Kurikulum kolegija *Kemometrika*

SVRHA KOLEGIJA: Kemometrika je relativno mlada disciplina kemije koja u posljednje vrijeme doživljava znatan uspon zahvaljujući sve bržem razvoju računala. Stoga je za modernu nastavu kemije svakako poželjno da studenti steknu osnovna znanja iz te discipline, odnosno da ovladaju matematičkim i statističkim metodama koje se najčešće rabe pri obradi kemijskih podataka te se upoznaju s odgovarajućim računalnim programskim paketima.

SADRŽAJ KOLEGIJA: Repetitorij elementarne statistike: osnove teorije vjerojatnosti, progresija pogreške, najčešće univarijatne i multivarijatne raspodjele, lokacijski i disperzijski statistički testovi. Analiza varijancije: jednosmjerna, višesmjerna (fiksni učinak, sa i bez ponavljanja). Regresijska analiza (metoda najmanjih kvadrata): linearni modeli (testovi značajnosti regresijskih parametara, vagani modeli, test skladnosti); polinomna regresija. Nelinearni modeli (osnovna načela i metode). Optimizacija nacrta pokusa: metoda politopa (simpleksa). Izglađivanje i filtriranje podataka. Rojna (*cluster*) analiza. Analiza glavnih komponenata (PCA), PCR, PLS, faktorska analiza. Neparametarska statistika: lokacijski i korelacijski testovi, analiza varijancije, usporedba raspodjela.

OPĆENITI ISHODI UČENJA:
Studenti bi trebali u potpunosti vladati lokacijskim i disperzijskim statističkim testovima te jednosmjernom i višesmjernom analizom varijancije (u manjoj mjeri i odgovarajućim neparametarskim verzijama). Linearna regresijska analiza zasigurno je među kemičarima najčešće korištena metoda obrade podataka. Stoga se očekuje da studenti solidno razumiju njezinu teorijsku osnovicu, kao i primjenu u slučaju jednostavne, multilinearne (matrični oblik) i polinomne regresije. Također bi trebali znati statističke kriterije i testove koji omogućuju procjenu adekvatnosti pretpostavljenog modela. Sve navedeno vrijedi i za nelinearnu regresijsku analizu, pri čemu se podrazumijeva poznavanje iterativnih metoda za pronalaženje regresijskih parametara (s tim je povezana i optimizacijska metoda simpleksa). Od multivarijatne analize podataka, očekuje se prvenstveno razumijevanje postavke modela te osnova analize glavnih komponenata i tehnika koje se na njoj baziraju. Nakon položenog ispita studenti bi u vlastitim istraživanjima trebali moći samostalno izabrati i primijeniti najpogodniju metodu analize eksperimentalnih podataka (kao i odgovarajuće računalne programe) te kritički interpretirati dobivene rezultate, što je važno ne samo u kemiji, nego i u svim ostalim područjima znanosti. Budući da su dužni napisati i prezentirati seminarski rad na zadanu ili odabranu temu, to im omogućava stjecanje vještine pisanja znanstvenog/stručnog teksta, kao i njegove prezentacije.

SPECIFIČNI ISHODI UČENJA/RAZRADA SADRŽAJA/RAZINA:
Opažajne ljestvice: nominalna, redna, intervalna, omjerna.
Teorija vjerojatnosti: definicija vjerojatnosti; vjerojatnost neovisnih događaja; slučajna varijabla (= varijata = statistik); vjerojatnosna funkcija diskretne varijate; gustoća vjerojatnosti kontinualne varijate; očekivana vrijednost slučajne variable i njezine proizvoljne analitičke funkcije; kumulativna raspodjela.
Iskustveni opis raspodjele (razdiobe) varijate: histogram, frekvencijski poligon (razredni i kumulativni); mjere središnje težnje (aritmetički, geometrijski i harmonijski prosjek, median, mod); mjere rasapa (variancija i standardno odstupanje); propagacija pogreške.
Procjena (estimacija) parametara raspodjele: aritmetički prosjek i standardno odstupanje kao estimatori μ, odnosno σ (na primjeru univarijatne Gaussove raspodjele); standardna pogreška aritmetičkog prosjeka; središnji granični poučak.
Načela statističkih testova: značajnost razlike aritmetičkih prosjeka dvaju uzoraka; parametarski lokacijski testovi: normalni i Studentov *t*-test; disperzijski testovi: raspodjela Fisherova statistika *F*; Fisherov *F*-test jednakosti dviju procjena iste varijancije; definicija i rapodjela varijate χ^2; Bartlettova kušnja homoskedastičnosti za >2 procjene varijancije.
Analiza variancije ('ANOVA'): test lokacijskih razlika >2 uzorka istovrsne varijate; hierarhijski model; faktorski modeli (fiksni učinak, sa i bez ponavljanja); test najmanje značajne razlike (*least significant difference*).
Linearna regresijska analiza: jednostavan model (pravac): normalne jednadžbe; testovi značajnosti odsječka i nagiba; signifikantnost regresije (analiza varijancije); pravac kroz zadanu točku (npr. ishodište); standardne pogreške i granice pozdanosti procjena parametara; multilinearna i polinomna regresija: normalne jednadžbe, matrični prikaz i rješenje; matrica varijancije-kovarijancije i granice pouzdanosti regresijskih parametara; modeli s uravnoteženom varijancijom (statističko vaganje); usporedba modela (parcijalni *F*-test); provjera valjanosti modela: test skladnosti (*goodness-of-fit test*), *lack-of-fit test*; koeficijent determinacije, korelacija, linearni korelacijski koeficijent; multivarijatni modeli.
Nelinearna regresijska analiza: metoda najmanjih kvadrata i normalne jednadžbe u nelinearnoj regresiji; Gauss-Newtonova metoda linearizacije; pristup nelinearnoj regresiji pomoću metode simplexa te optimizacija pokusa tom metodom.
Mulrivarijatna analiza podataka: rojna analiza**:** hijerarhijsko (dendrogram) i nehijerarhijsko grupiranje, mjere sličnosti, algoritmi grupiranja; analiza glavnih komponenata: manifestne i latentne varijable, redukcija dimenzionalnosti,

dijagonalizacija disperzijske matrice, dekompozicija po singularnim vrijednostima (SVD), *score* i *loading* prikazi, kriteriji i testovi za procjenu efektivnog ranga (pseudoranga) podatkovne matrice (spektrofotometrijski primjeri); faktorska analiza, iterativni algoritam avolucijske faktorske analize; temelji metoda PCR i PLS.
Izglađivanje i filtriranje podataka:* Fourierova i Hadamardova transformacija; konvolucija; metoda putujućeg prosjeka, polinomno izglađivanje (Savitzky–Golay), eksponencijalno izglađivanje; korekcija s obzirom na baznu liniju.
Neparametarska statistika:* lokacijski testovi (testovi predznaka, Wilcoxon, Mann-Whitney); korelacija (Spearmanov R, Kendallov τ); jednosmjerna (Kruskal-Wallis) i dvosmjerna analiza varijancije (Friedman); usporedba raspodjela (Kolmogorov-Smirnov).

* dijelom izborno gradivo koje može biti temom studentskih seminarkih radova

LITERATURA ZA KOLEGIJ:

1. D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. de Jong, P. J. Lewi, J. Smeyers-Verbeke: *Handbook of Chemometrics and Qualimetrics, Part A*, Elsevier, Amsterdam 1997.
2. B. G. M. Vandeginste, D. L. Massart, L. M. C. Buydens, S. de Jong, P. J. Lewi, J. Smeyers-Verbeke: *Handbook of Chemometrics and Qualimetrics, Part B*, Elsevier, Amsterdam 1998.
3. N. R. Draper, H. Smith: *Applied Regression Analysis*, Wiley, New York 1976 (1. izd.), 1981 (2. izd.).
4. B. E. Cooper, *Statistics for Experimentalists*, Pergamon Press, Oxford 1969.

DODATNA LITERATURA ZA KOLEGIJ:

1. W. R. Dillon, M. Goldstein: *Multivariate Analysis: Methods and Aplplications*, Wiley, New York 1984.
2. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, L. Kaufman: *Chemometrics: A Textbook*, Elsevier, Amsterdam 1988.
3. S. N. Deming, S. L. Morgan: *Experimental Design: A Chemometric Approach*, Elsevier, Amsterdam 1987.
4. I. Pavlić, *Statistička teorija i primjena*, 4. izd., Tehnička knjiga, Zagreb 1988.
5. W. H. Press, B. P. Flannery S. A. Teukolsky, W. T. Vetterling: *Numerical Recipes*, Cambridge University Press, Cambridge 1986 (i kasnija izdanja).