

::: statistična obdelava podatkov :::
...: vsota velikosti vseh trdih diskov v računalniku ...:

1....: uvod :..

Statistika je veda, ki proučuje množične pojave. Z zbiranjem, urejanjem, grupiranjem, povzemanjem, prikazovanjem in analiziranjem številskih podatkov o teh pojavih skuša odkriti njihove splošne zakonitosti in nato pridobljena spoznanja izkoristi za oblikovanje ustreznih napovedi oziroma odločitev.

Pri obdelavi se srečamo z različnimi pojmi. To so:

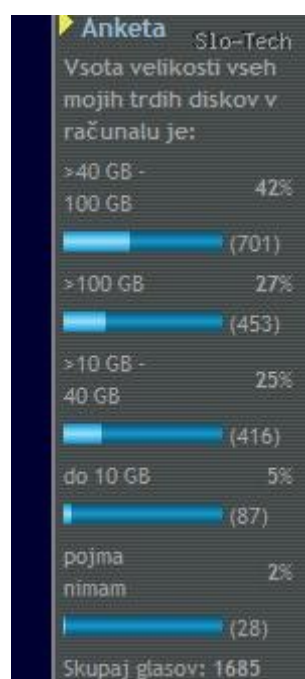
populacija oz. statistična množica - končna ali neskončna množica, ki jo preučujemo. Sestavljena je lahko iz živih bitij, predmetov, dogodkov, oziroma iz česar koli drugega

statistična enota - posamezen element populacije

statistična spremenljivka - značilnosti populacije, ki nas pri proučevanju zanimajo

vzorec - primerno izbrana končna podmnožica celotne populacije, ko preučujemo ustrezno statistično spremenljivko

Sam sem se odločil statistično obdelati anketo, ki sem jo našel na računalniški internetni strani <http://www.slo-tech.com>, ki jo redno obiskujem. Anketa sprašuje, koliko gigabajtov razpoložljivega prostora imajo uporabniki na svojih trdih diskih v računalniku.



2....: vsota velikosti vseh trdih diskov v računalniku - podatki :..

...: skupina :..	...: število :..	...: procentualno :..
pojma nimam	28	1.70%
do 10 GB	87	5.20%
od 10 do 40 GB	416	24.60%
od 40 do 100 GB	701	41.60%
nad 100 GB	453	26.90%
skupaj	1685	100.00%

3....: absolutna frekvenca, relativna frekvenca, komulativa frekvenc, komulativa relativnih frekvenc :..

Razred k	Vsota prostora na diskih – v GB	Sredina (y_k)	f_k	f_k^0	F_k	F_k^0
1	0	0	28	0.017	0	0.000
2	0 - 10	5	87	0.052	28	0.017
3	10 - 40	25	416	0.246	115	0.069
4	40 - 100	70	701	0.416	531	0.315
5	nad 100	/	453	0.269	1232	0.731
		skupaj	1685	1.000	1685	1.000

f_k – **absolutna frekvenca** nam pove kolikokrat je proučevana spremenljivka vsote prostora na trdih diskih dejansko zavzela vsako od podanih vrednosti.

f_k^0 – **relativna frekvenca** nam pove, kako so vrednosti vsot prostora na diskih porazdeljene po posameznih razredih.

F_k – **komulativa absolutnih frekvenc** nam pri izbrani vrednosti vsote prostora na diskih pove, na koliko enotah populacije je spremenljivka zavzela manjšo vrednost.

F_k^0 – **komulativa relativnih frekvenc** nam pove delež enot v populaciji z manjšo vrednostjo vsot prostora na diskih.

Še nekaj formul, s katerimi sem računal:

$y_k = (z_k + s_k) / 2$ - sredina

$f_k^0 = f_k / N$ - relativna frekvenca

$F_k = F_{k-1} + f_{k-1}$ ($k = 2, 3, \dots, r$) - komulativa absolutnih frekvenc

$F_k^0 = F_{k-1}^0 + f_{k-1}^0$ ($k = 2, 3, \dots, r$) - komulativa relativnih frekvenc

4...:: grafična predstavitev ...

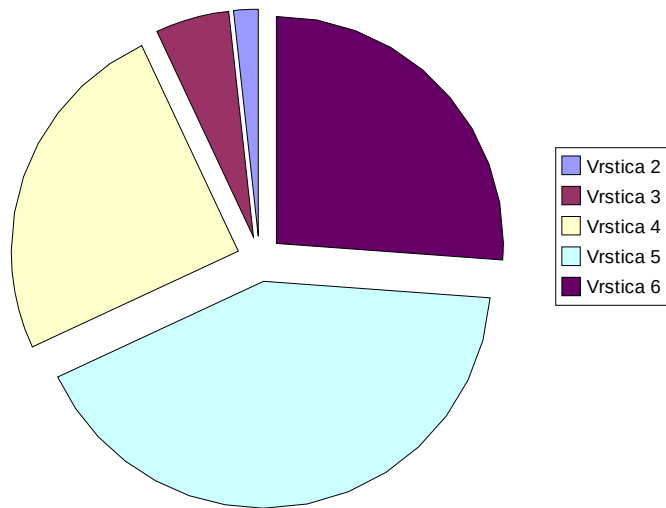
... stolpčni diagram ...



legenda: vrstica 2 : 0
vrstica 3 : 0 - 10 GB
vrstica 4 : 10 - 40 GB
vrstica 5 : 40 – 100 GB
vrstica 6 : nad 100 GB

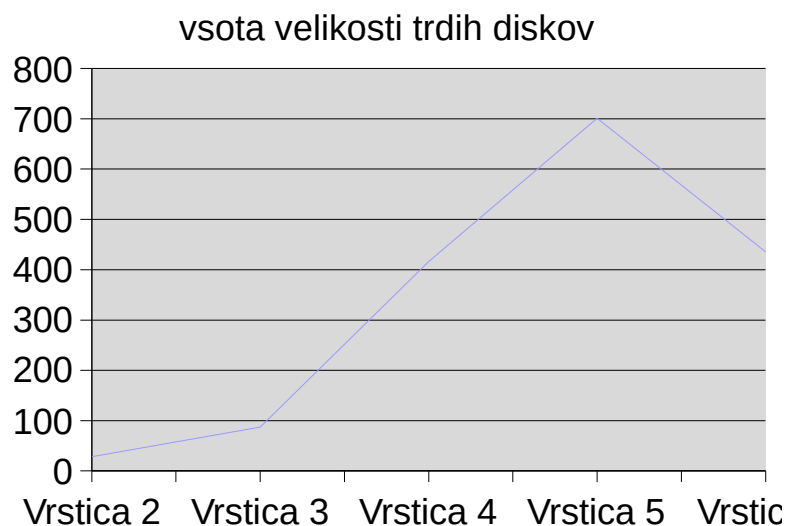
...: tortni diagram ...

vsota velikosti trdih diskov



legenda: vrstica 2 : 0
vrstica 3 : 0 - 10 GB
vrstica 4 : 10 - 40 GB
vrstica 5 : 40 – 100 GB
vrstica 6 : nad 100 GB

...: histogram ...



legenda: vrstica 2 : 0
vrstica 3 : 0 - 10 GB
vrstica 4 : 10 - 40 GB
vrstica 5 : 40 - 100 GB
vrstica 6 : nad 100 GB

5...:: aritmetična sredina, varianca, standardni odklon ::..

Za izračun **aritmetične sredine**, bomo potrebovali produkt absolutnih frekvenc, torej f_k in sredino y_k , ki smo ju zapisali že prej.

$$\mu = \frac{1}{N} \sum_{i=1}^N y_i$$

Razred k	Vsota prostora na diskih – v GB	Sredina (y_k)	f_k	$f_k \cdot y_k$
1	0	0	28	0
2	0 - 10	5	87	435
3	10 - 40	25	416	10400
4	40 - 100	70	701	49070
5	nad 100	/	453	/
		skupaj	1685	59905

Za izračun **variance** bomo potrebovali produkt absolutne frekvence f_k in kvadrat sredine y_k .

σ^2 – **varianca** je merilo za razpršenost oz. variabilnost podatkov

Razred k	Vsota prostora na diskih – v GB	Sredina (y_k)	f_k	$f_k \cdot y_k^2$
1	0	0	28	0
2	0 - 10	5	87	37845
3	10 - 40	25	416	4326400
4	40 - 100	70	701	34398070
5	nad 100	/	453	/
		skupaj	1685	38762315

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2$$

Po tej formuli izračunamo varianco.

Pridemo do rešitve: 21740,53875

σ – **standardni odklon** ali **standardna deviacija** je kvadratni koren variance in ima to lastnost, da ima enako mersko enoto kot jo imajo izhodiščni podatki.

$$\text{Sqrt}(\sigma^2)$$

$$= \text{sqrt}(21740,53875) = 147,4477492$$

6....:: zaključek ::...

Za statistiko je potrebno dosti zbranosti in predvsem urejenost podatkov. Dopustiti si ne smemo napak, kar se pa hitro lahko zgodi.