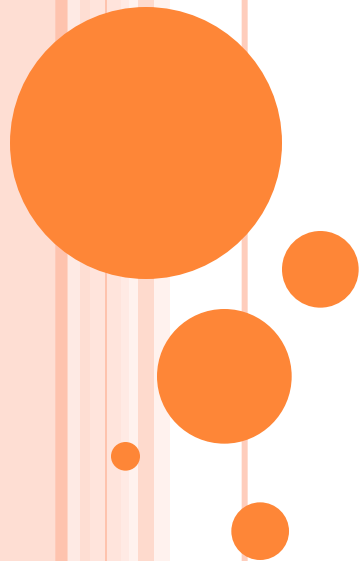


# STATISTIKA

Doc.dr. Tadeja Kraner Šumenjak





- **Literatura**

Jože Nemeč: Statistika

Katarina Košmelj: Uporabna statistika

- **Obveznosti**

**IZPIT LAHKO OPRAVIM S  
SPROTNIM DELOM**

Dva kolokvija (računanje, teorija in SPSS).

Povprečje obeh kolokvijev mora biti vsaj 50 %.

Opravljene vaje (pozitivna ocena).

Končna ocena je 75 % ocene pisnega izpita in 25 % ocene iz vaj.

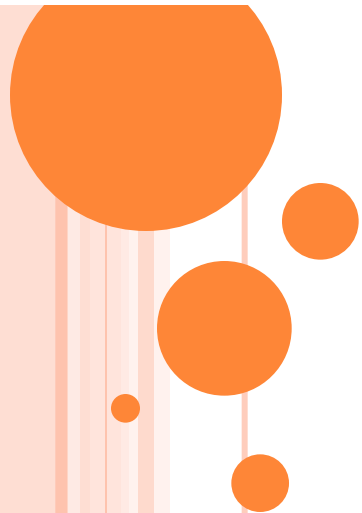
**REDNI IZPITNI ROKI**

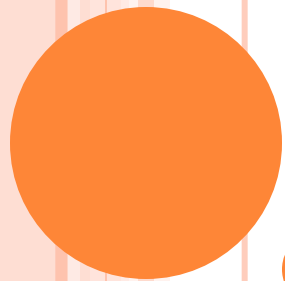
Pogoj za pristop k ustnemu izpitu je uspešno opravljen pisni izpit.

**pisno: ustno=50:50**

# PROGRAMSKI PAKETI

- SPSS
- EXCEL
- STATISTIX
- STATGRAPH...





# UVOD

Kaj je statistika?

Osnovni pojmi.



# KAJ JE STATISTIKA?

- **Statistika** je veda, ki proučuje **pojave**, ki se kažejo v velikem številu v določenem času in prostoru. Rečemo jim **množični pojavi**.
- Statistika je veda, ki se po eni strani ukvarja z zbiranjem, urejanjem, analiziranjem množičnih podatkov, ki opisujejo proučevani pojav. Po drugi strani pa jemlje vzorce podatkov iz populacije in na osnovi teh sklepa o populaciji. Včasih beseda statistika lahko pomeni tudi same podatke.
- Je interdisciplinarna veda. Je del mnogih naravoslovnih in družboslovnih ved. Npr. del bioloških in medicinskih ved je **BIOSTATISTIKA**.



## OPIŠNA STATISTIKA

- Zbiranje, urejanje in analiziranje podatkov.
- Izdelava tabel, grafikonov.
- Namen opisne statistike je na osnovi izračunanih podatkov okarakterizirati večjo količino podatkov.

## INFERENČNA STATISTIKA

- Statistično sklepanje iz vzorca na populacijo.
- Ocenjevanje parametrov.
- Testiranje hipotez.



- Primer **množičnega pojava** je študij na Univerzi v Mariboru. Statistično proučevanje bi lahko bilo usmerjeno v proučevanje uspešnosti brucev.
- Cilj statistike je nepregledno množico podatkov predelati v manjše skupine informacij in sicer tako, da bi se pri tem čim manj podatkov izgubilo.



# OSNOVNI POJMI

- **Statistična populacija** je podana s tremi opredelitvami:
  - **stvarna opredelitev**: koga oz. kaj proučujemo?
  - **krajevna opredelitev**: kje?
  - **časovna**: kdaj?
  - Oznaka  $N$  označuje število enot v populaciji.
- Primer populacije so: študenti vpisani v 1. letnik FKBV v šolskem letu 2010/11.
- Populacijo sestavljajo **enote**. Del populacije imenujemo **vzorec**.





- Za statistiko je vzorec del populacije, katerega enote izbiramo z namenom, da ocenimo stanje v populaciji.
- Vzorec, ki dobro posreduje lastnosti celotne populacije, imenujemo **represzentativen vzorec**. Oznaka  $n$  označuje število enot v vzorcu ( $n < N$ ).
- Več boste o vzorčenju slišali kasneje.



- **Statistična spremenljivke** so lastnosti enot, ki jih proučujemo.
- Npr. pri proučevanju uspešnosti brucev nas zanima:
  - leto rojstva,
  - kraj rojstva,
  - spol,
  - smer študija,
  - opravljena srednja šola,...



- Vsaka spremenljivka ima določene vrednosti. Glede na to, kako se vrednosti spremenljivke izražajo ločimo:
- **OPIŠNE** (atributivne, kvalitativne)  
spremenljivke: vrednosti se izražajo z opisi (besede, znaki).
- **ŠTEVILSKE** (numerične, kvantitativne)  
spremenljivke: vrednosti se izražajo s števili, s katerimi lahko računamo.



- **PRIMER:**

- Opisne spremenljivke so: spol, kraj rojstva, smer študija, izobrazba očeta.
- Številске spremenljivke so: leto rojstva, število otrok v družini, uspeh v osnovni šoli.



- Numerične spremenljivke delimo v dve skupini:
- **DISKRETNE:** Za diskretno spremenljivko velja, da zavzame le določene vrednosti. Njene vrednosti so pogosto cela števila. Npr. število otrok v družini, število opravljenih izpitov.
- **ZVEZNE:** Spremenljivka je zvezna, če med poljubnima dvema vrednostima na njenem definicijskem območju vedno obstaja tretja vrednost. Npr. starost, višina, stopnja inflacije...



- Za statistično spremenljivko določimo njeno **mersko lestvico**. Le ta je opredeljena glede na urejenost vrednosti spremenljivk:
  - ❖ V vrednostih ni nobene urejenosti (smer študija, kraj rojstva, barva oči) → **IMENSKA ALI NOMINALNA SPREMENLJIVKA**.
  - ❖ Vrednosti lahko uredimo v smiselni vrstni red (izobrazba) → **UREJENOSTNA ALI ORDINALNA**.
  - ❖ Vrednosti lahko odštevamo, ne pa delimo (temperatura, leto rojstva) → **RAZMIČNA ALI INTERVALNA**.
  - ❖ Vrednosti lahko delimo (masa, starost) → **RAZMERNOSTNA** .



- Temperatura ni razmernostna spremenljivka, kajti ne moremo reči, da je pri temperaturi 20 stopinj celzija petkrat topleje kot pri temperaturi 4 stopinje celzija. Razmernostne spremenljivke so pogoste na naravoslovnih področjih.
- V anketah pa pogosto srečamo ordinalne spremenljivke. Pogosta je 5 stopenjska ordinalna lestvica.

PRIMER: Ali ste zadovoljni s študijem?

(zelo nezadovoljen, nezadovoljen, niti zadovoljen-niti nezadovoljen, zadovoljen, zelo zadovoljen)



- **Zbiranje podatkov in viri:** ločimo **primarne** in **sekundarne** podatke. Prve je potrebno šele pridobiti, druga pa so že zbrani.
  
- **Napake** pri numeričnem izračunavanju statističnih parametrov:
  - a) zaradi napačnih podatkov
  - b) zaradi zaokrožitvenih napak
  - c) zaradi napačne metode





- Npr. masa bika je odvisna od natančnosti tehtnice, točnosti odčitavanja, obnašanja bika na tehtnici in drugih.
- Če med računanjem površno zaokrožujemo vmesne rezultate, lahko dobimo zelo netočen rezultat.
- Tretja vrsta napak, ki nastane zaradi izbire napačne metode je običajno posledica neizkušenosti ali pa neznanja.



- **Statistična raziskava in njen potek:**  
planiranje, zbiranje podatkov, priprava podatkov,  
njihova analiza in interpretacija rezultatov.



# OPISNA STATISTIKA

## RELATIVNA ŠTEVILA

Strukture.

Koeficienti.

Indeksi.

## FREKVENČNA PORAZDELITEV

Razvrščanje števil v razrede.

Grafično prikazovanje frekvenčnih porazdelitev.

Kumulativne frekvenčne porazdelitve.

Kvantili.

Mere centralne tendence.

Mere variabilnosti.

Relativne mere variabilnosti.

Mere asimetrije in sploščenosti.

# RELATIVNA ŠTEVILA

- Podatki postanejo zanimivi, ko jih med seboj primerjamo. **Relativna števila** primerjajo dva podatka z njunim razmerjem. Glede na to, v kakšni vsebinski povezavi sta podatka, ki ju primerjamo, imamo tri vrste relativnih števil:

- ➔ **strukture**
- ➔ **koeficiente**
- ➔ **indekse**



# STRUKTURE

- Celota je razdeljena v  $K$  skupin. Naj  $f_i$  določa število enot v  $i$ -ti skupini,  $i=1,2,3,\dots,K$ , to število imenujemo **frekvenca**. Pri izračunu **strukture** primerjamo število enot, ki jih ima posamični del celote, s številom enot, ki jih ima celota. Strukture izražamo v **deležih** ali **odstotkih**.
- **DELEŽ** enot v  $i$ -ti skupini je:

$$f_i^0 = \frac{f_i}{\sum_{i=1}^K f_i}$$



- **ODSTOTEK** enot v  $i$ -ti skupini je:

$$f_i \% = \frac{f_i}{\sum_{i=1}^K f_i} \cdot 100$$

Deleži so med 0 in 1. Vsota deležev je 1. Odstotki so med 0 in 100. Vsota odstotkov je 100. V rezultatih navadno navajamo odstotke zaokrožene na eno decimalko.

Frekvence izražene v deležih ali odstotkih imenujemo **RELATIVNE FREKVENCE**.



- Da pojasnimo, kaj določa skupine pri izračunu strukture, govorimo o strukturi po spolu, struktura po smeri študija.
- Za grafični prikaz strukture uporabljamo **strukturni stolpec** in **strukturni krog**. Eno strukturo prikažemo s strukturnim stolpcem ali krogom.
- Zahtevnejše, toda informacijsko veliko bogatejše, so večrazsežne strukture, ki so predstavljene v večrazsežnih ali kombinacijskih tabelah. Poseben primer večrazsežne tabele dveh opisnih spremenljivk je **kontingenčna tabela**. V primeru dveh spremenljivk in več svetujemo prikaz s stolpci.



- KOT, KI USTREZA POSAMEZNEMU STRUKTURNEMU DELEŽU

$$\varphi_k = \frac{360 \cdot d_k}{\sum d_k}$$

$\varphi_k$  kot  $k$  - tega deleža

$d_k$   $k$  - ti delež





## PRIMER

V tabeli so podatki za število študentov po smeri študija in po spolu .

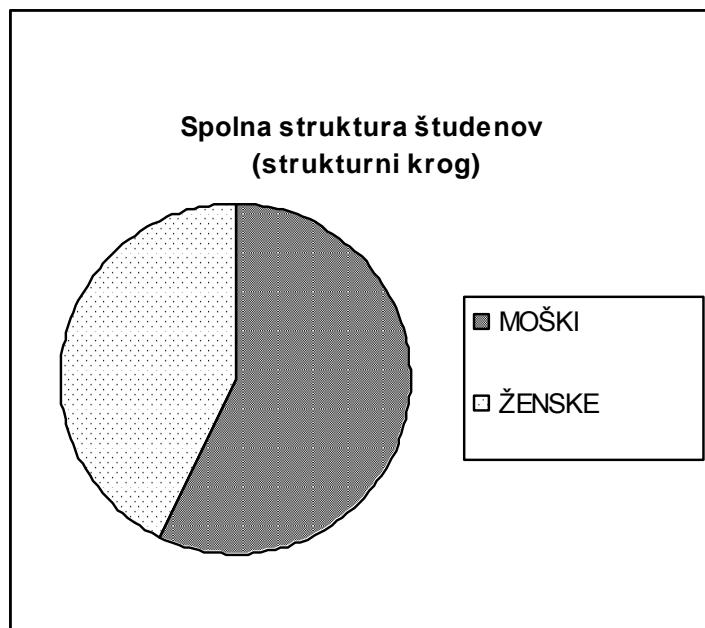
<b>SMER ŠTUDIJA</b>	<b>MOŠKI ŽENSKE</b>	<b>SKUPAJ</b>
<b>Agronomija (AG)</b>	<b>103 117</b>	<b>220</b>
<b>Biologija (BI)</b>	<b>26 62</b>	<b>88</b>
<b>Gozdarstvo (GO)</b>	<b>98 24</b>	<b>122</b>
<b>Krajinska arhitektura (KA)</b>	<b>11 25</b>	<b>36</b>
<b>Lesarstvo (LE)</b>	<b>188 21</b>	<b>209</b>
<b>Mikrobiologija (MI)</b>	<b>19 47</b>	<b>66</b>
<b>Zootehnika (ZO)</b>	<b>63 63</b>	<b>126</b>
<b>Živilska tehnologija (ŽT)</b>	<b>34 46</b>	<b>80</b>
<b>SKUPAJ</b>	<b>542 405</b>	<b>947</b>



- Prikaži spolno strukturo študentov s strukturnim stolpcem in strukturnim krogom.
- Prikaži strukturo stanja po smereh študija s strukturnim stolpcem.
- Prikaži strukturo stanja po smereh študija s strukturnim krogom.
- Prikaži spolno strukturo študentov po posameznih študijskih smereh.



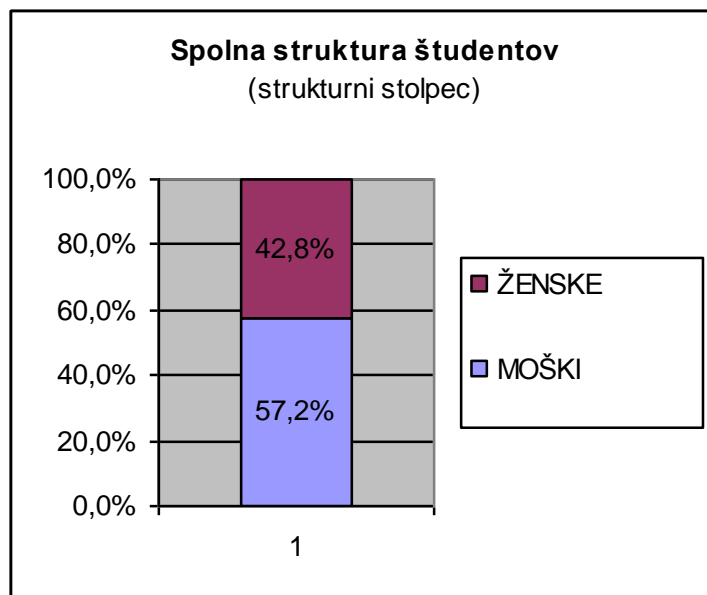
# SPOLNA STRUKTURA ŠTUDENTOV S STRUKTURNIM KROGOM



	število	kot krožnega izseka
moški		
ženske		



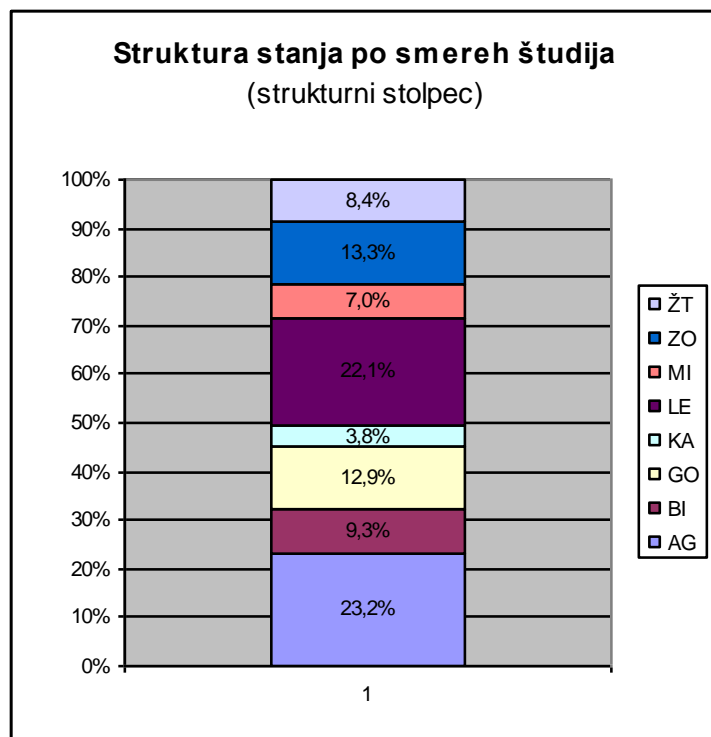
# SPOLNA STRUKTURA ŠTUDENTOV S STRUKTURNIM STOLPCEM



	število	odstotek
moški		
ženske		



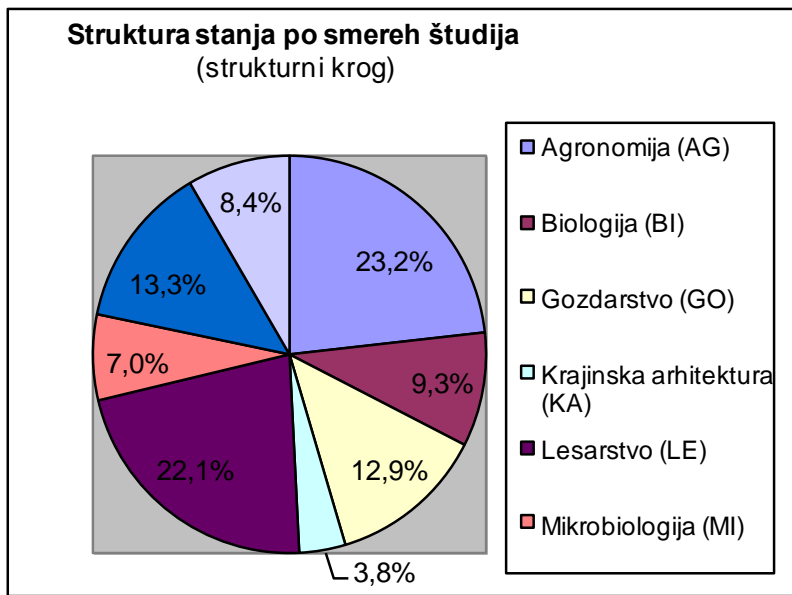
# STRUKTURA STANJA PO SMEREH ŠTUDIJA S STRUKTURNIM STOLPCEM



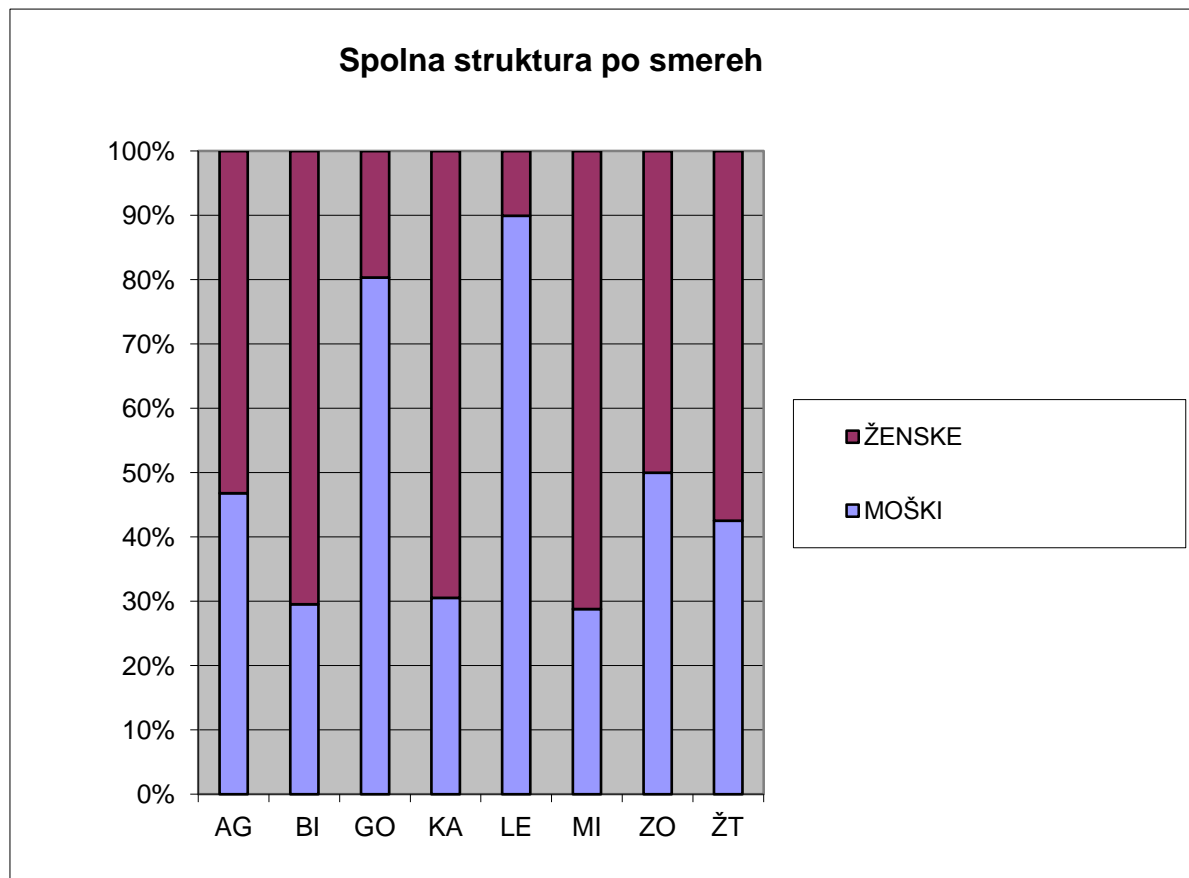
	Štev.	Odst.
ŽT		
ZO		
MI		
LE		
KA		
GO		
BI		
AG		



# STRUKTURA STANJA PO SMEREH ŠTUDIJA S STRUKTURNIM KROGOM



# SPOLNA STRUKTURA PO SMEREH ŠTUDIJA S STRUKTURNIMI STOLPCI



Spodaj je večrazsežna tabela porazdelitve učencev po barvi las in po spolu.

	rjava	črna	rdeča	blond
moški	3	28	9	3
ženske	14	20	15	8
skupaj	17	48	24	11

- Zapišite spolno strukturo po barvi las (spol v legendi).
- Strukturo barve las učencev po spolu (barva las v legendi).
- Obe porazdelitvi prikažite še grafično.





# KOEFICIENTI

- Koeficient je razmerje dveh podatkov, ki sta vsebinsko povezana. Pogosto uporabljeni koeficienti so:
  - gostota prebivalstva
  - število bolnikov na zdravnika
  - poraba goriva na 100 km
  - pridelek na ha
  - letna stopnja rodnosti
  - letna stopnja smrtnosti



**ČASOVNA VRSTA** je niz istovrstnih pojavov, ki se nanašajo na zaporedne časovne razmike ali trenutke.

**PRIMERI:**

število padavin po mesecih,

število konj po letih,

število govedi po letih,

inflacija po mesecih.



# INDEKSI

## ○ Indeksi stalno osnovo

Neki podatek v časovni vrsti izberemo za osnovo označimo ga z  $Y_0$ . Vse podatke primerjamo na ta podatek. Za osnovo ne izbiramo neobičajnih obdobj.

$$I_j = \frac{Y_j}{Y_0} \cdot 100$$

$Y_j$  podatek za  $j$ -to leto

$I_j$  indeks s stalno osnovo za  $j$ -to leto



## ○ Verižni indeks

Če za osnovo vzamemo predhodni podatek v časovni vrsti, dobimo **verižne indekse**.

$$V_j = \frac{Y_j}{Y_{j-1}} \cdot 100$$

$V_j$  verižni indeks za  $j$  - to leto

$Y_j$  podatek za  $j$  - to leto

$Y_{j-1}$  podatek za  $(j - 1)$  leto



- Pri interpretaciji verižnih indeksov uporabljamo **stopnja rasti**. Dobimo jo tako, da od verižnega indeksa odštejemo 100:

$$S_j = V_j - 100$$

Stopnja rasti je lahko pozitivna, negativna ali ničelna.



- **Preračunavanje verižnih indeksov v indekse s stalno osnovo (od baznega leta naprej)**

$$I_j = \frac{V_j \cdot I_{j-1}}{100}$$

$V_j$  verižni indeks za  $j$  - to leto

$I_j$  indeks s stalno osnovo za  $j$  - to leto

$I_{j-1}$  indeks s stalno osnovo za  $(j-1)$  leto



- **Preračunavanje verižnih indeksov v indekse s stalno osnovo (pred baznim letom)**

$$I_{j-1} = \frac{I_j}{V_j} \cdot 100$$

$V_j$  verižni indeks za  $j$  - to leto

$I_j$  indeks s stalno osnovo za  $j$  - to leto

$I_{j-1}$  indeks s stalno osnovo za  $(j-1)$  leto



- Denimo, da nas zanima velikost indeksa, ki bi bil v vseh letih enak in bi proizvodnja iz začetnega leta narasla enakomerno do vrednosti v končnem letu. Temu indeksu **pravimo povprečni indeks**.
- Povprečni indeks ali povprečni koeficient dinamike je geometrijska sredina pripadajočih verižnih indeksov ( $k$  je število vseh verižnih indeksov):

$$I = \sqrt[k]{V_1 V_2 \dots V_k}$$

- Če imamo podatke o časovni vrsti, je boljša pot za izračun povprečnega indeksa formula, ki upošteva prvi in zadnji podatek ter dolžino časovne vrste, ki je  $k$ :

$$I = 100 \cdot \sqrt[k-1]{\frac{Y_k}{Y_1}}$$





V tabeli prikazujemo število brezposelnih v Sloveniji po letih v obdobju 1984-1993.

- Izračunajte in grafično prikažite indekse s stalno osnovo 1984 za število brezposelnih.
- Izračunajte in grafično prikažite verižne indekse.

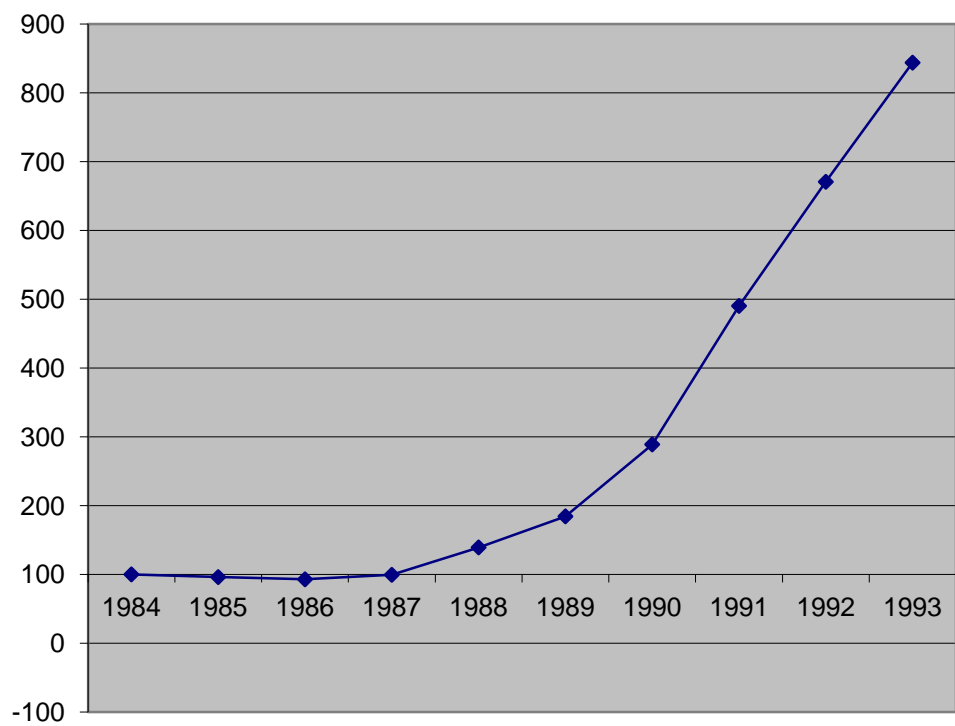
leto	Število brezposelnih
1984	15,3
1985	14,7
1986	14,2
1987	15,2
1988	21,3
1989	28,2
1990	44,2
1991	75,0
1992	102,6
1993	129,1



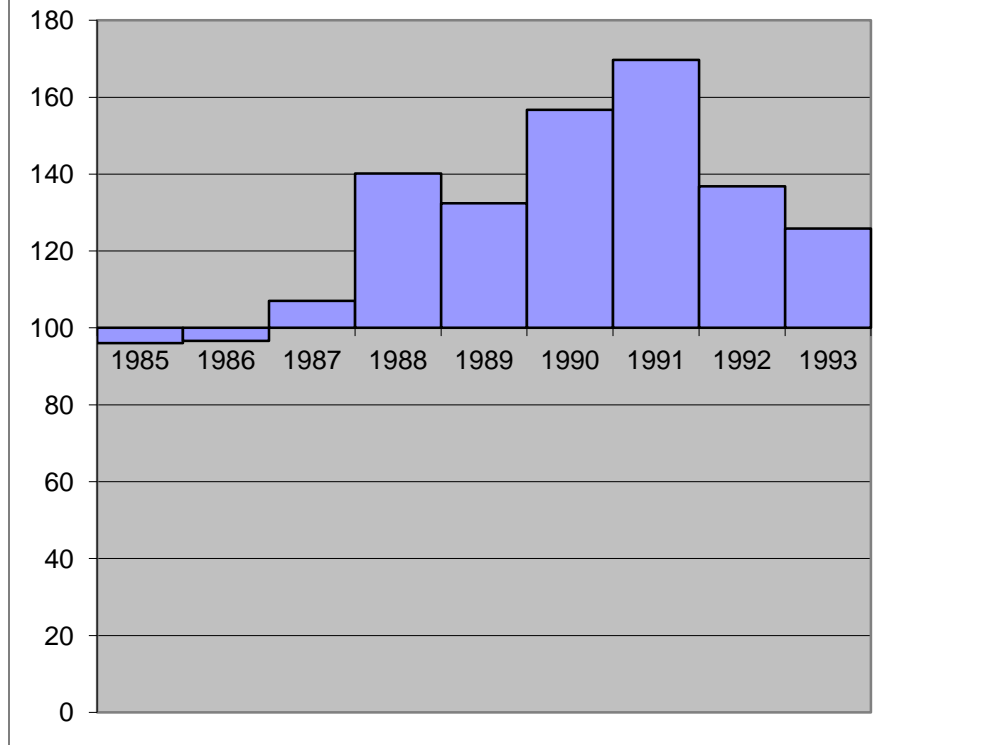
leto	Število brezposelnih	I84	Vk
1984	15,3	100	
1985	14,7	96,07843	96,07843
1986	14,2	92,81046	96,59864
1987	15,2	99,34641	107,0423
1988	21,3	139,2157	140,1316
1989	28,2	184,3137	132,3944
1990	44,2	288,8889	156,7376
1991	75	490,1961	169,6833
1992	102,6	670,5882	136,8
1993	129,1	843,7908	125,8285



**Indeksi s stalno osnovo (1984)**



### Verižni indeksi



V tabeli so podani verižni indeksi za pridelek vina. Izračunajte indeks s stalno osnovo za leto 1994 in povprečni indeks.

<b>Leto</b>	1992	1993	1994	1995	1996
$V_k$	-	117,2	113,6	78,8	146,6
$I_{1994}$					

