ログアウト
サイトトップへ
医学統計学

TOP > 医学統計学講座 > 第2回

目からウロコの医学統計学講座

株式会社サンテック 統計解析室室長
足立 堅一 先生

第2回: データの表示法と分布

～～～～～～ CONTENTS ～～～～～～

NEXT

印刷される場合には、こちら(PDF版)をご利用下さい。
PDFファイルをご覧になるには「推奨動作環境」をご覧ください。

PAGETOP

目からウロコの医学統計学講座

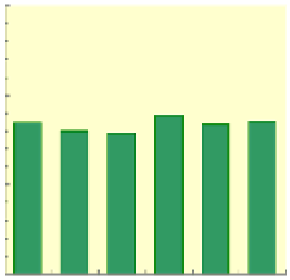
第2回：データの表示法と分布

株式会社サンテック　統計解析室室長
足立 堅一先生

NEXT ▶▶

今回は、データの背景に存在する「分布の形状」を知るための視覚的データ表示法として最も基本的な**histogram(柱状図)**、**幹葉図(stem and leaf)**などを解説する。
また、**「絶対」度数分布よりも「相対」度数分布が重要であること、「究極的相対度数分布」を「母集団」と考えることの薦めとその有用性について解説する。**

先ず、第1回の例題2で考察した視覚に訴える方法を学習しよう。



2.1 度数分布表とhistogram

例題1	サイコロを1000回繰り返し振ったときの、1～6の目の出方についての度数分布表とhistogramを作成すること。このとき、それぞれの目は何回出ると期待されるか。 解答

CONTENTS NEXT
PAGETOP

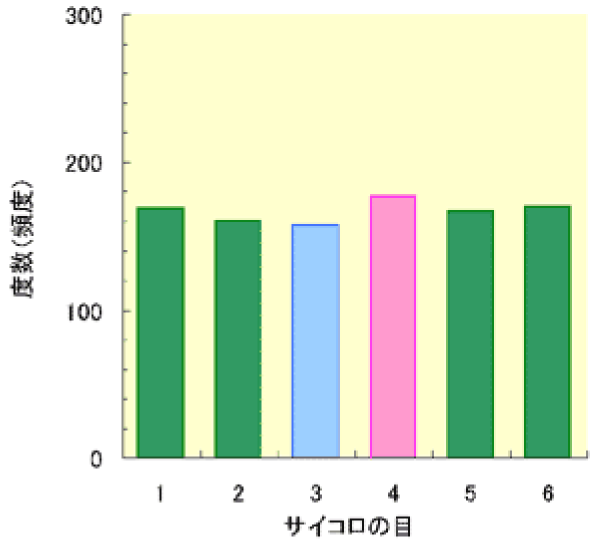
例題1　解答

各目の出ると期待される回数(期待度数)は1000/6＝167。
実例を表2-1(度数分布表)に示した。この場合は最低度数が「3の目」の157回、最高度数が「4の目」の177回。histogramも図2-1に示した。図では横軸が1～6の各目を、縦軸が度数(頻度、frequency)を示す。

表2-1　度数分布表

サイコロの目	度数 (頻度)	相対度数 (相対頻度)	累積度数	累積相対度数 (累積相対頻度)
1	169	0.169	169	0.169
2	160	0.160	329	0.329
3	**157**	0.157	486	0.486
4	**177**	0.177	663	0.663
5	167	0.167	830	0.830
6	170	0.170	1000	1.000

図2-1　histogram



CLOSE

ログアウト
サイトトップへ
医学統計学

TOP > 医学統計学講座 > 第2回

目からウロコの医学統計学講座

第2回：データの表示法と分布

株式会社サンテック 統計解析室室長
足立 堅一先生

BACK NEXT

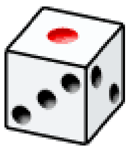
2.2 度数の相対化の必要性 ― 相対度数（相対頻度）

サイコロの例で考えると容易に理解できるが、1の目が10回出たと言っても、12回(12試行)中の10回であれば、サイコロに仕掛けがあるのではと疑いたくなるが、逆に60回中の10回であれば疑念は生じない。

これを2つのサイコロA・Bの比較の問題として捉えれば、極めて歴然とする。つまり、サイコロAもBも1の目が10回出たと言っても、分母、つまり繰り返し数、が同じであればまだしも、違えば比較不能である。分母を明示してAの10/12とBの10/60として初めて、比較する場が整う。Aでは0.83と異常に高い度数、Bは1/6で安心できる度数である。

つまり、**こうした事象の確率を比較するには、絶対度数では不都合で、相対度数化が必要になる。**換言すれば、絶対度数は、試行回数(繰り返し数)に依存するために、事象間の起こり易さ／難さを比較するときには不適であり、試行回数に依存していない指標としての相対度数に変換する必要がある。

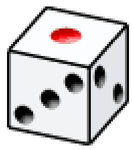
サイコロA：



が10回

12回振って

サイコロB：



が10回

60回振って

例題2	サイコロを1000回繰り返し振ったときの、1～6の目の出方についての度数分布表(例題1)より、それぞれの目の相対度数を計算し、これらの度数の和が1になることを確認すること。 解答

BACK CONTENTS NEXT

PAGETOP

例題2　解答

度数分布表

サイコロの目	度数 (頻度)	相対度数 (相対頻度)	累積度数	累積相対度数 （累積相対頻度）
1	169	**0.169**	169	0.169
2	160	**0.160**	329	0.329
3	157	**0.157**	486	0.486
4	177	**0.177**	663	0.663
5	167	**0.167**	830	0.830
6	170	**0.170**	1000	1.000
合　計		**1.000**		

CLOSE

目からウロコの医学統計学講座

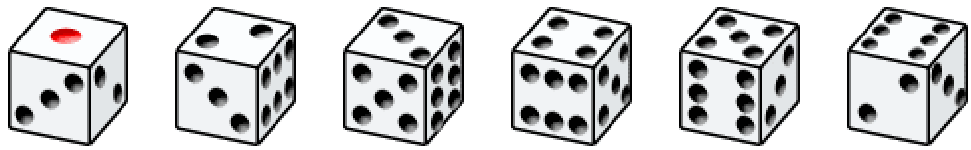
第2回：データの表示法と分布

株式会社サンテック　統計解析室室長
足立 堅一先生

BACK　NEXT

2.3 究極の相対度数分布 ― 確率密度関数(確率分布)

例題1においては、試行は1000回であった。このとき、それぞれの目の出方の相対度数は、最小で0.157、最大でも0.177となり、我々が想像・期待する0.167に十分に近い値(確率)と言える。
試行回数をさらに増大させて、1万回、10万回、. . . 、1億回、. . . 、1兆回、. . . とするのは、実行は困難としても、想像はいとも簡単である。そして、我々は試行回数を増やせば1/6に接近し、∞では1/6になることも容易に想像できる。
こうした「究極の」理論的度数分布を**確率密度関数(probability density function)・確率分布(probability distribution)**と把握すると理解し易い。また、**サイコロの目1～6の場合は、それぞれ一様に確率1/6となるので、「一様分布(uniform distribution)」と呼ばれる。**こうした理論分布は、一様分布でなくても、いろいろな形の分布でも誘導できる。



設問1

サイコロを4回振って、これを1試行としたときの、この4回の内で偶数の目が出る回数は0～4回が考えられる。2回が最もありそうで、0, 4回が最もありそうにない。0～4回の確率がどうなるか計算法を考えること。また、同じことを実際の試行をして実験する方法を考えてみること。設問なので余裕のある者は取り組むこと。

HINT 拙著「らくらく生物統計学、中山書店、1998年」の40p 表3.2を読んで計算してみること。実験する方法としては、実際のサイコロで実験しても良いが、パソコンを所有していれば、乱数関数と簡単なプログラム作成でもっとelegantに達成できる。

ここで、注目すべきは、**十分多くの繰り返しによる実験でもこうした理論分布を想像可能なだけでなくて、理論的にも誘導・確認可能である**ことである。これは、数学的確率(先験的確率)と統計的確率(経験的確率)の話とも関連する。確率の概念は、数学的さらには哲学的にその定義の妥当性や意味・意義についていろいろ議論されるようであるが、一般統計学教科書同様、本講座でも余り深入りしない。

BACK　CONTENTS　NEXT
PAGETOP

目からウロコの医学統計学講座

第2回：データの表示法と分布

株式会社サンテック　統計解析室室長
足立 堅一先生

BACK NEXT

2.4 確率分布 ー もう1つの観点からの「母集団」

究極的な相対度数分布を「母集団」と把握すると、しばしば利用される数学的発想に対する無理解や理解の困難さが氷解するであろう。数学者が想定する「母集団」とはしばしばこうした仮想的な「究極的な相対度数分布(確率分布)」である。先ず、設問1を例にとり、その発想の意味するところや意義や「母集団 vs. 標本」の構図などを学習する。

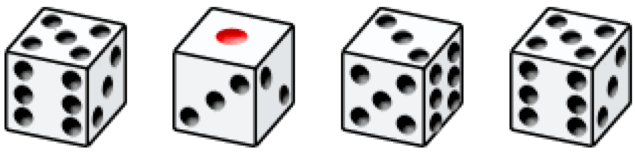
例題3	度数分布を「母集団」として想定するというが、それでは設問1の場合の「標本」とは何を指すか。「母集団」と想定する意義や利点は何か。 解答

BACK CONTENTS NEXT
PAGETOP

例題3　解答

我々が手にする「標本」とは、各試行の結果であり、この場合は4回繰り返し／試行と解釈できる。そしてしばしば、1試行が1実験・研究に該当する。
つまり、4回サイコロを振り、偶数が0回(＝奇数が4回)であれば、研究者はもしかしてサイコロに仕掛けがあるのではと疑い出すかも知れない。

この場合、さらに偶数の確率＝0.5を有効率0.5に、1繰り返しを1例と置換すれば、試験そのものになる。「うたい文句が有効率50%の薬なのに効いたのが4人中0例とはなんだ！？」と。
そのとき、偶数の確率＝0.5の理想的なサイコロで試行を無限回実施した相対度数を念頭において、それと1回の試行で得た結果とを照し合せれば、当該結果がどの程度珍しい出来事かを判断できる。
サイコロでの無限回試行は不可能だし、またサイコロも厳密には形や重心に多少の不均衡はあると考えられるが、重要な事実は、我々は試行することなく、理想的なサイコロでどのような結果になるかを想像し算出することさえ可能であるということだ。



CLOSE

目からウロコの医学統計学講座

第2回：データの表示法と分布

株式会社サンテック　統計解析室室長
足立 堅一先生

2.5　正規分布　—　連続量に関する最も頻出する最もpopularな分布

正規分布という言葉は、ほとんどの読者が一度は耳にしたであろう。「身長：♂ vs. ♀」を初めとして**様々な(自然)現象・事象が近似的に正規分布(左右対称形の釣鐘型の分布)することが経験的に知られている。**そのために、さらには数学的に一般に扱い易いとの理由から、以後学習する検定の多くもこれに立脚している。ここでは、次の例題で架空的人類♂♀について相対度数分布や幹葉図(stem and leaf)を作成しながら正規分布を学習しよう。

例題4　架空的人類の成人♂♀において**表2－2**のような身長のデータが得られた。度数分布図や幹葉図を作成すること。

表2－2　架空的人類♂♀の身長
50人の♂の身長(height)と50人の♀の身長の表。
♂♀別に意図的に低い順に並べ変えた。

♂	166.3	167.1	167.1	167.4	167.6	168.1	168.4	168.4	168.6	168.8
	168.8	168.9	169.0	169.2	169.2	169.2	169.2	169.3	169.5	169.5
	169.5	169.5	169.7	169.8	170.0	170.1	170.2	170.2	170.3	170.6
	170.6	170.7	170.8	170.9	170.9	170.9	170.9	171.3	171.7	171.7
	171.8	171.9	172.5	172.7	173.0	173.4	173.5	173.6	174.7	175.3
♀	161.5	162.7	163.4	163.6	163.7	164.0	164.0	164.1	164.6	164.7
	164.8	164.8	164.8	164.8	164.9	165.0	165.0	165.1	165.4	165.7
	165.7	165.7	166.0	166.0	166.1	166.1	166.2	166.2	166.3	166.4
	166.4	166.6	166.8	166.9	167.0	167.1	167.2	167.3	167.6	167.7
	167.7	167.7	167.7	168.1	168.2	169.1	169.6	169.8	170.1	170.2

解答

ここで注意すべきは先程のサイコロの度数分布と異なる箇所として、サイコロの離散量に対して身長は**連続量**であることであろう。つまり4回中「偶数の目」の出る回数は0～4回との意味は、4回中で「偶数の目」が1.5回出たなんてことはあり得ない。0,1,2,3,4の中間に落ちる値は全く存在しない。逆に、身長の160cmと161cmの間には無数の値が存在する。

最近ではExcelなどで自動的に描いてくれると思われるが、参考までに描き方の要点を示す。実例は例題4解答を参照のこと。

例題4　解答

架空的人類の成人♂♀をそれぞれ50例(表2-2)を抽出して、度数分布(図2-2、図2-3は絶対度数分布ではなく、50で除算して算出した相対度数分布である)と幹葉図(図2-4、2-5)を描いた。

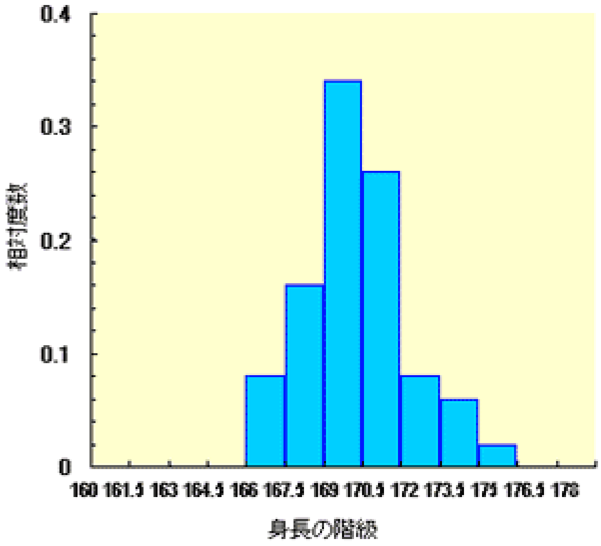


図2-2 histogram
50人の♂の身長(height)の図
横軸が身長の階級区分、縦軸が相対度数を示す

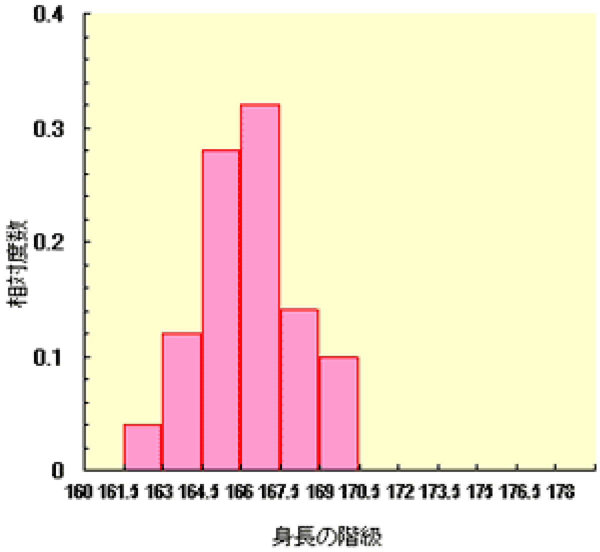


図2-3 histogram
50人の♀の身長(height)の図
横軸が身長の階級区分、縦軸が相対度数を示す

Quantiles (Def=5) (♂)

100%	Max	175.3	99%	175.3
75%	Q3	171.3	95%	173.6
50%	Med	170.05	90%	173.2
25%	Q1	169	10%	167.85
0%	Min	166.3	5%	167.1
			1%	166.3

Range	9
Q3-Q1	2.3
Mode	169.2
N	50
Mean	170.246

Stem	Leaf	#
175	3	1
174	7	1
173	0456	4
172	57	2
171	37789	5
170	0122366789999	13
169	022223555578	12
168	1446889	7
167	1146	4
166	3	1

図2－4　幹葉図
50人の♂の身長(height)の図
横軸が度数(頻度)、縦軸(Stem Leaf)が身長の階級区分を示す

Quantiles (Def=5) (♀)

100%	Max	170.2	99%	170.2
75%	Q3	167.3	95%	169.8
50%	Med	166.1	90%	168.65
25%	Q1	164.8	10%	163.85
0%	Min	161.5	5%	163.4
			1%	161.5

Range	8.7
Q3-Q1	2.5
Mode	164.8
N	50
Mean	166.122

Stem	Leaf	#
170	12	2
169	168	3
168	12	2
167	012367777	9
166	001122344689	12
165	0014777	7
164	0016788889	10
163	467	3
162	7	1
161	5	1

図2－5　幹葉図
50人の♀の身長(height)の図
横軸が度数(頻度)、縦軸(Stem Leaf)が身長の階級区分を示す

CLOSE

目からウロコの 医学統計学講座

第2回：データの表示法と分布

株式会社サンテック 統計解析室室長
足立 堅一先生

BACK NEXT

2.6 連続量に関するデータの表示法
1） 幹葉図(stem and leaf)とその作成法

(1)データの根幹的部分で先ず大きく分類した後、(2)細部について個数の分だけ細部の値をpick-upして積み上げる。
♀のデータについて述べると、(1)整数部の値で分類(これが**幹(stem)**)する。(2)小数点1桁の値をそれぞれの落ちる分類において積み上げる(これが**葉(leaf)**)。例えば、163cmは、163.4、163.6、163.7であるので、4、6、7を積み上げる。したがって、積み上げた高さが頻度になる。

Stem	Leaf	#
170	12	2
169	168	3
168	12	2
167	012367777	9
166	001122344689	12
165	0014777	7
164	0016788889	10
163	467	3
162	7	1
161	5	1

BACK CONTENTS NEXT
PAGETOP

目からウロコの医学統計学講座

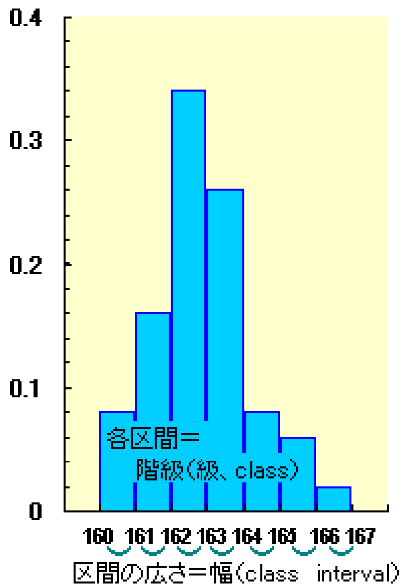
第2回：データの表示法と分布

株式会社サンテック 統計解析室室長
足立 堅一先生

BACK NEXT

2.6 連続量に関するデータの表示法
2） 度数分布図とその作成法

身長データの性格上、160cm代、161cm代、…のように適当に区間を設定する必要がある。そうした各区間を**階級(級、class)**、区間の広さを**幅(class interval)**という。160cm代、161cm代、…の例では幅は1cmである。



例題5	架空的人類の成人♀の(相対)度数分布において階級の数、幅はいくらか。 解答

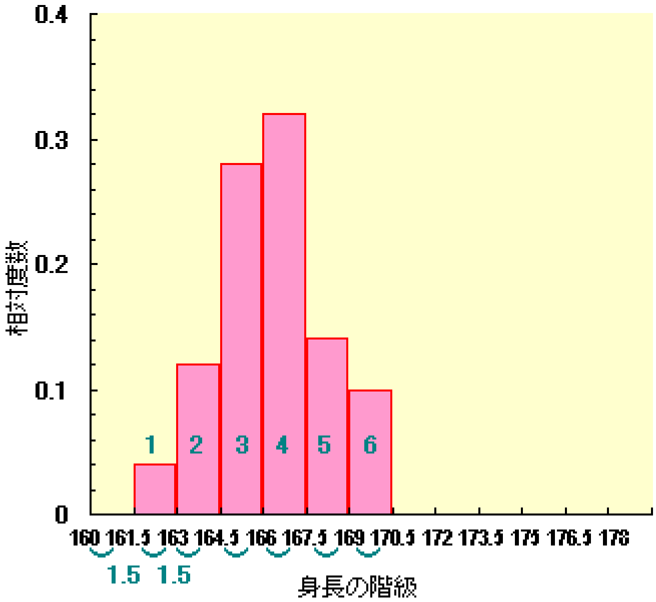
例題6	因みに、架空的人類の成人♀の(相対)度数分布における階級の数を例題5解答中の式(1)で計算すること。 解答

BACK CONTENTS NEXT
PAGETOP

例題5　解答

階級数は6、幅は1.5cm単位である。
なお、階級数nはデータの総数Nの大きさに応じて、7～15程度とするのが一般的と言われる。目安としては、次の**Sturges' ruleなる経験則**もある。

$$n=1+(\log_{10} N / \log_{10} 2) \quad \cdots\cdots(1)$$



CLOSE

例題6　解答

$n = 1 + (\log_{10} 50 / \log_{10} 2) = 6.6$。

CLOSE

目からウロコの医学統計学講座

第2回：データの表示法と分布

株式会社サンテック 統計解析室室長
足立 堅一先生

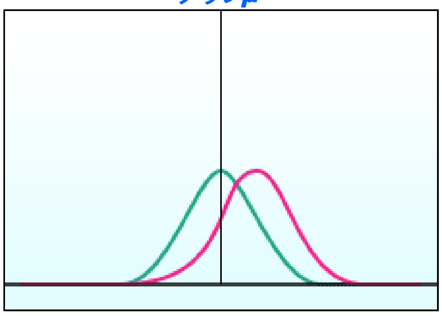
BACK

2.7 正規分布の形状とそれを決めるもの(parameter)

ー 位置のparameterと広がりのparameter

度数分布図からも判明する通り、正規分布の形状は釣鐘型である。分布の頂点が中央にあり、それを軸として左右対称に裾を引いている。分布の頂点を与える変量の値が平均値で、これが位置のparameter（母数）と呼ばれ、♂♀の身長が違えばこの頂点が乖離・移動する。事実、♂の平均値170.2cmと♀の平均値166.1cmとでは4cm程乖離して見える。

グラフμ

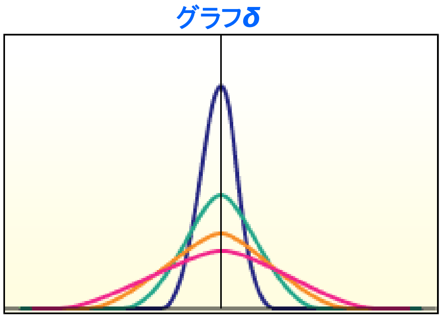


μ=0 μ=1 ALL

ボタンをクリックすると画像が変わります。

裾の長さはバラツキを反映し、長い程バラツキが大きいことを意味し、これは広がりのparameterである分散(variance)を開平した標準偏差(standard deviation)などで表現される。

グラフδ



δ=1 δ=2 δ=3 δ=4 ALL

ボタンをクリックすると画像が変わります。

BACK CONTENTS

PAGETOP