STA 302f16 Assignment Five[1]

Except for Problem **??**, these problems are preparation for the quiz in tutorial on Thursday October 20th, and are not to be handed in. As usual, **at times you may be asked to prove something that is not true**. In this case you should say why the statement is not always true. Please bring your printout for Problem **??** to the quiz. Do not write anything on the printout in advance of the quiz, except possibly your name and student number.

1. 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: $\sum_{i=1}^n (c_i - \overline{c})^2 = \sum_{i=1}^n c_i^2 - \frac{(\sum_{i=1}^n c_i)^2}{n}$.

   This is the simplest case of the Gauss-Markov Theorem. See your class notes.

2. For the general linear model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, suppose we want to estimate the linear combination $\boldsymbol{\ell}'\boldsymbol{\beta}$ 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 answer to Question **??**.

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

   (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 $\boldsymbol{\ell}'\boldsymbol{\beta}$. 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.

      i. What is $Var(\mathbf{c}'\mathbf{y})$? Show your work.

      ii. What is $Var(\mathbf{c}_0'\mathbf{y})$? Show your work.

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

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

(f) Show that if $\mathbf{c}$ satisfies $E(\mathbf{c}'\mathbf{y}) = \boldsymbol{\ell}'\boldsymbol{\beta}$ for all $\boldsymbol{\beta} \in \mathbb{R}^{k+1}$, then $H\mathbf{c} = \mathbf{c}_0$. That is, $\mathbf{c}_0$ is the projection of $\mathbf{c}$ onto the space spanned by the columns of the X matrix.

(g) By direct calculation, show $Var(\mathbf{c}'\mathbf{y}) \geq Var(\mathbf{c}_0'\mathbf{y})$. This means that no linear combination can have a smaller variance than $\mathbf{c}_0'\mathbf{y}$.

(h) Show that if $Var(\mathbf{c}'\mathbf{y}) = Var(\mathbf{c}_0'\mathbf{y})$, then $\mathbf{c} = \mathbf{c}_0$. This means that no other linear combination of $\mathbf{y}$ values can even tie the variance of $\mathbf{c}_0'\mathbf{y}$.

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

3. 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 $b = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$.

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

   (b) Let $b_2 = \frac{\overline{y}_n}{\overline{x}_n}$.

      i. Is b_2 an unbiased estimator of β? Answer Yes or No and show your work.

      ii. Is b_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 $Var(b_2)$?

      iv. How do you know $Var(b) < Var(b_2)$? No calculations are necessary.

   (c) Let $b_3 = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{x_i}$.

      i. Is b_3 an unbiased estimator of β? Answer Yes or No and show your work.

      ii. Is b_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 $Var(b_3)$?

      iv. How do you know $Var(b) < Var(b_3)$? No calculations are necessary.

4. In practice, $\mathbf{e}$ will never be zero. Why? It may help to think of the least-squares line on a two-dimensional scatterplot.

5. Show that *if* the hat matrix H has an inverse, then $\mathbf{e} = \mathbf{0}$. Start by calculating $H\mathbf{e}$.

6. We will now see that H cannot have an inverse if there are more observations than β parameters. Let $n > k + 1$. Since the space spanned by the columns of X is of dimension $k + 1$, there is an $n \times 1$ vector $\mathbf{v}$ that is *not* in the space spanned by the columns of X. That is, $\mathbf{v} \neq X\mathbf{a}$ for any $\mathbf{a} \in \mathbb{R}^{k+1}$.

   (a) The projection of $\mathbf{v}$ onto the space spanned by the columns of X is $H\mathbf{v}$. Show that the $n \times 1$ vector $\mathbf{v} - H\mathbf{v} \neq \mathbf{0}$. Notice how $\mathbf{v} - H\mathbf{v}$ is like an $\mathbf{e}$?

(b) Show that the columns of H are linearly dependent.

(c) Show that H has no inverse.

7. True or False: The sum of residuals is always equal to zero.

8. True or False: The sum of *expected* residuals is always equal to zero.

9. True or False: The sum of residuals is always equal to zero if the model has an intercept.

10. Sometimes one can learn by just playing around. Suppose we fit a regression model, obtaining $\mathbf{b}$, $\widehat{\mathbf{y}}$, $\mathbf{e}$ and so on. Then we fit another regression model with the same independent variables, but this time using $\widehat{\mathbf{y}}$ as the dependent variable instead of $\mathbf{y}$.

    (a) Denote the vector of estimated regression coefficients from the new model by $\mathbf{b}_2$. Calculate $\mathbf{b}_2$ and simplify. Should you be surprised at this answer?

    (b) Calculate $\widehat{\widehat{\mathbf{y}}} = X\mathbf{b}_2$. Why is this not surprising if you think in terms of projections?

11. Now do the same thing as in the preceding question, but with $\mathbf{e}$ as the dependent variable. Can you understand this in terms of projections?

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

    (a) Let $\mathbf{y} = A\mathbf{x}$, where 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 $\mathbf{y}$.

13. 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 $\mathbf{y} = \Sigma^{1/2}\mathbf{z} + \boldsymbol{\mu}$, where Σ 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.

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

(a) Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \Sigma)$ and $\mathbf{w} = A\mathbf{y}$, where A is an $r \times p$ matrix of constants. What is the distribution of $\mathbf{w}$? Use moment-generating functions to prove your answer.

(b) Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \Sigma)$ and $\mathbf{w} = \mathbf{y} + \mathbf{c}$, where $\mathbf{c}$ is an $p \times 1$ vector of constants. What is the distribution of $\mathbf{w}$? Use moment-generating functions to prove your answer.

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

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \qquad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

Using moment-generating functions, show y_1 and y_2 are independent.

16. Let $x = (x_1, x_2, x_3)'$ be multivariate normal with

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 0 \\ 6 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let $y_1 = x_1 + x_2$ and $y_2 = x_2 + x_3$. Find the joint distribution of y_1 and y_2.

17. 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.

18. Here are some distribution facts that you will need to know without looking at a formula sheet in order to follow the proofs. You are responsible for the proofs of these facts too, but here you are just supposed to write down the answers.

(a) Let $x \sim N(\mu, \sigma^2)$ and $y = ax + b$, where a and b are constants. What is the distribution of y?

(b) Let $x \sim N(\mu, \sigma^2)$ and $z = \frac{x-\mu}{\sigma}$. What is the distribution of z?

(c) Let $x_1, \ldots, x_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. What is the distribution of $y = \sum_{i=1}^n x_i$?

(d) Let $x_1, \ldots, x_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. What is the distribution of the sample mean $\overline{x}$?

(e) Let $x_1, \ldots, x_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. What is the distribution of $z = \frac{\sqrt{n}(\overline{x}-\mu)}{\sigma}$?

(f) Let $x_1, \ldots, x_n$ be independent random variables, with $x_i \sim N(\mu_i, \sigma_i^2)$. Let $a_1, \ldots, a_n$ be constants. What is the distribution of $y = \sum_{i=1}^n a_i x_i$?

(g) Let $x_1, \ldots, x_n$ be independent random variables with $x_i \sim \chi^2(\nu_i)$ for $i = 1, \ldots, n$. What is the distribution of $y = \sum_{i=1}^n x_i$?

(h) Let $z \sim N(0, 1)$. What is the distribution of $y = z^2$?

(i) Let $x_1, \ldots, x_n$ be random sample from a $N(\mu, \sigma^2)$ distribution. What is the distribution of $y = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - \mu)^2$?

(j) Let $y = x_1 + x_2$, where x_1 and x_2 are independent, $x_1 \sim \chi^2(\nu_1)$ and $y \sim \chi^2(\nu_1 + \nu_2)$, where ν_1 and ν_2 are both positive. What is the distribution of x_2?

19. The `statclass` data consist of Quiz average, Computer assignment average, Midterm score and Final Exam score from a statistics class, long ago. At the R prompt, type

```
statclass = read.table("http://www.utstat.utoronto.ca/~brunner/data/legal/LittleStatclassdata.txt")
```

You now have access to the `statclass` data, just as you have access to the `trees` data set used in lecture, or any other R data set.

(a) Calculate $\mathbf{b}$ with the `lm` function. What is b_2? The answer is a number on your printout.

(b) What is the predicted Final Exam score for a student with a Quiz average of 8.5, a Computer average of 5, and a Midterm mark of 60%? The answer is a number. Be able to do this kind of thing on the quiz with a calculator. My answer is 63.84144.

(c) For any fixed Quiz Average and Computer Average, a score one point higher on the Midterm yields a predicted mark on the Final Exam that is _____ higher.

(d) For any fixed Quiz Average and Midterm score, a Computer average that is one point higher yields a predicted mark on the Final Exam that is _____ higher. Or is it lower?
