MATH 423/533 – THE SUMS OF SQUARES DECOMPOSITION

The sums of squares decomposition that follows in a linear regression model is

$$\begin{array}{ccccc} \sum_{i=1}^{n}(y_i-\overline{y})^2 & = & \sum_{i=1}^{n}(y_i-\widehat{y}_i)^2 & + & \sum_{i=1}^{n}(\widehat{y}_i-\overline{y})^2 \\ \mathrm{SS_T} & = & \mathrm{SS_{Res}} & + & \mathrm{SS_R} \end{array}$$

which may be re-written

$$\mathbf{y}^\top(\mathbf{I}_n-\mathbf{H}_1)\mathbf{y} = \mathbf{y}^\top(\mathbf{I}_n-\mathbf{H})\mathbf{y} + \mathbf{y}^\top(\mathbf{H}-\mathbf{H}_1)\mathbf{y}$$

where

$$\mathbf{H}_1 = \mathbf{1}_n(\mathbf{1}_n^\top\mathbf{1}_n)^{-1}\mathbf{1}_n^\top \qquad\qquad \mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top.$$

In random variable form, we wish to study

$$\mathrm{SS_R} = \mathbf{Y}^\top(\mathbf{H}-\mathbf{H}_1)\mathbf{Y}$$

where, under the linear model formulation

$$\mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\mathbf{Y}|\mathbf{X}] = \mathbf{X}\beta \qquad\qquad \mathbb{V}\mathrm{ar}_{\mathbf{Y}|\mathbf{X}}[\mathbf{Y}|\mathbf{X}] = \sigma^2\mathbf{I}_n.$$

Recall the result for the expectation of a quadratic form that if $\mathbf{V}$ is a k-dimensional random vector with $\mathbb{E}[\mathbf{V}]=\mu$, $\mathbb{V}\mathrm{ar}[\mathbf{V}]=\Sigma$, then for $k\times k$ matrix $\mathbf{A}$, we have

$$\mathbb{E}[\mathbf{V}^\top\mathbf{A}\mathbf{V}] = \mathrm{trace}(\mathbf{A}\Sigma) + \mu^\top\mathbf{A}\mu$$

Using these results, we showed that

$$\begin{aligned} \mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\mathrm{SS_T}|\mathbf{X}] &= (n-1)\sigma^2 + \beta^\top\mathbf{X}^\top(\mathbf{H}-\mathbf{H}_1)\mathbf{X}\beta \\ \mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\mathrm{SS_{Res}}|\mathbf{X}] &= (n-p)\sigma^2 \\ \mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\mathrm{SS_R}|\mathbf{X}] &= (p-1)\sigma^2 + \beta^\top\mathbf{X}^\top(\mathbf{H}-\mathbf{H}_1)\mathbf{X}\beta \end{aligned}$$

For $\mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\mathrm{SS_R}|\mathbf{X}]$, we showed

$$\mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\mathbf{Y}^\top(\mathbf{H}-\mathbf{H}_1)\mathbf{Y}|\mathbf{X}] = \mathrm{trace}\{\sigma^2(\mathbf{H}-\mathbf{H}_1)\} + \beta^\top\mathbf{X}^\top(\mathbf{H}-\mathbf{H}_1)\mathbf{X}\beta = \sigma^2(p-1) + \beta^\top\mathbf{M}\beta$$

say, where

$$\mathbf{M} \;=\; \mathbf{X}^\top(\mathbf{H}-\mathbf{H}_1)\mathbf{X} = \mathbf{X}^\top(\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top)\mathbf{X} - \mathbf{X}^\top\mathbf{1}_n(\mathbf{1}_n^\top\mathbf{1}_n)^{-1}\mathbf{1}_n^\top\mathbf{X}$$

Now if we write $\mathbf{X}=[\mathbf{1}_n\ \mathbf{X}_{\mathrm{R}}]$, then

$$\begin{aligned} \mathbf{M} &= \begin{bmatrix} \mathbf{1}_n^\top \\ \mathbf{X}_{\mathrm{R}}^\top \end{bmatrix} \left(\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}_n(\mathbf{1}_n^\top\mathbf{1}_n)^{-1}\mathbf{1}_n^\top\right)[\mathbf{1}_n\ \mathbf{X}_{\mathrm{R}}] \\ &= \begin{bmatrix} \mathbf{1}_n^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}_n^\top\mathbf{1}_n(\mathbf{1}_n^\top\mathbf{1}_n)^{-1}\mathbf{1}_n^\top \\ \mathbf{X}_{\mathrm{R}}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{X}_{\mathrm{R}}^\top\mathbf{1}_n(\mathbf{1}_n^\top\mathbf{1}_n)^{-1}\mathbf{1}_n^\top \end{bmatrix}[\mathbf{1}_n\ \mathbf{X}_{\mathrm{R}}] \\ &= \begin{bmatrix} \mathbf{0}_n^\top \\ \mathbf{X}_{\mathrm{R}}^\top - \mathbf{X}_{\mathrm{R}}^\top\mathbf{1}_n(\mathbf{1}_n^\top\mathbf{1}_n)^{-1}\mathbf{1}_n^\top \end{bmatrix}[\mathbf{1}_n\ \mathbf{X}_{\mathrm{R}}] \end{aligned}$$

as the identity

$$\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top = \mathbf{X}^\top \qquad \Longrightarrow \qquad \begin{bmatrix} \mathbf{1}_n^\top \\ \mathbf{X}_{\text{R}}^\top \end{bmatrix} \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top = \begin{bmatrix} \mathbf{1}_n^\top \\ \mathbf{X}_{\text{R}}^\top \end{bmatrix}.$$

Hence

$$\mathbf{M} \;=\; \begin{bmatrix} 0 & \mathbf{0}_{p-1}^\top \\ \mathbf{0}_{p-1}^\top & \mathbf{X}_{\text{R}}^\top\mathbf{X}_{\text{R}} - \mathbf{X}_{\text{R}}^\top\mathbf{1}_n(\mathbf{1}_n^\top\mathbf{1}_n)^{-1}\mathbf{1}_n^\top\mathbf{X}_{\text{R}} \end{bmatrix}$$

The non-zero block matrix can be written

$$\mathbf{X}_{\text{R}}^\top\mathbf{X}_{\text{R}} - \mathbf{X}_{\text{R}}^\top\mathbf{1}_n(\mathbf{1}_n^\top\mathbf{1}_n)^{-1}\mathbf{1}_n^\top\mathbf{X}_{\text{R}} = \mathbf{X}_{\text{C}}^\top\mathbf{X}_{\text{C}}$$

where $\mathbf{X}_{\text{C}}$ is constructed from $\mathbf{X}_{\text{R}}$ by subtracting from each column the column mean. Hence in general if $\beta = (\beta_0, \beta_1, \ldots, \beta_{p-1})^\top$, we have that

$$\mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\text{SS}_{\text{R}}|\mathbf{X}] = (p-1)\sigma^2 + (\beta_1, \ldots, \beta_{p-1})^\top\mathbf{X}_{\text{C}}^\top\mathbf{X}_{\text{C}}(\beta_1, \ldots, \beta_{p-1})$$

For simple linear regression with $p = 2$,

$$\mathbf{X}_{\text{C}}^\top\mathbf{X}_{\text{C}} \equiv (x_{11} - \overline{x}_1, \ldots, x_{n1} - \overline{x}_1)^\top(x_{11} - \overline{x}_1, \ldots, x_{n1} - \overline{x}_1) = \sum_{i=1}^n (x_{i1} - \overline{x}_1)^2 = S_{xx}.$$

Thus for simple linear regression we have (with $p = 2$) that

$$\begin{aligned} \mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\text{SS}_{\text{T}}|\mathbf{X}] &= (n-1)\sigma^2 + \beta_1^2 S_{xx} \\ \mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\text{SS}_{\text{Res}}|\mathbf{X}] &= (n-p)\sigma^2 \\ \mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\text{SS}_{\text{R}}|\mathbf{X}] &= (p-1)\sigma^2 + \beta_1^2 S_{xx} \end{aligned}$$

Furthermore, under the assumption of normality of the residual errors ϵ, if $\beta_1 = 0$, we have

$$\begin{aligned} \frac{\text{SS}_{\text{T}}}{\sigma^2} &= \frac{\mathbf{Y}^\top(\mathbf{I}_n - \mathbf{H}_1)\mathbf{Y}}{\sigma^2} \sim \chi^2_{n-1} \\ \frac{\text{SS}_{\text{Res}}}{\sigma^2} &= \frac{\mathbf{Y}^\top(\mathbf{I}_n - \mathbf{H})\mathbf{Y}}{\sigma^2} \sim \chi^2_{n-p} \\ \frac{\text{SS}_{\text{R}}}{\sigma^2} &= \frac{\mathbf{Y}^\top(\mathbf{H} - \mathbf{H}_1)\mathbf{Y}}{\sigma^2} \sim \chi^2_{p-1} \end{aligned}$$

with SS_{Res} and SS_{R} independent. Thus, similarly to the result quoted earlier

$$\begin{aligned} \mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\text{SS}_{\text{T}}|\mathbf{X}] &= (n-1)\sigma^2 \\ \mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\text{SS}_{\text{Res}}|\mathbf{X}] &= (n-p)\sigma^2 \\ \mathbb{E}_{\mathbf{Y}|\mathbf{X}}[\text{SS}_{\text{R}}|\mathbf{X}] &= (p-1)\sigma^2 \end{aligned}$$

with $p = 2$ hold under the assumption that $\beta_1 = 0$.

Note: the assumption (hypothesis) that $\beta_1 = 0$ implies that x is not a relevant predictor of y.

- This is the context within which we will use the sums of squares decomposition most often.

If $\beta_1 \neq 0$, we have

$$
\begin{aligned}
\frac{\mathrm{SS_T}}{\sigma^2} &\sim \chi^2_{n-1}(\lambda) \\
\frac{\mathrm{SS_{Res}}}{\sigma^2} &\sim \chi^2_{n-p} \\
\frac{\mathrm{SS_R}}{\sigma^2} &\sim \chi^2_{p-1}(\lambda)
\end{aligned}
$$

where

$$\lambda = \frac{\beta_1^2 S_{xx}}{\sigma^2}$$

and $\chi^2_\nu(\delta)$ denotes the **non-central chisquared distribution** with ν degrees of freedom and noncentrality parameter δ. This distribution is a two parameter extension of the ordinary chi-squared distribution (which has a single parameter, the degrees of freedom).

Code for the non-central chi-squared distribution

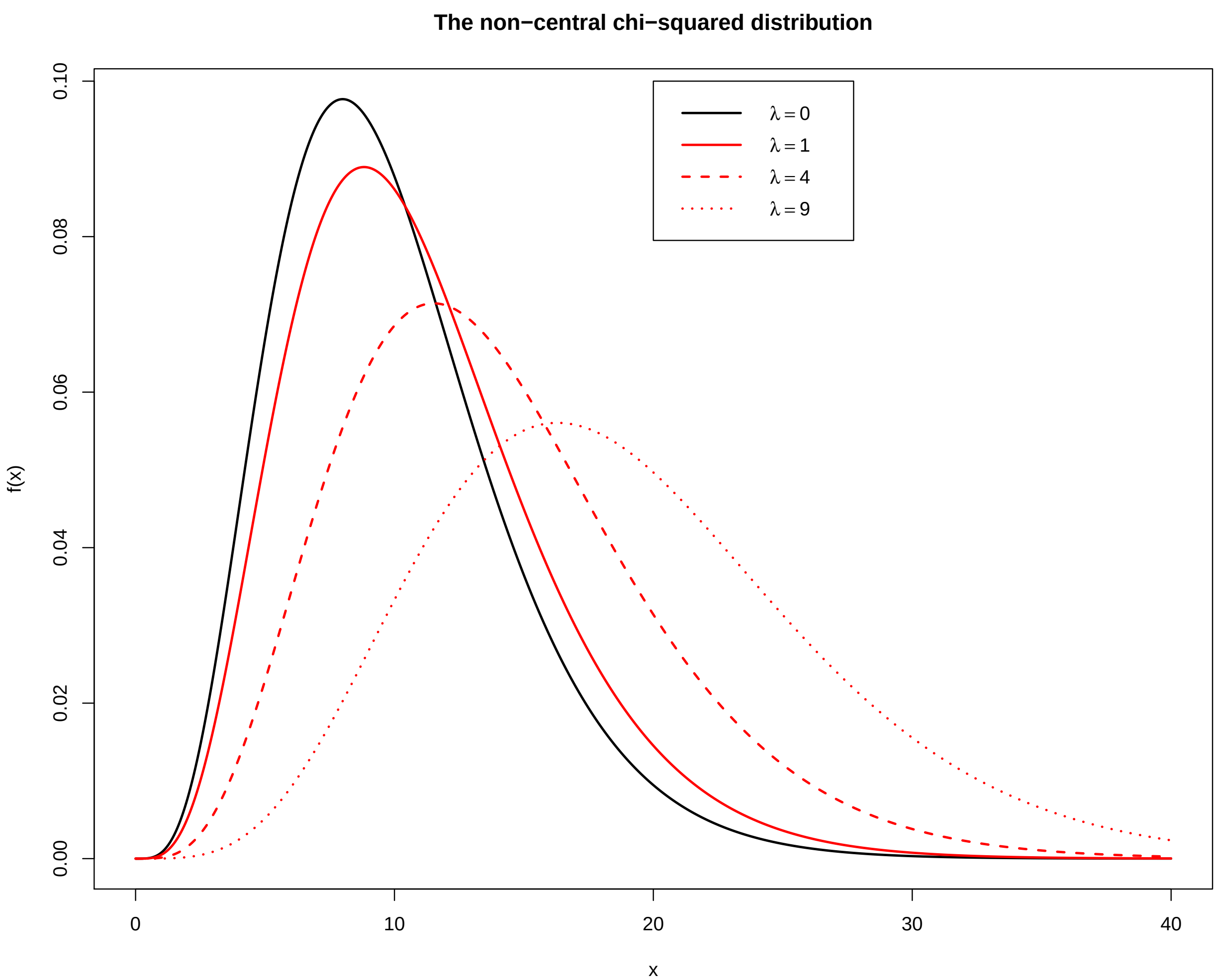
```
#Non-central chi-squared distribution

xv<-seq(0,40,by=0.01)

y0<-dchisq(xv,df=10,ncp=0)

y1<-dchisq(xv,df=10,ncp=1.0^2)
y2<-dchisq(xv,df=10,ncp=2.0^2)
y3<-dchisq(xv,df=10,ncp=3.0^2)


plot(xv,y0,type='l',lwd=2,xlab='x',ylab='f(x)')
lines(xv,y1,col='red',lwd=2)
lines(xv,y2,col='red',lwd=2,lty=2)
lines(xv,y3,col='red',lwd=2,lty=3)
legend(20,0.1,c(expression(lambda==0),expression(lambda==1),
                expression(lambda==4),expression(lambda==9)),
                lty=c(1,1,2,3),lwd=2,col=c('black','red','red','red'))
title('The non-central chi-squared distribution')
```


The non−central chi−squared distribution
λ = 0
λ = 1
λ = 4
λ = 9
f(x)
0.00
0.02
0.04
0.06
0.08
0.10
0
10
20
30
40
x