KLJUČNI POJMOVI 1:

1. **Populacija** – je potpun skup mogućih mjerenja ili podataka o nekom kvalitativnom svojstvu kojem treba dati zaključak.
2. **Uzorak** – predstavlja skup mjerenja (podataka) koja su sprovedena u toku istrazivanja.
3. **Kvalitativna dimenzija skupa** – karakteristika koju želimo analizirati. Obilježje uvijek zavisi od cilja istraživanja. U jednostavnijim slučajevima može biti samo jedno, kao broj živorođene djece (ako nas ne zanima njihov spol), a u složenijim slučajevima više njih.
4. **Kvantitativna dimenzija skupa** – Svaka pojava ima neko svoje kvantitaivno obilježje izraženo jedinicom mjere i količinom a u statistici ga nazivamo pokazateljem. Svakom kvalitativnom obilježju mora biti pridružen određeni pokazatelj kojem se to obilježje može ili treba kvantificirati.
5. **Prostorna dimenzija skupa** – Većina statističkih skupova vezana je za neku lokaciju u prostoru. Svjetska statistika stanovništva ili proizvodnje obuhvata ili barem nastoji obuhvatiti cijelu zemaljsku kuglu. Pri sadašnjem stanju stvari ovdje nije moguće računati s popisom koji bi se izveo u istom trenutku na cijeloj zemaljskoj kugli ili niti s jedinstvenim informacionim sistemom za cio svijet. Taj system nije dovoljno izgrađen, pa se zato koriste statistički izvori pojedinih zemalja, ali to ne mijenja suštinu : da prostor na koji se odnosi statistička pojava – mora biti definisan unaprijed.
6. **Vremenska dimenzija skupa.** Vrijeme se u statistici javlja dvojako: u smislu trenutka u kom se pojava analizira ili u smislu vremenskog perioda na koji se odnose podaci. Kad se radi o trenutku snimanja želi se doći do informacija o stanju skupa koji posmatramo na određeni dan. Te informacija nam omogućavaju da dođemo do osnovnih obilježja skupa i do structure poojave. Vremenski period nužno je definisati kad se radi o pojavama koje traju duzi vremenski period. Kao osnovicu ovdje mozemo odabrati i kraće periode ali kako god ih odabrali moramo koristiti isti period za iskazivanje pojave.
7. **Nominalno obilježje** – je manje ili više određeno njenim nazivom. Pojam muškarac je takvo obilježje i ono može biti dovoljno za utvrđivanje ukupnog broja muškaraca. Ali ako se treba utvrditi prosječna starost stanovništva moraće biti utvrđen i broj žena, sabrane sve godine i sva lica oba pola, a zatim podjeljen ukupan broj godina s ukupnim brojem stanovnika. Grupisanje mora biti izvedeno potpuno, tako da obuhvaćena lica budu raspoređena u neku od grupa. U slučajevima kad je takvih grupa srazmjerno mnogo obično se u posebne grupe izdvoje najčešći modaliteti a ostatak unese u modalitet „ostali".
8. **Numeričko obilježje** – nekad se može grupisati srazmjerno lako. Kod rada u smjenama moguće je govoriti o jednoj, dvije ili tri smjene. Međutim, ako treba utvrditi prosječna primanja, tada bi precizna računica zahtijevala da se formira neprekidni niz od minimalnog do maksimalnog primanja, što bi znatno komplikovalo račun. U takvim slučajevima moguća primanja podjeljena su u razrede od kojih nanjiži obično definišemo kao " do 100 KM mjesečno", najviši kao "preko 2000 KM", a između tih dviju veličina koristimo raspon "100-200", "200-300" itd.
9. **Prostorno obilježje** – je vezano za lokaciju, odnosno za teritorijalni raspored stanovništva i privredne aktivnosti. Struktura obilježja može biti definisana različito. Cilj prostorne definicije obilježja je da omogući praćenje promjena analizirane pojave na datom prostoru, poređenje s drugim prostornim jedinicama i analizu dinamike u toj prostornoj jedinici.
10. **Vremensko obilježje** – može se odnositi na trenutak ili period. Za vitalnu statistiku evidencija se izvodi dnevno, ali se podaci obrađuju po mjesecima iz kojih se dobije podatak za godinu dana ili duže. Privredna statistika uglavnom se odnosi na period od godinu dana, za koji su produzeća dužna dostavljati društvu svoje poslovne bilanse. Kad je potrebno raspolagati detaljnijim informacijama o promjenama npr. radi uvida u sezonske oscilacije, statistički uredi organizuju posebna istraživanja koja često imaju redovan karakter.
11. **Obilježje ranga** – grupišu se prema rangu. Rangirana može biti npr. veličina porodice, obrazovni nivo stanovništva ili kvalifikacioni sastav zaposlenih. Kod obrade podatala, pa i u anketama, često se ovo primjenjuje kao numerički opis nominalnih modaliteta datog obilježja.
12. **Anketa** – je procedura ispitivanja većeg broja ispitanika koja počiva na uniformiranim obrascima – upitnicima. Upitnik sadrži unaprijed definisanu listu pitanja na koja treba dobiti odgovor od svakog pripadnika uzorka ili populacije. Pitanja se obično formulišu tako da nude moguće odgovore u dvije ili više alternativa. Kada je dobro napravljen, upitnik daje vrlo pouzdane informacije, ali njegovo sastavljanje zahtijeva visok strepen stručnosti.
13. **Intervju** – je oblik razgovora s ispitanikom u kom se od ispitivača očekuje da je dobar poznavalac materije, a za ispitanika se pretpostavlja da je jedan od ključnih informatora. Ispitivač obično ima pripremljena pitanja na koja želi odgovor, ali ih tokom intervjua može reoblikovati kako bi došao do pouzdane informacije o pojavi i njenim dimenzijama (npr. novinarski).
14. **Fokus grupa** – je oblik prikupljanja informacija u kojem mala grupa članova tima analizira problem te idetifikuje obilježja i njihove dimenzije tokom jedne ili više sesija. Kada kreativno pristupi analizi, grupa može znatno doprinijeti kvalitetu prikupljenih informacija.
15. **Direktno posmatranje pojave** – primjenjuje se obično u timu stručnjaka od kojih su svi dobro upućeni u pojavu i mogu, na osnovu zajedničkog koncepta, dovesti istraživanje do kraja. Ovaj tip istraživanja može se izvoditi paralelno u različitim laboratorijama i kada se rezultati podudaraju zaključci mogu biti jako pouzdani. Metoda nema izvještačenosti kao kod ostalih metoda i daje procjenitelju dobar uvid u pojavu, ali može biti skupa. Metodu je također, teško standardizovati.
16. **Institucionalni izvori podataka** – su podaci koje postojeće institucije obično prikupljaju kao statističku građu koja je najčešće javnog karaktera. To mogu biti statistički uredi, odjeli za planiranje pri ministarstvima ili opštinama. Podaci se prikupljaju i obrađuju sistematski i periodično tako da se mogu izraditi vremenske, prostorne ili drugačije formulisane serije podataka. Prednost je u tome što se takve informacije srazmjerno lako dobivaju i mogu biti obimne. potrebno specijalističko znanje.
17. **Dokumentarni izvori podataka** – predstavljaju dokumetaciju o posmatranoj pojavi koja se može nalaziti u obliku izvještaja u bibliotekama, kod specijalizovanih agencija, zvaničnika ili ranijih istraživača. U ovakvim slučajevima građa je prikupljena, izvršena je sistematska analiza podataka i eventualno postoji izvještaj koji predstavlja interpretaciju statističkih podataka. Prednost ovog izvora je u postojanju građe. Nedostatak je obično u tome što je takava građa nepotpuna, a može biti nedostupna.
18. **Ključni informator** – su informatori o ključnim karakteristikama problema koji se istražuje i o načinu na koji je moguće ili poželjno pribaviti statističku građu od strane lica ili institucija koje su u pozicijj da dobro poznaju problem koji statistički treba obraditi.

19. **Pokazatelji dinamike –** su statistički pokazatelji kojima se nastoji utvrditi da li i koliko se neka pojava promjenila tokom vremena, pa ih – u tom smislu – možemo nazvati i vremenskim. Ovdje se najčešće koriste prosti indeksi i proste stope promjene.

20. **Pokazatelji strukture –** je statistički pokazatelj kojim se nastoji kvantitaivno izraziti neka promjena kod pojava koje se sastoje od više elemenata (složenih pojava) gdje se pri analizi gotovo isključivo koriste koeficijenti i procenti.

21. **Bazni indeks –** je omjer kojim se utvrđuje bazni period koji može biti bilo koji period, prethodni ili kasniji.

22. **Lančani indeks –** postoji onda ako su indeksi cijena izračunati samo za tekuću godinu u odnosu na prethodnu, a potrebno je da obradimo podatke koji se odnose na seriju dužu od dvije godine,a pa se oni još nazivaju i godišnjim indeksima.

23. **Kumulativni indeks –** je indeks koji se odnosi na više perioda ili godina.

24. **Prosti indeks –** je omjer između veličine jedne pojave u tekućem u odnosu na prethodni period, kojeg obično nazivamo baznim.

25. **Opšti indeks –** koristi se kod pojava koje sadrže brojne komponente kao što su cijene, inflacija i slično.

26. **Prosta stopa promjene –** je najčešće korišćena kategorija u tumačenju kretanja. Stopa je zapravo izvedena iz indeksa, prema definiciji koja kaže da iz indeksa dobijamo stopu promjene tako što indeksu oduzmemo 1, indeks tako što stopi promjene dodamo 1. Za stopu promjene možemo reći da ona predstavlja omjer između priraštaja neke pojave i same pojave.

27. **Prosječna stopa promjene –** je stopa koja se izračunava iz kumulativnog indeksa u istom periodu. Tu treba voditi računa o broju godina odnosno dužini perioda na koji se odnosi i jedan i drugi pokazatelj.

28. **Opšta stopa promjene –** je stopa koja se izračunava kod složenih pojava primjenom odgovarajućih pondera ili težina.

29. **Prirodni priraštaj –** je razlika uzmeđu živorođenih i umrlih lica.

30. **Vitalni indeks –** je omjer izemđu živorođenih i umrlih lica.

31. **Bruto domaći proizvod –** je sinonim za ostvareni obim proizvodnje i usluga i, kada se iz njega isključi inflacija uzima se da on održava fizički obim proizvodnje u datom periodu.

32. **Sistem nacionalnih računa –** predstavlja raspored svih subjekata kao što su školstvo, zdravstvo, javna uprava i drugi u neku od djelatnosti utvrđenih standardnom međunarodnom klasifikacijom.

KLJUČNI POJMOVI 2 :

33. **Startni prozor –** prozor predstavlja desktop našeg računara na kojem obično postavi tzv. kratica u našem slučaju Excela, a i mnogih drugih programa.

34. **Radni prozor –** predstavlja prozor koji nam se prikaže prilikom pokretanja Excela.

35. **Radna knjiga –** je predpostavljeni naziv koji se pojavljuje pri svakom otvaranju novog dokumenta.

36. **Radni list –** predstavlja radnu površinu naše radne knjige defaultni naziv mu je Sheet1, Sheet2…

37. **Radna ćelija –** predstavlja presjek između kolone i reda u koju upisujemo naš podatak

38. **Referenca apsolutna –** nam pomaže da uz pomoć znaka $ fiksiramo red, kolonu ili ćeliju i da bez obzira u kojoj se ćeliji nalazila formula uvijek treba koristiti podatak fiksiran apsolutnom referencom.

39. **Referenca relativna –** ima klizni karakter. Kad se uklone znaci dolara ispred reda i kolone u ćeleiji referenca postaje relativna. Ideja kopiranja formula iz jedne u drugu ćeliju ovdje je iskorištena tako da se ista formula ne mora unositi svaki put.

40. **Referenca mješovita –** ima red ili kolonu definisanu kao apsolutan.

41. **Referenca trodimenzionalna –** se reference se koriste kad se žele analizirati podaci u istim ćelijama ili njihovoj seriji iz više radnih listova unutar radne knjige.

42. **R1C1 –** jedan od načina na koji je moguće adresirati ćelije da slovom R označimo red a slovom C kolonu, iza kojih dolazi broj reda odnosno kolone.

43. **Operator aritimetički –** U aritimetičke operatore spadaju sabiranje(+), Oduzimanje ili Negacija(-), Množenje(*), Dijeljenje(/), Procenat(%), Potenciranje(^)

44. **Operator poređenja –** Jednako(=), Manje od(<),Vece od(>), Jednako ili vece od(=>), Jednako ili manje od(=<), Nije jednako(<>);

45. **Operatror referenci –** (Colon) kreira jednu reference za sve ćelije u zadanom rasponu, (Coma) Operator unije kombinuje više referenci u jednu, (Space) Operator presjeka referenca za presjek između drugih dviju refernci.

46. **Operator tekstovni –** (&) Povezuje dvije vrijednosti da se dobije neprekidna tekstualna vrijednost

47. **Ugniježdena funkcija –** U nekim slučajevima postoji potreba da se unese funkcija na mjesto njenog argument ili neke druge funkcije npr.
= IF(AVERAGE(F2:F5)>50,SUM(G2:G5))

48. **Formula sa naslaganim nazivima**- U formulama je moguće koristiti više naziva u istom ili različitom redu ili koloni. Ako je npr ćeliji C5 dat naziv<<West<<, a ćeliji E6<<Projected<< njihov zbir se može dobiti unosom formule =SUM(West Projected). Ovo se u Excelu naziva naslaganim formulama.

KLJUČNI POJMOVI 3 :

49. **Slučajni uzorak –** je izbor koji počiva na principu da svaki elemenat populacije ima jednaku šansu da bude izabran uzorak. Obično se to radi putem slučajnih brojeva.

50. **Namjerni uzorak –** je izbor u kojem istraživač iz populacije namjerno u uzorak odabere one elemente koje smatra tipičnim za populaciju. Prednost ovog metoda izbora je u tome što se od ovako odabranih ispitanika obično može lakše doći a potrebne informacije je jednostavnije prikupiti.

51. **Kvota uzorak –** je uzorak u kojem svaki anketar – uz uputstvo o načinu anketiranja – dobije broj ili kvotu ispitanika koje treba anketirati. Uputstvo obično sadrži i karakteristike koje anketirano lice treba posjedovati.

52. **Prigodni uzorak –** je onaj uzorak kada koristimo pogodnost postojanja nekih spiskova (telefonski imenik, adresar…) i iz njih biramo jedinice koje ćemo anketirati. Izbor jedinica može biti po principu slučajnosti, ali tada sui z ankete isključeni svi koji se ne nalaze na tom spisku, pa uzorak sadrži sistematsku grešku, koja deformiše sadržaj populacije.

53. **Stratifikovani uzorak –** je uzorak izdijeljen u grupe (stratume) od kojih se svaka razlikuje od druge po svom

kvantitavnom obilježju. Raspon tih razlika mora biti takav da pokriva cijelu strukturu osnovnog skupa.

54. **Uzorak skupina** – primjenjuje se u slučajevima kada je metodom prostog, slučajnog ili stratifikovanog uzorka suviše složen za izvođenje ili zahtijeva vioke izdatke. U takvim slučajevima možemo se odlučiti da odaberemo skupinu elemenata umjesto elemenata. Više takvih skupina, koje se međusobno isključuju, daće nam uzorak na osnovu kojeg procjenjujemo ponašanje populacije.

55. **Statistički korak** – se koristi kod biranja elementa slučajnim uzorkom iz tablice nakon što se izvrši identifikacija prvog elementa, nakon čega se vrši biranje ostalih elmenata ovim korakom.

56. **Tablica slučajnih brojeva** – se koristi kada se radi o ozbiljnijem istraživanju i možemo je naći u Excelu. Da bismo nju mogli koristiti moramo poznavati broj elemenata u uzorku. Koji god broj izvučemo iz tablice koristićemo ga za idetifikaciju elemenata iz uzorka. Na primjer ako treba da izvučemo 300 elemenata iz skupa veličine 3000 jedinica znači da treba da obuhvatimo svaki 10. elemenat skupa. Nakon što smo identifikovali prvi elemenat uzorka uzećemo svaki naredni deseti elemenat uzorka, unaprijed i unazad.

57. **RAND** – je funkcija Excela za dobijanje slučajnih uzoraka tako što se dobijeni broj može pomnožiti sa 10, 100 ili bilo kojom konstantom da bismo ga prilagodili veličini uorka.

KLJUČNI POJMOVI 4:

58. **Prosta aritmetička sredina** – dobija se tako što se svi članovi datog niza saberu i podijele s brojem članova.

59. **Ponderisana aritmetička sredina** – je sredina koja se izračunava iz frekvencijskog skupa, jer broj ponavljanja ustvari predstavlja frekvenciju tog člana u skupu.

60. **Frekvencijska sredina** – je sredina koja se izračunava iz frekvencijskog skupa, jer broj ponavljanja ustvari predstavlja frekvenciju tog člana u skupu.

61. **Opšta aritmetička sredina** – je sredina koja se izračunava iz frekvencijskog skupa, jer broj ponavljanja ustvari predstavlja frekvenciju tog člana u skupu.

62. **Razredna sredina** – je sredina koja se iračunava kada je uzorak sistematizovan tako da članovi skupa budu rapoređeni u klase ili razrede na neki način. Obično je to prema apsolutnoj vrijednosti člana

63. **Poziciona sredina** – postoji kada se u distribuiranim i homogenim skupovima utvrđuje srednja veličina u skupu a time je definisan i njen položaj (u sredini članova skupa, poredanih po njihovoj veličini).

64. **Lokaciona sredina** – postoji kada se u distribuiranim i homogenim skupovima utvrđuje srednja veličina u skupu a time je definisan i njen položaj (u sredini članova skupa, poredanih po njihovoj veličini).

65. **Modus** – je veličina člana koji se najčešće javlja u nizu. U uzorcima za koje tražimo sredinu često se članovi niza ponavljaju. Postupak utvrđivanja modusa je : **a)** poredati sve članove niza po veličini, **b)** odabrati veličinu u nizu koja se najčešće ponavlja. Ako u nizu postoji više članova s jednakim frekvencijama, dok ostali članovi imaju manju frekvenciju, tada se kaže da postoji više modusa. Ako se nijedan broj u nizu ne ponavlja smatra se da modusa nema.

66. **Medijana** – je broj koji se u skupu brojeve uređenih po veličini, nalazi u sredini. Dobija se tako što se : **a)** brojevi u nizu poredaju po veličini; a zatim **b)** ako je broj članova u nizu neparan, medijana predstavlja broj koji se nalazi tačno u sredini skupa, **c)** ako je broj članova u nizu paran, uzmu se dva srednja broja i podijele sa dva.

67. **Matematičko očekivanje** –javlja se kod skupova i predstavlja osobinu aritimetičke sredine. Aritimetička sredina, načelno, to bolje prestavlja skup što su manja odstupanja članova skupa od njegove sredine. Ta odstupanja ili dispreziju greške možemo uzeti kao pouzdano mjerilo reprezentativnosti uzroka njegovom sredinom.

68. **Geometrijska sredina** – se primjenjuje kada skup sadržava takve elemente koji, kada se poredaju po apsolutnoj vrijednosti, pokazuju tendenciju ubrzanog rasta. Tako mogu izgledati i stratumi u uređenom skupu, a tu karakteristiku ima i složena kamatna stopa.

69. **Harmonijska sredina** – se koristi za utvrđivanje sredine iz skupova u kojima su članovi na neki način recipročni. To može biti slučaj kod produktivnosti rada (broj proizvedenih jedinica po satu ili radniku) ili kod prosjeka relativnih brojeva iz kojih se može izvaditi ponderisana aritimetička sredina.

70. **Frekvencija** – je broj ponavljanja članova u skupu.

71. **Histogram frekvencija** – su linije između apcise, na kojoj se nalaze svi članovi skupa i ordinate frekvencije svakog konkretnog člana u svakom od uzoraka, koje spajaju frekvencije u svakom od datih uzoraka.

72. **Momenti distribucije uzorka**- je prosjek svih odstupanja od sredine što predstavlja mjerilo ditribucije uzorka. To mjerilo općenito nazivamo mometnom disperzije. Statistika poznaje momente oko sredine, oko nule ili nekog drugog broja.

73. **Varijansa uzorka** – je jako korištena mjera u statistici i predstavlja drugi momenat disperzije oko sredine. Varijansa uzorka nastupa nakon prvog momenta disperzije oko sredine koji je poznati kriterij sredine gdje suma odstupanja mora biti jednaka nuli.

74. **Varijansa populacije** – je razlika između momenata disperzije oko nule i za njeno izračunavanje koriste se originalni podaci. Veliki skupovi se računaju pomoću varijanse populacije.

75. **Standardna devijacija** – je linearni izraz varijanse. Budući da sve empirijske veličine u skupu manje ili više odstupanja od sredine, postoji neka srednja veličina tih odstupanja to je stand. devijacija.

76. **Koeficijent varijacije** – predstavlja omjer između standardne devijacije i aritmetičke sredine, koje imaju istu jedinicu mjere. Ovaj pokazatelj je relativnog karaktera- neutralan je – i može se koristiti za poređenje raznorodnih skupova, što mu daje prednost. On takođe govori o tome koliko je rasturanje, ali to izražava koeficijentom, procentima ili promilima.

77. **Granice povjerenja** – je ocjena pokazatelja koji se uzimaju kao vjerodostojni a to su sredina i standardno odstupanje od nje. Kada se određuje ova granica korisno je znati raspon u kom se veličine u skupu kreću. Ne smatraju se jako informativnim jer se često događa da su najmanja i najveća vrijednost pokazatelja ekstremne u skupu u smislu da je susjedna najmanja odnosno najveća veličina poprilično udaljena od krajnjih. Zbog toga se ove krajnosti odbacuju kao pokazatelji a uzimaju se susjedne. Najvjerodostojniji pokazatelji pouzdanosti i granice povjerenja su upravo sredina i standardno odstupanje.

78. **Stratumi** – su naziv za uzorke u statistici čiji elementi, kada se poredaju po njihovoj vrijednosti, imaju tendenciju rasta i mogu se grupisati u četiri, pet ili više grupa i time dobiti grupe podataka koje također odražavaju tendenciju rasta.

79. **Istinita vrijednost** – je vrijednost izmjerena najpreciznijim instrumentom, za sve druge instrumente i pomagala.

80. **Uslovno istinita vrijednost** – je dužina određena najpreciznijim mogućim instrumentom, za sve druge instrumente ili pomagala za svako mjerenje izvršeno nekim drugim, manje preciznim, instrumentom ili pomagalom.

KLJUČNI POJMOVI 5:

81. **Apsolutna frekvencija** – predstavlja broj elemenata u svakoj pojedinoj grupi
82. **Relataivna frekvencija**- predstavlja broj elemenata u svakoj pojedinoj grupi
83. **Teoretska frekvencija** – opisuje kvalitet aritimetičke sredine i zajedno sa varijansom, standardnom devijacijom, koeficijentom varijacije i karakteristike datog skupa.
84. **Vjerovatnoća a priori** – je vjerovatnoća koju unaprijed možemo odrediti. To najbolje možemo vidjeti na primjeru novčića, on može samo pasti na grb ili broj. Tada je vjerovatnoća da padne na grb ½ jer se radi jednoj od ukupno dvije mogućnosti.
85. **Vjerovatnoća a postoreiori** – je vjerovatnoća koje se utvrđuje iz velikog broja mjerenja. Ako u posudi imamo nepoznat broj crnih i bijelih kuglica, nećemo unaprijed ništa tvrditi osim da postoje dvije vrste kuglica.
86. **Protivna vjerovatnoća** – je druga ili druge mogućnsti koje isključuju prvu mogućnost kod a priori vjerovatnoće. Na primjeru bacanja kocke postoji šest mogućnosti, a vjerovatnoća da iz jednog bacanja dobijemo broj 2 je 1/6, sve ostale mogućnosti predstavljaju protivnu vjerovatnoću.
87. **Diskretna varijabla** – je varijabla koja može primiti samo određene vrijednosti za varijablu X kao što je broj upisanih studenata(cio broj) jer nije moguce imati 2,3 studenta, broj soba u kuci sl.
88. **Kontinuirana varijabla** – je varijabla koja može primiti bilo koju vrijednost u posmatranom intervalu.
89. **Distribucija frekvencija** – sve frekvencije imaju distribuciju koja opisuje kvalitet aritimetičke sredidne.
90. **Binomna distribucija** – je polazište distribucije koje je odredio Bernoulli principom da svako nagađanje može da se razumije kako pogodak ili promašaj, što sve događaje svodi na svega dvije mogućnosti.
91. **Bernoulijeva distribucija** – je polazište distribucije koje je odredio Bernoulli principom da svako nagađanje može da se razumije kako pogodak ili promašaj, što sve događaje svodi na svega dvije mogućnosti.
92. **Normalna distribucija** – Najćešće primjenjiva u statistici. Radi se o neprekidnoj funkciji koja kao i prekidne može da se izrazi u dvije varijante. Kao funkcija gustine funkcija kumulativne vjerovatnoće. Gaus je ovu tehniku prvi primjenio pa se jos zove Gaussova distribucija.
93. **Gaussova distribucija** – Najćešće primjenjiva u statistici. Radi se o neprekidnoj funkciji koja kao i prekidne može da se izrazi u dvije varijante. Kao funkcija gustine funkcija kumulativne vjerovatnoće. Gaus je ovu tehniku prvi primjenio pa se jos zove Gaussova distribucija.
94. **Standardizovana distribucija** – predstavlja normalnu – gausovu distribuciju koja se može standardizirati tako što se obilježje X linearno transformiše u X=μ+z*δ.
95. **Standardizovano obiljezje** – je standardizovano odstupanje datog skupa od njegove sredine.
96. **Simetrična distribucija** – predstavlja binomnu distribuciju kad je vjerovatnoća nekog događaja ista kao protivna vjerovatnoća.
97. **Asimetrična distribucija** – predstavlja binomnu distribuciju kad je vjerovatnoća nekog događaja različita od protivne vjerovatnoće.
98. **Gustina vjerovatnoće** – pomuću nje određene su mase vjerovatnoće koje slučajna varijabla X može primiti pri različitom broju pokušaja.
99. **Kumulativ vjerovatnoće** – predstavlja zbir vjerovatnoća pi pri čemu Xi može primiti sve vrijednosti i€[0,n]
100. **Pascalov trougao** – je trougao koji formiraju krajnosti dviju vjerovatnoća i simetrične su prema sredini
101. **Granica tolerancije** –
102. **Deformacija uzorka** – slučaj kad ima mnogo veću standardnu devijaciju u odnosu na populaciju iz koje je uzet
103. **BINOMDIST** – ova funkcija ima zadatak da računa binomnu vjerovatnoću distribucije u problemima s fiksnim brojem procjena ili testova, u uslovima kad su procjene samo tačne ili pogrešne, kad su nepristrasne i kad je vjerovatnoća uspjeha konstanta
104. **NEGBINOMDIST** – računa negativnu binomnu distribuciju, tj. vjerovatnoću da da će doći do određenog broja promašaja prije nego se postigne pogodatk
105. **POISSON** – Excel funkcija da bi se predvidio broj nekih događaja tokom zadanog vremena
106. **LOGINV** – ova funkcija izračunava inverziju lognormalne funkcije distribucije od x u kojoj ln(x) normalno distribuirano sa sredinom iz parametara i standardnom devijacijom.
107. **LOGNORMDIST** – izračunava kumulativnu logaritmsku normalnu distribuciju od x u čemu je ln(x) normalno distribuirano s parametarskom sredinom i standardnom devijacijom
108. **NORMSDIST** – izračunava standardizovanu normalnu kumulativnu funkciju distribucije
109. **NORMINV** – izračunava se inverzija norlmane kumulativne distribucije za poznatu sredinu i standardnu devijaciju
110. **NORMDIST** – izračunava se gustina i kumulativ funkcije za poznatu sredinu i standadnu devijaciju
111. **NORMSINV** – izračunava inverziju standardizovane normalne kumulativne distribucije, čija aritmetička sredina je 0, a standardna devijacija 1.

KLJUČNI POJMOVI 6:

112. **Slučajnost elementa** – predstavlja slučajan izbor elemenata iz uzorka kojima možemo vjerovati
113. **Slučajnost skupa** – treba da odgovara normalnoj teoretskoj distribuciji i nuzno je Raspologat nekim pokazateljem o tome da li je i u kojoj mjeri deformisan
115. **Kritična vrijednost** – je veličina koja razgraničava područje prihvatiljivosti od područja neprihvatljivosti
116. **Jednostrani interval** – je interval povjerenja u kom je granica povjerenja određena samo s jedne strane u smislu da sredina uzorka ne može biti manja od neke veličine ili da ne može biti veća od neke veličine
118. **Relataivna frekvencija**- predstavlja broj elemenata u svakoj pojedinoj grupi
119. **Empirijska frekvencija** – je frekvencija kod je koje je aritimetička sredina različita od nule, a varijansa različita od jedinice
121. **Broj stepeni slobode** – da bi postojala bilo kakva sredina nužno je imati najmanje dvije varijable u skupu ali to nije

dovoljno. Sav višak iznad tog nužnoj broja opažanja ili varijabli nazivamo brojem stepeni slobode

122. Signifikantnost varijable – je vjerovatnoća da je X^2 veće od neke konstante α, a α jednaka je toj konstanti

127. Hi-kvadrat distribucija – je distribucija slučajinih varijabli koje se sastoje od sume kvadrata drugih slučajinh varijabli, Xi moraju biti takvi d aim je aritimetička sredina jednaka nuli, a varijansa jednaka jedinic

128. T distribucija – služi za ocjenu repreznetativnosti nekog uzorka u odnosu na populaciju koju predstavlja. Može se koristiti kad treba donijeti odluku o aritimetičkoj sredini populacije ako je njena standardna devijacija nepoznata, također podesna u poređenju dvaju uzoraka od kojih jedan služi kao kontrolni

129. F distribucija – osnovni zadatak ove distribucije je da ustanovi razlike između dviju distribucija korištenjem omjera među varijansama.

130. Asmietričnost – predstavlja osobinu skupa kad njegove frekvencije nisu simetrično raspoređene oko njegove sredine.

131. Spljoštenost- predstavlja osobinu skupa koja podrazumijeva široku disperziju Varijabli oko aritimetičke sredineI visokoj koncentraciji varijabli oko sredine

132. Bowleyev model – kao mjera simetrije koriste se kvaritili poređeni s medijanom. Polazi se od činjenice da je simetričnoj distribuciji gornji kvartil udalje on medijane onoliko koliko je medijana udaljena od donjeg kvartila

133. Pearsonov model – kao mjera simetrije koju je iskoristio Karl Pearson kao omjer između odstupanja modusa od arit. sredine na jednoj strani i standarne devijacije na drugoj

135. CONFIDENCE – služi za izračunavanje granica povjerenja. Zadatak ove funkcije je da izračunava interval povjerenja u prosjek uzorka, to je raspon sa obje strane sredine u uzorku

135. FDIST – izračunava vjerovatnoću F distribucije. Služi za utvrđivanje da li dva skupa podataka imaju različit stepen rasturanja

136. FINV- izračunava inverziju F distribucije. Koristi se za izračunavanje za utvrđivanje kritične vjerovatnoće iz F distribucije

137. CHIDIST – izračunava jednostranu vjerovatnoću Hi-kvadrat distribucije

138. CHIINV – je inverzija funkcije CHIDIST s njom se dobija kritična vrijednost Hi-kvadrat distribucije

139. CHITEST – vrši testiranje nezavisnosti. Izračunava kritičnu vrijednost iz Hi-kvadrat distribucije za statistički uzorak i odabrani broj stepeni slobode

140. TDIST – služi za izračunavanje procentualne vjerovatnoće Studentove t distribucije u kojoj se numerička veličina x uzima kao kalkulativna veličina od t Koristi se umjesto tablica kritičnih veličina u t-distribuciji

141. TINV – izračunava se kritična vrijednost studentove t distribucije, kao funkcija vjerovatnoće i broja stepeni slobode

142. TTEST – ima zadatak da Studentovim t-testom provjeri da li dva uzorka dolaze iz iz istih populacija koje imaju jednaku sredinu

143. SKEW – služi za mjerenje asimetrije negog skupa oko njegove sredine

144. KURT – izračunava spljoštenost distribucije u odnosu na normalnu distribuciju

KLJUČNI POJMOVI 7:

145. Matematička osnova regresije – za izražavanje međusavisnosti dviju ili više varijabli koristi se matematička aparatura

147. Matematičko očekivanje regresije – linearna regresija počiva na matematičkoj linearnoj jednačini

148. Područje definicije regresije – regresije služe za statističku analizu međusavisnosti dviju ili više varijabli

148. Regresija uzorka – regresiju uzorka računamo prema formuli Y = a+b*X

149. Regresija populacije – regresiju populacije računamo prema formuli Y = α+β*X

150. Ekstremi funkcije – linerna regresija počiva na matematičkoj linearnoj jednačini, tako da u ovom slučaju zavisne i nezavisna varijabla predstavlja ekstreme funkcije

151. Parametri – Obzirom da je prosta linearna regresija data izrazom Y=a+b*X, pri čemu su a i b parametri, pri čemu nam b opisuje za koliko će se promijeniti zavisna varijabla ako se nezavisna promjeni za 1 ili za koliko će se procenata promjeniti nezavisna varijabla ako se zavisna promjeni za 1%

152. Varijansa i totalna varijansa – nam pokazuje ocjenu tačnosti regresije i računa se po formuli $δ^2=Σ(Yi'-Yi)^2$ / v

153. Objašnjena komponenta varijanse – varijasna se može razložiti na objašnjenu i neobjašnjenu komponentu, računa se po formuli $δ^2=ΣΔYi''^2$ / v

154. Neobjašnjena komponenta varijanse – varijasna se može razložiti na objašnjenu i neobjašnjenu komponentu, računa se po formuli $δ^2=ΣΔYi'^2$ / v

155. Standardna devijacija – regresije računa se po formuli $δ=\sqrt{}Σ(Yi'-Yi)^2$ / v

156. Koeficijent varijacije – računamo po formuli r = δ / y'

157. Standardizovana greška – Ako je parameter b pribiližno normalmo distribuiran tada je vjerovatnoća odstupanja parametra b odneke uslovno istinite vrijednosti normalno distribuirana sa nekom standardnom greskom projcene.

158. Koeficijent korelacije – služi nam sa ocjenu regresije. Utvrđujemo korelaciju između varijabli, kreće se u rasponu od -1 do 1

159. Kovarijansa – nam služi za ocijenu regresije i predstavlja zajedničko odstupanje varijabli x i y

160. Koeficijent determinacije – može se tumačiti kao mjera tačnosti regresije jer se njim definiše stepen objašnjenosti odstupanja u regresiji. On nije neposredni pokazatelj tačnosti nego pokazatelj mjere u kojoj se jedna varijabla objašnjena regresijom i aritimetickom sredinom uzetim zajedno

164. Testiranje hipoteze – Hipotezu za regresiju u kojoj varijabla X predstvlja nezavisnu a varijabla Y zavisnu možemo testirati pomoću koeficijenta smjera β populacije koju regresija predstavlja. Koeficijenti β i b su koeficijenti smjera i možemo postaviti hipotezu da varijabla Y ne zavisi od varijable X time što ćemo za β reći da je jednak 0 I u tom slčaju Y je jednak nekoj konstanti i ne mijenja se bez obzira koju vrijednost prima X

165. INTERCPET – izračunava tačku u kojoj neka linija siječe y-osu koristeći postojeće poznate veličine x i y

166. SLOPE – izračunava nagib linearne regresije iz poznatih podataka x i y. To je vertikalni razmak podijeljen s horizontalnim između bilo kojih dviju tačaka linije

171. FORECAST – izračunava ili predviđa buduću vrijednost varijable y za datu varijablu x. Poznate veličine su y i x iz regresije, a tražena veličina y dobija se korištenjem linearne regresije

172. **STEYX** – izračunava standardnu grešku neke predviđene vrijednosti y za svaki x u regresiji. Služi kao mjera veličine greške u predviđanjima y za svako pojedinačno x

173. **PEARSON** – izračunava korelacioni koeficijent product momenta, r indeks bez dimenzija u rasponu od -1 do 1 uključuuići njegove granice, a odražava širinu linearnog odnosa između podataka dva skupa

174. **RSQ** – izračunava kvadrat Pearsonovog koeficijenta korelacije

175. **FTEST** – izračunava rezultat F-testa ili jednostavne vjerovatnoće da varijanse u seriji 1 i seriji 2 nisu jako različite. Koristi se za provjeru da li se ispitne ocjene privatnih škola imaju veće rasturanje od javnih ili obratno

176. **TTEST** – izračunava empirijsku vjerovatnoću prema Studentovom t-testu. Služi za provjeru da li dva uzorka dolaze iz istih populacija koje imaju jednaku sredinu

KLJUČNI POJMOVI 7.2:

177. **Višestruka regresija** – uključuje više od jednog faktora u analizu što predstavlja bolje sredstvo za analizu pojava

178. **Dimenzije regresije** – su dimenzije geometrijskog prostora u kom možemo prikazati regresiju. Prosta linearna regresija ima jedan factor a dvije varijable koje se mogu prikazati u ravni a nju nazivamo dvodimenzionalnim prostorom, dvofaktorsku regresiju prikazujemo u trodimenzionalnom prostoru jer u njoj postoje tri varijable

179. **Faktori regresije** – faktorima nazivamo nezavisne varijable regresije. Radi se o faktorima koji utiču na ponasanje zavisne varijable.

181. **Izokvanta** – predstavlja liniju koja spaja jednake veličine outputa

182. **Marginalna stopa supstitucije** – predstavlja koeficijent kojim se dva faktora X i Y mogu supstituirati jedan drugim

183. **Marginalni prinosi faktora** – predstavlja koeficijent b koji opisuje porast zavisne varijable(z), za (Δz) ako nezavisna varijabla x poraste za neko (Δx)

184. **Multiplikatori prinosa** – veličina (Δz) dobijena iz gornjih jednakosti

186. **F-test** – služi za provjeru regresije u cjelini, predstavlja poređenje empirijske vrijednosti sa nekom kritičnom vrijednošću

187. **F-vrijednost regresije** – pokazuje da li je uočeni odnos između zavisne i nezavisnih varijabli pojavljuje slučajno

189. **Kritična t-vrijednost** – označva vrijednost koju t pokazatelj mora na nivou određeneog broja stepeni slobode i signifikasnosti nadmašiti da bi se odbacila hipoteza da je vrijednost posmatranog parametra jednaka nuli

189. **LINEST** – izračunava paramaetre za liniju koja metodom najmanjih kvadrata aproksimira empirijske podatke. Funkcija računa više veličina, pa je istu potrebno unijeti kao niz

190. **Koeficijent determinacije** – njime se upoređuju empirijske s procjenjenim veličinama u regresiji. Koeficijent može imati vrijednost između 0(regresija nije podesna za predviđanje vrijednosti y) i 1(savršena aproksimacija)

191. **Rezidualna suma kvadrata** – predstavlja kvadrat razlike u regresionoj analizi između stvarnih i procjenjenih veličina y

195. **Suma kvadrata regresije** – predstavlja sumu kvadrata između stvarnih veličina y i njihovog prosjeka koja je nazvana totalnom sumom kvadrata

KLJUČNI POJMOVI 7.3:

199. **Trend – pojam** – predstavlja oblik regresije u kojima se kao nezavisna varijabla uzima vrijeme, a zavisna varijabla može biti bilo koja druga odabrana varijabla. U pitanju je prosta regresija u kojima egzistiraju dvije varijable a vrijeme se uvijek uzima kao nezavisna. Također pokazuju projekciju posmatrane pojave na neki budući datum

200. **Trend – linearni** – počiva na seriji tačaka Mi od kojih je svaka definisana parom empirijskih veličina za koje predpostavljamo da uz neku grešku postoji linearna jednkost $Y = a+b*x$

201. **Trend – nelinearni**

202. **Trend – polinomni** – počiva ne seriji tačaka Mi od kojih je svaka definisana parom empirijskih veličina za koje predpostavljamo uz neku grešku veza između varijabli oblik polinoma $Y = a+b*x+c*x^2 +d*x^3$

203. **Trend – potencijalni** – je trend u kojem je zavisna varijabla dignuta na neku potenciju. U tom slučaju jednačina se najprije logaritmira da bi dobila linearnu formu iz koje se dobivaju jednačine odstupanja. Formula za računanje $Y= Y_0 * x^a$

204. **Trend – eksponencijani** – se računa po formuli $Y= Y_0 * e^{b*x}$

205. **Trend – veza s regresijom** – ako raspolažemo trendovima nekih varijabli za koje znamo da su međuzavisne možemo ih pretvoriti u odgovarajuću regresiju i obratno. pri tome su parametri regresije određeni parametrima trendova iz kojih je izvedena regresija.

209. **TREND** – izračunava vrijednost duž linearnog trenda koristeći se metodom najmanjih kvadrata a na osnovu poznatih y i x.

210. **GROWTH** – ima zadatak da procijeni rast nezavisne varijable za neke zadane veličine zavisne varijable.

211. **Normalne jednačine**

212. **MDETERM** – izračunava detereminantu matrice u rješavanju sistema matematičkih Jednačina s više varijabli koristeći se Kramerovim pravilom. Uslov za Izračunavanje je das u broj kolona i redova jednaki

213. **MINVERSE** – izračunava inverziju zadane matrice. Koristi se u rješavanju jednačina s više varijabli

KLJUČNI POJMOVI 8:

214. **Looreznova kriva** – daje nam jasniju sliku i omogućava analizu tokom vremena u prostoru o dohotku i socijalnom statusu stanovništva

215. **Ginijev koeficijent** – Corrado Ginni je oslanjajući se na Lorenzovu krivu Formulisao koeficijent koncentracije bogatstva

216. **Kuznetsov indeks nejednakosti** – Simon Kuznets je prvi izvršio poređenja dohotka različitih socijalnih grupa. On je formulisao raspon odstupanja i indeks nejednakosti u raspodjeli. Kuznetsov indeks nejednakosti predstavlja omjer između dohotka najbogatijih i najsiromašnijih 10% ili 20% stanovnika

217. **Kuznetsova inverzna kriva** – Ako posmatramo Kuznetsov koeficijent određen omjerom mase društvenog proizvoda koja pripada sloju od 20% najbogatijih stanovnika prema društvenom proizvodu donjih 40% najsimormasnijih na panelu zemalja na raznim nivoima razvijenosti dobija se kriva koja liči na izvrnuto slovo U.

Statistic	Description
se1,se2,...,sen	The standard error values for the coefficients m1,m2,...,mn.
seb	The standard error value for the constant b (seb = #N/A when *const* is FALSE).
r2	The coefficient of determination. Compares estimated and actual y-values, and ranges in value from 0 to 1. If it is 1, there is a perfect correlation in the sample — there is no difference between the estimated y-value and the actual y-value. At the other extreme, if the coefficient of determination is 0, the regression equation is not helpful in predicting a y-value. For information about how r2 is calculated, see "Remarks," later in this topic.
sey	The standard error for the y estimate.
F	The F statistic, or the F-observed value. Use the F statistic to determine whether the observed relationship between the dependent and independent variables occurs by chance.
df	The degrees of freedom. Use the degrees of freedom to help you find F-critical values in a statistical table. Compare the values you find in the table to the F statistic returned by **LINEST** to determine a confidence level for the model. For information about how df is calculated, see "Remarks," later in this topic. Example 4 shows use of F and df.
ssreg	The regression sum of squares.
ssresid	The residual sum of squares. For information about how ssreg and ssresid are calculated, see "Remarks," later in this topic.

The following illustration shows the order in which the additional regression statistics are returned.

	A	B	C	D	E	F
1	m_n	m_{n-1}	...	m_2	m_1	b
2	se_n	se_{n-1}	...	se_2	se_1	se_b
3	r_2	se_y				
4	F	d_f				
5	ss_{reg}	ss_{resid}				

Asimetrija skupa

Koeficijent α_3. Preciznija mjera asimetričnosti je jednačina trećeg momenta oko proste aritmetičke sredine, podijeljena trećom potencijom standardne devijacije:

$$6.4.d):\quad \alpha_3 = \frac{m_{s3}}{\sigma^3} = \frac{\sum (x_i-\mu)^3}{n\cdot\sigma^3}$$

a oko ponderisane sredine je:

$$6.4.e):\quad \alpha_3 = \frac{m_{s3}}{\sigma^3} = \frac{\sum f_i\cdot(x_i-\mu)^3}{n\cdot\sigma^3}$$

Mjerilo spljoštenosti skupa

Opšte mjerilo spljoštenosti

Koeficijent α_4 (**zove se i β_2**) a definisan je:

a) Kvocijentom između četvrtog momenta oko sredine (m_{s4}) i četvrte potencije standardne devijacije ili
b) Omjerom četvrtog momenta oko sredine i kvadrata drugog momenta oko sredine tj:

$$6.5.a):\quad \alpha_4 = \beta_2 = \frac{m_{s4}}{\sigma^4} = \frac{m_{s4}}{m_{s2}^{\ 2}}.$$

Kod neuređenog skupa (proste aritmetičke sredine) on glasi:

$$6.5.b):\quad \alpha_4 = \frac{\sum(\mu-x_i)^4}{n\cdot\sigma^4}$$

a kod frekvencijskog skupa (ponderisane sredine):

$$6.5.c):\quad \alpha_4 = \frac{\sum f_i\cdot(\mu-x_i)^4}{n\cdot\sigma^4}.$$

Standardizacija normalne distribucije

Funkcija 5.10.a) normalne distribucije može se **standardizovati** tako što uvedemo smjenu:

$$5.10.b):\ z = \frac{x-\mu}{\sigma}$$

Korelacija varijabli x i y u regresiji

Pokazatelji:

1. **Koeficijent korelacije** – međusobni položaj varijabli ili korelacija,
2. **Kovarijansa -** mjera međusobnog odstupanja od odgovarajućih sredina,
3. **Koeficijent determinacije** - mjera objašnjenosti regresije.

Kovarijansa (C): zajedničko odstupanje varijabli x i y. Dva načina utvrđivanja:

$$C_{xy} = \frac{\sum(x_i-\mu_x)\cdot(y_i-\mu_y)}{n-1};\qquad ili\qquad C_{xy} = \frac{\sum x\cdot y}{n-1} - \mu_x\cdot\mu_y.$$

Korelacija varijabli x i y u regresiji

Koeficijent korelacije: može se kretati u rasponu od -1 do 1, tj.: **-1** $\le \rho \le$ **1**.
Ako dvije varijable, koje koristimo u regresiji imaju:
ρ=1, varijable su pozitivne i međusobno potpuno koreliraju tako da nema neobjašnjenog dijela odstupanja i standardna devijacija regresije je jednaka nuli, $\sigma_r=0$. Kad je **0<ρ<1** postoji pozitivna korelacija između varijabli, $\sigma_r>0$ i ako $\sigma_r\Rightarrow\infty$, tada $\rho\Rightarrow0$.
ρ=-1, jedna od varijabli je negativna a druga pozitivna, i potpuno koreliraju jedna s drugom, pa je $\sigma_r=0$. Kad je **-1<r<0**, postoji negativna korelacija, a $\sigma_r>0$ i ako $\sigma_r\Rightarrow\infty$, tada $\rho\Rightarrow0$. Primjer: odnos između negativnih temperatura i potrošnje energije za grijanje.
ρ=0, varijable nemaju nikakvu linearnu vezu i ne objašnjavaju jedna drugu.

Koeficijent korelacije i parametar b (nagib regresije) - postoji i odnos:

$$b=\rho\cdot\frac{\sigma_y}{\sigma_x};\quad ili\quad \rho=b\cdot\frac{\sigma_x}{\sigma_y},$$

u čemu su:

$$\sigma_x=\sqrt{\frac{\sum\Delta x_i^2}{n-1}}=\sqrt{\frac{(x_i-\mu_x)^2}{n-1}};\qquad \sigma_y=\sqrt{\frac{\sum\Delta y_i^2}{n-1}}=\sqrt{\frac{(y_i-\mu_y)^2}{n-1}}$$

Napomena: Koordinatne razlike Δx_i, Δy_i odnose na odstupanja empirijskih vrijednosti varijabli x, y od njihove aritmetičke sredine (**μ_x, μ_y**), a ne od regresije (y'), koju obično označavamo sa $\Delta x_i'$ odnosno $\Delta y_i'$.

Pokazatelji tačnosti regresije

Pokazatelji: varijansa, standardna devijacija, koeficijent varijacije i standardizovana greška

Varijansa regresije (σ^2): $\boxed{\sigma^2=\frac{\sum(y_i'-y_i)^2}{\nu}=\frac{\sum(\Delta y_i')^2}{\nu}}$

u čemu su: $y_i'-y_i=\Delta y'$= razlike između regresionih (y_i') i empirijskih vrijednosti (y_i) varijable Y, i
ν= broj stepeni slobode regresije (u konkretnom slučaju n-2).

Neki udžbenici: umjesto ν u nazivniku koriste n ili n-1 (neispravno)
Varijansa (σ^2) se naziva i ***totalnom varijansom*** a može razložiti na ***objašnjenu*** (σ_o^2) i ***neobjašnjenu*** komponentu (σ_n^2), analogno koeficijentu determinacije:

$\boxed{\sigma^2=\sigma_o^2+\sigma_n^2}$ u čemu su: $\boxed{\sigma^2=\frac{\sum\Delta y_i^2}{\nu};\ \sigma_o^2=\frac{\sum\Delta y''^2_i}{\nu};\ i\ \sigma_n^2=\frac{\sum\Delta y'^2_1}{\nu}.}$

Standardna devijacija (σ): $\boxed{\sigma=\sqrt{\frac{\sum(y'-y_i)^2}{\nu}}=\sqrt{\frac{\sum(\Delta y')^2}{\nu}}}$

Koeficijent varijacije (ρ): $\boxed{\rho=\sigma/y'}$

Standardizovana greška regresije (se$_r$): $\boxed{se_r=\sigma/\sum(\Delta y')^2}$

631. Područje povjerenja, tablice i funkcije iz Excela

	Slučaj	**Šta se ispituje?**	ΣP(z)= 95%; α_{rep}=0,05; n=10,	
Dvostrani interval povjerenja	Sredina i varijansa populacije poznati	**Raspon oko sredine uzorka** $\mu_p=\mu_u\pm z_{rep}\cdot\frac{\sigma}{\sqrt{n}}$	**Z Tablice** P(z) 0,025 $z_{0,025}$ 1,960	**Excel** 0,975 =normsdist(E5) 1,960 =normsinv(0,975) -1,960 =normsinv(0,025)
	Sredina i varijansa populacije nepoznati	**Raspon oko sredine uzorka** $\mu_p=\mu_u\pm t_{rep}\cdot\frac{\sigma}{\sqrt{n}}$	ΣP(t) = 95%; α_{rep}=0,05; ν=9 **t-tablice** P(t) 0,025 $t_{0,025}$ 2,262	**Excel** 0,025006 =tdist(E10;9;1) 2,262 =tinv(2*E9;9)
	2 nezavisna uzorka, 2 populacije poznate	**Raspon oko sredine uzorka** $\Delta\mu_p=\Delta\mu_u\pm z_{rep}\cdot\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	ΣP(z)=95%; α_{rep}=0,05; n_1; n_2; ν=9 **Z Tablice** P(z) 0,025 $z_{0,025}$ 1,96	**Excel** 0,975 =normsdist(E14) 1,959964 = - normsinv(E13)
	2 nezavisna uzorka, 2 populacije nepoznate	**Raspon oko sredine uzorka** $\Delta\mu_p=\Delta\mu_u\pm t_{rep}\cdot\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	ΣP(t)=95%; α_{rep}=0,05; ν=(n_1-1)+n_2-1) =9 **t-tablice** P(t) 0,025 $t_{0,025}$ 2,262	**Excel** 0,025006 =tdist(E18;9;1) 2,262 =tinv(2*E17;9)
	Spareni uzorak	**Raspon oko sredine uzorka** $\Delta\mu=D\pm t_{rep}\cdot\frac{\sigma_\mu}{\sqrt{n}}$	ΣP(t)=95%; α_{rep}=0,05; ν=n-1=9 **t-tablice** P(t) 0,025 $t_{0,025}$ 2,262	**Excel** 0,025006 =tdist(E22;9;1) 2,262 =tinv(2*E21;9)
Jednostrani interval povjerenja	Sredina i varijansa populacije poznati	**Donja granica povjerenja** $\mu_p>\mu_u-z_{rep}\cdot\frac{\sigma}{\sqrt{n}}$	ΣP(z)= 95%; α_{rep}=0,05; n=10. **Z Tablice** P(z) 0,05 $z_{0,05}$ 1,645	**Excel** 0,950 =normsdist(E26) 1,645 =normsinv(0,95) -1,645 =normsinv(0,05)
	Sredina i varijansa populacije nepoznati	**Donja granica povjerenja** $\mu_p>\mu_u-t_{rep}\cdot\frac{\sigma}{\sqrt{n}}$	ΣP(t) = 95%; α_{rep}=0,05; ν=9 **t-tablice** P(t) 0,05 $t_{0,05}$ 1,833	**Excel** 0,050009 =tdist(E31;9;1) 1,833 =tinv(2*E30;9)
	2 nezavisna uzorka, 2 populacije poznate	**Gornja granica povjerenja** $\Delta\mu_p<\Delta\mu_u+z_{rep}\cdot\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	ΣP(z)=95%; α_{rep}=0,05; n_1; n_2; ν=9 **Z Tablice** P(z) 0,05 $z_{0,05}$ 1,645	**Excel** 0,950 =normsdist(E35) 1,644854 = - normsinv(E34)
	2 nezavisna uzorka, 2 populacije nepoznate	**Gornja granica povjerenja** $\Delta\mu_p<\Delta\mu_u+t_{rep}\cdot\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	ΣP(t)=95%; α_{rep}=0,05; ν=(n_1-1)+n_2-1) =9 **t-tablice** P(t) 0,05 $t_{0,05}$ 1,833	**Excel** 0,050009 =tdist(E39;9;1) 1,833 =tinv(2*E38;9)
	Spareni uzorak	**Gornja granica povjerenja** $\Delta\mu<D+t_{rep}\cdot\frac{\sigma_D}{\sqrt{n}}$	ΣP(t)=95%; α_{rep}=0,05; ν=n-1=9 **t-tablice** P(t) 0,05 $t_{0,05}$ 1,833	**Excel** 0,050009 =tdist(E43;9;1) 1,833 =tinv(2*E42;9)

Funkcije za Studentovu t-analizu distribucije uzorka								
Varijable uzorka (x_i)	**5**	**8**	**10**	**2**	**6**	**4**	**3**	$t=\dfrac{x\sqrt{\nu}}{\sqrt{\chi^2}}$
Aritmetička sredina:	5,43		Varijansa:	7,95		Stand.dev.	2,82	
Isti uzorak, z varijable	**-0,152**	**0,912**	**1,621**	**-1,216**	**0,203**	**-0,507**	**-0,861**	
Aritmetička sredina	0,00		Varijansa:	1,00		Stand.dev.	1,00	

t-funkcije u Excelu		
TTEST, jednostrani, sparen	0,000109726	Vjerovatnoća da uzorci iz redova 2 i 4 dolaze iz iste populacije.
TTEST, jednostrani, homosked.	0,00021667	Vjerovatnoća da uzorci iz redova 2 i 4 dolaze iz iste populacije.
TTEST, jednostrani, heterosked.	0,000815809	Vjerovatnoća da uzorci iz redova 2 i 4 dolaze iz iste populacije.
TTEST, dvostrani, sparen	0,000219451	Vjerovatnoća da uzorci iz redova 2 i 4 dolaze iz iste populacije.
TTEST, dvostrani, homoskedast.	0,00043334	Vjerovatnoća da uzorci iz redova 2 i 4 dolaze iz iste populacije.
TTEST, dvostrani, heterosked.	0,001631619	Vjerovatnoća da uzorci iz redova 2 i 4 dolaze iz iste populacije.

TDIST, jednostrano p(t), na x_i:	0,001226171	0,0001	2,89599E-05	0,04621	0,00048	0,00356	0,012	Masa t-vjerovatnoće
TDIST, dvostrano p(t), na x_i:	0,002452342	0,0002	5,79198E-05	0,09243	0,00096	0,00712	0,02401	Masa t-vjerovatnoće

Područje povjerenja	
1. Iz Excela	
TINV, $t_{0,95}$	0,065374001
TINV, $t_{0,05}$	2,446911846
TINV, $t_{0,025}$	2,968686684

2. Iz tablica	$t_{0,95}$	$t_{0,05}$	$t_{0,025}$
Vrijednost t_{krit}:		1,943	2,447
Stdev uzorka		2,82	2,82
Broj članova (n)		7	7
$Co=t_{krit}*\sigma/(n^{1/2})=$		2,07096	**2,60816**

$$\Delta x_{krit} = t_{krit}\frac{\sigma_u}{\sqrt{n}}$$

$$x_{dozv} \in \left[x_{\min}; x_{\max}\right]:$$

$$x_{dozv} = \mu \pm \Delta x_{krit}$$

95%-tna "t" vjerovatnoća za interval:	x_{min}=	**2,82**	μ=	5,43	x_{max}=	**8,04**

Drugo tumačenje: Područje pouzdanosti je raspon unutar kojeg se nalazi nepoznata sredina populacije kojoj uzorak pripada, pri odabranoj vjerovatnoći.

Zaključci o t-funkcijama Excela:

1. TTEST funkcija daje direktan odgovor na pitanje da li par uzoraka dolazi iz iste populacije, pri čemu treba specificirati traži li se jednostrana ili dvostrana vjerovatnoća, te da li je u pitanju spareni ili nezavisni uzorak, odnosno o jednakim ili različitim aritmetičkim sredinama.
2. TDIST daje masu ili gustinu vjerovatnoće na nekoj zadanoj varijabli uzorka x ili t_{rep}. Kritičnom α odgovara dvostrani rep distribucije.
3. TINV daje varijablu t_{krit} na nekoj zadanoj vjerovatnoći $\boldsymbol{\alpha}_{rep}$.

Dvostruki oprez:

a) U primjeni koristi $\boldsymbol{\alpha}_{rep}$ pa ga treba razlikovati od NORMSINV, koja koristi (1-α)

b) Za razliku od tablica u kojima se kritično alfa obično polovi, Excel ga koristi cijelo za izračunavanje t_{rep}.

632.TESTIRANJE HIPOTEZA - KLASIČAN TEST

1. Uzorak i populacija

z- i t-vrijednosti uzorka za odabrano **α**:

$$\mu_u-\mu_{hip}=z_{uz}\cdot\frac{\sigma}{\sqrt{n}};\ \Rightarrow\ z_{uz}=\frac{\left(\mu_u-\mu_{hip}\right)\cdot\sqrt{n}}{\sigma_p};$$

$$\mu_u-\mu_{hip}=t_{uz}\cdot\frac{\sigma}{\sqrt{n}};\ \Rightarrow\ t_{uz}=\frac{\left(\mu_u-\mu_{hip}\right)\cdot\sqrt{n}}{\sigma_u}$$

Testiranje pomoću tablica (Excel - analogno intervalu povjerenja)

Postupak testiranja:

1. Izračunamo z- ili t-vrijednost uzorka. Za dobijeno z_{uz} ili t_{uz} tražimo $P(z_{uz})$ ili $P(t_{uz})$.

a) Za H_0: $\mu_u = \mu_p$ (razlika je neznatna ⇒ jednostruka P-vrijednost):

2a). Dobijenu P-vrijednost poredimo s kritičnim **α**.

b) Za H_1: $\mu_u > \mu_p$ (razlika postoji i znatna je ⇒ dvostruka P-vrijednost):

2b). Dobijenu P-vrijednost množimo sa 2 i poredimo s kritičnim **α**.

	H_0: $\mu_u \approx \mu_p$		H_1: $\mu_u > \mu_p$		**Zaključak:**
H_0	Ztest	$z_{uz}>z_{teor}$	ili	$P(z_{uz})<P(z_{teor})$	H_0 se odbacuje u korist alternativne hipoteze, uz vjerovatnoću $P(z_{teor})$ da griješimo.
H_1	Ztest	$z_{uz}<z_{teor}$	ili	$P(z_{uz})>P(z_{teor})$	H_1 se odbacuje u korist H_0, uz vjerovatnoću $(1- P(z_{teor}))$.
H_0	t-test	$t_{uz}>t_{teor}$	ili	$P(t_{uz})<P(t_{teor})$	H_0 se odbacuje u korist alternativne hipoteze, uz vjerovatnoću $P(t_{teor})$ da griješimo.
H_1	t-test	$t_{uz}<t_{teor}$	ili	$P(t_{uz})>P(t_{teor})$	H_1 se odbacuje u korist H_0, uz vjerovatnoću $(1- P(z_{teor}))$.

2. Dva uzorka i dvije populacije

z- i t-vrijednosti uzorka za odabrano **α**:

$$\Delta\mu_u-\Delta\mu_{hip}=z_{uz}\cdot\sigma_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}};\ \Rightarrow\ z_{uz}=\frac{\left(\Delta\mu_u-\Delta\mu_{hip}\right)}{\sigma_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}};$$

$$\Delta\mu_u-\Delta\mu_{hip}=t_{uz}\cdot\sigma_u\sqrt{\frac{1}{n_1}+\frac{1}{n_2}};\ \Rightarrow\ t_{uz}=\frac{\left(\Delta\mu_u-\Delta\mu_{hip}\right)}{\sigma_u\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$$

Postupak testiranja:

1. Izračunamo z- ili t-vrijednost uzorka. Za dobijeno z_{uz} ili t_{uz} tražimo $P(z_{uz})$ ili $P(t_{uz})$.

a) Za H_0: $\Delta\mu_u = \Delta\mu_p$ (razlika je neznatna ⇒ jednostruka P-vrijednost):

2a). Dobijenu P-vrijednost poredimo s kritičnim **α**.

b) Za H_1: $\Delta\mu_u > \Delta\mu_p$ (razlika postoji i znatna je ⇒ dvostruka P-vrijednost):

2b). Dobijenu P-vrijednost množimo sa 2 i poredimo s kritičnim **α**.

				$\Delta\mu_p$	**Zaključak:**
H_0	Ztest	$z_{uz}>z_{teor}$	ili	$P(z_{uz})<P(z_{teor})$	H_0 se odbacuje u korist alternativne hipoteze, uz vjerovatnoću $P(z_{teor})$ da griješimo.
H_1	Ztest	$z_{uz}<z_{teor}$	ili	$P(z_{uz})>P(z_{teor})$	H_1 se odbacuje u korist H_0, uz vjerovatnoću $(1- P(z_{teor}))$.
H_0	t-test	$t_{uz}>t_{teor}$	ili	$P(t_{uz})<P(t_{teor})$	H_0 se odbacuje u korist alternativne hipoteze, uz vjerovatnoću $P(t_{teor})$ da griješimo.
H_1	t-test	$t_{uz}<t_{teor}$	ili	$P(t_{uz})>P(t_{teor})$	H_1 se odbacuje u korist H_0, uz vjerovatnoću $(1- P(z_{teor}))$.