Simple Linear Regression – Matrix Form

Q.1. A foam beverage insulator (beer hugger) manufacturer produces their product for firms that want their logo on beer huggers for marketing purposes. The firm's cost analyst wants to estimate their cost function. She interprets β_0 as the fixed cost of a production run, and β_1 as the unit variable cost (or marginal cost). Based on n = 5 production runs she observes the following pairs (X_i, Y_i) where X_i is the number of beer huggers produced in the i^{th} production run (in 1000s), and Y_i was the total cost of the run (in $1000).

i	X_i	Y_i
1	3.00	4.00
2	5.00	6.50
3	4.00	5.00
4	6.00	7.00
5	7.00	7.50

Obtain the following matrices and vectors: $\mathbf{X}, \quad \mathbf{Y}, \quad \mathbf{X'X}, \quad \mathbf{X'Y}, \quad (\mathbf{X'X})^{-1}, \quad \hat{\boldsymbol{\beta}} \quad \hat{\mathbf{Y}}, \quad \mathbf{e}$

Q.2. Give the expeccted values of $SS\mu = \mathbf{Y'PY}$ and $SSE = \mathbf{Y'}(\mathbf{I} - \mathbf{P})\mathbf{Y}$

Q.3. For the matrix form of simple linear regression:

p.3.a. Derive the least squares estimator of $\boldsymbol{\beta}$

p.3.b. Give the mean vector and variance-covariance matrix for the estimator in p.3.a.For

Q.4. For the matrix form of simple linear regression:

p.4.a. Write $\hat{\mathbf{Y}}$ and $\mathbf{e}$ as linear functions of $\mathbf{Y}$

p.4.b. Show that $\mathbf{P} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$ is symmetric and idempotent

p.4.c. Use p.4.a. and p.4.b. to show that $\sum_{i=1}^{n} \hat{Y}_i e_i = \hat{\mathbf{Y}}'\mathbf{e} = 0$

Q.5. An engineer is interested in the relationship between steel thickness (X) and its breaking strength (Y). She obtains the following matrices from a matrix computer package:

$$\mathbf{X'X} = \begin{bmatrix} 12 & 60 \\ 60 & 360 \end{bmatrix} \qquad \mathbf{X'Y} = \begin{bmatrix} 120 \\ 800 \end{bmatrix} \qquad \mathbf{Y'}(\mathbf{I} - \mathbf{P})\mathbf{Y} = 20 \qquad \mathbf{Y'}\left(\mathbf{P} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y} = 250$$

p.5.a. Compute $\hat{\boldsymbol{\beta}}$ and $\hat{V}\left\{\hat{\boldsymbol{\beta}}\right\}$

p.5.b. Give a 95% Confidence Interval for β_1

p.5.c. Test $H_0: \beta_1 = 0 \quad H_A: \beta_1 \neq 0$ at α = 0.05 significance level.

Q.6. Write out $\frac{SS\mu}{\sigma^2}$ and $\frac{SSE}{\sigma^2}$ as quadratic forms involving **Y**. Obtain the sampling distributions for each quantity and show they are independent.

Q.7. Write **e** as a linear function of **Y**, and use that to derive the mean vector and covariance matrix of **e**. Are the residuals uncorrelated?

Q.8. For SLR in matrix form, derive the least squares estimate for **β**

Q.9. For SLR in matrix form, obtain $SS\text{Model} = \sum_{i=1}^{n}\left(\hat{Y}_i\right)^2$ and $SSE = \sum_{i=1}^{n} e_i^2$ in matrix form (quadratic forms in **Y**). Obtain the distributions of the two sum of squares divided by σ^2 (be specific with regard to their family of distributions, degrees of freedom, and non-centrality parameters).

Q.10. For SLR in matrix form: Derive the Covariance of the vector of sample means and the vector of fitted values:

$$\overline{\mathbf{Y}} = \overline{Y}1_n = \frac{1}{n}\mathbf{J}_n\mathbf{Y} \qquad \hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'Y} = \mathbf{PY}$$

Q.11. Consider the simple linear regression model:

$$Y_i = \beta_0 + \beta_1\left(X_i - \overline{X}\right) + \varepsilon_i \quad i = 1,...,n \quad \varepsilon_i \sim NID\left(0, \sigma^2\right)$$

p.11.a. Give the corresponding **X** matrix **Y** vector, and **β** (symbolically):

p.11.b. Give the following matrices and vectors (symbolically): $\mathbf{X'X}, \quad \left(\mathbf{X'X}\right)^{-1} \quad \mathbf{X'Y} \quad \hat{\boldsymbol{\beta}}$

Q.12. Consider the MODEL Sum of Squares for Model 2. $SS(\text{Model}) = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} = \mathbf{Y'AY}$

p.12.a.Give the defining matrix **A** and show that it is idempotent.

p.12.b. Derive the degrees of freedom for SS(Model)

p.12.c. Derive the non-centrality parameter for the non-central chi-square distribution of SS(Model)/σ^2

p.12.d. Show that SS(Model)/σ^2 and SS(Residual)/σ^2 are independent (state SS(Residual) as quadratic form)

Q.13. A simple linear regression was fit, relating the modulus of a tire (Y) to the amount of weeks (X) heated at 125^0, with results given below:

Weeks (X)	Modulus(Y)
0	2.3
1	4.2
2	5.2
4	5.9
6	6.3
15	7.2

p.13.a. For SLR in matrix form, obtain **X**, **Y**, **X'X**, $\mathbf{(X'X)^{-1}}$, and **X'Y**

p.13.b. Compute $\hat{\boldsymbol{\beta}}, \quad \hat{\mathbf{Y}}, \quad \mathbf{e}, \quad MSE, \quad \hat{V}\left\{\hat{\boldsymbol{\beta}}\right\}$

p.13.c. Compute a 95% Confidence Interval for the change in the mean of modulus as weeks (X) increases by 1

Q.14. A simple regression model is fit, with 1 predictor and an intercept. Define the projection matrix as:
$\mathbf{P} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'}$.

p.14.a. Show that **P** is symmetric and idempotent. (Hint: $\mathbf{(X'X)^{-1}}$ is symmetric)

p.14.b. What does $\sum_{j=1}^{n} P_{ij}$ equal? Where P_{ij} is the element of **P** in the i^{th} row and j^{th} column

p.14.c. What does $\sum_{j=1}^{n} P_{ij}^2$ equal?

p.14.d. What does $\sum_{j=1}^{n}\left(P_{ij} - \overline{P}_{i\bullet}\right)^2$ equal?

Q.15. The total salaries (X, in millions of pounds) and the number of points earned (Y) for the n = 20 English Premier League teams in 1995/6 are used to fit a simple linear regression model. For this problem, we will treat this as a sample from a population of all possible league teams.

Team	Int	WageK	Points(Y)			X'X			X'Y
Arsenal	1	10.1	63			20	170.9		1042
Aston Villa	1	7.7	63			170.9	1745.15		9870.3
Blackburn	1	10.8	61						
Bolton	1	3.3	29						
Chelsea	1	9.2	50						
Coventry	1	5.8	38						
Everton	1	10.1	61						
Leeds	1	10.1	43						
Liverpool	1	13.2	71						
Manchester City	1	6.4	38						
Manchester United	1	13.3	82						
Middlesbrough	1	6.5	43						
Newcastle	1	19.7	78						
Nottingham Forest	1	8.5	58						
Queens Park	1	5	33						
Sheffield	1	6.4	40						
Southampton	1	4.1	38						
Tottenham	1	9.8	61						
West Ham United	1	6.2	51						
Wimbledon	1	4.7	41						

p.15.a. Compute $\left(\mathbf{X'X}\right)^{-1}$ and $\hat{\boldsymbol{\beta}}$

p.15.b. Compute the fitted value and residual for Arsenal.

p.15.c. Given $\overline{Y} = 52.1$ and $\sum_{i=1}^{20}\left(Y_i - \overline{Y}\right)^2 = 4367.8$ Compute SSR and SSE

p.15.d. Obtain a 95% Confidence Interval for β_1

Q.16. For the Analysis of Variance in SLR, with n observations and 1 predictor, complete the following parts.

p.16.a. Write the Regression and Residual sums of squares as quadratic forms.

p.16.b. Derive the distributions of SSR/σ^2 and SSE/σ^2

p.16.c. Show that SSR/σ^2 and SSE/σ^2 are independent

Q.17. The following measurements were obtained at n=5 locations on the earth's surface, where Y=measured gravity, and X= latitude. Consider fitting SLR, containing an intercept. Note this includes the fitted values and residuals from results of p.4.a, and makes use of both scalar model and matrix model.

Obs#	X(Lat)	Y(Grav)	Y-hat	e
1	20	1.10	1.08	0.02
2	30	1.60	1.54	0.06
3	40	1.90	2.00	-0.10
4	50	2.40	2.46	-0.06
5	60	3.00	2.92	0.08

p.17.a. Obtain the following matrices and vectors.

X'X			X'Y
INV(X'X)			Beta-hat

p.17.b. Compute the residual sum of squares, and give its degrees of freedom (based on scalar model).

p.17.c. Compute the regression sum of squares, and give its degrees of freedom (based on scalar model).

p.17.d. Conduct the F-test, testing H_0: $\beta_1 = 0$ vs H_A: $\beta_1 \neq 0$ at $\alpha = 0.05$ significance level.

p.17.d.i. Test Statistic:

p.17.d.ii. Rejection Region: ____

Q.18. In the matrix form of the simple linear regression model, the least squares estimator for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$ where the elements of $\mathbf{X}$ are fixed constants in a controlled laboratory experiment.

p.18.a. Derive $E\left\{\hat{\boldsymbol{\beta}}\right\}$ **Show all work**

p.18.b. Derive $V\left\{\hat{\boldsymbol{\beta}}\right\}$ **Show all work**

Q.19. An experiment is conducted, relating weekly sales for a food delivery (Y) service to the amount of advertising (X) during the week. The results for a sample of n = 6 weeks are given below. Fit SLR in matrix form by filling in the following matrices.

Week	1	2	3	4	5	6
Ad Spend (X)	2	2	4	4	6	6
Sales (Y)	20	30	40	50	70	60

$\mathbf{X} = \begin{bmatrix} \quad \end{bmatrix}$ $\quad \mathbf{Y} = \begin{bmatrix} \quad \end{bmatrix}$ $\quad \mathbf{X'X} = \begin{bmatrix} \quad \end{bmatrix}$ $\quad \mathbf{X'Y} = \begin{bmatrix} \quad \end{bmatrix}$

$(\mathbf{X'X})^{-1} = \begin{bmatrix} \quad \end{bmatrix}$ $\quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} \quad \end{bmatrix}$ $\quad \hat{\mathbf{Y}} = \begin{bmatrix} \quad \end{bmatrix}$ $\quad \mathbf{e} = \begin{bmatrix} \quad \end{bmatrix}$

$SSE = \mathbf{e'e} =$

Q.20. For the simple regression model, with an intercept term, complete the following parts

p.20.a. $\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{e} \quad \hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \quad \overline{\mathbf{Y}} = \frac{1}{n}\mathbf{J}_n\mathbf{Y}$ prove that $\left(\mathbf{Y}-\overline{\mathbf{Y}}\right)'\left(\mathbf{Y}-\overline{\mathbf{Y}}\right) = \left(\hat{\mathbf{Y}}-\overline{\mathbf{Y}}\right)'\left(\hat{\mathbf{Y}}-\overline{\mathbf{Y}}\right) + \mathbf{e'e}$

p.20.b. Derive the sampling distributions of $\hat{\mathbf{Y}}\text{-}\overline{\mathbf{Y}}$ and $\mathbf{e}$

Q.21. Based on Model 2, with $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$:

p.21.a. Derive the sampling distributions of the vectors $\hat{\mathbf{Y}}, \quad \mathbf{e}, \quad \overline{\mathbf{Y}}$

p.21.b. Derive the sampling distributions of $\frac{SSE}{\sigma^2}$ and $\frac{SSR}{\sigma^2}$ and show they are independent.

p.21.c. Give the sampling distribution of $F = MSR/MSE$ under H_0: $\beta_1 = 0$.

Q.22. Consider regression through the origin in matrix form where **1** is an *n*x1 vector of 1[s].

$$\mathbf{Y}=\begin{bmatrix} Y_1 \\ \dots \\ Y_n \end{bmatrix} \quad \mathbf{X}=\begin{bmatrix} X_1 \\ \dots \\ X_n \end{bmatrix} \quad \boldsymbol{\beta}=[\beta_1] \quad \mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \quad E\{\boldsymbol{\varepsilon}\}=\mathbf{0} \quad V\{\boldsymbol{\varepsilon}\}=\sigma^2\mathbf{I} \quad \hat{\boldsymbol{\beta}}=(\mathbf{X'X})^{-1}\mathbf{X'Y} \quad \tilde{\boldsymbol{\beta}}=(\mathbf{1'X})^{-1}\mathbf{1'Y}$$

Derive the means, variances, and covariance of

$$\hat{\beta} \text{ and } \tilde{\beta} \quad \text{where} \quad \mathrm{COV}\{\mathbf{AY},\mathbf{BY}\}=E\{(\mathbf{AY}-\mathbf{A}\boldsymbol{\mu}_{\mathbf{Y}})(\mathbf{BY}-\mathbf{B}\boldsymbol{\mu}_{\mathbf{Y}})'\}$$

Q.23. A simple linear regression (with an intercept) is fit, relating *Y* to *X*. The data are given below.

X	0	0	2	2	4	4
Y	12	8	14	16	23	17

p.23.a. Give the following matrices and vectors. $\mathbf{X}, \quad \mathbf{Y}, \quad \mathbf{X'X}, \quad \mathbf{X'Y}, \quad \mathbf{Y'Y}, \quad (\mathbf{X'X})^{-1}, \quad \hat{\boldsymbol{\beta}}, \quad \hat{\mathbf{Y}}, \quad \mathbf{e}$

p.23.b. Compute the following sums of square: SS Model, SSμ, SS Regression, and SS Error

Q.24. Based on Model 2, with $\hat{\boldsymbol{\beta}}=(\mathbf{X'X})^{-1}\mathbf{X'Y}$:

p.24.a. Derive the sampling distributions of the vectors $\hat{\mathbf{Y}}, \quad \mathbf{e}, \quad \overline{\mathbf{Y}}, \quad \mathbf{Y}-\overline{\mathbf{Y}}$ **Show all work**

p.24.b. Derive the sampling distributions of $\frac{TSS}{\sigma^2}$ and $\frac{SS\mu}{\sigma^2}$ where $SS\mu=n\overline{Y}^2=\mathbf{Y'A_{\boldsymbol{\mu}}Y}$ and show they are independent. **Show all work**

Q.25. A simple linear regression model is fit, where n=143 cats weighing over 2 kilograms were observed. The response variable was heart weight (Y, in grams) and the predictor was body weight (X, in kilograms). The estimated regression coefficient vector and its variance covariance matrix are given below.

$$\hat{\boldsymbol{\beta}}=\begin{bmatrix} 0.12 \\ 3.85 \end{bmatrix} \qquad \hat{V}\{\hat{\boldsymbol{\beta}}\}=\begin{bmatrix} 0.4489 & -0.1561 \\ -0.1561 & 0.0576 \end{bmatrix} \qquad \overline{X}=2.71$$

p.25.a. Obtain the predicted heart weight for a cat weighing 3.0 kilograms.

p.25.b. Compute a 95% Confidence Interval for the mean heart weight for all cats weighing 3 kilograms.

Q.26. A simple linear regression model was fit, relating the number of airplane tires consumed (Y) to the number of landings (X, in 100s) at an airfield over a series of periods. Based on Model 2, the following results were obtained.

X'X			X'Y		Y'Y
20	95.91		1051		59867
95.91	502.7151		5439.48		
INV(X'X)			Beta-hat		
0.58758	-0.11210		7.7761		
-0.11210	0.02338		9.3366		

p.26.a. How many periods were observed by the researchers? ___________

p.26.b. How many total aircraft landings were there during the study? ________________

p.26.c. How many tires were consumed over the study? ____________________

p.26.d. Compute: $\mathbf{Y'IY}$ $\qquad \mathbf{Y'PY}$ $\qquad \mathbf{Y'}\frac{1}{n}\mathbf{JY}$

p.26.e. Compute $SSE = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$ and $SSR = \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2$

p.26.f. Compute $\hat{V}\left\{\hat{\boldsymbol{\beta}}\right\}$

Q.27. A study in Beijing, China related heights (Y, cm) to mean foot length (X, cm) in a sample of n = 174 male college students. The sample means, standard deviations and correlations are given below. Complete the following table for the simple linear regression relating Height (Y) to mean foot length (X).

$$\overline{X} = 24.93 \quad \overline{Y} = 177.65 \quad s_X = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}} = 0.903 \quad s_Y = 4.810 \quad r_{XY} = 0.483$$

Regression Statistics					
R Square					
Residual Std Error					
Observations					
ANOVA					
	df	*SS*	*MS*	*F*	*F(.05)*
Regression					
Residual					
Total					
	Coefficients	*andard Err*	*t Stat*	*Lower*	*Upper*
Intercept					
X					