1 Useful Background Information

In this section of the notes, various definitions and results from calculus, linear/matrix algebra, and least-squares regression will be summarized. I will refer to these items at various times during the semester.

1.1 Taylor Series

1. Let $\eta^{(k)}(x)$ denote the k^{th} derivative of function $\eta(x)$. For function η and x_0 in some interval I, define

$$P_n(x,x_0) \;=\; \eta(x_0) \;+\; \eta^{(1)}(x_0)(x-x_0) \;+\; \eta^{(2)}(x_0)(x-x_0)^2 \;+\; \cdots \;+\; \eta^{(n)}(x_0)(x-x_0)^n$$
$$R_n(x,c) \;=\; \frac{(x-c)^{n+1}}{(n+1)!}\,\eta^{(n+1)}(c).$$

   Then, there exists some number z between x and x_0 such that

$$\eta(x) \;=\; P_n(x,x_0) \;+\; R_n(x,z)$$

2. **Taylor Series for functions of one variable**: If η is a function that has derivatives of all orders throughout an interval I containing x_0 and if $\lim_{n\to\infty} R_n(x,x_0) = 0$ for every x_0 in I, then $\eta(x)$ can be represented by the Taylor series about x_0 for any x_0 in I. That is,

$$\eta(x) \;=\; \eta(x_0) \;+\; \sum_{k=1}^{\infty} \frac{(x-x_0)^k}{k!}\,\eta^{(k)}(x_0)$$

3. Note that $P_n(x,x_0)$ is a polynomial of degree n. Thus, $P_n(x,x_0)$ is an n^{th}-order Taylor series approximation of $\eta(x)$ because $R_n(x,x_0)$ vanishes as n increases.

4. Practically, this means that even if the true form of $\eta(x)$ is unknown, we can use a polynomial $f(x) = P_n(x,x_0)$ to approximate it with the approximation improving as n increases.

5. In statistics, we may fit a linear model

$$f(x) \;=\; \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n.$$

   What we are actually doing is fitting

$$f(x) \;=\; P_n(x,0) \;=\; \eta(0) \;+\; \eta^{(1)}(0)x \;+\; \eta^{(2)}(0)x^2 \;+\; \cdots \;+\; \eta^{(n)}(0)x^n$$

   where $\beta_0 = \eta_0$ and $\beta_i = \eta^{(i)}(0)x^i$ for $i = 1,2,\ldots,n$ and we assume the remainder $R_n(x,0)$ is negligible.

6. Taylor series can be generalized to higher dimensions. I will only review the 2-dimensional case.

7. For function $\eta(x,y)$ let $\dfrac{\partial\eta^n}{\partial x^k \partial y^{n-k}}$ be the nth-order partial derivative with derivation taken k times with respect to x and $(n-k)$ times with respect to y.

8. If η is a function of (x,y) that has partial derivatives of all orders inside a ball B containing $\boldsymbol{p}_0$ and if $\lim_{n\to\infty} R_n(\boldsymbol{p},\boldsymbol{p}_0) = 0$ for every $\boldsymbol{p}_0$ in B, then $\eta(\boldsymbol{p})$ can be represented by the 2-variable Taylor series about $\boldsymbol{p}_0$ for any $\boldsymbol{p}_0$ in B.

9. For function $\eta(x,y)$ and $\boldsymbol{p}_0 = (x_0, y_0)$ in some open ball B containing $\boldsymbol{p}_0$, define $\boldsymbol{p} = (x,y)$ and

$$
\begin{aligned}
P_n(\boldsymbol{p},\boldsymbol{p}_0) &= \eta(\boldsymbol{p}_0) + \frac{(x-x_0)}{1!}\left.\frac{\partial\eta}{\partial x}\right|_{\boldsymbol{p}_0} + \frac{(y-y_0)}{1!}\left.\frac{\partial\eta}{\partial y}\right|_{\boldsymbol{p}_0} \\
&+ \frac{(x-x_0)^2}{2!}\left.\frac{\partial^2\eta}{\partial x^2}\right|_{\boldsymbol{p}_0} + \frac{(x-x_0)(y-y_0)}{1!1!}\left.\frac{\partial^2\eta}{\partial x\partial y}\right|_{\boldsymbol{p}_0} + \frac{(y-y_0)^2}{2!}\left.\frac{\partial^2\eta}{\partial y^2}\right|_{\boldsymbol{p}_0} \\
&+ \cdots \\
&+ \sum_{k=0}^{n-1}\frac{(x-x_0)^k(y-y_0)^{(n-1-k)}}{k!(n-1-k)!}\left.\frac{\partial^k\eta}{\partial^k x\partial^{(n-1-k)}y}\right|_{\boldsymbol{p}_0} \\
&+ \sum_{k=0}^{n}\frac{(x-x_0)^k(y-y_0)^{(n-k)}}{k!(n-k)!}\left.\frac{\partial^k\eta}{\partial^k x\partial^{(n-k)}y}\right|_{\boldsymbol{p}_0}
\end{aligned}
$$

$$
R_n(\boldsymbol{p},\boldsymbol{p}^*) = \sum_{k=0}^{n+1}\frac{(x-x_0)^k(y-y_0)^{(n+1-k)}}{k!(n+1-k)!}\left.\frac{\partial^k\eta}{\partial^k x\partial^{(n-k)}y}\right|_{\boldsymbol{p}^*}
$$

where $\boldsymbol{p}^*$ is a point on the line segment joining $\boldsymbol{p}$ and $\boldsymbol{p}_0$.

10. **Taylor Series for functions of two variables**: There exists some point $\boldsymbol{p}_z$ on the line segment joining $\boldsymbol{p}$ and $\boldsymbol{p}_0$ such that

$$\eta(\boldsymbol{p}) = P_n(\boldsymbol{p},\boldsymbol{p}_0) + R_n(\boldsymbol{p},\boldsymbol{p}_z)$$

11. Note that $P_n(\boldsymbol{p},\boldsymbol{p}_0)$ is a polynomial of degree n in variables x and y. Thus, $P_n(\boldsymbol{p},\boldsymbol{p}_0)$ is an n^{th}-order Taylor series approximation of $\eta(\boldsymbol{p})$ because $R_n(\boldsymbol{p},\boldsymbol{p}_0)$ vanishes as n increases.

12. Practically, this means that even if the true form of $\eta(\boldsymbol{p})$ is unknown, we can use a polynomial $f(\boldsymbol{p}) = P_n(\boldsymbol{p},\boldsymbol{p}_0)$ to approximate it with the approximation improving as n increases.

13. In statistics, we may fit a linear model

$$f(x,y) = \sum_{i=0}^{n}\sum_{j=0}^{n-i}\beta_{i,j}x^i y^j$$

What we are actually doing is fitting $f(x) = P_n(\boldsymbol{p},(0,0))$ where $\beta_{0,0} = \eta(0,0)$ and $\beta_{i,j} = \eta^{(i+j)}(0,0)x^i y^j$ for $i+j = 1,2,\ldots,n$, and we assume the remainder $R_n(\boldsymbol{p},(0,0))$ is negligible.

14. On the following page: $f_{12} = \dfrac{\partial^2 f}{\partial x\partial y} \qquad f_{11} = \dfrac{\partial^2 f}{\partial x^2} \qquad f_{22} = \dfrac{\partial^2 f}{\partial y^2}$.

Thus, $\Delta = \left(\dfrac{\partial^2 f}{\partial x\partial y}\right)^2 - \left(\dfrac{\partial^2 f}{\partial x^2}\right)\left(\dfrac{\partial^2 f}{\partial y^2}\right)$.

3.6 EXTREMAL PROBLEMS

Any continuous function f defined on a closed bounded set D attains a maximum (and a minimum) value at some point of D. If f is of class C' in D, and p_0 is an interior point of D at which f attains such an extremal value, then (Theorem 11, Sec. 3.3) all the first-order partial derivatives of f are 0 at p_0. This suggests that we single out the points in the domain of a function which have the last property.

Definition 4 *A **critical point** for a function f is a point p where*

$$f_1(p) = f_2(p) = \cdots = 0$$

The discussion in the first paragraph can be rephrased as asserting that the extremal points for the function f, which lie in the set D but do not lie on the boundary of D, are among the critical points for f in D. A critical point need not yield a local maximum or minimum value of f. However, since such a point is one where the directional derivative of f is 0 in every direction, the point is a stationary point for the function; this is reflected in the fact that the tangent hyperplane to the graph of f will be horizontal there. Since such a point can be one where the surface rises in one direction and falls in another, as in Fig. 3-9, a critical point need not correspond to either a maximum or a minimum. Such points are often called **saddle points,** or **minimax** points.

Before proceeding further, let us recall the facts about functions of one variable. A critical point is a solution of the equation $f'(x) = 0$, and corresponds to a point on the curve with equation $y = f(x)$ at which the tangent line is horizontal. The critical point may be an extremal point for f, or it may yield a point of inflection on the curve.

In the more familiar case of a function of one variable, the second derivative may be used to test the nature of a critical point.

Theorem 18 *Let f be of class C'' in the interval $[a, b]$ and let c be an interior point of this interval with $f'(c) = 0$. Then, in order that c be a local maximum point for f, it is necessary that $f''(c) \le 0$, and sufficient that $f''(c) < 0$; for c to be a minimum point, the conditions are the same with the inequality signs reversed.*

FROM 'ADVANCED CALCULUS' BUCK
McGRAW HILL

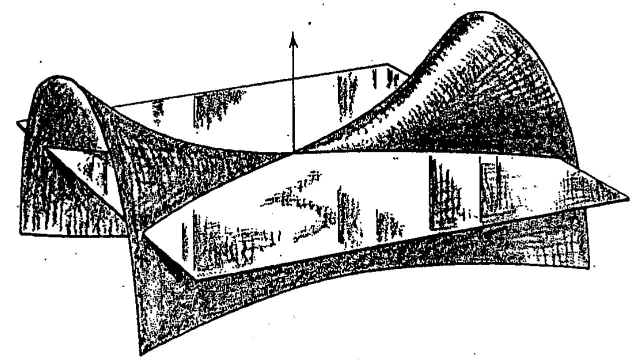

Figure 3-9 Horizontal tangent plane at saddle point.

What is the corresponding statement for functions of two variables? Clearly, if p_0 is a critical point for f lying interior to a set D, and if f has an extremal value at p_0, then this must also be extremal if we examine the values of f on any curve passing through p_0. In particular, approaching p_0 along the vertical and the horizontal directions, a necessary condition that p_0 be a maximum point for f is that $f_{11}(p_0) \le 0$ and $f_{22}(p_0) \le 0$. These conditions are not sufficient, nor are the conditions obtained by removing the equal signs (see Exercise 5). For example, the function given by $f(x, y) = xy$ has $(0, 0)$ for a critical point and

$$f_{11}(0, 0) = f_{22}(0, 0) = 0$$

but $(0, 0)$ is neither a maximum point nor a minimum point. The shape of the graph of f is again like the saddle shown in Fig. 3-9. The name "saddle point" or "minimax" is given to a critical point for a function which does not yield either a local maximum or a local minimum value for the function. A simple condition which is sufficient to ensure that a critical point p_0 be a *saddle point* is that $f_{11}(p_0)f_{22}(p_0)$ be strictly negative, since this implies that f has a local maximum at p_0 when p_0 is approached along one axis direction, and a local minimum when p_0 is approached along the other. A more general criterion can also be obtained.

Theorem 19 *Let f be of class C'' in a neighborhood of the critical point p_0, and let*

$$\Delta = (f_{12}(p_0))^2 - f_{11}(p_0)f_{22}(p_0)$$

Then,

(i) *If $\Delta > 0$, p_0 is a saddle point for f.*
(ii) *If $\Delta < 0$, p_0 is an extremal point for f, and is a maximum if $f_{11}(p_0) < 0$ and a minimum if $f_{11}(p_0) > 0$.*
(iii) *If $\Delta = 0$, the nature of p_0 is not determined by this test.*

In condition (ii), $f_{22}(p_0)$ may also be used to distinguish between maxima and minima.

1.2 Matrix Theory Terminology and Useful Results

15. If $\boldsymbol{x} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1k} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2k} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3k} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{nk} \end{bmatrix}$ then the symmetric matrix $\boldsymbol{X'X}$ can be written as

$$\boldsymbol{X'X} = \begin{bmatrix} \sum_{p=1}^{n} x_{p1}^2 & \sum_{p=1}^{n} x_{p1}x_{p2} & \sum_{p=1}^{n} x_{p1}x_{p3} & \cdots & \sum_{p=1}^{n} x_{p1}x_{pk} \\ & \sum_{p=1}^{n} x_{p2}^2 & \sum_{p=1}^{n} x_{p2}x_{p3} & \cdots & \sum_{p=1}^{n} x_{p2}x_{pk} \\ \text{symmetric} & & \sum_{p=1}^{n} x_{p3}^2 & \cdots & \sum_{p=1}^{n} x_{p3}x_{pk} \\ & & & \cdots & \cdots \\ & & & & \sum_{p=1}^{n} x_{pk}^2 \end{bmatrix}$$

16. Transpose of a product of two matrices: $(AB)' = B'A'$.
17. Transpose of a product of k matrices: If $B = A_1A_2\cdots A_{k-1}A_k$ then $B' = A_k'A_{k-1}'\cdots A_2'A_1'$.
18. The **trace** of a square matrix A, denoted $\text{tr}(A)$, is the sum of the diagonal elements of A.
19. For two k-square matrices A and B, $\text{tr}(A \pm B) = \text{tr}(A) \pm \text{tr}(B)$.
20. Given an $m \times n$ matrix A and an $n \times m$ matrix B, then $\text{tr}(AB) = \text{tr}(BA)$.
21. The rank of a matrix A, denoted $\text{rank}(A)$, is the number of linearly independent rows (or columns) of A.
22. If the determinant is nonzero for at least one matrix formed from r rows and r columns of matrix A but no matrix formed from $r+1$ rows and $r+1$ columns of A has nonzero determinant, then the rank of A is r.
23. Consider a k-square matrix A with $\text{rank}(A) = k$. The k-square matrix A^{-1} where $AA^{-1} = A^{-1}A = I_k$ is called the **inverse** matrix of A.
24. A k-square matrix A is **singular** if A is not invertible. This is equivalent to saying $|A| = 0$ or $\text{rank}(A) < k$.
25. Any nonsingular square matrix (i.e., its determinant $\neq 0$) will have a unique inverse.
26. In the use of least squares as an estimation procedure, it is often required to invert matrices which are symmetric. The inverse matrix is also important as a means of solving sets of simultaneous independent linear equations. If the set of equations is not independent, there is no unique solution.
27. The set of k linearly independent equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1k}x_k &= g_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2k}x_k &= g_2 \\ \cdots \quad \cdots \quad \cdots & \\ a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kk}x_k &= g_k \end{aligned}$$

can be written in matrix form as $\boldsymbol{Ax} = \boldsymbol{g}$. Thus, the solution is $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{g}$.

28. If $A = \text{diag}(a_1, a_2, \cdots, a_k)$ is a diagonal matrix with nonzero diagonal elements $a_1, a_2, \cdots, a_k$, then $A^{-1} = \text{diag}(1/a_1, 1/a_2, \cdots, 1/a_k)$ is a diagonal matrix with diagonal elements $1/a_1, 1/a_2, \cdots, 1/a_k$.

29. If S is a nonsingular symmetric matrix, then $(S^{-1})' = S^{-1}$. Thus, the inverse of a nonsingular symmetric matrix is itself symmetric.

30. A square matrix A is **idempotent** if $A^2 = A$.

31. A nonsingular k-square matrix $\boldsymbol{P}$ is **orthogonal** if $\boldsymbol{P}' = \boldsymbol{P}^{-1}$, or equivalently, $\boldsymbol{P}\boldsymbol{P}' = I_k$.

32. Suppose $\boldsymbol{P}$ is a k-square orthogonal matrix, $\boldsymbol{x}$ is a $k \times 1$ vector, and $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{x}$ is a $k \times 1$ vector. The transformation $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{x}$ is called an **orthogonal transformation**.

33. If $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{x}$ is an orthogonal transformation then $\boldsymbol{y}'\boldsymbol{y} = \boldsymbol{x}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{x}$.

1.3 Eigenvalues, Eigenvectors, and Quadratic Forms

34. If A is a k-square matrix and λ is a scalar variable, then $A - \lambda I_k$ is called the **characteristic matrix** of A.

35. The determinant $|A - \lambda I_k| = h(\lambda)$ is called the **characteristic function** of A.

36. The roots of the equation $h(\lambda) = 0$ are called the **characteristic roots** or **eigenvalues** of A.

37. Suppose λ^* is an eigenvalue of a k-square matrix $\boldsymbol{A}$, then an **eigenvector** associated with λ^* is defined as a column vector $\boldsymbol{x}$ which is a solution to $\boldsymbol{A}\boldsymbol{x} = \lambda^*\boldsymbol{x}$ or $(\boldsymbol{A} - \lambda^* I_k)\boldsymbol{x} = 0$.

38. An important use of eigenvalues and eigenvectors in response surface methodology is in the application to problems of finding optimum experimental conditions.

39. The quadratic form in k variables $x_1, x_2, \ldots, x_k$ is

$$\boldsymbol{Q} \;=\; \sum_{i=1}^{k} b_{ii}x_i^2 \;+\; 2\sum\sum_{i<j} b_{ij}x_ix_j \tag{1}$$

where we assume the elements b_{ij} $(i = 1, \ldots, k \quad j = 1, \ldots, k)$ are real-valued.

40. In matrix notation: $\boldsymbol{Q} = \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$ where

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_k \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ & b_{22} & \cdots & b_{2k} \\ & & \cdots & \cdots \\ \text{symmetric} & & & b_{kk} \end{bmatrix}$$

41. $\boldsymbol{B}$ and $|\boldsymbol{B}|$ are, respectively, called the matrix and determinant of the quadratic form $\boldsymbol{Q}$.

42. If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the eigenvalues of the symmetric matrix $\boldsymbol{B}$, then there exists an orthogonal transformation $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{w}$ with $\boldsymbol{w} = (w_1, w_2, \ldots, w_k)'$ such that the quadratic form $\boldsymbol{Q} = \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$ is transformed to the **canonical form**

$$\boldsymbol{Q} \;=\; \lambda_1 w_1^2 + \lambda_2 w_2^2 + \cdots + \lambda_k w_k^2 \;=\; \boldsymbol{w}'\boldsymbol{\Lambda}\boldsymbol{w} \tag{2}$$

where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)$. That is, the quadratic form $\boldsymbol{Q}$ can be transformed to one whose matrix $\boldsymbol{\Lambda}$ is diagonal whose elements are the eigenvalues of $\boldsymbol{B}$.
A manipulation of this type is extremely useful in describing the nature of a response surface and locating regions of optimum conditions.

43. The rank of a quadratic form $\boldsymbol{Q} = \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$ is defined to be rank($\boldsymbol{B}$) = the number of nonzero eigenvalues of $\boldsymbol{B}$.

44. An **indefinite quadratic form** $\boldsymbol{Q} = \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$ is one whose canonical form given in (2) contains both positive and negative coefficients, or equivalently, $\boldsymbol{B}$ has both positive and negative eigenvalues.

45. Suppose $\boldsymbol{B}$ is full (rank($\boldsymbol{B}$) $= k$).

    (a) If all eigenvalues are positive, then the quadratic form $\boldsymbol{Q}$ is **positive definite**.

    (b) If all eigenvalues are negative, then the quadratic form $\boldsymbol{Q}$ is **negative definite**.

46. Suppose $\boldsymbol{B}$ is less than full rank (rank($\boldsymbol{B}$) $< k$). That is, suppose at least one eigenvalue is zero.

    (a) If all nonzero eigenvalues are positive, then the quadratic form $\boldsymbol{Q}$ is **positive semidefinite**.

    (b) If all nonzero eigenvalues are negative, then the quadratic form $\boldsymbol{Q}$ is **negative semidefinite**.

47. The sign of a quadratic form $\boldsymbol{Q} = \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$ and the quadratic form type (i.e., positive definite, negative definite, etc.) are linked in the following way:

    (a) An indefinite quadratic form is positive for some $(x_1, x_2, \ldots, x_k)$, and negative for others.

    (b) A positive definite quadratic form is positive for all $(x_1, x_2, \ldots, x_k) \neq (0, \ldots, 0)$.

    (c) A negative definite quadratic form is negative for all $(x_1, x_2, \ldots, x_k) \neq (0, \ldots, 0)$.

    (d) A positive semidefinite quadratic form is nonnegative (≥ 0) for all real values of $x_1, x_2, \ldots, x_k$.

    (e) A negative semidefinite quadratic form is nonpositive (≤ 0) for all real values of $x_1, x_2, \ldots, x_k$.

48. All of these definitions also apply to the symmetric matrix $\boldsymbol{B}$ in the quadratic form $\boldsymbol{Q}$.

49. Theorem 1: If $\boldsymbol{X}$ is an $n \times p$ matrix ($p < n$) with rank($\boldsymbol{X}$) $= p$ (i.e., full column rank), then the $p \times p$ matrix $\boldsymbol{X'X}$ is positive definite and the $n \times n$ matrix $\boldsymbol{XX'}$ is positive semidefinite.

1.4 Matrix Differentiation

50. The column vector of partial derivatives of $f(\boldsymbol{z})$ with respect to $\boldsymbol{z}$ is

$$\partial f/\partial \boldsymbol{z} \;=\!= \begin{bmatrix} \partial f/\partial z_1 \\ \partial f/\partial z_2 \\ \cdots \\ \partial f/\partial z_k \end{bmatrix} \qquad \text{where } \boldsymbol{z} = \begin{bmatrix} z_1 \\ z_2 \\ \cdots \\ z_k \end{bmatrix}.$$

51. $\partial f/\partial \boldsymbol{z}'$ is the row vector of partial derivatives.

52. Rule 1: If $\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_k \end{bmatrix}$ is column vector of k constants, and if $f(\boldsymbol{z}) = \boldsymbol{a}'\boldsymbol{z} = \sum a_i z_i$, then

$$\partial(\boldsymbol{a}'\boldsymbol{z})/\partial \boldsymbol{z} = \boldsymbol{a}$$

53. Rule 2: If $f(\boldsymbol{z}) = \boldsymbol{z}'\boldsymbol{z} = \sum z_i^2$, then

$$\partial(\boldsymbol{z}'\boldsymbol{z})/\partial\boldsymbol{z} = 2\boldsymbol{z}$$

54. Rule 3: If $f(\boldsymbol{z}) = \boldsymbol{z}'\boldsymbol{B}\boldsymbol{z}$ for a k-square matrix $\boldsymbol{B}$, then

$$\partial(\boldsymbol{z}'\boldsymbol{B}\boldsymbol{z})/\partial\boldsymbol{z} \;=\; (\boldsymbol{B}+\boldsymbol{B}')\boldsymbol{z}$$

55. Rule 4: If $\boldsymbol{B}$ is a *symmetric* k-square matrix, then by Rule 3:

$$\partial(\boldsymbol{z}'\boldsymbol{B}\boldsymbol{z})/\partial\boldsymbol{z} \;=\; 2\boldsymbol{B}\boldsymbol{z}$$

1.5 Means, Variances, and Covariances

56. Let $y_1, y_2, \ldots, y_k$ be k random variables whose means are given by $E(y_i) = \mu_i$ for $i = 1, 2, \ldots, k$. In vector form we write

$$E(\boldsymbol{y}) \;=\; E\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_k \end{bmatrix} \;=\; \boldsymbol{\mu} \;=\; \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_k \end{bmatrix}$$

That is, the expectation of a vector is the vector of expectations.

57. The same applies more generally to matrices. The expectation of a matrix of random variables is a matrix containing the expected values of the individual random variables.

58. We can use matrix notation to describe the variances and covariances of the elements of a vector of random variables. Suppose the variances of the y_i are given by

$$\text{var}(y_i) \;=\; E[(y_i - \mu_i)^2] \;=\; \sigma_i^2 \qquad \text{for } i = 1, 2, \ldots, k$$

and the covariances are given by

$$\text{cov}(y_i, y_j) \;=\; E[(y_i - \mu_i)(y_j - \mu_j)] \;=\; \sigma_{ij} \qquad \text{for } i, j = 1, 2, \ldots, k \text{ and } i \neq j.$$

59. The **variance-covariance matrix** $\boldsymbol{\Sigma}$ is the symmetric matrix which contains the variances (σ_i^2) on the main diagonal and the covariances (σ_{ij}) as the off-diagonal elements. That is,

$$\boldsymbol{\Sigma} \;=\; E[(\boldsymbol{y}-\boldsymbol{\mu})(\boldsymbol{y}-\boldsymbol{\mu})'] \;=\; \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1k} \\ & \sigma_2^2 & \cdots & \sigma_{2k} \\ \text{sym}- & & \cdots & \cdots \\ \text{metric} & & & \sigma_k^2 \end{bmatrix}$$

$\boldsymbol{\Sigma}$ is also referred to as cov$(\boldsymbol{y})$ or var$(\boldsymbol{y})$.

60. If the vector of random variables $\boldsymbol{y} = (y_1, y_2, \ldots, y_k)'$ are jointly normally distributed with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, we write $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

61. For the special case where the random variables are uncorrelated ($\sigma_{ij} \equiv 0$) and have equal variances ($\sigma_i^2 \equiv \sigma^2$), we write $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \sigma^2\boldsymbol{I}_k)$.

62. Rule E1: If a vector $\boldsymbol{y}$ is a vector of k random variables with $E(\boldsymbol{y}) = \boldsymbol{\mu}$, then $E(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{\mu}$ where $\boldsymbol{A}$ is any $n \times k$ matrix of constants.

63. Rule E2: Rule E1 can be generalized to the case of a $k \times p$ matrix $\boldsymbol{X}$ of random variables x_{ij}. That is, if $E(\boldsymbol{X}) = \boldsymbol{M}$, then

$$E(\boldsymbol{AX}) \;=\; \boldsymbol{A}E(\boldsymbol{X}) \;=\; \boldsymbol{AM}$$

where $\boldsymbol{A}$ is any $n \times k$ matrix of constants.

64. Rule E3: Let $\boldsymbol{y}$ be a vector of random variables with $E(\boldsymbol{y}) = \boldsymbol{0}$ and variance-covariance matrix $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_k$. Then for a real symmetric matrix $\boldsymbol{B}$

$$E(\boldsymbol{y}'\boldsymbol{B}\boldsymbol{y}) \;=\; \sigma^2 \text{trace}(\boldsymbol{B}).$$

65. Rule E4: Let $\boldsymbol{y}$ be a vector of random variables with $E(\boldsymbol{y}) = \boldsymbol{\mu}$ and $\text{cov}(\boldsymbol{y}) = \boldsymbol{\Sigma}$. If $\boldsymbol{A}$ is an $n \times k$ matrix of constants, and $\boldsymbol{z} = \boldsymbol{Ay}$, then

$$\text{cov}(\boldsymbol{z}) \;=\; \text{cov}(\boldsymbol{Ay}) \;=\; \boldsymbol{A\Sigma A}'$$

66. Rule E5: A special case of Rule E4 is the situation for finding $\text{var}(\boldsymbol{a}'\boldsymbol{y}) = \text{var}(\sum a_i y_i) =$ the variance of a linear combination of random variables. The variance is given by the quadratic form

$$\text{var}(\boldsymbol{a}'\boldsymbol{y}) \;=\; \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} \;=\; \sum_{i=1}^{k} a_i^2 \sigma_i^2 \;+\; 2\sum\sum_{i<j} a_i a_j \sigma_{ij}.$$

where $\boldsymbol{y}$ be a vector of random variables with variance-covariance matrix $\text{cov}(\boldsymbol{y}) = \boldsymbol{\Sigma}$, and $\boldsymbol{a}$ is a vector of k constants.

1.6 Least Squares

67. Assume the response of interest y can be approximated by a low-order polynomial $f(x_1, x_2, \ldots, x_k)$ where $x_1, x_2, \ldots, x_k$ are k independent variables. Suppose that n experimental runs are taken for various combinations of the x's which were determined by the experimenter. The data is written in the form

$$\begin{array}{lcccccc} y_1 & & x_{11} & x_{21} & x_{31} & \cdots & x_{k1} \\ y_2 & & x_{12} & x_{22} & x_{32} & \cdots & x_{k2} \\ y_3 & & x_{13} & x_{23} & x_{33} & \cdots & x_{k3} \\ \cdots & & \cdot & \cdots & \cdots & \cdots & \cdots \\ y_n & & x_{1n} & x_{2n} & x_{3n} & \cdots & x_{kn} \end{array}$$

where $n > k$. The plan of experimental levels of the x's is called the **experimental design**.

68. If the approximating model assumed the experimenter can be written as

$$y_i \;=\; \beta_0 \;+\; \beta_1 x_{1i} \;+\; \beta_2 x_{2i} \;+ \cdots +\; \beta_k x_{ki} \;+\; \epsilon_i \qquad (i = 1, 2, \ldots, n)$$

where ϵ_i is a random variable. It is assumed that the ϵ_i are independent from run to run and $\epsilon_i \sim (0, \sigma^2)$.

69. In matrix form we can write $\;\boldsymbol{y} \;=\; \boldsymbol{X\beta} \;+\; \boldsymbol{\epsilon}\;$ where

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \cdots \\ \beta_k \end{bmatrix} \quad \text{and} \;\; \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ 1 & x_{13} & x_{23} & \cdots & x_{k3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdots \\ \epsilon_n \end{bmatrix}$$

and $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and $\text{cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n$. This model is referred to as the **general linear model**.

70. The general linear model can be applied to polynomial models of degree higher than one. For example, suppose the assumed model is quadratic in two variables x_1 and x_2. That is, the response for the i^{th} run involving x_{1i} and x_{2i} is given by

$$y_i \;=\; \beta_0 \;+\; \beta_1 x_{1i} \;+\; \beta_2 x_{2i} \;+\; \beta_{11} x_{1i}^2 \;+\; \beta_{22} x_{2i}^2 \;+\; \beta_{12} x_{1i} x_{2i} \;+\; \epsilon_i$$

where $i = 1, 2, \ldots, n$ with $n \geq 6$. For this example

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{11} \\ \beta_{22} \\ \beta_{12} \end{bmatrix} \qquad \text{and } \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{11}^2 & x_{21}^2 & x_{11}x_{21} \\ 1 & x_{12} & x_{22} & x_{12}^2 & x_{22}^2 & x_{12}x_{22} \\ 1 & x_{13} & x_{23} & x_{13}^2 & x_{23}^2 & x_{13}x_{23} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} & x_{1n}^2 & x_{2n}^2 & x_{1n}x_{2n} \end{bmatrix}$$

71. Given the design matrix $\boldsymbol{X}$ and a vector $\boldsymbol{y}$ of responses, the **method of least squares** yields an estimate $\boldsymbol{b}$ of $\boldsymbol{\beta}$ which minimizes L, the sum of squares of the errors (or deviations) of the observed responses from the estimated values:

$$L \;=\; \sum_{i=1}^{n} e_i^2 \;=\; \boldsymbol{e}'\boldsymbol{e} \qquad \text{where } e_i = y_i - \boldsymbol{x_i b}$$

or, equivalently,

$$L \;=\; (\boldsymbol{y} - \boldsymbol{Xb})'(\boldsymbol{y} - \boldsymbol{Xb}) \;=\; \boldsymbol{y}'\boldsymbol{y} \;-\; 2\,\boldsymbol{b}'\boldsymbol{X}'\boldsymbol{y} \;+\; \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}.$$

72. To find the $\boldsymbol{b}$ which minimizes L we first note that $\boldsymbol{X'X}$ is symmetric and use differentiation rules (Rule 1 and Rule 4 on page 13):

$$\partial L / \partial \boldsymbol{b} \;=\; -2\,\boldsymbol{X}'\boldsymbol{y} \;+\; 2\,(\boldsymbol{X'X})\boldsymbol{b}.$$

Setting the partial derivatives to 0 and solving for $\boldsymbol{b}$ yields $(\boldsymbol{X'X})\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y}$. These equations are called the **normal equations**.

73. Assuming $\boldsymbol{X'X}$ is nonsingular, we have the **least squares estimator**

$$\boldsymbol{b} \;=\; (\boldsymbol{X'X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$

74. $E(\boldsymbol{b}) = \boldsymbol{\beta}$. That is, the least squares estimator $\boldsymbol{b} = (\boldsymbol{X'X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ is unbiased, or equivalently, each element in $\boldsymbol{b}$ is unbiased for the parameter it is estimating.

75. In the development of experimental designs for response surface methodology, it is important to investigate the effect of the design on the variance-covariance matrix of $\boldsymbol{b}$:

$$\text{cov}(\boldsymbol{b}) \;=\; E[(\boldsymbol{b} - \boldsymbol{\beta})(\boldsymbol{b} - \boldsymbol{\beta})'] \;=\; \sigma^2(\boldsymbol{X'X})^{-1}.$$

This implies that the variances of the estimators in $\boldsymbol{b}$ are given by the main diagonal elements of $(\boldsymbol{X'X})^{-1}$ multiplied by σ^2, and the covariances between elements of $\boldsymbol{b}$ are the off-diagonal elements of $(\boldsymbol{X'X})^{-1}$ multiplied by σ^2.

1.7 Hypothesis Testing

76. If the additional assumption is made that ϵ_i is normally distributed, that is, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, then the y_i's are also normally distributed as $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}_n)$.

77. This also implies that $\boldsymbol{b} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X'X})^{-1})$.

78. <u>Sums of squares in the regression</u> Let $\boldsymbol{J}_n$ be a $n \times 1$ vector of ones.

    (a) *Total Sum of Squares*:

$$S_{yy} = SS_T = \boldsymbol{y'y} - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = \boldsymbol{y'}\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\boldsymbol{J}_n{}'\right)\boldsymbol{y}$$

    (b) *Regression Sum of Squares*:

$$SS_R = \boldsymbol{b'X'y} - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = \boldsymbol{y'}\left(\boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'} - \frac{1}{n}\boldsymbol{J}_n\boldsymbol{J}_n{}'\right)\boldsymbol{y}$$

    (c) *Error Sum of Squares*:

$$SS_E = \boldsymbol{y'y} - \boldsymbol{b'X'y} = \boldsymbol{y'}\left(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'}\right)\boldsymbol{y}$$

    Note: $SS_T = SS_R + SS_E$.

79. <u>ANOVA Table for significance of the regression</u>:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Regression	SS_R	k	MS_R	MS_R/MS_E
Error	SS_E	$n-k-1$	MS_E	
Total	S_{yy}	$n-1$		

    where $k =$ the number of parameters in the model excluding the intercept β_0.

80. To find the expectation $E(SS_E)$ note that

$$SS_E = \boldsymbol{y'}\left(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'}\right)\boldsymbol{y} = \boldsymbol{y'Py}$$

    where $\boldsymbol{P} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'}$. Thus, SS_E is a quadratic form in the y's. Then

$$\begin{aligned} E(SS_E) &= E(\boldsymbol{y'Py}) = E(\boldsymbol{\epsilon'P\epsilon}) \\ &= \sigma^2 \text{trace}(\boldsymbol{P}) = \sigma^2(n-k-1) \end{aligned}$$

    Thus, $MS_E = \dfrac{SS_E}{n-k-1}$ is unbiased estimator of σ^2. Notationally, we write $\widehat{\sigma}^2 = MS_E$.

81. Test for significance of the regression: We can write $\boldsymbol{\beta} = [\,\beta_0 \,|\, \boldsymbol{\beta}^*\,]'$ where $\boldsymbol{\beta}^*$ is the parameter vector excluding β_0. To test for the significance of the regression, that is, to test

$$\mathrm{H}_o : \boldsymbol{\beta}^* = \mathbf{0} \quad \text{against} \quad \mathrm{H}_a : \text{at least one parameter in } \boldsymbol{\beta}^* \text{ does not equal } 0$$

determine the p-value by comparing F_0 to the $F(p, n-p-1)$ distribution — its reference distribution under H_0.

82. Let β_j ($j \neq 0$) be a parameter in $\boldsymbol{\beta}$. From (63.) we know that $\boldsymbol{b} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X'X})^{-1})$. Thus, if $\beta_j = 0$ then $b_j \sim N(0, \sigma^2 C_{jj})$ where C_{jj} is the diagonal element of $(\boldsymbol{X'X})^{-1}$ corresponding to β_j.

83. Tests on individual regression coefficients: To test for the significance of β_j, that is, to test

$$\mathrm{H}_o : \beta_j = 0 \quad \text{against} \quad \mathrm{H}_a : \beta_j \neq 0,$$

first note that

$$t_0 \;=\; \frac{b_j}{\mathrm{se}(b_j)} \;=\; \frac{b_j}{\sqrt{\widehat{\sigma}^2 C_{ii}}}$$

follows a $t(n-k-1)$ distribution if $\mathrm{H}_o : \beta_j = 0$ is true. Therefore, the p-value of this test is determined by comparing t_0 to the $t(n-p-1)$ distribution — its reference distribution under H_0.