$\hat{\sigma}^2 = \frac{RSS}{n-p}$, show $\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$

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

To intuitively understand why is χ^2_{n-p}:

If we know β exactly, $\frac{1}{\sigma^2}(Y - X\beta)^T(Y - X\beta) = \frac{e^T e}{\sigma^2}$, $\frac{1}{\sigma^2} e^T e$ is the sum of n z^2, where $z \sim N(0, 1)$, and thus $\frac{1}{\sigma^2} e^T e \sim \chi^2_n$.

But we don't know β, so we use $\hat{\beta}$. Estimating $\hat{\beta}$ loses p dfs, so $RSS \sim \chi^2_{n-p}$

To formally prove $\frac{RSS}{\sigma^2}$ is the sum of $(n-p)$ z^2, we need to do axis transformation.

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

① Let $\xi = X\beta$, then $Y \sim N(\xi, \sigma^2 I_n)$ and $\xi \in C(X)$ (the column space of X)

② Let $Y = OZ$, then $Z = O^T Y \sim N(O^T \xi, \sigma^2 I_n)$

let $\eta = O^T \xi$.

$O = (V_1, V_2, \cdots, V_r, V_{r+1}, \cdots, V_n)$, where $V_1 - V_r$ is the orthonormal basis for $C(X)$, and $V_{r+1} - V_n$

are chosen to be orthogonal to $v_1 - v_r$ and normalized.

$r = \text{rank}(X)$,

Since $\xi \in C(X)$, we have that $\eta = \begin{pmatrix} v_1^T\xi \\ \vdots \\ v_r^T\xi \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

③ By checking the density of Z , we can show that the MLE and LSE of η is : $\hat{\eta}_i = z_i$

$$\text{So } \hat{\xi} = O\hat{\eta} = (v_1, \cdots, v_r, v_{r+1}, \cdots, v_n)\begin{pmatrix} \hat{\eta}_1 \\ \vdots \\ \hat{\eta}_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= \sum_{i=1}^{r} z_i v_i$$

$$RSS = (Y - \hat{\xi})^T (Y - \hat{\xi})$$

$$= \left(\sum_{i=1}^{n} z_i v_i - \sum_{i=1}^{r} z_i v_i\right)^T \left(\sum_{i=1}^{n} z_i v_i - \sum_{i=1}^{r} z_i v_i\right)$$

$$= \left(\sum_{i=r+1}^{n} z_i v_i\right)^T \left(\sum_{i=r+1}^{n} z_i v_i\right)$$

$$= \sum_{i=r+1}^{n} z_i^2 \quad (\because v_i^T v_i = 1,\ \underset{j \neq i}{v_i^T v_j = 0})$$

$$\sim \chi^2_{n-r}\, \sigma^2 \quad (\because z_i \sim N(0, \sigma^2),\ i \geq r+1)$$