General Linear and Subspace Testing (Lab 6)

BST 235: Advanced Regression and Statistical Learning

Alex Levis, Fall 2019

1 Review of general linear hypothesis testing

In the linear model

$$\mathbb{E}_P(Y \mid \mathbf{X}) = \mathbf{X}^T \boldsymbol{\beta}(P),$$

assuming homoscedastic normal data (i.e., $Y - \mathbb{E}_P(Y \mid \mathbf{X}) \mid \mathbf{X} \sim \mathcal{N}(0, \sigma^2)$), one may wish to perform inference on any wild function of $\boldsymbol{\beta}(P)$, e.g., $\boldsymbol{\chi}(\boldsymbol{\beta}) = \left(\sin(\beta_1), \sum_{j=1}^{d} \beta_j^4\right)$. In general, it is very difficult to achieve *exact* inference in this setting, without further restrictions on $\boldsymbol{\chi}(\,\cdot\,)$.

Perhaps unsurprisingly, there is one class of functions of the regression parameters with very well-characterized distributions — linear functions! In particular, let $A \in \mathbb{R}^{q \times d}$, and consider

$$\boldsymbol{\chi}(\boldsymbol{\beta}) = A\boldsymbol{\beta}.$$

Given the sample least squares estimator $\widehat{\boldsymbol{\beta}}$, a natural estimator of $\boldsymbol{\chi}(\boldsymbol{\beta}(P))$ is $\widehat{\boldsymbol{\chi}} = A\widehat{\boldsymbol{\beta}}$. We have already seen under assumptions (A') and (B), (C), (D), that $\widehat{\boldsymbol{\beta}} \mid \mathbb{X} \sim \mathcal{N}_d(\boldsymbol{\beta}(P), \sigma^2(\mathbb{X}^T\mathbb{X})^{-1})$, so by properties of normal random vectors,

$$\widehat{\boldsymbol{\chi}} = A\widehat{\boldsymbol{\beta}} \mid \mathbb{X} \sim \mathcal{N}_q\left(A\boldsymbol{\beta}(P), \sigma^2 A(\mathbb{X}^T\mathbb{X})^{-1}A^T\right).$$

Last lecture, we encountered the following problem: how should we test the null hypothesis

$$H_0 : A\boldsymbol{\beta}(P) = \mathbf{0}_q,$$

under the typical assumptions? This is known as a *general linear hypothesis* test, since we are asking whether there is additional linear structure among the components of $\boldsymbol{\beta}(P)$, encoded in the q rows of A. Assuming $\text{rank}(A) = q$, and defining $\Sigma = \sigma^2 A(\mathbb{X}^T\mathbb{X})^{-1}A^T$, we saw that Σ is strictly positive definite (i.e., invertible and positive semi-definite / "non-negative definite"). Therefore, using what we developed last lab, we could consider the rescaled random vector

$$\Sigma^{-1/2}A\widehat{\boldsymbol{\beta}} \mid \mathbb{X} \sim \mathcal{N}_q(\Sigma^{-1/2}A\boldsymbol{\beta}(P), I_q) \overset{H_0}{\equiv} \mathcal{N}_q(\mathbf{0}_q, I_q).$$

This is convenient as taking the squared norm of a standard multivariate normal vector, we get a (central) chi-squared distribution:

$$\begin{aligned}
&\left(\Sigma^{-1/2}A\widehat{\boldsymbol{\beta}}\right)^T \Sigma^{-1/2}A\widehat{\boldsymbol{\beta}} \\
&= \widehat{\boldsymbol{\beta}}^T A^T \Sigma^{-1} A\widehat{\boldsymbol{\beta}} \\
&= \frac{\widehat{\boldsymbol{\beta}}^T A^T \left(A(\mathbb{X}^T\mathbb{X})^{-1}A^T\right)^{-1} A\widehat{\boldsymbol{\beta}}}{\sigma^2} \mid \mathbb{X} \overset{H_0}{\sim} \chi_q^2(0).
\end{aligned}$$

Note, however, that since σ^2 is typically unknown, we are still not quite ready to use this to construct a hypothesis test.

One obvious idea is to plug in the (unbiased) estimator $\widehat{\sigma}_u^2 = \frac{1}{n-d}\|\mathbf{Y} - \mathbb{X}\widehat{\boldsymbol{\beta}}\|^2 = \frac{1}{n-d}\mathbf{Y}^T(I_n - \widehat{P}_{\mathbb{X}})\mathbf{Y}$, and see if we can characterize the distribution of the statistic under H_0. Recall that

$$\frac{1}{\sigma^2}\mathbf{Y}^T(I_n - \widehat{P}_{\mathbb{X}})\mathbf{Y} = \frac{1}{\sigma^2}\boldsymbol{\epsilon}^T(I_n - \widehat{P}_{\mathbb{X}})\boldsymbol{\epsilon} \,|\, \mathbb{X} \sim \chi^2_{n-d}(0),$$

where $\boldsymbol{\epsilon} = \mathbf{Y} - \mathbb{X}\boldsymbol{\beta}(P)$, since $(I_n - \widehat{P}_{\mathbb{X}})$ is symmetric, idempotent, and has rank $n-d$ when the columns of $\mathbb{X}$ are linearly independent. In other words, $\frac{(n-d)\widehat{\sigma}_u^2}{\sigma^2} \,|\, \mathbb{X} \sim \chi^2_{n-d}(0)$. We have seen that by Fisher-Cochran, the linear function $\widehat{\boldsymbol{\beta}}$ and the quadratic function $\widehat{\sigma}_u^2$ are independent (conditional on the covariates $\mathbb{X}$). Therefore,

$$\begin{aligned}
F &:= \frac{\widehat{\boldsymbol{\beta}}^T A^T \left(A(\mathbb{X}^T\mathbb{X})^{-1}A^T\right)^{-1} A\widehat{\boldsymbol{\beta}}/q}{\mathbf{Y}^T(I_n - \widehat{P}_{\mathbb{X}})\mathbf{Y}/(n-d)} \\
&= \frac{\widehat{\boldsymbol{\beta}}^T A^T \left(A(\mathbb{X}^T\mathbb{X})^{-1}A^T\right)^{-1} A\widehat{\boldsymbol{\beta}}/q}{\widehat{\sigma}_u^2} \\
&= \left\{\frac{1}{q} \cdot \frac{\widehat{\boldsymbol{\beta}}^T A^T \left(A(\mathbb{X}^T\mathbb{X})^{-1}A^T\right)^{-1} A\widehat{\boldsymbol{\beta}}}{\sigma^2}\right\} \Big/ \left\{\frac{1}{n-d} \cdot \frac{(n-d)\widehat{\sigma}_u^2}{\sigma^2}\right\} \\
&\overset{H_0}{\sim} \frac{\chi^2_q(0)/q}{\chi^2_{n-d}(0)/(n-d)} \equiv F_{q,n-d}(0), \text{ as the two chi-squared variables are independent.}
\end{aligned}$$

Given this test statistic, we can construct a standard F-test, that rejects H_0 with probability α under the null hypothesis. Specifically, we should reject when $F > F_{q,n-d,1-\alpha}(0)$, where $F_{q,n-d,1-\alpha}(0)$ is the $(1-\alpha)$-th quantile of the central F distribution with degrees of freedom q and $n-d$.

2 An alternative perspective

The linear model, equivalently stated in terms of the n observations in our sample, is

$$\mathbb{E}_P(\mathbf{Y} \,|\, \mathbb{X}) = \mathbb{X}\boldsymbol{\beta}(P) = \sum_{j=1}^d \mathbf{X}^{(j)}\beta_j(P) \in \mathcal{C}(\mathbb{X}) \subseteq \mathbb{R}^n.$$

As we noted above, though, the null hypothesis, $H_0 : A\boldsymbol{\beta}(P) = \mathbf{0}_q$, imposes q additional linear constraints on the parameter vector $\boldsymbol{\beta}(P)$. That is, under the null hypothesis, the mean of $\mathbf{Y}$ given $\mathbb{X}$ lies in a linear subspace of $\mathcal{C}(\mathbb{X})$:

$$V_0 := \left\{\mathbb{X}\boldsymbol{\beta} \,\middle|\, \boldsymbol{\beta} \in \mathbb{R}^d, A\boldsymbol{\beta} = \mathbf{0}_q\right\} = \{\mathbb{X}\boldsymbol{\beta} \,|\, \boldsymbol{\beta} \in \mathcal{N}(A)\} \subseteq \mathcal{C}(\mathbb{X}).$$

The null hypothesis is therefore equivalent to $H_0 : \mathbb{E}_P(\mathbf{Y} \,|\, \mathbb{X}) \in V_0$. This is an instance of a general subspace test setting, as we wish to know whether the (conditional) mean of the outcome lies in a particular subspace (i.e., V_0) of a larger assumed space (i.e., $\mathcal{C}(\mathbb{X})$). We study the abstract problem in the next section.

3 General subspace hypothesis testing

Consider the unconditional homoscedastic normal data setting, $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \sigma^2 I_n)$. Let $V_0, V \subseteq \mathbb{R}^n$ be two linear subspaces of $\mathbb{R}^n$ such that $V_0 \subseteq V$. We will consider testing

$$H_0 : \boldsymbol{\mu} \in V_0 \text{ versus } H_1 : \boldsymbol{\mu} \in V \setminus V_0.$$

The key idea will be to compare residuals under the null hypothesis with residuals from the model with no restrictions beyond $\boldsymbol{\mu} \in V$. Specifically, let

$$\boldsymbol{e}^{(0)} = \mathbf{Y} - P_{V_0}(\mathbf{Y}) = P_{V_0^\perp}(\mathbf{Y}), \text{ and } \boldsymbol{e}^{(1)} = \mathbf{Y} - P_V(\mathbf{Y}) = P_{V^\perp}(\mathbf{Y}),$$

and we will consider $\|\boldsymbol{e}^{(0)} - \boldsymbol{e}^{(1)}\|^2$ being large as evidence against H_0. To understand the distribution of this quantity, the following exercise is crucial.

Exercise 1. Let U, W be finite-dimensional subspaces of vector space V, with $U \subseteq W$. Show that

$$W = U \oplus (W \cap U^\perp).$$

To do this, recall from the first homework that for any vector space V, $V_0 \subseteq V$ a finite-dimensional linear subspace, $V = V_0 \oplus V_0^\perp$. Now, replace the larger vector space V with W and the subspace V_0 with U. Be careful with the symbol $\perp$.

Note the following corollaries to Exercise 1, which follow from results in the first homework:

(a) $V_0 \subseteq V$ means $V = V_0 \oplus (V \cap V_0^\perp)$, and $V^\perp \subseteq V_0^\perp$ means $V_0^\perp = V^\perp \oplus (V \cap V_0^\perp)$;

(b) $P_V = P_{V_0} + P_{V \cap V_0^\perp}$, and $P_{V_0^\perp} = P_{V^\perp} + P_{V \cap V_0^\perp}$ (see also Exercise 4 of Lab 4);

(c) $\dim(V) = \dim(V_0) + \dim(V \cap V_0^\perp) \implies \dim(V \cap V_0^\perp) = \dim(V) - \dim(V_0)$.

As a consequence of corollary (b), we find

$$\|\boldsymbol{e}^{(0)} - \boldsymbol{e}^{(1)}\|^2 = \|P_V(\mathbf{Y}) - P_{V_0}(\mathbf{Y})\|^2 = \|P_{V \cap V_0^\perp}(\mathbf{Y})\|^2 = \mathbf{Y}^T \widehat{P}_{V \cap V_0^\perp} \mathbf{Y}.$$

Moreover, again by corollary (b),

$$\mathbf{Y}^T \widehat{P}_{V \cap V_0^\perp} \mathbf{Y} = \mathbf{Y}^T \widehat{P}_{V_0^\perp} \mathbf{Y} - \mathbf{Y}^T \widehat{P}_{V^\perp} \mathbf{Y} = \|\boldsymbol{e}^{(0)}\|^2 - \|\boldsymbol{e}^{(1)}\|^2.$$

Let $\boldsymbol{\epsilon} = \mathbf{Y} - \boldsymbol{\mu}$, then

$$\|\boldsymbol{e}^{(0)} - \boldsymbol{e}^{(1)}\|^2 = \|P_{V \cap V_0^\perp}(\boldsymbol{\mu} + \boldsymbol{\epsilon})\|^2 \overset{H_0}{=} \|P_{V \cap V_0^\perp}(\boldsymbol{\epsilon})\|^2 = \boldsymbol{\epsilon}^T \widehat{P}_{V \cap V_0^\perp} \boldsymbol{\epsilon},$$

since $\boldsymbol{\mu} \in V_0$ under H_0. Combining these facts, and using corollary (c), we see that

$$\frac{\|\boldsymbol{e}^{(0)}\|^2 - \|\boldsymbol{e}^{(1)}\|^2}{\sigma^2} \overset{H_0}{=} \left(\frac{\boldsymbol{\epsilon}}{\sigma}\right)^T \widehat{P}_{V \cap V_0^\perp} \left(\frac{\boldsymbol{\epsilon}}{\sigma}\right) \sim \chi^2_{\mathrm{rank}(\widehat{P}_{V \cap V_0^\perp})}(0) \equiv \chi^2_{\dim(V) - \dim(V_0)}(0).$$

Exercise 2. In the above setting, show that the unbiased estimator of σ^2,

$$\widehat{\sigma}_u^2 = \frac{\|\mathbf{Y} - P_V(\mathbf{Y})\|^2}{n - \dim(V)} = \frac{\|\boldsymbol{e}^{(1)}\|^2}{n - \dim(V)},$$

is independent of $\|\boldsymbol{e}^{(0)} - \boldsymbol{e}^{(1)}\|^2$. Use this to justify the F-test of H_0 based on

$$F_{V,V_0} = \frac{\left(\|\boldsymbol{e}^{(0)}\|^2 - \|\boldsymbol{e}^{(1)}\|^2\right)/(\dim(V) - \dim(V_0))}{\widehat{\sigma}_u^2}. \tag{1}$$

4 Return to linear models

As argued in Section 2 above, the general linear hypothesis test can be stated as a general subspace hypothesis of

$$H_0 : \mathbb{E}_P(\mathbf{Y} \,|\, \mathbb{X}) \in V_0 \text{ versus } H_1 : \mathbb{E}_P(\mathbf{Y} \,|\, \mathbb{X}) \in V \setminus V_0, \tag{2}$$

where $V_0 = \{\mathbb{X}\boldsymbol{\beta} \,|\, \boldsymbol{\beta} \in \mathcal{N}(A)\} \subseteq V \subseteq \mathbb{R}^n$, and $V = \mathcal{C}(\mathbb{X})$.

Lemma 1. Assuming $\mathbb{X}$ has full column rank, $\dim(V_0) = d - q$ and $\dim(V) = d$.

Proof. That $\dim(V) = \dim(\mathcal{C}(\mathbb{X})) = d$ is an assumption of the lemma, so we need only show the other equality. By rank-nullity, we know

$$d = \operatorname{rank}(A) + \dim(\mathcal{N}(A)) \implies \dim(\mathcal{N}(A)) = d - q,$$

since we have assumed $\operatorname{rank}(A) = q$. Let $\mathbf{b}_1, \ldots, \mathbf{b}_{d-q} \in \mathbb{R}^d$ be a basis for $\mathcal{N}(A)$. It suffices to show that $\mathbb{X}\mathbf{b}_1, \ldots, \mathbb{X}\mathbf{b}_{d-q}$ is a basis for V_0. Clearly these vectors span V_0, since $\mathbf{b}_1, \ldots, \mathbf{b}_{d-q}$ spans $\mathcal{N}(A)$. It remains to establish linear independence. To that end, let $\alpha_1, \ldots, \alpha_{d-q} \in \mathbb{R}$ satisfy

$$\mathbf{0}_n = \sum_{j=1}^{d-q} \alpha_j \mathbb{X}\mathbf{b}_j = \mathbb{X}\left(\sum_{j=1}^{d-q} \alpha_j \mathbf{b}_j\right).$$

Then we know $\sum_{j=1}^{d-q} \alpha_j \mathbf{b}_j \in \mathcal{N}(\mathbb{X})$, but by rank-nullity

$$\dim(\mathcal{N}(\mathbb{X})) = d - \operatorname{rank}(\mathbb{X}) = d - d = 0.$$

This implies that $\mathcal{N}(\mathbb{X}) = \{\mathbf{0}_d\}$, so

$$\sum_{j=1}^{d-q} \alpha_j \mathbf{b}_j = \mathbf{0}_d \implies \alpha_1 = \cdots = \alpha_{d-q} = 0,$$

since $\mathbf{b}_1, \ldots, \mathbf{b}_{d-q}$ are linearly independent. Thus, $\mathbb{X}\mathbf{b}_1, \ldots, \mathbb{X}\mathbf{b}_{d-q}$ are linearly independent, as claimed. ☐

In order to derive the test statistic F_{V,V_0} in this setting, it remains to find more explicit forms for $\|\boldsymbol{e}^{(0)} - \boldsymbol{e}^{(1)}\|^2$ and $\|\boldsymbol{e}^{(1)}\|^2$. The latter term is easy, since

$$\|\boldsymbol{e}^{(1)}\|^2 = \|P_{V^\perp}(\mathbf{Y})\|^2 = \|P_{\mathcal{C}(\mathbb{X})^\perp}(\mathbf{Y})\|^2 = \mathbf{Y}^T(I_n - \widehat{P}_{\mathbb{X}})\mathbf{Y} = (n-d)\widehat{\sigma}_u^2.$$

Lemma 2. When $\text{rank}(\mathbb{X}) = d$, $\text{rank}(A) = q$, $\|\boldsymbol{e}^{(0)} - \boldsymbol{e}^{(1)}\|^2 = \mathbf{Y}^T \widehat{P}_U \mathbf{Y}$, where $U = \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}A^T$, and $\widehat{P}_U$ is the matrix corresponding to projection onto $\mathcal{C}(U)$.

Proof. (Caution: tricky proof!) Since

$$\|\boldsymbol{e}^{(0)} - \boldsymbol{e}^{(1)}\|^2 = \mathbf{Y}^T \widehat{P}_{V \cap V_0^\perp} \mathbf{Y},$$

we need only show that $\mathcal{C}(U) = V \cap V_0^\perp$.

By the form of U we must have $\mathcal{C}(U) \subseteq \mathcal{C}(\mathbb{X}) = V$, so by Exercise 1,

$$\mathcal{C}(U) \subseteq V \implies V = \mathcal{C}(U) \oplus (V \cap \mathcal{C}(U)^\perp).$$

But by corollary (a), $V = V_0 \oplus (V \cap V_0^\perp)$. We claim that it is sufficient to show

$$V \cap \mathcal{C}(U)^\perp = V_0. \tag{3}$$

To see this, note that this would imply $V_0 \subseteq \mathcal{C}(U)^\perp \iff \mathcal{C}(U) \subseteq V_0^\perp$, and

$$V = V_0 \oplus (V \cap V_0^\perp) = V_0 \oplus \mathcal{C}(U).$$

In turn, this would imply the desired equality $V \cap V_0^\perp = \mathcal{C}(U)$: the inclusion $\mathcal{C}(U) \subseteq V \cap V_0^\perp$ is already shown, and for any $w \in V \cap V_0^\perp$, its unique representation is $w = x + z \in V_0 \oplus \mathcal{C}(U)$ for one direct sum, and $w = 0 + w \in V_0 \oplus (V \cap V_0^\perp)$ for the other — as $\mathcal{C}(U) \subseteq V \cap V_0^\perp$, the two representations are equal and $w = z \in \mathcal{C}(U)$.

We finish by proving (3), which is equivalent to $V \cap \mathcal{N}(U^T) = V_0$. First, for $v \in V_0$, there exists $\boldsymbol{\beta} \in \mathcal{N}(A)$ such that $v = \mathbb{X}\boldsymbol{\beta}$. We must then have $\boldsymbol{\beta} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T v$, so $\mathbf{0}_q = A\boldsymbol{\beta} = U^T v$, implying $v \in \mathcal{N}(U^T)$. As v belongs to V trivially, $v \in V \cap \mathcal{N}(U^T)$. Conversely, for $v \in V \cap \mathcal{N}(U^T)$, there exists $\boldsymbol{\beta} \in \mathbb{R}^d$ such that $v = \mathbb{X}\boldsymbol{\beta}$ and $\mathbf{0}_q = U^T v$. Hence $\mathbf{0}_q = A(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{X}\boldsymbol{\beta} = A\boldsymbol{\beta}$, so $v \in V_0$. □

Exercise 3. Given the facts we showed above, derive the form of the test statistic F_{V,V_0} from (1) in this example. How does this compare to the statistic F derived at the end of Section 1?