STATISTIKA IN ANALIZA PODATKOV

**STATISTIKA** je veda, ki proučuje množične pojave in se ukvarja z zbiranjem, predstavitvijo, analizo in interpretacijo podatkov.

**ENOTA** je posamezni proučevani element (redni študent na Univerzi v Lj v študijskem letu 1994/95)

**POPULACIJA** je množica vseh proučevanih elementov; pomembna je časovna in prostorska opredelitev populacije (vsi redni študentje na Univerzi v Lj v študijskem letu 1994/95)

**VZOREC** je podmnožica populacije, na osnovi katere ponavadi sklepamo o lastnostih cele populacije (slučajni vzorec 300 rednih študentov na Univerzi v Lj v študijskem letu 1994/95)

**SPREMENLJIVKA** je lastnost enot; označujemo jih z X, Y (spol, uspeh iz matematike v zadnjem letniku srednje šole, izobrazba matere in višina mesečnih dohodkov staršev študenta)

**Vrste spremenljivk:**

* **glede na tip izražanja vrednosti: -** *opisne, atributivne*; vrednosti lahko opišemo z

besedami (poklic, uspeh)

 **-** *številske, numerične*; vrednosti lahko izrazimo s števili (starost)

* **glede na tip merjenja: -** *nominalne*; vrednosti lahko le razlikujemo med seboj;

 dve vrednosti sta enaki ali različni (spol)

 - *ordinalne*; vrednosti lahko uredimo od najmanjše do

 največje (uspeh)

 - *intervalne*; lahko primerjamo razlike med vrednostimi

 dvojic enot (temperatura)

 - *razmernostne*; lahko primerjamo razmerja med

 vrednostim dvojic enot (starost)

**Statistična analiza je:**

* *opisna statistika*; statistična analiza zbranih podatkov brez teženj, da bi iz teh podatkov posploševali čez njihov obseg
* *inferenčna statistika*; statistično sklepanje iz vzorca (dela populacije) na populacijo: - ocenjevanje značilnosti populacije

 - preverjanje domnev

* *univariantna*; analiza ene spremenljivke
* *bivariantna*; analiza dveh spremenljivk
* *multivariantna*; analiza več spremenljivk

**Koraki statistične analize:**

1. **Določitev vsebine in namena** statističnega proučevanja; opredelitev predmeta opazovanja (enote in populacije) in vsebine opazovanja (spremenljivk)
2. **Statistično opazovanje**; vrste opazovanj: - opazovanje cele populacije (popisi, tekoče registracije)

 - opazovanje vzorca (ankete)

1. **Osnovna obdelava**: - urejanje

 - razvrščanje podatkov

 - izračun osnovnih statističnih karakteristik

1. **Analitična obdelava**

**Frekvenčna porazdelitev spremenljivke** je tabela, ki jo določajo vrednosti ali skupine vrednosti in njihove frekvence.

**Razredi** so skupine vrednosti številskih spremenljivk.

xi,max = zgornja meja i-tega razreda

xi,min = spodnja meja i-tega razreda

**Širina** i-tega razreda je

 di = xi,max – xi,min

**Sredina** i-tega razreda je

**Kumulativa (Fi)** je frekvenca do spodnje meje določenega razreda.

**fi%** je relativna frekvenca = strukturni odstotki v i-tem razredu

**Grafično prikazovanje frekvenčnih porazdelitev:**

* *histogram*; drug poleg drugega rišemo stolpce pravokotnika, katerih višina je sorazmerna frekvenci v razredu. Širina pravokotnikov je enaka, ker so razredi enako široki.
* *poligon*; v koordinatnem sistemu zaznamujemo točke (xi, fi), kjer je xi sredina i-tega razreda in fi njegova frekvenca. K tem točkam dodamo še točki (x0, 0) in (xk+1,0), če je v frekvenčni porazdelitvi k razredov. Točke zvežemo z daljicami.
* *Ogiva*; grafična predstavitev kumulativne frekvenčne porazdelitve s polgonom, kjer v koordinatnem sistem nanašamo točke (xi,min,Fi).

**Oblika frekvenčnih porazdelitev:**

* *normalna*, ki je unimodalna (ima en vrh), simetrična in zvonaste oblike
* *bimodalna*, če ima dva vrha
* *polimodalna* z več vrhovi
* *asimetrična v levo*, če se rep vlečena levo
* *asimetrična v desno*, če se rep vleče v desno
* bolj *koničasta* in *sploščena* od normalne porazdelitve
* *J* in *U* oblike

Za grafično predstavitev s krogi je potrebno izračunati še **stopinje** fi0 :



**Ranžirna vrsta:** enote z ustreznimi vrednostnimi spremenljivke uredimo od tiste z najmanjšo vrednostjo do tiste z največjo vrednostjo.

**Rang R**: vsaki enoti v ranžirni vrsti priredimo zaporedno mesto.

**Kvantilni rang P** pove, na katerem delu celotnega ranžirnega razmika leži določena enota oziroma koliki del celotnega razmika ima manjše vrednosti od dane vrednosti.



**Kvantil** je vrednost spremenljivke, ki pripada določenemu kvantilnemu rangu.

Običajni kvantili:

 - **mediana**  Me(P=0.5)

* **kvartili** Q1(P=0.25), Q2(P=0.50), Q3(P=0.75)
* **decili** D1(P=0.1), D2(P=0.2),… D9(P=0.9)
* **centili**  C1(P=0.01), C2(P=0.02),… C99(P=0.99)

**Linearna interpolacija:**



**Srednje vrednosti:**

1. **Mediana** (Me) je tista vrednost spremenljivke, od katere je ravno toliko manjših vrednosti od nje, kolikor jih je večjih od nje. Zato je mediana vrednost spremenljivke, ki pripada kvantilnemu rangu 0.5.
* Če je liho število enot N=2m+1, je mediana (m+1) vrednost v ranžirni vrstici.
* Če pa je sodo število enot N=2m, je mediana

Iz frekvenčne porazdelitve pa izračunamo mediano tako, da izračunamo vrednost spremenljivke, ki pripada kvantilnemu rangu P=0.5. Ranžirno vrsto s pripadajočimi rangi predstavljajo spodnje meje razredov in ustrezne kumulative.

1. **Modus** (M0) je vrednost spremenljivke, ki se v populaciji najpogosteje pojavlja. Modus lahko razumemo kot vrednost spremenljivke, okoli katere se vrednosti najbolj gostijo.
2. **Aritmetična sredina** ali povprečje je vsota vseh vrednosti deljena s številom enot v populaciji.
3. **Geometrijska sredina** je enaka N-temu korenu iz produkta N vrednosti številske spremenljivke (pogoj xi>0).
4. **Harmonična sredina** je enaka recipročni vrednosti aritmetične sredine.

**MERE RAZPRŠENOSTI:**

1. **Variacijski razmik:**  R = xmax - xmin
2. **Kvartilni odklon:**
3. ** Povrečni absolutni odklon:**

**** Negrupirani podatki: Grupirani podatki:

****

1. **Varianca:**

****

1. **Standardni odklon**

**Sheppardov popravek:**



→ *Relativne mere razpršenosti* so absolutne mere deljene z ustrezno srednjo vrednostjo:

* relativni variacijski razmik:
* relativni kavrtilni odklon:
* relativni povprečni absolutni odklon:
* relativni standardni odklon – koeficient variacije:



→ *Normalna porazdelitev*: Denimo, da se spremenljivka X porazdeljuje normalno. Zanjo je izračunana aritmetična sredina μ in standardni odklon σ. Tedaj velja, da v razmiku:

♦[μ-σ; μ+σ] leži 68.3% enot

♦[μ-2σ; μ+2σ] leži 95.4% enot

♦[μ-3σ; μ+3σ] leži 99.7% enot

→ *Meri asimetrije*:

→ *Mere sploščenosti*:

→ *Meri asimetrije in sploščenosti s centralnimi momenti:*

L-ti centralni moment je:



♦Koeficienti asimetrije:

♦Koeficienti sploščenosti:



Permutacija je vsaka preureditev n elementov.

Variacija reda r iz n elementov je, če iz množice n elementov vzamemo r elementov in jih na nek način razporedimo.

Osnovni izrek kombinatorike: Imejmo izbor, sestavljen iz k delnih izborov. Prvič izbiramo med n1 možnostmi, drugič med n2 možnostmi,… in k-tič med nk možnostmi. Pri tako sestavljenem izboru je vseh možnosti:

 n=n1\*n2\*…\*nk

Število variacij, permutacij in kombinacij:

1. Število variacij reda r iz n elementov s ponavljanjem:
2. Število variacij reda r iz n elementov brez ponavljanja:
3. Število premutacij:
4. Število kombinacij:



# VERJETNOSTNI RAČUN

→ Poskus je realizacija neke množice skupaj nastopajočih dejstev.

→ Dogodek je pojav, ki v množico skupaj nastopajočih dejstev ne spada in se lahko v posameznem poskusu zgodi ali pa ne.

 **Dogodek je lahko**:

* gotov dogodek G; ob vsaki ponovitvi poskusa se zgodi
* nemogoč dogodek N; nikoli se ne zgodi
* slučajen dogodek; včasih se zgodi, včasih ne

**Računanje z dogodki:**

1. Dogodek A je **način** dogodka B (A⊂B), če se vsakič, ko se zgodi dogodek A, zagotovo zgodi tudi dogodek B.
2. Če je dogodek A način dogodka B in sočasno dogodek B način dogodka A, sta dogodka **enaka**: A⊂B∧B⊂A⇔A=B
3. **Vsota dogodkov** A in B (A∪B) je, če se zgodi vsaj eden od dogodkov A in B. Velja: A∪B=B∪A; A∪N=A;

A∪G=G, A∪A=A

1. **Produkt dogodkov** A in B (A∩B) se imenuje dogodek, če se zgodita A in B hkrati.

Velja: A∩B=B∩A; A∩N=N;

 A∩G=A; A∩A=A

1. Dogodku A **nasproten dogodek**‾A imenujemo negacija dogodka.

Velja: A∩‾A=N; A∪‾A=G;

 ‾N=G; ‾A=A

1. Dogodka A in B sta **nezdružljiva**, če se ne moreta zgoditi hkrati, njun produkt je torej nemogoč dogodek, A∩B=N

Velja: A∩‾A=N∧A∪‾A=G

1. Če lahko dogodek A izrazimo kot vsoto nezdružljivih in mogočih dogodkov, rečemo, da je A **sestavljen** dogodek. Dogodek, ki ni sestavljen, imenujemo **elementaren** dogodek.
2. Množico dogodkov S={A1, A2,…,An} imenujemo **popoln sistem dogodkov**, če se v vsaki ponovitvi poskusa zgodi natanko eden od dogodkov iz množice S.

→ Verjetnost

 **Statistična definicija verjetnosti:** verjetnost dogodka A v danem poskusu je število P(A), pri katerem se navadno ustali relativna frekvenca dogodka A v velikem številu ponovitev tega poskusa.

 **Osnovne lastnosti verjetnosti:**

1. Ker je relativna frekvenca vedno negativna, je verjetnost P(A) ≥ =
2. P(G) = 1
3. Naj bosta dogodka A in B nezdružljiva. Pokaže se lahko, da velja:

P(A∪B)=P(A)+P(B)

1. Za združljiva dogodka A in B (A∩B≠N) velja:

P(A∪B)=P(A) + P(B) – P(A∩B)

1. P(‾A)=1 – P(A)

→ Pogojna verjetnost:

 P'(A)=P(A/B)

Za neodvisna dogodka velja: P(A∩B) = P(A)\*P(B)