

2.5.2 Estimadores dos parâmetros

Suponha que as variáveis Y e ϵ estejam relacionadas por $Y = \mu + \epsilon$ e que $\mu \in L_1$, onde L_1 é um modelo linear de dimensão k_1 . Seja L_1^\perp o complemento ortogonal de L_1 tal que $\mathbb{R}^n = L_1 \oplus L_1^\perp$. Usando o teorema 3, segue que

$p_{L_1}(\epsilon)$ e $p_{L_1^\perp}(\epsilon)$ são independentes.

$$\begin{aligned} p_{L_1^\perp}(Y) &= Y - p_{L_1}(Y) = \mu + \epsilon - p_{L_1}(\mu + \epsilon) = \\ &= \mu + \epsilon - p_{L_1}(\mu) - p_{L_1}(\epsilon) = \epsilon - p_{L_1}(\epsilon) \text{ pois } \\ p_{L_1}(\mu) &= \mu \text{ já que } \mu \in L_1. \end{aligned}$$

Além disso, $p_{L_1}(Y) = p_{L_1}(\mu + \epsilon) = p_{L_1}(\mu) + p_{L_1}(\epsilon) = \mu + p_{L_1}(\epsilon)$.

Como $p_{L_1}(\epsilon)$ e $p_{L_1^\perp}(\epsilon) = \epsilon - p_{L_1}(\epsilon)$ são independentes segue que $p_{L_1}(Y)$ e $p_{L_1^\perp}(Y)$ são independentes, e em particular, $\hat{\mu} = p_{L_1}(Y)$ e $\tilde{\sigma}^2 = \frac{1}{n-k} \|Y - p_{L_1}(Y)\|^2$ são independentes.

Além disso,

$$\|Y - p_{L_1}(Y)\|^2 = \|\epsilon - p_{L_1}(\epsilon)\|^2 \sim \sigma^2 \chi_{(n-k_1)}^2.$$

Logo, $\tilde{\sigma}^2 \sim \sigma^2 \frac{\chi_{(n-k_1)}^2}{n-k_1}$ e, em particular,

$$E\tilde{\sigma}^2 = \sigma^2 \text{ e } Var(\tilde{\sigma}^2) = \frac{2\sigma^4}{n-k_1}.///$$

2.5.3 O teste F

Seja agora $L_2 \subseteq L_1$ uma segunda hipótese linear e suponha que L_2 seja verdadeiro tal que $\mu \in L_2$. Sejam p_1 e p_2 as projeções sobre L_1 e L_2 e $\hat{\mu}_i = p_i(y)$, $i = 1, 2$.

A estatística F para L_2 sob L_1 é

$$F(y) = \frac{\|p_1(y) - p_2(y)\|^2 / (k_1 - k_2)}{\|y - p_1(y)\|^2 / (n - k_1)}$$

$$\models F(Y) \sim F_{(k_1 - k_2, n - k_1)}, \quad \text{se } \mu \in L_2$$

O teste de L_2 sobre L_1 ao nível de significância α é, então realizado rejeitando-se $H_0 : \mu \in L_2$ se $F(y) > F_{1-\alpha}(k_1-k_2, n-k_1)$.

Para a demonstração deste resultado seja

$L_1 \ominus L_2$ o complemento ortogonal de L_2 em L_1 e escreva

$$\mathbb{R}^n = L_2 \oplus (L_1 \ominus L_2) \oplus L_1^\perp$$

Os subespaços acima têm dimensões k_2 , k_1-k_2 e $n-k_1$, respectivamente. As projeções sobre os três são p_2 , $p_1 - p_2$ e $y - p_1(y)$.

Escrevendo $Y = \mu + \epsilon$, $\mu \in L_2$ temos

$$\begin{aligned} p_1(Y) - p_2(Y) &= p_1(\mu + \epsilon) - p_2(\mu + \epsilon) = p_1(\mu) + \\ & p_1(\epsilon) - p_2(\mu) - p_2(\epsilon) = \mu + p_1(\epsilon) - \mu - p_2(\epsilon) \\ &= p_1(\epsilon) - p_2(\epsilon) \end{aligned}$$

Segue, do teorema 3, que $p_2(\epsilon)$, $p_1(\epsilon) - p_2(\epsilon)$ e $\epsilon - p_1(\epsilon)$ são independentes.

Além disso, $\|p_1(Y) - p_2(Y)\|^2 \sim \sigma^2 \chi_{(k_1 - k_2)}^2$ e

$\|Y - p_1(Y)\|^2 \sim \sigma^2 \chi_{(n - k_1)}^2$ e, como são independentes, $F(Y) \sim F_{(k_1 - k_2, n - k_1)}$.

Observe que como $\hat{\mu}_2$ é independente de

$(p_1(Y) - p_2(Y), Y - p_1(Y))$, segue que

$\hat{\mu}_2$ e $F(Y)$ são independentes.

2.6 A relação entre os testes t e F

Considere a hipótese linear $L_1 = \text{span}\{x_1, \dots, x_k\}$ onde x_1, \dots, x_k forma uma base, não necessariamente ortogonal, para L_1 . Neste caso, para $\mu \in L_1$, podemos escrever

$$\mu = \sum_{j=1}^k x_j \beta_j$$

e, buscamos estimadores para os β_j 's, os testes, etc.

Consideremos o teste da hipótese

$H_2 : \beta_k = 0$ que é uma hipótese linear dada por $L_2 = \text{span}\{x_1, \dots, x_{k-1}\}$. Esta hipótese reduz o número de parâmetros por 1, comparada com L_1 e, mostraremos que neste caso, o teste F é equivalente ao teste t .

Sejam p_1 e p_2 as projeções sobre L_1 e L_2 e considere $z = x_k - p_2(x_k)$.

Pelas propriedades de p_2 ,

$$z = x_k - p_2(x_k) \perp v, \forall v \in L_2.$$

Podemos escrever

$$\mu = \sum_j x_j \beta_j = \sum_{j=1}^{k-1} x_j \beta_j + (z + p_2(x_k)) \beta_k = \sum_{j=1}^{k-1} x_j \tilde{\beta}_j + z \beta_k,$$

para $\tilde{\beta}_j$ adequado, onde aqui usamos o fato de que $p_2(x_k) \in L_2$ tal que $p_2(x_k)$ é uma combinação linear de x_1, \dots, x_{k-1} .

Agora, como $z \perp L_2$ encontramos que $\hat{\mu}_1 = p_1(y) = p_2(y) + \frac{z \cdot y}{\|z\|^2} z \Rightarrow$

$$\hat{\beta}_k = \frac{z \cdot y}{\|z\|^2}, \hat{\mu}_2 = p_2(y) \text{ e}$$

$$\tilde{\sigma}_1^2 = \frac{1}{n-k} \|y - \hat{\mu}_1\|^2.$$

Para calcular a estatística F observe que

$$||\hat{\mu}_1 - \hat{\mu}_2||^2 = \hat{\beta}_k^2 ||z||^2.$$

Assim, a estatística F para L_2 sob L_1 é

$$F(y) = \frac{\hat{\beta}_k^2 ||z||^2}{\tilde{\sigma}_1^2} \sim F_{(1, n-k)}, \text{ sob } L_2.$$

Derivemos agora o teste t para $\beta_k = 0$. Para isto, observe que $\hat{\beta}_k$ é uma combinação linear de Y_1, \dots, Y_n e, assim, é normalmente distribuído.

$$E[z.Y] = \sum_{i=1}^n z_i E[Y_i] = z.\mu =$$

$$= z. \left(\sum_{j=1}^{k-1} x_j \tilde{\beta}_j + z \beta_k \right) = (z.z) \beta_k = ||z||^2 \beta_k$$

Assim, $E[\hat{\beta}_k] = \beta_k$ tal que $\hat{\beta}_k$ é um estimador não tendencioso de β_k . Similarmente,

$$Var(z.Y) = \sigma^2 ||z||^2 \quad \Rightarrow \quad Var(\hat{\beta}_k) = \frac{\sigma^2}{||z||^2}.$$

Logo, $\hat{\beta}_k \sim N(\beta_k, \sigma^2 / \|z\|^2)$.

O erro padrão de $\hat{\beta}_k$ é $s.e.(\hat{\beta}_k) = \tilde{\sigma}_1 / \|z\|$ e,

a estatística de teste para $\beta_k = 0$ é

$$t(y) = \frac{\hat{\beta}_k}{s.e.(\hat{\beta}_k)}.$$

Logo, vale a relação $F(y) = t^2(y)$.

Segue que um teste t bilateral que rejeita H_0 para $|t(y)| > c_1$ é equivalente ao teste F que rejeita H_0 para $F(y) > c_2$, onde $c_2 = c_1^2$.

Vantagem do teste t : contém a informação sobre o sinal do desvio da hipótese nula. Em particular, podemos definir um teste t unilateral para a hipótese $\beta_k = 0$ contra $\beta_k > 0$ que rejeita H_0 para $t(y) > t_{1-\alpha(n-k)}$ ou contra a hipótese $\beta_k < 0$ que rejeita H_0 para $t(y) < -t_{1-\alpha(n-k)}$. Porém, o teste t é restrito ao caso onde $\dim L_1 - \dim L_2 = 1$.

$$\models t(Y) \sim t_{(n-k)}$$

Sob $L_2(\beta_k = 0)$, temos que $\hat{\beta}_k/(\sigma/\|z\|) \sim N(0, 1)$. Vimos que $\tilde{\sigma}_1/\sigma \sim \sqrt{\chi_{n-k}^2/(n-k)}$. Como $\hat{\beta}_k$ é uma função de $p_1(y) - p_2(y)$ e $\tilde{\sigma}_1$ é uma função de $y - p_1(y)$, segue que o numerador e o denominador da estatística t são independentes. Logo, pela definição da distribuição t tem-se que $t(Y) \sim t_{(n-k)}$, sob $\beta_k = 0$.

2.7 Hipótese Afins e Regiões de Confiança

2.7.1 Elipsóides de Confiança

Considere a hipótese $L = \text{span}\{X\}$ onde X é uma matriz $n \times k$ de posto completo. Sob L , temos $\mu = X\beta$. Derivemos então uma região de confiança para β .

Consider a hipótese $H_2 : \beta = \beta_0$. Observe que este não é um modelo linear a menos que $\beta_0 = 0$.

$Y = \mu + \epsilon$ e, assim, $Y - X\beta_0 = \mu - X\beta_0 + \epsilon = X(\beta - \beta_0) + \epsilon$

Assim, o vetor $Y' = Y - X\beta_0$ segue um modelo linear L com parâmetro $\beta' = \beta - \beta_0$. Defina $L_2 = \{0\}$, o subespaço nulo que corresponde à $\beta' = 0$, uma hipótese equivalente à H_2 .

Assim, podemos testar L_2 sob $L(= L_1)$ por um teste F . Temos $\hat{\mu}_2 = 0 \Rightarrow \|\hat{\mu}_1 - \hat{\mu}_2\|^2 = \|X\hat{\beta}'\|^2$.

Como $\beta' = \beta - \beta_0$ temos $\hat{\beta}' = \hat{\beta} - \beta_0$ tal que $\|\hat{\mu}_1 - \hat{\mu}_2\|^2 = \|X(\hat{\beta} - \beta_0)\|^2 =$
 $= (\hat{\beta} - \beta_0)^T X^T X (\hat{\beta} - \beta_0)$.

Precisamos agora achar $\tilde{\sigma}^2$ sob L_2 . Porém,

$$\|y' - X\beta'\|^2 = \|y - X\hat{\beta}\|^2$$

tal que $\tilde{\sigma}^2$ para Y' é o mesmo que $\tilde{\sigma}^2$ para Y . Assim, a estatística F para L_2 é

$$F(Y) = \frac{\|X(\hat{\beta} - \beta_0)\|^2/k}{\tilde{\sigma}^2} = \frac{(\hat{\beta} - \beta_0)^T X^T X (\hat{\beta} - \beta_0)/k}{\tilde{\sigma}^2}$$

com distribuição $F_{(k, n-k)}$ para $\beta = \beta_0$. A estatística F acima pode ser usada para obtermos uma região de confiança para β .

Seja $\alpha \in (0, 1)$. Então o teste é rejeitado para $F(y) > F_{1-\alpha}(k, n-k)$. Assim, o conjunto de valores de β que são aceitos pelo teste F é dado por

$$\left\{ \beta \in \mathbb{R}^n : \frac{(\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta) / k}{\tilde{\sigma}^2} \leq F_{1-\alpha}(k, n-k) \right\}$$

Esta região define um elipsóide centrado em $\hat{\beta}$, especificando uma região de confiança para β com coeficiente de confiança de $1 - \alpha$.

Observe que se as colunas de X forem ortogonais $X^T X$ será uma matriz diagonal e, os eixos de confiança serão paralelos aos eixos coordenados pois a região de confiança tomará a forma

$$\left\{ \beta \in \mathbb{R}^n : \frac{\sum_{j=1}^k (\hat{\beta}_j - \beta_j)^2 \|x_j\|^2 / k}{\tilde{\sigma}^2} \leq F_{1-\alpha}(k, n-k) \right\}.$$

2.7.2 Intervalos de Confiança

A região de confiança definida na seção anterior é difícil de ser visualizada quando há mais de dois parâmetros. Frequentemente, calculamos IC's separados para cada parâmetro do modelo.

Considere novamente o modelo linear $L_1 = \text{span}\{x_1, \dots, x_k\}$ e a hipótese $H_0 : \beta_k = \beta_0$, onde agora $\beta_0 \in \mathbb{R}$ é alguma constante dada. Por argumentos similares aos usados para o teste F obtemos a estatística de teste $t(Y) = \frac{\hat{\beta}_k - \beta_0}{s.e.(\hat{\beta}_k)} \sim t_{(n-k)}$ sob H_0 ,

que pode ser usada para testar H_0 . O teste acima pode ser rearrumado para fornecer uma região de confiança para β_k dada por

$$IC(\beta_k, 1 - \alpha) : \hat{\beta}_k \pm t_{1-\frac{\alpha}{2}(n-k)} s.e.(\hat{\beta}_k)$$

Este é o IC padrão para um parâmetro univariado e pode ser obtido a partir do estimador e seu erro padrão.

Encontramos a expressão para $s.e.(\hat{\beta}_k)$, agora encontraremos as expressões para $s.e.(\hat{\beta}_i)$.

Vimos que $\hat{\beta} \sim N_k(\beta, \sigma^2 C)$, onde $C = (X^T X)^{-1}$. Assim, a variância de $\hat{\beta}_j$ é, expressa em termos dos elementos c_{ij} de C , $Var(\hat{\beta}_j) = \sigma^2 c_{jj}$, tal que $s.e.(\hat{\beta}_j) = \tilde{\sigma} c_{jj}^{1/2}$.

Assim, comparando com os resultados obtidos para $\hat{\beta}_k$ concluimos que $c_{kk} = ||z||^{-2}$. Também podemos verificar que $Cov(\hat{\beta}_j, \hat{\beta}_l) = \sigma^2 c_{jl}$ tal que $corr(\hat{\beta}_j, \hat{\beta}_l) = \frac{c_{jl}}{\sqrt{c_{jj}c_{ll}}}$.

2.7.3 Hipóteses Afim

A técnica usada anteriormente para testar hipóteses que são lineares após a subtração de um vetor constante pode ser facilmente estendida.

Considere um modelo linear L_1 e seja $x_0 \in L_1$ um vetor fixado.

Seja L_2 um subespaço linear de L_1 e considere a hipótese

$H_2 : \mu \in x_0 + L_2$ tal que H_2 seja uma sub-hipótese de L_1 .

Como o espaço $x_0 + L_2$ é um subespaço afim de L_1 , a hipótese é chamada hipótese afim.

Para testar H_2 consideramos o novo vetor de observações $Y' = Y - x_0$. Em termos de Y' , a hipótese L_1 permanece inalterada e a hipótese H_2 corresponde à $\mu_2 \in L_2$, onde $\mu' = E[Y']$.

O estimador para σ^2 sob L_1 baseado em Y' é $\tilde{\sigma}^2 = \frac{1}{n-k_1} \|Y' - \hat{\mu}_1\|^2$, com $k_1 = \dim L_1$, onde $\hat{\mu}_1 = p_1(Y') = p_1(Y) - x_0$, p_1 e p_2 sendo as projeções sobre L_1 e L_2 .

Assim, $\tilde{\sigma}^2 = \frac{1}{n-k_1} \|Y - x_0 - p_1(Y) + x_0\|^2 = \frac{1}{n-k_1} \|Y - p_1(Y)\|^2$, que é o mesmo estimador de σ^2 para Y sobre L_1 .

Para o numerador da estatística F , observe que o estimador de μ' sob H_2 é $p_2(Y - x_0) = p_2(Y) - p_2(x_0)$. Assim, a estatística F para H_2 é

$$F(Y) = \frac{\|p_1(Y) - p_2(Y) - (x_0 - p_2(x_0))\|^2 / (k_1 - k_2)}{\tilde{\sigma}^2}$$

onde $k_2 = \dim L_2$. Sob H_2 , $F(Y) \sim F_{(k_1 - k_2, n - k_1)}$.

2.8 Suficiência

Considere a função de verossimilhança dada por

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \|y - \mu\|^2 \right\}$$

Considere a questão de suficiência para os pares (β, σ^2) sob $L_1 = \text{span}\{X\}$ onde $\mu = X\beta$ com X de posto completo k . Temos, para $\mu = X\beta$,

$$\|y - \mu\|^2 =$$

$$(y - X\beta)^T (y - X\beta) = y^T y - 2y^T X\beta + \beta^T X^T X\beta.$$

Assim, podemos escrever,

$$L(X\beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{y^T y}{2\sigma^2} + \frac{y^T X\beta}{\sigma^2} - \frac{\beta^T X^T X\beta}{2\sigma^2} \right\}$$

A f.v. depende dos dados Y somente através da estatística $(Y^T Y, Y^T X)$, que é, portanto, suficiente. Os coeficientes são $(-\frac{1}{2\sigma^2}, \frac{\beta}{\sigma^2})$ cujo domínio é $(-\infty, 0) \times \mathbb{R}^k$ contém um conjunto aberto. Assim, a estatística é completa e portanto, suficiente minimal para (β, σ^2) .

Vimos que $\hat{\beta}$ é uma função 1 a 1 de $X^T Y$ pois a matriz $X^T X$ é não-singular. Além disso, temos que

$$\tilde{\sigma}^2 = \frac{1}{n-k} \{y^T y - \hat{\beta}^T X^T X \hat{\beta}\}$$

Assim, para $\hat{\beta}$ fixado, $\tilde{\sigma}^2$ é uma função 1 a 1 de $(Y^T Y, Y^T X)$ e portanto, $(\hat{\beta}, \tilde{\sigma}^2)$ também é uma estatística suficiente minimal.

(Capítulo 1): o fato de estimarmos μ e σ^2 juntos, tem o efeito de produzir um viés negativo no estimador de σ^2 pois $\hat{\mu}$ é o valor que minimiza $\|y - \mu\|$, empurrando o EMV $\hat{\sigma}^2 = \frac{1}{n}\|y - \mu\|^2$ para zero.

O estimador modificado $\tilde{\sigma}^2 = \frac{1}{n-k}\|y - \hat{\mu}\|^2$ é maior que $\hat{\sigma}^2$, compensando o viés negativo e fazendo $\tilde{\sigma}^2$ um estimador não-tendencioso de σ^2 .

Além disso, $\tilde{\sigma}^2$ é uma função da estatística suficiente completa $(Y^T Y, Y^T X)$ e, portanto, pelo teorema de Lehmann-Scheffé este é o estimador não tendencioso de variância mínima.

O viés negativo de $\hat{\sigma}^2$ pode ser compensado de muitas outras formas. A pergunta que se coloca aqui é por que escolhemos $\tilde{\sigma}^2$ e não um estimador que satisfaça por exemplo $E[\tilde{\sigma}_1] = \sigma$ com variância mínima?

Para justificar o estimador $\tilde{\sigma}^2$, introduziremos um conceito estendido de suficiência chamado L -suficiência.

O perfil da log. verossimilhança para σ^2 é

$$\tilde{L}(\sigma^2) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \|y - \hat{\mu}\|^2 \right\}$$

Por definição, uma estatística é dita L -suficiente se ela pode ser obtida aplicando-se o critério da fatoração de Neyman-Fisher ao perfil da log. verossimilhança (tratando \tilde{L} como se fosse a verdadeira verossimilhança).

No presente caso, o perfil da verossimilhança depende de Y somente através da função desvio minimizada $D(\hat{\mu}) = \|y - \hat{\mu}\|^2$ que é L -suficiente para σ^2 .

O princípio da suficiência, que também se estende para a L -suficiência, diz-nos que devemos usar a distribuição marginal de $D(\hat{\mu})$, a estatística L -suficiente, para inferências sobre σ^2 .

Em particular, devemos estimar σ^2 usando a verossimilhança marginal correspondente à $D(\hat{\mu})$ que fornece um estimador de MV marginal para σ^2 .

$$||y - \hat{\mu}||^2 \sim \sigma^2 \chi_{(n-k)}^2$$

Para derivar o EMV marginal para σ^2 considere a distribuição gama $Ga(\alpha, \lambda)$ com fdp

$$p(y; \alpha, \lambda) = \frac{(\lambda/\alpha)^\lambda}{\Gamma(\lambda)} y^{\lambda-1} e^{-\lambda y/\alpha}$$

α denota a média e λ é o parâmetro de forma.

Lembre da propriedade de mudança de escala:
 $cGa(\alpha, \lambda) = Ga(c\alpha, \lambda)$, $c > 0$. Como $\chi^2_{(f)} = Ga(f, f/2)$ segue que

$$\|y - \hat{\mu}\|^2 \sim Ga(\sigma^2(n - k), (n - k)/2).$$

Para obter o EMV de para α baseado na distribuição Gama, devemos maximizar a função $f(\alpha) = \alpha^{-\lambda} e^{-\lambda y/\alpha}$. O máximo é alcançado para $\hat{\alpha} = y$.

Para $\alpha = \sigma^2(n - k)$ e $\lambda = (n - k)/2$ o EMV marginal para σ^2 é $\tilde{\sigma}^2 = \frac{1}{n-k} \|y - \hat{\mu}\|^2$.

Assim, mostramos que $\tilde{\sigma}^2$ é o EMV marginal para σ^2 baseado em $\|y - \hat{\mu}\|^2$.

2.9 Distribuições não-centrais

Consideraremos agora as distribuições χ^2 , t e F não-centrais e mostraremos como elas podem ser usadas para obter as funções de poder dos testes t e F .

2.9.1 Qui-quadrado não-central

Sejam Y_1, \dots, Y_n independentes com $Y_i \sim N(\xi, 1)$, $i = 1, \dots, n$ onde $\xi = (\xi_1, \dots, \xi_n)^T \in \mathbb{R}^n$.

Estudaremos aqui a distribuição de $U = \sum_{i=1}^n Y_i^2$.

Se $\xi = 0$, a distribuição de U é a distribuição qui-quadrado padrão (central).

Para $\xi \neq 0$, a distribuição de U é chamada uma distribuição de qui-quadrado não-central com n graus de liberdade e parâmetro de não-centralidade $\delta = \|\xi\|$, e é denotada por $\chi^2_{(n;\delta)}$.

Em particular, a distribuição de qui-quadrado central é $\chi^2_{(n)} = \chi^2_{(n;0)}$.

Vamos demonstrar agora que a distribuição de U depende de ξ somente através de $\|\xi\|$.

Seja $Y = (Y_1, \dots, Y_n)^T$ tal que $U = \|Y\|^2$. Decomponha Y tal que

$Y = \xi + \epsilon$ onde as componentes de $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ têm distribuição $N(0, 1)$, $i = 1, \dots, n$. Seja A , uma matriz ortogonal. Então,

$$\|AY\|^2 = Y^T A^T A Y = Y^T Y = \|Y\|^2$$

Assim, $P(U > u) = P(\|Y\|^2 > u) = P(\|AY\|^2 > u) = P(\|A\xi + A\epsilon\| > u)$.

Pelo teorema 2, a variável $A\epsilon$ tem a mesma distribuição de ϵ tal que

$$P(U > u) = P(\|A\xi + \epsilon\|^2 > u).$$

Podemos escolher A tal que para $\delta = \|\xi\|$ temos $A\xi = (\delta, 0, \dots, 0)^T$. Segue que $P(U > u)$ depende de ξ somente através de $\delta = \|\xi\|^2$.

Podemos tomar uma nova coordenada do sistema definida por $e_1 = \frac{\xi}{\|\xi\|}$ e e_2, \dots, e_n ortogonais à e_1 tal que e_1, \dots, e_n seja uma base ortonormal para \mathbb{R}^n .

Seja $X = (X_1, \dots, X_n)^T$ as coordenadas de Y na base (e_1, \dots, e_n) . Então,

$$X_i = e_i Y \text{ e } EX_1 = \frac{\xi}{\|\xi\|} \xi = \|\xi\| = \delta$$

$$EX_i = e_i \xi = 0, \quad i = 2, \dots, n.$$

Mas, $U = \|Y\|^2 = \|X\|^2 = X_1^2 + \sum_{i=2}^n X_i^2 = (\delta + \varphi_1)^2 + \sum_{i=2}^n \varphi_i^2$

onde $\varphi_1, \varphi_2, \dots, \varphi_n$ são independentes com distribuição $N(0, 1)$.

Assim, decompomos U em uma soma de duas variáveis independentes, uma $\chi_{(1;\delta)}^2$ e outra $\chi_{(n-1)}^2$.

Obtemos $E[(\varphi_1 + \delta)^2] = Var(\varphi_1 + \delta) + \{E[\delta + \varphi_1]\}^2 = 1 + \delta^2$.

Usando $E[\chi_{(n-1)}^2] = n - 1$ obtemos $EU = \delta^2 + n$. Usando fórmulas para o quarto momento da distribuição normal obtemos $Var(U) = 4\delta^2 + 2n$.

A seguir, apresenta-se uma fórmula de convolução para a distribuição de Qui-quadrado não central.

Sejam U_1 e U_2 independentemente distribuídas com $U_i \sim \chi^2_{(n_i; \delta_i)}$, $i = 1, 2$. Então, $U_1 + U_2 \sim \chi^2_{(n_1+n_2; (\delta_1^2+\delta_2^2)^{1/2})}$.

Sejam Y_1 e Y_2 definidos da seguinte forma:

$$Y_1 = \begin{pmatrix} \delta_1 + \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{n_1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ e } Y_2 = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \delta_2 + \epsilon_{n_1+1} \\ \epsilon_{n_1+2} \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Podemos escrever $U_1 = \|Y_1\|^2$ e $U_2 = \|Y_2\|^2$, onde $\epsilon_1, \dots, \epsilon_n$ são iid's $N(0, 1)$. Como $Y_1 \perp Y_2$, segue, do Teorema de Pitágoras que,

$$U_1 + U_2 = \|Y_1\|^2 + \|Y_2\|^2 = \|Y_1 + Y_2\|^2.$$

Além disso, $Y_1 + Y_2 = \xi + \epsilon$ onde $\epsilon^T = (\epsilon_1, \dots, \epsilon_n)$ e $\xi^T = (\delta_1, 0, \dots, 0, \delta_2, 0, \dots, 0)$. Assim, $U_1 + U_2 \sim \chi^2_{(n_1+n_2; (\delta_1^2+\delta_2^2)^{1/2})} \Rightarrow U \sim \chi^2_{(n; \delta)}$.

2.9.2 Distribuições t e F não-centrais.

Sejam U_1 e U_2 independentes e tais que $U_1 \sim \chi^2_{(f_1;\delta)}$ e $U_2 \sim \chi^2_{(f_2)}$. Então a distribuição de $F = \frac{U_1/f_1}{U_2/f_2}$ é chamada distribuição F não-central e denotada por $F \sim F_{(f_1, f_2; \delta)}$.

Similarmente, se Y e U são independentes com $Y \sim N(\delta, 1)$ e $U \sim \chi^2_{(f)}$, então a distribuição de $t = \frac{Y}{\sqrt{U/f}}$ é chamada distribuição t não-central, denotada por $t \sim t_{(f; \delta)}$.

Generalizando aqui a relação entre as distribuições t e F tem-se que $t^2_{(f; \delta)} = F_{(1, f; \delta)}$.

2.9.3 O poder dos testes t e F

Considere o modelo padrão onde Y_1, \dots, Y_n são independentes com $Y_i = \mu_i + \epsilon_i$, $\mu^T = (\mu_1, \dots, \mu_n)$, $Y^T = (Y_1, \dots, Y_n)$. Sejam $L_2 \subseteq L_1$ duas hipóteses lineares e considere a estatística F do teste de L_2 sob L_1

$$F(y) = \frac{\|\hat{\mu}_1 - \hat{\mu}_2\|^2 / (k_1 - k_2)}{\|y - \hat{\mu}_1\|^2 / (n - k_1)}$$

Sob L_2 , $F(Y) \sim F_{(k_1 - k_2, n - k_1)}$. Em particular, a distribuição de $F(Y)$ não depende do valor de μ para $\mu \in L_2$.

Considere agora a distribuição de $F(Y)$ sob $L_1 \setminus L_2$. Primeiro note que $Y - p_1(Y) = \mu + \epsilon - p_1(\mu) - p_1(\epsilon) = \epsilon - p_1(\epsilon)$.

Assim, $\|Y - p_1(Y)\|^2 = \|\epsilon - p_1(\epsilon)\|^2 \sim \chi_{(n - k_1)}^2$.

Para o vetor $p_1(Y) - p_2(Y)$ o resultado é diferente.

$$p_1(Y) - p_2(Y) = p_1(\mu) + p_1(\epsilon) - p_2(\mu) - p_2(\epsilon) = p_1(\epsilon) - p_2(\epsilon) + \mu - p_2(\mu).$$

Como $p_1(\epsilon) - p_2(\epsilon)$ é independente de $\epsilon - p_1(\epsilon)$ segue que $p_1(Y) - p_2(Y)$ e $Y - p_1(Y)$ são independentes.

$$\frac{1}{\sigma^2} \|Y - p_1(Y)\|^2 \sim \chi_{(n-k_1)}^2$$

$$\begin{aligned} & \frac{1}{\sigma^2} \|p_1(Y) - p_2(Y)\|^2 = \\ & \|p_1(\epsilon/\sigma) - p_2(\epsilon/\sigma) + \frac{1}{\sigma}\{\mu - p_2(\mu)\}\|^2 \end{aligned}$$

Como $\epsilon_i/\sigma \sim N(0, 1)$ segue que $\frac{1}{\sigma^2} \|p_1(Y) - p_2(Y)\|^2 \sim \chi_{(k_1-k_2;\delta)}^2$, onde $\delta = \|\frac{1}{\sigma}\{\mu - p_2(\mu)\}\| = \frac{1}{\sigma} \|\mu - p_2(\mu)\|$.

Observe que para provar que $\frac{1}{\sigma^2} \|Y - p_1(Y)\|^2 \sim \chi^2_{(n-k_1)}$ temos que representar $p_1(\epsilon/\sigma) - p_2(\epsilon/\sigma)$ como um vetor em $\mathbb{R}^{k_1-k_2}$ com componentes iid's $N(0, 1)$. O parâmetro de não centralidade mede assim, em unidades de σ , a distância ortogonal entre μ e L_2 .

Pela definição da distribuição F não-central, obtemos

$$F(Y) \sim F_{(k_1-k_2, n-k_1; \delta)}$$

Podemos agora calcular o poder do teste F . Seja α o nível do teste tal que rejeitamos L_2 se $F(y) > F_{1-\alpha}(k_1-k_2, n-k_1)$.

O poder do teste F é então uma função de δ :

$$Poder_\alpha(\delta) = P\left(F(Y) > F_{1-\alpha}(n-k_1, k_1-k_2; \delta)\right)$$

onde $F(Y) \sim F_{(n-k_1, k_1-k_2; \delta)}$.

Mas,

$$E \left[\frac{\|Y - p_1(Y)\|^2}{n - k_1} \right] = \sigma^2$$

para $\mu \in L_2$ e $\mu \in L_1 \ominus L_2$.

$$E \left[\frac{\|p_1(Y) - p_2(Y)\|^2}{k_1 - k_2} \right] = \sigma^2 \left(1 + \frac{\delta^2}{k_1 - k_2} \right)$$

Isto confirma que a média do numerador da estatística F é uma função crescente de δ . Para $\delta = 0$, F é a razão de duas quantidades independentes, ambas com média σ^2 . Assim, sob L_2 , esperamos que F seja próxima de 1.

O poder do teste t bilateral pode ser obtido a partir do poder do teste F pois $t_{(f)}^2 = F_{(1,f)}$.

Considere então o poder do teste de $H_0 : \beta_k = \beta_0$, cuja estatística de teste é $t(Y) = \frac{\hat{\beta}_k - \beta_0}{s.e.(\hat{\beta}_k)}$. Sob H_0 , $\hat{\beta}_k - \beta_0 \sim N(\beta_k - \beta_0, \sigma^2 c_{jj})$.

Assim, como $\hat{\beta}_k$ e $s.e.(\hat{\beta}_k)$ são independentes, $t(Y) \sim t_{\left(n-k; (\beta_k - \beta_0)/(\sigma c_{jj}^{1/2})\right)}$.

Suponha que $H_A : \beta_k > \beta_0$. Então o teste rejeita H_0 se $t(y) > t_{1-\alpha(n-k)}$. Logo, o poder do teste será dado por

$$P(t(Y) > t_{1-\alpha(n-k)}) \text{ onde } t(Y) \sim t_{\left(n-k; (\beta_k - \beta_0)/(\sigma c_{jj}^{1/2})\right)}.$$