STATISTIKA

NEPARAMETARSKI TESTOVI

χ^2-TEST

NEPARAMETARSKI TESTOVI

- Za proveru statističkih hipoteza neparametarski testovi se koriste u slučajevima:

a) kada statističke parametre nije moguće izračunati zbog prirode obeležja (atributivno obeležje);

b) kada su vrednosti obeležja rangirane;

c) kada se izračunati statistički parametri odnose na osnovni skup čiji podaci nisu distribuirani po modelu normalne raspodele.

- Procedura provere hipoteze je ista kao kod parametarskih testova.

NEPARAMETARSKI TESTOVI

- Prednosti neparametarskih testova u odnosu na parametarske su:

a) Mogu da se izvode na skupovima različitih oblika raspodele;

b) Imaju primenu kod vrlo malih uzoraka, koji ne prelaze šest jedinica;

c) Jednostavni su.

Od postojećih neparametarskih testova najčešće se koristi χ^2 -test (hi-kvadrat test).

χ^2-TEST

- **χ^2-test** je neparametarski ***metod za ispitivanje značajnosti razlike apsolutnih frekvencija***.
- On se može primenjivati na nezavisne uzorke sa numeričkim ili atributivnim obeležjima za proveru većeg broja hipoteza.
- **χ^2-test je zasnovan na χ^2 distribuciji i koristi se za rešavanje nekoliko problema:**

1. **Test prilagođenosti;**
2. **Tabele kontingencije;**
3. **Test homogenosti.**

χ^2-TEST

- **<u>Test prilagođenosti</u>** je statistički test koji daje odgovor na pitanje u kojoj meri su empirijski podaci prilagođeni ili odgovaraju nekom teorijskom rasporedu verovatnoće (binomnom, Poasonovom, normalnom).
- Kod testa prilagođenosti nulta i alternativna hipoteza glase:

  H_o: Populacija je normalno raspoređena (ili Populacija je raspoređena po binomnom rasporedu ili Populacija je raspoređena po Poasonovom rasporedu)

χ^2-TEST

H_1: Populacija nije normalno raspoređena (ili Populacija nije raspoređena po binomnom rasporedu ili Populacija nije raspoređena po Poasonovom rasporedu).

- **Tabele kontingencije** omogućavaju da se ispita da li između dva obeležja elemenata jednog skupa postoji veza i da li je ta veza statistički značajna.
- Tabele kontingencije se formiraju tako što se podaci jednog uzorka razvrstavaju prema dva obeležja koja imaju najmanje dva modaliteta.

χ^2-TEST

- Tabele kontingencije mogu imati nekoliko redova i kolona, a najmanje dva reda i dve kolone.
- Nulta i alternativna hipoteza prilikom analize tabela kontingencije odnose se na testiranje nezavisnosti dva obeležja jedne populacije i glase:

  H_o: dva obeležja su međusobno nezavisna

  H_1: dva obeležja su međusobno zavisna

χ^2-TEST

- **Test homogenosti** predstavlja test jednakosti ili razlike proporcija za više populacija.
- Prilikom poređenja r populacija nulta i alternativna hipoteza glase:

  H_o: $\pi_1 = \pi_2 = \ldots = \pi_r$

  H_1: sve proporcije π_i, i=1,2,...,r nisu međusobno jednake
- Postupak testiranja je sledeći: Za svaku pojedinu populaciju uzima se u obzir opažena (empirijska) frekvencija (f_i) kao broj elemenata sa određenom osobinom.

χ^2-TEST

- Zatim se izračunavaju očekivane (teorijske) frekvencije (f_i’ ili f_i*) pod pretpostavkom da nema značajne razlike (da postoji jednakost) između posmatranih populacija prema učešću elemenata sa određenom osobinom.
- Nakon toga se računa vrednost χ^2 statistike testa prema fomuli:

$$\chi^2 = \sum_{i=1}^{r} \frac{(f_i - f_i^*)^2}{f_i^*}$$

pri čemu su fi opažene (empirijske) frekvencije, a f_i*očekivane (teorijske) frekvencije, a r predstavlja broj grupa frekvencija.

- U nastavku se određuje p-vrednost i izvodi zaključak.

χ^2-TEST

Uslovi za primenu χ^2 testa

1. χ^2 test se primenjuje samo prilikom testiranja apsolutnih frekvencija;
2. Zbir empirijskih i teorijskih frekvencija mora biti jednak $\sum f_i = \sum f_i'$;
3. Treba uzeti u obzir svako pojavljivanje i nepojavljivanje određene osobine da se ne bi narušio prethodni uslov. (Tako npr, ako se testiranjem neke pojave javljaju odgovori “da” i “ne”, to znači da se uz frekvenciju odgovora “da” mora pridružiti i frekvencija odgovora “ne”.);

χ^2-TEST

4. Frekvencije u pojedinim ćelijama moraju biti nezavisne;

5. Teorijske frekvencije ne smeju biti suviše male, odnosno manje od 5. Ako se u tabeli pojave male teorijske frekvencije tada je potrebno da se dodaju prethodnoj teorijskoj frekvenciji do ispunjenja ovog uslova. Zbog uporedivosti moraju se sažeti i odgovarajuće grupe empirijskih frekvencija. Sažimanje intervala se odražava na broj stepeni slobode.

χ^2-TEST

- **PRIMER:** Ispitivano je mišljenje građana o uvođenju novog načina naplate komunalnih usluga u 8 gradova. Postavljeno je pitanje: Da li su za takav način naplate? Dobijeni su sledeći podaci:

GRAD	BROJ ISPITANIKA	ODGOVORILI DA SU ZA NOV NAČIN NAPLATE	%
1	150	135	90,0
2	240	152	63,3
3	50	38	76,0
4	120	87	72,5
5	90	48	53,3
6	130	75	57,7
7	220	175	79,5
8	160	109	68,1
	1160		70,6

0,706

χ^2-TEST

- Na nivou značajnosti od 0,05 ispitati da li se stavovi građana po pitanju uvođenja novog načina naplate u ovih 8 gradova međusobno razlikuju.
- **<u>REŠENJE:</u>**
- H_0: $\pi_1 = \pi_2 = \ldots = \pi_8$

  H_1: sve proporcije π_i, i=1,2,…8 nisu međusobno jednake
- U opštem slučaju teorijske frekvencije dobijaju se tako što se ustanovi opšta (generalna) proporcija kao odnos zbira broja elemenata iz svih uzoraka koji imaju određenu osobinu i zbira broja elemenata svih uzoraka.

χ^2-TEST

- U našem slučaju opšta proporcija je 0,706.
- Opšta proporcija se množi sa svakim pojedinim obimom uzorka da bi se dobile teorijske frekvencije čiji zbir mora biti jednak zbiru empirijskih frekvencija.
- U ovom primeru teorijske frekvencije dobijaju na sledeći način:

$$f_1^* = 0{,}706 \cdot 150 = 105{,}90$$

$$f_2^* = 0{,}706 \cdot 240 = 169{,}44$$

$$f_3^* = 0{,}706 \cdot 50 = 35{,}30$$

$$f_4^* = 0{,}706 \cdot 120 = 84{,}72$$

χ^2-TEST

$$f_5^* = 0{,}706 \cdot 90 = 63{,}54$$

$$f_6^* = 0{,}706 \cdot 130 = 91{,}78$$

$$f_7^* = 0{,}706 \cdot 220 = 155{,}32$$

$$f_8^* = 0{,}706 \cdot 160 = 112{,}96$$

- U nastavku se izračunava vrednosti χ^2 statistike prema formuli:

$$\chi^2 = \sum \frac{(f_i - f_i^*)^2}{f_i^*}$$

- Neophodni elementi su dati u radnoj tabeli.

χ^2-TEST

Radna tabela

GRAD	ODGOVORILI DA SU ZA NOV NAČIN NAPLATE f_i	f_i^*	$(f_i - f_i^*)$	$(f_i - f_i^*)^2$	$\frac{(f_i - f_i^*)^2}{f_i^*}$
1	135	105,90	29,10	846,8100	7,9963
2	152	169,44	-17,44	304,1536	1,7951
3	38	35,30	2,70	7,2900	0,2065
4	87	84,72	2,28	5,1984	0,0614
5	48	63,54	-15,54	241,4916	3,8006
6	75	91,78	-16,78	281,5684	3,0679
7	175	155,32	19,68	387,3024	2,4936
8	109	112,96	-3,96	15,6816	0,1388
Σ	**819**				**19,5602**

χ^2-TEST

- Sada smo u mogućnosti da oderedimo p-vrednost na osnovu koje ćemo izvesti zaključak.
- Broj stepeni slobode u ovom primeru je v = r-1=8-1=7
- U tablicama χ^2 raspodele čitaju se kritične vrednosti za v =7 između kojih se nalazi realizovana vrednost χ^2 statistike iz ovog primera. To su vrednosti $\chi^2_{0,01} = 18{,}475$ i $\chi^2_{0,005} = 20{,}278$.
- Na osnovu čega zaključujemo da se p-vrednost nalazi u intervalu:

$$0{,}005 < \text{p-vrednost} < 0{,}01$$

χ^2-TEST

- Iz čega sledi da je p-vrednost manja od zadatog nivoa značajnosti (α=0,05), pa se može zaključiti da se postavljena nulta hipoteza odbacuje.
- Dakle, stavovi građana po pitanju uvođenja novog načina naplate u ovih 8 gradova međusobno se statistički značajno razlikuju.