トマ・ピケティ『*21* 世紀の資本』に見る数学教材２

統計資料から作られる総合指数の問題点

渡邊 勝

名寄市立大学短期大学部非常勤講師

2015 年９月５日

札幌市大通り高校

0. 問題意識

前編で触れたように経済格差が大きな社会問題として浮上している当にその時代に応じた書物が表れ多くの人々がこの問題、この書物、そして著者に関心を持っています。

ピケティのこの書物の中には、数学の教材になりそうな記述が見られます。前編で紹介した人口動態と指数関数は「相性」が良い関係にあります。数教協の実践でバクテリアの増殖が指数関数の導入教材として取り扱われています。バクテリアの増殖は話が見えやすいのですが、人口増加は身につまされる実感があります。それをすくい取って教材化しようとしました。

国富と所得比が貯蓄率と成長率の比になる経済理論は、漸化式あるいは階差方程式の問題です。高校生には、漸化式の「応用問題」として提示する教材化をしました。

日本の数学教科書には、フランス、米国と較べて、自然科学、社会科学で使われる話題あるいは例題、問題が少ないようです。数教協が標榜する「量から入る数学」の観点からも、この種の実践が待ち望まれるところです。そこで、話題性に乗って教材化をすることにしました。

1. 統計資料から作られる「総合指数」の問題点:ジニ係数を例として

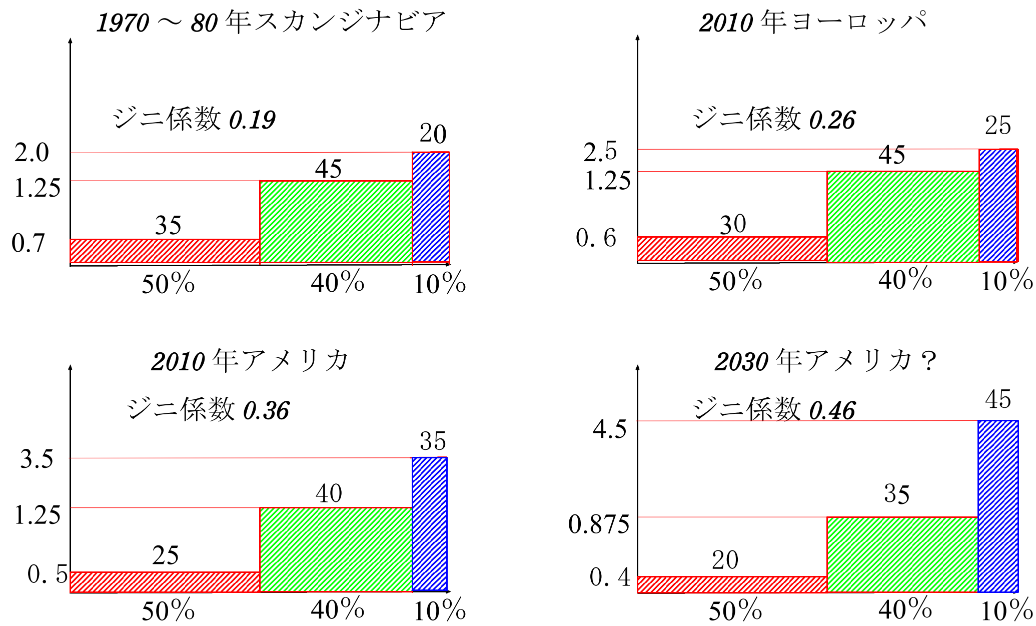
ピケティ氏は下の表を示して格差分析を行っています。本著書 *P.258*（渡辺がより見やすくなるように手を入れています。）

表 *7–1*　時間空間的に見た労働所得格差

	低格差	中格差	高格差	超高格差
	1970–80 年代、	*2010* 年、	*2010* 年、	*2030* 年、
	スカンジナビァ	ヨーロヅパ	米国	米国*?*
トップ *10* ％／上流階級	*20%*	*25* ％	*35* ％	*45* ％
トップ *1* ％／支配階級	*5* ％	*7* ％	*12* ％	*17* ％
残余 *9* ％／富裕階級	*15* ％	*18* ％	*23* ％	*28* ％
中間 *40* ％／中流階級	*45* ％	*45* ％	*40* ％	*35* ％
底辺 *50* ％／下流階級	*35* ％	*30* ％	*25* ％	*20* ％
ジニ係数（合成格差指数）	*0.19*	*0.26*	*0.36*	*0.46*

この表を柱状グラフにすると次のようになります。

横軸が「貧しい」→「富める」即ち「下流階級」「中流階級」「上流階級」の順に左から右へその人数比率を幅にして区間を定める。縦軸は各人の所得額指数。



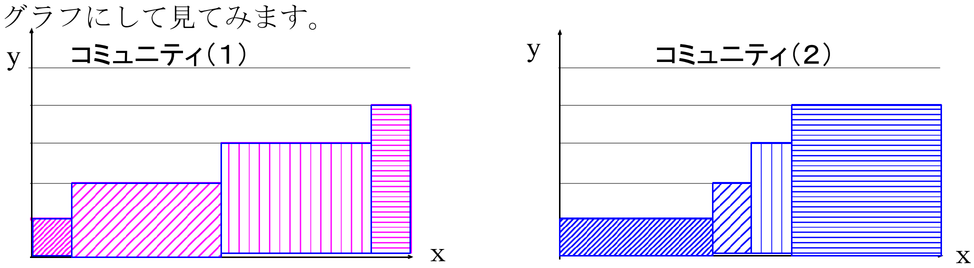
格差（所得格差、試験における得点格差など）を指数化する手法として、ローレンツ曲線とそれに基づくジニ係数があります。いまその簡単なモデルを示します。

10万人が居る二つのコミュニティがあり、それぞれの社会では下の表のような所得分布を持っています。コミュニティ(1)が比較的均等ですが、コミュニティ(2)は、格差が大きい社会です。

	コミュニティ(1)				コミュニティ(2)			
所得（百万円）：y	1	2	3	4	1	2	3	4
人数（万人）：x	1	4	4	1	4	1	1	4

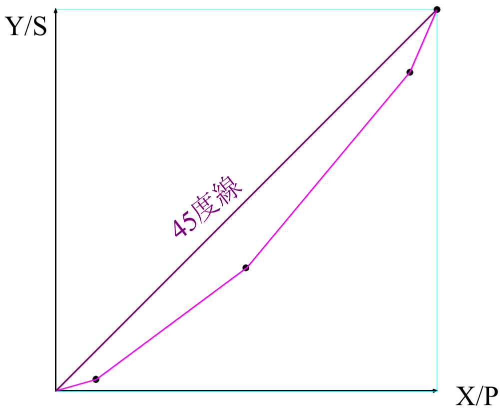
ｙとXの関係をグラフにして見てみます。

所得分布図

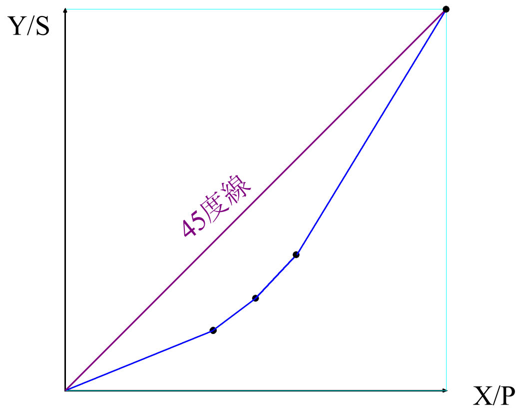


次に下表を作り「ローレンツ曲線（折れ線)」と呼ばれる折れ線を描きます。

	コミュニティ(1)				コミュニティ(2)			
所得（百万円）：y	1	2	3	4	1	2	3	4
人数（万人）：x	1	4	4	1	4	1	1	4
階層別所得総額：yx	1	8	１２	4	4	2	3	１６
上記累計：Y	1	9	２１	２５＝S	4	6	9	２５＝S
Y／S	0.04	0.34	0.84	1	0.16	0.24	0.34	1
人数累計：X	1	5	9	１０＝P	4	5	6	１０＝P
X／P	0.1	0.5	0.9	1	0.4	0.5	0.6	1



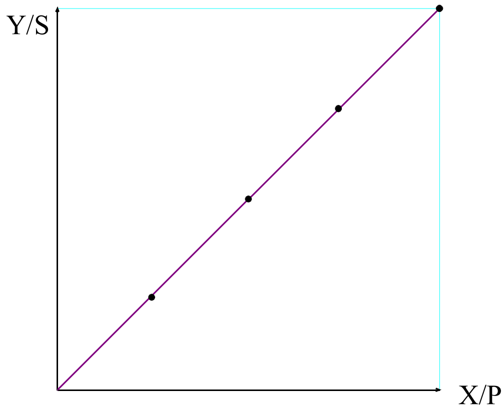
コミュニティ（1）



コミュニティ（2）

もし、格差のない均等な分布をしていると、下の表のようになります。

	均等コミュニティ			
所得（百万円）：y	2.5	2.5	2.5	2.5
人数（万人）：x	2.5	2.5	2.5	2.5
階層別所得総額：yx	6.25	6.25	6.25	6.25
上記累計：Y	6.25	12.50	18.75	25.00 ＝ S
Y／S	0.25	0.50	0.75	1
人数累計：X	2.5	5.0	7.5	10.00 ＝ P
X／P	0.25	0.5	0.75	1



この表に基づいて「ローレンツ曲線」を描くと右上のグラフになります。このように、均等分布を表現する曲線は「45 度線」になります。この「45 度線」を「均等分布線」と呼ぶことがあります。

① ローレンツ曲線（折れ線）と横軸、$X/P = 1$ で囲まれた部分の面積を求めます。
三角形と台形の面積を求めることになります。それぞれ

コミュニティ（1） $\frac{1}{2}\times 0.04\times 0.1+\frac{1}{2}\times(0.04+0.34)\times 0.4+\frac{1}{2}\times(0.34+0.84)\times 0.4+\frac{1}{2}\times(0.84+1)\times 0.1=0.406$

コミュニティ（2） $\frac{1}{2}\times 0.18\times 0.4+\frac{1}{2}\times(0.18+0.24)\times 0.1+\frac{1}{2}\times(0.24+0.34)\times 0.1+\frac{1}{2}\times(0.34+1)\times 0.4=0.354$

② 45 度線、横軸と $X/P = 1$ で囲まれた三角形の面積 0.5 から上記の値を引きます。45 度線とローレンツ曲線で囲まれた領域の面積が算出されます。
コミュニティ（1） 0.5－0.406＝0.094
コミュニティ（2） 0.5－0.354＝0.146

③ 0 から1の数値で指数化するために上記数値を2倍します。
コミュニティ（1） 0.094×2＝0.188
コミュニティ（2） 0.151×2＝0.292

この指数を「ジニ指数」と言います。

中間層が多いコミュニティ（1）のジニ係数は0.188、一方貧富の差が大きく中間層が薄いコミュニティ（2）で

は、0.302となって、両者の比較に一応の目安を与えてくれます。

なお、ローレンツ曲線を創案したM．O．ローレンツ(*1876* － *1959*)は、米国の経済学者で母集団の所得分布を見るために、*1905* 年にこの曲線を発表しました。

しかし、ジニ係数についてトマ・ピケティは以下のように言っています。

> 総合指数の問題点
>
> これまでに提起された問題に答えるために、国別の格差の歴史的動向の検証に入るが、その前に方法論上の問題を論じておかなければならない。特に、表 *7–1–7–3* には、検討されたさまざまな分配に対応するジニ係数も書かれている。ジニ係数―――イタリア人統計学者コンラド・ジニ(*1884* － *1965*)にちなんだ名称―――は、かなり一般的に使われる格差の総合指標で、公式報告や一般討論でもよく見かける。その仕組みにより０から１までの数値で表される指数だ。完全に平等であれば０になり、絶対的不平等、つまりごく少数のグループが入手可能なすべての資産を所有している場合は *1* になる。
>
> 実際には、現実社会における労働所得分布のジニ係数はおよそ *0.2* から *0.4* になり、資本所有分布では *0.6* から *0.9*、総所得の格差については、*0.3* から *0.5* になる。*1970* 年、*1980* 年代のスカンジナビアでは、労働所得のジニ係数は *0.19* で完全平等からそう遠くない。これとは逆に、ベル・エポツク期のヨーロツパにおける富の分布は、ジニ係数が *0.85* で、完全不平等に近い。
>
> これらの係数―――他にも、タイル係数などがある―――は役に立つこともあるが、問題も多い。それらはある分布が格差について言えることをすぺて―――階層の最底辺と中間層の格差、そして中間層と最上位、あるいは最上とその中のさらに上位の格差―――ひとつの数値指標に集約できると主張する。これは一見とてもシンプルで魅力的だが、いささか誤解を招くことは必至だ。多面的な現実を一次元の指標に集約しつつ、過度に事象を単純化せず、本来一緒に扱うべきでないことを一緒くたにしないですませるなど、実際には不可能だ。杜会的現実と格差の政治経済的重要性は、その分配の中での水準ごとにまったくちがうから、それらを個別に分析することが重要だ。加えて、労働の格差と資本の格差では、機能する経済メカニズムや規範によるその格差の正当化手段がまったくちがうのに、ジニ係数などの総合指数はそれを混同しがちだ。こういった理由から、格差を分析するならジニ係数のような総合指標を利用するよりも、総所得、国富におけるさまざまな十分位、百分位のシェアを示す分布表を使うほうがずっとよいと私は考えた。
>
> 分布表がもうひとつ優れているのは、既存の階層を構成するさまざまな社会グループの所得と富の水準にみんなが確実に注目するという点だ。そうした水準は、解読しにくい人為的な統計単位ではなく金額（もしくは当該国の平均所得と財産レペルに対するパーセント）で示されている。分布表は、社会的格差に関する具体的かつ直感的な理解を与えてくれると同時に、これらの問題の研究に使えるデータの評価とそれらのデータの限界を教えてくれる。これに対し、ジニ係数のような総合指標は、格差について抽象的で生気のない視点は与えてくれるが、現在の階層での自分の位置を人々が掴みにくくなってしまう（自分の位置を知るのはいつも有益なことだ。特に分配の百分位の上位に属しているくせにそれを忘れがちな人々はぜひやってほしい。経済学者たちはしぱしぱそういう人々の典型になっている)。指標はしぱしぱ、元データに例外や矛盾があるこを、あるいは国同士やちがう時期のデータが直接比較できないという事実（たとえぱ、分配の最上位が端折られたり、あるいは資本所得が国によって除外されたり含まれたりするため〕をあいまいにする。分布表を使うことで、一貫性を持つ透明性のある考察が避けられなくなるのだ。

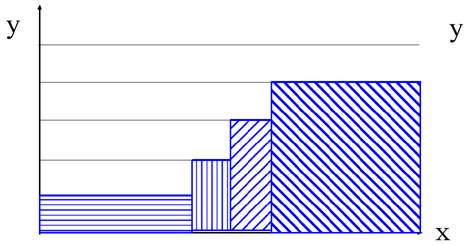
（下線は渡辺が付けました。）

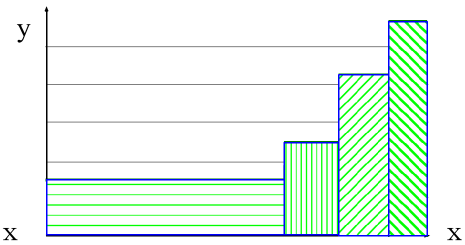
ピケティ氏の指摘のように、総合指数のみでは判断を誤ることが想定されます。如何にその具体例を示します。

前述のコミュニティ(2)とコミュニティ(3)を比較してみましょう。

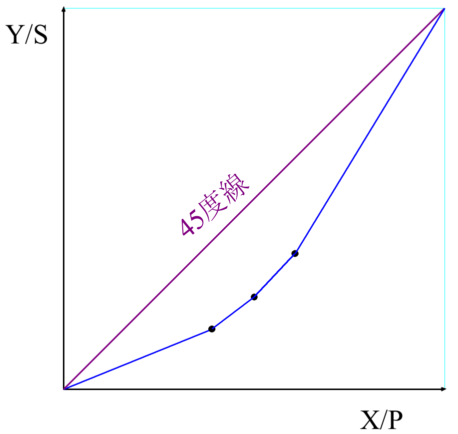
	コミュニティ(2)				コミュニティ(3)			
所得（百万円）: y	1	2	3	4	*1.5*	*2.5*	*4.2*	*5.6*
人数（万人）: x	4	1	1	4	*6.6*	*1.0*	*0.6*	*1.8*
階層別所得総額 : yx	4	2	3	１６	*10.0*	*2.5*	*2.5*	*10.0*
上記累計 : Y	4	6	9	２５＝S	*10.0*	*12.5*	*15.0*	*25.0* ＝S
Y／S	*0.16*	*0.24*	*0.34*	*1*	*0.4*	*0.5*	*0.6*	*1*
人数累計 : X	4	5	6	１０＝P	*6.6*	*7.6*	*8.2*	*10* ＝P
X／P	*0.4*	*0.5*	*0.6*	*1*	*0.66*	*0.76*	*0.82*	*1*

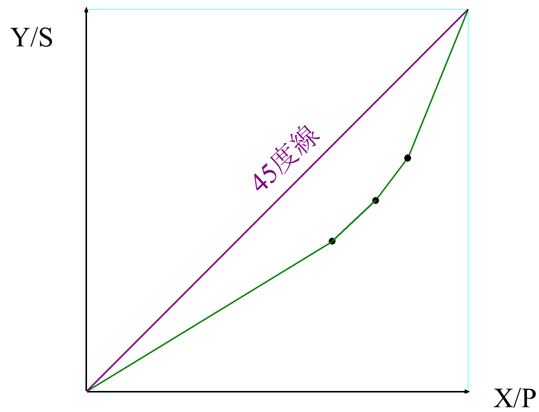
ｙとＸの関係をグラフにして見てみます。所得分布図とローレンツ曲線は以下のようになります。





コミュニティ(2)　　コミュニティ(3)





コミュニティ(2)　　コミュニティ(3)

① ローレンツ曲線（折れ線）と横軸、$X/P = 1$ で囲まれた部分の面積を求めます。

コミュニティ(2)　$\frac{1}{2}\times 0.18\times 0.4+\frac{1}{2}\times(0.18+0.24)\times 0.1+\frac{1}{2}\times(0.24+0.34)\times 0.1+\frac{1}{2}\times(0.34+1)\times 0.4=0.354$

コミュニティ(3)　$\frac{1}{2}\times 0.4\times 0.66+\frac{1}{2}\times(0.4+0.5)\times 0.1+\frac{1}{2}\times(0.5+0.6)\times 0.06+\frac{1}{2}\times(0.6+1)\times 0.18=0.354$

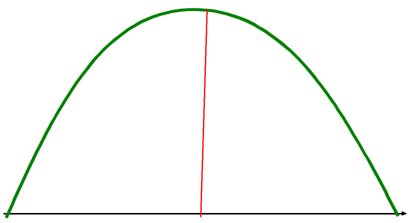
② *45* 度線、横軸と $X/P = 1$ で囲まれた三角形の面積 *0.5* から上記の値を引きます。*45* 度線とローレンツ曲線で囲まれた領域の面積が算出されます。上記数値が一致していますので

コミュニティ(2)とコミュニティ(3)は同じ値をとります。　0.5－0.354＝0.146

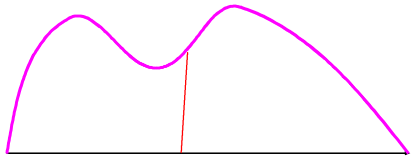
③ 0 から1の数値で指数化するために上記数値を2倍します。

コミュニティ(2)とコミュニティ(3)は同じ値をとります。　0.172×2＝0.292

明らかに、所得分布とローレンツ曲線が異なるのに、ジニ指数は一致しています。従って、所得分布の問題の場合、ジニ係数のみに依存して判断は出来ないことが分かります。
（実は**コミュニティ（３）のローレンツ曲線はコミュニティ（２）のそれを対称移動して作りました。**）
　一般に、統計資料の代表値の扱いについて、ピケティ氏が指摘することが当てはまります。有名な例として、得点分布が双峰分布の場合の平均値問題があります。



例(*1*)



例(*2*)

例(*1*)と例(*2*)は同じ平均値を持っています。例(*1*)の場合は資料が採られた集団はあるまとまりを持っていることが推測されますが、例(*2*)の場合は、この集団には二つの部分集団があることを覗わせます。従って、平均値だけで二つの集団を比較することは大事な点を見落とすことになります。

参考文献

１．高橋長太郎『所得分布の変動様式』一橋大学経済研究叢書、岩波書店、*1955* 年／ *p.　33* ～
２．大阪市立大学経済研究所『経済学辞典』、岩波書店、*1979* 年／ *p.　1381*
３．田中勝人『経済統計』現代経済学入門、岩波書店、*1996* 年／ *p.　47* ～
４．*Graham Upton, Ian Cook,* 白旗慎吾監訳『統計学辞典』、共立出版、*2011* 年／ *p.　448* ～

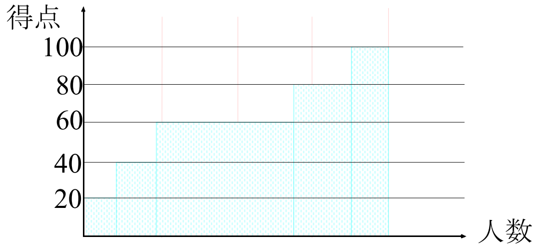
2. 授業書

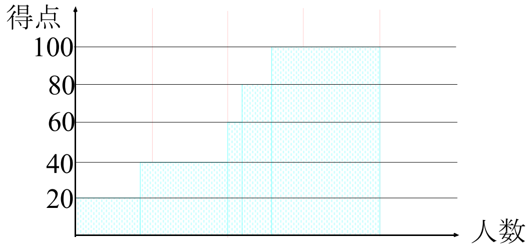
（0） ある学校の二つの学級で同じ試験が行われました。その結果は下表のようでした。

	クラス *A*					クラス *B*				
得点	*0* ～ *20*	*21* ～ *40*	*41* ～ *60*	*61* ～ *80*	*81* ～ *100*	*0* ～ *20*	*21* ～ *40*	*41* ～ *60*	*61* ～ *80*	*81* ～ *100*
人数	*4*	*5*	*19*	*7*	*5*	*8*	*12*	*2*	*4*	*14*

この二つのクラスを比較してみよう。

（1） 上の資料から柱状グラフ *histogram* を作成してみよう。





（2） 下表空欄を埋めながら平均値を求めてみよう。

得点		クラス *A*		クラス *B*	
階級	*X* 中央値	*f* 人数	*Xf* 積	*f* 人数	*Xf* 積
0 ～ *20*	*10*	4	*40*	*8*	*80*
21 ～ *40*	*30*	5	*150*	*12*	*360*
41 ～ *60*	*50*	*19*	*950*	*2*	*100*
61 ～ *80*	*70*	7	*490*	*4*	*280*
81 ～ *100*	*90*	5	*450*	*14*	*1260*
	合計	*40*	*2080*	*40*	*2080*

クラス *A* の得点平均値

m_A ＝ *2080 / 40 = 52*

クラスＢの得点平均値

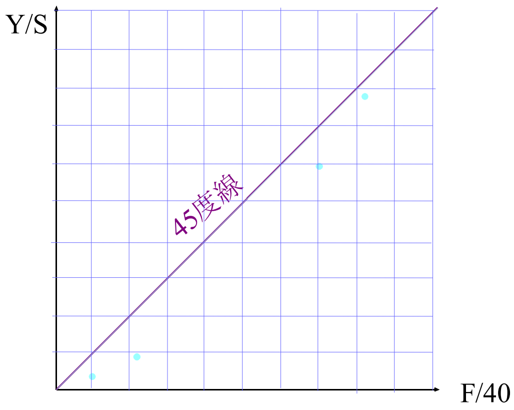
m_B ＝ *2080 / 40 = 52*

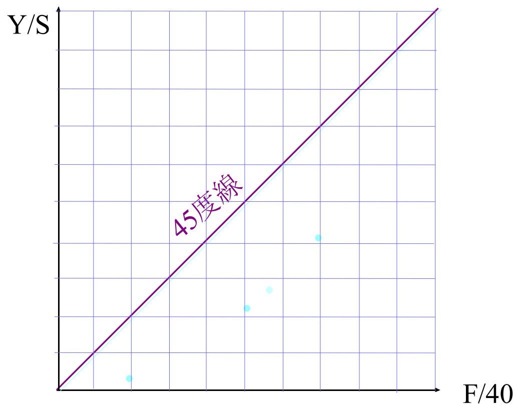
（3） アメリカの統計学者ローレンツ *Max C Lorenz*（*1876–1959*）は格差をグラフ化する手法を *1905* 年に編み出しました。彼の手順によって表を作り、それをグラフにしてみよう。

最初に以下の表を作ります。空欄を埋めてください。

	クラス *A*					クラス *B*				
得点：*X*（中央値）	*10*	*30*	*50*	*70*	*90*	*10*	*30*	*50*	*70*	*90*
人数：*f*	4	*5*	*19*	*7*	*5*	*8*	*12*	*2*	*4*	*14*
階層別得点総額：*Xf*	*40*	*150*	*950*	*490*	*450*	*80*	*360*	*100*	*280*	*1260*
上記累計：*Y*	*40*	*190*	*1140*	*1630*	*2080*	*80*	*440*	*540*	*820*	*2080*
Y ／ *S*（*S* は *Y* の最終値）	0.019	0.091	0.548	0.784	1	0.038	0.211	0.260	0.394	1
人数累計：*F*	*4*	*9*	*28*	*35*	*40*	*8*	*20*	*22*	*26*	*40*
F ／ *40*（各クラス人数は *40*）	0.1	0.225	0.7	0.875	1	0.2	0.5	0.55	0.65	1

（4）　***F*／*40*** を横軸に、***Y*／*S*** を縦軸にして（3）の表からグラフを作って見よう。



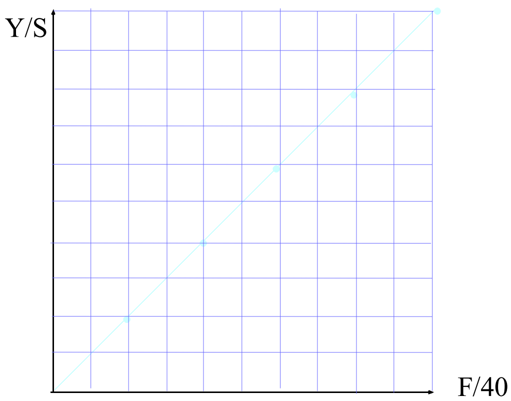


（5）　上のグラフから読み取れるものは何ですか？

（6）　もし、格差のない均等な分布をしていると、下の表のようになります。空欄を埋めてください。

	均等得点クラス				
得点：***X***	*60*	*60*	*60*	*60*	*60*
人数：***f***	*8*	*8*	*8*	*8*	*8*
階層別得点総額：***Xf***	*480*	*480*	*480*	*480*	*480*
上記累計：***Y***	*480*	*960*	*1440*	*1920*	*2400*
Y*／*S（***S*** は ***Y*** の最終値）	*0.2*	*0.4*	*0.6*	*0.8*	*1*
人数累計：***F***	*8*	*16*	*24*	*32*	*40*
F*／*40（各クラス人数は ***40***）	*0.2*	*0.4*	*0.6*	*0.8*	*1*

上の表に基づいてグラフを描いて見ましょう。



グラフの形状からどのような結論が得られますか？

（7）　イタリアの統計学者 *C.*ジニ：*Corrado　Gini*（*1884–1903*）は、格差の大小を一つの数値で表す方法を考えました。それを現代風にローレンツ曲線との関連でみると、次のような手法になります。

［ローレンツ曲線と *45* 度線で出来る領域の面積を比較する。そのために、横軸、縦軸ともに標準化して範囲を *0* から *1* までとする］というものです。

①　ローレンツ曲線と横軸、$F／40＝1$ の直線によって囲まれた多角形の面積を求める。

②　ローレンツ曲線と *45* 度線で出来る領域の面積を求める。

③　比較する数値「ジニ係数」を *0* から *1* にするために、②で求めた数値を *2* 倍する。

上記手順に従って二つのクラスの得点に関するジニ係数を求めてみよう。

解　①

クラスA　*0.39*

クラスB　*0.33*

②

クラスA　*0.11*

クラスB　*0.17*

③

クラスA　*0.22*

クラスB　*0.34*

（8）　ジニ係数は格差を正確に表すにふさわしいかどうか考察してみよう。

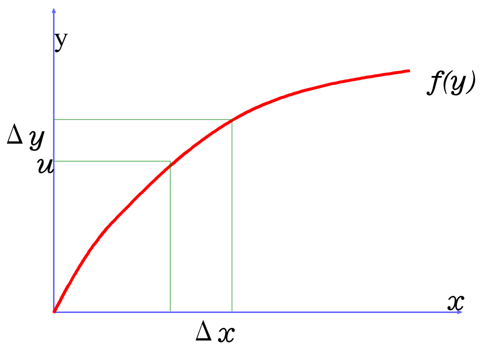
所得分布表は、人を「貧しい」から「富める」によって順番に並べたとき、所得 y を持つ人の「位置（順位）」が対応するものです。

所得分布表をグラフにすると一般には折れ線になります。

もし、滑らかな線のとき、 所得 y →人の位置 $x＝f(y)$ とします。 f は微分可能な関数です。

所得が u までの人数 p は、$p=\int_0^u f(y)dy$ 、これを $F(u)$ とします。総人数 P は、$P=\int_0^U f(y)dy$

ただし、U は最高の所得額。



所得が u から $u＋Δy$ の人々の総所得 $Δw$ は

$Δw \sim u×Δx$

ところで、$Δx＝F(u＋Δy)－F(u)$

よって、$Δx \sim F'(u)Δy＝f(u)Δy$

極限をとって、$dx＝f(u)dy$、

$dw＝udx=uf(u)dy$

所得 *0* から u までの人々の総所得合計 w は

$w=\int_0^u dw=\int_0^u yf(y)dy$

全所得総額 W は　$W=\int_0^U dw=\int_0^U yf(y)dy$　　ローレンツ曲線は、横軸に $p／P$

縦軸に $w／W$　として、点($p／P$、$w／W$) をとって作ります。

（9）あるコミュニティは、総人口 P が 2 単位人（1 単位 1 万）あり、所得分布は 0 の人から 2 単位円（1 単位 1 千万）まで広がっています。コミュニティの成員を貧しい方から富める方へ並べたときの位置 x は 所得金額が y に依って決まり、$y \to x = f(y) = y^2$ でした。

例題１：所得が 500 万円即ち 0.5 単位円までの人数はいくらでしょうか？

［解］ その人数を p とすると、$p = \int_0^{0.5} f(y)dy = \int_0^{0.5} y^2\,dy = \left[\frac{1}{3}y^3\right]_0^{0.5} = \frac{1}{3}\times 0.5^3 = \frac{0.125}{3} \cong 0.0417$

$p = 0.0417$ 単位人、すなわち 417 人

問題：所得が 8 百万円までの人数はいくらでしょうか？

［解］ その人数を p とすると、$p = \int_0^{0.8} f(y)dy = \int_0^{0.8} y^2\,dy = \left[\frac{1}{3}y^3\right]_0^{0.8} = \frac{1}{3}\times 0.8^3 = \frac{0.512}{3} \cong 0.1707$

$p = 0.1707$ 単位人、すなわち 1707 人

例題２：所得が 500 万円即ち 0.5 単位円までの人々が得た所得総額はいくらでしょうか？

[解]その[金額×人数]を w とすると

$$w = \int_0^{0.5} yf(y)dy = \int_0^{0.5} y^3\,dy = \left[\frac{1}{4}y^4\right]_0^{0.5} = \frac{1}{4}\times 0.5^4 = \frac{0.0625}{4} \cong 0.0156$$

$w = 0.0156$[単位円×単位人]、すなわち　0.0156 ×[1000 万円× 1 万人]＝ 0.0156[千億円人]

＝ 15 億 6 千「円人」

問題２：所得が 8 百万円までの人々が得た所得総額はいくらでしょうか。

[解]その[金額×人数]を w とすると

$$w = \int_0^{0.8} yf(y)dy = \int_0^{0.8} y^3\,dy = \left[\frac{1}{4}y^4\right]_0^{0.8} = \frac{1}{4}\times 0.8^4 = \frac{0.4096}{4} \cong 0.1024$$

$w = 0.1024$[単位円×単位人]、すなわち　0.1024 ×[1000 万円× 1 万人]＝ 0.1024[千億円人]

＝ 102 億 4 千「円人」