UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
Departamento de Estatística

ESTRUTURA TEORICA DOS COEFICIENTES TIPO KAPPA

Autor: ÁLVARO VIGO
Orientador: Professora JANDYRA M. G. FACHEL

Monografia apresentada para a obtenção do título de Bacharel em Estatística.

Porto Alegre, agosto de 1989.

A G R A D E C I M E N T O S

Ao Professor **Paulo S. Belmonte de Abreu**, do Departamento de Psiquiatria e Medicina Legal da Faculdade de Medicina desta Universidade, por ter motivado esta monografia, ao trazer ao Núcleo de Assessoria Estatística o problema. Também por ter fornecido parte da bibliografia utilizada, por ter participado das discussões teóricas sobre o assunto e finalmente pela autorização para usar seus dados como exemplo nesta monografia.

Ao Moacir, pela digitação, e aos colegas do **NAE** que colaboraram para a realização do presente trabalho.

Em especial, à amiga e Professora **Jandyra M. G. Fachel** pelo apoio e incentivo recebido e pela orientação prestada à presente monografia, cujas contribuições foram da mais valiosa importância.

Manifesto também meu reconhecimento ao trabalho que está sendo desenvolvido pelos professores João Riboldi, Elsa Mundstock e Jandyra Fachel, na estruturação e ampliação do Núcleo de Assessoria Estatística - **NAE**, através do qual adquire-se uma formação Estatística adequada à realidade do mercado de trabalho.

S U M Á R I O

1. INTRODUÇÃO

1.1. Apresentação

A origem da presente monografia está ligada à consulta feita ao Núcleo de Assessoria Estatística do Departamento de Estatística, pelo Professor Paulo S. Belmonte de Abreu, do Departamento de Psiquiatria e Medicina Legal da Faculdade de Medicina, quanto à aplicação do coeficiente de concordância kappa, para múltiplos juízes por sujeito, no caso de diagnósticos médicos.

Tendo em vista a diversidade de coeficientes kappa, aplicáveis tanto no contexto de concordância, quanto no contexto de associação, bem como à falta de referência bibliográfica nos livros tradicionais da área de Estatística, optou-se por um trabalho de cunho mais teórico, sem esquecer-se do enfoque aplicado, quando possível. No tocante à bibliografia, cabe ressaltar que o material disponível é oriundo, na sua maioria, das publicações da área médica.

Assim, o objetivo da monografia é resumir a informação disponível num único documento, apresentar de forma clara e concisa e tornar didático o uso do coeficiente kappa.

Alguns tópicos parecerão demasiadamente complexos aos leitores que não têm formação estatística, principalmente aqueles relativos à determinação dos Estimadores de Máxima Verossimilhança e Estimadores Jackknife. A esses leitores, aconselha-se desconsiderar os aspectos que forem excessivamente matemáticos.

Salienta-se, ainda, que embora tenha sido apresentado o coeficiente kappa 2 x 2 para o contexto de associação, a abordagem central desse trabalho é o contexto de concordância. Neste sentido cabe esclarecer, desde já, que estaremos nos referindo à concordância entre os juízes, com respeito às categorias de respostas (diagnósticos), nas quais eles (juízes) classificaram os indivíduos extraídos da população através de uma amostra aleatória.

1.2. Conceito de Fidedignidade

Os sistemas de classificação de diagnósticos, segundo Spitzer e Fleiss (1974), têm duas propriedades primárias: Fidedignidade e Validade. Fidedignidade diz respeito à consistência com a qual os sujeitos são classificados e validade refere-se à utilidade do sistema para atingir os objetivos propostos. No caso de diagnósticos psiquiátricos, o propósito do sistema de classificação é a obtenção de informações sobre as características clínicas, manifestação, exclusão e tratamento da doença. Uma condição necessária para que um sistema de classificação seja válido é que ele seja fidedigno. Não existem garantias de que um sistema fidedigno é válido, mas, certamente, se não for fidedigno será inválido.

A fidedignidade do sistema de diagnóstico pode ser analisada em sob três aspectos distintos:

- concordância entre os juízes independentes, que examinarem os mesmos pacientes;
- estabilidade dos diagnósticos ao longo do tempo;
- similaridade nas frequências de diagnósticos para amostras comparáveis.

Contudo, o primeiro aspecto (concordância entre juízes) é de fundamental importância e será o aspecto abordado nesta monografia.

Segundo Lord e Novick (1968), a fidedignidade do teste é uma medida do grau da variação dos escores verdadeiros, em relação à variação dos escores observados. Assim, temos:

$$X = T + E$$

onde X é o escore observado; T é o escore verdadeiro e E é o erro. Disso resulta a seguinte definição do coeficiente de fidedignidade da teoria clássica:

$$\rho^2_{XT} = \frac{\sigma^2_T}{\sigma^2_X} = 1 - \frac{\sigma^2_E}{\sigma^2_X}$$

1.3. Algumas Medidas de Concordância para 2 Juízes

Consideremos a situação na qual cada juiz pode classificar os indivíduos, selecionados aleatóriamente de uma população, numa das K categorias de respostas (diagnósticos), conforme mostra a tabela a seguir:

Tabela 1.1 - Forma Geral
das Frequências de Resposta

Juiz 1	Juiz 2: 1	2	...	K	Total
1	x_{11}	x_{12}	...	x_{1K}	x_1
2	x_{21}	x_{22}	...	x_{2K}	x_2
:	:	:		:	:
K	x_{K1}	x_{K2}	...	x_{KK}	x_K
Total	$x._1$	$x._2$	...	$x._K$	$x.. = n$

onde: x_{ij} = número de indivíduos classificados, independetemente, na categoria i pelo Juiz 1 e na categoria j pelo Juiz 2.

$i, j = 1, 2, \ldots, K$

$x..$ = n = número de indivíduos selecionados na população.

Queremos determinar a concordância existente entre os juízes, no que diz respeito à classificação dos indivíduos nas diferentes categorias de diagnósticos. Podemos considerar "concordância" como um caso especial de associação. A distinção entre concordância e associação é que para duas respostas concordarem, elas precisam estar em categorias idênticas, enquanto que para elas estarem perfeitamente associadas, basta que possamos predizer a categoria de uma resposta através do conhecimento de outra categoria de resposta. Desta forma, uma tabela de contingência pode apresentar um alto grau de associação e, ao mesmo tempo, uma pequena concordância.

Cabe ressaltar, ainda, que o número de linhas da Tabela 1.1 é igual ao número de colunas e, principalmente, que a ordem das categorias para as linhas precisa ser idêntica à ordem que aparece nas colunas.

Vamos considerar uma amostra de n indivíduos selecionados aleatoriamente de uma população. Suponha que cada indivíduo é classificado independentemente pelos mesmos dois juízes numa das K categorias de diagnósticos. Tomemos como exemplo os dados hipotéticos da Tabela 1.2, apresentados por Fleiss (1981). Cada cela corresponde a proporção dos indivíduos, em relação ao total, classificados em cada uma das K=3 categorias de diagnósticos. Assim, 5 % dos indivíduos foram classificados como neuróticos pelo Juiz A e psicóticos pelo Juiz B.

Tabela 1.2 - Proporção dos Indivíduos Classificados nas Categorias de Diagnósticos pelos dois Juízes (n=100)

Juiz A	Juiz B Psicótico	Neurótico	Orgânico	Total
Psicótico	0,75	0,01	0,04	0,80
Neurótico	0,05	0,04	0,01	0,10
Orgânico	0	0	0,10	0,10
Total	0,80	0,05	0,15	1,00

Queremos determinar o grau de concordância para cada categoria separadamente. Para tanto, precisamos separar a tabela original (K x K) em tabelas 2 x 2, nas quais as categorias de diagnósticos que não são de interesse imediato são combinadas numa única categoria (somando as respectivas proporções) chamada de "Outras". Dessa forma, podemos apresentar a forma geral dessas K tabelas 2 x 2:

Tabela 1.3 - Forma Geral das Tabelas 2 x 2 Resultante da Combinação das Categorias

Juiz A	Juiz B Categoria de Interesse	Outras	Total
Categoria de Interesse	a	b	p_1
Outras	c	d	q_1
Total	p_2	q_2	1

onde $p_2 = a + c$; $q_2 = 1 - p_2$; $p_1 = a + b$; $q_1 = 1 - p_1$.

Assim, para determinar o grau de concordância entre os juízes, no tocante a categoria de diagnóstico "Neurótico", temos a seguinte tabela:

Tabela 1.4 - Tabela para a Medida de Concordância da Categoria de Diagnóstico "Neurótico"

Juiz A	Juiz B		Total
	Neurótico	Outros	
Neurótico	0,04	0,06	0,10
Outros	0,01	0,89	0,90
Total	0,05	0,95	1,00

O índice de concordância mais simples é a proporção de concordância total, definido por

$$p_o = a + d \qquad (1.1)$$

Para a categoria de diagnóstico "Neurótico", temos:

$$p_o = 0{,}04 + 0{,}89 = 0{,}93$$

De forma similar, podemos determinar a proporção de concordância total para as categorias de diagnósticos "Psicótico" e "Orgânico". Esses valores são apresentados na Tabela 1.5, e através deles podemos concluir que existe uma boa concordância entre os juízes para as categorias de diagnósticos.

Holley, Guilford e Maxwell, sugeriram utilizar a variante $2p_o - 1$ como índice de concordância.

Contudo, se a categoria de interesse é relativamente rara, então a proporção d, que representa ausência de concordância, será grande, inflacionando o valor de p_o. Vários índices de concordância baseados somente nas proporções a,b e c, têm sido propostos. Entretanto, somente o índice p_s, chamado de proporção de concordância especifica, tem uma interpretação probabilística sensível, o qual é definido por:

$$p_s = \frac{2a}{2a + b + c} = a / \bar{p} \qquad (1.2)$$

onde $\bar{p} = (p_1 + p_2) / 2$.

Esse índice foi inicialmente proposto por Dice como uma medida de similaridade e tem a seguinte interpretação: um dos dois juízes é selecionado aleatoriamente e observemos os sujeitos classificados na categoria de interesse. Assim, p_s é a probabilidade condicional de que o segundo juiz também classificará os indivíduos na categoria, dado que o primeiro foi selecionado aleatoriamente.

Para a categoria "Neurótico" do exemplo sob estudo, temos:

$$p_s = \frac{2 \times 0{,}04}{2 \times 0{,}04 + 0{,}06 + 0{,}01} = 0{,}53$$

Os valores do índice de concordância específica p_s para as três categorias são apresentados na Tabela 1.5. Claramente podemos observar que as conclusões baseadas no índice p_o são diferentes daquelas baseadas no índice p_s.

Vamos definir agora o índice de concordância λ_r, proposto por Goodman e Kruskal.

Seja $\bar{q} = 1 - \bar{p}$, ou $\bar{q} = 1/2\,(q_1 + q_2) = d + (b + c)/2$ (1.3)

e suponhamos que $\bar{q} > \bar{p}$. Então,

$$\lambda_r = \frac{(a + d) - \bar{q}}{1 - \bar{q}} = \frac{2a - (b + c)}{2a + (b + c)} \qquad (1.4)$$

O índice λ_r assume seu valor máximo $\lambda_r = 1$ quando existe completa concordância e seu valor mínimo $\lambda_r = -1$ sempre que $a = 0$, independentemente do valor de d.

Para a categoria de diagnóstico "Neurótico", temos

$$\lambda_r = \frac{2 \times 0,04 - (0,06 + 0,01)}{2 \times 0,04 + (0,06 + 0,01)} = 0,06$$

e os valores de λ_r para as três categorias também são mostrados na Tabela 1.5. Como $\lambda_r = 2p_s - 1$, as conclusões são as mesmas apresentadas para o índice p_s.

A proporção de concordância específica ignora a proporção d. Se, ao invés disso, quiséssemos ignorar a proporção a, calcularíamos o índice correspondente, definido por:

$$p'_s = d / \bar{q} = 2d / 2d + b + c \qquad (1.5)$$

onde $\bar{q} = 1 - \bar{p}$.

Assim, para "Neurótico", temos:

$$p'_s = \frac{2 \times 0,89}{2 \times 0,89 + 0,06 + 0,01} = 0,96$$

Os valores de p'_s, mostrados na Tabela 1.5, dizem respeito a concordância, em relação à ausência do diagnóstico.

Rogot e Goldberg propuseram a utilização da média aritmética dos índices p_s e p'_s como um indicador de concordância, o qual foi denotado por:

$$A = \frac{1}{2} (p_s + p'_s) = \frac{a}{p_1 + p_2} + \frac{d}{q_1 + q_2} \qquad (1.6)$$

Para "Neurótico", temos

$$A = \frac{0,04}{0,10 + 0,05} + \frac{0,89}{0,90 + 0,95} = 0,75$$

também sendo mostrados na Tabela 1.5 os valores de A para as três categorias de diagnósticos.

Tabela 1.5 - Valores dos Índices de Concordância para os dados da Tabela 1.2

Categoria de Diagnóstico	p_o	p_s	λ_r	p_s'	A	κ
Psicótico	0,90	0,94	0,88	0,75	0,84	0,69
Neurótico	0,93	0,53	0,06	0,96	0,75	0,50
Orgânico	0,95	0,80	0,60	0,97	0,89	0,77

Analisando os valores das medidas de concordância para as diferentes categorias de diagnósticos, observamos que elas nos levariam a conclusões distintas. Nos próximos capítulos abordaremos a medida de concordância kappa, que também está sendo apresentada na Tabela 1.5 para o exemplo em questão, a qual é tema da presente monografia.

2. COEFICIENTE KAPPA 2 x 2

2.1. Introdução

O Coeficiente kappa foi originalmente proposto como uma estatística descritiva, indicando o grau de concordância, além de chance, entre dois escores por sujeito, com base numa variável dicotômica. Contudo, o seu contexto foi expandido para mais do que duas categorias de respostas, mais do que dois escores por sujeito, para respostas que não são estritamente categóricas e para a situação na qual não está em questão a concordância.

Existe uma diversidade de coeficiente kappa, os quais nem estimam o mesmo parâmetro e nem têm a mesma distribuição amostral. Esses coeficientes tem sido usados não somente como uma estatística descritiva, mas também para inferência estatística.

2.2. Concordância x Associação

Para que possamos concluir se dois juízes concordam ou discordam em relação a um determinado diagnóstico, não somente é necessário que a questão seja a mesma, mas também que eles entendam e respondam a mesma questão, sob instruções idênticas.

Na análise estatística, podemos ter uma medida de concordância como um tipo especial de associação. Usamos uma medida com o objetivo de descrever o grau de concordância entre os escores de um juiz para um grupo de indivíduos e as observações de outro juiz para o mesmo grupo de indivíduos (fidedignidade entre os juízes). Também para medir quanto os escores de um juiz para um grupo de indivíduos, concorda com os escores do mesmo juiz e mesmos indivíduos em diferentes instantes de tempo (fidedignidade teste-reteste). No caso de escores binários (dicotômicos), dois escores concordam se ambos os juízes dão a mesma resposta para a mesma questão. As discrepâncias nas respostas que são devidas à especificação da questão, por um viés, escala ou escolha de critério, não são ignoradas no contexto de concordância, mas são ignoradas no contexto de associação.

2.3. O Contexto de Concordância: Coeficiente Kappa Intraclasse

A utilização de um índice de concordância é apropriado se pudermos assumir que, na população dos indivíduos, os m ($m \geq 2$) escores (X_1, X_2, ..., X_m) de cada indivíduo têm distribuição invariante sob todas as permutações dos índices (1, 2, ..., m). Os juízes respondem a mesma questão (daí estarmos representando as respostas por X) e qualquer discrepância entre as respostas para o mesmo indivíduo são consideradas erros.

Bloch e Kraemer (1989) propuseram o seguinte modelo de concordância para o caso de escores independentes onde X é uma variável dicotômica que pode assumir respostas X+ e X−.
Seja $P(X+) = p_i$ e $p'_i = 1 - p_i$, para todos os escores de cada sujeito i. Na população de sujeitos, seja $E(p_i) = P$ e $var(p_i) = \sigma^2_P$. Então, o coeficiente kappa intraclasse é definido por

$$\kappa_I = \frac{\sigma^2_P}{P\,P'} \; ; \qquad \text{onde } P' = 1 - P. \tag{2.1}$$

Na Tabela 2.1 apresentamos o modelo teórico para uma variável resposta dicotômica quando existem somente dois escores independentes por indivíduo. A probabilidade de que exista concordância entre os juízes para o indivíduo i é $p_i^2 + (1 - p_i)^2$ e para a população de indivíduos, a concordância esperada é

$$E\left[p_i^2 + (1 - p_i)^2\right] = 2\sigma_P^2 + P'^2.$$

Tabela 2.1 - Modelo Teórico para
o Coeficiente Kappa Intraclasse 2 x 2

Juiz 2	Juiz 1 X+	Juiz 1 X-	Total
X+	$E\,(p_i^2)$	$E\,(p_i\ p'_i)$	P
X-	$E\,(p_i\ p'_i)$	$E\,(p'^2_i)$	P'
Total	P	P'	1

$p'_i = 1 - p_i$, $P' = 1 - P$

Se a probabilidade de concordância é $P^2 + P'^2$, então dizemos que existe apenas concordância aleatória (por chance). Assim, o coeficiente Kappa usualmente definido como

$$\frac{(\text{Concordância Esperada - Concordância Aleatória})}{(\text{Máxima Concordância Esperada - Concordância Aleatória})} = \frac{\sigma_P^2}{P\ P'}$$

é equivalente para a população definida acima. O coeficiente de correlação Kappa Intraclasse, κ_I, é recomendado quando os indivíduos são independentemente julgados através do mesmo instrumento de medida para avaliar a fidedignidade ou reprodutividade do instrumento.

Definimos em (2.1) o coeficiente kappa intraclasse, como

$$\kappa_I = \frac{\sigma_P^2}{P\,P'} = \frac{E\,(p_i^2) - P^2}{P\,P'} \;; \qquad P' = 1 - P.$$

Fazendo,

$$\text{Var}\,(p_i) = E\,(p_i^2) - \left[\,E\,(p_i)\,\right]^2 = E\,(p_i^2) - P^2$$

temos: $E\,(p_i^2) = P^2 + \kappa_I P\,P' = P(P + P'\kappa_I)$.

$$E\,(p_i\;p_i') = P\,P'(1 - \kappa_I)$$

$$E\,(p_i'\;p_i') = P'^2 + \kappa_I P\,P'.$$

Assim, podemos definir o modelo da Tabela 2.1, através da seguinte forma equivalente:

Tabela 2.2 - Probabilidade Esperada para as Respostas Conjuntas: Modelo de Concordância

Juiz 2	Juiz 1: X+	Juiz 1: X-	Total
X+	$P^2 + \kappa_I PP'$	$PP'\,(1 - \kappa_I)$	P
X-	$PP'\,(1 - \kappa_I)$	$P'^2 + \kappa_I PP'$	P'
Total	P	P'	1

Vamos supor, por exemplo, que através de uma amostra de n indivíduos foram obtidas as seguintes frequências de respostas:

Tabela 2.3 - Frequências Observadas

Juiz 2	Juiz 1		Total
	X+	X-	
X+	n_1	n_2	$n_1 + n_2$
X-	n_3	n_4	$n_3 + n_4$
Total	$n_1 + n_3$	$n_2 + n_4$	n

A partir dessa tabela de frequências podemos obter os estimadores de máxima verossimilhança para P e κ_I, como segue:

$$L\left(P,\kappa_I / n_1,n_2,n_3,n_4\right) = \left[P^2 + \kappa_I PP'\right]^{n_1} + \left[PP'(1-\kappa_I)\right]^{n_2+n_3} + $$
$$+ \left[P'^2 + \kappa_I PP'\right]^{n_4} \qquad (2.2)$$

Tomando o logaritmo da função de máxima verossimilhança, obtemos

$$\ln L\left(P,\kappa_I/n_1,n_2,n_3,n_4\right)=n_1 \ln\left[P^2+\kappa_I PP'\right]+$$
$$+ (n_2 + n_3) \ln \left[PP'(1-\kappa_I)\right] +$$
$$+ n_4 \ln \left[P'^2 + \kappa_I PP'\right] \qquad (2.3)$$

Então, derivando a equação (2.3) em relação a κ_I e P e igualando a zero, obtemos os estimadores de máxima verossimilhança, $\hat{\kappa}_I$ e $\hat{P}$, de κ e P, respectivamente. Isto é,

$$\frac{\partial \ln L}{\partial P} = \frac{n_1 \left[2P(1-\kappa_I) + \kappa_I\right]}{P^2 + \kappa_I PP'} + \frac{(n_2+n_3)\left[(1-2P)(1-\kappa_I)\right]}{PP'(1-\kappa_I)} -$$

$$- \frac{n_4 \left[2P'(1-\kappa_I) + \kappa_I\right]}{P'^2 + \kappa_I PP'} = 0 \qquad (2.4)$$

então, $$\hat{p} = \frac{2n_1 + n_2 + n_3}{2n} \tag{2.5}$$

e, $$\frac{\partial \ln L}{\partial \kappa_I} = \frac{n_1 PP'}{P^2 + \kappa_I PP'} - \frac{(n_2 + n_3)PP'}{PP'(1 - \kappa_I)} + \frac{n_4 PP'}{P'^2 + \kappa_I PP'} = 0 \tag{2.6}$$

então, $$\hat{\kappa}_I = \frac{4(n_1 n_4 - n_2 n_3) - (n_2 - n_3)^2}{(2n_1 + n_2 + n_3)(2n_4 + n_2 + n_3)} \tag{2.7}$$

Podemos re-escrever o coeficiente de concordância kappa intraclasse, como

$$\kappa_I = \frac{P_o - P_c}{P_{max} - P_c} \tag{2.8}$$

onde: $p_o = \dfrac{n_1 + n_4}{n}$, $\quad p_c = \hat{p}^2 - \hat{p}'^2$ e $p_{max} = 1$.

De acordo com Winer, o estimador $\hat{\kappa}_I$ é idêntico ao estimador do coeficiente de correlação intraclasse para variáveis dicotômicas e foi proposto por Scott como uma medida de concordância entre dois juízes quando eles respondem à mesma questão com instruções idênticas.

Com o objetivo de determinar a variância de κ_I, usaremos o seguinte resultado proposto por Fisher. Seja $T(n_1, n_2, \ldots, n_k)$ qualquer função das frequências observadas de uma amostra de tamanho n, com distribuição k-nomial com parâmetros $w_1, w_2, \ldots, w_k$ $\left[\sum n_i = n \text{ e } \sum w_i = 1\right]$. Então, assintoticamente,

$$\frac{1}{n} \operatorname{Var}(T) = \sum_{i=1}^{k} w_i \left[\frac{\partial T}{\partial n_i}\right]^2 - \left[\frac{\partial T}{\partial n}\right]^2$$

com $n_i = n\, w_i$.

De acordo com (2.7), temos $T(n_1, n_2, n_3, n_4) = \kappa_I$.
Calculando $\frac{\partial T}{\partial n} = 0$, obtemos:

$$n \operatorname{Var}(\hat{\kappa}_I) = (1 - \kappa_I) \left[(1 - \kappa_I)(1 - 2\kappa_I) + \frac{\kappa_I(2 - \kappa_I)}{2PP'} \right] \qquad (2.9)$$

Transformação para Estabilização da Variância

Kendall e Stuart propuseram a seguinte transformação para a estabilização da variância de $\hat{\kappa}_I$:

$$z(\kappa_I) = \int [\operatorname{Var}(\kappa_I)]^{-1/2} \, d\kappa_I.$$

Adaptando os resultados de Kraemer (1986) para $\hat{p} \neq 1/2$, definimos:

$$\kappa_L = -\operatorname{Min}(\hat{p} / \hat{p}', \hat{p}' / \hat{p});$$

$$t = 2(3 - 10\,\hat{p}\hat{p}') / (1 - 4\,\hat{p}\hat{p}');$$

$$\kappa_o = \frac{\left(t - \sqrt{t^2 - 24}\right)}{6}$$

$$V(\kappa) = (1 - \kappa) \left[(1 - \kappa)(1 - 2\kappa) + \frac{\kappa(2 - \kappa)}{\hat{p}\,\hat{p}'} \right];$$

$$V(\kappa_L) = V_L;$$

$$V(\kappa_o) = V_o;$$

$$S_U = 1 / (1 - \kappa_o);$$

$$S_L = \left[1 - V_L / V_o\right]^{1/2} / (\kappa_o - \kappa_L).$$

Então, uma transformação aproximada para a estabilização da variância de $\hat{\kappa}_I$ é dada por:

$$Z = \begin{cases} 1 \Big/ \left[S_U \sqrt{V_o}\,\right] \text{arcsen} \left[S_U (\hat{\kappa}_I - \kappa_o)\right], \text{ se } \hat{\kappa}_I \geq \kappa_o \\ 1 \Big/ \left[S_L \sqrt{V_o}\,\right] \text{arcsen} \left[S_L (\hat{\kappa}_I - \kappa_o)\right], \text{ se } \hat{\kappa}_I \leq \kappa_o \end{cases} \qquad (2.10)$$

se $\hat{p} = 1/2$, então

$$Z = \text{arcsen} (\hat{\kappa}_I) \qquad (2.11)$$

Com a utilização do método delta (Bishop ,Fienberg e Holland, 1975, 14.6) para calcular a variância de Z para ordem $1/n$, foi constatado que a variância de $\hat{\kappa}_I$ depende grandemente dos valores de P e κ_I, enquanto que a variância de Z é completamente estável para valores de κ e P que não sejam extremos (Bloch e Kraemer, 1989).

Um Estimador Jackknife de κ_I

Jackknife é uma técnica estatística proposta originalmente por Quenouille em 1956, a qual é usada para reduzir o viés na estimação de parâmetros e para estimar o erro padrão desses parâmetros. Através desta técnica, podemos obter um estimador de κ_I, dado por:

$$\hat{\kappa}_J = \sum_{i=1}^{4} \frac{n_i}{n} J_i$$

onde

$$J_1 = \frac{\hat{\kappa}_I \left[1 - \frac{n}{n-1}\left(\hat{p}'/\hat{p}\right)\right] + \hat{p}'/\hat{p}}{1 - \frac{1}{n-1}\left(\hat{p}'/\hat{p}\right)},$$

$$J_2 = J_3 = \frac{\hat{\kappa}_I \left(1 - \frac{1}{4\hat{p}\hat{p}'}\right) - \left(\frac{n-1}{2n-1}\right)\left(1 - \frac{1}{4n\hat{p}\hat{p}'}\right)}{\frac{n}{2n-1} - \frac{1}{4n\hat{p}\hat{p}'}},$$

$$\text{e,} \quad J_4 = \frac{\hat{\kappa}_I \left[1 - \frac{n}{n-1}\left(\hat{p}/\hat{p}'\right)\right] + \hat{p}/\hat{p}'}{1 - \frac{1}{n-1}\left(\hat{p}/\hat{p}'\right)}$$

e onde $\hat{p}' = 1 - \hat{p}$ e $\hat{p}$ é o estimador de máxima verossimilhança de P, dado pela equação (2.5).

J_1, por exemplo, é obtido pela omissão de um dos n_1 indivíduos para os quais ambos os juízes respondem X+ e então calcula-se κ_I com os n - 1 indivíduos restantes.

A variância de $\hat{\kappa}_J$ é estimada por:

$$\hat{\text{Var}}\,(\hat{\kappa}_J) = \sum_{i=1}^{4} n_i \left(J_i - \hat{\kappa}_J\right)^2 \Big/ n\,(n-1) \qquad (2.13)$$

Pode-se mostrar que

$$\lim_{n \to \infty} n\,\hat{\text{Var}}\,(\hat{\kappa}_J) = (1 - \kappa_I)\left[(1 - \kappa_I)(1 - 2\kappa_I) + \kappa_I(2 - \kappa_I)/(2PP')\right]$$

Bloch e Kraemer(1989), comentam as situações nas quais ocorrerão problemas com os estimadores $\hat{\kappa}_I$ e $\hat{\kappa}_J$ e apresentam os resultados das simulações realizadas para verificar as propriedades desses estimadores com pequenas amostras. Ocorrerão problemas quando o tamanho da amostra é pequeno, o valor de P está próximo de 0 ou 1 ou κ_I está próximo de 0 ou 1.

Se as n observações amostrais estão numa cela da diagonal principal, o coeficiente $\hat{\kappa}_I$ não está definido, caracterizando um resultado degenerado. Se todas estiverem na diagonal principal, mas nem todas na mesma cela, o valor de $\hat{\kappa}_I = 1$ e a aproximação da distribuição amostral pela distribuição normal, não é adequada.

Portanto, quando o valor de P estiver próximo de 0 ou 1 ou κ_I estiver próximo de 0 ou 1, devemos usar uma aproximação da distribuição amostral de $\hat{\kappa}_I$ ou $\hat{\kappa}_J$ pela distribuição normal (assintoticamente), somente se o tamanho da amostra for suficientemente grande, tal que a ocorrência de uma valor extremo ou degenerado da estatística κ seja altamente improvável.

No tocante aos resultados das simulações, cabe resaltar: o viés do estimador Jackknife, $\hat{\kappa}_J$, é menor do que o viés do estimador da máxima verossimilhança, $\hat{\kappa}_I$, o qual tende a subestimar κ_I quando o tamanho da amostra é pequeno. Outras comparações, com respeito a assimetria, curtose e a transformação Z para a estabilização da variância de $\hat{\kappa}_I$, podem ser obtidas com maiores detalhes em Bloch e Kraemer (1989).

Ainda segundo Block e Kraemer (1989), se estivérmos trabalhando com pequenas amostras, é recomendada a utilização do estimador Jackknife de κ_I, não somente porque seu viés é menor, mas também porque a estimativa da variância amostral por ele obtida é mais precisa. Procedimentos que requerem a distribuição teórica, tais como intervalos de confiança, determinação do poder do teste ou formulações do teste, são melhores se usarmos uma aproximação normal da transformação Z de kappa.

2.4. O contexto de Associação: Coeficiente Kappa Ponderado

Quando existem dois escores (possivelmente independentes) por indivíduo, isto é: X (com categorias de resposta X+ e X-) e Y (com respostas Y+ e Y-), para cada indivíduo i, seja:

$p_i = P\ (X+)$, $P = E\ (p_i)$ e

$q_i = P\ (Y+)$, $Q = E\ (q_i)$.

Então, para escores independentes, a probabilidade que ambos os juízes atribuam uma resposta "+" para o indivíduo i é $E(p_i\ q_i)$.

Na Tabela 2.4 consta o modelo geral, apresentado por Bloch e Kraemer (1989), para o contexto de associação.

Tabela 2.4 - Modelo de Associação para uma tabela 2x2

Juiz 2	Juiz 1		Total
	Y+	Y-	
X+	$E\ (p_i\ q_i)$	$E\ (p_i\ q'_i)$	P
X-	$E\ (p'_i\ q_i)$	$E\ (p'_i\ q'_i)$	P'
Total	Q	Q'	1

onde $p'_i = 1 - p_i$, $q'_i = 1 - q_i$, $P' = 1 - P$ e $Q' = 1 - Q$.

A família dos estimadores kappa ponderados, a qual incorpora as perdas relativas dos dois tipos de discordâncias num número $(0 \leq r \leq 1)$ como uma medida de associação aproximada para o modelo é abordagem em Bloch e Kraemer (1989). Suponha que para cada um dos quatro resultados existe alguma utilidade associada, que pode ser expressa em termos de dinheiro, importância, etc., conforme a Tabela 2.5. Nessa forma, a utilidade de ambos os juízes darem uma resposta

"+" é A, e assim por diante. Necessitamos que A, B, C e D sejam tais que $A \geq B$ e $D \geq C$; $A - B - C + D > 0$ e P é fixado.

Tabela 2.5 - Utilidade de Respostas Associadas

Juiz 2	Juiz 1		Total
	Y+	Y−	
X+	A	B	P
X−	C	D	P'
Total	Q	Q'	1

O máximo de utilidade de associação esperada é obtida se todas as observações estiverem na diagonal principal, dada por:

Utilidade Esperada Máxima = PA + P'D.

Se a associação entre as variáveis é devida ao acaso, então a utilidade esperada é dada por:

Utilidade Aleatória = APQ + BPQ' + CP'Q + DP'Q'.

A utilidade esperada dada pelas Tabelas 2.4 e 2.5 é

$$\text{Util. Esperada} = A \cdot E(p_i\ q_i) + B\ E(p_i\ q_i') + C\ E(p_i'\ q_i) + D\ E(p_i'\ q_i').$$

Portanto, podemos definir o coeficiente kappa no contexto de associação, como segue:

$$\frac{(\text{Utilidade Esperada} - \text{Utilidade Aleatória})}{(\text{Utilidade Esperada Máxima} - \text{Utilidade Aleatória})},$$ o qual é igual

a $$\kappa(r) = \frac{E\,(p_i\ q_i) - PQ}{PQ'r + P'Qr'} \qquad (2.14)$$

onde $r = \dfrac{A - B}{A - B - C + D}$ $\qquad r' = 1 - r.$

Se, ao invés, exigirmos que $A \geq C$ e $D \geq B$; $A - C - B + D > 0$ e Q fixado, então temos $\kappa(r)$ dado por (2.14), com

$$r = \frac{D - B}{A - B - C + D} \quad .$$

Assim, o índice r reflete a importância relativa das duas classificações da diagonal cruzada (X-Y+) versus (X+Y-), quando o objetivo é avaliar a associação de X em relação ao padrão de Y, ou vice-versa.

Podemos apresentar o coeficiente $\kappa(r)$ como uma média ponderada de $\kappa(1)$ e $\kappa(0)$, para qualquer r:

$$\kappa(r) = w(r)\ \kappa(1) + \left[1 - w(r)\ \kappa(0)\right]$$

onde $w(r) = \dfrac{PQ'r}{PQ'r + P'Qr'}$.

De acordo com Bloch e Kraemer (1989), muitas das medidas de associação para tabelas 2 x 2, com 0 indicando associação aleatória e 1 indicando associação perfeita, corresponde a um ou outro parâmetro kappa ponderado.

Descrevendo o modelo apresentado na Tabela 2.4 em termos de $\kappa(r)$, obtemos a tabela abaixo:

Tabela 2.6 - Probabilidade Esperada de Respostas Conjuntas para o Modelo de Associação

Juiz 2	Juiz 1		Total
	Y+	Y-	
X+	$PQ + \kappa(r)\left[rPQ' + r'P'Q\right]$	$PQ' - \kappa(r)\left[rPQ' + r'P'Q\right]$	P
X-	$QP' - \kappa(r)\left[rPQ' + r'P'Q\right]$	$P'Q' + \kappa(r)\left[rPQ' + r'P'Q\right]$	P'
Total	Q	Q'	1

Tabela 2.7 - Frequências Observadas

Juiz 2	Juiz 1		Total
	Y+	Y-	
X+	n_1	n_2	$n_1 + n_2$
X-	n_3	n_4	$n_3 + n_4$
Total	$n_1 + n_3$	$n_2 + n_4$	n

Se nos n escores independentes, as frequências observadas são dadas como na Tabela 2.7, então a função de verossimilhança é dada por:

$$L\left[P,\ Q,\ \kappa(r) / n_1,\ n_2,\ n_3,\ n_4\right] = \left[PQ + \kappa(r)\,(rPQ' + r'P'Q)\right]^{n_1} +$$

$$+ \left[PQ' - \kappa(r)\,(rPQ' + r'P'Q)\right]^{n_2} + \left[QP' - \kappa(r)\,(rPQ' + r'P'Q)\right]^{n_3} +$$

$$+ \left[P'Q' + \kappa(r)\ (rPQ' + r'P'Q)\right]^{n_4} . \qquad (2.15)$$

Fazendo o logaritmo da função de verossimilhança, obtemos:

$$\ln L = n_1 \ln \left[PQ + \kappa(r)\ (rPQ' + r'P'Q)\right] +$$

$$+ n_2 \ln \left[PQ' - \kappa(r)\ (rPQ' + r'P'Q)\right] +$$

$$+ n_3 \ln \left[QP' - \kappa(r)\ (rPQ' + r'P'Q)\right] +$$

$$+ n_4 \ln \left[P'Q' + \kappa(r)\ (rPQ' + r'P'Q)\right]. \qquad (2.16)$$

Afim de obter os estimadores de máxima verossimilhança de P, Q e $\kappa(r)$, deriva-se ln L em relação a P,Q e $\kappa(r)$, respectivamente, igualando a zero, isso é: $\frac{\partial}{\partial P} \ln L = 0$, $\frac{\partial}{\partial Q} \ln L = 0$ e $\frac{\partial}{\partial \kappa(r)} \ln L = 0$. Daí, obtém-se

$$\hat{p} = (n_1 + n_2)/n , \qquad (2.17)$$

$$\hat{q} = (n_1 + n_3)/n , \text{ e} \qquad (2.18)$$

$$\hat{\kappa}(r) = \frac{(n_1 n_4 - n_2 n_3)}{r\ (n_1 + n_2)(n_2 + n_4) + r'\ (n_3 + n_4)(n_1 + n_3)} \qquad (2.19)$$

O coeficiente $\hat{\kappa}(r)$ é chamado de "estatística kappa ponderado". Para o caso no qual $r = 1/2$, temos $\hat{\kappa}(1/2)$ que é conhecido como "kappa de Cohen", o qual será abordado mais adiante.

Bloch e Kraemer (1989) descrevem alguns resultados assintóticos para este caso.

2.5. Coeficiente Kappa de Cohen e outros Coeficientes Kappa

De acordo com Bloch e Kraemer (1989), o coeficiente kappa tem sido descrito como uma "medida de concordância corrigida por chance", sendo geralmente definido por:

$$\kappa = \frac{(p_o - p_c)}{(p_m - p_c)} \tag{2.20}$$

onde p_o é a probabilidade de uma resposta X+Y+ ou X-Y-, p_c define um valor por chance ou aleatório e p_m o valor máximo.

Para os valores das marginais (P,Q) fixados, para uma decisão aleatória teríamos:

$$p_c = PQ + P'Q'$$

e o valor máximo para p_o seria $p_m = 1 - |P - Q|$.
Substituido esses valores na equação (2.20), temos:

i) se $P > Q$, $\kappa = \kappa(0)$;

ii) se $P < Q$, $\kappa = \kappa(1)$;

iii) se $P = Q$, $\kappa = \kappa(r)$, para qualquer r.

Mais genericamente, seja $\Omega = (P^*, Q^*)$, tal que (P,Q) esteja em Ω. Seja

$$p_c^\Omega = \underset{\Omega}{\mathrm{Max}} \left[P^* Q^* + P^{*'} \; Q^{*'} \right],$$

$$p_m^\Omega = \underset{\Omega}{\mathrm{Max}} \left[1 - |P^* - Q^*|\right] \quad \text{e,}$$

$$\kappa_\Omega = \frac{p_o - p_c^\Omega}{p_m^\Omega - p_c^\Omega}.$$

Consideremos alguns exemplos:

1) Se Ω foi definido como um valor constante, $P = Q$, então

$$p_c^{\Omega} = P^2 + P'^2 = Q^2 + Q'^2 \ , \quad p_m^{\Omega} = 1 \quad e \quad \kappa_{\Omega} = \kappa_I$$

2) Se Ω foi definido por uma série de valores (P^*, Q^*), com

$P^* + Q^* = P + Q$, então

$$p_c^{\Omega} = \frac{(P + Q)^2}{4} + \frac{(P' + Q')^2}{4} \ , \qquad p_m^{\Omega} = 1$$

e
$$\kappa_{\Omega} = \frac{E\,(p_i q_i) - \left(\frac{P + Q}{2}\right)^2}{\left(\frac{P + Q}{2}\right)\left(\frac{P' + Q'}{2}\right)} .$$

A estimativa amostral deste kappa seria idêntica para $\hat{\kappa}_I$, mas como P e Q não são iguais a teoria amostral diferirá.

3) Se Ω foi definido como a série de todos (P^*, Q^*), com

$P^*Q^* + P^{*'}Q^{*'} \leq PQ + P'Q'$, onde P e Q são iguais ou encontram-se no mesmo lado em relação a $1/2$ (cujo caso $PQ + P'Q' \geq 1/2$), então

$$p_c^{\Omega} = PQ + P'Q' \quad , \quad p_m^{\Omega} = 1$$

e, portanto, $\kappa_{\Omega} = \kappa(1/2)$ o qual é estimador do coeficiente kappa. Este estimador é o mais comumente usado e é chamado kappa de Cohen.

4) Se Ω foi definido como a série de todos (P^*, Q^*), com $P^*Q^* + P^{*'}Q^{*'} \leq PQ + P'Q'$, onde P e Q encontram-se em lados

opostos em relação a 1/2 (cujo caso $PQ + P'Q' < 1/2$), então

$$p_c^{\Omega} = PQ + P'Q' \quad , \qquad p_m^{\Omega} = 1 - \sqrt{1 - 2(PQ + P'Q')} \quad ,$$

e portanto,

$$\kappa_{\Omega} = \kappa(r) \ , \quad \text{para} \quad r = \frac{PQ' - \sqrt{1 - 2(PQ + P'Q')}}{(P - Q)} \quad ,$$

o qual não é o coeficiente kappa de Cohen, pois $r \neq 1/2$.

5) Se Ω foi definido como a série de todos (P^*, Q^*), com

$P^* - Q^* = P - Q$, então

$$p_c^{\Omega} = \left[1 - (P - Q)^2\right] \Big/ 2 \quad ; \qquad p_m^{\Omega} = 1 - |P - Q| .$$

Neste caso κ_{Ω} não é igual nem a κ_I e nem a $\kappa(r)$, para qualquer r.

Dessa forma, para cada definição de Ω, obteremos diferentes coeficientes kappa, que podem corresponder a $\kappa(r)$ para algum valor de r, ou corresponder a κ_I, ou diferir completamente deles.

Ainda segundo Bloch e Kraemer (1989), quando o contexto é de concordância recomenda-se usar o coeficiente $\hat{\kappa}_I$ e, quando existe uma definição apropriada de utilidade de associação, recomenda-se usar o coeficiente $\hat{\kappa}(r)$, para algum valor relevante de r. Se a associação está sendo avaliada contra todas as possíveis associações com as marginais fixadas, recomenda-se utilizar o coeficiente $\hat{\kappa}(0)$ se $P \geq Q$ e $\hat{\kappa}(1)$ se $P \leq Q$.

No capítulo 1 foram apresentadas várias medidas de concordância aplicadas ao exemplo da Tabela 1.2, cujos valores são mostrados na Tabela 1.5. Como mencionado anteriormente, essas medidas de concordância nos levariam a conclusões distintas; portanto não são totalmente adequadas. Foi apresentada, também, uma medida de concordância kappa, que será discutida a seguir.

De acordo com Fleiss (1981), exceto nas situações extremas (nas quais $p_1=q_2=0$ ou $p_2=q_1=0$), algum grau de concordância seria esperado somente por chance, como mostra o modelo da Tabela 2.8. Por exemplo, se um juiz A utiliza uma série de critérios para decidir sobre a ausência ou presença de uma condição e um juiz B utiliza outro grupo de critérios distintos e independentes, então toda a concordância observada seria explicada somente por chance (aleatória).

Tabela 2.8 - Proporção Esperada por Chance das Respostas Conjuntas dos 2 Juízes, para o Modelo da Tabela 1.3

	Juiz B		
Juiz A	Categoria de Interesse	Outras	Total
Categoria de Interesse	p_1p_2	p_1q_2	p_1
Outras	p_2q_1	q_1q_2	q_1
Total	p_2	q_2	1

Ainda segundo Fleiss (1981), uma maneira natural de introduzir a "correção por chance" é a seguinte: consideremos um índice que assume o valor o 1 quando existe perfeita concordância, denotado por I_o, cujo valor observado é calculado através das proporções da Tabela 1.3. Seja I_e o valor da concordância esperada somente por chance, calculado a partir das proporções da Tabela 2.8. A concordância obtida, além daquela esperada por chance, é denotada por $I_o - I_e$ e o máximo de concordância observada, além daquela por chance, é denotada por $1 - I_e$. Assim, Fleiss definiu uma medida de concordância, denominada kappa, como

$$\hat{\kappa} = \frac{I_o - I_e}{1 - I_e} \tag{2.21}$$

Se existe perfeita concordância, então $\hat{\kappa} = +1$. Se a concordância observada é maior ou igual a concordância por chance, então $\hat{\kappa} \geq 0$ e, se for menor ou igual do que aquela por chance, então $\hat{\kappa} \leq 0$. O valor máximo de $\hat{\kappa}$ depende das proporções marginais da Tabela 2.8. Se $I_e = 1/2$, então o valor mínimo será $\hat{\kappa} = -1$, caso contrário, estará compreendido entre -1 e 0. Pode-se mostrar que $\hat{\kappa}$ definido em (2.21) é idêntico a

$$\hat{\kappa} = \frac{2(ad - bc)}{p_1 q_2 + p_2 q_1} \tag{2.22}$$

Dessa forma, para a categoria de diagnóstico "Neurose" do exemplo apresentado no capítulo anterior, temos:

Tabela 2.9 - Proporção Esperada por Chance das Respostas Conjuntas para a Categoria de Diagnóstico "Neurótico"

Juiz A	Juiz B		Total
	Neurótico	Outros	
Neurótico	0,005	0,095	0,10
Outros	0,045	0,855	0,90
Total	0,05	0,95	1,00

Então, das proporções da Tabela 1.4, temos

$$\hat{\kappa} = \frac{2(0{,}04 \times 0{,}89 - 0{,}06 \times 0{,}01)}{0{,}10 \times 0{,}95 + 0{,}05 \times 0{,}90} = 0{,}50$$

Este valor e os outros dois, relativos às categorias de diagnóstico "Psicótico" e "Orgânico" são mostrados na Tabela 1.5, indicando que a concordância é melhor no tocante à desordem orgânica, menor em psicose e pobre no que se refere à neurose.

De acordo com Fleiss, esta estatística kappa pode ser interpretada como um coeficiente de correlação intraclasse, o qual é grandemente usado como medida de fidedignidade entre juízes para o caso de escores quantitativos.

Landis e Koch (1977a) propuseram a seguinte interpretação para os valores de kappa, com respeito ao grau de concordância além de chance:

Valor da Estatística Kappa	Poder da Concordância
< 0,00	Muito Pobre, Fraco
0,00 a 0,20	Pequeno
0,21 a 0,40	Médio, Regular, Razoável
0,41 a 0,60	Moderado
0,61 a 0,80	Substancial, Considerável
0,81 a 1,00	Quase Perfeito

3. EXTENSÃO DO COEFICIENTE KAPPA

3.1. Coeficiente Kappa para 2 Juízes e K Diagnósticos

Vimos no item 2.5 a definição da estatística kappa proposta por Fleiss (1981). Para o exemplo da Tabela 1.2, determinamos os valores do coeficientes de concordância kappa, para as três categorias de diagnósticos, os quais são mostrados na Tabela 1.5. Contudo, as vezes queremos uma medida de concordância através de todas categorias de diagnósticos. Para tanto, Fleiss (1981) sugeriu um valor "kappa total", definido como uma média ponderada dos kappas individuais, no qual os pesos são os denominadores dos kappas individuais $\left[\text{isso é, a quantidade } p_1q_2 + p_2q_1 \text{ em } (2.22)\right]$. Contudo, através da Tabela 3.1, podemos chegar a uma fórmula mais sugestiva:

Tabela - Proporções dos Escores Conjuntos
para 2 Juízes, com K Categorias de Diagnósticos

Juiz A	Juiz B 1	2	. . .	K	Total
1	p_{11}	p_{12}	. . .	p_{1K}	$p_{1.}$
2	p_{21}	p_{22}	. . .	p_{2K}	$p_{2.}$
.	.	.	. . .	.	.
.	.	.	. . .	.	.
.	.	.	. . .	.	.
K	p_{K1}	p_{K2}	. . .	p_{KK}	p_K
Total	$p_{.1}$	$p_{.2}$	. . .	$p_{.K}$	1

Seja a proporção total de concordância observada, p_o,

$$p_o = \sum_{i=1}^{K} p_{ii} \qquad (3.1)$$

e seja p_c a proporção total de concordância esperada por chance,

$$p_e = \sum_{i=1}^{K} p_{i.}\, p_{.i} \tag{3.2}$$

Então, o valor de kappa total (Fleiss, 1981) é definido por

$$\hat{\bar{\kappa}} = \frac{p_o - p_e}{1 - p_e} \tag{3.3}$$

Para testar a hipótese de que os escores são independentes, Fleiss, Cohen e Everitt mostraram que o erro padrão de kappa pode ser estimado por

$$s.e._o(\hat{\bar{\kappa}}) = \frac{1}{(1 - p_e)\sqrt{n}} \sqrt{p_e + p_e^2 - \sum_{i=1}^{K} p_{i.}\, p_{.i}(p_{i.} + p_{.i})} \tag{3.4}$$

A hipótese pode ser testada contra a alternativa de que a concordância é maior do que somente por chance, através da estatística de teste

$$z = \frac{\hat{\bar{\kappa}}}{s.e._o(\hat{\bar{\kappa}})} \sim N(0,1) \tag{3.5}$$

Um teste unilateral é mais apropriado do que um teste bilateral. Assim, temos

$H_o : \bar{\kappa} = 0$

$H_a : \bar{\kappa} > 0$

Aplicando aos dados do exemplo da Tabela 1.2, com uma amostra de 100 indivíduos, temos:

$$p_o = \sum_{i=1}^{3} p_{ii} = 0{,}75 + 0{,}04 + 0{,}10 = 0{,}89$$

e

$$p_e = \sum_{i=1}^{3} p_{i.}\, p_{.i} = 0{,}80 \times 0{,}80 + 0{,}10 \times 0{,}05 + 0{,}10 \times 0{,}15 = 0{,}66$$

daí,

$$\hat{\bar{\kappa}} = \frac{p_o - p_e}{1 - p_e} = \frac{0{,}89 - 0{,}66}{1 - 0{,}66} = 0{,}68$$

e,

$$s.e._o(\hat{\bar{\kappa}}) = \frac{1}{(1 - 0{,}66)\sqrt{100}} \sqrt{0{,}66 + (0{,}66)^2 - 1{,}0285} = 0{,}076$$

Precisamos, agora, realizar o teste de hipótese a fim de verificar se o valor de $\bar{\kappa}$ é significativamente diferente de zero, isto é, se a concordância observada é muito maior do que aquela que seria esperada por chance.

Temos:

H_o : Os escores são independentes, isto é, a concordância observada é somente esperada por chance.

H_a : A concordância observada é superior àquela esperada somente por chance.

ou, H_o: $\bar{\kappa} = 0$ vs. H_a: $\bar{\kappa} > 0$

sob H_o, temos,

$$z = \frac{\hat{\bar{\kappa}}}{s.e._o(\hat{\bar{\kappa}})} = \frac{0{,}68}{0{,}076} = 8{,}95$$

Portanto, rejeitamos H_o: $\bar{\kappa} = 0$, isto é, que os escores são independentes. Isso indica que o grau de concordância é superior àquela esperada por chance.

No caso particular em que K = 2 categorias de diagnósticos, se aplicarmos as fórmulas (3.1) - (3.5), estaremos estudando a fidedignidade de cada categoria. Assim, novamente para os dados do exemplo da Tabela 1.2 e, juntando as categorias de diagnósticos que não aquela de interesse, foram obtidos os seguintes valores:

Tabela 3.2 - Valor de Kappa para cada Categoria Individual e Kappa total para o exemplo da Tabela 1.2

Categoria de Diagnóstico	p_o	p_e	$\hat{\kappa}$	$s.e._o(\hat{\kappa})$	z
Psicótico	0,90	0,68	0,69	0,100	6,90
Neurótico	0,93	0,86	0,50	0,093	5,38
Orgânico	0,95	0,78	0,77	0,097	7,94
Total	0,89	0,66	0,68	0,076	8,95

Assim, por exemplo, para a categoria de diagnóstico "Neurótico", através da Tabela 1.4, podemos obter:

$$p_o = 0,04 + 0,89 = 0,93;$$

$$p_e = 0,05 \times 0,10 + 0,95 \times 0,90 = 0,86;$$

$$\hat{\kappa} = \frac{0,93 - 0,86}{1 - 0,86} = 0,50;$$

$$s.e._o(\hat{\kappa}) = \frac{1}{(1 - 0,86)\sqrt{100}} \sqrt{0,86 + (0,86)^2 - 1,58} = 0,093;$$

$$\text{e,} \quad z = \frac{0,50}{0,093} = 5,38.$$

A interpretação desses valores, bem como daqueles das categorias de diagnósticos "Psicótico" e "Orgânico", apresentados na Tabela 3.2, sugerem que a concordância observada é superior àquela esperada somente por chance.

Cabe ser observado, ainda, que o valor da estatística kappa total $(\hat{\bar{\kappa}})$, pode ser obtida pela soma das diferenças individuais $p_o - p_e$, dividida pela soma das diferenças individuais $1 - p_e$, confirmando que $\hat{\bar{\kappa}}$ é uma média ponderada dos coeficientes individuais $\hat{\kappa}$ de cada categoria, isto é:

$$\hat{\bar{\kappa}} = \frac{(0,90 - 0,68)+(0,93 - 0,86)+(0,95 - 0,78)}{(1 - 0,68)+(1 - 0,86)+(1 - 0,78)} = 0,68 \ .$$

Fleiss (1981) apresenta, também, a estatística de teste para o caso no qual queremos testar se o valor observado de kappa [tanto kappa total $(\hat{\bar{\kappa}})$ quando os kappas individuais $(\hat{\kappa})$] é igual a um valor pré fixado κ $(\kappa \neq 0)$, conforme segue:

$$s.e.(\hat{\kappa}) = \frac{\sqrt{A + B - C}}{(1-P_e)\sqrt{n}} \qquad (3.6)$$

onde $$A = \sum_{i=1}^{K} p_{ii} \left[1 - (p_{i.} + p_{.j})(1 - \hat{\kappa})\right]^2, \qquad (3.7)$$

$$B = (1 - \hat{\kappa})^2 \sum_{i \neq j}\sum p_{ij} (p_{.i} + p_{j.})^2 \ , \qquad (3.8)$$

e, $$C = \left[\hat{\kappa} - p_e (1 - \hat{\kappa})\right]^2 \qquad (3.9)$$

A hipótese será rejeitada se o valor da estatística de teste

$$z = \frac{| \hat{\kappa} - \kappa |}{s.e.(\hat{\kappa})} \sim N(0,1) \qquad (3.10)$$

for significativamente maior do que o valor tabelado. Um intervalo de 100 $(1 - \alpha)$% de confiança, para κ, é dado por:

$$\hat{\kappa} - C_{\alpha/2}\ s.e.(\hat{\kappa}) \leq \kappa \leq \hat{\kappa} + C_{\alpha/2}\ s.e.(\hat{\kappa}) \ . \qquad (3.11)$$

Assim, supondo que para o exemplo da Tabela 1.2 o valor de kappa total da população é $\bar{\kappa}$ = 0,80, temos:

$$H_o:\ \bar{\kappa} = 0{,}80 \quad vs. \quad H_a:\ \bar{\kappa} \neq 0{,}80 \quad com \quad \alpha = 0{,}05$$

$$A = 0{,}2995 \quad ; \quad B = 0{,}0079$$

$$C = \left[0{,}68 - 0{,}66\ (1 - 0{,}68)\right]^2 = 0{,}2198$$

$$s.e.(\hat{\bar{\kappa}}) = \frac{\sqrt{0{,}2995 + 0{,}00079 - 0{,}2198}}{(1 - 0{,}66)\sqrt{100}} = 0{,}087$$

$$z = \frac{|\ 0{,}68 - 0{,}80|}{0{,}087} = 1{,}38.$$

Portanto, não rejeitamos a hipótese de que o valor de kappa total da população é igual a $\bar{\kappa}$ = 0,80, ao nível de significância de 5 % e essa amostra. Isso indica que há evidências de que existe uma considerável "concordância total" na população.

Para o intervalo com 95 % de confiança, temos:

$$0{,}68 - 1{,}96 \times 0{,}087 \leq \bar{\kappa} \leq 0{,}68 + 1{,}96 \times 0{,}087$$

$$0{,}5095 \leq \bar{\kappa} \leq 0{,}8505$$

Isso significa que a probabilidade de que o intervalo (0,5095 ; 0,8505) contenha o verdadeiro valor do coeficiente de concordância kappa total é 0,95.

Fleiss (1981), apresenta também uma medida de concordância entre Juízes para o caso onde a seriedade relativa de cada possível discordância poderia ser quantificada. Suponhamos que, independentemente dos dados coletados, os pesos de concordância w_{ij} $(i = 1, 2, \ldots, K)$ são associados com um motivo clínico ou racional para as K^2 celas. Os pesos estão compreendidos no intervalo $0 \leq w_{ij} \leq 1$ e são tais que:

$w_{ij} = 1$; isto é, para perfeita concordância é dado o peso máximo;

$0 \leq w_{ij} < 1$; para $i \neq j$; isto é, para todos os desacordos são dados menos do que o peso máximo;

$w_{ij} = w_{ji}$; isto é , dois juízes são considerados simétricos.

A proporção de concordância ponderada observada é dada por

$$p_{o(w)} = \sum_{i=1}^{K} \sum_{j=1}^{K} w_{ij} \; p_{ij} \, , \tag{3.12}$$

e a proporção de concordância ponderada esperada por chance, é

$$p_{e(w)} = \sum_{i=1}^{K} \sum_{j=1}^{K} w_{ij} \; p_{i.} \; p_{.j} \tag{3.13}$$

Daí, segue que o coeficiente kappa ponderado é dado por

$$\hat{\kappa}_w = \frac{p_{o(w)} - p_{e(w)}}{1 - p_{e(w)}} \tag{3.14}$$

Como foi salientado por Fleiss (1981), quando $w_{ij} = 0$ para todo $i \neq j$, isto é, todos os desacordos são considerados igualmente sérios, então o coeficiente kappa ponderado é idêntico ao κappa total dado pela equação (3.3). A interpretação da magnitude do coeficiente $\hat{\kappa}_w$ é a mesma proposta por Landis e Koch (1977a).

Ainda segundo Fleiss (1981), o coeficiente de correlação intraclasse é idêntico ao coeficiente de concordância kappa ponderado (exceto por um termo de ordem $1/n$), quando os pesos são dados por

$$w_{ij} = 1 - \frac{(i - j)^2}{(K - 1)^2} \qquad (3.15)$$

O erro padrão de $\hat{\kappa}_w$ para testar a hipótese de que $\kappa_w = 0$, é dado por

$$s.e._o(\hat{\kappa}_w) = \frac{1}{(1 - p_{e(w)})\sqrt{n}} \times$$

$$\times \sqrt{\sum_{i=1}^{K} \sum_{j=1}^{K} p_{i.} p_{.j} \left[w_{ij} - (\bar{w}_{i.} + \bar{w}_{.j}) \right]^2 - p_{e(w)}^2} \qquad (3.16)$$

onde $$\bar{w}_{i.} = \sum_{j=1}^{K} p_{.j} w_{ij}, \qquad (3.17)$$

$$\bar{w}_{.j} = \sum_{i=1}^{K} p_{i.} w_{ij}. \qquad (3.18)$$

Sob H_o, temos a estatística de teste

$$z = \frac{\hat{\kappa}_w}{s.e._o(\hat{\kappa})} \sim N(0,1). \qquad (3.19)$$

Para testarmos a hipótese de que o valor do coeficiente kappa ponderado, κ_w, é igual a um valor pré-especificado $\kappa_w \neq 0$, devemos utilizar a seguinte fórmula de erro padrão de $\hat{\kappa}_w$:

$$s.e.(\hat{\kappa}_w) = \frac{1}{(1 - p_{e(w)})\sqrt{n}} \times$$

$$\times \sqrt{\sum_{i=1}^{K} \sum_{j=1}^{K} p_{ij} \left[w_{ij} - (\bar{w}_{i.} - \bar{w}_{.j})(1 - \hat{\kappa}_v)\right]^2 - \left[\hat{\kappa}_v - p_{e(v)}(1 - \hat{\kappa}_v)\right]^2}$$

(3.20)

e sob H_0 temos a estatística de teste

$$z = \frac{|\hat{\kappa}_v - \kappa_v|}{s.e.(\hat{\kappa}_v)} \sim N(0,1). \qquad (3.21)$$

Fleiss (1981) comenta que o erro padrão do coeficiente kappa ponderado, dados pelas equações (3.3) e (3.6) são casos especiais do erro padrão do coeficiente kappa ponderado, dados, respectivamente, pelas equações (3.16) e (3.20), quando $w_{ii} = 1$ e $w_{ij} = 0$, para todo $i \neq j$.

3.2. Múltiplos Escores (Juízes) por Indivíduo

Fleiss (1981) apresenta a situação na qual o número de juízes por sujeito é maior do que dois. Retira-se uma amostra de n indivíduos, com m_i sendo o número de escores para o sujeito i e, os juízes que avaliam um indivíduo, não são os mesmos que avaliam outro. Seja K = 2, isto é, os indivíduos serão classificados, por cada um dos juízes, numa das duas categorias de diagnósticos e x_i o número de escores positivos do sujeito i, tal que $m_i - x_i$ é o número de escores negativos desse indivíduo. Assim, temos:

m_i : número de escores do indivíduo i

x_i : número de escores positivos do sujeito i

K : número de categoria de diagnósticos (K = 2).

Com a finalidade de obter uma estatística kappa, foi explorada a identidade entre o coeficiente de correlação intraclasse e o coeficiente kappa, aplicando a técnica one-way da análise de variância, codificando os escores positivos com 1 e os escores negativos com 0.

Seja $\bar{p}$ a proporção total de escores positivos, dada por

$$\bar{p} = \frac{\sum_{i=1}^{n} x_i}{n\,\bar{m}} \qquad (3.22)$$

onde
$$\bar{m} = \frac{\sum_{i=1}^{n} m_i}{n} \qquad (3.23)$$

é o número médio de escores por indivíduo. Se o tamanho da amostra é relativamente grande ($n \geq 20$), o quadrado médio entre sujeitos (BMS) é aproximadamente igual a

$$BMS = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - m_i\,\bar{p})^2}{m_i} \qquad (3.24)$$

e o quadrado médio dentro dos sujeitos é igual a

$$WMS = \frac{1}{n\,(\bar{m} - 1)} \sum_{i=1}^{n} \frac{x_i\,(m_i - x_i)}{m_i}\,. \qquad (3.25)$$

O coeficiente de correlação intraclasse é estimado por

$$r = \frac{BMS - WMS}{BMS + (m_o - 1)\,WMS} \qquad (3.26)$$

onde $$m_o = \bar{m} - \frac{\sum_{i=1}^{n} (m_i - \bar{m})^2}{n(n-1)\bar{m}} . \qquad (3.27)$$

Contudo, se n é grande, m_o e $\bar{m}$ são muito próximos. Assim, se fizérmos $m_o = \bar{m}$ na equação (3.26), a expressão resultante é a estatística kappa, dada por:

$$\hat{\kappa} = \frac{BMS - WMS}{BMS + (\bar{m} - 1)\ WMS} = 1 - \frac{\sum_{i=1}^{n} \frac{x_i(m_i - x_i)}{m_i}}{n(\bar{m} - 1)\bar{p}\ \bar{q}} \qquad (3.28)$$

onde $\bar{q} = 1 - \bar{p}$.

Segundo Fleiss (1981), o coeficiente de concordância kappa definido acima, tem as seguintes propriedades: se não existe variação na proporção de escores positivos entre os indivíduos (isto é, se $\bar{p} = x_i/m_i$, $\forall$ i, com $\bar{p} \neq 0$, $\bar{p} \neq 1$), então existe mais discordância dentro dos sujeitos do que entre os sujeitos. Neste caso, $\hat{\kappa}$ assumirá seu valor mínimo em $-1/(\bar{m} - 1)$.

Se as diversas proporções x_i/m_i variam exatamente como proporções binomiais com parâmetros m_i e probabilidade comum $\bar{p}$, então existe similaridade tanto entre como dentro dos sujeitos e o valor será $\hat{\kappa} = 0$. Se cada proporção x_i/m_i assume o valor 0 ou o valor 1, então existe perfeita concordância dentro dos sujeitos e o coeficiente $\hat{\kappa}$ assumirá o valor $\hat{\kappa} = 1$.

Para testar a hipótese de que o valor de kappa é igual a zero (H_0: $\kappa = 0$), Fleiss (1981) apresenta uma aproximação para o erro padrão de $\hat{\kappa}$, como segue.

Seja $\bar{m}_H$ a média harmônica do número de escores por indivíduo, isto é

$$\bar{m}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{m_i}} \qquad (3.29)$$

O erro padrão de $\hat{\kappa}$ é dado por

$$s.e._o(\hat{\kappa}) = \frac{1}{(\bar{m} - 1)\sqrt{n\,\bar{m}_H}} \sqrt{2(\bar{m}_H - 1) + \frac{(\bar{m} - \bar{m}_H)(1 - 4\bar{p}\,\bar{q})}{\bar{m}\,\bar{p}\,\bar{q}}} \qquad (3.30)$$

e a hipótese (H_0: $\kappa = 0$) será testada pela estatística de teste

$$z = \frac{\hat{\kappa}}{s.e._o(\hat{\kappa})} \sim N(0,1). \qquad (3.31)$$

Vamos considerar, agora, o caso em que o número de categorias de diagnósticos é $K \geq 2$. Denotaremos por $\bar{p}_j$ a proporção de escores na categoria j e por $\hat{\kappa}_j$ o valor do coeficiente kappa para a categoria j ($j = 1, 2, \ldots, K$). Landis e Koch (1977b) propuseram a média ponderada dos coeficientes kappa individuais como uma medida da concordância total entre os juízes, como segue:

$$\hat{\bar{\kappa}} = \frac{\sum_{j=1}^{K} \bar{p}_j\,\bar{q}_j\,\hat{\kappa}_j}{\sum_{j=1}^{K} \bar{p}_j\,\bar{q}_j} \qquad (3.32)$$

onde $\bar{q}_j = 1 - \bar{p}_j$.

Quando o número de escores por sujeito é constante e igual a m, Fleiss (1981) apresenta os estimadoresde κ_j e $\bar{\kappa}$, denotados respectivamente por $\hat{\kappa}_j$ e $\hat{\bar{\kappa}}$, bem como o erro padrão de $\hat{\kappa}_j$ e $\hat{\bar{\kappa}}$, como segue. Seja x_{ij} o número de escores do sujeito i (i = 1,2,...,n) e categoria j (j = 1, 2, ..., K), tal que:

$$\sum_{j=1}^{K} x_{ij} = m \quad , \forall\, i \tag{3.33}$$

A igualdade da equação (3.33) refere-se ao número constante de juízes para todos os indivíduos. Cabe ressaltar que esses juízes não podem ser os mesmos de um indivíduo para outro, como mencionado anteriormente.

Então, segundo Fleiss (1981), o valor de $\hat{\kappa}_j$ é dado por

$$\hat{\kappa}_j = 1 - \frac{\sum_{i=1}^{n} x_{ij}(m - x_{ij})}{n\, m(m - 1)\bar{p}_j\, \bar{q}_j} \tag{3.34}$$

$$\hat{\bar{\kappa}} = 1 - \frac{nm^2 - \sum_{i=1}^{n} \sum_{j=1}^{K} x_{ij}^2}{n\, m(m - 1) \sum_{j=1}^{K} \bar{p}_j\, \bar{q}_j} \tag{3.35}$$

Para testar a hipótese de κ_j e $\bar{\kappa}$ é igual a zero, foram sugeridas os seguintes resultados assintóticos:

$$s.e._o(\hat{\bar{\kappa}}) = \frac{\sqrt{2}}{\sum_{j=1}^{K} \bar{p}_j \bar{q}_j \sqrt{nm(m-1)}} \times$$

$$\times \sqrt{\left(\sum_{j=1}^{K} \bar{p}_j \bar{q}_j\right)^2 - \sum_{j=1}^{K} \bar{p}_j \bar{q}_j (\bar{q}_j - \bar{p}_j)} \qquad (3.36)$$

e,

$$s.e._o(\hat{\kappa}_j) = \sqrt{\frac{2}{n\, m(m-1)}} \; . \qquad (3.37)$$

Fleiss (1981) salienta que $s.e._o(\hat{\kappa}_j)$ é independente de $\bar{p}_j$ e $\bar{q}_j$, sendo um caso especial de (3.30), quando o número de juízes por sujeito são iguais, donde temos $\bar{m} = \bar{m}_H = m$.

Para este último modelo, foi desenvolvido um programa em linguagem BASIC e, posteriormente, serão apresentadas algumas aplicações.

3.3. Classificação do Individuo em mais de uma Categoria de Diagnóstico, pelo mesmo Juiz.

Segundo Kraemer (1980), o coeficiente kappa tem sido desenvolvido como uma medida de concordância entre os juízes, corrigida por chance, aplicada principalmente para verificar a fidedignidade ou reprodutividade de uma categorização de diagnósticos de pacientes. É uma estatística baseada numa amostra de n indivíduos de uma grande população, com m_i observações independentes para cada indivíduo $i(i = 1, 2, \ldots, n)$. Cada observação ou escore é uma classificação do sujeito numa (ou mais) das K categorias de diagnósticos. Já foi extendido em diversas direções, dentre as quais:

i) a forma ponderada a qual quantifica a seriedade diferencial dos erros de má classificação;

ii) quando existem múltiplas categorias de respostas, com um número igual de observações (juízes) por indivíduo, maior do que dois;

iii) quando existem duas categorias de respostas e um número desigual de observações por sujeito;

iv) quando são permitidas múltiplas escolhas de categorias de resposta por indivíduo; e,

v) generalizações baseadas numa extensão do conceito dos modelos de componentes de variância para dados categorizados, junto com aproximações para análise de tabelas de contingência multimensionais (Landis e Koch, 1977b, c).

A restrição de que a classificação seja somente numa categoria de resposta, frequentemente é inadequada. Esses sistemas de categorização de diagnósticos permitem o assinalamento de categorias de respostas primárias, secundária e terciária, refletindo que os indivíduos poderiam enquadrar-se em duas ou mais categorias de diagnósticos.

De acordo com Kraemer (1980), uma forma de kappa apropriada para múltiplas escolhas está baseada numa suposição que frequentemente é inaceitável. Se A e B são duas das K categorias de respostas, essa aproximação trata a situação (A / B: igualmente A ou B) como em concordância total com cada uma das duas seguintes respostas ordenadas (AB: primeiro A, segundo B; BA: primeiro B, segundo A), tratando as respostas ordenadas como em concordância com cada outra. Considerando como respostas equivalentes estas duas últimas alternativas, pode, em algumas circunstâncias, obscurecer distinções reais e vitais.

Outro problema encontrado nesta generalização de kappa é o de termos igual número de observações por sujeito.

Kraemer (1980) sugere uma extensão da definição do coeficiente kappa, o qual trata apropriadamente com múltiplas respostas por observação e o qual permite múltipla e não necessariamente igual número de observações por sujeito.

Definição de κ_o de Kraemer.

Como foi apresentado por Kraemer (1980), uma observação sobre um sujeito é considerada como uma ordenação das K categorias de respostas. Assim, uma escolha simples, A, impõe um rank de 1,0 na categoria A e 1/2(K + 2) $\left[\text{a média dos ranks } 2, 3, \ldots, K\right]$ nas outras K - 1 categorias de respostas. Uma respostas ambígua, A/B, impõe um rank de 1,5 nas categorias A e B e de 1/2(k + 3) nas outras K - 2 categorias de respostas. Uma resposta ordenada, AB, impõe um rank 1,0 na categoria A; 2,0 na B e 1/2(k + 3) nas outras K - 2 respostas. Obviamente, as observações A, A/B, AB e BA não são consideradas equivalentes. A concordância entre duas observações pode ser medida pelo coeficiente de correlação por ranks de Spearman, cujos índices de concordância são mostrados na Tabela 3.3. O grau de concordância depende do número de categorias de respostas disponíveis, bem como da natureza de cada observação. Concordância perfeita, com valor 1,0 , é indicado somente quando duas observações são idênticas em todos os aspectos.

Tabela 3.3 - Grau de Concordância entre Observações Ordenadas (AB, BA), Ambíguas (A/B) e e única (A), como Função do Número Categorias de Respostas (K)

Resposta(s)	Rank de Spearman	K = 2	3	5	10	∞
A vs. A/B	$\left[(K-2)/2(K-1)\right]^{1/2}$	0,00	0,50	0,61	0,67	0,71
A vs. AB	$\left[K/(K-1)\right]^{1/2}$	1,00	0,87	0,79	0,74	0,71
A vs. AB	$\left[K^2(K-3)/2(K-1)^3\right]^{1/2}$	−1,00	0,00	0,40	0,58	0,71
A/B vs. AB ou BA	$\left[K(K-2)/(K-1)^2\right]^{1/2}$	0,00	0,87	0,97	0,99	1,00
AB vs. BA	$(K^2-2K-1)/(K-1)^2$		0,50	0,88	0,98	1,00

Seja $r_i (i = 1, 2, \ldots, n)$ a média dos coeficientes de correlação por rank de Spearman entre os $1/2\ m_i(m_i - 1)$ pares de observações do indivíduo i e seja r_I a média de $r_1, r_2, \ldots, r_n$, isto é, a média dos coeficientes de correlação intrasujeitos. Seja, ainda, r_T a média dos coeficientes de correlação por rank de Spearman entre todos os pares de observações, sendo que existem $m.(m. - 1)$ pares, onde $m. = \sum_{i=1}^{n} m_i$. Então, κ_o é definido por

$$\kappa_o = \frac{r_I - r_T}{1 - r_T} \tag{3.38}$$

De acordo com Kraemer (1980), sob a hipótese de nulidade o valor esperado de κ_o é igual a zero. Por outro lado, $\kappa_o = 1$ se e somente se existe absoluta concordância entre todas as observações de qualquer indivíduo sozinho (isto é, $r_I = 1$) e também alguma heterogeneidade entre indivíduos (isto é, $r_T \neq 1$).

Seja R_{ij} o rank médio atribuído para a categoria de resposta j, pelas m_i observações do sujeito i. Se ocorrerem empates, usamos a correção $\sum (t^3 - t)$, onde t é o número de categorias com um rank empatado, somando para todos os grupos de empates de cada observação. Por exemplo, se as observações (em ordem crescente dos ranks) de um indivíduo foi (1,5; 1,5;4;4;4;6), então existem dois grupos de empates, um de tamanho t = 2 e outro com t = 3 e a correção de empates para esta observação é $(2^2 - 2) + (3^3 - 3) = 30$. A correção média de empates das m_i observações do sujeito i é notada por T_i. Seja S_i^2 a variância amostral das quantidades R_{ij}, dada por

$$S_i^2 = \frac{\sum_{j=1}^{K} \left(R_{ij} - \bar{R}\right)^2}{K - 1}, \text{ onde } \bar{R} = 1/2(K + 1)$$

O coeficiente de concordância de Kendall para os m_i ranks do indivíduo i é

$$W_i = \frac{12(K - 1)S_i^2}{(K^3 - K - T)}, \text{ podendo ser verificado que}$$

$$r_i = \frac{m_i W_i - 1}{m_i - 1}$$

Similarmente, seja $\bar{R}_{.j}$ a média total dos ranks da categoria j e T. a média ponderada dos T_i, isto é,

$$\bar{R}_{.j} = \sum_{i=1}^{n} \frac{R_{ij}}{n} \quad , \quad T. = \sum_{i=1}^{n} \frac{m_i T_i}{m.}$$

Seja S_T^2 a variância amostral das quantidades $\bar{R}_{.j}$, dada por

$$S_T^2 = \frac{\sum_{j=1}^{K} \left(\bar{R}_{.j} - \bar{R}\right)^2}{K - 1} .$$

Então o coeficiente de concorddncia total de Kendall é

$$W_T = \frac{12(K - 1)S_T^2}{K^3 - K - T.} \qquad \text{e, consequentemente,}$$

$$r_T = \frac{m.W_T - 1}{m. - 1} \tag{3.39}$$

Dessa forma, o cálculo de κ_o pode ser obtido através de uma análise de variância dos ranks.

Ainda de acordo com Kraemer (1980), no caso de uma escolha de uma única categoria por observação e com um número igual de observações por sujeito, m_o, pode ser mostrado que

$$r_I = \frac{K\bar{P} - 1}{K - 1} \tag{3.40}$$

onde $\bar{P}$ é a proporção média de pares concordantes, para todos os pares possíveis de observação de cada sujeito. Ainda,

$$r_T = \frac{n\, m_o K\, P_e - n\, m_o - K + 1}{(n\, m_o - 1)(K - 1)} \tag{3.41}$$

onde $P_e = \sum_{j=1}^{K} p_j^2$ e p_1, p_2, , p_K são as proporções totais de observações nas quais as diversas categorias de respostas foram selecionadas. Assim,

$$\kappa_o = \frac{\bar{P} - P_e}{1 - P_e} + \frac{(1 - \bar{P})}{n\, m_o (1 - P_e)} \tag{3.42}$$

Como o número e observações por sujeito, m_o, ou o número de sujeitos, n, geralmente é grande, o último termo da equação (3.42) pode ser desconsiderado. Daí,

$\kappa_o \geq (\bar{P} - P_e)/(1 - P_e) = \kappa_F$ assintoticamente tem distribuição semelhante a κ_F, sugerido por Fleiss para o caso de múltiplos juízes.

No caso especial em que $m_o = 0$, temos $r_I = 2p_o - 1$, onde p_o é a proporção de concordância observada entre o Juiz 1 e o Juiz 2 e, a proporção de concordância esperada por chance é $p_e = P_e - \sum_{j=1}^{K} \Delta_j^2$, onde Δ_j é a diferença na frequência relativa da seleção da categoria de resposta j pelo Juiz 1 e Juiz 2.

Disso resulta o coeficiente proposto originalmente por Cohen, denotado aqui por

$$\kappa_c = \frac{p_o - p_e}{1 - p_e} \quad .$$

Segundo Kraemer (1980), se os dois juízes para um indivíduo forem selecionados aleatoriamente de um grande grupo de juízes, ou se os juízes forem fixados mas não for observado um viés entre eles, as frequências marginais seriam consistentemente diferentes (isto é, $p_e \cong P_e$). Assim, para grandes amostras, κ_F , κ_c e κ_o tendem a coincidir.

Distribuição Amostral de κ_o

De acordo com Kraemer (1980), existem duas aproximações para a avaliação da distribuição amostral do coeficiente kappa: uma na qual a distribuição marginal é fixada (r_T é fixado) e outra na qual a amostragem dos sujeitos e das observações são incondicionais. Para o caso em que o r_T é fixado, o erro padrão de κ_o é dado por

$$s.e.(\kappa_o) = \frac{S_r}{\sqrt{n}\,\left(1 - r_T\right)} \tag{3.43}$$

onde $S_r^2 = \dfrac{\sum_{i=1}^{n} (r_i - r_I)^2}{n - 1}$ é a variância sobre os r_i sujeitos. Para amostras com um tamanho de amostra, n, relativamente grande, podemos utilizar a distribuição t para determinar um intervalo de $(1 - \alpha)100\%$ de confiança para o valor da população, utilizando $n - 1$ graus de liberdade, como segue:

$$\kappa \geq \kappa_o - t_{\alpha(n-1)}\, s.e.(\kappa_o) \tag{3.44}$$

onde $t_{\alpha(n-1)}$ é o valor crítico unilateral ao nível de α % de significância, para a distribuição t com $n - 1$ graus de liberdade.

Kraemer (1980) apresenta, ainda, uma distribuição amostral do coeficiente kappa obtida através da técnica de estimação jackknife, a qual pode ser usada para obter intervalos de confiança robustos para um parâmetro, quando o tamanho de amostra é pequeno. Assim, calculamos os coeficientes kappa κ_1, κ_2, ..., κ_n, omitindo cada um dos n indivíduos e, para cada um desses coeficientes, calculamos um "pseudovalor"

$$J_i(\kappa) = n\,\kappa_o - (n-1)\kappa_i \quad , \qquad i = 1, 2, \ldots, n.$$

A estimativa jackknife de κ é a média dos pseudovalores, dada por

$$J(\kappa) = \sum_{i=1}^{n} J_i(\kappa)/n. \tag{3.45}$$

Em geral, tem sido encontrado que

$$\frac{\sqrt{n}\,\left[J(\kappa) - \kappa\right]}{S_J}$$ tem aproximadamente distribuição t com

n - 1 graus de liberdade, onde

$$S_J^2 = \frac{\sum_{i=1}^{n}\left[J_i(\kappa) - \kappa\right]^2}{n-1}.$$

Assim, o intervalo de confiança para κ é aproximadamente

$$\kappa \geq J(\kappa) - t_{n-1}\,\frac{S_J}{\sqrt{n}}. \tag{3.46}$$

O coeficiente de correlação médio, entre os indivíduos, omitindo o indivíduo i, é dado por

$$r_I^{(i)} = \frac{n\,r_I - r_i}{n-1}$$ e, sem o sujeito i,

a média dos ranks total é

$$\bar{R}_{.j}^{(i)} = \left[m.\,\bar{R}_{.j} - m_i R_{ij}\right] \Big/ (m. - m_i) \qquad \begin{matrix} i = 1, 2, \ldots, n \\ j = 1, 2, \ldots, K \end{matrix}$$

e, então, a variância amostral de $\bar{R}_{.j}^{(i)}$ é

$$S_T^{2(i)} = \frac{m.^2\ S_T^2 - 2\ m_i m.\ S_T S_i C_i + m_i^2 S_i^2}{(m. - m_i)^2}$$

onde C_i são os coeficientes de correlação de produto momentos sobre os pares $(R_{ij}, \bar{R}_{.j})$, $j = 1, 2, \ldots, K$.

Disso resulta que

$$W_T^{(i)} = \frac{12(K - 1)\ S_T^{2(i)}}{K^3 - K\ T.^{(i)}} ,$$

onde $T.^{(i)} = (m.T. - m_i T_i)/(m. - m_i)$; e,

$$r_T^{(i)} = \frac{(m. - m_i)\ W_T^{(i)} - 1}{m. - m_i - 1} .$$

O coeficiente kappa, omitindo o indivíduo i, é dado por

$$\kappa_i = \frac{r_I^{(i)} - r_T^{(i)}}{1 - r_T^{(i)}} , \quad i = 1, 2, \ldots, n. \qquad (3.47)$$

Assim, para utilizar o procedimento jackknife para determinar κ_o, falta determinar os coeficientes de correlação C_1, C_2, ..., C_n, entre os ranks médios para cada indivíduo e a média dos ranks para todos os indivíduos.

Um Exemplo de Aplicação para o Caso de Múltiplos Diagnósticos por Juiz.

Consideremos o exemplo apresentado por Mezzich, Kraemer e Worthington (1981). Trinta psiquiatras classificaram, independentemente, 27 casos de crianças com problemas psíquicos, sendo que cada psiquiatra julgou 3 casos e cada caso voltou a ser julgado por 3 ou 4 psiquiatras. A Tabela 3.4 mostra os resultados das 90 formulações de múltiplos diagnósticos. As categorias de diagnóticos são ordenadas como primária, secundária e terciária, quando existirem.

Na classificação realizada pelo psiquiatra 1 para o caso 1, um rank 1 foi dado para o diagnóstico 9 (Retardo Mental) e um rank 2 para o diagnóstico 11 (Desordem de Falta de Atenção) e para as 18 categorias de diagnósticos restantes foi atribuído o rank médio igual a 11,5. Para as outras 89 formulações de diagnósticos foi repetido o mesmo processo.

A seguir, devemos determinar a média dos coeficientes de correlação por ranks de Spearman entre todos pares de diagnósticos de cada caso. Para os três primeiros casos, temos: $r_1 = 0{,}58$ e $r_2 = 0{,}17$ e $r_3 = -0{,}06$. Calculamos, também, a média dos coeficientes de correlação entre as crianças (r_I) e a média dos coeficientes de correlação de Spearman entre todos os pares de psiquiatras sobre todos os casos, a qual é dada por r_T.

Assim, para este exemplo, temos $r_I = 0{,}40$, $r_T = 0{,}09$ e $S_r = 0{,}28$ é o desvio padrão dos coeficientes de correlação entre as crianças. Daí, das equações (3.38) e (3.43), resulta

$$\kappa_o = \frac{r_I - r_T}{1 - r_T} = \frac{0,40 - 0,009}{1 - 0,09} = 0,34;$$

$$s.e.(\kappa_o) = \frac{S_r}{\sqrt{n}\,(1 - r_T)} = \frac{0,28}{\sqrt{27}\,(1 - 0,09)} = 0,06 \text{ ; e,}$$

$$t = \frac{0,34}{0,06} = 5,7.$$

Portanto para um teste unilateral com (27 - 1) graus de liberdade, a significância de $\kappa_o > 0$ é $p < 0,001$.

A seguir será apresentada a Tabela 3.4, sendo que no interior da mesma constam as categorias de diagnósticos nas quais as crianças foram classificadas pelos juízes. Os nomes dessas categorias serão mantidas no original, a fim de evitar possíveis inadequações na tradução.

1 - Organic mental disorders; 2 - Substance use disorders; 3 - Schizophrenic and paranoid disorders; 4 - Schizoaffective disorders; 5 - Affective disorders; 6 - Psychoses not elsewhere classified; 7 - Anxiety factitious, somatoform and dissociative disorders; 8 - Psychosexual disorder; 9 - Mental retardation; 10 - Pervasive developmental disorder; 11 - Attention deficit disorders; 12 - Conduct disorders; 13 - Anxiety disorders of childhood or adolescence; 14 - Other disorders of childhood or adolescence, speech and stereotyped movement disorders, disorders characteristic of late adolescence; 15 - Eating disorders; 16 - Reactive disorders not elsewhere classified; 17 - Disordersof

impulse control not elsewhere classified; 18 - Sleep and other disorders; 19 - Conditions not attributable to a mental disorder; 20 - No diagnosis on Axis I.

Tabela 3.4 - Múltiplos Diagnósticos para 27 Casos de Crianças com Problemas Psiquiátricos Usando DMS - III Axis I

Casos	J U Í Z E S			
	1	2	3	4
1	9,11	11,9,14	16,9	11,9
2	16	16,14	12	14,5
3	17	12	7,8	13
4	16,13	13,16,14	16	
5	7	7,12,13	13	
6	10	10	10	
7	7,16	13	16	
8	1,14	13	16,13	
9	5	20	13,14	
10	12,13,14	12,14,13	12,11,14	
11	13	18	16	
12	5,18	1,5,18	1	
13	14,13	14,7,7	14,16	
14	11,16	14,11,16	11,13	
15	10	3,18	10,11	
16	14,5	5,16	14	
17	12	12,11	12	
18	20	16	16	
19	13	14	14	
20	9,14,10	9,11,14	10,9	
21	12,11	11,14	11	
22	17	12	12	12,17,15
23	16,13	12	14	13
24	12	12	16	12
25	13	20	13	13
26	13	13,16	13	16
27	10,9	9,10	9	9,10

4. APLICAÇÕES

4.1. Apresentação do Programa "KAPPA"

Com a finalidade de atingir um dos objetivos inicialmente propostos, relativo a consulta feita pelo Professor Paulo Belmonte de Abreu, do Departamento de Psiquiatria e Medicina Legal, foi realizado um programa em linguagem BASIC, para determinar o coeficiente kappa de cada categoria de diagnóstico, o coeficiente kappa total e seus respectivos erros-padrão [fórmulas (3.34) - (3.37)]. Este modelo refere-se à situação na qual temos um número de escores por indivíduo constante e igual a m. Deve-se salientar, novamente, que os juízes que classificam um indivíduo não são supostos os mesmos para cada um dos outros indivíduos.

O programa consiste da entrada das informações relativas à definição do modelo, da entrada dos dados (resultado da classificação dos indivíduos pelos juízes) através de uma matriz, da consistência dessa matriz no tocante ao número de juízes por sujeito, do cálculo das estimativas de κ_j e $\bar{\kappa}$ e, finalmente, do output, que pode ser via terminal de vídeo ou impressora. Estas etapas serão discutidas de forma detalhada, com a apresentação das telas de entrada e saída e dos relatórios.

Desenvolvido em linguagem BASIC, posteriormente o programa foi transformado num arquivo "executável" através da compilação, passando assim a denominar-se KAPPA.EXE. Isto visou facilitar sua utilização pelo usuário que não tem experiência prévia da linguagem, bem como evitar danos ao programa fonte por possíveis acidentes.

Assim sendo, pode ser utilizado em qualquer microcomputador compatível com IBM/PC (XT, AT, 286 e 386), bastando digitar "KAPPA" e teclar "Enter" ou "Return".

No **Anexo I** encontra-se uma listagem do programa-fonte. As pessoas interessadas em obter cópia do programa executável em disquete, deverão dirigir-se ao Núcleo de Assessoria Estatística do Departamento de Estatística da UFRGS.

A seguir serão abordadas alguns aspectos de entrada e saída do programa KAPPA.

- Definição do Modelo: identificada pela **Tela 1** do **Anexo II**, esta etapa refere-se a entrada dos valores iniciais do modelo, os quais são:

  i) Número de Categorias de Diagnósticos;
  ii) Número de Juízes; e,
  iii) Número de Indivíduos.

- Entrada de Dados: a **Tela 2** mostra as opções que o usuário possui para digitar a matriz de dados, que pode ser através dos valores individuais ou pela entrada de uma ou mais linhas da matriz.

- Demonstração dos Resultados: tanto a matriz de dados quanto os resultados da análise estão disponível via impressora ou terminal de vídeo. Na **Tela 3** apresenta os resultados da análise de um modelo e o usuário é perguntado se deseja imprimir esses resultados, os quais são mostrados no **Relat 1** do **Anexo III**. O relatório intitulado **Relat 2**, refere-se à matriz de dados, com saída pala impressora.

4.2. Aplicação do Programa KAPPA ao Exemplo Apresentado por Fleiss

Fleiss (1981 - Pág. 230) apresentou um exemplo com dados hipotéticos para o modelo de m juízes por sujeito (m constante), conforme segue:

Número de Juízes por Sujeito: m = 5

Número de Categorias de Diagnósticos: K = 3

Número de Indivíduos: n = 10

A matriz de dados e os resultados da análise encontram-se no Anexo III. Para a interpretação destes resultados obtivemos os seguintes valores de z para os 3 diagnósticos

Diagnóstico 1 : z = 2,917 ($p < 0{,}01$)
Diagnóstico 2 : z = 6,711 ($p < 0{,}0000$)
Diagnóstico 3 : z = 3,490 ($p < 0{,}001$)
e o kappa total tem um valor associado de z = 5,832 ($p < 0{,}000$)

Embora estamos rejeitando a hipótese de que $\kappa_1 = 0$, $\kappa_2 = 0$, $\kappa_3 = 0$ e $\bar{\kappa} = 0$, somente o coeficiente de concordância $\hat{\kappa}_2 = 0{,}671$ apresenta um grau de fidedignidade de diagnóstico "considerável". O coeficiente $\bar{\kappa} = 0{,}418$ indica um grau de concordância "moderado:, enquanto que $\hat{\kappa}_1 = 0{,}292$ e $\hat{\kappa}_3 = 0{,}349$ indicam apenas um grau de concordância "regular".

4.3. Aplicação a um Caso Real

O exemplo a seguir trata do caso onde temos 11 psiquiatras (juízes) diagnosticando 20 pacientes de hospitais brasileiros. As categorias de diagnósticos possíveis são 10, de acordo com o sistema de diagnóstico internacional. A matriz de dados e os resultados são mostrados a seguir. Obteve-se um coeficiente kappa total $\hat{\bar{\kappa}} = 0{,}492$, o qual é significante e é comparável em grandeza aos valores obtidos para o caso americano. O objetivo principal deste estudo era medir a fidedignidade do sistema de diagnóstico utilizado internacionalmente. Portanto, podemos concluir que o sistema de diagnóstico é preciso, visto que a fidedignidade foi alta tanto no caso brasileiro, como no caso americano.

U F R G S - INSTITUTO DE MATEMATICA
NUCLEO DE ASSESSORIA ESTATISTICA
Coeficiente de Concordancia KAPPA

MATRIZ DE DADOS DO MODELO: TABELA C

Sujeito (i)	Categoria de Diagnosticos 1	2	3	4	5	6	7	8	9	10
1	0	0	0	9	1	0	1	0	0	0
2	1	0	0	10	0	0	0	0	0	0
3	5	0	0	0	5	0	1	0	0	0
4	0	0	0	1	2	0	8	0	0	0
5	3	7	0	1	0	0	0	0	0	0
6	2	1	3	4	0	0	1	0	0	0
7	7	0	3	1	0	0	0	0	0	0
8	1	0	9	1	0	0	0	0	0	0
9	0	0	0	0	2	0	1	0	0	8
10	0	0	0	10	1	0	0	0	0	0
11	1	0	0	0	1	0	1	1	2	5
12	0	0	4	5	0	0	2	0	0	0
13	0	0	2	9	0	0	0	0	0	0
14	0	0	0	0	1	0	0	0	3	7
15	1	0	1	4	1	0	4	0	0	0
16	0	0	0	0	1	1	0	0	0	9
17	0	0	0	0	0	0	1	10	0	0
18	0	0	11	0	0	0	0	0	0	0
19	0	0	0	2	0	9	0	0	0	0
20	0	0	11	0	0	0	0	0	0	0

U F R G S - INSTITUTO DE MATEMATICA
NUCLEO DE ASSESSORIA ESTATISTICA
Coeficiente de Concordancia KAPPA

Resultados da Analise para o Modelo: TABELA C

Diagnostico (j)	p Medio	KAPPA (j)	Erro Padrao de KAPPA(j)	Valor de z
1	0.095	0.263	0.030	8.722
2	0.036	0.507	0.030	16.818
3	0.200	0.653	0.030	21.671
4	0.259	0.526	0.030	17.459
5	0.068	0.099	0.030	3.268
6	0.045	0.707	0.030	23.437
7	0.091	0.285	0.030	9.452
8	0.050	0.809	0.030	26.819
9	0.023	0.140	0.030	4.659
10	0.132	0.603	0.030	19.993
KAPPA total		0.492	0.012	40.522

Numero de Diagnosticos: 10

Numero de Juizes: 11

Numero de Individuos: 20

5. CONCLUSÃO

Na presente monografia foram abordados diversos coeficientes kappa, tanto no contexto de concordância quanto no contexto de associação. Contudo, alguns tipos de coeficientes kappa não foram explorados, como, por exemplo, o coeficiente RE, o qual é recomendado quando assumimos que a concordância por chance ocorre de uma maneira totalmente aleatória (Maxwell, 1977).

Acreditamos ter contribuído para a divulgação e utilização dos coeficientes kappa e propomos que numa futura pesquisa sejam estudados aqueles não abordados nesta monografia, bem como a realização da comparação numérica desses coeficientes.

Outros tópicos, tais como a determinação da fidedignidade de sistemas de diagnóstico através dos modelos da análise de variância, também ficam como proposta para um futuro estudo, estruturação e apresentação.

6. ANEXOS

6.1. Anexo I - Listagem do Programa KAPPA.BAS

```
' NOME DO PROGRAMA: KAPPA.BAS              30/06/89
' AUTOR: VIGO
'
'
' Calcula os Coeficientes de Concordancia Kappa para
' o caso de Multiplos Juizes por Sujeito
'
COLOR 0,2
CLS
DO WHILE CONT3 <> 13
   LOCATE 5,8 : PRINT "KK    KK       AAAAAAAAA     PPPPPPPPP     PPPPPPPPP     AAAAAAAAA "
   LOCATE 6,8 : PRINT "KK   KK        AA      AA    PP      PP    PP      PP    AA      AA "
   LOCATE 7,8 : PRINT "KK KK          AA      AA    PP      PP    PP      PP    AA      AA"
   LOCATE 8,8 : PRINT "KKKK           AAAAAAAAA     PPPPPPPPP     PPPPPPPPP     AAAAAAAAA"
   LOCATE 9,8 : PRINT "KK KK          AA      AA    PP            PP            AA      AA"
   LOCATE 10,8 : PRINT "KK  KK        AA      AA    PP            PP            AA      AA"
   LOCATE 11,8 : PRINT "KK   KK       AA      AA    PP            PP            AA      AA"
   LOCATE 16,20 : PRINT "UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL"
   LOCATE 17,20 : PRINT "Nucleo de Assessoria Estatistica"
   LOCATE 18,20 : PRINT "Autor:  VIGO"
   COLOR 16,2
   LOCATE 23,50 : PRINT "Tecle  ENTER  para Continuar"
   CONT3=0 : CONT3$=""
   DO UNTIL CONT3=13
      CONT3$=INKEY$
      IF CONT3$ <> "" THEN CONT3=ASC(CONT3$)
   LOOP
LOOP
COLOR 15,0
CONT$=""
DO WHILE CONT$ <> "N"
   CALL TELA
   LOCATE 6,30 : PRINT "DEFINICAO  DO  MODELO"
   LOCATE 7,30 : PRINT STRING$(21,"_")
   OP$="" : ALT$=""
   DO WHILE ALT$ <> "N"
      LOCATE 10,46 : PRINT "     "
      LOCATE 10,20 : INPUT "Numero  de  Diagnosticos: ";K
      LOCATE 10,46 : PRINT " "
      LOCATE 12,40 : PRINT "     "
      LOCATE 12,20 : INPUT "Numero  de  Juizes: ";M
      LOCATE 12,40 : PRINT " "
      LOCATE 14,44 : PRINT "     "
      LOCATE 14,20 : INPUT "Numero  de  Individuos: ";N
      LOCATE 14,44 : PRINT " "
      LOCATE 23,35 : PRINT "           No  maximo  15  caracteres"
      LOCATE 17,40 : PRINT  STRING$(20," ")
      LOCATE 17,15 : LINE INPUT "Identificacao do Modelo: ";NOME$
      NOME1$=UCASE$(NOME$)
      LOCATE 23,35: PRINT "Deseja fazer alguma alteracao (S/N):   ";
      OP$=""
      DO UNTIL OP$ = "S" OR OP$ = "N"
         OP$ = UCASE$(INKEY$)
```

```
   LOOP
   ALT$=OP$
LOOP
OPTION BASE 1
DIM MAT(N+1,K+1) , PB(K) , KAPA(K) , ZKAPA(K)
'
' Entrada de Dados na Matriz
'
CALL TELA
LOCATE 6,29: PRINT "ENTRADA  DE  DADOS"
LOCATE 7,29 : PRINT STRING$(18,"_")
OP$=""
LOCATE 12,15 : PRINT "1.  Entrada com Valores Individuais"
LOCATE 14,15 : PRINT "2.  Entrada por Linha"
LOCATE 18,30 : PRINT "Opcao:  "
DO UNTIL OP$="1" OR OP$="2"
   OP$=UCASE$(INKEY$)
LOOP
IF OP$="2" THEN
   CALL TELA
   DIM VET(K)
   CONF$=""
   DO WHILE CONF$ <> "S"
      NSUJ=0 : TOTSUJ=0
      DO WHILE TOTSUJ < N
         CALL TELA
         LOCATE 4,21 : PRINT "Numero de Observacoes de cada Categoria"
         LOCATE 21,17 : PRINT "Numero de Linhas Ja Entradas: ";STR$(TOTSUJ)
         LOCATE 21,57 : PRINT "Restantes: ";STR$(N - TOTSUJ)
         IF K > 4 THEN
            RESTO1=(K MOD 4) : DIVID1=(K - RESTO1)/4
            L2=0 : REPET1=1 : CATG=0
            DO WHILE REPET1 <= DIVID1
               COL=0
               FOR J=1 TO 4
                   LOCATE 7+L2,J+COL:PRINT "Categoria";STR$(J+CATG);":";
                   INPUT VET(J+CATG)
                   COL=COL + 20
               NEXT J
               L2=L2 + 2
               CATG=4*REPET1
               REPET1=REPET1 + 1
            LOOP
            COL=0
            FOR J=(4*DIVID1+1) TO 4*DIVID1 + RESTO1
                J1=1
                LOCATE 7+L2,J1+COL:PRINT "Categoria";STR$(J);":";
                INPUT VET(J)
                COL=COL + 20
                J1=J1+1
            NEXT J
         ELSE
            COL=0
            FOR J=1 TO K
                LOCATE 10,J+COL : PRINT "Categoria";STR$(J);":";
```

```
                INPUT VET(J)
                COL=COL + 20
            NEXT J
         END IF
         LOCATE 18,8 : INPUT "====>>>> Numero de Sujeitos: ";NSUJ
         IF (NSUJ + TOTSUJ) <= N THEN
            FOR I=TOTSUJ+1 TO TOTSUJ + NSUJ
                FOR J=1 TO K
                    MAT(I,J) = VET(J)
                NEXT J
            NEXT I
            TOTSUJ=TOTSUJ + NSUJ
         ELSE
            LOCATE 21,1 : PRINT STRING$(80," ")
            LOCATE 21,17 : PRINT "Numero de Linhas Maior do que o Especificado"
            LOCATE 23,50 : PRINT "Tecle  ENTER  para Continuar"
            CONT1=0 : CONT1$=""
            DO UNTIL CONT1=13
               CONT1$=INKEY$
               IF CONT1$ <> "" THEN CONT1=ASC(CONT1$)
            LOOP
         END IF
      LOOP
      CALL TELA
      LOCATE 12,28 : PRINT "Entrada de Dados Concluida"
      LOCATE 23,57 : PRINT "Confirma: (S/N)"
      CONF$=""
      DO UNTIL CONF$ = "S" OR CONF$="N"
         CONF$=UCASE$(INKEY$)
      LOOP
   LOOP
ELSE
   CALL TELA
   FOR I=1 TO N
       FOR J=1 TO K
           LOCATE 12,20 : PRINT "Elemento (";STR$(I);",";STR$(J);") da Matriz ";
           INPUT MAT(I,J)
           LOCATE 12,46 : PRINT "       "
       NEXT J
   NEXT I
END IF
CALL TELA
'
' Conferindo a soma das linhas da matriz
'
SOMAL$=""
FOR I=1 TO N
    SOMAL=0
    FOR J=1 TO K
        SOMAL=SOMAL + MAT(I,J)
    NEXT J
    IF SOMAL <> M THEN
       SOMAL$="DIF"
    END IF
NEXT I
```

```
IF SOMAL$ = "DIF" THEN
   LOCATE 23,30 : PRINT "====>>>>  Voce Precisa Conferir a Matriz de Dados"
END IF
'
' Imprimindo a Matriz de Dados
'
LOCATE 12,27 : PRINT "Deseja Imprimir a Matriz: (S/N)"
OP$=""
DO UNTIL OP$="S" OR OP$="N"
   OP$=UCASE$(INKEY$)
LOOP
IF OP$ = "S" THEN
   LOCATE 23,30 : PRINT "                    Impressora  OK ? (S)    "
   OP$=""
   DO UNTIL OP$="S"
      OP$=UCASE$(INKEY$)
   LOOP
   LPRINT TAB(3);"U F R G S - INSTITUTO DE MATEMATICA"
   LPRINT TAB(3);"NUCLEO DE ASSESSORIA ESTATISTICA"
   LPRINT TAB(3);"Coeficiente de Concordancia  KAPPA"
   LPRINT:LPRINT:LPRINT:LPRINT
   LPRINT TAB(20);"MATRIZ  DE  DADOS  DO  MODELO:  ";NOME1$
   LPRINT TAB(3);STRING$(74,"-")
   LPRINT:LPRINT
   LPRINT TAB(10);"Sujeito        Categoria de Diagnosticos"
   LPRINT TAB(12);"(i)";
   FOR J=1 TO K
       COL=3*J + 25
       LPRINT TAB(COL);STR$(J);
   NEXT J
   LPRINT:LPRINT
   FOR I=1 TO N
       LPRINT TAB(12);STR$(I);
       FOR J=1 TO K
           COL=3*J + 25
           LPRINT TAB(COL);STR$(MAT(I,J));
       NEXT J
   NEXT I
   LPRINT : LPRINT TAB(3);STRING$(74,"-")
   LPRINT CHR$(12)
END IF
IF SOMAL$ = "DIF" THEN
   LOCATE 23,30 : PRINT "====>>>>  Voce Precisa Conferir a Matriz de Dados"
END IF
LOCATE 12,15 : PRINT "Deseja Conferir a Matriz de Dados: (S/N)        "
OP$=""
DO UNTIL OP$="S" OR OP$="N"
   OP$=UCASE$(INKEY$)
LOOP
IF SOMAL$="DIF" THEN
   OP$="S"
END IF
'
' Imprimindo a matriz na tela
'
```

```
IF OP$="S" THEN
   IF N <= 12 THEN
      LIN=N-12 : NLIN=0 : DIVID=1
   ELSE
        RESTO=(N MOD 12) : DIVID=(N - RESTO)/12
        IF RESTO=0 THEN
           NLIN=0 : LIN=0
        ELSE
            NLIN=0 : DIVID=DIVID+1 : LIN=0
        END IF
   END IF
   REPET=1
   DO WHILE REPET <= DIVID
      CALL CONFMAT
      L1=1
      FOR I=1+NLIN TO 12 +LIN
          LOCATE 8+L1,10 : PRINT STR$(I)
          FOR J=1 TO K
              LOCATE 8+L1,3*J+15 : PRINT MAT(I,J)
          NEXT J
          L1=L1+1
      NEXT I
      NLIN=NLIN + 12 : LIN=LIN + 12
      IF N > 12 THEN
         IF RESTO <> 0 THEN
            IF (12 + LIN - N) > 0 THEN
               IF N > 24 THEN
                  LIN=REPET*12 + RESTO - 12
               ELSE
                  LIN=RESTO
               END IF
            END IF
         END IF
      END IF
      AUX$="" : OP$=""
      DO WHILE AUX$ <> "N"
         LOCATE 22,15 : PRINT "               Deseja Fazer Alguma Alteracao: (S/N)"
         DO UNTIL OP$ = "S" OR OP$ = "N"
            OP$=UCASE$(INKEY$)
         LOOP
         AUX$=OP$
         LOCATE 23,1 : PRINT STRING$(78," ")
         IF AUX$="S" THEN
            L=0 : C=0
            DO UNTIL (L >= 1 AND L <= N)
               LOCATE 22,1 : PRINT STRING$(70," ")
               LOCATE 22,15 : INPUT "Linha: ";L
            LOOP
            DO UNTIL (C >=1 AND C <= K)
               LOCATE 22,43 : INPUT "Coluna: ";C
            LOOP
            LOCATE 23,5 : INPUT "====>>>>  Novo Valor da Matriz: ";MAT
            LOCATE 23,60 : PRINT "Confirma: (S/N)"
            CONF$=""
            DO UNTIL CONF$ = "S" OR CONF$ = "N"
```

```
            CONF$ = UCASE$(INKEY$)
         LOOP
         IF CONF$ = "S" THEN
            MAT(L,C)=MAT
         END IF
         LOCATE 23,1 : PRINT STRING$(78," ")
         OP$=""
      END IF
   LOOP
   REPET=REPET+1
  LOOP
LOCATE 22,1 : PRINT STRING$(79," ")
END IF
'
' Imprimindo a Matriz de Dados
'
LOCATE 23,30 : PRINT "    Deseja Imprimir a Matriz: (S/N)             "
OP$=""
DO UNTIL OP$="S" OR OP$="N"
   OP$=UCASE$(INKEY$)
LOOP
IF OP$ = "S" THEN
   LOCATE 23,1 : PRINT STRING$(79,"  ")
   LOCATE 23,55 : PRINT "Impressora  OK ? (S) "
   OP$=""
   DO UNTIL OP$="S"
      OP$=UCASE$(INKEY$)
   LOOP
   LPRINT TAB(3);"U F R G S - INSTITUTO DE MATEMATICA"
   LPRINT TAB(3);"NUCLEO DE ASSESSORIA ESTATISTICA"
   LPRINT TAB(3);"Coeficiente de Concordancia  KAPPA"
   LPRINT:LPRINT:LPRINT:LPRINT
   LPRINT TAB(20);"MATRIZ  DE  DADOS  DO  MODELO:  ";NOME1$
   LPRINT TAB(3);STRING$(74,"-")
   LPRINT:LPRINT
   LPRINT TAB(10);"Sujeito          Categoria de Diagnosticos"
   LPRINT TAB(12);"(i)";
   FOR J=1 TO K
       COL=3*J + 25
       LPRINT TAB(COL);STR$(J);
   NEXT J
   LPRINT:LPRINT
   FOR I=1 TO N
       LPRINT TAB(12);STR$(I);
       FOR J=1 TO K
           COL=3*J + 25
           LPRINT TAB(COL);STR$(MAT(I,J));
       NEXT J
   NEXT I
   LPRINT : LPRINT TAB(3);STRING$(74,"-")
   LPRINT CHR$(12)
END IF
FOR I=1 TO N
    FOR J=1 TO K
        SQX=SQX + MAT(I,J)^2
```

```
        NEXT J
        MAT(I,J)=SQX : SSQX=SSQX + MAT(I,J)
        SQX=0
      NEXT I
   FOR J=1 TO K
        FOR I=1 TO N
            SOMAC=SOMAC + MAT(I,J)
        NEXT I
        MAT(I,J)=SOMAC
        SOMAC=0
   NEXT J
   MAT(N+1,K+1)=SSQX
   FOR I=1 TO K
        PB(I)=MAT(N+1,I)/(N*M)
        SPQ=SPQ + PB(I)*(1 - PB(I))
   NEXT I
'
'  Calculo do Coeficiente Kappa Total
'
   KAPAB=1 - (N*M^2 - SSQX)/(N*M*(M-1)*SPQ)
'
'  Calculo dos Coeficientes Kappa de cada categoria
'
   FOR J=1 TO K
        SXMX=0
        FOR I=1 TO N
            SXMX=SXMX + MAT(I,J)*(M - MAT(I,J))
        NEXT I
        KAPA(J)=1 - SXMX/(N*M*(M-1)*PB(J)*(1-PB(J)))
   NEXT J
'
'  Calculo do Erro Padrao de Kappa e Kappa Total
'
   SEK=SQR(2/(N*M*(M-1)))
   FOR I=1 TO K
        SPQSE=SPQSE + PB(I)*(1-PB(I))*((1-PB(I)) -PB(I))
   NEXT I
   SEKB=SQR(2*(SPQ^2 - SPQSE))/(SPQ*SQR(N*M*(M-1)))
'
'  Calculo dos valores de Zkapa(i) e Zkapab
'
   ZKAPAB=KAPAB/SEKB
   FOR J=1 TO K
        ZKAPA(J)=KAPA(J)/SEK
   NEXT J
'
'  Mostrando os resultados na tela
'
   IF K <= 12 THEN
      LINK=K - 12 : NLINK=0 : DIVIDK=1
   ELSE
      RESTOK=(K MOD 12) : DIVIDK=(K - RESTOK)/12
      IF RESTOK=0 THEN
         NLINK=0 : LINK=0
      ELSE
```

```
      NLINK=0 : DIVIDK=DIVIDK + 1 : LINK=0
   END IF
END IF
REPETK=1
DO WHILE REPETK <= DIVIDK
   CALL TELA
   LOCATE 5,6 : PRINT "Diagnostico     p       KAPPA    Erro Padrao     Valor de"
   LOCATE 6,11 : PRINT "(j)       Medio      (j)      de KAPPA(j)        z"
   LK=1
   FOR J=1+NLINK TO LINK+12
        LOCATE 7+LK,11 : PRINT STR$(J)
        LOCATE 7+LK,19 : PRINT USING "##.###";PB(J)
        LOCATE 7+LK,28 : PRINT USING "##.###";KAPA(J)
        LOCATE 7+LK,40 : PRINT USING "####.###";SEK
        LOCATE 7+LK,54 : PRINT USING "###.###";ZKAPA(J)
        LK=LK + 1
   NEXT J
   NLINK=NLINK + 12 : LINK=LINK + 12
   IF K > 12 THEN
      IF RESTOK <> 0 THEN
         IF (12 + LINK - K) > 0 THEN
            IF K > 24 THEN
               LINK=REPETK*12 + RESTOK - 12
            ELSE
               LINK=RESTOK
            END IF
         END IF
      END IF
   END IF
   REPETK=REPETK + 1
   LOCATE LK+8,6 : PRINT "KAPPA total"
   LOCATE LK+8,28 : PRINT USING "##.###";KAPAB
   LOCATE LK+8,40 : PRINT USING "####.###";SEKB
   LOCATE LK+8,54 : PRINT USING "###.###";ZKAPAB
   IF K > 12 THEN
      IF REPETK-1 < DIVIDK THEN
         LOCATE 23,50 : PRINT "Tecle  ENTER  para Continuar"
         CONT2=0 : CONT2$=""
         DO UNTIL CONT2=13
            CONT2$=INKEY$
            IF CONT2$ <> "" THEN CONT2=ASC(CONT2$)
         LOOP
      END IF
   END IF
LOOP
LOCATE 23,45 : PRINT "Deseja  Imprimir (S/N):          "
OP$="" : ALT$=""
DO UNTIL OP$="S" OR OP$="N"
   OP$=UCASE$(INKEY$)
LOOP
ALT$=OP$
IF ALT$="S" THEN
   ALT$="" : OP$=""
   DO WHILE ALT$ <> "S"
      LOCATE 23,45 : PRINT "          Impressora  OK ? ( S ) "
```

```
            DO UNTIL OP$ = "S"
               OP$=UCASE$(INKEY$)
            LOOP
            ALT$=OP$
         LOOP
         LPRINT TAB(3);"U F R G S - INSTITUTO DE MATEMATICA"
         LPRINT TAB(3);"NUCLEO DE ASSESSORIA ESTATISTICA"
         LPRINT TAB(3);"Coeficiente de Concordancia  KAPPA"
         LPRINT:LPRINT:LPRINT:LPRINT
         LPRINT TAB(10);"Resultados da Analise para o Modelo: ";NOME1$
         LPRINT TAB(3);STRING$(75,"-")
         LPRINT TAB(8);"Diagnostico      p      KAPPA    Erro Padrao    Valor de"
         LPRINT TAB(11);"(j)          Medio     (j)     de KAPPA(j)       z"
         LPRINT TAB(3);STRING$(75,"-")
         LPRINT
         FOR J=1 TO K
             LPRINT TAB(11);STR$(J);
             LPRINT STRING$(9," ");
             LPRINT USING "##.###";PB(J);
             LPRINT STRING$(3," ");
             LPRINT USING "##.###";KAPA(J);
             LPRINT STRING$(5," ");
             LPRINT USING "####.###";SEK;
             LPRINT STRING$(6," ");
             LPRINT USING "###.###";ZKAPA(J)
         NEXT J
         LPRINT
         LPRINT TAB(8);"KAPPA total";
         LPRINT STRING$(12," ");
         LPRINT USING "##.###";KAPAB;
         LPRINT STRING$(6," ");
         LPRINT USING "###.###";SEKB;
         LPRINT STRING$(6," ");
         LPRINT USING "###.###";ZKAPAB
         LPRINT : LPRINT TAB(3);STRING$(75,"-")
         LPRINT:LPRINT:LPRINT
         LPRINT TAB(10);"Numero de Diagnosticos:  ";K :LPRINT
         LPRINT TAB(10);"Numero de Juizes:  ";M : LPRINT
         LPRINT TAB(10);"Numero de Individuos:  ";N
         LPRINT CHR$(12)
      END IF
      CALL TELA
      CONT$=""
      LOCATE 11,28 : PRINT "Deseja  Continuar (S/N) :"
      DO UNTIL CONT$ = "S" OR CONT$ = "N"
         CONT$=UCASE$(INKEY$)
      LOOP
      IF CONT$ = "S" THEN
         CLEAR
      END IF
   LOOP
   SUB TELA
       CLS
       LOCATE 1,23 : PRINT "COEFICIENTE  DE  CONCORDANCIA  KAPPA"
       LOCATE 2,25 : PRINT "(Para Numero de Juizes Constante)"
```

```
    LOCATE 3,1: PRINT STRING$(80,"_")
    LOCATE 22,1: PRINT STRING$(80,"_")
END SUB
SUB CONFMAT
    CLS
    LOCATE 1,23 : PRINT "COEFICIENTE  DE  CONCORDANCIA  KAPPA"
    LOCATE 2,25 : PRINT "(Para Numero de Juizes Constante)
    LOCATE 3,1: PRINT STRING$(80,"_")
    LOCATE 4,30 : PRINT "MATRIZ  DE  DADOS"
    LOCATE 6,8 : PRINT "Sujeito      Categoria de Diagnosticos"
    LOCATE 7,10 : PRINT "(i)"
    FOR J=1 TO K
        LOCATE 7,3*J+15 : PRINT J
    NEXT J
    LOCATE 21,1: PRINT STRING$(80,"_")
END SUB
SYSTEM
```

6.2. Anexo II - Telas do Programa KAPPA

Tela 1

COEFICIENTE DE CONCORDANCIA KAPPA
(Para Numero de Juizes Constante)

DEFINICAO DO MODELO

Numero de Diagnosticos: 3

Numero de Juizes: 5

Numero de Individuos: 10

Identificacao do Modelo: Fleiss (1981)

No maximo 15 caracteres

Tela 2

COEFICIENTE DE CONCORDANCIA KAPPA
(Para Numero de Juizes Constante)

ENTRADA DE DADOS

---

1. Entrada com Valores Individuais

2. Entrada por Linha

Opcao:

Tela 3

COEFICIENTE DE CONCORDANCIA KAPPA
(Para Numero de Juizes Constante)

Diagnostico (j)	p Medio	KAPPA (j)	Erro Padrao de KAPPA(j)	Valor de z
1	0.400	0.292	0.100	2.917
2	0.240	0.671	0.100	6.711
3	0.360	0.349	0.100	3.490
KAPPA total		0.418	0.072	5.832

Deseja Imprimir (S/N):

5.3. Anexo III - Relatórios do Exemplo Apresentado por Fleiss

Relat 1

U F R G S - INSTITUTO DE MATEMATICA
NUCLEO DE ASSESSORIA ESTATISTICA
Coeficiente de Concordancia KAPPA

Resultados da Analise para o Modelo: FLEISS (1981)

Diagnostico (j)	p Medio	KAPPA (j)	Erro Padrao de KAPPA(j)	Valor de z
1	0.400	0.292	0.100	2.917
2	0.240	0.671	0.100	6.711
3	0.360	0.349	0.100	3.490
KAPPA total		0.418	0.072	5.832

Numero de Diagnosticos: 3

Numero de Juizes: 5

Numero de Individuos: 10

U F R G S - INSTITUTO DE MATEMATICA
NUCLEO DE ASSESSORIA ESTATISTICA
Coeficiente de Concordancia KAPPA

MATRIZ DE DADOS DO MODELO: FLEISS (1981)

Sujeito (i)	Categoria de Diagnosticos 1	2	3
1	1	4	0
2	2	0	3
3	0	0	5
4	4	0	1
5	3	0	2
6	1	4	0
7	5	0	0
8	0	4	1
9	1	0	4
10	3	0	2

7. REFERÊNCIAS BIBLIOGRÁFICAS

Bloch, D. A., e Kraemer, H. C. (1989). 2 x 2 Kappa coefficients: Measures of agreement or association. Biometrics **45**, 269-287.

Carey., G. e Gottesman, I. I. (1978). Reliability and validity in binary ratings. *Arch. Gen. Psychiatry* 35:1454-1459.

Fleiss, J. L. (1981). The measurement of interrater agreement *Statistical methods for rates and proportions*, (2a ed, pp 212-236). New York: Wesley.

Kraemer, H. C. (1980). Extension of the kappa coefficient. *Biometrics* **36**, 207-216.

Landis, J. R. e Koch, G. G. (1977a). The measurement of observer agreement for categorical data. *Biometrics* **33**, 159-174.

Landis, J. R. e Koch, G. G. (1977b). A one-way components of variance model for categorial data. *Biometrics* **33**, 671-679.

Landis, J. R. e Koch, G. G. (1977c). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* **33**, 363-374.

Lord, F. M. e Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley Publishing Company.

Maxwell, A. E. (1977). Coefficients of agreement between observers and their interpretation. *Brit. J. Psychiat.*, **130**, 79-83.

Mezzich, J. E. et al (1981). Assessment of agreement among several raters formulating multiple diagnoses. *Journal of Psychiatric Research*, Vol. **16**, 29-39.

Spitzer, R. L. e Fleiss, J. L. (1974). A re-analysis of reliability of psychiatric diagnosis. *Br. J. Psychiatry* **125**, 341-347.

Romanoski, A. J. et al (1988). Interobserver reliability of a "Standardized Examination" (SPE) for case ascertainment (DMS-III). *The Journal of Nervous and Mental Disease*, Vol. **176**, No. 2, 63-71.