

# Transformacije podataka

# Teorijske osnove

---

Sadržaj:

- 1 rangiranje
- 2 standardizacija podataka (z-vrijednosti)
- 3 utvrđivanje vjerojatnosti putem normalne distribucije
- 4 ostale transformacije podataka (skala školskih ocjena skala T – skorova)

# Teorijske osnove

---

## 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*)

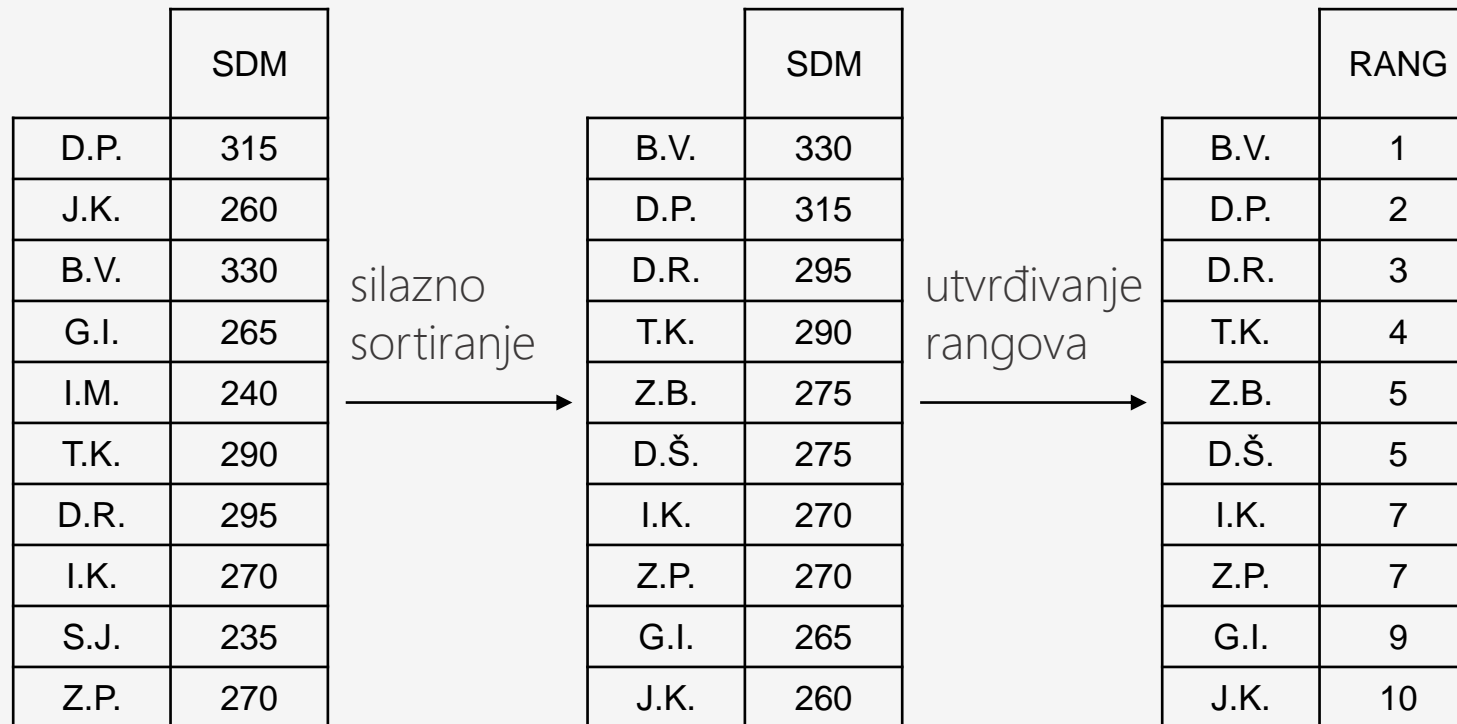
**Obrnuto skalirana varijabla** - najmanja numerička vrijednost predstavlja najbolji rezultat (npr. *Sprint 100 metara*)

# Teorijske osnove

---

## Rangiranje

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



# Teorijske osnove

---

## 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.

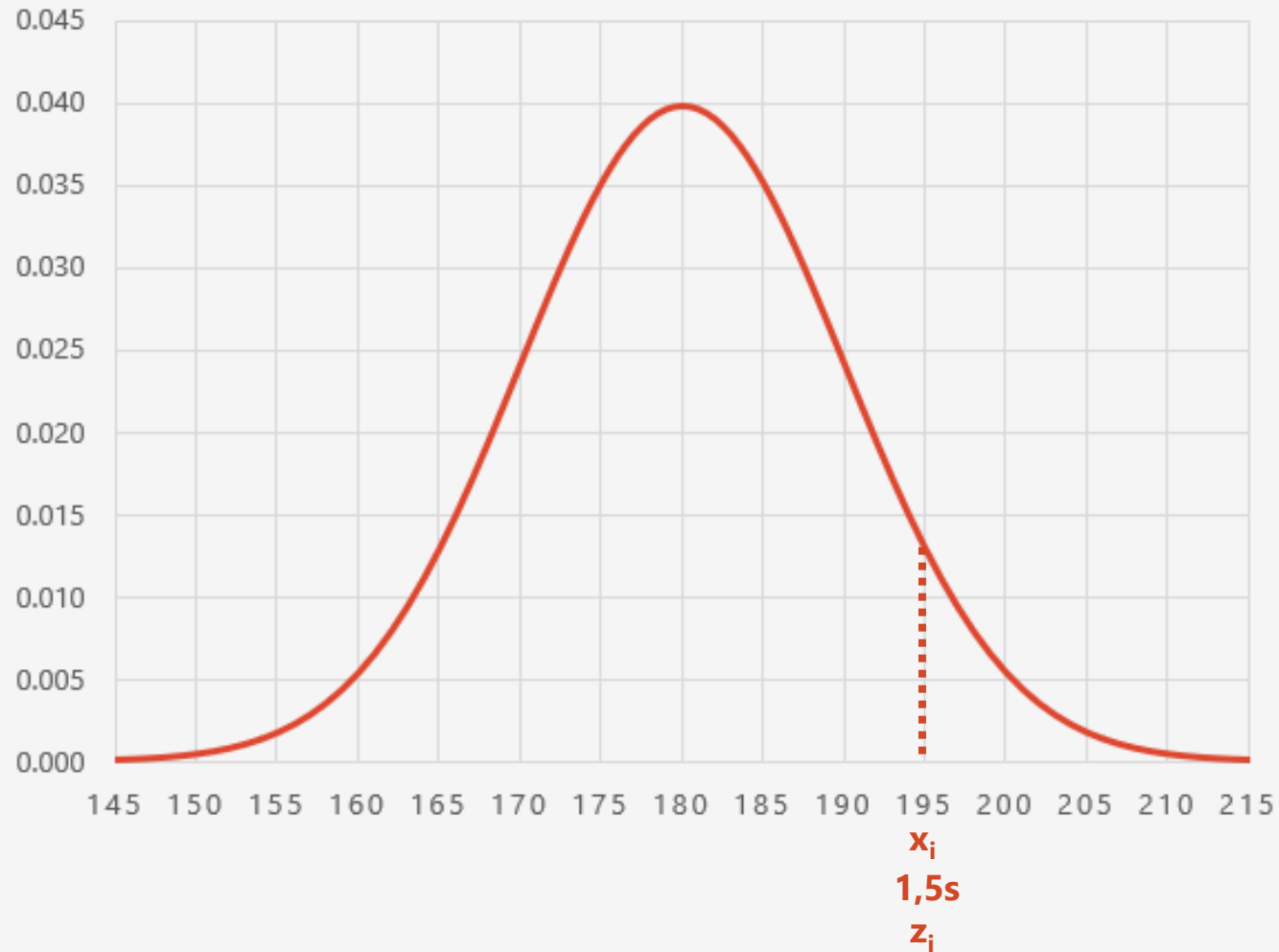
# Teorijske osnove

## Z - vrijednosti

$\bar{x} = 180$  cm  
 $s = 10$  cm  
 $x_i = 195$  cm

---

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



# Teorijske osnove

---

## 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.

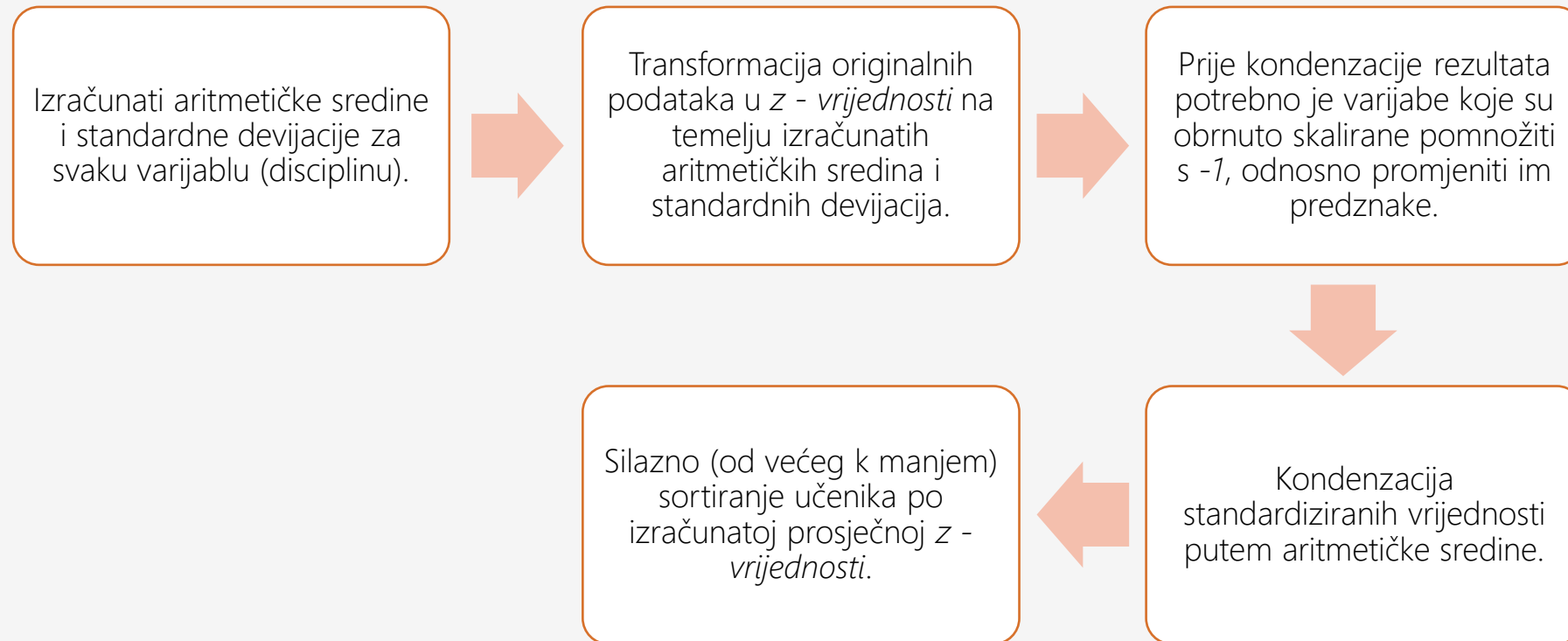
Potrebno je utvrditi ukupan poredak (rang) ovog natjecanja ?

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

# Teorijske osnove

---

## Z - vrijednosti





# Teorijske osnove

## 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

# Teorijske osnove

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

# Teorijske osnove

---

## Z - vrijednosti

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

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

# Teorijske osnove

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

# Teorijske osnove

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

# Teorijske osnove

---

## Z - vrijednosti

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

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

# Teorijske osnove

---

## 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):  $T - skor_i = 50 + 10 \cdot z_i$

# Microsoft Excel



**Zadatak 1:** U datoteci Judo.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



# Microsoft Excel



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

Napomena: Varijable ONT i OUZ 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)			

# Microsoft Excel



**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					