University of Colorado
Department of Civil, Environmental and Architectural Engineering
CVEN 5454 Quantitative Methods

Homework #5 Due by December 9, 2025

Topics: Regression Chapters 11, 12, Helsel and Hirsch Chapters 10,11,12 and 15. Bivariate and conditional PDFs – Kernel density estimators and Copulas. Using R-commands for these will make it easier!

---

1. Bivariate PDFs – Copula/Kernel Density Estimators

For the Lees Ferry monthly flow

(i) Plot the bivariate Normal PDF (X = Mar flow and Y = May flow)

(ii) Plot the conditional PDF for Y, for X = 0.6MaF and X = 0.8MaF

(iii) Repeat (i) and (ii) using Kernel density estimators

(iv) ***Optional:*** Use Copula to obtain the bivariate PDF/CDF.

2. *Correlation/Mutual Information (MI)*

Select May flow and one of the monthly flows (Jan, Feb, Mar or Apr)

For this pair compute

(a) Correlation coefficient

(b) Kendall's tau

(c) Spearman's rank correlation.

Perform significance test on all of them, report the p-value and summarize.

(d) Compute the Mutual Information (MI) using Kernel Density Estimators and its significance. Compare with the correlation coefficients above.

3. *Simple Regression / Robust regression/trend line*

Consider the data in Problem 12.4 of Helsel and Hirsch. To this data perform the following.

(a) Fit a robust or nonparametric line – i.e., Kendall-Theil line between Calcium (dependent variable) and %Carbon-14 (independent variable).

(b) Also fit a linear regression line and show them with the scatterplot of the data.

(b) Perform model diagnostics for these two models along with a significance test on their slopes and compare them.

4. *Multivariate Regression*

Problem 11.1 of Helsel and Hirsch. Obtain the *best model* for LOGTN. To this end, the following steps need to be performed. You can get the excel file with the data from (https://www.sciencebase.gov/catalog/item/5bf30260e4b045bfcae0c205)

(a) Scatterplot LOGTN with all the potential predictor variables and apply a default smoother in R to smooth the scatterplot and comment on the relationships you see (linear, nonlinear etc.).

(b) Using PRESS or AIC or BIC to obtain the best model and perform model diagnostics including ANOVA. (As ***optional exploration*** try all three to see if the best model differs).

(c) Plot the historical and modeled values with 95% confidence interval along with 1:1 line for visual inspection of the model performance.

(d) Estimate the skill of the model by repeatedly dropping 10% of points - drop 10% of points, fit the best model to the rest and predict the dropped points; compute skill measures – R and RMSE; repeat. Boxplot the skill measures.

(e) Comment on the model shortcomings