

ОИМЕ И ПРЕЗИМЕ: Илија Вишинов

БРОЈ НА ИНДЕКС: 161078

1. (70 поени) Дизајнирајте научен експеримент што би дал одговор на едно прашање што ве интересира. Експериментот треба да е квантитативен, но не мора да биде реалистичен или лесно изводлив. Опишете ја методологијата на истражувањето во следните категории:

а) Која е хипотезата што ја тестирате? (Хипотезата треба да е потврдна реченица)

Дијабетесот, односно високо ниво на шеќер во крвта измерено преку glycated hemoglobin (HbA1C) или моментални мерења на шеќер се карактеризираат со помал HRV (heart rate variability) во работата на срцето.

б) Кои величини/квантитети ќе ги измерите како дел од експериментот?

На лицата вклучени во студијата би им се измериле HbA1C вредностите, континуирано мерење на едноканален (со поедноставен сензор), двоканален или повеќеканален ЕКГ од кој со дигитално процесирање на тие сигнали би се извлекле типовите на битови и соодветните ритам епизоди кои се јавуваат во сигналот односно работата на срцето. Од ова би се извлекле растојанијата помеѓу соодветните битови за дадени ритам епизоди и од тие RR, NN или други типови на интервали би се пресметале вредностите за HRV. Овие податоци може да се искористат така што за посебни случаи и сериозности на болеста би се разгледувале посебни податочни множества и зависности помеѓу работата на срцето и нивото на шеќер во крвта, или па би се испитувал некој генерален заклучок.

в) Кој статистички метод ќе го користите за да ја тестирате хипотезата?

Доколку проблемот се разгледува низ призмата на зависност од некој степен помеѓу вредностите на моментално ниво на гликоза во крв или HbA1C со HRV, можат да се искористат Пирсонова корелација која би ја пресметала јачината на линеарната зависност помеѓу двете вредности, или па Спирманова ранк корелација која би ја отсликала монотоната нелинеарна врска помеѓу двата параметри.

Доколку пак, имаме класи од лицата кои учествуваат во студијата, како дијабетичари наспроти недијабетичари, дијабетичари наспроти пред-дијабетичари, пред-дијабетичари наспроти недијабетичари, или па се класификувани според некои вредности на параметрите, може да се применат Point-biserial корелација и некој статистички тест како на пример t-test. Point-biserial корелацијата би ја пресметала сепарабилноста на

распределбата на класите (би се пресметало всушност Пирсонова корелација помеѓу вредноста на HRV и две енкодирани класи, но има и посебен метод за point-biserial корелација која ги дава идентичните резултати). Т-тестот би се искористил за да се види дали има значајна разлика меѓу распределбите на HRV за двете класи со тоа што би се пресметало дали двете распределби имаат значајна разлика на средните вредности.

Сите методи би дале р-вредност која би ја дала значајноста на тестот за корелација или разлика на средната вредност. Оптималната граница за прифаќање или отфрлање на хипотеза обично се зема да е 0.05, но во зависност од колку треба да сме сигурни во исходите на експериментите оваа вредност може и треба да се поместува по потреба.

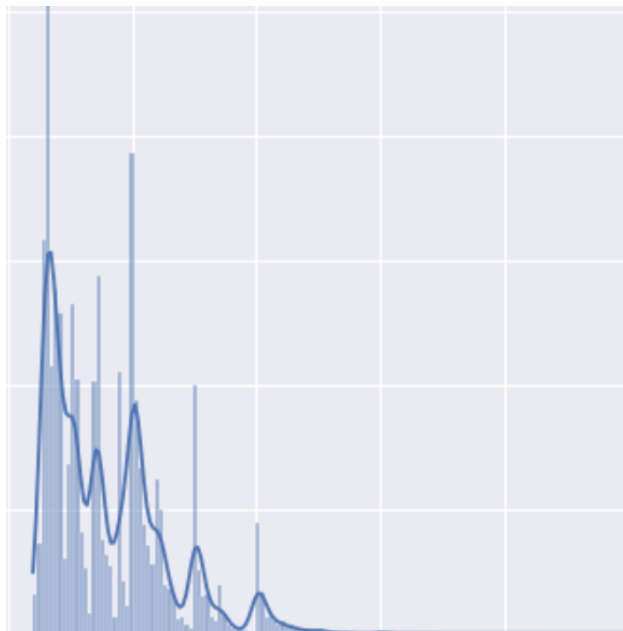
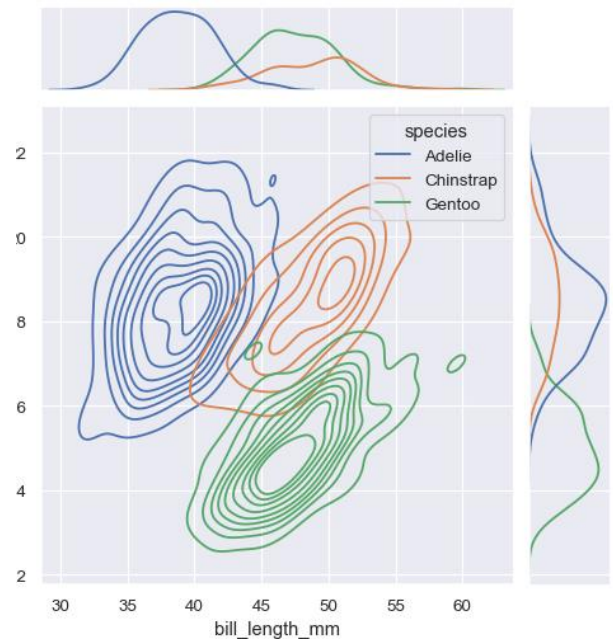
г) Како ќе ги визуелизирате мерењата од б)? Предложете скица во која ќе бидат претставени величините од мерењата. Скицата можете да ја пратите во прилог како дигитална фотографија.

За визуелизација може да се искористат хистограми за да се види распределбата на класите во една димензија, box-whiskers графици за да се сумаризира распределбата и се видат outliers и формата на распределбата, еднодимензионални и дводимензионални kernel-density estimation графици за да се апроксимираат тие распределби за да ја доловат поголемата слика кога би имале повеќе податоци, scatter графици за да се видат точните точки.

Освен на сировите податоци може да се применат и мапирања во други простори и да се визуелизираат таму податоците и да се пресметаат статистиките како на пример со Principal Components Analysis (PCA).

Резултатите од корелациите и нивните р-вредности може да се прикажат во heatmaps или некакви custom корелациони матрици.

Примери:



25 - 30	.14	.14	.18	.23	.15	.2	.22
20 - 25	.12	.12	.17	.13	.15	.16	.26
15 - 20	.12	.12	.17	.11	.13	.15	.12
10 - 15	.11	.11	.13	.11	.11	.16	.15
5 - 10	.11	.11	.13	.11	.11	.16	.15
1 - 5	.11	.11	.13	.11	.11	.16	.15
0 - 1	.11	.11	.13	.11	.11	.16	.15
-1 - 0	.11	.11	.13	.11	.11	.16	.15
-5 - -1	.15	.2	.15	.2	.21	.19	.23
-10 - -5	.15	.16	.15	.16	.17	.11	.26
-15 - -10	.15	.16	.15	.16	.17	.11	.26
-20 - -15	.15	.16	.15	.16	.17	.11	.26
-25 - -20	.15	.16	.15	.16	.17	.11	.26
-30 - -25	.15	.16	.15	.16	.17	.11	.26
-35 - -30	.15	.16	.15	.16	.17	.11	.26
-40 - -35	.15	.16	.15	.16	.17	.11	.26
-45 - -40	.15	.16	.15	.16	.17	.11	.26
-50 - -45	.15	.16	.15	.16	.17	.11	.26
-55 - -50	.15	.16	.15	.16	.17	.11	.26
-60 - -55	.15	.16	.15	.16	.17	.11	.26
SDMN	.13	.15	.13	.15	.14	.16	.22
SDMN-1	.13	.15	.13	.15	.14	.16	.22
ASDMN	.13	.15	.13	.15	.14	.16	.22
ASDMN-1	.13	.15	.13	.15	.14	.16	.22
ASDMN-2	.13	.15	.13	.15	.14	.16	.22
ASDMN-3	.13	.15	.13	.15	.14	.16	.22
SDAMN-2	.13	.15	.13	.15	.14	.16	.22
SDAMN-3	.13	.15	.13	.15	.14	.16	.22
MN50	.13	.15	.13	.15	.14	.16	.22
MN50-1	.13	.15	.13	.15	.14	.16	.22
PMN50	.13	.15	.13	.15	.14	.16	.22
PMN50-1	.13	.15	.13	.15	.14	.16	.22
RMSD	.13	.15	.13	.15	.14	.16	.22
RMSD-1	.13	.15	.13	.15	.14	.16	.22
SD1	.13	.15	.13	.15	.14	.16	.22
SD2	.13	.15	.13	.15	.14	.16	.22
SD1/SD2	.13	.15	.13	.15	.14	.16	.22

2. (30 поени) Како следните ресурси ќе ви помогнат да го направите горното истражување репродуцибилно? Одговорот не треба да биде општ туку да се однесува на конкретниот експеримент од претходната секција.

a) Dashboard

Dashboard би се применил така што со скрипти кои периодично би се повикувале, новите податоци би се додавале на претходните и би се апдејтирале plotly графициите кои се наоѓаат на dashboard-от. Со долги испитувања би можело да се надгледува како одредени медикаменти и протоколи вилијаат на HRV, глукоза или HbA1C. Исто би се добивале дополнителни статистики од кои би произлегувале нови идеи за подобро спроведување на студијата. Сите овие визуелизации ќе можат да се набљудувани од нови истражувачи, за да им дадат вовед во се што е изработено, како тие би донеле нови идеи и нови експерименти врз постоечките.

б) Контејнеризација

Со контејнеризација би можело многу полесно да се соработува со други истражувачи и тестирање на нивните експерименти и алгоритми без да мора да се троши непотребно многу време на справување со околината во која тој алгоритам е подготвен.

в) Version control

Со вклученоста на повеќе учесници во студијата и развивање на инфраструктурата на проектот кој го содржи pipeline-от на прибирање податоци, препроцесирање, визуелизација, тестови, методи, машинско учење итн., version control би помогнал да сето тоа остане организирано, со recovery points во случај на проблеми, и колаборација на заеднички библиотеки и модули кои би биле реискористливи. Новите истражувачи кои би го репродуцирале истраживањето и би го надградувале можат да ја земат веќе постоечката архитектура, да проверат дали во нејзе постојат алгоритми кои имаат грешки и да се вклучат во нејзино надградување така што ќе допринесат на истражувачкиот процес.

3. (20 поени) За ова прашање ќе треба да најдете оригинален истражувачки труд на сајтот:

[Scholar.google.com](https://scholar.google.com)

Трудот треба да има секција за методи (најчесто поднаслов Methods или Methodology) и да има јасна хипотеза. Бидејќи голем дел од трудовите се достапни само со плаќање (paywalled), на час ви кажавме како да пристапите до нив бесплатно. Прашањето 4 е поврзано со истиот труд, така што посветете доволно време во изборот на трудот за да можете полесно да ги одговорите сите прашања и задачи.

На час не ви кажавме како да цитирате труд, така што ова ќе треба сами да го дознаете. Цитирајте го избраниот труд користејќи го IEEE стилот на цитирање!

ОДГОВОР:

Mancini, Matteo, Agah Karakuzu, Thomas Nichols,
Julien Cohen-Adad, Mara Cercignani, and Nikola Stikov.

"The quest for measuring myelin with MRI-An interactive
meta-analysis of quantitative comparisons with histology." *BioRxiv* (2020).

Методологијата е објаснета во тетратката.

Освен наведениот, има повеќе стилови на IEEE цитирање:

- | | |
|-----------|--|
| MLA | Mancini, Matteo, et al. "The quest for measuring myelin with MRI-An interactive meta-analysis of quantitative comparisons with histology." <i>BioRxiv</i> (2020). |
| APA | Mancini, M., Karakuzu, A., Nichols, T., Cohen-Adad, J., Cercignani, M., & Stikov, N. (2020). The quest for measuring myelin with MRI-An interactive meta-analysis of quantitative comparisons with histology. <i>BioRxiv</i> . |
| Chicago | Mancini, Matteo, Agah Karakuzu, Thomas Nichols, Julien Cohen-Adad, Mara Cercignani, and Nikola Stikov. "The quest for measuring myelin with MRI-An interactive meta-analysis of quantitative comparisons with histology." <i>BioRxiv</i> (2020). |
| Harvard | Mancini, M., Karakuzu, A., Nichols, T., Cohen-Adad, J., Cercignani, M. and Stikov, N., 2020. The quest for measuring myelin with MRI-An interactive meta-analysis of quantitative comparisons with histology. <i>BioRxiv</i> . |
| Vancouver | Mancini M, Karakuzu A, Nichols T, Cohen-Adad J, Cercignani M, Stikov N. The quest for measuring myelin with MRI-An interactive meta-analysis of quantitative comparisons with histology. <i>BioRxiv</i> . 2020 Jan 1. |

Трудот може и да се цитира во LaTeX ако со BibTeX синтакса се додаде во .bib библиотека на следниот начин:

```
@article{mancini2020quest,  
  title={The quest for measuring myelin with MRI-An interactive meta-analysis of quantitative  
  comparisons with histology},  
  author={Mancini, Matteo and Karakuzu, Agah and Nichols, Thomas and Cohen-Adad, Julien  
  and Cercignani, Mara and Stikov, Nikola},  
  journal={BioRxiv},  
  year={2020},  
  publisher={Cold Spring Harbor Laboratory}  
}
```

4. (130 поени) Направете Jupyter тетратката поврзана со трудот од претходното прашање и прикачете ја на GitHub (доколку немате профил креирајте го, ќе ви треба). Линкот од вашиот Github hero мора да биде испратен до 23.59 часот на 11 февруари (сите промени по овој краен рок нема да бидат прифатени). Исто така нема да прифаќаме тетратки хостирани на било кое друго место освен на Github.

а) Тетратката треба да започне со краток опис на трудот (напишан во Markdown). Краткиот опис треба во стотина зборови да објасни зошто е овој труд значаен.

б) Остатокот од тетратката го оставаме на вас. Не заборавате дека испитите ќе бидат рангирани, така што тие кои ќе имаат најквалитетна тетратка ќе добијат најмногу поени. За да биде кандидат за максимална оценка, тетратката треба да ги содржи повеќето од овие карактеристики:

- Формули од избраниот труд напишани во LaTeX
- Ќелии со код од избраниот труд кои може да се егзекутираат (полесно е ова да се направи доколку податоците и кодот од трудот се јавно достапни)
- Интерактивна визуелизација (Plotly, ipywidgets или други алатки)
- Вметнато лого на журналот во кој е објавен трудот
- Ембедиран мултимедијален запис поврзан со трудот (YouTube видео, podcast, ...)

Целта на ова прашање е да бидете креативни. Понудете ни тетратка која го надополнува оригиналниот PDF и го прави истражувањето да биде покорисно. Доколку трудот ги споделува податоците, тогаш можете да направите и сосема нова визуелизација. Изненадете нè!

P.S. Вашите одговори на испитот треба да бидат прикачени на GitHub (во PDF или друг електронски формат) заедно со Jupyter тетратката.

