付録 A　PC を用いた確率計算実習 [Mathematica]

Mathematica は，Wolfram 社の販売する数式処理ソフトです．グラフ化やシミュレーションまで，今後の学習・研究に役立つことでしょう．

- 情報科学部ではサイトライセンス契約で利用可能となっています．
- 利用方法の詳細は，次のページを参照してください．（講義ページからリンクされています）．
  `http://www.oit.ac.jp/is/shinkai/lecture/mathematica.html`

演習室の PC からの起動方法

- Windows から，普通のソフトウェアと同じように，「プログラム」から「Wolfram Mathematica 8」を選択する．（下の方の「Mathematica カーネル」ではない．）
- 起動したら，5! と入力し，「shift+ 中央の enter」または「右端の enter」とキーボードを押して，120　と計算されるかどうか確かめよう．

A.1　課題

$\boxed{\text{A.1}}$「徹底攻略　常微分方程式」（真貝著，共立出版，2010 年 8 月）の p204–209 を熟読しながら，すべて実行して確かめよ[*1]．(p210-211 は不要)

$\boxed{\text{A.2}}$（誕生日問題）N 人集まったとき，誕生日が同じ人がいる確率について考える．

(1) 自分と同じ誕生日の人がいる確率を求め，N を横軸としてグラフにせよ．確率が 0.5 を超えるのは，何人以上のときか．

(2) N 人のなかで，同じ誕生日の人が少なくとも 1 組以上いる確率を求め，N を横軸としてグラフにせよ．確率が 0.5 を超えるのは，何人以上のときか．

$\boxed{\text{A.3}}$（サイコロをふる）

(1) 上記ウェブページより，`dice.txt` を参照し，関数 `dice[n_, m_]` を Mathematica で定義せよ[*2]．例えば，5 回サイコロをふる試行を行うには，`dice[1,5]` を実行する．この試行を 10 回繰返すには `dice[10,5]` を実行すればよい．

(2) 20 回サイコロを振り，1– 6 の目について出現回数を記録せよ．

(3) 上記ウェブページより，`dice1.txt` を参照し，関数 `dice1[n_]` を Mathematica で定義せよ．例えば，`dice1[100]` を実行すると，100 回サイコロをふったときの目の出た頻度が表示される．

(4) サイコロをふって出る目の平均値が，試行回数を増やすと 3.5 に近づいていくことを示せ．

$\boxed{\text{A.4}}$（2 項分布）酔歩問題やサイコロ問題で登場する 2 項分布 $B(n,p)$ について．

(1) 2 項分布 $B(n,p)$ で，$X = k$ のときの確率を表示する関数を作成せよ．

(2) 酔歩問題．左右どちらかに 1/2 の確率で歩く酔っぱらいが 10 歩後にいる位置の確率を求めよ．

(3) (2) をグラフにせよ．

[*1] 「微積分学 I」「微分方程式」で行った練習と同じ．該当部分は（学内からなら）web ページで pdf が手に入る．

[*2] 平嶋先生作のスクリプト．

$\boxed{\text{A.5}}$（標準正規分布）

(1) 標準正規分布曲線 $f(z)$ をグラフで示せ．

(2) $f(z)$ の $\alpha \leq z \leq 1000$ の部分の面積を計算する関数を作り，値を標準正規分布表と比較せよ．

(3) 知能指数が 125 以上の人は，何 % いるだろうか．

(4) 知能指数が 200 以上の人は，何 % いるだろうか．

$\boxed{\text{A.6}}$（Buffon の針）平行線が $2h$ の間隔で無数に引かれている平面に，長さ 2ℓ（ただし $\ell < h$）の針を無作為に落とすとき，針が平行線と交わる確率を p とする．p を用いると円周率 π が計算できる．

(1) 上記ウェブページより，`buffon1.txt` を参照し，関数 `buffon1[n_, length_]` を Mathematica で定義せよ．`buffon1[100, 0.7]` と実行せよ．

(2) 講義で解説した例題 1.24 によれば，$\pi = 2\ell/ph$ である．シミュレーションを繰り返すと円周率 π が得られるだろうか．（`buffon1` では，$h = 1$ である）．

A.2 方針とヒント

$\boxed{\text{A.2}}$（誕生日問題）

(1) 2 人なら，相手が同じ誕生日の確率は，$\dfrac{1}{365}$．3 人なら，$\dfrac{1}{365} + \dfrac{1}{365}$．
グラフをプロットするのは，`Plot[....]`

(2) 余事象を考えることで，確率は $1 - \dfrac{{}_{365}\mathrm{P}_N}{365^N}$．これをプロットすればよい．
${}_{365}\mathrm{P}_N$ を関数として組んでみよう．
`f[N_]:=....`

$\boxed{\text{A.3}}$（サイコロをふる）

(1) 発生する乱数は毎回異なるので，何回か試してみよう．

(4) `dice1[ ]` のプログラムを使っても良いが，次のコマンドを用いてグラフにすることもできる．
乱数を 1 から 6 までの範囲で，10 個並べる命令文の例．（ついでにグラフも）

```
t1 = Table[Random[Integer, {1, 6}], {10}]
ListPlot[t1]
```

上記の `Table` に格納された数字の和を求めるためには，

```
Sum[t1[[i]], {i, 1, 10}]
```

付録 B　PC を用いた統計計算実習 [Excel または Mathematica]

データ処理の 1 例として,「父親／母親の身長」と「息子／娘の身長」の相関を調べよう．データは，Excel または Mathematica 形式で，授業ページからリンクして置いてある．

B.1　課題

次のデータは，父親・母親とその成人した子供の身長データ (cm) である．2019–2012 年度の講義中に行ったアンケートから，兄妹あるいは姉弟の組み合わせの家族構成をもつデータ 163 家族分を抽出した．

$\boxed{\text{B.1}}$ (1) 父親データの標本平均，標本分散，標準偏差を求めよ．

(2) 母親データの標本平均，標本分散，標準偏差を求めよ．

(3) 息子データの標本平均，標本分散，標準偏差を求めよ．

(4) 娘データの標本平均，標本分散，標準偏差を求めよ．

$\boxed{\text{B.2}}$ 父親・母親・息子・娘　のうちから 2 者を取り出し，相関係数をそれぞれ求めよ．

$\boxed{\text{B.3}}$ (1) 父親・母親・息子・娘　のうちから 2 者を取り出し，回帰直線を求めよ．

(2) 父親・母親の身長に対する息子の身長，および娘の身長について，重回帰分析をせよ．

	x	y	z	w
	父	母	息子	娘
1	165	156	168	154
2	169	163	166	155
3	173	159	174	159
4	174	161	173	158
5	175	158	172	162
6	174	150	171	153
7	175	155	160	155
8	171	163	169	164
9	175	158	170	155
10	178	148	165	164
11	163	163	161	152
12	170	155	165	150
13	175	167	171	161
14	172	161	170	164
15	174	161	176	158
16	173	160	175	162
17	175	165	175	165
18	173	152	163	162
19	169	152	167	153
20	180	165	174	165
21	172	150	162	153
22	161	157	171	154
23	168	150	170	153
24	172	162	174	160
25	178	155	172	166
26	165	154	173	163
27	163	158	172	160
28	160	150	170	150
29	177	156	172	156
30	169	150	169	160
31	175	155	168	160
32	170	158	172	162
33	167	158	171	162
34	168	160	175	162
35	170	155	164	152
36	163	155	158	155
37	170	165	178	170
38	168	165	179	168
39	171	164	177	170
40	174	157	168	157
41	171	164	178	159
42	170	150	170	160
43	164	160	180	163
44	175	155	170	157
45	175	158	173	175
46	171	170	172	155
47	155	160	177	154
48	175	150	178	170
49	171	165	160	158
50	165	153	172	154
51	175	156	174	164
52	175	163	179	167
53	168	160	173	166
54	160	165	170	165
55	165	150	167	155
56	170	164	167	165
57	166	153	163	153
58	164	166	173	165
59	161	153	165	149
60	175	155	168	166
61	175	152	168	157
62	160	155	165	160
63	163	154	173	160
64	170	154	167	152
65	175	165	173	160
66	167	153	160	136
67	170	165	169	156
68	172	147	169	145
69	172	168	160	153
70	170	160	171	157
71	174	170	185	165
72	170	161	169	155
73	168	158	168	158
74	172	165	164	157
75	169	162	172	165
76	175	165	171	162
77	169	162	180	163
78	178	159	175	160
79	189	170	179	169
80	175	165	169	165
81	164	166	172	163
82	175	160	178	163
83	155	145	159	150
84	175	164	164	158
85	173	163	171	165
86	165	160	170	165
87	160	150	166	153
88	167	153	172	165
89	175	154	171	150
90	176	150	174	168
91	178	163	173	165
92	180	165	175	165
93	167	155	170	158
94	168	160	163	161
95	165	159	160	155
96	168	150	165	158
97	167	165	176	160
98	164	150	160	160
99	170	164	170	167
100	155	165	152	160
101	174	160	175	162
102	160	165	180	155
103	174	153	170	156
104	170	160	163	150
105	171	158	172	153
106	173	156	178	152
107	170	150	165	150
108	168	154	172	154
109	165	155	176	160
110	175	163	175	159
111	175	150	170	159
112	175	150	178	170
113	165	150	166	160
114	170	170	175	152
115	160	162	171	152
116	170	160	171	185
117	179	170	182	165
118	176	165	174	167
119	176	160	180	160
120	168	148	169	160
121	175	160	175	158
122	165	154	169	148
123	166	160	161	152
124	170	155	166	152
125	180	147	160	155
126	170	146	170	155
127	173	160	175	170
128	165	165	170	150
129	173	162	186	163
130	169	157	173	160
131	174	162	178	160
132	176	156	168	158
133	170	160	170	158
134	180	165	170	170
135	173	154	172	156
136	178	158	172	156
137	170	165	177	160
138	170	160	168	157
139	164	160	171	161
140	176	160	172	165
141	175	160	170	163
142	173	155	160	155
143	178	164	186	166
144	170	165	173	160
145	175	160	174	161
146	170	160	172	165
147	170	150	170	160
148	175	165	176	165
149	169	158	165	155
150	187	164	180	171
151	160	154	166	158
152	173	156	180	156
153	175	163	165	140
154	165	155	170	155
155	170	162	172	160
156	159	160	167	164
157	171	156	166	154
158	174	167	177	158
159	175	160	162	164
160	170	160	163	153
161	173	164	173	165
162	182	162	170	161
163	168	150	166	150

B.2　ヒント

■平均，積和

n 個のデータ (x_i, y_i), $(i = 1, \cdots, n)$ が与えられているとき，平均と積和は，次のように定義される．

$$\overline{x} = \frac{1}{n}\sum_i x_i, \quad S_{xx} = \sum_i (x_i - \overline{x})^2$$
$$\overline{y} = \frac{1}{n}\sum_i y_i, \quad S_{xy} = \sum_i (x_i - \overline{x})(y_i - \overline{y})$$

■相関係数

x と y の対からなる標本 (x_i, y_i) $(i = 1, \cdots, n)$ の**相関係数** r は

$$r = \frac{\displaystyle\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\left[\displaystyle\sum_i (x_i - \overline{x})^2\right]\left[\displaystyle\sum_i (y_i - \overline{y})^2\right]}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \tag{付録 B.1}$$

■最小２乗法による回帰直線解析

n 個のデータ (x_i, y_i), $(i = 1, \cdots, n)$ が与えられているとき，これらのデータ分布を，もっとも良く近似する直線（**回帰直線**）を

$$y(x) = ax + b \tag{付録 B.2}$$

とすると，

$$a = \frac{S_{xy}}{S_{xx}}, \qquad b = \overline{y} - a\overline{x} \tag{付録 B.3}$$

■2 変数に対する回帰方程式（重回帰解析）

データ y を，2 個のデータ (x_1, x_2) を用いて表現する回帰方程式

$$y(x_1, x_2) = a_1x_1 + a_2x_2 + b \tag{付録 B.4}$$

の係数は次のように求められる．

$$a_1 = \frac{\begin{vmatrix} S_{x_1y} & S_{x_1x_2} \\ S_{x_2y} & S_{x_2x_2} \end{vmatrix}}{\begin{vmatrix} S_{x_1x_1} & S_{x_1x_2} \\ S_{x_1x_2} & S_{x_2x_2} \end{vmatrix}}, \qquad a_2 = \frac{\begin{vmatrix} S_{x_1x_1} & S_{x_1y} \\ S_{x_1x_2} & S_{x_2y} \end{vmatrix}}{\begin{vmatrix} S_{x_1x_1} & S_{x_1x_2} \\ S_{x_1x_2} & S_{x_2x_2} \end{vmatrix}}, \qquad b = \overline{y} - a_1\overline{x_1} - a_2\overline{x_2} \tag{付録 B.5}$$

ただし，$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$

付録 C　解答例

C.1　確率問題　解答例

A.2 （誕生日問題）

(1) `Plot[x/365, {x, 1, 365}]`

(2)
```
f[N_] := 365!/(365 - N)!
Plot[1 - f[x]/(365^x), {x, 1, 60}]
Plot[1 - f[x]/(365^x), {x, 22, 23}]
```

A.3 （サイコロをふる）

(4)
```
tmax = 1000;
t1 = Table[Random[Integer, {1, 6}], {tmax}];
t2 = Table[Sum[t1[[i]], {i, 1, j}]/j, {j, tmax}];
plot1 = ListPlot[t2, PlotJoined -> True, PlotRange -> {3, 4}]
N[t2[[tmax]]]
```

A.4 （2 項分布）

(1) `binomialdist[n_, p_, k_] := ( n!/ (n-k)!  / k!)  p^k (1-p)^(n-k)`

(2) `Do[Print[binomialdist[10, 1/2, k]], {k, 0, 10}]`

(3) `Plot[binomialdist[10, 1/2, k], {k, 0, 10}]`

A.5 （標準正規分布）

(1)
```
normaldist[x_] := Exp[-x^2/2]/Sqrt[2 Pi]
Plot[normaldist[x], {x, -5, 5}]
```

(2) `bunpu[a_, b_] := NIntegrate[normaldist[x], {x, a, b}]`

(3) `bunpu[(125 - 100)/15, 1000]`

(4) `bunpu[(200 - 100)/15, 1000]`

C.2　統計問題　解答例（2019 年版データ）

$\boxed{\text{B.1}}$ 平均 $\overline{x} = \dfrac{1}{163}\displaystyle\sum_{i=1}^{163} x_i$, 分散 $S_x^2 = \dfrac{1}{163}\displaystyle\sum_{i=1}^{163}(x_i - \overline{x})^2$，標準偏差 $\sigma_x = \sqrt{S_x^2}$ などから求める．

	標本平均	標本分散	標準偏差
父親	$\overline{x}$ =170.63	S_x^2 =32.28	σ_x =5.68
母親	$\overline{y}$ =158.54	S_y^2 =32.72	σ_y =5.72
息子	$\overline{z}$ =170.58	S_z^2 =33.49	σ_z =5.78
娘	$\overline{w}$ =159.22	S_w^2 =40.76	σ_w =6.38

$\boxed{\text{B.2}}$ 例えば父親と母親の相関係数 r_{xy} は,

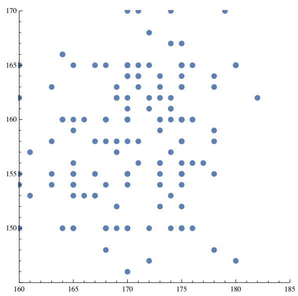
$$S_{xx} = \sum_{i=1}^{163}(x_i - \overline{x})^2 = 5229.91, \qquad S_{yy} = \sum_{i=1}^{163}(y_i - \overline{y})^2 = 5300.49,$$

$$S_{xy} = \sum_{i=1}^{163}(x_i - \overline{x})(y_i - \overline{y}) = 1079.39 \text{ より，} r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = 0.205.$$

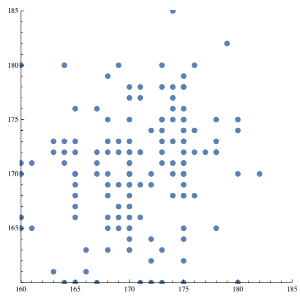
表形式にして相関係数をまとめると，次のようになる．

	父親 x	母親 y	息子 z	娘 w
父親 x	—	0.205	0.323	0.324
母親 y		—	0.364	0.334
息子 z			—	0.460
娘 w				—

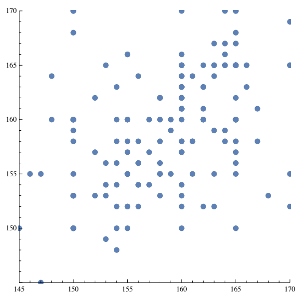
父親と母親のデータ分布図



父親と息子のデータ分布図



母親と娘のデータ分布図

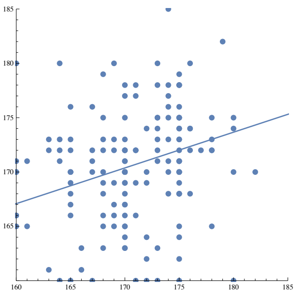


$\boxed{\text{B.3}}$ (1) 例えば，父親 x と息子 z の身長データの回帰直線 $z = ax + b$ は，(??) より，$a = S_{xz}/S_{xx} = 0.329$, $b = \overline{z} - a\overline{x} = 114.47$ となる．

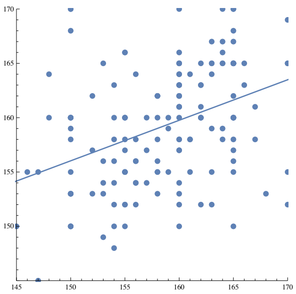
回帰直線は，次のようになる．

(父親 x, 母親 y)	$y = +0.206\,x + 123.32$	
(父親 x, 息子 z)	$z = +0.329\,x + 114.47$	図参照
(父親 x, 娘 w)	$w = +0.364\,x + \ \ 97.03$	
(母親 y, 息子 z)	$z = +0.368\,y + 112.21$	
(母親 y, 娘 w)	$w = +0.373\,y + 100.11$	図参照
(息子 z, 娘 w)	$w = +0.507\,z + \ \ 72.67$	図参照

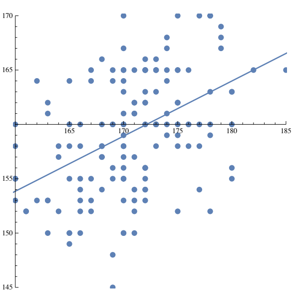
父親と息子のデータ分布図



母親と娘のデータ分布図



息子と娘のデータ分布図



(2) 父親 x と母親 y の身長に対する息子の身長 z の回帰方程式は，(??) にならって，$z = a_1 x + a_2 y + b$ とおける．(??) より，係数を定めると，

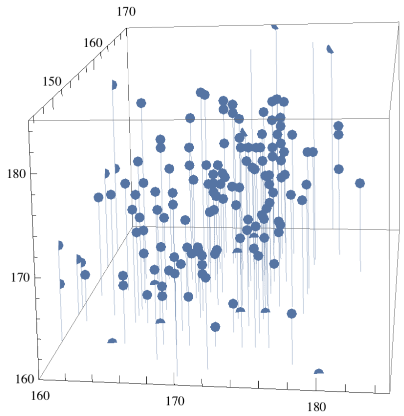
$$z = 0.264\,x + 0.314\,y + 75.68 \cdots\cdots (\#)$$

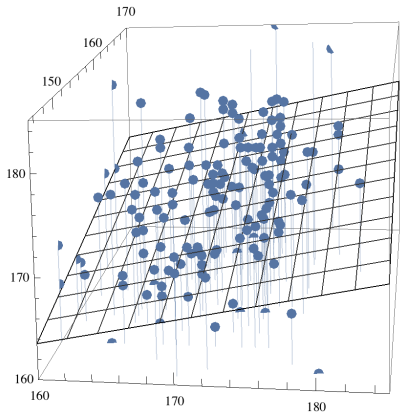
となる．(x_i, y_i, z_i) の 3 次元データプロットを下図に示す．得られた回帰方程式は，(x, y, z) 空間内の平面の方程式である．データの点に平面を重ねた図も示す．データのちらばり具合を代表する平面になっていることがわかる．

父親 x と母親 y の身長に対する娘の身長 w の回帰方程式は，同様に

$$w = 0.300\,x + 0.312\,y + 58.59 \cdots\cdots (\flat)$$

となる．





- (2) で得られた (#) は父親の身長 x と母親の身長 y から息子の身長 z を予測する式，$(\flat)$ は娘の身長 w を予測する式であるといえる．サンプル数が少ないので誤差は多いと考えられるが，このデータからは，息子も娘も父親の影響を強く受けて身長が決まるという結果になった．
- ちなみに，web で「身長 予測 Height Predictor」などで検索すると，次のような式が見つかる．

$$\begin{cases} \text{息子の身長} &= (\text{父親の身長} + \text{母親の身長}) \times 1.08/2 \\ \text{娘の身長} &= (\text{父親の身長} \times 0.923 + \text{母親の身長})/2 \end{cases}$$

$$\begin{cases} \text{息子の身長} &= (\text{父親の身長} + \text{母親の身長} + 13)/2 \\ \text{娘の身長} &= (\text{父親の身長} + \text{母親の身長} - 13)/2 \end{cases}$$

はてさて真偽は??