STA 305s14 Regular Assignment One[1]

This assignment is strictly review. Problems are preparation for Term Test One on Feb. 3d, and will not be handed in. Use the formula sheet, which is posted on the course home page.

1. 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 term test or final exam, you will need to get most of them right (say 8 out of 10) 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 that the independent and dependent variable are related.

   (g) ______ The greater the p-value, the stronger the evidence against the null hypothesis.

   (h) ______ If $p > .05$ we reject the null hypothesis at the .05 level.

   (i) ______ If $p < .05$ we reject the null hypothesis at the .05 level.

   (j) ______ 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.

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

   (l) ______ When you add another independent variable in multiple regression, R^2 cannot go down.

   (m) ______ We observe $r = 0.50$, $p = .002$. This means that 50% of the variation in the dependent variable is explained by a linear relationship with the independent variable.

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

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

2. In the following, X and Y are random variables, while a and b are fixed constants. Using the definitions of variance covariance (see formula sheet) along with familiar properties of expected value, show the following:

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

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

   (c) $Var(aX + b) = a^2 Var(X)$

   (d) $Var(a) = 0$

   (e) $Cov(X + a, Y + b) = Cov(X,Y)$

   (f) $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab Cov(X,Y)$

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

3. 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 $m = \overline{y}$.

4. Let $Y_1, \ldots, Y_n$ be independent random variables with $E(Y_i) = \mu$ and $Var(Y_i) = \sigma^2$ for $i = 1, \ldots, n$. Let $a_1, \ldots, a_n$ be constants and define the linear combination L by $L = \sum_{i=1}^n a_i Y_i$.
   (a) What is $E(L)$? Show your work. Do you use independence? Answer Yes or No. If the answer is Yes, indicate where you use it by drawing an arrow to one of the equals signs, and writing "I use independence here."
   (b) What is $Var(L)$? Show your work. Do you use independence? Answer Yes or No. If the answer is Yes, indicate where you use it by drawing an arrow to one of the equals signs, and writing "I use independence here."
   (c) A statistic T is an *unbiased estimator* of a parameter θ if $E(T) = \theta$. Suppose that L is an unbiased estimator of μ. Does this mean that $a_i = \frac{1}{n}$ for $i = 1, \ldots, n$? Answer Yes or No. If the answer is Yes, prove it. If the answer is No, give another set of constants $a_1, \ldots, a_n$ that make L unbiased.

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

6. Let $\mathbf{X}$ be a random vector with expected value $\boldsymbol{\mu}_x$, and let $\mathbf{Y}$ be a random vector with expected value $\boldsymbol{\mu}_y$. Using the definitions on the formula sheet,
   (a) Show $cov(\mathbf{Y}) = E\{\mathbf{Y}\mathbf{Y}'\} - \boldsymbol{\mu}_y\boldsymbol{\mu}_y'$. Why is it incorrect (and worth zero marks) to say $E\{\mathbf{Y}^2\} - \boldsymbol{\mu}_y^2$?
   (b) Let $\mathbf{A}$ be a matrix of constants (of the right dimensions). Show $cov(\mathbf{AY}) = \mathbf{A}cov(\mathbf{Y})\mathbf{A}'$. Why is it incorrect (and worth zero marks) to say $\mathbf{A}^2 cov(\mathbf{Y})$?

7. In this course, you will not need to prove any distribution facts using moment-generating functions. But you have to know some standard results that are not directly on the formula sheet. Just write down the answers to the following.
   (a) Let $Y_1, \ldots, Y_k$ be independent chi-squared random variables with respective parameters $\nu_1, \ldots, \nu_k$. What is the distribution of $Y = \sum_{j=1}^k Y_j$?
   (b) Let $Y \sim N(\mu, \sigma^2)$. What is the distribution of $\frac{Y-\mu}{\sigma}$?
   (c) $Z \sim N(0,1)$. What is the distribution of Z^2?
   (d) Let $Y_1, \ldots, Y_n$ be a random sample (meaning they are independent and identically distributed) from a normal distribution with expected value μ and variance σ^2.
      i. What is the distribution of $Y = \sum_{i=1}^n Y_i$?
      ii. What is the distribution of $\overline{Y}$?

8. Again, let $Y_1, \ldots, Y_n$ be a random sample from a univariate normal distribution with expected value μ and variance σ^2. Using material from the formula sheet (not moment-generating functions), *prove* the distribution of $Y = \sum_{i=1}^n Y_i$. Let $\mathbf{1}$ denote an $n \times 1$ vector of ones.

9. Show that if $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where the covariance matrix $\boldsymbol{\Sigma}$ is strictly positive definite, then $W = (\mathbf{Y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$ has a chi-squared distribution with p degrees of freedom. It will help to start by finding the distribution of $\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu})$.

10. For the general linear regression model in which $\mathbf{X}$ is an n by $(k+1)$ matrix of constants,

    (a) Give $E(\mathbf{Y})$. Show the calculation.
    (b) Give $cov(\mathbf{Y})$. Show the calculation.
    (c) What is the distribution of $\mathbf{Y}$? Just write down the answer. What fact on the formula sheet lets you do this?
    (d) Give $E(\widehat{\boldsymbol{\beta}})$. Show the calculations. Simplify.
    (e) Give $cov(\widehat{\boldsymbol{\beta}})$. Show the calculations. Simplify.
    (f) What is the distribution of $\widehat{\boldsymbol{\beta}}$? Just write down the answer. What fact on the formula sheet lets you do this?

11. Let $Y_i = \beta x_i + \epsilon_i$ for $i = 1, \ldots, n$, where $\epsilon_1, \ldots, \epsilon_n$ are a random sample from a distribution with expected value zero and variance σ^2, and β and σ^2 are unknown constants. The numbers $x_1, \ldots, x_n$ are known, observed constants. This is a special case of the general linear model given on the formula sheet.

    (a) What is $\mathbf{X}'\mathbf{X}$?
    (b) What is $\mathbf{X}'\mathbf{Y}$?
    (c) $(\mathbf{X}'\mathbf{X})^{-1}$?
    (d) What is $\widehat{\boldsymbol{\beta}}$?

12. Let $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ for $i = 1, \ldots, n$, where $\epsilon_1, \ldots, \epsilon_n$ are a random sample from a distribution with expected value zero and variance σ^2, and β_0, β_1 and σ^2 are unknown constants. The numbers $x_1, \ldots, x_n$ are known, observed constants. This is a special case of the general linear model given on the formula sheet.

    (a) What is $\boldsymbol{\beta}$?
    (b) What is $\mathbf{X}'\mathbf{X}$?
    (c) What is $\mathbf{X}'\mathbf{Y}$?