# PS699: Problem Set 1 -- Matrix Algebra and Calculus

This assignment and several subsequent ones use the data set, "Political Structure in Developed Democracies," created for this class from the Political Data Handbook of OECD Countries by Jan-Erik Lane, David McKay, and Kenneth Newton. The data (in several formats) and a data codebook (in .rtf format) are on my web page.

1) In your spreadsheet, create the data matrix of Scandinavian countries' as the rows/observations (in this order: Denmark, Finland, Iceland, Norway, Sweden) and the following as the columns/variables (in this order) Lower-house seats (LSEATS), Proportionality Index (LPROP), Number of Governments Postwar (NGOVPW), and Social Security Spending as a % of total government expenditure (SSSPEND). Call this matrix **A**. (You could do this by hand without too much difficulty, but the point here is to practice data manipulation, and visualize matrices, in spreadsheets.)

[In the interest of conserving paper, and as good practice for the future (in terms of presentation of data and work), try to arrange parts a-e neatly on one page and parts f-h on no more than two pages.]

a) Print the data matrix **A**. Label it (in the spreadsheet, appealingly) so an outsider would know what it is.

b) What are the dimensions of **A**? (indicate this clearly and appealingly too, using the spreadsheet)

c) What is its transpose, $\mathbf{A}'$? what are the dimensions of $\mathbf{A}'$? (transpose **A** in the spreadsheet & print it, clearly labeled; include the dimensions clearly indicated.) (Transpose is a "Paste Special" in Excel)

d) In matrix **A**, what are the positional coordinates of the number of governments in Sweden's postwar history? of LPROP in Denmark? (Highlight or border the cell containing these data & label their coordinates)

e) Using your spreadsheet's "column average" function to calculate the row vector of means, $\mathbf{m}'$. Show $\mathbf{m}'$, labeled clearly.

f) Use the spreadsheet to calculate $\mathbf{A}$-$\mathbf{m}'$, call that **B**. Show **B**, labeled clearly.

g) Use the spreadsheet to calculate $(1/n)(\mathbf{B}'\mathbf{B})$; call the result **V**. Neatly arrange one page containing $\mathbf{B}'$, **B**, $\mathbf{B}'\mathbf{B}$, and **V**. Write the steps (including spreadsheet formulae) you employed to obtain the result on that page. Make sure each matrix is clearly labeled on the printout.

(Matrix multiplication in Excel is a pain. The command is "mmult", but the procedure is not very intuitive. Try help mmult, and then follow the instructions. You will need to know the dimensions of the solution.)

h) Do you recognize any statistical meanings for the elements of the solution to *g*? What are these elements?

2) In your spreadsheet, create two vectors, one equal to all 23 countries' observations on the average duration of governments in the postwar era, DGOVPW, and the other equal to the average number of parties in government in the postwar era, NPGOVPW. Subtract the mean of DGOVPW across these countries from every observation on DGOVPW; call the resulting vector **y**. Do the same for NPGOVPW, and call the resulting vector **x**. Use the scalar formula derived in class to determine the coefficient, $b$, on the vector, **x**, that minimizes $\mathbf{y} - b\mathbf{x}$.

Arranged appealingly on one page:

a) Present a graph scatter-plotting **y** against **x**. Put the line given by $\mathbf{y} = b\mathbf{x}$ on this graph too. Label the graph appropriately and appealingly.

b) Write the formula you used to obtain $b$.

c) Give a substantive interpretation of $b$.

d) Now create another vector, $\mathbf{x}_2$, equal to a $(23x1)$ column of ones. Create a matrix **X**, by putting the column of ones left of the column NPGOVPW (without its means subtracted). Using the multivariate formula given in class, find the coefficient vector **b**, which minimizes the distance from the $(23x1)$ vector **y**, given by DGOVPW (without its mean subtracted), to a line given by $\mathbf{Xb}$. Print $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ (showing the vector **b**, then the "=", then $(\mathbf{X}'\mathbf{X})^{-1}$, then $\mathbf{X}'\mathbf{y}$. Notice anything about the answers to a-c & d? Why did that occur?

3) In the notes, I proved the commutative and associative properties of matrix addition.

a) Now prove the distributive property of transposition over addition: $(\mathbf{A}+\mathbf{B})' = \mathbf{A}' + \mathbf{B}'$.

b) Using these three properties, prove that $[(\mathbf{A}+\mathbf{B})+\mathbf{C}]' = \mathbf{C}' + \mathbf{A}' + \mathbf{B}'$.

c) Does all this mean that we can add matrices in any order we like, transposing them as directed whenever?

4) Prove (ABC)′ = C′B′A′

a) First show that the expression on the right-hand side is conformable for multiplication if the expression on the left-hand side is.

b) Now do the proof (Hint: first prove that (**AB**)′ = **B′A′**; then use that result.)

5) In the mathematics appendix, Greene notes that the angle between **x***b* and **y**, $\theta$, could be considered a measure of how far **x***b* is from **y**. Its cosine varies from 0-1, being equal to 0 when **x***b* and **y** are orthogonal and equal to 1 when **x***b* and **y** are on the same line. This cosine is given by $cos\ \theta = (\mathbf{x}b'\mathbf{y}) / (\|\mathbf{x}b\| \cdot \|\mathbf{y}\|)$.

a) Substitute the formula for $\|\mathbf{v}\|$, the norm or length of a vector, **v**, into the cosine equation above & rewrite it.

b) Does the right-hand side look similar to any regression quantity you recognize? What? (Hint: call $\mathbf{x}b=\hat{\mathbf{y}}$.)

6) $\mathbf{C} = \begin{bmatrix} 1 & 3 \\ 5 & 6 \end{bmatrix}$ ; $\mathbf{E} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 2 & 1 \\ 1 & 3 \end{bmatrix}$ ; $\mathbf{F} = \begin{bmatrix} 1 & 3 & 7 \\ 2 & 1 & 4 \\ 3 & 1 & 5 \end{bmatrix}$

a) what is $\mathbf{CE}'$

b) what is $\mathbf{C}^{-1}$

c) what is $(\mathbf{E}'\mathbf{E})^{-1}$

d) what is $\mathbf{F}^{-1}$

7) In each of the following, how much does $y=f(x,\cdot)$ change for a marginal increase in $x$ (*i.e.*, take the partial derivative $\partial y/\partial x$. Unless otherwise specified, no other variables besides $y$ is a function of $x$. Also, $e$ here is the natural.)

a) $y = b_0 + b_1x + b_2z + \varepsilon$ b) $y = e^{a+bx}$ c) $y = wxz$

d) $y = b_0 + b_1x + b_2z + b_3xz + \varepsilon$ e) $y = ln(72x^{3e})$ f) $y = [e^{a+bx}]/[wxz]$

8) In each of the following, solve the indicated integral:

a) $\int_{-\infty}^{1} e^{a+bx}dx$ b) $\int_{a}^{c} (b_0+b_1x+b_2x^2)dx$