1. **predavanje RVM – Kvantitativne metode**

Rok Strašek, Koper 7.5.2010

Statistične metode – z njimi proučujemo največ pojavov. Zato so najbolj splošne, generalne.

IZPIT
Sestavljen je iz dveh delov:

- pisni del (piše se na enkratnem srečanju. Deli se na dva dela)

- **v sklopu 1. dela se preverja računske spretnosti ali znanja. Ta je računski del. Rešuje se par nalog, ki bodo predelane na predavanjih bodisi v sklopu vaj.**
  - **4. naloge (1 iz Straškovega področja, 3 naloge Kodrič. 60% ocene (4x15%). Vsaka naloga 15 %).**
- **v sklopu 2 je teoretični del (20 vprašanj oz. trditev, izjav). Izmed teh bo nekaj odgovorov in zgolj eden od njih bo pravilen. (30 minut časa).**

- ustni del (služil bo bolj, da se potrdi oceno, ki je bila dosežena na pisnem izpitu. Pisni del je najprej potrebno opraviti s pozitivno oceno, potem še ustno Strašek, da preveri ali smo res sami pisali izpit. Vprašanja o čem smo se pogovarjali, nato se oblikuje ocena).

LITERATURA

- zapiski s predavanj

a) literatura (predlog Strašek)

- R.R. Pagano; **Understanding Statistics in the Behavioral Sciences**, 7-th edition, Thomon – Wadswoorth, Belmont 2004. (*knjiga kjer je vse na enem mestu; vse povedano in bistveno več*)
- L. Pfajfar, F. Arh; **Statistika I**, Ekonomska fakulteta, Lj 2003.
- F. Avsec, A. Cokan, I. Pucelj; **Kombinatorika, verjetnostni račun in statistika**, DZS, Lj 1986. (srednješolska knjiga, za gimnazije, ogromno primerov, primeri so tudi od tu vzeti, na začetku vsako poglavje strnjeno, povzeta vsebina, potem pa primeri)

b) literatura (predlog Kodrič)
- B. Košmelj; **Statistika 2 – 1.del**, Skripta, Ekonomska fakulteta, Ljubljana 2003.
- B. Košmelj; **Analiza odvisnosti za vzorčenja podatkov**, Ekonomska fakulteta Lj.

2. **Delitev statistike**
   a. **OPISNA STATISTIKA** (zbiranje, urejanje, razvrščanje, opisovanje, predstavljanje nekih podatkov)
      To je le del statistike. Če gremo proučevati nek vzorec govorimo o opisni statistiki. (Recimo imamo skupino študentov. Kaj bi bila opisna statistika ? Da bi se recimo prešteli, razvrstili po spolu, povprečna višina, povprečna teža, najmanjša teža, največja teža, itd.)

b. **ANALITIČNA STATISTIKA** (podatke uporabimo pri sklepanju nekih zakonitosti. Recimo vzamemo višino in težo prej omenjene skupine študentov. Iščemo zvezo med višino in težo prisotnih. Morda izpeljemo celo enačbo. V sklopu te statistike ločimo dve vrsti statističnega proučevanja:

   i. **UPORABNA STATISTIKA**
   Ukvarjali se bomo pretežno s tem delom statistike. Zadoščale nam bodo obče znane metode, ki so jih znanstveniki in strokovnjaki v preteklosti razvili. Ta orodja nam že služijo v obliki računalniških programov ali so zapisana v knjigah. Sprejeli jih bomo kot dejstva, da je to sprejemljivo orodje in da se s pomočjo tega da stvari izračunati.

   ii. **MATEMATIČNA STATISTIKA**
   S to vrsto se ukvarja skupina ljudi, strokovnjakov, ki te metode razvijajo, izpeljujejo, dokazujejo ali jih postavijo. To so matematični statistiki. To bo morda v bodoče najbolj cenjen poklic. Zaposljivost bo pa v zavarovalniških in bančnih sferah, itn.

3. Vzroki napak pri rabi statistike

1. Računske implikacije napak

Če nekaj preberemo v časopisu ali knjigi še ni rečeno da je to res. Lahko gre za tiskarsko napako. Do podatkov nasploh moramo biti kritični.

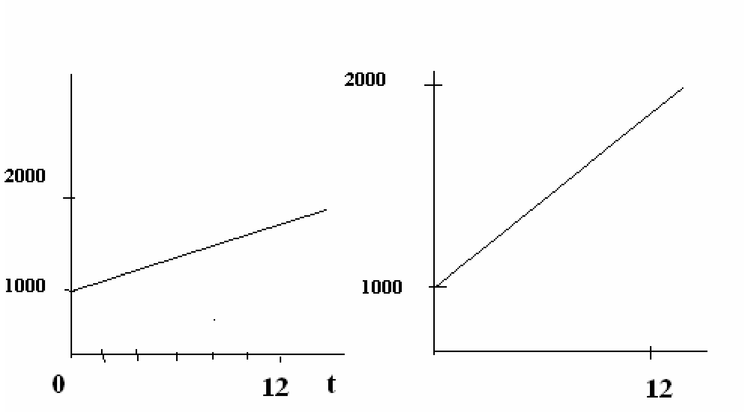
2. Napačni sklepi povezani z odstotki

Recimo pri razprodajah se to pogosto uporablja.

3. Lažna natančnost rezultatov

Povprečno število študentov, ki so obiskovali predavanja je bilo 7,39 študentov. Potrebno je zaokrožiti na smiselno vrednost. Velja načelo da naj bo takšna natančnost kot je ob zbiranju podatkov tudi pri predstavitvi rezultatov.

4. Kreativna raba grafov pri prikazovanju rezultatov



Predstavitev prodaje v lanskem letu. Na abscisi so meseci. Grafa predstavljata isto stvar, predstavljeno na dva različna načina. Drugi graf je bolj impresiven.

5. Prikazovanje istega podatka na različne načine

(Primer: V lanskem letu smo pridelali za 1000 EUR dobička. To je lahko isti podatek kot če bi rekel. V lanskem letu smo imeli za 1Mio EUR prihodkov. Tisti, ki ne loči dobička od prihodkov, bo mislil, da je podatek 1Mio EUR o prihodku kot zelo dober. Tisti ki pozna oba podatka pa ve, da je to precej slab dobiček, če gre na drugi strani za 999.000 EUR odhodkov).

6. ***Nepopolni podatki, manjkajoči odstotki o velikosti razredov v populaciji***

(Primer: 23% prometnih nesreč povzročijo vozniki, ki so mlajši od 20 let in samo 2% prometnih nesreč povzročijo starejši od 80 let. Na prvi pogled zgleda kot da so 80-letniki izvrstni vozniki. Kaj je tu manjkajoč podatek ? Število osemdesetletnikov, ki vozijo avto. Če jih je zelo malo je kumulativno gledano tistih 2% lahko zelo veliko. Pri anketah je ključni podatek pod grafikoni, tisto kar je najmanjše napisano. Gre za N, ki pomeni numerus, število ljudi, ki je bilo vključenih v anketo. Če je podatek 10 ipd. potem to ni nek relevanten podatek).

7. Nepravilno vzorčenje

Telefonske ankete. Vzorec naj bi najbolj odražal naravo celotne populacije.

8. Nerazumevanje med eksplicitno resnico, implicitno resnico in verjetno resnico.

Kaj je **EKSPLICITNA RESNICA** ? Nekaj kar je dejstvo, kar je res.

IMPLICITNA RESNICA je kar se lahko na nekaj sklepa. (Če ugasnemo luč bo zelo verjetno da ne bomo nič videli). Ni samo po sebi jasno je pa zelo verjetno. Recimo da nas je 11 prisotnih. V tem letu bomo imeli vsi rojstni dan. Teden ima sedem dni (pon. – ned.). Verjetna implicitna resnica je izjava, da imata vsaj dva prisotna študenta rojstni dan na isti dan v tednu. Če bi nas bilo samo sedem bi se lahko zgodilo, da ima vsak na svoj dan rojstni dan. In prejšnja izjava ne bi bila resnična. V trenutku pa ko jih je osem, je izjava gotovo resnična. Četudi jih bo sedem imelo rojstni dan na svoj dan bo pa osmi imel na tisti dan, ki je že zaseden. enajsti pa gotovo. Ko smo povedali izjavo stvar ni samo po sebi umevna, ko pa premislimo in napravimo sklep pa to sledi.

VERJETNA RESNICA

Vremenska napoved. Kakšno bo jutri vreme ?

9. Seštevanje verjetnosti kjer to ni možno

Verjetnost je neka količina s pomočjo katere se da celo računati. Izkaže se, da verjetnosti lahko odštevamo, seštevamo in celo množimo. Verjetnost nemogočega dogodka je 0 (0%). Verjetnost gotovega dogodka je pa 1 (100%).

10. Zamenjevanje korelacije in vzročnosti

Recimo da proučujemo kako je nek pojav odvisen od drugega. Nato se išče povezavo med zelo neverjetnimi stvarmi. (Primer: Proučujemo temperaturo v Mariboru in količino prodanega sladoleda v Kopru. Smiselno je proučevati povezavo med tistimi pojavi, ki so vzročno povezani.

4. Osnovni pojmi statističnega proučevanja

POPULACIJA

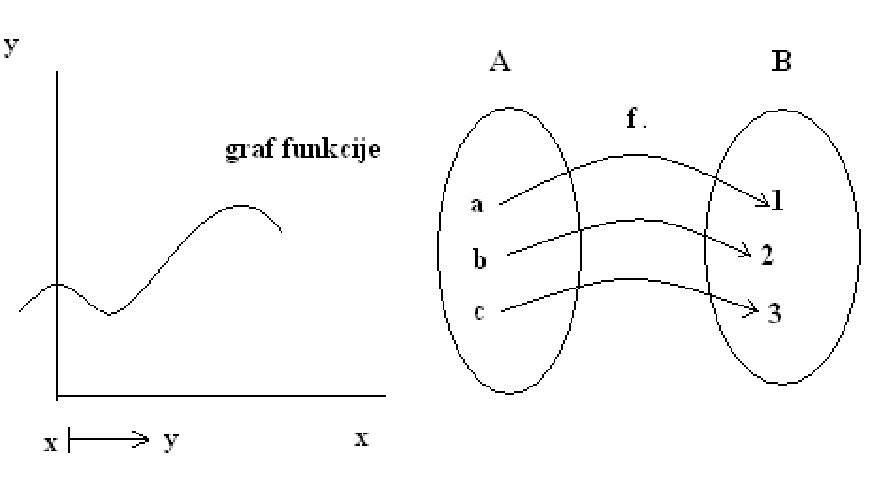
To je tista množica nekih enot na katerih nek statistični pojav proučujemo. Recimo vsi Slovenci smo populacija, prebivalci EU so populacija. Lahko so pa tudi manjše množice kot populacija. Vsi prisotni študentje, vsi Koprčani, itd. Na začetku je treba opredeliti populacijo. Vsaka podmnožica te populacije, neka skupina, nek del celotne populacije je **VZOREC**.

Vzorci so lahko primerni in neprimerni.

STATISTIČNA ENOTA **je najmanjši, nedeljiv del populacije ali vzorca**.

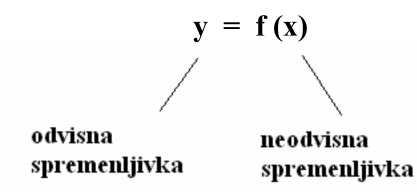
SLUČAJNA SPREMENLJIVKA

Znan nam je pojem spremenljivke.
Imamo neko množico A in vsakemu elementu iz te množice priredimo nek element množice B. Vsakega preslikamo. Temu rečemo da smo napravili funkcijo. Funkcija je nek predpis, preslikava iz množice A v množico B.



Funkcije so zelo različnih oblik.
Upodabljanje funkcije s pomočjo puščičnega diagrama. V primeru ko sta množica A in B z zelo velikim številom elementov, rečemo da elemente množice A upodobimo na x (abscisni) osi. Elemente B pa na y (ordinatni) osi.

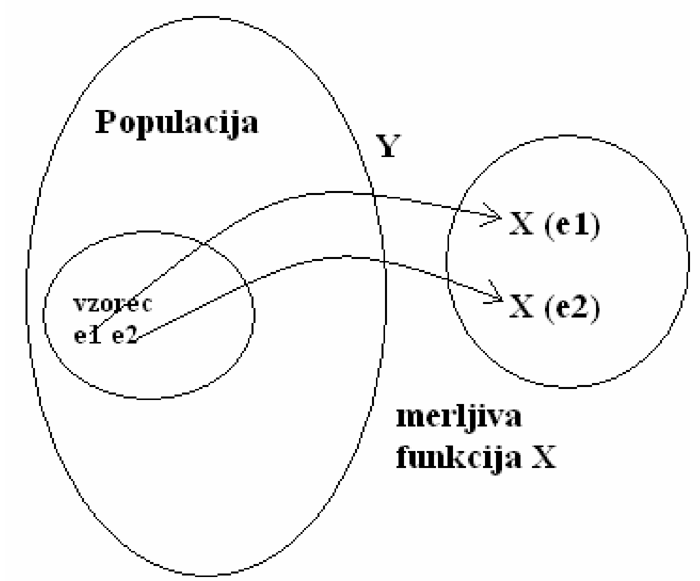
Funkcija podana z enačbo:



Slučajna spremenljivka je nekaj povsem drugega.
Kaj lahko naredimo na slučajni spremenljivki ?
Slučajna spremenljivka je nekaj drugega kot neodvisna spremenljivka. **To je nekaj kar se nanaša na statistiko**.

Slučajna spremenljivka se nanaša na statistično proučevanje nekega pojava oz. na populacijo oz. vzorec.
Vzamemo neko populacijo. To je neka množica elementov na katerih proučujemo neko lastnost. Če te lastnosti ne moremo proučevati na celotni populaciji, vzamemo vzorec. To je neka podmnožica populacije. **Najmanjši del vzorca je** STATISTIČNA ENOTA.
Proučujemo neko lastnost. Kaj bi lahko bila lastnost populacije? Populacija bi lahko bili avtomobili in proučujemo barvo avtomobilov. Rečemo: kakšno barvo ima prvi, kakšno drugi, tretji, itd. Vsaki statistični enoti bi priredili neko merljivo funkcijo. Merljivo funkcijo označimo recimo z X. Vsaki **statistični enoti e1, e2 priredimo vrednost X(e1), X(e2)**, itn.

(glej sliko spodaj). Vrednosti merljive funkcije na različnih statističnih enotah so lahko enake. Lahko imamo dva različna avta, recimo Opla in Audija z isto barvo. Imamo dve različni statistični enoti in izmerljiva funkcija barve preslika v isti rezultat. Podatek ima isto mero.



Najmanjši del vzorca je statistična enota (e).

Tako kot če bi merili težo bi se lahko zgodilo, da bi dve različni statistični enoti imeli isto težo. Na posameznem vzorcu ali populaciji vpeljemo različne merljive funkcije. Enkrat je to lahko višina, teža, inteligenčni kvocient ali pri avtih prostornina motorja, barva, teža, itn.

Merljivi funkciji pri statističnem proučevanju, v kolikor imamo opravka samo z eno, niti ne dodajamo posebnega imena. **Če imamo merljivih funkcij, ki jih proučujemo več, potem rečemo, da imamo opravka s spremenljivkami. Če izberem poljubno izmed teh spremenljivk potem rečemo, da smo imeli opravka z NAKLJUČNO SPREMENLJIVKO.**

Dogovorjeno pravilo je, da v sklopu statističnega proučevanja, **MERLJIVE FUNKCIJE** poimenujemo oz. smatramo kot **NAKLJUČNE SPREMENLJIVKE**.

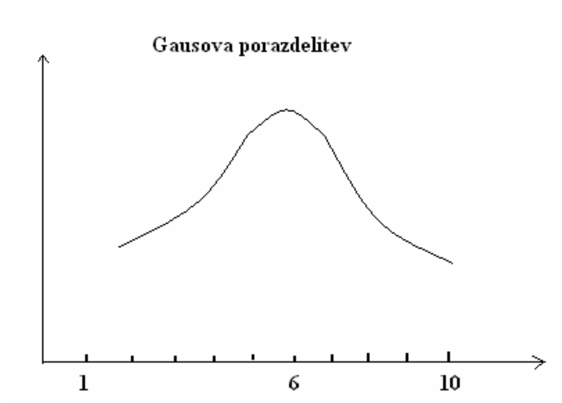
Rečemo: Višina je naključna spremenljivka, teža naključna spremenljivka, inteligenčni kvocient, naključna spremenljivka. Če bi nas kdo vprašal koliko smo visoki, bi recimo rekli 179. In rečemo naključna spremenljivka višine nam je priredila vrednost 179. Tonetu je naključna spremenljivka teže priredila vrednost 88kg. Marku je naključna spremenljivka inteligenčnega kvocienta priredila vrednost 122. **Naključna spremenljivka je v bistvu neke vrste lastnost ali funkcija**.

V sklopu statističnega proučevanja proučujemo lahko različne stvari.

A) proučevanje statističnih spremenljivk oz. proučevanje njihovih lastnosti. Ukvarjali bi se zgolj z **eno merljivo funkcijo oz. slučajno spremenljivko**. Kaj bi lahko z njo proučevali ? Izražanje statističnih spremenljivk.

Kako se izraža spremenljivka inteligenčnega kvocienta ? V obliki števila (98, 100, 120, 110).
Če pa rečemo, da na naš vzorec vpeljemo slučajno spremenljivko barve las. Kaj so potem vrednosti slučajne spremenljivke ? (siva, modra, rdeča, blond, itn.). **Spremenljivke so lahko**

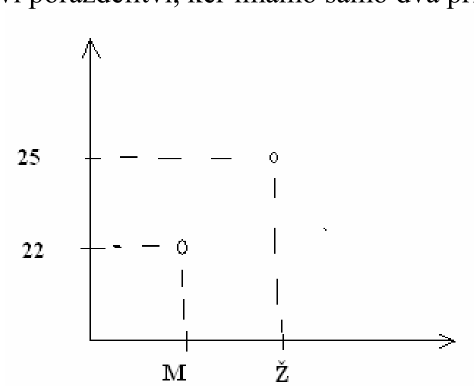
opisne ali numerične. Ene so primerjalne. Pogovarjamo se recimo o tem kdo je starejši. Tega ne gledamo v smislu števila let ampak rečemo ta je mlajši od tega, itd. (učbenik Statistika I).



Recimo da gledamo ocenjevanje od 1 do 10. Največ študentov je imelo oceno 6 nato je proti 10 zadeva nekoliko padla. Podobno se je zgodilo na drugo stran (5,6). Ocene so bile porazdeljene z Gaussovo porazdelitvijo. Ocene so v glavnem Gaussovo porazdeljene v skladu z Gaussovo porazdelitvijo.

Za spol ne pričakujemo da gre za Gaussovo porazdelitev, ker imamo samo dve vrednosti. **Gaussova porazdelitev je teoretično tista, ki ima obliko zvonca**.

Če pa gledamo spol dobimo, za moške številčno vrednost in za ženske neko vrednost. Ne moremo govoriti o Gaussovi porazdelitvi, ker imamo samo dva primera.



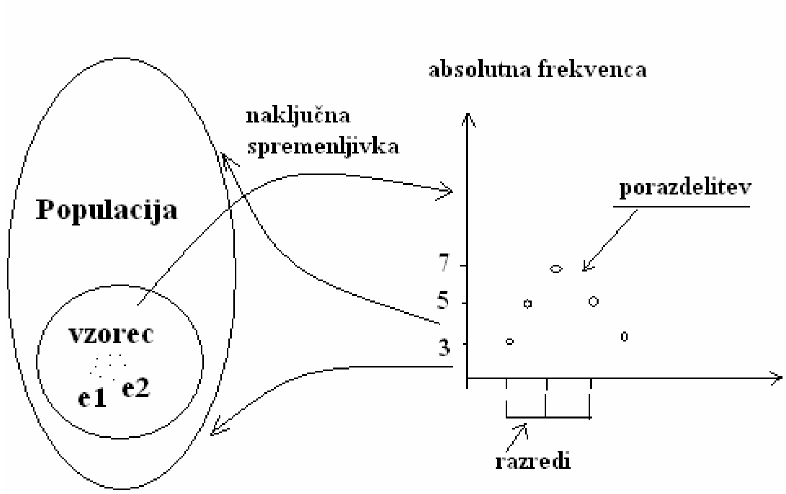
Spol: M – moški, Ž – ženski

B) proučevanje zvez med statističnimi spremenljivkami. Opravka imamo z dvema spremenljivkama. Recimo spremenljivki x in y. In bi proučevali povezavo med dvema naključnima spremenljivkama na isti populaciji. Proučevali bi recimo zvezo med slučajno spremenljivko teža in inteligenčnim kvocientom. Ali so njune vrednosti med seboj povezane? Ali recimo lahko sklepamo, da ima tisti, ki je bolj debel tudi višji inteligenčni kvocient? Lahko se zgodi da ni nobene povezave. Lahko pa ugotovimo, da smo tako specifična populacija, da bi se ta povezava našla. Pravimo da je **povezava pozitivna**.

Če bi hoteli proučevati zelo veliko populacijo na kateri ne moremo sklepati oz. ne moremo narediti analize na celotni populaciji. Zato izberemo nek vzorec. Ga proučimo. Ugotovimo značilnosti na tem vzorcu in na osnovi funkcije oz. lastnosti vzorca, poskušamo in rečemo:

Ker je vzorec reprezentativen lahko posplošimo in rečemo. To kar smo ugotovili na vzorcu bomo posplošili na celotno populacijo.

Če smo na osnovi dvajsetih ljudi v vzorcu dobili tako formulo in lahko posplošimo to formulo na celotno populacijo potem bi lahko rekli: Ti si eden izmed te populacije. Ker je vzorec reprezentativen povej koliko si težak in povedal ti bom koliko imaš inteligenčni kvocient.
TO JE BISTVO STATISTIČNEGA PROUČEVANJA.



Imamo populacijo. Na populaciji želimo proučevati neko lastnost oz. proučujemo neko NAKLJUČNO SPREMENLJIVKO. Le redko se zgodi, da lahko izvedemo raziskavo na celotni populaciji. Zato vzamemo vzorec. To isto **naključno spremenljivko ne analiziramo na populaciji ampak na vzorcu**. In na osnovi naključne spremenljivke in proučevanja poizkusimo temu vzorcu oz. tej naključni spremenljivki prirediti t.i. porazdelitev. Kaj ta porazdelitev pomeni ? To je ključno vprašanje, ki ga poskušamo razvozlati.

Če na osnovi vzorca za naključno spremenljivko ugotovimo kakšno ima porazdelitev, potem v skladu s to porazdelitvijo zelo enostavno sklepamo nazaj na celotno populacijo.

Konkretni primer za porazdelitev:
Imamo vzorec. Vzorec ima vedno končno število statističnih enot. Končno pomeni, da jih lahko preštejemo. To je tudi bistvo vzorca. Da iz velikega števila reduciramo na manj. **Potem vzorcu preko naključne spremenljivke priredimo vrednosti te naključne spremenljivke**.

Recimo da gre za inteligenčni kvocient. Vrednosti te naključne spremenljivke, za vsako statistično enoto, nanesem, na x os. Lahko se zgodi, da bi vsaka statistična enota imela različno vrednost naključne spremenljivke. Je to možno ?
Če merimo recimo višino na mm natančno. Tudi če bi imeli 1000 različnih osebkov bi se zelo verjetno zgodilo, da bi skoraj vsak imel različno višino. Če bi merili višino tako natančno bi vsaka vrednost naključne spremenljivke imela svojo vrednost. In koliko bi bila vrednost na y osi ? Ena (1). Bi bila ta informacija dobra? Ta informacija bi bila pravzaprav zelo dobra, ker bi res za vsakega posameznika vedeli natančno koliko je njegova višina. Če bi hoteli sklepati bi vzeli vzorec 1000 Slovencev in ugotovili, da ima med temi 1000 ljudmi vsak svojo višino. Iz tega bi lahko napravili samo relevanten sklep, da bi na vseh 2Mio prebivalcih dobili 2Mio različnih višin. Ta sklep ne bi bil dober.

Zato napravimo tako, da v bistvu vrednosti naključne spremenljivke smiselno razdelimo oz. jih združimo v neke razrede.
Vzamemo razred od prve do druge višine, nato razred od druge do tretje, itn. Potem bi posameznemu razredu prešteli koliko osebkov je prišlo, z vrednostjo naključne spremenljivke, v ta razred. Za prvi razred dobimo neko novo točko, prav tako za ostale razrede. Ali na ta način pridobimo ali zgubimo na natančnosti podatkov? Zgubimo. Zakaj? Do sedaj nam bi ta podatek povedal (3, 5, 7, 11), da imajo višino v razredu od 150 do 165 trije osebki. Prej, če bi za vsakega posebej izmerili bi vedeli za vsakega natančno koliko ima. Tu pa ne vemo. Vemo samo da so od 150 do 165 trije. S tem izgubimo sled. Ne vemo več kdo je imel koliko. Lahko so vsi trije imeli 164, ali pa 151. Je pa ta stvar bolj uporabna, z vidika posplošitve iz vzorca na celotno populacijo.
Kolikšno je optimalno število razredov ? Prvi kriterij je po občutku. Velja nepisani kriterij, da če imamo sorazmerno veliko podatkov, napravimo nekje med 10 in 20 razredov.
V splošnem obstajajo različne formule, ki računajo te razrede.

Obstaja področje matematike, ki bi lahko, glede na te podatke in te pike, oblikovali neko funkcijo, neko neprekinjeno, približno črto tako, da bi najbolje povezovala. Ali so podatki 3, 5, 7 11 primerne količine? S pomočjo teh podatkov ne moremo sklepati na celotno populacijo. Če te vrednosti preračunamo na odstotke. Pred tem nam te vrednosti predstavljajo absolutne frekvence, lahko pa jih preračunamo na relativne. Relativna pomeni kakšen delež oz. odstotek to število predstavlja v celotnem številu osebkov. Ali bi se graf kaj spremenil, če bi šli iz absolutnih frekvenc na relativne frekvence? Če bi na ordinati imeli 100 enot potem bi to predstavljalo 3/100. Če stvar preračunamo v relativne frekvence potem smo že bližje. Ta vrednost nam potem predstavlja koliko odstotkov vzorca pade v posamezen razred. Ko imamo relativne frekvence je samo stvar, da ugotovimo, koliko je ta vzorec reprezentativen. In če je reprezentativen, lahko relativne frekvence oz. odstotke zelo elegantno posplošimo na populacijo. Rečemo recimo: 15% ljudi iz vzorca, ki je reprezentativen, je v razredu od 175 do 180. Ker je vzorec reprezentativen lahko sklepamo, da je teh 15% ljudi celotne populacije v tem razredu.
Temu grafu, obliki teh točk kako so razvrščene, pravimo **PORAZDELITEV**.

Statistiki naredijo celo korak naprej. Rekli smo, da smo naredili neko napako, s tem ko smo podatke iz vzorca zmetali skupaj v razrede. Potem se je v neki toči raziskovanja pojavila ideja ali bi lahko s katero od statističnih metod oz. analitičnih metod, morda vendarle, s tem ko smo naredili združevanje razredov, nazaj preoblikovali in zagotovili večjo natančnost. Dejansko se to ne da narediti. Se pa da s posebnimi metodami aproksimativno poiskati neko funkcijo, ki se bo tem podatkom optimalno prilegala. (metoda najmanjših kvadratov, najmanjših odklonov, itn.)

Samo statistično proučevanje poteka ravno na ta način. Iz populacije vzamemo vzorec. Iz vzorca proučujemo naključno spremenljivko in napravimo frekvenčno porazdelitev najprej absolutnih frekvenc potem napravimo relativne frekvence in s pomočjo katere od aproksimativnih metod napravimo iz diskretne zvezno funkcijo. Ko imamo zvezno lahko sklepamo iz vzorca na celotno populacijo.

Ključni problem je to, da vzorcu priredimo ustrezno razvrstitev relativnih frekvenc. Kaj pa je porazdelitev ?
Relativna frekvenca v matematiki ni ravno najboljši primer. O relativnih frekvencah govorimo ko imamo opravka s pikami, s končnim številom točk.

Če bi nadalje iz relativnih frekvenc prišli na zvezno, neprekinjeno funkcijo, potem **ne govorimo več o relativnih frekvencah ampak jim pravimo verjetnosti.**

PORAZDELITEV JE V BISTVU FUNKCIJA OZ. PREDPIS, KI VSAKI VREDNOSTI NAKLJUČNE SPREMENLJIVKE PRIREDI NJENO VERJETNOST.

POJEM VERJETNOSTI

V statističnem proučevanju je verjetnost zelo pomemben pojem.
V kakšni zvezi sta statistika in verjetnost?

statistika ←———→ verjetnost

Izkaže se, da dobimo neke verodostojne podatke in najbolj kredibilne sklepe, pri proučevanju družboslovnih pojavov, če sočasno vključimo tako statistiko kot verjetnost.

Najlepše ločitev med statistiko in verjetnostjo napravimo na naslednji način:
O statistiki govorimo, ko želimo povedati o nekem pojavu, kaj se je z njim dogajalo potem, ko smo izvedli neko meritev, raziskavo, ko smo dobili podatke o populaciji in vzorcu in nato poskušamo o populaciji oz. vzorcu in pojavu, ki smo ga proučevali spregovoriti na osnovi podatkov, ki smo jih dobili.

Imamo populacijo, vzorec, izberemo neko naključno spremenljivko, s katero od metod to izmerimo, proučujemo, izračunamo najpomembnejše parametre in na osnovi teh parametrov dobimo neke sklepe, lastnosti o tem pojavu.
Po drugi strani imamo pa verjetnost.
V sklopu tega v naših mislih napravimo model in potem na osnovi tega modela skušamo sklepati kaj se bo oz. bi zgodilo, če bi nek preizkus oz. pojav izvedli.

<u>Konkretni primer</u>
Imamo kovanec in ga vržemo. S kakšno verjetnostjo bo padla cifra? 50%. Kam uvrstimo ta sklep? Ustvarili smo nek miselni model. (kovanec, dve strani, vržemo, pade na eno ali na drugo stran, dva možna izida, eden se bo zgodil, ½ , 0,5, 50%, miselni proces).

Kako bi prišli do te verjetnosti, če bi se je lotili s pomočjo statistike?

Vzeli bi kovanec in rekli, da ga vržemo 100 krat pa bomo šteli kolikokrat je padel na cifro kolikokrat na grb. Če bi 100 krat vrgel in bi padlo 44 krat bi rekli, verjetnost je 44%, da pri metu pade cifra. Pri tem imamo statistično proučevanje, kjer na osnovi podatkov, meritev, sklepamo o pojavu, na drugi strani pa na osnovi nekega modela.

Koliko je verjetnost da pade pri metu igralne kocke šestica?
1/6. Sodi kam ? Izračunamo s pomočjo nekega modela**. Število ugodnih s številom možnih.**
Ugoden je ena šestica, vseh možnih je šest. Model, 1/6.

Po drugi strani bi pa lahko vprašali. Ali je ta kocka poštena ? Če tega ne vemo je treba kocko metati. Poskus poizkušamo velikokrat ponoviti. Vržemo 100 krat, 1000 krat prešteli in zapisali bi kolikokrat je padla šestica in bi prišli do sklepa. **Ta sklep sodi v kategorijo statistike**.

Kaj je idealna kombinacija? Ko pogoje povežemo. To počne predvsem ekonomija. Kaj napravi? Spremljamo nek pojav, inflacijo, na osnovi minulih podatkov. Podatke spremljamo več mesecev, več let, spremljamo gibanje, dinamiko in ugotovimo kako se inflacija giblje. Pridemo do sedanjosti. Podatkov za prihodnost nimamo. **Skušamo napraviti sklep za naprej na osnovi zbranih podatkov, na osnovi modela, teorije verjetnosti. Govorimo o regresiji.** Napravimo neko premico ali funkcijo, ki se optimalno prilega podatkom in jo potegnemo v prihodnost. Nato rečemo, da se bo zelo verjetno tako gibalo.

Statistična definicija verjetnosti dogodka je neko število pri katerem se stabilizira relativna frekvenca dogodka A v velikem številu ponovitev tega poskusa.

$$\mathbf{P(A) \approx \frac{k(A)}{n}}$$

Ta definicija ima neko slabost. Katero? Slaba je ta definicija v matematičnem smislu zato ker piše v velikem številu ponovitev. To ni dovolj natančno. Kaj je to veliko število ponovitev? Je to 50000, 100000, 200000? Tega ne vemo. Zato definicija ni najbolj natančna.

Klasična definicija verjetnosti

$$\mathbf{P(A)\ \frac{\text{število ugodnih izidov za A}}{\text{število vseh izidov}}}$$

Kakšna je verjetnost, da pri metu **dveh kock** pade **vsota cifer 7** ?

a) Koliko je vseh možnih izidov ? Koliko je pravzaprav možnih vseh vsot, če vržemo dve kocki ?
   Najnižja vsota je **2 (11)**, sledijo **3, 4, 5, 6, 7, 8, 9, 10, 11, 12**, to je **11** različnih vsot.
   Enkrat je pa vsota 7. S tega vidika je verjetnost **1/11**.
   Izhodišče je bilo, da smo kot možne izide vzeli vse mogoče vsote. Lahko bi vzeli tudi

b) 11 12 13 14 15 **16**
   22 23 24 **25** 26 (21 je isto kot 12, če kock ne ločimo)
   33 **34** 35 36
   44 45 46
   55 56
   66

   Koliko je možnih? **21**
   Koliko je ugodnih? **3**
   Verjetnost **3/21** kar se krajša s 3 na **1/7**.

Katera verjetnost je sedaj prava? Za isti pojav imamo dva različna izida. Kaj pa če sta ti dve kocki različnih barv? Potem je 12 nekaj drugega kot 21. Potem je imamo 36 možnih, ugodnih je pa (16, 25, 34, 43, 52, 61), torej 6. Verjetnost je 6/36=1/6. Sedaj imamo že tri različne izide. Kateri je pravi?
Tudi ta definicija ni dovolj natančna. Dopušča za isto vprašanje različne odgovore. Ni enolično določeno. Kdaj bi bila ta definicija dobra?
Če bi imeli vsi izidi enake možnosti.

Izid je pravzaprav vsota pik. V primeru b se 7 in 6 lahko zgodita na tri načine. 5 se pa lahko zgodi samo na dva načina. Primer b ne bi bil primeren za to obravnavo, ker ni enakih možnosti (6, 7 in 5 nimajo enake možnosti).

"Poštena kocka" bi morala biti iz neke homogene snovi.

Matematiki so za primere, kjer ni enakih možnosti, vpeljali različne definicije verjetnosti.
Omenili bomo **statistično, klasično in aksiomatično definicijo**. Niso pa edine.
Aksiomatična je zelo kompleksna, je pa značilna in uporabna za vse pojave. Tudi take kjer imamo opravka z izidi, ki imajo vsi enako možnost. V sklopu te definicije se vpelje aritmetiko na ta način, da so osnovni objekti te teorije dogodki, vsakemu dogodku pa potem pripišemo neko verjetnost oz. neko število med 0 in 1. Nato pa po nekih pravilih tudi računamo.
Teorija verjetnosti hodi z roko v roki s teorijo o množicah. Računske operacije pri množicah (unija, presek, razlika, kartezični produkt, itn.)
Med dva dogodka vpeljemo računsko operacijo, ki je primerljiva z računsko operacijo preseka. Pri preseku rečemo, da je presek dveh množic množica v kateri so elementi, ki so hkrati v množici A in hkrati v množici B. Lahko rečemo, da je dogodek C produkt dogodkov A in B, če se poleg dogodka A hkrati zgodi tudi B.

Kako se lahko število vseh izidov računa? Dostikrat moramo v sklopu verjetnosti število vseh izidov in število ugodnih izidov izračunati. Imamo različne primere pojavov s pomočjo katerih lahko izide računamo. To računamo s pomočjo sklopa teorije, ki ji rečemo **elementarna verjetnost** (zelo dobro obrazloženo v knjigi **Kombinatorika, verjetnostni račun in statistika**).

5. PERMUTACIJE

O njih govorimo ko nas zanima število različnih razvrstitev **n** elementov na **n** mest.

$$n! = n * (n-1) * (n-2) \dots (3*2*1)$$

imamo množico A, ki ima tri števila A {1, 2, 3}

Koliko različnih trimestnih števil lahko iz tega tvorimo? **3! = 3*2*1 = 6**
Katera pa so to: 123, 132, 231, 213, 321, 312
To so permutacije. Bistveno pri permutacijah je **n elementov na n mest**.

Vprašanje je lahko. Imamo pet različnih črk. Koliko različnih petmestnih besed lahko tvorimo? 5!=5*4*3*2*1=120

6. PERMUTACIJE S PONAVLJANJEM

Zanima nas število različnih razvrstitev **n elementov na n mest pri čemer se elementi lahko ponovijo**.
Če imamo množico A in je nekaj elementov med seboj enakih

A = {1, 1, 2}

Če bi imeli razvrstitev teh treh elementov na tri mesta bi imeli naslednje situacije:
112, 121, 211 pri permutacijah s ponavljanjem enk med sabo ne ločimo. 11 je isto kot 11, če zamenjamo enke. Teh števil med sabo ne ločimo.

$$\frac{n!}{n1!\ n2!\ \ldots nk!}$$

Pomnožimo oz. delimo s fakulteteto tistih, ki so med sabo enaki.
Za gornjo množico imamo:
Števila so tri (3) → 3!, deljeno z
enki sta dve(2) → 2!
Dvojk je pa ena (1) → 1!

Imamo: $$\frac{3!}{2!*1!} = \frac{3*2*1}{(2*1)*(1)} = 3$$

7. VARIACIJE

Zanima nas število razvrstitev **n elementov na k mest (n ≥k),** pri čemer je vrstni red pomemben.

$$V_k^n = \frac{n!}{(n-k)!}$$

A = {1,2,3,4} Tvorimo dvomestna števila. Koliko jih je.

$$V_2^4 = \frac{4!}{2!} = \frac{4*3*2*1}{2*1} = 12$$

12	21	31	41
13	23	32	42
14	24	34	43

8. VARIACIJE S PONAVLJANJEM (dovoljeno ponavljanje)

$$^{p}V_k^n = n^k$$

$$4^2 = 16$$

Kaj še manjka ? če dovolimo ponavljanje so to 11, 22, 33, 44 kar znese 12+4=16

9. KOMBINACIJE

Zanimive so zaradi igre na srečo, ki jo imenujemo loto.
n elementov na k mest, pri čemer je vrstni red nepomemben.
Pri lotu je 12 isto kot 21.

$$C^{n}_{k} = \binom{n}{k} = \frac{n!}{(n-k)!\ \ k!}$$

Binomski simbol (isto kot variacije
samo da so ti, ker jih je manj, zmanjšani za k!)

A = {1, 2, 3, 4} imamo dvomestni loto

$$C^{4}_{2} = \binom{4}{2} = \frac{4!}{2!\ \ 2!} = \frac{4 * 3 * 2 * 1}{2 * 1 * 2 * 1} = 6$$

Kombinacije teh štirih števil množice so : **12, 13, 14, 23, 24, 34**

10. KOMBINACIJE (s ponavljanjem)
Dovolimo da se številke ponovijo.

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{(n-1)!\ \ k!}$$

Slovenski loto ima 39 številk, 7 se jih izžreba.

$$^{p}C^{n}_{k} = C^{39}_{7} = \frac{39!}{32!\ 7!} =$$

$$\frac{39*38*37*\cancel{36}*\cancel{35}*34*33*\cancel{32}*\cancel{31}}{\cancel{32}*\cancel{31}*\cancel{30}*\ldots\ \cancel{7}*\cancel{6}*\cancel{5}*4*\cancel{3}*\cancel{2}*\cancel{1}} =$$

$$\frac{39*38*37*..34*33}{4} = 15380937$$

Verjetnost zadetka : **1 / 15380937** = **0,000000065**

Če stane ena kombinacija 20 centov. Da bi vplačali vse kombinacije potrebujemo 3.076.187,4 EUR.