3章　度数分布とローレンツ曲線

1. 度数分布表

教科書 48-51ページ

(1) データの表し方

- 例：20人のテストの得点のデータ（48ページ）
  43,20,18, 38,32,33,91,9,12,26,41,
  53,25,65,29,37,36,43,33,57
- データの個数の一般的な表し方：n
  - n：**データの個数（データ数，標本の大きさ）**
  - 上の例では，n=20
- データをみただけでは特徴がわからない

→大きさの順に並べる（ソートする）のも1つの方法
  - nが小さいときには有効（大きいとうまくいくとは限らない）

(2) 度数分布表

- データの特徴とは
  - 真ん中がどのくらいか
  - 何点から何点ぐらいに散らばっているか
  - 何点ぐらいを取っている人が多いか
  - ・・・

⇒表3-1のような表（**度数分布表**）にまとめる

- 度数分布表の作成方法
  ① 一定の幅をもった区間（**階級**;class）を作る
  ② 各区間（階級）に含まれるデータの個数（**度数**または頻度; frequency）を数える

(3) 度数，相対度数，累積度数

- 度数：各階級に含まれるデータの個数
  - 第1階級の度数：n_1
  - 第2階級の度数：n_2
  ・・・
  - 第 i 階級の度数：n_i
  ・・・
  - 第m階級（最後の階級）の度数：n_m

$$n = n_1 + n_2 + \cdots + n_m$$

m:階級数

階級			度数
0点以上	～	10点未満	1
10	～	20	2
20	～	30	4
30	～	40	6
40	～	50	3
50	～	60	2
60	～	70	1
70	～	80	0
80	～	90	0
90	～	100	1
100	～		0
合計			20

相対度数

- 各階級の度数の割合

$$相対度数 = \frac{各階級の度数}{データの個数} = \frac{n_i}{n}$$

$$i = 1,2,\cdots,m$$

- 相対度数の合計＝1
- 異なった分布を比較するときなどに便利
  ex. データ数が異なる場合の分布の比較

相対度数による比較

階級			度数		相対度数	
			数学	英語	数学	英語
以上		未満				
0点以上	～	10点未満	1	2	0.05	0.08
10	～	20	2	3	0.10	0.12
20	～	30	4	5	0.20	0.20
30	～	40	6	6	0.30	0.24
40	～	50	3	3	0.15	0.12
50	～	60	2	2	0.10	0.08
60	～	70	1	1	0.05	0.04
70	～	80	0	1	0.00	0.04
80	～	90	0	1	0.00	0.04
90	～	100	1	1	0.05	0.04
合計			20	25	1.00	1.00

累積度数:ある階級以下に含まれる度数の合計

階級			度数	相対度数	累積度数	累積相対度数
0点以上	～	10点未満	1	0.05	1	0.05
10	～	20	2	0.10	3	0.15
20	～	30	4	0.20	7	0.35
30	～	40	6	0.30	13	0.65
40	～	50	3	0.15	16	0.80
50	～	60	2	0.10	18	0.90
60	～	70	1	0.05	19	0.95
70	～	80	0	0.00	19	0.95
80	～	90	0	0.00	19	0.95
90	～	100	1	0.05	20	1.00
100	～		0	0.00	20	1.00
合計			20	1.00	–	–

- 第１階級の累積度数R_1＝第１階級の度数n_1
- 第２階級の累積度数R_2
  ＝第１階級の度数n_1+第２階級の度数n_2

第３階級の累積度数R_3

$$=\underbrace{\text{第１階級の度数}n_1+\text{第２階級の度数}n_2}+\text{第３階級の度数}n_3$$

＝第２階級の累積度数R_2 ＋第３階級の度数n_3

…

累積度数の公式と利用

- 累積度数の公式
- $R_i = n_1 + n_2 + \cdots + n_{i-1} + n_i$
  $= R_{i-1} + n_i$
  - 第 i 階級の累積度数 R_i
  ＝第 i-1 階級の累積度数 R_{i-1} ＋第 i 階級の度数 n_i
- **最後の階級の累積度数＝データ数**
- 順位をみるときに便利
  ex. 順位が真ん中に対応するデータ

- **累積相対度数**:相対度数の累積度数
  - **最後の階級の累積相対度数＝1**
  - 一定の割合に対応するデータをみるときに便利
    ex. 上位10%
    真ん中(50%)
- 度数分布表
  - 度数だけを表にする場合
  - 度数だけでなく，相対度数，累積度数，累積相対度数なども含めて表にする場合

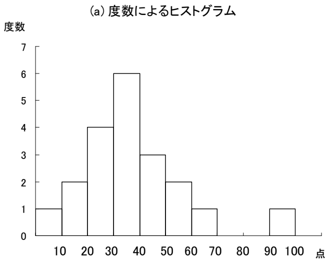
2. ヒストグラム

教科書 52-59ページ

(1) ヒストグラム
- 度数分布を視覚的に表現する
- 度数分布表を棒グラフにしたものが**ヒストグラム**
  - 縦軸:度数(または相対度数)
  - 横軸:変数(階級に対応)
- 注意点
  - 横軸，縦軸が何かを明示すること(特に横軸の単位)
  - 柱の間隔は0(連続データの場合)

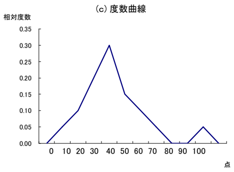
得点のヒストグラム

(a) 度数によるヒストグラム



*cf.***度数曲線**:
ヒストグラムの棒の真ん中を結んで，分布のだいたいの形を把握する

(c) 度数曲線



(2) 階級の決め方

① **階級数**
- 多すぎず，少なすぎず(一度作成してから，検討する)
- 参考:階級数決定の公式(69ページ)
  - **スタージスの公式**
  - テレルとスコットの公式など

② **階級幅**
- できれば等間隔に
- 状況に応じて幅は異なってもよい(*ex.*表3-4)

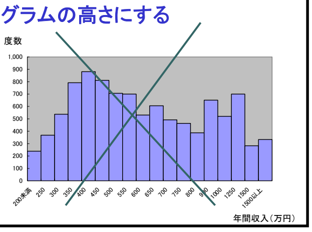
③ **級限界**(階級の両端)
- 通常は，###以上 ･･･未満 (*cf.*Excelの分析ツール)
- ###:下限，･･･:上限
- 最後の階級は，**オープンエンドの階級**になる場合がある(###以上だけで，上限を与えない)

(3) ヒストグラムにおける階級幅の調整

- 階級の幅が異なるときのヒストグラムの描き方
  (56ページ 例題3.1参照)
  - 階級幅で度数を調整する
    - 階級幅に反比例させる：階級幅が2倍→棒の高さを1/2
  - **度数÷階級幅(×a)をヒストグラムの高さにする**

※ オープンエンドの階級(最初と最後の階級)における調整
⇒**階級値**(階級の真ん中の値)を利用する

$$\text{階級値}=\frac{a+b}{2}$$



(4) クロス集計

- 表3-1などの度数分布表
⇒**単純集計**
- 表3-5のような男女別の得点の度数分布表
⇒**クロス集計**
  - クロス集計するための属性などによる違いをみることができる
    - 男女，年齢，地域・・・
    - 量的データでも可能

3. ローレンツ曲線とジニ係数

教科書 60-69ページ

(1) 所得格差の問題

- ローレンツ曲線，ジニ係数
  - **所得分布，所得分配の不平等度**を分析する方法
- 日本の所得分配
  - 高度成長期→所得分配が平等化
  - バブル崩壊後→所得の不平等度が拡大？
  - 論点 (60ページ参照)
    - 近年，格差は拡大しているのか
    - 何が格差を大きくしたのか
    - 格差によって，不平等が生じているのではないか
      - 「結果の平等」と「機会の平等」
      - 教育の問題
      - 正社員と非正規社員

(2) ローレンツ曲線

【遺産配分の仮設例】4000万円の遺産を5人の子供で分配する場合
① 完全な平等→1人800万円ずつ分配

	人数	金額(万円)	比率		累積比率	
			人数	金額	人数	金額
					0	0
四男	1	800	0.2	0.20	0.20	0.20
三男	1	800	0.2	0.20	0.40	0.40
次男	1	800	0.2	0.20	0.60	0.60
長男	1	800	0.2	0.20	0.80	0.80
五男	1	800	0.2	0.20	1.00	1.00
合計	5	4000	1	1		

平等である⇒
人数の比率と金額の比率が一致する
人数の累積比率と金額の累積比率が一致する

② 不平等な配分の例

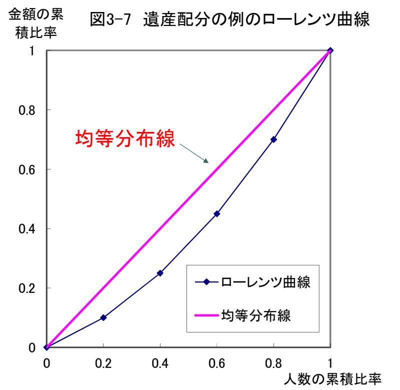
四男：400万円 (10%) 三男：600万円 (15%) 次男：800万円 (20%)
長男：1000万円 (25%) 五男：1200万円 (30%)

	人数	金額(万円)	比率		累積比率	
			人数	金額	人数	金額
					0	0
四男	1	400	0.2	0.10	0.20	0.10
三男	1	600	0.2	0.15	0.40	0.25
次男	1	800	0.2	0.20	0.60	0.45
長男	1	1000	0.2	0.25	0.80	0.70
五男	1	1200	0.2	0.30	1.00	1.00
合計	5	4000	1	1		

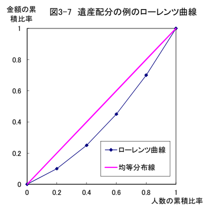
2つの比率をグラフにする

不平等である⇒
– 人数の比率と金額の比率が一致しない
– 人数の累積比率と金額の累積比率が一致しない

⇒**ローレンツ曲線**



ローレンツ曲線の描き方

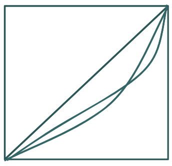


- ローレンツ曲線のグラフ
  - 横軸：人数の累積比率（累積相対度数）
  - 縦軸：所得の累積比率
  - 正方形のなかにグラフを描く
    - 横軸，縦軸とも最大値は1.0
  - 各階級について，両者の値をプロットして，線で結ぶ
- 対角線（45度線）＝**均等分布線**または**完全平等線**
  - 完全に平等な場合のローレンツ曲線

ローレンツ曲線の解釈と注意点

- 完全に均等な場合：ローレンツ曲線＝均等分布線
- 最も不平等な場合：ローレンツ曲線＝正方形の右側の外枠
- ローレンツ曲線と均等分布線の近さ
  - ローレンツ曲線が均等分布線に近いほど分配は平等
  - ローレンツ曲線が均等分布線から離れるほど（外枠に近いほど）分配は不平等
- 1本のローレンツ曲線で不平等度は測れない
  - 図3-8参照
  - 過去との比較（時系列）
  - 国別の比較（クロスセクション）
- ローレンツ曲線は，均等分布線の上に位置することはない
  - 所得を低いほうから並べるため
- 階級ごとに度数が異なっている場合は表3-7などを参照

(3) ジニ係数



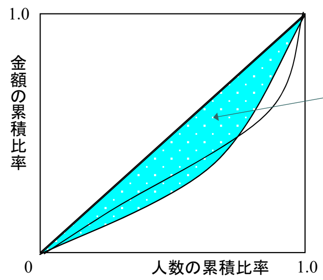
① ジニ係数の目的

- ローレンツ曲線での不平等度の比較
  - 視覚的だが，主観的になりやすい
  - ローレンツ曲線が交わったり，接したりした場合の比較が困難
  - かなり近かったり，多くの比較を行う場合（たとえば時系列比較）も困難

→数値による不平等の比較が必要：**ジニ係数**

＝ローレンツ曲線の均等分布線への近さを測定する指標

② ジニ係数（G）の定義



距離で近さを測るのは困難

面積で測る

ローレンツ曲線と均等分線で囲まれた面積が小さい方が，両者は近いと判断する

ジニ係数G＝ローレンツ曲線と均等分布線によって囲まれる面積×2

③ ジニ係数の解釈

- 囲まれた面積を2倍する理由
  - ジニ係数の範囲：0≦G≦1

（囲まれた面積だと0～0.5）

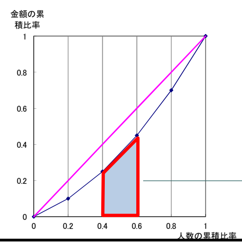
- G=0：完全な平等（均等分布線と一致）

G=1：完全な不平等（外枠に一致）

  - ジニ係数が小さいほど（0に近いほど），分配は平等
  - ジニ係数が大きいほど（1に近いほど），分配は不平等

④ジニ係数の計算

	人数	金額（万円）	比率		累積比率		台形の面積
			人数	金額	人数	金額	
					0	0	
四男	1	400	0.2	0.1	0.20	0.10	0.01
三男	1	600	0.2	0.15	0.40	0.25	0.035
次男	1	800	0.2	0.2	0.60	0.45	0.07
長男	1	1000	0.2	0.25	0.80	0.70	0.115
五男	1	1200	0.2	0.3	1.00	1.00	0.17
合計	5	4000	1	1			0.4



G=(0.5-0.4)×2=0.2
または
G=1-0.4×2=0.2

台形の面積=（上底＋下底）×高さ÷2

$(0.25+0.45)\times 0.2 \div 2 = 0.07$

4章　データの代表値

1. 平均値
   (1) 平均値（算術平均値）
   (2) 2つのグループの平均値
   (3)度数分布表からの平均値の計算
   (4)平均値の定義と∑
   (5) 平均値の性質
2. メディアン（中央値）
   (1) メディアンの定義と性質
   (2)度数分布表からのメディアンの計算
3. モード（最頻値）
4. 平均値・メディアン・モードの関係
5. その他の平均値

1. 平均値

教科書 76-84ページ

(1) 平均値（算術平均値）

- 平均値の目的
  - データの中心を測る尺度：代表値
  - 平均値は代表値の1つ
- 平均値 $\bar{x}$（エックス・バー）の定義
  - 平均値＝データの合計÷データ数

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

データの表し方

- 一般的なデータの表現：

$$x_1, x_2, \cdots, x_n$$

または x_i $(i = 1,2,\cdots,n;\ n\text{はデータ数})$

- 例題4.1　330, 280, 230, 240, 390, 290, 340, 1580

$$\downarrow\ \downarrow\ \downarrow\ \downarrow\ \downarrow\ \downarrow\ \downarrow\ \downarrow$$

$$x_1\ x_2\ x_3\ x_4\ x_5\ x_6\ x_7\ x_8$$

  - 8人の場合と7人の場合の代表性
  - 外れ値

(2) 2つのグループの平均値

- 48ページの得点の例題（男女別の得点）

	平均点	度数	計
男	33	11	363
女	42	9	378
全体		20	741

$$\cancel{\frac{33+42}{2} = 37.5}$$

単純平均

$$\frac{11\times33+9\times42}{11+9} = \frac{11}{20}\times33 + \frac{9}{20}\times42 = 37.05$$

加重平均

ウェイト　$\frac{n_i}{n}$

(3) 度数分布表からの平均値の計算

- 50ページの例
  （20人の得点）

n_i　x_i^*

階級 以上		階級 未満	度数	階級値	度数×階級値
0	～	10	1	5	5
10	～	20	2	15	30
20	～	30	4	25	100
30	～	40	6	35	210
40	～	50	3	45	135
50	～	60	2	55	110
60	～	70	1	65	65
70	～	80	0	75	0
80	～	90	0	85	0
90	～	100	1	95	95
合計			20		750

$$\bar{x}^* = 750 \div 20 = 37.5$$

本当の平均との違い

- 度数分布表から計算した平均

度数をウェイトにした加重平均

$$\frac{1\times5+2\times15+\cdots+1\times95}{1+2+\cdots+1} = \frac{750}{20} = 37.5$$

- もとのデータから直接計算した平均

異なる

$$\frac{43+20+18+\cdots+57}{1+2+\cdots+1} = \frac{741}{20} = 37.05$$

ただし違いは小さい

(4) 平均値の定義と総和記号∑

① 平均値の定義 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

② 総和記号Σの定義

Σ(シグマ):合計

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

合計の最後

合計の最初

ex. $\sum_{i=1}^{3} x_i = x_1 + x_2 + x_3$,

$\sum_{i=3}^{6} x_i = x_3 + x_4 + x_5 + x_6$

③ Σの計算例

$x_1 = 330, x_2 = 280, x_3 = 230, x_4 = 240, x_5 = 390, \cdots$

$\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4 = 330 + 280 + 230 + 240 = 1080$

$\sum_{i=3}^{5} x_i = x_3 + x_4 + x_5 = 230 + 240 + 390 = 860$

$\sum_{i=1}^{4} 5x_i = 5x_1 + 5x_2 + 5x_3 + 5x_4 = 1650 + 1400 + 1150 + 1200 = 5400$

$\sum_{i=1}^{4} (x_i + 10) = (x_1 + 10) + (x_2 + 10) + (x_3 + 10) + (x_4 + 10)$

$= 340 + 290 + 240 + 250 = 1120$

$\sum_{i=1}^{3} (x_i - 250)^2 = (x_1 - 250)^2 + (x_2 - 250)^2 + (x_3 - 250)^2$

$= (330 - 250)^2 + (280 - 250)^2 + (230 - 250)^2$

$= 80^2 + 30^2 + (-20)^2 = 6400 + 900 + 400 = 7700$

④ Σの性質

定義 $\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$

(1) $\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i$

(2) $\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

または $\sum_{i=1}^{n} (x_i - y_i) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i$

(3) $\sum_{i=1}^{n} a = na$

証明は
練習問題12

Σの性質の数値例

i	x_i	$5x_i$	y_i	x_i+y_i	$a(=5)$
1	330	1650	10	340	5
2	280	1400	20	300	5
3	230	1150	30	260	5
合計	840	4200	60	900	15

$\sum_{i=1}^{n} x_i$　$\sum_{i=1}^{n} 5x_i = 5\sum_{i=1}^{n} x_i$　$\sum_{i=1}^{n} y_i$　$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$　$\sum_{i=1}^{n} 5 = 5 \times 3$

(5) 平均値の性質

1. 合計＝平均×データ数 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \Rightarrow n\bar{x} = \sum_{i=1}^{n} x_i$

2. 平均からの偏差:各データが平均からどれだけ離れているか

平均からの偏差＝各データ－平均 $x_i - \bar{x}$

平均からの偏差の合計は0

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

例: $x_1 = 50, x_2 = 50, x_3 = 70, x_4 = 80, x_5 = 90$

(平均からの)偏差

	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - 70)^2$
	50	-18	324	400
	50	-18	324	400
	70	2	4	0
	80	12	144	100
	90	22	484	400
合計	340	0	1280	1300
平均	68			

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

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

3. **平均からの偏差の2乗和は最小**(証明は97ページ)

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 \to \text{最小}$$ $\boxed{a\text{からの偏差}: x_i - a}$

4. **データに一定値を加えた平均＝もとの平均＋一定値**

$$\boxed{\overline{x+c} = \bar{x} + c}$$

ex.あるテストで全員の得点に10点加えると，平均点も10点上がる

5. **データの一定倍の平均＝もとの平均×一定値**

$$\boxed{\overline{ax} = a\bar{x}}$$

ex. あるテストで全員の得点を2倍すると，平均点も2倍になる

※$ax_i + c$の平均値：$\overline{ax+c} = a\bar{x} + c$　証明は練習問題14

6. **平均は極端に大きい(小さい)値の影響を受けやすい**

- 例題4.1

330,280,230,240,390,290,340,1580

$n=8$のとき $\bar{x}=460$
- 平均より多い人：1人
- 平均より少ない人：7人

→平均＝まん中ではない
(平均が中心としての意味がない)

$n=7$のとき $\bar{x}=300$
- 平均より多い人：3人
- 平均より少ない人：4人

→平均が中心としての意味をもつ

cf. trimmed mean (刈り込み平均)

2. メディアン(中央値)

教科書 85-86ページ

(1) メディアンの定義と性質

- **メディアン(中央値)**：データを大きさの順に並べたとき(**順序統計量**)，順位が真ん中に対応するデータの値
  - 例1　10, 20, 30, 80, 100
  - 例2　10, 20, 30,80
- データ数nが奇数の場合
  - メディアン＝ $\dfrac{n+1}{2}$ 番目のデータ
- データ数nが偶数の場合
  - メディアン＝ $\dfrac{n}{2}$ 番目と $\dfrac{n}{2}+1$ 番目の平均

メディアンと平均

- 次の6人のテストの得点のデータについて，メディアンと平均を求めよ.

【データ】0, 5, 10, 15, 20, 100

メディアン＝

平均＝

- 100点を除いた5人のメディアンと平均を求めよ.

メディアン＝

平均＝

- どんなことがわかるか

メディアンの性質

- **メディアンは外れ値の影響を受けない**
  - 「まん中」の値として平均よりも適当
  - 平均≠まん中
- **メディアンより大きいデータの個数＝メディアンより小さいデータの個数**

→データの中央値(中位数)

中心の尺度としては，平均よりもメディアンの方が望ましい

(2) 度数分布表からのメディアンの求め方

- メディアンの定義から，メディアンに対応するデータ(順位が真ん中)が含まれる階級を決めて，
  ① メディアンの含まれる階級の階級値
  ② メディアンの含まれる階級を比例配分

のいずれかで求める

- (例)57ページ　表3-4のメディアン
  - メディアンの含まれる階級　500～550万円

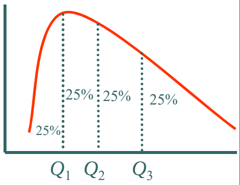
①の方法･･･メディアン=525万円(または表より522万円)

②の方法･･･ $\text{メディアン} = 500 + \dfrac{5000-4334}{700} \times 50$

$= 548$

(3) 相対貧困率と四分位数

- 相対貧困率
  - 所得水準のメディアンの1/2以下の人が全人口に占める割合
  - 日本は比較的高い数値
- 分位数
  - メディアン＝二分位数（データを50%ずつに分ける点）
  - **四分位数**＝データを25%ずつに分ける点
    - 第1四分位数(Q_1)，第2四分位数(Q_2)，第3四分位数(Q_3)
  - 他にも五分位数，十分位数など



3．モード

教科書 89-91ページ

- **モード**：度数分布表で最大の度数に対応
  - ヒストグラムの頂点に対応する階級の階級値
  - 階級の幅が異なる場合には，階級の幅で調整した度数に対応する階級値
- 数値例
  - 表3－1
  - 表3－4
- 利用方法
  - 世帯人員のモード
  - 出生率（合計特殊出生率）と子供の数のモード
  - 平均は，現実の値を表わさないことが多い

世帯人員別世帯数の推移

世帯人員（10区分）	昭和45年 1970	昭和55年 1980	平成2年 1990	平成12年 2000	平成17年 2010	平成27年 2015
世帯数						
総数	30, 297, 014	35, 823, 609	40, 670, 475	46, 782, 383	51, 842, 307	53, 331, 797
1人	6, 137, 443	7, 105, 246	9, 389, 660	12, 911, 318	16, 784, 507	18, 417, 922
2	4, 183, 902	6, 001, 075	8, 370, 087	11, 743, 432	14, 125, 840	14, 876, 547
3	5, 321, 911	6, 475, 220	7, 350, 639	8, 810, 437	9, 421, 831	9, 364, 781
4	6, 884, 785	9, 070, 100	8, 787, 908	7, 924, 827	7, 460, 339	7, 069, 141
5	3, 907, 031	3, 981, 763	3, 805, 147	3, 167, 227	2, 571, 743	2, 403, 060
6	2, 285, 353	2, 032, 848	1, 903, 065	1, 448, 960	984, 751	811, 735
7	982, 787	843, 249	814, 631	594, 352	359, 325	280, 442
8	386, 814	235, 880	198, 932	144, 907	100, 655	79, 779
9	134, 855	55, 354	38, 309	27, 856	23, 721	20, 845
10人以上	72, 133	22, 874	12, 097	9, 067	9, 595	7, 545
世帯人員	103, 350, 641	115, 450, 540	121, 545, 271	124, 724, 660	124, 724, 660	124, 724, 660
1世帯当たり人員（平均値）	**3. 41**	**3. 22**	**2. 99**	**2. 67**	**2. 46**	**2. 34**
メディアン	3	3	3	2	2	2

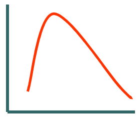
出所：総務省統計局「国勢調査」

4．平均値・メディアン・モードの関係

教科書 91-93ページ

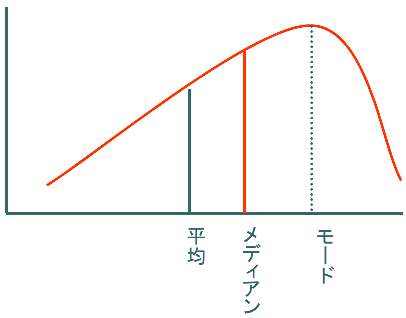
―分布の形の平均値・メディアン・モードの大きさ―

- 例題4.6：表3-7 (57ページ)の年間収入階級別世帯数の中心の尺度
  - 平均値：　640万円
  - メディアン：525万円
  - モード：　375万円



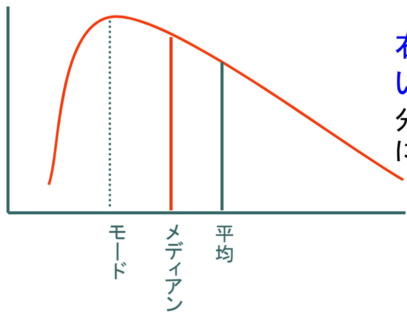
- モード＜メディアン＜平均

→分布の山が左に偏った，右スソの長い分布をとることが多い
  - 社会科学のデータに多い



左スソが長い分布＝
分布の山が右に偏った分布

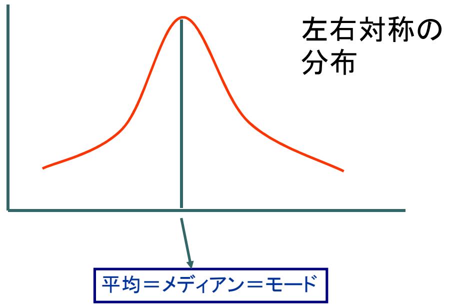
平均＜メディアン＜モード ←→ 左スソが長い分布，（左に歪んだ分布）



右スソが長い分布＝
分布の山が左に偏った分布

モード＜メディアン＜平均 ←→ 右スソが長い分布（右に歪んだ分布）

（例）貯蓄残高のデータ



モード＝メディアン＝平均 ←→ 左右対称の分布（自然科学のデータなど）

平均が最良の中心の尺度ではない

- 中心の尺度としてなら，メディアンの方が適切な代表値
- にもかかわらず，なぜ平均の方がよく用いられるか
  - メディアンにはない望ましい性質をもっている

⇒母集団と標本の関係（統計学IIの問題）

5. その他の平均値

教科書 93-97ページ

- **幾何平均**＝ $\sqrt[n]{x_1 \times x_2 \times \cdots \times x_n}$
  - 成長率の平均
  - 複利計算
  - 物価指数･･･
- **移動平均**
  - 3か月（3項）移動平均 $\dfrac{x_{t-1}+x_t+x_{t+1}}{3}$
  - 5か月（5項）移動平均 $\dfrac{x_{t-2}+x_{t-1}+x_t+x_{t+1}+x_{t+2}}{5}$
  - ･･･
    - 平滑化（不規則変動の除去）
    - 項数が長いほど平滑化される
    - 期間の最初と最後のデータが利用できない

3章　度数分布とローレンツ曲線　練習問題

１．次の表は，1996 年から 2015 年までの経済成長率（%）の推移を示している．

年度	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
経済成長率(%)	2.7	0.0	−0.8	0.7	2.5	−0.6	0.9	2.1	1.5	2.1
年度	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
経済成長率(%)	1.4	1.2	−3.5	−2.2	3.2	0.5	0.9	2.6	−0.4	1.3

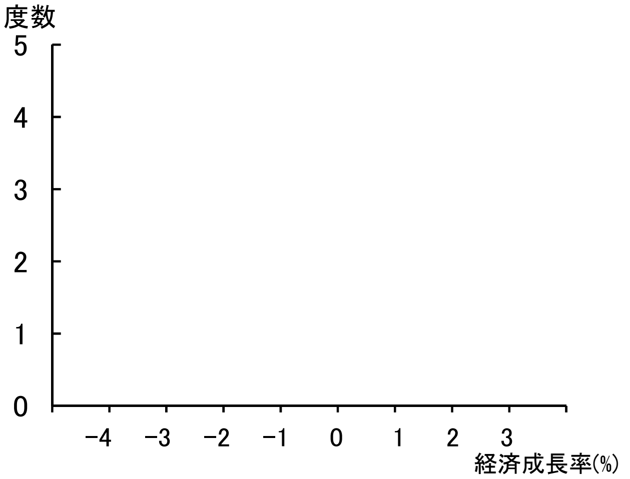
出所：内閣府「国民経済計算」

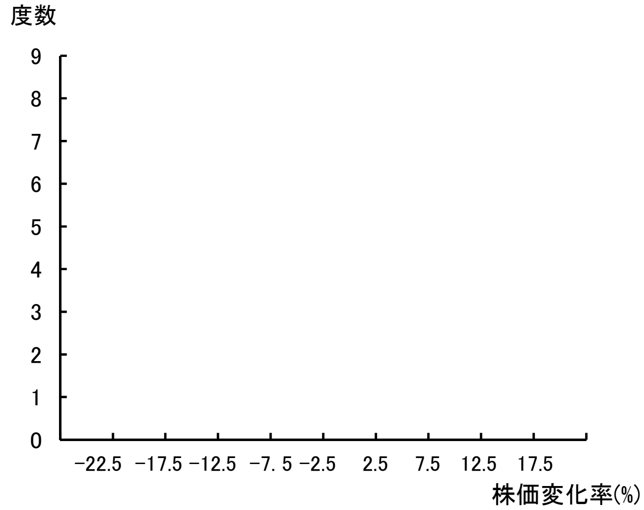
(1) 経済成長率とは何か.

(2) 下の度数分布表を完成させ，ヒストグラムを描け.

（4章練習問題用）

経済成長率			度数	相対度数	累積度数	累積相対度数	階級値	度数×階級値
以上		未満						
−4%	～	−3%						
−3	～	−2						
−2	～	−1						
−1	～	0						
0	～	1						
1	～	2						
2	～	3						
3		4						
合計							−	





２．次の表は，2015 年 3 月から 2017 年 3 月までのトヨタ自動車の株価変化率（月次，%）の推移を示している．下の度数分布表を完成させ，ヒストグラムを描け（前ページ右）．

年月	2015年3月	2015年4月	2015年5月	2015年6月	2015年7月	2015年8月	2015年9月	2015年10月	2015年11月
株価変化率(%)	4.0	−0.3	2.9	−4.7	0.6	−12.9	−3.1	7.1	2.5

年月	2015年12月	2016年1月	2016年2月	2016年3月	2016年4月	2016年5月	2016年6月	2016年7月	2016年8月
株価変化率(%)	−2.2	−3.8	−18.1	0.9	−5.0	2.3	−12.6	16.7	5.8

年月	2016年9月	2016年10月	2016年11月	2016年12月	2017年1月	2017年2月	2017年3月
株価変化率(%)	−7.4	5.2	9.4	3.4	−4.3	−3.3	−5.1

（4章練習問題用）

株価変化率			度数	相対度数	累積度数	累積相対度数		階級値	度数×階級値
以上		未満							
−22.5%	〜	−17.5%							
−17.5	〜	−12.5							
−12.5	〜	−7.5							
−7.5	〜	−2.5							
−2.5	〜	2.5							
2.5	〜	7.5							
7.5	〜	12.5							
12.5	〜	17.5							
合計					−	−		−	

３．テキスト 3 章練習問題　２　（70 ページ）

４．テキスト 3 章練習問題　４(1)　（70 ページ）

【交通事故死者数の階級は 0-50，50-100，100-150，…】

【平均気温の階級は 7.5-10.0，10-12.5，12.5-15.0，…】

５．次のデータは，ある企業に勤める 20 人の年間収入のデータである（単位：万円）．

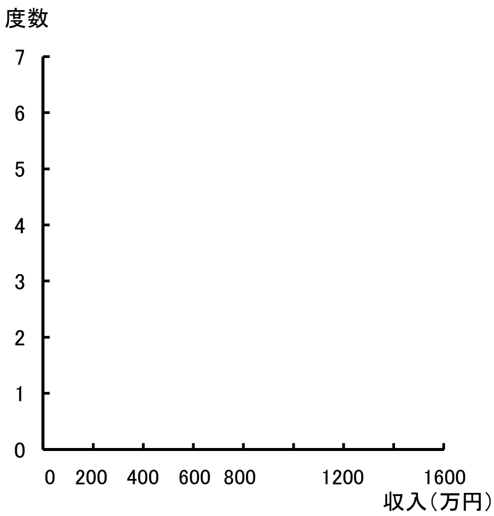
180, 220, 250, 300, 320, 350, 380, 480, 500, 550

580, 650, 720, 780, 820, 950, 1000, 1150, 1350, 1500

下の度数分布表を完成させ，ヒストグラムを描け．

（4章練習問題用）

収入階級（万円）			度数	相対度数	累積度数	累積相対度数		階級値	度数×階級値
以上		未満							
0	〜	200							
200	〜	400							
400	〜	600							
600	〜	800							
800	〜	1200							
1200	〜	1600							
合計					—	—		—	



６．テキスト 64 ページの遺産の分配２（900，800，700，600，1000 万円と配分）の場合，下の表を埋め，ローレンツ曲線を描け．またジニ係数を算出せよ．

	人数	金額（万円）	比率		累積比率		台形の面積
			人数	金額	人数	金額	
					0	0	
A	1	600					
B	1	700					
C	1	800					
D	1	900					
E	1	1000					
合計	5	4000					

６のローレンツ曲線

７のローレンツ曲線

７．1500 万円を４人に分配することを考える．

(1) 150 万円，300 万円，450 万円，600 万円と分配したとき，(1)の表を完成させ，ローレンツ曲線を描き，ジニ係数を求めよ．

(2) 300 万円，300 万円，450 万円，450 万円と分配した場合のローレンツ曲線を(1)と同じグラフに描け．またジニ係数を求め，どちらの分配がより平等（格差が小さい）か述べよ．

（1）	データ		比率		累積比率		台形の面積
	人数	収入	人数	収入	人数	収入	
	1	150					
	1	300					
	1	450					
	1	600					
計	4	1500					

（2）	データ		比率		累積比率		台形の面積
	人数	収入	人数	収入	人数	収入	
	1	300					
	1	300					
	1	450					
	1	450					
計	4	1500					

８．2000 万円を A～E の５人に配分するのに次の２通りの配分を考える

配分①　A:200 万円，B:300 万円，C:400 万円，D:500 万円，E:600 万円

配分②　A:100 万円，B:300 万円，C:400 万円，D:500 万円，E:700 万円

それぞれの場合について下の表を完成させ，１つのグラフにローレンツ曲線を描け．また ジニ係数を計算せよ．どちらの配分がより不平等か.

配分①			比率		累積比率		
	人数	配分額	人数	配分額	人数	配分額	台形の面積
A	1	200					
B	1	300					
C	1	400					
D	1	500					
E	1	600					
合計	5	2000					
						G	

配分②			比率		累積比率		
	人数	配分額	人数	配分額	人数	配分額	台形の面積
A	1	100					
B	1	300					
C	1	400					
D	1	500					
E	1	700					
合計	5	2000					
						G	

８のローレンツ曲線

９．下の表は，1986 年と 2017 年の年間収入五分位階級別の１か月あたりの実収入のデータである（二人以上の世帯のうち勤労者世帯．適当に四捨五入して，数値を丸めてある）．このデータについて，下の表を埋め，ローレンツ曲線を描き，ジニ係数を計算せよ．ただし，ローレンツ曲線は１つのグラフに描くこと．その結果どのようなことがわかるか．ただし，実収入の比率は小数３位を四捨五入して，小数第２位まで求めよ（台形の面積は小数第３位まで）．

1986			比率		累積比率		
階級	世帯数	実収入	世帯数	実収入	世帯数	実収入	台形の面積
		万円					
Ⅰ	2000	30					
Ⅱ	2000	40					
Ⅲ	2000	40					
Ⅳ	2000	50					
Ⅴ	2000	70					
合計							

（丸める前の数字は，Ⅰ：250,540，Ⅱ：346,455，Ⅲ：426,196，Ⅳ：522,993，Ⅴ：718,528，単位は円）

2017			比率		累積比率		
階級	世帯数	実収入	世帯数	実収入	世帯数	実収入	台形の面積
		万円					
Ⅰ	2000	30					
Ⅱ	2000	40					
Ⅲ	2000	50					
Ⅳ	2000	60					
Ⅴ	2000	90					
合計							

（丸める前の数字は，Ⅰ：304,777，Ⅱ：396,979，Ⅲ：500,064，Ⅳ：604,015，Ⅴ：863,267，単位は円）
出所：総務省統計局「家計調査」

９のローレンツ曲線

10．テキスト３章　練習問題７（71 ページ）（各階級は，度数を 2000 ずつとしておく（同じ度数であれば，いくつにしても構わない）

11．【要ＰＣ】３章　練習問題６（71 ページ）

12．【要ＰＣ】表 3-9（従業員規模別事業所数，71 ページ）の各産業についてローレンツ曲線を描き，ジニ係数を計算せよ．

4章 データの代表値:平均 練習問題

1．あるクラスの男子 24 人の平均点は 50 点，女子 16 人の平均点は 60 点であった．クラス全体の平均点を求めよ．

2．ある科目の試験で，１年生は 200 人受験し，その平均点は 75 点，２年生は 60 人受験して，その平均点は 72 点，３年生以上は 40 人受験して，その平均点は 60 点であったという．全体の平均点を求めよ．

3．テキスト４章練習問題　3

4．20 歳代 2000 人を調査したところ失業率は 10%，30 歳代 3000 人を調査したところ失業率は 8%であった．20 歳代と 30 歳代全体の失業率は何%か.

5．３章練習問題１，２，５について，作成した度数分布表よりそれぞれ平均値を求めよ．

6. 右の表は, ある会社の従業員の給与に関する度数分布表である. この表から, 給与の平均値を求めよ.

給与（万円）			度数	階級値	度数×階級値
以上		未満			
0	～	200	3		
200	～	400	8		
400	～	600	7		
600	～	800	4		
800	～	1000	1		
1000	～	1400	2		
合計				―	

7．次の値を計算せよ．ただし， x, y は以下の値をとるものとする．

$$x_1 = 330, x_2 = 280, x_3 = 230, x_4 = 240, x_5 = 390, x_6 = 290, x_7 = 340, x_8 = 1580$$
$$y_1 = 10, y_2 = 20, y_3 = 30$$

(1) $\sum_{i=1}^{3} x_i$　(2) $\sum_{i=4}^{7} x_i$　(3) $\sum_{i=1}^{3} x_i^2$

(4) $\sum_{i=1}^{3} (x_i + 20)$　(5) $\sum_{i=1}^{3} 5x_i$　(6) $\sum_{i=1}^{3} (x_i - 250)$

(7) $\sum_{i=1}^{3} (x_i - 250)^2$　(8) $\sum_{i=1}^{3} (x_i + y_i)$　(9) $\sum_{i=1}^{3} x_i y_i$

8． $x_1 = 10, x_2 = 20, x_3 = 30,\ y_1 = 5, y_2 = 0, y_3 = -1$ のとき，次の値を求めよ(n=3).

(1) $\sum_{i=1}^{n} x_i$　(2) $\sum_{i=1}^{n} y_i$　(3) $\sum_{i=2}^{3} x_i$　(4) $\sum_{i=1}^{n} 3x_i$

(5) $\sum_{i=1}^{n} -3y_i$　(6) $\sum_{i=1}^{n} x_i^2$　(7) $\sum_{i=1}^{n} (x_i + y_i)$

(8) $\sum_{i=1}^{n} (x_i - y_i)$　(9) $\sum_{i=1}^{n} (x_i - 20)$　(10) $\sum_{i=1}^{n} (x_i - 20)^2$

9．テキスト４章練習問題　1（97-98 ページ）

10. テキスト4章練習問題　2（98ページ）

11. $\sum_{i=1}^{3} x_i = 840, \sum_{i=1}^{3} y_i = 60$ を利用して，次の値を求めよ．

(1) $\sum_{i=1}^{3} 5x_i$　(2) $\sum_{i=1}^{3} 5y_i$　(3) $\sum_{i=1}^{3}(x_i + y_i)$　(4) $\sum_{i=1}^{3}(x_i - y_i)$

(5) $\sum_{i=1}^{3}(5x_i + 5y_i)$　(6) $\sum_{i=1}^{3}(x_i + 20)$　(7) $\sum_{i=1}^{3}(x_i - 250)$

12. 次の式を証明せよ．

(1) $\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i$

（証明）$\sum_{i=1}^{n} ax_i =$

(2) $\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

（証明）$\sum_{i=1}^{n}(x_i + y_i) =$

(3) $\sum_{i=1}^{n} a = na$

（証明）$\sum_{i=1}^{n} a =$

(4) $\sum_{i=1}^{n}(x_i - y_i) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i$

（証明）$\sum_{i=1}^{n}(x_i - y_i) =$

13. 20,40,40,100 という4つのデータに関して，

(1) 平均を求め，平均からの偏差の合計が 0 になることを確かめよ．

(2) すべてのデータが 10 下がったときの平均を求めよ．

(3) すべてのデータが 1/4 になったときの平均を求めよ．

(4) 平均からの偏差2乗和を求めよ．さらに，40 からの偏差2乗和を求めよ．

i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - 40)^2$
1	20			
2	40			
3	40			
4	100			
合計				
平均				

14. 次の式を証明せよ．

(1) $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

（証明）$\sum_{i=1}^{n}(x_i - \bar{x}) =$

(2) $\overline{x+c} = \bar{x} + c$

（証明）$\overline{x+c} = \frac{1}{n}\sum_{i=1}^{n}(x_i + c) =$

(3) $\overline{ax} = a\bar{x}$

（証明）$\overline{ax} = \frac{1}{n}\sum_{i=1}^{n} ax_i =$

15. ４章練習問題　6

16. ４章練習問題　7

17. 次のデータは，20 世帯の貯蓄残高のデータである（単位：万円）．

50,120,180,220,230,240,300,310,320,350,350,420,450,500,520,550,620,700,980,1100

(1) これらのデータを度数分布表にまとめよ（階級は 0-200, 200-400,･･･,1000-1200）．

(2) ヒストグラムを作成せよ．

(3) 度数分布表により，平均とモードを求めよ．またメディアンをもとのデータから求めよ

(4) (3)の結果から，このデータの分布の形について，理由を簡単につけて述べよ．

貯蓄残高（万円）			度数	相対度数	累積度数	累積相対度数		階級値	度数×階級値
以上		未満							
0	～	200							
200	～	400							
400	～	600							
600	～	800							
800	～	1000							
1000	～	1200							
合計					―	―		―	

18. 次のデータは，あるクラスの 25 人の英語のテストの得点のデータである（単位: 点).

5, 13, 25, 30, 37, 41, 47, 52, 53, 57, 58, 61, 63, 63, 65,
66, 67, 70, 75, 77, 78, 82, 82, 83, 95

(1) これらのデータを度数分布表にまとめよ（階級は，0-20,20-40,⋯,80-100).
(2) ヒストグラムを作成せよ.
(3) 度数分布表により，平均とモードを求めよ．またメディアンをもとのデータから求めよ
(4) (3)の結果から，このデータの分布の形について，理由を簡単につけて述べよ.

得点（点）			度数	相対度数	累積度数	累積相対度数		階級値	度数×階級値
以上		未満							
0	～	20							
20	～	40							
40	～	60							
60	～	80							
80	～	100							
合計					―	―		―	

19. 次のデータは，ある株の 20 か月の変化率のデータである（単位:%).

0.1, 0.2, 0.3, 0.8, 1.2, 2.1, 2.2, 2.2, 2.5, 3.2, 3.2, 3.5, 3.8, 3.9, 4.2, 4.5, 5.1, 6.2, 6.3, 9.8

(1) これらのデータを度数分布表にまとめよ（階級は，0-2,2-4,⋯,8-10).
(2) ヒストグラムを作成せよ.
(3) 度数分布表により，平均とモードを求めよ．またメディアンをもとのデータから求めよ
(4) (3)の結果から，このデータの分布の形について，理由を簡単につけて述べよ.

変化率（%）			度数	相対度数	累積度数	累積相対度数		階級値	度数×階級値
以上		未満							
0	～	2							
2	～	4							
4	～	6							
6	～	8							
8	～	10							
合計					―	―		―	