STA 302f16 Assignment One[1]

Please do these review questions in preparation for Quiz One; they are not to be handed in. This material will be on the final exam only indirectly. The following formulas will be supplied with Quiz One. You may use them without proof.

$E(x) = \sum_x x\, p_{_x}(x)$	$E(x) = \int_{-\infty}^\infty x f_{_x}(x)\, dx$
$E(g(x)) = \sum_x g(x)\, p_{_x}(x)$	$E(g(\mathbf{x})) = \sum_{x_1} \cdots \sum_{x_p} g(x_1, \ldots, x_p)\, p_{\mathbf{x}}(x_1, \ldots, x_p)$
$E(g(x)) = \int_{-\infty}^\infty g(x)\, f_{_x}(x)\, dx$	$E(g(\mathbf{x})) = \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty g(x_1, \ldots, x_p)\, f_{\mathbf{x}}(x_1, \ldots, x_p)\, dx_1 \ldots dx_p$
$E(\sum_{i=1}^n a_i x_i) = \sum_{i=1}^n a_i E(x_i)$	$Var(x) = E\left((x-\mu_{_x})^2 \right)$
$Cov(x,y) = E\left((x-\mu_{_x})(y-\mu_{_y}) \right)$	$Corr(x,y) = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}$

The first three problems are very elementary, but they may help to clarify some basic concepts. Please recall the following. Suppose the discrete random variable $y = g(x)$. To find the probability distribution of y, list the possible values of y and then add up the x probabilities corresponding to each value.

1. The random variable x is uniformly distributed on the integers $\{-3, -2, -1, 0, 1, 2, 3\}$, meaning $P(x = -1) = P(x = -2) = \cdots = P(x = 3) = \frac{1}{7}$. Let $y = x^2$.
   (a) What is $E(x)$? The answer is a number. Show your work.
   (b) Calculate the variance of x. The answer is a number. Show your work.
   (c) What is $P(y = -1)$?
   (d) What is $P(y = 9)$?
   (e) What is the probability distribution of y? Give the y values with their probabilities.
   (f) What is $E(y)$? The answer is a number. Did you already do this question?

2. The discrete random variables x and y have joint distribution

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

   (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.

[1] Copyright information is at the end of the last page.

(c) Are x and y independent? Answer Yes or No and show some work.

(d) Calculate $E(x)$. Show your work.

(e) Denote a "centered" version of x by $x_c = x - E(x) = x - \mu_x$.

   i. What is the probability distribution of x_c? Give the values with their probabilities.

   ii. What is $E(x_c)$? Show your work.

   iii. What is the probability distribution of x_c^2? Give the values with their probabilities.

   iv. What is $E(x_c^2)$? Show your work.

(f) What is $Var(x)$? If you have been paying attention, you don't have to show any work.

(g) Calculate $E(y)$. Show your work.

(h) Calculate $Var(y)$. Show your work. You may use Question 6a if you wish.

(i) Calculate $Cov(x, y)$. Show your work. You may use Question 6b if you wish.

(j) Let $Z_1 = g_1(x, y) = x + y$. What is the probability distribution of Z_1? Show some work.

(k) Calculate $E(Z_1)$. Show your work.

(l) Do we have $E(x + y) = E(x) + E(y)$? Answer yes or No. Note that the answer *does not require independence*.

(m) Let $Z_2 = g_2(x, y) = xy$. What is the probability distribution of Z_2? List the values with their probabilities. Show some work.

(n) Calculate $E(Z_2)$. Show your work.

(o) Do we have $E(xy) = E(x)E(y)$? Answer yes or No. The connection to independence is established in Question 5.

3. Here is another joint distribution. The point of this question is that you can have zero covariance without independence.

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

   (a) Calculate $Cov(x, y)$. Show your work. You may use Question 6b if you wish.

   (b) Are x and y independent? Answer Yes or No and show some work.

4. Let x be a discrete random variable and let a be a constant. Using the expression for $E(g(x))$ at the beginning of this assignment, show $E(a) = a$. Is the result still true if x is continuous?

5. Let x_1 and x_2 be continuous random variables that are *independent*. Using the expression for $E(g(\mathbf{x}))$ at the beginning of this assignment, show $E(x_1 x_2) = E(x_1)E(x_2)$. Draw an arrow to the place in your answer where you use independence, and write "This is where I use independence." Because x_1 and x_2 are continuous, you will need to integrate. Does your proof still apply if x_1 and x_2 are discrete?

6. Using the definitions of variance covariance along with the linear property $E(\sum_{i=1}^n a_i y_i) = \sum_{i=1}^n a_i E(y_i)$ (no integrals), show the following:

   (a) $Var(y) = E(y^2) - \mu_y^2$

   (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 5. Important: Does this contradict Question 3?

7. Let x be a random variable and a be a constant. Show

   (a) $Var(ax) = a^2 Var(x)$.

   (b) $Var(x + a) = Var(x)$.

8. Show $Var(x + y) = Var(x) + Var(y) + 2Cov(x,y)$.

9. Let x and y be random variables, and let a and b be constants. Show $Cov(x+a, y+b) = Cov(x,y)$.

10. Let x and y be random variables, with $E(x) = \mu_x$, $E(y) = \mu_y$, $Var(x) = \sigma_x^2$, $Var(y) = \sigma_y^2$, $Cov(x,y) = \sigma_{xy}$ and $Corr(x,y) = \rho_{xy}$. Let a and b be non-zero constants.

    (a) Find $Cov(ax, y)$.

    (b) Find $Corr(ax, y)$. Do not forget that a could be negative.

11. Let $y_1, \ldots, y_n$ be numbers, and $\overline{y} = \frac{1}{n}\sum_{i=1}^n y_i$. Show

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

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

    (c) The sum of squares $Q_m = \sum_{i=1}^n (y_i - m)^2$ is minimized when $m = \overline{y}$.

12. 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 $\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^n x_i y_i \; - \; n\overline{x}\,\overline{y}$.

13. 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.

    (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. This is very quick.

(e) Let $a_1, \ldots, a_n$ be constants and define the linear combination L by $L = \sum_{i=1}^n a_i y_i$. Show that if $\sum_{i=1}^n a_i = 1$, then L is an unbiased estimator of μ.

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

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

14. Here is a simple linear regression model. Let $y = \beta_0 + \beta_1 x + \epsilon$, where β_0 and β_1 are constants (typically unknown), x is a known, observable constant, and ϵ is a random variable with expected value zero and variance σ^2.

(a) What is $E(y)$?

(b) What is $Var(y)$?

(c) Suppose that the distribution of ϵ is normal, so that it has density $f(\epsilon) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\epsilon^2}{2\sigma^2}}$. Find the distribution of y. Show your work. Hint: differentiate the cumulative distribution function of y.

(d) Suppose there are two equations:

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_1 + \epsilon_1 \\
y_2 &= \beta_0 + \beta_1 x_2 + \epsilon_2
\end{aligned}
$$

with $E(\epsilon_1) = E(\epsilon_2) = 0$, $Var(\epsilon_1) = Var(\epsilon_2) = \sigma^2$ and $Cov(\epsilon_1, \epsilon_2) = 0$. Note that x_1 and x_2 are constants, not random variables. What is $Cov(y_1, y_2)$? You don't need to show any work. Just refer to a problem you solved earlier in this assignment.

15. Let $\mathbf{A} = \left(\begin{array}{rr} 2 & 5 \\ 1 & -4 \\ 0 & 3 \end{array} \right)$ and $\mathbf{B} = \left(\begin{array}{rr} 1 & 0 \\ 2 & 3 \\ -1 & 3 \end{array} \right)$. Which of the following are possible to compute? Don't do the calculations. Just answer each one Possible or Impossible.
    (a) $\mathbf{A}+\mathbf{B}$
    (b) $\mathbf{A}-\mathbf{B}$
    (c) $\mathbf{AB}$
    (d) $\mathbf{A}'\mathbf{B}$
    (e) $\mathbf{B}'\mathbf{A}$
    (f) $\mathbf{A}/\mathbf{B}$
    (g) $\mathbf{B}/\mathbf{A}$

16. For the matrices of Question 15, verify that
$$\mathbf{A}'\mathbf{B} = \left(\begin{array}{rr} 4 & 3 \\ -6 & -3 \end{array} \right) \quad \text{and} \quad \mathbf{B}'\mathbf{A} = \left(\begin{array}{rrr} 2 & 19 & 13 \\ 1 & -10 & -13 \\ 0 & 9 & 9 \end{array} \right).$$

17. Let $\mathbf{c} = \left(\begin{array}{r} 2 \\ 1 \\ 0 \end{array} \right)$ and $\mathbf{d} = \left(\begin{array}{r} 1 \\ 2 \\ -1 \end{array} \right)$. Verify that
$$\mathbf{c}'\mathbf{d} = 4 \quad \text{and} \quad \mathbf{cd}' = \left(\begin{array}{rrr} 2 & 4 & -2 \\ 1 & 2 & -1 \\ 0 & 0 & 0 \end{array} \right).$$

18. Matrix multiplication does not commute. That is, if $\mathbf{A}$ and $\mathbf{B}$ are matrices, in general it is *not* true that $\mathbf{AB} = \mathbf{BA}$ unless both matrices are 1×1. Establish this important fact by making up a simple numerical example in which $\mathbf{A}$ and $\mathbf{B}$ are both 2×2 matrices. Carry out the multiplication, showing $\mathbf{AB} \neq \mathbf{BA}$.

19. Let $\mathbf{A}$ be a square matrix with the determinant of $\mathbf{A}$ (denoted $|\mathbf{A}|$) equal to zero. What does this tell you about $\mathbf{A}^{-1}$? No proof is required here.

20. Recall that $\mathbf{A}$ symmetric means $\mathbf{A} = \mathbf{A}'$. Let $\mathbf{X}$ be an n by p matrix. Prove that $\mathbf{X}'\mathbf{X}$ is symmetric.

21. Let $\mathbf{X}$ be an n by p matrix with $n \neq p$. Why is it incorrect to say that $(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{X}^{-1}\mathbf{X}'^{-1}$?
