Fundamentos de Econometría

Ignacio Lobato

ITAM

Normal Multivariada

- Ahora la función de distribución conjunta del vector $\underset{n\times 1}{Y}$ se distribuye normal
- Su función de distribución es:

$$f\left(Y\right) = \left(2\pi\right)^{-n/2} \left|\Sigma\right|^{-1/2} e^{-w/2}$$

- Donde $w = \left(Y - \mu\right)' \Sigma^{-1} \left(Y - \mu\right)$ es un escalar
- $\underset{n\times 1}{\mu}$ es un vector de medias
- $\underset{n\times n}{\Sigma}$ es una matriz positiva definida de varianzas y covarianzas

Normal Multivariada

- Para una normal multivariada Y, partimos el vector en 2 subvectores $\underset{n_1\times 1}{Y_1}$ y $\underset{n_2\times 1}{Y_2}$
- Se cumple que: $E(Y) = \mu$ y $V(Y) = \Sigma$
- La distribución marginal de Y_1 es normal: $Y_1 \sim N(\mu_1, \Sigma_{11})$, esto se cumple para Y_2 o cualquier subvector de Y
- $Y_2|Y_1 \sim N(\mu_2^*, \Sigma_{22}^*)$
  - $\mu_2^* = E(Y_2|Y_1) = \alpha + B'Y_1$
  - $B = (\mathit{\Sigma}_{11})^{-1}\Sigma_{12}$
  - $\alpha = \mu_2 - B'\mu_1$
  - $\Sigma_{22}^* = V(Y_2|Y_1) = \Sigma_{22} - B'\Sigma_{11}B$

Normal Multivariada

- **Teorema:** para una normal multivariada, ausencia de correlación implica independencia
- Demostración:
  - Y_1 y Y_2 no están correlacionadas, entonces $\Sigma_{12} = 0$
  - Entonces B=0, por tanto $\mu_2^* = \alpha = \mu_2$ y $\Sigma_{22}^* = \Sigma_{22}$
  - En consecuencia $Y_2 | Y_1 \sim N(\mu_2, \Sigma_{22})$ para todo Y_1
  - La distribución condicional es igual a la marginal, por lo tanto Y_1 y Y_2 son independientes

Normal Multivariada

- Funciones lineales de un vector multinormal se distribuyen normal
- Para un vector multinormal Y, $z = g + HY$, se cumple:

$$z \sim N\left(g + H\mu, H\Sigma H'\right)$$

- H debe ser no estocástica y de rango completo para garantizar que la matriz de varianzas y covarianzas no sea singular

Vector Normal Estándar

- Si $\underset{n\times 1}{z} \sim N(0, I)$, decimos que z es un vector normal estándar
- Si $\underset{n\times 1}{z}$ es un vector normal estándar:

$$w = z'z \sim \chi^2(n)$$

- Si $w_1 \sim \chi^2(m)$ y $w_2 \sim \chi^2(n)$ son independientes:

$$v = \frac{w_1/m}{w_2/n} \sim F(m, n)$$

- Si $z \sim N(0, 1)$ y $w \sim \chi^2(n)$ son independientes:

$$u = \frac{z}{\sqrt{w/n}} \sim t(n)$$

Resultados Asintóticos

- Muestras grandes, $n \to \infty$
- Si $v \sim F(m, n)$, $mv \to^D \chi^2(m)$
  - Escribimos $v = \frac{w_1/m}{w_2/n}$, entonces $mv = \frac{w_1}{w_2/n}$
  - $\frac{w_2}{n} = \frac{1}{n}\sum_i z_i^2$, es una media muestral de la variable z^2 que tiene media 1
  - $\frac{w_2}{n} \to^p 1$, entonces $\frac{w_1}{1} \to^p \chi^2(m)$
- Si $u \sim t(n)$, $u \to^D N(0, 1)$
  - Escribimos $u = \frac{z}{\sqrt{(w/n)}}$
  - De la prueba anterior vimos que $\frac{w}{n} \to^p 1$, entonces u converge a la misma distribución que z

Formas Cuadráticas

- **Teorema:** sea $\underset{n\times 1}{Y} \sim N(\mu, \Sigma)$, entonces
  $w = (Y-\mu)' \Sigma^{-1} (Y-\mu) \sim \chi^2(n)$
  - Como Σ es positiva definida, la podemos escribir como $\Sigma = C\Lambda C'$, donde Λ es diagonal con los eigenvalores de Σ y C es ortonormal con los eigenvectores de Σ como columnas
  - Sea $\Lambda^* = \sqrt{\Lambda^{-1}}$
  - Sea $H = C\Lambda^* C'$, entonces $H' = H$, $H'H = C\Lambda^{-1}C' = \Sigma^{-1}$ y $H\Sigma H' = I$
  - Sea $\epsilon = Y - \mu$, entonces $\epsilon \sim N(0, \Sigma)$
  - Por último, sea $z = H\epsilon$, entonces $z \sim N(0, I)$
  - Definimos $w = \epsilon'\Sigma^{-1}\epsilon = \epsilon' H' H\epsilon = (H\epsilon)'(H\epsilon) = z'z$

- **Teorema:** sea $\underset{n\times 1}{z} \sim N(0, I)$ y $\underset{n\times n}{M}$ una matriz no estocástica de rango $r \leq n$, entonces $w = z'Mz \sim \chi^2(r)$
  - Como M es simétrica e idempotente, $M = C\Lambda C'$, donde C es ortonormal y Λ es diagonal
  - Los eigenvalores de M son r 1's y n-r 0's:

$$\Lambda = \begin{pmatrix} I_{r\times r} & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{pmatrix}$$

  - Partimos C como $C = \left(\underset{n\times r}{C_1}, \underset{n\times(n-r)}{C_2} \right)$
  - $\underset{r\times 1}{u} = C_1'z$ multinormal con media 0 y varianza I
  - $w = z'Mz = z'(C_1C_1')z = (z'C_1)(C_1'z) = u'u$

Inferencia

- Varios tipos de inferencia
- Dependiendo la cantidad de variables:
  - Univariada
  - Multivariada
- Dependiendo los supuestos sobre las distribuciones:
  - Aproximada
  - Exacta

Inferencia Exacta

- Hacemos supuestos sobre las distribuciones de la muestra
- Las distribuciones de los estadísticos estimados son exactas
- Por ejemplo, bajo el modelo clásico:

$$Y \sim^{E} N\left(X\beta, \sigma^2 I\right), \ X \, no \, estocástica$$

- Adicionalmente se asume que la varianza es conocida

Prueba de Hipótesis

- Univariada: se refiere a que se hace inferencia sobre un único parámetro o combinación lineal de parámetros
- $\hat{\beta}_j \sim^E N\left(\beta, \sigma^2 q_{ii}\right) \beta_j$ y el ICA al 95 % para β_j es:

$$\beta_j \in \left(\hat{\beta}_j \pm 1{,}96\sigma\sqrt{q_{ii}}\right)$$

- Para el i-ésimo regresor, se desea probar la hipótesis $H_0 : \beta_i = \beta_i^0$ vs $H_1 : \beta_i \neq \beta_i^0$
- El estadístico es:

$$z_i^0 = \left(b_j - \beta_j^0\right) / \hat{\sigma}\sqrt{q_{ii}} \sim N\left(0, 1\right)$$

- Comparamos $\left|z_i^0\right|$ con el valor crítico de la normal correspondiente al nivel de significancia deseado
- Por ejemplo, al 95 %:
  - Si $\left|z_i^0\right|$ >1.96, se rechaza la hipótesis nula $H_0 : \beta_i = \beta_i^0$
  - Si $\left|z_i^0\right| \leq 1.96$, se acepta la hipótesis nula $H_0 : \beta_i = \beta_i^0$

Intervalo de Confianza Exacto

- La formula para el intervalo de confianza es:

$$b_j \pm z_\alpha \hat{\sigma} \sqrt{q_{ii}}$$

- Es decir, estimación puntual $\pm$ z_α(error estándar)
- Como se asumió distribución normal, el error estándar es conocido
- Si β_i^0 está dentro del intervalo de confianza, se acepta la hipótesis nula

Inferencia Multivariada

- Ahora queremos estimar un conjunto de p restricciones lineales entre los coeficientes de β
- Por ejemplo:

$$\beta_1 = 0$$
$$\beta_2 + \beta_3 = 1$$

- En otras palabras, queremos estimar $\underset{p\times 1}{\theta} = \underset{p\times k}{R}\,\underset{k\times 1}{\beta} - \underset{p\times 1}{r}$, donde H es una matriz no estocástica de rango p
- La hipótesis nula es $H_0 : \theta = 0$
- Para nuestro ejemplo: $R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$, $r = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

Prueba de Hipótesis

- El estadístico que se propone es:

$$w^0 = (\theta_n - \theta)' \left[\sigma^2 R \left(X'X\right)^{-1} R'\right]^{-1} (\theta_n - \theta) \sim \chi^2_p$$

- Se compara con c_p, el valor crítico superior (cola derecha) de la distribución χ^2 con p grados de libertad
- Si $w^0 > c_p$, se rechaza la hipótesis nula $H_0 : \theta = 0$
- Si $w^0 \leq c_p$, se acepta la hipótesis nula $H_0 : \theta = 0$

Elipses de Confianza

- Similar a los intervalos, pero ahora con múltiples dimensiones
- Ahora son regiones en $\mathbb{R}^p$
- La forma de la región es:

$$(\theta_n - \theta)' \left[\sigma^2 R \left(X'X\right)^{-1} R'\right]^{-1} (\theta_n - \theta) \leq c_p$$

- Si 0 cae dentro de la región, se acepta la hipótesis nula

Potencia de la Prueba

- La potencia de la prueba se refiere a la probabilidad de rechazar una hipótesis nula como función del verdadero parámetro
- Para una prueba al 95 %, $P\left[\left(w^0 > c\right)|\theta\right] \geq ,05$ con igualdad ssi $\theta = \theta^0$
- La potencia de a prueba es mayor al nivel de significancia excepto en el valor de la hipótesis nula
- La potencia es una función creciente en la distancia entre el valor de la hipótesis θ^0 y el valor real θ

σ^2 Desconocida

- Eliminamos el supuesto de varianza conocida
- Ahora la estimamos de forma insesgada con $\hat{\sigma}^2 = e'e/\left(n-k\right)$
- Para el caso univariado, el estadístico es:

$$\hat{z}_i = \frac{(b_i - \beta_i)}{\hat{\sigma}\sqrt{q_{ii}}}$$

- Para el caso multivariado:

$$\hat{w} = (\theta_n - \theta)' \left[p\hat{\sigma}^2 R \left(X'X\right)^{-1} R'\right]^{-1} (\theta_n - \theta)$$

σ^2 Desconocida

- **Teorema:** $w_0 = e'e/\sigma^2 \sim \chi^2_{n-k}$
- Demostración:
  - Sea $\epsilon = Y - X\beta$, entonces $\epsilon \sim N\left(0, \sigma^2 I\right)$
  - Definimos $u = \frac{1}{\sigma}\epsilon$, entonces $\underset{n\times 1}{u} \sim N(0, I)$
  - Ahora, $Y = X\beta + \epsilon = X\beta + \sigma u$, entonces $e = MY = M(X\beta + \sigma u) = \sigma M u$
  - Recordemos que M es idempotente, no estocástica y de rango n-k
  - Entonces $e'e = \sigma^2 u' M u$ y $w_0 = \frac{e'e}{\sigma^2} = u' M u \sim \chi^2_{n-k}$

σ^2 Desconocida

- **Teorema:** $\hat{w} = (\theta_n - \theta)' \left[p\hat{\sigma}^2 R \left(X'X \right)^{-1} R' \right] (\theta_n - \theta) \sim F_{p,n-k}$
- Demostración:
  - $\frac{\hat{\sigma}^2}{\sigma^2} = \left[e'e / (n-k) \right] / \sigma^2 = \left(e'e/\sigma^2 \right) / (n-k) = w_0 / (n-k)$
  - $\hat{w}/p = (w/p) / \left(\hat{\sigma}^2/\sigma^2 \right) = (w/p) / \left[w_0 / (n-k) \right]$
  - $w \sim \chi^2_p$ es independiente de $w_0 \sim \chi^2_{n-k}$

σ^2 Desconocida

- **Teorema:** $u_i = (b_i - \beta_i) / \hat{\sigma}\sqrt{q_{ii}} \sim t_{n-k}$
- Demostración:
  - $z_i = (b_i - \beta_i) / \hat{\sigma}\sqrt{q_{ii}}$
  - $\hat{\sigma}_{bi} / \sigma_{bi} = \sqrt{(\hat{\sigma}^2 q_{ii}) / (\sigma^2 q_{ii})} = \sqrt{\hat{\sigma}^2 / \sigma^2} = \sqrt{w_0 / (n-k)}$
  - $z_i \sim N(0,1)$ es independiente de $w_0 \sim \chi^2_{n-k}$

Intervalos de Confianza

- Seguimos utilizando la formula estimación puntual $\pm t_{\frac{\alpha}{2},(n-k)}$ error estándar:

$$b_i \pm t_{\frac{\alpha}{2},(n-k)} \hat{\sigma} \sqrt{q_{ii}}$$

- La distribución t depende del tamaño muestral
- Ahora se utiliza el valor crítico de la distribución t
- Definimos la región A como:

$$\begin{aligned} A &= \left\{ \beta_i - t_{\frac{\alpha}{2},(n-k)} \hat{\sigma} \sqrt{q_{ii}} \leq b_i \leq \beta_i + t_{\frac{\alpha}{2},(n-k)} \hat{\sigma} \sqrt{q_{ii}} \right\} \\ &= \left\{ \left| \frac{b_i - \beta_i}{\hat{\sigma} \sqrt{q_{ii}}} \right| \leq t_{\frac{\alpha}{2},(n-k)} \hat{\sigma} \sqrt{q_{ii}} \right\} \end{aligned}$$

- El intervalo es $P(A) = 1 - \alpha$, donde α es el nivel de significancia
- $(b_i - \beta_i) / \hat{\sigma} \sqrt{q_{ii}} \sim t_{n-k}$, entonces $c = t_{\alpha/2}$
- En una prueba de hipótesis, si el valor de la hipótesis nula $\beta_i = \beta_i^0$ cae dentro del intervalo, se acepta

Elipses de Confianza

- Similar que el intervalo, pero ahora para un vector $\underset{p\times 1}{\theta} = R\beta - r$ de parámetros
- La región queda dada por:

$$(\theta_n - \theta)' \frac{\left[pR\left(X'X\right)^{-1} R'\right]^{-1}}{\hat{\sigma}^2} (\theta_n - \theta) \leq F_{p,n-k}$$

- d se obtiene de la distribución $F_{p,n-k}$
- Si 0 cae dentro de la región, se acepta la hipótesis nula

Inferencia Aproximada

- Ahora relajaremos los supuestos que hemos hecho sobre la distribución de la muestra
- En la práctica no se conocen las distribuciones
- Los resultados ya no son exactos
- Nos basaremos ahora en propiedades asintóticas para obtener las distribuciones aproximadas

Estadístico de Wald

- Eliminamos el supuesto de normalidad
- Utilizamos el mismo estimador $\hat{\theta}$=Rb-r
- Nuestro estadístico para la prueba de hipótesis ahora es:

$$W_n = (Rb - r)' \left[\tilde{\sigma}_n^2 R \left(X'X \right)^{-1} R' \right]^{-1} (Rb - r)$$

- $\tilde{\sigma}_n^2 = \frac{e'e}{n}$, la estimación de la varianza ahora no tiene que ser insesgada, basta que sea consistente
- $W_n \to^D \chi_p^2$, la distribución ahora es aproximada

Estadístico de Wald

- Definimos los residuales de la regresión bajo la hipótesis nula como $e^{*\prime} e^{*}$
- Una manera alternativa de escribir el estadístico de Wald es:

$$F_n = \frac{e^{*\prime} e^{*} - e' e}{p e' e} = \frac{\sigma^{*2} - \tilde{\sigma}^2}{p \sigma^2} \rightarrow^{D} \frac{\chi^2_p}{p}$$

- Nuevamente, la distribución es aproximada
- Cuando n es grande, inferencia exacta e inferencia aproximada llevan a resultados similares