Matematikai statisztikai elemzések 1.

A statisztika alapfogalmai, feladatai, táblázatok, grafikus ábrázolás. Leíró statisztika, osztályozás, gyakorisági sorok

Prof. Dr. Závoti, József

Matematikai statisztikai elemzések 1.: A statisztika alapfogalmai, feladatai, táblázatok, grafikus ábrázolás. Leíró statisztika, osztályozás, gyakorisági sorok

Prof. Dr. Závoti, József
Lektor: Bischof , Annamária

Ez a modul a TÁMOP - 4.1.2-08/1/A-2009-0027 „Tananyagfejlesztéssel a GEO-ért" projekt keretében készült. A projektet az Európai Unió és a Magyar Állam 44 706 488 Ft összegben támogatta.

v 1.0

Publication date 2010

Kivonat

Ez a modul a statisztika alapfogalmaival és fő feladataival ismerteti meg az olvasót. Megismerheti a leíró és következtetéses statisztika legfontosabb eszközeit, elsajátíthatja gyakorisági sorok kezelését. A statisztikai skálák bevezetése segít eligazodni az adatok megadási módjaiban. A modul lényeges eleme a grafikus ábrázolások értelmezésének elsajátítása. Nagy hangsúlyt fektetünk az adatok osztályokba sorolásának gyakorlati megvalósítására.

Tartalom

1. fejezet - A statisztika alapfogalmai, feladatai, táblázatok, grafikus ábrázolás. Leíró statisztika, osztályozás, gyakorisági sorok

1. 1.1 Bevezetés

Jelen modul a Matematikai statisztikai elemzések tárgy első fejezete, modulja. Amennyiben az itt leírtak nem elegendőek a témakör megértéshez, akkor forduljon a szerzőhöz segítségért.

Jelen modul célja, hogy az Olvasó megismerkedjen a statisztika alapfogalmaival, és képessé váljon adatok osztályokba sorolására, összetettebb számítási feladatok elvégzésére.

Ez a modul egyszerű módon ismerteti meg az olvasót a statisztika legfontosabb eszközeivel, amelyek a mindennapi életben történő események megértéséhez nélkülözhetetlenek. A statisztika egy empirikus tudomány, a gyakorlati élet hozta létre. Legegyszerűbb eszközeit (táblázatok, grafikonok,...) az adatrendszerek szemléltetésére használjuk. Nagyon fontos, hogy nagy mennyiségű adatot mi magunk is osztályokba tudjunk sorolni, azért, hogy az adatrendszerhez valamilyen értelmezést tudjunk rendelni.

2. 1.2 A statisztika alapfogalmai és főbb feladatai

2.1. 1.2.1 Leíró és következtetéses statisztika

A **matematikai statisztika** a véletlen tömegjelenségek statisztikai törvény-szerű-ségeit vizsgálja.

A statisztika egyszerűbb problémáit a **leíró statisztika** keretein belül lehet kezelni:

- Adatok ábrázolása
- Grafikonok szerkesztése
- Táblázatok készítése
- Egyszerű paraméterek számolása (átlagértékek, szóródás)
- Indexszámítás
- Koncentráció-számítás

A matematikai statisztika gyakorlati használhatósága a valószínűségszámítás elméleti alapjain nyugszik. **Minőségellenőrzés** során sokszor nincs mód a teljes sokaságot átvizsgálni, hanem csak egy n-elemű **minta** alapján kell következtetéseket levonnunk. A mintaelemekből célszerű olyan függvényeket konstruálni, amelyek jó információt nyújtanak az egész eloszlásra. Tapasztalati adatokból, u.n. mintából következtetünk események valószínűségeire, vagy valószínűségi változók ismeretlen eloszlás-és sűrűségfüggvényeire.

A statisztika másik nagy területe az **induktív statisztika** (következtetéses statisztika):

- Becslések
- Tesztek
- Döntéselmélet
- Többváltozós statisztikai módszerek

A statisztika felhasználási területe az **extrapoláció** (predikció): A jelenlegi adatok alapján a jövőre nézve statisztikai prognózisokat lehet készíteni, feltételezve, hogy a körülmények, feltételek azonosak maradnak. Ilyen prognózisok készülnek a következő évi energia felhasználásra, az adóbevételre, a népességszám alakulására, a munkanélküliségre, stb.

2.2. 1.2.2 A statisztikai adatok forrásai és feldolgozása

Az **adatok forrása** szerint megkülönböztetünk:

- hivatalos statisztikai adatokat, amelyeket a Központi Statisztikai Hivatal (KSH) évkönyvben, folyóiratokban tesz közzé
- nem hivatalos statisztikai adatokat, ilyenek az ipari és kereskedelmi kamarák jelentései, a különböző közvélemény-kutató intézetek felmérései, a nagyvállalatok mérlegei.

A statisztikai **adatfeldolgozás lépései**:

1. Tervezés
2. Mintavételezés (elsődleges - másodlagos)
    - Kérdőív: olcsó – de általában kevés jön vissza
    - Interjú: drága – kvalifikált személyek szükségesek a felméré-sekhez
    - Megfigyelés: pl.: forgalomszámlálás
    - Kísérlet: pl.: a közgazdaságtanban az áruteszt
    - Automatikus rögzítés: vonalkódok a bevásárlóközpontokban vagy a telefonközpontok működése
3. Előkészítés: táblázat – grafikon szerkesztése
4. Analízis: matematikai statisztikai módszerek bevetése
5. Interpretáció: eredmények értékelése

Alapadatok: Sokaság (populáció)

Pl.: egy cég számlái 2009. szeptember 10-én, halálos balesetek száma 2008-ban

2.3. 1.2.3 Statisztikai skálák, információ szintek:

1. Nominális skála: nincs természetes sorrend, mellérendeltség

pl.: vallások, nemek, színek

1. Ordinális (Rang, sorrendi) skála: van sorrend, létezik rendezés

pl.: iskolai jegyek, futball bajnokság

1. Intervallum skála: nullpont választása önkényes

pl.: hőmérséklet ($20°C \neq 2 \cdot 10°C$), időszámítás

1. Arány skála**:** létezik abszolút 0 pont

pl.: magasság, kor, jövedelem

Az adatokat transzformációnak vethetjük alá úgy, hogy a meglévő viszonyok nem változnak.

2.4. 1.2.4 A statisztikai sokaság

Sokaság: a vizsgálat tárgyát képező egységek összessége, vagyis azon jelenségek összessége, amire a megfigyelés irányul.

Csoportosítási lehetőségek:

1. Álló (időpont, stock) és mozgó (időtartam, flow) sokaság:

Példák:

Álló sokaság:

- Magyarország lakossága 2010. jan. 1-jén
- Raktár állománya 2010. szept. 10-én
- Pénztári bevétel 2010 szept. 10-én

Mozgó sokaság

- Születések száma 1 év alatt
- Egy bankba 1 nap alatt befizetett összegek

1. Véges és végtelen sokaság:

Példák:

Véges sokaságok:

- Egy teremben egy órán ülő hallgatók száma
- Gépkocsi állomány Magyarországon 2010-ben

Végtelen sokaság:

- Égen lévő csillagok száma

1. Alapsokaság és részsokaság

Példák:

alapsokaság: Magyarország lakossága

részsokaság: Magyarország férfi lakosainak száma, Vas megye lakosainak száma, angolul beszélő magyarok száma

1. Teljes- és mintasokaság:

Példák:

teljes sokaság: Magyarország lakossága

mintasokaság: háztartási jövedelmi felvétel során megkérdezett lakosok

1. Valós (létező) és fiktív (elképzelt) sokaság:

Példák:

valós sokaság: 2010-ben született gyerekek száma

fiktív sokaság: 2010-ben megszület*hető* gyerekek száma

1. Aggregált sokaság: Különböző fajta, minőségileg eltérő, de együtt vizsgált elemek

Aggregátum: értékben összegzett mennyiség pl.: húsfogyasztás 2009-ben.

2.5. 1.2.5 Statisztikai ismérvek

Definíció:

Ismérvek (karakterisztikus tulajdonságok): a sokaság egyedeinek jellemző tulajdonsága, olyan vizsgálati szempontok, amelyek alapján egy sokaság át nem fedő részekre bontható.

Példa:

Ismérv:	**Ismérvadat:**
kor	év
nem	férfi
magasság	cím
kereset	Ft

Ismérvek fajtái:

- Időbeli pl.: születési idő, érettségi ideje
- Területi pl.: születési hely, nyaralás helyszíne
- Tárgyi:- minőségi pl.: szemszín, iskolai végzettség

- mennyiségi pl.: magasság, testvérek száma

2.6. 1.2.6 Diszkrét és folytonos ismérv adatok

Megkülönböztethetünk **diszkrét** és **folytonos ismérv** adatokat.

Példák:

Diszkrét ismérv pl.: hallgatók száma, üzem dolgozói

Folytonos ismérv pl.: egy asztal hossza, súly stb.

3. 1.3 Statisztikai tábla

Definíció:

Statisztikai tábla: Statisztikai sorok összefüggő rendszere a megfelelő külső formával együtt.

Statisztikai sor: Olyan sor, amit akkor kapunk, ha a sokaságot valamilyen ismérv szerint sorba rendezünk.

Statisztikai tábla **dimenzió**ja: az a szám, amely jelzi, hogy a tábla egye-egy adata hány statisztikai sorhoz tartozik.

Statisztikai táblák csoportosítása:

1. egyszerű tábla: összehasonlító vagy leíró sorokat tartalmaz, nincs összesen sor/oszlop
2. csoportosító tábla: egy irányban van csoportosítás valamely ismérv alapján, ebben az irányban egy összesen sor/oszlop is megtalálható

3. kombinációs tábla: legalább két irányban csoportosít, azaz minimum két összesen sor/oszlop van

4. 1.4 Grafikus ábrázolási módok

A rendelkezésünkre álló adatok legegyszerűbb elemzési módja, az ábrázolás. Az ember a vizuális élményt könnyebben feldolgozza, mint a „száraz” számokat, valamint azon kívül, hogy látványos, igen informatív is tud lenni.

4.1. 1.4.1 Mértani alakzatok

Grafikon: a változók kapcsolatának képszerű bemutatása

Az adatok grafikonok segítségével való ábrázolása történhet koordináta rendszeren belül és kívül is. A koordináta rendszer sok plusz információval szolgálhat, ám vannak olyan esetek, amikor nem feltétlenül van rá szükség.

A leggyakrabban használt grafikon típusok:

1) Oszlopdiagram (oszlopgrafikon) – a különböző oszlopok nem érhetnek össze, az oszlopok jelölik az adott mutató(k) nagyságát

2) Hisztogram – az oszlopoknak össze **kell** érniük

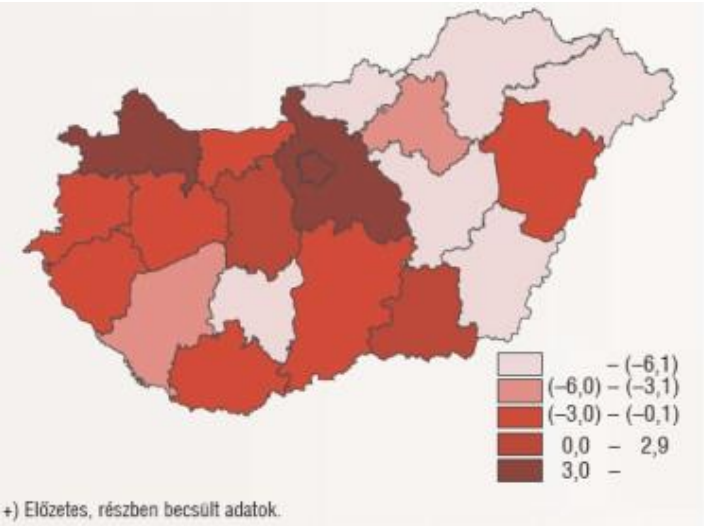
3) Pontdiagram – az adott *x* értékhez tartozó *y* értékek

4) Vonaldiagram (vonalgrafikon)

5) Kördiagram/tortadiagram – általában a relatív gyakoriságok ábrázolására használatos

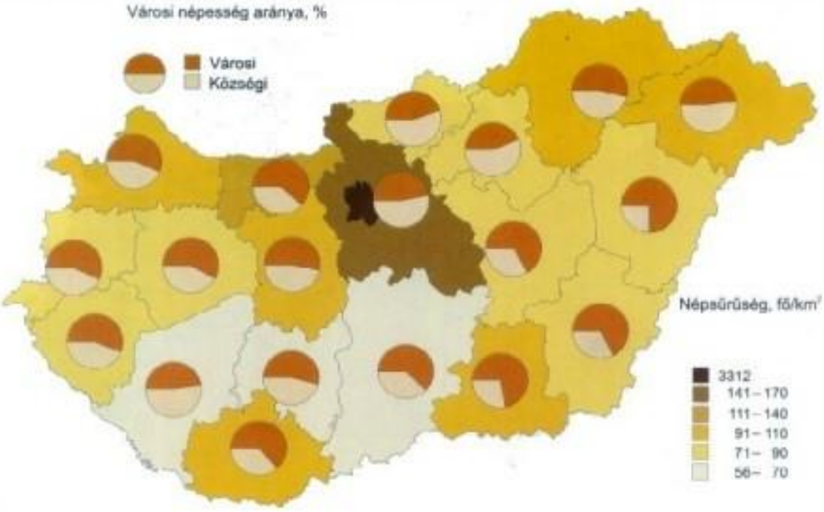
4.2. 1.4.2 Térképen alapuló ábrázolás

1) Kartogram – a színezés jelzi az adott mutató nagyságát



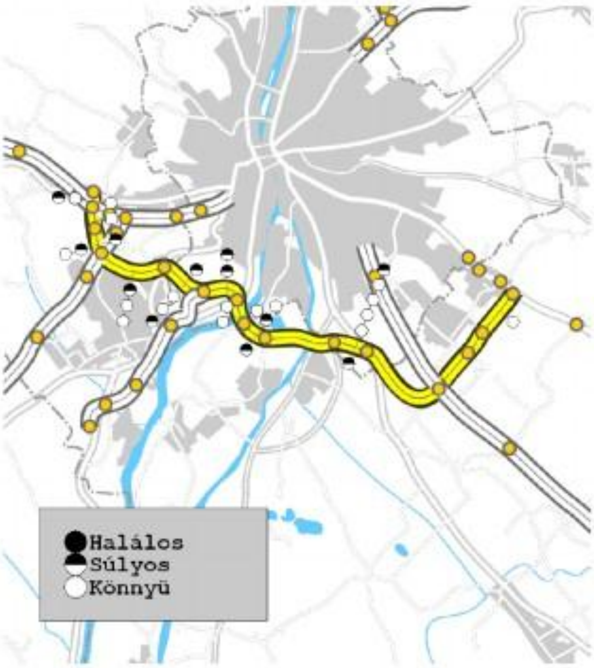
1. ábra: Ezer lakosra jutó belföldi vándorlási különbözet 2009., Forrás: Statisztikai tükör 2010/42

2) Kartodiagram – A térkép különböző pontjain diagramokat helyeznek el, ezzel szemléltetve a mutató nagyságát



2. ábra: Népsűrűség és városi népesség aránya 2002.január 1. Forrás: sdt.sulinet.hu

3) Ponttérkép – egyfajta sűrűségi mutató, a pontok mennyisége jelzi a mutató nagyságát



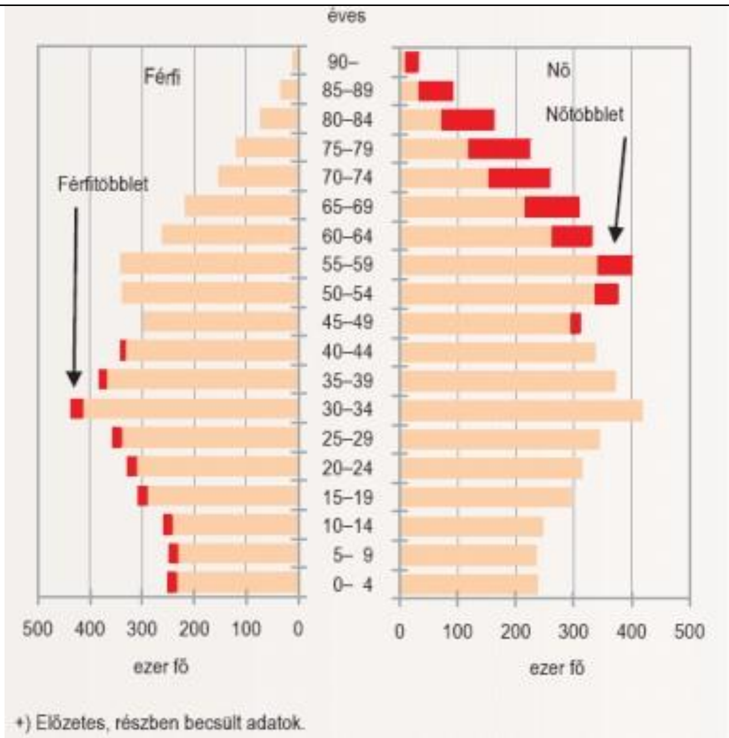
3. ábra: M0 autóbaleseti ponttérkép 2007. I-VIII.hó Forrás: www.police.hu

4.3. 1.4.3 Piktogram

Piktorgam(piktográf) – a nagyság szemlélteti a mutató értékét, vagy egy egység egy piktogram és a piktogramok száma a mutató értékével egyezik meg

4.4. 1.4.4 Korfa

Korfa – a népesség nem és kor szerinti megoszlását szemléltető grafikus eszköz, az ábra jobb oldalán a nők, a bal oldalán pedig a férfiak adott korévhez tartozó számát megjelenítő oszlopdiagramokon.



4. ábra: Forrás: Statisztikai tükör 2010/42

5. 1.5 Gyakorisági eloszlások

Legyen adott *N* elemű sokaság metrikus skálán.

Legyenek adottak a következő ismérvek: $x_1, x_2, \dots, x_k$

Jelölje $f_1, f_2, \dots, f_k,$ ahányszor x_i előfordul

Abszolút gyakoriság: $f_i \; i = 1, 2, \dots, k \; 0 \le f_i \le N \mid f_1 + f_2 + \dots + f_k = N$

Relatív gyakoriság: $g_i = \dfrac{f_i}{N} \; i = 1, 2, \dots, k \; 0 \le g_i \le 1$

Ekkor fennáll: $\sum_{i=1}^{k} g_i = 1$

Abszolút gyakorisági összeg (kumulált gyakorisági sor):

$$F_i = f_1 + \dots + f_i = \sum_{j=1}^{i} f_j \quad i = 1, 2, \dots, k$$

F_i azon elemek száma, amelyek legfeljebb x_i ismérvvel rendelkeznek.

Relatív gyakorisági összeg:

$$G_i = g_1 + \dots + g_i = \sum_{j=1}^{i} g_j \quad i = 1, 2, \dots, k$$

$$G_i = \frac{F_i}{N}$$

Kapcsolat a gyakorisági és a kumulált gyakorisági sorok között:

$$f_i = F_i - F_{i-1}$$

$$g_i = G_i - G_{i-1}$$

$$F_0 = G_0 = 0$$

6. 1.6 Gyakorisági eloszlás osztályokba sorolt adatokra

Ha nagyon sok ismérv van, vagy folytonos ismérvek vannak, akkor célszerű **osztályokat képezni** az ismérvekre.

Egy adatrendszer feldolgozásánál alapvető probléma, hogy hány osztály képezzünk? Általában célszerű 20 osztálynál kevesebbet választani. Ha sok osztályt választunk, töredezett a hisztogram, ha kevés osztállyal dolgozunk, akkor pedig durva lesz a felbontás.

Az osztályhatárokat úgy kell megválasztani, hogy minden elem belekerüljön valamelyik osztályba (teljes), minden elem csak egy osztályba kerüljön (diszjunkt, átfedésmentes), és lehetőleg homogén osztályok legyenek.

Az **osztályok számának (k) meghatározására** a szakirodalomban általában kétféle módszert javasolnak. Megállapodás kérdése, hogy melyiket választjuk. Kevés adatszámra mindkét módszer közel azonos osztályszámot szolgáltat.

1. legkisebb *k*, amelyre $2^k > N$
2. Sturges-képlet: $k = [1 + 3.3 \lg N]$,

ahol *N* az osztályozni kívánt adatok száma (minta, vagy sokaság elemszáma)

Egyenközű osztályszélesség esetén minden **osztály hossza**: $\Delta x = \frac{x_{max} - x_{min}}{k}$.

De választhatunk különböző osztályszélességeket is.

Jelölések:

x_i^a : az *i*-edik osztály alsó határa

x_i^f : az *i*-edik osztály felső határa

Az **osztályhatárok meghatározása** történhet az alábbi szabályok szerint:

$$x_1^a = x_{min} - \frac{adatpontoság}{2} \quad x_1^f = x_1^a + \Delta x = x_2^a$$

$$x_k^a = x_1^a + (i-1) \cdot \Delta x \quad i = 2, 3, \ldots, k \quad x_k^f = x_k^a + \Delta x = x_1^a + k \cdot \Delta x$$

Az **osztályközép** képzési szabálya: $x_i' = \frac{x_i^f + x_i^a}{2} \quad i = 1, 2, \ldots, k$

Gyakorisági és sűrűséghisztogram

Gyakorisági hisztogramm szerkesztése

Tegyük fel, hogy az *a,b* intervallum lefedi a mintaterjedelmet.

Osszuk fel az *a,b* intervallumot n részre:

$$a = d_0 < d_1 < \ldots < d_n = b$$

A részintervallumok *n* számára nincs általános szabály, általában 6-12 részintervallumot képezzünk.

Adjuk meg az egyes d_{i-1},d_i részintervallumba eső mintaelemek k_i számát ($i=1,2,\ldots,n$) és mindegyik részintervallumra rajzoljunk az oda eső mintaelemek gyakoriságával arányos magasságú téglalapot: az *i*-edik részintervallumra rajzolt téglalap magassága legyen

$$\frac{k}{d_i - d_{i-1}}$$

Ekkor a téglalapok területeinek összege *n*.

Sűrűséghisztogram szerkesztése

Az egyes intervallumokra rajzolt téglalapok magasságát az oda eső mintaelemek relatív gyakoriságával adjuk meg, azaz az *i*-edik részintervallum magassága legyen

$$\frac{k}{n \cdot (d_i - d_{i-1})}$$

Az így kapott lépcsős függvény a tapasztalati sűrűségfüggvény, amely közelíti az ismeretlen elméleti sűrűségfüggvényt. Ha ez a hisztogram egy haranggörbét közelít, akkor az eloszlást jó közelítésben normálisnak tekinthetjük.

Tapasztalati eloszlásfüggvény:

Az $x_1,x_2,\ldots,x_n$ mintához tartozó tapasztalati eloszlásfüggvény az x-tengellyel párhuzamos szakaszokból álló lépcsős függvény, amelynek minden egyes felvett x_i értékénél *1/n* ugrás van, ha x_i-t egyszer kaptuk a mintában; *k/n* ugrás van, ha *k*-szor fordul elő x_i a mintában. A minta eloszlásfüggvénye a minta elemszámának növelésével minden x-re egyenletesen konvergál az elméleti eloszlásfüggvényhez.

7. 1.7 Összefoglalás

1. A 2009-es beruházások nemzetgazdasági ágak szerint Magyarországon (Forrás:KSH):

251.5, 13.4, 907.9, 205.3, 117.0, 106.7, 337.1, 787.5, 41.2, 182.4, 47.7, 981.1, 68.4, 86.1, 158.5, 88.2, 87.7, 40.0, 25.4

Számítsa ki és értelmezze a helyzetmutatókat (átlag, módusz, medián)!

1. A 2009-es beruházások nemzetgazdasági ágak szerint Magyarországon (Forrás: KSH):

251.5, 13.4, 97.9, 205.3, 117.0, 106.7, 337.1,787.5, 41.2, 183.4, 47.7, 981.1, 68.4, 86.1, 158.5, 88.2, 57.7, 40.0, 25.4

Végezze el az osztályba sorolást! Készítsen hisztogramot!

1. Az Express újságban 1995. 10. 04.-én eladásra kínált 70 m2 körüli lakások ára (mFt):

2.0, 4.0, 3.1, 3.4, 4.2, 6.0, 3.6, 3.1, 2.6, 3.3, 3.4, 3.5, 2.4, 3.2, 3.8, 3.1, 5.3, 2.5, 3.6, 3.0, 3.5, 3.5, 4.1.

Végezze el az osztályba sorolást! Készítsen hisztogramot!

1. Egy vállalkozásnál az azonos termékeket előállító dolgozók napi teljesítménye db-ban:

90, 98, 92, 94, 101, 103, 99, 96, 94, 100, 98, 92, 96, 91, 104, 100, 99, 97, 93, 102, 101, 96, 93, 88, 97

Végezze el az osztályba sorolást! Készítsen hisztogramot!

1. Egy közkedvelt gyorsétterem-hálózat egyik egységében megfigyelték a kiszolgálási időt (mp):

45	48	49	56	61	66	66	66	70	72
72	75	78	79	81	81	83	95	102	135

Határozza meg ugyanezen értékeket osztályozással! Készítsen hisztogramot!

Irodalomjegyzék

Csanády V., Horváth R., Szalay L.: *Matematikai statisztika,* EFE Matematikai Intézet, Sopron, 1995

Csernyák L.: *Valószínűségszámítás,* Nemzeti Tankönyvkiadó, Budapest, 1990

Denkinger G. : *Valószínűségszámítás,* Nemzeti Tankönyvkiadó, Budapest, 1978

Hunyadi-Vita : *Statisztika közgazdászoknak,* KSH, Budapest, 2002,

Keresztély-Sugár-Szarvas: *Statisztika példatár közgazdászoknak,* BKE, Nemzeti Tankönyvkiadó, 2005

Korpás A.: *Általános statisztika I-II.,* Nemzeti Tankönyvkiadó, Budapest, 1996

Obádovics J. Gy.: *Valószínűségszámítás és matematikai statisztika,* Scolars Kiadó, Budapest, 2003

Reimann J. - Tóth J.: *Valószínűségszámítás és matematikai statisztika,* Tankönyvkiadó, Budapest, 1991

Solt Gy.: *Valószínűségszámítás,* Műszaki Könyvkiadó, Budapest, 1971

Závoti-Polgárné-Bischof (2009): *Statisztikai képletgyűjtemény és táblázatok,* NYME Kiadó, Sopron, 2009