

Univerza v Ljubljani

Filozofska fakulteta

Oddelek za bibliotekarstvo, informacijsko znanost in knjigarstvo

STATISTIKA - VRSTE OBDELAVE PODATKOV

Seminarska naloga pri predmetu

Uvod v znanstveno delo

Mentorica: dr. Vlasta Zabukovec

Avtorici: Katja Šuler, Manca Lončarič

Ljubljana, 2013

Kazalo vsebine

UVOD.....	2
KAJ JE STATISTIKA.....	2
OSNOVNI POJMI.....	2
PRIMER FREKVENC IN ODSOTKOV.....	3
MERE SREDNJIH VREDNOSTI.....	3
Aritmetična sredina.....	3
Mediana.....	4
Modus.....	4
PRIMERI SREDNJIH VREDNOSTI.....	4
Aritmetična sredina.....	4
Mediana.....	4
Modus.....	4
MERE VARIABILNOSTI oziroma RAZPRŠENOSTI.....	5
Variacijski razmik.....	5
Medčetrtnski razmik.....	5
Standardni odklon in varianca.....	5
PRIMERI MER VARIABILNOSTI.....	5
Medčetrtnski razmik, standardni odklon ter varianca.....	5
Variacijski razmik.....	6
KOEFICIENT KORELACIJE.....	6
PRIMERI KOEFICIENTA KORELACIJE.....	6
ZAKLJUČEK.....	6
VIRI.....	7

Kazalo tabel

UVOD

V seminarski nalogi bova predstavili statistično obdelavo podatkov, oziroma splošno opredelile statistiko, navedle njene osnovne pojme, predstavile mere srednjih vrednosti, mere razpršenosti ter koeficient korelacije. Namen seminarske naloge je, da vas čim bolj seznaniva s pojmom statistike in njene uporabe.

Ključne besede: statistika, aritmetična sredina, modus, mediana, standardni odklon

KAJ JE STATISTIKA

Statistika je znanost, ki razvija metode o zbiranju statističnih podatkov, njihovi analizi in predstavitvi. Statistika se uporablja kot znanstvena veda in disciplina v ekonomiji, medicini, psihologiji, šolstvu, na področju varnosti. Opravlja se jo tako za širši krog uporabnikov ali za svoje potrebe (Brvar, 2007).

Rezultat statistične analize oziroma proučevanja so posamezne statistične informacije ali bolj ali manj obsežna poročila z ugotovitvami, pojasnili,... V statistikah, namenjenih najširšemu krogu uporabnikov, so podatki največkrat prikazani v obliki tabel in grafikonov s pojasnili (Brvar, 2007).

»Bistven je cilj, namen statistike, to pa je dati številkam pomen« (Sirkin, 1995: 27).

OSNOVNI POJMI

Množico vseh elementov pojava, ki se proučuje s statističnimi metodami, imenujemo populacija ali statistična populacija. Populacije so na primer kazniva dejanja, prijavljena policiji, prebivalci, rojstva, smrti itd (Brvar, 2007).

Spremenljivka je lastnost statistične enote, ki lahko zavzame katerokoli vrednost iz določene množice možnih vrednosti. Spremenljivke enot iz prejšnjega odstavka so na primer vrsta, datum in kraj kaznivega dejanja (Brvar, 2007).

Vzorec je del populacije, ki mora biti po strukturi čim bolj podoben populaciji. Temu pravimo reprezentativnost vzorca. Vzorec dejansko analiziramo, populacijo pa le ocenjujemo. Raziskovanje pojavov na vzorcu ima veliko praktično vrednost, saj prihranimo veliko denarja in časa, pri zbiranju podatkov potrebujemo manj sodelavcev, hitreje obdelamo podatke ter ker velikokrat ne moremo zajeti v raziskavo vseh enot populacije, lahko pojav raziščemo samo na delu populacije. Po velikosti ločimo male in velike vzorce (Brvar, 2007).

Statistični parametri so številčne in kakovostne značilnosti populacije. Mednje sodi tudi numerus (N), ki predstavlja moč vzorca oziroma število enot v populaciji ali vzorcu. Med parametre sodijo tudi relativna števila (odstotki), mere srednjih vrednosti (aritmetična sredina, modus, mediana), mere razpršenosti (variacijski razmik, medčetrtnski razmik, standardni odklon in varianca), korelacijski koeficient (Brvar, 2007).

Frekvenca (f) ali pogostost je število enakih vrednosti v populaciji ali vzorcu. Absolutna frekvenca predstavlja število enot, ki imajo določeno vrednost, relativna frekvenca pa predstavlja odstotek enot, ki imajo neko določeno vrednost (Brvar, 2007).

PRIMER FREKVENC IN ODSOTOKOV

Tabela : Frekvenčna tabela spremenljivke klasifikacija prometne nesreče

Klasifikacija prometne nesreče	Frekvenca (f)	Strukturni delež (p %)
Brez poškodbe	20346	64,45
Lahka telesna poškodba	9940	31,49
Huda telesna poškodba	1050	3,33
Smrt	233	0,74
Skupaj	31569	100,01

Vir: Brvar, B. (2007). *Statistika*. Ljubljana: Fakulteta za varnostne vede, Univerza v Mariboru.

Frekvenca nam v tabeli pokaže število različnih klasifikacij prometne nesreče, strukturni delež pa nam pokaže odstotek teh klasifikacij.

MERE SREDNJIH VREDNOSTI

Poznamo več mer srednjih vrednosti, uporabljamo pa zlasti tri: aritmetično sredino, mediano in modus (Sagadin, 1982).

Aritmetična sredina

Aritmetično sredino imenujemo tudi povprečje. Označujemo jo s črko M (*angl. mean*: povprečje) ali tudi z "x prečno". Prednosti aritmetične sredine so, da je najbolj natančna, zanesljiva in občutljiva mera srednje vrednosti, saj pri njenem izračunavanju upoštevamo vrednosti čisto vseh rezultatov (Sagadin, 1982).

Mediana

Mediano ali središčnico označujemo s kratico Me (*lat. medianus*: središčen) in je srednja vrednost po položaju. Polovica enot ima večjo vrednost, polovica manjšo. Je točka. Prednost Me je v tem, da ekstremni rezultati nimajo vpliva na njeno vrednost, saj upoštevamo le vrednost v sredini ranžirne vrste. Njena pomanjkljivost pa je v tem, da je manj občutljiva in natančna kot aritmetična sredina, ker ne upošteva vrednosti vseh rezultatov ter precej odvisna od velikosti vzorca (Sagadin, 1982).

Modus

Modus ali gostiščnica je tista vrednost spremenljivke, okrog katere so vrednosti oziroma rezultati najbolj gosti. Modus je rezultat, ki se največkrat ponavlja. Prednost modusa je v tem, da ga lahko določamo tudi pri spremenljivkah, ki se pojavljajo v ločenih kategorijah. Pomanjkljivost modusa je v tem, da je najbolj groba ter najmanj zanesljiva in zato tudi najmanj uporabna mera srednje vrednosti (Sagadin, 1982).

PRIMERI SREDNJIH VREDNOSTI

Aritmetična sredina

Primer podatkov o doseženih točkah, ki so jih dosegli policisti na testu znanja o uporabi prisilnih sredstev. Rezultati so bili naslednji:

67, 45, 91, 56, 87, 33, 61, 72, 60, 49, 55, 29, 67, 44, 43, 44, 59, 66, 83, 72, 88, 95, 39, 54, 86, 77, 88, 74, 82, 73, 66, 42, 89, 95, 33, 61, 57, 67, 34, 49.

Povprečna vrednost doseženih točk na testu znaša 63,3 točke. Do te vrednosti smo prišli tako, da smo sešteli vseh 40 vrednosti in vsoto delili s 40 (Brvar, 2007).

Mediana

Vzamemo enak primer kot pri aritmetični sredini, le da moramo podatke urediti v urejeno vrsto po velikosti.

29, 33, 33, 34, 39, 42, 43, 44, 44, 45, 49, 49, 54, 55, 56, 57, 59, 60, 61, 61, 66, 66, 67, 67, 67, 72, 72, 73, 74, 77, 82, 83, 86, 87, 88, 88, 89, 91, 95, 95.

Mediana znaša 63,5 oz 64 točk, kar deli populacijo na dva dela - 50% policistov je doseglo manj ali enako 64 točk, 50% policistov pa več ali enako 64 točk (Brvar, 2007).

Modus

Tabela : Frekvenčna tabela spremenljivke razširjenost pojava korupcije in pojava sprejemanja podkupnine v javnih službah

Razširjenost pojava korupcije in pojava sprejemanja podkupnine v javnih službah	Frekvenca (f)	Strukturni delež (p%)
Ne vem	21	2,58
Nihče ne počne tega	347	42,58
To jih počne nekaj	370	45,40
To počnejo skoraj vsi	77	9,45
Skupaj	815	100,01

Vir: Brvar, B. (2007). *Statistika*. Ljubljana: Fakulteta za varnostne vede, Univerza v Mariboru.

Modus lahko takoj razberemo, saj je odgovor, ki ima največjo frekvenco. To je odgovor To jih počne nekaj (Brvar, 2007).

MERE VARIABILNOSTI oziroma RAZPRŠENOSTI

Poznamo več mer razpršenosti, ki nam povedo, koliko so rezultati razpršeni okrog njihove srednje vrednosti oziroma, koliko se individualni rezultati odklanjajo (variirajo) od srednje vrednosti populacije. Ene od teh so variacijski razmik, medčetrtnski razmik, standardni odklon in varianca (Kompore idr., 2011).

Variacijski razmik

Je najpogostejša mera razpršenosti, ki nam pove, kolikšen je razpon celotnega niza ali kolikšna je absolutna razlika med najnižjo in najvišjo vrednostjo v nizu (Kompore idr., 2011). Njegova prednost je v tem, da je preprost za izračunavanje, vendar pa je hkrati zelo groba, nestabilna in nezanesljiva mera, ker ga določata le ekstremni vrednosti (Kompore idr., 2011).

Medčetrtnski razmik

Je razlika med največjo in najmanjšo vrednostjo v naboru podatkov (Kompore idr., 2011).

Standardni odklon in varianca

Sta najbolj zanesljivi meri razpršenosti, ki upoštevata vse vrednosti v nizu. Standardni odklon označimo SD ali grško črko sigma (σ) in varianco z σ^2 (Sagadin, 1982). Standardni odklon nam pove, kolikšna so povprečna odstopanja posameznih rezultatov od aritmetične sredine (Sagadin, 1982).

PRIMERI MER VARIABILNOSTI

Medčetrtnski razmik, standardni odklon ter varianca

Primer podatkov o doseženih točkah, ki so jih dosegli policisti na testu znanja o uporabi prisilnih sredstev. Rezultati so bili naslednji:

67, 45, 91, 56, 87, 33, 61, 72, 60, 49, 55, 29, 67, 44, 43, 44, 59, 66, 83, 72, 88, 95, 39, 54, 86, 77, 88, 74, 82, 73, 66, 42, 89, 95, 33, 61, 57, 67, 34, 49.

Medčetrtnski razmik izračunamo po matematični formuli. Najprej izračunamo kvartila Q_1 in Q_3 . Kvartil Q_1 znaša 45, kvartil Q_3 pa 77. Medčetrtnski razmik je 77-45, kar znaša 32. Medčetrtnski razmik torej predstavlja 32 točk, ki so jih dosegli policisti (Brvar, 2007).

Izračunajmo varianco in standardni odklon za rezultate, ki so jih policisti dosegli na testu znanja o uporabi prisilnih sredstev. V prejšnjih primerih smo že izračunali razlike med posameznimi vrednostmi in aritmetično sredino. Te razlike kvadriramo in seštejemo. Varianca po tem izračunu znaša 354,51 in predstavlja točke. Ker pa si točke težko predstavljamo, izračunamo še standardni odklon. Standardni odklon znaša 18,8 točk, in to pomeni, da se vrednosti spremenljivke odklanjajo od aritmetične sredine v povprečju za 18,8 točk (Brvar, 2007).

Variacijski razmik

Vzeli bomo enak primer kot pri prejšnjih primerih, to je primer podatkov o doseženih točkah, ki so jih dosegli policisti na testu znanja o uporabi prisilnih sredstev. Rezultati so bili naslednji:

67, 45, 91, 56, 87, 33, 61, 72, 60, 49, 55, 29, 67, 44, 43, 44, 59, 66, 83, 72, 88, 95, 39, 54, 86, 77, 88, 74, 82, 73, 66, 42, 89, 95, 33, 61, 57, 67, 34, 49.

Variacijski razmik znaša 66, to je razlika med največjim in najmanjšim številom točk, ki so jih dosegli policisti na testu znanja o uporabi prisilnih sredstev.

KOEFICIENT KORELACIJE

Za označevanje povezave oziroma soodvisnosti med dvema pojavoma uporabljamo izraz korelacija. V statistiki se ta izraz nanaša na to, kako močno in v kateri smeri sta dva pojava med seboj povezana. Ločimo pozitivno in negativno korelacijo. O pozitivni korelaciji govorimo takrat, ko dve spremenljivki skupaj naraščata, o negativni korelaciji pa, ko se spremenljivki gibljeta v nasprotni smeri (Sagadin, 1982).

PRIMERI KOEFICIENTA KORELACIJE

Pozitivna korelacija: Odnos med težavnostjo učne snovi pri nekem predmetu in časom, ki ga učenci potrebujejo, da se to snov naučijo. Z naraščanjem težavnosti snovi narašča tudi količina za učenje potrebnega časa.

Negativna korelacija: Odnos med količino zaužitega alkohola in zmožnostjo reševanja logičnih problemov. Z naraščanjem količine zaužitega alkohola se posameznikova zmožnost reševanja logičnih problemov zmanjša.

ZAKLJUČEK

Statistika nas spremlja vse življenje, prisotna je na vseh področjih naše aktivnosti. Njena naloga je raziskovanje zakonitosti, ki se pojavljajo v množičnih pojavih, procesih. Po pojavu osebnih računalnikov in še dodatno po nastanku spleta se je uporabnost statistike zelo povečala.

VIRI

Brvar, B. (2007). *Statistika*. Ljubljana: Fakulteta za varnostne vede, Univerza v Mariboru.

Dogša, I., Curk, J., Kompare, A., Stražišar, M., Vec, T. (2011). *Psihologija, spoznanja in dileme, učbenik za psihologijo v 4. letniku gimnazijskega izobraževanja*. Ljubljana: DZS.

Sagadin, J. (1982). *Osnovne statistične metode za pedagoge - druga, dopolnjena izdaja*. Ljubljana: Filozofska fakulteta Univerze Edvarda Kardelja v Ljubljani.

Sirkin, M. (1995). *Statistics for the Social Sciences*. London: Sage publications.