Assignment 2: Data Visualization

Due beginning of class: Monday September 25

6. **Graduate students** (bonus undergraduates): Suppose we have n-dimensional real and linearly independent vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$ and $\mathbf{y}$. The vector $\mathbf{y}$ is the sum of two n-dimensional real vectors μ and $\mathbf{r}$

$$\mathbf{y} = \mu + \mathbf{r}$$

where μ is restricted to be a linear combination of the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$. That is

$$\mu = \theta_1 \times \mathbf{x}_1 + \theta_2 \times \mathbf{x}_2 + \cdots + \theta_p \times \mathbf{x}_p$$

for some unknown real constants $\theta_1, \theta_2, \ldots, \theta_p$, or equivalently

$$\mu = \mathbf{X}\theta$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_p]$ is an $n \times p$ matrix and $\theta = (\theta_1, \theta_2, \ldots, \theta_p)^T$ is a $p \times 1$ vector.

(a) (5 marks) For any $\mathbf{y}$, neither μ nor $\mathbf{r}$ are uniquely defined. Suppose we choose particular vectors $\widehat{\mu}$, and $\widehat{\mathbf{r}}$ (with $\mathbf{y} = \widehat{\mu} + \widehat{\mathbf{r}}$) to be such that they are orthogonal to one another (whatever values any θ_i take). That is, $\widehat{\mu}^T\widehat{\mathbf{r}} = 0$.

Prove that this additional constraint implies that

$$\widehat{\mu} = \mathbf{P}\mathbf{y}$$

where $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ and hence show that $\widehat{\mathbf{r}} = (\mathbf{I}_n - \mathbf{P})\mathbf{y}$.

$\widehat{\mu}^T\widehat{\mathbf{r}} = 0$ (whatever values any θ_i take) $\implies$ $\widehat{\mathbf{r}}$ is orthogonal to $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$. Hence:

$$\begin{aligned}
\mathbf{X}^T\widehat{\mathbf{r}} &= \begin{bmatrix} \mathbf{x}_1^T\widehat{\mathbf{r}} \\ \cdots \\ \mathbf{x}_p^T\widehat{\mathbf{r}} \end{bmatrix} \\
&= \mathbf{0} \\
\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\widehat{\mathbf{r}} &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{0} \\
\mathbf{P}\widehat{\mathbf{r}} &= \mathbf{0} \\
\mathbf{P}\mathbf{y} - \mathbf{P}\widehat{\mathbf{r}} &= \mathbf{P}\mathbf{y} \\
\mathbf{P}\widehat{\mu} &= \mathbf{P}\mathbf{y}
\end{aligned}$$

But we can simplify $\mathbf{Pu}$ to:

$$\begin{aligned}
\mathbf{P}\mu &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mu \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\theta \\
&= \mathbf{X}\mathbf{I}_p\theta \\
&= \mathbf{X}\theta \\
&= \widehat{\mu}
\end{aligned}$$

$\therefore \widehat{\mu} = \mathbf{P}\mathbf{y}$. For the second proof:

$$\begin{aligned}
\mathbf{P}\mathbf{y} &= \widehat{\mu} \\
&= \mathbf{y} - \widehat{\mathbf{r}} \\
\widehat{\mathbf{r}} &= \mathbf{y} - \mathbf{P}\mathbf{y} \\
&= (\mathbf{I}_n - \mathbf{P})\mathbf{y}
\end{aligned}$$

(b) (2 marks) Show that $\mathbf{P}$ is an idempotent matrix, that is that $\mathbf{P}^2 = \mathbf{P}$.

$$\begin{aligned}
\mathbf{P}^2 &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{I}_p\mathbf{X}^T \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\
&= \mathbf{P}
\end{aligned}$$

$\therefore \mathbf{P}$ is an idempotent matrix.

(c) (2 marks) Show that if $\mathbf{P}$ is an idempotent matrix, then so must be $ ({**I}__n -{P**})$.

Assume that $\mathbf{P}$ is an idempotent matrix.

$$
\begin{aligned}
(\mathbf{I}_n - \mathbf{P})^2 &= (\mathbf{I}_n - \mathbf{P})(\mathbf{I}_n - \mathbf{P}) \\
&= \mathbf{I}_n^2 - \mathbf{I}_n\mathbf{P} - \mathbf{P}\mathbf{I}_n + \mathbf{P}^2 \\
&= \mathbf{I}_n - \mathbf{P} - \mathbf{P} + \mathbf{P} \qquad \text{since } \mathbf{P} \text{ is idempotent} \\
&= \mathbf{I}_n - \mathbf{P}
\end{aligned}
$$

$\therefore (\mathbf{I}_n - \mathbf{P})$ is an idempotent matrix.

(d) (2 marks) Show that $\widehat{\mathbf{r}}$ is in fact orthogonal to $\widehat{\mu}$.

We first show that P is symmetric:

$$
\begin{aligned}
\mathbf{P}^T &= (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T \\
&= (\mathbf{X}^T)^T(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1})^T \\
&= \mathbf{X}((\mathbf{X}^T\mathbf{X})^{-1})^T\mathbf{X}^T \\
&= \mathbf{X}((\mathbf{X}^T\mathbf{X})^T)^{-1}\mathbf{X}^T \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \\
&= \mathbf{P}
\end{aligned}
$$

$\therefore \mathbf{P}$ is symmetric.

$$
\begin{aligned}
\widehat{\mu}^T\widehat{\mathbf{r}} &= \widehat{\mu} \cdot \widehat{\mathbf{r}} \\
&= \mathbf{P}\mathbf{y} \cdot (\mathbf{I}_n - \mathbf{P})\mathbf{y} \\
&= \mathbf{y} \cdot \mathbf{P}(\mathbf{I}_n - \mathbf{P})\mathbf{y} \\
&= \mathbf{y}^T(\mathbf{P} - \mathbf{P}^2)\mathbf{y} \\
&= \mathbf{y}^T\mathbf{0}\mathbf{y} \\
&= \mathbf{0}
\end{aligned}
$$

$\therefore \widehat{\mu}$ is orthogonal to $\widehat{\mathbf{r}}$.

(e) (5 marks) Show that $\widehat{\mu} = \mathbf{P}\mathbf{y}$ is the choice of μ which minimizes the squared length of $\mathbf{r}$. That is it minimizes $\mathbf{r}^T\mathbf{r} = (\mathbf{y} - \mu)^T(\mathbf{y} - \mu)$.