1. Suppose we observe pairs of data values \( x_i \) and \( y_i \) for \( i = 1, \ldots, n \) individuals (for example, \( x_i \) could be the length of a student’s shoe and \( y_i \) the height of the student). It is common to assume a linear model for the relationship between the mean of a random variable \( Y_i \) and some observed covariate \( x_i \):

\[
\mu_i = E(Y_i|X_i = x_i) = \beta_0 + \beta_1 x_i.
\]

a) The least squares estimates for \( \beta_0 \) and \( \beta_1 \) are the values \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) that minimize the sum of squared deviations between the observed \( y_i \)'s and the fitted means, \( \hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \). Derive expressions for \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \).

b) A Normal linear model assumes \( Y_i|X_i = x_i \) \( \sim \) \( N(\mu_i, \sigma^2) \). Show that the least squares estimates are also the maximum likelihood estimates (assuming the \( x_i \)'s are observed and hence known) under this model. Also derive the MLE for \( \sigma^2 \).

c) Find an unbiased estimate for \( \sigma^2 \) and show that it is independent of \( \hat{\beta} \).

d) Import the height and shoe length data into R (the first column is 0 for females and 1 for males, but you can ignore this for now). Fit a simple linear regression of height on shoe length and interpret the results. Use the \texttt{lm} (linear model) function. Here are commands to get a summary, a graph of the data and a residual plot.

```r
out = lm(height ~ shoe); summary(out)
par(mfrow=c(2,1)) ### this allows two graphs in the same window
plot(shoe, height); abline(out)
plot(out$fitted, out$resid)
```

2. In a certain hypothetical community, the correlation between years of education for husband and wife pairs is 0.5, and the average and standard deviation are 12 and 3 years, respectively, for both husbands and wives.

a) Show that the least squares estimates (regardless of whether or not you specify Normal errors) can be expressed in terms of the sample means, standard deviations and the sample correlation coefficient \( r_{xy} \):

\[
r_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right).
\]

Write out the relationship between the standardized fitted means, \( (\hat{\mu}_i - \bar{y})/s_y \), and the standardized \( x \) values, \( (x_i - \bar{x})/s_x \). Use this expression to justify the term “regression to the mean.”

b) Write out the least squares equation for predicting a husband’s education from his wife’s education.

c) What years of education would you predict for the husband of a wife with 18 years of education? What years of education would you predict for the wife of a husband with 15 years of education?
d) Apparently well-educated women tend to marry men with less education, who in turn tend to marry women with even less education. Explain this apparent paradox. Give a graphical as well as a theoretical explanation.

e) Show how to get the regression estimates for predicting height from shoe length and for predicting shoe length from height from the averages, sample standard deviations and the correlation coefficient (use \( \text{cor}(x, y) \) to get the correlation).

3. Suppose \( X \) and \( Y \) follow a bivariate Normal distribution.

a) Derive the conditional distribution of \( Y \mid X = x \) and relate this to the linear regression prediction equations. By what fraction is the variance of \( Y \) reduced when \( X = x \) is observed?

b) For the linear model \( Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \), with errors \( \epsilon_i \) having mean 0, show that the least squares estimates are unbiased for \( \beta_0 \) and \( \beta_1 \), even if the \( \epsilon_i \)'s are not Normal.

c) Explain how you could find reasonable estimates for the parameters of the multiplicative model

\[
Y_i = \alpha x_i^\beta \xi_i, \quad i = 1, \ldots, n,
\]

where the errors \( \xi_i \) are independent and follow a distribution with median 1; i.e., given \( X_i = x_i \), \( P(Y_i \leq \alpha x_i^\beta) = P(X_i \geq \alpha x_i^\beta) = 0.5 \).

4. Suppose \( \theta_1, \ldots, \theta_k \stackrel{i.i.d.}{\sim} N(\mu, A) \) and \( X_i \mid \theta_i \sim N(\theta_i, V), \; i = 1, \ldots, k \). For example, the \( \theta_i \)'s might represent the mean scores on a test for the entire \( i \)th school district, and \( X_i \) is the average for a random sample of \( n \) scores from that district. If the distribution of scores for individual students in district \( i \) is \( N(\theta_i, \sigma^2) \), then the average \( X_i \) has variance \( V = \sigma^2/n \).

a) What is the marginal distribution of the \( X_i \)'s? What is the correlation between \( X_i \) and \( \theta_i \)?

b) We are most interested in the \( \theta_i \)'s but typically only observe the \( X_i \)'s. What is the conditional distribution of \( \theta_i \mid X_i = x_i \)? Relate this to the simple regression equation.

c) In a simple regression situation, we observe \( \{x, y\} \) pairs and fit a line that summarizes their relationship. But here we wish to fit a line based only on the \( x_i \)'s. What other information can we take advantage of? Demonstrate the James-Stein estimate for a data set of your choosing (or I can help you find one).

5. Turn in (From the 2004 honor exam) Suppose we have \( n \) data points \( \{x_i, y_i\} \) with \( i = 1, \ldots, n \). Assume that the data follow the model:

\[
y_i = \beta x_i + \epsilon_i
\]

where \( \epsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2) \) are random errors and the \( x_i \)'s are assumed to be fixed constants.

a) Find the least squares estimate \( \hat{\beta} \).

b) What is the sampling distribution of \( \hat{\beta} \)?

c) Suppose \( \sigma^2 \) is known. Find an expression for a 100(1 - \( \alpha \))% confidence interval for \( \beta \).

d) Suppose \( \sigma^2 \) is unknown. Find an expression for a 100(1 - \( \alpha \))% confidence interval for \( \beta \).