UM ESTUDO ESTATÍSTICO DE DOIS PROBLEMAS PRÁTICOS

ROSA MARIA INGA SANTIVAÑEZ

Orientador:
Prof. Dr. PUSHPA NARAYAN RATHIE

Dissertação apresentada no Instituto de Matemática, Estatística e Ciência da Computação, como requisito parcial para obtenção do título de Mestre em Estatística.

NOVEMBRO/1983

Aos meus pais.

AGRADECIMENTOS

Desejo apresentar meus sinceros agradecimentos a várias pessoas que contribuíram na elaboração desta tese:

Ao Professor Pushpa Narayan Rathie pela segura e objetiva orientação durante a elaboração deste trabalho e pela grande dedicação e amizade;

Às minhas amigas Irene Ferreira e Regina Moran pelo incentivo e apoio que me prestaram,

E aos meus familiares pela formação e carinho que me deram.

SUMÁRIO

Neste trabalho objetivamos selecionar um conveniente tratamento estatístico para o estudo dos seguintes problemas:

i) Integração dos imigrantes na cidade de São Paulo, e

ii) Modos de obtenção dos alimentos nas aldeias Alantesu e Juina.

Inicialmente apresentamos as metodologias existentes para a análise desses problemas e posteriormente selecionar os métodos mais adequados para serem aplicados.

Os métodos, de distância de Matusita, teste T e testes de Kullback, estes dois últimos utilizando a medida de informação, foram os que apresentaram melhores resultados sobre os outros.

Por conseguinte, o estudo dos dois problemas mencionados são exemplos práticos da Teoria de Informação.

SUMMARY

The present work aims to select a convenient statistical treatment for the study of the following problems:

i) Integration of immigrants in the city of São Paulo, and

ii) Ways of obtaining food in the villages Alantesu and Juina.

In the beginning we present the existant methodologies for analysing these problems and later on we select the proper methods to be applied.

The methods of Matusita's distance, T test and Kullback tests, the later two making use of the information measures, were the methods presenting more advantages over the others.

Thus, the study of above mentioned problems are practical examples of Information Theory.

ÍNDICE

INTRODUÇÃO

Informação, como uma definição tecnicamente aplicável a estatística, foi introduzida pela primeira vez por Fisher, em seu trabalho sobre teoria da estimação, em 1925. Nesse trabalho Fisher definiu uma medida para a quantidade de informação, contida numa amostra sobre um parâmetro desconhecido.

Shannon e Wiener, independentemente, publicaram, em 1948, trabalhos apresentando, uma medida logarítmica de informação para uso em teoria da comunicação. Esses trabalhos estimularam uma grande quantidade de estudos, nos círculos ligados a engenharia, sobre o objetivo de teoria de informação.

Teoria de informação é um ramo da teoria matemática de probabilidade e estatística e como tal, suas formulações abstratas são aplicáveis a qualquer sistema de observação probabilístico ou estatístico. Consequentemente, encontramos a teoria de informação aplicada numa grande variedade de áreas e de modo relevante na inferência estatística. Ela proporciona uma unificação de resultados conhecidos e leva a uma natural generalização e derivação de novos resultados. A teoria de informação é também bastante utilizada por engenheiros de comunicação, psicólogos, físicos e biólogos, sendo que nesse último caso sua aplicação ocorre especialmente na área de Ecologia, pela aplicação do índice de diversidade, ou entropia de Shannon, para o estudo de ecosistemas.

Devido à grande aplicabilidade da Teoria de Informação, fomos motivados a desenvolvê-la no estudo da integração dos imigrantes na cidade de São Paulo e também no estudo dos modos de obtenção dos alimentos em duas aldeias, Alantesu e Juina, ou seja, dois problemas ligados a Demografia e Ecologia respectivamente.

Para isso dividimos nosso trabalho da seguinte forma:

No primeiro capítulo, apresentamos os dois problemas que serão estudados.

No segundo capítulo, fazemos uma breve exposição da metodologia utilizada para a análise dos problemas propostos, além de apresentar outras metodologias existentes que também podem ser utilizadas nesses casos.

O terceiro capítulo contém os resultados obtidos na aplicação da metodologia e finalmente apresentamos as conclusões obtidas utilizando-se a teoria da informação na análise dos problemas.

CAPÍTULO I

APRESENTAÇÃO DOS PROBLEMAS

Neste capítulo, vamos apresentar os dois problemas que serão estudados e discutidos no decorrer deste trabalho.

1.1. INTEGRAÇÃO DOS IMIGRANTES NA CIDADE DE SÃO PAULO

O objetivo de nosso trabalho é estudar o fenômeno da integração entre os grupos de imigrantes na cidade de São Paulo e a integração destes com os brasileiros, no período de 1930 a 1981. Para isso tomamos como variável indicadora da integração: o número de casamentos.

É importante ressaltar que vamos trabalhar apenas com os quatro principais grupos de imigrantes, a saber: Italianos, Espanhóis, Portugueses e Japoneses; e que vamos considerar como imigrante toda pessoa que não se naturalizou brasileira até o momento do casamento. Como brasileiros entendemos as pessoas com nacionalidade brasileira no momento do casamento.

Contamos para nossa pesquisa com dados cedidos pela Fundação Sistema Estadual de Análise de Dados (SEADE), apresentados na tabela 1.1.

Esses dados nos fornecem para cada grupo étnico:

- Número de casamentos dentro do grupo (Endogamia);
- Número de casamentos de homens fora do grupo (Hfg);
- Número de casamentos de mulheres fora do grupo (Mfg).

Numa primeira análise dos dados, observamos que em todos os grupos étnicos ocorre, quase sempre, uma diminuição no número de casamentos nos anos bissextos, e como esse fenômeno ocorre tanto no número de casamentos de homens e mulheres fora do grupo (brasileiro, italiano, espanhol, português, japonês) como no número de casamentos dentro do grupo, decidimos agrupar os dados de 4 em 4 anos (um ano depois e dois antes do ano bissexto) e trabalhar com as médias dos números de casamentos nesses períodos. Isso nos permite além de trabalhar com todos os dados, evitar possíveis perturbações causadas por esse fenômeno.

Os dados assim agrupados estão na tabela 1.2 e através dela estudaremos:

- O comportamento da integração dos homens e mulheres analisado separada e conjuntamente.
- Existência de homogeneidade entre os grupos.
- Existência de homogeneidade entre os grupos, condicionados ao tipo de casamento (endogamia, casamento de homens fora do grupo, casamento de mulheres fora do grupo, casamento fora do grupo).
- Existência de interação e independência entre os fatores: Período (4 anos), Nacionalidade (Brasileiro, Italiano, Espanhol, Português, Japonês) e Tipo de casamento.

Para maiores detalhes sobre as definições e significados de homogeneidade, interação e independência, etc., ver o capítulo 2 e Kullback (1959).

TABELA 1.1

NÚMERO DE CASAMENTOS DE BRASILEIROS, DE IMIGRANTES ITALIANOS, ESPANHOIS, PORTUGUESES E JAPONESES DE ACORDO COM O TIPO DE CASAMENTO (1930-1981) NA CIDADE DE SÃO PAULO.

ANO	BRASILEIROS			ITALIANOS			ESPANHOIS			PORTUGUESES			JAPONESES		
	E	Hfg	Mfg	E	Hfg	Mfg	E	Hfg	Mfg	E	Hfg	Mfg	E	Hfg	Mfg
1981	65.191	412	1.056	5	158	57	6	123	43	86	609	270	63	179	55
1980	63.335	477	1.117	15	168	53	8	145	56	86	664	302	76	146	72
1979	64.289	581	1.297	18	210	79	12	182	64	113	732	361	76	191	96
1978	61.529	607	1.283	4	214	83	12	198	73	140	704	384	75	191	91
1977	60.701	628	1.377	14	214	102	13	188	78	174	796	392	73	199	76
1976	57.710	619	1.359	20	262	98	20	208	91	160	722	371	103	188	80
1975	61.333	787	1.639	30	292	138	21	244	106	252	920	514	91	216	62
1974	57.806	816	1.670	22	347	165	17	235	119	239	949	500	107	172	65
1973	55.584	867	1.760	29	339	178	25	274	113	316	951	547	147	235	69
1972	46.014	678	1.484	42	303	117	35	266	114	305	793	433	131	162	54
1971	44.082	806	1.646	48	332	177	28	236	114	360	923	495	169	190	55
1970	38.723	699	1.513	56	298	138	50	273	136	405	809	437	108	183	38
1969	36.987	646	1.572	43	308	160	48	280	100	441	843	392	120	189	43
1968	31.441	553	1.217	57	268	113	69	197	109	387	662	350	107	147	38
1967	31.310	532	1.515	76	312	143	72	257	81	536	796	340	76	214	32
1966	30.680	577	1.477	86	327	144	95	257	115	557	755	342	88	205	43
1965	29.586	600	1.486	106	319	139	125	240	95	661	721	359	105	252	53
1964	27.708	476	1.404	102	368	132	157	237	72	560	658	296	80	201	36
1963	32.254	589	1.794	165	457	143	160	300	98	700	813	369	99	293	48
1962	29.667	567	1.714	170	461	141	157	292	81	738	766	351	100	254	53
1961	30.854	536	1.827	198	482	120	168	290	79	687	836	349	105	276	45
1960	23.997	370	1.505	186	417	92	154	260	61	557	654	249	96	238	32
1959	28.513	523	1.830	236	517	135	187	270	93	715	830	318	112	286	49
1958	25.982	464	1.618	258	476	115	199	219	61	618	732	271	108	236	62

Fonte: FUNDAÇÃO SEADE

TABELA 1.1 (CONTINUAÇÃO)

ANO	BRASILEIROS			ITALIANOS			ESPANHÓIS			PORTUGUESES			JAPONESES		
	E	Hfg	Mfg	E	Hfg	Mfg	E	Hfg	Mfg	E	Hfg	Mfg	E	Hfg	Mfg
1957	25.340	500	1.478	260	460	129	135	202	60	584	648	296	123	210	57
1956	22.256	341	1.258	228	395	86	117	167	59	483	568	184	112	171	55
1955	26.705	424	1.401	245	459	88	102	171	54	575	621	282	129	194	44
1954	24.426	368	1.141	152	358	88	75	133	56	404	507	209	158	169	41
1953	22.837	347	1.032	110	350	73	48	118	37	261	425	212	165	152	38
1952	18.483	266	824	84	251	51	22	82	48	175	368	147	145	143	40
1951	20.685	315	890	59	228	78	16	116	33	156	455	197	181	116	32
1950	18.324	316	841	45	226	82	15	117	42	129	417	181	131	99	29
1949	17.276	325	935	38	261	74	17	167	62	141	427	194	142	106	21
1948	13.046	254	800	35	218	71	11	137	52	121	392	142	122	79	15
1947	15.990	366	990	25	278	110	17	173	59	138	506	214	123	63	13
1946	13.588	367	897	37	339	104	9	174	83	150	461	194	98	49	12
1945	11.968	360	954	24	249	106	34	161	82	210	531	195	102	45	9
1944	9.874	290	818	37	194	90	24	136	64	163	376	145	71	30	9
1943	10,661	392	884	55	243	125	32	163	106	222	470	185	75	41	9
1942	9.151	359	825	64	246	118	53	187	103	209	405	172	52	27	6
1941	9.048	352	913	47	263	123	47	207	100	214	454	168	35	36	8
1940	7.103	333	758	64	225	114	37	189	124	215	375	143	25	20	3
1939	8.494	382	1.014	94	301	144	104	226	120	255	539	185	22	18	3
1938	7.259	379	1.011	81	398	119	73	227	139	211	515	162	20	13	1
1937	7.152	376	1.109	81	320	142	69	240	120	275	606	180	7	11	2
1936	6.035	356	1.017	60	336	137	85	197	126	258	552	162	17	13	12
1935	6.391	418	1.229	88	406	169	113	264	156	354	660	197	3	3	0
1934	5.260	383	1.142	116	376	150	113	241	133	372	616	195	9	5	1
1933	4.818	379	1.038	118	359	134	126	176	140	462	577	182	10	3	0
1932	3.484	280	798	82	274	110	78	135	107	440	452	128	3	2	0
1931	3.564	292	820	135	303	120	110	151	122	434	446	130	6	1	1
1930	3.148	319	835	128	311	126	120	167	105	482	431	163	3	2	1

Fonte: FUNDAÇÃO SEADE

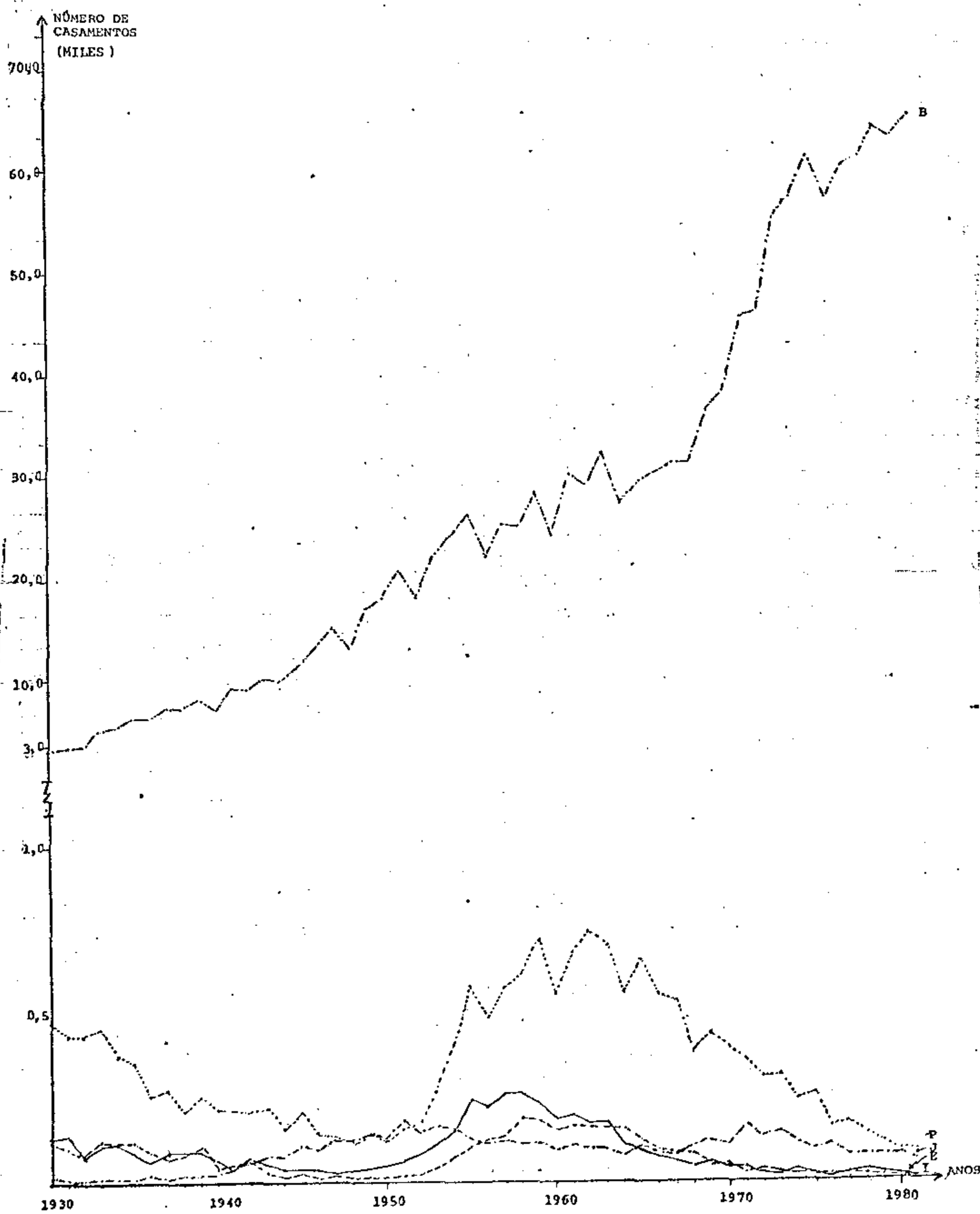

FIGURA 1.1 ENDOGAMIA

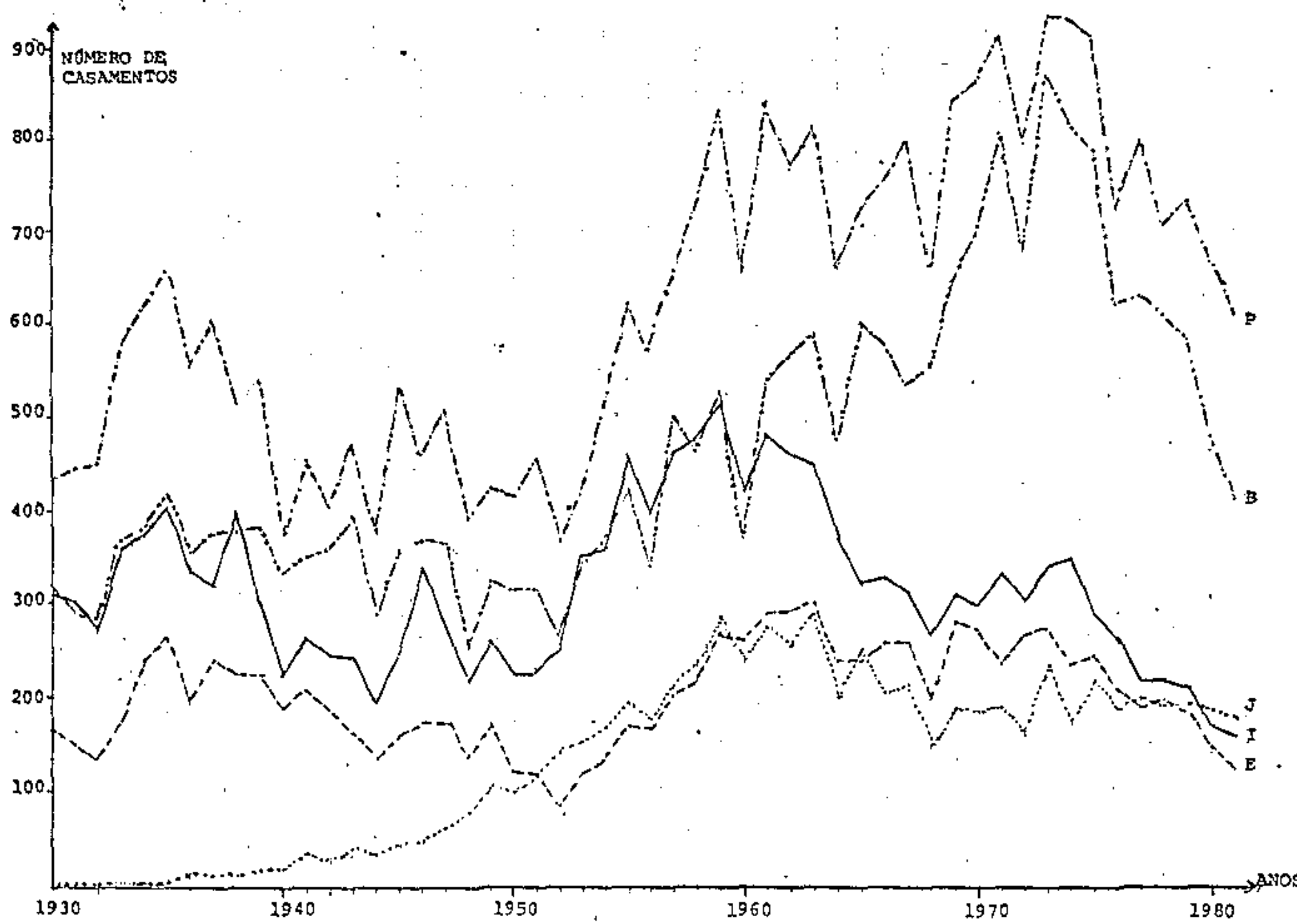

FIGURA 1.2 HOMENS CASADOS FORA DO GRUPO

FIGURA 1.3 MULHERES CASADAS FORA DO GRUPO

TABELA 1.2

NÚMERO DE CASAMENTOS DE BRASILEIROS, DE IMIGRANTES ITALIANOS, ESPANHOIS, PORTUGUESES E JAPONESES DE ACORDO COM O TIPO DE CASAMENTO (1930-1981) NA CIDADE DE SÃO PAULO.

PERIODO	BRASILEIROS			ITALIANOS			ESPANHOIS			PORTUGUESES			JAPONESES		
	E	Hfg	Mfg	E	Hfg	Mfg	E	Hfg	Mfg	E	Hfg	Mfg	E	Hfg	Mfg
78/81	53586,00	519,25	1138,25	10,50	187,50	68,00	9,50	162,00	59,00	106,25	677,25	329,25	72,50	176,75	78,50
74/77	53387,50	712,50	1521,25	21,50	278,75	125,75	17,75	218,75	98,50	206,25	846,75	444,25	93,50	133,75	264,50
70/73	46100,75	762,50	1600,75	43,75	318,00	152,50	34,50	262,25	119,25	346,50	869,00	478,00	138,75	192,50	53,75
66/63	32604,50	577,00	1445,25	65,50	303,75	140,00	71,00	247,75	101,25	480,25	764,00	356,00	97,75	168,75	39,00
62/65	29803,75	558,00	1599,50	135,75	401,25	138,75	149,75	267,25	66,50	664,75	739,50	343,75	96,00	250,00	47,50
58/61	27336,50	473,25	1695,00	219,50	473,00	115,50	177,00	259,75	73,50	644,25	763,00	296,75	105,25	259,00	47,00
54/57	24681,75	408,25	1319,25	221,25	418,00	97,75	107,25	1[illegible],25	57,25	511,50	586,00	242,75	130,50	186,00	49,25
50/53	20032,25	311,00	896,75	74,50	263,75	71,00	25,25	108,25	40,00	180,25	416,25	184,25	155,5	127,50	34,75
46/43	14975,00	328,00	905,50	33,75	274,00	69,75	13,50	162,75	189,00	137,50	446,50	186,00	121,25	74,25	15,25
42/45	10413,50	350,25	845,25	45,00	233,00	109,75	35,75	161,75	89,75	201,00	445,50	174,25	75,00	35,75	6,25
38/41	7376,00	361,50	924,00	71,50	296,75	125,00	65,25	212,25	120,75	223,75	470,75	164,50	25,50	21,75	3,75
34/37	6209,50	383,25	1124,25	86,25	359,50	149,50	95,00	235,50	133,75	314,75	608,50	163,50	9,00	6,00	3,75
30/33	3753,50	317,50	872,75	115,75	311,75	122,50	106,00	157,25	118,50	454,50	476,50	150,75	5,50	2,00	0,50

Fonte: FUNDAÇÃO SEADE.

1.2. ESTUDO DOS MODOS DE OBTENÇÃO DOS ALIMENTOS NAS ALDEIAS ALANTESU (FLORESTA) E JUINA (CERRADO)*

Um segundo objetivo de nosso trabalho é analisar as influências regionais e climáticas sobre os modos de obtenção dos alimentos em duas aldeias, Alantesu e Juina, situadas respectivamente no Vale de Guaporé, zona de floresta, e Chapada de Peris, zona de cerrado, ambas no município de Vila Bela da Santa Trinidade, Mato Grosso.

O período amostrado foi:

TABELA 1.3

ESTAÇÃO \ ALDEIA	ALDEIA ALANTESU	ALDEIA JUINA
Seca	02/08/79 - 05/09/79	16/09/79 - 15/10/79
Chuvosa	16/12/79 - 15/01/79	26/01/79 - 01/03/80

E a frequência dos ítens alimentares por dia numa determinada estação, foi assim convencionada:

- Unitária: quando o item surgiu em pelo menos uma refeição,

(*) Os dados para este estudo foram retirados da dissertação de Mestrado de Eleonore Setz, intitulada: "Ecologia Alimentar em um grupo indígena: Comparação entre aldeias Nambiquara de floresta e de cerrado", apresentada em 1983.

de um indivíduo qualquer, na aldeia ou na caçada.

Nula: quando não ocorreu em nenhuma refeição de nenhum indivíduo na aldeia ou na caçada.

Observemos que a frequência máxima possível para cada item, numa aldeia, é igual ao número de dias observados na estação. Isto é, se na aldeia Alantesu foram feitas observações, durante 27 dias na estação seca, então a frequência máxima para qualquer item alimentar dessa aldeia, nessa estação, é 27. As frequências observadas dos ítens alimentares por aldeia e por estação são dadas na tabela 1.5.

As frequências máximas são:

TABELA 1.4

ESTAÇÃO / ALDEIA	SECA	CHUVOSA
ALANTESU	27	27
JUINA	29	30

As frequências observadas dos ítens alimentares agrupados segundo seu modo de obtenção, por aldeia e por estação, são dadas na tabela 1.6, onde, por convenção, definimos:

- Caçar: sair para procurar, emboscar, perseguir, aprisionar e abater animais terrestres ou aéreos, usando arco e flechas ou espingardas e voltar para a aldeia ou acampamento.

- Coletar: sair para procurar, cavoucar, colher e obter ítens alimentares e medicinais, animais ou vegetais, manualmente ou utilizando machado, pá ou pau de cavar ou de rastrear buraco e transportá-los para a aldeia ou acampamento.

- Pescar: sair para apanhar peixes e tartarugas em corpo d'água usando arco e flechas, linha e anzol ou cesto ou mergulhando, e retornar a aldeia ou ao acampamento.

- Colheitar: sair para roça e colher os alimentos cultivados pela aldeia.

TABELA 1.5

FREQUÊNCIA DOS ITENS ALIMENTARES NAS ALDEIAS ALANTESU (A) E JUINA (J) NAS DUAS ÉPOCAS SECA(s) E CHUVA (c)

COLHIDO-AMILÁCEOS	As	Ac	Js	Jc
Mandioca-mansa	25	22	15	19
Mandioca-brava-beiju			29	30
Batatas-doces	19	18		
Milho	8	24		
Bananas	14	11		
Arroz		19		
Mandioca-mansa-puba	10	1		
Feijão			3	6
Feijão-fava	4	4		
Inhame	6		1	1
Mandioca-mansa-beiju	1	6		
Araruta	1			5
	88	86	48	80
total	174		128	

COLETADOS-AMILÁCEOS	As	Ac	Js	Jc
Cará-do-mato			1	
total	0		1	

COLHIDOS-FRUTOS	As	Ac	Js	Jc
Mamão	10	2		
Melancia		6		
Amendoim	2	1		2
Pepino		5		
	12	14	0	2
total	26		2	

COLETADOS-INSECTOS	As	Ac	Js	Jc
Mel	17	3	14	12
Tucura			6	22
Larvas de marimbondo	1	2	14	2
Corõ Cs			4	
Corõ Cc				4
Larvas	3			
Gafanhoto verde			3	
Corõ Cla				3
Gafanhotões				3
Carregador a		2		
Carregador j				2
Larvas		1		
Percevejo			1	
Corõ de besouro				1
Gafanhoto diferente				1
	21	8	42	50
total	29		92	

COLETADOS-VERTEBRADO	As	Ac	Js	Jc
Teiú	2		10	2
Ovos	3		9	
Tatu-cascudo			5	6
Lagartixo			5	2
Tatu-galinha (7c)			1	6
Tatu-galinha (9c)	2	2		2
Jabuti	3	3		
Pedreira			5	
Tatu (11c)			4	
Calango	1		2	1
Corujas			4	
Pica-pau			4	
Papagaios			3	
Gambá			1	2
Ratinhos			1	2
Lacraia (ave)			1	1
Morcego 1	1			
Passarinho			1	
Morcego 2			1	
Sapo			1	
Lagarto-vermelho			1	
Jibóia			1	
Arara-pequena				1
Paca				1
Rã				1
Ratão				1
	12	5	59	27
total	17		86	

PESCADOS-VERTEBRADOS	As	Ac	Js	Jc
Peixes a	27	18		
Peixes j			5	9
Peixitos			2	
Tartaruga a		1		
Tartaruga b		1		
	27	20	7	9
total	47		16	

TABELA 1.5 (CONTINUAÇÃO)

COLETADOS-FRUTOS	As	Ac	Js	Jc
Coquinho	10	4	2	1
Marmelada		15		
Jatobá	4		10	
Castanha-de-caju			10	
Jabuticaba			9	
Fruto amarelo 1		6		
Caju			6	
Mangaba			4	1
Sadãosu				5
Piqui			4	
Tarauãquisu	3			
Coco-inajá	1	2		
Frutão		3		
Marolo			2	1
Buriti			1	2
Cacau		2		
Bacaba			2	
Abacaxi				2
Dũntequisu	1			
Frutos a	1			
Ingá		1		
Jabuticaba-do-mato			1	
Atĩnquisu			1	
Fruto j			1	
Fruto-lobinho			1	
Dãiriquisu			1	
Cainregãtsu			1	
Caroço de paina			1	
Wãisiquisu				1
Frutinhos j				1
Maracujá				1
	20	33	57	15
total	53		72	

CAÇADOS-VERTEBRADOS	As	Ac	Js	Jc
Porcos	17	8	6	1
Tamanduá-bandeira			9	6
Macaco-prego	5	6		
Carno	1		3	6
Veado j			1	7
Ave	1	1	4	
Macaco-bugio		5		
Macuco	2	2	1	
Tatu-canastra			3	2
Jacu	1		4	
Arara-canindé	2		1	1
Macaco-preto		3		
Paca			2	
Capivara	2			
Tamanduá-mirim		2		
Arara-canga	1	1		
Bugio-caraya			2	
Anta	1			
Parauacu	1			
Papagaio	1			
Bugio-preto		1		
Couro			1	
Ema			1	
Andorinha			1	
Onça-parda			1	
Seriema			1	
Pomba			1	
Gavião-branco			1	
Gavião-preto			1	
Urutau			1	
Crai-crai			1	
Sagüi				1
Cutia				1
Morcego 3				1
	35	34	47	27
total	69		74	

Utilizaremos a tabela 1.6 para estudarmos:

- Existência de homogeneidade entre as duas aldeias.

- Existência de homogeneidade entre as duas aldeias condicionado a estação (seca, chuvosa).

- Existência de interação e independência entre os 3 fatores: Modo de obtenção (caçar, coletar, pescar, colheitar), Aldeia (Alantesu, Juina), Estação (seca, chuvosa).

TABELA 1.6

MODO DE OBTENÇÃO DOS ÍTENS ALIMENTARES	ALANTESU			JUINA		
	SECA	CHUVA	TOTAL ESTAÇÃO	SECA	CHUVA	TOTAL ESTAÇÃO
Colhido	100	100	200	48	82	130
Coletado	53	46	99	159	92	251
Caçado	35	34	69	47	27	74
Pescado	27	20	47	7	9	16

CAPÍTULO II

METODOLOGIA

Faremos aqui, uma breve exposição da metodologia que será utilizada na análise dos problemas propostos. O teste de independência para tabela de contingência utilizando a distância de Matusita, o teste Qui-Quadrado e ANOVA são apresentados aqui, apenas a título de informação e não serão empregados na análise. A justificativa para isso está dada na secção 2.8.

2.1. DISTÂNCIA DE MATUSITA

Consideremos duas distribuições de probabilidade dadas por:

$$P \equiv (p_1,\ldots,p_n), \quad p_i \geq 0, \quad i = 1,\ldots,n, \quad \sum_{i=1}^{n} p_i = 1$$

$$Q \equiv (q_1,\ldots,q_n), \quad q_i \geq 0, \quad i = 1,\ldots,n, \quad \sum_{i=1}^{n} q_i = 1$$

A distância de Matusita entre P e Q é definida por:

$$d_n(P:Q) = [\sum_{i=1}^{n} (p_i^{1/2} - q_i^{1/2})^2]^{1/2}$$

$$= [2(1 - \sum_{i=1}^{n} p_i^{1/2} q_i^{1/2})]^{1/2}$$

E a medida de afinidade de Matusita entre as populações P e Q é dada por:

$$A_n(P:Q) = \sum_{i=1}^{n} (p_i\ q_i)^{1/2}$$

[ver Mathai e Rathie (1971)].

Dentre as propriedades da distância de Matusita, destacamos as seguintes:

P1. Simetria

$d_n(P:Q) = d_n(Q:P)$

P2. Desigualdade

$0 \leq d_n(P:Q) \leq \sqrt{2}$

Observemos que $d_n(P:Q) = 0$ quando $A_n(P:Q) = 1$, ou seja, as duas distribuições coincidem $(P \equiv Q)$ (isto é, $p_i = q_i$; $\forall\ i = 1,\ldots,n$) e $d_n(P:Q) = \sqrt{2}$ quando $A_n(P:Q) = 0$, e isto ocorre quando $p_i q_i = 0$ $\forall\ i = 1,\ldots,n$, ou seja, quando as duas distribuições estão bem distantes.

Para maiores informações a respeito das propriedades da distância de Matusita e sobre outras medidas de informação ver Mathai e Rathie (1975).

2.2. ÍNDICE DE DIVERSIDADE

Uma medida que caracteriza uma coleção de organismos é sua "diversidade". Entendemos que uma coleção não apresenta diversidade quando seus membros pertencem a uma só espécie. Por outro la do, encontramos um índice de diversidade máximo, em coleções que apresentam vários indivíduos em diferentes espécies, ou seja, o índice de diversidade nos indica como estão distribuídas as quantidades observadas, ou indivíduos, nas categorias ou espécies.

A medida de diversidade, dada por Shannon é:

$$H = - \sum_{i=1}^{\hat{s}} p_i \log p_i$$

onde:

s = Número de categorias.

p_i = Probabilidade que um sujeito pertença a i-ésima categoria $(\sum_{i=1}^{s} p_i = 1)$.

com o logarítmo na base 2, 10 ou e.

Essa medida é também conhecida como entropia e, como "diversidade de espécie" quando é aplicada a ecosistemas. (Ver Pielou (1965); Zar (1974); Lloyd, Zar, Karr (1968).)

Um estimador natural de H é dado por:

$$\hat{H} = (n \log n - \sum_{i=1}^{s} f_i \log f_i) / n$$

tomando-se como estimadores dos p_i, $i = 1, \ldots, s$

$$\hat{p}_i = (f_i) / n$$

onde:

f_i = Número de observações na i-ésima categoria.

n = Tamanho da amostra.

Demonstra-se (ver Basharin (1959)) que $\hat{H}$ é um estimador viciado e consistente de H, com distribuição assintoticamente normal de parâmetros:

$$E(\hat{H}) = H - \frac{s-1}{2N} \log_2 e + o(\frac{1}{N^2}) .$$

$$Var(\hat{H}) = \frac{1}{N} [\sum_{i=1}^{s} p_i \log_2^2 p_i - H^2] + o(\frac{1}{N^2})$$

onde $o(1/N^{\alpha})$ designa a orden $N^{-\alpha}$ $(\alpha > 0)$.

A variância estimada de $\hat{H}$ é

$$s_H^2 = (\sum_i f_i \log^2 f_i - (\sum_i f_i \log f_i)^2 / n)/n^2 ,$$

tomando-se ainda $\hat{p}_i = (f_i)/n, \quad \forall\, i = 1,\ldots,s.$

2.3. TESTE PARA DIFERENÇA ENTRE DOIS ÍNDICES DE DIVERSIDADE (TESTE T)

Suponhamos que temos duas amostras com índices de diversidade H_1 e H_2 e que queremos testar a hipótese nula, que as diversidades das amostras populacionais são iguais. Hutchenson propôs em (1970) um Teste T, obtendo a estatística.

$$t = \frac{H_1 - H_2}{\sqrt{s^2_{H_1} + s^2_{H_2}}}$$

que, como um caso particular de um resultado de Welch (1938), segue uma distribuição T com v graus de liberdade,

$$v = \frac{(s^2_{H_1} + s^2_{H_2})^2}{\frac{(s^2_{H_1})^2}{n_1} + \frac{(s^2_{H_2})^2}{n_2}} .$$

2.4. TESTES UTILIZANDO A INFORMAÇÃO INTEGRAL DE KULLBACK PARA ANALISAR TABELAS DE CONTINGÊNCIA

Nesta secção apresentamos uma metodologia para a análise de tabelas de contingência baseada em testes que utilizam a informação integral de Kullback (ver Kullback (1959), Wilks (1962)). A estes testes denominaremos testes de Kullback e que também são conhecidos como teste G (ver Sokal, Rohlf (1969)).

Aplicaremos também, as análises de McGill (1954) para tabelas de contingência, no estudo da interação entre os fatores, mas antes de apresentarmos os testes para o caso de uma tabela de três entradas, enunciaremos algumas definições que serão úteis.

DEFINIÇÕES:

D1. TABELA DE CONTINGÊNCIA: uma tabela de contingência é essencialmente uma amostra de uma população multivariada com as várias probabilidades e partições de categorias, sujeitas às restrições da distribuição multinomial.

TABELA DE CONTINGÊNCIA DE 3 FATORES

Suponhamos que temos N observações independentes e que a cada uma dessas observações associamos uma categoria em cada um dos 3 fatores: Linha (L), Coluna (C) e Depth (D).

Denominaremos:

r : número de categorias do fator Linha.

c : número de categorias do fator Coluna.

d : número de categorias do fator Depth.

r x c x d : número de células da tabela de contingência de 3 fatores.

A tabela de contingência pode então, ser representada da seguinte maneira:

C	C_1					C_j					C_c			
L	D_1			D_d	......	D_1		D_k	D_d	...	D_1			D_d
L_1														
L_2														
⋮														
L_i								x_{ijk}						
⋮														
L_r														

onde:

X_{ijk} = frequência de ocorrência na i-ésima Linha, da j-ésima Coluna da k-ésima categoria Depth, e

$\sum_{ijk}\sum\sum X_{ijk} = N.$

Denotemos:

$$X_{i..} = \sum_j \sum_k X_{ijk}, \quad X_{.j.} = \sum_i \sum_k X_{ijk} \quad X_{..k} = \sum\sum_{ij} X_{ijk}$$

$$X_{ij.} = \sum_k X_{ijk} \quad X_{i.k} = \sum_j X_{ijk} \quad X_{.jk} = \sum_i X_{ijk}$$

$$\sum_k X_{..k} = \sum_i X_{i..} = \sum_j X_{.j.} = N$$

$$\sum\sum_{ij} X_{ij.} = \sum\sum_{ik} X_{i.k} = \sum\sum_{jk} X_{.jk} = N.$$

D2. A INFORMAÇÃO MÉDIA DE DISCRIMINAÇÃO (OU INFORMAÇÃO INTEGRAL DE KULLBACK): Suponhamos duas hipóteses, H_1 e H_2, especificando a distribuição de probabilidade de duas populações c-valoradas, isto é, suponhamos:

$$H_1: p_{11},\ldots,p_{1c}, \quad \text{com} \quad \sum_{i=1}^{c} p_{1i} = 1 \quad \text{(Hipótese alternativa)}$$

$$H_2: p_{21},\ldots,p_{2c}, \quad \text{com} \quad \sum_{i=1}^{c} p_{2i} = 1 \quad \text{(Hipótese nula).}$$

Definimos informação média de discriminação de H_1 contra H_2, por

$$I(H_1 : H_2) = \sum_{i=1}^{c} p_{1i} \log \left(\frac{p_{1i}}{p_{2i}}\right)$$

com o logaritmo na base e.

E a informação média de discriminação de H_1 contra H_2 sobre uma amostra de tamanho N, (O_N), por

$$I(H_1 : H_2, O_N) = N\, I(H_1 : H_2) .$$

D3. MÍNIMA INFORMAÇÃO MÉDIA DE DISCRIMINAÇÃO E DISTRIBUIÇÃO CONJUGADA: Suponhamos, H_1 e H_2, duas hipóteses especificando a função densidade de probabilidade, f_1 e f_2, de duas populações; λ uma medida de probabilidade, $Y = T(X)$, uma estatística, tal que:

$$\theta = \int T(X)\, f_1(X)\, d\lambda(X) \quad \text{e} \quad M_2(\tau) = \int f_2(X)\, e^{\tau T(X)}\, d\lambda(X),$$

exista para τ, em algum intervalo.

Sabemos, por Kullback (1959), que a informação média de discriminação de H_1 contra H_2 satisfaz:

$$I(H_1 : H_2) \geq \theta\tau - \log M_2(\tau)$$

com

$$\theta = \frac{\partial}{\partial\tau} \log M_2(\tau) .$$

Definimos, então, a mínima informação média de discriminação, por:

$$I(* : H_2) = \theta\tau - \log M_2(\tau).$$

Observemos que:

$$I(H_1 : H_2) = I(* : H_2)$$

se, e somente se,

$$f_1(X) = f^*(X) = e^{\tau T(X)} f_2(X) / M_2(\tau), \quad \lambda\text{-a.c.}$$

onde $f^*(X)$ é denominada distribuição conjugada.

Aplicando os resultados anteriores no caso de uma população com distribuição multidimensional, c-valorada obtemos as seguintes expressões para $M_2(\tau)$, $T(X)$ e $P^*(X)$, a distribuição conjugada:

$$M_2(\tau) = (p_1 e^{\tau_1} + \ldots + p_c e^{\tau_c})^N$$

$$T(X) = (X_1, \ldots, X_c)$$

$$\tau = (\tau_1, \ldots, \tau_c)$$

$$p^*(X) = \frac{N!}{X_1! \ldots X_c!} (p_1^*)^{X_1} \ldots (p_c^*)^{X_c} \qquad (2.1)$$

onde

$$p_i^* = \frac{p_i e^{\tau_i}}{p_1 e^{\tau_1} + \ldots + p_c e^{\tau_c}}$$

Notemos que a distribuição conjugada, $p^*(X)$, é multinomial com parâmetros, p_i^*, $i = 1,\ldots,c$, dados por:

$$N\, p_i^* = E^*(X_i) = \theta_i \tag{2.2}$$

com

$$\theta_i = \frac{\partial}{\partial \tau_i} \log\,(p_1\, e^{\tau_1}+\ldots+p_c\, e^{\tau_c})^N, \quad i = 1,\ldots,c.$$

E que os valores para τ_i, $i = 1,\ldots,c$, nesse caso, são obtidos das relações:

$$\frac{\theta_i}{\theta_j} = \frac{p_i\, e^{\tau_i}}{p_j\, e^{\tau_j}}, \quad i,j = 1,\ldots,c$$

resultando

$$\tau_i = \log \frac{\theta_i}{Np_i} + \log k$$

com $k = p_1\, e^{\tau_1}+\ldots+p_c\, e^{\tau_c} > 0$.

Dessas duas últimas igualdades, obtemos:

a) $\tau_i = \log \dfrac{\theta_i}{Np_i} \qquad i = 1,\ldots,c$

para $k = 1$.

b) $\tau_i = \begin{cases} \log \dfrac{\theta_i p_c}{\theta_c p_i} & i = 1,\ldots,c-1 \\ 0 & i = c \end{cases}$

para $X_c = N-X_1-\ldots-X_{c-1}$.

D4. AMOSTRAS HOMOGÊNEAS: Amostras de populações com o mesmo parâmetro.

A. TESTE DE INDEPENDÊNCIA ENTRE TRÊS FATORES

Consideremos uma tabela de três entradas definida como em D1, na qual queremos testar, a hipótese alternativa

$$H_1 : p_{ijk} \neq p_{i..}\,p_{.j.}\,p_{..k}, \quad \text{para pelo menos um } (i,j,k), \quad \sum_{ijk}\sum\sum p_{ijk} = 1$$

versus a hipótese nula

$$H_2 : p_{ijk} = p_{i..}\,p_{.j.}\,p_{..k}, \qquad i=1,\ldots,r \quad j=1,\ldots,c \quad k=1,\ldots,d$$

$$\sum_i p_{i..} = \sum_j p_{.j.} = \sum_k p_{..k} = 1$$

$$p_{i..} > 0, \quad p_{.j.} > 0, \quad p_{..k} > 0$$

onde p_{ijk}, $i = 1,\ldots,r$; $j = 1,\ldots,c$; $k = 1,\ldots,d$, são as probabilidades associadas as categorias i,j,k dos fatores L, C e D respectivamente (ver D1).

A estatística proposta por Kullback, para testar essa hipótese nula é duas vezes a Mínima informação média de discriminação estimada, $2\hat{I}((p)^*:(p))$, que segue uma distribuição qui-quadrado com $(r\,c\,d-1)$ graus de liberdade.

A demonstração desse fato é dada a seguir.

Primeiramente calculamos a informação média de discriminação de H_1 versus H_2, aplicando D2. Obtemos então,

$$I(H_1 : H_2;\, O_N) = N \sum_{ijk}\sum\sum p_{ijk} \log \left(p_{ijk} / (p_{i..}\,p_{.j.}\,p_{..k})\right) \qquad (2.3)$$

A mínima informação média de discriminação é obtida substituindo-se, $p(\underset{\sim}{X})$ de (2.3) pela sua distribuição conjugada $p^*(\underset{\sim}{X})$, obtida em D3, resultando

$$I((p)^* : (p)) = N \sum_{ijk}\sum\sum p^*_{ijk} \log (p^*_{ijk} / (p_{i..} p_{.j.} p_{..k})) \qquad (2.4)$$

Por (2.2) temos:

$$E(X_{ijk}) = \theta_{ijk} = N\, p^*_{ijk}$$

onde $\hat{p}^*_{ijk} = X_{ijk}/N$ é um estimador não viciado.

E novamente substituindo, em (2.4), p^*_{ijk} por seu estimador $\hat{p}^*_{ijk}$, obtemos que a mínima informação média de discriminação de H_1 versus H_2 estimada, é:

$$\hat{I}((p)^* : (p)) = \sum_{ijk}\sum\sum X_{ijk} \log (X_{ijk} / (N\, p_{i..} p_{.j.} p_{..k})) \; .$$

Observemos que a seguinte aproximação

$$\log (a/b) \simeq (a^2 - b^2) / 2ab \qquad (2.5)$$

pode ser aplicada ao membro do lado direito da equação anterior e essa aproximação é sensivelmente boa quando

$$X_{ijk} / N \simeq p_{i..}\; p_{.j.}\; p_{..k} \; .$$

Utilizando-a então, podemos concluir que o estimador, $2\,\hat{I}((p)^* : (p))$ tem distribuição qui-quadrado com $(r\,c\,d-1)$ graus de liberdade.

Mas, geralmente não dispomos dos valores $p_{i..}$, $p_{.j.}$, $p_{..k}$ para achar a estatística $2\,\hat{I}((p)^* : (p))$ e por isso vamos calcular uma estatística que dependa somente dos dados da amostra.

Reescrevendo, então, a última equação obtemos

$$\hat{I}((p)^* : (p)) = \sum_{ijk}\sum\sum X_{ijk} \log \left(\frac{X_{ijk}}{N\, \tilde{p}_{i..}\tilde{p}_{.j.}\tilde{p}_{..k}}\right) + \tag{2.6}$$

$$+ N \sum_{ijk}\sum\sum \tilde{p}_{i..}\tilde{p}_{.j.}\tilde{p}_{..k} \log \frac{\tilde{p}_{i..}\tilde{p}_{.j.}\tilde{p}_{..k}}{p_{i..}p_{.j.}p_{..k}} +$$

$$+ N \sum_{ijk}\sum\sum \left[\frac{X_{ijk}}{N} - \tilde{p}_{i..}\tilde{p}_{.j.}\tilde{p}_{..k}\right] \log \left[\frac{\tilde{p}_{i..}\tilde{p}_{.j.}\tilde{p}_{..k}}{p_{i..}p_{.j.}p_{..k}}\right].$$

Sob a hipótese nula, H_2, existem

(r-1) parâmetros independentes: $p_{1..}, \ldots, p_{(r-1)..}$

(c-1) parâmetros independentes: $p_{.1.}, \ldots, p_{.(c-1).}$

(d-1) parâmetros independentes: $p_{..1}, \ldots, p_{..(r-1)}$

onde

$$\tilde{p}_{ijk} = p_{ijk}(\tilde{p}_{1..}, \ldots, \tilde{p}_{(r-1)..}, \tilde{p}_{.1.}, \ldots, \tilde{p}_{.(c-1).}, \tilde{p}_{..1}, \ldots, \tilde{p}_{..(d-1)})$$

e $\tilde{p}_{i..} = \sum_{jk}\sum \tilde{p}_{ijk}$, $\tilde{p}_{.j.} = \sum_{ik}\sum \tilde{p}_{ijk}$, $\tilde{p}_{..k} = \sum_{ij}\sum \tilde{p}_{ijk}$.

Calculemos, agora, os valores de $\tilde{p}_{ijk}$ que anulam o último termo de (2.6), isto é, que tornam a informação aditiva:

$$\sum_{ijk}\sum\sum \frac{X_{ijk}}{N} \log\left(\frac{\tilde{p}_{i..}\tilde{p}_{.j.}\tilde{p}_{..k}}{p_{i..}p_{.j.}p_{..k}}\right) = \sum_{ijk}\sum\sum \tilde{p}_{i..}\tilde{p}_{.j.}\tilde{p}_{..k} \log\left(\frac{\tilde{p}_{i..}\tilde{p}_{.j.}\tilde{p}_{..k}}{p_{i..}p_{.j.}p_{..k}}\right)$$

Sob a hipótese nula, H_2, obtemos

$$\tilde{p}_{i..} = \frac{X_{i..}}{N} \qquad i = 1,\ldots,r$$

$$\tilde{p}_{.j.} = \frac{X_{.j.}}{N} \qquad j = 1,\ldots,c$$

$$\tilde{p}_{..k} = \frac{X_{..k}}{N} \qquad k = 1,\ldots,d.$$

E finalmente, substituindo os valores de $\tilde{p}_{i..}$, $\tilde{p}_{.j.}$, $\tilde{p}_{..k}$ na equação (2.6), e multiplicando tudo por 2 obtemos

$$\begin{aligned}
2\hat{I}((p)^* : (p)) &= 2\sum_i X_{i..} \log\left(\frac{X_{i..}}{N\,p_{i..}}\right) + 2\sum_j X_{.j.} \log\left(\frac{X_{.j.}}{N\,p_{.j.}}\right) + \\
&+ 2\sum_k X_{..k} \log\left(\frac{X_{..k}}{N\,p_{..k}}\right) + 2\sum_{ijk}\sum\sum X_{ijk} \log\left(\frac{N^2 X_{ijk}}{X_{i..}X_{.j.}X_{..k}}\right) \\
&= 2\hat{I}(\tilde{p}_{i..} : p_{i..}) + 2\hat{I}(\tilde{p}_{.j.} : p_{.j.}) + 2\hat{I}(\tilde{p}_{..k} : p_{..k}) + \\
&+ 2\hat{I}(H_1 : H_2(L \times C \times D)) .
\end{aligned}$$

Para testar as hipóteses correspondentes às três primeiras estatísticas precisaríamos de $p_{i..}, p_{.j.}, p_{..k}$ respectivamente, mas em geral não dispomos dessa informação. Por outro lado a última estatística, $2\hat{I}(H_1:H_2(L \times C \times D))$, nos permite testar a hipótese H_1 versus H_2, a qual não contém nenhum parâmetro desconhecido, e segue uma distribuição qui-quadrado com $rcd-1-(r-1)-(c-1)-(d-1)$ graus de liberdade. Isso pode verificar-se facilmente aplicando-se em $2\hat{I}(H_1:H_2(L \times C \times D))$ a aproximação dada em (2.5).

Observemos que $2\hat{I}(H_1:H_2(L \times C \times D))$, também chamado componente de independência, pode ser analisado através de seus componentes aditivos, ou seja

$$2\hat{I}(H_1:H_2(L\times C\times D)) = 2 \sum_{i}\sum_{j} X_{.jk} \log \left(\frac{N X_{.jk}}{X_{.j.} X_{..k}}\right) + 2 \sum_{i}\sum_{j}\sum_{k} X_{ijk} \log \left(\frac{N X_{ijk}}{X_{i..} X_{.jk}}\right)$$

$$= 2\hat{I}(H_1:H_2(C\times D)) + 2\hat{I}(H_1:H_2(L\times CD))$$

onde:

$2\hat{I}(H_1:H_2(C \times D))$ é a estatística para testar a hipótese,

$H_1 : p_{.jk} \neq p_{.j.} p_{..k}$, para pelo menos um (j,k)

versus

$H_2 : p_{.jk} = p_{.j.} p_{..k}$, $\quad j = 1,\ldots,c \quad k = 1,\ldots,d$

a qual segue uma distribuição qui-quadrado com $(c-1)(d-1)$ graus de liberdade.

E $2\hat{I}(H_1:H_2(L \times CD))$ é a estatística para testar a hipótese,

$H_1: p_{ijk} \neq p_{i..}p_{.jk}$, para pelo menos (i,jk), $\sum\sum\sum_{ijk} p_{ijk} = 1$

versus

$$H_2: p_{ijk} = p_{i..}p_{.jk}, \quad i = 1,\dots,r \quad j = 1,\dots,c \quad k = 1,\dots,d$$

$$\sum_i p_i = \sum\sum_{jk} p_{.jk} = 1$$

a qual segue uma distribuição qui-quadrado com $(r-1)(cd-1)$ graus de liberdade.

A análise de $2\hat{I}(H_1: H_2(L \times C \times D))$ através de seus componentes aditivos vem do fato de que $H_2(L \times C \times D) \Leftrightarrow H_2(L \times CD) \cap H_2(C \times D)$, isto é, os 3 fatores são independentes se e somente se o fator Linha é independente de par Coluna, Depth e o fator Coluna é independente do fator Depth, já que,

$$p_{ijk} = p_{i..}p_{.jk} \quad \text{e} \quad p_{.jk} = p_{.j.}p_{..k} \Rightarrow p_{ijk} = p_{i..}p_{.j.}p_{..k}$$

e

$$p_{ijk} = p_{i..}p_{.j.}p_{..k} \Rightarrow p_{.jk} = p_{.j.}p_{..k} \quad \text{ou} \quad p_{ijk} = p_{i..}p_{.jk} .$$

Nossa tabela para análise da independência é dada a seguir:

TABELA 2.1

COMPONENTE	INFORMAÇÃO	GRAUS DE LIBERDADE
COLUNA x DEPTH $H_2(C \times D)$	$2 \sum\sum_{jk} X_{.jk} \log \left(\frac{N X_{.jk}}{X_{.j.} X_{..k}}\right)$	$(c-1)(d-1)$
LINHA x (COLUNA, DEPTH) $H_2(L \times CD)$	$2 \sum\sum\sum_{ijk} X_{ijk} \log \left(\frac{N X_{ijk}}{X_{i..} X_{.jk}}\right)$	$(r-1)(cd-1)$
INDEPENDÊNCIA $H_2(R \times C \times D)$	$2 \sum\sum\sum_{ijk} X_{ijk} \log \left(\frac{N^2 X_{ijk}}{X_{i..} X_{.j.} X_{..k}}\right)$	$rcd-r-c-d+2$

Lembremos que podemos realizar outras análises com respeito à independência usando os seguintes fatos:

$$H_2(L \times C \times D) \Leftrightarrow H_2(C \times LD) \cap H_2(L \times D)$$

pois,

$$p_{ijk} = p_{.j.} p_{i.k} \quad e \quad p_{i.k} = p_{i..} p_{..k} \Rightarrow p_{ijk} = p_{i..} p_{.j.} p_{..k}$$

e

$$p_{ijk} = p_{i..} p_{.j.} p_{..k} \Rightarrow p_{i.k} = p_{i..} p_{..k} \quad ou \quad p_{ijk} = p_{.j.} p_{i.k}$$

e

$$H_2(L \times C \times D) \Leftrightarrow H_2(LC \times D) \cap H_2(L \times C)$$

pois

$$p_{ijk} = p_{ij.}p_{..k} \quad e \quad p_{ij.} = p_{i..}p_{.j.} \Rightarrow p_{ijk} = p_{i..}p_{.j.}p_{..k}$$

e

$$p_{ijk} = p_{i..}p_{.j.}p_{..k} \Rightarrow p_{ij.} = p_{i..}p_{.j.} \quad ou \quad p_{ijk} = p_{ij.}p_{..k} \; .$$

Ou ainda que

$$H_2(L \times CD) \Leftrightarrow H_2((L/D) \times (C/D)) \cap H_2(L \times D) \; .$$

Isto é, o fator Linha é independente dos fatores (Coluna, Depth) se e somente se os fatores Linha e Coluna são condicionalmente independentes, dado o fator Depth e se os fatores Linha e Depth são independentes, o que ocorre já que:

$$p_{ijk} = p_{i.k}p_{.jk}/p_{..k} \quad e \quad p_{i.k} = p_{i..}p_{..k} \Rightarrow p_{ijk} = p_{i..}p_{.jk}$$

e

$$p_{ijk} = p_{i..}p_{.jk} \Rightarrow p_{i.k} = p_{i..}p_{..k} \quad ou \quad p_{ijk} = p_{i.k}p_{.jk}/p_{..k} .$$

Podemos também, decompor $2\hat{I}(H_1:H_2(L \times CD))$ da seguinte forma:

$$2\hat{I}(H_1:H_2(L\times CD)) = 2\sum_{ik}\sum X_{i.k} \log \left(\frac{N\,X_{i.k}}{X_{i..}X_{..k}}\right) + 2\sum_{ijk}\sum\sum X_{ijk} \log\left(\frac{X_{ijk}X_{..k}}{X_{i.k}X_{.jk}}\right)$$

$$= 2\hat{I}(H_1:H_2(L\times D)) + 2\hat{I}(H_1:H_2((L/D)\times(C/D)))$$

onde

$2\hat{I}(H_1:H_2(L \times D))$ é a estatística para testar a hipótese,

H_1: $p_{i.k} \neq p_{i..}p_{..k}$, para pelo menos (i,k)

versus

H_2: $p_{i.k} = p_{i..}p_{..k}$, $i = 1,\dots,r$ $k = 1,\dots,d$

a qual segue uma distribuição qui-quadrado com $(r-1)(d-1)$ graus de liberdade.

E $2\hat{I}(H_1:H_2((L/D) \times (C/D)))$ é a estatística para testar se os fatores Linha e Coluna são condicionalmente independentes dado o fator Depth, isto é, $2\hat{I}(H_1:H_2((L/D) \times (C/D)))$ é a estatística para testar:

$$H_1: p_{ijk} \neq \frac{p_{i.k}\, p_{.jk}}{p_{..k}}, \quad \text{para pelo menos um } (i,j), \quad \sum_{i}\sum_{j} p_{ijk} = 1$$

versus

$$H_2: p_{ijk} = \frac{p_{i.k}\, p_{.jk}}{p_{..k}}, \; i=1,\dots,r,\; j=1,\dots,c,\; k=1,\dots,d, \quad p_{..k} = \sum_{i}\sum_{j} p_{ijk}$$

a qual segue uma distribuição qui-quadrado com $d(r-1)(c-1)$ graus de liberdade.

Logo, nossa tabela para análise de independência será:

TABELA 2.2

COMPONENTE	INFORMAÇÃO	GRAUS DE LIBERDADE
LINHA x DEPTH $H_2(L \times D)$	$2 \sum\sum_{ik} X_{i.k} \log \left(\frac{N X_{i.k}}{X_{i..} X_{..k}}\right)$	$(r-1)(d-1)$
(LINHA/DEPTH) x (COLUNA/DEPTH) $H_2((L/D) \times (C/D))$	$2 \sum\sum\sum_{ijk} X_{ijk} \log \left(\frac{X_{ijk} X_{..k}}{X_{i.k} X_{.jk}}\right)$	$d(r-1)(c-1)$
LINHA x (COLUNA, DEPTH) $H_2(L \times CD)$	$2 \sum\sum\sum_{ijk} X_{ijk} \log \left(\frac{N X_{ijk}}{X_{i..} X_{.jk}}\right)$	$(r-1)(cd-1)$

B. TESTES DE HOMOGENEIDADE

Suponhamos que queremos testar a hipótese nula, que r amostras independentes de uma tabela $c \times d$ são homogêneas, isto é, que as r amostras de populações multinomiais com cd categorias são homogêneas (ver D4).

Podemos abordar esse problema considerando-se essas r amostras como uma tabela de 3 fatores $(r \times c \times d)$ e nessa situação o teste de hipótese é dado por:

$$H_1 : p_{ijk} \neq p_{.jk}, \quad \sum_{j=1}^{c} \sum_{k=1}^{d} p_{ijk} = 1 \qquad i = 1,\dots,r$$

$$H_2 : p_{ijk} = p_{.jk} \qquad i = 1,\ldots,r \qquad j = 1,\ldots,c \qquad k = 1,\ldots,d \, .$$

$$\sum_{jk}\sum p_{.jk} = 1$$

A estatística proposta por Kullback para testar essa hipótese é:

$$2\hat{I}((p)^*:(p)) = 2 \sum_{ijk}\sum\sum X_{ijk} \log \left(\frac{X_{ijk}}{X_{i..}\,p_{.jk}}\right)$$

a qual segue uma distribuição qui quadrado com $r(c\,d-1)$ graus de liberdade (o procedimento para deduzir esta estatística é similar ao apresentado no caso da estatística para independência entre fatores).

E como, em geral não conhecemos os valores de $p_{.jk}$, podemos seguir o mesmo raciocínio feito anteriormente, obtendo:

$$\hat{I}((p)^*:(p)) = \sum_{ijk}\sum\sum X_{ijk} \log \left(\frac{X_{ijk}}{X_{i..}\,\tilde{p}_{.jk}}\right) + N \sum_{ijk}\sum\sum \tilde{p}_{.jk} \log \left(\frac{\tilde{p}_{.jk}}{p_{.jk}}\right) \qquad (2.7)$$

$$+ N \sum_{ijk}\sum\sum \left[\frac{X_{ijk}}{N} - \tilde{p}_{.jk}\right] \log \left(\frac{\tilde{p}_{.jk}}{p_{.jk}}\right) \, .$$

Ainda de maneira análoga ao caso de independência, observamos, que, sob a hipótese nula, H_2, existem $c\,d-1$ parâmetros independentes $(p_{.11},\ldots,p_{.c(d-1)})$ onde:

$$\tilde{p}_{ijk} = p_{ijk}(\tilde{p}_{.11},\ldots,\tilde{p}_{.c(d-1)})$$

e

$$\tilde{p}_{.jk} = \sum_i \tilde{p}_{ijk} \, .$$

Então, sob a hipótese nula obtemos:

$$\tilde{p}_{.jk} = X_{.jk}/N \quad j = 1,\ldots,c \quad k = 1,\ldots d \, .$$

E substituindo os valores de $\tilde{p}_{.jk}$ em (2.7), temos:

$$2\hat{I}((p)^*:(p)) = 2\sum_{jk}\sum X_{.jk} \log\left(\frac{X_{.jk}}{N\,p_{.jk}}\right) + 2\sum_{ijk}\sum\sum X_{ijk} \log\left(\frac{N\,X_{ijk}}{X_{i..}\,X_{.jk}}\right)$$

$$= 2\hat{I}(H_1:H_2(\tilde{p}_{.jk}:p_{.jk})) + 2\hat{I}(H_1:H_2 \text{ (Homogeneidade (C,D)))} \, .$$

Para obtermos a primeira estatística precisaríamos do valor de $p_{.jk}$, que em geral é desconhecido. Mas por outro lado, obtemos, $2\hat{I}(H_1:H_2$ (Homogeneidade (C,D))), que é a estatística para testar H_1 versus H_2, e que segue uma distribuição qui quadrado com $(r-1)(c\,d-1)$ graus de liberdade (este último resultado é obtido aplicando-se a aproximação dada em (2.5) em $2\hat{I}(H_1:H_2$ (Homogeneidade (C,D))).

Podemos realizar ainda, outras análises dentro do componente de Homogeneidade (C,D) através de seus componentes aditivos. Isto vem do fato de que

Homogeneidade (C,D) $\Leftrightarrow$ Homogeneidade (D/C) $\cap$ Homogeneidade (C)

já que,

$$p_{ijk}/p_{ij.} = p_{.jk}/p_{.j.} \qquad i=1,\dots,r \quad j=1,\dots,c \quad k=1,\dots,d$$

e $p_{ij.} = p_{.j.} \Rightarrow p_{ijk} = p_{.jk}$, $\quad i=1,\dots,r, \quad j=1,\dots,c, \quad k=1,\dots,d.$

E $\quad p_{ijk} = p_{.jk} \qquad i = 1,\dots,r \quad j = 1,\dots,c \quad k = 1,\dots,d$

$\Rightarrow p_{ij.} = p_{.j.} \qquad i = 1,\dots,r \quad j = 1,\dots,c$ e $p_{ijk}/p_{ij.} = p_{.jk}/p_{.j.}$

Decompondo $2\hat{I}(H_1:H_2$ (Homogeneidade (C,D))) temos:

$2\hat{I}(H_1:H_2$ (Homogeneidade (C,D))) =

$$= 2 \sum_{ij}\sum X_{ij.} \log \left(\frac{N X_{ij.}}{X_{i..} X_{.j.}}\right) + 2 \sum_{ijk}\sum\sum X_{ijk} \log \left(\frac{X_{ijk}}{\frac{X_{ij.} X_{.jk}}{X_{.j.}}}\right)$$

$= 2\hat{I}(H_1:H_2$ (Homogeneidade (C))) + $2\hat{I}(H_1:H_2$ (Homogeneidade Condicional (D/C))) ,

onde $2\hat{I}(H_1:H_2$ (Homogeneidade (C))) é a estatística para testar a hipótese

$H_1 : p_{ij.} \neq p_{.j.}$, para pelo menos (i,j) $\quad i=1,\dots,r \quad j=1,\dots,c$

$H_2 : p_{ij.} = p_{.j.}$, $\quad i = 1,\dots,r \quad j = 1,\dots,c$

a qual segue uma distribuição qui quadrado com $(r-1)(c-1)$ graus de liberdade.

E $2\hat{I}(H_1:H_2$ (Homogeneidade Condicional (D/C))) é a estatística para testar se as classificações (ou categorias) Depth são condicionalmente homogêneas dado o fator Coluna. O teste de hipótese correspondente é:

$$H_1 : p_{ijk}/p_{ij.} \neq p_{.jk}/p_{.j.} \qquad i=1,\ldots,r \qquad j=1,\ldots,c \qquad \sum_k p_{ijk}=p_{ij.}$$

$$H_2 : p_{ijk}/p_{ij.} = p_{.jk}/p_{.j.}, \qquad i=1,\ldots,r \qquad j=1,\ldots,c \qquad k=1,\ldots,d$$

$$\sum_k p_{.jk} = p_{.j.}$$

a qual segue uma distribuição qui quadrado com $c(r-1)(d-1)$ graus de liberdade.

Pode-se fazer uma análise mais detalhada analisando se as classificações (ou categorias) Depth são condicionalmente homogêneas dada j-ésima classificação Coluna. O teste de hipótese correspondente é:

$$H_1 : \frac{p_{ijk}}{p_{ij.}} \neq \frac{p_{.jk}}{p_{.j.}}, \qquad i=1,\ldots,r \qquad \sum_k p_{ijk} = p_{ij.}$$

$$H_2 : \frac{p_{ijk}}{p_{ij.}} = \frac{p_{.jk}}{p_{.j.}}, \qquad i=1,\ldots,r \qquad k=1,\ldots,d \qquad \sum_k p_{.jk} = p_{.j.}$$

e a estatística para testar a hipótese é

$$2\hat{I}(H_1:H_2) = 2 \sum_{ik}\sum X_{ijk} \log \left(\frac{X_{ijk}X_{.j.}}{X_{ij.}X_{.jk}}\right)$$

a qual segue uma distribuição qui quadrado com (r-1)(d-1) graus de liberdade.

Logo nossa tabela para a análise da homogeneidade é:

TABELA 2.3

COMPONENTE	INFORMAÇÃO	GRAUS DE LIBERDADE
HOMOGENEIDADE C	$2 \sum_{ij}\sum X_{ij.} \log \left(\frac{N X_{ij.}}{X_{i..} X_{.j.}}\right)$	(r-1)(c-1)
HOMOGENEIDADE CONDICIONAL (D/C)	$2 \sum\sum_{ijk}\sum X_{ijk} \log \left(\frac{X_{ijk}X_{.j.}}{X_{ij.} X_{.jk}}\right)$	c(r-1)(d-1)
HOMOGENEIDADE CONDICIONAL (D / C = j)	$2 \sum_{ik}\sum X_{ijk} \log \left(\frac{X_{ijk}X_{.j.}}{X_{ij.} X_{.jk}}\right)$	(r-1)(d-1)
⋮		
HOMOGENEIDADE (C,D)	$2 \sum\sum_{ijk}\sum X_{ijk} \log \frac{N X_{ijk}}{X_{i..} X_{.jk}}$	(r-1)(cd-1)

C. TESTES DE INTERAÇÃO

Podemos analisar o componente de homogeneidade Condicional (D/C) dentro de dois componentes aditivos. Isto vem do fato que:

$$H_2 \text{ (Homog. Condicional (D/C))} \Leftrightarrow H_2 \text{ (Interação LD)} \cap H_2 \text{ (Interação (LD,C))}$$

já que

$$p_{i.k} = \sum_{j=1} \frac{p_{ij.}p_{.jk}}{p_{.j.}} \quad \text{e} \quad p_{ijk} = \frac{p_{i.k}p_{ij.}p_{.jk}}{\left(\sum_j \frac{p_{ij.}p_{.jk}}{p_{.j.}}\right) p_{.j.}} \Leftrightarrow \frac{p_{ijk}}{p_{ij.}} = \frac{p_{.jk}}{p_{.j.}} .$$

Decompondo $2\hat{I}(H_1 : H_2$ (Homogeneidade Condicional (D/C))) temos:

$$2\hat{I}(H_1 : H_2 \text{ (Homogeneidade Condicional (D/C)))} =$$

$$= 2 \sum_{ik}\sum X_{i.k} \log \left(\frac{X_{i.k}}{y_{i.k}}\right) + 2 \sum_{ijk}\sum\sum X_{ijk} \log \left(\frac{X_{ijk}y_{i.k}X_{.j.}}{X_{i.k}X_{ij.}X_{.jk}}\right)$$

$$= 2\hat{I}(H_1 : H_2 \text{ (Interação RD))} + \hat{I}(H_1 : H_2 \text{ ((Interação (LD,C)))}$$

onde

$$y_{i.k} = \sum_{j=1}^{c} \frac{X_{ij.}X_{.jk}}{X_{.j.}} .$$

E $2\hat{I}(H_1 : H_2$ (Interação LD)) é a estatística para testar a hipótese se existe interação (ou relação) entre os fatores L,D; a qual segue uma distribuição qui quadrado com $(r-1)(d-1)$ graus de liberdade. Por outro lado $2\hat{I}(H_1:H_2$ (Interação (LD,C)) é a estatística para testar se existe interação (ou relação) entre o fator C e os fatores L,D.

Logo, nossa tabela de análise será:

TABELA 2.4

COMPONENTE	INFORMAÇÃO	GRAUS DE LIBERDADE
INTERAÇÃO LD	$2 \sum\sum_{ij} X_{i.k} \log \left(\frac{X_{i.k}}{Y_{i.k}}\right)$	$(r-1)(d-1)$
INTERAÇÃO (LD,C)	$2 \sum\sum\sum_{ijk} X_{ijk} \log \left(\frac{X_{ijk} Y_{i.k} X_{.j.}}{X_{i.k} X_{ij.} X_{.jk}}\right)$	$(r-1)(c-1)(d-1)$
HOMOG. CONDICIONAL (D/C)	$2 \sum\sum\sum_{ijk} X_{ijk} \log \left(\frac{X_{ijk} X_{.j.}}{X_{ij.} X_{.jk}}\right)$	$c(r-1)(d-1)$

Por outro lado o componente de homogeneidade condicional (D/C) pode também ser analisado algebricamente

$$2\hat{I}(H_1:H_2 \text{ (Homogeneidade Condicional (D/C)))} =$$

$$= 2 \sum\sum_{ik} X_{i.k} \log \left(\frac{N X_{i.k}}{X_{i..} X_{..k}}\right) + 2 \sum\sum\sum_{ijk} X_{ijk} \log \left(\frac{X_{ijk} X_{i..} X_{.j.} X_{..k}}{N X_{ij.} X_{i.k} X_{.jk}}\right)$$

$$= 2\hat{I}(H_1:H_2 \text{ (Homogeneidade D))} + 2\hat{I}(H_1:H_2 \text{ (Interação LCD))} .$$

Nem sempre o componente de homogeneidade D é menor que o componente de homogeneidade condicional (D/C). Se o componente de homogeneidade D é maior que o componente de homogeneidade (D/C) então a interação LCD é negativa e portanto não segue uma distribuição qui quadrado (já que ela é uma diferença de dois qui quadrados e essa diferença pode ser negativa). No apêndice mostramos sua função densidade supondo independência das variáveis (para maior informação ber Kullback (1959), McGill (1954), Springer (1979), Pachares (1952), Pearson, Stouffer e David (1932)).

Na continuação, apresentamos a tabela de análise.

TABELA 2.5

COMPONENTE	INFORMAÇÃO	GRAUS DE LIBERDADE
HOMOGENEIDADE D	$2 \sum\sum_{ik} X_{i.k} \log \left(\frac{N X_{i.k}}{X_{i..} X_{..k}}\right)$	$(r-1)(d-1)$
INTERAÇÃO LCD	$2 \sum\sum\sum_{ijk} X_{ijk} \log \left(\frac{X_{ijk} X_{i..} X_{.j.} X_{..k}}{N X_{ij.} X_{i.k} X_{.jk}}\right)$	FUNÇÃO DE WHITTAKER
HOMOG. CONDICIONAL (D/C)	$2 \sum\sum\sum_{ijk} X_{ijk} \log \left(\frac{X_{ijk} X_{.j.}}{X_{ij.} X_{.jk}}\right)$	$c(r-1)(d-1)$

2.5. TESTE DE INDEPENDÊNCIA PARA TABELA DE CONTINGÊNCIA UTILIZANDO A DISTÂNCIA DE MATUSITA

Consideremos uma tabela de 3 fatores definida como em D1, na qual queremos testar, a hipótese alternativa

$$H_1 : p_{ijk} \neq p_{i..}\,p_{.j.}\,p_{..k}, \quad \text{para pelo menos um } (i,j,k)$$

$$\text{e} \quad \sum_{ijk}\sum\sum p_{ijk} = 1, \quad p_{ijk} > 0$$

versus a hipótese nula:

$$H_2 : p_{ijk} = p_{i..}\,p_{.j.}\,p_{..k}, \quad i = 1,\dots,r \quad j = 1,\dots,c \quad k = 1,\dots,d$$

$$\sum_i p_{i..} = \sum_j p_{.j.} = \sum_k p_{..k} = 1$$

$$p_{i..} > 0, \quad p_{.j.} > 0, \quad p_{..k} > 0 .$$

A estatística proposta por Goldsteim, Wolf, Dillon (1976), para testar essa hipótese é:

$$\hat{\rho} = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{d} (X_{ijk}\, X_{i..}\, X_{.j.}\, X_{..k} / N^4)^{1/2}$$

onde $\hat{\rho}$ é assintoticamente $1-(1/8N)\,\chi^2_{(rcd - r - c - d+2)}$.

Por outro lado, para testar a hipótese:

$$H_1 : p_{ijk} \neq p_{ij.}\,p_{..k}, \quad \text{para pelo menos } (ij,k), \quad \sum_{ijk}\sum\sum p_{ijk} = 1$$

$$H_2 : p_{ijk} = p_{ij.}\,p_{..k}, \quad i = 1,\dots,r \quad j = 1,\dots,c \quad k = 1,\dots,d$$

$$\sum_{ij}\sum p_{ij.} = \sum_k p_{..k} = 1$$

eles propuseram a seguinte estatística:

$$\hat{\rho} = \sum_{ijk}\sum\sum (X_{ijk}\, X_{ij.}\, X_{..k} / N^3)^{1/2}$$

onde $\hat{\rho}$ é assintoticamente $1-(1/8N)\,.\,\chi^2_{(rc-1)(d-1)}$.

2.6. TESTE QUI QUADRADO PARA TESTAR INDEPENDÊNCIA DE FATORES EM TABELAS DE CONTINGÊNCIA

Novamente consideremos a tabela de Contingência de 3 fatores definida em D1 e consideremos o teste de hipótese

$H_1 : p_{ijk} \neq p_{i..}p_{.j.}p_{..k}$, para pelo menos (i,j,k)

$H_2 : p_{ijk} = p_{i..}p_{.j.}p_{..k}$, $i = 1,\ldots,r$ $j = 1,\ldots,c$ $k = 1,\ldots,d.$

O teste Qui Quadrado para testar essa hipótese é:

$$x^2 \doteq \sum_{ijk}\sum\sum (X_{ijk} - X_{i..}X_{.j.}X_{..k} / N^2)^2 \;/\; (X_{i..}X_{.j.}X_{..k} / N^2)$$

o qual segue uma distribuição qui quadrado com $rcd-r-c-d+2$ graus de liberdade.

De forma similar ao anterior se podem efetuar outros testes de independência.

2.7. ANÁLISE DE VARIÂNCIA (ANOVA)

Para efeito da análise da tabela de contingência de 3 fato-

UNICAMP
BIBLIOTECA CENTRAL

res, podemos utilizar modelos log-linear (ver Sokal, Rohlf (1969), Everit (1977)), da seguinte forma:

$$\log \hat{f}_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}$$

onde

$\hat{f}_{ijk}$ = frequência esperada na i-ésima Linha, j-ésima Coluna, k-ésima Depth da tabela de contingência.

μ = a média dos logaritmos de frequências esperadas.

α_i = efeito da categoria e do fator Linha.

β_j = efeito da categoria j do fator Coluna.

γ_k = efeito da categoria k do fator Depth.

$\alpha\beta_{ij}$ = termo de interação que expressa a dependência da categoria i do fator Linha e a j do fator Coluna.

$\alpha\gamma_{ik}$ = termo de interação que expressa a dependência da categoria i do fator Linha e k do fator Depth.

$\beta\gamma_{jk}$ = termo da interação que expressa a dependência da categoria j do fator Coluna e k do fator Depth.

$\alpha\beta\gamma_{ijk}$ = termo de interação que expressa a dependência da categoria i do fator Linha e j do fator Coluna e k do fator Depth.

2.8. DISCUSSÃO SOBRE OS TESTES

Podemos observar que a estatística $2\hat{I}$ do teste de Kullback, com sua propriedade de aditividade é bastante rica, no sentido de abrir novos campos para análise.

Isso acontece porque seus componentes aditivos nos fornecem novas estatísticas, que nos permitem testar mais hipóteses do que qualquer outro teste. Por exemplo, com o teste de Kullback, podemos examinar, numa tabela de contingência de 3 fatores, as possíveis interações de dois fatores, as interações de dois fatores com um terceiro; e as interações entre os três fatores.

O teste de Kullback substitui ainda o teste χ^2 e ANOVA nos casos em que estes não se aplicam, isto é, quando as esperanças de cada célula numa tabela de contingência, são pequenas.

O teste de Kullback além de apresentar grande vantagem computacional, no estudo da independência entre fatores, sobre o teste de independência utilizando a distância de Matusita; testa também interação e homogeneidade entre fatores, os quais não podem ser testados utilizando-se a distância de Matusita.

Em resumo o teste de Kullback é mais completo e eficiente computacionalmente, que outros testes pois examina independência, interações e homogeneidade entre os fatores de uma tabela de contingência, utilizando para ele expressões da forma n log n. (Para mais informações a respeito ver McGill (1954); Fisher (1950); Sokal, Rohlf (1969); Everit (1977).)

CAPÍTULO III

APLICAÇÃO DA METODOLOGIA

Neste capítulo apresentamos os resultados obtidos na aplicação da metodologia aos problemas expostos no capítulo 1.

3.1. APLICAÇÃO DA METODOLOGIA NO PROBLEMA DA INTEGRAÇÃO DE IMIGRANTES NA CIDADE DE SÃO PAULO

Para o estudo e análise do fenômeno da integração dos imigrantes na cidade de São Paulo, aplicaremos os três seguintes métodos; a distância de Matusita, o teste T e testes de Kullback.

Seus resultados serão apresentados nas sub-secções 3.1.1, 3.1.2 e 3.1.3 respectivamen_e.

3.1.1. APLICAÇÃO DA DISTÂNCIA DE MATUSITA

Numa primeira análise utilizaremos a Distância Matusita, para ver quais foram os grupos mais integrados e qual foi o grau de integração de cada um deles ao longo de todos os períodos de tempo estudados. O procedimento e os resultados obtidos são dados a seguir.

Baseados nos dados da tabela 1.2, calculamos a distância de Matusita, para cada um dos 13 grupos de 4 anos, comparando o comportamento dos casamentos entre os cinco grupos étnicos segundo o sexo e em conjunto, nessa ordem: brasileiros e italianos (B-I), brasileiros e espanhóis (B-E), brasileiros e portugueses (B-P), brasileiros e japoneses (B-J), italianos e espanhóis (I-E), italianos e portugueses (I-P), italianos e japoneses (I-J), espanhóis e portugueses (E-P), espanhóis e japoneses (E-J), portugueses e japoneses (P-J).

Ordenamos os grupos étnicos de acordo com o grau de integração na cidade de São Paulo, entendendo por maior grau, aquele que apresenta maior proporção de casamentos com elementos de outros grupos.

Na continuação apresentamos o estudo do comportamento de integração das mulheres, dos homens e dos grupos.

a) ESTUDO DO COMPORTAMENTO DE INTEGRAÇÃO DAS MULHERES

Fixamos, primeiro, o grupo mais integrado, ou seja, aquele que tem menor porcentagem de endogamia. Por exemplo, para o período 78/81 da tabela 1.2, obtemos as seguintes porcentagens de endogamia:

PERÍODO	B	I	E	P	J
78/81	0,9816	0,1337	0,1386	0,2439	0,4801

Vemos que a menor porcentagem de endogamia corresponde ao grupo das imigrantes italianas e portanto no período de 78/81 podemos concluir que o grupo das imigrantes italianas foi o mais integrado.

Examinando todas as distâncias de Matusita relacionadas com o grupo das italianas no período 78/81 (ver tabela 3.1), temos que o grupo mais próximo destas, é o das espanholas, já que apresenta a menor distância. O segundo mais próximo é o das portuguesas seguido das japonesas e brasileiras, ou seja, no período 78/81 a ordem por grau de integração é IEPJB.

Este mesmo procedimento é feito para os 13 períodos de tempo e os resultados obtidos são apresentados na tabela 3.1. Através deles observamos que o grupo menos integrado foi o das brasileiras mas isso se deve ao fato de que o número de brasileiras é muito grande comparado com os demais grupos. Outros grupos menos integrado foram das japonesas e o das portuguesas, sendo que o grupo das japonesas sofreu uma alteração de comportamento no período 58/61 passando a ocupar o terceiro lugar. Por outro lado as italianas e espanholas, durante quase todos os períodos apresentaram-se como as mais integradas.

A figura 3.1 nos apresenta graficamente as alterações na ordem de integração tomando-se a distância dos grupos, com respeito ao grupo mais integrado situado na origem. Em outras palavras, a distância zero significa a distância do grupo mais integrado consigo mesmo.

TABELA 3.1 DISTÂNCIA DE MATUSITA PARA MULHERES

PERÍODO	B-I	B-E	B-P	B-J	I-E	I-P	I-J	E-P	E-J	P-J	ORDEM
78/81	1,012	1,005	0,886	0,657	0,007	0,142	0,389	0,135	0,382	0,248	IEPJB
74/77	0,977	0,969	0,792	0,551	0,009	0,206	0,459	0,196	0,450	0,256	IEPJB
70/73	0,865	0,864	0,668	0,370	0,002	0,213	0,516	0,211	0,515	0,308	IEPJB
66/69	0,745	0,654	0,498	0,354	0,097	0,259	0,405	0,163	0,309	0,147	IEPJB
62/65	0,556	0,419	0,393	0,383	0,141	0,167	0,176	0,026	0,037	0,010	IEPJB
58/61	0,381	0,327	0,350	0,343	0,055	0,031	0,038	0,024	0,017	0,007	IPJEB
54/57	0,357	0,504	0,374	0,322	0,044	0,017	0,036	0,028	0,080	0,052	EPIJB
50/53	0,558	0,677	0,574	0,233	0,126	0,018	0,330	0,108	0,454	0,348	EPIJB
46/49	0,760	1,018	0,610	0 100	0,288	0,160	0,667	0,445	0,931	0,514	EIPJB
42/45	0,708	0,712	0,468	0,043	0,004	0,251	0,668	0,255	0,672	0,426	EIPJB
38/41	0,586	0,599	0,378	0,038	0,014	0,214	0,550	0,228	0,563	0,341	EIPJB
34/37	0,513	0,464	0,249	0,171	0,051	0,268	0,346	0,218	0,296	0,079	IEPJB
30/33	0,348	0,362	0,073	0,156	0,014	0,276	0,501	0,290	0,515	0,229	EIPJB

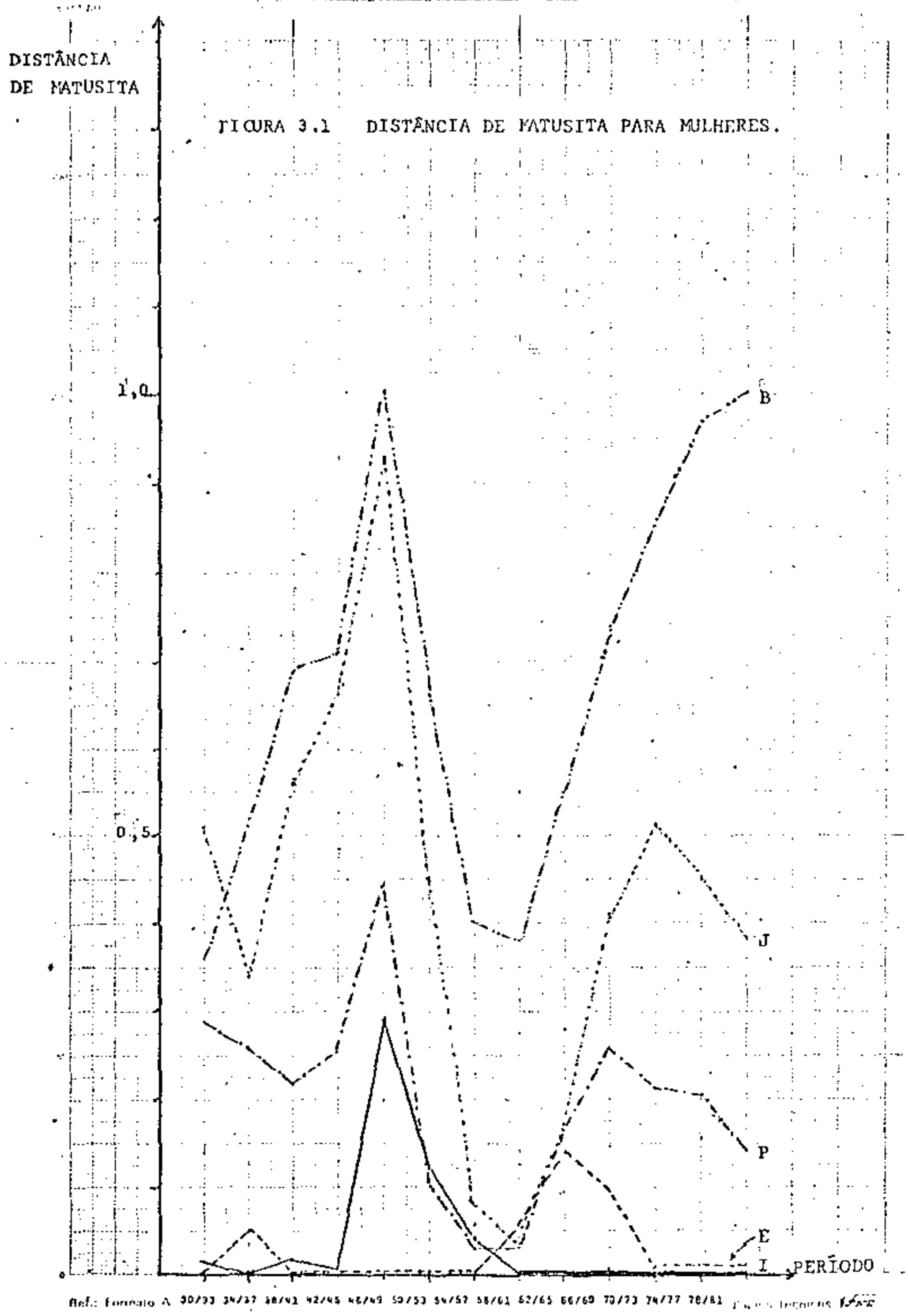

FIGURA 3.1 DISTÂNCIA DE MATUSITA PARA MULHERES.

b) ESTUDO DO COMPORTAMENTO DE INTEGRAÇÃO DOS HOMENS

Para a ordenação segundo o grau de integração dos homens procedemos de forma análoga ao caso anterior. Os resultados são apresentados na tabela 3.2, e deles podemos concluir que o grupo menos integrado foi o grupo dos brasileiros. Esse fato se deve também ao número muito grande de brasileiros comparado com o número de homens nos outros grupos.

Nesse período observamos que o comportamento dos japoneses muda bastante através dos anos pois, de 1930 a 1953, se apresentam como um grupo pouco integrado, passando a ter maior integração a partir de 1954 e chegando a ser o grupo mais integrado em 1958/1961.

A partir de 1970 passa, novamente, a ser um grupo pouco integrado. Os portugueses, ao contrário, apresentam-se como um grupo pouco integrado através de todo o período considerado e os italianos e espanhóis como os mais integrados durante quase todos os períodos.

c) ESTUDO DO COMPORTAMENTO DE INTEGRAÇÃO DOS GRUPOS

A tabela 3.3 nos apresenta a ordem de integração nos diferentes períodos. Nela observamos que os grupos dos italianos e espanhóis se apresentam como os mais integrados em quase todos os períodos, e que o grupo dos japoneses apresenta um alto grau de integração de 1954 a 1965; sobressaindo-se no período de 1958/61 .

TABELA 3.2 DISTÂNCIA DE MATUSITA PARA HOMENS

PERÍODO	B-I	B-E	B-P	B-J	I-E	I-P	I-J	E-P	E-J	P-J	ORDEM
78/81	1,169	1,165	1,048	0,880	0,005	0,145	0,336	0,139	0,331	0,192	IEPJB
74/77	1,122	1,116	0,962	0,829	0,007	0,187	0,335	0,181	0,328	0,149	IEPJB
70/73	1,035	1,041	0,852	0,722	0,007	0,208	0,347	0,215	0,354	0,141	EIPJB
66/69	0,962	0,912	0,749	0,792	0,057	0,235	0,189	0,179	0,132	0,047	IEJPB
62/65	0,877	0,772	0,663	0,852	0,116	0,231	0,028	0,116	0,088	0,204	IJEPB
58/61	0,817	0,732	0,683	0,845	0,092	0,145	0,030	0,053	0,122	0,175	JIEPB
54/57	0,792	0,750	0,678	0,728	0,045	0,122	0,068	0,078	0,023	0,054	IEJPB
50/53	0,922	0,956	0,838	0,603	0,039	0,093	0,345	0,132	0,383	0,252	EIPJB
46/49	1,034	1,082	0,885	0,511	0,057	0,169	0,562	0,226	0,616	0,398	EIPJB
42/45	0,937	0,915	0,777	0,420	0,025	0,177	0,545	0,142	0,521	0,373	IEPJB
38/41	0,874	0,829	0,739	0,530	0,050	0,147	0,367	0,097	0,318	0,221	IEPJB
34/37	0,844	0,743	0,689	0,507	0,110	0,168	0,357	0,058	0,248	0,191	IEPJB
30/33	0,724	0,591	0,509	0,259	0,140	0,226	0,476	0,086	0,339	0,254	IEPJB

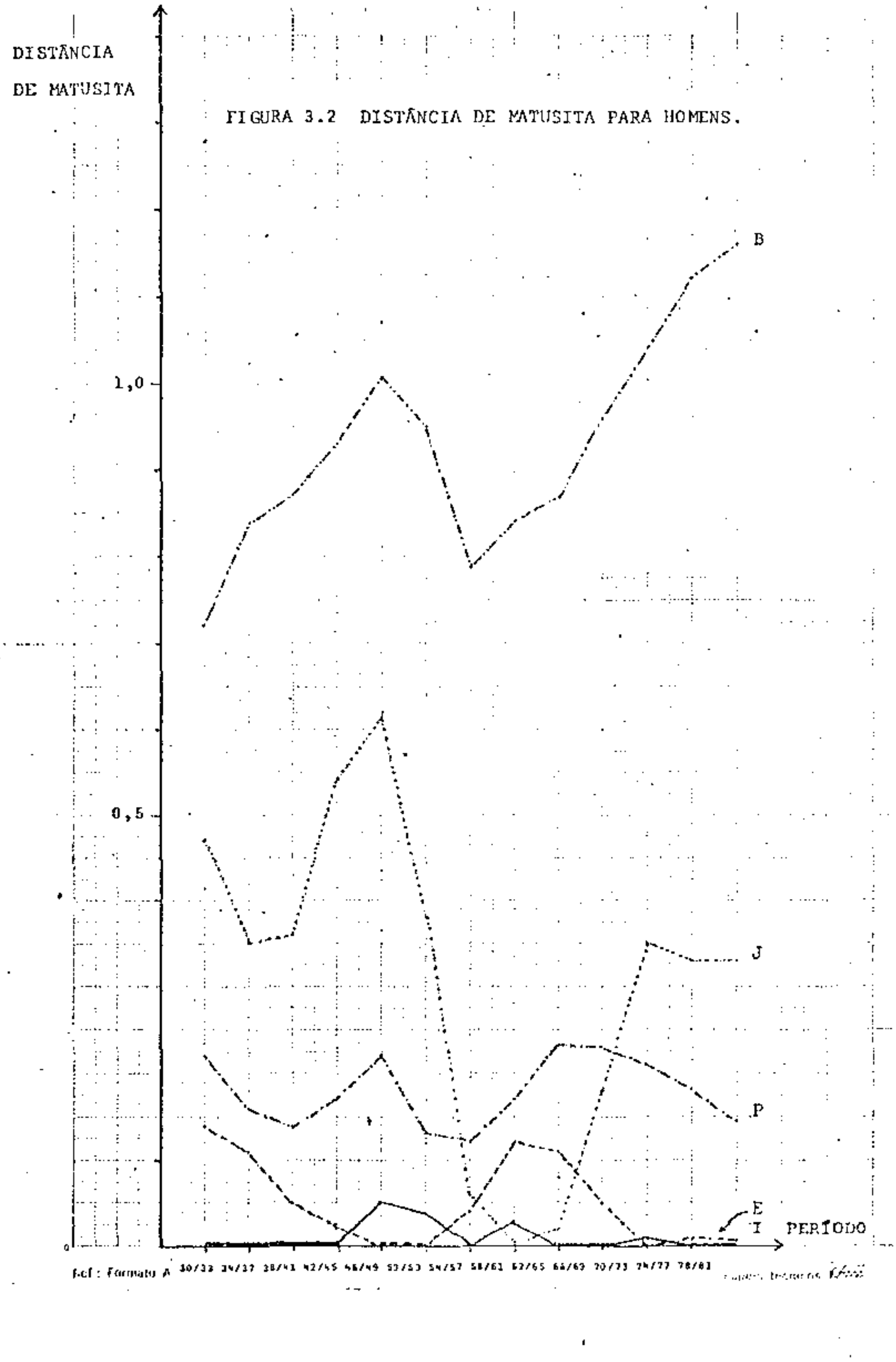
DISTÂNCIA
DE MATUSITA
FIGURA 3.2 DISTÂNCIA DE MATUSITA PARA HOMENS.
1,0
0,5
B
J
P
E
I
PERÍODO

TABELA 3.3 DISTÂNCIA DE MATUSITA PARA HOMENS E MULHERES

PERÍODO	B-I	B-E	B-P	B-J	I-E	I-P	I-J	E-P	E-J	P-J	ORDEM
78/81	1,136	1,133	1,040	0,887	0,004	0,114	0,289	0,110	0,284	0,175	IEPJB
74/77	1,090	1,085	0,958	0,818	0,006	0,154	0,309	0,148	0,303	0,156	IEPJB
70/73	1,004	1,008	0,851	0,690	0,004	0,173	0,346	0,177	0,350	0,174	EIPJB
66/69	0,923	0,873	0,730	0,730	0,057	0,213	0,213	0,156	0,156	0,001	IEPJB
62/65	0,818	0,715	0,632	0,771	0,112	0,199	0,052	0,088	0,060	0,148	IJEPB
58/61	0,731	0,654	0,623	0,747	0,081	0,114	0,018	0,032	0,099	0,132	JIEPB
54/57	0,716	0,694	0,635	0,659	0,024	0,086	0,060	0,062	0,036	0,026	IEJPB
50/53	0,860	0,905	0,806	0,548	0,049	0,060	0,332	0,110	0,381	0,273	EIPJB
46/49	0,955	1,044	0,829	0,427	0,102	0,140	0,558	0,242	0,655	0,422	EIPJB
42/45	0,866	0,855	0,711	0,325	0,013	0,170	0,562	0,156	0,549	0,397	IEPJB
38/41	0,777	0,753	0,642	0,401	0,026	0,145	0,392	0,119	0,367	0,249	IEPJB
34/37	0,707	0,633	0,544	0,392	0,079	0,172	0,327	0,093	0,249	0,156	IEPJB
30/33	0,573	0,498	0,351	0,080	0,078	0,228	0,496	0,150	0,419	0,272	IEPJB

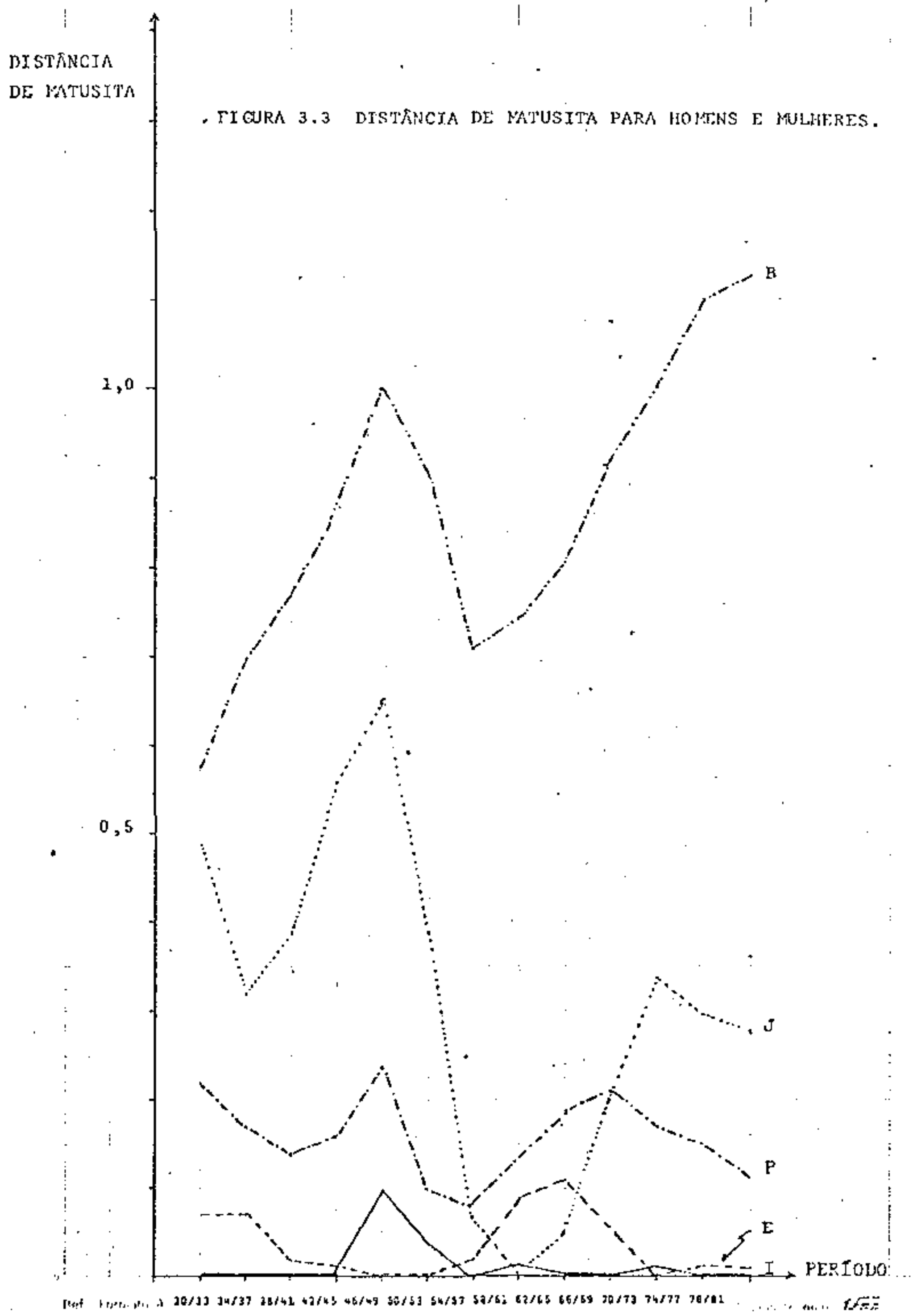

FIGURA 3.3 DISTÂNCIA DE MATUSITA PARA HOMENS E MULHERES.

Os portugueses se apresentam como um grupo pouco integrado seguido dos brasileiros.

Como vemos, estas conclusões coincidem com as já mencionadas, no caso do sexo masculino.

Se nas tabelas 3.1, 3.2 e 3.3 ignorássemos o grupo dos brasileiros, a ordenação quanto ao grau de integração dos imigrantes não seria, em geral, afetada.

3.1.2 APLICAÇÃO DO TESTE T

Aplicaremos o teste T para analisar quais grupos apresentaram o mesmo comportamento de integração considerando-se os 4 tipos de casamentos: endogamia, casamentos de homens fora do grupo; casamentos de mulheres fora do grupo; e casamentos de homens e mulheres fora do grupo; isto é, aplicaremos o teste T para testar se existe diferença significativa entre os índices de diversidade para 2 grupos étnicos, considerando-se os diversos tipos de casamento.

CASO 1

H_o : Não existe diferença entre os índices de diversidade nos períodos para os grupos (i,j) considerando-se casamentos dentro do grupo (endogamia).

Para o cálculo da estatística T, vamos utilizar a seguinte informação obtida com os dados da tabela 1.2.

TABELA 3.4

NACIONALIDADE	BRASILEIRO	ITALIANO	ESPANHÓIS	PORTUGUESES	JAPONESES
$\Sigma f_i \ell n f_i$	3621939,50	5425,44	4109,04	26745,72	5233,21
$\Sigma f_i (\ell n f_i)^2$	37950295,00	26219,11	18987,77	161092,41	24539,54
n_i	346910,50	1144,50	907,50	4471,25	1126,00

As estatísticas T calculadas para esta hipótese são apresentadas na tabela 3.8. Através delas chegamos às seguintes conclusões: a um nível de significância de 5% não existe diferença significativa entre os índices de diversidade nos períodos para os grupos brasileiros e italianos considerando casamento dentro do grupo, ou seja, a diversidade nos períodos, no grupo de brasileiros é igual a dos italianos.

Da mesma maneira, vemos que não existe diferença significativa entre os índices de diversidadade nos períodos para os grupos brasileiros e espanhóis, e entre italianos e espanhóis.

CASO 2

H_o : Não existe diferença entre os índices de diversidade nos períodos para os grupos (i,j) considerando-se casamento de homens fora do grupo.

Para o cálculo da estatística T vamos utilizar a seguinte informação (calculada através dos dados da tabela 1.2).

TABELA 3.5

NACIONALIDADE	BRASILEIRO	ITALIANO	ESPANHÓIS	PORTUGUESES	JAPONESES
$\Sigma f_i \ln f_i$	37529,37	23834,37	14006,27	52447,10	8879,13
$\Sigma f_i (\ln f_i)^2$	232894,38	138143,96	74920,89	339704,03	46388,71
n_i	6062,25	4119,00	2623,75	8109,50	1716,00

As estatísticas T são apresentadas na tabela 3.8. Nela podemos observar o seguinte: a um nível de significância de 5% não existe diferença significativa entre os índices de diversidade nos períodos para os grupos de italianos e espanhóis considerando casamento de homens fora do grupo.

A um nível de significância de 5% vemos que não existe diferença significativa entre os índices de diversidade nos períodos para os grupos italianos e portugueses considerando-se casamento de homens fora do grupo. De igual maneira, a um nível de 5% não existe diferença significativa entre os índices de diversidade para os grupos espanhóis e portugueses.

A um nível de significância de 1% vemos que não existe diferença significativa entre os índices de diversidade nos períodos para os grupos brasileiros e espanhóis. A um nível de significância de 0,02% vemos que não existe diferença significativa entre os índices de diversidade nos períodos para os grupos brasileiros e portugueses.

CASO 3

H_o : Não existe diferença entre os índices de diversidade nos períodos para os grupos (i,j) considerando casamento de mulheres fora do grupo.

De modo análogo aos casos anteriores temos a informação dada por:

TABELA 3.6

NACIONALIDADE	BRASILEIRA	ITALIANA	ESPANHOLAS	PORTUGUESAS	JAPONESAS
$\Sigma\ f_i\ \ell n\ f_i$	113762,75	7197,20	5999,62	20071,78	1748,96
$\Sigma\ f_i\ (\ell n\ f_i)^2$	813503,27	34478,17	28162,82	114513,65	6907,92
n_i	15928,00	1505,75	1286,00	3534,00	452,00

E pela tabela 3.8 na linha referente ao caso 3, podemos observar que nós aceitamos a hipótese nula em dois casos: no primeiro, a um nível de significância de 5%, aceitamos que não existe diferença significativa entre os índices de diversidade nos períodos para os grupos brasileiras e italianas, considerando casamento de mulheres fora do grupo; no outro caso, a um nível de significância de 5%, aceitamos que não existe diferença significativa entre os índices de diversidade nos períodos para os grupos espanholas e portuguesas, considerando casamento de mulheres fora do grupo.

Para os outros pares de grupos, a hipótese nula é rejeitada.

CASO 4

H_o : Não existe diferença significativa entre os índices de diversidade nos períodos para os grupos (i,j) considerando o total de casamentos fora do grupo.

Para o cálculo da estatística T vamos utilizar os dados da tabela 3.7 (calculada através dos dados obtidos da tabela 1.2).

TABELA 3.7

NACIONALIDADE	BRASILEIRO	ITALIANO	ESPANHOL	PORTUGUÊS	ESPANHOL
$\Sigma f_i \ell n f_i$	164181,66	34266,35	22418,51	79629,36	11720,92
$\Sigma f_i(\ell n f_i)^2$	1227200,40	208995,93	128739,70	545525,42	63917,05
n_i	21990,25	5624,75	3909,75	11643,50	2168,00

As estatísticas T são apresentadas na quarta linha da tabela 3.8 e nela observamos que aceitamos a hipótese nula, a um nível de significância de 0.1% aceitamos que não existe diferença significativa entre os índices de diversidade nos períodos para os grupos brasileiros e italianos, e entre os espanhóis e portugueses considerando-se o total de casamentos fora do grupo.

A um nível de significância de 5% nós aceitamos a hipótese nula de que não existe diferença significativa entre os índices de diversidade nos períodos dos brasileiros e espanhóis, e italianos e espanhóis, considerando-se os casamentos fora do grupo e a um nível de significância de 0.5% aceitamos a hipótese nula para brasileiros e portugueses.

TABELA 3.8 ESTATISTICA T

CASO \ GRUPOS (i,j)	B - I	B - E	B - P	B - J	I - E	I - P	I - J	E - P	E - J	P - J
1 ENDOGAMIA (CASAMENTOS DENTRO DO GRUPO)	T= 0,71 v= 1231 $t_{0,05,v}$= 1,96 então ns	T= 1,54 v= 917 $t_{0,05,v}$=1,963 então ns	T= 14,28 v= 4153 $t_{0,05,v}$= 1,96 então s	T= 4,76 v= 1141 $t_{0,05,v}$=1,96 então s	T= 0,67 v= 1562 $t_{0,05,v}$=1,96 então ns	T= 5,82 v= 1436 $t_{0,05,v}$= 1,96 então s	T= 3,25 v= 1353 $t_{0,05,v}$=1,96 então s	T= 6,19 v= 1135 $t_{0,05,v}$= 1,96 então s	T= 3,63 v= 1533 $t_{0,05,v}$= 1,96 então s	T= 3,53 v= 1926 $t_{0,05,v}$= 1,96 então s
2 HOMENS CASADOS FORA DO GRUPO	T= 3,31 v= 10219 $t_{0,05,v}$= 1,96 então s	T= 2,44 v= 6213 $t_{0,05,v}$= 1,96 então s $t_{0,01,v}$= 2,57 então ns	T= 2,95 v= 11448 $t_{0,05,v}$= 1,96 então s $t_{0,002,v}$= 3,09 então ns	T= 19,54 v= 2075 $t_{0,05,v}$= 1,96 então s	T= 0,47 v= 5494 $t_{0,05,v}$= 1,96 então ns	T= 0,78 v= 5575 $t_{0,05,v}$= 1,96 então ns	T= 20,50 v= 2027 $t_{0,05,v}$= 1,96 então s	T= 0,14 v= 4563 $t_{0,05,v}$= 1,96 então ns	T= 19,50 v= 2222 $t_{0,05,v}$= 1,96 então s	T= 22,62 v= 1491 $t_{0,05,v}$= 1,96 então s
3 MULHERES CASADAS FORA DO GRUPO	T= 0,60 v= 1914 $t_{0,05,v}$= 1,96 então ns	T= 3,60 v= 1387 $t_{0,05,v}$= 1,96 então s	T= 6,40 v= 4217 $t_{0,05,v}$= 1,96 então s	T= 10,99 v= 457 $t_{0,05,v}$=1,965 então s	T= 2,68 v= 2271 $t_{0,05,v}$= 1,96 então s	T= 5,38 v= 4532 $t_{0,05,v}$= 1,96 então s	T= 10,90 v= 497 $t_{0,05,v}$=1,965 então s	T= 0,29 v= 2266 $t_{0,05,v}$= 1,96 então ns	T= 9,87 v= 595 $t_{0,05,v}$=1,964 então s	T= 9,17 v= 507 $t_{0,05,v}$=1,965 então s
4 TOTAL DE CASAMENTOS FORA DO GRUPO	T= 3,26 v= 10412 $t_{0,05,v}$= 1,96 então s $t_{0,001,v}$=3,29 então ns	T= 1,17 v= 5868 $t_{0,05,v}$= 1,96 então ns	T= 2,77 v= 21395 $t_{0,05,v}$= 1,96 então s $t_{0,005,v}$= 2,81 então ns	T= 21,47 v= 2277 $t_{0,05,v}$= 1,96 então s	T= 1,23 v= 7913 $t_{0,05,v}$= 1,96 então ns	T= 5,55 v= 14265 $t_{0,05,v}$= 1,96 então s	T= 21,97 v= 2451 $t_{0,05,v}$= 1,96 então s	T= 3,33 v= 8552 $t_{0,05,v}$= 1,96 então s $t_{0,001,v}$=3,29 então ns	T= 20,07 v= 2547 $t_{0,05,v}$= 1,96 então s	T= 22,10 v= 2437 $t_{0,05,v}$= 1,96 então s

OBSERVAÇÃO: s: Significativo
ns: Não significativo

3.1.3 APLICAÇÃO DOS TESTES DE KULLBACK

Até aqui estudamos o problema da integração entre os imigrantes comparando apenas dois grupos étnicos.

Façamos agora uma análise de integração a nível dos cinco grupos aplicando para isso o teste de homogeneidade de Kullback bem como os testes de interação e independência de Kullback para verificarmos se os fatores Período, Nacionalidade e Tipo de casamento, serão independentes ou se existe algum tipo de interação. Para isso, olhemos as duas tabelas de contingência obtidas da tabela 1.2.

Na primeira tabela os fatores considerados são:

Período de tempo, L, com 13 categorias; Nacionalidade, C,com 5 categorias: 1- brasileiro, 2- italiano, 3- espanhol, 4- português, 5- japonês; Tipo de casamento, D, com 3 categorias: 1- casamento dentro do grupo, 2- casamento de homens fora do grupo, 3- casamento de mulheres fora do grupo.

E aplicando os testes de Kullback nos dados dessa tabela de contingência, obtemos os resultados que se seguem (H, HC, I significam homogeneidade, homogeneidade condicional e interação respectivamente, e S e NS, significativa e não significativa respecticamente).

TABELA 3.9

COMPONENTE	INFORMAÇÃO	G. L. (v)	$\chi^2_{v,0.05}$	$\chi^2_{v,0.01}$
H - NACIONALIDADE	12805,00	48	65,17 S	73,68 S
H.C. - TIPO CASAMENTOS / NACIONALIDADE	9271,80	120	146,57 S	158,95 S
H.C. - TIPO DE CASAMENTO / BRASILEIRO	7234,96	24	36,42 S	42,98 S
H.C. - TIPO DE CASAMENTO / ITALIANO	379,89	24	36,42 S	42,98 S
H.C. - TIPO DE CASAMENTO / ESPANHOL	510,52	24	36,42 S	42,98 S
H.C. - TIPO DE CASAMENTO / PORTUGUÊS	915,74	24	36,42 S	42,98 S
H.C. - TIPO DE CASAMENTO / JAPONÊS	230,69	24	36,42 S	42,98 S
H - (NACIONALIDADE, TIPO CASAMENTO)	22076,80	168	168,61 S	181,84 S

CONCLUSÕES: A um nível de significância de 1% e 5%

1) Não existe homogeneidade entre as nacionalidades;

2) Os resultados de Nacionalidade, Tipo de casamento não são homogêneos;

3) Não existe homogeneidade condicional para o Tipo de casamento dado a Nacionalidade;

4) Não existe homogeneidade condicional para os tipos de casamento dado cada uma das nacionalidades.

TABELA 3.10

COMPONENTE	INFORMAÇÃO	G.L. (v)	$\chi^2_{v,0.05}$	$\chi^2_{v,0.01}$
H - TIPO DE CASAMENTO	13231,60	24	36,41 S	42,98 S
H.C. - NACIONALIDADE / TIPO CASAMENTO	8845,20	144	168,61 S	181,84 S
H.C. - NACIONALIDADE / ENDOGAMIA	7041,90	48	65,17 S	73,68 S
H.C. - NACION. / CAS. HOMENS F. DO G.	1076,11	48	65,17 S	73,78 S
H.C. - NACION. / CAS. MULHERES F. DO G.	727,19	48	65,17 S	73,68 S
H - NACIONALIDADE / TIPO DE CASAMENTO	22076,80	168	ANALISADO	TAB 3.9

CONCLUSÕES: A um nível de significância de 1% e 5%

1) Os resultados dos tipos de casamentos não são homogêneos;

2) Não existe homogeneidade condicional para as nacionalidades dado o Tipo de casamento;

3) Não existe homogeneidade condicional entre as nacionalidades dado a endogamia;

4) Não existe homogeneidade condicional entre as nacionalidades dado o casamento de homens fora do grupo;

5) Não existe homogeneidade condicional entre as nacionalidades dado casamento de mulheres fora do grupo.

As três últimas conclusões complementam as obtidas pelo teste T .

TABELA 3.11

COMPONENTE	INFORMAÇÃO	G.L. (v)	$\chi^2_{v,0.05}$	$\chi^2_{v,0.01}$
H_2 (NACIONALIDADE x TIPO DE CASAMENTO)	90948,40	8	14,06 S	18,47 S
H_2 (PERÍODO x (NACION., TIPO CASAMENTO))	22076,80	168	168,61 S	181,84 S
H_2 (PERÍODO x NACION. x TIPO CASAMENTO)	113025,20	176	168,61 S	181,84 S

CONCLUSÕES: A um nível de significância de 1% e 5%

1) Rejeitamos a hipótese de independência entre Nacionalidade e Tipo de casamento;

2) Rejeitamos a independência entre os 3 fatores;

3) Também rejeitamos a hipótese de independência entre o fator Período e o par (Nacionalidade, Tipo de casamento).

TABELA 3.12

COMPONENTE	INFORMAÇÃO	G.L. (v)	$\chi^2_{v,0.05}$	$\chi^2_{v,0.01}$
H_2 (PERÍODO x TIPO DE CASAMENTO)	13231,60	24	36,41 S	42,98 S
H_2 ((PERÍODO / TIPO CASAMENTO) x (NACION. / TIPO CASAMENTO))	8845,20	144	168,61 S	181,84 S
H_2 (PERÍODO x (NACIONALIDADE, TIPO CASAMENTO))	22076,80	168	ANALISADO	TAB. 3.11
H_2 (PERÍODO x NACIONALIDADE)	12805,00	48	65,17 S	73,68 S
H_2 ((PERÍODO / NACIONALIDADE) x (TIPO CASAM. / NACIONALIDADE))	9805,00	120	146,56 S	158,95 S

CONCLUSÃO: A um nível de significância de 1% e 5%

1) Rejeitamos a hipótese de independência entre o Período e Tipo de casamento;

2) Rejeitamos a hipótese de independência condicional entre o Período e Nacionalidade dado o fator Tipo de casamento;

3) Rejeitamos a hipótese de independência entre o Período e Nacionalidade;

4) Rejeitamos a hipótese de independência condicional entre o Período e Tipo de casamento dado o fator Nacionalidade.

TABELA 3.13

COMPONENTE	INFORMAÇÃO	G.L. (v)	$\chi^2_{v,0.05}$	$\chi^2_{v,0.01}$
I - (PERÍODO, TIPO DE CASAMENTO)	3365,00	24	38,88 S	45,64 S
I - ((PERÍODO, TIPO CASAMENTO), NACION.)	5906,80	96	119,87 S	131,14 S
H.C. - (TIPO CASAMENTO / NACIONALIDADE)	9271,80	120	ANALISADO	TAB. 3.9

CONCLUSÃO: A um nível de significância de 1% e 5%

1) Rejeitamos a hipótese de que existe interação entre Período de tempo e Tipo de casamento;

2) Rejeitamos a hipótese de que existe interação entre o par (Período, Tipo de casamento) e Nacionalidade.

TABELA 3.14

COMPONENTE	INFORMAÇÃO	G.L. (v)	$\chi^2_{v,0.05}$	$\chi^2_{v,0.01}$
I - (PERÍODO, NACIONALIDADE)	3047,20	48	65,17 S	73,68 S
I - ((PERÍODO, NACION.), TIPO CASAMENTO)	5798,00	96	119,87 S	131,14 S
H.C. - NACIONALIDADE / TIPO CASAMENTO	8845,20	144	ANALISADO	TAB.3.10

CONCLUSÃO: A um nível de significância de 1% e 5%

1) Rejeitamos a hipótese de que existe interação entre o Período e Nacionalidade;

2) Rejeitamos a hipótese de que existe interação entre o par (Período, Nacionalidade) e Tipo de Casamento.

Consideremos agora os seguintes fatores e suas categorias para uma tabela de contingência:

Período de Tempo, L, com 13 categorias; Nacionalidade, C, com 5 categorias: 1- brasileiro, 2- italiano, 3- espanhol, 4- português, 5- japonês; Tipo de Casamento, D, com 2 categorias: 1- casamentos dentro do grupo, 2- total de casamentos fora do grupo.

Nesse caso obtemos os seguintes resultados aplicando o teste de homogeneidade de Kullback:

TABELA 3.15

COMPONENTE	INFORMAÇÃO	G.L. (v)	$\chi^2_{v,0.05}$	$\chi^2_{v,0.01}$
H - TIPO DE CASAMENTO	13216,60	12	21,02 S	26,22 S
H.C. - NACIONALIDADE / TIPO CASAMENTO	8439,60	96	119,87 S	131,14 S
H.C. - NACIONALIDADE / ENDOGAMIA	7041,80	48	65,17 S	73,68 S
H.C. - NACIONALIDADE / CASAM. FORA GRUPO	1397,80	48	65,17 S	73,68 S
H - NACIONALIDADE, TIPO CASAMENTO	21656,20	108	133,25 S	145,09 S

CONCLUSÃO: A um nível de significância 1% e 5%

1) A conclusão 2 da tabela 3.9 e as conclusões 1, 2, 3 da tabela 3.10 são confirmadas aqui;

2) Não existe homogeneidade condicional entre as nacionalidades dado casamento fora do grupo. (Esta informação completa a informação já obtida mediante o Teste T.)

3.2. APLICAÇÃO DA METODOLOGIA NO ESTUDO DOS MODOS DE OBTENÇÃO DOS ALIMENTOS NAS ALDEIAS ALANTESU E JUINA

Para o estudo e análise dos modos de obtenção dos ítens alimentares nas aldeias Alantesu e Juina aplicamos o Teste T e os testes de Kullback. Seus resultados serão apresentados nas sub-secções 3.2.1 e 3.2.2.

3.2.1. APLICAÇÃO DO TESTE T

De modo análogo ao problema anterior aplicamos aqui o teste T a fim de detectar possíveis influências das estações nos modos de obtenção dos alimentos.

Para o cálculo da estatística T vamos fazer uso da seguinte informação obtida dos dados da tabela 1.6.

TABELA 3.16

ALDEIAS	ALANTESU			JUINA		
ESTAÇÃO	SECA	CHUVOSA	TOTAL ESTAÇÃO	SECA	CHUVOSO	TOTAL ESTAÇÃO
n_i	215,00	200,00	415,00	261,00	210,00	471,00
$\Sigma f_i \ell n f_i$	384,07	351,57	863,24	515,22	384,83	1034,71
$\Sigma f_i (\ell n f_i)^2$	696,34	640,77	1817,94	1042,62	718,64	2308,00

Essa estatística será agora utilizada para a análise do problema alimentar considerando-se as estações: seca; chuvosa; seca e chuvosa, isto é, queremos testar se existe diferença entre os índices de diversidade nos modos de obtenção dos alimentos considerando-se as diversas estações.

CASO 1

H_o : Não existe diferença entre os índices de diversidade nos modos de obtenção dos ítens alimentares das aldeias Alantesu e Juina, considerando-se a estação seca.

A estatística T e seus respectivos graus de liberdade são dados por $T = 4,2369016$ e $v = 464$. E a um nível de significância de 5% temos: $T_{0.05(2),464} = 1.965$.

CONCLUSÃO: A um nível de significância de 5% nós rejeitamos a hipótese nula, H_o, ou seja, na estação seca existe diferença significativa entre os índices de diversidade nos modos de obtenção dos ítens alimentares das aldeias Alantesu e Juina.

CASO 2

H_o : Não existe diferença entre os índices de diversidade nos modos de obtenção dos ítens alimentares das aldeias Alantesu e Juina considerando-se a estação chuvosa.

A estatística T e seus graus de liberdade são: $T = 1,5603835$ e $v = 410$. A um nível de significância de 5% temos: $T_{0.05(2),410} = 1.966$.

CONCLUSÃO: A um nível de significância de 5% nós aceitamos , H_o, ou seja, na estação chuvosa não existe diferença significativa entre os índices de diversidade.

CASO 3

H_o : Não existe diferença significativa entre os índices de diversidade nos modos de obtenção dos ítens alimentares das aldeias Alantesu e Juina considerando-se a variável estação (seca + chuvosa).

A estatística T e seus respectivos graus de liberdade são : $T = 3{,}6467994$ e $v = 885$. A um nível de significância de 5%, temos: $T_{0.05(2),885} = 1.963$.

CONCLUSÃO: A um nível de significância de 5% nós rejeitamos H_0, ou seja, existe diferença significativa entre os Índices de diversidade nos modos de obtenção dos ítens alimentares das duas aldeias considerando-se a estação seca e chuvosa.

3.2.2. APLICAÇÃO DOS TESTES DE KULLBACK

Aplicaremos agora os testes de Kullback para examinarmos a existência de homogeneidade, independência e interações entre os seguintes fatores:

L : Modo de obtenção dos ítens, com quatro categorias;

C : Aldeia, com 2 categorias: Alantesu (1) e Juina (2);

D : Estação, com 2 categorias: Seca (1) e Chuvosa (2).

Na aplicação dos testes de Kullback obtivemos os seguintes resultados:

TABELA 3.17

COMPONENTE	INFORMAÇÃO	G.L. (v)	$\chi^2_{v,0.05}$	$\chi^2_{v,0.01}$
H_2 (ALDEIA x ESTAÇÃO)	1,15	1	3,84 NS	6,63 NS
H_2 (MODO OBTENÇÃO x (ALDEIA,ESTAÇÃO))	124,10	9	16,92 S	21,67 S
H_2 (MODO OBTENÇÃO x ALDEIA x ESTAÇÃO)	125,25	10	18,31 S	23,21 S

CONCLUSÃO: A um nível de significância de 1% e 5%

1) Não existe evidência significativa para rejeitar a hipótese de independência entre os fatores Aldeia e Estação;

2) Rejeitamos a hipótese de independência entre o fator Modo de obtenção dos ítens alimentares e o par (Aldeia, Estação);

3) Rejeitamos a hipótese de independência entre os três fatores.

TABELA 3.18

COMPONENTE	INFORMAÇÃO	G.L. (v)	$\chi^2_{v,0.05}$	$\chi^2_{v,0.01}$
H_2 (MODO OBTENÇÃO x ESTAÇÃO)	17,84	3	7,81 S	[illegible]
H_2 ((MODO OBTENÇÃO / ESTAÇÃO) x (ALDEIA / ESTAÇÃO))	106,25	6	12,59 S	16,81 S
H_2 (MODO OBTENÇÃO x (ALDEIA, ESTAÇÃO))	124,09	9	ANALISADO	TAB. 3.17
H_2 (MODO OBTENÇÃO x ALDEIA)	95,79	3	7,81 S	11,30 S
H_2 ((MODO OBTENÇÃO / ALDEIA) x (ESTAÇÃO / ALDEIA))	28,30	6	12,59 S	16,81 S

CONCLUSÃO: A um nível de significância de 1% e 5%

1) Rejeitamos a hipótese de independência entre Modo de obtenção dos ítens alimentares e Estação;

2) Rejeitamos a hipótese de independência condicional entre Modo de obtenção dos ítens alimentares e Aldeia dado o fator Estação;

3) Rejeitamos a hipótese de independência condicional entre Modo de obtenção de alimentos e Aldeia;

4) Rejeitamos a hipótese de independência condicional entre Modo de obtenção dos alimentos e Estações dado o fator Aldeia.

TABELA 3.19

COMPONENTE	INFORMAÇÃO	G.L. (v)	$\chi^2_{v,0.05}$	$\chi^2_{v,0.01}$
H - (ALDEIA)	95,79	3	7,81 S	11,30 S
H.C. - (ESTAÇÕES / ALDEIA)	28,30	6	12,60 S	16,80 S
H.C. - (ESTAÇÕES/ALANTESU)	1,01	3	7,81 S	11,30 S
H.C. - (ESTAÇÕES / JUINA)	27,29	3	7,81 S	11,30 S
H - (ALDEIA, ESTAÇÕES)	124,09	9	16,90 S	21,70 S
H - (ESTAÇÕES)	17,84	3	7,81 S	11,30 S
H.C. - (ALDEIA / ESTAÇÕES)	106,25	6	12,60 S	16,80 S
H.C. - (ALDEIA / SECA)	84,00	3	7,81 S	11,30 S
H.C. - (ALDEIA / CHUVOSA)	22,25	3	7,81 S	11,30 S

CONCLUSÃO: A um nível de significância de 1% e 5%

1) Os resultados das Aldeias não são homogêneos;

2) Os resultados de Aldeia, Estações não são homogêneos;

3) Não existe homogeneidade condicional entre as estações dado o fator Aldeia;

4) Não existe homogeneidade condicional entre as estações dado a aldeia Alantesu:

5) Não existe homogeneidade condicional entre as estações dado a aldeia Juina;

6) Os resultados das estações não são homogêneos;

7) Não existe homogeneidade condicional entre as aldeias dado o fator Estação;

8) Não existe homogeneidade condicional entre os resultados das aldeias dado Seca;

9) Não existe homogeneidade condicional entre as aldeias dado a estação chuvosa.

TABELA 3.20

COMPONENTE	INFORMAÇÃO	G.L. (v)	$\chi^2_{v,0.05}$	$\chi^2_{v,0.01}$
I - (MODO OBTENÇÃO, ESTAÇÕES)	15,54	3	7,81 S	11,30 S
I - ((MODO OBTENÇÃO,ESTAÇÃO), ALDEIA)	12,76	3	7,81 S	11,30 S
H.C. - (ESTAÇÕES / ALDEIA)	28,30	6	ANALISADO	TAB. 3.19

CONCLUSÃO: A um nível de significância de 1% e 5%

1) Rejeitamos a hipótese de que existe interação entre o Modo de obtenção dos ítens alimentares e Estações;

2) Rejeitamos a hipótese de que exista interação entre o par (Modo de obtenção dos ítens alimentares, Estação) e Aldeia.

TABELA 3.21

COMPONENTE	INFORMAÇÃO	G.L. (v)	$\chi^2_{v,0.05}$	$\chi^2_{v,0.01}$
I - (MODO DE OBTENÇÃO, ALDEIA)	93,39	3	7,81 S	11,30 S
I - ((MODO OBTENÇÃO, ALDEIA), ESTAÇÕES)	12,86	3	7,81 S	11,30 S
H.C. - (ALDEIA / ESTAÇÃO)	106,25	6	ANALISADO	TAB. 3.19

CONCLUSÃO: A um nível de significância de 1% e 5%

1) Rejeitamos a hipótese de que exista interação entre o Modo de obtenção dos ítens alimentares e Aldeia;

2) Rejeitamos a hipótese de que exista interação entre o par (Modo de obtenção dos ítens alimentares, Aldeia) e Estações.

TABELA 3.22

COMPONENTE	INFORMAÇÃO	G.L. (v)	$\chi^2_{v,0.05}$	$\chi^2_{v,0.01}$
H - ESTAÇÕES	17,84	3	ANALISADO	TAB. 3.19
I - (MODO OBTENÇÃO - ALDEIA - ESTAÇÃO)	10,46	3	7,81	11,30
H.C. - (ESTAÇÃO / ALDEIA)	28,30	6	ANALISADO	TAB. 3.19

CONCLUSÃO:

1) A um nível de significância de 5% nós rejeitamos a hipótese de que existe interação entre os três fatores. Por outro lado a um nível de significância de 1% não existe evidência significativa para rejeitar a hipótese de interação entre os três fatores.

CONCLUSÕES

Neste vamos apresentar as conclusões a que chegamos para os dois problemas tratados neste capítulo.

INTEGRAÇÃO DOS IMIGRANTES NA CIDADE DE SÃO PAULO

De um modo geral, podemos descrever o comportamento dos diversos grupos de imigrantes, no período de 1930 a 1981 por alto índice de integração para os italianos e espanhóis, seguido pelos japoneses e portugueses.

O grupo de homens japoneses foi o que apresentou maior oscilação quanto ao grau de integração, apresentando no período de 1954 a 1965 seu maior nível e chegando mesmo a ser o grupo mais integrado em 1958/61. E ao contrário do que se pode imaginar dos portugueses eles se mantiveram quase sempre no terceiro lugar de integração com exceção de 1957 a 1969 quando ocuparam o quarto lugar.

Por outro lado o grupo dos brasileiros sempre se apresentou como o menos integrado (isto foi devido ao número muito grande de brasileiros comparado com os demais grupos).

No grupo das mulheres o comportamento é semelhante havendo apenas uma troca de posições entre portuguesas e japonesas. O grupo das mulheres japonesas apresentou um comportamento estável de integração, mantendo-se quase sempre no quarto lugar com exceção de 1958/61 na qual ocuparam o terceiro lugar.

Considerando-se o total de casamentos fora do grupo, a um nível de significância de 5% não existe diferença significativa entre os índices de diversidade nos períodos para os grupos italianos e espanhóis. Isto também ocorre entre os brasileiros e espanhóis.

A um nível de significância de 0.1% não existe diferença significativa entre os índices de diversidade nos períodos para os grupos espanhóis e portugueses; e entre os brasileiros e italianos.

Finalmente, a um nível de significância de 0.5% não existe diferença significativa entre os índices de diversidade nos períodos para os grupos brasileiros e portugueses.

Fazendo uma análise individual, para homens e mulheres, observamos que nos casamentos de mulheres fora do grupo, não existe diferença significativa entre os índices de diversidade nos períodos para os grupos, das brasileiras e italianas. Isto também acontece entre as espanholas e portuguesas.

Considerando casamento de homens fora do grupo observamos que não existe diferença significativa entre os índices de diversidade nos períodos para os grupos de italianos e espanhóis, italianos e portugueses, espanhóis e portugueses, brasileiros e espanhóis e os brasileiros e portugueses.

Comparando os cinco grupos, observamos que as diferentes nacionalidades não são homogêneas e que também não existe homogeneidade condicional entre os grupos de diferentes nacionalidades dado o casamento de homens fora do grupo ou dado o casamento de mulheres fora do grupo ou dado o casamento total fora do grupo. Isso nos leva a concluir que não existe homogeneidade condicional entre os grupos étnicos dado o fator Tipo de casamento.

Examinando a homogeneidade entre os grupos dado o Tipo de casamento, observamos que não existe interação entre os fatores Período e Nacionalidade, e que não existe interação entre Tipo de casamento e o par (Período, Nacionalidade). Concluímos ainda que os tipos de casamento não são homogêneos e que também não são condicionalmente homogêneos dado o fator Nacionalidade ou dado uma nacionalidade especificamente.

Vimos que não existe interação entre Período e Tipo de casamento ou entre Nacionalidade e o par (Período, Tipo de casamento).

Por outro lado verificamos que os fatores Nacionalidade e Tipo de casamento; Período e Tipo de casamento; Período e Nacionalidade não são independentes e que também não existe uma independência condicional entre eles e nem entre Período e o par (Nacionalidade, Tipo de casamento).

Portanto, concluímos que não existe independência entre os três fatores, Período, Nacionalidade e Tipo de casamento.

ESTUDOS DOS MODOS DE OBTENÇÃO DOS ALIMENTOS NAS ALDEIAS ALANTESU E JUINA

Através dos resultados apresentados no capítulo anterior podemos concluir que não existe independência entre os pares de fatores, Modo de obtenção e Aldeia; Modo de obtenção e Estações, e nem mesmo independência condicional de cada par dado o terceiro fator.

Por outro lado concluímos que os fatores Aldeia e Estações são independentes mas, Aldeia, Estação e Modo de obtenção dos alimentos não são independentes.

Concluímos ainda que o comportamento das frequências durante a Seca e Chuva não são homogêneas e que também não são condicionalmente homogêneas dado o fator Aldeia ou dado uma aldeia especificamente.

Por outro lado, as aldeias Alantesu e Juina não são homogêneas e também não são condicionalmente homogêneas dado o fator Estação ou dado a estação Seca. A homogeneidade entre as aldeias Alantesu e Juina, dada a estação chuvosa não está bem caracterizada já que obtivemos resultados contraditórios nos dois testes aplicados, (Teste T e Teste de Homogeneidade de Kullback), mas, apesar disso, tendemos a acreditar que, de fato, não existe homogeneidade condicional uma vez que não obtivemos homogeneidade entre as aldeias dado o fator Estação e dado a estação Seca.

Do fato de não existir interação entre o Modo de obtenção dos alimentos e as Estações e entre as Aldeias e o par de fatores (Modo de obtenção dos alimentos e Estações), concluímos que não existe homogeneidade condicional entre as estações dado o fator Aldeia. Da mesma forma concluímos que não existe homogeneidade condicional entre as aldeias dado o fator Estação já que não existe interação entre os Modos de obtenção dos alimentos e Aldeia; e entre as Estações e o par (Modo de obtenção dos alimentos e Aldeias).

Finalizando, concluímos que existe interação entre os três fatores.

APÊNDICE

A DISTRIBUIÇÃO DA DIFERENÇA DE DOIS QUI QUADRADOS INDEPENDENTES

Neste apêndice, nós vamos provar o seguinte resultado.

RESULTADO:

Seja $Z = X_1 - X_2$, onde X_1 e X_2 são v.a$_s$. independentes que seguem uma distribuição qui quadrado com n_1 e n_2 graus de liberdade respectivamente. Então a função densidade de Z é dada por

$$g(z) = \begin{cases} \dfrac{|z|^{\frac{n_1+n_2}{4} - 1}}{2^{\frac{n_1+n_2}{2}} \Gamma(\frac{n_2}{2})} W_{(\frac{n_2-n_1}{4}),(\frac{n_1+n_2-2}{4})}(|z|), & z < 0 \\[2ex] \dfrac{z^{\frac{n_1+n_2}{4} - 1}}{2^{\frac{n_1+n_2}{4}} \Gamma(\frac{n_1}{2})} W_{(\frac{n_1-n_2}{4}),(\frac{n_1+n_2-2}{4})}(z) & z > 0 \end{cases}$$

onde W é a função de Whittaker (ver Erdélyi (1954)) definida por

$$W_{k,\mu}(z) = e^{-\frac{z}{2}} \sum_{\mu,-\mu} \frac{\Gamma(-2\mu)}{\Gamma(\frac{1}{2}-\mu-k)} z^{\frac{1}{2}+\mu} \; {}_1F_1 \left(\frac{1}{2}+\mu-k;\ 2\mu+1;\ z\right)$$

onde ${}_1F_1\ (\frac{1}{2}+\mu-k\ ;\ 2\mu+1;\ z)$ é a série hipergeométrica de Kummer, definida por

$$
{}_1F_1\ (\frac{1}{2}+\mu-k;\ 2\mu+1;\ z) = \sum_{k=0}^{\infty} . \frac{(\frac{1}{2}+\mu-k)_k}{(2\mu+1)_k} . \frac{z^k}{k!}
$$

PROVA:

Sejam X_1 e X_2 v.a$_s$. independentes tais que $X_1 \sim \chi^2_{(n_1)}$ e $X_2 \sim \chi^2_{(n_2)}$.

Logo

$$
f(x_1,x_2) = \frac{x_1^{\frac{n_1}{2}-1} e^{-\frac{x_1}{2}}}{2^{\frac{n_1}{2}}\ \Gamma(\frac{n_1}{2})} \quad \frac{x_2^{\frac{n_2}{2}-1} e^{-\frac{x_2}{2}}}{2^{\frac{n_2}{2}}\ \Gamma(\frac{n_2}{2})}, \qquad x_1 > 0, \qquad x_2 > 0 .
$$

Seja a seguinte transformação

$$
Z = X_1 - X_2 \qquad e \qquad V = X_2
$$

de onde obtemos que a função de densidade conjunta de (Z,V) é

$$
g(v,z) = \frac{(z+v)^{\frac{n_1}{2}-1}\ v^{\frac{n_2}{2}-1}\ e^{-\frac{z}{2}-v}}{2^{\frac{n_1+n_2}{2}}\ \Gamma(\frac{n_1}{2})\ \Gamma(\frac{n_2}{2})}
$$

Logo a função de densidade marginal de Z será

$$g(z) = \begin{cases} \dfrac{e^{-\frac{z}{2}}}{2^{\frac{n_1+n_2}{2}}\,\Gamma(\frac{n_1}{2})\,\Gamma(\frac{n_2}{2})} \displaystyle\int_{-z}^{\infty} v^{\frac{n_2}{2}-1}\, e^{-v}\, (v+z)^{\frac{n_1}{2}-1}\, dv & \text{para } -v < z < 0 \\[2ex] \dfrac{e^{-\frac{z}{2}}}{2^{\frac{n_1+n_2}{2}}\,\Gamma(\frac{n_1}{2})\,\Gamma(\frac{n_2}{2})} \displaystyle\int_{0}^{\infty} v^{\frac{n_2}{2}-1}\, e^{-v}\, (v+z)^{\frac{n_1}{2}-1}\, dv & \text{para } z > 0,\ v > 0 \end{cases}$$

Mediante as tabelas de integral transformadas de Erdélyi (1954), obtemos o resultado.

□

CASO ESPECIAL:

Sejam X_1, X_2 v.a$_s$. i.i.d. que seguem uma distribuição qui quadrado com n graus de liberdade, e seja $Z = X_1 - X_2$, então a função densidade de Z é dada por

$$g(z) = \frac{|z|^{\frac{n-1}{2}}}{\sqrt{\pi}\; 2^n\; \Gamma(\frac{n}{2})}\; K_{(\frac{n-1}{2})}\left(\frac{|z|}{2}\right) \qquad -\infty < Z < \infty$$

porque,

$$W_{o,\frac{n-1}{2}}(|z|) = \left(\frac{|z|}{\pi}\right)^{1/2} K_{\frac{n-1}{2}}\left(\frac{|z|}{2}\right)$$

onde $K_{\frac{n-1}{2}}\left(\frac{|z|}{2}\right)$ é a função de Bessel modificada de ordem $(n-1)/2$.

Fazendo a seguinte transformação

$$T = Z / 2$$

obtemos que

$$g(t) = \frac{|t/2|^{\frac{n-1}{2}}}{\sqrt{\pi}\ \Gamma(\frac{n}{2})} K_{\frac{n-1}{2}}(|t|) .$$

Este resultado coincide com o obtido por Pachares (1952) e por Springer (1979).

BIBLIOGRAFIA

[1] BASHARIN, G.P. (1959): On a statistical estimate for the entropy of a sequence of independent random variables. Theory Probability Application, 4, 333-336.

[2] COCHRAN, W.G. (1954): Some methods for strengthening, the commom χ^2 tests. Biometrics, 10, 417-451.

[3] CRAMER, H. (1945): Mathematical methods of statistics. Princeton University Press.

[4] ERDÉLYI, A. (1954): Tables of integral transforms. McGraw-Hill Book Company, Inc.

[5] EVERIT, B.S. (1977): The analysis contingency tables. John Wiley and Sons, Inc., Halsted Press, New York.

[6] FISHER, R.A. (1936): The Fiducial argument in Statistical inference. Annals Eugenic. IV, 391-398.

[7] GOLDSTEIN, M.; WOLF, E.; DILLON, W. (1976): On a test of independence for contingency tables. Communication Statistics. Theor. Meth., A5(2), 159-169.

[8] GOOD, I.J. (1957): Saddle-point methods for the multinomial distribution. Annals Mathematical Statistics, 28, 861-881.

[9] HUTCHESON, K. (1970): A test for comparing diversities based

on the Shannon formula. Journal of Theoretical Biology, 29, Nº 1, 151-154.

[10] KULLBACK, S. (1959): Information theory and Statistics. Jonh Wiley and Sons, Inc. London.

[11] LLOYD, L.; ZAR, J.; KARR, J. (1968): On the calculation of information theoretical measures of diversity. The American Midland Naturalist, 19, Nº 2, 257-272.

[12] MATHAI, A.M.; RATHIE, P.N. (1972): Characterization of Matusita's measure of affinity. Annals of the Institute of Statistical Mathematics, 24, Nº 3, 473-483.

[13] MATHAI, A.M.; RATHIE, P.N. (1975): Basic concepts in information theory and statistics. Wiley, Halsted Press, New York.

[14] McGILL, W.J. (1954): Multivariate information transmission. Psychometrika. 19, Nº 2, 97-116.

[15] PACHARES, J. (1952): The distribution of the difference of two independent Chi-Squares. Annals Mathematical Statistics, 23, pag. 639.

[16] PEARSON, K.; STOUFFER, S.A.; DAVID, F.N. (1932): Further applications in statistics of the $T_n(X)$ Bessel function. Biometrika, 14, 19-350.

[17] PIELOU, E.C. (1966): Species diversity and pattern diversity in the study of Ecology succession. Journal of Theoretical

Biology, 10, 370-383.

[18] SETZ, E. (1983): Ecologia alimentar em um grupo indígena: comparação entre aldeias Nambiquara de floresta e de cerrado. Tese de Mestrado. Instituto de Biológia - UNICAMP.

[19] SOKAL, R.R.; ROHLF, F.J. (1969): Biometry. San Francisco , Freeman.

[20] SPRINGER, M.D. (1979): The algebra of Random variables. John Wiley and Sons, Inc., Printed in the United States of America.

[21] VERDUGO, A.; RATHIE, P.N. (1979): An information measure and integration of immigrants. Anais 10ª Conferência Internacional de Biometria.

[22] WELCH, B.L. (1936): The specification of rules for rejecting too variable a product, with particular reference to an electric lamp problem. Journal Royal Statistics Sociate. Supplement III, Nº 1, 29-48.

[23] WELCH, B.L. (1938): The significance of the difference between two means when the population variances are unequal. Biometrika, 29, 350-362.

[24] WILKS, S.S. (1935): The likelihood of independence in contingency tables. Annals Mathematical Statistics, 6, 190-196.

[25] WILKS, S.S. (1962): Mathematical Statistics. John Wiley and