The F-Test for Comparing Reduced vs. Full Models

Model and Hypotheses

Assume the Gauss-Markov Model with normal errors:

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

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

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

- 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

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

- The reduced model says
- The full model says

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,

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

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

is equivalent to

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

Test Statistic

For the general case, consider the test statistic

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

- 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,

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

Thus,

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

This implies that

$$\begin{aligned}(\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{X}_0})^\top = \boldsymbol{P}_{\boldsymbol{X}_0}^\top \quad &\Longrightarrow \quad \boldsymbol{P}_{\boldsymbol{X}_0}^\top\boldsymbol{P}_{\boldsymbol{X}}^\top = \boldsymbol{P}_{\boldsymbol{X}_0}^\top \\ &\Longrightarrow \quad \boldsymbol{P}_{\boldsymbol{X}_0}\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{P}_{\boldsymbol{X}_0}. \quad \square\end{aligned}$$

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

Thus,

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

$\square$

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

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

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

Deriving the Distribution of F

Now back to determining the distribution of

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

An important first step is to note that

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

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

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

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 $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Longrightarrow \boldsymbol{y}^\top \boldsymbol{A y} \sim \chi^2_m(\boldsymbol{\mu}^\top \boldsymbol{A\mu}/2)$.

Distribution of the Numerator

For the numerator of our F statistic, we have

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

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

(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

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

Distribution of the Numerator

Therefore,

Distribution of the Numerator

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

where

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

Distribution of the Denominator

Denominator:

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

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

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

Why?

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

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 $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta} \in \mathcal{C}(\boldsymbol{X}_0)$, then the noncentrality parameter θ is 0 because

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

  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.

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

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

Noncentrality Parameter

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

$$
\begin{aligned}
&\boldsymbol{\beta}^\top \boldsymbol{X}^\top (\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}_0}) \boldsymbol{X}\boldsymbol{\beta} \\
&\quad = \; \boldsymbol{\beta}^\top \boldsymbol{X}^\top (\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}_0})^\top (\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}_0}) \boldsymbol{X}\boldsymbol{\beta} \\
&\quad = \; || \, (\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}_0}) \boldsymbol{X}\boldsymbol{\beta} \, ||^2 = \; || \, \boldsymbol{P}_{\boldsymbol{X}} \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{P}_{\boldsymbol{X}_0} \boldsymbol{X}\boldsymbol{\beta} \, ||^2 \\
&\quad = \; || \, \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{P}_{\boldsymbol{X}_0} \boldsymbol{X}\boldsymbol{\beta} \, ||^2 = \; || \, \mathrm{E}(\boldsymbol{y}) - \boldsymbol{P}_{\boldsymbol{X}_0} \mathrm{E}(\boldsymbol{y}) \, ||^2 \, .
\end{aligned}
$$

Note that

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

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

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

Equivalence of F-Tests

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

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

with an appropriately chosen $\boldsymbol{C}$ and $\boldsymbol{d}$.

The equivalence of these tests is proved in STAT 6110.

Example: F-Test for Lack of Linear Fit

Suppose a balanced, completely randomized design is used to assign 1, 2, or 3 units of fertilizer to a total of 9 plots of land.

The yield harvested from each plot is recorded as the response.

Let y_{ij} denote the yield from the jth plot that received i units of fertilizer ($i, j = 1, 2, 3$).

Suppose all yields are independent and $y_{ij} \sim \mathcal{N}(\mu_i, \sigma^2)$ for all $i, j = 1, 2, 3$.

$$\text{If } \boldsymbol{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{31} \\ y_{32} \\ y_{33} \end{bmatrix}, \text{ then } \mathrm{E}(\boldsymbol{y}) = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_3 \end{bmatrix}.$$

Suppose we wish to determine whether there is a linear relationship between the amount of fertilizer applied to a plot and the expected value of a plot's yield.

In other words, suppose we wish to know if there exists real numbers β_1 and β_2 such that

$$\mu_i = \beta_1 + \beta_2(i) \text{ for all } i = 1, 2, 3.$$

Consider testing

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

$$\boldsymbol{X}_0 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Note $H_0 : \mathrm{E}(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X}_0) \iff \exists\, \boldsymbol{\beta} \in \mathbb{R}^2 \ni \mathrm{E}(\boldsymbol{y}) = \boldsymbol{X}_0\boldsymbol{\beta} \iff$

$$\exists \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \in \mathbb{R}^2 \ni \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \beta_1 + \beta_2(1) \\ \beta_1 + \beta_2(1) \\ \beta_1 + \beta_2(1) \\ \beta_1 + \beta_2(2) \\ \beta_1 + \beta_2(2) \\ \beta_1 + \beta_2(2) \\ \beta_1 + \beta_2(3) \\ \beta_1 + \beta_2(3) \\ \beta_1 + \beta_2(3) \end{bmatrix}$$

$\iff \mu_i = \beta_1 + \beta_2(i)$ for all $i = 1, 2, 3.$

Note $\mathrm{E}(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X}) \iff \exists\, \boldsymbol{\beta} \in \mathbb{R}^3 \ni \mathrm{E}(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta} \iff$

$$\exists \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \in \mathbb{R}^3 \ni \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_1 \\ \beta_1 \\ \beta_2 \\ \beta_2 \\ \beta_2 \\ \beta_3 \\ \beta_3 \\ \beta_3 \end{bmatrix}.$$

This condition clearly holds with $\beta_i = \mu_i$ for all $i = 1, 2, 3$.

The alternative hypothesis

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

is equivalent to

H_A : There do not exist $\beta_1, \beta_2 \in \mathbb{R}$ such that

$$\mu_i = \beta_1 + \beta_2(i) \qquad \forall\ i = 1, 2, 3.$$

Because the lack of fit test is a reduced vs. full model F test, we can also obtain this test by testing

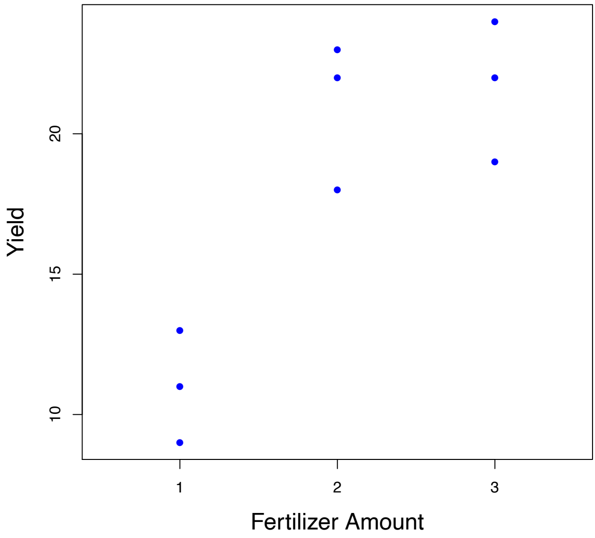
$$H_0 : \boldsymbol{C\beta} = \boldsymbol{d} \quad \text{vs.} \quad H_A : \boldsymbol{C\beta} \neq \boldsymbol{d}$$

for appropriate $\boldsymbol{C}$ and $\boldsymbol{d}$.

$$\boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \qquad \boldsymbol{C} = ? \qquad \boldsymbol{d} = ?$$

R Code and Output

```
> x=rep(1:3,each=3)
> x
[1] 1 1 1 2 2 2 3 3 3
>
> y=c(11,13,9,18,22,23,19,24,22)
>
> plot(x,y,pch=16,col=4,xlim=c(.5,3.5),
+      xlab="Fertilizer Amount",
+      ylab="Yield",axes=F,cex.lab=1.5)
> axis(1,labels=1:3,at=1:3)
> axis(2)
> box()
```


Yield
10
15
20
1
2
3
Fertilizer Amount


```
> X0=model.matrix(~x)
> X0
  (Intercept) x
1           1 1
2           1 1
3           1 1
4           1 2
5           1 2
6           1 2
7           1 3
8           1 3
9           1 3
```

```
> X=model.matrix(~0+factor(x))
> X
  factor(x)1 factor(x)2 factor(x)3
1          1          0          0
2          1          0          0
3          1          0          0
4          0          1          0
5          0          1          0
6          0          1          0
7          0          0          1
8          0          0          1
9          0          0          1
```

```
> proj=function(x){
+   x%*%ginv(t(x)%*%x)%*%t(x)
+ }
>
> library(MASS)
> PX0=proj(X0)
> PX=proj(X)
```

```
> Fstat=(t(y)%*%(PX-PX0)%*%y/1)/
+       (t(y)%*%(diag(rep(1,9))-PX)%*%y/(9-3))
> Fstat
         [,1]
[1,] 7.538462
>
> pvalue=1-pf(Fstat,1,6)
> pvalue
           [,1]
[1,] 0.03348515
```



```
> reduced=lm(y~x)
> full=lm(y~0+factor(x))
>
> rvsf=function(reduced,full)
+ {
+   sser=deviance(reduced)
+   ssef=deviance(full)
+   dfer=reduced$df
+   dfef=full$df
+   dfn=dfer-dfef
+   Fstat=(sser-ssef)/dfn/
+         (ssef/dfef)
+   pvalue=1-pf(Fstat,dfn,dfef)
+   list(Fstat=Fstat,dfn=dfn,dfd=dfef,
+        pvalue=pvalue)
+ }
```

```
> rvsf(reduced,full)
$Fstat
[1] 7.538462

$dfn
[1] 1

$dfd
[1] 6

$pvalue
[1] 0.03348515
```



```
> anova(reduced,full)
Analysis of Variance Table

Model 1: y ~ x
Model 2: y ~ 0 + factor(x)
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1      7 78.222
2      6 34.667  1    43.556 7.5385 0.03349 *
---
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1
```

```
> test=function(lmout,C,d=0){
+   b=coef(lmout)
+   V=vcov(lmout)
+   dfn=nrow(C)
+   dfd=lmout$df
+   Cb.d=C%*%b-d
+   Fstat=drop(
      t(Cb.d)%*%solve(C%*%V%*%t(C))%*%Cb.d/dfn)
+   pvalue=1-pf(Fstat,dfn,dfd)
+   list(Fstat=Fstat,pvalue=pvalue)
+ }
> test(full,matrix(c(1,-2,1),nrow=1))
$Fstat
[1] 7.538462
$pvalue
[1] 0.03348515
```

SAS Code and Output

```
data d;
  input x y;
  cards;
1 11
1 13
1 9
2 18
2 22
2 23
3 19
3 24
3 22
;
run;
```

```
proc glm;
  class x;
  model y=x;
  contrast 'Lack of Linear Fit' x 1 -2 1;
run;
```


The SAS System

The GLM Procedure

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	214.2222222	107.1111111	18.54	0.0027
Error	6	34.6666667	5.7777778		
Corrected Total	8	248.8888889			

R-Square	Coeff Var	Root MSE	y Mean
0.860714	13.43684	2.403701	17.88889

Source	DF	Type I SS	Mean Square	F Value	Pr > F
x	2	214.2222222	107.1111111	18.54	0.0027

Source	DF	Type III SS	Mean Square	F Value	Pr > F
x	2	214.2222222	107.1111111	18.54	0.0027

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Lack of Linear Fit	1	43.55555556	43.55555556	7.54	0.0335