Ma 6266
Problems I
Spring 1999

1. Let $Y_1, \dots, Y_n$ be independent

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

where ε_i are iid N(0,1) and β_0 and β_1 are unknown. The x_i are assumed to be known and not all equal.

(a) Find the most powerful test of

$$H_0 : \beta_0 = 0, \ \beta_1 = 2 \quad \text{against } H_A : \ \beta_0 = 0, \ \beta_1 = 3$$

of size α.

(b) Is there a uniformly most powerful test of

$$H_0 : \beta_0 = 0, \ \beta_1 = 2 \quad \text{against } H_A : \ \beta_0 = 0, \ \beta_1 > 2$$

of size α ? If so, give the test.

(c) Give the likelihood ratio test of

$$H_0 : \beta_0 = 0 \quad \text{against } H_A : \ \beta_0 \neq 0$$

of size α.

2. Let V be a vector space and denote by sp{ $v_1, \dots, v_k$ } the linear span of the set of vectors {$v_1, \dots, v_k$ }.

a) Show that sp{$v_1, \dots, v_k$} is a subspace of V.

b) Prove that if v   sp{$v_1, \dots, v_k$} and {$v_1, \dots, v_k$} is a linearly independent set then {$v_1, \dots, v_k$, v} is a linearly independent set.

c) A vector space V is said to have dimension n (dim(V) = n) if every set containing n+1 vectors is a linearly dependent set and there is a linearly independent set of n elements. It can be shown that if dim(V) = n then every basis for V has n elements. Assuming this fact, prove that if dim(V) = n, if k < n, and if { $v_1, \dots, v_k$ } are linearly independent, then there are vectors {$v_{k+1}, \dots, v_n$} in V such that {$v_1, \dots, v_k, v_{k+1}, \dots, v_n$} is a basis for V.

3. The mix moisture x_1 and the relative humidity x_2 are considered to have an effect on the density Y of a finished product. An experiment yielded the following coded data.

x_1	x_2	y
4.7	-1	3
5.0	1	3
5.2	2.5	4
5.2	-1	5
5.9	-3	10
4.7	1.5	2

Write out the Z matrix for the cases

a) $E[Y] = \beta_0 + \beta_1 x_1 + \beta_2 x_2$,

b) $E[Y] = \beta_0 + \beta_1 x_1 + \beta_{12} x_1 x_2$.

c) Find the numerical values of the least squares estimates in each of the two cases above for this data set.

4. (Scheffé) Let (a) and (A) denote the models

(a): $Y_i = \alpha + \beta x_i + \varepsilon_i$, $E[e_i] = 0, E[\varepsilon_i \varepsilon_{i'}] = \sigma^2 \delta_{ii'}$

(A): $Y_i = \alpha + \beta x_i + \gamma x_i^2 + \varepsilon_i$, $E[e_i] = 0, E[\varepsilon_i \varepsilon_{i'}] = \sigma^2 \delta_{ii'}$.

Find, by using ordinary differentiation techniques from elementary calculus, the least squares estimates of α and β under (a) and express the least squares estimates of α, β, and γ under (A) using determinants.

5. In problem 4. find the variances and covariances of the estimates of α and β under (a). Show that if we take $\delta + \beta(x_i - \bar{x})$ in place of $\alpha + \beta x_i$ under (a), then $\hat{\delta} = \bar{y}$ and $\text{cov}(\hat{\delta}, \hat{\beta}) = 0$.

6. In problem 4. express $\text{var}(\hat{\gamma})$ using determinants.

7. Prove the following lemma: If $Y' = (Y_1, \ldots, Y_n)$, $E[Y] = \eta$, $\varepsilon = Y - \eta$, and $Q(y)$ is a quadratic form in y, then $E[Q(Y)] = Q(\eta) + E[Q(e)]$. Note that $Q(\eta)$ may be evaluated by replacing the y_i by their expected values in $Q(Y)$, and that $E[Q(\varepsilon)]$ is the value of $E[Q(Y)]$ when $\eta = 0$.

8. Suppose $Y, X_1, \ldots, X_k$ are random variables with joint probability density function f and Y has a variance. The best predictor $g(x_1, \ldots, x_k)$ of Y in terms of mean square error is the function g which minimizes $E[(Y - g(X_1, \ldots, X_K))^2$.

(a) Show that

$$g(x_1, \ldots, x_k) = E[Y \mid X_1 = x_1, \ldots, X_k = x_k].$$

The function g( ) is called the regression function of y on $x_1, \ldots, x_k$.

b) If

$$f(x,y) = \begin{cases} 2 & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le x \\ 0 & \text{otherwise} \end{cases}$$

find the regression function of Y on X.

c) In b) find $\min E[(Y - g(X))^2]$.

d) Consider linear predictors $\lambda(x) = a + bx$ of Y based on X . Find a and b, which minimize

$$E[(Y - \lambda(X))^2].$$

e) Express $\min E[(Y - \lambda(X))^2]$ in terms of σ_X^2, σ_Y^2, and ρ_{XY}.

f) If $f(y, x_1 \ldots, x_k) = (\sum_{j=1}^{k} x_j) \prod_{j=1}^{k} x_j \exp\{-(1+y)(\sum_{j=1}^{k} x_j)\}$ for all x's and $y > 0$, find the regression function of y on the x's.

g) Show that close to the point $(\bar{x}_1, \ldots, \bar{x}_k)$ the regression function is approximately linear in that there are constants $\beta_0, \ldots, \beta_k$ such that

$$| g(x) - (\beta_0 + \sum_{j=1}^{k} \beta_j x_j) | = o(|| x - \bar{x} ||)$$

and find the constants.