	2,10	2,0,	1,95	1,89	1,84	1,76	1,68
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67

b) v_1 a számláló, v_2 a nevező szabadságfokát jelöli.

Az F-eloszlás kvartilis értékei (folytatás)

$$1-\alpha = 0,95$$

v_2	v_1								
	20	30	40	60	100	120	200	500	∞
1	248,0	250,0	251,0	252,0	253,0	253,0	254,0	254,0	254,0
2	19,40	19,50	19,50	19,50	19,49	19,50	19,49	19,50	19,50
3	8,66	8,62	8,59	8,57	8,56	8,55	8,54	8,54	8,53
4	5,80	5,75	5,72	5,69	5,66	5,66	5,65	5,64	5,63
5	4,56	4,50	4,46	4,43	4,40	4,40	4,38	4,37	4,36
6	3,87	3,81	3,77	3,74	3,71	3,70	3,69	3,68	3,67
7	3,44	3,38	3,34	3,30	3,28	3,27	3,25	3,24	3,23
8	3,15	3,08	3,04	3,01	2,98	2,97	2,96	2,94	2,93
9	2,94	2,86	2,83	2,79	2,76	2,75	2,73	2,72	2,71
10	2,77	2,70	2,66	2,62	2,59	2,58	2,56	2,55	2,54
11	2,65	2,57	2,53	2,49	2,45	2,45	2,42	2,41	2,40
12	2,54	2,47	2,43	2,38	2,35	2,34	2,32	2,31	2,30
13	2,46	2,38	2,34	2,30	2,26	2,25	2,24	2,22	2,21
14	2,39	2,31	2,27	2,22	2,19	2,18	2,16	2,14	2,13
15	2,33	2,25	2,20	2,16	2,12	2,11	2,10	2,08	2,07
16	2,28	2,19	2,15	2,11	2,07	2,06	2,04	2,02	2,01
17	2,23	2,15	2,10	2,06	2,02	2,01	1,99	1,97	1,96
18	2,19	2,11	2,06	2,02	1,98	1,97	1,95	1,93	1,92
19	2,16	2,07	2,03	1,98	1,94	1,93	1,91	1,90	1,88
20	2,12	2,04	1,99	1,95	1,90	1,90	1,87	1,85	1,84
21	2,10	2,01	1,96	1,92	1,87	1,87	1,84	1,82	1,81
22	2,07	1,98	1,94	1,89	1,84	1,84	1,81	1,80	1,78
23	2,05	1,96	1,91	1,86	1,82	1,81	1,79	1,77	1,76
24	2,03	1,94	1,89	1,84	1,80	1,79	1,76	1,74	1,73
25	2,01	1,92	1,87	1,82	1,77	1,77	1,74	1,72	1,71
26	1,99	1,90	1,85	1,80	1,76	1,75	1,72	1,70	1,69
27	1,97	1,88	1,84	1,79	1,74	1,73	1,71	1,68	1,67
28	1,96	1,87	1,82	1,77	1,72	1,71	1,69	1,67	1,65
29	1,94	1,85	1,81	1,75	1,71	1,70	1,68	1,65	1,64
30	1,93	1,84	1,79	1,74	1,69	1,68	1,66	1,64	1,62
40	1,84	1,74	1,69	1,64	1,59	1,58	1,55	1,53	1,51
60	1,75	1,65	1,59	1,53	1,48	1,47	1,44	1,41	1,39
100	1,68	1,57	1,51	1,45	1,39	1,35	1,33	1,30	1,28
200	1,62	1,52	1,45	1,41	1,32	1,31	1,26	1,22	1,19
400	1,60	1,49	1,42	1,36	1,28	1,29	1,22	1,16	1,13
1000	1,58	1,47	1,41	1,34	1,26	1,28	1,19	1,13	1,08
∞	1,57	1,46	1,39	1,32	1,24	1,22	1,17	1,11	1,00

Az F-eloszlás kvartilis értékei

$$1 - \alpha = 0,99$$

v_2	v_1											
	1	2	3	4	5	6	7	8	9	10	12	14
1	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6106	6142
2	98,49	99,00	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40	99,42	99,43
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,05	26,92
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,37	14,24
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,89	9,77
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,60
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,47	6,35
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,67	5,56
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,11	5,00
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,71	4,60
11	9,65	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,40	4,29
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,16	4,05
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,85
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,70
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,56
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,45
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,45	3,35
18	8,28	6,01	5,09	4,58	4,25	4,01	3,85	3,71	3,60	3,51	3,37	3,27
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,19
20	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37	3,23	3,13
21	8,02	5,78	4,87	4,37	4,04	3,81	3,65	3,51	3,40	3,31	3,17	3,07
22	7,94	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	3,02
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,97
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,25	3,17	3,03	2,93
25	7,77	5,57	4,68	4,189	3,86	3,63	3,46	3,32	3,21	3,13	2,99	2,89
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,17	3,09	2,96	2,86
27	7,68	5,49	4,60	4,11	3,79	3,56	3,39	3,26	3,14	3,06	2,93	2,83
28	7,64	5,45	4,57	4,07	3,76	3,53	3,36	3,23	3,11	3,03	2,90	2,80
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,08	3,00	2,87	2,77
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,06	2,98	2,84	2,74
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,88	2,80	2,66	2,56
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,40
100	6,90	4,82	3,98	3,51	3,20	2,99	2,82	2,69	2,59	2,51	2,36	2,26
200	6,76	4,71	3,88	3,41	3,11	2,90	2,73	2,60	2,50	2,41	2,28	2,17
400	6,70	4,66	3,83	3,36	3,06	2,85	2,69	2,55	2,46	2,37	2,23	2,12
1000	6,66	4,62	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,20	2,09
∞	6,64	4,60	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	2,07

Az F-eloszlás kvartilis értékei (folytatás)

$$1 - \alpha = 0,99$$

	v_2										
1	6169	6208	6234	6258	6286	6302	6323	6334	6352	6361	6366
2	99,44	99,45	99,46	99,47	99,48	99,48	99,49	99,49	99,49	99,50	99,50
3	26,83	26,69	26,60	26,50	26,41	26,35	26,27	26,23	26,18	26,14	26,12
4	14,15	14,02	13,93	13,83	13,74	13,69	13,61	13,57	13,52	13,48	13,46
5	9,68	9,55	9,47	9,38	9,29	9,24	9,17	9,13	9,07	9,04	9,02
6	7,52	7,39	7,31	7,23	7,14	7,09	7,02	6,99	6,94	6,90	6,88
7	6,27	6,15	6,07	5,98	5,90	5,85	5,78	5,75	5,70	5,67	5,65
8	5,48	5,36	5,28	5,20	5,11	5,06	5,00	4,96	4,91	4,88	4,86
9	4,92	4,80	4,73	4,64	4,56	4,51	4,45	4,41	4,36	4,33	4,31
10	4,52	4,41	4,33	4,25	4,17	4,12	4,05	4,01	3,96	3,93	3,91
11	4,21	4,10	4,02	3,94	3,86	3,80	3,74	3,70	3,66	3,62	3,60
12	3,98	3,86	3,78	3,70	3,61	3,56	3,49	3,46	3,41	3,38	3,36
13	3,78	3,67	3,59	3,51	3,42	3,37	3,30	3,27	3,21	3,18	3,16
14	3,62	3,51	3,43	3,34	3,26	3,21	3,14	3,11	3,06	3,02	3,00
15	3,48	3,36	3,29	3,20	3,12	3,07	3,00	2,97	2,92	2,89	2,87
16	3,37	3,25	3,18	3,10	3,01	2,96	2,89	2,86	2,80	2,77	2,75
17	3,27	3,16	3,08	3,00	2,92	2,86	2,79	2,76	2,70	2,67	2,65
18	3,19	3,07	3,00	2,91	2,83	2,78	2,71	2,68	2,62	2,59	2,57
19	3,12	3,00	2,92	2,84	2,76	2,70	2,63	2,60	2,54	2,51	2,49
20	3,05	2,94	2,86	2,77	2,69	2,63	2,56	2,53	2,47	2,44	2,42
21	2,99	2,88	2,80	2,72	2,63	2,58	2,51	2,47	2,42	2,38	2,36
22	2,94	2,83	2,75	2,67	2,58	2,53	2,46	2,42	2,37	2,33	2,31
23	2,89	2,78	2,70	2,62	2,53	2,48	2,41	2,37	2,32	2,28	2,26
24	2,85	2,74	2,66	2,58	2,49	2,44	2,36	2,33	2,27	2,23	2,21
25	2,81	2,70	2,62	2,54	2,45	2,40	2,32	2,29	2,23	2,19	2,17
26	2,77	2,66	2,58	2,50	2,41	2,36	2,28	2,25	2,19	2,15	2,13
27	2,74	2,63	2,55	2,47	2,38	2,33	2,25	2,21	2,16	2,12	2,10
28	2,71	2,60	2,52	2,44	2,35	2,30	2,22	2,18	2,13	2,09	2,06
29	2,68	2,57	2,49	2,41	2,32	2,27	2,19	2,15	2,10	2,06	2,03
30	2,66	2,55	2,47	2,38	2,29	2,24	2,16	2,13	2,07	2,03	2,01
40	2,49	2,37	2,29	2,20	2,11	2,05	1,97	1,94	1,88	1,84	1,81
60	2,32	2,20	2,12	2,03	1,93	1,87	1,79	1,74	1,68	1,63	1,60
100	2,19	2,06	1,98	1,94	1,79	1,73	1,64	1,59	1,51	1,46	1,43
200	2,09	1,97	1,88	1,79	1,69	1,62	1,53	1,48	1,39	1,33	1,28
400	2,04	1,92	1,84	1,74	1,64	1,57	1,47	1,42	1,32	1,24	1,19
1000	2,01	1,89	1,81	1,71	1,61	1,54	1,44	1,38	1,28	1,19	1,11
∞	1,99	1,87	1,79	1,69	1,59	1,52	1,41	1,36	1,25	1,15	1,00
v_2	16	20	24	30	40	50	75	100	200	500	∞