



# 確率と統計

中山クラス  
第3週



# 第2回レポートについて

- ◆締切 10月31(金)17:00
- ◆提出場所: 1号館2階1-205庶務課前 レポート提出箱
- ◆レポート内容: 次のスライドに示す内容を作成し, 別紙で作成した表紙とともに1つにまとめて左上をステープル留めして提出のこと.
- ◆レポート作成上の注意とヒント: 第1回レポートと同じ
- ◆レポート作成例: 講義WebPageからダウンロード  
ただし, 使用しているデータは異なる

# 第2回レポート課題

- I. 以下に示す用語の意味を説明せよ.  
相関, 連関, 共分散, 相関係数, クロス集計表,  
ファイ係数
- II. 第3章 練習問題, 及び, 以下の項目に対する解答を作成せよ. Rの出力を含める場合はR出力部分をコピー&ペーストして良い. 記述部分は手書き.
  - (1) で作成した散布図から分かることを述べよ.
  - (2) の結果から分かることを述べよ.
  - (3) の結果から分かることを述べよ.
  - (4) の結果から分かることを述べよ.

# 今日の内容

## 第3章 2つの変数の記述統計

2つの変数の関係について、量的変数どうし、及び、質的変数どうしの関係について学ぶ。

3.1 2つの変数の関係

3.2 散布図

3.3 共分散

3.4 相関係数

3.5 クロス集計表

3.6 ファイ係数

教科書に書いてあることを実際自分の手で打ち込みながら、Rで各種統計量を計算してみよう。

# csvファイルの活用

使用方法は第2週の講義で説明(講義資料参照)

R Console画面で入力したデータは作業を終了すると消える. 作業スペースを保存すると残るが, 管理が面倒. 特に, データの種類が増えてくるとファイルで管理した方が便利.

内容が分かるようにファイル名をつける.

Rで作業する時には, 別の変数に入力して使用

```
> aaa <- read.csv("xxx.csv")
```

# 3.1 2つの変数の関係

第2章: 1つの変数の統計量(平均や分散, 標準偏差)

第3章: 2つの変数の関係を統計量で表す.

**相関:** 量的変数どうしの関係

例)「国語の得点が高い人ほど英語の得点が高い」

**連関:** 質的変数どうしの関係

例)「洋食が好きな人には甘党が多い」

## 3.2 散布図

**相関**は2変数の関係の強さを表す

2つの変数 $x$ と $y$ の関係は大まかにいって次の3パターン.

**正の相関**:  $x \rightarrow$ 大のとき $y \rightarrow$ 大 相関が強い(大きい)

**負の相関**:  $x \rightarrow$ 大のとき $y \rightarrow$ 小 逆相関が強い(大きい)

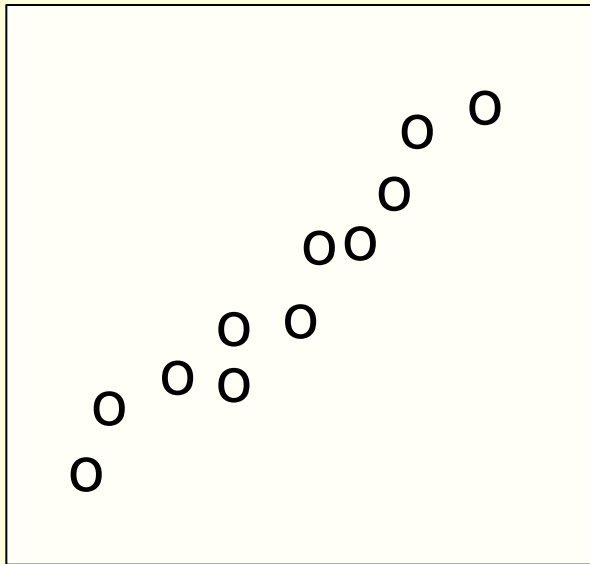
**無相関**: 上記のような関係がない 相関がない(弱い)

**散布図**: 2つの変数を縦軸と横軸にとり, データを点で表した図

→散布図で, 相関の様子が視覚的に把握できる.

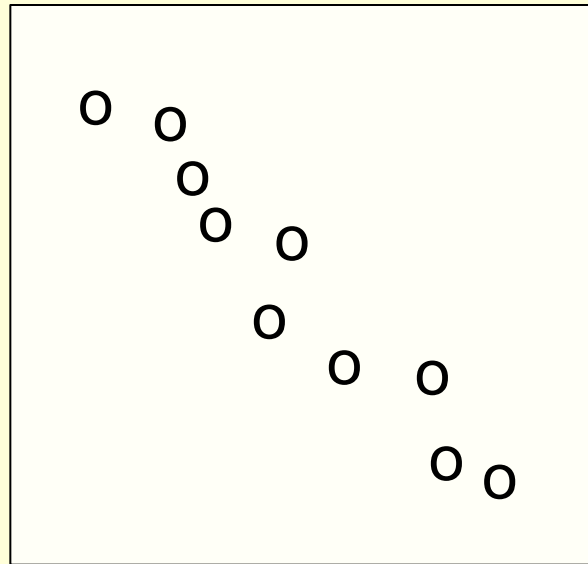
# 散布図の例

正の相関



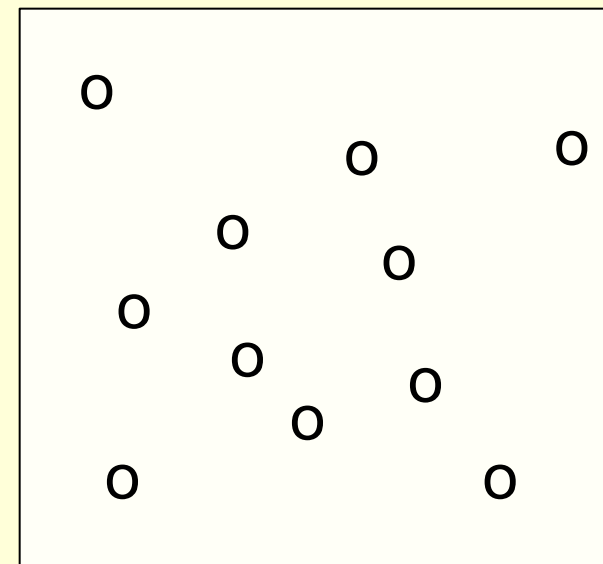
英語の点数

負の相関



国語の点数

相関がない



社会の点数

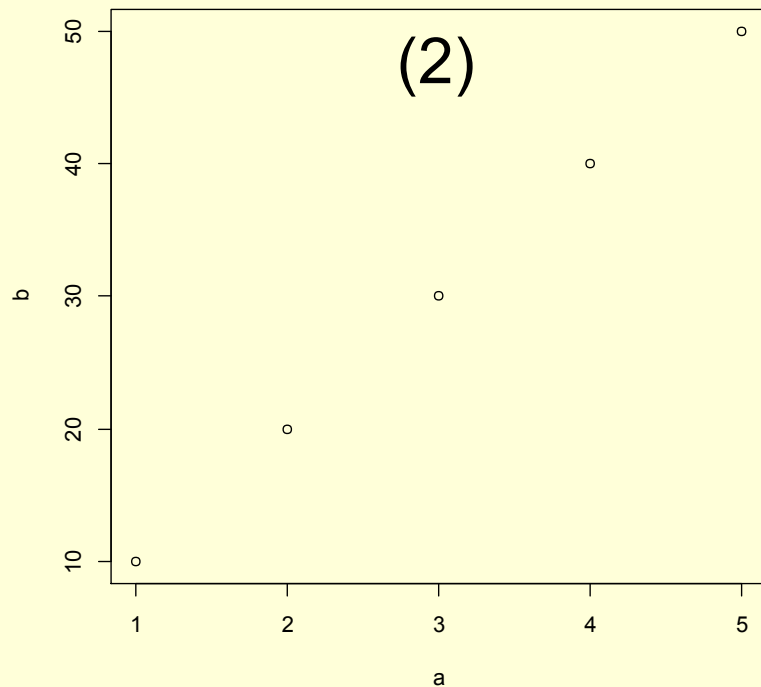
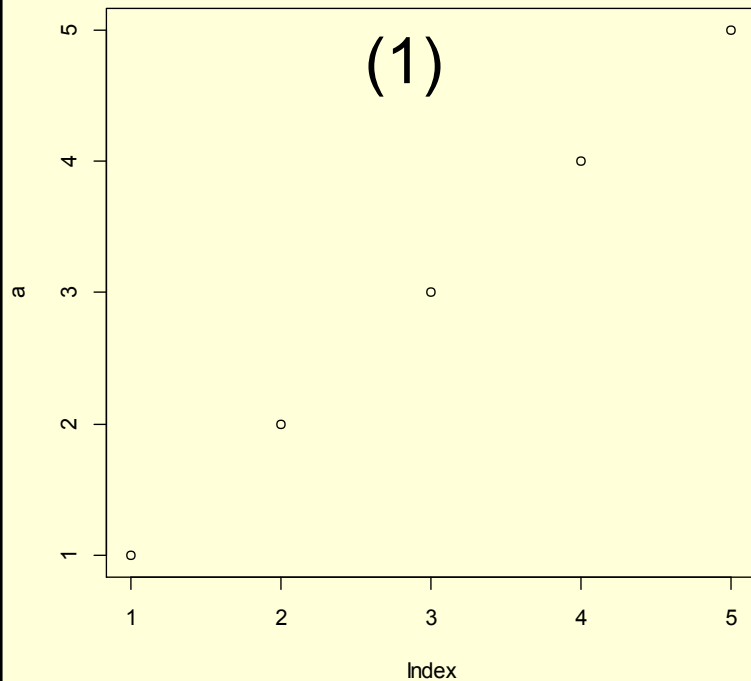
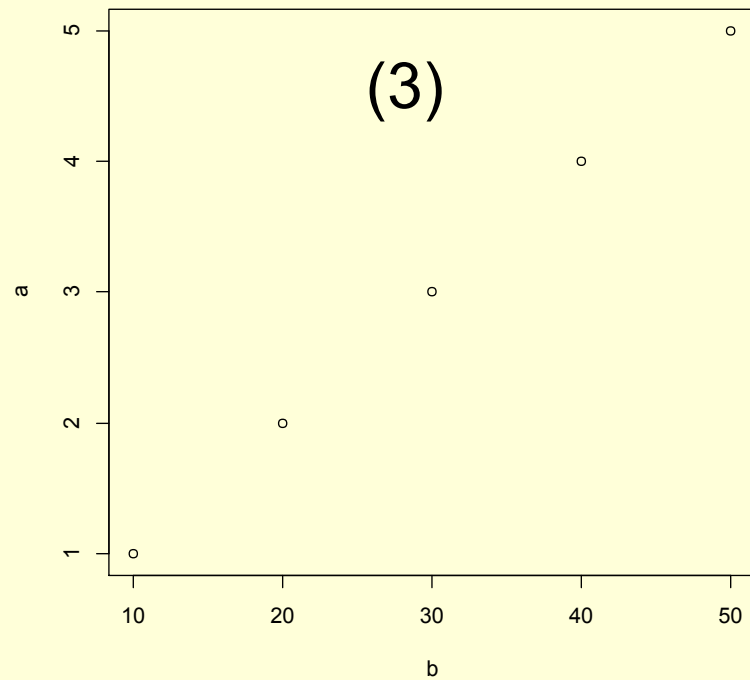
Rのコマンド

```
> plot(国語の点数, 数学の点数)
```

( , )は半角



```
> a <- c(1,2,3,4,5)
> plot(a) #(1)
> b <- c(10,20,30,40,50)
> plot(a,b) #(2)
> plot(b,a) #(3)
```



# plot( )図のコピー&ペースト

plot( )で作成された図のウィンドウで右クリック  
「メタファイルにコピー」

または

「ビットマップにコピー」

を選択

コピーしたいファイル(R以外)に移動して「貼り付け」  
図がコピーされる.

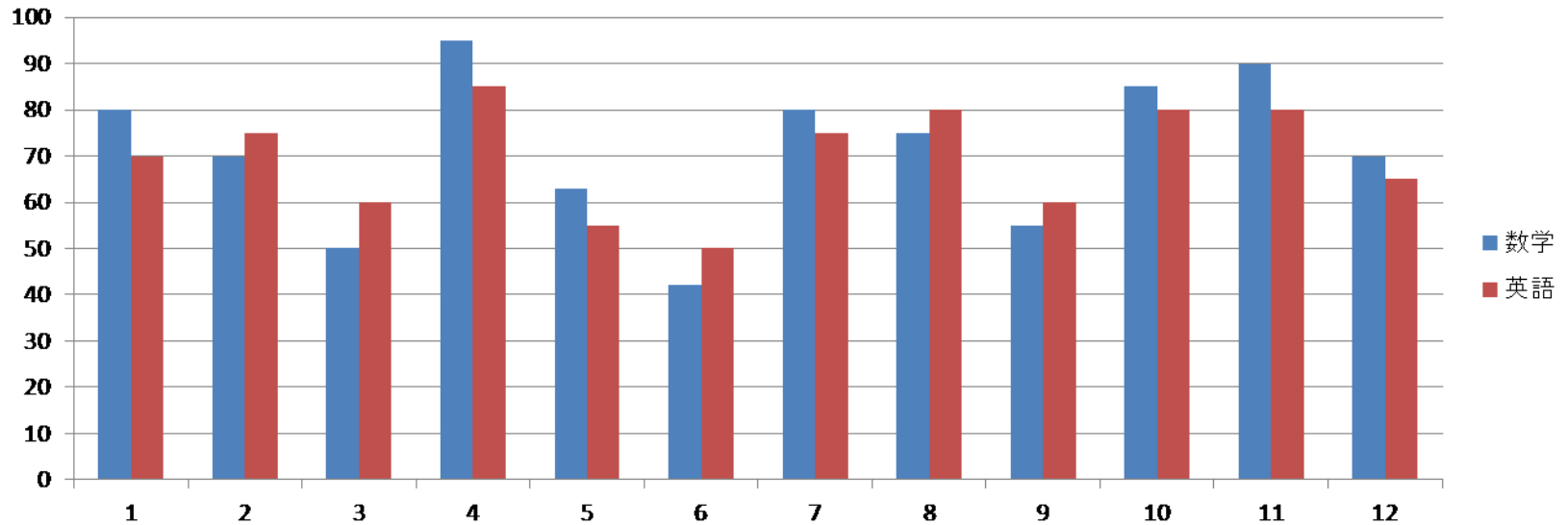
メタファイル: 図のデータを数式で保存

→伸縮してもきれい

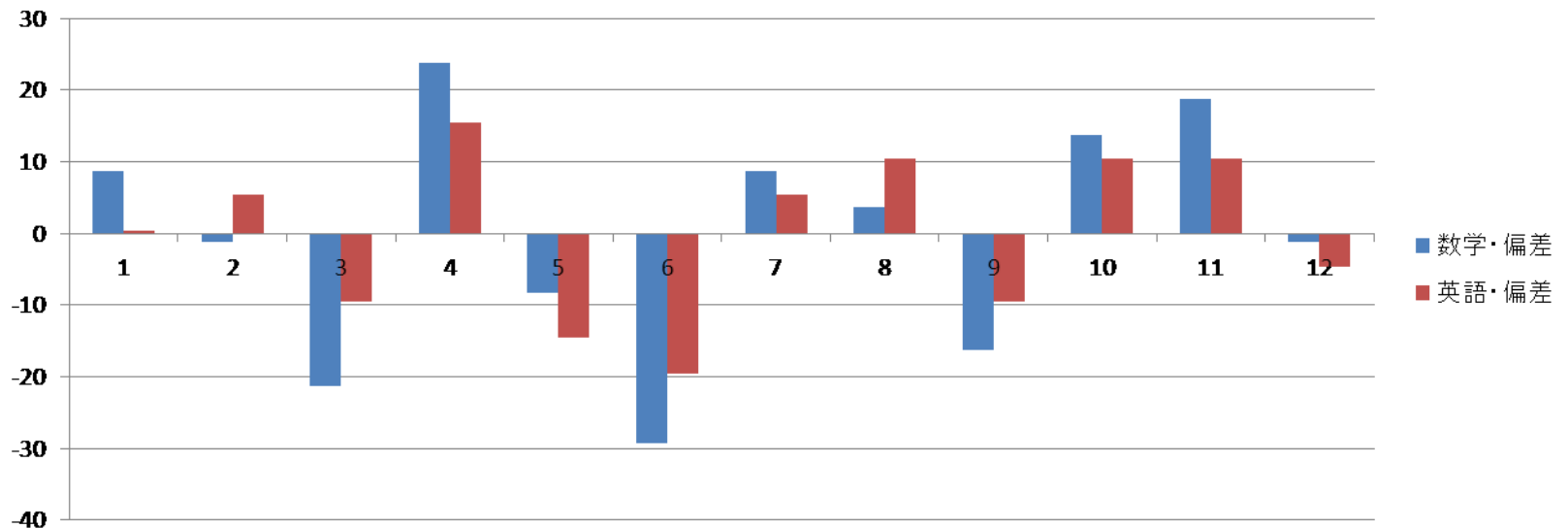
ビットマップ: 画像そのままをマス目データとして保存

→拡張するとぼける

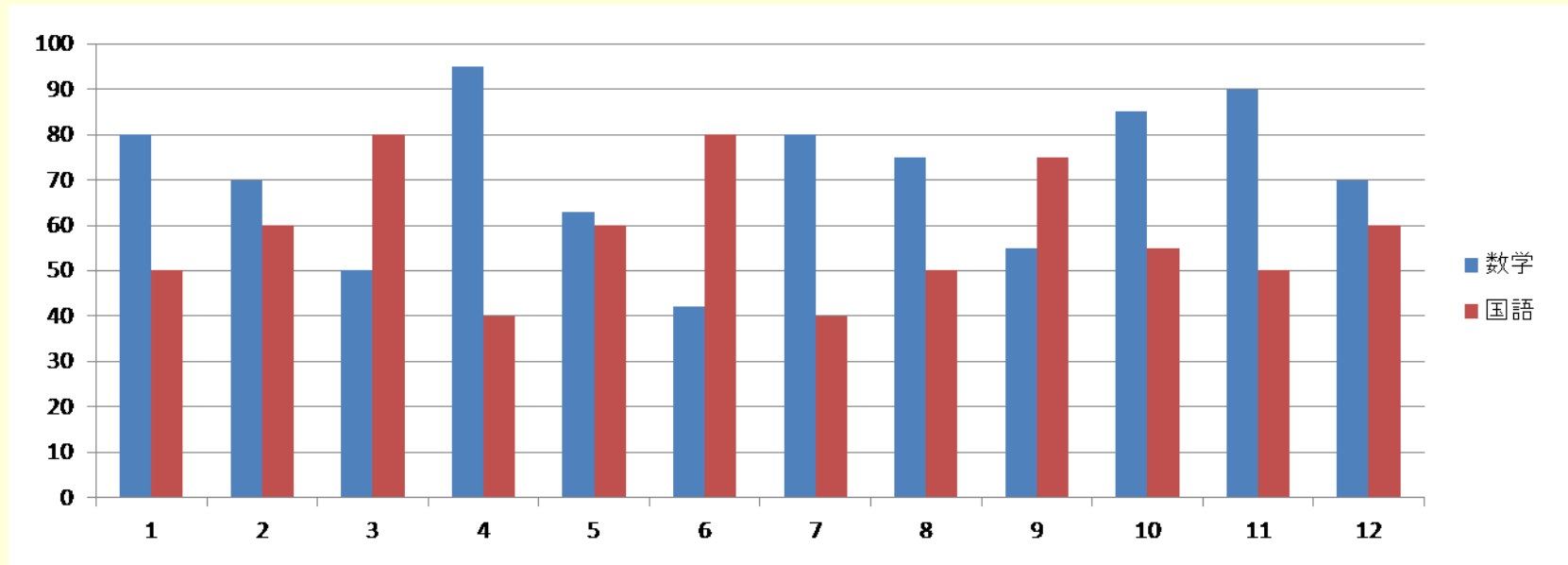
# 相関の例(数学と英語)



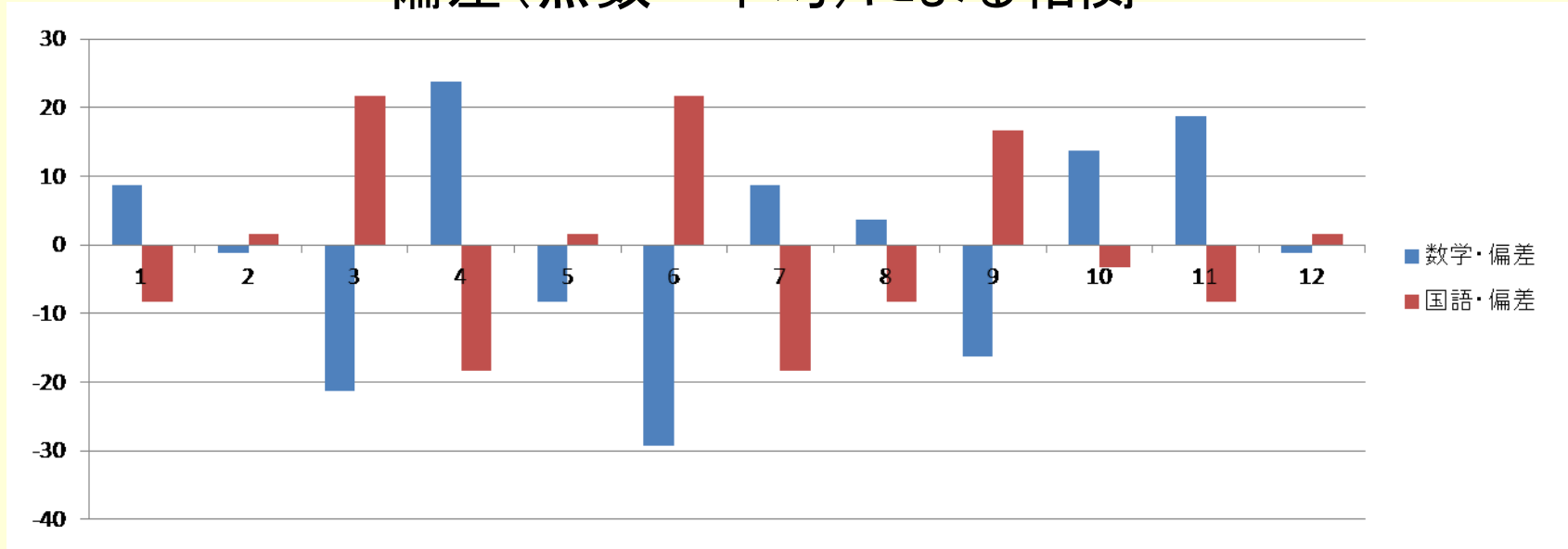
## 偏差(点数-平均)による相関



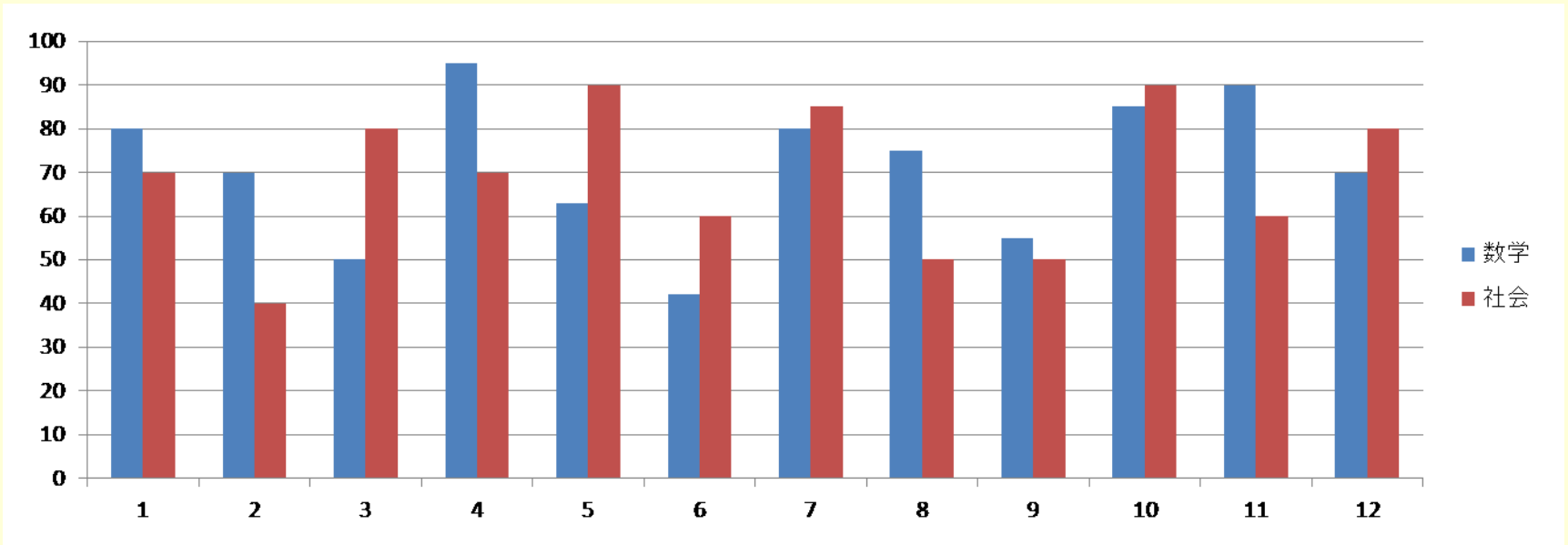
# 相関の例(数学と国語)



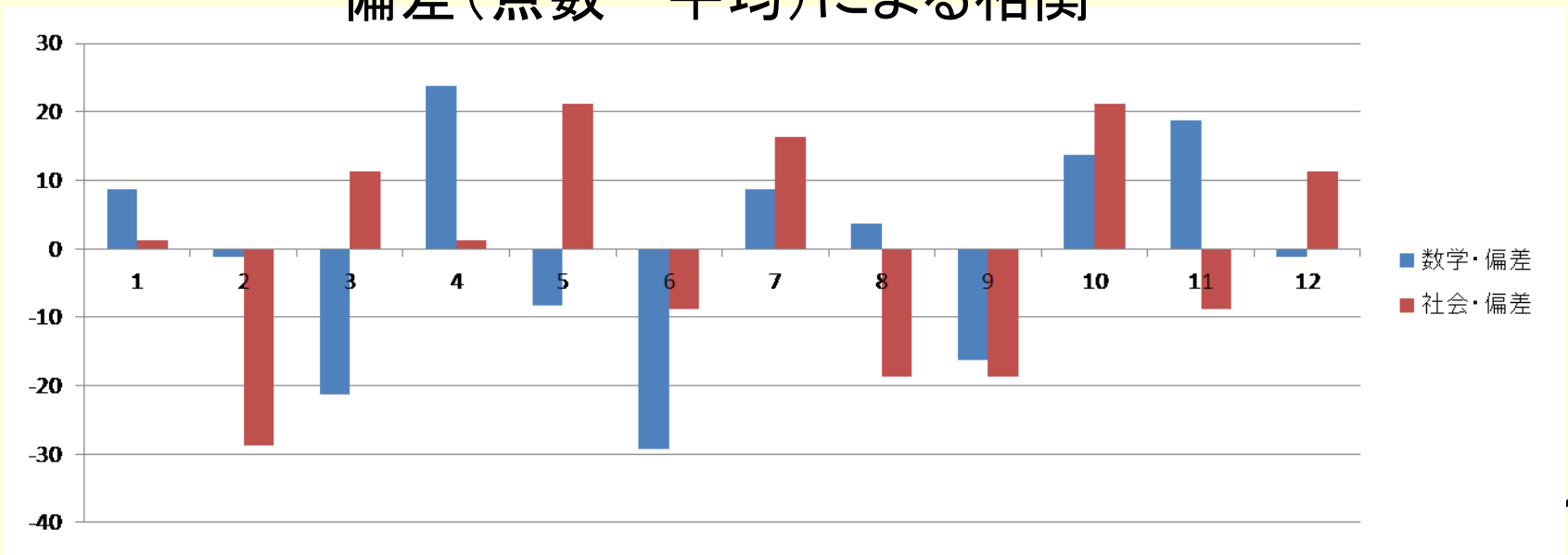
## 偏差(点数-平均)による相関



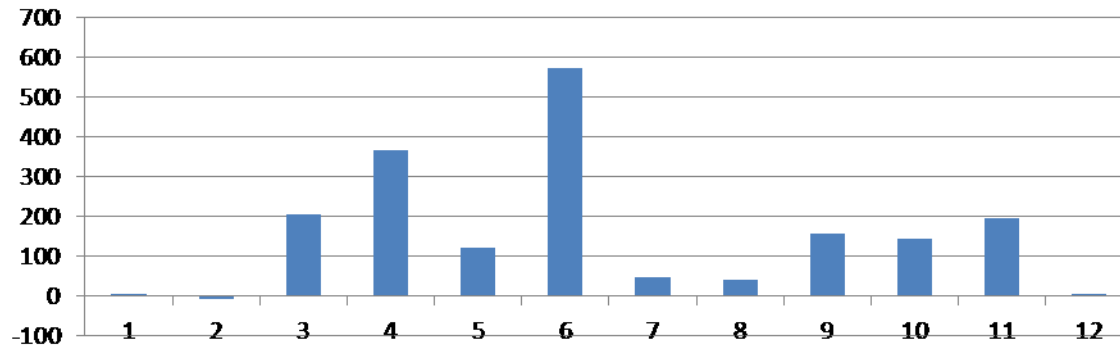
# 相関の例 (数学と社会)



## 偏差 (点数 - 平均) による相関

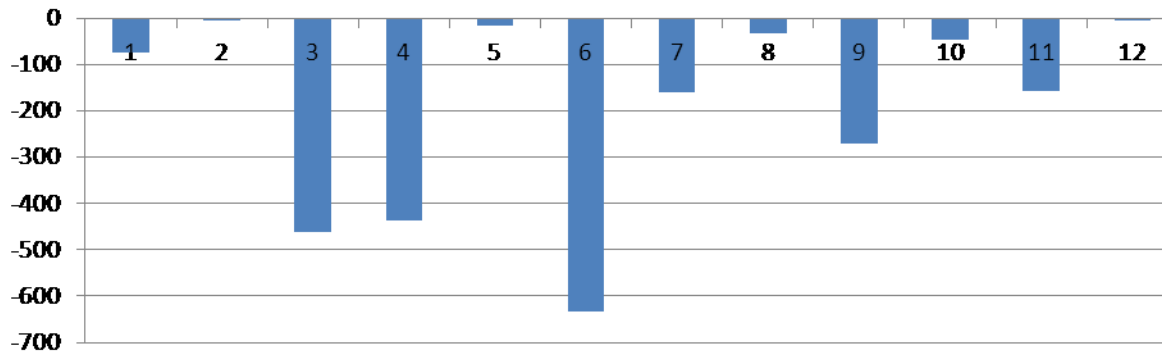


## 数学・英語の偏差の積



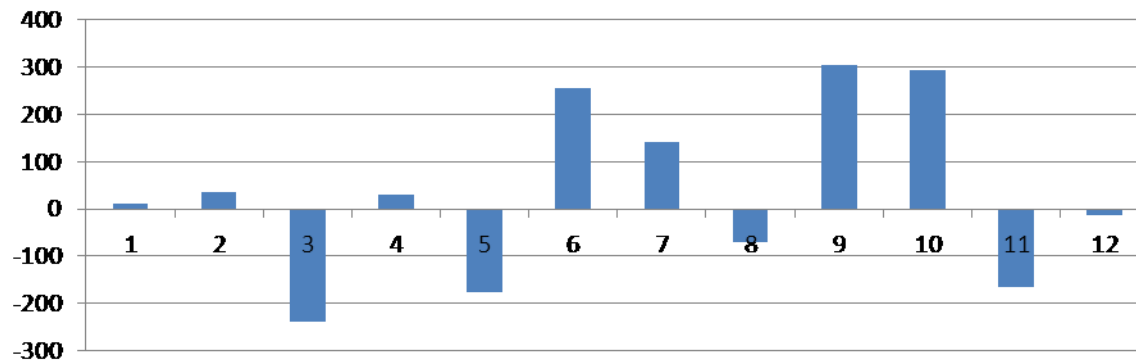
■ 数学・英語の偏差の積

## 数学・国語の偏差の積



■ 数学・国語の偏差の積

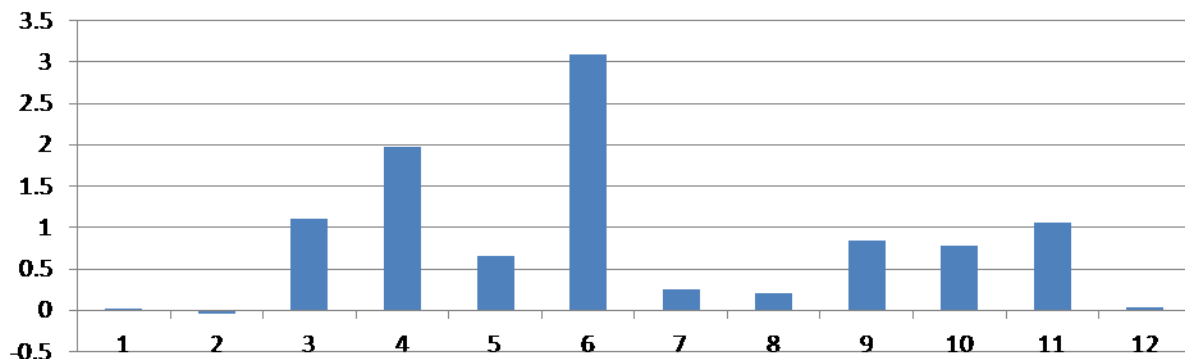
## 数学・社会の偏差の積



■ 数学・社会の偏差の積

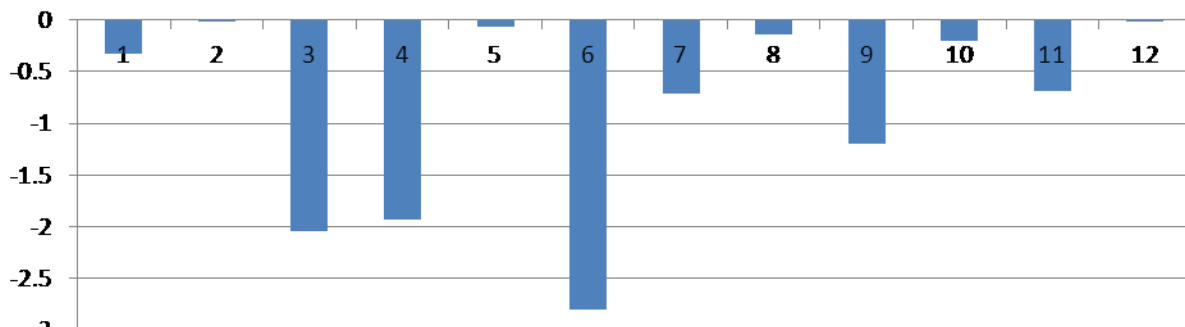
点数の大きさの  
影響を受ける

## 数学・英語／標準偏差



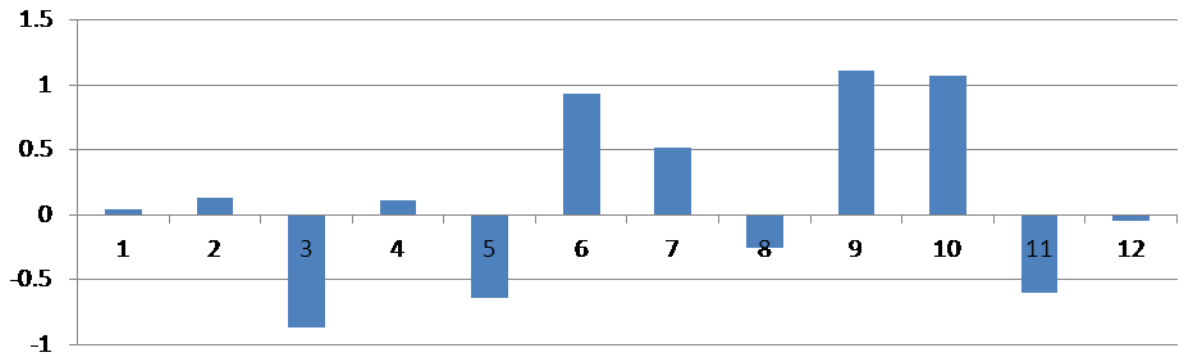
■ 数学・英語／標準偏差

## 数学・国語／標準偏差



■ 数学・国語／標準偏差

## 数学・社会／標準偏差



■ 数学・社会／標準偏差

点数の大きさの  
影響を受けない

## 3.3 共分散(1)

**相関**: 2つの変数の相関は偏差(データの値-平均)により強調される.

**共分散**: 2つの変数の偏差の積の平均

(参考)分散

$$S_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$V = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Rによる計算式

```
> cov12 <- sum((T1-mean(T1))*(T2-mean(T2)))/length(T1)
```



# Rによるベクトル計算(1)

$$c(1, 2, 3)+4=c(1+4, 2+4, 3+4)$$

$$c(1, 2, 3)*4=c(1*4, 2*4, 3*4)$$

$$c(1, 2, 3)/4=c(1/4, 2/4, 3/4)$$

$$c(1, 2, 3)+c(4, 5, 6)=c(1+4, 2+5, 3+6)$$

$$c(1, 2, 3)*c(4, 5, 6)=c(1*4, 2*5, 3*6)$$

$$c(1, 2, 3)/c(4, 5, 6)=c(1/4, 2/5, 3/6)$$

$$c(1, 2, 3, 4)+c(1, 2)=c(1+1, 2+2, 3+1, 4+2)$$

$$c(1, 2, 3, 4)/c(1, 2)=c(1/1, 2/2, 3/1, 4/2)$$

$$c(1, 2)/c(1, 2, 3, 4)=c(1/1, 2/2, 1/3, 2/4)$$

データの個数が整数倍であれば、計算されるが、  
計算内容を理解することが重要

# Rによるベクトル計算(2)

```
> T1 <- c(1,2,3,4,5)
```

```
> Tm <- mean(T1)
```

```
> Tm
```

```
[1] 3
```

```
> T1-Tm
```

```
[1] -2 -1 0 1 2
```

```
> T2 <- c(1,2,3,4,5)
```

```
> T1*T2
```

```
[1] 1 4 9 16 25
```

```
> T3 <- c(2,3,4,5,6)
```

```
> T1+T3
```

```
[1] 3 5 7 9 11
```

```
> T1^2
```

```
[1] 1 4 9 16 25
```

```
> T1+T1
```

```
[1] 2 4 6 8 10
```

## 3.3 共分散(2)

分散: 分母が $n-1 \rightarrow$ 不偏分散

共分散: 分母が $n-1 \rightarrow$ 不偏共分散

Rの共分散関数: cov  $\rightarrow$  不偏共分散 ・・・注意

共分散はデータ(数値)の大きさによって変わる

また, 単位を変える(例: m  $\rightarrow$  cm)とデータの大きさが変わる  
るので共分散も変わる.

データの大きさに関係なく相関を表すために, 共分散を  
標準偏差で正規化する  $\rightarrow$  相関係数

## 3.4 相関係数(1)

相関係数:

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

← (共分散)  
← (xの標準偏差) (yの標準偏差)

Rによる計算式

- > cov(T1, T2)/sd(T1)\*sd(T2) 共分散と標準偏差で計算
- > cor(T1, T2) Rの関数を使用

Rでは共分散, 標準偏差とも不偏分散に基づいているが, 相関係数では分母と分子の両方がn-1で割った形になっているので, 相殺される.

## 3.4 相関係数(2)

$-1 \leq \text{相関係数} \leq 1$

相関係数 = 1      ←  $x(n) = y(n)$

相関係数 = -1   ←  $x(n) = -y(n)$

相関係数が1に近い → 強い正の相関がある

    -1に近い → 強い負の相関がある

    絶対値が小さい → 相関が弱い

相関係数の目安 → 表3.1

2変数の関係を調べるときは、必ず散布図を確認する  
習慣をつけること.

## 3.5 クロス集計表

**相関**: 2つの量的変数の関係を示す.

**連関**: 2つの質的変数の関係を示す.

**クロス集計表**: 2つの質的変数を縦横にとって  
度数を集計した表

```
> 数学 <- c("嫌い","嫌い","好き","好き",...)
> 統計 <- c("好き","好き","嫌い","好き",...)
> table(数学,統計)
```

		統計	
		嫌い	好き
数学	嫌い	10	4
	好き	2	4

## 3.6 ファイ係数(1)

**相関係数**: 2つの量的変数の関係を示す.

**ファイ係数**: 2つの質的変数の関係(連関)を示す.

1と0の2値変数で計算される**相関係数**  
クロス集計表を記述する指標

好き=1, 嫌い=0と置き換えて**相関係数**を計算

→ **ファイ係数**

ファイ係数=正で値が大きい → 正の強い相関

負で絶対値が大きい → 負の強い相関

ファイ係数だけでなく, クロス集計表も使うこと.

## 3.6 ファイ係数(2)

Rによる計算

<データ入力／csvファイルの読み込み>

```
> 数学 <- c("嫌い","嫌い","好き","好き",...)
```

```
> 統計 <- c("好き","好き","嫌い","好き",...)
```

<好き=1, 嫌い=0への変換>

```
> 数学10 <- ifelse(数学=="好き",1,0)
```

```
> 統計10 <- ifelse(統計=="好き",1,0)
```

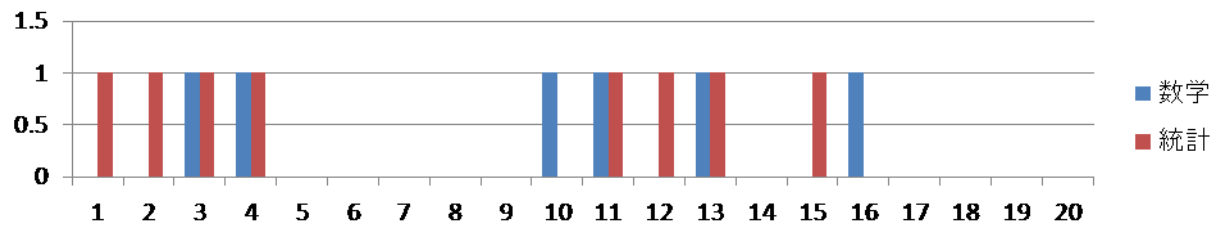
<2値データの相関係数を計算>

```
> cor(数学10,統計10)
```

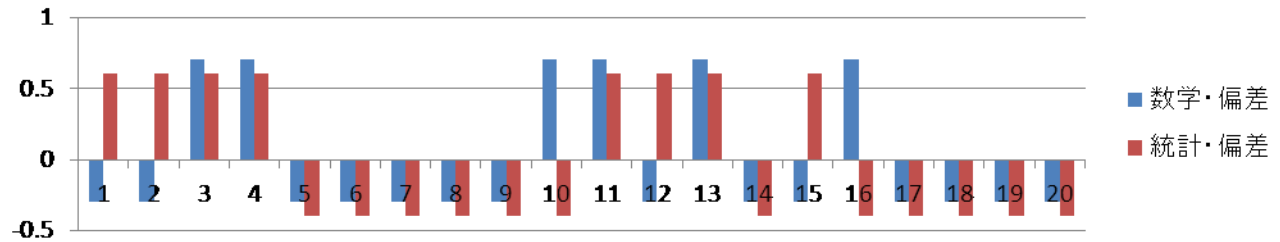
```
[1] 0.356
```



## 2値データ

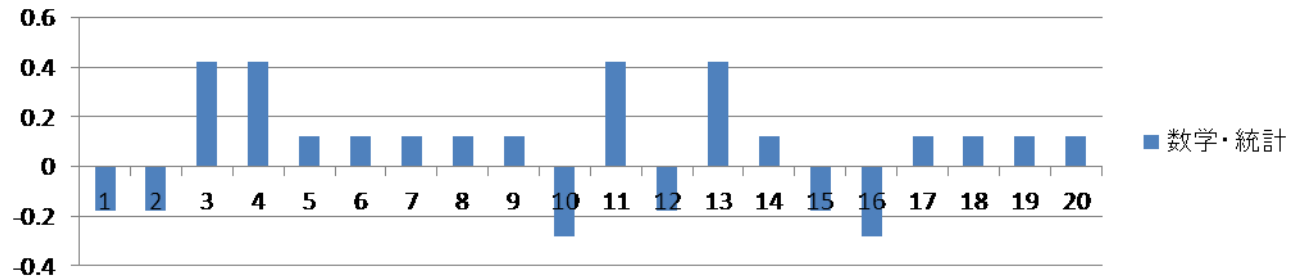


## 2値データ ー平均



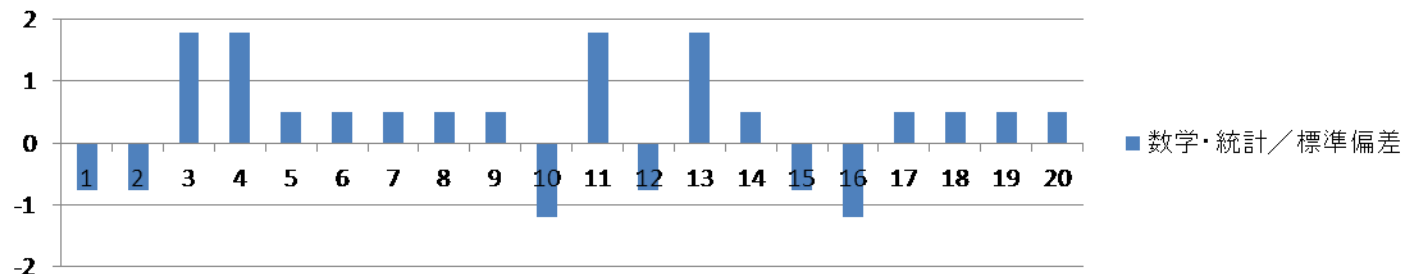
## 数学・統計

## (2値データ ー平均)の積



## 数学・統計／標準偏差

## [(2値データ ー平均)の積]／ 標準偏差



# 第3章の関数のまとめ

p.68 表3.3を参照

(参考)

table(X)	Xの度数分布表	第2章
table(X,Y)	XとYのクロス集計表 X:列方向, Yが行方向	第3章