

基本的な関数



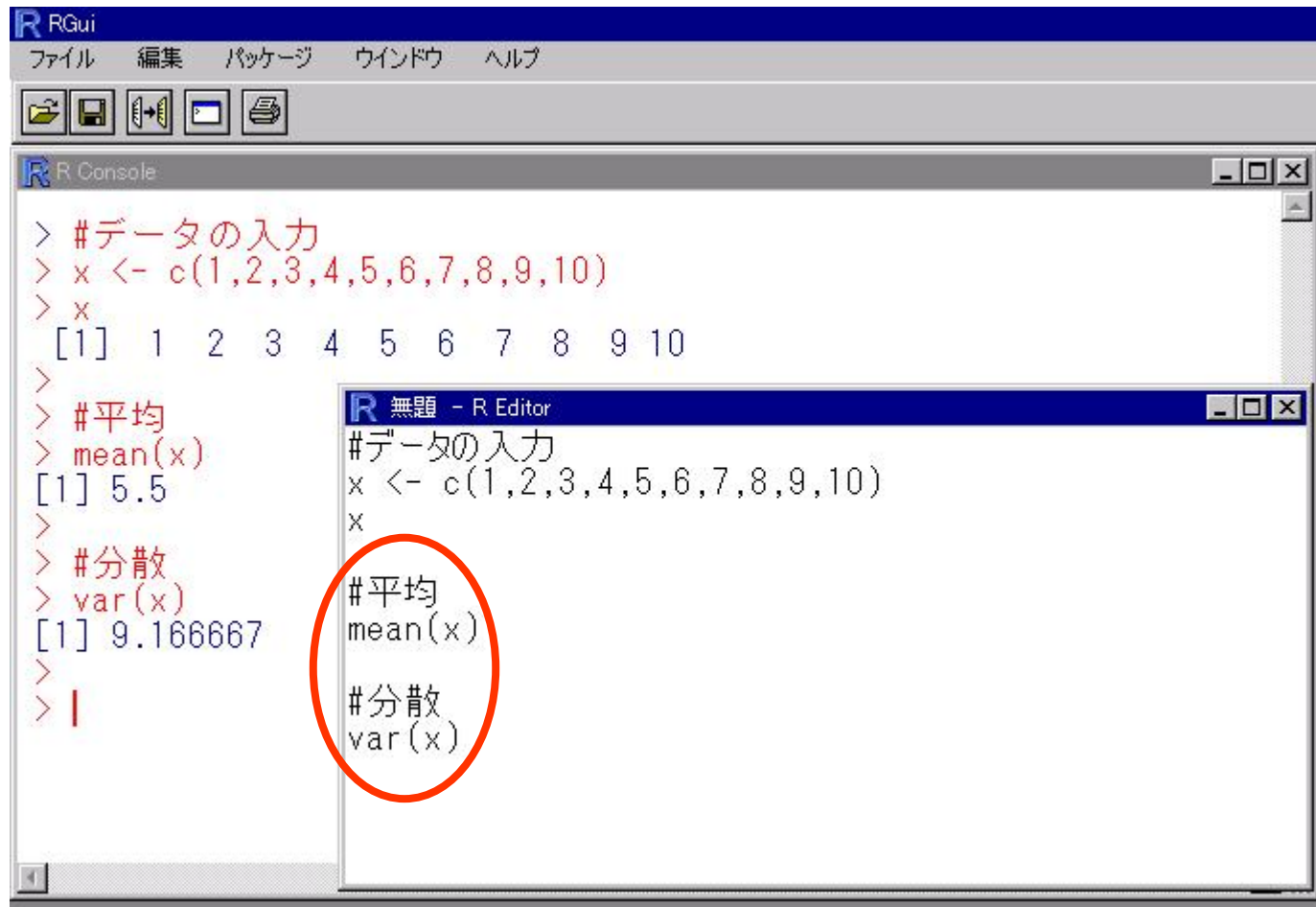
平均・分散の計算

- 「R」では、平均・分散を簡単に計算することができる
 - 平均を求める関数 : `mean()`
 - 分散を求める関数 : `var()`

- 「R」では分散は「不偏分散」として定義されていることに注意

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

「R」での平均・分散の計算



The image shows a screenshot of the RGui interface. The main window is titled "R Console" and contains the following code and output:

```
> #データの入力
> x <- c(1,2,3,4,5,6,7,8,9,10)
> x
[1] 1 2 3 4 5 6 7 8 9 10
>
> #平均
> mean(x)
[1] 5.5
>
> #分散
> var(x)
[1] 9.166667
>
> |
```

An inset window titled "R 無題 - R Editor" is overlaid on the console, showing the same code and output. The lines for the mean and variance calculations are circled in red:

```
#データの入力
x <- c(1,2,3,4,5,6,7,8,9,10)
x
#平均
mean(x)

#分散
var(x)
```

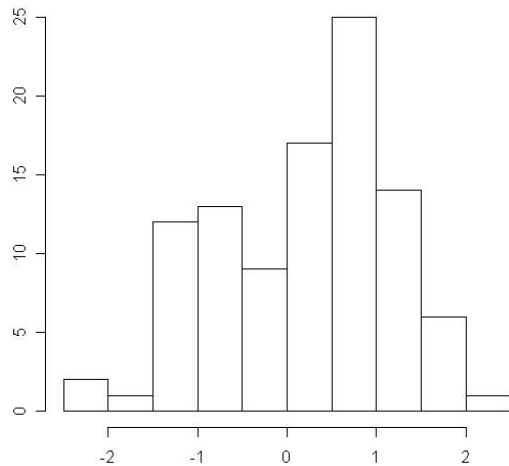
データの視覚化

- 統計学では、データを数字の並びとして眺めるだけでなく、ヒストグラムや散布図を描いて**視覚化**することにより、見通しが良くなることもある
- Rでデータを視覚化するための主な関数
 - ヒストグラム : `hist()`
 - 散布図 : `plot()`
 - 確率分布 : `curve()`

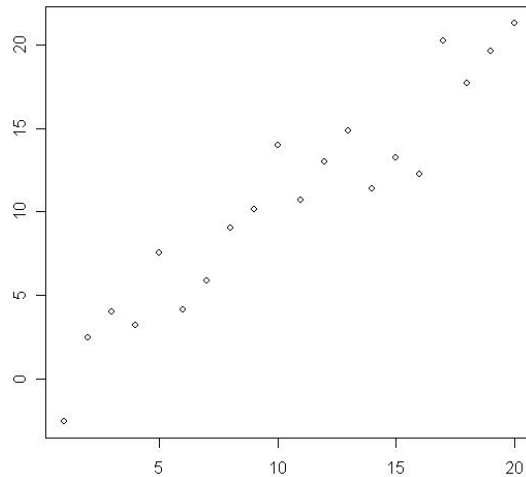
代表的な関数

□ 次のような図が描けるようになります。

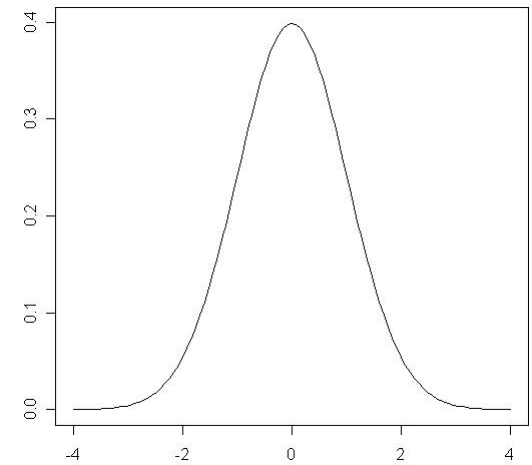
ヒストグラム



散布図



確率分布



ヒストグラム



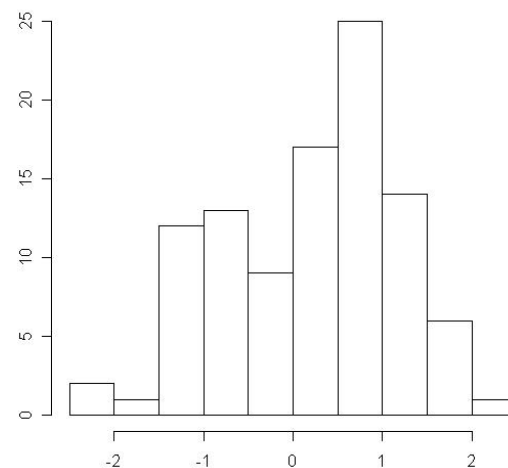
ヒストグラム

□ 1変量のデータに対するグラフ表現の1つ

- データの分布状況の把握
- 外れ値の検出
- データの分布状況の把握
 - 対称形かどうか、偏りはあるか

□ hist(): ヒストグラムを描く関数

- hist()で利用できる主な引数
 - breaks : 階級幅の設定
 - prob : Y軸の取り方(頻度、確率)
 - col : 色の設定



成績データ

□ 成績データ

- 2つの高校A、Bにおいて、3年生の数学の学力に差があるかどうかを調べるため、A高校から8人、B高校から7人を無作為に選んで、実力テストを行ったところ、次のような結果を得た。

| | | | | | | | | |
|------|----|----|----|----|----|----|----|----|
| A 高校 | 70 | 67 | 81 | 92 | 78 | 62 | 85 | 73 |
| B 高校 | 66 | 75 | 48 | 58 | 80 | 57 | 50 | |

成績データのヒストグラム

- 右のヒストグラムは、次のプログラムで描くことができる。

```
A <- c(70,67,81,92,78,62,85,73)
```

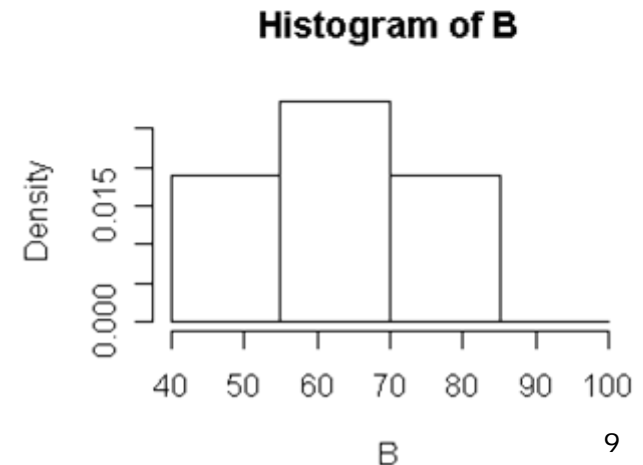
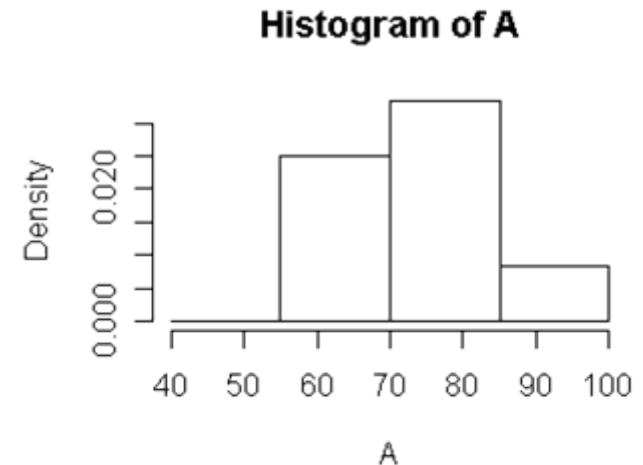
```
B <- c(66,75,48,58,80,57,50)
```

```
par(mfrow=c(2,1))
```

```
br <- seq(40,100,15)
```

```
hist(A, breaks=br, prob=T)
```

```
hist(B, breaks=br, prob=T)
```



乱数の生成

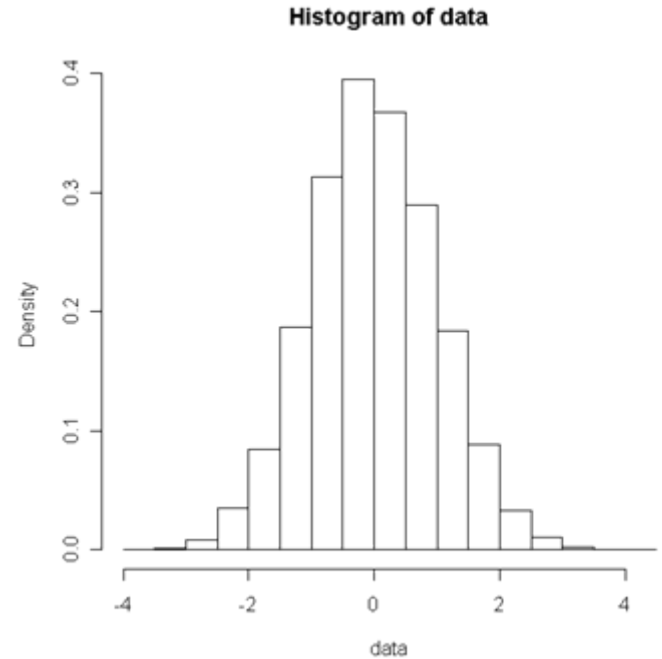
- Rでは、種々の乱数を発生させることができます。
 - 参考URLの下記の箇所が参考になります。
 - R-Tips : 「第60節」 [確率分布と乱数](#)
 - JIN'S PAGE : [Rと確率分布](#)
 - 今回は、正規乱数を発生させて、そのヒストグラムを描いてみます。
 - 正規乱数を発生させる場合と同じ要領で、様々な乱数を発生させることができます。
正規乱数以外の乱数については、参考URLを見て下さい。
 - 正規乱数を発生させるための関数

`rnorm(n=標本数, mean=平均, sd=標準偏差)`

正規乱数とヒストグラム

- 右のヒストグラムは、次のプログラムで描くことができます。

```
data <- rnorm(n=1000,mean=0,sd=1)
hist(data, prob=T)
```

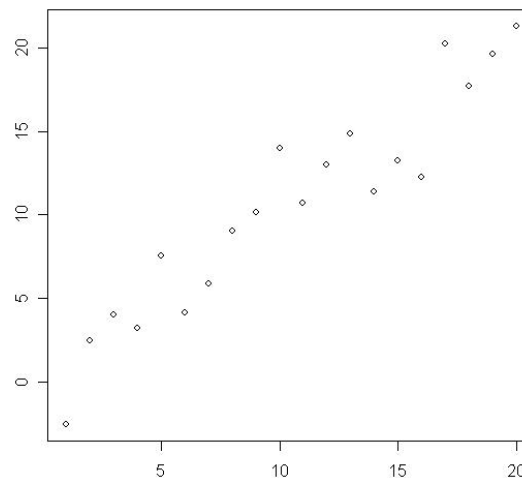


散布図



散布図

- 2変量データに対するグラフ表現の1つ
 - データは直線上に分布しているか。曲線上か。
 - 相関の強さ
 - 外れ値の有無
 - クラスタリングの状況
 - 散布密度



「R」で散布図を描く

□ 基本的な文法

- `plot(x, y)`

□ 面積と家賃のデータ

- ある不動産で取扱っている賃貸物件から無作為に抽出した10件の1DKの物件の面積と1か月の家賃である。

| | | | | | | | | | |
|----|-------|-----|-------|-------|------|------|-------|-------|-------|
| 面積 | 30.16 | 18 | 27.92 | 25.64 | 29.6 | 40 | 35.56 | 35.74 | 40.09 |
| 家賃 | 5.5 | 5.7 | 8 | 8.2 | 11 | 12.3 | 12.9 | 13.7 | 16.3 |

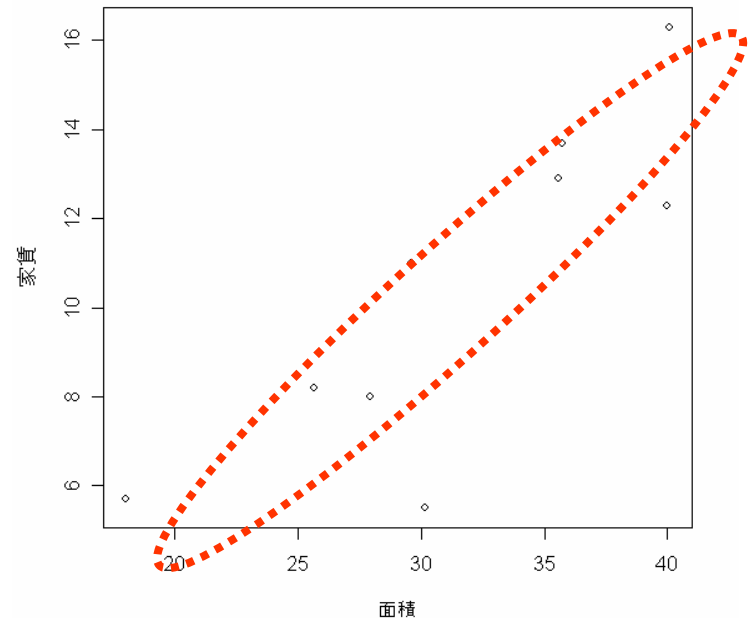
面積と家賃のデータのプロット

- 右の散布図は、次のプログラムで描くことができる。

```
x <- c(30.16, 18, 27.92, 25.64, 29.6,  
       40, 35.56, 35.74, 40.09)
```

```
y <- c(5.5, 5.7, 8, 8.2, 11, 12.3, 12.9,  
       13.7, 16.3)
```

```
plot(x, y, xlab="面積", ylab="家賃")
```



確率分布



確率分布

□ Rでは、種々の確率分布を描くことができます。

■ 参考URLの下記の箇所が参考になります。

□ R-Tips : 「第60節」 [確率分布と乱数](#)

□ JIN'S PAGE : [Rと確率分布](#)

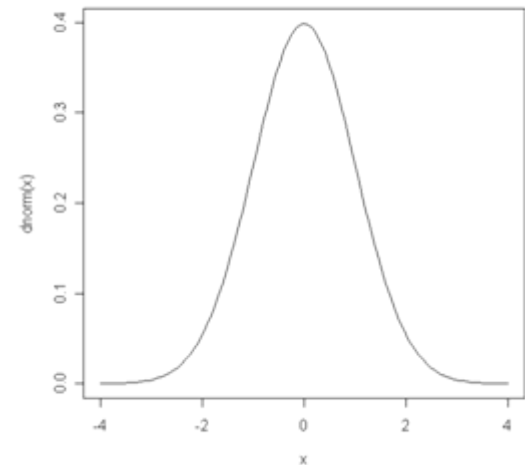
■ 今回は、正規分布を描いてみます。

□ 正規分布を描く場合と同じ要領で、様々なグラフを描くことができます。その他の分布については、参考URLを見て下さい。

正規分布

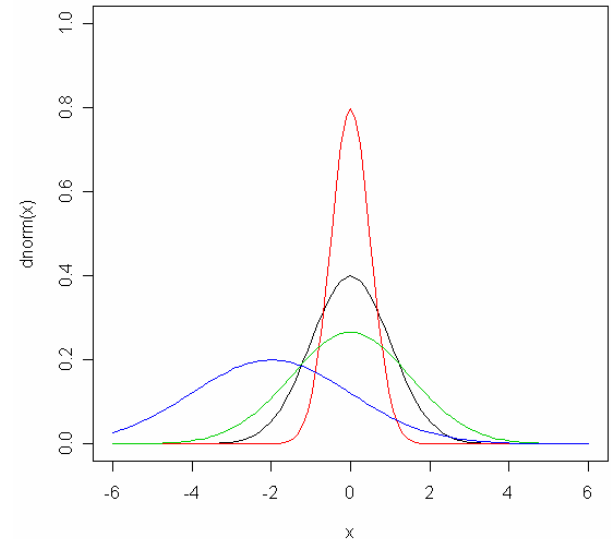
- 右の図は、次のプログラムで描くことができます。

```
curve(dnorm(x), from=-4, to=4)
```



図の重ねがき

- `curve`関数に限っては、“`add=T`”で重ねがきをすることができます。
- 次のプログラムを実行すると、右の図を描くことができます。



```
curve(dnorm(x),from=-6,to=6,ylim=c(0,1),col=1)
curve(dnorm(x,mean=0,sd=0.5),col=2,add=T)
curve(dnorm(x,mean=0,sd=1.5),col=3,add=T)
curve(dnorm(x,mean=-2,sd=2),col=4,add=T)
```