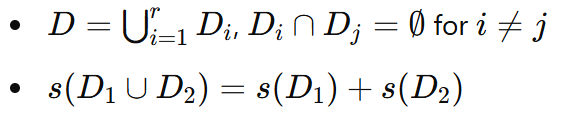
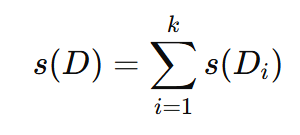
***3.11.1 Conceptual***

3.1. Suppose that s is a statistic and that for any two sets D1 and D2 that form a partition of the set D, s(D1 ∪ D2) = s(D1) + s(D2). Suppose that r > 2 and D1, D2, . . . , Dr is also a partition of D. Argue that s(D) = s(D1)+ s(D2) + · · · + s(Dr).

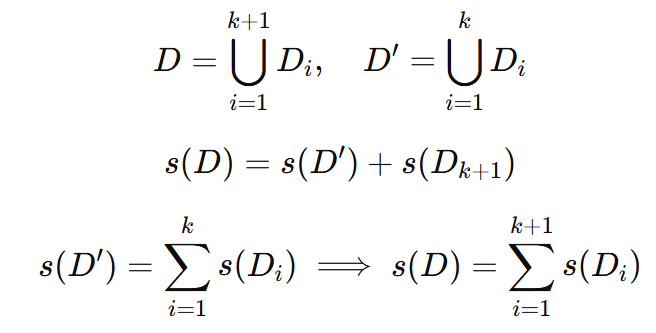
1. Assumptions:

2. Proof by Induction:

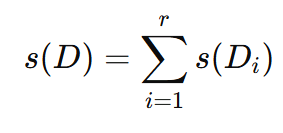
* Base Case:
* Inductive Hypothesis: For r = k:



* Inductive Step: For r = k + 1:

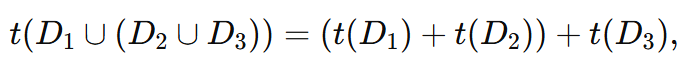


3. Conclusion:

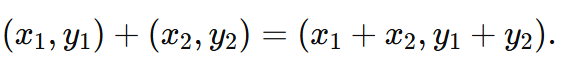


3.2. Show that the statistic t(D) = (A, z) (Eqs.(3.22) and (3.32)) is an associative statistic. Recall that pairs are added component-wise; hence, (x1, y1) + (x2, y2) = (x1 + x2, y1 + y2).

To prove that t(D)=(A, z) is an associative statistic, we need to show that for any three disjoint datasets D1,D2,D3 the following holds:



where t(D)=(A, z) and the addition of pairs is defined as:



1. Expression for t(D1∪(D2∪D3)):

t(D1​∪(D2​∪D3​)) = t(D1​) + t(D2​∪D3​)

Using the property of t for unions:

t(D2​∪D3​) = t(D2​) + t(D3​)

Substituting:

t(D1​∪(D2​∪D3​)) = t(D1​) + (t(D2​) + t(D3​))

**2. Expression for (t(D1) + t(D2)) + t(D3):** Using the pairwise addition:

(t(D1​) + t(D2​)) + t(D3​) = (t(D1​) + t(D2​)) + t(D3​)

**3. Associativity of Addition for Pairs:** Since addition of pairs is component-wise, it is associative:



Therefore:

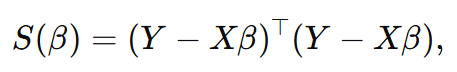
t(D1​∪(D2​∪D3​)) = (t(D1​) + t(D2​)) + t(D3​)

**Conclusion:**

The statistic t(D) = (A, z) is associative because pairwise addition is associative.

3.3. (Requires multivariable calculus.) Show that the least squares estimator is the solution to the normal equations (3.28). Specifically, minimize the objective function S(β) (Eq.(3.25)) with respect to β. Differentiate S(β) with respect to β. Set the vector of partial derivatives equal to 0(q×1) and solve for β. The solution is the least squares estimator.

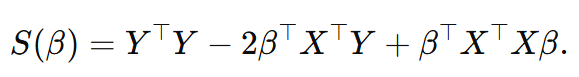
1. **Objective Function S(β):**



where:

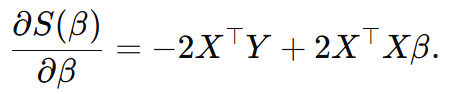
* Y is the n×1 vector of observed values,
* X is the n× q matrix of predictors,
* β is the q×1 vector of coefficients.

2. **Expand S(β):**

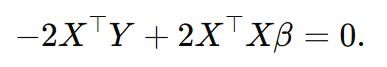
****

**3. Differentiate with respect to β:**

The derivative of S(β) with respect to β (a q×1 vector) is:



**4. Set the Gradient to Zero:**

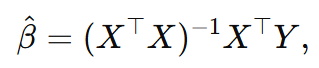
To minimize S(β), set the gradient equal to the zero vector:

Simplify:

5. **Solve for β:**

assuming X^⊤ X is invertible.

**Conclusion:**

The least squares estimator β^​ is:

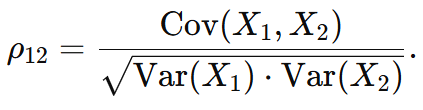
which solves the normal equations:

3.6. Suppose that X2 = aX1 + b, where X1 is a random variable with finite mean and variance and a != 0 and b are real numbers. Show that the population correlation coefficient ρ12 = 1.

To show that the population correlation coefficient ρ12​=1, we use the definition of the correlation coefficient and the relationship X2=aX1+b, where a≠0 and b are real numbers.

**1. Definition of Correlation Coefficient:**

The population correlation coefficient between X1​ and X2​ is defined as:



**2. Relationship Between X2​ and X1​:**

Given X2=aX1+b, calculate the variance and covariance.

**Conclusion:**

**If X2=aX1+b, the population correlation coefficient ρ12=1.**

**a. Covariance:**

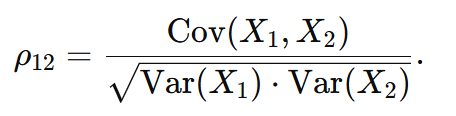
Since b is a constant, Cov(X1,b)=0, and Cov(X1,aX1)=aVar(X1​):

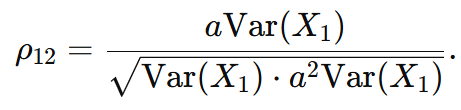
b. Variance of X2​:

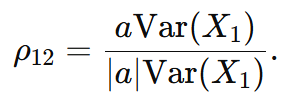
Since b is a constant:

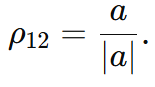


**3. Substitute into ρ12​:**

Using the definition of ρ12​:

Substitute Cov(X1,X2)=aVar(X1) and Var(X2)=a2Var(X1​):

Simplify:

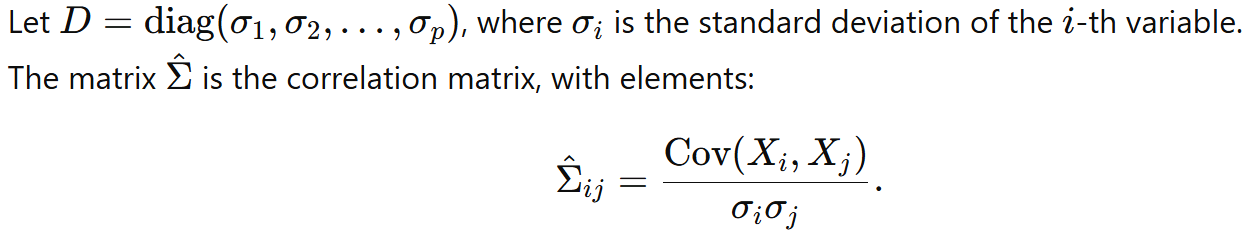
Since a≠0:

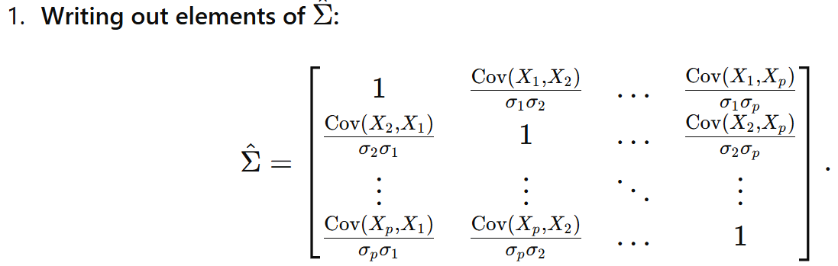
The term a/∣a∣ equals 1 when a>0 and −1 when a<0. By definition, the population correlation coefficient takes the absolute value:

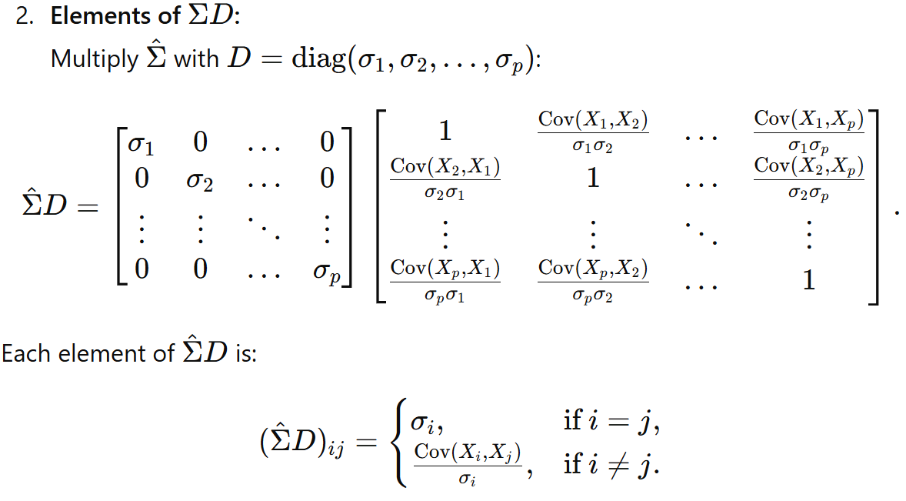


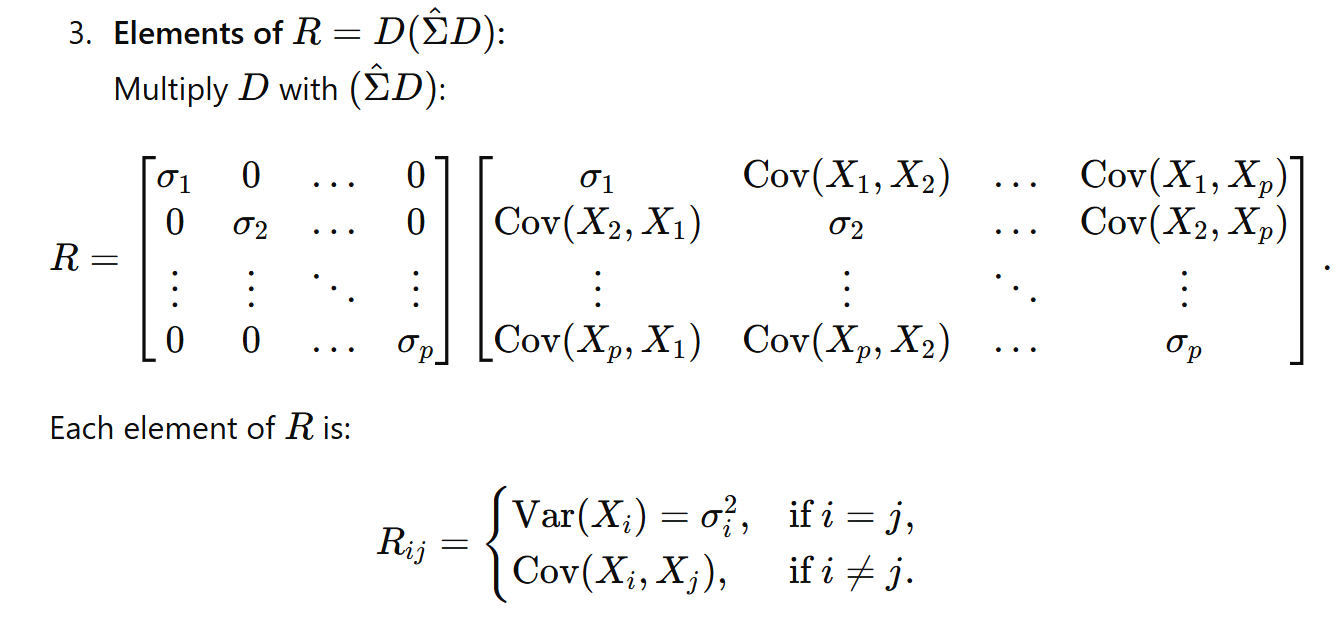
3.7. Show that

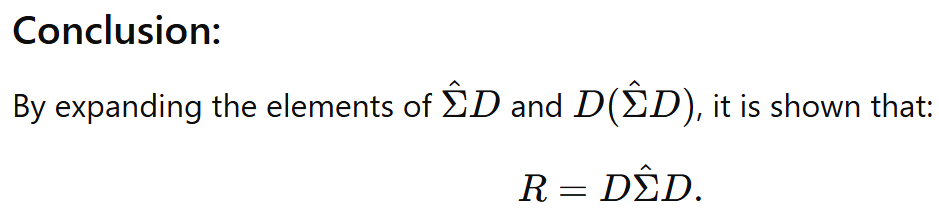
R=DΣ^D

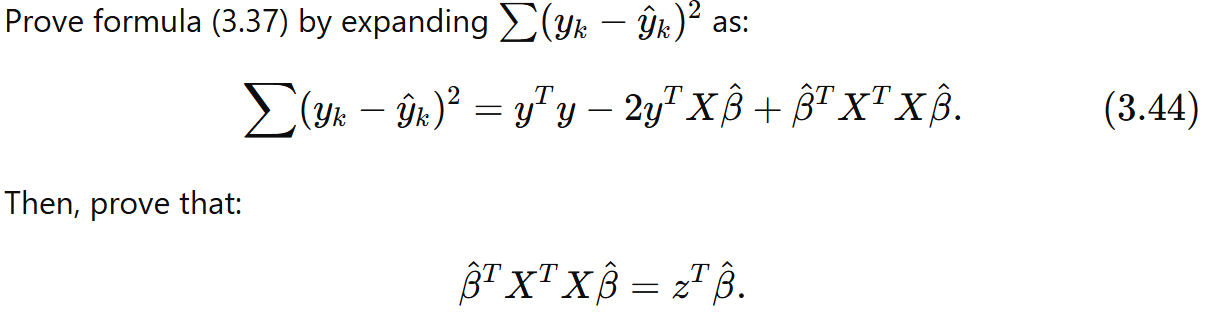
by writing out some of the elements of Σ^ and then the elements of the product D(Σ^D).









3.8.

