

STA 2101/442 Assignment Four¹

One version of the general linear model with fixed effects is $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where

- X is an $n \times p$ matrix of known constants with $n > p$ and the columns of X linearly independent.
- $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown constants.
- $\boldsymbol{\epsilon}$ is an $n \times 1$ random vector with $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{\epsilon}) = \sigma^2 I_n$.
- $\sigma^2 > 0$ is an unknown constant.

The least-squares estimate of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{y}$, the vector of predicted y values is $\hat{\mathbf{y}} = X\hat{\boldsymbol{\beta}}$, and the vector of residuals is $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$.

1. Give the dimensions (number of rows and number of columns) of the following matrices.

- (a) \mathbf{y}
- (b) $\boldsymbol{\beta}$
- (c) $X\boldsymbol{\beta}$
- (d) $(X^\top X)^{-1}$
- (e) $\hat{\boldsymbol{\beta}}$
- (f) $\hat{\mathbf{y}}$
- (g) \mathbf{e}
- (h) $\mathbf{e}^\top \mathbf{e}$
- (i) $\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top$
- (j) $X^\top \mathbf{e}$

2. Show $X^\top \mathbf{e} = \mathbf{0}$.

3. Why does $X^\top \mathbf{e} = \mathbf{0}$ tell you that if a regression model has an intercept, the residuals must add up to zero?

4. Let $\mathcal{S} = (\mathbf{y} - X\boldsymbol{\beta})^\top (\mathbf{y} - X\boldsymbol{\beta})$. Note that this is the sum of squared differences between the y_i observations and their expected values, and the β_j values that minimize it are the least squares estimates.

- (a) Show that $\mathcal{S} = (\mathbf{y} - X\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - X\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top (X^\top X)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Hint: Add and subtract $\hat{\mathbf{y}}$.
- (b) Why does this imply that the minimum of $\mathcal{S}(\boldsymbol{\beta})$ occurs at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$?
- (c) The columns of X are linearly independent. Why does linear independence guarantee that the minimum is unique?
- (d) Isn't it nice to be able to do this without calculus?

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

5. The “hat” matrix is $H = X(X^\top X)^{-1}X^\top$.
- What are the dimensions (number of rows and columns) of H ?
 - Show that H is symmetric.
 - Show that H is idempotent, meaning $H = H^2$.
 - Show that $\hat{\mathbf{y}} = H\mathbf{y}$.
 - Using $\text{tr}(AB) = \text{tr}(BA)$, find $\text{tr}(H)$. (That’s the trace, sum of the diagonal elements.)
 - Show that $I - H$ is symmetric.
 - Show that $I - H$ is idempotent.
 - Show that $X^\top(I - H) = \mathbf{0}$.
 - Show that $\mathbf{e} = (I - H)\mathbf{y}$.
 - Show that $\mathbf{e} = (I - H)\boldsymbol{\epsilon}$.
 - Find $\text{tr}(I - H)$.
6. In simple regression through the origin, there is one explanatory variable and no intercept. The model is $y_i = \beta_1 x_i + \epsilon_i$.
- Find the least squares estimator of β_1 with calculus.
 - What is the X matrix?
 - What is $X^\top X$?
 - What is $X^\top \mathbf{y}$?
 - What is $(X^\top X)^{-1}$?
 - What is $\hat{\beta}_1 = (X^\top X)^{-1}X^\top \mathbf{y}$? Compare this with your answer to [6a](#).
7. There can even be a regression model with an intercept and no explanatory variables. In this case the model would be $y_i = \beta_0 + \epsilon_i$.
- Find the least squares estimator of β_0 with calculus. What’s a least-squares estimator again? Find the parameter value(s) that make the y_i observations as close as possible to their expected values.
 - What is the X matrix?
 - What is $X^\top X$?
 - What is $X^\top \mathbf{y}$?
 - What is $(X^\top X)^{-1}$?
 - What is $\hat{\beta}_0 = (X^\top X)^{-1}X^\top \mathbf{y}$? Compare this with your answer to [7a](#).

8. The set of vectors $\mathcal{V} = \{\mathbf{v} = X\mathbf{a} : \mathbf{a} \in \mathbb{R}^{k+1}\}$ is the subset of \mathbb{R}^n consisting of linear combinations of the columns of X . That is, \mathcal{V} is the space *spanned* by the columns of X . The least squares estimator $\hat{\boldsymbol{\beta}} = (X^\top X)^{-1}X^\top \mathbf{y}$ was obtained by minimizing $(\mathbf{y} - X\mathbf{a})^\top(\mathbf{y} - X\mathbf{a})$ over all $\mathbf{a} \in \mathbb{R}^{k+1}$. Thus, $\hat{\mathbf{y}} = X\hat{\boldsymbol{\beta}}$ is the point in \mathcal{V} that is *closest* to the data vector \mathbf{y} . Geometrically, $\hat{\mathbf{y}}$ is the *projection* (shadow) of \mathbf{y} onto \mathcal{V} . The hat matrix H is a *projection matrix*. It projects the image on any point in \mathbb{R}^n onto \mathcal{V} . Now we will test out several consequences of this idea.
- The shadow of a point already in \mathcal{V} should be right at the point itself. Show that if $\mathbf{v} \in \mathcal{V}$, then $H\mathbf{v} = \mathbf{v}$.
 - The vector of differences $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ should be perpendicular (at right angles) to each and every basis vector of \mathcal{V} . How is this related to Question 2?
 - Show that the vector of residuals \mathbf{e} is perpendicular to any $\mathbf{v} \in \mathcal{V}$.
 - If \mathbf{u} is a general point in \mathbb{R}^n show that $(I - H)\mathbf{u}$ is perpendicular to any $\mathbf{v} \in \mathcal{V}$ (this includes \mathbf{e}).
9. The linear regression model with intercept can be written in scalar form as $y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_{p-1} x_{i,p-1} + \epsilon_i$. Defining $SSTO = \sum_{i=1}^n (y_i - \bar{y})^2$, $SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ and $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$, show $SSTO = SSE + SSR$.
10. In this question you will show that $MSE = SSE/(n - p)$ is an unbiased estimator of the error variance σ^2 , using $tr(AB) = tr(BA)$. Start with the fact that the trace of a 1×1 matrix is just the matrix itself, like this: $E(\mathbf{e}^\top \mathbf{e}) = E(tr(\mathbf{e}^\top \mathbf{e})) \dots$ Please continue; Question 5j is helpful.
11. Recall that if $\mathbf{w} \sim N_p(\boldsymbol{\mu}, \Sigma)$, then $A\mathbf{w} + \mathbf{c} \sim N_q(A\boldsymbol{\mu} + \mathbf{c}, A\Sigma A^\top)$. If $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 I_n)$ in the general linear model, give the distributions of the following random vectors. Simplify as much as possible.
- \mathbf{y}
 - $\boldsymbol{\beta}$
 - $\hat{\boldsymbol{\beta}}$
 - $\hat{\mathbf{y}}$
 - \mathbf{e}
 - $X^\top \mathbf{e}$

12. Let $\mathbf{X} = (X_1, X_2, X_3)^\top$ be multivariate normal with

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

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

13. Let X_1 be $\text{Normal}(\mu_1, \sigma_1^2)$. Let X_2 be $\text{Normal}(\mu_2, \sigma_2^2)$, 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? Do it the easy way.

14. High School History classes from across Ontario are randomly assigned to either a discovery-oriented or a memory-oriented curriculum in Canadian history. At the end of the year, the students are given a standardized test and the median score of each class is recorded. Please consider a regression model with these variables:

X_1 Equals 1 if the class uses the discovery-oriented curriculum, and equals 0 if the class uses the memory-oriented curriculum.

X_2 Average parents' education for the classroom.

X_3 Average family income for the classroom.

X_4 Number of university History courses taken by the teacher.

X_5 Teacher's final cumulative university grade point average.

Y Class median score on the standardized history test.

The full regression model (as opposed to the reduced models for various null hypotheses) implies

$$E[Y|X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5.$$

For each question below, please give

- The null hypothesis in terms of β values.
 - $E[Y|X]$ for the reduced model you would use to answer the question. Don't re-number the variables.
- (a) If you control for parents' education and income and for teacher's university background, does curriculum type affect test scores? (And why is it okay to use the word "affect?")
- (b) Controlling for parents' education and income and for curriculum type, is teacher's university background (two variables) related to their students' test performance?
- (c) Controlling for teacher's university background and for curriculum type, are parents' education and family income (considered simultaneously) related to students' test performance?
- (d) Controlling for curriculum type, teacher's university background and parents' education, is parents' income related to students' test performance?
- (e) Here is one final question. Assuming that X_1, \dots, X_5 are random variables (and I hope you agree that they are), would you expect X_1 to be related to the other explanatory variables? Would you expect the other explanatory variables to be related to each other?

15. The U.S. Census Bureau divides the United States into small pieces called census tracts; lots of information is collected about each census tract. The census tracts are grouped into four geographic regions: Northeast, North Central, South and West. In one study, the cases were census tracts, the explanatory variables were Region and average income, and the response variable was crime rate, defined as the number of reported serious crimes in a census tract, divided by the number of people in the census tract.
- (a) Write $E(Y|x)$ for a regression model with parallel regression lines. You do not have to say how your dummy variables are defined. You will do that in the next part.
 - (b) Make a table showing how your dummy variables are set up. There should be one row for each region, and a column for each dummy variable. Add a wider column on the right, in which you show $E(Y|x)$. Note that the *symbols* for your dummy variables will not appear in this column. There are examples of this format in the lecture slides and the text for each region.
 - (c) For each of the following questions, give the null hypothesis in terms of the β parameters of your regression model. We are not doing one-tailed tests, regardless of how the question is phrased.
 - i. Controlling for income, does average crime rate differ by geographic region?
 - ii. Controlling for income, is average crime rate different in the Northeast and North Central regions?
 - iii. Controlling for income, is average crime rate different in the Northeast and Western regions?
 - iv. Controlling for income, is the crime rate in the South more than the average of the other three regions?
 - v. Controlling for income, is the average crime rate in the Northeast and North Central regions different from the average of the South and West?
 - vi. Controlling for geographic region, is crime rate connected to income?

This assignment was prepared by [Jerry Brunner](#), Department of Statistics, University of Toronto. It is licensed under a [Creative Commons Attribution - ShareAlike 3.0 Unported License](#). Use any part of it as you like and share the result freely. The L^AT_EX source code is available from the course website: <http://www.utstat.toronto.edu/~brunner/oldclass/appliedf16>