

## STA 431s17 Assignment One<sup>1</sup>

1. In the following,  $X$  and  $Y$  are scalar random variables, while  $a$  and  $b$  are fixed constants. For each pair of statements below, one is true and one is false (that is, not true in general). State which one is true, and prove it. Zero marks if you prove both statements are true, even if one of the proofs is correct. Please use only expected value signs in your answers, not integrals or summation.
  - (a)  $Var(aX) = aVar(X)$ , or  $Var(aX) = a^2Var(X)$ .
  - (b)  $Var(aX + b) = a^2Var(X) + b^2$ , or  $Var(aX + b) = a^2Var(X)$ .
  - (c)  $Var(a) = 0$ , or  $Var(a) = a^2$ .
  - (d)  $Cov(aX, bY) = abCov(X, Y)$ , or  $Cov(aX, bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$ .
  - (e)  $Cov(X + a, Y + b) = Cov(X, Y) + ab$ , or  $Cov(X + a, Y + b) = Cov(X, Y)$ .
  - (f)  $Var(aX + bY) = a^2Var(X) + b^2Var(Y)$ , or  $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$ .
2. Let  $X_1, \dots, X_n$  be random variables, let  $a_1, \dots, a_n$  be constants, and let  $Y = \sum_{i=1}^n a_i X_i$ . Derive a general formula for  $Var(Y)$ . Show your work; it's a lot easier with the centering rule; see Question 8. Now give the useful special case that applies when  $X_1, \dots, X_n$  are independent.
3. Let  $X_1, \dots, X_n$  be independent and identically distributed random variables (the standard model of a random sample with replacement). Denoting  $E(X_i)$  by  $\mu$  and  $Var(X_i)$  by  $\sigma^2$ ,
  - (a) Show  $E(\bar{X}) = \mu$ ; that is, the sample mean is unbiased for  $\mu$ .
  - (b) Find  $Var(\bar{X})$ .
  - (c) Letting  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ , show that  $E(S^2) = \sigma^2$ . That is, the sample variance is an unbiased estimator of the population variance. Consider adding and subtracting  $\mu$ .
4. In this regression model, the explanatory variables are random. Independently for  $i = 1, \dots, n$ , let  $Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \epsilon_i$ , where  $E(X_{i,1}) = \mu_1$ ,  $E(X_{i,2}) = \mu_2$ ,  $E(\epsilon_i) = 0$ ,  $Var(\epsilon_i) = \sigma^2$ ,  $\epsilon_i$  is independent of both  $X_{i,1}$  and  $X_{i,2}$ , and

$$cov \begin{pmatrix} X_{i,1} \\ X_{i,2} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{pmatrix}$$

- (a) What is  $Var(Y_i)$ ? You may be able to just write down the answer.
- (b) What is  $Cov(X_{i,1}, Y_i)$ ? Show your work. The centering rule definitely helps; see Question 8.
- (c) What is  $Cov(X_{i,2}, Y_i)$ ?

5. Let

$$\mathbf{x}_i = \begin{pmatrix} x_{i,1} \\ \vdots \\ x_{i,p} \end{pmatrix} \text{ and } \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{pmatrix}.$$

Let the  $p \times p$  matrix  $\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$ . Give a *scalar* formula for element (2, 3) of  $\widehat{\Sigma}$ . If you get stuck, an example with  $p = 3$  should help.

6. Let the  $p \times 1$  random vector  $\mathbf{X}$  have mean  $\boldsymbol{\mu}$  and variance-covariance matrix  $\boldsymbol{\Sigma}$ , let  $\mathbf{A}$  be an  $r \times p$  matrix of constants, and let  $\mathbf{c}$  be an  $r \times 1$  vector of constants. Starting from the definition of a covariance matrix on the formula sheet, find  $cov(\mathbf{A}\mathbf{X} + \mathbf{c})$ . Show your work.

---

<sup>1</sup>This assignment was prepared by [Jerry Brunner](#), Department of Statistical Sciences, University of Toronto. It is licensed under a [Creative Commons Attribution - ShareAlike 3.0 Unported License](#). Use any part of it as you like and share the result freely. The L<sup>A</sup>T<sub>E</sub>X source code is available from the course website: <http://www.utstat.toronto.edu/~brunner/oldclass/431s17>

7. Let  $\mathbf{X}$  and  $\mathbf{Y}$  be random vectors, and let  $\mathbf{A}$  and  $\mathbf{B}$  be matrices of constants. Starting from the definition on the formula sheet, find  $cov(\mathbf{AX}, \mathbf{BY})$ . Show your work. Of course we are assuming that the matrices are the right size.
8. This question takes you through the useful *centering rule*; see the lecture slides or Appendix A of the text. Denote the centered version of the random vector  $\mathbf{X}$  by  $\overset{c}{\mathbf{X}} = \mathbf{X} - \boldsymbol{\mu}_x$ . Most of the following are very quick.
- Show  $E(\overset{c}{\mathbf{X}}) = \mathbf{0}$ .
  - Show  $cov(\overset{c}{\mathbf{X}}) = cov(\mathbf{X})$ .
  - Let  $\mathbf{L} = \mathbf{A}_1\mathbf{X}_1 + \cdots + \mathbf{A}_m\mathbf{X}_m + \mathbf{b}$ , where the  $\mathbf{A}_j$  are matrices of constants, and  $\mathbf{b}$  is a vector of constants. Show  $\overset{c}{\mathbf{L}} = \mathbf{A}_1\overset{c}{\mathbf{X}}_1 + \cdots + \mathbf{A}_m\overset{c}{\mathbf{X}}_m$ .
  - Show  $cov(\mathbf{L}) = E(\overset{c}{\mathbf{L}}\overset{c}{\mathbf{L}}^\top)$ .
  - Show  $cov(\mathbf{L}_1, \mathbf{L}_2) = E(\overset{c}{\mathbf{L}}_1\overset{c}{\mathbf{L}}_2^\top)$ .
9. High School History classes from across Ontario are randomly assigned to either a discovery-oriented or a memory-oriented curriculum in Canadian history. At the end of the year, the students are given a standardized test and the median score of each class is recorded. Please consider a regression model with these variables:
- $X_1$  Equals 1 if the class uses the discovery-oriented curriculum, and equals 0 if the class uses the memory-oriented curriculum.
  - $X_2$  Average parents' education for the classroom.
  - $X_3$  Average family income for the classroom.
  - $X_4$  Number of university History courses taken by the teacher.
  - $X_5$  Teacher's final cumulative university grade point average.
  - $Y$  Class median score on the standardized history test.

The full regression model (as opposed to the reduced models for various null hypotheses) implies

$$E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5.$$

For each question below, please give the null hypothesis in terms of  $\beta$  values, and  $E[Y|X]$  for the reduced model you would use to answer the question. Don't re-number the variables.

- If you control for parents' education and income and for teacher's university background, does curriculum type affect test scores? (And why is it okay to use the word "affect?")
- Controlling for parents' education and income and for curriculum type, is teacher's university background (two variables) related to their students' test performance?
- Controlling for teacher's university background and for curriculum type, are parents' education and family income (considered simultaneously) related to students' test performance?
- Controlling for curriculum type, teacher's university background and parents' education, is parents' income related to students' test performance?
- Here is one final question. Assuming that  $X_1, \dots, X_5$  are random variables (and I hope you agree that they are), would you expect  $X_1$  to be related to the other explanatory variables? Would you expect the other explanatory variables to be related to each other?