

カイ二乗検定(chi-square test, χ^2 -test)

市政に対する住民の評価について、市民1万人から無作為に抽出した市民1,000人を対象に調査を行ったところ、次のような結果になった(回答者475人)。

評価する	評価しない	合計
235名(49.5%)	240名(50.5%)	475名(100.0%)

回答のあった475名を居住地区によって分類してみた。(単位:人)

区分	東地区	西地区	合計
評価する	128	107	235
評価しない	116	124	240
合計	244	231	475

果たして住民の居住地区によって市政に対する評価に差があるといえるだろうか？

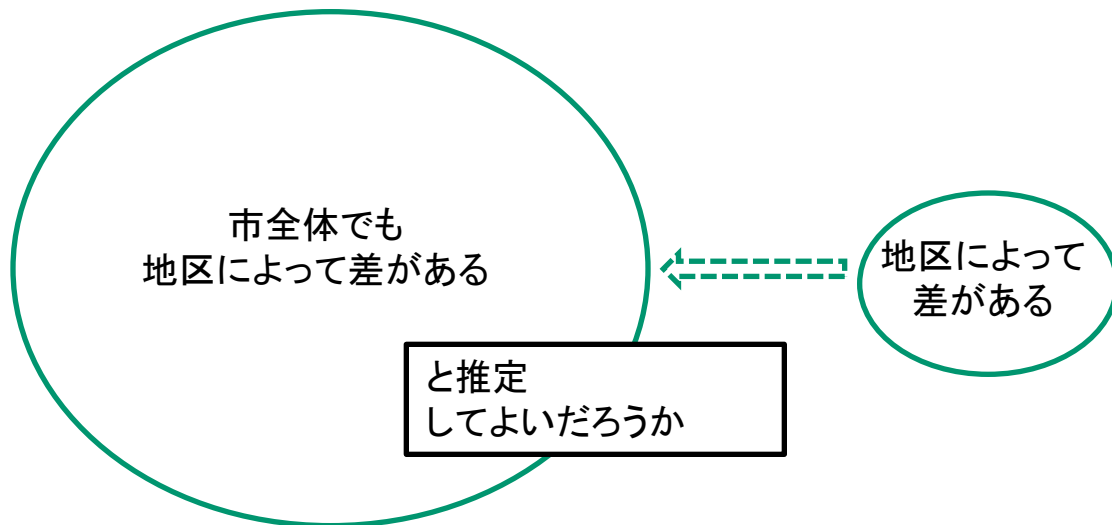
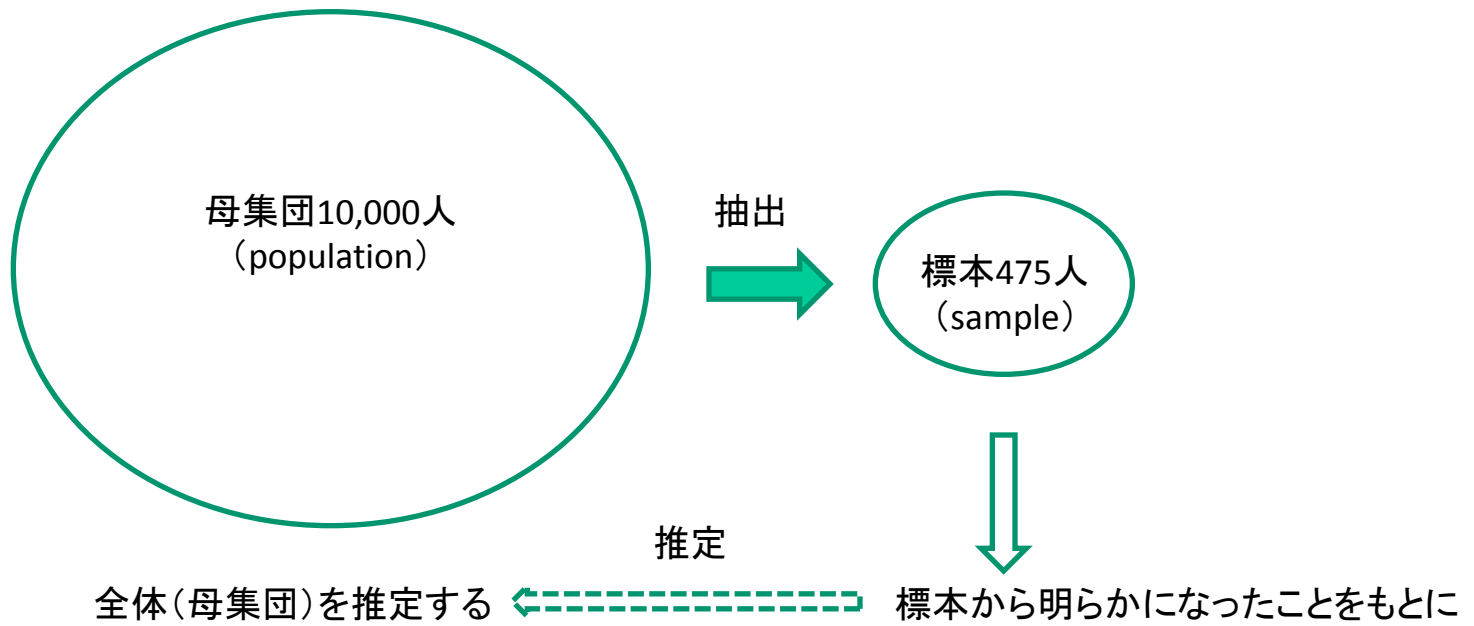
【百分率構成比】

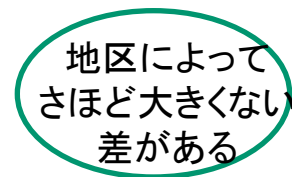
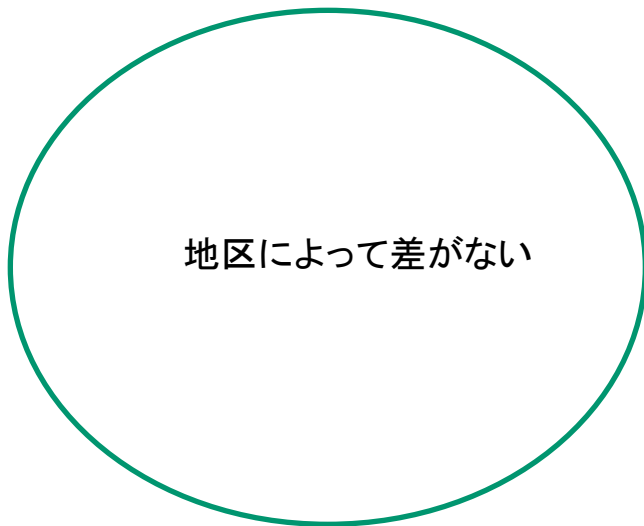
「東地区の方が西地区より評価が
高いような気がする」

(単位: %)

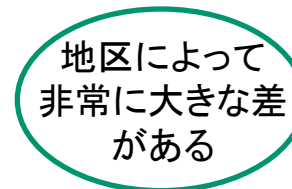
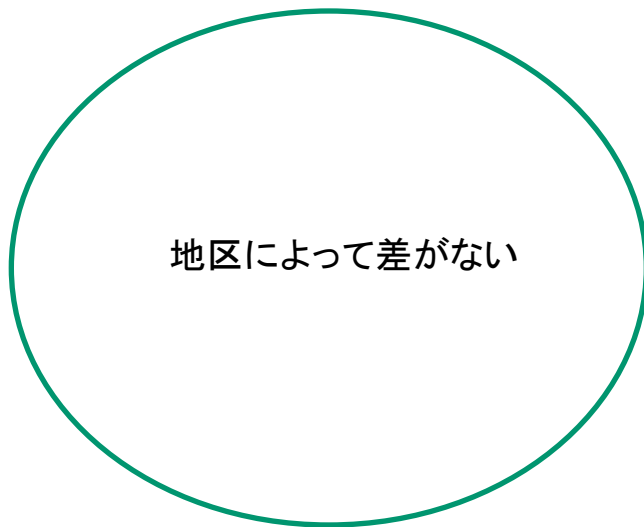
区分	東地区	西地区	合計
評価する	52.5	46.3	49.5
評価しない	47.5	53.7	50.5
合計	100.0	100.0	100.0

「でもそんなに大きな差じゃないよ。
両方ともだいたい半々、同じでしょ。」

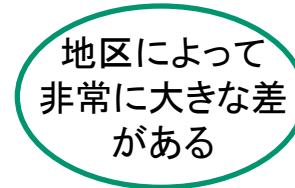
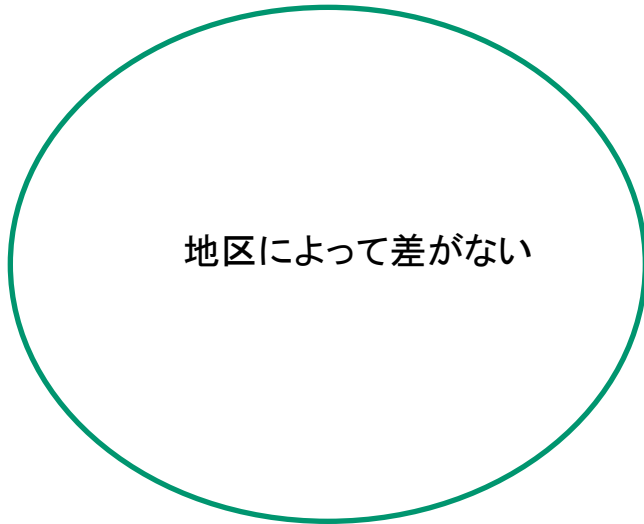




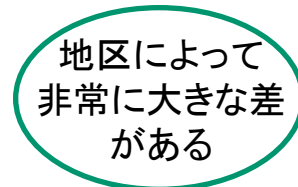
こういうことは普通にありうる



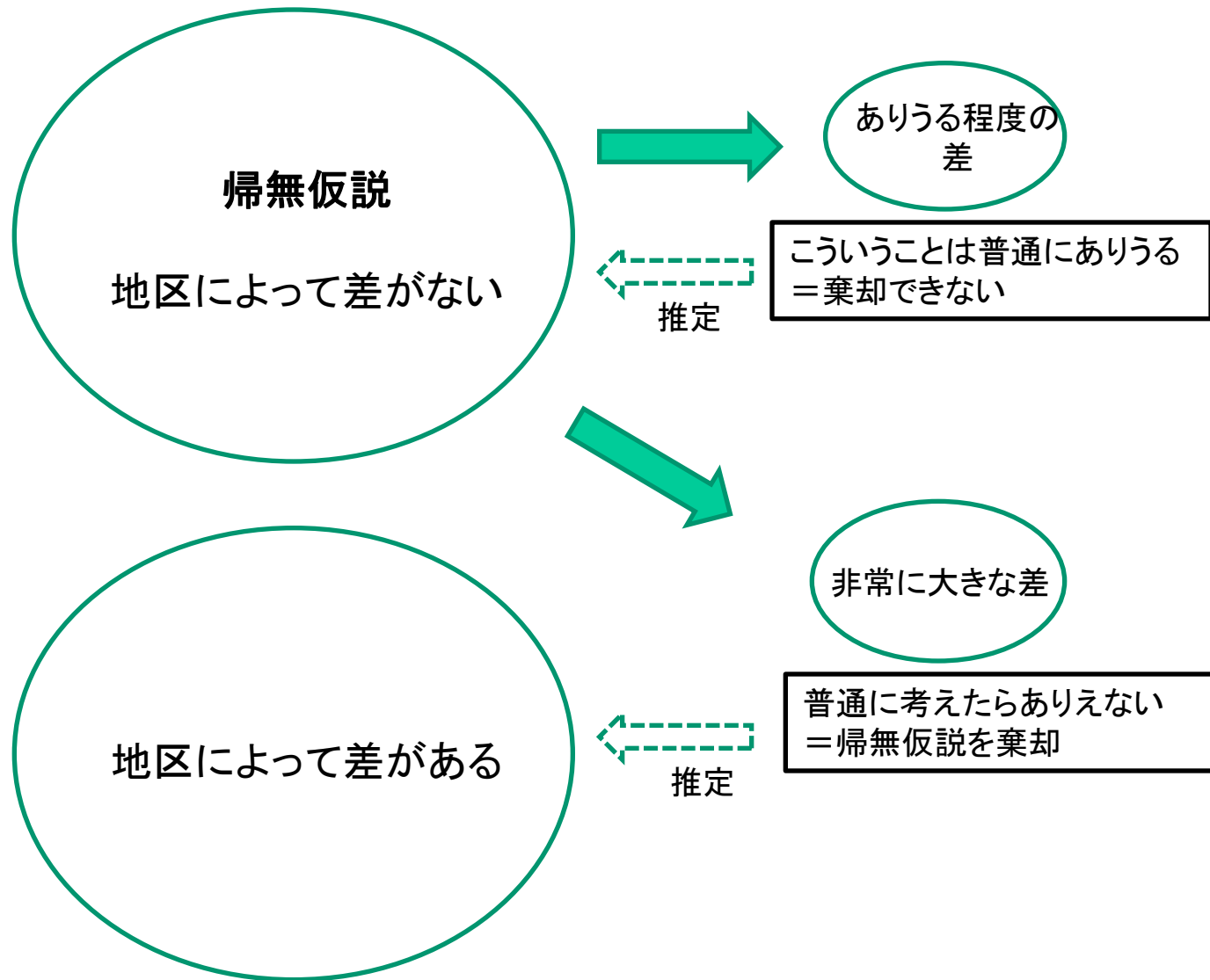
あまりありえない



あまりありえない



こう考えた方が自然



カイ二乗値 (chi-square statistic) の計算

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

(f_o = 実測値、 f_e = 期待値)

(単位: %)

区分	東地区	西地区	合計
評価する	52.5	46.3	49.5
評価しない	47.5	53.7	50.5
合計	100.0	100.0	100.0

市全体での評価は、「評価する＝49.5%」、「評価しない＝50.5%」。

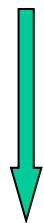
もし住民の居住地区によって評価に差が無いとしたら、それぞれの評価の比率は市全体の比率と同じになる。



(単位: %)

区分	東地区	西地区	合計
評価する	49.5	49.5	49.5
評価しない	50.5	50.5	50.5
合計	100.0	100.0	100.0

区分	東地区	西地区	合計
評価する	49.5	49.5	49.5
評価しない	50.5	50.5	50.5
合計	100.0	100.0	100.0



もし居住地区によって評価に差が無いとしたら、
調査結果はこのようになると期待される

区分	東地区	西地区	合計
評価する	120.7	114.3	235
評価しない	123.3	116.7	240
合計	244	231	475

=期待値

(単位:人)

区分	東地区	西地区	合計
評価する	120.7	114.3	235
評価しない	123.3	116.7	240
合計	244	231	475

=期待値

(単位:人)

区分	東地区	西地区	合計
評価する	128	107	235
評価しない	116	124	240
合計	244	231	475

=実測値
(観察値)

期待値と実測値の差(乖離値)を指数化したもの=カイ二乗値

(単位:人)

区分	東地区	西地区	合計
評価する	2,475	2,475	4,950
評価しない	2,525	2,525	5,050
合計	5,000	5,000	10,000

地区によって
評価に差が
無いとしたら、

区分	東地区	西地区	合計
評価する	120.7	114.3	235
評価しない	123.3	116.7	240
合計	244	231	475

=期待値

差が小さい

区分	東地区	西地区	合計
評価する	123	112	235
評価しない	121	119	240
合計	244	231	475

=実測値

乖離値→カイ二乗値が小さい＝地区によって評価に差が無いとしても、
これくらいの差が現れることは普通にある

⇒地区によって評価に差が無い確率が高い

(単位:人)

区分	東地区	西地区	合計
評価する	2,475	2,475	4,950
評価しない	2,525	2,525	5,050
合計	5,000	5,000	10,000

地区によって
評価に差が
無いとしたら、

区分	東地区	西地区	合計
評価する	120.7	114.3	235
評価しない	123.3	116.7	240
合計	244	231	475

=期待値

差が大きい

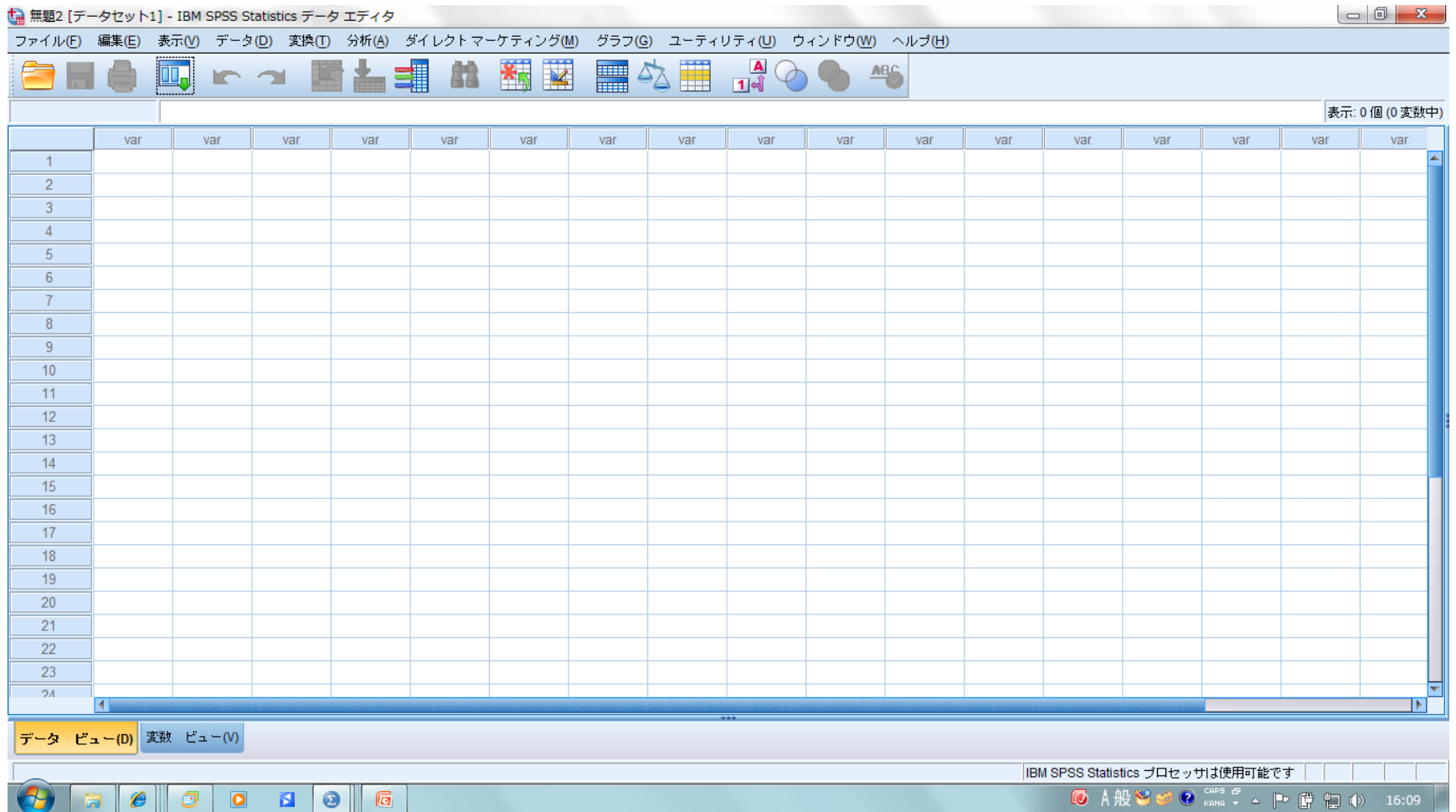
区分	東地区	西地区	合計
評価する	38	197	235
評価しない	206	34	240
合計	244	231	475

=実測値

乖離値→カイ二乗値が大きい＝地区によって評価に差が無いとしたら、
こんな大きな差が現れることは普通に考えたら、まずありえない

⇒地区によって評価に差が無い確率は低い

SPSSによるカイニ乗検定



IBM SPSS Statistics データ エディタ

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) ダイレクトマーケティング(M) グラフ(G) ユーティリティ(U) ウィンドウ(W) ヘルプ(H)

	名前	型	幅	小数桁数	ラベル	値	欠損値	列	配置	尺度	役割
1	地区	文字列	8	0		なし	なし	8	≡ 左	名義(N)	入力
2	評価	文字列	8	0		なし	なし	8	≡ 左	名義(N)	入力
3	人数	数値	8	0		なし	なし	8	≡ 右	スケール(S)	入力
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											

変数の型

☐ 数値(N)
☐ カンマ(C) 文字(H): 8
☐ ドット(D)
☐ 科学的表記法(S)
☐ 日付(A)
☐ ドル記号 (\$) (L)
☐ 通貨フォーマット (U)
☒ 文字列(R)

OK キャンセル ヘルプ

データ ビュー(D) 変数 ビュー(V)

IBM SPSS Statistics プロセッサは使用可能です 重み付き オン

16:05

*無題1 [データセット0] - IBM SPSS Statistics データ エディタ

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) ダイレクトマーケティング(M) グラフ(G) ユーティリティ(U) ウィンドウ(W) ヘルプ(H)

表示: 3 個 (3 変数中)

	地区	評価	人数	var	var	var	var	var	var	var	var	var	var	var	var	var
1	東	する	128													
2	東	しない	116													
3	西	する	107													
4	西	しない	124													
5																
6																
7																
8																
9																
10																
11																
12																
13																
14																
15																
16																
17																
18																
19																
20																
21																
22																
23																
24																

5: 人数

変数プロパティの定義(V)...
不明の測定レベルを設定...
データ プロパティのコピー(C)...
新しいユーザー指定の属性(B)...
日付の定義(E)...
多重回答グループの定義(M)...
重複ケースの特定(U)...
ケースの並べ替え(O)...
変数の並べ替え(B)...
行と列の入れ換え(N)...
ファイルの結合(G)...
再構成(R)...
グループ集計(A)...
データセットをコピー(D)...
ファイルの分割(F)...
ケースの選択(S)...
ケースの重み付け(W)...

データ ビュー(D)

データ ビュー(D) 変数 ビュー(V)

ケースの重み付け(W)...

現在の状態: 重み付け変数は人数

重み付き オン

16:08

ケースの重み付け

☐ ケースの重み付けなし(N)
☒ ケースの重み付け(W)
 度数変数(F): 人数

現在の状態: 重み付け変数は人数

OK 貼り付け(P) 戻す(R) キャンセル ヘルプ

IBM SPSS Statistics データ エディタ

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) ダイレクトマーケティング(M) グラフ(G) ユーティリティ(U) ウィンドウ(W) ヘルプ(H)

5:人数

	地区	評価	人数
1	東	する	1
2	東	しない	1
3	西	する	1
4	西	しない	1
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			

分析(A) > 記述統計(E) > クロス集計表(C)...

クロス集計表: 統計量の指定

☒ カイ 2 乗(H) ☐ 相関係数(R)

名義データ

- ☐ 分割係数(O)
- ☐ Phi および Cramer V(P)
- ☐ ラムダ(L)
- ☐ 不確定性係数(U)

順位データ

- ☐ ガンマ(G)
- ☐ Somers の d(S)
- ☐ Kendall のタウ b
- ☐ Kendall のタウ c

間隔尺度の名義

- ☐ イータ(E)
- ☐ カッパ(K)
- ☐ 相対リスク(I)
- ☐ McNemar(M)

☐ Cochran と Mantel-Haenszel の統計量(A)

共通オッズ比の検定値(T): 1

続行 キャンセル ヘルプ

クロス集計表

人数

行(O): 評価

列(C): 地区

層 1 / 1

前(V) 次(N)

☐ テーブル層に層変数を表示する

ラスタ棒グラフの表示(B)

クロス集計表の非表示(T)

OK 貼り付け(P) 戻す(R) キャンセル ヘルプ

*出力1 [ドキュメント1] - IBM SPSS Statistics ビューア

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 挿入(I) 書式(O) 分析(A) ダイレクトマーケティング(M) グラフ(G) ユーティリティ(U) ウィンドウ(W) ヘルプ(H)

出力
 ログ
 クロス集計表
 表題
 記録
 アクティブデータ
 処理したケースの
 評価と地区のクロス表
 カイ2乗検定

クロス集計表
 [データセット0]

処理したケースの要約

	ケース					
	有効数		欠損		合計	
	N	パーセント	N	パーセント	N	パーセント
評価 * 地区	475	100.0%	0	.0%	475	100.0%

評価と地区のクロス表

度数

		地区		合計
		西	東	
評価	しない	124	116	240
	する	107	128	235
合計		231	244	475

カイ2乗検定

	値	自由度	漸近有意確率 (両側)	正確有意確率 (両側)	正確有意確率 (片側)
Pearson のカイ2乗	1.789 ^a	1	.181		
連続修正 ^b	1.552	1	.213		
尤度比	1.790	1	.181		
Fisher の直接法				.199	.106
有効なケースの数	475				

a. 0 セル (.0%) は期待度数が 5 未満です。最小期待度数は 114.28 です。
 b. 2x2 表に対してのみ計算

IBM SPSS Statistics プロセッサは使用可能です

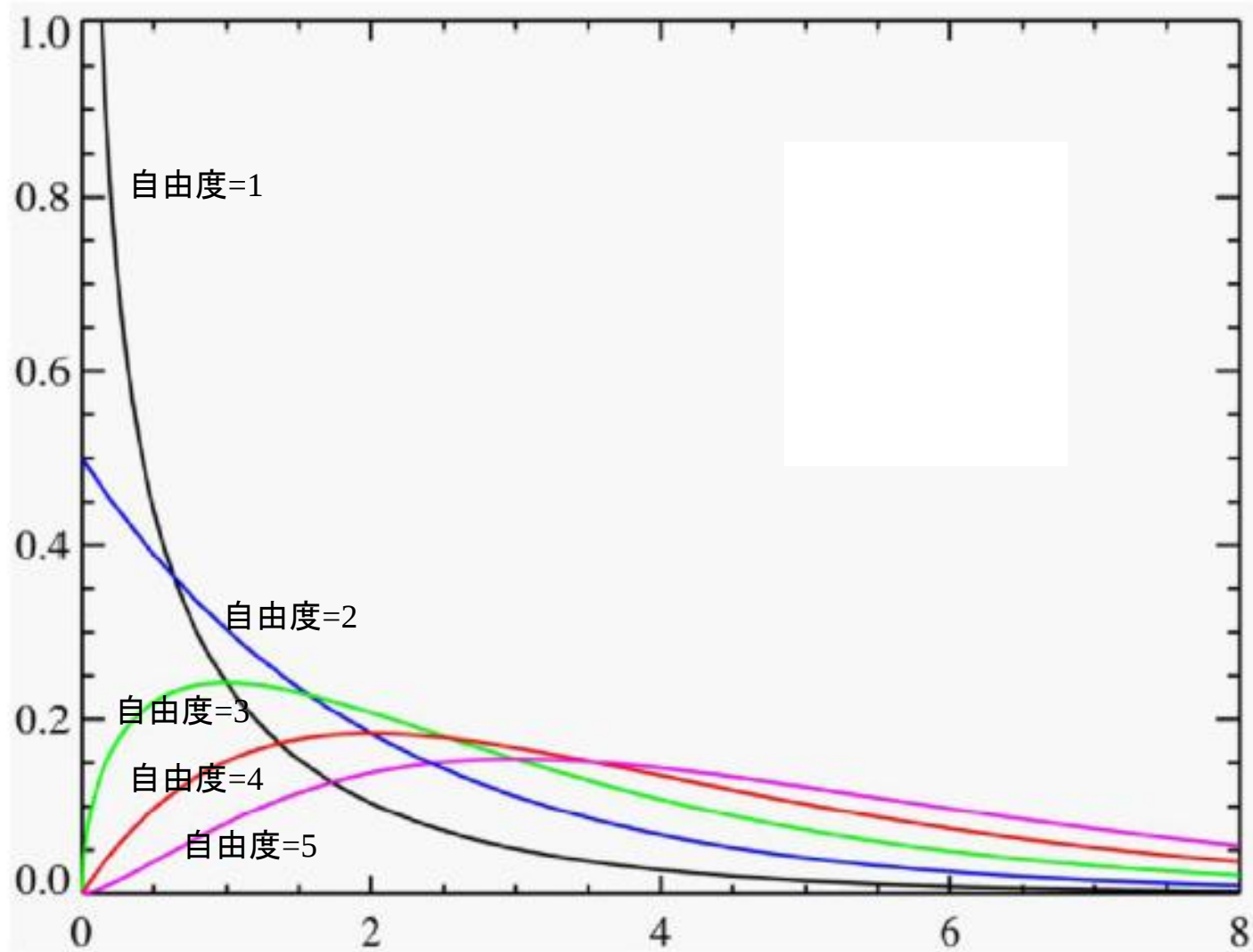
A 般 CAPS KANA 16:00

Table I Distribution of χ^2

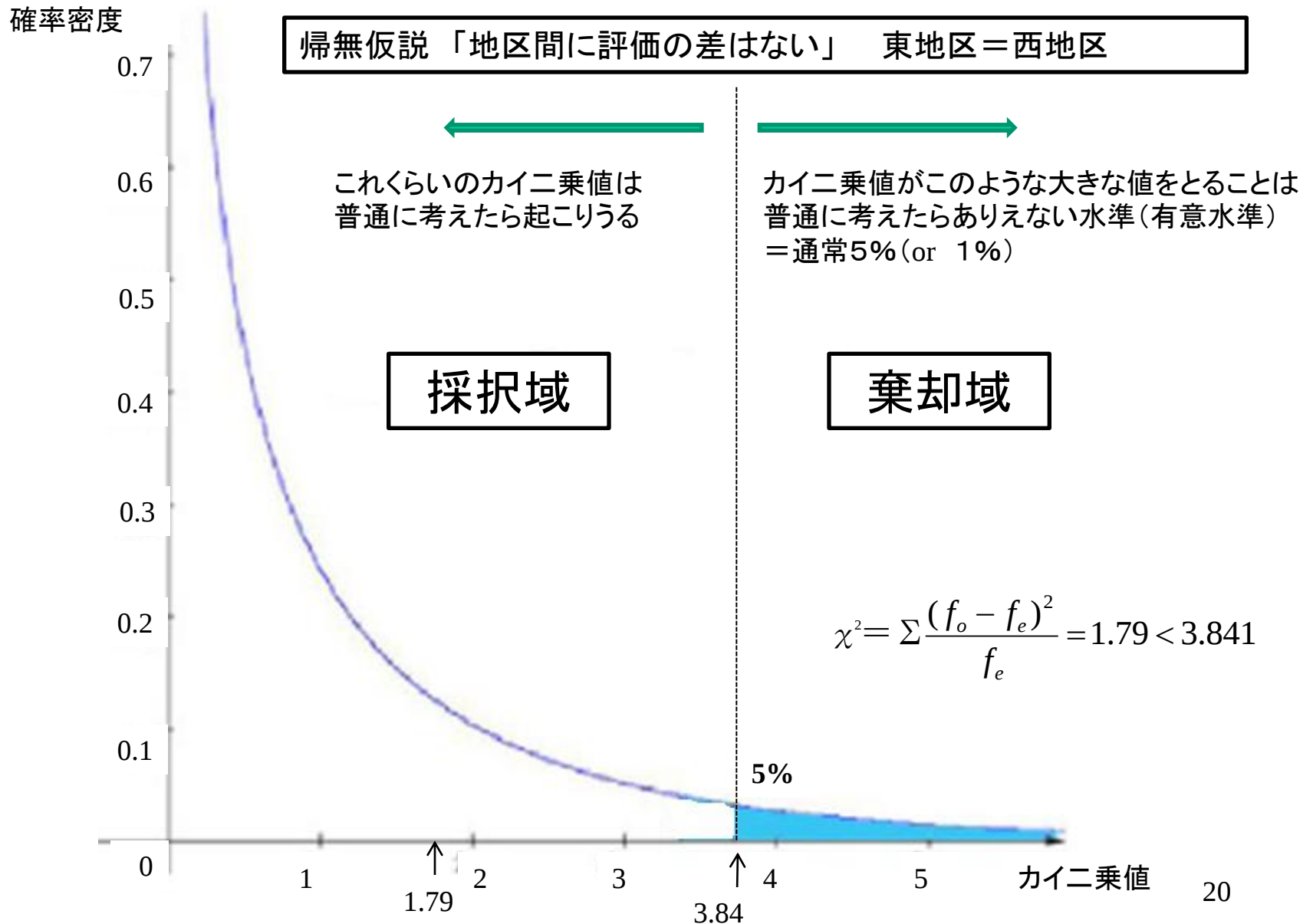
Probability

df	.99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	.0157	.0268	.0393	.0540	.0832	.1038	.1455	.1847	.2147	.2706	.3341	.4114	.4835	10.827
2	.0201	.0404	.0540	.0717	.1005	.1244	.1847	.2386	.2719	.3386	.4013	.4835	.5646	13.815
3	.115	.185	.235	.296	.390	.455	.675	.838	.934	1.213	1.437	1.754	2.079	16.268
4	.297	.429	.538	.675	.872	.1.064	1.649	2.195	2.357	3.006	3.357	4.078	4.608	18.465
5	.554	.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086	20.517
6	.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22.457
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.125
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980	51.179
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314	52.620
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642	54.052
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963	55.476
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278	56.893
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588	58.302
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59.703

カイ二乗分布(確率密度)



自由度1のカイ二乗分布(確率密度)



数量分析ホームワーク

- (1) ある市では道路建設に際して、沿線住民の意見を調査するため、住民の意識調査を行った。A 地区は沿線からやや離れた地区、B 地区は沿線沿いに所在する地区である。この調査結果（表 1）から、住民の居住地区によって道路建設に対する意見に差があるといえるかどうか、カイ二乗検定を行い分析してください。

【表 1】

(単位：人)

区 分	A 地区	B 地区	市全体
賛成	1 1 8	8 2	2 0 0
反対	1 0 2	1 1 5	2 1 7
合 計	2 2 0	1 9 7	4 1 7

- (2) 下の表（表 2）は、路上禁煙条例に対する住民の意見を職業別に分類したものです。この結果から、回答者の職業によって回答内容に違いがあるといえるかどうか、カイ二乗検定を行い分析してください。

【表 2】

(単位：人)

区 分	会社員	自営業	その他	合 計
賛成	1 0 1	8 2	1 8	2 0 1
反対	6 9	9 3	1 1	1 7 3
どちらともいえない	7 9	3 6	1 1	1 2 6
合 計	2 4 9	2 1 1	4 0	5 0 0