1. (a) Show $\frac{1}{n}J$ idempotent

$$(\tfrac{1}{n}J)(\tfrac{1}{n}J) = \frac{1}{n^2}(\underset{\sim}{1}\cdot\underset{\sim}{1}')(\underset{\sim}{1}\cdot\underset{\sim}{1}')$$
$$= \frac{1}{n^2}\underset{\sim}{1}(\underset{\sim}{1}'\cdot\underset{\sim}{1})\underset{\sim}{1}'$$
$$= \frac{1}{n^2}\underset{\sim}{1}(n)\underset{\sim}{1}'$$
$$= \frac{1}{n}\underset{\sim}{1}\underset{\sim}{1}'$$
$$= \frac{1}{n}J$$

Show $H-\frac{1}{n}J$ idempotent

$$(H-\tfrac{1}{n}J)(H-\tfrac{1}{n}J) = H - \tfrac{1}{n}JH - \tfrac{1}{n}HJ + \tfrac{1}{n}J$$
$$= H - \tfrac{1}{n}J(X(X'X)^{-1}X') - \tfrac{1}{n}(X(X'X)^{-1}X')J + \tfrac{1}{n}J$$
$$= H - \tfrac{1}{n}J(XX^{-1}(X^{-1})'X') - \tfrac{1}{n}(XX^{-1}(X^{-1})'X')J + \tfrac{1}{n}J$$
$$= H - \tfrac{1}{n}J(I(XX^{-1})') - \tfrac{1}{n}(I(XX^{-1})')J + \tfrac{1}{n}J$$
$$= H - \tfrac{1}{n}J - \tfrac{1}{n}J + \tfrac{1}{n}J$$
$$= H - \tfrac{1}{n}J$$

Show $I-H$ idempotent

$$(I-H)(I-H) = II - HI - IH + H \quad (H \text{ idempotent})$$
$$= I - H - H + H$$
$$= I - H$$

Show $\frac{1}{n}J$ and $H-\frac{1}{n}J$ pairwise orthogonal

$$\tfrac{1}{n}J(H-\tfrac{1}{n}J) = \tfrac{1}{n}JH - \tfrac{1}{n}J$$
$$= \tfrac{1}{n}J - \tfrac{1}{n}J$$
$$= 0$$

Show $\frac{1}{n}J$ and $I-H$ pairwise orthogonal

$$\tfrac{1}{n}J(I-H) = \tfrac{1}{n}J - \tfrac{1}{n}JH$$
$$= \tfrac{1}{n}J - \tfrac{1}{n}J$$
$$= 0$$

(c) $SSE = \underset{\sim}{Y}'(I-H)\underset{\sim}{Y}$

$\because \underset{\sim}{Y} \sim N(X\underset{\sim}{\beta}, I\sigma^2)$

$\therefore \dfrac{SSE}{\sigma^2} = \underset{\sim}{Y}'\left(\dfrac{I-H}{\sigma^2}\right)\underset{\sim}{Y}$

According to lecture 12, if $Z \sim N(\underset{\sim}{\mu}, V\sigma^2)$ for a nonsingular matrix V, then a quadratic form $\underset{\sim}{Z}'\left(\dfrac{A}{\sigma^2}\right)\underset{\sim}{Z}$ is distributed as noncentral chisquare distribution with $df = r(A)$ and $\Omega = \dfrac{\underset{\sim}{\mu}'A\underset{\sim}{\mu}}{2\sigma^2}$

In this case $\underset{\sim}{\mu} = X\underset{\sim}{\beta}$, $V = I$ is nonsingular matrix,

$A = I-H$

$r(A) = r(I-H) = r(I) - r(H) = n - p'$

$$\Omega = \frac{\underset{\sim}{\mu}'A\underset{\sim}{\mu}}{2\sigma^2} = \frac{\underset{\sim}{\beta}'X'(I-H)X\underset{\sim}{\beta}}{2\sigma^2} = \frac{\hat{\underset{\sim}{Y}}'(I-H)\hat{\underset{\sim}{Y}}}{2\sigma^2}$$

$$\begin{aligned}
\hat{\underset{\sim}{Y}}'(I-H)\hat{\underset{\sim}{Y}} &= \underset{\sim}{Y}'H'(I-H)H\underset{\sim}{Y} \\
&= (\underset{\sim}{Y}'HI - \underset{\sim}{Y}'HH)H\underset{\sim}{Y} \\
&= (\underset{\sim}{Y}'H - \underset{\sim}{Y}'H)H\underset{\sim}{Y} \\
&= 0 \cdot H\underset{\sim}{Y} = 0
\end{aligned}$$

$\because AV = (I-H)I = I-H$ is idempotent

$\therefore \underset{\sim}{Y}'\left(\dfrac{I-H}{\sigma^2}\right)\underset{\sim}{Y} = \dfrac{\underset{\sim}{Y}'(I-H)\underset{\sim}{Y}}{\sigma^2} = \dfrac{SSE}{\sigma^2}$ is distributed as $\chi^2_{(n-p')}$

2(b) For multiple regression model with $X_1, \dots, X_p$ and interception, Let design matrix be X^*

$$(X^*)'X^* = \begin{bmatrix} n & \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{2i} & \cdots & \sum_{i=1}^n X_{pi} \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_{1i}^2 & \sum_{i=1}^n X_{1i}X_{2i} & \cdots & \sum_{i=1}^n X_{1i}X_{pi} \\ \sum_{i=1}^n X_{2i} & \sum_{i=1}^n X_{1i}X_{2i} & \sum_{i=1}^n X_{2i}^2 & \cdots & \sum_{i=1}^n X_{2i}X_{pi} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^n X_{pi} & \sum_{i=1}^n X_{1i}X_{pi} & \sum_{i=1}^n X_{2i}X_{pi} & \cdots & \sum_{i=1}^n X_{pi}^2 \end{bmatrix}_{p \times p}$$

If we remove the column for interception from design matrix, we will get this for $X'X$ due to property of matrix multiplication.

$$X'X = \begin{bmatrix} \sum_{i=1}^n X_{1i}^2 & \sum_{i=1}^n X_{1i}X_{2i} & \cdots & \sum_{i=1}^n X_{1i}X_{pi} \\ \sum_{i=1}^n X_{1i}X_{2i} & \sum_{i=1}^n X_{2i}^2 & \cdots & \sum_{i=1}^n X_{2i}X_{pi} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^n X_{1i}X_{pi} & \sum_{i=1}^n X_{2i}X_{pi} & \cdots & \sum_{i=1}^n X_{pi}^2 \end{bmatrix}_{p \times p}$$

In our case, all variables are categorical variables, so

Let $\#(E_i = 1) = n_i$, $i = 1, \dots, p$

$$X'X = \begin{bmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & n_p \end{bmatrix}_{p \times p}$$

(b/c $\sum_{i=1}^n E_{ji}E_{ki} = 0$ for $j = 1, \dots, p$ $i = 1, \dots, p$ $k = 1, \dots, p$ $j \neq k$ due to property of categorical variable)

is diagonal matrix

(d) WTS: $\frac{SSR}{\sigma^2}$ and $\frac{SSE}{\sigma^2}$ are independent

$\Leftrightarrow$ WTS: SSR and SSE are independent

$$\frac{SSR}{\sigma^2} = \frac{\underset{\sim}{Y}'(H - \frac{1}{n}J)\underset{\sim}{Y}}{\sigma^2} \qquad \frac{SSE}{\sigma^2} = \frac{\underset{\sim}{Y}'(I - H)\underset{\sim}{Y}}{\sigma^2}$$

According to lecture 12, $\underset{\sim}{Z}'A\underset{\sim}{Z}$ and $\underset{\sim}{Z}'B\underset{\sim}{Z}$ are independent if $AVB=0$.

Let $SSR=\underset{\sim}{Y}'(H-\frac{1}{n}J)\underset{\sim}{Y}$ be $\underset{\sim}{Z}'A\underset{\sim}{Z}$, let $SSE=\underset{\sim}{Y}'(I-H)\underset{\sim}{Y}$ be $\underset{\sim}{Z}'B\underset{\sim}{Z}$, $V=I$

$$
\begin{aligned}
AVB &= (H-\tfrac{1}{n}J)I(I-H) \\
&= (H-\tfrac{1}{n}J)(I-H)
\end{aligned}
$$

From part (a), we proved that $H-\frac{1}{n}J$ and $I-H$ are orthogonal, so $AVB=0$. Hence $SSR=\underset{\sim}{Y}'(H-\frac{1}{n}J)\underset{\sim}{Y}$ and $SSE=\underset{\sim}{Y}'(I-H)\underset{\sim}{Y}$ are independent, and $\frac{SSR}{\sigma^2}$ and $\frac{SSE}{\sigma^2}$ are independent

(e) WTS: F-test for $\beta_1=\beta_2=\cdots=\beta_p=0$ is a particular case of general linear hypothesis test.

F-test: $H_0: \beta_1=\beta_2=\cdots=\beta_p=0$

H_a: At least one $\beta_i\neq 0$, $i=1,\dots,p$

Let $K=\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}_{p'\times p'}=I$, $\underset{\sim}{\beta}=\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_p\end{bmatrix}_{p'\times 1}$ $\underset{\sim}{m}=\begin{bmatrix}\beta_0\\ 0\\ \vdots\\ 0\end{bmatrix}_{p'\times 1}$

$\therefore H_0: K'\underset{\sim}{\beta}=\underset{\sim}{m}$

$H_a: K'\underset{\sim}{\beta}\neq\underset{\sim}{m}$

2.(a) $Y=\beta_1 E_1+\beta_2E_2+\cdots+\beta_pE_p+\varepsilon_i$

with assumption $\varepsilon_i\sim(0,\sigma^2)$

$E(E_1)=E_1$, $E(E_2)=E_2$, $\dots$, $E(E_p)=E_p$

$Var(E_j)=0$, $j=1,\dots,p$

Interpretation: β_j = the expectation of Y at $X=j$, $j=1,\dots,p$

(b)

$$(X'X)^{-1} = \left[\begin{array}{cccc|cccc} n_1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & n_p & 0 & 0 & \cdots & 1 \end{array}\right]$$

$$= \left[\begin{array}{cccc|cccc} 1 & 0 & \cdots & 0 & \frac{1}{n_1} & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \frac{1}{n_2} & \cdots & 0 \\ \vdots & \vdots & & & \vdots & \vdots & & \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & \frac{1}{n_p} \end{array}\right]$$

Let $\sum_{i=1}^{n} Y_{i(X=j)}$ denotes sum(Y_i correspoding to $X=j$)

$$X'\underset{\sim}{Y} = \begin{bmatrix} \sum_{i=1}^{n} Y_{i(X=1)} \\ \sum_{i=1}^{n} Y_{i(X=2)} \\ \vdots \\ \sum_{i=1}^{n} Y_{i(X=p)} \end{bmatrix}_{p\times 1}$$

$$\underset{\sim}{\hat{\beta}} = (X'X)^{-1}X'\underset{\sim}{Y} = \begin{bmatrix} \frac{1}{n_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{n_2} & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & \frac{1}{n_p} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} Y_{i(X=1)} \\ \sum_{i=1}^{n} Y_{i(X=2)} \\ \vdots \\ \sum_{i=1}^{n} Y_{i(X=p)} \end{bmatrix}$$

where $\overline{Y_i}$: (the sum of Y correponding to $X=i$)/#($X=i$), $i=1,\cdots,p$

$$= \begin{bmatrix} \sum_{i=1}^{n} Y_{i(X=1)} \\ \sum_{i=1}^{n} Y_{i(X=2)} \\ \vdots \\ \sum_{i=1}^{n} Y_{i(X=p)} \end{bmatrix} = \begin{bmatrix} \overline{Y}_1 \\ \overline{Y}_2 \\ \vdots \\ \overline{Y}_p \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$$

Q2(c).

$Var(\hat{\beta}) = \sigma^2 \cdot (X'X)^{-1}$

$= \sigma^2 \cdot \begin{bmatrix} \frac{1}{n_1} & & & 0 \\ & \frac{1}{n_2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{n_p} \end{bmatrix} = \begin{bmatrix} \frac{\sigma^2}{n_1} & & & 0 \\ & \frac{\sigma^2}{n_2} & & \\ & & \ddots & \\ 0 & & & \frac{\sigma^2}{n_p} \end{bmatrix}$

$var(\hat{\beta}_j) = \frac{\sigma^2}{n_j}$

$cov(\hat{\beta}_i, \hat{\beta}_j) = 0 \quad (i \neq j)$

$\rightarrow cor(\hat{\beta}_i, \hat{\beta}_j) = 0$

Q2. (d). usual t-test.

$H_0: \beta_1 = 0$

$H_0: \beta_2 = 0$

$\vdots$

$H_0: \beta_p = 0$

ANOVA:

$H_0: \beta_1 = \beta_2 = \cdots \beta_p = 0.$

$H_0: \mu_1 = \mu_2 = \cdots = \mu_p = 0.$ (μ is mean) $Y = \beta_1 Z_1 + \cdots \beta_p Z_p.$