STA 302f14 Assignment One[1]

The following formulas will be supplied with Quiz One.

$E(g(X)) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$	$E(g(\mathbf{X})) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_p)\, f_{\mathbf{X}}(x_1, \ldots, x_p)\, dx_1 \ldots dx_p$
$Var(Y) = E[(Y - \mu_Y)^2]$	$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$

1. Let X be a continuous random variable and let a be a constant. Using the expression for $E(g(X))$ above, show $E(a) = a$.

2. Let X_1 and X_2 be continuous random variables. Using the expression for $E(g(\mathbf{X}))$ above, show $E(X_1 + X_2) = E(X_1) + E(X_2)$. If you assume independence, you get a zero.

3. Let Y_1 and Y_2 be continuous random variables that are *independent*. Using the expression for $E(g(\mathbf{Y}))$, show $E(Y_1 Y_2) = E(Y_1)E(Y_2)$. Draw an arrow to the place in your answer where you use independence, and write "This is where I use independence."

4. Using the definitions of variance covariance along with familiar properties of expected value (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 3.

5. In the following, X and Y are 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. Use definitions and familiar properties of expected value, not integrals.
   (a) $Var(aX) = aVar(X)$ or $Var(aX) = a^2 Var(X)$
   (b) $Var(aX + b) = a^2 Var(X) + b^2$ or $Var(aX + b) = a^2 Var(X)$
   (c) $Var(a) = 0$ or $Var(a) = a^2$
   (d) $Cov(X + a, Y + b) = Cov(X,Y) + ab$ or $Cov(X + a, Y + b) = Cov(X,Y)$
   (e) $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$ or $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X,Y)$

6. 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 $m = \overline{Y}$.

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

7. 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 $E(\sum_{i=1}^n Y_i)$.

   (b) Find $Var\left(\sum_{i=1}^n Y_i\right)$. Show your work. Draw an arrow to the place in your answer where you use independence, and write "This is where I use 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 $L = \sum_{i=1}^n a_i Y_i$. What condition on the a_i values makes L 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 $Var(L)$?

8. Let $Y_1, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2, so that $T = \frac{\sqrt{n}(\overline{Y}-\mu)}{S} \sim t(n-1)$. This is something you don't need to prove, for now.

   (a) Derive a $(1-\alpha)100\%$ confidence interval for μ. "Derive" means show all the high school algebra. Use the symbol $t_{\alpha/2}$ for the number satisfying $Pr(T > t_{\alpha/2}) = \alpha/2$.

   (b) A random sample with $n = 23$ yields $\overline{Y} = 2.57$ and a sample variance of $S^2 = 5.85$. Using the critical value $t_{0.025} = 2.07$, give a 95% confidence interval for μ. The answer is a pair of numbers.

   (c) Test $H_0 : \mu = 3$ at $\alpha = 0.05$.

      i. Give the value of the T statistic. The answer is a number.
      ii. State whether you reject H_0, Yes or No.
      iii. Can you conclude that μ is different from 3? Answer Yes or No.
      iv. If the answer is Yes, state whether $\mu > 3$ or $\mu < 3$. Pick one.

   (d) Show that using a t-test, $H_0 : \mu = \mu_0$ is rejected at significance level α if and only the $(1-\alpha)100\%$ confidence interval for μ does not include μ_0. The problem is easier if you start by writing the set of $Y_1, \ldots, Y_n$ values for which H_0 is *not* rejected.

   (e) In Question 8b, does this mean $Pr\{1.53 < \mu < 3.61\} = 0.95$? Answer Yes or No and briefly explain.

9. Label each statement below True or False. Write "T" or "F" beside each statement. Assume the $\alpha = 0.05$ significance level. If there are True-False questions on the quiz, you will need to get most of them right (say 4 out of 5) in order to get any credit.

   (a) _____ The p-value is the probability that the null hypothesis is true.

   (b) _____ The p-value is the probability that the null hypothesis is false.

   (c) _____ In a study comparing a new drug to the current standard treatment, the null hypothesis is rejected. This means the new drug is ineffective.

   (d) _____ We observe $r = -0.70$, $p = .009$. We conclude that high values of X tend to go with low values of Y and low values of X tend to go with high values of Y.

   (e) _____ The p-value is the probability of failing to replicate significant results in a second independent random sample of the same size.

   (f) _____ The greater the p-value, the stronger the evidence against the null hypothesis.

   (g) _____ If $p > .05$ we reject the null hypothesis at the .05 level.

   (h) _____ If $p < .05$ we reject the null hypothesis at the .05 level.

   (i) _____ In a study comparing a new drug to the current standard treatment, $p > .05$. We conclude that the new drug and the existing treatment are not equally effective.

   (j) _____ The 95% confidence interval for β_3 is from -0.26 to 3.12. This means $P\{-0.26 < \beta_3 < 3.12\} = 0.95$.

   (k) _____ The p-value is the maximum significance level α such that the null hypothesis is rejected.

   (l) _____ The p-value is the minimum significance level α such that the null hypothesis is rejected.

10. In *Linear models in statistics*, do problems 2.3, 2.4 and 2.14 parts a, b, and h. Review Chapter 2 or your linear algebra text as necessary. The answers are in the back of the book. The *trace* of a square matrix $\mathbf{A}$, denoted $tr(\mathbf{A})$, is the sum of the diagonal elements.

11. Which statement is true? (Quantities in boldface are matrices of constants.)

   (a) $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$

   (b) $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{BA}+\mathbf{CA}$

   (c) Both a and b

   (d) Neither a nor b

12. Which statement is true?

   (a) $a(\mathbf{B}+\mathbf{C}) = a\mathbf{B}+a\mathbf{C}$

   (b) $a(\mathbf{B}+\mathbf{C}) = \mathbf{B}a+\mathbf{C}a$

   (c) Both a and b

   (d) Neither a nor b

13. Which statement is true?

    (a) $(\mathbf{B}+\mathbf{C})\mathbf{A} = \mathbf{AB}+\mathbf{AC}$
    (b) $(\mathbf{B}+\mathbf{C})\mathbf{A} = \mathbf{BA}+\mathbf{CA}$
    (c) Both a and b
    (d) Neither a nor b

14. Which statement is true?

    (a) $(\mathbf{AB})' = \mathbf{A}'\mathbf{B}'$
    (b) $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$
    (c) Both a and b
    (d) Neither a nor b

15. Which statement is true?

    (a) $\mathbf{A}'' = \mathbf{A}$
    (b) $\mathbf{A}''' = \mathbf{A}'$
    (c) Both a and b
    (d) Neither a nor b

16. Suppose that the square matrices $\mathbf{A}$ and $\mathbf{B}$ both have inverses. Which statement is true?

    (a) $(\mathbf{AB})^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1}$
    (b) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
    (c) Both a and b
    (d) Neither a nor b

17. Which statement is true?

    (a) $(\mathbf{A}+\mathbf{B})' = \mathbf{A}'+\mathbf{B}'$
    (b) $(\mathbf{A}+\mathbf{B})' = \mathbf{B}'+\mathbf{A}'$
    (c) $(\mathbf{A}+\mathbf{B})' = (\mathbf{B}+\mathbf{A})'$
    (d) All of the above
    (e) None of the above

18. Which statement is true?

    (a) $(a+b)\mathbf{C} = a\mathbf{C}+b\mathbf{C}$
    (b) $(a+b)\mathbf{C} = \mathbf{C}a+\mathbf{C}b$
    (c) $(a+b)\mathbf{C} = \mathbf{C}(a+b)$
    (d) All of the above
    (e) None of the above

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. Recall that an inverse of the matrix $\mathbf{A}$ (denoted $\mathbf{A}^{-1}$) is defined by two properties: $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ and $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. Prove that inverses are unique, as follows. Let $\mathbf{B}$ and $\mathbf{C}$ both be inverses of $\mathbf{A}$. Show that $\mathbf{B} = \mathbf{C}$.

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