h (1) $Y = X\beta + \varepsilon$ 最小二乘 $\hat{\beta} = (X^TX)^{-1}X^TY$

$\varepsilon_i \overset{i.i.d}{\sim} N(0, \varepsilon^2)$, $i = 1 \cdots n$

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\varepsilon_i^2\right)$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{\varepsilon^T\varepsilon} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-(Y-X\beta)^T(Y-X\beta)\frac{1}{2\sigma^2}}$$

对给定的 $\sigma^2$, $\max\limits_{\beta} L(\beta, \sigma^2) \Longleftrightarrow \underset{\beta}{\overset{\min}{\cancel{\max}}} (Y-X\beta)^T(Y-X\beta)$

这相当于求解最小二乘问题 $\Rightarrow \hat{\beta} = (X^TX)^{-1}X^TY$

(2) 设 $a^TY$ 是 $C^T\beta$ 的一个线性无偏估计

$$E(a^TY) = E[a^T(X\beta+\varepsilon)] = a^TX\beta = C^T\beta \quad \text{对} \ \forall \beta \ \text{成立}$$

$$\Rightarrow \quad a^TX = C^T$$

令 $C^T\hat{\beta} = C^T(X^TX)^{-1}X^TY = \hat{a}^TY$ , $\hat{a} = X(X^TX)^{-1}C$

$$Var(a^TY) = Var(a^T\varepsilon) = \sigma^2 a^Ta = \sigma^2(a-\hat{a}+\hat{a})^T(a-\hat{a}+\hat{a})$$

$$= \sigma^2[(a-\hat{a})^T(a-\hat{a}) + \hat{a}^T\hat{a} + 2(a-\hat{a})^T\hat{a}]$$

式中 $(a-\hat{a})^T\hat{a} = a^TX(X^TX)^{-1}C - C^T(X^TX)^{-1}X^TX(X^TX)^{-1}C$

$$= C^T(X^TX)^{-1}C - C^T(X^TX)^{-1}C = 0$$

2) $Var(a^TY) = \sigma^2[(a-\hat{a})^T(a-\hat{a}) + \hat{a}^T\hat{a}]$

易见，当 $a=\hat{a}$ 时 $Var(a^TY)$ 最小，亦即 $C^T\hat{\beta}$ 是 $C^T\beta$ 的 BLUE

(3) $\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = (y-\hat{y})^T(y-\hat{y})$

$\hat{y} = X\hat{\beta} = X(X^TX)^{-1}X^Ty$ ， $(y-\hat{y})^T(y-\hat{y}) = \left(I - X(X^TX)^{-1}X^T\right)y$

$= \left(I - X(X^TX)^{-1}X^T\right)(X\beta+\varepsilon)$

$= \left(I - X(X^TX)^{-1}X^T\right)\varepsilon$

$\Rightarrow (y-\hat{y})^T(y-\hat{y}) = \varepsilon^T\left(I - X(X^TX)^{-1}X^T\right)\varepsilon$ ，易见 $I - X(X^TX)^{-1}X^T$ 是幂等阵

$= tr\left(\varepsilon^T\left(I - X(X^TX)^{-1}X^T\right)\varepsilon\right)$

$= tr\left(\left(I - X(X^TX)^{-1}X^T\right)\varepsilon\varepsilon^T\right)$

由： $E(\varepsilon\varepsilon^T) = Var\,\varepsilon = \sigma^2 I$

$E(tr(W)) = tr(EW)$ （因为 $E(trW) = EW_{11} + \cdots + EW_{nn} = tr(EW)$）

得到： $E(y-\hat{y})^T(y-\hat{y}) = tr\left(\left(I - X(X^TX)^{-1}X^T\right)E\varepsilon\varepsilon^T\right)$

$= \sigma^2 tr(I) - \sigma^2 tr\left(X(X^TX)^{-1}X^T\right)$

$= \sigma^2 tr(I_n) - \sigma^2 \varepsilon\, tr\left[(X^TX)^{-1}X^TX\right]$

$= \sigma^2(n - tr(I_p)) = \sigma^2(n-p)$

$\Rightarrow E\hat{\sigma}^2 = E\frac{1}{n-p}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = \sigma^2$

(4) $\hat{\beta} = (X^TX)^{-1}X^Ty = (X^TX)^{-1}X^T(X\beta+\varepsilon) = \beta + (X^TX)^{-1}X^T\varepsilon$

$\varepsilon \sim N(0, \sigma^2 I_N)$ $\Rightarrow$ $\hat{\beta}$ 服从正态分布

$E\hat{\beta} = \beta$, $Var(\hat{\beta}) = (X^TX)^{-1}X^T E\varepsilon\varepsilon^T X(X^TX)^{-1} = \sigma^2(X^TX)^{-1}$

$\Rightarrow \hat{\beta} \sim N(\beta, \sigma^2(X^TX)^{-1})$

由上一个问，$(y-\hat{y})^T(y-\hat{y}) = \varepsilon^T(I - X(X^TX)^{-1}X^T)\varepsilon$，$I - X(X^TX)^{-1}X^T$ 为幂等阵

$tr(I - X(X^TX)^{-1}X^T) = N-p$ $\Rightarrow$ 其特征值 $\lambda_1 = \cdots = \lambda_{N-p} = 1$, $\lambda_{N-p+1} = \cdots = \lambda_N = 0$

$I - X(X^TX)^{-1}X^T = e_1e_1^T + \cdots + e_{N-p}e_{N-p}^T$ （$e_i$为特征向量）

$\varepsilon^T(I - X(X^TX)^{-1}X^T)\varepsilon = (e_1^T\varepsilon)^T(e_1^T\varepsilon) + \cdots + (e_{N-p}^T\varepsilon)^T(e_{N-p}^T\varepsilon)$

$e_i^T\varepsilon \sim N(0, \sigma^2 e_i^Te_i) = N(0, \sigma^2)$

$Cov(e_i^T\varepsilon, e_j^T\varepsilon) = e_i^T E\varepsilon\varepsilon^T e_j = \sigma^2\delta_{ij} = \begin{cases}\sigma^2 & i=j \\ 0 & i\neq j\end{cases}$

故 $e_i^T\varepsilon \overset{i.i.d}{\sim} N(0, \sigma^2)$ $(i=1\cdots N-p)$

$\Rightarrow \underbrace{\varepsilon^T(I - X(X^TX)^{-1}X^T)\varepsilon}_{\displaystyle \parallel \atop \displaystyle (N-p)\hat{\sigma}^2} \sim \sigma^2\chi^2(N-p)$

(5) 设 $X \sim N(\mu, \sigma^2)$

$$Ee^X = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\mu+\frac{1}{2}\sigma^2} e^{-\frac{1}{2\sigma^2}[(x-\mu)-\sigma^2]^2} dx$$

令 $y = x - \mu$

$$Ee^X = e^{\mu+\frac{1}{2}\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(y-\sigma^2)^2} \frac{1}{\sqrt{2\pi}\sigma} dy = e^{\mu+\frac{1}{2}\sigma^2}$$

本题中：$\varepsilon \sim N(0, \sigma^2)$，$\ln y \sim N(X^T\beta, \sigma^2)$

$$EY = Ee^{\ln Y} = e^{X^T\beta + \frac{1}{2}\sigma^2}$$

(6). 记 $\hat{\varepsilon} = y - \hat{y}$

$$y^T y = (y - \hat{y} + \hat{y})^T (y - \hat{y} + \hat{y}) = (\hat{y} + \hat{\varepsilon})^T (\hat{y} + \hat{\varepsilon}) = \hat{y}^T \hat{y} + \hat{\varepsilon}^T \hat{\varepsilon}$$

其中 $\hat{y}^T \hat{\varepsilon} = \hat{\beta}^T X^T \hat{\varepsilon} = 0$，因为 $X^T \hat{\varepsilon} = X^T (y - \hat{y}) = X^T (I - X(X^T X)^{-1} X^T) y = X^T y - X^T y = 0$

$X^T \hat{\varepsilon} = 0$，考虑 $X$ 第一列为 $[1, \cdots, 1]^T$，有：$\sum_{i=1}^{n} \hat{\varepsilon}_i = \sum_{i=1}^{n} (y_i - \hat{y}_i) = 0$

$$\Rightarrow \bar{y} = \bar{\hat{y}}$$

$$\Rightarrow \quad y^T y - n(\bar{y})^2 = \hat{y}^T \hat{y} - (\bar{\hat{y}})^2 n + \hat{\varepsilon}^T \hat{\varepsilon}$$

$$\Rightarrow \sum_{j=1}^{n} (y_j - \bar{y})^2 = \sum_{j=1}^{n} (\hat{y}_j - \bar{y})^2 + \sum_{j=1}^{n} \hat{\varepsilon}_j$$

即 $TSS = ESS + RSS$