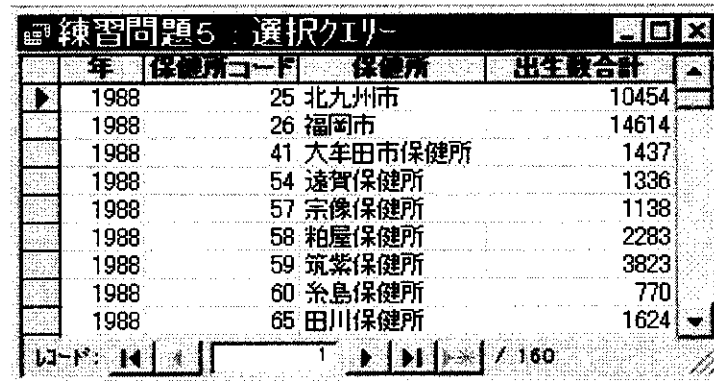


練習問題5

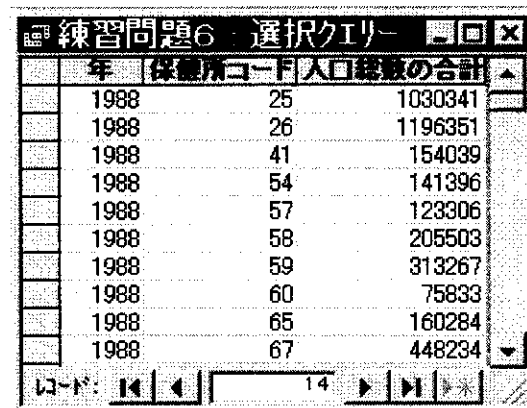
「集計1」クエリー (5.1 で作成) を基に、年別・保健所別の出生数合計を集計する『練習問題5』クエリーを作成して下さい。クエリーを新規作成、「集計1」クエリーを追加した後、作成して下さい。並び替えは、「年」「保健所コード」の順です。



年	保健所コード	保健所	出生数合計
1988	25	北九州市	10454
1988	26	福岡市	14614
1988	41	大牟田市保健所	1437
1988	54	遠賀保健所	1336
1988	57	宗像保健所	1138
1988	58	粕屋保健所	2283
1988	59	筑紫保健所	3823
1988	60	糸島保健所	770
1988	65	田川保健所	1624

練習問題6


【人口総数】テーブルを基に、年別・保健所コード別に人口総数を集計する『練習問題6』クエリーを作成して下さい。並び替えは、「年」「保健所コード」の順です。



年	保健所コード	人口総数の合計
1988	25	1030341
1988	26	1196351
1988	41	154039
1988	54	141396
1988	57	123306
1988	58	205503
1988	59	313267
1988	60	75833
1988	65	160284
1988	67	448234

練習問題7

「練習問題5」クエリーと「練習問題6」クエリーから、年別・保健所別の出生率（千人対）を計算して下さい。クエリーを新規作成、「練習問題5」クエリーと「練習問題6」クエリーを追加した後、作成して下さい。並び替えは、「年」「保健所コード」の順です。



年	保健所コード	保健所	出生率
1988	25	北九州市	101
1988	26	福岡市	122
1988	41	大牟田市保健所	9.3
1988	54	遠賀保健所	9.4
1988	57	宗像保健所	9.2
1988	58	粕屋保健所	11.1
1988	59	筑紫保健所	12.2
1988	60	糸島保健所	10.2
1988	65	田川保健所	10.1
1988	67	久留米保健所	11.3

保健情報処理研修会

人口動態統計解析

平成11年度

福岡県保健環境研究所

片岡 恭一郎

目 次

1	人口動態統計解析とは	115
2	人口動態統計の歴史	115
3	人口動態統計システムについて	116
4	人口動態統計に用いられる指標	118
5	死亡率の推定と検定	120
6	地域診断データベースの操作	123
7	地域診断データベースを用いた人口動態解析演習	126
付録	平成9年 人口動態統計上巻から抜粋	141

1 人口動態統計解析とは

人口動態統計解析とは人口動態事象(出生・死亡・死産・婚姻・離婚)の統計値 について分からない事柄を筋道たてて明らかにすることである。

例えば、

- ① 生まれてくる赤ちゃんはこの頃多いのかな、少ないのかな?→時間的経過
- ② 私の住んでいる地域ではどうなんだろう?→地理的分布
- ③ 私の住んでいる地域では肝臓がんで死亡する人が多いと聞いたけれど本当かしら?→母比率の検定
- ④ 婚姻件数が多い地域は生まれ来る赤ちゃんも多いのかな?→相関関係

などの分からない事柄について、解明することである。

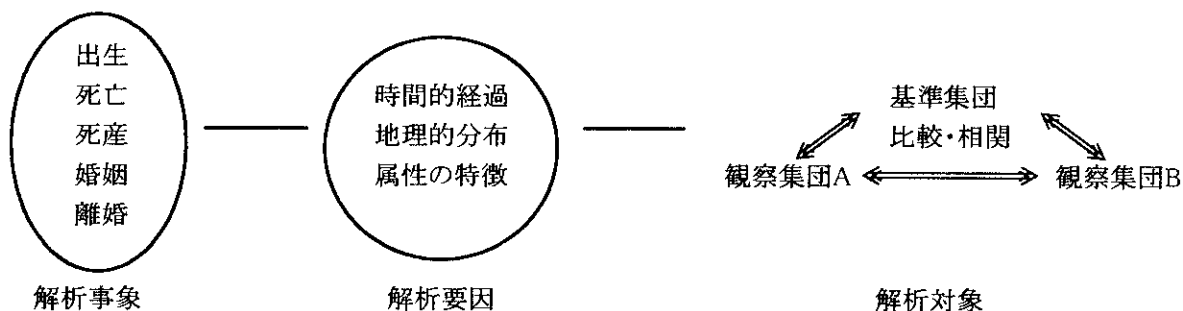


図1 人口動態統計解析の事象、要因及び対象

また、図1に示すように「観察集団の人口動態事象について、時間的経過、地理的分布、属性の特徴が、基準集団あるいは観察集団間と比較してどうであるかを評価すること」とも言える。

2 人口動態統計の歴史

(1) 「数えよ」^{1),2)}

John Graunt (英、1620-1674)は、1662年に「死亡表に関する自然的ならびに政治的諸観察」を著し、そのなかで、ロンドンにおける死亡票及び洗礼の記録を10年ごとに分析し、出生数、死亡数がともに女より男に多いこと、乳児の死亡率が高いこと、死亡率に季節変動があることなどを明らかにした。また、ロンドン市内のペストの流行を数量的に扱い、流行の年度について気象その他生態的特徴を調べている。さらに、生物統計学の2つの手法である人口の推定と生命表の作成も初めて行った。(参照ファイル:グラント扉.bmp、グラント表.bmp)

グラントは「生物現象を大数として取り扱ふと、恒常性を持ち、予測しうることを示した。疫学の先覚者といわれている。

(2) 日本の人口動態統計制度²⁾

645年 大化改新において、籍に関する制度が定められた。

1871年 明治4年 戸籍の法が制定される。

1872年 明治5年 戸籍簿の作成。男女別の出生数と死亡数の調査始まる。

1898年 明治31年 戸籍法の制定。

1899年 明治32年 人口動態調査方式の確立。

① 地方分査方式→中央集査方式。

② 表方式→出生・死亡・婚姻・離婚の各1件ごとの調査小票方式

1902年 明治35年 第1回生命表の作成。明治24年～31年の死亡統計に基づくもの。

1946年 昭和21年 人口動態調査令及び死産の届出に関する規定の整備。

1947年 昭和22年 人口動態調査が総理庁統計局から厚生省(昭和13年創設)へ移管。

参考文献

1) 新しい疫学 重松逸造、柳川洋監修 (財)日本公衆衛生協会 1991

2) 厚生統計テキストブック第2版 (財)厚生統計協会 1990

3 人口動態統計システムについて

(1) システムの一般モデル¹⁾

システムとは、複数の要素を持ち、その要素間に相互関連性があり、共通の目的／機能を持つ集合体を言う。

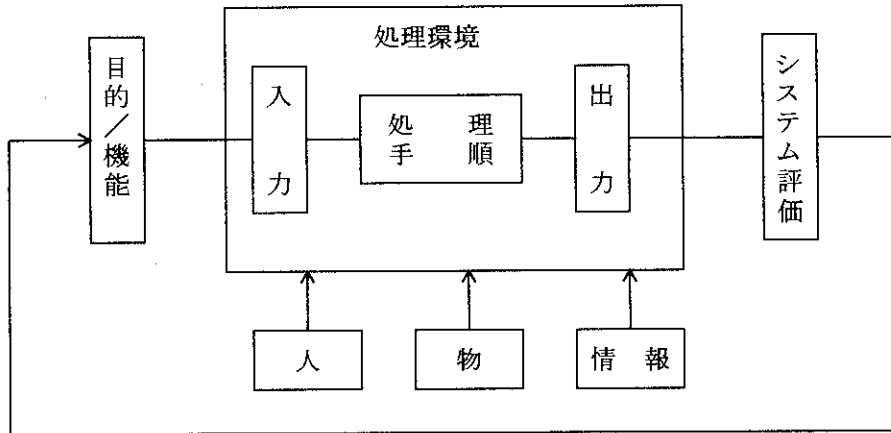


図2 システムの一般モデル

(2) 人口動態統計システム

ア 目的／機能

(ア) 人口推計のためのデータを提供する。

人口推計は、国勢調査人口を基にして、人口動態統計の出生と死亡、国内外への入出国、さらに都道府県間の移動を加味して作られる。

(イ) 各種行政のためのデータを提供する。

母子保健…合計特殊出生率、低体重児割合、乳児死亡率、周産期死亡率、死産率、婚姻率
公衆衛生…結核等の感染症による死亡、がん・脳血管疾患・心疾患等の生活習慣病による死亡率
福祉行政…離婚率
保健福祉の総合的指標…平均余命

(ウ) 保健所管内の保健行政推進のためのデータを提供する。

人口動態調査令施行細則第7条[保健所における利用]…人口動態調査票等の利用と保存。

(エ) 指定統計調査票の目的外使用としてデータを提供する。

統計法第15条…目的外使用の禁止と総務庁長官による目的外使用の承認。がん登録患者の死亡確認。特定死因と危険因子との関連分析等々。

イ 入力

入力票の詳細については付録1頁を参照のこと。届出の特記事項については下記のとおりである。

(ア) 出生票

航海中—船長が航海日誌に記載し、その謄本を送付。

病院、監獄等—長又は管理人が届け出る。

棄児発見—発見者又は警察官(24時間以内)、市区村長が調書作成する。父又は母が棄児を引き取ったときは1カ月以内に出生届。

(イ) 死亡票

航海中—船長が航海日誌に記載し、その謄本を送付する。

水難、火災、事変による死亡—取り調べをした官公署が死亡地の市区町村長へ報告。

死刑—監獄の長、所在地の市区町村長。

失踪—判決確定後10日以内に裁判請求者が届け出る。

(ウ) 死産票

航海中—船長が航海日誌に記載し、その謄本を送付する。

ウ 処理手順

(ア) 届出・登録システム

付録2頁参照のこと。

(イ) 年報処理システム

人口動態統計処理のシステムフローチャートを図3に示す。

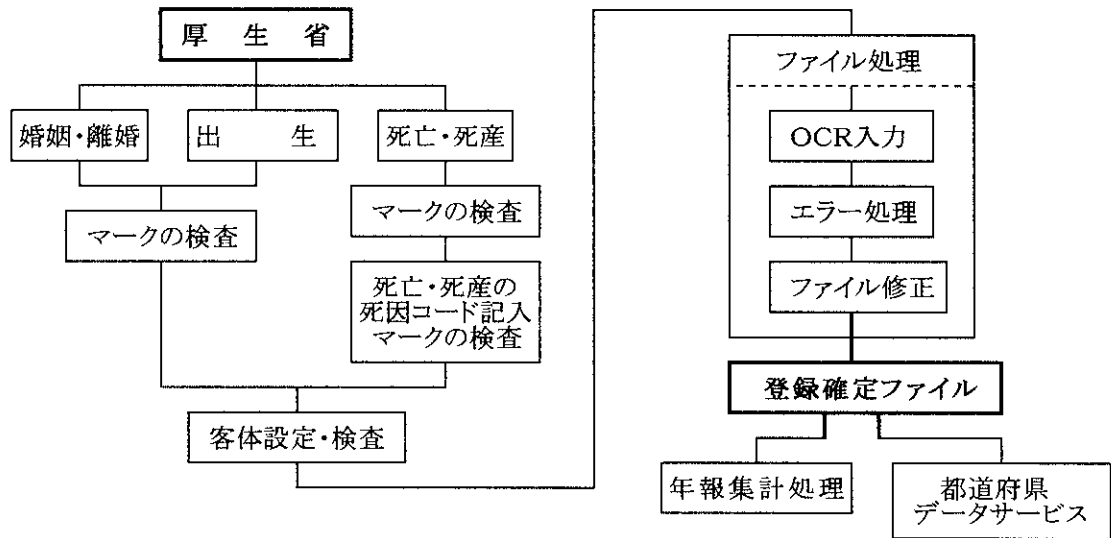


図3 人口動態統計年報処理のシステムフローチャート

エ 出力

(ア) 人口動態統計 上・中・下巻

(イ) WHO送付用表・テープ

(ウ) 登録確定ファイル

出生 昭和49年(1974年)～ 死亡 昭和44年(1969年)～ 死産 昭和44年(1969年)～
 婚姻 昭和48年(1973年)～ 離婚 昭和48年(1973年)～ 周産期死亡 昭和48年(1973年)～

調査年	調査番号	客体設定	届出年月	届出地	事件種別	住所	性別	出生年月日時分	死亡年月
年	月	年	月	都道府県	市区町村	NO.	性別	年	月
97	02	16	09	01	01	101~	1	01	01
1	2	3	4	5	6	7	8	9	0

時分	国内	国外	配偶関係	死亡	原因	外国符号	路上交通事故	手術の有無	1歳未満の病死	母の
時	分	年	月	年	年	年	年	年	年	年
10	00	00	01	01	00	00	00	00	00	01
1	2	3	4	5	6	7	8	9	0	1

生年月日	母の	前回の妊娠	事件本人の年齢	日	時	分	死	乳児	感	死	WHO
年	年	年	年	日	時	分	因	死	染	年	一般
01	01	01	00	00	00	00	00	00	00	00	00
1	2	3	4	5	6	7	8	9	0	1	2

WHO	子の数	妊
乳児小児	出生	産
0	0	0
1	2	3

図4 都道府県データサービス用の磁気テープレイアウトフォームの例(死亡票)²⁾

参考文献

- 1) 情報学の常識77 日比野省三 福村出版 1984
- 2) 平成9年人口動態調査テープレイアウト及びコード一覧 厚生省大臣官房統計情報部編 1998

4 人口動態統計に用いられる指標

人口動態事象を評価するためには評価尺度(指標)が必要である。指標は調べようとする目的、価値感によって異なるものである(主な指標は、「人口動態統計 上巻」に掲載されている。付録4頁を参照のこと)。

例えば、目的が「母子手帳を配布」することであれば、その地域でどれだけ赤ちゃんが産まれるかを知らないといけないので、その指標は「出生数」であろうし、目的が「A地域とB地域の出生を比べる」ことであれば、出生数は地域の人口に比例すると考えられるからその指標は「出生率」が適しているだろう。このように解析する目的に合わせた指標を選択することが重要である。

人口動態統計で用いられる多くの指標は「率」である。そのほか、「比」、「割合」、「百分率」などがある。

(1) 比、割合、百分率、率の区別

比、割合、百分率、率の指標は群別に計数(count)されるデータに用いられる。離散的(discrete)あるいは定性的(qualitative)データともいわれる。

ア 比(ratio)

A事象とB事象があったとき、 A/B あるいは B/A のことをAとBの比という。

例…出生性比、死亡性比、乳児死亡性比、死産性比など。

イ 割合(propotion)

A事象とB事象があったとき、 $A/(A+B)$ あるいは $B/(A+B)$ のことを総数に対するAの割合あるいはBの割合という。必ず1より小さい値になる。

例…低体重児割合

ウ 百分率(percentage,%)

割合を100倍した値をいう。

エ 率(rate)

(ア) 割合を1,000倍、10,000倍あるいは100,000倍した値をいう。

例…乳児死亡率、新生児死亡率、死産率、周産期死亡率など。

(イ) (ある期間の計数) / (ある時点の統計量)であらわしたものをそれを1,000倍(あるいは10,000倍、100,000倍)した値をいう。

例…出生率、死亡率、自然増加率、婚姻率、離婚率など。

計数データは比、割合、百分率、率などに変換して解析されることが多いが、必ず実際の数字も挙げておかなければならない。次のような考え方もある。¹⁾

① 百分率の分母が2500未満の値ならば%の値は小数点以下を丸めて整数値で表すこと。

$D[P]=\sqrt{P(1-P)/n}$ からPが10~90%のとき標準誤差が0.5%以上である。したがって、推定値の計算で小数点以下の数字を出しても、標準誤差を考えると無意味である。(参照ファイル:標準誤差.xls)

② 分母が100未満の百分率は出さなくて、調査実数を記すのがよい。百分率を出す場合は必ず実数も付記すること。

③ 分母が25未満のときは決して百分率を出さないこと。標準誤差が6%以上となり、標準誤差を考えると百分率の10の位の数字すら不確かともいえる。

【練習1】 比率区分.xlsを開き問題を解いて下さい。

(2) 主な指標の解説

ア 合計特殊出生率について

15歳から49歳までの女子の年齢別出生率を合計したもので、1人の女子が仮にその年次の年齢別出生率で一生涯の間に生むとした時の子ども数に相当する。

$$\text{計算式は} \frac{15\text{歳の母からの出生数}}{15\text{歳の女子人口}} + \frac{16\text{歳の母からの出生数}}{16\text{歳の女子人口}} + \dots + \frac{49\text{歳の母からの出生数}}{49\text{歳の女子人口}}$$

であるが、県あるいは市区町村などで各歳ごとの出生数や人口が入手困難な場合は5歳階級の数値を用い、次のように計算する。

$$\frac{15\sim 19\text{歳の母からの出生数}}{15\sim 19\text{歳の女子人口}} \times 5 + \dots + \frac{45\sim 49\text{歳の母からの出生数}}{45\sim 49\text{歳の女子人口}} \times 5$$

* 特殊出生率…粗出生率の分母人口が10月1日現在の日本人人口を用いるのに対して、分母に女子人口あるいは妊娠可能年齢人口等の特殊な人口を用いるという意味で特殊出生率という。

イ 粗死亡率、年齢調整死亡率及びSMRについて

(ア) 粗死亡率(crude mortality rate)

通常、死亡率は $\frac{\text{ある地域の1年間の日本における日本人の死亡数}}{\text{死亡が起こった地域の10月1日現在日本人人口}} \times 1,000$

で表される。

例えば、福岡県の平成9年10月1日現在の日本人人口は、4,944,000人であり、平成9年1月1日から12月31日までの死亡数が36,884人だったとき、粗死亡率は、

$$\frac{36,884}{4,944,000} \times 1,000 = 7.46$$

人口1,000人当たり7.46となる。

粗死亡率は、地域の年齢構成に大きく影響される。すなわち、高齢者の多い地域ほど粗死亡率は高くなる。

例えば、図1のような異なる年齢構成を持つA、Bの地域があるとき、それぞれの年齢階級では同じ死亡率であっても、高齢者の多いB地域の方が粗死亡率が高くなる。

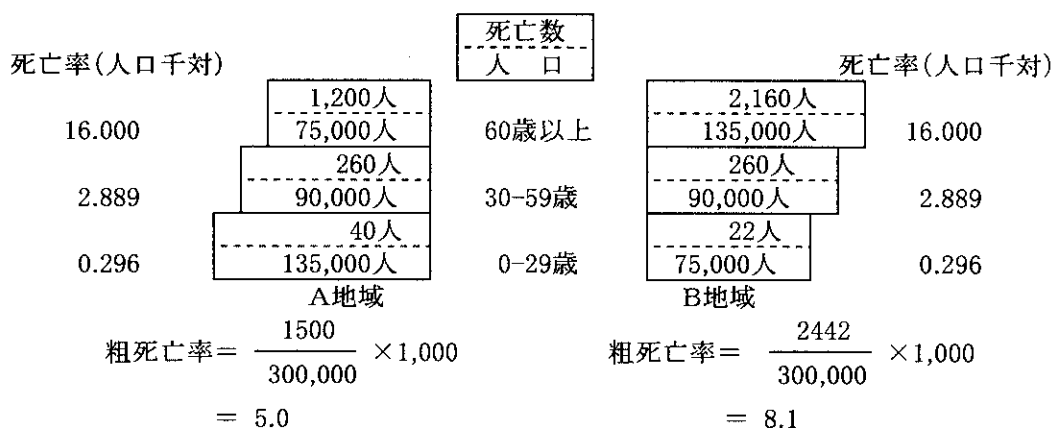


図5 年齢構成が異なる地域の粗死亡率

(イ) 年齢調整死亡率(age-adjusted mortality rate)

年齢構成を考慮した死亡率が年齢調整死亡率である。年齢調整死亡率には、観察集団の年齢構成が基準人口集団と同一であると仮定して計算する直接法と、基準人口集団での年齢別の死亡率を用いて計算する間接法とがある。

a 直接法

計算の定義式は付録5頁の年齢調整死亡率の項を参照のこと。

基準人口は死亡率の計算では通常「昭和60年モデル人口」を用いているが、比較をする目的によっては別の基準人口を設定する。例えば、特定の要因に曝露された群と曝露されなかった群を比較する場合は、曝露群における発生を見たいのであるから、曝露群における人口構成を「基準人口」として用いることがある。従って、年齢調整死亡率の絶対値は用いた基準人口によって変わることは銘記すべきである。

年齢調整死亡率(直接法)の考え方を図5で示したA地域を例に取り示す。

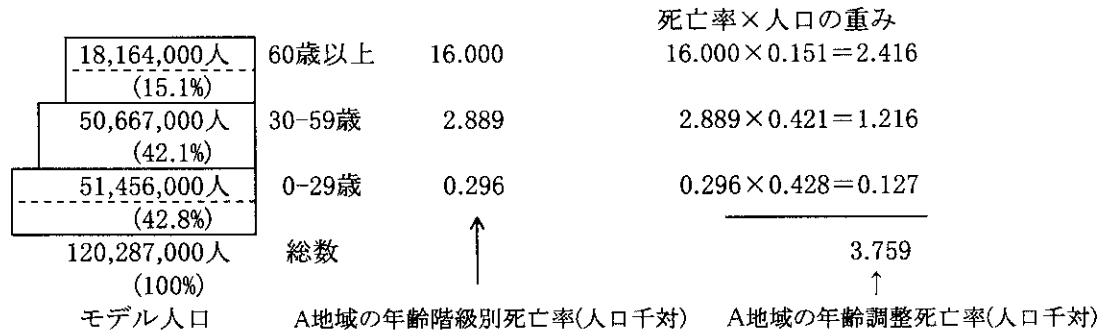


図6 年齢調整死亡率(直接法)の考え方

b 間接法

定義式は、(基準集団の粗死亡率)×SMR で表される。

多くはSMR(standardized mortality ratio:標準化死亡比)のみで解析に用いる。

c SMR

定義式は $\frac{\text{観察集団の実際の死亡数}}{\text{期待死亡数}}$ である。通常100倍したものをを用いる。

期待死亡数={基準集団の各年齢(年齢階級)死亡率×観察集団の各年齢(年齢階級)人口}の各年齢(年齢階級)の総和

SMRは基準とした集団(多くは全国値)の年齢階級別死亡率を観察集団の年齢階級別人口に乗じて総和したものを期待死亡数として、それに対する観察集団の実際の死亡数の比で表したもの。

【例題】 図5で示したA地域を基準集団として、B地域のSMRを計算する。

$$\begin{aligned} \text{SMR} &= \frac{2,442}{0.016 \times 135,000 + 0.002889 \times 90,000 + 0.000296 \times 75,000} \times 100 \\ &= \frac{2,442}{2,442} \times 100 = 100 \end{aligned}$$

参考文献

- 1) 医統計テキスト 遠藤和男・山本正治 西村書店 1992

5 死亡率の推定と検定

(1) 年齢調整死亡率の場合

ア 推定

観察集団の年齢階級別死亡数がポアソン分布に従うと仮定すると、年齢調整死亡率の分散は次のように推定される¹⁾。

$$\text{分散} = \sum \frac{\left[\frac{\text{基準集団の各年齢階級人口}}{\text{基準集団の人口総数}} \right]^2 \times (\text{観察集団の各年齢階級死亡数})}{(\text{観察集団の各年齢階級人口})^2}$$

したがって年齢調整死亡率の95%信頼区間は近似的に次のようになる。

$$\text{年齢調整死亡率} \pm 1.96 \sqrt{\text{分散}}$$

【例題】 図5で示したA地域の年齢調整死亡率の95%信頼区間を求める。

$$\text{分散} = \frac{0.151^2 \times 1,200}{75,000^2} + \frac{0.421^2 \times 260}{90,000^2} + \frac{0.428^2 \times 40}{135,000^2} = 0.0000001094$$

$$\text{年齢調整死亡率の95\%信頼区間(人口千対)} = 3.759 \pm 0.205$$

イ 検定

観察集団での真の年齢調整死亡率が基準集団での年齢調整死亡率に等しいかどうかを検定する¹⁾。

帰無仮説: 観察集団の年齢調整死亡率 = 標準集団の年齢調整死亡率

$$\text{検定統計量: } Z = \frac{|\text{観察集団の年齢調整死亡率} - \text{基準集団の年齢調整死亡率}|}{\sqrt{\text{年齢調整死亡率の分散}}}$$

ただし、年齢調整死亡率の分散

$$= \sum \frac{\left(\frac{\text{基準集団の各年齢階級人口}}{\text{基準集団の人口総数}} \right)^2 \times (\text{基準集団の各年齢階級別死亡率})}{(\text{観察集団の各年齢階級人口})}$$

基準集団の各年齢階級別死亡率が不明のときは、観察集団の各年齢階級別死亡率を用いる。そのときは、年齢調整死亡率の分散は、(1)のアで示した分散と同じである。

判定: Zが近似的に正規分布に従うことを利用し、検定する。

Z > 1.960のとき危険率5%で有意差があると判定する。

Z > 2.576のとき危険率1%で有意差があると判定する。

【例題】 福岡県の年齢調整死亡率の95%信頼区間の計算と全国の年齢調整死亡率との検定を行う。

	0~29歳	30~59歳	60歳以上	合計
モデル人口の割合(昭和60年)	0.428	0.421	0.151	1
全国の人口(平成7年)	46,739,063	51,919,607	25,640,277	124,298,947
全国の死亡数(平成7年)	22,504	126,596	772,402	921,502
福岡県の人口(平成7年)	1,881,839	2,000,449	1,014,163	4,896,451
福岡県の死亡数(平成7年)	892	5,026	31,239	37,157
全国年齢階級死亡率(人口千対)	0.481	2.438	30.125	
全国年齢階級死亡率×重み	0.206	1.026	4.549	5.781
			↑	全国年齢調整死亡率
福岡県年齢階級死亡率(人口千対)	0.474	2.512	30.803	
福岡県年齢階級死亡率×重み	0.203	1.058	4.651	5.912
			↑	福岡県年齢調整死亡率

年齢調整死亡率の分散(観察集団の死亡率を使用)

$$= 0.428^2 \times 892 / 1,881,839^2 + 0.421^2 \times 5,026 / 2,000,449^2 + 0.151^2 \times 31,239 / 1,014,163^2$$

$$= 0.0000000009613$$

よって、福岡県年齢調整死亡率の95%信頼区間は

$$(5.912 - 1.96 \times \sqrt{0.0000000009613} \times 1,000, 5.912 + 1.96 \times \sqrt{0.0000000009613} \times 1,000)$$

$$= (5.851, 5.973)$$

次に、Zの値は $Z = \frac{|0.005781 - 0.005912|}{\sqrt{0.0000000009613}} = \frac{0.000130}{0.000031} = 4.206$

となり、福岡県の年齢調整死亡率は全国の年齢調整死亡率に比べて危険率1%で有意に高いと判定される。

【練習2】 死亡率検定.xlsを開き【例題】を試して下さい。

(2) SMRの場合

ア 推定

(ア) 正規近似による推定

SMRの分散は次のように推定される¹⁾。

$$\text{分散} = \frac{\text{観察集団の実際死亡数}}{(\text{期待死亡数})^2}$$

したがってSMRの95%信頼区間(有意水準(α)が0.05のときの信頼区間)は近似的に次のようになる。

$$\text{SMR} \pm 1.96 \sqrt{\text{分散}}$$

ただし、1.96は正規分布の上側確率 0.05/2=0.025の標準化偏差の値。

【例題】 (1)のイで示した全国の年齢階級死亡率を基準とし、福岡県のSMRを計算する。

また、その95%信頼区間を求める。

期待死亡数 = 1,881,839 × 0.000481 + 2,000,449 × 0.002438 + 1,014,163 × 0.030125 = 36,335

福岡県のSMR = 37,157 / 36,335 = 1.023

SMRの分散 = 37,157 / 36,334² = 0.0000281

SMRの95%信頼区間 = 1.023 ± 1.96 × √0.0000281 = 1.023 ± 1.96 × 0.005305 = 1.023 ± 0.010

(イ) ポアソン分布による推定²⁾

有意水準(α)が5%のときの信頼区間(100-α)、95%信頼区間を推定する。

実際死亡数(Obs)がポアソン分布に従うとき、その死亡数の95%信頼区間の下限値をλ_L、上限値をλ_Uとおくと、ポアソン分布とχ²分布との関係から次の関係が導かれる。

$$P(\lambda_U) = P_r\{\chi^2_{2 \times Obs} \leq 2\lambda_U\} = \alpha/2$$

$$1 - P(\lambda_L) = P_r\{\chi^2_{2 \times (Obs+1)} \leq 2\lambda_L\} = 1 - \alpha/2$$

このことから、SMR_L(下限値) = λ_L/Exp、SMR_U(上限値) = λ_U/Expは次のように求められる。ただし、Expは期待死亡数。

$$SMR_L = \frac{CHIINV(1-\alpha/2, 2 \times Obs)/2}{Exp}$$

$$SMR_U = \frac{CHIINV(\alpha/2, 2 \times Obs + 2)/2}{Exp}$$

CHIINVはMicrosoft Excelの統計関数である。

CHIINV 関数項目
カイ2乗(χ ²)分布の逆関数を返します。つまり、 確率 = CHIDIST(χ)であるとき、CHIINV(確率) = χ という関係が成り立ちます。χ ² 検定は、実測値と期待値を比較して、仮説の妥当性を検定するために使います。
書式 CHIINV(確率 ; 自由度)
確率 χ ² 分布に伴う確率を指定します。
自由度 自由度を指定します。
解説
<ul style="list-style-type: none">引数に数値以外の値を指定すると、エラー値 #VALUE! が返されます。確率 < 0 または 確率 > 1 である場合、エラー値 #NUM! が返されます。自由度 に整数以外の値を指定すると、小数点以下が切り捨てられます。自由度 < 1 または 自由度 ≥ 1010 である場合、エラー値 #NUM! が返されます。CHIINV 関数の計算には、反復計算の手法が使用されます。確率 の値が指定されると、計算結果の精度が ±3×10⁻⁷ 以内になるまで反復計算が行われます。反復計算を 100 回行っても結果が収束しない場合は、エラー値 #N/A が返されます。
使用例
CHIINV(0.05,10) = 18.30703

<注意> Microsoft Excelでは、自由度が1,367までしか計算できない。つまり、実際死亡数が682人を超えるような場合は計算できない。SASやSPSSのような汎用統計パッケージでは対応可能である。

実際死亡数が多い場合は正規近似による推定でよい。

イ 検定

(ア) 正規分布を仮定した検定

観察集団での死亡数が基準集団から求められた期待死亡数に等しいかどうかを検定する¹⁾。

帰無仮説: 観察集団の実際の死亡数 = 観察集団の期待死亡数

$$\text{検定統計量: } Z = \frac{|\text{観察集団の実際の死亡数} - \text{観察集団の期待死亡数}|}{\sqrt{\text{観察集団の期待死亡数}}}$$

判定: Zが近似的に正規分布に従うことを利用し、検定する。

Z > 1.960のとき危険率5%で有意差があると判定する。

Z > 2.576のとき危険率1%で有意差があると判定する。

【例題】 (2)のアの(ア)で求めた福岡県のSMRの検定を行う。

$$Z = |37,157 - 36,334| / \sqrt{36,334} = 823 / 190.6 = 4.3 > 2.56$$

福岡県の実際の死亡数は危険率1%で期待死亡数より有意に高い

(イ) ポアソン分布を仮定した検定

死亡数が少ない場合には、Zの正規分布の近似が良くないのでポアソン分布から計算する¹⁾。

a 実際の死亡数(D) < 期待死亡数(E)のとき

$$p = \sum_{d=0}^D \frac{E^d}{d!} e^{-E}$$

Microsoft Excelでは $p = \text{POISSON}(D, E, \text{TRUE})$ で求められる。

を計算し、pが0.025より小さいとき5%で有意、0.005より小さいとき1%で有意と判定する。

b 実際の死亡数(D) > 期待死亡数(E)のとき

$$p = 1 - \sum_{d=0}^{D-1} \frac{E^d}{d!} e^{-E}$$

Microsoft Excelでは $p = 1 - \text{POISSON}(D-1, E, \text{TRUE})$ で求められる。

を計算し、pが0.025より小さいとき5%で有意、0.005より小さいとき1%で有意と判定する。

<p>POISSON 関数項目</p> <p>ポアソン確率分布の値を返します。通常、ポアソン分布は一定の時間内に起きる事象の数を予測するために利用されます。たとえば、ポアソン分布を使って、高速道路の料金所を1分間に通過する自動車の台数を予測することができます。</p> <p>書式</p> <p>POISSON(イベント数, 平均, 累積形式)</p> <p>イベント数 生じる事象の数を指定します。</p> <p>平均 一定の時間内に起きる事象の平均値を指定します。</p> <p>累積形式 確率分布を計算する関数形式を、論理値で指定します。累積形式に TRUE を指定すると、生じるランダムな事象の数が0から イベント数 の範囲であるような累積ポアソン確率分布関数が計算されます。また、累積形式に FALSE を指定すると、生じる事象の数が正確に イベント数 となるようなポアソン確率密度関数が計算されます。</p> <p>解説</p> <ul style="list-style-type: none"> • イベント数に小数点以下の値を指定しても切り捨てられます。 • イベント数 または 平均 に数値以外の値を指定すると、エラー値 #VALUE! が返されます。 • イベント数 ≤ 0 の場合、エラー値 #NUM! が返されます。 • 平均 ≤ 0 の場合、エラー値 #NUM! が返されます。 • POISSON 関数は、次のように計算されます。 <p>累積形式 = FALSE の場合</p> $POISSON = \frac{e^{-\lambda} \lambda^x}{x!}$ <p>累積形式 = TRUE の場合</p> $CUMPOISSON = \sum_{k=0}^x \frac{e^{-\lambda} \lambda^k}{k!}$ <p>使用例</p> <p>POISSON(2,5,FALSE) = 0.084224</p> <p>POISSON(2,5,TRUE) = 0.124652</p>

【練習3】 死亡率検定.xlsを開き【例題】を試して下さい。

参考文献

- 1) 新しい疫学 重松逸造、柳川洋監修 (財)日本公衆衛生協会 1991
- 2) Jian Sun, et al., A Simple Method for Calculating the Exact Confidence Interval of the Standardized Mortality Ratio with SAS Function. J Occup Health 38 196-197 1996

6 地域診断データベースの操作

(1) データファイル(起動ソフトはMicrosoft Access97)の種類

ア 人口動態総覧

昭和55年～平成9年。福岡県保健所一市区町村別。

イ 死因・性・年齢階級別死亡数

昭和53年～平成9年。全国及び福岡県保健所一市区町村別。

ウ 性・年齢階級別人口

(ア) 日本人人口 昭和53年～平成9年。全国。

(イ) 国勢調査人口

a 外国人を含む 昭和55、60、平成2、7年。福岡県保健所一市区町村別。

b 日本人のみ 平成2、7年。福岡県保健所一市区町村別。

エ モデル人口

昭和60年。全国。

オ 福岡県生命表

昭和55、60、平成2、7年。

(2) CD-Rの使用条件

- ア OSはWindows95以上であること。
- イ 使用するコンピュータにMicrosoft Access 97がインストールされていること。
- ウ 使用するコンピュータのハードディスクに48MB以上の空き容量があること。
- エ CD-RからハードディスクにAccessファイルをコピーして使用すること。

CD-Rから直接呼び出しても機能の多くは使用可能であるが、死亡数の検索のとき、死因の選択ができな
いため(CD-Rには選択した死因が書き込めない)、正しい検索結果が得られない。

(3) Accessファイル「地域診断1999.mdb」の準備

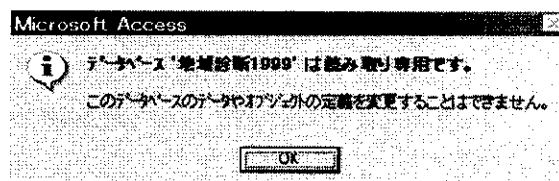
CD-Rの「地域診断1999.mdb」をハードディスクにコピーする。

(デスクトップにコピーする場合の操作法を以下に示す。)

- ア 「地域診断データ ～1999年版～」のCD-RをCD-R読み取り装置にセットする。
- イ マイコンピュータのアイコン¹を左ダブルクリックし、マイコンピュータのウィンドウを開く。
- ウ CD-Rのアイコン「Tiiki」²を左ダブルクリックし、「地域診断1999.mdb」のアイコン³を表示させる。
- エ ⁴にマウスポインターを持っていき、左クリックしながら、デスクトップのウィンドウまでドラッグする。→
コピーが開始される。数十秒待つて、コピー終了。
- オ コピー元のCD-Rは閉じ、CD-R読み取り装置から取り出しておく。
- カ デスクトップ上の「地域診断1999.mdb」のアイコンを右クリックし、プロパティを左クリックする。
- キ 属性のところの読み取り専用の項目がチェックされているので、そのところにマウスポインターを持っていき、
左クリックする。→OKボタンを左クリックする。

【注意】 CD-Rから「地域診断1999.mdb」

のアイコンを直接左ダブルクリックすると、右のメッセージ
が表示される。OKのボタンをクリックすると、地域診断デ
ータのメニュー画面が表示され、使用可能となる。ただ
し、読み取り専用のため、死亡数検索の時の死因選択は
できない。



(4) Accessファイル「地域診断1999.mdb」の操作方法

- ア 「地域診断1999.mdb」のアイコン¹を左ダブルクリックし、メニュー画面(図1)を表示させる。
各項目のボタンを左クリックすることによって、検索作業を進める。

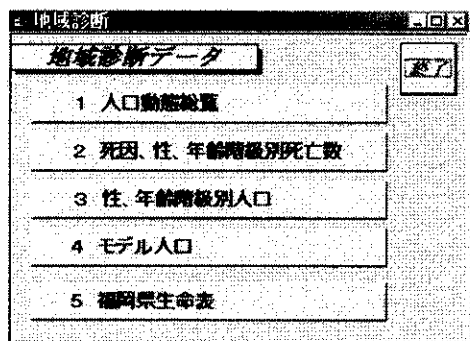


図1 地域診断のメニュー画面

イ 死亡数の検索例

ここでは、筑紫保健所の筑紫野市の男の死亡総数、悪性新生物総数及び各部位の平成4年から平成5年ま
での死亡数を検索し、エクセルファイルに保存する場合の操作例を説明する。

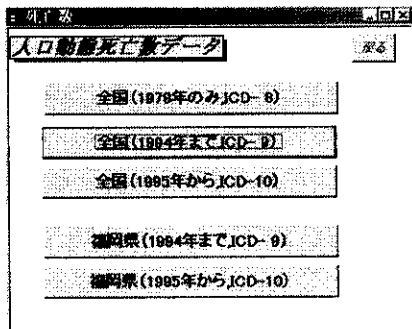


図2 死亡数データ画面

- (ア) 図1の画面から「2 死因、性、年齢階級別死亡数」のボタンを左クリックする。
- (イ) 図2の画面から「福岡県(1994年まで,ICD- 9)」のボタンを左クリックする。

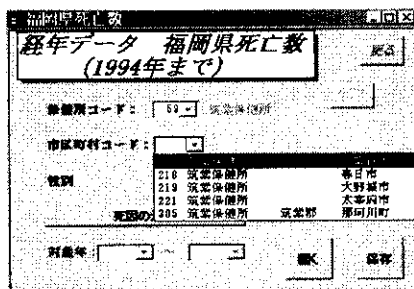


図3 死亡数検索画面

- (ウ) 図3の画面が表示されるので、保健所コードのボタンを左クリックすると保健所一覧が表示されるので、その中から筑紫保健所を左クリックする。
- (エ) 市区町村コード、性別、対象年(ウ)と同様な操作により、筑紫野市、男、平成4年、平成5年を選ぶ。
- (オ) 死因の選択は、「死因の選択」ボタンを左クリックすると、図4の死因選択の画面が表示される。

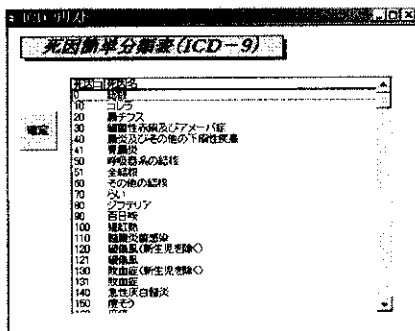


図4 死因選択画面

死因選択の一般的方法を以下に示す。

a 連続している死因を選択するとき

例えば、0 総数から41 胃腸炎までを選択したいときは、総数を左クリックしたまま、胃腸炎までドラッグする。または、総数を左クリックをし、次にマウスポインターを胃腸炎の行に持っていき、shiftキーを押しながら左クリックする。

b 連続していない死因を複数選択するとき

例えば、0 総数、20 腸チフス、50 呼吸器系の結核、51 全結核を選択したいときは、総数を左クリックする。次に、Ctrlキーを押しながら、腸チフス、呼吸器系の結核、全結核の行を左クリックする。ただし、必ずしも総数を第1番目にクリックする必要はない。また、2番目以降の死因を選択するときはCtrlキーを押しながら左クリックすることに気を付ければ、死因をクリックする順序は前後してもかまわない。

総数並びに悪性新生物の総数及び各部位を選択するときは、まず総数のところを左クリックする。次に、死因名の列の右側に付いているスクロールバーにマウスポインターを持っていき左クリックし、悪性新生物の死因名を表示させて、Ctrlキーを押しながら、悪性新生物の各部位を左クリックしていく。

(カ) 死因を選択し終わったならば、確定のボタンを左クリックする。

(キ) 図3の死亡数検索画面に戻るので、開くのボタンを左クリックすると、図5のような検索結果が表示される(データ量が多いため検索に時間がかかる)。

年	性別	保健所	市区町村	年齢	死因	人数	1	2
1992	平成 4年	59 筑紫保健所	217 筑紫野市	1 男	0 総数	197	1	1
1992	平成 4年	59 筑紫保健所	217 筑紫野市	1 男	270 悪性新生物	67	0	0
1992	平成 4年	59 筑紫保健所	217 筑紫野市	1 男	280 食道	5	0	0
1992	平成 4年	59 筑紫保健所	217 筑紫野市	1 男	290 胃	11	0	0
1992	平成 4年	59 筑紫保健所	217 筑紫野市	1 男	300 直腸、直腸S状結腸移行部及び肛門	1	0	0
1992	平成 4年	59 筑紫保健所	217 筑紫野市	1 男	310 肝	14	0	0
1992	平成 4年	59 筑紫保健所	217 筑紫野市	1 男	320 肺	3	0	0
1992	平成 4年	59 筑紫保健所	217 筑紫野市	1 男	330 気管、気管支及び肺	14	0	0
1992	平成 4年	59 筑紫保健所	217 筑紫野市	1 男	360 白血病	2	0	0
1992	平成 4年	59 筑紫保健所	217 筑紫野市	1 男	370 その他	17	0	0
1993	平成 5年	59 筑紫保健所	217 筑紫野市	1 男	0 総数	244	4	0
1993	平成 5年	59 筑紫保健所	217 筑紫野市	1 男	270 悪性新生物	87	0	0

図5 検索結果画面(部分)

- (ク) 図5の右上の閉じるボタンを~~X~~左クリックすると、図3の死亡数検索画面に戻る。
- (ケ) エクセルファイルに保存するとき、保存のボタンを~~X~~左クリックすると、ファイルへの出力画面(図6)が表示されるので、ファイル名を入力し、OKボタンを左クリックすると保存先(この例では、My Documents)に保存される。

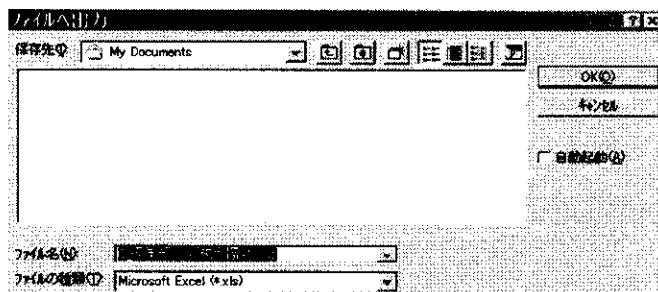


図6 ファイル出力画面

7 地域診断データベースを用いた人口動態解析演習

- (1) 玄海町の低体重児%の年次推移を例に取り、母比率と標本比率との検定や傾向性の検定及び時系列グラフを描く。

ア 1999人口動態解析研修ホルダー^{1999人口動態解析研修}内のKAISEKI.xls及び地域診断研修用.mdbを開き、図7のように、両方のファイルが見えるように、画面を調整する。

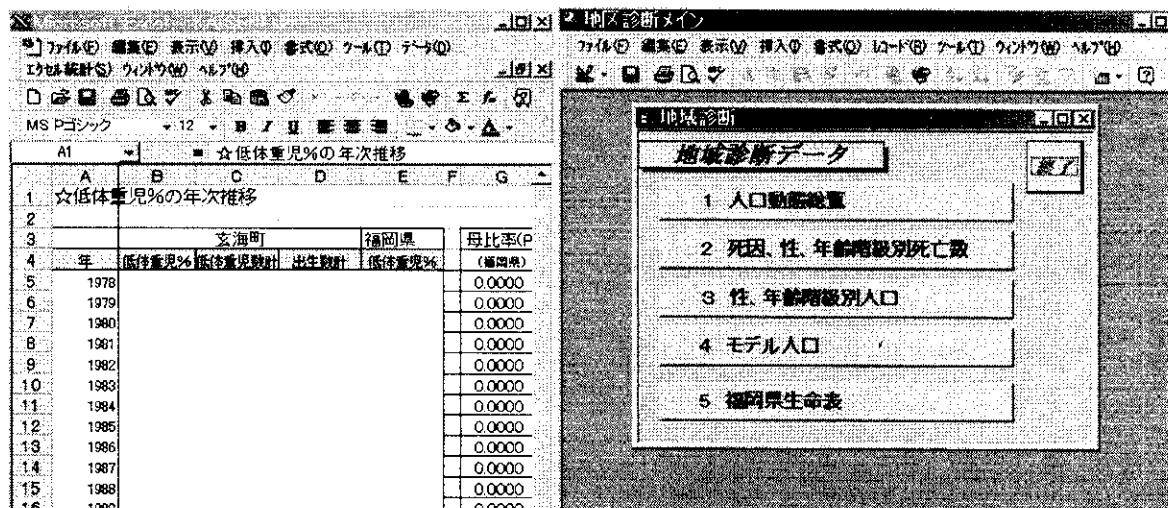


図7 KAISEKI.xlsと地域診断研修用.mdbの表示(一部分)

イ 地域診断研修用.mdbから玄海町の低体重児%、低体重児数計、出生数計及び福岡県の低体重児%データを検索し、KAISEKI.xlsのシートにコピーする。

- (ア) **1 人口動態総覧** クリック→ **経年データ(市区町村)** クリック→図8のように市区町村コード(玄海町)、対象年(昭和53年、平成9年)を選択し、開くボタンをクリックする。

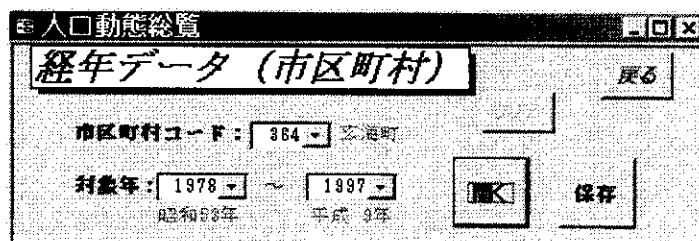


図8 市区町村コード等の入力画面

(イ) 玄海町の低体重児%のコピー&貼り付け操作

- a ①にマウスポインタを置く→マウスポインタが↓印に変わる→クリックする。→図9に示すように黒く反転する→**CTRL** + **C** (コピー)する。

年	低体重児%	低体重児数計	出生数計	低体重児%	母比率P (福岡県)
1978	0.0000	0.43	6	3	0.0000
1979	0.0000	2.64	3	0	0.0000
1980	0.0000	-1.87	3	0	0.0000
1981	0.0000	2.09	4	1	0.0000
1982	0.0000	0.00	4	1	0.0000
1983	0.0000	-1.34	9	4	0.0000
		-4.82	6	6	
		-2.10	3	3	

図9 玄海町の低体重児%のコピー&貼り付けの手順

- b KAISEKI.xlsをアクティブにし(タイトルバーが青い状態)、マウスポインタを②に置き、クリックする→**CTRL** + **V** (貼り付け)する。

(ウ) 玄海町の低体重数計のコピー&貼り付け操作

- a ③にマウスポインタを置きクリックする。→黒く反転する→**CTRL** + **C** (コピー)する。
 b KAISEKI.xlsをアクティブにし(タイトルバーが青い状態)、マウスポインタを④に置き、クリックする→**CTRL** + **V** (貼り付け)する。

(エ) 玄海町の出生数計のコピー&貼り付け操作

データベースから出生数計の列を見つけたのち、(イ)あるいは(ウ)と同様な操作をし、玄海町の出生数計のカラムに貼り付ける。

(オ) 福岡県の低体重児%のコピー&貼り付け操作

- a 図10の画面まで戻す。

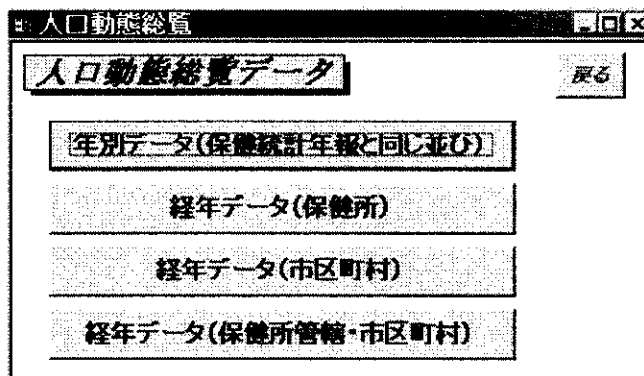


図10 人口動態総覧データメニュー画面

- b **経年データ(保健所)** クリック→図11のように保健所コード、対象年を選択→開くボタンをクリックする。

図11 福岡県の保健所コード及び対象年の入力

- c 図12のように①の列を[CTRL] + [C] (コピー)する。
 d KAISEKI.xlsをアクティブにし(タイトルバーが青い状態)、②の列に[CTRL] + [V] (貼り付け)する。

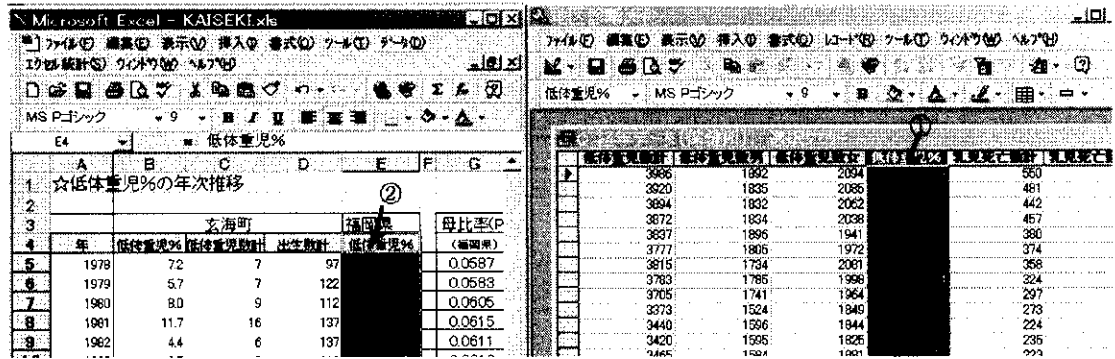


図12 福岡県の低体重児%のコピー&貼り付けの手順

ウ 母比率と標本比率との検定の解説¹⁾

検定の手順

(ア) 帰無仮説 H_0 と対立仮説 H_1 をたてる。

$H_0: p_s = P$ 標本比率(p_s)と母比率(P)は等しい。→玄海町の低体重児%と福岡県の低体重%は等しい。

$H_1: p_s \neq P$ 標本比率と母比率は等しくない。→玄海町の低体重児%と福岡県の低体重%は等しくない。

(イ) 有意水準を決める

多くの場合は有意水準(α) = 5%(0.05、両側)でよいが、次のような考え方もある。

- a 採用する対立仮説が既存の概念や知識と異なる場合、 $\alpha = 5\%$ では説得力がなく、 $\alpha = 1\%$ (0.01、両側)とすべきである。
- b 標本数が少ないが、常識にかなった結果を得ている場合、これ以上実験を続けるかどうかの判断としては5%では厳しすぎる。 $\alpha = 10\%$ (0.1、両側)で設定して、第1報(preliminary report)として発表してよい。

(ウ) 検定統計量を求める。

標本数を n 、母比率を P ($Q = 1 - P$)とすると、次の3とおりに別れる。

- a $nP \geq 5$ のとき→正規近似による。 (a)
- b $nP < 5$
 - $n < 50$ のとき→F検定式による。 (b)
 - $n \geq 50$ のとき→ポアソン分布による。 (c)

$nP =$ 標本数 \times 母比率で、標本集団における期待値を示す。

(a) 正規近似による。

$$z_0 = \frac{|p_s - P| - \frac{1}{2n}}{\sqrt{\frac{P(1-P)}{n}}} = \frac{|x - nP| - \frac{1}{2}}{\sqrt{nP(1-P)}} \quad \text{ただし、} p_s = \frac{x}{n}$$

(b) F検定式による。

① $x > nP$ のとき

$$F_0 = \frac{n_2 Q}{n_1 P} \quad \text{ただし、} n_1 = 2(n - x + 1), n_2 = 2x$$

② $x < nP$ のとき

$$F_0 = \frac{n_2 P}{n_1 Q} \quad \text{ただし、} n_1 = 2(x + 1), n_2 = 2(n - x)$$

(c) ポアソン分布による。

① $x > nP$ のとき

$$p = 1 - \sum_{i=0}^{x-1} \frac{e^{-\lambda} \times \lambda^i}{i!} \quad \text{ただし、} \lambda = nP$$

② $x < nP$ のとき

$$p = \sum_{i=0}^x \frac{e^{-\lambda} \times \lambda^i}{i!} \quad \text{ただし、} \lambda = nP$$

(エ) Microsoft Excel等の関数を利用し、確率を求める。

a 正規近似のとき。

$$p = 1 - \text{NORMSDIST}(z_0)$$

NORMSDIST 関数項目
標準正規累積分布関数の値を返します。この分布は、平均が 0(ゼロ) で標準偏差が 1 である正規分布に対応します。正規分布表の代わりにこの関数を使用することができます。
書式 NORMSDIST(x)
x 関数に代入する値を指定します。
解説
<ul style="list-style-type: none"> x に数値以外の値を指定すると、エラー値 #VALUE! が返されます。 正規分布の標準密度関数は次の式で定義されます。 $f(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
使用例
NORMSDIST(1.333333) = 0.908789

b F検定式のとき。

$$p = \text{FDIST}(F_0, n_1, n_2)$$

FDIST 関数項目
F 確率分布を返します。この関数を使用すると、2 組のデータを比較して、ばらつきが両者で異なるかどうかを調べることができます。たとえば、テストの成績を男女別に分析して、男子生徒の成績と女子生徒の成績のばらつきが異なるかどうかを検定することができます。
書式 FDIST(x, 自由度1, 自由度2)
x 関数に代入する値を指定します。
自由度1 自由度の分子を指定します。
自由度2 自由度の分母を指定します。
解説
<ul style="list-style-type: none"> x に数値以外の値を指定すると、エラー値 #VALUE! が返されます。 x < 0 である場合、エラー値 #NUM! が返されます。 自由度1, 自由度2 に整数以外の値を指定すると、小数点以下が切り捨てられます。 自由度1 < 1 または 自由度1 ≥ 1010 である場合、エラー値 #NUM! が返されます。 自由度2 < 1 または 自由度2 ≥ 1010 である場合、エラー値 #NUM! が返されます。 FDIST 関数は、数式 FDIST=P(F<x) で表されます (F は F 確率分布を持つ任意の実数)。
使用例
FDIST(15.20675, 6, 4) = 0.01

c ポアソン分布のとき。

(a) $x > nP$ のとき

$$p = 1 - \text{POISSON}(x-1, nP, \text{TRUE})$$

(b) $x < nP$ のとき

$$p = \text{POISSON}(x, nP, \text{TRUE})$$

POISSON	
関数項目	
	ポアソン分布の値を返します。通常、ポアソン分布は一定の時間内に起きる事象の数を予測するために利用されます。たとえば、ポアソン分布を使って、高速道路の料金所を1分間に通過する自動車の台数を予測することができます。
書式	
	POISSON(イベント数, 平均, 累積形式)
	イベント数 生じる事象の数を指定します。
	平均 一定の時間内に起きる事象の平均値を指定します。
	累積形式 確率分布を計算する関数形式を、論理値で指定します。累積形式に TRUE を指定すると、生じるランダムな事象の数が 0 から イベント数の範囲であるような累積ポアソン分布関数が計算されます。また、累積形式に FALSE を指定すると、生じる事象の数が正確に イベント数 となるようなポアソン確率密度関数が計算されます。
解説	
	<ul style="list-style-type: none"> イベント数に小数点以下の値を指定しても切り捨てられます。 イベント数または平均に数値以外の値を指定すると、エラー値 #VALUE! が返されます。 イベント数 ≤ 0 の場合、エラー値 #NUM! が返されます。 平均 ≤ 0 の場合、エラー値 #NUM! が返されます。 POISSON 関数は、次のように計算されます。
	累積形式 = FALSE の場合
	$POISSON = \frac{e^{-\lambda} \lambda^x}{x!}$
	累積形式 = TRUE の場合
	$CUMPOISSON = \sum_{k=0}^x \frac{e^{-\lambda} \lambda^k}{k!}$
使用例	
	POISSON(2,5,FALSE) = 0.084224
	POISSON(2,5,TRUE) = 0.124652

(オ) 仮説H₀を採用するか棄却するか判定する。

- a p × 2 < α のときH₀を棄却する。すなわち危険率αで有意差ありと判定する。p < αと表示する。
- b p × 2 ≥ α のときは有意差なしとする。n.s.(not significant)と表示する。

エ カイ2乗一傾向性の検定の解説¹⁾

分類カテゴリが順序尺度(年次、年齢階級などカテゴリの順序が入れ替えられないような尺度)の時は、χ²傾向性の検定も考慮する。

年次	低体重児数	出生数
B ₁	x ₁	y ₁
B ₂	x ₂	y ₂
B ₃	x ₃	y ₃
⋮	⋮	⋮
B _i	x _i	y _i
⋮	⋮	⋮
B _j	x _j	y _j
計	n ₁	n ₂

(ア) 全体のχ²を計算する。

- a H₀:年次による低体重児%の差はない。
- b 有意水準 α = 0.05

c 検定統計量

$$\chi^2 = \frac{n_2}{n_1(n_2 - n_1)} \left(\frac{x_1^2}{y_1} + \frac{x_2^2}{y_2} + \dots + \frac{x_j^2}{y_j} - \frac{n_1^2}{n_2} \right)$$

d Microsoft Excel関数を利用して確率を求める。

$$p = \text{CHIDIST}(\chi^2, j-1)$$

e 仮説H₀を採用するか棄却するか判定する。

p < α のとき、有意差あり。

p ≥ α のとき、有意差なし。

CHIDIST
 関数項目

片側カイ2乗 (χ^2) 分布の確率を返します。 χ^2 分布は χ^2 検定と関連しています。 χ^2 検定は、実測値と期待値を比較するときに使用します。たとえば、ある植物の遺伝子実験で、次の世代の母口は一定の色の組み合わせが発生するとい仮説を立てたします。ここで、予測された色と観察の結果を比較することにより、仮説の妥当性を検証することができます。

書式
CHIDIST(*x*, *自由度*)

x 分布を評価する値を指定します。
自由度 自由度を指定します。

解説

- 引数に数値以外の値を指定すると、エラー値 #VALUE! が返されます。
- x* に負の数を指定すると、エラー値 #NUM! が返されます。
- 自由度* に整数以外の値を指定すると、小数点以下が切り捨てられます。
- 自由度* < 1 または *自由度* \geq 10¹⁰ である場合、エラー値 #NUM! が返されます。
- CHIDIST 関数は、数式 CHIDIST = P(0<*x*) で表されます ($\alpha = \chi^2$ 確率変数)。

使用例
 CHIDIST(18.307,10) = 0.050001

(イ) 回帰による χ^2 を計算する。

年次	低体重児数	出生数	得点	$a_i \times x_i$	$a_i \times y_i$	$a_i^2 \times y_i$
B ₁	x ₁	y ₁	a ₁	a ₁ × x ₁	a ₁ × y ₁	a ₁ ² × y ₁
B ₂	x ₂	y ₂	a ₂	a ₂ × x ₂	a ₂ × y ₂	a ₂ ² × y ₂
B ₃	x ₃	y ₃	a ₃	a ₃ × x ₃	a ₃ × y ₃	a ₃ ² × y ₃
⋮	⋮	⋮	⋮	⋮	⋮	⋮
B _i	x _i	y _i	a _i	a _i × x _i	a _i × y _i	a _i ² × y _i
⋮	⋮	⋮	⋮	⋮	⋮	⋮
B _j	x _j	y _j	a _j	a _j × x _j	a _j × y _j	a _j ² × y _j
計	n ₁	n ₂		m ₁	m ₂	m ₃

- H₀: 回帰による傾向は認められない。
- 有意水準 $\alpha = 0.05$
- 検定統計量

$$\chi^2 = \frac{n_2(n_2 \times m_1 - n_1 \times m_2)^2}{n_1(n_2 - n_1)(n_2 \times m_3 - m_2^2)}$$
- Microsoft Excel関数を利用して確率を求める。
 $p = \text{CHIDIST}(\chi^2, 1)$
- 仮説H₀を採用するか棄却するか判定する。
 $p < \alpha$ のとき、傾向性あり。
 $p \geq \alpha$ のとき、傾向性が認められない。

(ウ) 仮想的な回帰線からの偏り χ^2_{U} を計算する。

- H₀: 回帰直線からの偏りはない。
- 有意水準 $\alpha = 0.05$
- 検定統計量

$$\chi^2_{\text{U}} = \chi^2 - \chi^2$$
- Microsoft Excel関数を利用して確率を求める。
 $p = \text{CHIDIST}(\chi^2_{\text{U}}, j-2)$
- 仮説H₀を採用するか棄却するか判定する。
 $p < \alpha$ のとき、偏りがある。
 $p \geq \alpha$ のとき、偏りが認められない。