HW2.

1. From OLS, $\hat{\beta} = (X^T X)^{-1} X^T y$.

From MLE, $\hat{\beta} = \text{argmax}_\beta \log L(\beta)$.

According to (A5) assumption, $\varepsilon \sim N(0, \sigma^2 I)$.

$L(\beta) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^T (y - X\beta)\right\}$

$\log L(\beta) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)^T (y - X\beta)$

$\hat{\beta} = \text{argmax}_\beta \left\{-(y - X\beta)^T (y - X\beta)\right\}$

$= \text{argmin}_\beta \left\{(y - X\beta)^T (y - X\beta)\right\}$.

对 $(y - X\beta)^T (y - X\beta)$ 求导，得 $-2X^T (y - X\beta)$.

令 $-2X^T (y - X\beta) = 0$

则 $X^T y = X^T X\beta$，即 $\hat{\beta} = (X^T X)^{-1} X^T y$, the same as the OLS estimator.

2. 假设 $\beta' = c^T y = \sum_i c_i y_i$ 是 y 的任一线性无偏估计量，$\hat{\beta} = (X^T X)^{-1} X^T y$ 是 OLS 得到的.

$E(\beta') = E(c^T y) = c^T X\beta = \beta$. $\Rightarrow$ $c^T X = I$.

$Cov(\beta') = Cov(c^T y) = c^T Var(y) c = \sigma^2 c^T c$

$Cov(\hat{\beta}) = Cov[(X^T X)^{-1} X^T y] = \sigma^2 (X^T X)^{-1} = \sigma^2 c^T X (X^T X)^{-1} X^T c$.

$Cov(\beta') - Cov(\hat{\beta}) = \sigma^2 c^T [I - X(X^T X)^{-1} X^T] c$.

$= \sigma^2 c^T [I - H] c$

其中 $H = X(X^T X)^{-1} X^T$，$HH = X(X^T X)^{-1} X^T X (X^T X)^{-1} X = H$ $\Rightarrow$ H 为幂等矩阵.

$(I - H)(I - H) = I - H - H + H = I - H$ $\Rightarrow$ $I - H$ 为幂等矩阵. $\Rightarrow$ $I - H$ 半正定.

$\therefore$ $Cov(\beta') - Cov(\hat{\beta})$ 一定半正定.

$\therefore$ $Var(z^T \beta') \geq Var(z^T \hat{\beta})$

$\therefore$ 得证，$\hat{\beta} = (X^T X)^{-1} X^T y$ 有最小方差.

3. $E(\hat{\sigma}^2)=E\left\{\frac{1}{N-p}\sum_{i=1}^{N}e_i^2\right\}$.

$$=\frac{1}{N-p}\sum_{i=1}^{N}(Ee_i^2)$$

$$=\frac{1}{N-p}\sum_{i=1}^{N}(Var(e_i)+\overset{=0}{E^2(e_i)})$$

$$=\frac{1}{N-p}\sum_{i=1}^{N}Var(e_i).$$

$Var(e_i)=Var(e)$的对角元素.

$$Var(e)=Var(y-\hat{y})=Var(y)+Var(\hat{y})-2Cov(y,\hat{y})$$

$$=\sigma^2 I_{n\times n}+Var(X\hat{\beta})-2Cov(y,X\hat{\beta}).$$

$$=\sigma^2 I_{n\times n}+X(X^TX)^{-1}X^T\sigma^2-2X(X^TX)^{-1}X^T\sigma^2$$

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

$$=(I-H)\sigma^2$$ 其中 $H=X(X^TX)^{-1}X^T$.

$Var(e_i)=(1-h_{ii})\sigma^2$ 其中h_{ii}为H的对角元.

$$E(\hat{\sigma}^2)=\frac{1}{N-p}\sum_{i=1}^{N}(1-h_{ii})\sigma^2=\frac{N}{N-p}\sigma^2-\frac{1}{N-p}\sum_{i=1}^{N}h_{ii}\sigma^2$$

$$\sum_{i=1}^{N}h_{ii}=tr(H)=tr(X(X^TX)^{-1}X^T)=tr((X^TX)^{-1}X^TX)=tr(I_p)=p.$$

$\therefore E(\hat{\sigma}^2)=\frac{N-p}{N-p}\sigma^2=\sigma^2$. 得证.

4. (1) $E(\hat{\beta})=E[(X^TX)^{-1}X^Ty]=E[(X^TX)^{-1}X^TX\beta]=\beta$.

$Cov(\hat{\beta})=Cov[(X^TX)^{-1}X^Ty]=(X^TX)^{-1}X^TCov(y)X(X^TX)^{-1}=\sigma^2(X^TX)^{-1}$.

由$\hat{\beta}=(X^TX)^{-1}X^Ty$可知$\hat{\beta}$是对$y$进行线性变换后得到.

$\because \varepsilon\sim N(0,\sigma^2 I_n)$.

$\therefore y\sim N(X\beta,\sigma^2 I_n)$

$\therefore \hat{\beta}$也服从正态分布, 即$\hat{\beta}\sim N(\beta,\sigma^2(X^TX)^{-1})$.

(2) $(N-p)\hat{\sigma}^2 = \sum_{i=1}^{N}(y_i-\hat{y}_i)^2$

$= (y-X\hat{\beta})^T(y-X\hat{\beta})$

$= [(I-H)y]^T[(I-H)y]$

$= y^T(I-H)^T(I-H)y$ 　　(I-H是幂等矩阵)

$= y^T(I-X(X^TX)^{-1}X^T)y$

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

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

$\frac{(N-p)\hat{\sigma}^2}{\sigma^2} = (\frac{\varepsilon}{\sigma})^T(I-X(X^TX)^{-1}X^T)(\frac{\varepsilon}{\sigma})$.

根据(A4)，$\varepsilon \sim N(0,\sigma^2 I) \Rightarrow \frac{\varepsilon}{\sigma} \sim N(0,I)$.

$I-X(X^TX)^{-1}X^T$为幂等阵，故$I-H$的特征值一定为0或1. $\Rightarrow tr(I-H) = rank(I-H)$.

$I-H$为实对称阵 $\Rightarrow$ 存在正交阵Q使$I-H=QDQ^T$, $D=diag(1,\dots,1,0,\dots,0)$.

$rank(H) = tr(I) - tr[X(X^TX)^{-1}X^T] = n - tr(H) = n-p$. $\Rightarrow$ D中1的个数为$n-p$.

$\therefore \frac{(N-p)\hat{\sigma}^2}{\sigma^2} = (\frac{\varepsilon}{\sigma}Q^T)^T D(\frac{\varepsilon}{\sigma}Q^T)$.

又 $\because \frac{\varepsilon}{\sigma}Q^T \sim N(0,QQ^T) \sim N(0,I)$.

$\therefore \frac{(N-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$. 即$(N-p)\hat{\sigma}^2 \sim \sigma^2\chi^2_{n-p}$，得证.

5. $\log(y) \sim N(x^T\beta, \sigma^2)$.

令$z=\log(y)$，则$f_z(z) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(z-x^T\beta)^2}{2\sigma^2}\right\}$，$y=e^z$，$g^{-1}(y)=\log(y)$.

$f_y(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(\log y-x^T\beta)^2}{2\sigma^2}\right\}\cdot\frac{1}{y}$. 　　$y>0$.

$E(y) = \int_0^{+\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(\log y-x^T\beta)^2}{2\sigma^2}\right\}dy$

$= \int\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(z-x^T\beta)^2}{2\sigma^2}+z\right\}dz$.

$= \int\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\left[z^2-2(x^T\beta+\sigma^2)z+(x^T\beta+\sigma^2)^2-2\sigma^2x^T\beta-\sigma^4\right]\right\}dz$.

$= \exp\left\{x^T\beta+\frac{1}{2}\sigma^2\right\}$. 　　(后面一串积分相当于求$N(x^T\beta+\sigma^2,\sigma^2)$的pdf的积分，为1).

$\therefore E(y) = \exp\left\{x^T\beta+\frac{1}{2}\sigma^2\right\}$.

6. $RSS=\sum_i (y_i-\hat{y}_i)^2$.

$ESS=\sum_i (\hat{y}_i-\bar{y})^2$.

$TSS=\sum_i (y_i-\bar{y})^2=\sum_i (y_i-\hat{y}_i)^2+\sum_i (\hat{y}_i-\bar{y})^2+2\sum_i (y_i-\hat{y}_i)(\hat{y}_i-\bar{y})$.

$\quad\quad = RSS+ESS+2\sum_i (y_i-\hat{y}_i)(\hat{y}_i-\bar{y})$.

问题转化为证明 $\sum_i (y_i-\hat{y}_i)(\hat{y}_i-\bar{y})=0$.

$\sum_i y_i\hat{y}_i-\sum_i y_i\bar{y}-\sum_i \hat{y}_i^2+\sum_i \hat{y}_i\bar{y}$

$=\sum_i (y_i-\hat{y}_i)\hat{y}_i$

$=\sum_i e_i x_i^T\hat{\beta}$ (因为 OLS 中 $\sum_i e_i x_i^T=0$).

$=0$.

$\therefore TSS=ESS+RSS$, 得证.