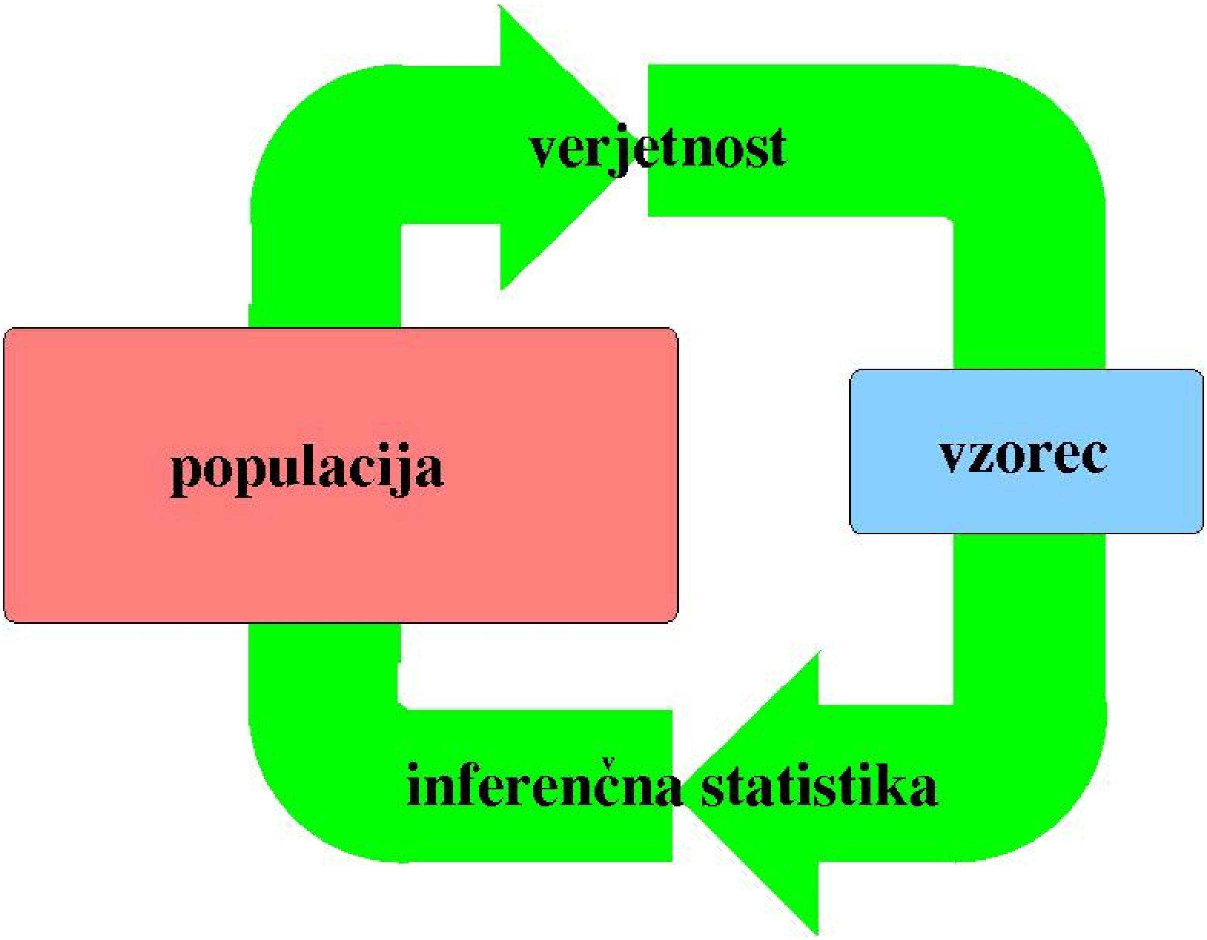
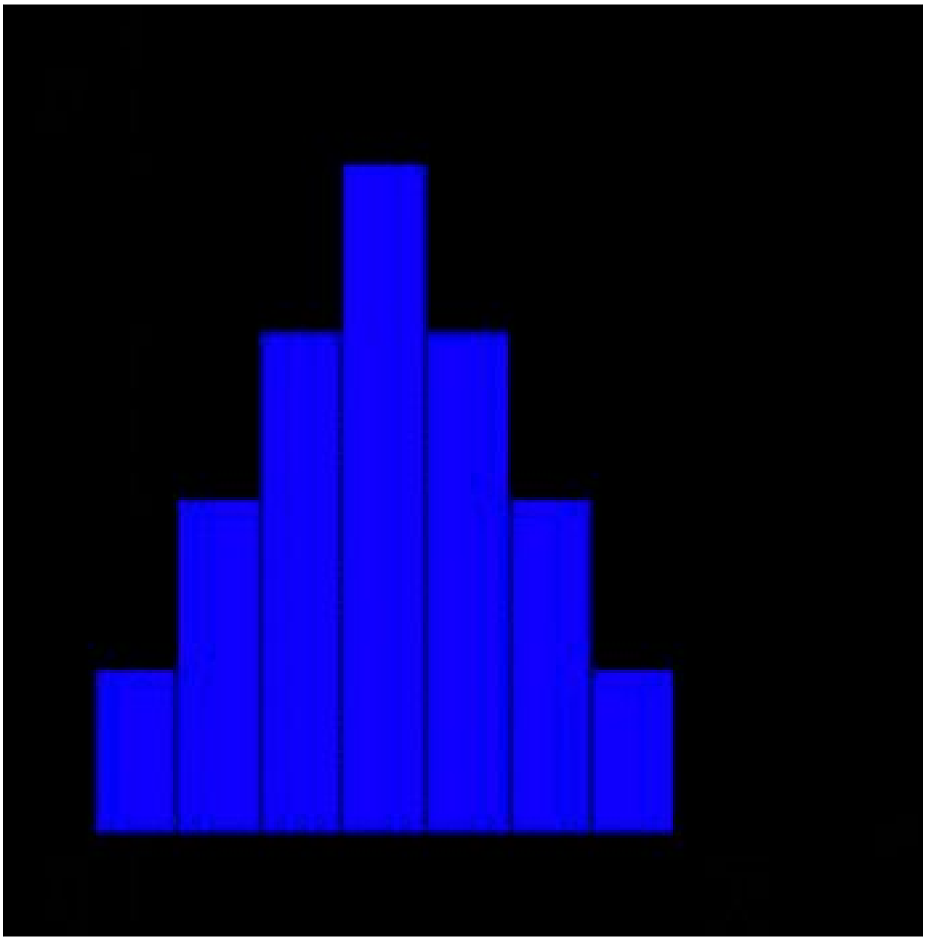
OSNOVE VERJETNOSTI IN STATISTIKE
(Predznanje): **Osnove kombinatorike** (sami oz. na vajah ponovite pravilo vsote, pravilo produkta, permutacije, permutacije s ponavljanjem, kombinacije, binomski izrek in Pascalov trikotnik).
(1. poglavje) **Uvod:** Igre na srečo, kaj je statistika.

- Predstavitev predmeta in ekipe
- Zakaj obravnavamo verjetnost in statistiko hkrati?

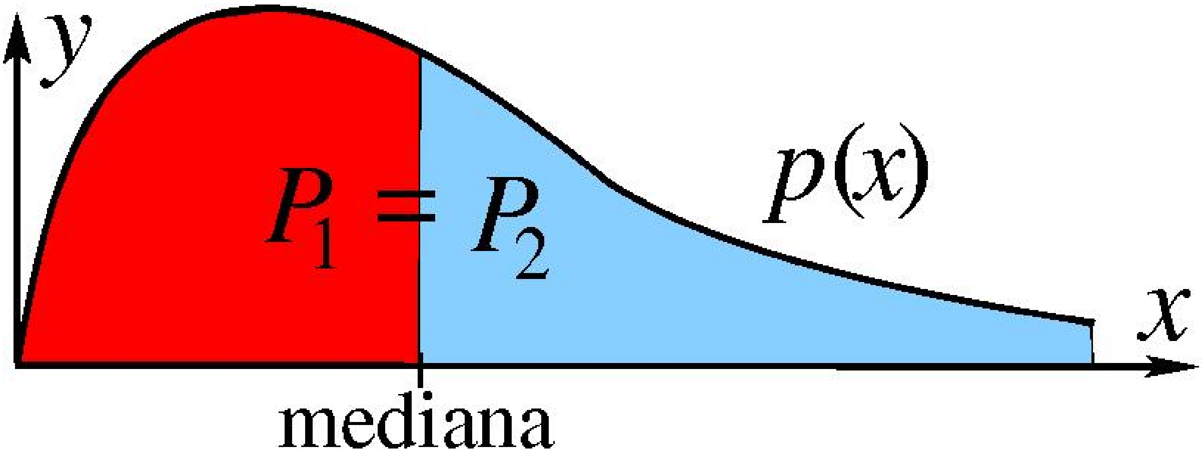


(10. poglavje) **Opisna statistika**

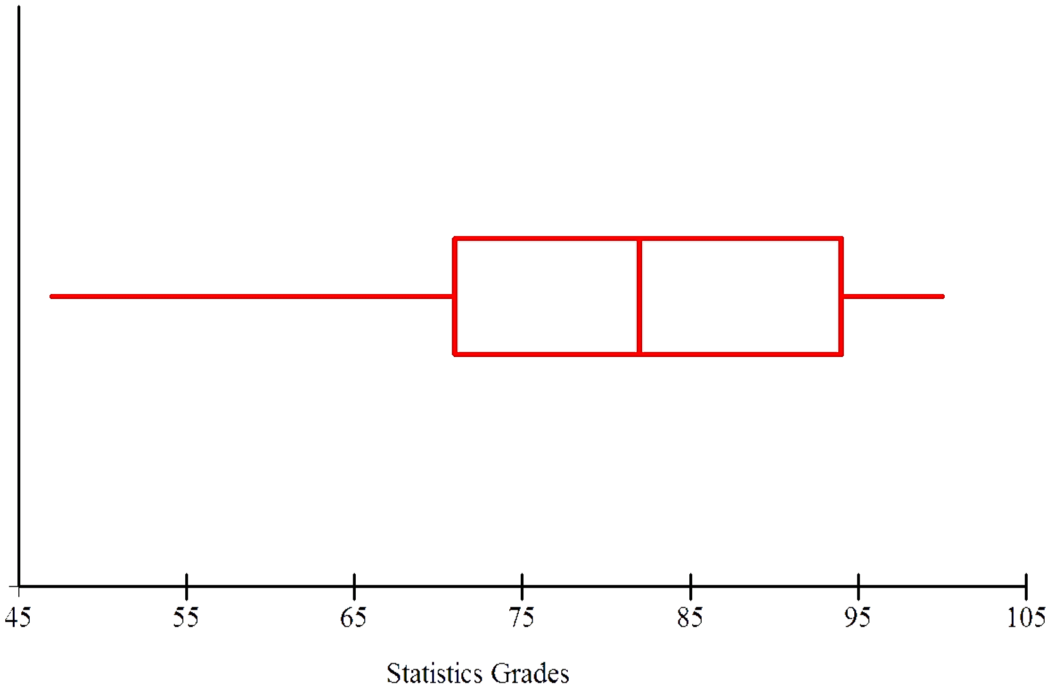
- (10.1) Vrste spremenljivk oziroma podatkov
- (10.2) Grafična predstavitev kvantitivnih podatkov: histogrami



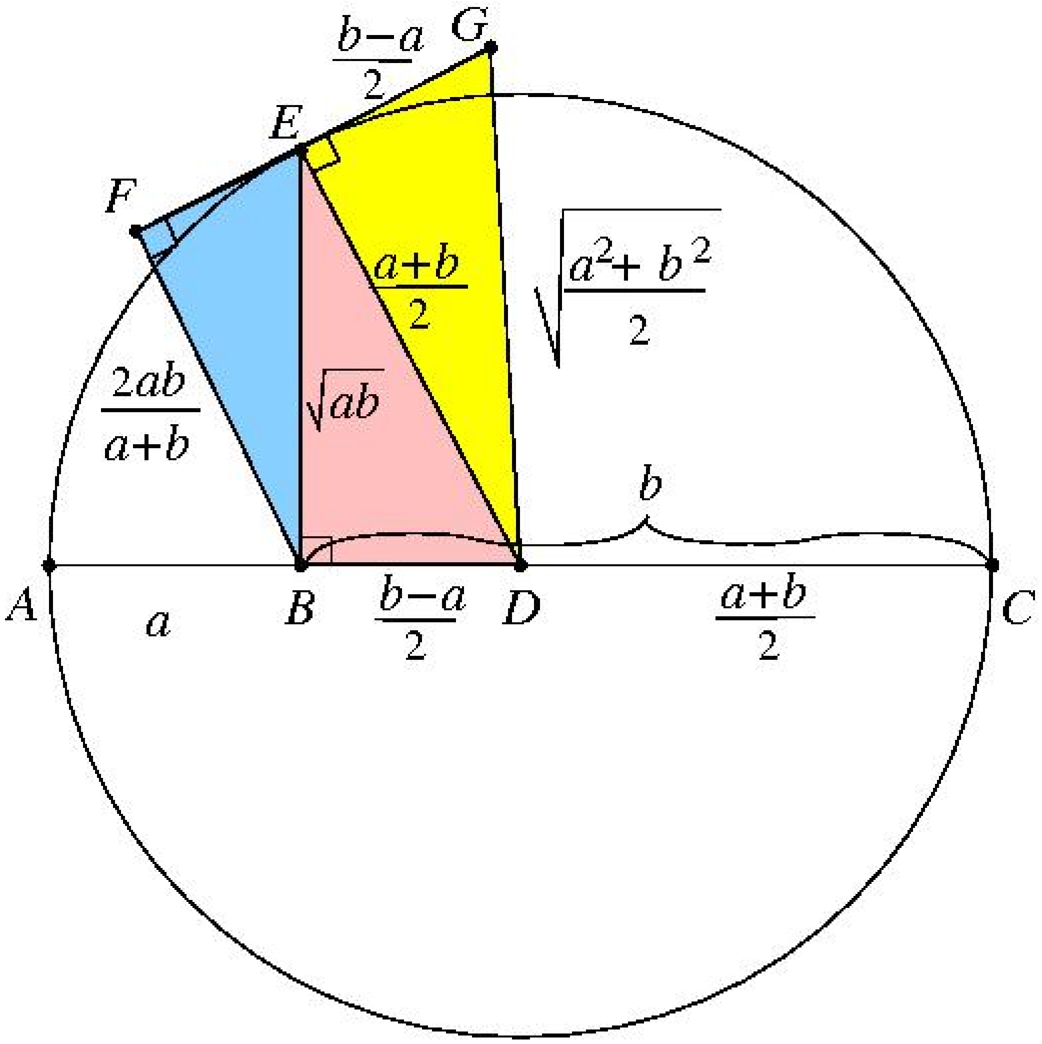
- (10.3) Mere za lokacijo in razpršenost (srednje vrednosti - modus/mediana/povprečje), razpon,



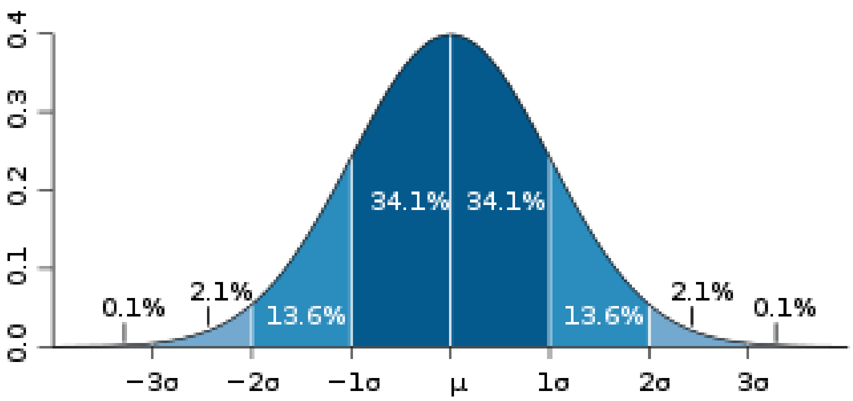
-
- Centili/kvartili, škatla z brki



-
- povprečje in druge (A.8) sredine dveh nenegativnih števil (od najmanjše do največje): kvadratna, aritmetična=povprečje, geometrijska, harmonična;



-
- S potenčno sredino nenegativnih števil k-te stopnje lahko definiramo za
    - $k=1$ povprečje, ki mu pravimo tudi aritmetična sredina,
    - $k=2$ kvadratno sredino,
    - $k=-1$ harmonično srednino ter
    - v limiti, ko gre $k\to 0$, celo geometrijsko sredino
- (še vedno velja ista razvrstitev, katerikoli pa sta enaki natanko tedaj, ko so števila enaka)
- Varianca/disperzija/razperšenost, tj. kvadrat standardnega odklona (ki je kvadratna sredina vseh odklonov)
- Zvonasta oblika in empirična pravila (npr. 68-95-99.7), mere oblike
- Momenti, odklon in empirična pravila povezana z porazdelitvijo (ali podatki) zvonaste oblike



(10.4) Standardizacija ($z=(x-m)/s$), odštevanje premakne izhodišče koordinatnega sistema v desno oz. os x v levo za m, deljanje pa predstavlja razteg oz. skrčitev - odvisno od tega ali je količina s manjša ali večja od 1)

(2. poglavje) **Poskusi, dogodki in definicija verjetnosti:**

- Osnovni pojmi (gotov in nemogoč dogodek)
- Računanje z dogodki: je način dogodka, vsota in produkt dveh dogodkov, nasproten dogodek, (v zapiskih si lahko ogledate še matematično strukturo, ki ji pravimo grupa)
- Popoln sistem dogodkov (najenostavnejši primer: črno-beli svet - naloga o miški).
- Statistična, klasična in geometrijska definicija verjetnosti
- Osnovne lastnosti verjetnosti (ne pozabite tudi na zakon vključitve in izključitve) $$ P(A+B)=P(A)+P(B)-P(AB) $$ za poljubna dogodka A in B
- Vsota (unija) in produkt (presek) treh in več dogodkov
- (2.5)* Aksiomi Kolmogorova

(3. poglavje) **Pogojna verjetnost:**

- (3.1) Avto in dve kozi (Monty Hall problem)
- (3.2) Definicija (praktično enaka običajni verjetnosti, le seznamu pogojev za veljavnost poskusa dodamo, da se mora zgoditi še dogodek B)
- Formula za pogojno verjetnost (ekvivalentna relaciji $P(B)P(A/B)=P(AB)$ seveda le če verjetnost $P(B)\ne 0$),
- Kontingenčna tabela
- Neodvisnost dogodkov (dveh $P(AB)=P(A)P(B)$ ali več)
- Grafični pristop k pogojni verjetnosti in diagram dogodkov
- Naloga o nalezljivi bolezni (našli smo način z uporabo "dvojne" relacije $P(B)P(A/B) =P(AB) =P(A)P(B/A)$ za poljubna dogodka A in B ter seveda prve relacije (pa ne pozabite na relacijo "marmornega kolača": $$ P(B)=P(AB)+P(\bar{A}B) $$ (mi ji bomo rekli formula za totalno verjetnost), kjer smo za popoln sistem dogodkov vzeli $\{A,\bar{A}\}$ - vse skupaj smo morali formulo za pogojno verjetnost uporabiti trikrat).
- Naloga o nalezljivi bolezni (tudi grafična predstavitev reševanja - je nekoliko bolj urejena).
- (3.3) Obrazec za popolno verjetnost (razbitje - posplošitev marmornega kolača) in
- Večstopenjski poskusi in Bayesov obrazec, ki da odgovor na nalogo o nalezljivi bolezni v enem koraku.

(4. poglavje) **Bernoullijevo zaporedje neodvisnih poskusov.**

- (4.1) Računanje $P_i(k)$.
- (4.2) Usojena krivulja v Pascalovem trikotniku.
- (A.6) Stirlingova formula, (A.5) Število e
- Študija oblike histograma, ki ga dobimo iz n-te vrstice Pascalovega trikotnika (A.6)
- Usojena krivulja: \`višino' kvadratov 1x1 pomnožimo s kvocientom $\sqrt{n}/2^n$, širino pa s kvocientom $1/\sqrt{n}$, kar pomeni, da je površina celotnega histograma, ki ga sestavljajo novi pravokotniki, enaka 1, višina pa gre proti $\sqrt{2/\pi}$. S katero funkcijo opišemo obliko tako dobljenega histograma (za dovolj velike n)?
- Histogram je simetričen glede na sredino - pri $n/2$, iz konkavne v konveksno ali obratno preide pri $$ {n\over 2}\pm {\sqrt{n}\over 2} $$ tako da vpeljemo standardizacijo $x={k-(n/2) \over \sqrt{n}/2}$ in z nekaj matematičnega truda dobimo formulo za normalno krivuljo ${1\over \sqrt{2\pi}} e^{-x^2/2}$
- De Moivrov točkovni obrazec (za $p=1/2$) ter njegovo posplošitev Laplaceov točkovni obrazec (za vrednosti p, ki so blizu 1/2): $$ P_n(k) \approx {1\over \sqrt{2\pi npq}} e^{-(k-np)^2/2npq} $$

Zanimivost: Ste vedeli, da se da vseh pet najpomembnejših števil:

- 0 (enota za seštevanje oz. nevtralni element),
- 1 (enota za množenje),
- π (razmerje med premerom in obsegom kroga, oz. med ploščino enotskega kroga in enotskega kvadrata, je iracionalno število $3.14\dots$),
- i (imaginarna enota),

- e (slednje imenujemo tudi Eulerjeva konstanta oz. osnova naravnega logaritma in je enako $\lim_{n\to \infty} (1+{1\over n})^n$, je tudi irracionalno število $2.71\ldots$)
-

povezati v eno samo zvezo na naslednji način $$ e^{i \pi} + 1 = 0? $$ Zares neverjetna zveza (ki ji pravimo **Eulerjeva identiteta**).
(5. poglavje) **Slučajne spremenljivke in porazdelitve:**

- Porazdelitvena funkcija in njene lastnosti
- (5.1) diskretne slučajne spremenljivke: verjetnostna tabela, indikatorska slučajna spremenljivka
- Pričakovana vrednost slučajne spremenljivke X (kot uteženo povprečje): $$ \mu=E(X) \ \ \ \ (\mbox{glavna junakinja našega predmeta}) $$
- Pričakovana vrednost za binomsko porazdelitev $B(n,p)$ je np (privzamemo aditivnost $E(X+Y)=E(X)+E(Y)$ in jo predstavimo kot vsoto indikatorskih slučajnih spremenljivk, ki imajo pričakovano vrednost p)
- Poissonova porazdelitev in povezava z binomsko porazdelitvijo, preko pričakovane vrednosti $\lambda=np$: Poissonov točkovni obrazec (za za velike n in majhne vrednosti p): $$ P_n(k) \approx {(np)^k \over k!} e^{-np} = {\lambda^k \over k!} e^{-\lambda} $$
- Pascalova porazdelitev (poseben primer - geometrijska porazdelitev, $G(p)$, s pričakovano vrednostjo $1/p$)
- Geometrijska porazdelitev na praktičnem primeru - zbiranje kuponov (angl. coupon collector's problem ter occupancy problem, v literaturi se omenja tudi povezava s quality control) - zopet smo uporabili aditivno lastnost pričakovane vrednosti
- Hipergeometrijska porazdelitev (ima tri parametre)
- (5.2) Ponovitev: integrali (štetje kvadratkov ali preko prostornine vode)
- (5.3) Zvezne slučajne spremenljivke (gostota verjetnosti).
- Enakomerna porazdelitev zvezne slučajne spremeljivke na intervalu $[a,b]$, $a\le b$, oznaka ${\cal U}(a,b)$ (angl. uniform)
- Normalna ali Gaussova porazdelitev $N(\mu,\sigma)$, kjer je prvi parameter enak pričakovani vrednosti, drugi pa standardnemu odklonu (katerih pomena že poznamo)
- Standardna normalna porazdelitev, funkcija napake
- Laplaceov intervalski obrazec (raba: binomska porazdelitev $B(n,p)$ je za velike n in p blizu 1/2 dobro aproksimirana z normalno porazdelitvijo $N(np,\sqrt{np(1-p)})$).
- Bernoullijev zakon velikih števil, primer uporabe in dokaz
- Porazdelitev Poissonovega toka - eksponentna,
- Funkcija Gama (označimo jo z g, potem je $g(1)=1$, zadovoljuje rekurzivno zvezo torej je $g(x)=(x-1)g(x-1)$, torej je $g(n)=(n-1)!$ za naravno število n, kar pomeni, da je dovolj, da poznamo njene vrednosti npr. na intervalu (0,1)
- Hi-kvadrat porazdelitev

(6. poglavje) **Slučajni vektorji in neodvisnost slučajnih spremenljivk:**

- Definicije
- Robna porazdelitvena funkcija
- Diskretne večrazsežne porazdelitve, stopničasta ploskev
- (6.1) Polinomska porazdelitev
- Dvojni integral predstavlja prostornino pod ploskvijo: ponazorimo jo lahko z rjuho, ki jo v vsaki točki ravnine dvignemo na določeno višino, le-to predpisuje vrednost $z=p(x,y)$ (če smo prej šteli kvadratke, moramo sedaj šteti kocke).
- Kako na grobo oceniti število kock na sliki (z običajnim štetjem ali z integriranjem)?
- (6.2) Intermezzo: ponovitev dvojnih in dvakratnih integralov
- (6.3) Zvezne večrazsežne porazdelitve
- Robni verjetnostni gostoti
- Primer
- Večrazsežna normalna porazdelitev
- Neodvisnost slučajnih spremenljivk

(7. poglavje) **Funkcije slučajnih spremenljivk in vektorjev**

- (7.1) Funkcije slučajnih spremenljivk
- (7.2) Funkcije in neodvisnost (funkcije slučajnih vektorjev, vsota slučajnih spremenljivk, konvolucija, primeri slučajnih spremenljivk, ki si porazdeljene normalno ali pa po hi-kvadrat)
- (7.3)* Funkcije slučajnih vektorjev (Jaccobijevo determinanto in primer s polarnimi koordinatami lahko predelate samo, če vas zanima in želite pridobiti več izkušenj)
- (7.4) Pogojne porazdelitve (pogojna porazdelitvena funkcija,
- Pogojna verjetnostna funkcija
- Pogojna gostota (v primeru 2-razsežne normalne porazdelitve dobimo enorazsežno normalno porazdelitev)

(8. poglavje) **Momenti in kovarianca:**

- (8.1) Matematično upanje (pričakovana vrednost - njena povezava s povprečjem, kdaj obstaja
- Slučajna spremenljivka, ki ima Cauchyjevo porazdelitev nima mat. upanja
- Pričakovana vrednost vsote dveh slučajnih spremenljivk homogenost (linearen funkcional)
- Študija na primeru binomske porazdelitve,
- Linearnost (raba: pričakovana vrednost populacije μ je enaka pričakovani vrednosti vzorčnega povprečja)
- Matematično upanje produkta in koreliranost
- (A.9) Cauchyjeva neenakost
- (8.2) Disperzija (definicija in lastnosti)
- (8.3) Standardizacija spremenljivke
- (8.4) Kovarianca (definicija, lastnosti, korelacijski koeficient)

Uvod v CLI (metanje dveh/treh kock)

- Parametri populacije: pričakovana vrednost (μ), odklon (σ), delež (π), parameter, ki povezuje dve populaciji - korelacijski koeficient (ρ).
- Na vzorcu: vzorčno povprečje $\bar{x}$, vzorčni odklon s, vzorčni delež (p.
- Slučajne spremenljivke $X_1,X_2,\dots,X_n$ spremljajo dogajanje na vzorcu $(x_1,x_2,\dots,x_n)$, ki mora biti izbran naključno in neodvisno (to pomeni, da so te spremenljike paroma neodvisne), X prečna pa označuje njihovo povprečje in je tudi slučajna spremenljivka.
- (8.5) Pogojno matematično upanje (regresijska funkcija)
- Kovariančna matrika
- (8.6) Višji momenti (začetni moment, centralni moment, asimetričnost, sploščenost, kvantili, mediana, kvartili, kvartilni razmik)

(9. poglavje) **Karakteristične funkcije in limitni izreki:**

- (9.1)* Karakteristična funkcija
- (9.2)* Limitni izreki (verjetnostno konvergira, skoraj gotovo konvergira)
- Preverili, da povprečje neodvisnih slučajnih spremenljivk $X_1,X_2,\dots,X_n$, ki so porazdeljene normalno, limitira k slučajni spremenljivki, ki je tudi normalna
- Če imajo vse slučajne spremenljivke X_i enak začetni moment z_1, potem ima tak moment tudi slučajna spremenljivka, ki predstavlja njihovo povprečje, medtem, ko gre (pada) drugi centralni moment proti 0 z naraščujočim n
- (9.2)* Šibki in krepki zakon velikih števil
- (9.3) Centralni limitni izrek (CLI)
- Čebiševa neenakost (z dokazom za zvezne slučajne spremenljivke) (lahko bi navedli tudi nekaj njenih posledic: izreka Markova in Čebiševa ter Bernoullijev zakon velikih števil)

Uporaba verjetnosti:

- Zamenjalna šifra (premešalka)
- Indeks ujemanja
- Izziv iz 2. svetovne vojne
- Paradoks rojstnih dnevov (v poljubni skupini 23ih ljudi je verjetnost, da imata vsaj dva skupni rojstni dan večja od 1/2)
- Raba v kriptografiji (zgoščevalne funkcije, DLP)

- Raba v teoriji kodiranja (kode za odpravljanje napak: ponavljanje in večinsko pravilo, Hammingova koda - štirim bitom dodamo tri kontrolne in pokažemo, da lahko odpravimo eno napako, glavni mejniki teorije kodiranja)
- Ramseyeva teorija (popoln nered je nemogoč, vsaka dovolj velika struktura vsebuje urejeno podstrukturo, SIM igra, zgornje in spodnje meje za Ramseyjeva števila,...)
- V praksi pogosto potrebujemo učinkovito računanje v kolobarju ostankov pri deljenju z naravnim številom n, tj. $Z_n=\{0,1,...,n-1\}$ in operacijama +, * mod n
    - Deljenje doseżemo z reševanjem linearne Diophantske enačbe $ax+by=c$
    - Z Evklidovim algoritmom (EA) poiščemo $D(a,b)$, ki mora deliti c (to pa ni samo zadosten, pač pa tudi potreben pogoj za obstoj rešitev). Pri tem za naravni števili a, b ($a\ge b$), iščemo celi števili q in r, za kateri velja $a=qb+r$ in $0\le r< b$.
    - Iskanje kvocienta q je računsko zamudna operacija (z njim pa je lahko izračunati ostanek $r=a-bq$), zato naključno izbiramo pare (a,b), ki so manjši od fiksnega naravnega števila M.
    - Slučajna spremenljivka Q sledi vrednost kvocienta q.
    - Zanima nas njena porazdelitev, tj. verjetnosti
        - $P(Q=1)$,
        - $P(Q=2)$,
        - $P(Q=3)$,
        - ...,
    - ki jih brez težav izračunamo z geometrijskim pristopom: $P(Q=m)=1/(m(m+1))$ oziroma
        - $P(Q=1)=1/2$,
        - $P(Q=2)=1/6$,
        - $P(Q=3)=1/12$,
        - ...
    - To pomeni, da se izplača izračunati $a-b$ in preveriti, če je dobljena razlika manjša od b,... (v polovici primerov bo).

(10. poglavlje) **Opisna statistika:** ponovitev in standardizacija.

- CLI za delež

(11. poglavje) **Vzorčenje:**

- Ponovitev [CLI in njegova zgodba pri vzorčenju]
- hitrost centralne tendence pri CLI
- (11.3) Porazdelitev vzorčnih povprečij
- (11.4) Vzorčna statistika:
    - (A) vzorčno povprečje
    - (B) vzorčna disperzija
- Ponovitev variante CLI-ja za vzorčno disperzijo
- (11.5) Nove porazdelitve (Studentova, Fisherjeva)
- (11.6) Cenilke: osnovni pojmi, točkovne cenilke
- (11.6) Cenilke - simetrične funkcije/formule s katerimi iz meritev vzorca ocenimo določen parameter populacije (doslednost, nepristranskost)
- metoda momentov in metoda največjega verjetja;
- Cenilka C_n parametra ζ je nepristranska, če je $E(C_n)= ζ$ (za vsak n)
- (11.7) Vzorčna statistika (nadaljevanje): (C) Vzorčne aritmetične sredine (nas ima res večina zelo podobne sposobnosti?)

CLI za delež,

- (D) Vzorčni deleži
- (E) Razlika vzorčnih aritmetičnih sredin
- (F) Razlika vzorčnih deležev

Na grafično nazoren način si [oglejte z-vrednosti], ki predstavljajo (s procenti od 1):

1. "0 to z" funkcijo napake $Φ(z)$,

2. "Up to z" porazdelitveno funkcijo $F(z)$ (oz. x_p je kvantil p-tega reda, kadar je $F(x_p)=p$) in
3. "z onwards" polovico tveganja, tj. $P(Z \ge z_{\alpha/2})=\alpha/2=P(Z\le -z_{\alpha/2})$

(da boste potem lažje uporabljali tabele za $N(0,1)$ na kolokviju/izpitu)
Pregledna predstavitev vseh izrekov tipa CLI, ki smo jih spoznali doslej (CLI za pričakovano vrednosti, za odklon, za delež in za razlike pričakovanih vrednosti ter deležev).
(12. poglavje) **Intervali zaupanja:**

- (12.1) Pomen stopnje tveganja pri intervalih zaupanja;
- (12.2) Intervalsko ocenjevanje parametrov za:
  - pričakovano vrednost
  - delež
  - odklon
  - razliko pričakovanih vrednosti (tudi v primeru ujemajočih se parov), deležev in za
  - kvocient odklonov
- (12.3) Izbira velikosti vzorca

(13. poglavje) **Preverjanje domnev:**

- *ničelna domneva* H_0 (želimo jo ovreči - lahko je (ne)prametrična, enostavna/sestavljena)
- *alternativna domneva* H_a ali H_1 (nazdružljiva z ničelno, v primeru enostavne parametrične domneve imamo zanjo tri možnosti - katere?)
- *zavrnitveni kriterij* (odvisen od stopnje tveganja α)
- *kritično območje* označimo s K_α
- (13.1) Ilustrativni primer (sodišče)
- Osnovne definicije
  - P-*vrednost* (bistvena stopnja značilnosti/signifikantnosti, tj. največji α, ki ga je vodja eksperimenta pripravljen sprejeti glede na dan vzorec,
  - če je $P\mbox{-vrednost} \gt \alpha$ → FTR H_0, tj. ne moremo zavrniti ničelne domneve, če pa je $P\mbox{-vrednost} \lt \alpha$ → zavrni H_0)
  - *napaka 1. vrste* (zavrnitev ničelne domneve, če je le-ta pravilna - njena verjetnost je α)
  - *napaka 2. vrste* (nismo zavrnili ničelne domneve, čeprav je bila le-ta v resnici napačna - verjetnost tega dogodka označimo z β)
  - *moč testa* 1-β (verjetnost zavrnitve ničelne domneve, ko je le-ta v resnici napačna)
- (13.6) Formalen postopek za preverjanje domnev;
- (13.7) Domneve za povprečje (znan odklon, neznan odklon in velik vzorec);
- (13.8) Domneve za razliko dveh povprečij
- (13.9) Domneve za povprečje razlik
- (13.10) Domneve za delež
- (13.11) Razlika deležev dveh populacij
- (13.12) Analiza variance (preverjanje domneve o varianci populacije in preverjanje domneve o kvocijentu varianc in neodvisnih vzorcih
- (13.13) Domneve o porazdelitvi spremenljivke (enakomerna in normalna).

(14. poglavje) **Bivariatna analiza:**

- (14.1) Povezanost dveh nominalnih spremenljivk
- (14.2) Koeficient asociacije
- (14.3) Povezanost dveh ordinalnih spremenljivk
- (14.4) Povezanost dveh številskih spremenljivk
- (spustili: (14.5) Parcialna korelacija)
- (14.7) Regresijska analiza (linearni model)
- statistično sklepanje o regresijskemu koeficientu, pojasnjena varianca (ANOVA)

(15. poglavje) **Časovne vrste in trendi:**

- (15.1) Primerljivost členov v časovni vrsti
- (15.2) Grafični prikaz časovne vrste
- (15.3) Indeksi
- (15.4) Sestavine dinamike v časovnih vrstah

(16. poglavje) **Načrtovanje eksperimentov**

Zaključki:

- multivariantna analiza
- diskriminacijska analiza
- analiza faktorjev
- naključni sprehodi
- vizualizacija in analiza slik
- ponovno vzorčenje
- kvaliteta podatkov
- inovacija
- komunikacija
- timsko delo