Ec. 2120, Fall 2022 Due: September 16, 2022 at 5pm

PROBLEM SET 1

There are a number of problems below, but you are to write up and hand in solutions to only some of them; the rest may be useful for additional study, etc.

Please hand in solutions for problems 1, 2, 3, 5, 9, 11, 12, 13.

It is fine to discuss the questions with others and form groups of no larger than 4. You can submit one homework per group. It is preferable that the groups have the same composition till the end of the semester.

1. Let $\mathcal{R}$ denote the real numbers and let H be a (real) vector space. A *semi-inner product* on H is a function $\langle \cdot, \cdot \rangle$ from $H \times H$ into $\mathcal{R}$ such that

(i) $\langle cf + g, h \rangle = c\langle f, h \rangle + \langle g, h \rangle$ for all c in $\mathcal{R}$ and f, g in H.

(ii) $\langle f, g \rangle = \langle g, f \rangle$ for all f, g in H.

(iii) $\langle \cdot, \cdot \rangle$ is nonnegative definite: that is, $\langle f, f \rangle \geq 0$ for all f in H.

A semi-inner product is a (i) bilinear, (ii) symmetric, (iii) nonnegative definite form on $H \times H$. A semi-inner product is an *inner product* if $\langle f, f \rangle = 0$ implies $f = 0$. A reference for the projection theorem in a linear vector space with an inner product is D. Luenberger, *Optimization By Vector Space Methods*, 1969, John Wiley & Sons, chapter 3.

(a) Suppose that H is a linear space of random variables with $E(X^2) < \infty$ for all random variables X in H. Define, for X and Y in H,

$$\mathrm{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(\tilde{X}\tilde{Y}),$$

with $\tilde{X} = X - E(X)$, $\tilde{Y} = Y - E(Y)$. Show that

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

(b) Consider using $\langle X, Y \rangle = \mathrm{Cov}(X, Y)$. Does this define a semi-inner product? Does it define an inner product? Explain.

2. As in question 1, let H be a vector space with an inner product $\langle \cdot, \cdot \rangle$. In this problem, we will not appeal to the projection theorem, but instead provide direct derivations for some of its implications.

(a) Given f and g in H, consider the problem

$$\min_{c \in \mathcal{R}} ||f - cg||^2. \tag{1}$$

Suppose that the first-order condition

$$\frac{\partial}{\partial c} ||f - cg||^2 = 0$$

is satisfied by $c = \beta$. Define $\hat{f} = \beta g$. Show that

$$\langle f - \hat{f}, g \rangle = 0.$$

Show that

$$||f - cg||^2 = ||f - \hat{f}||^2 + (c - \beta)^2 ||g||^2$$

and conclude that $c = \beta$ is a solution to (1).

(b) Let H_1 be a (linear) subspace of H. (H_1 need not have finite dimension). Given $f \in H$, consider the problem

$$\min_{h \in H_1} ||f - h||^2. \tag{2}$$

Suppose that $\hat{f} \in H_1$ satisfies

$$\langle f - \hat{f}, h \rangle = 0 \quad \text{for all} \quad h \in H_1.$$

Show that $\hat{f}$ is the unique solution to (2).

(c) As in (b), let H_1 be a subspace of H. Given f and g in H, suppose that $\hat{f}$ and $\hat{g}$ in H_1 satisfy

$$\langle f - \hat{f}, h \rangle = 0 \quad \text{and} \quad \langle g - \hat{g}, h \rangle = 0$$

for all $h \in H_1$. Given scalars $c_1, c_2 \in \mathcal{R}$, consider the problem

$$\min_{h \in H_1} ||c_1 f + c_2 g - h||^2.$$

What is the solution to this problem? Explain your reasoning.

3. The correlation between the random variables X and Y is defined as

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{[\mathrm{Var}(X)\mathrm{Var}(Y)]^{1/2}},$$

(with $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$). Let

$$E^*(Y \,|\, 1, X) = \beta_0 + \beta_1 X$$

denote the (minimum mean-square error) linear predictor of Y given $1, X$. The population R^2 for Y on X is defined by

$$R^2_{\text{pop}} = 1 - \frac{||Y - E^*(Y \,|\, 1, X)||^2}{||Y - E^*(Y \,|\, 1)||^2}.$$

(with inner product $\langle X, Y \rangle = E(XY)$ and norm $||X|| = \langle X, X \rangle^{1/2}$.)

(a) Show that

$$\mathrm{Cor}(X, Y) = \beta_1 \frac{[\mathrm{Var}(X)]^{1/2}}{[\mathrm{Var}(Y)]^{1/2}}.$$

(b) Show that

$$\mathrm{Cor}(X,Y)^2 = R^2_{\mathrm{pop}}.$$

(c) Show that

$$-1 \le \mathrm{Cor}(X,Y) \le 1.$$

4. Consider the following $n \times 1$ data matrices:

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

The sample covariance between x and y is defined by

$$\mathrm{Cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}),$$

with

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

The sample correlation between x and y is defined by

$$\mathrm{Cor}(x,y) = \frac{\mathrm{Cov}(x,y)}{[\mathrm{Var}(x)\mathrm{Var}(y)]^{1/2}},$$

(with $\mathrm{Var}(x) = \mathrm{Cov}(x,x)$). Provide properties for the sample correlation that correspond to the properties for the population correlation in (a), (b), (c) of question 3. Briefly indicate how the demonstration of these properties can follow the arguments used in your answer to question 3.

5. Let

$$E^*(Y \mid 1, X_1, \ldots, X_K) = \beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K$$

denote the population linear predictor of Y given $1, X_1, \ldots, X_K$. We can write this as

$$E^*(Y \mid X) = X'\beta$$

with

$$X = \begin{pmatrix} 1 \\ X_1 \\ \vdots \\ X_K \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}.$$

Define the prediction error

$$U = Y - X'\beta.$$

Show that

$$\text{Var}(Y) = \text{Var}(X'\beta) + \text{Var}(U).$$

6. Consider the data matrices

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad x = \begin{pmatrix} 1 & x_{11} & \dots & x_{1K} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nK} \end{pmatrix},$$

where y is $n \times 1$ and x is $n \times (K+1)$. Let b denote the $(K+1) \times 1$ matrix of least squares coefficients, and define the fitted values

$$\hat{y} = xb$$

and the residuals

$$e = y - \hat{y}.$$

Show that

$$\text{Var}(y) = \text{Var}(\hat{y}) + \text{Var}(e).$$

(Here Var refers to sample variance.)

7. Let $E(Y \mid Z)$ denote the conditional expectation of Y given Z, where Y and Z are random variables with a joint distribution. Let

$$U = Y - E(Y \mid Z)$$

denote the prediction error. The conditional variance of Y given Z is defined as

$$\text{Var}(Y \mid Z) = E[(Y - E(Y \mid Z))^2 \mid Z] = E(U^2 \mid Z).$$

(a) Show that

$$\text{Var}(U) = E(U^2) = E[\text{Var}(Y \mid Z)].$$

(b) Show that

$$\text{Var}(Y) = \text{Var}[E(Y \mid Z)] + E[\text{Var}(Y \mid Z)].$$

This formula is known as the (population) analysis of variance decomposition.

8. Suppose that Z_1 and Z_2 are binary random variables: Z_1 takes on only the values 0 and 1, and Z_2 takes on only the values 0 and 1. Consider the (population) linear predictor of Y given $1, Z_1, Z_2, Z_1 \cdot Z_2$:

$$E^*(Y \mid 1, Z_1, Z_2, Z_1 \cdot Z_2) = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_1 \cdot Z_2.$$

(a) Does

$$E(Y \mid Z_1, Z_2) = E^*(Y \mid 1, Z_1, Z_2, Z_1 \cdot Z_2)?$$

Explain.

(b) Suppose that data (y_i, z_{i1}, z_{i2}) are available from a random sample of $i = 1, \ldots, n$ individuals. The following four sample means have been tabulated:

$$\bar{y}_{00}, \quad \bar{y}_{01}, \quad \bar{y}_{10}, \quad \bar{y}_{11},$$

where

$$\bar{y}_{lm} = \frac{\sum_{i=1}^{n} y_i 1(z_{i1} = l, z_{i2} = m)}{\sum_{i=1}^{n} 1(z_{i1} = l, z_{i2} = m)} \qquad (l, m = 0, 1).$$

($1(B)$ is the indicator function that equals 1 if the event B occurs and equals 0 otherwise.) Use these means to provide an estimate of β_3.

9. Let Y = log earnings, Z_1 = years of education, Z_2 = years of work experience. A random sample of n individuals has provided the data (y_i, z_{i1}, z_{i2}) for $i = 1, \ldots, n$. This data has been used to obtain the following least squares fit:

$$\hat{y}_i = b_0 + b_1 z_{i1} + b_2 z_{i2} + b_3 z_{i1}^2 + b_4 z_{i1} \cdot z_{i2} + b_5 z_{i2}^2.$$

(a) We are interested in the following partial (predictive) effect of education on log earnings:

$$\theta = E(Y \mid Z_1 = 16, Z_2 = 20) - E(Y \mid Z_1 = 12, Z_2 = 20).$$

Explain how to use the least-squares estimates (the b's) to obtain an estimate of θ.

(b) Now consider the average partial effect:

$$\gamma = E[E(Y \mid Z_1 = 16, Z_2) - E(Y \mid Z_1 = 12, Z_2)],$$

where the outer expectation is over the marginal distribution of Z_2. Explain how to use the least-squares estimates and the z data to obtain an estimate of γ.

10. Consider the following regression function:

$$E(Y \mid Z_1, Z_2) = \theta Z_1 + g(Z_2),$$

where the function $g(\cdot)$ is an unknown function, which is not restricted. This regression function does impose the restriction that Z_1 enters linearly and there is no interaction between Z_1 and Z_2. The random variable Z_2 is discrete and takes on only the values $\delta_1, \delta_2, \delta_3$. A random sample provides the data (y_i, z_{i1}, z_{i2}) for $i = 1, \ldots, n$. Suggest an estimator for θ, based on a least-squares fit of y on $x_1, \ldots, x_K$, where

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad x_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{pmatrix} \qquad (j = 1, \ldots, K).$$

Be explicit on how the x_j are constructed from the data on (z_{i1}, z_{i2}) for $i = 1, \ldots, n$.

11. The partial covariance between the random variables Y and X_K given $X_1, \ldots, X_{K-1}$ is defined as

$$\mathrm{Cov}^*(Y, X_K \,|\, X_1, \ldots, X_{K-1}) = E(\tilde{Y}\tilde{X}_K),$$

where

$$\tilde{Y} = Y - E^*(Y \,|\, 1, X_1, \ldots, X_{K-1}), \quad \tilde{X}_K = X_K - E^*(X_K \,|\, 1, X_1, \ldots, X_{K-1}).$$

(If the list of conditioning variables is empty, this reduces to $\mathrm{Cov}^*(Y, X_K) = \mathrm{Cov}(Y, X_K)$, because $E^*(Y \,|\, 1) = E(Y)$ and $E^*(X_K \,|\, 1) = E(X_K)$.) The partial variance of Y is

$$\mathrm{Var}^*(Y \,|\, X_1, \ldots, X_{K-1}) = \mathrm{Cov}^*(Y, Y \,|\, X_1, \ldots, X_{K-1}) = E(\tilde{Y}^2).$$

The partial correlation between Y and X_K is

$$\mathrm{Cor}^*(Y, X_K \,|\, X_1, \ldots, X_{K-1}) = \frac{E(\tilde{Y}\tilde{X}_K)}{[E(\tilde{Y}^2)E(\tilde{X}_K^2)]^{1/2}}.$$

(a) Show that the absolute value of the partial correlation is less than or equal to one.

(b) Consider the linear predictor

$$E^*(Y \,|\, 1, X_1, \ldots, X_K) = \beta_0 + \beta_1 X_1 + \ldots + \beta_K X_K.$$

Use our residual regression results to show that

$$\beta_K = \frac{\mathrm{Cov}^*(Y, X_K \,|\, X_1, \ldots, X_{K-1})}{\mathrm{Var}^*(X_K \,|\, X_1, \ldots, X_{K-1})}.$$

(c) Suggest definitions for sample partial covariance and for sample partial correlation.

12. The conditional covariance between the random variables Y and Z_1 given Z_2 is defined as

$$\mathrm{Cov}(Y, Z_1 \,|\, Z_2) = E(\tilde{Y}\tilde{Z}_1 \,|\, Z_2),$$

where

$$\tilde{Y} = Y - E(Y \,|\, Z_2), \quad \tilde{Z}_1 = Z_1 - E(Z_1 \,|\, Z_2).$$

(a) Discuss briefly the difference between conditional covariance and partial covariance.

13. Consider the following long and short linear predictors in the population:

$$E^*(Y \,|\, 1, X_1, X_2, X_3) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3,$$
$$E^*(Y \,|\, 1, X_1, X_2) = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2.$$

(a) Provide a formula relating α_2 and β_2.

(b) Suppose that X_3 is uncorrelated with X_1 and with X_2. Does $\alpha_2 = \beta_2$? Explain.

(c) Suppose that

$$\mathrm{Cov}(X_2, X_3) = 0, \quad \mathrm{Cov}(X_1, X_3) \neq 0, \quad \mathrm{Cov}(X_1, X_2) \neq 0.$$

Does $\alpha_2 = \beta_2$? Explain.

14. We observe $W_i = (Y_i, S_i, Z_i)$ for $i = 1, \ldots, n$ individuals. $Y_i = \log(\text{earnings})$. S_i = years of schooling. $Z_i = 1$ if the individual is given encouragement, perhaps through a subsidy, to obtain additional schooling; $Z_i = 0$ if the encouragement is not offered. The earnings are measured after the individuals have completed their schooling. We shall assume that $D_i = (W_i, A_i)$ is independent and identically distributed according to some unknown distribution; the latent variable A_i is not observed, and is meant to capture unmeasured aspects of initial ability or pre-school human capital, which may be correlated with S_i. The latent variable model has the following key assumptions:

$$E(Y_i \,|\, S_i, Z_i, A_i) = \gamma_1 + \gamma_2 S_i + \gamma_3 A_i,$$

so that the encouragement does not have a direct effect on earnings, conditional on schooling and initial ability;

$$E^*(A_i \,|\, 1, Z_i) = E(A_i),$$

so that the encouragement is uncorrelated with A_i (which would be very plausible if the encouragement were randomly assigned). In addition, we shall assume that the encouragement does affect schooling, in that

$$E^*(S_i \,|\, 1, Z_i) = \lambda_1 + \lambda_2 Z_i$$

with $\lambda_2 > 0$.

Let the notation for the linear predictor of Y_i given a constant and Z_i be

$$E^*(Y_i \,|\, 1, Z_i) = \beta_1 + \beta_2 Z_i.$$

Show that

$$\gamma_2 = \beta_2 / \lambda_2.$$

15. Consider the following model for measurement error in cross-section data:

$$E^*(Y_i \,|\, 1, Z_i^*, Z_i) = \gamma_0 + \gamma_1 Z_i^*$$
$$E^*(Z_i \,|1,\, Z_i^*) = Z_i^*,$$

where Z_i is a noisy measurement on the true value Z_i^*. The population model is expressed in terms of the vector of random variables

$$D_i = (Y_i, Z_i, Z_i^*).$$

Assume that the D_i are independent and identically distributed (i.i.d.) according to some unknown distribution. We have observations on

$$W_i = (Y_i, Z_i)$$

for $i = 1, \ldots, n$. Data on Z_i^* are not available. Assume that $\gamma_1 > 0$.

(a) Work out the covariance matrix for (Y_i, Z_i) as a function of γ_1, $\mathrm{Var}(Z_i^*)$, and some additional parameters that you will need to define.

(b) Consider the linear predictor

$$E^*(Y_i \,|\, 1, Z_i) = \beta_0 + \beta_1 Z_i.$$

Show that

$$0 < \beta_1 \leq \gamma_1.$$

(Recall that we are assuming $\gamma_1 > 0$.)

(c) Consider the reverse linear predictor

$$E^*(Z_i \,|\, 1, Y_i) = \alpha_0 + \alpha_1 Y_i.$$

Let $\lambda_1 = 1/\alpha_1$ and show that

$$\lambda_1 \geq \gamma_1.$$