

Теме за семинарски из Увода у статистику

Април 2019

1. Оцењивање моде расподеле. Представити оцену из литературе и емпиријски испитати њену непристрасност и постојаност.

[Литература](#)

2. Упоредјивање моћи Вилкоксоновог и т-теста (кад су задовољени услови за т-тест и кад нису).

3. Конструкција различитих интервала поверења и њихово упоређивање. Циљ је утврдити какви се интервали поверења добијају када се за њихову конструкцију користе различите стожерне величине, с једне стране, као и како се мења дужина интервала када се његове границе мењају.

4. Статистичка анализа музичких композиција

[Литература 1](#)

[Литература 2](#)

5. Осмислите истраживање (и спроведите га) чији је циљ да се утврди како на "сналажење" на Математичком факултету утиче врста средње школе из које су дошли, као и то да ли су у средњој и (или) основној похађали додатну наставу.

6. Осмислите истраживање (и спроведите га) чији је циљ да се утврди веза између следећих обележја посматраних на студентима Математичког факултета: пролазност на испиту, оцене, имунитет, друштвени живот, унутрашње задовољство.

7. Цензурисани подаци ([линк: Chapter 11](#))

Објаснити шта је десно цензурисан узорак. Приказати како се у том случају оцењује функција расподеле посматраног обележја и неки нумерички параметри расподеле. Илустровати примером (уз графички приказ).

Све описано применити на неким реалним подацима.

8. Цензурисани подаци ([линк: Chapter 11](#))

Објаснити шта је лево цензурисан узорак. Приказати како се у том случају оцењује функција расподеле посматраног обележја и неки нумерички параметри расподеле. Илустровати примером (уз графички приказ).

Све описано применити на неким реалним подацима.

9. Оцена густине језгрима. (Линк, 15.3.)

Описати метод и илустровати самосталним примером.

10. Линк

Да ли нормализација и стандардизација података у линеарној регресији мења резултате? Објаснити ту појаву на случају са једним и више предикторима.

11. Проверити Муртонову теорему (*Morton's theorem*) на генерисаним узорцима. Објаснити резултате.

12. Циљ истраживања је да се закључи који од три класична теста (Колмогоров–Смирновљев, Крамер–вон Мизесов и Андерсон–Дарлинггов) експоненцијалности са непознатим параметрима, против алтернативних расподела (Вејбулове (1.4) и Гама расподела $\Gamma(2, 1)$ и $\Gamma(4, 1)$) је најмоћнији. Емпиријске моћи оценити за обим узорка $n = 10$, $n = 20$ и $n = 50$. Дискутовати резултате.

13. Циљ истраживања је да се закључи који од три класична теста (Колмогоров–Смирновљев, Крамер–вон Мизесов и Андерсон–Дарлинггов) нормалности са непознатим параметрима, против алтернативних расподела (Логистичке $\mathcal{L}(0, 1)$ и Кошијеве $\mathcal{C}(0, 1)$) је најмоћнији. Емпиријске моћи оценити за обим узорка $n = 10$, $n = 20$ и $n = 50$. Дискутовати резултате.

Напомена: Потребно је направити одговарајуће генераторе Нормалне, Логистичке и Кошијеве расподеле на основу Униформне расподеле. Поред тога потребно је испрограмирати функције којим се рачунају потребне тест статистике.

14. Линк 1

Тестови симетрије. Одабрати три теста симетрије око координатног почетка. Испитати да ли расподеле тест статистика, под нултом хипотезом, зависе од полазне расподеле обележја. Затим, одредити емпиријске моћи одабраних тестова у случају неколико алтернатива које су симетричне око неке друге тачке, за различите обиме узорка. Један пример алтернативе је нормална $\mathcal{N}(m, \sigma^2)$ расподела када је $m \neq 0$.

15. Поређење методе момената и методе максималне веродостојности за оцењивање.

За поређење одабрати обележје X из померене експоненцијалне расподеле са густином

$$f(x) = \lambda e^{-\lambda(x-\mu)}, \text{ за } x \geq \mu.$$

- а) Испитати (коришћењем Монте Карло метода) колико су оцене отпорне на присуство аутлајера.
- б) Испитати (коришћењем Монте Карло метода) који метод нам даје бољу оцену у средње квадратном смислу.
16. Једна од функција које једнозначно одређују расподелу је и момент генераторна функција. Дефинисана је са $M_X(t) = Ee^{tX}$. Њена непристрасна оцена (показати то), на основу простог случајног узорка је $M_n(t) = \frac{1}{n} \sum_{i=1}^n e^{tX_i}$. Одредити момент-генераторну функцију случајне величине са Пуасономовом $\mathcal{P}(\lambda)$. На основу тога предложите неку тест статистику за тестирање H_0 : обележје X има $\mathcal{P}(\lambda_0)$ расподелу, као и облик критичне области за тестирање. Одредити емпиријске моћи теста, за $\alpha = 0.05$, када је $\lambda_0 = 1$ алтернативна расподела исто $\mathcal{P}(1)$ расподела, а затим и моћи против $\mathcal{P}(1.1)$, $\mathcal{P}(1.2)$, $\mathcal{P}(2)$, $\mathcal{P}(0.8)$, $\mathcal{P}(1.5)$ алтернатива, за $n = 10$, $n = 20$ и $n = 50$. Приликом оцењивања моћи користите $N = 10000$ Монте Карло понављања.
17. Крускал–Валисов тест (непараметарска ANOVA)
[Литература, 14.4.](#)
 Описати Крускал–Валисов тест и по чему се разликује од ANOVA теста и применити га на податке из примера у истом поглављу у књизи из литературе.
18. А/В тестирање ([текст пројекта](#))
 Извршити тестирање да ли је промена коју је направио сајт Udacity на својој платформи допринела бољем искуству корисника. Ово је завршни пројекат за [курс А/В тестирања](#). На интернету постоје решења неких од полазника курса, која могу бити од помоћи.
19. [Приходи од коцкарских игара у Лас Вегасу](#)
 На основу сирових података у извештаја (Raw data – табеле на странама 7-21) репродуковати резултате у извештају и нацртати графике из извештаја. За учитавање података из табела може помоћи програм [Tabula](#). Пожељно је дати и неке сопствене закључке.
20. Зигурат алгоритам
[Литература](#)
 Описати и имплементирати Зигурат алгоритам за генерисање нормалне и експоненцијалне расподеле. Објаснити предност овог алгоритма у односу на друге алгоритме за генерисање.

21. Нумеричко израчунавање оцена максималне веродостојности

Литература: Knight, Keith. Mathematical statistics. Boca Raton: Chapman & Hall/CRC Press, 2000, одељак 5.7.

Описати примену Њутновог метода на налазење оцене максималне веродостојности, као и Фишерову модификацију истог. Применити оба метода на примеру Кошијеве расподеле (пример из књиге) и упоредити им перформансе (брзину конвергенције, осетљивост на одабир почетне тачке...). Такође илустровати примену на некој расподели са више параметара.

22. Бајесов метод оцењивања

Литература, део 5.8.

Описати идеју Бајесовог метода оцењивања параметара и илустровати примером (може и из књиге) поступак добијања тачкасте оцене за параметар.

Напомене:

- Свака тема је предвиђена за 3 особе.
- Пријављивање за теме се врши попуњавањем форме на <http://uus-matf.online/prijava>
- Можете програмирати у језику/програму који желите (C, R, Excel, Matlab, Python, ...).
- Потребно је послати комплетну документацију као и семинарски откуцан у Latex-у на мејл uvodustatistiku@gmail.com. (**Послати пдф фајл и код који је коришћен за добијање резултата**).
- Рад на крају мора садржати списак коришћене литературе (која не мора бити ограничена на ону дату у теми). [Користан линк](#).
- **Рок за предају је 23. мај. Одбране ће бити у недељи после.** За рад и одбране се може освојити највише 20 поена (10 + 10). Поени за рад су јединствени за све чланове групе, а на одбрани се поени стичу индивидуално.
- **Питања се могу постављати до 20. маја.** Сва питања у вези са семинарским радом слати искључиво на мејл uvodustatistiku@gmail.com, која ћемо редовно проверавати.
- Ко жели да задржи поене са семинарског од претходне године (скалиране на 20 поена) потребно је да се јави мејлом на адресу uvodustatistiku@gmail.com.