

TRANSFORMACIJE PODATAKA

Vježba 4



- 1 Rangiranje
- 2 Z - vrijednosti
- 3 Utvrđivanje vjerojatnosti putem normalne distribucije
- 4 Ostale transformacije podataka (skala školskih ocjena skala T – skorova)

Rangiranje

Rangiranje je transformacija kvantitativne (intervalne ili omjerne) varijable u ordinalnu, odnosno zamjena rezultata odgovarajućim rangovima.

Rang je redni broj entiteta utvrđen sortiranjem prema nekoj kvantitativnoj varijabli.

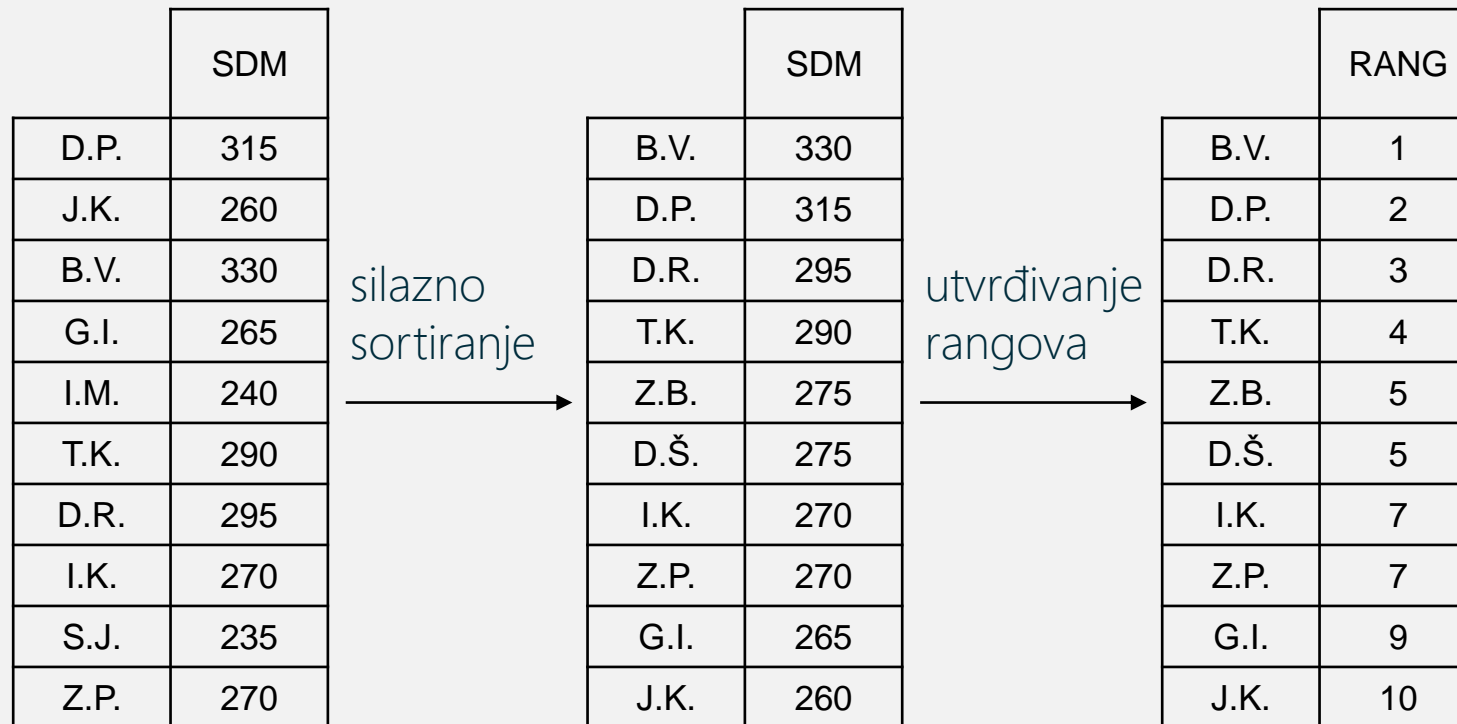
Uobičajeno je da se najbolji rezultat zamjenjuje rangom 1. Sukladno tome rang se entiteta u *normalno skaliranim varijablama* utvrđuje silaznim sortiranjem, a u *obrnuto skaliranim varijablama* uzlaznim sortiranjem.

Normalno skalirana varijabla - najveća numerička vrijednost predstavlja najbolji rezultat (npr. *Skok uvis*)

Obrnuto skalirana varijabla - najmanja numerička vrijednost predstavlja najbolji rezultat (npr. *Sprint 100 metara*)

Rangiranje

Primjer: 10 sportaša je izmjereno testom Skok udalj s mjesta. Rangiranje je izvršeno sljedećim postupkom:



Z - vrijednosti

Postupak standardizacije provodi se pomoću formule

$$z_i = \frac{x_i - \bar{x}}{s}$$

gdje je:

z_i – standardizirani rezultat entiteta i

x_i – originalna vrijednost entiteta i

\bar{x} – aritmetička sredina

s – standardna devijacija.

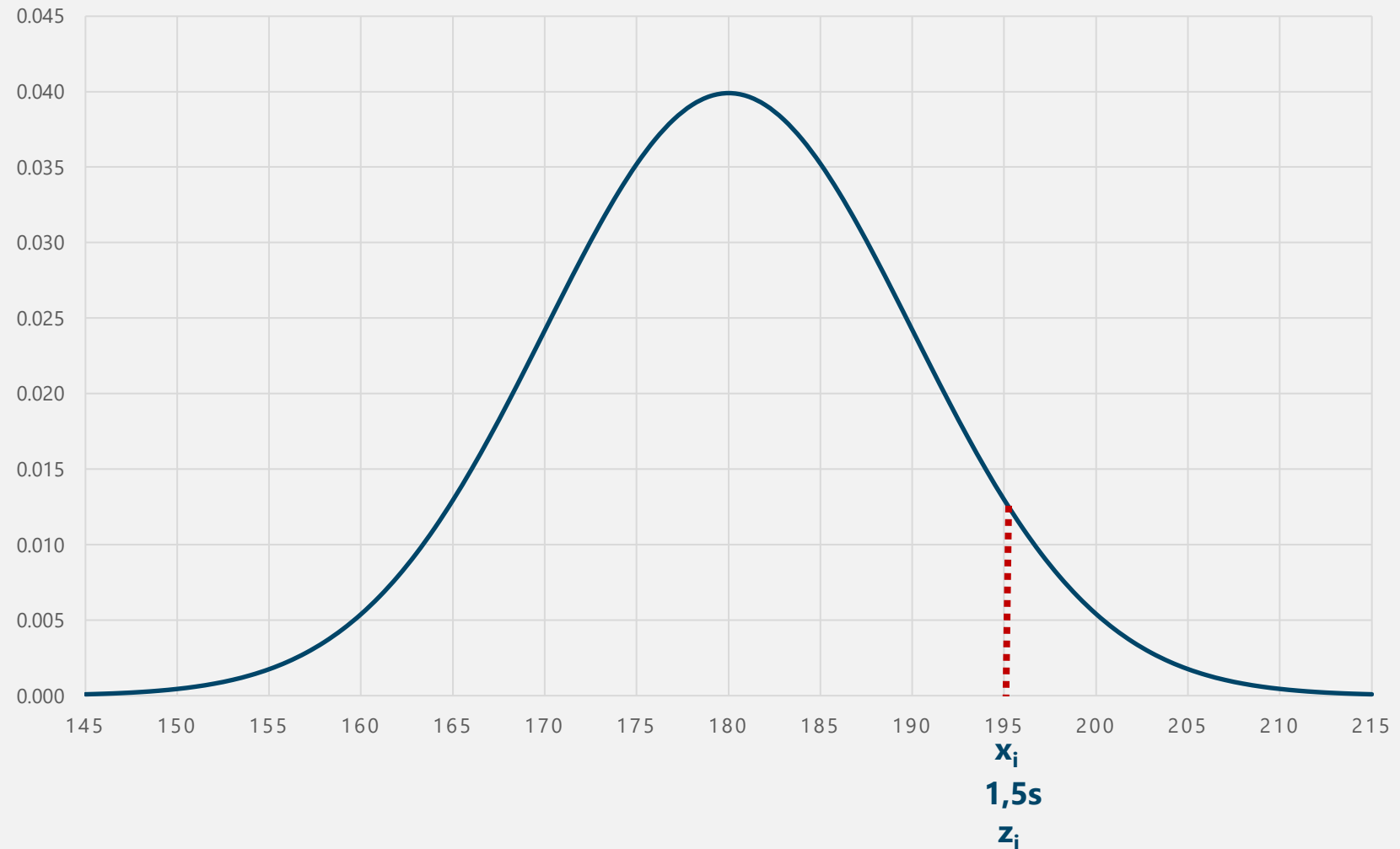
Z - vrijednosti

$$\bar{x} = 180 \text{ cm}$$

$$s = 10 \text{ cm}$$

$$x_i = 195 \text{ cm}$$

$$z_i = \frac{195 - 180}{10} = 1,5s$$



Z - vrijednosti

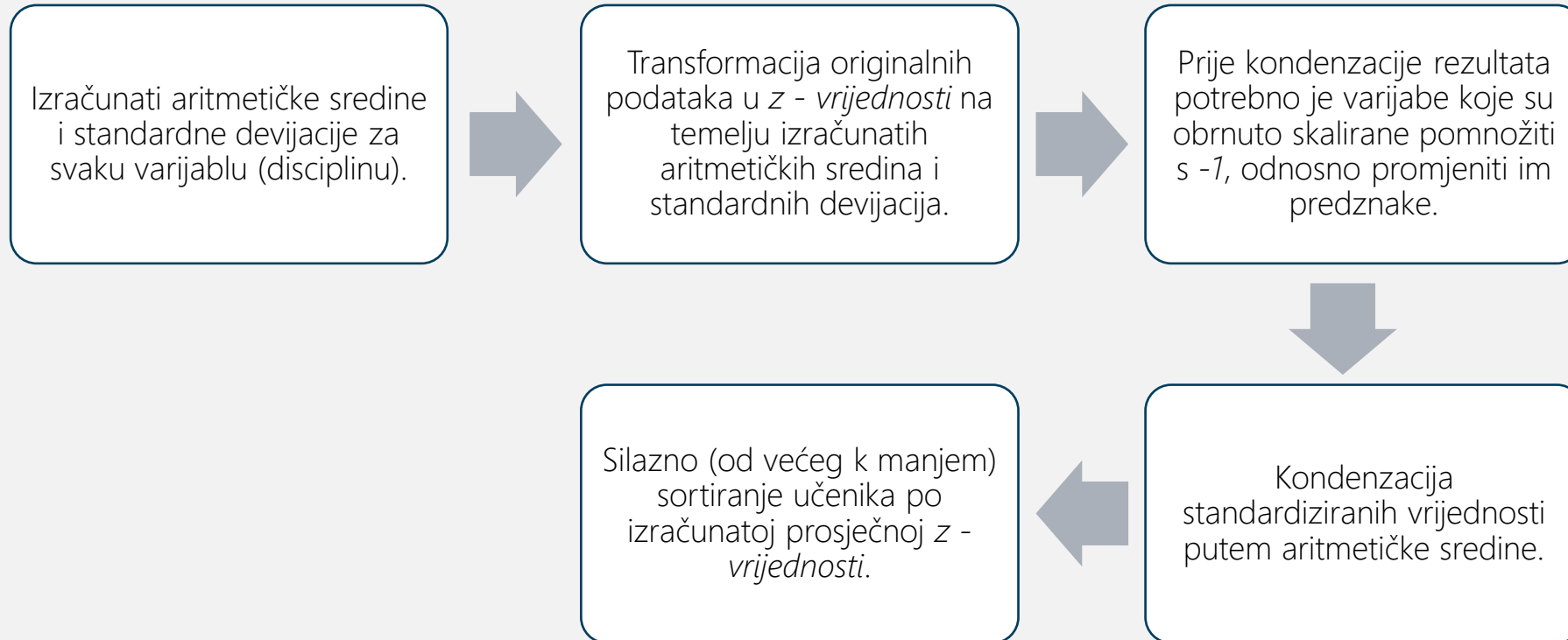
Primjer: Deset učenika natjecalo se u tri atletske discipline:

- skok udalj (SD),
- trčanje na 100 metara (T100m) i
- bacanje kugle (BK) i postiglo rezultate navedene u tablici.

Potrebno je utvrditi ukupan poredak (rang) ovog natjecanja ?

Učenik	SD	T100m	BK
AB	359	13,6	561
DF	321	13,9	550
JG	346	13,7	538
KL	332	14,0	490
DD	450	12,2	518
ED	314	14,1	551
TB	410	12,5	589
ZN	425	12,3	602
RG	369	13,5	547
EN	378	13,8	510

Z - vrijednosti



Z - vrijednosti

Prvi korak: Izračunati aritmetičke sredine i standardne devijacije za svaku varijablu (disciplinu).

	SD	T100m	BK
\bar{x}	370,4	13,36	545,6
s	45,66	0,73	34,21

Učenik	SD	T100m	BK
AB	359	13,6	561
DF	321	13,9	550
JG	346	13,7	538
KL	332	14,0	490
DD	450	12,2	518
ED	314	14,1	551
TB	410	12,5	589
ZN	425	12,3	602
RG	369	13,5	547
EN	378	13,8	510

Z - vrijednosti

Drugi korak: Transformacija originalnih podataka u z - vrijednosti na temelju izračunatih aritmetičkih sredina i standardnih devijacija.

	SD	T100m	BK
\bar{x}	370,4	13,36	545,6
s	45,66	0,73	34,21

$$z_{AB,SD} = \frac{359 - 370,4}{45,66} = -0,25s$$

Učenik	SD	T100m	BK
AB	359	13,6	561
DF	321	13,9	550
JG	346	13,7	538
KL	332	14,0	490
DD	450	12,2	518
ED	314	14,1	551
TB	410	12,5	589
ZN	425	12,3	602
RG	369	13,5	547
EN	378	13,8	510

Z - vrijednosti

Treći korak: Prije kondenzacije rezultata varijable koje su obrnuto skalirane pomnožiti s -1 , odnosno rezultatima promijeniti predznake.

Učenik	SD	T100m	BK
AB	-0,25	0,33	0,45
DF	-1,08	0,74	0,13
JG	-0,53	0,46	-0,22
KL	-0,84	0,87	-1,63
DD	1,74	-1,58	-0,81
ED	-1,24	1,01	0,16
TB	0,87	-1,17	1,27
ZN	1,20	-1,44	1,65
RG	-0,03	0,19	0,04
EN	0,17	0,60	-1,04

Z - vrijednosti

Četvrti korak: Kondenzacija standardiziranih vrijednosti putem aritmetičke sredine.

$$\begin{aligned}\bar{z}_{AB} &= \frac{z_{AB,SD} + z_{AB,T100m} + z_{AB,BK}}{3} = \\ &= \frac{-0,25 + (-0,33) + 0,45}{3} = -0,04\end{aligned}$$

Učenic	SD	T100m	BK	\bar{z}
AB	-0,25	-0,33	0,45	-0,04
DF	-1,08	-0,74	0,13	
JG	-0,53	-0,46	-0,22	
KL	-0,84	-0,87	-1,63	
DD	1,74	1,58	-0,81	
ED	-1,24	-1,01	0,16	
TB	0,87	1,17	1,27	
ZN	1,20	1,44	1,65	
RG	-0,03	-0,19	0,04	
EN	0,17	-0,60	-1,04	

Z - vrijednosti

Četvrti korak: Kondenzacija standardiziranih vrijednosti putem aritmetičke sredine.

$$\begin{aligned}\bar{z}_{AB} &= \frac{z_{AB,SD} + z_{AB,T100m} + z_{AB,BK}}{3} = \\ &= \frac{-0,25 + (-0,33) + 0,45}{3} = -0,04\end{aligned}$$

Učenic	SD	T100m	BK	\bar{z}
AB	-0,25	-0,33	0,45	-0,04
DF	-1,08	-0,74	0,13	-0,56
JG	-0,53	-0,46	-0,22	-0,41
KL	-0,84	-0,87	-1,63	-1,11
DD	1,74	1,58	-0,81	0,84
ED	-1,24	-1,01	0,16	-0,70
TB	0,87	1,17	1,27	1,10
ZN	1,20	1,44	1,65	1,43
RG	-0,03	-0,19	0,04	-0,06
EN	0,17	-0,60	-1,04	-0,49

Z - vrijednosti

Peti korak: Silazno (od većeg k manjem) sortiranje učenika po izračunatoj prosječnoj z - vrijednosti.

Učenik	SD	\bar{z}
ZN	1	1,43
TB	2	1,10
DD	3	0,84
AB	4	-0,04
RG	5	-0,06
JG	6	-0,41
EN	7	-0,49
DF	8	-0,56
ED	9	-0,70
KL	10	-1,11

Ostale transformacije

Standardizirane rezultate moguće je transformirati na različite načine, zavisno o potrebi. Najčešće se transformiraju u vrijednosti na sljedećim skalama:

- skala školskih ocjena (1-5): $ocjena_i = 3 + 0,83 \cdot z_i$
- skala T – skorova (20-80): $T - skor_i = 50 + 10 \cdot z_i$



Zadatak 1: U datoteci *JUDO3F.xls* utvrdite rangove judaša na temelju rezultata u varijabli ONT (okretnost na tlu). Iskoristite funkciju Rank.

1

UTVRĐIVANJE RANGA: Utvrđivanje rangova vrši se pomoću funkcije *Rank*. Funkcija se unosi u označeno polje matrice odabirom opcije *Function...* U traku *Order* dijaloškog okvira za unos ove funkcije potrebno je upisati *1* ako se utvrđuje rang u obrnuto skaliranoj varijabli.

	A	B	C	D	E	F	G
1	ENTITETI	ONT	OUZ	NEB	SKL	(...)0 - Descending	
2	Marko	25,4	5,4	4	7	(...)1 - Ascending	
3	Mate	14,7	4,9	14	12	20	132



Zadatak 2: U datoteci *JUDO3F.xls* izračunajte z-vrijednosti za sve varijable i rangirajte entitete po prosječnoj z-vrijednosti svih varijabli.

Napomena: Varijable ONT, OUZ i T20m su obrnuto skalirane.

2

RAČUNANJE Z - VRIJEDNOSTI: Izračunavanje z-vrijednosti vrši se pomoću funkcije *Standardize*. Funkcija se unosi u označeno polje matrice odabirom opcije *Function...* Prethodno je potrebno izračunati aritmetičku sredinu (funkcija *Average*) i standardnu devijaciju (funkcija *Stdev*) varijable.

K	L	M	N	O
	ONT			
	=STANDARDIZE(
	STANDARDIZE(x; mean; standard_dev)			



Zadatak 3: 257 dječaka je izmjereno testom za procjenu eksplozivne snage *Skok udalj s mjesta*. Aritmetička sredina iznosi 215 cm, a standardna devijacija 12 cm. Učenik XY postigao je rezultat 230 cm. Potrebno je procijeniti postotak (%) i broj učenika koji postižu lošije i bolje rezultate od učenika XY.

$$z_{XY} = \frac{230-215}{12} = 1,25s$$

$$p = \text{NORM.S.DIST}(1,25; \text{TRUE})$$

$$p = 0,8944 \rightarrow 98,44\% \text{ lošijih}$$

$$1 - p = 0,1056 \rightarrow 10,56\% \text{ boljih}$$

$$p \cdot n = 0,1056 \cdot 257 \approx 27 - \text{boljih}$$

$$257 - 27 = 230 \text{ lošijih}$$

3

RAČUNANJE POVRŠINE ISPOD NORMALNE DISTRIBUCIJE:

Površina ispod normalne distribucije od x do lijevog kraja distribucije računa se formulom:

= NORM.DIST(x; as; sd; TRUE) i predstavlja vjerojatnost postizanja većih rezultata od x.

	A	B	C	D	E	F	G
7							
8	as =	215					
9	sd =	12					
10	x =	230					
11	p =	0,8944					



Zadatak 4: Aritmetička sredina svih službenih skokova udalj nekog atletičara iznosi 8,5 metara, a standardna devijacija 0,1 metar. Putem normalne distribucije procijenite kolika je vjerojatnost da ovaj atletičar na natjecanju izvede skok kraći od 8,3 metara (klasifikacijska norma za OI) i kolika je vjerojatnost da izvede skok dulji od 8,95 (svjetski rekord)

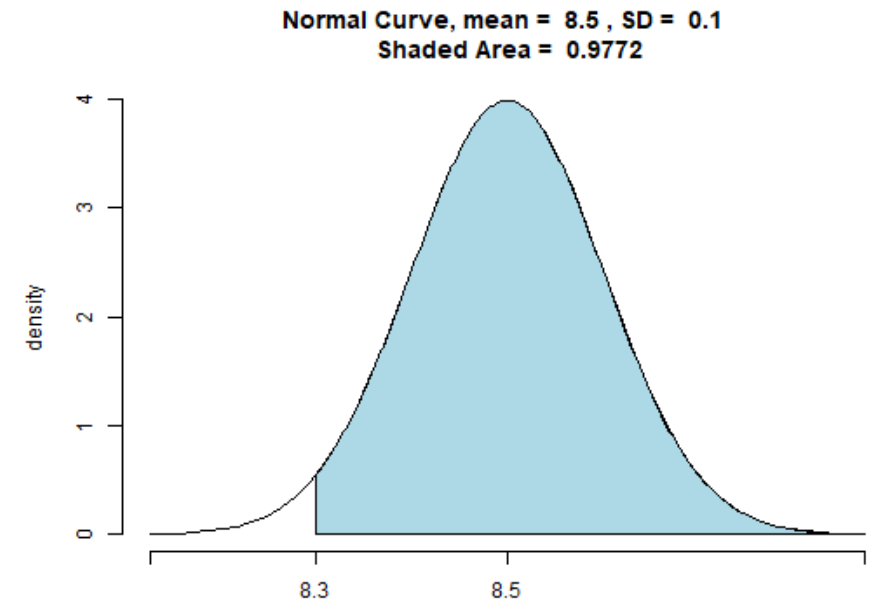
Normal (Gauss) Distribution - Probability Calculator

Normal (Gauss) Distribution - Probability Calculator

Mean

Standard Deviation

Value (X)





Zadatak 5: U datoteci *JUDO3F.csv* izvršite standardizaciju svih varijabli.

Z - value

Select Variables:

- ONT
- OUZ
- NEB
- SKL
- TRB
- CUC
- SDM
- BML
- T20m

Copy

	ONT ↕	OUZ ↕	NEB ↕	SKL ↕	TRB ↕	CUC ↕	SDM ↕	BML ↕	T20m ↕
Marko	2.72	2.75	-1.93	-1.09	-1.44	-1	-1.02	-0.22	2.08
Mate	-0.35	1.83	1.24	-0.52	-1.2	-0.7	-2.51	-2.15	2.48
Šime	0.13	0	-0.03	-0.41	0.74	-0.54	-0.13	0.21	0.54
Mile	-0.18	0.55	0.29	-0.29	-0.6	-0.41	-0.43	1.29	-0.27
Jure	0.42	0.36	-2.25	-0.75	-1.63	-0.92	-1.32	-0.11	-0.16
Ante	-0.27	-0.37	-0.03	0.39	-0.29	-0.73	-0.43	-0.32	0.37
Ive	-0.27	0.18	-1.93	-0.41	-0.48	-0.89	-0.73	-1.51	0.72
Stipe	1	1.46	-0.67	-1.54	-1.38	-0.4	-1.62	0.54	0.34
Tin	0.45	0.18	-0.35	0.39	-0.23	-0.48	-1.32	-1.72	1.63
Dino	-0.35	1.1	-0.98	-0.63	-0.9	-0.78	1.05	-0.11	0.62
Darko	0.59	1.46	1.55	-0.97	-0.84	-0.19	-0.13	1.07	-0.59
Stanko	-0.21	0.73	-0.67	-0.52	-0.72	-0.18	-1.62	0.96	0.71
Branko	0.02	0.36	-2.57	0.73	-0.17	-0.03	0.16	-0.32	0.34
Žarko	-0.64	-0.19	0.6	1.53	1.4	0.89	1.94	1.72	-1.83
Sanimir	0.65	1.46	-0.03	-1.54	-0.9	-0.92	0.16	0.64	-0.67
Toni	-0.01	0.18	-0.03	-0.41	-0.35	-0.09	1.65	0.11	-1.42