

経済情報処理

平均 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

例：100 人に対してテストを行った結果

平均点は、全員の点数を足し合わせ、それを人数（100）で割ったもの
全員が 50 点であった場合も、100 点が 50 人、0 点が 50 人の場合も平均点は 50 点
得点の分布が次のような場合

得点	100	90	80	70	60	50	0
人数	10	10	10	10	10	10	40

平均点は

$$\frac{100 \times 10 + 90 \times 10 + 80 \times 10 + 70 \times 10 + 60 \times 10 + 50 \times 10 + 0 \times 40}{100} = \frac{4500}{100} = 45$$

50 点を取った人は平均点を上回ってはいるが、順位は 51 位（51～60 位が同点）

50 位の人の得点を考える場合も → メジアン(中央値)

厳密には、100 人の場合、真ん中の人はいないので、50 番目の人の得点と 51 番目の人の得点の平均をメジアンとする。 → この場合メジアンは 55 点

AVERAGE(数値 1, 数値 2, ...)

数値 1, 数値 2, ... 平均を求める数値データを指定。引数は 30 個まで。数値 はセル範囲に対する参照であってもかまわない。

AVERAGE は対象が文字列等の場合は無視され、AVERAGEA では文字列の値は 0 として計算される。

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

それぞれの値が、平均からどのくらい離れているか → $x_i - \bar{x}$

この平均を考えればよいのだが、どんな値の場合も $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) = 0$ となってしまう。

絶対値(注)を使う方法も $\frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$ → 数学的な取り扱いが面倒

そこで、絶対値の代わりにそれぞれの 2 乗の平均とする(どんな実数も 2 乗すれば正または 0)。

結局分散は、 $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ と定義される。

VAR.P(数値 1, 数値 2, ...) (2003 以前は VARP)

数値 1, 数値 2, ... 母集団全体に対応する数値を指定。引数は 30 個まで。数値はセル範囲に対する参照であってもかまわない。数値として文字列、論理値、空白セルの参照を指定すると、エラーになる。

VAR.S (2003 以前は VAR) という関数もあるが、こちらは n ではなく $n-1$ で割ったもの。また、文字列データを 0 として計算する VARPA、VARA という関数もある。

標準偏差 $\sigma = \sqrt{\sigma^2}$

分散は 2 乗をしたものの平均をとっているのので、元の値に比べ、大きなものとなる。同じ程度の大きさとなるように、分散の平方根をとったものが標準偏差である。

STDEV.P(数値 1, 数値 2, ...) (2003 以前は STDEVP)

使い方は VAR.P と同様。

これについても、STDEV.S、STDEV.PA、STDEV.SA があり、意味は分散と同様。

変動係数 $\frac{\sigma}{\bar{x}}$

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

データの標準化 $u_i = \frac{x_i - \bar{x}}{\sigma}$

データは平均が 0、分散（標準偏差）が 1 となるように変換される。

偏差値は、平均が 50、標準偏差が 10 となる用にしたもの。

$$u_i = \frac{x_i - \bar{x}}{\sigma} \times 10 + 50$$

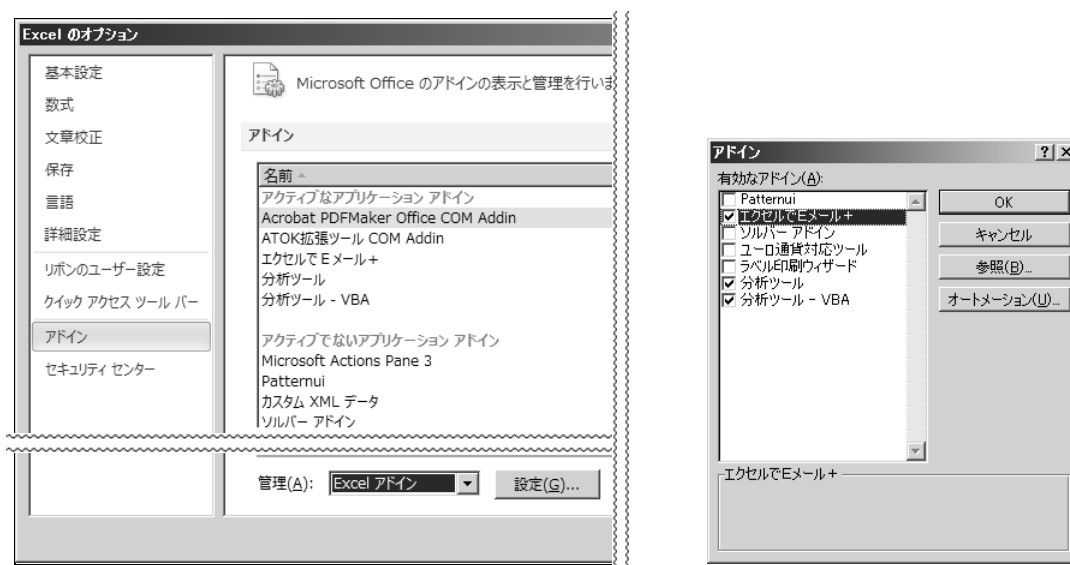
STANDARDIZE(x, 平均, 標準偏差)

X 標準化変量を計算する数値
平均 対象となる分布の算術（相加）平均
標準偏差 対象となる分布の標準偏差

ヒストグラム（度数分布）

ヒストグラムをエクセルで作成するには、始めに以下の作業を行う必要がある。

ファイル オプション をクリックすると下図左に示す Excel のオプション というダイアログボックスが表示される。この左側部分で **アドイン** を指定する（下図左はこの状態）。これにより表示されるものの下部に、**管理(A)** というボックスがあるので、その内容を **Excel アドイン** とし、**設定(G)...** のボタンをクリックする。これで下図右に示す **アドイン** というダイアログボックスが表示されるので、**分析ツール** 及び **分析ツール - VBA** という部分にチェックマークを付け、**OK** をクリックする。この作業の結果、**データ** のリボン右端に **データ分析** というものが表示されるようになる。



準備

ヒストグラム作成の対象となるデータを準備しておくことはいままでの準備でもないであろう。このデータは一般的に 1 列に間を置かないで並べておく。データはどのような順に並んでも問題ない。

データとは別の場所にどのようなデータ区間（階級）でヒストグラムを作成するかを指定する。例えばある列に上から順に、 $a_0, a_1, a_2, a_3, \dots, a_{n-1}, a_n$ という値を入力しておいたとする。後に、これをデータ区間として指定した場合には、対象データを、 a_0 以下、 a_0 より大きく a_1 以下、 a_1 より大きく a_2 以下、 a_2 より大きく a_3 以下、 \dots 、 a_{n-1} より大きく a_n 以下、 a_n より大きいという区間ごとに度数（出現回数、頻度）を計算しヒストグラムが作成されることになる。

注意. 例えば 100 点満点の試験の点数に関するヒストグラムを作る際、1 桁、10 点台、20 点台、 \dots といった区切りでヒストグラムを作るのであれば、データ区間を 10、20、30、 \dots などとしてはダメで（これでは 0~10、11~20、21~30、 \dots となってしまう）、9、19、29、 \dots としておかなければならない。

ヒストグラムの作成

上記の準備が済んだならば、いよいよヒストグラムの作成である。データにある分析ツールをクリックすると右のダイアログボックスが表示されるので、左側のボックスからヒストグラムを選択しOKをクリックする。これにより右に示すヒストグラムのダイアログボックスが表示される。この図ではボックスに書込がなされているが、始めはこれらのボックスは全て空欄となっている。

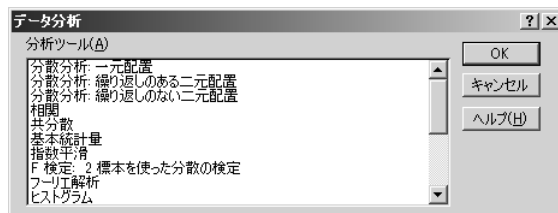
まず、入力範囲(I)のボックスに、ヒストグラムを作成しようとしているデータの範囲を入力する。入力方法としては、このボックスをクリックし、ボックス内にカーソルが表示されるようにしてから、データが入っているセルをドラッグし、範囲指定を行えばよい。

データ区間(B)のボックスも、入力範囲(I)のボックスと同様にして、データ区間を表すセル範囲を指定すればよい。

出力オプションの部分では、まず、出力先(O)、新規又は次のワークシート(P)、新規ブック(W)のいずれかを選択する。出力先(O)を選択した場合には、既存のシートの指定した場所に結果が表示されるようになる。具体的には、出力先(O)を選択した上で、その右のボックスをクリックしてカーソルを移動させてから（出力先を選択しただけでは、カーソルの移動が行われないので注意すること）、適当なセルを指定（ワークシート上のそのセルをクリックすればよい）すれば、その位置が左上となるように出力が行われる。通常はこれを使えば良いだろう。

最後に、パレート図(A)、累積度数分布の表示(M)、グラフ作成(C)について、必要なものにチェックマークを付ける（その部分をクリック）。パレート図を指定した場合には、度数の多い順にデータ区間を並べ替えたものが出力され、後述するグラフもその順序となる。累積度数分布の表示では、順に（パレート図を指定した場合には、度数の多い順、指定していない場合には、データ区間の並べられた順）累積度数の百分率が出力され、グラフを表示させた場合には、累積度数の百分率のグラフも折れ線グラフとして表示される。グラフ作成を指定すると、ヒストグラムが表示される。

以上の指定を行った上で、OK ボタンをクリックすれば、指定された位置に、度数分布表及びグラフ（グラフ作成が指定され場合のみ）が表示される。グラフについては、当初の表示は小さいので、適当な大きさに拡大する必要がある。また、見やすくなるように適当にグラフの設定を変えた方がいいだろう。

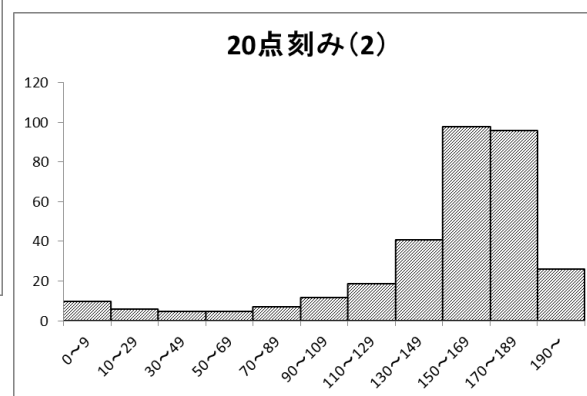
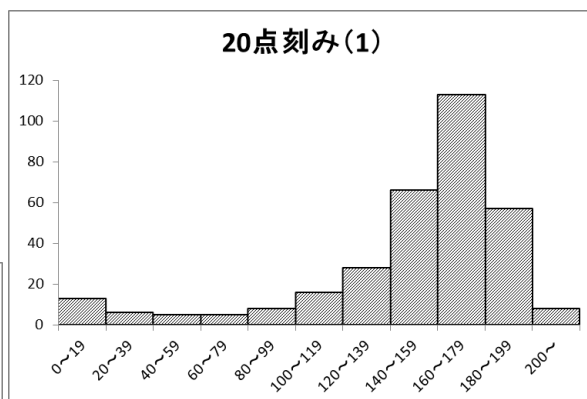
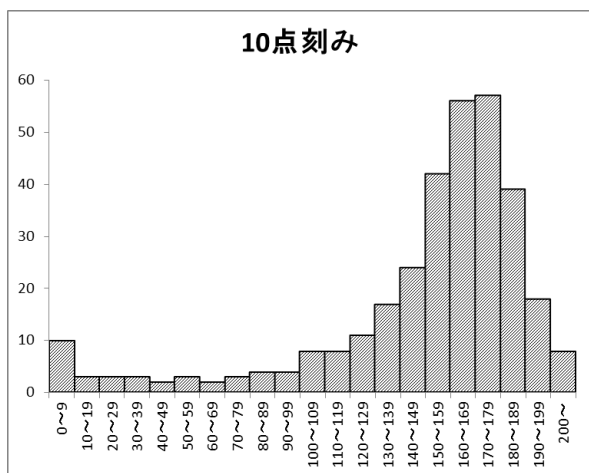


10点刻み

20点刻み(1)

20点刻み(2)

データ区間	頻度	構成比	データ区間	頻度	構成比	データ区間	頻度	構成比
0～9	10	3.1%	0～19	13	4.0%	0～9	10	3.1%
10～19	3	0.9%	20～39	6	1.8%	10～29	6	1.8%
20～29	3	0.9%	40～59	5	1.5%	30～49	5	1.5%
30～39	3	0.9%	60～79	5	1.5%	50～69	5	1.5%
40～49	2	0.6%	80～99	8	2.5%	70～89	7	2.2%
50～59	3	0.9%	100～119	16	4.9%	90～109	12	3.7%
60～69	2	0.6%	120～139	28	8.6%	110～129	19	5.8%
70～79	3	0.9%	140～159	66	20.3%	130～149	41	12.6%
80～89	4	1.2%	160～179	113	34.8%	150～169	98	30.2%
90～99	4	1.2%	180～199	57	17.5%	170～189	96	29.5%
100～109	8	2.5%	200～	8	2.5%	190～	26	8.0%
110～119	8	2.5%	合計	325	100.0%	合計	325	100.0%
120～129	11	3.4%						
130～139	17	5.2%						
140～149	24	7.4%						
150～159	42	12.9%						
160～169	56	17.2%						
170～179	57	17.5%						
180～189	39	12.0%						
190～199	18	5.5%						
200～	8	2.5%						
合計	325	100.0%						



課題 5

提出期限 6月2日(金) 午後2時まで

課題 2 で用いた成績.xlsx を読み込み、以下のことを行い、学務情報システムを使って完成したファイルを提出せよ。保存する際はファイル名を「基本統計量」という名前に変更しておくとい。

1. 修正点の部分が正しくない場合はそれを直しておく。
2. 素点及び修正点について、平均、分散 1 (データそのものの分散)、分散 2 (データが標本となっていると考え、母集団の分散)、標準偏差 1、標準偏差 2 (1、2 は分散と同様)、標準偏差 1 を用いた変動係数を求める。
3. 素点と修正点それぞれについて、標準偏差 1 を用いてデータの標準化を行ったもの及び偏差値を求める。
4. 素点について 10 点刻み、20 点刻み (1) (20 点未満、20 点以上 40 点未満、40 点以上 60 点未満、… としたもの) 及び 20 点刻み (2) (10 点未満、10 点以上 30 点未満、30 点以上 50 点未満、… としたもの) を求めよ。各データ区間の構成比も求めよ。グラフも作成し、グラフについては、見栄えの良いように手を加えておく。

	A	B	C	D	E	F	G
1	NO.	素点	修正点	素点 (標準化)	修正点 (標準化)	素点 (偏差値)	修正点 (偏差値)
2	001	168	74	0.416	0.383	54.2	53.8
3	002	98	45	-1.123	-1.057	38.8	39.4
4	003	172	76	0.504	0.482	55.0	54.8
5	004	142	64	-0.156	-0.114	48.4	48.9
6	005	141	63	-0.178	-0.163	48.2	48.4
7	006	169	74	0.438	0.383	54.4	53.8
8	007	193	86	0.966	0.979	59.7	59.8
9	008	188	84	0.856	0.879	58.6	58.8
10	009	168	74	0.416	0.383	54.2	53.8
11	010	169	74	0.438	0.383	54.4	53.8

325
326	324	151	67	0.042	0.035	50.4	50.4
327	325	164	72	0.328	0.284	53.3	52.8
328							
329	平均	149.1	66.3				
330	分散1	2067.2	405.6				
331	分散2	2073.6	406.8				
332	標準偏差1	45.5	20.1				
333	標準偏差2	45.5	20.2				
334	変動係数	0.3	0.3				