Econometrics I

Problem Set 1

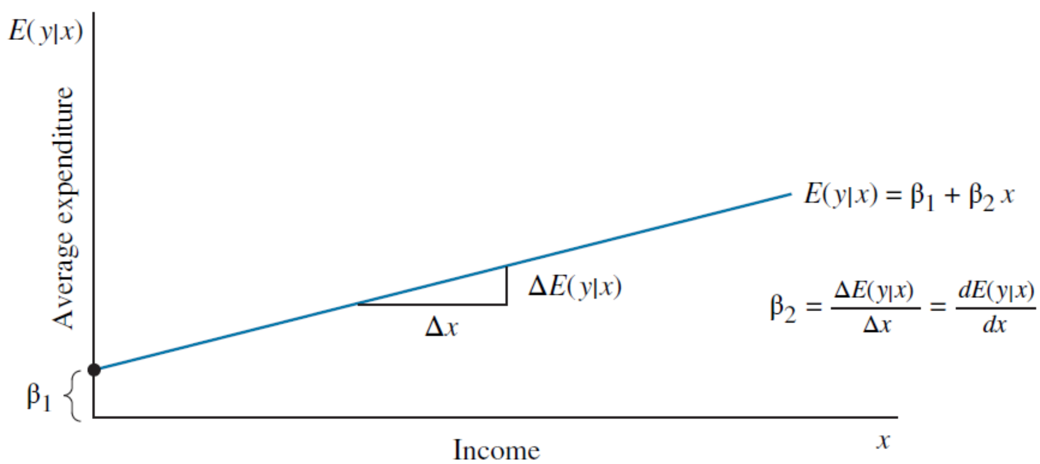
Matías Cabello
University Halle-Wittenberg
Chair of Economic Growth and Development

October 16, 2025

1. **Deriving the univariate OLS estimator**

   (a) Draw $\mathbb{E}[y_i|x] = \beta_1 + \beta_2 x_i$ on a plane and explain what β_1 and β_2 are.

Solution: β_1 stands for the intercept (the value when $x = 0$) and β_2 stands for the slope (how much y changes when x changes).



   (b) Explain the difference between population and sample regression.

Solution: The *sample regression* or *fitted values* stand for the regression line $\hat{y}_i + \hat{\beta}_1 + \hat{\beta}_2 x_i$ estimated with a given sample (that was taken out of the population). The *population regression*, by contrast, is the expected value $\mathbb{E}[y_i|x_i] = \beta_1 + \beta_2 x_i$ of the *data generating process* $y_i = \beta_1 + \beta_2 x_i + u_i$.

(c) Show that the slope of the OLS estimate is

$$\hat{\beta}_2 = \frac{\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n} \left(\sum x_i\right)^2} \tag{OLS-sums}$$

Solution: The simple linear regression model is:

$$y_i = \beta_1 + \beta_2 x_i + u_i \tag{1}$$

For a sample of n observations, the fitted regression line is:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i \tag{2}$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimators of the unknown parameters β_1 and β_2. The least squares principle chooses the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ to minimize the sum of squared residuals (SSR):

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2 \tag{3}$$

To find the values of $\hat{\beta}_1$ and $\hat{\beta}_2$ that minimize SSR, we use calculus. We take the partial derivatives of SSR with respect to $\hat{\beta}_1$ and $\hat{\beta}_2$ and set them equal to zero. This gives us the "first order conditions."

The first order conditions are:

$$\frac{\partial SSR}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0 \tag{4}$$

$$\frac{\partial SSR}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^{n} \left[x_i (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) \right] = 0 \tag{5}$$

Dividing both equations by -2 and simplifying, we get the "normal equations":

$$\sum y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum x_i \tag{6}$$

$$\sum x_i y_i = \hat{\beta}_1 \sum x_i + \hat{\beta}_2 \sum x_i^2 \tag{7}$$

We now have two linear equations with two unknowns, $\hat{\beta}_1$ and $\hat{\beta}_2$. To solve for $\hat{\beta}_2$, multiply equation (6) by $\sum x_i$ and equation (7) by n:

$$\left(\sum x_i\right)\left(\sum y_i\right) = n\hat{\beta}_1\left(\sum x_i\right) + \hat{\beta}_2\left(\sum x_i\right)^2 \tag{8}$$

$$n\sum x_i y_i = n\hat{\beta}_1\left(\sum x_i\right) + \hat{\beta}_2\, n\sum x_i^2 \tag{9}$$

Now subtract equation (8) from equation (9):

$$n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right) = \hat{\beta}_2\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right] \tag{10}$$

Solving for $\hat{\beta}_2$, we obtain the formula for the least squares estimator of the slope:

$$\hat{\beta}_2 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \tag{11}$$

For completeness, let use derive $\hat{\beta}_1$ too. Once we have $\hat{\beta}_2$, we can solve for $\hat{\beta}_1$ using the first normal equation (6). Dividing (6) by n:

$$\bar{y} = \hat{\beta}_1 + \hat{\beta}_2\bar{x} \tag{12}$$

Rearranging, we get the formula for the least squares intercept estimator:

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x} \tag{13}$$

(d) Show that the slope of the OLS estimate can also be written as

$$\hat{\beta}_2 = \frac{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\mathrm{cov}(x, y)}{\mathrm{var}(x)}. \tag{OLS-cov}$$

Use the covariance formula

$$\mathrm{cov}(x, y) = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

and the fact that $\bar{x} = \frac{1}{n}\sum_i x$.

Solution: We basically need to proof two claims.

First claim: $\sum(x_i - \bar{x})^2 = \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2$

Proof:

$$\begin{aligned}
\sum_i (x_i - \bar{x})^2 &= \sum_i (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\
&= \sum_i x_i^2 - 2\bar{x}\sum_i x_i + n\bar{x}^2 \\
&= \sum_i x_i^2 - 2\bar{x}\bar{x}n + n\bar{x}^2 \\
&= \sum_i x_i^2 - n\bar{x}^2 \\
&= \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2
\end{aligned}$$

Second claim: $\sum(x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i$

Proof:

$$\begin{aligned}
\sum (x_i - \bar{x})(y_i - \bar{y}) &= \sum (x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y}) \\
&= \sum x_i y_i - \bar{y}\sum x_i - \bar{x}\sum y_i + n\bar{x}\bar{y} \\
&= \sum x_i y_i - \bar{y}n\bar{x} - \bar{x}n\bar{y} + n\bar{x}\bar{y} \\
&= \sum x_i y_i - \bar{y}n\bar{x} \\
&= \sum x_i y_i - n\tfrac{1}{n}\sum_i y_i \tfrac{1}{n}\sum_i x_i \\
&= \sum x_i y_i - \tfrac{1}{n}\sum_i y_i \sum_i x_i
\end{aligned}$$

(e) What are the meanings of $\text{var}(x)$ and $\text{cov}(x, y)$?

Solution: The variance formula is

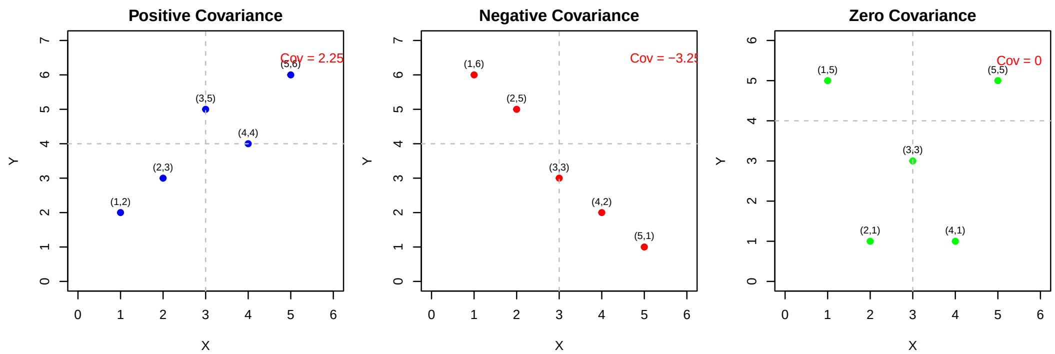
$$\text{var}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

It captures the dispersion of x_i around its mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. For example, the variance of $x = (0, 1, 2, 3, 4)$ is 2, that of $x = (-1, 1, 2, 3, 5)$ is 4, that of $x = (3, 3, 3, 3, 3)$ is is 0.

The covariance formula is

$$\text{cov}(x, y) = \text{cov}(y, x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).$$

As shown in the figure below, it captures how the observations are distributed across the four quadrants that result from dividing the each plot by the means of x and y.



Note, by looking at the formula:

1. If datapoints are mostly in quadrants upper-right and lower-left, then the sum will contain mostly positive elements (resulting from a multiplication of positive with positive, or of negative with negative).
2. If datapoints are mostly in quadrants lower-right and upper-left, then the sum will contain mostly positive elements (resulting from a multiplication of positive with negative, or of negative with positive).
3. If points are scattered randomly in all quadrants, then the covariance is zero

(f) What is the only requirement for the OLS method to yield a result for $\hat{\beta}_2$?

Solution: The only requirement is $\text{var}(x) \neq 0$. In other words, the explanatory variable x cannot be a constant; it needs to variate. If $\text{var}(x) = 0$, then $\hat{\beta}_2$ is undefined.

2. **Deriving the multivariate OLS estimator**

(a) Consider now the multivariate regression

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_{2,i} + \ldots + \hat{\beta}_k x_{k,i} + \hat{u}_i \qquad \text{for all } i = 1, ..., n$$

with k different regressors (the first of which is just the constant). Show that it can be expressed in matrix form as

$$\underset{(n\times 1)}{y} = \underset{(n\times k)}{X}\,\underset{(k\times 1)}{\hat{\beta}} + \underset{(n\times 1)}{\hat{u}} \qquad \text{(m-form)}$$

Solution: when we have k explanatory variables (also known as *regressors*) and n observations, the model can be written as:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_{2,i} + \ldots + \hat{\beta}_k x_{k,i} + \hat{u}_i \qquad \text{for all } i = 1, ..., n \tag{14}$$

Since this relationship holds for each observation i (e.g., each year in our example), we can express it using vector notation:

$$\underset{(n\times 1)}{y} = \hat{\beta}_1 \underset{(n\times 1)}{\mathbb{1}} + \hat{\beta}_2 \underset{(n\times 1)}{x_2} + \hat{\beta}_3 \underset{(n\times 1)}{x_3} + \ldots + \hat{\beta}_k \underset{(n\times 1)}{x_k} + \underset{(n\times 1)}{\hat{u}}$$

Where:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \hat{\beta}_1 + \begin{bmatrix} x_{2,1} \\ x_{2,2} \\ \vdots \\ x_{2,n} \end{bmatrix} \hat{\beta}_2 + \ldots + \begin{bmatrix} x_{k,1} \\ x_{k,2} \\ \vdots \\ x_{k,n} \end{bmatrix} \hat{\beta}_k + \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix}$$

The explanatory variables can be grouped into a single $n \times k$ matrix, denoted as X. Similarly, the regression coefficients are grouped into a $k \times 1$ column vector $\hat{\beta}$. Thus, we can write:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & x_{2,1} & x_{3,1} & \cdots & x_{k,1} \\ 1 & x_{2,2} & x_{3,2} & \cdots & x_{k,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2,n} & x_{3,n} & \cdots & x_{k,n} \end{bmatrix}}_{X} \cdot \underbrace{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}}_{\hat{\beta}} + \underbrace{\begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix}}_{\hat{u}}$$

This leads to the matrix representation of the multivariate model:

$$\underset{(n\times 1)}{y} = \underset{(n\times k)}{X} \underset{(k\times 1)}{\hat{\beta}} + \underset{(n\times 1)}{\hat{u}} \tag{E}$$

(b) Show that the sum of squared residuals can be expressed as

$$
\begin{aligned}
SSR(\hat{\beta}, y, X) = \sum_{i=1}^{n} \hat{u}_i^2 = \hat{u}'\hat{u} &= (y - X\hat{\beta})'(y - X\hat{\beta}) \\
&= y'y - 2y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}
\end{aligned}
\tag{SSR}
$$

Solution:

Despite being a scalar, the sum of squared residuals also has a matrix representation. This is:

$$
\begin{aligned}
\sum_{i=1}^{n} \hat{u}_i^2 = \hat{u}'\hat{u} &= (y - X\hat{\beta})'(y - X\hat{\beta}) \\
&= \left(y' - (X\hat{\beta})'\right)\left(y - X\hat{\beta}\right) && \text{using } (A+B)' = A' + B' \\
&= \left(y' - \hat{\beta}'X'\right)\left(y - X\hat{\beta}\right) && \text{using } (AB)' = B'A' \\
&= y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} \\
&= y'y - 2y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}
\end{aligned}
$$

The last equality makes use of $\hat{\beta}'X'y = (\hat{\beta}'X'y)' = y'X\hat{\beta}$, which holds because a scalar is equal to its transpose.

(c) Show that the derivative of (SSR) takes the form

$$
\begin{aligned}
\frac{\partial \hat{u}'\hat{u}}{\partial \hat{\beta}} &= -2\frac{\partial y'X\hat{\beta}}{\partial \hat{\beta}} + \frac{\partial \hat{\beta}'X'X\hat{\beta}}{\partial \hat{\beta}} \\
&= -2X'y + 2X'X\hat{\beta} = \underset{k\times 1}{0}
\end{aligned}
$$

and can be solved for the OLS estimator

$$
\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = (X'X)^{-1}X'y. \tag{$\hat{\beta}$-OLS}
$$

Make use of the differentiation rules stated in the appendix.

Solution: The idea now is to find a vector $\hat{\beta}$ that contains the coefficients which minimize this expression. That is, we seek

$$\hat{\beta} = \arg\min_{\hat{\beta}}[\hat{u}'\hat{u}] = \arg\min_{\hat{\beta}}\left[y'y - 2y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}\right]$$

The minimization can be achieved by differentiating $\hat{u}'\hat{u}$ with respect to each of the k coefficients contained in $\hat{\beta}$, setting the derivatives to zero, and solving the resulting system of k equations with k unknowns. Fortunately, this first-order condition can also be derived using a couple of matrix differentiation rules.

First, note that the derivative of $y'y$ with respect to $\hat{\beta}$ must be zero, so that

$$\frac{\partial \hat{u}'\hat{u}}{\partial \hat{\beta}} = -2\frac{\partial y'X\hat{\beta}}{\partial \hat{\beta}} + \frac{\partial \hat{\beta}'X'X\hat{\beta}}{\partial \hat{\beta}} = 0.$$

Then, if we consider the $1 \times k$ vector, $a = y'X$, we can apply the general rule $\frac{\partial a_{(1\times k)} z_{(k\times 1)}}{\partial z_{(k\times 1)}} = a'(1 \times k)$ (the proof is referred to the appendix), which is equivalent to $-2X'y$ for the first term. The second term involves a slightly more complex derivative. First, we will note that the term $X'X$ is a symmetric matrix:

$$\begin{aligned}
X'X &= \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k,1} & x_{k,2} & \cdots & x_{k,n} \end{bmatrix} \begin{bmatrix} x_{1,1} & x_{2,1} & \cdots & x_{k,1} \\ x_{1,2} & x_{2,2} & \cdots & x_{k,2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,n} & x_{2,n} & \cdots & x_{k,n} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n x_{1,i}^2 & \sum_{i=1}^n x_{1,i}x_{2,i} & \cdots & \sum_{i=1}^n x_{1,i}x_{k,i} \\ \sum_{i=1}^n x_{2,i}x_{1,i} & \sum_{i=1}^n x_{2,i}^2 & \cdots & \sum_{i=1}^n x_{2,i}x_{k,i} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{k,i}x_{1,i} & \sum_{i=1}^n x_{k,i}x_{2,i} & \cdots & \sum_{i=1}^n x_{k,i}^2 \end{bmatrix}
\end{aligned}$$

Since $X'X$ is symmetric (note that $\sum_{i=1}^n x_{1,i}x_{2,i} = \sum_{i=1}^n x_{2,i}x_{1,i}$, etc.), we can use the rule $\frac{\partial z'Az}{\partial z} = 2Az$, valid whenever $A_{k\times k}$ is symmetric (see proof in the appendix).

With the derivative solved, it only remains to solve for the vector $\hat{\beta}$ from

$$\frac{\partial \hat{u}'\hat{u}}{\partial \hat{\beta}} = -2X'y + 2X'X\hat{\beta} = \underset{k\times 1}{0} \tag{15}$$

to obtain the vector of estimated coefficients resulting from the system of k equations represented by (15). The solution is:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = (X'X)^{-1}X'y \tag{16}$$

(d) What condition is required for the estimates vector (16) to exist?

Solution: In analogy to the requirement $\text{var}(x) \neq 0$, here we need $rank(X) = k$, which means that each column (each explanatory variable) is an independent linear condition of the remaining columns (of the remaining regressor). Note that, in the univariate case, $\text{var}(x) = 0$ would imply that x is a linear representation of the constant, violating $rank(X) = k$.

Matrix math appendix

Basic Properties

- In general, $\mathbf{AB} \neq \mathbf{BA}$.
- Let $\mathbf{A} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{bmatrix}$, then $\mathbf{A}' = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}$

  and $\mathbf{A}'\mathbf{A} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{bmatrix} = \begin{bmatrix} a_1^2 + a_2^2 + a_3^2 & a_1b_1 + a_2b_2 + a_3b_3 \\ a_1b_1 + a_2b_2 + a_3b_3 & b_1^2 + b_2^2 + b_3^2 \end{bmatrix}$
- Let vector $a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$ and vector $b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$, then $a' \cdot b = a_1b_1 + a_2b_2 + a_3b_3$.
- Inverses: $\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{A}^{-1} \cdot \mathbf{A} = \mathbf{I}_n$, where $\mathbf{A}$ is an $n \times n$ square matrix.
- Distributivity:

$$\begin{aligned} \mathbf{A}(\mathbf{B} + \mathbf{C}) &= \mathbf{AB} + \mathbf{AC} \\ (\mathbf{A} + \mathbf{B})\mathbf{C} &= \mathbf{AC} + \mathbf{BC} \end{aligned}$$

- Associativity: $\mathbf{ABC} = \mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$
- Exponents: $\mathbf{A}^k = \underbrace{\mathbf{AA}\cdots\mathbf{A}}_{k,\text{times}}$
- Transposition:

$$\begin{aligned} (\mathbf{A} + \mathbf{B})' &= \mathbf{A}' + \mathbf{B}' \\ (\mathbf{AB})' &= \mathbf{B}'\mathbf{A}' \end{aligned}$$

Some Differentiation Rules

1. When we differentiate a scalar or a (scalar) function y with respect to a vector $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}'$, we have that:

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{bmatrix}'.$$

   That is, the result of the derivative is a vector containing the derivatives with respect to each element of $\mathbf{x}$.

2. Let $a' = [a_1\ a_2\ \dots\ a_n]$ and let $z = [z_1\ z_2\ \dots\ z_n]'$, then $a'z = \sum_{i=1}^n a_i z_i$ and

$$\frac{\partial(a'z)}{\partial z} = \begin{bmatrix} \frac{\partial a'z}{\partial z_1} \\ \frac{\partial a'z}{\partial z_2} \\ \vdots \\ \frac{\partial a'z}{\partial z_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

3. Let there be a matrix $z'Az$ such that:

$$z'Az = [z_1 z_2 \dots z_n] \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

   where A is a symmetric matrix (that is, $a_{ij} = a_{ji}$). Then,

$$\frac{\partial(z'Az)}{\partial z} = 2Az.$$

   Proof:

$$\begin{aligned} z'Az &= \begin{bmatrix} z_1 & z_2 & \dots & z_n \end{bmatrix} \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \\ &= \begin{bmatrix} \sum_i^n z_i a_{i1} & \sum_i^n z_i a_{i2} & \dots & \sum_i^n z_i a_{in} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \sum_j^n \sum_i^n z_i a_{ij} z_j \end{aligned}$$

Thus, since $a_{ij} = a_{ji}$, for example $\frac{\partial(z'Az)}{\partial z_1} = 2a_{ii}z_1 + z_2a_{21} + ... + z_2a_{12} + ... = 2\sum_i^n a_{1i}z_i$. Therefore,

$$\frac{\partial(z'Az)}{\partial z} = \begin{bmatrix} \frac{\partial(z'Az)}{\partial z_1} \\ \frac{\partial(z'Az)}{\partial z_2} \\ \vdots \\ \frac{\partial(z'Az)}{\partial z_n} \end{bmatrix} = \begin{bmatrix} 2\sum_i^n a_{1i}z_i \\ 2\sum_i^n a_{2i}z_i \\ \vdots \\ 2\sum_i^n a_{ni}z_i \end{bmatrix} = 2Az$$