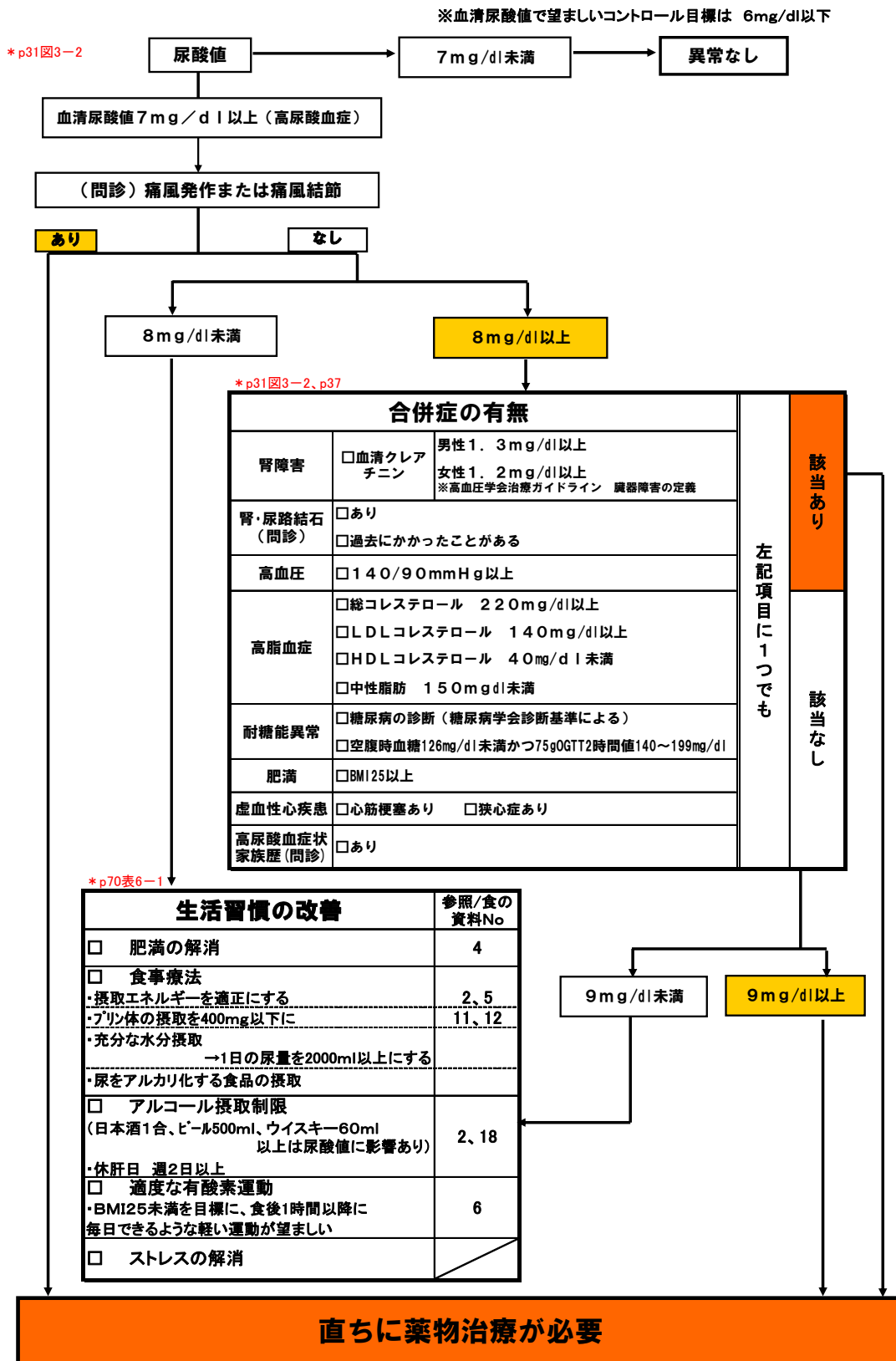


④高尿酸血症予防と治療の階層化フロー図



* は日本痛風・核酸代謝学会編「高尿酸血症・痛風治療ガイドライン」参照箇所

『健診データの活用と分析』(青森県版)

目次

1. はじめに	p1 (P265)
2. 健診データから分かること	p1
3. データ分析の実際	p2
(1)表形式のデータベースの準備	p2
(2)エクセルのデータベース機能を用いた分析演習 p3	
a)オートフィルタを用いる分析	p4
b)ピボットテーブルを用いた分析	p6
(3)統計学的検定	p14
4. 最後に	p16
(参考)	p16

1. はじめに

保健戦略を立てるためには、地域の課題を正確に把握し、精度の高い地域診断を行うことが求められます。健診結果はその意味で、地域の健康指標の特性を知り、対策・事業の根拠を求めることや、介入の効果を見ることができる格好な資料です。ここでは、健診機関から提供された電子媒体を活用して、健診データから得られる分析を例示します。ここに示すのはほんの一例ですが、これを参考にしつつ、健診データを幅広く活用して頂きたいと思います。

なお、これは平成17、18年度に青森県上十三保健所が管内8市町村の保健担当者及び保健所職員を対象に行った研修事業「老人保健事業による基本健康診査のデータ分析と地域診断」の内容の一部を教材化したものです。教材の内容は、健診データについて、エクセルをデータベースとして用いる場合の初歩を解説したものとなっています。

CD-ROM 版ではエクセルファイル「健診データ分析演習事例」を収録してありますので、ここで示された分析例をすべて実際に行ってみることができます。ぜひ試してみてください。

2. 健診データから分かること

健診データから種々の地域の健康指標を分析できますが、概略、以下のことを得られるでしょう。

- ・地域全体の特性を探る

他市町村・県・国と比較することで地域特性を明確にできる
対策・事業の根拠とする

- ・地域の変化を時系列的(年次的)に把握する
- ・地域内のグループ毎の比較ができる(地区、職業、性別等々)
- ・事業等による介入の効果を見る

比較して差異が認められたとしても、その差が統計的に有意であるかどうかを見るには検定が必要です。また、比較にあたっては厳密には年齢で標準化する必要があります。でも大事なことは、あまり厳格に考えず、とにかく分析にトライしてみることです。その過程でさまざまな課題が浮き彫りにされてくるでしょう。

3. データ分析の実際

(1) 表形式のデータベースの準備

表計算ソフトの Excel 2003 を用います。健診結果をエクセルで開けるようになった状態から始めます。健診機関からの生データをエクセルで開くまでの手順については巻末に(参考)で紹介しておきます。

エクセルはデータベースとしては制約がありますが(次ページ[コラム]参照)、「抽出」、「並べ替え」、「集計」、「ピボットテーブル」などの機能は使い勝手がよく、簡単な検定もサポートしており、レセプトや健診結果の分析には適しています。特に制約が大幅に少なくなった Excel 2007 は更にお薦めです。

エクセルをデータベースとして用いる場合、第1行目に項目名を入れます。項目名は空白があってははいけません。各項目名は列の名前になります(データベース上、列は「フィールド」で、列の名前は「フィールド名」といいます)。各人のデータは第2行目からとして下さい。1人のデータは1行になります(行は「レコード」といいます)。

ちなみに、第1行目に表題、第2行目を空白、第3行目に項目名、第4行目からレコードという形でも差し支えありません。その場合でも項目名は1行のみ使用してください。見やすくするため項目に2, 3行使い、階層化するという工夫は、データベース機能を使う場合は障害となります。

なお、第1行目の項目名を眺めてみて、分かりにくい名前、あるいは他と同じ名前などあれば、一つずつ区別できる名前にした方がよいでしょう。例えば「血液検査11」を「HDL-C」に替えておくなど工夫します。

今回分析の対象としたデータでは、年齢の項目がなく、8桁の「西暦生年月日」と、同じく8

桁の「受診年月日」から年齢を出す必要がありました。そこで年齢を算出する関数式をお示ししておきます。いろいろな方法がありますが、ここではより簡便に概算的に「=INT((検査日セル-誕生日セル)/10000)」で算出します。殆ど正確で実用上はこれでまず問題ありません。すなわち 1 列挿入し 1 行目に項目名を付加し、2 行目にこの式を入れ、概算年齢(殆ど正確である)を出し、後はドラッグするなどしてコピーします(図1)。

①列を増やして、1 行目に項目名を付加します

②2 行目に年齢を出す式を入れます

③ドラッグでコピーします(裏技的に、セルの右下隅にカーソルを置いて+を表示させ、そのままダブルクリックするといっぺんにコピーできます)

	A	F	G	H	I	性別
1	受診年月日	生年元号	和暦生年月	西暦生年月	年齢	
2	20040705	3	120203	19370203	67	
3	20040705	3	140910	19390910		
4	20040705	3	130101	19380101		
5	20040705	3	120929	19370929		
6	20040705	3	120515	19370515		
7	20040705	3	120531	19370531		

図1

これでエクセルのデータベース機能を使う準備が整いました。では、早速データの分析を行ってみましょう。

[コラム]エクセルの制約について

エクセルは Excel 2003 までは、列数が 256 に限られ、行数も 6 万 5536 という制限がありました。従って、健診データに予備項目をたくさん入れると 256 を超えてしまうことがあります。現に今回利用した青森県の健診機関の基本健康診査結果は列セル数が 275 あり、そのままでは、エクセルで読み込むと情報が欠けてしまいました。そのため、エクセルで読みこめるよう前処理が必要でした。また、行の制約によって、分析する健診者が多くなると全員のデータを読み込めないことが起こります。

しかし、19 年 1 月に発売された Excel 2007 ではその点が大幅に改善され、1 万 6384 列、104 万 8576 行が使えます。これはざっと 1 万項目、100 万人のデータベースを構築できるということですから、殆どの市町村では十分な数字です。今回の分析は普及度を考えて Excel 2003 を用いましたが、可能な人は Excel 2007 を使って分析すると良いでしょう。

(2) エクセルのデータベース機能を用いた分析演習

ここではエクセルの代表的なデータベース機能である、オートフィルタとピボットテーブルを用いた分析を例行的に行ってみます。データはある町のもので、ある年の基礎健診受診者は1594名でした。このうち65歳以上の割合は61.7%です。一人ずつの項目(レコード)は選択して40個(列ではAからANまで)に絞っています。このデータについて分析していきます。

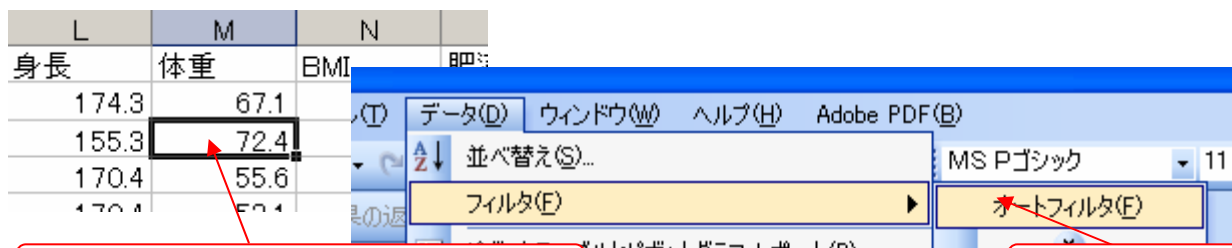
a) オートフィルタを用いる分析

オートフィルタは複数の条件を持った人(例えば、メタボリック・シンドロームの人)を抽出するのに大変便利な機能です。例題を参考に自市町村の健診データを使って、いろいろな角度から抽出作業を行ってみましょう。オリジナルデータは腹囲がないので、肥満に関してはBMIを指標にしています。

〔例題1〕 肥満(BMI25以上)、喫煙、高血圧傾向(最高血圧130以上)のすべてを持った人はどのくらいいますか。

〔解説〕

オートフィルタを使うには、データの入っているセルのいずれかをクリックしておいて、「データ」から「オートフィルタ」をチェックします(図2)。

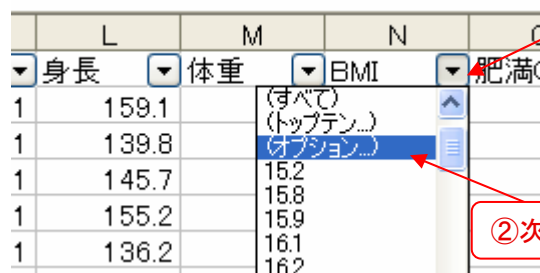


①まずデータセルのどれかをクリックし

②クリックします

図2

まず、BMI25以上の人を抽出してみます。項目BMIの横の を ックし、オプションを選んで(図3)、さらにオプションを選択し、図4のようなオプション選択画面が表示されたら、25以上という抽出条件を指定します。



①まずここをクリックし

②次に、ここをクリックします

図3

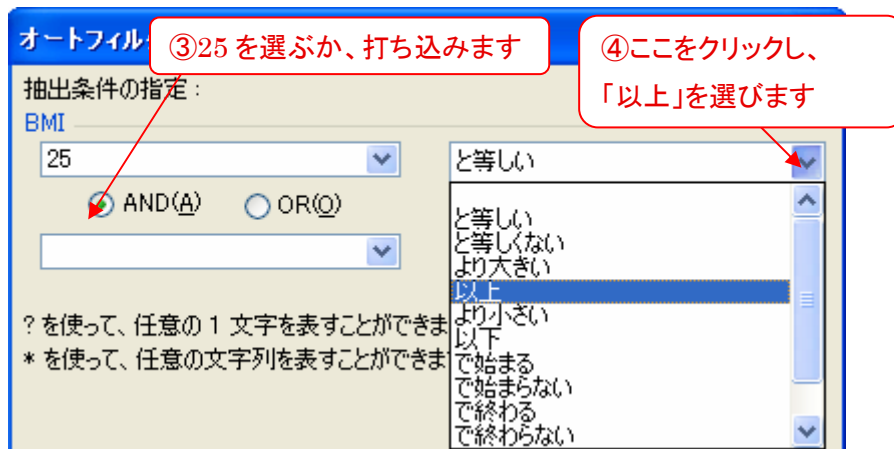


図4

BMI の条件で抽出できたら、続けて現在喫煙している人(「喫煙分類」3)を同じように抽出し、更に最高血圧 130 以上の人を抽出します。そうすると、25 人が抽出されました。全体に対する比率は 1.6%と少ないものですが、例えば図5のように、年齢順に並べてみると、比較的若年層も多いなどの具体的な問題点が見えてくるでしょう。