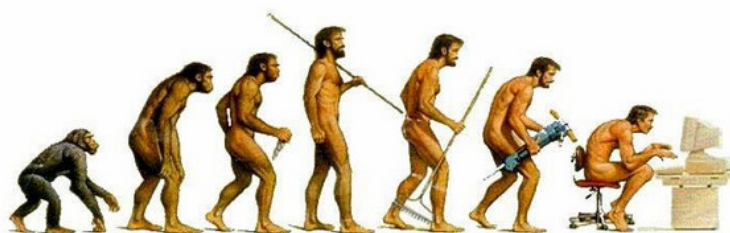


Masarykova univerzita v Brně
Přírodovědecká fakulta

APLIKOVANÁ STATISTIKA PRO ANTROPOLOGY

Zadání domácího úkolu – Skupina A






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.

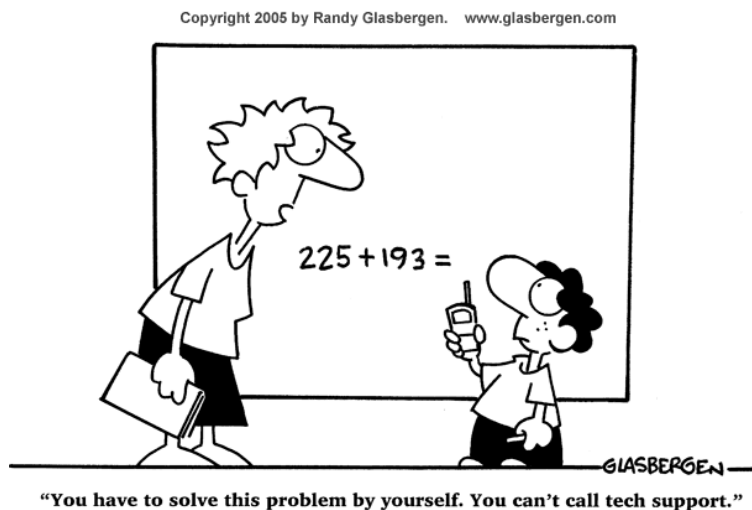
Aby byl úkol uznán za splněný, je potřeba získat alespoň 120 bodů (75 %). Pokud nebude dodatečně stanoveno jinak, má student na vyřešení domácího úkolu 14 dní počínaje dnem zadání domácího úkolu. Pokud student potřebných 120 bodů nezíská, bude mu úkol navrácen a student dostane jeden týden na dořešení 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 nebo 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 :).

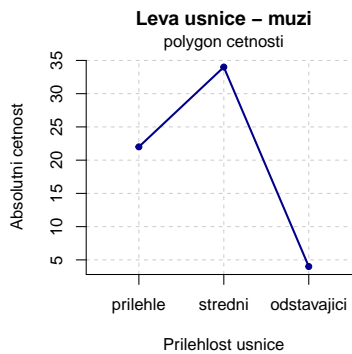
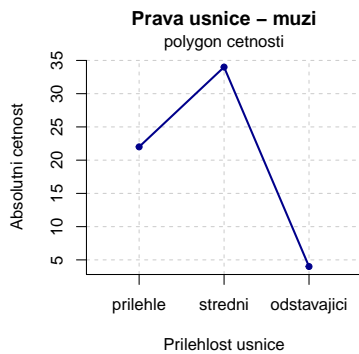
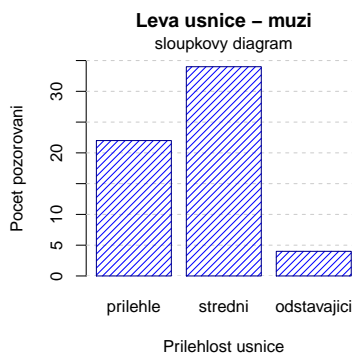
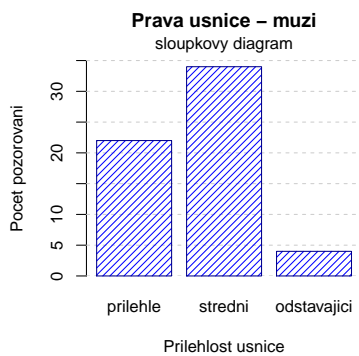


Příklad 1 (20 b). Datový soubor 101-usnice.txt obsahuje údaje o přilehlosti levé a pravé ušnice u 120 studentů jisté vysoké školy. U každého z těchto 120 studentů byla klasifikována přilehlost levé a pravé ušnice do jedné ze tří kategorií (1 – přilehlé, 2 – středně přilehlé, 3 – odstávající). U každého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

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

```
##      nj Nj      pj      Fj
## přilehle  22 22  0.3667  0.3667
## středni  34 56  0.5667  0.9333
## odstávajici  4 60  0.0667  1.0000
```

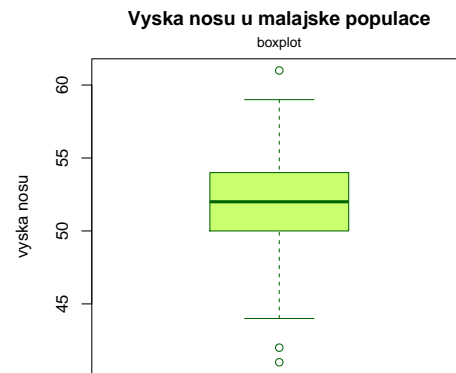
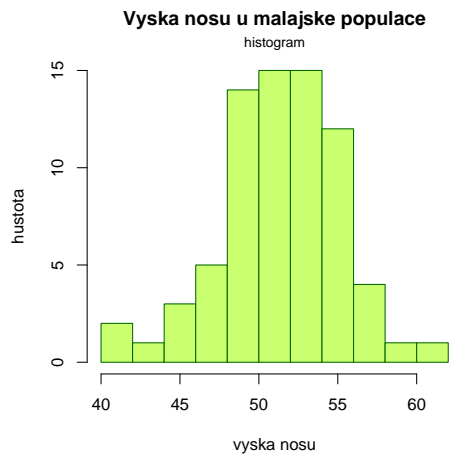
```
##      nj Nj      pj      Fj
## přilehle  22 22  0.3667  0.3667
## středni  34 56  0.5667  0.9333
## odstávajici  4 60  0.0667  1.0000
```



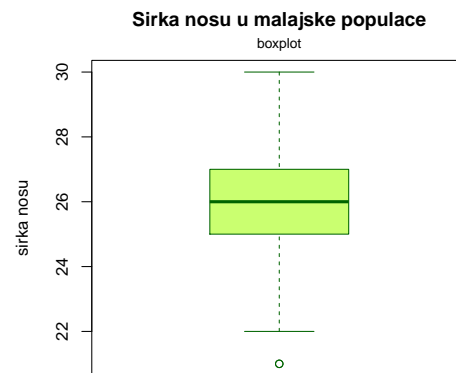
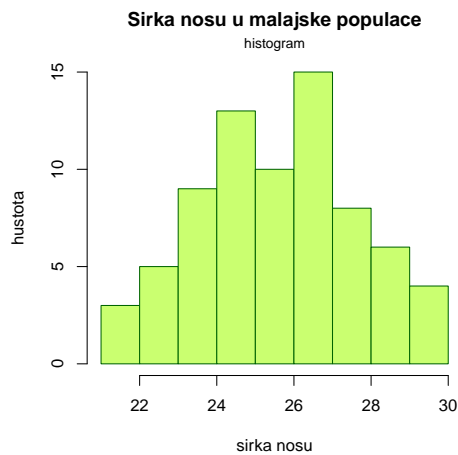
Příklad 2 (30 b). Načtěte datový soubor 102-delka-a-sirka-nosu.txt obsahující údaje o délce a šířce nosu malajské a peruánské populace.

1. Stanovte medián, dolní kvartil a horní kvartil pro délku a šířku nosu malajské populace.
2. Stanovte aritmetický průměr, směrodatnou odchylku, koeficient šikmosti a koeficient špičatosti pro délku a šířku nosu malajské populace;
3. Výsledky z bodů 1. a 2. vložte do přehledné tabulky.
4. Pro oba znaky (délka nosu, šířka nosu) malajské 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 malajské populace. Hodnotu korelačního koeficientu řádně interpretujte.
7. Nakreslete tečkový graf zachycující vztah mezi délkou a šířkou nosu malajské populace.

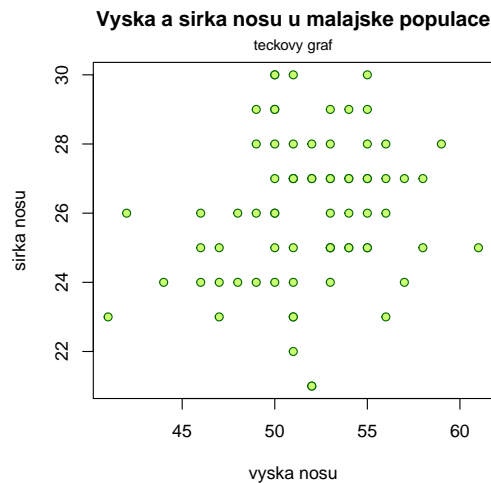
##	median	d.kvartil	h.kvartil	prumer	sm.odch	sikmost	spicatost
## 50%	52	50	54	51.8219	3.7098	-0.3758	0.6635



##	median	d.kvartil	h.kvartil	prumer	sm.odch	sikmost	spicatost
## 50%	26	25	27	26.0685	2.1089	-0.1547	-0.3521



```
## [1] 0.1889101
```



Příklad 3 (10 b). Pravděpodobnost výskytu dermatoglyfického vzoru 'vír' na palci pravé ruky u mužů je 0.533. Jaká je pravděpodobnost, že mezi 100 náhodně vybranými muži bude mít vzor víru na palci

1. nejvýše 35 mužů;
2. 45 – 55 mužů (včetně hodnoty 45 a 55);
3. 60 a více mužů.

```
## [1] 0.000173664
```

```
## [1] 0.6306127
```

```
## [1] 0.1066598
```

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 $X \sim N(np, np(1 - p))$.

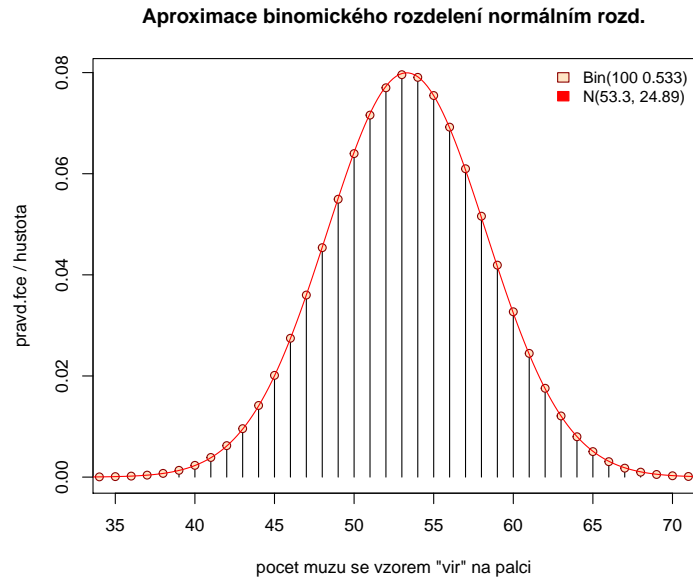
1. Se znalostí této informace vypočítejte přibližnou pravděpodobnost, že mezi 100 náhodně vybranými muži bude mít vzor víru na palci
 - (a) nejvýše 35 mužů;
 - (b) 45 – 55 mužů (včetně hodnoty 45 a 55);
 - (c) 60 a více mužů.
2. Porovnejte tyto hodnoty s hodnotami z předchozího příkladu. Na kolik des. míst se shodují?
3. Teoretické okénko:
 - (a) Jak se nazývá věta, která nám umožňuje aproximovat binomické rozdělení normálním rozdělením?
 - (b) Co bychom museli udělat, aby se přibližné pravděpodobnosti vypočítané pomocí normálního rozdělení více přiblížily skutečným hodnotám pravděpodobností vypočítaných pomocí binomického rozdělení?

```
## [1] 0.0001222287
```

```
## [1] 0.6021951
```

```
## [1] 0.126625
```

Příklad 5 (10 b). 1. Nakreslete pravděpodobnostní funkci binomického rozdělení $\text{Bin}(n, p)$ 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

Vypočtěte 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 rozptyl výšky lebky žen starověké egyptské populace.
2. Vypočítejte oboustranný 90% intervalový odhad pro rozptyl výšky lebky žen starověké egyptské populace.

```
## [1] "s^2 = 21.6528"  
## [1] "dh = 17.5164"  
## [1] "hh = 27.5732"
```

Příklad 8 (20 b). Opět budeme pracovat s datovým souborem 103-lebky-muzi-zeny.txt.

1. Vhodným testem ověřte normalitu dat.
2. Na hladině významnosti $\alpha = 0.05$ testujte nulovou hypotézu, že střední hodnota výšky lebky u žen starověké egyptské populace je stejná, jako střední hodnota výšky lebky žen novověké egyptské populace, jejíž odhad byl stanoven jako $m_f = 126.94$. Testování proveďte dvěma libovolnými způsoby.
3. Postup celého příkladu řádně okomentujte.

```
## [1] 0.0989
```

a) Testování pomocí kritického oboru:

```
## [1] "t0 = -2.796"  
## [1] -1.982597  
## [1] 1.982597
```

b) Testování pomocí intervalu spolehlivosti:

```
## [1] 124.7904  
## [1] 126.5741
```

c) Testování pomocí p-hodnoty:

```
## [1] 0.006144932
```