STATISTIKA USMENI

Dodat stat nizove i stat tabele

**STATISTIKA** – znanstvena metoda koja se bavi prikupljanjem, uređivanjem, analizom i tumačenjem podataka.

* DESKRIPTIVNA – u okviru deskriptivne statistike zaključci se donose na temelju *svih podataka*. Ona obuhvaća postupke uređivanja, grupiranja, tabeliranja, grafičkog prikazivanja te izračunavanja različitih statističko-analitičkih veličina
* INFERENCIJALNA – u sklopu inferencijalne statistike zaključci se odnose na temelju *dijela podataka (uzoraka)*. Temelji se na teoriji vjerojatnosti

Kvantili-numerički niz uređen po veličini dijele na jednakobrojne dijelove.Medijan spada među kvantile(kvartil,decil,percentil)

**NORMALNA (GAUSSOVA) DISTRIBUCIJA** – najvažnija distribucija vjerojatnosti.

**-** dvoparametarska funkcija (određena s 2 parametra) : očekivana vrijednost i varijanca

- zvonolika je i simetrična

- budući da aritmetička sredina i standardna devijacija ovise o mjernim jedinicama varijable uvodi se jedinična (standardizirana) normalna distribucija

Distribucije vjerojatnosti diskretne slučajne varijable: (najčešće se koriste binomna i Poissonova)

1. BINOMNA
2. POISSONOVA
3. HIPERGEOMETRIJSKA
4. UNIFORMNA

Binomna distribucija- njena je definicija povezana sa Bernaulijevim pokusom. Bernoullijev pokus je slučajni pokus sljedećih svojstava:        - ima dva ishoda (uspjeh i neuspjeh)

- vjerojatnost ishoda uspjeh je p, a neuspjeh q=1-p

- pokusi su neovisni

Slučajni pokus-pokus je slučajan ako se u definiranim uvjetima može ponavljati, ako postoje barem 2 različita ishoda te ako se ishodi ne mogu predvidjeti sa sigurnošću

Poissonova distribucija**–** granični slučaj binomne distribucije. Prikladna je za opis rijetkih događaja, tj. događaja koji se javljaju s malom vjerojatnošću.

Linearni trend-Model linearnog trenda identičan je modelu jednostavne linearne regresije u kojemu je vrijeme nezavisna varijabla. Prikladan je kada se vremenska pojava mijenja od razdoblja do razdoblja za približno isti apsolutni iznos, tj. kada su prve diferencije približno konstantne.

Eksponencijalni trend-Prikladan je kada se vremenska pojava mijenja od razdoblja do razdoblja za približno isti relativni iznos, tj. kada su verižni indeksi približno konstantni.

Verižni indeksi-njima se prati razvoj pojave u uzastopnim vremenskim razdobljima. Verižni indeks Vt razdoblja t dobije se tako da se vrijednost toga razdoblja podijeli s vrijednošću prethodnog razdoblja te se pomnoži sa sto

Stope promjene-od verižnog indeksa se odbije sto.

Skupni indeks-njima se prati dinamika skupine pojava u vremenu npr.proizvodnja, uvoz, izvoz....

Regresijska analiza-njom se ispituje ovisnost jedne varijable o drugoj varijabli ili o više drugih varijabli, npr.proizvodnja o broju zaposlenih, potrošnja o visini plaće

Procjena aritmetičke sredine– može biti brojem i intervalom

Klasična vjerojatnost(vjerojatnost a priori) – pretpostavlja se da se pokus ponavlja konačan broj puta, pri čemu se vjerojatnost određuje kao omjer povoljnog broja ishoda m i ukupnog broja ishoda n

Statistička teorija vjerojatnosti-

STATISTIČKA VJEROJATNOST (vjerojatnost a posteriori) – broj ponavljanja pokusa je beskonačan, a vjerojatnost se aproksimira relativnom frekvencijom, tj.omjerom apsolutne frekvencije opsega i opsega statističkog skupa

Standardizirana varijabla-linearna transformacija numeričke varijable x. Određuje se tako da se odstupanja numeričke varijable od njezine aritmetičke sredine podijele sa standardnom devijacijom, tj. da se izraze u jedinicama standardnih devijacija. Aritmetička sredina standardizirane varijable jednaka je nuli, a standardna devijacija jednaka je jedan

Ginijev koeficijent-temelj za njegovo utvrđivanje je površina između pravca jednolike raspodjele i Lorenzove krivulje. Što je koncentracija veća to se Lorenzova krivulja više udaljuje od toga pravca

Lorenzova krivulja-grafički prikaz koji se koristi u analizi koncentracije za maksimalnu raspodjelu. To je krivulja koja grafički prikazuje stupanj (ne)jednakosti u distribuciji neke varijable.

Test hipoteza aritmetičkih sredina-

**TESTIRANJE HIPOTEZA O PRETPOSTAVLJENOJ VRIJEDNOSTI ARITMETIČKE SREDINE OSNOVNOG SKUPA –** ovo se testiranje provodi na temelju slučajnog uzorka od N članova

Skupni test-testira se značajnost svih nezavisnih varijabli u modelu. Uvijek se provodi kao f-test.

Snaga testa-vjerojatnost odbacivanja lažne nulte hipoteze.

Koeficijent varijacije-prosječno odstupanje empirijskih vrijednosti od regresijskih izraženo relativno.

Testiranje proporcije

**PROPORCIJA OSNOVNOG SKUPA – parametar koji predstavlja omjer broja članova osnovnog skupa s određenim modalitetom obilježja i opsega statističkog skupa**

Model jednostavne linearne regresije

**MODEL JEDNOSTAVNE LINEARNE REGRESIJE** – njime se izražava odnos među dvjema pojavama. Model sadrži jednu zavisnu i jednu nezavisnu varijablu. Primjenjuje se za varijable koje su u linearnom statističkom odnosu.

Kada se koristi model linearnog trenda?Kada su prve diferencije jednake.

ANOVA - od čega se sastoji? -jednostavna regresija

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| izvor varijacije | stupnjevi slobode | zbroj kvadrata | sredina kvadrata | empirijski t-omjer |
| protumačen modelom | 1 | SP | SP/1 | (SP/1)/SR/(n-2) |
| neprotumačena (rezidualna) odstupanja | n-2 | SR | SR/(n-2) | / |
| UKUPNO | n-1 | ST | / | / |

Grafičko prikazivanje vremenskog niza:·

* INTERVALNI NIZOVI prikazuju se *površinskim i linijskim grafikonima*.
* TRENUTNI NIZOVI prikazuju se samo *linijskim grafikonima*

Radi lakšeg praćenja u grafikon se ucrtava mreža. Prikaz je u pravokutnom koordinatnom sustavu s aritmetičkim mjerilima na osima. Na osi apscisa je mjerilo za varijablu vrijeme, a na osi ordinata za članove vremenskog niza.

Vremenski niz-nastaje kronološkim nizanjem podataka o nekoj pojavi (proizvodnja,uvoz,izvoz)

**PROCJENA TOTALA OSNOVNOG SKUPA**

**TOTAL** – zbroj vrijednosti numeričkog obilježja. Taj je parametar povezan s aritmetičkom sredinom.

  🡪 

Zbog toga se postupak procjenjivanja totala svodi na postupak procjenjivanja aritmetičke sredine.

* procjena totala brojem - 

 - 

* intervalna procjena za veliki uzorak - 

 - koeficijent pouzdanosti

 - standardna pogreška procjene totala

 - razina pouzdanosti (povjerenja)

* intervalna procjena za mali uzorak – kao koeficijent korisnosti koristi se .

Procjena varijance brojem:  

α i β – nepoznati parametri populacije

e – nepoznate vrijednosti slučajne varijable

Komponenta trenda-predstavlja osnovnu tendenciju razvoja pojave u vremenu. Izražava se nekom funkcijom vremena. S obzirom na tu funkciju vremena trend može biti *linearni i eksponencijalni.*

Razina signifikantnosti(razina značajnosti)-alfa – vjerojatnost odbacivanja istinite nulte hipoteze

Srednje potpune vrijednosti

**SREDNJE VRIJEDNOSTI STATISTIČKOG NIZA –** konstante kojima se predstavljaju nizovi varijabilnih podataka.

* POTPUNE – računaju se na temelju svih podataka. U njih se ubrajaju aritmetička, geometrijska i harmonijska sredina.
* POLOŽAJNE – u pravilu su jednake jednom modalitetu statističke varijable. U njih se ubrajaju MOD i MEDIJAN

Procjena proporcija

**PROCJENA PROPORCIJE OSNOVNOG SKUPA**

**PROPORCIJA OSNOVNOG SKUPA** – parametar koji predstavlja omjer broja članova osnovnog skupa s određenim modalitetom obilježja i opsega statističkog skupa

procjenitelj procjene brojem:  m - broj elemenata s određenim modalitetom obilježja u uzorku

 n - broj elemenata uzorka

Sampling distribucija proporcija uzoraka ima oblik binomne ili hipergeometrijske distribucije. Za dovoljno veliki uzorak aproksimira se normalnom distribucijom.

intervalna procjena za veliki uzorak: 

p-proporcija osnovnog skupa

 - koeficijent pouzdanosti

 - standardna pogreška procjene totala

 - razina pouzdanosti (povjerenja)

Kontinuirane slučajne varijable

DISTRIBUCIJA VJEROJATNOSTI

• diskretne slučajne varijable – skup uređenih parova različitih vrijednosti slučajne varijable xi i pripadajućih vjerojatnosti p(xi). Ima slijedeća svojstva:

 - p(xi) ≥ 0

 - ∑p(xi) = 1

Kumulativna funkcija ili funkcija distribucije F(xi) – pokazuje kolika je vjerojatnost da slučajna

varijabla x poprimi vrijednost xi ili manju.

• kontinuirane slučajne varijable – opisuje razdiobu vjerojatnosti na intervalu vrijednosti varijable. Njena svojstva:

- f(x) ≥ 0

- 

Višestruka regresija (ANOVA)

**TESTIRANJE HIPOTEZA O MODELU VIŠESTRUKE REGRESIJE** – najčešće se koriste slijedeći testovi:

1. PARCIJALNI TEST – testira se značajnost podskupa nezavisne varijable
2. POJEDINAČNI TEST – testira se značajnost jedne nezavisne varijable. Može se

 provesti i kao f-test i kao t-test

1. SKUPNI TEST – testira se značajnost svih nezavisnih varijabli u modelu. Uvijek

 se provodi kao f-test.



U hipotezi H0 stoji da nijedna nezavisna varijabla nije značajna za model.

U hipotezi H1 stoji da je barem jedna nezavisna varijabla značajna za model.

Test veličina je empirijski f-omjer iz tabele ANOVA.

Odluka se donosi usporedbom empirijskog f-omjera sa teorijskom vrijednošću koja se očitava iz tablica

Kako se interpretira medijan?
Kako se interpretira varijanca?

Geometrijska mjera disperzije

Aritmetička mjera disperzije

**HARMONIJSKA SREDINA** – recipročna vrijednost aritmetičke sredine recipročnih vrijednosti varijable x

* negrupirani pojedinačni podaci 🡪 
* grupirani podaci 🡪 

Harmonijska sredina manja je od aritmetičke i geometrijske sredine.

Definicije vjerojatnosti

1. KLASIČNA DEFINICIJA (vjerojatnost a priori) – pretpostavlja se da se pokus ponavlja konačan broj puta, pri čemu se vjerojatnost određuje kao omjer povoljnog broja ishoda m i ukupnog broja ishoda n 

2. STATISTIČKA VJEROJATNOST (vjerojatnost a posteriori) – broj ponavljanja pokusa je beskonačan, a vjerojatnost se aproksimira relativnom frekvencijom, tj.omjerom apsolutne frekvencije opsega i opsega statističkog skupa.

3. SUBJEKTIVNA VJEROJATNOST – to je broj iz intervala [0,1] određen na temelju prosudbe okolnosti relevantnih za nastup slučajnog događaja

Vremenski niz - skup kronološki uređenih vrijednosti koje predstavljaju neku pojavu (proizvodnja, uvoz, izvoz).

Rezidualna odstupanja- osnove za mjerenje reprezentativnosti regresije

**slučajni uzorci** – svaki element ima vjerojatnost izbora veću od nule. Kod ovih uzoraka moguće je izračunati grešku. Analiziraju se metodama inferencijalne statistike.

* + *jednostavni slučajni uzorak* – svaki element ima jednaku vjerojatnost izbora. primjenjuje se kod homogenih skupova.
	+ *stratificirani uzorak –* prikladniji je kod skupova koji pokazuju veći stupanj varijabilnosti. Jedinice se izabiru iz homogenih dijelova osnovnog skupa koji se nazivaju STRATUMI.
	+ *uzorak skupina –* u uzorak se ne izabiru pojedini elementi nego njihove skupine

**REGRESIJSKA ANALIZA –** njom se ispituje ovisnost jedne varijable o drugoj varijabli ili o više drugih varijabli, npr. proizvodnja o broju zaposlenih, potrošnja o visini plaće

**MJERE ASIMETRIJE –** njima se mjeri način rasporeda podataka prema aritmetičkoj sredini ili nekoj drugoj vrijednosti. Najvažnije su:

1. Koeficijent asimetrije α3 - potpuna mjera
2. Pearsonova mjera nepotpune
3. Bowleyeva mjera mjere

**1. KOEFICIJENT ASIMETRIJE α3**

**-**MOMENTI OKO SREDINE – aritmetičke sredine odstupanja vrijednosti numeričke varijable od njezine aritmetičke sredine podignuti na neku potenciju

- s obzirom na veličinu r govori se o nultom, prvom, drugom, trećem ili četvrtom momentu oko sredine

 - koeficijent asimetrije α3 je omjeru trećeg momenta oko sredine i standardne devijacije podignute na treću potenciju 🡪  obično se kreće u intervalu ±z, a u određenim slučajevima može biti izvan toga intervala:

α3 = 0 🡪 simetrična distribucija

α3 nije jednako 0 🡪 asimetrična distribucija

2.Pearsonova mjera-definira se kao standardizirano odstupanje vrijednosti medijana ili moda od aritmetičke sredine, temelji se na odnosu srednjih vrijednosti u distribucijama frekvencija

3.Bowleyeva mjera-mjera asietrije koja se temelji na odnosu medijana i kvartila, kreće se u intervalu ±1

Proporcije testiranja

**SREDNJE VRIJEDNOSTI STATISTIČKOG NIZA –** konstante kojima se predstavljaju nizovi varijabilnih podataka.

* POTPUNE – računaju se na temelju svih podataka. U njih se ubrajaju aritmetička, geometrijska i harmonijska sredina.
* POLOŽAJNE – u pravilu su jednake jednom modalitetu statističke varijable. U njih se ubrajaju MOD i MEDIJAN.

Testiranje o razlici proporcija-nismo radili

Kada se upotrebljava eksponencijalni trend?Kada su verižni indeksi približno konstantni.

KOMPONENTA TRENDA – predstavlja osnovnu tendenciju razvoja pojave u vremenu. Izražava se nekom funkcijom vremena. S obzirom na tu funkciju vremena trend može biti *linearni i eksponencijalni*

Sezonska komponenta-posljedica je klimatskih faktora, ritma, proizvodnje, potrošnje… Očituje se onda kada se vremenska pojava obnavlja na približno isti način unutar jedne godine.

CIKLIČKA KOMPONENTA – očituje se onda kada se vremenska pojava obnavlja na približno isti način s periodom od 2 ili više godine.

Koeficijent korelacije

**KORELACIJSKA ANALIZA –** njome se određuju jakosti veze. Mjera jakosti veze je koeficijent korelacije.

Koeficijent korelacije drugi je korijen iz koeficijenta determinacije: 

**MODELI TRENDA –** njima se statistički opisuje dugoročna kovarijacija pojave sa vremenom



1. ADITIVNI MODEL: 
2. MULTIPLIKATIVNI MODEL: 



1. MODEL LINEARNOG TRENDA: 
2. MODEL EKSPONENCIJALNOG TRENDA: 

Distribucije vjerojatnosti kontinuirane varijable i objasniti najvažnije

Distribucije vjerojatnosti kontinuirane slučajne varijable:

1. NORMALNA (GAUSSOVA)
2. STUDENTOVA (T-DISTRIBUCIJA)
3. (HI-KVADRAT)
4. F-DISTRIBUCIJA
5. UNIFORMNA KONTINUIRANA
6. EKSPONENCIJALNA
7. **NORMALNA (GAUSSOVA) DISTRIBUCIJA** – najvažnija distribucija vjerojatnosti.

**-** dvoparametarska funkcija (određena s 2 parametra) : očekivana vrijednost i varijanca🡪

- zvonolika je i simtrična

- budući da aritmetička sredina i standardna devijacija ovise o mjernim jedinicama varijable uvodi se jedinična (standardizirana) normalna distribucija 🡪  

 Jedinična normalna distribucija je tabelirana. U poljima tabele nalaze se površine koje predstavljaju vjerojatnosti. U pred-stupcu tabele su sve vrijednosti izražene kao brojevi s jednom decimalom. Druga decimala nalazi se u zaglavlju.Budući da je distribucija simetrična u tabeli su dane samo vrijednosti z.



🡪 označena površina označava vjerojatnost da slučajna varijabla poprimi vrijednost iz intervala od 0 do z.

1. **STUDENTOVA DISTRIBUCIJA**

- njen oblik je određen veličinom n

- za n>30 (veliki uzorak) distribucija se po obliku približava normalnoj distribuciji

- za n<30 (mali uzorak) distribucija je više razvučena na obje strane uzduž apscise

- i ona je zvonolika i simetrična te tabelirana

- u pred-stupcu tablice nalaze se stupnjevi slobode, u zaglavlju su vjerojatnosti. U poljima tabele su

 kritične vrijednosti.



 Vjerojatnost da će slučajna varijabla koja se ravna po Studentovoj distribuciji poprimiti vrijednost veću od tα iznosi α.

ANOVA-višestruka regresija

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| izvor varijacije | stupnjevi slobode | zbroj kvadrata | sredina kvadrata | empirijski t-omjer |
| protumačen modelom | k | SP | SP/k | SP/k/(SR/(n-2)) |
| neprotumačena (rezidualna) odstupanja | n-(k+1) | SR | SR/(n—(k+1)) | / |
| UKUPNO | n-1 | ST | / | / |