ML E TESTES WALD, LM E RAZÃO DE ML

Prof. Denisard Alves

1. **Princípio da Maximaverossimilhança (MV)**

Já discutimos em aulas anteriores o princípio da maximaverossimilhanaça-MV. Agora vamos dar um tratamento mais completo.
Nos anos recentes tem sido comum o uso de testes com base nos enfoques de Wald e do multiplicador de Lagrange. São testes baseados no enfoque de MV.
Suponha um vetor de observações:
$y' = [y_1, y_2, \ldots, y_n]$ de n observações amostrais dependentes de k+1 parâmetros desconhecido $\theta' = [\theta_0, \theta_1, \ldots, \theta_k]$. Vamos definir a função de densidade conjunta das observações da amostra como $f(y;\theta)$, que indica a dependência de θ.Essa densudade pode ser interpretada de duas formas diferentes. Para um dado θ, ela representa a probabilidade do conjunto de valores amostrais de y ser obtido. Alternativamente ela pode ser interpretada como dado o conjunto de valores da amostra qual a probanilidade da amostra ter sido gerada para valores diferentes de θ. Nesta última interpretação ela é chamada de função de verossimilhança, ou seja,

Função de verossimilhança = $\boldsymbol{L(\theta; y) = f(y; \theta)}$

É usual reverter símbolos da função de verossimilhança para enfatizar que o que está variando, está definindo a probabilidade dos valores da amostra, agora é o vetor de parâmetros θ. Maximizar a função de verossimilhança com respeito a θ, significa encontrar o conjunto de valores para θ que maximiza a probabilidade de se obter os valores observados da amostra. Então $\hat{\theta}$, o vetor de parâmetros estimados que maximiza essa probabilidade ou a verossimilhança é chamado de Estimador de Maximaverossemelhança (MLE)[1]. Em geral, é mais simples maximizar logaritmo da função de verossimilhança, ou seja:

$$\boldsymbol{l = \ln L(\theta; y)}$$

Logo,

$$\boldsymbol{\frac{\partial l}{\partial \theta} = \frac{1}{L}\frac{\partial L}{\partial \theta}} \text{ e o } \boldsymbol{\hat{\theta}}, \text{ que maximiza } \boldsymbol{l} \text{ também maximiza } \boldsymbol{L}.$$

As condições de primeira ordem de máximo, derivadas de l com relação a θ, que chama-se por $\boldsymbol{s(\theta; y)}$, também é chamada de *score function*, quando igualadas a zero fornecem as k+1 equações necessárias para se encontrar o vetor dos k+1 parâmetros estimados $\hat{\theta}$, que maximizam a verossimilhança, ou seja $\hat{\theta}$ é obtido pela solução das CPO:

[1] MLE do inglês Maximum Likelihood Estimator

$$s(\boldsymbol{\theta}; \boldsymbol{y}) = \frac{\partial \boldsymbol{l}}{\partial \boldsymbol{\theta}}$$

O uso do MLE é devido a suas propriedades desejáveis, que a seguir são sumarizadas.

2. Propriedades do MV

A atratividade do MLE decorre de suas propriedades desejáveis, em especial suas propriedades de grandes amostras ou assintóticas. Elas são obtidas sob condições bem gerais.

2.1 Consistência

Para um melhor entendimento do conceito de consistência que envolve o conceito de convergência em probabilidade, Farei uma breve revisão, sumarizando, os aspectos essenciais apresentadas ho livro do Amemiya[2], para isso começaremos com convergência de números reais.

Definição 2.1.1 Convergência de sequência de números reais) A sequência de números reais, $(\alpha_n\}, n = 1, 2, \ldots,$ converge para um número real α se para qualquer $\epsilon > 0$, existe um número inteiro *N* tal que para todo *n* > *N* temos

$$|\alpha_n - \alpha| < \epsilon.$$

Então, escrevemos $\alpha_n \to \alpha,$ quando $n \to \infty$ ou $\lim_{n\to\infty} \alpha_n = \alpha.$

Para levar este conceito para sequências de variáveis aleatórias ocorrem algumas diferenças. No caso de variáveis aleatórias 2.1 as vezes é verdadeira as vezes não, pois estamos tratando de variáveis aleatórias, bem diferente de convergência de números reais. Agora só podemos falar da probabilidade de 2.1 ser verdadeira e não afirmar que ela é verdadeira. Tal constatação nos sugere que 2.1 deve ser definida em outros termos, ou seja, pode-se afirmar que 2.1 ocorre com probabilidade se aproximando de 1 quando $n \to \infty$. Então nós temos:

Convergência em Probabilidade:

Uma sequência de variáveis aleatórias $\{y_n\}$, n=1.2,...se diz que converge *em probabilidade* para uma variável aleatória *y* se para qualquer $\epsilon > 0$ e $\delta > 0$ existe um número inteiro N para todo n > N, onde a $P(|y_n - y| < \epsilon) > 1 - \delta$. Escreve-se $y_n \xrightarrow{p} y$ quando $n \to \infty$ ou $\mathrm{plim}_{n\to\infty} y_n = y$, ou, alternativamente e mais usual, $P(|y_n - y| < \epsilon) = 1$ para qualquer $\epsilon > 0$

Diferentemente do caso de sequência de constantes onde apenas um tipo de convergência, como estabelecido em 2.1 é suficiente, no caso de sequências de variáveis aleatórias precisamos de dois tipos de convergência: ***convergência em média quadrática*** e ***convergência em distribuição***.[3]

Definição 2.1.2 (Convergência em Média Quadrática)

Diremos que uma variável aleatória y_n converge em média quadrática para y se $\lim_{n\to\infty} E(y_n - y)^2 = 0$ e dizemos que

$\boldsymbol{y_n} \xrightarrow{M} \boldsymbol{y}.$

---

[2] Veja AMEMIYA, T., Introcucyion to Statistics and Econometrics, Harvard University Press, Cambridge, MA, 1994, pp.100 -110.

[3] Existe um outro modo de convergência que é a *almost sure convergence*. Que não será discutida aqui, pois os dois tipos de convergência mencionados acima serão suficientes para a discussão sobre propriedades de grandes amostras desta nota.

Definição 2.1.3 (Convergência em Distribuição)
Uma sequência de variáveis aleatórias $\{y_n\}$ converge para y, em distribuição se a função distribuição de probabilidade dessa sequência, definida por $\mathbf{F_n}$, for igual a cada ponto de **F,** função de distribuição de probabilidades de y, e escrevemos $y_n \xrightarrow{d}$ y, e **chamamos F** a distribuição limite da sequência $\{y_n\}$.

Definição 2.1.4 Leis dos Grandes Números (LLN)
As LLN são teoremas sobre *convergência em probabilidade* (ou "*almost sure converegence"*) no caso especial onde $\{y_n\}$ é uma média amostral ou quando $y_n = \bar{y}_n$, onde:

$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i$.
Note que y_i aqui é uma notação geral para uma variável aleatória. Observe que, no contexto do modelo de regressão linear ela não representa a variável explicativa. No caso de regressão, y_i= $\boldsymbol{x_i u_i}$.
A LLN é uma forma mais fácil de se chegar ao plim do que as definições de convergência em probabilidade e convergência em média quadrática discutidas em 2.1.1 e 2.1.2 acima. As LLN são de grande uso em econometria, pois elas tratam de convergência de médias amostrais e os estimadores envolvem médias.
2.1.4 LLN (Lei dos Grandes Números Fraca)
Uma LLN fraca especifica condições sobre os termos individuais y_i em $\bar{y}_n$, sob as quais

$$\bar{y}_n - E(\bar{y}_n) \xrightarrow{p} 0$$ [4]

Aplicando-se a convergência em probabilidade ou a a convergência em média quadrática, ou mesmo a LLN fraca, dada as hipóteses do modelo de MV, com o estimadoe sendo função de ariáveis iid e que formam médias, temos garantido q

$$\text{plim}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$$

2.2 Normalidade Assintótica

No caso do plim, a LLN fraca dá respaldo a convergência de $\hat{\theta}$ para a média populacional θ. No caso da convergência em distribuição além de se supor que a média existe também é necessário a existência da variância, no caso em que a convergência em distribuição é para uma distribuição assintótica requer também a existência da variância-covariância assintótica.
Importante observar que consistência requer uma distribuição degenerada, ou seja a variância vai para zero quando aumenta o tamanho da amostra e o estimador converge para o verdadeiro valor do parâmetro. Neste caso não há como fazer inferência, a variância é zero, a distribuição se degenera em cima do verdadeiro valor do parâmetro, quando $n \to \infty$. É necessário fazer com que

[4] Uma LLN forte estas condições sobre os elementos y_i que compõem $\bar{y}_n$ levam a *almost sure convergence*.

a distribuição não se degenere para permitir a inferência estatística. Para isso será necessário magnificar ou fazer um reescalonamento na variável aleatória ou o estimador para que a convergência em distribuição possa ocorrer evitando que a distribuição se degenere quando quando $n \to \infty$.

É usualmente usado $\sqrt{n}$ como fator para reescalonar r a variável aleatória para se evitar que a distribuição se degenere. Então, considere $y_n = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$**.** y_n pode ter uma função de distribuição acumilada F_n complicada. Mas. Como qualquer outra função, F_n converge para uma função limite, como vimos na definição 2.1, onde convergência é não estocástica, no sentido matemático. Aplicando-se a definição 2.1.3, vista acima, temos se $\mathbf{F_n}$ for definida agora como a função de **distribuição de probabilidade** y_n**,** $\mathbf{F_n}$, for igual a cada ponto de continuidade de **F,** função de distribuição de probabilidades de y, e escrevemos $y_n \xrightarrow{d}$ y, e chamamos **F** a distribuição limite da sequência $\{y_n\}$.

Em geral $y_n \xrightarrow{d}$ y implica em $y_n \xrightarrow{p}$ y, mas o reverso não é verdadeiro.

Pela LLN ou pela convergência em probabilidade y_n converge para uma constante: a distribuição se degenera e $\lim_{n \to \infty} \bar{y}_n = \lim_{n\to\infty} E(\bar{y}_n) < \infty$[5].

Para eliminarmos o problema em que $\lim_{n\to\infty} [\bar{y}_n - E(\bar{y}_n)]$ser zero, rescalonamos $[y_n - E(y_n)]$ pelo desvio padrão, o que garante à variável aleatória reescalonada a existência de variância igual a 1.

Teorema do Limite Central

Dado

$$z_n = \frac{\bar{y}_n - E(\bar{y}_n)}{\sqrt{Var(\bar{y}_n)}}$$

O $\bar{y}_n$ é a média amostral. O Teorema do Limite Central especifica as condições sobre y_i para que z_n convirja assintoticamente para a $N(0,1)$ *ou* $z_n \xrightarrow{d} N(0,1)$.

No caso de MV e MQO como veremos, para $n \to \infty$ MQO, independente da distribuição dos erros, converge para a distribuição assontótica de MV.

No cado da MV a distribuição de $\widehat{\boldsymbol{\theta}}$ é normal e a distribuição assintótica é dada por:

$$\hat{\boldsymbol{\theta}} \overset{a}{\sim} N(\boldsymbol{\theta}, \boldsymbol{I}^{-1}(\boldsymbol{\theta}))$$

Com essa definição fica estabelecido que a distribuição assintótica de $\widehat{\boldsymbol{\theta}}$ tem média $\boldsymbol{\theta}$ e variância $\boldsymbol{I}^{-1}(\boldsymbol{\theta})$**,** que é a inversa **da *Matriz de Informação.*** A matriz de informação é definida de duas formas equivalentes:

$$\boldsymbol{I}(\boldsymbol{\theta}) = \boldsymbol{E}\left[\left(\frac{\partial l}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial l}{\partial \boldsymbol{\theta}}\right)'\right] = -\boldsymbol{E}\left[\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]$$

5 Supondo que a média exista e seja finita, a LLN adequada, no caso a de Kolmogorov. Veja Cameron,C em Cameron: Assintotic Theory for OLS, Notas de Aula E240, Theorem 18 e Definição 13.

Usualmente é mais *simples fazer os cálculos dos elementos da matriz de informação usando a segunda alternativa.* $\boldsymbol{\theta}$ *é k x 1*, portanto,

$$\frac{\partial l}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \partial l/\partial \theta_1 \\ \partial l/\partial \theta_2 \\ \vdots \\ \partial l/\partial \theta_k \end{bmatrix}$$

Cada um dos elementos desse vetor gradiente ou score, é em si mesmo uma função de $\boldsymbol{\theta}$, o que nos permite diferenciar cada elemento desse vetor com relação a cada elemento de $\boldsymbol{\theta}$, por exemplo, com relação ao primeiro elemento de $\boldsymbol{\theta}$ temos:

$$\frac{\partial[\partial l/\partial \boldsymbol{\theta}]}{\partial \theta_1} = \begin{bmatrix} \frac{\partial^2 l}{\partial \theta_1^2} & \frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 l}{\partial \theta_1 \partial \theta_k} \end{bmatrix}$$

Se a derivada for feita para cada elemento do vetor teremos a matriz kxk de derivadas segundas de $\boldsymbol{l}$**,** conhecida por matriz

Hessiana,

$$\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{bmatrix} \frac{\partial^2 l}{\partial \theta_1^2} & \frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 l}{\partial \theta_1 \partial \theta_k} \\ \frac{\partial^2 l}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 l}{\partial \theta_2^2} & \cdots & \frac{\partial^2 l}{\partial \theta_2 \partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \theta_k \partial \theta_1} & \frac{\partial^2 l}{\partial \theta_k \partial \theta_2} & \cdots & \frac{\partial^2 l}{\partial \theta_k^2} \end{bmatrix}$$

a. **Eficiência Assintót**ica

$\widehat{\boldsymbol{\theta}}$ é o MLE do vetor θ com $k+1$ parâmetros desconhecidos,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{a} N(\mathbf{0}, \boldsymbol{V})$$

onde $\boldsymbol{V}$ é a matriz positiva definida – PD de variância-covariância. Qualquer outro estimador também consistente, assintoticamente normal, com matriz de variância-covariância $\widetilde{\boldsymbol{V}}$ também PD. Então $\widetilde{\boldsymbol{V}}$**-** $\boldsymbol{V}$ é positiva semidefinida, o que garante a eficiência assintótica para o $\widehat{\boldsymbol{\theta}}$ MLE. Como exemplificaremos com o modelo de regressão linear clássico nesta nota,
$\boldsymbol{V} = \boldsymbol{I}^{-1}(\boldsymbol{\theta})$**,** que estabelece o *"Cramer-Rao Lower Bound"*, ou limite mínimo de Cramer-Rao, já apresentado em nota de aula anterior sobre propriedades dos estimadores.

c.1 Derivação da Matriz de Informação

Se tomarmos $\boldsymbol{s(\theta; y)} = \frac{\partial l}{\partial \theta}$, que é constituído de observações de uma amostra aleatória, que são funções de $\boldsymbol{\theta}$ tendo, pois uma distribuição de densidade de probabilidade, nos permite, então calcular encontrar o vetor de *k* médias zero e a matriz de variância, $I(\boldsymbol{\theta})$, *k* x *k,* destes *k* vetores[6]. Para demostrar a média zero, tomemos a função de distribuição de probabilidades conjuntas das das n observações da amostra, que será igual a 1:

$$\int \cdots \int f(y_1, y_2, \ldots, y_n; \theta) \boldsymbol{d} y_1 \cdots \boldsymbol{dy_n} = \int \cdots \int \boldsymbol{Ldy} = \mathbf{1}$$

Derivando ambos os lados da igualdade com relação a $\boldsymbol{\theta}$ obtemos:

$$\int \cdots \int \frac{\partial L}{\partial \theta} dy = 0$$

Mas,

$$\boldsymbol{E(s)} = \int \cdots \int \frac{\partial l}{\partial \theta} \boldsymbol{L dy}$$

E,

$$\boldsymbol{Var(s)} = \boldsymbol{E(ss')} = \boldsymbol{E}\left[\left(\frac{\partial \boldsymbol{l}}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \boldsymbol{l}}{\partial \boldsymbol{\theta}}\right)'\right] = \boldsymbol{I(\theta)}$$

Ou seja, temos então a média e variância do gradiente ou *score function.*

3. Estimador de MV do modelo de regressão clássico

Como vimos, em aulas anteriores, o modelo linear de regressão é dado por:

$\boldsymbol{y = X\beta + u}$,

onde $\boldsymbol{y}$ é $\boldsymbol{n\ x\ 1}$, $\boldsymbol{X}$ é $\boldsymbol{n\ x\ (k+1)}$, $\boldsymbol{\beta}$ é $\boldsymbol{(k+1) x\ 1\ e\ u}$ é $\boldsymbol{n}$ **x** $\mathbf{1}$.

Adicionando-se ao modelo linear as 5 hipóteses de Gauss-Markov teremos as condições para o modelo ser BLUE e se, além delas adicionarmos a hipótese 6 de normalidade para o termo erro:

H6 $\boldsymbol{u \sim \mathrm{N}(0, \sigma^2 I_n)}$.

A função densidade multivariada normal de **u** é:

$$\boldsymbol{f(u)} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)(\boldsymbol{u'u})}$$

A função multivariada de

$$f(\boldsymbol{y} \mid \boldsymbol{X}) = f(\boldsymbol{u}) \left| \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{y}} \right|$$

[6] Farei aqui uma adaptação da apresentação do Maximum Likelihood Estimation with Stata, Gold, W., j. Pitblado, B. Poi, 4TH Edition, Stata Press, 2004, cap 1.

onde $|\frac{\partial u}{\partial y}|$ é o *Jacobiano* da transformação de $\boldsymbol{u}$ em $\boldsymbol{y}$ e é dado pelo determinante da seguinte matriz n x n:

$$\begin{bmatrix} \frac{\partial u_1}{\partial y_1} & \frac{\partial u_1}{\partial y_2} & \cdots & \frac{\partial u_1}{\partial y_n} \\ \frac{\partial u_2}{\partial y_1} & \frac{\partial u_2}{\partial y_2} & \cdots & \frac{\partial u_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_n}{\partial y_1} & \frac{\partial u_n}{\partial y_2} & \cdots & \frac{\partial u_n}{\partial y_n} \end{bmatrix}$$

No caso do modelo de regressão linear clássico essa matriz é a identidade, pois $\frac{\partial u_i}{\partial y_j} = 1, \text{se } i = j;$ e $= 0$ para $i \neq j$. A razão é simples; cada observação da amostra é independente da outra, logo não existe entre correlação entre o erro de uma observação e a variável dependente de outra.
A log da verossimilhança é:

$$l = \ln f(\boldsymbol{y} \mid \boldsymbol{X}) = \ln f(\boldsymbol{u}) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\boldsymbol{u}'\boldsymbol{u}$$

$$= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})$$

O vetor de parâmetros desconhecidos, $\boldsymbol{\theta}'$ neste caso terá k + 2 parâmetros desconhecidos, k+1, $\beta s \ e \ \sigma^2$.

$$\boldsymbol{\theta}' = [\boldsymbol{\beta}', \sigma^2]$$

Tomando derivadas parciais com relação aos parâmetros encontramos as equações do gradiente que quando igualadas a zero para se obter os estimadores dos parâmetros se tornam em condições de 1ª ordem-CPO, necessárias para o máximo da função log da verossimilhança:

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = -\frac{1}{\sigma^2}(-\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{X}'\boldsymbol{X\beta})$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})$$

As soluções dessas CPO nos rendem:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

$$\hat{\sigma}^2 = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})/n$$

Como os elementos de fora da diagonal principal de

Como vemos o estimador obtido pau o vetor $\widehat{\boldsymbol{\beta}}$ é exatamente igual a solção de MQO para o modelo de regressão linear clássico. Logo este $\widehat{\boldsymbol{\beta}}$ de ML apresenta todas as propriedades desejáveis vistas para o estimador MQO. Mas, o mesmo não ocorre para o estimador $\hat{\sigma}^2$, que, como se observa é viesado, pois sua $E(\hat{\sigma}^2)$=$\sigma^2 \frac{n-k-1}{n}$,tomando a derivadas segundas do gradiente com relação a β e σ^2, temos

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\frac{\boldsymbol{X}'\boldsymbol{X}}{\sigma^2}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \sigma^2} = -\frac{\boldsymbol{X}'\boldsymbol{u}}{\sigma^4}$$

$$\frac{\partial^2 l}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{\boldsymbol{u}'\boldsymbol{u}}{\sigma^6}$$

Tomando o negativo da esperança matemática, obtemse os elementos da matriz de informação:

$$-E\left(\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right) = \frac{\boldsymbol{X}'\boldsymbol{X}}{\sigma^2}$$

$$-E\left(\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \sigma^2}\right) = \mathbf{0}$$

$$-E\left(\frac{\partial^2 l}{\partial (\sigma^2)^2}\right) = \frac{n}{2\sigma^4}$$

Para se obter o último resultado lembre-se que $E(u ú) = E(\sum u_i^2 = n\sigma^2$.
E a matriz de informação é:

$$\boldsymbol{I}(\boldsymbol{\theta}) = \boldsymbol{I}\begin{pmatrix} \boldsymbol{\beta} \\ \sigma^2 \end{pmatrix} = \begin{bmatrix} \frac{1}{\sigma^2}(\boldsymbol{X}'\boldsymbol{X}) & \mathbf{0} \\ \mathbf{0} & \frac{n}{2\sigma^4} \end{bmatrix}$$

A sua inversa nos fornecerá a matriz de variância-covariância do estimador ML que coincide com o *Cramer-Rao Lower Bound:*

$$\boldsymbol{I}^{-1}\begin{pmatrix} \boldsymbol{\beta} \\ \sigma^2 \end{pmatrix} = \begin{bmatrix} \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{2\sigma^4}{n} \end{bmatrix}$$

Como é possível observar em $\boldsymbol{I}^{-1}\binom{\boldsymbol{\beta}}{\sigma^2}$ os elementos fora da diagonal principal são zeros, indicando que $\beta\ e\ \sigma^2$são independentemente distribuídos.Type equation here.
Substituindo os estimadores de MV na função de verossimilhança, obtemos;
$L(\hat{\beta}\ \hat{\sigma}^2)$.
Substituindo os valores de $\hat{\beta}$ e de $\hat{\sigma}^2$nafunção de verossimilhança e tomando-a em termos exponenciais obtemos:

$$\begin{aligned} L\left(\hat{\beta}\ \hat{\sigma}^2\right) &= (2\pi \mathrm{e})^{-\frac{n}{2}} (\hat{\sigma}^2)^{-\frac{n}{2}} \\ &= \left(\frac{2\pi \mathrm{e}}{n}\right)^{-\frac{n}{2}} (\hat{u}'\hat{u})^{-\frac{n}{2}} \\ &= \text{Constante.}\ (\hat{u}'\hat{u})^{-\frac{n}{2}} \end{aligned}$$

No caso a Constante não depende de nenhum parâmetro do modelo.

O elemento (1,1) que é a matriz de variância-covariância de $\widehat{\boldsymbol{\beta}}$ de ML, que coincide com a de MQO é o limite ínimo de Cramer-Rao, o que implicitamente demonstra que MQO é assintoticamente eficiente.

4. Testes de LR, WALD e LM

Vamos ilustrar estes testes no âmbito de hipóteses lineares sobre os βs. Elas terão o seguinte formato:

H_0: $\boldsymbol{R\beta} = \boldsymbol{r}$

Onde $\boldsymbol{R}$ é uma matriz conhecida $q\ x\ k$+1, $\boldsymbol{\beta}$ é um vetor (k+1) x 1 e $\boldsymbol{r}$ é um vetor também de elementos conhecidos $q\ x$ 1. E q < k+1.

Os testes serão os testes LR, W e LM.

4.1 Teste de LR

Os estimadores de MV desenvolvidos acima definem o máximo da função de verossimilhança sem impor nenhuma restrição sobre os parâmetros. Vamos representar o máximo da função de verossimilhança como:

$L(\widehat{\boldsymbol{\beta}}; \hat{\sigma}^2)$. É importante que seja observado que $\hat{\sigma}^2$ é expresso como função da soma dos resíduos sem restrições ao quadrado, $\hat{u}'\hat{u}$. O estimador de MV a semelhança do estimador de MQO, também pode ser obtido através da maximização do log da verossimilhança impondo Ho: $R\beta = r$. Ou seja $max_\beta(\tilde{l}) = l - \gamma'(R\beta - r)$, onde é um vetor q x 1 de multiplicadores de Lagrange,

Assim obtém-se $L(\tilde{\beta}\,; \tilde{\sigma}^2)$, que corresponde a soma de quadrados dos res;iduos do estimador ML com restrições, $\tilde{u}'\tilde{u}$.

É fácil entender que $L(\tilde{\beta}\,; \tilde{\sigma}^2) \leq L(\widehat{\boldsymbol{\beta}}; \hat{\sigma}^2)$. Quando se restringe os parâmetros a função de verossimilhança jamais terá um valor máximo superior ao da função sem restrições.

A estatística de LR é obtida pela razão entre a função sem restrição sobre o valor da que é obtido pela função quando se restringe o valor dos parâmetros especificados na hipótese nula. A estatística então é:

$\lambda = \frac{L(\widehat{\boldsymbol{\beta}}; \hat{\sigma}^2)}{L(\tilde{\beta}\,; \tilde{\sigma}^2)}$.

Esta estatística necessita de transformações especiais no caso de amostras finitas ficam bem complexas e tem que ser encontrada a distribuição amostral de λ caso a caso. Mas, para grandes amostras a estatística de teste tem distribuição amostral definida e é dada por:

$$\mathrm{LR} = -2\ln\lambda = 2[\ln L(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) - \ln L(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2)] \overset{a}{\sim} \chi^2(q)$$

O estimador de MV restrito é obtido pelo $max_\beta(\tilde{l})$ e substituindo

$L(\hat{\beta}\ \hat{\sigma}^2) = \text{Constante}.\,(\hat{u}'\hat{u})^{-\frac{n}{2}}$ e

Fazendo o mesmo para o máximo da função de verossimilhança restrita;

$$L(\tilde{\beta}\ \tilde{\sigma}^2) = \text{Constante}.\,(\tilde{u}'\tilde{u})^{-\frac{n}{2}}$$

Substituindo esses valores em:

$LR = n\left[ln\ L\left(\hat{\beta}\ \hat{\sigma}^2\right) - ln\ L\left(\tilde{\beta}\ \tilde{\sigma}^2\right)\right]$

Obtemos;

$LR = n[\ln\ (\tilde{u}'\tilde{u}) - ln\ (\hat{u}'\hat{u})]$,

Colocado o LR de uma forma mais adequada para mostrarmos. Posteriormente, que WAL$D\ \geq LR\ \geq LM$, temos:

$$LR = n\left[ln\frac{\tilde{u}'\tilde{u}}{\hat{u}'\hat{u}}\right] = n[\ln\left(\frac{\hat{u}'\hat{u}}{\hat{u}'\hat{u}} + \frac{\tilde{u}'\tilde{u}}{\hat{u}'\hat{u}} - \frac{\hat{u}'\hat{u}}{\hat{u}'\hat{u}}\right)]$$

$$= n[\ln\left(1 + \frac{\tilde{u}'\tilde{u} - \hat{u}'\hat{u}}{\hat{u}'\hat{u}}\right)]$$

$= n[\ln\left(1 + \frac{\tilde{u}'\tilde{u} - \hat{u}'\hat{u}}{\hat{u}'\hat{u}}\right)]$, ou

$$LR = n\left[\ln\ (\frac{\hat{u}'\hat{u}}{\tilde{u}'\tilde{u}})^{-1}\right] = n\left[\ln\left(\frac{1}{\frac{\tilde{u}'\tilde{u}}{\tilde{u}'\tilde{u}} + \frac{\hat{u}'\hat{u}}{\tilde{u}'\tilde{u}} - \frac{\tilde{u}'\tilde{u}}{\tilde{u}'\tilde{u}}}\right)\right]$$

$$= n\left[\ln\left(\frac{1}{1 - \frac{(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})}{\tilde{u}'\tilde{u}}}\right)\right]$$

Como vemos o cálculo da estatística LR requer a estimação dos modelos sem e com restrições para se obter a soma dos quadrados dos resíduos dos dois modelos.

4.2 Teste de Wald (W)[7]

No teste **de W**ald apenas o vetor $\hat{\beta}$ de MV da regressão sem restrições é estimado. A partir dos coeficientes estimados aplica-se as restrições, dadas por Ho: $R\beta = r$, e verifica se a redução da função de verossimilhança é "pequena" ou "grande", sendo que pequena ou grande terá métrica definida pela distribuição da estatística de teste, W.
Como:

$\hat{\beta} \sim N[\hat{\beta}, I^{-1}(\beta)]$, sob $\boldsymbol{Ho}$, $(\boldsymbol{R\hat{\beta} - r}) \sim \boldsymbol{N}[\boldsymbol{0}, \boldsymbol{R}I^{-1}(\beta)R']$, onde $\boldsymbol{I^{-1}(\beta) = \sigma^2(X'X)^{-1}}$. Como vimos acima a matriz de informaçãp é bloco diagonal logo podemos centrar a atenção na sub-matriz relativa a β. Segue então que:

$$\left(\boldsymbol{R\hat{\beta} - r}\right)'[\boldsymbol{RI^{-1}(\beta)R'}]^{-1}\left(\boldsymbol{R\hat{\beta} - r}\right) \overset{a}{\sim} \boldsymbol{\chi^2_{(q)}},$$

se distribui assintoticamente como uma Chi$^{\mathbf{2}}$, com q graus de liberdade.

---

[7] Veja, Wooldridge.J., Introductory Econometrics: A modern Approach, 6[th] , Edition, Cengage, Appendix e, pp. 737-31 para uma discussão excelente dos testes assintóticos

onde q é o número de linhas da matriz R ou o número de restrições impostas sobre o vetor β. A distribuição assintótica ainda se mantém quando substituímos $I^{-1}(\beta)$ por $\hat{\boldsymbol{\sigma}}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

$$\mathrm{W} = \frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})}{\hat{\boldsymbol{\sigma}}^2} \overset{a}{\sim} \chi^2(q)$$

Com as hipóteses H1 a H6[8] é fácil de ver que o Wald se reduz ao teste de F, quando dividida por q:

$$\boldsymbol{W/q} = \frac{\dfrac{(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})'}{\boldsymbol{q}}}{\dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\boldsymbol{n}}} \sim \boldsymbol{F(q, n-k-1)}$$

Nós já vimos acima que o numerador da estatística Wald pode ser expresso como $(\tilde{u}'\tilde{u} - \widehat{\hat{u}'u})$, que é a diferença da soma dos quadrados da regressão com restrição menos a soma de quadrados da regressão sem restrições. Portanto, $\boldsymbol{W}$

$$\boldsymbol{W} = \frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})}{\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\boldsymbol{n}}} \overset{a}{\sim} \boldsymbol{\chi^2_{(q)}}$$

$$\boldsymbol{W} = \frac{(\tilde{\mathbf{u}}'\tilde{\mathbf{u}} - \widehat{\hat{\mathbf{u}}'\mathbf{u}})}{\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\boldsymbol{n}}} = \frac{n(\tilde{\mathbf{u}}'\tilde{\mathbf{u}} - \widehat{\hat{\mathbf{u}}'\mathbf{u}})}{\hat{\mathbf{u}}'\hat{\mathbf{u}}} \overset{a}{\sim} \boldsymbol{\chi^2_{(q)}}$$

4.3 Teste de LM[9]

O teste de LM é conhecido como o teste do s*core*, pois ele se baseia nas condições de primeira ordem para maximização da MV:

$$\boldsymbol{s(\theta)} = \frac{\partial \ln L}{\partial \boldsymbol{\theta}} = \frac{\partial l}{\partial \boldsymbol{\theta}}$$

---

[8] Sob H6 a estatística Wald tem distribuição Chi[2] exata, logo quando dividida pelo número de restrições terá a distribuição F exata obtida para amostras finitas.

[9] Veja, Wooldridge, J., *op. cit.*, Chap. 5, onde só o resultado $LM = nR^2 \sim \chi^2_{(q)}$ é apresentado.

O estimador sem restrições, $\hat{\theta}$, que é obtido através da soloção as CPO: $\boldsymbol{s(\theta)} = \mathbf{0},$ que fornece o *score vector* avaliado em $\hat{\theta}$. Quando se impõem restrições sobre os parâmetros, para se obter o estimador com restrições, o $\boldsymbol{s(\theta)}$, quando avaliado em $\tilde{\theta}$ não será zero. Mas, se as restrições forem válidas, o valor do máximo restrito $l(\tilde{\theta}) \approx l(\hat{\theta})$, e $\boldsymbol{s}(\tilde{\theta}) \approx 0$. Como vimos anteriormente, o $\boldsymbol{s(\theta)}$ tem média zero e variância $\boldsymbol{I(\theta)}$. A forma quadrática $\boldsymbol{s'(\theta)I^{-1}(\theta)s(\theta)}$ que está [ponderada pela matriz de variância-covariância, $\boldsymbol{I^{-1}(\theta)}$, tem uma distribuição χ^2. Avaliando esta forma quadrática em $\theta = \tilde{\theta}$ fornece um test4 para se testar a hipótese nula. O resultado importante é que sob a hipótese nula com *q* restrições nos coeficientes, resulta que:

$$\boldsymbol{LM = s'(\tilde{\theta})I^{-1}(\tilde{\theta})s(\tilde{\theta})} \overset{a}{\sim} \chi^2_{(q)}$$

Em contraste com o teste de Wald onde se estimava a regressão sem restrições agora basta estimar a regressão com restrições para se construir a estatística LM. Ë uma das razões para a popularidade do LM, mas a principal está na facilidade em aplica-lo, pois em muitas situações é bem mais simples estimar a regressão com restrições e aplicar o teste. Vimos acima que o vetor *score* ou as CPO é dado por:

$$\boldsymbol{s(\theta)} = \begin{bmatrix} \dfrac{\partial l}{\partial \boldsymbol{\beta}} \\ \dfrac{\partial l}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sigma^2}\boldsymbol{X'u} \\ -\dfrac{n}{2\sigma^2} + \dfrac{\boldsymbol{u'u}}{2\sigma^4} \end{bmatrix}$$

Para se avaliar o vetor *score* nos valores do estimador com restrições basta substituirmos θ por $\tilde{\theta}$ em $\boldsymbol{s(\theta)}$ e substituirmos **o** $\boldsymbol{u}$ por $\boldsymbol{\tilde{u} = y - X\tilde{\beta}}$ **e** $\boldsymbol{\sigma^2}$ por $\boldsymbol{\tilde{\sigma}^2}$ **,** ou seja,

$$\boldsymbol{s(\tilde{\theta})} = \begin{bmatrix} \frac{1}{\tilde{\sigma}^2}\boldsymbol{X'\tilde{u}} \\ \mathbf{0} \end{bmatrix}.$$

O inverso da matriz de informação, como vimos. Avaliando-a em $\tilde{\theta}$ temos:

$$\boldsymbol{I}^{-1}(\tilde{\boldsymbol{\theta}}) = \begin{bmatrix} \tilde{\sigma}^2(\boldsymbol{X'X})^{-1} & \mathbf{0} \\ \mathbf{0} & \dfrac{2\tilde{\sigma}^4}{n} \end{bmatrix}$$

Substituindo $\boldsymbol{s}$ e $\boldsymbol{I^{1}}$ na estatística LM, temos:

$LM = \begin{bmatrix} \frac{1}{\tilde{\sigma}^2}\boldsymbol{X'\tilde{u}} & \mathbf{0} \end{bmatrix}' \begin{bmatrix} \tilde{\sigma}^2(X'X)^{-1} & 0 \\ 0 & \frac{\sigma^4}{n} \end{bmatrix} \begin{bmatrix} \frac{1}{\tilde{\sigma}^2}X'\tilde{u} \\ 0 \end{bmatrix}$, fazendo o produto dos vetores e matriz, obtemos:

$$\boldsymbol{LM} = \frac{\boldsymbol{n[\tilde{u}'X(X'X)^{-1}X'\tilde{u}}}{\tilde{u}'\tilde{u}}$$

Mas, $\boldsymbol{\tilde{u}'X(X'X)^{-1}X'\tilde{u}} = \boldsymbol{\tilde{u}'X(X'X)^{-1}X'X(X'X)^{-1}X'\tilde{u}}$, pois $\boldsymbol{X\tilde{u}'X(X'X)^{-1}X'X(X'X)^{-1}X'\tilde{u}}$ é uma matriz simétrica e idempotente. Então

$\boldsymbol{\tilde{u}'X(X'X)^{-1}X'X(X'X)^{-1}X'\tilde{u} = \hat{\alpha}'X'X\hat{\alpha}}$**,** que nada mais é do que a soma dos quadrados explicada da regressão de $\tilde{u}$ na matriz com k+1 variáveis $\boldsymbol{X}$. Observe que neste caso, $\tilde{u}'\tilde{u}$ nada mais é do que a soma dos quadrados a ser explicada, no caso, a soma dos quadrados da variável dependente $\tilde{u}$. Desta forma a estatística LM é simplesmente dada por:

$$LM = n\left[\frac{Soma\ dos\ Quadrados\ Explicada(SQE)}{Soma\ dos\ Quadrados\ Total(SQT}\right]$$

$$\boldsymbol{LM = nR^2 \sim \chi^2_{(q)}}$$

PASSOS (para se testar $H_0 = \beta_1 = 0, \dots, \beta_q = 0)$

1º) Faz-se a regressão;

$y_i = \beta_0' + \beta_{q+1}x_{q+1,i} + \cdots + \beta_k x_{k,i} + u_i$, que é a regressão sob H_0, donde se obtém:

$\tilde{u}_i$=$y_i - \tilde{\beta}_0' - \tilde{\beta}_{q+1}x_{q+1,i} - \cdots - \tilde{\beta}_k x_{k,i}$

2º) Taz-se a regressão de $\tilde{u}_i\ em\ x_1, x_2, \dots, x_k$, com as K variáveis explicativas;

3º) Calcula-se

$LM = nR^2$ que tem distribuição $\chi^2_{(q)}$ e testa-se para se rejeitar ou não H_0.

Também é simples mostrar que

$\boldsymbol{LM = \frac{n[\tilde{u}'X(X'X)^{-1}X'\tilde{u}]}{\tilde{u}'\tilde{u}} = n\frac{(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})}{\tilde{u}'\tilde{u}}}$.

Para mostrar como se chega a este resultado temos que mostrar que:

$[\boldsymbol{\tilde{u}'X(X'X)^{-1}X'\tilde{u}}\ ] = (\tilde{u}'\tilde{u} - \hat{u}'\hat{u})$.

Considere:

$\tilde{u}'\tilde{u} - \boldsymbol{\tilde{u}'X(X'X)^{-1}X'}\tilde{u}'$=$\tilde{u}'[I_n - X(X'X)^{-1}X']\tilde{u}$=$\tilde{u}'M\tilde{u}$

, onde

$M = [I_n - X(X'X)^{-1}X']$ é uma matriz simétrica e idempotente, tal que $MM' = MM = M$.

Então temos que mostrar que $\tilde{u}'M\tilde{u} = \hat{u}'\hat{u}$, o que ocorrerá se $M\tilde{u} = \hat{u}$. Por definição

$\tilde{u} = y - X\tilde{\beta}$, onde $\tilde{\beta}$ é o estimador com restrições satisfazendo

$R\tilde{\beta} = r$.

Portanto, $M\tilde{u} = M(y - X\tilde{\beta}) = My - MX\tilde{\beta}$, mas, é fácil verificar que $MX = 0$, logo $M\tilde{u} = My = M(X\hat{\beta} + \hat{u}) = M\hat{u}$=$\hat{u}$, pois

$[I_n - X(X'X)^{-1}X']\hat{u} = [\hat{u} - X(X'X)^{-1}X'\hat{u}] = \hat{u}$, pelas condições de primeira ordem $X'\hat{u} = 0$.

Portanto $[\boldsymbol{\tilde{u}'X(X'X)^{-1}X'\tilde{u}}\ ] = (\tilde{u}'\tilde{u} - \hat{u}'\hat{u})$ e

$$\boldsymbol{LM} = \frac{\boldsymbol{n}(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})}{\tilde{u}'\tilde{u}}$$

5. Relação entre W, LR e LM

Podemos agora demonstrar a famosa desigualdade entre os 3 testes:

$$W \geq LR \geq LM$$

Sabemos que:

$\ln(1 + x) \cong x - \frac{x^2}{2!}$+....[10]

---

10 Expansão por Série de Taylor no ponto x = 0, para aproximar

$f(x) = f(0) + \frac{x}{1!}\frac{\partial f}{x}|_{x=0} + \frac{x^2}{2!}\frac{\partial^2 f}{\partial x^2}|_{x=0} + \cdots$

A 2ª formula para LR derivada acima, nos fornece:
$LR = c$, logo aplicando a expansão acima até o 2º termo temos:
$LR = n\left(\frac{\tilde{u}'\tilde{u} - \hat{u}'\hat{u}}{\hat{u}'\hat{u}}\right) - \frac{n}{2}\left(\frac{\tilde{u}'\tilde{u} - \hat{u}'\hat{u}}{\hat{u}'\hat{u}}\right)^2$.
Mas, como visto acima $n\left(\frac{\tilde{u}'\tilde{u} - \hat{u}'\hat{u}}{\hat{u}'\hat{u}}\right) = W$, conclui-se pois que

$$W \geq LR$$

De forma similar, mas agora usando a 3ª expressão para LR, temos

$$LR = -n\ln\left(1 - \frac{(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})}{\tilde{u}'\tilde{u}}\right)$$

$$LR = -n\left(-\frac{(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})}{\tilde{u}'\tilde{u}}\right) - (-\frac{n}{2})\left(-\frac{(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})}{\tilde{u}'\tilde{u}}\right)^2$$

$LR = n\left(\frac{(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})}{\tilde{u}'\tilde{u}}\right) + \frac{n}{2}\left(\frac{(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})}{\tilde{u}'\tilde{u}}\right)^2$ ou

$LR = LM + \frac{n}{2}\left(\frac{(\tilde{u}'\tilde{u} - \hat{u}'\hat{u})}{\tilde{u}'\tilde{u}}\right)^2$, concluímos então que

$LR \geq LM$, ou seja, comparando os três testes temos:

$W \geq LR \geq LM.$
Os testes são assintoticamente equivalentes, mas para amostras finitas, podem ser diferentes, mas mantendo a ordem de grandeza acima demonstrada.

6. Exemplo

Usarei o exemplo do Wooldridge, onde ele especifica uma relação para explicar o número de vezes que jovens nascidos em 1960, dão detidos pela polícia em 1986. Vou copiar aqui o Cap5.do, colocado no STOA.

```
* Stata Do-file


* setup
version 14.2
capture log close
set more off

* open log
log using Chap5Asyntotics, replace text
```

$f(x) = \ln(1+x) = \ln(1+0) + \frac{x}{1!}(1+0) + \frac{x^2}{2!}[-1(1+0)^{-2}] + \cdots$

$\ln(1+x) = x - \frac{x^2}{2} + \ldots$

```
* open data
 use "/Users/denisardalves/Desktop/EAE-324/DATA SET/CRIME1.DTA", clear
/*Para voces verem que narr86, mesmo com n=2724 está muito longe da dist. normal, vou usar o comando kdensity em narr86*/

kdensity narr86, bwidth(0.2) normal
* Vou selecionar uma amostra, com reposição, das 2725 obs dos dados de CRIME1.DTA. Para isso uso o comando bsample com n = 500. Quero mostrar que as diferenças entre os testes de WALD, LR e LM, diminuem a medida que o n aumenta.*/
bsample 500
* estimar regressão sem restrições
reg narr86 avgsen tottime qemp86 black hispan inc86

estimates store regsr
ereturn list
scalar sqrsr = e(rss)
*estimar regressão com restrições, supondo que raça não afeta criminalidade
reg narr86 avgsen tottime qemp86 black hispan inc86

estimates store regcr
ereturn list
scalar sqrcr = e(rss)
* compute LR
scalar LR=e(N)*(log(sqrcr) - log(sqrsr))
* se usarmos LR2=nln(1+(sqrcr-sqrsr)/sqrsr). temos

scalar LR2=e(N)*(log(1+(sqrcr-sqrsr)/sqrsr))

di LR2
ˆ* computo da estatística WALD
scalar WALD=e(N)*(sqrcr-sqrsr)/sqrsr
di wALD
* computo do LM
*1o Passo:

quietly reg narr86 avgsen tottime qemp86 inc86
predict rescr, residuals
*2o Passo:

reg rescr avgsen tottime qemp86 black hispan inc86
*3o Passo:

 scalar LM=e(N)*e(r2)
 di LM
 di LR
 di WALD
```

```
 scalar chic = invchi2tail(2,.05)
di "Chi-square(2) 95th percentile = " chic
scalar pvalue = chi2tail(2,WALD)
di pvalue
log close

log close
```

OBS: Para vocês observarem como os testes se aproximam cada vez mais, é só apagar o comando bsample e rodar o do file(mude o nome do do file) que você estará usando as 2725 e não 500 com0 selecionadas pelo bsample, n=500.