STA 6207 — Section 8907 — Fall 2008

Assignment 3

1. Let $\boldsymbol{Z}$ be a random column vector. Using the definition

$$V(\boldsymbol{Z}) \;=\; E\big[(\boldsymbol{Z} - E(\boldsymbol{Z}))(\boldsymbol{Z} - E(\boldsymbol{Z}))'\big]$$

and properties of matrices and the matrix expected value, show that

$$V(\boldsymbol{Z}) \;=\; E(\boldsymbol{Z}\boldsymbol{Z}') - E(\boldsymbol{Z})E(\boldsymbol{Z})'.$$

2. A car rental agency would like to model its monthly volume of business (Y_i, in thousands of rentals) in a certain city in terms of the city's volume of business-related travel (X_{i1}, thousands) and volume of other tourism (X_{i2}, thousands) for the same month (indexed by i). Use the linear regression model

$$Y_i \;=\; \beta_0 \,+\, \beta_1 X_{i1} \,+\, \beta_2 X_{i2} + \epsilon_i, \qquad \epsilon_1, \ldots, \epsilon_8 \,\sim\, i.i.d.\, N(0, \sigma^2)$$

and the 8-month data set listed after your name below to complete the parts that follow.

	X11	X12	Y1	X21	X22	Y2	X31	X32	Y3	X41	X42	Y4
BAI LEI	11	14	2.5	12	20	2.9	10	24	2.9	9	25	2.8
BROWN STEPHEN V	12	23	3.0	13	25	3.0	9	15	2.2	12	23	3.2
CHANG CHE-SHUN	9	18	2.5	10	17	2.7	7	13	1.9	13	26	3.2
CHEN OU	12	17	3.0	13	29	3.5	9	14	2.4	12	22	3.3
DONOHOE MICHAEL	9	19	2.8	12	13	2.5	10	14	2.1	8	16	2.4
GAO HAIBING	8	17	2.5	9	14	2.5	8	13	2.4	10	16	2.8
GARG DIVYA	8	11	2.4	11	26	3.4	11	18	2.9	13	28	3.4
GLUCK MATHEW R	10	20	2.5	12	20	3.1	12	19	2.9	10	19	2.8
GORDON ROBERT F	13	18	2.8	10	25	2.6	10	15	2.4	8	10	2.4
GUCI LEDIA	15	24	3.4	9	17	2.5	8	14	1.8	12	20	2.9
HARIHARAN POOJA	10	24	3.1	8	15	2.3	13	27	3.3	8	10	2.1
HUANG LEI	9	14	2.4	9	15	2.2	8	15	2.4	12	30	3.4
KIM CHANMIN	7	13	1.9	7	18	2.3	8	13	2.3	11	13	2.7
KIRPICH ALEX	10	14	2.4	8	16	2.4	9	17	2.7	12	19	2.6
LEARY EMILY V	9	19	2.8	13	19	3.2	9	13	2.4	10	13	2.4
LI KE	12	25	3.1	12	22	2.8	10	19	2.7	12	19	2.7
LIN TONG	9	14	2.6	7	13	2.0	9	18	2.5	9	17	2.3
LIU MINZHAO	8	13	1.7	12	17	2.7	8	25	2.7	10	16	2.7
LU CUIE	9	14	2.4	7	13	2.1	10	15	2.4	7	10	1.7
LUO XUAN	14	25	3.5	16	27	3.5	11	14	2.5	14	23	3.3
MA LU	10	19	2.6	7	14	1.9	11	21	2.5	12	16	2.8
MALLICK PRANJAL	9	10	2.1	9	13	2.1	11	19	2.7	11	25	2.6
MARCUS GABRIEL	9	17	2.6	10	19	3.0	13	16	3.0	10	23	3.1
NAMKOONG YOUNG	9	15	2.7	11	19	2.8	8	15	2.3	12	18	3.1
NEAL DANIEL W	14	20	3.2	9	19	2.7	14	29	3.5	10	18	2.5
PETTERSON SONIA	14	21	3.1	12	17	3.1	9	16	2.6	10	21	2.8
PRANO BRIJIDA A	11	15	2.6	10	23	2.5	11	17	2.7	14	29	3.5
SHAO ANQI	10	22	2.7	11	20	2.7	12	32	3.8	13	26	3.2
SINHA AMIT	11	23	2.6	10	15	2.6	8	14	2.0	12	19	2.9

THAYER LAURA K	12	19	2.8	8	15	2.0	12	23	3.1	8	14	2.4
YE RONGZHONG	7	10	1.9	13	19	3.1	9	13	2.4	11	19	2.7
ZHOU ZHUO	13	33	3.3	15	26	3.3	12	21	3.1	8	16	2.1
ZHU XIAOYU	11	23	3.1	10	19	2.8	8	17	2.1	10	20	3.0
demo	11	26	3.0	9	15	2.5	15	20	3.2	11	12	2.6

CONTINUED ...

	X51	X52	Y5	X61	X62	Y6	X71	X72	Y7	X81	X82	Y8
BAI LEI	13	32	3.8	9	18	2.4	7	16	2.2	9	18	2.3
BROWN STEPHEN V	9	23	2.6	8	15	2.5	12	20	3.0	10	25	3.1
CHANG CHE-SHUN	15	35	3.8	8	17	2.4	12	23	3.2	13	21	3.3
CHEN OU	9	16	2.3	9	15	2.4	8	15	2.3	12	16	3.1
DONOHOE MICHAEL	13	31	3.4	13	19	2.8	11	29	3.3	9	16	2.4
GAO HAIBING	12	23	3.3	9	16	2.2	8	15	2.3	9	14	2.2
GARG DIVYA	10	18	2.5	14	31	3.4	11	21	2.9	11	22	2.6
GLUCK MATHEW R	8	13	2.3	7	16	2.1	11	14	2.5	12	15	2.7
GORDON ROBERT F	8	19	2.4	9	20	2.8	8	8	2.1	7	9	2.1
GUCI LEDIA	11	24	3.0	10	17	2.4	11	22	2.4	8	16	2.6
HARIHARAN POOJA	11	16	2.4	9	23	2.8	8	14	2.4	7	12	2.0
HUANG LEI	10	20	2.8	12	28	3.2	10	16	2.5	10	14	2.5
KIM CHANMIN	7	12	1.8	14	25	3.8	10	20	2.9	13	25	3.0
KIRPICH ALEX	12	20	2.7	8	11	2.1	10	15	2.3	13	27	3.6
LEARY EMILY V	8	19	2.3	11	32	3.5	10	17	2.7	9	12	2.2
LI KE	13	30	3.3	9	11	2.4	11	17	2.9	8	19	2.4
LIN TONG	10	18	2.6	8	16	2.0	11	17	2.8	8	17	2.6
LIU MINZHAO	13	25	3.6	9	14	2.5	7	10	1.9	12	31	3.7
LU CUIE	11	17	2.7	13	25	2.9	14	22	3.3	15	28	3.6
LUO XUAN	13	26	3.8	10	18	2.5	11	17	2.9	14	32	3.7
MA LU	10	17	2.4	11	12	2.5	12	21	3.2	13	21	3.6
MALLICK PRANJAL	8	15	1.9	12	22	2.8	9	19	2.5	10	13	2.8
MARCUS GABRIEL	9	10	2.1	7	22	2.6	12	26	3.1	8	17	2.2
NAMKOONG YOUNG	10	19	2.6	9	17	2.4	8	14	2.2	15	18	3.2
NEAL DANIEL W	12	27	3.4	8	13	2.6	15	25	3.4	7	10	1.9
PETTERSON SONIA	8	20	2.8	11	17	2.4	9	17	2.4	11	16	2.6
PRANO BRIJIDA A	12	18	2.9	18	34	3.8	10	22	3.1	10	15	2.3
SHAO ANQI	11	21	2.8	9	13	2.3	11	19	2.7	10	20	2.9
SINHA AMIT	9	17	2.5	9	14	2.1	9	10	1.9	13	28	3.4
THAYER LAURA K	11	17	2.9	9	13	2.2	10	16	2.6	12	27	3.0
YE RONGZHONG	10	17	2.6	11	20	2.6	10	20	2.5	11	18	2.8
ZHOU ZHUO	11	22	3.1	10	19	2.6	13	26	3.4	18	36	4.6
ZHU XIAOYU	9	14	2.4	8	12	2.3	11	14	2.3	8	14	2.4
demo	10	18	2.5	12	19	3.1	11	22	3.2	13	25	3.5

(a) Obtain the following using SAS® `PROC IML` or R:

(i) $\boldsymbol{X}$ and $\boldsymbol{Y}$

(ii) $\boldsymbol{X}'\boldsymbol{X}$

(iii) $\boldsymbol{X}'\boldsymbol{Y}$

(iv) $\boldsymbol{Y}'\boldsymbol{Y}$

(v) $|\boldsymbol{X}'\boldsymbol{X}|$

(vi) $(\boldsymbol{X}'\boldsymbol{X})^{-1}$

(vii) $\widehat{\boldsymbol{\beta}}$

(viii) $\widehat{\boldsymbol{Y}}$ and $\boldsymbol{e}$

(ix) s^2

(x) $\boldsymbol{s^2}(\widehat{\boldsymbol{\beta}})$

Submit *both* your commands (`PROC IML` or R) and the resulting output.

(b) Write out the estimated regression equation. Using only your results from part (a) and a calculator, compute an ANOVA table that includes the corrected total, regression, and residual sources of variation. Also compute the coefficient of determination.

(You will use the same data set on Assignment 4, so please retain a copy of your results, which you may use to complete Assignment 4, provided they are correct.)

3. Perform the following exercises from the textbook, Section 4.8:

   (a) Exercise 4.15

   (b) Exercise 4.16 [Note that the result of Exercise 4.16 shows that formula (4.30) reduces to the simple linear regression formula given in lecture for $E(\text{MS(Regr)})$.]

4. The *sample correlation* between $\widehat{Y}$ and Y is defined as

$$\widehat{\rho} \;=\; \frac{\sum_{i=1}^{n}\left(\widehat{Y}_i - \sum_{k=1}^{n}\widehat{Y}_k/n\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(\widehat{Y}_i - \sum_{k=1}^{n}\widehat{Y}_k/n\right)^2 \cdot \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}$$

assuming the denominator is nonzero. Complete the following steps to show that, for a multiple linear regression with intercept, $\widehat{\rho} = \sqrt{R^2}$.

Some of these statements may have been partially derived during lecture, but you should write out the full derivation (adding any steps that may have been omitted).

   (a) Show that $\sum_{k=1}^{n}\widehat{Y}_k/n = \overline{Y}$. [Hint: The result from Exercise 4.15(a) may be useful.]

   (b) Write $\widehat{\rho}$ entirely in terms of $\widehat{\boldsymbol{Y}} - \mathbf{1}\overline{Y}$ and $\boldsymbol{Y} - \mathbf{1}\overline{Y}$.

   (c) Show that $\widehat{\boldsymbol{Y}} - \mathbf{1}\overline{Y} = (\boldsymbol{P} - \boldsymbol{J}/n)\boldsymbol{Y}$ and $\boldsymbol{Y} - \mathbf{1}\overline{Y} = (\boldsymbol{I} - \boldsymbol{J}/n)\boldsymbol{Y}$.

   (d) Show that $(\boldsymbol{P} - \boldsymbol{J}/n)(\boldsymbol{I} - \boldsymbol{J}/n) = (\boldsymbol{P} - \boldsymbol{J}/n)$. [Hint: A result from Exercise 4.15 may be useful.]

   (e) Using the previous parts, show that

$$\widehat{\rho}^2 \;=\; \frac{\boldsymbol{Y}'(\boldsymbol{P} - \boldsymbol{J}/n)\boldsymbol{Y}}{\boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{J}/n)\boldsymbol{Y}} \;=\; R^2.$$

   (f) Show that $(\widehat{\boldsymbol{Y}} - \mathbf{1}\overline{Y})'(\boldsymbol{Y} - \mathbf{1}\overline{Y}) \geq 0$, and therefore that $\widehat{\rho} \geq 0$.

5. Complete the following steps to show that

$$E(\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}) \;=\; \text{tr}(\boldsymbol{A}\boldsymbol{V}_y) + \boldsymbol{\mu}_y'\boldsymbol{A}\boldsymbol{\mu}_y$$

where $\boldsymbol{A}$ is an $n \times n$ constant matrix, $\boldsymbol{Y}$ is an $n \times 1$ random vector, $\boldsymbol{\mu}_y = E(\boldsymbol{Y})$, and $\boldsymbol{V}_y = V(\boldsymbol{Y})$.

   (a) Show that $E(\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}) = \text{tr}\left(\boldsymbol{A}\,E(\boldsymbol{Y}\boldsymbol{Y}')\right)$. [Hint: What is the trace of a scalar? Properties of the trace and the matrix expected value might be useful.]

   (b) Complete the derivation, with the aid of part (a) and Problem 1.