MLDS 401 (Fall 2023): Lab 1 – 09/26/2023

2.3 (Weighted least squares) Suppose that the observations y_i have different precisions and so we would like to weight them using different weights $w_i > 0$. For example, each y_i may be a sample mean of n_i i.i.d. observations so that their variances are inversely proportional to the n_i. Hence we use the n_i as the weights. Show that the **weighted least squares (WLS)** estimator of the slope β for regression through the origin, obtained by minimizing the LS criterion $Q = \sum_{i=1}^{n} w_i(y_i - \beta x_i)^2$, equals

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} w_i x_i y_i}{\sum_{i=1}^{n} w_i x_i^2}.$$

2.10 (Price elasticities of steaks) Data file `steakprices.csv` gives time series data on the prices and quantities sold of three types of beef steaks, chuck, porterhouse and rib eye,
(*Source*: `http://www.aabri.com/manuscripts/08118.pdf`).

a) Estimate the price elasticities of all three steaks. Given that chuck is the least expensive cut and porter house is the most expensive cut of beef among these three cuts, are their price elasticities in the expected order?

b) Estimate how much the demand will change if the price is increased by 10% for each cut.

2.11 (Smoking versus cancer) Data file `smoking-cancer.csv` contains data from 43 states and Washington, D.C. on the average number of cigarettes smoked (hundreds/capita) and number of deaths per 100,000 population due to four types of cancer (bladder, lung, kidney and leukemia).

a) Make scatter plots of the number of deaths due to each type of cancer versus cigarettes smoked to see what types of relationships (linear, nonlinear) exist and if there are any outliers.

b) Perform tests on the correlations to see which type of cancer deaths are most significantly correlated with cigarette smoking.

Question 3

The sample data below relates y=ammonium concentration (mg/L) in soil to x=transpiration (ml/) of plant stock.

x	5.8	8.8	11.0	13.6	18.5	21.0	23.7	26.0	28.3	31.9	36.5	38.2	40.4
y	7.8	8.2	6.9	5.3	4.7	4.9	4.3	2.7	2.8	1.8	1.9	1.1	0.4

Given the calculated quantities:

$n = 13$, $\sum x_i = 303.7$, $\sum y_i = 52.8$ $S_{xx} = 1585.230769$, $S_{xy} = -341.959231$, $S_{yy} = 77.270769$

(a) What is the equation of the estimated regression line?

(b) What is a point value prediction of ammonium concentration for x=25?

(c) What is a point value prediction of ammonium concentration for x=45? Does it make sense to calculate this point estimate?

(d) Calculate and interpret the coefficient of determination

(e) Calculate a 99% confidence interval for β_1

(f) Calculate a 95% confidence interval for the mean ammonium concentration (mg/L) for $x = 25$

(g) Calculate a 95% prediction interval for a single observation of y for $x = 25$