Análise de Regressão Múltipla – Inferência

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k + u$

Inferência Estatística Clássica

- Inferência é a área que descreve os procedimentos por meio dos quais usamos as observações para *tirar conclusões* sobre a população ou sobre o processo gerador dos dados

Inferência Estatística Clássica

- O pressuposto é que existe um processo gerador dos dados desconhecido
- Esse processo pode ser descrito por uma distribuição de probabilidade
- Esta distribuição de probabilidade pode ser descrita por um conjunto de parâmetros
- D. Normal – desconhecidos $\boldsymbol{\mu}$ e $\boldsymbol{\sigma}^2$.

Inferência Estatística Clássica

- Em geral a inferência estatística pode ser classifica por dois tópicos:
- Inferência Clássica
- Inferência Bayesiana

Inferência Estatística Clássica

- A inferência estatística clássica possui duas premissas:
  1. Os dados amostrais são as únicas informações relevantes.
  2. A construção e avaliação dos diferentes procedimentos para a inferência se baseiam em comportamentos de longo prazo sob circunstâncias essencialmente similares.

Inferência Estatística Bayesiana

- Combinamos informação amostral com informação prévia
- Suponha uma amostra aleatória *n* de uma população normalmente distribuida com $\boldsymbol{\mu}$ e $\boldsymbol{\sigma^2}$ desconhecidas
- Queremos fazer inferência sobre o valor de $\boldsymbol{\mu}$

Inferência Estatística Bayesiana

- Em inferência clássica tomamos a média amostral como o estimador de $\boldsymbol{\mu}$ -- nesse caso sua variância será $\boldsymbol{\sigma^2/n}$
- O inverso dessa variância é conhecido como **precisão amostral** $\boldsymbol{(n/\sigma^2)}$
- Em inferência Bayesiana temos informação prévia sobre o valor de $\boldsymbol{\mu}$

Inferência Estatística Bayesiana

- Esse valor prévio de $\boldsymbol{\mu}$ é expresso em termos de uma distribuição de probabilidade conhecida como *distribuição a priori*
- Suponha uma distribuição a priori, $N(\boldsymbol{\mu_0}, \boldsymbol{\sigma_0^2})$ com precisão $\boldsymbol{1/\sigma^2}$
- Combinando com a informação amostral obtemos a *distribuição a posteriori* de $\boldsymbol{\mu}$

Estimando a média

- A média estimada é uma ponderação com a média amostral e a média *a priori*

$$\mu_{Bayesiana} = \frac{w_1 \bar{y} + w_2 \mu_0}{w_1 + w_2}$$

$$w_1 = n/\sigma^2$$

$$w_2 = 1/\sigma_0^2$$

Média Bayesiana

- Exemplo:
  - Média amostral = 20 e $\boldsymbol{\sigma}^2$ = 4
  - Média a priori = 10 e $\boldsymbol{\sigma_0}^2$ = 2
- Então teremos:
  - Média a posteriori =

$$\frac{\frac{1}{4}(20)+\frac{1}{2}(10)}{\frac{1}{4}+\frac{1}{2}}=\frac{10}{3/4}=13.33$$

Inferência Estatística Clássica

- Estimativa pontual
- Estimativa intervalar
- Teste de hipótese

Estimativa Pontual

- Suponha que a distribuição de probabilidade envolva um parâmetro θ e que tenhamos uma amostra n de ($y_1, \ldots, y_n$) dessa distribuição
- Faça $g(y_1, \ldots, y_n)$ o estimador de θ
  - Estimador: $g(y_1, \ldots, y_n)$
  - Estimativa: θ

Estimativa Pontual

- Interpretação: o estimador é uma V.A. e a estimativa um valor dessa V.A.
- Por exemplo, se usamos a média amostral, 4, para estimar o valor de θ, então essa é a estimava de θ

Teste de Hipótese

- Em teste de hipótese sugerimos uma hipótese sobre o valor estimado de θ
- Por exemplo, $\theta = 4$, e examinamos o grau de evidência na amostra em favor da hipótese, pensando se aceitamos ou rejeitamos esta hipótese

Inferência em Regressão

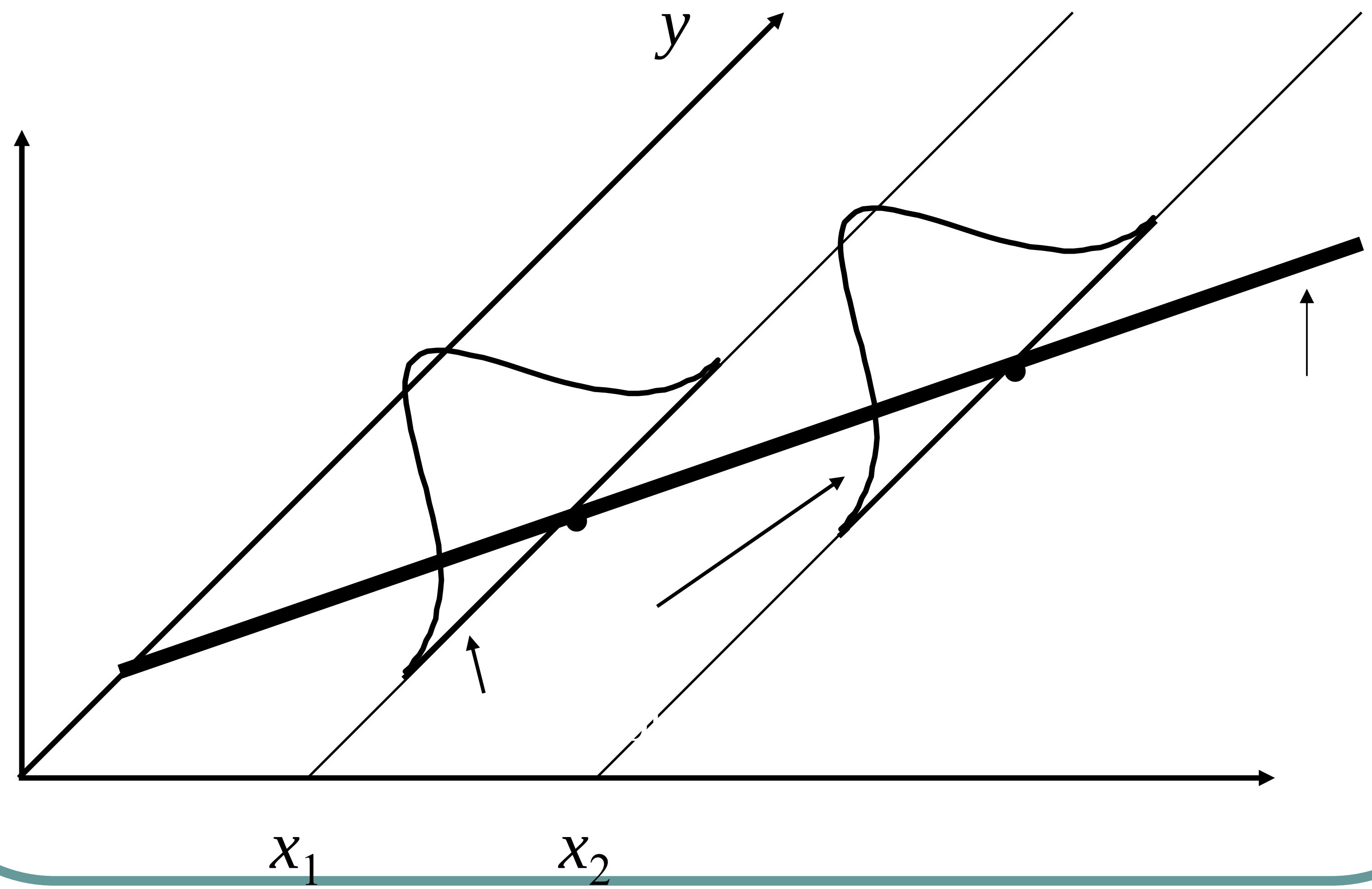
Hipóteses do Modelo Linear Clássico (MLC)

- Até agora,sabiamos que dadas as hipóteses Gauss-Markov, o estimador OLS é BLUE
- Para usar testes de hipóteses, precisamos adicionar outra hipótese
- Assuma que u é independente de x_1, $x_2, \ldots, x_k$ e u é normalmente distribuida com média zero e variância σ^2: $u \sim N(0,\sigma^2)$

Hipóteses MLC

- Sob MLC, OLS não é apenas BLUE, mas é o estimador não-viesado da menor variância
- Podemos resumir as hipóteses da população do MLC como segue
- $y|\boldsymbol{x} \sim \mathrm{N}(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k, \sigma^2)$
- Por enquanto assumimos apenas normalidade
- Grandes amostras nos permitem excluir essa hipótese de normalidade

A distribuição normal homocedastica com uma única variável explicativa



Distribuições de Amostragem Normais

Sob as hipóteses MLC, condicional sobre os valores amostrais das variáveis independentes

$\hat{\beta}_j \sim \text{Normal}\left[\beta_j, Var\left(\hat{\beta}_j\right)\right]$, então

$$\left(\hat{\beta}_j - \beta_j\right) \Big/ SD\left(\hat{\beta}_j\right) \sim \text{N}\left(0,1\right)$$

$\hat{\beta}_{\text{j}}$ é distribuida normalmente porque é uma combinação linear dos erros

O Teste *t*

Sob as hipóteses do MLC

$$\left(\hat{\beta}_j - \beta_j\right) \Big/ SE\left(\hat{\beta}_j\right) \sim t_{n-k-1}$$

Esta é uma distribuição *t*

porque temos que estimar σ^2 por $\hat{\sigma}^2$

Note que os graus de liberdade são: $n-k-1$

O Teste *t*

- Conhecendo a distribuição amostral para o estimador padronizado nos permite conduzir testes de hipóteses
- Começe com uma hipótese nula
- Por exemplo, H_0: $\beta_j = 0$
- Se aceitarmos a hipótese nula, então aceitamos que x_j não possui efeito sobre y, controlando para os outros x's

O Teste *t*

Para realizar o teste primeiro precisamos formar a estatística t para $\hat{\beta}_j$: $t_{\hat{\beta}_j} \equiv \hat{\beta}_j \big/ SE\left(\hat{\beta}_j\right)$

Nós iremos usar a estatística t em conjunto com uma regra de rejeição para determinar quando aceitar a hipótese nula, H_0

Teste *t* : Alternativas

- Além da hipótese nula, H_0, precisamos de uma hipótese alternativa, H_1, e um nível de significância
- H_1 pode ser uni-caudal ou bi-caudal
- H_1: $\beta_j > 0$ ou H_1: $\beta_j < 0$ são uni-caudais
- H_1: $\beta_j \neq 0$ é uma alternativa bi-caudal
- Se desejamos ter apenas 5% de probabilidade de rejeitar H_0 se ele é realmente verdade, então dizemos que o nível de significância é de 5%

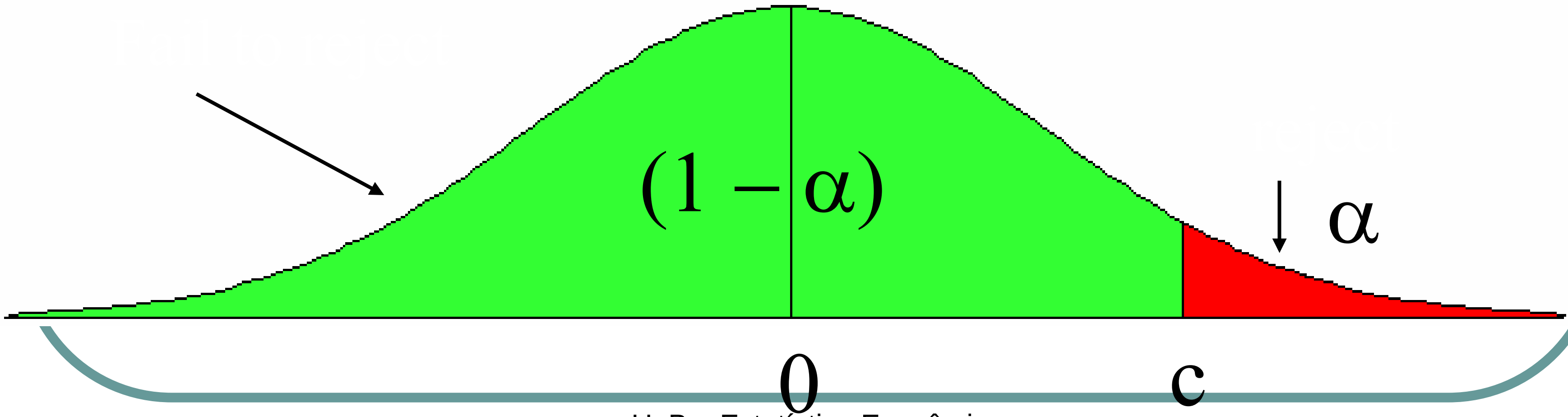
Alternativa Uni-caudal

- Tendo selecionado um nível de significância, α, procuramos o percentil $(1 - \alpha)$ em uma distribuição t com $n - k - 1$ graus de liberdade e df – chamamos isto de valor crítico
- Nós podemos rejeitar a hipótese nula se a estatística t é maior do que o valor crítico
- Se a estatística t é menor do que o valor crítico então falhamos em rejeitar a hipótese nula

Alternativa Uni-caudal

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i$$

$H_0: \beta_j = 0$ $\qquad H_1: \beta_j > 0$



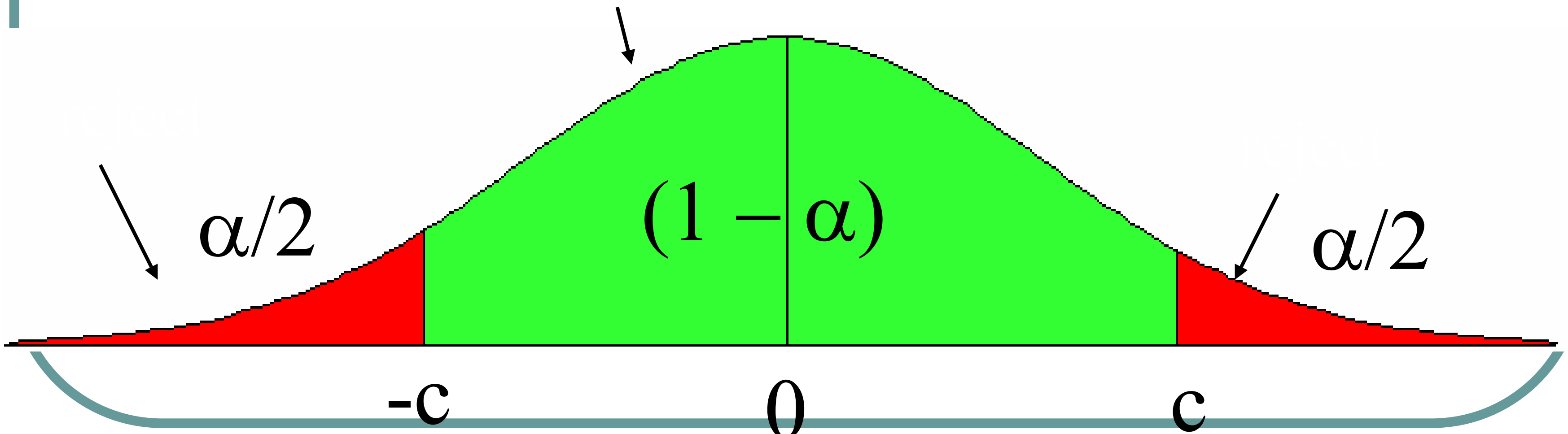
Uni-caudal *vs* bi-caudal

- Como a distribuição *t* é simétrica, testando H_1: $\beta_j < 0$ é direto. O valor crítico é apenas o valor negativo anterior
- Nós podemos rejeitar a nula se a estatística *t* $< -c$, e se a estatística *t* $>$ do que $-c$ então falhamos em rejeitar a hipótese nula
- Para um teste bi-caudal, nós determinamos o valor crítico baseado em $\alpha/2$ e rejeitamos H_1: $\beta_j \neq 0$ se o valor absoluto da estatística $t > c$

Alternativas Bi-caudais

$$y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_k X_{ik} + u_i$$

$H_0: \beta_j = 0$ $\qquad$ $H_1: \beta_j \neq 0$



Sumário para H_0: $\beta_j = 0$

- A não ser que seja explicitado, assumimos que a alternativa seja bi-caudal
- Se nós rejeitarmos a nula, tipicamente dizemos "x_j é estatisticamente significante ao nível de α %"
- Se nós falhamos em rejeitar a hipótese nula, dizemos tipicamente que "x_j é estatisticamente insignificante ao nível de α %"

Testando outras hipóteses

- Uma forma mais geral de testar a estatística *t* reconhece que desejamos testar algo como H_0: $\beta_j = a_j$
- Nesse caso, a estatística *t* apropriada é

$$t = {\left(\hat{\beta}_j - a_j\right)} \Big/ {SE\left(\hat{\beta}_j\right)}, \text{ onde}$$

$$a_j = 0 \text{ para o teste padrão}$$

Intervalos de Confiança

- Outra forma de usar testes estatísticos é construir um intervalo de confiança usando o mesmo valor crítico como foi usado para o teste bi-caudal
- Um intervalo de confiaça de (1 - α) % é definido como:

$$\hat{\beta}_j \pm c \bullet SE\left(\hat{\beta}_j\right), \text{ tal que c é o percential } \left(1\text{-}\frac{\alpha}{2}\right)$$

em uma distribuição t_{n-k-1}

Computando valores-*p* para testes *t*

- Uma alternativa à versão clássica do teste é perguntar: “qual é o menor nível de significância ao qual a hipótese nula seria rejeitada?”
- Então, cacule a estatística *t*, e então procure qual o percentil é apropriado à distribuição *t* – este é o valor-*p*
- O valor-*p* é a probabilidade da estatística *t* aceitar a hipótese nula

Testes no computador

- A maioria dos programas computa o valor-p, assumindo um teste bi-caudal
- Se você realmente deseja uma alternativa unicaudal, então deve dividir o valor-p por 2
- Stata fornece a estatística *t*, *p*-value, e intervalo de confiança de confidence 95% para H_0: $\beta_j = 0$, nas colunas "**t**", "**P > |t|**" e "**[95% Conf. Interval]**"

Preços Hedônicos

- Modelo de preços em termos de suas características
- Decomposição do preço na disposição a pagar por cada característica
- Mercado imobiliário residencial
  - Atributos do imóvel
  - Atributos da vizinhança (peer effect)
  - Distância do trabalho

Testando uma combinação linear

- Suponha que ao invés de testar se β_1 é igual a uma constante , você deseja testar se ele é igual a outro parâmetro, isto é $H_0 : \beta_1 = \beta_2$
- Use o mesmo procedimento para formar uma estatística t

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{SE\left(\hat{\beta}_1 - \hat{\beta}_2\right)}$$

Testando uma combinação linear

Uma vez que

$$SE\left(\hat{\beta}_1 - \hat{\beta}_2\right) = \sqrt{Var\left(\hat{\beta}_1 - \hat{\beta}_2\right)}, \text{ então}$$

$$Var\left(\hat{\beta}_1 - \hat{\beta}_2\right) = Var\left(\hat{\beta}_1\right) + Var\left(\hat{\beta}_2\right) - 2Cov\left(\hat{\beta}_1, \hat{\beta}_2\right)$$

$$SE\left(\hat{\beta}_1 - \hat{\beta}_2\right) = \left\{\left[SE\left(\hat{\beta}_1\right)\right]^2 + \left[SE\left(\hat{\beta}_2\right)\right]^2 - 2s_{12}\right\}^{1/2}$$

onde s_{12} é uma estimativa da $Cov\left(\hat{\beta}_1, \hat{\beta}_2\right)$

Testando uma combinação linear

- Então, para usar a fórmula precisamos de s_{12}, o que não está na tela dos programas
- Mas vários programas possuem uma opção para fazer isto
- No Stata, após o comando de regressão **reg y x1 x2 … xk** você deve escrever **test x1 = x2** para ter um p-valor para este teste
- Genericamente falando, você sempre pode re-escrever o problema para o teste que deseja fazer

Exemplo:

- Suponha que você esteja interessado no efeito de gastos de uma campanha de gastos sobre votos
- Modelo é $voteA = \beta_0 + \beta_1\log(expendA) + \beta_2\log(expendB) + \beta_3 prtystrA + u$
- H_0: $\beta_1 = - \beta_2$, or H_0: $\theta_1 = \beta_1 + \beta_2 = 0$
- $\beta_1 = \theta_1 - \beta_2$, então subst. e re-arrange $\Rightarrow$ $voteA = \beta_0 + \theta_1\log(expendA) + \beta_2\log(expendB - expendA) + \beta_3 prtystrA + u$

Exemplo:

- Este é o mesmo modelo original, mas agora temos o desvio padrão para $\beta_1 + \beta_2 = \theta_1$ diretamente a partir da regressão básica
- Qualquer combinação linear de parâmetros poderia ser testado de uma forma similar
- Outros exemplos de hipóteses sobre combinação linear de parâmetros:
  - $\beta_1 = 1 + \beta_2$; $\beta_1 = 5\beta_2$; $\beta_1 = -1/2\beta_2$; etc

Múltiplas Restrições Lineares

- Tudo que fizemos até o momento envolvia testar uma única restrição linear, (e.g. $\beta_1 = 0$ ou $\beta_1 = \beta_2$)
- Entretanto, podemos desejar testar conjuntamente testes de hipóteses múltiplas sobre nossos parâmetros
- Um exemplo típico de teste são as “restrições de exclusão” – desejamos saber se um grupo de parâmetros é igual a zero

Testando Restrições de Exclusão

- Agora a hipótese nula poderia ser algo como H_0: $\beta_{k-q+1} = 0, \ldots, \beta_k = 0$
- A alternativa é apenas H_1: H_0 não é verdadeira
- Não podemos apenas checar cada estatística *t* separadamente, porque nós desejamos saber se os *q* parâmetros são conjuntamente significantes a um dado nível – é possível que individualmente nenhum termo seja significante

Restrições de Exclusão

- Para testar precisamos estimar o **modelo restrito** sem a inclusão de x_{k-q+1}, …, x_k, bem como o **modelo não-restrito** com todos os x's incluídos
- Intuitivamente, nós desejamos saber se a mudança em SSR é grande o suficiente para garantir a inclusão de x_{k-q+1}, …, x_k

$$F \equiv \frac{\left(SSR_r - SSR_{ur}\right)/q}{SSR_{ur}/\left(n-k-1\right)}, \text{ tal que}$$

$$r \text{ é restrito e } ur \text{ é o não restrito}$$

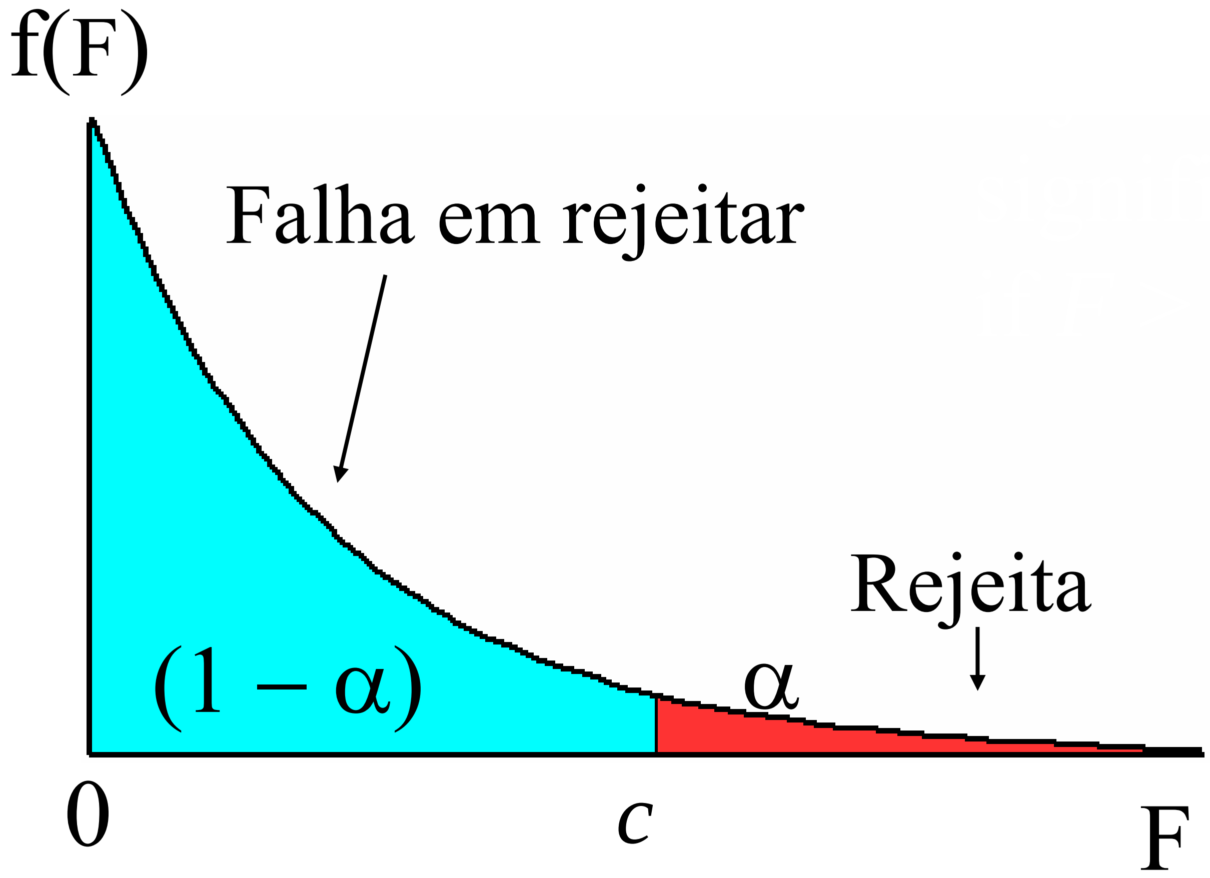
A Estatística *F*

- A estatística *F* é sempre positiva, uma vez que a SSR de um modelo restrito não pode ser menor do que a SSR do modelo não-restrito
- Essencialmente, a estatística *F* está medindo o aumento relativo na SSR quando movendo do modelo não restrito para o modelo restrito
- q = # de restrições, ou $df_r - df_{ur}$
- $n - k - 1 = df_{ur}$

A Estatística F

- Para decidir se o incremento na SSR é grande o suficiente para rejeitar as exclusões quando mudamos para um modelo restrito, precisamos conhecer um pouco da distribuição amostral da nossa estatística F
- Não supreendentemente, $F \sim F_{q,n-k-1}$, tal que q é referida como o numerador dos graus de liberdade e $n - k - 1$ como o denominador dos graus de liberdade

A Estatística *F*



O R^2 forma a Estatística F

- Como as SSR podem ser grandes e estranhas, uma fórmula alternativa pode ser útil
- Nós usamos o fato que SSR = SST(1 – R^2) para qualquer regressão, então podemos substituir para SSR_u e SSR_{ur}

$$F \equiv \frac{\left(R^2_{ur} - R^2_{r}\right)/q}{\left(1 - R^2_{ur}\right)/\left(n-k-1\right)}$$

Significância Total

- Um caso especial de exclusão de restrições é testar H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$
- Dado que o R^2 de um modelo com apenas um intercepto será zero, a estatística *F* é simplesmente

$$F = \frac{R^2 / k}{\left(1 - R^2\right) / \left(n - k - 1\right)}$$

Restrições Lineares Gerais

- A forma básica da estatística *F* irá funcionar para qualquer conjunto de restrições lineares
- Primeiro estime o modelo não-restrito e depois o restrito
- Em cada caso, guarde os valores da SSR
- Impor restrições pode ser difícil – isso pode requerer que as variáveis tenham que ser redefinidas novamente

Exemplo:

- Vamos usar o mesmo modelo de votos
- O modelo é $voteA = \beta_0 + \beta_1 \log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystrA + u$
- agora a hipótese nula H_0: $\beta_1 = 1$, $\beta_3 = 0$
- Substitutindo as restrições: $voteA = \beta_0 + \log(expendA) + \beta_2 \log(expendB) + u$, então
- Use $voteA - \log(expendA) = \beta_0 + \beta_2 \log(expendB) + u$ como o modelo restrito

Sumário Estatística *F*

- Apenas como a estatística t, valores-p podem ser calculados procurando pelo percentil apropriado na distribuição F
- Stata realiza isso pelos comandos: **display fprob(*q*, *n* – *k* – 1, *F*)**, onde os valores apropriados de F, q, e $n - k - 1$ devem ser usados
- Se apenas uma exclusão está sendo testada, então $F = t^2$, e os valores p serão os mesmos