

# Histogram a jeho príbuzní

## Stĺpcový a koláčový diagram

Agentúrna správa<sup>1</sup> hovorí:

*Počas ôsmich mesiacov tohto roku (2007) sa tak zaregistrovalo 53 578 áut ...*

*Poradie desiatich najžiadanejších spomedzi všetkých značiek nových osobných a malých nákladných automobilov dodávaných na Slovensko vedie za osem mesiacov Škoda s trhovým podielom 22 percent. Nasleduje Peugeot, ktorý dosiahol viac než sedem percent, a tretí je Citroën s výsledkom taktiež viac než sedem percent na trhu. Aj ďalšie poradie je tesné: 4. Volkswagen (7,2 percenta), 5. Renault (5,5), 6. Suzuki (5,5), 7. Ford (4,9), 8. Kia (4,9), 9. Opel (4,5) a prvú desiatku uzatvára kórejský Hyundai s podielom 4,3 percenta.*

Zostavme na základe týchto údajov tabuľku početností a relatívnych početností pre 10 najpredávanejších značiek. Keďže plánujeme tieto údaje neskôr aj kresliť, hneď si ich pripravíme v Matlabe.

*Poradie značiek – por; Názvy značiek – nz; Percentuálny podiel – pp; Počet predaných kusov jednotlivých značiek – pk:*

```
» por=1:10;  
» nz={'Skoda', 'Peugeot', 'Citroen', 'VW', 'Renault', 'Suzuki', 'Ford', 'Kia', 'Opel', 'Hyundai'};  
» pp=[0.22, 0.076, 0.074, 0.072, 0.0552, 0.0548, 0.0492, 0.0488, 0.045, 0.043];  
» z=53578; pk=pp*z;
```

Výsledky<sup>2</sup> (por', nz', pk', pp')

| Poradie | Značka    | Počet predaných ks      | Podiel na trhu          |
|---------|-----------|-------------------------|-------------------------|
| 1       | 'Skoda'   | 1.1787160000000000e+004 | 2.2000000000000000e-001 |
| 2       | 'Peugeot' | 4.0719280000000000e+003 | 7.6000000000000000e-002 |
| 3       | 'Citroen' | 3.9647720000000000e+003 | 7.4000000000000000e-002 |
| 4       | 'VW'      | 3.8576160000000000e+003 | 7.2000000000000000e-002 |
| 5       | 'Renault' | 2.9575056000000000e+003 | 5.5200000000000000e-002 |
| 6       | 'Suzuki'  | 2.9360744000000000e+003 | 5.4800000000000000e-002 |
| 7       | 'Ford'    | 2.6360376000000000e+003 | 4.9200000000000000e-002 |
| 8       | 'Kia'     | 2.6146064000000000e+003 | 4.8800000000000000e-002 |
| 9       | 'Opel'    | 2.4110100000000000e+003 | 4.5000000000000000e-002 |
| 10      | 'Hyundai' | 2.3038540000000000e+003 | 4.3000000000000000e-002 |

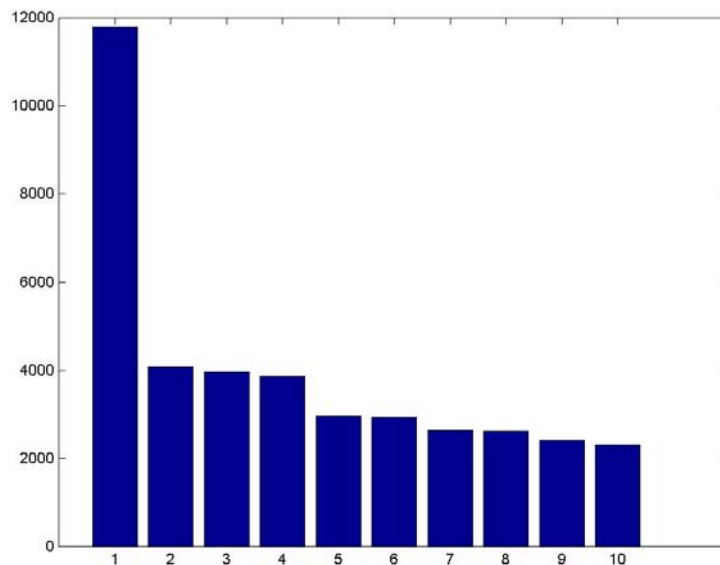
<sup>1</sup> <http://auto.hnonline.sk/c1-22036640-skoda-stale-s-vysokym-podielom-na-trhu>

Uvedená správa slúži len ako ilustratívny príklad číselných hodnôt z praxe. Viaceré hodnoty nie sú uvedené dostatočne presne, preto si ich spresnenie bez ostychu vymyslíme.

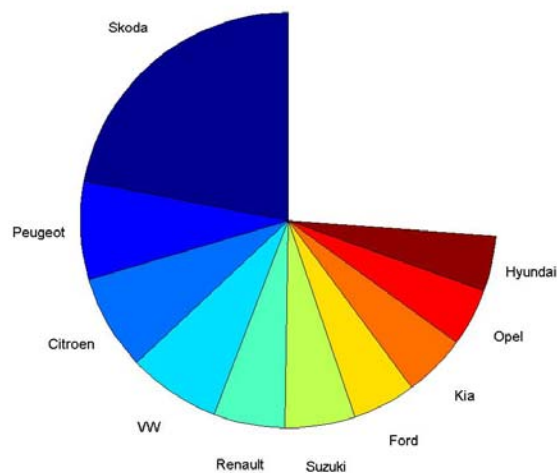
<sup>2</sup> Uvedená tabuľka nie je priamo výstupom z Matlabu, ale výsledky sú upravené vo Worde.

Zobrazíme počty predaných vozidiel stĺpcovým a podiel na trhu koláčovým diagramom:

» `bar(pk)`

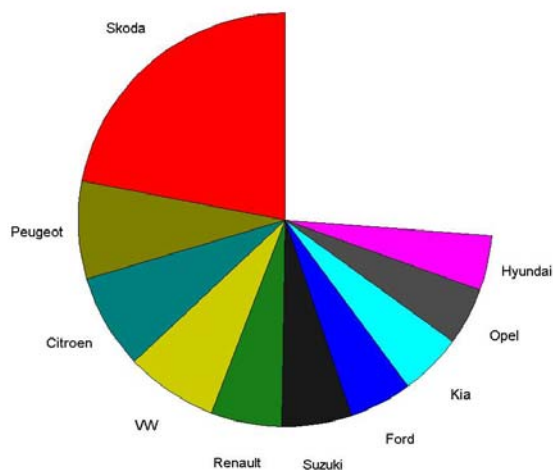


» `pie(pp,nz)`



- Ak nezadáme názvy značiek `nz`, Matlab vypíše percentá. Skúste – `pie(pp)`
- Ak je súčet hodnôt v `pp` menší ako 1, Matlab to považuje za percentá a nakreslí neúplný kruh. Medzi našimi hodnotami chýbajú počty značiek na 11. a vyššom mieste.
- Ak je súčet hodnôt `pp` väčší ako 1, Matlab to chápe ako početnosti, ich súčet považuje za celok a sám si podľa toho vypočíta percentá. Kruh bude úplný. Skúste – `pie(pk, nz)`
- Farby jednotlivých výsekov z kruhu si Matlab volí sám a vidíme, že v tom má pomerne prehľadný systém. Ak chceme mať iné farby, musíme si ich namiešať vektorom `[r g b]`, kde hodnoty `r`, `g`, `b` sú z intervalu `[0 1]` a predstavujú množstvo červenej, zelenej a modrej:

» `farby = [1 0 0; 0.5 0.5 0; 0 0.5 0.5; 0.8 0.8 0; 0.1 0.5 0.1; 0.1 0.1 0.1; 0 0 1; 0 1 1; 0.3 0.3 0.3; 1 0 1];`  
» `pie(pp, nz), colormap(farby)`

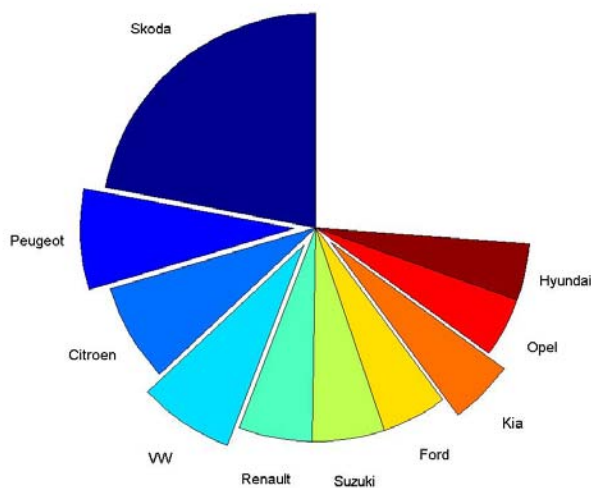


Matlab to predtým spravil nepochybne krajšie, ale hlavne že je po našom...

Skúste namiešať iné farby podľa toho, ako je nalakovaný voz príslušnej značky parkujúci najbližšie k Vášmu bydlisku (táto úloha nebude na skúške).

– Zvýrazníme a „nenápadne“ ponúkneme značky, ktoré sa momentálne (2007) aspoň čiastočne vyrábajú na Slovensku:

» `sk=[0 1 0 1 0 0 0 1 0 0]; pie(pp, sk, nz)`



Podobne skúste „ponúknuť“ značky vyrábané v Česku, Maďarsku, Nemecku, apod.

– Zopakujte kreslenie niektorých predošlých diagramov s použitím príkazu `pie3` (syntax sa nelíši).

## Histogram

Krajina Zxcvbnmsko má 1 933 556 obyvateľov s pravidelnými príjmami. Minimálna mzda – a ani dôchodok pod túto hranicu nesmie klesnúť – je 1200 dg (dg = deravý groš). Najvyššie príjmy sú síce štátnym tajomstvom, ale iste nepresahujú hodnotu 100 000 dg.

Očíslujme si najprv obyvateľov:

```
» x=1:1933556;
```

Ich príjmy  $z$  sú takéto:

```
» t=rand(1,1933556); t=t.^4; z=t*98800+1200;
```

Radi by sme si urobili obraz o rozložení príjmov obyvateľstva. K dispozícii máme zatiaľ príjmy všetkých jednotlivcov, čo sú síce údaje kompletne, ale ťažko sa v nich vyznať. Na lepšiu orientáciu musíme definovať príjmové triedy. Pohľad to bude hrubší, ale rozhodne prehľadnejší.

|                           |                |
|---------------------------|----------------|
| 1. Najnižšie príjmy:      | 1200 – 2000    |
| 2. Slabé príjmy:          | 2000 – 6000    |
| 3. Ako-také príjmy:       | 6000 – 12000   |
| 4. Podpriemerné príjmy:   | 12000 – 16000  |
| 5. Priemerné príjmy:      | 16000 – 22000  |
| 6. Nadpriemerné príjmy:   | 22000 – 30000  |
| 7. Také už lepšie príjmy: | 30000 – 42000  |
| 8. Elitné príjmy:         | 42000 – 100000 |

Hranice tried umiestnime do vektora  $ht$ :

```
» ht=[1200, 2000, 6000, 12000, 16000, 22000, 30000, 42000, 100000];
```

Opýtame sa, koľko ľudí bude v jednotlivých triedach, a odpoveď nazveme  $pt$ . Pomôže nám funkcia `histc` (histogram + count).<sup>3</sup>

```
» pt=histc(z,ht)
```

```
pt = 597437    336887    210360    93654    110271    113954    132669    338324    0
```

Posledný výsledok<sup>4</sup> určuje počet ľudí s príjmom rovných 100000 dg. Vo všeobecnosti to číslo pripočítavame k predposlednej hodnote vektora  $pt$ .

```
» n=length(pt); pt=[pt(1:n-2), pt(n-1)+pt(n)]
```

Početnosť má byť vyjadrená plochou stĺpca, preto musíme zistiť výšky stĺpcov a to delením početností tried ich šírkou:

---

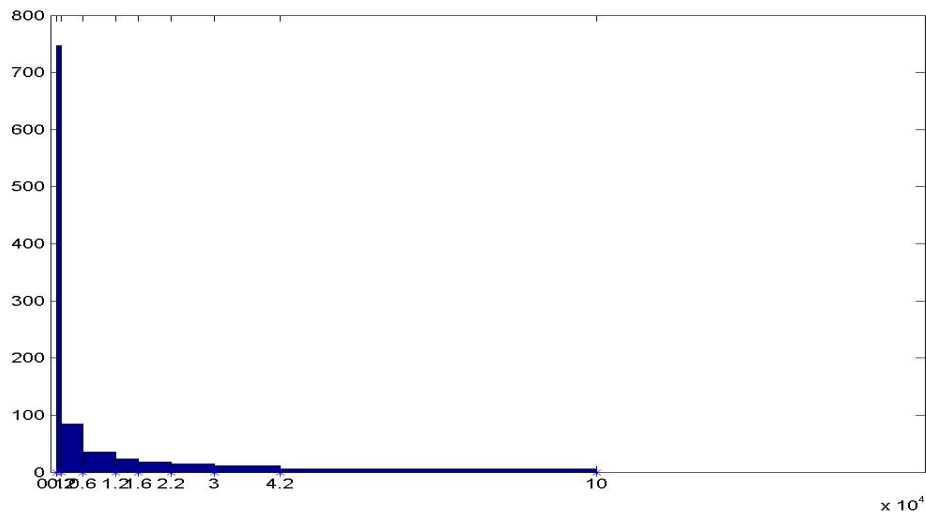
<sup>3</sup> Výpis výsledkov je upravený tak, aby zaberol menej miesta.

<sup>4</sup> Ostatné hodnoty hovoria o počtoch v  $[ht(i), ht(i+1))$ .

```
» ns=length(ht); st=ht(2:ns)-ht(1:ns-1); vt=pt./st;
```

Na samotné kreslenie si matlab vyžaduje vektor *pt* rovnakej dĺžky ako je počet hraničných bodov, teda niečo také, ako bola pôvodná odpoveď na príkaz *histc*. Pridáme tam teda nulu. Na vykreslenie slúži príkaz *bar( . , . , 'histc')*. Posledným parametrom *histc* nám známy príkaz *bar* prepíname do iného režimu a preto sa netreba pozastavovať nad tým, že parametre do neho ukladáme inak ako sme to robili pri bežnom stĺpcovom diagrame:

```
» bar(ht, [vt,0], 'histc')
```



### Poznámky:

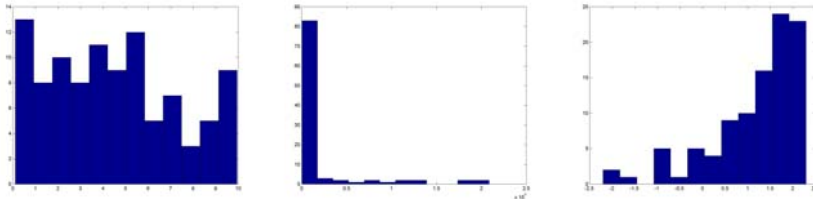
Pozrite tiež:

<http://en.wikipedia.org/wiki/Histogram>

## Dodatok – rýchly histogram:

Ak nenástojíme na tom, aby sme triedy určovali my, a bude nám stačiť, ak Matlab sám rozdelí údaje do  $m$  tried s rovnakou šírkou, môžeme kreslenie histogramu urýchliť príkazom `hist(z,m)`, ktorý hneď vykreslí obrázok. Ak nezadáme číslo  $m$ , tj. príkaz bude len `hist(z)`, Matlab predpokladá  $m=10$ .

```
» x=rand(1,100)*10; hist(x,12)
» y=exp(x); hist(y,12)
» w=log(x); hist(w,12)
```



Porovnajme všetky tri obrázky. Histogram pre  $x$  je rozložený zhruba rovnomerne, pre  $y$  je vychýlený doľava a pre  $w$  doprava. Pozrime sa na hodnoty priemeru a mediánu:

```
» [mean(x), median(x)]
ans = 4.4772 4.1713
» [mean(y), median(y)]
ans = 1894.9 65.6
» [mean(w), median(w)]
ans = 1.1690 1.4275
```

Úloha: Vysvetlite, aký je vzťah medzi nakreslenými histogramami a vzdialenosťou hodnoty mediánu od priemeru.

---

### Rozptyl

```
» var(x,1)
ans = 7.9891
Opravený (1/n-1):
» var(x)
ans = 8.0698
```

### Priemerná odchýlka

```
» d=sum(abs(x-median(x)))/length(x)
d = 2.3845
```

Úloha: Vypočítajte to isté pre  $y$  a  $w$ , ako aj pre  $z$  z predošlej úlohy.