

Biologický korespondenční seminář



Biozvěst

Ročník 1

Serie 1

Milí přátelé,

Po zahřívacím nultém ročníku vstupuje náš seminář do první série. Na co se můžete těšit? Každé dva měsíce pro Vás připravíme sadu pěti úloh, od klasických témat ryze „přírodovědných“ až po nejnovější modní výstřelky z pole molekulární biologie. Doufáme, že si řešením našich úloh rozšíříte své obzory a možná pochopíte staré známé věci ve zcela nových souvislostech. Naprosto vyžít se budete moci v úloze experimentální a doufáme, že něco nového Vám přinese seriál, který bude letos pojednávat o biostatistice.

A nejen to! Během roku pro Vás uspořádáme dvě terénní soustředění a několikrát do roka budete pozváni strávit Fluorescenční noc na Viničné, v budově slavné Přírodovědecké fakulty v Praze. Jako řešitelé semináře budete zvýhodněni při přijímání na toto setkání plné mikroskopování, výletů a přednášek.

Šíření moudrosti našich Biozvěstů podpořil finančně program EU Mládež v akci a mediálně PrF UK.

Tento projekt byl realizován za finanční podpory Evropské unie. Za obsah sdělení odpovídá výlučně autor. Sdělení nereprezentuje názory Evropské komise a Evropská komise neodpovídá za použití informací, jež jsou jeho obsahem.



**Mládež
v akci**

Jak řešit

Veškeré pokyny k řešení semináře získáte na internetové stránce Biozvěstu

www.studiumbiologie.cz/biozvest

(nebo zadejte „Biozvěst“ do Google). Na stránce také naleznete přihlášku, kterou vyplíte. Úlohy Vám budeme zasílat prostřednictvím Google skupiny „Řešitelé Biozvěstu“

Biozvest-resitele@googlegroups.com

<https://groups.google.com/d/forum/biozvest-resitele>

ke které se přihlašte a nastavte, aby Vám od nás přicházely všechny e-maily. Alternativně se k nám můžete připojit prostřednictvím Facebooku, skupina „Biozvěst“

<https://www.facebook.com/groups/175384482597684/>

Odhlášení ze semináře lze provést jednoduše odhlášením ze skupiny. V těchto skupinách máte také prostor pro otázky a diskuse ohledně úloh.

Vaše řešení úloh nám posílejte na adresu:

biozvest@gmail.com

Nejpraktičtější formou řešení bude prostý text v e-mailu, ale přijímáme veškeré formáty příloh. Každou úlohu pište do samostatného e-mailu a v předmětu uveďte

Ročník-Série-Úloha-Jméno_Příjmení,

např. 1-1-2-Bioslav_Biomilný v případě druhé úlohy první série aktuálního ročníku.

Uzávěrka 1. série: pondělí 4.11.2013 o půlnoci.

V případě opožděného odevzdání úloh se strhává za každý celý den jeden bod s výjimkou zvláště závažných a omluvených situací. V případě, že byste se ocitli bez internetu, můžete využít i klasickou poštu

Stanislav Vosolsobě

Katedra experimentální biologie rostlin

Přírodovědecká fakulta Univerzity Karlovy v Praze

Viničná 5

128 44 Praha 2

Vyhodnocení Vašich řešení dostanete e-mailem.

Nelekejte se, když Vám přijdou úlohy na první pohled příliš těžké, ponořte se do informačních zdrojů a uvidíte, že na vše lze někde nalézt odpověď. Dobré tipy k řešení naleznete také na stránce Biozvěstu v sekci „Návody“. Není nutné abyste kompletně vyřešili všechny úlohy a asi se to ani nikomu nepodaří, stačí odeslat libovolně velký fragment. Oceníme, pokud přiložíte jakékoliv připomínky (např. úloha byla příliš lehká/těžká, nesrozumitelná, nudná), úlohy se pokusíme tvořit k Vaší maximální spokojenosti.

Veškeré dotazy či připomínky směrujte na adresy biozvest@gmail.com či vosolsob@natur.cuni.cz

Mnoho zdaru při řešení Vám za kolektiv autorů přeje

Stanislav Vosolsobě

Úloha 1: Zmrzlinový pohár*

Autor: Stanislav Vosolsobě

Počet bodů: 20

Bioslav byl koncem zimy na lyžařském kurzu v Tatrách a náhodou slavil svátek (31.4.). Spolužáci mu připravili zmrzlinový pohár. Neboť byly na horách odříznuty od civilizace, museli si vystačit s tím, co okolní příroda dala a tak použili k přípravě poháru různobarevný sníh z přírody. Pohár se skládal z těchto složek, které jsme pracovníě pojmenovali podle vzhledu (Bioslav snědl vždy jen část a zbytek podrobil analýze):

Stracciatella

Obsahovala tisíce drobných tmavých organismů, které od pohledu patřili do skupiny Pancrustacea. Bioslav si již dříve všiml, že tyto živočichové poskakují všude po sněhu v obrovských počtech.

Melounová

Pod mikroskopem byly patrné jednotlivé oranžové buňky. Bioslav izoloval ze zmrzliny pigmenty, rozdělil je chromatograficky a určil absorpční spektrum. Identifikoval dvě látky s absorpčními maximy 663 a 430 nm, respektive 648 a 463 nm a velké množství látky s absorpčními mezi 450 a 500 nm. Jiný zástupce tohoto rodu je již delší dobu kompletně osekvenován.

Citronová

Bioslav měl silné tušení hned od prvního pohledu. Pod mikroskopem nic nepozoroval. Nechal vzorek odpařit a získal bezbarvou látku tvořenou jehličkovitými krystaly. Své tušení potvrdil poté, co trochu látky natavil nad vařičem a poté přidal k Fehlingovu činidlu - vzniklo temně modré zabarvení. Při zahřívání roztoku nenatavené látky ve

* Úloha pro velký úspěch recyklována z minulého ročníku

Fehlingově činidle se nic nestalo.

Čokoládová

Pod mikroskopem Bioslav spatřil neidentifikovatelnou změť. Potom preparát obarvil barvivem DAPI a pod fluorescenčním mikroskopem s ultrafialovou excitací spatřil velké množství drobných modrých teček (objektiv 60x) a občas i větší kolečka. Bioslav zajásal, provedl izolaci DNA a amplifikoval různé mitochondriální geny pomocí PCR. Toto zjistil sekvencí:

>sekvence 1

```
AACTATTCCCTGATCTCTCTCCCCACATTTTAATTCATATATT
TAATAACATCTACTGTACCTCCCCAGTATGTACTCTTTCCACCCC
CTATGTATATCGTGCATTAATGGTTTGCCCATGCATATAAGCATG
TACATACTATGCTTGATTTTGCAATTCGTGCACCTACCTAGACCA
CGAGCTTGATCACCAAGCCTCGAGAAACCATCAACCCCTTGCTAGA
CGTGTACCTCTTCTCGCTCCGGGCCCATAGCATGTGGGGGTTTCT
AGCCTGAAACTATACCTGGCATCTGGTTCTTACTTCAGGGCCATG
AAAGTCCTCAATCCAATCTACTAACCTCTCAAATGGGACATCTC
GATGGACTAATGACTAATCAGCCCATGATCACACATAACTGTGGTG
TCATGCATTTGGTATTTTTTAATTTTATAGGGGGGGGGGACTGGTA
TCAGCTGAGCTATGGCCGTAAAGGCCTCGTAGCAGTCAAATAACTT
GTAGCTGGGCTTATCCTTCATCATT
```

>sekvence 2

```
ATGACAAACATTCGAAAAACACACCCCTACTAAAAATTATTAAT
CACTCTTTTCATCGACTTACCAGCCCCATCCAATATCTCATCATGA
TGAAACTTTGGCTCCCTCTTAGGAATCTGCCTAATAATCCAAATC
CTTAGAGGTCTATTTCTAGCAATACACTATACATCAGACACAATAA
CAGCATTCTCTTCAGTAACTCATATCTGCCGAGACGTAAATTATG
GATGACTAATTCGATATTTACACGCAACGGAGCCTCAATATTCT
TCATTTGCTTATTTCTACATGTAGGACGAGGAATATATTACGGAT
CATATACCTTCATAGAAACATGAAACATCGGTGTAATTCTCCTATT
CGCAGTAATAGCCACAGCATTATAGGTTATGTCTTCCATGGGG
ACAAATATCTTCTGAGGAGCAACAGTAATTACAAATCTATTATCA
GCAATTCCATACATCGGAACCTAGTAGAATGAATCTGAGGA
GGATTCTCAGTAGATAAAGCTACACTAACACGTTTCTTCGCCTTC
CACTTTATCTTACCATT
CATCATTTGCCGCCCTAGTAATCGTCCATCTCCTATTTCTCCATGA
AATGGATCAAATAACCCCTACAGGCCTTAACCTCAGACGCCGACAA
AATTCCATTTACCCGTACTATACAATTAAGACATTCTAGGAGTA
CTCATAATAGTGTCTTCTTAATAACCTTAGTTCTCTTCTTCCCA
GACCTTCTAGGTGACCCGACAACCTATATACCTGCCAACCCTACTA
AATACCCACCATATTAACCAGAAATGGTACTTCTCTATTTGCA
TAGCAATTCTACGATCCATCCCTAACAAATTAGGTGGAGTCTTA
GCCCTAATTTTATCTATCCTTATTCTAGCCCTATTACCATTCCTA
CACACCTCTAAACAACGCGACCTAATATTCCGTCCAATCACTCAA
ACCCTATACTGAATCTTAGTTGCTAACCTTCTCGTCTAACTTGA
ATTGGAGGTCAACAGTAGAACACCCATTATCATCATTTGGCCAA
CTAGCCTCAATCAGCTACTTCTTATTATTCTAATCCTCATACCC
ATCTCAGGCATTATTGAAG
ATAAAATA
```

Limetková

Tentokrát Bioslav našel organismu příbuzný organismu z melounové zmrzliny, který se i v přírodě vyskytuje v podobném prostředí. Akorát nebyly útvary kulaté a oranžové, ale zelené rohlíčkovité s dlouze zašpičatělými konci.

1. Který organismus byl ve zmrzlině stracciatella a jak se tam dostal?
2. U melounové zmrzliny se pokuste určit, o jaké mohlo jít pigmenty a která skupina s převahou jednobuněčných organismů jimi disponuje.
3. Vyhledejte si na internetu seznamy kompletně osekvenovaných organismů a nalezněte prvně sekvenovaného zástupce z výše jmenované skupiny.
4. Který druh tohoto rodu je typický ekologií popsanou v úloze, jak vypadá jeho ekologická nika (čili „kde, kdy, jak, proč a s kým“)?

5. Pokuste se objasnit poněkud nestandardní aplikaci Fehlingova činidla na citronovou zmrzlinu a co bylo „esencí“ této zmrzliny.
6. Jak první umělá syntéza hlavní látky z citronové zmrzliny změnila náhled na živý svět?
7. Co jsou nejpravděpodobněji malé tečky a větší kolečka obarvená barvivem DAPI v čokoládové zmrzlině a kde se tam vzali?
8. Zdůvodněte přítomnost dvou zmíněných sekvencí v čokoládové zmrzlině (při identifikaci sekvencí využijte zkušenosti z 5. úlohy 2. série 0. ročníku, na webu Biozvěstu v sekci „Archiv“).

Úloha 2: Mléko

Autor: Stanislav Vosolsobě

Počet bodů: 12

Produkce mléka je ústřední evoluční novinkou ssawců. Neméně významnou roli sehrála tato kapalina ve vývoji indoevropské populace lidstva – jak by asi vypadala naše krajina a strava, kdyby neexistovalo pastevectví za účelem mléčné výroby? Než se začala využívat ropa, bylo mléko bezpochyby druhou nejvýznamnější tekutinou pro lidskou civilizaci, hned po vodě. Přestaňme ale filosofovat.

1. Nejprve se podíváme na mléko z evolučního pohledu. Propátrejte, jaké jsou teorie evoluce mléčných žláz, z jakých vlastností plazů mohla vzniknout produkce mléka? Pokuste se probádat původní odborné práce, rady jak s nimi pracovat naleznete na webu Biozvěstu v sekci „Návody>Literární zdroje“.
2. Jakým obecně známým procesem na buněčné úrovni je produkováno mléko?
3. Popište, kde v buňce se syntetisují mléčné proteiny (a také všechny buněčné proteiny, které podléhají procesu z otázky 2) a jak putují po buňce, k čemu je tzv. „signální peptid“?
4. Zjistěte, který protein je v mléce nejhojnější. Získejte jeho sekvenci a doložte, že se skutečně v buňce syntetisuje procesem nastíněným v otázce 3. Pomůže Vám SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP/>). Na stránce SignalP nemusíte nic přenastavovat, pouze vložte čistou sekvenci proteinu, pokud budete vkládat ve FASTA formátu, odmažte složitou hlavičku, kterou mají sekvence ze serveru NCBI, jinak SignalP hlásí chybu. Jak najít žádanou sekvenci a jak s ní naložit Vám ozřejmí první díl seriálu o bioinformtice v nultém ročníku Biozvěstu, naleznete jej na webu v sekci „Archiv“.

Úloha 4: Vegetariáni

Autor: Albert Damaška

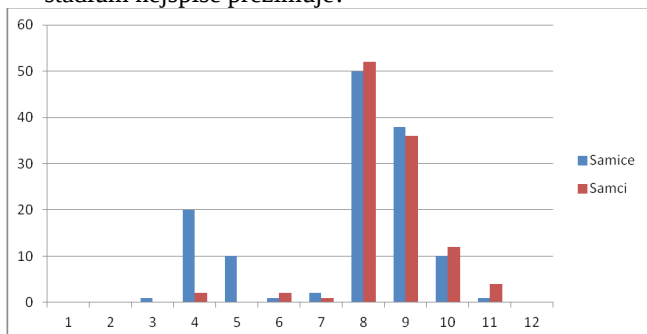
Počet bodů: 20

Někdo vnímá vegetariány jako osvícené průkopníky zdravé výživy, jiný zase za blázny s poruchou příjmu potravy. Člověk totiž není specialisován pouze na příjem rostlinné stravy. Řada jiných živočichů se však živí pouze rostlinami. Říkáme jim *herbivoři* nebo také *fytofágové*.

1. Když v přírodě pojídáte nějaké rostliny, můžete se specializovat buďto pouze na jeden konkrétní druh, nebo vybírat z širšího spektra. Živočichům, vázaným na kon-

krétní druh rostliny, říkáme monofágové, zatímco těm, kteří tak vybírají nejsou a živí se spoustou druhů napříč skupinami rostlin, říkáme polyfágové. Najděte nějaké příklady monofágního a polyfágního druhu hmyzu.

2. Zamyslete se nad tím, čím může být pro zvíře výhodné být monofágem nebo polyfágem a jaké to naopak přináší nevýhody. Uveďte pro každou ze strategií positiva a negativa.
3. Být monofágem (nebo tzv. oligofágem – druhem, specializovaným jen na určitou skupinu vzájemně si blízce příbuzných druhů rostlin) není jen výsadou hmyzu. Uveďte příklady úzce specializovaných herbivorních savců a ke každému rostlinu, kterou se živí.
4. Rostliny se samozřejmě brání tomu, aby byly ožírány. Vytvářejí proto řadu obranných strategií. Účinným způsobem, jak se ubránit nepříteli z řad hmyzu, je použití chemických zbraní. Uveďte nějaké chemikálie, které rostliny používají v boji proti hmyzu, a ke každé i rostlinu, která ji produkuje. Jaké další prostředky by proti hmyzu mohly být účinné?
5. V poslední době vymírají či už vymřeli v ČR někteří motýli, jako například žlutásek barvoměnný (*Colias myrmidone*). Mohlo by se zdát, že za tím stojí vymizení jejich živné rostliny, u zmíněného žlutásky konkrétně čilimníku. Rostlina se ale u nás stále vyskytuje. Co jiného mohlo zapříčinit vymření tohoto motýla?
6. Následující graf zachycuje hypotetickou situaci. Na živných rostlinách na jedné lokalitě jsme každý měsíc hledali imaga jednoho druhu dřepčíka. Graf ukazuje, kolik jedinců kterého pohlaví jsme v jakém měsíci našli. Vysvětlíte tvar grafu (růsty a poklesy). Které stadium nejspíše přezimuje?



7. Získávat efektivně energii z rostlin není jednoduché, neboť je jí hodně vázáno ve složitém polysacharidu celulóze, základní to stavební složce rostlinných pletiv. Rozkládána může být sice enzymem celulasou, ten ale živočichové syntetizovat sami neumí. Přesto však mnozí z nich dokáží energii z celulósy získat. Jak to dělají? Mimochodem, někteří termity si vlastní celulasu vytvářet dovedou. Jak u nich tato schopnost mohla vzniknout?
8. O některých druzích býložravého hmyzu často slyšíme v médiích. Je to například mandelinka *Diabrotica virgifera*, nosatec *Ips typographus* nebo bělásek *Pieris brassicae*. Proč se média zajímají zrovna o ně (a ne třeba o dřepčíky rodu *Clavicornaltica*)? Čím je způsobeno to, co dělají (a co tak zajímá ona média)?
9. V současnosti už v přírodě téměř nevzniká žádné nové uhlí. Existuje domněnka, podle které za to mohou termity. Jak mohou tyto švábi tak výrazně narušovat proces vzniku uhlí?

Úloha 4: Ejhle houba!

Autor: Stanislav Vosolsobě, Celie Korittová, Houby na Vůznicí

Počet bodů: 16

K podzimu neodmyslitelně patří houby. A protože to jsou nesmírně zajímavé organismy, podíváme se na ně podrobněji i v našem semináři.



1. Často v lese nacházíme houby malé a velké. Zkuste na vybraném druhu houby zjistit, jak rychle dochází k růstu plodnic, zda některé houby mají předurčeno zůstat malé a jiné vyrůst veliké, či zda všechny houby mohou dorůst do maximální velikosti, pokud mají dostatek času.
2. Pokuste se stanovit, ve které fázi vývoje plodnice dojde k uvolňování spor, tedy v jakém stádiu plodnice již neznemožníme její reprodukční poslání sběrem.
3. Hlavním záporným dopadem nadměrného sběru plodnic není zdaleka jen znemožnění reprodukce hub samotných, ale především destrukce biotopu pro řadu mnohdy vzácných druhů hmyzu, které označujeme jako „červy“. Pokuste se vychovat z houbových červů dospělé jedince, zdokumentujte je (nakreslete či vyfotografujete) a pokuste se je určit.

Úloha 5: Selský pohled na základní statistické veličiny

Autor: Stanislav Vosolsobě a Jiří Hadrava

Počet bodů: 9

Seriál v letošním ročníku se Vám pokusí přiblížit statistiku, která je nezbytná pro exaktní rozhodování v biologii. Pokusíme se na statistiku pohlédnout z takové strany, aby byla podstata jejího uvažování vždy opřena pouze o předpoklady, které jsou intuitivně zřejmé. Ústřední statistické otázky nastíníme na následujících příkladech:

- 1) V lese je několik set stromů, každý jinak vysoký. Existuje nějaký způsob, jak odpovědět na triviální otázku „Jak je ten les vysoký?“
- 2) Změřili jsme pouze několik stromů z tohoto lesa. Můžeme nějak odhadnout odpověď na výše uvedenou otázku?

3) Máme větší počet lesů. Můžeme rozhodnout, zda se liší výškou?

4) Jen výška k popisu lesa nestačí. Jak odpovědět na generální otázku „Čím se tyto lesy liší?“ a je možné zjistit, zda vzhled lesů nějak souvisí s vybranými podezřelými faktory, jako je třeba podloží, srážky a podobně?

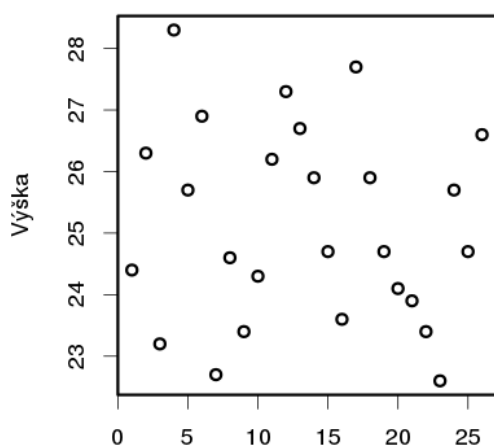
Každému ze statistických problémů bude věnován jeden díl našeho seriálu.

„Jak je ten les vysoký?“

Výchozím podkladem pro každé statistické zpracování je vždy soubor dat, v případě našeho lesa to bude jednoduchá tabulka výšky (l) typu:

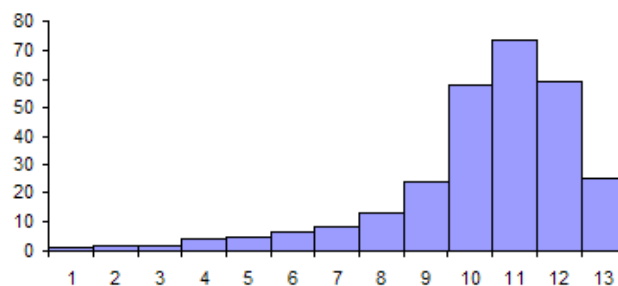
Číslo stromu i	Výška l (m)
1	24,4
2	26,3
3	23,2
...	...
...	...
poslední strom (n -tý)	28,3

Na prvním obrázku je tabulka převedena do jednoduchého grafu závislosti výšky na pořadovém čísle stromu. Vidíme, že výšky stromů se pohybují v rozmezí 22 – 29 m.



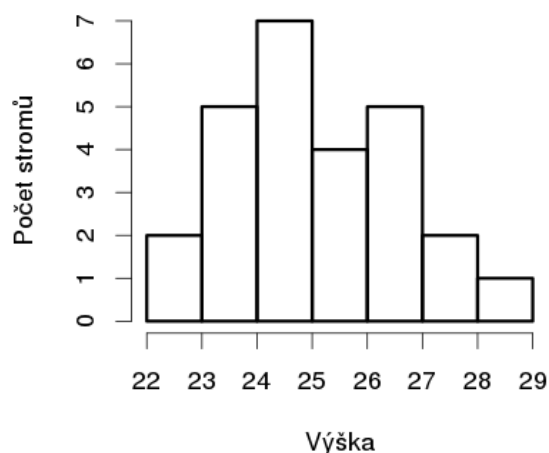
Index stromu

Grafické zobrazování je ve statistice velice důležité, neboť lze již z grafu usoudit mnoho závislostí. Prvním důležitým typem grafu je **histogram**, zkonstruuje se tak, že si dle potřeby data rozdělíme do intervalů, v našem případě po jednom metru, a vyneseme graf počtu stromů v jednotlivých výškových kategoriích. Z grafu jsou patrné všechny hlavní statistické parametry dat – poloha nejčetnějších hodnot (**modus**) a charakter rozložení dat, který lze popsat parametry **špičatosti** (špičaté rozložení znamená, že většina hodnot se pohybuje v úzkém intervalu okolo nejčastější hodnoty a okolo je jen malé množství velmi rozptýlených dat) a **šikmostí** (typické šikmé rozložení má třeba výška rostlin – kontinuum od semenáčků po dospělé rostliny, které dosahují určité maximální velikosti, kterou přesahuje jen málo jedinců). Příklad takového šikmého rozložení vidíte na obrázku. Rozdělení s nulovou šikmostí poznáme podle toho, že je symetrické dle středu (nejčastější hodnoty)



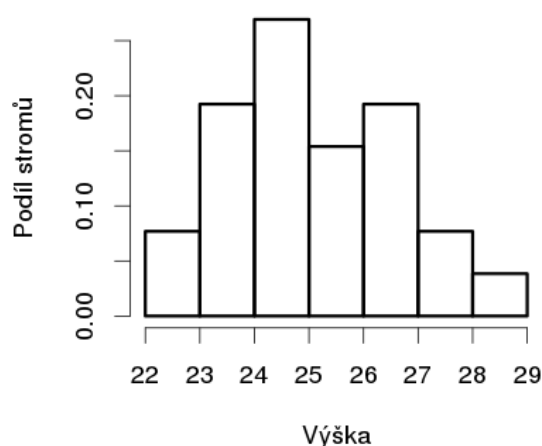
Na následujícím obrázku je již konkrétní histogram pro náš les. Sešikmení je v tomto případě mírné směrem k vysokým hodnotám.

Histogram četnosti



Místo prostého počtu změřených hodnot spadajících do jednotlivých intervalů histogramu můžeme vynášet četnost relativní, neboli podíl z celkového počtu hodnot. Po vynásobení stem získáváme klasická procenta. Údaje v procentech můžeme číst i jako **pravděpodobnost**, v tomto případě třeba že strom s výškou mezi 23 a 24 m nalezneme v lese zhruba s dvacetiprocentní pravděpodobností.

Histogram hustoty



Vraťme se ale k naší úvodní otázce: **Jak získat několik málo čísel, která by nám o našem lese řekla co nejvíce a nemuseli bychom prezentovat celý soubor dat?** Vзьijme se do postu lesníka, pro něhož je důležité odhadnout, kolik dřeva v daném lese stojí. Počet klád získaných vykácením

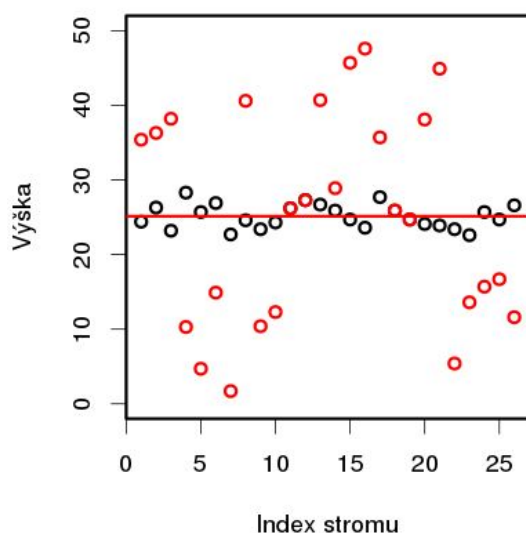
lesa bude odpovídat celkové délce všech stromů v lese (L), kterou vypočítáme takto:

$$L = \sum_{i=1}^n l_i$$

Řecký symbol sigma znamená „sčítej jednotlivé délky stromů z nějaké tabulky (množiny), kde jsme si je označili obecně jako l_i (tedy l_1, l_2 až l_n pokud máme n údajů v tabulce. Rozsah sčítání se značí podivnými kódy nad a pod sigma - sčítej počínaje stromem s indexem $i = 1$ až po poslední n -tý strom“. Potom je intuitivní zavést **aritmetický průměr** tak, že vydělíme celkový součet délek všech stromů jejich počtem, neboli:

$$\bar{l} = \frac{1}{n} \sum_{i=1}^n l_i$$

Aritmetický průměr $\bar{l}$ (čteme „l s pruhem“) můžeme také nazývat **efektivní hodnotou** výšky, neboť kdybychom nahradily všechny stromy jedinci s výškou $\bar{l}$, jevil by se nám les v mnoha ohledech identicky. Pojem efektivní hodnota se často používá v souvislosti s vykonanou prací, tam je význam zcela zřejmý. Aritmetický průměr má i fyzikální význam **těžiště** dat, kdyby se graf dat či histogram podložil osou v místě průměru.



Teď si však představte dva různé lesy: Jeden bude dobře pěstěný (tzv. „smrčák jak sviňa“) s téměř jednotnou výškou (hodnota špičatosti rozložení bude vysoká), druhý bude „zanedbaný“ prales s obrovskými stromy a různověkým podrostem, jak je zřejmé z grafu. Dejme tomu, že průměrná výška obou lesů bude stejná (na obrázku červená linie). Jak popsat rozdíl mezi lesy, zase nejlépe jedním parametrem? K odvození tohoto parametru si Vás dovoluji pozvat na drobnou myšlenkovou procházku. Představme si, že graf lesa je objektem z kuliček v ploše (co datový bod, to jedna kulička), který uchopíme za osu v místě aritmetického průměru. Rozdíl v rozložení kuliček se neprojeví, pokud bude objekt v klidu, ale pokud do objektu udeříme, aby se roztočil, zjistíme, že roztočit objekt s kuličkami daleko od sebe bude mnohem náročnější, než objekt s kuličkami blízko od osy (vyzkoušejte si sami s koulí zavěšenou na provázku). Fyzikálně to lze vyjádřit za pomoci jednoduchých vzorečků. Máme-li za daný čas t opsat kružnici o poloměru r (neboli otočit grafem kolem dokola), bude muset být rychlost pohybu jednotlivých kuliček tím

větší, čím budou dále od osy otáčení (to je jako na kolotoči, čím dále jsme od středu, tím rychleji se pohybujeme), neboli

$$v = \frac{2\pi r}{t}$$

a její kinetická (pohybová) energie, kterou musíme vykonat na roztočení kuličky, tudíž závisí kvadraticky na poloměru rotace (vzpomeňte na základní vzoreček pro kinetickou energii)

$$E = \frac{1}{2}mv^2 = \frac{2m\pi^2}{t^2}r^2$$

Vzhledem k tomu, že čas i hmotnost všech kuliček při rotaci objektem (grafem) budou stejné, můžeme označit celý zlomek jako konstantu k a vztah zapíšeme jednoduše

$$E = kr^2$$

Máme tu tedy veličinu s významem energie, pomocí které dokážeme jednoduše popsat rozložení bodů v prostoru. Ve statistice se analogie rotační energie nazývá **rozptyl** a pro celý datový soubor se nejdříve vypočte jeho celková hodnota, která je úměrná energetické náročnosti hypotetického uvedení dat do rotačního pohybu

$$R = \sum_{i=1}^n (l_i - \bar{l})^2$$

kde rozdíl v závorce je vzdálenost od rotační osy (těžiště, čili v tomto případě aritmetického průměru) a pak obdobně jako v případě průměru vydělíme celkovou veličinu počtem datových bodů

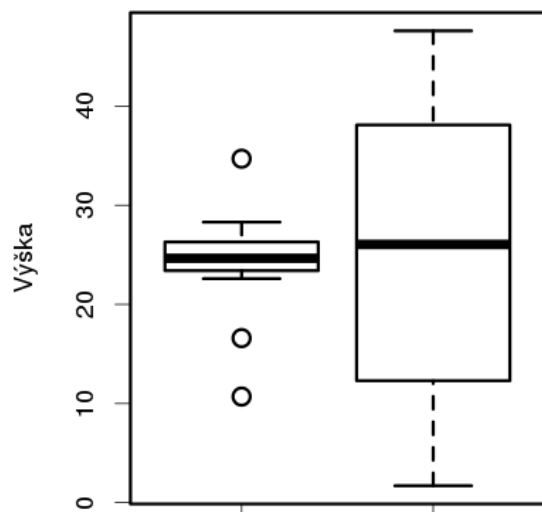
$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (l_i - \bar{l})^2$$

Velčinu σ^2 označujeme jako **rozptyl** či **variance**, její odmocnina σ zvaná **směrodatná odchylka** potom udává, v jaké vzdálenosti od průměru by měly být rozloženy všechny datové body, aby z hlediska „energetické náročnosti rotačního pohybu“ se choval datový soubor stejně jako stávající.

Pro pochopení principů rotačního pohybu existují jednoduché experimenty, například pirueta – krasobruslař se roztočí běžnou rychlostí s rozpráženými rukama a poté připaží – okamžitě to způsobí zrychlení jeho rotace, neboť rotace s rukama daleko od těžiště byla více energeticky náročná. Podobně na kolotoči můžete měnit rychlost otáčení tím, že se přibližujete či vzdalujete ke středu.

Díky druhé mocnině při výpočtu rozptylu zvýrazňujeme příspěvek velkých odchylek, my už však intuitivně tušíme, že to není jen jakýsi výmysl „praotce statistika“, ale vychází to z fyzikálního pohledu na problém jak popsat data.

Pro úplnost našeho textu je nezbytné zmínit také statistický parametr zvaný **kvantil**, který si vysvětlíme na 50% kvantilu zvaném **medián**. V našem příkladě je to výška, pro kterou platí, že polovina stromů z lesa je vyšší, než je hodnota mediánu a polovina stromů bude nižší. **Dolní kvartil** je název pro 25% kvantil, ohraničuje výšku první čtvrtiny nejnižších stromů, **horní kvartil** naopak odřezává čtvrtinu nejvyšších stromů. Obecným pojmem je **percentil**, například desátý percentil je výška, která dělí soubor na 10 % nejnižších stromů a zbytek.

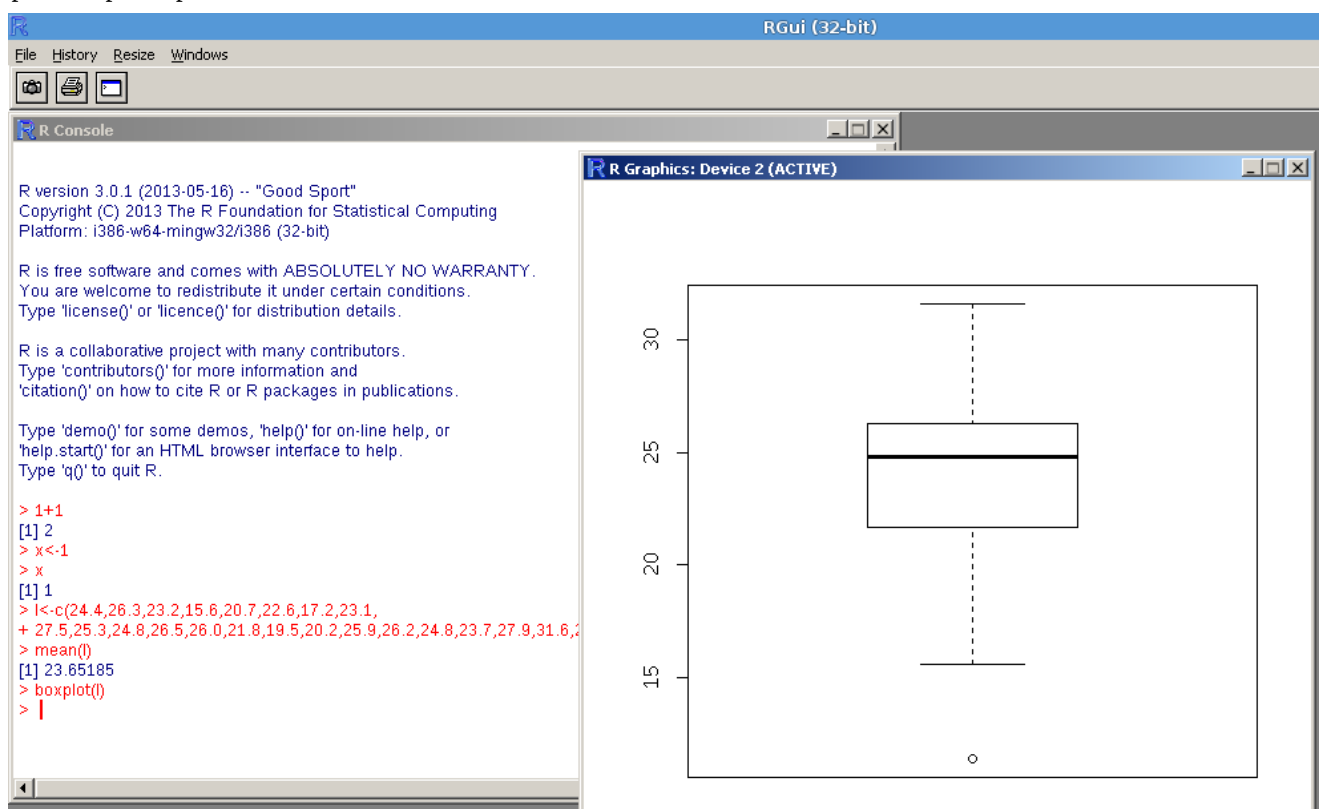


Na principu kvantilů je založen asi nejpraktičtější graf ve statistice zvaný boxplot. Pokud máme více lesů a chceme nějak vizualizovat všechny najednou, tak pomocí histogramů, které jsme používali dosud, by byl výsledek velmi nepřehledný. Za tímto účelem se používá tzv. krabicový diagram neboli boxplot. V něm je zachována většina informací, které můžeme vyčíst z histogramu. Je tvořen obdélníčkem nakresleným podél y-ové osy mezi hodnotami obou kvartilů. Uprostřed obdélníku je tlustá čára v místě mediánu a na obě strany od obdélníku vedou úsečky, které standardně dosahují k datovému bodu, který je nejvýše 1.5x dále od okraje obdélníku, než je délka obdélníku. Pokud jsou v souboru dat hodnoty ještě odlehlejší, kreslí se do grafu kolečkem, pokud je rozsah dat užší, vymezují pak úsečky přesně maximum a minimum. Z boxplotu lze vyčíst rozptýl rozložení podle délky obdélníku, špičatost podle poměru obdélníku a úseček a sešikmenost

dat podle posice mediánu v rámci obdélníku. Aritmetický průměr dat může ležet kdekoli v rámci boxplotu, záleží na rozložení dat. Na obrázku vidíme porovnání boxplotů pro oba naše lesy (u pěstěného lesa byly oproti prvním grafům dodány navíc odlehle hodnoty).

A nyní přejdeme do praxe a povíme si, jak se dá statistika počítat, naučíme se pracovat s mocným nástrojem **R**.

Biologové pro statistickou analýzu svých dat nejčastěji používají software jménem „R“. Jedná se o poměrně univerzální nástroj, kterým lze nahradit jak obyčejnou kalkulačku, tak i třeba tabulkové editory typu Excel. Můžeme v něm také programovat. Hlavní výhodou tohoto software je jeho snadná dostupnost - zdarma si jej můžete stáhnout na stránkách <http://www.r-project.org/> navíc jde o tzv. open source software. to znamená, že jeho zdrojové kódy jsou volně dostupné, a jakýkoli programátor se může podílet na jeho vylepšování. Díky tomu všichni programátorsky zdatnější uživatelé „eRka“ po celém světě mohou tento software nadále rozvíjet tím, že do něj budou doplňovat nové a nové funkce, které je poté možné stáhnout v nějakém doplňkovém balíčku (*package*). Ačkoli to tak může znít, nejedná se jen o nějakou utopistickou představu, ale tento systém skutečně funguje. Mnoho programátorů nejen z řad biologů vydává nové a nové balíčky včetně těch specializovaných na poměrně úzká témata. Mezi více než čtyřmi a půl tisíci balíčky, jejichž seznam si můžete prohlédnout na <http://mirror.fcaglp.unlp.edu.ar/CRAN/>, tak naleznete nástroje na práci např. s genetickými daty, s potravními sítěmi, s fytoecologickými snímky, s geografickými mapami, nebo třeba se záznamy ptačích zpěvů. My se však prozatím zaměříme pouze na základní užívání tohoto software, na běžné matematické úkony a na jednodušší statistické analýzy.



Program R nainstalujete ze stránky

<http://cran.r-project.org/>

uživatelé Windows kliknou na „Download R for Windows“ a poté na **install R for the first time** a **Download R 3.0.1 for Windows**. Po spuštění instalačního programu nic nezkazíte, pokud jen odklikáte „Další“ a necháte výchozí nastavení.

Spustíme-li R, na monitoru nám pod informacemi o užívané verzi software naběhne znak „>“, za nějž můžeme začít psát příkazy. Ano, s Rkem se pracuje pomocí psaní příkazů, což s sebou na jednu stranu nese nutnost naučit se komunikovat jeho jazykem (jazyk, který používá R, se nazývá „S“), na druhou stranu to však přináší i řadu výhod a nových možností. Oproti standardním klikacím grafickým programům typu Excel nám zde nehrozí, že na něco jen tak omylem klikneme, něco se stane, a my nevíme, co. Rko udělá jen to, co mu napíšeme a odklepeme enterem, a každý příkaz, který jsme mu takto napsali, si můžeme zpětně prohlédnout a zkontrolovat. Z toho vyplývá i další obrovská výhoda: seznam příkazů, které použijeme, si můžeme vykopírovat do vedlejšího textového souboru, tzv. **skriptu**, v němž tímto máme kompletně zaznamenaný sebesložitější pracovní postup. To by v programu ovládaném klikáním myši nebylo možné. Budeme-li někdy v budoucnosti chtít provést tytéž úkony pouze s nějakou drobnou modifikací či s jinými vstupními daty, v Rku stačí ve skriptu přepsat příslušný řádek a spustit jej. Oproti tomu v Excelu bychom museli od začátku do konce provádět všechno klikání znovu ručně. Ponechme teď stranou úvahy, jak by asi byl přehledný klikací program, který by nabízel tolik funkcí, kolik nabízí Rko, a vrhněme se rovnou do výuky jazyku Rka.

Průběh výpočtů vidíte na obrázku, v okně programu je další okno, tzv. konsole, což je obecně řečeno okno, kam se píšou příkazy, místo abyste tam na něco klikali. Na obrázku také vidíte, co se stalo po zadání 1+1 (viz další text), červeně jsou příkazy zadávané uživatelem a modře je odpověď Rka. Nápis [1] před dvojkou ignorujte, je jen indexem výstupu, kdyby Rko vypisovalo dlouhatánský vektor čísel přes více řádků, byl by na každém řádku index prvního čísla z tohoto řádku.

Kromě okna s konsolí je na obrázku okno s grafem, to se otevře poté, co zadáte příslušný příkaz ke tvorbě grafu. Ale nepředbíhejme - první řadě můžeme R využít jako obyčejnou kalkulačku. Zkuste napsat nějaké příklady a enterem vyzvěte Rko k jejich spočítání. Můžete používat používat standardní operátory + a -, k násobení použijte symbol * k dělení /. Mocnění se provádí pomocí symbolu ^, závorkovat můžete pomocí obyčejných kulatých závorek. **Pokud se budete chtít vrátit k některému z předcházejících příkazů, abyste jej upravili nebo spustili znovu, použijte šipku nahoru**, tato vlastnost konsoly Vám ušetří velmi mnoho práce.

Rko nemusí čísla jenom číst a vypisovat, ale může si je také zapamatovat. Můžete si zvolit název nějaké **proměnné** (např. x, ale název může být libovolný) a té přiřadit určitou hodnotu, s níž můžete nadále pracovat. Například přiřazení hodnoty 1 do proměnné x uděláte pomocí příkazu

```
x <- 1
```

Symbol <- má podobný význam jako klasické „=“, dokonce lze v Rku používat i =, ale znak šipky je správnější

(nezkoumejme proč), pro psaní znaků jako je < si přepněte na anglickou klávesnici. Mezery mezi symboly psát můžete, ale Rko si stejně dobře poradí i se zápisem x<-1. Hodnotu proměnné vypíšete zadáním jejího jména a stisknutím klávesy enter. Jakým způsobem lze s proměnnými pracovat nejlépe zjistíte, budete-li si s nimi chvíli hrát, zkoušet proměnné přepisovat výsledky příkladů obsahující jiné proměnné či sebe sama (např. $x \leftarrow 2 \cdot x + 1$).

Pracovat nemusíme pouze s proměnnými obsahujícími jedno číslo, ale i s proměnnými obsahujícími celou řadu hodnot. Takovým proměnným se říká **vektory**. Vektory tvoříme pomocí funkce c, které do kulatých závorek (jako tzv. **argumenty funkce**) zadáme čárkami oddělený seznam hodnot, jež má vektor obsahovat. Například vektor obsahující hodnoty od 1 do 5 vložíme do proměnné x příkazem

```
x <- c(1, 2, 3, 4, 5)
```

Budeme-li nyní provádět jakékoli akce s proměnnou x, budeme je provádět s celým vektorem čísel. Vyzkoušejte Rku zadat pár příkladů obsahujících vektor. Třeba napište $5 \cdot x + 1$

Vyzkoušejte si taky, jak funguje řetězení vektorů, např.

```
y <- c(x, x+5, 11, 12)
```

Abychom nemuseli čísla vypisovat ručně, můžeme rovnou zadat, že chceme například vektor čísel od 1 do 12. Ten vytvoříme příkazem c(1:12). Pokud bychom nechtěli pracovat s celým vektorem, ale podívat se pouze na nějakou jeho část, využijeme tzv. **indexování** pomocí hranatých závorek. Na n-tý prvek vektoru x se podíváme příkazem x[n]. Chceme-li tedy například 4. prvek vektoru y nahradit hodnotou 2, uděláme to příkazem

```
x[4] <- 2
```

Indexování v Rku nabízí i daleko zajímavější a komplikovanější možnosti použití. Pro jejich vysvětlování tu však nemáme dost prostoru. Nicméně zájemci si je mohou sami nastudovat metodou pokus-omyl. Jako motivační příklad uvedu třeba možnost nechat si vypsat ty hodnoty z vektoru y, které jsou na těch pozicích, v nichž jsou hodnoty vektoru x rovny dvěma:

```
y[x==2]
```

V zadání je skutečně dvojité rovnítko, které zde doslova znamená otázku „je x identické se dvojkou?“ a Rko si na ni odpovídá „PRAVDA“ či „NEPRAVDA“. Zkuste do Rka napsat jednoduše 1==1, odpoví Vám „TRUE“. Naproti tomu jednoduché „=“ by Vám ve výrazu x=2 zapsalo do proměnné x číslo 2.

A jak je to s tou statistikou?

Rko umožňuje statisticky analyzovat data vložená do vektorů. Udělejme si nyní vektor l obsahující hodnoty výšek stromů, s nimiž jsme se setkali v teoretickém úvodu. Pozor, místo desetinných čárek je třeba používat tečky, neboť čárky jsou již vyhrazeny pro oddělování jednotlivých čísel.

```
l <- c(24.4, 26.3, 23.2, 15.6, 20.7, 22.6, 17.2, 23.1, 27.5, 25.3, 24.8, 26.5, 26.0, 21.8, 19.5, 20.2, 25.9, 26.2, 24.8, 23.7, 27.9, 31.6, 21.6, 11.4, 24.8, 27.7, 28.3)
```

Histogram výšek stromů nyní vytvoříme příkazem hist(l). Aby se nám histogram rozsekal do sloupečků po jednotlivých metrech a ne po pěti metrech, jak to udělá automaticky, zadáme funkci hist kromě datového vektoru ještě jeden argument, a to vektor obsahující místa zlomů mezi jednotlivými sloupečky. Celý příkaz pro vytvoření

histogramu pak bude mít podobu
`hist(l,breaks=c(1:32))`

Analogicky si můžeme vykreslit také krabicový diagram, k tomu použijeme příkaz
`boxplot(l)`

Součet hodnot vektoru l (v teoretickém úvodu označen pomocí velkého písmene L) spočítáme příkazem
`sum(l)`

Aritmetický průměr hodnot vektoru l zjistíme příkazem
`mean(l)`

Rozptyl počítá funkce
`var(l)`

směrodatnou odchylku můžeme spočítat buďto jako odmocninu z rozptylu funkcí `sqrt`, tedy `sqrt(var(l))`, nebo přímo funkcí počítající směrodatnou odchylku `sd(l)`

Užitečnou funkcí je také funkce `summary`: příkaz
`summary(l)`
 nám vypíše nejnížší a nejvyšší hodnotu vektoru l , jejich medián a kvartily. Chceme-li některou z těchto hodnot samostatnou, můžeme si ji oddělit opět pomocí indexování (např. dolní kvartil nám Rko vypíše na zadání příkazu `summary(l)[2]`).

O dalších funkcích a možnostech jejich využití se dozvíte v dalších dílech seriálu. Pokud byste se však nemohli dočkat a chtěli si nějakou další funkci najít už teď, využít můžete velmi kvalitní nápovědy, která je součástí Rka. Vyhledávat v ní můžete pomocí dvou otazníků a následného napsání klíčového slova. Zajímá-li vás například název funkce, která by vracela akorát nejvyšší hodnotu vektoru, zkuste napsat `??maximal`. A hle, taková funkce existuje, pro zjištění maxima si není třeba komplikovat život indexováním kombinovaným s používáním funkce `summary`. Rko v nápovědě našlo funkci `base::max`, což znamená, že v základním balíčku (který se jmenuje `base` a který máte nainstalovaný bez přidávání některého z dalších tisíců doplňkových balíčků), je funkce `max`, která dělá to, co jste hledali. Když znáte název funkce, můžete se detaily o jejích možnostech a použití dozvědět v nápovědě ke konkrétním funkcím. Po vyskočení z vyhledávací nápovědy stisknutím klávesy `q` (jako *quit*) ji vyvoláte pomocí jednoho otazníku a zadání jména funkce. Příkaz

`?max`

vám takto otevře stránku, kde máte vysvětleno, jak fungují funkce `min` a `max`, máte tam vypsané všechny argumenty, které jim lze zadat, najdete tam odkazy na další podobné funkce, najdete tam příklady využití těchto funkcí a také citaci autora, který tuto funkci pro R naprogramoval.

Pokud budete propadat při rozhovoru s programem panice, pamatujte na základní postuláty programování a práce s počítačem vůbec

1) Můj problém vyřešit jde.

2) Já na to přijdu.

3) Radu, jak na to, určitě někde najdu (a to buď v nápovědě k programu, viz příkaz `?něco`, či najdu nějaký návod na

internetu (Googlujte, googlujte, googlujte!) a nebo se zeptám zkušenějšího – neváhejte a pište na fóra Biozvěstu či přímo autorům)

A na závěr úloha:

1. Vyberte si libovolný soubor objektů v existujícím Vesmíru a změřte u nich nějakou vlastnost, tak jako byly zkoumány lesy. Udělejte proměnnou v R.
2. Spočítejte průměr, rozptyl a směrodatnou odchylku
3. Vytvořte histogram a boxplot. Výstup nám pošlete třeba tak, že celou pracovní plochu v programu R včetně grafů zkopírujete stiskem klávesy PrintScreen (PrtSc), načež si můžete otevřít program jako třeba Word, PowerPoint, Malování či jiný grafický editor a stiskem kláves Ctr+V vložíte obrázek a uložíte.

