MATH 484 & 564: Homework 1
Due on Sept 08, 2022, 11:59 pm

Please follow the guidelines in "Course Syllabus Slides" to prepare and submit your homework assignment. Late submission will not be accepted except for reasonable excuse.

Total points=60, 10 pts for each problem.

1. Problem 1 data link 1.27

*1.27. **Muscle mass.** A person's muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selected 15 women from each 10-year age group, beginning with age 40 and ending with age 79. The results follow; X is age, and Y is a measure of muscle mass. Assume that first-order regression model (1.1) is appropriate.

i:	1	2	3	...	58	59	60
X_i:	43	41	47	...	76	72	76
Y_i:	106	106	97	...	56	70	74

a. Obtain the estimated regression function. Plot the estimated regression function and the data. Does a linear regression function appear to give a good fit here? Does your plot support the anticipation that muscle mass decreases with age?

b. Obtain the following: (1) a point estimate of the difference in the mean muscle mass for women differing in age by one year, (2) a point estimate of the mean muscle mass for women aged $X = 60$ years, (3) the value of the residual for the eighth case, (4) a point estimate of σ^2.

2. Problem 2 data link 1.28

1.28. **Crime rate.** A criminologist studying the relationship between level of education and crime rate in medium-sized U.S. counties collected the following data for a random sample of 84 counties; X is the percentage of individuals in the county having at least a high-school diploma, and Y is the crime rate (crimes reported per 100,000 residents) last year. Assume that first-order regression model (1.1) is appropriate.

i:	1	2	3	...	82	83	84
X_i:	74	82	81	...	88	83	76
Y_i:	8,487	8,179	8,362	...	8,040	6,981	7,582

a. Obtain the estimated regression function. Plot the estimated regression function and the data. Does the linear regression function appear to give a good fit here? Discuss.

b. Obtain point estimates of the following: (1) the difference in the mean crime rate for two counties whose high-school graduation rates differ by one percentage point, (2) the mean crime rate last year in counties with high school graduation percentage $X = 80$, (3) ε_{10}, (4) σ^2.

3. Problem 3

1.39. Two observations on Y were obtained at each of three X levels, namely, at $X = 5$, $X = 10$, and $X = 15$.

a. Show that the least squares regression line fitted to the *three* points $(5, \bar{Y}_1)$, $(10, \bar{Y}_2)$, and $(15, \bar{Y}_3)$, where $\bar{Y}_1$, $\bar{Y}_2$, and $\bar{Y}_3$ denote the means of the Y observations at the three X levels, is identical to the least squares regression line fitted to the original six cases.

Simple Linear Regression

b. In this study, could the error term variance σ^2 be estimated without fitting a regression line? Explain.

4. Problem 4 data link 1.42

1.42. **Typographical errors.** Shown below are the number of galleys for a manuscript (X) and the dollar cost of correcting typographical errors (Y) in a random sample of recent orders handled by a firm specializing in technical manuscripts. Assume that the regression model $Y_i = \beta_1 X_i + \varepsilon_i$ is appropriate, with normally distributed independent error terms whose variance is $\sigma^2 = 16$.

i:	1	2	3	4	5	6
X_i:	7	12	4	14	25	30
Y_i:	128	213	75	250	446	540

a. State the likelihood function for the six Y observations, for $\sigma^2 = 16$.

b. Evaluate the likelihood function for $\beta_1 = 17$, 18, and 19. For which of these β_1 values is the likelihood function largest?

c. The maximum likelihood estimator is $b_1 = \sum X_i Y_i / \sum X_i^2$. Find the maximum likelihood estimate. Are your results in part (b) consistent with this estimate?

d. Using a computer graphics or statistics package, evaluate the likelihood function for values of β_1 between $\beta_1 = 17$ and $\beta_1 = 19$ and plot the function. Does the point at which the likelihood function is maximized correspond to the maximum likelihood estimate found in part (c)?

5. Problem 5: Prove that $\sum_{i=1}^{n} e_i x_i = 0$ and $\sum_{i=1}^{n} e_i = 0$.

6. Problem 6: Prove that $\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2\right)$.