The F-test for Comparing Full and Reduced Models

- Suppose the normal theory Gauss-Markov model holds.

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

- Suppose $\mathcal{C}(\boldsymbol{X}_0) \subset \mathcal{C}(\boldsymbol{X})$ and we wish to test

$$H_0 : \mathrm{E}(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X}_0) \text{ vs. } \mathrm{H_A} : \mathrm{E}(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X}) \setminus \mathcal{C}(\boldsymbol{X}_0).$$

- The “reduced” model corresponds to the null hypothesis and says that $\mathrm{E}(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X}_0)$, a specified subspace of $\mathcal{C}(\boldsymbol{X})$.
- The “full” model says that $\mathrm{E}(\boldsymbol{y})$ can be anywhere in $\mathcal{C}(\boldsymbol{X})$.
- For example, suppose

$$\boldsymbol{X}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \text{ and } \boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

- In this case, the reduced model says that all 6 observations have the same mean.

- The full model says that there are three groups of two observations. Within each group, observations have the same mean. The three group means may be different from one another.

- For this example,

$$H_0 : \mathrm{E}(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X}_0) \text{ vs. } \mathrm{H_A} : \mathrm{E}(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X}) \setminus \mathcal{C}(\boldsymbol{X}_0)$$

is equivalent to

$$H_0 : \mu_1 = \mu_2 = \mu_3 \text{ vs. } H_A : \mu_i \neq \mu_j, \text{ for some } i \neq j$$

if we use μ_1, μ_2, μ_3 to denote the elements of $\boldsymbol{\beta}$ in the full model, i.e.

$$\boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}.$$

- For the general case, consider the test statistic

$$F = \frac{\boldsymbol{y}'(\boldsymbol{P}_X - \boldsymbol{P}_{X_0})\boldsymbol{y}/\left[\text{rank}(\boldsymbol{X}) - \text{rank}(\boldsymbol{X}_0)\right]}{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_X)\boldsymbol{y}/\left[n - \text{rank}(\boldsymbol{X})\right]}$$

- To show that this statistic has an F distribution, we will use the following fact:

$$\boldsymbol{P}_{X_0}\boldsymbol{P}_X = \boldsymbol{P}_X\boldsymbol{P}_{X_0} = \boldsymbol{P}_{X_0}.$$

- There are many ways to see that this is true.

$$\forall\, \boldsymbol{a} \in \mathbb{R}^n, \qquad \boldsymbol{P}_{\boldsymbol{X}_0}\boldsymbol{a} \in \mathcal{C}(\boldsymbol{X}_0) \subset \mathcal{C}(\boldsymbol{X}).$$

- Thus, $\forall\, \boldsymbol{a} \in \mathbb{R}^n, \quad \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{X}_0}\boldsymbol{a} = \boldsymbol{P}_{\boldsymbol{X}_0}\boldsymbol{a}$

- This implies $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{X}_0} = \boldsymbol{P}_{\boldsymbol{X}_0}$.

- Transposing both sides of this equality and using symmetry of projection matrices yields

$$\boldsymbol{P}_{\boldsymbol{X}_0}\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{P}_{\boldsymbol{X}_0}.$$

- Alternatively,

$$\begin{aligned}\mathcal{C}(\boldsymbol{X}_0) \subset \mathcal{C}(\boldsymbol{X}) \quad &\Longrightarrow \quad \text{Every column of } \boldsymbol{X}_0 \in \mathcal{C}(\boldsymbol{X}) \\ &\Longrightarrow \quad \boldsymbol{P}_X \boldsymbol{X}_0 = \boldsymbol{X}_0.\end{aligned}$$

- Thus,

$$\begin{aligned}\boldsymbol{P}_X \boldsymbol{P}_{X_0} = \boldsymbol{P}_X \boldsymbol{X}_0 (\boldsymbol{X}_0' \boldsymbol{X}_0)^- \boldsymbol{X}_0' \quad &= \quad \boldsymbol{X}_0 (\boldsymbol{X}_0' \boldsymbol{X}_0)^- \boldsymbol{X}_0' \\ &= \quad \boldsymbol{P}_{X_0}.\end{aligned}$$

- This implies that

$$\begin{aligned}(\boldsymbol{P}_X \boldsymbol{P}_{X_0})' = \boldsymbol{P}_{X_0}' \quad &\Longrightarrow \quad \boldsymbol{P}_{X_0}' \boldsymbol{P}_X' = \boldsymbol{P}_{X_0}' \\ &\Longrightarrow \quad \boldsymbol{P}_{X_0} \boldsymbol{P}_X = \boldsymbol{P}_{X_0}. \quad \square\end{aligned}$$

- Alternatively, $\mathcal{C}(\boldsymbol{X}_0) \subset \mathcal{C}(\boldsymbol{X}) \implies \boldsymbol{XB} = \boldsymbol{X}_0$ for some $\boldsymbol{B}$ because every column of $\boldsymbol{X}_0$ must be in $\mathcal{C}(\boldsymbol{X})$.

- Thus,

$$
\begin{aligned}
\boldsymbol{P}_{\boldsymbol{X}_0}\boldsymbol{P}_{\boldsymbol{X}} &= \boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^-\boldsymbol{X}_0'\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^-(\boldsymbol{XB})'\boldsymbol{P}_{\boldsymbol{X}} \\
&= \boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^-\boldsymbol{B}'\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^-\boldsymbol{B}'\boldsymbol{X}' \\
&= \boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^-(\boldsymbol{XB})' = \boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^-\boldsymbol{X}_0' = \boldsymbol{P}_{\boldsymbol{X}_0}. \\
\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{X}_0} &= \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^-\boldsymbol{X}_0' = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{XB}(\boldsymbol{X}_0'\boldsymbol{X}_0)^-\boldsymbol{X}_0' \\
&= \boldsymbol{XB}(\boldsymbol{X}_0'\boldsymbol{X}_0)^-\boldsymbol{X}_0' = \boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^-\boldsymbol{X}_0' = \boldsymbol{P}_{\boldsymbol{X}_0}
\end{aligned}
$$

□

- Note that $\boldsymbol{P_X} - \boldsymbol{P_{X_0}}$ is a symmetric and idempotent matrix:

$$
\begin{aligned}
(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})(\boldsymbol{P_X} - \boldsymbol{P_{X_0}}) &= \boldsymbol{P_X P_X} - \boldsymbol{P_X P_{X_0}} - \boldsymbol{P_{X_0} P_X} + \boldsymbol{P_{X_0} P_{X_0}} \\
&= \boldsymbol{P_X} - \boldsymbol{P_{X_0}} - \boldsymbol{P_{X_0}} + \boldsymbol{P_{X_0}} \\
&= \boldsymbol{P_X} - \boldsymbol{P_{X_0}}
\end{aligned}
$$

Now back to determining the distribution of

$$\frac{\boldsymbol{y}'(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{y} / \left[\text{rank}(\boldsymbol{X}) - \text{rank}(\boldsymbol{X}_0)\right]}{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} / \left[n - \text{rank}(\boldsymbol{X})\right]}.$$

$$\frac{\boldsymbol{y}'(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{y}}{\sigma^2} \sim \chi^2_{\text{rank}(\boldsymbol{X})-\text{rank}(\boldsymbol{X}_0)}(\boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{X}\boldsymbol{\beta}/\sigma^2)$$

because

$$\left(\frac{\boldsymbol{P_X} - \boldsymbol{P_{X_0}}}{\sigma^2}\right)(\sigma^2\boldsymbol{I}) = \boldsymbol{P_X} - \boldsymbol{P_{X_0}}$$

is idempotent and

$$\begin{aligned} \text{rank}(\boldsymbol{P_X} - \boldsymbol{P_{X_0}}) &= tr(\boldsymbol{P_X} - \boldsymbol{P_{X_0}}) = tr(\boldsymbol{P_X}) - tr(\boldsymbol{P_{X_0}}) \\ &= \text{rank}(\boldsymbol{P_X}) - \text{rank}(\boldsymbol{P_{X_0}}) \\ &= \text{rank}(\boldsymbol{X}) - \text{rank}(\boldsymbol{X}_0). \end{aligned}$$

- $\frac{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y}}{\sigma^2} \sim \chi^2_{n-\text{rank}(\boldsymbol{X})}$ by previous work.
- $(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{y}$ is independent of $\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}$ because

$$\begin{aligned}(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})(\sigma^2\boldsymbol{I})(\boldsymbol{I} - \boldsymbol{P_X}) &= \sigma^2(\boldsymbol{P_X} - \boldsymbol{P_X}\boldsymbol{P_X} - \boldsymbol{P_{X_0}} + \boldsymbol{P_{X_0}}\boldsymbol{P_X}) \\ &= \sigma^2(\boldsymbol{P_X} - \boldsymbol{P_X} - \boldsymbol{P_{X_0}} + \boldsymbol{P_{X_0}}) = \boldsymbol{0}.\end{aligned}$$

- Thus, $\boldsymbol{y}'(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{y} = [(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{y}]'\,(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{y}$ is independent of $\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}$.

- Thus, it follows that

$$
\begin{aligned}
F &= \frac{\boldsymbol{y}'(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{y} / [\text{rank}(\boldsymbol{X}) - \text{rank}(\boldsymbol{X}_0)]}{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} / [n - \text{rank}(\boldsymbol{X})]} \\
&\sim F_{\text{rank}(\boldsymbol{X}) - \text{rank}(\boldsymbol{X}_0), n - \text{rank}(\boldsymbol{X})}(\boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{X\beta}/\sigma^2).
\end{aligned}
$$

- If H_0 is true, i.e. if $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X\beta} \in \mathcal{C}(\boldsymbol{X}_0)$, then the noncentrality parameter is 0 because

$$
\begin{aligned}
(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{X\beta} &= \boldsymbol{P_X X\beta} - \boldsymbol{P_{X_0} X\beta} \\
&= \boldsymbol{X\beta} - \boldsymbol{X\beta} = \boldsymbol{0}
\end{aligned}
$$

- In general, the noncentrality parameter quantifies how far the mean of $\boldsymbol{y}$ is from $\mathcal{C}(\boldsymbol{X}_0)$:

$$\boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}_0})\boldsymbol{X}\boldsymbol{\beta}/\sigma^2$$

$$= \boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}_0})'(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}_0})\boldsymbol{X}\boldsymbol{\beta}/\sigma^2$$

$$= \| (\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}_0})\boldsymbol{X}\boldsymbol{\beta} \|^2 /\sigma^2 = \| \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{P}_{\boldsymbol{X}_0}\boldsymbol{X}\boldsymbol{\beta} \|^2 /\sigma^2$$

$$= \| \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{P}_{\boldsymbol{X}_0}\boldsymbol{X}\boldsymbol{\beta} \|^2 /\sigma^2 = \| \mathrm{E}(\boldsymbol{y}) - \boldsymbol{P}_{\boldsymbol{X}_0}\mathrm{E}(\boldsymbol{y}) \|^2 /\sigma^2$$

- Note that

$$\begin{aligned}
\boldsymbol{y}'(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{y} &= \boldsymbol{y}'\left[(\boldsymbol{I} - \boldsymbol{P_{X_0}}) - (\boldsymbol{I} - \boldsymbol{P_X})\right]\boldsymbol{y} \\
&= \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_{X_0}})\boldsymbol{y} - \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} \\
&= SSE_{\text{REDUCED}} - SSE_{\text{FULL}}.
\end{aligned}$$

- Also rank$(\boldsymbol{X})-$rank$(\boldsymbol{X}_0)$

$$\begin{aligned}
&= [n - \text{rank}(\boldsymbol{X}_0)] - [n - \text{rank}(\boldsymbol{X})] \\
&= DFE_{\text{REDUCED}} - DFE_{\text{FULL}}
\end{aligned}$$

where $DFE =$ Degrees of Freedom for Error.

- Thus, the F statistic has the familiar form

$$\frac{(SSE_{\text{REDUCED}} - SSE_{\text{FULL}})/(DFE_{\text{REDUCED}} - DFE_{\text{FULL}})}{SSE_{\text{FULL}}/DFE_{\text{FULL}}}.$$

- It turns out that this reduced vs. full model F test is equivalent to the F test we learned for testing

$$H_0 : \boldsymbol{C\beta} = \boldsymbol{d} \qquad \text{vs.} \qquad H_A : \boldsymbol{C\beta} \neq \boldsymbol{d},$$

where $\boldsymbol{C}\beta$ is testable.

- Proving the equivalence of these F tests is covered in detail in 611.
- We will look at a few examples in 511.