

# 経営統計学基礎

経営情報学科

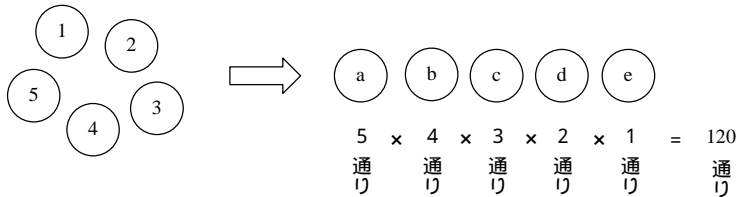
福井正康

# 1章 場合の数

## 1.1 順列

例

5個の番号の付いたボールを1列に並べる場合の数



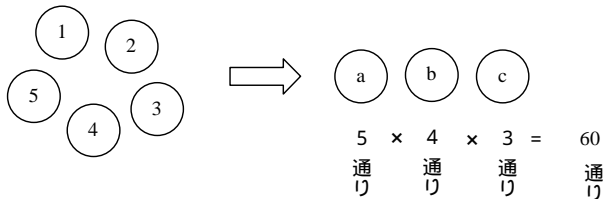
$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

$n$ 個の異なるものを1列に並べる場合の数

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$$

例

5個の番号の付いたボールから3個を取り出し1列に並べる場合の数



$${}_5P_3 = 5 \times 4 \times 3 = \frac{5!}{2!} = \frac{5!}{(5-3)!}$$

$n$ 個の異なるものから $r$ 個取り出し1列に並べる場合の数

$${}_nP_r = n \times (n-1) \times (n-2) \times \cdots \times (n-r+1) = \frac{n!}{(n-r)!}$$

問題

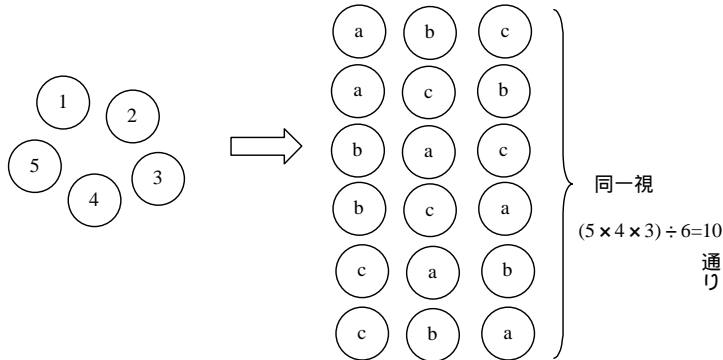
以下の場合の数求めよ。但し、MS-Excelでは、 $n! = \text{fact}(n)$ で与えられる。

- 1) 6個の異なるものを1列に並べる場合の数
- 2) 7個の異なるものから3個取り出して並べる場合の数
- 3) 10個の異なるものから4個取り出して並べる場合の数

## 1.2 組合せ

例

5 個の番号の付いたボールから 3 個取り出す場合の数



$${}_5C_3 = \frac{{}_5P_3}{3!} = \frac{5!}{2!3!} = \frac{5!}{(5-3)!3!}$$

$n$  個の異なるものから  $r$  個取り出す場合の数

$${}_nC_r = \frac{{}_nP_r}{r!} = \frac{n!}{(n-r)!r!}$$

問題

以下の場合の数を求めよ。但し、MS-Excel では  ${}_nC_r = \text{combin}(n, r)$  で与えられる。

- 1) 10 個の異なるものから 4 個取り出す場合の数
- 2) 10 個の異なるものから 6 個取り出す場合の数

問題

${}_nC_r$  と  ${}_nC_{n-r}$  は等しいことを示せ。

## 1.3 その他の場合の数

例

コインもサイコロも区別できるものとして以下について答えよ。

- 1) コインを 3 枚投げる場合の表と裏の出方の場合の数
- 2) コインを 3 枚投げる場合に表が 2 枚出る場合の数
- 3) サイコロを 2 つ振る場合の目の出方の場合の数
- 4) サイコロを 2 つ振る場合に目の合計が 5 になる場合の数

## 2章 確率

### 2.1 事象と集合

例

さいころを投げる場合の出る目の事象

$$1, 2, 3, 4, 5, 6 \quad U = \{1, 2, 3, 4, 5, 6\} : \text{全事象 (集合の言葉では全体集合)}$$

さいころを2個投げる場合の出る目の事象

$$U = \{(1,1), (1,2), (1,3), \dots, (6,5), (6,6)\} : \text{全事象 (全体集合)}$$

注) さいころは区別できるとして求める

例

さいころの目が奇数となる事象  $A$

$$A = \{1, 3, 5\} \quad \bar{A} = \{2, 4, 6\} : \text{余事象 (集合の言葉では補集合)}$$

さいころの目が3以下となる事象  $B$

$$B = \{1, 2, 3\} \quad \bar{B} = \{4, 5, 6\}$$

さいころの目が奇数で「かつ」3以下となる事象

$$A \cap B = \{1, 3\} : \text{積事象 (積集合)}$$

さいころの目が奇数「または」3以下となる事象

$$A \cup B = \{1, 2, 3, 5\} : \text{和事象 (和集合)}$$

さいころの目が4または6となる事象を  $C = \{4, 6\}$  とすると、

$$A \cap C = \phi : \text{空事象 (空集合)}$$

積事象が空事象となるお互いの事象を排反事象という。

### 2.2 確率とは

#### 1. 統計的確率

ある試行を  $n$  回繰り返して、事象  $A$  が  $r$  回起こったとき、 $\lim_{n \rightarrow \infty} r/n = p$  ならば、 $p$  を事象  $A$  の起こる (統計的) 確率という。

#### 2. 数学的確率

ある試行の全事象  $U$  に含まれる要素の数を  $n(U)$  とし、これらは同等に起こるものとする。事象  $A$  に含まれる要素の数を  $n(A)$  とするとき、事象  $A$  の起こる (数学的) 確率  $p'$  を  $p' = n(A)/n(U)$  と定義する。

#### 3. 統計的確率と数学的確率の同等性

$$p = p' \quad \text{大数の法則}$$

注) 確率の値  $0 \leq p \leq 1$

例

さいころの目が奇数となる確率

$$p = \frac{n(A)}{n(U)} = \frac{3}{6} = \frac{1}{2}$$

さいころの目が奇数でかつ 3 以下となる確率

$$p = \frac{n(A \cap B)}{n(U)} = \frac{2}{6} = \frac{1}{3}$$

3 枚の硬貨を同時に投げるとき、1 枚が表である確率

$$p = \frac{3}{2^3} = \frac{3}{8} \quad \text{硬貨は区別できるとして求める}$$

a と b の 2 つのサイコロを振って、a が偶数、b が 2 以下となる確率

$$p = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

このように事象 A と事象 B とが同時に起こる場合、その生起確率が  $P = P_A \times P_B$  で与えられる場合、2 つの事象は互いに独立であるという。

## 2.3 期待値

例

じゃんけんで勝ったら 1000 円もらい、負けたら 2000 円支払う。自分がもらうお金の期待値を求めよ。あなたはこの賭けをやりますか？

$$E(X) = \frac{1}{2} \times 1000 + \frac{1}{2} \times (-2000) = -500 \text{ 円}$$

$X$ : 1000, -700, -200 の値を確率的に持つ変数 (確率変数)

問題

ある企業に 100 万円投資すると、確率 0.7 で 140 万円の収入があり、確率 0.2 で 50 万円の収入となり、確率 0.1 で収入がなくなる。このような投資は有利と考えるか？

解答

$$E(X) = 0.7 \times 140 + 0.2 \times 50 + 0.1 \times 0 = 108 > 100$$

期待値が投資額を上回っているため、不利な投資ではないと思われる。

公式

一般に、確率変数  $X$  の実現値  $x_i$  の生起確率を  $p_i$  として、期待値  $E(X)$  は以下となる。

$$E(X) = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n = \sum_{i=1}^n p_i x_i$$

### 3章 データと集計

#### 3.1 データの種類

量的データ	数値の大きさに意味のあるデータ 例 身長, 体重, 企業の売上高等	値が重要
質的データ	カテゴリとして分類されるデータ 例 アンケートの選択問題 支持する 支持しない 階級分割 分からない	割合が重要

注) 量的データでも階級に分けると質的データとなる。

#### 3.2 データのグラフ表現

数を比較する	棒グラフ, 積み重ね棒グラフ
割合を比較する	円グラフ
値の変化を見る	折れ線グラフ
2つの量の関係を見る	散布図

#### 3.3 データの集計方法

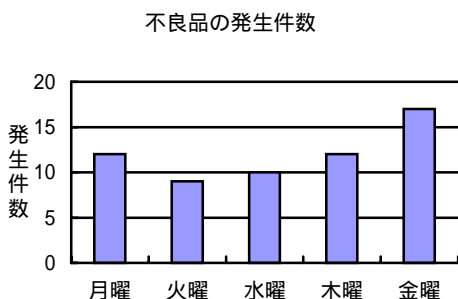
##### 1. 質的データ

###### 単純集計

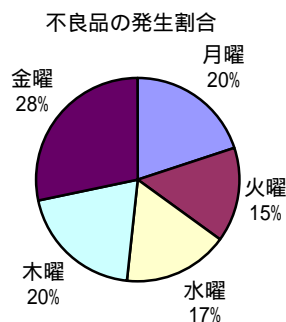
###### 1次元分割表

曜日	月曜	火曜	水曜	木曜	金曜	合計
発生件数	12	9	10	12	17	60

棒グラフ (数を比較する)



円グラフ (割合を比較する)

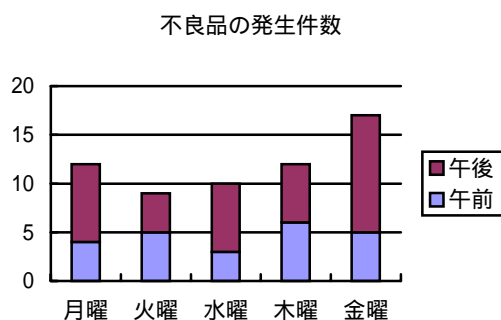


## クロス集計

### 2次元分割表

曜日	月曜	火曜	水曜	木曜	金曜
午前	4	5	3	6	5
午後	8	4	7	6	12

### 積み重ね棒グラフ



## 問題

20人にアンケートを取ったところ、以下のような結果が得られた。以下の問いに答えよ。

性別	回答	性別	回答	性別	回答	性別	回答
1	1	2	3	1	2	2	1
2	1	1	1	1	2	1	2
1	2	2	2	2	1	1	1
1	2	2	3	2	1	2	3
2	1	1	1	1	3	1	1

注) 性別 1:男 2:女

回答 1:はい 2:いいえ 3:どちらとも言えない

- 1) 回答に関する1次元分割表を描け。
- 2) 性別と回答に関する2次元分割表を描け。
- 3) 1)の分割表を用いて棒グラフと円グラフを描け。
- 4) 2)の分割表を用いて積み重ね棒グラフを描け。

## 2. 量的データ

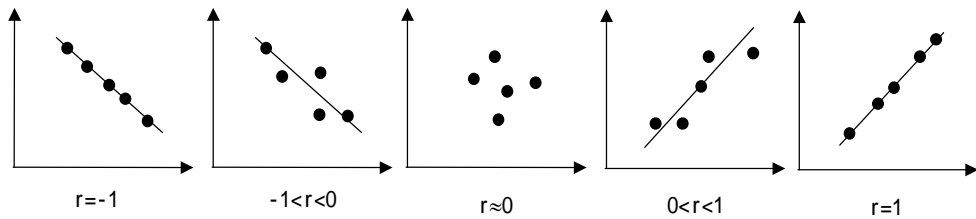
### 単純集計・群別集計

度数分布表，ヒストグラム，代表値（平均，分散，標準偏差等）

後述

### クロス集計

散布図（分布図，相関図），相関係数

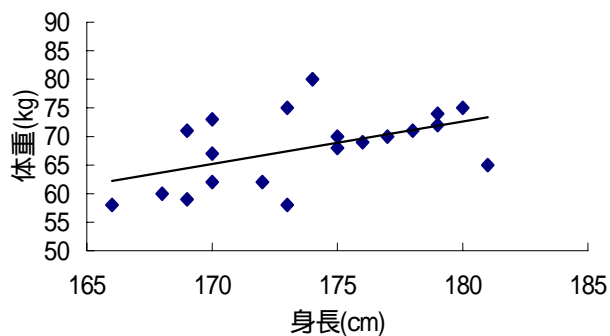


例

以下のデータから、散布図と相関係数を求めよ。

身長(cm)	169	175	170	179	176	174	173	181	179	178
体重(kg)	71	68	67	72	69	80	75	65	74	71
身長(cm)	170	180	177	175	172	166	168	173	169	170
体重(kg)	62	75	70	70	62	58	60	58	59	73

身長と体重の相関



$$\begin{aligned} \text{相関係数} &= \text{correl}(\text{範囲 1}, \text{範囲 2}) \\ &= 0.513047 \end{aligned}$$



## 4章 度数分布

### 4.1 度数分布表

例 20人の学生の試験を調べたら、以下の成績であった。データの分布を調べよ。

74, 68, 70, 69, 54, 63, 81, 72, 78, 65, 57, 65, 74, 67, 75, 54, 84, 51, 65, 84

度数分布表

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

階級の幅：10

階級値： $60 < x \leq 70$  の場合  $(60+70)/2=65$  階級の中央値

### 4.2 ヒストグラム (histogram)

例 上の度数分布表をヒストグラムに表わせ。

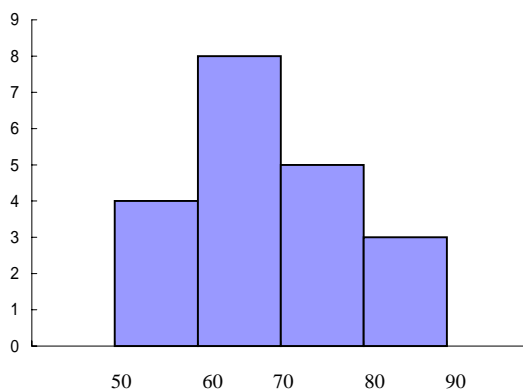


図 試験結果のヒストグラム

注意

1. ヒストグラムは面積グラフである。(例えば、175-185を1つの階級にする)
2. ヒストグラムの階級数は大体 $\sqrt{n}$  ( $n$ はデータ数)にすると見易い。

## 問題

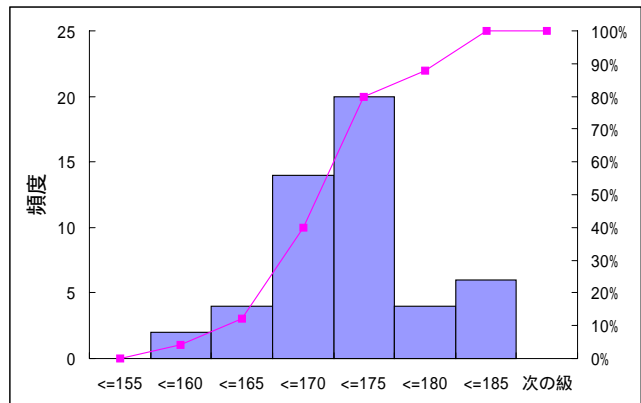
Excel の分析ツールを用いて、以下の 50 人の身長データ(cm)で度数分布表を作り、ヒストグラムを描け。

184.9, 170.6, 164.7, 165.3, 165.1, 170.5, 171.2, 159.8, 167.2, 170.2,  
165.9, 170.3, 171.9, 183.7, 158.5, 169.8, 181.5, 172.2, 171.7, 164.5,  
166.0, 171.1, 178.5, 173.6, 180.4, 165.1, 169.4, 172.4, 174.2, 164.6,  
176.9, 180.6, 170.4, 178.7, 166.2, 172.5, 172.6, 166.2, 170.2, 170.2,  
165.0, 175.3, 165.6, 174.8, 169.7, 169.3, 169.6, 174.0, 180.5, 172.2

## 解答

Excel 使用例

データ区間	頻度	累積 %
<=155	0	.00%
<=160	2	4.00%
<=165	4	12.00%
<=170	14	40.00%
<=175	20	80.00%
<=180	4	88.00%
<=185	6	100.00%
次の級	0	100.00%



## 5章 分布の代表値

### 5.1 分布の中心を表わす代表値

データ：  $x_1, x_2, \dots, x_n$

165, 168, 170, 170, 172, 173, 176, 177, 180

#### 1) 平均値 (average, mean)

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$$

$$= \frac{1}{n} \sum_{i=1}^n x_i$$

平均値 = average(範囲)

#### 2) 中央値 (中間値, メジアン median)

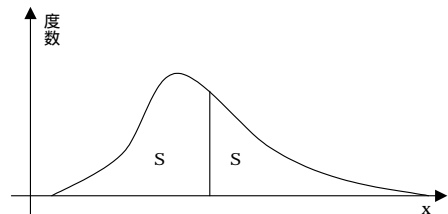
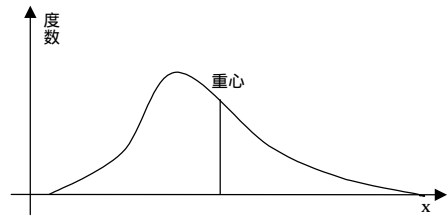
データを小さい方から大きい方に並べて、中間の値

例 以下のデータの中央値を求める。

2, 3, 3, 5, 6, 9, 13

2, 3, 3, 5, 6, 9, 11, 13

中央値 = median(範囲)

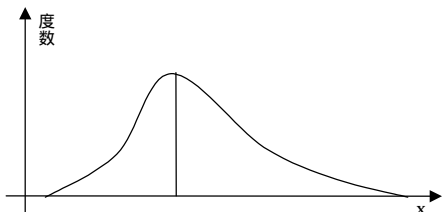


特に、度数分布表が与えられている場合

#### 3) 最頻値 (モード mode)

度数分布表において最も頻度の高い値

最頻値 = mode(範囲) : 使わない方がよい



### 5.2 分布の広がりを表わす代表値

#### 1) レンジ (range)

$R = \text{最大値} - \text{最小値}$       レンジ = max(範囲) - min(範囲)

#### 2) 分散 (variance)      意味の説明と式の変形

$$s^2 = \frac{1}{n} [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2]$$

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

分散 = varp(範囲)

### 3) 不偏分散

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

不偏分散 = var(範囲) 分散から求めてもよい

### 4) 標準偏差 (standard deviation)

$$s = \sqrt{\text{分散}} \quad \text{または} \quad u = \sqrt{\text{不偏分散}}$$

標準偏差 = stdevp(範囲) または =stdev(範囲)

### 問題

Excel の関数を使って、以下のデータの平均値、中央値、レンジ、分散、不偏分散、それぞれの標準偏差を求めよ。

身長(cm) 171, 181, 172, 166, 172, 175, 168, 174, 171, 170

### 解答

平均値	172	中央値	171.5	レンジ	15
分散	15.2	標準偏差	3.898718		
不偏分散	16.88889	標準偏差	4.109609		

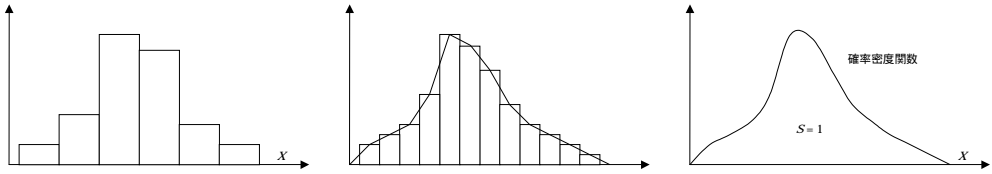
### 重要

- 1) 平均は分布の中心である。
- 2) 分散は分布のばらつきである。
- 3) 標準偏差は分布の拡がりである。
- 4) 分散は標準偏差の2乗である。

## 6章 正規分布 (normal distribution)

### 6.1 連続的データの確率分布

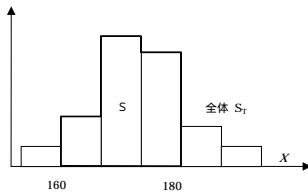
データ数を十分多く取ったヒストグラムの上端をつなぎ、全体の面積が1になるように、目盛りを付けたものを確率密度関数と呼ぶ。



$y = f(X)$  : 確率密度関数  
 $X$  : 確率変数

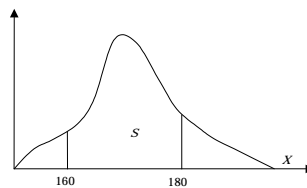
確率密度関数の意味

ヒストグラム

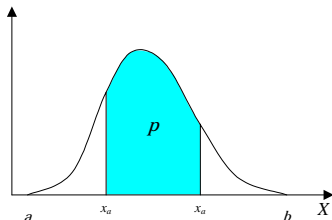


割合 =  $S/S_T$

確率密度関数



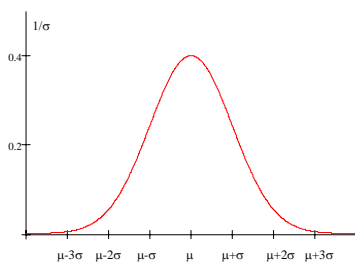
割合 =  $S$



確率  $p = P(x_a \leq X \leq x_b)$   
 $X$  が  $x_a \leq X \leq x_b$  となる確率  
 全確率  $P(a \leq X \leq b) = 1$

### 6.2 平均 $\mu$ 分散 $\sigma^2$ の正規分布 : $N(\mu, \sigma^2)$ 分布

正規分布とは偶発的なデータのゆらぎによる分布 (量的データの基本となる分布)

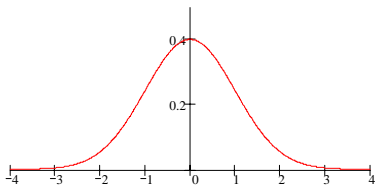


$N(\mu, \sigma^2)$  分布の密度関数

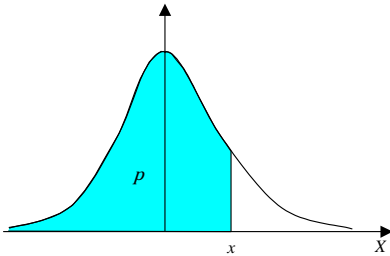
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{注) } f(0) = \frac{0.3989\dots}{\sigma}$$

### 6.3 標準正規分布(standard normal distribution) [ 平均 0 分散 1 の正規分布 : $N(0,1)$ 分布 ]

標準正規分布の密度関数



$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{注) } f(0) = 0.3989\dots$$



値と確率

Excel  $p = \text{normsdist}(x)$

$x = \text{normsinv}(p)$

#### 問題

標準正規分布に対して以下の確率を求めよ。

- 1 )  $P(X \leq 2) =$
- 2 )  $P(X \geq 2) =$
- 3 )  $P(X \geq 1) =$
- 4 )  $P(X \leq -1, X \geq 1) =$
- 5 )  $P(-1 \leq X \leq 2) =$

## 6.4 正規分布の性質

### 1. 標準正規分布への変換

$X \sim N(\mu, \sigma^2)$  分布 のとき、

$$X' = X + d \sim N(\mu + d, \sigma^2) \text{ 分布}$$

$$X' = cX \sim N(c\mu, c^2\sigma^2) \text{ 分布}$$

変換後も正規分布になる

まとめて

$$X' = cX + d \sim N(c\mu + d, c^2\sigma^2) \text{ 分布}$$

標準正規分布へ

$$X' = \frac{X - \mu}{\sigma} \sim N(0, 1) \text{ 分布}$$

例

$X \sim N(170, 100)$  分布のとき、以下の分布を求める。

1 )  $X' = X + 10 \sim N(180, 100)$  分布

2 )  $X' = 2X \sim N(340, 400)$  分布

3 )  $X' = \frac{X - 170}{10} \sim N(0, 1)$  分布

### 2. 正規分布の合成

$X_1 \sim N(\mu_1, \sigma_1^2)$  分布 ,  $X_2 \sim N(\mu_2, \sigma_2^2)$  分布 のとき、

$$X = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \text{ 分布}$$

変換後も正規分布になる

互いに独立な確率変数  $X_1$  ,  $X_2$  が、 $X_1 \sim N(10, 9)$  分布 ,  $X_2 \sim N(7, 16)$  分布であるとするとき、以下の確率変数  $X$  の分布を求めよ。

1 )  $X = X_1 + X_2 \sim N(10, 25)$  分布

2 )  $X = 2X_1 + X_2 \sim N(27, 52)$  分布

3 )  $X = X_1 - X_2 \sim N(3, 25)$  分布

問題

$X \sim N(170, 100)$  分布のとき、以下の値は、標準正規分布ではいくつに相当するか。

- 1)  $X = 170$
- 2)  $X = 180$
- 3)  $X = 160$

問題

平均 25.5、標準偏差 3.2 のとき、以下の確率を求めよ。

- 1)  $P(X \leq 20) =$
- 2)  $P(X \geq 30) =$

問題

ある商品の製造は3つの工程からなり、各工程に要する日数は、以下のような正規分布に従うとする。

	平均	標準偏差
第1工程	3	1
第2工程	10	3
第3工程	5	2

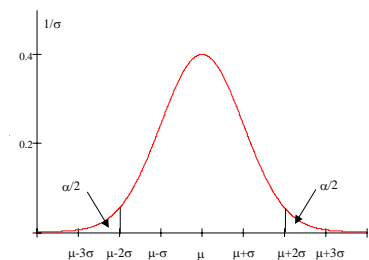
- 1) 完成までに要する時間の平均と標準偏差を求めよ。
- 2) 納期を20日とすると、納期に遅れる確率を求めよ。

解答

- 1)
- 2)



### 3. 確率の概数



平均  $\mu$ 、分散  $\sigma^2$  が何であっても、

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.683 \quad \alpha = 0.317 (32\%)$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.954 \quad \alpha = 0.046 (5\%)$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.997 \quad \alpha = 0.003 (0.3\%)$$

特に、平均から  $\pm 2\sigma$  の外側に 5% !

#### 問題

ある集団の身長分布は、平均 170cm、標準偏差 10cm の正規分布であった。以下の確率の概数を求めよ。

1 )  $P(160 \leq X \leq 180) =$

2 )  $P(150 \leq X \leq 190) =$

3 )  $P(X \geq 190) =$

### 4. 偏差値について

点数の平均  $\bar{x}$ 、標準偏差  $s$  のとき、 $x$  の偏差値は以下で与えられる。

$$\text{偏差値} = 50 + 10 \times \frac{x - \bar{x}}{s}$$

例えば、偏差値 60 の人の上には 16%、70 の人の上には 2.5% 程度の人がいる

#### 問題

1000 人が受験した試験の成績分布は、平均 60 点、標準偏差 15 点の正規分布であった。A, B, C 君の点数が 75 点、90 点 45 点であるとき以下の問いに答えよ。

1 ) A 君の偏差値を求めよ。

2 ) B 君の偏差値を求めよ。

3 ) C 君の偏差値を求めよ。

4 ) B 君の順位はおよそ何番か。

5 ) C 君の順位はおよそ何番か。

### 5. 中心極限定理

よく利用される公式

独立な確率変数、 $X_1, X_2, \dots, X_n$  が、平均  $\mu$ 、分散  $\sigma^2$  の確率分布に従うとき、

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \underset{n \rightarrow \infty}{\sim} N(\mu, \sigma^2/n) \text{ 分布}$$

分布によらない関係

平均を取ることの意義

例

独立な確率変数  $X_i$  が、平均 100、分散 10 のとき、以下の変数の分布を求めよ。

$$1) \bar{X} = \frac{1}{5}(X_1 + X_2 + \cdots + X_5)$$

$$2) \bar{X} = \frac{1}{10}(X_1 + X_2 + \cdots + X_{10})$$

## 8章 標準正規分布から導かれる分布

### 8.1 $\chi^2$ 分布

$X_i \sim N(0, 1)$  分布で独立なとき、 $\chi^2 = \sum_{i=1}^n X_i^2 \sim \chi_n^2$  分布 (自由度  $n$  の  $\chi^2$  分布)

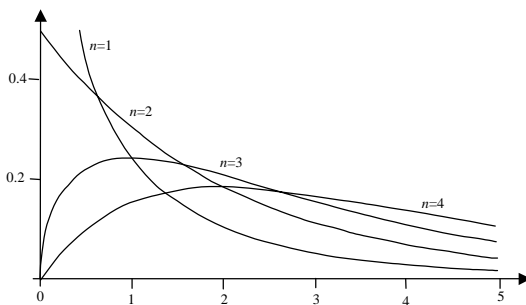
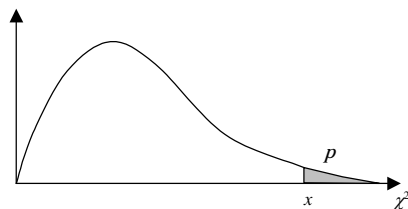


図 自由度 1,2,3,4 の  $\chi^2$  分布の確率密度関数



$$p = \text{chidist}(x, n)$$

$$x = \text{chiinv}(p, n)$$

例

自由度 5 ,  $\chi^2$  値 10 のときの上側確率

$$\text{chidist}(10, 5) = 0.075235$$

自由度 10 , 上側確率 0.05 のときの  $\chi^2$  値

$$\text{chiinv}(0.05, 10) = 18.30703$$

### 8.2 F 分布

$\chi_1^2 \sim \chi_{n_1}^2$  分布 ,  $\chi_2^2 \sim \chi_{n_2}^2$  分布で独立なとき、

$$F = \frac{\chi_1^2/n_1}{\chi_2^2/n_2} \sim F_{n_1, n_2} \text{ 分布 (自由度 } n_1, n_2 \text{ の F 分布)}$$

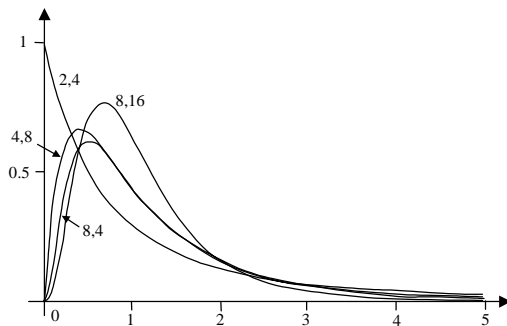
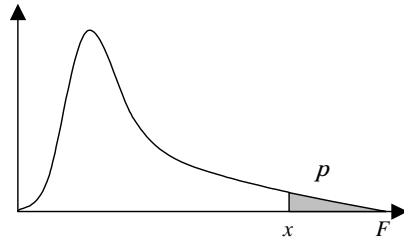


図 自由度(2,4), (4,8), (8,16), (8,4)の F 分布の確率密度関数



$$p = fdist(x, n_1, n_2)$$

$$x = finv(p, n_1, n_2)$$

例

自由度(8, 4), F 値 10 のときの上側確率  $fdist(10, 8, 4) = 0.020592$

自由度(10, 5), 上側確率 0.05 のときの F 値  $finv(0.05, 10, 5) = 4.73505$

### 8.3 t 分布

$X \sim N(0,1)$  分布,  $\chi^2 \sim \chi_n^2$  分布で独立なとき、

$$t = \frac{X}{\sqrt{\chi^2/n}} \sim t_n \text{ 分布 (自由度 } n \text{ の } t \text{ 分布)} \quad \text{注) } t^2 \sim F_{1,n} \text{ 分布}$$

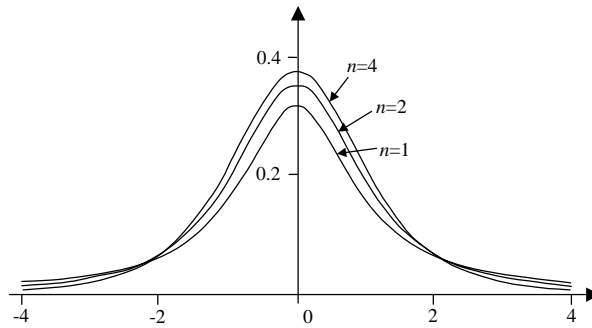
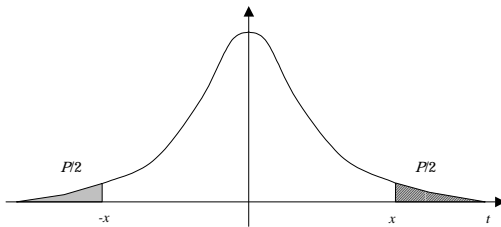


図 自由度 1, 2, 4 の t 分布の確率密度関数

注)  $n \rightarrow \infty$  で  $N(0,1)$  分布



上側確率  $p/2 = tdist(x, n, 1)$

両側確率  $p = tdist(x, n, 2)$

$x = tinv(p, n)$

例

自由度 10, t 値 2 のときの上側確率  $p/2$   $tdist(2, 10, 1) = 0.036694$

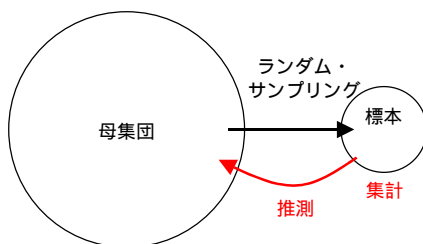
自由度 10, t 値 2 のときの両側確率  $p$   $tdist(2, 10, 2) = 0.073388$

自由度 10, 両側確率 0.05 のときの t 値  $tinv(0.05, 10) = 2.228139$

注) 片側確率は両側確率の 1/2 で指定する。

## 9章 検定の基礎

### 9.1 母集団と標本



母集団：日本人全体・日本の中小企業等（事実上全数調査不能）

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。

### 9.2 検定とは

例 超能力を持つという人にコインの裏表を当てる実験をしてもらい、100回の試行で70%の正解率を得た。この人には本当に超能力があると考えられるか？

有意水準を5%として判定せよ。20回の試行ではどうか。

有意水準（危険率）：超能力があると判定して間違える確率

70%の正解率は確かに超能力があつて起こったものか、偶然に起こったものか、判定する。

試行回数に依存する。

答  $\chi^2$  検定を用いる。

試行回数 100 回

$$\chi^2 = \frac{(70-50)^2}{50} + \frac{(30-50)^2}{50} = 2 \times \frac{400}{50} = 16 \quad (\sim \chi_1^2 \text{ 分布})$$

$p = \text{chidist}(16,1) = 0.0000633 < 0.05$  より、超能力があるといえる。

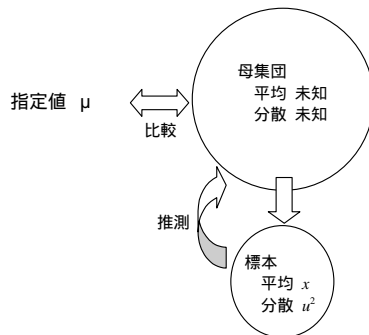
試行回数 20 回

$$\chi^2 = \frac{(14-10)^2}{10} + \frac{(6-10)^2}{10} = 2 \times \frac{16}{10} = 3.2$$

$p = \text{chidist}(3.2,1) = 0.073638 > 0.05$  より、超能力があるといえない。

### 9.3 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。



例

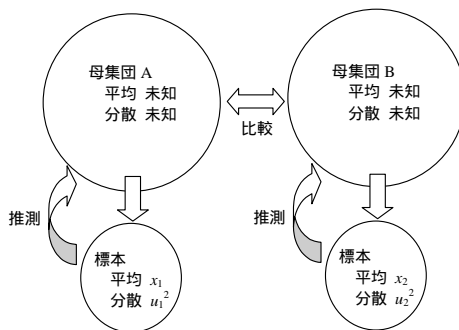
量的データの比較：標本調査世帯と全国平均との所得の比較

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

この標本はこの母集団から抽出されたものかどうかを判定する。

または、ある設定値と標本を比較する問題ととらえてもよい。

2) いくつかの母集団のある指標を比較する。



例

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）

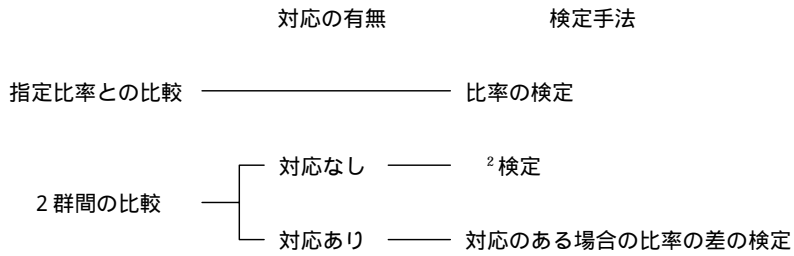
標本店における宣伝前後の売り上げ比較（対応がある場合）

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）

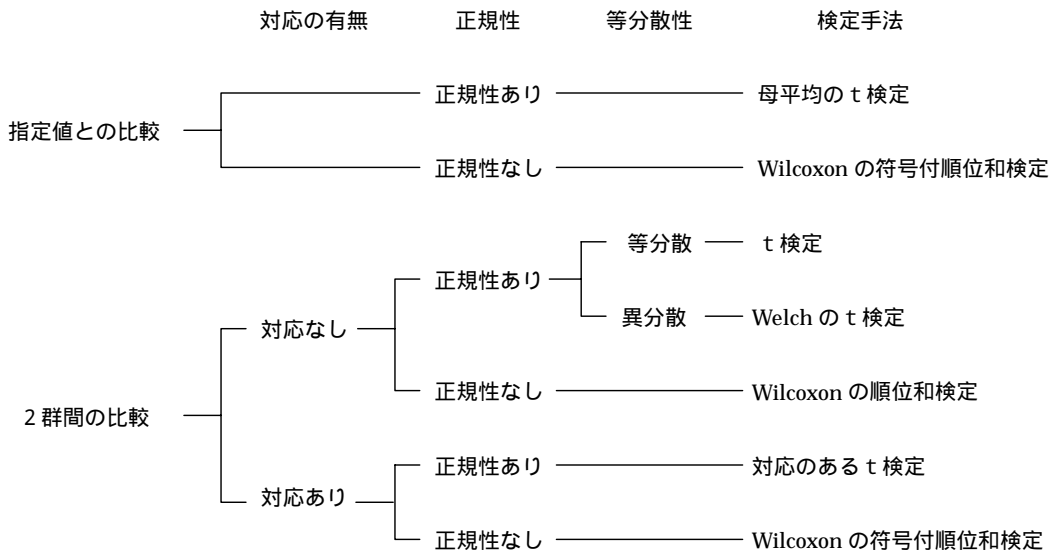
標本店における従業員教育前後の評判の変化（対応がある場合）

## 9.4 検定選択ツリー

### 質的データ



### 量的データ



これらの検定手法について今後詳細に見て行く。

## 旧6章 確率分布

### 6.1 離散的データの確率分布

確率変数  $X$

変数  $x_1, x_2, \dots, x_n$  ( $n \rightarrow \infty$ の場合もある。)

確率  $p_1, p_2, \dots, p_n$

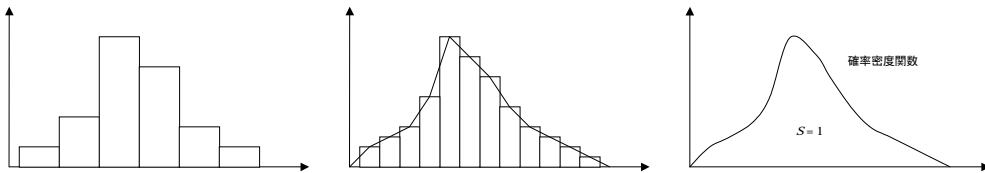
1) 確率変数の平均  $E(X) = \sum_{i=1}^n x_i p_i = \mu$

2) 確率変数の分散  $V(X) = E((X - \mu)^2) = \sum_{i=1}^n p_i (x_i - \mu)^2 = \sum_{i=1}^n p_i x_i^2 - \mu^2$

サイコロの目の平均と分散を求めよ。

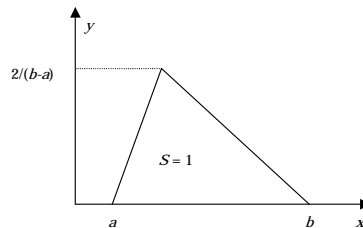
### 6.2 連続的データの確率分布

データ数を十分多く取ったヒストグラムの上端をつなぎ、全体の面積が1になるように、目盛りを付けたものを確率密度関数と呼ぶ。

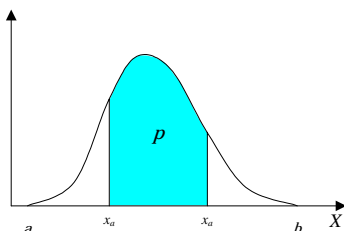


$y = f(x)$  : 確率密度関数

例 密度関数の目盛の付け方



確率密度関数の意味



確率  $P(x_a \leq X \leq x_b) = S = \int_{x_a}^{x_b} f(x) dx$

全確率  $P(a \leq X \leq b) = \int_a^b f(x) dx = 1$



## 6.3 確率変数の平均と分散の一般的性質

### 1. 線形変換

確率変数  $X$ ,  $E(X) = \mu$ ,  $V(X) = \sigma^2$

$X' = X + d$  について、

$$E(X') = \mu + d$$

$$V(X') = \sigma^2$$

$X' = cX$  について、

$$E(X') = c\mu$$

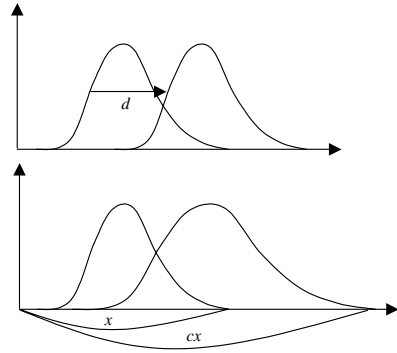
$$V(X') = c^2\sigma^2$$

$X' = cX + d$  について、

$$E(X') = c\mu + d$$

$$V(X') = c^2\sigma^2$$

のとき、



離散の場合の証明

$$\begin{aligned} E(X') &= \sum_{i=1}^n (cx_i + d)p_i = c \sum_{i=1}^n x_i p_i + d \sum_{i=1}^n p_i \\ &= cE(X) + d = c\mu + d \end{aligned}$$

$$\begin{aligned} V(X') &= \sum_{i=1}^n (cx_i + d - c\mu - d)^2 p_i \\ &= c^2 \sum_{i=1}^n (x_i - \mu)^2 p_i = c^2 V(X) = c^2 \sigma^2 \end{aligned}$$

### 2. 確率変数の合成

互いに独立な確率変数  $X_1, X_2$

$$E(X_1) = \mu_1, E(X_2) = \mu_2,$$

$$V(X_1) = \sigma_1^2, V(X_2) = \sigma_2^2$$

のとき、

$X = X_1 + X_2$  について、

$$E(X) = \mu_1 + \mu_2$$

$$V(X) = \sigma_1^2 + \sigma_2^2$$

問題

上記確率変数の合成で、 $X = c_1 X_1 + c_2 X_2$  の場合、平均と分散はどうなるか。

解答

$$E(c_i X_i) = c_i \mu_i, V(c_i X_i) = c_i^2 \sigma_i^2 \quad (i=1, 2) \text{ より、}$$

$$E(X) = c_1\mu_1 + c_2\mu_2, \quad V(X) = c_1^2\sigma_1^2 + c_2^2\sigma_2^2$$

### 問題

平均  $\mu$  , 分散  $\sigma^2$  の互いに独立な確率変数  $X_i$  ( $i = 1, 2, \dots, n$ ) を使って、標本平均に相当する新しい確率変数  $X = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$  を作った。新しい確率変数の平均と分散を求めよ。

### 解答

$$\text{平均 } \mu \quad \text{分散 } \frac{\sigma^2}{n}$$