TESTES NÃO-PARAMÉTRICOS PARA ANÁLISE DE ALGUNS EXPERIMENTOS COM MEDIDAS REPETIDAS

Este exemplar corresponde a redação final da teses devidamente corregida e defendida pelo Sr. MIGUEL ANGEL URIBE OPAZO e aprovada pela Comissão Julgadora.

Campinas, 2 de Agosto de 1991.

Prof. Dr. ~~BELMER~~ GARCIA NEGRILLO

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de Mestre em ESTATISTICA.

TESTES NÃO-PARAMÉTRICOS PARA ANÁLISE DE ALGUNS EXPERIMENTOS COM MEDIDAS REPETIDAS

MIGUEL ANGEL URIBE OPAZO

Orientador :

Prof. Dr. BELMER GARCIA NEGRILLO

IMECC-UNICAMP

1991

AGRADECIMENTOS

A Deus, que nos concedeu a vida, a inspiração e a paciência necessárias para chegar ao final deste trabalho.

Aos meus familiares, que sempre me incentivaram para continuar, principalmente nos momentos difíceis e que souberam compreender os momentos em que fiquei ausente do seu convívio.

Ao meu Orientador e amigo Belmer Garcia Negrillo, que sirviu-me de exemplo profissional e me brinda com um grande apoio para que eu siga adiante.

Ao Professor Dalton pelo apoio com que me brindou.

Aos amigos, que me ap o iaram e que de alguma forma contribuíram comigo.

Aos Professores, que me acompanharam desde o início de minha formação profissional.

Y en especial, a mi Esposa ESTANIS, que la Amo Mucho.

A todos, obrigado por tudo!

SUMÁRIO

INTRODUÇÃO

Muitos dados provenientes dos mais diversos experimentos, usados para análise estatística, são confrontados com resultados que não satisfazem suposições tais como a normalidade da distribuição probabilística dos dados, aditividade dos efeitos presentes no modelo.

Isto tem gerado problemas de decisão no que diz respeito aos testes elaborados, como por exemplo, como agir se não é satisfeita a normalidade.

A teoria dos Métodos Não-Paramétricos é essencialmente relacionada com o desenvolvimento de procedimentos de inferência estatística, que não exigem qualquer suposição explícita acerca da distribuição de probabilidades das observações amostrais. Isto significa dizer que, para aqueles dados que não atendem a pressuposição de normalidade, a estatística não-paramétrica surge como uma solução.

A importância deste trabalho em relação aos demais está no fato de utilizar escores selecionados pela análise exploratória de dados obtendo testes mais eficientes e localmente mais poderosos, pois a maioria dos trabalhos anteriores só utiliza os escores de Wilcoxon levando-nos a cometer erros, em muitos casos, superiores que aqueles cometidos utilizando a teoria paramétrica quando não são satisfeitas as pressuposições mencionadas anteriormente.

Este trabalho propõem o uso de técnicas não-paramétricas para a análise de experimentos em modelos hierárquicos com dois fatores fixos, blocos aleatórios não balanceados, modelos mistos e modelo em parcelas subdivididas com medidas repetidas. O desenvolvimento está mais confinado à inferência estatística para grandes amostras.

No capítulo I, apresentaremos o desenvolvimento da teoria Não-Paramétrica, tais como estatística linear de postos e testes localmente mais poderosos, que utilizaremos nos outros capítulos.

No capítulo II, apresentaremos a teoria Não-Paramétrica usando a estatística linear de postos para a análise de experimentos hierárquicos com dois fatores fixos, análise do delineamento em blocos aleatórios não balanceados e comparações múltiplas em cada análise.

No capítulo III, apresentaremos a análise Não-Paramétrica usando a estatística linear de postos, em modelos mistos em experimentos com medidas repetidas , experimento em "SPLIT PLOT" completamente casualizado e modelo "SPLIT PLOT" em experimento com medidas repetidas. Cabe ressaltar que Gary Koch (1968, 1969 e 1985), trabalhou estes tipos de análises para o caso de usar o escore de Wilcoxon; com base nisto, nós generalizamos para uma função escore que satisfaz as condições dadas no capítulo I.

No capítulo IV, apresentaremos algumas aplicações dos modelos mencionados nos capítulos anteriores, usando para a análise computacional os recursos do CM-SOC da EMBRAPA e o SENP da UNICAMP-ESALQ/USP (CIAGRI).

Para um leitor menos acostumado com o formalismo estatístico que deseje utilizar as aplicações desenvolvidas neste trabalho, a leitura do capítulo I poderá ser omitida, passando diretamente para o capítulo II.

CAPÍTULO I

DESENVOLVIMENTO DA TEORIA NÃO-PARAMÉTRICA

Com a finalidade de desenvolver a teoria não-paramétrica apresentaremos as seguintes definições e teoremas mais importantes que serão utilizados nos capítulos seguintes.

1.1 - DEFINIÇÕES

Definição 1.1.1- PERMUTAÇÕES.

Dado um número inteiro $N \geq 2$, podemos considerar todos os pontos $r = \{ r_1, r_2, \ldots, r_N \}$ obtido de $\varepsilon = \{ 1,2,3,\ldots,N \}$ através de redistribuição das coordenadas, chamamos r uma permutação.

O conjunto de permutações $r = \{r_1, r_2,\ldots,r_N\}$ forma um espaço que denotaremos por $\mathbb{R}$. O número de elementos de $\mathbb{R}$ é $N!$.

Definição 1.1.2.- COMPOSIÇÃO DE PERMUTAÇÕES

Se $v, r, s \in \mathbb{R}$, a permutação v é uma composição das permutações r e s, que denotaremos $v = r \circ s$, se $v_i = r(s_i)$, $1 \leq i \leq N$.

A composição de permutações não é comutativa, isto é, $r \circ s \neq s \circ r$.

Definição 1.1.3.- PERMUTAÇÃO INVERSA

Dizemos que d é uma permutação inversa de r se : $d \circ r = r \circ d = \varepsilon$, isto é, $d(r_i) = i$, para $1 \le i \le N$, e que podendo ser notada por r^{-1}.

1.2 - DISTRIBUIÇÃO SOBRE O ESPAÇO DE PERMUTAÇÕES

Seja $R = (R_1, R_2, \ldots, R_N)$ um vetor aleatório com valores em $\mathbb{R}$, dizemos que R tem distribuição uniforme sobre $\mathbb{R}$ se:

$$P[\ R = r] = \frac{1}{N!} \quad , \text{ para todo } r \in \mathbb{R} \tag{1.2.1.1}$$

TEOREMA 1.2.1

Se $R = (R_1, R_2, \ldots, R_N)$ é uniformemente distribuída sobre $\mathbb{R}$, então:

1) $P\ [\ R_i = k\ ] = \dfrac{1}{N} \qquad 1 \le i, k \le N$

2) $P\ [\ R_i = k\ ,\ R_j = h\ ] = \dfrac{1}{N(N-1)} \qquad i \ne j\ ,\ k \ne h$

$1 \le i, j, k, h \le N$

3) $P\ [\ R_i = k\ ,\ R_j = k\ ] = 0 \qquad 1 \le i \ne j,\ k \le N$

Prova

1) $\text{Card}\ (R = (R_1, R_2, \ldots, R_N)\ |\ R_i = k) = (N-1)!$

$$P[\,R_i = k\,] = \frac{(N-1)!}{N!} = \frac{1}{N} \qquad 1 \le i\ ,\ k \le N$$

2) $\text{Card}\ (R = (R_1, R_2, \ldots, R_N)\ |\ R_i = k\ ,\ R_j = h\ ,\ i \neq j\ ) = (N-2)!$

assim temos:

$$P\ [R_i = k\ ,\ R_j = h\ ] = \frac{(N-2)!}{N!}$$

$$P\ [\ R_i = k\ ,\ R_j = h\ ] = 1/(N(N-1)) \qquad 1 \le i \neq j,\ h \neq k \le N$$

3) $P\ [\ R_i = k\ ,\ R_j = k\ ] = P\ [\text{evento impossível}\ ]$

$$= P\ [\phi\ ] = 0. \qquad 1 \le i \neq j\ ,\ k \le N$$

TEOREMA 1.2.2

Se R é uniformemente distribuída em $\mathbb{R}$ então :

1) $$E(R_i) = \frac{N+1}{2}$$

2) $$VAR(R_i) = \frac{N^2-1}{12}$$

3) $$COV(R_i\ , R_j) = -\frac{N+1}{12} \qquad 1 \le i \neq j \le N$$

PROVA

1) $E(R_i) = \sum_{j=1}^{N} j \; P\,[R_i = j] = \sum_{j=1}^{N} j \; \frac{1}{N}$

$$= \frac{1}{N} \sum_{j=1}^{N} j = \frac{1}{N} \; \frac{N(N+1)}{2}$$

o que résulta:

$$E(R_i) = \frac{N+1}{2}$$

2) $VAR(R_i) = E\,[R_i - E(R_i)]^2$

$$= E(R_i^2) - 2E[\,R_i E(R_i)\,] + [E(R_i)]^2$$

$$= E\,(R_i^2) - [E(R_i)]^2$$

$$E(R_i^2) = \sum_{j=1}^{N} j^2 \; P\,[R_i = j]$$

$$= \sum_{j=1}^{N} j^2 \; \frac{1}{N} = \frac{1}{N} \sum_{j=1}^{N} j^2$$

$$= \frac{1}{N} \; \frac{N(N+1)(2N+1)}{6}$$

$$E(R_i^2) = \frac{(N+1)(2N+1)}{6}$$

de (1), concluimos que:

$$VAR(R_i) = \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4}$$

$$Var(R_i) = \frac{N^2-1}{12}$$

3) $$COV(R_i,R_j) = \sum_{h \neq k}^{N}\sum^{N} (R_h - E(R_h))(R_k - E(R_k)) \; P[R_i=h, R_j=k]$$

$$= \frac{1}{N(N-1)} \sum_{h \neq k}^{N}\sum^{N} (R_h - E(R_h))(R_k - E(R_k))$$

sabemos que : $$(\sum_{i=1}^{N} x_i)^2 = \sum_{i=1}^{N} x_i^2 + \sum_{i \neq j}^{N}\sum^{N} x_i x_j$$

$$\Rightarrow \quad \sum_{i \neq j}^{N}\sum^{N} x_i x_j = (\sum_{i=1}^{N} x_i)^2 - \sum_{i=1}^{N} x_i^2$$

do que resulta:

$$COV(R_i,R_j) = \frac{1}{N(N-1)} \left[\left(\sum_{i=1}^{N} (R_i - E(R_i)) \right)^2 - \sum_{i=1}^{N} (R_i - E(R_i))^2 \right]$$

$$= \frac{1}{N(N-1)} \left[0 - \sum_{i=1}^{N} (R_i - E(R_i))^2 \right]$$

$$= - \frac{1}{N(N-1)} \sum_{i=1}^{N} (R_i - E(R_i))^2$$

$$= - \frac{1}{N(N-1)} \; \frac{N^2 - 1}{12}$$

$$COV(R_i, R_j) = - \frac{N - 1}{12}$$

1.3 - ESTATÍSTICA LINEAR DE POSTOS

Se $X_1, X_2, \ldots, X_N$ é uma amostra aleatória com função distribuição F absolutamente contínua e se $R_1, R_2, \ldots, R_N$ são os postos respectivos de $X_1, X_2, \ldots, X_N$, então a estatística:

$$S = \sum_{i=1}^{N} c_i a(R_i) \qquad (1.3.1.1)$$

é chamada estatística linear de postos.

onde:

c_i são constantes, chamadas constantes de regressão, que servem para identificar as variações experimentais ou amostrais, isto é, a qual sub-amostras pertencem.

$a(R_i)$ é a função escore, avaliada nos postos R_i para todo $i=1,2,\ldots,N$, que será analisada na seção (1.5).

A estatística linear de postos **S** é utilizada para fazer inferência concernente a uma seqüência de observações $X_1, X_2, \ldots, X_N$, onde cada observação é obtida, geralmente, sob diferentes condições experimentais e desejamos decidir se estas variações causam modificações nas distribuições de observações individuais.

A seleção de escores depende da forma da distribuição adjacente e dos parâmetros, que são sensíveis às variações nas condições experimentais; enquanto as constantes de regressão depende da relação de regressão entre os mencionados parâmetros e algumas características numéricas das condições experimentais.

Para determinar a distribuição adjacente dos dados é recomendável fazer uma análise exploratória a qual será apresentada na seção (1.6).

1.3.1.-Distribuição Assintótica da Estatística Linear de Postos S

Agora vamos determinar a função de distribuição assintótica de S, para o qual daremos as condições nas funções escores e constrastes de regressão para que a distribuição de S seja simétrica e assintótica normal.

TEOREMA 1.3.1

Se R é uniformemente distribuída sobre $\mathbb{R}$ e $S = \sum_{i=1}^{N} c_i a(R_i)$ então :

1) $E(S) = N\,\bar{c}\,\bar{a}$.

2) $\operatorname{Var}(S) = \dfrac{1}{N-1} \sum_{i=1}^{N} (c_i - \bar{c})^2 \sum_{i=1}^{N} (a(R_i) - \bar{a})^2$.

onde :

$$\bar{a} = \frac{1}{N} \sum_{i=1}^{N} a(R_i)$$

$$\overline{c} = \frac{1}{N} \sum_{i=1}^{N} c_i$$

Prova

$$1)\ E(S) = E\left(\sum_{i=1}^{N} c_i a(R_i)\right) = \sum_{i=1}^{N} c_i E(a(R_i))$$

$$= \sum_{i=1}^{N} c_i \sum_{i=1}^{N} a(R_i)\ P\ [\ R_i = j\ ]$$

$$= \sum_{i=1}^{N} c_i \sum_{i=1}^{N} a(R_i)\ \frac{1}{N}$$

$$= \sum_{i=1}^{N} c_i\ \overline{a}$$

$$= N\ \overline{c}\ \overline{a}\ .$$

$$2)\ VAR(S) = \sum_{i=1}^{N} c_i VAR\ [a(R_i)] + \sum_{i \neq j}^{N} \sum^{N} c_i c_j COV\ [a(R_i), a(R_j)\ ] \qquad (1.3.1.1)$$

determinemos primeiro:

$$VAR\ (a(R_i)) = \sum_{j=1}^{N} (a(R_i) - \overline{a}\ )^2\ P\ [\ R_i = j\ ]$$

$$= \sum_{j=1}^{N} (a(R_i) - \overline{a}\ )^2\ \frac{1}{N} \qquad (1.3.1.2)$$

$$COV(a(R_i),\ a(R_j)) = \sum_{i \neq j}^{N} \sum^{N} (a(R_i) - \overline{a}\ )(a\ (R_j - \overline{a}\ )P\ [R_i = h\ , R_j = k\ ]$$

$$= \sum_{i\neq j}^{N}\sum^{N} (a(R_i) - \overline{a})(a\ (R_j - \overline{a}\ )\ \frac{1}{N(N-1)}$$

sabemos que : $(\sum_{i=1}^{N} x_i\ )^2 = \sum_{i=1}^{N} x_i^2 + \sum_{i\neq j}^{N}\sum^{N} x_i x_j$ (1.3.1.3)

$$COV(a(R_i, a(R_j)) = \frac{1}{N(N-1)} \left[\left(\sum_{i=1}^{N} (a(R_i) - a(R_i)) \right)^2 - \sum_{i=1}^{N} (a(R_i) - a(R_i))^2 \right]$$

$$= \frac{1}{N(N-1)} \left[\ 0 - \sum_{i=1}^{N} (a(R_i) - a(R_i))^2 \right] \quad (1.3.1.4)$$

substituindo (1.3.1.2) e (1.3.1.4) em (1.3.1.1) temos:

$$VAR(S) = \frac{1}{N} \sum_{i=1}^{N} c_i^2 \sum_{i=1}^{N} (a(R_i) - \overline{a}\ )^2 - \left[\sum_{i\neq j}^{N}\sum^{N} c_i c_j \sum_{i=1}^{N} (a(R_i) - \overline{a}\ )^2 \right] \frac{1}{N(N-1)}$$

$$= \frac{1}{N} \sum_{i=1}^{N} (a(R_i) - \overline{a}\ )^2 - \left[\sum_{i=1}^{N} c_i^2 - \sum_{i\neq j}^{N}\sum^{N} c_i c_j\ \frac{1}{N-1} \right]$$

$$= \frac{1}{N(N-1)} \sum_{i=1}^{N} (a(R_i) - \overline{a}\ )^2 \left[N \sum_{i=1}^{N} c_i^2 - \left(\sum_{i=1}^{N} c_i^2 + \sum_{i\neq j}^{N}\sum^{N} c_i c_j \right) \right]$$

de (1.3.1.3) temos :

$$VAR(S) = \frac{1}{N(N-1)} \sum_{i=1}^{N} (a(R_i) - \overline{a}\ )^2 \left[N \sum_{i=1}^{N} c_i^2 - \left(\sum_{i=1}^{N} c_i \right)^2 \right]$$

$$VAR(S) = \frac{1}{N(N-1)} \sum_{i=1}^{N} (a(R_i) - \bar{a})^2 \left[N \sum_{i=1}^{N} c_i^2 - N^2 \bar{c}^{\,2} \right]$$

$$= \frac{1}{N-1} \sum_{i=1}^{N} (a(R_i) - \bar{a})^2 \, N \left[\sum_{i=1}^{N} c_i^2/N - \bar{c}^{\,2} \right]$$

$$= \frac{1}{N-1} \sum_{i=1}^{N} (a(R_i) - \bar{a})^2 \, N \left[\sum_{i=1}^{N} (c_i - \bar{c})^2/N \right]$$

$$VAR(S) = \frac{1}{N-1} \sum_{i=1}^{N} (a(R_i) - \bar{a})^2 \sum_{i=1}^{N} (c_i - \bar{c})^2$$

TEOREMA 1.3.2 (Simetria da Estatística Linear de Postos)

Se R é uniformemente distribuída em $\mathbb{R}$ com :

$a(R_i) + a(N - R_i + 1) = k \qquad 1 \leq i \leq N$

ou

$c_i + c_{N-i+1} = p \qquad 1 \leq i \leq N$

onde k e p são constantes, então, a distribuição da estatística linear de posto S é simétrica em E (S) , isto é,

$P[\, S = E(S) + s \,] = P[\, S = E(S) - s \,]$, para todo s real.

PROVA

Suponhamos que:

$$a(R_i) + a(N - R_i + 1) = k \quad , \quad k \text{ constante} \tag{1.3.2.1}$$

somando para $i=1,2,...,N$ temos que:

$$\sum_{i=1}^{N} a(R_i) + \sum_{i=1}^{N} a(N - R_i + 1) = \sum_{i=1}^{N} k$$

$$2\sum_{i=1}^{N} a(R_i) = Nk$$

pelo que : $k = 2\,\overline{a}$

em (1.3.2.1) temos :

$$a(R_i) + a(N - R_i + 1) = 2\,\overline{a}$$

reagrupando convenientemente temos:

$$a(R_i) - \overline{a} = \overline{a} - a(N-R_i + 1) \tag{1.3.2.2}$$

considere-se a estatística

$$T = \sum_{i=1}^{N} c_i a(N-R_i+1) \tag{1.3.2.3}$$

que tem a mesma distribuição que S ,

assim obtemos:

$$P[\,S = E(S) + s\,] = P[\,T = E(T) + s\,]$$

para provar a simetria, é suficiente provar que os eventos $S = E(S) - s$ e $T = E(T) + s$ são equivalentes.

Dado $R = r$, o evento $T = E(T) + s$ pode ser escrito como:

$$\sum_{i=1}^{N} c_i a(N - r_i + 1) = \bar{a}\sum_{i=1}^{N} c_i + s \qquad (1.3.2.4)$$

$$\sum_{i=1}^{N} c_i [\, a(N - r_i + 1) - \bar{a} \,] = s$$

por (1.3.2.2) temos :

$$a(N - r_i + 1) - \bar{a} = \bar{a} - a(r_i)$$

do que resulta:

$$\sum_{i=1}^{N} c_i [\, \bar{a} - a(r_i) \,] = s$$

$$- \sum_{i=1}^{N} c_i a(r_i) + \bar{a}\sum_{i=1}^{N} c_i = s$$

$$\sum_{i=1}^{N} c_i a(r_i) = \bar{a}\sum_{i=1}^{N} c_i - s = E(S) - s \qquad (1.3.2.5)$$

$$S = E(S) - s$$

de (1.3.2.4) e (1.3.2.5) temos que os eventos $T = E(T) + s$ e $S = E(S) - s$ são equivalentes, isto é :

$$P\,[\, S = E(S) + s \,] = P\,[\, S = E(S) - s \,].$$

Com a finalidade de estudar a distribuição da estatística linear de postos S, apresentaremos algumas suposições:

1.3.2 - Suposições Importantes

A1) A constante de regressão satisfaz a condição de "Noether":

$$\lim_{N \to \infty} \frac{\max\limits_{1 \le i \le N} (c_i - \bar{c})^2}{\sum\limits_{i=1}^{N} (c_i - \bar{c})^2} = 0$$

onde $\bar{c} = \dfrac{1}{N} \sum\limits_{i=1}^{N} c_i$

A2) Se os escores $a(R_i)$ podem ser escritos na forma :

$$a(R_i) = u + v\,\varphi\,(R_i /(N+1)) \qquad 1 \le i \le N$$

onde $u \ge 0$, $v \ne 0$ são constantes arbitrárias e φ uma função, então, dizemos que os escores $a(R_i)$ sao gerados pela função φ .

A3) O escore $a(R_i)$ satisfaz:

$$\frac{1}{N} \sum_{i=1}^{N} [\, a\, (R_i) - E\, (\, \varphi\, (\, U_{(i)}))]^2 \to 0$$

onde $U_{(i)}$ denota a i-ésima estatística de ordem de uma amostra de tamanho N proveniente de uma distribuição Uniforme em (0,1).

1.3.3 - Normalidade Assintótica da Estatística Linear de Postos S

Nesta seção estudaremos a distribuição assintótica da estatística linear de posto S, definida na seção (1.3).

Seja $X_1, X_2, \ldots, X_N$ uma amostra aleatória de uma distribuição arbitrária, possivelmente discreta. Então, o conjunto de estatísticas de ordem $X_{(i)}$ decompõe-se dentro de **g** grupos:

$$X_{(1)} = \ldots = X_{(\tau_1)} < X_{(\tau_1+1)} = \ldots = X_{(\tau_1+\tau_2)} < \ldots <$$
$$< X_{(\tau_1+\tau_2+\ldots+\tau_{g-1}+1)} = \ldots = X_{(N)}$$

A ocorrência é caracterizada pelo vetor $\tau = (\tau_1, \tau_2, \ldots, \tau_g)$ onde tanto g como τ_i's são aleatórios.

Seja $$a(i,\tau) = \frac{1}{\tau_k} \sum_{j=T_{k-1}+1}^{T_k} a(j) \qquad (1.3.3.1)$$

com $T_{k-1} = \tau_1 + \tau_2 + \ldots + \tau_{k-1} < i \leq \tau_1 + \tau_2 + \ldots + \tau_k = T_k$

então a estatística linear de postos S, é definida como:

$$S = \sum_{i=1}^{N} c_i a(R_i, \tau) \qquad (1.3.3.2)$$

Se a distribuição de X_i's não é contínua, então, não é verdade que $R=(R_1,R_2, \ldots,R_N)$ é uma permutação de $(1,2,\ldots,N)$ com probabilidade 1.

De qualquer modo, a distribuição condicional de S dado τ é a mesma como se R fosse uniformemente distribuida sobre o espaço de permutações em virtude da especial forma de $a(R_i)$. (Hájek and Šidák : 1967).

A distribuição assintótica da estatística linear de postos S pode tomar três formas diferentes:

Forma 1

A estatística S é assintoticamente normal com parâmetros (E(S), VAR(S)) onde VAR(S) é independente de τ, mas depende da função de distribuição F dos X_i's. Se F não é absolutamente contínua . (Vorkičková : 1970).

Forma 2

A distribuição condicional de S dado τ difere da distribuição normal com parametros $(E(S),VAR(S|\tau))$ na distância de Kolmogorov , para $\varepsilon > 0$ pequeno, com probabilidade maior ou igual a $1-\varepsilon$ se $N \geq N(\varepsilon)$. (Vorkičková :1970).

Forma 3

Estabelece-se a região W_N tal que a distância de Kolmogorov entre o limite $L(S|\tau)$ e $N(E(S),VAR(S|\tau))$ é pequena se ε e $\tau \in W_N$ e $N \geq N(\varepsilon, \{W_N\})$

Teorema 1.3.3 (caso sem empates)

Se R é uniformemente distribuida em $\mathbb{R}$ e S pode ser escrito como:

$$S = \sum_{i=1}^{N} c_i [\ u + v\ \varphi\ (R_i/(N+1))]\ ,$$

onde $u \geq 0$, $v \neq 0$ são constantes (1.3.3.4)

se φ é uma função não-decrescente no intervalo (0,1) ou não-crescente em (0,a) e não-decrescente em (a,1) para algum $a \in (0,1)$ e que satisfaz:

$$0 < \int_0^1 (\ \varphi(t) - \overline{\varphi}\ )^2\ dt < \infty \tag{1.3.3.5}$$

onde:

$$\overline{\varphi} = \int_0^1 \varphi(t)\ dt$$

então, $\forall\ \varepsilon > 0\ ,\ \exists\ \delta = \delta(\varepsilon,\varphi)$. tal que

$$\max_{1 \leq i \leq N} \{\ (c_i - \overline{c}\ )^2\ \} < \delta \sum_{i=1}^{N} (\ c_i - \overline{c})^2 \tag{1.3.3.6}$$

e temos que :

$$\sup_{-\infty<\Delta<\infty} \left| P [S < \Delta] - \phi\left(\frac{\Delta - E(S)}{\sqrt{VAR(S)}}\right) \right| < \varepsilon \qquad (1.3.3.7)$$

onde ϕ é a distribuicão normal padronizada.

Isto é, a estatística S é assintótica normal com paramêtros E(S) e VAR(S).

PROVA (Hájek and Šidák : 1967)

Teorema 1.3.3B (Caso de empates)

Seja $a(R_i)$ que satisfaz (1.3.2.4), onde φ não é constante e é integravel.

Seja $X_1, X_2, \ldots, X_N$ a amostra aleatória independente e identicamente distribuida, então, para $\varepsilon > 0$ e $\nu > 0$ existe um $\delta = \delta(\varepsilon,\nu)$ tal que :

$$\max_{1\leq i\leq N} (c_i - \overline{c})^2 < \delta \sum_{i=1}^{N} (c_i - \overline{c})^2$$

com $\displaystyle\sum_{i=1}^{N} (a(R_i) - \overline{a})^2 > \nu \sum_{i=1}^{N} (a(R_i, \tau) - \overline{a})^2$

implica

$$\sup_{-\infty<\Delta<\infty} \left| P (S \leq E(S) + \Delta (VAR(S|\tau))^{1/2} \mid \tau) - \phi(\Delta) \right| < \varepsilon$$

Prova (Hajék : 1970)

Como vamos trabalhar com várias amostras, precisaremos da estatística Q , cujo resultado apresentaremos no seguinte teorema:

TEOREMA 1.3.4 (Distribuição da Estatística Q)

Suponhamos que uma amostra de tamanho $N = \sum_{j=1}^{k} n_j$ é retirada de uma população, onde n_j =1,2,..,k.

Seja $S_j = \sum_{j=1}^{N} c_{ij} \; a(R_i,\tau)$, a estatística linear de postos da j-ésima amostra,

onde $c_{ij} = \begin{cases} 1, & \text{se } X_i \text{ pertence a j-ésima amostra} \\ 0, & \text{c.c.} \end{cases}$

Sobre a condição do teorema (1.3.3B), a estatística Q dada por:

$$Q = (N-1)\left[\sum_{i=1}^{N} (a(R_i,\tau) - \overline{a})^2\right]^{-1} \left[\sum_{j=1}^{k} \left(\frac{S_j^2}{n_j}\right) - N\,\overline{a}^{\,2}\right] \qquad (1.3.3.8)$$

então, $\forall\, \varepsilon > 0$ e $\varphi > 0$ existe $\delta = \delta(\varepsilon,\varphi)$ tal que

$$\sum_{i=1}^{N} (a(R_i) - \overline{a})^2 > \varphi \sum_{i=1}^{N} (a(R_i, \tau)) - \overline{a})^2$$

e $\mathrm{Min}(n_1, n_2, \ldots, n_k) > \delta$

temos que:

$$\sup_{-\infty < q < \infty} \left| P[\, Q \le q \mid \tau \,] - P[\, \chi^2_{k-1} \le q \,] \right| < \varepsilon$$

onde χ^2_{k-1} possui distribuição Qui-Quadrado com k-1 graus de liberdade.

Prova : (Hájek:1970)

Como temos observado nos teoremas anteriores, as restrições são feitas principalmente sobre as funções escores, assim vamos selecionar estas funções que satisfaçam as suposições dos teoremas e que também possamos obter testes localmente mais poderosos.

1.4 - TESTES LOCALMENTE MAIS PODEROSOS DE TAMANHO α

Apresentaremos o teste localmente mais podroso; detamanho α, somente para o caso de duas amostras, pois, para casos de amostras de tamanho maior que dois (mais de duas amostras), os testes tem o mesmo comportamento.

1.4.1 - Testes de Locação para Duas Amostras

Sejam $x_1, x_2, \ldots, x_m$ e $y_1, y_2, \ldots, y_n$ duas amostras aleatórias de distribuições absolutamente contínuas, que denominaremos como $G(x)$ e $H(y)$, com densidade $g(x)$ e $h(y)$ respectivamente, e seja $R_{(i)}$ o posto de $y_{(i)}$ nas observações combinadas.

TEOREMA 1.4.1 (HOEFFDING : 1951)

Se $h(x) > 0$ implica $g(x) > 0$

então:

$$P\left[R_{(1)} = r_1,\ R_{(2)} = r_2, \ldots, R_{(n)} = r_n \right] = \frac{1}{\binom{m+n}{m}} E\left[\prod_{i=1}^{n} \frac{h(V_{(r_i)})}{g(V_{(r_i)})} \right] \tag{1.4.1.1}$$

onde $V_{(1)} \leq V_{(2)} \leq \ldots \leq V_{(m+n)}$, são estatísticas de ordem de uma amostra de tamanho $m+n = N$ de G.

PROVA

$$P\left[R_{(1)} = r_1,\ R_{(2)} = r_2, \ldots, R_{(n)} = r_n \mid y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n)} \right] =$$

tem distribuição multinomial

$$= \frac{m!}{\prod_{j=1}^{n} (r_{j+1} - r_j - 1)!} G(y_{(1)})^{r_1 - 1} [G(y_{(2)}) - G(y_{(1)})]^{r_2 - r_1 - 1} [1 - G(y_{(n)})]^{m+n-r_n}$$

$$P\left[R_{(1)} = r_1,\ R_{(2)} = r_2, \ldots, R_{(n)} = r_n ,\ y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n)} \right]$$

$$= P\left[R_{(1)} = r_1,\ R_{(2)} = r_2, \ldots, R_{(n)} = r_n \right] n! \prod_{i=1}^{n} h(y_{(i)})$$

multiplicando e dividindo por:

$$(m+n)! \prod_{i=1}^{n} g(y_{(i)})$$

temos:

$$P\left[R_{(1)} = r_1,\ R_{(2)} = r_2, \ldots, R_{(n)} = r_n\right] =$$

$$= \int \cdots \int \frac{m!\,n!}{(m+n)!} \; \frac{\prod_{j=1}^{n} h(y_{(j)})}{\prod_{j=1}^{n} g(y_{(j)})} \left\{ \frac{(m+n)!}{m!\,n!} \; \frac{1}{\prod_{j=1}^{n} (r_{j+1} - r_j - 1)} \quad G(y_{(1)})^{r_1 - 1} \right.$$

$$\left. [G(y_{(2)}) - G(y_{(1)})]^{r_2 - r_1 - 1} [1 - G(y_{(n)})]^{m+n-r} \prod_{j=1}^{n} g(y_{(j)}) \right\} dy_{(1)} \ldots dy_{(n)}$$

então:

$$P\left[R_{(1)} = r_1,\ R_{(2)} = r_2, \ldots, R_{(n)} = r_n\right] = \frac{1}{\binom{m+n}{m}} \quad E\left[\prod_{i=1}^{n} \frac{h(y_{(i)})}{g(y_{(i)})}\right]$$

No caso do modelo de locação com :

$G(x) = F(x)$ e $H(y) = F(x - \Delta)$,

onde:

$F \in \mathcal{F} = \{ F : F$ é absolutamente contínua e $F(0) = 1/2$, única $\}$

$F(y) = F(x-\Delta)$ e $G(x) = F(x)$, significa que entre X e Y existe diferença só de locação.

Sob $H_0 : \Delta = 0$ a distribuição de $R_{(1)} \leq R_{(2)} \leq \ldots \leq R_{(n)}$ é Uniforme, istó é,

$$P\left[R_{(1)} = r_1,\ R_{(2)} = r_2, \ldots, R_{(n)} = r_n\right] = \frac{1}{\binom{m+n}{m}} \quad E\left[\prod_{i=1}^{n} \frac{h(y_{(i)})}{g(y_{(i)})}\right]$$

$$= \frac{1}{\binom{m+n}{m}} \qquad (1.4.1.2)$$

se k é um inteiro tal que:

$$\frac{k}{\binom{m+n}{m}} = \alpha \qquad (1.4.1.3)$$

então, qualquer conjunto C de k vetores de postos $(r_1, r_2, \ldots, r_n)$ é uma região crítica de tamanho α para testar H_0: $\Delta = 0$, onde, determinaremos a melhor região crítica sobre algum critério.

A potência da região crítica C de tamanho α, é dada por

$$\beta(\Delta) = \sum_{(r_1, r_2, \ldots, r_n) \in C} \frac{1}{\binom{m+n}{m}} \; E\left[\prod_{i=1}^{n} \frac{f(V_{(r_i)} - \Delta)}{f(V_{(r_i)})} \right] \qquad (1.4.1.4)$$

onde $V_{(1)} \leq V_{(2)} \leq \ldots \leq V_{(m+n)}$ são estatísticas de ordem de uma amostra de tamanho m+n de $F(x)$.

<u>Definição 1.4.1</u>

O teste localmente mais poderoso de tamanho α é dado pela região crítica C' de tamanho α, tal que $\beta'(0)$ é máximo , onde $\beta'(0)$ é a derivada de $\beta(\Delta)$ avaliada em $\Delta=0$.

TEOREMA 1.4.2

Dado $F \in \mathcal{F}$, e supondo que a derivada sobre a esperança exista, então o teste localmente mais poderoso de tamahno α que rejeita $H_0 : \Delta = 0$ em favor de $H_A : \Delta > 0$, é dado por:

$$V = -\sum_{i=1}^{n} E\left[\frac{f'(V_{(r_i)})}{f(V_{(r_i)})} \right] \geq c \qquad (1.4.1.5)$$

onde c é determinado por $P\ [V \geq c\ ] = \alpha$ sob H_0.

Prova

Para encontrar a estatística que maximiza o poder;

$$\beta'(\Delta) = \frac{\partial}{\partial \Delta} \beta(\Delta)$$

$$= \sum_{(r_1, r_2, \ldots, r_n) \in C} \frac{1}{\binom{m+n}{m}} \quad E\left[\frac{-[f'(V_{(r_i)} - \Delta) \prod_{j \neq i}^{n} f\ (V_{(r_j)} - \Delta\ )]}{\prod_{i=1}^{n} f(V_{(r_i)})} \right]$$

$$\beta'(0) = \sum_{(r_1, r_2, \ldots, r_n) \in C} \frac{1}{\binom{m+n}{m}} \left(-\sum_{i=1}^{n} E\left[\frac{f'(V_{(r_i)})}{f(V_{(r_i)})} \right] \right)$$

onde C é qualquer conjunto de k vetores de postos $(r_1, r_2, \ldots, r_n)$ tal que:

$$\frac{k}{\binom{m+n}{m}} = \alpha$$

Para maximizar $\beta'(0)$ podemos construir C com aqueles vetores de postos $(r_1, r_2, \ldots, r_n)$ que rendem os k maiores valores de

$$V = -\sum_{i=1}^{n} E\left[\frac{f'(V_{(r_i)})}{f(V_{(r_i)})}\right] \qquad (1.4.1.6)$$

Consequentemente podemos encontrar uma constante c tal que $V \geq c$, dá a região crítica de tamanho α que maximiza $\beta'(0)$ por definição. Então, V é a estatística que oferece o teste localmente mais poderoso.

como $S = \sum_{i=1}^{N} c_i a(R_i)$, com $c_1 = c_2 = \ldots = c_n = 1$

e $c_{n+1} = c_{n+2} = \ldots = c_{n+m} = 0$

temos:

$$S = \sum_{i=1}^{n} a(R_i)$$

onde R_i é o posto de y_i na amostra combinada.

Os escores $a(R_i)$ são gerados:

$$a(R_i) = -E\left[\frac{f'(V_{(r_i)})}{f(V_{(r_i)})}\right] \qquad (1.4.1.7)$$

sendo $V_{(1)} \leq V_{(2)} \leq \ldots \leq V_{(m+n)}$ estatísticas de ordem de $F(x)$.

Então, o teste localmente mais poderoso é obtido pela estatística:

$$S = V = -\sum_{i=1}^{n} E\left[\frac{f'(V_{(r_i)})}{f(V_{(r_i)})}\right] = \sum_{i=1}^{n} a(R_i) \qquad (1.4.1.8)$$

O resultado concernente à distribuição univariada da estatística linear de posto pode ser extendido para o caso multivariado, que se apresentará a seguir. (Puri and Sen : 1969).

1.5 - DISTRIBUIÇÃO ASSINTÓTICA DA ESTATÍSTICA LINEAR DE POSTOS MULTIVARIADA.-

Seja $X_{\alpha} = (X_{1\alpha}, X_{2\alpha}, \ldots, X_{p\alpha})$ $1 \leq \alpha \leq N_v$, $v \geq 1$ um conjunto de vetores aleatórios estocasticamente independentes com função de distribuição $F(x) \in \mathcal{F}$, e $x \in \mathcal{R}^p$, espaço real p-dimensional .

Seja $R_{vj\alpha}$ o posto de $X_{j\alpha}$ no conjunto de $\{X_{j1}, X_{j2}, \ldots, X_{jN_v}\}$

para $1 \leq j \leq p$, $1 \leq \alpha \leq N_v$, $v \geq 1$

Definimos a estatística linear multivariada de postos da seguinte maneira:

$$S_{vj} = \sum_{\alpha=1}^{N_v} c_{v\alpha}\ a_{vj}(R_{vj\alpha})$$

onde:

$c_{v\alpha}$, $1 \leq \alpha \leq N_v$, são constantes de regressão e satisfaz a condição de "Noether" (suposição A1)

e os escores $a_j(\alpha)$, $1 \leq j \leq p$, $1 \leq \alpha \leq N_v$ converge quando $v \to \infty$, para cada j, em média quadrática a uma função integrável não constante $\varphi(u)$, $0 < u < 1$.

A distribuição assintótica condicional conjunta e a incondicional desta estatística , são de notável interesse para o teste não-paramétrico multivariado.

1.5.1 - Suposições Básicas

Hájek and Šidák (1967), aproveita a distribuição assintótica da estatística para S_{vj}, $1 \leq j \leq p$.

Seja f_{0v} a família de todas as N_v funções densidades p-dimensional tal que :

$$H_0 : h_v(x_1, x_2, \ldots, x_{N_v}) = \prod_{\alpha=1}^{N_v} f(x_\alpha) \quad , \quad v \geq 1 \qquad (1.5.1.1)$$

onde $x_\alpha \in \mathcal{R}^p$ e $f(.)$ é uma função densidade p-dimensional.

Se $h_v \in f_{0v}$, então, as observações dos vetores X_α, $1 \leq \alpha \leq N_v$ são independentes e identicamente distribuídas de acordo com a mesma função densidade f(.) p-dimensional.

Seja $\underset{\sim}{X}'_v = (X_1 \; , X_2 \; , \ldots , X_{N_v})$, uma seqüência de variáveis aleatórias p-dimendional, com função densidade $h_v \in f_{0v}$.

Denotaremos $R_{vj\alpha}$, o posto de $X_{j\alpha}$ no conjunto de $\{X_{j1}, X_{j2}, \ldots, X_{jN_V}\}$; $1 \leq j \leq p$, e seja o vetor real $(c_{v1}, c_{v2}, \ldots, c_v)$
tal que:

$$\sum_{\alpha=1}^{N_v} (c_{v\alpha} - \overline{c}_v)^2 > 0 \qquad (1.5.1.2)$$

onde :

$$\overline{c}_v = \frac{1}{N_v} \sum_{\alpha=1}^{N_v} c_{v\alpha}$$

supondo a condição de "Noether" (A1)

$$\lim_{v \to \infty} \left\{ \frac{\sum_{\alpha=1}^{N_v} (c_\alpha - \overline{c}_v)^2}{\max\limits_{1 \leq \alpha \leq N_v} (c_\alpha - \overline{c}_v)^2} \right\} = \infty \quad , \qquad (1.5.1.3)$$

é satisfeita. Reescrevendo a estatística linear de postos S_{vj} da seguinte maneira:

$$S_{vj} = \sum_{\alpha=1}^{N_v} (c_{v\alpha} - \bar{c}_v)\ a_j(R_{vj\alpha}) \ , \quad 1 \le j \le p \quad \text{(centrada)} \quad (1.5.1.4)$$

assumindo que existe alguma função integrável quadrática $\varphi_j(u)$, $1 \le j \le p$, tal que

$$\sigma^2(\varphi_j) = \int_0^1 \{\ \varphi_j(u) - \bar{\varphi}_j\ \}^2\ du \ > 0$$

$$\bar{\varphi}_j = \int_0^1 \varphi_j(u)\ du \qquad (1.5.1.5)$$

e

$$\lim_{v\to\infty} \int_0^1 \{\ a_j\ (1 + [\ un_v]) - \varphi_j(u)\ \}^2\ du \ = 0 \qquad (1.5.1.6)$$

notando que [t] é a função maior inteiro de t. (Hájek : 1967). Estas suposições (1.5.1.1) a (1.5.1.6) serão utilizadas para estudar a normalidade assintótica da estatística S_{vj}.

1.5.2 - Normalidade Assintótica

Sejam $x_\alpha \in \mathcal{R}^p$, $1 \le \alpha \le N_v$ vetores. Então, para qualquer seqüência $x_v = (x_1\ , x_2\ , \ldots, x_{N_v})$ de vetores, corresponde uma outra seqüência de vetores $x_v^{(\cdot)} = (x^{(1)}, x^{(2)}, \ldots, x^{(N_v)})$ constituido do mesmo vetor, mas reordenado de forma crescente. Denotaremos esta correspondência por $X^{(\cdot)}(x_v)$.

Se X_v é regida pela função densidade $h_v \in f_{0v}$ definida na seção (1.5.1), então :

$$P_v [X_v = x_v \mid X_v^{(.)} = x_v^{(.)}] = \frac{1}{N_v!} \quad , \text{ se } \quad X^{(.)}(x_v) = x_v^{(.)} \in \mathbb{X}_v^{(.)},$$

onde $\mathbb{X}_v^{(.)}$ denota o sub-espaço de $\mathbb{X}_v = \mathcal{R}^{N_v p}$, espaço real $N_v p$-dimensional, contendo pontos $x_v^{(.)}$.

Podemos agora verificar a variância condicional

$$VAR(S_{vj} \mid X_v^{(.)} = x_v^{(.)}) = \frac{1}{N_v - 1} \sum_{\alpha=1}^{N_v} (c_{v\alpha} - \bar{c}_v)^2 \sum_{\alpha=1}^{N_v} (a_{vj}(R_{vj\alpha}) - \bar{a}_{vj})^2$$

onde $\bar{a}_{vj}$ é o valor medio dos escores de uma variavél j, $j=1,2,...,p$.

E a covariânçia condicional,

$$COV(S_{vi}, S_{vj} \mid X_v^{(.)} = x_v^{(.)}) =$$

$$= \frac{1}{N_v - 1} \sum_{\alpha=1}^{N_v} (c_{v\alpha} - \bar{c}_v)^2 \sum_{\alpha=1}^{N_v} (a_{vi}(R_{vi\alpha}) - \bar{a}_{vi})(a_{vj}(R_{vj\alpha}) - \bar{a}_{vj})$$

$$\text{onde } S_{vj} = \sum_{\alpha=1}^{N_v} (c_{v\alpha} - \bar{c}_v)\, a_j\, (R_{vj\alpha}) \qquad 1 \le j \le p$$

Denotaremos a correlação condicional de S_{vi} e S_{vj} por γ_{vij} e definidemos a matriz de correlação condicional

$$\underset{\sim}{\gamma}_v = ((\gamma_{vij})) \qquad (1.5.2.1)$$

sendo

$$\sigma(\varphi_i,\varphi_j) = \int_0^1 \int_0^1 (\varphi_i(u) - \overline{\varphi}_i)(\varphi_j(s) - \overline{\varphi}_j) \; dP \; [\; U_i \leq u \;,\; U_j \leq s \;]$$
$$1 \leq i,j \leq p, \qquad (1.5.2.2)$$

onde $U_j = F_j(X_j)$, sendo $F_j(.)$ a função distribuição marginal de X_j, definamos a matriz de correlação

$$\underset{\sim}{\gamma} = ((\gamma_{ij})) .$$

onde :

$$\gamma_{ij} = \frac{\sigma(\varphi_i,\varphi_j)}{\sigma(\varphi_i)\sigma(\varphi_j)} \qquad (1.5.2.3)$$

<u>LEMA 1.5.1</u>

Seja $X_v = (X_1, X_2, \ldots, X_{N_v})$ variáveis aleatórias regidas pela função densidade $h_v \in f_{0v}$ dada por

$$h_v(x_1,x_2,\ldots,x_{N_v}) = \prod_{\alpha=1}^{N_v} f(x_\alpha) \qquad , \; v \geq 1$$

e supondo (1.5.1.6) , temos que $\underset{\sim}{\gamma}_v$ converge em probabilidade a $\underset{\sim}{\gamma}$ quando $v \to \infty$

PROVA (Hájek and Šidák: 1973).

Denotaremos a distribuição condicional conjunta de Z_{vj} , $1 \leq j \leq p$ por $G_v(\Delta_1,\Delta_2,\ldots,\Delta_p \mid X_v^{(\cdot)})$,
onde:

$$Z_{vj} = \frac{S_{vj}}{VAR(S_{vj}|X_v^{(\cdot)})^{1/2}}$$

a distribuição condicional de $G_v(. \mid X_v^{(\cdot)})$ é uma função de $X_V^{(\cdot)}$ com valores $G_v(.|x_v^{(\cdot)})$, quando $X_v^{(\cdot)} = x_v^{(\cdot)}$

TEOREMA 1.5.2

Seja $X_v=(X_1,X_2,\ldots,X_{N_v})$ um vetor aleatório com função densidade de probabilidade $h_v \in f_{0v}$, e supondo que a matriz $\underset{\sim}{\gamma} = ((\gamma_{ij}))$ é positiva definida e (1.5.1.6) é verdadeira, então, para qualquer $\eta > 0$, existe K_η tal que

$$\sum_{\alpha=1}^{N_v} (c_{v\alpha} - \overline{c}_v)^2 > K_\eta \max_{1\leq\alpha\leq N_v} (c_{v\alpha} - \overline{c}_v)^2, \qquad (1.5.2.4)$$

implica que:

$$P[\sup_{\Delta_1,\ldots,\Delta_p} |G_v(\Delta_1,\ldots,\Delta_p \mid X_v^{(\cdot)}) - \phi(\Delta_1,\ldots,\Delta_p| \underset{\sim}{0}, \underset{\sim}{\gamma})| < \eta] > 1-\eta \qquad (1.5.2.5)$$

onde $\phi\,(.|\underset{\sim}{0},\ \underset{\sim}{\gamma})$ denota a função distribuição normal p-variável com vetor de media $\underset{\sim}{0}$ e matriz de correlação $\underset{\sim}{\gamma}$.

PROVA (Hájek and Šidák : 1973).

O seguinte teorema é dado para estudar a distribuição normal conjunta incondicional.

TEOREMA 1.5.3

Seja S_{vj} , $1\leq j \leq p$ definida em (1.5.1.4); então sob a suposição do teorema 1.5.2 , temos que para algum $\eta > 0$, existe K_η tal que se cumpre (1.5.2.4) , implica

$$\sup_{\Delta_1,\ldots,\Delta_p} \left| P\left[S_{v1} \leq \Delta_1 \sigma_{v1}, \ldots, S_{vp} \leq \Delta_p \sigma_{vp} \right] - \phi(\Delta_1,\ldots,\Delta_p \mid \underset{\sim}{0},\gamma) \right| < \eta$$

onde:

$$\sigma^2_{vj} = \sum_{\alpha=1}^{N_v} (c_{v\alpha} - \overline{c}_v)^2\, \sigma^2(\varphi_j) , \qquad 1\leq j \leq p$$

PROVA (Hájek and Šidák :1973).

Como a função escore deve satisfazer certas condições, para que a distribuição da estatística linear de postos S, tenha distribuição assintótica Normal, e também obter testes localmente mais podero, apresentaremos algumas regras de seleção para obter tais escores.

1.6 - OBTENÇÃO DE ESCORES PARA ALGUMAS DISTRIBUIÇÕES ADJACENTES

1.6.1 .- Testes De Locação.

Se é conhecida a função densidade adjacente, determinaremos qual é o melhor escore que nos dá testes localmente mais poderosos de tamanho α.

Segundo o teorema (1.4.2) os escores são gerados por:

$$a(R_i) = - E\left[\frac{f'(V_{(r_i)})}{f(V_{(r_i)})}\right] \qquad (1.6.1.1)$$

onde $V_{(1)} \leq V_{(2)} \leq . . . \leq V_{(m+n)}$ são estatísticas de ordem de $F(x)$ e $m+n = N$.

Na continuação apresentaremos alguns resultados :

a) <u>Distribuição Normal</u>

Se $f(x) = \dfrac{1}{(2\pi)^{1/2}\sigma} \exp - \left\{ \dfrac{(x - \mu)^2}{2\sigma^2} \right\}$ $\qquad - \infty < x < \infty$

temos que, $\dfrac{f'(x)}{f(x)} = -x$

desenvolvendo (1.4.1.7) temos :

$$a(R_i) = - E \left[\frac{f'(V_{(r_i)})}{f(V_{(r_i)})} \right] = E(V_{(r_i)})$$

onde $V_{(r_i)}$ são as N estatísticas de ordem da normal estandardizada N(0,1), assim, os escores da distribuição são dados por:

$$a(R_i) = \phi^{-1}\left(\frac{R_i}{N+1} \right) , \qquad (1.6.1.2)$$

conhecido como e escore de VAN DER WAERDEN .

Se a distribuição adjacente das observações é a distribuição Normal, então a estatística que oferece o teste localmente mais poderoso de tamanho α para o caso de duas amostras independentes é:

$$S = \sum_{i=1}^{n} \phi^{-1}\left(\frac{R_i}{N+1} \right)$$

b) Distribuição Logística

$$f(x) = \frac{\exp\{-x\}}{(1 + \exp\{-x\})^2} \qquad -\infty < x < \infty$$

então :

$$F(x) = \frac{1}{(1 + e^{-x})}$$

e $\quad f(x) = F(x)(1-F(x))$

ademais: $-\dfrac{f'(x)}{f(x)} = 2F(x) - 1$

desenvolvendo (1.4.1.7) temos :

$$a(R_i) = -E\left[\frac{f'(V_{(r_i)})}{f(V_{(r_i)})}\right] = E\{2F(V_{(r_i)}) - 1\}$$

como: $E(V_{(r_i)}) = \dfrac{R_i}{N+1}$

então: $\quad a(R_i) = \dfrac{2R_i}{N+1} - 1$,

conhecida como o escore de MAM WHITHER.

Simplificando, já que a função escore é invariante sob transformação linear, temos:

$$a(R_i) = \frac{R_i}{N+1} \quad \text{ou} \quad a(R_i) = R_i \qquad (1.6.1.3)$$

é conhecida como o escore de WILCOXON.

Se a distribuição adjacente das observações é a distribuição Logística, então a estatística que oferece o teste localmente mais Poderoso de tamanho α para o caso de duas amostras independentes é:

$$S = \sum_{i=1}^{n} R_i \,/N+1 \qquad \text{ou} \qquad S = \sum_{i=1}^{n} R_i$$

c) Distribuição Dupla Exponencial

$$f(x) = \frac{\exp\{|x|\}}{2}, \qquad -\infty < x < \infty$$

Temos : $\dfrac{f'(x)}{f(x)} = \text{Sinal}(x)$, para todo $x \neq 0$

desenvolvendo (1.4.1.7) temos :

$$a(R_i) = -E\left[\frac{f'(V_{(r_i)})}{f(V_{(r_i)})}\right] = E(\text{sinal}(V_{(r_i)}))$$

observemos que $x > 0$, se $(2F(x) - 1) > 0$, assim,

$$a(R_i) = E(\text{ sinal } (V_{(r_i)}))$$
$$= E(\text{sinal } (2\ F(V_{(r_i)}) - 1))$$
$$= \text{sinal}(2E (F(V_{(r_i)}) -1))$$
$$= \text{sinal } (2 \frac{R_i}{N+1} - 1)$$
$$a(R_i) = \text{sinal } (R_i - \frac{N+1}{2})$$

então:

$$a(R_i) = \begin{cases} 1\ , & \text{se } R_i > (N+1)/2 \\ 0\ , & \text{se } R_i \leq (N+1)/2 \end{cases} \qquad (1.6.1.4)$$

é conhecido como o escore da MEDIANA.

d) <u>Distribuição Uniforme</u>

$$f(x) = \frac{1}{b - a} \qquad , \qquad a < x < b$$

de GASTWIRTH (1966), temos:

$$a(R_i) = a^*(R_i) \ / \ N+1 \qquad (1.6.1.5)$$

$$\text{onde: } a^*(R_i) = \begin{cases} R_i - g - 1/2 & , \text{ se } R_i \leq g \\ R_i - N + g - 1/2 & , \text{ se } R_i \geq N-g+1 \\ 0 & , \text{ c.c.} \end{cases}$$

com $g = [(N + 3) / 4]$ e $[\ . \ ]$ é a função valor inteiro.

é conhecido como o escore de GASTWIRTH .

1.6.2 -Regra de Decisão para Seleção de Escores em Testes de Locação quando a Função Densidade Adjacente Não é Conhecida

Na busca de obter estimadores mais eficientes e testes localmente mais poderosos, a seleção de escores deve ser a mais adequada, porque a seleção errada dos escores pode acarretar erros maiores que aqueles que se cometeriam utilizando testes paramétricos, quando os erros não tem distribuição normal.

Para selecionar os escores de cada teste devemos estudar simetria e longitude das caudas do conjunto de dados amostrais. Com o fim de analisar a simetria do conjunto de dados amostrais definiremos a função seletora φ_2 :

$$\varphi_2 = \frac{\overline{U}(0,05) - m(0,25)}{m(0,25) - \overline{L}(0,05)} \qquad (1.6.2.1)$$

onde:

$\overline{U}(\eta)$: é a média das $[N\eta + 1]$ estatísticas de ordem superior .

$\overline{L}(\eta)$: é a média das $[N\eta + 1]$ estatísticas de ordem inferior.

$m(\eta) = \dfrac{1}{h} \sum_{i=[N\eta]+1}^{N-[N\eta]} X_{(i)}$, estatísticas de ordem centraes.

$h = N - 2[N\eta]$,

e $[\ t\ ]$ é o valor inteiro de t .

No qual temos a seguinte regra de decisão:

Se $\varphi_2 < 1/2$ a distribuição das observações é assimétrica à esquerda.

Se $1/2 \leq \varphi_2 \leq 2$ a distribuição das observações é simétrica

Se $\varphi_2 > 2$ a distribuição das observações é assimétrica à direita.

Para estudar a longitude das caudas do conjunto de dados amostrais, consideremos o caso em que a distribuição seja simétrica ($1/2 \leq \varphi_2 \leq 2$), definindo o seguinte indicador desenvolvido por NEGRILLO (1989).

$$\varphi_1 = \frac{\overline{U}(0,05) - \overline{L}(0,05)}{\overline{U}(0,5) - \overline{L}(0,5)} \qquad (1.6.2.2)$$

Para qual temos a seguinte regra de decisão:

φ_1	Distribuição	Escore de Locação
$\varphi_1 < 2$	Uniforme Caracteristica: cauda curta	Escore de Gastwirth (1.6.1.5)
$2 \le \varphi_1 \le 2,92$	Normal Caracteristica: cauda média -	Escore de Van der Waerden (1.6.1.2)
$2,92 < \varphi_1 \le 3,8$	Logística Caracteristica: cauda média +	Escore de Wilcoxon (1.6.1.3)
$\varphi_1 > 3,8$	Exponencial Dupla Caracteristica: cauda comprida	Escore da Mediana (1.6.1.4)

No caso de que a distribuição seja assimétrica temos:

φ_2	Distribuição	Escore de Locação
$\varphi_2 < 1/2$	Assimétrica a Esquerda	$a(R_i) = \begin{cases} R_i/(N+1) & , \text{ se } R_i > \frac{N+1}{2} \\ 1/2 & , \text{ c.c.} \end{cases}$
$\varphi_2 > 2$	Assimétrica a Direita	$a(R_i) = \begin{cases} R_i/(N+1) & , \text{ se } R_i \le \frac{N+1}{2} \\ 1/2 + 1/(N+1), & \text{ c.c.} \end{cases}$

No caso de mais de duas amostras: para a i-ésima amostra determinamos φ_1 e φ_2 denotandos por $\varphi_{1.i}$ e $\varphi_{2.i}$ respectivamente, para $i=1,2,\ldots,c$ e calculamos:

$$\overline{\varphi}_1 = \sum_{i=1}^{c} n_i \, \varphi_{1.i} \, / \, N \qquad \text{e} \qquad \overline{\varphi}_2 = \sum_{i=1}^{c} n_i \, \varphi_{2.i} \, / \, N$$

onde :

$$N = \sum_{i=1}^{c} n_i$$

sendo n_i o tamanho da i-ésima amostra , $i=1,2$, ...,c.
e aplicaremos as regras mencionadas anteriormente.

1.7 - EFICIÊNCIA

1.7.1 - Eficiência de Pitman.

É um criterio objetivo que pode ser utilizado para eleger, entre dois ou mais testes, que são comparáveis de maneira bem definida. Na teoria de estimação puntual, a eficiência de dois estimadores não-viciados para um determinado parâmetro, é definida como a razão de suas variâncias. Em algumas situações o valor limite desta razão pode interpretar-se como o número relativo de observações adicionais, que é necessario para obter-se a mesma eficiência que o estimador de menor eficiência. A idéia de eficiência entre duas estatísticas de prova está fortemente ligada a definição anterior, onde a potência é considerada como uma medida equivalente, e tem muitas variáveis na hipótese do teste.

A maneira mais comum de comparar dois testes é fazer equivalência a todos os fatores, excetuando-se o tamanho amostral.

1.7.2 - Eficiência Relativa Assintótica.(ARE)

Nas seções prévias fez-se um rigoroso desenvolvimento da distribuição assintótica da estatística linear de postos; nesta seção discutiremos a eficiência relativa assintótica de testes que utilizam estatística linear de postos.

A família de testes utilizando estatítica linear de postos e testes de razão de verossimilhança, possuem algumas propriedades em comum, sob a hipótese nula, tal como, a distribuição assintótica normal.

Supondo que tenhamos duas estatísticas consistentes denotadas como Q_N e $T_{N'}$, para testar a mesma hipótese nula H_0 versus uma alternativa H_1.

Supomos ainda que a estatística do teste Q_N é baseada na estatística linear de postos definida no teorema (1.3.4), de tamanho N, e $T_{N'}$ é obtida em função da razão verossilihança, de tamanho N'. Então, os dois testes tem o mesmo poder assintótico, se e somente se:

$$\lim_{N, N' \to \infty} \frac{N^{1/2} \sigma^{-1}(\varphi)\ \gamma(u,f)}{(N')^{1/2} \sigma^{-1}} = 1 \qquad (1.7.2.1)$$

onde :

σ^2 é a variância da estatistica $T_{N'}$

$$\sigma^2(\varphi) = \int_0^1 (\varphi(u) - \overline{\varphi})^2 du$$

$$\overline{\varphi} = \int_0^1 \varphi(u)\, du$$

$$\gamma(u,f) = \int_0^1 \varphi(u)\varphi(u,f)\, du$$

$$\varphi(u,f) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}, \qquad u \in (0,1)$$

desenvolvendo (1.7.3.1) temos:

$$\lim_{N,N'\to\infty} (N' / N) = \frac{\sigma^2\, \gamma^2(u,f)}{\sigma^2(\varphi)} \qquad (1.7.2.2)$$

pela definição (1.7.2) a Eficiência Relativa Assintótica de Q_N com respeito a $T_{N'}$ é :

$$\text{ARE}(Q_N, T_N) = \frac{\sigma^2\, \gamma^2(u,f)}{\sigma^2(\varphi)} \qquad (1.7.3.3)$$

Se o erro básico da função distribuição F(.) é completamente conhecido, então, a eficiência relativa assintótica é conseguida quando tomamos:

$$\varphi(u) = \varphi(u,f) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}$$ (Mansouri :1990).

Na tabela (1.7), Piere and Rauch (1984) apresentam a Eficiência Relativa Assintotica de Pitman (ARE) na comparação dos métodos Não-Paramétricos e os métodos Paramétricos, para o modelo fatorial 3x4 com interação.
Usando ARE, para varias distribuições adjacentes temos os seguintes resultados:

TABELA 1.7.- ARE(NP,P) Para Várias Distribuições Adjacentes.

Distribuições	ARE(NP,P)
Normal	0.9549
Normal Contaminada	1.840
Exponencial Dupla	1.50
Exponencial	3.00
Cauchy	∞

onde: NP :representa métodos não-paramétricos, usando o escore de Wilcoxon.

P :representa métodos paramétricos .

Na distribuição normal contaminada foi usado 5% de contaminação com desvio padrão igual a 5.

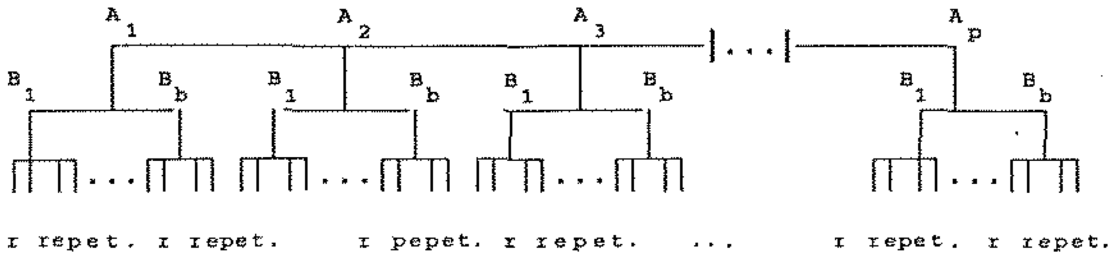
CAPÍTULO II

TESTES NÃO-PARAMÉTRICOS PARA A ANÁLISE DE EXPERIMENTOS HIERÁRQUICOS COM DOIS FATORES FIXOS E BLOCOS ALEATÓRIOS NÃO BALANCEADOS

Neste capítulo consideraremos a distribuição assintótica da estatística linear de postos definida no capítulo I, para a análise de experimentos Hierárquicos com dois fatores fixos e para a análise de experimentos em blocos aleatórios não balanceados. As aplicações serão dadas no capítulo IV.

2.1 - TESTES NÃO-PARAMÉTRICOS PARA EXPERIMENTOS HIERÁRQUICOS COM DOIS FATORES FIXOS

No planejamento de experimentos, muitas vezes se tem um experimento com dois fatores fixos A e B classificados de forma hierárquica, sendo B um sub fator de A, e que é representado no seguinte diagrama:



Exemplo:

Uma oficina tem quatro tipos de máquinas e cinco funcionarios por máquina. Deseja-se estudar se existe diferença entre a produtividade de cada máquina. Também deseja-se saber se os funcionarios em cada máquina têm a mesma produtividade.

Este é um experimento hierárquico onde os fatores são:

A : Máquina (fator fixo), com quatro níveis.

B: Funcionário (fator fixo), com cinco níveis.

O modelo matemático associado a este experimento hierárquico é definido da seguinte maneira:

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + e_{ijk}$$

$i = 1,2,\dots,p$ nível do fator A

$j = 1,2,\dots,b$ nível do fator B

$k = 1,2,\dots,r$ repetições

$N = pbr$.

Sujeito ás seguintes restrições nos parâmetros :

$$\sum_{i=1}^{p} \alpha_i = 0 \ , \qquad e \quad \sum_{j=1}^{b} \beta_{j(i)} = 0 \quad \text{para todo } i.$$

onde :

y_{ijk} : é a resposta associada ao j-ésimo nível do fator B, dentro do i-ésimo nível do fator A, na k-ésima repetição.

μ : é a média geral.

α_i : é o efeito do i-ésimo nível do fator A.

$\beta_{j(i)}$: é o efeito do j-ésimo nível do fator B dentro do i-ésimo nível do fator A.

e_{ijk} : é o erro aleatório.

2.1.1 - Suposições:

Nos métodos não-paramétricos os erros e_{ijk} são variáveis aleatórias independentes e identicamente distribuída com função distribuição absolutamente contínua e $E(e_{ijk}) = 0$.

Assim, temos que y_{ijk} tem função distribuição absolutamente contínua.

2.1.2 - Hipóteses

As hipóteses que vamos testar neste tipo de experimento são:

$H_{01} : \alpha_1 = \alpha_2 = . . . = \alpha_p = 0$ (não existe efeito do fator A).

e

$H_{02} : \beta_{1(i)} = \beta_{2(i)} = . . . = \beta_{b(i)} = 0$ para todo i. (não existe efeito do fator B dentro do i-ésimo nível do fator A para todo i).

2.1.3 - Transformação de Alinhamento

A distribuição da estatística do teste, para um determinado fator depende das hipóteses nulas dos outros fatores. A dependência é eliminada através de transformações de alinhamento , obtidas pela subtração no modelo geral dos efeitos estimados dos outros fatores.

a) Para testar a existência de efeito do fator A

$$x_{ijk} = y_{ijk} - \hat{\beta}_{j(i)}$$

que resulta:

$$x_{ijk} = y_{ijk} - \tilde{y}_{ij.} + \tilde{y}_{i..}$$

b) Para testar a existência de efeito de fator B dentro do i-ésimo nível do fator A

$$x_{ijk} = y_{ijk} - \hat{\alpha}_i$$

que resulta:

$$x_{ijk} = y_{ijk} - \tilde{y}_{i..} + \tilde{y}_{...}$$

onde a notação $\tilde{y}$ pode denotar a média aritmética, mediana ou outra estatística de locação não-paramétrica.

No caso que $\tilde{y}$ seja a média aritmética temos:

$$\tilde{y}_{i..} = \frac{1}{rb} \sum_{j=1}^{b} \sum_{k=1}^{r} y_{ijk}$$

$$\tilde{y}_{ij.} = \frac{1}{r} \sum_{k=1}^{r} y_{ijk}$$

$$\tilde{y}_{...} = \frac{1}{prb} \sum_{i=1}^{p} \sum_{j=1}^{b} \sum_{k=1}^{r} y_{ijk}$$

Para cada transformação de alinhamento definimos R_{ijk} o posto de x_{ijk} na amostra conjunta, os escores $a(R_{ijk})$ são escolhidos de acordo as funções seletoras φ_1 e φ_2 definidas na seção (1.6.2), que mede a longitude das caudas e a simetria dos dados originais y_{ijk}.

2.1.4 - Estatística do Teste

a) Para testar H_{01}

definimos a estatistica linear de posto $S_{i..}$ da seguinte maneira:

$$S_{i..} = \frac{1}{br} \sum_{j=1}^{b} \sum_{k=1}^{r} c_{ijk} a(R_{ijk}) \quad , \; i=1,2,\ldots,p$$

onde:

$$c_{ijk} = \begin{cases} 1 \text{ , se } y_{ijk} \text{ pertence a i-ésima casela} \\ 0 \text{ , caso contrario.} \end{cases}$$

Se a(.) satisfaz as condições dos teoremas de simetria (1.3.2) e normalidade (1.3.3), temos que a estatística linear de posto $S_{i..}$ tem distribuição assintótica Normal com parâmetros :

$$S_{i..} \sim N\ (\ E(S_{i..})\ ,\ VAR(S_{i..})\ )$$

do teorema (1.3.1) temos:

$$E(S_{i..}) = \overline{a}_{...} = \overline{S}_{...}$$

onde:

$$\overline{a}_{...} = \frac{1}{N} \sum_{i=1}^{p} \sum_{j=1}^{b} \sum_{k=1}^{r} a(R_{ijk})$$

e

$$VA(S_{i..}) = \frac{(N-br)br}{N} \Delta^2$$

A estatística do teste H_{01} é da forma : (Kubinger :1986).

$$Q_1 = \frac{SQ(A)}{\Delta^2}$$

onde:

$$SQ(A) = br \sum_{i=1}^{p} (S_{i..} - \overline{S}_{...})^2$$

$$\Delta^2 = \frac{1}{N-1} \sum_{i=1}^{p} \sum_{j=1}^{b} \sum_{k=1}^{r} (a(R_{ijk}) - \bar{a}_{...})^2$$

sob H_{01}, a estatística Q_1 tem distribuição assintótica Qui Quadrado com p-1 graus de liberdade (teorema 1.3.4), isto é,

$$Q_1 \sim \chi^2_{(p-1)}$$

A regra de decisão é a seguinte :

Se $Q_1 > \chi^2_{(p-1)}(1-\alpha)$, então, a Hipótese H_{01} é rejeitada ao nível de significância α.

2.1.4.1 - Comparações Múltiplas.

Se para o fator A, a hipótese nula H_{01} é rejeitada ao nível de significância α, então, admitimos que pelo menos dois níveis do fator A tem efeitos diferentes e para se determinar estes níveis, realizamos o seguinte teste :

$H_0 : \alpha_i = \alpha_{i'}$, para $i < i' = 2,3,\ldots,p$

$H_1 : \alpha_i \neq \alpha_{i'}$, para algum $i \neq i'$

com a seguinte regra de decisão:

rejeitamos a hipótese nula H_0 se :

$$| D_{ii'} | \geq Z_{\alpha'/2} \mathrm{Var}(D_{ii'})^{1/2} ,$$

onde $\alpha' = 2\alpha / (p (p-1))$

sendo: $\phi(Z_{\alpha'/2}) = 1 - \alpha'/2$

Ou ainda pode-se também usar a seguinte estatística (Sen and Purim: 1985), rejeitamos a hipótese H_0 se:

$$| D_{ii'} | > [\mathrm{VAR}(D_{ii'}) \; \chi^2_{(p-1)}(1-\alpha/2)]^{1/2}$$

onde:

$$D_{ii'} = (S_{i..} - S_{i'..}) \qquad , \quad i,i'=1,2,\ldots,p$$

$$n_i = br$$

e

$$\mathrm{VAR}(D_{ii'}) = \left(\frac{1}{n_i} + \frac{1}{n_{i'}} \right) \Delta^2$$

$$\Delta^2 = \frac{1}{pbr-1} \sum_{i=1}^{p} \sum_{j=1}^{b} \sum_{k=1}^{r} (a(R_{ijk}) - \bar{a}_{...})^2$$

b) <u>Para testar H_{02}</u>

Definimos $S_{ij.} = \frac{1}{r} \sum_{k=1}^{r} c_{ijk}\, a(R_{ijk})$, $i=1,2,\ldots,p$; $j=1,2,\ldots,b$

onde:

$$c_{ijk} = \begin{cases} 1 \text{ , se } y_{ijk} \text{ pertence a ij-ésima casela} \\ 0 \text{ , caso contrário.} \end{cases}$$

Se a(.) satisfaz as condições dos teoremas de simetria (1.3.2) e normalidade (1.3.3), temos que a estatística linear de posto $S_{ij.}$ tem distribuição assintótica Normal com parâmetros :

$$S_{ij.} \sim N\ (\ E(\ S_{ij.}\ )\ ,\ VAR(\ S_{ij.}\ )\ )$$

do teorema (1.3.1) temos:

$$E(S_{ij.}) = \bar{a}_{i..} = \bar{S}_{ij.}$$

onde:

$$\bar{a}_{i..} = \frac{1}{br} \sum_{j=1}^{b} \sum_{k=1}^{r} a(R_{ijk}) = \bar{S}_{ij.}$$

e

$$VAR(S_{ij.}) = \frac{(N-r)r}{N} \Delta^2$$

onde:

$$\Delta^2 = \frac{1}{N-1} \sum_{i=1}^{p} \sum_{j=1}^{b} \sum_{k=1}^{r} (a(R_{ijk}) - \bar{a}_{...})^2$$

A estatística do teste é da forma: (Kubinger 1986)

$$Q_2 = \frac{SQ_\beta(A)}{\Delta^2}$$

onde:

$$SQ_\beta(A) = r \sum_{i=1}^{p} \sum_{j=1}^{b} (S_{ij.} - \bar{S}_{ij.})^2$$

sob H_{02}, a estatística Q_2 tem distribuição assintótica Qui Quadrado com p(b-1) graus de liberdade, isto é,

$$Q_2 \sim \chi^2_{p(b-1)}$$

A regra de decisão é a seguinte:
ao nível de significância α, se $Q_2 > \chi^2_{p(b-1)}(1-\alpha)$, então, a Hipótese H_{02} é rejeitada .

Caso rejeitemos a hipótese H_{02}, podemos testar dentro de quais níveis de A ocorre diferença significativa entre os níveis de B.

Para isto, testaremos isoladamente a hipótese:

$$H_{02i} : \beta_1(i) = \beta_2(i) = \ldots = \beta_b(i) = 0$$

A estatística do teste é da forma:

$$Q_{\beta}(A_i) = \frac{SQ_{\beta}(A_i)}{\Delta^2}$$

onde :

$$SQ_{\beta}(A_i) = r \sum_{j=1}^{b} (S_{ij.} - \overline{S}_{ij.})^2$$

Sob H_{02i}, a estatística $Q_{\beta}(A_i)$ tem distribuição assintótica Qui-Quadrado com b-1 graus de liberdade, isto é,

$$Q_{\beta}(A_i) \sim \chi^2_{(b-1)}$$

A regra de decisão é a seguinte:

rejeitamos H_{02i}, ao nível de significância α se:

$$Q_{\beta}(A_i) > \chi^2_{(b-1)}(1-\alpha) \ ,$$

Como podemos observar, para cada classe de hipótese existem diferentes transformações e consequentemente diferentes estatísticas Δ^2, $S_{i..}$, $S_{ij.}$, $\overline{a}_{...}$, etc.

Caberia futuramente, analizar comparacões múltiplas para se determinar que níveis do fator B dentro do i-ésimo nível de A são diferentes.

2.2.- TESTES NÃO-PARAMÉTRICOS PARA DELINEAMENTO EM BLOCOS ALEATÓRIOS NÃO BALANCEADOS

Em experimentação, por várias razões, nem sempre é possível usar blocos de tamanhos que possibilitem acomodar todos os tratamentos. Se todos os tratamentos não são alocados em todos os blocos, tem-se um delineamento em blocos incompletos. O caso de blocos incompletos balanceados já foi apresentado em análise não-paramétrica por Negrillo (1987).

Consideremos o experimento com n blocos aleatórios, envolvendo k tratamentos. Suponhamos que nem todos os blocos recebem os k tratamentos, isto é, o número de tratamentos no j-ésimo bloco é k_j , onde $2 \leq k_j \leq k$.

Seja X_{ij} a observação do i-ésimo tratamento do j-ésimo bloco onde $i \in \beta_j$; $j=1,2,\ldots,n$, e β_j é o conjunto de índices do j-ésimo bloco, correspondente àqueles tratamentos, nos quais as observações estão presente.

Seja $\nu_i \leq n$, o conjunto de índices de àqueles blocos que tem observações do i-ésimo tratamento.

Consideremos o seguinte modelo matemático associado ao experimento:

$$X_{ij} = u_j + \tau' Z_{ij} + \sigma e_{ij}$$

sujeito à seguinte restrição nos parâmetros:

$$\sum_{i=1}^{k} \tau_i = 0$$

onde:

X_{ij} : resposta associada ao i-ésimo tratamento no j-ésimo bloco,

$\tau' = (\tau_1, \tau_2, \ldots \tau_k)$, vetor que contém k efeitos de tratamento de interesse, todos identificáveis,

$Z'_{ij} = (z_{1i}, z_{2i}, \ldots, z_{ki})$ é um indicador ;

onde $z_{pi} = \begin{cases} 1 , & \text{se } i=p \\ 0 , & \text{se } i \neq p \end{cases}$ para $i, p = 1, 2, \ldots, k$.

u_j e σ são respectivamente os parâmetros de locação e dispersão do j-ésimo bloco;

e_{ij} é o erro aleatório.

2.2.1 - Suposições

Os erros e_{ij} são variáveis aleatórias independentes identicamente distribuidas com função de distribução absolutamente contínua, e $E(e_{ij}) = 0$.

A resposta X_{ij} tem função distribuição absolutamente contínua.

2.2.2 - Hipótese de Interesse

A hipótese que se testará neste tipo de experimento é:

H_0: $\tau_1 = \tau_2 = \ldots = \tau_k = 0$, (não existe efeito de tratamento)

H_1:pelo menos algum $\tau_i \neq 0$.

2.2.3 - Estatística do Teste

A estatística do teste é derivada utilizando a conhecida estatística linear de postos que descreveremos a seguir :
seja R_{ij} o posto de X_{ij} dentro do j-ésimo bloco, de acordo com a função indicadora de simetria $\bar{\varphi}_2 = \sum_i n_i \varphi_{2.i}$ / N, e de longitude das caudas $\overline{\varphi}_1 = \sum_i n_i \varphi_{1.i}$ /N , definidas na seção (1.6.2), onde:

n_i é o número de observações do i-ésimo tratamento;

N é o número total de observações;

$\varphi_{2.i}$ e $\varphi_{1.i}$ são calculados oara cada tratamento,

selecionamos a função escore a(.), e neste caso, utilizaremos os escores transformados dados por Burnett and Willan (1988):

$$a^*(R_{ij}) = \begin{cases} \dfrac{(a(R_{ij}) - \bar{a}_j)}{\bar{a}_j} & , \text{ se } \overline{a}_j \neq 0 \\ a(R_{ij}) & , \text{ se } \overline{a}_j = 0 \end{cases}$$

onde :

$$\dot{a}_j = \sum_{i\in\beta_j} a(R_{ij}) \;/\; k_j \qquad , \; j=1,2,\ldots,n$$

assim temos que:

$$E\;[a^*(R_{ij})] = 0.$$

Desta forma a estatística linear de postos para i-ésimo tratamento é definida:

$$S_i = \sum_{j\in\upsilon_i} a^*(R_{ij}) \qquad \text{para } i=1,2,\ldots,k$$

onde :

υ_i é o conjunto de índices daqueles blocos que tem observações do i-ésimo tratamento.

Consideremos o vetor $S'=(S_1,S_2,\ldots,S_k)$, que sob H_0, tem $E(S)=0$ e matriz de covariância $V = [(v_{it})]$; $i,t=1,2,\ldots,k$ onde:

$$\begin{cases} v_{it} = \sum_{j\in\upsilon_i} \left(\sum_{p\in\beta_j} (a^*(R_{pj}))^2/k_j \right) \;; & \text{para } i=t=1,2,\ldots,k \\ v_{it} = -\sum_{j\in\upsilon_i\cap\upsilon_t} \left[\sum_{p\in\beta_j} a^*(R_{pj})^2 \;/(k_j(k_j-1)) \right] \;; & \text{para } i\neq t=1,2,\ldots,k \end{cases}$$

Como os escores $a^{*}(R_{ij})$ são uniformemente limitados (Woolson and Loncherbruch: 1981) e satisfazem a condição de simetria do teorema (1.3.2) e da seção (1.5), temos, que a estatística linear de posto multivariada $n^{-1/2}S$ tem distribuição assintótica normal multivariada com vetor de médias zero e matriz de covariâcia :

$$V = ((v_{it})), \quad i,t=1,2,...k.$$

Sob H_0, a forma quadrática $Q = S'V^{-1}S$, tem distribuição assintótica Qui-Quadrado com k-1 graus de liberdade. (Bruner and Willian :1988). E a estatística Q fornece um teste localmente mais poderoso sob H_0.

2.2.4 - Comparações Múltiplas

Se a hipótese H_0 é rejeitada ao nível de significância α, e se desejamos determinar quais são os tratamentos que são iguais ou diferentes sob o ponto de vista estatístico, podemos fazer comparações múltiplas, isto é, as hipóteses são dadas por:

$H_{0i} : \tau_i = \tau_{i'}$, para todo $i \neq i' = 1,2,...,k$

$H_{1i} : \tau_i \neq \tau_{i'}$ para algum $i \neq i'$

Seja a estatística $D_{ii'} = (\overline{S}_i - \overline{S}_{i'})$, para $i \neq i'$

onde :

$$\overline{S}_i = S_i / n_i$$

e n_i é o número de observações do i-ésimo tratamento.
A variável $D_{ii'}$ sob H_0, tem esperança zero e variância :

$$VAR(D_{ii'}) = \frac{1}{n_i^2} \sum_{j \in \upsilon_i} (A_j / k_j) + \frac{1}{n_{i'}^2} \sum_{j \in \upsilon_{i'}} (A_j / k_j) + $$
$$+ \frac{2}{n_i n_{i'}} \sum_{j \in \upsilon_i \cap \upsilon_{i'}} (A_j / (k_j (k_j - 1)))$$

onde:

$$A_j = \sum_{p \in \beta_j} a^*(R_{pj})^2 \text{ , para } j=1,2,\ldots,n$$

A hipótese nula H_{0i}, é rejeitada ao nível de significância α se:

$$| D_{ii'} | = | \overline{S}_i - \overline{S}_{i'} | \geq \{ \chi^2_{(\hat{k}-1)} (1-\alpha/2) \ VAR(D_{ii'}) \}^{1/2}$$

onde:

$$\hat{k} = \min_{1 \leq j \leq n} k_j \ .$$

CAPÍTULO III

TESTES NÃO-PARAMÉTRICOS EM EXPERIMENTOS COM MEDIDAS REPETIDAS

Neste capítulo apresentaremos a análise estatística de dados provenientes de um planejamento, com medidas repetidas, usando a teoria Não-Paramétrica.

Geralmente certos aspectos de análise de Variância de medidas repetidas podem ser justificadas supondo que as observações tem distribuição normal. Quando esta suposição não é satisfeita, então os testes Não-Paramétricos, como se viu anteriormente, são complemento viáveis.

3.1. - TE S TES NÃO-PARAMÉTRICOS PARA MODELO MISTO EM EXPERIMENTOS COM MEDIDAS REPETIDAS

Apresentaremos um experimento com medidas repetidas utilizando modelos mistos, envolvendo n sujeitos selecionados aleatoriamente, que respondem a um tratamento em cada uma das p condições de avaliação, definidas previamente, e estamos interessados em determinar se existe ou não efeitos das condições de avaliação; como por exemplo tempo, temperatura, dosagem, etc .

Exemplo:

Um pesquisador deseja verificar se existe diferenças no tempo, para um exame de sensibilidade auditiva em pacientes. Escolhe ao acaso um número determinado de pessoas e realiza o exame em diferentes períodos e tempo, e mede o tempo médio de resposta. Neste exemplo, a condição de avaliação é o tempo.

Tabela 3.1 - ESTRUTURA DE DADOS BÁSICA PARA ESTUDOS DE PLANEJAMENTOS COM MEDIDAS REPETIDAS.

	condições de avaliação				
Sujeitos	C1	C2	C3	. . .	Cp
S1	y_{11}	y_{12}	y_{13}	. . .	y_{1p}
S2	y_{21}	y_{22}	y_{23}	. . .	y_{2p}
S3	y_{31}	y_{32}	y_{33}	. . .	y_{3p}
⋮	⋮				⋮
Sn	y_{n1}	y_{n2}	y_{n3}	. . .	y_{np}

Assim, o modelo considera que os efeitos do sujeito são fatores aleatórios e os efeitos das condições de avaliação são fatores fixos.

O principal objetivo será estudar o comportamento das respostas em diferentes condições de avaliação.

Seja :

y_{ij} : a resposta do i-ésimo sujeito à j-ésima condição de avaliação. $i=1,2,\dots,n$ $j=1,2,\dots,p$.

Desta forma, cada sujeito responde a p distintas condições de avaliação.

Definimos a matriz $\underset{\sim}{Y} = (y_{ij})$, formada de n vetores linha estocasticamente independentes

$$\underset{\sim}{y}_i' = (y_{i1}, y_{i2}, \dots, y_{ip}) \quad , i=1,2,\dots,n.$$

3.1.1 - Suposições

Com a finalidade de aplicar a teoria não-paramétrica desenvolvida no capítulo I, apresentaremos as seguintes suposições:

A1).- $\underset{\sim}{y}_i$ tem função de distribuição acumulada p-variada absolutamente contínua $F_i(y)$.

onde:

$$F_i(y) = G_i(x - \mu_i)$$

$\underset{\sim}{\mu}_i' = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{ip})$,vetor de parâmetros de posição

onde: $\mu_{ij} = \beta_i + \tau_j$ $\quad i=1,2,\ldots,n$; $j=1,2,\ldots,p$ $\quad$ (3.1.1.1)

sendo:

β_i : o efeito do i-ésimo sujeito.

τ_j : o efeito da j-ésima condição de avaliação.

A2).- A distribuição conjunta de qualquer conjunto linearmente independente de contrastes, entre as observações em qualquer sujeito particular, é **simétrica diagonal**.

Esta suposição significa:

Se temos o seguinte constraste entre as observações:

$$\underset{(p-1)\times 1}{\underset{\sim}{U}_i} = \underset{(p-1)\times p}{\underset{\sim}{C}} \; \underset{p\times 1}{\underset{\sim}{y}_i} \qquad (3.1.1.2)$$

onde:

$$\begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{i(p-1)} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ 1 & 0 & -1 & 0 & \dots & 0 \\ \vdots & & & & & \vdots \\ 1 & 0 & 0 & 0 & \dots & -1 \end{bmatrix} \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{ip} \end{bmatrix}$$

$\underset{\sim}{C}$ é uma matriz (p-1)xp de constrastes linearmente independente por linha , isto é, $\underset{\sim}{C} \times \underset{\sim}{1}_p = \underset{\sim}{0}$

$\underset{\sim}{1}'_p = (1,1,\dots,1)$ vetor unitário px1

$\underset{\sim}{0}$ é um vetor nulo (p-1) x 1

e definimos o seguinte contraste:

$$\underset{(p-1)\times 1}{\underset{\sim}{\theta}} = \underset{(p-1)\times p}{\underset{\sim}{C}} \; \underset{(p-1)\times 1}{\underset{\sim}{\mu}_i} = \underset{(p-1)\times p}{\underset{\sim}{C}} \; \underset{(p-1)\times 1}{\underset{\sim}{\tau}} \qquad (3.1.1.3)$$

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{(p-1)} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ 1 & 0 & -1 & 0 & \dots & 0 \\ \vdots & & & & & \vdots \\ 1 & 0 & 0 & 0 & \dots & -1 \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_p \end{bmatrix}$$

sendo $\underset{\sim}{\mu}_i$ e $\underset{\sim}{\tau}' = (\tau_1, \tau_2, \dots, \tau_p)$, definidos em (3.1.1.1)

então:

A distribuição de $(\underset{\sim}{U}_i - \underset{\sim}{\theta})$ e $(\theta - \underset{\sim}{U}_i)$, para i=1,2,...,n é a mesma; por tanto é simétrica diagonal.

3.1.2 - Suposições Adicionais

Ante a necessidade de realizar a análise estatística a este tipo de planejamento utilizando o modelo Misto, apresentaremos duas suposições adicionais que podem ou não ser impostas.

A3) A aditividade dos efeitos do sujeito. Isto implica que

$$G_1 = G_2 = \ldots = G_n = G;$$

A4) Estrutura de Uniformidade na matriz de covariância Σ, da forma :

$$\Sigma = \rho\sigma^2\mathbf{J} + (1-\rho)\sigma^2\mathbf{I}.$$

sendo $\mathbf{J}_{pxp}$ matriz de 1's e $\mathbf{I}_{pxp}$ matriz identidade. Isto implica que a variabilidade da resposta do sujeito é independente das condições de avaliação.

3.1.3 - Hipótese de Interesse

De acordo com a suposição (A1) a hipótese nula de interesse pode ser escrita da seguinte forma:

$H_0 : \tau_1 = \tau_2 = \ldots = \tau_p$ (não existe efeito de condições de avaliação).

versus

H_1: ao menos um dos τ_i's são diferentes .

Para testar esta hipótese temos a seguintes situações:

3.1.4 - CASO 1

Ante a necessidade de efetuar o teste da hipótese nula de interesse, temos neste primeiro caso, que a suposição A3) não é satisfeita e a estrutura de covariância é geral.

Segundo estas condições, desenvolvemos o seguinte procedimento: se temos o seguinte contraste entre as observações:

$$\underset{(p-1)x1}{\underset{\sim}{U}_i} = \underset{(p-1)Xp}{\underset{\sim}{C}} \ \underset{px1}{\underset{\sim}{y}_i} \qquad \text{definido em (3.1.1.2)}$$

$$\begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{i(p-1)} \end{bmatrix} = \begin{bmatrix} y_{i1} - y_{i2} \\ y_{i1} - y_{i3} \\ \vdots \\ y_{i1} - y_{ip} \end{bmatrix}$$

de (3.1.1.1) sabemos que: $\mu_{ij} = \beta_i + \tau_j$, para $i=1,2,\ldots,n$

$j=1,2,\ldots,p$.

Seja $\underset{\sim}{\theta}$ o vetor definido em (3.1.1.3), supõe que $(\underset{\sim}{U}_i - \underset{\sim}{\theta})$ e $(\underset{\sim}{\theta} - \underset{\sim}{U}_i)$ tem a mesma distribuição,

sob a hipótese nula de interesse H_0: $\underset{\sim}{\theta} = \underset{\sim}{0}$, então, $\underset{\sim}{U}_i$ e $-\underset{\sim}{U}_i$ têm a mesma distribuição, para $i=1,2,\ldots,n$.

Segundo o parágrafo anterior, testar

H_0: $\tau_1 = \tau_2 = \ldots = \tau_p$; é equivalente a testar:

$H_{01}: \underset{\sim}{\theta} = \underset{\sim}{0}$. Como $\underset{\sim}{\theta} = \underset{\sim}{C}\,\underset{\sim}{\tau}$ então :

$$H_{01}: \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ 1 & 0 & -1 & 0 & \dots & 0 \\ \vdots & & & & & \vdots \\ 1 & 0 & 0 & 0 & \dots & -1 \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

observamos que esta suposição é menos restrita que a usual suposição de multinormalidade de $\underset{\sim}{U}_i$. (Koch and Sen : 1968).

Para testar H_{01} , procedemos da seguinte maneira:

Seja R_{ij} o posto de y_{ij} na amostra conjunta $\{y_{i1}, y_{i2}, \dots, y_{ip}\}$ isto é, os postos são determinados por sujeito. Para $i=1,2,\dots,n$ e $j=1,2,\dots,p$.

Definimos a matriz de postos $\underset{\sim}{R}$ da seguinte maneira:

$$\underset{n\times p}{\underset{\sim}{R}} = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1p} \\ R_{21} & R_{22} & \dots & R_{2p} \\ \vdots & & & \vdots \\ R_{n1} & R_{n2} & \dots & R_{np} \end{bmatrix}$$

de acordo com as funções seletoras φ_2 e φ_1 definidas na seção (1.6.2), determinamos a função escore a(.), obtendo a matriz de escores $\underset{\sim}{a}$

$$\underset{n\times p}{\underset{\sim}{a}} = \begin{bmatrix} a(R_{11}) & a(R_{12}) & \dots & a(R_{1p}) \\ a(R_{21}) & a(R_{22}) & \dots & a(R_{2p}) \\ \vdots & & & \vdots \\ a(R_{n1}) & a(R_{n2}) & \dots & a(R_{np}) \end{bmatrix}$$

Definimos a estatística linear de postos S_j da seguinte maneira:

$$S_j = \frac{1}{n}\sum_{i=1}^{n} a(R_{ij}) \quad , \qquad j=1,2,\ldots,p$$

formando o vetor $\underset{\sim}{S}' = (S_1, S_2, \ldots, S_p)$.

Se a(.) satisfaz as condições dos teoremas de simetria (1.3.2) e normalidade (1.3.3), temos que a estatística linear de postos S_j tem distribuição assintótica Normal com parâmetros:

$$S_j \sim N(\, E(S_j)\, , VAR(S_j)\,)$$

sendo:

$E(S_j) = \bar{a}_{..}$

onde:

$$\bar{a}_{..} = \frac{1}{np}\sum_{i=1}^{n}\sum_{j=1}^{p} a(R_{ij}) = \bar{S}_.$$

e

$$COV(S_j, S_{j'}) = v_{jj'}$$

onde:

$$v_{jj'} = \frac{1}{n^2}\sum_{i=1}^{n}(\, a(\, R_{ij}) - \bar{a}_{..})\,(\, a(\, R_{ij'}) - \bar{a}_{..})$$

para todo $j,j' = 1,2,\ldots,p$. Assim, definimos a matriz de covariância

$$V = (\, (\, v_{jj'}\,)\,).$$

A estatística do teste é uma estatística denominada estatística de WALD e é dada por:

$$W_1 = \underset{\sim}{S}'C'(CVC')^{-1}C\,\underset{\sim}{S}\,. \qquad (3.1.4.1)$$

Sob H_0, a estatística W_1 tem uma distribuição assintótica Qui-Quadrado com p-1 graus de liberdade, isto é,

$$W_1 \sim \chi^2_{(p-1)}$$

A regra de decisão é a seguinte:

Se $W_1 > \chi^2_{(p-1)}(1-\alpha)$, então , a hipótese nula H_{01} é rejeitada ao nível de significância α .

Visto que a estatística W é baseada na estatística linear de posto, esta é claramente não influenciada pela aditividade dos efeitos. Também a estatística de posto é menos vulnerável a observações discrepantes.

3.1.4.1 -Comparações Múltiplas

No caso de rejeitar H_{01}, teríamos que estudar que condição de avaliação tem efeito significativo, para isso, trabalharemos com as comparações múltiplas utilizando a matriz de contrastes C de posto(C) = p-1, definido em (3.1.1.2).

Exemplo: Considere-se a seguinte hipótese nula:

$H_0 : \tau_1 = \tau_2 = \tau_3$,

que é equivalente a testar $H_0 : \tau_1 - \tau_2 = 0$ e $\tau_1 - \tau_3 = 0$. Então

$$H_0 : \quad \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & -1 & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \vdots \\ \tau_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

neste caso a matriz de constraste $C = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & -1 & 0 & 0 & \dots & 0 \end{bmatrix}$

E utiliza-se a estatística W_1 definida em (3.1.4.1).

3.1.5 - CASO 2

Neste segundo caso consideremos:

1.-Não é conhecida a aditividade dos efeitos ;

2.-asumimos a suposição que a matriz de covariância tem estrutura uniforme.

Sob H_0, e sob estas suposições, $\underset{\sim}{y}_i$ é formada de p-variáveis aleatórias intercambiáveis para todo $i=1,2,\ldots,n$; assim, a distribuição de $\underset{\sim}{y}_i$ é invariante, sobre qualquer permutação das coordenadas para $i=1,2,\ldots,n$ isto é,

$$f(y_{i1},y_{i2},\ldots,y_{ip}) \equiv f(y_{i\alpha_1},y_{i\alpha_2},\ldots,y_{i\alpha_p}).$$

Seja R_{ij} o posto de y_{ij} na amostra conjunta $\{y_{i1},y_{i2},\ldots,y_{ip}\}$ para $i=1,2,\ldots,n$

definimos a matriz de postos $\underset{\sim}{R}$ da seguinte maneira:

$$\underset{n\times p}{\underset{\sim}{R}} = \begin{bmatrix} R_{11} & R_{12} & \ldots & R_{1p} \\ R_{21} & R_{22} & \ldots & R_{2p} \\ \vdots & & & \vdots \\ R_{n1} & R_{n2} & \ldots & R_{np} \end{bmatrix}$$

Segundo as funções seletoras φ_2 e φ_1 definidas na seção (1.6.2), determinamos a função escore a(.), obtendo a matriz de escores $\underset{\sim}{a}$

$$\underset{n\times p}{\underset{\sim}{a}} = \begin{bmatrix} a(R_{11}) & a(R_{12}) & \ldots & a(R_{1p}) \\ a(R_{21}) & a(R_{22}) & \ldots & a(R_{2p}) \\ \vdots & & & \vdots \\ a(R_{n1}) & a(R_{n2}) & \ldots & a(R_{np}) \end{bmatrix}$$

desta forma determinamos a estatística linear de postos S_j

onde : $$S_j = \frac{1}{n} \sum_{i=1}^{n} a(R_{ij}) \quad , \text{ para } j=1,2,\ldots,p$$

formando o vetor $\underset{\sim}{S}' = (S_1, S_2, \ldots, S_p)$.

Se a(.) satisfaz as condições dos teoremas de simetria (1.3.2) e normalidade (1.3.3), temos que a estatística linear de postos S_j tem distribuição assintótica Normal com parâmetros:

$$S_j \sim N\left(E(S_j) \ , \ VAR(S_j) \right)$$

sendo:

$$E(S_j) = \overline{a}_{..}$$

tal que:

$$\overline{a}_{..} = \frac{1}{np} \sum_{i=1}^{n} \sum_{j=1}^{p} a(R_{ij}) = \overline{S}_{.}$$

e

$$COV(S_j \ , \ S_{j'}) = \frac{p(\delta_{jj'}) - 1}{(p-1)n} \quad \sigma^2$$

onde:

$$\delta_{jj'} = \begin{cases} 1 & , \text{ se } j=j' \\ 0 & , \text{ se } j \neq j' \end{cases}$$

$$\sigma^2 = \frac{1}{np} \sum_{i=1}^{n} \sum_{j=1}^{p} \left(a\left(R_{ij}\right) - \overline{a}_{..} \right)^2 .$$

A estatística do teste é :(Koch and Sen: 1968)

$$W_2 = \sum_{j=1}^{p} (S_j - \overline{S}_{.})^2 \left[\frac{n\ (\ p\ -\ 1)}{p\sigma^2} \right]$$

Sob H_0, a estatística W_2 tem uma distribuição assintótica Qui-Quadrado com (p-1) graus de liberdade, isto é,

$$W_2 \sim \chi^2_{(p-1)}$$

A regra de decisão é a seguinte:

Se $W_2 > \chi^2_{(p-1)}(1-\alpha)$, então, a hipótese nula H_0 é rejeitada ao nível de significância α.

3.2.5.1 - Comparações Múltiplas

Se a hipótese nula H_0 é rejeitada ao nível α, estamos admitindo que pelo menos dois níveis do fator tem efeitos diferentes e para determinar os níveis que são diferentes, podemos realizar o teste:

H_0: $\tau_j = \tau_{j'}$, para todo $j \neq j' = 1, 2, \ldots, p$
H_1: $\tau_j \neq \tau_{j'}$, para algum $j \neq j' = 1, 2, \ldots, p$
a estatística do teste é da seguinte forma:

$$D_{jj'} = (S_j - S_{j'})$$

com a seguinte regra de decisão:

rejeitamos a hipótese nula H_0, se

$$|D_{jj'}| > [\chi^2_{(p-1)}(1-\alpha/2)(VAR(D_{jj'}))]^{1/2}$$

onde

$VAR(D_{jj'}) = (1/n + 1/n)\delta^2$

sendo:

$$\Delta^2 = \frac{1}{np-1} \sum_{i=1}^{n} \sum_{j=1}^{p} (a(R_{ij}) - \bar{a}_{..})^2$$

3.1.6 - CASO 3

Neste caso consideraremos a seguinte situação:

1.- Aditividade dos efeitos. (Cumpre a suposição A3).

2.- A matriz de covariância tem uma estrutura geral.

Se é conhecida a aditividade dos efeitos ,então, esta informação permite aumentar o poder do teste , e podemos escrever a resposta do i-ésimo sujeito à j-ésima condição de avaliação da seguinte maneira:

$$y_{ij} = \mu + \beta_i + \tau_j + e_{ij}$$

sujeito a:

$$\sum_{j=1}^{p} \tau_j = 0$$

onde :

y_{ij}: resposta do i-ésimo sujeito na j-ésima condição de avaliação.

μ : média geral.

β_i : efeito aleatório do i-ésimo sujeito.

τ_j : efeito fixo do j-ésima condição de avaliação.

e_{ij} : erro aleatório.

Supondo que :

$\underset{\sim}{e}'_i = (e_{i1}, e_{i2}, \ldots, e_{ip})$, para $i=1,2,\ldots,n$ são n-vetores aleatórios p-variados independentes, com função de distribuição absolutamente contínua e simétrica em torno de zero.

Seja a hipótese nula de interesse :

$H_0 : \tau_1 = \tau_2 = \ldots = \tau_p = 0$ (não existe efeito de condição de avaliação).

Considere-se a transformação de alinhamento, para testar se existem os efeitos das condições de avaliação, pois, através dela eliminamos os efeitos aleatórios do sujeito.

$$z_{ij} = y_{ij} - \hat{\beta}_i$$

então:

$$z_{ij} = y_{ij} - \overline{y}_{i.} + \overline{y}_{..} \qquad , \ i=1,2,\ldots,n \quad e \quad j=1,2,\ldots,p$$

tendo em conta que a transformação de alinhamento também pode realizar-se com a mediana ou outra estatística não-paramétrica.

Formando a matriz $\underset{\sim}{Z}$

$$\underset{n\times p}{\underset{\sim}{Z}} = \begin{bmatrix} z_{11} & z_{12} & \ldots & z_{1p} \\ z_{21} & z_{22} & \ldots & z_{2p} \\ \vdots & & & \vdots \\ z_{n1} & z_{n2} & \ldots & z_{np} \end{bmatrix}$$

se definimos : $\underset{\sim}{Z}_i = (z_{i1}, z_{i2}, \ldots, z_{ip})$, para $i=1,2,\ldots,n$

$= \underset{\sim}{y}_i (\underset{\sim}{I}_p - p^{-1} \underset{\sim}{1}'_p \underset{\sim}{1}_p)$

onde:

$\underset{\sim}{y}_i = (y_{i1}, y_{i2}, \ldots, y_{ip})$

$\underset{\sim}{I}_p$ matriz identidade pxp

$\underset{\sim}{1}'_p = (1,1,\ldots,1)$ vetor unitário px1, então,

sob H_0, $\underset{\sim}{Z}_i = \underset{\sim}{e}_i (\underset{\sim}{I}_p - p^{-1} \underset{\sim}{1}'_p \underset{\sim}{1}_p)$, $i=1,2,\ldots,n$

onde pela suposição (A2) de simetria das diagonais da distribuição de $\underset{\sim}{e}_i$ e a simetria de ($\underset{\sim}{I}_p - \underset{\sim}{1}'_p \underset{\sim}{1}_p$), a distribuição de $\underset{\sim}{Z}_i$ é também simétrica diagonal en torno do vetor $\underset{\sim}{0}$.

Isto é, sob H_0, $\underset{\sim}{Z}_i$ e - $\underset{\sim}{Z}_i$ tem a mesma distribuição.

Para testar a hipótese nula H_0, procedemos da seguinte maneira:

Seja R_{ij} o posto de z_{ij} na amostra conjunta $\{z_{11}, z_{12}, \ldots, z_{np}\}$ para $i=1,2,\ldots,n$ e $j=1,2,\ldots,p$.

Determinando a matriz de postos $\underset{\sim}{R}$

$$\underset{n\times p}{\underset{\sim}{R}} = \begin{bmatrix} R_{11} & R_{12} & & R_{1p} \\ R_{21} & R_{22} & \ldots & R_{2p} \\ \vdots & & & \vdots \\ R_{n1} & R_{n2} & \ldots & R_{np} \end{bmatrix}$$

A função escore a(.) é definida segundo os indicadores φ_2 e φ_1 definidos na seção (1.6.2), que mede a simetria e longitude das caudas das respostas y_{ij}.

Assim, obtemos a matriz de escores $\underset{\sim}{a}$

$$\underset{n\times p}{\underset{\sim}{a}} = \begin{bmatrix} a(R_{11}) & a(R_{12}) & \dots & a(R_{1p}) \\ a(R_{21}) & a(R_{22}) & \dots & a(R_{2p}) \\ \vdots & & & \vdots \\ a(R_{n1}) & a(R_{n2}) & \dots & a(R_{np}) \end{bmatrix}$$

Definimos a estatística linear de postos S_j

onde: $$S_j = \frac{1}{n} \sum_{i=1}^{n} a(R_{ij}) \quad , \text{ para } j=1,2,\dots,p$$

formando o vetor $\underset{\sim}{S}' = (S_1, S_2, \dots, S_p)$.

Se $a(.)$ satisfaz as condições dos teoremas de simetria (1.3.2) e normalidade (1.3.3), temos que a estatística linear de postos S_j tem distribuição assintótica Normal com parâmetros:

$$S_j \sim N\ (\ E(S_j)\ ,\ VAR(S_j\ )\ )$$

onde:

$E(S_j) = \bar{a}_{..}$

sendo:

$$\bar{a}_{..} = \frac{1}{np} \sum_{i=1}^{n} \sum_{j=1}^{p} a(R_{ij}) = \bar{S}.$$

e

$$COV(S_j, S_{j'}) = \frac{1}{n^2} \sum_{i=1}^{n} (a(R_{ij}) - \bar{a}_{..})\ (a(R_{ij'}) - \bar{a}_{..}) = v_{jj'}$$

assim, definimos a matriz de covariância $V = [(v_{jj'})]$ e a estatística do teste é : (MEYDRECH : 1986)

$$W_3 = (S - \bar{S}.\underset{\sim}{1}_p) \left[(\underset{\sim}{I}_p - p^{-1}\underset{\sim}{1}'_p\ \underset{\sim}{1}_p)\ V\ (\underset{\sim}{I}_p - p^{-1}\underset{\sim}{1}'_p\ \underset{\sim}{1}_p)' \right]^{-} (S - \bar{S}.\underset{\sim}{1}_p)'$$

Sob H_0, a estatística W_3 tem uma distribuição assintótica Qui-Quadrado com p-1 graus de liberdade, isto é,

$$W_3 \sim \chi^2_{(p-1)}$$

com a seguinte regra de decisão:

rejeitamos H_0, se $W_3 > \chi^2_{(p-1)}(1-\alpha)$, ao nível de significância α.

3.1.7 - **Caso 4**

Neste último caso consideraremos as seguintes : cumpre-se a suposição (A3) e (A4).

O modelo segundo a aditividade dos efeitos é a seguinte:

$$y_{ij} = \mu + \beta_i + \tau_j + e_{ij}$$

$$i=1,2,\ldots,n \quad e \quad j=1,2,\ldots,p$$

sujeito a:

$$\sum_{j=1}^{p} \tau_j = 0$$

onde :

y_{ij} : resposta do i-ésimo sujeito na j-ésima condição de avaliação
μ : média geral.
β_i : efeito aleatório do i-ésimo sujeito.
τ_j : efeito fixo do j-ésima condição de avaliação.
e_{ij} : erro aleatório.

Supondo que $\underset{\sim}{e}'_i = (e_{i1}, e_{i2}, \ldots, e_{ip})$, para $i=1,2,\ldots,n$ são n-vetores aleatórios p-variados independentes , com função de distribuição absolutamente contínua e simétrica em torno de zero.

Seja a hipótese nula de interesse:

$H_0 : \tau_1 = \tau_2 = \ldots = \tau_p = 0$ (não existe efeito de condição de avaliação).

Considere-se a transformação de alinhamento para testar se existem os efeitos das condições de avaliação

$z_{ij} = y_{ij} - \hat{\beta}_i$ isto é, eliminamos o efeito do fator sujeito.

onde:

$$\hat{\beta}_i = \bar{y}_{i.} - \bar{y}_{..}$$

então:

$$z_{ij} = y_{ij} - \bar{y}_{i.} + \bar{y}_{..} \qquad , i=1,2,\ldots,n \text{ e } j=1,2,\ldots,p$$

assim, formamos a matriz $\underset{\sim}{Z}$

$$\underset{n\times p}{\underset{\sim}{Z}} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & & & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{bmatrix}$$

Seja R_{ij} o posto de z_{ij} na amostra conjunta $\{z_{11}, z_{12}, \ldots, z_{np}\}$ para $i=1,2,\ldots,n$ e $j=1,2,\ldots,p$, definindo a matriz de postos $\underset{\sim}{R}$

$$\underset{n\times p}{\underset{\sim}{R}} = \begin{bmatrix} R_{11} & R_{12} & & R_{1p} \\ R_{21} & R_{22} & \cdots & R_{2p} \\ \vdots & & & \vdots \\ R_{n1} & R_{n2} & \cdots & R_{np} \end{bmatrix}$$

A função escore a(.) é definida segundo os indicadores φ_2 e φ_1 definidos na seção (1.6.2), que mede a simetria e longitude das caudas das respostas y_{ij} ,e obtemos a matriz de escores :

$$\underset{n\times p}{\underset{\sim}{a}} = \begin{bmatrix} a(R_{11}) & a(R_{12}) & \cdots & a(R_{1p}) \\ a(R_{21}) & a(R_{22}) & \cdots & a(R_{2p}) \\ \vdots & & & \vdots \\ a(R_{n1}) & a(R_{n2}) & \cdots & a(R_{np}) \end{bmatrix}$$

assim, definimos a estatística linear de postos S_j

onde: $$S_j = \frac{1}{n} \sum_{i=1}^{n} a(R_{ij}) \quad , \text{ para } j=1,2,\ldots,p.$$

Se a(.) satisfaz as condições dos teoremas de simetria (1.3.2) e normalidade (1.3.3), temos que a estatística linear de postos S_j tem distribuição assintótica Normal com parâmetros:

$$S_j \sim N\left(E(S_j)\ ,\ VAR(S_j) \right)$$

onde:

$E(S_j) = \overline{a}_{..}$

sendo:

$$\overline{a}_{..} = \frac{1}{np} \sum_{i=1}^{n} \sum_{j=1}^{p} a(R_{ij}) = \overline{S}_{.}$$

e

$$VAR(S_j) = \frac{1}{np-1} \sum_{i=1}^{n} \sum_{j=1}^{p} (a(R_{ij}) - \overline{a}_{..})^2 = \Delta^2$$

A estatística do teste é : (Kubinger : 1986)

$$W_4 = SQ(\tau)/\Delta^2$$

onde:

$$SQ(\tau) = n \sum_{j=1}^{p} (S_j - \overline{S}_{.})^2$$

e

$$\Delta^2 = \frac{1}{np-1} \sum_{i=1}^{n} \sum_{j=1}^{p} (a(R_{ij}) - \overline{a}_{..})^2$$

Sob H_0, a estatística W_4 tem distribuição assintótica Qui-Quadrado com p-1 graus de liberdade. isto é,

$$W_4 \sim \chi^2_{(p-1)}$$

A regra de decisão é a seguinte:

rejeitamos H_0, se $W_4 > \chi^2_{(p-1)}(1-\alpha)$, ao nível de significância α.

No caso de rejeitar H_0, teríamos que estudar quais condições de avaliação tem efeito significativo; para isso, trabalharemos com as comparações múltiplas.

3.1.7.1 - Comparações Múltiplas

Se para o fator fixo, a hipótese nula H_0 é rejeitada ao nínel α, estamos admitindo que pelo menos dois níveis do fator tem efeitos diferentes e para determinar os níveis que são diferentes, podemos realizar o teste:

H_0: $\tau_j = \tau_{j'}$, para todo $j \neq j' = 1, 2, \ldots, p$

H_1: $\tau_j \neq \tau_{j'}$, para algum $j \neq j' = 1, 2, \ldots, p$

a estatística do teste é dada por:

$$D_{jj'} = (S_j - S_{j'})$$

Regra de decisão:

rejeitamos a hipótese nula H_0, se

$$|D_{jj'}| > \{ \chi^2_{(p-1)}(1-\alpha/2)(VAR(D_{jj'}))\}^{1/2}$$

onde:

$$VAR(D_{jj'}) = (1/n + 1/n)\Delta^2$$

sendo:

$$\Delta^2 = \frac{1}{np-1} \sum_{i=1}^{n} \sum_{j=1}^{p} (a(R_{ij}) - \bar{a}_{..})^2$$

É importante mencionar que quando o número de observações é pequeno ($n \leq 15$), para o caso 1 e 2, deve-se utilizar a distribuição exata, e quando o número de observações é maior de 15, trabalhar com a aproximação assintotica desenvolvida nesta seção. (Meydrech : 1986).

3.2 - TESTE NÃO-PARAMÉTRICO PARA EXPERIMENTOS "SPLIT PLOT" COMPLETAMENTE CASUALIZADOS.

Entre os delimeamentos clássicos, os que utilizam unidades subdivididas ocupam um lugar proeminente. Eles são provavelmente os mais simples entre os que possuem vários níveis de erro e permitem estudar com clareza os efeitos de diversas formas de aleatorização.

Quando se quer comparar dois tratamentos, o experimentador pode atribuir um tratamento que denotaremos T e denominaremos tratamento primário, ás unidades maiores de acordo com um plano experimental particular. Cada uma dessas unidades, que chamaremos de parcela, pode ser subdividida, por sua vez, em unidades menores ou sub-parcelas, aplicando-se a elas o outro tratamento, que denominaremos tratamento secundário e denotaremos por B, de acordo com um segundo plano específico.
O delineamento resultante é dito em parcelas subdivididas ou em "SPLIT PLOT" e seu princípio de construção tem sido utilizado para comparar dois ou mais fatores.

<u>Exemplo:</u>

Um experimento de preparação de biscoitos de chocolate, é feito em três recipientes para a preparação da massa, onde cada recipiente tem uma caracteristica diferente; deseja-se comparar em temperaturas diferentes a textura dos biscoitos para o qual mede-se o número de biscoitos que não tem boa qualidade.

Tabela 3.2 - ESTRUTURA DE DADOS BÁSICA PARA ESTUDO COM PLANEJAMENTO EM ESTRUTURA "SPLIT PLOT "

sub-população ou tratamento primário	unidades experimentais	Tratamento Secundário				
		B_1	B_2	B_3		B_p
T_1	1	y_{111}	y_{112}	y_{113}	. . .	y_{11p}
	2	y_{121}	y_{122}	y_{123}	. . .	y_{12p}
	⋮	⋮				⋮
	n	y_{1n1}	y_{1n2}	y_{1n3}	. . .	y_{1np}
T_2	1	y_{211}	y_{212}	y_{213}	. . .	y_{21p}
	2	y_{221}	y_{222}	y_{223}	. . .	y_{22p}
	⋮	⋮				⋮
	n	y_{2n1}	y_{2n2}	y_{2n3}	. . .	y_{2np}
⋮	⋮	⋮				⋮
T_c	1	y_{c11}	y_{c12}	y_{c13}	. . .	y_{c1p}
	2	y_{c21}	y_{c22}	y_{c23}	. . .	y_{c2p}
	⋮	⋮				⋮
	n	y_{cn1}	y_{cn2}	y_{cn3}	. . .	y_{cnp}

Considere-se uma situação experimental envolvendo N unidades experimentais, asinaladas ao acaso a **c** tratamentos primários e respondem a cada um dos **p** tratamentos secundários definidas previamente.

Seja:

Y_{ijk}: resposta da j-ésima unidade experimental no i-ésimo tratamento primário do k-ésimo tratamento secundário.

$i = 1,2,\ldots,c$	nível dos tratamentos primários
$j = 1,2,\ldots,n$	unidades experimentais
$k = 1,2,\ldots,p$	nível dos tratamentos secundários
$N = cn$	

Definimos a matriz $\underset{\sim}{Y} = ((\, y_{ij} \,))$, formada de N vetores linha estocasticamente independentes.

$y'_{ij} = (\, y_{ij1} \,,\, y_{ij2} \,,\ldots, y_{ijp} \,)$, que é o vetor de respostas do j-ésimo indivíduo ao i-ésimo tratamento primário.

Supõe-se que y_{ij} tem distribuição absolutamente contínua p-variada com função distribuição acumulada $F_i(y)$,

onde :

$$F_i(y) = G\,(y-\mu_i),$$

sendo:

$\mu'_i = (\, \mu_{i1} \,,\, \mu_{i2} \,,\, \ldots \,,\, \mu_{ip} \,)$, vetores do parâmetros de médias e são indicativos de locação.

3.2.1 - Suposições

Com a finalidade de fazer uma análise estatística deste experimento aplicando teoria Não-Paramétrica, presupõe-se:

B1).- Não-interação entre tratamentos primários e tratamentos secundários, isto significa, que cada tratamento primário tem o mesmo efeito independente do tratamento secundário.

B2).-Estrutura Uniforme na matriz de covariância, dada pela seguinte forma:

$$\Sigma = \rho\sigma^2\mathbf{J} + (1-\rho)\sigma^2\mathbf{I} , \qquad (3.2.2.1)$$

sendo $\mathbf{J}_{pxp}$ uma matriz de 1's e $\mathbf{I}_{pxp}$ uma matriz identidade.

3.2.2 - Hipóteses de Interesse

Os principais objetivos são:

i) verificar a existência ou não do efeito de diferentes tratamentos primários;

ii) verificar aexistência ou não do efeito de diferentes tratamentos secundários;

iii) verificar a presença da interação entre tratamento primário e tratamento secundário.

Sabemos que y_{ij} tem distribuição absolutamente contínua p-variada com função distribuição acumulada $F_i(y)$,

onde :

$F_i(y) = G\ (y-\mu_i)$,

$\mu_i' = (\mu_{i1} , \mu_{i2} , \ldots , \mu_{ip})$ vetor de parâmetros de posição.

Assumindo que y_{ijk} tem a siguente estrutura aditiva :

$$y_{ijk} = \mu + \tau_i + \gamma_k + \tau\gamma_{ik} + e_{ijk}$$

$i=1,2,\ldots,c$, $k=1,2,\ldots,p$, $j=1,2,\ldots n$.

onde:

μ : media geral

τ_i :efeito do i-ésimo tratamento primário.

γ_k : efeito da k-ésimo tratamento secundário.

$\tau\gamma_{ik}$: efeito de interação do i-ésimo tratamento primário com o efeito do k-ésimo tratamerrnto secundário.

e_{ijk} : o erro aleatório, com $E(e_{ijk}) = 0$.

sujeito a:

$$\sum_{i=1}^{c} \tau_i = \sum_{k=1}^{p} \gamma_k = 0$$

e

$$\sum_{i=1}^{c} \tau\gamma_{ik} = \sum_{k=1}^{p} \tau\gamma_{ik} = 0$$

a) <u>Teste de Não-interação.-</u>

A validade da suposição de não-interação entre tratamento primário e tratamento secundário pode-se testar da seguinte maneira:

H_{01} : $\tau\gamma_{11} = \ldots = \tau\gamma_{cp} = 0$.

Para testar esta hipótese efetuamos a seguinte transformação de alinhamento, com a finalidade de eliminar os efeitos que não serão testados nesta hipótese :

$$z_{ijk} = y_{ijk} - \overline{y}_{i..} - \overline{y}_{..k} + 2\overline{y}_{...}$$

onde :

$$\overline{y}_{..k} = \frac{1}{cn} \sum_{j=1}^{n} \sum_{i=1}^{c} y_{ijk}$$

$$\overline{y}_{i..} = \frac{1}{np} \sum_{k=1}^{p} \sum_{j=1}^{n} y_{ijk}$$

$$\overline{y}_{...} = \frac{1}{npc} \sum_{i=1}^{c} \sum_{j=1}^{n} \sum_{k=1}^{k} y_{ijk}$$

Seja R_{ijk} o posto de z_{ijk} na amostra conjunta $\{z_{111}, z_{112}, \ldots, z_{11p}, \ldots, z_{1n1}, \ldots, z_{1np}, \ldots, z_{cn1}, \ldots, z_{cnp}\}$.

Os escores a(.) são escolhidos de acordo com as funções seletoras φ_2 e φ_1 obtidas das observações y_{ijk}, definidas na seção (1.6.2).

A estatística do teste é :

$$SQ_{\tau\gamma} = \frac{Q_{\tau\gamma}}{\Delta^2}$$

onde:

$$SQ_{\tau\gamma} = n \sum_{i=1}^{c} \sum_{k=1}^{p} (\overline{S}_{i.k} - \overline{S}_{i..} - \overline{S}_{..k} + \overline{S}_{...})^2$$

$$\sigma^2 = \frac{1}{ncp-1} \sum_{i=1}^{c} \sum_{j=1}^{n} \sum_{k=1}^{p} (a(R_{ijk}) - \overline{a}_{...})^2$$

$$\overline{a}_{...} = \frac{1}{ncp} \sum_{i=1}^{c} \sum_{j=1}^{n} \sum_{k=1}^{p} a(R_{ijk}) = \overline{S}_{...}$$

$$\overline{S}_{i..} = \frac{1}{np} \sum_{k=1}^{p} \sum_{j=1}^{n} a(R_{ijk}) \qquad i=1,2, \ldots, c$$

$$\overline{S}_{..k} = \frac{1}{nc} \sum_{i=1}^{c} \sum_{j=1}^{n} a(R_{ijk}) \qquad k=1, 2, \ldots, p$$

$$\overline{S}_{i.k} = \frac{1}{n} \sum_{j=1}^{n} a(R_{ijk})$$

Sob H_{01}, a estatística $Q_{\tau\gamma}$ tem distribuição assintótica Qui-Quadrado com $(c-1)(p-1)$ graus de liberdade, isto é,

$$Q_{\tau\gamma} \sim \chi^2_{(c-1)(p-1)}$$

<u>Regra de decisão:</u>

A hipótese nula H_{01}, é rejeitada ao nível de significância α, se:

$$Q_{\tau\gamma} > \chi^2_{(c-1)(p-1)}(1-\alpha)$$

No caso de confirmar a suposição (B1) de não-interação entre tratamento primário e tratamento secundário, pasaremos a testar as seguintes hipóteses :

Na segunda hipótese de interesse desejamos testar se não existe efeito de tratamento primário, dada por:

$$H_{02} : \tau_1 = \tau_2 = \ldots = \tau_c = 0$$

Na tercera hipótese de interesse desejamos testar se não existe efeito do tratamento secundário, dada por:

$$H_{03} : \gamma_1 = \gamma_2 = \ldots = \gamma_p = 0$$

b) Para testar H_{02} utilizaremos a seguinte transformação de alinhamento, com a finalidade de eliminar os efeitos do tratamento secundário, que não serão testados nesta hipótese:

$$z_{ijk} = y_{ijk} - \bar{y}_{i.k} + \bar{y}_{i..}$$

onde:

$$\bar{y}_{i.k} = \frac{1}{n} \sum_{j=1}^{n} y_{ijk}$$

$$\bar{y}_{i..} = \frac{1}{np} \sum_{k=1}^{p} \sum_{j=1}^{n} y_{ijk}$$

Seja R_{ijk} o posto de z_{ijk} na amostra conjunta $\{z_{111}, z_{112}, \ldots, z_{11p}, \ldots, z_{in1}, \ldots, z_{inp}, \ldots, z_{cn1}, \ldots, z_{cnp}\}$.
Os escores a(.) são escolhidos de acordo com as funções seletoras φ_2 e φ_1 obtidas das observações y_{ijk}, definida na seção (1.6.2).

Definimos a estatística linear de posto da seguinte forma:

$$\overline{S}_{i..} = \frac{1}{np} \sum_{j=1}^{n} \sum_{k=1}^{p} a(R_{ijk}) \;, \quad i=1,2,\ldots,c$$

Se a(.) satisfaz a condições dos teoremas de simetria (1.3.2) e normalidade (1.3.3), a estatística linear de postos tem distribuição assintótica Normal com parâmetros:

$$\overline{S}_{i..} \;\hat{\sim}\; N\left(E(\overline{S}_{i..}) \;,\; VAR(\overline{S}_{i..}) \right)$$

onde:

$E(\overline{S}_{i.}) = \overline{a}_{...}$

sendo:

$$\overline{a}_{...} = \frac{1}{ncp} \sum_{i=1}^{c} \sum_{j=1}^{n} \sum_{k=1}^{p} a(R_{ik}) = \overline{S}_{...}$$

e

$$VA(\overline{S}_{i..}) = \frac{1}{ncp-1} \sum_{i=1}^{c} \sum_{j=1}^{n} \sum_{k=1}^{p} \left(a(R_{ijk}) - \overline{a}_{...} \right)^2 = \Delta^2$$

A estatística do teste é dada por:

$$Q_\tau = \frac{SQ_\tau}{\Delta^2}$$

onde:

$$SQ_\tau = np \sum_{i=1}^{c} (\bar{S}_{i..} - \bar{S}_{...})^2$$

$$\Delta^2 = \frac{1}{ncp-1} \sum_{i=1}^{c} \sum_{j=1}^{n} \sum_{k=1}^{p} (a(R_{ijk}) - \bar{a}_{...})^2$$

sob H_{02}, a estatística Q_τ tem distribuição assintótica Qui-Quadrado com c-1 graus de liberdade, isto é,

$$Q_\tau \sim \chi^2_{(c-1)}$$

Regra de decisão:

A hipótese H_{02}, é rejeitada ao nível de significância α, se:

$$Q_\tau > \chi^2_{(c-1)}(1-\alpha)$$

c) Para testar $H_{03} : \gamma_1 = \gamma_2 = \ldots = \gamma_p = 0$ (não existe efeito de tratamento secundário) utilizaremos a seguinte transformação de alinhamento, com a finalidade de eliminar os efeitos primários que não serão testados nesta hipótese:

$$z_{ijk} = y_{ijk} - \overline{y}_{i.k} + \overline{y}_{..k}$$

sendo :

$$\overline{y}_{i.k} = \frac{1}{n} \sum_{j=1}^{n} y_{ijk}$$

$$\overline{y}_{..k} = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} y_{ijk}$$

Seja R_{ijk} o posto de z_{ijk} na amostra conjunta $\{z_{111}, z_{112}, \ldots, z_{11p}, \ldots, z_{1n_1 1}, \ldots, z_{1n_1 p}, \ldots, z_{cn_c 1}, \ldots, z_{cn_c p}\}$

Os escores a(.) são escolhidos de acordo com as funções seletora φ_2 e φ_1 obtidas das observações y_{ijk}, definidas na seção (1.6.2).

A estatística linear de postos é definida da seguinte forma :

$$\overline{S}_{..k} = \frac{1}{nc} \sum_{i=1}^{c} \sum_{j=1}^{n} a(R_{ijk}) \qquad , \text{ para } k=1,2,\ldots,p$$

se a(.) satisfaz a condições do teoremas de simetria (1.3.2) e normalidade (1.3.3), a estatística linear de postos tem distribuição assintótica Normal com parâmetros:

$$\overline{S}_{..k} \sim N\left(E(\overline{S}_{..k}) \, , \, VAR(\overline{S}_{..k}) \right)$$

onde:

$$E(\overline{S}_{..k}) = \overline{a}_{...}$$

sendo:

$$\bar{a}_{...} = \frac{1}{ncp} \sum_{i=1}^{c} \sum_{j=1}^{n} \sum_{k=1}^{p} a(R_{ijk}) = \bar{S}_{...}$$

e

$$VAR(\bar{S}_{..k}) = \frac{1}{ncp-1} \sum_{k=1}^{p} \sum_{j=1}^{n} \sum_{i=1}^{c} (a(R_{ijk}) - \bar{a}_{...})^2 = \Delta^2$$

Assim, a estatística do teste é dada por:

$$Q_\gamma = \frac{SQ_\gamma}{\Delta^2}$$

onde:

$$SQ_\gamma = nc \sum_{k=1}^{p} (\bar{S}_{..k} - \bar{S}_{...})^2$$

Sob H_{03}, a estatística Q_γ tem distribuição assintótica Qui-Quadrado com p-1 graus de liberdade, isto é,

$$Q_\gamma \sim \chi^2_{(p-1)}$$

<u>Regra de decisão:</u>

A hipótese nula H_{03}, é rejeitada ao nível de significância α, se:

$$Q_\gamma > \chi^2_{(p-1)}(1-\alpha)$$

- Comparações Múltiplas

Se para cada fator , a hipótese nula H_0 é rejeitada ao nível de significância α, estamos admitindo que pelo menos dois níveis do fator tem efeitos diferentes e, para determinar os níveis que são diferentes, podemos realizar os seguintes testes:

a1) Tratamento Primário:

H_{0i} : $\tau_i = \tau_{i'}$, para $i \neq i' = 1, 2, \ldots, c$

versus

A_{1i} : $\tau_i \neq \tau_{i'}$ para algum $i \neq i'$

Regra de decisão:

rejeitamos H_{0i} , se:

$$| \bar{S}_{i..} - \bar{S}_{i'..} | > [\ (2/np)\Delta^2\ \chi^2_{(c-1)}(1-\alpha/2)\ ]^{1/2}$$

b1) Tratamento Secundário:

H_{0k} : $\gamma_k = \gamma_{k'}$, para $k \neq k' = 1, 2, \ldots, p$

versus

A_{1k} : $\gamma_k \neq \gamma_{k'}$ para algum $k \neq k'$

Regra de decisão:

rejeitamos H_{0k} , se:

$$| \overline{S}_{..k} - \overline{S}_{..k'} | > [(2/nc)\Delta^2 \chi^2_{(p-1)} (1-\alpha/2)]^{1/2}$$

c1) <u>Interação entre Tratamento primário e Tratamento secundário</u>.

H_{0ik}: $\tau\gamma_{ik} = \tau\gamma_{lg}$, para $i,l=1,2,...,c$ e $k,g=1,2,...,p$

versus

A_{1ik}: $\tau\gamma_{ik} \neq \tau\gamma_{lg}$ para algum $i\neq l$ ou $k\neq g$

Regra de decisão:

rejeitamos H_{0ik}, se:

$$| \overline{S}_{i.k} - \overline{S}_{l.g} | > [(2/n) \Delta^2 \chi^2_{(c-1)(p-1)} (1-\alpha/2)]^{1/2}$$

Como podemos observar, para cada classe de hipótese existem diferentes transformações e conseqüentemente diferentes Δ^2, $\overline{S}_{i..}$, $\overline{S}_{..k}$, $\overline{S}_{i.k}$, $\overline{S}_{...}$, etc.

3.3 - TESTE NÃO-PARAMÉTRICO PARA MODELOS "SPLIT PLOT" EM EXPERIMENTOS COM MEDIDAS REPETIDAS.

Tabela 3.3 - ESTRUTURA DE DADOS BÁSICA PARA ESTUDO DO MODELOS "SPLIT PLOT" COM MEDIDAS REPETIDAS

sub-população ou tratamento primário	unidades experimentais	condição de avaliação				
		B_1	B_2	B_3		B_p
T_1	1	y_{111}	y_{112}	y_{113}	. . .	y_{11p}
	2	y_{121}	y_{122}	y_{123}	. . .	y_{12p}
	⋮	⋮				⋮
	n_1	y_{1n_11}	y_{1n_12}	y_{1n_13}	. . .	y_{1n_1p}
T_2	1	y_{211}	y_{212}	y_{213}	. . .	y_{21p}
	2	y_{221}	y_{222}	y_{223}	. . .	y_{22p}
	⋮	⋮				⋮
	n_2	y_{2n_21}	y_{2n_22}	y_{2n_23}	. . .	y_{2n_2p}
⋮	⋮	⋮				⋮
T_c	1	y_{c11}	y_{c12}	y_{c13}	. . .	y_{c1p}
	2	y_{c21}	y_{c22}	y_{c23}	. . .	y_{c2p}
	⋮	⋮				⋮
	n_c	y_{cn_c1}	y_{cn_c2}	y_{cn_c3}	. . .	y_{cn_cp}

Considere-se uma situação experimental envolvendo N unidades experimentais, asinaladas ao acaso a **c** tatamentos primários e respondem a cada um das **p** condições de avaliação definidas previamente.

Seja:

Y_{ijk}: resposta da j-ésima unidade experimental no i-ésimo tratamento primário da k-ésima condição de avaliação.

$i = 1, 2, \ldots, c$ nível dos tratamentos primários

$j = 1, 2, \ldots, n_i$ unidades experimentais

$k = 1, 2, \ldots, p$ nível das condições de avaliação.

$$N = \sum_{i=1}^{c} n_i$$

Definimos a matriz $\underset{\sim}{Y} = ((\ y_{ij}\ ))$, composta de N vetores linha estocasticamente independentes.

$y_{ij}' = (\ y_{ij1}\ ,\ y_{ij2}\ , \ldots, y_{ijp}\ )$, que é o vetor de respostas do j-ésimo indivíduo ao i-ésimo tratamento primário.

Supõe-se que y_{ij} tem distribuição absolutamente contínua p-variada com função distribuição acumulada $F_i(y)$,

onde :

$$F_i(y) = G\ (y - \mu_i)$$

sendo:

$\mu_i' = (\ \mu_{i1}\ ,\ \mu_{i2}\ ,\ \ldots\ ,\ \mu_{ip}\ )$, vetores do parâmetros de médias e são indicativos de locação.

3.3.1 - Suposições

Com a finalidade de fazer uma análise estatística deste experimento aplicando teoria Não-Paramétrica, presupõe-se:

B1).- Não-interação entre tratamentos primários e condições de avaliação.

B2).- Estrutura Uniforme na matriz de covariância definida em (3.2.2.1).

Tendo em conta ou não as suposições (B1) e (B2) , teremos três situações para testar: se existe efeito de tratamento primário, e se existe efeitos das condições de avaliação.

3.3.2 - **CASO 1**

Ante a necessidade de efetuar os testes das hipóteses nulas de interesse, neste primeiro caso, não assumiremos as suposições (B1) e (B2), isto é:

1.- Não é conhecida a não-interação entre tratamento primário e condição de avaliação.

2.- A matriz de covariância Σ tem estrutura geral.

Segundo estas suposições, desenvolvemos os seguintes procedimentos para testar:

a)- Primera hipótese nula de interesse:

H_{01}: $\mu_1 = \mu_2 = \ldots = \mu_c$ (não existe efeito dos tratamentos primários).

Para cada condição de avaliação da matriz $\underset{\sim}{Y}$, determinamos seu posto independente das demais, isto é,

seja R_{ijk} o posto de y_{ijk} na amostra conjunta

$\{y_{11k}, y_{12k}, \ldots, y_{1n_1k}, y_{21k}, \ldots, y_{2n_2k}, \ldots, y_{c1k}, \ldots, y_{cn_ck}\}$

desta maneira, formamos a matriz de postos $\underset{\sim}{R}$

onde:

$$\underset{\sim}{R}_{N\times p} = \begin{bmatrix} R_{111} & R_{112} & R_{113} & \ldots & R_{11p} \\ \vdots & \vdots & \vdots & & \vdots \\ R_{1n_11} & R_{1n_12} & R_{1n_13} & \ldots & R_{1n_1p} \\ R_{211} & R_{212} & R_{213} & \ldots & R_{21p} \\ \vdots & \vdots & \vdots & & \vdots \\ R_{1n_21} & R_{1n_22} & R_{1n_23} & \ldots & R_{1n_2p} \\ \vdots & \vdots & \vdots & & \vdots \\ R_{cn_c1} & R_{cn_c2} & R_{cn_c3} & \ldots & R_{cn_cp} \end{bmatrix}$$

de acordo com as funções seletoras φ_2e φ_1, definidas na seção (1.6.2) que nos indica a simetria e longitude das caudas dos dados originais, determinamos a função escore a(.), e construimos a matriz $\underset{\sim}{a}$

$$\underset{N\times p}{\tilde{a}} = \begin{bmatrix} a(R_{111}) & a(R_{112}) & a(R_{113}) & \ldots a(R_{11p}) \\ \vdots & \vdots & \vdots & \vdots \\ a(R_{1n_1 1}) & a(R_{1n_1 2}) & a(R_{1n_1 3}) & \ldots a(R_{1n_1 p}) \\ a(R_{211}) & a(R_{212}) & a(R_{213}) & \ldots a(R_{21p}) \\ \vdots & \vdots & \vdots & \vdots \\ a(R_{1n_2 1}) & a(R_{1n_2 2}) & a(R_{1n_2 3}) & \ldots a(R_{1n_2 p}) \\ \vdots & \vdots & \vdots & \vdots \\ a(R_{cn_c 1}) & a(R_{cn_c 2}) & a(R_{cn_c 3}) & \ldots a(R_{cn_c p}) \end{bmatrix}$$

desta forma, determinamos a estatística linear de postos $\overline{S}_{i.k}$:

$$\overline{S}_{i.k} = \frac{1}{n_i} \sum_{j=1}^{n_i} a(R_{ijk})$$

$$i = 1,2,\ldots,c$$

$$k = 1,2,\ldots,p$$

assim, definimos o vetor $\overline{S}_i$ como:

$$\overline{S}_i' = (\overline{S}_{i.1}, \overline{S}_{i.2}, \ldots, \overline{S}_{i.p}) , \quad \text{para } i=1,2,\ldots,c$$

Para cada condição de avaliação obtemos a média aritmética da estatística

$$\overline{S}_{..k} = \frac{1}{c} \sum_{i=1}^{c} \overline{S}_{i.k}$$

Se $n = n_i$, então:

$$\overline{S}_{..k} = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n} a(R_{ijk})$$

observando que $\overline{S}_{..k} = \overline{a}_{..k}$

onde :

$$N = \sum_{i=1}^{c} n_i$$

Se a(.) satisfaz as condições dos teoremas de simetria (1.3.2) e normalidade (1.3.3), temos que a estatística linear de postos $\overline{S}_{i.k}$ tem distribuição assintótica Normal com parâmetros:

$$\overline{S}_{i.k} \sim N\left(E(\overline{S}_{i.k}) \ , \ VAR(\overline{S}_{i.k}) \right)$$

onde :

$E(\overline{S}_{i.k}) = \overline{S}_{..k}$

e

$COV(\overline{S}_{i.k} \ , \ \overline{S}_{i'.k'}) =$

$$= \frac{(N(\delta_{ii'}) - n)}{n(N-1)} \ \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n} (a(R_{ijk}) - \overline{a}_{..k})(a(R_{i'jk'}) - \overline{a}_{..k'})$$

para $n = n_i = n_{i'}$, onde $i,i' = 1,2,\ldots,c$

$k,k' = 1,2,\ldots,p$

$$COV(\overline{S}_{i.k} \ , \ \overline{S}_{i'.k'}) = \frac{(N(\delta_{ii'}) - n)}{n(N-1)} \ v_{kk'}$$

onde:

$$v_{kk'} = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n} (a(R_{ijk}) - \overline{a}_{..k})(a(R_{i'jk'}) - \overline{a}_{..k'})$$

$$\delta_{ii'} = \begin{cases} 1 & , \text{ se } i=i' \\ 0 & , \text{ se } i \neq i' \end{cases}$$

Seja a matriz de covariância $\underset{p \times p}{\underset{\sim}{V}} = ((v_{kk'}))$, $k,k'=1,2,\ldots,p$ considerando para obtenção da estatística do teste , $\underset{\sim}{V}$ positiva definida.

Deste modo , a estatística do teste é dada pela forma quadrática (Koch:1969):

$$Q_1 = \frac{N-1}{N} \sum_{i=1}^{c} n_i \left(\overline{S}_i - \overline{S} \right)' V^{-1} \left(\overline{S}_i - \overline{S} \right)$$

onde:

$$\overline{S}_i' = (\overline{S}_{i.1}, \overline{S}_{i.2}, \ldots, \overline{S}_{i.p}) \qquad , \qquad i = 1,2,\ldots,c$$

$$\overline{S}' = (\overline{S}_{..1}, \overline{S}_{..2}, \ldots, \overline{S}_{..p})$$

Sobre H_{01} , a estatística Q_1 tem distribuição assintótica Qui-Quadrado com p(c-1) graus de liberdade , isto é,

$$Q_1 \sim \chi^2_{p(c-1)}$$

Regra de decisão:

Se $Q_1 > \chi^2_{p(c-1)}(1-\alpha)$, então, rejeitamos a hipótese nula H_{01}, ao nível de significância α.

Se a matriz V é singular então, utilizaremos a inversa geralizada V^- e se o posto de V é p'.

então, $Q_1 \sim \chi^2_{p'(c-1)}$

b) - <u>Segunda hipótese nula de interesse</u>

$H_{02} = \mu_{i1} = \mu_{i2} = \ldots = \mu_{ip}$, para $i=1,2,\ldots,c$ (não existe efeito de condição de avaliação).

A matriz de dados pode ser vista como um conjunto de modelos mistos. Cada um dos quais corresponde a um particular tratamento (Whole plot).

Supõe-se simetria das diagonais:

seja $\underset{(p-1)\times 1}{\underset{\sim}{U}_{ij}} = \underset{(p-1)\times p}{\underset{\sim}{C}} \underset{p\times 1}{Y_{ij}}$, um contraste,

onde:

$Y'_{ij} = (y_{ij1}, y_{ij2}, \ldots, y_{ijp})$ $\qquad i=1,2,\ldots,c$

$\qquad j=1,2,\ldots,n_i$

C é a matriz de contrastes linearmente independente por linha, isto é,

$$\underset{(p-1)\times p}{\underset{\sim}{C}} \times \underset{\sim}{1}_p = \underset{(p-1)\times 1}{\underset{\sim}{0}}$$

sendo:

$\underset{\sim}{1}'_p = (1,1,\ldots,1)$, vetor de 1's px1

e $\underset{\sim}{0}$ (vetor nulo p-1 x 1)

se temos outro contraste da forma:

$$\underset{(p-1)\times 1}{\underset{\sim}{\theta}_i} = \underset{\sim}{C}\, \mu_i \quad , \text{ com } \quad \mu_i' = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{ip})$$

então:

o vetor ($\underset{\sim}{U}_{ij}$ - $\underset{\sim}{\theta}_i$) e ($\underset{\sim}{\theta}_i$ - $\underset{\sim}{U}_{ij}$) tem a mesma distribuição.

sob H_{02}, $\underset{\sim}{\theta}_i = \underset{\sim}{0}$

desta maneira:

$\underset{\sim}{U}_{ij}$ e (-1) $\underset{\sim}{U}_{ij}$ tem a mesma distribuição para $i=1,2,\ldots,c$
$j=1,2,\ldots,n_i$

isto significa que a distribuição é simétrica em torno de zero.

então, testar:

$$H_{02}: \mu_{i1} = \mu_{i2} = \ldots = \mu_{ip}$$

equivale a testar:

$$H_{02^*} : \underset{\sim}{\theta}_i = \underset{\sim}{0} .$$

Como $Y_{ij} \sim F_i(y) = G\,(y-\mu_i)$

onde $\mu'_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{ip})$.

Para testar: $H_{02^*} : \underset{\sim}{\theta}_i = \underset{\sim}{0}$

procedemos da seguinte maneira :

seja R_{ijk} o posto de y_{ijk} na amostra conjunta $\{ y_{ij1}, y_{ij2}, \ldots, y_{ijp} \}$, isto é, os postos são dados por unidade experimental em diferentes condicões de avaliação.

para $i=1,2,\ldots,c$ $\qquad N = \sum_{i=1}^{c} n_i$
$j=1,2,\ldots,n_i$
$k=1,2,\ldots,p$

obtendo-se uma matriz de postos $\underset{\sim}{R}$

$$\underset{N\times p}{\underset{\sim}{R}} = \begin{bmatrix} R_{111} & R_{112} & R_{113} & \cdots & R_{11p} \\ \vdots & \vdots & \vdots & & \vdots \\ R_{1n_1 1} & R_{1n_1 2} & R_{1n_1 3} & \cdots & R_{1n_1 p} \\ R_{211} & R_{212} & R_{213} & \cdots & R_{21p} \\ \vdots & \vdots & \vdots & & \vdots \\ R_{1n_2 1} & R_{1n_2 2} & R_{1n_2 3} & \cdots & R_{1n_2 p} \\ \vdots & \vdots & \vdots & & \vdots \\ R_{cn_c 1} & R_{cn_c 2} & R_{cn_c 3} & \cdots & R_{cn_c p} \end{bmatrix}$$

De acordo com as funções seletoras φ_2 e φ_1 definidas na seção (1.6.2), determinamos a função escore a(.), assim, construimos a matriz $\underset{\sim}{a}$

$$\underset{N\times p}{\underset{\sim}{a}} = \begin{bmatrix} a(R_{111}) & a(R_{112}) & a(R_{113}) & \cdots a(R_{11p}) \\ \vdots & \vdots & \vdots & \vdots \\ a(R_{1n_1 1}) & a(R_{1n_1 2}) & a(R_{1n_1 3}) & \cdots a(R_{1n_1 p}) \\ a(R_{211}) & a(R_{212}) & a(R_{213}) & \cdots a(R_{21p}) \\ \vdots & \vdots & \vdots & \vdots \\ a(R_{1n_2 1}) & a(R_{1n_2 2}) & a(R_{1n_2 3}) & \cdots a(R_{1n_2 p}) \\ \vdots & \vdots & \vdots & \vdots \\ a(R_{cn_c 1}) & a(R_{cn_c 2}) & a(R_{cn_c 3}) & \cdots a(R_{cn_c p}) \end{bmatrix}$$

onde $N = \sum_{i=1}^{c} n_i$

A estatística linear de postos é definida com :

$$\overline{S}_{..k} = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n_i} a(R_{ijk}) \qquad , \text{ onde } N = \sum_{i=1}^{c} n_i$$

$$k=1,2,\ldots,p$$

Se a(.) satisfaz a condições dos teoremas de simetria (1.3.2) e normalidade(1.3.3), temos que a estatística linear de postos tem distribuição assintótica Normal com parâmetros:

$$\overline{S}_{..k} \sim N\left(E(\overline{S}_{..k}),\ VAR(\overline{S}_{..k}) \right)$$

onde :

$$E(\overline{S}_{..k}) = \overline{a}_{...}$$

sendo, $$\overline{a}_{...} = \frac{1}{Np} \sum_{i=1}^{c} \sum_{j=1}^{n_i} \sum_{k=1}^{p} a(R_{ijk}) = \overline{S}_{...}$$

e

$$COV(\overline{S}_{..k}, \overline{S}_{..k'}) = \frac{1}{N^2} \sum_{i=1}^{c} \sum_{j=1}^{n_i} (a(R_{ijk}) - \overline{a}_{...})(a(R_{ijk'}) - \overline{a}_{...})$$

$$COV(\overline{S}_{..k}, \overline{S}_{..k'}) = v_{kk'}$$

onde:

$$v_{kk'} = \frac{1}{N^2} \sum_{i=1}^{c} \sum_{j=1}^{n_i} (a(R_{ijk}) - \overline{a}_{...})(a(R_{ijk'}) - \overline{a}_{...})$$

definimos a matriz de covariância $V = ((v_{kk'}))$ para $k,k' = 1,2,..p$.

A estatística do teste é forma : (Koch: 1969)

$$Q_2 = \underset{\sim}{S}'\underset{\sim}{C}'[\underset{\sim}{C}\ V\ \underset{\sim}{C}']^{-1}\ \underset{\sim}{C}\ S$$

onde:

$S' = (\overline{S}_{..1}, \overline{S}_{..2}, \ldots, \overline{S}_{..p})$.

$\underset{\sim}{C}$ matriz de contrastes linearmente independente por linha.

Sob $H_{02}*$, a estatística Q_2 tem distribuição assintótica Qui-Quadrado com p-1 graus de libedade, isto é,

$$Q_2 \sim \chi^2_{(p-1)}$$

<u>Regra de decisão:</u>

se $Q_2 > \chi^2_{(p-1)}(1-\alpha)$, então, a hipótese $H_{02}*$, é rejeitada ao nível de significância α.

b.1) <u>Comparações Múltiplas</u>

No caso de rejeitar H_{02}, teríamos que estudar qual condição de avaliação tem efeito significâtivo, para os quais, trabalharemos com as comparações múltiplas, utilizando a matriz de contrastes C, com posto de C igual a p-1 .

Exemplo: Testar:

$H_{02}: \mu_{i1} = \mu_{i2} = \mu_{i3p}$

é equivalente a testar :

$H_{02}: \mu_{i1} - \mu_{i2} = 0$ e $\mu_{i1} - \mu_{i3} = 0.$

utilizando a matriz de constrastes

$$\underset{2\times p}{C} = \begin{bmatrix} 1 & -1 & 0 & 0 & \ldots 0 \\ 1 & 0 & -1 & 0 & \ldots 0 \end{bmatrix}$$

temos:

$$H_{02}: \quad \begin{bmatrix} 1 & -1 & 0 & 0 & \ldots 0 \\ 1 & 0 & -1 & 0 & \ldots 0 \end{bmatrix} \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{ip} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

e utilizamos a estatística Q_2 para testar H_{02}.

3.3.3 - CASO 2

Neste segundo caso consideraremos que a suposições (B1) e (B2), como verdadeiras, isto é:

-É conhecida a não-interação entre tratamento e condição de avaliação.

- A matriz de covariância Σ, tem estrutura uniforme.

Então, para resolver este segundo caso aplicamos a análise da seção (3.2), pois com estas duas suposições teríamos um experimento analizado como "SPLIT PLOT".

3.3.4.- CASO 3

Neste terceiro caso temos as seguintes considerações:

1).- É conhecida a não-interação entre tratamento e condição de avaliação.

2).- A matriz de covariância Σ, tem estrutura geral.

Segundo estas condições, desenvolvemos os seguintes procedimentos para testar:

a) Primeira hipótese nula de interesse :

H_{01}: $\mu_1 = \mu_2 = \dots = \mu_c$ (não existe efeito dos tratamentos primários)

como se viu na seção (3.2), presupõe-se:

y_{ijk} tem a seguinte estrutura aditiva:

$$y_{ijk} = \mu + \tau_i + \gamma_k + e_{ijk}$$

$$i=1,2,\dots,c, \quad j=1,2,\dots,n_i \quad k=1,2,\dots,p$$

supondo $n_i=n$, para todo $i=1,2,\dots,c$ (tamanhos iguais)

sujeito a:

$$\sum_{i=1}^{c} \tau_i = \sum_{k=1}^{p} \gamma_k = 0$$

onde :

μ : media geral;

τ_i :efeito do i-ésimo tratamento primário;

γ_k : efeito da k-ésima condição de avaliação;

e_{ijk}: erro aleatório, com $E(e_{ijk}) = 0$.

Podemos rescrever a hipótese nula da seguinte maneira:

$H_{01^*} : \tau_1 = \tau_2 = \dots = \tau_c = 0$

Para testar H_{01^*} utilizaremos a seguinte transformação de alinhamento com a finalidade de eliminar os efeitos das condições de avaliação que não serão testados nesta hipótese:

$$z_{ijk} = y_{ijk} - \overline{y}_{i.k} + \overline{y}_{i..}$$

onde: $\overline{y}_{i.k} = \dfrac{1}{n} \sum_{j=1}^{n} y_{ijk}$

$$\overline{y}_{i..} = \frac{1}{np} \sum_{k=1}^{p} \sum_{j=1}^{n} y_{ijk}$$

Seja R_{ik} é o posto de z_{ijk} na amostra conjunta $\{z_{111}, \dots, z_{11p}, \dots, z_{in_1 1}, \dots, z_{1n_1 p}, \dots, z_{cn_1 1}, \dots, z_{cn_c p}\}$

Os escores a(.) são escolhidos de acordo com a funções seletoras φ_2 e φ_1 obtidas das observações y_{ijk}, definida na seção (1.6.2).

Definimos a estatística linear de posto da seguinte forma:

$$\overline{S}_{i..} = \frac{1}{np} \sum_{j=1}^{n} \sum_{k=1}^{p} a(R_{ijk}) \quad , \qquad i=1,2,\dots,c$$

Se a(.) satisfaz a condições dos teoremas de simetria (1.3.2) e normalidade (1.3.3), temos que a estatística linear de postos tem distribuição assintótica Normal com parâmetros:

$$\overline{S}_{i..} \sim N(E(\overline{S}_{i..}) , VAR (\overline{S}_{i..}))$$

onde:

$$E(\bar{S}_{i..}) = \bar{a}_{...}$$

sendo :

$$\bar{a}_{...} = \frac{1}{ncp} \sum_{i=1}^{c} \sum_{k=1}^{p} a(R_{ijk}) = \bar{S}_{...}$$

e

$$VA(\bar{S}_{i..}) = \frac{1}{ncp-1} \sum_{i=1}^{c} \sum_{j=1}^{n} \sum_{k=1}^{p} (a(R_{ijk}) - \bar{a}_{...})^2 = \Delta^2$$

A estatística do teste é dada por:

$$Q_\tau = \frac{SQ_\tau}{\Delta^2}$$

onde:

$$SQ_\tau = np \sum_{i=1}^{c} (\bar{S}_{i..} - \bar{S}_{...})^2$$

e

$$\Delta^2 = \frac{1}{ncp-1} \sum_{i=1}^{c} \sum_{j=1}^{n} \sum_{k=1}^{p} (a(R_{ijk}) - \bar{a}_{...})^2$$

sob $H_{01}*$, a estatística Q_τ tem distribuição assintótica Qui-Quadrado com c-1 graus de liberdade, isto é,

$$Q_\tau \sim \chi^2_{(c-1)}$$

Regra de decisão:

A hipótese H_{01}, é rejeitada oa nível de significância α, se:

$$Q_\tau > \chi^2_{(c-1)}(1-\alpha)$$

Comparações Múltiplas

Se a hipótese nula H_{01} é rejeitada ao nível de significância α, estamos admitindo que pelo menos dois níveis do fator tem efeitos diferentes e para determinar os níveis que são diferentes, realizaremos o teste:

H_{01i}: $\tau_i = \tau_{i'}$, para $i \neq i' = 1,2,\ldots,c$ (efeito de tratamento primário)

versus

A_{1i}: $\tau_i \neq \tau_{i'}$ para algum $i \neq i'$

Regra de decisão:

Rejeitamos $H_{01i'}$, se:

$$| \overline{S}_{i\,..} - \overline{S}_{i'\,..} | > [\ (2/np)\sigma^2 \ \chi^2_{(c-1)} (1-\alpha/2) \ ]^{1/2}$$

b) Segunda hipótese nula de interesse:

H_{02}: $\mu_{i1} = \mu_{i2} = \ldots = \mu_{ip}$ (não existe efeito de condição de avaliação).

A matriz de dados pode ser vista como um conjunto de modelos mistos. Cada um dos quais corresponde a um particular tratamento (Whole plot).

Supõe-se simetria das diagonais como no caso (1) seção (3.2.3), então, testar :

$$H_{02}: \mu_{i1} = \mu_{i2} = \ldots = \mu_{ip}$$

equivale a testar :

$$H_{02}{*} : \underset{\sim}{\theta}_i = \underset{\sim}{0}$$

onde :

$$\underset{(p-1)\times 1}{\underset{\sim}{\theta}_i} = \underset{\sim}{C}\,\mu_i \quad , \text{ com} \qquad \mu_i{}' = (\mu_{i1}\ ,\ \mu_{i2}\ ,\ldots,\mu_{ip})$$

e C uma matriz de constrastes (p-1)xp linearmente inndependente por linha.

Para testar a nova hipótese nula, procedemos da seguinte maneira:

seja R_{ijk} o posto de y_{ijk} na amostra conjunta $\{y_{ij1}, y_{ij2}, \ldots, y_{ijp}\}$, isto é, os postos são dados por unidade experimental em diferentes ou condicões de avaliação.

para $i=1,2,\ldots,c$

$j=1,2,\ldots,n_i \qquad N = \sum_{i=1}^{c} n_i$

$k=1,2,\ldots,p$

obtendo-se uma matriz de postos $\underset{\sim}{R}$

$$\underset{N\times p}{\underset{\sim}{R}} = \begin{bmatrix} R_{111} & R_{112} & R_{113} & \ldots & R_{11p} \\ \vdots & \vdots & \vdots & & \vdots \\ R_{1n_1 1} & R_{1n_1 2} & R_{1n_1 3} & \ldots & R_{1n_1 p} \\ R_{211} & R_{212} & R_{213} & \ldots & R_{21p} \\ \vdots & \vdots & \vdots & & \vdots \\ R_{1n_2 1} & R_{1n_2 2} & R_{1n_2 3} & \ldots & R_{1n_2 p} \\ \vdots & \vdots & \vdots & & \vdots \\ R_{cn_c 1} & R_{cn_c 2} & R_{cn_c 3} & \ldots & R_{cn_c p} \end{bmatrix}$$

de acordo com as funções seletoras $\overline{\varphi}_2$ e $\overline{\varphi}_1$ definidas na seção (1.6.2), determinamos a função escore a(.),

assim, construimos a matriz

$$\underset{N\times p}{\underset{\sim}{a}} = \begin{bmatrix} a(R_{111}) & a(R_{112}) & a(R_{113}) & \ldots a(R_{11p}) \\ \vdots & \vdots & \vdots & \vdots \\ a(R_{1n_1 1}) & a(R_{1n_1 2}) & a(R_{1n_1 3}) & \ldots a(R_{1n_1 p}) \\ a(R_{211}) & a(R_{212}) & a(R_{213}) & \ldots a(R_{21p}) \\ \vdots & \vdots & \vdots & \vdots \\ a(R_{1n_2 1}) & a(R_{1n_2 2}) & a(R_{1n_2 3}) & \ldots a(R_{1n_2 p}) \\ \vdots & \vdots & \vdots & \vdots \\ a(R_{cn_c 1}) & a(R_{cn_c 2}) & a(R_{cn_c 3}) & \ldots a(R_{cn_c p}) \end{bmatrix}$$

$$\text{onde } N = \sum_{i=1}^{c} n_i$$

A estatística linear de postos é definida com :

$$\overline{S}_{..k} = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n_i} a(R_{ijk}) \,, \qquad \text{onde } N = \sum_{i=1}^{c} n_i$$

$$k=1,2,\ldots,p$$

Se a(.) satisfaz a condições dos teoremas de simetria (1.3.2) e normalidade(1.3.3), temos que a estatística de postos tem distribuição assintótica Normal com parâmetros:

$$\overline{S}_{..k} \sim N(E(\overline{S}_{..k}), VAR(\overline{S}_{..k}))$$

onde:

$E(\bar{S}_{..k}) = \bar{a}_{...}$

sendo, $\bar{a}_{...} = \dfrac{1}{Np} \sum_{i=1}^{c} \sum_{j=1}^{n_i} \sum_{k=1}^{p} a(R_{ijk}) = \bar{S}_{...}$

e

$COV(\bar{S}_{..k}, \bar{S}_{..k'}) = v_{kk'}$

onde:

$$v_{kk'} = \frac{1}{N^2} \sum_{i=1}^{c} \sum_{j=1}^{n_i} (a(R_{ijk}) - \bar{a}_{...})(a(R_{ijk'}) - \bar{a}_{...})$$

definimos a matriz de covariância $V = ((v_{kk'}))$, para $k,k'=1,2,..,p$.

A estatística do teste é :

$$Q_2 = \underset{\sim}{S}'\underset{\sim}{C}'[\underset{\sim}{C} V \underset{\sim}{C}']^{-1} \underset{\sim}{C} S$$

onde:

$S' = (\bar{S}_{..1}, \bar{S}_{..2}, \ldots, \bar{S}_{..p})$.

$\underset{\sim}{C}$ matriz de contrastes lineammente independente por linha.

Sob $H_{02}*$, a estatística Q_2 tem distribuição assintótica Qui-Quadrado com p-1 graus de libedade, isto é,

$$Q_2 \sim \chi^2_{(p-1)}$$

<u>Regra de decisão:</u>

Se $Q_2 > \chi^2_{(p-1)}(1-\alpha)$, então, a hipótese H_{02}, é rejeitada ao nível de significância α.

Comparações Múltiplas

No caso de rejeitar H_{02}, teríamos que estudar qual tratamento tem efeito significâtivo, para isso trabalharemos com as comparações múltiplas, utilizando a matriz de contrastes C com posto p-1 .

c) -Teste de Interação

A validade da suposição de não-interação entre tratamento primário e condição de avaliação pode-se testar da seguinte maneira:

Seja $R_{ij}^{(kk')}$ o posto de $(y_{ijk} - y_{ijk'})$ na amostra conjunta

$\{(y_{11k} - y_{11k'}), \ldots, (y_{1n_1k} - y_{1n_1k'}), \ldots, (y_{cn_ck} - y_{cn_ck})\}$

para $k \neq k' = 1, 2, \ldots, p$

de acordo com as funções seletoras φ_2 e φ_1 definidas na seção (1.6.2), determinamos a função escore a(.).

Seja $a(R_{ij}^{(k)}) = \sum_{k'=1}^{p} a(R_{ij}^{(kk')})$ uma função dos postos $R_{ij}^{(kk')}$

$$\underset{N\times p}{\underset{\sim}{a}} = \begin{bmatrix} a(R_{11}^{(1)}) & a(R_{11}^{(2)}) & a(R_{11}^{(3)}) & \ldots a(R_{11}^{(p)}) \\ \vdots & \vdots & \vdots & \vdots \\ a(R_{1n1}^{(1)}) & a(R_{1n1}^{(2)}) & a(R_{1n1}^{(3)}) & \ldots a(R_{1n1}^{(p)}) \\ a(R_{21}^{(1)}) & a(R_{212}^{(2)}) & a(R_{213}^{(3)}) & \ldots a(R_{21}^{(p)}) \\ \vdots & \vdots & \vdots & \vdots \\ a(R_{1n_2}^{(1)}) & a(R_{1n_2}^{(2)}) & a(R_{1n_2}^{(3)}) & \ldots a(R_{1n_2}^{(p)}) \\ \vdots & \vdots & \vdots & \vdots \\ a(R_{cn_c}^{(1)}) & a(R_{cn_c}^{(2)}) & a(R_{cn_c}^{(3)}) & \ldots a(R_{cn_c}^{(p)}) \end{bmatrix}$$

desta forma, determinamos a estatística linear de postos $\overline{S}_{i.k}$

$$\overline{S}_{i.k} = \frac{1}{n_i} \sum_{j=1}^{n_i} a(R_{ij}^{(k)}) \qquad , \text{ onde } N = \sum_{i=1}^{c} n_i$$

$$k=1,2,\ldots,p$$

Assim, obtemos o vetor $\overline{S}_i$

$$\overline{S}_i' = (\overline{S}_{i.1}, \overline{S}_{i.2}, \ldots, \overline{S}_{i.p}), \text{ para } i=1,2,\ldots,c$$

Se a(.) satisfaz a condições dos teoramas de simetria (1.3.2) e normalidade (1.3.3), temos que a estatística linear de posto tem distribuição assintótica Normal com parâmetros:

$$\overline{S}_{i.k} \sim N(\; E(\overline{S}_{i.k}),\; VAR(\overline{S}_{i.k}))$$

$$E(\overline{S}_{i.k}) = \overline{a}_{...}$$

sendo, $\overline{a}_{...} = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n_i} a(R_{ij}^{(k)}) = \overline{S}_{...}$, para todo k

e

$$COV(\overline{S}_{i.k}, \overline{S}_{i.k'}) =$$

$$= \frac{(Np-N)}{(Np-1)p} \; \frac{1}{N^2} \sum_{i=1}^{c} \sum_{j=1}^{n_i} (a(R_{ij}^{(k)}) - \bar{a}_{...})(a(R_{ij}^{(k')}) - \bar{a}_{...})$$

$$COV(\bar{S}_{..k}, \bar{S}_{..k'}) = \frac{(p-1)}{(Np-1)p} V_{kk'}$$

onde:

$$V_{kk'} = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n_i} (a(R_{ij}^{(k)}) - \bar{a}_{...})(a(R_{ij}^{(k')}) - \bar{a}_{...})$$

definimos a matriz de covariâncias $V = ((V_{kk'}))$

para $k, k' = 1, 2, ..p$

A estatística do teste é :(Koch: 1969)

$$Q_3 = \frac{N-1}{N} \sum_{i=1}^{c} n_i \, \bar{S}_i' C' [CVC']^{-1} C \, \bar{S}_i$$

onde:

$\bar{S}_i' = (\bar{S}_{i.1}, \bar{S}_{i.2}, ..., \bar{S}_{i.p})$, para $i=1,2,...,c$

e C é uma matriz (p-1)xp de contrastes linealmente independente por linha.

Sob a hipótese nula H_{03}, a estatística Q_3 tem distribuição assintótica Qui-Quadrado com (c-1)(p-1) graus de liberdade.

isto é, $Q_3 \sim \chi^2_{(c-1)(p-1)}$

Regra de decisão:

Rejeitamos a hipótese H_{03}, se $Q_3 > \chi^2_{(c-1)(p-1)}(1-\alpha)$, ao nível α de significância .

CAPÍTULO IV

APLICAÇÕES

Neste capítulo apresentaremos problemas práticos na qual aplicaremos testes Não-Paramétricos desenvolvidos nos capítulos II e III para a solução dos problemas.

4.1.- Problema 01. (CLASSIFICAÇÃO HIERÁRQUICA).

Um Químico deseja verificar se existe diferença de coloração das massas produzidas por diferentes processos. Para isto ele mede a mistura contida nelas. Os processos de coloração são classificados em 15 níveis: A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8, A_9, A_{10}, A_{11} , A_{12} ,A_{13} , A_{14} e A_{15}.

Ele deseja também conhecer, se o produto em cada processo tem a mesma mistura, em períodos de tempo diferentes. Para isto verifica a coloração em dois períodos de tempo de cada um dos processos, efetuando duas repetições.

Na tabela (4.1.1) apresentamos os resultados deste experimento para o qual utilizaremos um delineamento hierárquico, no qual temos 15 processos de coloração (Fator A), provados em dois períodos de tempo (Fator B), com duas repetições por período (r).

TABELA 4.1.1. Dados Observados Contendo a Mistura de Coloração de 15 Processos em dois Períodos de Tempo.

	A_1		A_2		A_3		A_4		A_5		A_6	
r	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2
1	40	30	26	25	29	14	30	24	19	17	33	26
2	39	30	28	26	28	15	31	24	20	17	32	24
$\bar{y}_{ij.}$	39.5	30	27	25.5	28.5	14.5	30.5	24	19.5	17	32.5	25
$\bar{y}_{i..}$	34.75		26.25		21.50		27.25		18.25		28.75	

	A_7		A_8		A_9		A_{10}		A_{11}		A_{12}	
r	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2
1	23	32	34	29	27	31	13	27	25	25	29	31
2	24	33	34	29	27	31	16	24	23	27	29	32
$\bar{y}_{ij.}$	23.5	32.5	34	29	27	31	14.5	25.5	24	26	29	31.5
$\bar{y}_{i..}$	28.00		31.50		29.00		20.00		25.00		30.25	

	A_{13}		A_{14}		A_{15}	
r	B_1	B_2	B_1	B_2	B_1	B_2
1	19	29	23	25	39	26
2	20	30	24	25	37	28
$\bar{y}_{ij.}$	19.5	29.5	23.5	25.	38	27
$\bar{y}_{i..}$	24.5		24.25		32.5	

$\bar{y}_{...} = 26.78$

Fonte: Box Hunter & Hunter (1978).

O modelo matemático associado a este experimento hierárquico é definido da seguinte maneira:

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + e_{ijk}$$

$i=1,2,\ldots,15.$ (nível do fator A)

$j=1,2.$ (nível do fator B)

$k=1,2.$ (repetições)

N=60 observações.

onde:

y_{ijk} : resposta ao j-ésimo nível do fator B, dentro do i-ésimo nível do fator A, na k-ésima repetição.

α_i : efeito do i-ésimo fator A.

$\beta_{j(i)}$: efeito do j-ésimo nível do fator B dentro do i-ésimo nível do fator A.

μ : média geral.

e_{ijk} : erro aleatório.

Sujeito às seguintes restrições nos parâmetros:

$$\sum_{i=1}^{15} \alpha_i = 0 \quad , \quad \sum_{j=1}^{2} \beta_{j(i)} = 0 \text{ , para todo } i$$

4.1.1.- Suposição:

Os erros e_{ijk} são variáveis aleatórias independentes, identicamente distribuidas com função de distribuição absolutamente contínua e $E(e_{ijk}) = 0$.

4.1.2.- Hipóteses nulas de interesse:

As hipóteses nulas de interesse a testar são as seguintes:

1) H_{01}: $\alpha_1 = \alpha_2 = \ldots = \alpha_{15} = 0$ (Não existe efeito do fator A)

2) H_{02}: $\beta_{1(i)} = \beta_{2(i)} = 0$ para todo i. (não existe efeito do fator B dentro do i-ésimo nível do fator A).

Segundo nosso enfoque não-paramétrico realizaremos as trasformações de alinhamento definida na seção (2.1.3), para testar se existe os efeitos em cada hipótese de interesse por separado. Assim:

a) para testar se existe efeito do fator A, temos a seguinte transformação de alinhamento:

$$x_{ijk} = y_{ijk} - \hat{\beta}_j(i)$$

então:

$$x_{ijk} = y_{ijk} - \bar{y}_{ij.} + \bar{y}_{i..}$$

sendo:

$\bar{y}_{ij.}$: efeito médio de $B_j(i)$;

$\bar{y}_{i..}$: efeito médio de A_i.

Onde resulta:

Tabela 4.1.2.- Dados Transformados para Efetuar o Teste de Não Existência dos Efeitos do Fator A.

A_1		A_2		A_3		A_4		A_5	
35.25	34.75	25.25	25.75	22	21	26.75	27.25	17.75	18.25
34.25	34.75	27.25	26.75	21	22	27.75	27.25	18.75	18.25

A_6		A_7		A_8		A_9		A_{10}	
29.25	29.75	27.50	27.5	31.5	31.5	29.0	29.0	18.50	21.50
28.25	27.75	28.50	28.5	31.5	31.5	29.0	29.0	21.50	18.50

A_{11}		A_{12}		A_{13}		A_{14}		A_{15}	
26.0	24.0	30.25	29.75	24.0	24.	23.75	24.25	33.5	31.5
24.0	26.0	30.25	30.75	25.0	25.	24.75	24.25	31.5	33.5

b) Para testar se existe efeito do fator B dentro do i-ésimo nível do fator A, temos a seguinte transformação de alinhamento:

$$x_{ijk} = y_{ijk} - \hat{\alpha}_i$$

então:

$$x_{ijk} = y_{ijk} - \overline{y}_{i..} + \overline{y}_{...}$$

sendo:

$\overline{y}_{i..}$: efeito médio de A_i ;

$\overline{y}_{...}$: média geral.

onde resulta:

Tabela 4.1.3.- Dados Transformados para Efetuar o Teste de Não Existência dos Efeitos do Fator B Dentro do i-ésimo Nínel do Fator A.

A_1		A_2		A_3		A_4		A_5	
B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2
32.03	22.03	26.53	25.53	34.28	19.28	29.53	23.53	27.53	25.53
31.03	22.03	29.53	26.53	33.28	20.28	30.53	23535	28.53	25.53

A_6		A_7		A_8		A_9		A_{10}	
B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2
31.03	24.03	21.78	30.78	29.28	24.28	24.78	28.78	19.78	33.78
30.03	22.03	22.78	31.78	29.28	24.28	24.78	28.78	22.78	30.78

A_{11}		A_{12}		A_{13}		A_{14}		A_{15}	
B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2
26.78	26.78	25.53	27.53	21.28	31.28	25.53	27.53	33.28	20.28
24.78	28.78	25.53	28.53	22.28	32.28	26.53	27.53	31.28	22.28

Para cada transformação de alinhamento definimos R_{ijk} o posto de x_{ijk} na amostra conjunta, e de acordo com as funções seletoras, $\varphi_2 = 0.9501$, que indica que a distribuição dos dados originais y_{ijk} tem um comportamento simétrico, e $\varphi_1 = 2.730$ que estuda as longitudes das caudas, assim, escolhemos a função escore de VAN DER WAERDER:

$$a(R_{ijk}) = \phi^{-1}\left(\frac{R_{ijk}}{N + 1}\right)$$

para o qual obtemos:

Tabela 4.1.4 .- Escores Para Efetuar o Teste de Não-Existência dos Efeitos do Fator A.

A_1		A_2		A_3		A_4		A_5	
2.130	1.747	-0.313	-0.270	-0.88	-1.160	-0.12	-0.020	-2.13	-1.747
1.509	0.516	-0.020	-0.120	-1.16	-0.880	0.165	-0.020	-1.29	-1.747

A_6		A_7		A_8		A_9		A_{10}	
0.538	0.611	0.082	0.082	1.019	1.019	0.423	0.423	-1.45	-1.012
0.227	0.165	0.291	0.291	1.019	1.019	0.423	0.423	-1.01	-1.450

A_{11}		A_{12}		A_{13}		A_{14}		A_{15}	
-0.206	-0.660	0.713	0.611	-0.660	-0.660	-0.790	-0.510	1.341	1.019
-0.662	-0.200	0.713	0.715	-0.370	-0.370	-0.440	-0.520	1.019	1.341

Tabela 4.1.5.- Escores para Efetuar o Teste de Não-Existência dos Efeitos do Fator B Dentro do i-ésimo Nínel do Fator A.

B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2
1.290	-1.123	-0.06	-0.249	2.130	-2.130	0.611	-0.710	0.165	-0.249
0.946	-1.123	0.611	-0.061	1.580	-1.581	0.740	-0.710	0.291	-0.249

B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2
0.946	-0.538	-1.290	0.824	0.515	-0.611	-0.440	0.401	-1.841	1.841
0.687	-1.123	-0.824	1.201	0.515	-0.611	-0.440	0.401	-0.824	0.824

B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2	B_1	B_2
0.041	0.041	-0.249	0.165	-1.390	1.084	-0.249	0.160	1.581	-1.581
-0.440	0.401	-0.249	0.291	-0.940	1.391	-0.061	0.160	1.084	-0.946

Para tais conjuntos de dados, aplicou-se os testes não paramétricos desenvolvidos na seção (2.1.4), cujos resultados são apresentados a seguir na tabela (4.1.6).

Tabela 4.1.6.- Análise Não-Paramétrica e Comparações Múltiplas para o Estudo da Mistura do Produto.

		NÃO-PARAMÉTRICA Escore de VAN DER WAERDER		
FATOR	G.L.	Q	p-value	
A	14	59.17	0	*
B(A)	15	55.93	0	*
$B(A_1)$	1	5.59	0.0181	*
$B(A_2)$	1	0.21	0.6500	
$B(A_3)$	1	15.33	0	*
$B(A_4)$	1	2.14	0.1425	
$B(A_5)$	1	0.25	0.6100	
$B(A_6)$	1	3.02	0.0822	
$B(A_7)$	1	4.77	0.0289	*
$B(A_8)$	1	1.41	0.2250	
$B(A_9)$	1	0.79	0.3700	
$B(A_{10})$	1	7.91	0.0081	*
$B(A_{11})$	1	0.20	0.6555	
$B(A_{12})$	1	0.25	0.6170	
$B(A_{13})$	1	6.43	0.0111	*
$B(A_{14})$	1	0.11	0.7400	
$B(A_{15})$	1	7.51	0.0220	*

onde:

Q : Estatística do teste não-paramétrica usando os escores de Van Der Waerder.

Observamos que utilizando o método não-paramétrico, rejeitamos a não existência do efeito do fator A com p-value=0, embora o teste Qui-Quadrado tenha sido altamente significativo, as comparações múltiplas, para α=0.05 só acusa diferença entre A_1 e A_5.
Também rejeitamos a não existência do efeito do fator B dentro do i-ésimo efeito do fator A , com um p-value = 0.

Pelos resultados apresentados na tabela (4.1.6), concluimos que não existe efeito do fator B (tempo) dentro dos fatores A_2, A_4, A_5, A_6,A_8, A_9, A_{11}, A_{12}, A_{14} (processo de coloração).

4.2.- <u>Problema 02.</u> (DELINEAMENTO EM BLOCOS ALEATÓRIOS NÃO BALANCEADOS).

Num experimento agrônomico deseja-se verificar se existe diferença entre oito tipos de fertilizantes na produção de arroz. Para isto utiliza-se seis blocos, onde por questões de espaço os blocos 1 a 4 recebem todos os tratamentos e os blocos 5 e 6 só recebem seis tratamentos.
Os dados da tabela (4.2.1), são os resultados do experimento de blocos aleatorios não balanceados.

Tabela 4.2.1.- Produção de arroz, em Kg/ha, em solo Cerrado.

BLOCOS	TRATAMENTOS							
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
B_1	500	1670	1460	1580	1380	1460	880	880
B_2	1250	1420	1870	1790	1520	2120	1850	1460
B_3	1190	2420	2590	2250	1250	2830	1980	2520
B_4	460	1660	1420	1080	670	2080	1330	1000
B_5	1290	1620	1500	1670	1420	2330	*	*
B_6	790	1380	1540	1460	790	1380	*	*

Fonte :2º Simpósio de Estatística Aplicada à Experimentação Agronômica Londrina-PR.(1987).

Considere-se o seguinte modelo matemático associado ao experimento:

$$x_{ij} = u_j + \tau_i + \sigma e_{ij}$$

$i = 1,2,\ldots,8$ (nível do fator tatamento)

$j = 1,2,\ldots,6$ (blocos)

Sujeito a:

$$\sum_{i=1}^{8} \tau_i = 0$$

onde:

x_{ij}: é a observação do i-ésimo tratamento no j-ésimo bloco.

τ_i : é o efeito do i-ésimo tratamento.

u_j : é o efeito do j-ésimo bloco.

e_{ij} : o erro aleatório.

σ: parâmetro de dispersão do j-ésimo bloco.

4.2.1.-Suposição:

Os erros e_{ij} são variáveis aleatórias independentes, identicamente distribuidas com função de distribuição absolutamente contínua e $E(e_{ij}) = 0$.

4.2.2.- Hipótese de Interesse:

A hipótese que se testará neste tipo de experimento é:

H_0: $\tau_1 = \tau_2 = \ldots = \tau_8 = 0$ (não existe efeito de tratamento).

H_1: existe pelo menos dois τ_i diferente.

Utilizando os métodos Não-Paramétricos, determinamos R_{ij} o posto de cada observação x_{ij} dentro de cada bloco. Assim, temos a seguinte matriz de postos:

$$\underset{\sim}{R} = \begin{bmatrix} 1 & 8 & 5\text{-}6 & 7 & 4 & 5\text{-}6 & 2\text{-}3 & 2\text{-}3 \\ 1 & 2 & 7 & 5 & 4 & 8 & 6 & 3 \\ 1 & 5 & 7 & 4 & 2 & 8 & 3 & 6 \\ 1 & 7 & 6 & 4 & 2 & 8 & 5 & 3 \\ 1 & 4 & 3 & 5 & 2 & 6 & * & * \\ 1\text{-}2 & 3\text{-}4 & 6 & 5 & 1\text{-}2 & 3\text{-}4 & * & * \end{bmatrix}$$

De acordo com a funçãos indicadora $\overline{\varphi}_2$=2.225, nos diz que a distribuição dos dados é assimétrica a direita. Então, selecionamos a função escore :

$$a(R_{ij}) = \begin{cases} R_{ij} / (k_j + 1) & , \text{ se } R_{ij} \leq (k_j+1)/2 \\ 1/2 + 1/(k_j + 1) & , \quad \text{c.c.} \end{cases}$$

e utilizando a transformação de Burnett and Willan , definida na seção (2.2.3), tendo a seguinte matriz de escores:

$$\underset{6 \times 8}{\underset{\sim}{a^*}} = \begin{bmatrix} -0.740 & 0.427 & 0.427 & 0.427 & 0.037 & 0.427 & -0.500 & -0.500 \\ -0.750 & -0.500 & 0.376 & 0.376 & 0 & 0.376 & 0.376 & -0.250 \\ -0.750 & 0.376 & 0.376 & 0 & -0.500 & 0.376 & -0.250 & 0.376 \\ -0.750 & 0.376 & 0.376 & 0 & -0.500 & 0.376 & 0.376 & -0.250 \\ -0.693 & 0.386 & -0.075 & 0.386 & -0.386 & 0.386 & * & * \\ -0.537 & 0.155 & 0.386 & 0.386 & -0.537 & 0.155 & * & * \end{bmatrix}$$

Assim, o vetor que nos fornecem as estatísticas lineares de postos é:

S'=(-4.220 , 1.220, 1.866, 1.575, -1.886, 2.096, 0.002, -0.624)

e matriz de covariância:

$$V=\begin{bmatrix} 1.097 & -0.177 & -0.177 & -0.177 & -0.177 & -0.177 & -0.110 & -0.110 \\ -0.177 & 1.097 & -0.177 & -0.177 & -0.177 & -0.177 & -0.110 & -0.110 \\ -0.177 & -0.177 & 1.097 & -0.177 & -0.177 & -0.177 & -0.110 & -0.110 \\ -0.177 & -0.177 & -0.177 & 1.097 & -0.177 & -0.177 & -0.110 & -0.110 \\ -0.177 & -0.177 & -0.177 & -0.177 & 1.097 & -0.177 & -0.110 & -0.110 \\ -0.177 & -0.177 & -0.177 & -0.177 & -0.177 & 1.097 & -0.110 & -0.110 \\ -0.110 & -0.110 & -0.110 & -0.110 & -0.110 & -0.110 & 0.763 & -0.110 \\ -0.110 & -0.110 & -0.110 & -0.110 & -0.110 & -0.110 & -0.110 & 0.763 \end{bmatrix}$$

4.2.3.- Estatística do Teste e Decisão:

A estatística do teste é, $Q=S'V^{-1}S=$ 26.524 com p-value=0.0004, então, segundo este resultado podemos concluir que existe evidência significativa que pelo menos dois tratamentos diferem entre si.

4.2.4.- Comparações Múltiplas:

Para determinar os tratamentos que diferem entre si, realizaremos as comparações múltiplas . Nossa hipótese de interesse será:

$H_{0i} : \tau_i = \tau_{i'}$ para todo $i \neq i' = 1, 2, \ldots 8$.

$H_{1i'} : \tau_i \neq \tau_{i'}$ para algum $i \neq i'$

obtendo os seguintes resultados, apresentados na tabela (4.2.2).

Tabela 4.2.2.- Comparações Múltiplas no Estudo do Ensaio de Adubação de 8 tratamentos na Produção de Arroz em Solo Cerrado, Utilizando Análise Não-Paramétrica.

Tratamentos	Estatística $\lvert\bar{S}_i - \bar{S}_{i'}\rvert$	p-value	Tratamentos	Estatística $\lvert\bar{S}_i - \bar{S}_{i'}\rvert$	p-value
$\tau_1 = \tau_2$	0.906	0.006*	$\tau_2 = \tau_8$	0.359	0.866
$\tau_1 = \tau_3$	1.014	0.001*	$\tau_3 = \tau_4$	0.048	0.999
$\tau_1 = \tau_4$	0.965	0.002*	$\tau_3 = \tau_5$	0.625	0.176
$\tau_1 = \tau_5$	0.389	0.706	$\tau_3 = \tau_6$	0.038	0.999
$\tau_1 = \tau_6$	1.052	0.001*	$\tau_3 = \tau_7$	0.310	0.924
$\tau_1 = \tau_7$	0.703	0.207	$\tau_3 = \tau_8$	0.467	0.674
$\tau_1 = \tau_8$	0.547	0.057	$\tau_4 = \tau_5$	0.576	0.259
$\tau_2 = \tau_3$	0.108	0.998	$\tau_4 = \tau_6$	0.086	0.999
$\tau_2 = \tau_4$	0.059	0.999	$\tau_4 = \tau_7$	0.262	0.962
$\tau_2 = \tau_5$	0.517	0.387	$\tau_4 = \tau_8$	0.418	0.770
$\tau_2 = \tau_6$	0.146	0.994	$\tau_5 = \tau_6$	0.663	0.125
$\tau_2 = \tau_7$	0.202	0.988	$\tau_5 = \tau_7$	0.313	0.920
$\tau_6 = \tau_7$	0.348	0.880	$\tau_5 = \tau_8$	0.158	0.996
$\tau_6 = \tau_8$	0.504	0.593	$\tau_7 = \tau_8$	0.156	0.993

Observamos que os tratamentos τ_1 e τ_2 , τ_1 e τ_3, τ_1 e τ_4, τ_1 e τ_6 diferem significativamente.

4.3.- Problema 03. (MODELOS MISTOS COM MEDIDAS REPETIDAS).

Um Pesquisador deseja verificar se existe diferença no tempo para um exame de sensibilidade auditiva em pacientes. escolhe ao acaso 11 pessoas e realiza o exame em cinco períodos de tempo. E mede o tempo médio de resposta em cada período.

Os dados da tabela (4.3.1), apresentam estes resultados.

Tabela 4.3.1- Reação Média de Tempo para um Exame de Sensibilidade Auditiva em Cinco Períodos de Tempo.

Sujeitos	Sinais Positivas					
	P_1	P_2	P_3	P_4	P_5	Media
S_1	51	36	50	35	42	42.8
S_2	27	20	26	17	27	23.4
S_3	37	22	41	37	30	33.4
S_4	42	36	32	34	27	34.2
S_5	27	18	33	14	29	24.2
S_6	43	32	43	35	40	38.6
S_7	41	22	36	25	38	32.4
S_8	38	21	31	20	16	25.2
S_9	36	23	27	25	28	27.8
S_{10}	26	31	31	32	36	31.2
S_{11}	29	20	25	26	25	25.0
Media	36.09	25.55	34.09	27.27	30.73	30.75

Fonte : Tim (1980) Multivariate Analysis of Repeated Measurements North-Holland Publishing Company.

Nosso objetivo é determinar se a reação média de tempo é a mesma nos 5 períodos, para o qual consideramos que os efeitos do sujeito são aleatórios e os efeitos dos períodos (condição de avaliação) são fixos.

Seja y_{ij} a resposta do i-ésimo sujeto a j-ésima período.

$i=1,2,...,11$

$j=1,2,...,5$

Definimos $y_i'=\{y_{i1}, y_{i2}, ..., y_{i5}\}$, vetor de resposta por sujeito, onde temos 11 vetores respostas estocasticamente independentes.

4.3.1.- Suposições:

Considere-se as suposições (A1) e (A2) verdadeiras, definidas na seção (3.1.1), que se refere a distribuição de y_i e contrastes.

4.3.2.- Hipótese de Interesse:

A hipótese de interesse para este tipo de experimento é saber se a reação média de tempo entre os cinco períodos é a mesma, para o qual testaremos a seguinte hipótese:

$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5$ (não existe diferença de reação de tempo entre os cinco períodos).

H_1: pelo menos dois períodos, existe diferença de reação.

4.3.3.- Análise do modelo a usar:

Para realizar a análise estatística deste experimento com medidas repetidas, utilizando um modelo misto, analisaremos se são satisfeitas as condições de um dos quatro casos desenvolvidos na seção (3.1), isto é, estudaremos a aditividade dos efeitos e estrutura de covariância das observações.

a) Teste de Não-aditividade.-

Tabela 4.3.2.-Quadro de Análise de Variância para o teste de Não Aditividade.

Fonte	G.L.	S.Q.	Q.M.	F	p-value
Não-Aditividade Transformavel	1	0.80573	0.80573	0.03352	0.8556
Resto	39	937.267	24.032		
Residuo	40	938.073			

Não existe evidência de falta de aditividade, na qual o modelo aditivo parece razoável.

b) Estrutura de Covariância.-

Assumimos a mesma estrutura de covarância para todos os vetores resposta $y_i = (y_{i1}, y_{i2}, \ldots, y_{i5})$, e assumimos estrutura de uniformidade na matriz de covariância, por ser o experimento associado com um modelo misto univariado, isto é, pressupõe-se que as variâncias dos elemento y_i são iguais e que as correlações entre dois quaisquer elementos de y_i são constantes e iguais a ρ. (Andrade D.and Singer J.:1986).

Segundo os resultados do teste de Aditividade e o estudo da estrutura de covariância utilizaremos o Caso 4 definido na seção (3.1.7) para a resolução do problema.

Então, o modelo asociado ao experimento é:

$$y_{ij} = \mu + \beta_i + \tau_j + e_{ij}$$

$$i=1,2,\ldots,11 \qquad j=1,2,\ldots,5$$

Sujeito a:

$$\sum_{j=1}^{5} \tau_j = 0$$

Assumindo que o vetor de erros $e_i' = (e_{i1}, e_{i2}, \ldots, e_{i5})$, para $i=1,2,..,11$ são vetores aleatórios 5-variados independentes, com função distribuição absolutamente contínua e esperança em torno de zero.

4.3.4.- Transformação de Alinhamento

Considerando a transformação de alinhamento para testar se existem os efeitos das condições de avaliação (períodos).

$$z_{ij} = y_{ij} - \overline{y}_i + \overline{y}_{...} \quad \text{(eliminamos o efeito do fator sujeito).}$$

onde resulta, a seguinte matriz:

$$\underset{\sim}{Z} = \begin{bmatrix} 38.95 & 23.95 & 37.95 & 22.95 & 29.95 \\ 34.35 & 27.35 & 33.35 & 24.35 & 34.35 \\ 34.35 & 19.35 & 38.35 & 34.35 & 27.35 \\ 38.55 & 32.55 & 28.55 & 30.55 & 23.55 \\ 33.55 & 24.55 & 39.55 & 20.55 & 35.55 \\ 35.15 & 24.15 & 35.15 & 27.15 & 32.15 \\ 39.35 & 20.35 & 34.35 & 23.35 & 36.35 \\ 43.55 & 26.55 & 36.55 & 25.55 & 21.55 \\ 38.75 & 25.95 & 29.95 & 27.95 & 30.95 \\ 25.55 & 30.55 & 30.55 & 31.55 & 35.55 \\ 34.75 & 25.75 & 30.75 & 31.75 & 30.75 \end{bmatrix}$$

Seja R_{ij} o posto de z_{ij} na amostra conjunta $\{z_{11}, \ldots, z_{115}\}$;
Definimos a matriz de postos $\underset{\sim}{R}$:

$$R = \begin{bmatrix} 51\text{-}52 & 8 & 48 & 5 & 22\text{-}23 \\ 36\text{-}40 & 18\text{-}19 & 34 & 10 & 36\text{-}40 \\ 36\text{-}40 & 1 & 49 & 36\text{-}40 & 18\text{-}19 \\ 50 & 33 & 21 & 24\text{-}26 & 7 \\ 35 & 11 & 54 & 3 & 44\text{-}45 \\ 42\text{-}43 & 9 & 42\text{-}43 & 17 & 32 \\ 53 & 2 & 36\text{-}40 & 6 & 46 \\ 55 & 16 & 47 & 12\text{-}13 & 4 \\ 51\text{-}52 & 15 & 22\text{-}23 & 20 & 29 \\ 12\text{-}13 & 24\text{-}26 & 24\text{-}26 & 30 & 44\text{-}45 \\ 41 & 14 & 27\text{-}28 & 31 & 27\text{-}28 \end{bmatrix}$$

Para determinar a função escore calculamos quanto valem os indicadores de simetria φ_2 e longitude das caudas φ_1.

φ_2 = 1.7626 , então , os dados tem um comportamento simétrico, e

φ_1 = 2.1622, então, segundo a seção (1.6.2), a função escore será:

$$a(R_{ij}) = \phi^{-1}(R_{ij}/N+1) \quad \text{escore de VAN DER WAERDEN}$$

Obtendo a matriz de escores $\underset{\sim}{a}$:

$$
\underset{\sim}{a} = \begin{bmatrix}
1.405 & -1.067 & 1.067 & -1.345 & -0.248 \\
0.464 & -0.439 & 0.271 & -0.920 & 0.464 \\
0.464 & -2.100 & 1.150 & 0.464 & -0.439 \\
1.241 & 0.225 & -0.318 & -0.134 & -1.150 \\
0.318 & -0.854 & 1.802 & -1.611 & 0.823 \\
0.703 & -0.991 & 0.703 & -0.514 & 0.180 \\
1.611 & -1.802 & 0.464 & -1.241 & 0.920 \\
2.100 & -0.565 & 0.991 & -0.761 & -1.465 \\
1.405 & -0.619 & -0.248 & -0.366 & 0.044 \\
-0.761 & -0.134 & -0.134 & 0.089 & 0.823 \\
0.619 & -0.674 & -0.022 & 0.134 & -0.022
\end{bmatrix}
$$

Assim, determinamos o vetor das estatíticas lineares de postos $S' = (0.87024 \quad -0.82039 \quad 0.52067 \quad -0.56423 \quad 0.006296)$

$VAR(S_j) = 0.89425$

e

$SQ(\tau) = 11(2.019849) = 22.21819$

então: $W = SQ(\tau) / VAR(S_j) = 24.8457.$

4.3.5.- <u>Estatística do Teste e Decisão:</u>

Sob H_0, a estatística do teste é, $W = 24.8457$ com p-value = 0. Pela qual, chegamos a conclusão de que a hipótese nula é rejeitada, isto é, existe diferença pelo menos em dois períodos das reações médias de tempo .

4.3.6.- Comparações Múltiplas

Para determinar quais períodos serão diferentes efetuamos as seguintes comparações múltiplas:

$H_0 : \tau_j = \tau_{j'}$ para todo $j \neq j' = 1, 2, \ldots, 5$

$H_1 : \tau_j \neq \tau_{j'}$ para algum $j \neq j'$

obtendo os seguintes resultados apresentados na tabela (4.3.3).

Tabela 4.3.3.- Comparações Múltiplas no Estudo da Reação Média de Tempo para o Exame de Sensibilidade Auditiva, Utilizando Teste não-paramétrico.

Provas	Estatística	p-vale
$\tau_1 = \tau_2$	17.57	0 *
$\tau_1 = \tau_3$	0.751	0.944
$\tau_1 = \tau_4$	12.650	0.013 *
$\tau_1 = \tau_5$	4.724	0.316
$\tau_2 = \tau_3$	11.050	0 *
$\tau_2 = \tau_4$	0.403	0.982
$\tau_2 = \tau_5$	4.075	0.395
$\tau_3 = \tau_4$	7.230	0.124
$\tau_3 = \tau_5$	1.700	0.789
$\tau_4 = \tau_5$	1.914	0.751

Pelo resultados apresentados na tabela (4.3.3) concluimos que as reações médias de tempo nos períodos 1 e 2, 2 e 3, 1 e 4 diferem significativamente.

4.4.- <u>PROBLEMA 04</u> (CLASSIFICAÇÃO EM PARCELA SUBDIVIDIDA "SPLIT PLOT")

Num experimento de preparação de biscoitos de chocolate, é realizado em 3 recipientes para a preparação da massa, onde cada recipiente tem caracteristicas diferentes quanto à forma, e deseja-se comparar em 6 diferentes temperaturas a estrutura dos biscoitos, para o qual se mede o número de biscoitos que não tem boa qualidade. Os resultados do experimento são apresentados na tabela (4.4.1).

Tabela 4.4.1.- Número de Biscoitos que não têm Boa Qualidade Segundo Tipo de Recipiente por Temperatura.

Recipiente	Rep	Temperatura 175°	185°	195°	205°	215°	225°
I	1	42	46	47	39	53	42
	2	47	29	35	47	57	45
	3	32	32	37	43	45	45
	4	26	32	35	24	39	26
	5	28	30	31	37	41	47
	6	24	22	22	29	35	26
	7	26	23	25	27	33	35
	8	24	33	23	32	31	34
	9	24	27	28	33	34	23
	10	24	33	27	31	30	33
	11	33	39	33	28	33	30
	12	28	31	27	39	35	43
	13	29	28	31	29	37	33
	14	24	40	29	40	40	31
	15	26	28	32	25	37	33
II	1	39	46	51	49	55	42
	2	35	46	47	39	52	61
	3	34	30	42	35	42	35
	4	25	26	28	46	37	37
	5	31	30	29	35	40	36
	6	24	29	29	29	24	35
	7	22	25	26	26	29	36
	8	26	23	24	31	27	37
	9	27	26	32	28	32	33
	10	21	24	24	27	37	30
	11	20	27	33	31	28	33
	12	23	28	31	34	31	29
	13	32	35	30	27	35	30
	14	23	25	22	19	21	35
	15	21	21	28	26	27	20

segue....

III	1	46	44	45	46	48	63
	2	43	43	43	46	47	58
	3	33	24	40	37	41	38
	4	38	41	38	30	36	35
	5	21	25	31	35	33	23
	6	24	33	30	30	37	35
	7	20	21	31	24	30	33
	8	24	23	21	24	21	35
	9	24	18	21	26	28	28
	10	26	28	27	27	35	35
	11	28	25	26	25	38	28
	12	24	30	28	35	33	28
	13	28	29	43	28	33	37
	14	19	22	27	25	25	35
	15	21	28	25	25	31	25

Fonte : William G.Cocrhan & Gertrude Cox. Experimental Designs (1957).

Consideremos que o experimento tem 45 unidades experimentais, que são designadas ao acaso a 3 recipientes (tratamento primário) e respondem a 6 temperaturas diferentes (condição de avaliação).

Seja :

y_{ijk} : Número de biscoitos quebrados na j-ésima unidade experimental no i-ésimo recipiente da k-ésima temperatura.

$i=1,2,3$ (nível do Recipientes);

$j=1,2,..,15$ (unidades experimentais);

$k=1,2,...6$ (condições de avaliação);

$N=45$.

$y_{ij}=(y_{ij1}, y_{ij2}, ..., y_{ij6})$ o vetor resposta da j-ésima unidade experimental ao i-ésimo recipiente.

4.4.1.-Suposições:

1) Supondo que y_{ij} tem distribuição absolutamente contínua 6-variada.

2) Por tratar-se de um experimento em "Split Plot" na qual as subparcelas são aleatorizadas, temos que a matriz de covariância tem estrutura uniforme.

4.4.2.-Análise de Perfis Médios.

Realizando uma análise exploratória dos dados, para estudar os perfis médios de respostas dos três recipentes nas 6 temperaturas.

Tabela 4.4.2.- Médias Amostrais observadas no estudo da Qualidade de Biscoitos Segundo Recipiente, por Temperatura.

	Temperatura					
Recipiente	175^0	185^0	195^0	205^0	215^0	225^0
I	29.1	31.5	30.8	33.5	38.7	35.1
II	26.9	29.4	31.7	32.1	34.5	35.3
III	27.9	28.9	31.7	30.9	34.4	35.7

Figura 4.4.1 .Representação gráfica dos perfis médios observados de resposta correspondentes a uma situação com 3 recipentes e 6 temperaturas (condições de avaliação).

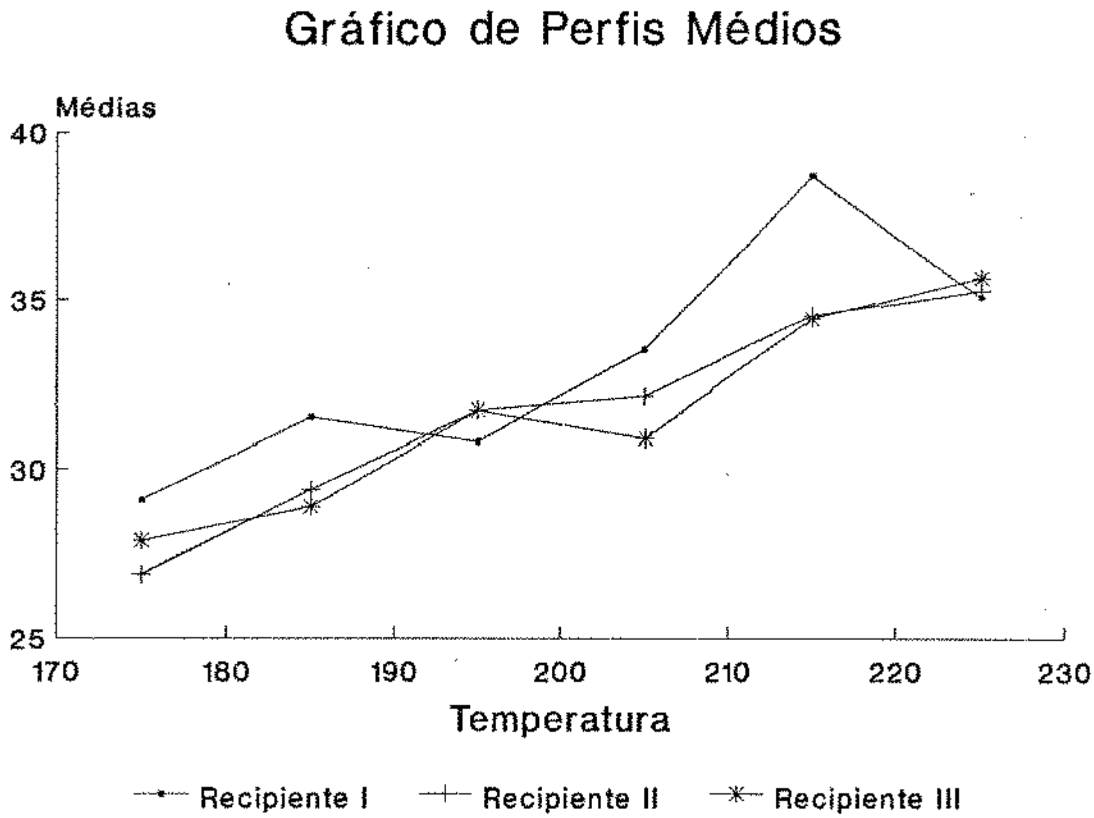


Gráfico de Perfis Médios
Médias
40
35
30
25
170
180
190
200
210
220
230
Temperatura
Recipiente I
Recipiente II
Recipiente III

Segundo as suposições feitas e a análise de perfis de médias, utilizaremos o CASO 2 definido na seção (3.3.3), para resolver este problema pela análise Não-Paramétrica .

Na seleção de escores, obtem-se o indicador $\overline{\varphi}_2$=2.69, no qual nos diz que a distribuição dos dados é assimétrica direita.

4.4.3.-Análise Exploratoria dos Dados.

Para comprovar este resultado de assimetria, efetuaremos o análise exploratoria dos dados, na qual realizaremos o diagrama de Ramos-e-Folhas, estudaremos o 2^0- Coeficiente de Pearson e a curtose dos dados amostrais.

i) Ramos-e-Folhas

Apresentaremos o diagramna de Ramos-e-Folhas na figura (4.4.2), para estudar a forma da distribuição dos dados amostrais. (Tukey: 1977).

Figura 4.4.2 Diagrama de Ramos-e folhas dos dados do número de biscoitos que não tem boa qualidade

```
18 |0
19 |00
20 |000
21 |0000000000
22 |00000
23 |00000000
24 |000000000000000000
25 |0000000000000
26 |00000000000000
27 |0000000000000
28 |000000000000000000000
29 |00000000000
30 |0000000000000
31 |00000000000000
32 |00000000
33 |0000000000000000000
34 |0000
35 |00000000000000000000
36 |000
37 |00000000000
38 |0000
39 |000000
40 |00000
41 |000
42 |00000
43 |000000
44 |0
45 |0000
46 |0000000
47 |000000
48 |0
49 |0
50 |
51 |0
52 |0
53 |0
54 |
55 |0
56 |
57 |0
58 |0
59 |
60 |
61 |0
62 |
63 |0
```

ii) 2º Coeficiente de Pearson

O 2º Coeficiente de Pearson é utilizado para indicar-nos a forma da distribuição dos dados amostrais.

Seja : M_e a Mediana,

Q_i o i-ésimo quartil,

então, o 2º Coeficiente de Pearson é:

$$As = \frac{Q_3 + Q_1 - 2M_e}{Q_3 - Q_1} = \frac{26 + 37 - 2(31)}{11} = 0.09$$

Este resultado As=0.09, nos da um indicativo que a distribuição dos dados amostrais tem um comportamento assimetrico a Direita.

iii) Medida de Curtose

O coeficiente de curtose é dado da seguinte maneira:

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{37 - 26}{44} = 0.25$$

diremos que a curva correspondente à distribuição de freqüência é leptocúrtica.

4.4.4.- Escolha da Função Escore

Segundo o indicador de simetria $\overline{\varphi}_2$, e a análise exploratória dos dados, escolheremos a função escore:

$$a(R_{ijk}) = \begin{cases} R_{ijk}/(npc+1) & , \text{ se } R_{ijk} \leq (ncp+1)/2 \\ 1/2 + 1/(ncp+1) & , \text{ se } R_{ijk} > (ncp+1)/2 \end{cases}$$

(assimétrica a direita) definida na seção (1.6.2).

Os resultados da aplicação do teste Não-Paramétrica aos dados é apresentado na tabela (4.4.3).

Tabela 4.4.3.- Análise Não-Paramétrica para o Modelo Proposto.

Fonte	G.L.	Q	p-value
Recipiente	2	7.55	0.022
Temperatura	5	42.83	0
RecxTemp	10	7.67	0.66

Q : estatística do teste utilizando o escore da distribuição assimétrica a direita.

Segundo a tabela (4.4.3), observamos que existe o efeito do fator Recipiente, então, para determinar quais dos recipientes são diferentes, efetuamos as comparações múltiplas, obtendo-se que o recipiente 1 e 3 são diferentes.

Também observamos que existe o efeito temperatura na estrutura dos biscoitos, então, para determinar quais das temperaturas são diferentes, realisamos as comparações múltiplas obtendo os seguintes resultados:

Tabela 4.4.4.- Comparações dois a dois no Estudo da Temperatura na Qualidade dos Biscoitos Utilizando Teste Não-Paramétrica.

Significância

Temperaturas	10%	5%	1%
175^0 e 185^0	NS	NS	NS
175^0 e 195^0	S	S	S
175^0 e 205^0	S	S	S
175^0 e 215^0	S	S	S
175^0 e 225^0	S	S	S
185^0 e 215^0	S	S	NS
185^0 e 225^0	S	S	NS

NS :Diferença não significativa

S : Diferença significativa

Outrossim, obteve-se que as demais comparações das temperaturas não são significativas.

É interessante notar que este problema 04, as observações não tem distribuição normal, mas foi possível analisar-se pela estatística não-paramétrica.

CONCLUSÕES

As principais vantagens do uso de testes não-paramétricos devem-se ao fato de serem menos exigentes e mais eficientes que os métodos paramétricos, quando os dados da população não têm distribuição normal. São úteis no caso em que é difícil estabelecer uma escala de valores quantitativos para os dados e geralmente estes testes Não-Paramétricos, baseados nos postos, são robustos ante valores discrepantes. Por isso, podemos afirmar que os métodos Não-Paramétricos podem ser uma boa complementação aos métodos Paramétricos.

Quando o número de observações é pequeno deve-se usar a distribuição exata, para o qual é necessario realizar muitos cálculos , combinações e simulações. Este tipo de problema poderá ser abordado em outros trabalhos posteriores.

A análise exploratória dos dados, assim como os testes de aditividade e a análise da estrutura de covariância tem um fator importante para decidir os métodos e casos a aplicar.

O uso das tranformações de alinhamento apresentadas neste trabalho são de grande utilidade no desenvolvimento dos testes, quando existe aditividade dos efeitos e está sendo incorporado nos ultimos trabalhos de análises não-paramétricas.

É importante mencionar que está em desenvolvimento o pacote SENP que conterá recursos para a aplicação da maioria dos testes apresentados neste trabalho. O SENP está sendo coordenado pelo Prof. Belmer Garcia Negrillo (UNICAMP) com a colaboração do autor deste trabalho e de professores do centro de computação de ESALQ/USP (CIAGRI).

REFERÊNCIAS BILBIOGRÁFICAS

AGRESTI A and PENDERGAST J , (1986) Comparing Mean Ranks For Repeated Measures Data , Commun. Statist. Theor. Meth 15(5) 1417-1433.

BURNETT, R. and WILLAN A., (1988) Linear Rank For Randomized Block Designs . Commun. Statist.-Theory Meth. 17(8) 2455-2470.

BRUNNER and NEUMANN, (1982) Rank Tests for Correlated Randon Variables. Biometric 24 373-389.

COCHRAN W. and COX G. (1957) Experimental Designs. Ed. John Wiley & Sons,Inc. Londom.

ANDRADE D. and SINGER J. (1986) Análise de Dados Longitudinais. VII Simpósio Nacional de Probabilidade e Estatística.

HÁJEK, J. (1961) Some extensions of the Wald-Wolfowitz-Noether Theorem, Math. Statist.32 506-523.

HÁJEK, J. and SIDAK, Z. (1967) Theory of Rank Tests. Academia, Prague. Academi Press New York.

HÁJEK J. (1969) Noparametric Statistics. Academia Holdenday San Francisco California.

HÁJEK J.(1970) Miscellaneous Problems of Rank Test Noparametric Techniques in Statistical Inferencial. ed. Puri M.L. 3-17 .

HÁJEK, and SIDAK . (1973) Approach to the Asymptotic Distribuition of Multrivariate Rank Order Statistics. Journal of Multivariante Analysis 3, 57-70.

HILL and PADMANABHAN (1988) Adaptive Nonparametric Procedures and Applications , Appl.Statist. 37(2) 205-218.

HOGG R.V. (1982) On Adaptive StatisticalInference. Commum. Statist. Theor. Meth. 11(22) pag 2531-2542.

HOGG,R.V.(1974) Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory. Journal of the American Statistical Associations 69(348) 909-927.

HOEFFDING, W. (1951) Optimum nonparametric tests. IN Proc. Second Berkeley Symp. Math. Statist. Probab., Vol 1 (Ed: J. Neyman). Univ. of California Press. 203-220.

KOCH, G. and SEN P. (1968) Some Aspects of the Statistical Analysis of the Mixed Model ,Biometrics, 24 27-47.

KOCH, G.G. (1969) Some Aspects of the Statistical Analysis of "Split Plot" Experiments im Completely Randomized Layouts, American Statistical Association Journal, 64 486-505.

KOCH, G.G. (1970) The use of noparametric methods in the Statistical Analysis of complex Split-plot Experiment. Biometrics, 26, 105-128.

KOCH, G.G. (1980) Some Views on Parametric and Non-Parametric Analysis for Repeated Measurements and Selected Bibliography. Biometrics 249-265.

KOCH,G.G. (1985) Repeated Measurements Studies, Design and Analysis. Biometrics.

KUBINGER K.D. (1986) A Note on Non-Parametric Tests for Interaction in Two-Way Lanjuts. Biometrics 28 67-72.

PIRIE W. and HOWARD L. (1984) Simulated Efficiencies of Tests and Estimators from General Linear Models Analysis based on Rank: The Two-Way Layout. Journal Statist. Comput. Simul. 20 197-204.

PURI, M.L. and SEN. (1969) A Class of Rank Order Test for a General Linear Hipothesis. Amn. Math. Statist. 40 135-1343.

PURI, M.L. And SEN. (1971) Noparametric Methods in Multivariate Analysis. New York: John Wiley .

PURI, M.L. And SEN. (1985) Noparametric Methods in General Linear Models .John Wiley .

PURI, M.L. and SEOH M. (1989) Central Limit Theorems Under Alternatives for a Broad Class of Nonparametric Statistics, Journal of Statistical Planning and Inference 22 271-294.

MANSOURI.H .(1990) Rank Tests for ordered alternatives in Analysis of Variance Journal of Statistic Planning and Inference. 24 107-1117.

MEHROTRA K.G. and MICHALEK J.E. THOMAS WRITE (1983) Score Computation forLinear Rank Procedures, Journal Statist. Comput. Simul. 16 201-211.

MEYDRECH E.F (1986) On a Nonparametric Test For The Mixed-Model. Commum. Statst. Theor. Meth. 15(5), 1587-1596.

NEGRILLO, B.G.(1987) Método Não Paramétrico Uni e Multivariado. Relatório Interno IMECC/UNICAMP.

NEGRILLO , B.G.(1989) Um Modelo Linear Geral Não-Paramético. Relatório Técnico 34/89 IMECC/UNICAMP.

NEGRILLO, B.G. (1990) Tecnicas Não-Paramétricas para Experimentos Completamente Casualizados para Três Fatores, com Interação. Relatório Técnico 30/90 IMECC/UNICAMP.

RAZ J. (1989) Analysis of Repeated Mesurements Using Nonparametric Smoothers and Randomization Tests.Biometrics 45 851-871.

SHIRAISHI T,(1990) R-Estimators and Confidence Regions in One-Way MANOVA Journal of Statistical Planning and Inference. 24 203-213.

TIMM N. H.(1980) Multivariate Analysis of Repeated Measurements North-Holland Publishing Company 41-87.

TUKEY J.W. (1977) Exploratory Data Analysis. Reading, Addison Wesley, Massachusetts.

VAN DER LAAN and VERDOOREN L,R. (1986) Classical Analysis of Variance Methods and Noparametric . Counterparts Statistica Neerlandica 635-663.

VORLIČKOVÁ. D. (1970) Asymptotic Properties of Rank Tests of Symetry Under Discrete Distribuitions.AMS 43 275-289.

WOOLSON, R.F. and LOCHENBRUCH, P.A. (1981) Rank Tests for Censored Randomized Block Designs. BiometriKa, 68 427-435.