1: Regression

(a) Reason:

A student working on a summer internship in the economic research department of a large corporation studied the relation between sales of a product (Y, in milion dollars) and population (X, in million persons) in the firm's 50 marketing districts. The normal error regression model was employed. The student first wished to test whether or not a linear association between Y and X existed. The student accessed a simple linear regression program and obtained the following information on the regression coefficients:

Parameter	Estimated value	95% confidence limits	
Intercept	7.43119	-1.18518	16.0476
Slope	.755048	.452886	1.05721

- The student concluded from these results that there is a linear association between Y and X. Is the conclusion warranted? What is the implied level of significance?
- Someone questioned the negative lower confidence limit for the intercept, pointing out that dollar sales cannot be negative even if the population in a district is zero. Discuss.

(b) Reason:

In a test of the alternatives $H_0 : \beta_1 \leq 0$ versus $H_a : \beta_1 > 0$, an analyst concluded H_0. Does this conclusion imply that there is no linear association between X and Y? Explain.

(c) Reason:

A member of a student team playing an interactive marketing game received the following computer output when studying the relation between advertising expenditures (X) and sales (Y) for one of the team's products:

- Estimated regression equation: $\hat{Y} = 350.7 - .18X$
- Two-sided P–value for estimated slope: .91

The student stated: "The message I get here is that the more we spend on advertising this product, the fewer units we sell!" Comment.

(d) Solve:

In the regression $Y_i = \beta_0 + \beta_1 X_i + u_i$ suppose we multiply each X value by a constant, say, 2. Will it change the residuals and fitted values of Y? Explain. What if we add a constant value, say, 2, to each X value?

(e) R commands: (with interpretation)

Grade point average. The director of admissions of a small college selected 120 students at random from the new freshman class in a study to determine whether a student's grade point average (OPA) at the end of the freshman year (Y) can be predicted from the ACT test score (X). The results of the study follow. Assume that first–order regression model is appropriate.

Data are available in the homework page (the link is near this homework).

- Obtain the least squares estimates of β_0 and β_1 and state the estimated regression function.
- Plot the estimated regression function and the data. "Does the estimated regression function appear to fit the data well?"
- Obtain a point estimate of the mean freshman GPA for students with ACT test score $X = 30$.
- What is the point estimate of the change in the mean response when the entrance test increases by one point?
- Obtain a 99 percent confidence interval for β_1. Interpret your confidence interval. Does it include zero? Why might the director of admissions be interested in whether the confidence interval includes zero?
- Test, using the test statistic t^*, whether or not a linear association exists between student's ACT score (X) and GPA at the end of the freshman year (Y). Use a level of significance of .01. State the alternatives, decision rule, and conclusion.
- What is the P–value of your test in part (b)? How does it support the conclusion reached in part (b)?

NOTE: Solve the exercise in the following two ways:

- exploiting the proper regression R functions
- using R as a calculator (without exploiting internal functions for regression, but using only the arithmetic operators and the mean function).