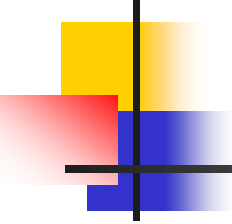


# 4. NEPARAMETRIČNE STATISTIKE

## 4.1. Uvod

- 
- Neparametrične statistike uporabljamo v naslednjih okoliščinah:
    - Merilna lestvica je razvrstitvena (ordinalna) in ne sorazmerna ali intervalna.
    - Merilna lestvica je sorazmerna ali intervalna, a frekvenčna porazdelitev odstopa od normalne. Preden se odrečemo parametričnim metodam, skušamo podatke s transformacijo prilagoditi normalni porazdelitvi.
    - Število opazovanj je izredno majhno.

# 4.1.1. Uvod

- Neparametrične metode temeljijo mediani in razvrstitvi (ranku) vrednosti podatkov.
- Razvrstitveni podatki ohranijo vso informacijo, sorazmerni in intervalni pa je nekaj izgubijo.

PODATKI	PARAMETRIČNA	NEPARAMETRIČNA
Normalni, velik $n$ , sorazmernostni ali intervalni	Najbolj natančni in zanesljivi rezultati	Zavrnamo $H_0$ : izid enak parametričnemu Sprejmemo $H_0$ : nezanesljiv izid
Nenormalni, majhen $n$ , ordinalni	Popolnoma nezanesljivi rezultati, ki pogosto vodijo do neupravičene zavrnitve $H_0$	Najboljši možni rezultati glede na kakovost podatkov



## 4.1.2. Uporaba razvrstitve

---

- Razvrstitev (rank) je mesto vrednosti v zaporedju, urejenem od najnižje proti najvišji (ali obratno) vrednosti podatkov.
- Vrednost podatka  $x_i$  predstavimo z razvrstitvijo  $R(x_i)$ .
- Primer:

$$x_1=4,3 \quad x_2=9,3 \quad x_3=0,3 \quad x_4=2,9 \quad x_5=3,2 \quad x_6=7,7 \quad x_7=5 \quad x_8=0,4$$

$$R(x_1)=5 \quad R(x_2)=8 \quad R(x_3)=1 \quad R(x_4)=3 \quad R(x_5)=4 \quad R(x_6)=7 \quad R(x_7)=6 \quad R(x_8)=2$$

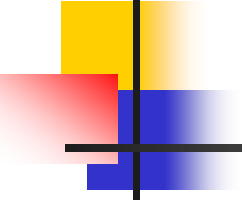
## 4.1.2. Uporaba razvrstitve

- Kadar obstaja več enakih vrednosti, njihovo razvrstitev združimo, tako da je njihova razvrstitev povprečje razvrstitve skupine.
- Primer:

$x_i$	56	42	61	61	42	55	35	42	39	65
$R(x_i)$	7	4	8,5	8,5	4	6	1	4	2	10
	35	39	42	42	42	55	56	61	61	65

## 4.2. Univariatne metode

### 4.2.1. Test predznaka hipotetične mediane

- 
- Enakovreden je parametričnemu t-testu ujemanja srednje vrednosti z neko predpostavljeno vrednostjo.
  - $H_0$  trdi, da podatki izhajajo iz populacije z določeno predpostavljeno mediano.
  - S + in – ponazorimo, ali je vsak posamezen podatek od predpostavljene Me višji ali nižji.
  - Bolj kot se hipotetična Me in Me vzorca razlikujeta, večje je neravnotežje med + in – ter s tem manjša verjetnost, da vzorec izhaja iz te populacije.

## 4.2.1. Test predznaka hipotetične mediane

- $H_0: Me_{\text{populacija}} = Me_{\text{predpostavljena}}$   
 $H_1: Me_{\text{populacija}} \neq Me_{\text{predpostavljena}}$
- Zavrni  $H_0$  na ravni zaupanja  $\alpha = 0,05$ , če je predpostavljena mediana izven razpona vrednosti, ki ustreza razponu
$$t_{0,05,n-1} \cdot \sqrt{\frac{n}{2}}$$
- Na obe strani mediane vzorca.
- Pri  $n < 20$  uporabimo tabele binomne porazdelitve.

## 4.2.2. Mann – Whitneyev test

- Je neparametrična različica t-testa enakosti dveh populacijskih povprečij.
- $H_0: Me_x = Me_y$   
 $H_1: Me_x \neq Me_y$
- Poiščemo skupne razvrstitve za podatke obeh primerjanih vzorcev, ter izračinamo testno statistiko T:

$$T = \sum_{i=1}^n R(x_i) - \frac{n \cdot (n + 1)}{2}$$

## 4.2.2. Mann – Whitneyev test

- $R(x_i)$  je razvrstitev vzorca  $x$  in  $n$  njegova velikost.
- Razvrstitev vzorca  $y$  za test ne potrebujemo.
- Velikost vzorca  $y$  je  $m$ . Potrebujemo jo za primerjavo izračunanega  $T$  s kritično vrednostjo iz tabel.
- $H_0$  zavrnamo, kadar je  $T_{\text{izračunani}} > T_{\text{tabelirani}}$



## MANN-WHITNEY TEST ( $\alpha = 5\%$ )

<i>m</i>	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	25	30	35		
<i>n</i>																					
3	1	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	11	13	16		
	2	2	2	3	3	4	5	5	6	7	7	8	9	9	10	11	15	18	22		
4	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13	19	23	27		
	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18	24	29	34		
5	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20	27	33	39		
	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25	32	40	47		
6		5	6	8	10	11	13	14	16	17	19	21	22	24	25	27	35	43	51		
		7	8	10	12	14	16	17	19	21	23	25	26	28	30	32	42	51	60		
7			8	10	12	14	16	18	20	22	24	26	28	30	32	34	44	54	64		
			10	13	15	17	19	21	24	26	28	30	33	35	37	39	131	62	73		
8				13	15	17	19	22	24	26	29	31	34	36	38	41	53	65	77		
				15	18	20	23	26	28	31	33	36	39	41	44	47	63	74	87		
9					17	20	23	26	28	31	34	37	39	42	45	48	62	76	90		
					21	24	27	30	33	36	39	42	45	48	51	54	70	89	100		
10						23	26	29	33	36	39	42	45	48	52	55	71	87	103		
						27	31	34	37	41	44	48	51	55	58	62	79	97	114		
11							30	33	37	40	44	47	51	55	58	62	80	98	116		
							34	38	42	46	50	54	57	61	65	69	89	109	128		
12								37	41	45	49	53	57	61	65	69	89	109	129		
								42	47	51	55	60	64	68	72	77	99	120	142		
13									45	50	54	59	63	67	72	76	98	120	143		
									51	56	61	65	70	75	80	84	109	132	156		
14										55	59	64	67	74	78	83	108	132	156		
										61	66	71	77	82	87	92	118	144	170		
15											64	70	75	80	85	90	117	143	169		
											72	77	83	88	94	100	128	156	184		
16												75	81	86	92	98	126	155	183		
												83	89	95	101	107	138	168	198		
17													87	93	99	105	136	166	197		
													96	102	109	115	148	180	213		
18														99	106	112	145	177	210		
														109	116	123	158	192	227		
19															113	119	154	189	224		
															123	130	168	204	241		
20																127	164	201	237		
																138	178	216	256		
25																	201	259	306		
																	227	277	327		
30																			317	376	
																			338	400	
35																					445

Upper and lower entries are for two-tailed and one-tailed tests respectively. As the distribution is symmetrical in (*m*,*n*) repeated values have been omitted.

## 4.2.3. Kruskal – Wallisev test



---

- Je preprosta razširitev Mann – Whitneyevega testa.
- Analogen je enojni analizi variance.
- $H_0$ : Me vseh populacij so enake  
 $H_1$ : Me vsaj ene od populacij se razlikuje

## 4.2.3. Kruskal – Wallisev test

- Testna statistika je  $H$ .

$$H = \frac{12}{N \cdot (N + 1)} \cdot \sum_{j=1}^k \frac{\left( \sum_{i=1}^{n_j} R(x_{ij}) \right)^2}{n_j} - 3 \cdot (N + 1)$$

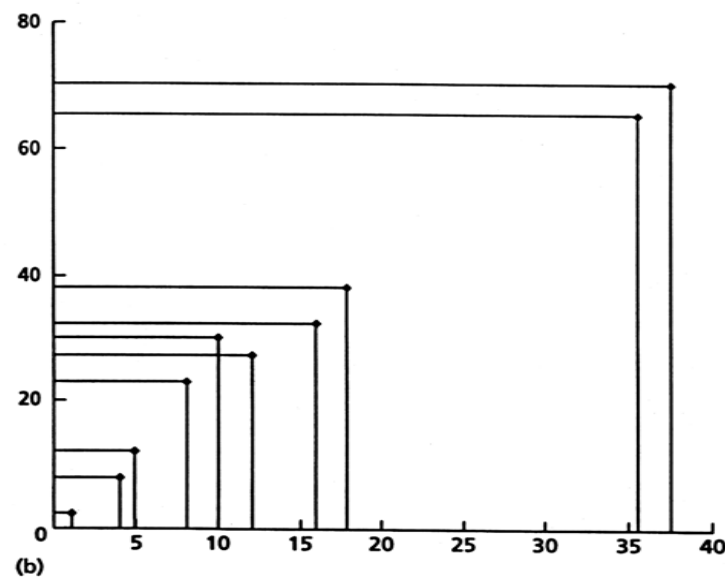
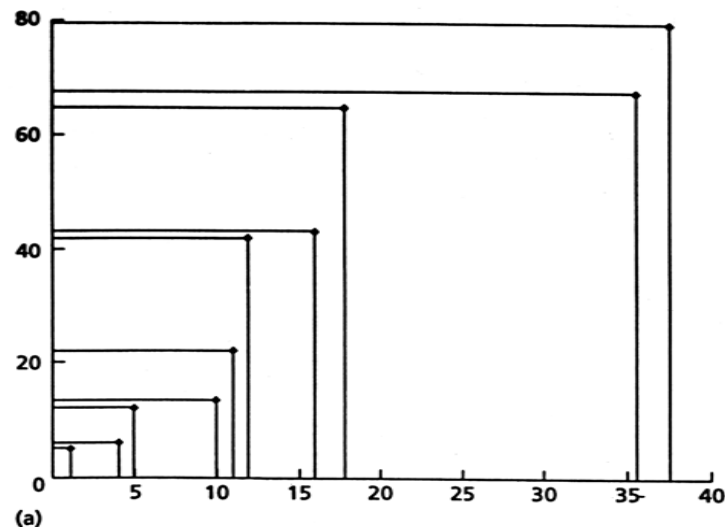
- $k$  vzorcev velikosti  $n_1, n_2, \dots, n_k$  daje skupno velikost vzorca  $N$ .  $R(x_{ij})$  je razvrstitev  $i$ -tega podatka  $j$ -tega vzorca.

- Testno statistiko  $H$  primerjamo s kritičnimi vrednostmi  $X^2_{\alpha, k-1}$ .

# 4.3. Bivariatne metode

## 4.3.1. Neparametrična korelacija

- Spearmanov razvrstitveni korelacijski koeficient ( $r_s$ ) je neparametrični različek Paersonovega korelacijskega koeficienta ( $r$ ).
- Za spremenljivki  $x$  in  $y$  ločeno poiščemo njune razvrstitve  $R(x_i)$  in  $R(y_i)$ .
- Kadar je za vsako opazovanje  $i$ , razvrstitev  $x$  enaka razvrstitvi  $y$ , je razvrstitvena korelacija popolna.

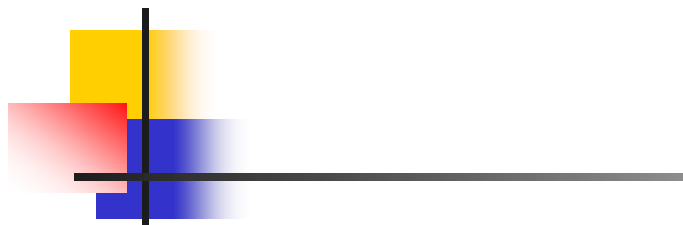


# 4.3.1. Neparametrična korelacija

- $H_0: \rho' = 0$       korelacije ni
- $H_1: \rho' \neq 0$       korelacija obstaja
- Spermanov razvrstitveni korelacijski koeficient temelji na vsoti razlik med odgovarjajočima razvrstitvama  $x$  in  $y$ .

$$r' = 1 - \frac{6 \cdot \sum_{i=1}^n (R(x_i) - R(y_i))^2}{n \cdot (n^2 - 1)}$$

- Izračunano vrednost primerjamo s tabelirano.
- $H_0$  zavrnamo, če  $r'_{\text{izračunani}} > r'_{\text{tabelirani}}$ .



	Significance, $\alpha$ , for One-tailed Test					
	.10	.05	.025	.01	.005	.001
	Significance, $\alpha$ , for Two-tailed Test					
	.20	.10	.05	.02	.01	.002
<i>n</i>						
4	.8000	.8000				
5	.7000	.8000	.9000	.9000		
6	.6000	.7714	.8286	.8857	.9429	
7	.5357	.6786	.7450	.8571	.8929	.9643
8	.5000	.6190	.7143	.8095	.8571	.9286
9	.4667	.5833	.6833	.7667	.8167	.9000
10	.4424	.5515	.6364	.7333	.7818	.8667
11	.4182	.5273	.6091	.7000	.7455	.8364
12	.3986	.4965	.5804	.6713	.7273	.8182
13	.3791	.4780	.5549	.6429	.6978	.7912
14	.3626	.4593	.5341	.6220	.6747	.7670
15	.3500	.4429	.5179	.6000	.6536	.7464
16	.3382	.4265	.5000	.5824	.6324	.7265
17	.3260	.4118	.4853	.5637	.6152	.7083
18	.3148	.3994	.4716	.5480	.5975	.6904
19	.3070	.3895	.4579	.5333	.5825	.6737
20	.2977	.3789	.4451	.5203	.5684	.6586
21	.2909	.3688	.4351	.5078	.5545	.6455
22	.2829	.3597	.4241	.4963	.5426	.6318
23	.2767	.3518	.4150	.4852	.5306	.6186
24	.2704	.3435	.4061	.4748	.5200	.6070
25	.2646	.3362	.3977	.4654	.5100	.5962
26	.2588	.3299	.3894	.4564	.5002	.5856
27	.2540	.3236	.3822	.4481	.4915	.5757
28	.2490	.3175	.3749	.4401	.4828	.5660
29	.2443	.3113	.3685	.4320	.4744	.5567
30	.2400	.3059	.3620	.4251	.4665	.5479

# 4.3.1. Neparametrična korelacija

- Raje kot o linearnem odnosu, govorimo o visoki enakomernosti med spremenljivkama  $x$  in  $y$ .
- Dokaj pogosto je Pearsonov koeficient visok, Spearmanov pa nizek, ali obratno.
- Tudi v primeru normalno porazdeljenih podatkov parametričen test ni bistveno boljši od neparametričnega.

