# General Linear Hypothesis Test

November 1, 2021

**Notation:** Write $n \times (p+1)$ matrix $\mathbf{X} = (\mathbf{1}, \mathbf{x}_1, \cdots, \mathbf{x}_p)$, where $\mathbf{1}$ is $n$-dimensional vector with all entries being 1 and $\mathbf{x}_j \in \mathbb{R}^n$. Denote the column space of $\mathbf{X}$ by $\mathcal{C}_{\mathbf{X}}$, i.e. $\mathcal{C}_{\mathbf{X}} = \{c_0\mathbf{1} + c_1\mathbf{x}_1 + \cdots + c_p\mathbf{x}_p : c_0, c_1, \ldots, c_p \in \mathbb{R}\}$. For $\mathbf{Y} \in \mathbb{R}^n$, let $\Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}})$ denote the projection of $\mathbf{Y}$ onto $\mathcal{C}_{\mathbf{X}}$. Note $\mathbf{X}$ is full rank as in usual assumption. Suppose $\mathbf{A}$ is linear subspace in $\mathbb{R}^k$. Let $\mathbf{A}^{\perp} = \{\mathbf{x} \in \mathbf{R}^n : \mathbf{x}^t\mathbf{a} = \mathbf{0} \text{ for all } \mathbf{a} \in \mathbf{A}\}$.

Let $\mathbf{A}$ be $(p+1) \times q$ matrix of full column rank and $\mathbf{c}$ be $q$-dimensional vector($q < p+1$). Suppose we want to test

$$H_0 : \mathbf{A}^t\boldsymbol{\beta} = \mathbf{c} \text{ versus } H_1 : \mathbf{A}^t\boldsymbol{\beta} \neq \mathbf{c}$$

Under the constraint $\mathbf{A}^t\boldsymbol{\beta} = \mathbf{c}$, we have to find minimizer of $g(\boldsymbol{\beta}) = ||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||^2$. Let $\hat{\boldsymbol{\beta}}_r = \underset{\mathbf{A}^t\boldsymbol{\beta}=\mathbf{c}}{\arg\min} ||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||^2$. We first consider the case when $\mathbf{c} = \mathbf{0}$.

**Claim:** Let $\mathbf{Z}$ be $q \times n$ matrix. Regard $\mathbf{Z}$ as a linear map $\mathbf{Z} : \mathbb{R}^n \to \mathbb{R}^q$. Then, $\text{Null}(\mathbf{Z}) = \text{Range}(\mathbf{Z}^t)^{\perp}$ so that $\mathbb{R}^n = \text{Range}(\mathbf{Z}^t) \oplus \text{Null}(\mathbf{Z})$.

*Proof.* Take any $\mathbf{v} \in \mathbb{R}^q$. If $\mathbf{u} \in \text{Null}(\mathbf{Z})$, $\mathbf{u}^t\mathbf{Z}^t\mathbf{v} = (\mathbf{Z}\mathbf{u})^t\mathbf{v} = \mathbf{0}^t\mathbf{v} = \mathbf{0}$. Hence $\text{Null}(\mathbf{Z}) \subseteq \text{Range}(\mathbf{Z}^t)^{\perp}$. Now choose any $\mathbf{w} \in \text{Range}(\mathbf{Z}^t)^{\perp}$. Then $\mathbf{w}^t(\mathbf{Z}^t\mathbf{v}) = (\mathbf{Z}\mathbf{w})^t\mathbf{v} = 0$ for any $\mathbf{v} \in \mathbb{R}^q$. Since this holds for all $\mathbf{v} \in \mathbf{R}^q$, $\mathbf{Z}\mathbf{w} = 0$ and thus $\mathbf{w} \in \text{Null}(\mathbf{Z})$, which implies $\text{Range}(\mathbf{Z}^t)^{\perp} \subseteq \text{Null}(\mathbf{Z})$.Therefore, we conclude that $\text{Null}(\mathbf{Z}) = \text{Range}(\mathbf{Z}^t)^{\perp}$. $\square$

Assuming $\mathbf{c} = \mathbf{0}$, $\mathbf{A}^t\boldsymbol{\beta} = \mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$. Note that because $\mathbf{A}$ and $\mathbf{X}$ are full rank, $\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}$ is invertible. Using the result of claim, one can clearly see that $\mathbf{X}\hat{\boldsymbol{\beta}}_r$ is the projection of $\mathbf{Y}$ onto

$$\mathcal{C}_{\mathbf{X}} \cap \text{null}(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t) = \mathcal{C}_{\mathbf{X}} \cap \text{Range}(\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{\perp} = \mathcal{C}_{\mathbf{X}} \cap \mathcal{C}^{\perp}_{\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}}$$

Since $\mathcal{C}_{\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}} \subseteq \mathcal{C}_{\mathbf{X}}$,

$$\begin{aligned}
\mathbf{X}\hat{\boldsymbol{\beta}}_r &= \Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}} \cap \mathcal{C}^{\perp}_{\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}}) \\
&= \Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}}) - \Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}} \cap \mathcal{C}_{\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}}) \\
&= \Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}}) - \Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}}) \\
&= \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-1}\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} \\
&= \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-1}\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}
\end{aligned}$$

Thus, we obtain $\hat{\boldsymbol{\beta}}_r = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} - (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-1}\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} \cdots (*)$. Now, we generalize this result for all $\mathbf{c}$. Since $\mathbf{A}^t\boldsymbol{\beta} = \mathbf{A}^t\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c}$, $\mathbf{A}^t(\boldsymbol{\beta} - \mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c}) = \mathbf{0}$, if we let $\boldsymbol{\gamma} = \boldsymbol{\beta} - \mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c}$,

$$||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||^2 = ||\mathbf{Y} - \mathbf{X}(\boldsymbol{\gamma} + \mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c})||^2 = ||\mathbf{Y} - \mathbf{X}\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c} - \mathbf{X}\boldsymbol{\gamma}||^2$$

with $\mathbf{A}^t\boldsymbol{\gamma} = \mathbf{0}$. Because $\hat{\boldsymbol{\beta}}_r = \hat{\boldsymbol{\gamma}} + \mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c}$, it sufficies to find $\hat{\boldsymbol{\gamma}}$, where $\hat{\boldsymbol{\gamma}} = \underset{\mathbf{A}^t\boldsymbol{\gamma}=\mathbf{0}}{\arg\min} ||\mathbf{Y} - \mathbf{X}\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c} - \mathbf{X}\boldsymbol{\gamma}||^2$. But $\hat{\boldsymbol{\gamma}}$ can be easily found by replacing $\mathbf{Y}$ with $\mathbf{Y} - \mathbf{X}\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c}$ in (*). Hence,

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_r &= \hat{\gamma} + \mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c} \\
&= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t(\mathbf{Y} - \mathbf{X}\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c}) - (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-1}\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t(\mathbf{Y} - \mathbf{X}\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c}) \\
&\quad + \mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c} \\
&= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} - \mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c} - (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-1}\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t(\mathbf{Y} - \mathbf{X}\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c}) \\
&\quad + \mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c} \\
&= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} - (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-1}\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t(\mathbf{Y} - \mathbf{X}\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c}) \\
&= \hat{\boldsymbol{\beta}} - (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-1}\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t(\mathbf{Y} - \mathbf{X}\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c})
\end{aligned}
$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$. To derive test, note that if $\mathbf{A}^t\boldsymbol{\beta} = \mathbf{c}$ is not true, $\mathbf{R}(\boldsymbol{\beta}_{-r}|\boldsymbol{\beta}_r) = ||\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_r||^2 - ||\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}||^2$ tends to get larger. Let $\hat{\mathbf{u}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-1}\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t(\mathbf{Y}-\mathbf{X}\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c})$.

$$
\begin{aligned}
\mathbf{R}(\boldsymbol{\beta}_{-r}|\boldsymbol{\beta}_r) &= ||\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}\hat{\mathbf{u}}||^2 - ||\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}||^2 \\
&= ||\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}||^2 + 2(\mathbf{X}\hat{\mathbf{u}})^t(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + ||\mathbf{X}\hat{\mathbf{u}}||^2 - ||\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}||^2 \\
&= 2\hat{\mathbf{u}}^t\mathbf{X}^t(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + ||\mathbf{X}\hat{\mathbf{u}}||^2 = 2\hat{\mathbf{u}}^t\mathbf{X}^t(\mathbf{I} - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t)\mathbf{Y} + ||\mathbf{X}\hat{\mathbf{u}}||^2 \\
&= 2\hat{\mathbf{u}}^t(\mathbf{X}^t - \mathbf{X}^t)\mathbf{Y} + ||\mathbf{X}\hat{\mathbf{u}}||^2 = ||\mathbf{X}\hat{\mathbf{u}}||^2
\end{aligned}
$$

.
With simple calculation,

$$
\begin{aligned}
||\mathbf{X}\hat{\mathbf{u}}||^2 &= (\mathbf{Y} - \mathbf{X}\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c})^t\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-1}\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t(\mathbf{Y} - \mathbf{X}\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c}) \\
&= (\mathbf{A}^t\hat{\boldsymbol{\beta}} - \mathbf{c})^t(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-1}(\mathbf{A}^t\hat{\boldsymbol{\beta}} - \mathbf{c}) = ||(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-\frac{1}{2}}(\mathbf{A}^t\hat{\boldsymbol{\beta}} - \mathbf{c})||^2
\end{aligned}
$$

We know that $\hat{\boldsymbol{\beta}} \sim N_{p+1}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^t\mathbf{X})^{-1})$. Hence, $\mathbf{A}^t\hat{\boldsymbol{\beta}} \sim N_q(\mathbf{A}^t\boldsymbol{\beta}, \sigma^2\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})$. Under $H_0$, $\mathbf{A}^t\hat{\boldsymbol{\beta}} \sim N_q(\mathbf{c}, \sigma^2\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})$. So $(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-\frac{1}{2}}(\mathbf{A}^t\hat{\boldsymbol{\beta}}-\mathbf{c})/\sigma \sim N_q(\mathbf{0}_q, \mathbf{I}_q)$ and thus $\mathbf{R}(\boldsymbol{\beta}_{-r}|\boldsymbol{\beta}_r)/\sigma^2 \sim \chi^2(q)$.

In the lecture, we've seen that $\text{SSE}/\sigma^2 = ||\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}||^2/\sigma^2 \sim \chi^2(n-p-1)$. By direct computation, it can be verified that

$$
\begin{aligned}
||\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}||^2 &= \mathbf{Y}^t(\mathbf{I} - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t)\mathbf{Y} \\
&= (\mathbf{Y} - \mathbf{X}\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c})^t(\mathbf{I} - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t)(\mathbf{Y} - \mathbf{X}\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{c})
\end{aligned}
$$

Recall that if $\mathbf{z} \sim N_k(\mu, \Sigma)$ and $\mathbf{P}, \mathbf{Q}$ are $k \times k$ symmetric, idempotent matrices, $\mathbf{z}^t\mathbf{P}\mathbf{z}$ and $\mathbf{z}^t\mathbf{Q}\mathbf{z}$ are independent if and only if $\mathbf{P}\Sigma\mathbf{Q} = \mathbf{0}$. Because

$$
\begin{aligned}
&\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-1}\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t(\mathbf{I} - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t) \\
&= \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-1}\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}(\mathbf{X}^t - \mathbf{X}^t) = \mathbf{0}
\end{aligned}
$$

and $\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A}(\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{A})^{-1}\mathbf{A}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$, $\mathbf{I}-\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ are symmetric, idempotent, $\mathbf{R}(\boldsymbol{\beta}_{-r}|\boldsymbol{\beta}_r)/\sigma^2$ and $\text{SSE}/\sigma^2$ are independent. Therefore,

$$
F \equiv \frac{\frac{\mathbf{R}(\boldsymbol{\beta}_{-r}|\boldsymbol{\beta}_r)}{q\sigma^2}}{\frac{\text{SSE}}{(n-p-1)\sigma^2}} = \frac{\mathbf{R}(\boldsymbol{\beta}_{-r}|\boldsymbol{\beta}_r)/q}{\text{SSE}/(n-p-1)} \sim F(q, n-p-1)
$$

So we reject $H_0$ if $F > F_\alpha(q, n-p-1)$.