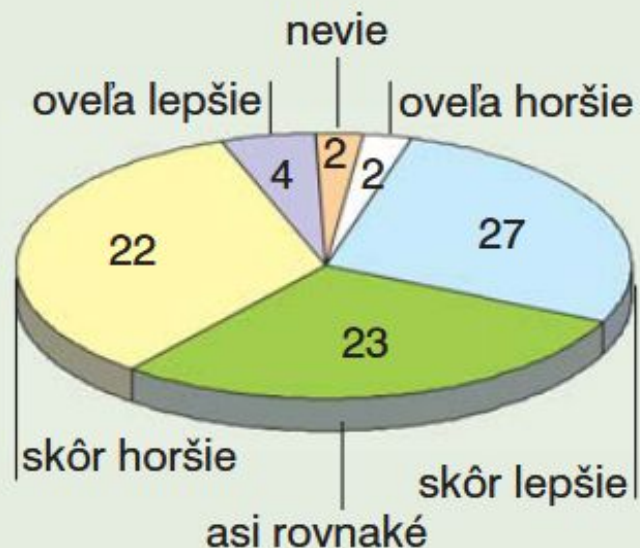


PaŠ pre informatikov (cv 01)

Beáta Stehlíková, FMFI UK Bratislava

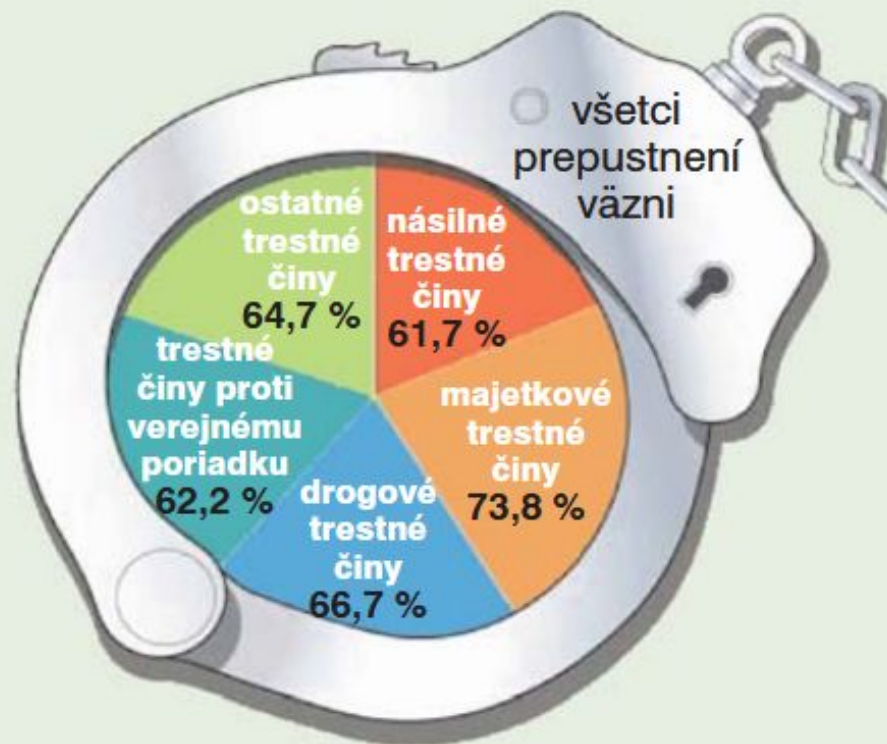
- 15.** Na internete existuje viacero stránok, ktoré komentujú nesprávne použitie grafov v médiách a odbornej literatúre. Našli sme tam aj nasledujúce tri koláčové diagramy. Diskutujte o tom, kvôli akým chybám sa tam dostali. Navrhnite tiež, ako tieto chyby napraviť – ak je náprava možná.

a) Graf znázorňuje percentuálne zastúpenie jednotlivých odpovedí na otázku: *Aké sú vaše očakávania v nasledujúcom roku v oblasti životnej úrovne?* položenú v rámci prieskumu verejnej mienky.



b) Graf znázorňuje výsledky štúdie, ktorá v 15 štátoch USA zisťovala podiel recidivistov medzi väzňami. Štúdia skúmala, koľko z 272 111 väzňov prepustených z väzenia v r. 1994 bolo do troch rokov znova odsúdených do väzenia za nové trestné činy. Počty percent sú uvádzané pre jednotlivé skupiny väzňov podľa trestného činu, za ktorý boli odsúdení.

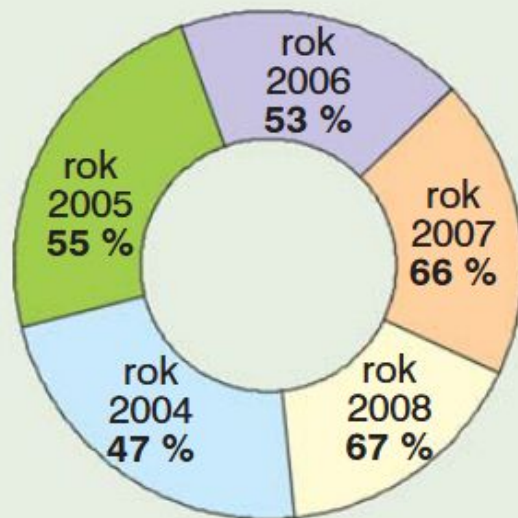
Podiel recidivistov medzi väzňami prepustenými v r. 1994



Rozsah súboru:

272 111 väzňov v 15 štátoch USA

c) Graf znázorňuje počet percent obyvateľov Slovenska (zistený každoročným prieskumom verejnej mienky), ktorí v danom roku dôverovali EÚ.



Tatiana Kossacká: Pravdepodobnosť a štatistika pre informatikov, 2024

Úloha 6 Herné pole hazardnej hry ruleta pozostáva z 37 políčok, ktoré sú rozdelené do troch farieb: červená, čierna a zelená. Každé políčko má presne jedno číslo zo množiny $\{0, 1, 2, \dots, 36\}$ a platí:

- Čísla na červených políčkach $R = \{1, 3, 5, 7, 9, 12, 14, 16, 18, 19, 21, 23, 25, 27, 30, 32, 34, 36\}$.
- Čísla na čiernych políčkach $B = \{2, 4, 6, 8, 10, 11, 13, 15, 17, 20, 22, 24, 26, 28, 29, 31, 33, 35\}$.
- Čísla na zelených políčkach $G = \{0\}$.
- Párne čísla $E = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36\}$.

Existuje možnosť stavať peniaze na červenú, čiernu, párne, nepárne čísla, jednotlivé čísla atď. Počas hry stavíte rovnakú sumu na **červenú a párne čísla**. To znamená, že ak loptička pristane na čiernom nepárnom čísle alebo na nule, prehrávate (nula sa v rulete nepovažuje ani za párne, ani za nepárne číslo). Zjednodušene platí, že ak vyhráte, dostanete stávkku späť dvojnásobne. Určte kombináciu množín pre nasledujúce udalosti a vypíšte príslušné čísla.

- a) Prehráte celú stávkku.
- b) Nemáte žiadny zisk.
- c) Nemáte žiadnu stratu.
- d) Nemáte ani zisk, ani stratu.
- e) Získate výhru.

```
import random

mince = random.choices(["H", "Z"], k=10)
print(mince)
```

```
['H', 'H', 'H', 'Z', 'H', 'H', 'H', 'H', 'H', 'Z']
```

```
def simulacia_longest_run(pocet_minci):  
    mince = random.choices(["H", "Z"], k=pocet_minci)  
    return longest_run(mince)  
  
simulacie1 = [simulacia_longest_run(10) for _ in range(10**5)]
```

```
import pandas as pd  
  
pocetnosti1 = pd.Series(simulacie1).value_counts().sort_index()  
print(pocetnosti1)
```

```
1      201  
2    17020  
3    36161  
4    24716  
5    12499  
6     5557  
7     2285  
8      932  
9     417  
10     212  
Name: count, dtype: int64
```

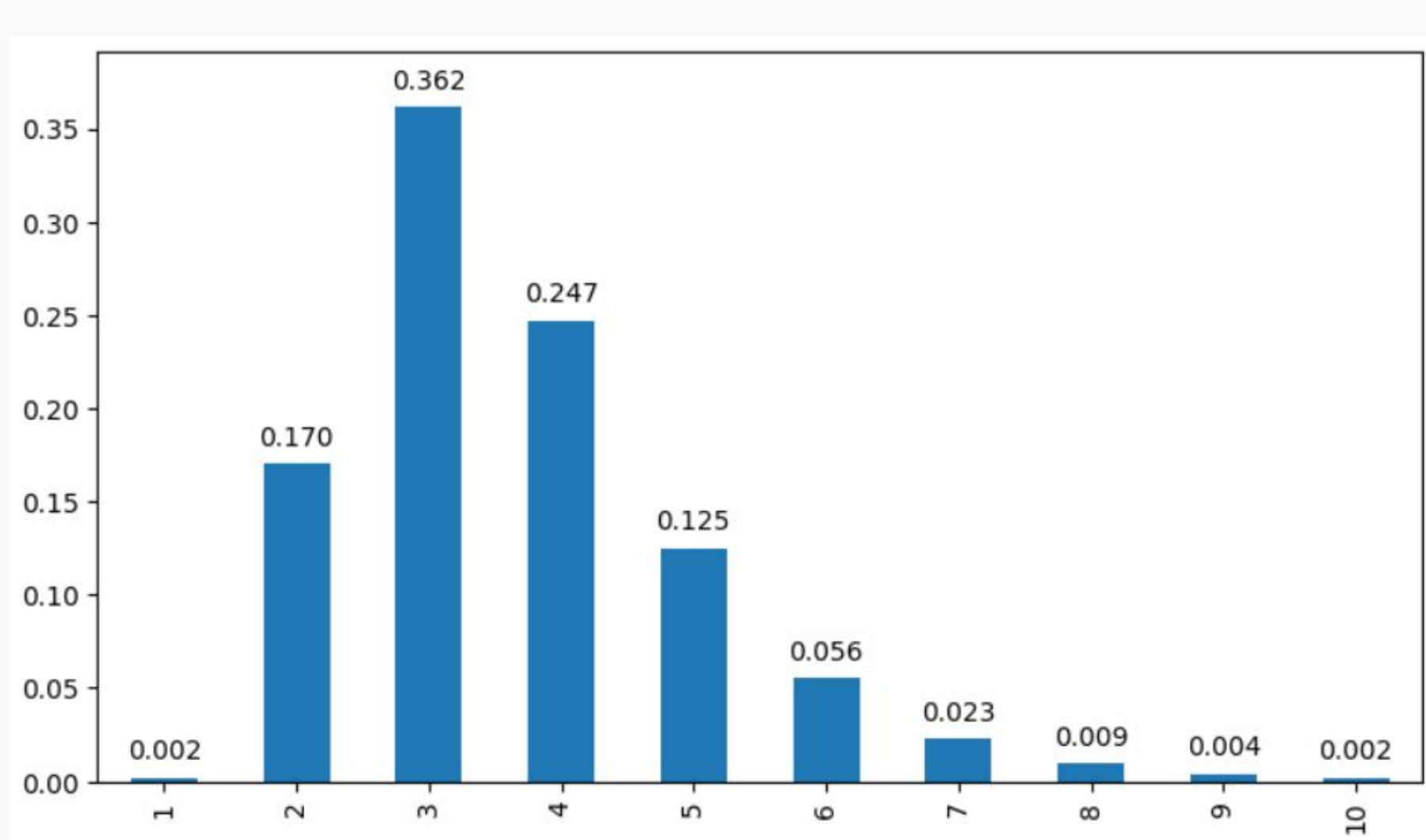
```
sum([n >= 5 for n in simulacie1])/len(simulacie1)
```

```
0.21902
```

```
rel_pocetnosti1 = pd.Series(simulacie1).value_counts(normalize=True).sort_index()  
print(rel_pocetnosti1)
```

```
1      0.00201  
2      0.17020  
3      0.36161  
4      0.24716  
5      0.12499  
6      0.05557  
7      0.02285  
8      0.00932  
9      0.00417  
10     0.00212
```

```
Name: proportion, dtype: float64
```



```
print(rel_pocetnosti1)
```

1	0.00201
2	0.17020
3	0.36161
4	0.24716
5	0.12499
6	0.05557
7	0.02285
8	0.00932
9	0.00417
10	0.00212

Name: proportion, dtype: float64

```
import numpy as np
```

```
print(np.cumsum(rel_pocetnosti1))
```

1	0.00201
2	0.17221
3	0.53382
4	0.78098
5	0.90597
6	0.96154
7	0.98439
8	0.99371
9	0.99788
10	1.00000

Name: proportion, dtype: float64

```
import numpy as np
```

```
print(np.mean(simulacie1))
```

```
print(np.median(simulacie1))
```

```
3.66749
```

```
3.0
```

```
print(np.quantile(simulacie1, [0.25, 0.5, 0.75]))
```

```
[3. 3. 4.]
```

```
print(np.quantile(simulacie1, 0.95))
```

```
6.0
```

M. Schilling, The longest run of heads, *College Mathematics Journal* 21:3 (1990) 196–207.

The two sequences shown below each purportedly represent the results of 200 tosses of a fair coin. One of these is an actual sequence obtained from coin tossing, while the other sequence is artificial. Can you decide, in sixty seconds or less, which of the sequences is more likely to have arisen from actual coin tossing and which one is the imposter?

Sequence #1

T H H H H T T T T H H H H T H H H H H H H H T T T H H T T H H H H H T T T T T T H H T H H T H H H T
T T H T T H H H H T H T T T H T T T H H T T T T H H H H H H T T T H H T T H H H T H H H H H T T T T
T H T T T H H T T H T T H H T T T H H T T T H H T H H T T T T T H H T H H H H H H T H T H T T
H T H T T H H H T T H H T H T H H H H H H H H T T H T T H H H T H H T T H T T T T T T H H H T H H H

Sequence #2

T H T H T T T H T T T T T H T H T T T H T T H H H T H H T H T H T H T T T T H H T T H H T T H H H T
H H H T T H H H T T T H H H T H H H H T T T H T H T H H H H T H T T T H H H T H H T H T T T H H T H
H H T H H H H T T H T H H T H H H T T T H T H H H T H H T T T H H H T T T T H H H T H T H H H H T H
T T H H T T T T H T H T H T T H T H H T T H T T T H T T T T H H H H T H T H H H T T H H H H H T H H

The above challenge is based on a classroom experiment originally performed by Révész [14]. The class is divided into two groups. In the first group, each student is instructed to toss a coin 200 times and record the resulting sequence of heads and tails. Each student in the second group is merely to write down a sequence of heads and tails that the student believes is a reasonable *simulation* of 200 tosses of a fair coin. Given the combined results of the two groups, Révész claims that the students can be classified back into their original groups with a surprising degree of accuracy by means of a very simple criterion: In students' simulated patterns, the longest run of consecutive heads or consecutive tails is almost invariably *too short* relative to that which tends to arise from actual coin tossing.

The real coin tossing sequence above is #1, which has a longest run of eight heads (twice), while the longest run found in Sequence #2 is only five heads long.

