

Djalma Pessoa
Estatística
Não-Paramétrica

COPYRIGHT © by DJALMA PESSOA (1977)

Nenhuma parte deste livro pode ser reproduzida,
por qualquer processo, sem a permissão do autor.

INSTITUTO DE MATEMÁTICA PURA E APLICADA

Rua Luiz de Camões, 68
20.000 - Rio de Janeiro - RJ

Prefácio

Estas notas visam apresentar, em nível elementar, algumas técnicas e idéias básicas da Estatística Não-Paramétrica. Os procedimentos abordados são os mais conhecidos e em número bastante limitado, apenas o estritamente necessário para introduzir o enfoque característico da Estatística Não-Paramétrica. Isto pode acarretar uma falsa idéia de limitação dos métodos não-paramétricos. Na verdade, técnicas não paramétricas já foram desenvolvidas para análise de dados nos mais diversos tipos de problemas estatísticos. Uma parcela mínima destas técnicas é apresentada neste texto. Daí, talvez, a utilidade do mesmo ser diminuta para aqueles, principalmente, interessados em aplicações de técnicas estatísticas.

A abordagem dos métodos não paramétricos tem, recentemente, sido utilizada como uma via didática de introdução dos conceitos básicos da Estatística. Tal esquema, em comparação com o clássico, reduz bastante os pré-requisitos da Teoria das Probabilidades. Para a leitura destas notas, ao contrário, é desejável algum conhecimento prévio de Estatística. Basicamente as noções, em nível elementar, de teste de Hipótese e Estimativas pontual e

por intervalo, contidas nos textos de introdução à Estatística. Um resumo destes pré-requisitos é apresentado no Apêndice no final das notas.

Infelizmente, o tempo disponível para a redação deste texto não permitiu a inclusão de exemplos de aplicações interessantes das técnicas abordadas. Estas são bastante utilizadas em Educação, Psicologia, Biologia, etc. Um levantamento de dados em periódicos nacionais, nestas áreas, fornecem um farto material para aplicações. Quando estas revistas apresentam alguma análise dos dados experimentais, geralmente utilizam técnicas paramétricas (hipótese de Normalidade). Alguns textos disponíveis de Estatística Não-Paramétrica apresentam uma boa quantidade de aplicações a outras áreas. Por considerá-las remotamente ligadas à nossa realidade decidimos, ao invés de reproduzi-las, utilizar dados simulados. A vantagem destes é permitir o conhecimento prévio do provável resultado da análise estatística.

Não-Paramétrica é uma das áreas da Estatística de maior desenvolvimento recente e a qual se tem dedicado grande esforço de pesquisa. Com a publicação deste texto elementar esperamos despertar o interesse pelo seu estudo mais aprofundado.

Rio de Janeiro, junho de 1977

DJALMA PESSOA

ÍNDICE

	pag.
Capítulo I - Modelo Estatístico	1
Capítulo II - Comparação de dois Tratamentos; Problema de duas Amostras	
§1. Introdução	14
§2. Distribuição de W_S quando $F = G$	18
§3. Distribuição aproximada de W_S sob a hipótese	23
§4. Determinação de $E_H(W_S)$ e $Var_H(W_S)$	25
§5. Caso de Empates	27
§6. Aproximação Normal para a distribuição de W_S^*	33
§7. Potência do Teste de Wilcoxon	36
§8. Cálculo de potência do Teste de Wilcoxon	38
§9. Estimação do Efeito do Tratamento	44
§10. Intervalos de Confiança para o efeito Δ de tratamento	50
§11. Teste de Escores Normais	52
§12. Comentários sobre os testes apresentados	56
§13. Testes para diferença entre escalas	59
§14. Teste de Komogorov-Sminorv	64
Exercícios - Capítulo II	75

Capítulo III - Comparações Emparelhadas -
Problema de uma amostra

§1. Introdução	81
§2. Teste do sinal	83
§3. Aproximação normal para $\pi(p)$	86
§4. Teste de Wilcoxon dos postos sinalizados	90
§5. Cálculo de $E_H(V_S)$ e de $Var_H(V_S)$	95
§6. Aproximação normal para distribuição de V_S	96
§7. Caso de Empates	99
§8. Cálculo de $E_H V_S^*$ e de $Var_H V_S^*$	102
§9. Aproximação Normal para distribuição de V_S^* ..	104
§10. Potência do Teste de Postos sinalizados de Wilcoxon	108
§11. Aproximação para a Potência do Teste de Postos Sinalizados de Wilcoxon	111
§12. Cálculo de $E(V_S)$ e $Var(V_S)$ quando L não é simétrica em torno de zero	112
§13. Estimação do efeito de tratamento	119
§14. Intervalo de confiança para o efeito de tratamento	122
Exercícios - Capítulo II	127

Capítulo IV - Problema de s-amostras - Blocos
Aleatorizados Completos

§1. Introdução	129
§2. Teste de Kruskal-Wallis	132
§3. Teste de Jonckheere-Terpstra	137

	pag.
§4. Blocos Aleatorizados Completos	139
§5. Teste de Friedman	141
§6. Postos alinhados	147
Exercícios - Capítulo IV	150
APÊNDICE	155
TABELAS	177
REFERÊNCIAS	195

* * *

Capítulo I

Modelo Estatístico

Um aspecto de fundamental importância nas aplicações da Estatística é a construção de um modelo Estatístico adequado. Toda informação acerca do processo de obtenção dos dados, bem como, características relevantes da situação estudada devem ser sumarizados no modelo.

Usualmente, este modelo é especificado por uma família \mathcal{F} de funções de distribuição, e o espaço de todos os resultados possíveis para o vetor de dados. Este, consiste de um conjunto de números $x = (x_1, \dots, x_n)$, observados em um experimento. O relacionamento entre x e \mathcal{F} é feito através de um vetor aleatório $X = (X_1, \dots, X_n)$. No modelo, consideramos o vetor de dados x como o valor de um vetor aleatório X cuja distribuição F é um elemento da família \mathcal{F} . Na sua forma mais simples, este modelo considera, ainda, as variáveis X_1, \dots, X_n independentes e identicamente distribuídos. A distribuição do vetor fica então, especificada a partir da distribuição comum, univariada, de suas variáveis aleatórias componentes. Este é o modelo adequado para descrever o processo

de amostragem aleatória simples, onde todos os subconjuntos de n indivíduos de uma população tem a mesma chance de serem retirados, e neles são observados características numéricas x_1, \dots, x_n . Sendo este o caso que vai nos interessar, passaremos a denotar por F a distribuição (univariada) comum dos X 's e por \mathfrak{F} a família de distribuições (univariadas) considerada no modelo.

Comumente, nos referimos à F como distribuição subjacente, ou verdadeira distribuição, dos X 's. Outras vezes, em linguagem mais informal, dizemos que os dados x_1, \dots, x_n são "gerados" pela distribuição F . É importante observar que os dados x_1, \dots, x_n , neste modelo, não são simplesmente encarados como números. Ao contrário, o relacionamento dos dados com os outros componentes do modelo nos permite extrair a informação contida naquelas. Abaixo, esquematizamos os componentes do modelo e o relacionamento entre os mesmos.

dados $v.$ a's i.i.d. família de distribuições
 $(x_1, \dots, x_n) \rightarrow (X_1, \dots, X_n) \rightarrow \mathfrak{F}$, onde

- a) x_1, \dots, x_n são valores das variáveis aleatórias X_1, \dots, X_n .
- b) As variáveis aleatórias X_1, \dots, X_n têm distribuição comum F .
- c) A distribuição F pertence a \mathfrak{F} .

A família \mathfrak{F} de distribuições, da qual F é um de seus elementos, é especificada a partir de propriedades do processo ou fenômeno em estudo. O objetivo geral da análise estatística é a partir dos dados fazer afirmações sobre F . Usualmente, no sentido de estabelecer a sua inclusão em alguma subfamília \mathfrak{F}_0 de \mathfrak{F} , ou seja, afirmações do tipo $F \in \mathfrak{F}_0 \subset \mathfrak{F}$, melhor especificando F .

A seguir, através de um exemplo, ilustramos a construção de um modelo estatístico.

Exemplo 1: Consideremos uma grandeza física desconhecida

θ , que pode ser, por exemplo, um comprimento.

São feitas n medições x_1, \dots, x_n desta grandeza e a partir destes números desejamos estimar θ .

Em Física, na escola secundária, aprendemos (aprendíamos) que uma estimativa de θ é $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Uma justificativa apresentada na adoção desta estimativa era

(i) a soma dos resíduos $\sum_{i=1}^n (x_i - \bar{x})$ é nula;

(ii) a soma dos resíduos quadráticos $\sum_{i=1}^n (x_i - \bar{x})^2$ é

minimizado tomando $\hat{\theta}(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i$. Pa-

ra isto, basta notar que $\sum_{i=1}^n (x_i - \hat{\theta})^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\hat{\theta} - \bar{x})^2$.

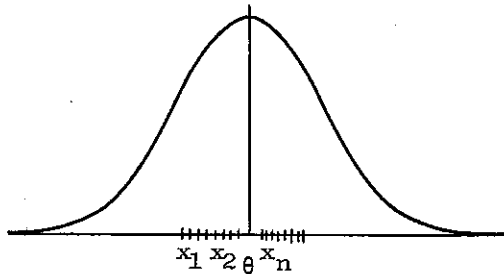
A propriedade (ii) é uma forma particular do princípio dos mínimos quadrados.

Nesta abordagem, os dados (x_1, \dots, x_n) são enca-

rados simplesmente como números. A relação destes com a grandeza θ estimada não fica clara. É de se esperar, por exemplo, que valores grandes de θ produzam valores grandes de x_1, \dots, x_n . Consideremos, agora, o modelo estatístico descrito acima, para esta situação. Neste modelo, x_1 é o valor de uma variável aleatória X_1 ; x_2 é o valor de uma variável aleatória X_2 , etc., x_n é o valor de X_n . As variáveis X_1, \dots, X_n são independentes e com distribuição comum F , desconhecida. Qual, então, a relação entre os dados (x_1, \dots, x_n) e o valor de θ ? No modelo tomamos a distribuição F com o dependente de θ . Para explicitar esta dependência, escrevemos F_θ no lugar de F . A grandeza estimada θ serve como parâmetro da família $\{F_\theta\}_{\theta \in \Theta}$ de distribuições dos X 's. Vamos, agora, tomar $x_i = \theta + \epsilon_i, \dots, x_n = \theta + \epsilon_n$, onde os ϵ_i são os erros na medição, considerados aleatórios. No modelo, $\epsilon_1, \dots, \epsilon_n$ são, respectivamente, valores de variáveis aleatórias E_1, \dots, E_n independentes e com distribuição comum F_0 .

Vamos, agora, supor que estes erros sejam, na sua maioria, pequenos em valor absoluto, e que as frequências de erros positivos e negativos, de iguais valores absolutos, sejam iguais. Suponhamos, ainda, que erros muito grandes sejam muito raros. Diante destas características, é razoável tomar a distribuição F_0 como sendo $N(0, \sigma_0^2)$,

onde supomos σ_0^2 conhecido. Deste modo, a distribuição das variáveis aleatórias observadas, X_1, \dots, X_n , pertence à família das normais $\{N(\theta, \sigma_0^2)\}_{\theta \in \Theta}$. O valor desconhecido $\theta = \theta_0$, que estamos estimando, é justamente aquele que especifica a distribuição das variáveis aleatórias X_1, \dots, X_n . De que forma os dados (x_1, \dots, x_n) estão relacionados com θ ? Os valores x_1, \dots, x_n tendem a se aglomerar em torno de θ , e valores grandes de θ tendem a "gerar" valores grandes, de x_1, \dots, x_n .



No modelo, tomamos $\mathfrak{F} = \{N(\theta, \sigma_0^2)\}_{\theta \in \Theta}$. Esta hipótese é, geralmente, muito forte. Uma hipótese menos restritiva é supor $\mathfrak{F}' =$ família de todas as distribuições com densidade e simétricas em torno de algum valor θ . A distribuição dos erros E_1, \dots, E_n , neste caso, satisfaz $P[E_i \in (y_1, y_2)] = P[E_i \in (-y_1, -y_2)]$ com $y_1, y_2 > 0$ e $y_1 < y_2$, ou ainda, $F_{E_i}(y) = 1 - F_{E_i}(-y) \forall y \in \mathbb{R}$. Assim, temos $\mathfrak{F} \subset \mathfrak{F}'$. Uma hipótese menos restritiva, ainda, sobre a família de distribuições subjacentes, é supor ape-

nas que a frequência de erros positivos é igual a de erros negativos, ou seja, $P(E_1 < 0) = P(E_1 > 0)$. Designando por \mathfrak{F}'' a família de todas as distribuições com densidade e tais que $P(X_1 - \theta < 0) = P(X_1 - \theta > 0)$, temos $\mathfrak{F} \subset \mathfrak{F}' \subset \mathfrak{F}''$.

Observe, considerando a família \mathfrak{F} , que cada elemento da família é caracterizado por um valor de θ . Dizemos então que θ é um parâmetro da distribuição ou que a família é parametrizada por θ . O modelo, neste caso, é dito paramétrico. No modelo paramétrico, no caso geral, a família de distribuições \mathfrak{F} pode ser indexada por um subconjunto Θ do R^n , para algum n . No caso de \mathfrak{F}' , o valor de θ não fixa uma distribuição única. A forma funcional da distribuição é livre, e podemos ter como distribuições simétricas em torno de θ : Normais, Cauchy, exponencial duplas, etc. Ao contrário de \mathfrak{F} , \mathfrak{F}' não pode ser parametrizada, de modo natural, por um subconjunto Θ de R^n , para algum n . O modelo neste caso é dito não paramétrico. Teríamos, também, um modelo não paramétrico quando a família de distribuições subjacentes, adotada no modelo, fosse \mathfrak{F}'' . A questão - qual das três famílias deve ser utilizada no modelo? - não tem resposta fácil. Por um lado, a adoção de \mathfrak{F} no modelo, permite fazer inferências mais precisas sobre a verdadeira distri

buição dos X 's, a partir dos dados x_1, \dots, x_n . Por outro lado, a adoção de \mathcal{F} é uma restrição muito forte e a verdadeira distribuição dos X 's pode não ser um elemento de \mathcal{F} .

Procedimentos estatísticos, com desempenho ótimo para a família \mathcal{F} , podem ser muito pobres, quando a distribuição subjacente é um elemento de \mathcal{F}' , sem pertencer a \mathcal{F} . A média das observações $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, no Exemplo 1, é o estimador não viciado de mínima variância quando a família de distribuições subjacentes é \mathcal{F} . O seu desempenho é simplesmente desastroso, quando trabalhamos com \mathcal{F}' , para o caso de F ser Cauchy.

De fato, raramente o estatístico dispõe, em situações concretas, de informações sobre o processo observado que tornem legítima a adoção de \mathcal{F} . No Exemplo 1, esta utilização é justificada, por alguns autores, do seguinte modo: o erro total é a soma de um grande número de pequenos erros independentes e, portanto, pelo teorema central do Limite, tem distribuição normal. Este argumento é considerado, por outros autores, pouco mais que um apelo ao fantasma de Gauss (autor da Normal). Outro argumento razoável é que qualquer distribuição simétrica em torno de zero com densidade "suave" e unimodal, pode, na sua parte central, ser aproximada por uma $N(0, \sigma^2)$. Isto, porém, não elimina a possibilidade de que os erros maiores possam

ter uma distribuição totalmente diversa da normal.

Os procedimentos estatísticos que apresentamos nestas notas são apropriados para trabalhar com as famílias "grandes" de distribuições, do tipo de \mathfrak{F}' e \mathfrak{F}'' no exemplo anterior.

É surpreendente como alguns destes procedimentos perdem muito pouca eficiência, quando comparados com o melhor procedimento para a família \mathfrak{F} . Outra vantagem é que eles são baseados, apenas, na ordem ou posto das observações e não nos seus valores numéricos. Assim, podem ser utilizados em situações onde se conhece apenas qual a maior observação, a segunda maior, etc. Muitos destes procedimentos envolvem, como veremos, estatísticas cujas distribuições são independentes da distribuição subjacente (distribution free).

Famílias de distribuições mais utilizadas.

Distribuições discretas:

- a) Distribuição Binomial com parâmetros n e p ; n inteiro positivo, $0 \leq p \leq 1$, $q = 1-p$.

$$P(X=x) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, \dots, n$$

$$E(X) = np \quad e \quad \text{Var}(X) = npq.$$

- b) Distribuição hipergeométrica com parâmetros N , d e n .

$$P(X=x) = \binom{d}{x} \binom{N-d}{n-x} / \binom{N}{n} \quad \begin{matrix} x = 0, 1, \dots, d \\ \max(0, d+n-N) \leq x \leq \min(d, n) \end{matrix} \quad e$$

$$E(X) = n \frac{d}{N} \quad e \quad \text{Var}(X) = n \frac{d}{N} \left(1 - \frac{d}{N}\right)$$

Distribuições Contínuas:

São caracterizadas por terem a função de distribuição,

$F(x) = P(X \leq x)$, contínua, ou equivalentemente, $P(X=x) = 0$ para todo x . Uma importante propriedade de variáveis aleatórias com distribuição contínua é a seguinte:

Teorema: Se X_1, \dots, X_n são n variáveis aleatórias independentes com distribuições contínuas

F_1, \dots, F_n , então $X_i \neq X_j$, $1 \leq i \neq j \leq n$, com probabilidade 1.

Prova: V. Hájek, Capítulo 2, pág. 20.

As distribuições contínuas que apresentamos abaixo têm a propriedade de possuir densidade, isto é, existe $f \geq 0$ tal que $F(x) = \int_{-\infty}^x f(t) dt$. A densidade de cada distribuição listada abaixo é apresentada na forma padrão.

Distribuição Normal

A densidade da distribuição normal é

$$f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < +\infty$$

Distribuição de Cauchy

A densidade da distribuição de Cauchy é

$$f_{0,1}(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < +\infty$$

Distribuição Logística

A distribuição logística tem densidade

$$f_{0,1}(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \quad -\infty < x < +\infty$$

Exponencial dupla

A distribuição exponencial dupla tem densidade

$$f_{0,1}(x) = \frac{1}{2} e^{-|x|}, \quad -\infty < x < +\infty.$$

Para cada uma destas densidades, podemos gerar uma família de distribuições a partir de $f(x;a,b) = \frac{1}{a} f_{0,1}\left(\frac{x-b}{a}\right)$, $a > 0$. Estas densidades recebem, ainda, o mesmo nome que as correspondentes f_0 . Para cada f_0 fixo, com a e b variando, o conjunto de densidades obtidas forma uma família de locação-escala. O parâmetro a mede a escala e b a locação. (Veja Apêndice).

As densidades apresentadas acima, na sua forma padrão, são simétricas em torno de zero; isto é, satisfazem $F(X) = 1-F(-x)$, para qualquer $x \in \mathbb{R}$.

Duas outras famílias de locação-escala importantes são a uniforme e a exponencial.

Distribuição Uniforme

A densidade (padrão) é dada por

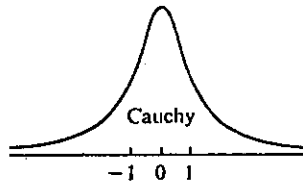
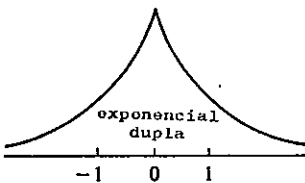
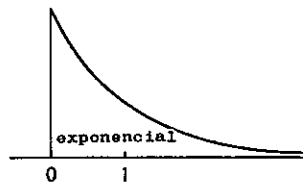
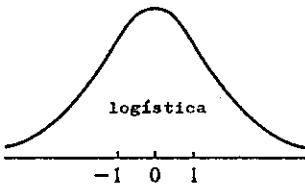
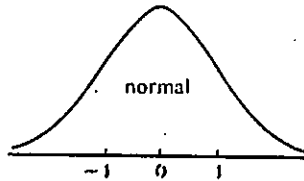
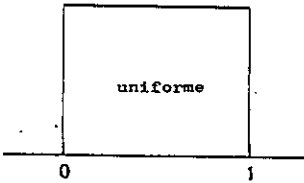
$$f_{0,1}(x) = 1 \quad \text{se} \quad 0 \leq x \leq 1$$
$$= 0 \quad \text{caso contrário}$$

Distribuição Exponencial

A densidade da exponencial (padrão) é

$$f_{0,1}(x) = \exp(-x) \quad x \geq 0$$

Os gráficos destas densidades são apresentados na figura abaixo.



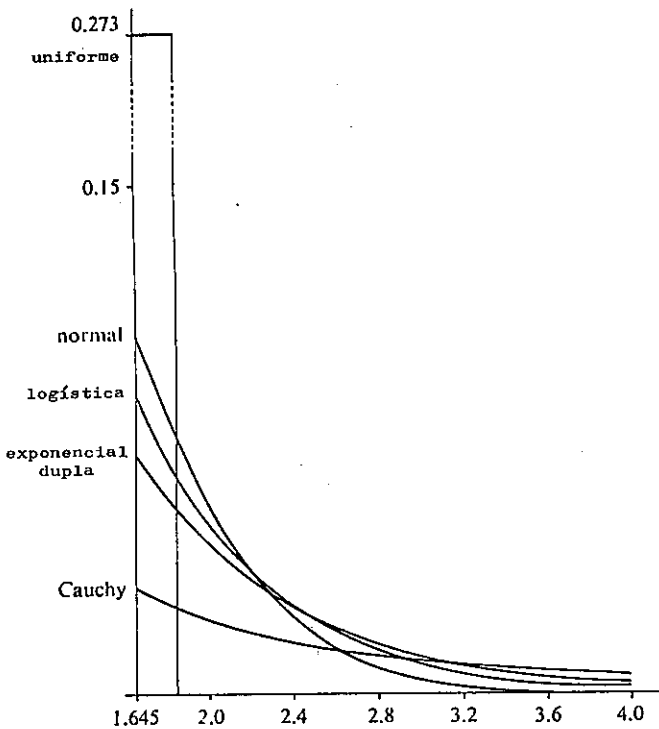
Uma característica importante nestas densidades é a velocidade com que tendem para zero quando $x \rightarrow +\infty$ ou $x \rightarrow -\infty$. A densidade da normal, por exemplo, tende para zero extremamente rápido, da ordem de $e^{-x^2/2}$. Esta propriedade, torna muitas vezes pouco realística a adoção da família normal em um modelo estatístico. É frequente, na prática, o aparecimento dos chamados erros grosseiros.

Pelo gráfico das densidades, vemos que algumas têm o mesmo aspecto geral: simétricas, unimodais. Este gráfico, contudo, não permite fazer distinções entre as caudas. Abaixo, reproduzimos a comparação, entre as caudas das distribuições, apresentada em Hájek (pag. 43). Com a finalidade de tornar as caudas comparáveis, os parâmetros a e b foram calculados de modo que

$$F\left(-\frac{b}{a}\right) = \frac{1}{2} \quad \text{e} \quad F\left(\frac{1,645-b}{a}\right) = 0.95$$

para cada uma das distribuições. Os tipos representativos têm, portanto, a mesma mediana e o mesmo 95% percentil que a normal padrão.

Caudas das Distribuições



A partir do gráfico, observamos que a distribuição com cauda mais leve é a uniforme, depois e em ordem vêm a normal, a logística, a exponencial dupla e, por fim, a Cauchy que tem a cauda mais grossa. Caso seja provável a inclusão na amostra de valores excêntricos (outliers), devemos adotar uma densidade de caudas mais grossa. Tal escolha torna o modelo mais realístico.

Capítulo II

Comparação de Dois Tratamentos; Problema de duas Amostras

§1. Introdução.

Em muitas situações, temos o problema de verificar se um novo tratamento apresenta melhores resultados que um tratamento padrão, ou mesmo, de comparar dois novos tratamentos. Como exemplos temos as comparações de dois analgésicos utilizando o tempo decorrido até o alívio da dor; de duas marcas de gasolina pela quilometragem por litro consumido; de dois métodos de treinamento físico, etc...

Evidentemente, certas precauções devem ser tomadas na realização do experimento, no sentido de evidenciar diferenças exclusivamente devidas aos diferentes tratamentos. Diferenças atribuíveis a outras causas podem mascarar completamente os efeitos dos tratamentos.

Suponhamos que uma amostra aleatória de tamanho $N = m+n$ é retirada de uma população onde n indivíduos são submetidos ao tratamento e m servem de controle.

Consideremos o seguinte modelo: X_1, \dots, X_m , as

respostas dos controles, são variáveis aleatórias independentes com distribuição comum F .

Y_1, \dots, Y_n , as respostas dos indivíduos tratados, são variáveis aleatórias independentes com distribuição comum F . Isto é $P(X_i \leq x) = F(x)$ e $P(Y_i \leq y) = G(y)$ além disto, consideremos X_i e Y_j independentes para qualquer par (i, j) .

A hipótese de nenhum efeito de tratamento é traduzida em termos do modelo, por:

$$H: G = F.$$

Vamos nos restringir à alternativa de que o tratamento tem o efeito de aumentar as respostas. Isto é, com relação a G , F gera valores maiores com maior frequência e menores com menor frequência. Isto pode ser expresso pela desigualdade

$$G(x) \leq F(x), \quad \forall x$$

com desigualdade estrita para algum x .

Definição 1 - Uma variável aleatória Y com distribuição G é estocasticamente maior que uma variável aleatória X com distribuição F , se $G(x) \leq F(x) \quad \forall x$ e existe x_0 tal que $G(x_0) < F(x_0)$. Neste caso dizemos também que G é estocasticamente maior do que F .

Exemplo 1: Seja $\Delta > 0$ e $G(x) = F(x-\Delta) \quad \forall x$, isto é,

G é obtida deslocando-se F de Δ . Este é o chamado modelo de deslocamento, que traduz matematicamente a hipótese do efeito do tratamento adicionar a constante Δ à resposta do controle, independentemente desta resposta.

Retornemos, agora, ao problema de testar:

H: $F = G$ contra K: G estocasticamente maior que F.

Para solucionar este problema vamos determinar um subconjunto do espaço de todas as observações possíveis e adotar a regra de que se o vetor de observações

$(x_1, \dots, x_m; y_1, \dots, y_n)$ pertencer a este conjunto rejeitaremos a hipótese. Valores muito maiores dos y's do que dos x's favorecem a rejeição de H. Deste modo, vamos tomar a amostra conjunta dos x's e dos y's e ordenar os seus valores. A partir daí podemos definir o posto de cada observação como o número de observações menores ou iguais a ela. Por exemplo, seja $m = n = 3$ e $X_1 = 1, X_2 = 3, X_3 = 5, Y_1 = 2, Y_2 = 4, Y_3 = 6$ então $\text{Posto}(X_1) = 1, \text{Posto}(X_2) = 3, \text{etc.}$

Denotemos por $S_1 < S_2 < \dots < S_n$ os postos dos Y's na amostra conjunta. Valores grandes dos S's favorecem a rejeição de H. Vamos então considerar a estatística $W_s = S_1 + \dots + S_n$ e tomar como região de rejeição $W_s \geq c$, onde c será determinado a partir do nível do teste. Obtemos então o teste de Wilcoxon: Rejeite H: $F=G$,

se $W_S \geq c$ onde c é determinado por $P_H(W_S \geq c) = \alpha$, sendo α o nível de significância do teste. Para completar a especificação do teste precisamos determinar c .

A notação P_H utilizada acima, indica que estamos considerando a distribuição de W_S quando $F = G$.

Teorema 1 - Se $X_1, \dots, X_m; Y_1, \dots, Y_n$ são independentes com distribuição comum F contínua então:

$$(2.1) \quad P_H(S_1 = x_1, \dots, S_n = s_n) = \frac{1}{\binom{N}{n}}$$

onde $s_1 < s_2 < \dots < s_n$ e $s_i \in \{1, 2, \dots, N\}$.

Prova: Considere (s_1, \dots, s_n) , então algum subconjunto de n variáveis entre as $N = n+m$ variáveis terá como vetor de postos (s_1, \dots, s_n) . Por outro lado, sendo a distribuição conjunta de $(X_1, \dots, X_m; Y_1, \dots, Y_n)$ simétrica nas N variáveis, qualquer subconjunto de n variáveis tem a mesma probabilidade de ter como postos (s_1, \dots, s_n) . Existem $\binom{N}{n}$ diferentes grupos de n variáveis escolhidas entre as N variáveis $(X_1, \dots, X_m; Y_1, \dots, Y_n)$. Portanto, cada um destes subconjuntos tem probabilidade $1/\binom{N}{n}$ de ter os postos (s_1, \dots, s_n) .

Observações: 1) Em todas as considerações feitas acima foi suposto não haver valores iguais entre os x 's e os y 's. Posteriormente trataremos do ca-

so em que houver "empates".

2) Podemos garantir, com probabilidade 1, a ausência de empates se F e G são contínuas.

3) O valor do teorema acima deriva do fato da distribuição conjunta de (S_1, \dots, S_n) ser independente de F . Isto não ocorre quando F não é contínua.

A igualdade (2.1) nos diz que os valores dos S 's poderiam ser obtidos de modo equivalente (em termos de distribuição de probabilidades), tomando-se uma amostra aleatória de tamanho n de população $\pi = \{1, 2, \dots, N\}$, e designando-se por S_1 o menor elemento de amostra, S_2 o segundo menor, etc.... e S_n o maior. Este experimento pode ser executado com ajuda de uma tabela de números aleatórios, sem o conhecimento de F (contínua).

§2. Distribuição de W_S quando $F = G$.

—
Sendo $W_S = S_1 + \dots + S_n$, a partir da distribuição conjunta de (S_1, \dots, S_n) podemos determinar a distribuição de W_S . Para isto, seja $\#(w; n, m)$ o número de diferentes amostras de tamanho n retiradas de população $\pi = \{1, \dots, N\}$ tendo a soma dos elementos iguais a w .
Então

$$(2.2) \quad P_H(W_S=w) = \frac{\#(w;n,m)}{\binom{N}{n}} .$$

Exemplo 2: Vamos achar a distribuição de W_S para $m = n = 3$. Para isto, enumeremos todas as amostras possíveis de tamanho 3 da população

$\pi = \{1,2,3,4,5,6\}$ e as somas correspondentes dos elementos das amostras.

(s_1, s_2, s_3)	(1,2,3)	(1,2,4)	(1,2,5)	(1,2,6)	(1,3,4)	(1,3,5)
w	6	7	8	9	8	9

(s_1, s_2, s_3)	(1,3,6)	(1,4,5)	(1,4,6)	(1,5,6)	(2,3,4)	(2,3,5)
w	10	10	11	12	9	10

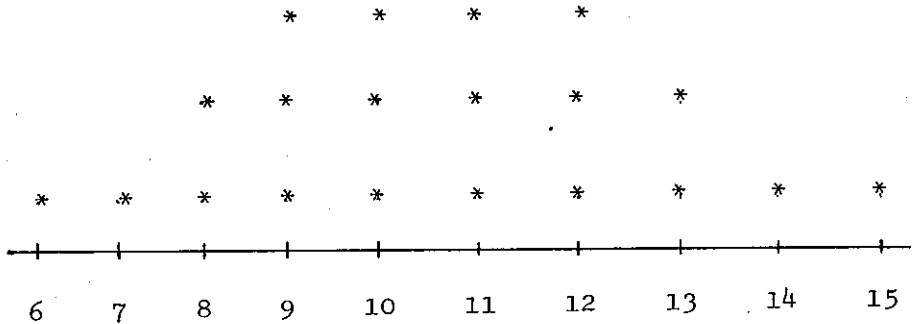
(s_1, s_2, s_3)	(2,3,6)	(2,4,5)	(2,4,6)	(2,5,6)	(3,4,5)	(3,4,6)
w	11	11	12	13	12	13

(s_1, s_2, s_3)	(3,5,6)	(4,5,6)
w	14	15

A distribuição de W_S é dada por:

w	6	7	8	9	10	11	12	13	14	15
$P(W_S=w)$	1/20	1/20	2/20	3/20	3/20	3/20	3/20	2/20	1/20	1/20

Diagrama de distribuição de
frequências de W_S



Observe que a distribuição de frequências de W_S é simétrica em torno de 10.5.

Quando m e n crescem, é impraticável a enumeração de todos os casos. É possível, então, apelar para as várias tabelas existentes que, usualmente, apresentam a distribuição de $W_{XY} = W_S - 1/2 n(n+1)$ no lugar da W_S . Ao contrário de W_S , a distribuição de W_{XY} é a mesma quando, por exemplo, $m = k_1$, $n = k_2$ e $m = k_2$, $n = k_1$. Outra vantagem é que o menor valor de W_{XY} é zero para quaisquer valores de m , n , simplificando a construção da tabela.

O teste de Wilcoxon pode ser definido, de modo equivalente, utilizando os postos $R_1 < \dots < R_m$ dos contro

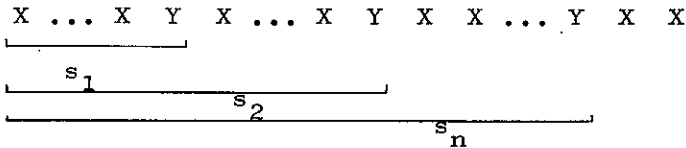
les. Neste caso, o teste rejeitaria quando $W_R = R_1 + \dots + R_m$ fosse pequeno, pois $W_R + W_S = 1/2 N(N+1)$. Considere, agora, as estatísticas $W_{XY} = W_S - 1/2 n(n+1)$ e $W_{YX} = W_R - 1/2 m(m+1)$.

Teorema 2 - Sejam W_{XY} e W_{YX} definidos acima, então

$$(2.3) \quad W_{XY} = \text{número de pares } (X_i, Y_j) \text{ tais que } X_i < X_j$$

$$(2.4) \quad W_{YX} = \text{número de pares } (X_i, Y_j) \text{ tais que } Y_j < X_i$$

Prova: Vamos provar (2.3). Como os pontos dos Y's são s_1, \dots, s_n temos uma disposição do seguinte tipo:



Então:

- número de X's menores que o primeiro Y = $S_1 - 1$
- número de X's menores que o segundo Y = $(S_2 - 1) - 1$
- número de X's menores que o último Y = $(S_n - 1) - (n-1)$

Logo:

$$\begin{aligned}
 & \text{o número total de pares } (X_i, Y_j) \text{ com } X_i < X_j = \\
 & = (S_1 - 1) + (S_2 - 2) + \dots + (S_n - n) = W_S - 1/2 n(n+1) = W_{XY}.
 \end{aligned}$$

Nota: W_{XY} e W_{YX} são chamadas estatísticas de Mann-Whitney e podemos usá-las para definir o teste de Wilcoxon.

Teorema 3 - A distribuição nula (sob a hipótese) de W_S é simétrica em torno de $1/2 n(N+1)$.

Prova: Considere os indivíduos ordenados em disposição inversa da considerada na obtenção dos postos. O indivíduo que, na primeira disposição tinha posto 1 agora terá N, o que tinha 2 agora terá N-1, etc... e em geral o que tinha S agora terá N-S+1. Os postos nesta nova ordenação são chamados anti-postos e designados por $(S'_1 \dots S'_n)$. A uma dada énupla de anti-postos corresponde uma e só uma énupla de postos, portanto:

$$P_H(S'_1 = s_1, \dots, S'_n = s_n) = \frac{1}{\binom{N}{n}}.$$

Como os vetores (S_1, \dots, S_n) e (S'_1, \dots, S'_n) têm a mesma distribuição, então $W_{S'} = S'_1 + \dots + S'_n$ tem a mesma distribuição que W_S .

Ou seja, $\sum_{i=1}^n (N-S_i+1) = n(N+1) - W_S$ tem a mesma distribuição de W_S .

Queremos provar que a distribuição de W_S , sob a hipótese, atribui a mesma massa a pontos equidistantes de $1/2 n(N+1)$, isto é:

$$(2.5) \quad P_H[W_S = 1/2 n(N+1)+k] = P_H[W_S = 1/2 n(N+1)-k]$$

Como W_S e $n(N+1)-W_S$ têm a mesma distribuição:

$$P_H[W_S = 1/2 n(N+1)+k] = P_H[n(N+1)-W_S = 1/2 n(N+1)+k]$$

$$= P_H[W_S = 1/2 n(N+1)-k].$$

Teorema 4 - W_{XY} e W_{YX} definidos acima têm a mesma distribuição.

Prova: Temos,

$$P_H[W_{XY}=k] = P_H[W_S-1/2 n(n+1)=k] =$$

$$= P[W_S-1/2 n(n+1)-1/2 mn=k-1/2 mn] =$$

$$= P_H[W_S-1/2 n(N+1)=k-1/2 mn] = P_H[W_S-1/2 n(N+1)=1/2 mn-k]$$

esta última passagem pela simetria da distribuição de W_S em torno de $1/2 n(N+1)$.

Por outro lado, $W_S+W_R = 1/2 (m+n)(N+1) \Leftrightarrow W_S-1/2 n(N+1) - 1/2 mn = 1/2 m(m+1)-W_R$, logo $P_H[W_{XY}=k] = P_H[W_{YX}=k]$.

§3. Distribuição aproximada de W_S sob a hipótese.

Já vimos que sob a hipótese $F = G$ os postos (S_1, \dots, S_n) formam uma amostra aleatória da população $\pi = (1, 2, \dots, N)$. Vamos considerar agora um contexto mais geral que fornecerá resultados a serem utilizados posteriormente. Seja uma sequência de populações:

$$\pi_1 = \{v_{11}\}, \quad \pi_2 = \{v_{21}, v_{21}\}, \dots, \quad \pi_N = \{v_{N1}, v_{N2}, \dots, v_{NN}\}.$$

De cada população é retirada uma amostra de tamanho $n=n(N)$.

Designemos por A_{N1}, \dots, A_{Nn} os valores que constituem a amostra. Sejam $S_N = A_{N1} + \dots + A_{Nn}$ e $v_N = \frac{v_{N1} + \dots + v_{Nn}}{N}$.

Teorema 5 - Uma condição suficiente para que a variável padronizada $S_N^* = [S_N - E(S_N)] / \sqrt{\text{Var}(S_N)}$ tenha distribuição limite $N(0,1)$ é que n e $m = N-n$ tendam ambos para infinito e que

$$(2.6) \quad \frac{\max(v_{Ni} - v_N)^2}{\sum (v_{Nj} - v_N)^2} \max\left(\frac{N-n}{n}, \frac{n}{N-n}\right) \rightarrow 0 \quad \text{quando } N \rightarrow \infty.$$

Prova: Vide Lehmann pag. 353.

Aplicação do Teorema: Seja $\pi_N = \{1, 2, \dots, N\}$, isto é, $v_{N1} = 1, \dots, v_{Nn} = N$ para todo N . Neste caso:

$$v_N = \frac{1 + \dots + N}{N} = 1/2 (N+1)$$

$$\max(v_{Ni} - v_N)^2 = [N - 1/2(N+1)]^2 = \frac{(N-1)^2}{4}$$

$$\begin{aligned} \sum (v_{Ni} - v_N)^2 &= \sum v_{Ni}^2 - N v_N^2 = \sum_{i=1}^N i^2 - \frac{N(N+1)^2}{4} = \\ &= N(N^2 - 1)/12. \end{aligned}$$

A condição (2.6) fica,

$$\frac{(N-1)^2/4}{N(N^2-1)/12} \max\left(\frac{N-n}{n}, \frac{n}{N-n}\right) \rightarrow 0.$$

Suponhamos, sem perda de generalidade, $n > N-n$, então $\max\left(\frac{N-n}{n}, \frac{n}{N-n}\right) = \frac{n}{N-n}$ e a condição (2.6) pode ser escrita como segue:

$$\frac{3(N-1)}{(N+1)} \cdot \frac{n}{N} \cdot \frac{1}{N-n} \rightarrow 0 \text{ quando } N \rightarrow \infty.$$

Os dois primeiros fatores são limitados e $N-n$ tende para infinito, logo a condição é satisfeita. Esta aplicação prova que,

$$(2.7) \quad P_H \left[\frac{W_S - E_H(W_S)}{\sqrt{\text{Var}_H W_S}} \leq a \right] \rightarrow \Phi(a) \text{ quando } m \text{ e } n \text{ tendem para o infinito,}$$

onde $\Phi(a) = \int_{-\infty}^a \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ é a função de distribuição de uma $N(0,1)$ calculada no ponto a .

Este resultado é de grande utilidade, pois podemos a partir dele obter aproximações para $P_H(W_S \leq a)$ quando m e n são grandes. Para isto, precisamos calcular $E_H(W_S)$ e $\text{Var}(W_S)$.

§4. Determinação de $E_H(W_S)$ e $\text{Var}_H(W_S)$.

Vamos considerar uma situação mais geral. Seja uma população consistindo de N números v_1, \dots, v_N . Se retirarmos um número V ao acaso desta população, temos

$$E(V) = \frac{1}{N} \cdot v_1 + \frac{1}{N} \cdot v_2 + \dots + \frac{1}{N} \cdot v_N = \bar{v}$$

$$\text{Var}(V) = \frac{1}{N} \sum_{i=1}^N (v_i - \bar{v})^2 = \frac{1}{N} \sum_{i=1}^N v_i^2 - \bar{v}^2 = \tau^2$$

No caso particular em que $v_i = i$, obtemos

$$\bar{v} = \frac{1 + \dots + N}{N} = 1/2 (N+1)$$

e

$$\tau^2 = \frac{1}{N} \sum_{i=1}^N v_i^2 - \bar{v}^2 = \frac{1}{N} \sum i^2 - 1/4 (N+1)^2 = \frac{N^2 - 1}{12}.$$

Retiremos agora aleatoriamente n números V_1, \dots, V_n da população $\{v_1, \dots, v_N\}$. Seja $T = V_1 + \dots + V_n$, como os V_i 's têm a mesma distribuição:

$$(2.8) \quad ET = n\bar{v}.$$

$$\text{Var } T = \sum_{i=1}^n \text{Var}(V_i) + \sum_{i \neq j} \text{Cov}(V_i, V_j).$$

A variância comum dos V_i 's é τ^2 e os pares (V_i, V_j) têm a mesma covariância, pois a distribuição conjunta dos pares é a mesma: $P(V_i = v_i, V_j = v_j) = \frac{1}{N(N-1)}$, $\forall (i, j)$. Seja λ o valor desta covariância, então

$$\text{Var } T = n\tau^2 + N(N-1)\lambda.$$

Tomando agora $n=N$, obtemos $T = v_1 + \dots + v_N$, constante,

e

$$\text{Var } T = N\tau^2 + N(N-1)\lambda = 0 \Rightarrow \lambda = -\frac{\tau^2}{N-1}$$

Substituindo este valor,

$$(2.9) \quad \text{Var } T = \frac{n(N-n)}{N-1} \tau^2.$$

Para $v_i = i$, temos: $ET = 1/2 n(N+1)$ e

$$(2.10) \quad \text{Var } T = \frac{n(N-n)(N+1)}{12}.$$

A aproximação (2.7) pode ser, em geral, melhorada introduzindo-se uma "correção de continuidade" devido ao fato de estarmos aproximando uma distribuição discreta pela distribuição Normal. Com esta correção fica

$$(2.11) \quad P(W_S \leq c) \approx \Phi \left[\frac{c-1/2 - n(N+1)+1/2}{\sqrt{mn(N+1)/12}} \right].$$

Exemplo 3: Sejam $m = n = 7$.

O valor exato de $P[W_S \leq 59]$ é 0.8086 enquanto que o valor obtido, a partir da aproximação normal sem correção de continuidade, é .80; com a correção de continuidade é 0.81. A aproximação normal com correção de continuidade é razoável a partir de $\min(m,n) = 4$.

§5. Caso de Empates.

Uma das hipóteses feitas anteriormente foi a de não ocorrência de empates. Em dados reais estes surgem devido ao arredondamento de medidas ou quando trabalhamos com distribuições discretas. Outras vezes, os postos são atribuídos a partir de um julgamento e não através de medidas. Neste caso, pode não ser possível decidir qual o

melhor entre dois resultados. Podemos querer, apesar de ocorrência de empates, aplicar o teste de Wilcoxon.

Consideremos, por exemplo, o caso em $m=3$ e $n=2$ e suponhamos que os três menores valores das respostas sejam empatadas. Temos, então, os pontos 1,2,3 para serem atribuídos às três menores respostas. Como devemos fazer esta atribuição? Uma possibilidade que é razoável é atribuir a cada uma das 3 menores respostas o posto médio, no caso, igual a 2. A tabela abaixo indica as diversas configurações possíveis e a soma dos postos dos Y's correspondentes. Denotemos por W_S^* a soma dos postos médios dos tratamentos (dos Y's).

Configurações dos Y's

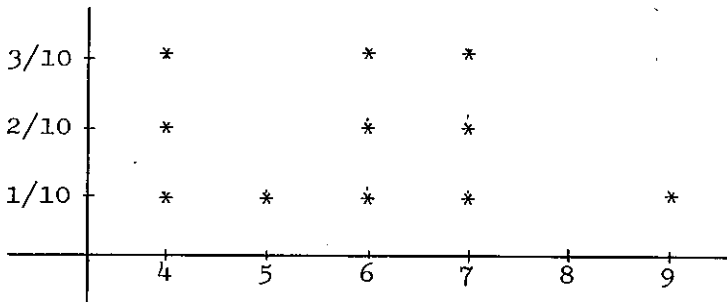
2 (1)	2 (2)	2 (3)	4	5	W_S^*
Y	Y				4 (3)
Y		Y			4 (4)
Y			Y		6 (5)
Y				Y	7 (6)
	Y	Y			4 (5)
	Y		Y		6 (6)
	Y			Y	7 (7)
		Y	Y		6 (7)
		Y		Y	7 (8)
			Y	Y	9 (9)

Obs.: Os postos circulados correspondem aos casos sem empates.

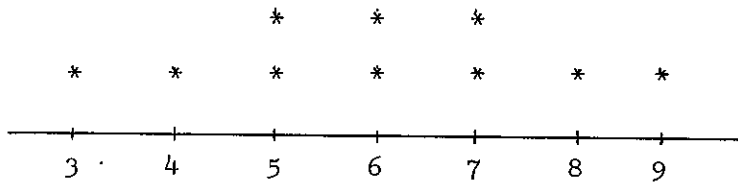
Distribuição de frequências de W_S^*

W^*	4	6	7	9
W_S^*	3/10	3/10	3/10	1/10

Caso de empates



Caso sem empates



No caso de empates, a distribuição de W_S^* concentra massa em menos valores, é menos dispersa e não é simétrica.

De modo geral, denotamos por $S_1^* < S_2^* < \dots < S_n^*$ os postos médios dos tratamentos e por $W_S^* = S_1^* + \dots + S_n^*$, a estatística de Wilcoxon. O teste de Wilcoxon é dado por: Rejeite a hipótese se

$$W_S^* \geq c$$

onde c é determinado a partir do nível de significância α do teste pela condição: $P_H(W_S^* \geq c) = \alpha$.

A distribuição nula de (S_1^*, \dots, S_n^*) não é mais a dada em (2.1), os valores tomados pelos S^* 's podem ser diferentes dos tomados pelos S 's. Apesar do número de escolhas dos n postos médios e das configurações dos Y 's continuarem sendo as mesmas, a distribuição da soma dos postos médios é alterada. A distribuição de W_S^* , sob a hipótese, depende também do número de observações empatadas em cada valor, ao contrário da de W_S que só depende de m e n .

Suponhamos que temos d_1 observações empatadas no menor valor, d_2 observações empatadas no 2º menor valor, etc. de empatadas no maior valor, com as observações tomando e valores distintos. Chamamos o vetor $(e; d_1, \dots, d_e)$ configuração de empates. Para uma dada configuração de empates, a distribuição de W_S^* pode ser determinada como no exemplo anterior, a partir das diferentes escolhas de posições para os Y 's. Entretanto, a distribuição de $(e; d_1, \dots, d_e)$ depende (sob a hipótese) de F o que faz com que a distribuição de W_S^* dependa também de F .

Ilustramos os pontos acima através do seguinte exemplo apresentados em Lehmann pag. 59:

Exemplo 5: Consideremos F discreta com massas p e q ,

respectivamente, nos pontos a e b , com $a < b$, e sejam $m=2$, $n=1$. Para este caso temos as seguintes configurações de empates:

X_2	X_2	Y_1	d_1	d_2	Prob.
b	b	b	0	3	q^3
a	b	b	1	2	$3pq^2$
a	a	b	2	1	$3p^2q$
a	a	a	3	0	p^3

$$P(e=1) = p^3 + q^3$$

$$P(e=2) = 3pq(p+q) = 3pq.$$

Por outro lado, para uma dada configuração de empates a distribuição de W_S^* independe de F :

$$P(W_S^* = 2 \mid d_1 = 0) = 1$$

$$P(W_S^* = 1 \mid d_1 = 1) = 1/3 \qquad P(W_S^* = 2,5 \mid d_1 = 1) = 2/3$$

$$P(W_S^* = 1,5 \mid d_1 = 2) = 2/3 \qquad P(W_S^* = 3 \mid d_1 = 2) = 1/3$$

$$P(W_S^* = 2 \mid d_1 = 3) = 1$$

Obtemos então:

w	1	1.5	2	2.5	3
$P(W_S^*=w)$	pq^2	$2p^2q$	p^3+q^3	$2pq^2$	p^2q

Neste exemplo, a distribuição dos postos médios (S_1^*, \dots, S_n^*) para uma dada configuração de empates independe de F .

Permanece assim o esquema de seleção de uma amostra aleatória de tamanho n da população de postos médios. Da mesma forma, para dado (d_1, \dots, d_e) , a distribuição de W_S^* independe de F . Por outro lado, vemos que as distribuições de $(e; d_1, \dots, d_e)$, (S_1^*, \dots, S_n^*) e W_S^* (não condicional) dependem de p e q e portanto de F . A propriedade de independer da distribuição subjacente é perdida pela distribuição de W_S^* . Podemos, contudo, neste caso, considerar o teste condicional dado a configuração de empates, com $C(e; d_1, \dots, d_e)$ determinado de modo que o nível de significância condicional seja o mais próximo possível de α :

$$P_H[W_S^* \geq c(e; d_1, \dots, d_e) | e; d_1, \dots, d_e] = \alpha,$$

e tomando-se então

$$W_S^* \geq C(e; d_1, \dots, d_e)$$

como região de rejeição. Tabelas para a distribuição de W_S^* dependeriam, então, da configuração de empates sendo, desta forma, pouco prática. Felizmente, a aproximação normal é válida, também, neste caso.

§6. Aproximação Normal para a distribuição de W_S^* .

Antes de provarmos a aproximação normal, vamos calcular $E_H W_S^*$ e $Var_H W_S^*$ para uma dada configuração $(e; d_1, \dots, d_e)$ (na realidade $E(W_S | (e; d_1, \dots, d_e))$ e $Var(W_S | (e; d_1, \dots, d_e))$).

Cálculo de $E_H W_S^*$ e $Var_H W_S^*$.

Sejam d_1, \dots, d_e tais que $d_1 + \dots + d_e = N$. Consideremos a população dos números $\{v_1, v_2, \dots, v_N\}$ onde

$$v_1 = \dots = v_{d_1} = \frac{1+2+\dots+d_1}{d_1} = \frac{1}{2}(d_1+1)$$

é o posto médio das d_1 observações empatadas no menor valor.

$$v_{d_1+1} = \dots = v_{d_1+d_2} = \frac{(d_1+1)+\dots+(d_1+d_2)}{d_2} = d_1 + \frac{1}{2}(d_2+1)$$

$$v_{d_1+d_2+1} = \dots = v_{d_1+d_2+d_3} = d_1 + d_2 + \frac{1}{2}(d_3+1) \text{ etc...}$$

Agora podemos aplicar a fórmula (2.8) $ET = n\bar{v}$, onde \bar{v} é a média dos v 's. Observe que a soma dos v 's, no primeiro grupo, é $1 + \dots + d_1$; no segundo $(d_1+1) + \dots + (d_1+d_2)$ etc. De modo que a soma total dos v 's é a soma dos inteiros $1+2 + \dots + N = \frac{N(N+1)}{2}$ e novamente $\bar{v} = \frac{N+1}{2}$. Para o cálculo da variância temos

$\text{Var } T = \frac{n(N-n)}{N-1} \tau^2$, pela fórmula (2.9), onde

$$\tau^2 = \frac{1}{N} \sum_{i=1}^N v_i^2 - \bar{v}^2.$$

Cálculo de τ^2 :

Vamos partir da identidade:

$$\sum_{i=1}^k a_i^2 = \sum_{i=1}^k (a_i - \bar{a})^2 + k\bar{a}^2.$$

Fazendo $a_1 = 1, \dots, a_{d_1} = d_1$, obtemos $\bar{a} = v_1 = \dots = v_{d_1}$,

$$\sum_{i=1}^{d_1} (a_i - \bar{a})^2 = \sum_{i=1}^{d_1} \left(i - \frac{d_1+1}{2}\right)^2 = \frac{d_1(d_1^2-1)}{12}, \text{ e}$$

$$\sum_{i=1}^{d_1} i^2 = \frac{d_1(d_1^2-1)}{12} + \sum_{i=1}^{d_1} v_i^2.$$

Aplicando a identidade para $a_1 = d_1+1, \dots, a_{d_2} = d_1+d_2$,

obtemos

$$\sum_{i=d_1+1}^{d_1+d_2} i^2 = \frac{d_2(d_2^2-1)}{12} + \sum_{i=d_1+1}^{d_1+d_2} v_i^2$$

assim por diante. Somando as expressões acima,

$$\begin{aligned} \sum_{i=1}^N v_i^2 &= \sum_{i=1}^N i^2 - \sum_{j=1}^e \frac{d_j(d_j^2-1)}{12} = \\ &= \frac{N(N+1)(2N+1)}{6} - \sum_{j=1}^e \frac{d_j(d_j^2-1)}{12} \end{aligned}$$

logo

$$\tau^2 = \frac{N^2-1}{12} - \frac{\sum_{i=1}^e d_i(d_i^2-1)}{12}$$

e

$$(2.12) \quad \text{Var } T = \frac{n(N-n)(N+1)}{12} - \frac{n(N-n) \sum_{i=1}^e d_i (d_i^2 - 1)}{12 N(N-1)}.$$

Para uma determinada configuração de empates $(e; d_1, \dots, d_e)$ e sob a hipótese $F = G$ temos o seguinte teorema de aproximação.

Teorema 6 - Sob as condições mencionadas acima

$$P \left[\frac{W_S^* - E(W_S^*)}{\sqrt{\text{Var } W_S^*}} \leq a \right] \rightarrow \Phi(a) \quad \text{quando } m \text{ e } n$$

tendem para infinito e existe $0 < \epsilon < 1$ tal que

$$\max_{i=1, \dots, e} \left(\frac{d_i}{N} \right) \leq 1 - \epsilon \quad \text{quando } N \rightarrow \infty.$$

Prova: Basta aplicar o Teorema 5 deste capítulo, tomando

$$v_{N1}, \dots, v_{NN} \quad \text{como sendo os postos médios, isto é,}$$

$$v_{Ni} = v_i. \quad \text{Temos } \max(v_{Ni} - v_N)^2 \leq \left[N - \frac{1}{2}(N+1) \right]^2 = 1/4 \binom{N-1}{4}^2,$$

$$\Sigma (v_{Nj} - v_N)^2 = \frac{N(N^2-1)}{12} - \frac{\Sigma (d_i^2 - d_i)}{12} = \frac{N^3}{12} - \frac{\Sigma d_i^3}{12}.$$

Por outro lado, $\frac{d_i}{N} \leq 1 - \epsilon \quad \forall i \Rightarrow d_i \leq (1 - \epsilon)N$ e

$$\Sigma d_i^3 = \Sigma d_i^2 \cdot d_i \leq (1 - \epsilon)^2 N^2 \Sigma d_i = (1 - \epsilon)^2 N^3 \leq (1 - \epsilon)N^3$$

então

$$\Sigma (v_{Nj} - v_{N1})^2 \geq \frac{N^3}{12} - \frac{(1 - \epsilon)N^3}{12} = \frac{N^3}{12} \epsilon.$$

Para verificar a condição do Teorema 5 basta mostrar que

$$\frac{1/4 \binom{N-1}{4}^2}{N^3 \epsilon} \max \left(\frac{N-n}{n}, \frac{n}{N-n} \right) \rightarrow 0 \quad \text{quando } N \rightarrow \infty.$$

Se $n > N-n$, $\max\left(\frac{N-n}{n}, \frac{n}{N-n}\right) = \frac{n}{N-n}$ e $\frac{3(N-1)^2}{\epsilon N^2} \cdot \frac{n}{N} \cdot \frac{1}{N-n} \rightarrow 0$
 quando $N \rightarrow \infty$ pois $N-n \rightarrow \infty$.

Observação: É importante observar que o teorema de aproximação foi demonstrado para o caso $F = G$ e para uma configuração de empates dada. Poderia então, ser aplicado para determinar a constante $C(\epsilon; d_1, \dots, d_e)$ do teste condicional mencionado acima. Demonstra-se, contudo, que a distribuição (não condicional) de

$\frac{W_S^* - EW_S^*}{\sqrt{\text{Var } W_S^*}} \rightarrow N(0,1)$. Isto mostra que no limite a constante crítica $C(\epsilon; d_1, \dots, d_e)$ define um teste incondicional de nível α .

§7. Potência do Teste de Wilcoxon.

Para estudar a potência do teste de Wilcoxon vamos considerar o modelo de deslocamento que supõe $G(x) = F(x-\Delta) \quad \forall x$ com $\Delta > 0$. A função potência do teste é dada por $\Pi_F(\Delta) = P_{\Delta}(W_{XY} \geq c)$, onde o símbolo P_{Δ} indica que a distribuição dos Y 's é deslocada de Δ em relação à distribuição dos X 's. Vamos estudar esta função para $-\infty < \Delta < +\infty$. O caso em que $\Delta > 0$ corresponde a respostas maiores do tratamento, para $\Delta = 0$ temos o ca-

so de nenhum efeito de tratamento (hipótese) e $\Pi_F(0)$ é o nível de significância do teste. Para $\Delta < 0$, temos o caso do tratamento decrescer as respostas. No contexto acima, estamos interessados em testar $H: \Delta = 0$ contra $K: \Delta > 0$. Para um nível de significância α dado determinamos $c = c(\alpha; m, n)$. O bom desempenho do teste é caracterizado por valores grandes de $\Pi_P(\Delta)$ quando $\Delta > 0$ (comparada com a potência, para o mesmo Δ , de outros testes de nível α). Independente da comparação com outros testes, é razoável exigir-se que a capacidade de detectar o efeito do tratamento aumente quando Δ cresce.

Teorema 7 - A função potência do teste de Wilcoxon,

$\Pi_F(\Delta) = P_{\Delta}(W_{XY} \geq c)$, é uma função não decrescente de Δ .

Prova: Seja $\Delta_0 < \Delta_1$. Queremos mostrar que $P_{\Delta_0}(W_{XY} \geq c) \leq P_{\Delta_1}(W_{XY} \geq c)$. Observando que $P_{\Delta}(Y_i \leq y) = F(y - \Delta)$, então,

$$\begin{aligned} P_{\Delta_1}(Y_i \leq y) &= F(y - \Delta_1) = F[y - (\Delta_1 - \Delta_0) - \Delta_0] = P_{\Delta_0}(Y_i \leq y - (\Delta_1 - \Delta_0)) \\ &= P_{\Delta_0}(Y_i + (\Delta_1 - \Delta_0) \leq y). \end{aligned}$$

Esta igualdade mostra que $Y_i + (\Delta_1 - \Delta_0)$, sob a alternativa Δ_0 , tem a mesma distribuição que Y_i , sob a alternativa Δ_1 . Defina $V_i = Y_i + (\Delta_1 - \Delta_0)$, então,

$$P_{\Delta_1}(W_{XY} \geq c) = P_{\Delta_0}(W_{XV} \geq c),$$

onde W_{XY} denota o número de pares (X_i, Y_j) tal que $X_i < Y_j$; W_{XV} o número de pares (X_i, V_j) com $X_i < V_j$. Como $V_j \geq Y_j \forall j$, então, $W_{XY} \leq W_{XV}$ e portanto $P_{\Delta_0}(W_{XY} \geq c) \leq P_{\Delta_0}(W_{XV} \geq c)$.

Consequências do Teorema 7:

1. $\Pi_F(\Delta) \geq \alpha$, $\forall F$ contínua e $\Delta > 0$, pois $\Pi_F(0) = \alpha$.

Portanto, o teste de Wilcoxon é não viciado contra alternativas de deslocamento.

2. Considere o problema de testar $H': \Delta \leq 0$ contra

$K: \Delta > 0$. Pelo Teorema 7, o teste de Wilcoxon com $\Pi_F(0) = \alpha$ satisfaz $\sup_{\Delta \geq 0} \Pi_F(\Delta) = \alpha$ e, portanto, tem nível α para testar H' .

§8. Cálculo de Potência do Teste de Wilcoxon.

A aproximação normal, para a distribuição de W_{XY} , válida quando $F = G$, pode ser demonstrada sob certas hipóteses quando $F \neq G$. Neste caso sendo impraticável a utilização de tabelas, esta aproximação se faz ainda mais necessária.

Vamos estudar o problema da aproximação normal para a distribuição de W_{XY} em um contexto mais geral.

Observe que podemos escrever $W_{XY} = \sum_i \sum_j \varphi(X_i, Y_j)$, onde

$$\varphi(x, y) = \begin{cases} 1 & \text{se } x < y \\ 0 & \text{caso contrário} \end{cases}$$

Consideremos, agora, uma classe maior de estatísticas tomando $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}$, qualquer.

Definição - Sejam $X_1, \dots, X_m; Y_1, \dots, Y_n$ independentes, tendo os X 's a mesma distribuição F , e os Y 's a distribuição comum G . Então $U = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \varphi(X_i, Y_j)$, com $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}$, é chamada estatística U .

A demonstração da normalidade assintótica de U é dificultada pela dependência de suas parcelas. A idéia da prova é "aproximar" U por uma soma de variáveis aleatórias independentes do tipo $S = \sum a_i(X_i) + \sum b_j(Y_j)$; onde S é assintoticamente equivalente à estatística U . A escolha adequada das funções a_i e b_j é feita utilizando o método das projeções de Hajek. Para maiores detalhes vide Lehmann pag. 362. A seguir, enunciamos o seguinte teorema de aproximação:

Teorema 8 - A distribuição de $\frac{W_{XY} - E(W_{XY})}{\sqrt{\text{Var}(W_{XY})}}$ tende para a distribuição normal padrão quando m e n tendem para infinito, e quando as distribuições F e G são tais que $0 < P(X < Y) < 1$.

Prova: Vide Lehmann, pag. 365, exemplo 20.

Nas aplicações deste teorema, precisamos calcular $E(W_{XY})$ e $\text{Var}(W_{XY})$ sob a alternativa, isto é, quando $F \neq G$.

Cálculo de $E(W_{XY})$ e $\text{Var}(W_{XY})$ quando $F \neq G$.

Podemos escrever $W_{XY} = \sum_{i=1}^m \sum_{j=1}^n \varphi(X_i, Y_j)$, onde

$\varphi(x, y) = 1$ se $x < y$ e $\varphi(x, y) = 0$, caso contrário.

Como os pares (X_i, Y_j) têm a mesma distribuição conjunta, as parcelas $\varphi(X_i, Y_j)$ têm a mesma esperança, e

$$(2.13) \quad EW_{XY} = mn E\varphi(X, Y) = mn P[X < Y] = mn p_1,$$

onde $p_1 = P[X < Y]$. Para a variância, temos

$$(2.14) \quad \text{Var } W_{XY} = \sum_{i=1}^m \sum_{j=1}^n \text{Var } \varphi(X_i, Y_j) + \\ + \sum_i \sum_j \sum_k \sum_l \underset{(i, j) \neq (k, l)}{\text{cov}[\varphi(X_i, Y_j), \varphi(X_k, Y_l)]}.$$

O primeiro termo, fácil de calcular pois $\text{Var } \varphi(X_i, Y_j)$ in depende de (i, j) , é igual a $mn \text{Var } \varphi(X, Y)$, portanto,

$$(2.15) \quad \text{Var } \varphi(X, Y) = p_1(1-p_1).$$

Consideremos, agora, os vários tipos de termos que entram no cálculo da soma das covariâncias. Primeiro, se $i \neq k$ e $j \neq l$ então as variáveis aleatórias $\varphi(X_i, Y_j)$ e $\varphi(X_k, Y_l)$ são independentes e conseqüentemente $\text{cov}[\varphi(X_i, Y_j), \varphi(X_k, Y_l)] = 0$. Consideremos agora o caso

em que $i \neq k$ e $j = l$, que fornece termos do tipo $\text{cov}[\varphi(X_i, Y_j), \varphi(X_k, Y_j)]$. O número de termos deste tipo é $m(m-1)n$. O outro caso possível é quando $i = k$ e $j \neq l$, ou seja, termos do tipo $\text{cov}[\varphi(X_i, Y_j), \varphi(X_i, Y_l)]$ em número de $n(n-1)m$. Por outro lado

$$(2.16) \quad \text{cov}[\varphi(X_i, Y_j), \varphi(X_i, Y_l)] = P[\varphi(X_i, Y_j) = 1, \varphi(X_i, Y_l) = 1] - p_1^2 = P[X_i < Y_j \text{ e } X_i < Y_l] - p_1^2 = p_2 - p_1^2,$$

onde $p_2 = P[X_i < Y_j \text{ e } X_i < Y_l]$. Da mesma forma,

$$(2.17) \quad \text{cov}[\varphi(X_i, Y_j), \varphi(X_k, Y_j)] = P[X_i < Y_j \text{ e } X_k < Y_j] - p_1^2 = p_3 - p_1^2,$$

onde $p_3 = P[X_i < X_j \text{ e } X_k < Y_j]$. Substituindo estes valores em (2.14) obtemos:

$$(2.18) \quad \text{Var}(W_{XY}) = mn p_1 (1 - p_1) + n(n-1)m(p_2 - p_1^2) + m(m-1)n(p_3 - p_1^2).$$

Exemplo 6: Calcular $E(W_{XY})$ e $\text{Var}(W_{XY})$ quando

$$F = N(\mu, \sigma^2) \text{ e } G = N(\mu + \Delta, \sigma^2).$$

Basta aplicar as fórmulas (2.13) e (2.18). Para isto precisamos calcular p_1 , p_2 e p_3 .

$$p_1 = P(X < Y) = P(X - Y < 0) = P\left[\frac{X - (Y - \Delta)}{\sigma\sqrt{2}} < \frac{\Delta}{\sigma\sqrt{2}}\right].$$

Como X e Y são normais independentes, $X - (Y - \Delta)$ tem

distribuição normal com média $E[X-(Y-\Delta)] = 0$ e variância $2\sigma^2$, $\frac{X-(Y-\Delta)}{\sigma\sqrt{2}}$ tem distribuição normal padrão.

Logo, $p_1 = \Phi\left(\frac{\Delta}{\sigma\sqrt{2}}\right)$, valor que podemos obter utilizando uma tabela da distribuição normal padrão.

Para p_2 e p_3 temos

$$p_2 = P[X < Y \text{ e } X < Y'] = P\left[\frac{X-(Y-\Delta)}{\sigma\sqrt{2}} < \frac{\Delta}{\sigma\sqrt{2}} \text{ e } \frac{X-(Y'-\Delta)}{\sigma\sqrt{2}} < \frac{\Delta}{\sigma\sqrt{2}}\right]$$

com X , Y e Y' independentes. As variáveis

$Z = \frac{X-(Y-\Delta)}{\sigma\sqrt{2}}$ e $Z' = \frac{X-(Y'-\Delta)}{\sigma\sqrt{2}}$ são normais padrões, não

sendo, contudo, independentes. O coeficiente de correlação entre eles é

$$\rho(Z, Z') = \frac{\text{Cov}(Z, Z')}{\sqrt{\text{Var } Z} \sqrt{\text{Var } Z'}} = \text{EZZ}' ,$$

onde

$$\text{EZZ}' = \frac{1}{2\sigma^2} E[X^2 - X(Y'-\Delta) - X(Y-\Delta) + (Y-\Delta)(Y'-\Delta)] = \frac{EX^2}{2\sigma^2} = \frac{1}{2} .$$

Para a determinação de p_2 e p_3 precisamos, então, utilizar uma tabela de Normal Bivariada com $\rho = 1/2$.

Exemplo 7: Calcular $E'W_{XY}$ e $\text{Var } W_{XY}$ quando

$$F = U[\theta - 1/2, \theta + 1/2] \text{ e } G = U[\theta + \Delta - 1/2, \theta + \Delta + 1/2].$$

Basta considerar o caso em que $0 < \Delta < 1$. Para $\Delta > 1$ os Y 's são maiores que os X 's com probabilidade 1 e

$W_{XY} = mn$ com probabilidade 1.

Designando por I_A a função dada por $I_A(x) = 1$

se $x \in A$; $I_A(x) = 0$ se $x \in A^c$, temos

$$p_1 = P[X < Y] = \int_{-\infty}^{\infty} \frac{I(x)}{[\theta - 1/2, \theta + 1/2]} dx \int_x^{+\infty} \frac{I(y)}{[\theta + \Delta - 1/2, \theta + \Delta + 1/2]} dy$$

Por outro lado,

$$\int_x^{+\infty} \frac{I(y)}{[\theta + \Delta - 1/2, \theta + \Delta + 1/2]} dy = \begin{cases} 1 & \text{se } x \leq \theta + \Delta - 1/2 \\ (\theta + \Delta + 1/2) - x & \text{se } \theta + \Delta - 1/2 < x < \theta + \Delta + 1/2 \\ 0 & \text{se } x \geq \theta + \Delta + 1/2 \end{cases}$$

Substituindo estes valores temos

$$p_1 = \int_{-\infty}^{\theta + \Delta - 1/2} \frac{I(x)}{[\theta - 1/2, \theta + 1/2]} dx + \int_{\theta + \Delta - 1/2}^{\theta + \Delta + 1/2} (\theta + \Delta + 1/2 - x) \frac{I(x)}{[\theta - 1/2, \theta + 1/2]} dx =$$

$$= \Delta + 1/2 (1 - \Delta^2) = \Delta + 1/2 - \Delta^2/2.$$

Para p_2 , temos

$$p_2 = P[X < Y \text{ e } X < Y'] =$$

$$= \int_{-\infty}^{+\infty} \frac{I(x)}{[\theta - 1/2, \theta + 1/2]} dx \int_x^{+\infty} \frac{I(y)}{[\theta - 1/2 + \Delta, \theta + 1/2 + \Delta]} dy \int_x^{+\infty} \frac{I(y')}{[\theta - 1/2 + \Delta, \theta + 1/2 + \Delta]} dy'$$

$$= \int_{-\infty}^{+\infty} \frac{I(x)}{[\theta - 1/2, \theta + 1/2]} dx \left(\int_x^{+\infty} \frac{I(y)}{[\theta - 1/2 + \Delta, \theta + 1/2 + \Delta]} dy \right)^2 =$$

$$= \int_{-\infty}^{\theta+\Delta-1/2} \frac{I(x)}{[\theta-1/2, \theta+1/2]} dx + \int_{\theta+\Delta-1/2}^{\theta+\Delta+1/2} \frac{(\theta+\Delta+1/2-x)^2 I(x)}{[\theta-1/2, \theta+1/2]} dx$$
$$= \Delta - \left(\frac{\Delta^3}{3} + \frac{1}{3}\right) = \frac{1}{3} (1 + 3\Delta - \Delta^3).$$

A distribuição F é, no caso, simétrica e $p_3 = p_2$. A partir dos valores de p_1 e p_2 , podemos calcular $E W_{XY}$ e $\text{Var } W_{XY}$ utilizando as fórmulas (2.13) e (2.18).

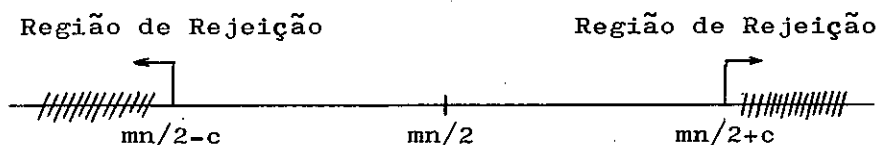
§9. Estimação do Efeito do Tratamento.

Em muitas situações, apenas testar a hipótese de nulidade do efeito não é satisfatório. Quando a hipótese é rejeitada, podemos desejar, ainda, estimar o efeito do tratamento. Quando o modelo utilizado é o de deslocamento, um parâmetro que quantifica, de modo bem natural este efeito, é o deslocamento Δ provocado pelo tratamento na distribuição das respostas.

Nesta secção vamos abordar o problema de estimação de Δ no modelo de deslocamento. É bastante conhecida a relação existente entre os problemas estatísticos de teste de Hipóteses e Estimação por Intervalos de Confiança. É possível, a partir da região de aceitação de

um teste, obtermos uma região de confiança para o parâmetro sendo testado e vice-versa. Neste caso, existe uma relação entre o nível de significância do teste e o nível de confiança da região. Menos conhecida é, porém, a obtenção de estimadores a partir de testes. Este "processo de inversão" aplicado ao teste de Wilcoxon fornece o estimador de Hodges-Lehmann. Consideremos a hipótese H de nenhum efeito ($\Delta=0$) contra a alternativa bilateral de efeito não nulo ($\Delta\neq 0$). A extensão natural do teste de Wilcoxon, que apresentamos, rejeita H quando W_{XY} é muito grande ou muito pequeno.

A distribuição de W_{XY} , sob a hipótese H , é simétrica em torno de $mn/2$. O teste bilateral de Wilcoxon (Mann-Whitney) rejeita, então, a hipótese quando $W_{XY} \geq mn/2 + c$ ou $W_{XY} \leq mn/2 - c$, sendo c determinado a partir do nível de significância.



Os valores de W_{XY} que apoiam mais fortemente a aceitação da hipótese são os mais próximos de $mn/2$. Em particular, $W_{XY} = mn/2$ é o valor que dá maior indicação favorável à aceitação de $F = G$.

Queremos estimar o valor Δ de modo que X_1, \dots, X_m e $Y_1 - \Delta, \dots, Y_n - \Delta$ tenham a mesma distribuição. O estimador natural é, portanto, $\hat{\Delta}$ tal que $W_{X, Y - \hat{\Delta}} = mn/2$. Por outro lado, $W_{X, Y - \hat{\Delta}}$ é o número de pares (X_i, Y_j) tal que $X_i < Y_j - \hat{\Delta}$ ou, equivalentemente, que $Y_j - X_i > \hat{\Delta}$. Devemos, então, determinar $\hat{\Delta}$ de modo que para a metade dos pares (X_i, Y_j) as diferenças $Y_j - X_i$ sejam menores e para a outra metade as diferenças sejam maiores que ele. Isto é obtido tomando-se $\hat{\Delta} = \text{mediana}\{(Y_j - X_i); i=1, \dots, m; j=1, \dots, n\}$. Designando-se por $D_{(1)} < D_{(2)} < \dots < D_{(mn)}$ as diferenças $Y_j - X_i$ ordenadas: $\hat{\Delta} = 1/2 (D_k + D_{(k+1)})$ para o caso em que $mn = 2k$ (par) e $\hat{\Delta} = D_{(k+1)}$ para o caso em que $mn = 2k+1$ (ímpar).

Se m e n são grandes, a determinação de $\hat{\Delta}$ se torna muito trabalhosa, envolve a ordenação de mn diferenças $(Y_j - X_i)$. Pode-se, contudo, simplificar esta tarefa utilizando-se o método gráfico apresentado em Wolff e Hollander pag. 80.

Propriedades do estimador $\hat{\Delta}$.

Em geral, as propriedades mais relevantes de um estimador são relativas à centralização e dispersão de sua distribuição em torno do parâmetro estimado. É desejável que o estimador tenha distribuição bem centrada e

pouco dispersa em torno do parâmetro estimado. Quando uma distribuição é simétrica em torno de algum valor, este é, naturalmente, escolhido como centro da distribuição. Em outros casos pode-se tomar como centro o valor esperado ou a mediana da distribuição. Estes valores coincidem com o centro de simetria no caso da distribuição simétrica. Demonstra-se que o estimador $\hat{\Delta}$ tem distribuição simétrica em torno de Δ quando a distribuição F é simétrica em torno de algum ponto. A distribuição de $\hat{\Delta}$ é, também, simétrica em torno de Δ quando $m = n$. Nos casos acima, $E(\hat{\Delta}) = \Delta$ e o estimador $\hat{\Delta}$ é não viciado. Além disto, Δ é a mediana da distribuição de $\hat{\Delta}$, isto é, $P(\hat{\Delta} < \Delta) = P(\hat{\Delta} > \Delta) = 1/2$, e $\hat{\Delta}$ tem a mesma probabilidade de subestimar que a de superestimar Δ . Mais ainda, esta propriedade é válida sempre que mn é ímpar e é aproximadamente satisfeita quando mn é par (Vide Lehmann pag. 86-87).

A dispersão de uma distribuição é usualmente medida pela sua variância. No caso da distribuição de $\hat{\Delta}$, é mais simples utilizar-se uma outra medida. Adota-se neste caso $P(|\hat{\Delta} - \Delta| \leq a)$, isto é, a probabilidade de que o estimador difira do valor verdadeiro Δ de menos que uma quantidade dada a . É desejável que esta probabilidade seja grande, indicando que o estimador tem alta probabi-

lidade de estar próximo do valor estimado.

O estimador clássico de Δ é a diferença $\bar{\Delta} = \bar{Y} - \bar{X}$. Quando F e G são normais este é o estimador não viciado de mínima variância. Uma comparação entre os estimadores $\hat{\Delta}$ e $\bar{\Delta}$ pode ser feita como segue.

Seja $n_{\Delta}^{\wedge}(\pi, a)$ o número de observações requeridas pelo estimador $\hat{\Delta}$ a fim de que $P(|\hat{\Delta} - \Delta| \leq a) = \pi$ e $n_{\Delta}^{-}(\pi, a)$ o número de observações de modo que $P(|\bar{\Delta} - \Delta| \leq a) = \pi$. Quando $a \rightarrow 0$ então $n_{\Delta}^{\wedge}(\pi, a) \rightarrow \infty$ e $n_{\Delta}^{-}(\pi, a) \rightarrow \infty$ e a relação $\frac{n_{\Delta}^{-}}{n_{\Delta}^{\wedge}} \rightarrow 3/\pi = 0.95$. (Vide Lehmann pag. 89). Isto mostra que o estimador $\hat{\Delta}$ perde muito pouca "eficiência" em relação a $\bar{\Delta}$, mesmo no caso mais favorável a este último. Para outros casos, em geral $\hat{\Delta}$ é mais eficiente que $\bar{\Delta}$, aumentando esta eficiência na medida em que as caudas de F são mais "grossas".

O exemplo seguinte ilustra como $\bar{\Delta}$ é muito mais sensível que $\hat{\Delta}$ ao aparecimento de valores excêntricos (outliers).

Exemplo 8: Sejam as respostas do tratamento: 5,4; 1,5; 8,6; 6,1; 0,5; 4,1; 2,8; 1,8; 4,1; 8,3 e as respostas dos controles: 6,6; 1,9; 8,2; 1,1; 1,1; 8,8. Obtemos então: $\bar{\Delta} = \bar{Y} - \bar{X} = 4,32 - 4,63 = -0,31$, enquanto que $\hat{\Delta} = \frac{1}{2} [D_{(30)} + D_{(31)}] = -0.15$. Suponhamos agora que quando se anotou a última resposta do tratamento a vírgula foi esquecida, ficando 83 no lugar de 8,3. Neste caso

obteríamos: $\bar{\Delta} = 7.16$ e $\hat{\Delta} = 0$.

Um estimador cujo valor varia pouco com o aparecimento de valores excêntricos (outliers) é dito robusto.

Pode-se mostrar, ainda, que o estimador $\hat{\Delta}$ é consistente, isto é, $\forall \epsilon > 0$, $P(|\hat{\Delta}_{m,n} - \Delta| < \epsilon) \rightarrow 1$ quando $m, n \rightarrow \infty$. Aqui, $\hat{\Delta}_{m,n}$ representa o estimador de Hodges-Lehmann baseado em m resposta dos controles e n dos tratamentos.

No caso geral do problema de duas amostras, sem a utilização do modelo de deslocamento, podemos tomar $p_1 = P(X < Y)$ como uma medida da tendência do tratamento de aumentar as respostas. Sabemos que $E(W_{XY}) = mnp_1$, desta forma W_{XY}/mn é um estimador não viciado de p_1 . É fácil mostrar que W_{XY}/mn é um estimador consistente de $P(X < Y)$. Basta utilizar a desigualdade de Chebyshev:

$$P(|W_{XY}/mn - p_1| \geq \epsilon) \leq \frac{1}{\epsilon^2} \text{Var}(W_{XY}/mn),$$

por (2.18) sabemos que

$$\text{Var}(W_{XY}/mn) = \frac{p_1(1-p_1)}{mn} + \frac{(n-1)}{mn} (p_2 - p_1^2) + \frac{(m-1)}{mn} (p_3 - p_1^2).$$

Quando $m, n \rightarrow \infty$, $\text{Var}(W_{XY}/mn) \rightarrow 0$ o que acarreta a consistência do estimador.

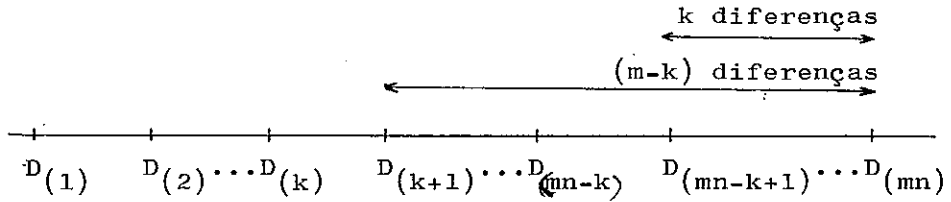
§10. Intervalos de Confiança para o efeito Δ de tratamento.

Vamos considerar, inicialmente, o problema de testar a hipótese $H: \Delta = \Delta_0$ contra a alternativa $K: \Delta \neq \Delta_0$. Tomando-se as variáveis X_1, \dots, X_m ; $Y'_1 = Y_1 - \Delta_0$, $Y'_2 = Y_2 - \Delta_0, \dots, Y'_n = Y_n - \Delta_0$, este problema é equivalente ao de testar $H_0: \Delta = 0$ contra $K: \Delta \neq 0$. A região de aceitação do teste de Mann-Whitney para a hipótese acima é dada por $mn/2 - c < W_{XY'} < mn/2 + c$ onde c é determinado por $P_{\Delta_0}(mn/2 - c < W_{X, Y - \Delta_0} < mn/2 + c) = 1 - \alpha$. Como $W_{X, Y - \Delta_0}$ só toma valores inteiros, determinamos k tal que $P_{\Delta_0}(k < W_{X, Y - \Delta_0} < mn - k) \approx 1 - \alpha$. É importante observar que:

$$P_{\Delta_0}(k < W_{X, Y - \Delta_0} < mn - k) = P_0(k < W_{X, Y} < mn - k),$$

sendo portanto k e $mn - k$, determinados a partir da distribuição de $W_{X, Y}$ sob a hipótese de nenhum efeito, que independe de F (contínua).

Vamos, agora, achar os valores de Δ_0 , que satisfazem a condição $k < W_{X, Y - \Delta_0} < mn - k$. Para isto consideremos as mn diferenças $Y_j - X_i$ ordenadas crescentemente: $D_{(1)} < D_{(2)} < \dots < D_{(mn)}$. A condição $W_{X, Y - \Delta_0} > k$ equivale a dizer que:



$\#\{(i, j); Y_j - X_i > \Delta_o\} > k$ ou que $\Delta_o < D_{mn-k}$.

Por outro lado $W_{X, Y - \Delta_o} < mn-k \Leftrightarrow \#\{(i, j); Y_j - X_i > \Delta_o\} < m-k$
 $\Leftrightarrow \Delta_o \geq D_{(k+1)}$. Portanto temos:

$$(2.19) \quad P_{\Delta_o} (D_{(k+1)} \leq \Delta_o \leq D_{mn-k}) = 1-\alpha \quad \forall \Delta_o \text{ e } \forall F \text{ cont nua}$$

e $[D_{k+1}, D_{mn-k}]$   um intervalo de confian a para Δ com n vel de confian a $(1-\alpha)$.

Exemplo 9: Para os dados do exemplo 8, ache um intervalo de confian a com n vel de confian a de 0.90.

Basta determinar k tal que $P_o(W_{XY} \leq k) = 0.05$. O valor mais aproximado   $k = 14$ com $P_o(W_{XY} \leq 14) = 0.0467$. Desta forma, (D_{15}, D_{45})   um intervalo de confian a com n vel de confian a de $1-2 \times 0,0467 = 0.907$. A partir dos dados obtemos: $D_{15} = -4,1$ e $D_{45} = 3,0$.

§11. Teste de Escores Normais.

O teste clássico utilizado no problema de duas amostras, modelo de deslocamento, com F normal, é o teste t -Student. Este teste é, nesta situação, mais poderoso que o teste de Wilcoxon. Contudo na maioria dos problemas de aplicação a hipótese de normalidade não é justificável, sendo prática corrente a utilização do teste t -Student em casos onde a distribuição subjacente é muito diferente da normal. Para distribuições com caudas grossas, (densidade $f(x)$ tende para zero quando $x \rightarrow \pm\infty$ muito mais devagar que na normal) o teste t -student tem desempenho bastante fraco. Uma medida comparativa deste desempenho, para dois testes alternativos de uma mesma hipótese, é dada pela eficiência relativa assintótica. A seguinte definição de eficiência relativa assintótica é também chamada eficiência de Pitman de um teste em relação a outro. Consideremos, por exemplo, os testes t -Student e o teste de Wilcoxon para testar a hipótese $H: \Delta = 0$ no problema de duas amostras, modelo de deslocamento. Para ambos os testes tomemos o mesmo nível de significância. Sejam $n_w(\Delta_0)$ e $n_t(\Delta_0)$ os tamanhos de amostras requeridas, respectivamente, pelo teste de Wilcoxon e pelo teste t -Student, de mesmo nível de signi-

ficância α , para alcançar a mesma potência $\alpha < \pi < 1$ contra a alternativa Δ_0 . A eficiência de Pitman do teste-t em relação ao teste de Wilcoxon é dada por:

$$\lim_{\Delta_0 \rightarrow 0} \frac{n_t(\Delta_0)}{n_w(\Delta_0)}$$
. Em muitos casos, este limite existe e é independente de α e de π . Denotaremos esta eficiência por $e_{w,t}(F)$, indicando sua dependência da distribuição subjacente F . Mostra-se que quando F é normal esta eficiência é $3/\pi = 0.955$. Apesar deste caso ser favorável ao teste-t, a eficiência perdida pelo teste de Wilcoxon é baixa, cerca de 5%. Para outros casos importantes a comparação é bastante favorável ao teste de Wilcoxon. Por exemplo, para F exponencial, a eficiência de Pitman é igual a 3.

O teste de Wilcoxon tem ainda a importante propriedade de ter nível de significância independente de F (contínua). Esta propriedade é preservada por qualquer teste baseado apenas nos postos das observações. Tendo em vista o bom desempenho do teste-t no caso normal, é natural tentar um teste baseado nos postos e que imite o teste-t. Designemos, como antes, por $r_1 < r_2 < \dots < r_m$ e $s_1 < s_2 < \dots < s_n$, respectivamente, os postos dos controles e dos tratamentos na amostra conjunta dos X 's e dos Y 's. Vamos a partir do conhecimento destes postos tentar predizer as observações. Sendo F normal e sob a hipóte

se H , as observações X 's e as Y 's formam uma amostra aleatória de tamanho $N = m+n$ de uma $N(0,1)$. Designemos esta amostra conjunta por Z_1, \dots, Z_N . Conhecendo-se o posto s de uma observação, nesta amostra, como poderíamos predizê-la? A observação de posto s é a estatística de ordem $Z^{(s)}$. Um preditor desta é $a_N(s) = E_{\phi}(Z^{(s)})$, valor esperado da s -ésima estatística de ordem de uma amostra de tamanho N da $N(0,1)$. Os valores $a_N(1), \dots, a_N(N)$ são denominados escores normais e tabela dos largamente (V.Owen). Desta forma, as observações Y 's são reconstruídas por $a_N(s_1), \dots, a_N(s_n)$ e as observações X 's por $a_N(r_1), \dots, a_N(r_m)$. Vamos substituir estas observações reconstruídas na expressão do teste- t . Este teste pode ser escrito na forma:

$$(2.20) \quad (\bar{Y} - \bar{X}) / \sqrt{\frac{N}{\sum_{i=1}^N (z_i - \bar{Z})^2}} \geq c.$$

Temos, $\bar{Z} = \sum_{i=1}^N a_N(i) / N$. Vamos mostrar que $\sum_{i=1}^N a_N(i) = 0$.

Para isto, basta mostrar que $a_N(s) = -a_N(N+1-s)$. Faça $z'_i = -z_i$ e seja $a'_N(s) = E(z'^{(s)})$. Como (z'_1, \dots, z'_N) é ainda uma amostra aleatória de tamanho N de uma $N(0,1)$, $a'_N(s) = a_N(s)$. Por outro lado, $Z'(s) = -Z^{(N+1-s)}$ e $a'_N(s) = a_N(s) = E(Z'(s)) = -E(Z^{(N+1-s)}) = -a_N(N+1-s)$, acarretando $\sum_{i=1}^N a_N(s) = 0$.

Substituindo este valor em (2.20) o denominador fica

$\sum_{i=1}^N a_N^2(i)$ que só depende de N , portanto constante.

A expressão do teste fica reduzida a $\frac{1}{n} \sum_{j=1}^N a_N(s_j) - \frac{1}{m} \sum_{i=1}^m a_N(s_i) \geq c'$, levando em conta que $\sum_{i=1}^N a_N(i) = 0$, o teste fica:

$$(2.21) \quad \sum a_N(s_j) \geq c.$$

Este é o chamado teste de Escores Normais.

Propriedades do teste de Escores Normais:

1. A potência do teste é uma função não decrescente de Δ .
Desta forma, o teste é não viciado e tem nível α para testar $H': \Delta \leq 0$.
2. A eficiência de Pitman do teste de Escores Normais em relação ao teste-t é no caso Normal igual a 1, e em outros casos nunca inferior a 1.

Teste de Van Der Warden

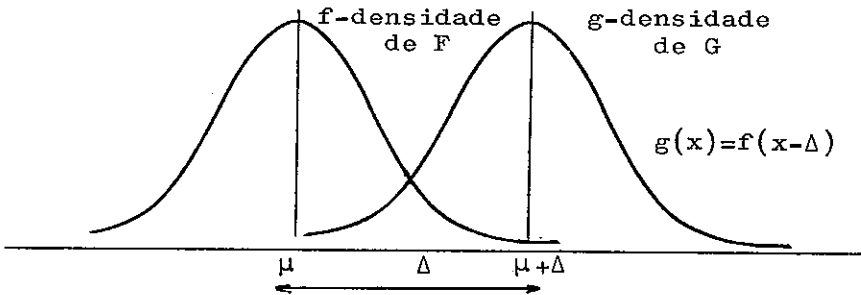
Apesar do excelente desempenho do teste de Escores Normais quando F é normal, a disponibilidade de uma tabela de Escores Normais limita sua utilização. Para contornar esta dificuldade é introduzida uma aproximação para os escores normais. Sendo $Z^{(1)}, \dots, Z^{(N)}$ as estatísticas de ordem definidas na secção anterior, os esco-

res normais são dados por $a_N(i) = E_{\Phi} Z^{(i)}$, onde Φ é a função de distribuição da normal padrão. É fato conhecido que se X tem uma distribuição F , contínua, então, $Y = F(X)$ tem distribuição uniforme no intervalo $[0,1]$. Consequentemente, se Y tem distribuição uniforme, $F^{-1}(Y)$ tem distribuição F . Por outro lado, F^{-1} preserva a ordem das observações: se $U^{(i)}$ é a i -ésima estatística de ordem de uma distribuição uniforme em $[0,1]$, $\Phi^{-1}(U^{(i)})$ é a i -ésima estatística de ordem de uma $N(0,1)$. Temos, então, $a_N(i) = E_{\Phi}(z^{(i)}) = E_U \Phi^{-1}(U^{(i)})$. Aproximando, agora, $U^{(i)}$ por $EU^{(i)}$ que tem a expressão simples $EU^{(i)} = \frac{i}{N+1}$, obtemos $a_N(i) \approx \Phi^{-1}(\frac{i}{N+1})$. Substituindo $a_N(i)$, $i=1, \dots, N$, por estes valores em (2.21), obtemos o teste de Van der Warden. A aplicação deste teste é mais simples, basta utilizar uma tabela da função de distribuição $N(0,1)$. Além disto, o teste de Van der Warden é assintoticamente equivalente ao teste de escores normais.

§12. Comentários sobre os testes apresentados.

Os testes de Wilcoxon e de Escores Normais são correspondentes não paramétricos do teste t-Student. Com estes testes queremos discriminar, por exemplo, f

de g como na figura abaixo. Intuitivamente, g tem uma tendência a gerar valores maiores. Um bom discrimina



dor pode então ser obtido a partir dos postos (S_1, \dots, S_n) das observações geradas por g , na amostra global de todas as observações geradas por g e f . "Valores grandes" de S'_1, \dots, S_n apoiam a situação da figura acima em contraposição a $f = g$. Com a finalidade de caracterizar "valores grandes" de (S_1, \dots, S_n) , podemos, em princípio, utilizar qualquer função $a_N: \{1, \dots, N\} \rightarrow \mathbb{R}$ tal que $a_N(i) \leq a_N(i+1)$. Definida esta função, basta considerar $\sum_{i=1}^n a_N(s_i) \geq c$ como região de rejeição da hipótese $f=g$. Uma escolha simples de a_N é tomar $a_N(i) = i$ que corresponde ao teste de Wilcoxon. No caso do teste de escores normais, $a_N(i) = E_{\Phi}(Z^{(i)})$ e como $Z^{(i)} \leq Z^{(i+1)}$, temos satisfeita a condição $a_N(i) \leq a_N(i+1)$. Os $a_N(i)$'s acima, são chamados de escores. Como devem ser escolhidos estes escores? A maneira como apresentamos o teste de escores normais indica que podemos tomar partido nesta escolha da distribuição particular f envolvida. Os tes

tes obtidos, sendo funções dos postos, têm nível de significância independente de f e g . Contudo, a potência depende de f e, obviamente de Δ . Uma característica importante de f que tem influência na potência é o "peso das caudas". A distribuição normal, por exemplo, tem densidade tendendo para zero, quando $x \rightarrow \pm\infty$ muito velozmente: da ordem de e^{-x^2} . O mesmo não ocorre com a distribuição de Cauchy onde a velocidade desta convergência é $1/x^2$. Uma comparação entre as caudas das diferentes distribuições é apresentada no gráfico da pag. 13, Capítulo I destas notas. Desta comparação, obtemos, em ordem crescente dos pesos das caudas a seguinte ordenação: Uniforme, Normal, Logística, Exponencial Dupla, Cauchy.

O efeito das caudas pesadas é gerar valores excêntricos (outliers) com maior frequência. A idéia é então definir os escores dando menor peso a estas observações. Desta forma, o teste de escores normais tem melhor desempenho que o de Wilcoxon para o caso Normal e para distribuições com caudas mais finas. Por outro lado, o teste de Wilcoxon é particularmente bom no caso da logística, apresentando melhor desempenho que o de escores normais para distribuições com caudas mais grossas como a de Cauchy. Neste último caso, o teste que apresenta proteção máxima é o chamado teste da Mediana. Os escores para estes testes são definidos por: $a(i) = 0$ se $i \leq 1/2(N+1)$

e $a_N(i) = 1$ se $i > 1/2 (N+1)$, sendo então $\sum_{i=1}^n a_N(S_i)$ o número de observações Y's maiores que a mediana da amostra conjunta dos X's e dos Y's. Para maiores detalhes sobre o teste da mediana vide Hájek pag. 56.

§13. Testes para diferença entre escalas.

Os testes apresentados até agora visam, particularmente, detectar a diferença de locação entre duas distribuições. Na realidade, estes são utilizados para testar a alternativa mais geral de que Y é estocasticamente maior do que X. Evidentemente, F e G podem diferir em outros aspectos, por exemplo, terem parâmetros de escalas distintos. No caso de locação, podemos colocar o modelo estatístico da seguinte forma: $N = m+n$ observações; X_1, \dots, X_m e Y_1, \dots, Y_n , onde

$$X_i = e_i + \mu_1 \quad i=1, \dots, m$$

$$Y_j = e_{m+j} + \mu_2 \quad j=1, \dots, n$$

Observamos os X's e Y's sendo e_{m+1}, \dots, e_{m+n} variáveis aleatórias não observáveis e $\Delta = \mu_2 - \mu_1$ é o deslocamento na locação, parâmetro de interesse, no caso. Além disto supomos que os e's são todos independentes e com distri

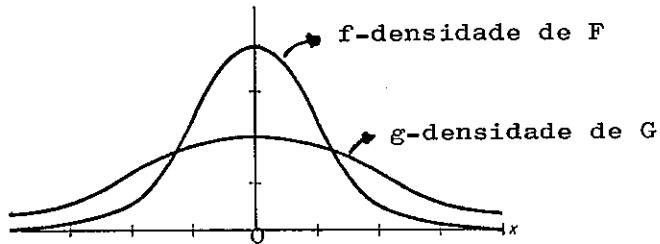
buição contínua. Queremos testar, por exemplo, $\Delta = 0$ contra $\Delta > 0$. No caso de escala o modelo é o seguinte: $N = m+n$ observações; X_1, \dots, X_m e Y_1, \dots, Y_n ,

$$X_i = \sigma_1 e_i + \mu \quad i=1, \dots, m$$

$$Y_j = \sigma_2 e_{m+j} + \mu \quad j=1, \dots, n$$

Onde os e 's são variáveis aleatórias independentes, com distribuição comum contínua e de mesma locação zero (por exemplo, mediana igual a zero), e μ um parâmetro desconhecido. A hipótese de interesse, no caso, é acerca da razão σ_1/σ_2 .

A figura abaixo ilustra as densidades f e g de duas distribuições F e G , com o mesmo parâmetro de locação e diferentes parâmetros de escala:



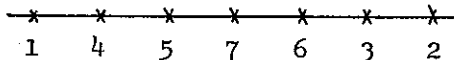
Densidades com o mesmo parâmetro de locação e diferentes parâmetros de escala

Na situação ilustrada na figura, o parâmetro de escala

σ_2 de F é menor que o parâmetro de escala σ_1 de G. Supondo agora que temos m observações (x_1, \dots, x_m) , geradas por F, e n observações (y_1, \dots, y_n) , geradas por G, como poderíamos detectar a diferença entre os parâmetros de escala a partir das observações? Vejamos o que ocorre se tentarmos o teste de Wilcoxon. Raciocinando com a figura acima, a distribuição G tenderia a gerar os valores nos extremos correspondentes aos postos mais altos e mais baixos. Os postos altos compensariam os postos baixos na soma, e o valor da estatística de Wilcoxon não seria significativa. Desta forma, não rejeitaríamos a hipótese $\sigma_1 = \sigma_2$. O teste de Wilcoxon não é, portanto, adequado para detectar este tipo de diferença entre F e G.

Teste de Siegel-Tukey

A idéia neste teste é atribuir postos do seguinte modo: menores postos para os extremos, aumentando os postos à medida que os valores se aproximam do centro. A figura abaixo ilustra este método de atribuição de postos.



Aqui atribuímos posto 1 à menor observação; posto 2 à maior; posto 3 à segunda maior, etc. Sejam S_1, \dots, S_n os postos do tratamento e $W_S = \sum_{i=1}^n S_i$. Valores grandes de W_S são significantes, rejeitando-se a hipótese de nenhum efeito de tratamento quando $W_S \geq c$. A distribuição de W_S é dada pelo Teorema 1, sendo portanto válida, a mesma aproximação normal vista anteriormente.

Exemplo 10: Sejam $m=5$ e $n=6$ e as respostas dos tratamentos: -1,5; -0,7; -0,4; 0; 0,1; 1,3 e dos controles: -1,2; -0,9; -0,8

Observações	Y	X	X	X	Y	Y	Y	Y	X	X	Y
	-1,5	-1,2	-0,9	-0,8	-0,7	-0,4	0	0,1	0,3	1	1,3
Postos	1	4	5	8	9	11	10	7	6	3	2

$$W_r = 4 + 5 + 8 + 6 + 3 = 26$$

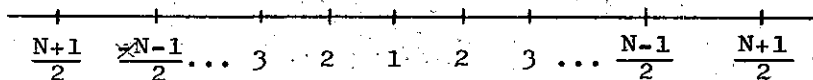
$$W_S = 1 + 9 + 11 + 10 + 7 + 2 = 40$$

Rejeitamos a hipótese $H: \sigma_1 = \sigma_2$ em favor da alternativa $K: \sigma_2 < \sigma_1$, quando W_S é grande, ou equivalentemente, quando W_r é pequeno. Temos: $P(W_r \leq 26) = P(W_{YX} \leq 11) = 0.27$ o que mostra que a diferença não é significativa ao nível de 5%.

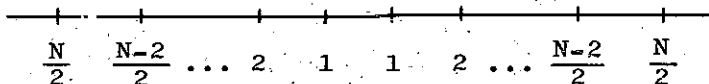
Teste de Ansari-Bradley

Observe que no lugar da atribuição de postos utilizada, seria igualmente razoável, atribuir posto 1 à maior observação e posto 2 à menor, etc. Na maioria das vezes, as decisões tomadas (rejeitar ou aceitar a hipótese), a partir das duas atribuições de postos, coincidem. É possível, porém, construir exemplos onde as decisões são contrárias. Uma maneira de evitar esta ambiguidade é atribuir a cada observação o posto médio dos postos obtidos pelos dois processos de atribuições. Isto é equivalente à atribuição de postos ilustrada abaixo:

Caso $N = 2k+1$ (ímpar)



Caso $N = 2k$ (par)



A distribuição da soma dos postos dos Y's sob a hipótese não é, contudo, a distribuição de Wilcoxon.

Tabelas da distribuição da estatística deste teste são dadas em Wolf e Hollander

§14. Teste de Kolmogorov-Sminorv.

Este teste destina-se a detectar alternativas gerais, isto é, qualquer tipo de diferença entre F e G . Os testes apresentados até este ponto objetivaram detectar diferença na locação, ou, diferença na escala. Tais características não são, contudo, as únicas que distinguem as duas distribuições F e G . O teste de Kolmogorov-Sminorv se baseia na comparação das funções de distribuições empíricas das duas amostras. Consideremos

(X_1, \dots, X_m) uma amostra aleatória da distribuição F e (Y_1, \dots, Y_n) , amostra aleatória da distribuição G .

Sejam $X_{(1)} < X_{(2)} < \dots < X_{(m)}$ e $Y_{(1)} < Y_{(2)} < \dots < Y_{(n)}$ as correspondentes amostras ordenadas. A função de distribuição empírica da amostra (X_1, \dots, X_m) é dada por:

$$(2.22) \quad F_m(x) = \begin{cases} 0 & \text{se } x < X_{(1)} \\ \frac{k}{m} & \text{se } X_{(k)} \leq x < X_{(k+1)} \quad k=1, 2, \dots, m-1 \\ 1 & \text{se } x \geq X_{(m)} \end{cases}$$

As funções de distribuição empíricas F_m e G_n são estimadores das funções de distribuição F e G . Desta forma, diferenças entre F e G se refletem, através

das observações, em diferenças entre F_m e G_n . Se a hipótese $H_0: G(x) = F(x) \quad \forall x$ é verdadeira, (X_1, \dots, X_m) e (Y_1, \dots, Y_n) são amostras da mesma população. A menos portanto, de variações amostrais, sob H_0 , devemos ter uma concordância razoável entre as duas distribuições empíricas. O teste bilateral de K-S utiliza a estatística: $D_{m,n} = \max_x |G_n(x) - F_m(x)|$. A alternativa de interesse é: $K: F(x) \neq G(x)$ para algum x . O teste rejeita a hipótese H_0 em favor de K quando $D_{m,n} \geq c_\alpha$, onde c_α é calculado a partir de: $P_H(D_{m,n} \geq c_\alpha) = \alpha$.

Antes de tentarmos obter a distribuição de $D_{m,n}$ sob a hipótese $H: F = G$ (F e G contínuas), vamos fazer algumas considerações acerca da estatística $D_{m,n}$. Primeiro, apesar de não ser imediato, $D_{m,n}$ só depende dos postos das observações. Para mostrar este fato, vamos introduzir alguma notação. Designemos por

$T: R^N \rightarrow \{1, 2, \dots, N\}$ a função tal que $T(x_1, \dots, x_N) = (\text{Posto}(x_1), \text{Posto}(x_2), \dots, \text{Posto}(x_N))$. Seja G o conjunto das transformações $g_\varphi: R^N \rightarrow R^N$ tal que $g_\varphi(x_1, \dots, x_N) = (\varphi(x_1), \dots, \varphi(x_N))$ onde $\varphi: R \rightarrow R$ é contínua, estritamente crescente e sobre R .

Lema 1 - Com T e G definidas acima temos:

(a) $T(g_\varphi(x_1, \dots, x_N)) = T(x_1, \dots, x_N)$

(b) Se $T(x_1, \dots, x_N) = T(x'_1, \dots, x'_N)$ então existe

$\varepsilon_\varphi \in G$ tal que $(x'_1, \dots, x'_N) = \varepsilon_\varphi(x_1, \dots, x_N)$.

Prova: Parte (a) é evidente, pois a aplicação de φ a cada coordenada preserva a ordem.

Se $T(x_1, \dots, x_N) = T(x'_1, \dots, x'_N)$ então a mesma permutação de $(1, \dots, n)$ ordena as coordenadas crescentemente:

$$x_{i_1} < x_{i_2} < \dots < x_{i_N} \quad \text{e} \quad x'_{i_1} < x'_{i_2} < \dots < x'_{i_N}.$$

Vamos agora construir φ tal que $x'_i = \varphi(x_i) \quad \forall i$ onde $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ é sobre \mathbb{R} , contínua e estritamente crescente. Para isto, defina $\varphi(x) = x + (x'_{i_1} - x_{i_1})$ para $x \leq x_{i_1}$, $\varphi(x) = x + (x'_{i_N} - x_{i_N})$ para $x \geq x_{i_N}$ e linear entre x_{i_k} e $x_{i_{k+1}}$ para $k=1, \dots, n-1$.

Lema 2 - Seja $U: \mathbb{R}^N \rightarrow \mathbb{R}$ tal que $U(\varepsilon_\varphi(x_1, \dots, x_N)) = U(x_1, \dots, x_N)$, então existe $h: \mathbb{R} \rightarrow \mathbb{R}$ tal que $U(x_1, \dots, x_N) = h[T(x_1, \dots, x_N)]$.

Prova: Consideremos (x_1, \dots, x_N) e (x'_1, \dots, x'_N) tais que $T(x_1, \dots, x_N) = T(x'_1, \dots, x'_N)$ então $(x'_1, \dots, x'_N) = \varepsilon_\varphi(x_1, \dots, x_N)$ para algum $\varepsilon_\varphi \in G$ e portanto $U(x'_1, \dots, x'_N) = U(\varepsilon_\varphi(x_1, \dots, x_N)) = U(x_1, \dots, x_N)$.

Observação: Para pontos (x_1, \dots, x_N) onde haja coordenada com valores iguais deve-se considerar o posto médio.

Os lemas acima nos dão o resultado desejado, bastando para isto, mostrar que $D_{m,n}$ é invariante por transformações $g_\varphi \in G$.

Lema 3 - Seja $D_{m,n}(x_1, \dots, x_m; y_1, \dots, y_n)$ a estatística de Kolmogorov-Sminorv, então

$$D_{m,n}(x'_1, \dots, x'_m; y'_1, \dots, y'_n) = D_{m,n}(x_1, \dots, x_m; y_1, \dots, y_n),$$

onde $x'_i = \varphi(x_i)$ e $i=1, \dots, m$ e $y'_j = \varphi(y_j)$, $j=1, \dots, n$.

Prova: Fácil. Temos:

$$D_{m,n}(x_1, \dots, x_m; y_1, \dots, y_n) = \sup_{x \in \mathbb{R}} |F_m(x) - G_n(x)| = \sup_{x \in \mathbb{R}} |F_m(\varphi^{-1}(x)) - G_n(\varphi^{-1}(x))|;$$

(o supremo para $x \in \mathbb{R}$ porque $\varphi(\mathbb{R}) = \mathbb{R}$).

Designando por F'_m e G'_n , respectivamente as funções de distribuições empíricas de $\varphi(X_1), \dots, \varphi(X_m)$ e de $\varphi(Y_1), \dots, \varphi(Y_n)$, temos

$$F_m(\varphi^{-1}(x)) = \frac{\text{número de observações } x_i \text{'s } \varphi^{-1}(x)}{n} = \frac{\text{número de observações } \varphi(x_i) \text{'s } \leq x}{n} = F'_m(x)$$

e igualmente: $G_n^{-1}(\varphi^{-1}(x)) = G'_n(x)$. Logo,

$$\sup_{x \in \mathbb{R}} |F_m(\varphi^{-1}(x)) - G_n(\varphi^{-1}(x))| = \sup_{x \in \mathbb{R}} |F'_m(x) - G'_n(x)| = D_{m,n}(x'_1, \dots, x'_m; y'_1, \dots, y'_n).$$

Teorema 10 - Se $X_1, \dots, X_m; Y_1, \dots, Y_n$ são independentes com distribuição comum F , contínua, a distribuição de $D_{m,n}$ independe de F .

Prova: É uma consequência, imediata, dos Lemas 1, 2 e 3.

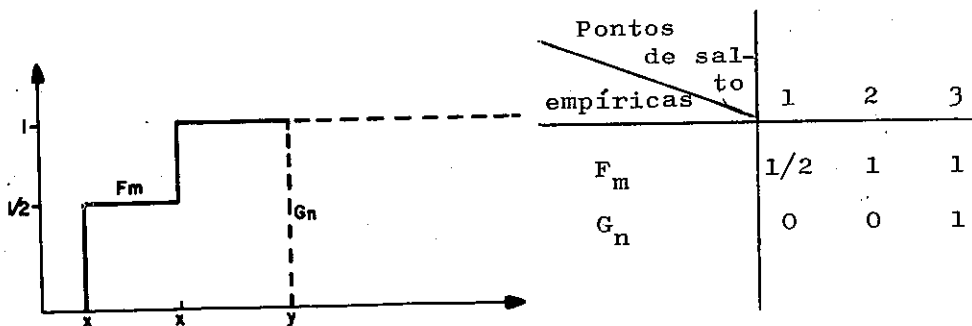
Deste teorema, decorre que a distribuição de $D_{m,n}$ sob a hipótese $F = G$ (F contínua), pode ser obtida a partir de (2.1).

Exemplo 11: Consideremos $m=2, n=1$. Os postos neste caso podem ser: $R_1 = 1, R_2 = 2, S_1 = 3$;

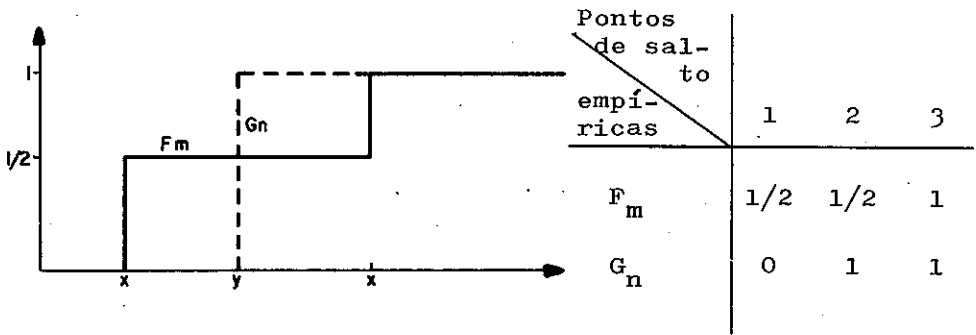
$R_1 = 1, R_2 = 3, S_1 = 2$; $R_1 = 2, R_2 = 3, S_1 = 1$. Sob H , cada uma destas possibilidades tem probabilidade $1/3$.

Precisamos determinar o valor de $D_{m,n}$ em cada um destes casos.

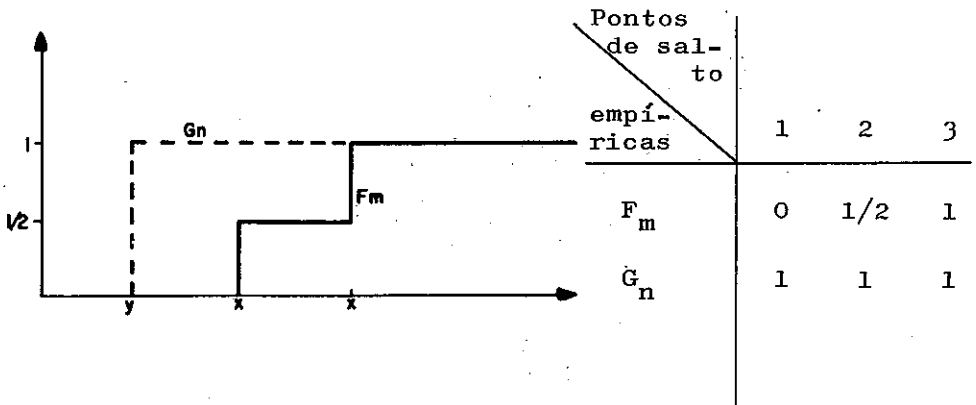
1º Caso: $R_1 = 1, R_2 = 3, S_1 = 3$, correspondente a uma disposição: X X Y



2º Caso: $R_1 = 1, R_3 = 3, S_1 = 2$, correspondente a uma disposição: X Y X.

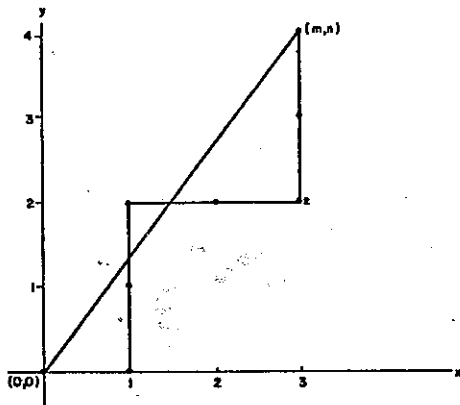


3º Caso: $R_1 = 2, R_2 = 3, S_1 = 1$, correspondente a uma disposição: Y X X.

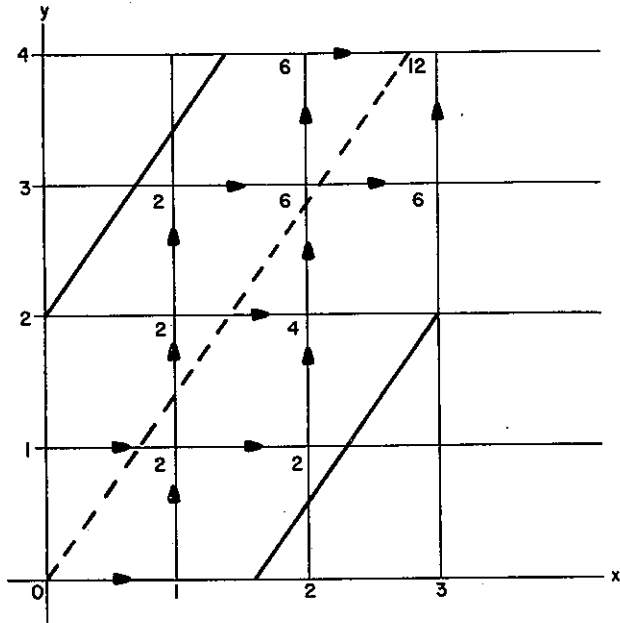


Juntando os resultados acima, temos a distribuição de $D_{m,n}$:
 $P(D_{m,n} = 1/2) = 1/3$ $P(D_{m,n} = 1) = 2/3$.

Apresentamos a seguir um método útil para calcular a probabilidade de significância do teste de Kolmogorov-Sminorv. Sendo d o valor de $D_{m,n}$ observado na amostra, queremos calcular a probabilidade de obter um valor tão ou mais extremo do que o valor observado d , ou seja, $P(D_{m,n} \geq d)$. Valores baixos desta probabilidade favorecem a rejeição da hipótese. O seguinte método é apresentado em Hodges. Primeiro, colocamos a amostra combinada dos x 's e y 's em ordem crescente de magnitude. Consideremos, por exemplo, uma disposição do tipo $xyyxyy$. Podemos representar no plano XY , este arranjo por um caminho que move um passo para a direita para cada observação x e um passo para cima para cada observação y . O gráfico abaixo mostra a representação de $xyyxyy$.



As coordenadas (u,v) , com u e v inteiros representam, respectivamente, os valores de $m F_m(x)$ e $n G_n(x)$. O valor d da estatística de K-S para a amostra é dado pela maior das diferenças $|u/m-v/n| = |nu-mv|/mn$. A equação da reta ligando $(0,0)$ a (m,n) é $nx-my = 0$. A distância vertical de um ponto no caminho a esta reta é: $|v-nu/m|$. Desta forma, o ponto que fornece a maior diferença $|u/m-v/v| = d$ é o ponto mais afastado da reta que liga $(0,0)$ a (m,n) . No gráfico este ponto é representado por Q e sua distância a esta reta é $nd = 2$, portanto, $d = 2/4$. O número total de arranjos das m variáveis X 's e das n variáveis Y 's é $\binom{m+n}{m}$, cada um destes arranjos correspondendo a um caminho entre $(0,0)$ e (m,n) . Sob a hipótese H_0 estes caminhos são igualmente prováveis. Basta, portanto, contar o número de caminhos tendo pontos a uma distância maior ou igual a nd da reta unindo $(0,0)$ a (m,n) . Para esta enumeração, desenhemos duas retas a uma mesma distância vertical nd de OM . O gráfico seguinte representa estas duas retas.



Seja $A(m,n)$ o número de caminhos de $(0,0)$ até (m,n) que estão inteiramente dentro destas linhas de fronteira. A probabilidade procurada é:

$$(2.23) \quad P(D_{m,n} \geq d) = 1 - P(D_{m,n} < d) = 1 - \frac{A(m,n)}{\binom{m+n}{m}}$$

A figura indica a maneira de contar $A(m,n)$, partindo de $(0,0)$ até (m,n) . O número $A(u,v)$ em cada interseção satisfaz a fórmula de recursão:

$$A(u,v) = A(u-1,v) + A(u,v-1)$$

com condições de fronteira: $A(0,v) = A(u,0) = 1$. O valor de $A(u,v)$ é a soma dos valores nos pontos anterior-

res do caminho respeitando-se a condição de este estar sempre dentro das fronteiras. Para o exemplo da disposição $xyyxyy$, onde, $nd = 2$ e $A(3,4) = 12$, temos:

$$P(D_{3,4} \geq 0.5) = 1 - \frac{12}{\binom{7}{4}} = 0.343.$$

A distribuição assintótica, sob a hipótese, de $D_{m,n}$ quando $m, n \rightarrow \infty$ e m/n mantém-se constante, é dada por:

$$(2.24) \quad \lim_{m,n \rightarrow \infty} P\left(\sqrt{\frac{mn}{m+n}} D_{m,n} \leq z\right) = L(z),$$

onde

$$(2.25) \quad L(z) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 z^2}.$$

Alternativas unilaterais podem ser testadas utilizando-se a estatística: $D_{m,n}^+ = \max_x (F_m(x) - G_n(x))$. A região de rejeição para a alternativa de que X é estocasticamente menor que X é dada por $D_{m,n}^+ \geq c_{\alpha}$. A distribuição assintótica de $D_{m,n}$ é dada por

$$(2.26) \quad \lim_{m,n} P\left(\sqrt{\frac{mn}{m+n}} D_{m,n}^+ \leq z\right) = 1 - e^{-2z^2}.$$

Abaixo apresentamos alguns resultados úteis para amostras grandes. Se m e n são os tamanhos das amostras, rejeite ao

$$\text{nível } \alpha = 0,10 \text{ se } d \geq 1.22 \sqrt{\frac{1}{n} + \frac{1}{m}} ;$$

$$\text{nível } \alpha = 0,05 \text{ se } d \geq 1.36 \sqrt{\frac{1}{n} + \frac{1}{m}} ;$$

$$\text{nível } \alpha = 0,05 \text{ se } d \geq 1.63 \sqrt{\frac{1}{n} + \frac{1}{m}} .$$

Comentário: O teste de Kolmogorov-Sminorv destina-se a detectar qualquer tipo de diferença entre as distribuições F e G. O seu desempenho para alternativas específicas, por exemplo de deslocamento, é pobre quando comparado ao teste de Wilcoxon. O teste de Wilcoxon concentra atenção apenas nos arranjos dos postos que sugerem aumento das respostas. O teste de Kolmogorov-Sminorv aloca os arranjos de modo a detectar diversos tipos de alternativas. No caso de se ter idéia de como o tratamento tende a modificar as respostas, é preferível usar um teste específico para o tipo de alternativa em questão.

Exercícios - Capítulo II

1. Para $m=2$, $n=3$, determine:

$$P_H(S_1=1, S_2=2, S_3=3); P_H(S_1=1, S_2=3, S_3=5); P_H(S_1=1);$$
$$P_H(R_1=1) \text{ e } P_H(S_1=2).$$

2. Para $m=2$, $n=3$, determine:

$$P_H(S_1+S_2+S_3=7); P_H(S_1+S_2+S_3 \leq 7); P_H(S_1+2S_2+3S_3 \leq 20) \text{ e}$$
$$P_H(S_1^2+S_2^2+S_3^2 \leq 16).$$

3. Determine a distribuição nula de W_{XY} para os casos

(i) $m=2$, $n=4$; (ii) $m=2$, $n=5$; (iii) $m=n=3$. Desenhe o histograma da distribuição de W_{XY} em cada um dos casos acima.

4. Em um estudo envolvendo $2m$ indivíduos, m recebem tratamento e m servem de controle. As respostas ordenadas forneceram a seguinte disposição: XYXY ... XY, de modo que os postos do tratamento foram $2, 4, \dots, 2m$. Supondo que valores grandes de W_S são significantes, utilize a aproximação normal para obter a probabilidade de significância aproximada, quando m é grande.

5. Determine a distribuição nula de W_S^* e desenhe o seu histograma quando $m=n=3$ e (i) $d_1=2$, $d_2=2$, $d_3=1$, $d_4=1$; (ii) $d_1=1$, $d_2=2$, $d_3=3$; (iii) $d_1=d_2=d_3=2$;

(iv) $d_1=d_2=d_3=1$, $d_4=3$; (v) $d_1=d_2=3$.

6. Por enumeração, determine (i) $P_H(W_S^* \geq 22)$ quando $m=n=4$ e $d_1=d_2=d_3=1$, $d_4=3$, $d_5=d_6=1$; (ii) $P_H(W_R^* \leq 28)$ quando $m=n=6$ e $d_1=2$, $d_2=1$, $d_3=3$, $d_4=2$, $d_5=1$, $d_6=2$, $d_7=1$; (iii) $P_H(W_R^* \leq 86,5)$ quando $m=n=10$ e $d_1=5$, $d_2=9$, $d_3=6$.

7. Para m e n arbitrários determine (i) $P_H[W_S = \frac{1}{2} n(n+1)]$; (ii) $P_H[W_S = \frac{1}{2} n(n+1)+1]$; (iii) $P_H[W_S = \frac{1}{2} n(n+1)+2]$, supondo $m \geq 2$.

8. Obtenha a seguinte aproximação para a potência do teste de nível α de Wilcoxon:

$$\pi(F, G) \approx \Phi \left[\frac{mn(p_1 - 1/2) - U_\alpha \sqrt{mn(N+1)/12}}{\sqrt{\text{Var}(W_{XY})}} \right]$$

para m e n grandes.

9. Partindo da aproximação do problema 8, mostre que para

Δ pequeno, $\pi_F(\Delta) \approx \Phi \left[\sqrt{\frac{12mn}{N+1}} f^*(0)\Delta - U_\alpha \right]$ onde f^* é a densidade da distribuição F^* da diferença de duas variáveis independentes tendo distribuição comum F ; U_α é tal que $\Phi(U_\alpha) = 1-\alpha$.

Sugestão: Considere o desenvolvimento de F^* em torno de zero, $F^*(\Delta) = F^*(0) + \Delta f^*(0)$. Aproxime a variância $\text{Var } W_{XY}$ por seu valor quando $\Delta = 0$.

10. Para o caso em que F é Normal com $\sigma^2 = 4$, determine o tamanho aproximado da amostra para obter, no teste de Wilcoxon com $\alpha = 0,02$, a potência $\pi = 0,9$ na alternativa $\Delta = 1$.

11. Considere a seguinte generalização de W_{XY} , para o caso em que todas as observações não são distintas:

$$W_{XY}^* = [\text{número de pares } (X_i, Y_j) \text{ tal que } X_i < Y_j] + \\ + 1/2 [\text{números de pares } (X_i, Y_j) \text{ tal que } X_i = Y_j].$$

Mostre que $W_{XY}^* = W_S^* - 1/2 n(n+1)$.

12. Seja X_1, \dots, X_m uma amostra aleatória de uma distribuição F e Y_1, \dots, Y_n uma amostra aleatória de uma distribuição G . Suponhamos que tanto F como G concentram toda a massa de probabilidade em dois pontos a e b ($a < b$). Designando por A_1, B_1 e A_2, B_2 o número de X 's iguais a a, b e o número de Y 's iguais a a, b respectivamente:

	a	b	
X	A_1	B_1	m
Y	A_2	B_2	n
	d_1	d_2	N

a) Determine W_{XY}^* (V. Exercício 11)

b) Determine, sob a hipótese $H: F=G$, a distribuição conjunta de A_1, A_2, B_1, B_2 .

c) Determine, sob a hipótese $H: F=G$; a distribuição condicional de B_2 dados d_1 e d_2 .

d) Mostre que o teste bilateral de Wilcoxon (condicional) rejeita H quando B_2 é muito grande ou muito pequeno.

13. Considere uma população $\pi = \{v_1, \dots, v_N\}$, onde r dos v 's são iguais a 1 e os restantes $N-r$ são iguais a zero. Da população π é retirada uma amostra de tamanho n e D é a soma dos elementos da amostra. Deste modo, D é o número de 1's incluídos na amostra e tem distribuição hipergeométrica. Mostre que $ED = \frac{nr}{N}$ e $\text{Var } D = \frac{n(N-n)}{N-1} \frac{r}{N} (1-\frac{r}{N})$. Verifique o resultado calculando a variância de B_2 no problema 12, utilizando, para este cálculo, a relação existente entre B_2 e W_{XY}^* .

14. O teste χ^2 para testar homogeneidade entre duas distribuições, aplicado no problema 12, rejeita $H: F=G$ se $D = \frac{N(a_1 b_2 - a_2 b_1)^2}{m n d_1 d_2}$ é grande onde D tem aproximadamente distribuição χ^2 com 1 g.l. Estude a relação entre este teste e o teste de Wilcoxon para a mesma hipótese.

15. Considere o estimador $\hat{\Delta}$, definido na seção 9. Mostre que a distribuição do erro do estimador $\hat{\Delta} - \Delta$ é independente de Δ .

16. Considere o estimador $\hat{\Delta} = \hat{\Delta}(X_1, \dots, X_m; Y_1, \dots, Y_n)$ definido na seção 9. Mostre que

$$\begin{aligned} \hat{\Delta}(aX_1+b, \dots, aX_m+b; aY_1+b, \dots, aY_n+b) &= \\ &= a \hat{\Delta}(X_1, \dots, X_m; Y_1, \dots, Y_n). \end{aligned}$$

17. Mostre que se $X_1, \dots, X_m; Y_1, \dots, Y_n$ são independentes com uma mesma distribuição F , simétrica em torno de zero, então a distribuição de $\hat{\Delta}$ é simétrica em torno de zero.

18. Utilizando os exercícios 15, 16, 17 mostre que a distribuição de $\hat{\Delta}$ é simétrica em torno de Δ quando F é simétrica em torno de algum μ .

19. Mostre que $\hat{\Delta}(X_1, \dots, X_m; Y_1, \dots, Y_n) = -\hat{\Delta}(Y_1, \dots, Y_n; X_1, \dots, X_m)$.

20. Utilizando os exercícios 15 e 19 mostre que a distribuição de $\hat{\Delta}$ é simétrica em torno de Δ quando $m=n$.

21. Considere os seguintes estimadores para o parâmetro de deslocamento Δ , no problema de duas amostras: $\hat{\Delta} = \text{med}\{Y_j - X_i\}$; $\bar{\Delta} = \bar{Y} - \bar{X}$ e $\tilde{\Delta} = \text{med}(Y_j) - \text{med}(X_i)$. Designando por δ um destes estimadores, mostre que

$$\begin{aligned} \delta(aX_1+b, \dots, aX_m+b; aY_1+c, \dots, aY_n+c) &= a\delta(X_1, \dots, X_m; Y_1, \dots, Y_n) + \\ &+ (c-b). \end{aligned}$$

22. Utilizando o Exercício 21 mostre a simetria de distribuição de \bar{A} e \tilde{A} nos seguintes casos:

- (i) F é simétrica em torno de algum μ .
- (ii) $m = n$.

Capítulo III

Comparações Emparelhadas - Problema de uma amostra

§1. Introdução.

No capítulo anterior estudamos a comparação de dois tratamentos a partir de duas amostras independentes. Uma amostra da distribuição F e outra da distribuição G . Este tipo de comparação tende a ser ineficiente quando as respostas dos indivíduos são altamente variáveis. Esta variação tende a esconder o efeito do tratamento. Neste caso, podemos dividir os indivíduos em blocos mais homogêneos e comparar as respostas correspondentes ao grupo de tratamento e ao grupo de controle, apenas, dentro de cada bloco. Em muitas situações experimentais é possível tomar blocos de tamanho dois. Como exemplos: estudos com gêmeos; os dois braços ou os dois olhos do mesmo indivíduo. Mesmo no caso em que não haja emparelhamento natural, o esquema pode ser adotado tentando-se cuidadosamente assemelhar os indivíduos relativamente às suas características mais importantes.

Suponhamos que uma amostra aleatória, de tamanho

N , de pares de indivíduos é retirada aleatoriamente de uma população. Fazemos aleatoriamente a atribuição de um indivíduo para receber o tratamento e do outro para servir de controle. Além disto, as atribuições para pares distintos devem ser independentes.

Sejam x_i e y_i respectivamente as respostas do tratamento e controle no i -ésimo par. O modelo estatístico utilizado para descrever esta situação considera as respostas como valores de N pares de variáveis aleatórias $(X_1, Y_1) \dots (X_N, Y_N)$, independentes e tendo a mesma distribuição bivariada. Vamos, contudo, concentrar atenções nas diferenças $Z_i = Y_i - X_i$. A hipótese de nenhum efeito é expressa pela igualdade entre as distribuições de $Z_i = Y_i - X_i$ e $-Z_i = X_i - Y_i$. Designando por L a distribuição comum dos Z_i 's: $L(z) = 1 - L(-z)$, isto é, a distribuição dos Z_i 's é, sob a hipótese, simétrica em torno de zero. A alternativa, por exemplo, do tratamento aumentar as respostas, resultaria numa maior frequência de valores positivos dos Z 's. A distribuição L neste caso concentraria mais massa de probabilidade nos valores positivos. Como no modelo de duas amostras, vamos considerar o caso particular, mais simples, do modelo de deslocamento. O efeito do tratamento, neste caso, é dado por um acréscimo de Δ (constante) nas respostas dos controles.

Denotemos por $E(z)$ a distribuição de $Z_i = Y_i - X_i$ sob a hipótese de nenhum efeito, que como vimos acima é simétrica em torno de zero. Portanto, a distribuição de $(Y_i - \Delta) - X_i$ é E , e $P(Z_i \leq z) = P((Y_i - \Delta) - X_i \leq z - \Delta) = E(z - \Delta)$, onde E é simétrica em torno de zero. Neste modelo o problema se reduz a testar $\Delta = 0$ contra alternativas $\Delta > 0$.

Problema de uma amostra

O modelo descrito acima, introduzido a partir da consideração dos pares (X_i, Y_i) , é utilizado, também, quando consideramos N medições de uma grandeza Δ . Neste caso, sendo Z'_1, \dots, Z'_N as medições, consideramos $Z'_i = \Delta + Z_i$ onde os Z_i 's são identicamente distribuídos com distribuição simétrica em torno de zero. Neste contexto, fazer inferências acerca de Δ é chamado problema de uma amostra. Testar, por exemplo, $\Delta = \Delta_0$ reduz-se a testar $\Delta = 0$, no modelo anterior, considerando $Z_i = Z'_i - \Delta_0$.

§2. Teste do sinal.

Consideremos o problema de testar a hipótese $H: \Delta = 0$ contra a alternativa $K: \Delta > 0$. O teste do sinal

se baseia no número S_N de diferenças $Z_i = Y_i - X_i$ positivas. Vamos supor que a distribuição L das Z_i 's é contínua. Deste modo, a probabilidade de diferenças nulas (empates) é zero. As distribuições dos Z_i 's são simétricas em torno de zero, quando a hipótese é verdadeira. Logo, a probabilidade de Z_i ter valor positivo é igual a probabilidade de ter valor negativo, ambos iguais a $1/2$. Por outro lado, os sinais das diferenças Z_i 's para diferentes i 's são independentes, visto que os Z_i 's são independentes. Portanto, supondo a hipótese verdadeira, $S_N =$ número de diferenças $Z_i = Y_i - X_i$ positivas ($i=1, \dots, N$), tem distribuição binomial $B(N; 1/2)$, isto é, com $p = 1/2$. O teste do sinal rejeita a hipótese de nenhum efeito de tratamento quando $S_N \geq c$, onde c é determinado a partir de $P_H(S_N \geq c) = \alpha$, e α é o nível de significância fixado para o teste. Utilizando-se uma tabela de distribuição binomial com $p = 1/2$, acha-se o valor de c .

Quando o valor de N é grande, toma-se a aproximação normal. No caso: $ES_N = N/2$ e $\text{Var } S_N = N/4$, então,

$$P_H\left(\frac{S_N - 1/2 N}{\sqrt{N}/2} \leq a\right) \rightarrow \Phi(a), \text{ quando } N \rightarrow \infty.$$

A determinação da potência do teste do sinal ainda se baseia na distribuição binomial, visto que os Z 's

e em consequência seus sinais são independentes quando alguma alternativa é verdadeira. A probabilidade de $Z_i > 0$ é, contudo, diferente de $1/2$ sob a alternativa:

$P(Z_i > 0) = 1 - L(0) = p$. O número S_N de observações positivas tem, sob a alternativa, distribuição $B(N, p)$. A potência do teste do sinal é dada por

$$(3.1) \quad \pi(p) = P_p(S_N \geq c).$$

Teorema 1 - A potência do teste do sinal é uma função não decrescente de p .

Prova: Consideremos variáveis aleatórias U_1, \dots, U_N independentes e identicamente distribuídas, cada uma tendo distribuição uniforme em $(0, 1)$. Os eventos $[U_i \in (0, p)]$, $i=1, \dots, N$, têm probabilidade p e são independentes. Se $p < p'$ então $[U_i \in (0, p)] \subset [U_i \in (0, p')]$ e a ocorrência de pelo menos c dos eventos $[U_i \in (0, p)]$ implica na ocorrência de pelo menos c dos eventos $[U_i \in (0, p')]$. Por outro lado, a potência na alternativa p é igual à probabilidade de ocorrência de pelo menos c eventos $[U_i \in (0, p)]$, logo crescente com p .

Consequências do Teorema 1:

1. O teste do sinal é não viciado para todas as alternati-

vas tais que $p \geq 1/2$.

2. No caso do modelo de deslocamento o teste do sinal é não viciado para testar $H: \Delta=0$ contra $K: \Delta>0$, pois $p = 1-E(-\Delta) = E(\Delta)$ e portanto $p \geq 1/2$ quando $\Delta > 0$.

3. O teste do sinal de nível α para testar $H: \Delta=0$ é também de nível α para testar $H': \Delta \leq 0$, admitindo-se a possibilidade de efeito negativo do tratamento.

§3. Aproximação normal para $\pi(p)$.

Como vimos, $\pi(p) = P_p(S_N \geq c)$ pode ser obtido a partir de uma tabela da distribuição $B(N, p)$. Quando N é grande podemos aproximar a distribuição de $(S_N - Np)/\sqrt{Npq}$ pela distribuição normal padrão. Deste modo,

$$\pi(p) = P_p(S_N \geq c) = P\left(\frac{S_N - Np}{\sqrt{Npq}} \geq \frac{c - Np}{\sqrt{Npq}}\right),$$

e introduzindo a correção de continuidade temos,

$$(3.2) \quad \pi(p) \approx 1 - \Phi\left(\frac{c - 1/2 - Np}{\sqrt{Npq}}\right).$$

Exemplo 1: Sejam Z_1, \dots, Z_N , N variáveis aleatórias independentes com distribuição comum $N(\Delta, \sigma^2)$.

Considere o teste do sinal para testar $H: \Delta=0$ contra $K: \Delta>0$ e calcule o número de observações necessárias para obter o nível α e a detecção de uma alternativa $\Delta > 0$ com probabilidade $1-\alpha$.

Solução: Vamos utilizar a aproximação normal. A condição do nível α , dá a região de rejeição:

$$\frac{S_N - N/2}{1/2 \sqrt{N}} \geq U_\alpha \quad \text{onde} \quad U_\alpha = \Phi^{-1}(1-\alpha).$$

Logo a constante crítica c do teste é: $c \approx N/2 + 1/2 U_\alpha \sqrt{N}$. O valor de p sob a alternativa Δ_0 é: $p_0 = E(\Delta_0) = \Phi(\Delta_0/\sigma)$. A potência para esta alternativa p_0 , tem valor:

$$\pi(p_0) \approx 1 - \Phi\left[\frac{c - 1/2 - Np_0}{\sqrt{Np_0q_0}}\right] = 1 - \Phi\left[\frac{1/2 U_\alpha - \sqrt{N}(p_0 - 1/2)}{\sqrt{p_0q_0}}\right].$$

A potência na alternativa p_0 é $1-\alpha$:

$$1-\alpha \approx 1 - \Phi\left[\frac{1/2 U_\alpha - \sqrt{N}(p_0 - 1/2)}{\sqrt{p_0q_0}}\right],$$

logo

$$-U_\alpha = \frac{1/2 U_\alpha - \sqrt{N}(p_0 - 1/2)}{\sqrt{p_0q_0}} \quad c$$

$$(3.3) \quad N = \frac{U_\alpha (1/2 + \sqrt{pq})}{p_0 - 1/2}.$$

Tomando-se $\alpha = 5\%$, por exemplo, obtemos a partir de (3.3) os seguintes valores de N , para diferentes valores de Δ/σ :

Δ/σ	0,13	0,25	0,38	0,52	0,67	0,84	1,04
p	0,55	0,60	0,65	0,70	0,75	0,80	0,85
N	1077	265	114	62	37	24	16

Suponhamos, agora, que a ocorrência de empates entre os elementos do par ou que $Z_i = 0$ tenha probabilidade positiva. Sejam $P(Z_i < 0) = p_-$, $P(Z_i = 0) = p_0$ e $P(Z_i > 0) = p_+$. Denotemos os números de Z_i 's positivos, zeros e negativos, respectivamente, por N_+ , N_0 e N_- . A distribuição conjunta de (N_+, N_0, N_-) é a trinomial dada por:

$$P(N_- = a, N_0 = b, N_+ = c) = \frac{N!}{a!b!c!} p_-^a p_0^b p_+^c$$

com $a+b+c = N$.

A hipótese de simetria da distribuição dos Z_i 's em torno de zero, agora implica que $p_+ = p_-$ e não como antes que $p_+ = 1/2$. Nestas condições, a distribuição de N_+ , sob a hipótese, depende de p_0 e é

$$\begin{aligned} P_H(N_+ = c) &= \frac{N!}{c!(N-c)!} p_+^c (1-p_+)^{N-c} = \\ &= \frac{N!}{c!(N-c)!} \frac{1}{2^N} (1-p_0)^c (1+p_0)^{N-c}. \end{aligned}$$

Podemos efetuar o teste condicionalmente dado N_0 . Neste caso, a distribuição do número de Z_i 's positivos, dado N_0 , é independente de p_0 :

$$P(N_+ = c / N_0 = b) = \frac{P(N_+ = c, N_0 = b)}{P(N_0 = b)} = \frac{\frac{N!}{a!b!c!} p_-^a p_0^b p_+^c}{\frac{N!}{b!(N-b)!} p_0^b (1-p_0)^{N-b}} =$$

$$= \frac{(N-b)!}{a!c!} \left(\frac{p_-}{p_- + p_+}\right)^a \left(\frac{p_+}{p_- + p_+}\right)^c.$$

Ou seja a distribuição condicional de N_+ dado N_0 é binomial com $N' = N - N_0$ e $p'_+ = \frac{p_+}{p_- + p_+}$. Sob a hipótese H , $p_+ = p_-$ e $p'_+ = 1/2$. O teste do sinal, nesta situação condicional, reduz-se ao anterior desprezando os empates e tomando o número de observações $N' = N - N_0$, isto é, o número de observações diferentes de zero.

Exemplo 2: Suponhamos que os valores dos Z_i 's sejam:

0,46; 0,00; 1,49; 1,02; 1,39; 0,14; -2,53; -0,35; -0,47; -0,56 e queremos testar $H: \Delta = 0$ contra $K: \Delta > 0$. Temos um valor Z'_i igual a zero. Aplicamos o teste do sinal com $N=9$, aos valores diferentes de zero. O número de valores positivos observados é 5, portanto, a probabilidade de significância do teste é: $P(S_9 \geq 5) > 0,50$. A hipótese não seria rejeitada nem a 5% nem a 1% de nível de significância.

§4. Teste de Wilcoxon dos postos sinalizados.

O teste do sinal utiliza apenas o sinal dos Z_i 's, desprezando, portanto, parte da informação fornecida pelos dados. Assim, uma diferença $Z_i = -2$, por exemplo, é considerada no teste com o mesmo peso de uma diferença $Z_i = -9$. Uma forma de considerar, além do sinal, o valor absoluto dos Z_i 's é dada pelo teste de Wilcoxon dos postos sinalizados. Sejam N pares de indivíduos (ou N indivíduos no problema de 1 amostra) com designação aleatória de tratamento e controle dentro de cada par. Consideremos as diferenças de respostas, nos N pares, entre o tratamento e o controle. Seja $N_+ = n$ o número destas diferenças nos N pares que são positivas e $N_- = m = N - n$ o número de diferenças negativas. Consideremos, agora, os postos dos valores absolutos das diferenças e tomemos cada posto precedido do sinal da diferença correspondente. Suponhamos, por enquanto, que não haja empates entre os valores absolutos das diferenças. Designemos por $R_1 < \dots < R_m$ os postos correspondentes às diferenças negativas e por $S_1 < \dots < S_n$ os correspondentes às diferenças positivas. Deste modo, os R_i 's e os S_j 's, conjuntamente, são os inteiros $(1, \dots, N)$. Para caracterizar uma das 2^N possí-

veis combinações de postos sinalizados $\pm 1, \pm 2, \dots, \pm N$, basta conhecer (S_1, \dots, S_n) , subconjunto dos postos das diferenças positivas. Por exemplo, se $n=3$; $S_1=1, S_2=2, S_3=3$, temos $(+1, +2, +3, -4, \dots, -N)$. Quando $n=0$ o conjunto $(S_1 \dots S_n)$ é vazio. No teste da hipótese H de nenhum efeito de tratamento, contra a alternativa do tratamento aumentar as respostas, aquela é rejeitada quando obtemos muitas diferenças positivas e, além disto, as diferenças positivas são maiores em valor absoluto que as negativas. Deste modo, no teste de postos sinalizados de Wilcoxon as diferenças positivas maiores têm peso maior. A estatística do teste é a soma dos postos sinalizados positivos: $V_S = S_1 + \dots + S_n$. Rejeitamos a hipótese de nenhum efeito quando V_S é suficientemente grande: $V_S \geq c$. É importante observar que o número n de parcelas de V_S é o valor de uma variável aleatória, que toma valores no conjunto $(0, 1, \dots, N)$.

Exemplo 3: Consideremos $N = 10$ pares cujas diferenças entre as respostas do tratamento e controle são: $-1,8; -0,3; 1,2; 1,0; -0,5; -1,6; -0,1; -0,4; -0,6; 0,7$. Neste caso, $N_+ = n = 3$ e temos:

Valores das diferenças	-1,8	-0,3	1,2	1,0	-0,5	-1,6	-0,1	-0,4	-0,6	0,7
Postos sinalizados	-10	-2	+8	+7	-4	-9	-1	-3	-5	+6

$(S_1, S_2, S_3) = (6, 7, 8)$; $(R_1, R_2, R_3, R_4, R_5, R_6, R_7) = (1, 2, 3, 4, 5, 9, 10)$
e $V_S = 6+7+8 = 21$.

Vamos, agora, mostrar que, sob a hipótese de nenhum efeito, a distribuição de V_S é independente da distribuição dos Z_i 's. Para isto, basta mostrar que a distribuição de $(N_+; S_1, \dots, S_n)$ independe, sob a hipótese, da distribuição dos Z_i 's.

Teorema 2 - Se Z_1, \dots, Z_N são independentes com distribuição contínua L , simétrica em torno de zero, então:

$$(3.4) \quad P_H(N_+ = n; S_1 = s_1, \dots, S_n = s_n) = (1/2)^N$$

onde N_+ denota o número de Z 's positivos e $s_1 < \dots < s_n$ com $s_i \in \{1, 2, \dots, N\}$.

Prova: A expressão (3.4) é equivalente a afirmar que, sob H , cada uma das 2^N possíveis combinações de sinais $\pm 1, \pm 2, \dots, \pm N$ dos postos sinalizados, tem a mesma probabilidade $(1/2)^N$. Considere a função sinal:

$$U(x) = \begin{cases} 1 & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$

Por hipótese a distribuição L (contínua) é simétrica em torno de zero, logo $P(U(Z_i)=1) = P(Z_i \geq 0) = 1/2$. Por ou-

tro lado, $U(Z_i)$ é independente de $|Z_i|$, pois temos:

$$\begin{aligned} P(U(Z_i)=1, |Z_i| \leq z) &= P(0 \leq Z_i \leq z) = 1/2 P(|Z_i| \leq z) = \\ &= P(U(Z_i)=1) P(|Z_i| \leq z). \end{aligned}$$

Este fato e a independência dos Z_i 's implicam a independência entre os vetores $(U(Z_1), \dots, U(Z_N))$ e $(|Z_1|, \dots, |Z_N|)$. Em particular, $(U(Z_1), \dots, U(Z_N))$ é independente do vetor de postos $(\text{Posto } |Z_1|, \dots, \text{Posto } |Z_N|)$, visto que este último só depende de $(|Z_1|, \dots, |Z_N|)$. Seja agora (I_1, \dots, I_N) a permutação de $(1, \dots, N)$ que ordena os $|Z_i|$'s: $|Z_{I_1}| < \dots < |Z_{I_N}|$. O vetor aleatório (I_1, \dots, I_N) é função somente dos postos $(\text{Posto } |Z_1|, \dots, \text{Posto } |Z_N|)$ e portanto também independente de $(U(Z_1), \dots, U(Z_N))$. Desta forma, para uma dada permutação $I_1 = i_1, \dots, I_N = i_N$, temos:

$$\begin{aligned} P(U(Z_{I_2}) = t_1, \dots, U(Z_{I_N}) = t_N \mid I_1 = i_1, \dots, I_N = i_N) &= \\ P(U(Z_{i_1}) = t_1, \dots, U(Z_{i_N}) = t_N \mid I_1 = i_1, \dots, I_N = i_N) &= \\ P(U(Z_{i_1}) = t_1, \dots, U(Z_{i_N}) = t_N) &= (1/2)^N, \end{aligned}$$

onde $t_i = 1$ ou 0 , $i = 1, \dots, N$. Sendo esta igualdade válida para qualquer permutação $I_1 = i_1, \dots, I_N = i_N$, temos:

$$P(U(Z_{I_1}) = t_1, \dots, U(Z_{I_N}) = t_N) = (1/2)^N.$$

Exemplo 4: Determine a distribuição de V_S , sob a hipótese de nenhum efeito de tratamento (L simétrica em torno de zero) quando $N = 3$.

Solução: Os possíveis vetores de postos sinalizados e correspondentes valores v 's de V_S são dados na tabela abaixo:

Postos sinalizados	-1,-2,-3	+1,-2,-3	-1,+2,-3	-1,+2,+3
v	0	1	2	3
Postos sinalizados	+1,+2,-3	+1,-2,+3	-1,+2,+3	+1,+2,+3
v	3	4	5	6

Pelo Teorema 3.1 cada conjunto de postos sinalizados tem probabilidade $1/8$, e a distribuição de V_S é dada por:

v	0	1	2	3	4	5	6
$P_H(V_S=v)$	$1/8$	$1/8$	$1/8$	$2/8$	$1/8$	$1/8$	$1/8$

Observe que a distribuição de V_S é, no exemplo acima, simétrica em torno de 3.

A distribuição de V_S , no caso geral, pode ser obtida por enumeração. Designando por $\#(v;N)$ o número de combinações de postos sinalizados, cuja soma dos pos-

tos positivos é igual a v , obtemos

$$(3.5) \quad P_H(V_S=v) = \frac{\#(v;N)}{2^N}$$

Várias tabelas são disponíveis com os valores destas probabilidades. É importante observar, para facilitar o uso destas tabelas, que a distribuição de V_S é simétrica em torno de $N(N+1)/4$. Além disto, sendo $V_R = R_1 + \dots + R_m$, então V_R tem a mesma distribuição que V_S . A simetria da distribuição de V_S em torno de $1/4 N(N+1)$ é fácil provar a partir da igualdade entre as distribuições de V_S e V_R e de $V_R + V_S = 1/2 N(N+1)$.

Quando N é grande, utiliza-se a aproximação Normal para distribuição de V_S . Antes de discutirmos esta aproximação normal, vamos calcular a esperança e a variância de V_S , sob a hipótese de nenhum efeito do tratamento.

§5. Cálculo de $E_H(V_S)$ e de $Var_H(V_S)$.

Como vimos, sob a hipótese de nenhum efeito, cada um dos postos $1, 2, \dots, N$ tem sinal positivo com probabilidade $1/2$ e além disto os sinais são independentes para os diferentes postos. As parcelas de V_S formam, en-

tão, uma amostra da população $1, 2, \dots, N$, sendo, cada elemento incluído na amostra com probabilidade $1/2$ e as decisões de incluir, distintos elementos, independentes. Seja $I_k = 1$ se k é incluído na amostra e $I_k = 0$ caso contrário. Portanto, temos: $P(I_k=1) = 1/2$ e $P(I_k=0) = 1/2$, $k=1, \dots, N$ e I_k independente de I_i ; se $i \neq j$. Logo:

$$V_S = \sum_{k=1}^N k I_k, \quad E(V_S) = \sum_{k=1}^N k E(I_k)$$

$$\text{Var}(V_S) = \sum_{k=1}^N k^2 \text{Var} I_k.$$

Sendo I_k ($k=1, \dots, n$) uma variável de Bernoulli com $p = 1/2$, temos $E I_k = 1/2$ e $\text{Var} I_k = 1/4$, $k=1, \dots, N$. Substituindo estes valores, obtemos

$$(3.6) \quad E(V_S) = \frac{1}{2} \sum_{k=1}^N k = N(N+1)/4$$

$$(3.7) \quad \text{Var}(V_S) = \frac{1}{4} \sum_{k=1}^N k^2 = N(N+1)(2N+1)/24.$$

§6. Aproximação Normal para distribuição de V_S .

Na secção anterior vimos que $V_S = \sum_{k=1}^N k I_k$, com os I_k 's independentes e igualmente distribuídos. Esta ex

pressão de V_S é válida sob a hipótese de nenhum efeito de tratamento, isto é, quando a distribuição dos Z_i 's é simétrica em torno de zero. Observe que V_S , na forma acima, é igual a uma soma de variáveis aleatórias independentes. As parcelas não são, contudo, identicamente distribuídas. O Teorema Central do Limite, na sua forma mais simples, não é aplicável neste caso. Vamos a seguir apresentar a versão do Teorema Central, devida a Lindberg, para a soma de variáveis aleatórias independentes e não necessariamente identicamente distribuídas. Trataremos do caso particular onde as parcelas são variáveis aleatórias discretas.

Sejam X_1, X_2, \dots variáveis aleatórias discretas. Suponhamos que X_1 toma os valores a_1, a_2, \dots ; X_2 os valores b_1, b_2, \dots ; etc. Designemos por μ_i e σ_i^2 respectivamente a média e a variância de X_i . A variância da soma $S_n = X_1 + \dots + X_n$ é $\sigma_1^2 + \dots + \sigma_n^2$ onde $\sigma_1^2 = \sum_j P(X_1=a_j)(a_j-\mu_1)^2$; $\sigma_2^2 = \sum_j P(X_2=b_j)(b_j-\mu_2)^2$, etc. A idéia do teorema é controlar a contribuição, devida às massas nas caudas das distribuições de $X_1, X_2, \dots, X_n, \dots$, na variância de S_n . Para isto consideremos na variância de X_1 , $\sigma_1^2 = \sum_j P(X_1=a_j)(a_j-\mu_1)^2$, a contribuição dos termos que satisfazem: $(a_j-\mu_1)^2 > t(\sigma_1^2 + \dots + \sigma_n^2)$. Denotemos esta contribuição por $\tau_{n1}^2(t) = \sum_{j \in A_1^n(t)} P(X_1=a_j)(a_j-\mu_1)^2$ onde

$$A_1^n(t) = \{j; (a_j - \mu_1)^2 > t(\sigma_1^2 + \dots + \sigma_n^2)\}.$$

Do mesmo modo podemos definir a contribuição dada pelas caudas de X_2 :

$$(3.8) \quad \tau_{n2}^2(t) = \sum_{j \in A_2^n(t)} P(X_2 = b_j) (b_j - \mu_2)^2$$

onde

$$A_2^n(t) = \{j; (b_j - \mu_2)^2 > t(\sigma_1^2 + \dots + \sigma_n^2)\},$$

analogamente definimos $\tau_{n3}^2(t), \dots, \tau_{nn}^2(t)$.

Teorema 3 (Lindberg) - Sejam X_1, \dots, X_n independentes com médias $E(X_i) = \mu_i$ e variâncias $\text{Var}(X_i) = \sigma_i^2$. Consideremos as somas padronizadas

$$S_n^* = \frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}}.$$

Uma condição suficiente para que a distribuição limite de S_n^* seja $N(0,1)$ é que para cada t

$$\frac{\tau_{n1}^2(t) + \dots + \tau_{nn}^2(t)}{\sigma_1^2 + \dots + \sigma_n^2} \rightarrow 0 \text{ quando } u \rightarrow \infty.$$

Prova: V. Gnedenko pag. 260.

Como aplicação deste teorema vamos provar a norma

lidade assintótica de $V_S = \sum_{k=1}^N k I_k$, quando $N \rightarrow \infty$.

Fazendo $X_k = k I_k$, temos, $\mu_k = k/2$, $\sigma_k^2 = k^2/4$ e daí

$$\sigma_1^2 + \dots + \sigma_n^2 = \frac{1}{4} \sum_{i=1}^n k^2 = \frac{1}{24} n(n+1)(2n+1).$$

Como cada X_k toma apenas dois valores, ou seja, $X_k = 0$ ou $X_k = k$, temos $\tau_{nk}^2(t) = 0$ se

$$\max \{ (k - \mu_k)^2, (0 - \mu_k)^2 \} < tn(n+1)(2n+1)/24.$$

Como $(k - \mu_k)^2 = k^2/4$, $(0 - \mu_k)^2 = k^2/4$ e $k^2/4 < n^2$, para $k=1, \dots, n$, podemos para cada t determinar $n_0(t)$ tal que $n^2 < tn(n+1)(2n+1)/24$, se $n > n_0(t)$. Logo, para $n > n_0(t)$, $\tau_{n1}^2(t) = \dots = \tau_{nn}^2(t) = 0$ e a condição do teorema é satisfeita.

§7. Caso de Empates.

A definição dada de postos sinalizados, supõe a não ocorrência de empates entre os valores absolutos das diferenças e que nenhuma das diferenças é igual a zero. Quando supomos a distribuição L contínua, a probabilidade de ocorrência destes eventos é zero. Valores empatados, contudo, ocorrem na prática devido à aproximação limitada, utilizada na medição das respostas. Neste caso,

consideramos os postos médios dos valores absolutos das respostas, do modo como foi feito no teste de Wilcoxon para duas amostras. Quando a distribuição L não é contínua a probabilidade de empates é positiva. O teste neste caso é válido condicionalmente, dada a configuração de empates.

Exemplo 5: Na comparação de dois métodos de ensino, 10 pares de alunos foram utilizados. Os alunos em cada par tinham aproximadamente o mesmo QI, mesmo nível sócio-econômico mesmo nível de conhecimentos, etc. Em cada par um aluno foi selecionado ao acaso sendo a ele aplicado o método I, e ao outro elemento do par aplicado o método II. No fim do treinamento, os alunos foram submetidos a um exame e classificados segundo os graus A, B, C, D, F. Supondo que não haja diferença entre os dois métodos, determine a distribuição dos postos sinalizados.

Vamos supor, por exemplo, a ocorrência da seguinte configuração de empates:

$d_0 = 3$ com mesmo grau para ambos elementos do par.

$d_1 = 2$ valores absolutos de diferenças iguais ao menor valor diferente de zero.

$d_2 = 3$ valores absolutos de diferenças iguais ao segundo menor valor diferente de zero.

$d_3 = 1$ valor absoluto de diferença igual ao maior valor.

Os postos dos valores absolutos das diferenças são 0,0,0,1,1,2,2,3. Os postos médios correspondentes são 2;2;2;4,5;4,5;6,5;6,5,8.

Após a determinação dos postos médios, vamos restringir a atenção apenas aos postos das diferenças não nulas. No teste do sinal desprezamos os pares com diferenças de respostas nulas. No teste de postos sinalizados de Wilcoxon, estas, são somente consideradas na determinação dos postos. Temos então os seguintes grupos (s_1^*, \dots, s_n^*) de postos sinalizados positivos, respectivas somas V_s^* e probabilidades de ocorrências.

$(s_1^* \dots s_n^*)$	v_s^*	$P(V_S^* = v_s^*)$	v_s^*	(s_1^*, \dots, s_n^*)
nenhum	0	1/32	30	4,5;4,5;6,5;6,5;8
4,5	4,5	2/32	25,5	4,5;6,5;6,5;8
6,5	6,5	2/32	23,5	4,5;4,5;6,5;8
8	8	1/32	22	4,5;4,5;6,5;6,5
4,5;4,5	9	1/32	21	6,5;6,5;8
4,5;6,5	11	4/32	19	4,5;6,5;8
4,5;8	12,5	2/32	17,5	4,5;6,5;6,5
6,5;6,5	13	1/32	17	4,5;4,5;8
6,5;8	14,5	2/32	15,5	4,5;4,5;6,5

As probabilidades na coluna central referem-se aos valores de v_S^* da esquerda e da direita. Estas probabilidades foram obtidas por simples enumeração, que mostramos para alguns casos na tabela abaixo:

	4,5	4,5	6,5	6,5	8	v_S^*
						4,5
	+					4,5
		+				11
			+			11
	+			+		11
		+	+			11
		+			+	11
			+		+	14,5
				+	+	14,5

Observe que a distribuição de V_S^* é simétrica, no exemplo acima, em torno de 15. No caso geral a distribuição de V_S^* é simétrica em torno de seu valor esperado $E V_S^*$.

§8. Cálculo de $E_H V_S^*$ e de $Var_H V_S^*$.

A distribuição de V_S^* para uma dada configuração de empates independe da distribuição L , geradora dos z_i 's. Vamos fixar a configuração de empates por d_0, d_1, \dots, d_e tal que $\sum_{i=1}^e d_i = N$, os d 's definidos no

exemplo anterior. Desprezando, agora, os pares empatados, temos $N' = N - d_0$ valores absolutos de diferenças não nulas cujos postos médios são: $u_1, \dots, u_{N'}$ (as diferenças nulas são consideradas no valor dos postos médios). Entre os postos médios $u_1, \dots, u_{N'}$, d_1 são iguais a $d_0 + 1/2(d_1+1)$; d_2 são iguais a $d_0 + d_1 + 1/2(d_2+1)$ e etc. Pelo mesmo raciocínio utilizado no cálculo de $E_{H V_S}$ e $Var_{H V_S}$, temos:

$$E_{H V_S}^* = 1/2 (u_1 + \dots + u_{N'})$$

$$Var_{H V_S}^* = 1/4 (u_1^2 + \dots + u_{N'}^2).$$

Observe que o conjunto dos u 's aumentado dos postos médios das diferenças nulas, d_0 valores iguais a $1/2(d_0+1)$ coincide com o conjunto de postos médios $\{v_1, \dots, v_N\}$ definido no cálculo de $E W_S^*$, na página 33. Portanto, temos:

$$\sum_{i=1}^{N'} u_i + 1/2 d_0(d_0+1) = \sum_{j=1}^N v_j = \frac{N(N+1)}{2}$$

e

$$(3.9) \quad E_H(V_S^*) = 1/4 [N(N+1) - d_0(d_0+1)].$$

Do mesmo modo,

$$\begin{aligned} \sum_{i=1}^{N'} u_i^2 + 1/4 d_0(d_0+1)^2 &= \sum_{j=1}^N v_j^2 = \\ &= \frac{N(N+1)(2N+1)}{6} - \frac{\sum_{i=0}^e d_i(d_i^2-1)}{12} \end{aligned}$$

Logo:

$$(3.10) \text{Var}_H(V_S^*) = \frac{N(N+1)(2N+1) - d_o(d_o+1)(2d_o+1)}{24} - \frac{\sum_{i=1}^e d_i(d_i^2-1)}{48}.$$

§9. Aproximação Normal para distribuição de V_S^* .

Para demonstrar a normalidade assintótica (sob a hipótese) da estatística de Wilcoxon V_S , fizemos $X_k = k I_k$ e o resultado foi obtido por aplicação do teorema de Lindberg. No caso de empates podemos escrever $V_S^* = \sum_{k=1}^N u_k I_k$, onde u_o destes u_k 's são nulos e os restantes $N' = N - n_o$ os postos médios, sendo d_1 dos u_k 's iguais a $d_o + 1/2(d_1+1)$; d_2 iguais a $d_o + d_1 + 1/2(d_2+1)$ e etc. Queremos, agora, obter a normalidade assintótica de V_S^* . Será que podemos aplicar o Teorema 3 deste capítulo? Uma diferença essencial entre os casos (empates e sem empates), mencionados acima, fornece a resposta. Suponhamos, por exemplo, $N=3$ e seja o caso sem empates, então podemos escrever $V_S = I_1 + 2I_2 + 3I_3$. Quando passamos de N para $N+1$, temos $V_S = I_1 + 2I_2 + 3I_3 + 4I_4$. Observe que nos dois casos as variáveis

$X_k = kI_k$ $k=1, \dots, N$ são as mesmas e, portanto, suas distribuições independem de N . Vejamos, agora, o caso de empates. Seja, por exemplo, $N=4$; $d_0=1$, $d_1=2$ e $d_2=1$, então, $V_S^* = 0I_1 + 2,5I_2 + 2,5I_3 + 4I_4$. Vamos passar de N para $N+1$, introduzindo um Z_i tal que: $N=5$; $d_0=1$, $d_1=3$ e $d_2=1$ e neste caso $V_S^* = 0I_1 + 3I_2 + 3I_3 + 3I_4 + 5I_5$. Deste modo, quando passamos de N para $N+1$ as variáveis $X_k = u_{k-k} I_k$ para $k=1, \dots, N$, têm distribuições diferentes. Esta situação não é coberta pelo Teorema 3. Podemos utilizar uma extensão do Teorema 3 para o seguinte esquema triangular:

X_{11} - com distribuição F_{11}

X_{21}, X_{22} - independentes com distribuições F_{21}, F_{22}

X_{31}, X_{32}, X_{33} - independentes com distribuições F_{31}, F_{32}, F_{33}

.....

$X_{n1} \dots X_{nn}$ - independentes com distribuições $F_{n1} \dots F_{nn}$.

Denotemos por μ_{ni} e σ_{ni}^2 , respectivamente, a média e a variância de X_{ni} e seja

$$S_n^* = \frac{\sum_{i=1}^n (X_{ni} - \mu_{ni})}{\sqrt{\sum_{i=1}^n \sigma_{ni}^2}} .$$

Para a n -ésima linha deste esquema triangular, definimos

$\tau_{n1}^2(t), \dots, \tau_{nn}^2(t)$ como anteriormente.

Teorema 4 - Se para cada $t > 0$

$$(3.11) \quad \frac{\tau_{n1}^2(t) + \dots + \tau_{nn}^2(t)}{\sigma_{n1}^2 + \dots + \sigma_{nn}^2} \rightarrow 0 \quad \text{quando } n \rightarrow \infty$$

a distribuição limite de S_n^* é a distribuição normal padrão $N(0,1)$.

Uma aplicação deste teorema fornece a prova da normalidade assintótica de V_S^* . Podemos, de modo mais apropriado,

escrever $V_S^* = \sum_{k=1}^n v_{nk} I_k$, onde d_0 dos v_{nk} 's são iguais a zero e os $N' = N - d_0$ restantes são os postos médios. Fazendo $X_{nk} = v_{nk} I_k$, temos $EX_{nk} = u_{nk} = v_{nk}/2$ e

$\text{Var } X_{nk} = \sigma_{nk}^2 = v_{nk}^2/4$. Como $X_{nk} = 0$ ou $X_{nk} = v_{nk}$ e $(0 - v_{nk}/2)^2 = (v_{nk} - v_{nk}/2)^2 = 1/4 v_{nk}^2$, basta mostrar que

dado $t > 0$ podemos tomar n suficientemente grande tal que $1/4 v_{nk}^2 < t \sum_{k=1}^n \sigma_{nk}^2$, para $k=1, \dots, n$, ou equivalentemente $\max_{k=1, \dots, n} v_{nk}^2 < t \sum_{k=1}^n v_{nk}^2$. Para isto basta mostrar

$$\lim_{n \rightarrow \infty} \frac{\max_{k=1, \dots, n} v_{nk}^2}{\sum_{k=1}^n v_{nk}^2} = 0. \quad \text{Temos, a partir de (3.10), } \sum_{k=1}^n v_{nk}^2 =$$

$$= \sum_{i=1}^{N'} u_i^2 = \frac{n(n+1)(2n+1) - d_0(d_0+1)(2d_0+1)}{6} - \frac{\sum_{i=1}^e d_i(d_i^2-1)}{12}.$$

Por outro lado, $\sum_{i=1}^e d_i(d_i^2-1) < \sum_{i=1}^e d_i^3 \leq \left(\sum_{i=1}^e d_i\right)^3 = (n-d_0)^3$

portanto,

$$\sum_{k=1}^n v_{nk}^2 \geq \frac{n(n+1)(2n+1) - d_0(d_0+1)(2d_0+1)}{6} - \frac{(n-d_0)^3}{12}.$$

Temos ainda, $\max_{1 \leq k \leq n} v_{nk}^2 \leq n^2$ e

(3.12)

$$\frac{\max_{nk} v_{nk}^2}{\sum_{k=1}^n v_{nk}^2} \leq \frac{n^2}{\frac{n(n+1)(2n+1) - d_0(d_0+1)(2d_0+1)}{6} - \frac{(n-d_0)^3}{12}}.$$

Se $n-d_0 \rightarrow \infty$ quando $n \rightarrow \infty$, então o segundo membro de (3.12) tende para zero quando $n \rightarrow \infty$. Isto pode ser verificado como segue:

$$\frac{n^2}{\frac{n(n+1)(2n+1) - d_0(d_0+1)(2d_0+1)}{6} - \frac{(n-d_0)^3}{12}} = \frac{n^2}{\sum_{k=d_0+1}^n k^2 - \frac{(n-d_0)^3}{12}}.$$

Agora, majoramos o denominador do segundo membro:

$$k^2 \geq \int_{k-1}^k t^2 dt, \text{ implicando } \sum_{k=d_0+1}^n k^2 \geq \int_{d_0}^n t^2 dt = \frac{n^3 - d_0^3}{3}.$$

Além disto $\frac{(n-d_0)^3}{12} \leq \frac{n^3 - d_0^3}{12}$, pois $d_0 \leq n$. Portanto,

$$\sum_{k=d_0+1}^n k^2 - \frac{(n-d_0)^3}{12} \leq \frac{n^3 - d_0^3}{3} - \frac{n^3 - d_0^3}{12} = \frac{n^3 - d_0^3}{4}.$$

Introduzindo em (3.12) o majorante obtido, temos

$$\frac{\max_{1 \leq n \leq k} v_{nk}^2}{\sum_{k=1}^n v_{nk}^2} \leq \frac{4n^2}{n^3 - d_0^3} = \frac{4n^2}{(n-d_0)(n^2 + nd_0 + d_0^2)} \leq \frac{4}{n-d_0} \rightarrow 0, \text{ quan-}$$

do $n \rightarrow \infty$, pois por hipótese $n-d_0 \rightarrow \infty$ quando $n \rightarrow \infty$.

A condição 3.11 do Teorema 4 é, então, satisfeita e

$$(3.13) \quad \frac{V_S^* - E(V_S^*)}{\sqrt{\text{Var } V_S^*}} \xrightarrow{D} N(0,1).$$

§10. Potência do Teste de Postos sinalizados de Wilcoxon.

Consideremos o modelo de deslocamento descrito anteriormente. O efeito do tratamento, neste caso, é deslocar a distribuição de $Z_i = Y_i - X_i$ de uma quantia fixa Δ . A distribuição $E(z)$ de Z_i , $i=1, \dots, N$, sob a hipótese de efeito nulo, é simétrica em torno de zero. Sob uma alternativa Δ a distribuição de Z_i é $E(z-\Delta)$, $i=1, \dots, N$. A função potência do teste de postos sinalizados de Wilcoxon é definida por: $\pi_E(\Delta) = P_\Delta(V_S \geq c)$. O índice Δ indica, aqui, a alternativa sob consideração. Para a determinação da potência do teste, seria preciso conhecer a distribuição dos postos sinalizados quando a distribuição dos Z_i 's é simétrica em torno de $\Delta \neq 0$. Esta tarefa é bastante complicada, sendo contornada pela utilização da aproximação

mação normal para a distribuição de V_S , quando $\Delta \neq 0$.
Antes de mostrarmos a validade desta aproximação, apresentamos o seguinte:

Teorema 5 - No modelo de deslocamento, a função potência do teste de postos sinalizados de Wilcoxon,

$\pi_E(\Delta) = P_{\Delta}(V_S \geq c)$ é uma função não decrescente de Δ .

Prova: Seja $\Delta_0 < \Delta_1$. No modelo de deslocamento

$$P_{\Delta}(Z_i \leq z) = E(z - \Delta). \text{ Logo,}$$

$$\begin{aligned} P_{\Delta_1}(Z_i \leq z) &= E(z - \Delta_1) = E[z - (\Delta_1 - \Delta_0) - \Delta_0] = P_{\Delta_0}(Z_i \leq z - (\Delta_1 - \Delta_0)) = \\ &= P_{\Delta_0}(Z_i + (\Delta_1 - \Delta_0) \leq z). \end{aligned}$$

Logo, sob a alternativa Δ_0 , $Z_i + (\Delta_1 - \Delta_0)$ tem a mesma distribuição que Z_i , sob a alternativa Δ_1 .

Seja $Z'_i = Z_i + (\Delta_1 - \Delta_0)$ e denotemos por V'_S a soma dos postos sinalizados positivos calculados a partir dos Z_i 's.

Então

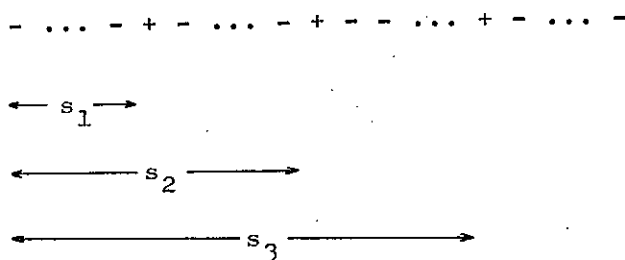
$$P_{\Delta_1}(V_S \geq c) = P_{\Delta_0}(V'_S \geq c).$$

Para a conclusão prova basta obtermos outra expressão para V_S . Para isto, consideremos o conjunto das médias

$1/2 (Z_i + Z_j)$ com $i \leq j$. É fácil verificar que $V_S =$ número de médias positivas $1/2 (Z_i + Z_j)$ com $i \leq j$.

Para mostrar esta igualdade, consideremos os Z 's ordenados segundo a ordem crescente dos seus valores absolutos. Su

ponhamos que os sinais dos Z 's, assim ordenados, fornecem a seguinte disposição:



Nesta ordenação, cada soma do primeiro Z positivo com um dos Z 's que o precede ou consigo próprio é positiva. O número total de tais somas é s_1 . O mesmo raciocínio fornece s_2 somas para o segundo Z positivo, e assim por diante. Neste modo, temos um total de V_S pares (i, j) com $Z_i + Z_j > 0$, ou seja,

$$(3.14) \quad V_S = \text{número de semi-somas } \frac{Z_i + Z_j}{2} > 0, \text{ com } i \leq j.$$

A prova do teorema é concluída observando as implicações:
 $\Delta_0 < \Delta_1 \Rightarrow Z'_i > Z_i \quad i=1, \dots, N \Rightarrow \text{número de } \frac{Z'_i + Z'_i}{2} > 0, i \leq j$
 $\geq \text{número de } \frac{Z'_i + Z'_j}{2} > 0 \text{ com } i \leq j \Rightarrow V'_S \geq V_S \Rightarrow (V_S \geq c \Rightarrow V'_S \geq c)$
 $\Rightarrow P_{\Delta_1}(V_S \geq c) = P_{\Delta_0}(V'_S \geq c) \geq P_{\Delta_0}(V_S \geq c).$

Consequências do Teorema 3

1. $\pi_E(\Delta) \geq \alpha, \forall F$ contínua e $\Delta > 0$, pois $\pi_E(0) = \alpha$.

Logo o teste de postos sinalizados de Wilcoxon é não vi-
ciado para alternativas de deslocamento.

2. O teste de postos sinalizados de Wilcoxon de nível α
para testar $H_0: \Delta=0$ tem, também nível α para tes-
tar $H'_0: \Delta \leq 0$.

§11. Aproximação para a Potência do Teste de Postos Sina-
lizados de Wilcoxon.

Para o cálculo da potência do teste vamos utili-
zar a aproximação normal que, pelo teorema abaixo, é vá-
lida.

Teorema 6 - A distribuição de $\frac{V_S - E(V_S)}{\sqrt{\text{Var}(V_S)}}$ tende para a
distribuição normal padrão, quando $N \rightarrow \infty$,
se a distribuição L dos Z 's satisfaz $0 < P(Z < 0) < 1$.

Prova: V. Lehmann, pag. 368, Exemplo 21.

Observação: Nos casos extremos onde $P(Z_1 > 0) = 0$ ou
 $P(Z_1 > 0) = 1$, V_S é constante com probabilidade
de 1. Obviamente, o teorema não se aplica neste caso.

Nas aplicações do Teorema 3, precisamos calcular

$E(V_S)$ e a $Var(V_S)$ sob a hipótese alternativa, isto é, quando L não é simétrica em torno de zero. Este cálculo é feito a seguir.

§12. Cálculo de $E(V_S)$ e $Var(V_S)$ quando L não é simétrica em torno de zero.

Suponhamos que a distribuição L dos Z 's é contínua. Já vimos em (3.14) que a estatística V_S do teste de Wilcoxon pode ser escrita

$$V_S = \text{número de médias } 1/2(Z_i + Z_j) > 0 \quad i \leq j.$$

Seja $\varphi(x,y) = 1$ se $x+y > 0$ e $\varphi(x,y) = 0$ caso contrário, então, $V_S = \sum_{i \leq j} \varphi(Z_i, Z_j)$. Com $p_1 = P(Z_1 > 0)$ e $p'_1 = P(Z_i + Z_j > 0)$, temos

$$\begin{aligned} (3.15) \quad E(V_S) &= E\left[\sum_{i < j} \varphi(Z_i, Z_j) + \sum_{i=1}^N \varphi(Z_i, Z_i)\right] = \\ &= \binom{N}{2} p'_1 + N p = 1/2 N(N-1) p'_1 + N p. \end{aligned}$$

A variância de V_S é dada por

$$(3.16) \quad Var(V_S) = \sum_{i \leq j} Var \varphi(Z_i, Z_j) + \sum_{\substack{i \leq j \quad k \leq l \\ (i,j) \neq (i,k)}} cov[\varphi(Z_i, Z_j), \varphi(Z_k, Z_l)].$$

Decompondo o primeiro somatório de (3.16),

$$\sum_{i \leq j} \text{Var } \varphi(Z_i, Z_j) = \sum_{i < j} \text{Var } \varphi(Z_i, Z_j) + \sum_{i=1}^N \text{Var } \varphi(Z_i, Z_i),$$

e sendo $\text{Var } \varphi(Z_i, Z_j) = p'_1(1-p'_1)$ para $i \neq j$ e $\text{Var } \varphi(Z_i, Z_i) = p(1-p)$, então, o primeiro somatório de (3.16) é igual a: $\binom{N}{2} p'_1(1-p'_1) + Np(1-p)$. Vamos, agora, calcular o segundo somatório de (3.16). Analisemos os diversos tipos de termos tendo valor comum. Consideremos, primeiro, os termos onde todos os índices i, j, k, ℓ são distintos. Como, neste caso, $\varphi(Z_i, Z_j)$ e $\varphi(Z_k, Z_\ell)$ são independentes e as suas covariâncias são nulas. Passemos ao segundo tipo de termo, com $i < j; k < \ell$, e onde três dos índices i, j, k, ℓ são distintos. Cada tripla ($i < j < k$) de inteiros distintos, neste caso, gera os seguintes pares: $(i, j), (i, k); (i, k), (i, j); (i, k), (j, k); (j, k), (i, k); (i, j), (j, k); (j, k), (i, j)$. O número de triplas distintas é $\binom{N}{3}$, cada uma gerando seis termos do somatório e fornecendo o total $6 \times \binom{N}{3} = N(N-1)(N-2)$ termos. É fácil verificar que as correspondentes covariâncias são iguais, sendo esta parte do somatório igual a $N(N-1)(N-2) \text{cov}[\varphi(Z_i, Z_j), \varphi(Z_k, Z_\ell)]$. Vejamos, agora, os termos onde $i=j$ ou $k=\ell$. Se $i=j$ e $k=\ell$ com $i = j = k = \ell$ então $(i, j) = (k, \ell)$ e o termo não está incluído no somatório. Se $i=j, i \neq k$ e $i \neq \ell$ então $\text{cov}[\varphi(Z_i, Z_i), \varphi(Z_k, Z_\ell)] = 0$ pois Z_i e (Z_k, Z_ℓ) são in-

dependentes. Restam, portanto, os termos onde um dos pares tem os elementos iguais e o outro um só elemento comum com aquele. Neste caso, cada par $(i < k)$ gera os pares: $(i, i), (i, k); (i, k), (i, i); (i, k), (k, k); (k, k), (i, k)$. O número total de termos deste tipo é $4 \times \binom{N}{2} = 2N(N-1)$.

Portanto, a soma total de termos deste tipo é $2N(N-1) \text{Cov}(V_{ij}, V_{ii})$, visto que as covariâncias são iguais nestes termos. Juntando os resultados obtidos acima, temos

$$\text{Var } V_S = \binom{N}{2} p_1' (1-p_1') + Np(1-p) + N(N-1)(N-2) \text{Cov}[\varphi(Z_i, Z_j), \varphi(Z_k, Z_j)] + 2N(N-1) \text{cov}[\varphi(Z_i, Z_j), \varphi(Z_i, Z_i)].$$

Calculamos a seguir o valor de cada covariância na expressão da $\text{Var } V_S$ acima:

$$\begin{aligned} & \text{cov}[\varphi(Z_i, Z_j), \varphi(Z_k, Z_j)] = \\ &= \text{Exp}(Z_i, Z_j) \varphi(Z_k, Z_j) - \text{Exp}(Z_i, Z_j) \text{Exp}(Z_k, Z_j) = \\ &= P[Z_i + Z_j > 0 \text{ e } Z_k + Z_j > 0] - P[Z_i + Z_j > 0] P[Z_k + Z_j > 0] = \\ &= p_2' - p_1'^2, \end{aligned}$$

onde $p_2' = P[Z_i + Z_j > 0 \text{ e } Z_k + Z_j > 0]$

$$\begin{aligned} \text{cov}[\varphi(Z_i, Z_j), \varphi(Z_i, Z_i)] &= P[Z_i + Z_j > 0 \text{ e } Z_i > 0] - P[Z_i + Z_j > 0] P[Z_i > 0] \\ &= p_2'' - p_1' p, \end{aligned}$$

onde $p_2'' = P[Z_i + Z_j > 0 \text{ e } Z_i > 0]$.

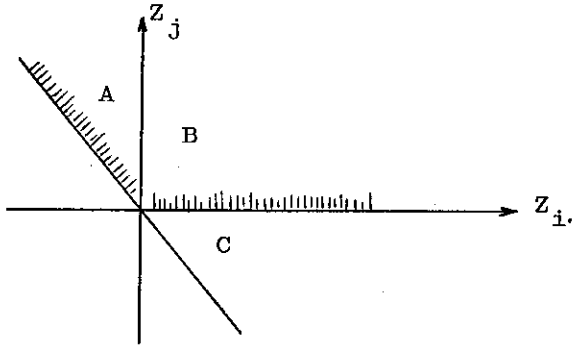
Podemos, ainda, escrever p''_2 em função de p'_1 e p .

Para isto observe que no gráfico abaixo, p''_2 representa a probabilidade de (Z_i, Z_j) tomar valores na área hachurada. Temos

$$p''_2 = P[Z_i + Z_j > 0 \text{ e } Z_i > 0] = P(A) + P(B)$$

$$P(B) = P[Z_i > 0 \text{ e } Z_j > 0] = P[Z_i > 0]P[Z_j > 0] = p^2$$

$$p'_1 = P[Z_i + Z_j > 0] = P(A) + P(B) + P(C).$$



Por outro lado:

$$P(A) = P[Z_i + Z_j > 0 \text{ e } Z_i < 0] \text{ e}$$

$$P(C) = P[Z_i + Z_j > 0 \text{ e } Z_j < 0]$$

Como Z_i e Z_j são independentes e têm a mesma distribuição $P(A) = P(C)$ o que acarreta $p''_2 = 1/2 (p'_1 + p^2)$.

A expressão final da $\text{Var}(V_S)$ é

(3.17)

$$\begin{aligned} \text{Var } V_S &= \frac{N(N-1)}{2} p'_1(1-p'_1) + Np(1-p) + N(N-1)(N-2)(p'_2 - p'_1)^2 + \\ &\quad + 2 N(N-1) [1/2 (p'_1 + p^2) - p'_1 p] = \\ &= N(N-1)(N-2)(p'_2 - p'_1)^2 + \frac{N(N-1)}{2} [2(p-p'_1)^2 - 3p'_1(1-p'_1)] \\ &\quad + Np(1-p). \end{aligned}$$

Exemplo 6: Mostre, a partir da fórmula (3.17), que se a distribuição dos Z_i 's é simétrica em torno de zero então $\text{Var } V_S = N(N+1)(2N+1)/24$.

Vamos supor que a distribuição E de Z_i tem densidade e . Para aplicar a fórmula (3.17) precisamos achar p_1 , p'_1 e p'_2 . Como a distribuição de Z_i é simétrica em torno de zero, $p_1 = P(Z_i > 0) = 1/2$. Para o cálculo de p'_1 , observe na Figura 1 a região de integração. Temos,

$$\begin{aligned} p'_1 = P(Z_i + Z_j > 0) &= \int_{-\infty}^{+\infty} \int_{-z_i}^{+\infty} e(z_j) e(z_i) dz_j dz_i = \\ &= \int_{-\infty}^{+\infty} [1 - E(-z_i)] e(z_i) dz_i = \int_{-\infty}^{+\infty} E(z_i) e(z_i) dz_i \\ &= 1/2 [E(z_i)]^2 \Big|_{-\infty}^{+\infty} = 1/2. \end{aligned}$$

No cálculo acima foi utilizado: $E(+\infty) - E(-z_i) = 1 - E(-z_i) = E(z_i)$, pois a distribuição E é simétrica em torno de zero.

Determinação de $p'_2 = P(Z_1 + Z_j > 0 \text{ e } Z_k + Z_j > 0)$:

$$\begin{aligned} p'_2 &= \int_{-\infty}^{+\infty} \int_{-z_i}^{+\infty} \int_{-z_i}^{+\infty} e(z_j) e(z_k) e(z_i) dz_j dz_k dz_i = \\ &= \int_{-\infty}^{+\infty} (1 - E(-z_i))^2 e(z_i) dz_i = \int_{-\infty}^{+\infty} [E(z_i)]^2 e(z_i) dz_i = \\ &= 1/3 [E(z_i)]^3 \Big|_{-\infty}^{+\infty} = 1/3. \end{aligned}$$

Substituindo estes valores na fórmula (3.17), obtemos $\text{Var } V_S = N(N+1)(2N+1)/24$. Este resultado é ainda válido quando E não tem densidade. Neste caso, teríamos de utilizar a integral de Riemann-Stieltjes em lugar da integral de Riemann.

Exemplo 7: Calcule, para o modelo de deslocamento e sendo a distribuição dos Z's uniformes, a esperança e a variância de V_S .

Neste caso, a distribuição sob a hipótese é $U[-1/2, 1/2]$, e sob a alternativa é $U[\Delta-1/2, \Delta+1/2]$.

Designemos por I a função indicadora de $[\Delta-1/2, \Delta+1/2]$, dada por $I(x) = 1$ se $x \in [\Delta-1/2, \Delta+1/2]$ e $I(x) = 0$ caso contrário.

Cálculo de p , p'_1 e p''_2 :

$$p = P[Z_1 > 0] = \int_0^{+\infty} I(z_1) dz_1 = \Delta + 1/2, \quad \text{para } 0 < \Delta < 1/2,$$

$$p'_1 = P[Z_1 + Z_j > 0] = \int_{-\infty}^{+\infty} \int_{-z_i}^{+\infty} I(z_j) I(z_i) dz_j dz_i \quad e$$

$$\int_{-z_i}^{+\infty} I(z_j) dz_j = \begin{cases} 0 & \text{se } z_i \leq -\Delta - 1/2 \\ \Delta + 1/2 + z_i & \text{se } -\Delta - 1/2 < z_i < 1/2 - \Delta \\ 1 & \text{se } z_i \geq 1/2 - \Delta \end{cases}$$

Logo,

$$p'_1 = \int_{-\infty}^{-\Delta-1/2} 0 \cdot I(z_i) dz_i + \int_{-\Delta-1/2}^{1/2-\Delta} (\Delta + 1/2 + z_i) I(z_i) dz_i + \int_{1/2-\Delta}^{+\infty} 1 \cdot I(z_i) dz_i.$$

Por outro lado,

$$\begin{aligned} \int_{-\Delta-1/2}^{1/2-\Delta} (\Delta + 1/2 + z_i) I(z_i) dz_i &= \int_{-\Delta-1/2}^{1/2-\Delta} (\Delta + 1/2 + z_i) dz_i = \\ &= 1/2 - 2\Delta^2 \end{aligned}$$

e

$$\int_{1/2-\Delta}^{+\infty} I(z_i) dz_i = \int_{1/2-\Delta}^{\Delta+1/2} dz_i = 2\Delta \quad \text{e substituindo estes}$$

valores, obtemos $p'_1 = 1/2 + 2\Delta(1-\Delta)$.

Finalmente, para p'_2 temos

$$\begin{aligned}
 p'_2 &= P[Z_i + Z_j > 0 \text{ e } Z_k + Z_j > 0] = \\
 &= \int_{-\infty}^{+\infty} \int_{-z_i}^{+\infty} \int_{-z_i}^{+\infty} I(z_j) I(z_k) I(z_i) dz_j dz_k dz_i \\
 &= \int_{-\infty}^{+\infty} \left(\int_{-z_i}^{+\infty} I(z_j) dz_j \right)^2 I(z_i) dz_i = \int_{-\infty}^{-\Delta-1/2} 0 \cdot I(z_i) dz_i + \\
 &+ \int_{-\Delta-1/2}^{1/2-\Delta} (\Delta+1/2+z_i)^2 I(z_i) dz_i + \int_{1/2-\Delta}^{+\infty} 1 \cdot I(z_i) dz_i = \\
 &= \int_{\Delta-1/2}^{1/2-\Delta} (\Delta+1/2+z_i)^2 dz_i + \int_{1/2-\Delta}^{\Delta+1/2} dz_i = 1/3 + 2\Delta(1-4\Delta^2).
 \end{aligned}$$

A partir dos valores de p , p'_1 e p'_2 , utilizando (3.15) e (3.17), calculamos EV_S e $Var V_S$.

§13. Estimação do efeito de tratamento.

Suponhamos que as variáveis aleatórias Z_1, \dots, Z_N , do modelo descrito anteriormente, tenham distribuição L , simétrica em torno de θ . Queremos estimar θ , centro de simetria da distribuição comum de Z_1, \dots, Z_N . O estima-

dor clássico utilizado para resolver este problema, quando a distribuição comum L dos Z 's é normal, é

$$\bar{\theta} = \frac{Z_1 + \dots + Z_N}{N}.$$

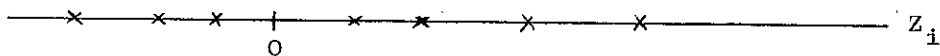
Neste caso, $\bar{\theta}$ é um estimador não viciado e uniformemente de mínima variância. Este estimador é, contudo, muito sensível a valores "excentricos" (outliers) e pouco eficiente para distribuições com caudas grossas. Quando L é a distribuição de Cauchy, por exemplo, a distribuição de $\bar{\theta}$, para qualquer N , é a mesma que de uma só observação Z_i . Desta forma, estimar θ por $\bar{\theta}$ é tão preciso quanto estimá-lo por uma única observação Z_i , desprezando toda informação dada pelas outras observações.

A seguir, utilizando o método de inversão de testes apresentado no Capítulo 2, obtemos estimadores para θ ; a partir do teste do sinal e do teste de postos sinalizados de Wilcoxon.

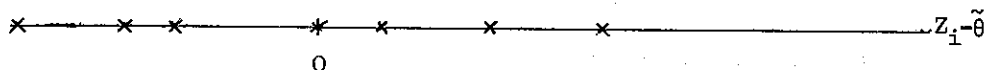
Estimador derivado do teste do sinal.

As variáveis $Z_1 - \theta, \dots, Z_n - \theta$ são simétricas em torno de zero. Tomemos, então, para estimador de θ o valor $\tilde{\theta}$, de modo que $Z_1 - \tilde{\theta}, \dots, Z_n - \tilde{\theta}$ forneçam, de acordo com teste do sinal, o máximo de evidência em favor da hipótese de simetria da distribuição de $Z_i - \theta$ em torno de zero. Tal ocorre quando o número de valores $Z_i - \tilde{\theta}$ po-

sitivos é igual ao número de valores $Z_i - \tilde{\theta}$ negativos. Ou seja, quando $\tilde{\theta} = \text{med}(Z_i)$. Para ilustrar a idéia acima, considere a disposição dos Z_i 's dada na figura abaixo.



A disposição abaixo fornece, segundo o teste do sinal, máximo apoio à hipótese de simetria em torno de zero.



Estimador derivado do teste de postos sinalizados de Wilcoxon.

Neste caso, procuramos o valor $\hat{\theta}$ que fornece para a estatística de Wilcoxon V_S , baseada em $Z_1 - \hat{\theta}, \dots, Z_n - \hat{\theta}$, o valor de maior evidência em favor da hipótese de simetria em torno de zero. Isto é equivalente a tomar $\hat{\theta}$ de modo que metade das somas $(Z_i - \hat{\theta}) + (Z_j - \hat{\theta})$ $i \leq j$ sejam positivas e metade negativas. Como $(Z_i - \hat{\theta}) + (Z_j - \hat{\theta}) = (Z_i + Z_j) - 2\hat{\theta}$, basta tomar $2\hat{\theta} = \text{med}_{i \leq j} [Z_i + Z_j]$ ou

$$(3.18) \quad \hat{\theta} = \text{med}_{i \leq j} [1/2(Z_i + Z_j)].$$

Comentários sobre os estimadores $\hat{\theta}$ e $\tilde{\theta}$.

1. Quando a distribuição L de Z_1, \dots, Z_N é simétrica em torno de θ , as distribuições de $\hat{\theta}$ e $\tilde{\theta}$ são simétricas em torno de θ . (Fácil de demonstrar).
2. Os estimadores $\hat{\theta}$ e $\tilde{\theta}$ são pouco sensíveis a valores excêntricos (outliers), portanto, mais robustos que $\bar{\theta}$.
3. Uma comparação entre as concentrações das distribuições dos estimadores $\bar{\theta}$, $\tilde{\theta}$ e $\hat{\theta}$ em torno de θ é feita considerando as probabilidades: $P[|\bar{\theta}-\theta|<a]$, $P[|\tilde{\theta}-\theta|<a]$ e $P[|\hat{\theta}-\theta|<a]$, para diferentes valores de a . O resultado da comparação é o mesmo obtido na comparação entre os testes correspondentes: t-Student, Sinal e Wilcoxon. (V. Lehmann pag. 173-174).

§14. Intervalo de confiança para o efeito de tratamento.

Um intervalo de confiança para θ , o centro de simetria da distribuição L dos Z 's, pode ser derivado a partir do teste do sinal ou do teste de postos sinalizados de Wilcoxon. Para isto podemos utilizar o mesmo método adotado no problema de duas amostras: construção do

intervalo a partir da região de aceitação do teste.

De início, consideremos o problema de testar a hipótese $H_0: \theta = \theta_0$ contra a alternativa $H_A: \theta \neq \theta_0$. Tal problema reduz-se ao de testar $H'_0: \theta = 0$ contra $H'_A: \theta \neq 0$, considerando-se as observações transformadas: $Z_1 - \theta_0, \dots, Z_N - \theta_0$. Denotemos por $S(Z - \theta_0)$ a estatística do teste do sinal calculada a partir dos valores de $Z_1 - \theta_0, \dots, Z_N - \theta_0$. A região de aceitação do teste do sinal para a hipótese $H_0: \theta = \theta_0$ contra a alternativa $H_A: \theta \neq \theta_0$ é dada por

$$(3.19) \quad k < S_N(Z - \theta_0) < N - k,$$

onde $1 \leq k \leq N$ e $k < N - k$. O nível de significância α deste teste satisfaz

$$(3.20) \quad P_{\theta_0} [k < S_N(Z - \theta_0) < N - k] = 1 - \alpha.$$

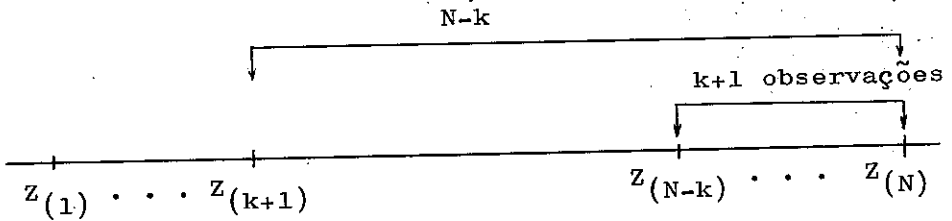
Vamos agora, determinar os valores de θ_0 que satisfazem a desigualdade (3.19) acima. Antes porém, é importante observar que a igualdade (3.20) independe de θ_0 :

$$P_{\theta_0} [k < S_N(Z - \theta_0) < N - k] = P_0 [k < S_N(Z) < N - k], \quad \forall \theta_0.$$

A notação P_{θ_0} indica que a distribuição dos Z's é simétrica em torno de θ_0 ; P_0 indica que esta distribuição é simétrica em torno de zero..

Sejam $Z_{(1)} < Z_{(2)} < \dots < Z_{(N)}$ os valores dos Z's ordenados crescentemente. A desigualdade $S_N(Z - \theta_0) > k$ é sa-

tisfeita se no mínimo $k+1$ diferenças $(Z_i - \theta_0)$ são positivas; equivalentemente, no mínimo $(k+1)$ dos Z_i 's são maiores que θ_0 . Isto ocorre quando $\theta_0 < Z_{(N-k)}$.



Por outro lado, $S_N(Z - \theta_0) < N-k$ indica que o número de Z_i 's maiores que θ_0 deve ser no máximo $N-k-1$. Isto ocorre quando $\theta_0 \geq Z_{(k+1)}$. Juntando as condições acima, obtemos

$$(3.21) \quad P_{\theta_0} [Z_{(k+1)} \leq \theta_0 < Z_{(N-k)}] = P_{\theta_0} [k < S_N(Z - \theta_0) < N-k] \\ = P_{\theta_0} [k < S_N(Z) < N-k] = 1-\alpha, \quad \forall \theta_0, \quad \forall L \text{ contínua e}$$

simétrica em torno de θ_0 . Deste modo, $(Z_{(k+1)}; Z_{(N-k)})$ é um intervalo de confiança com nível de confiança $(1-\alpha)$ para θ .

Exemplo 8: Os dados abaixo foram obtidos a partir de uma tabela de desvios normais, somando-se 10 a cada valor selecionado. Determine um intervalo de confiança de aproximadamente 95% de nível de confiança para o centro de simetria da distribuição dos Z 's, cujos valores

observados são 10,46; 10,14; 12,45; 9,68; 9,93; 10,30; 9,72; 11,30; 10,24; 9,04; 10,06; 7,47; 9,47; 9,81; 8,44; 10,19. Para $N = 16$, obtemos de uma tabela binomial com $p = 1/2$: $P(S_N \leq 4) = 0,0384$ e $P(S_N \leq 3) = 0,0106$, portanto, $P[4 < S_N < 12] = 1 - 2 \times 0,0384 = 0,923$. O intervalo de confiança com 92% de nível de confiança é $(Z_{(5)}, Z_{(12)})$. Os valores de $Z_{(5)}$ e $Z_{(12)}$ na amostra são 9,68 e 10,24. Considerando-se $(Z_{(4)}, Z_{(13)})$ obtemos um nível de confiança de $1 - 2 \times 0,0106 = 0,978$, ou seja, aproximadamente 98%. Escrevendo as observações em ordem crescente determinamos $Z_{(4)}$ e $Z_{(13)}$: $Z_{(4)} = 9,47$ e $Z_{(13)} = 10,30$.

A partir do teste dos postos sinalizados de Wilcoxon podemos, também, construir intervalos de confiança para θ . Como antes, consideremos o seguinte problema de teste de hipótese: testar $H_0: \theta = \theta_0$ contra $H_A: \theta \neq \theta_0$. Denotemos por $V_S(Z - \theta_0)$ a estatística do teste de postos sinalizados de Wilcoxon calculado a partir dos valores de $Z_1 - \theta_0, \dots, Z_N - \theta_0$. A região de aceitação do teste da hipótese $H_0: \theta = \theta_0$ contra a alternativa $H_A: \theta \neq \theta_0$ é dada por $i < V_S(Z - \theta_0) < M - i$, onde $M = N(N+1)/2$ e $0 \leq i < M - i$. O nível de significância deste teste é α , satisfazendo

$$(3.22) \quad P_{\theta_0} [i < V_S(Z - \theta_0) < M - i] = 1 - \alpha.$$

A distribuição de $Z_i - \theta_0$ é simétrica em torno de zero quando a de Z_i é simétrica em torno de θ_0 . Desta forma, a igualdade acima independe de θ_0 e temos,

$$(3.23) \quad P_{\theta_0} [i < V_S(Z - \theta_0) < M - i] = P_0 [i < V_S(Z) < M - i] = 1 - \alpha, \forall \theta_0$$

e $\forall L$ contínua; L simétrica em torno de θ_0 . Denotemos por $A_{(1)} < A_{(2)} < \dots < A_{(M)}$ as $M = N(N+1)/2$ semi-somas $(Z_j + Z_k)/2$, $j \leq k$, ordenadas crescentemente. A

desigualdade $V_S(Z - \theta_0) > i$ equivale a dizer que o número de semi-somas $\left[\frac{(Z_j - \theta_0) + (Z_k - \theta_0)}{2} \right]_{j \leq k}$ positivas é maior

que i . Ou seja, o número de semi-somas $\left(\frac{Z_j + Z_k}{2} \right)_{j \leq k}$ maiores que θ_0 é no mínimo $i+1$, portanto, devemos ter

$\theta_0 < A_{(M-i)}$. Por outro lado, $V_S(Z - \theta_0) < M - i$ implica que

$\theta_0 \geq A_{(i+1)}$. Colocando juntas as duas condições, obtemos

o intervalo de confiança procurado: $(A_{(i+1)}, A_{(M-i)})$.

Exercícios - Capítulo III

1. Obtenha a distribuição nula de S_N , estatística do teste do sinal, quando (i) $N=4$; (ii) $N=5$. Desenhe os histogramas correspondentes.
2. Determine quando $N=5$ as probabilidades
(i) $P_H(n=3; S_1=1, S_2=2, S_3=3)$; (ii) $P_H(n=3; S_1=1)$;
(iii) $P_H(n \geq 1; S_1=1)$.
3. Determine a probabilidade $P_H(n=3; S_1 + \dots + S_n=10)$ quando (i) $N=5$; (ii) $N=6$; (iii) $N=7$.
4. Determine quando $N=5$ as probabilidades
(i) $P_H(S_1 + \dots + S_n=10)$; (ii) $P_H(S_1+2S_1 + \dots + nS_n=10)$.
5. Considere o teste de postos sinalizados de Wilcoxon, para testar nulidade de efeito contra tendência de aumentar as respostas. Utilizando aproximação normal, determine a probabilidade de significância do teste, sabendo que $N=10$ e a soma dos postos sinalizados do tratamento é 47.
6. Prove que as distribuições nulas de V_S e V_r são iguais.
7. Utilizando o resultado do problema 6, mostre que a distribuição nula de V_S é simétrica em torno de $N(N+1)/4$.

8. Determine a distribuição nula de V_S^* e desenhe o seu histograma quando $N=5$ e (i) $d_0=0, d_1=d_2=2, d_3=1$;
(ii) $d_0=1, d_1=2, d_2=d_3=1$; (iii) $d_0=2, d_1=d_2=d_3=1$;
(iv) $d_0=1, d_1=d_2=2$.
9. Mostre que $P_H(V_S^* \leq 4) = 4/32$ quando $d_0=1, d_1=2, d_2=1, d_3=2$.
10. Determine o número de observações necessárias para obter resultado significativo com probabilidade de 0.90 se $\Delta=0.25$, no teste unilateral do sinal tendo $\alpha=0.05$. Suponha $F = U[-1/2, 1/2]$, distribuição uniforme entre $-1/2$ e $1/2$.
11. Determine o número de observações necessárias para o teste de postos sinalizados de Wilcoxon, nas condições do Exercício 10.
12. Sejam Z_1, \dots, Z_N independentes e com distribuição comum L simétrica em torno de θ . Considere os seguintes estimadores de θ : $\hat{\theta} = \text{med} [1/2(Z_i + Z_j)]$;
 $\tilde{\theta} = \text{med}_{i=1, \dots, N} (Z_i)$ e $\bar{\theta} = \frac{1}{N} \sum_{i=1}^N Z_i$. Mostre que as distribuições dos erros $\hat{\theta} - \theta$, $\tilde{\theta} - \theta$ e $\bar{\theta} - \theta$ independem de θ .
13. Nas condições do Problema 12, mostre que as distribuições de $\hat{\theta}$, $\tilde{\theta}$ e $\bar{\theta}$ são simétricas em torno de θ .
Sugestão: Sem perda de generalidade faça $\theta=0$. Verifique que $\hat{\theta}(-Z_1, \dots, -Z_N) = -\hat{\theta}(Z_1, \dots, Z_N)$.

Capítulo IV

Problema de s-amostras - Blocos Aleatorizados Completos

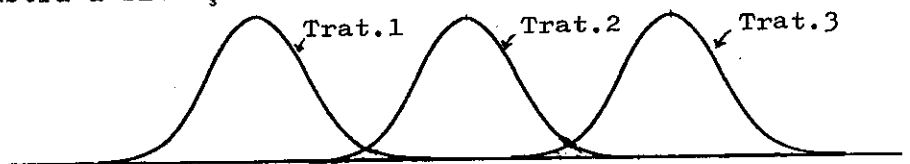
§1. Introdução.

Neste capítulo abordamos, de maneira bem sucinta, o problema da comparação de mais de dois tratamentos. Este problema é uma generalização natural do de duas amostras. Surge, por exemplo, quando queremos comparar três ou mais marcas de gasolina; três ou mais dietas; vários métodos de aprendizagem, etc.

Uma questão simples que pode ser levantada nesta comparação é a da existência de qualquer diferença entre os tratamentos. Antes de introduzirmos um modelo estatístico para analisar esta questão, vamos supor que diferenças entre os tratamentos se manifestam no nível geral das respostas. Isto é, no caso de haver diferenças, existe uma ordem: um tratamento tende a fornecer maiores respostas; outro as próximas maiores respostas, etc. Na construção do modelo estatístico somos ainda mais específicos, considerando, apenas, diferenças nas locações das distribuições das respostas.

Suponhamos que $N = n_1 + \dots + n_s$ indivíduos são retirados de uma população por meio de uma amostra aleatória simples e n_1 destes recebem o tratamento 1, n_2 o tratamento 2, etc.. Consideremos o seguinte modelo estatístico: as respostas do tratamento 1 são valores de variáveis aleatórias independentes X_{11}, \dots, X_{1n_1} com distribuição comum F_1 ; as do tratamento 2 são os valores de variáveis aleatórias independentes X_{21}, \dots, X_{2n_2} , com distribuição comum F_2 e assim por diante.

Mais especificamente, vamos tomar $X_{ij} = \alpha + \mu_i + \epsilon_{ij}$, onde α é um efeito geral; μ_i o efeito do i -ésimo tratamento, $i=1, \dots, s$ e os ϵ_{ij} $i=1, \dots, s$; $j=1, \dots, N$, são variáveis aleatórias independentes, identicamente distribuídas e com distribuição contínua. Deste modo, fica clara a semelhança com o modelo mais simples de análise de variância (one way-layout), onde ainda é feita a hipótese dos ϵ_{ij} terem distribuição $N(0, \sigma^2)$. Em termos das distribuições F_1, \dots, F_s , este modelo supõe que o efeito de cada tratamento é adicionar uma constante às respostas dos controles, independentemente do valor desta resposta, ou seja: $F_i(x) = F_0(x + \mu_i)$. A figura abaixo ilustra a situação.



Queremos testar a hipótese de nenhuma diferença entre os s tratamentos,

$H : \mu_1 = \mu_2 = \dots = \mu_s$, contra a alternativa

H_A : os μ_i 's não são todos iguais.

Sejam R_{11}, \dots, R_{1n_1} os postos dos n_1 indivíduos que receberam o tratamento 1; R_{21}, \dots, R_{2n_2} os postos dos que receberam o tratamento 2, e assim por diante. Todos estes, postos determinados na amostra global de $N = n_1 + \dots + n_s$, indivíduos. Vamos upor, ainda, que dentro de cada grupo estes postos estão ordenados: $R_{11} < \dots < R_{1n_1}, R_{21} < \dots < R_{2n_2}$, etc.

O número de maneiras possíveis de alocar n_1 dos N postos para o tratamento 1, n_2 para o tratamento, etc., é

$$\binom{N}{n_1} \binom{N-n_1}{n_2} \dots \binom{N-n_1-\dots-n_{s-2}}{n_{s-1}}, \text{ que denotamos por } \binom{N}{n_1, \dots, n_s}, \text{ e é chamado coeficiente multinomial.}$$

Sob a hipótese H , a distribuição dos postos é

$$(4.1)$$

$$P_H(R_{11}=r_{11}, \dots, R_{1n_1}=r_{1n_1}; \dots; R_{s1}=r_{s1}, \dots, R_{sn_s}=r_{sn_s}) = \frac{1}{\binom{N}{n_1, \dots, n_s}}.$$

Estamos supondo, aqui, que sob a hipótese a distribuição comum $F_1 = F_2 = \dots = F_s = F$ é contínua e, por-

tanto, a probabilidade de empates é igual a zero. Esta fórmula expressa que todas as $\binom{N}{n_1, \dots, n_s}$ partições possíveis do conjunto $\{1, \dots, N\}$ em s conjuntos de tamanhos n_1, \dots, n_s são igualmente prováveis, sob a hipótese H . A demonstração é idêntica à do caso de duas amostras.

§2. Teste de Kruskal-Wallis.

No modelo descrito, supomos a existência de uma ordenação entre os tratamentos. Uma indicação da posição do i -ésimo tratamento, nesta ordenação, é dada pela média de seus postos: $R_{i.} = \frac{R_{i1} + \dots + R_{in_i}}{n_i}$.

Uma evidência contra a hipótese H é fornecida por uma grande dispersão entre estas médias; espera-se, sob a hipótese H que estas médias sejam próximas umas das outras e, portanto, próximas da média geral:

$$R_{..} = \frac{(R_{11} + \dots + R_{1n_1}) + \dots + (R_{s1} + \dots + R_{sn_s})}{N} = \frac{(N+1)}{2}$$

Um índice desta proximidade entre os $R_{i.}$'s e $R_{..}$ é dado por $\sum c_i [R_{i.} - 1/2(N+1)]^2$ onde os c_i 's são pesos adequados. Dentro desta idéia, foi proposta por Kruskal -Wallis a seguinte estatística:

$$(4.2) \quad K = \frac{12}{N(N+1)} \sum_{i=1}^s n_i \left(R_{i.} - \frac{N+1}{2} \right)^2.$$

O teste de Kruskal -Wallis rejeita a hipótese H de igualdade de efeitos dos tratamentos quando

$$K \geq c,$$

onde c é determinado por $P_H(K \geq c) = \alpha$; α é o nível de significância fixado para o teste. A determinação da constante crítica do teste c é feita a partir da distribuição de K , sob a hipótese H. Esta distribuição pode ser obtida da distribuição (4.1) dos postos

$R_{11}, \dots, R_{1n_1}; \dots; R_{s1}, \dots, R_{sn_s}$ enumerando para cada valor de K os casos favoráveis à sua ocorrência. Esta tarefa, mesmo para N pequeno, é bastante trabalhosa. Algumas tabelas são disponíveis para pequenos valores de N , suplementadas pela seguinte propriedade de limite:

$$P_H(K < c) \rightarrow \chi_{s-1}^2(c), \text{ quando } \min\{n_1, \dots, n_s\} \rightarrow \infty, \text{ onde}$$

χ_{s-1}^2 é a função de distribuição comutativa de uma qui-quadrado com $s-1$ graus de liberdade.

Uma forma de K mais apropriada para os cálculos é

$$(4.3) \quad K = \frac{12}{N(N+1)} \sum_{i=1}^s \frac{R_i^2}{n_i} - 3(N+1),$$

onde $R_i = \sum_{j=1}^{n_i} R_{ij}$ e $i=1, \dots, s$.

Até este ponto, supomos a não ocorrência de empates. Quando estes ocorrem podemos usar os postos médios, definidos como a média dos postos das respostas empatadas. Consideremos a seguinte configuração de empates: as N respostas tomam e valores distintos sendo d_1 destas iguais ao menor valor; d_2 iguais ao próximo menor, etc. Denotemos os postos médios das respostas do i -ésimo tratamento por $R_{i1}^*, \dots, R_{in_i}^*$ e sua soma por R_i^* . A estatística de teste apropriada, no caso, é

$$(4.4) \quad K^* = \frac{[12/N(N+1)] \sum R_i^{*2}/n_i - 3(N+1)}{1 - \sum (d_i^3 - d_i)/(N^3 - N)}$$

Rejeita-se a hipótese de nenhuma diferença entre os tratamentos quando $K^* \geq c$. A distribuição nula (sob H) de K^* pode ser obtida calculando-se a probabilidade de cada arranjo possível dos postos e o correspondente valor de K^* . Obviamente, este processo de enumeração exaustiva é muito trabalhoso. A distribuição nula de K^* tem também como limite uma distribuição $\chi^2_{(s-1)}$.

Observação: Quando todas as respostas são distintas,

$d_1 = \dots = d_e = 1$, $\sum (d_i^2 - d_i) = 0$ e a expressão (4.4) se reduz a (4.3).

Exemplo 1: Os valores indicados abaixo, para as respostas dos tratamentos I, II, III e IV foram simulados

segundo uma distribuição $N(30;100)$. (V. pag. 69 Cochran e Snedecor).

Tratamentos

I	II	III	IV
33	32	39	17
53	31	34	22
34	11	33	20
29	30	33	19
39	19	33	3
57	24	39	21
12	53	36	3
24	44	32	25
39	19	32	40
36	30	30	21

Vamos testar a hipótese de igualdade de efeitos entre os tratamentos. A tabela abaixo indica os postos médios das respostas.

Postos Médios

I	II	III	IV
25,5	22	33,5	5
38,5	20	28,5	12
28,5	3	25,5	9
16	18	25,5	7
33,5	7	25,5	1,5
40	13,5	33,5	10,5
4	38,5	30,5	1,5
13,5	37	22	15
33,5	7	22	36
30,5	18	18	10,5
263,5	184	264,5	108

Uma verificação pode ser feita notando que

$$R_1^* + R_2^* + R_3^* + R_4^* = 1/2 N(N+1) = 820.$$

O numerador de K^* é

$$\frac{12}{N(N+1)} \sum_i \frac{R_i^{*2}}{n_i} - 3(N+1) = 12,302.$$

Temos os seguintes valores dos d_i 's: 11 valores iguais a 1, 6 valores iguais a 2, 3 valores iguais a 3 e 2 valores iguais a 4. Deste modo,

$$\frac{\sum (d_i^3 - d_i)}{N^3 - N} = 0,0036$$

e obtemos $K^* = 12,346$. Utilizando uma tabela qui-quadrado com 3 g.l. obtemos a probabilidade de significância do teste: $0,0059 < P(K^* \geq 12,346) < 0,0074$. A hipótese seria rejeitada, mesmo, ao nível de 1%.

O teste de Kruskal -Wallis reduz-se, no caso $s=2$, ao teste bilateral de Wilcoxon, do problema de duas amostras. Em seguida, apresentamos no caso de s -amostras, um procedimento adequado para detectar alternativa do tipo unilateral. No caso de s -amostras, uma alternativa unilateral é dada por uma ordenação fixa dos tratamentos, tal como: o 2º melhor que o 1º, o terceiro melhor que o 2º, etc. Em termos do modelo estatístico apresentado, o problema correspondente é de testar a hipótese

$$H_0: \mu_1 = \mu_2 = \dots = \mu_s \quad \text{contra alternativas da forma}$$

$$H_A: \mu_1 \leq \mu_2 \leq \dots \leq \mu_s.$$

O teste de Kruskal -Wallis não é apropriado, nesta situação, pois rejeita H_0 sempre que as médias R_i dos postos são bastante desiguais, não levando em consideração a sua ordem. Aqui, só tendências crescentes dos R_i seriam evidências favoráveis à alternativa H_A .

§3. Teste de Jonckheere-Terpstra.

Sejam X_{i1}, \dots, X_{in_i} e X_{j1}, \dots, X_{jn_j} , respectivamente, as respostas do i -ésimo e do j -ésimo tratamentos. Consideremos a estatística de Mann-Whitney: W_{ij} = número de pares (k, ℓ) tal que $X_{ik} < X_{j\ell}$, para os tratamentos i e j . A estatística do teste de Jonckheere-Terpstra é a soma

$$(4.5) \quad W = \sum_{i < j} W_{ij}.$$

O teste rejeita H_0 para valores grandes de W ou, equivalentemente, para valores grandes de $J = 2W - \sum_{i < j} n_i n_j$, cuja distribuição encontra-se tabelada para $n_1 = \dots = n_s = n$ e várias combinações de valores de s e n . Quando N é grande, a distribuição de W pode ser aproximada pela

distribuição normal. Neste caso, é útil saber que

$$(4.6) \quad E_{HW} = \frac{1}{2} \sum_{i < j} n_i n_j = \frac{(N^2 - \sum n_i^2)}{4}$$

e

$$(4.7) \quad \text{Var}_H(W) = \frac{N^2(2N+3) - \sum n_i^2(2n_i+3)}{72}$$

Exemplo 2: As respostas dos tratamentos indicados nas tabelas abaixo são valores simulados de distribuições: $N(4,1)$; $N(5,1)$ e $N(6,1)$. Testar $H_0: \mu_1 = \mu_2 = \mu_3$ contra $H_A: \mu_1 \leq \mu_2 \leq \mu_3$.

Tratamentos

I	II	III
3,49	5,00	6,33
3,00	3,72	7,11
4,97	3,21	7,07
4,98	4,01	6,77
2,90	3,64	6,23
4,25	4,12	5,70
5,26	4,84	7,06
3,07	4,17	6,16
3,77	4,19	5,87
3,42	3,65	4,80

Os valores da estatística de Mann-Whitney são $W_{12} = 58$,

$W_{13} = 97$ e $W_{23} = 98$. Substituindo os valores $N = 30$ e $n_1 = n_2 = n_3 = 10$, calculamos $EW = 150$ e $\text{Var } W = 691.67$. A probabilidade de significância do teste é

$$P(W \geq 253) \approx 1 - \Phi\left(\frac{253-150}{26.30}\right),$$

praticamente igual a zero.

§4. Blocos Aleatorizados Completos.

Na comparação de s tratamentos, tratada na seção anterior, a grande variabilidade das respostas pode causar o mesmo tipo de dificuldade já discutida para o caso de duas amostras. A dificuldade foi, então, contornada utilizando-se o esquema experimental de emparelhamentos. De modo análogo, no caso de s -tratamentos, podemos dividir os indivíduos em blocos homogêneos de s -indivíduos e fazer a comparação dos tratamentos entre os indivíduos de cada bloco. Como exemplos práticos, temos: os animais de uma mesma ninhada; canteiros de uma plantação situados próximos; diferentes laboratórios; diferentes professores, etc. Na verdade, o que tratamos nesta seção é o análogo não-paramétrico do modelo de análise de variância com dois fatores controlados (Two-way Layout).

Suponhamos, então, que sN indivíduos são escolhidos aleatoriamente de uma população, s destes em cada um dos N blocos. Dentro de cada bloco, indivíduos são selecionados ao acaso, sendo um para cada tratamento, e submetidos aos s -tratamentos. Vamos considerar o seguinte modelo estatístico:

		Blocos			
Tratamentos		1	2	. . .	N
1		X_{11}	X_{12}	. . .	X_{1N}
2		X_{21}	X_{22}	. . .	X_{2N}
.		.			
.		.			
.		.			
s		X_{s1}	X_{s2}	. . .	X_{sN}

com

$$(4.8) \quad X_{ij} = \delta + \alpha_i + \beta_j + e_{ij} \quad \begin{matrix} i = 1, \dots, s \\ j = 1, \dots, N \end{matrix}$$

onde, δ é a média geral, desconhecida; α_i é o efeito do i -ésimo bloco e β_j o efeito do j -ésimo tratamento.

Podemos tomar $\sum_{i=1}^s \alpha_i = 0$ e $\sum_{j=1}^N \beta_j = 0$. Os erros e_{ij} são independentes e têm a mesma distribuição contínua.

Queremos, então, testar a hipótese:

$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_s$ contra a alternativa

H_A : os α 's não são todos iguais.

§5. Teste de Friedman.

No modelo considerado, supomos que os efeitos dos tratamentos afetam preponderantemente o nível geral das respostas. Deste modo, existe uma ordenação entre os tratamentos, um tendendo a produzir as menores respostas; outro as próximas menores, etc.

Uma indicação de posição de cada tratamento nesta ordenação é dada pela média de seus postos, nos N diferentes blocos:

$$R_{i.} = \frac{R_{i1} + \dots + R_{iN}}{N},$$

onde $R_{i,j}$ é o posto do i -ésimo tratamento dentro do j -ésimo bloco.

Diferenças entre os efeitos dos tratamentos se refletem em diferenças entre os $R_{i.}$'s. Desta forma, valores próximos dos $R_{i.}$'s apoiam H_0 ; valores muito distintos fornecem evidência a favor de H_A . No primeiro caso, estes valores seriam, também, próximos de

$$R_{..} = \frac{(R_{11} + \dots + R_{1N}) + \dots + (R_{s1} + \dots + R_{sN})}{sN} =$$

$$= \frac{N \times 1/2 s(s+1)}{sN} = 1/2 (s+1).$$

Uma medida da proximidade entre os valores dos $R_{i.}$'s é a estatística

$$(4.9) \quad Q = \frac{12N}{s(s+1)} \sum_{i=1}^s [R_{i.} - 1/2 (s+1)]^2.$$

O valor de Q é igual a zero quando os $R_{i.}$'s são iguais entre si, e é grande quando existe diferença substancial entre os $R_{i.}$'s. Assim sendo, rejeitamos H_0 quando

$$Q \geq c.$$

Uma expressão alternativa para Q , mais simples para os cálculos, é

$$(4.10) \quad Q = \frac{12}{Ns(s+1)} \sum_{i=1}^s R_i^2 - 3N(s+1), \quad \text{onde } R_i = \sum_{j=1}^N R_{ij}$$

A determinação da constante crítica c é feita a partir da distribuição de Q , sob a hipótese H_0 . Pode-se mostrar que sob a hipótese H_0 , e sendo os s indivíduos, dentro de cada bloco, selecionados aleatoriamente para a aplicação dos s -tratamentos - todos os $s!$ permutações de $(1, 2, \dots, s)$ são valores equiprováveis de (R_{1j}, \dots, R_{sj}) . Como as atribuições dos tratamentos em diferentes blocos

são independentes, temos

(4.11)

$$P_H(R_{11} = r_{11}, \dots, R_{s1} = r_{s1}; \dots; R_{1N} = r_{1N}, \dots, R_{sN} = r_{sN}) = \left(\frac{1}{s!}\right)^N$$

onde (r_{1j}, \dots, r_{sj}) é uma permutação de $(1, 2, \dots, s)$, para $j = 1, \dots, N$.

A distribuição de Q sob a hipótese H_0 é obtida a partir de (4.11), calculando-se para cada alocação possível de postos, o valor de Q correspondente e sua probabilidade.

Exemplo 3: Considere o caso $N=2$ (2 blocos) e $s=4$ (4 tratamentos). Abaixo, enumeramos algumas configurações de postos e correspondentes valores de Q .

(a)	Blocos	
	trat.	1 2
I	1	1
II	2	2
III	3	3
IV	4	4

$$Q = 6$$

(b)	Blocos	
	trat.	1 2
I	1	1
II	2	2
III	3	4
IV	4	3

$$Q = 5.4$$

(c) trat.	Blocos	
	1	2
I	1	1
II	2	3
III	3	4
IV	4	2

$$Q = 4.2$$

(d) trat.	Blocos	
	1	2
I	1	1
II	2	3
III	3	2
IV	4	4

$$Q = 5.4$$

O número de possíveis configurações de postos é, neste caso, $(4!)^2 = 576$. É fácil ver que várias destas configurações fornecem o mesmo valor de Q . Desta forma,

$$P_H(Q=q) = \frac{\text{número de configurações com } Q=q}{576}$$

Esta distribuição de Q é tabelada para diferentes valores de s e N , quando estes são pequenos. Para valores grandes de N , aqui também, dispõe-se de uma aproximação. Temos o seguinte resultado: quando $N \rightarrow \infty$,

$$P_H(Q \leq c) \rightarrow \chi_{s-1}^2(c),$$

onde, χ_{s-1}^2 é a função cumulativa de uma qui-quadrado com $s-1$ graus de liberdade.

No caso de ocorrência de empates entre respostas de um mesmo bloco, a estatística do teste é convenientemente modificada, introduzindo-se uma correção por empates. Seja R_{ij}^* o posto médio do indivíduo recebendo tratamento é no

bloco j . Denotemos por R_i^* a soma dos postos médios dos indivíduos recebendo o i -ésimo tratamento, isto é,

$$R_i^* = \sum_{j=1}^N R_{ij}^*.$$

A estatística apropriada no caso de empates é

$$(4.12) \quad Q^* = \frac{[12/Ns(s+1)] \sum R_i^{*2} - 3N(s+1)}{1 - \sum_{j=1}^N \sum_{i=1}^{e_j} (d_{ij}^3 - d_{ij}) / Ns(s^2 - 1)},$$

onde e_j denota o número de valores distintos tomados pelas observações do j -ésimo bloco; d_{1j} destas observações são iguais ao menor valor, d_{2j} ao segundo menor, etc. O teste rejeita H_0 quando: $K^* \geq c$. A distribuição de K^* é, novamente, aproximada por uma qui-quadrado com $s-1$ graus de liberdade.

No Capítulo 3 estudamos o problema da comparação emparelhada de dois tratamentos. Foram, então, apresentados os testes do sinal e dos postos sinalizados de Wilcoxon. A comparação de s tratamentos no caso $s = 2$ se reduz ao da comparação por pares. É interessante verificar o que ocorre com o teste de Friedman neste caso. Designando por A o número de blocos onde o tratamento 1 obtém o posto 1 e o tratamento 2 o posto 2, temos

$$R_{1.} = \frac{A + 2(N-A)}{N} = 2 - \frac{A}{N}; \quad R_{2.} = \frac{N-A + 2A}{N} = 1 + \frac{A}{N}.$$

Substituindo estes valores na expressão de Q , vem

$$Q = \frac{12N}{2 \times 3} \left[\left(2 - \frac{A}{N} - \frac{3}{2} \right)^2 + \left(1 + \frac{A}{N} - \frac{3}{2} \right)^2 \right] = 4N \left(\frac{A}{N} - \frac{1}{2} \right)^2.$$

De acordo com o teste de Friedman, rejeitamos a hipótese de nenhuma diferença entre tratamentos quando Q é suficientemente grande ou, no caso, quando $\left| \frac{A}{N} - \frac{1}{2} \right|$ é suficientemente grande. Por outro lado, A representa o número de pares onde a resposta do segundo tratamento menos a resposta do primeiro é positiva. A condição de rejeição se reduz, portanto, no caso $s=2$ à do teste bilateral do sinal. O desempenho do teste do sinal é, geralmente, mais fraco que o do teste de postos sinalizados de Wilcoxon. Na verdade, aquele deve ser utilizado tão somente, nos casos onde se conhece, apenas, o tratamento de maior efeito no par. No teste dos postos sinalizados de Wilcoxon é levado em consideração o valor da diferença entre as respostas dos dois tratamentos no par. Isto, em geral, aumenta bastante a capacidade de detectar diferenças entre os dois tratamentos. É natural, no caso de mais de dois tratamentos, do mesmo modo, procurar melhorar o desempenho do teste de Friedman. A seguir, daremos uma breve descrição de como fazê-lo.

§6. Postos alinhados.

No teste de Friedman, a atribuição de postos aos tratamentos é feita dentro de cada bloco. Tal procedimento torna o teste, principalmente quando s é pequeno, muito pouco sensível. Por outro lado, não tem sentido a comparação entre blocos, devido ao efeito de bloco. A idéia é eliminar, de certa forma, o efeito do bloco, subtraindo de cada observação deste, uma estimativa de sua locação. Para este fim, pode-se utilizar, por exemplo, a média ou mediana das observações do bloco. O método acima é chamado de alinhamento das observações.

Designemos por \hat{R}_{ij} o posto alinhado, ou seja, o posto de observação alinhada do i -ésimo tratamento no j -ésimo bloco no conjunto de todas as sN observações alinhadas. Uma motivação análoga à descrita para o teste de Friedman, fornece a seguinte estatística de teste:

$$(4.13) \quad \hat{Q} = \frac{s-1}{\sum \sum (\hat{R}_{ij} - \hat{R}_{.j})^2} \sum_{i=1}^s [N\hat{R}_{i.} - 1/2 N(sN+1)]^2,$$

onde, $\hat{R}_{i.} = \sum_{j=1}^N R_{ij}/N$ é a média dos postos do i -ésimo tratamento nos N blocos e $\hat{R}_{.j} = \sum_{i=1}^s R_{ij}/s$ é a média dos postos dos s tratamentos no j -ésimo bloco.

Para valores grandes de sN , a estatística \hat{Q} é aproximadamente distribuída como uma qui-quadrado com $s-1$ graus

de liberdade.

Exemplo 4: A tabela abaixo (Snedecor, pag. 297) fornece os valores das observações para um experimento simulado de acordo com o modelo da página 140. Na geração, os erros ϵ_{ij} são $N(0;25)$.

$\mu=30$

Tratamentos	Blocos			Soma dos Postos (R_i)
	$\beta_1=1$	$\beta_2=-4$	$\beta_3=3$	
$\alpha_1=10$	30 10 1 -11	30 10 -4 -7	30 10 3 3	8
	$x_{11}=30$ (2)	$x_{12}=29$ (2)	$x_{13}=46$ (4)	
$\alpha_2=3$	30 3 1 1	30 3 -4 5	30 3 3 -3	11
	$x_{21}=35$ (4)	$x_{22}=34$ (4)	$x_{23}=33$ (3)	
$\alpha_3=0$	30 0 1 0	30 0 -4 4	30 0 3 -1	8
	$x_{31}=31$ (3)	$x_{32}=30$ (3)	$x_{33}=32$ (2)	
$\alpha_4=-13$	30 -13 1 -2	30 -13 -4 -2	30 -13 3 1	3
	$x_{41}=16$ (1)	$x_{42}=11$ (1)	$x_{43}=21$ (1)	

Temos então $s=4$ tratamentos e $N=3$ blocos.

Utilizando o teste de Friedman vamos testar a hipótese

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$. Os dados acima fornecem

$$R_1 = R_{11} + R_{12} + R_{13} = 2 + 2 + 4 = 8$$

$$R_2 = R_{21} + R_{22} + R_{23} = 4 + 4 + 3 = 11$$

$$R_3 = R_{31} + R_{32} + R_{33} = 3 + 3 + 2 = 8$$

$$R_4 = R_{41} + R_{42} + R_{43} = 1 + 1 + 1 = 3$$

Os postos se encontram, na tabela acima, entre parêntesis.

$$Q = \frac{12}{3 \times 4 \times 5} \times 258 - 3 \times 3 \times 5 = 6,60$$

A probabilidade de significância do teste é

$P(\chi_3^2 \geq 6,60) = 0.0858$; a hipótese não seria rejeitada para $\alpha = 5\%$.

Para este mesmo conjunto de dados vamos considerar os postos alinhados. Utilizaremos a média em cada bloco para alinhar as observações.

Observações alinhadas

Tratamentos	Blocos			$\hat{R}_{i.}$
	1	2	3	
1	2 (6)	3 (7.5)	13 (12)	8.50
2	7 (10)	8 (11)	0 (5)	8.67
3	3 (7.5)	4 (9)	-1 (4)	6.83
4	-12 (2.5)	-15 (1)	-12 (2.5)	2.00
$\hat{R}_{.j}$	6.50	7.13	5.58	

Os pontos, no conjunto de todas as observações, encontram-se na tabela acima entre parêntesis. Temos $\hat{Q} = 9.11$ e a probabilidade de significância do teste é $P(\chi_3^2 \geq 9.11) \approx 0.0237$. A hipótese seria rejeitada pelo teste \hat{Q} ao nível de 5%.

Exercícios - Capítulo IV

1. Utilizando a fórmula (4.1), determine, para o caso em que $n_1=2, n_2=3, n_3=3$, as probabilidades

(i) $P_H(R_{11}=1, R_{12}=2; R_{21}=3, R_{22}=4, R_{23}=5)$

(ii) $P_H(R_{11}=1, R_{12}=3; R_{21}=5, R_{22}=6, R_{23}=7)$

(iii) $P_H(R_{11}=1, R_{21}=2)$

(iv) $P_H(R_{11}=1, R_{12}=2)$

(v) $P_H(R_{11}=1, R_{21}=4)$

(vi) $P_H(R_{11}=1, R_{21}=2, R_{31}=3)$.

2. Ache a distribuição nula da estatística K para os casos (i) $s=3, n_1=n_2=1, n_3=6$; (ii) $s=3, n_1=n_2=n_3=2$.

3. Considere a tabela de contingência $2 \times t$ abaixo

Tratamentos	Respostas				Total
	A_1	A_2	\dots	A_t	
A	A_1	A_2	\dots	A_t	m
B	B_1	B_2	\dots	B_t	n
	d_1	d_2	\dots	d_t	N

Considere, agora, as t respostas como t -tratamentos e os 2 tratamentos A e B como as duas respostas possíveis de cada um dos tratamentos. Atribua escore 1 ao tra

tamento A e escore 2 ao tratamento B. Mostre que a soma dos postos médios correspondentes à i -ésima resposta é

$$i) \quad R_i^* = A_i \left(\frac{m+1}{2} \right) + B_i \left(m + \frac{n+1}{2} \right)$$

ii) Mostre que a estatística de "Kruskal -Wallis" para os t "tratamentos" com duas respostas é dada por

$$\bar{K}^* = \frac{N(N-1)}{mn} \left(\sum \frac{A_i^2}{d_i} - \frac{m^2}{N} \right).$$

4. Amostras de tamanhos m e n ($m+n = N$) são retiradas de uma população e designadas para os tratamentos A e B, respectivamente. Suponhamos que as respostas possíveis são $1, 2, \dots, t$. Sejam F_A e F_B as distribuições das respostas, respectivamente, sob tratamento A e B. Considerando a tabela de contingência do Exercício 3, mostre que

$$(i) \quad P(A_1=a_1, \dots, A_t=a_t) = \binom{m}{a_1, \dots, a_t} \pi_1^{a_1} \dots \pi_t^{a_t},$$

onde π_1, \dots, π_t são, respectivamente, as probabilidades das respostas $1, \dots, t$ sob a distribuição F_A .

(ii) Sendo π_1, \dots, π_t como definidas em (i) e

π'_1, \dots, π'_t , respectivamente, as probabilidades das respostas $1, \dots, t$ sob a distribuição F_B , então, sob a hipótese $H: \pi_1 = \pi'_1, \dots, \pi_t = \pi'_t$, temos

$$P_H(A_1=a_1, \dots, A_t=a_t; B_1=b_1, \dots, B_t=b_t) = \\ = \binom{m}{a_1, \dots, a_t} \binom{n}{b_1, \dots, b_t} \pi_1^{a_1+b_1} \dots \pi_t^{a_t+b_t}.$$

(ii) A distribuição condicional de $A_1, \dots, A_t; B_1, \dots, B_t$, dada a configuração de empates (d_1, \dots, d_t) é

$$P_H(A_1=a_1, \dots, A_t=a_t; B_1=b_1, \dots, B_t=b_t | d_1, \dots, d_t) = \\ = \frac{\binom{m}{a_1, \dots, a_t} \binom{n}{b_1, \dots, b_t}}{N} \\ (d_1, \dots, d_t)$$

com $a_i + b_i = d_i$, $i=1, \dots, t$.

5. Amostras de tamanhos d_1, \dots, d_t ($d_1 + \dots + d_t = N$) são retiradas de uma população e designadas para os tratamentos $1, \dots, t$. Suponhamos que as respostas possíveis são A e B. Sejam F_i , $i=1, \dots, t$ as distribuições das respostas dos tratamentos, onde F_i atribui probabilidade p_i a A e q_i a B. Considerando a tabela de contingência do Exercício 3, mostre que

(i) $P_H(A_1=a_1, \dots, A_t=a_t) = \binom{d_1}{a_1} \dots \binom{d_t}{a_t} p^m \cdot q^n$, sob a hipótese $H: F_1 = \dots = F_t$, sendo p o valor comum dos p_i 's e q o valor comum dos q_i 's.

(ii) A distribuição condicional de $A_1, \dots, A_t; B_1, \dots, B_t$

dada a configuração de empates (m, n) é

$$P_H(A_1=a_1, \dots, A_t=a_t; B_1=b_1, \dots, B_t=b_t | (m, n)) = \\ = \binom{d_1}{a_1} \dots \binom{d_t}{a_t} / \binom{N}{n}$$

onde $a_1 + \dots + a_t = m$ e $b_1 + \dots + b_t = N-m$.

(iii) A probabilidade obtida no ítem (ii) deste exercício é igual à obtida no ítem (ii) do Exercício 4.

6. Observe, no texto, o esquema experimental, na definição da estatística K^* de Kruskal -Wallis. Compare este com o utilizado na definição da estatística \bar{K}^* , no ítem (ii) do Exercício 3. A partir do resultado obtido no ítem (iii) do Exercício 5 que conclusão podemos tirar sobre as distribuições de K^* e \bar{K}^* ?

7. Na derivação da expressão de \bar{K}^* , no Exercício 3, mostre que esta não se altera quando atribuímos, no lugar dos escores 1 e 2, escores a e b para os tratamentos A e B.

8. Mostre que o teste de Kruskal -Wallis se reduz ao teste bilateral de Wilcoxon, no caso $s=2$.

9. Mostre que a distribuição da estatística J , definida no texto, é simétrica em torno de zero.

10. Considere três tratamentos A, B e C e suponhamos que na comparação dos mesmos foram utilizados 4 blocos. Ache, sob a hipótese de nenhum efeito de tratamento, a probabilidade do tratamento A obter:

- (i) posto 3 nos 4 blocos
- (ii) posto 3 nos três primeiros blocos e posto 2 no último
- (iii) posto 3 em três blocos e posto 2 no outro restante.

11. Com s tratamentos e N blocos qual a probabilidade, sob a hipótese de nenhum efeito de tratamento, do primeiro tratamento ter posto 1 em todos os N blocos? Qual a probabilidade do primeiro tratamento ter posto 1 e o segundo posto 2 em todos os N blocos?

APÊNDICE

Incluimos neste apêndice, de modo bem sucinto, algumas definições e conceitos necessários à compreensão do texto. Na verdade, o material é apresentado a título de recordação, sendo considerado como pré-requisito para o curso de Não-Paramétrica.

Dizemos que (X_1, \dots, X_n) é uma amostra aleatória de uma distribuição F , se X_1, \dots, X_n são variáveis aleatórias independentes com a mesma função de distribuição F , isto é,

$$P(X_i \leq x) = F(x) = F(x) \quad \forall x \in R, \quad i=1, \dots, n \quad e$$

$$(A.1) \quad F(x_1, \dots, x_n) = F(x_1) \dots F(x_n),$$

onde F_{X_1, \dots, X_n} é função de distribuição conjunta de (X_1, \dots, X_n) .

Na prática, dois tipos de funções de distribuições são de interesse, em Estatística:

Função escada correspondente à variável aleatória tomando valores discretos $a_1 < a_2 < \dots$ em número finito ou enumerável com probabilidades $p_1, p_2, \dots, \sum p_i = 1$, respectivamente. A função de distribuição apresenta saltos nos pontos

tos a_1, a_2, \dots e permanece constante em cada um dos intervalos $[a_i, a_{i+1})$. O salto em a_i é igual a p_i . A função de distribuição fica então definida a partir da função $p(a_i) = P(X=a_i) = p_i$, chamada função de probabilidade de X; $F(x) = \sum_{a_i \leq x} p(a_i)$.

O segundo tipo de função de distribuição é chamado absolutamente contínua. A função F , neste caso, admite uma representação da forma

$$(A.2) \quad F(x) = \int_{-\infty}^x f(x) dx,$$

onde f é chamada de densidade de probabilidade. Se f é contínua no ponto x , temos $F'(x) = f(x)$. Alguns exemplos de distribuições discretas e de distribuições tendo densidade são apresentados no Capítulo I. A condição de independência (A.1), das v.a's X_1, \dots, X_n , no caso discreto, fica

$$(A.3) \quad p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p(x_1) \cdot p(x_2) \cdot \dots \cdot p(x_n), \\ \forall (x_1, \dots, x_n) \in \mathbb{R}^n,$$

onde $p_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1=x_1, \dots, X_n=x_n)$ é a função de probabilidade conjunta dos v.a's X_1, \dots, X_n . No caso de F ter densidade, a densidade conjunta de X_1, \dots, X_n é

$$(A.4) \quad f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f(x_1) \cdot f(x_2) \dots f(x_n) \\ \forall (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Dizemos que a distribuição de uma variável aleatória X é simétrica em torno de zero se X e $-X$ têm a mesma distribuição: $P(X \leq x) = P(-X \leq x)$, $\forall x \in \mathbb{R}$. Quando F , função de distribuição de X , é contínua

$$(A.5) \quad F(x) = 1 - F(-x), \quad \forall x \in \mathbb{R}.$$

No caso discreto, X tem distribuição simétrica em torno de zero se

$$(A.6) \quad P(X=x) = P(X=-x), \quad \forall x \in \mathbb{R}.$$

No caso de densidade, a condição (A.5) é equivalente a

$$f(x) = f(-x) \quad \forall x \in \mathbb{R}.$$

Exemplos de distribuições simétricas em torno de zero:

$N(0,1)$, Exponencial dupla padrão, Cauchy padrão. As densidades destas distribuições são dadas no Capítulo I.

Observe que estas distribuições são, ainda, simétricas em torno de zero quando o parâmetro de locação $a=0$ e o parâmetro de escala b tem um valor qualquer (V. pag.10 do Capítulo I).

Uma variável aleatória X tem distribuição simétrica em

torno de a se $X-a$ tem distribuição simétrica em torno de zero.

Exemplos: Normal, Cauchy, exponencial dupla com parâmetro de locação a .

Os elementos distintos de uma família de distribuição são caracterizados por terem valores distintos de alguma característica numérica, denominada parâmetro da distribuição. Como exemplo, vimos, no Capítulo I, os parâmetros de locação e escalas a e b . Abaixo discutimos a geração de famílias de distribuições a partir destes pa râmetros.

Consideremos uma variável aleatória X com distribuição F_0 . Seja agora a v.a. $Y = X+a$, obtida por uma mudança de locação. Denotemos a distribuição de Y por F_a , e consideremos a família de todas estas distribuições com $-\infty < a < \infty$. Esta família é chamada família de locação gerada pela distribuição F_0 , e a é um parâmetro de locação. Assim, a distribuição F_a é dada por $F_a(x) = P(Y \leq x) = P(X+a \leq x) = P(X \leq x-a) = F_0(x-a)$, e nos referimos à $\{F_0(x-a)\}_{a \in \mathbb{R}}$ como família de locação com parâmetro de locação a . Se $F_a(x)$ tem densidade $f_a(x)$, então

$$f_a(x) = \frac{d}{d} F_0(x-a) = f_0(x-a)$$

e a família é dada por $\{f_0(x-a)\}_{a \in \mathbb{R}}$. Os elementos desta

família têm densidades com a mesma forma e são obtidas umas das outras apenas por deslocamento da posição. Observe, ainda, que se μ é a média de F_0 , então $\mu+a$ é a média de F_a . A variância fica, contudo, inalterada,

$$\text{Var}(Y) = \text{Var}(X-a) = \text{Var } X \quad \forall a \in \mathbb{R},$$

para todas as distribuições da família.

No caso de famílias com parâmetro multidimensional (a,b) , chamamos a parâmetro de locação da família se esta é da forma $F(x-a,b)$, ou, no caso de densidades, da forma $f(x-a,b)$. Isto equivale a dizer que se a distribuição de X tem parâmetros (a,b) , então a distribuição de $X-a$ tem a distribuição com parâmetros $(0,b)$, pois

$$P(X-a \leq x) = P(X \leq x+a) = F(x+a-a, b) = F(0, b).$$

Exemplos: 1. Normal: $f_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, μ é pa-

râmetro de locação, pois podemos escrever

$f_{\mu, \sigma^2}(x)$ na forma $f(x-\mu, \sigma^2)$ com f dada por

$$f(t, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-t^2/2\sigma^2}.$$

2. Exponencial dupla: $f_{\alpha, \beta}(x) = \frac{1}{2\beta} e^{-|x-\alpha|/\beta}$, α é pa-

râmetro de locação, pois podemos escrever $f_{\alpha, \beta}(x) = f(x-\alpha, \beta)$ com f dada por $f(t, \beta) = \frac{1}{2\beta} e^{-|t|/\beta}$.

De modo análogo ao que foi feito para família de locação, podemos gerar uma família de escala a partir de uma distribuição dada. Seja X uma v.a. com função de distribuição F_1 e seja $b > 0$. Denotemos por F_b a função de distribuição de $Y = bX$. O conjunto $\{F_b\}_{b>0}$ é chamado família de escala gerada por F_1 , e a é o parâmetro de escala. A relação entre as funções de distribuições de X e Y é $F_b(x) = P(Y \leq x) = P(bX \leq x) = P(X \leq x/b) = F_1(x/b)$. No caso de F ter densidade, temos

$$f_b(x) = \frac{dF_b(x)}{dx} = \frac{1}{b} f_1(x/b).$$

Deste modo, as funções de distribuições podem ser escritas na forma $F_1(x/b)$, e no caso de densidade, na forma $1/b f_1(x/b)$.

Exemplo: Normal $N(0, \sigma^2)$, neste caso, σ é parâmetro de escala, pois

$$f_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}},$$

basta tomar $f_1(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ e obtemos $f_\sigma(x) = \frac{1}{\sigma} f_1(x/\sigma)$.

Podemos definir parâmetro de escala também no caso de famílias com dois parâmetros. Neste caso, dizemos que b é um parâmetro de escala se as funções de distribuição de família são da forma $F_1(x/b, a)$ ou as densida

des da forma $\frac{1}{b} f_1(x/b, a)$. Isto equivale a dizer que se X tem distribuição com parâmetros (a, b) , X/b tem distribuição com parâmetros $(a, 1)$.

Exemplo: $N(\mu, \sigma^2)$, neste caso σ é um parâmetro de escala. Tome $b = \sigma$, $a = \mu/\sigma$ e $f_1(u, v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(u-v)^2}{2}}$. É fácil verificar, de outra maneira, que se X tem distribuição com parâmetro $(\mu/\sigma, 1)$. Os dois processos de geração de famílias de distribuições podem ser combinados. Para isto, consideremos uma v.a. X com distribuição $F_{0,1}$ e façamos a transformação linear: $Y = a+bX$, $b > 0$. A família $(F_{a,b})_{\substack{a \in \mathbb{R} \\ b > 0}}$ das distribuições de Y com a e b variando, forma o que chamamos de família de locação-escala, gerada da distribuição padrão $F_{0,1}$. A relação entre $F_{a,b}$ e F é dada por

$$F_{a,b}(x) = P(a+bX \leq x) = P(X \leq \frac{x-a}{b}) = F_{0,1}(\frac{x-a}{b}),$$

Portanto, outra caracterização desta família é de ter as distribuições da forma

$$(A.7) \quad F_{(0,1)}(\frac{x-a}{b})$$

ou, no caso de densidades, da forma

$$(A.8) \quad \frac{1}{b} f_{(0,1)}(\frac{x-a}{b}).$$

De maneira equivalente, se a distribuição de Y é um elemento da família com parâmetros a e b , então

$$X = \frac{Y-a}{b}$$

tem a distribuição padrão $F_{0,1}$.

As famílias de distribuições apresentadas no Capítulo I, foram obtidas pelo método de geração acima.

Outros parâmetros importantes da distribuição F são a média ou variância da população:

Média

$$\mu_X = EX = \int_{-\infty}^{+\infty} xf(x)dx \quad \text{onde } f \text{ é a densidade de } F.$$

Variância

$$\sigma_X^2 = \text{Var } X = E(X - \mu_X)^2 = EX^2 - \mu_X^2$$

Observações: 1. O parâmetro de locação é igual a média quando a distribuição padrão da família de locação tem média zero: $\mu_Y = a + \mu_X$ e $\mu_Y = a$ se $\mu_X = 0$.

2. O parâmetro de escala é igual a variância quando a distribuição padrão da família de escala tem variância 1: $\sigma_Y^2 = b^2 \sigma_X^2$ e $\sigma_Y^2 = b^2$ se $\sigma_X^2 = 1$.

A média e a variância de uma combinação linear de V. a.'s, satisfazem

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i)$$

$$\text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + \sum_{i \neq j} a_i a_j \text{cov}(X_i, X_j)$$

onde a_1, \dots, a_n são constantes, e a covariância entre X_i e X_j é dada por

$$\text{Cov}(X_i, X_j) = E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})] = E(X_i X_j) - \mu_{X_i} \mu_{X_j}$$

Se as v.a.'s aleatórias X_1, \dots, X_n são independentes, então $E(X_i X_j) = E(X_i)E(X_j) = \mu_{X_i} \mu_{X_j}$ e

$\text{cov}(X_i, X_j) = 0$, logo

$$\text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i)$$

O coeficiente de correlação entre as variáveis X e Y , é

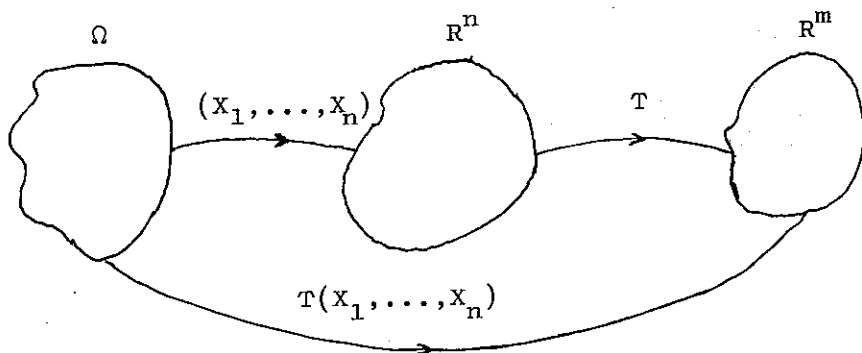
$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Outro paâmetro importante, relativo à locação da distribuição, é a mediana:

$m =$ mediana de X , é o valor que satisfaz

$$P(X < m) \leq 1/2 \quad \text{e} \quad P(X > m) \leq 1/2$$

Uma estatística é uma função das variáveis aleatórias amostrais. O esquema a seguir, esclarece a definição:



$T(X_1, \dots, X_n)$ é uma estatística.

Estatísticas utilizadas frequentemente são:

1. Média Amostral: $\bar{X} = 1/n \sum_{i=1}^n X_i$.

2. Desvio padrão Amostral: $S^2 = 1/n \sum (X_i - \bar{X})^2$.

3. i-ésima estatística de ordem: Considere $T(x_1, \dots, x_n) = (x_{(1)}, \dots, x_{(n)})$ onde $(x_{(1)}, \dots, x_{(n)})$ é a permutação de (x_1, \dots, x_n) tal que $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ (supondo os x_i 's distintos). Tome, agora $T(X_1, \dots, X_n) = (X_{(1)}, \dots, X_{(n)})$. A i-ésima estatística de ordem é $X_{(i)}$.

Observação: Não é verdade que $X_{(i)} = X_j$ para algum

$$1 \leq j \leq n; \text{ o que temos é } X_{(i)}(w) = X_j(w)$$

para um dado w e algum j . Quando variamos w , j pode variar.

4. Mediana das observações: Se $n = 2k+1$, $\text{med}(X_1, \dots, X_n) = X_{(k+1)}$; se $n=2k$, $\text{med}(X_1, \dots, X_n) = \frac{1}{2} (X_{(k)} + X_{(k+1)})$.

5. Combinações lineares das observações: $a_1 X_1 + \dots + a_n X_n$ com a_1, \dots, a_n constantes.

6. Combinações lineares das estatísticas de ordem:

$$a_1 X_{(1)} + \dots + a_n X_{(n)}.$$

Observações: a) Não é verdade que $a_1 X_1 + \dots + a_n X_n = a_1 X_{(1)} + \dots + a_n X_{(n)}$ (Por que?).

b) A média é uma estatística do tipo 5, com $a_1 = \dots = a_n = 1/n$ e é também do tipo 6.

c) A mediana é uma estatística do tipo 6 com $a_{k+1} = 1$ e $a_i = 0$ se $i \neq k+1$, no caso ímpar; $a_k = 1/2$, $a_{k+1} = 1/2$ e $a_i = 0$ se $i \neq k$ e $i \neq k+1$, no caso par.

7. Postos das observações: O posto de X_i é o número R_i de observações menores ou iguais a X_i . Considere

$$u(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases}$$

Podemos escrever $R_i = \sum_{j=1}^n u(X_i - X_j)$, $1 \leq i \leq N$. Se a distribuição dos X_i 's é contínua a probabilidade de observações com o mesmo valor é zero. Os valores dos Postos R_1, \dots, R_n , neste caso, formam uma permutação de $(1, \dots, N)$.

8. Estatísticas do tipo: $a_1 c_{R_1} + a_2 c_{R_2} + \dots + a_n c_{R_n}$ onde a_1, \dots, a_n e c_1, \dots, c_n são constantes. Os a_i 's são chamados escores e os c_i 's constantes de regressão.

Exemplo: Considere duas amostras (X_1, \dots, X_m) e (Y_1, \dots, Y_n) . Seja $(Z_1, \dots, Z_m; Z_{m+1}, \dots, Z_{m+n})$ a amostra conjunta, onde $Z_i = X_i$ $i=1, \dots, m$ e $Z_i = Y_i$ para $i = m+1, \dots, m+n$ e R_1, \dots, R_{m+n} os postos dos Z_i 's. Tomando $a_i = i$, $c_i = 0$ se $1 \leq i \leq m$ e $c_i = 1$ se $m+1 \leq i \leq m+n$, obtemos $W_S = \sum_{i=m+1}^{m+n} R_i$ que é a estatística de Wilcoxon, definida no Capítulo II.

Estimadores são funções das variáveis aleatórias $\hat{\theta}(X_1, \dots, X_n)$, ou seja estatísticas, utilizados para estimar, θ , parâmetro desconhecido da distribuição F_θ de (X_1, \dots, X_n) . Em princípio, qualquer função $\hat{\theta}(X_1, \dots, X_n)$ de X_1, \dots, X_n poderia ser tomada como estimador de θ . O que desejamos, porém, é determinar estimadores satisfazendo certas propriedades.

Estimadores não viciados:

$\hat{\theta}(X_1, \dots, X_n)$ é um estimador não viciado de $g(\theta)$ se

$$E_\theta \hat{\theta}(X_1, \dots, X_n) = g(\theta), \quad \forall \theta \in \Theta,$$

onde E_θ é a esperança calculada a partir da distribuição F_θ e Θ é o conjunto de valores do parâmetro.

Exemplo: Seja X_1, \dots, X_n uma amostra de uma distribuição tendo densidade $f(x; \theta)$, então $\bar{X} = 1/n \sum_{i=1}^n X_i$ é um estimador não viciado de $g(\theta) = \int_{-\infty}^{+\infty} xf(x; \theta) dx$.

Estimador Consistente

$\hat{\theta}_n(X_1, \dots, X_n)$ é um estimador consistente de θ se $\lim_{n \rightarrow \infty} P_{\theta}(|\hat{\theta}_n - \theta| > \epsilon) = 0$, para qualquer $\epsilon > 0$.

Estimador com erro médio quadrático mínimo

$\hat{\theta}^*$ é um estimador de θ com erro médio quadrático mínimo se $E_{\theta}(\hat{\theta}^* - \theta)^2 \leq E_{\theta}(\hat{\theta} - \theta)^2$, $\forall \theta \in \Theta$ onde $\hat{\theta}$ é qualquer outro estimador.

Se $\hat{\theta}$ é um estimador não viciado de θ , o erro médio quadrático se reduz à variância

$$E_{\theta}(\hat{\theta} - \theta)^2 = \text{Var}_{\theta} \hat{\theta}.$$

Exemplo: Se (X_1, \dots, X_n) é uma amostra de uma $N(\mu, 1)$ então $\bar{X} = 1/n \sum_{i=1}^n X_i$ é um estimador não viciado de mínima variância de μ .

Seja T_n uma sequência de variáveis aleatórias e F uma distribuição contínua. Então T_n tem distribuição limite F se $P(T_n \leq t) \rightarrow F(t)$ quando $n \rightarrow \infty$.

Quando a distribuição de T_n tende para a distribuição de T , denotamos por $T_n \xrightarrow{D} T$.

Estimador Assintoticamente Normal

$\hat{\theta}_n$ é um estimador de θ , assintoticamente normal se $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N(0, \sigma^2(\theta))$.

Exemplo: Sob certas condições de regularidade, estimadores de Máxima Verossimilhança são consistentes e assintoticamente normais.

Uma comparação assintótica entre estimadores pode ser feita utilizando a variância assintótica $\sigma^2(\theta)$. Sob certas condições de regularidades os estimadores de Máxima Verossimilhança são eficientes, isto é, têm variância assintótica mínima.

Os estimadores comentados acima são chamados estimadores pontuais. Outro processo de estimação, que ao mesmo tempo fornece o estimador e a sua precisão, é por intervalos de confiança. Neste caso, o estimador é dado por intervalo de extremos (U, V) , onde U e V são funções das v.a.'s amostrais.

Um intervalo de confiança para θ com coeficiente de confiança $(1-\alpha)$ é um intervalo aleatório (U, V) , tal que

$$P_{\theta}(U < \theta < V) = 1-\alpha \quad \forall \theta \in \Theta.$$

Teste de Hipótese

Seja X_1, \dots, X_n uma amostra de uma distribuição

F. Suponhamos que o modelo estatístico especifica que $F \in \mathcal{F}$, onde \mathcal{F} é uma dada família de distribuições. Uma hipótese estatística é uma afirmação do tipo $F \in \mathcal{F}_0$, onde $\mathcal{F}_0 \subset \mathcal{F}$. Suponhamos, para fixar idéias, que \mathcal{F} seja parametrizada por um conjunto Θ , isto é, cada elemento $\theta \in \Theta$ especifica um só elemento de \mathcal{F} . Neste contexto, uma hipótese estatística é uma afirmação do tipo $\theta \in \Theta_0$ onde $\Theta_0 \subset \Theta$. A hipótese estatística $\theta \in \Theta_0$ é usualmente denominada hipótese nula e representada por $H_0: \theta \in \Theta_0$. A hipótese $\theta \in \Theta - \Theta_0$ é denominada hipótese alternativa e representada por $H_a: \theta \in \Theta - \Theta_0$. Muitas vezes, no lugar de H_a utiliza-se, H_1 , K , etc. Quando o conjunto Θ_0 é unitário, $\Theta_0 = \{\theta_0\}$, a hipótese $H_0: \theta = \theta_0$ é dita simples, caso contrário é composta.

Exemplo: X_1, \dots, X_n i.i.d. $N(\theta, 1)$

$$H_0: \theta \leq \theta_0 \text{ contra } H_a: \theta > \theta_0.$$

Neste exemplo $\Theta = \mathbb{R}$, $\Theta_0 = (-\infty, \theta_0]$.

Um teste de uma hipótese estatística é uma regra que possibilita tomar a decisão de aceitar ou rejeitar H_0 , com base no valor de uma estatística. Esta é denominada estatística do teste.

A regra de decisão é dada pela fixação de um subconjunto R no contradomínio da estatística de teste T , denomina

do região crítica ou região de rejeição. Se $T(x_1, \dots, x_n) \in R$, então rejeitamos a hipótese.

Exemplo: X_1, \dots, X_n i.i.d $N(\theta, 1)$.

$$H_0: \theta \leq 0 \text{ contra } H_a: \theta > 0.$$

Considere o teste: rejeite se $\bar{X} \geq 2$.

Neste caso, $T(X_1, \dots, X_n) = \bar{X} = 1/n \sum_{i=1}^n X_i$ é a estatística do teste e $(2, \infty)$ é a região crítica. O valor 2, neste exemplo, é chamado valor crítico.

A escolha da região crítica é feita levando em consideração as probabilidades dos erros:

erro do tipo I - é o erro de rejeitar a hipótese H_0 quando esta é verdadeira

erro do tipo II - é o erro de deixar de rejeitar H_0 quando esta é falsa.

As probabilidades destes erros são, respectivamente,

$$\alpha(\theta) = P_\theta(T \in R) \text{ para } \theta \in \Theta_0$$

$$\beta(\theta) = P_\theta(T \notin R) \text{ para } \theta \in \Theta - \Theta_0.$$

O tamanho do teste é dado por $\sup_{\theta \in \Theta_0} \alpha(\theta)$.

Exemplo: No exemplo anterior, $\alpha(\theta) = P_\theta(\bar{X} \geq 2) = 1 - \Phi(\sqrt{n}(2-\theta))$, $\theta \leq 0$, onde Φ é a função de distribuição de uma normal padrão. Para o erro do tipo II,

$$\beta(\theta) = P_{\theta}(\bar{X} < 2) = \Phi(\sqrt{n}(2-\theta)), \quad \theta > 0.$$

O tamanho do teste, neste caso, é

$$\sup_{\theta \leq 0} \alpha(\theta) = 1 - \Phi(2\sqrt{n}).$$

O nível de significância do teste é um limitante nominal selecionado para $\alpha(\theta)$, $\theta \in \Theta_0$.

No caso discreto este nível não é usualmente atingido, fato que ocorre nos testes não paramétricos apresentados.

Dizemos que o teste tem nível de significância exato α , quando $\max_{\theta \in \Theta_0} \alpha(\theta) = \alpha$.

Consideremos agora o problema de testar

$$H_0: \theta = \theta_0 \quad \text{contra} \quad H_A: \theta > \theta_0.$$

Suponhamos que o teste estatístico utilizado para testar a hipótese acima tenha a forma

$$\text{Rejeite } H_0 \text{ se } T(X_1, \dots, X_n) \geq t_0.$$

Seja $t = T(x_1, \dots, x_n)$ o valor observado da estatística de teste. A probabilidade

$$\hat{\alpha} = P_{\theta_0}(T(X_1, \dots, X_n) \geq t)$$

é denominada probabilidade de significância do teste. Se o nível de significância exato do teste $\alpha = P_{\theta_0}(T \geq t_0)$

é maior que α rejeitamos a hipótese H_0 ao nível de α . Observe que se $\hat{\alpha} = P_{\theta_0}(T \geq t) \leq P_{\theta_0}(T \geq t_0) = \alpha$ é porque $t \geq t_0$ e devemos rejeitar H_0 ao nível α . Desta forma, basta calcular a probabilidade de significância e comparar com o nível de significância nominal adotado. Não é necessário determinar t_0 correspondente ao nível α . A probabilidade de significância fornece a probabilidade de, sob a hipótese, obter-se um valor tão extremo ou mais extremo que o observado. Se esta probabilidade é pequena, a hipótese deve ser rejeitada. Ou, ainda, dizemos que o valor t é significativo ao nível $\alpha \geq \hat{\alpha}$.

A função potência de um teste é $\pi: \Theta \rightarrow R$ dada por

$$\pi(\theta) = P_{\theta}(T \in R), \quad \theta \in \Theta,$$

ou seja, a probabilidade de rejeitar H_0 quando o valor do parâmetro é θ . Temos ainda

$$\begin{aligned} \pi(\theta) &= \alpha(\theta) \quad \text{para } \theta \in \Theta_0 \quad \text{e} \\ \pi(\theta) &= 1 - \beta(\theta) \quad \text{para } \theta \in \Theta - \Theta_0. \end{aligned}$$

Exemplo: No exemplo anterior, a função potência do teste é dada por

$$\pi(\theta) = P_{\theta}(X \geq 2) = 1 - \Phi(\sqrt{n}(2 - \theta)), \quad -\infty < \theta < \infty.$$

Um teste é não viciado se

$$\pi(\theta) \geq \sup_{\theta \in \Theta_0} \alpha(\theta) \quad \forall \theta \in \Theta - \Theta_0,$$

isto é, o valor da potência do teste para valores do parâmetro na alternativa é maior que o tamanho do teste.

Um teste de tamanho α é mais poderoso na alternativa $\theta_1 \in \Theta - \Theta_0$, se nenhum outro teste, de tamanho no máximo α , tem potência maior contra a mesma alternativa. Isto é, a potência π^* do teste mais poderoso em θ_1 , satisfaz:

$$\pi^*(\theta_1) \geq \pi(\theta_1)$$

onde π é qualquer outro teste tal que $\sup_{\theta \in \Theta_0} \pi(\theta) \leq \alpha$ e $\sup_{\theta \in \Theta} \pi^*(\theta) \geq \alpha$.

Se o teste tem tamanho α , e é mais poderoso contra qualquer alternativa $\theta \in \Theta - \Theta_0$, dizemos que ele é uniformemente mais poderoso de tamanho α .

Um teste é consistente na alternativa $\theta \in \Theta - \Theta_0$ se sua potência π satisfaz

$$\lim_{n \rightarrow \infty} \pi(\theta) = 1,$$

onde a estatística do teste é baseada em n v.a.'s amostrais X_1, \dots, X_n .

Relação entre testes e regiões de confiança

Consideremos a hipótese estatística

$$H(\theta_0): \theta = \theta_0 \text{ contra } H_A: \theta \neq \theta_0.$$

Quando θ_0 varia obtemos uma família de hipóteses esta-

tísticas $\{H(\theta_0)\}_{\theta_0 \in \Theta}$. Seja T a estatística do teste, e para cada hipótese $H(\theta_0)$, consideremos a região crítica $R(\theta_0)$ correspondente ao nível de significância α , isto é,

$$P_{\theta_0}(T \in R^c(\theta_0)) = 1-\alpha \quad \forall \theta_0 \in \Theta,$$

onde $R^c(\theta_0)$ é a complementar de $R(\theta_0)$, ou seja, a região de rejeição.

Seja, agora, para uma dada enupla de valores observados x_1, \dots, x_n , o conjunto

$$S(x_1, \dots, x_n) = \{\theta_0; T(x_1, \dots, x_n) \in R^c(\theta_0)\},$$

como $\theta \in S(x_1, \dots, x_n) \Leftrightarrow T(x_1, \dots, x_n) \in R^c(\theta_0)$ então

$$P_{\theta}(\theta \in S(X_1, \dots, X_n)) = P_{\theta}(T \in R^c(\theta)) = 1-\alpha, \quad \forall \theta \in \Theta$$

e $S(X_1, \dots, X_n)$ é uma região de confiança de nível de confiança $1-\alpha$ para θ .

Exemplo: Considere X_1, \dots, X_n v.a.'s i.i.d $N(\theta, \sigma^2)$

com σ^2 desconhecido. Seja testar a hipótese

$H(\theta_0): \theta = \theta_0$ contra a alternativa $H_A: \theta \neq \theta_0$. O teste

clássico para esta hipótese é o teste t-Student cuja região de aceitação é dada por

$$-t_{\alpha/2} < \frac{\bar{X} - \theta_0}{S/\sqrt{n}} < t_{\alpha/2}$$

onde α é o nível de significância do teste

$$s^2 = \Sigma (x_i - \bar{x})^2 / (n-1)$$

$t_{\alpha/2}$ - é o valor tal que $P(T_{n-1} > t_{\alpha/2}) = \alpha/2$ com T_{n-1} tendo distribuição t-Student com $n-1$ graus de liberdade. Deste modo

$$P_{\theta_0} \left(-t_{\alpha/2} < \frac{\bar{x} - \theta_0}{s/\sqrt{n}} < t_{\alpha/2} \right) = 1 - \alpha \quad \forall \theta_0 \in \Theta$$

e

$$P_{\theta_0} \left(\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \theta_0 < \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right) = 1 - \alpha \quad \forall \theta_0 \in \Theta$$

logo $(\bar{x} - t_{\alpha/2} s/\sqrt{n}, \bar{x} + t_{\alpha/2} s/\sqrt{n})$ é um intervalo de confiança para θ com nível de confiança $(1-\alpha)\%$.



T A B E L A S
= = = = = = =



TABELA A - Wilcoxon Duas Amostras

A tabela abaixo fornece, para os níveis mais próximos dos indicados, a probabilidade $P_H(W_{XY} \geq c)$.

Exemplo: Para $n=3$ e $m=5$ temos

$$P_H(W_{XY} \leq 8) = P_H(W_{XY} \geq 19) = 0,0714$$

Observação: Os valores das probabilidades estão multiplicados por 10^4 .

$n = 3$					
m	0.10	0.05	0.025	0.005	
3	(7, 14) 1000	(6, 15) 0500			
4	(7, 17) 0571	(6, 18) 0286			
	(8, 16) 1143	(7, 17) 0571	(6, 18) 0286		
5	(8, 19) 0714	(7, 20) 0357	(6, 21) 0179		
	(9, 18) 1250	(8, 19) 0714	(7, 20) 0357		
6	(9, 21) 0833	(8, 22) 0476	(7, 23) 0238		
	(10, 20) 1310	(9, 21) 0833	(8, 22) 0476	(6, 24) 0119	
7	(10, 23) 0917	(8, 25) 0333	(7, 26) 0167		
	(11, 22) 1333	(9, 24) 0583	(8, 25) 0333	(6, 27) 0083	
8	(11, 25) 0970	(9, 27) 0424	(8, 28) 0242		
	(12, 24) 1394	(10, 26) 0667	(9, 27) 0424	(6, 30) 0061	
$n = 4$					
m					
4	(12, 24) 0571	(11, 25) 0286	(10, 26) 0143		
	(13, 23) 1000	(12, 24) 0571	(11, 25) 0286	(10, 26) 0143	
	(14, 22) 1714				
5	(14, 26) 0952	(12, 28) 0317	(11, 29) 0159		
	(15, 25) 1429	(13, 27) 0556	(12, 28) 0317	(10, 30) 0079	
6	(15, 29) 0857	(13, 31) 0333	(12, 32) 0190	(10, 34) 0048	
	(16, 28) 1286	(14, 30) 0571	(13, 31) 0333	(11, 33) 0095	
7	(16, 32) 0818	(14, 34) 0364	(13, 35) 0212	(10, 38) 0030	
	(17, 31) 1152	(15, 33) 0545	(14, 34) 0364	(11, 37) 0061	
8	(17, 35) 0768	(15, 37) 0364	(14, 38) 0242	(11, 41) 0040	
	(18, 34) 1071	(16, 36) 0545	(15, 37) 0364	(12, 40) 0081	

(continua)

Tabela A - Wilcoxon Duas Amostras.

(continuação)

$\frac{n = 5}{m}$	0.10		0.05		0.025		0.005	
5	(20, 35)	0754	(19, 36)	0476	(17, 38)	0159	(15, 40)	0040
	(21, 34)	1111	(20, 35)	0754	(18, 37)	0278	(16, 39)	0079
6	(22, 38)	0887	(20, 40)	0411	(18, 42)	0152	(16, 44)	0043
	(23, 37)	1234	(21, 39)	0628	(19, 41)	0260	(17, 43)	0087
7	(23, 42)	0745	(21, 44)	0366	(20, 45)	0240	(16, 49)	0025
	(24, 41)	1010	(22, 43)	0530	(21, 44)	0366	(17, 48)	0051
8	(25, 45)	0855	(23, 47)	0466	(21, 49)	0225	(17, 53)	0031
	(26, 44)	1111	(24, 46)	0637	(22, 48)	0326	(18, 52)	0054
$\frac{n = 6}{m}$								
6	(30, 48)	0898	(28, 50)	0465	(26, 52)	0206	(23, 55)	0043
	(31, 47)	1201	(29, 49)	0660	(27, 51)	0325	(24, 54)	0076
7	(32, 52)	0903	(29, 55)	0367	(27, 57)	0175	(24, 60)	0041
	(33, 51)	1171	(30, 54)	0507	(28, 56)	0256	(25, 59)	0070
8	(34, 56)	0906	(31, 59)	0406	(29, 61)	0213	(25, 65)	0040
	(35, 55)	1142	(32, 58)	0539	(30, 60)	0296	(26, 64)	0063
$\frac{n = 7}{m}$								
7	(41, 64)	0825	(39, 66)	0487	(36, 69)	0189	(32, 73)	0035
	(42, 63)	1043	(40, 65)	0641	(37, 68)	0265	(33, 72)	0055
8	(44, 68)	0946	(41, 71)	0469	(38, 74)	0200	(34, 78)	0047
	(45, 67)	1159	(42, 70)	0603	(39, 73)	0270	(35, 77)	0070
$\frac{n = 8}{m}$								
8	(55, 81)	0974	(51, 85)	0415	(49, 87)	0249	(43, 93)	0035
	(56, 80)	1172	(52, 84)	0524	(50, 86)	0325	(44, 92)	0052

(continua)

TABELA 2 - Wilcoxon 1-amostra

A tabela abaixo fornece, para os níveis mais próximos de 1%, 2,5%, 5% e 10%, a probabilidade $P_H(V_S \geq c)$, onde V_S é a soma dos postos sinalizados.

Exemplo: Para $N=8$, $P_H(V_S \geq 28) = 0,027$.

N	C	P	N	C	P
1	1	0.500	12	58	0.010
2	3	0.250		50	0.026
3	6	0.125		44	0.046
4	10	0.062		34	0.102
	8	0.125	13	65	0.011
5	15	0.031		57	0.024
	13	0.062		49	0.047
	11	0.094		39	0.095
6	21	0.016	14	73	0.010
	19	0.031		63	0.025
	17	0.047		53	0.052
	13	0.109		43	0.097
7	28	0.008	15	80	0.011
	24	0.023		70	0.024
	20	0.055		60	0.047
	16	0.109		46	0.104
8	32	0.012	16	88	0.011
	28	0.027		76	0.025
	24	0.055		64	0.052
	20	0.098		52	0.096
9	39	0.010	17	97	0.010
	33	0.027		83	0.025
	29	0.049		71	0.049
	23	0.102		55	0.103
10	45	0.010	18	105	0.010
	39	0.024		91	0.024
	33	0.053		77	0.049
	27	0.097		61	0.098
11	52	0.009	19	114	0.010
	44	0.027		98	0.025
	38	0.051		82	0.052
	30	0.013		66	0.098
			20	124	0.010
				106	0.024
				90	0.049
				70	0.101

TABELA C - Kruskal -Wallis

A tabela abaixo fornece $P_H(K \geq h) = P$ para valores de h tendo o correspondente P próximo de uma dos valores: 0,99; 0,95; 0,90; 0,75; 0,50; 0,25; 0,10; 0,05 e 0,01.

Exemplo: Para $n_1=3, n_2=3, n_3=1$, temos

$$P_H(K \geq 3.10) = 0,267$$

N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P	N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P	
4	2	1	1	25.5	0.30	1.	6	3	2	1	73.8	0.10	1.	
				28.0	1.80	0.833					74.3	0.24	0.933	
				29.5	2.70	0.500					75.0	0.43	0.900	
5	3	1	1	46.3	0.53	1.	6	2	2	2	77.8	1.24	0.700	
				47.0	0.80	0.800					81.0	2.14	0.533	
				50.3	2.13	0.700					84.3	3.10	0.267	
				53.0	3.20	0.300					88.5	4.29	0.100	
5	2	2	1	45.0	0	1.	7	5	1	1	113.2	0.26	1.	
				46.0	0.40	0.933					114.0	0.43	0.905	
				46.5	0.60	0.867					117.2	1.11	0.762	
				48.5	1.40	0.733					122.8	2.31	0.524	
				50.0	2.00	0.600					125.2	2.83	0.333	
				51.0	2.40	0.467					130.0	3.86	0.143	
				52.5	3.00	0.333								
54.0	3.60	0.200												
6	4	1	1	74.0	0.14	1.								
				76.2	0.79	0.933								
				77.0	1.00	0.800								
				82.2	2.50	0.467								
				86.0	3.57	0.200								

(continua)

TABELA C - Kruskal -Wallis
(continuação)

N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P	N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P
7	4	2	1	112.0	0	1.	8	4	3	1	162.3	0.06	1.
				113.2	0.27	0.933					163.2	0.21	0.950
				113.5	0.32	0.895					164.6	0.43	0.900
				117.0	1.07	0.743					167.2	0.88	0.743
				121.5	2.04	0.495					173.0	1.83	0.514
				125.5	2.89	0.267					178.6	2.76	0.229
				130.8	4.02	0.114					186.3	4.06	0.093
				134.5	4.82	0.057					193.2	5.21	0.050
7	3	3	1	112.7	0.14	0.986	8	4	2	2	162.0	0	1.
				113.3	0.29	0.957					162.8	0.12	0.971
				114.7	0.57	0.871					164.0	0.33	0.890
				117.3	1.14	0.743					166.0	0.67	0.757
				121.3	2.00	0.514					172.8	1.79	0.514
				126.7	3.14	0.243					180.8	3.12	0.248
				133.3	4.57	0.100					188.8	4.46	0.100
				136.0	5.14	0.043					192.8	5.12	0.052
7	3	2	2	112.0	0	1.	8	3	3	2	162.2	0.03	1.
				112.8	0.18	0.971					163.3	0.22	0.946
				113.0	0.21	0.895					163.5	0.25	0.896
				115.3	0.71	0.743					166.2	0.69	0.757
				119.5	1.61	0.524					172.8	1.81	0.511
				128.0	3.43	0.248					180.8	3.14	0.243
				132.8	4.46	0.105					189.3	4.56	0.100
				134.0	4.71	0.048					192.8	5.14	0.061
8	6	1	1	162.5	0.08	1.	9	7	1	1	226.1	0.15	1.
				164.7	0.44	0.964					227.0	0.27	0.944
				165.2	0.53	0.893					229.6	0.61	0.917
				170.0	1.33	0.714					233.0	1.07	0.750
				174.5	2.08	0.500					241.6	2.21	0.500
				180.7	3.11	0.250					245.6	2.74	0.306
				186.5	4.08	0.107					257.0	4.27	0.083
				8	5	2					1	162.3	0.05
163.2	0.20	0.940	226.2				0.16	0.960					
164.7	0.45	0.905	227.7				0.36	0.889					
168.0	1.00	0.750	231.2				0.82	0.730					
173.5	1.92	0.488	238.5				1.80	0.500					
178.8	2.80	0.286	247.2				2.96	0.230					
187.2	4.20	0.095	256.5				4.20	0.095					
192.0	5.00	0.048	261.2				4.82	0.048					
193.5	5.25	0.036	267.0				5.60	0.024					

(continua)

TABELA C - Kruskal -Wallis
(continuação)

N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P	N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P				
9	5	3	1	225.5	0.07	0.992	9	3	3	3	225.7	0.09	0.993				
				226.3	0.18	0.952					227.0	0.27	0.929				
				227.5	0.34	0.889					227.7	0.36	0.879				
				231.1	0.82	0.750					231.0	0.80	0.721				
				238.3	1.78	0.488					237.7	1.69	0.511				
				246.3	2.84	0.258					249.0	3.20	0.254				
				255.1	4.02	0.095					259.7	4.62	0.100				
				261.5	4.87	0.052					267.0	5.60	0.050				
				273.0	6.40	0.012					273.7	6.49	0.011				
9	5	2	2	225.7	0.09	0.984	10	8	1	1	303.0	0.05	1.				
				226.0	0.13	0.937					305.5	0.33	0.933				
				226.8	0.24	0.913					307.0	0.49	0.889				
				230.8	0.77	0.759					313.0	1.15	0.733				
				237.7	1.69	0.495					321.1	2.03	0.489				
				248.5	3.13	0.254					329.5	2.95	0.244				
				257.8	4.37	0.090					343.0	4.42	0.067				
				262.8	5.04	0.056					10	7	2	1	302.8	0.03	1.
				274.0	6.53	0.008									305.1	0.28	0.939
			305.3	0.30	0.900												
9	4	4	1	225.5	0.07	0.987					310.8	0.90	0.744				
				226.2	0.17	0.968					319.0	1.80	0.494				
				227.2	0.30	0.911					328.5	2.84	0.244				
				231.5	0.87	0.759					341.0	4.20	0.100				
				238.2	1.77	0.498					345.6	4.71	0.050				
				245.2	2.70	0.260					356.5	5.89	0.017				
				255.5	4.07	0.102	10	6	3	1	303.0	0.05	0.983				
				262.2	4.97	0.048					304.2	0.18	0.952				
				275.0	6.67	0.010					305.5	0.33	0.921				
9	4	3	2	225.6	0.08	0.987					309.7	0.78	0.752				
				225.8	0.11	0.944					318.8	1.78	0.502				
				227.1	0.28	0.902					329.5	2.95	0.255				
				230.2	0.70	0.756					338.3	3.91	0.095				
				237.1	1.61	0.502					347.0	4.85	0.050				
				248.2	3.10	0.251					362.8	6.58	0.012				
				258.3	4.44	0.102											
				265.5	5.40	0.051											
				272.2	6.30	0.011											

(continua)

TABELA C - Kruskal -Wallis

(continuação)

N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P	N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P
10	6	2	2	303.2	0.07	0.984	10	4	3	3	302.9	0.05	0.984
				303.5	0.11	0.946					304.0	0.16	0.941
				305.2	0.29	0.911					305.6	0.34	0.895
				309.2	0.73	0.743					308.9	0.70	0.764
				318.5	1.75	0.500					317.3	1.62	0.497
				330.2	3.02	0.244					329.6	2.95	0.253
				343.2	4.44	0.108					345.6	4.70	0.101
				348.5	5.02	0.051					355.0	5.73	0.050
				362.5	6.55	0.011					364.3	6.75	0.010
10	5	4	1	303.0	0.06	0.983	11	9	1	1	397.1	0.10	1.
				304.2	0.19	0.952					400.4	0.40	0.945
				305.2	0.29	0.906					401.1	0.46	0.909
				310.0	0.82	0.762					407.8	1.07	0.745
				318.8	1.78	0.498					416.4	1.86	0.509
				329.0	2.90	0.251					426.4	2.77	0.255
				338.8	3.96	0.102					437.1	3.74	0.127
				347.0	4.86	0.056					446.0	4.55	0.055
				365.2	6.84	0.011							
10	5	3	2	303.1	0.07	0.981	11	8	2	1	396.6	0.06	0.990
				303.7	0.13	0.951					397.5	0.14	0.958
				305.0	0.28	0.901					400.1	0.38	0.901
				309.0	0.71	0.743					406.1	0.92	0.739
				317.3	1.61	0.502					415.6	1.78	0.501
				329.8	2.98	0.252					426.6	2.78	0.251
				343.7	4.49	0.101					440.1	4.01	0.093
				350.6	5.25	0.049					446.6	4.60	0.053
				365.0	6.82	0.010					463.5	6.14	0.012
10	4	4	2	303.0	0.05	0.988	11	7	3	1	396.5	0.04	0.994
				304.2	0.19	0.940					397.9	0.17	0.955
				305.0	0.27	0.893					400.3	0.39	0.903
				309.2	0.74	0.757					405.9	0.90	0.765
				317.5	1.64	0.510					415.1	1.74	0.503
				330.5	3.05	0.239					427.9	2.90	0.230
				344.2	4.55	0.098					438.3	3.84	0.105
				350.5	5.24	0.052					450.5	4.95	0.047
				365.5	6.87	0.011					469.1	6.65	0.011

(continua)

TABELA C - Kruskal -Wallis

(continuação)

<i>N</i>	<i>n</i> ₁	<i>n</i> ₂	<i>n</i> ₃	$\sum R_i^2/n_i$	<i>h</i>	<i>P</i>	<i>N</i>	<i>n</i> ₁	<i>n</i> ₂	<i>n</i> ₃	$\sum R_i^2/n_i$	<i>h</i>	<i>P</i>
11	7	2	2	396.6	0.06	0.990	11	5	4	2	396.4	0.04	0.992
				397.0	0.09	0.958					397.6	0.14	0.952
				398.6	0.24	0.895					398.8	0.25	0.902
				403.3	0.66	0.746					402.8	0.62	0.749
				413.6	1.60	0.509					413.5	1.59	0.499
				429.8	3.07	0.251					428.0	2.91	0.249
				445.8	4.53	0.099					445.7	4.52	0.101
				452.6	5.14	0.044					454.0	5.27	0.051
				468.6	6.60	0.011					474.3	7.12	0.010
11	6	4	1	396.4	0.04	0.993	11	5	3	3	396.5	0.05	0.994
				397.7	0.15	0.955					397.9	0.17	0.948
				399.9	0.36	0.895					398.7	0.24	0.902
				405.9	0.90	0.757					403.5	0.68	0.765
				415.2	1.75	0.499					412.8	1.53	0.505
				426.7	2.79	0.254					428.7	2.97	0.242
				440.4	4.04	0.094					445.9	4.53	0.097
				450.4	4.95	0.047					456.7	5.52	0.051
				473.9	7.08	0.010					472.8	6.98	0.011
11	6	3	2	396.5	0.05	0.992	11	4	4	3	396.5	0.05	0.993
				397.5	0.14	0.951					397.6	0.14	0.959
				398.7	0.24	0.900					399.6	0.33	0.890
				403.5	0.68	0.747					403.3	0.67	0.742
				414.0	1.64	0.496					413.0	1.55	0.503
				428.0	2.91	0.253					428.2	2.93	0.250
				446.0	4.55	0.101					446.0	4.55	0.099
				453.5	5.23	0.052					457.3	5.58	0.051
				472.7	6.97	0.009					474.6	7.14	0.010
11	5	5	1	396.4	0.04	0.994	12	10	1	1	509.6	0.20	0.985
				398.0	0.18	0.944					509.9	0.22	0.955
				400.4	0.40	0.885					513.9	0.53	0.909
				406.0	0.91	0.752					521.1	1.08	0.742
				415.2	1.75	0.493					531.5	1.88	0.500
				428.0	2.91	0.242					544.4	2.88	0.273
				440.4	4.04	0.105					557.6	3.89	0.106
				450.0	4.91	0.053					567.5	4.65	0.046
				476.4	7.31	0.009							

(continua)

TABELA C - Kruskal -Wallis

(continuação)

N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P	N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P
12	9	2	1	507.8	0.06	0.985	12	7	3	2	507.6	0.05	0.987
				509.5	0.19	0.958					508.3	0.10	0.954
				511.3	0.33	0.894					510.0	0.23	0.902
				517.8	0.83	0.748					516.0	0.69	0.752
				531.1	1.85	0.491					527.3	1.56	0.494
				545.5	2.96	0.248					546.0	3.00	0.251
				557.8	3.91	0.100					566.0	4.54	0.101
				569.9	4.84	0.048					576.5	5.34	0.050
				589.5	6.35	0.009					595.9	6.84	0.010
12	8	3	1	507.5	0.04	0.989	12	6	5	1	507.5	0.04	0.990
				509.8	0.22	0.946					509.5	0.19	0.950
				511.1	0.32	0.902					511.2	0.32	0.900
				517.5	0.80	0.746					518.0	0.84	0.752
				529.8	1.76	0.508					529.9	1.76	0.505
				544.1	2.86	0.256					543.3	2.79	0.263
				559.1	4.01	0.099					558.0	3.92	0.104
				570.5	4.88	0.048					569.9	4.84	0.051
				595.5	6.80	0.009					598.0	7.00	0.010
12	8	2	2	507.6	0.05	0.990	12	6	4	2	507.4	0.03	0.993
				508.5	0.12	0.952					508.7	0.13	0.951
				509.6	0.20	0.911					510.0	0.23	0.909
				516.6	0.74	0.757					516.4	0.72	0.743
				529.0	1.69	0.494					528.4	1.65	0.502
				546.6	3.05	0.251					546.0	3.00	0.252
				565.0	4.46	0.101					565.4	4.49	0.100
				572.6	5.05	0.053					575.4	5.26	0.050
				591.6	6.51	0.010					602.4	7.34	0.010
12	7	4	1	507.5	0.04	0.989	12	6	3	3	507.3	0.03	0.990
				509.5	0.20	0.947					508.5	0.12	0.962
				511.2	0.33	0.897					510.7	0.28	0.888
				517.6	0.81	0.754					516.0	0.69	0.761
				529.8	1.76	0.505					526.7	1.51	0.503
				543.5	2.81	0.260					545.8	2.99	0.249
				559.2	4.02	0.103					566.7	4.59	0.098
				571.3	4.95	0.052					580.0	5.62	0.050
				597.8	6.99	0.010					600.5	7.19	0.010

(continua)

TABELA C - Kruskal -Wallis

(continuação)

N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P	N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P	
12	5	5	2	507.6	0.05	0.988	13	9	3	1	638.1	0.07	0.987	
				509.2	0.17	0.947					639.4	0.16	0.950	
				510.3	0.25	0.896					642.3	0.35	0.894	
				516.0	0.69	0.749					649.0	0.79	0.754	
				528.4	1.65	0.496					664.8	1.83	0.496	
				546.3	3.02	0.243					680.3	2.86	0.247	
				565.6	4.51	0.100					698.8	4.07	0.099	
				575.2	5.25	0.051					710.8	4.86	0.052	
				601.5	7.27	0.010					737.0	6.59	0.010	
12	5	4	3	507.4	0.03	0.990	13	9	2	2	637.6	0.04	0.993	
				508.5	0.12	0.953					639.4	0.16	0.944	
				510.3	0.26	0.900					640.5	0.23	0.905	
				515.4	0.64	0.754					646.8	0.64	0.755	
				526.6	1.51	0.495					661.1	1.59	0.497	
				545.3	2.95	0.251					684.9	3.16	0.248	
				566.1	4.55	0.099					704.3	4.44	0.101	
				580.2	5.63	0.050					714.0	5.08	0.051	
				603.8	7.44	0.010					738.4	6.69	0.011	
12	4	4	4	507.5	0.04	0.994	13	8	4	1	638.1	0.07	0.987	
				509.0	0.15	0.941					639.1	0.14	0.951	
				510.5	0.27	0.913					642.0	0.33	0.901	
				516.5	0.73	0.746					649.5	0.82	0.748	
				526.5	1.50	0.510					664.2	1.80	0.502	
				545.0	2.92	0.252					679.4	2.79	0.255	
				567.5	4.65	0.097					698.2	4.04	0.099	
				581.0	5.69	0.049					712.2	4.96	0.050	
				605.0	7.54	0.011					742.5	6.96	0.010	
13	11	1	1	638.1	0.07	1.	13	8	3	2	637.5	0.03	0.995	
				641.4	0.29	0.962					639.0	0.13	0.947	
				642.8	0.38	0.910					641.0	0.26	0.896	
				654.5	1.15	0.744					647.0	0.66	0.752	
				666.5	1.94	0.500					660.5	1.55	0.500	
				682.8	3.02	0.231					682.8	3.02	0.253	
				698.1	4.03	0.090					704.5	4.45	0.101	
				709.0	4.75	0.038					717.6	5.32	0.050	
											743.0	6.99	0.010	
13	10	2	1	637.6	0.04	0.993								
				639.4	0.16	0.946								
				641.6	0.30	0.904								
				649.4	0.82	0.751								
				664.9	1.84	0.513								
				680.5	2.87	0.247								
				697.6	4.00	0.105								
				710.4	4.84	0.049								
				734.5	6.43	0.009								

(continua)

TABELA C - Kruskal -Wallis

(continuação)

N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P	N	n_1	n_2	n_3	$\sum R_i^2/n_i$	h	P
13	7	5	1	638.1	0.08	0.988	13	6	5	2	637.7	0.04	0.986
				638.9	0.13	0.954					639.4	0.16	0.942
				641.8	0.32	0.904					640.3	0.22	0.903
				649.3	0.81	0.749					646.7	0.64	0.751
				664.2	1.79	0.500					660.5	1.55	0.499
				679.3	2.79	0.251					682.7	3.01	0.250
				698.2	4.04	0.098					704.9	4.47	0.100
				712.8	5.00	0.051					717.7	5.32	0.051
			742.9	6.99	0.010					747.7	7.30	0.010	
13	7	4	2	637.6	0.04	0.986	13	6	4	3	637.5	0.03	0.988
				638.6	0.11	0.949					638.8	0.12	0.950
				640.5	0.23	0.899					640.8	0.25	0.907
				646.6	0.63	0.753					646.9	0.65	0.748
				660.5	1.55	0.502					660.2	1.53	0.499
				682.3	2.99	0.251					680.9	2.90	0.251
				706.0	4.55	0.100					706.8	4.60	0.100
				718.3	5.36	0.051					722.0	5.60	0.050
			748.0	7.32	0.010					750.2	7.47	0.010	
13	7	3	3	637.5	0.03	0.996	13	5	5	3	637.5	0.04	0.989
				638.9	0.13	0.948					638.7	0.11	0.951
				641.3	0.28	0.895					640.6	0.24	0.902
				648.0	0.72	0.747					647.0	0.66	0.751
				660.0	1.51	0.498					659.9	1.51	0.497
				681.6	2.94	0.249					681.5	2.94	0.246
				705.6	4.52	0.101					705.9	4.55	0.100
				722.0	5.60	0.050					722.3	5.63	0.051
			746.6	7.23	0.010					751.4	7.54	0.010	
13	6	6	1	638.2	0.08	0.987	13	5	4	4	637.4	0.03	0.996
				640.0	0.20	0.946					638.8	0.12	0.952
				642.2	0.34	0.897					640.4	0.23	0.903
				649.0	0.79	0.744					646.7	0.64	0.745
				664.2	1.79	0.504					660.2	1.53	0.501
				679.0	2.77	0.256					681.2	2.92	0.249
				697.7	4.00	0.098					707.0	4.62	0.100
				710.7	4.86	0.051					722.2	5.62	0.050
			744.2	7.07	0.010					754.7	7.76	0.009	

(continua)

TABELA D - Friedman

A tabela abaixo fornece $P_H(Q \geq c) = P$, para valores de c tendo o correspondente P próximo de um dos valores: 0,01; 0,05; 0,10; 0,25; 0,50; 0,75 e 0,95.

Exemplo: Para $N=3$ (blocos) e $s=5$ (tratamentos)

$$P_H(Q \geq 1,60) = 0,522$$

s	c	P	s	c	P	s	c	P	s	c	P	
2	0.	1.	7	0.29	0.964	11	0.18	0.976	15	0.13	0.982	
	1.00	0.833		0.86	0.768		0.73	0.732		0.93	0.711	
	3.00	0.500		2.00	0.486		1.64	0.470		1.60	0.513	
	4.00	0.167		3.43	0.237		2.91	0.256		2.80	0.267	
3	0.67	0.944	8	4.57	0.112	12	4.91	0.100	8.93	4.93	0.096	
		2.00		0.528	6.00		0.051	6.54		0.043	6.40	0.047
		4.67		0.194	8.86		0.008	8.91		0.011	8.93	0.010
		6.00		0.028	0.25		0.967	0.17		0.978	0.75	0.751
4	0.50	0.931	9	0.75	0.794	13	1.50	0.500	8.00	0.62	0.767	
		1.50		0.653	1.75		0.531	3.17		0.249	3.17	0.249
		2.00		0.431	4.75		0.120	4.67		0.108	4.67	0.108
		3.50		0.273	6.25		0.047	6.17		0.050	6.17	0.050
		4.50		0.125	9.00		0.010	8.67		0.011	8.67	0.011
		6.50		0.042	0.22		0.971	0.15		0.980	0.62	0.767
5	0.40	0.954	10	0.67	0.814	14	1.38	0.527	9.00	0.57	0.781	
		1.20		0.691	1.56		0.569	2.92		0.278	2.92	0.278
		1.60		0.522	2.89		0.278	4.67		0.107	4.77	0.098
		3.60		0.182	6.22		0.048	6.00		0.050	6.00	0.050
		5.20		0.093	8.67		0.010	8.67		0.012	8.67	0.012
		6.40		0.039	0.20		0.974	0.14		0.981	0.80	0.710
6	0.33	0.956	8.00	0.80	0.710	9.00	0.57	0.781	9.00	0.57	0.781	
		1.00		0.740	1.80		0.436	1.71		0.489	1.71	0.489
		2.33		0.430	3.20		0.222	3.00		0.242	3.00	0.242
		3.00		0.252	5.00		0.092	5.14		0.089	5.14	0.089
		5.33		0.072	6.20		0.046	6.14		0.049	6.14	0.049
		6.33		0.052	8.60		0.012	9.00		0.010	9.00	0.010
9.00	0.008											

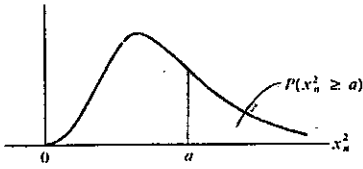
(continua)

TABELA D - Friedman

(continuação)

N=4

	s	c	P	s	c	P
2	0.60	0.958		8	0.45	0.957
	1.80	0.792			1.35	0.754
	3.00	0.542			2.55	0.500
	4.80	0.208			4.20	0.247
	6.00	0.042			6.30	0.098
3	0.60	0.958		7.65	0.049	
	1.80	0.727		10.35	0.010	
	2.60	0.524				
	4.20	0.293				
	6.60	0.075				
	7.00	0.054				
	8.20	0.017				
4	0.60	0.930				
	1.50	0.753				
	2.70	0.513				
	4.50	0.237				
	6.00	0.106				
	7.50	0.054				
	9.30	0.011				
5	0.60	0.944				
	1.32	0.769				
	2.52	0.520				
	4.20	0.266				
	6.12	0.102				
	7.80	0.049				
6	9.96	0.009				
	0.40	0.952				
	1.40	0.779				
	2.60	0.517				
	4.20	0.259				
	6.20	0.109				
	7.60	0.043				
10.20	0.010					
7	0.43	0.964				
	1.46	0.754				
	2.49	0.533				
	4.37	0.239				
	6.26	0.101				
	7.63	0.051				
	10.37	0.009				



Graus de Liberdade	Níveis de Probabilidade						
	0.40	0.30	0.20	0.10	0.05	0.01	0.001
1	0.71	1.07	1.64	2.71	3.84	6.63	10.83
2	1.83	2.41	3.22	4.61	5.99	9.21	13.82
3	2.95	3.66	4.64	6.25	7.81	11.34	16.27
4	4.04	4.88	5.99	7.78	9.49	13.28	18.47
5	5.13	6.06	7.29	9.24	11.07	15.09	20.52
6	6.21	7.23	8.56	10.64	12.59	16.81	22.46
7	7.28	8.38	9.80	12.02	14.07	18.48	24.32
8	8.35	9.52	11.03	13.36	15.51	20.09	26.12
9	9.41	10.66	12.24	14.68	16.92	21.67	27.88
10	10.47	11.78	13.44	15.99	18.31	23.21	29.59
11	11.53	12.90	14.63	17.28	19.68	24.72	31.26
12	12.58	14.01	15.81	18.55	21.03	26.22	32.91
13	13.64	15.12	16.98	19.81	22.36	27.69	34.53
14	14.69	16.22	18.15	21.06	23.68	29.14	36.12
15	15.73	17.32	19.31	22.31	25.00	30.58	37.70
16	16.78	18.42	20.47	23.54	26.30	32.00	39.25
17	17.82	19.51	21.61	24.77	27.59	33.41	40.79
18	18.87	20.60	22.76	25.99	28.87	34.81	42.31
19	19.91	21.69	23.90	27.20	30.14	36.19	43.82
20	20.95	22.77	25.04	28.41	31.41	37.57	45.31
21	21.99	23.86	26.17	29.62	32.67	38.93	46.80
22	23.03	24.94	27.30	30.81	33.92	40.29	48.27
23	24.07	26.02	28.43	32.01	35.17	41.64	49.73
24	25.11	27.10	29.55	33.20	36.42	42.98	51.18
25	26.14	28.17	30.68	34.38	37.65	44.31	52.62
26	27.18	29.25	31.79	35.56	38.89	45.64	54.05
27	28.21	30.32	32.91	36.74	40.11	46.96	55.48
28	29.25	31.39	34.03	37.92	41.34	48.28	56.89
29	30.28	32.46	35.14	39.09	42.56	49.59	58.30
30	31.32	33.53	36.25	40.26	43.77	50.89	59.70

(continua)

Graus de Liberda- de	Níveis de Probabilidade						
	0.99	0.95	0.90	0.80	0.70	0.60	0.50
1	0.00	0.00	0.02	0.06	0.15	0.27	0.45
2	0.02	0.10	0.21	0.45	0.71	1.02	1.39
3	0.11	0.35	0.58	1.01	1.42	1.87	2.37
4	0.30	0.71	1.06	1.65	2.19	2.75	3.36
5	0.55	1.15	1.61	2.34	3.00	3.66	4.35
6	0.87	1.64	2.20	3.07	3.83	4.57	5.35
7	1.24	2.17	2.83	3.82	4.67	5.49	6.35
8	1.65	2.73	3.49	4.59	5.53	6.42	7.34
9	2.09	3.33	4.17	5.38	6.39	7.36	8.34
10	2.56	3.94	4.87	6.18	7.27	8.30	9.34
11	3.05	4.57	5.58	6.99	8.15	9.24	10.34
12	3.57	5.23	6.30	7.81	9.03	10.18	11.34
13	4.11	5.89	7.04	8.63	9.93	11.13	12.34
14	4.66	6.57	7.79	9.47	10.82	12.08	13.34
15	5.23	7.26	8.55	10.31	11.72	13.03	14.34
16	5.81	7.96	9.31	11.15	12.62	13.98	15.34
17	6.41	8.67	10.09	12.00	13.53	14.94	16.34
18	7.01	9.39	10.86	12.86	14.44	15.89	17.34
19	7.63	10.12	11.65	13.72	15.35	16.85	18.34
20	8.26	10.85	12.44	14.58	16.27	17.81	19.34
21	8.90	11.59	13.24	15.44	17.18	18.77	20.34
22	9.54	12.34	14.04	16.31	18.10	19.73	21.34
23	10.20	13.09	14.85	17.19	19.02	20.69	22.34
24	10.86	13.85	15.66	18.06	19.94	21.65	23.34
25	11.52	14.61	16.47	18.94	20.87	22.62	24.34
26	12.20	15.38	17.29	19.82	21.79	23.58	25.34
27	12.88	16.15	18.11	20.70	22.72	24.54	26.34
28	13.56	16.93	18.94	21.59	23.65	25.51	27.34
29	14.26	17.71	19.77	22.48	24.58	26.48	28.34
30	14.95	18.49	20.60	23.36	25.51	27.44	29.34

REFERÊNCIAS

- GNEDENKO, B.V. - The theory of probability, Moscou, MIR, 1969.
- LEHMANN, E.L. - Nonparametrics: Statistics methods based on ranks. San Francisco, Holden-Day, 1975.
- HOLLANDER, M. & WOLFE, D.A. - Nonparametric statistical methods. New York, John Wiley, 1973.
- HÁJEK, J. - A course in nonparametric statistics. San Francisco, Holden-Day, 1969.
- COCHRAN, W.G. & SNEDECOR, G.W. - Statistical methods. 5 ed. Iowa, The Iowa State College Press, 1956.
- HODGES, J.L. Jr. - The significance probability of the Sminorv. Two sample test. Arkiv Mat., 3: 469-486, 1958.
- OWEN, D.B. - Handbook of statistical tables. Reading, Mass., Addison Wesley, 1962.

