どうする分布表？　　横山 徹

【はじめに】

　平均や分散が大きく異なるデータで度数分布表を生徒に作成させてみる。

> 課題
> 　下記のデータについて、階級の幅を適切にとり、２つのデータを比較する度数分布表を作りなさい。

【参考】

　個別のデータでは階級の幅の取り方について様々な方法がある。

番号	A	B
1	30	29
2	72	47
3	47	-4
・	・	・
・	・	・
99	49	-15
100	20	-79
最大	82	132
最小	-25	-123
平均	25.04	-2.11
標準偏差	19.67	47.07

スタージェスの式

$K=1+\log_2 N$

　K：求めたい階級の数

　N：データ数

スコットの選択　$w=\dfrac{3.5\sigma}{\sqrt[3]{n}}$

フリードマン＝ダイアコニスの選択

$$w=\frac{2(Q_3-Q_1)}{\sqrt[3]{n}}$$

w：階級の幅　　n：データ数

σ：標準偏差　　Q：四分位数

A.Bそれぞれで適切な階級の幅を求めると

スタージェスの式では $K=7.64$ から

$$w_A=14.0 \quad w_B=33.5$$

スコットの選択では

$$w_A=\frac{3.5\times19.7}{\sqrt[3]{100}}=14.9 \quad w_B=\frac{3.5\times47.1}{\sqrt[3]{100}}=35.5$$

フリードマン＝ダイアコニスの選択では

$$w_A=\frac{2(35-12)}{\sqrt[3]{100}}=9.9$$

$$w_B=\frac{2(29.5+30.5)}{\sqrt[3]{100}}=25.9$$

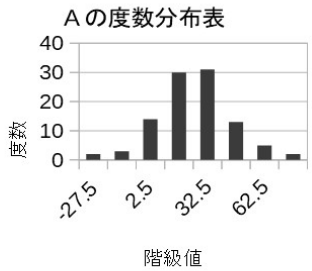
　個別にAとBの度数分布表を作成するときの適切な階級の幅はそれぞれ１０～１５と３０前後で、２倍ほどの違いがある。Aを１５、Bを３０の幅でそれぞれ度数分布表を作成すると次のようになる。

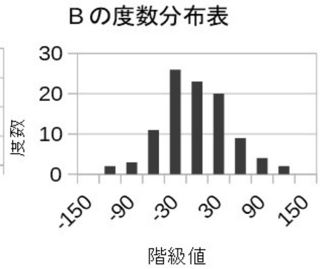
Aの度数分布表

階級 以上	未満	階級値	度数
-35～	-20	-27.5	2
-20～	-5	-12.5	3
-5～	10	2.5	14
10～	25	17.5	30
25～	40	32.5	31
40～	55	47.5	13
55～	70	62.5	5
70～	85	77.5	2

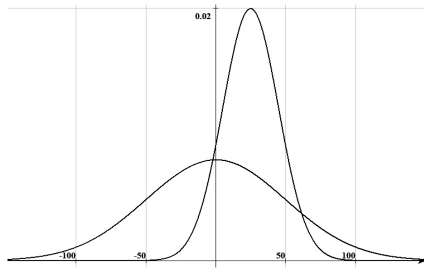
Bの度数分布表

階級 以上	未満	階級値	度数
-165～	-135	-150	0
-135～	-105	-120	2
-105～	-75	-90	3
-75～	-45	-60	11
-45～	-15	-30	26
-15～	15	0	23
15～	45	30	20
45～	75	60	9
75～	105	90	4
105～	135	120	2
135～	165	150	0





　Aは平均25、標準偏差20、Bは平均0、標準偏差50として正規乱数を発生させたデータである。



【おわりに】

　階級の幅を決めるときに参考となる３つの公式を理解し、正規分布らしさが残る概形を示すことができれば良いが、試行錯誤させることに意義があると思う。適切な処理をすることは、逆に提示されたデータが適切かどうかを考えることにもなる。

　余談になるが、度数分布表を考えると「的」が頭に浮かんできて、階級の幅を変えるように、的の同心円の半径を変えたらどうなるか気になってきた。

【余談１】

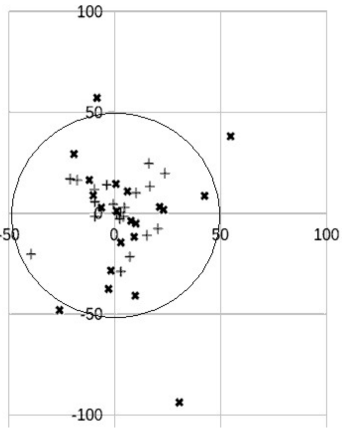
矢を的に放つとき、的の中心を原点とし、矢が当たった場所を極形式を用いて

$(r\cos\theta, r\sin\theta)$　 $-\infty < r < \infty$、 $0 \leqq \theta \leqq \pi$

で表す。$r<0$ は下半平面上の点になる。

このときθは一様分布、勝敗に関係するrは平均 0 の正規分布になると考えると上級者ほど r の標準偏差は小さくなる。

表1　AとBの実力差



ＡとＢの r の分布ががN$_A(0,20^2)$ とN$_B(0.40^2)$ に従うとして、２０個のデータを乱数を用いて発生させると表１のようになる。 X_A をAの r の値をとる確率変数として

$P(|X_A|<20)$

$P(20 \leqq |X_A| < 40)$

$P(40 \leqq |X_A|)$

としてBも同様に求めると表２となる。階級の取り方を変えて表４を作り、表２，表４から同時分布を求めると表３，表５になる。

表2

階級	A	B
0-20	0.6826	0.383
20-40	0.2718	0.2996
40以上	0.0456	0.3174

表3

A\B	0-20	20-40	40以上
0-20	0.2614358	0.20450696	0.21665724
20-40	0.1040994	0.08143128	0.08626932
40以上	0.0174648	0.01366176	0.01447344

表4

階級	A	B
0-30	0.8664	0.5468
30-60	0.131	0.3196
60以上	0.0026	0.1336

表5

A\B	0-30	30-60	60以上
0-20	0.2614358	0.32996884	0.09119536
20-60	0.1205684	0.15217432	0.04205728
60以上	0.0009958	0.00125684	0.00034736

表３と表５の同時分布からAとBの勝つ確率を求めると表６になる。

表6

	A勝ち	B勝ち	引き分け	A勝/B勝
表3	0.50743352	0.13522596	0.35734052	3.7525
表5	0.41015408	0.07388344	0.51596248	5.5514

階級の取り方は的の規格に該当し、勝敗の決め方、ポイントの付与方法によっても勝つ確率は変化することが予想される。大会と娯楽では求められるルールも違ってくる。

国際大会では実力が僅差の競技者に対して勝敗を決めることが必要であり、娯楽では実力差の違いがあっても楽しめるルールが必要になる。また娯楽であても職場の親睦会と父親が子供と遊ぶ場合では実力差は大きく異なる。

【問題】

ＡとＢの勝つ確率がどう変化するかに着目してゲームのルールを考え、適用するルールが競技の目的や対象に対して妥当かを説明しなさい。

$N_A(0,20^2)$ と $N_B(0.40^2)$ の正規乱数を１万個発生させて表１と表３の階級に合わせて勝敗をカウントしたものが表７と表８で、これらの割合を表６と比較したものが表９である。

表7

			勝ち A	勝ち B	引分
1	36.5	43.8	1		
2	-19.8	-111.8	1		
3	-8.9	-57.2	1		
4	-7.8	32.1	1		
5	-19.5	-132	1		
	.	.			
	.	.			
9999	-1.2	3.5			1
10000	-9.4	74.4	1		
		合計	5061	1330	3609

表8

			勝ち A	勝ち B	引分
1	36.5	43.8			1
2	-19.8	-111.8	1		
3	-8.9	-57.2	1		
4	-7.8	32.1	1		
5	-19.5	-132	1		
	.	.			
	.	.			
9999	-1.2	3.5			1
10000	-9.4	74.4	1		
		合計	4094	751	5155

表9

	A勝ち	B勝ち	引き分け
表7	0.5061	0.133	0.3609
	0.50743352	0.13522596	0.35734052
表8	0.4094	0.0751	0.5155
	0.41015408	0.07388344	0.51596248

＊下段の数値は表6の確率を表す

引き分けを減らすルールを考えるとAとBそれぞれの勝つ確率は増えると予想される。そこで発生させた乱数を用いて中心からの距離の近い方を勝ちとしてシミュレーションを行うと表１０になり、Aの勝つ確率は0.7、Bの勝つ確率は0.3の付近となることが予想される。

表10

			勝ち A	勝ち B	引分
1	36.5	43.8	1		
2	-19.8	-111.8	1		
3	-8.9	-57.2	1		
4	-7.8	32.1	1		
5	-19.5	-132	1		
			.	.	
			.	.	
9999	-1.2	3.5	1		
10000	-9.4	74.4	1		
		合計	7041	2936	23

余談２
　計算を簡単にするためにＡとＢの勝つ確率をそれぞれ0.8、0.2としよう。その後の調査で、Ａは本番に弱く、勝った次の試合では勝つ確率は0.6まで下がり、負けた次の試合では勝つ確率は0.4になるという。このまま毎日試合を繰り返すとＡとＢの確率がどうなるか。但し、初回の試合でＡとＢの勝つ確率は0.8、0.2とする。

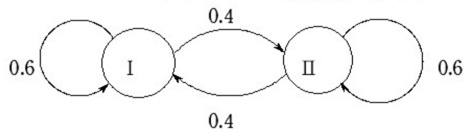
　５日後まで表計算ソフトで計算すると表１１のようになる。確率は0.5に近付くか、それとも逆転するであろうか？

表11

第n日	勝つ確率	
	A	B
初日	0.8	0.2
1	0.56	0.44
2	0.512	0.488
3	0.5024	0.4976
4	0.50048	0.49952
5	0.500096	0.499904

　Ａが勝つ状態をⅠ、Bが勝つ状態をⅡとし、ⅠとⅡの状態の変化を図にすると表１２のようになる。

図１２　状態推移図



　n日後にＡが勝つ確率を p_n、Bが勝つ確率を q_n とすると $p_0=0.8$　$q_0=0.2$ で

$$p_n=0.6p_{n-1}+0.4q_{n-1} \quad \cdots\cdots①$$
$$q_n=0.4p_{n-1}+0.6q_{n-1} \quad \cdots\cdots②$$

（１）３項間漸化式による解法

①から $q_{n-1}=\dfrac{5}{2}p_n-\dfrac{3}{2}p_{n-1}$

$$q_n=\frac{5}{2}p_{n+1}-\frac{3}{2}p_n \cdots\cdots③$$

となり、②に代入して整理すると

$$5p_{n+1}-6p_n+p_{n-1}=0$$

特性方程式を解くと $\lambda=1,\ \dfrac{1}{5}$ になるから

$\lambda=1$のとき

$p_{n+1}-p_n=\dfrac{1}{5}(p_n-p_{n-1})$ で $p_1-p_0=-\dfrac{6}{25}$

となり $p_n-p_{n-1}=-\dfrac{6}{25}\cdot\dfrac{1}{5^n}$ ……④

$\lambda=\dfrac{1}{5}$のとき

$p_{n+1}-\dfrac{1}{5}p_n=p_n-\dfrac{1}{5}p_{n-1}$ で $p_1-\dfrac{5}{5}p_0=\dfrac{2}{5}$

となり $p_n-\dfrac{1}{5}p_{n-1}=\dfrac{2}{5}$ ……⑤

④⑤から $p_n=\dfrac{1}{2}+\dfrac{3}{10}\cdot\dfrac{1}{5^n}$　これを③に代入し

$$q_n=\frac{1}{2}-\frac{3}{10}\cdot\frac{1}{5^n}$$

試合を毎日繰り返せばA,Bの勝つ確率 p と q は

$p=q=\dfrac{1}{2}$ になる。

（２）ハミルトン･ケーリーの定理を用いた解法

$\begin{pmatrix}p_n\\q_n\end{pmatrix}=\begin{pmatrix}0.6&0.4\\0.4&0.6\end{pmatrix}\begin{pmatrix}p_{n-1}\\q_{n-1}\end{pmatrix}$ で表し、$P=\begin{pmatrix}0.6&0.4\\0.4&0.6\end{pmatrix}$ とおき計算を繰り返すと

$\begin{pmatrix}p_n\\q_n\end{pmatrix}=P^n\begin{pmatrix}0.8\\0.2\end{pmatrix}$ となる。行列 P について

　ハミルトン･ケーリーより $P^2-1.2P+0.2E=0$ が成り立つ。また剰余の定理で

$x^n=\dfrac{1}{5}(x-1)(5x-1)Q(x)+ax+b$ とすると

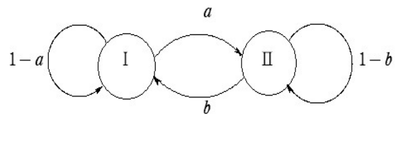
$a=\dfrac{5}{4}\left(1-\dfrac{1}{5^n}\right)$　$b=\dfrac{1}{4}\left(1-\dfrac{5}{5^n}\right)$ になるから

$$\begin{aligned}P^n&=(P^2-1.2P+0.2E)Q(P)+aP+bE\\&=\frac{5}{4}\left(1-\frac{1}{5^n}\right)P-\frac{1}{4}\left(1-\frac{5}{5^n}\right)E\\&=\begin{pmatrix}\dfrac{1}{2}+\dfrac{1}{2}\cdot\dfrac{1}{5^n}&\dfrac{1}{2}-\dfrac{1}{2}\cdot\dfrac{1}{5^n}\\\dfrac{1}{2}-\dfrac{1}{2}\cdot\dfrac{1}{5^n}&\dfrac{1}{2}+\dfrac{1}{2}\cdot\dfrac{1}{5^n}\end{pmatrix}\end{aligned}$$

となり $p_0=0.8$　$q_0=0.2$ として計算すると（１）と同様の結果を得る。

研究
　状態ⅠからⅡ、ⅡからⅠになる確率をそれぞれａ、ｂとするとき、ＡとＢがそれぞれ勝つ確率とａ，ｂとの関係を考察しなさい。

図１３



$$P=\begin{pmatrix}1-a&b\\a&1-b\end{pmatrix}$$

　初期値を同じ $p_0=0.8$　$q_0=0.2$とし、Pの値を変えて図１４～１７の場合を表計算ソフトを使って計算してみよう。

図１４

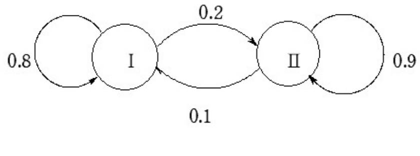


図１５

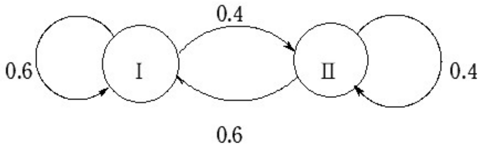


図１６

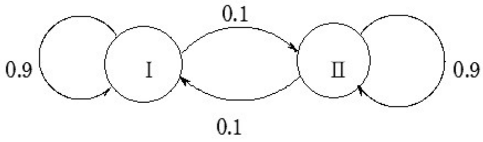


図１７

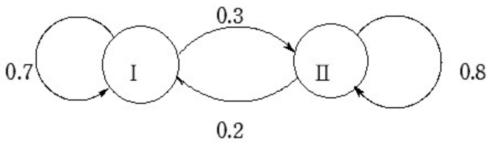


図14の確率

第n日	勝つ確率		A-B
	A	B	
初日	0.8	0.2	0.6
1	0.66	0.34	0.32
2	0.562	0.438	0.124
3	0.4934	0.5066	-0.01
4	0.44538	0.55462	-0.11
5	0.411766	0.588234	-0.18
	・	・	
	・	・	
29	0.33334836	0.66665164	-0.33
30	0.33334385	0.66665615	-0.33

図15の確率

第n日	勝つ確率		A-B
	A	B	
初日	0.8	0.2	0.6
1	0.6	0.4	0.2
2	0.6	0.4	0.2
3	0.6	0.4	0.2
4	0.6	0.4	0.2
5	0.6	0.4	0.2
6	0.6	0.4	0.2
7	0.6	0.4	0.2
8	0.6	0.4	0.2
9	0.6	0.4	0.2

図16の確率

第n日	勝つ確率		A-B
	A	B	
初日	0.8	0.2	0.6
1	0.74	0.26	0.48
2	0.692	0.308	0.384
3	0.6536	0.3464	0.307
4	0.62288	0.37712	0.246
5	0.598304	0.401696	0.197
	・	・	
	・	・	
29	0.50046423	0.49953577	0.001
30	0.50037138	0.49962862	0.001

図17の確率

第n日	勝つ確率		A-B
	A	B	
初日	0.8	0.2	0.6
1	0.6	0.4	0.2
2	0.5	0.5	0
3	0.45	0.55	-0.1
4	0.425	0.575	-0.15
5	0.4125	0.5875	-0.18
6	0.40625	0.59375	-0.19
7	0.403125	0.596875	-0.19
8	0.4015625	0.5984375	-0.2
9	0.40078125	0.59921875	-0.2

【行列の分類】

一般論として $P=\begin{pmatrix}1-a & b\\ a & 1-b\end{pmatrix}$ を分類すればよい。

特別な場合として最初に $(a\quad b)=(0\quad 0)$, $(1\quad 1)$ のときを考える。

（１） $a=b=0$ のとき　（２） $a=b=1$ のとき

$$P=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\qquad P=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$$

図１８

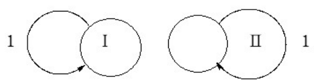
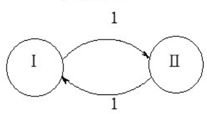


図１９



（３） $(a\quad b)\fallingdotseq(1\quad 1)$, $(0\quad 0)$ のときを考えよう。

$0\leqq a,b\leqq 1$ かつ $(a,b)\fallingdotseq(0,0),(1,1)$ として

$P=\begin{pmatrix}1-a & b\\ a & 1-a\end{pmatrix}$ で $|P-\lambda E|=0$ から固有値を求めると $\lambda_1=1\quad \lambda_2=1-a-b$ で、$\lambda_1\quad\lambda_2$ に対する固有ベクトルはそれぞれ $\begin{pmatrix}b\\ a\end{pmatrix}$ と $\begin{pmatrix}1\\ -1\end{pmatrix}$ となる。

$Q=\begin{pmatrix}b & 1\\ a & -1\end{pmatrix}$ とおくと $ab\fallingdotseq 0$ であるから逆行列が存在して　$Q^{-1}=\dfrac{1}{a+b}\begin{pmatrix}1 & 1\\ a & -b\end{pmatrix}$

$Q^{-1}PQ=\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}$ から $a+b=c$ とおくと

$$P^n=Q\begin{pmatrix}{\lambda_1}^n & 0\\ 0 & {\lambda_2}^n\end{pmatrix}Q^{-1}$$

$$=\frac{1}{c}\begin{pmatrix}b+a(1-c)^n & b-b(1-c)^n\\ a-a(1-c)^n & a+b(1-c)^n\end{pmatrix}$$

$0<a+b=c<2$ であるから $0\leqq|1-c|<1$ で

$$\lim_{n\to\infty}P^n=\frac{1}{a+b}\begin{pmatrix}b & b\\ a & a\end{pmatrix}$$

初期値 p_0、 q_0 は $p_0+q_0=1$ を満たすから

$$\lim_{n\to\infty}\begin{pmatrix}p_n\\ q_n\end{pmatrix}=\lim_{n\to\infty}P^n\begin{pmatrix}p_0\\ q_0\end{pmatrix}=\frac{1}{a+b}\begin{pmatrix}b(p_0+q_0)\\ a(p_0+q_0)\end{pmatrix}$$

$$=\frac{1}{a+b}\begin{pmatrix}b\\ a\end{pmatrix}$$

となり、 p_n と q_n の極限は $(a\quad b)\fallingdotseq(1\quad 1)$, $(0\quad 0)$ の場合には初期値 p_0、 q_0 に依存しないことも分かる。