

Biologický korespondenční seminář



Biozvěst

Ročník 1

Série 2

Milí přátelé,

Moc nás potěšilo Vaše zapálení pro řešení úloh první série, obdrželi jsme i řešení, která by dokonalostí vypracování předčila i schopnosti Biozvěstů. V nové sérii naleznete o úlohu více, neboť povídání o programu R nám z pomůcky k počítání statistiky přerostlo v samostatný seriál. Přejeme Vám, abyste s Biozvěstem příjemně strávili závěr roku a užili si čas Vánoc.

Šíření moudrosti našich Biozvěstů podpořil finančně program EU Mládež v akci a mediálně PrF UK. Díky tomu pro Vás můžeme uspořádat dvě terénní soustředění a několikrát do roka budete pozváni strávit Fluorescenční noc na Viničné, v budově slovné Přírodovědecké fakulty v Praze. Jako řešitelé semináře budete zvýhodněni při přijímání na toto setkání plné mikroskopování, výletů a přednášek. Absolventi Biozvěstu budou mít nárok získat tzv. Pas mládeže, evropsky uznávaný doklad získaných dovedností.

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 vyplňte. 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.

Svá řešení úloh nám pošlete 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_Přijmení,

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

Uzávěrka 2. série: pondělí 6.1.2014 o půlnoci.

V případě opožděného odevzdání úloh se strhává za každý

celý den jeden bod (rozhodující je den, kdy je zaslána poslední úloha) 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 svých 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“.

Jako správní vědeckí adeпти se snažte vždy o co nejučenější formulaci odpovědi, vždy popište všechny úvahy, jak jste ke konkrétním závěrům přišli. U netriviálních skutečností doporučujeme i uvést citaci, stačí jednoduše (např. „Wikipedie: Translace“) či stručný název publikace ze které čerpáte („Alberts, Zákl. bun. biol.“).

Formálně správně provedené zpracování hodnotíme zejména u praktické úlohy, rady a příklady naleznete 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 směřujte na adresy

biozvest@gmail.com či vosolsob@natur.cuni.cz

Mnoho zdaru při řešení Vám za kolektiv autorů přeje

Stanislav Vosolsobě

Úloha 1: Mykorhiza

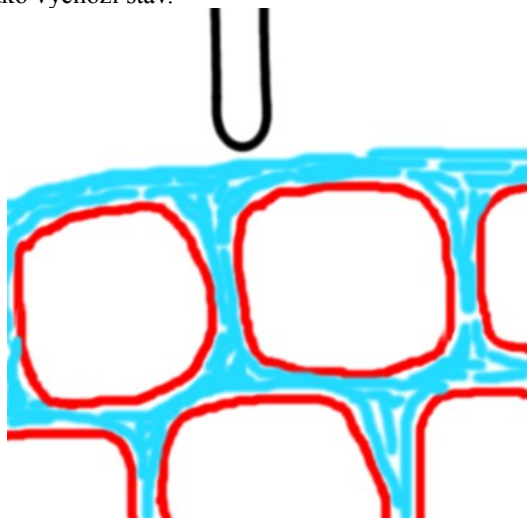
Autor: Stanislav Vosolsobě

Počet bodů: 12

Podzemní soužití hub a rostlin je víc, než jen že se potkaly dva organismy a dohodly se na společné cestě životem. Bez mykorhizy by totiž nebylo zřejmě ani hub, ani vyšších rostlin. Podobně jako po setkání lidoopa s vlkem vznikl člověk lovec a pes, přítel člověka.

1. Evolučně nejstarší mykorhiza je arbuskulární. Pokuste se zjistit, u kterých archaických rostlin již byla doložena a jaká skupina hub ji vytváří. Kde leží tato skupina na fylogenetickém stromě hub a co z toho vyplývá z hlediska evoluce hlavních skupin hub a rostlin?
2. Zástupci této skupiny hub jsou výhradně mykorhizní. Jaké problémy to přináší, pokud bychom chtěli tyto houby kultivovat (např. na živném médiu)?
3. U kterých rostlin nacházíme arbuskulární mykorhizu?
4. Mnohem mladší je ektomykorhiza. Jaká skupina hub a jaké rostliny se podílí na této symbiose? Vyjmenujte i konkrétní příklady druhů!
5. Do přibuzenstva ektomykorhizních hub patří oproti arbuskulárním mnoho dalších nemykorhizních. Jaké zdroje živin tyto houby využívají?

- Na obrázku vidíte situaci, kdy hyfy houby (černě) právě potkala kořen rostliny. Dokreslete do obrázku (stačí velice schematicky, třeba i v programu malování), co se stane po několika dnech, až se ustaví mykorhizní symbiosa. Nakreslete jeden obrázek pro variantu arbuskulární mykorhizy, druhý pro ektomykorhizu. Dbejte na správné zakreslení buněčných stěn (modře) a plasmatické membrány rostlinné buňky (červeně).
- To, že tečou sacharidy z rostlin k houbám výměnou za vodu a minerály je obecně platné. V jakých situacích se však směr toku sacharidů může obracet? Vymyslete jednak situaci vyskytující se v běžné ektomykorhizní symbiose kdy u daného druhu rostliny a houby ve většině ostatních situací tečou sacharidy směrem k houbám a pak příklady rostlin, které mají obrácený tok sacharidů jako výchozí stav.



Úloha 2: Geografem v říši zvířat I

Autor: Albert Damaška

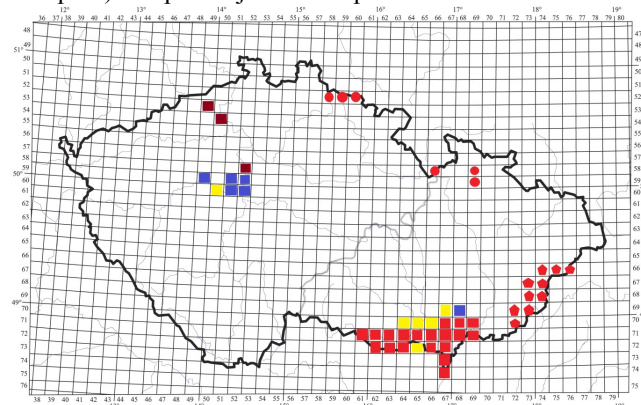
Počet bodů: 12

Jedním z přístupů, jakým zkoumat živočichy, je jejich výzkum na regionální úrovni. Výsledná data z takových výzkumů mohou posloužit k mnohým zajímavým vědeckým cílům, ale hodí se i pro účely ochrany přírody. Metoda regionálního zkoumání fauny, kterou se v této úloze budeme zabývat, se nazývá **faunistika**. Na tomto místě je dobré zmínit, že přístupů k regionálnímu výzkumu biodiversity je více a v botanice mají o něco větší vědeckou váhu, než v zoologii. My se ale v této úloze budeme zabývat jen zvířaty a faunistikou (a botanický protějšek, floristiku, necháme na jindy). Faunistika je metoda, která nám přináší data o rozšíření daného druhu v rámci jeho areálu (pak mluvíme o faunistickém mapování) nebo data o diversitě druhů na konkrétní lokalitě (pak jde o faunistický průzkum).

- Mapování je metodou velmi efektivní v případě, že do něj lze začlenit širokou veřejnost. Jaké vlastnosti musí mít mapované zvíře, aby se do mapování dalo začlenit co nejvíce lidí? Uveďte příklady takových zvířat či případně rovnou nějakých mapovacích programů ze současné doby, do kterých je začleněna informovaná veřejnost.
- Průzkum, na rozdíl od mapování, je spíše záležitostí dlouhodobé práce jednotlivce, případně více lidí, kteří

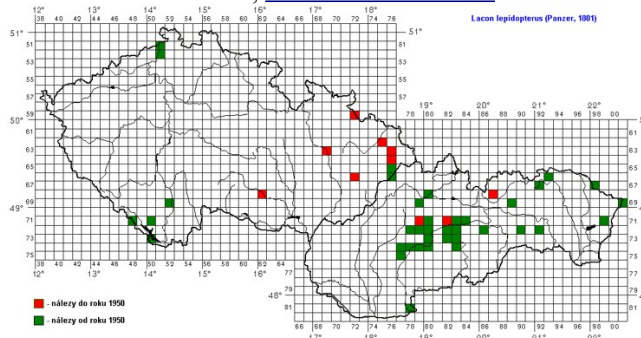
lokalitu dlouho navštěvují. K výzkumu jakých jevů je možné data z faunistických průzkumů použít? K čemu se výsledky průzkumů hodí prakticky (např. při ochraně přírody)?

- Výsledky faunistických mapování se často zana- menávají do tzv. síťových map. Jednu takovou hypote- tickou síťovou mapu pro tři různé druhy máte na prvním obrázku (jeden zobrazen čtverečky, druhý kolečky, třetí pětiúhelníky). Co můžeme o jednotlivých druzích z mapy vyčíst (ekologické nároky, charakteristika areálu apod.)? Doporučujeme si mapu na monitoru zvětšit.



- Barevné odlišení jednotlivých čtverečků má u „čtvereč- kového druhu“ zásadní význam. Červená barva znázor- ňuje čtverce, kde byl nalezen do roku 2009. Žlutá pak zobrazuje nové nálezy roku 2010, modrá nové nálezy roku 2011 a tmavě červená nové nálezy roku 2012. Jak lze vysvětlit náhlé rozšíření druhu? Je tento trend ojed- inělý, nebo se posun severní hranice rozšíření druhů v poslední době děje často? Pokud považujete za správnou druhou možnost, své tvrzení dokažte a uveďte několik druhů živočichů, které se ve 21. století v ČR takto rozšířily. Proč se asi nešířil „kolečkový druh“?

- Následující faunistická mapa je na rozdíl od té horní skutečná a mapuje výskyt kovařika *Lacon lepidopterus*. Vidíme, že tento kovařík byl po roku 1950 nově nalezen na několika místech v Čechách. Jde o druh, vázaný na staré stromy a mrtvé dřevo. Stojí snad za objevením druhu v Čechách to, že se po roce 1950 náhle rozšířil? Pokud soudíte, že nikoli, co by mohlo za zdánlivou dispersi stát? A pokud s náhlým rozšířením druhu sou- hlasíte, co by mohlo takové šíření způsobit? Zdroj mapy: Dušánek et Mertlik, www.elateridae.com.



- Faunistické články se obvykle nepublikují v impak- tovaných časopisech, ale nacházejí své místo v těch periodikách, která se zabývají otázkami regionálního charakteru. Někdy se jim říká *regionální sborníky*. Vy- chází ve vašem kraji také nějaké regionální pří-

rodovědné sborníky? Často je vydávají správy CHKO či jiné orgány AOPK ČR, nebo správy NP, případně regionální musea. Pokuste se ve vašem kraji nějaký takový sborník vypátrat a napište, jak se jmenuje. Vyšel v posledních letech v tomto sborníku nějaký faunistický článek? O čem byl a kdo byl autorem?

Úloha 3: Křídla

Autor: Michael Mikát, Jiří Hadrava

Počet bodů: 20

Hmyz je co do počtu druhů velmi bohatou skupinou živočichů. Přibližně tři čtvrtiny popsáných druhů všech organismů patří mezi hmyz. S výjimkou moří dokázal hmyz osídlit snad všechna zákoutí naší planety od chladné tundry přes vyhřáté pouště, hluboké jeskyně a dna horských potoků až po tropické deštné lesy. Hmyz je také nesmírně rozmanitý po stránce potravních strategií. U žádné jiné skupiny než u hmyzu nenajdeme tak komplexní sociální společenství, jaká známe například od mravenců či termitů.

Jednou z nejnápadnějších vlastností, kterou můžeme na hmyzu pozorovat, je schopnost aktivního letu. Hmyz je vůbec první skupinou, která kdy umění létat ovládla. První hmyz se do vzduchu vznesl již v období zvaném devon, tedy dávno před tím, než si letové schopnosti osvojili obratlovci (pterosauři, ptáci a netopýři). Křídla, která ovládaný let umožnila, se v evoluci hmyzu vyvinula pouze jednou. Toto tvrzení zároveň implikuje, že všechny dnešní okřídlené hmyz je potomkem tohoto prvního letce. Skupinu, resp. evoluční linii, do níž všechny křídlaté hmyz patří, nazýváme Pterygota. Ne všechna dnešní Pterygota však křídla využívají, mnohé skupiny patří mezi Pterygota o křídla druhotně přišly nebo je mají modifikované do podob, které jim slouží k jiným účelům než letu.



Kromě Pterygot do hmyzu patří i řada dalších, takzvaně primárně bezkřídle skupin hmyzu, tedy skupin bezkřídleho hmyzu, jejichž předci nikdy žádná křídla neměli (jejich poslední společný předek sdílený s kterýmkoli zástupcem křídlatého hmyzu žil v siluru, tedy dříve, než se předkům křídlatých křídla vyvinula). Primárně bezkřídle skupinami hmyzu jsou rybenky (Zygentoma) a chvostnatky (Archaeognatha). Mezi hmyz v širším slova smyslu můžeme zařadit i další primárně bezkřídle skupiny, jako jsou vidličnatky (Diplura), chvostoskoci (Collembola) a hmyzenky (Protura). Všechny tyto skupiny spolu s pravým létajícím hmyzem (Pterygota) po svém společném předku zdědili

svou šestinohost, proto evoluční linii, do níž všechna tato zvířata patří, nazýváme skupinou Hexapoda.

1. Vědci dlouho spekulovali nad tím, z jakého orgánu se křídla vyvinula. Jednou ze starších teorií je například, že vznikla postupným protahováním výrůstků (paranot) na boku hrudi. Dnes je ale na základě paleontologických dokladů a podkladů z evolučně-vývojové biologie považována za nejpravděpodobnější jiná hypotéza. Napište, z čeho se s největší pravděpodobností křídlo vyvinulo. Na formulaci této hypotézy se významně podílela paleontoložka českého původu žijící v Kanadě. Napište její jméno.
2. Křídla dnešních druhů hmyzu mají různý tvar - mohou být různě protažená, zkrácená či dokonce třepená. Rovněž se mohou lišit svým vzhledem - mohou být čirá či zbarvená, s řídkou či hustou žilnatinou, případně pokrytá chloupky nebo šupinkami. Nakreslete, jak si myslíte, že mohlo vypadat křídlo posledních společného předka všech dnešních Pterygot.

Většina zástupců hmyzu má dva páry křídel vyrůstajících z hrudi. U některých skupin došlo k redukci jednoho či obou párů křídel, u některých vymřelých skupin se naopak vyskytovaly tři páry křídel.

3. Hrud' hmyzu se skládá ze tří článků. Jak se nazývá článek, který křídla obvykle nese? A jak se nazývají články, které křídla obvykle nesou?
4. Tři páry křídel se vyskytovaly u některých zástupců jedné skupiny hmyzu, které žila v prvohorách - v karbonu a permu. Jak se tato skupina nazývá?
5. Tři skupiny hmyzu mají v současnosti pouze jeden pár křídel. Jmenujte tyto skupiny. Napište rovněž, který hrudní článek tato křídla nese.
6. U skupin, které jste vyjmenovali v otázce 5, je redukován pár křídel přece jen alespoň trochu zachovalý, byť nemůže rozhodně sloužit k letu. Jak se takovým redukováným křídlům říká a k čemu slouží?
7. Jestli na daném článku vyroste nebo nevyroste křídlo (ale rovněž i třeba to, zda a kde vyroste kráčivá končetina či tykadlo), ovlivňují geny z rodiny Hox. Tyto geny obecně ovlivňují pozice orgánů podél předozadní osy těla, a to nejen u hmyzu, ale u i většiny dalších živočichů včetně člověka. Jaký Hox gen konkrétně ovlivňuje, kolik párů křídel hmyzu vyroste?

U některých skupin hmyzu došlo k úplné ztrátě křídel, nebo jejich naprosté nefunkčnosti. To je obvykle spojené se způsobem života, při kterém by křídla překážela, nebo by bylo zbytečné a příliš nákladné je vytvářet.

8. Jmenujte dvě skupiny hmyzu, které ztratily křídla v souvislosti s parazitismem.
9. Jmenujte tři skupiny hmyzu, u nichž došlo k redukci křídel (či letových schopností) v důsledku skrytého způsobu života.
10. U řady skupin hmyzu je přítomnost křídel v rámci druhu polymorfni - tedy někteří jedinci křídla mají a jiní nikoli. Vyjmenujte alespoň tři důvody, proč může být pro druh výhodné, aby část jedinců křídla měla a část nikoli. Uveďte ke každému důvodu konkrétní příklad druhu, u něhož daný důvod hraje významnou roli.

U některých skupin hmyzu má jedno pohlaví výrazně lepší letové schopnosti než druhé (např. jedno je okřídlené a druhé nikoli).

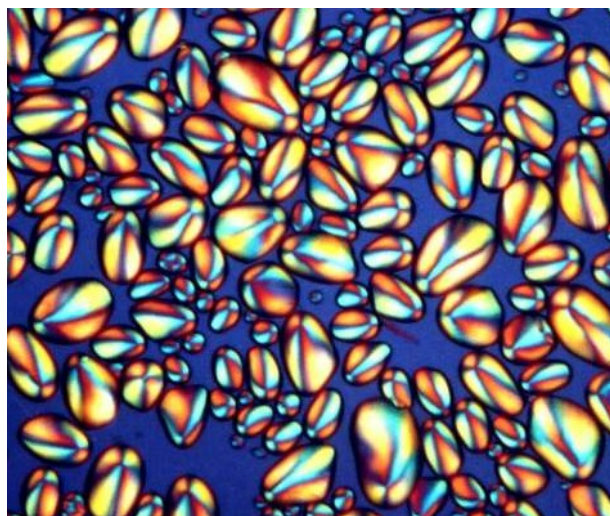
11. Které pohlaví bývá obvykle lépe letuschopné a proč tomu tak je?
12. U kterých skupin hmyzu tomu tak je?
13. Najdete nějaký druh hmyzu, který je výjimkou, tedy že má lepší letové schopnosti druhé pohlaví?
14. Naprostá většina druhů hmyzu má křídla pouze ve stádiu dospělce. U jedné skupiny hmyzu má ale křídla i poslední larvální stádium. Jak se toto stádium nazývá a u kterého hmyzího řádu se s ním můžeme setkat?

Úloha 4 (praktická): Hladovění

Autor: Stanislav Voslosobě

Počet bodů: 12

Během dne rostliny provádějí fotosyntezu a zajišťují si tak energii ve formě sacharidů. Avšak v noci fotosyntéza probíhat nemůže a rostliny musí mít vytvořenu dostatečnou zásobu sacharidů, které se ukládají jako škrobová zrna a v noci jsou odbourávána aerobním dýcháním, tak jako to dělají živočichové. Zatímco v zásobních orgánech, jako je například kořen, skladují rostliny energetické zásoby na dlouhou dobu, z listů jsou sacharidy vytvářené fotosyntézou průběžně odváděny a v listech se vytváří pouze nejnutnější zásoba nutná právě pro přečkání noci.



1. Vaším úkolem bude právě zjistit, jak dlouho vydrží list ve tmě bez fotosyntézy. Využijete skutečnosti, že škrob se dá v listu snadno detekovat látkou běžně dostupnou v lékárně. Jako experimentální materiál doporučujeme pokojovou rostlinu s měkkými listy (případně si vysejte třeba salát). Nechte rostlinu den na světle a potom během noci sledujte, zda-li se mění obsah škrobu. Druhý den nechte rostlinu ve tmě a zjistěte, zda se postupně zásoba zcela vyčerpá.
2. Abyste mohli škrob detekovat, musíte list poněkud macerovat a odstranit chlorofyl. popište podrobně celou vaši metodiku.
3. Škrobová zrna působí jako neživé krystaly. Z čeho však vznikají (popište na úrovni buněčné struktury).

Pokud byste váhali, jak správně sepsat protokol, podívejte se do sekce „Návody“ na webu Biozvěstu!

Úloha 5: R - funkce, podmínky a cyklus

Autor: Jiří Hadrava

Počet bodů: 6

V první sérii letošního Biozvěstu jsme se naučili základům práce se statistickým softwarem R. Naučili jsme se jej využívat jako **kalkulačku**, naučili jsme se využívat **proměnné**, naučili jsme se počítat nejen s jednotlivými čísly, ale i s celými číselnými **vektory** a ukázali jsme si několik základních statistických a grafických **funkcí**. Nyní se pokusme trochu teoretičtěji zamyslet nad tím, co vlastně přesně to slovo *funkce* znamená. Jistě jej znáte z hodin středoškolské matematiky a v obecnější podobě i z běžného jazyka. V Rku jsme se s ním poprvé setkali při popisu funkce `c()`, funkce, které když zadáme seznam čísel, vrátí nám vektor. Obecně je funkce něco, co nám na zadání určitého **vstupu** odpoví vrácením příslušného **výstupu**. Matematické funkce, které znáte ze středních škol, jsou předpisy, které vám na základě zadání určitého čísla (vstupu), nejběžněji značeného jako x , vrátí číslo jiné, jeho funkční hodnotu $f(x)$ (výstup).

Funkce v R je určitá část tohoto programu, která vám po zadání vstupu (*argumentu funkce*) vrátí výsledek určitého procesu, který daná funkce reprezentuje. Setkali jsme se již například s funkcí `mean()`, která po zadání vektoru čísel vrátí hodnotu jejich průměru, nebo s funkcí `boxplot()`, která po zadání vektoru čísel vrátí obrázek zobrazující jejich krabíkový graf. Všimněte si shody Rkového jazyka s jazykem matematickým, v obou případech se vstup funkce píše do kulatých závorek za název funkce a celek pak má význam výsledku procesu, který daná funkce s daným vstupem provede. Díky tomu pak můžeme výstupy z funkcí nejen nechat vypisovat, ale také třeba přiřazovat do proměnných nebo je řetězit s dalšími funkcemi tím, že výstup jedné funkce použijeme jako vstup jiné - např. `mean(c())`. Většina toho, co se v budoucnu budete v Rku dále učit, budou právě nové funkce a možnosti jejich propojování.

Jaké všechny vstupní argumenty lze které funkci zadat, nejlépe dohledáte v nápovědě pro konkrétní funkci. Argumenty funkcím můžete zadávat „jen tak“ jejich vypsáním do kulatých závorek. Jestliže ale funkce pracuje s větším množstvím různých vstupů, bylo by příliš náročné pamatovat si, v jakém pořadí se tyto argumenty zapisují. V těchto případech můžeme argument zadat tak, že napíšeme jméno argumentu, rovná se a hodnotu, kterou danému argumentu přiřazujeme. Vzpomeňte si, toto jsme v minulé sérii použili při přenastavování šířky sloupců v histogramu - `hist(1,breaks=c(1:32))`. Funkci `hist()` totiž kromě vstupních dat můžeme zadat také spoustu dalších informací o tom, jak má nastavit měřítka, jak má popsat osy, jakou barvou má histogram vykreslit apod. Nejčastěji však budeme chtít většinu těchto technických parametrů ponechat tak, jak jsou nastaveny automaticky, a když jsme minule potřebovali změnit právě a pouze šířky intervalů, do nichž se má vstupní vektor rozškatulkovat, bylo by značně nepraktické muset ve správném pořadí vypsát i hodnoty ostatních možných argumentů.

Nové funkce si také můžete sami vytvářet skládáním z funkcí existujících a z elementárních matematických operátorů. K tomu slouží funkce `function()`, jejíž výstup můžete uložit do proměnné, a tím z ní udělat novou funkci. Argumentem funkce `function()` jsou vstupy budoucí funkce. Proces, který vaše funkce s těmito vstupy provede,

napišete do složených závorek následujících za těmi kulatými. Pokud tedy chcete vytvořit například funkci, která ze vstupu vezme dva vektory, každý zprůměruje a jako výstup vrátí součin těchto průměrů, uděláte to takto:

```
moje_funkce <- function(x,y)
{
  z <- mean(x)*mean(y)
  return(z)
}
```

Případně lze celý proces zapsat i úsporněji bez využívání pomocné proměnné *z* takto:

```
moje_funkce <- function(x,y)
{
  return(mean(x)*mean(y))
}
```

Funkci pak můžete používat tak, jako každou jinou, např. příkaz

```
moje_funkce(c(1,2,3),c(3,3))
```

vám vrátí hodnotu 6.

Pokud má mít některý ze vstupů naší nové funkce určitou hodnotu automaticky předdefinovanou tak, jako jsme to viděli např. u technických parametrů grafických funkcí, tuto *defaultní* hodnotu argumentu zadefinujeme hned při vytváření funkce. Chceme-li tedy například, aby druhý vektor funkce `moje_funkce()` nebylo nutné zadávat a aby v případě jeho nezadání tato funkce pracovala např. s vektorem `c(2,3,4)` jako s vektorem *y*, funkci naší definujeme takto:

```
moje_funkce <- function(x, y = c(2,3,4))
{
  z <- mean(x)*mean(y)
  return(z)
}
```

Často se nám může hodit použít nějakou funkci či sérii příkazů opakovaně, k tomu využijeme tzv. **cykly**. Cykly jsou nástrojem velmi hojně užívaným programátory, často je však použijeme například při konstrukci nějakých složitějších biologických modelů.

Jakými všemi způsoby lze cykly v R utvářet, se dozvíte v nápovědě zadáte-li `?Control` (pozor, jak už jste možná zjistili, R rozlišuje velká a malá písmena, zde je použito písmeno velké, což u názvů funkcí bývá spíše vzácné). Zde ukážu pouze jeden z cyklů, a to tzv. **for-cyklus**. Slovo „for“ v názvu pochází z anglického výrazu odpovídajícímu českému „pro“, funkce `for()` totiž nese význam: „Následující sled příkazů proved' pro každou hodnotu v daném rozmezí“. For-cyklus prostě několikrát provede zadaný kus skriptu. Kolikrát jej provede zadáváme jako argument funkce ve tvaru

```
for(i in 1:100)
```

kde *i* je název **indexu**, tedy proměnné, pomocí které si cyklus počítá, kolikrát již příkazy provedly. Hodnoty indexu zadáme jako rozmezí hodnot, kterých bude index postupně nabývat. V našem případě tedy cyklus proběhne stokrát, poprvé pro hodnotu $i=1$, podruhé $i=2$ atd. až do posledního cyklu, stého běhu, v němž se bude *i* rovnat stu. Místo *i* lze samozřejmě použít i jiný název proměnné pro indexování, jiný název dokonce budeme muset nutně zvolit třeba v případě, když budeme potřebovat vnořit dva for-cykly do sebe (tzn. jeden for-cyklus spouštějící opakovaně nějaký příkaz

se bude opakovaně spouštět jiným for-cyklem).

Seznam příkazů, které for-cyklus pro každou z hodnot indexu provede, píšeme obdobně jako u funkce `function()` do složených závorek až za příkaz této funkce. V tomto sledu příkazů samozřejmě můžeme využívat i indexující proměnné, velmi se však doporučuje do ní žádná hodnota nezapisovat. V případě chyby bychom totiž mohli poměrně snadno vytvořit nekonečný cyklus, a počítač tak docela zavařit (např. příkaz `for(i in 1:3){i<-2}` by nám nikdy nedoběhl do konce.

Pokud budeme chtít uvnitř cyklu zapisovat data do nějakého vektoru, který před spuštěním cyklu neexistuje, nepůjde to. Je třeba si jej nejprve vytvořit příkazem

```
a = NULL
```

kde *a* je jméno proměnné, kterou jsme si tímto inicializovali. Jednoduchou ukázkou for-cyklu budiž následující kód:

```
a = NULL
for(i in 1:100)
{
  a[i] <- 2
}
```

Tento skript vytvoří vektor jménem *a* a ve sto krocích jej postupně od první do sté pozice zaplňuje hodnotami 2. Ačkoli je for-cyklus poměrně elegantní a univerzální nástroj, mnoho operací, k nimž bychom jej mohli využít, lze provést i bez něj. Např. výše vypsáný třířádkový skript bychom mohli nahradit jediným příkazem, pokud bychom znali funkci `rep()`. S tou bychom celý proces mohli zapsat jako `a <- c(rep(2,100))`.

Druhým užitečným nástrojem, kterého jste i již v nápovědě `?Control` mohli všimnout, jsou **podmínky**. Ty nám umožňují na základě nějakého vstupního stavu nebo předchozího výsledku rozhodnout, jaký příkaz (nebo série příkazů) má být provedena. Funkce, která splněnost podmínky kontroluje, se nazývá `if()` (z anglického „když“). Seznam příkazů, které mají být provedeny, je-li podmínka splněna, se opět zapisuje do složených závorek následujících po funkci. Do argumentu funkce zadáváme, jaká podmínka má být splněna. Chceme-li například spočítat průměr z vektoru *a*, ale to pouze v případě, že jeho první prvek má hodnotu větší než dva, uděláme to příkazem

```
if(a[1]>2)
{
  mean(a)
}
```

Pokud se má nějaký jiný příkaz provést v případě, že podmínka splněna není, po uzavření složených závorek napíšeme `else` (z anglického „jinak“) a alternativní příkaz zadáme do dalších složených závorek následujících za tímto slovem, např. tedy

```
if(a[1]>2) mean(a) else min(a)
```

(pokud je první prvek vektoru *a* větší než 2, vypíše průměr tohoto vektoru, pokud ne, vypíše jeho nejmenší hodnotu; zde jsme si také ukázali příklad, kdy za příkazem pro funkci stojí jen jeden příkaz pro vykonání - potom není nutné jej uzavírat do složených závorek). Jako podmínku můžeme tak, jako v ukázce výše, poměřovat, zda je hodnota větší (>), menší (<), či rovna (== pozor, zde se píše rovnítka dvě). Tato znaménka můžeme také kombinovat (např. `>=` znamená je větší nebo rovno). Chceme-li vyjádřit nerovnost,

uděláme to pomocí znaků `!=`. Můžeme také kombinovat více podmínek dohromady, symbolem pro logický operátor `a` je `&` a symbolem pro *nebo* je `|`. Složitější podmínka pak může mít třeba podobu

```
if (a[1]<=2 & a[2]!=a[3])
{
  mean(a)
}
```

Na závěr si ještě ukážeme několik praktických funkcí. Funkce `length()` vrací hodnotu délky vektoru, který se této funkci zadá jako argument. Funkci `ls()` lze použít bez jakéhokoli argumentu a jejím výstupem je seznam všech objektů, které jste si vytvořili. Funkce `unique()` vrací hodnoty ze zadaného vektoru tak, že každou napíše pouze jednou. Funkcí `round()` můžete zaokrouhlovat čísla, funkcí `sort()` můžete přeházet čísla ve vektoru tak, aby byly seřazeny od nejmenšího po největší, funkce `rank()` zase místo každého z prvků vstupního vektoru vrátí pořadové číslo pozice, na které by daná hodnota ležela, kdyby byl vektor seřazen. Pokud některé z těchto funkcí budete někdy potřebovat, najděte si podrobnosti v nápovědě.

1. Napište funkci, která jako vstup přijme vektor, číslo a parametr n . Funkce přepíše n -tou pozici vektoru zadaným číslem a z výsledného vektoru spočítá průměr a rozptyl, které vypíše jako výstup. Pokud parametr n nezadáme, funkce bude přepisovat první pozici ve vektoru.
2. Napište funkci, jejímž vstupem bude jedno celé číslo (n) a výstupem vektor, který bude obsahovat právě všechna prvočísla od 1 do n . Možná se vám k řešení tohoto úkolu bude hodit matematická operace *modulo*, jejíž implementace v R je následující: Příkaz `x%%y` vám vrátí zbytek po dělení čísla x číslem y .

Pokud vám bude R při řešení úkolů hlásit chyby (což je velmi pravděpodobné), snažte se z chybové hlášky vyčíst, kde ve vašem skriptu nastává problém a daný úsek opravit. V případě, že to nepůjde, doporučuji zkusit vymyslet nějaký jiný postup řešení. Pokud by nic z toho váš problém neřešilo a nevěděli jste si rady, můžete se na mě obrátit, poslat mi e-mailem na hadravajirka_zavinac_seznam.cz svůj skript a pokusím se vám poradit.



Úloha 6: Jak vzniká rozdělení veličin v přírodě?

Autor: Jan Smyčka

Počet bodů: 12

V minulém díle seriálu o statistice jsme se zabývaly tzv. popisnou statistikou, tedy na modelovém příkladě lesa otázkou

„jak je vysoký les“. Komprimace dat pomocí průměrů, směrodatných odchylek a dalších charakteristik je velmi užitečný nástroj, nicméně vynechává a dokonce trochu zastírá podstatu pro vědce mnohem palčivější otázky: Jak vznikají stromy, které potom můžeme změřit a vytvořit z nich data? Jak lze popsat proces, na jehož počátku jsou podmínky pro růst a na konci je výška stromu?

Představme si, že jsme lesníci: máme na starost les a víme například, že strom rostoucí na kamenité půdě bývá obvykle vysoký 10 metrů, ale na jílovém podkladě jsou vesměs o 2 metry vyšší stromy než na půdě kamenité. Pomocí této znalosti bychom mohli předpovídat výšku stromu rovnici, pokud víme v jakých půdních poměrech strom roste:

$$\text{výška_stromu} = 10 + 2 * \text{jílovitost}$$

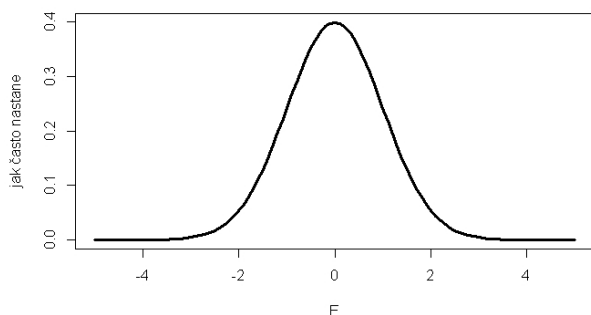
kde „jílovitost“ je proměnná dosahující hodnoty 0 pro strom rostoucí v kamenité půdě a 1 pro strom rostoucí na jílu¹. To co se vyskytuje na levé straně rovnice však není přesně výška stromu, ale nějaký její odhad. Ostatně dva stromy rostoucí ve stejné půdě budou sice podobně velké, ale úplně stejně nejspíš ne. Proto by se slušelo napsat spíš:

$$\text{výška_stromu} = 10 + 2 * \text{jílovitost} + E$$

kde E je nějaké nám neznámé číslo (může být i záporné) vyjadřující, jak se **náhoda** uplatnila na výšce každého konkrétního stromu. Jak vzniká tato náhoda je do značné míry filosofická či dokonce teologická otázka. Může se jednat o nějaké procesy v lese, o kterých nic nevíme, i když jsou teoreticky předpověditelné (například množství srážek, sluneční osvit, chemismus půdy a spousta dalších), chyba způsobená nepřesností měření, nebo může jít o skutečnou náhodu ve smyslu „vyšší moc chtěla, aby tento strom byl vyšší“. Ať už za existencí této náhody stojí cokoli, její důležitou vlastností je pro nás to, že u jednoho opakování (například konkrétního stromu) nejsme schopni dopředu odhadnout její hodnotu. Nicméně pokud máme opakování více, tedy například jsme změřili celý les stromů, obvykle zjistíme, že některé hodnoty E nastávají častěji, než jiné. Můžeme pak přibližně určit její **pravděpodobnostní rozdělení**, tedy funkci ukazující jak často nastávají které konkrétní hodnoty E , pokud bychom je zkoumali na nekonečně mnoha (pro představu stačí velmi mnoha) stromech.

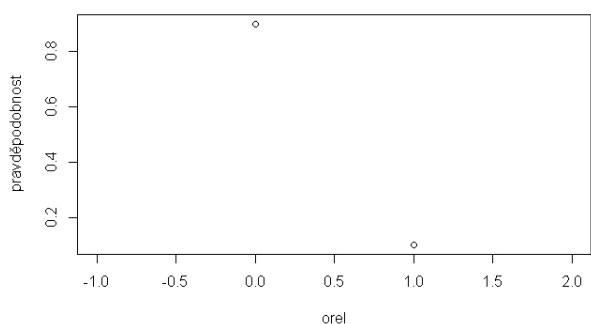
V praxi například velmi často platí, že většina číselných hodnot E je nahromaděna blízko 0 a jen velmi málo jich „ulítné“ daleko od ní (pro pravděpodobnostní rozdělení takového E viz obr 1). Pokud by tomu tak bylo v případě našeho lesa, byla by to dobrá zpráva, protože by znamenala, že s použitím naší předpovědní rovnice pro určení výšky stromu uděláme nejčastěji velmi malou chybu a k velkým chybám dojde jen občas. Nicméně v tuto chvíli není úplně zřejmé, proč by náhodná odchylka ve výškách stromů takovouto vlastnost měla mít. Abychom si to ozřejmili, musíme si nejdříve povědět něco o různých typech náhody a pravděpodobnostních rozděleních, která se k nim vážou.

¹ Analogicky bychom mohli na pravou stranu přičítat další známé efekty, například funkci popisující, jak roste výška stromu se srážkami, nebo osluněním.



Obr 1: Funkce zobrazující, jak často nastanou které odchylky velikosti stromu od našeho modelu. Nejčastější jsou v tomto případě odchylky blízké nule.

Když se řekne náhoda, většina lidí si představí házení mincí (nebo kostkou, což je složitější příklad téhož). Pro házení mincí platí, stejně jako u náhodné složky výšky stromu, že u konkrétního hodu nejsme schopni určit jeho výsledek, nicméně pokud budeme házet mincí delší dobu, získáme určitou představu, jak často padá panna a jak často orel. Pokud máme spravedlivou minci, budou panny a orlové přibližně v poměru 1:1, pokud si necháme rub odlít z olova a líc z hliníku, může být poměr výrazně vychýlený, například 1:9. Každopádně, pokud máme (ideálně nekonečně dlouhou) zkušenost s házením nějakou mincí, můžeme si k ní sestavit funkci pravděpodobnostního rozdělení (někdy také zvanou distribuční funkce). Ta bude definovaná pouze v hodnotách 0 (panna) a 1 (orel) a každé z těchto hodnot bude přiřazena pravděpodobnost, že padne (viz obr 2). Takovému pravděpodobnostnímu rozdělení se říká **alternativní** a je určeno jedním parametrem (p), kterým je pravděpodobnost, že nastane jedna z možností (můžeme se arbitrálně rozhodnout která).



Obr 2: Distribuční funkce zobrazující pravděpodobnosti jevů (panna, orel) z alternativního rozdělení, kde pravděpodobnost panny je 0.9 a pravděpodobnost orla 0.1.

Jak by to vypadalo, pokud bychom alternativní rozdělení chtěli použít pro modelování stromů v lese? Představme si, že kromě typu podkladu výšku stromů v našem lese určuje jediný náhodný faktor (ostatní buďto zanedbáme, nebo máme vpravdě pohádkový les). Tím faktorem je, že včera přišla vichřice a některé stromy zlomila. Pravděpodobnost zlomení stromu je 0.3 a zlomený strom je vždy o pět metrů menší. Rovnice pro odvození výšky stromu bude mít tvar:

$$\text{výška_stromu} = 10 + 2 * \text{jilovitost} - 5 * \text{zlomený}$$

kde proměnnou „zlomený“ nabývající hodnoty 0 a 1 pro jednotlivé stromy neznáme, ale víme, že pokud budeme mít velké množství stromů, bude v 70% případů nabývat hodnoty 0 a ve 30% hodnoty 1.

Jako první úkol si můžeme zkusit namodelovat les po bouřce v Rku. Začneme tím, že si vytvoříme prázdný vektor les.

```
les = NULL
```

Dále si vytvoříme vektor `jil` určující, které stromy v našem lese budou růst na jílovitěm a kamenitěm podkladě (návod na jeho vytvoření je v předchozí sérii). Ten bude obsahovat 5krát hodnotu 1 a pětkrát hodnotu 0. A spočteme si, jak vyřezal les před bouřkou.

```
les <- 10 + 2 * jil
```

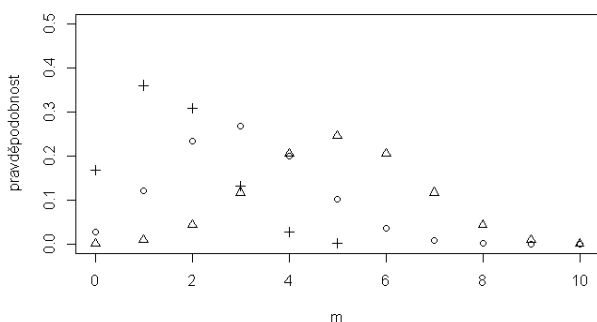
Nyní vytvoříme vektor bouřky. Aby si vědci nemuseli házet mincí, má v sobě Rko zabudované tzv. generátory náhodných čísel. Z alternativního rozdělení se dá generovat pomocí příkazu

```
rbinom(n, size, prob)
```

kde argument `n` určuje počet hodnot, které chceme generovat (v našem případě 10), argument `prob` pravděpodobnost se kterou nastane jednička (pro nás 0.3). Argument `size` musí být nastaven na hodnotu 1, pro jiné hodnoty generuje příkaz z binomického rozdělení, které je složitějším příbuzným alternativního (viz dále). Sami prozkoumejte, jak generátor funguje (jistě se vám bude hodit funkce `hist()`, která maluje histogramy, viz předchozí série). Potom s jeho pomocí vygenerujte vektor `zlomeny`, který bude obsahovat deset hodnot, které budou s pravděpodobností 0.3 rovné 1, jinak 0. Pomocí vektoru `zlomeny` modifikujte `les` na les po vichřici.

```
les <- les - 5 * zlomeny
```

V předchozím textu jsme se již zmínili o **binomickém rozdělení**. Jedná se o rozdělení pravděpodobnosti, že v n -tici pokusů z alternativního rozdělení (např. hodů mincí) bude m výsledků jednoho typu (např. orlů). Například pokud budu házet spravedlivou ($p=0.5$) mincí dvakrát ($n=2$), můžu určit, že pravděpodobnost dvou ($m=2$) orlů je 0.25, jednoho ($m=1$) orla 0.5 a žádného ($m=0$) orla 0.25. Distribuční funkce binomického rozdělení na rozdíl od alternativního není definovaná pouze pro 0 a 1, ale pro všechna přirozená čísla mezi 0 až n (ve dvou hodech mincí může nastat žádný, 1 nebo 2 orlí) a je určena počtem pokusů (n) a pravděpodobností výsledku v jednotlivém pokusu (p). Pro ukázkou distribuční funkce binomického rozdělení s různými parametry viz obr 3. V Rku se z binomického rozdělení generuje opět pomocí funkce `rbinom()`, přičemž argument `size` určuje počet pokusů (n). Sami si vyzkoušejte, jak generátor funguje.



Obr 3: Distribuční funkce binomického rozdělení s různými parametry. Kolečka mají parametry $n=10$, $p=0.3$ (pravděpodobnosti, že m stromů z deseti zlomí vichřice); trojúhelníky $n=10$, $p=0.5$ (pravděpodobnosti, že m krát z deseti hodů mincí padne orel); křížky $n=5$, $p=0.3$ (pravděpodobnosti, že m stromů z pěti zlomí vichřice). Všimněte si, že jednotlivé distribuční funkce nejsou definovány pro hodnoty větší, než n .

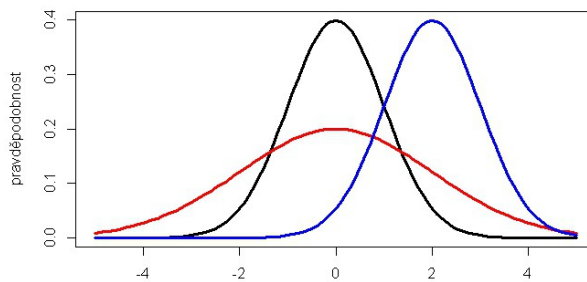
Dalším rozdělením, které je extrémně důležité pro biologa je **Poissonovo rozdělení**. Jedná se vlastně o speciální příklad binomického rozdělení, kde n je velmi (nekonečně) velké a p je velmi (nekonečně) malé. Jeho distribuční funkce je definována pro všechna přirozená čísla a 0 (tedy opět vlastně 0 až n) a její tvar je určen parametrem λ , který je odpovídá střední hodnotě (průměrné hodnotě, které veličiny z takového rozdělení nabývají), rozptylu tohoto rozdělení a zároveň těžko představitelnému součinu „nekonečně malého“ a „nekonečně velkého“ čísla $n \cdot p$. V praxi se Poissonovo rozdělení využívá k popisu řídkých jevů. Například jak často člověk za život pokálí pták, když jednotlivých okamžiků v životě (n) je nekonečně mnoho, ale pravděpodobnost pokálení v každém okamžiku (p) je velmi malá. Na tomto příkladě si lze dobře představit, že různí lidé žijí různě dlouho a různě často nad nimi létají ptáci, takže jednotlivé součiny $\lambda = n \cdot p$ se mohou pro různé lidi lišit, i když jde vždy o součin „nuly“ a „nekonečna“. Poissonovo rozdělení má také široké využití v prostorové statistice, kde se využívá kupříkladu k modelování veličin jako je počet bodů v buňce při náhodném pokládání bodů do jednotlivých buněk v prostoru rozděleném mřížkou. V Rku se jeho generátor volá `rpois(n, lambda)`. Opět si sami vyzkoušejte, jak funguje.

Již jsme se zmínili o celkem velkém množství pravděpodobnostních rozdělení, nicméně stále ještě jsme si neprozradili, jak vzniká to, které vidíme na obrázku 1. A protože nevíme, jak vzniká, nedokážeme se ani rozhodnout, jestli je vhodné pro popis náhody v našem lese. Na příkladu alternativního a binomického rozdělení jsme si ukázali, že pravděpodobnostní rozdělení lze konstruovat tak, že se vezme proces jednodušší (s alternativním rozdělením), výsledky se posčítají a součet má rozdělení složitější (binomické). Obdobně se matematici pokusili zkonstruovat rozdělení, které vzniká, když se udělá průměr (což je vlastně opět součet, jen dělený počtem sčítanců) z velmi (nekonečně) mnoha výsledků z alternativního rozdělení o nějaké konkrétní hodnotě p . A zjistili, že rozdělení, které takto vzniklo má velmi zajímavé vlastnosti.

První z nich je, že není definované pouze pro některé hodnoty (třebas přirozená čísla), ale pro všechna reálná čísla.

Jeho distribuční funkce je tedy spojitá (nejedná se jednotlivé body, ale je to křivka, kterou lze namalovat jedním tahem). To nakonec není příliš překvapivé a pro racionální čísla, tedy všechna čísla zapsatelná zlomkem, se dokonce dá snadno odvodit, proč tomu tak je.

Daleko zajímavější vlastností tohoto rozdělení je, že jej (nebo velmi podobné rozdělení, lišící se pouze střední hodnotou a rozptylem, viz dále) získáme pokud zprůměrujeme výsledky procesů i z jiného rozdělení, než alternativního, za určitých předpokladů². Dokonce platí, že můžeme v tomto průměru namíchat i výsledky z různých rozdělení a stále získáme stejné (nebo opět velmi podobné) rozdělení, jaké vzniká průměrem nekonečně mnoha výsledků z alternativního rozdělení³. Toto extrémně zajímavé rozdělení se nazývá **normální**. Je určeno dvěma parametry, μ a σ^2 , která odpovídají střední hodnotě a rozptylu rozdělení. V praxi μ určuje pozici vrcholu zvonovité křivky distribuční funkce a σ^2 její roztažení do stran (viz obr 4). Generátor normálního rozdělení se v Rku volá příkazem `rnorm(n, mean, sd)`, kde `mean` určuje μ a `sd` určuje σ .



Obr 4: Distribuční funkce normálního rozdělení s různými parametry. Černá křivka je pro $\mu=0$, $\sigma^2=1$; modrá pro $\mu=2$, $\sigma^2=1$; červená pro $\mu=0$, $\sigma^2=1.41$.

Normální rozdělení má velmi široké využití právě z toho důvodu, že je možné na něj pohlížet jako na průměr mnoha procesů z jiných rozdělení. Představme si nyní znovu náš les. Je velmi pravděpodobné, že náhodný prvek (E) ve velikosti stromu vznikl kombinací mnoha nám neznámých efektů a událostí, jako byly výše zmíněné bouřky, ale třeba i kůrovcové kalamity nebo fakt, že některé stromy (a nevíme které), rostou na více osluněných stanovištích. Pokud věříme, že těchto událostí bylo velmi mnoho a že jejich efekty je možné (alespoň povětšinou) sčítat, je skutečně rozumné popisovat náhodu (E) pomocí normálního rozdělení. Takové normální rozdělení samozřejmě nebude mít $\mu=0$, protože je velmi nepravděpodobné, aby se výše zmíněné efekty v průměru vyloučily. Nicméně toto μ je již součástí naší znalosti o běžné výšce stromu v lese (v příkladu 10 m), takže je v rovnici již zahrnuto a u E skutečně můžeme uvažovat $\mu=0$.

1. První úlohou je generování lesa po bouřce, viz text. Řešením se rozumí kompletní kód, vygenerované vektory a histogramy stromů rostoucích na jilu a na kamenité půdě.

² Tímto předpokladem je konečný rozptyl původních rozdělení.

³ Tuto vlastnost normálního rozdělení je možné odvodit na základě tzv. Centrální limitní věty.

2. Krátce zdůvodněte, proč může náhodná veličina z normálního rozdělení nabývat hodnoty jakéhokoliv racionálního (zlomkem zapsatelného) čísla. Nechceme exaktní matematický důkaz, ale jen krátkou úvahu na toto téma, která bude dávat smysl.
3. Pokuste se vygenerovat les s pěti stromy na kamenité půdě a pěti na jílu s chybami z normálního rozdělení se střední hodnotou 0 a rozptylem 1. Řešením se rozumí kompletní kód, vygenerované vektory a histogramy stromů rostoucích na jílu a na kamenité půdě.
4. Napište funkci (viz další část návodu na používání Rka), která bude vytvářet les o zadané základní výšce, počtu stromů, poměru stromů na jílu a kamenité půdě a síle náhody (rozptyl normálního rozdělení), a navíc namaluje histogram.
5. Vytvořte průměr z 1000 výsledků procesu z alternativního rozdělení s $p=0.5$. Takovýchto průměrů vytvořte 1000 a jejich histogram srovnajte s histogramem 1000 hodnot z normálního rozdělení se střední hodnotou 0.5 a rozptylem 0.25.

