§１ どこで最小・最大となるか？

<table>
<tr><td colspan="4">MIN2=.2.1 1.5 2.9 3 3.3 3.7 4.7 4.2 4.8 4.9 3 2.9 1.1 1.8
MIN2=:MIN2,2 3.7 5 3.3 0.5 0.9 0.5 _1.4 _0.4 2.3 3.9 8.4 7.2 6.7
MIN3=:6.7 4.6 5.2 4 5 6.7 13.4 9.1 6.6 4.6 7.7 6.1 4.6 9.2 6.2 4.3
MIN3=:MIN3,3.5 5.6 6.9 8 6.3 8.4 5.4 4.2 3 10.4 8.4 6.3 12 10.4 7.9
MIN4=:8 11.1 10.5 10.7 13.3 12.6 10.9 10.3 12.9 12.6 10.5 8.8 9.5 12 10.6
MIN4=:MIN4,8.9 9.9 11.2 11 12.1 13.5 14.5 9.9 9.6 11.1 11.8 11.3 14.1 13.2 16.1
【東京地区の平成９年２月、３月、４月の最低気温のデータ】</td></tr>
<tr><td><./ MIN2
_1.4</td><td><./ MIN3
3</td><td><./ MIN4
8</td><td>「<./」は右引数の最小値を出力する。</td></tr>
<tr><td colspan="3">]MF2=:(]=<./)MIN2
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
]MF3=:(]=<./)MIN3
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
0
]MF4=:(]=<./)MIN4
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td><td>右引数で与えたデータの最小値に一致する“位置”に「1」を与え、他の位置には「0」を与える。</td></tr>
<tr><td colspan="3">]MH2=:(=<./)MIN2
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
]MH3=:(=<./)MIN3
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
0
]MH4=:(=<./)MIN4
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td><td>上の一連の結果と同じである。

「]=<./」はフォーク
「=<./」はフック</td></tr>
<tr><td>MF2 -: MH2
1</td><td>MF3 -: MH3
1</td><td>MF4 -: MH4
1</td><td>「-:」は左右が一致すれば「1」を返す。</td></tr>
<tr><td>MF2 # 1+i.28
22</td><td>MF3 # 1+i.31
25</td><td>MF4 # 1+i.30
1</td><td>左引数で１に対応する右引数の値</td></tr>
<tr><td colspan="3">]MX2=:(=>./)MIN2
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
]MX3=:(=>./)MIN3
0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0
]MX4=:(=>./)MIN4</td><td></td></tr>
</table>

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1			
MX2 # 1+i.28 26	MX3 # 1;i.31 7	MX4 # 1+i.30 30	2 月の最低気温が最大の日は 26 日

§ 2 統計数値の切捨て・切上げ・四捨五入

<table>
<tr><td colspan="3">　]M=:2 14$MIN2
2.1 1.5 2.9 　3 3.3 3.7 4.7 　4.2 　4.8 4.9 　3 2.9 1.1 1.8
　2 3.7 　5 3.3 0.5 0.9 0.5 _1.4 _0.4 2.3 3.9 8.4 7.2 6.7
【東京地区の平成９年２月の最低気温のデータを２×14 の形のテーブルにしてＭに挿入】</td></tr>
<tr><td colspan="2">　<. M
2 1 2 3 3 3 4 　4 　4 4 3 2 1 1
2 3 5 3 0 0 0 _2 _1 2 3 8 7 6
【Ｍのデータの小数点以下の値を切捨て】</td><td>　>. M
3 2 3 3 4 4 5 　5 5 5 3 3 2 2
2 4 5 4 1 1 1 _1 0 3 4 9 8 7
【Ｍのデータの小数点以下の値を切上げ】</td></tr>
<tr><td colspan="2">　(<.@+&0.5) M
2 2 3 3 3 4 5 　4 5 5 3 3 1 2
2 4 5 3 1 1 1 _1 0 2 4 8 7 7
　]C=:0": M
2 2 3 3 3 4 5 　4 　5 5 3 3 1 2
2 4 5 3 1 1 1 _1 　0 2 4 8 7 7
　]D=:". 0": M
2 2 3 3 3 4 5 　4 5 5 3 3 1 2
2 4 5 3 1 1 1 _1 0 2 4 8 7 7</td><td>Ｍのデータの小数点以下を四捨五入している。

上の結果と同じに見えるが、書式演算子「":」の結果は数値でなく”文字”である。

「":」の逆演算である「". 」という演算子により“数値化”すれば、「(<.@+&0.5) M」の結果と同じになる。</td></tr>
<tr><td>　1+C
|domain
error
|　1　+C</td><td>　1+D
3 3 4 4 4 5 6 5 6 6 4 4 2 3
3 5 6 4 2 2 2 0 1 3 5 9 8 8</td><td>「1+C」は数値と文字の和であるから、“domain error”となる。「1+D」は数値同士の和であるからエラーにならない。</td></tr>
</table>

…………「Ｊ」言語メモ ……………………

<table>
<tr><td>　+&0.5 1.4
1.9
　0.5&+ 1.4
1.9</td><td>　(]+0.5"_)1.4
1.9</td><td>「+」という演算子に定数「0.5」を「&」という接続詞で連結すると「0.5 を加える」という動詞になる。定数と動詞を逆順にしても「]+0.5"_」というフォークでも同じ結果</td></tr>
<tr><td>　(<.@+&0.5) 1.4
1</td><td>　(<.@+&0.5) 1.5
2</td><td>「(<.@+&0.5)」という関数は「小数点以下を四捨五入する」という動詞になる。</td></tr>
<tr><td colspan="2">　5.2": X=:1.23456 3.14159
1.23 3.14</td><td>書式関数「":」の左に“5.2”といった数値を入力すると、5 のスペースに小数点２桁で四捨五入した数値を表示(結果は文字！)</td></tr>
</table>

7.4″: X 1.2346 3.1416	7のスペースに小数点4桁で四捨五入

§３ データの分類と分類されたデータの平均と分散の計算

```
clsfy=:3 :'(0.45+~.u),:+/"1=u=./:~<.y.'
meanc=:([:+/*/)%+/@{: NB. 分類されたデータの平均を求める関数
varc=:([:+/{:*[:*:{.-meanc)%+/@{: NB. 分類されたデータの分散
```

MIN4=:10+(_20 11 5 7 33 26 9 3 29 26 5 _12 _5 20 6)%10 MIN4=:MIN4,10+(_11 _1 12 10 21 35 45 _1 _4 11 18 13 41 32 61)%10 【東京地区の平成９年４月の最低気温のデータ】

]D=:10{.MIN4 8 11.1 10.5 10.7 13.3 12.6 10.9 10.3 12.9 12.6]A=:/:~D 8 10.3 10.5 10.7 10.9 11.1 12.6 12.6 12.9 13.3	MIN4 から最初の 10 個(1 日から 10 日まで)を取り出しＤに挿入している。 Ｄを大小順に並べ直したものをＡに挿入している。
]B=:<.A 8 10 10 10 10 11 12 12 12 13]F=:+/"1= B 1 4 1 3 1]X=:0.45+~.B 8.45 10.45 11.45 12.45 13.45	Ａの数値の小数点以下を“切り落とし”て、整数値にしてＢに挿入している。 同じ数値ごとの度数を与えている。「=」の片側形は「~.=/]」のように演算する。 「~.」は“重複要素を排除する”演算子で、Ｂの重複要素を除き「0.45」を加えている。
X,:F 8.45 10.45 11.45 12.45 13.45 1 4 1 3 1]XF=:clsfy D 8.45 10.45 11.45 12.45 13.45 1 4 1 3 1	X(分類されたクラスの代表値)にＦ(度数)を下に付加したテーブルである。なお、10 度台のデータは 9.95-10.94 の範囲の計測値が記述されるので、中間点は“10.45”となる。 「clsfy」は、分類幅が“１”という特殊な場合の度数分布表を与える関数である。

(+/X*F)%+/F 11.25	meanc XF 11.25	mean D 11.29	分類されたデータ XF の平均とデータＤから計算した平均は多少異なる。
(+/F**:X-M)%+/F 1.96	varc XF 1.96	var D 2.3029	分類されたデータ XF の分散とデータＤから計算した分散は多少異なる。

]XF=:clsfy MIN4	４月の 30 日間の全データ MIN4 を分類した結果の度数分布表である。

8.45 9.45 10.45 11.45 12.45 13.45 14.45 16.45 3 4 6 6 5 3 2 1		
meanc XF 11.4167	varc XF 3.63222	度数分布表からの平均と分散の計算結果で、(Dから直接計算は11.4167;3.23)

§４ 統計データの分類と集計（級間隔を任意に与えた場合）

```
acum=:[:+/"1[:=[:/:~<.@(%~)

number=:(1:~:*@<./@])+>.@((>./-<./)@]%[)  NB.  value の補助関数
value=:([:(]-**0:=2:|])<.@<./)@]+[:+:[:i.number
table=:value,:acum [ meanc=:+/@:*/%+/@:{:
 NB. 左引数で指定した級間隔で右引数のデータを分類する関数
```

`2 14 $ D=:2 %~ M2`
`1.05 0.75 1.45  1.5 1.65 1.85 2.35  2.1  2.4 2.45  1.5 1.45 0.55  0.9`
`   1 1.85  2.5 1.65 0.25 0.45 0.25 _0.7 _0.2 1.15 1.95  4.2  3.6 3.35`
［東京の平成９年２月の最低気温のデータ M2 を２で割った D を２×14 のテーブルで表示］	

`<. D`
`1 0 1 1 1 1 2 2 2 2 1 1 0 0 1 1 2 1 0 0 0 _1 _1 1 1 4 3 3`
`]C=:/:~ <. D`
`_1 _1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 3 3 4`
	D を切捨てて整数値にしている。
	
	上の結果を「/:~」で大小順に並べている。

`]T=:=C`
`1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0`
`0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0`
`0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0`
`0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0`
`0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0`
`0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1`
	「=」という演算子の片側形は「~.=/]」と同じ演算を行う。「~.」は「重複要素を排除する」という演算子である。

`+/"1 T` `2 6 12 5 2 1`	`2 value M2` `_1 1 3 5 7 9`	T というテーブルを 1-セルについて加えて「度数分布」と「級代表値」を与えている。

`]D2=:2 table M2`
`_1 1  3 5 7 9`
` 2 6 12 5 2 1`
	`]D3=:2 table M3`
	`3  5  7 9 11 13`
	`2 10 10 5  2  2`
		`]D3=:2 table M4`
		`7  9 11 13 15`
		`7 12  8  2  1`
			「table」という関数によって、2,3,4 月の気温の度数分布

`(mean M2);meanc D2`	`(mean M3);meanc D3`	`(mean M4);meanc D4`	

3.08929 3.14286	6.79677 7.06452	11.4167 9.53333
]D5=:(1+{.D4),:{:D4 8 10 12 14 16 7 12 8 2 1	meanc D5 10.5333	４月の気温に関しては、級代表値の与え方にやや不適切さがあったために、元データからの平均とずれが生じた。

§５ 幹葉図 ：（片側形・両側形)のプログラム

```
   sld_m=:3 :0
integer=:[:+/[:*/'0'"_=[:":,.
decimal=:[:-[:{:[:>./([:#;._1'.'"_,":)"0
grid=:integer`decimal@.('.'"_ e.":)
digit=:(0:>._0.1"_*grid)":,. NB. digitize after listing
leaf=:_1:`([:-1:+grid)@.(grid>.0:){"1[:digit/:~
drop=:_1:`(_2:)`(_1:-grid)@.(1:+[:*1:+grid)
stem0=:drop}."1 digit NB. stem part(not ordered)
stem=:[:stem0/:~ NB. ordered stem
add0=:(0:>.1:+grid)$'0'"_
sld=:(~.@stem,"1 add0);=@stem#leaf
'program set of sld(monadic)'
)
   sld_d=:3 :0
stemb=:#@[({.;}.)[:stem0[:;[:/:~&.>;
stemd=:[:~.[:stem,
shaped=:([:<"1 stemd)=/L:1[:<"1 L:0 stemb
leafd=:shaped#L:0[:leaf L:0;
sldd=:([:|."1@>@{.leafd);stemd;{:@leafd
'program set of sld(dyadic)'
)
```

```
MIN2=.2.1 1.5 2.9 3 3.3 3.7 4.7 4.2 4.8 4.9 3 2.9 1.1 1.8
MIN2=:MIN2,2 3.7 5 3.3 0.5 0.9 0.5 _1.4 _0.4 2.3 3.9 8.4 7.2 6.7
MIN3=:6.7 4.6 5.2 4 5 6.7 13.4 9.1 6.6 4.6 7.7 6.1 4.6 9.2 6.2 4.3
```

```
MIN3=:MIN3,3.5 5.6 6.9 8 6.3 8.4 5.4 4.2 3 10.4 8.4 6.3 12 10.4 7.9
MIN4=:8 11.1 10.5 10.7 13.3 12.6 10.9 10.3 12.9 12.6 10.5 8.8 9.5 12 10.6
MIN4=:MIN4,8.9 9.9 11.2 11 12.1 13.5 14.5 9.9 9.6 11.1 11.8 11.3 14.1 13.2
16.1
```

【東京地区の平成９年２月、３月、４月の最低気温のデータ】

```
   D=:10{.MIN4
8  11.1  10.5  10.7  13.3  12.6  10.9  10.3  12.9
12.6
   ]A=:/:~D
8  10.3  10.5  10.7  10.9  11.1  12.6  12.6  12.9
13.3
```

MIN4 のデータから最初の 10 個を取り出しＤに挿入している。

Ｄを大小順に並べ直したもの。
「/:~D」は「D/:D」と同じである。

digit A	stem A
8.0	8
10.3	10
10.5	10
10.7	10
10.9	10
11.1	11
12.6	12
12.6	12
12.9	12
13.3	13
【データの文字化】	【幹を分離】

```
   leaf A
0357916693
```

(~.@stem,"1 add0)A	(=@stem#leaf)A
8	0
10	3579
11	1
12	669
13	3

sld D

8	0
10	3579
11	1
12	669
13	3

sld MIN2	MIN2 sldd MIN3	MIN3 sldd MI

_1	4
_0	4
0	559
1	158
2	01399
3	0033779
4	2789
5	0
6	7
7	2
8	4

4	_1	
4	_0	
955	0	
851	1	
99310	2	
9773300	3	05
9872	4	023666
0	5	0246
7	6	12336779
2	7	79
4	8	044
	9	12
	10	44
	12	0
	13	4

50	3	
666320	4	
6420	5	
97763321	6	
97	7	
440	8	089
21	9	5699
44	10	97763321
	11	011238
0	12	01669
4	13	235
	14	15
	16	1

§６　２項係数とパスカルの三角形

```
bic=:i.&>:!] NB. ２項係数を求める関数
pascal=:":@bic"0@i.@>: NB. パスカルの三角形
```

i.&>: 4 0 1 2 3 4	右引数の値に「>:」で１を加えてから「i.」により整数値を生成している。
(0 1 2 3 4)!4 1 4 6 4 1 bic 4 1 4 6 4 1	4の場合の全ての２項係数を与えている。(「i.&>:4)！4」のように演算するフォーク)
(<bic 3),<bic 4 │1 3 3 1│1 4 6 4 1│ (bic 3);bic 4 │1 3 3 1│1 4 6 4 1│	3､4の場合２項係数を「<」という演算子により、ボックスで囲んで表示している。 「;」の両側形は左右をボックスで囲みながら表示する。
bic"0(3 4) 1 3 3 1 0 1 4 6 4 1 ":@bic"0(3 4) 1 3 3 1 1 4 6 4 1	ランク０の要素(つまりアトム)ごとに実行。ただ２つの結果の形を整えるために、3のときの結果の最後に「0」が付いてしまう。 「":」という演算子により文字化しながら２項係数を求めると、結果のないところには「空白」が付くことで、０が表われない。
<pascal 4 1 1 1 1 2 1 1 3 3 1 1 4 6 4 1	めでたくn=4までのパスカルの三角形を求めボックスで囲みながら表示している。 なおこの結果は文字列のテーブルであるから、演算に用いるときには「":」の逆演算である「".」により数値化する必要がある。
,.<"1 pascal 4 │1│ │1 1│ │1 2 1│ │1 3 3 1│ │1 4 6 4 1│	「<"1」によりランク１の要素(リスト)に関してボックスをかけて（「, .」によってテーブル化して）みると、空欄が同じ長さだけついていることが確かめられる。

§７　２項分布

```
binom=:4 :'(k!x.)*(y.^k)*(1-y.)^|.k=.i.>:x.'
```

NB.　２項分布の確率関数を与える両側形の関数

```
bden=:bic@[*(]^i.@>:@[)*-.@]^[:|.i.@>:@[ [ bic=:i.&>:!]
```

NB.　２項分布の確率関数の関数型関数(Tacit Definition)

<table>
<tr><td colspan="2">   ]K=.i.>:4
0 1 2 3 4
   |.K
4 3 2 1 0</td><td>0から始まる4に1を加えた５個の連続した整数値を変数Kに挿入している。

「|.」は右引数の数列を“逆順にする演算子である。</td></tr>
<tr><td colspan="2">   (1-0.6)^|.K
0.0256 0.064 0.16 0.4 1
   ]A=:(-.0.6)^|.K
0.0256 0.064 0.16 0.4 1
   ]B=:0.6^K
1 0.6 0.36 0.216 0.1296
   ]C=:K!4
1 4 6 4 1
   A*B*C
0.0256 0.1536 0.3456 0.3456 0.1296</td><td>（1-0.6）の4乗,3乗,2乗,1乗,0乗の値を求めている。

上と同じ結果で、Aに挿入して表示している。
「-.y」は“1-y”と同じである。
0.6の0乗、1乗、2乗、3乗、4乗の値を求めている。
nが4の場合の２項係数を与えている。

上の３種類の演算結果を要素ごとに掛けている。</td></tr>
<tr><td colspan="2">   4 binom 0.6
0.0256 0.1536 0.3456 0.3456 0.1296
   4 bden 0.6
0.0256 0.1536 0.3456 0.3456 0.1296</td><td>上の結果と一致。つまり２項分布の確率関数の明示型定義(Explicit Definition)。

関数型定義(Tacit Definition)でも同じ結果が得られる。</td></tr>
<tr><td>   +/ 4 binom 0.6
1</td><td>   +/ 4 bden 0.6
1</td><td>確率関数であるから、合計は当然１である。</td></tr>
<tr><td colspan="2">   4 binom"0(0.6 0.5)
0.0256 0.1536 0.3456 0.3456 0.1296
0.0625   0.25 0.375   0.25 0.0625
   4 bden"0(0.6 0.7)
0.0256 0.1536 0.3456 0.3456 0.1296
0.0081 0.0756 0.2646 0.4116 0.2401</td><td>右引数にPの値を複数個挿入することも可能である。但しその場合には、0のランクを指定することが必要である。</td></tr>
</table>

§８ (標準)正規分布

```
nden=:([:^-@-:@*:)%(%:@o.2)"_  NB. 標準正規分布の確率密度関数
ndf0=:3 :'+/d*-:(}:+}.)nden,(i.1001)*d=.y.%1000'
ndf=:0.5" _+ndf0 NB. ndfs=:[:-/[:ndf"0|. NB. 分布関数
ndfs=:[:-/[:ndf0"0|. NB. 右引数で与えた区間での積分値を与える。
```

]D=:%:o.2 2.50663	$\sqrt{2\pi}$ の値を与えている。「%:」は平方根、「o.」は円周率 π を与える演算子。
]A=:-@-:@*: 0 0.5 1 2 0 _0.125 _0.5 _2]B=:^A 1 0.882497 0.606531 0.135335 B % D 0.398942 0.352065 0.241971 0.053991	「0,0.5,1,2」の２乗の半分の値の符号を変えた値。「*:」は平方値、「-:」は半分にする演算子「$e^{-0}, e^{-0.125}., e^{-0.5}., e^{-1}.$」の値を与えている。 「^」は指数の値を出力する演算子。 上の値を D=$\sqrt{2\pi}$ で割った値を与えている。
nden 0 0.5 1 2 0.398942 0.352065 0.241971 0.053991	標準正規分布の確率密度関数の「0,0.5,1,2」での値を示している。

ndf0"0(0.5 1 1.65 1.96) 0.191462 0.341345 0.450528 0.475002 ndf0"0(_1 _0.5) _0.341345 _0.191462 ndf"0(_1 _0.5 1.65 1.96) 0.158655 0.308538 0.950528 0.975002		標準正規分布の確率密度関数の(0,x)の範囲での積分(x は右引数で入力)。 マイナスの値を入力すると、値も負になる。 標準正規分布の分布関数の値を出力する。
(ndf0 1)+ndf0 0.5 0.532807 ndfs _0.5 1 0.532807		(_0.5,1)の範囲での積分 分布関数「ndfs」という関数を用いても積分の値が求められる。
(ndf0 1)-ndf0 0.5 0.149882 ndfs 0.5 1 0.149882		(0.5,1)の範囲での積分 分布関数「ndfs」を用いた場合には、右引数に区間の数値を入力すればよい。
ndfs _1 _0.5	ndfs _0.5 _1	(_1,_0.5)の範囲での積分

0.149882	_0.149882	(_0.5,_1)と入力すれば負の値を出力する。

§１０ 標本比率の区間推定

```
width=:([:%:([+>:@[*-~)%])%]+4: NB. 標本比率の標準誤差を与える(両側関数)
ratio=:(([+2:)%]+4:)-+:@(,-)@width NB. 比率に対する信頼度 95%の信頼区間
```

275 -~ 400 125	右引数の値から左引数の値を引いている。「~」は左右の引数を反対にする副詞である
275([+>:@[*-~)400 34775	「x([+>:@[*-~]y」は「x+(>:x)*(y-x))」という値を求めている。
275([:%:([+>:@[*-~)%]+4:)400 9.27775	前の結果を右引数の値で割ってから平方根をとっている。
275(([:%:([+>:@[*-~)%])%]+4:)400 0.0230793	前の結果を右引数の値に 4 を加えた値で割算し、標本比率の標準誤差を求めている。
]W=:275 width 400 0.0230793	標本比率に対する標準誤差が与えられ、結果をＷという変数に挿入して表示する。
(,-)W 0.0230793 _0.0230793 275 +:@(,-)@width 400 0.0461586 _0.0461586	「(, -)w」はフックで、「W,-W」という結果と同じである 「width」に「+:@(, -)」を接続した結果で、前の結果の２倍の値が与えられる。
275 (([+2:)%]+4:) 400 0.685644	左引数に 2 を加えた値([+2:]を右引数に 4 を加えた値(]+4:)で割算している。
275((([+2:)%]+4:)-+:@(,-)@width)400 0.639485 0.731802 275 ratio 400 0.639485 0.731802	直上の値からその前の値を引いている。 前の結果と同じで、「ratio」の演算結果は、比率に対する信頼度 95%の信頼区間を与える。この関数の左引数にはある属性をもつサンプル数、右引数には観測数を入力する

…………「Ｊ」言語メモ ……………………

上で用いた「4:」や「2:」などの動詞は、常に同じ定数を取り出すという関数で、_9 から 9 までの整数値にコロン(：)をつけて定義され、結構、重宝な「動詞」である。さらに、他の任意の定数を“動詞化”するには、「10”_」,「(o.2)”_」といったように、右に(“_)をつけてやればよい。また「&」という接続詞を用いて以下のようにしてもよい。

275(2&+@[,4&+@])400	275(([+2:),]+4:)400

277 404	277 404

§１１ (正規分布の)平均の推定

```
mean_sd=:[:%:([:mean*:@(-mean))%# NB. 標本平均の標準誤差を与える関数
mean_est=:mean@]([;[-+:@(,-)@])mean_sd@] NB. 平均の点推定と区間推定を与える。
nrnd0=:3 :'_6+(+/?(12,y.)$1000)%1000' NB. 標準正規分布の乱数を発生
nrand=:[:([:".0:":])5:+2:*nrnd0 NB. ほぼ 0-10 の範囲の整数型正規乱数を発生
```

] D=.nrand 10 6 5 4 5 9 6 6 4 3 6	「nrand」という関数は、平均が 5 で分散が 4 の(離散型)正規乱数を生成する。
]M=:(mean=:+/%#)D 5.4	D というデータの平均を求め M に挿入している。
(-mean)D 0.6 _0.4 _1.4 _0.4 3.6 0.6 0.6 _1.4 _2.4 0.6	D というデータの平均からの偏差を求めている。
([:mean*:@(-mean))D 2.44 ([:%:([:mean*:@(-mean))%#)D 0.493964]S=:mean_sd D 0.493964	上の値の平方値の平均で、つまり「分散」を求めている。 上で求めた分散の値をデータの個数で割ってから平方根をとっている。 標準偏差をデータ数の平方根で割った値で、いわゆる「標準誤差」に他ならない。
+:@(,-) S 0.987927 _0.987927	「(,-)S＝S,-S」はフックで、「+:」という演算子で 2 倍の結果が得られる。
M([;[-+:@(,-)@])S 5.4 │ 4.41207 6.38793 mean_est D 5.4 │ 4.41207 6.38793	ボックスの左側に平均値 M、右側には M - 2S と M+2S の値(信頼区間)が与えられている。「；」という演算子は、「左右の値をボックスで囲みながら連結する」という動詞である。

mean_est nrand 10 4.7 │ 3.40231 5.99769 mean_est nrand 10 5.3 │ 4.0655 6.5345 mean_est nrand 50 4.9 │ 4.27199 5.52801	乱数の値がその都度異なるから、結果は実験の都度変化する。 「nrand」は平均の整数乱数を発生させ、信頼区間はいずれも 5 をカバーしている。 サンプル数が多くなると、信頼区間の幅は狭くなる。

§１２ 相関係数に対する区間推定

```
sdev=:[:%:[:mean*:@dev=:-mean=:+/%# NB. 偏差と標準偏差を与える関数
cor=:[:mean[:*/(dev%sdev)&> NB. 左引数と右引数で与えたデータの相関係数
corest_d=:4 :'(<:%>:)^(^.(>:%-.)y.)-(,-)4%%:x.-3' NB. 相関係数の信頼区間
corest_m=:([:{.#&>)corest_d cor NB. 右引数にボックス形で与える片側関数
```

(,-)%: 100-3 9.84886 _9.84886	「n-3=97」の平方根とその反対符号を与えている。「(,-)y=y,-y」はフックである。
]W=:4%(, -)%: 97 0.406138 _0.406138	4を上の値で割ったものをWに挿入している。つまり推定幅を与えている。
(>:%-.)0.5 3	「-.」は補数(つまり1 - y)を与える演算子で、(1+y)/(1 - y)という値を与える。
]Q=:^.@(>:%-.)0.5 1.09861	上の値の自然対数をQに挿入する。つまり、フィッシャー変換を行なっている。
]S=:Q - W 0.692474 1.50475	Qの値に推定幅をマイナス・プラスして、変数Sに挿入する。
(<:@^%>:@^) S 0.333034 0.636564	「($(e^s - 1)\%(e^s + 1)$)」という値を求めている。
100 cor_est 0.5 0.333034 0.636564	上記の一連の演算を行なっている。つまり、標本数を左引数に標本相関係数を右引数に入力すると、相関係数に対する信頼度95%の信頼区間を出力している。

SM=:9820 13836 11506 8330 8761 11744 12769 15356 12316 13540 NB. 男子自殺者数

SW=:6491 8641 8637 6114 6967 8231 7773 8027 7772 6976 NB. 女子自殺者数

RUE=:1.2 2.5 1.7 1.2 1.1 1.9 2.0 2.6 2.1 2.5 NB. 完全失業率(%)

cor SM ; RUE 0.968921	男子自殺者数と完全失業率の相関係数を求めている。
10 corest_d cor SM ; RUE 0.866391 0.993063 corest_m SM;RUE 0.866391 0.993063	(男子自殺者数と完全失業率の)相関係数に対する信頼区間を出力する両側形の関数。 (男子自殺者数と完全失業率の)相関係数に対する信頼区間を出力する片側形の関数。
corest_m SW ; RUE	女子自殺者数と完全失業率の信頼区間では、結果

_0.0642458 0.895202	が 0 を含んでいるので無相関の可能性すらある。

§１３ 観測比率の差に関する検定

```
ratio_dif=:3 :'(-/%/&>y.)%%:(+/%{:&>y.)*(*-.)%/+/>y.'
NB. ボックス形で与えた２組の観測比率の差に関する検定統計量
ratio_t=:([:-/%/&>)%[:%:+/@([:%{:&>)*(*-.)@([:%/+/@:>)
```

]RD=:69 300 ; 81 300 69 300 81 300　　# RD 2		ボックスで囲まれたデータをRDという変数に入力している。 RDのアイテム数は2で、ボックスで囲まれた１つ１つが１個の要素とみなされる。
>RD 69 300 81 300	; RD 69 300 81 300	「<. 」という演算子により、ボックスを開くと、テーブルの形の変数になる。「；」で開くと、リストになる。
+/>RD 150 600	%/+/>RD 0.25	2組のデータを合わせたときの比率を与えている。
]V=:(*-.) 0.25 0.1875		「0.25＊(1-0.25)」と計算を行なうフックで、差がないとしたときの分散の推定値
{:&> RD 300 300	]R=:+/%{:&> RD 0.00666667	ボックスごとに末尾の要素を取り出す。 上の結果の逆数を求めてから足している。
%:R * V 0.0353553		上の結果に分散を掛けてから平方根をとったもので、「標準誤差」を与えている。
]S=:%:(+/%{:&>RD)*(*-.)%/+/>RD 0.0353553 -/%/&> RD _0.04 (-/%/&> RD)%S _1.13137		標準誤差を求めて、Sという変数に挿入している。 ボックスごと求めた比率の差を求めている。 上の値を標準誤差のSで割った値で、差がないという帰無仮説の下ではｔ－分布。
ratio_dif RD _1.13137	ratio_t RD _1.13137	上記の一連の演算を行なっている。つまり、比率の差の検定統計量を与えている。

]SD=:456 1200 ; 504 1200 456 1200　504 1200		内閣の支持率調査の2回の結果(架空のデータ)から「人気は盛り返したか？」
ratio_dif SD	Ratio_t SD	「支持率に変化なし」とする“帰無仮説”は

_2	_2	棄却される。

§１４ 平均の差に関する検定

```
ssdev=:[:+/*:@(-mean=:+/%#) NB. 偏差平方和を与える関数。
mean_dif=:3 :'(-/mean&>y.)%%:(+/ssdev&>y.)*(+/%*/*+/-2:)#&>y.'
meandif=:-/@:(mean&>)%[:%:(+/%*/*+/-2:)@:(#&>)*[:+/ssdev&>
```

<table>
<tr><td colspan="2">M1=:11 13 12 9 2 2 1 0 0 1 0 0
M2=:10 7 14 5 1 11 11 6 14 13 11
M2=:M2,14 13 15 7 13 9 2</td><td>M1 は水戸泉(幕内力士)の平成 3･4 年の 12 場所の番付位置のデータ(0 は三役)

M2 は平成 5 年以降 18 場所の番付位置</td></tr>
<tr><td colspan="2">   M1 #&>@; M2
12 18
   (*/*+/-2:)12 18
6048
   (+/%*/*+/-2:)12 18
0.00496032
   ]C=:M1([:(+/%*/*+/-2:)#&>@;)M2
0.00496032</td><td>M1, M2 という 2 組のデータ数を同時に算出している。

「(*/12 18)*(+/-2:)12 18＝216*(30-2)」の値である。
右引数の総和を上の演算結果の 6048 で割った結果が与えられる。
上の一連の結果を C という変数に挿入している。</td></tr>
<tr><td>   M1 mean&>@;M2
4.25 9.77778</td><td>   M1 ssdev&>@; M2
308.25 307.111</td><td>M1, M2 という 2 組のデータの平均と「偏差平方和」を同時に算出している。</td></tr>
<tr><td colspan="2">   ]S=:%:C*+/(M1 ssdev&>@; M2)
1.74711

   (M1-/@(mean&>@;)M2)%S
_3.16396</td><td>2 組の分散を加えて C を掛け平方根をとった値を S に挿入。これは、平均に差が無いという仮説の場合の「標準誤差」を出力。

平均の差の値を標準誤差で割った値で、帰無仮説のもとではｔ－分布に従う。</td></tr>
<tr><td colspan="2">   M1 mean_dif M2
_3.16396
   meandif M1;M2
_3.16396
【関数型定義でも同じ結果が得られる】</td><td>上記の一連の演算を行なっていて、平均の差に関するｔ－統計量の値を与えている。

水戸泉の平成 3･4 年の 12 場所の番付位置と平成 5 年以降の位置とは有意に異なっている(番付の下位のほうで低迷している！)。</td></tr>
</table>

mean_dif MIN2;MIN3 _5.97993 mean_dif MIN3;MIN4 _8.21659 mean_dif MIN2;MIN4 _15.5776	meandif MIN2;MIN3 _5.97993 meandif MIN3;MIN4 _8.21659 meandif MIN2;MIN4 _15.5776	東京地区の平成９年２月と３月の最低気温の間には明らかな違いがある。また３月と４月とではさらなる顕著な違いがある。 ２月と４月の間では､｢差がない｣とする仮説は検定するまでもない！

§１５ 一様性の検定

```
chi_test=:([:+/*:@(-mean))%mean NB. 一様性検定のためのカイ２乗統計量
```

DICE 10 14 10 9 10 7	サイコロ投げの 60 回の実験結果で､1 から 6 の目までの出現回数を示している。
DICEB 5 14 8 5 8 20	6 の目の出る確率を 2/7 にしたサイコロの実験結果を示している。
(-mean)DICE 0 4 0 _1 0 _3 *:@(-mean)DICE 0 16 0 1 0 9 ([:+/*:@(-mean))DICE 26 (([:+/*:@(-mean))%mean)DICE 2.6	10 という平均からの偏差を求めている。 平均からの偏差の平方値を与えている。 上の結果の総和で、偏差平方和を与えている。 偏差平方和を平均値で割った値が与えられる。
chi_test DICE 2.6 chi_test DICEB 17.4	上記の一連の演算を行なっていて、いわゆる「カイ２乗統計量」の値を示している。 偏ったサイコロの場合には、カイ２乗統計量が大きな値をとり、一様性の仮説が棄却される(自由度 5 のカイ 2 乗分布の 5%点は 11.07 である)］。
H=:28 25 22 24 19 16 14 12(架空のデータ) chi_test H 11.3	160 レースの枠ごとの優勝回数のデータ。 自由度 7 のカイ 2 乗分布の 5%点は 14.067 であるから､一様性の仮定は棄却できない。(一見外枠のほうが回数が少ない感じ！)

…………「Ｊ」言語メモ ……………………

上で用いた DICE や DICEB といったデータは、次のような関数を利用して得た結果である

```
dice=:1:+?@($6:) NB. 等確率のサイコロ投げの実験結果

diceb=:6:<.1:+?@($7:) NB. 偏ったサイコロ投げの実験結果
```

]D=:+/"1(>:i.6)=/ dice 60 8 11 7 16 10 8 chi_test D 5.4	]DB=:+/"1(>:i.6)=/ diceb 60 10 8 5 5 11 21 chi_test DB 17.6

§１６ 適合度検定

```
gf_test=:[:+/([:*:]-(*+/))%(*+/)

 NB. 左引数で与えた理論分布との適合性のためのカイ２乗統計量
```

]E=:0.5 0.5(*+/)480 420 450 450 0.5 0.5 * +/480 420 450 450 0.5 0.5(]-(*+/))480 420 30 _30 (0.5 0.5([:*:]-(*+/))480 420)%E 2 2 0.5 0.5(([:*:]-(*+/))%(*+/))480 420 2 2	左引数の理論比率を観測値の総数に掛けた理論度数をＥという変数に挿入し表示している。「x(*+/)y」は「x*(+/y)」のように演算を行なう“両側フック”である。 右引数から直上の値を引いて、理論値からの偏差を与えている。 上の結果の平方値(偏差平方値)を理論度数Ｅの値で割っている。
0.5 0.5([:+/([:*:]-(*+/))%(*+/))480 420 4	上の結果の合計値で、偏差平方和を理論度数で割った値である。
0.5 0.5 gf_test 480 420 4	上の一連の演算結果と同じで、左引数で与えた理論分布との適合性を示すカイ２乗統計量を与えている。
(6$%6)gf_test D=: 8 11 7 16 10 8 5.4 (6$%6)gf_test DB=: 5 14 8 5 8 20 17.4 ((5$%7),2%7)gf_test DB 6.96667	サイコロ投げの 60 回の実験結果で、1 から 6 の目までの出現回数を示している。 6 の目の出る確率を 2/7 にしたサイコロの実験結果の検定統計量の値を示している。 DICEB というデータを得たのと同じ確率を与えてやると、統計量の値は小さくなる。

…………統計学メモ ……………………

カイ２乗統計量の大小を判定するための上側 5%点は、自由度によって異なる。自由度が 1 から 10 までの上側 5%点は、次のように与えられている：

3.841 5.991 7.815 9.488 11.070 12.590 14.070 15.510 16.920 18.312

特に□で囲ってある値が自由度 5 の場合の値である。正しいサイコロ投げの実験結果の D に対しては“一様性の仮定”が棄却できないが、偏ったサイコロ投げの実験結果の DB に対

しては一様でないことは明らかである(99%点の 15.086 よりも大きい)。

§１７ 傾向性の検定

```
trend=:4 :'(+/y.**:p-q)%(*-.)q=.mean p=.x.%y.'
```

NB.左引数の右引数に対する比率の傾向性(一様か否か)を調べる検定統計量

X=:>:i.5 [Y=:4 8 6 8 10	XとYに値を挿入する。
]P=:X % Y 0.25 0.25 0.5 0.5 0.5]Q=:mean P 0.4 +/Y**:P-Q 0.51	「X÷Y」の値をPに挿入し結果を表示している。 Pの平均値をQに挿入し結果を表示している。 $Y \times (P-Q)^2$ の総和を与えている。
(+/Y**:P-Q)%(*-.)Q 2.125 X trend Y 2.125	上の値を「Q(1-Q)＝0.24」という値で割っている。 比率には傾向性がないという仮説の下でのカイ２乗統計量を与えている。
H=:150 250 264 302 238 176 36 N=:534 746 784 705 443 299 70 H trend N 158.702	Hは不眠症を訴えた人の数。 年齢階級別成人女子の不眠症数 Nは年齢階級ごとの調査人数。 傾向性がないという仮説の下でのカイ２乗値がかなり大きく、年齢と共に不眠症の比率が高くなるという傾向がある。

…………「統計学」メモ ……………………

観測比率 $P_i = x_i / n_i (i=1,2,\cdots\cdots\cdots,k)$ に傾向性がないとする仮説の下でのカイ２乗値は

$$CHI0(k) = \sum_{i=1}^{k} n_i \frac{(P_i - p_{0i})^2}{p_{0i}(1-p_{0i})} = \sum_{i=1}^{k} \frac{(x_i - n_i p_{0i})^2}{n_i p_{0i}(1-p_{0i})}$$

と表され、これが近似的に自由度 k のカイ２乗分布に従う変量である。

ところでこの値は、理論比率 P_{0i} が分からないと計算できないので、その値を

$\overline{P} = \{P_1 + P_2 + \cdots\cdots\cdots + P_k\}/k$ という値で代用することにすると

$$CHI\left(k\right) = \sum_{i=1}^{k} \frac{n_i\left(P_i - \overline{P}\right)^{2_i}}{\overline{P}\left(1 - \overline{P}\right)}$$

も、自由度$\left(k - 1\right)$のカイ２乗分布で近似できることになる。

§１８ 分割表の独立性の検定

```
cont_test=:+/@,@(*:@-%])+/"1*/+/%+/@,
```

NB. 右引数で与えた分割表の独立性の仮定の下でのカイ２乗統計量

]C=:2 2$45 15 25 15 45 15 25 15	2×2の分割表のデータを変数Cに挿入して表示している。
(+/%+/@,)C 0.7 0.3 +/"1 C 60 40	Cのアイテムについての総和をデータの合計100で割っている。 Cの１セルであるリスト(横方向)についての総和を与えている。
]P=:(+/"1*/+/%+/@,)C 42 18 28 12 C(*:@-%])P 0.214286　0.5 0.321429 0.75 +/,C(*:@-%])P 1.78571 cont_test C 1.78571	「0.7 0.3」「60 40」との外積(掛け算に関するクロス表)をPに挿入する。 「(C-P)の平方」をPで割ったテーブルを与えている。 上で求めた値の総和を与えている。つまり「$\sum (c_{ij} - p_{ij})^2 / p_{ij}$」の値である。 独立性の仮説の下でのカイ２乗統計量を与える。

…………「J」言語メモ ……………………

,C 45 15 25 15		「，」という演算子の片側形は右引数の要素を「リスト化」する。
,.45 15 45 15	,. i.2 2 2 0 1 2 3 4 5 6 7	「，．」の片側形は全ての要素を「テーブル化」する。
$,: i.2 1 2 $,: i.2 2		「，：」の片側形は,見かけは変わらないがランクを１つ上げたアレイを生成する。

1 2 2	

§１９ 分割表モデルの最尤推定

```
MLED= MLED=:%+/@,: NB. 分割表の従属モデルのパラメータの最尤推定
MLEI=:(+/"1*/+/)@(%+/@,) NB. 左引数に理論比率を与えたときのモデルの
AIC
```

]C=:2 2$45 15 25 15 45 15 25 15]B=:(%+/@,)C 0.45 0.15 0.25 0.15	2×2 の分割表のデータを変数 C に挿入して表示している。 C の各要素を要素の総数 100 で割った値を示している。
MLED C 0.45 0.15 0.25 0.15	上と同じ結果で、従属モデルのパラメータの最尤推定値を与える。
]E=:+/"1 B 0.6 0.4]F=:+/ B 0.7 0.3 (+/"1*/+/)B 0.42 0.18 0.28 0.12	B の 1-セル(横方向)に対する合計値を求めて E に挿入。 B のアイテム(縦方向)に対する合計値を求めて E に挿入。 「E=:+/ “ B＝0.60 0.4」と「F=:+/B=0.7 0.3」の外積を作成している。
MLEI C 0.42 0.18 0.28 0.12	上と同じ結果で、独立モデルのパラメータの推定値を与える。

…………「統計学」メモ ……………………

pq	$p(1-q)$	p
$(1-p)q$	$(1-p)(1-q)$	$1-p$
q	$1-q$	1

【独立モデル】

p_{11}	p_{12}	p
p_{21}	p_{22}	$1-p$
p	$1-q$	1

【従属モデル】

従属モデルの場合の最大対数尤度は

$$MLLD = a\log(a/n) + b\log(b/n) + c\log(c/n) + d\log(d/n) \quad (n = a + b + c + d)$$

さらに独立モデルの場合の最大対数尤度も、

$$MLLI = a\log PQ + b\log P(1-Q) + c\log(1-P) + d\log(1-P)(1-Q)$$
$$(P = (a+b)/n\,, Q = (a+c)/n \quad ; \quad n = a+b+c+d)$$

のように与えられる。

§２０ 分割表モデルの最大対数尤度と情報量規準

```
MLLI=:[:+/[:,]*^.@MLEI NB. 2×2分割表の独立モデル最大対数尤度
MLLD=:[:+/[:,]*^.@MLED NB. 2×2分割表の従属モデル最大対数尤度
AICI=:3 :'+:(+/<:$ y.)-MLLI y.' NB.分割表の独立モデルの情報量規準(AIC)
AICD=:3 :'+:(<:+/$ y.)-MLLD y.' NB.分割表の従属モデルの情報量規準(AIC)
```

]A=:MLEI C 0.42 0.18 0.28 0.12	独立モデルのパラメータの推定値を示している。
+/,C*^.A _128.388	上のＡという値の対数にＣを掛けた結果の各要素の合計である。
]A1=:(MLLI=:[:+/[:,]*^.@MLEI)C _128.388	上と同じ結果で、独立モデルの下での最大対数尤度を与えている。
+/<:$ C 2	Cの各ランクから１を引いた値の合計で、独立モデルのパラメータ数を与えている。
+:(+/<:$ C)-A1 260.775	パラメータ数から最大対数尤度を引いた値を２倍していて、AICの値である。
AICI C 260.775	上の結果と同じで、分割表の独立モデルの下での情報量規準を与えている
]B=:MLED C 0.45 0.15 0.25 0.15	従属モデルのパラメータの推定値を示している。
+/,C*^.B _127.504	上のＢという値の対数にＣを掛けた結果の各要素の合計である。
]B1=:(MLLD=:[:+/[:,]*^.@MLED)C _127.504	上と同じ結果で、従属モデルの下での最大対数尤度を与えている。
<:*/$ C 3	Ｃの要素の数から１を引いた従属モデルのパラメータ数を与えている。
+:(<:+/$ C)-B1 261.008	パラメータ数から最大対数尤度を引いた値を２倍していて、AICの値である。

AICD C 261.008	上と同じ結果で、分割表の従属モデルの下での情報量規準を与えている。