8. Adatsokaságok jellemzői, a valószínűség számítás elemei

Vázlat:

1. Mit jelent a statisztika szó ? Mi a statisztikus feladata?
2. Statisztikai sokaság, minta
   a) Adatsokaságok jellemzői: a statisztikai adat, relatív gyakoriság,
   b) Táblázatok . Osztályba sorolás
   c) gyakoriság diagramok (vagy hisztogramok) kördiagram sávdiagram oszlopdiagram vonaldiagram
   d) Statisztikai mutatók: középértékek**(**átlag, adatsokaság mértani közepe, $H \leq G \leq A \leq N$ medián , módusz, a szórások (Terjedelem, átlagos abszolút eltérés, szórás)
3. A valószínűség: Esemény, eseménytér, relatív gyakoriság, esemény valószínűsége, a valószínűség tulajdonságai**,** valószínűségi változó,

4. Bizonyítható tételek: Számtani –mértani közép közti egyenlőtlenség, Valószínűség re vonatkozó tételek közül valamelyik

5. Alkalmazások

Mit jelent a statisztika szó ? Szűkebb értelemben információk rendezett összessége, tágabb értelemben a véletlen tömegjelenségek vizsgálatára vonatkozó módszerek összessége

Mi a statisztikus feladata? Kap egy feladatot , megtervezi az adatgyűjtést , mintavételt, adatokat gyűjt illetve másokkal gyűjtet, kiértékelés módját megtervezi (A Központi Statisztikai Hivatalt hazánkban 1867-ben hozták létre , azóta végeznek népszámlálást, adatgyűjtést, személyes adatokra vonatkozó adatgyűjtést csak törvényben lehet elrendelni- adatszolgáltatás önkéntes. A statisztikai munka során egy nagy elemszámú halmaz elemeinek tulajdonságairól kívánunk tájékozódni.
Statisztikai sokaság, vagy populáció –azoknak a dolgoknak, egyedeknek csoportja amelyekről adatokat gyűjtünk. A sokaság elemei az egyedek. Nem mindig van arra lehetőség, hogy egy vizsgálandó területen minden egyedről felvegyük az adatokat. Ilyenkor a vizsgálandó egyedek közül egy **mintát**, vagyis az eredeti statisztikai sokaság egy részhalmazát választjuk ki, s ezt vizsgáljuk, mint statisztikai sokaságot. Olyan mintavétel az ideális, ahol a vizsgált tulajdonság előfordulása közelítőleg olyan arányban fordul elő a mintában, mint az eredeti statisztikai sokaságban. Az ilyen mintát nevezik: **reprezentatív mintának** (például: közvéleménykutatásoknál használják). A mintavételi eljárások közül a legkézenfekvőbb módszer: a **véletlenszerű mintavétel**. Ilyenkor a statisztikai sokaság minden eleme a kiválasztás során ugyanakkora eséllyel kerülhet be a kiválasztott mintába (ilyen például a lottó).Minta: A statisztikai sokaságból kiválasztott olyan rész amelyektől adatokat kapunk. Reprezentatív a minta, a hűen tükrözi a sokaság összetételét. Ismérv: A z egyedek vizsgált tulajdonsága. Az ismérv lehet kvalitatív(

minősítéses) pl. hajszín, vagy lehet kvantitatív (méréses),pl. testmagasság osztályzat, életkor… Az ismérv konkrét értéke az adat Pl, vörös ,szőke,….

Adatsokaságok jellemzői:

A statisztikai elemzések során egyedek (például: emberek, állatok, termékek stb…) összegyűjtött adatait vizsgálják. Ezeknek az egyedeknek halmaza: **a statisztikai sokaság**. A vizsgált tulajdonságok: az **ismérvek/ változók**. A sokaság egyes egyedeinek az ismérv szerinti tulajdonsága a **statisztikai adat**. Például: a statisztikai sokaság lehet egy táncmulatságon résztvevők halmaza. Az egyedek az egyes résztvevők. Az ismérv a résztvevők életkora(ez méréses ismérv). Az adatok ebben az esetben az egyes életkorok . Az egyes adatfajták előfordulási száma az adat gyakorisága. A gyakoriságot az összes megfigyelés számával osztva kapjuk **a relatív gyakoriságot.**
A statisztikus feladata a kiválasztott mintából kapott adatok alapján megbecsülje a statisztikai sokaságból kapható hasonló választ.

Az adatok összegyűjthetők táblázatban:

A leíró statisztika fontos eszközei a **táblázatok** . A táblázatokban az egyes értékek áttekinthetően láthatóak.

Osztályba sorolás: Sokszor előfordul, hogy nagy mennyiségű adat esetén az adatokat nem soroljuk fel egyenként, hanem osztályokba soroljuk őket, így jobban áttekinthetők. Egy adat két osztályban nem jelenhet meg, de mindegyik megjelenik valamelyikben. Egy osztályköz hossza az osztály felső és alsó határának különbsége. Osztályközépnek nevezzük az osztály alsó és felső határának számtani közepét. Úgy tekintjük, mintha minden az osztályba tartozó adat értéke az osztályközép lenne .Az egyes osztályokba tartozó adatok száma az osztály kumulált gyakorisága.

(például: a foglalkoztatottak száma esetén például az olyan vállalatokat, amelyek 100 ezernél kevesebb embert foglalkoztatnak egy osztályba soroljuk. Azokat a vállaltokat, ahol a foglalkoztatottak száma 100 ezernél több és 200 ezernél kevesebb, egy másik osztályba soroljuk stb.)

A táblázat adatai ábrázolhatók , szemléletessé tehetők különböző diagramokon. Nagyobb méretű táblázatnak általában csak kisebb részeit ábrázolják diagramon.

A **diagramok** az adatok gyors áttekintését és az egyes értékek ránézésre történő összehasonlítását teszik lehetővé. A **gyakoriság diagramok (vagy hisztogramo**k) a táblázatoknál rendszerint áttekinthetőbben mutatják a gyakoriság eloszlását

1. **Kördiagram**: alkalmas a %-ban megadott adatok ábrázolására. A teljes kör jelenti a 100%-ot. Akkor használjuk leginkább, ha az adatoknak az egészhez viszonyított arányát akarjuk szemléltetni. Nem érdemes kördiagramot alkalmazni, ha túl sok adat van, mert ebben az esetben a középponti szögek nagyon kicsik lehetnek, így nem könnyű őket összehasonlítani. Rajz!!!!
2. **Sávdiagram Rajz!!!**
3. **Oszlopdiagram**: nem érdemes oszlopdiagramot használni, ha az adatok között van egy olyan kiugróan nagy, amely nem férne rá a grafikonra, vagy egy olyan kicsi, amely összehasonlíthatatlanul eltörpülne a többi adathoz képest. Rajz!!!
4. Szemléletesebb képet ad a változásról a: **töröttvonal-grafikon (vagy vonaldiagram**). Ez leginkább akkor hasznos, ha az adatok időbeli változását vagy egymáshoz való

viszonyát szeretnénk ábrázolni. valamilyen mennyiség időbeli változásának leírására jól használható. Rajz!!!

Az adatok jellemzésére valók a statisztikai mutatók.

A statisztikai számsokaságokat különböző **statisztikai mutatókkal** szokás jellemezni.. Statisztikai mutatók a **középértékek(** helyzetparaméterek) **és** az ezektől való eltérések mérésére való **a szórások** (szórás, átlagos abszolúteltérés)

Középértékek, illetve helyzetparaméterek:

1. **Átlag**: egy számsokaság **számtani közepe (vagy átlaga)** úgy kapható meg, ,hogy a számsokaság összegét elosztjuk a számsokaság darabszámával $x_{átl.} = \frac{x_1 + x_2 + ... + x_n}{n}$.

Ha az egyes adatok gyakorisága k_1 , k_2, … k_n akkor az adatok átlagát megadhatjuk súlyozott számtani középként is. $x_{átl.} = \frac{k_1x_1 + k_2x_2 + ... + k_nx_n}{k_1 + k_2 + ... + k_n}$

Előnye.: A nála nagyobb adatoktól való eltéréseinek összege ugyanannyi, mint a nála kisebb adatoktól való eltéréseinek összege

Hátránya: Egy –egy kiugró adat nagyon eltorzíthatja.

2. Adatsokaság **mértani közepe**: Két pozitív valós szám **mértani közepe** a szorzatuk négyzetgyöke.(Két szám mértani közepe ugyanannyiszorosa az egyik számnak amekkora része a másik számnak.)

Beszélhetünk n darab pozitív valós szám (adatsokaság) négyzetgyökéről is, ez az adott a_1, a_2… a_n számok szorzatának n-ik gyöke.

Tétel: Két pozitív szám (a;b) számtani közepe nagyobb a két szám mértani közepénél vagy egyenlő vele.

Bizonyítás: $\frac{a+b}{2} \geq \sqrt{a \cdot b}$ ezt kell bizonyítani. Ez pontosan akkor igaz,

ha $\frac{a+b}{2} - \sqrt{a \cdot b} \geq 0$ Ez pedig igaz, mert $\frac{a+b}{2} - \sqrt{a \cdot b} = \frac{a + b - \sqrt{ab}}{2} = \frac{(\sqrt{a} - \sqrt{b})^2}{2}$

. A kapott kifejezés pedig nyilvánvalóan nem negatív, mivel a tört számlálója teljes négyzet, nevezője pedig pozitív szám. A bizonyításból látszik egyenlőség pontosan akkor van ha a=b.

A Thálesz-tétel és a derékszögű háromszögben érvényes magasság-tétel felhasználásával az egyenlőtlenség geometria úton is bizonyítható. Thálesz-kör átmérője legyen a+b sugár nagyobb vagy egyenlő, mint a kör bármely húrjának fele.

(Értelmeztük 2 pozitív valós szám négyzetes közepét és harmonikus közepét is.
Definíció: két pozitív valós szám négyzetes közepe a két szám négyzete számtani közepének a négyzetgyöke $\sqrt{\frac{a^2 + b^2}{2}}$ (n darab valós szám négyzetes közepéről is beszélhetünk. Ez az adott számok négyzetei, számtani közepének négyzetgyöke.)

Definíció: két pozitív valós szám harmonikus közepe a reciprokuk számtani közepének reciproka $\frac{1}{\frac{\frac{1}{a}+\frac{1}{b}}{2}}=\frac{2}{\frac{1}{a}+\frac{1}{b}}=\frac{2ab}{a+b}$ (ez is értelmezhető n darab poz valós számra is).

A közepek összehasonlítása bebizonyítható: $H \le G \le A \le N$ „=” akkor és csak akkor ha a vizsgált számok egyenlők.

2. **Medián**: Az adatsokaságok mediánja a nagyság szerint rendezett adatok közül a középső, ha az adatok száma páratlan, és ha az adatok száma páros, akkor a medián a két középső szám számtani közepe.(2n+1 db adat mediánja az n+1-edik adat, 2n db adat mediánja az n. és n+1. átlaga.
Előnye: ugyanannyi adat kisebb nála, mint amennyi nagyobb
A mediánnak az adatoktól mért távolságainak összege minimális
(pl. egy vállalatnál a fizetések átlagát eltorzíthatja a vezetők magas bére, így a fizetések középértékét a medián jobban jellemzi.
3. **Módusz**: A statisztikai sokaságban leggyakrabban előforduló érték a módusz.
Előnye: könnyen meghatározható, jó eséllyel lehet tippelni az adatokra.
Hátránya : Egy adatot kiemel, többiről nem mond semmit. Nem használható, ha az adatok előfordulási száma nem jellemző

A középértékek sokszor félrevezető információkat adnak egy adathalmazról már csak azért is, mert magából a középértékből nem látszik, hogy az egyes értékek hogyan helyezkednek el a középértékek körül, épp erről adnak információt a **szóródási mértékek**.

Terjedelem, átlagos abszolút eltérés, szórás

A szóródást jellemzi valamennyire az adathalmaz legnagyobb adata, a legkisebb adata és a terjedelme. Egy adathalmaz **terjedelme** az adathalmaz legnagyobb adatának és legkisebb adatának különbsége. Minél kisebb a minta terjedelme, annál jobban jellemzi az adatsokaság középértéke a mintát. Előnye: könnyen meghatározható. Hátránya: szélsőséges adatok eltorzíthatják-elszokták hagyni gyakran a minta alsó és felső negyedét

Egy adathalmaz egyes adataiból egy adathalmaz valamely középértékét kivonva kapjuk az adatoknak a középértéktől való **eltérését.** Ezek abszolút értéke az **abszolút eltérés**. Az így kapott értékek számtani közepe az **átlagos abszolút eltérés**.
Egy adathalmaz elemei legyenek x_1;x_2; ….x_n, az adathalmaz valamely középértéke pedig :k - ekkor az adathalmaz átlagos abszolút eltérése : $S_n=\frac{|x_1-k|+|x_2-k|+....+|x_n-k|}{n}$

Sn minimális, ha k a medián.
Egy adathalmaz egyes adatai és az adathalmaz **átlaga** különbségei négyzeteinek a számtani közepe a **szórásnégyzet**. $D_n{}^2=\frac{(x_1-x_{átl})^2+(x_2-x_{átl})^2+...+(x_n-x_{átl})^2}{n}$

A szórás (Dn) a szórásnégyzet négyzetgyöke: (képlet)
(Minta szórásnégyzete az adatsokaság átlagától való eltérések négyzetének átlaga.)

A valószínűség:

Tapasztalataink mutatják, hogy a véletlen jelenségek körében is érvényesülnek bizonyos törvényszerűségek. Ezeket akkor észleljük, ha ugyanazt a jelenséget nagyon sokszor lényegében ugyanolyan körülmények között figyeljük meg. A valószínűségszámítás e véletlen tömegjelenségek vizsgálatával foglalkozik.

Esemény, eseménytér:

Egy valószínűségi kísérlet lehetséges kimenetelei: **az elemi események**.
Az elemi események halmaza az **eseménytér.** A kísérlet eseményei az eseménytér részhalmazai. Például: ha két pénzérmét feldobunk, és azt nézzük, hogy fejre (F), vagy írásra(I) esik akkor 4 elemi esemény van: FF, FI, IF, II., kockadobásnál 6 db elemi esemény,
Eseménynek nevezzük az eseménytér részhalmazait , pl kockadobásnál: esemény: páros számot dobunk. Egy esemény bekövetkezik, ha a kísérlet kimenetele az eseménynek megfelelő részhalmazba tartozó elemi esemény.
Lehetetlen esemény az az esemény amely sohasem következik be.
Biztos esemény egy kísérletnél az az esemény amely mindenképpen bekövetkezik, Pl, A= legfeljebb 6-ost dobunk jele :I
Egy A esemény **komplementer eseménye** az A(komplementer) esemény, amely pontosan akkor következik be, amikor az A esemény nem következik be.
Tetszőleges A és B események esetén a két esemény ***összege*** az A+B esemény, amely pontosan akkor következik be, ha az A és B esemény közül legalább az egyik bekövetkezik.
Különbsége az A-B esemény, amely pontosan akkor következik be, ha az A bekövetkezik de a B nem.
Szorzata az AB, amely pontosan akkor következik be, ha az A és B esemény is bekövetkezik.
Ha AB=0, a két eseményt ***egymást kizáró eseményeknek nevezzük***

Egy kísérletet n-szer elvégezve az A esemény k-szor következik be, akkor definiálható az A esmény bekövetkezésének gyakorisága és relatív gyakorisága. Gyakoriság: k

Relatív gyakoriság $\frac{k}{n}$

Tulajdonságok:
$0 \leq k \leq n,$
$n > 0$ a biztos esemény relatív gyakorisága 1.
$0 \leq \frac{k}{n} \leq 1$

Egy adott A esemény valószínűségének azt a számot nevezzük, ami körül a relatív gyakoriság ingadozik. Jele: P(A).

A valószínűség tulajdonságai:

1. Tetszőleges A esemény esetén: $0 \leq P(A) \leq 1$
2., P(I)= 1
3. Ha az A és B két tetszőleges egymást kizáró esemény (AB=0),akkor P(A+B)= P(A) +P(B).
Bebizonyíthatók a fenti 3 tulajdonság alapján a következők:

- P(0)=0 Bizonyítás: A biztos esemény és lehetetlen esemény szorzata a lehetetlen esemény . Összegük a biztos esemény, erre alkalmazva a 3. tulajdonságot adódik az állítás
- Ha az A és B két tetszőleges esemény, akkor P(A+B) =P(A)+P(B)- P(AB).
- P(A)+P(A komplementer)=1 vagyis P(A komplementer)=1 –P(A)Biz: A és A komplementer szorzata a lehetetlen esemény. A és A komplementer összegének valószínűsége a 3. tulajdonság alapján….. Felhasználva, hogy A és komplementerének összege a biztos esemény ,adódik az állítás)
-

Egy kísérlet során minden elemi eseményhez hozzárendelhetünk egy valós számot, ily módon az elemi események halmazán értelmeztünk egy függvényt. Ezt a függvényt nevezzük valószínűségi változónak. Ha véges eseményrendszert képez le a valós számok halmazába: diszkrét valószínűségi változónak nevezzük. A kimenetelekhez rendelt számokat a valószínűségi változó értékeinek nevezzük. Ha megadjuk a valószínűségi változó egyes értékei mellé azok bekövetkezéseinek valószínűségeit, akkor a valószínűségi változó eloszlásait adtuk meg. Tanultunk egyenletes, binomiáli, hipergeometrikus eloszlású valószínűségi változóról A valószínűségi változónak értelmeztük a várható értékét, szórását A statisztika és a valószínűségelmélet egyes fogalmai között szoros kapcsolat van-. Relatív gyakoriság- valószínűség, átlag- várható érték.

Alkalmazások:

A mértani közepet használjuk például a magasság-tételben. Derékszögű háromszögben az átfogóhoz tartozó magasság mértani közepe az átfogó két szeletének. Illetve a befogó-tételben.
Különböző gyakorlati problémák esetén használjuk még például 120m kerítéssel legfeljebb mekkora területű téglalap alakú telket lehet körülkeríteni.
A harmonikus közép alkalmazása: átlagsebesség kiszámítása.
Valószínűségszámítás alkalmazása: biológiában öröklődésnél van szerepe, genetikai számításoknál. Minőség-ellenőrzés területén: termelésben megszabják, h mennyi lehet a hibás áru. +szerencsejátékosok használják a kaszinókban.

Kidolgozója: *Katanics Kinga* 12.D