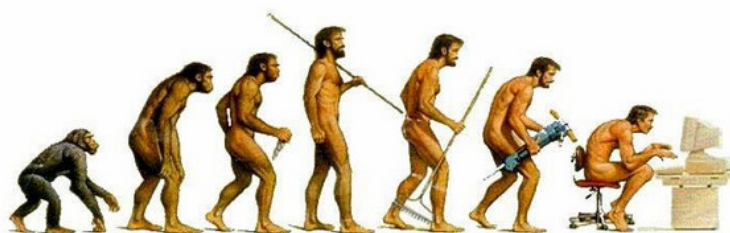


Masarykova univerzita v Brně  
Přírodovědecká fakulta

# APLIKOVANÁ STATISTIKA PRO ANTROPOLOGY

Zadání domácího úkolu – Skupina B






Brno, 2016

## Pokyny k řešení domácího úkolu

Domácí úkol sestává z osmi příkladů. Za vyřešení příkladů lze získat  $20+30+10+15+10+25+15+25=150$  bodů + 10 bodů za celkovou úpravu úkolu, detailní komentáře k postupům, apod. Celkem lze tedy získat 160 bodů.

Aby byl úkol uznán za splněný, je potřeba získat alespoň 120 bodů (75 %). Pokud nebude dodatečně stanoveno jinak, má student na vyřešení domácího úkolu 14 dní počínaje dnem zadání domácího úkolu. Pokud student potřebných 120 bodů nezíská, bude mu úkol navrácen a student dostane jeden týden na dořešení příkladů. Pokud ani potom student kýženého počtu nedosáhne, bude psát na konci semestru prověřovací písemku na látku obsaženou v tomto úkolu.

Řešení domácího úkolu zasílejte, prosím, na e-mail vyučující, tedy [375612@seznam.cz](mailto:375612@seznam.cz) nebo [375612@mail.muni.cz](mailto:375612@mail.muni.cz). Po převzetí úkolu Vám bude zaslán stručný e-mail jako potvrzení, že řešení úkolu v pořádku dorazilo. :)

Jako kompletní řešení se považuje dodání **zcela funkčního** -Skriptu (před odesláním skriptu vyučující si vyčistěte workspace a všechny příkazy si finálně ještě projděte, abyste měli jistotu, že vám vše funguje, jak má) a dodání textového souboru (Word) obsahujícího všechny potřebné komentáře, popisy postupů, apod. Jako alternativu je možné odeslat také pouze funkční -Skript obsahující popisy postupů formou -kových komentářů (za symbolem #). V tom případě není nutné psát komentáře do Wordu.

Domácí úkoly budou opravovány hromadně po obdržení všech (nebo alespoň významné většiny) řešení úkolů, přičemž na opravení úkolů si vyučující vyhrazuje 14 denní lhůtu :).

Přeji vám hodně zdaru při řešení příkladů :).

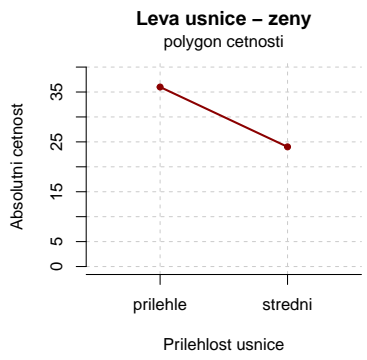
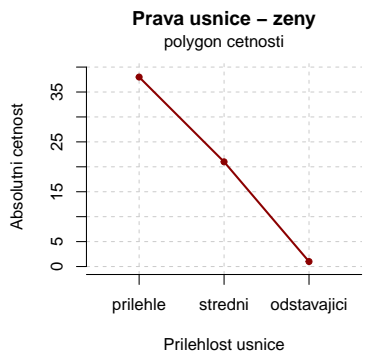
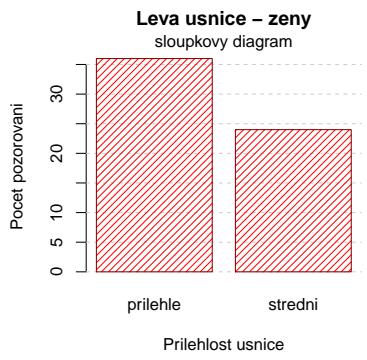
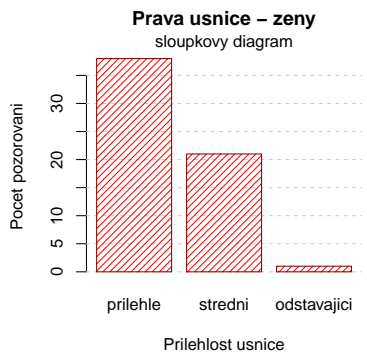


**Příklad 1 (20 b).** Datový soubor 101-usnice.txt obsahuje údaje o přilehlosti levé a pravé ušnice u 120 studentů jisté vysoké školy. U každého z těchto 120 studentů byla klasifikována přilehlost levé a pravé ušnice do jedné ze tří kategorií (1 – přilehlé, 2 – středně přilehlé, 3 – odstávající). U každého studenta byl dále zaznamenán údaj o jeho pohlaví (0 – muž, 1 – žena).

1. Načtete datový soubor 101-usnice.txt.
2. V datové tabulce přiřadíte jednotlivým kategoriím přilehlosti ušnic jejich názvy (1 – přilehle, 2 – středni, 3 – odstávající) a ve sloupci sex změňte kódování 0, 1 na muž a žena. Názvy znaků potom změňte na prava, leva, pohlavi. Nadále pracujte s touto datovou tabulkou.
3. Sestrojte variační řadu pro přilehlost
  - (a) pravé ušnice u žen;
  - (b) levé ušnice u žen.
4. Nakreslete sloupkový graf a polygon četností pro přilehlost
  - (a) pravé ušnice u žen;
  - (b) levé ušnice u žen.

```
##          nj Nj  pj  Fj
## přilehle  38 38 0.6333 0.6333
## středni   21 59 0.3500 0.9833
## odstávající  1 60 0.0167 1.0000
```

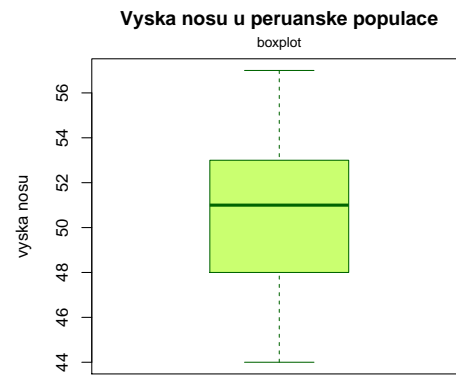
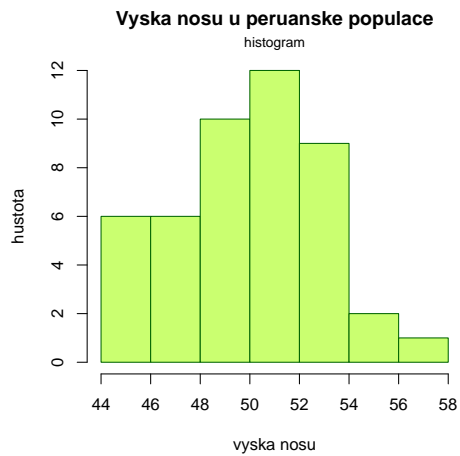
```
##          nj Nj  pj  Fj
## přilehle  36 36 0.6 0.6
## středni   24 60 0.4 1.0
```



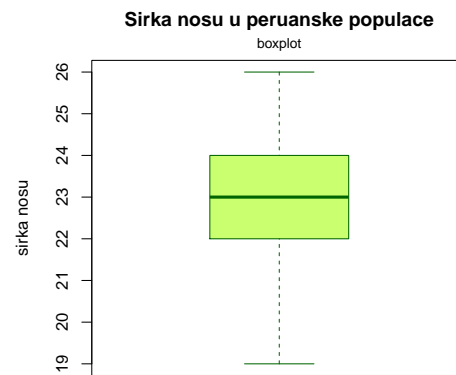
**Příklad 2 (30 b).** Načtěte datový soubor 102-delka-a-sirka-nosu.txt obsahující údaje o délce a šířce nosu malajské a peruánské populace.

1. Stanovte medián, dolní kvartil a horní kvartil pro délku a šířku nosu peruánské populace.
2. Stanovte aritmetický průměr, směrodatnou odchylku, koeficient šikmosti a koeficient špičatosti pro délku a šířku nosu peruánské populace;
3. Výsledky z bodů 1. a 2. vložte do přehledné tabulky.
4. Pro oba znaky (délka nosu, šířka nosu) peruánské populace sestrojte krabicové grafy a histogramy.
5. Všechny zadané výsledky i grafy řádně okomentujte.
6. Vypočítejte vhodný korelační koeficient pro vztah mezi délkou a šířkou nosu peruánské populace. Hodnotu korelačního koeficientu řádně interpretujte.
7. Nakreslete tečkový graf zachycující vztah mezi délkou a šířkou nosu peruánské populace.

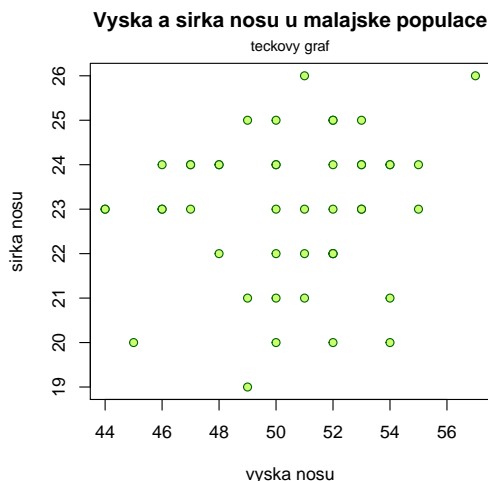
```
##      median d.kvartil h.kvartil  prumer sm.odch sikmost spicatost
## 50%      51         48         53 50.4565  3.0267 -0.3134  -0.4115
```



```
##      median d.kvartil h.kvartil  prumer sm.odch sikmost spicatost
## 50%      23         22         24 22.9783  1.6483 -0.4761  -0.2598
```



## [1] 0.1370691



**Příklad 3 (10 b).** V nemocnicích jednoho blíže nespecifikovaného kraje byl zkoumán poměr zastoupení pohlaví druhorozených dětí. Na základě dlouhodobého výzkumu bylo zjištěno, že pravděpodobnost, že narození druhého dítěte – holčičky v tomto kraji nastává s pravděpodobností 0.4652, kdežto narození chlapce jako druhorozeného dítěte nastává s pravděpodobností 0.5348. Vypočítejte, jaká pravděpodobnost, že z 200 druhorozených dětí narozených v tomto kraji bude

- alespoň 75 holčiček (včetně hodnoty 75);
- 60 až 80 holčiček (včetně hodnot 60 a 80);
- nejvýše 75 holčiček.

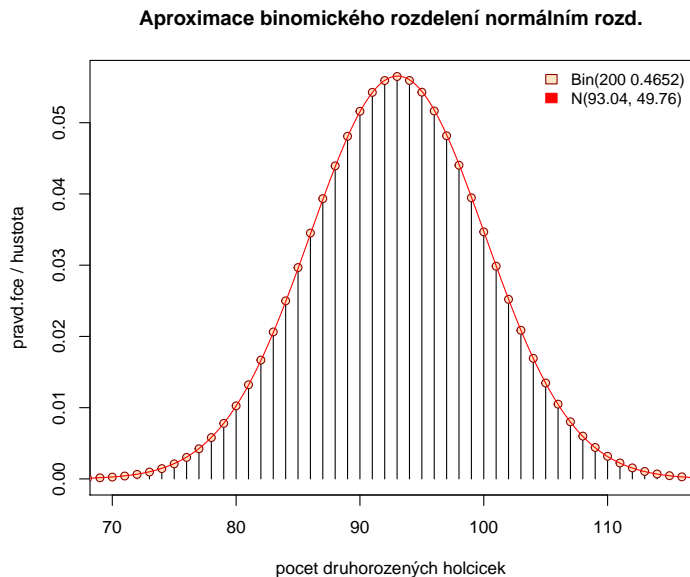
## [1] 0.9959142  
## [1] 0.03729584  
## [1] 0.006196344

**Příklad 4 (15 b).** Za předpokladu, že náhodná veličina pochází z binomického rozdělení  $X \sim \text{Bin}(n, p)$ , se tato veličina  $X$  asymptoticky (= pro dostatečně velké  $n$ ) řídí také normálním rozdělením se střední hodnotou  $\mu = np$  a rozptylem  $\sigma^2 = np(1 - p)$ , tedy  $X \sim N(np, np(1 - p))$ .

- Se znalostí této informace vypočítejte přibližnou pravděpodobnost, že mezi 200 druhorozenými dětmi narozenými v tomto kraji bude
  - alespoň 75 holčiček (včetně hodnoty 75);
  - 60 – 80 holčiček (včetně hodnoty 60 a 80);
  - nejvýše 75 holčiček.
- Porovnejte tyto hodnoty s hodnotami z předchozího příkladu. Na kolik des. míst se shodují?
- Teoretické okénko:
  - Jak se nazývá věta, která nám umožňuje aproximovat binomické rozdělení normálním rozdělením?
  - Jak bychom docílili toho, aby se pravděpodobnosti vypočítané pomocí normálního rozdělení více přiblížily skutečným hodnotám pravděpodobností vypočítaných pomocí binomického rozdělení?

```
## [1] 0.9965248
## [1] 0.03225584
## [1] 0.00527224
```

**Příklad 5 (10 b).** 1. Nakreslete pravděpodobnostní funkci binomického rozdělení  $Bin(n, p)$  z příkladu 3 a superponujte ji křivkou normálního rozdělení  $N(np, np(1 - p))$  tak, jak je uvedeno na následujícím obrázku.



**Příklad 6 (25 b).** Zkoumali jsme potomky kosmanů. Náhodná veličina  $X$  udává počet manželských potomků, které samice porodila a náhodná veličina  $Y$  počet nemanželských potomků, které samice porodila. Je známa simultánní pravděpodobnostní funkce  $\pi(x, y)$  diskrétního náhodného vektoru  $(X, Y)$ :

Tabulka simultánní pstní fce $\pi(X, Y)$			
$X$ - počet manž.p.	$Y$ - počet nemanž.p.		
	1	2	3
1	0.2	0.04	0.01
2	0.15	0.36	0.09
3	0.05	0.1	0.0

Vypočtete koeficient korelace manželských a nemanželských potomků. Hodnotu koeficientu korelace řádně interpretujte.

```
##           EX  EY  DX  DY  CXY  RXY
## Charakteristiky 1.9 1.7 0.39 0.41 0.11 0.2751
```

**Příklad 7 (15 b).** Načtete datový soubor 103-lebky-muzi-zeny.txt obsahující údaje o výšce lebky 215 mužů a 107 žen ze starověké egyptské populace.

1. Vypočítejte bodový odhad pro střední hodnotu výšky lebky mužů starověké egyptské populace;
2. Vypočítejte oboustranný 95% intervalový odhad střední hodnoty výšky lebky mužů starověké egyptské populace.

```
## [1] "m = 131.8977"  
## [1] "dh = 131.2476"  
## [1] "hh = 132.5477"
```

**Příklad 8 (20 b).** Načtete opět datový soubor 103-lebky-muzi-zeny.txt.

1. Vhodným testem ověřte normalitu dat.
2. Na hladině významnosti  $\alpha = 0.01$  testujte nulovou hypotézu, že rozptyl výšky lebky mužů starověké egyptské populace je stejný, jako rozptyl výšky lebky mužů novověké egyptské populace, jejíž odhad byl stanoven jako  $s_m^2 = 26.74$ . Testování proveďte dvěma libovolnými způsoby.
3. Postup celého příkladu řádně okomentujte.

```
## [1] 0.1263
```

a) Testování pomocí kritického oboru:

```
## [1] "t0 = 187.126"  
## [1] 164.4685  
## [1] 271.0372
```

b) Testování pomocí intervalu spolehlivosti:

```
## [1] 18.46149  
## [1] 30.42375
```

c) Testování pomocí p-hodnoty:

```
## [1] 0.1851161
```