



## Excelによる演習3の作業手順



千葉県衛生研究所  
健康疫学研究室  
2008.8.19

1

## 既存データの活用 演習3

### 【目的3】 データ集計の基本作業の習得

- 1) 連続量を含むデータの要約・作図ができる
- 2) カテゴリー分けを適切にできる
- 3) 複数の条件を持つ人を抽出できる  
(オートフィルターの活用)
- 4) クロス集計ができる(ピボットテーブルの活用)

性・年齢階級別に、肥満や生活習慣病のハイリスク者の割合を求め、対策を考える資料を作成する。

2

### 演習3 連続量を含むデータの分析

演習3のシートを開いてください  
データは、A市で行った基本健康診査の測定結果の一部です

\* このデータを用いて、生活習慣病対策の資料を作成したい。

#### 対象者の性別の特性を表す

1. 男性、女性の人数を数える
2. 年度末年齢を求め、性別に記述統計で示す
3. 年度末年齢を10歳階級に分け、度数分布を示す
4. BMIを小数点1桁で求める(計算式によるデータの追加)  
$$\text{BMI} = \frac{\text{体重(kg)}}{\text{身長(m)}^2}$$
5. BMIの判定区分(3区分)をデータに追加(カテゴリーの作成)
6. 50歳代で、BMI判定区分が「肥満」または「やせ」を抽出する  
(オートフィルターの利用による対象の抽出)
7. 50歳代女性において、BMI判定区分が「肥満」と「やせ」の人の収縮期血圧平均値と標準偏差をそれぞれ求め、作図する。

3

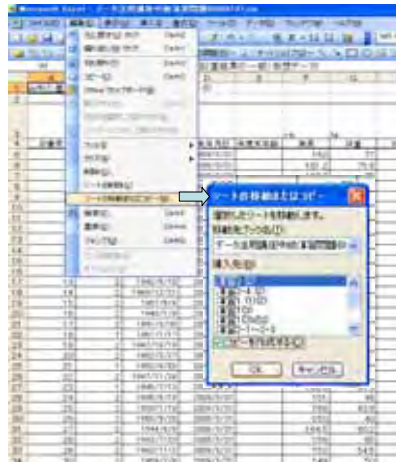
### 演習3 連続量を含むデータの分析

8. 性・年齢階級別のBMI判定区分の分布を作表する  
(ピボットテーブルを使ったクロス集計表の作成)
9. 作成したクロス表を使い、「40~59歳」「60~79歳」の2群について、BMI判定区分が「肥満」または「やせ・標準」の性別の数を表示させる
10. 性・年齢・肥満の有無別に、血圧、中性脂肪の記述統計(人数、平均、標準偏差、最大値、最小値)を求める。  
(ピボットテーブルを使い、層別に記述統計を作成する)
11. 総コレステロールとHDLコレステロール値の関係をみるため、性別に散布図を作成し、相関関係があるかを考える  
(散布図の作成)

4

準備 単位の行を表から切り離しておく

演習3のシートを開いてください



これから作業を行うため、シートのコピーを2枚作成します。

編集  
↓  
シートの移動またはコピー  
↓  
演習3を選択し、コピーを作成するにチェック  
↓  
OK

\* 演習3のシートの前に演習3(2)のシートができます。同様に、もう1枚コピーし、演習3(3)のシートを作ります。

準備 作成したコピーの演習3(2)を利用する

変数の説明の行(最初の行)と、変数名以下を切り離す



3行目と4行目の間に1行空白の行を入れます。

これで、変数名1行、データ1ケース1行のデータベースができました。

演習3-1 該当するカテゴリに入る数を数える

男性、女性の人数を数える

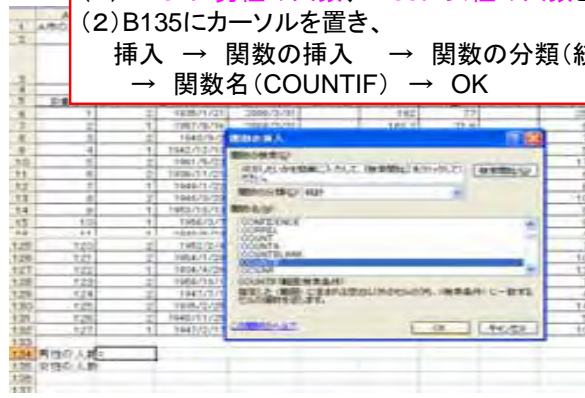
1) COUNTIF関数を使う

(1) A134に男性の人数、A135に女性の人数と入力

(2) B135にカーソルを置き、

挿入 → 関数の挿入 → 関数の分類(統計)

→ 関数名(COUNTIF) → OK とする。



演習3-1 該当するカテゴリに入る数を数える

男性、女性の人数を数える

1) COUNTIF関数を使う

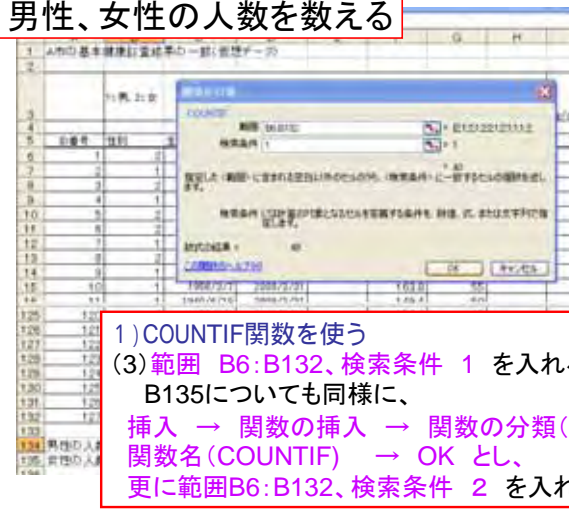
(3) 範囲 B6:B132、検索条件 1 を入れる。

B135についても同様に、

挿入 → 関数の挿入 → 関数の分類(統計) →

関数名(COUNTIF) → OK とし、

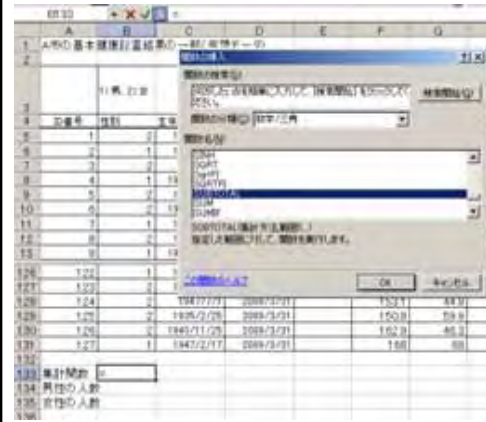
更に範囲B6:B132、検索条件 2 を入れる。



### 演習3-1 該当するカテゴリに入る数を数える

男性、女性の人数を数える

演習3(3)のシートを使います

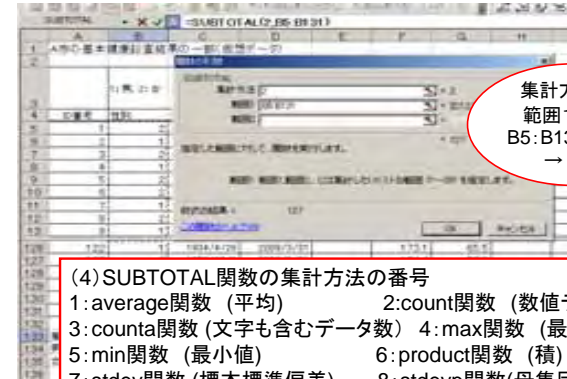


- 2)抽出と集計機能を使う  
 (1)演習3(3)のシートを開きます。  
 (2)A133に”集計関数“、A134に”男性の人数“、A135に”女性の人数“と記入します。  
 (3)B133のセルに、カーソルを置き、SUBTOTAL関数で集計様式を決めます。

### 演習3-1 該当するカテゴリに入る数を数える

男性、女性の人数を数える

演習3(3)のシートを使います



- (4)SUBTOTAL関数の集計方法の番号  
 1:average関数(平均) 2:count関数(数値データ数)  
 3:counta関数(文字も含むデータ数) 4:max関数(最大値)  
 5:min関数(最小値) 6:product関数(積)  
 7:stdev関数(標本標準偏差) 8:stdevp関数(母集団標準偏差)  
 9:sum関数(総和) 10:var関数(標本の分散)  
 11:varp関数(分散)

### 演習3-1 該当するカテゴリに入る数を数える

男性、女性の人数を数える

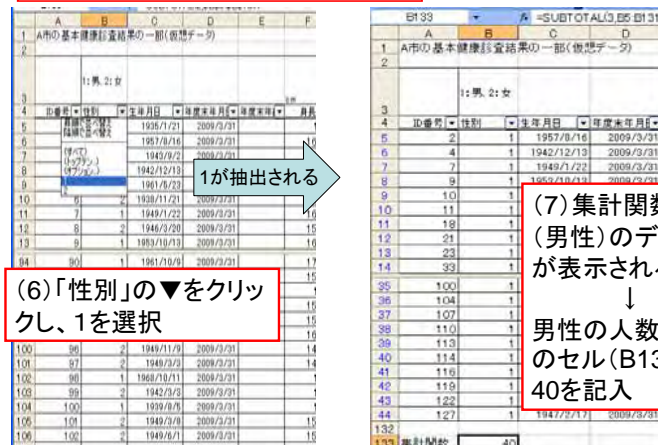
演習3(3)のシートを使います



- (5)性別データセルの1つを選択し、  
 データ → フィルタ → オートフィルタ を選択

### 演習3-1 該当するカテゴリに入る数を数える

男性、女性の人数を数える



- (6)「性別」の▼をクリックし、1を選択

- (7)集計関数に1(男性)のデータ数が表示される  
 ↓  
 男性の人数の隣のセル(B134)に40を記入



### 演習3-1 該当するカテゴリに入る数を数える

氏名	性別	生年月日
1		
2		
3		
4	男	2/1935/1/21
5	女	2/1949/2/1
6	女	2/1961/6/5
7	女	2/1939/1/11
8	女	2/1946/2/2
9	女	2/1942/8/12
10	女	2/1940/6/13
11	女	2/1965/12/14
12	女	2/1957/2/15
13	女	2/1988/2/11
14	女	2/1940/3/2
15	女	2/1926/2/11
16	女	2/1967/9/21
17	女	2/1965/6/13
18	女	2/1952/2/12
19	女	2/1954/1/1
20	女	2/1959/1/23
21	女	2/1947/2/12
22	女	2/1925/2/19
23	女	2/1940/1/11
24	合計関数	87
25	男性の人数	41
26	女性の人数	87

男性、女性の人数を数える

(8)「性別」の▼をクリックし、2を選択

集計関数に2(女性)のデータ数が表示される

女性の人数の隣のセル(B135)に87を記入

オートフィルタでの抽出内容を変えると、SUBTOTAL関数の集計値が変化し、フィルタ抽出された分だけの結果が表示されます。

13

### 演習3-2 年齢を求め、性別に記述統計で示す

年度末年齢を計算する

演習3(2)のシートを出します。

氏名	性別	生年月日	年度末年齢
1			
2			
3			
4	男	2/1935/1/21	2009/3/31
5	女	2/1949/2/1	2009/3/31
6	女	2/1961/6/5	2009/3/31
7	女	2/1939/1/11	2009/3/31
8	女	2/1946/2/2	2009/3/31
9	女	2/1942/8/12	2009/3/31
10	女	2/1940/6/13	2009/3/31
11	女	2/1965/12/14	2009/3/31
12	女	2/1957/2/15	2009/3/31
13	女	2/1988/2/11	2009/3/31
14	女	2/1940/3/2	2009/3/31
15	女	2/1926/2/11	2009/3/31
16	女	2/1967/9/21	2009/3/31
17	女	2/1965/6/13	2009/3/31
18	女	2/1952/2/12	2009/3/31
19	女	2/1954/1/1	2009/3/31
20	女	2/1959/1/23	2009/3/31
21	女	2/1947/2/12	2009/3/31
22	女	2/1925/2/19	2009/3/31
23	女	2/1940/1/11	2009/3/31

(1)年齢を求める式をセルに入れる  
=datedif(起算日, 終了日, 表示形式)  
表示形式・Y(年)、M(月)、D(日)

(2)セルE6を選択し、コピー  
E7からE132まで選択し、貼り付ける。

14

### 演習3-2 年齢を求め、性別に記述統計で示す

男性の年齢の記述統計を求める

(1)性別に並び替える(オートフィルタまたはソートを使用)  
(2)分析ツールを使って「基本統計量」を求める  
ツール → 分析ツール  
→ 基本統計量 → OK

15

### 演習3-2 年齢を求め、性別に記述統計で示す

男性の年齢の記述統計を求める

(3)入力範囲 E6:E45  
出力先 新規ワークシート  
出力内容 統計情報 と指定→OK

項目	値
平均	6015
標準偏差	1.630066
中央値	62
総数	69
標準偏差 (σ)	10.30944
分散	106.2846
尖度	-0.65465
歪度	-0.62496
範囲	94
最小	40
最大	74
合計	2406
標本数	40
最大値(σ)	74
最小値(σ)	40

列1を年齢の男性の年齢に書き換える。女性も同様に年齢の記述統計を求める

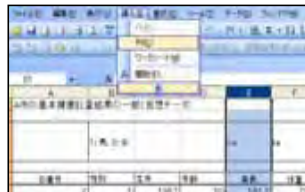
男女それぞれの年齢の記述統計を見て、分かったことを列記しなさい。

16

演習3-3 年齢を10歳階級に分け、度数分布を示す

年齢を10歳階級に分ける

10歳階級の作り方



- 1) 年齢の右に列を挿入  
挿入 → 列
- 2) セルF5に「年齢階級」と変数名を入力する

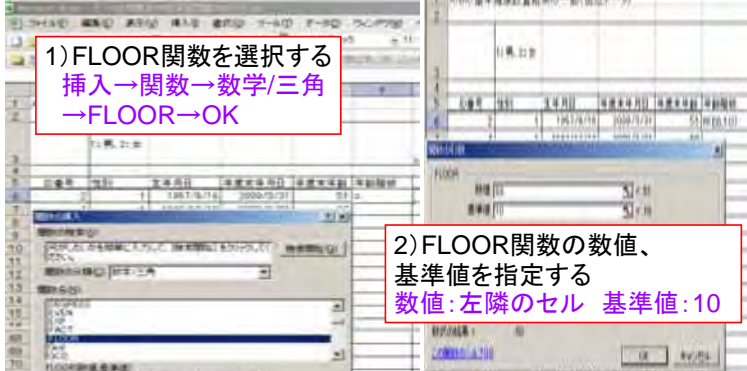
INT関数: 切り捨てて整数にする  
TRUNC関数: 数値の小数部を切り捨て、整数または指定した桁数に変換

- 1) 年齢を10で割る。  
例:  $74 \text{歳} = 70$   
 $= \text{INT}(\text{年齢} / 10)$   
 $= \text{TRUNC}(\text{年齢} / 10)$   
注: 70歳代を70と標記するには、式の/10の後に\*10として、結果を10倍する。
- 2) FLOOR関数を使う  
指定された基準値倍数のうち、最も近い値に数値を切り捨てる。  
例:  $\text{FLOOR}(54, 10)$ は54を最も近い10の倍数に切り捨てる = 50  
注: 50歳代は5と標記するには、結果を10で割る。

17

演習3-3 年齢を10歳階級に分け、度数分布を示す

10歳階級を作成する



- 1) FLOOR関数を選択する  
挿入→関数→数学/三角  
→FLOOR→OK

- 2) FLOOR関数の数値、基準値を指定する  
数値: 左隣のセル 基準値: 10

- 3) セルF6を、F7からF132までにコピー

18

演習3-3 年齢を10歳階級に分け、度数分布を示す

度数分布の作成



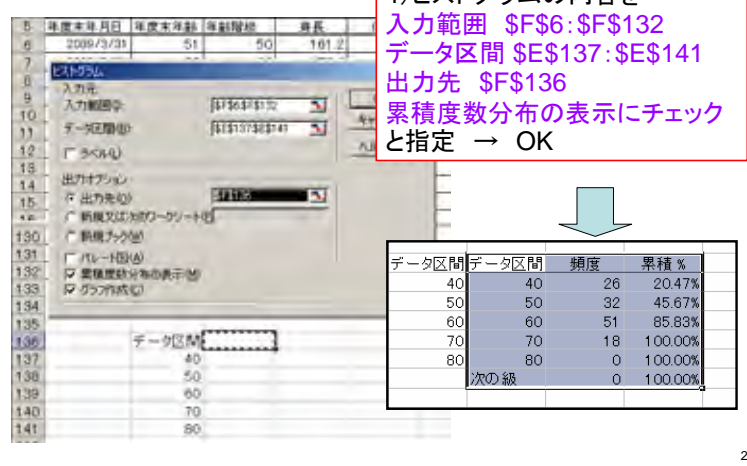
- 1) ヒストグラム関数を使用するため、データ区間をあらかじめ作成する。  
10歳階級のデータの下の(F136)にデータ区間と入力し、F137からF146に40~80を縦に入力

- 2) 度数分布表を作成  
分析ツール  
→ヒストグラム

19

演習3-3 年齢を10歳階級に分け、度数分布を示す

度数分布の作成



- 4) ヒストグラムの内容を  
入力範囲 \$F\$6:\$F\$132  
データ区間 \$E\$137:\$E\$141  
出力先 \$F\$136  
累積度数分布の表示にチェックと指定 → OK

データ区間	データ区間	頻度	累積%
40	40	26	20.47%
50	50	32	45.67%
60	60	51	85.83%
70	70	18	100.00%
80	80	0	100.00%
	次の級	0	100.00%

20

### 演習3-3 年齢を10歳階級に分け、度数分布を示す

#### 作成された度数分布表

データ区間	頻度	累積%
40	26	20.47%
50	32	45.67%
60	51	85.83%
70	18	100.00%
80	0	100.00%
次の級	0	100.00%

年齢の度数分布表を見て、分かったことを列記しなさい。

#### 応用問題: 5歳階級を作成するにはどうしたら良いか

##### 1) 年齢の割り算を使う場合

44歳 40歳代(40)、45歳 45歳代(45)、49歳 45歳代(45)となるように計算方法を考えること。

##### 2) FLOOR関数を使う場合

FLOOR関数の意味が分かれば、式は書けます。

21

### 演習3-4 BMIを小数点1桁で求める

#### BMIを計算する

The screenshot shows an Excel spreadsheet with a formula bar containing `=ROUND(H6/(G6/100)^2,1)`. Below it is a table with columns for '年齢階級', '身長', '体重', 'BMI', and 'Tb'. The BMI values are rounded to one decimal place.

1) BMIの計算式をI6のセルに記入する

$$=H6/(G6/100)^2$$

$$=H6/(G6 * G6) * 10000$$

体重(kg)/身長(m)<sup>2</sup>

2) このままだと、小数点以下の桁数が多すぎるので、

有効数字は小数点第2位を四捨五入して小数点第1位とする。

四捨五入で小数点1位で求めるには、ROUND関数を利用  
`=ROUND(H6/(G6/100)^2, 1)`

ROUND関数  
 ROUND(数値, 桁数)で、指定の小数点以下の桁数になるよう四捨五入

#### 有効数字の考え方

測定値には誤差が含まれるので、誤差を含む可能性がある位から下の位を近似して無視するのが有効数字です。身長、体重の測定精度は小数1位なので、BMIも小数1位で表すのが適切です。割り算のままではダメです。

22

### 演習3-5 BMIの判定区分(3区分)をデータに追加

#### BMIの判定区分を追加する

The screenshot shows the formula bar with `=ROUND(H6/(G6/100)^2,1)` and a portion of the data table with columns for '年齢階級', '身長', '体重', 'BMI', and 'BMI区分'.

3) I6をI7からI132までコピーする。

年齢階級	身長	体重	BMI	BMI区分
50	161.2	71.6	27.6	
60	153.4	50.5	21.5	
60	161.5	55.4	21.2	
50	165.9	56.7	20.6	
50	163.8	55	20.5	
60	149.4	50	22.4	
50	151.1	51.5	22.6	
50	168.8	63.9	22.4	
60	166.9	61.3	22	
60	173.2	83.6	27.9	
40	175	70	22.9	
50	154.4	53.9	22.6	
60	153.1	44.9	19.2	
70	150.9	59.9	26.3	
60	162.9	46.3	17.4	

4) I列の右に1列挿入し、BMI区分と変数名を入れる

23

### 演習3-5 BMIの判定区分(3区分)をデータに追加

#### BMIの判定区分を追加する

やせ = 1: 18.5未満    標準 = 2: 18.5以上、25未満  
 肥満 = 3: 25以上

BMIの値を3つに分けるには、論理関数のIFを利用

IF(論理式, 真の場合, 偽の場合)

指定された条件が TRUE (真) のとき真の場合を返し、FALSE (偽) のとき偽の場合を返します。

IF(論理式, 真の場合, IF(論理式, 真の場合, IF(論理式, 真の場合, 偽の場合))) と IF文の中にIF文があっても良い(最大7つまで)

BMIの値が18.5未満→1、18.5以上、25未満→2、25以上→3と指定するには、0より大きい場合→18.5未満なら1、(18.5以上)で25未満なら2、(25以上)なら3、0以下は何も入れない(ヌル)と書いてみよう【注】ヌルは”(半角)を2つ並べて表す→””

$$=IF(I6>0, IF(I6<18.5, 1, IF(I6<25, 2, 3)), ””)$$

24



### 演習3-5 BMIの判定区分(3区分)をデータに追加

#### BMIの判定区分を追加する

Excel formula: `=IF(I6>0,IF(I6<18.5,1,IF(I6<25.2,3)), "")`

性別	生年月日	年度末年月日	年度末年齢	年齢階級	身長	体重	BMI	BMI区分	TG	HDL-C
男	1957/8/16	2009/3/31	51	50	161.2	71.0	27.0	3	2	161
男	1953/10/13	2009/3/31	55	50	153.4	50.5	21.5	1	2	153
男	1949/1/22	2009/3/31	60	60	161.5	55.4	21.2	1	2	161
男	1956/3/7	2009/3/31	53	50	165.9	56.7	20.6	2	2	165
男	1948/8/19	2009/3/31	60	50	149.4	50	22.4	2	2	149
男	1951/1/17	2009/3/31	58	50	163.8	55	20.5	1	2	163
男	1952/6/20	2009/3/31	56	50	168.8	63.9	22.4	2	2	168
男	1945/7/13	2009/3/31	63	60	166.9	61.3	22	2	2	166
男	1940/3/3	2009/3/31	69	60	173.2	83.6	27.9	3	2	173
男	1952/2/9	2009/3/31	57	50	164.5	57.5	21.3	2	2	164
男	1950/9/26	2009/3/31	59	50	167.7	76.2	27.1	3	2	167

セルJ6にIF文の式を入れ、J7からJ132までコピー

#### 注意!!

有効数字表示にしていないと、判定が異なる場合があります

No.111は、身長151cm、体重57kgです。BMIをそのまま計算すると、24.998903となり、判定区分は「標準(2)」です。

しかし、誤差を考えると、24.998903は25.0と同等であり、その方が正しい判断です。したがって、No.111のBMIは有効数字で表現した25.0とし、判定は「肥満」となります。

25

### 演習3-6 50歳代で「肥満」または「やせ」の人を抽出

ある条件にあうケースを抽出する

→ オートフィルタが便利

1) データ → フィルタ  
→ オートフィルタ を選択

オートフィルタのボタン▼が表示される

▼をクリックすると並べ替え方法・抽出の方法、セルに含まれる内容等が表示される

26

### 演習3-6 50歳代で「肥満」または「やせ」の人を抽出

ある条件にあうケースを抽出する

→ オートフィルタが便利

2) 年齢階級の▼をクリックし、50歳台を表す「50」を選択

3) BMI区分の▼をクリックし、オプションを選択

50だけが表示される

27

### 演習3-6 50歳代で「肥満」または「やせ」の人を抽出

ある条件にあうケースを抽出する

→ オートフィルタが便利

4) 「やせ」は1、「肥満」は3なので、「1と等しい」OR「3と等しい」となるように適語を選択 → OK

BMI区分が1または3、年齢階級が50の人が表示される。

28

演習3-7 50歳代女性において、BMI判定区分が「肥満」と「やせ」の人の収縮期血圧平均値と標準偏差をそれぞれ求め、作図する。

1) 演習3-6の結果の表から、女性だけを抽出する。  
性別の▼をクリックし、2を選択

ID番号	性別	生年月日	年度生年月	年度生年月	年齢	身長	体重	BMI	BMI区分	TG
9	1	1953/10/13	2009/3/31	55	50	165.9	56.7	20.6	2	63
10	10	1956/9/7	2009/3/31	53	50	163.8	55	20.5	2	92
12	18	1951/1/17	2009/3/31	58	50	151.1	51.5	22.6	2	70
13	21	1952/6/20	2009/3/31	56	50	168.8	63.9	22.4	2	49
17	46	1952/2/9	2009/3/31	57	50	164.5	57.5	21.2	2	82
37	104	1955/4/19	2009/3/31	53	50	162	47	17.9	1	115
62	26	1950/9/26	2009/3/31	58	50	150	40	17.8	1	34
69	34	1958/4/18	2009/3/31	50	50	156.3	44	18	1	54
70	35	1951/4/18	2009/3/31	57	50	158.8	54.9	21.5	2	84
77	43	1954/1/4	2009/3/31	55	50	152.5	40.3	17.3	1	48
78	44	1950/4/6	2009/3/31	58	50	150	52.8	23.5	2	132
83	51	1950/6/17	2009/3/31	58	50	156.2	54.5	22.3	2	67
82	62	1950/3/31	2009/3/31	59	50	157	56	22.7	2	79
93	67	1956/3/12	2009/3/31	53	50	155.1	53	22	2	153
99	75	1952/1/27	2009/3/31	57	50	155.1	48	20	2	58
100	77	1957/4/10	2009/3/31	51	50	154	44	18.6	2	160
101	82	1954/5/21	2009/3/31	54	50	148.5	44	20	2	93
112	96	1949/1/9	2009/3/31	59	50	147.5	52.5	24.1	2	82
116	102	1949/6/1	2009/3/31	59	50	155.8	42.7	17.6	1	48
117	103	1957/5/10	2009/3/31	51	50	155.5	53	21.9	2	99
118	105	1949/4/8	2009/3/31	59	50	151.8	44.8	19.4	2	74
125	117	1957/9/24	2009/3/31	51	50	157	56	22.7	2	132
126	118	1956/6/22	2009/3/31	52	50	155	51	21.2	2	129
127	120	1952/2/4	2009/3/31	57	50	154.2	45	19.9	2	72
128	121	1954/1/28	2009/3/31	55	50	152	56	24.2	2	104
128	123	1958/10/1	2009/3/31	50	50	154.4	53.9	22.6	2	72

演習3-7 50歳代女性において、BMI判定区分が「肥満」と「やせ」の人の収縮期血圧平均値と標準偏差をそれぞれ求め、作図する。

- 2) ID番号、性別、BMI区分、収縮期血圧、拡張期血圧の表示部分を選択  
→ コピー → 「sheet1」を開き、A25にカーソルを置く → 貼り付け

ID番号	性別	BMI区分	収縮期血圧	拡張期血圧
15	2	3	102	72
26	2	1	104	60
31	2	3	152	98
34	2	3	120	75
43	2	1	106	61
102	2	3	124	88
111	2	3	106	70
116	2	1	115	67
108	2	3	124	88
111	2	3	106	70

演習3-7 50歳代女性において、BMI判定区分が「肥満」と「やせ」の人の収縮期血圧平均値と標準偏差をそれぞれ求め、作図する。

2) SUBTOTAL関数を使って平均・標準偏差を求める  
(1) 収縮期血圧平均値  
=SUBTOTAL(1,D26:D33)

(2) 収縮期血圧の標本標準偏差を求めるSUBTOTAL関数を入力  
=SUBTOTAL(7,D26:D33)

1) 計算結果を記入する欄を作成する

演習3-7 50歳代女性において、BMI判定区分が「肥満」と「やせ」の人の収縮期血圧平均値と標準偏差をそれぞれ求め、作図する。

- 3) 「やせ」の血圧平均値、標準偏差を求める。  
オートフィルタのBMI区分で1を選択
- 4) 表記された計算値を「やせ」の欄に転記。  
5) オートフィルタのBMI区分の選択を3に変える。  
肥満者の血圧の平均・標準偏差が表示される。

ID番号	性別	BMI区分	収縮期血圧	拡張期血圧
25	2	3	102	72
26	2	1	104	60
27	2	3	152	98
28	2	3	120	75
29	2	1	106	61
31	2	3	124	88
32	2	3	106	70
33	2	3	106	70
35	2	1	115	67
36	2	1	111	61
37	2	3	124	88
38	2	3	124	88
39	2	3	106	70
40	2	3	106	70
41	2	3	106	70



演習3-7 50歳代女性において、BMI判定区分が「肥満」と「やせ」の人の収縮期血圧平均値と標準偏差をそれぞれ求め、作図する。

SUBTOTAL関数以外の方法

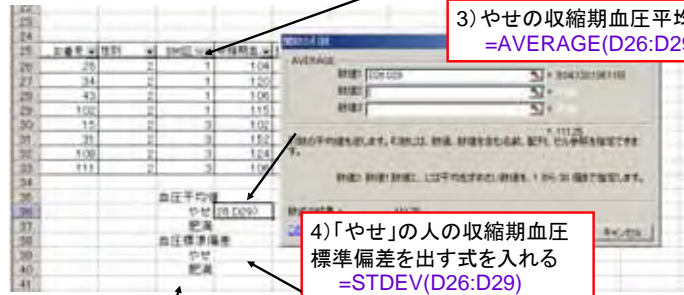
2) BMIで並べ替え

3) やせの収縮期血圧平均値  
=AVERAGE(D26:D29)

4) 「やせ」の人の収縮期血圧  
標準偏差を出す式を入れる  
=STDEV(D26:D29)

1) 計算結果を記入する欄を作成する

5) 「肥満」も同様に、収縮期血圧の平均値、標準偏差の関数式を入れる

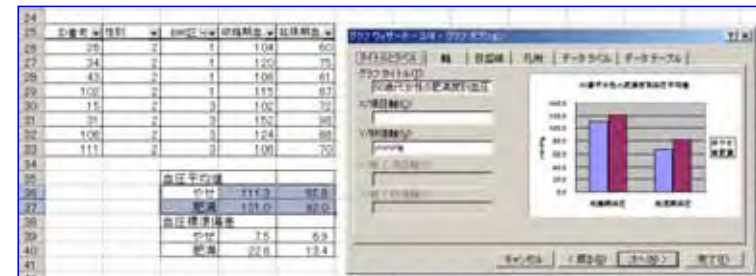


演習3-7 50歳代女性において、BMI判定区分が「肥満」と「やせ」の人の収縮期血圧平均値と標準偏差をそれぞれ求め、作図する。

「やせ」と「肥満」の収縮期血圧を比較するグラフを作成する

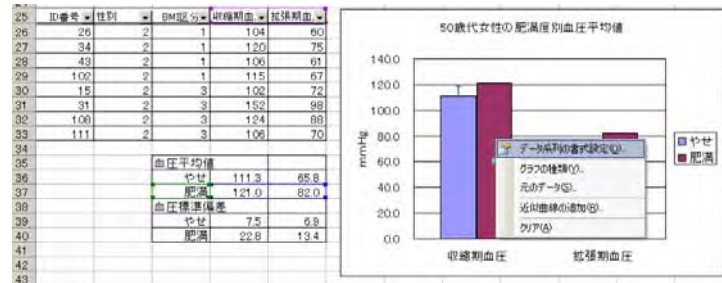
グラフタイトル  
50歳代女性の肥満度別血圧平均値  
Y/数値軸: mmHg

血圧平均値	やせ	111.3	65.8
肥満	121.0	82.0	
血圧標準偏差	やせ	7.5	6.9
肥満	22.8	13.4	



演習3-7 50歳代女性において、BMI判定区分が「肥満」と「やせ」の人の収縮期血圧平均値と標準偏差をそれぞれ求め、作図する。

出来上がった図を見やすくする

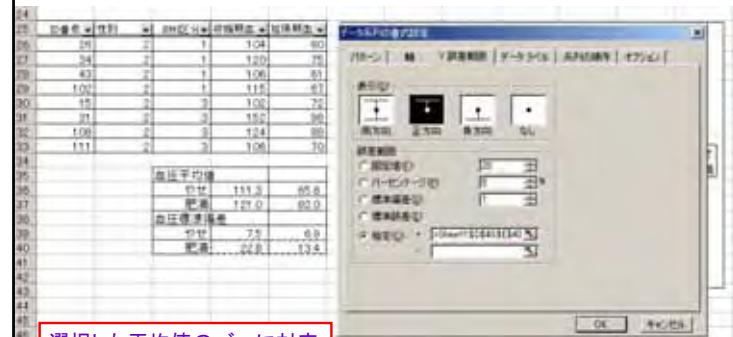


1) 背景を白くする  
プロットエリアの書式設定  
→ 輪郭「自動」、領域「なし」

2) 標準偏差を加える  
平均値のバーを選択  
→ 右クリック  
→ データ系列の書式の設定

演習3-7 50歳代女性において、BMI判定区分が「肥満」と「やせ」の人の収縮期血圧平均値と標準偏差をそれぞれ求め、作図する。

標準偏差の追加



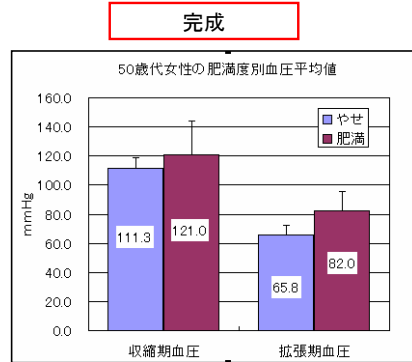
選択した平均値のバーに対応する標準偏差の範囲を指定  
→ OK

演習3-7 50歳代女性において、BMI判定区分が「肥満」と「やせ」の人の収縮期血圧平均値と標準偏差をそれぞれ求め、作図する。

3) 棒の間隔を変更する



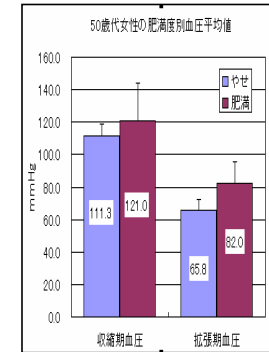
4) 数値を入れる



37

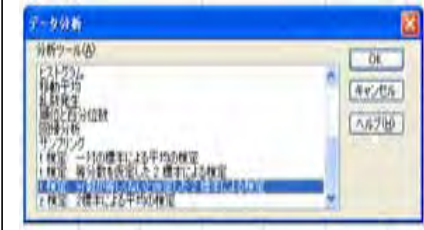
ちょっと進んで...

時間があれば、「やせ」と「肥満」の収縮期血圧に、統計的な差があるかを検討してみましょう。



平均値	収縮期	拡張期
やせ	111.3	65.8
肥満	121.0	82.0
標準偏差		
やせ	7.5	6.9
肥満	22.8	13.4

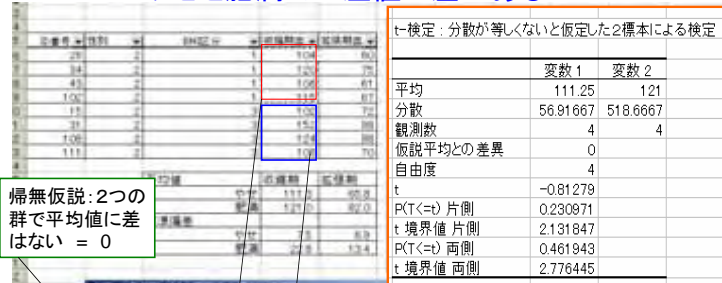
標準偏差の違いが大きいので、分散は等しくないと判断される



38

ちょっと進んで...

やせと肥満で血圧値に差があるか?



帰無仮説: 2つの群で平均値に差はない = 0

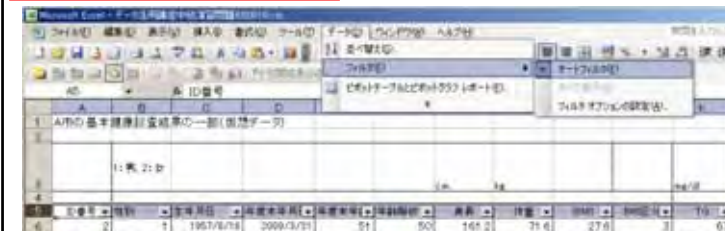
両側検定で  $p > 0.05$  なので、「有意差はない」

結論

演習3-8 性・年齢階級別のBMI判定区分の分布を作表する

ピボットテーブルを使ってクロス表を作成する

オートフィルターを解除する



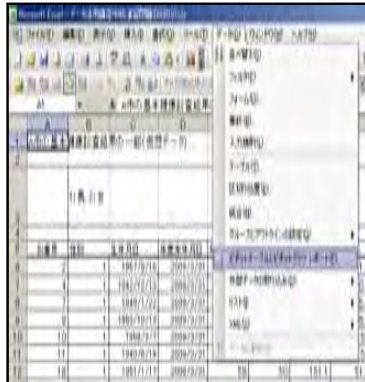
オートフィルタをクリック → 全てのデータが表示される

ID番号	性別	生年月日	年度生年月日	年度年齢	年齢階級	身長	体重	BMI	BMI区分	TG
5	2	1957/8/16	2009/3/31	51	50	161.2	71.6	27.6	3	
6	4	1942/12/13	2009/3/31	66	60	153.4	50.5	21.5	2	
7	1	1949/1/22	2009/3/31	60	60	161.5	55.4	21.2	2	

40

演習3-8 性・年齢階級別のBMI判定区分の分布を作表する  
ピボットテーブルを使ってクロス表を作成する

1)ピボットテーブルを開く



2)この状態で次へ



3)範囲を指定(A5:Z132)



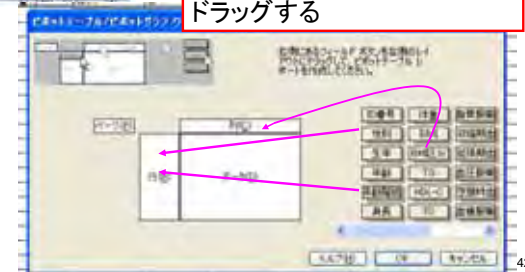
41

演習3-8 性・年齢階級別のBMI判定区分の分布を作表する  
ピボットテーブルを使ってクロス表を作成する

4) 新規ワークシートを選び、レイアウトをクリック



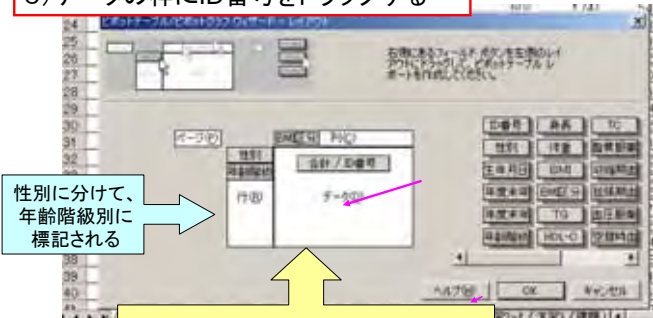
5)表の行、列に必要な項目をドラッグする



42

演習3-8 性・年齢階級別のBMI判定区分の分布を作表する  
ピボットテーブルを使ってクロス表を作成する

6)データの枠にID番号をドラッグする



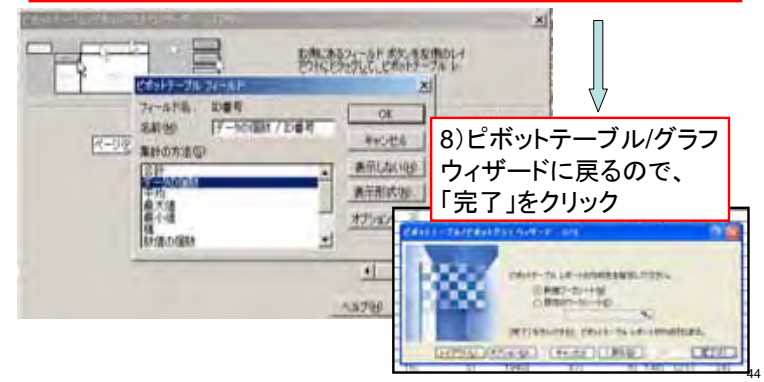
性別に分けて、年齢階級別に  
標記される

「合計」だと、該当するセルに入るID番号の合計値が標記される。  
今回は、合計ではなく、「データの個数」が必要なので、標記させる内容を変更する。

43

演習3-8 性・年齢階級別のBMI判定区分の分布を作表する  
ピボットテーブルを使ってクロス表を作成する

7)「合計/ID番号」をダブルクリックし、ピボットテーブルフィールドを表示させ、「データの個数」を選択 → OK



8)ピボットテーブル/グラフ  
ウィザードに戻るので、  
「完了」をクリック

44



演習3-7 性・年齢階級別のBMI判定区分の分布を作表する

ピボットテーブルを使ってクロス表を作成する

9) 新しいシートに下のような結果が表示される

データの個数 / ID番号	BMI区分				
性別	年齢階級	1	2	3	総計
1	40	6	1	7	
	50	1	5	2	8
	60	1	16	2	19
	70	5	1	6	
1 集計	2	32	6	40	
2	40	1	14	4	19
	50	4	16	4	24
	60	5	19	8	32
	70	1	6	5	12
2 集計	11	55	21	87	
総計	13	87	27	127	

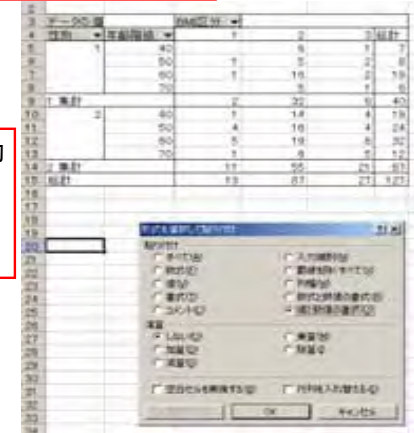
45

演習3-8 性・年齢階級別のBMI判定区分の分布を作表する

10) できたクロス表をコピーして、形を整える

①クロス表を選択し、  
編集 → コピー

②コピー先にカーソルを移動  
編集  
→ 形式を選択して貼り付け  
→ 値と数値の書式  
→ OK



46

演習3-8 性・年齢階級別のBMI判定区分の分布を作表する

ピボットテーブルを使ってクロス表を作成する

コピーした状態では、罫線等は表示されない

データの個数 / ID番号	BMI区分				
性別	年齢階級	1	2	3	総計
1	40	6	1	7	
	50	1	5	2	8
	60	1	16	2	19
	70	5	1	6	
1 集計	2	32	6	40	
2	40	1	14	4	19
	50	4	16	4	24
	60	5	19	8	32
	70	1	6	5	12
2 集計	11	55	21	87	
総計	13	87	27	127	

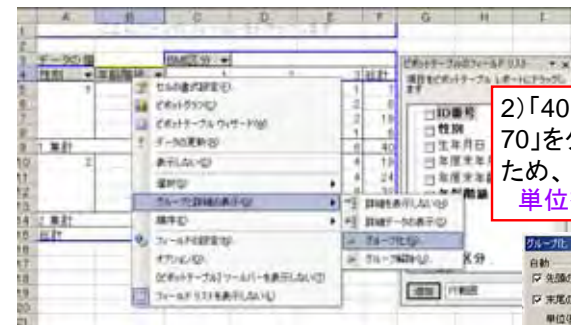
罫線、凡例などを記入して、形を整える。

データの個数 / ID番号	BMI区分	標準	肥満	総計
22 性別	年齢階級	やせ		
23 男性	40歳代	0	6	1
	50歳代	1	5	2
	60歳代	1	16	2
	70歳代	0	5	1
	合計	2	32	6
28 女性	40歳代	1	14	4
	50歳代	4	16	4
	60歳代	5	19	8
	70歳代	1	6	5
	合計	11	55	21
32 総計	合計	13	87	27

47

演習3-9 作成したクロス表を使い、「40～50歳代」「60歳代以上」の2群について、BMI判定区分が「肥満」または「やせ・標準」の性別の数を表示させる

1) 年齢階級を右クリックし、コマンドを表示する。  
グループと詳細の表示をクリック → グループ化を選択

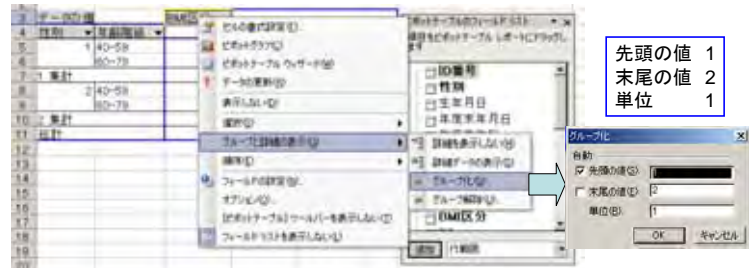


2) 「40～50」と「60～70」をグループ化するため、  
単位を20にし、OK

48

演習3-9 作成したクロス表を使い、「40～50歳代」「60歳代以上」の2群について、BMI判定区分が「肥満」または「やせ・標準」の性別の数を表示させる

3) 年齢が40～59と60～79の2群に変わる。  
BMI区分も同様に右クリックをしてグループ化「1と2」「2より大きい」の2カテゴリーにする。

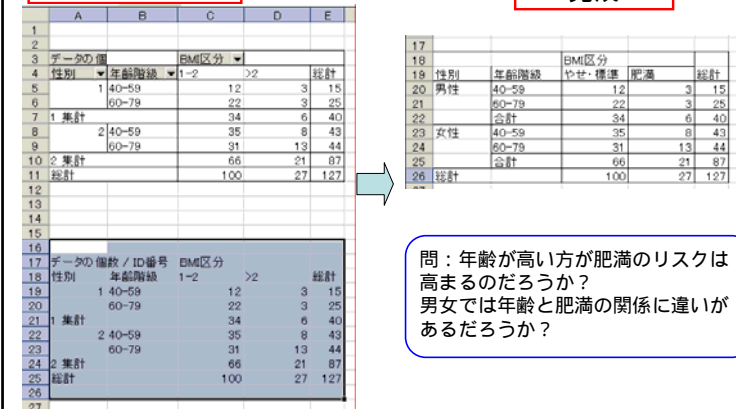


4) できたクロス表をコピーし、形を整える

演習3-9 作成したクロス表を使い、「40～59歳」「60～79歳」の2群について、BMI判定区分が「肥満」または「やせ・標準」の性別の数を表示させる

4) 形を整える

完成



問：年齢が高い方が肥満のリスクは高まるのだろうか？  
男女では年齢と肥満の関係に違いがあるだろうか？

演習3-9の応用 (オッズ比)

完成

	性別	年齢階級	BMI区分		総計
19	男性	40-59	やせ・標準	12	3 15
20	男性	60-79	肥満	22	3 25
21	合計			34	6 40
22	女性	40-59	やせ・標準	35	8 43
23	女性	60-79	肥満	31	13 44
24	合計			66	21 87
25	総計			100	27 127

オッズ比  
ある事象の起こりやすさを2つの群で比較して示す統計的な尺度

	症例	対照	
暴露あり	a	b	a+b
暴露なし	c	d	c+d

暴露ありの疾患オッズ=a/b  
暴露なしの疾患オッズ=c/d  
オッズ比(OR)=ad/bc  
95%信頼区間(+=上限値、-=下限値)  
=(OR)×exp(±1.96×SQRT(1/a+1/b+1/c+1/d))  
\* 95%信頼区間の上限と下限の値が1を挟まなければ統計的に有意な関係があると判断される。

問(余力があり、統計が分かる人向け)  
年齢が高い方が肥満のリスクは高いか？  
男女では年齢と肥満の関係に違いがあるか？  
同じ年代の場合、肥満リスクに男女差はあるか？  
オッズ比、95%信頼区間を求めて検討しなさい。

演習3-9の応用(カイ2乗検定)

	性別	年齢階級	BMI区分		総計
19	男性	40-59	やせ・標準	12	3 15
20	男性	60-79	肥満	22	3 25
21	合計			34	6 40
22	女性	40-59	やせ・標準	35	8 43
23	女性	60-79	肥満	31	13 44
24	合計			66	21 87
25	総計			100	27 127

カイ2乗検定  
ある事象の分布に群間差があるかを検討する統計的な尺度

測定値

	症例	対照	
暴露あり	a	b	a+b
暴露なし	c	d	c+d
	a+c	b+d	

カイ2乗検定のp値が0.05未満であれば有意差あり

期待値

	症例	対照	
暴露あり	(a+b)(a+c) (a+b+c+d)	(a+b)(b+d) (a+b+c+d)	a+b
暴露なし	(c+d)(a+c) (a+b+c+d)	(c+d)(b+d) (a+b+c+d)	c+d
	a+c	b+d	

問  
男女それぞれについて、肥満の分布は年齢差があるかを検討しなさい。

同年代の場合、肥満の分布に男女差はあるかを検討しなさい。

カイ2乗検定のp値=CHITEST(実測値の範囲,期待値の範囲)

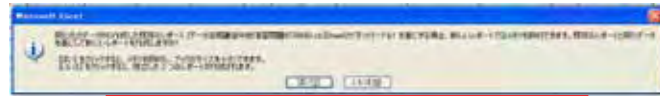
演習3-10 性・年齢・肥満の有無別、血圧、中性脂肪の記述統計  
(人数、平均、標準偏差、最大値、最小値)

ピボットテーブルを使い、層別に記述統計を作成する

1) シート演習3(2)に戻り、ピボット  
テーブルウィザードを開く → 次へ



2) 範囲が正しく指定されてい  
れば(A5:Z132) OK



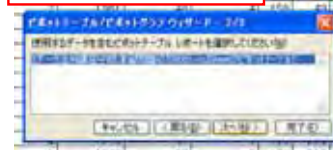
3) データに変更がないので、「はい」を選択

53

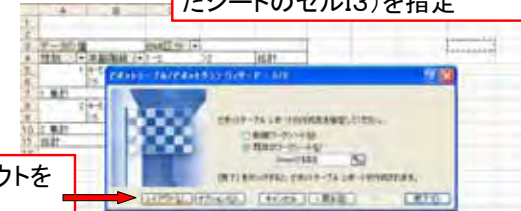
演習3-10 性・年齢・肥満の有無別、血圧、中性脂肪の記述統計  
(人数、平均、標準偏差、最大値、最小値)

ピボットテーブルを使い、層別に記述統計を作成する

4) このまま「次へ」



5) 既存のワークシートに出力さ  
せるため、出力場所(先に作っ  
たシートのセルI3)を指定



6) レイアウトを  
クリック

54

演習3-10 性・年齢・肥満の有無別、血圧、中性脂肪の記述統計  
(人数、平均、標準偏差、最大値、最小値)を求める。

ピボットテーブルを使い、層別に記述統計を作成する

7) 行に性別、年齢階級、列にBMI区  
分、データフィールドに、収縮期血圧  
をドラッグ



8) 収縮期血圧の集計内容を  
確認する。  
① 合計をデータの個数に変  
更 → ダブルクリックして、  
データの個数に変更  
② 平均値の集計用に収縮期  
血圧を再度ドラッグ  
合計を同様にして平均に変更  
以下、同様にし、  
③ 標本標準偏差  
④ 最大値  
⑤ 最小値  
の集計方法を選ぶ。

55

演習3-10 性・肥満の有無別に、血圧、中性脂肪の記述統計  
(人数、平均、標準偏差、最大値、最小値)を求める。

ピボットテーブルを使い、層別に記述統計を作成する

9) 収縮期血圧の指定が終わったら、OK



10) 出力先指定画面に  
戻るので、完了

56



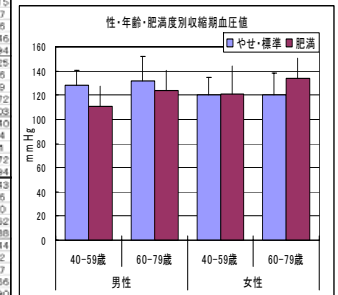
演習3-10 性・年齢・肥満の有無別に、血圧、中性脂肪の記述統計(人数、平均、標準偏差、最大値、最小値)を求める。

12)コピーをして、形を整える

性別	年齢階級	データ	BMI区分	統計		
男性	40-59	データの個数 / 収縮期血圧	1-2	12	3	
		平均 / 収縮期血圧		128.1666667	111	
		標準標準偏差 / 収縮期血圧		12.59749855	16.00347694	
		最大値 / 収縮期血圧		146	146	
	60-79	データの個数 / 収縮期血圧		22	3	
		平均 / 収縮期血圧		131.5	124	
		標準標準偏差 / 収縮期血圧		20.47238721	16.37070554	
		最大値 / 収縮期血圧		172	172	
	女性	40-59	データの個数 / 収縮期血圧		34	6
			平均 / 収縮期血圧		130.3235294	117.5
			標準標準偏差 / 収縮期血圧		17.96064038	16.17096163
			最大値 / 収縮期血圧		172	172
60-79		データの個数 / 収縮期血圧		109	94	
		平均 / 収縮期血圧		120.405714	121.125	
		標準標準偏差 / 収縮期血圧		14.32251562	23.2713036	
		最大値 / 収縮期血圧		162	152	

演習3-10 性・年齢・肥満の有無別、血圧、中性脂肪の記述統計(人数、平均、標準偏差、最大値、最小値)

結果を図にすると・・・



どうやったら、このような、図が書けるだろうか？各自で書いてみよう。

演習3-10 性・年齢・肥満の有無別、血圧、中性脂肪の記述統計(人数、平均、標準偏差、最大値、最小値)を求める。

一歩進んだ学習・・・作成したテーブルを利用して、別の表を作る

ドラッグして、戻す

全体についての、肥満有無別の表ができる

演習3-10 性・年齢・肥満の有無別、血圧、中性脂肪の記述統計(人数、平均、標準偏差、最大値、最小値)を求める。

一歩進んだ学習・・・作成したテーブルを利用して、別の表を作る

中性脂肪(TG)をフィールドにドラッグ

同様にTGをフィールドにドラッグし、平均、標準標準偏差、最大値、最小値の集計方法を設定する

合計/TGを右クリックして集計方法を「合計」から「データの個数」に変更

演習3-10 性・年齢・肥満の有無別、血圧、中性脂肪の記述統計  
(人数、平均、標準偏差、最大値、最小値)を求める。

中性脂肪(TG)  
の集計結果の  
完成

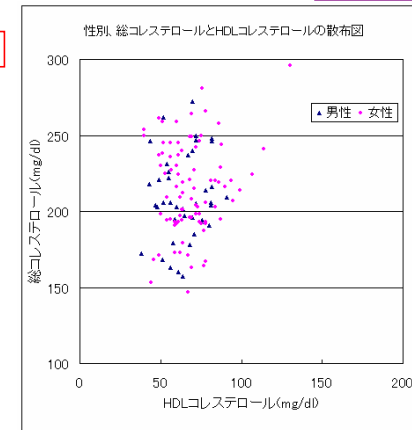
性別	年齢階級	データ	EM区分	統計	
1	140-59	データの個数 / TG	12	3	15
		平均 / TG2	98	93.66666667	97.13933333
		標準標準偏差 / TG3	38.21708328	40.27819923	37.18268926
		最大値 / TG	175	140	175
		最小値 / TG2	49	67	49
	60-79	データの個数 / TG	22	3	25
		平均 / TG2	106.3181818	398	119.52
		標準標準偏差 / TG3	59.20616632	19.28730152	65.08079594
		最大値 / TG	250	222	250
		最小値 / TG2	42	106	42
1 データの個数 / TG		34	6	40	
1 平均 / TG2		103.9823529	150.8939393	110.5	
1 標準標準偏差 / TG3		52.28593571	68.69764673	56.68084287	
1 最大値 / TG		250	222	250	
1 最小値 / TG2		42	67	42	
2	140-59	データの個数 / TG	35	8	43
		平均 / TG2	90.4	126.125	97.04651163
		標準標準偏差 / TG3	35.41452012	86.27934283	49.53640003
		最大値 / TG	177	276	276
		最小値 / TG2	34	32	32
	60-79	データの個数 / TG	31	13	44
		平均 / TG2	89.09677419	114	96.45454545
		標準標準偏差 / TG3	44.00784388	54.7813837	48.17432229
		最大値 / TG	219	259	259
		最小値 / TG2	44	61	44
2 データの個数 / TG		66	21	87	
2 平均 / TG2		89.78787879	118.6190476	96.74712644	
2 標準標準偏差 / TG3		39.37414816	66.65168879	48.5681536	

61

演習3-11 総コレステロールとHDLコレステロール値の関係をみるため、性別に散布図を作成し、相関関係があるかを考える

散布図の作成

このような図を作成する



62

演習3-11 総コレステロールとHDLコレステロール値の関係をみるため、性別に散布図を作成する

散布図の作成

- 1) 新しいワークシートを挿入し、ID番号、性別、総コレステロール、HDLコレステロールの値をコピーする。
- 2) データを性別に並び替える  
データ → 並べ替え → 最優先されるキー(性別) → OK

	A	B	C	D	E	F
1	ID番号	性別	HDL-C	TG		
2	1	2	60	230		
3	2					
4	3					
5	4					
6	5					
7	6					
8	7					
9	8					
10	9					
11	10					
12	11					
13	12					
14	13					
15	14					
16	15					
17	16					
18	17	2	78	193		

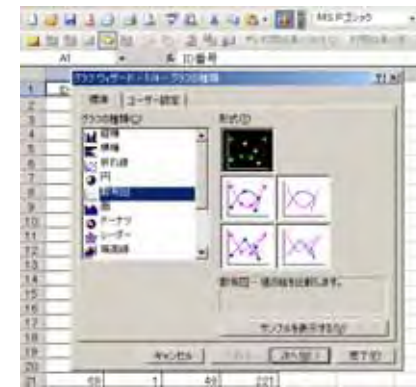
63

演習3-11 総コレステロールとHDLコレステロール値の関係をみるため、性別に散布図を作成する

散布図の作成

- 3) ツールバーのグラフをクリックし、グラフウィザードを立ち上げる

散布図を選択

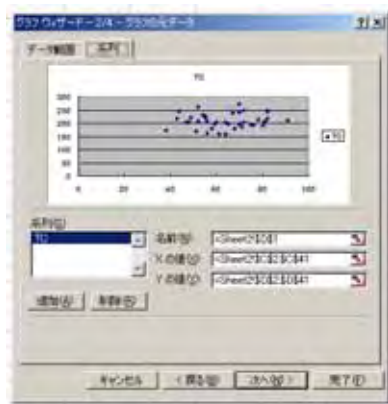


64

演習3-11 総コレステロールとHDLコレステロール値の関係を見るため、性別に散布図を作成する

散布図の作成

3) 男性のデータ範囲を指定する

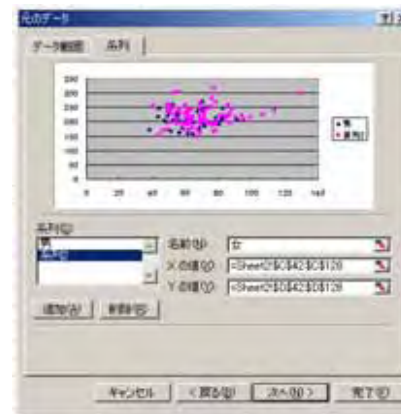


- ①Yの値に男性の総コレステロールのデータ範囲を指定
- ②Xの値に男性のHDLコレステロールのデータ範囲を指定
- ③名前に男性と入れ、追加をクリック

演習3-11 総コレステロールとHDLコレステロール値の関係を見るため、性別に散布図を作成する

散布図の作成

4) 追加をクリックした後、女性のデータ範囲を指定する

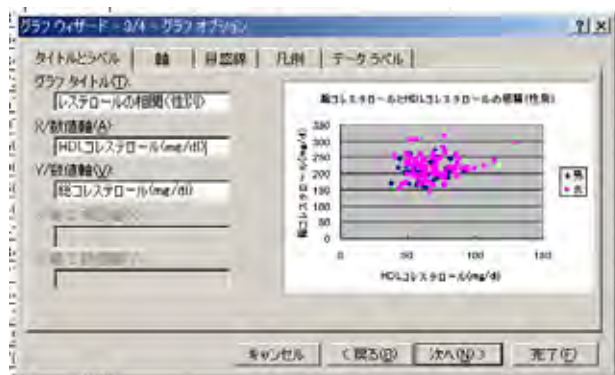


- ①Yの値に女性の総コレステロールのデータ範囲を指定
- ②Xの値に女性のHDLコレステロールのデータ範囲を指定
- ③名前に女性と入れ、次へをクリック

演習3-11 総コレステロールとHDLコレステロール値の関係を見るため、性別に散布図を作成する

4) グラフのタイトル、軸の名前等を入力 → 次へ → 完了

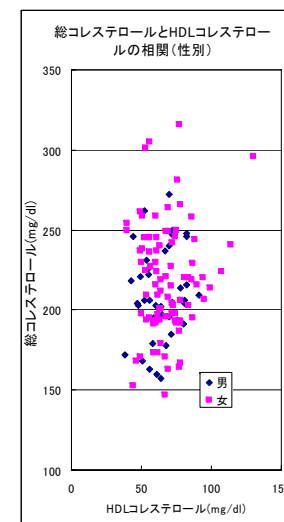
性別、総コレステロールとHDLコレステロールの散布図



演習3-11 総コレステロールとHDLコレステロール値の関係を見るため、性別に散布図を作成する

5) グラフのスタイルを整える

散布図では、X軸、Y軸の目盛間隔が同じになるように調整する





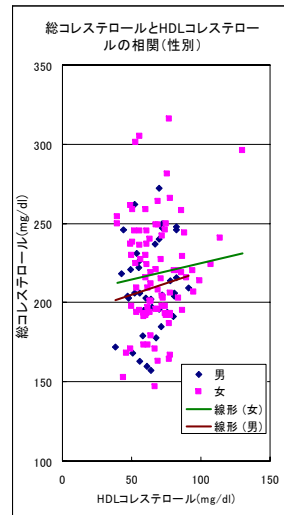
演習3-11 総コレステロールとHDLコレステロール値の関係を見るため、性別に散布図を作成する

5) グラフのスタイルを整える

プロットしているマークの形・色・大きさを調整し、散布の状況が分かるようにする

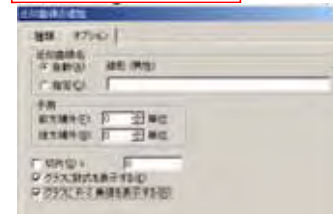
マークを右クリックし、データ系列の書式設定を選んでマークの形、色、大きさを変更

近似曲線の追加を選択し、近似曲線と決定係数(R2乗)を表示させる



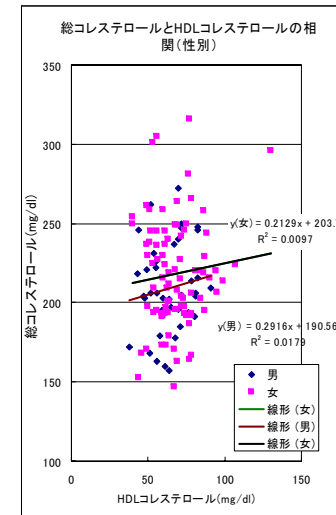
演習3-11 総コレステロールとHDLコレステロール値の関係を見るため、性別に散布図を作成する

6) 近似曲線を追加する



近似曲線の追加を選択し、近似曲線の種類を選び(線形)、オプション画面で数式の表示とR-2乗値の表示にチェックを入れる

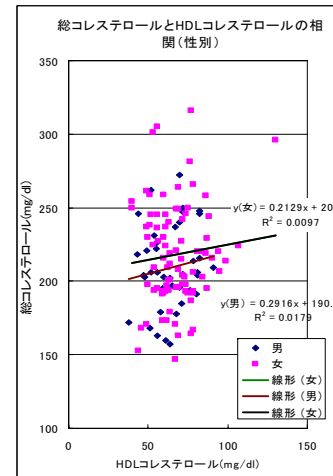
近似曲線式の位置等を調整し、見やすい図を完成させる



課題

- 1) 演習3のシートの方には、各検査値を判定基準にしたがってカテゴリ化した結果を入れる列があります。それぞれの基準に従って、測定値をカテゴリ化し、結果を記入してください。
- 2) 性・年齢階級別に、各測定値をカテゴリ化した結果が何人ずつになるか、ピボットテーブルを使って表を作ってください。
- 3) オートフィルタを使って、BMIが25以上、収縮期血圧が130mmHg以上、中性脂肪150mg/dl以上、空腹時血糖100mg/dl以上の人を抽出してみましょう。

演習3-11 総コレステロールとHDLコレステロール値の関係を見るため、性別に散布図を作成する



R-2乗は、相関係数の2乗値  
この大きさは、XとYの値の関係の強さを示す。  
相関係数の値と関係の強さの目安は、  
相関係数(r)の絶対値が  
 $|r| < 0.3$  ..無相関  
 $0.3 \leq |r| < 0.7$  ..弱い相関  
 $0.7 \leq |r|$  ..強い相関

問: 男女それぞれについて、総コレステロール値とHDLコレステロール値には相関関係があると言えるだろうか?