Lecture 07
Hypothesis Testing with Multivariate Regression

23 September 2015

Taylor B. Arnold
Yale Statistics
STAT 312/612

Yale

Goals for today

1. Review of assumptions and properties of linear model
2. The multivariate T-test
3. Hypothesis test simulations
4. The multivariate F-test

Review from Last Time

Classical linear model assumptions

I. Linearity $Y = X\beta + \epsilon$

II. Strict exogeneity $\mathbb{E}\left(\epsilon|X\right) = 0$

III. No multicollinearity $\mathbb{P}\left[\text{rank}(X) = p\right] = 1$

IV. Spherical errors $\mathbb{V}\left(\epsilon|X\right) = \sigma^2\mathbb{I}_n$

V. Normality $\epsilon|X \sim \mathcal{N}(0, \sigma^2\mathbb{I}_n)$

Finite sample properties

Under assumptions I-III:

(A) $\mathbb{E}(\widehat{\beta}|X) = \beta$

Under assumptions I-IV:

(B) $\mathbb{V}(\widehat{\beta}|X) = \sigma^2(X^tX)^{-1}$

(C) $\widehat{\beta}$ is the best linear unbiased estimator (Gauss-Markov)

(D) $Cov(\widehat{\beta}, r|X) = 0$

(E) $\mathbb{E}(s^2|X) = \sigma^2$

Under assumptions I-V:

(F) $\widehat{\beta}$ achieves the Cramér–Rao lower bound

The T-test

Hypothesis tests

Consider testing the hypothesis $H_0 : \beta_j = b_j$.

Hypothesis tests

Consider testing the hypothesis $H_0 : \beta_j = b_j$.

Under assumptions I-V we have the following:

$$\widehat{\beta}_j - b_j \Big| X, H_0 \sim \mathcal{N}(0, \sigma^2 \left((X^t X)^{-1}_{jj}\right))$$

Z-test

This suggests the following test statistic:

$$z = \frac{\widehat{\beta}_j - b_j}{\sqrt{\sigma^2 \left((X^t X)^{-1}_{jj} \right)}}$$

With,

$$z \mid X, H_0 \sim \mathcal{N}(0, 1)$$

T-test

As in the simple linear linear regression case, we generally need to estimate σ^2 with s^2.

T-test

As in the simple linear linear regression case, we generally need to estimate σ^2 with s^2.

This yields the following test statistic:

$$
\begin{aligned}
t &= \frac{\widehat{\beta}_j - b_j}{\sqrt{s^2 \left((X^t X)^{-1}_{jj} \right)}} \\
&= \frac{\widehat{\beta}_j - b_j}{\text{S.E.}(\widehat{\beta}_j)}
\end{aligned}
$$

T-test

The test statistic has a T-distribution with $(n - p)$ degrees of freedom under the null hypothesis:

$$t | X, H_0 \sim t_{n-p}$$

T-test

The test statistic has a T-distribution with $(n - p)$ degrees of freedom under the null hypothesis:

$$t \mid X, H_0 \sim t_{n-p}$$

This time around, we'll actually prove this.

To start, we re-write the test statistic as:

$$t = \frac{\widehat{\beta} - b}{\sqrt{\sigma^2 \left((X^t X)^{-1}_{jj}\right)}} \cdot \sqrt{\frac{\sigma^2}{s^2}}$$

To start, we re-write the test statistic as:

$$\begin{aligned} t &= \frac{\widehat{\beta} - b}{\sqrt{\sigma^2 \left((X^t X)^{-1}_{jj} \right)}} \cdot \sqrt{\frac{\sigma^2}{s^2}} \\ &= \frac{z}{\sqrt{s^2/\sigma^2}} \end{aligned}$$

To start, we re-write the test statistic as:

$$
\begin{aligned}
t &= \frac{\widehat{\beta} - b}{\sqrt{\sigma^2 \left((X^t X)_{jj}^{-1}\right)}} \cdot \sqrt{\frac{\sigma^2}{s^2}} \\
&= \frac{z}{\sqrt{s^2/\sigma^2}} \\
&= \frac{z}{q/(n-p)}
\end{aligned}
$$

To start, we re-write the test statistic as:

$$
\begin{aligned}
t &= \frac{\widehat{\beta} - b}{\sqrt{\sigma^2 \left((X^t X)^{-1}_{jj} \right)}} \cdot \sqrt{\frac{\sigma^2}{s^2}} \\
&= \frac{z}{\sqrt{s^2 / \sigma^2}} \\
&= \frac{z}{q / (n - p)}
\end{aligned}
$$

Where $q = r^t r / \sigma^2$.

To start, we re-write the test statistic as:

$$\begin{aligned} t &= \frac{\widehat{\beta} - b}{\sqrt{\sigma^2 \left((X^t X)^{-1}_{jj} \right)}} \cdot \sqrt{\frac{\sigma^2}{s^2}} \\ &= \frac{z}{\sqrt{s^2/\sigma^2}} \\ &= \frac{z}{q/(n-p)} \end{aligned}$$

Where $q = r^t r/\sigma^2$.

We need to show that (1) $q|X \sim \chi^2_{n-p}$ and (2) $z \perp\!\!\!\perp q|X$.

Lemma If B is a symmetric idempotent matrix and $u \sim \mathcal{N}(0, \mathbb{I}_n)$, then $u^t B u \sim \chi^2_{\mathrm{tr(B)}}$.

Proof: The symmetric matrix B can be written as by its eigen-decompostion; for some orthonormal Q matrix and diagonal matrix Λ:

$$B = Q^t \Lambda Q$$

Notice that as B is idempotent:

$$(Q^t \Lambda Q) = (Q^t \Lambda Q)(Q^t \Lambda Q)$$

Notice that as B is idempotent:

$$
\begin{aligned}
(Q^t \Lambda Q) &= (Q^t \Lambda Q)(Q^t \Lambda Q) \\
&= Q^t \Lambda Q Q^t \Lambda Q
\end{aligned}
$$

Notice that as B is idempotent:

$$\begin{aligned}
(Q^t \Lambda Q) &= (Q^t \Lambda Q)(Q^t \Lambda Q) \\
&= Q^t \Lambda Q Q^t \Lambda Q \\
&= Q^t \Lambda^2 Q
\end{aligned}$$

Notice that as B is idempotent:

$$\begin{aligned}(Q^t \Lambda Q) &= (Q^t \Lambda Q)(Q^t \Lambda Q) \\ &= Q^t \Lambda Q Q^t \Lambda Q \\ &= Q^t \Lambda^2 Q\end{aligned}$$

Since $Q^t = Q^{-1}$:

$$\Lambda = \Lambda^2$$

Notice that as B is idempotent:

$$\begin{aligned}(Q^t \Lambda Q) &= (Q^t \Lambda Q)(Q^t \Lambda Q) \\ &= Q^t \Lambda Q Q^t \Lambda Q \\ &= Q^t \Lambda^2 Q\end{aligned}$$

Since $Q^t = Q^{-1}$:

$$\Lambda = \Lambda^2$$

Therefore all the elements of Λ are 0 or 1.

By rotating the columns of Q, we can actually write:

$$\Lambda = \begin{pmatrix} \mathbb{I}_{\mathrm{rank}(B)} & 0 \\ 0 & 0 \end{pmatrix}$$

By rotating the columns of Q, we can actually write:

$$\Lambda = \begin{pmatrix} \mathbb{I}_{\mathrm{rank}(B)} & 0 \\ 0 & 0 \end{pmatrix}$$

Notice that this also shows that:

$$\begin{aligned} \mathrm{tr}(B) &= \mathrm{tr}(Q^t \Lambda Q) \\ &= \mathrm{tr}(\Lambda Q Q^t) \\ &= \mathrm{tr}(\Lambda) \\ &= \mathrm{rank}(B) \end{aligned}$$

Now let $v = Q^t u$. We can see that the mean of v is zero:

$$\begin{aligned} \mathbb{E}v &= \mathbb{E}Q^t u \\ &= Q^t \mathbb{E}u \\ &= 0 \end{aligned}$$

Now let $v = Q^t u$. We can see that the mean of v is zero:

$$\begin{aligned} \mathbb{E}v &= \mathbb{E}Q^t u \\ &= Q^t \mathbb{E}u \\ &= 0 \end{aligned}$$

And the variance of v is:

$$\begin{aligned} \mathbb{E}vv^t &= Q^t \mathbb{E}(uu^t) Q \\ &= Q^t \mathbb{I}_n Q \\ &= \mathbb{I}_n \end{aligned}$$

Now we calculate the quadtratic form u^tBu by the transform of v. Substitute that $u = (Q^t)^{-1}v = Qv$:

$$u^tBu = v^tQBQ^tv$$

Now we calculate the quadtratic form u^tBu by the transform of v. Substitute that $u = (Q^t)^{-1}v = Qv$:

$$\begin{aligned} u^tBu &= v^tQBQ^tv \\ &= v^tQ(Q^t\Lambda Q)Q^tv \end{aligned}$$

Now we calculate the quadtratic form $u^t Bu$ by the transform of v. Substitute that $u = (Q^t)^{-1} v = Qv$:

$$\begin{aligned} u^t Bu &= v^t QBQ^t v \\ &= v^t Q(Q^t \Lambda Q)Q^t v \\ &= v^t QQ^t \Lambda QQ^t v \end{aligned}$$

Now we calculate the quadtratic form u^tBu by the transform of v. Substitute that $u = (Q^t)^{-1}v = Qv$:

$$\begin{aligned} u^tBu &= v^tQBQ^tv \\ &= v^tQ(Q^t\Lambda Q)Q^tv \\ &= v^tQQ^t\Lambda QQ^tv \\ &= v^t\Lambda v \end{aligned}$$

Now we calculate the quadtratic form $u^t B u$ by the transform of v. Substitute that $u = (Q^t)^{-1} v = Qv$:

$$\begin{aligned} u^t B u &= v^t Q B Q^t v \\ &= v^t Q (Q^t \Lambda Q) Q^t v \\ &= v^t Q Q^t \Lambda Q Q^t v \\ &= v^t \Lambda v \\ &= v^t \begin{pmatrix} \mathbb{I}_{\mathrm{tr}(B)} & 0 \\ 0 & 0 \end{pmatrix} v \end{aligned}$$

Now we calculate the quadtratic form $u^t B u$ by the transform of v. Substitute that $u = (Q^t)^{-1} v = Qv$:

$$\begin{aligned} u^t B u &= v^t Q B Q^t v \\ &= v^t Q (Q^t \Lambda Q) Q^t v \\ &= v^t Q Q^t \Lambda Q Q^t v \\ &= v^t \Lambda v \\ &= v^t \begin{pmatrix} \mathbb{I}_{\mathrm{tr}(B)} & 0 \\ 0 & 0 \end{pmatrix} v \\ &= \sum_{i=1}^{\mathrm{tr}(B)} v_i^2 \end{aligned}$$

Now we calculate the quadtratic form u^tBu by the transform of v. Substitute that $u = (Q^t)^{-1}v = Qv$:

$$\begin{array}{rcl} u^tBu & = & v^tQBQ^tv \\ & = & v^tQ(Q^t\Lambda Q)Q^tv \\ & = & v^tQQ^t\Lambda QQ^tv \\ & = & v^t\Lambda v \\ & = & v^t\begin{pmatrix} \mathbb{I}_{\mathrm{tr}(B)} & 0 \\ 0 & 0 \end{pmatrix}v \\ & = & \displaystyle\sum_{i=1}^{\mathrm{tr}(B)} v_i^2 \\ & \sim & \chi^2_{\mathrm{tr}(B)} \end{array}$$

Which completes the lemma.

(1) We know that $r^t r = \epsilon^t M \epsilon$, so:

$$q = \frac{r^t r}{\sigma^2}$$

(1) We know that $r^t r = \epsilon^t M \epsilon$, so:

$$\begin{aligned} q &= \frac{r^t r}{\sigma^2} \\ &= \frac{\epsilon^t}{\sigma} \cdot M \cdot \frac{\epsilon}{\sigma} \end{aligned}$$

(1) We know that $r^t r = \epsilon^t M \epsilon$, so:

$$\begin{aligned} q &= \frac{r^t r}{\sigma^2} \\ &= \frac{\epsilon^t}{\sigma} \cdot M \cdot \frac{\epsilon}{\sigma} \end{aligned}$$

Under Assumption V, $\frac{\epsilon}{\sigma} \sim \mathcal{N}(0, \mathbb{I}_n)$. Therefore from the lemma, then $q \sim \chi^2_{\mathrm{tr(M)}}$; last time we showed the tr(M) $= n - p$, which finishes the first part of the proof.

(2) The random variables $\widehat{\beta}$ and r are both linear combinations of ϵ, and are therefore jointly normally distributed. As they are uncorrelated (problem set 2), this implies that they are independent. Finally, this implies that $z = f(\widehat{\beta})$ and $q = g(r)$ and themselves independent.

So, therefore, we have:

$$
\begin{aligned}
t &= \frac{\widehat{\beta} - b}{\sqrt{\sigma^2 \left((X^t X)^{-1}_{jj} \right)}} \cdot \sqrt{\frac{\sigma^2}{s^2}} \\
&= \frac{z}{q/(n-p)} \\
&\sim t_{n-p}
\end{aligned}
$$

Simulation

F-Test

F-test

Now consider the hypothesis test $H_0 : D\beta = d$ for a matrix D with k rows and rank k.

We'll form the following test statistic:

$$F = \frac{(D\widehat{\beta} - d)^t [D(X^t X)^{-1} D^t]^{-1} (D\widehat{\beta} - d)/k}{s^2}$$

And prove that is has an F-distribution with k and $n - p$ degrees of freedom.

We can re-write the test statistics as:

$$F = \frac{w/k}{q/(n-p)}$$

Where:

$$w = (D\widehat{\beta} - d)^t[\sigma^2 D(X^t X)^{-1} D^t]^{-1}(D\widehat{\beta} - d)$$
$$q = r^t r/\sigma^2$$

We can re-write the test statistics as:

$$F = \frac{w/k}{q/(n-p)}$$

Where:

$$\begin{aligned} w &= (D\widehat{\beta} - d)^t[\sigma^2 D(X^tX)^{-1}D^t]^{-1}(D\widehat{\beta} - d) \\ q &= r^t r/\sigma^2 \end{aligned}$$

We've already shown that $q|X \sim \chi^2_{n-p}$, and can see that $q \perp\!\!\!\perp w$ by the same argument as before. All that is left is to show that $w \sim \chi^2_k$.

Let $v = D\widehat{\beta} - d$. Under the null hypothesis $v = D(\widehat{\beta} - \beta)$.

Let $v = D\widehat{\beta} - d$. Under the null hypothesis $v = D(\widehat{\beta} - \beta)$. Therefore:

$$\mathbb{V}(v|X) = \mathbb{V}(D(\widehat{\beta} - \beta)|X)$$

Let $v = D\widehat{\beta} - d$. Under the null hypothesis $v = D(\widehat{\beta} - \beta)$. Therefore:

$$\begin{aligned} \mathbb{V}(v|X) &= \mathbb{V}(D(\widehat{\beta} - \beta)|X) \\ &= D\mathbb{V}(\widehat{\beta} - \beta|X)D^t \end{aligned}$$

Let $v = D\widehat{\beta} - d$. Under the null hypothesis $v = D(\widehat{\beta} - \beta)$. Therefore:

$$\begin{aligned} \mathbb{V}(v|X) &= \mathbb{V}(D(\widehat{\beta} - \beta)|X) \\ &= D\mathbb{V}(\widehat{\beta} - \beta|X)D^t \\ &= \sigma^2 D(X^t X)^{-1} D^t \end{aligned}$$

Let $v = D\widehat{\beta} - d$. Under the null hypothesis $v = D(\widehat{\beta} - \beta)$. Therefore:

$$\begin{aligned} \mathbb{V}(v|X) &= \mathbb{V}(D(\widehat{\beta} - \beta)|X) \\ &= D\mathbb{V}(\widehat{\beta} - \beta|X)D^t \\ &= \sigma^2 D(X^t X)^{-1} D^t \end{aligned}$$

Now, for the multivariate normally distributed v with zero mean, we have:

$$w = v^t \mathbb{V}(v|X)^{-1} v$$

Now, decompose $\mathbb{V}(v|X)^{-1} = \Sigma^{-1}$ as $Q^t \Lambda Q$. Notice that this implies:

$$\Sigma^{-1} = Q^t \Lambda^{1/2} \Lambda^{1/2} Q$$

Now, decompose $\mathbb{V}(v|X)^{-1} = \Sigma^{-1}$ as $Q^t \Lambda Q$. Notice that this implies:

$$\begin{aligned} \Sigma^{-1} &= Q^t \Lambda^{1/2} \Lambda^{1/2} Q \\ Q\Sigma^{-1} &= \Lambda^{1/2} \Lambda^{1/2} Q \end{aligned}$$

Now, decompose $\mathbb{V}(v|X)^{-1} = \Sigma^{-1}$ as $Q^t \Lambda Q$. Notice that this implies:

$$\begin{aligned}
\Sigma^{-1} &= Q^t \Lambda^{1/2} \Lambda^{1/2} Q \\
Q\Sigma^{-1} &= \Lambda^{1/2} \Lambda^{1/2} Q \\
\Lambda^{-1/2} Q \Sigma^{-1} &= \Lambda^{1/2} Q
\end{aligned}$$

From here, we can write w as the inner product of a suitably defined vector u:

$$w = v^t \Sigma^{-1} v$$

From here, we can write w as the inner product of a suitably defined vector u:

$$
\begin{aligned}
w &= v^t \Sigma^{-1} v \\
&= v^t Q^t \Lambda^{1/2} \Lambda^{1/2} Q v
\end{aligned}
$$

From here, we can write w as the inner product of a suitably defined vector u:

$$\begin{aligned} w &= v^t \Sigma^{-1} v \\ &= v^t Q^t \Lambda^{1/2} \Lambda^{1/2} Q v \\ &= u^t u \end{aligned}$$

From here, we can write w as the inner product of a suitably defined vector u:

$$\begin{aligned} w &= v^t\Sigma^{-1}v \\ &= v^tQ^t\Lambda^{1/2}\Lambda^{1/2}Qv \\ &= u^tu \end{aligned}$$

With a mean of zero:

$$\begin{aligned} \mathbb{E}(u|X) &= \mathbb{E}(\Lambda^{1/2}Qv|X) \\ &= \Lambda^{1/2}Q\mathbb{E}(v|X) \\ &= 0 \end{aligned}$$

And variance of

$$\mathbb{E}(uu^t|X) = \mathbb{E}(\Lambda^{1/2}Qvv^tQ^t\Lambda^{1/2}|X)$$

And variance of

$$\begin{aligned}\mathbb{E}(uu^t|X) &= \mathbb{E}(\Lambda^{1/2}Qvv^tQ^t\Lambda^{1/2}|X) \\ &= \Lambda^{1/2}Q\mathbb{E}(vv^t|X)Q^t\Lambda^{1/2}\end{aligned}$$

And variance of

$$\begin{aligned}\mathbb{E}(uu^t|X) &= \mathbb{E}(\Lambda^{1/2}Qvv^tQ^t\Lambda^{1/2}|X) \\ &= \Lambda^{1/2}Q\mathbb{E}(vv^t|X)Q^t\Lambda^{1/2} \\ &= \Lambda^{1/2}Q\Sigma Q^t\Lambda^{1/2}\end{aligned}$$

And variance of

$$\begin{aligned}
\mathbb{E}(uu^t|X) &= \mathbb{E}(\Lambda^{1/2}Qvv^tQ^t\Lambda^{1/2}|X) \\
&= \Lambda^{1/2}Q\mathbb{E}(vv^t|X)Q^t\Lambda^{1/2} \\
&= \Lambda^{1/2}Q\Sigma Q^t\Lambda^{1/2} \\
&= \Lambda^{-1/2}Q\Sigma^{-1}\Sigma Q^t\Lambda^{1/2}
\end{aligned}$$

And variance of

$$
\begin{aligned}
\mathbb{E}(uu^t|X) &= \mathbb{E}(\Lambda^{1/2}Qvv^tQ^t\Lambda^{1/2}|X) \\
&= \Lambda^{1/2}Q\mathbb{E}(vv^t|X)Q^t\Lambda^{1/2} \\
&= \Lambda^{1/2}Q\Sigma Q^t\Lambda^{1/2} \\
&= \Lambda^{-1/2}Q\Sigma^{-1}\Sigma Q^t\Lambda^{1/2} \\
&= \Lambda^{-1/2}\Lambda^{1/2}
\end{aligned}
$$

And variance of

$$\begin{aligned}
\mathbb{E}(uu^t|X) &= \mathbb{E}(\Lambda^{1/2}Qvv^tQ^t\Lambda^{1/2}|X) \\
&= \Lambda^{1/2}Q\mathbb{E}(vv^t|X)Q^t\Lambda^{1/2} \\
&= \Lambda^{1/2}Q\Sigma Q^t\Lambda^{1/2} \\
&= \Lambda^{-1/2}Q\Sigma^{-1}\Sigma Q^t\Lambda^{1/2} \\
&= \Lambda^{-1/2}\Lambda^{1/2} \\
&= \mathbb{I}_k
\end{aligned}$$

And, finally:

$$w = uu^t \sim \chi^2_k$$

Which finishes the proof.

An alternative F-test

Now, consider the following estimator:

$$\widetilde{\beta} = \arg\min_{b} ||y - Xb||_2^2, \quad \text{s.t.} \quad Db = d$$

An alternative F-test

Now, consider the following estimator:

$$\widetilde{\beta} = \arg\min_b ||y - Xb||_2^2, \quad \text{s.t.} \quad Db = d$$

We then define the restricted residuals $\widetilde{r} = y - X\tilde{\beta}$.

An alternative F-test

An alternative expression for the F-test statistic is:

$$F = \frac{(\tilde{r}^t\tilde{r} - r^t r)/k}{r^t r/(n-p)}$$

Conceptually, it should make sense that this is large whenever the null hypothesis is false.

An alternative F-test

If we let SSR_U be the sum of squared residuals of the unrestricted model ($r^t r$) and SSR_R be the sum of squared residuals of the restricted model, then this can be re-written as:

$$F = \frac{(\text{SSR}_U - \text{SSR}_R)/k}{\text{SSR}_U/(n-p)}$$

This is the way it is written in the homework and in the Fumio Hayashi text. It is left for you to prove that this is equivalent to the other F test.

Application