

# Új kurzus regisztrálása a „Molekuláris orvostudományok” doktori iskolában

---

A kurzus neve: „Adatelemzés és programozás orvosbiológusoknak II.”

A modern orvos-biológiai kutatások során a kutatók egyre gyakrabban kerülhetnek kapcsolatba olyan mennyiségű adathalmazzal melyek „kézi” feldolgozása gyakorlatilag lehetetlen. Ilyen adatok pl. az mRNS microarray illetve szekvenálási adatok, szekvencia adatbázisok, beteg adatbázisok, de ilyen lehet akár képek, videók analízise (pl. kísérleti állat mozgásának követése). Ezen adatok feldolgozásához egyre fontosabb, hogy a kísérleti területen dolgozó kutatók is rendelkezzenek alapvető programozási és „adat analízis” kompetenciákkal. Az adatok feldolgozására számos esetben léteznek kereskedelmi vagy ingyenes szoftverek, azonban ezek az adatok feldolgozásában limitált lehetőséget biztosítanak az új vagy egyéni szempontok alapján történő feldolgozásra. Programozási alapismeretekkel azonban bármilyen szempontok alapján történő feldolgozás lehetővé válik, akár nagy adathalmazok esetén is. Ez egyrészt növeli a kutatók hatékonyságát az adatfeldolgozásban, másrészt lehetőséget teremt kreatív és újszerű megoldásokra.

A kurzus célja olyan alap programozási ismeretek átadása, ami alapján a hallgatók önnállóan fejleszthetik tovább programozási ismereteiket online elérhető anyagok alapján a saját érdeklődési területüknek megfelelően. A kurzus nem igényel előzetes programozási ismereteket.

A kurzus a python és az R programozási nyelvekre koncentrál. Ezek a nyelvek a bioinformatikában és azon túl is széles körben használtak, ingyenesek, kiválóan alkalmasak kezdő programozók részére a könnyen érthető szerkezet és egyszerűség miatt, ugyanakkor komolyabb programok megírására is alkalmasak.

A kurzus problémaorientált, a nyelvek elemeit egyszerű példákon keresztül tanulják a diákok. Később valódi biológiai adatsorokon dolgozva tanulják meg az összetettebb feladatok megoldását. A kurzus során a tananyaghoz kapcsolódó heti feladatokat kapnak, amiket következő alkalommal megbeszélünk. A hallgatókat bátorítjuk a feladatok csoportos megoldására, illetve van lehetőség kutatási feladataikhoz kapcsolódó adatsorokon is dolgozni. Az értékelés a részfeladatok és a projekten végzett munka alapján történik.

A kurzus 14 előadásból áll, heti egy-egy alkalommal 90 perces időtartamban. A kurzus elvégzéséhez szükség van számítógépre, erről a hallgatóknak maguknak kell gondoskodniuk. A kurzus folytatása az őszi félév „Adatelemzés és programozás orvosbiológusoknak I” kurzusának. Az előadásokat Dr Cserző Miklós, Dr Sebestyén Endre, Dr Szalai Bence és Dr Turu Gábor tartja a következő címekkel:

1. Python ismétlés I.
2. Python ismétlés II.
3. Python adatfeldolgozás I. (Pandas)
4. Python adatfeldolgozás II. (Pandas)
5. Python adatfeldolgozás III. (Pandas)
6. Python adatvizualizáció I. (matplotlib)
7. Python adatvizualizáció II. (seaborn)
8. R és parancssor ismétlés
9. RNAseq adatok kiértékelése I. (R, parancssor, sleuth, kalisto)
10. RNAseq adatok kiértékelése II. (R, parancssor, sleuth, kalisto)
11. RNAseq adatok kiértékelése III. (R)
12. Exploratív adatanalízis (Python)
13. Modellelés, 'machine learning' I. (Python)
14. Modellelés, 'machine learning' II. (Python)