1. J.J. Thomson (1856–1940), discovered the electron while investigating the basic nature of cathode rays. In laboratory experiments Thomson isolated negatively charged particles for which he could determine the mass-charge ratio. This ratio appeared to be constant over a wide variety of experimental conditions and to be a characteristic of these new particles. Thomson obtained the following results with two different cathode ray tubes, using air as the gas (state any assumption you made to answer this question):

Tube 1	0.57	0.34	0.43	0.32	0.48	0.40	0.40
Tube 2	0.53	0.47	0.47	0.51	0.63	0.61	0.48

   (a) Construct a 95% confidence interval for the difference in the means of the two tubes. Do the two tubes appear to produce consistent results?

   (b) Establish a 99% confidence interval for the mean mass-charge ratio, treating Thomson's two sets of measurements as one sample of size 14.

2. Identify the independent variable x and the dependent variable y in each of the following situations:

   (a) A training director wishes to study the relationship between the duration of training for new recruits and their performance in a skilled job.

   (b) A chemist wishes to study the relationship between the drying time of a paint and the concentration of a chemical additive that is supposed to accelerate the drying process.

   (c) A market analyst wishes to relate the expenditures incurred in promoting a product in test markets and the subsequent amount of product sales.

3. A morning newspaper lists the following used-car prices for a compact, with age x measured in years and selling price y measured in thousands dollars:

x	10	11	11	12	12	13	15	16	17	19
y	2.45	1.80	2.00	2.00	1.70	1.20	1.15	0.69	0.60	0.47

   (a) Plot the scatter diagram.

   (b) Determine the equation of the least squares regression line and draw this line on the scatter diagram.

   (c) Test $H_0 : \beta_1 = 0$ vs. $H_A : \beta_1 \neq 0$.

4. File `gross.dat` contains gross national product y in real dollars for 26 recent years, $x = 1, 2, \ldots, 26$. Fit a linear regression and check the model assumptions. Which assumption(s) for a linear regression model appear to be seriously violated by the data? (NOTE: Regression methods are usually not appropriate for this type of data.)

5. **A test for the presence of one type of serial correlation:** In a multiple regression model, for instance one with three predictor variables

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + e_i, \ i = 1, 2, \ldots, n$$

it is assumed that the errors e_i are independently distributed as $N(0, \sigma)$. However, the errors may be associated in time, so that those adjacent in time have the correlation ρ. Further, if the special model holds in which errors two time units apart have correlation ρ^2, those three units apart $\rho^3, \ldots,$ those k units apart ρ^k, it is possible to test for independence. This test, called the *Durbin-Watson test*, is based on the statistic

$$d = \frac{\sum_{t=2}^{n} (u_t - u_{t-1})^2}{\sum_{t=1}^{n} u_t^2}$$

where $u_t = y_t - \hat{y}_t,\ t = 1, \ldots, n$ are the residuals arranged in time order. The distribution of d is complicated. However, two critical values d_L and d_U are tabulated in Durbin, J. and Watson, G. S., Testing for Serial Correlation in Least Squares Regression, II [*Biometrika* (1951), Vol. 38, 159-78] for different sample sizes n and for several numbers of predictor variables in the model. In testing $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$, the procedure has an unusual structure that enables us to conclude that:
(i) d is significant if $d < d_L$ or $4 - d < d_L$
(ii) d is not significant if $d > d_U$ and $4 - d > d_U$
(iii) The test is inconclusive otherwise.
To test the two-sided alternative at level α, both d_L and d_U are obtained from the above mentioned table with $\alpha/2$. One sided tests can be constructed by noting that $d < d_L$ indicates a positive correlation and that $4 - d < d_L$ indicates a negative correlation. Although correlation among the errors e_i is a serious violation of assumptions, it usually requires 20 or more observations to detect even moderately large values of ρ in this special model.

(a) Calculate d for the data on gross national product given in question 6. And make a conclusion. $(d_L, d_U) = (1.30, 1.46)$ for 26 observations and 1 predictor.

6. In an experiment involving stored-product beetles (Trogoderma glabrum) and their sex-attractant pheromone, the pheromone is placed in a pit-trap in the centers of identical square arenas. Marked beetles are then released along the diagonals of each square at various distances from the pheromone source. After 48 hours, the pit-traps are inspected. Control pit-traps containing no pheromone captured no beetles.

Release distance (centimeters)	No. of beetles captured out of 8
6.25	5,3,4,6
12.5	5,2,5,4
25	4,5,3,0
50	3,4,2,2
100	1,2,2,3

(a) Plot the original data with y =number of beetles captured. Repeat with $x = log(distance)$.

(b) Fit a straight line by least squares to the appropriate graph in (a).

(c) Test $H_0 : \beta_1 = 0$ vs. $H_A : \beta_1 \neq 0$.

(d) Establish a 95% confidence interval for the mean at a release distance of 18 cm.

(e) Check the residuals. Do you need to modify your model?