2a. $X = \overset{p\times 1}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}}$

$$XX' = \overset{p\times 1}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}} \overset{1\times p}{\begin{bmatrix} x_1 & x_2 & \cdots & x_p \end{bmatrix}}$$

$$= \overset{p\times p}{\begin{bmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_p \\ x_2x_1 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_px_1 & \cdots & & x_p^2 \end{bmatrix}}$$

$$\beta = \overset{p\times 1}{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}}$$

$$\mathcal{E}(XX')\beta = \overset{p\times p}{\begin{bmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_p \\ x_2x_1 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_px_1 & \cdots & & x_p^2 \end{bmatrix}} \overset{p\times 1}{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}}$$

$$= \overset{p\times 1}{\begin{bmatrix} x_1^2\beta_1 + x_1x_2\beta_2 + \ldots + x_1x_p\beta_p \\ x_2x_1\beta_1 + x_2^2\beta_2 + \ldots + x_2x_p\beta_p \\ \vdots \\ x_px_1\beta_1 + x_px_2\beta_2 + \ldots + x_p^2\beta_p \end{bmatrix}}$$

$$= \overset{p\times 1}{\begin{bmatrix} x_1(\beta_1x_1 + \beta_2x_2 + \ldots + \beta_px_p) \\ x_2(\beta_1x_1 + \beta_2x_2 + \ldots + \beta_px_p) \\ \vdots \\ x_p(\beta_1x_1 + \beta_2x_2 + \ldots + \beta_px_p) \end{bmatrix}}$$

$$\mathcal{E}(XY) = \overset{p\times 1}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}} \overset{1\times 1}{[Y]}$$

$$= \overset{p\times 1}{\begin{bmatrix} x_1Y \\ x_2Y \\ \vdots \\ x_pY \end{bmatrix}}$$

we know $Y = (\beta_1x_1 + \beta_2x_2 + \ldots + \beta_px_p)$

Therefore $\mathcal{E}(XX')\beta = \mathcal{E}(XY)$.

2b. $X = \overset{p\times 1}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}}$

$$XX' = \overset{p\times 1}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}} \overset{1\times p}{\begin{bmatrix} x_1 & x_2 & \cdots & x_p \end{bmatrix}}$$

$$= \overset{p\times p}{\begin{bmatrix} x_1x_1 & x_2x_1 & \cdots & x_1x_p \\ x_2x_1 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_px_1 & \cdots & & x_px_p \end{bmatrix}}$$

$$\alpha \in \mathbb{R}^p = \overset{p\times 1}{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}}$$

$$\mathcal{E}(XX')\alpha = \overset{p\times p}{\begin{bmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_p \\ x_2x_1 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_px_1 & \cdots & & x_p^2 \end{bmatrix}} \overset{p\times 1}{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}}$$

$$= \overset{p\times 1}{\begin{bmatrix} x_1^2a_1 + x_1x_2a_2 + \ldots + x_1x_pa_p \\ x_2x_1a_1 + x_2^2a_2 + \ldots + x_2x_pa_p \\ \vdots \\ x_px_1a_1 + x_px_2a_2 + \ldots + x_p^2a_p \end{bmatrix}} = \begin{bmatrix} x_1(x_1a_1 + x_2a_2 + \ldots + x_pa_p) \\ x_2(x_1a_1 + x_2a_2 + \ldots + x_pa_p) \\ \vdots \\ x_p(x_1a_1 + x_2a_2 + \ldots + x_pa_p) \end{bmatrix}$$

If $\mathcal{E}(XX')\alpha = 0$, then each row of $\mathcal{E}(XX')\alpha = 0$.

therefore, $(x_1a_1 + x_2a_2 + \ldots + x_pa_p)$ must $= 0$.

we can see this is true if $a_1 = -x_p,\ a_2 = -x_{p-1},\ a_3 = -x_{p-2},\ \ldots$

$a_p = x_1,\ a_{p-1} = x_2,\ a_{p-2} = x_3$.

Moreover, as long as we multiply all $a_1 \ldots a_p$ by the same scalar $n \in \mathbb{R}$, it will remain true that $\mathcal{E}(XX')\alpha = 0$.

Therefore S contains this $\alpha = (a_1, \ldots a_p)$, as well as any product of A and any $n \in \mathbb{R}$. And as $\mathbb{R}$ is infinitely large, there are infinitely many n, making the set S infinitely large.

(c)

No. Given that we know β_0 is the output of a regression using OLS, it is the result of a minimization problem so there should not be any other coefficient (not equal to 0) that also satisfies the equation.

(d)

i)

Y is wages

ii)

With in X there are two explanatory variables:

X_1 is age

X_2 is the years of experience in the workforce

iii)

X_1 (Age) has an interesting relationship with wages because as you get older, you tend to have spent longer at any given job or company, and you are also seen as more mature and wise. This loyalty and those respected traits usually result in higher wages.

X_2 (Experience) also has an interesting relationship with wages because the more experience you have, the more time you spend perfecting the skills related to the field. If you have spent a longer time within a certain industry too, you are also likely to be seen as more capable and worthy of promotions within the industry too. These factors both make one a more attractive employee, and typically, a more highly paid employee as well.

iv)

It is unlikely that, even knowing the probability distribution of (X,Y), we could measure the "true" population regression coefficient. Regressions can never truly measure the exact coefficient, but just do their best to estimate. However, because there is multicollinearity in our model, it is even more difficult to measure the "true" population regression coefficient. With multicollinearity, there is less independent variation for our explanatory variables, ie. we can't see the effect of age without the effect of experience (and vice versa). The estimators for age and experience are in a sense "diluted" as their correlation means the effect of age on wage likely captures a portion of the effect of experience on wage because it is hard to find changes in age without corresponding changes in experience (and vice versa). For example we might learn that years of education and years of experience are both positively correlated with wages. But because the variance of the estimators are high we don't know exactly how much each contributes to increasing or decreasing wages.

Problem 3

3a.

$$\underset{n\times p}{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{n1} & \cdots & & x_{np} \end{bmatrix}$$

$$X'X = \underset{p\times n}{\begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{1p} & \cdots & & x_{np} \end{bmatrix}} \underset{n\times p}{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{n1} & \cdots & & x_{np} \end{bmatrix}}$$

$$= \underset{p\times p}{\begin{bmatrix} (x_{11}^2 + x_{21}^2 + \dots + x_{n1}^2) & (x_{11}x_{12} + \dots + x_{n1}x_{n2}) & \cdots & (x_{11}x_{1p} + \dots + x_{n1}x_{np}) \\ (x_{12}x_{11} + \dots + x_{n2}x_{n1}) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ (x_{1p}x_{11} + \dots + x_{np}x_{n1}) & \cdots & & (x_{1p}^2 + \dots + x_{np}^2) \end{bmatrix}}$$

$$x_i x_i' = \underset{p\times 1}{\begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}} \underset{1\times p}{\begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{ip} \end{bmatrix}}$$

$$= \underset{p\times p}{\begin{bmatrix} x_{i1}^2 & x_{i1}x_{i2} & \cdots & x_{i1}x_{ip} \\ x_{i2}x_{i1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{ip}x_{i1} & \cdots & & x_{ip}^2 \end{bmatrix}}$$

$$\Rightarrow \sum_{i=1}^{n} x_i x_i' = \begin{bmatrix} (x_{11}^2 + \dots + x_{n1}^2) & (x_{11}x_{12} + \dots + x_{n1}x_{n2}) & \cdots & (x_{11}x_{1p} + \dots + x_{n1}x_{np}) \\ (x_{12}x_{11} + \dots + x_{n2}x_{n1}) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ (x_{1p}x_{11} + \dots + x_{np}x_{n1}) & \cdots & & (x_{1p}^2 + \dots + x_{np}^2) \end{bmatrix}$$

$$\Rightarrow \sum_{i=1}^{n} x_i x_i' = X'X$$

3b.

$$\underset{k\times 1}{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$$

$$\underset{k\times T}{\mathbb{X}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1t} \\ x_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{k1} & \cdots & & x_{kt} \end{bmatrix}$$

$$\mathbb{X}\mathbb{X}' = \underset{k\times T}{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1t} \\ x_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{k1} & \cdots & & x_{kt} \end{bmatrix}} \underset{T\times k}{\begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{1t} & \cdots & & x_{kt} \end{bmatrix}}$$

$$= \underset{k\times k}{\begin{bmatrix} (x_{11}^2 + x_{12}^2 + x_{1t}^2) & (x_{11}x_{21} + x_{12}x_{22} + \dots + x_{1t}x_{2t}) & \cdots & (x_{11}x_{k1} + x_{12}x_{k2} + \dots + x_{1t}x_{kt}) \\ (x_{21}x_{11} + x_{22}x_{12} + \dots + x_{2t}x_{1t}) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ (x_{k1}x_{11} + x_{k2}x_{12} + \dots + x_{kt}x_{1t}) & \cdots & & (x_{k1}^2 + x_{k2}^2 + \dots + x_{kt}^2) \end{bmatrix}}$$

$$= \underset{k\times k}{\begin{bmatrix} \sum_{i=1}^{t} (x_{1t}^2) & \sum_{i=1}^{t} (x_{1t}x_{2t}) & \cdots & \sum_{i=1}^{t} (x_{1t}x_{kt}) \\ \sum_{i=1}^{t} (x_{2t}x_{1t}) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \sum_{i=1}^{t} (x_{kt}x_{1t}) & \cdots & & \sum_{i=1}^{t} (x_{kt}^2) \end{bmatrix}}$$

$$\mathcal{E}(\mathbb{X}\mathbb{X}') = \frac{1}{t} \begin{bmatrix} \sum_{i=1}^{t} (x_{1t}^2) & \sum_{i=1}^{t} (x_{1t}x_{2t}) & \cdots & \sum_{i=1}^{t} (x_{1t}x_{kt}) \\ \sum_{i=1}^{t} (x_{2t}x_{1t}) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \sum_{i=1}^{t} (x_{kt}x_{1t}) & \cdots & & \sum_{i=1}^{t} (x_{kt}^2) \end{bmatrix}$$