[STAT409] Homework 2

1. In the multiple linear regression model:
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$
where the definition of vectors are given in the lecture note.
   (a) Show that LSE of $\boldsymbol{\beta}$, the minimizer of $S(\boldsymbol{\beta})$ is given by
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{1}$$
   (b) Prove the following decomposition (a.k.a ANOVA decomposition) in the multiple linear regression.
$$\underbrace{(\mathbf{y}-\bar{\mathbf{y}})^T(\mathbf{y}-\bar{\mathbf{y}})}_{\text{SST}} = \underbrace{(\mathbf{y}-\hat{\mathbf{y}})^T(\mathbf{y}-\hat{\mathbf{y}})}_{\text{SSE}} + \underbrace{(\hat{\mathbf{y}}-\bar{\mathbf{y}})^T(\hat{\mathbf{y}}-\bar{\mathbf{y}})}_{\text{SSR}}$$

2. In the simple linear regression $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, we have the following OLS slope esitmators:
$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}, \tag{2}$$
   (a) Compute $E(\hat{\beta}_1)$ and $Var(\hat{\beta}_1)$.
   (b) It is well known that $\hat{\beta}_1$ is normally distributed when ϵ_i's are normaly distributed. Justify this.

3. For a given $y_1, \cdots, y_n \overset{iid}{\sim} F$, show that the sample median is the minimizer of $A_n(c)$ where
$$A_n(c) = \frac{1}{n}\sum |y_i - c|$$
(hint: You may use the fact that the derivative of $|x|$ is simply $\text{sign}(x)$ and ignore its nondifferentiability at $x = 0$.)

4. Suppose $X_1 \sim Bionmial(n_1, \theta)$ and $X_2 \sim Binomial(n_2, \theta)$ are independent random variables.
   (a) Write down the log likelihood of θ, denoted by $\ell(\theta)$.
   (b) Compute the maximum likelihood estimator of θ, denoted by $\hat{\theta}$.
   (c) Compute the information of θ:
$$I(\theta) = E\left[\left\{\frac{\partial \ell(\theta)}{\partial \theta}\right\}^2\right] = -E\left[\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right]$$