STA 431s23 Assignment One[1]

These problems are not to be handed in. They are practice for the Quiz on Friday January 20.

1. Look at the formula sheet and answer the following True of False for general random variables X and Y.
   (a) $Var(X) = 0$.
   (b) $Cov(X, Y) = Cov(Y, X)$.
   (c) $Cov(X, X) = Var(X)$.

2. The discrete random variables X and Y have joint distribution

	$x = 1$	$x = 2$	$x = 3$
$y = 1$	$3/12$	$1/12$	$3/12$
$y = 2$	$1/12$	$3/12$	$1/12$

   (a) What is the marginal distribution of X? List the values with their probabilities.
   (b) What is the marginal distribution of Y? List the values with their probabilities.
   (c) Calculate $E(X)$. Show some work.
   (d) Calculate $E(Y)$. Show some work.
   (e) Let $Z = g(X, Y) = XY$. What is the probability distribution of Z? List the values with their probabilities. Show some work.
   (f) Calculate $E(Z) = E(XY)$. Show your work.
   (g) Do we have $E(XY) = E(X)E(Y)$? Answer Yes or No.
   (h) Using the well-known formula $Cov(X, Y) = E(XY) - E(X)E(Y)$ (to be proved later), what is $Cov(X, Y)$?
   (i) Are X and Y independent? Answer Yes or No and show your work. Note that for discrete random variables, X and Y independent means $P(X = x, Y = y) = P(X = x)\, P(Y = y)$ for all real x and y. So to prove independence, you need to establish it for all x and y, while to prove lack of independence, you only need to find one exception.

3. Let X and Y be *continuous* random variables that are independent, meaning $f_{x,y}(x, y) = f_x(x)\, f_y(y)$ for all real x and y. Using the expression for $E(g(\mathbf{x}))$ on the formula sheet, show $E(XY) = E(X)E(Y)$. Draw an arrow to the place in your answer where you use independence, and write "This is where I use independence." Because X and Y are continuous, you will need to integrate. Does your proof still apply if X and Y are discrete?

[1]This assignment was prepared by Jerry Brunner, Department of Statistical Sciences, University of Toronto. The LaTeX source code is available from the course website: http://www.utstat.toronto.edu/brunner/oldclass/431s23

4. This question clarifies the meaning of $E(a)$ and $Var(a)$ when a is a constant.

   (a) Let X be a discrete random variable with $P(X = a) = 1$ (later we will call this a *degenerate* random variable). Using the definitions on the formula sheet, calculate $E(X)$ and $Var(X)$.

   (b) Let a be a real constant and X be a continuous random variable with density $f(x)$. Let $Y = g(X) = a$. Using the formula for $E(g(X))$ on the formula sheet, calculate $E(Y)$ and $Var(Y)$. This reminds us that the change of variables formula (which is a very big theorem) applies to the case of a constant function.

5. Using the definitions of variance and covariance along with the linear property $E(\sum_{i=1}^n a_iY_i) = \sum_{i=1}^n a_iE(Y_i)$ (no integrals), show the following:

   (a) $Var(Y) = E(Y^2) - (E(Y))^2$

   (b) $Cov(X, Y) = E(XY) - E(X)E(Y)$

   (c) If X and Y are independent, $Cov(X, Y) = 0$. Of course you may use Problem 3.

6. 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) $Cov(aX, bY) = ab\,Cov(X, Y)$, 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) $Var(X + Y) = Var(X) + Var(Y)$, or $Var(X + Y) = Var(X) + Var(Y) + 2\,Cov(X, Y)$.

7. Let $E(X) = \mu_x$, $E(Y) = \mu_y$, and $E(Z) = \mu_z$. Show $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$.

8. Show $Cov(X_1+X_2, Y_1+Y_2) = Cov(X_1, Y_1)+Cov(X_1, Y_2)+Cov(X_2, Y_1)+Cov(X_2, Y_2)$. There is a generalization of this fact on the formula sheet. Can you find it?

9. Let X and Y be random variables, with $E(X) = \mu_x$, $E(Y) = \mu_y$, $Var(X) = \sigma_x^2$, $Var(Y) = \sigma_y^2$, and $Cov(X, Y) = \sigma_{xy}$. Let a be a non-zero constant. Find $Corr(aX, Y)$. Do not forget that a could be negative. Use $sign(a)$ to denote the sign of a. Show your work.

10. Let $y_1, \ldots, y_n$ be numbers (not necessarily random variables), and $\overline{y} = \frac{1}{n}\sum_{i=1}^n y_i$. Show

    (a) $\sum_{i=1}^n (y_i - \overline{y}) = 0$

    (b) $\sum_{i=1}^n (y_i - \overline{y})^2 = \sum_{i=1}^n y_i^2 \; - \; n\overline{y}^2$

11. Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be numbers, with $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$ and $\overline{y} = \frac{1}{n}\sum_{i=1}^n y_i$. Show $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^n x_i y_i \; - \; n\overline{x}\,\overline{y}$.

12. Let $Y_1, \ldots, Y_n$ be *independent* random variables with $E(Y_i) = \mu$ and $Var(Y_i) = \sigma^2$ for $i = 1, \ldots, n$. For this question, please use definitions and familiar properties of expected value, not integrals or sums.

    (a) Find $E(\sum_{i=1}^n Y_i)$. Are you using independence?

    (b) Find $Var\left(\sum_{i=1}^n Y_i\right)$. What earlier questions are you using in connection with independence?

    (c) Using your answer to the last question, find $Var(\overline{Y})$.

    (d) A statistic T is an *unbiased estimator* of a parameter θ if $E(T) = \theta$. Show that $\overline{Y}$ is an unbiased estimator of μ.

    (e) Let $a_1, \ldots, a_n$ be constants and define the linear combination L by $L = \sum_{i=1}^n a_i Y_i$. What condition on the a_i values makes L an unbiased estimator of μ? Show your work.

    (f) Is $\overline{Y}$ a special case of L? If so, what are the a_i values?

    (g) What is $Var(L)$?

13. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables (the standard model of a random sample with replacement). Denote $E(X_i)$ by μ and $Var(X_i)$ by σ^2.

    (a) Letting $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$, show that $E(S^2) = \sigma^2$. That is, the sample variance is an unbiased estimator of the population variance. Consider adding and subtracting μ.

    (b) Find $Cov(X_j, \overline{X})$. Show the calculation.

    (c) Find $Cov(\overline{X}, X_j - \overline{X})$. Show the calculation.

14. The pairs of random variables $(X_1, Y_1), \ldots, (X_n, Y_n)$ are a random sample from a bivariate distribution with $E(X_i) = \mu_x$, $E(Y_i) = \mu_y$, $Var(X_i) = \sigma_x^2$, $Var(Y_i) = \sigma_y^2$, and $Cov(X_i, Y_i) = \sigma_{xy}$. Because $(X_1, Y_1), \ldots, (X_n, Y_n)$ are a random sample, they are independent for $i = 1, \ldots, n$. However, X_i and Y_i are not necessarily independent, because σ_{xy} might not equal zero.

    (a) Find $Cov(\overline{X}, \overline{Y})$. Show the calculation.

    (b) Find $Corr(\overline{X}, \overline{Y})$. Show the calculation.

15. Let

$$\begin{aligned} Y_1 &= \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \epsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_2, \end{aligned}$$

where $E(X_1) = \mu_1$, $E(X_2) = \mu_2$, $Var(X_1) = \phi_{11}$, $Var(X_2) = \phi_{22}$, $Cov(X_1, X_2) = \phi_{12}$, $E(\epsilon_1) = E(\epsilon_2) = 0$, $Var(\epsilon_1) = \sigma_{11}$, $Var(\epsilon_2) = \sigma_{22}$, $Cov(\epsilon_1, \epsilon_2) = \sigma_{12}$, and also, X_1 and X_2 are independent of ϵ_1 and ϵ_2. Calculate $Cov(Y_1, Y_2)$. Show your work. My answer is $\alpha_1\beta_1\phi_{11} + \alpha_2\beta_2\phi_{22} + (\alpha_1\beta_2 + \alpha_2\beta_1)\phi_{12} + \sigma_{12}$.

16. 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

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

For each question below, please give the null hypothesis in terms of β values. The terms Controlling, Correcting, Holding constant and Adjusting all mean the same thing.

(a) 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?")

(b) Correcting for parents' education and income and for curriculum type, is teacher's university background (two variables) related to their students' test performance?

(c) Holding teacher's university background and curriculum type constant, are parents' education and family income (considered simultaneously) related to students' test performance?

(d) Adjusting for curriculum type, teacher's university background and parents' education, is parents' income related to students' test performance?

(e) Here is one final question. Assuming that $X_1, \ldots, X_5$ are random variables (and I hope you agree that they are), would you expect X_1 ro be related to the other explanatory variables? Would you expect the other explanatory variables to be related to each other?