STA 302f13 Assignment Five[1]

These problems are preparation for the quiz in tutorial on Friday October 17th, and are not to be handed in.

1. The "hat" matrix is given by $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. It's called the hat matrix because it puts a hat on $\mathbf{Y}$ by $\mathbf{HY} = \widehat{\mathbf{Y}}$. The hat matrix is special.
   (a) What are dimensions (number of rows and columns) of the hat matrix?
   (b) Show that the hat matrix is symmetric.
   (c) Show that the hat matrix is *idempotent*, meaning $\mathbf{H}^{1/2} = \mathbf{H}$.
   (d) Show that $(\mathbf{I} - \mathbf{H})$ is also symmetric and idempotent.
   (e) Write $\widehat{\boldsymbol{\epsilon}}$ in terms of the hat matrix (it's a function of $\mathbf{I} - \mathbf{H}$).
   (f) Write SSE in terms of the hat matrix. Simplify.
   (g) From the last assignment, recall that the the subset of $\mathbb{R}^n$ spanned by the columns of the $\mathbf{X}$ matrix is $\mathcal{V} = \{\mathbf{v} = \mathbf{Xb} : \mathbf{b} \in \mathbb{R}^{k+1}\}$. Also recall that $\widehat{\mathbf{Y}}$, being the closest point in $\mathcal{V}$ to the data vector $\mathbf{Y}$, is the orthoganal projection of $\mathbf{Y}$ onto $\mathcal{V}$. Since $\mathbf{Y}$ could be any point in $\mathbb{R}^n$, multiplication by the hat matrix $\mathbf{H}$ is the operation that projects any point in $\mathbb{R}^n$ onto $\mathcal{V}$. It's like the light bulb above the point that you turn on in order to cast a shadow onto $\mathcal{V}$.

       All this talk implies that if a point is already in $\mathcal{V}$, its shadow is the point itself. Verify that $\mathbf{Hv} = \mathbf{v}$ for any $\mathbf{v} \in \mathcal{V}$.

2. The first parts of this question were in Assignment One. Let $Y_1, \ldots, Y_n$ be independent random variables with $E(Y_i) = \mu$ and $Var(Y_i) = \sigma^2$ for $i = 1, \ldots, n$.
   (a) Write down $E(\overline{Y})$ and $Var(\overline{Y})$.
   (b) Let $c_1, \ldots, c_n$ be constants and define the linear combination L by $L = \sum_{i=1}^n c_i Y_i$. What condition on the c_i values makes L an unbiased estimator of μ? Recall that L unbiased means that $E(L) = \mu$ for *all* real μ. Treat the cases $\mu = 0$ and $\mu \neq 0$ separately.
   (c) Is $\overline{Y}$ a special case of L? If so, what are the c_i values?
   (d) What is $Var(L)$?
   (e) Now show that $Var(\overline{Y}) < Var(L)$ for every unbiased $L \neq \overline{Y}$. Hint: Add and subtract $\frac{1}{n}$.

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

3. For the general linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, suppose we want to estimate the linear combination $\mathbf{a}'\boldsymbol{\beta}$ based on sample data. The Gauss-Markov Theorem tells us that the most natural choice is also (in a sense) the best choice. This question leads you through the proof of the Gauss-Markov Theorem. Your class notes should help. Also see your solution of Question 2.

   (a) What is the most natural choice for estimating $\mathbf{a}'\boldsymbol{\beta}$?

   (b) Show that it's unbiased.

   (c) The natural estimator is a *linear* unbiased estimator of the form $\mathbf{c}_0'\mathbf{Y}$. What is the $n \times 1$ vector $\mathbf{c}_0$?

   (d) Of course there are lots of other possible linear unbiased estimators of $\mathbf{a}'\boldsymbol{\beta}$. They are all of the form $\mathbf{c}'\mathbf{Y}$; the natural estimator $\mathbf{c}_0'\mathbf{Y}$ is just one of these. The best one is the one with the smallest variance, because its distribution is the most concentrated around the right answer. What is $Var(\mathbf{c}'\mathbf{Y})$? Show your work.

   (e) We insist that $\mathbf{c}'\mathbf{Y}$ be unbiased. Show that if $E(\mathbf{c}'\mathbf{Y}) = \mathbf{a}'\boldsymbol{\beta}$ for *all* $\boldsymbol{\beta} \in \mathbb{R}^{k+1}$, we must have $\mathbf{X}'\mathbf{c} = \mathbf{a}$.

   (f) So, the task is to minimize $Var(\mathbf{c}'\mathbf{Y})$ by minimizing $\mathbf{c}'\mathbf{c}$ over all $\mathbf{c}$ subject to the constraint $\mathbf{X}'\mathbf{c} = \mathbf{a}$. As preparation for this, show $(\mathbf{c} - \mathbf{c}_0)'\mathbf{c}_0 = 0$.

   (g) Using the result of the preceding question, show

$$\mathbf{c}'\mathbf{c} = (\mathbf{c} - \mathbf{c}_0)'(\mathbf{c} - \mathbf{c}_0) + \mathbf{c}_0'\mathbf{c}_0.$$

   (h) Since the formula for $\mathbf{c}_0$ has no $\mathbf{c}$ in it, what choice of $\mathbf{c}$ minimizes the preceding expression? How do you know that the minimum is unique?

   The conclusion is that $\mathbf{c}_0'\mathbf{Y} = \mathbf{a}'\widehat{\boldsymbol{\beta}}$ is the Best Linear Unbiased Estimator (BLUE) of $\mathbf{a}'\boldsymbol{\beta}$.

4. The model for simple regression through the origin is $Y_i = \beta x_i + \epsilon_i$, where $\epsilon_1, \ldots, \epsilon_n$ are independent with expected value 0 and variance σ^2. In previous homework, you found the least squares estimate of β to be $\widehat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$.

   (a) What is $Var(\widehat{\beta})$?

   (b) Let $\widehat{\beta}_2 = \frac{\overline{Y}_n}{\overline{x}_n}$.

      i. Is $\widehat{\beta}_2$ an unbiased estimator of β? Answer Yes or No and show your work.

      ii. Is $\widehat{\beta}_2$ a linear combination of the Y_i variables, of the form $L = \sum_{i=1}^n c_i Y_i$? Is so, what is c_i?

      iii. What is $Var(\widehat{\beta}_2)$?

      iv. How do you know $Var(\widehat{\beta}) \leq Var(\widehat{\beta}_2)$? No calculations are necessary.

(c) Let $\widehat{\beta}_3 = \frac{1}{n}\sum_{i=1}^n \frac{Y_i}{x_i}$.

i. Is $\widehat{\beta}_3$ an unbiased estimator of β? Answer Yes or No and show your work.

ii. Is $\widehat{\beta}_3$ a linear combination of the Y_i variables, of the form $L = \sum_{i=1}^n c_i Y_i$? Is so, what is c_i?

iii. What is $Var(\widehat{\beta}_3)$?

iv. How do you know $Var(\widehat{\beta}) \leq Var(\widehat{\beta}_3)$? No calculations are necessary.

5. The joint moment-generating function of a p-dimensional random vector $\mathbf{X}$ is defined as $M_{\mathbf{X}}(\mathbf{t}) = E\left(e^{\mathbf{t}'\mathbf{X}}\right)$.

(a) Let $\mathbf{Y} = \mathbf{AX}$, where $\mathbf{A}$ is a matrix of constants. Find the moment-generating function of $\mathbf{Y}$.

(b) Let $\mathbf{Y} = \mathbf{X}+\mathbf{c}$, where $\mathbf{c}$ is a $p\times 1$ vector of constants. Find the moment-generating function of $\mathbf{Y}$.

6. Let $Z_1, \ldots, Z_p \stackrel{i.i.d.}{\sim} N(0,1)$, and

$$\mathbf{Z} = \left(\begin{array}{c} Z_1 \\ \vdots \\ Z_p \end{array}\right).$$

(a) What is the joint moment-generating function of $\mathbf{Z}$? Show some work.

(b) Let $\mathbf{Y} = \boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}$, where $\boldsymbol{\Sigma}$ is a $p \times p$ symmetric *non-negative definite* matrix and $\boldsymbol{\mu} \in \mathbb{R}^p$.

i. What is $E(\mathbf{Y})$?

ii. What is the variance-covariance matrix of $\mathbf{Y}$? Show some work.

iii. What is the moment-generating function of $\mathbf{Y}$? Show your work.

7. We say the p-dimensional random vector $\mathbf{Y}$ is multivariate normal with expected value $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, and write $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, when $\mathbf{Y}$ has moment-generating function $M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}$.

(a) Let $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{W} = \mathbf{AY}$, where $\mathbf{A}$ is an $r \times p$ matrix of constants. What is the distribution of $\mathbf{W}$? Show your work.

(b) Let $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{W} = \mathbf{Y} + \mathbf{c}$, where $\mathbf{A}$ is an $p \times 1$ vector of constants. What is the distribution of $\mathbf{W}$? Show your work.

8. Let $\mathbf{Y} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with

$$\mathbf{Y} = \left(\begin{array}{c} Y_1 \\ Y_2 \end{array}\right) \quad \boldsymbol{\mu} = \left(\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right) \quad \boldsymbol{\Sigma} = \left(\begin{array}{cc} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{array}\right)$$

Using moment-generating functions, show Y_1 and Y_2 are independent.

9. Let $\mathbf{X} = (X_1, X_2, X_3)'$ be multivariate normal with

$$\boldsymbol{\mu} = \left[\begin{array}{c} 1 \\ 0 \\ 6 \end{array}\right] \mbox{ and } \boldsymbol{\Sigma} = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{array}\right].$$

Let $Y_1 = X_1 + X_2$ and $Y_2 = X_2 + X_3$. Find the joint distribution of Y_1 and Y_2.

10. Let X_1 be Normal(μ_1, σ_1^2), and X_2 be Normal(μ_2, σ_2^2), independent of X_1. What is the joint distribution of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$? What is required for Y_1 and Y_2 to be independent? Hint: Use matrices.

11. Show that if $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma}$ positive definite, then $Y = (\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ has a chi-square distribution with p degrees of freedom.

12. Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution.
    (a) Show $Cov(\overline{X}, (X_j - \overline{X})) = 0$ for $j = 1, \ldots, n$.
    (b) Show that $\overline{X}$ and S^2 are independent.
    (c) Show that
    $$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$
    where $S^2 = \frac{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}{n-1}$. Hint: $\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n \left(X_i - \overline{X} + \overline{X} - \mu\right)^2 = \ldots$

13. Recall the definition of the t distribution. If $Z \sim N(0,1)$, $W \sim \chi^2(\nu)$ and Z and W are independent, then $T = \frac{Z}{\sqrt{W/\nu}}$ is said to have a t distribution with ν degrees of freedom, and we write $T \sim t(\nu)$. As in the last question, let $X_1, \ldots, X_n$ be random sample from a $N(\mu, \sigma^2)$ distribution. Show that $T = \frac{\sqrt{n}(\overline{X} - \mu)}{S} \sim t(n-1)$.
