

4 章 度数分布

4.1 度数分布表

3.4 節で予告をしましたが、ここでは得られた量的データがどのようなばらつきをしているのか調べる方法について例を用いて考えましょう。

今 20 人の学生の試験を調べたら、以下の成績でした。これからデータのばらつきの状況を調べてみましょう。このばらつきの状況を分布といいます。

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

まず点数をいくつかの階級に分け、その階級ごとに何人の人が入っているか調べます。以下の表 4.1 を見て下さい。

表 4.1 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		

この表では分割の幅を 10 点に設定して、区切りの良い位置で階級に分けています。各階級を代表する値を階級値といい、通常その階級の中央の値によって表わします。例えば、階級が $60 < x \leq 70$ の場合、階級値は $(60+70)/2=65$ となります。

この分割に対して、それぞれの階級にいくつのデータが属するか(何人の人が含まれるか)を度数という名称で表わしています。もちろん各階級の度数の合計は全体の人数 20 人となります。このように階級ごとに度数を表示した表を度数分布表といいます。度数分布表をさらに分かり易くするために、相対度数、累積度数、累積相対度数等を追加する場合もあります。相対度数は全度数に対するある階級の度数の割合を表わし、累積度数はその階級以下の度数の合計です。累積相対度数は累積度数の全度数に対する割合を表わします。

問題

Excel を用いて以下のような度数分布表（上の例で与えたもの）を描け。但し、網掛けの部分には計算式を用いること。

階級	度数	相対度数	累積度数	累積相対度数
$50 < x \leq 60$	4	20.0%	4	20.0%
$60 < x \leq 70$	8	40.0%	12	60.0%
$70 < x \leq 80$	5	25.0%	17	85.0%
$80 < x \leq 90$	3	15.0%	20	100.0%
計	20	100.0%		

解答

省略

4.2 ヒストグラム

度数分布表は分かり易い集計の方法ですが、さらに視覚化するために、ヒストグラムと呼ばれるグラフを描く場合があります。ここでは上の例を元にこのグラフの特徴を説明しましょう。

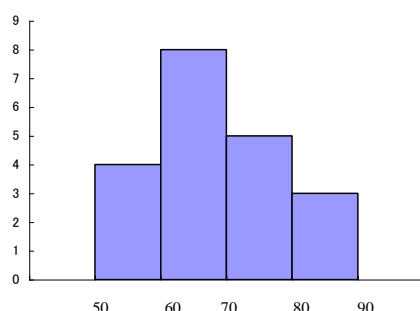


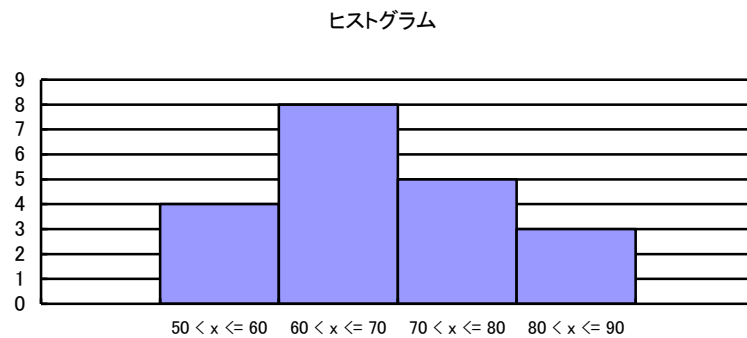
図 4.1 試験結果のヒストグラム

図 4.1 を見て下さい。横軸には変数の値（テストの点数）、縦軸には度数を取ります。横軸の上に階級幅で度数に見合った高さの棒を立てて行きます。階級間には隙間の無いようにします。これがヒストグラムと呼ばれるグラフです。ヒストグラムは一見棒グラフのように見えますが、少し違うところを持っています。例えば 10 点区切りだったデータを、何らかの理由で 70 点から 90 点まで 20 点区切りにして、1 つの階級にしなければならなくなったとします。そのとき棒の高さはいくらになるのでしょうか。通常の棒グラフだと 2 つの度数を合わせて、8 としたいところですが、これではもとのグラフから大きくかけ離れてしまいます。元のグラフの形を考慮するなら半分の 4 にするのが妥当なように思えます。階級幅が倍になった分だけ、高さを半分にするというのは、ヒストグラムが高さを重視しているのではなく、棒の面積を重視していることに相当します。ヒストグラムは通常の棒グラフと異なり、面積で度数を表わすグラフとなっているのです。

さて、これまでは階級幅については初めから与えられたものとしてきましたが、ヒストグラムを描く際、見た目に分かり易いものを描きたいものです。階級について、全データ数を n とした場合、分割数を \sqrt{n} 程度にするのが見易く描けるようです。

問題

前節の問題で与えた度数分布表を利用して、Excel を用いて以下のようなヒストグラムを描け。



解答

省略

問題

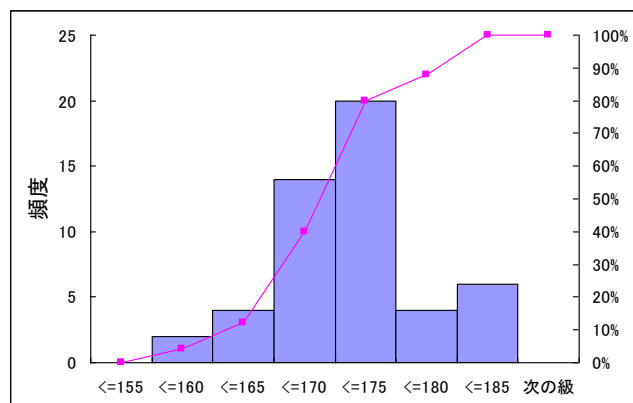
以下の 50 人の身長データ(cm)で度数分布表を作り、ヒストグラムを描け。ここでは Excel の分析ツールを利用せよ。

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%



注) データ区間に書かれた不等号は後で追加した。