Some specific results for multiple linear regression models.

R1 In the linear regression model $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

with $\mathrm{Cov}(\underline{\varepsilon}) = \sigma^2 \underline{I}$ an estimator for σ^2 is given by

$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{n-(k+1)}$ where k is the number of regression variables.

Proof $\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = (\underline{Y} - \hat{\underline{Y}})^T (\underline{Y} - \hat{\underline{Y}}) = \hat{\underline{\varepsilon}}^T \hat{\underline{\varepsilon}}$

and $\hat{\underline{\varepsilon}} = \underline{Y} - \hat{\underline{Y}} = (\underline{I} - \underline{H})\underline{Y} = (\underline{I} - \underline{H})(\underline{Y} - \underline{X}\underline{\beta}) = (\underline{I} - \underline{H})\underline{\varepsilon}$

since $(\underline{I} - \underline{H})\underline{X}\underline{\beta} = \underline{X}\underline{\beta} - \underline{X}\underline{\beta} = 0$

Therefore $E[\hat{\underline{\varepsilon}}^T \hat{\underline{\varepsilon}}] = E[\underline{\varepsilon}^T (\underline{I} - \underline{H})(\underline{I} - \underline{H})\underline{\varepsilon}] = E[\underline{\varepsilon}^T (\underline{I} - \underline{H})\underline{\varepsilon}]$

$= tr[(\underline{I} - \underline{H})\sigma^2 \underline{I}] = \sigma^2 tr[(\underline{I} - \underline{H})] = \sigma^2 [tr(\underline{I}) - tr(\underline{H})]$

$tr(\underline{H}) = \cancel{tr(X)}\ tr(\underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T) = tr[\underline{X}^T\underline{X}(\underline{X}^T\underline{X})^{-1}] = tr(\underline{I}_{(k+1)\times(k+1)}$

$\underline{X}^T\underline{X}$ is a $(k+1)\times(k+1)$ matrix

Hence $E[\hat{\underline{\varepsilon}}^T \hat{\underline{\varepsilon}}] = \sigma^2(n - (k+1))$ and

$$E[\hat{\sigma}^2] = \sum_{i=1}^{n} \frac{E[\hat{\underline{\varepsilon}}^T \hat{\underline{\varepsilon}}]}{n-(k+1)} = \frac{\sigma^2(n-(k+1))}{n-(k+1)} = \sigma^2.$$

R2

In the linear regression model $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$, $\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 \underline{I})$

a)
$$\frac{SS_E}{\sigma^2} = \frac{(\underline{Y} - \underline{X}\hat{\underline{\beta}})^T(\underline{Y} - \underline{X}\hat{\underline{\beta}})}{\sigma^2} \sim \chi^2(n-k-1)$$

Proof. $(\underline{I} - \underline{H})(\underline{Y} - \underline{X}\underline{\beta}) = \underline{Y} - \underline{X}\hat{\underline{\beta}} - \underline{X}\underline{\beta} + \underline{X}\underline{\beta} = \underline{Y} - \underline{X}\hat{\underline{\beta}}$

Hence
$$\frac{SS_E}{\sigma^2} = \frac{(\underline{Y} - \underline{X}\underline{\beta})^T(\underline{I} - \underline{H})(\underline{I} - \underline{H})(\underline{Y} - \underline{X}\underline{\beta})}{\sigma^2} = \frac{(\underline{Y} - \underline{X}\underline{\beta})^T(\underline{I} - \underline{H})(\underline{Y} - \underline{X}\underline{\beta})}{\sigma^2}$$

which is $\chi^2(\text{rank}(\underline{I} - \underline{H}))$ and $\underset{\underset{tr(\underline{I}-\underline{H})}{\parallel}}{\text{rank}(\underline{I} - \underline{H})} = n-k-1$

R2

b) Assume $\beta_1 = \beta_2 = \dots = \beta_k = 0$ i.e. $Y_i = \beta_0 + \varepsilon_i$, $i = 1, 2, \dots, n$

Then
$$\frac{SS_T}{\sigma^2} = \frac{(\underline{Y} - \bar{\underline{Y}})^T(\underline{Y} - \bar{\underline{Y}})}{\sigma^2} \sim \chi^2(n-1)$$

and
$$\frac{SS_R}{\sigma^2} = \frac{(\hat{\underline{Y}} - \bar{\underline{Y}})^T(\hat{\underline{Y}} - \bar{\underline{Y}})}{\sigma^2} \sim \chi^2(k)$$

Proof In this case $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}\beta_0 + \underline{\varepsilon} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_0 \end{bmatrix} + \underline{\varepsilon}$

Therefore $(\underline{I} - \frac{1}{n}\underline{1}\underline{1}^T)(\underline{Y} - \underline{X}\underline{\beta}) = \underline{Y} - \bar{\underline{Y}} - \underline{\beta_0} + \underline{\beta_0} = \underline{Y} - \bar{\underline{Y}}$

and $(\underline{H} - \frac{1}{n}\underline{1}\underline{1}^T)(\underline{Y} - \underline{X}\underline{\beta}) = \underline{X}\hat{\underline{\beta}} - \bar{\underline{Y}} - \underline{\beta_0} + \underline{\beta_0} = \hat{\underline{Y}} - \bar{\underline{Y}}$

and we get
$$\frac{SS_T}{\sigma^2} = \frac{(\underline{Y} - \underline{X}\underline{\beta})^T(\underline{I} - \frac{1}{n}\underline{1}\underline{1}^T)(\underline{I} - \frac{1}{n}\underline{1}\underline{1}^T)(\underline{Y} - \underline{X}\underline{\beta})}{\sigma^2}$$

$$= \frac{(\underline{Y} - \underline{X}\underline{\beta})^T(\underline{I} - \frac{1}{n}\underline{1}\underline{1}^T)(\underline{Y} - \underline{X}\underline{\beta})}{\sigma^2} \sim \chi^2(\text{rank}(\underline{I} - \tfrac{1}{n}\underline{1}\underline{1}^T))$$
$$\sim \chi^2(n-1)$$

and

$$\frac{SS_R}{\sigma^2} = \frac{(\underline{Y}-\underline{X}\underline{\beta})^T(\underline{H}-\frac{1}{n}\underline{1}\underline{1}^T)(\underline{H}-\frac{1}{n}\underline{1}\underline{1}^T)(\underline{Y}-\underline{X}\underline{\beta})}{\sigma^2}$$

$$= \frac{(\underline{Y}-\underline{X}\underline{\beta})^T(\underline{H}-\frac{1}{n}\underline{1}\underline{1}^T)(\underline{Y}-\underline{X}\underline{\beta})}{\sigma^2} \sim \chi^2(k)$$

since $tr(\underline{H}-\frac{1}{n}\underline{1}\underline{1}^T) = k+1-1 = k.$

We also have $(\underline{I}-\underline{H})(\underline{H}-\frac{1}{n}\underline{1}\underline{1}^T) = H-H-\frac{1}{n}\underline{1}\underline{1}^T+\frac{1}{n}\underline{1}\underline{1}^T = \underline{0}$

from which it follows that SS_E and SS_R are independent.

$\underline{R3}$. Assume $\underline{\varepsilon} \sim N(0, \sigma^2\underline{I})$. Then $\hat{\underline{\beta}}$ and $\hat{\sigma}^2$ are independent.

$\underline{\text{Proof}}$. $Cov(\hat{\underline{\beta}}, \underline{Y}-\underline{X}\hat{\underline{\beta}}) = E[(\hat{\underline{\beta}}-\underline{\beta})(\underline{Y}-\underline{H}\underline{Y})^T] = E[(\hat{\underline{\beta}}-\underline{\beta})\hat{\underline{\varepsilon}}^T]$

$= ~~E[(X^TX)^{-1}X^T(Y-X\beta)\ldots]~~$

$$= E[(\underline{X}^T\underline{X})^{-1}\underline{X}^T(\underline{Y}-\underline{X}\underline{\beta})\underline{\varepsilon}^T(\underline{I}-\underline{H})] = E[(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{\varepsilon}\underline{\varepsilon}^T(\underline{I}-\underline{H})]$$

$$= \sigma^2[(\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{I}(\underline{I}-\underline{H})] = \sigma^2[(\underline{X}^T\underline{X})^{-1}\underline{X}^T - (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T] = \underline{0}$$

The Fisher ($\mathcal{F}$) distribution and the analysis of variance table for linear regression.

Assume $X \sim \chi^2(\nu_1)$ and $Y \sim \chi^2(\nu_2)$ and X and Y are independent. Then $\dfrac{X/\nu_1}{Y/\nu_2}$ is Fisher distributed with ν_1 and ν_2 degrees of freedom, writes $\mathcal{F}_{\nu_1,\nu_2}$.

Note $\mathcal{F} \sim \mathcal{F}_{\nu_1,\nu_2} \Rightarrow \dfrac{1}{\mathcal{F}} \sim \mathcal{F}_{\nu_2,\nu_1}$

Now define f_{α,ν_1,ν_2} by $P(\mathcal{F}_{\nu_1,\nu_2} \leq f_{\alpha,\nu_1,\nu_2}) = \alpha$

Since $P(f_{1-\frac{\alpha}{2},\nu_1,\nu_2} \leq \mathcal{F} \leq f_{\frac{\alpha}{2},\nu_1,\nu_2}) = 1-\alpha$

$$\Leftrightarrow P\left(\frac{1}{f_{\frac{\alpha}{2},\nu_1,\nu_2}} \leq \frac{1}{\mathcal{F}} \leq \frac{1}{f_{1-\frac{\alpha}{2},\nu_1,\nu_2}}\right) = 1-\alpha$$

We have $\dfrac{1}{f_{\frac{\alpha}{2},\nu_1,\nu_2}} = f_{1-\frac{\alpha}{2},\nu_2,\nu_1}$ and $f_{1-\frac{\alpha}{2},\nu_1,\nu_2} = \dfrac{1}{f_{\frac{\alpha}{2},\nu_2,\nu_1}}$ $\left(f_{\frac{\alpha}{2},\nu_2,\nu_1} = \dfrac{1}{f_{1-\frac{\alpha}{2},\nu_1,\nu_2}}\right)$

The density function in the $\mathcal{F}$ distribution is given by

$$f(x) = k_{\nu_1,\nu_2} \cdot x^{\frac{\nu_1}{2}-1}\left(1+\frac{\nu_1}{\nu_2}x\right)^{-\frac{\nu_1+\nu_2}{2}}, \quad x > 0$$

where $$k_{\nu_1,\nu_2} = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\cdot\Gamma\left(\frac{\nu_2}{2}\right)} \cdot \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}$$

$E[\mathcal{F}] = \dfrac{\nu_1}{\nu_2-2}$ and $Var(\mathcal{F}_{\nu_1,\nu_2}) = \dfrac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$ provided $\nu_2 > 4$