METHODS QUALIFYING EXAM

AUGUST 2005

INSTRUCTIONS:

1. DO NOT put your NAME on the exam. Place the NUMBER assigned to you on the UPPER LEFT HAND CORNER of EACH PAGE of your exam.
2. Please start your answer to EACH QUESTION on a SEPARATE sheet of paper.
3. Answer all the questions.
4. Be sure to attempt all parts of every question. It may be possible to answer a later part of a question without having solved the earlier parts.
5. Be sure to hand in all of your exam. No additional material will be accepted once the exam has ended and you have left the exam room.

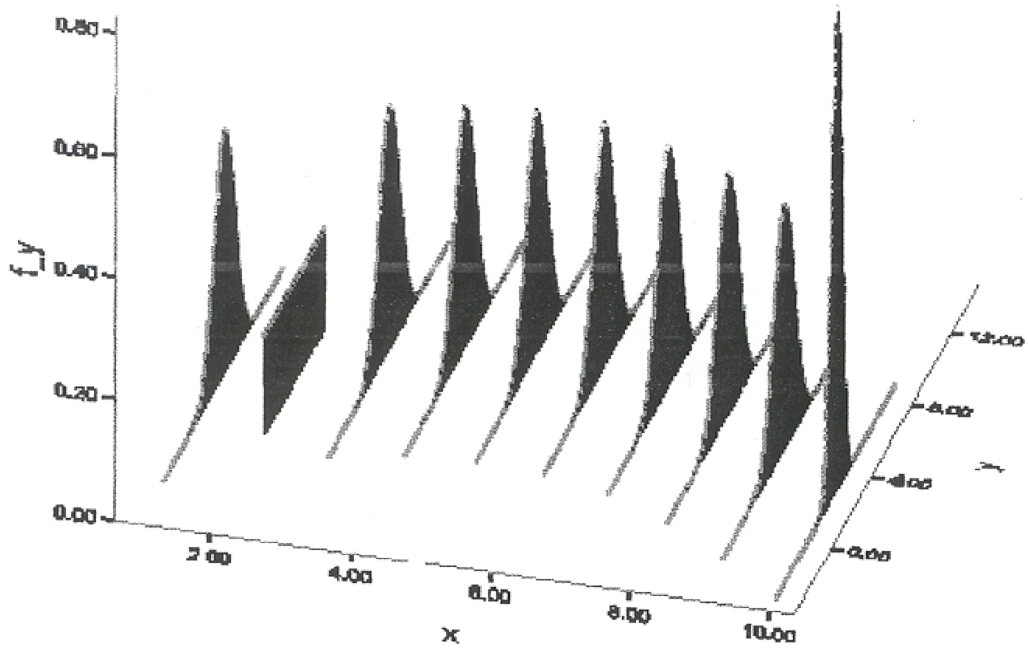
Problem 1. Regression Problem:

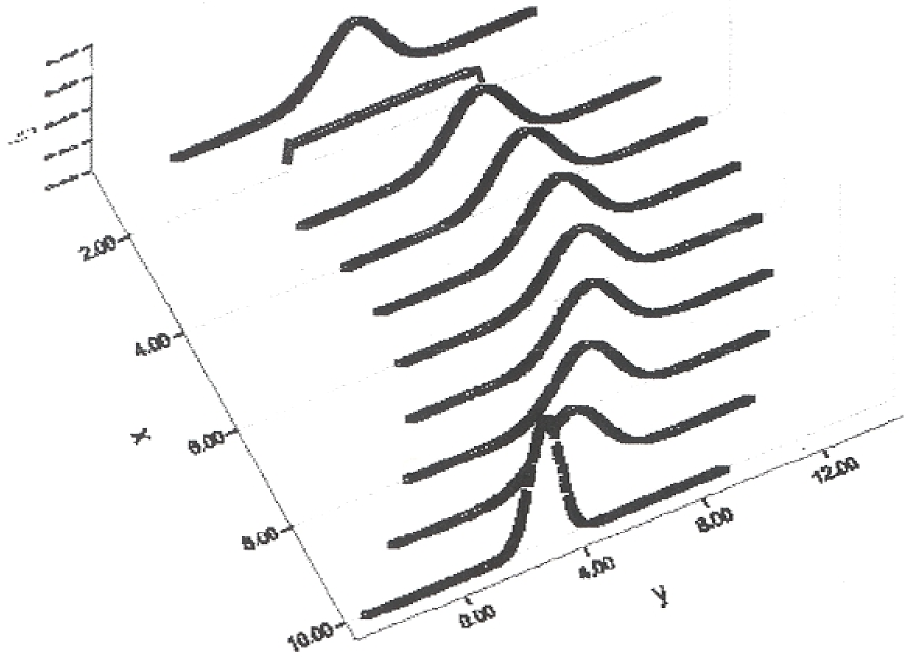
We have observed y = response (change in blood pressure) and x = dosage level of a drug. We assume a polynomial relationship between y and x.

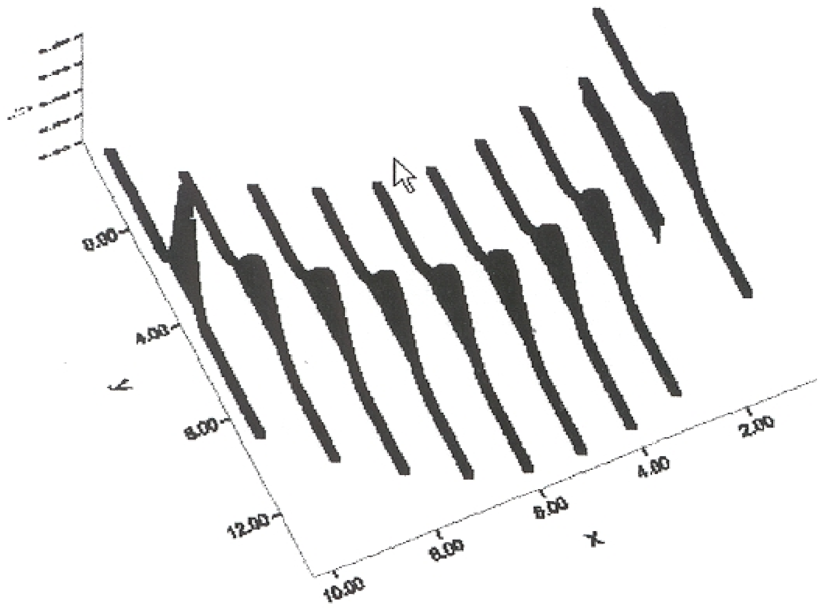
The three graphs are all the same; but have been rotated to give additional views.

1. What is the appropriate model?
2. What are the usual regression assumptions?
3. Answer the following (in detail where appropriate):

    a. Sketch E(y)
    b. Based on the graphs, make comments about the assumptions. Do they appear to be satisfied or violated?
    c. How many populations are represented by the graphs?
    d. For x = 10, sketch the pdf of y. Label the axis, approximate min and max values such that f_y > 0, indicate the mean and standard deviation.
    e. List all of the parameters in the model
    f. What is the effect of non-normal errors on hypothesis testing?
    g. What is the effect of non-constant variance on hypothesis testing?


0.80
0.60
0.40
0.20
0.00
f_y
2.00
4.00
6.00
8.00
10.00
x
y


2.00
4.00
x
6.00
8.00
10.00
0.00
4.00
y
8.00
12.00


0.00
4.00
y
8.00
12.00
10.00
8.00
6.00
x
4.00
2.00

Problem 2.

This problem investigates the effect of correlated errors on least squares estimators. In order to keep the algebraic complexity to a minimum, we consider the simple linear (no-intercept) regression model:

$$Y_t = \beta X_t + u_t \ ,$$

where we assume that the disturbance u_t follows the first-order autoregressive scheme:

$$u_t = \rho u_{t-1} + \varepsilon_t \ ,$$

where $|\rho| < 1$ and ε_t satisfies the assumptions:

$$\left.\begin{aligned} E(\varepsilon_t) &= 0 \\ E(\varepsilon_t \varepsilon_{t+s}) &= \sigma_\varepsilon^2 \quad \text{if } \ s = 0 \\ &= 0 \qquad \text{if } \ s \neq 0 \end{aligned}\right\} \text{ for all } t \ .$$

Assume that the independent variables, X_t, are random variables that we treat as "fixed" by conditioning on their observed values.

a) Prove heuristically (i.e., show the algebra, but neither prove convergence nor justify interchange of limits rigorously) that:

$$u_t = \Sigma_{r=0}^{\infty} \rho^r \varepsilon_{t-r}^{r} \ ,$$

and, therefore, that:

$$E(u_t) = 0 \ .$$

b) Using the representation for u_t given in part a), show (heuristically, as defined in part a)) that:

$$E(u_t u_{t-s}) = \rho^s \sigma_u^2 \ ,$$

where:

$$\sigma_u^2 = \frac{\sigma_\varepsilon^2}{1-\rho^2} \ .$$

c) Show that the least squares estimator (LSE) of β under this model is: (You may use without proof the matrix formula $\hat{\beta} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$, but define clearly the elements of β, X and $\boldsymbol{Y}$.)

$$\hat{\beta} = \frac{\sum_{t=1}^{n} X_t Y_t}{\sum_{t=1}^{n} X_t^2} .$$

d) Conditioning on the X_t's (i.e. treating them as "fixed"), show that:

$$Var\left(\hat{\beta}\right) = \left(\frac{\sigma_u^2}{\sum_{t=1}^{n} X_t^2}\right) \times \left(1 + 2\rho\frac{\sum_{t=1}^{n-1} X_t X_{t+1}}{\sum_{t=1}^{n} X_t^2} + 2\rho^2\frac{\sum_{t=1}^{n-2} X_t X_{t+2}}{\sum_{t=1}^{n} X_t^2} + \ldots + 2\rho^{n-1}\frac{X_1 X_n}{\sum_{t=1}^{n} X_t^2}\right) .$$

e) As a further algebraic simplification, suppose that the X_t's also follow a first-order autoregressive scheme with parameter ρ identical to that of the u_t's, so that for large n:

$$\frac{\sum_{t=1}^{n-s} X_t X_{t+s}}{\sum_{t=1}^{n} X_t^2} \doteq \rho^s .$$

Using this approximation (without proving it), show that:

$$Var\left(\hat{\beta}\right) \doteq \left(\frac{\sigma_u^2}{\sum_{t=1}^{n} X_t^2}\right) \times \left(\frac{1+\rho^2}{1-\rho^2}\right) .$$

Evaluate this approximate expression for the variance of the least LSE of β when $\rho = 0.8$. Comment on this effect of correlated errors *under this set of assumptions* on the LSE of the regression coefficient, β.

f) Under the same assumptions used in part e) to derive the approximate expression for $Var(\hat{\beta})$, it can be shown that (but you need <u>not</u> show this):

$$E\left(s^2\right) \doteq \left(\frac{\sigma_u^2}{n-1}\right)\left(n - \frac{1+\rho^2}{1-\rho^2}\right) ,$$

where s^2 is the usual LSE of variance, i.e.:

$$s^2 = \left(\frac{1}{n-1}\right)\sum_{t=1}^{n}\left(Y_t - \hat{\beta}X_t\right)^2 .$$

Evaluate this approximate expression for the expected value of the least LSE of variance when $\rho = 0.8$. Comment on this effect of correlated errors *under this set of assumptions* on the LSE of variance.

g) Taken together, comment on the results of parts e) and f) as they pertain to least squares inferences concerning β *under this set of assumptions*.

Problem 3.

A graduate student in a Statistics Department has extended their advisor's robust estimation method. The student is in the process of putting a research proposal together. The advisor wants the student to compare the modified method to some well-known methods in a simulation experiment. The advisor tells the student that they are expected to use a well-designed experiment. The advisor's method takes about 1 minute to calculate on a sample size of about 100 and increases linearly with sample size. The method works on univariate Y data having model $Y = X + \sigma\varepsilon$. Here the X and ε are independent random variables. The parameter σ is measurement error.

The advisor tells the student that the following factors need to be included in the study. Those factors are sample size, error distributions (at least the normal, a heavy tailed distribution, and an asymmetric distribution), different levels of measurement error variance and the advisor's guess is that the MSE of all estimators for this problem will increase as a quadratic function of error variance, and will depend on the distribution of the true unobserved X values (use at least 3 or more types). The student is to compare their new method to three standard robust estimation methods, one of which is the sample median.

Assume that you are to design the experiment.

1. List at least two important considerations in choosing the levels of each of the following factors:
   a. sample sizes
   b. error distributions
   c. measurement error variance
   d. distribution for the underlying X.
2. What general class of experiment would you choose and why? That is, would you choose a fractional factorial, repeated measure, D-optimal design, etc.
3. How many sample sizes might you choose and why?
4. How many error variances would you choose and why?
5. How would you choose the number of replicates?
6. Once the experiment is completed, what method(s) do you expect to use to compare the results, and why?

Problem 4.

It is desired to estimate the weight of each of 4 objects on a balance. Let $\beta_1, \beta_2, \beta_3, \beta_4$ be their actual weights. Weight measurements are subject to additive error, so that for any given object with true weight β, its measured weight Y is a random variable satisfying

$$Y = \beta + \epsilon$$

where $E(\epsilon) = 0$ and $Var(\epsilon) = \delta^2$. Suppose that each object is weighed separately once, the first and second objects are weighed together once and the third and fourth objects are weighed together once as well, yielding measurements $Y_1, Y_2, \ldots, Y_6$ satisfying

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_1 + \beta_2 \\ \beta_3 + \beta_4 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}$$

where $E(\epsilon_j) = 0$ and $Var(\epsilon_j) = \delta^2$ and $Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$.

(a) Find the best linear unbiased estimator of $\beta = (\beta_1, \ldots, \beta_4)$.

(b) For the estimators $\hat{\beta}_j, j = 1, \ldots, 4$ of part (a) calculate $\mathrm{Var}(\hat{\beta}_j)$.

(c) Suppose that you were told that $\beta_1 - \beta_2 = m_1$ and $\beta_3 - \beta_4 = m_2$ where m_1 and m_2 are known quantities. Provide a linear unbiased estimator of $\beta = (\beta_1, \ldots, \beta_4)$ that incorporates this additional knowledge. What is the variance of the new estimators of $\beta_j, j = 1, \ldots, 4$?