

INVESTIGAÇÕES SOBRE O EFEITO DE DIVERSOS DELINEAMENTOS AMOSTRAIS
SOBRE A DISTRIBUIÇÃO ASSINTÓTICA DA ESTATÍSTICA DE PEARSON
PARA INDEPENDÊNCIA EM TABELAS DE CONTINGENCIA

Este exemplar corresponde a redação
final da tese devidamente corrigida
e defendida pelo Sr. JOSE LUIS
LLANOS CARRILLO e aprovada pela
Comissão Julgadora.

Campinas, 29 de Setembro de 1987

Prof. Dr.



Sebastião de Amorim

Dissertação apresentada ao Institu-
to de Matemática, Estatística e
Ciência da Computação, UNICAMP, co-
mo requisito parcial para obtenção
do Título de Mestre em ESTATÍSTICA.

A mi padre (in memoriam),
madre y hermanos.

AGRADECIMENTOS

Meus sinceros agradecimentos ao Prof. Dr. Sebastião de Amorim, pela orientação.

Ao Cnpq e Capes pelo apoio financeiro.

A minha esposa Maura, pela paciência, compreensão e carinho, sem o qual não seria possível a realização deste trabalho.

A minha mãe Felícia e aos meus irmãos Laura, Guillermo, María e Silvia pelo apoio emocional.

A todos os amigos do IMECC pela colaboração direta ou indireta.

Ao pessoal do futebol pelos momentos de lazer e descontração.

ÍNDICE

INTRODUÇÃO	1
CAPÍTULO 1 - APRESENTAÇÃO DA TEORIA BÁSICA	8
1.1 - Tabelas de classificação cruzada	8
1.1.1 - Estatística Qui-Quadrado	11
1.1.2 - Distribuição assintótica	11
1.1.3 - Simulação Via Monte-Carlo	14
1.2 - Distorções devidas ao efeito do delineamento	19
1.2.1 - Efeito de delineamento	19
CAPÍTULO 2 - SIMULAÇÕES DE DIVERSOS DELINEAMENTOS	21
2.1 - Amostragem aleatória simples sem reposição	21
2.2 - Amostragem aleatória estratificada com alocação pro- porcional	26
2.3 - Amostragem aleatória por conglomerado	29
2.3.1 - Configuração com conglomerados homogêneos	30
2.3.2 - Configuração com conglomerados semelhantes aos da po- pulação	41
2.4 - Comentários e conclusões	54
CAPÍTULO 3 - CORREÇÕES PARA ALGUNS CASOS	55
3.1 - O teste de Wald	55

3.2 - Correção da Estatística de Pearson X^2	59
3.2.1 - Para o caso de independência na tabela rxc	61
3.2.2 - Distribuição assintótica de X^2	63
3.3 - Correções aproximadas para alguns casos	64
3.3.1 - Amostragem aleatória simples sem reposição	65
3.3.2 - Amostragem aleatória estratificada com alocação proporcional	70
3.3.3 - Amostragem aleatória por conglomerado	76
3.4 - Comentários e conclusões	87
CAPÍTULO 4 - CORREÇÃO PARA AMOSTRAGEM ALEATÓRIA POR CONGLOMERADOS	88
4.1 - Uma correção teórica	88
4.1.1 - Estimador da variância	94
4.2 - Matriz de efeito de delineamento	96
4.3 - Estimador de V_h pelo método de reamostragem "Bootstrap"	98
4.3.1 - Estimando V_h por reamostragem "Bootstrap"	99
4.4 - Resultados via simulação para X^2	99
4.4.1 - Resultados para conglomerados homogêneos	101
4.4.2 - Resultados para conglomerados semelhantes aos da população	120
4.5 - Comentários e conclusões	139

APÊNDICE	140
A.1 - Composição dos dados	140
A.2 - Prova da matriz G	142
A.3 - O programa	147
REFERÊNCIAS BIBLIOGRÁFICAS	161

INTRODUÇÃO

Na análise estatística de dados categóricos é comumente de interesse testar a hipótese de independência entre os fatores, em tabelas de contingência de duas entradas. Nestes casos, a estatística Qui-Quadrado de Pearson é frequentemente utilizada, caracterizando-se como um dos procedimentos estatísticos de maior popularidade. Sob suposições de amostragem multinomial, a Estatística de Pearson tem distribuição assintótica conhecida, a distribuição Qui-Quadrado.

Em particular, aplicando-se Amostragem Aleatória Simples com reposição (AAS-CR) em populações finitas, a suposição de multinomialidade é satisfeita, justificando a aplicação da Estatística de Pearson, desde que o tamanho da amostra não seja demasiado pequeno. Ao Delineamento de Amostragem Aleatória Simples com reposição (AAS-CR) denominaremos Delineamento Neutro.

Contudo em diversos contextos de pesquisa, seja por conveniência, flexibilidade ou economia, são frequentemente empregados delineamentos amostrais de maior ou menor complexidade, diferentes do delineamento neutro. Nestes casos, a suposição de distribuição multinomial para os dados é violada, podendo provocar profundas alterações na distribuição amostral da Estatística de Pearson.

Alguns delineamentos amostrais mais frequentes são os de Amostragem Aleatória Simples sem reposição (AAS-SR), Amostragem Aleatória Estratificada (AAE) e Amostragem Aleatória por Conglomerado (AAC). Todos estes delineamentos são não-neutros, no sentido de apresentarem um efeito do delineamento (deff) representado por uma matriz diferente da identidade, invalidando assim a suposição de distribuição multinomial para os dados.

Num primeiro passo, nosso objetivo neste trabalho tem um caráter pedagógico. Procuramos fundamentar parcialmente e ilustrar as distorções, ora suaves e perfeitamente aceitáveis, ora profundas, na distribuição assintótica da Estatística de Pearson. Investigamos também o comportamento da Estatística de Pearson para pequenas amostras, num contexto de delineamento neutro. Aqui chegamos a evidências de que, talvez, as restrições geralmente levantadas na literatura, (Battacharyya e Johnson 1977, Raktos e Hubert (1977), Yule e Kendall 1968) contra a aplicabilidade do teste Qui-Quadrado para pequenas amostras sejam um tanto exageradas.

Em seguida avaliamos o desempenho de algumas abordagens alternativas à Estatística de Pearson, como a estatística de Wald e os procedimentos corretivos para aquelas distorções, sugeridos em Rao e Scott (1981). Vemos que as correções propostas em Rao e Scott são efetivas para AAC, desde que a estrutura de variância entre e dentro de conglomerado satisfaça a certas condições mais ou menos rigorosas. Em situações opostas, também de ocorrência frequente, aquela correção na verdade aprofunda as distorções.

Propomos assim uma alternativa própria que consiste na aplicação de um esquema baseado no "bootstrap" para correção da Estatística de Pearson em AAC.

Para todos nossos estudos ilustrativos e de comparação, utilizamos uma população hipotética, de 600 indivíduos, classificados em termos de duas variáveis, que denominamos, por conveniência, sexo e idade, com 2 e 3 categorias respectivamente.

		I D A D E			TOTAL
		1	2	3	
S E X	1	40	100	60	200
	2	80	200	120	400
TOTAL		120	300	180	600

Os indivíduos foram distribuídos nas celas no sentido de se ter perfeita independência entre os dois fatores, caracterizando assim uma situação de hipótese nula verdadeira.

As ilustrações da distribuição da Estatística de Pearson, bem como os estudos comparativos entre alternativas diversas àquela abordagem, são feitos via Monte-Carlo, retirando-se um número adequado de amostras, de tamanhos diversos, daquela população sob vários delineamentos amostrais.

O primeiro delineamento não-neutro considerado foi o de Amostragem Aleatória Simples sem reposição (AAS-SR). As distorções na distribuição da Estatística de Pearson podem aqui ser associadas ao fator de correção por população finita (fcpf), podendo ser severas ou não, de acordo com aquele fator. A correção proposta em Rao e Scott usa o próprio fcpf, e corrige adequadamente as distorções.

Em seguida consideramos o delineamento de Amostragem Aleatória Simples Estratificada (AAE) com alocação proporcional, e AAS-CR dentro de cada estrato. Para isto consideramos a população como dividida em dois estratos fortemente diferenciados,

Estrato 1 :

		I D A D E			TOTAL
		1	2	3	
S E X O	1	30	20	30	80
	2	10	40	30	80
	TOTAL	40	60	60	160

Estrato 2 :

		I D A D E			TOTAL
		1	2	3	
S E X O	1	10	80	30	120
	2	70	160	90	320
	TOTAL	80	240	120	440

Conforme já sugerido em Nathan (1975), as distorções associadas àquele delineamento amostral foram extremamente suaves, não justificando maiores preocupações. Desenvolvemos, neste caso, uma explicação teórica parcial para estes resultados.

Trabalhando com AAC, criamos diversas distribuições espaciais da população, ordenando adequadamente os 600 indivíduos, no sentido de forçarmos estruturas pré-estabelecidas de variabilidade entre e intra-conglomerados. Assim, distribuindo os indivíduos sequencialmente de forma que indivíduos da mesma cela venham juntos, teremos que todos os conglomerados de tamanho submúltiplo de 20 terão perfeita homogeneidade interna, desde que a pré-divisão da população em conglomerados seja feita, sequencialmente, pela ordem espacial. Neste caso de perfeita homogeneidade, a transformação de Rao e Scott fornece resultados excelentes. Em verdade, a correção é exata se os conglomerados forem escolhidos por AAS-CR, conforme veremos. Em todos nossos trabalhos consideramos a população dividida em conglomerados de mesmo tamanho. Os planos amostrais serão sempre em 1 estágio, implicando que 1 conglomerado amostrado entre integralmente para a amostra.

No outro extremo, criamos distribuições espaciais no sentido de induzirmos uma alta variabilidade intra-conglomerados. Em particular, dividindo os 600 indivíduos em 20 subpopulações idênticas de tamanho 30 e ordenando internamente cada subpopulação ao acaso, criamos uma distribuição espacial em que cada conglomerado de tamanho igual a um submúltiplo de 30 será, a menos do fator de correção para população finita (f_{cpf}), equivalente a uma AAS-SR, de igual tamanho, da popula-

ção. Veremos que, nestes casos, a correção de Rao e Scott aprofunda as distorções na distribuição da Estatística de Pearson.

No apêndice A1 detalhamos melhor a construção destas distribuições espaciais alternativas dos indivíduos da população.

Nos casos dos conglomerados homogêneos, a alternativa "bootstrap" parece desempenhar-se tão bem quanto a correção de Rao e Scott. Já com conglomerados heterogêneos, com distribuição interna de Sexo e Idade semelhante à da população, o "bootstrap" continua apresentando performance satisfatória, enquanto a correção de Rao e Scott aprofunda, em vez de corrigir, as distorções na Estatística de Pearson o que, aliás, pode ser explicado facilmente.

CAPÍTULO 1

APRESENTAÇÃO DA TEORIA BÁSICA

Em muitos problemas de levantamento por amostragem em populações finitas, deseja-se testar hipóteses numa tabela de contingência de dupla entrada. Por exemplo, no caso do teste de independência de dois fatores é comum o uso da Estatística de Pearson, a qual se distribui assintoticamente segundo uma variável aleatória Qui-Quadrado (χ^2_v) com v graus de liberdade, se os dados são levantados segundo um Delineamento Amostral Neutro (AAS-CR). Neste capítulo apresentaremos a teoria conhecida.

1.1 - TABELAS DE CLASSIFICAÇÃO CRUZADA.

Também denominadas Tabelas de Contingência, são construídas classificando-se cada um dos indivíduos em estudo, simultaneamente, segundo duas variáveis categóricas. Uma tabela de contingência pode ser construída a partir de todos os indivíduos de uma população, ou a partir de uma amostra daquela população, quando então a denominamos tabela de contingência amostral.

Para estabelecermos a notação a ser utilizada ao longo deste trabalho, consideremos a seguinte tabela de contingência amostral,

		variável 2						marginal
		1	2	.	.	.	c	
variável 1	1	n_{11}	n_{12}	.	.	.	n_{1c}	n_{1+}
	2	n_{21}	n_{22}	.	.	.	n_{2c}	n_{2+}

	r	n_{r1}	n_{r2}	.	.	.	n_{rc}	n_{r+}
marginal		n_{+1}	n_{+2}	.	.	.	n_{+c}	n

Aqui, r e c representam os números de linhas e de colunas, respectivamente e n_{ij} é o número de indivíduos na (i,j) -ésima cela. Os totais por linha e por coluna são representados por n_{i+} e n_{+j} , respectivamente, para todo i e todo j . O total de indivíduos representados na tabela é n , o tamanho da amostra. Numa tabela de contingência cobrindo todos os indivíduos da população adotaremos a notação N_{ij} , N_{i+} , N_{+j} e N , para representarmos o número de indivíduos por cela, os totais por linha e por coluna e o tamanho da população, respectivamente.

As proporções populacionais de indivíduos por cela, por linha e por coluna denominaremos, respectivamente, p_{ij} , p_{i+} e p_{+j} . Analogamente, para as proporções amostrais, adotamos p_{ij} , p_{i+} e p_{+j} .

Naturalmente, temos as seguintes relações,

$$n_{i+} = \sum_{j=1}^c n_{ij} \quad (1.1)$$

$$n_{+j} = \sum_{i=1}^r n_{ij} \quad (1.2)$$

$$n = \sum_{i=1}^r \sum_{j=1}^c n_{ij} \quad (1.3)$$

e analogamente para N_{i+} , N_{+j} e N . Assim

$$p_{ij} = N_{ij} / N \quad (1.4)$$

$$p_{i+} = N_{i+} / N \quad (1.5)$$

$$p_{+j} = N_{+j} / N \quad (1.6)$$

com $\sum_{i=1}^r \sum_{j=1}^c p_{ij} = 1 \quad (1.7)$

e analogamente para as proporções amostrais.

Definimos ainda o vetor populacional

$$\underline{p} = (p_{11}, p_{12}, \dots, p_{r1}, p_{r2}, \dots, p_{rc-1})' \quad (1.8)$$

e analogamente o vetor amostral $\hat{\underline{p}}$.

1.1.2 - ESTATÍSTICA QUI-QUADRADO

Para se testar a independência entre as duas variáveis numa tabela de contingência, as hipóteses de interesse são :

$$H_0 : h_{ij} = p_{i+} \cdot p_{+j} = 0 \quad ; \quad i=1,2,\dots,r-1 \quad j=1,2,\dots,c-1 \quad (1.9)$$

$$H_1 : h_{ij} = p_{i+} \cdot p_{+j} \neq 0 \quad \text{para pelo menos um } i \text{ ou um } j$$

Karl Pearson em 1900, propôs a seguinte estatística :

$$X^2 = n \sum_{i=1}^r \sum_{j=1}^c \frac{(\hat{p}_{ij} - \hat{p}_{i+} \cdot \hat{p}_{+j})^2}{(\hat{p}_{i+} \cdot \hat{p}_{+j})} \quad (1.10)$$

denominada Estatística de Pearson.

Sob hipótese de amostragem multinomial, como no caso do delineamento neutro (AAS-CR), a Estatística de Pearson se distribui assintoticamente segundo uma variável aleatória Qui-Quadrado com $(r-1) \cdot (c-1)$ graus de liberdade.

1.1.3 - DISTRIBUIÇÃO ASSINTÓTICA

Se os dados provêm de um Delineamento Amostral Neutro (AAS-CR) com tamanho de amostra igual a n , então a distribuição do vetor $\tilde{n} = (n_{11}, n_{12}, \dots, n_{rc})$ é multinomial, isto é :

$$P(\tilde{n}) = \frac{n!}{\prod_{i=1}^r \prod_{j=1}^c n_{ij}} \prod_{i=1}^r \prod_{j=1}^c (p_{ij})^{n_{ij}} \quad (1.11)$$

com :

$$E(n_{ij}) = n \cdot p_{ij} \quad (1.13)$$

$$V(n_{ij}) = n \cdot p_{ij} (1 - p_{ij}) \quad (1.14)$$

Neste caso o estimador de máxima-verossimilhança para p_{ij} é dado por :

$$\hat{p}_{ij} = n_{ij} / n \quad (1.15)$$

Assim, para testar :

$$H_0 : p_{ij} = p_{i+} \cdot p_{+j}$$

a Estatística de razão de verossimilhança é :

$$\lambda = \frac{L(\underline{n}; \underline{p}^*)}{L(\underline{n}; \underline{\hat{p}})} = \frac{\frac{n!}{\prod_{i=1}^r \prod_{j=1}^c (p_{ij}^*)^{n_{ij}}} \prod_{i=1}^r \prod_{j=1}^c p_{ij}^n}{\frac{n!}{\prod_{i=1}^r \prod_{j=1}^c (\hat{p}_{ij})^{n_{ij}}} \prod_{i=1}^r \prod_{j=1}^c \hat{p}_{ij}^n} = \prod_{i=1}^r \prod_{j=1}^c \left[\frac{p_{ij}^*}{\hat{p}_{ij}} \right]^{n_{ij}} \quad (1.16)$$

onde \underline{p}^* é o vetor que maximiza a função de verossimilhança $L(\underline{n}; \underline{p})$ sob H_0 .

Tomando-se o logarítmo, temos

$$\log \lambda = \log \left[\prod_{i=1}^r \prod_{j=1}^c \left[\frac{p_{ij}^*}{\hat{p}_{ij}} \right]^{n_{ij}} \right] =$$

$$= \sum_{i=1}^r \sum_{j=1}^c n_{ij} \cdot \log \left(p_{ij}^* / \hat{p}_{ij} \right). \quad (1.17)$$

Temos aqui o seguinte resultado assintótico (Bickel e Doksum 1977, p. 229)

$$-2 \cdot \log \lambda = -2 \cdot \sum_{i=1}^r \sum_{j=1}^c n_{ij} \cdot \log \left(n \cdot p_{ij}^* / n_{ij} \right) \sim \chi_b^2 \quad (1.18)$$

com $b = (r-1) \cdot (c-1)$, quando $n \rightarrow \infty$.

Mas definindo,

$$E(n_{ij}) = n \cdot p_{ij}^* = m_{ij} \quad (1.19)$$

Logo

$$-2 \cdot \log \lambda = -2 \cdot \sum_{i=1}^r \sum_{j=1}^c n_{ij} \cdot \log \left(m_{ij} / n_{ij} \right) \quad (1.20)$$

a qual é equivalente assintoticamente à Estatística de Pearson (Wilks 1935).

1.1.4 - SIMULAÇÃO VIA MONTE-CARLO

Para determinarmos, via Monte-Carlo, a distribuição exata da Estatística de Pearson, sob diversas condições de delineamento amostral e de tamanho de amostra, criamos uma população hipotética de 600 indivíduos, classificados segundo sexo e idade :

$$\text{Sexo} \left\{ \begin{array}{l} 1 - \text{Feminino} \\ 2 - \text{Masculino} \end{array} \right. \quad \text{Idade} \left\{ \begin{array}{l} 1 - \text{Jovem} \\ 2 - \text{Adulto} \\ 3 - \text{Velho} \end{array} \right. \quad (1.21)$$

A distribuição dos 600 indivíduos segundo as duas variáveis classificatórias acima é dada pela tabela 1.1 .

Tabela 1.1 : Distribuição dos 600 indivíduos na população segundo sexo e faixa etária.

		I D A D E			TOTAL
		1	2	3	
S E X O	1	40	100	60	200
	2	80	200	120	400
	TOTAL	120	300	180	600

Como se pode observar da tabela acima, existe perfeita independência entre as duas variáveis classificatórias, por tanto a hipótese nula é verdadeira.

Inicialmente adotamos o delineamento amostral neutro (AAS-CR), com tamanhos de amostra iguais a 15, 20, 30, 40, 50, 100, 150, 200, 250 e 300 com 1000 repetições Monte-Carlo (MC), em cada caso.

Para compararmos a distribuição exata da Estatística de Pearson, com a distribuição χ^2_b estimamos, com base nas 1000 repetições MC, a probabilidade da Estatística de Pearson (X^2_I) superar o k-ésimo percentil, x_k , da distribuição Qui-Quadrado com o correspondente número de graus de liberdade. Por maior interesse tomamos k igual a 90, 95 e 99.

Os resultados são apresentados na tabela 1.2 . Em cada caso apresentamos também as estimativas da Esperança e da Variância de X^2_I .

Tabela 1.2 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAS-CR e diversos tamanhos amostrais, estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho da Amostra n	$P [X^2_I > x_k]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
50	0.095	0.047	0.006	2.08	3.60
100	0.103	0.050	0.011	1.98	4.09
150	0.110	0.053	0.009	2.03	4.25
200	0.102	0.050	0.008	2.02	4.03
250	0.112	0.052	0.013	2.06	4.32
300	0.088	0.045	0.011	2.00	3.92

Em nenhum dos casos se observou discordância significativa entre os valores observados e os previstos pela distribuição assintótica, como pode ser observado melhor na figura 1.1, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a estimativa da distribuição exata da Estatística de Pearson.

É comum na literatura a advertência contra a utilização do teste Qui-Quadrado quando n é pequeno. Em geral estabelece-se regras que desaconselham o uso do teste quando $n < 5$ para algum (i,j) (Battacharyya e Johnson 1977, Raktos e Hubert 1979, Yule e Kendall 1968). Parece que na inexistência de estudos extensivos sobre o comportamento da Estatística de Pearson para pequenas amostras, os autores têm procurado ficar do lado seguro.

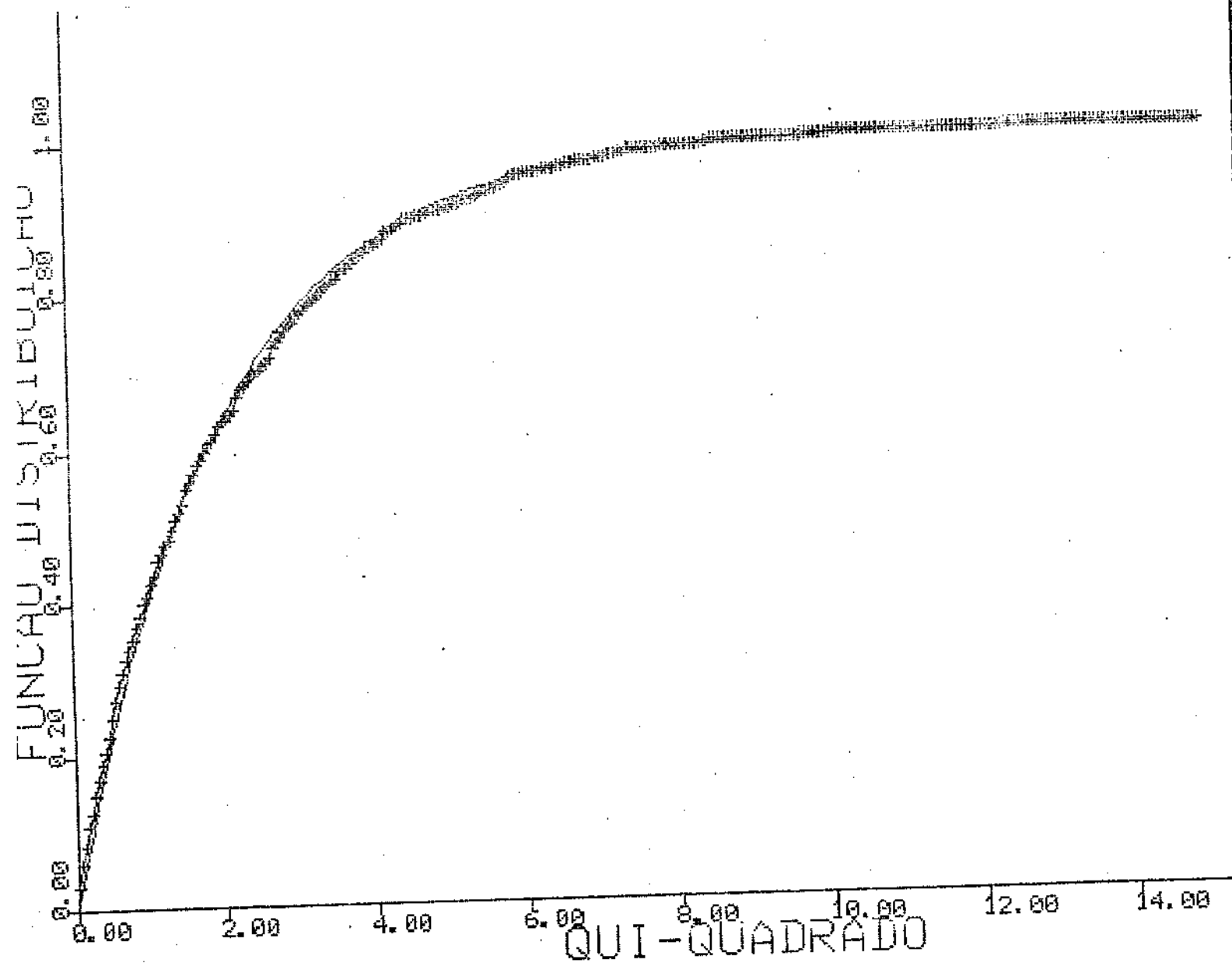
Nossos estudos de Monte-Carlo parecem sugerir que as advertências são um pouco exageradas. Pelo menos no caso particular em que trabalhamos, os resultados assintóticos se constituíram em aproximação bastante satisfatória até para tamanho amostral $n = 15$, o que implica em $E(n_{ij}) = 1$. Os resultados são apresentados na tabela 1.3. Note que estamos aqui falando da distribuição condicional da Estatística de Pearson, dado todos os totais marginais diferentes de zero.

Conforme ficará evidenciado mais adiante, o delineamento amostral, com base no qual os dados foram levantados, é que pode realmente provocar distorções graves, mesmo para amostras grandes. Curiosamente a literatura conhecida não chama muito a atenção para esta questão.

Tabela 1.3 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAS-CR e diversos tamanhos amostrais, estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho da Amostra n	$P \left[\sum_{i=1}^n X_i^2 > x_k \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
40	0.100	0.045	0.002	2.06	3.57
30	0.101	0.048	0.006	2.05	3.69
20	0.104	0.048	0.008	2.13	3.64
15	0.078	0.042	0.003	2.03	3.36

Figura 1.1 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAS-CR com $n = 200$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.



1.2 - DISTORÇÕES DEVIDAS AO EFEITO DO DELINEAMENTO

Geralmente, por razões de comodidade, conveniência, custos, etc., o levantamento dos dados é feito segundo delineamentos amostrais, de maior ou menor complexidade, diferentes do delineamento amostral neutro. Nestes casos $P(\tilde{n})$ pode não mais se ajustar a um modelo probabilístico multinomial, violando assim uma suposição básica para a convergência da Estatística de Pearson à distribuição Qui-Quadrado. Os testes decorrentes podem vir a ser, de acordo com cada caso, mais conservativos ou mais agressivos que o planejado, isto é, o α real menor ou maior que o α nominal, respectivamente.

1.2.1 - EFEITO DE DELINEAMENTO

No caso univariado (Kish 1965, seção 8.2) com a população dividida em 2 categorias, tem-se $\tilde{n} = (n_1, n_2)$, e o efeito do delineamento é dado por

$$deff = \frac{v_{d.a.}(\tilde{n}_1)}{v_{AAS-CR}(\tilde{n}_1)} = \frac{v(\tilde{n}_1)}{n.p(1-p)} \quad (1.22)$$

onde :

$v_{d.a.}$ - denota o operador variância com respeito ao delineamento amostral adotado.

$v_{AAS-CR}(\tilde{n}_1) = n.p.(1-p)$ - é a variância de \tilde{n}_1 sob delineamento amostral neutro (AAS-CR).

p - é a probabilidade de cair um elemento amostrado numa categoria.

Generalizando para o caso multivariado, com a população dividida em mais de 2 categorias, $D = P^{-1} \cdot V$ pode ser pensado como uma extensão natural multivariada do deff onde :

P = matriz de covariância sob o delineamento amostral neutro.

V = matriz de covariância sob o delineamento amostral adotado.

Mas

$$P = \text{diag}(\underline{p}) - \underline{p} \cdot \underline{p}' \quad (1.23)$$

matriz quadrada de ordem $rc-1$.

No caso univariado, o deff indicará, ao ser comparado com a unidade, qual variância é maior. No caso multivariado, a matriz de efeito de delineamento D , deve ser comparada com a matriz identidade e, através da média de seus autovalores e analogamente ao caso univariado, será possível saber qual matriz de variâncias é maior.

CAPÍTULO 2

SIMULAÇÕES DE DIVERSOS DELINEAMENTOS

Neste capítulo apresentaremos, via simulação Monte-Carlo, as distorções na distribuição da Estatística de Pearson para independência em tabelas de contingência, devidas ao delineamento amostral.

Trabalhando sempre com a população hipotética de 600 indivíduos, descrita anteriormente na tabela 1.1, consideramos diversos delineamentos amostrais de uso frequente. Em todos os casos empregamos simulações M.C. com 1000 repetições.

2.1 - AMOSTRAGEM ALEATÓRIA SIMPLES SEM REPOSIÇÃO

Neste caso, as distorções são associadas ao fator de correção por população finita, $1 - n/N$. Para ilustração consideramos, como na construção da tabela 1.2, diversos tamanhos amostrais construindo, de forma análoga, a tabela 2.1.

Tabela 2.1 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAS-SR e diversos tamanhos amostrais, estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho da Amostra n	$P \left[\chi^2_{\frac{1}{k}} > x \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
50	0.079	0.033	0.002	1.83	3.05
100	0.060	0.026	0.007	1.67	3.12
150	0.040	0.017	0.003	1.50	2.14
200	0.022	0.009	0.000	1.32	1.58
250	0.010	0.006	0.000	1.15	1.25
300	0.004	0.000	0.000	0.97	0.86

Verifica-se então, ao contrário do ocorrido para AAS-CR, fortes distorções nos níveis de significância reais do teste Qui-Quadrado. A aplicação da Estatística de Pearson em AAS-SR produz, assim, testes mais conservativos, isto é, com α real menor que o α nominal.

Os gráficos a seguir comparam a distribuição χ^2_2 (linha fina), com a distribuição estimada da Estatística de Pearson (linha grossa), para AAS-SR, com diversos tamanhos amostrais. As distorções, crescentes com n, são agora claramente visíveis.

Figura 2.1 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAS-SR com $n = 50$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

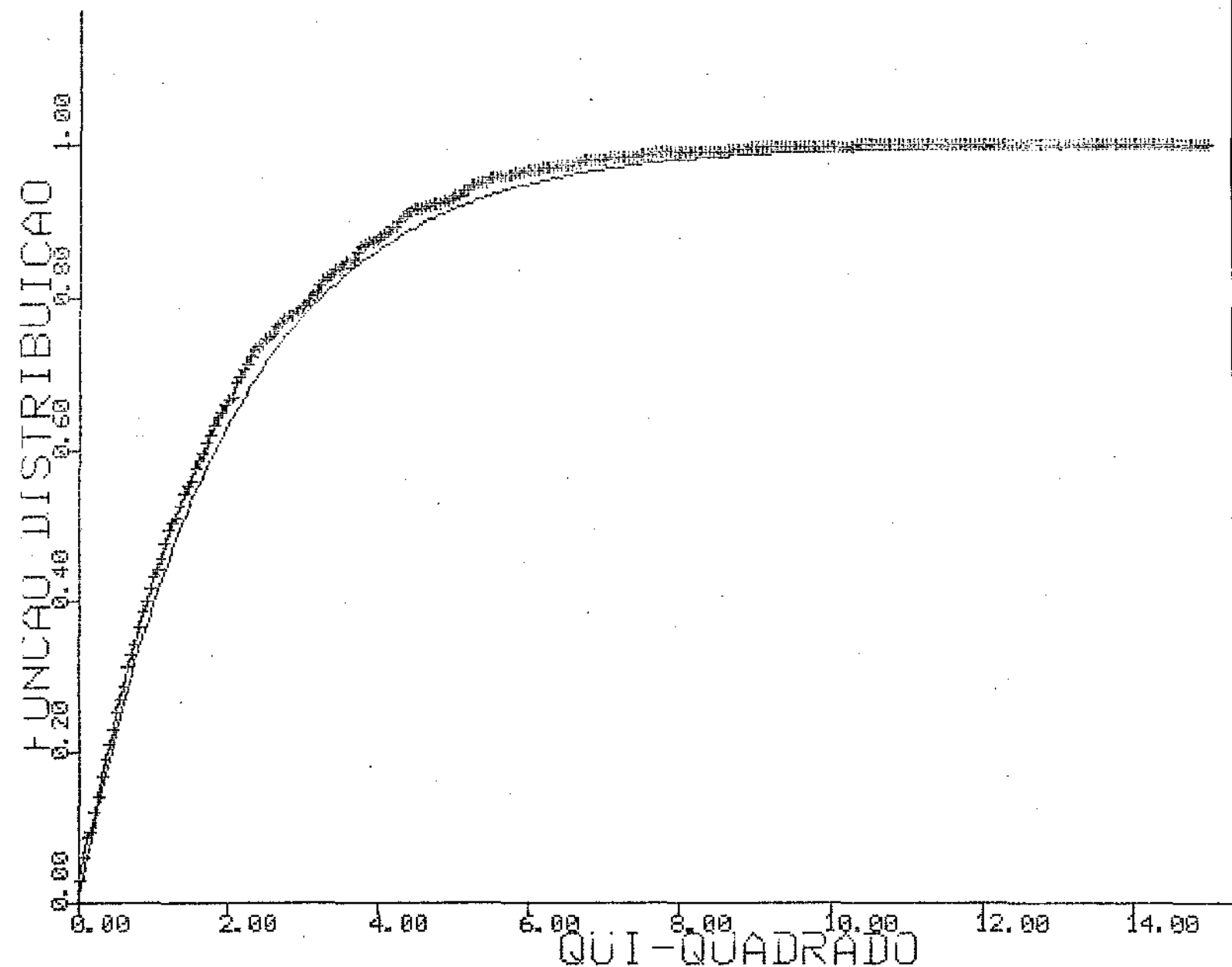


Figura 2.2 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAS-SR com $n = 200$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

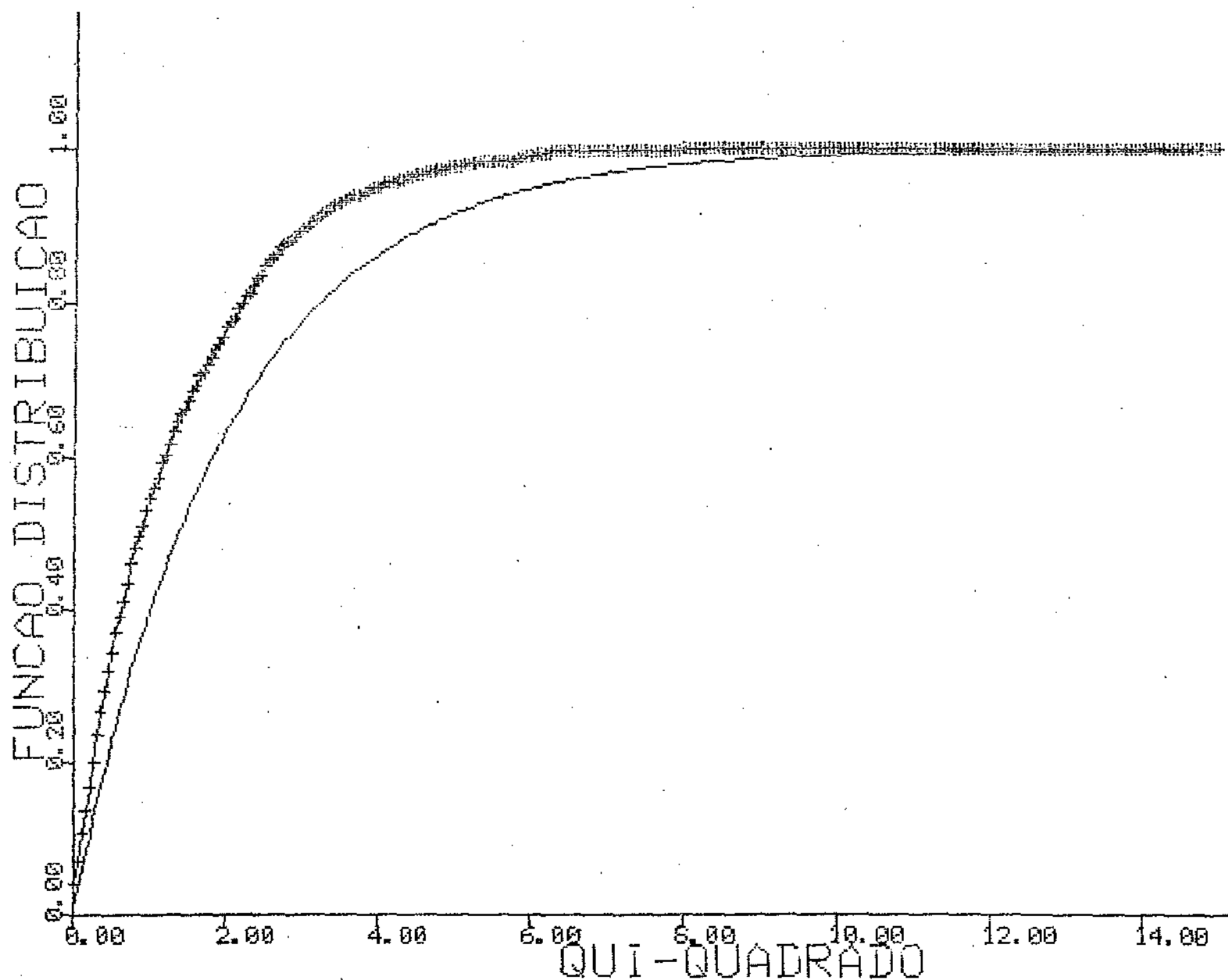
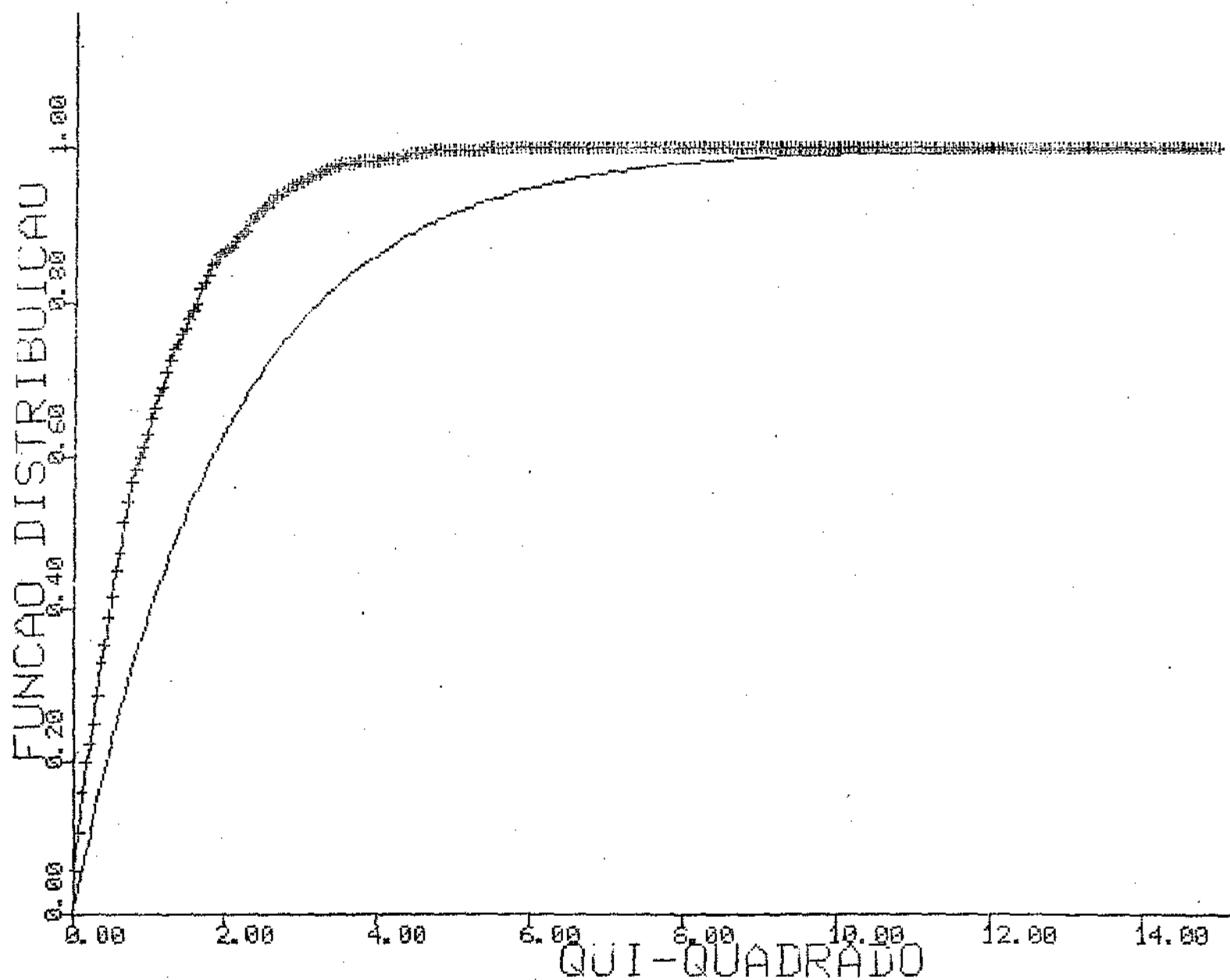


Figura 2.3 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAS-SR com $n = 300$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.



2.2 - AMOSTRAGEM ALEATÓRIA ESTRATIFICADA COM ALOCAÇÃO PROPORCIONAL

Para este caso, os 600 indivíduos da população de trabalho, apresentada na Tabela 1.1, foram divididos em 2 Estratos :

Estrato 1 :

		I D A D E			TOTAL
		1	2	3	
S E X O	1	30	20	30	80
	2	10	40	30	80
	TOTAL	40	60	60	160

Estrato 2 :

		I D A D E			TOTAL
		1	2	3	
S E X O	1	10	80	30	120
	2	70	160	90	320
	TOTAL	80	240	120	440

Para cada tamanho amostral efetuou-se 1000 repetições M.C. A alocação foi proporcional ao tamanho, com AAS-CR dentro de cada estrato.

Os resultados referentes ao comportamento das caudas das distribuições exatas da Estatística de Pearson são apresentadas na seguinte tabela.

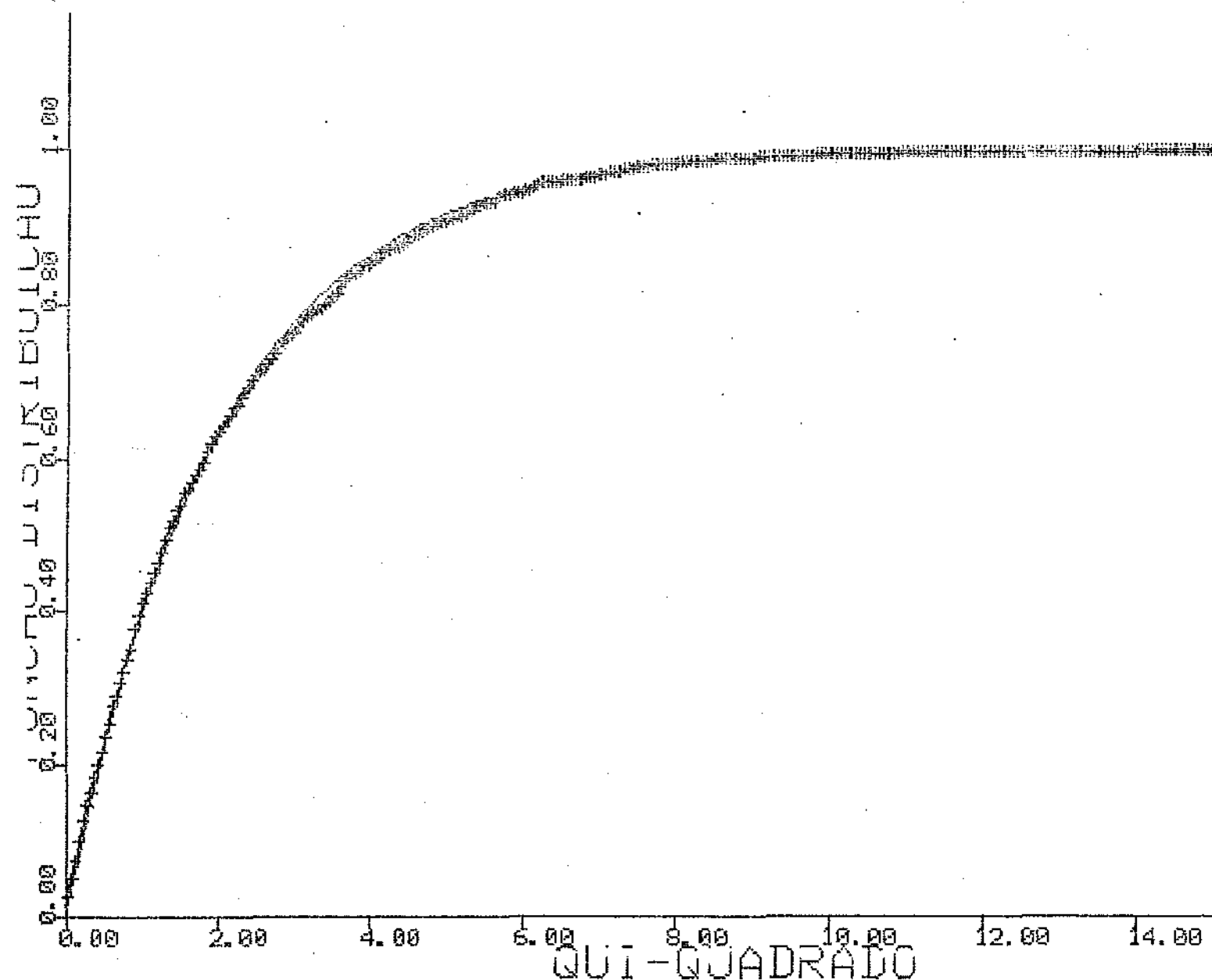
Tabela 2.2 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAE e diversos tamanhos amostrais, estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho da Amostra n	$P \left[\chi^2_I > \chi^2_k \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
50	0.099	0.037	0.006	1.95	3.46
100	0.097	0.052	0.011	1.99	4.02
150	0.081	0.042	0.004	1.94	3.38
200	0.096	0.047	0.010	1.99	3.84
250	0.105	0.043	0.009	1.99	4.03
300	0.108	0.054	0.010	2.04	4.04

Como vemos, apesar da acentuada diferença entre os dois estratos, as distorções nos níveis de significância não parecem importantes. As diferenças entre as frequências esperadas e as observadas não foram significativas ao nível de 5% , excetuando o caso para n=150 e k=99 .

Uma ilustração visual é dada pelo gráfico seguinte, onde a curva de linha fina é a função de distribuição da Qui-Quadrado com 2 graus de liberdade e a curva de linha grossa é a função de distribuição estimada da Estatística de Pearson.

Figura 2.4 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAE com $n = 300$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.



Como pode ser visto na tabela e no gráfico, a distribuição empírica da Estatística de Pearson, no caso de Amostragem Aleatória Estratificada com Alocação Proporcional ao tamanho do estrato, sofre pouca distorção com respeito à distribuição da Qui-Quadrado. Esta pouca distorção já havia sido observada por Nathan (1975), que a ilustra com base em estudos M.C., sugerindo que a perda de poder da Estatística de Pearson e do teste de razão de verossimilhança com respeito ao teste de Wald, neste caso, é pequena, podendo ser considerada desprezível para efeitos práticos.

Na seção 3.2 desenvolvemos uma fundamentação teórica parcial para estes resultados.

2.3 - AMOSTRAGEM ALEATÓRIA POR CONGLOMERADO

Trabalhando com a população hipotética criada, primeiramente ordenamos os indivíduos em sequência no sentido de se formar conglomerados perfeitamente homogêneos. Posteriormente a sequência foi alterada para se induzir conglomerados com configuração interna semelhante à da própria população. No apêndice A1, o processo de condicionamento dos conglomerados pela ordenação da população é descrita em detalhes.

Para as simulações consideramos a população dividida em conglomerados de mesmo tamanho sendo a seleção dos mesmos feita por AAS-CR.

Veremos que as diferentes configurações dos conglomerados gerarão distorções em sentidos diferentes na distribuição exata da Estatística de Pearson. Para ilustração das distorções efetuamos simulações M.C. para diversos tamanhos de conglomerado. O processo de amostragem foi

feito em 1 estágio, com o conglomerado selecionado entrando integralmente para a amostra.

2.3.1 - CONFIGURAÇÃO COM CONGLOMERADOS HOMOGÊNEOS

Esta configuração é uma situação limite para diversos casos de ocorrência prática, caracterizados por uma variabilidade intra conglomerados menor que a variabilidade para a população toda. Por exemplo, amostrando-se residências numa área urbana, é prática comum estabelecer-se um delineamento amostral em 2 estágios, com amostragem por conglomerados no 1º estágio. Nestes casos os conglomerados são quarteirões. É fato conhecido que existe uma certa homogeneidade em termos socio-econômicos e culturais dentro de um mesmo quarteirão que, via de regra, não revela a heterogeneidade existente na cidade toda (Demming 1960).

As simulações foram feitas para tamanhos de conglomerados iguais a 2, 4, 5, 10 e 20, tomando-se em cada caso, diversos tamanhos de amostra.

Os resultados apresentados nas tabelas a seguir revelam uma forte distorção na distribuição exata da Estatística de Pearson, sempre no sentido de se produzir testes muito mais agressivos (α real muito maior que o α nominal) especificados. Os gráficos apresentados ilustram de forma mais completa, as distorções.

Tabela 2.3 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do conglomerado (T.C.) = 2

número de conglo. amostra dos	$P \left[\sum_{i=1}^2 X_i > x \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
25	0.324	0.227	0.100	4.08	14.75
50	0.334	0.243	0.102	4.16	15.98
75	0.330	0.230	0.103	4.08	14.99
100	0.320	0.231	0.103	4.14	17.32
125	0.345	0.236	0.099	4.09	15.88
150	0.305	0.223	0.097	3.99	15.79

O gráfico é o seguinte :

Figura 2.5 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC com T.C. = 2 e nca = 100, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

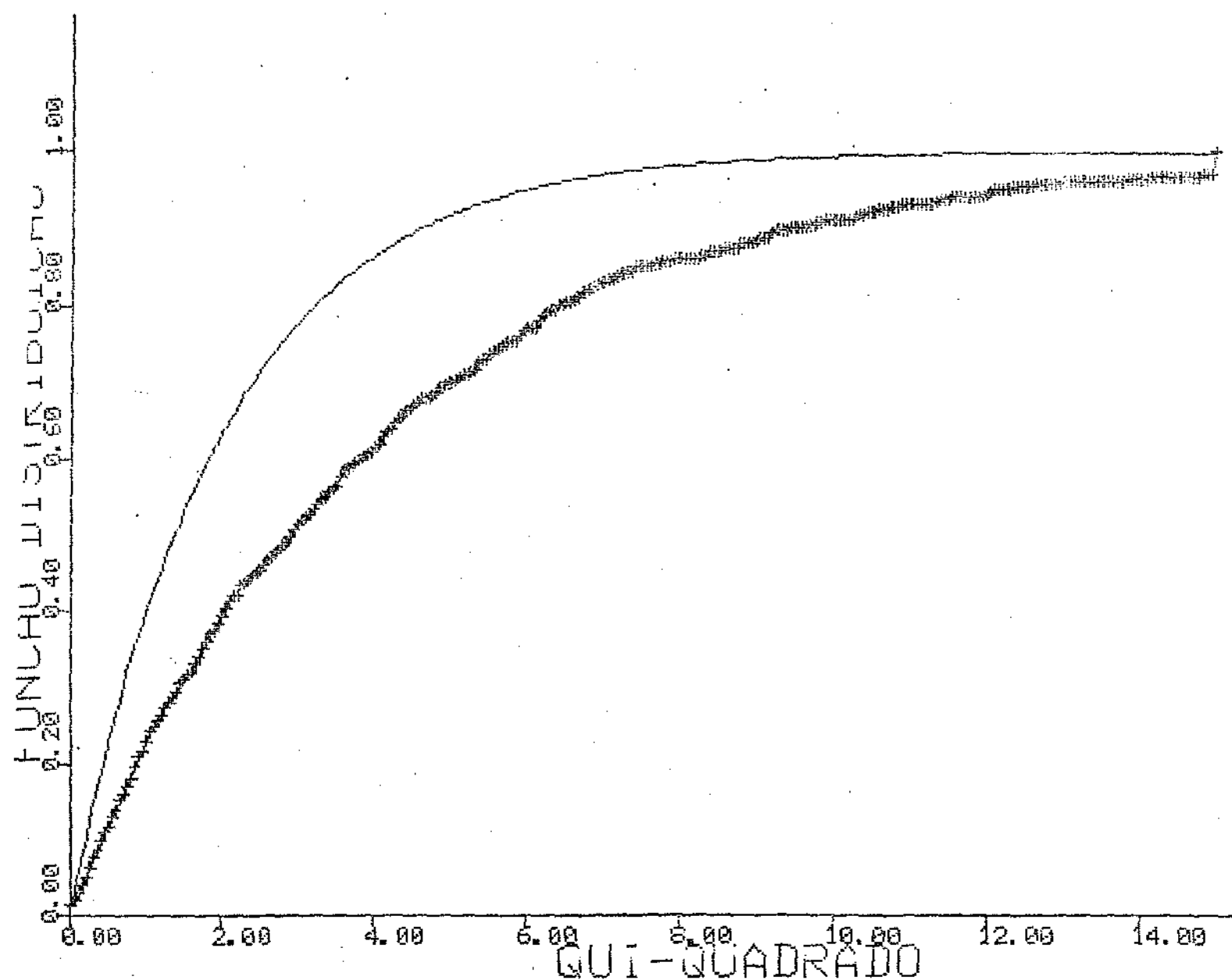


Tabela 2.4 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do Conglomerado (TC) = 4

número de conglo. amostra dos	$P \left[\sum_{i=1}^2 X_i > x \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
25	0.559	0.472	0.312	7.69	55.44
40	0.660	0.458	0.278	7.58	55.62
50	0.562	0.488	0.317	7.81	59.91
60	0.562	0.466	0.298	7.74	56.17
75	0.542	0.460	0.298	7.71	62.57

O gráfico é o seguinte :

Figura 2.6 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC com $T.C. = 4$ e $nca = 50$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

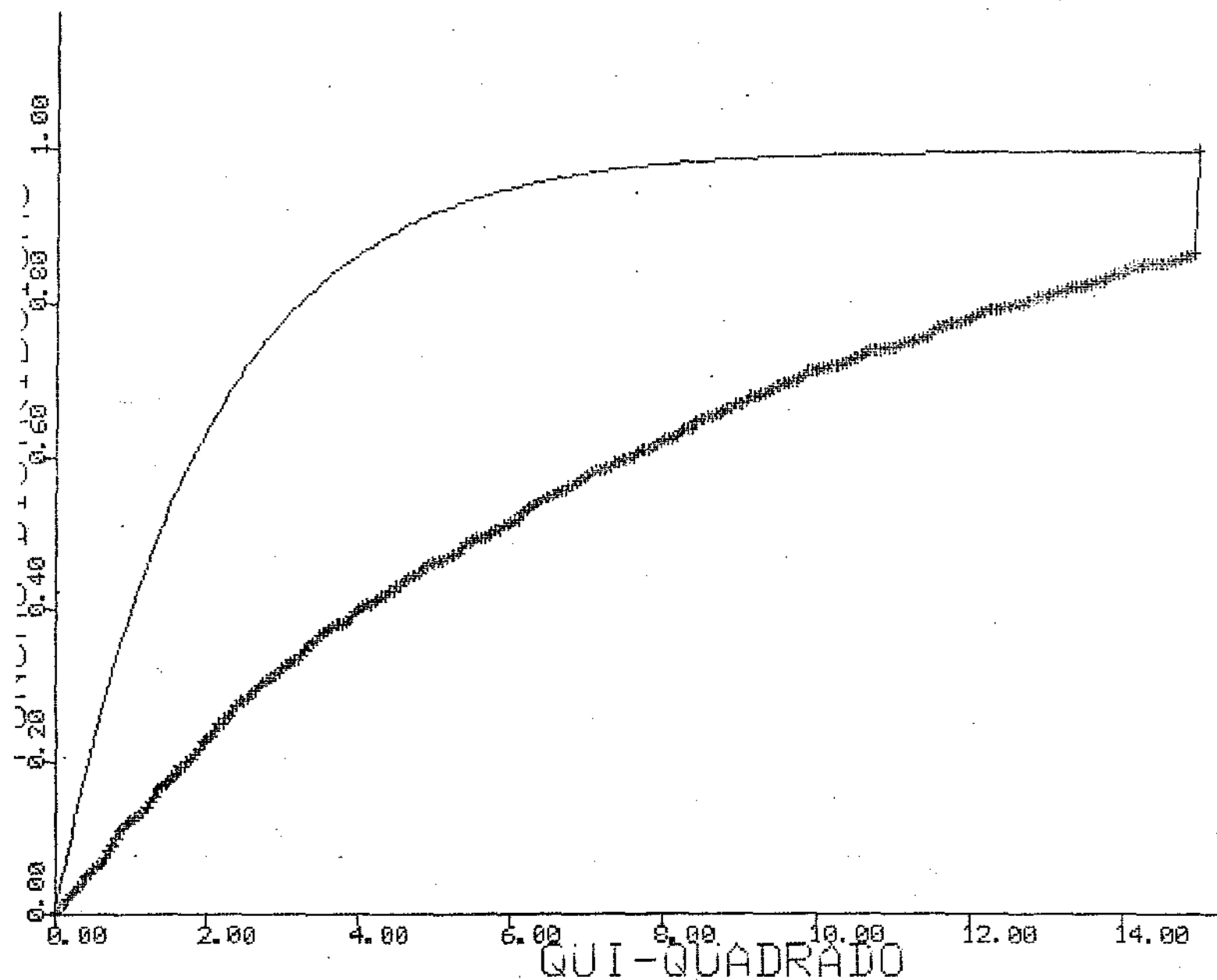


Tabela 2.5 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do Conglomerado (TC) = 5

número de conglo. amostra dos	$P \left[\chi^2_{\frac{2}{1}} > \chi^2_{\frac{2}{k}} \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
10	0.711	0.623	0.412	10.25	72.73
20	0.667	0.606	0.444	10.88	105.60
30	0.629	0.553	0.407	9.99	93.60
40	0.648	0.570	0.417	9.89	81.86
50	0.639	0.551	0.425	10.22	94.79
60	0.626	0.551	0.425	9.48	82.48

O gráfico é o seguinte :

Figura 2.7 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC com $T.C. = 5$ e $nca = 10$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

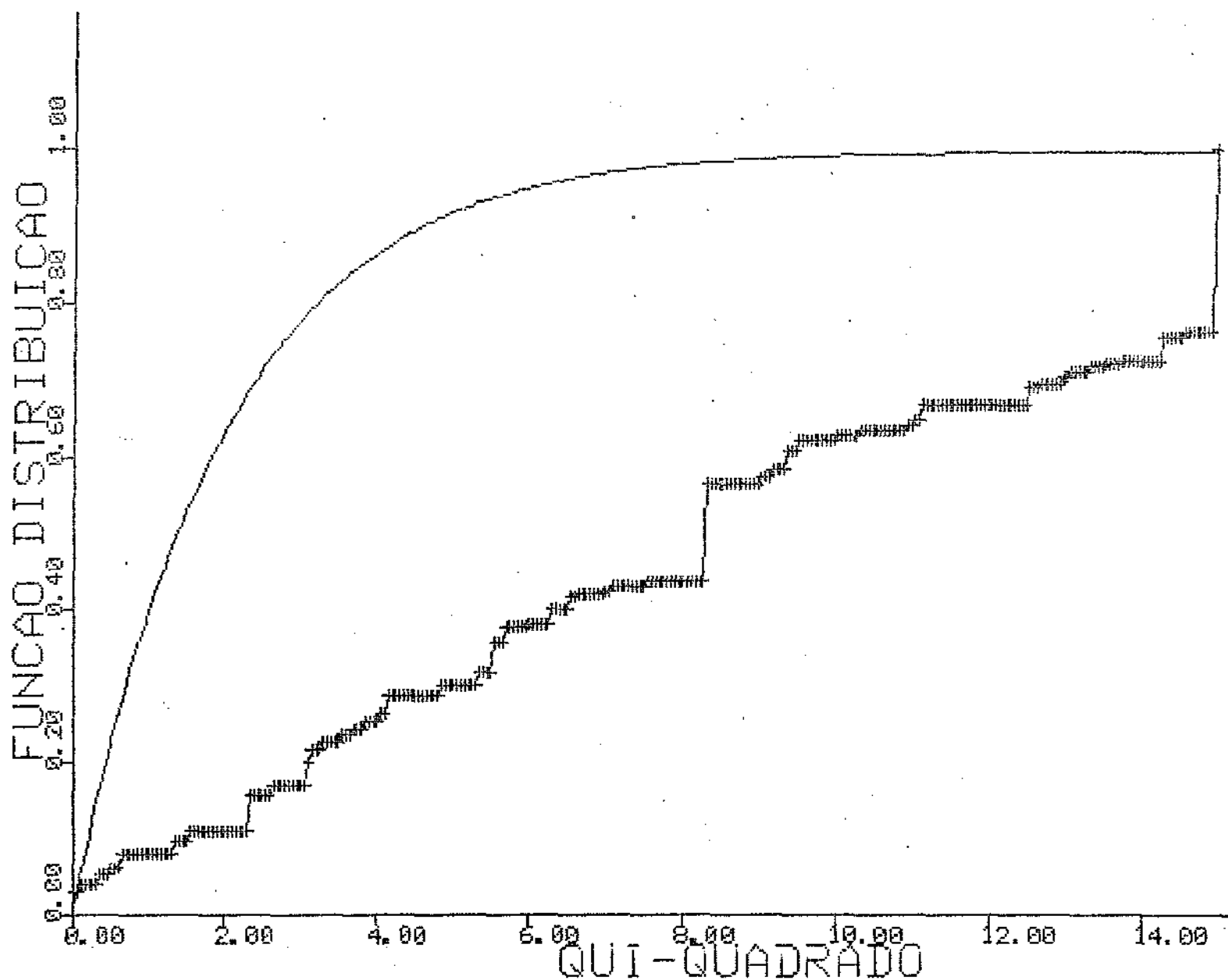


Tabela 2.6 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do Conglomerado (TC) = 10

número de conglo. amostra dos	$P \left[\sum_{i=1}^2 X_i > x \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
10	0.869	0.805	0.686	20.37	330.31
20	0.830	0.796	0.681	21.05	367.00
30	0.787	0.746	0.634	19.20	310.89

O gráfico é o seguinte :

Figura 2.8 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC com T.C. = 10 e nca = 10, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

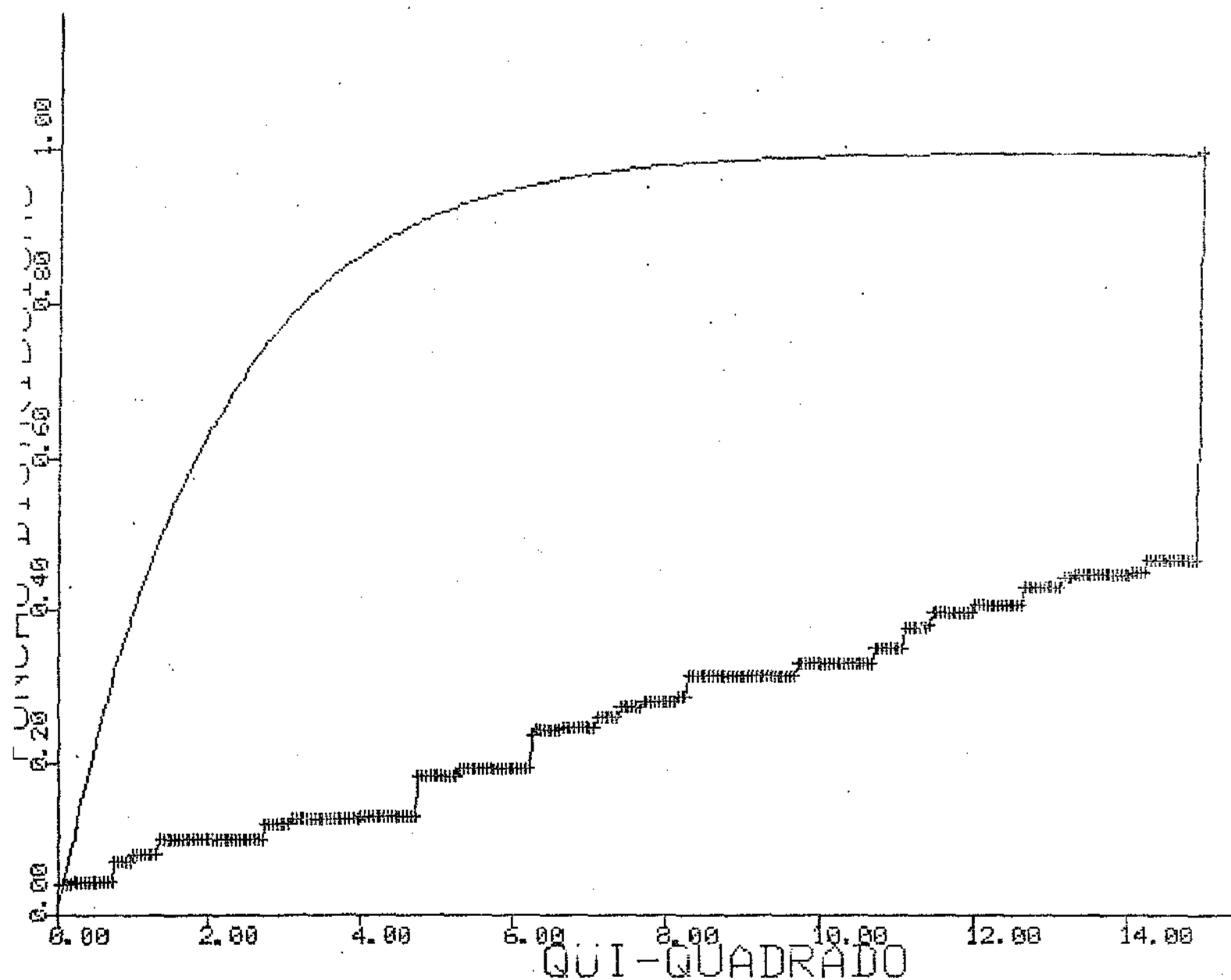


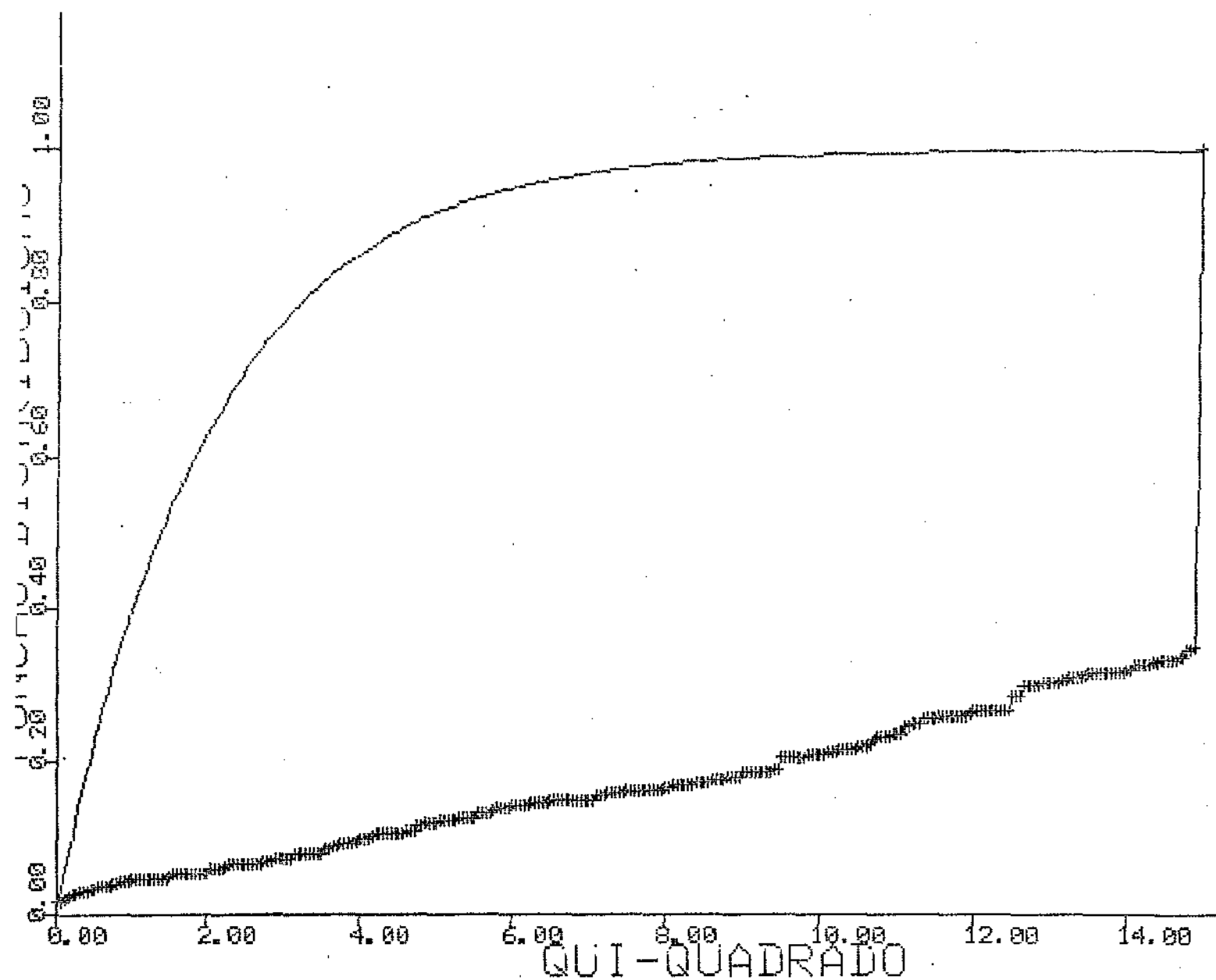
Tabela 2.7 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do Conglomerado (TC) = 20

número de conglo. amostra dos	$P \left[\sum_{i=1}^2 X_i > x_k \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
5	0.754	0.747	0.669	25.00	600.29
10	0.890	0.853	0.813	30.50	728.86
15	0.866	0.830	0.751	28.40	638.70

O gráfico é o seguinte :

Figura 2.9 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC com T.C. = 20 e nca = 10, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.



2.3.2 - CONFIGURAÇÃO COM CONGLOMERADOS SEMELHANTES AOS DA POPULAÇÃO

Também aqui temos uma situação limite para diversos casos de ocorrência na prática, onde cada conglomerado pode ser considerado, aproximadamente, como uma amostra representativa da população (ver Apêndice A1).

Consideramos conglomerados iguais a 3, 5, 6, 10, 15 e 20, tomando-se em cada caso, diversos tamanhos de amostra.

Os resultados apresentados nas tabelas a seguir revelam, ao contrário do ocorrido com conglomerados homogêneos, uma distorção na distribuição exata da Estatística de Pearson, sempre no sentido de se produzir testes mais conservativos que os valores especificados. Os gráficos apresentados ilustram de forma mais completa as distorções.

Tabela 2.8 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do Conglomerado (TC) = 3

número de conglo. amostrados	$P \left[X \geq \frac{2}{k} \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
15	0.081	0.034	0.002	1.54	2.55
35	0.078	0.035	0.005	1.21	2.52
50	0.079	0.035	0.007	1.39	2.55
65	0.080	0.038	0.003	1.56	2.54
85	0.079	0.034	0.009	2.31	2.71
100	0.096	0.051	0.013	2.93	3.19

O gráfico é o seguinte :

Figura 2.10 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC com T.C. = 3 e nca = 15, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

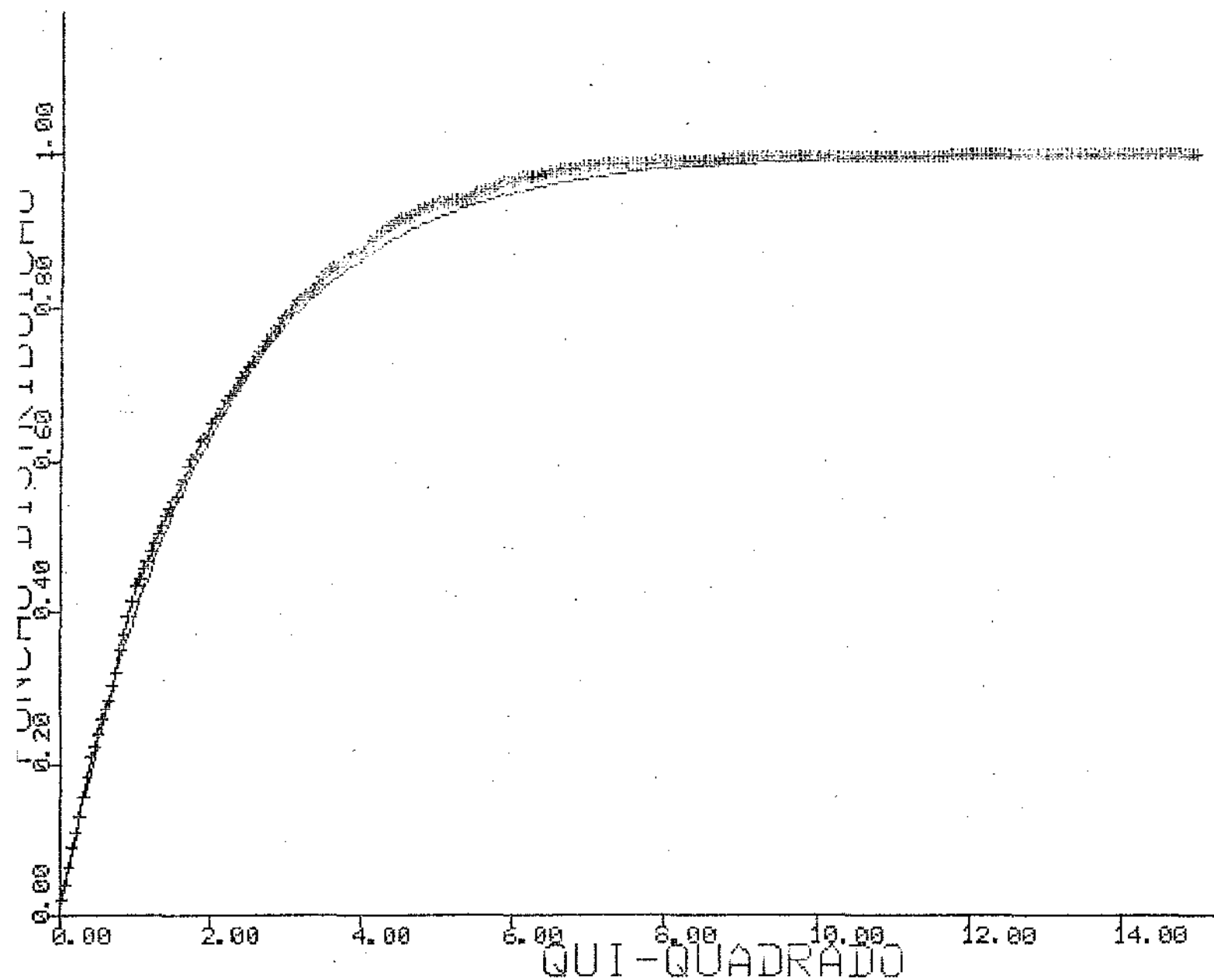


Tabela 2.9 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do Conglomerado (TC) = 5

número de conglo. amostrados	$P \left[\sum_{i=1}^2 X_i^2 > x_k \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
10	0.079	0.036	0.005	0.95	2.54
20	0.053	0.022	0.000	0.48	1.87
30	0.105	0.049	0.006	1.24	2.95
40	0.066	0.034	0.004	2.19	2.52
50	0.065	0.026	0.003	0.92	2.31
60	0.070	0.038	0.014	2.86	2.83

O gráfico é o seguinte :

Figura 2.11 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC com T.C. = 5 e nca = 20, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

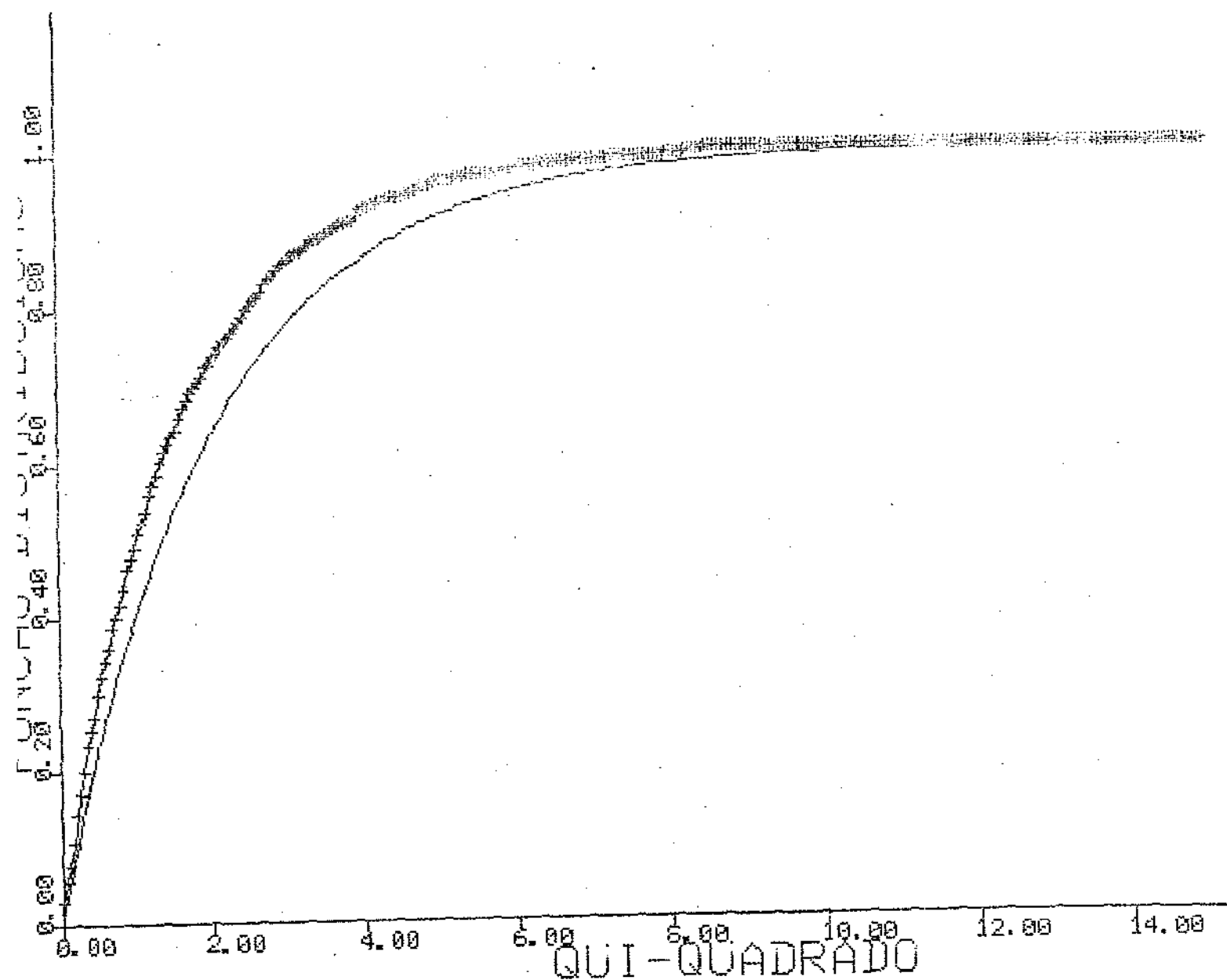


Tabela 2.10 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do Conglomerado (TC) = 6

número de conгло. amostra dos	$P \left[\chi^2_{\frac{2}{I}} > x_k \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
15	0.045	0.018	0.000	1.07	1.72
20	0.041	0.021	0.003	0.58	1.82
25	0.067	0.029	0.006	1.92	2.44
35	0.056	0.017	0.002	0.68	1.77
45	0.074	0.031	0.008	0.76	2.75
50	0.071	0.029	0.006	1.44	2.38

O gráfico é o seguinte :

Figura 2.12 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC com T.C. = 6 e nca = 20, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

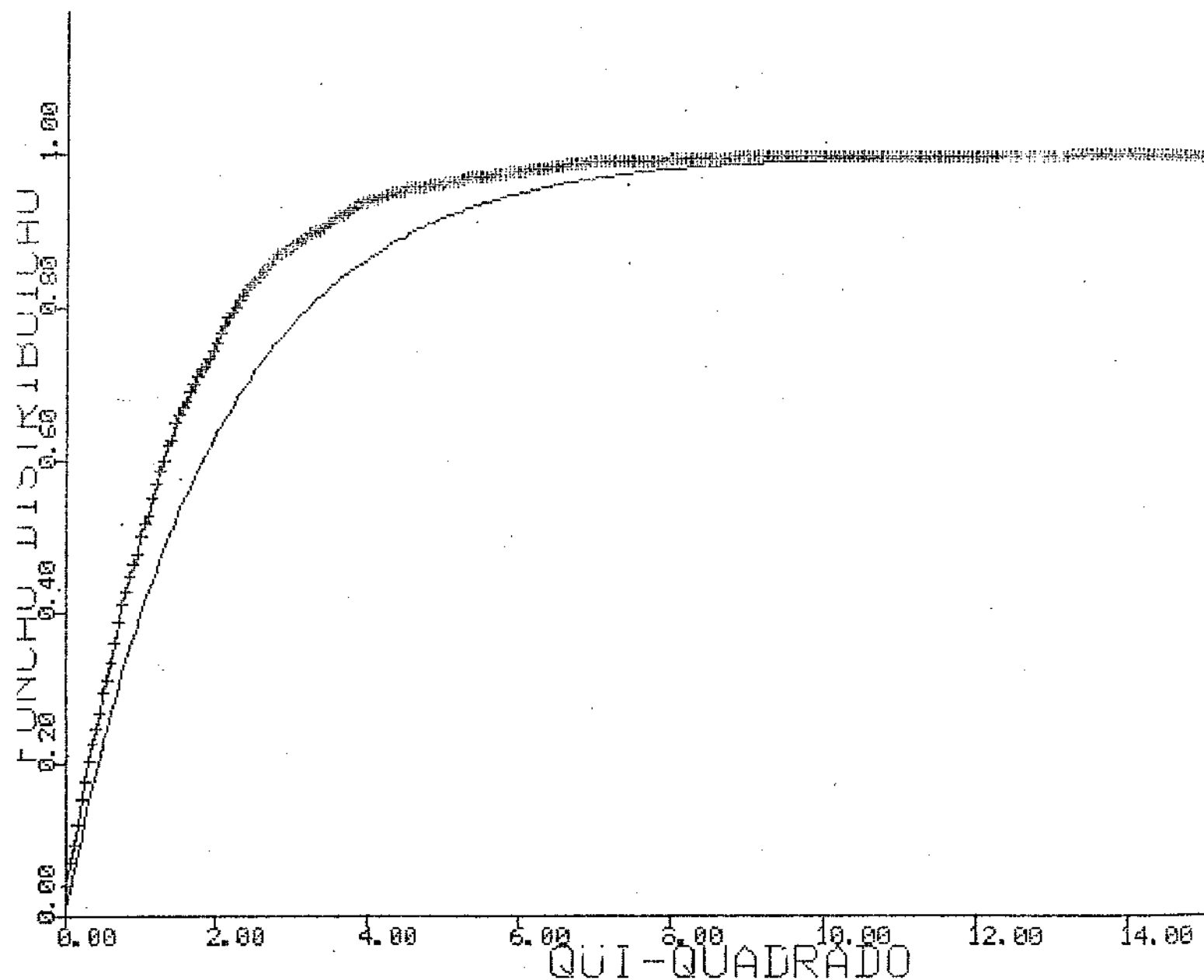


Tabela 2.11 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do Conglomerado (TC) = 10

número de conglo. amostra dos	$P \left[\chi^2_{\frac{2}{1}} > \chi^2_k \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
5	0.054	0.027	0.006	3.06	2.30
10	0.067	0.036	0.008	3.07	2.64
15	0.055	0.028	0.003	3.71	2.07
20	0.021	0.007	0.001	1.00	1.22
25	0.035	0.011	0.001	1.24	1.53
30	0.033	0.015	0.003	0.50	1.69

O gráfico é o seguinte :

Figura 2.13 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC com T.C. = 10 e nca = 20, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

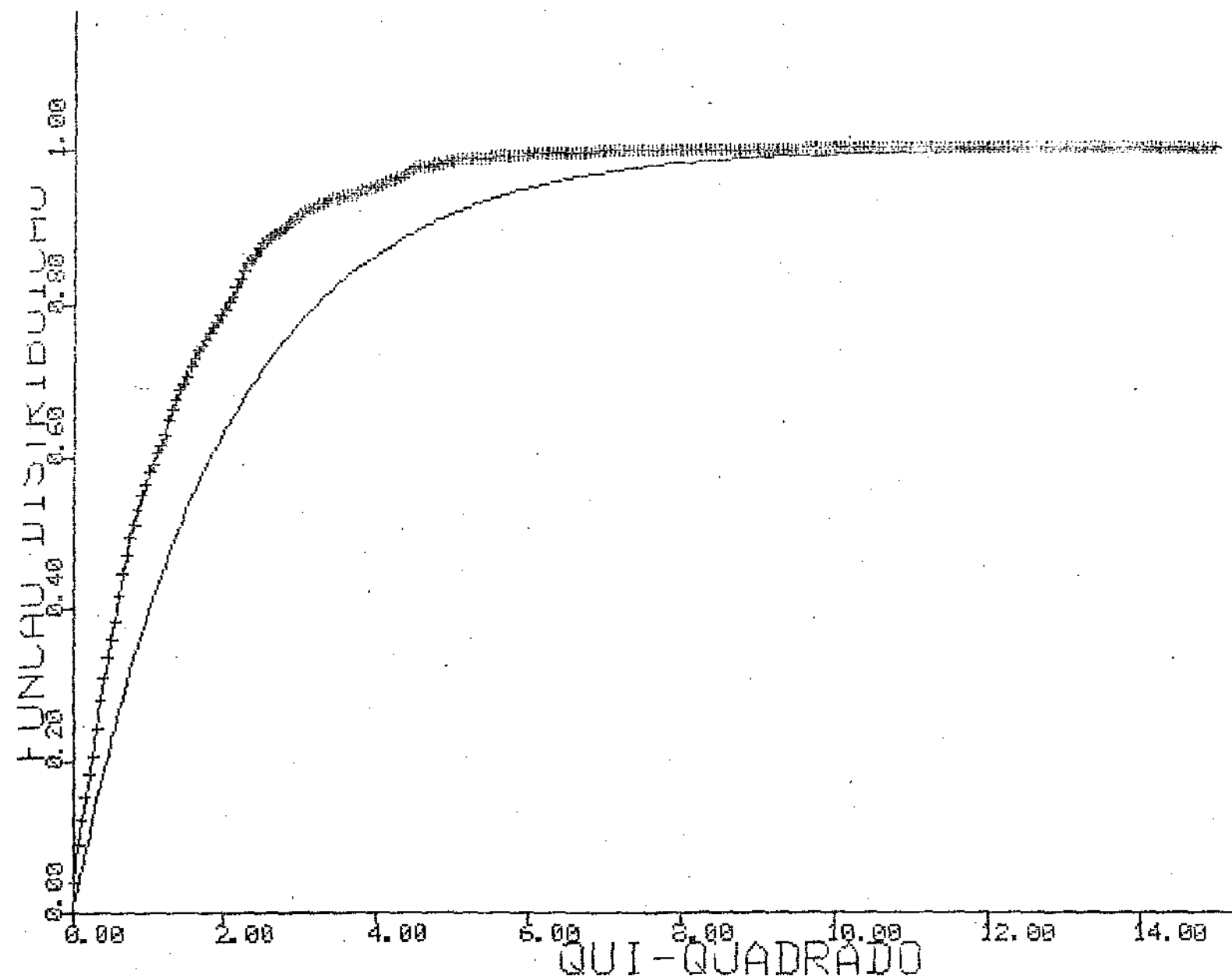


Tabela 2.12 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do Conglomerado (TC) = 15

número de conglo. amostra dos	$P \left[\sum_{i=1}^2 X_i^2 > x_k \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
5	0.007	0.003	0.000	0.68	0.67
7	0.018	0.007	0.002	0.90	1.10
10	0.014	0.005	0.000	1.06	0.96
14	0.007	0.001	0.000	0.52	0.68
17	0.003	0.001	0.000	1.05	0.47
20	0.011	0.003	0.000	0.46	0.83

O gráfico é o seguinte :

Figura 2.14 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC com T.C. = 15 e nca = 14, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

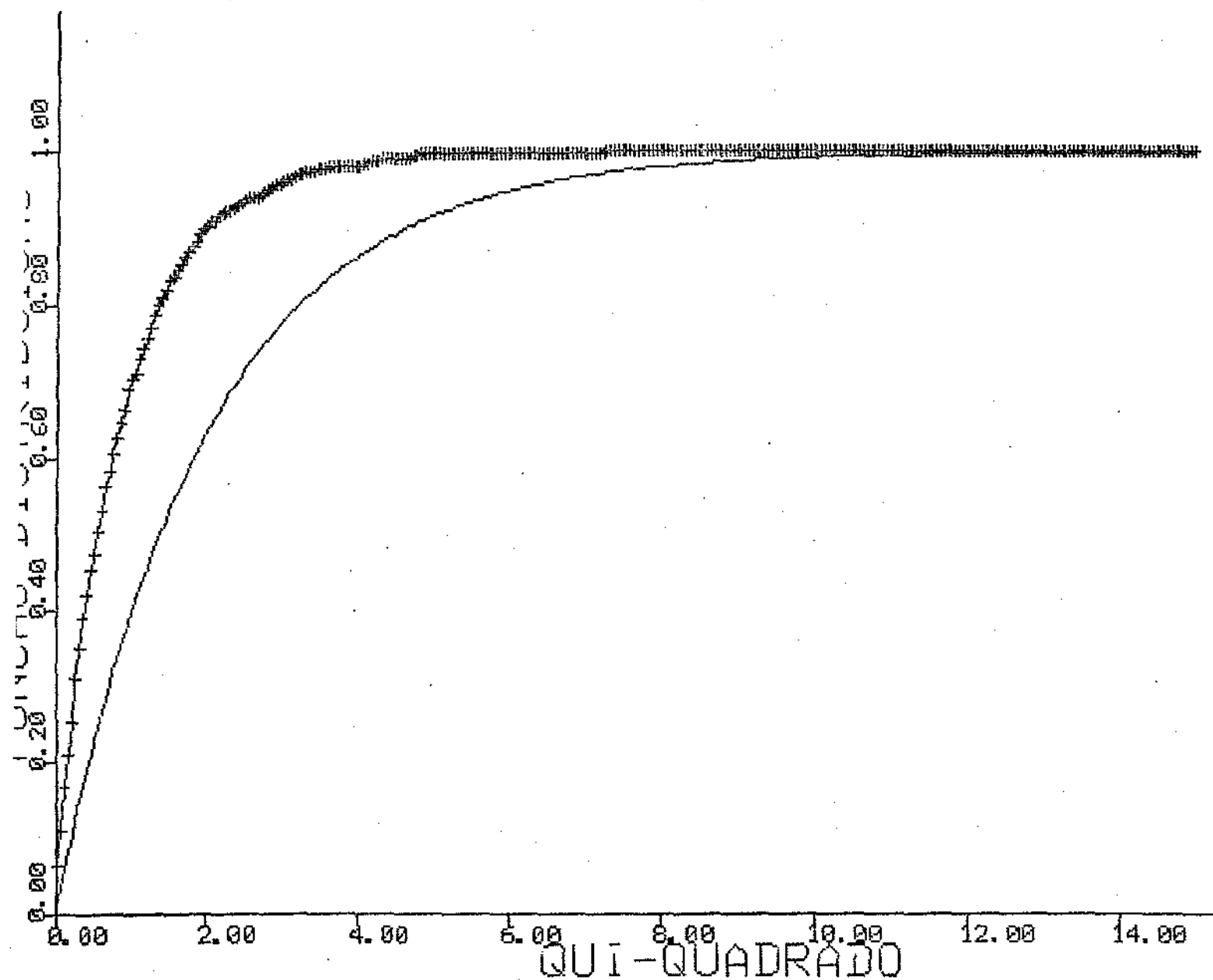


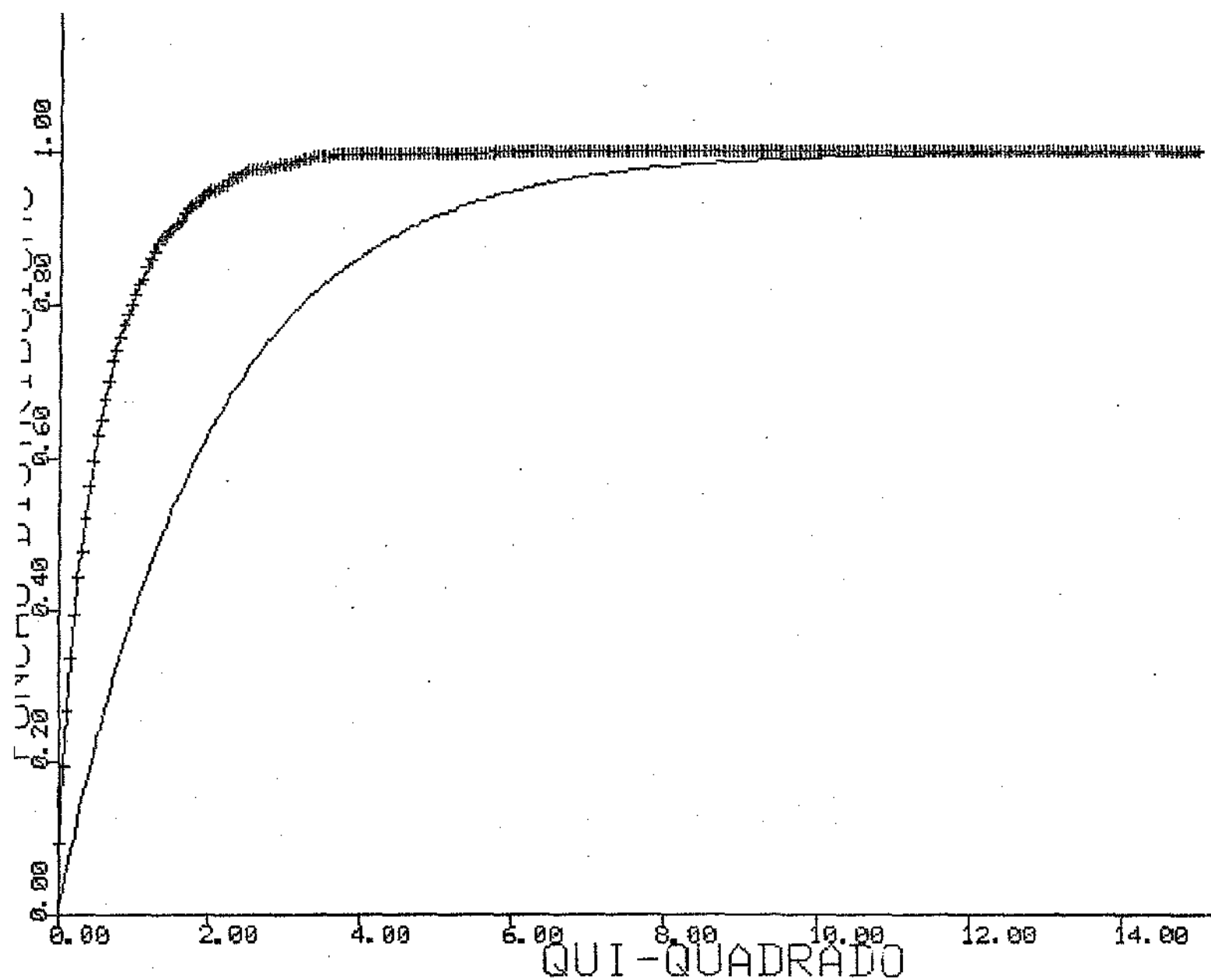
Tabela 2.13 Comparação da cauda da distribuição exata da Estatística de Pearson, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do Conglomerado (TC) = 20

número de conglo. amostra dos	$P \left[\sum_{i=1}^2 X_i > x_k \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
5	0.001	0.000	0.000	0.31	0.38
10	0.005	0.004	0.000	0.77	0.54
15	0.021	0.007	0.004	1.51	1.53

O gráfico é o seguinte :

Figura 2.15 Comparação da distribuição exata da Estatística de Pearson, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC com T.C. = 20 e nca = 5, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.



2.4 - COMENTARIOS E CONCLUSOES

Como vemos, trabalhando com duas situações extremas em termos da relação entre variabilidade dentro de e entre conglomerados, as distorções possíveis na distribuição exata da Estatística de Pearson com respeito à Qui-Quadrado, podem ser muito fortes e em ambas as direções: no sentido de produzir testes mais ou menos conservativos que o valor especificado. Assim, qualquer procedimento corretivo geral, que não leve em consideração a direção da distorção, se produzirá melhores resultados em alguns casos, certamente que produzirá resultados piores em outros.

Estes casos considerados se constituem limites extremos para situações reais encontráveis. Desta forma, as distorções aqui ilustradas são também limites extremos para as distorções que poderão ocorrer em situações práticas.

Efetuamos também alguma exploração MC em situações intermediárias. Em particular, partindo da população em conglomerados homogêneos descrita anteriormente, invertemos, aleatoriamente, a posição de m pares de indivíduos na população. Fazendo $m \rightarrow \infty$, teríamos a população ordenada de forma perfeitamente casual, com nenhuma associação entre indivíduos num mesmo conglomerado. Para m finito, igual a 100, por exemplo, cria-se uma situação parecida com o que ocorre na realidade, com conglomerados apresentando alguma - mas não perfeita - homogeneidade.

Nestes casos, as distorções verificadas se situavam entre a situação de não distorção (AAS-CR) e a situação extrema de AAC com conglomerados homogêneos, que apresentamos.

CAPÍTULO 3

CORREÇÕES PARA ALGUNS CASOS

Até agora temos mostrado empiricamente, por meio de tabelas e gráficos, que o delineamento amostral pode causar distorções na distribuição da Estatística de Pearson, aplicada à análise de independência de duas variáveis numa tabela de contingência em populações finitas. Neste capítulo se procurará estabelecer uma fundamentação teórica para estas distorções. Estudaremos também, para certos delineamentos mais comuns, algumas correções alternativas para as distorções. Os desempenhos destas correções alternativas serão comparados empiricamente.

3.1 - O TESTE DE WALD

Para se testar uma hipótese geral sobre \underline{p} , o vetor de proporções por celas, o teste de Wald é correto, desde que se disponha de estimares adequados, sob o delineamento amostral adotado, da matriz de covariância.

Suponhamos a seguinte hipótese nula :

$$H_0 : h_i(\underline{p}) = 0 \quad , \quad i=1,2, \dots ,b \quad (3.1)$$

contra a alternativa,

$$H_1 : h_i(\underline{p}) \neq 0 \quad , \quad \text{para pelo menos um } i \quad .$$

com $h_i(\underline{p})$ uma função real no domínio do vetor \underline{p} .

Assumindo que $d h_i(\underline{p}) / d \underline{p}_j$ é contínua numa vizinhança do ver-

ondeiro \underline{p} para $j = 1, 2, \dots, K-1$ e que a matriz :

$$G(\underline{p}) = \left[\frac{d h_i(\underline{p})}{d p_j} \right] \quad (3.2)$$

tem posto b , então segue-se do Teorema Central do Limite e pela linearização de Taylor , $\underline{h}(\underline{p}) = \underline{h}(\underline{p}) + G(\underline{p}).(\underline{p} - \underline{p}) + R$ (3.3)

que :

$$n^{1/2} . (\underline{h}(\underline{p}) - \underline{h}(\underline{p})) \xrightarrow{D} N (0 , GVG') \quad (3.4)$$

onde :

$$\underline{h}(\underline{p}) = (h_1(\underline{p}), h_2(\underline{p}), \dots, h_b(\underline{p}))'$$

$$G = G(\underline{p})$$

V = matriz de covariância do vetor \underline{p}

K = número de categorias

\underline{p} = estimador consistente de \underline{p}

Supondo disponível \hat{V} , um estimador consistente de V , será natural basear um teste sobre a estatística generalizada de Wald :

$$\chi^2_W(h) = n . \underline{h}'(\underline{p}) . (\hat{G} . \hat{V} . \hat{G}')^{-1} . \underline{h}(\underline{p}) \quad (3.5)$$

com $\hat{G} = G(\underline{p})$; a qual se distribui assintoticamente como χ^2_b sob H_0 .

Frequentemente, na análise de dados, estimadores de V ou de GVG' não são disponíveis, e no entanto é comum se proceder como se a amos-

tragem fosse aleatória com reposição, utilizando-se a matriz de covariância multinomial, P no lugar de V . Ou seja,

$$X^2_{\sim}(h) = n \cdot \hat{h}'(\hat{p}) \cdot (\hat{G}o \cdot \hat{P}o \cdot \hat{G}o')^{-1} \cdot \hat{h}(\hat{p}) \quad (3.6)$$

onde $\hat{G}o \cdot \hat{P}o \cdot \hat{G}o'$ é qualquer estimador de GPG' que seja consistente (sob o delineamento amostral empregado) quando H_0 é verdadeira.

A distribuição assintótica de $X^2_{\sim}(h)$, é determinada desde formas quadráticas conhecidas (Johnson e Kotz 1970).

TEOREMA 1.- Sob a hipótese nula $H_0 : h(p) = 0$,

$$X^2_{\sim}(h) \approx \sum_{i=1}^b \xi_{o_i} \cdot Z_i^2 \quad (3.7)$$

onde ξ_i 's são os autovalores da matriz :

$$D_h = (GPG')^{-1} \cdot (GVG')^{-1}, \quad \xi_1 > \xi_2 > \dots > \xi_b > 0, \quad (3.8)$$

D_h é a matriz de efeito de Delineamento, Z_1, Z_2, \dots, Z_b são v.a. χ^2_1 independentes e ξ_{o_i} é o valor de ξ_i sob H_0 .

Em geral então, $X^2_{\sim}(h)$ tem distribuição aproximadamente igual a uma soma ponderada, $\sum_{i=1}^b \xi_{o_i} \cdot Z_i^2$ de v.a. χ^2_1 independentes Z_i (Solomon e Stephens 1976).

Os seguintes resultados são consequências imediatas do teorema.

Corolário 1: Seja ϕ_1 o maior dos autovalores de D_h , então

$$\frac{\sum_{h=1}^2 \tilde{X}(h)}{\phi_1} \leq \sum_{i=1}^b Z_i \quad (3.9)$$

onde $\sum_{i=1}^b Z_i$ é distribuída assintoticamente como χ_b^2 sob H_0 .

Ou seja, se um autovalor maior pode ser especificado (ou um razoável limite superior pode ser encontrado), pode-se obter através deste resultado, um teste assintoticamente conservativo (Rao e Scott 1981) por tratar :

$$\frac{\sum_{h=1}^2 \tilde{X}(h)}{\phi^*} \text{ como } \chi_b^2, \text{ onde } \phi^* \geq \phi_1$$

Corolário 2 : Seja ϕ uma constante, então

$$\frac{\sum_{h=1}^2 \tilde{X}(h)}{\phi} \approx \chi_b^2 \quad (3.10)$$

para qualquer p , se e somente se $V = \phi.P$, isto é

$$\text{var}(\hat{p}_i) = \frac{\phi}{n} \hat{p}_i (1 - \hat{p}_i) \quad \text{e} \quad \text{cov}(\hat{p}_i, \hat{p}_j) = -\frac{\phi}{n} \hat{p}_i \hat{p}_j \quad (3.11)$$

o que equivale a dizer que o efeito do Delineamento Amostral é constante em cada cela.

Em particular (Rao e Scott 1981) :

$$\xi_1 = \sup_{\tilde{c}} \left[\frac{\tilde{c}' (GVG') \tilde{c}}{\tilde{c}' (GPG') \tilde{c}} \right] = \sup_{\tilde{c}} \left[\frac{V_{d.a.} \left[\sum_{l=1}^{rc-1} \tilde{c}_l \cdot \hat{a}_l \right]}{V_{AAS-CR} \left[\sum_{l=1}^{rc-1} \tilde{c}_l \cdot \hat{d}_l \right]} \right] \quad (3.12)$$

com \tilde{c} vetor de ordem $rc-1$, \hat{a}_l e \hat{d}_l elementos das matrizes GVG' e GPG' respectivamente e \hat{d}_l é determinado pelos n / n sob AAS-CR .

Assim ξ_1 é o maior deff possível tomado sobre todos os \hat{a}_l individuais e sobre todas as possíveis combinações lineares dos \hat{a}_l . Os outros ξ'_s podem ser chamados deff's generalizados que são consistentemente estimados pelos ξ'_s , os autovalores de $D = (GPG')^{-1} GVG'$.

3.2 - CORREÇÃO DA ESTATÍSTICA DE PEARSON X^2_1

Como se obervou na seção anterior, conhecendo-se os ξ_0 's ou um estimador consistente dos ξ'_s , pode-se obter aproximações precisas para os quantis da distribuição assintótica de $X^2(h)$ usando, por exemplo, o método de Solomon e Stephens (1977). Porém, conhecer os ξ_0 's ou os ξ'_s é essencialmente equivalente a conhecer a matriz de covariância V_0 sob H_0 ou $V = (v_{ij})$ e tendo isto, pode-se sempre construir a estatística generalizada de Wald. Na prática, gostaríamos de ter uma aproximação simples para a distribuição assintótica, que apenas requizesse limitado conhecimento sobre V .

Rao e Scott(1981) propõem uma simples modificação para a Estatística de Pearson.

$$\chi^2_c = \sum_{i=1}^b \frac{\hat{\phi}_i^2}{\hat{\phi}_.} \quad (3.13)$$

a qual é distribuída assintoticamente como :

$$Y = \sum_{i=1}^b \left[\hat{\phi}_{o_i} / \hat{\phi}_{o.} \right] Z_i \quad (3.14)$$

i.e, como uma v. a. χ^2_b sob H_0 onde :

$$\hat{\phi}_. = \frac{1}{b} \sum_{i=1}^b \hat{\phi}_i \quad (3.15)$$

$$\hat{\phi}_{o.} = \frac{1}{b} \sum_{i=1}^b \hat{\phi}_{o_i} \quad (3.16)$$

$$Z_i \sim \chi^2_1$$

e $\hat{\phi}_i$'s são os autovalores de $D = (GPG')^{-1} GVG'$ e são estimadores consistentes dos ϕ_i 's sob H_0 (Rao e Scott 1981).

Esperança de Y é a mesma de χ^2_b ;

$$E(Y) = \sum_{i=1}^b \left(\hat{\phi}_{o_i} / \hat{\phi}_{o.} \right) E(Z_i) = \sum_{i=1}^b \left(\hat{\phi}_{o_i} / \hat{\phi}_{o.} \right) .1 =$$

$$E(Y) = b \quad (3.17)$$

mas sua variância é :

$$\text{Var}(Y) = \text{Var} \left[\sum_{i=1}^b \left(\frac{\phi_{0i}}{\phi_{0.}} \right) Z_i \right] = \sum_{i=1}^b \left(\frac{\phi_{0i}}{\phi_{0.}} \right)^2 \text{Var}(Z_i) =$$

$$\text{Var}(Y) = 2b + 2 / \phi_{0.}^2 \left[\sum_{i=1}^b \left(\phi_{0i} - \phi_{0.} \right)^2 \right] \quad (3.18)$$

a qual é maior que $\text{Var}\left(\chi_b^2\right) = 2.b$ a menos que todos os ϕ_{0i} 's sejam iguais.

3.2.1 - PARA O CASO DE INDEPENDENCIA NA TABELA $r \times c$

Se temos uma tabela de r -linhas e c -colunas, a hipótese de interesse, no caso de independência é :

$$H_0 : h_{ij}(\tilde{p}) = p_{ij} - p_{i+} \cdot p_{+j} = 0 \quad ; \quad i=1,2, \dots, r-1 \quad j=1,2, \dots, c-1 \quad (3.19)$$

$$H_1 : h_{ij}(\tilde{p}) = p_{ij} - p_{i+} \cdot p_{+j} \neq 0 \quad \text{para pelo menos um } i \text{ ou um } j$$

A estatística usual para testar H_0 é :

$$\chi^2_I = n \cdot \sum_{i=1}^r \sum_{j=1}^c \left(\hat{p}_{ij} - \hat{p}_{i+} \cdot \hat{p}_{+j} \right)^2 / \left(\hat{p}_{i+} \cdot \hat{p}_{+j} \right) \quad (3.20)$$

a qual pode ser escrita como (Rao e Scott, 1981) :

$$\chi^2_I = n \cdot h'(\tilde{p}) \cdot \left(P_r^{-1} \otimes P_c^{-1} \right) \cdot h(\tilde{p}) \quad (3.21)$$

onde :

\otimes - denota o produto direto (de Kroneker), das matrizes.

\hat{p}_{ij} é o estimador de p_{ij} sob o delineamento amostral.

$\hat{h}(p)$ é o vetor coluna dos $\hat{h}_{ij}(p)$'s

$P_r = \text{diag}(\hat{p}_r) - \hat{p}_r \cdot \hat{p}_r'$ matriz quadrada de ordem $r-1$.

$P_c = \text{diag}(\hat{p}_c) - \hat{p}_c \cdot \hat{p}_c'$ matriz quadrada de ordem $c-1$.

n = tamanho da amostra

$\hat{p}_r = (p_{1+}, p_{2+}, \dots, p_{(r-1)+})'$

$\hat{p}_c = (p_{+1}, p_{+2}, \dots, p_{+(c-1)})'$

e \hat{P}_r e \hat{P}_c é o valor de P_r e P_c respectivamente, substituindo \hat{p} por \hat{p} .

A estatística generalizada de Wald para testar H_0 é dada por :

$$X_W^2 = n \cdot \hat{h}'(p) \cdot V_h^{-1} \cdot \hat{h}(p) \quad (3.22)$$

onde :

$\frac{1}{n} \cdot V_h$ é um estimador da matriz de covariância de $\frac{1}{n} \cdot V_h$ de $\hat{h}(p)$.

A Estatística de Wald, X_W^2 , é distribuída aproximadamente como uma v.a. χ_b^2 sob H_0 para n suficientemente grande com $b = (r-1)(c-1)$.

O estimador V_h pode ser obtido pelo método de linearização (Fellegi 1980) ou diretamente utilizando, por exemplo, o método de reamostragem Bootstrap, se o delineamento amostral o permite, ou outros métodos.

Uma estatística modificada pode ser :

$$X_c^2 = X_I^2 / \hat{\xi}. \quad (3.23)$$

onde :

$$\hat{\xi}. = \frac{1}{b} \cdot \text{tr} \left[\begin{pmatrix} -1 & -1 \\ P_r & P_c \end{pmatrix} \cdot V_h \right] = \frac{1}{b} \cdot \sum_{i=1}^r \sum_{j=1}^c v_{ij}(h) / \hat{p}_{i+} \cdot \hat{p}_{+j} =$$

$$\hat{\xi}. = \frac{1}{b} \cdot \sum_{i=1}^r \sum_{j=1}^c (1 - \hat{p}_{i+}) (1 - \hat{p}_{+j}) \cdot \hat{\xi}_{ij} \quad (3.24)$$

Aqui, $\frac{1}{n} \cdot v_{ij}(h)$ é um estimador da variância de $h_{ij}(\hat{p})$ e $\hat{\xi}_{ij}$ é o deff estimado de $h_{ij}(\hat{p})$, isto é :

$$\hat{\xi}_{ij} = \frac{v_{ij}(h)}{\hat{p}_{i+} \cdot \hat{p}_{+j} (1 - \hat{p}_{i+}) (1 - \hat{p}_{+j})} \quad (3.25)$$

$$\text{ou } deff = V_{AAS-CR}^{-1} \cdot V_D^{-1} = P^{-1} \cdot V_D^{-1} \quad (3.26)$$

3.2.2 - DISTRIBUIÇÃO ASSINTÓTICA DE X_c^2

A hipótese H_0 dada por 3.19 é um caso especial de 3.1 com $K=rc$ e $b=(r-1) \cdot (c-1)$. Pode-se mostrar que, quando H_0 é verdadeira, GPG' se reduz para $P \otimes P$ (Rao e Scott 1981). Assim, X_I^2 é da forma $X(h)$ dada por 3.6 e pelo teorema 1 temos que,

$$X_I^2 = \sum_{i=1}^b \xi_{oi} Z_i \quad (3.27)$$

sob H_0 e os $\hat{\phi}_i$'s são os autovalores de :

$$D = \begin{pmatrix} P_h & \emptyset & P_r & P_c \end{pmatrix}^{-1} (G.V.G') = \begin{pmatrix} P_h & \emptyset & P_r & P_c \end{pmatrix}^{-1} V \quad (3.28)$$

e $\hat{\phi}_0$ é o valor de $\hat{\phi}_i$ sob H_0 .

Um bom teste é obtido tratando-se a estatística modificada $X^2 / \hat{\phi}_i$, como uma variável aleatória χ^2_b , onde $\hat{\phi}_i = \sum_{i=1}^b \hat{\phi}_i / b$ e os $\hat{\phi}_i$'s são os autovalores de :

$$\hat{D} = \begin{pmatrix} \hat{P}_h & \emptyset & \hat{P}_r & \hat{P}_c \end{pmatrix}^{-1} \hat{V} \quad (3.29)$$

A estatística $X^2 / \hat{\phi}_i$ é distribuída assintoticamente como :

$$\sum_{i=1}^b (\hat{\phi}_{0i} / \hat{\phi}_{0i}) Z_i^2 \quad (3.30)$$

3.3 - CORRECÇÕES APROXIMADAS PARA ALGUNS CASOS

Os delineamentos amostrais não neutros causam distorções na Estatística de Pearson, como pôde ser observado no capítulo 2, tornando-se necessária alguma correção. Esta correção pode ser conseguida através da matriz de covariância do vetor $\tilde{h}(\tilde{p})$ sob o delineamento amostral empregado, de forma análoga às seções anteriores. Nós mostraremos para alguns casos, via simulação de Monte-Carlo, como obter este fator de correção e a melhora no desempenho decorrente da correção. Também apresentaremos tabelas e gráficos para ilustrar os efeitos das correções.

Trabalhamos ainda com a população hipotética de 600 indivíduos descrita anteriormente e considerando os mesmos delineamentos amostrais do capítulo 2. Em cada caso empregamos 1000 simulações e a mesma semente com a qual foram gerados os resultados do capítulo 2, podendo assim melhor comparar o desempenho da Estatística de Pearson.

3.3.1 - AMOSTRAGEM ALEATÓRIA SIMPLES SEM REPOSIÇÃO

Neste caso temos que :

$$V = (1 - n/N).P \quad (3.31)$$

onde :

N - tamanho da População finita.

n - tamanho da amostra.

P - Matriz de covariância Multinomial ou a Variância sob o delineamento amostral neutro (AAS-CR).

$(1 - n/N)$ é o fator de correção para população finita (fcpf)

então :

$$\begin{aligned} D_h &= (GPG')^{-1} \cdot V_h = (GPG')^{-1} \cdot GVG' = (GPG')^{-1} \cdot G(fcpf.P)G' = \\ &= fcpf.(G.P.G')^{-1} \cdot (G.P.G') = \\ D_h &= fcpf.I \end{aligned} \quad (3.32)$$

com I = matriz identidade.

Logo o fator de correção é $\phi = fcpf = (1 - n/N)$.

Aplicando este fator de correção na Estatística de Pearson, conforme visto na seção 3.2, para cada caso respectivo do capítulo 2 ob-

teve-se os seguintes resultados:

Tabela 3.1 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAS-SR e diversos tamanhos amostrais, estimada com base em 1000 repetições MC, com a correspondente distribuição Qui-Quadrado com 2 graus de libertad, com média teórica igual a 2 e variância teórica igual a 4. Esta tabela deve ser comparada com tabela 2.1.

$$fc_{pf} = 1 - n/N \; .$$

Tamanho da Amostra n	$P \left[\overset{2}{\underset{I}{X}} > \underset{k}{x} \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
50	0.104	0.042	0.007	2.00	3.62
100	0.104	0.052	0.011	1.99	4.49
150	0.093	0.046	0.010	1.99	3.81
200	0.094	0.045	0.007	1.99	3.70
250	0.098	0.035	0.009	1.97	3.66
300	0.103	0.043	0.004	1.95	3.44

Verifica-se uma sensível melhora no desempenho da Estatística de Pearson corrigida, com os níveis de significância reais aproximando-se bem dos nominais da distribuição Qui-Quadrado.

Os gráficos a seguir comparam a distribuição $\overset{2}{\underset{2}{X}}$ (linha fina), com a distribuição estimada da Estatística de Pearson (linha grossa), para AAS-SR corrigida com diversos tamanhos amostrais. As melhoras são agora claramente visíveis, comparando-se estes gráficos com os seus correspondentes nas figuras (2.1) a (2.3) .

Figura 3.1 Comparação da distribuição exata da Estatística de Pearson corrigida, estimada com base em 1000 repetições M.C. , com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAS-SR com $n = 50$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.1 .

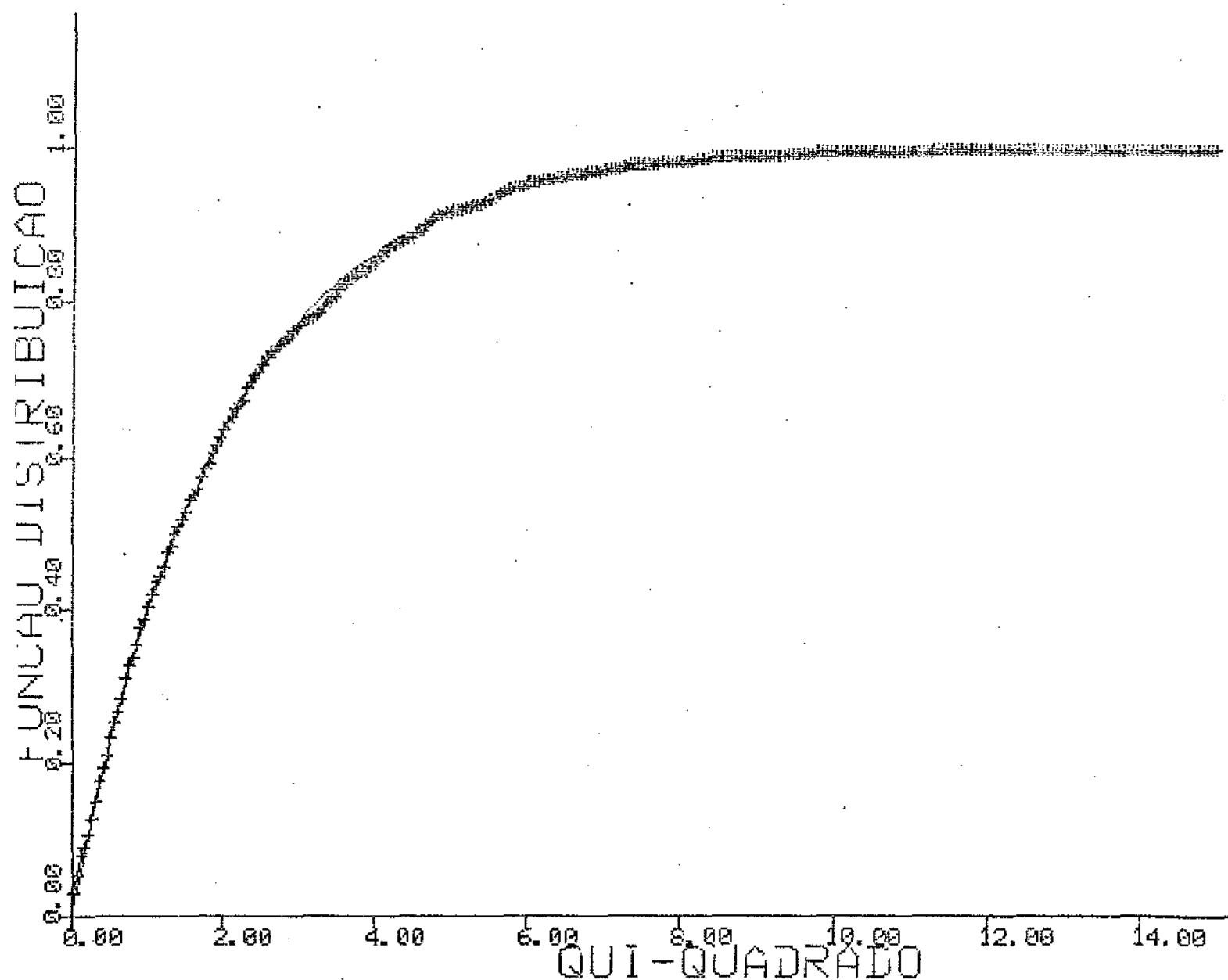


Figura 3.2 Comparação da distribuição exata da Estatística de Pearson corrigida, estimada com base em 1000 repetições M.C. , com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAS-SR com $n = 200$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.2 .

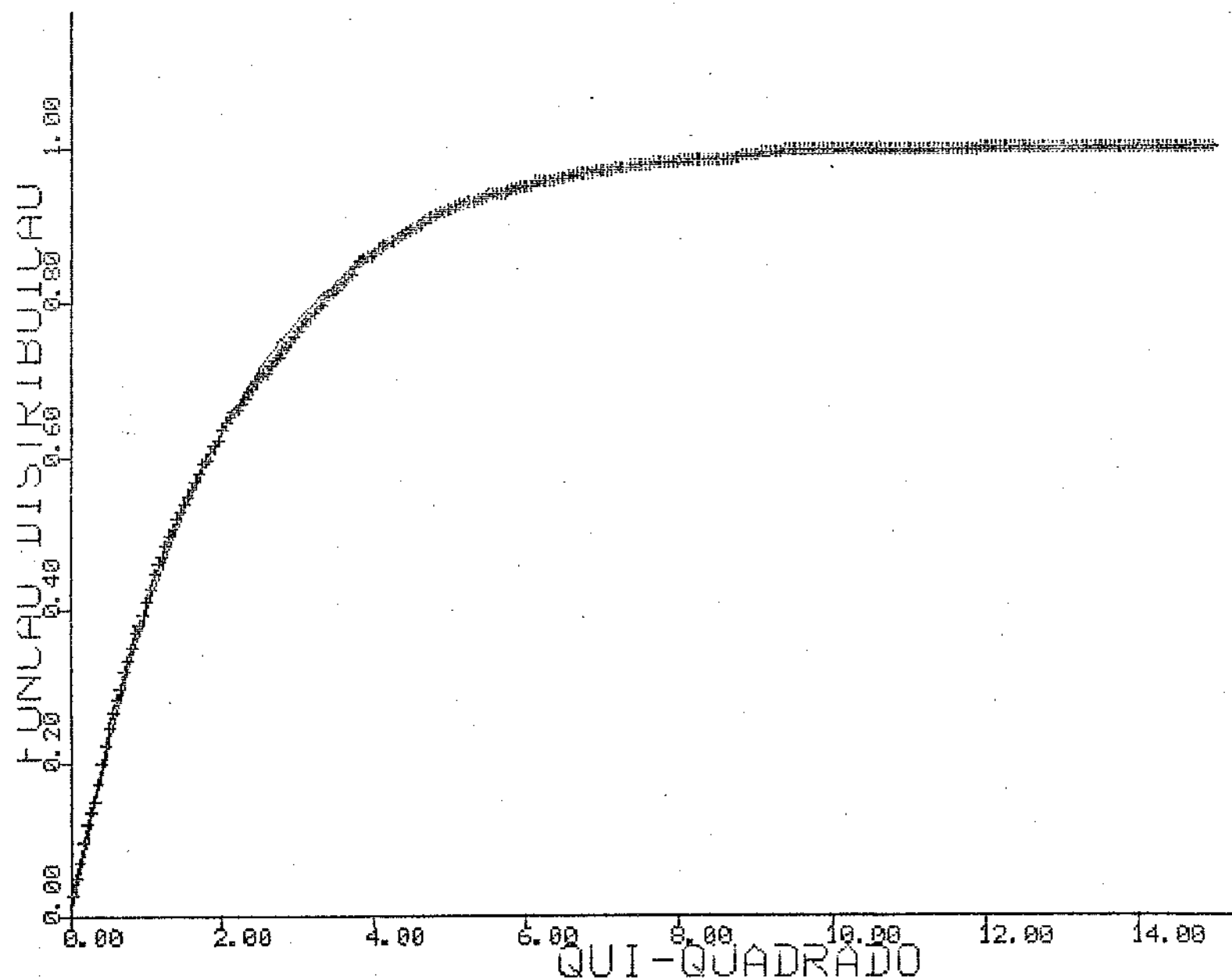
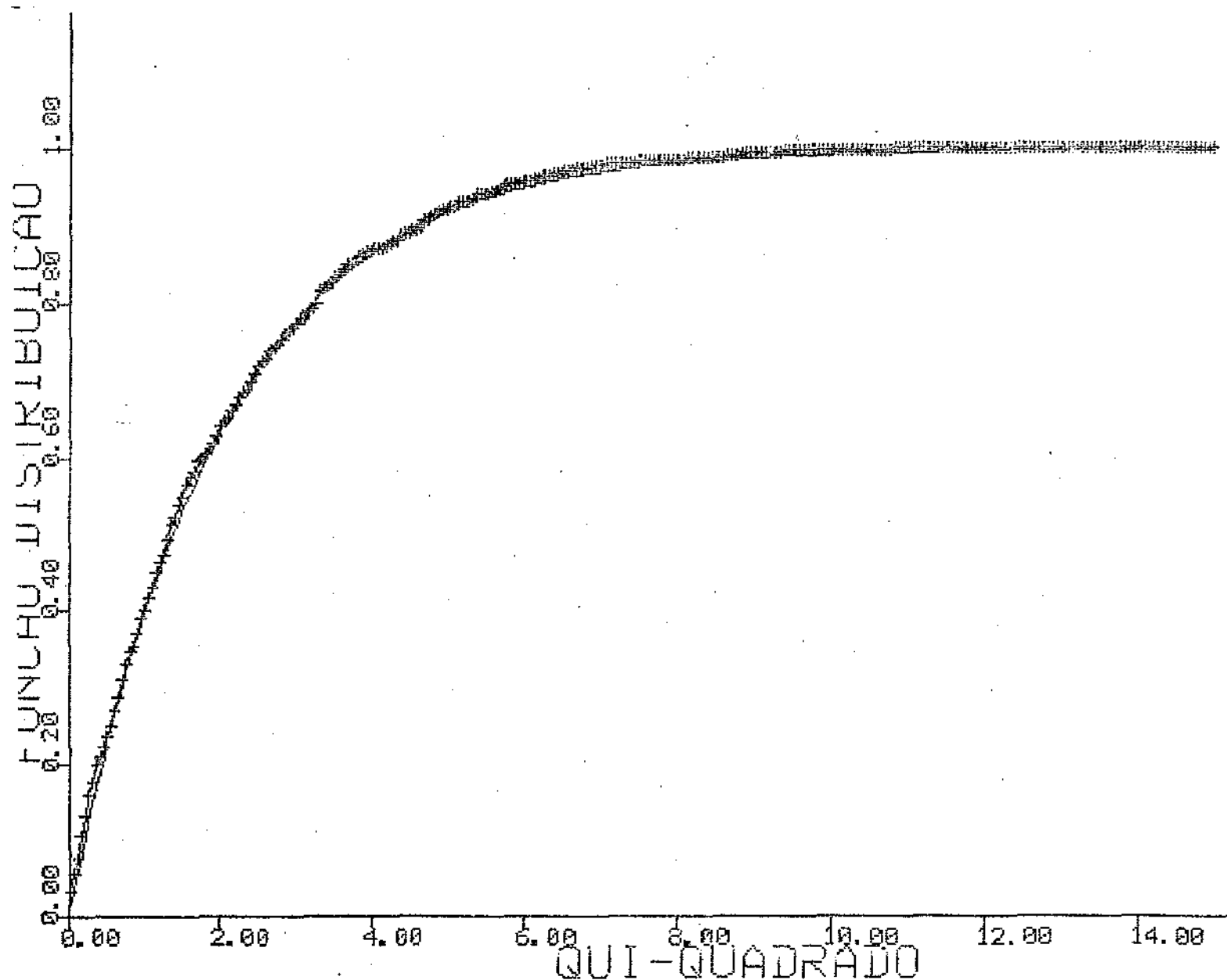


Figura 3.3 Comparação da distribuição exata da Estatística de Pearson corrigida, estimada com base em 1000 repetições M.C. , com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAS-SR com $n = 300$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.3 .



3.3.2 - AMOSTRAGEM ALEATÓRIA ESTRATIFICADA COM ALOCAÇÃO PROPORCIONAL.

Como foi dito na seção 2.2, a Estatística de Pearson sofre pouca distorção neste caso. Nathan (1975), mostrou via simulação, esta pouca perda de poder da Estatística de Pearson para a Estatística de Wald. Em nosso trabalho, também observou-se o mesmo efeito, e uma explicação parcial pode ser a seguinte.

Supondo a população dividida em L estratos.

$S = (S_1, S_2, \dots, S_L)$ uma amostra aleatória estratificada.

S_l — é uma amostra aleatória simples de tamanho n_l extraída com reposição do estrato l , $l=1,2,\dots,L$.

N — tamanho da população i.e. $N = \sum_{l=1}^L N_l$ (3.33)

N_l — tamanho de cada estrato na população.

$$n = \sum_{l=1}^L n_l \quad \text{e} \quad n_l = \sum_{i=1}^L n_{li} \quad (3.34)$$

n_{li} — denota a frequência por cela no estrato l ; $i=1,2,\dots,rc$

W_l — denota a proporção populacional de elementos no estrato l

p_{li} — denota a proporção de elementos do estrato l pertencendo à categoria i .

logo temos que,

$$p_i = \sum_l W_l \cdot p_{li} \quad \text{e} \quad \hat{p}_i = \sum_l W_l \cdot n_{li} / n_l \quad (3.35)$$

e \hat{p}_i reduz para $\hat{p}_i = n_i / n$ sob Alocação Proporcional pois $n_l = n \cdot W_l$.

Então o vetor de média é $\underline{p} = (p_1, p_2, \dots, p_{rc-1})'$ e a matriz de covariância é V/n onde, conforme apontado em Rao e Scott (1981),

$$V = P - \sum_{l=1}^L W_l \cdot (\underline{p}_l - \underline{p})(\underline{p}_l - \underline{p})' = \quad (3.36)$$

$$V = P - R \quad (3.37)$$

onde :

$$\underline{p}_l = (p_{1,1}, p_{1,2}, \dots, p_{1,(rc-1)})' \quad (3.38)$$

$$R = \sum_{l=1}^L W_l \cdot (\underline{p}_l - \underline{p})(\underline{p}_l - \underline{p})'$$

Desta forma a matriz deff é :

$$D_h = (GPG')^{-1} \cdot GVG' = (GPG')^{-1} \cdot [GPG' - GRG'] =$$

$$D_h = I - (GPG')^{-1} \cdot GRG' \quad (3.39)$$

e por 3.12 temos que :

$$0 \leq \underline{c}'(GVG')\underline{c} / \underline{c}'(GPG')\underline{c} = [\underline{c}'(GPG')\underline{c} - \underline{c}'(GRG')\underline{c}] / \underline{c}'(GPG')\underline{c} =$$

$$= 1 - \underline{c}'(GPG')\underline{c} / \underline{c}'(GPG')\underline{c} \leq 1$$

$$\text{ou seja } \phi_0 \leq 1 \text{ para qualquer } \underline{p}_0 \quad \text{e} \quad 0 \leq X_1^2 \leq \sum_{i=1}^b Z_i^2 = \chi_b^2.$$

Assim a Estatística de Pearson X_1^2 é sempre assintoticamente conservativa para Amostragem Aleatória Estratificada.

No caso de $L = 2$ estratos, que é a nossa situação, R pode ser escrita como (Rao e Scott 1981) :

$$R = W_1 . W_2 (p_{\sim 1} - p_{\sim 2}) (p_{\sim 1} - p_{\sim 2})' \quad (3.40)$$

onde :

$$W_2 = 1 - W_1$$

$$p_{\sim 1} = (p_{11}, p_{12}, \dots, p_{1(rc-1)})'$$

$$p_{\sim 2} = (p_{21}, p_{22}, \dots, p_{2(rc-1)})'$$

logo (3.39) fica :

$$D_h = I - (GPG')^{-1} . G . \left[W_1 . W_2 (p_{\sim 1} - p_{\sim 2}) (p_{\sim 1} - p_{\sim 2})' \right] . G' \quad (3.41)$$

assim, o efeito do delineamento vai ser maior ou menor segundo a variabilidade de um estrato para o outro, e também da alocação proporcional.

Em nosso caso, temos que :

$$W_1 = \frac{160}{600} = \frac{4}{15} = 0.267 \quad ; \quad W_2 = \frac{440}{600} = \frac{11}{15} = 0.733$$

$$p_{\sim 1} = \frac{1}{160} (30, 20, 30, 10, 40)'$$

$$= (0.1875, 0.1250, 0.1875, 0.0625, 0.2500)'$$

$$p_{\sim 2} = \frac{1}{440} (10, 80, 30, 70, 160)'$$

$$= (0.0227, 0.1818, 0.068, 0.159, 0.3636)'$$

$$W_{12} \cdot W_{26} = \frac{16}{60} \cdot \frac{44}{60} = 0.1957$$

$$R = \frac{1}{2} \times \begin{bmatrix} 841 & -290 & 609 & -493 & -580 \\ -290 & 100 & -210 & 170 & 200 \\ 609 & -210 & 441 & -357 & -420 \\ -493 & 170 & -357 & 289 & 340 \\ -580 & 200 & -420 & 340 & 400 \end{bmatrix}$$

60×44

agora a matriz G é :

$$G(\underline{p}) = \left[\frac{d h_{ij}(\underline{p})}{d p_{ij}} \right] \quad \begin{matrix} i=1,2,\dots,r-1 \\ j=1,2,\dots,c-1 \end{matrix}$$

onde :

$$h_{ij}(\underline{p}) = p_{ij} - p_{i+} \cdot p_{+j}$$

e :

$$\underline{p} = (p_{11}, p_{12}, \dots, p_{1c}, p_{21}, \dots, p_{2c}, \dots, p_{r1}, p_{r2}, \dots, p_{r(c-1)})'$$

com $r = 2$ e $c = 3$ temos :

$$\underline{h}(\underline{p}) = \begin{bmatrix} h_{11} \\ h_{12} \end{bmatrix} = \begin{bmatrix} p_{11} - p_{1+} \cdot p_{+1} \\ p_{12} - p_{1+} \cdot p_{+2} \end{bmatrix} =$$

e

$$G = \frac{d h(\underline{p})}{d \underline{p}} = \begin{bmatrix} 1-(p_{1+} + p_{+1}) & -p_{+1} & -p_{+1} & -p_{1+} & 0 \\ -p_{+2} & 1-(p_{1+} + p_{+2}) & -p_{+2} & 0 & -p_{1+} \end{bmatrix}$$

a prova de G será dada no Apêndice A2.

Portanto a matriz G fica :

$$G = \frac{1}{60} \begin{bmatrix} 28 & -12 & -12 & -20 & 0 \\ -30 & 10 & -30 & 0 & -20 \end{bmatrix}$$

logo

$$(GRG') = \frac{1}{\begin{matrix} 4 \\ 6.44 \end{matrix}} \begin{bmatrix} \frac{10404}{100} & -\frac{1224}{10} \\ -\frac{1224}{10} & 144 \end{bmatrix}$$

o vetor \underline{p} é :

$$\underline{p} = \frac{1}{600} (40, 100, 60, 80, 200)'$$

e conseqüentemente,

$$P = \frac{1}{60^2} \begin{bmatrix} 800 & -400 & -240 & -320 & -800 \\ -400 & 500 & -60 & -80 & -200 \\ -240 & -60 & 324 & -48 & -120 \\ -320 & -80 & -48 & 416 & -160 \\ -800 & -200 & -120 & -160 & 800 \end{bmatrix}$$

logo :

$$GPG' = \frac{1}{\frac{4}{60}} \begin{bmatrix} 1622016 & -668160 \\ -668160 & 201600 \end{bmatrix}$$

e

$$(GPG')^{-1} = \frac{\frac{4}{6}}{11943936} = \begin{bmatrix} 2201600 & 668160 \\ 668160 & 1622016 \end{bmatrix}$$

então :

$$(GPG')^{-1} \cdot (GRG') = \begin{bmatrix} -0.1157 & 0.136126 \\ -0.2455 & 0.288825 \end{bmatrix}$$

logo :

$$D_h = I - (GPG')^{-1} \cdot (GRG') = \begin{bmatrix} 1.1157 & -0.136126 \\ 0.2455 & 0.711175 \end{bmatrix}$$

e

$$\phi_o = \frac{\text{tr}(D)}{h} = 0.9134$$

Assim vemos que o efeito do Delineamento Populacional é $\phi_o = 0.9134$, i.e. , próximo a 1. Isto explica a pouca distorção, neste exemplo, que sofre a Estatística de Pearson no caso de amostragem aleatória estratificada com alocação proporcional, como foi visto no capítulo 2, apesar da acentuada diferença entre os dois estratos.

O efeito acima é equivalente ao de um delineamento por AAS-SR, com tamanho amostral $n = 52$.

3.3.3 - AMOSTRAGEM ALEATÓRIA POR CONGLOMERADO

Na seção 2.3 observamos que quando os dados são ordenados em conglomerados homogêneos e de mesmo tamanho, este delineamento amostral causa fortes distorções na distribuição da Estatística de Pearson à medida que o tamanho do conglomerado cresce.

Trabalhando com a população hipotética criada, introduzimos agora uma correção proposta por Rao e Scott (1981).

$$X_c^2 = X_l^2 / m_{os} \quad (3.42)$$

onde :

$$m_{os} = \frac{1}{n} \sum_{l=1}^{nca} m_l^2$$

R = número de conglomerados na população

N = tamanho da população

M_t = tamanho de cada conglomerado ; $t = 1, 2, \dots, R$

n = tamanho da amostra

m_l = tamanho da amostra em cada conglomerado ; $l = 1, 2, \dots, nca$

nca = número de conglomerados amostrados

para o nosso caso temos que :

$M_t = M \longrightarrow$ conglomerados de igual tamanho

$m_l = M \longrightarrow$ todas as unidades do conglomerado serão amostradas.

Logo :

$$m_{os} = \frac{1}{n} \sum_{l=1}^{nca} m_l^2 = \frac{1}{n} \sum_{l=1}^{nca} M^2 = \frac{nca \cdot M^2}{n} =$$

mas $n = nca \cdot M$

então :

$$m_{os} = M = \text{tamanho do conglomerado}$$

Os resultados obtidos usando esta correção ($FC = M$) via Monte-Carlo são apresentados nas tabelas e gráficos seguintes.

Tabela 3.2 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4. Esta tabela deve ser comparada com a tabela 2.3 .

Tamanho do conglomerado (TC) = 2

Fator de correção = $m_{os} = TC = 2$

número de conglo. amostrados	$P \left[\sum_{l=1}^2 X_l > x \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
25	0.095	0.054	0.007	2.04	3.69
50	0.102	0.057	0.006	2.08	3.99
75	0.073	0.043	0.007	2.05	3.50
100	0.103	0.054	0.014	2.07	4.33
125	0.098	0.053	0.009	2.05	3.97
150	0.096	0.051	0.090	1.99	3.95

O gráfico é o seguinte.

Figura 3.4 Comparação da distribuição exata da Estatística de Pearson corrigida, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC, com $TC = 2$ e $nca = 100$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.5.

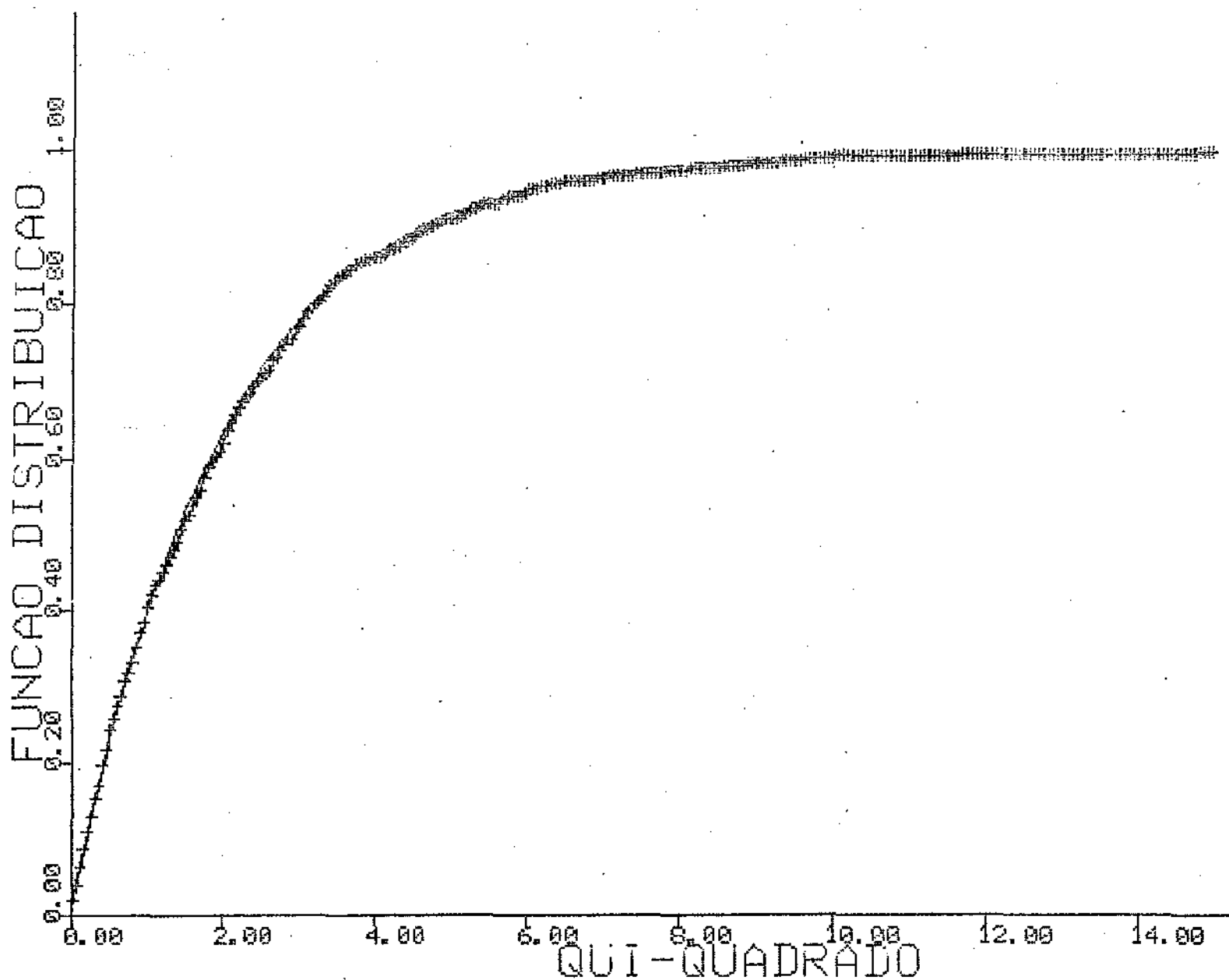


Tabela 3.3 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4. Esta tabela deve ser comparada com a tabela 2.4 .

Tamanho do conglomerado (TC) = 4

Fator de correção = $\frac{m}{os}$ = TC = 4

número de conglo. amostra dos	$P \left[\chi^2_{\frac{2}{I}} > x \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
25	0.079	0.042	0.008	1.92	3.46
40	0.087	0.047	0.006	1.89	3.48
50	0.082	0.038	0.007	1.95	3.74
60	0.097	0.043	0.007	1.93	3.51
75	0.069	0.031	0.003	1.93	3.91

O gráfico é o seguinte.

Figura 3.5 Comparação da distribuição exata da Estatística de Pearson corrigida, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC, com $TC = 4$ e $nca = 50$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.6 .

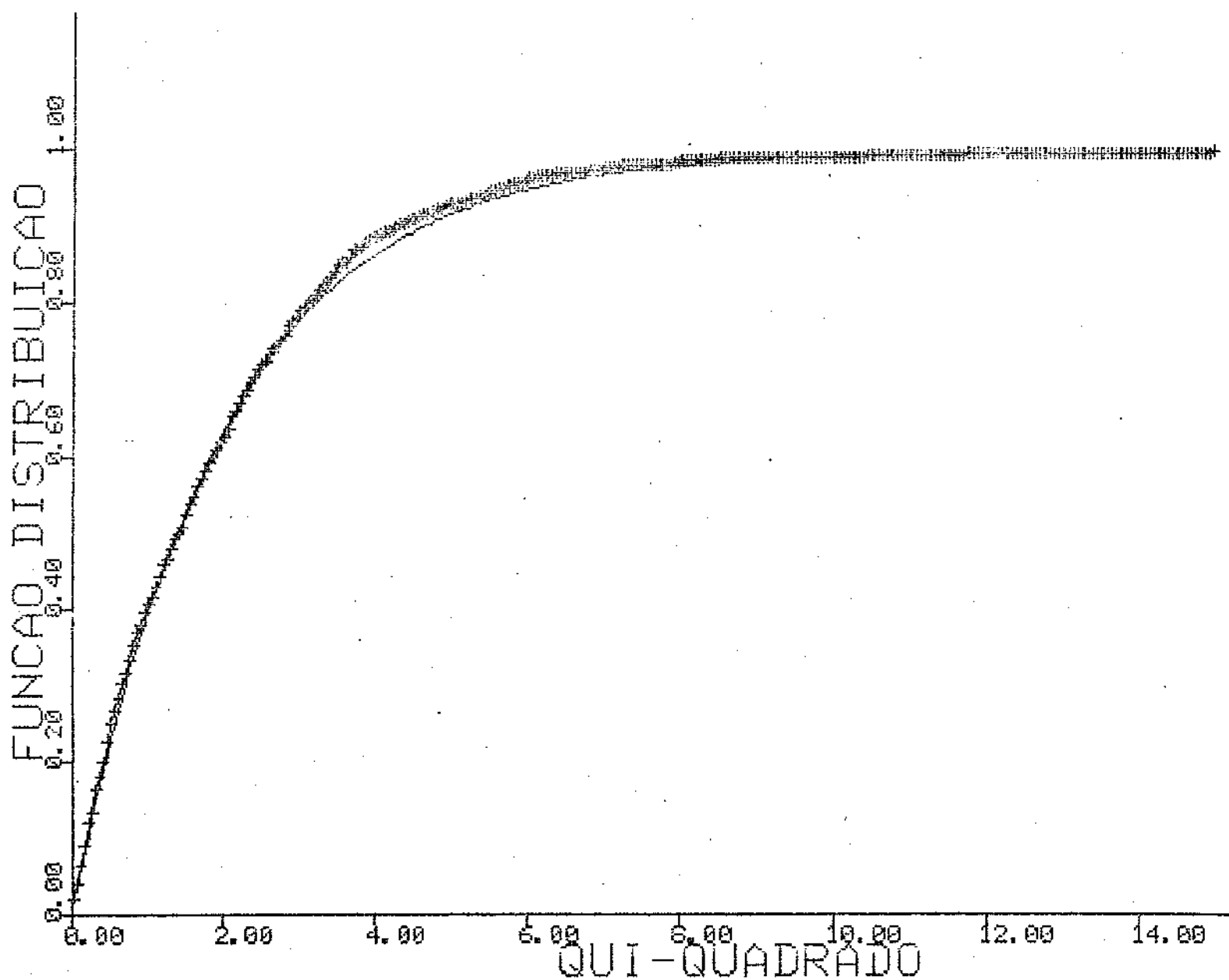


Tabela 3.4 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4. Esta tabela deve ser comparada com a tabela 2.5 .

Tamanho do conglomerado (TC) = 5

Fator de correção = $m \over os$ = TC = 5

número de conglo. amostra dos	$P \left[X \geq x \right]$ $I \quad k$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
10	0.073	0.003	0.007	2.05	2.90
20	0.117	0.057	0.013	2.18	4.22
30	0.103	0.050	0.004	1.99	3.74
40	0.087	0.044	0.007	1.98	3.27
50	0.100	0.038	0.007	2.04	3.79
60	0.076	0.038	0.007	1.90	3.30

O gráfico é o seguinte.

Figura 3.6 Comparação da distribuição exata da Estatística de Pearson corrigida, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC, com $TC = 5$ e $nca = 40$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.7 .

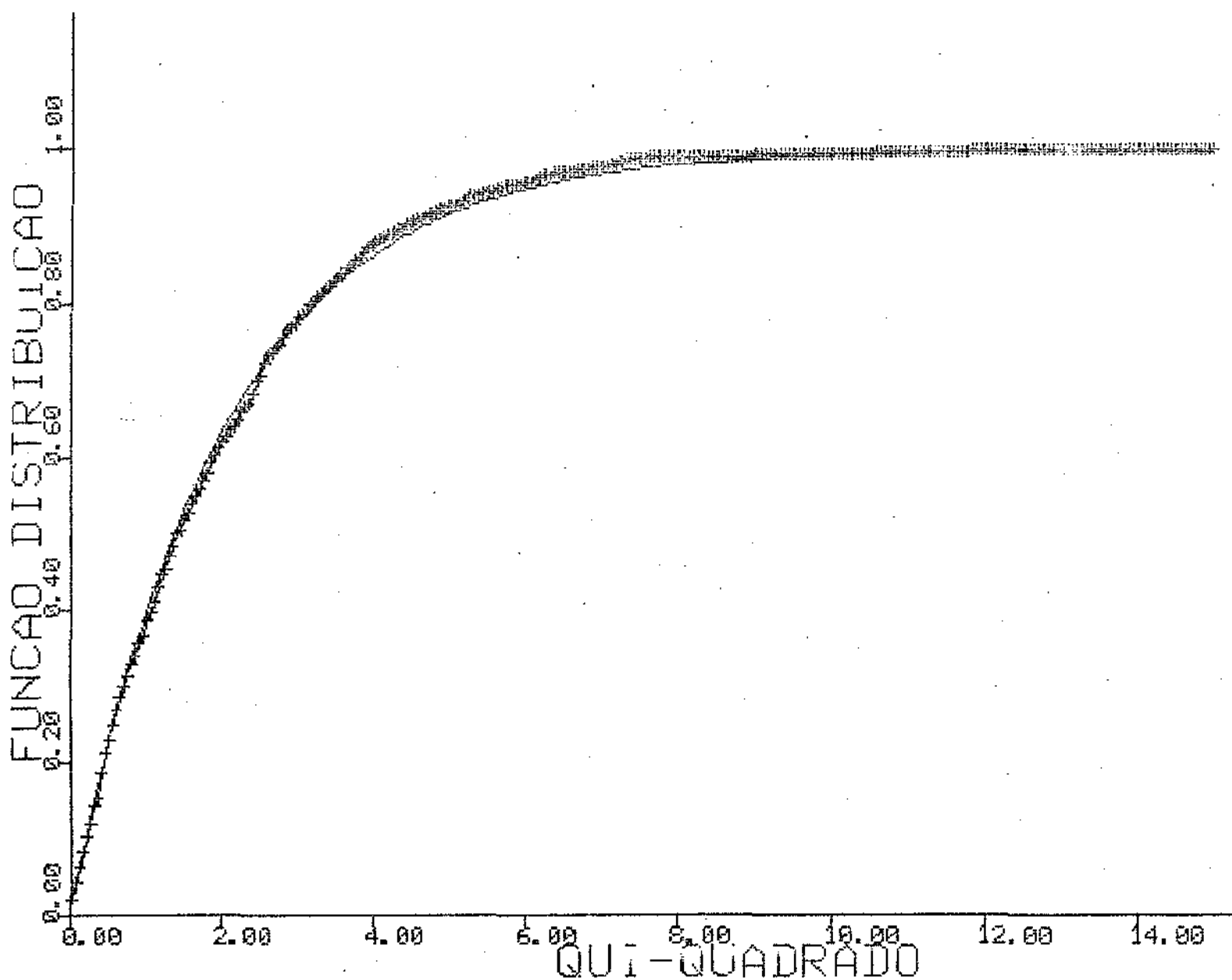


Tabela 3.5 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4. Esta tabela deve ser comparada com a tabela 2.6

Tamanho do conglomerado (TC) = 10

Fator de correção = m = TC = 10
os

número de conglo. amostra dos	$P \left[\frac{X^2}{I} > x \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
10	0.082	0.042	0.009	2.03	3.30
20	0.097	0.047	0.007	2.10	3.67
30	0.085	0.0335	0.004	1.92	3.11

O gráfico é o seguinte.

Figura 3.7 Comparação da distribuição exata da Estatística de Pearson corrigida, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC, com $TC = 10$ e $nca = 10$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.8 .

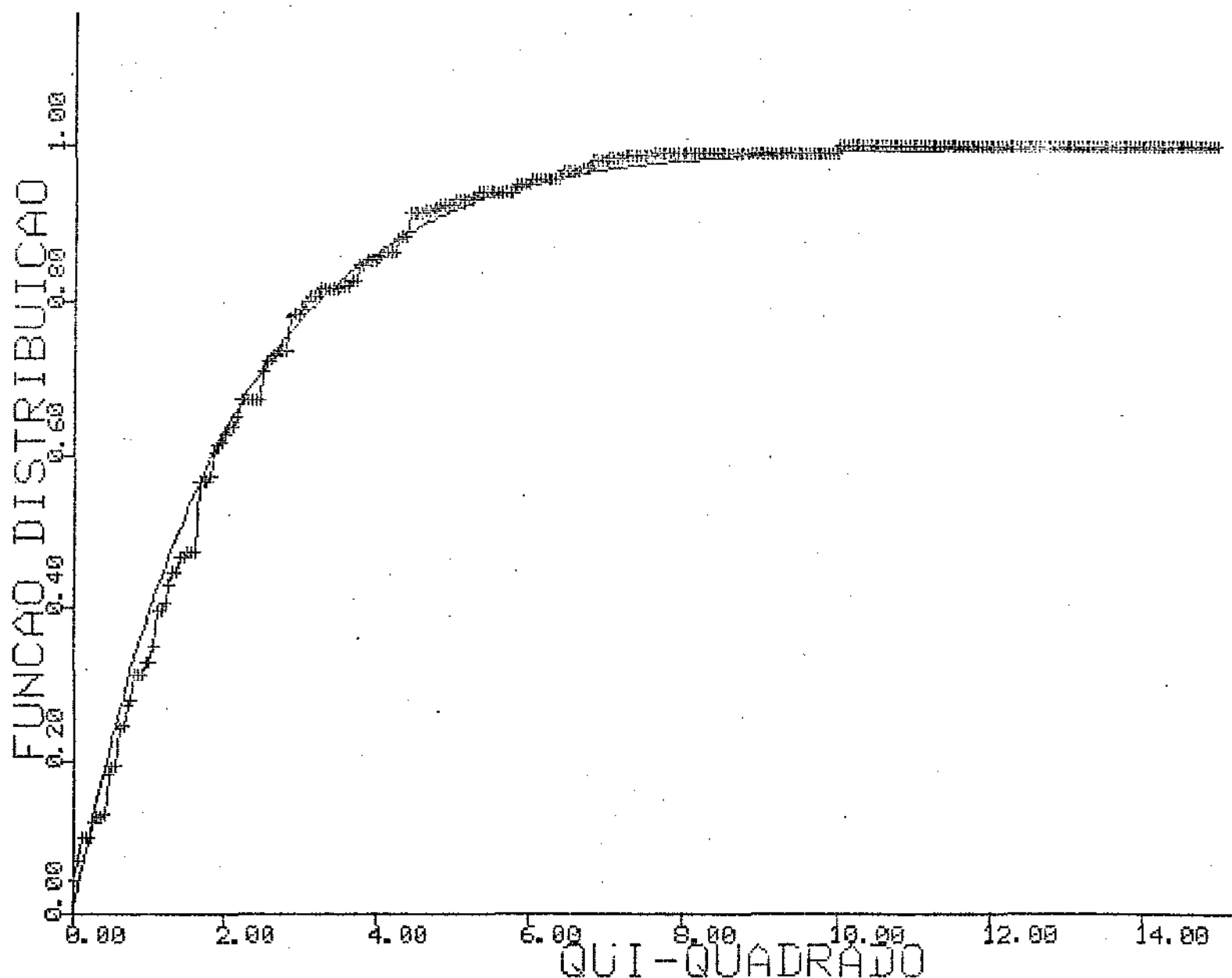


Tabela 3.6 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4. Esta tabela deve ser comparada com a tabela 2.7 .

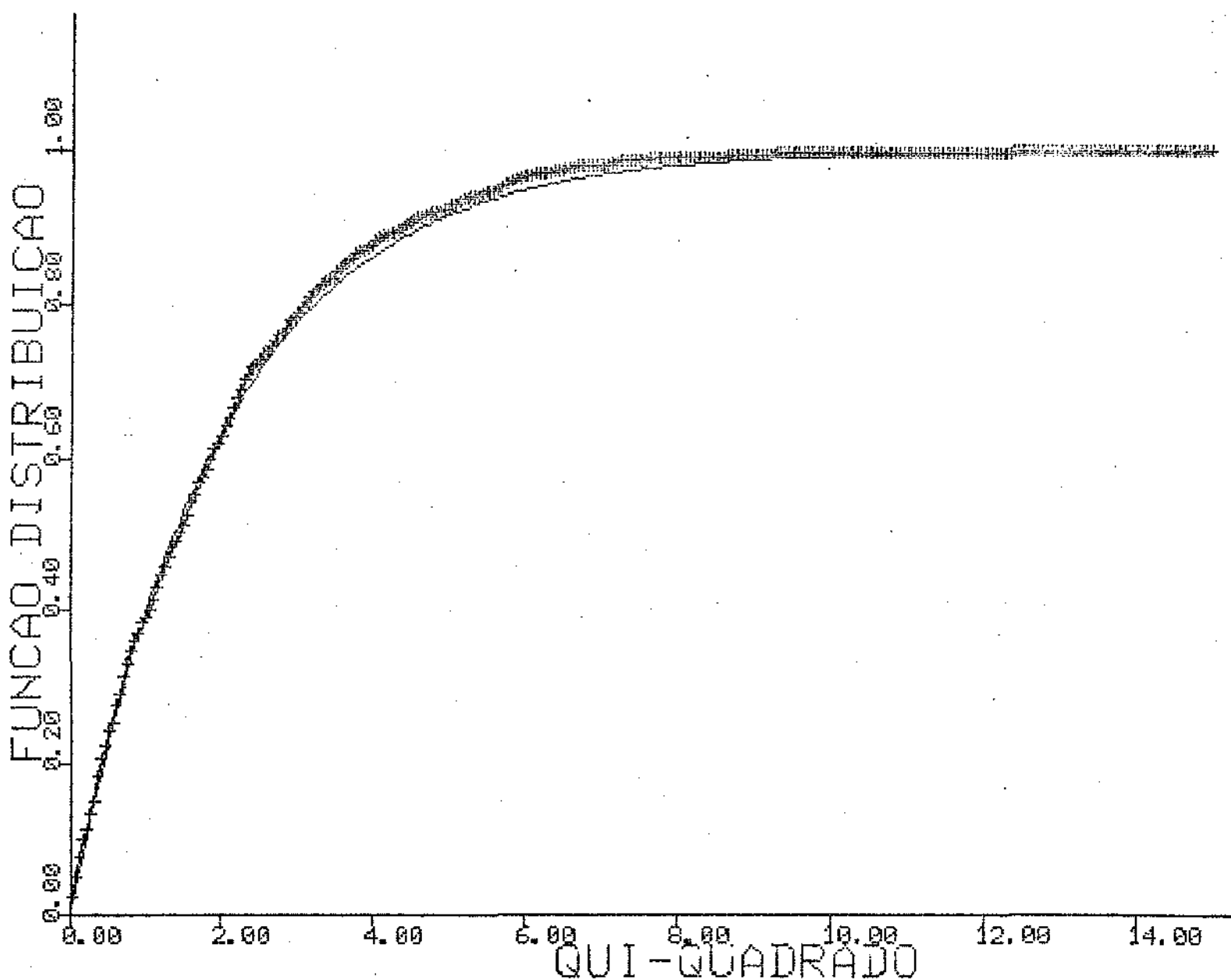
Tamanho do conglomerado (TC) = 20

Fator de correção = $\frac{m}{os}$ = TC = 20

número de conгло. amostra dos	$P \left[\chi^2_{\frac{2}{1}} > \chi^2_k \right]$			ESTIMATIVAS	
	k = 90	k = 95	k = 99	Média	Variância
5	0.190	0.000	0.000	2.51	1.93
10	0.085	0.036	0.010	2.20	3.04
15	0.105	0.050	0.009	2.16	3.55

O gráfico é o seguinte.

Figura 3.8 Comparação da distribuição exata da Estatística de Pearson corrigida, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC, com $TC = 20$ e $nca = 10$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.9 .



3.4 - COMENTÁRIOS E CONCLUSÕES

As tabelas e gráficos apresentados mostram que, no caso de conglomerados homogêneos, a correção parece eliminar as distorções na distribuição da Estatística de Pearson com respeito à distribuição Qui-Quadrado. De fato, é fácil perceber que AAS-CR de nca conglomerados homogêneos equivale a uma AAS-CR de nca indivíduos da população. Os dois delineamentos implicam, obviamente, na mesma distribuição para n , a menos de um fator de escala igual a M , o tamanho do conglomerado. Esta diferença faz com que a Estatística de Pearson, para o primeiro caso, tenha uma distribuição idêntica à do segundo caso, o delineamento neutro, a menos do fator de escala M . A correção proposta iguala assim as duas distribuições. Desta forma, a estatística corrigida não sofre o efeito do delineamento (AAC). Suas possíveis distorções serão devidas unicamente ao tamanho da amostra, nca . Se nca for suficientemente grande, e nós temos evidência de que, neste caso, 15 já é um bom tamanho, as distorções na distribuição condicional da Estatística de Pearson serão pequenas.

Por outro lado, no caso oposto, de conglomerados semelhantes à população, esta correção é obviamente inapropriada pois acentuaria, em vez de atenuar, as distorções.

Estes argumentos serão mais elaborados no próximo capítulo.

CAPÍTULO 4

CORREÇÃO PARA AMOSTRAGEM ALEATÓRIA POR CONGLOMERADOS

Neste capítulo se encontrará uma correção populacional exata para a Estatística de Pearson e também um estimador para esta correção no caso em que o delineamento amostral empregado é Amostragem Aleatória por Conglomerado, com reposição.

Propomos também uma alternativa própria, baseada no método "Bootstrap", para estimar esta correção. Foi feito um estudo comparativo, via simulação, para ver o desempenho de ambas as correções.

4.1 - UMA CORREÇÃO TEÓRICA

Para o caso de AAC, e seguindo o raciocínio da seção 3.2, o problema é determinar a matriz de variância V_{AAC} , devido a este delineamento amostral.

Para isto, definimos,

L = número de conglomerados na população

M = tamanho de cada conglomerado

$N = L.M$ tamanho da população

nca = número de conglomerados amostrados

$n = nca.M$ = tamanho da amostra

$N_{(i,j)1}$ = frequência populacional por cela no conglomerado i

$p_{(i,j)1} = N_{(i,j)1} / N$

com \tilde{N}_1 e \tilde{p}_1 vetores de proporções e de frequências no conglomerado 1, respectivamente.

Logo :

$$p_{1j} = \frac{1}{L} \sum_{l=1}^L p_{(i,j)l} \quad \text{e} \quad \hat{p}_{1j} = \frac{1}{nca.M} \sum_{l=1}^{nca} N_{(i,j)l} \quad (4.1)$$

com
$$\sum_{i=1}^r \sum_{j=1}^c p_{1j} = 1 \quad (4.2)$$

Seja o vetor bidimensional, $Y_{\sim 1k}$ contendo a informação de interesse relativa ao indivíduo $(1,k)$, para $l=1,2, \dots, L$ e $k=1,2, \dots, M$ e o vetor de média

$$\bar{Y}_{\sim 1} = \frac{1}{M} \sum_{k=1}^M Y_{\sim 1k} \quad (4.3)$$

Definimos a variável indicadora

$$I_{(i,j)lk} = \begin{cases} 1 & \text{se no } l\text{-ésimo conglomerado o } k\text{-ésimo} \\ & \text{elemento cai na cela } (i,j). \\ 0 & \text{c.c.} \end{cases} \quad (4.4)$$

e o vetor de indicadores como :

$$I_{\sim 1k} = \left[I_{(1,1)lk}, I_{(1,2)lk}, \dots, I_{(r,c-1)lk} \right] ; \begin{matrix} l=1,2,\dots,L \\ k=1,2,\dots,M \end{matrix} \quad (4.5)$$

com o total por conglomerados,

$$I_{\sim 1} = \sum_{k=1}^M I_{\sim 1k} = N_{\sim 1} \quad (4.6)$$

A média por conglomerados será dada por:

$$\bar{I}_{\sim 1} = \frac{1}{M} \sum_{k=1}^M I_{\sim 1k} = \frac{1}{M} I_{\sim 1} = p_{\sim 1} \quad (4.7)$$

A média geral é :

$$\bar{I}_{\sim} = \frac{1}{L.M} \sum_{l=1}^L \sum_{k=1}^M I_{\sim 1k} = p_{\sim} \quad (4.8)$$

Por outro lado sabemos que:

$$\begin{aligned} SQT &= \sum_{l=1}^L \sum_{k=1}^M (I_{\sim 1k} - \bar{I}_{\sim})(I_{\sim 1k} - \bar{I}_{\sim})' = \\ &= \sum_{l=1}^L \sum_{k=1}^M (I_{\sim 1k} - \bar{I}_{\sim 1})(I_{\sim 1k} - \bar{I}_{\sim 1})' + M \sum_{l=1}^L (\bar{I}_{\sim 1} - \bar{I}_{\sim})(\bar{I}_{\sim 1} - \bar{I}_{\sim})' = \end{aligned}$$

$$SQT = SQDC + M.SQEC \quad (4.9)$$

Encontrando a matriz de covariância para efeitos de comparação de cada delineamento amostral temos os seguintes casos:

1) O caso de Amostragem Aleatória por Conglomerado

Amostramos nca conglomerados com reposição, logo :

$$\hat{\bar{I}}_{\sim} = \frac{1}{nca} \sum_{l=1}^{nca} \bar{I}_{\sim 1} = \frac{1}{nca.M} \sum_{l=1}^{nca} \sum_{k=1}^M I_{\sim 1k} =$$

então

$$\hat{\bar{I}}_{\sim} = \hat{p}_{\sim} \quad (4.10)$$

Assim temos que :

$$E \left(\hat{\bar{I}} \right) = \hat{p} = \left[p_{11}, p_{12}, \dots, p_{rc-1} \right] \quad (4.11)$$

e

$$\begin{aligned} V_{AAC} \left(\hat{\bar{I}} \right) &= V_{AAC} \left(\hat{p} \right) = \frac{1}{nca} V_{AAC} \left(\bar{I}_{lk} \right) = \\ &= \frac{1}{nca} \cdot \frac{1}{L} \sum_{l=1}^L \left(\bar{I}_{\sim l} - \bar{I}_{\sim} \right) \left(\bar{I}_{\sim l} - \bar{I}_{\sim} \right)' = \\ V_{AAC} \left(\hat{\bar{I}} \right) &= \frac{1}{nca \cdot L} SQEC \end{aligned} \quad (4.12)$$

ii) O caso de Amostragem Aleatória Simples com reposição

Amostramos $n = nca \cdot M$ unidades com reposição, então :

$$\hat{\bar{I}}_{\sim} = \frac{1}{n} \sum_{l=1}^{nca} \sum_{k=1}^M \bar{I}_{\sim lk} = \frac{1}{nca \cdot M} \sum_{l=1}^{nca} \sum_{k=1}^M \bar{I}_{\sim lk}$$

com

$$E \left(\hat{\bar{I}}_{\sim} \right) = E \left(\hat{p}_{\sim} \right) = \hat{p}_{\sim} \quad (4.13)$$

e

$$\begin{aligned} V_{AAS-CR} \left(\hat{\bar{I}}_{\sim} \right) &= V_{AAS-CR} \left(\hat{p}_{\sim} \right) = \frac{1}{n} V_{AAS-CR} \left(\bar{I}_{\sim lk} \right) \\ V_{AAS-CR} \left(\hat{\bar{I}}_{\sim} \right) &= \frac{1}{nca \cdot M} \cdot \frac{SQT}{L \cdot M} \end{aligned} \quad (4.14)$$

mas de 4.9 vem que, $SQT = SQDC + M.SQEC$

$$SQEC = \frac{1}{M} (SQT - SQDC)$$

Portanto de 4.15 e 4.17, temos que

$$\begin{aligned} V_{AAC}(\bar{I}) &= \frac{1}{nca.L} SQEC = \frac{1}{nca.L.M} (SQT - SQDC) = \\ &= \frac{1}{nca.L.M} SQT (I - SQT^{-1}.SQDC) \end{aligned}$$

com I matriz identidade de ordem rc-1, logo

$$V_{AAC}(\bar{I}) = M.V_{AAS-CR}(\bar{I}). \left[I - SQT^{-1}.SQDC \right] \quad (4.15)$$

Introduzimos agora uma informação adicional. Sabe-se que :

$$SQDC = \sum_{l=1}^L \sum_{k=1}^M (I_{\sim lk} - \bar{I}_{\sim l})(I_{\sim lk} - \bar{I}_{\sim l})'$$

e por 4.7 temos que :

$$\begin{aligned} \sum_{k=1}^M (I_{\sim lk} - \bar{I}_{\sim l})(I_{\sim lk} - \bar{I}_{\sim l})' &= \sum_{k=1}^M I_{\sim lk}.I_{\sim lk}' - M.\bar{I}_{\sim l}.\bar{I}_{\sim l}' = \\ &= \text{diag}(N_{\sim l}) - M.p_{\sim l}.p_{\sim l}' \end{aligned}$$

$$\sum_{k=1}^M (I_{\sim lk} - \bar{I}_{\sim l})(I_{\sim lk} - \bar{I}_{\sim l})' = M. \left[\text{diag}(p_{\sim l}) - p_{\sim l}.p_{\sim l}' \right] \quad (4.16)$$

então :

$$SQDC = M. \sum_{l=1}^L \left[\text{diag}(p_{\sim l}) - p_{\sim l}.p_{\sim l}' \right] \quad (4.17)$$

E também :

$$SQEC = \sum_{l=1}^L (\bar{\underline{I}}_{\sim l} - \bar{\underline{I}})(\bar{\underline{I}}_{\sim l} - \bar{\underline{I}})' = \sum_{l=1}^L \bar{\underline{I}}_{\sim l} \cdot \bar{\underline{I}}_{\sim l}' - L \cdot \bar{\underline{I}} \cdot \bar{\underline{I}}'$$

$$SQEC = \sum_{l=1}^L \underline{p}_{\sim l} \cdot \underline{p}_{\sim l}' - L \cdot \underline{p} \cdot \underline{p}' \quad (4.18)$$

Portanto, de 4.17 e 4.18 segue-se que,

$$SQT = SQDC + M \cdot SQEC =$$

$$= M \cdot L \cdot \left[\text{diag}(\underline{p}) - \underline{p} \cdot \underline{p}' \right]$$

$$SQT = M \cdot L \cdot P \quad (4.19)$$

com

$P = \text{diag}(\underline{p}) - \underline{p} \cdot \underline{p}' =$ matriz de covariância multinomial

e de forma análoga temos que:

$$P_l = \text{diag}(\underline{p}_{\sim l}) - \underline{p}_{\sim l} \cdot \underline{p}_{\sim l}' \quad \text{para } l=1, 2, \dots, L$$

$P_l =$ matriz de covariância dentro de cada conglomerado.

Assim 4.18 fica :

$$\begin{aligned} V_{AAC}(\hat{\underline{I}}_{\sim}) &= M \cdot V_{AAS-CR}(\hat{\underline{p}}_{\sim}) \cdot \left[I - SQT^{-1} \cdot SQDC \right] = \\ &= M \cdot V_{AAS-CR}(\hat{\underline{p}}_{\sim}) \cdot \left[I - \left[M \cdot L \cdot P \right]^{-1} \cdot \left[M \cdot \sum_{l=1}^L P_l \right] \right] \end{aligned}$$

$$V_{AAC}(\hat{\bar{I}}) = M \cdot V_{AAS-CR}(\hat{p}) \cdot \left[I - \left[L \cdot P \right]^{-1} \cdot \left[\frac{L}{\sum_{l=1}^L} P_l \right] \right] \quad (4.20)$$

Mas,

$$V_{AAS-CR}(\hat{p}) = \frac{1}{nca \cdot M} \cdot \frac{1}{L \cdot M} \cdot SQT$$

$$V_{AAS-CR}(\hat{p}) = \frac{1}{nca \cdot M} \cdot P \quad (4.21)$$

Portanto :

$$V_{AAC}(\hat{p}) = \frac{1}{nca} \cdot P \cdot \left[I - \frac{1}{L} \cdot P^{-1} \cdot \sum_{l=1}^L P_l \right] \quad (4.22)$$

Note-se que a matriz P é calculada a partir dos dados da população, pois cada conglomerado entra integralmente na amostra.

4.1.1 - ESTIMADOR DA VARIANCIA

O estimador de máxima verossimilhança para a Variância é :

$$\hat{V}_{AAC}(\hat{p}) = \frac{1}{nca} \cdot \hat{P} \cdot \left[I - \frac{1}{nca} \cdot \hat{P}^{-1} \cdot \sum_{l=1}^{nca} P_l \right] \quad (4.23)$$

com :

nca = número de conglomerados amostrados

$$\hat{P} = \text{diag}(\hat{p}) - \hat{p} \cdot \hat{p}'$$

$$P_1 = \text{diag} (\hat{p}_1) - \hat{p}_1 \cdot \hat{p}_1'$$

$$\hat{p} = (\hat{p}_{11}, \hat{p}_{12}, \dots, \hat{p}_{rc-1})'$$

$$\hat{p}_1 = (\hat{p}_{(1,1)1}, \hat{p}_{(1,2)1}, \dots, \hat{p}_{(r,c-1)1})'$$

pois o estimador de máxima verossimilhança de p_{ij} é :

$$\hat{p}_{ij} = \frac{1}{n} \cdot n_{ij} \quad \text{para } i=1,2,\dots,r \\ j=1,2,\dots,c$$

e da matriz $V_{\text{AAS-CR}} (\hat{p})$ é ,

$$V_{\text{AAS-CR}} (\hat{p}) = \frac{1}{nca.M} \cdot P \quad (4.24)$$

4.2 - MATRIZ DE EFEITO DE DELINEAMENTO

Desde 4.22 a matriz D_h fica então :

$$D_h = (GPG_h')^{-1} \cdot G_h' G_h' = \quad (4.25)$$

$$D_h = (GPG_h')^{-1} \left[G_h' \left[M.P \left[I - \frac{1}{L} \cdot P^{-1} \left[\sum_{l=1}^L P_l \right] \right] \right] \cdot G_h' \right] \quad (4.26)$$

com

$$G_h = \frac{d h(p)}{d p} = \begin{cases} \frac{d h_{ij}(p)}{d p_{ab}} = 1 - (p_{i+} + p_{+j}) & \text{se } a=i, b=j \\ \frac{d h_{ij}(p)}{d p_{ab}} = -p_{+j} & \text{se } a=i, b \neq j \\ \frac{d h_{ij}(p)}{d p_{ab}} = -p_{i+} & \text{se } a \neq i, b=j \\ \frac{d h_{ij}(p)}{d p_{ab}} = 0 & \text{se } a \neq i, b \neq j \end{cases} \quad (4.27)$$

A prova de G_h será dada no Apêndice A2

Consequentemente , \hat{D}_h é dada por :

$$\hat{D}_h = (\hat{G} \cdot \hat{P} \cdot \hat{G}')^{-1} \cdot \hat{G} \cdot \hat{M} \cdot \hat{P} \cdot \left[I - \frac{1}{nca} \cdot \hat{P}^{-1} \cdot \left(\sum_{l=1}^{nca} \hat{P}_l \right) \right] \cdot \hat{G}' \quad (4.28)$$

que é o estimador de D_h sob Amostragem Aleatória por Conglomerado, com reposição de conglomerados, entrando o conglomerado selecionado integralmente na amostra, onde :

\hat{G} a mesma que G com \hat{p} em vez de p .

O fator de correção ϕ_o é :

$$\phi_o = \frac{1}{b} \text{tr}(\hat{D}_h) = \frac{1}{b} \sum_{i=1}^b \phi_{oi} = \frac{1}{b} \sum_{s=u}^d d_{hsu} \quad (4.29)$$

onde :

d_{hsu} = são os elementos da diagonal de D_h

$b = (r-1)(c-1)$

e portanto

$$\frac{\sum_{c=1}^2 X^2}{c} = \frac{\sum_{i=1}^2 X^2}{\phi_o} \quad (4.30)$$

No caso de se ter conglomerados homogêneos, a matriz P_1 é igual à matriz nula, ficando portanto 4.22 como,

$$V_{AAC}(\hat{\tilde{p}}) = M \cdot V_{AAS-CR}(\hat{\tilde{p}})$$

e assim a matriz,

$$D_h = M.I \quad (4.31)$$

e consequentemente $\phi_o = M$. Assim, este fator de correção é exato para a configuração de conglomerados homogêneos.

O estimador do Fator de Correção será :

$$\hat{\phi}_h = \text{tr}(\hat{D}_h) = \frac{1}{b} \sum_{i=1}^b \hat{\phi}_i = \frac{1}{b} \sum_{s=u}^b \hat{d}_{hsu} \quad (4.32)$$

onde \hat{d}_{hsu} são os elementos da diagonal da Matriz \hat{D}_h

e portanto

$$\frac{X^2}{c} = \frac{X}{\hat{\phi}_h} \quad (4.33)$$

4.3 - ESTIMADOR DE V_h PELO MÉTODO DE REAMOSTRAGEM "BOOTSTRAP"

Como foi visto na seção 4.1, devemos encontrar a matriz de covariância V_h do vetor $\tilde{h}(p)$ devido ao delineamento amostral empregado. Um estimador desta matriz, a partir de uma amostra, foi encontrada por máxima verossimilhança utilizando-se (4.27). Porém existem outros métodos, não paramétricos, para estimar esta matriz de covariância. Exploraremos em particular o método de reamostragem "Bootstrap", desenvolvendo uma adaptação do mesmo ao problema da estimação da matriz de covariância.

4.3.1 - ESTIMANDO V_h POR REAMOSTRAGEM "BOOTSTRAP"

O método de "Bootstrap" (Efron 1979) será aqui utilizado para se estimar V_h a partir dos dados amostrais.

O plano de reamostragem, para este caso foi o seguinte :

- i) Extraí-se uma amostra de nca conglomerados, com reposição de conglomerados.
- ii) Reamostra-se nca conglomerados com reposição desde a amostra de nca conglomerados, MRB vezes, construindo assim MRB amostras "Bootstrap".
- iii) Para cada reamostra "Bootstrap" calculamos o vetor $\hat{h}(p)$, da hipótese de interesse.
- iv) Calculamos a matriz de covariância para $\hat{h}(p)$.

4.4 - RESULTADOS VIA SIMULAÇÃO PARA X_c^2

Para todos os casos simulados, o número de repetições de Monte-Carlo e também de reamostras "Bootstrap" foi de 1000 vezes.

A simulação foi feita com os dados criados na tabela 1.1, os quais foram divididos em conglomerados de tamanho M. Os conglomerados foram gerados de maneira tal que a estrutura dentro de cada conglomerado seja de duas formas :

- 1) O conglomerado é homogêneo.
- 2) O conglomerado é semelhante à população.

Adotamos aqui a seguinte notação :

$$\chi^2_I = \text{Qui-Quadrado de Pearson.}$$

FCT = fator de correção teórica.

FCA = fator de correção amostral estimada por máxima verossimilhança.

FCB = fator de correção estimada por Bootstrap.

$$\chi^2_{CT} = \text{Qui-Quadrado de Pearson corrigida pelo FCT}$$

$$\chi^2_{CA} = \text{Qui-Quadrado de Pearson corrigida pelo FCA}$$

$$\chi^2_{CB} = \text{Qui-Quadrado de Pearson corrigida pelo FCB}$$

$$\chi^2_W = \text{Qui-Quadrado de Wald}$$

\overline{FCA} , \overline{FCB} são as médias aritméticas de FCA e FCB respectivamente.

4.4.1 - RESULTADOS PARA CONGLOMERADOS HOMOGÊNEOS

Trabalhando com a população hipotética descrita anteriormente, na qual tinha-se uma perfeita homogeneidade dentro de cada conglomerado, se fez uma ligeira alteração dentro de cada conglomerado, obtendo-se assim, conglomerados "quase homogêneos", no sentido de se ter num mesmo conglomerado (de tamanho menor ou igual a 20), indivíduos pertencendo, no máximo, a duas celas diferentes (ver Apêndice A1).

Nesta nova configuração dos conglomerados, introduzimos agora a correção proposta na seção 4.3 . Estimando-se o fator de correção pelos dois métodos: por máxima verossimilhança e por "Bootstrap".

Também calculamos o fator de correção populacional nesta nova configuração, apenas para efeito de comparação, pois na prática isto não ocorre. Ainda calculamos a estatística de Wald.

Consideramos os mesmos tamanhos de conglomerados da seção 2.3.1 .

Os resultados obtidos usando estas correções via Monte-Carlo são apresentados nas tabelas e gráficos seguintes.

Tabela 4.1 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do conglomerado (T.C.) = 3

Fator de correção teórico = 2.90

nca		$P [X^2 > x_k^2]$					ESTIMATIVAS	
		X_I^2	X_{CT}^2	X_{CA}^2	X_W^2	X_{CB}^2	\overline{FCA}	\overline{FCB}
15	k = 90	0.507	0.102	0.104	0.185	0.116	2.86	2.78
	k = 95	0.400	0.040	0.046	0.097	0.054		
	Média	6.23	2.14	2.17	2.88	2.23		
	Varian.	21.82	3.34	3.37	7.63	3.51		
35	k = 90	0.471	0.114	0.115	0.165	0.123	2.89	2.82
	k = 95	0.376	0.06	0.058	0.101	0.065		
	Média	6.12	2.11	2.12	2.44	2.17		
	Varian.	35.25	4.18	4.23	6.31	4.45		
50	k = 90	0.476	0.104	0.105	0.138	0.105	2.89	2.84
	k = 95	0.377	0.051	0.050	0.089	0.053		
	Média	6.06	2.09	2.09	2.40	2.13		
	Varian.	35.30	4.18	4.18	7.06	4.35		
65	k = 90	0.049	0.092	0.094	0.113	0.093	2.89	2.85
	k = 95	0.348	0.043	0.042	0.065	0.046		
	Média	5.58	1.92	1.93	2.10	1.95		
	Varian.	27.98	3.31	3.35	4.79	3.44		
85	k = 90	0.046	0.113	0.114	0.121	0.114	2.90	2.86
	k = 95	0.359	0.063	0.063	0.076	0.066		
	Média	5.98	2.06	2.06	2.22	2.08		
	Varian.	34.56	4.10	4.13	5.54	4.17		
100	k = 90	0.482	0.089	0.087	0.104	0.089	2.89	2.87
	k = 95	0.376	0.050	0.047	0.055	0.050		
	Média	5.89	2.03	2.03	2.14	2.04		
	Varian.	32.13	3.81	3.80	4.95	3.86		

Os gráficos são os seguintes :

Figura 4.1 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCA, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade para AAC, com T.C. = 3 e nca = 100, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

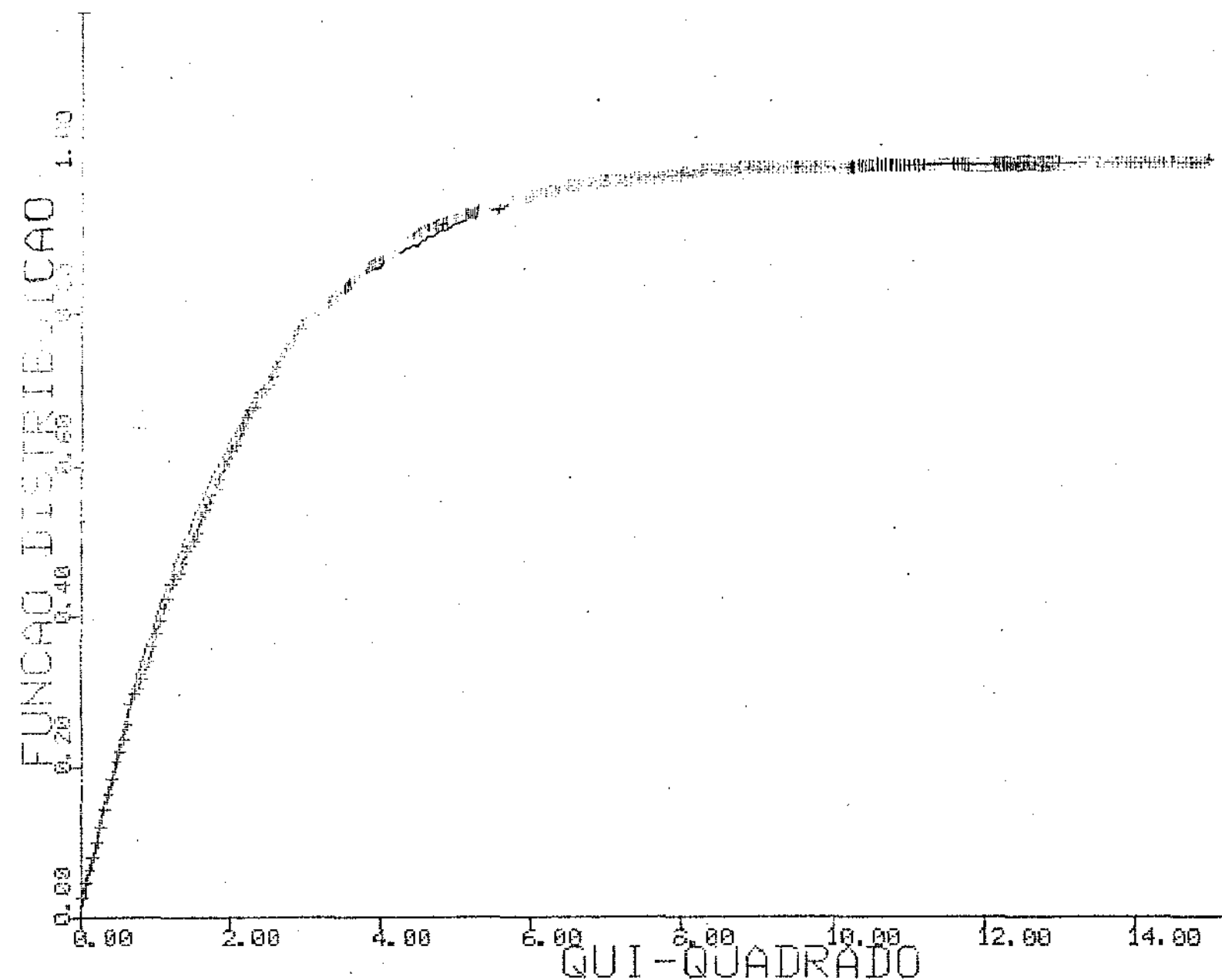


Figura 4.2 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCB, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade para AAC, com T.C. = 3 e nca = 100, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

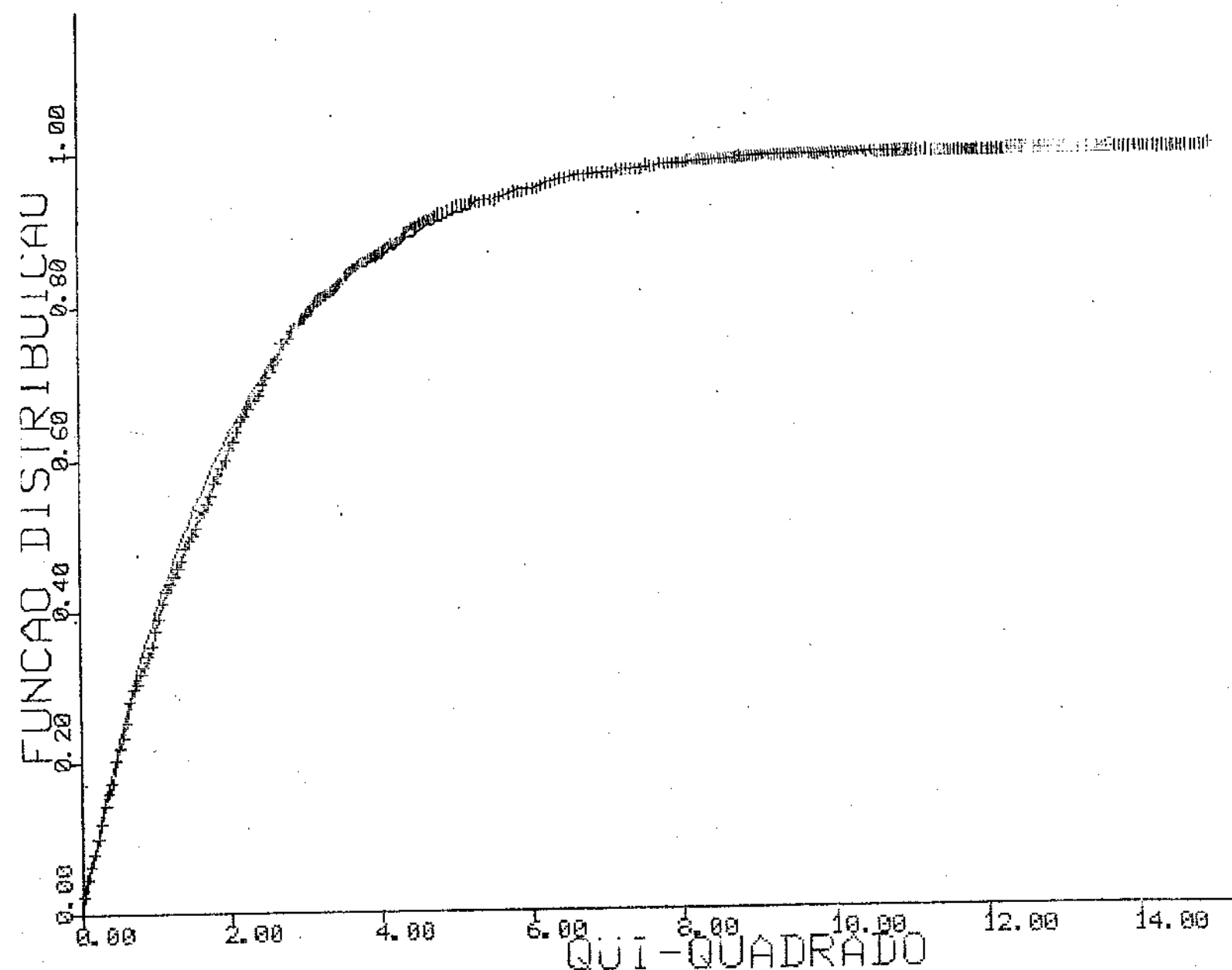


Tabela 4.2 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do conglomerado (T.C.) = 5

Fator de correção teórico = 5.00

nca		$P \left[\overset{2}{X} > \overset{2}{x} \right]$ k					ESTIMATIVAS	
		$\overset{2}{X}$ I	$\overset{2}{X}$ CT	$\overset{2}{X}$ CA	$\overset{2}{X}$ W	$\overset{2}{X}$ CB	\overline{FCA}	\overline{FCB}
10	k = 90	0.726	0.100	0.100	0.151	0.111	5.00	4.95
	k = 95	0.638	0.046	0.046	0.109	0.048		
	Média	11.04	2.20	2.20	2.99	2.23		
	Varian.	85.29	3.41	3.41	12.52	3.24		
20	k = 90	0.660	0.100	0.100	0.189	0.107	5.00	4.84
	k = 95	0.588	0.038	0.038	0.121	0.043		
	Média	10.22	2.04	2.04	2.76	2.10		
	Varian.	82.96	3.31	3.31	8.45	3.46		
30	k = 90	0.679	0.098	0.098	0.189	0.107	5.00	4.85
	k = 95	0.601	0.047	0.047	0.109	0.048		
	Média	10.51	2.10	2.10	2.63	2.15		
	Varian.	89.85	3.59	3.59	7.25	3.75		
40	k = 90	0.632	0.097	0.097	0.149	0.109	5.00	4.88
	k = 95	0.545	0.052	0.052	0.098	0.055		
	Média	9.86	1.97	1.97	2.33	2.01		
	Varian.	97.65	3.90	3.90	6.95	4.02		
50	k = 90	0.626	0.094	0.094	0.137	0.098	5.00	4.90
	k = 95	0.553	0.046	0.046	0.081	0.047		
	Média	10.10	2.02	2.02	2.28	2.05		
	Varian.	104.22	4.16	4.16	6.35	4.38		
60	k = 90	0.651	0.084	0.084	0.118	0.090	5.00	4.91
	k = 95	0.565	0.051	0.051	0.070	0.052		
	Média	10.15	2.03	2.03	2.24	2.07		
	Varian.	104.92	4.19	4.19	5.82	4.38		

Os gráficos são os seguintes :

Figura 4.3 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCA, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade para AAC, com T.C. = 5 e nca = 50, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

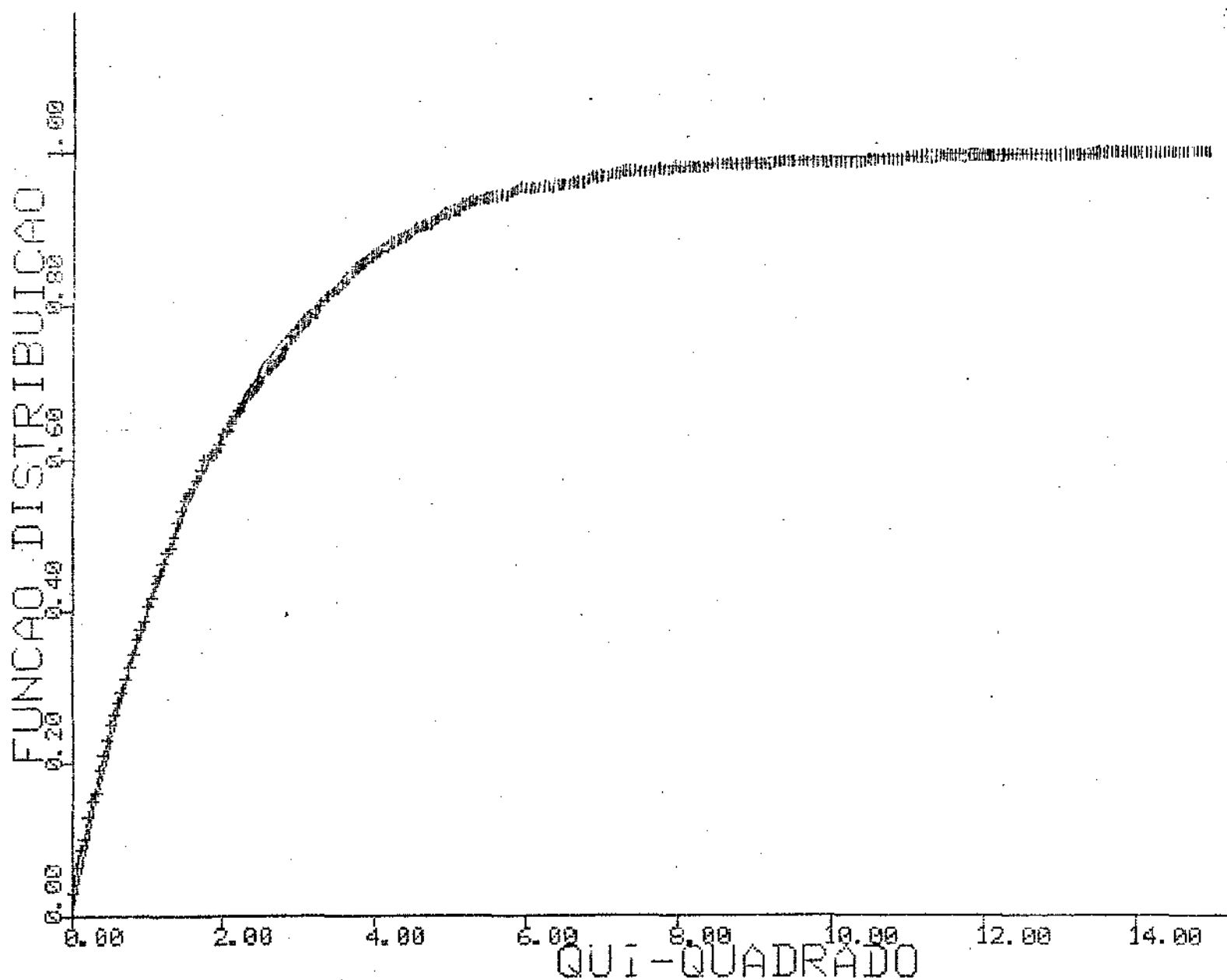


Figura 4.4 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCB, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade para AAC, com $T.C. = 5$ e $nca = 50$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

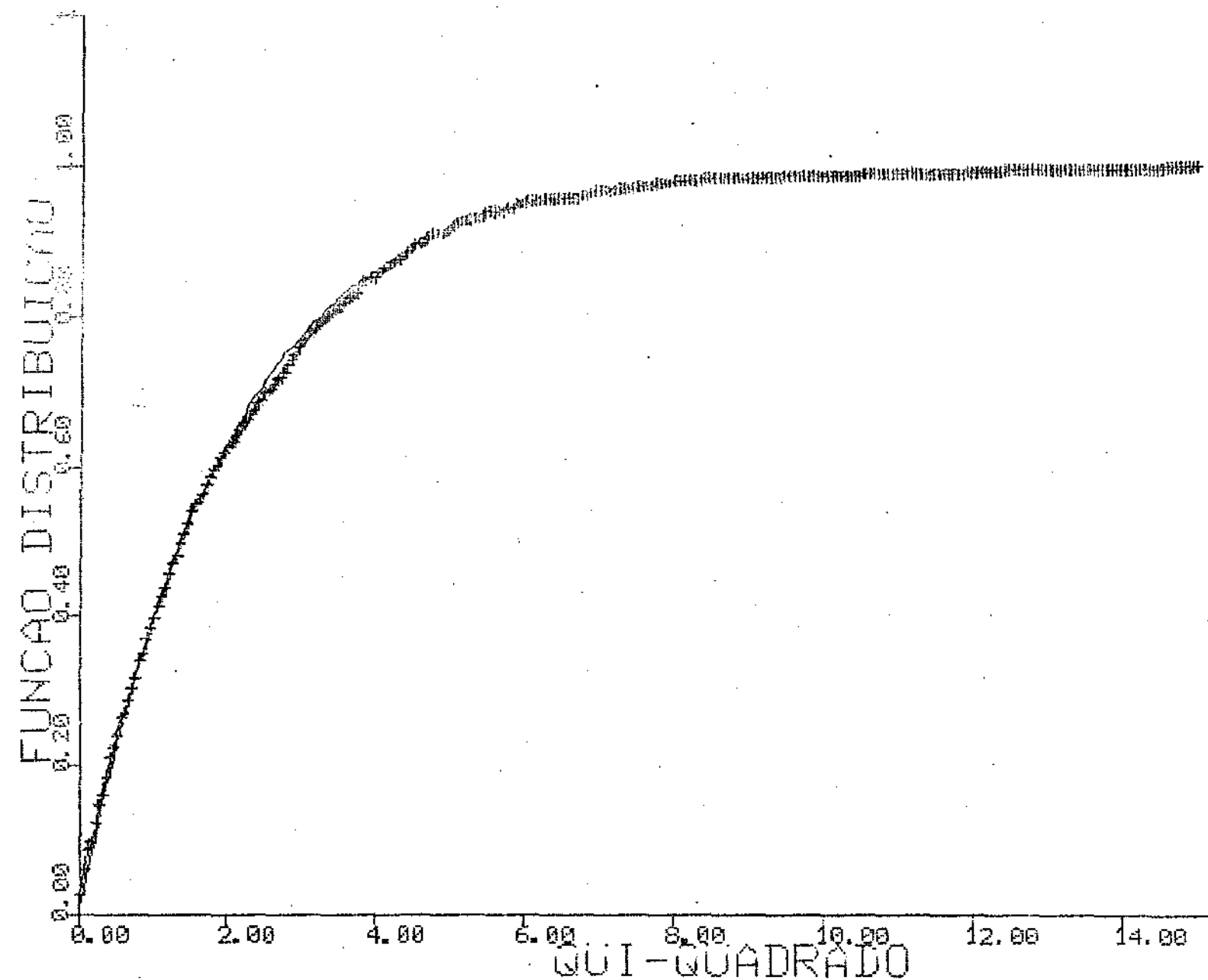


Tabela 4.3 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do conglomerado (T.C.) = 6

Fator de correção teórico = 5.61

nca		$P [X^2 > x_k^2]$					ESTIMATIVAS	
		X^2_I	X^2_{CT}	X^2_{CA}	X^2_W	X^2_{CB}	\overline{FCA}	\overline{FCB}
10	k = 90	0.757	0.070	0.098	0.168	0.102	5.42	5.56
	k = 95	0.709	0.042	0.043	0.097	0.054		
	Média	12.15	2.16	2.25	3.00	2.27		
	Varian.	90.28	2.86	3.14	8.78	3.10		
20	k = 90	0.718	0.101	0.105	0.164	0.110	5.50	5.32
	k = 95	0.657	0.037	0.047	0.098	0.053		
	Média	11.86	2.11	2.16	2.67	2.22		
	Varian.	110.24	3.49	3.76	6.33	3.90		
25	k = 90	0.697	0.092	0.097	0.153	0.102	5.51	5.35
	k = 95	0.638	0.044	0.049	0.086	0.049		
	Média	11.52	2.05	2.09	2.47	2.15		
	Varian.	114.29	3.62	3.91	5.82	4.08		
35	k = 90	0.680	0.093	0.090	0.150	0.098	5.55	5.41
	k = 95	0.605	0.043	0.040	0.084	0.050		
	Média	11.22	2.00	2.03	2.35	2.07		
	Varian.	124.55	3.95	4.07	6.59	4.19		
45	x = 90	0.659	0.105	0.106	0.157	0.115	5.57	5.46
	x = 95	0.591	0.049	0.049	0.094	0.052		
	Média	11.28	2.00	2.02	2.32	2.06		
	Varian.	124.37	3.94	3.99	6.35	4.12		
50	x = 90	0.675	0.096	0.100	0.137	0.109	5.57	5.47
	x = 95	0.607	0.051	0.054	0.087	0.057		
	Média	11.66	2.07	2.09	2.36	2.13		
	Varian.	132.39	4.20	4.27	6.73	4.40		

Os gráficos são os seguintes :

Figura 4.5 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCA, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade para AAC, com T.C. = 6 e nca = 25, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

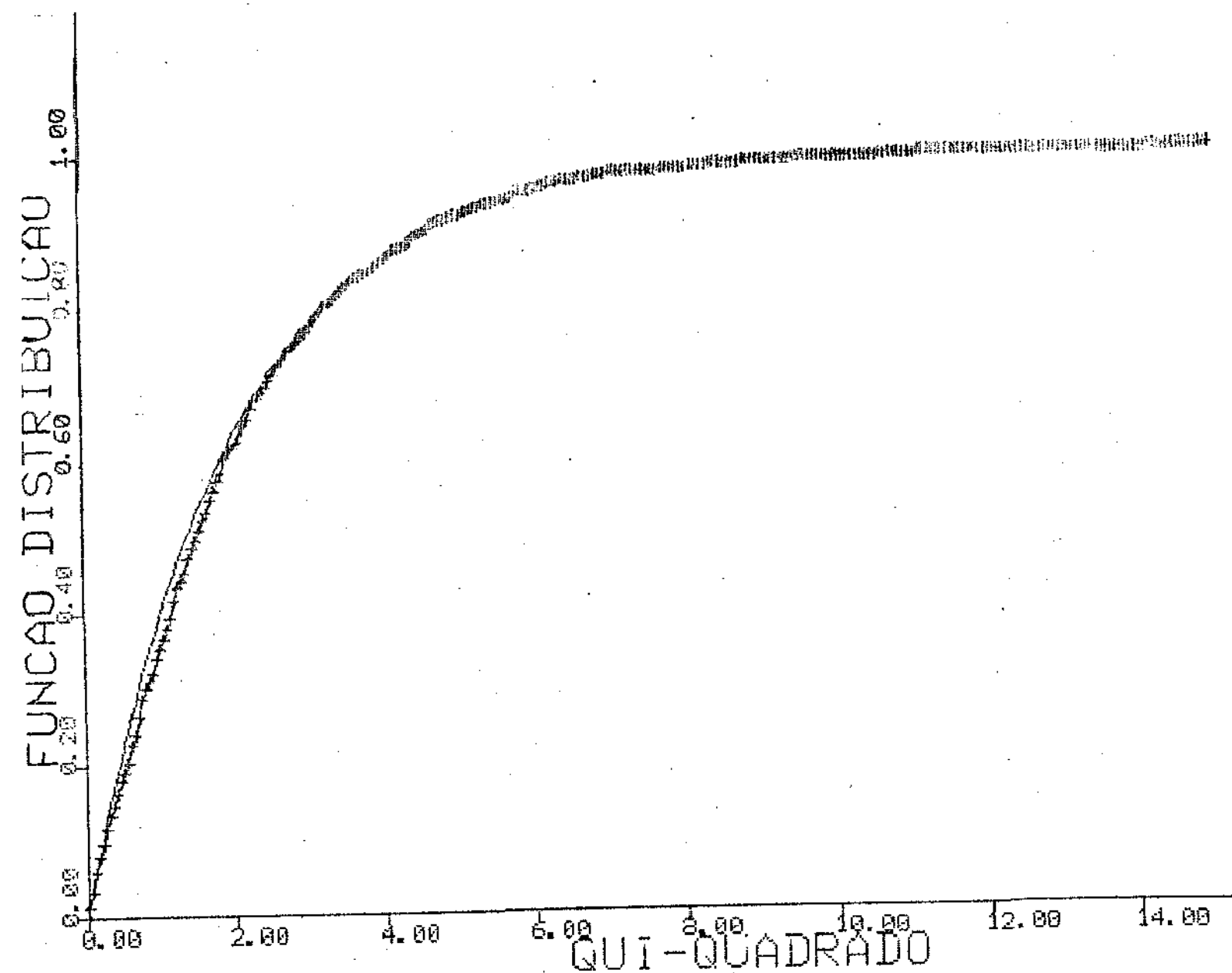


Figura 4.6 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCB, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade para AAC, com T.C. = 6 e nca = 100, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

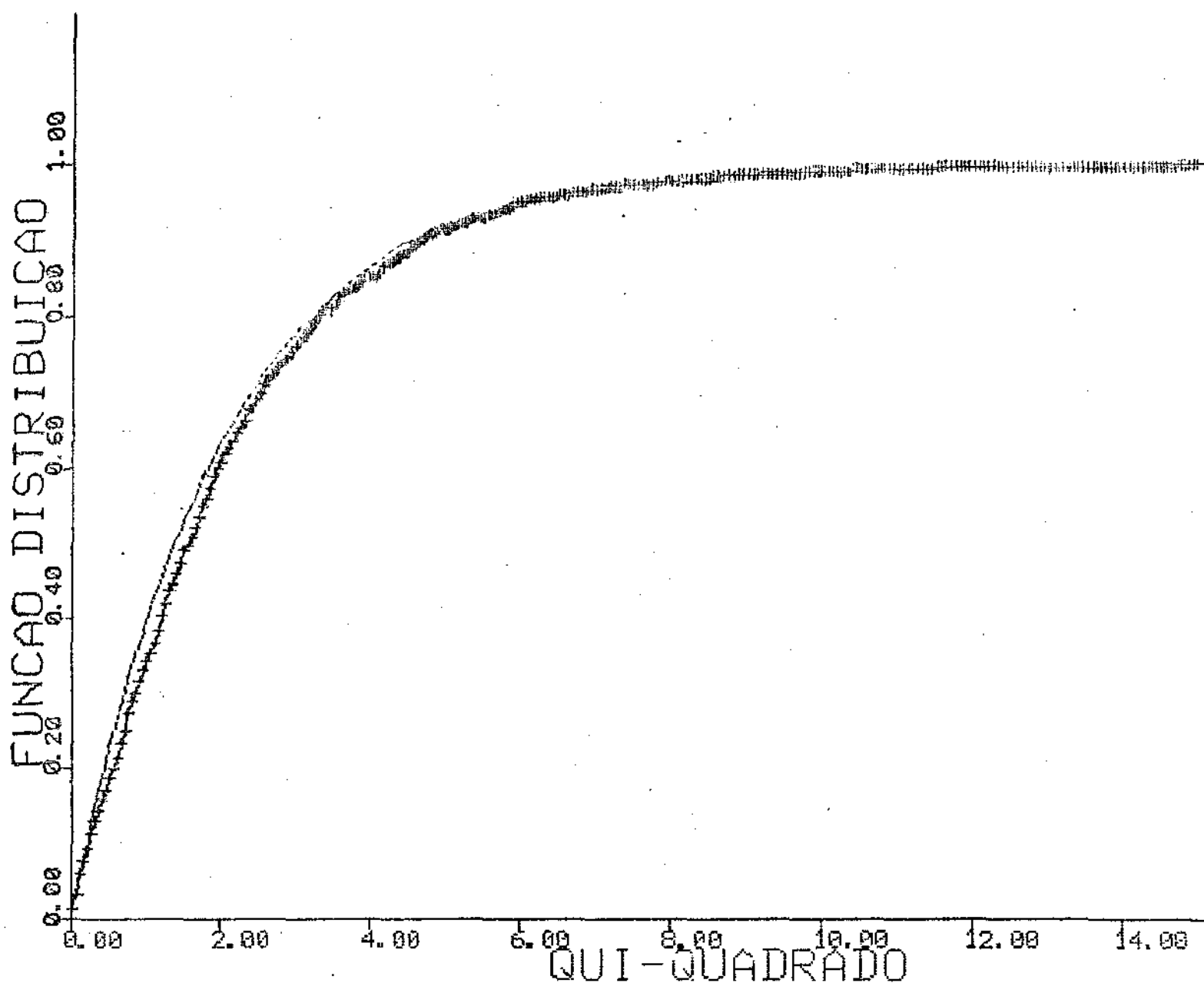


Tabela 4.4 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do conglomerado (T.C.) = 10

Fator de correção teórico = 10.00

nca		P [$\chi^2 > \chi^2_k$]					ESTIMATIVAS	
		χ^2_I	χ^2_{CT}	χ^2_{CA}	χ^2_W	χ^2_{CB}	\overline{FCA}	\overline{FCB}
5	k = 90	1.0	0.216	0.216	0.292	0.111	10.00	10.65
	k = 95	1.0	0.000	0.000	0.208	0.000		
	Média	24.95	2.49	2.49	5.22	2.31		
	Varian.	214.91	2.14	2.14	61.14	1.52		
10	k = 90	0.933	0.091	0.073	0.181	0.106	11.11	9.14
	k = 95	0.888	0.051	0.034	0.109	0.054		
	Média	15.98	1.59	1.43	1.68	1.66		
	Varian.	278.64	2.78	2.25	10.63	2.72		
15	k = 90	0.860	0.088	0.092	0.191	0.100	9.92	9.64
	k = 95	0.831	0.034	0.038	0.105	0.038		
	Média	21.01	2.10	2.08	2.78	2.14		
	Varian.	302.71	3.02	3.07	6.08	3.24		
20	k = 90	0.850	0.106	0.107	0.201	0.117	9.97	9.65
	k = 95	0.809	0.038	0.039	0.120	0.038		
	Média	21.55	2.15	2.15	2.85	2.22		
	Varian.	335.94	3.35	3.35	8.71	3.49		
25	k = 90	0.838	0.101	0.101	0.190	0.113	10.00	9.69
	k = 95	0.788	0.049	0.049	0.117	0.055		
	Média	21.69	2.16	2.16	2.76	2.23		
	Varian.	3374.14	3.74	3.74	6.85	3.95		
30	k = 90	0.707	0.093	0.093	0.165	0.100	10.00	9.72
	k = 95	0.755	0.037	0.037	0.092	0.040		
	Média	19.63	1.96	1.96	2.43	2.01		
	Varian.	348.27	3.48	3.48	6.31	3.67		

Os gráficos são os seguintes :

Figura 4.7 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCA, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade para AAC, com T.C. = 10 e nca = 30, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

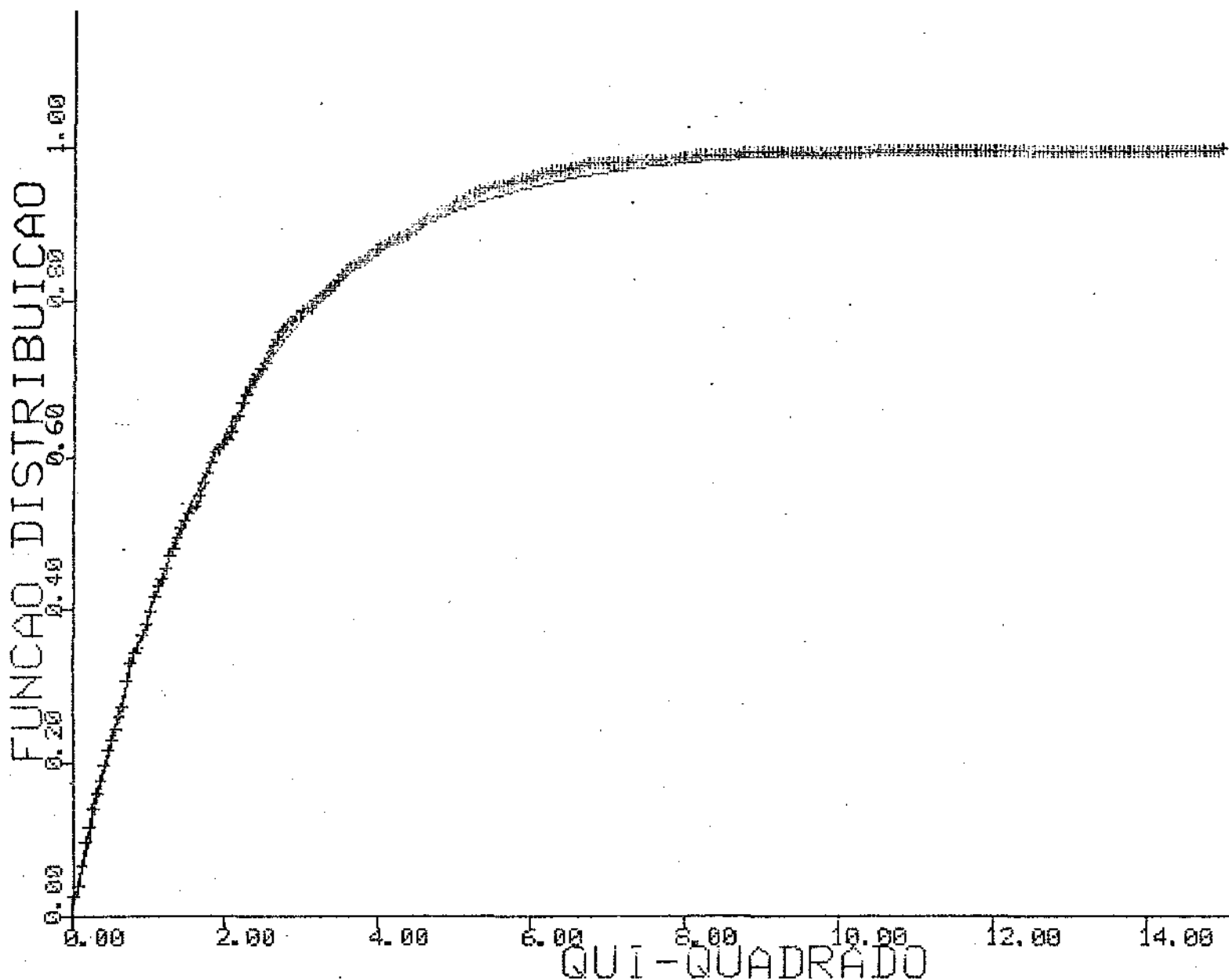


Figura 4.8 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCB, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade para AAC, com T.C. = 10 e nca = 30, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

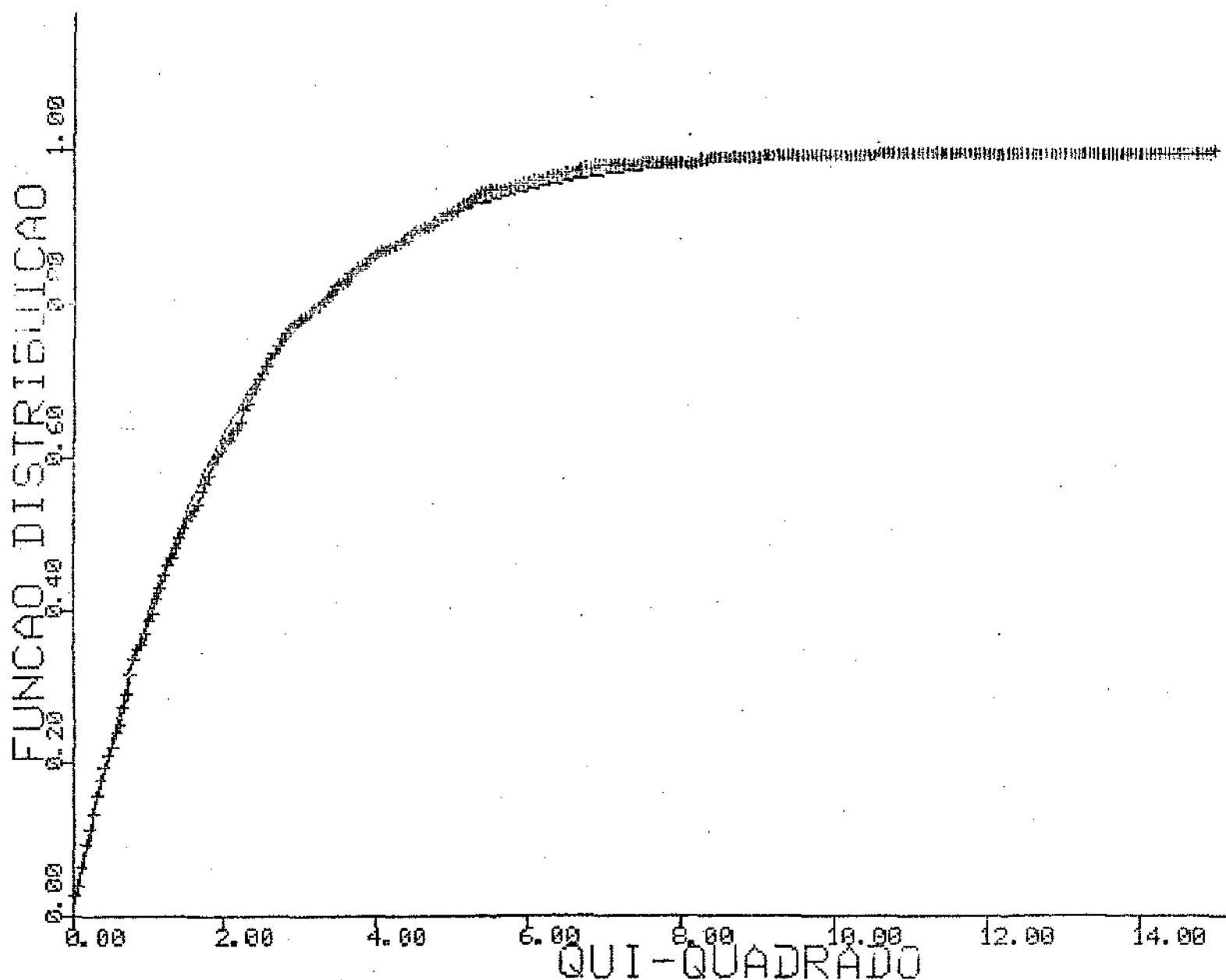


Tabela 4.5 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do conglomerado (T.C.) = 15

Fator de correção teórico = 12.58

nca		$P [X^2 > x_k^2]$					ESTIMATIVAS	
		X^2_1	X^2_{CT}	X^2_{CA}	X^2_W	X^2_{CB}	\overline{FCA}	\overline{FCB}
7	k = 90	0.902	0.103	0.143	0.224	0.134	10.76	11.21
	k = 95	0.864	0.047	0.095	0.151	0.079		
	Média	27.64	2.19	2.38	3.97	2.34		
	Varian.	574.94	3.63	5.13	65.40	6.00		
10	k = 90	0.882	0.112	0.141	0.177	0.141	11.60	11.59
	k = 95	0.847	0.060	0.081	0.114	0.079		
	Média	27.72	2.20	2.34	3.09	2.41		
	Varian.	618.51	3.90	5.86	12.49	4.77		
14	k = 90	0.844	0.106	0.122	0.181	0.134	11.95	11.56
	k = 95	0.811	0.049	0.055	0.099	0.063		
	Média	27.11	2.15	2.27	2.77	2.35		
	Varian.	560.45	3.53	4.00	6.91	4.27		
17	k = 90	0.860	0.091	0.100	0.145	0.111	12.10	11.66
	k = 95	0.832	0.035	0.045	0.084	0.050		
	Média	26.03	2.06	2.16	2.62	2.23		
	Varian.	605.60	3.82	4.16	9.17	4.25		
20	k = 90	0.874	0.092	0.102	0.147	0.107	11.99	11.60
	k = 95	0.834	0.051	0.055	0.077	0.061		
	Média	25.76	2.04	2.13	2.54	2.20		
	Varian.	602.04	3.80	3.92	6.26	4.16		

Os gráficos são os seguintes :

Figura 4.9 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCA, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade para AAC, com T.C. = 15 e nca = 20, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

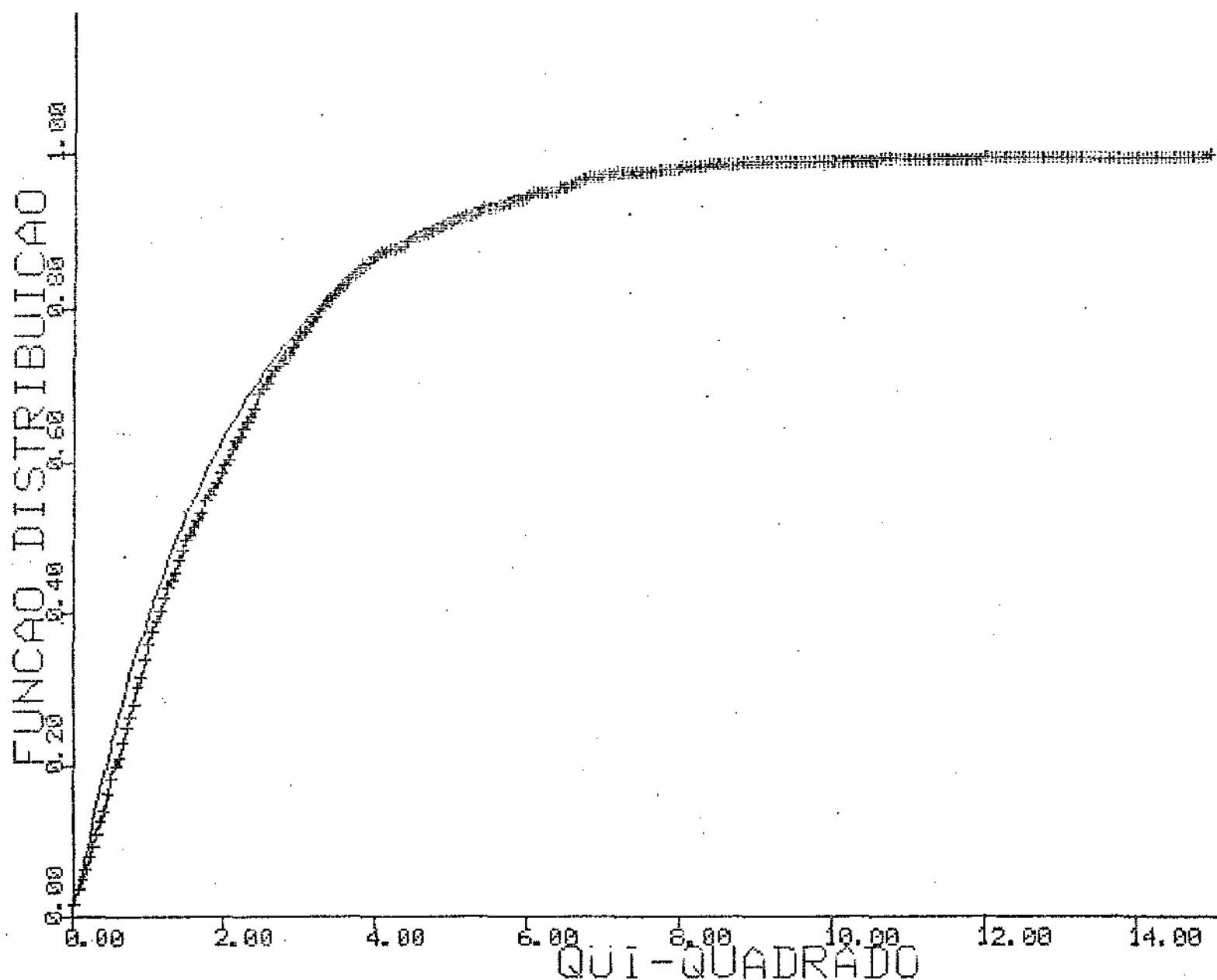


Figura 4.10 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCB, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade para AAC, com T.C. = 15 e nca = 20, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

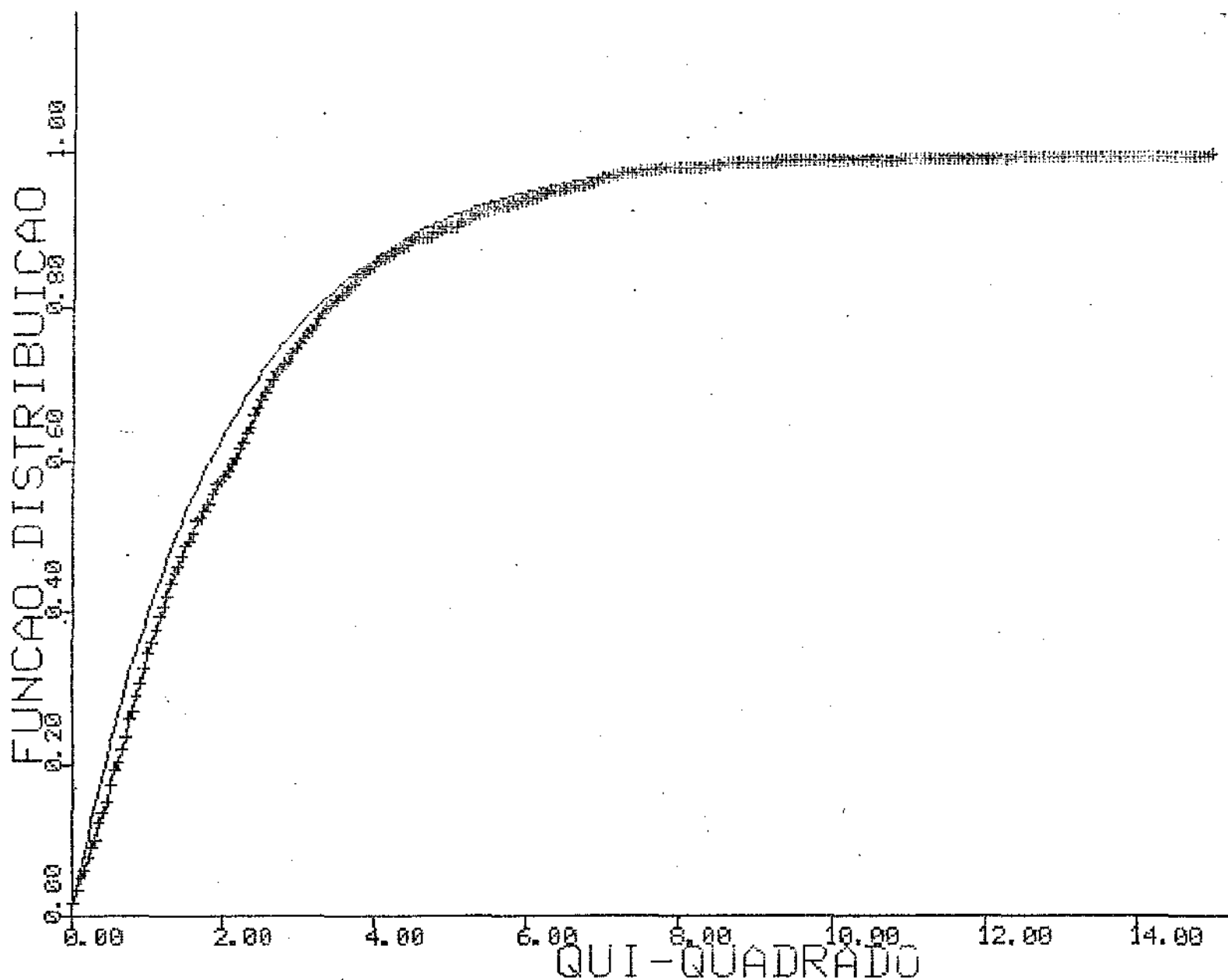


Tabela 4.6 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do conglomerado (T.C.) = 20

Fator de correção teórico = 14.86

nca		$P [X^2 > x_k^2]$					ESTIMATIVAS	
		X^2_I	X^2_{CT}	X^2_{CA}	X^2_W	X^2_{CB}	\overline{FCA}	\overline{FCB}
5	k = 90	0.942	0.066	0.188	0.325	0.105	11.04	21.01
	k = 95	0.942	0.047	0.110	0.239	0.050		
	Média	32.18	2.16	2.36	4.82	2.26		
	Varian.	567.00	2.56	4.66	56.98	5.37		
10	k = 90	0.903	0.087	0.101	0.157	0.108	13.91	13.66
	k = 95	0.871	0.041	0.049	0.098	0.050		
	Média	30.60	2.05	2.17	2.84	2.23		
	Varian.	757.59	3.43	3.62	9.87	3.61		
15	k = 90	0.896	0.109	0.101	0.157	0.109	14.27	13.80
	k = 95	0.864	0.055	0.053	0.086	0.059		
	Média	31.76	2.13	2.22	2.65	2.28		
	Varian.	826.522	3.74	4.01	5.91	4.16		

Os gráficos são os seguintes :

Figura 4.11 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCA, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC, com T.C. = 20 e nca = 10, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.

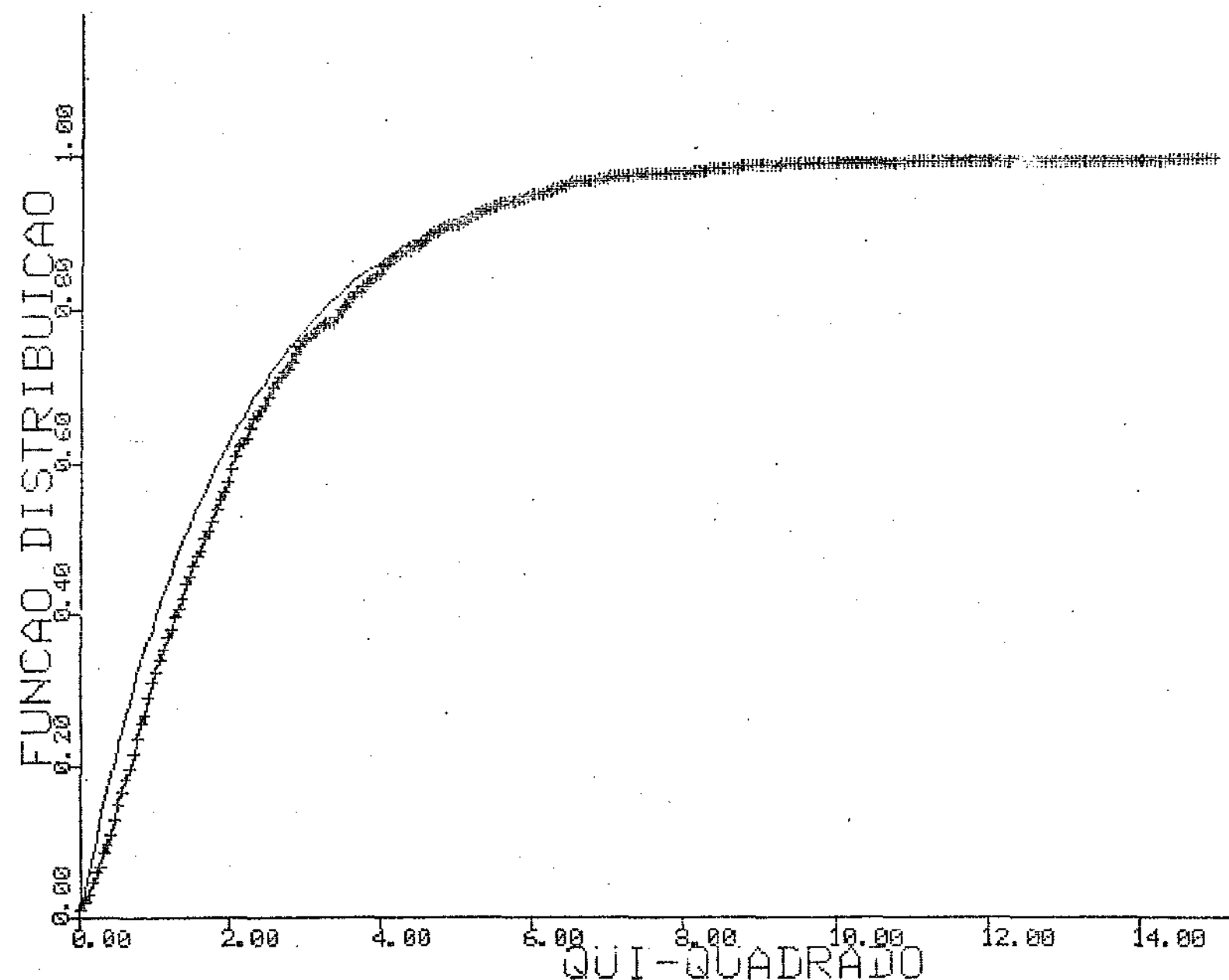
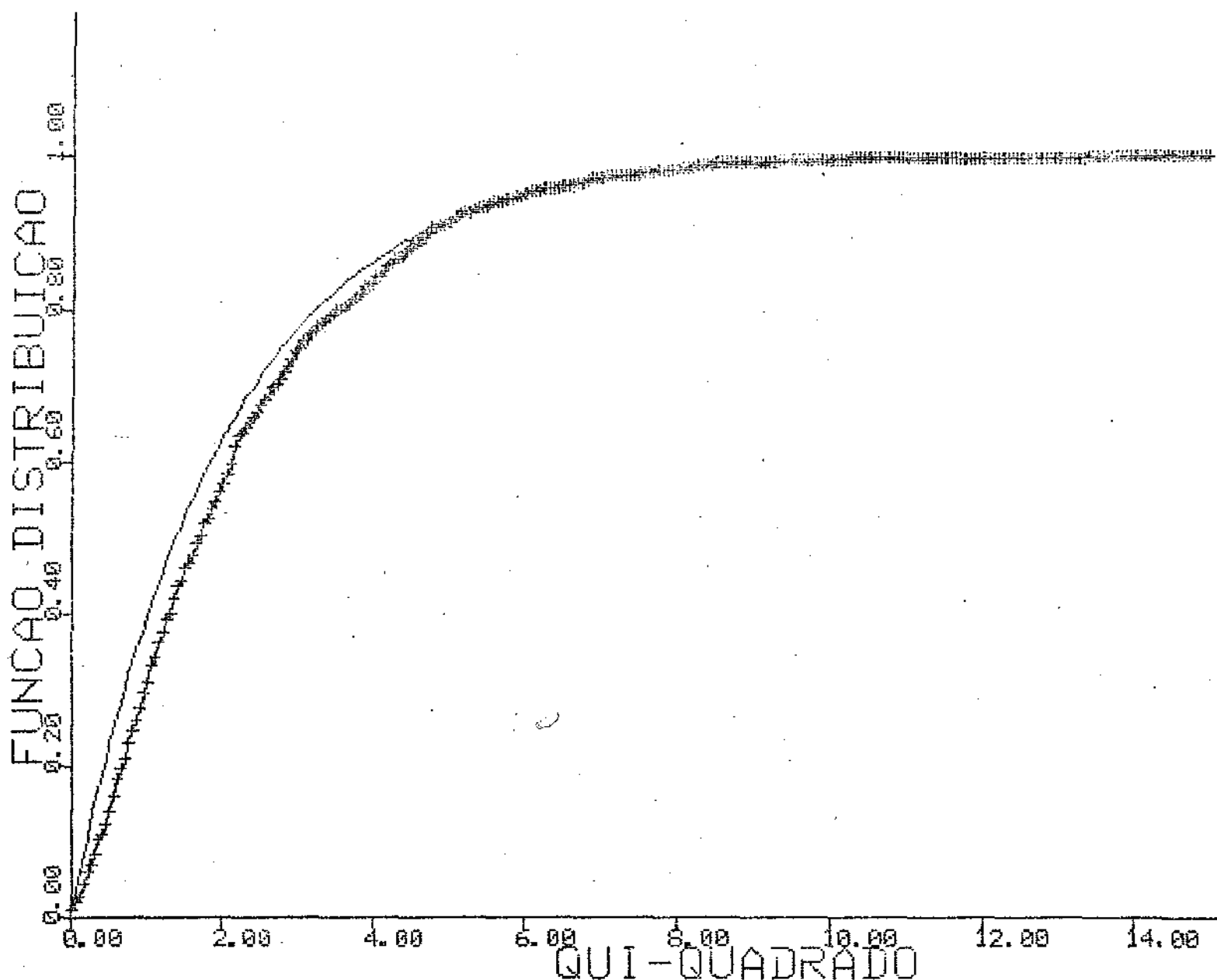


Figura 4.12 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCB, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, para AAC, com T.C. = 20 e nca = 10, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson.



4.4.2 - RESULTADOS PARA CONGLOMERADOS SEMELHANTES AOS DA POPULAÇÃO

Considerando-se os mesmos tamanhos dos conglomerados da seção 2.3.2, introduzimos a correção proposta na seção 4.3.2, estimando-se o fator de correção pelos dois métodos: por máxima verossimilhança e pelo método Bootstrap.

Apenas para efeitos de comparação, calculamos a estatística de Wald e o fator de correção populacional.

Os resultados via simulação Monte-Carlo, são apresentados nas seguintes tabelas e gráficos. Estes devem ser comparados com as tabelas e gráficos da seção 2.3.2.

Tabela 4.7 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do conglomerado (T.C.) = 3

nca		$P [\overset{2}{X} > \underset{k}{x}]$					FCT	ESTIMATIVAS	
		$\overset{2}{X}$ I	$\overset{2}{X}$ CT	$\overset{2}{X}$ CA	$\overset{2}{X}$ W	$\overset{2}{X}$ CB		\overline{FCA}	\overline{FCB}
15	k = 90	0.081	0.098	0.103	0.179	0.126	0.91	0.97	0.88
	k = 95	0.034	0.052	0.054	0.114	0.067			
	Média	1.54	1.68	1.71	2.12	1.87			
	Varian.	2.55	3.03	3.55	10.94	4.09			
35	k = 90	0.078	0.085	0.098	0.115	0.109	0.93	0.87	0.85
	k = 95	0.032	0.042	0.045	0.069	0.051			
	Média	1.21	1.29	1.49	1.66	1.53			
	Varian.	2.52	2.86	3.39	5.22	3.61			
50	k = 90	0.079	0.094	0.095	0.111	0.106	0.93	0.94	0.91
	k = 95	0.035	0.041	0.051	0.061	0.056			
	Média	1.39	1.48	1.61	1.80	1.62			
	Varian.	2.55	2.91	3.16	4.05	3.33			
65	k = 90	0.008	0.098	0.098	0.114	0.099	0.88	0.95	0.91
	k = 95	0.038	0.060	0.054	0.053	0.055			
	Média	1.56	1.76	1.73	1.66	1.79			
	Varian.	2.54	3.23	3.11	3.33	3.24			
85	k = 90	0.079	0.096	0.091	0.098	0.097	0.89	0.89	0.87
	k = 95	0.034	0.052	0.045	0.052	0.048			
	Média	2.31	2.57	2.69	2.93	2.71			
	Varian.	2.71	3.36	3.32	3.66	3.42			
100	k = 90	0.096	0.103	0.115	0.117	0.120	0.97	0.97	0.96
	k = 95	0.051	0.054	0.060	0.072	0.065			
	Média	2.93	3.01	3.09	3.28	3.08			
	Varian.	3.19	3.36	3.74	4.55	3.88			

Os gráficos são os seguintes :

Figura 4.13 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCA, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, com $T.C. = 3$ e $nca = 15$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.10.

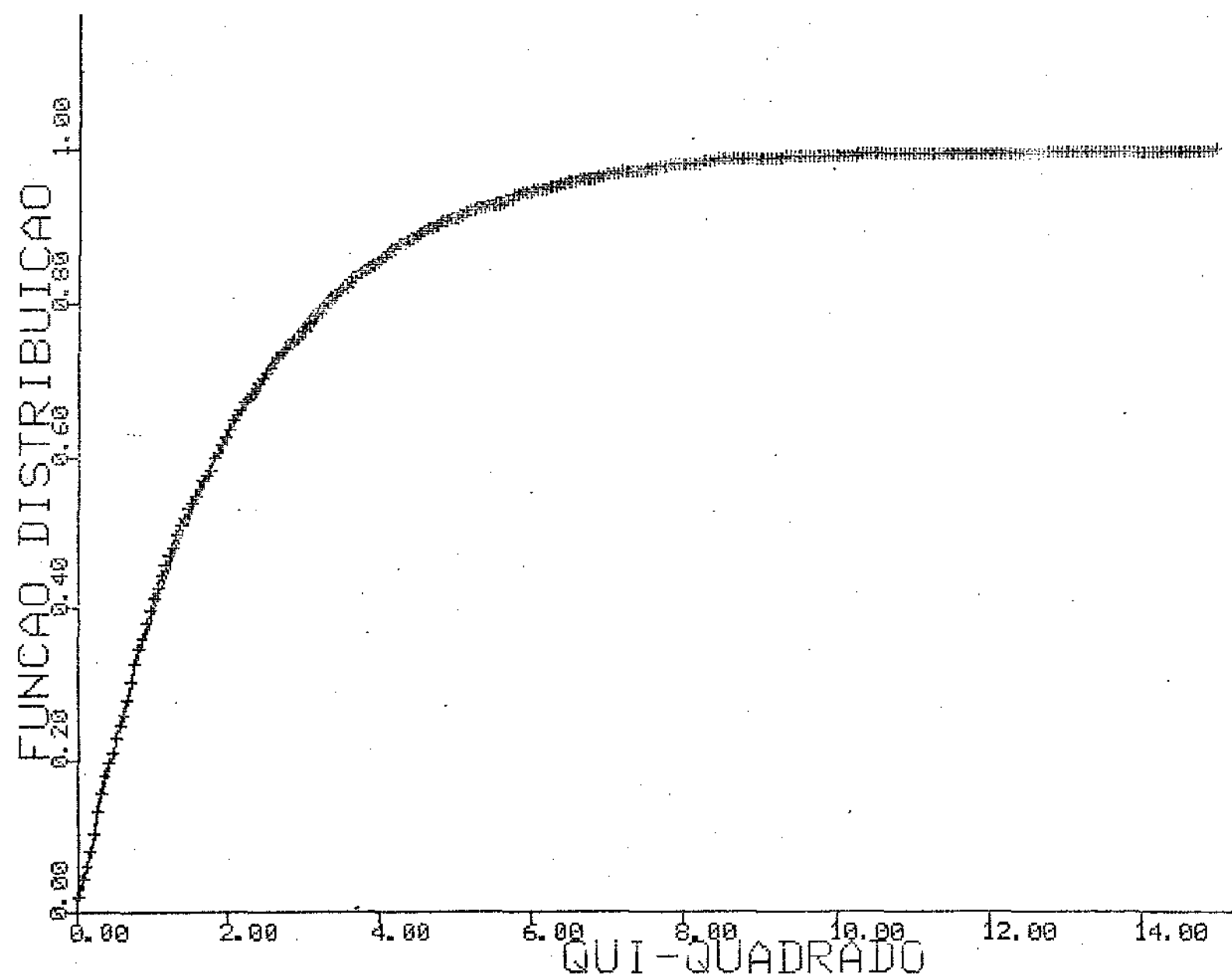


Figura 4.14 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCB, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, com $T.C. = 3$ e $nca = 15$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.10.

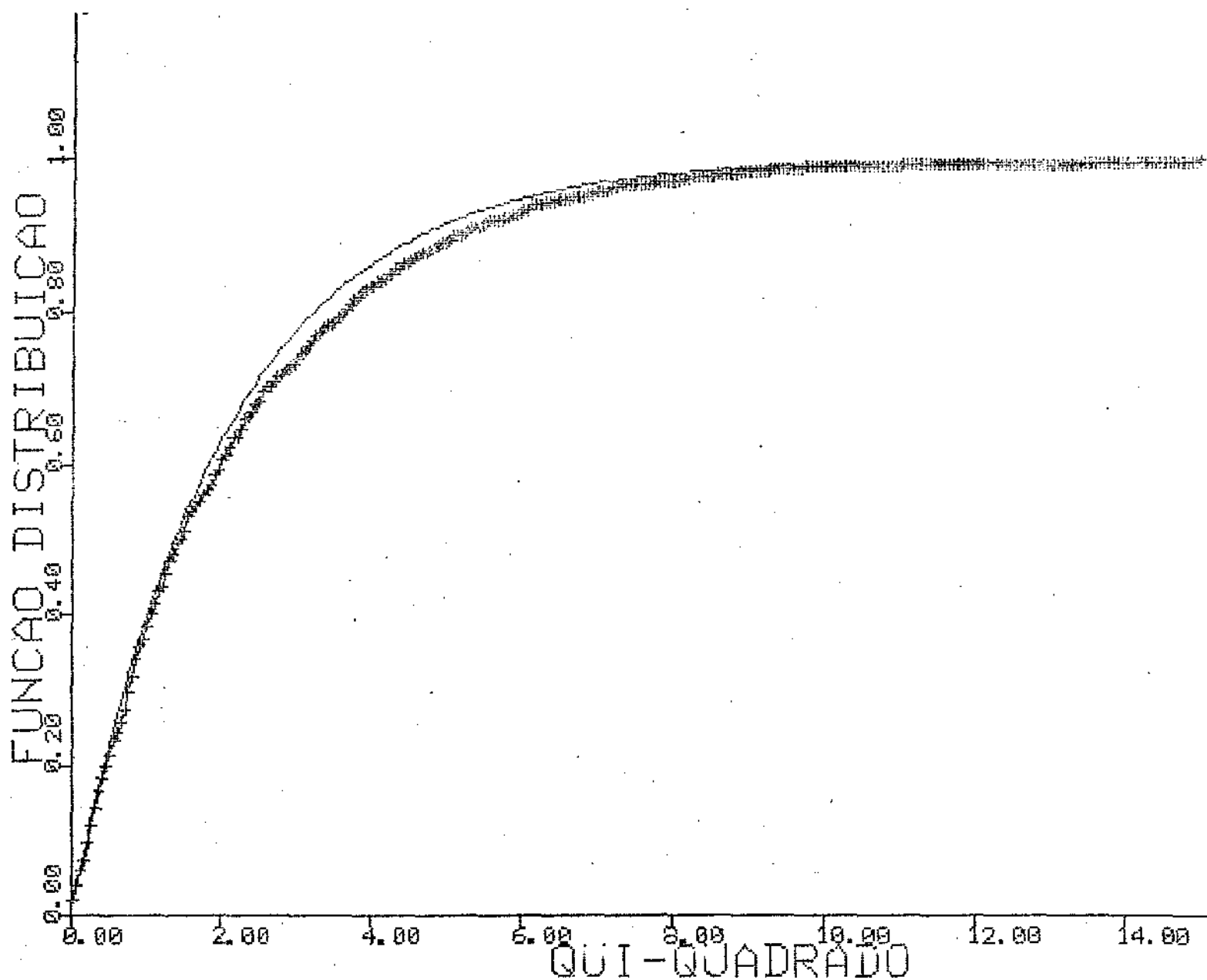


Tabela 4.8 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do conglomerado (T.C.) = 5

nca		$P \left[\chi^2 > x \right]$ k					FCT	ESTIMATIVAS	
		χ^2_1	χ^2_{CT}	χ^2_{CA}	χ^2_W	χ^2_{CB}		\overline{FCA}	\overline{FCB}
10	k = 90	0.079	0.117	0.115	0.179	0.140	0.84	0.93	0.84
	k = 95	0.036	0.061	0.074	0.144	0.088			
	Média	0.95	1.14	1.24	2.88	1.37			
	Varian.	2.54	3.60	4.40	17.54	5.54			
20	k = 90	0.053	0.124	0.114	0.154	0.137	0.69	0.66	0.62
	k = 95	0.022	0.067	0.060	0.955	0.072			
	Média	0.48	0.69	0.78	0.93	0.84			
	Varian.	1.87	3.89	3.81	6.81	4.21			
30	k = 90	0.105	0.133	0.137	0.160	0.147	0.86	0.98	0.95
	k = 95	0.049	0.082	0.078	0.101	0.084			
	Média	1.24	1.44	1.34	1.48	1.40			
	Varian.	2.95	3.99	3.95	6.13	4.33			
40	k = 90	0.066	0.129	0.098	0.105	0.107	0.76	0.89	0.88
	k = 95	0.034	0.067	0.048	0.060	0.049			
	Média	2.19	2.86	2.38	2.60	2.38			
	Varian.	2.52	4.3	3.49	4.42	3.67			
50	k = 90	0.065	0.110	0.108	0.128	0.116	0.80	0.79	0.78
	k = 95	0.026	0.062	0.059	0.071	0.063			
	Média	0.92	1.14	1.16	1.26	1.17			
	Varian.	2.31	3.54	3.82	4.71	4.05			
60	k = 90	0.070	0.093	0.094	0.116	0.099	0.91	0.89	0.88
	k = 95	0.038	0.046	0.053	0.060	0.053			
	Média	2.86	3.13	3.24	4.04	3.25			
	Varian.	2.83	3.39	3.51	3.88	3.68			

Os gráficos são os seguintes :

Figura 4.15 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCA, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, com $T.C. = 5$ e $nca = 20$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.11.

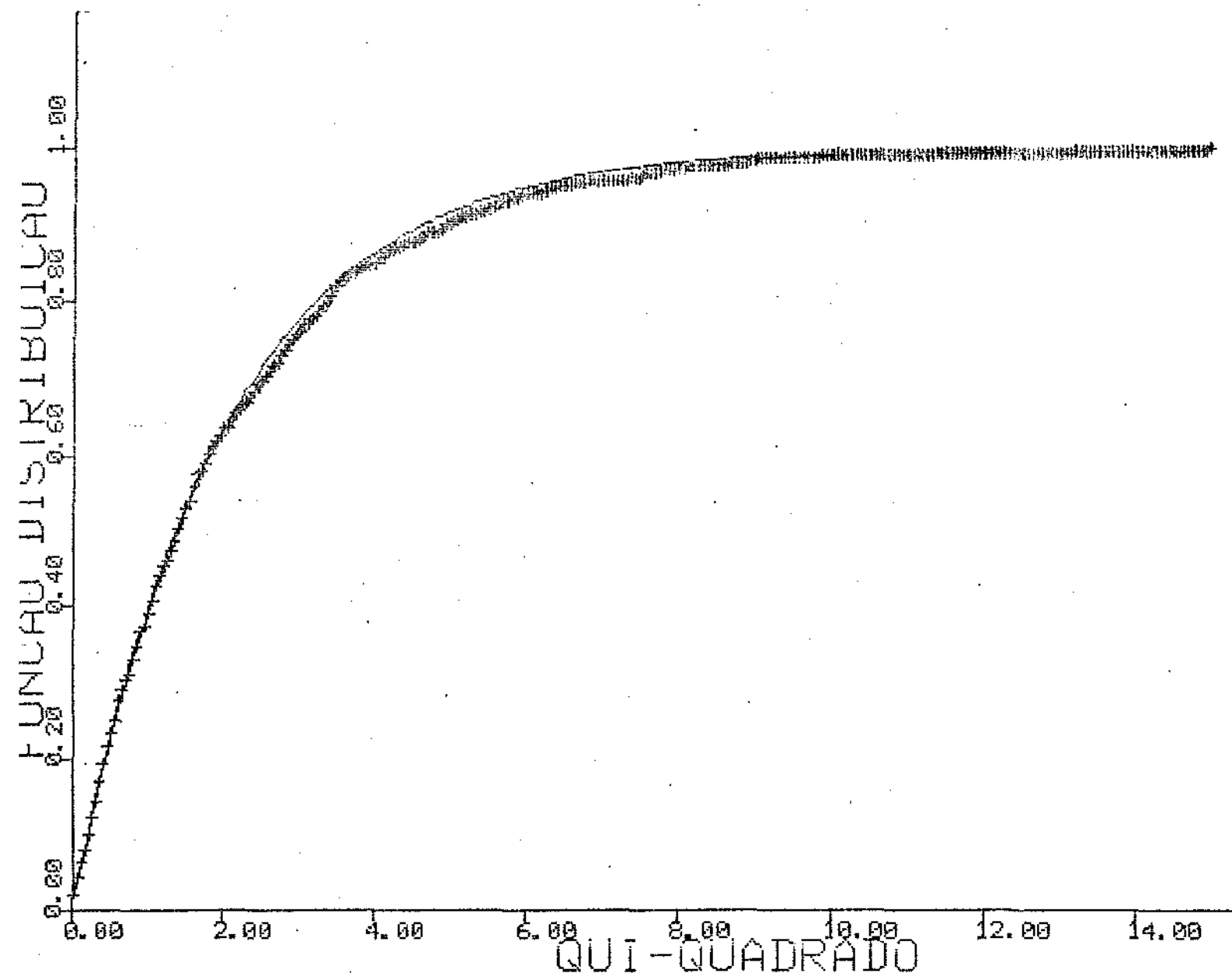


Figura 4.16 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCB, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, com $T.C. = 5$ e $nca = 20$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.11.

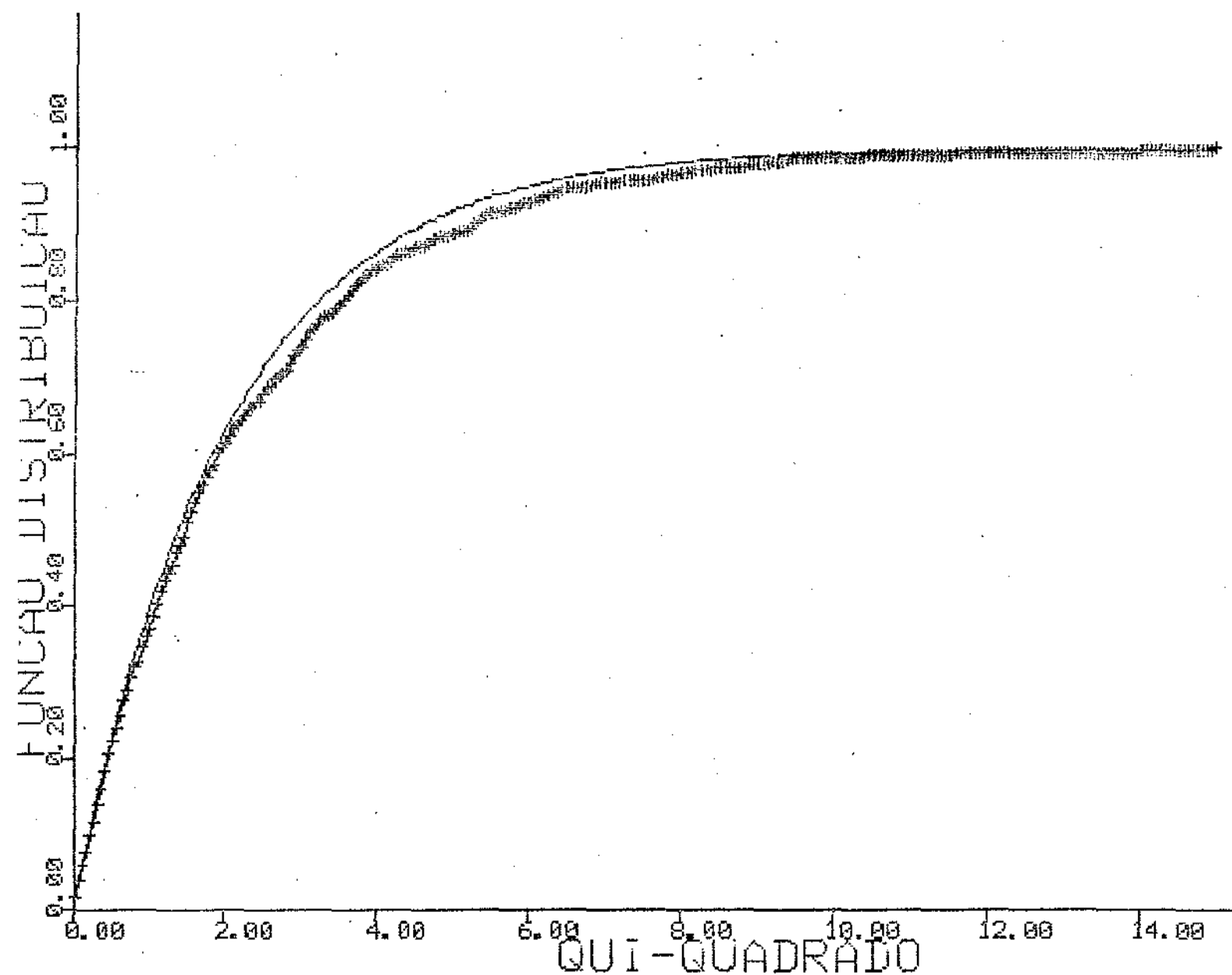


Tabela 4.9 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do conglomerado (T.C.) = 6

nca		$P [X > x_k]$					FCT	ESTIMATIVAS	
		x_I	x_{CT}	x_{CA}	x_W	x_{CB}		\overline{FCA}	\overline{FCB}
10	k = 90	0.045	0.065	0.141	0.219	0.167	0.87	0.74	0.66
	k = 95	0.018	0.035	0.087	0.157	0.108			
	Média	1.07	1.23	1.59	2.11	1.78			
	Varian.	1.72	2.24	8.44	30.46	9.34			
20	k = 90	0.041	0.106	0.107	0.145	0.115	0.69	0.63	0.61
	k = 95	0.021	0.054	0.057	0.092	0.068			
	Média	0.58	0.83	0.89	0.83	0.93			
	Varian.	1.82	3.75	4.20	7.04	4.66			
25	k = 90	0.067	0.101	0.107	0.126	0.119	0.85	0.74	0.72
	k = 95	0.029	0.052	0.059	0.073	0.067			
	Média	1.92	2.26	2.74	2.65	2.87			
	Varian.	2.44	3.36	3.89	5.78	4.33			
35	k = 90	0.056	0.112	0.095	0.119	0.099	0.72	0.78	0.75
	k = 95	0.017	0.065	0.051	0.064	0.053			
	Média	0.68	0.94	0.86	0.70	0.90			
	Varian.	1.77	3.32	2.99	4.55	3.20			
45	k = 90	0.074	0.113	0.102	0.112	0.106	0.83	0.80	0.77
	k = 95	0.031	0.059	0.055	0.068	0.061			
	Média	0.76	0.92	0.93	1.07	0.97			
	Varian.	2.75	3.95	3.87	4.47	4.10			
50	k = 90	0.071	0.109	0.096	0.105	0.103	0.80	0.82	0.81
	k = 95	0.029	0.061	0.060	0.065	0.060			
	Média	1.44	1.79	1.65	2.37	1.67			
	Varian.	2.38	3.70	3.52	4.09	3.70			

Os gráficos são os seguintes :

Figura 4.17 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCA, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, com T.C. = 6 e nca = 20, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.12.

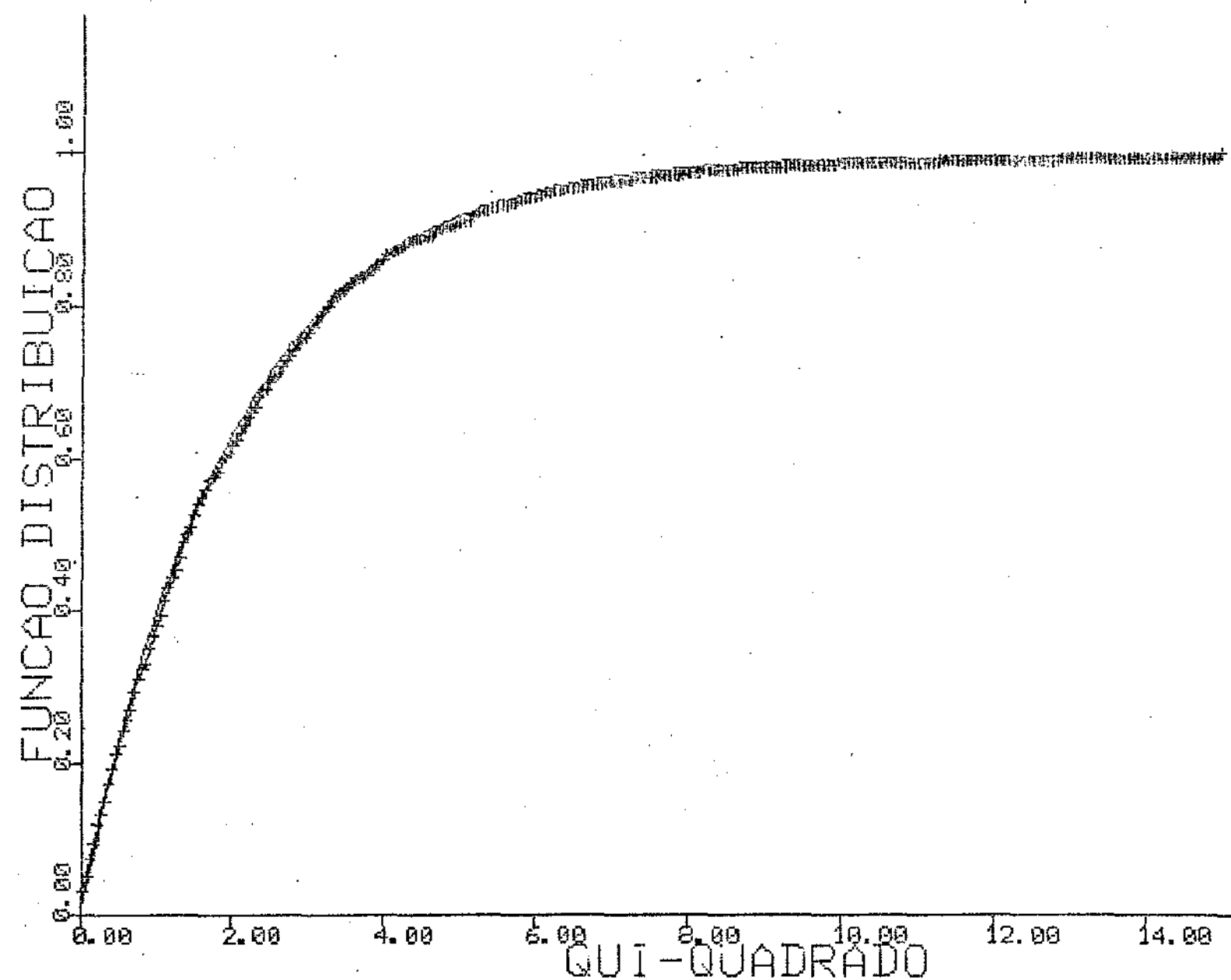


Figura 4.18 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCB, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, com $T.C. = 6$ e $nca = 20$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.12.

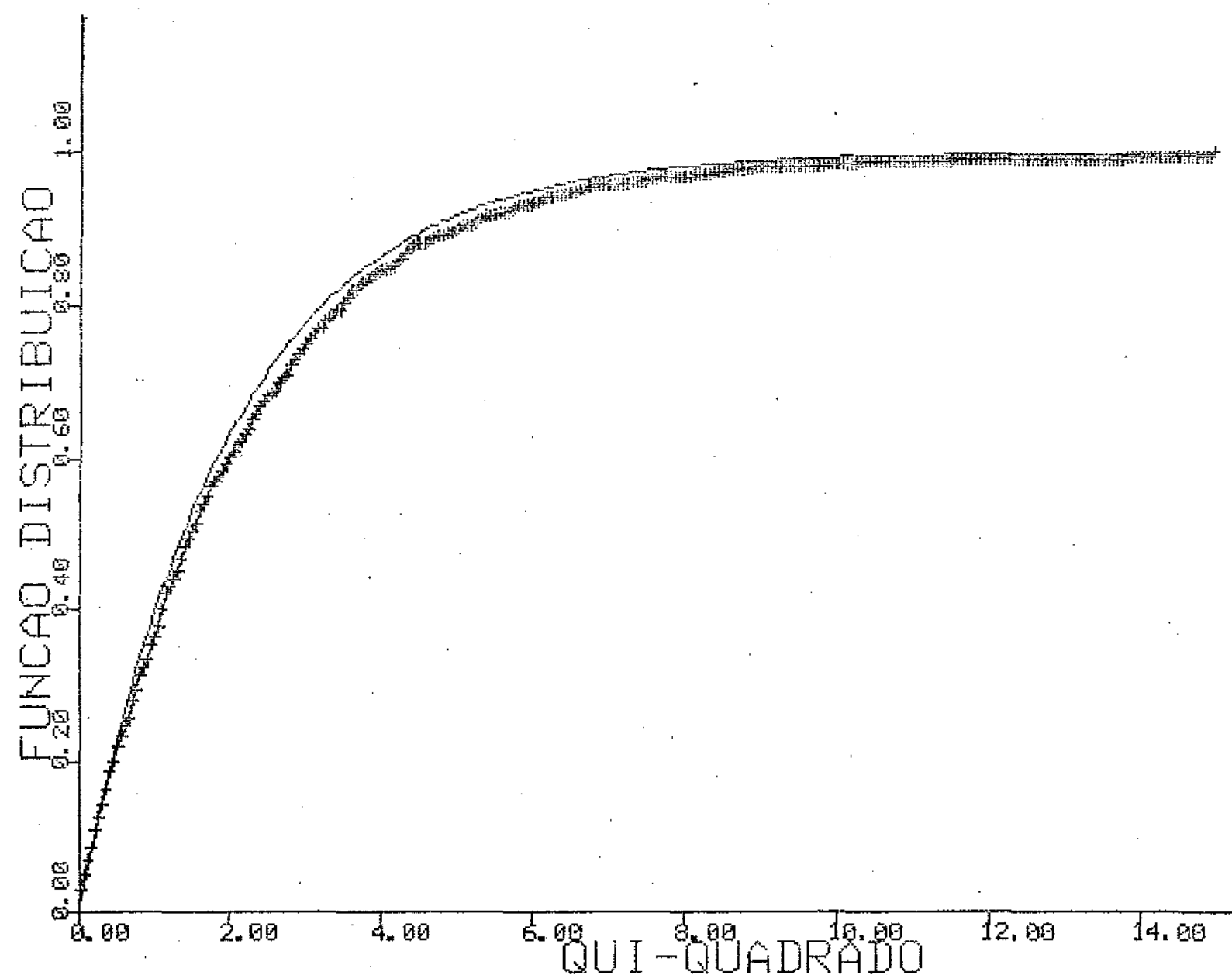


Tabela 4.10 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do conglomerado (T.C.) = 10

nca		$P [X^2 > x]$ k					FCT	ESTIMATIVAS	
		2	2	2	2	2		FCA	FCB
		X_I	X_{CT}	X_{CA}	X_W	X_{CB}			
5	K = 90	0.054	0.077	0.159	0.337	0.205	0.84	0.54	0.44
	k = 95	0.027	0.044	0.105	0.283	0.153			
	Média	3.06	3.63	7.17	8.78	8.71			
	Varian.	2.30	3.25	26.31	2414.9	41.96			
10	k = 90	0.067	0.138	0.126	0.195	0.145	0.72	0.56	0.50
	k = 95	0.036	0.080	0.082	0.131	0.102			
	Média	3.07	4.25	6.60	8.09	7.30			
	Varian.	2.64	5.07	5.88	23.19	7.18			
15	k = 90	0.055	0.168	0.119	0.163	0.135	0.61	0.67	0.62
	k = 95	0.028	0.102	0.060	0.102	0.076			
	Média	3.71	6.02	5.55	5.21	6.05			
	Varian.	2.07	5.43	4.54	9.73	5.26			
20	k = 90	0.021	0.123	0.105	0.142	0.114	0.54	0.61	0.57
	k = 95	0.007	0.071	0.059	0.077	0.062			
	Média	1.00	1.85	1.78	1.95	1.88			
	Varian.	1.22	4.13	4.21	5.20	4.64			
25	k = 90	0.035	0.140	0.115	0.141	0.122	0.60	0.66	0.63
	k = 95	0.011	0.075	0.064	0.082	0.070			
	Média	1.24	2.05	1.87	2.06	1.95			
	Varian.	1.53	4.17	4.01	5.25	4.36			
30	k = 90	0.033	0.081	0.094	0.114	0.105	0.74	0.73	0.72
	k = 95	0.015	0.039	0.046	0.057	0.050			
	Média	0.50	0.67	0.68	0.65	0.72			
	Varian.	1.69	3.07	3.26	4.19	3.50			

Os gráficos são os seguintes :

Figura 4.19 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCA, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, com $T.C. = 10$ e $nca = 20$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.13.

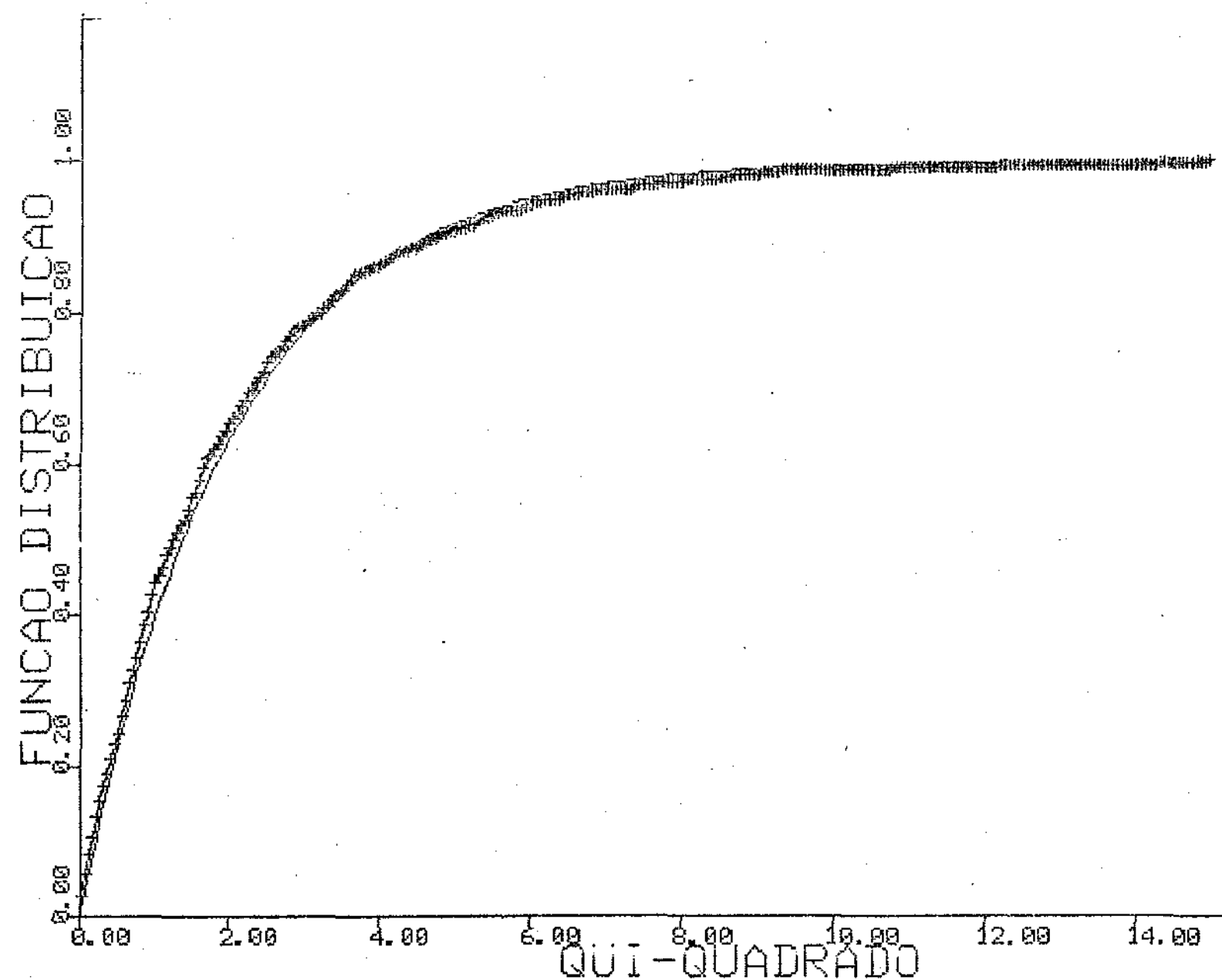


Figura 4.20 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCB, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, com $T.C. = 10$ e $nca = 20$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.13.

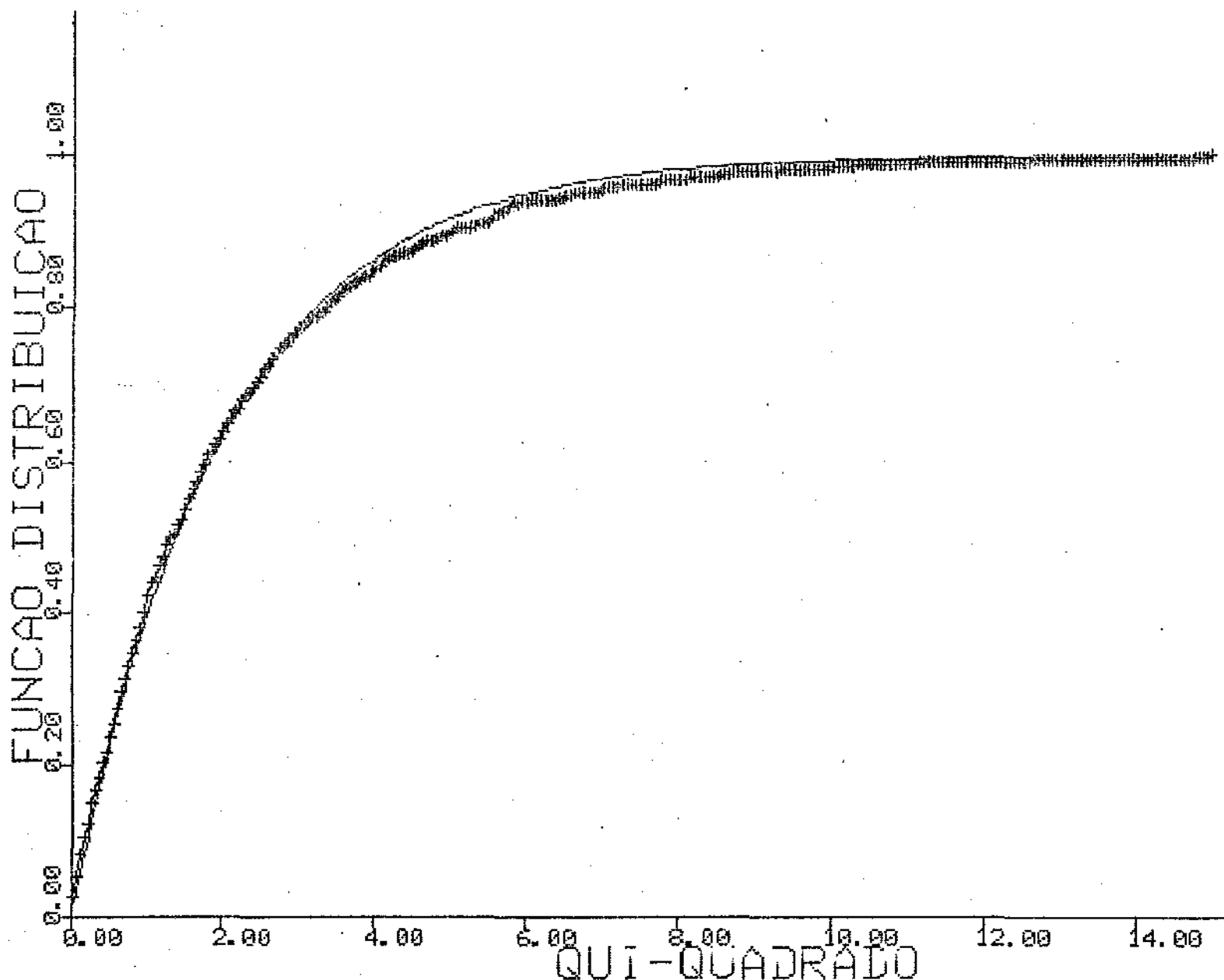


Tabela 4.11 Comparação da cauda da distribuição exata da Estatística de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do conglomerado (T.C.) = 15

nca		$P \left[\sum_{i=1}^k X_i^2 > x \right]$					FCT	ESTIMATIVAS	
		$\sum_{i=1}^k X_i^2$	$\sum_{i=1}^k X_{CT}^2$	$\sum_{i=1}^k X_{CA}^2$	$\sum_{i=1}^k X_W^2$	$\sum_{i=1}^k X_{CB}^2$		\overline{FCA}	\overline{FCB}
5	k = 90	0.007	0.072	0.183	0.311	0.240	0.52	0.50	0.41
	k = 95	0.003	0.036	0.116	0.250	0.172			
	Média	0.68	1.30	1.73	4.02	2.08			
	Varian.	0.67	2.46	12.34	4740.4	18.46			
7	k = 90	0.018	0.075	0.156	0.273	0.186	0.60	0.48	0.40
	k = 95	0.007	0.033	0.092	0.212	0.128			
	Média	0.90	1.50	2.72	4.48	3.30			
	Varian.	1.10	3.07	10.76	843.6	14.24			
10	k = 90	0.014	0.076	0.118	0.193	0.137	0.60	0.56	0.50
	k = 95	0.005	0.041	0.076	0.126	0.089			
	Média	1.06	1.73	2.02	2.07	2.25			
	Varian.	0.96	2.58	6.35	19.06	7.82			
14	k = 90	0.007	0.151	0.114	0.150	0.131	0.35	0.38	0.35
	k = 95	0.001	0.087	0.063	0.102	0.076			
	Média	0.52	1.47	1.15	1.51	1.27			
	Varian.	0.68	5.42	4.37	9.60	5.07			
17	k = 90	0.003	0.061	0.102	0.131	0.114	0.47	0.41	0.38
	k = 95	0.001	0.018	0.060	0.081	0.068			
	Média	1.05	2.20	2.77	3.34	3.02			
	Varian.	0.47	2.08	4.11	7.66	4.54			
20	k = 90	0.011	0.091	0.120	0.147	0.126	0.45	0.39	0.38
	k = 95	0.003	0.065	0.070	0.090	0.080			
	Média	0.46	1.01	1.66	1.21	1.72			
	Varian.	0.83	3.96	4.70	5.74	5.20			

Os gráficos são os seguintes :

Figura 4.21 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCA, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, com $T.C. = 15$ e $nca = 14$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.14.

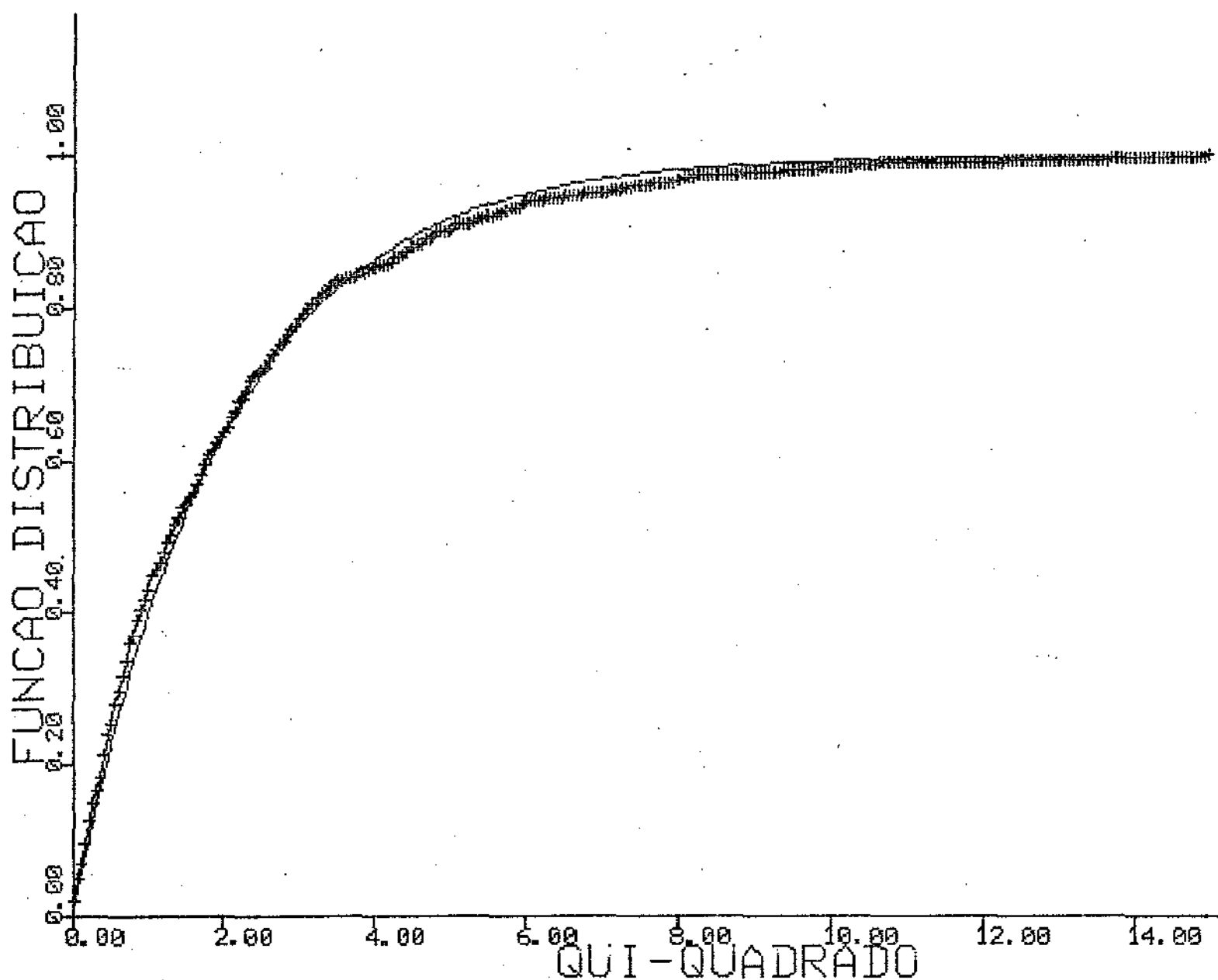


Figura 4.22 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCB, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, com $T.C. = 15$ e $nca = 14$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.14.

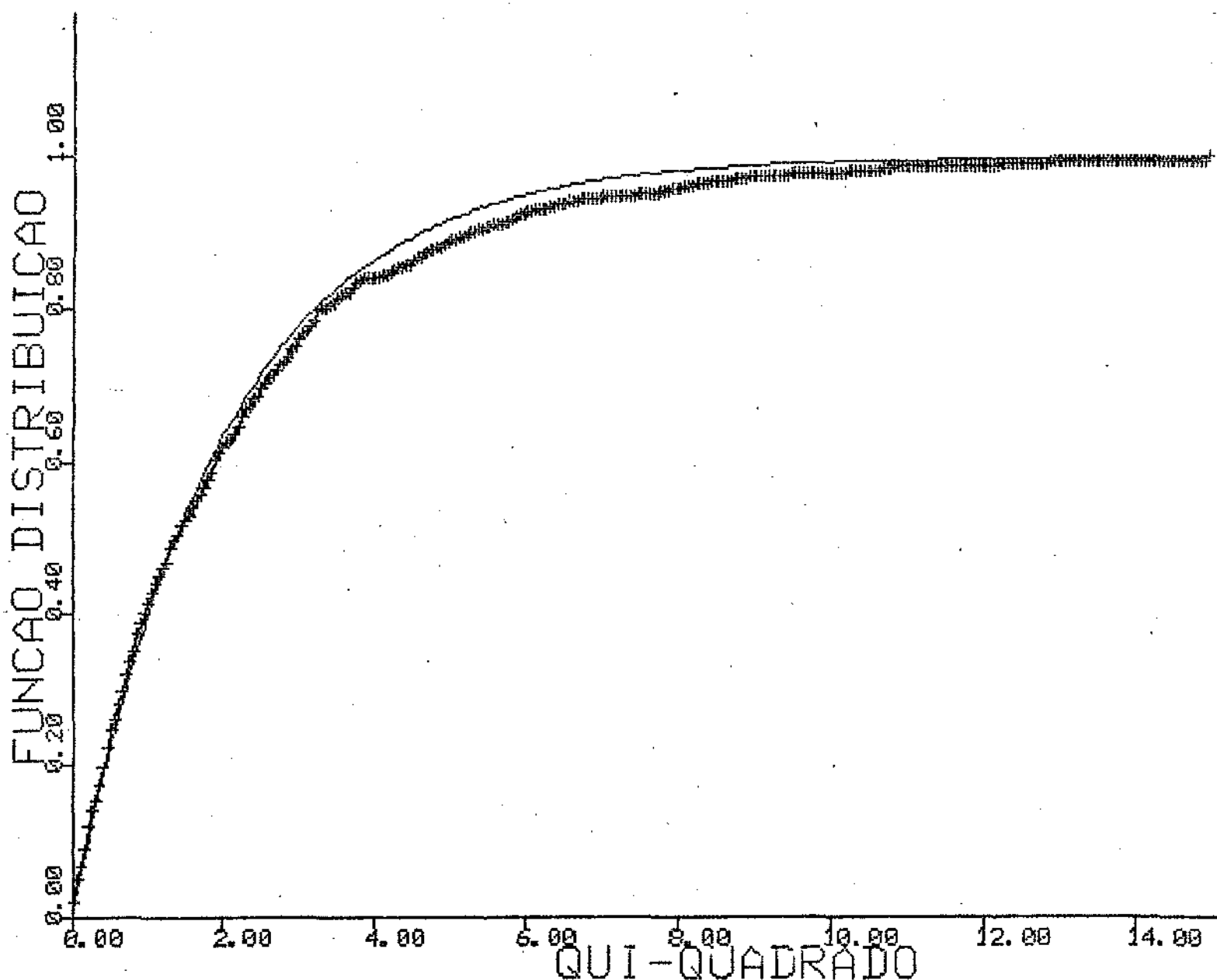


Tabela 4.12 Comparação da distribuição exata da Estatística de Pearson de Pearson corrigida, para AAC e diversos números de conglomerados amostrados (nca), estimada com base em 1000 repetições M.C., com a correspondente distribuição Qui-Quadrado com 2 graus de liberdade, com média teórica igual a 2 e variância teórica igual a 4.

Tamanho do conglomerado (T.C.) = 20

nca		$P \left[\chi^2 > \chi^2_k \right]$					FCT	ESTIMATIVAS	
		χ^2_1	χ^2_{CT}	χ^2_{CA}	χ^2_W	χ^2_{CB}		\overline{FCA}	\overline{FCB}
5	k = 90	0.001	0.016	0.166	0.305	0.199	0.62	0.36	0.28
	k = 95	0.000	0.003	0.119	0.254	0.156			
	Média	0.31	0.50	1.09	1.37	1.34			
	Varian.	0.38	0.98	13.39	2453.8	19.77			
10	k = 90	0.005	0.048	0.138	0.208	0.163	0.45	0.31	0.28
	k = 95	0.004	0.023	0.090	0.146	0.107			
	Média	0.77	1.68	2.76	3.46	3.18			
	Varian.	0.54	2.63	11.50	36.14	14.03			
15	k = 90	0.021	0.056	0.141	0.173	0.154	0.68	0.46	0.41
	k = 95	0.007	0.031	0.093	0.108	0.107			
	Média	1.51	2.21	3.63	3.34	4.05			
	Varian.	1.53	3.30	7.98	8.07	9.30			

Os gráficos são os seguintes :

Figura 4.23 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCA, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, com $T.C. = 20$ e $nca = 5$, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.15.

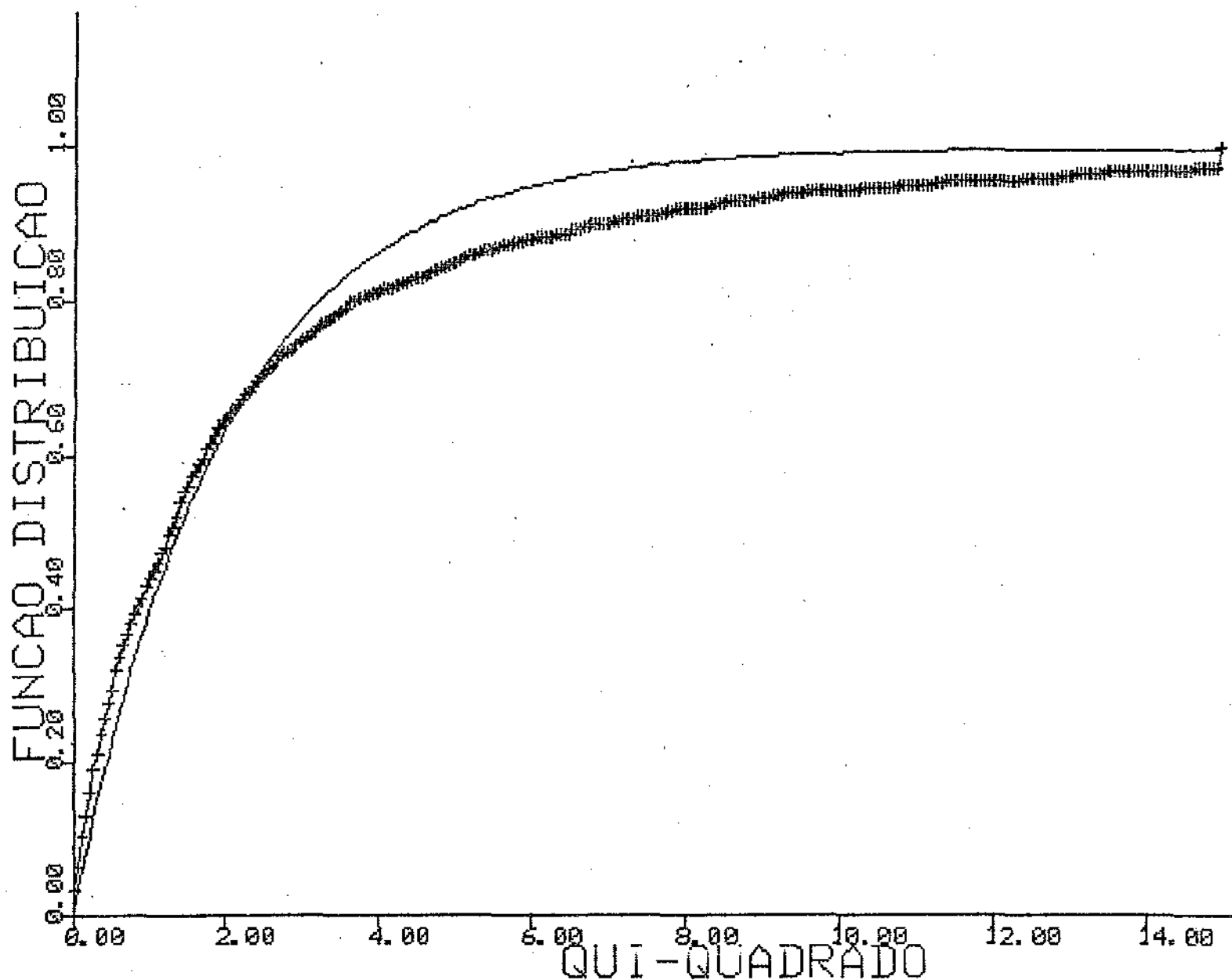
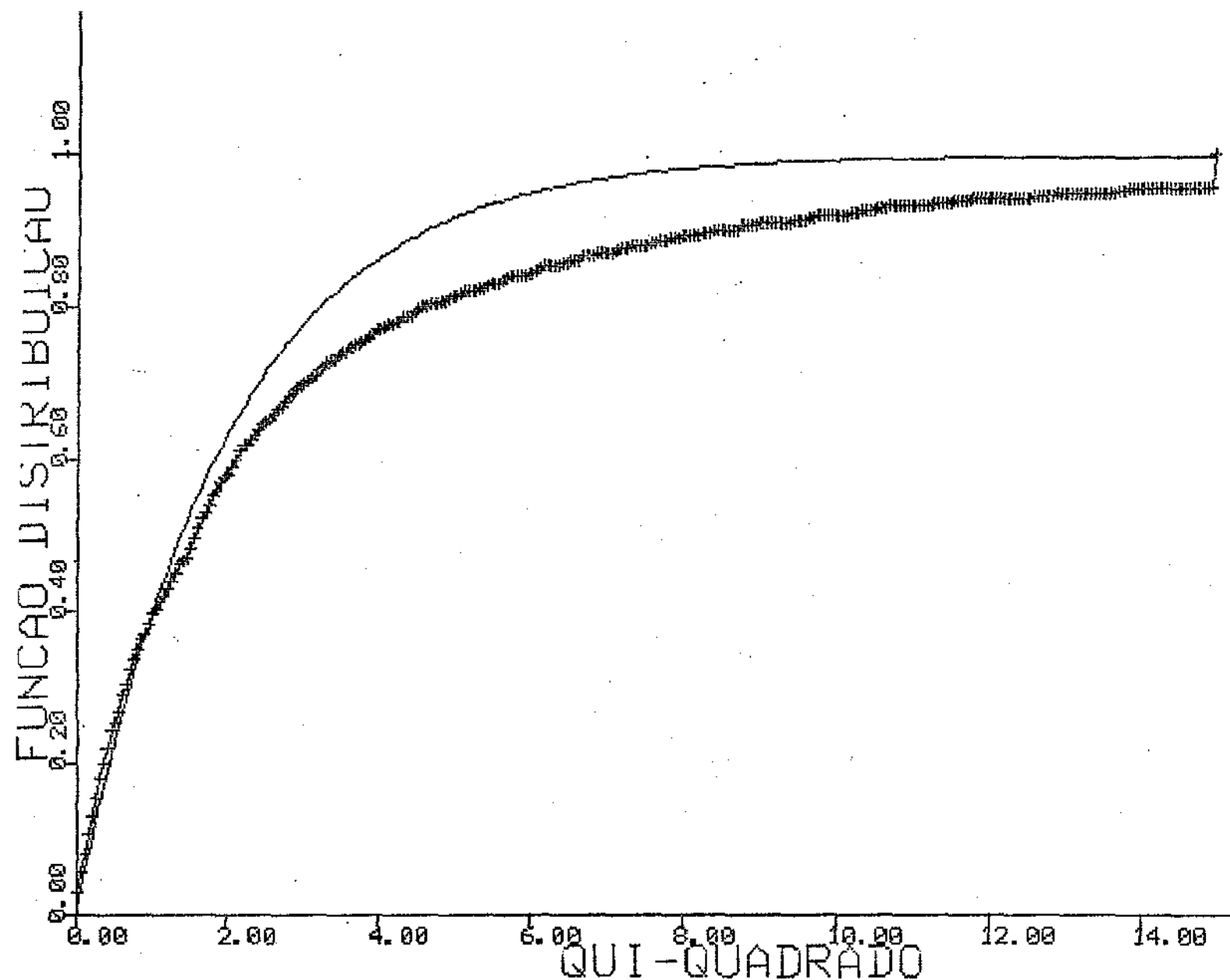


Figura 4.24 Comparação da distribuição exata da Estatística de Pearson corrigida pelo FCB, estimada com base em 1000 repetições MC, com a distribuição Qui-Quadrado com 2 graus de liberdade, com T.C. = 20 e nca = 5, onde a curva de linha fina é a distribuição teórica e a curva de linha grossa é a distribuição exata da Estatística de Pearson. Este gráfico deve ser comparado com a figura 2.15.



4.5 - COMENTARIOS E CONCLUSOES

Como pode-se observar nas tabelas e gráficos apresentados, a estatística corrigida parece apresentar um melhor desempenho.

No caso de conglomerados homogêneos, a correção estimada por reamostragem "Bootstrap" apresenta uma performance equivalente à correção de Rao e Scott. Também a correção estimada por máxima verossimilhança apresenta o mesmo desempenho.

No caso de conglomerados semelhantes à população, o desempenho da correção estimada por "Bootstrap" já não é tão bom como no caso anterior. A correção estimada por máxima verossimilhança parece apresentar um melhor desempenho para este caso.

A vantagem do método de reamostragem "Bootstrap", é que não se precisa de uma expressão formal para encontrar a matriz de covariância V_h , sob o delineamento amostral empregado, tornando-se assim, uma alternativa de fácil cálculo, desde que um computador seja disponível; enquanto que, a alternativa por máxima verossimilhança, apresenta o problema de que se deve ter disponível uma expressão completa para a matriz de covariância, a qual, sob delineamentos amostrais mais complexos, nem sempre é conhecida.

APENDICE

Aqui será dada alguma informação adicional para o esclarecimento de certos pontos.

A1 - COMPOSIÇÃO DOS DADOS

Para inserir os dados no programa de simulação foram criados dois vetores IY , IX os quais indicavam as variáveis: SEXO e IDADE.

Para a composição espacial dos dados, temos duas situações:

i) Configuração de conglomerados homogêneos

A estrutura de cada vetor foi a seguinte :

IY(i) = 1	,	IX(i) = 1	; i = 1, 2, ... , 30
IY(i) = 1	,	IX(i) = 2	; i = 31, 32, ... , 50
IY(i) = 1	,	IX(i) = 3	; i = 51, 52, ... , 80
IY(i) = 2	,	IX(i) = 1	; i = 81, 82, ... , 90
IY(i) = 2	,	IX(i) = 2	; i = 91, 92, ... , 130
IY(i) = 2	,	IX(i) = 3	; i = 131, 132, ... , 160
IY(i) = 1	,	IX(i) = 1	; i = 161, 162, ... , 170
IY(i) = 1	,	IX(i) = 2	; i = 171, 172, ... , 250
IY(i) = 1	,	IX(i) = 3	; i = 251, 252, ... , 280
IY(i) = 2	,	IX(i) = 1	; i = 281, 282, ... , 350
IY(i) = 2	,	IX(i) = 2	; i = 351, 352, ... , 510
IY(i) = 2	,	IX(i) = 3	; i = 511, 512, ... , 600

Assim, pode-se observar que foram criados segmentos de população

homogêneos até um tamanho de 10, e o programa de simulação, no momento de criar os conglomerados, os cria sequencialmente por tamanho do conglomerado, homogeneamente.

ii) Configuração com conglomerados semelhantes aos da população

Para este caso, os dados da tabela 1 podem ser também apresentados da seguinte forma:

		I D A D E			TOTAL
		1	2	3	
S E X O	1	2	5	3	10
	2	4	10	6	20
	TOTAL	6	15	9	30

O programa cria a estrutura de cada vetor IY e IX como segue:

$IY(i) = 1$, $IY(i) = 1$; $i = 1, 2$
 $IY(i) = 2$, $IY(i) = 1$; $i = 3, 4, \dots, 7$
 $IY(i) = 3$, $IY(i) = 1$; $i = 8, 9, 10$
 $IY(i) = 1$, $IY(i) = 2$; $i = 11, 12, \dots, 14$
 $IY(i) = 2$, $IY(i) = 2$; $i = 15, 16, \dots, 24$
 $IY(i) = 3$, $IY(i) = 2$; $i = 25, 26, \dots, 30$

Obtendo-se assim um segmento que representa a estrutura de toda a população, este segmento chamaremos de Mini-População. Repetindo este segmento 20 vezes se poderá obter os dados da população toda.

A2 - PROVA DA MATRIZ G

Na seção 4.2 temos o seguinte:

$$G = \frac{d \tilde{h}(\tilde{p})}{d \tilde{p}} = \left\{ \begin{array}{ll} \frac{d h_{ij}(\tilde{p})}{d p_{ab}} = 1 - (p_{i+} + p_{+j}) & \text{se } a=i, b=j \\ \frac{d h_{ij}(\tilde{p})}{d p_{ab}} = -p_{+j} & \text{se } a=i, b \neq j \\ \frac{d h_{ij}(\tilde{p})}{d p_{ab}} = -p_{i+} & \text{se } a \neq i, b=j \\ \frac{d h_{ij}(\tilde{p})}{d p_{ab}} = 0 & \text{se } a \neq i, b \neq j \end{array} \right. \quad (a.1)$$

Prova

A hipótese de interesse é :

$$H_0 : h_{ij} = p_{ij} - p_{i+} \cdot p_{+j} = 0 \quad \begin{array}{l} i=1,2,\dots,r-1 \\ j=1,2,\dots,c-1 \end{array}$$

então o vetor $\tilde{h}(\tilde{p})$ é da seguinte forma:

$$\tilde{h}(p) = \begin{bmatrix} h_{11} \\ h_{12} \\ \vdots \\ h_{(r-1)(c-1)} \end{bmatrix} = \begin{bmatrix} p_{11} - p_{1+} \cdot p_{+1} \\ p_{12} - p_{1+} \cdot p_{+2} \\ \vdots \\ p_{(r-1)(c-1)} - p_{(r-1)+} \cdot p_{+(c-1)} \end{bmatrix} \quad (a.2)$$

Logo, para encontrar a matriz G,

$$G = \frac{d \tilde{h}(p)}{d p} = \left[\frac{d h_{ij}(p)}{d p_{ab}} \right] =$$

temos os seguintes casos :

i) Se $a=i$ e $b=j$

$$\frac{d h_{ij}(p)}{d p_{ab}} = \frac{d}{d p_{ij}} \left[p_{ij} - p_{i+} \cdot p_{+j} \right] =$$

$$= \frac{d}{d p_{ij}} \left[p_{ij} - \left[\sum_{j=1}^r p_{ij} \right] \cdot \left[\sum_{i=1}^c p_{ij} \right] \right] =$$

$$= 1 - \left[\left[\frac{d}{d p_{ij}} \sum_{j=1}^r p_{ij} \right] \cdot \sum_{i=1}^c p_{ij} + \sum_{j=1}^r p_{ij} \cdot \left[\frac{d}{d p_{ij}} \sum_{i=1}^c p_{ij} \right] \right] =$$

$$= 1 - \left[1 \cdot \sum_{i=1}^c p_{iJ} + \sum_{j=1}^r p_{iJ} \cdot 1 \right] = 1 - (p_{i+} + p_{+J})$$

Assim :

$$\frac{d h_{iJ}^{(p)}}{d p_{iJ}} = 1 - (p_{i+} + p_{+J}) \quad (a.3)$$

ii) Se $a \neq i$ ou $b \neq j$

$$\begin{aligned} \frac{d h_{iJ}^{(p)}}{d p_{ab}} &= \frac{d}{d p_{ab}} \left[p_{iJ} - p_{i+} \cdot p_{+J} \right] = \\ &= \frac{d}{d p_{ab}} \left[p_{iJ} - \left[\sum_{j=1}^c p_{iJ} \right] \cdot \left[\sum_{i=1}^r p_{iJ} \right] \right] = \\ &= 0 - \left[\left[\frac{d}{d p_{ab}} \sum_{j=1}^c p_{iJ} \right] \cdot \sum_{i=1}^r p_{iJ} + \sum_{j=1}^c p_{iJ} \cdot \left[\frac{d}{d p_{ab}} \sum_{i=1}^r p_{iJ} \right] \right] = \end{aligned}$$

Aqui temos os seguintes casos :

a) se $a=i$ e $b \neq j$

$$\frac{d h_{iJ}^{(p)}}{d p_{ab}} = \frac{d h_{iJ}^{(p)}}{d p_{ib}} =$$

$$\begin{aligned}
 &= 0 - \left[\left[\frac{d}{d p_{ib}} \sum_{j=1}^c p_{ij} \right] \cdot \sum_{i=1}^r p_{ij} + \sum_{j=1}^c p_{ij} \cdot \left[\frac{d}{d p_{ib}} \sum_{i=1}^r p_{ij} \right] \right] = \\
 &= - \left[1 \cdot \sum_{i=1}^r p_{ij} + \sum_{j=1}^c p_{ij} \cdot 0 \right] = - \sum_{i=1}^r p_{ij}
 \end{aligned}$$

Assim :

$$\frac{d h_{ij}^{(p)}}{d p_{ib}} = -p_{+j} \quad (a.4)$$

b) se $a=i$ e $b \neq j$

$$\begin{aligned}
 \frac{d h_{ij}^{(p)}}{d p_{ab}} &= \frac{d h_{ij}^{(p)}}{d p_{aj}} = \\
 &= 0 - \left[\left[\frac{d}{d p_{aj}} \sum_{j=1}^c p_{ij} \right] \cdot \sum_{i=1}^r p_{ij} + \sum_{j=1}^c p_{ij} \cdot \left[\frac{d}{d p_{aj}} \sum_{i=1}^r p_{ij} \right] \right] = \\
 &= - \left[0 \cdot \sum_{i=1}^r p_{ij} + \sum_{j=1}^c p_{ij} \cdot 1 \right] = - \sum_{j=1}^c p_{ij} =
 \end{aligned}$$

Assim :

$$\frac{d h_{ij}^{(p)}}{d p_{aj}} = -p_{i+} \quad (a.5)$$

c) se $a \neq 1$ e $b \neq j$

$$\begin{aligned} \frac{d h_{ij}^{(p)}}{d p_{ab}} &= \\ &= 0 - \left[\left[\frac{d}{d p_{ab}} \sum_{j=1}^c p_{ij} \right] \sum_{i=1}^r p_{ij} + \sum_{j=1}^c p_{ij} \cdot \left[\frac{d}{d p_{ab}} \sum_{i=1}^r p_{ij} \right] \right] = \\ &= 0 - \left[0 \cdot \sum_{i=1}^r p_{ij} + \sum_{j=1}^c p_{ij} \cdot 0 \right] = 0 \end{aligned}$$

Assim :

$$\frac{d h_{ij}^{(p)}}{d p_{ab}} = 0 \quad (a.6)$$

Portanto desde (a.3) até (a.6) segue-se a matriz G.

A3 - O PROGRAMA

Aqui apresentamos o programa criado para simular o delineamento amostral por conglomerado. Para os outros delneamentos se utilizou variações deste programa.

```

C
C  PROGRAMA PARA SIMULACAO POR CONGLOMERADOS
C
  DIMENSION IXE(600),IYE(600),NPOP(2,3),NTAB(2,3),MX(3),MY(2)
  DIMENSION PMULTI(5,5),SOMAVAR(5,5),PMULTIE(5,5)
  DIMENSION H(2,5),HT(5,2),HE(2,5),HTE(5,2)
  DIMENSION NFX(300),NFXCT(300),NFXCE(300)
  DIMENSION FX(300),FXCT(300),FXCE(300)
  DIMENSION HVE(2,1),HVET(1,2),VARCOH(2,2),VARINV(2,2),RR(2,1)
  DIMENSION VARC(5,5),VARBA(2,2),NFXW(300),FXW(300)
  DIMENSION NBOTS(300,2,3),VARHVE(2,2),VARHVI(2,2),RRR(2,1)
  DIMENSION NFXB(300),FXB(300),HPH(2,2),DE(2)
  DOUBLE PRECISION X,PX,XCT,XCE,XW,XB,SOMAX,SOMAXCT,SOMAX2,SOMAXCT2
  DOUBLE PRECISION XMEDIO,XCTMEDIO,XCEMEDIO,XWMEDIO,XBMEDIO
  DOUBLE PRECISION SOMAXCE,SOMAXW,SOMAXB,SOMAXCE2,SOMAXW2,SOMAXB2
  DOUBLE PRECISION VARX,VARXCT,VARXCE,VARXW,VARXB,HVE,HVET,RR,VARINV
  DOUBLE PRECISION RNTAES,RES,PMULTI,SOMAVAR,HE,HTE,FCE,VARCOH,HPH
  DOUBLE PRECISION DE,VARHVE,FLB,FCT,BB,FCB,FLA,VARBA,PMULTIE,VARC

C
C  AMOSTRAGEM ALEATORIA POR CONGLOMERADOS COM REPOSICAO
C
  TYPE *, ' ENTRE COM DOIS SEMENTES'
  ACCEPT *, IS1,IS2
  NN=600
  TYPE *, ' PARA A GERACAO DOS DADOS TEMOS'
  TYPE *, ' CONGLOMERADOS HOMOGENEOS : TECLE 11'
  TYPE *, ' CONGLOMERADOS-MINIPOPULACAO : TECLE 22'
10  TYPE *, ' TECLE 11 OU 22 PARA CRIAR A POPULACAO'
  ACCEPT *,ICRIA
  IF (ICRIA.EQ.11) THEN
      TYPE *, 'CONGLOMERADOS HOMOGENEOS'
      CALL HOMOGNEO(ICRIA,IXE,IYE)
  ELSE IF (ICRIA.EQ.22) THEN
      TYPE *, ' CONGLOMERADOS-MINIPOPULACAO'
      CALL MINIPULACAO(ICRIA,IS1,IS2,IXE,IYE)
  ELSE
      GO TO 10
  END IF
  TYPE *, ' '
  TYPE *, ' ENTRE O TAMANHO DO CONGLOMERADO (SUBMULTIPLO DE 30)'
  ACCEPT *,M
  TYPE *, ' '
  TYPE *, ' QUANTOS CONGLOMERADOS SERAO AMOSTRADOS'
  ACCEPT *,NCA

C
C  TAMANHO DA AMOSTRA TOTAL 'NA'
C  NUMERO DE CONGLOMERADOS NA POPULACAO 'NCP'
C
  NCP=NN/M
  NA=M*NCA

C
C  MATRIZ DE DADOS DA POPULACAO NPOP(I,J)
C
  NPOP(1,1)=40

```

```
NPOP(1,2)=100
NPOP(1,3)=60
NPOP(2,1)=80
NPOP(2,2)=200
NPOP(2,3)=120
```

```
C
C  CALCULANDO A VARIANCIA POPULACIONAL P-MULTINOMIAL
C  A VARIANCIA POPULACIONAL POR CONGLOMERADOS
C  A MATRIZ DEFF E OS FATORES DE CORRECAO QUANDO SE
C  CONHECE A POPULACAO.
```

```
C  CALL VARMULTINOMIAL(NPOP,NN,H,HT,PMULTI)
```

```
C  INDI=1
C  CALL VARIANCONGLO(M,IS1,IS2,IXE,IYE,NCP,NCA,INDI,NTAB,VARC,
C  INBOTS)
```

```
C  CALL FATORES(PMULTI,SOMAVAR,M,NA,NCP,H,HT,INN,FC,VARBA,HPH)
C  IF (INN.EQ.0) GO TO 180
C  FCT=FC
```

```
C  ENTRANDO ALGUNS PARAMETROS DA SIMULACAO
```

```
SOMAX=0.0
SOMAXCT=0.0
SOMAXCE=0.0
SOMAXW=0.0
SOMAXB=0.0
SOMAX2=0.0
SOMAXCT2=0.0
SOMAXCE2=0.0
SOMAXW2=0.0
SOMAXB2=0.0
DO I=1,300
  NFX(I)=0
  NFXCT(I)=0
  NFXCE(I)=0
  NFXW(I)=0
  NFXB(I)=0
  FX(I)=0.0
  FXCT(I)=0.0
  FXCE(I)=0.0
  FXW(I)=0.0
  FXB(I)=0.0
```

```
END DO
```

```
C  TYPE *, 'QUANTAS REPETICOES BOOTSTRAP'
C  ACCEPT *,NRB
C  TYPE *, 'QUANTAS REPETICOES DE MONTECARLO'
C  ACCEPT *,NMC
```

```
C  COMECANDO A SIMULACAO MONTECARLO
```

```
C  DO 10000 IMC=1,NMC
```

```
C  ZERANDO A TABELA DE CONTINGENCIA E AS MARGINAIS
```

```
90 DO 100 I=1,2
  MY(I)=0
  DO 100 J=1,3
```

```

      MX(J)=0
100  CONTINUE
C
C      PROCESSA-SE A AMOSTRAGEM E CONSTROE-SE A TABELA
C      DE CONTINGENCIA
C      NCA E O NUMERO DE CONGLOMERADOS NA AMOSTRA
C      M E O TAMANHO DE CADA CONGLOMERADO
C      TAMBEM SE ESTIMA A VARIANCIA DE MAXIMA VEROSSIMILHANCA
C      PARA AMOSTRAGEM POR CONGLOMERADOS
C
      INDI=2
      CALL VARIANCONGLO(M,IS1,IS2,IXE,IYE,NCP,NCA,INDI,NTAB,VARC,
INBOTS)
C
C      CALCULANDO A ESTATISTICA QUI-QUADRADO DE PEARSON
C      PARA A TABELA DE CONTINGENCIA NTAB(I,J)
C
      DO 110 I=1,2
        DO 110 J=1,3
          MX(J)=MX(J)+NTAB(I,J)
          MY(I)=MY(I)+NTAB(I,J)
110  CONTINUE
C
C      CALCULO DO VETOR H DA HIPOTESE
C
      DO I=1,2
        HVE(I,1)=(1.0*NTAB(1,I))/(1.0*NA)-(1.0*MY(1)*MX(I))/(1.0*NA*NA)
        HVET(1,I)=HVE(I,1)
      END DO
      X=0.0
      XCT=0.0
      XCE=0.0
      XW=0.0
      XB=0.0
      DO 120 I=1,2
        DO 120 J=1,3
C
C      CALCULO DO VALOR ESPERADO NA POSICAO (I,J)
C
          RNTAES=(1.0*MY(I)*MX(J))/(1.0*NA)
          IF (RNTAES)115,120,115
115      RES=1.0*NTAB(I,J)-RNTAES
          PX=RES*RES/RNTAES
          X=X+PX
120  CONTINUE
          IF(X.LT.0.0) GO TO 90
C
C      CALCULANDO A VARIANCIA ESTIMADA POR MAXIMA VEROSSIMILHANCA
C      P-MULTINOMIAL PARA A.A.S.-C.R.
C
          CALL VARMULTINOMIAL(NTAB,NA,HE,HTE,PMULTIE)
C
C      CALCULANDO O FATOR DE CORRECAO ESTIMADO POR MAXIMA
C      VEROSSIMILHANCA
C
          CALL FATORES(PMULTIE,VARC,M,NA,NCP,HE,HTE,INN,FCE,VARCOH,HPH)
          IF (INN.EQ.0) GO TO 90
C
C      CALCULANDO A ESTATISTICA DE PEARSON CORRIGIDA
C

```

XCT=X/FCT
XCE=X/FCE
FLA=FLA+FCE

C
C
C
CALCULANDO A ESTATISTICA DE WALD PELO METODO DE LINEARIZACAO

IP=2
CALL HINVERSA(VARCOH,IP,INV,VARINV)
IF (INV.EQ.0) THEN
 XW=0.0
ELSE
 DO I=1,2
 RR(I,1)=0.0
 DO K=1,2
 RR(I,1)=RR(I,1)+VARINV(I,K)*HVE(K,1)
 END DO
 END DO
 DO J=1,2
 XW=XW+NA*HVET(1,J)*RR(J,1)
 END DO
 IF(XW.LT.0) GO TO 90
 IF(XW.GT.1600) GO TO 90
END IF

C
C
C
CALCULANDO A VARIANCIA BOOTSTRAP DO VETOR H DA HIPOTESE

CALL VARBOTS(NBOTS,IS1,IS2,NCA,NA,NRB,VARHVE)

C
C
C
AGORA CALCULO DO FATOR DE CORRECAO COM A VARIANCIA BOOTSTRAP
PRIMEIRO ACHAMOS A MATRIZ DEFF=DE(I,J)

DO 910 I=1,2
 DO 910 J=1,2
 DE(I,J)=0.0
 DO 910 K=1,2
 DE(I,J)=DE(I,J)+1.0*NA*HPH(I,K)*VARHVE(K,J)
910 CONTINUE

C
C
C
CALCULANDO OS VALORES DE DE(I,J) E ACHANDO A MEDIOA DELES E
IGUAL A ACHAR O TRASO DA MATRIZ=SOMA DA DIAGONAL DE(I,J)

BB=0.0
FCB=0.0
BB=DE(1,1)+DE(2,2)
FCB=BB/2.0
IF(FCB.EQ.0) GO TO 90
XB=X/FCB
FLB=FLB+FCB

C
C
C
ACUMULANDO O VALOR DE X,XCT,XCE,XW,XB

SOMAX=SOMAX+X
SOMAXCT=SOMAXCT+XCT
SOMAXCE=SOMAXCE+XCE
SOMAXW=SOMAXW+XW
SOMAXB=SOMAXB+XB
SOMAX2=SOMAX2+X*X
SOMAXCT2=SOMAXCT2+XCT*XCT
SOMAXCE2=SOMAXCE2+XCE*XCE
SOMAXW2=SOMAXW2+XW*XW

SOMAXB2=SOMAXB2+XB*XB

CONSTRUINDO A TABELA DE FREQUENCIA DA QUI-QUADRADO

```
TYPE *, 'IMC=', IMC
LX=INT(20*X)+1
IF (LX.GT.300) LX=300
  NFX(LX)=NFX(LX)+1
LXCT=INT(20*XCT)+1
  IF (LXCT.GT.300) LXCT=300
    NFXCT(LXCT)=NFXCT(LXCT)+1
LXCE=INT(20*XCE)+1
  IF (LXCE.GT.300) LXCE=300
    NFXCE(LXCE)=NFXCE(LXCE)+1
LXW=INT(20*XW)+1
  IF (LXW.GT.300) LXW=300
    NFXW(LXW)=NFXW(LXW)+1
LXB=INT(20*XB)+1
  IF (LXB.GT.300) LXB=300
    NFXB(LXB)=NFXB(LXB)+1
```

10000 CONTINUE

CONSTRUINDO A TABELA DA ACUMULADA

```
FX(1)=(1.0*NFX(1))/(1.0*NMC)
FXCT(1)=(1.0*NFXCT(1))/(1.0*NMC)
FXCE(1)=(1.0*NFXCE(1))/(1.0*NMC)
FXW(1)=(1.0*NFXW(1))/(1.0*NMC)
FXB(1)=(1.0*NFXB(1))/(1.0*NMC)
DO 150 I=2,300
  NFX(I)=NFX(I-1)+NFX(I)
  NFXCT(I)=NFXCT(I-1)+NFXCT(I)
  NFXCE(I)=NFXCE(I-1)+NFXCE(I)
  NFXW(I)=NFXW(I-1)+NFXW(I)
  NFXB(I)=NFXB(I-1)+NFXB(I)
  FX(I)=(1.0*NFX(I))/(1.0*NMC)
  FXCT(I)=(1.0*NFXCT(I))/(1.0*NMC)
  FXCE(I)=(1.0*NFXCE(I))/(1.0*NMC)
  FXW(I)=(1.0*NFXW(I))/(1.0*NMC)
  FXB(I)=(1.0*NFXB(I))/(1.0*NMC)
```

150

CONTINUE

```
FCEMEDIO=FLA/(1.0*NMC)
FCBMEDIO=FLB/(1.0*NMC)
XMEDIO=SOMAX/(1.0*NMC)
XCTMEDIO=SOMAXCT/(1.0*NMC)
XCEMEDIO=SOMAXCE/(1.0*NMC)
XWMEDIO=SOMAXW/(1.0*NMC)
XBMEDIO=SOMAXB/(1.0*NMC)
VARX=(SOMAX2-NMC*XMEDIO*XMEDIO)/(1.0*(NMC-1))
VARXCT=(SOMAXCT2-NMC*XCTMEDIO*XCTMEDIO)/(1.0*(NMC-1))
VARXCE=(SOMAXCE2-NMC*XCEMEDIO*XCEMEDIO)/(1.0*(NMC-1))
VARXW=(SOMAXW2-NMC*XWMEDIO*XWMEDIO)/(1.0*(NMC-1))
VARXB=(SOMAXB2-NMC*XBMEDIO*XBMEDIO)/(1.0*(NMC-1))
```

TYPE 155,ICRIA,M,NCA,NRB,NMC,FCT,FCEMEDIO,FCBMEDIO

WRITE(31,155)ICRIA,M,NCA,NRB,NMC,FCT,FCEMEDIO,FCBMEDIO

155

```
FORMAT(/,2X,'ICRIA=',I2,5X,'M=',I2,5X,'NCA=',I3,5X,'NRB=',I5,
15X,'NMC=',I5,/,2X,'FCT=',F10.6,10X,'FCE. MEDIO=',F10.6,10X,
2'FCB. MEDIO=',F12.6)
```

WRITE(31,160)XMEDIO,XCTMEDIO,XCEMEDIO,XWMEDIO,XBMEDIO

TYPE 160,XMEDIO,XCTMEDIO,XCEMEDIO,XWMEDIO,XBMEDIO

```

160     FORMAT(/,1X,'XM=',F10.6,2X,'XCTM=',F10.6,2X,'XCEM=',
1F10.6,2X,'XWM=',F10.6,'XBM=',F10.6)
      WRITE(31,170)VARX,VARXCT,VARXCE,VARXW,VARXB
      TYPE 170,VARX,VARXCT,VARXCE,VARXW,VARXB
170     FORMAT(/,1X,'V(X)=',F10.6,2X,'V(XCT)=',F12.6,2X,'V(XCE)=',
1F12.6,2X,'V(XW)=',F14.6,2X,'V(XB)=',F12.6)
      TYPE *, 'TECLE 3 P/GRAVAR RESULT ,C.C. QUALQUER TECLA'
      ACCEPT *,NGRAV
      IF (NGRAV.EQ.3) THEN
        DO I=1,300
          WRITE(32,175)FX(I),FXCT(I),FXCE(I),FXW(I),FXB(I)
175          FORMAT(5(F10.6,2X))
        END DO
      END IF
180     STOP
      END
C
C     SUBROTINA PARA A GERACAO DOS DADOS
C
      SUBROUTINE HOMOGENEO(ICRIAD,IX,IY)
      DIMENSION IX(600),IY(600)
C
C     GERANDO OS DADOS HOMOGENEAMENTE
C
      TYPE *, 'GERANDO OS DADOS'
      DO 3000 I=1,30
        IX(I)=1
3000      IY(I)=1
      DO 3010 I=31,50
        IX(I)=2
3010      IY(I)=1
      DO 3020 I=51,80
        IX(I)=3
3020      IY(I)=1
      DO 3030 I=81,90
        IX(I)=1
3030      IY(I)=2
      DO 3040 I=91,130
        IX(I)=2
3040      IY(I)=2
      DO 3050 I=131,160
        IX(I)=3
3050      IY(I)=2
      DO 3060 I=161,170
        IX(I)=1
3060      IY(I)=1
      DO 3070 I=171,250
        IX(I)=2
3070      IY(I)=1
      DO 3080 I=251,280
        IX(I)=3
3080      IY(I)=1
      DO 3090 I=281,350
        IX(I)=1
3090      IY(I)=2
      DO 3100 I=351,510
        IX(I)=2
3100      IY(I)=2
      DO 3110 I=511,600
        IX(I)=3

```

```

3110      IY(I)=2
C
TYPE *, ' GERAMOS CONGLOMERADOS HOMOGENEOS '
RETURN
END

C
C      SUBROTINA PARA A GERACAO DOS DADOS COMO UMA MINIPOPULACAO
C

SUBROUTINE MINIPOPULACAO(ICRIAD,ISD1,ISD2,IXED,IYED)
DIMENSION IX(30),IY(30),IXA(30),IYA(30)
DIMENSION IXED(600),IYED(600)
ICRIAD=22
TYPE *, ' GERANDO OS DADOS '
DO 4000 I=1,2
    IX(I)=1
    IY(I)=1
4000 DO 4010 I=3,7
    IX(I)=2
    IY(I)=1
4010 DO 4020 I=8,10
    IX(I)=3
    IY(I)=1
4020 DO 4030 I=11,14
    IX(I)=1
    IY(I)=2
4030 DO 4040 I=15,24
    IX(I)=2
    IY(I)=2
4040 DO 4050 I=25,30
    IX(I)=3
    IY(I)=2
4050 TYPE *, ' '
TYPE *, ' GERAMOS UMA MINIPOPULACAO DE TAMANHO 30 C/U '
TYPE *, ' '
4060 TYPE *, ' TECLE 1 P/EMBARALHAR O CC. '
ACCEPT *,INE
IF (INE.EQ.1) THEN
    TYPE *, ' EMBARALHANDO A MINIPOPULACAO '
    DO IEMB=1,20
        CALL EMBARALHA(IX,IY,ISD1,ISD2,IXA,IYA)
        KK=30*(IEMB-1)
        DO L=1,30
            IXED(L+KK)=IXA(L)
            IYED(L+KK)=IYA(L)
        END DO
    END DO
ELSE IF (INE.NE.0) THEN
    GO TO 4060
ELSE
    TYPE *, ' NAO EMBARALHA A MINIPOPULACAO '
    DO ICO=1,20
        K=30*(ICO-1)
        DO L=1,30
            IXED(L+K)=IX(L)
            IYED(L+K)=IY(L)
        END DO
    END DO
END IF
RETURN
END

```

C
C SUBROTINA QUE EMBARALHA OS DADOS EM CADA MINIPOPULACAO
C

SUBROUTINE EMBARALHA(IXD,IYD,ISS1,ISS2,IXAD,IYAD)
DIMENSION IXD(30),IYD(30),IXAD(30),IYAD(30),INDICA(30)
DO 5000 J=1,30
INDICA(J)=0
5000 CONTINUE
DO 5030 J=1,30
5010 CALL RANDU(ISS1,ISS2,U)
I=INT(30*U)+1
IF (INDICA(I)-1)5020,5010,5020
5020 INDICA(I)=1
IXAD(J)=IXD(I)
IYAD(J)=IYD(I)
5030 CONTINUE
RETURN
END

C
C SUBROTINA PARA CALCULAR A VARIANCIA P-MULTINOMIAL
C

SUBROUTINE VARMULTINOMIAL(NTABE,NND,HD,HTD,PMUL)
DIMENSION NTABE(2,3),PROMY(2),PROMX(3),PROPOR(2,3)
DIMENSION VETPROP(6),HD(2,5),HTD(5,2),PMUL(5,5)
DIMENSION PP(5,1),PPT(1,5),PDIAG(5,5),RE(5,5)
DOUBLE PRECISION PROPOR,PROMX,PROMY,VETPROP,HD,HTD,PP,PPT
DOUBLE PRECISION PDIAG,RE,PMUL

C
C CALCULANDO AS PROPORCOES DA TABELA NTABE(I,J)
C --AS MARGINAIS E O VETOR DE PROPORCOES
C

K=0
DO 1000 I=1,2
1000 PROMY(I)=0.0
DO 1010 J=1,3
1010 PROMX(J)=0.0
DO 1200 I=1,2
DO 1200 J=1,3
PROPOR(I,J)=(1.0*NTABE(I,J))/(1.0*NND)
PROMX(J)=PROMX(J)+PROPOR(I,J)
PROMY(I)=PROMY(I)+PROPOR(I,J)
K=K+1
VETPROP(K)=PROPOR(I,J)

1200 CONTINUE

C
C CALCULANDO A MATRIZ H(I,J)
C

HD(1,1)=1-(PROMY(1)+PROMX(1))
HD(1,2)=-PROMX(1)
HD(1,3)=-PROMX(1)
HD(1,4)=-PROMY(1)
HD(1,5)=0.0
HD(2,1)=-PROMX(2)
HD(2,2)=1-(PROMY(1)+PROMX(2))
HD(2,3)=-PROMX(2)
HD(2,4)=0.0
HD(2,5)=-PROMY(1)

C
C CALCULANDO A MATRIZ HT(I,J)
C


```

DO 1300 I=1,2
  DO 1300 J=1,5
    HTD(J,I)=HD(I,J)
1300  CONTINUE
C
C  CALCULANDO O VETOR DE PROPORCOES DE DIMENSION MENOR
C  E A TRANSPOSTA E COLOCANDO-A NA DIAGONAL
C
DO 1400 K=1,5
  PP(K,1)=VETPROP(K)
  PPT(1,K)=VETPROP(K)
  PDIAG(K,K)=VETPROP(K)
1400  CONTINUE
C
C  AGORA O PRODUTO DOS VETORES
C
DO 1500 I=1,5
  DO 1500 J=1,5
    RE(I,J)=0.0
    RE(I,J)=RE(I,J)+PP(I,1)*PPT(1,J)
1500  CONTINUE
C
C  CALCULANDO A MATRIZ DE VARIANCIA P-MULTINOMIAL
C
DO 1600 I=1,5
  DO 1600 J=1,5
    PMUL(I,J)=PDIAG(I,J)-RE(I,J)
1600  CONTINUE
  RETURN
END
C
C  SUBROTINA PARA CALCULAR A VARIANCIA POR CONGLOMERADOS
C
SUBROUTINE VARIANCONGLO(MM,ISS1,ISS2,IXXE,IYYE,NCPD,NCAD,INDI,
INTABD,VARCON,NBOT)
  DIMENSION IXXE(600),IYYE(600),NTABD(2,3),SOMVAR(5,5)
  DIMENSION NCON(2,3),PROPCO(2,3),PPCO(6),PPCOM(6),DIF(5)
  DIMENSION VARCON(5,5),NBOT(300,2,3)
  DOUBLE PRECISION PROPCO,PPCO,PPCOM
  DOUBLE PRECISION VARCON,SOMVAR
C
C  INDI=1 POPULACIONAL  INDI=2 AMOSTRAL
C
DO 6000 I=1,5
  DO 6000 J=1,5
    SOMVAR(I,J)=0.0
6000  CONTINUE
    IF (INDI.EQ.1) THEN
      LL=NCPD
    ELSE
      LL=NCAD
    END IF
DO 6002 L=1,300
  DO 6002 K=1,2
    DO 6002 J=1,3
      NBOT(L,K,J)=0
6002  CONTINUE
DO 6005 I=1,2
  DO 6005 J=1,3
    NTABD(I,J)=0

```

```

6005  CONTINUE
C
C  ZERANDO O VETOR DE MEDIA E MATRIZ DE COVARIANCIA
C
DO I=1,5
  DO J=1,5
    SOMVAR(I,J)=0.0
  END DO
END DO

C
C  COMECA A AMOSTRAGEM POR CONGLOMERADOS
C
DO 6100 ICP=1,LL
  DO 6020 I=1,2
    DO 6020 J=1,3
      NCON(I,J)=0
      PROPCO(I,J)=0.0
6020  CONTINUE

      IF (INDI.EQ.1) THEN
        INI=MM*(ICP-1)+1
        FIN=INI+MM-1
      ELSE
        CALL RANDU(ISS1,ISS2,U)
        K1=INT(NCPD*U)+1
        INI=MM*(K1-1)+1
        FIN=INI+MM-1
      END IF
      DO 6030 LK=INI,FIN
        IC=IXXE(LK)
        IL=IYYE(LK)
        NCON(IL,IC)=NCON(IL,IC)+1
6030  CONTINUE

      SOMANDO AS TABELAS AARA CONTRUIR A TABELA TOTAL
      E TAMBEM CONSTRUINDO A AMOSTRA PARA BOOTSTRAP

      DO 6040 I=1,2
        DO 6040 J=1,3
          NTABD(I,J)=NTABD(I,J)+NCON(I,J)
          NBOT(ICP,I,J)=NCON(I,J)
6040  CONTINUE

      CONSTRUINDO O VETOR DE PROPORCOES

      KC=0
      DO 6050 I=1,2
        DO 6050 J=1,3
          PROPCO(I,J)=(1.0*NCON(I,J))/(1.0*MM)
          KC=KC+1
          PPCO(KC)=PROPCO(I,J)
6050  CONTINUE

      SOMA CRUZADA DE QUADRADOS

      DO I=1,5
        DO J=I,5
          SOMVAR(I,J)=SOMVAR(I,J)+PPCO(I)*PPCO(J)
          SOMVAR(J,I)=SOMVAR(I,J)
        END DO

```

```

        END DO
6100  CONTINUE
C
C      VETOR P
C
      KD=0
      DO I=1,2
        DO J=1,3
          KD=KD+1
          PPCOM(KD)=(1.0*NTABD(I,J))/(1.0*MM*LL)
          VARCON(I,J)=0.0
        END DO
      END DO
C
C      MATRIZ DE VARIANCIA POR CONGLOMERADO
C
      DO 6080 I=1,5
        DO 6080 J=I,5
          VARCON(I,J)=(SOMVAR(I,J)-LL*PPCOM(I)*PPCOM(J))/(1.0*(LL-1))
          VARCON(J,I)=VARCON(I,J)
6080  CONTINUE
      RETURN
      END
C
C      SUBROTINA PARA CALCULAR O FATOR DE CORRECAO
C
      SUBROUTINE FATORES(PMULTD,VARCO,MD,NAD,NC,HD,HTD,IN,FCD,ABA,DAR)
      DIMENSION PMULTD(5,5),VARCO(5,5)
      DIMENSION VARPOP(5,5),VARCOPO(5,5)
      DIMENSION RES1(2,5),RES2(2,5),HD(2,5),HTD(5,2)
      DIMENSION ARRI(2,2),ABA(2,2),DAR(2,2),DEF(2,2)
      DOUBLE PRECISION SOMVARD,VARPOP,VARCOPO,PMULTD,HD,HTD
      DOUBLE PRECISION RES1,RES2,ARRI,ABA,DAR,DEF,B,FCD,VARCO
C
C      CALCULANDO A MATRIZ DE VARIANCIA DO VETOR DE PROPORCOES
C      VARPOP=VARIANCIA SOB A.A.S.-C.R.
C      VARCOPO=VARIANCIA SOB A.C.-C.R.
C      VARCON=VARIANCIA DO VETOR DE FREQUENCIAS
      DO 2000 I=1,5
        DO 2000 J=1,5
          VARPOP(I,J)=PMULTD(I,J)/(1.0*NAD)
          VARCO(I,J)=1.0*MD*VARCO(I,J)
          VARCOPO(I,J)=VARCO(I,J)/(1.0*NAD)
2000  CONTINUE
C
C      CALCULANDO OS PRODUTOS DE H(I,J)*P(I,J)*HT(I,J)
C      E DE H(I,J)*V(I,J)*HT(I,J)
C
      DO 2100 I=1,2
        DO 2100 J=1,5
          RES1(I,J)=0
          RES2(I,J)=0
        DO 2100 K=1,5
          RES1(I,J)=RES1(I,J)+HD(I,K)*PMULTD(K,J)
          RES2(I,J)=RES2(I,J)+HD(I,K)*VARCO(K,J)
2100  CONTINUE
      DO 2200 I=1,2
        DO 2200 J=1,2
          ARRI(I,J)=0.0
          ABA(I,J)=0.0

```

DO 2200 K=1,5

ARRI(I,J)=ARRI(I,J)+RES1(I,K)*HTD(K,J)

ABA(I,J)=ABA(I,J)+RES2(I,K)*HTD(K,J)

200 CONTINUE

CALCULANDO A INVERSA DE $H(I,J)*P(I,J)*HT(I,J)$

IP=2

CALL HINVERSA(ARRI,IP,INSD,DAR)

INSD=0 MATRIZ SINGULAR INSD=1 MATRIZ NAO SINGULAR

IF (INSD.EQ.0) THEN

IN=0

GO TO 2800

ELSE

IN=1

END IF

CALCULANDO A MATRIZ $DEFF=(H*P*HT)'*(H*V*HT)$

DO 2300 I=1,2

DO 2300 J=1,2

DEF(I,J)=0.0

DO 2300 K=1,2

DEF(I,J)=DEF(I,J)+DAR(I,K)*ABA(K,J)

300 CONTINUE

CALCULAR OS AUTOVALORES DE DEFF E ACHAR A MEDIA E⁻ IGUAL A

ACHAR O TRASO = SOMA DOS ELEMENTOS DA DIAGONAL E DIVIDIR POR 2

B=0.0

FCD=0.0

B=DEF(1,1)+DEF(2,2)

FCD=B/2.0

IF (FCD.EQ.0) THEN

IN=0

ELSE

IN=1

END IF

800 RETURN

END

SUBROTINA PARA CALCULAR A INVERSA DE UMA MATRIZ

SUBROUTINE HINVERSA(SP,IP,INS,Q)

DIMENSION SP(2,2),Q(2,2)

DOUBLE PRECISION DET,SP,Q

DET=SP(1,1)*SP(2,2)-SP(1,2)*SP(2,1)

IF (DET.LT.0.0) DET=-DET

IF (DET.EQ.0.0) THEN

INS=0

ELSE

INS=1

Q(1,1)=SP(2,2)

Q(1,2)=-SP(1,2)

Q(2,1)=-SP(2,1)

Q(2,2)=SP(1,1)

DO I=1,2

DO J=1,2

C
C
SOMA DE QUADRADOS (PRODUTOS CRUZADOS)

DO 8060 I=1,2
SOMAHV(I)=SOMAHV(I)+HV(I)
DO 8060 J=1,2
SOMACRU(I,J)=SOMACRU(I,J)+HV(I)*HV(J)

8060 CONTINUE

8100 CONTINUE

C
C
C
MEDIA DO VETOR

DO 8120 I=1,2
HVMEDIA(I)=SOMAHV(I)/(1.0*MCB)

8120 CONTINUE

C
C
C
CALCULANDO A MATRIZ DE VARIANCIA DO VETOR HV

DO 8130 I=1,2
DO 8130 J=1,2
RES(I,J)=0.0
RES(I,J)=HVMEDIA(I)*HVMEDIA(J)
VARHV(I,J)=(SOMACRU(I,J)-MCB*RES(I,J))/(1.0*(MCB-1))
VARHV(J,I)=VARHV(I,J)

8130 CONTINUE

RETURN

END

Foi utilizado o computador VAX 785/11 VMS, da UNICAMP. Para os gráficos foi utilizado o PLOTER do Instituto de Física da UNICAMP.

REFERENCIAS BIBLIOGRAFICAS

- BHATTACHARYYA,, G.K. & JOHNSON, R.A. (1977), Statistical Concepts and Methods, John Wiley, New York.
- BICKEL, P.J. & DOXSUM, K.A. (1977), Mathematical Statistics, San Francisco : Holden-Day.
- DEMING , W.E. (1960), Sample Design in Business Research, John Wiley, New York.
- EFRON, B. (1979), Bootstrap Methods : another look of the jackknife, Ann. Statist., 7, 1-26.
- FELLEGI, I.P. (1980), Approximate Tests of Independence and Goodness of Fit Based on Stratified Multistage Samples, Journal of the American Statistical Association, 75, 261-8.
- JOHNSON, N.L. & KOTZ, S. (1970), Continuous Univariate Distributions, Boston : Houghton Mifflin.
- KISH, L. (1965), Survey Sampling, John Wiley, New York.
- NATHAN, G. (1975), Tests of Independence in Contingency Tables from Stratified Proportional Samples, Sankhyā, Ser. C, 37, 77-87.
- RAKTOE, B.L. & HUBERT, J.J. (1979), Basic Applied Statistics, Marcel Dekker, New York.
- RAO, J.N.K. & SCOTT, A.J. (1981), The Analysis of Categorical Data from Complex Sample Surveys : Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables, Journal of the American

Statistical Association, 76, 221-30.

SOLOMON, H. & STEPHENS, M.A. (1977), Distribution of a Sum of Weighted Chi-Square Variables, Journal of American Statistical Association, 72, 881-5.

WILKS, S.S. (1935), The Likelihood Test of Independence in Contingency Tables, Annals of Mathematical Statistics, 6, 190-6.

YULE, G.U. & KENDALL, M.G. (1968), An Introduction to the Theory of Statistics, Hafner Publishing Company, New York.