数理統計学
(第2回)

機械工学科
塩幡　宏規

第1章　データの整理

1.1 度数分布

✓変量とは

ある特性を数量で表したもの

連続変量：身長，体重，温度

連続的な数値

離散変量：人数，個数，事故件数

計数として数えられる

✓度数分布表

変量をいくつかの区間に整理・分類し表にする

効果：集団に関する特性を総合的に概観的に把握できる
　　：平均値，標準偏差の計算が容易

階級(または級)：区間，度数：階級に属するデータの個数
階級値：階級の中央値

✓度数分布表の作り方

手順１　データの中の最大値と最小値を求める
手順２　階級の数kを求める
手順３　階級の幅hを求める
手順４　階級の境界値cを求める
手順５　階級値x_iを求める
手順６　各階級に入るデータの度数を数え、表1.3を作る
手順７　階級値を横軸、度数を縦軸としてヒストグラムを作成する

•階級の数kの目安

データの数　N	50未満	50～100	100～250	250以上
階級の数　k	5～7	6～10	7～12	10～20

•階級の幅h

＝(最大値－最小値)/階級の数

•階級の境界値c

下側の境界値$c_{l,n-1}$＝最小値－測定単位/2（n=1）
上側の境界値$c_{u,n}$＝下側の境界値+階級の幅

下側の境界値$c_{l,n-1}$ ＝ $c_{u,n-1}$（n=2，・・・，k）
上側の境界値$c_{u,n}$＝ $c_{u,n}$+階級の幅

•階級値x_n

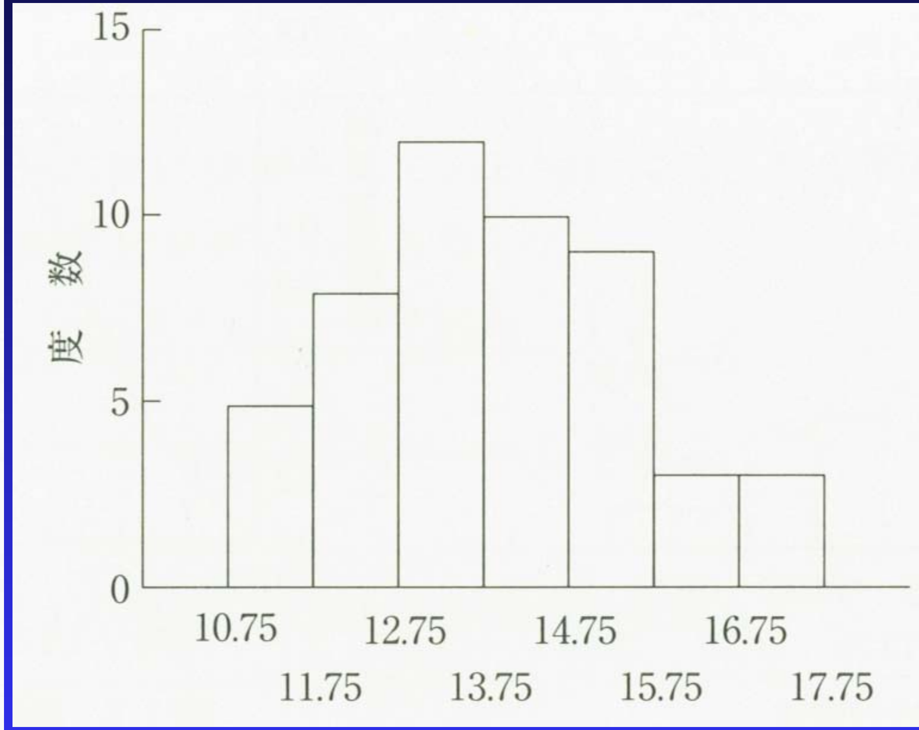
x_n＝（$c_{l,n-1}$+ $c_{u,n}$）/2 （n=1，・・・，k）

•例1.1

3歳児50人の体重

階　級 c	階級値 x	度数 f
10.75 ～ 11.75	11.25	5
11.75 ～ 12.75	12.25	8
12.75 ～ 13.75	13.25	12
13.75 ～ 14.75	14.25	10
14.75 ～ 15.75	15.25	9
15.75 ～ 16.75	16.25	3
16.75 ～ 17.75	17.25	3
計	—	50

（度数分布表）



（ヒストグラム）

✓相対度数

各階級iの度数f_i/全度数N　$\Rightarrow$　総和は1になる

✓累積度数

対象とする階級iまでの全度数Σf_i

✓累積相対度数

累積度数Σf_i/全度数N

表1.4　相対度数と累積相対度数

階級 （以上～未満）			階級値 x	度数 f	相対度数 f/N	累積度数 Σf	累積相対度数 $\Sigma f/N$
10.75	～	11.75	11.25	5	0.10	5	0.10
11.75	～	12.75	12.25	8	0.16	13	0.26
12.75	～	13.75	13.25	12	0.24	25	0.50
13.75	～	14.75	14.25	10	0.20	35	0.70
14.75	～	15.75	15.25	9	0.18	44	0.88
15.75	～	16.75	16.25	3	0.06	47	0.94
16.75	～	17.75	17.25	3	0.06	50	1.00
計			—	50	1.00	—	—

3歳児50人の体重

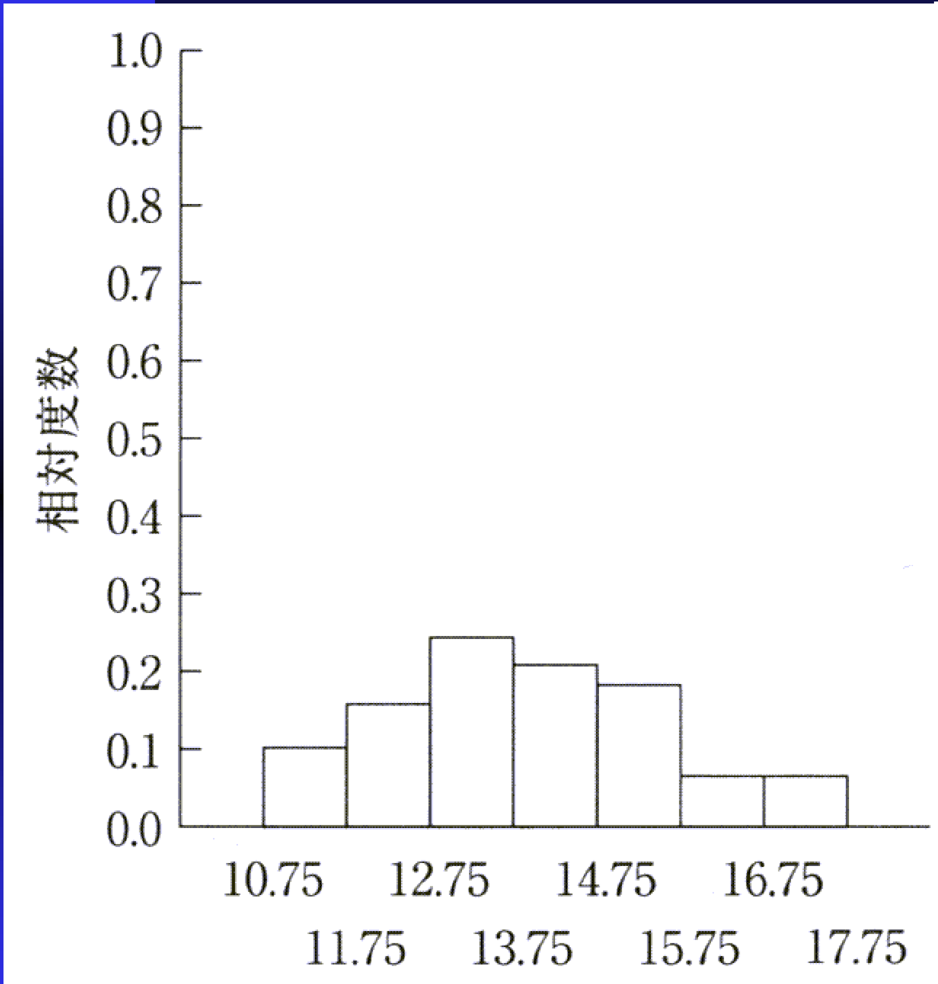


図1.2　相対度数

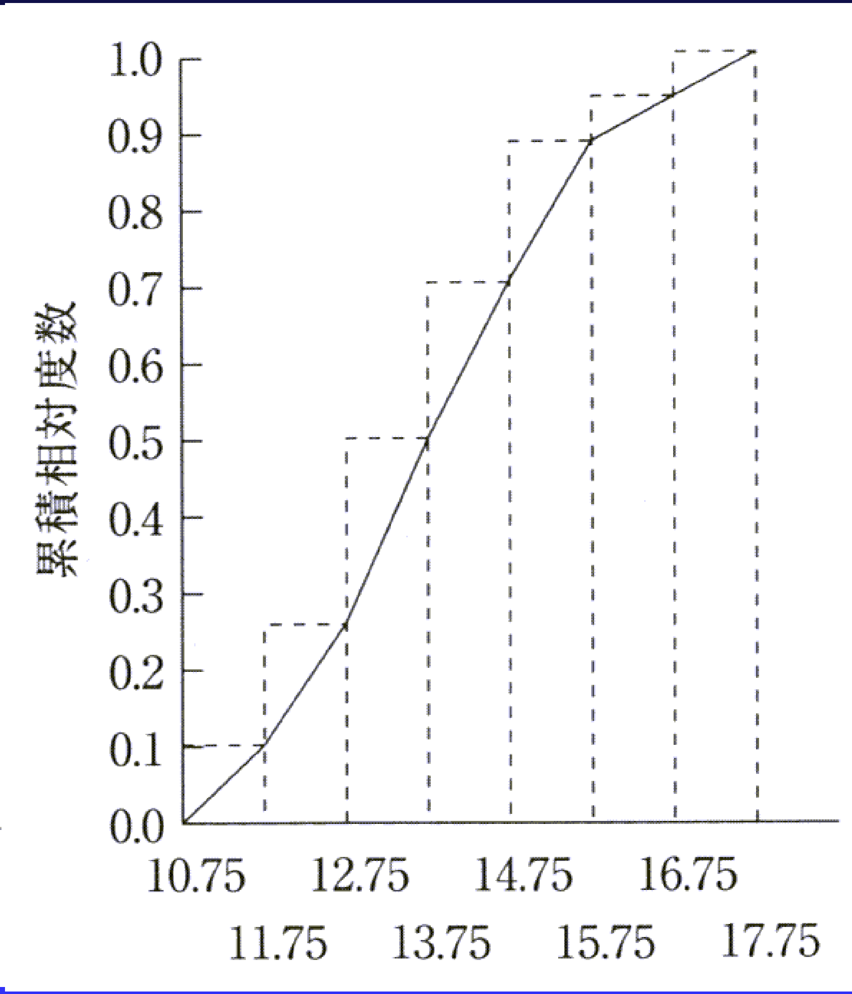


図1.3　累積相対度数

ヒストグラムの例

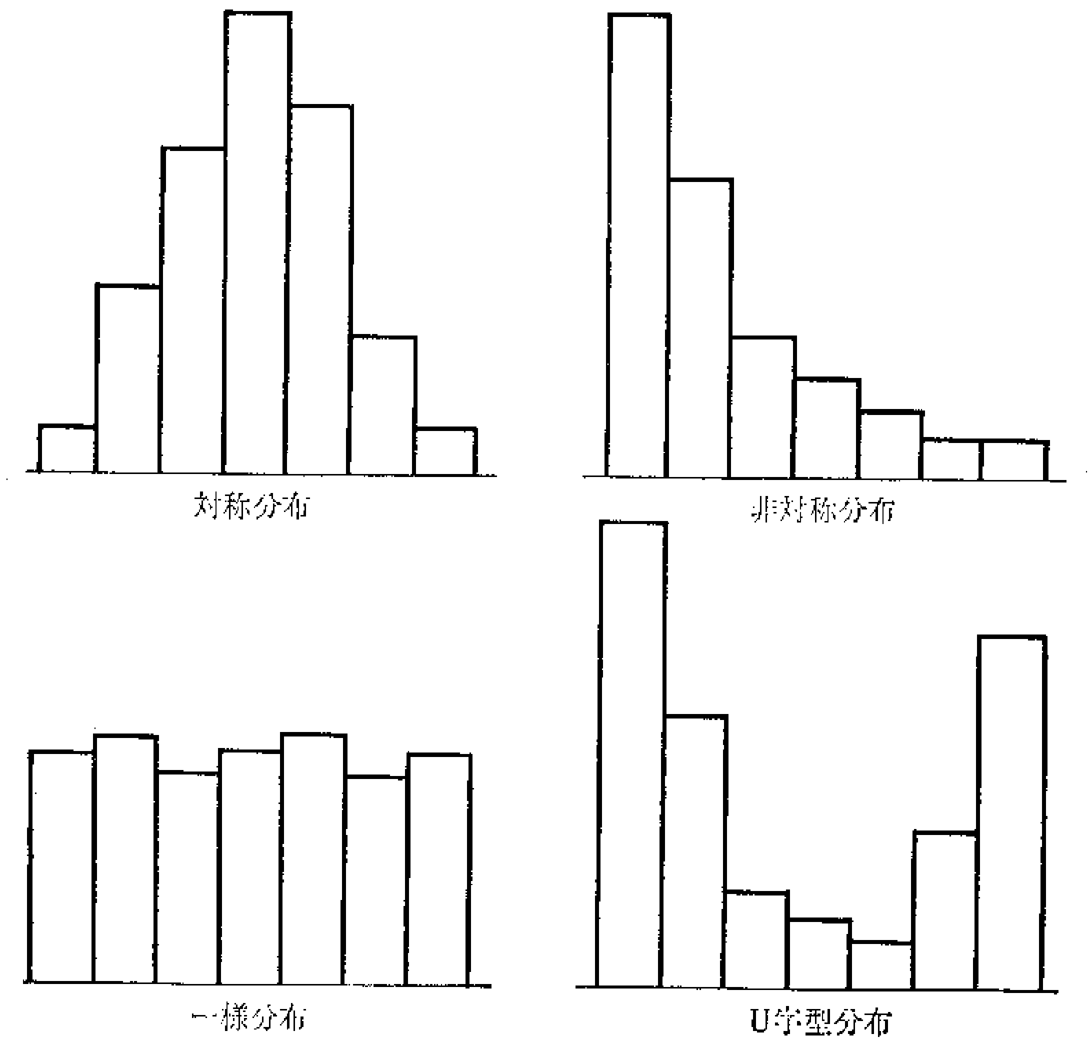


図 1.5　いろいろな分布

度数分布やヒストグラム

↓

全体の分布状態を知ることができる

全体の特性を1つの値で知ることができないか

1.2 代表値と散布度

•代表値:分布の中心的な値
　平均値，中央値，最頻値
•散布度:分布の代表値の周りのばらつき度合いを示す値
　分散，標準偏差

✓代表値

➢平均値:算術平均

➢中央値:変量を大きさ順に並べたときの中央の値(中位数ともいう)偶数個と奇数個で算出法が異なる.

➢最頻値:変量のうち最も高い度数で現れる値(モード，並数ともいう)

分布の中心的な位置を示す

➢平均値

- 算術平均

平均値：

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{N} = \frac{1}{N}\sum_{i=1}^{n} x_i$$

度数分布表に整理されている場合

平均値：

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_n f_n}{N} = \frac{1}{N}\sum_{i=1}^{n} x_i f_i$$

$$f_1 + f_2 + \cdots + f_n = \sum_{i=1}^{n} f_i = N$$

算術平均の例　：3歳児50人の体重

階級（以上～未満）			階級値x	度数f	$x_i f_i$
10.75	～	11.75	11.25	5	56.25
11.75	～	12.75	12.25	8	98.00
12.75	～	13.75	13.25	12	159.00
13.75	～	14.75	14.25	10	142.50
14.75	～	15.75	15.25	9	137.25
15.75	～	16.75	16.25	3	48.75
16.75	～	17.75	17.25	3	51.75
計			－	50	693.50

平均値(ヒストグラムから）:

$$\bar{x} = \frac{693.50}{50} = 13.87$$

平均値(実際の値から）＝13.77

- 中央値（Me）

度数総数Nが奇数の場合

$$Me = x_{(N+1)/2}$$

$x_1 \leq x_2 \leq \cdots \leq x_N$ のとき

度数総数Nが偶数の場合

$$Me = \frac{1}{2}\left(x_{N/2} + x_{N/2+1}\right)$$

算術平均　:階級値, 度数が大きい場合

階級値の仮平均x_0計算　　→　階級値との差u_i計算

表 1.5

級 (以上～未満)	階級値 x_i	度　数 f_i	u_i	$u_i f_i$
$c_0 \sim c_1$	x_1	f_1	u_1	$u_1 f_1$
$c_1 \sim c_2$	x_2	f_2	u_2	$u_2 f_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$c_{n-1} \sim c_n$	x_n	f_n	u_n	$u_n f_n$
計	—	N	—	$\sum u_i f_i$

$$u_i = \frac{x_i - x_0}{c}$$

c:階級幅

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{n} x_i f_i = \frac{1}{N}\sum_{i=1}^{n}\left(x_0 + cu_i\right) f_i = x_0 + c\cdot\frac{1}{N}\sum_{i=1}^{n} u_i f_i$$

$$= x_0 + c\bar{u}$$

ただし, $\bar{u} = \frac{1}{N}\sum_{i=1}^{n} u_i f_i$

算術平均の例3　:階級値，度数が大きい場合

表 1.6

級（以上～未満）	階級値 x_i	度　数 f_i	u_i	$u_i f_i$
1800～2000	1900	1	−6	−6
2000～2200	2100	1	−5	−5
2200～2400	2300	1	−4	−4
2400～2600	2500	2	−3	−6
2600～2800	2700	4	−2	−8
2800～3000	2900	5	−1	−5
3000～3200	3100	10	0	0
3200～3400	3300	12	1	12
3400～3600	3500	5	2	10
3600～3800	3700	6	3	18
3800～4000	3900	3	4	12
計	—	50	—	18

仮平均x_0＝3100

階級幅c ＝200

$$u_i = \frac{x_i - 3100}{200}$$

$$\overline{u} = \frac{1}{N}\sum_{i=1}^{n} u_i f_i = \frac{1}{50}\times 18 = 0.36$$

$$\therefore \overline{x} = x_0 + c \cdot \overline{u}$$

$$= 3100 + 200 \times 0.36 = 3172$$

中央値の例

（1）５，７，３，８，４，５，６ → ３，４，５，５，６，７，８

N=7（奇数）⇒N=（7+1）/2＝4⇒Me=5

（2）８，２，６，７，５，４，７，９ → ２，４，５，６，７，７，８，９

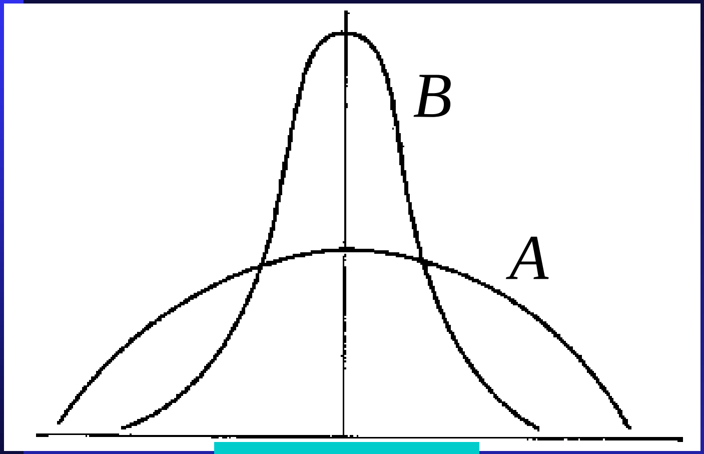
N=8（偶数）⇒N＝4，5⇒Me=（6+7）/2＝6.5

- 最頻値(モード)

: 変量のうち最も高い度数で現れる値(Mo)

　ある商品について金額別に分類したとき最も売れ行きの高い商品の金額が最頻値Moである.

✓散布度



$\bar{x}_A = \bar{x}_B$

代表値(平均値)はその分布の中心を示すだけで，分布の特徴を示せない

- 範囲(レンジ)R:最大値と最小値の差で示す
- 平方和:代表値に中央値を用いるときに使われる
- 分散:標準偏差とともによく用いられる
- 標準偏差:分散の平方根である

- 範囲（レンジ）: R

  R=x_{max}-x_{min}

- 平方和: S

データから

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

度数分布表から

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

➢分散S^2と標準偏差S

分散

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

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

標準偏差

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

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

度数分布表のデータを用いる場合

分散

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

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

標準偏差

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

$$s = \sqrt{\frac{1}{N}\sum_{i=1}^{n}x_i{}^2 f_i - \bar{x}^2}$$

➢不偏分散と標準偏差

分散

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

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

標準偏差

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

➢度数分布表の場合の不偏分散と標準偏差

平方和

$$S=\sum_{i=1}^{n}(x_i-\bar{x})^2 f_i=\sum_{i=1}^{n}x_i^2 f_i-\frac{\left(\sum_{i=1}^{n}x_i f_i\right)^2}{N}\left(\sum_{i=1}^{n}f_i=N\right)$$

分散

$$s^2=\frac{S}{N-1}=\frac{1}{N-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 f_i=\frac{1}{N-1}\sum_{i=1}^{n}x_i{}^2 f_i-\frac{\left(\sum_{i=1}^{n}x_i f_i\right)^2}{N}$$

標準偏差

$$s=\sqrt{\frac{1}{N-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 f_i}=\sqrt{\frac{1}{N-1}\left\{\frac{1}{N-1}\sum_{i=1}^{n}x_i{}^2 f_i-\frac{\left(\sum_{i=1}^{n}x_i f_i\right)^2}{N}\right\}}$$

平均値，分散･標準偏差：階級値，度数が大きい場合

階級値の仮平均x_0を利用

c：階級幅

表 1.9

階級値 x_i	度　数 f_i	u_i	$u_i f_i$	$u_i^2 f_i$
x_1	f_1	u_1	$u_1 f_1$	$u_1^2 f_1$
x_2	f_2	u_2	$u_2 f_2$	$u_2^2 f_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	f_n	u_n	$u_n f_n$	$u_n^2 f_n$
計	N	—	$\sum_{i=1}^{n} u_i f_i$	$\sum_{i=1}^{n} u_i^2 f_i$

$$u_i = \frac{x_i - x_0}{c}$$

$$\bar{x} = x_0 + c\bar{u}$$

$$x_i - \bar{x} = c(u_i - \bar{u})$$

$$s_x{}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot f_i = \frac{1}{n-1}\sum_{i=1}^{n} c^2 (u_i - \bar{u})^2 \cdot f_i$$

$$= c^2 \frac{1}{n-1}\sum_{i=1}^{n}(u_i - \bar{u})^2 \cdot f_i$$

$$= c^2 s_u{}^2 \qquad \therefore s_x = c s_u$$

$$Su = \sum_{i=1}^{n}(u_i - \bar{u})^2 f_i$$

$$= \sum_{i=1}^{n} u_i^2 f_i - \frac{\left(\sum_{i=1}^{n} u_i f_i\right)^2}{N}$$

✓積率，歪度，尖度

➢積率：

$$m_k = \frac{1}{N}\sum_{i=1}^{n} x_i{}^k f_i \quad \left(N = \sum_{i=1}^{n} f_i \right)$$

原点周りのk次の積率

$$M_k = \frac{1}{N}\sum_{i=1}^{n} (x_i - \bar{x})^k f_i$$

平均値周りのk次の積率

$$m_1 = \frac{1}{N}\sum_{i=1}^{n} x_i f_i \quad (= \bar{x})$$

平均

$$M_2 = \frac{1}{N}\sum_{i=1}^{n} (x_i - \bar{x})^2 f_i \quad \left(= s^2\right)$$

分散

平均値周りの2次の積率

$$M_3 = \frac{1}{N}\sum_{i=1}^{n} (x_i - \bar{x})^3 f_i$$

単位に影響される

$M_3>0$：右側が大きい分布

$M_3=0$：対称分布

$M_3<0$：左側が大きい分布

平均値周りの3次の積率

- 歪度:ヒストグラムが対象でない度合いを示す値

$$\alpha_3 = \frac{M_3}{s^3}$$ 歪度

α_3＞0:右に歪んだ分布
α_3 =0:左右対称な分布
α_3 <0:左に歪んだ分布

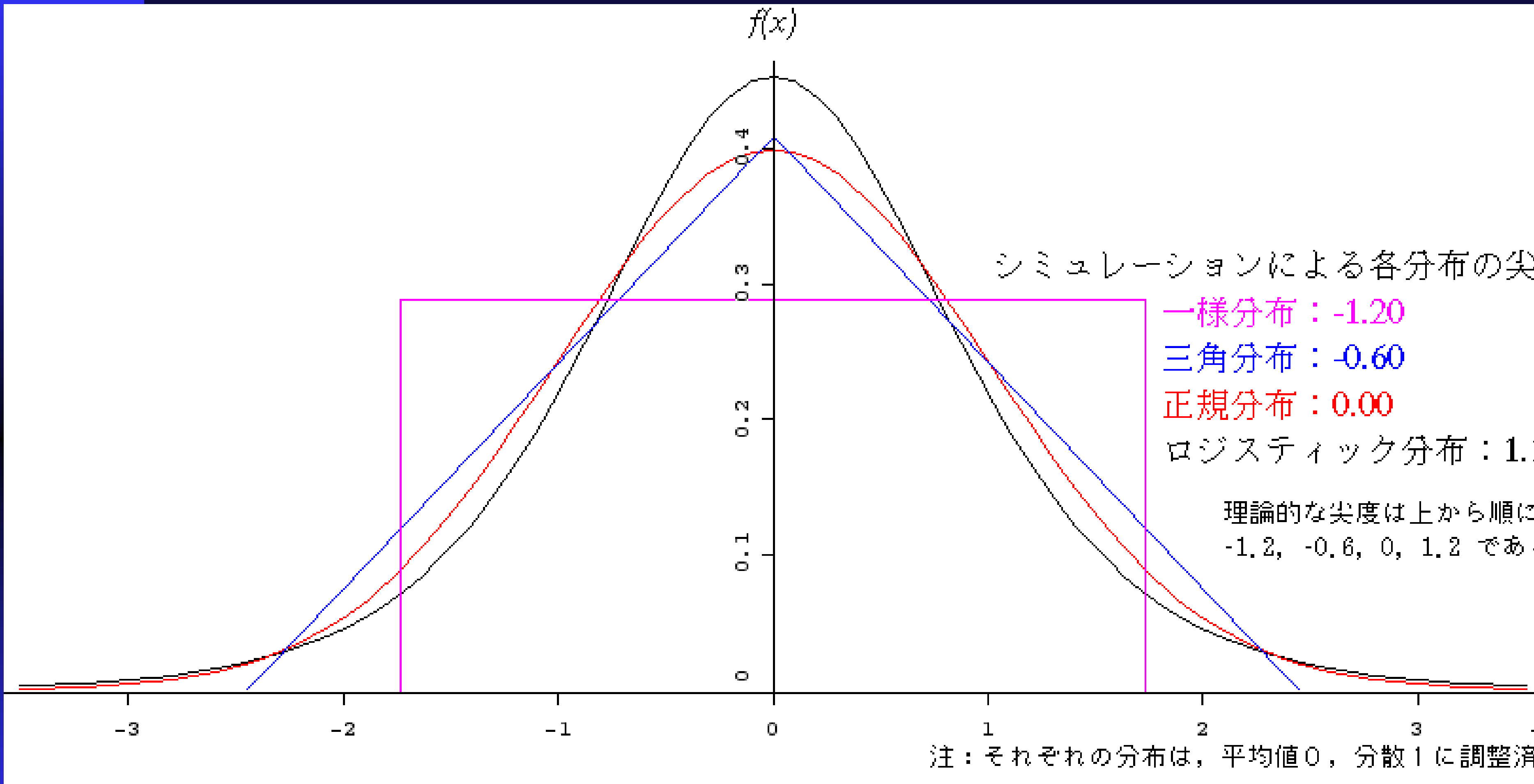
- 尖度:ヒストグラムで尖りの度合いを示す.

$$M_4 = \frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x})^4 f_i$$

平均値周りの4次の積率

$$\alpha_4 = \frac{M_4}{s^4}$$ 尖度

尖度の標準値は3, 3以上の分布は尖りで, 3以下であれば扁平
正規分布の尖度は3


f(x)
0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
シミュレーションによる各分布の尖
一様分布：-1.20
三角分布：-0.60
正規分布：0.00
ロジスティック分布：1.
理論的な尖度は上から順に
-1.2, -0.6, 0, 1.2 であ
注：それぞれの分布は，平均値０，分散１に調整済