

統計学 第3回

「データの尺度・データの図示」

中澤 港(なかざわ みなと)
内線 1453 , E-mail: minato@ypu.jp

尺度

- 尺度とは？
→データに何らかの値を対応させる基準である。
- 尺度は，
 - (1) 値の差に意味があるかどうか，
 - (2) 値の順序に意味があるかどうか，
 - (3) ゼロに意味があるかどうか，によって，名義尺度，順序尺度，間隔尺度，比尺度の4つに分類される。

尺度と変数

- 研究対象にとりあげる操作概念は変数という形をとる。変数は固有の尺度をもつ。
- 同じ生データでも，コーディングのやり方次第で変数の尺度は変わりうる。
- 対応する尺度の種類によって，変数は，
 - 図示の仕方も違うし，
 - 代表値も違うし，
 - 適用できる統計解析手法も違う。

名義尺度

- 値の差は意味をもたない
- 値の順序は意味をもたない
- 質的データに分類基準を与える
- 性別, 職業, 居住地など

順序尺度

- 値の差は意味をもたない
- 値の順序は意味をもつ
- 鉱物の強度, 地震の震度, 尿検査のタンパク検出, 「好き」「普通」「嫌い」, 出生順位など。
- もっともらしい仮定を導入して間隔尺度とみなすこともある

間隔尺度

- 値の差は意味をもつ
- 値の順序も意味をもつ
- ゼロに意味がない(=値の比は意味をもたない)
- 摂氏温度, 西暦年など

比尺度

- 値の差は意味をもつ
- 値の順序も意味をもつ
- ゼロに意味がある(=値の比は意味をもつ)
- 身長, 体重, 年齢など
- 大雑把に言えば, 客観的に測定可能な変数の多くは比尺度をもつ

データの図示

- 離散量の場合
 - 度数分布図
 - 積み上げ棒グラフ
 - 帯グラフ
 - 円グラフ(ドーナツグラフ)
- 連続量の場合
 - ヒストグラム
 - レーダーチャート
 - 幹葉表示 (stem and leaf plot)
 - 箱ヒゲ図 (box and whisker plot)
 - 散布図 (scatter plot)

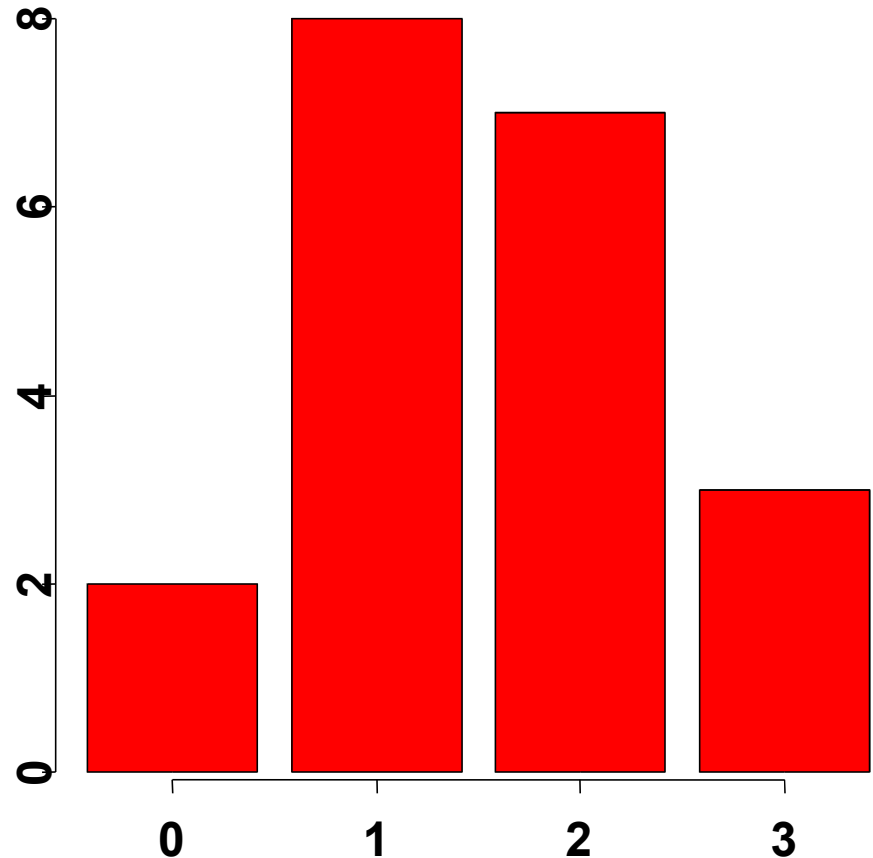
離散変数の図示の例1

- 夫婦の子ども数: 20組の夫婦について
2, 3, 1, 0, 3, 2, 2, 1, 1, 1, 2, 2, 1, 3, 2, 1, 0, 2, 1, 1
だった場合, どうまとめるか?

度数分布表としては,

子ども数	夫婦数
0	2
1	8
2	7
3	3
合計	20

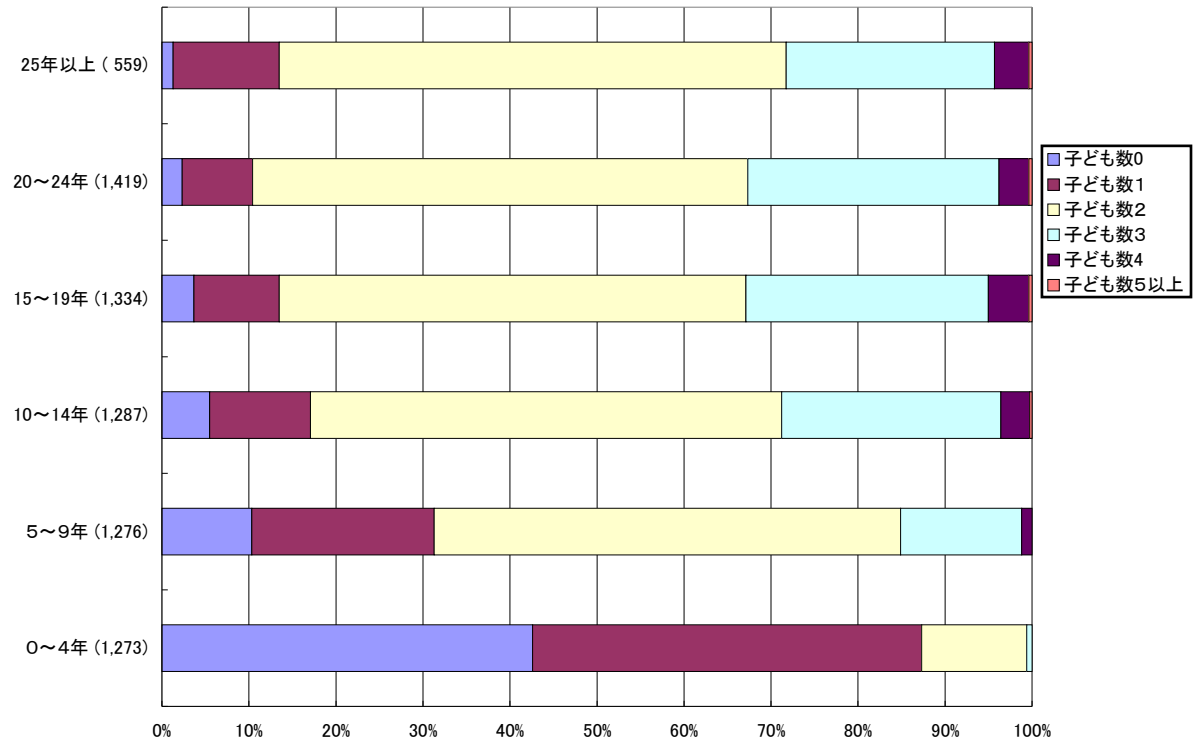
図示するとわかりやすい。



離散変数の図示の例2

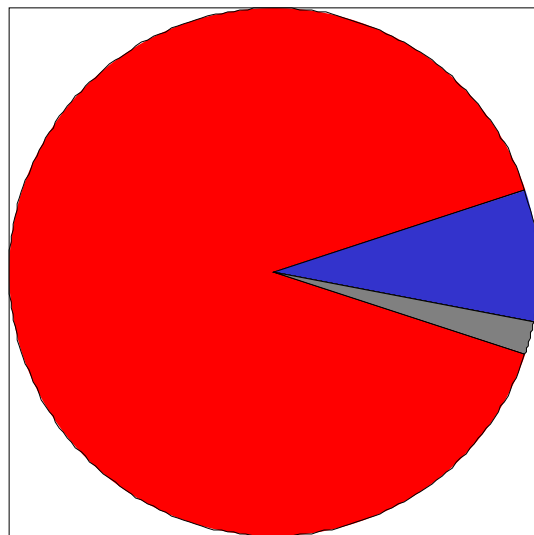
結婚持続期間	0人	1人	2人	3人	4人	5人以上
0～4年 (1, 273)	42.6	44.7	12.1	0.6	-	-
5～9年 (1, 276)	10.3	21.0	53.6	13.9	1.2	-
10～14年 (1, 287)	5.5	11.6	54.2	25.2	3.3	0.3
15～19年 (1, 334)	3.7	9.8	53.6	27.9	4.6	0.4
20～24年 (1, 419)	2.3	8.1	57.0	28.9	3.4	0.4
25年以上 (559)	1.3	12.2	58.3	24.0	3.9	0.4

国立社会保障
・人口問題研
究所が行った
「第11回出生
動向基本調査
・夫婦調査」か
ら、結婚持続
期間と子ども
数の関係

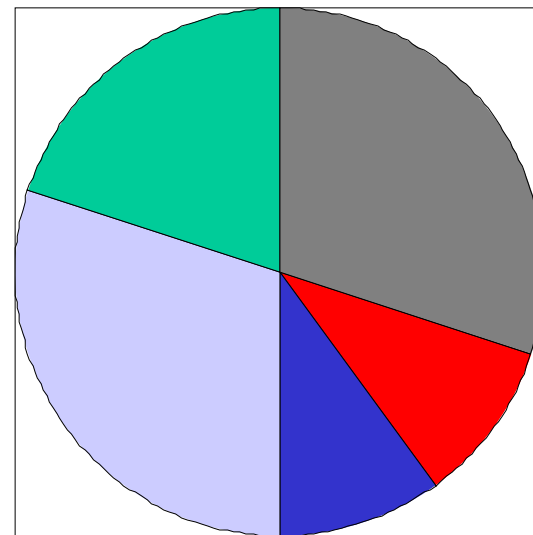
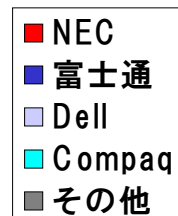


離散変数の図示の例3

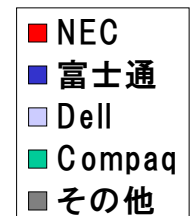
- マーケットシェアなどは、円グラフにするとわかりやすい。日本のパソコン市場におけるメーカー別売上高の1987年と2000年の比較をしたら、下図のようになる(数値は不正確)



1987年



2000年

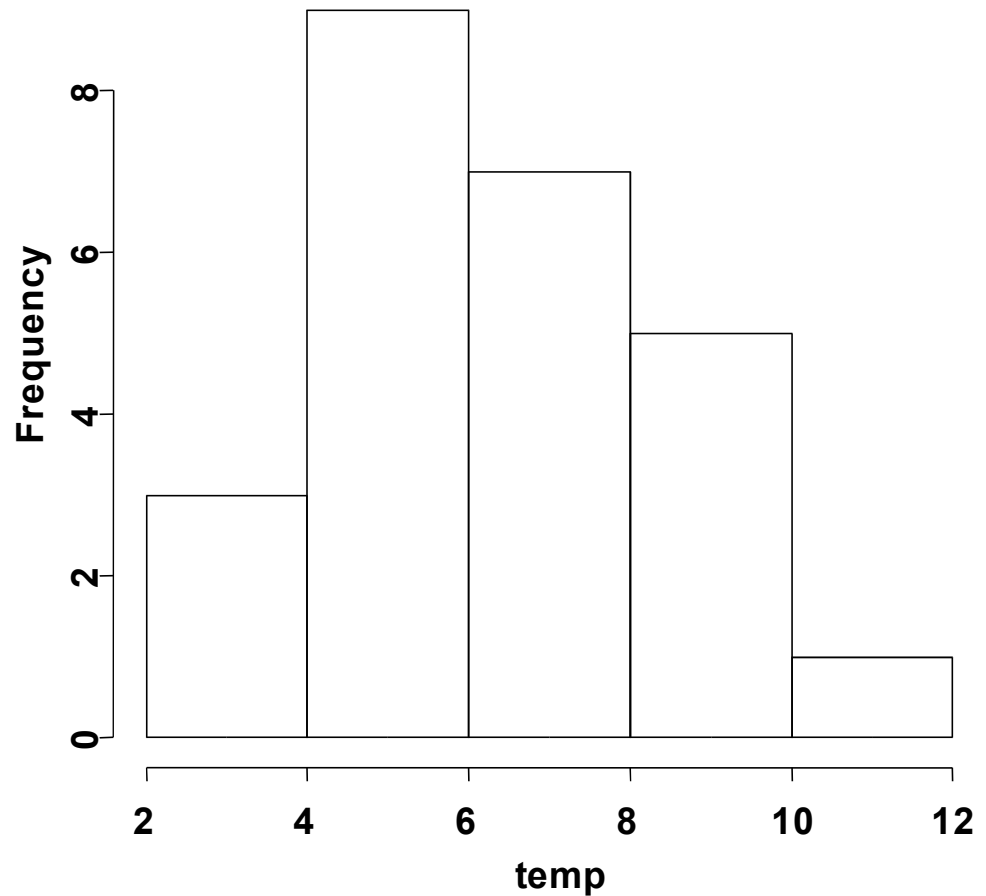


連続変数の図示の例1

平成元年3月9
日から4月2日
の東京地区の最
低気温

3.2, 3.1, 5.1, 4.8,
8.3, 9.8, 8.3, 6.6,
5.1, 3.8, 5.2, 5.6,
6.5, 5.7, 5.7, 7.4,
6.2, 7.0, 6.7, 5.7,
6.2, 6.0, 8.8,
10.7, 8.5

Lowest temperature in Tokyo Mar-9-1989 to Apr-2-1989



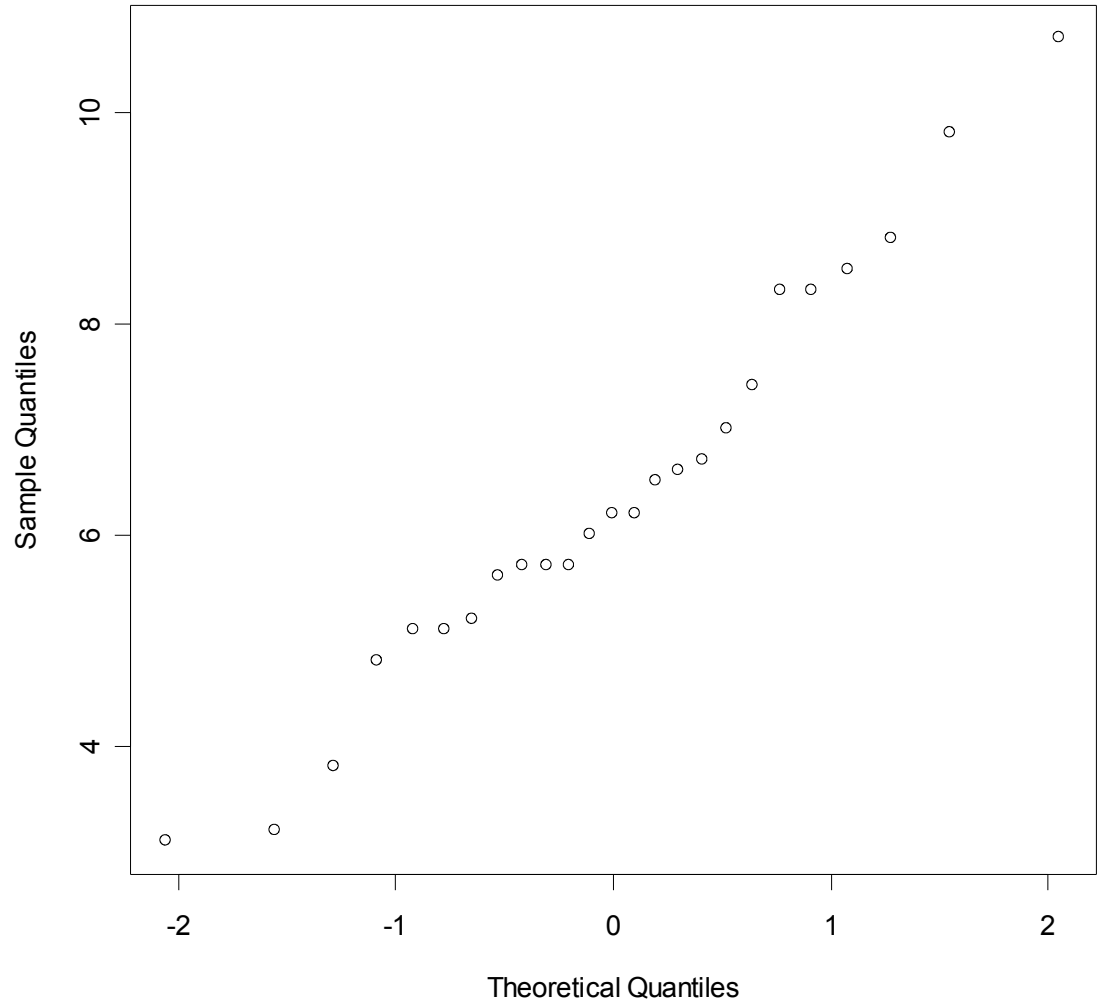
ヒストグラム

連続変数の図示の例2

平成元年3月9
日から4月2日
の東京地区の最
低気温

3.2, 3.1, 5.1, 4.8,
8.3, 9.8, 8.3, 6.6,
5.1, 3.8, 5.2, 5.6,
6.5, 5.7, 5.7, 7.4,
6.2, 7.0, 6.7, 5.7,
6.2, 6.0, 8.8,
10.7, 8.5

Normal QQplot



正規確率プロット

連続量変数の図示の例3

平成元年3月9
日から4月2日
の東京地区の最
低気温

3.2, 3.1, 5.1, 4.8,
8.3, 9.8, 8.3, 6.6,
5.1, 3.8, 5.2, 5.6,
6.5, 5.7, 5.7, 7.4,
6.2, 7.0, 6.7, 5.7,
6.2, 6.0, 8.8,
10.7, 8.5

幹葉表示 (Stem and leaf plot) では,

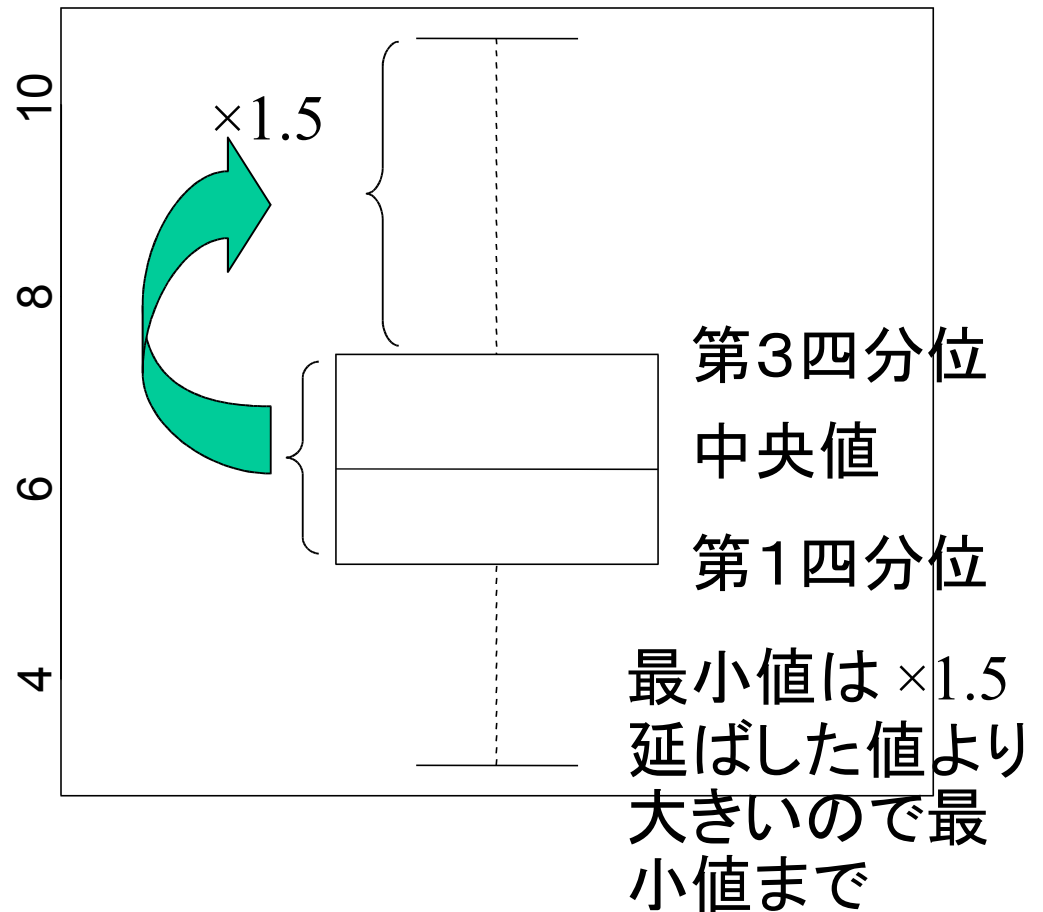
```
3 | 2 1 8
4 | 8
5 | 1 1 2 6 7 7 7
6 | 6 5 2 7 2 0
7 | 4 0
8 | 3 3 8 5
9 | 8
10 | 7
```

連続変数の図示の例4

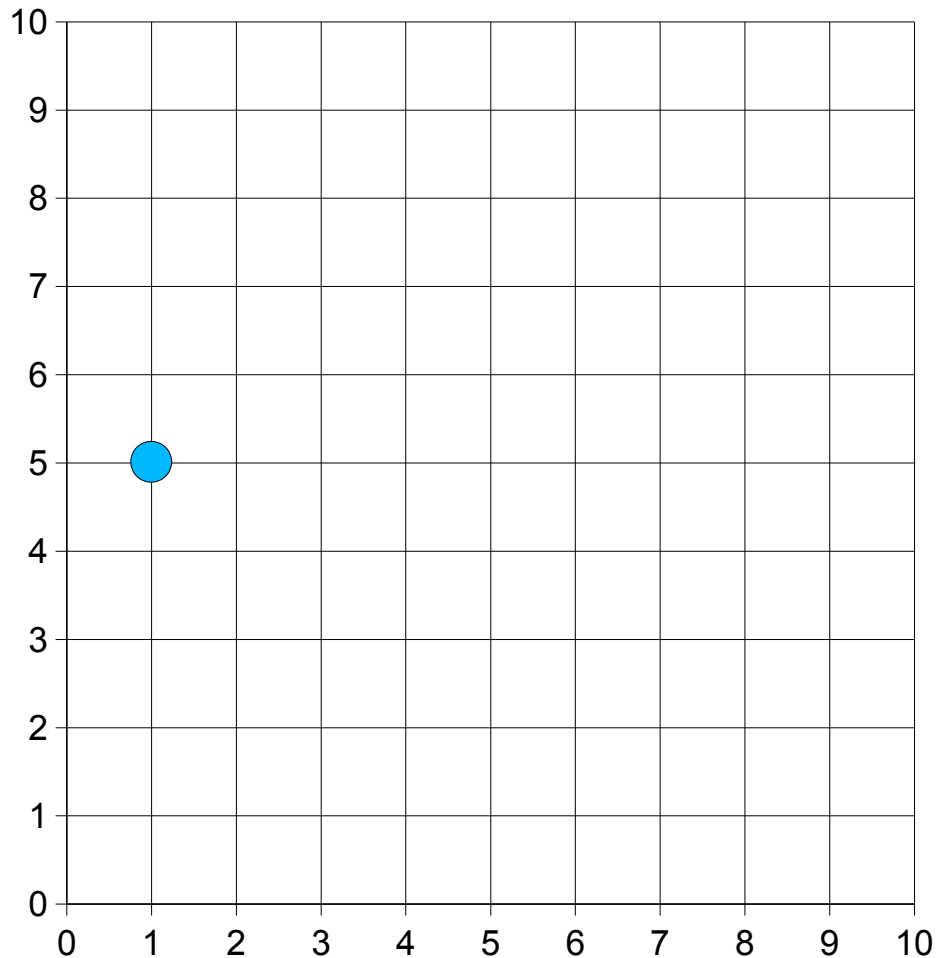
平成元年3月9
日から4月2日
の東京地区の最
低気温

3.2, 3.1, 5.1, 4.8,
8.3, 9.8, 8.3, 6.6,
5.1, 3.8, 5.2, 5.6,
6.5, 5.7, 5.7, 7.4,
6.2, 7.0, 6.7, 5.7,
6.2, 6.0, 8.8,
10.7, 8.5

箱ヒゲ図では,



散布図を描く練習



左の方眼に、次のデータ点をプロットしてみよう。

(1,5) : プロット済み

(2,4)

(10,8)

(1,3)

(4,2)

(7,8)

(9,7)

なお、R に書かせるなら、

```
x <- c(1,2,10,1,4,7,9)
```

```
y <- c(5,4,8,3,2,8,7)
```

```
plot(x,y)
```