1. Consider three beacons labelled A, B and C, located, respectively at $(45, 10)$, $(0, 30)$ and $(0, 0)$ in two-dimensional space (x = longitude, y = latitude). A boat has been observed

   (i) due north of beacon A,

   (ii) due east of beacon B,

   (iii) due north-east of beacon C.

   (a) Draw a diagram to show that these three observations do not uniquely identify the position of the boat. Please be as neat as you possibly can.

   (b) Suppose now that the true position of the boat is (x, y). Express the three observations as three equations in x and y. Confirm that these three equations are inconsistent.

   (c) Set up a linear model in matrix notation to represent the above situation.

   (d) Compute ordinary least-squares estimates of (x, y).

   (e) Compute the three residuals and the residual sum of squares.

   (f) Explain why the three residuals are equal in absolute value.

   (g) Suppose now that the three observations are independent and normally distributed with white noise.

      i. Write down the joint distribution of the least-squares estimator $\hat{x}, \hat{y})$. [You need not evaluate the covariance matrix explicitly.]

      ii. Show that
      $$Q = (x - \hat{x})^2 + (y - \hat{y})^2 - (x - \hat{x})(y - \hat{y})$$
      is distributed as a multiple of central chi-squared. What is the multiplier? What are the degrees of freedom? [Hint: Express the quadratic form in matrix notation and use your answer to (i).]

      iii. Explain why the distribution of Q is independent of the distribution of the residual sum of squares S_e.

      iv. Write down the distribution of the ratio Q/S_e.

      v. Obtain a 75%-confidence region for the true position of the boat using the answer to (iv): the resulting region will be an ellipse. Draw this ellipse in on your diagram (a). [Please ask for tables if you do not have any with you.]

2. Consider the following data which were collected in the Snake River watershed (Wyoming, USA) during the 17-year period 1919 through 1935. The numbers X represent the water content of snow on April 1, while Y denotes the water yield from April through July (both in inches).

$$\sum x_j = 511.5 \quad \sum x_i^2 = 16,628.65$$

$$\sum y_i = 267.1 \ \sum y_i^2 = 4,549.43$$
$$\sum x_j y_i = 8,653.45 \ n = 17.$$

Consider the data set (adapted from *Applied Regression Analysis* by Sanford Weisberg, pub. Wiley, New York, 1980, p. 128 ff) concerning the average brain and body weights for 10 species of mammals, where x denotes the logarithms of body weight (in kg.), and y denotes the logarithms of brain weight (in grams). Both logarithms to base 10.

i		x_i	y_i
1.	Lesser short-tailed shrew (musaraigne)	-2.301	-0.854
2.	Arctic fox	0.530	1.648
3.	Guinea pig (cobaye)	0.017	0.740
4.	Star-nosed mule	-1.222	0
5.	Man	1.792	3.121
6.	Kangaroo	1.544	1.748
7.	Asian elephant	3.406	3.663
8.	African elephant	3.823	3.757
9.	Desert hedgehog (hérisson)	-0.260	0.380
10.	Giraffe	2.723	2.833

Consider the simple linear regression model:

$$E(y) = \alpha + \beta_x,$$

as a special case of the model (∗) defined at the beginning of Problem 4 on page 4. (You do not need to solve Problem 4 in order to do this problem.)

(a) Plot the data.

(b) Estimate α and β by least squares. (Use 4 decimal places.)

(c) Draw in the least squares line.

(d) It has been suggested that the 5th observation (for man) may not belong to the same line as the other 9 observations. To assess this suggestion we will use the Studentized residual W_5 as defined in (b) of Problem 4.

   i. Identify the observation for man on the data plot.

   ii. Recall that for the simple linear regression model, the matrix

$$where$$

$C is the$ 10×10 centering matrix, and $\underline{x}$ is the column vector of x_i's. Evaluate m_5, the 5th diagonal element of M.

iii. Compute the statistic W_5.

iv. What do you conclude? Use a two-sided test at level 95%. (Percentage points available in the table on page 2.)

McGILL UNIVERSITY

FACULTY OF SCIENCE

FINAL EXAMINATION

MATHEMATICS 189-423A

REGRESSION & ANALYSIS OF VARIANCE

Examiner: Professor G.P.H. Styan
Associate Examiner: Professor

Date: Tuesday, December 8, 1998
Time: 9:00 A.M. - 12:00 Noon.

INSTRUCTIONS

This exam comprises the cover and ? pages of questions.