Hypothesis Testing

Throughout, we assume the normal-error linear model that is based on the model

$$y = \beta_1 x_1 + \cdots + \beta_p x_p + \text{noise}.$$

[Note the slight change in the notation.]

1. A test for one parameter

Suppose we want to test to see whether or not the $(\ell + 1)$st x-variable has a [linear] effect on the y variable. Of course, $1 \leq \ell \leq p$, so we are really testing the statistical hypothesis

$$H_0 : \ \beta_\ell = 0.$$

Since $\widehat{\boldsymbol{\beta}} \sim \mathrm{N}_p\left(\boldsymbol{\beta}\,, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$, it follows that

$$\widehat{\beta}_\ell \sim \mathrm{N}\left(\beta_\ell\,, \sigma^2\left[(\mathbf{X}'\mathbf{X})^{-1}\right]_{\ell,\ell}\right).$$

Because S is independent of $\widehat{\boldsymbol{\beta}}$ and hence $\widehat{\beta}_\ell$, and since $S^2/\sigma^2 \sim \chi^2_{n-p}/(n-p)$,

$$\frac{\widehat{\beta}_\ell - \beta_\ell}{S\sqrt{\left[(\mathbf{X}'\mathbf{X})^{-1}\right]_{\ell,\ell}}} = \frac{\sigma}{S} \cdot \frac{\widehat{\beta}_\ell - \beta_\ell}{\sigma\sqrt{\left[(\mathbf{X}'\mathbf{X})^{-1}\right]_{\ell,\ell}}} \sim t_{n-p}.$$

Therefore, it is now a routine matter to set up a t-test for $H_0 : \ \beta_\ell = 0$. As usual, testing has implications that are unattractive; it is much better to present a confidence interval [which you can then use for a test if you want, any way]: A $(1-\alpha) \times 100\%$ confidence interval for β_ℓ is

$$\widehat{\beta}_\ell \pm S t^{\alpha/2}_{n-p}\sqrt{\left[(\mathbf{X}'\mathbf{X})^{-1}\right]_{\ell,\ell}}.$$

If you really insist on performing a level-α test for β_ℓ, it suffices to check to see if this confidence interval contains 0. If 0 is not in the confidence interval then you reject. Otherwise, you do nothing.

2. Least-squares estimates for contrasts

We wish to study a more general problem. Recall that our model has the form

$$y = \beta_1 x_1 + \beta_1 x_1 + \cdots + \beta_p x_p + \text{noise}.$$

If we sample then the preceding becomes $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, as before. As part of model verification, we might ask to see if $(x_i)_{i\in J}$ should be excised from the model, where $J := \{\ell, \ell+1, \ldots, r\}$ is a subset of the index $\{1, \ldots, n\}$. In other words, we ask

$$H_0: \ \beta_\ell = \cdots = \beta_r = 0.$$

Note that we can translate the preceding, using the language of matrix analysis, as $H_0: \ \boldsymbol{A\beta} = \boldsymbol{0}$, where

$$\boldsymbol{A} := \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix},$$

where the identity matrix in the middle is $(r-\ell+1)\times(r-\ell+1)$; it starts on position (ℓ, ℓ) and runs $r-\ell$ units in rows and in columns.

Now we ask a slightly more general question [it pays to do this, as it turns out]: Suppose $\boldsymbol{A}$ is a $q \times p$ matrix of full rank $q \le p$, and we are interested in testing the hypothesis,

$$H_0: \ \boldsymbol{A\beta} = \boldsymbol{0}. \tag{1}$$

The first question to ask is, “how can we estimate $\boldsymbol{\beta}$”? The answer is given to us by the principle of least squares: We write [as before]

$$\boldsymbol{Y} = \boldsymbol{\theta} + \boldsymbol{\varepsilon}, \qquad \text{where } \boldsymbol{\theta} := \boldsymbol{X\beta},$$

and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$ are mean-zero random variables, and first find the least-squares estimate $\widehat{\boldsymbol{\theta}}_{H_0}$ of $\boldsymbol{\theta}$, under the assumption that H_0 is valid. That is, we seek to minimize $\|\boldsymbol{Y} - \boldsymbol{Xb}\|^2$ over all p-vectors $\boldsymbol{b}$ such that $\boldsymbol{Ab} = \boldsymbol{0}$. The optimal value yields $\widehat{\boldsymbol{\theta}}_{H_0} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{H_0}$.Then we obtain $\widehat{\boldsymbol{\beta}}_{H_0}$ by noticing that if X has full rank, then $\widehat{\boldsymbol{\beta}}_{H_0} = (X'X)^{-1}X'\widehat{\boldsymbol{\theta}}_{H_0}$.

Now it follows by differentiation [or just geometrically] that $\widehat{\boldsymbol{\theta}}_{H_0}$ is the projection of $\boldsymbol{Y}$ onto the subspace $\mathcal{G}$ of all vectors of the form $\boldsymbol{\vartheta} = \boldsymbol{Xb}$ that satisfy $\boldsymbol{Ab} = \boldsymbol{0}$, where $\boldsymbol{b}$ is a p-vector. We can simplify this description a little when $\boldsymbol{X}$ has full rank. Note that whenever $\boldsymbol{\vartheta} = \boldsymbol{Xb}$, we can solve to get $\boldsymbol{b} = (X'X)^{-1}X'\boldsymbol{\vartheta}$. Therefore, it follows that—when $\boldsymbol{X}$

has full rank— $\widehat{\boldsymbol{\theta}}_{H_0}$ is the projection of the observations vector $\boldsymbol{Y}$ onto the subspace $\mathcal{G}$ of all vectors of the form $\boldsymbol{\vartheta}$ that satisfy

$$\boldsymbol{A}_1\boldsymbol{\vartheta} = \mathbf{0}, \quad \text{where} \quad \boldsymbol{A}_1 := \boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

In other words, $\mathcal{G}$ is the subspace of $\mathcal{C}(\mathbf{X})$, whose every element $\boldsymbol{\vartheta}$ is orthogonal to every row of $\boldsymbol{A}_1$. In symbols,

$$\mathcal{G} = \mathcal{C}(\mathbf{X}) \cap \left[\mathcal{C}(\boldsymbol{A}_1')\right]^{\perp}.$$

Because $\boldsymbol{A}_1'\boldsymbol{b} = \mathbf{X}\boldsymbol{c}$ for $\boldsymbol{c} := (\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}\boldsymbol{b}$, it follows that $\mathcal{C}(\boldsymbol{A}_1')$ is a subspace of $\mathcal{C}(\mathbf{X})$. Therefore, we can apply the Pytheagorean property to see that

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{H_0} &= \boldsymbol{P}_{\mathcal{G}}\boldsymbol{Y} = \boldsymbol{P}_{\mathcal{C}(\mathbf{X})\cap[\mathcal{C}(\boldsymbol{A}_1')]^{\perp}}\boldsymbol{Y} \\ &= \boldsymbol{P}_{\mathcal{C}(\mathbf{X})}\boldsymbol{Y} - \boldsymbol{P}_{\mathcal{C}(\boldsymbol{A}_1')}\boldsymbol{Y} \\ &= \widehat{\boldsymbol{\theta}} - \boldsymbol{A}_1'(\boldsymbol{A}_1\boldsymbol{A}_1')^{-1}\boldsymbol{A}_1\boldsymbol{Y}.\end{aligned}$$

Now

$$\boldsymbol{A}_1\boldsymbol{A}_1' = \boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}' = \boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'.$$

Therefore,

$$\widehat{\boldsymbol{\theta}}_{H_0} = \widehat{\boldsymbol{\theta}} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\left[\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{Y}.$$

Aside: How do we know that $\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'$ is nonsingular? Note that $\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'$ is positive semidefinite. Now $\mathbf{X}'\mathbf{X}$ is positive definite; therefore, so is its inverse. Therefore, we can write $(\mathbf{X}'\mathbf{X})^{-1} = \boldsymbol{B}^2 = \boldsymbol{B}\boldsymbol{B}'$, where $\boldsymbol{B} := (\mathbf{X}'\mathbf{X})^{-1/2}$ is the square root of $(\mathbf{X}'\mathbf{X})^{-1}$. In this way we find that the rank of $\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'$ is the same as the rank of $\boldsymbol{A}\boldsymbol{A}'$. Since $\boldsymbol{A}$ has full rank, $\boldsymbol{A}\boldsymbol{A}'$ is invertible. Equivalently, full rank. Equivalently, $\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'$ is a full-rank positive definite matrix; hence nonsingular.

The vector $\widehat{\boldsymbol{\theta}}_{H_0} := \widehat{\boldsymbol{Y}}_{H_0} = \mathbf{X}\widehat{\boldsymbol{\beta}}_{H_0}$ is the vector of fitted values, assuming that H_0 is correct. Therefore, the least-squares estimate for $\boldsymbol{\beta}$—under H_0—is

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{H_0} &:= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\boldsymbol{\theta}}_{H_0} = \widehat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\left[\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{Y} \\ &= \widehat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\left[\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{A}\widehat{\boldsymbol{\beta}} \\ &= \left(\boldsymbol{I} - (\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\left[\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{A}\right)\widehat{\boldsymbol{\beta}}.\end{aligned}$$

This can be generalized further as follows: Suppose we wish to test

$$H_0: \ \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{c},$$

where $\boldsymbol{c}$ is a known q-vector [we just studied this in the case that $\boldsymbol{c} = \mathbf{0}$]. Then we reduce the problem to the previous one as follows: First find a known p-vector $\boldsymbol{\beta}_0$ such that $\boldsymbol{A}\boldsymbol{\beta}_0 = \boldsymbol{c}$. Then, create a new parametrization of our problem by setting

$$\boldsymbol{\gamma} := \boldsymbol{\beta} - \boldsymbol{\beta}_0,$$

and

$$\widetilde{\boldsymbol{Y}} := \boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \text{equivalently} \quad \widetilde{\boldsymbol{Y}} := \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}_0.$$

Since $\boldsymbol{A}\boldsymbol{\gamma} = \boldsymbol{0}$, we know the least-squares estimate $\widehat{\boldsymbol{\gamma}}_{H_0}$ is given by

$$\widehat{\boldsymbol{\gamma}}_{H_0} = \left(\boldsymbol{I} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{A}\right)\widehat{\boldsymbol{\gamma}},$$

where

$$\widehat{\boldsymbol{\gamma}} := (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widetilde{\boldsymbol{Y}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} - \boldsymbol{\beta}_0 = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0.$$

In other words,

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{H_0} - \boldsymbol{\beta}_0 &= \left(\boldsymbol{I} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{A}\right)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\\
&= \left(\boldsymbol{I} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{A}\right)\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{A}\boldsymbol{\beta}_0\\
&= \left(\boldsymbol{I} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{A}\right)\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}\left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{c}.
\end{aligned}$$

In this way, we have discovered the following:

Theorem 1. *Consider once again the general linear model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$. If $\mathbf{A}_{q\times p}$ and $\mathbf{c}_{q\times 1}$ are known, and $\mathbf{A}$ has full rank $q \le p$, then the least-squares estimate for $\boldsymbol{\beta}$—under the null hypothesis $H_0:\ \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{c}$—is*

$$\widehat{\boldsymbol{\beta}}_{H_0} = \boldsymbol{\Theta}\widehat{\boldsymbol{\beta}} + \boldsymbol{\mu},$$

where

$$\boldsymbol{\Theta} := \left(\boldsymbol{I} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{A}\right), \tag{2}$$

and

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{c}) := (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{c}, \tag{3}$$

provided that X has full rank.

3. The normal model

Now consider the same problem under the normal model. That is, we consider $H_0:\ \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{c}$ under the assumption that $\boldsymbol{\varepsilon} \sim \mathrm{N}_p(\boldsymbol{0}\,,\sigma^2\boldsymbol{I})$.

Theorem 2. *Consider the normal-error linear model* $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$. *If* $\boldsymbol{A}_{q\times p}$ *and* $\boldsymbol{c}_{q\times 1}$ *are known, and* $\boldsymbol{A}$ *has full rank* $q \le p$, *then the least-squares estimate for* $\boldsymbol{\beta}$*—under the null hypothesis* $H_0: \boldsymbol{A\beta} = \boldsymbol{c}$*—satisfies*

$$\widehat{\boldsymbol{\beta}}_{H_0} \sim \mathrm{N}_p\left(\boldsymbol{\beta}\,,\sigma^2\boldsymbol{\Theta}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\Theta}'\right),$$

provided that $\boldsymbol{X}$ *has full rank.*

Indeed, since

$$\widehat{\boldsymbol{\beta}} \sim \mathrm{N}_p\left(\boldsymbol{\beta}\,,\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) \quad \text{and} \quad \widehat{\boldsymbol{\beta}}_{H_0} = \boldsymbol{\Theta}\widehat{\boldsymbol{\beta}} + \boldsymbol{\mu},$$

it follows that

$$\widehat{\boldsymbol{\beta}}_{H_0} \sim \mathrm{N}_p\left(\boldsymbol{\Theta\beta} + \boldsymbol{\mu}\,,\sigma^2\boldsymbol{\Theta}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\Theta}'\right).$$

Therefore, it remains to check that $\boldsymbol{\Theta\beta} + \boldsymbol{\mu} = \boldsymbol{\beta}$ when $\boldsymbol{A\beta} = \boldsymbol{c}$. But this is easy to see directly.

Next we look into inference for σ^2. Recall that our estimation of σ^2 was based on $\mathrm{RSS} := \|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2$. Under H_0, we do the natural thing and estimate σ^2 instead by

$$\begin{aligned}\mathrm{RSS}_{H_0} &:= \left\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{H_0}\right\|^2 \\ &= \left\|\underbrace{\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}}_{\mathcal{T}_1} - \underbrace{\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}\left[\boldsymbol{c} - \boldsymbol{A}\widehat{\boldsymbol{\beta}}\right]}_{\mathcal{T}_2}\right\|^2.\end{aligned}$$

I claim that $\mathcal{T}_2$ is orthogonal to $\mathcal{T}_1$; indeed,

$$\begin{aligned}\mathcal{T}_2'\mathcal{T}_1 &= \left[\boldsymbol{c} - \boldsymbol{A}\widehat{\boldsymbol{\beta}}\right]'\left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{A}\overbrace{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}}^{\widehat{\boldsymbol{\beta}}} \\ &\qquad - \left[\boldsymbol{c} - \boldsymbol{A}\widehat{\boldsymbol{\beta}}\right]'\left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}\boldsymbol{A}\underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}}_{\boldsymbol{I}}\widehat{\boldsymbol{\beta}} \\ &= 0.\end{aligned}$$

Therefore, the Pythagorean property tells us that

$$\begin{aligned}\mathrm{RSS}_{H_0} &= \left\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right\|^2 + \|\mathcal{T}_2\|^2 \\ &= \mathrm{RSS} + \|\mathcal{T}_2\|^2.\end{aligned}$$

Next we compute

$$\|\mathcal{T}_2\|^2 = \mathcal{T}_2'\mathcal{T}_2$$

$$= \left[\boldsymbol{c} - \boldsymbol{A}\widehat{\boldsymbol{\beta}}\right]' \left[\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\right]^{-1} \underbrace{\underbrace{\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1} \underbrace{\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}}_{\boldsymbol{I}}\boldsymbol{A}'}_{\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'} \left[\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\right]^{-1}}_{\boldsymbol{I}} \left[\boldsymbol{c} - \boldsymbol{A}\widehat{\boldsymbol{\beta}}\right]$$

$$= \left[\boldsymbol{c} - \boldsymbol{A}\widehat{\boldsymbol{\beta}}\right]' \left[\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\right]^{-1} \left[\boldsymbol{c} - \boldsymbol{A}\widehat{\boldsymbol{\beta}}\right].$$

In other words,

$$\mathrm{RSS}_{H_0} = \mathrm{RSS} + \left[\boldsymbol{c} - \boldsymbol{A}\widehat{\boldsymbol{\beta}}\right]' \left[\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\right]^{-1} \left[\boldsymbol{c} - \boldsymbol{A}\widehat{\boldsymbol{\beta}}\right]. \tag{4}$$

Moreover, the two terms on the right-hand side are independent because $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$—hence $\widehat{\boldsymbol{\beta}}$ and $\mathrm{RSS} = \|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2$—are independent. Now we know the distribution of $\mathrm{RSS} := (n-p)S^2 \sim \sigma^2(n-p)\chi^2_{n-p}$. Therefore, it remains to find the distribution of the second term on the right-hand side of (4). But

$$\boldsymbol{A}\widehat{\boldsymbol{\beta}} \sim \mathrm{N}_q\left(\boldsymbol{A}\boldsymbol{\beta}, \sigma^2\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\right) \overset{H_0}{=} \mathrm{N}_q\left(\boldsymbol{c}, \sigma^2 \underbrace{\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'}_{:=\boldsymbol{\Sigma}}\right).$$

Therefore, $\boldsymbol{Z} := \sigma^{-1}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}) \sim \mathrm{N}_q(\boldsymbol{0}, \boldsymbol{I}_{q\times q})$. Also, we can write the second term on the right-hand side of (4) as

$$\left[\boldsymbol{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}\right]' \left[\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\right]^{-1} \left[\boldsymbol{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}\right] = \sigma^2\boldsymbol{Z}'\boldsymbol{Z} = \sigma^2\|\boldsymbol{Z}\|^2 \sim \sigma^2\chi^2_q.$$

Let us summarize our efforts.

Theorem 3. *Consider normal–error linear model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Suppose $\boldsymbol{A}_{q\times p}$ and $\boldsymbol{c}_{q\times 1}$ are known, and $\boldsymbol{A}$ has full rank $q \le p$. Then under the null hypothesis $H_0:\ \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{c}$, we can write*

$$\mathrm{RSS}_{H_0} = \mathrm{RSS} + W,$$

provided that $\mathbf{X}$ has full rank, where RSS *and W are independent, we recall that* $\mathrm{RSS} \sim \sigma^2(n-p)\chi^2_{n-p}$, *and $W \sim \sigma^2\chi^2_q$. In particular,*

$$\frac{(\mathrm{RSS}_{H_0} - \mathrm{RSS})/q}{\mathrm{RSS}/(n-p)} \overset{H_0}{\sim} \frac{\chi^2_q/q}{\chi^2_{n-p}/(n-p)} \qquad \text{[the two } \chi^2\text{'s are independent]}$$

$$= F_{q,n-p}.$$

See your textbook for the distribution of this test statistic under the alternative [this is useful for power computations]. The end result is a "noncentral F distribution."

4. Examples

1. A measurement-error model. For our first example, consider a random sample $Y_1, \ldots, Y_n \sim \mathrm{N}(\mu\,, \sigma^2)$; equivalently,

$$Y_i = \mu + \varepsilon_i \qquad (1 \le i \le n),$$

where $\boldsymbol{\varepsilon} \sim \mathrm{N}_n(\mathbf{0}\,, \sigma^2 \boldsymbol{I})$. This is a linear model with $p = 1$, $\boldsymbol{X} := \mathbf{1}_{n\times 1}$, and $\boldsymbol{\beta} := \mu$. Recall that $(\boldsymbol{X}'\boldsymbol{X})^{-1} = 1/n$ and hence $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = \bar{Y}$.

If we test $H_0 : \ \mu = \mu_0$ for a μ_0 the is known, then $\boldsymbol{A} = 1$ is a 1×1 matrix $(q = 1)$ and $\boldsymbol{A\beta} = \boldsymbol{c}$ with $\boldsymbol{c} = \mu_0$.

Given that H_0 is true, the least-squares estimator of μ $[\hat{\boldsymbol{\beta}}_{H_0}]$ is

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{H_0} := \hat{\mu}_{H_0} &= \hat{\boldsymbol{\beta}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}\left(\boldsymbol{c} - \boldsymbol{A}\hat{\boldsymbol{\beta}}\right)\\ &= \bar{Y} + \frac{1}{n}\cdot n \cdot (\mu_0 - \bar{Y}) = \mu_0.\end{aligned}$$

[Is this sensible?] And

$$\mathrm{RSS} = \|\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 = ns_y^2.$$

Therefore,

$$\begin{aligned}\mathrm{RSS}_{H_0} - \mathrm{RSS} &= (\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})'\left[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right]^{-1}\left(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c}\right)\\ &= n(\bar{Y} - \mu_0)^2.\end{aligned}$$

And

$$\frac{(\mathrm{RSS}_{H_0} - \mathrm{RSS})/q}{\mathrm{RSS}/(n-p)} = \frac{(\bar{Y} - \mu_0)^2}{s_y^2/(n-1)} \overset{H_0}{\sim} F_{1,n-1}.$$

But

$$\frac{\bar{Y} - \mu_0}{s_y/\sqrt{n-1}} \overset{H_0}{\sim} t_{n-1}.$$

Therefore, in particular, $t_k^2 = F_{1,k}$.

2. Simple linear regression. Here,

$$Y_i = \alpha + \beta x_i + \varepsilon_i \qquad (1 \le i \le n).$$

Therefore, $p = 2$,

$$\boldsymbol{\beta} = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}, \qquad \text{and} \qquad \boldsymbol{X} = \begin{pmatrix}1 & x_1\\ \vdots & \vdots\\ 1 & x_n\end{pmatrix}.$$

Recall that the least-squares estimates of α and β are

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}, \qquad \hat{\beta} = \frac{rs_y}{s_x}.$$

Now consider testing the hypothesis,

$$H_0: \ \beta = 0, \ \alpha = \mu_0,$$

where μ_0 is known.

Let $\boldsymbol{c} = (\mu_0\,, 0)'$ and $\boldsymbol{A} = \boldsymbol{I}_2$, so that $q = 2$. Then, H_0 is the same as $H_0: \ \boldsymbol{A\beta} = \boldsymbol{c}$. We have

$$\hat{\boldsymbol{\beta}}_{H_0} = \hat{\boldsymbol{\beta}} + (\boldsymbol{c} - \boldsymbol{A}\hat{\boldsymbol{\beta}}) = \boldsymbol{c} = \begin{pmatrix} \mu_0 \\ 0 \end{pmatrix}.$$

[Is this sensible?]

Now,

$$\text{RSS}_{H_0} - \text{RSS} = (\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})' \left[\boldsymbol{A}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{A}'\right]^{-1} (\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c}) = (\hat{\boldsymbol{\beta}} - \boldsymbol{c})'(\mathbf{X}'\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{c}).$$

Now,

$$\mathbf{X}'\mathbf{X} = n \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{pmatrix} \quad \Rightarrow \quad \hat{\boldsymbol{\beta}} - \boldsymbol{c} = \begin{pmatrix} \bar{Y} - \hat{\beta}\bar{x} - \mu_0 \\ \hat{\beta} \end{pmatrix}.$$

Therefore,

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{c})'(\mathbf{X}'\mathbf{X}) = n(\bar{Y} - \mu_0)(1\,, \bar{x}),$$

whence

$$\text{RSS}_{H_0} - \text{RSS} = n(\bar{Y} - \mu_0)^2.$$

Next we compute

$$\text{RSS} = \|\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 = \sum_{i=1}^{n} \left[Y_i - (\mathbf{X}\hat{\boldsymbol{\beta}})_i\right]^2.$$

Since

$$\mathbf{X}\hat{\boldsymbol{\beta}} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \bar{Y} - \hat{\beta}\bar{x} \\ \hat{\beta} \end{pmatrix} = \left(\bar{Y} + \hat{\beta}(x_i - \bar{x})\right)_{i=1}^{n} = \left(\bar{Y} + \frac{r s_y}{s_x}(x_i - \bar{x})\right)_{i=1}^{n},$$

it follows that

$$\begin{aligned} \text{RSS} &= \sum_{i=1}^{n} \left[(Y_i - \bar{Y})^2 + \frac{r^2 s_y^2}{s_x^2}(x_i - \bar{x})^2 - 2\frac{r s_y}{s_x}(Y_i - \bar{Y})(x_i - \bar{x})\right] \\ &= n s_y^2 + n r^2 s_y^2 - \frac{2 n r s_y}{s_x} \sum_{i=1}^{n} (Y_i - \bar{Y})(x_i - \bar{x}) \\ &= n s_y^2 + n r^2 s_y^2 - 2 n r^2 s_y^2 \\ &= n s_y^2 (1 - r^2). \end{aligned}$$

Therefore,

$$\frac{(\bar{Y} - \mu_0)^2}{s_y^2(1 - r^2)} \overset{H_0}{\sim} F_{2,n-2}.$$

3. Two-sample mean. Consider two populations: $\mathrm{N}(\mu_1\,, \sigma^2)$ and $\mathrm{N}(\mu_2\,, \sigma^2)$ with equal variances. We wish to know if $\mu_1 = \mu_2$. Take two independent random samples,

$$\begin{aligned} y_{1,1}, \ldots, y_{1,n_1} &\sim \mathrm{N}(\mu_1\,, \sigma^2),\\ y_{2,1}, \ldots, y_{2,n_2} &\sim \mathrm{N}(\mu_2\,, \sigma^2). \end{aligned}$$

We have a linear model: $p = 2$, $n = n_1 + n_2$,

$$\boldsymbol{Y} = \begin{pmatrix} y_{1,1} \\ \vdots \\ y_{1,n_1} \\ y_{2,1} \\ \vdots \\ y_{2,n_2} \end{pmatrix}, \qquad \boldsymbol{X} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{n_1 \times 1} & \mathbf{0}_{n_1 \times 1} \\ \mathbf{0}_{n_2 \times 1} & \mathbf{1}_{n_2 \times 1} \end{pmatrix}.$$

In particular,

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix} \quad \Rightarrow \quad (\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} n_1^{-1} & 0 \\ 0 & n_2^{-1} \end{pmatrix}.$$

So now consider

$$H_0 : \mu_1 = \mu_2 \qquad \Longleftrightarrow \qquad H_0 : \mu_1 = \mu_2 \qquad \Longleftrightarrow \qquad H_0 : \underbrace{(1\,, -1)}_{\boldsymbol{A}} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \mathbf{0}.$$

That is, $q = 1$, $\boldsymbol{A} := (1\,, -1)$, and $\boldsymbol{c} = 0$. In this way we find that

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = \begin{pmatrix} \bar{y}_{1,\bullet} \\ \bar{y}_{2,\bullet} \end{pmatrix}.$$

[Does this make intuitive sense?]

In order to find $\hat{\boldsymbol{\beta}}_{H_0}$, we first compute

$$\boldsymbol{A}\hat{\boldsymbol{\beta}} = (1\,, -1) \begin{pmatrix} \bar{y}_{1,\bullet} \\ \bar{y}_{2,\bullet} \end{pmatrix} = \bar{y}_{1,\bullet} - \bar{y}_{2,\bullet}.$$

Also,

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}' = \begin{pmatrix} n_1^{-1} & 0 \\ 0 & n_2^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} n_1^{-1} \\ -n_2^{-1} \end{pmatrix},$$

so that

$$\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}' = \frac{1}{n_2} + \frac{1}{n_2} = \frac{n}{n_1 n_2}.$$

Now we put things together:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{H_0} &= \hat{\boldsymbol{\beta}} + \begin{pmatrix} n_1^{-1} \\ -n_2^{-1} \end{pmatrix} \frac{n_1 n_2}{n} \left(\bar{y}_{1,\bullet} - \bar{y}_{2,\bullet}\right) \\ &= \hat{\boldsymbol{\beta}} + \begin{pmatrix} \frac{n_2}{2}\left(\bar{y}_{1,\bullet} - \bar{y}_{2,\bullet}\right) \\ \\ -\frac{n_1}{n}\left(\bar{y}_{1,\bullet} - \bar{y}_{2,\bullet}\right) \end{pmatrix} \\ &= \begin{pmatrix} \frac{n_1}{n}\bar{y}_{1,\bullet} + \frac{n_2}{n}\bar{y}_{2,\bullet} \\ \\ \frac{n_1}{n}\bar{y}_{1,\bullet} + \frac{n_2}{n}\bar{y}_{2,\bullet} \end{pmatrix}.\end{aligned}$$

Since $n_2\bar{y}_{2,\bullet} = \sum_{j=1}^{n_2} y_{j2,j}$ and $n_1\bar{y}_{1,\bullet} = \sum_{i=1}^{n_1} y_{1,j}$, it follows that

$$\hat{\boldsymbol{\beta}}_{H_0} = \begin{pmatrix} \bar{y}_{\bullet,\bullet} \\ \bar{y}_{\bullet,\bullet} \end{pmatrix}.$$

[Does this make sense?] Since

$$\boldsymbol{X}\hat{\boldsymbol{\beta}} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{y}_{1,\bullet} \\ \bar{y}_{2,\bullet} \end{pmatrix} = \begin{pmatrix} \bar{y}_{1,\bullet} \\ \vdots \\ \bar{y}_{1,\bullet} \\ \bar{y}_{2\,\bullet} \\ \vdots \\ \bar{y}_{2,\bullet} \end{pmatrix} = \begin{pmatrix} \bar{y}_{1\,\bullet}\mathbf{1}_{n_1\times 1} \\ \bar{y}_{2,\bullet}\mathbf{1}_{n_2\times 1} \end{pmatrix},$$

we have

$$\begin{aligned}\text{RSS} &= \sum_{j=1}^{n_1} \left(y_{1,j} - \bar{y}_{1\,\bullet}\right)^2 + \sum_{j=1}^{n_2} \left(y_{2,j} - \bar{y}_{2\,\bullet}\right)^2 \\ &= n_1 s_1^2 + n_2 s_2^2.\end{aligned}$$

I particular,

$$\frac{\text{RSS}}{n-p} = \frac{n_1}{n-2}s_1^2 + \frac{n_2}{n-2}s_2^2 := s_p^2$$

is the socalled "pooled variance."

Similarly,

$$\text{RSS}_{H_0} - \text{RSS} = \frac{n_1 n_2}{n}\left(\bar{y}_{1,\bullet} - \bar{y}_{2,\bullet}\right)^2.$$

Therefore,

$$\frac{\frac{n_1 n_2}{n}\left(\bar{y}_{1,\bullet} - \bar{y}_{2,\bullet}\right)^2}{s_p^2} \overset{H_0}{\sim} F_{1,n-2} \quad \Rightarrow \quad \frac{\bar{y}_{1,\bullet} - \bar{y}_{2,\bullet}}{s_p\sqrt{\frac{n}{n_1 n_2}}} \overset{H_0}{\sim} t_{n-2}.$$

4. ANOVA: One-way layout. Consider p populations that are respectively distributed as $\mathrm{N}(\mu_1\,,\sigma^2),\ldots,\mathrm{N}(\mu_p\,,\sigma^2)$. We wish to test

$$H_0:\ \mu_1 = \cdots = \mu_p.$$

We have seen that we are in the setting of linear models, so we can compute $\hat{\boldsymbol{\beta}}_{H_0}$ etc. that way. I will leave this up to you and compute directly instead: Sample $y_{j,1}\,,\ldots y_{j,n_k}$ i.i.d. $\mathrm{N}(\mu_j\,,\sigma^2)$ [independent also as j varies]. Then we vectorize:

$$\boldsymbol{Y} := \begin{pmatrix} y_{1,1} \\ \vdots \\ y_{1,n_1} \\ \vdots \\ y_{p,1} \\ \vdots \\ y_{p,n_p} \end{pmatrix}; \qquad \text{etc.}$$

Instead we now find $\hat{\boldsymbol{\beta}}$ directly by solving

$$\min_{\boldsymbol{\mu}} \sum_{i=1}^{p}\sum_{j=1}^{n_i} \left(y_{i,j} - \mu_i\right)^2.$$

That is, compute

$$\frac{\partial}{\partial \mu_i}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(y_{i,j}-\mu_i\right)^2 = -\sum_{j=1}^{n_i} 2\left(y_{i,j}-\mu_i\right) \equiv 0 \implies \hat{\mu}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{i,j} = \bar{y}_{i,\bullet}.$$

This yields

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \bar{y}_{1,\bullet} \\ \vdots \\ \bar{y}_{n,\bullet} \end{pmatrix}.$$

What about $\hat{\boldsymbol{\beta}}_{H_0}$? Under H_0, $\mu_1 = \cdots = \mu_p \equiv \mu$ and so $q = p - 1$. So we have

$$\min_{\mu} \sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(y_{i,j}-\mu\right)^2 \implies \hat{\boldsymbol{\beta}}_{H_0} = \begin{pmatrix} \bar{y}_{\bullet,\bullet} \\ \vdots \\ \bar{y}_{\bullet,\bullet} \end{pmatrix}.$$

Also,

$$\mathrm{RSS} = \sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(y_{i,j} - \bar{y}_{i,\bullet}\right)^2,$$

and

$$\begin{aligned}
\mathrm{RSS}_{H_0} &= \sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(y_{i,j}-\bar{y}_{\bullet,\bullet}\right)^2 = \sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(y_{i,j}-\bar{y}_{i,\bullet}+\bar{y}_{i,\bullet}-\bar{y}_{\bullet,\bullet}\right)^2 \\
&= \sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(y_{i,j}-\bar{y}_{i,\bullet}\right)^2 + 2\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(y_{i,j}-\bar{y}_{i,\bullet}\right)\left(\bar{y}_{i,\bullet}-\bar{y}_{\bullet,\bullet}\right) \\
&\qquad + \sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(\bar{y}_{i,\bullet}-\bar{y}_{\bullet,\bullet}\right)^2 \\
&= \sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(y_{i,j}-\bar{y}_{i,\bullet}\right)^2 + \sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(\bar{y}_{i,\bullet}-\bar{y}_{\bullet,\bullet}\right)^2 \\
&= \mathrm{RSS} + \sum_{i=1}^{p} n_i\left(\bar{y}_{i,\bullet}-\bar{y}_{\bullet,\bullet}\right)^2 .
\end{aligned}$$

It follows from the general theory that

$$\frac{\sum_{i=1}^{p} n_i\left(\bar{y}_{i,\bullet}-\bar{y}_{\bullet,\bullet}\right)^2/(p-1)}{\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(y_{i,j}-\bar{y}_{i,\bullet}\right)^2/(n-p)} \overset{H_0}{\sim} F_{p-1,n-p}.$$

"Statistical interpretation":

$$\frac{\sum_{i=1}^{p} n_i\left(\bar{y}_{i,\bullet}-\bar{y}_{\bullet,\bullet}\right)^2}{p-1} = \text{The variation between the samples;}$$

whereas

$$\frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(y_{i,j}-\bar{y}_{i,\bullet}\right)^2}{n-p} = \text{The variation within the samples.}$$

Therefore,

$$\mathrm{RSS}_{H_0} = \text{Variation between} + \text{Variation within} = \text{Total variation}.$$