The F-Test for Comparing Reduced vs. Full Models

Model and Hypotheses

Assume the Gauss-Markov Model with normal errors:

smaller model
less complex

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

larger model accommodating more parameters/complexity

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

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

- 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 $\mathcal{C}(\boldsymbol{X})$.

Model Matrix under each Hypothesis

For example, suppose

model matrix for an "intercept only" model

$$X_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \text{ and } X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

regardless of regressors, averaging all y-values is E(y)

- The reduced model says

  we assume same mean $E(y)=\mu$ for all observations

- The full model says

there are 3 distinct means: each group (of size 2) has its own mean

For this example, let μ_1, μ_2, and μ_3 be the elements of $\boldsymbol{\beta}$ in the full model, i.e., $\boldsymbol{\beta} = [\mu_1, \mu_2, \mu_3]^\top$. Then, for the full model,

$$\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X\beta} = \overset{X}{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}} \overset{\beta}{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \end{bmatrix}, \text{ and}$$

mean of y_1 & y_2

⋮

y_5 & y_6 have mean μ_3

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

is equivalent to

$H_0 : \theta = 0$ MCP vs. $H_a : \theta \neq 0$

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

Test Statistic

For the general case, consider the test statistic

non-central $\chi^2_{df=m}$

$$F = \frac{\boxed{\boldsymbol{y}^\top(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{y}}/[\text{rank}(\boldsymbol{X}) - \text{rank}(\boldsymbol{X}_0)]}{\boxed{\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}}/[n - \text{rank}(\boldsymbol{X})]}.$$

m

Central $\chi^2_{df=n-\text{rank}(X)}$

- When the reduced model is correct, the numerator and denominator of the F-statistic are both unbiased estimators of σ^2, so F should be close to 1.

- When the reduced model is not correct, the numerator of the F-statistic is estimating something larger than σ^2, so F should be larger than 1. Thus, values of F much larger than 1 are not consistent with the reduced model being correct.

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

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

There are many ways to see that this fact is true. First,

1

$$\mathcal{C}(\boldsymbol{X}_0) \subset \mathcal{C}(\boldsymbol{X}) \implies \text{Each column of } \boldsymbol{X}_0 \in \mathcal{C}(\boldsymbol{X})$$
$$\implies \boxed{\boldsymbol{P_X X}_0} = \boldsymbol{X}_0.$$

Projecting X_0 onto $C(X)$

Thus,

$$\boldsymbol{P_X P_{X_0}} = \underbrace{\boxed{\boldsymbol{P_X X}_0}}_{X_0}(\boldsymbol{X}_0^\top \boldsymbol{X}_0)^- \boldsymbol{X}_0^\top = \boldsymbol{X}_0(\boldsymbol{X}_0^\top \boldsymbol{X}_0)^- \boldsymbol{X}_0^\top$$
$$= \boldsymbol{P_{X_0}}.$$

This implies that

due symmetry property

$$(\boldsymbol{P_X P_{X_0}})^\top = \boldsymbol{P}_{\boldsymbol{X}_0}^\top \implies \boldsymbol{P}_{\boldsymbol{X}_0}^\top \boldsymbol{P}_{\boldsymbol{X}}^\top = \boldsymbol{P}_{\boldsymbol{X}_0}^\top$$
$$\implies \boldsymbol{P_{X_0} P_X} = \boldsymbol{P_{X_0}}. \quad \square$$

2

Alternatively,

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

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

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

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

Thus, 3

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

□

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

Symmetry:

$$\boxed{(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})^\top} = \boldsymbol{P_X^\top} - \boldsymbol{P_{X_0}^\top} = \boldsymbol{P_X} - \boldsymbol{P_{X_0}}.$$

idempotent:

P_X $\quad P_{X_0}$ $\quad P_{X_0}$

$$(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})(\boldsymbol{P_X} - \boldsymbol{P_{X_0}}) = \boldsymbol{P_X P_X} - \underline{\boldsymbol{P_X P_{X_0}}} - \underline{\boldsymbol{P_{X_0} P_X}}$$

P_{X_0}

$$+\boldsymbol{P_{X_0} P_{X_0}}$$

result from slide 6:

$$= \boldsymbol{P_X} - \boldsymbol{P_{X_0}} - \boldsymbol{P_{X_0}} + \boldsymbol{P_{X_0}}$$

$$\boxed{= \boldsymbol{P_X} - \boldsymbol{P_{X_0}}.}$$

Deriving the Distribution of F

Now back to determining the distribution of

$$F = \frac{\boldsymbol{y}^\top(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{y}/\,[\text{rank}(\boldsymbol{X}) - \text{rank}(\boldsymbol{X}_0)]}{\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}/\,[n - \text{rank}(\boldsymbol{X})]}.$$

m

An important first step is to note that

the σ^2 technically cancel out

$$F = \frac{\boldsymbol{y}^\top\left(\frac{\boldsymbol{P_X} - \boldsymbol{P_{X_0}}}{\sigma^2}\right)\boldsymbol{y}/\,[\text{rank}(\boldsymbol{X}) - \text{rank}(\boldsymbol{X}_0)]}{\boldsymbol{y}^\top\left(\frac{\boldsymbol{I} - \boldsymbol{P_X}}{\sigma^2}\right)\boldsymbol{y}/\,[n - \text{rank}(\boldsymbol{X})]}.$$

Now we can show that the numerator is a chi-squared random variable divided by its degrees of freedom, independent of the denominator, which is a central chi-squared random variable divided by its degrees of freedom. Once we show all these things, we will have established that the statistic F has an F distribution (see prerequisite knowledge material from day 1).

Deriving the Distribution of F

Our main assumption about the model is

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}) \quad \Longrightarrow \quad \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$$

Recall from the prerequisite knowledge material from day 1:

- Suppose $\boldsymbol{\Sigma}$ is an $n \times n$ positive definite matrix.
- Suppose $\boldsymbol{A}$ is an $n \times n$ symmetric matrix of rank m such that $\boldsymbol{A\Sigma}$ is idempotent (i.e., $\boldsymbol{A\Sigma A\Sigma} = \boldsymbol{A\Sigma}$).
- Then $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Longrightarrow \boxed{\boldsymbol{y}^\top \boldsymbol{A} \boldsymbol{y} \sim \chi^2_m(\boldsymbol{\mu}^\top \boldsymbol{A} \boldsymbol{\mu}/2)}$.

Distribution of the Numerator

For the numerator of our F statistic, we have

$$\boldsymbol{\mu} = \boldsymbol{X\beta}, \quad \boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}, \quad \boldsymbol{A} = \boxed{\left(\frac{\boldsymbol{P_X} - \boldsymbol{P_{X_0}}}{\sigma^2}\right),}$$ and

$$\begin{aligned} m &= \operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}\left(\frac{\boldsymbol{P_X} - \boldsymbol{P_{X_0}}}{\sigma^2}\right) = \operatorname{rank}(\boldsymbol{P_X} - \boldsymbol{P_{X_0}}) \\ &= \operatorname{tr}(\boldsymbol{P_X} - \boldsymbol{P_{X_0}}) = \operatorname{tr}(\boldsymbol{P_X}) - \operatorname{tr}(\boldsymbol{P_{X_0}}) \\ &= \operatorname{rank}(\boldsymbol{P_X}) - \operatorname{rank}(\boldsymbol{P_{X_0}}) = \operatorname{rank}(\boldsymbol{X}) - \operatorname{rank}(\boldsymbol{X}_0). \end{aligned}$$

$= m$

(Multiplying by a nonzero constant does not affect the rank of a matrix. Rank is the same as trace for idempotent matrices. Trace of a difference is the same as the difference of traces. The rank of a projection matrix is equal to the rank of the matrix whose column space is projected onto.)

Distribution of the Numerator

To verify that Σ is positive definite, note that for any $\boldsymbol{a} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$,

$$\boldsymbol{a}^\top \boldsymbol{\Sigma} \boldsymbol{a} = \boldsymbol{a}^\top \left(\sigma^2 \boldsymbol{I}\right) \boldsymbol{a} = \sigma^2 \boldsymbol{a}^\top \boldsymbol{a} = \sigma^2 \sum_{i=1}^{n} a_i^2 > 0.$$

To verify that $\boldsymbol{A\Sigma}$ is idempotent, we have

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

$A \Sigma A \Sigma = A \Sigma$

Distribution of the Numerator

Therefore,

Distribution of the Numerator

$$\boldsymbol{y}^\top(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}_0})\boldsymbol{y}/\sigma^2 \sim \chi^2_{(\text{rank}(\boldsymbol{X})-\text{rank}(\boldsymbol{X}_0))}(\theta),$$

where

$$\theta = \underbrace{\frac{1}{2}\boldsymbol{\beta}^\top\boldsymbol{X}^\top\left(\frac{\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}_0}}{\sigma^2}\right)\boldsymbol{X}\boldsymbol{\beta}.}_{(\mu^\top A \mu)/2}$$

Distribution of the Denominator

Denominator:

SSE / df

$$\text{MSE} = \boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} / [n - \text{rank}(\boldsymbol{X})]$$

Distribution of the Denominator

$$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}/\sigma^2 \sim \chi^2_{(n-\text{rank}(\boldsymbol{X}))}$$

- This distributional result holds regardless of whether or not the reduced model is correct.
- This distributional result follows from the same type of argument used to show the distribution of the numerator.

Independence of Numerator and Denominator

By the independence result at the end of the preliminary notes, we can show that $\boldsymbol{y}^\top(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{y}/\sigma^2$ is independent of $\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}/\sigma^2$ because it holds that

$$\underbrace{\left(\frac{\boldsymbol{P_X} - \boldsymbol{P_{X_0}}}{\sigma^2}\right)(\sigma^2\boldsymbol{I})\left(\frac{\boldsymbol{I} - \boldsymbol{P_X}}{\sigma^2}\right)}_{(**)} \overset{!}{=} 0. \qquad (**)$$

Why?

$$\begin{aligned}
(**) &= \frac{1}{\sigma^2}(\boldsymbol{P_X} - \overset{P_X}{\boldsymbol{P_X}\boldsymbol{P_X}} - \boldsymbol{P_{X_0}} + \overset{= P_{X_0}}{\underline{\boldsymbol{P_{X_0}}\boldsymbol{P_X}}}) \\
&= \frac{1}{\sigma^2}(\underline{\boldsymbol{P_X} - \boldsymbol{P_X}} - \underline{\boldsymbol{P_{X_0}} + \boldsymbol{P_{X_0}}}) = \mathbf{0}.
\end{aligned}$$

independence ✓

Distribution of F

Thus, it follows that

Distribution of F

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

$$\sim F_{\text{rank}(\boldsymbol{X})-\text{rank}(\boldsymbol{X}_0),n-\text{rank}(\boldsymbol{X})}(\theta),$$

where

$$\theta = \frac{\boldsymbol{\beta}^\top\boldsymbol{X}^\top(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{X}\boldsymbol{\beta}}{2\sigma^2}.$$

Noncentrality Parameter

- 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 θ is 0 because

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

  Hence,

$$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}/\sigma^2 \sim \chi^2_{\mathrm{rank}(\boldsymbol{X})-\mathrm{rank}(\boldsymbol{X}_0)},$$

  a central χ^2 distr.

end lecture 4
01-30-25

- If H_0 is false and $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X\beta} \notin \mathcal{C}(\boldsymbol{X}_0)$, then $(\boldsymbol{P_X} - \boldsymbol{P_{X_0}})\boldsymbol{X\beta} \neq 0$ and $\theta > 0$. Hence,

$$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}/\sigma^2 \sim \chi^2_{\mathrm{rank}(\boldsymbol{X})-\mathrm{rank}(\boldsymbol{X}_0)}(\theta),$$