MATH 204 - EXERCISES 4

These exercises are not for assessment

1. Define the following vectors and matrices:
   - $n \times 1$ vector $\underset{\sim}{y} = [y_1, \ldots, y_n]^\mathsf{T}$
   - $n \times 2$ matrix $\boldsymbol{X}$ given by
   $$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix}^\mathsf{T}$$
   - 2×1 Parameter estimate vector $\underset{\sim}{\beta} = \left[\widehat{\beta}_0, \widehat{\beta}_1\right]^\mathsf{T}$

   Remember that the transpose operator $^\mathsf{T}$ means "flipping" a $r \times c$ matrix into a $c \times r$ matrix.

   It can be shown that
   $$\underset{\sim}{\widehat{\beta}} = (\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^{-1}\boldsymbol{X}^\mathsf{T}\underset{\sim}{y}$$
   where
   $$\boldsymbol{X}^\mathsf{T}\boldsymbol{X} \qquad\qquad (\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^{-1}$$
   are a 2×2 matrix and it's inverse.

   Verify the formula for $\underset{\sim}{\widehat{\beta}}$ by computing $\boldsymbol{X}^\mathsf{T}\boldsymbol{X}$ and $(\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^{-1}$ and comparing with the formulae given in lectures; write out the matrix multiplication in full.

   Recall that if A and B are two matrices, where A is a $r_1 \times c$ matrix and B is a $c \times r_2$ matrix, then the product AB is an $r_1 \times r_2$ matrix, with (i, j)th element
   $$\sum_{k=1}^{c} a_{ik} b_{kj} \qquad\qquad i = 1, \ldots, r_1,\ j = 1, \ldots, r_2.$$

2. Data Analysis: The data in the SPSS data set **Mercedes.sav** are taken from the advertising pages of the London Sunday Times, presenting Mercedes cars for sale in the UK. The asking prices (in pounds sterling) are compared against various factors (type/model of car, age of car in six-month units based on date of registration), recorded mileage, and vendor)

   Interest lies in explaining the variation in price, due to these factors, identifying *outliers* (non-typical prices), and predicting prices. Data columns are:

   - Car number $1, \ldots, 54$.
   - Asking **price** in pounds.
   - **Type/Model** of car: this is a discrete **factor** taking five levels (0=model 500, 1=450, 2=380, 3=280, 4=200).
   - **Age** of car in six-month units, based on registration date; this is a continuous **covariate**.
   - Recorded **mileage** (in thousands); this is a continuous **covariate**.
   - **Vendor**; this is a discrete **factor** taking five levels (0,1,2,3 are different dealerships, 4 is "sale by owner").

   Using these data, use ANOVA and regression techniques to identify which factors are influential in explaining the variation in car price. Are there any anomalous car prices ?