

R бағдарламасында деректерді талдау

Жаз 2025

Санжар Қалқанбай

MIND MNU

s.kalkanbay@gmail.com

Last Update: 2025-06-12

Жалпы ақпарат

Сынып сағаттары

Бұл сынып сағаттары дүйсенбі/бейсенбі күндері 2 сағат жүреді. Жалпы, курс келесі форматта жүргізіледі (өзгеруі мүмкін):

- Курстың теориялық бөлігі сабақтың бірінші жарты сағатында лекция форматында ұсынылады.
- Практикалық бөлік (яғни, R) қалған уақытта өткізіледі. Студенттер компьютерлерін алып келуі талап етіледі.

Офистік Сағаттар

Офистік сағаттар - бұл ұсынылған материалды немесе басқа материалға қатысты қызығушылықтарыңызды талқылау үшін мұғаліммен кездесуге арналған уақыт. Курсқа қатысты талқылаулар қосымша көмек сұрауды, сынып материалдарын түсіндіруді, немесе сізді қызықтыратын аспектілерді қарастыруды қамтиды.

Офистік сағаттар алдын ала жазылу бойынша, сейсенбі/бейсенбі күндері 16:00-17:00 сағаттарында өтеді. Төмендегі URL мекенжайында жазылу керек:

<https://calendly.com/s-kalkanbay/ofistik-sagattar>

Әрбір кездесу уақыты 15 минуттан тұрады. Ұзағырақ кездесуді қаласаңыз, екі орынды брондаңыз. Броньды кездесу уақытынан *кемінде 4 сағат бұрын жасаңыз*. Егер берілген слоттар ыңғайсыз болса, маған қол жетімді уақыттарыңыздың ұсынысымен жазыңыз.

Пререквизиттер

Бұл курс студенттердің статистикалық бағдарламадар және/немесе бағдарламалау туралы алдын ала білімі жоқ деп санайды. Студенттер математикалық талдау және статистикалық терминдер туралы негізгі білімге ие, сонымен қатар ағылшын тілін төмен немесе орта деңгейде біледі деп санайды.

Курс мақсаттары

R - data analysis үшін, әсіресе әлеуметтік ғылымдарда кеңінен қолданылатын бағдарламалау тілі. Ол деректермен жұмыс істеу, статистикалық талдау жасау және визуализация жасау үшін икемді ортаны ұсынады. Бұл курс *tidyverse* және R көмегімен деректер дайындаудың практикалық дағдыларына,

визуализацияға және қарапайым статистикалық модельдеуге баса назар аудара отырып, R негіздерін үйретеді.

Осы курсты аяқтағаннан кейін студенттер келесі дағдыларды меңгереді деп күтілуде:

- R-дың негізгі синтаксисы мен құрылымын түсіну.
- R көмегімен деректерді импорттау, тазарту және манипуляциялау.
- Тартымды және ақпараттық визуализациялауды құру.
- Негізгі статистикалық талдауды орындау.
- Статистикалық модельдерді құру және талдау.
- Quarto көмегімен құжаттар мен презентациялар жасау.

Курс практикалық жұмысқа бағытталған және әлеуметтік ғылымдар бойынша тәуелсіз зерттеулер мен сандық талдау жүргізу үшін негізгі дағдыларды береді.

Статистикалық бағдарламалық жасақтама

Бұл курста біз R “open-source” статистикалық бағдарламалық жасақтамасын қолданамыз (<http://www.r-project.org>), сондай-ақ RStudio пайдаланушы интерфейсін (<http://www.rstudio.com/>) қолданамыз. Студенттерде R және R Studio орнатуға және іске қосуға қабілетті компьютерлер болады деп күтілуде.

Алғашқы сабақ басталар алдында орнату файлдарын жүктеу қажет!

- RStudio: <https://posit.co/download/rstudio-desktop/>
- R:
 - Windows үшін: <https://cran.r-project.org/bin/windows/base/R-4.5.1-win.exe>
 - Apple MacOS үшін: <https://cran.r-project.org/bin/macosx/big-sur-arm64/base/R-4.5.1-arm64.pkg>

Курстың Қысқаша Мазмұны

Сабақ

Күні

Тақырыптар

1 Сабақ

TBD

1. Курсқа кіріспе және курс құрылымы
2. R және RStudio орнату, жұмыс ортасын баптау
3. R калькулятор ретінде: негізгі синтаксис, айнымалылар, меншіктеу
4. Деректер түрлері және векторлар
5. Көмек алу жолдары

2 Сабақ

TBD

1. Деректер құрылымы: векторлар, матрицалар, тізімдер, деректер жиындары (data frame)
2. Іріктеу және индекстеу
3. ‘tibble’ және қарапайым деректерді өңдеуге кіріспе
4. CSV/Excel файлдарын оқу

3 Сабақ

TBD

1. ‘dplyr’ көмегімен деректерді тазалау
2. Филтрлау, таңдау, өзгерту, жиынтықтау
3. Топтау және агрегация
4. Таза және түсінікті скрипт жазу

4 Сабақ

TBD

1. ‘ggplot2’ кітапханасына кіріспе: графика грамматикасы
2. Нүктелік, бағаналық және гистограмма графиктерін жасау
3. Эстетикалық баптаулар және тақырыптар
4. Графиктерді экспорттау

5 Сабақ

TBD

1. Категориялық айнымалылармен (факторлармен) жұмыс істеу
2. Уақыт пен күндермен жұмыс істеу
3. ‘tidyr’ арқылы қайта құрылымдау (pivot, join)
4. Жоба ұйымдастыру бойынша кеңестер

6 Сабақ

TBD

1. R тілінде сипаттамалық статистика жасау
2. Сандық және категориялық деректерді жиынтықтау
3. Орташа мән мен өзгергіштікті түсіндіру
4. Кестелер мен диаграммаларда нәтижелерді көрсету

7 Сабақ

TBD

1. ‘lm()’ көмегімен сызықтық модельдеуге кіріспе
2. Қарапайым регрессия жүргізу
3. Нәтижелерді интерпретациялау
4. Регрессия нәтижелерін визуалдау

8 Сабақ

TBD

1. Quarto-ға кіріспе
2. Динамикалық құжаттар жасау
3. Кестелер, графиктер және ішкі статистикалар
4. Қорытынды: R-ды зерттеуде қолдану