機械システム応用実験

～データの統計的処理とデータの考察について～

I.　データの統計的処理

まず、データの統計的処理について学ぶ。ある物理量を測定する場合、測定値の誤差を考慮する必要がある。測定値の誤差には、間違い（mistake）や系統的誤差（systematic error）などがあるが、間違いや系統的誤差が無い場合にも、不明の原因によって生じる値のばらつき、すなわち偶然誤差（accidental error）が生じる。一般的に値のばらつきは、正規分布となる場合が多いので、以下では正規分布について学ぶが、その前に正規分布で重要となる平均値と標準偏差の考え方を学ぶ。

(1) 身長の平均値と標準偏差

1-1　クラス全員の身長について、平均値$\bar{x}$と標準偏差σを求めよ。平均値$\bar{x}$と標準偏差σの算出方法については、配付した資料を参照せよ。算出にあたっては、Excel などのソフトウェアを積極的に使用すること。

1-2　1-1 について、5cm 単位で級に分けた度数分布表を作成し、ヒストグラムを描け。

(2) 抵抗値の平均値と標準偏差

データのばらつきの分布と正規分布の関係を調べるために、以下の課題について調べよ。

2-1　表 1 は市販の 200Ωのカーボン抵抗を 1990 個､テスタにより抵抗値を実測した結果である。抵抗 15 個の抵抗値を実測し、表 1 に加えよ。（小数点第 1 位まで計れるように、テスタのレンジを適切に選択せよ。）またこの抵抗値のヒストグラムを示せ。

2-2　表 1 について、累積分布曲線を示せ。

2-3　表 1 より、このカーボン抵抗の抵抗値の平均値$\bar{x}$と標準偏差σを求めよ。

2-4　求めた平均値$\bar{x}$と標準偏差σをもとに、正規分布の曲線（理論値）を 2-1 で描かせたヒストグラムと同一のグラフに示せ。

2-5　表 1 のデータに基づき、平均値$\bar{x}$を中心に全体の 68.3%となる抵抗値と 95.4%となる抵抗値の級の範囲を求めよ。

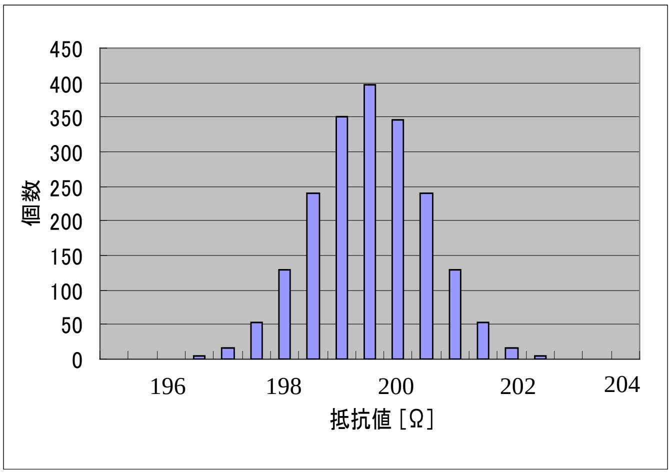
2-6　2-4 で求めた正規分布曲線について、$\bar{x}\pm\sigma$および$\bar{x}\pm 2\sigma$となる境界の抵抗値を算出せよ。

ヒント：正規分布の理論値は以下の式で表される。

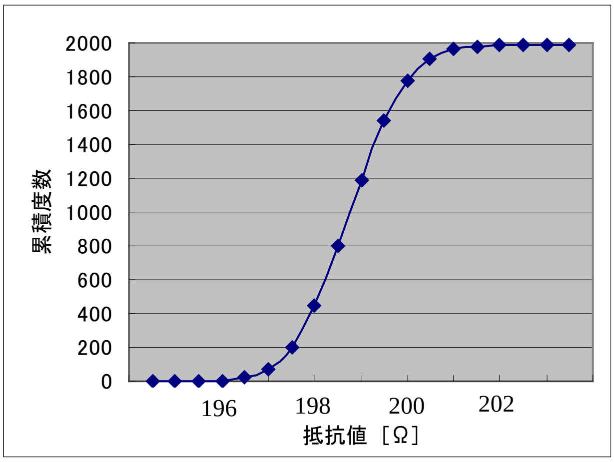
$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}exp\left(-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2\right)$$

表 1　市販の 200Ωの抵抗の抵抗値の分布

抵抗値 [Ω]	累積数	頻度数	頻度分布	100	200	300
204.0						
203.5						
203.0			一			
202.5			𝍵			
202.0			正正正丅			
201.5			正正正正正正正正正正 下			
201.0			正正正正正正正正正正 正正正正正正正正正正	正正正正正正丅		
200.5			正正正正正正正正正正 正正正正正正正正正正	正正正正正正正正正正 正正正正正正正正正正	正正正正正正正正一	
200.0			正正正正正正正正正正 正正正正正正正正正正	正正正正正正正正正正 正正正正正正正正正正	正正正正正正正正正正 正正正正正正正正正正	正正正正正正正正正正 一
199.5			正正正正正正正正正正 正正正正正正正正正正	正正正正正正正正正正 正正正正正正正正正正	正正正正正正正正正正 正正正正正正正正正正	正正正正正正正正正正 正正正正正正正正正丅
199.0			正正正正正正正正正正 正正正正正正正正正正	正正正正正正正正正正 正正正正正正正正正正	正正正正正正正正正正 正正正正正正正正正正	正正正正正正正正正正
198.5			正正正正正正正正正正 正正正正正正正正正正	正正正正正正正正正正 正正正正正正正正正正	正正正正正正正正一	
198.0			正正正正正正正正正正 正正正正正正正正正正	正正正正正下		
197.5			正正正正正正正正正正 下			
197.0			正正正丅			
196.5			𝍵			
196.0			一			
195.5						
195.0						



ヒストグラムの例



累積度数分布の例

II. データの考察

どのような関係にあるか不明な物理量について、その物理量をグラフに描き、関係を推測することは、実験結果の考察の第一歩である。また、推測により仮定した関係を物理的なモデルにより説明することが、最も多く見られる科学論文の作成手法である。今回は、音階と周波数特性を例に、このような作業を行ってみよう。（関係を知っている人も知らないつもりで考えること。）

(1) 実験 1（代表者に実験してもらい、代表者のデータを使ってグラフを作成する。）

1-1 電子楽器により、“ド”から 2 オクターブ高い“ド”までの、25 個の鍵盤（n = 1～25 の番号をつける。）の音声信号波形を観測し、それぞれの周期 T(n)を測定する。音階の番号 n と周期 T(n)の関係をまず線形グラフに示せ。また、音階の番号 n と周波数 *f*(n)の関係についても線形グラフに示せ。

1-2 1-1 で描いたグラフから推測して、音階の番号 n と周波数 *f*(n)の関係を最も良く表すことのできるグラフを示せ。

1-3 音階の番号 n と周波数 *f*(n)の関係を数式により示せ。

(2) 実験 2（代表者に実験してもらい、代表者のデータを使ってグラフを作成する。）

2-1 ギターのフレットはキーボードの鍵盤と同様の働きをする。フレットの番号を n として、各フレットにおける弦の長さ L(n)を測定し、フレットの番号（=音階の番号）n と弦の長さ L(n)の関係を線形グラフに示せ。

2-2 周波数と弦の長さの関係を示すグラフを描き、その関係を数式により示せ。

2-3 弦の振動について物理モデルを考案し、2-2 の関係が正しいことを示せ。（次ページのヒントを参照せよ。あくまで考察すること。“教科書などで調べた公式があてはまるから”というのは、考察ではない。） ***この課題は余裕があれば行うこと。***

2-4 2-3 のモデルをもとに、ギターの音階を決める要因とギターの構造について説明せよ。

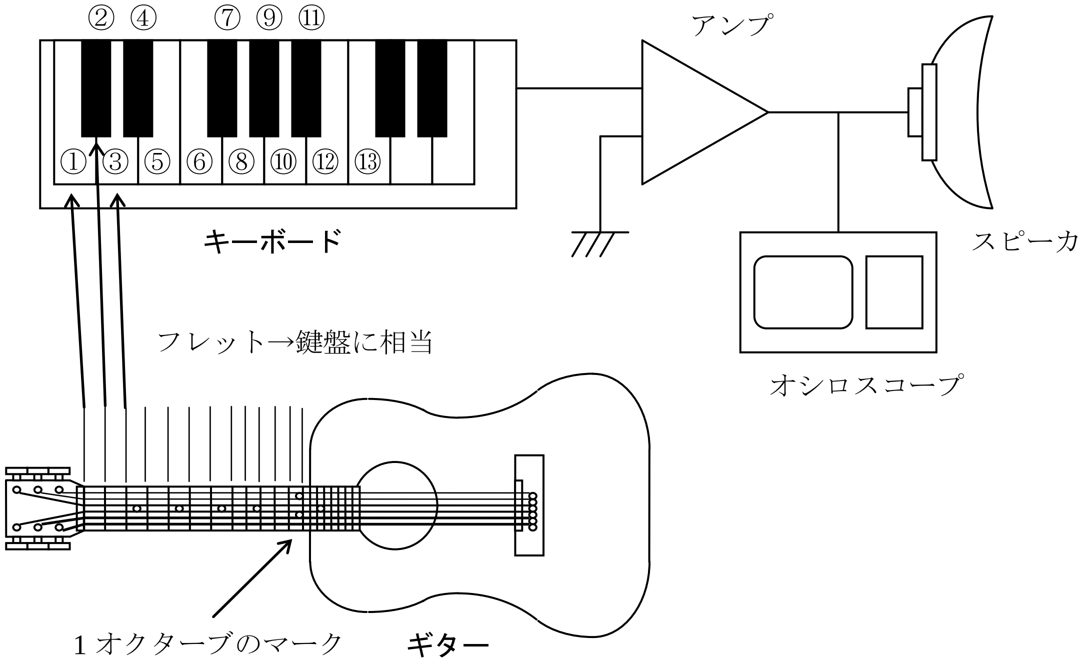


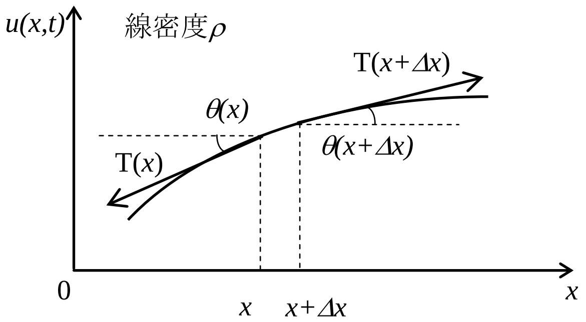
表 2　キーボードの鍵盤の番号と周波数、時間の関係

番号 n	時間 T_1 [ms]	時間 T_2 [ms]	周期 T [ms]	周波数 f [Hz]	番号 n	時間 T_1 [ms]	時間 T_2 [ms]	周期 T [ms]	周波数 f [Hz]
1(ド)					13（ド）				
2					14				
3					15				
4					16				
5					17				
6					18				
7					19				
8					20				
9					21				
10					22				
11					23				
12					24				
					25（ド）				

弦の長さと周波数の関係を考察するときのヒント

ヒント 1

以下は振動している弦の位置 x における変位 $u(x)$を示している。この弦の運動方程式をモデルとして考えよう。ただし張力 T は位置 x によらず一定であり、角度θは非常に小さく、$\sin\theta \fallingdotseq \tan\theta$が成り立つとする。



ヒント 2

$$\rho\frac{\partial^2 u}{\partial t^2} = T\frac{\partial^2 u}{\partial x^2}$$

上記の微分方程式の解のひとつは、以下の式で表される。

$u(x,t)$ = U(x)sin(ωt+ϕ)

上式で、U(x) = U_0sin(kx+ψ)

（$\frac{k^2}{\omega^2} = \frac{\rho}{T}$ となることを確かめよ。）

ヒント 3

弦の長さを L とすると、弦の両端が固定されていることより、どんな時刻 t においても、$u(0,t)=u(L,t)=0$ となる。（境界条件）この境界条件より、kL = nπであることがわかる。(確かめよ。)

このことと、$\omega=2\pi f$ の関係より、周波数 f と線密度ρ、張力 T、弦の長さ L の関係を導き出せ。

Break time

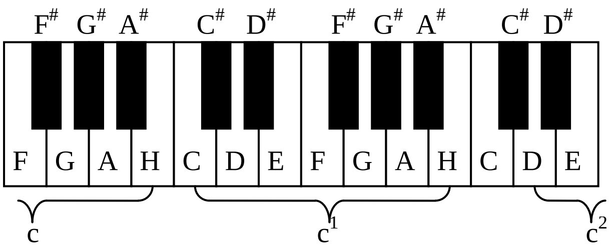
音と周波数の話

音階で 1 オクターブ上がるとか 1 オクターブ下がるということは周波数 f が 2 倍になったり、1/2 倍になったりすることである。音階の周波数は以下の表 1 のように定められており、1 オクターブ上がるごとに（$C_2 \rightarrow C_1 \rightarrow C \rightarrow c \rightarrow c^1 \rightarrow c^2 \rightarrow c^3 \rightarrow c^4$）周波数が倍になっていることが分かる。(表 1 の各行の値を比較してみれば分かる。) では、各音の関係はどうか？(つまり、$C \rightarrow C^\# \rightarrow D \rightarrow D^\# \rightarrow E \rightarrow E \rightarrow F \rightarrow F^\# \rightarrow G \rightarrow G^\# \rightarrow A \rightarrow A^\# \rightarrow H$ と変わるにつれて、各周波数がどう変化しているか？) 表 1 の十二平均率音階では、各周波数は均等倍の関係になっている。つまり $f_{C\#} = \alpha f_C$, $f_D = \alpha f_{D\#}$, ・・・の関係にある。ということは、$f_{cc2} = \alpha^{12} f_{cc1} = 2 f_{cc1}$ となり、$\alpha^{12}=2$ ということになる。ためしに、c^1 と c^2 について、各音の周波数を片対数グラフに描いてみよう。

（各音を線形軸に均等にとり、各音の周波数を対数軸にとる。）どのようなグラフになるだろうか？

表 1. 国際基準イ= a^1 = 440Hz に基く十二平均率音階　（単位：Hz）

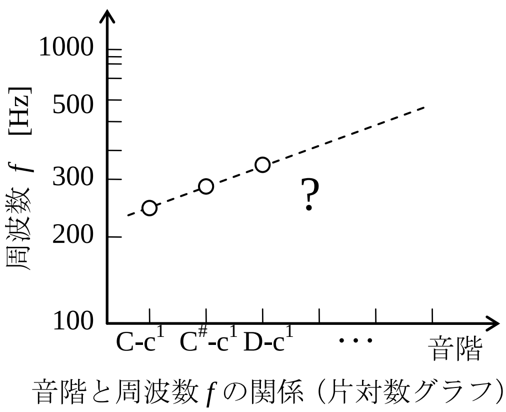
	C_2	C_1	C	c	c^1	c^2	c^3	c^4
C （ド）	16.352	32.703	65.406	130.81	261.63	523.25	1046.5	2093.0
C#	17.324	34.648	69.296	138.59	277.18	554.37	1108.7	2217.5
D （レ）	18.354	36.708	73.416	146.83	293.66	587.33	1174.7	2349.3
D#	19.445	38.891	77.782	155.56	311.13	622.25	1244.5	2489.0
E （ミ）	20.602	41.203	82.407	164.81	329.63	1318.5	1318.5	2637.0
F（ファ）	21.827	43.654	87.307	174.61	349.23	698.46	1396.9	2793.8
F#	23.125	46.249	92.499	185.00	369.99	739.99	1480.0	2960.0
G （ソ）	24.500	48.999	97.999	196.00	392.00	783.99	1568.0	3136.0
G#	25.957	51.913	103.83	207.65	415.30	830.61	1661.2	3322.4
A （ラ）	27.500	55.000	110.00	220.00	**440.00**	880.00	1760.0	3520.0
A#	29.135	58.270	116.54	233.08	466.16	932.33	1864.7	3729.3
H （シ）	30.868	61.735	123.47	246.94	493.88	987.77	1975.5	3951.1



ピアノの鍵盤

A-c^1(ラ)　: f_{Ac1}=440Hz
A-c^2　　: f_{Ac2}=880Hz
1 オクターブ上がると周波数は倍になる

NHK の時報
ピッ、ピッ、ピッ、ピーン
A-c^1,　A-c^1,　A-c^1,　A-c^2



音階と周波数 f の関係（片対数グラフ）

統計的処理に関する基礎

平均・分散と度数分布

算術平均

次の表１は、あるサッカーチームＡの選手１１名の身長である。

番号	1	2	3	4	5	6	7	8	9	10	11
身長（cm）	188	186	187	187	185	186	165	189	192	170	183

表１ サッカーチームＡの選手の身長

このデータを見ると 180cm 以上の長身選手がほとんどであることがわかる。このチームのいわゆる平均身長は、次のように計算される。

$$(188+186+187+187+185+186+165+189+192+170+183)\div 11$$
$$=2018\div 11=183.45...=183 \qquad (1)$$

これより、このチームＡの平均身長は 183cm であることがわかった。データから身長の平均値を計算すると、そのチームが平均的に背の高い「大型チーム」であることを確認できる。このように、データ x_i を合計し、データ数 N で割った式(2)の $\bar{x}$ を**算術平均**と呼ぶ。（$\bar{x}$ は x バーと読む）

$$\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i \qquad (2)$$

算術平均の Excel による計算

Excel に表１の表を作成し、番号１から１１までの選手の平均値を計算しよう。一番簡単な方法は身長が入力されているセルの合計を１１で割ることである（やってみよ）。Excel では平均などを計算するための便利な関数が用意されている。メニューの「挿入」の「関数」と選択すると関数を選ぶことができる。平均を計算するには、計算結果を入れるセルを選択しておいて、関数メニュー中の「統計」の中の AVERAGE を選択し、範囲を選べばよい（図１を参照）。一度使用した関数は次回から「最近使用した関数」で選ぶことができる。

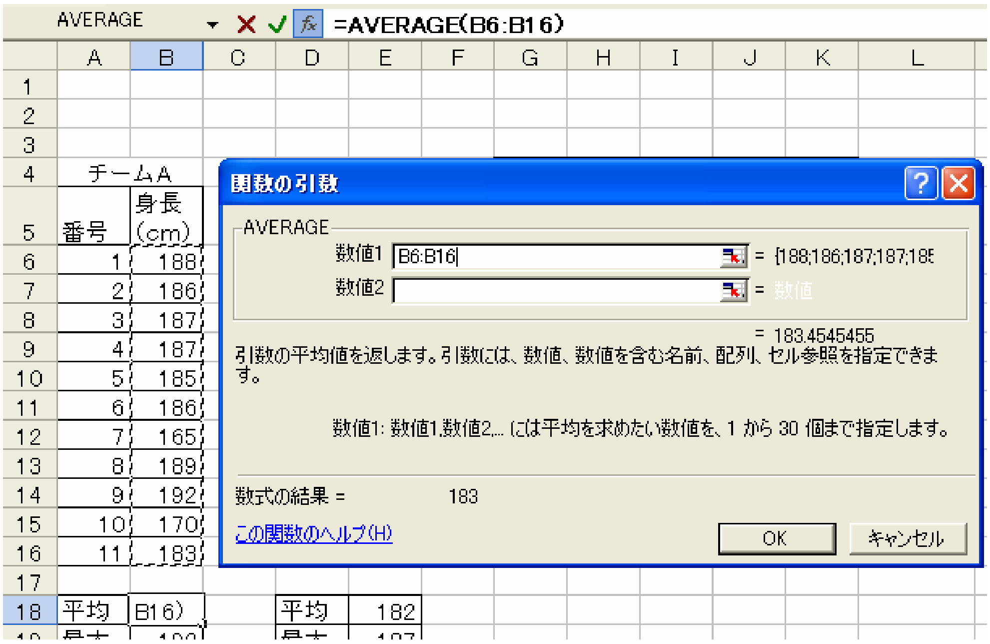
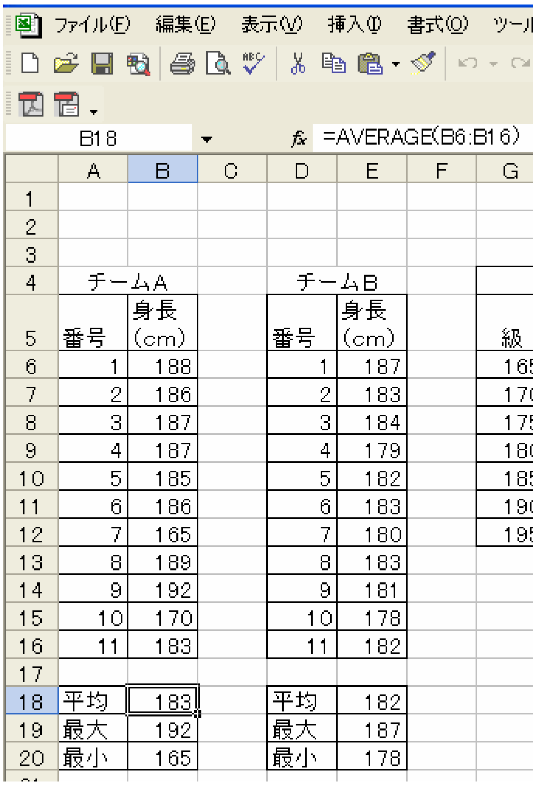


図１ 平均値の計算

例題 1：Excel で表 1 と、表 2 に示すサッカーチーム B の選手１１名の身長を入力し、それぞれの平均身長を計算してみよう（図 2 を参照）。

番号	1	2	3	4	5	6	7	8	9	10	11
身長(cm)	187	183	184	179	182	183	180	183	181	178	182

表 2　サッカーチーム B の選手の身長



	A	B	C	D	E	F	G
1							
2							
3							
4	チームA			チームB			
5	番号	身長(cm)		番号	身長(cm)		級
6	1	188		1	187		165
7	2	186		2	183		170
8	3	187		3	184		175
9	4	187		4	179		180
10	5	185		5	182		185
11	6	186		6	183		190
12	7	165		7	180		195
13	8	189		8	183		
14	9	192		9	181		
15	10	170		10	178		
16	11	183		11	182		
17							
18	平均	183		平均	182		
19	最大	192		最大	187		
20	最小	165		最小	178		

図 2 平均身長の計算

例題 2:チーム A と B それぞれのチームの最大･最小の身長を Excel の関数 MAX と MIN を用いて算出せよ。

例題 3：チーム A およびチーム B の各選手の身長を棒グラフで表せ。(図 3)

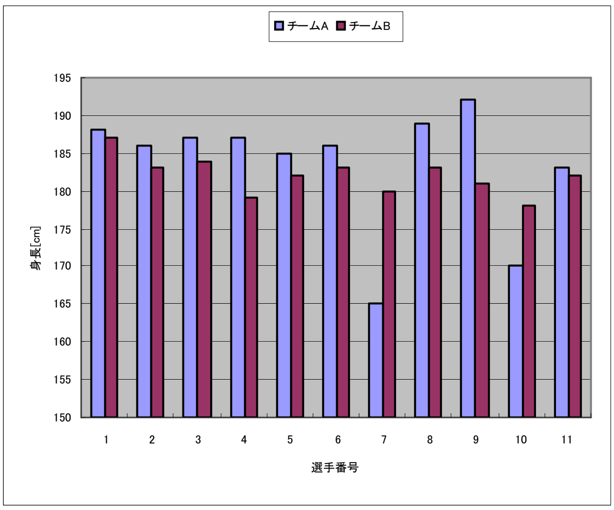


図 3 チーム A と B の各選手の身長比較

度数分布とヒストグラム

図 3 を見るとチーム A と比較してチーム B が全体的に身長が低いことがわかる。しかし身長の算術平均を計算するとチーム A が 183cm、チーム B は 182cm となり、あまり大きな違いはない。この原因を検討するには**度数分布**と**ヒストグラム**を用いるのが便利である。度数分布はデータを大きさによっていくつかの級に分け、おのおのの級に入るデータの数（度数）の分布を示したものである。5cm 単位で級に分けた場合のチーム A の度数分布は表 3 のようになる。

級(cm)	165	170	175	180	185	190	195
累積度数	1	2	2	2	4	10	11
度数	1	1	0	0	2	6	1

表 3 チーム A の選手身長の度数分布

表 3 をグラフで表すと図 4 のようになる。度数分布を棒グラフで表現したものをヒストグラムと呼ぶ。

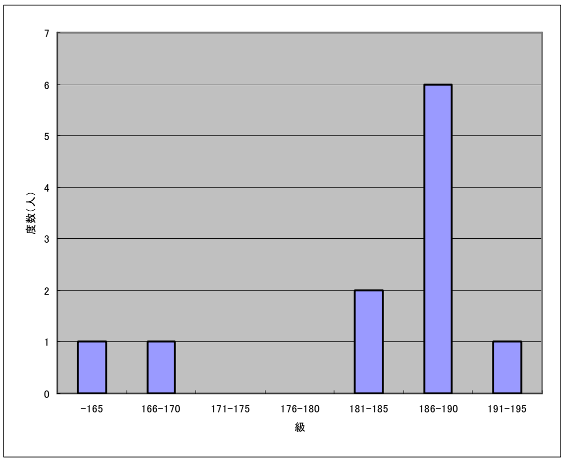


図 4 チーム A の選手身長のヒストグラム

例題 4：チーム A と B の身長の度数分布を同一のヒストグラムで表現せよ。(図 5)

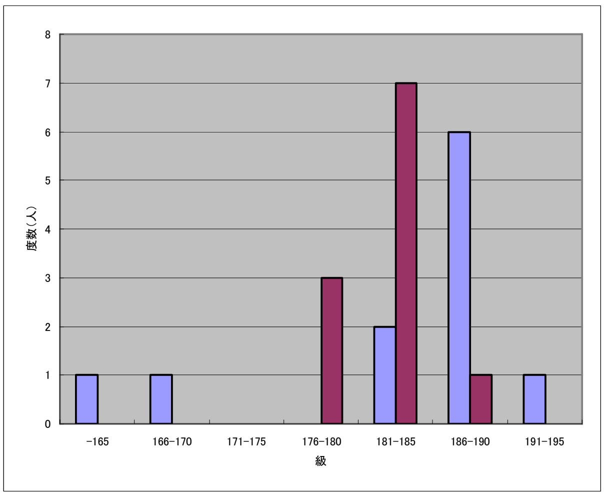


図 5 チーム A と B の選手身長のヒストグラム

<u>中位数（メディアン）と最頻値（モード）</u>

図 5 から、チーム A と B の平均身長に大きな差がないのは、チーム A のほとんどの選手はチーム B の選手の身長より大きいが、170cm 以下の選手がいるために算術平均値が下がっているためであることがわかる。このような場合に便利な別の平均は**<u>中位数（メディアン）</u>**と**<u>最頻値（モード）</u>**である。

- 中位数（メディアン）とはデータを大きさの順番に並べた場合に、ちょうど中央に位置するデータの値であり、チーム A は 186cm、チーム B は 183cm となる。なお、データの数が偶数の場合には、中央に位置する 2 個のデータの平均を取る。
- 最頻値（モード）とはもっともデータが集中している部分のことで、ヒストグラムなら山のもっとも高い部分である。チーム A と B は図 5 からそれぞれ 186～190 の級、181～185 の級になる。

例題 5：Excel でチーム A と B それぞれのメディアンを下記の 3 通りの方法で計算せよ。（図 6 を参照）

① メニューボタンの「降順で並び替え」

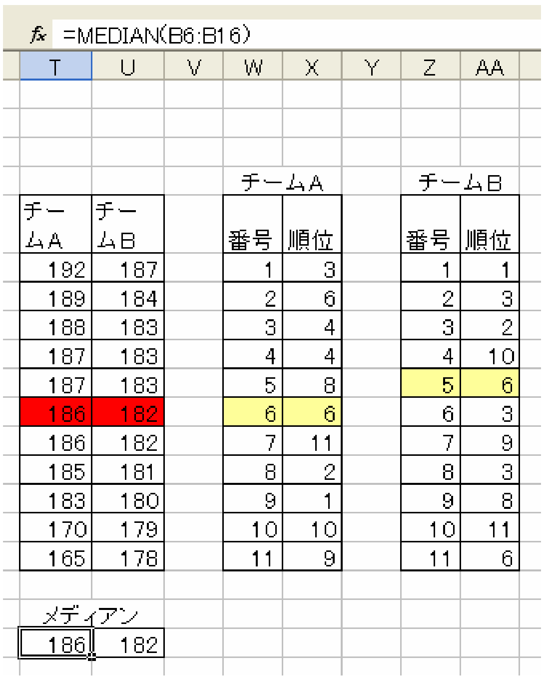
並び替えたいデータをコピーして選択した後に、「降順で並び替え」を押して中位の値（6 番目）をとる。

② メディアンを計算する関数 MEDIAN

③ 順位を計算する RANK 関数

RANK 関数の書式は、RANK(セル番号, 範囲, 順序)　である。

例：B7 のセルの、B6～B16 のデータ中の大きい順の順位を求めたい場合は、=RANK(B7,B6:B16, 0)、小さい順の順位なら、 =RANK(B7,B6:B16, 1)



fx =MEDIAN(B6:B16)

T	U	V	W	X	Y	Z	AA
			チームA			チームB	
チームA	チームB		番号	順位		番号	順位
192	187		1	3		1	1
189	184		2	6		2	3
188	183		3	4		3	2
187	183		4	4		4	10
187	183		5	8		5	6
186	182		6	6		6	3
186	182		7	11		7	9
185	181		8	2		8	3
183	180		9	1		9	8
170	179		10	10		10	11
165	178		11	9		11	6
メディアン							
186	182						

<u>図 6 並び替え、メディアン、順位の計算</u>

<u>標準偏差（母集団標準偏差）と分散（母集団分散）</u>

図 5 から、チーム B は選手の身長が 181～185 の級付近に集中しているのに対して、チーム A は 165cm から 190cm 超まで身長の分布が広がっていることがわかる。このような度数分布の広がり具合を表すために、<u>標準偏差</u>と<u>分散</u>が用いられる。標準偏差をσとおくと、N 個のデータ x_i の標準偏差σは次式で与えられる。

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} \tag{3a}$$

ここで $\bar{x}$ は(2)式の算術平均である。また分散は σ^2 であり、次式で与えられる。

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 \tag{3b}$$

すなわち、分散 σ^2 は平均値 $\bar{x}$ からの偏差 $x_i - \bar{x}$ の二乗和の平均値であり、標準偏差 σ はその平方根である。

チーム A と B について分散 σ^2 を計算すると、それぞれ 62 と 6 になり、選手の身長の分布が幅広いチーム A の分散が大きくなる。このように分散を計算すれば、分布の程度を定量的に把握できる。なお (5.3a)(5.3b)式は、正確にはそれぞれ、**母集団標準偏差**と**母集団分散**と呼ばれるものである。この他に用いられる分散として、次式(5.3c)の**標本標準偏差** $\hat{\sigma}$ と(5.3d)の**標本分散** $\hat{\sigma}^2$ がある。

$$\hat{\sigma} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} \tag{3c}$$

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 \tag{3d}$$

母集団標準偏差・分散と標本標準偏差・分散の違いについては統計の授業で詳しく学んでほしい。ここでは母集団標準偏差・分散のみを扱う。

例題 6

チーム A と B の分散（母集団分散）をそれぞれ下記の 2 通りの方法で計算し、結果を確認せよ。

① Excel で関数 VARP を用いて計算せよ。

② VARP を用いずに、(3b)式を用いて計算せよ。

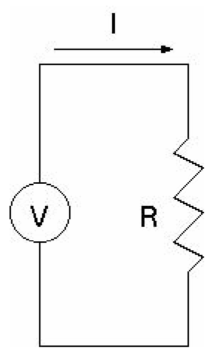


図 7 抵抗間電圧と電流

回帰直線と最小二乗法

図 7 に示す回路で電圧 V、電流 I、抵抗値 R の間には、オームの法則により次式の関係が成立する。

$$V = RI \tag{4}$$

したがって抵抗の抵抗値が未知の場合でも、V と I を一組計測すれば R を求めることができるはずである。しかし実際には様々な要因によって、V と I の一組だけから R を決定することは危険である。実際に、ある抵抗に対して、V と I を計測し、R を計算したところ表 4 の値になった。

データ番号	1	2	3	4	5	6	7	8	9	10
電圧[V]	1.00	1.50	2.00	2.50	3.00	3.50	4.00	4.50	5.00	5.50
電流[A]	6.10E-04	9.89E-04	1.27E-03	1.62E-03	1.94E-03	2.21E-03	2.54E-03	2.86E-03	3.17E-03	3.52E-03
抵抗値[Ω]	1.64E+03	1.52E+03	1.58E+03	1.54E+03	1.54E+03	1.59E+03	1.58E+03	1.57E+03	1.58E+03	1.56E+03
データ番号	11	12	13	14	15	16	17	18	19	
電圧[V]	6.00	6.50	7.00	7.50	8.00	8.50	9.00	9.50	10.00	
電流[A]	3.81E-03	4.08E-03	4.45E-03	4.75E-03	5.10E-03	5.35E-03	5.74E-03	6.00E-03	6.33E-03	
抵抗値[Ω]	1.58E+03	1.59E+03	1.57E+03	1.58E+03	1.57E+03	1.59E+03	1.57E+03	1.58E+03	1.58E+03	

表 4 ある抵抗に対する V と I および計算された R

このように R の値は実験毎に異なってしまう。V と I の関係をグラフに表したものが図 8 であるが、(4)のような直線関係になっていないことがわかる。これは V や I の値にノイズと呼ばれるさまざまな誤差要因が混入するためである。

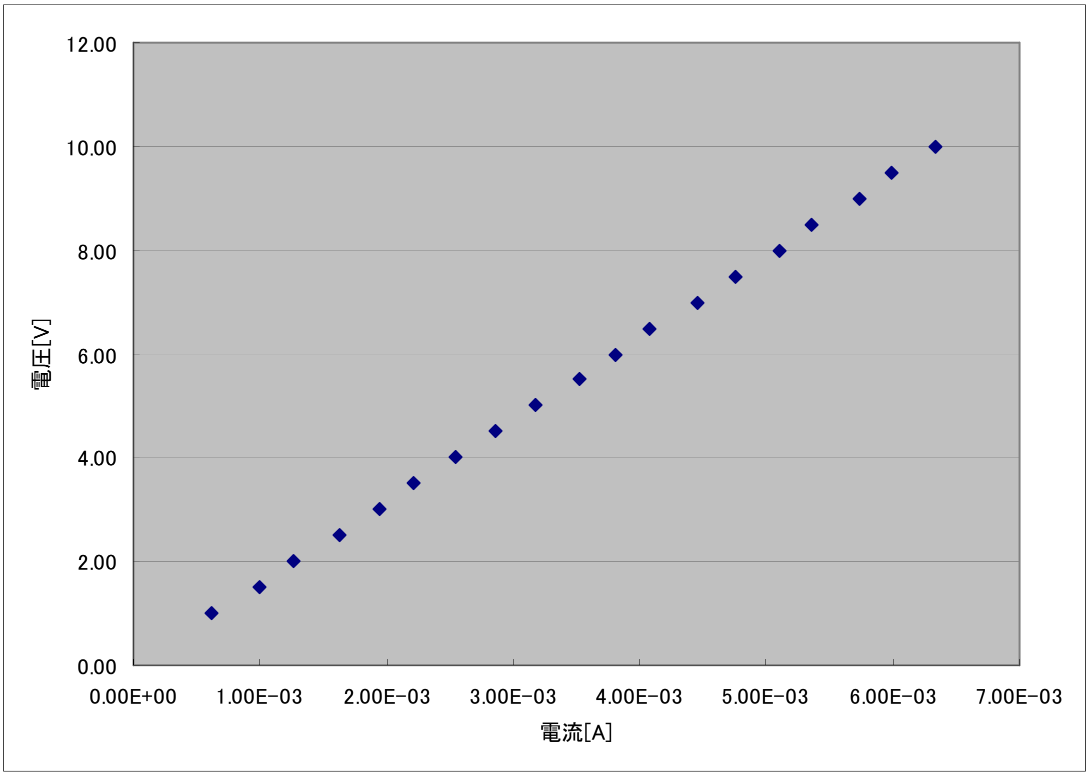


図 8 V と I の関係

このデータから R を計算するひとつの方法は、各々の実験データ(１番～１１番)で得られた抵抗値 R を平均することであるが、理論的な妥当性が乏しい。本節では、最小二乗法による回帰直線の計算方法を説明しよう。

回帰直線

データの組 (x, y) に対して

$$y = ax + b \tag{5}$$

のような直線の対応関係が成り立つとき、この直線(5)を回帰直線とよぶ。(4)式が成立するなら、y=V, x=I, a=R, b=0 となる。図 8 の実験データに対しては、分布するデータ点の中心を貫く直線となるはずである。ノイズに汚された第 i 回目の実験結果の組(x_i, y_i)から a と b の推定値 $\hat{a}$ と $\hat{b}$ を求める代表的な手法が最小二乗法である。

最小二乗法（1 次式の場合）

第 i 番目のデータの組 (x_i, y_i) に対して、次の E を最小にする $\hat{a}$ と $\hat{b}$ を求めたい。

$$E = \sum_{i=1}^{N} \left\{ y_i - \left(\hat{a} x_i + \hat{b} \right) \right\}^2 \tag{6}$$

E を最小にする $\hat{a}$ と $\hat{b}$ は次式で計算できることが知られている。

$$\hat{a} = \frac{\frac{1}{N} \sum_{i=1}^{N} (x - \overline{x})(y - \overline{y})}{\frac{1}{N} \sum_{i=1}^{N} (x - \overline{x})^2} \tag{7}$$

$$\hat{b} = \overline{y} - \hat{a} \overline{x} \tag{8}$$

(5.6)式の E は推定誤差 $y_i - \left(\hat{a} x_i + \hat{b} \right)$ の二乗和であり、これを最小にすることから最小二乗法と呼ばれている。

例題 7

表 4 と図 8 を Excel で作成せよ。

例題 8

表 4 のデータに対して Excel で回帰直線を計算することによって抵抗値 R を求めよ。なお、有効桁数は、V と I それぞれの有効桁数の短いほうになる。理由は各自で考えよ。Excel で回帰直線を求めるには、グラフのデータ部分を右クリックして、メニューの「近似曲線の追加」を選択し、種類で「線形近似」、オプションで「グラフに数式を表示する」を選択すると、自動計算されて表示される。(図 9)

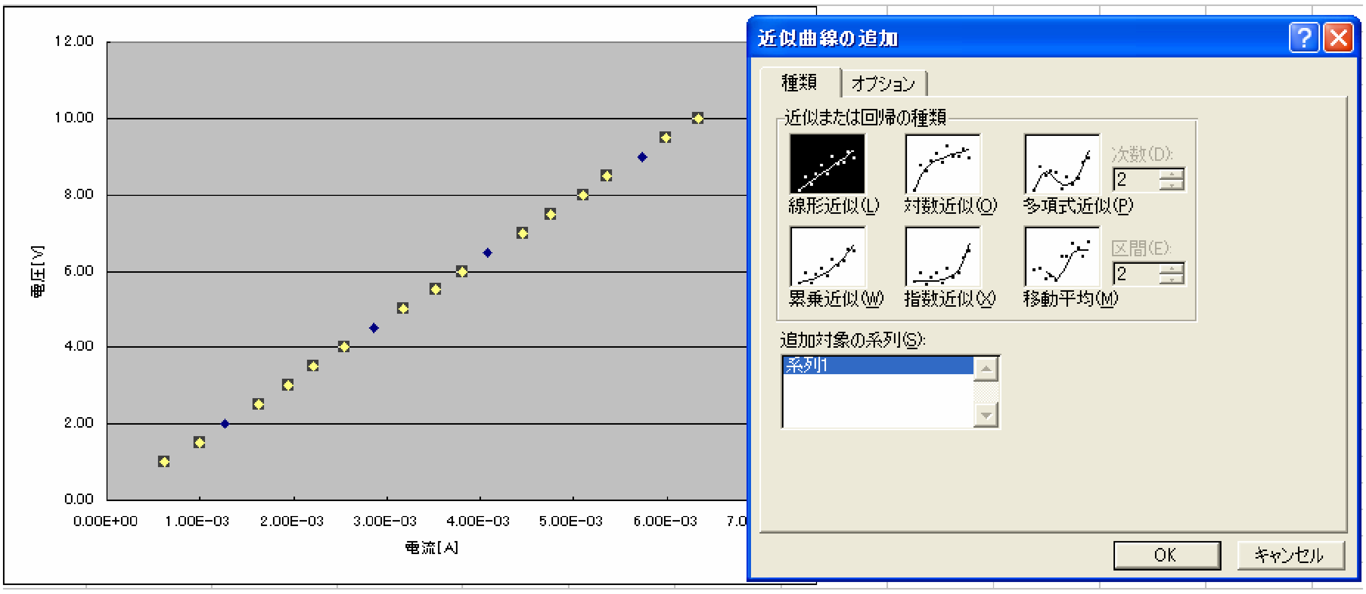


図 9 近似曲線の追加

例題 9

表 4 のデータに対して、**(7) と (8) を用いて**回帰直線を計算し、例題 8 の結果と比較せよ。