

# 教員免許状更新講習

## 学校保健と統計学

福山平成大学 福井正康

## 利用するデータについて（免許講習データ.txt）

- あなたの性別は。
  - 男性
  - 女性
- あなたの学年は。
  - 1年
  - 2年
  - 3年
- あなたの気になっていることについて、いくつでも○を付けて下さい。
  - 恋愛のこと
  - 勉強のこと
  - 友人のこと
- 学力テスト成績
- 心理テスト成績（値が高いほど不安がないとする）

## 1. データの集計

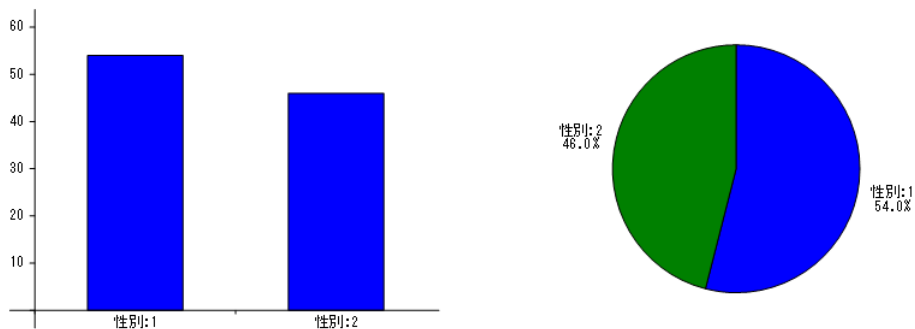
### 1.1 質的データの集計

#### 問題

- 性別に関する1次元分割表を描け。

男性	女性	合計
54	46	100

- 性別に関する棒グラフと円グラフを描け。



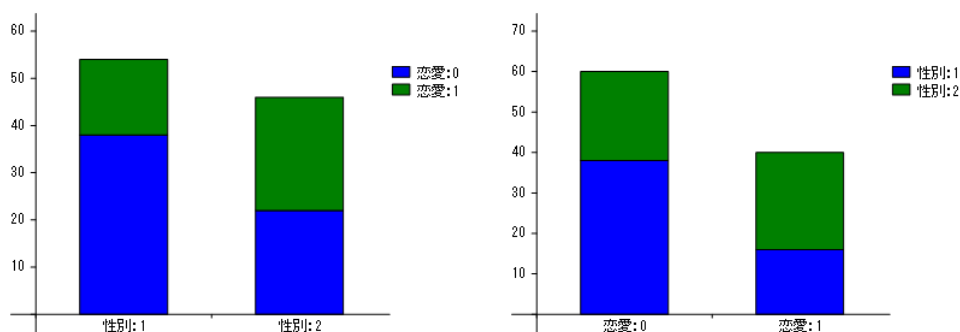
- 学年に関する1次元分割表を描け。

1年	2年	3年	合計
26	36	38	100

- 性別と恋愛（気にならない／気になる）に関する2次元分割表を描け。

	ならない	なる	合計
男性	38	16	54
女性	22	24	46
合計	60	40	100

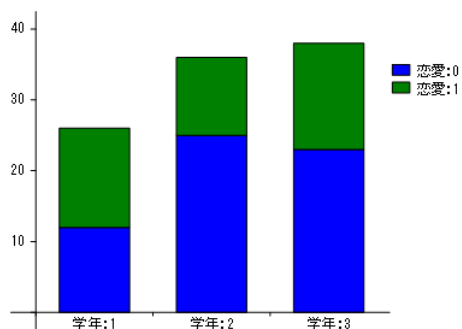
5) 性別と恋愛に関する積み重ね棒グラフを描け。



6) 学年と恋愛（気にならない／気になる）に関する2次元分割表を描け。

	ならない	なる	合計
1 年	12	14	26
2 年	25	11	36
3 年	23	15	38
合計	60	40	100

7) 学年と恋愛に関する積み重ね棒グラフを描け。



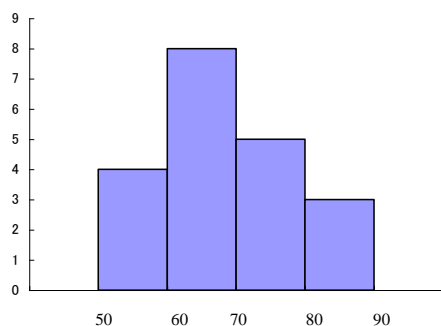
## 1.2 量的データの集計

### 度数分布表

階級	度数	相対度数 (%)	累積度数	累積相対度数 (%)
$50 \leq x < 60$	4	20	4	20
$60 \leq x < 70$	8	40	12	60
$70 \leq x < 80$	5	25	17	85
$80 \leq x < 90$	3	15	20	100
計	20	100		

注) 各階級の幅を階級幅、各階級の中央の値を階級値という。

## ヒストグラム



基本統計量（要約統計量） 【データ 3, 3, 4, 2, 8】

分布の中心を表わす統計量（代表値）

平均値（average, mean）

$$\text{平均値} = \frac{1}{5}(3+3+4+2+8) = 4$$

中央値（中間値, メジアン median）

データを小さい方から順番に並べて中間の値

$$2, 3, 3, 4, 8 \quad \rightarrow \quad 3$$

$$2, 3, 3, 4, 6, 8 \quad \rightarrow \quad (3+4)/2=3.5$$

最頻値（モード mode）

度数分布表やヒストグラムでまとめられている場合は、最大度数の階級値  
分布の広がりを表わす統計量（散布度）

レンジ（range）

$$R = \text{最大値} - \text{最小値} = 6$$

分散（variance）

$$s^2 = \frac{1}{5}[(3-4)^2 + (3-4)^2 + (4-4)^2 + (2-4)^2 + (8-4)^2] = 4.4$$

標準偏差（standard deviation）

$$s = \sqrt{\text{分散}} = 2.098$$

不偏分散

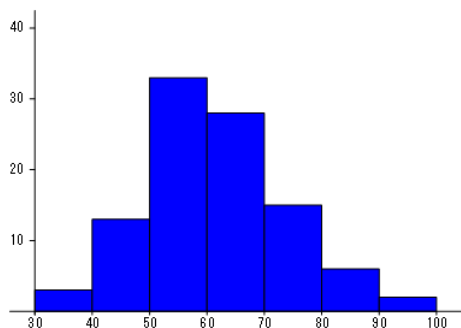
$$u^2 = \frac{1}{5-1}[(3-4)^2 + (3-4)^2 + (4-4)^2 + (2-4)^2 + (8-4)^2] = 5.5$$

標準偏差（standard deviation）

$$u = \sqrt{\text{不偏分散}} = 2.345$$

## 問題

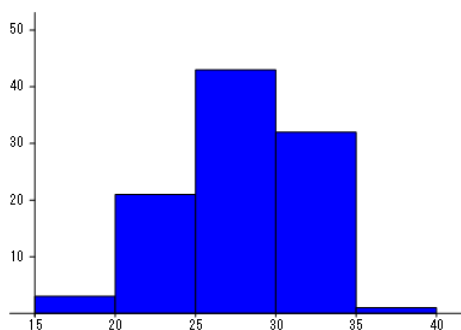
- 1) 学力テストについてヒストグラムを描け。



- 2) 学力テストについて基本統計量を求めよ。

平均値	中央値	標準偏差
61.030	60.500	12.081

- 3) 心理テストについてヒストグラムを描け。



- 4) 心理テストについて基本統計量を求めよ。

平均値	中央値	標準偏差
27.660	28.000	3.753

- 5) 男女別の学力テストについて基本統計量を求めよ。

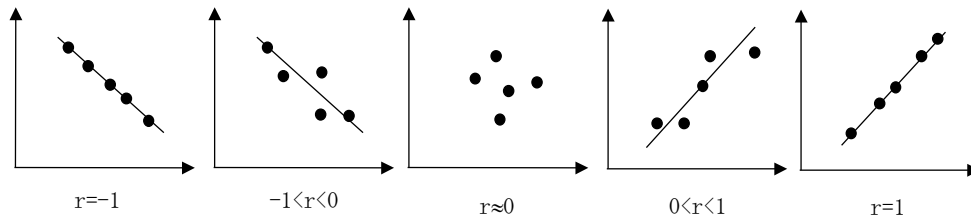
	平均値	中央値	標準偏差
男性	59.407	58.500	12.808
女性	62.935	64.000	11.000

- 6) 男女別の心理テストについて基本統計量を求めよ。

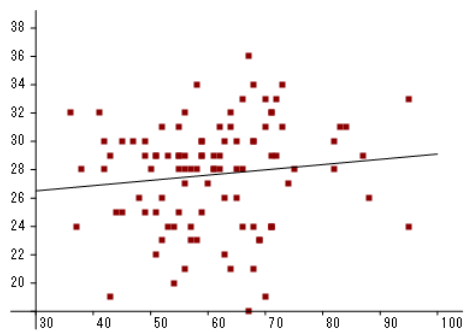
	平均値	中央値	標準偏差
男性	27.130	28.000	3.127
女性	28.283	29.500	4.329

### 1.3 相関

散布図（分布図，相関図）



1) 心理テスト（縦軸）と学力テストの散布図を描け。



2) 心理テストと学力テストの相関係数を求めよ。

相関係数    [ 0.119 ]

## 2. 検定

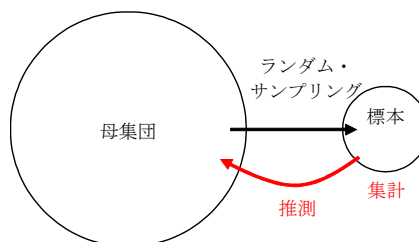
### 2.1 検定とは

#### 母集団と標本

母集団：調査の対象，日本人・日本の中小企業等  
(全数調査不可能な場合がある)

標本： 偏りがないように選抜（ランダムサンプリング）された実際に調査する対象

母集団の全数調査が不可能な場合、標本をとって母集団を推測する。



#### どんな検定があるか

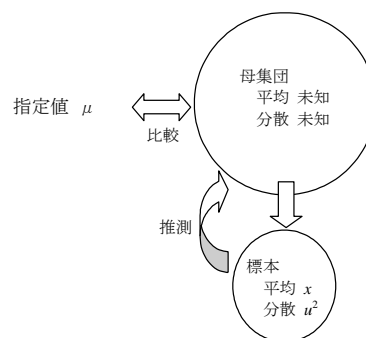
##### 1) 指定値と母集団のある指標を比較する。

質的データの比較：

標本調査の結果（割合）と期待される結果  
(割合) との比較

量的データの比較：

標本調査世帯と全国平均との所得の比較



##### 2) いくつかの母集団のある指標を比較する。

質的データの比較：

男女間での意識調査の結果（割合）の比較  
(対応がない場合)

標本店舗における従業員教育前後の評判の  
変化 (対応がある場合)

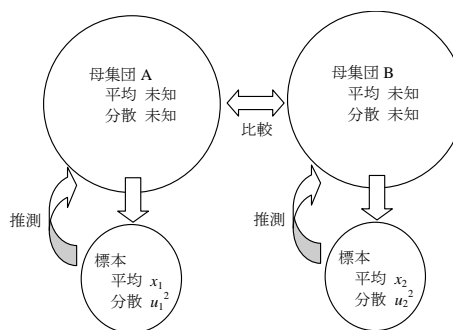
量的データの比較：

2つの標本調査世帯の所得の比較

(対応がない場合)

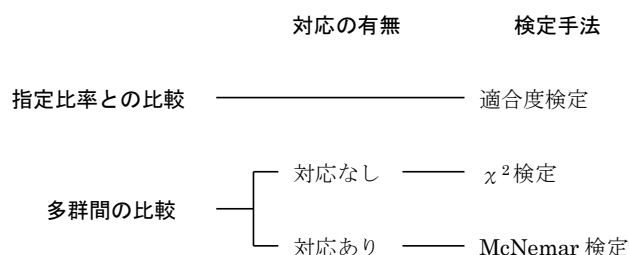
標本店舗における宣伝前後の売り上げ比較

(対応がある場合)



## 2.2 質的データの検定

### 検定選択ツリー



### 群間の比率の比較検定

- 1) 性別と恋愛（気にならない／気になる）に関する 2 次元分割表を描け。

	ならない	なる	合計
男性	38	16	54
女性	22	24	46
合計	60	40	100

- 2) 恋愛に関する心配事は男女による差があるといえるか。有意水準 5% で判定せよ。

検定名 [  $\chi^2$  検定 ]

検定確率 [ 0.0367 ]      差があると [ いえる ] ・ いえない

有意水準と呼ばれる基準に対して、検定確率が小さければ、差があると判定し、  
検定確率が大きければ、差がないと判定する。

- 3) 恋愛に関する心配事は学年による差があるといえるか。有意水準 5% で判定せよ。

検定名 [  $\chi^2$  検定 ]

検定確率 [ 0.2832 ]      差があると [ いえる ] ・ いえない

- 4) 勉強に関する心配事は男女による差があるといえるか。有意水準 5% で判定せよ。

検定名 [  $\chi^2$  検定 ]

検定確率 [ 0.6764 ]      差があると [ いえる ] ・ いえない

- 5) 勉強に関する心配事は学年による差があるといえるか。有意水準 5% で判定せよ。

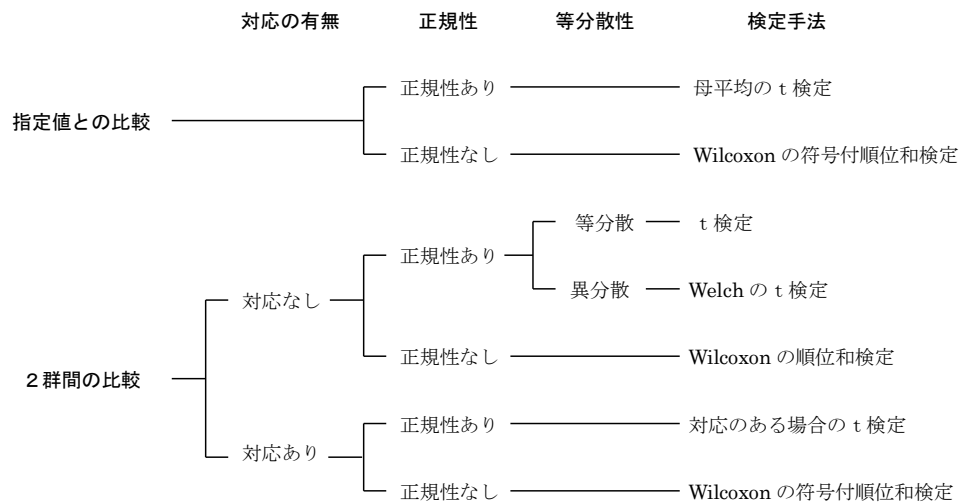
検定名 [  $\chi^2$  検定 ]

検定確率 [ 0.0003 ]      差があると [ いえる ] ・ いえない



## 2.3 量的データの検定

### 検定選択ツリー



### 2 群間の量的データの差の比較検定

1) 男女間で学力テストに差があるといえるか。有意水準 5% で判定せよ。

検定名 [ **t 検定** ]

検定確率 [ **0.1465** ]      差があると [ いえる・ ]

2) 男女間で心理テストに差があるといえるか。有意水準 5% で判定せよ。

検定名 [ **Wilcoxon 順位和検定** ]

検定確率 [ **0.0908** ]      差があると [ いえる・ ]

3) 恋愛が気になるかどうかで学力テストに差があるといえるか。有意水準 5% で判定せよ。

検定名 [ **t 検定** ]

検定確率 [ **0.2758** ]      差があると [ いえる・ ]

4) 恋愛が気になるかどうかで心理テストに差があるといえるか。有意水準 5% で判定せよ。

検定名 [ **t 検定** ]

検定確率 [ **0.2887** ]      差があると [ いえる・ ]

5) 勉強が気になるかどうかで学力テストに差があるといえるか。有意水準 5% で判定せよ。

検定名 [ **Wilcoxon 順位和検定** ]

検定確率 [ **0.0000** ]      差があると [  ・いえない ]

6) 勉強が気になるかどうかで心理テストに差があるといえるか。有意水準 5% で判定せよ。

検定名 [ **Wilcoxon 順位和検定** ]

検定確率 [ **0.0160** ]      差があると [  ・いえない ]

## 2.4 集計・検定結果の表示

- ・ 集計値の桁数は、平均・標準偏差等でデータ桁数より 1 桁か 2 桁程度多く表示する。  
例：171, 173, 174, … → 平均 172.7
- ・ 検定の際、検定統計量の値や自由度などはあまり書かれることがないが、検定手法の名前は書く場合もある（書かない場合もある）。
- ・ 検定確率値については、あまり具体的な数値を書くことはなく、n.s.,  $p<0.05$ ,  $p<0.01$ ,  $p<0.001$  (n.s., \*, \*\*, \*\*\* にしてもよい) のどれかの書き方にする。

### 補遺 プログラムの自宅パソコンでの実行

持ち帰られたプログラムを自宅パソコン（Windows の動くもの）で実行するときの注意を書いておきます。CAnalysis.exe をダブルクリックして動かす際、おそらく最初に左図のようなメッセージが出ます。ここで、「実行しない」を選ばずに、「詳細情報」をクリックします。すると右図のような表示に変わります。ここで「実行」をクリックすると C.Analysis が起動します。一度起動すると後は問題なく起動します。ウィルスソフトが実行を妨げる場合は、HP などを書いてある方法を試して下さい。

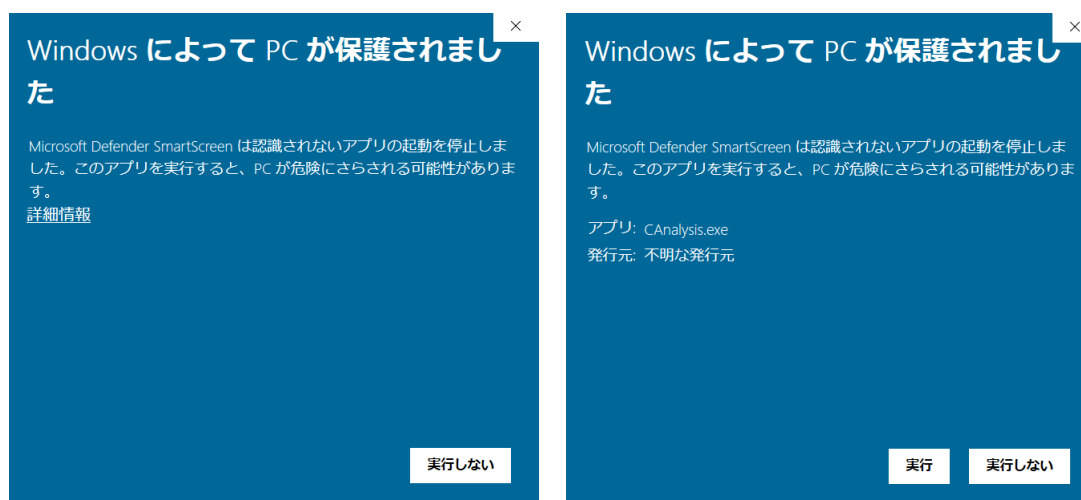


図 1 警告メッセージ