

Zápočtový domácí úkol
Pravděpodobnost a statistika (NMAI059)
Zimní semestr 2012-13

Simulace opravování zápočtových úloh

Dne 7. ledna byla zveřejněna zápočtová úloha na cvičení z předmětu Pravděpodobnost a statistika (NMAI059). Studenti mají na vypracování čas do 11. března, tedy 64 dní. Cvičení má tři paralelky, které mají 15, 20 a 25 studentů. Každá paralelka má svého cvičícího a příslušní studenti mu posílají úlohy emailem. Bylo vyzorováno, že v každé paralelce se čas odevzdání úlohy prvního studenta (tj. studenta, který odevzdá úkol jako první z celé paralelky) řídí rovnoměrným rozdělením na celém časovém intervalu. U druhého studenta se čas odevzdání řídí rovnoměrným rozdělením v době mezi odevzdáním prvního studenta a koncem. Obecně, u n -tého studenta se čas odevzdání řídí rovnoměrným rozdělením mezi odevzdáním $(n - 1)$ -ního studenta a koncem. Předpokládáme, že nakonec každý student úlohu odevzdá (v řádném termínu).

Doba, za kterou cvičící opraví úlohu, se řídí exponenciálním rozdělením se střední hodnotou dva dny. Cvičící opravuje vždy jen jednu úlohu. Když ji opraví, pokračuje na další v pořadí nebo počká, než mu přijde další úloha.

1. Předpokládejme, že každý cvičící opravuje úlohy svým studentům sám. Simulujte každoroční odevzdávání a opravování úloh během 60 let historie školy. Na základě výsledků svých simulací řešte následující úkoly:

- (a) Znázorněte a okomentujte rozdělení doby do ukončení opravování všech úloh celého ročníku.
- (b) Pro každou paralelku zvlášť odhadněte střední dobu do dokončení opravování a čas, kdy již budou všechny úlohy opraveny s pravděpodobností 90 %.
- (c) Pro každou paralelku zvlášť odhadněte pravděpodobnost, s jakou se podaří opravit všechny úlohy do 30 dnů po termínu odevzdání.
- (d) Pro každou paralelku zvlášť otestujte, zda je očekávaná doba do konce opravování významně vyšší než 30 dní po termínu.
Nezapomeňte uvést uvažovaný model, jakou hypotézu a jakou alternativu uvažujete, jaký test volíte a zda jsou splněny jeho předpoklady (příp. proč porušení předpokladů nevedí). Výsledky testů řádně interpretujte.
- (e) Odhadněte procento studentů, kteří pošlou úlohu během posledního dne (posledních 24 hod) před termínem odevzdání.

2. Cvičící se rozhodli, že budou úlohy opravovat společně, a došlé úlohy si proto seřadí do jedné společné fronty. Studenti úlohy posílají na email svému cvičícímu, odkud jsou obratem přeměrovány na společný email, kde se tvoří společná fronta. Cvičící odebírají úlohy z této fronty postupně, když úlohu doopraví, jdou na další. Rozdělení dob odevzdání je stejné jako v části 1. Proveďte novou simulaci pro tuto situaci a na základě výsledků řešte následující úkoly:

- (a) Znázorněte rozdělení doby do ukončení opravování všech úloh celého ročníku. Porovnejte výsledek s 1(a) a okomentujte.
- (b) Odhadněte pravděpodobnost, s jakou se všechny úlohy podaří opravit do 30 dnů po termínu.
- (c) Otestujte, zda se tímto postupem podaří úlohy opravit dříve, než když každý cvičící opravuje sám.
Nezapomeňte uvést uvažovaný model, jakou hypotézu a jakou alternativu uvažujete, jaký

test volíte a zda jsou splněny jeho předpoklady (příp. proč porušení předpokladů nevádí). Výsledek testu řádně interpretujte.

- (d) Studenty zajímá doba čekání na navrácení úlohy (tj. doba, která uplyne mezi tím, co student úlohu pošle, a než ji dostane opravenou). Zkonstruuje meze, mezi nimiž se tato doba čekání nachází s pravděpodobností 90 %.

Pozn.: Úloha není zadána jednoznačně – uveďte libovolné možné řešení.

- (e) Sestrojte 90% interval spolehlivosti pro střední hodnotu doby čekání na navrácení úlohy. *Jaké jsou předpoklady konstrukce tohoto intervalu? Jsou ve Vašem případě splněny? Ověřte alespoň graficky.*

Z R se Vám mohou hodit funkce `for`, `runif`, `rexp`, `hist`, `mean`, `quantile`, `qnorm`, `qt`, `t.test`, ... Nápovědu ke všem funkcím získáte příkazem `help(jmenofunkce)` nebo `?jmenofunkce`.

Pokyny k vypracování

- Vaše řešení musí obsahovat zejména **komentář** vašich statistických úvah s relevantními výstupy simulací a jejich interpretací – **maximálně** 3 strany textu + obrázky. Velikost použitého fontu (v hlavních částech dokumentu) se musí pohybovat od 10 do 12 pt.
- Komentář řešení vyžadujeme v **souvislých** větách. Celý dokument musí být (včetně popisek tabulek a grafů) napsán konzistentně v jednom jazyce – čtete česky, slovensky, anglicky. Dokument tvořte v čem chcete (doporučujeme \LaTeX , ale s MS Wordem či OO Writerem to také asi půjde), ale nám pošlete jenom **pdf**! Obrázky zakomponujte do dokumentu, matematiku sázejte v kvalitě běžné v 21. století.
- Dále nám zašlete **váš okomentovaný kód** použitý k analýzám a malování obrázků. Abychom si mohli vaše výsledky ověřit, nastavte `set.seed(ddmmyyyy)` podle vašeho data narození před prvním generováním náhodných čísel. Tento použitý `seed` nezapomeňte uvést ve svém řešení.
- Soubor s R kódem nazvěte `prijmeni_jmeno.R`, soubor s hlavním komentářem nazvěte `prijmeni_jmeno.pdf`. V názvech souborů **nepoužívejte** diakritiku! Oba soubory pošlete **e-mailem** svému cvičícímu. Mějte na paměti, že v dnešní době odeslání e-mailu ještě neznamená jeho přijetí adresátem. Cvičící v přiměřené době obvykle potvrdí přijetí (maily však nečteme kontinuálně 24 hodin denně a často ani o víkendech).
- **Deadline** pro odeslání práce je **pondělí 11. března 2013 (23:59 CET)**. Práce došlé po tomto datu budou mazány.
- Jedná se o **samostatnou** práci. Pokud usoudíme, že se některá řešení navzájem příliš podobají, odměníme autory bonusem minus 30 bodů.
- Za práci je možno získat maximálně 40 bodů, přičemž 20 bodů bude uděleno za statistiku, 10 bodů za komentář k R-kovému kódu a 10 bodů za celkový dojem z práce.