

鈴木義一郎著
J言語による統計分析
J807 対応版

第一章 データの縮約

鈴木義一郎先生の「J言語による統計解析」の数多くのスクリプトは、JAPLAのライブラリに残されていますが、ほとんどはJ600に対応したもので、最近のJ800以降では、部分的に正しく動作しない場合が多くなっています。

私も、この多数の統計解析関数スクリプトを利用して、J言語による時計解析のパッケージに応用させて頂いていますが、現在作成中のAndroid版の統計解析アプリ開発のために、勉強がてらスクリプトのために、デバッグしながら作業を進めています。

今回は第一章が終了したので、この内容を、Android 端末用の「J言語による統計分析・スクリプト」としてのデータベースの作成と並行して行っています。

J言語・データベース一覧 赤枠の [幾何平均] 選択	データベースの内容 青枠はスクリプト関数	データベースの内容(続き) 緑枠はJ言語式の解説
 <p>0005 /データ縮約/範囲と四分...</p> <p>0006 /データ縮約/3種類の平均...</p> <p>0007 /ータ縮約/3種類の平均値...</p> <p>0008 /データ縮約/3種類の平均...</p> <p>0009 /データ縮約/標準偏差/偏...</p> <p>0010 /データ縮約/標準偏差/標...</p>	 <p>大分類 データ縮約</p> <p>中分類 3種類の平均値</p> <p>小分類 幾何平均</p> <p>Script NB.幾何平均(Geometric mean)。 NB.相乗平均ともいう。つまり、yで与えたデータの累積の、yの累乗根を求めている。 meang=:#%:*/</p>	 <p>0007 /ータ縮約/3種類の平均値/幾何平均</p> <p>小分類 幾何平均</p> <p>Script NB.幾何平均(Geometric mean)。 NB.相乗平均ともいう。つまり、yで与えたデータの累積の、yの累乗根を求めている。 meang=:#%:*/</p> <p>NB.J言語解説 NB.これは「フォーク」で、(#%:*)y=(#y)%:(*/y)のように計算している。</p>

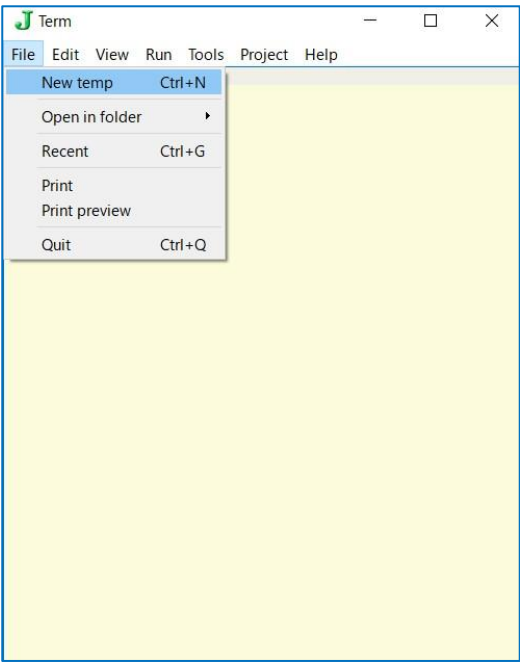
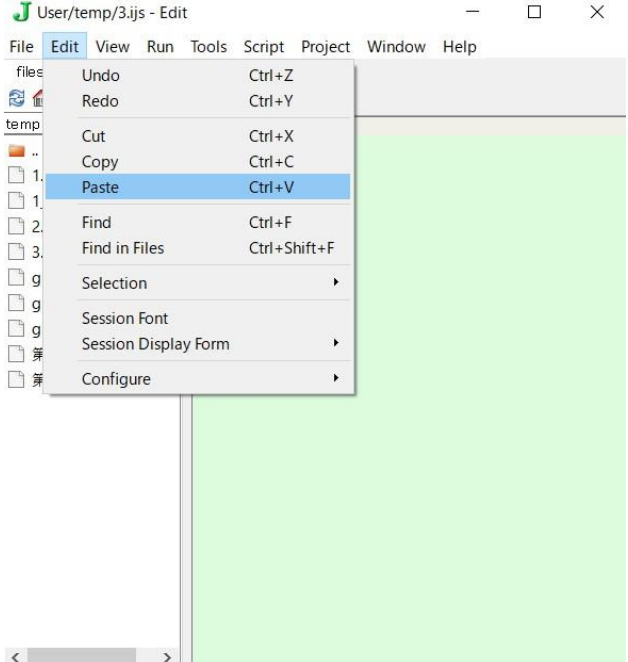
本データベースは、Android Apri [Memento DB](By Lucky Droid)を採用しています。内容は、第1章までです。Memento DBはGoogle PlayからDL出来ます。ご利用希望の方は、Webでsakura.japlaまでご連絡下さい。

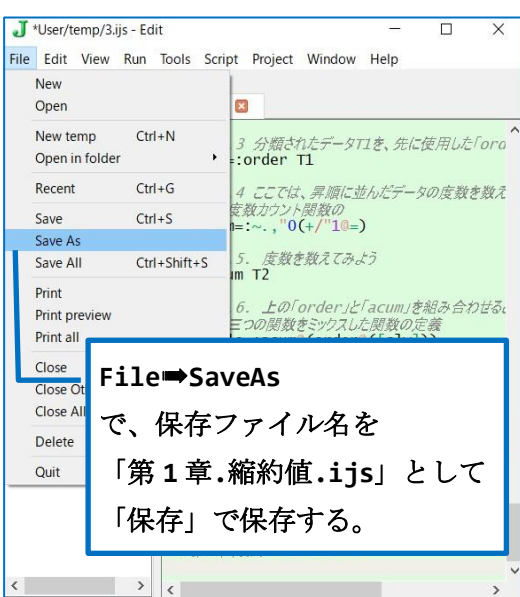
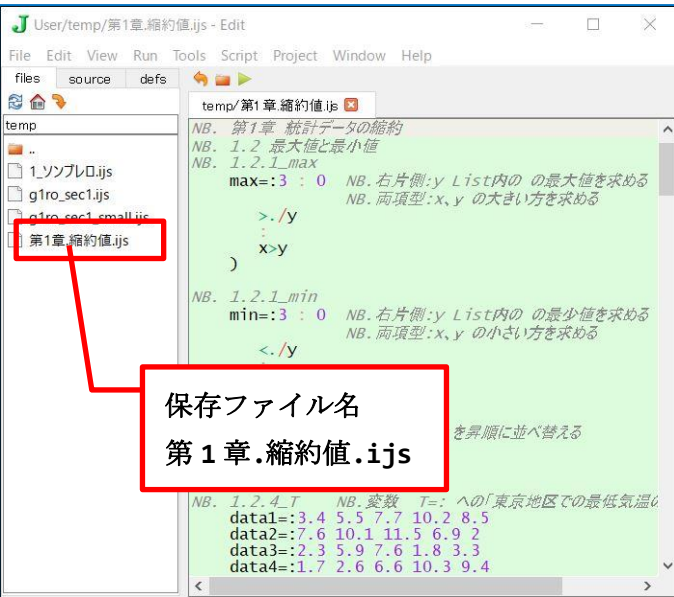
図1. データベースの例

J 言語による統計分析・スクリプト集

本スクリプト集の使い方。

本スクリプトは、小単位で実行できるように完結しています。各章のスクリプトをそのまま、J801 以上のバージョンの「ijs 編集画面」に貼付け、保存してください。

J の基本画面	J の編集画面
	
J を起動し、メニューの File ⇒ NewTemp で、J の編集画面を開く。	緑画面に J のスクリプト(ijs)を貼付ける。本書の章別のスクリプトをコピーし、本画面の Edit ⇒ Paste で、貼り付ける。

ijs ファイルを保存する。	保存された ijs ファイル
 <p>File⇒SaveAs で、保存ファイル名を「第1章.縮約値.ijs」として「保存」で保存する。</p>	 <p>保存ファイル名 第1章.縮約値.ijs</p>

1. 第一章.縮約値

1.1 スクリプト

行番号	第1章.統計データの性格と縮約値.ijs Script
1	NB.第1章.統計データの性格と縮約値.ijs
2	NB.japla
3	NB. 第1章 統計データの縮約
4	NB. 1.2 最大値と最小値
5	NB. 1.2.1_max
6	max=:3:0 NB.右片側:y List 内の の最大値を求める
7	NB.両項型:x、y の大きい方を求める
8	>./y
9	:
10	x>y
11)
12	
13	NB. 1.2.1_min
14	min=:3:0 NB.右片側:y List 内の の最少値を求める
15	NB.両項型:x、y の小さい方を求める
16	<./y
17	:
18	x<.y
19)
20	
21	NB. 1.2.3_order NB.List y を昇順に並べ替える
22	NB.並べ替え関数を定義する
23	order=:/~
24	
25	NB. 1.2.4_T NB.変数 T=: への「東京地区での最低気温のデータ」の代入
26	data1=:3.4 5.5 7.7 10.2 8.5
27	data2=:7.6 10.1 11.5 6.9 2
28	data3=:2.3 5.9 7.6 1.8 3.3
29	data4=:1.7 2.6 6.6 10.3 9.4
30	data5=:7.2 8.8 10.2 10.7 10.8
31	
32]Q=:data1,data2,data3,data4,data5
33	
34]T=:order Q
35	
36	NB. 1.3.1_rng NB.List y の Data Range(範囲)
37	NB.範囲関数を定義
38	rng=:>./-<./
39	
40	NB. 1.3.2.rngq NB.List y の四分位偏差を求める
41	NB.四分位偏差関数を定義
42	rngq=:3:0

```

43      r=:a,b,b+a=:<.-:b=:<.0.5+ -:#y
44      ([:{-}:+/"1(r{z}).(<:r){z=:/:~y
45      )
46
47      NB.四分位偏差の計算
48      ]rngq Q
49
50      NB. 1.4.1. 3 種類の平均値
51      NB.テストデータ
52      ]DATA=:>.i.5
53      NB. 3 つの平均値間の定義
54      mean=:+/%#          NB.算術平均関数定義
55      ]mean DATA
56
57      meang=:#%:*/        NB.幾何平均関数定義
58      ]meang DATA
59
60      NB.meanh=:mean&.%   NB.調和平均関数定義
61      meanh=:mean&.(%"_)
62      ]meanh DATA
63
64
65      NB. 1.5.1. 偏差平方和から標準偏差を求める
66      NB. 標準偏差等の関数を定義
67      ss=:+/&*:&(-mean)  NB.偏差平方和
68      ]ss DATA
69
70      sd=:+/%:&(ss%#)    NB.標準偏差
71      ]sd DATA
72
73      NB. 1.5.2. 平均と標準偏差を同時に求める
74      NB.平均と標準偏差が同時に計算する関数の定義
75      mesd=:mean,sd
76
77      NB.平均と標準偏差を同時に求める
78      ]mesd DATA
79
80      NB. 1.6.1. 相関比を求める
81      NB. 相関比関数を定義
82      corro=:3 : 0
83      m=:+/(k=:+/%#)&>y)*w=:n%/n=:#&>y
84      v=:(((+/"1*:(>y)-k)%n),:*:k-m)+/. * w
85      w=: 'within variance    ',8j2":{.v
86      b=: 'between variance   ',8j2":bv=:}.v
87      t=: 'total variance     ',8j2":v=:+ /v
88      w,b,t,: 'correration ratio ',8j2":bv%v

```

```

89     )
90
91 NB. 1.6.2. 国語の点数テストデータ 太郎 KT、次郎 KZ に代入
92     KT=:80 70 60 90 100
93     KZ=:50 60 70 80 90
94
95 NB. 1.6.3. 太郎、次郎の国語テスト得点の相関比を求める
96     ]corro KT;KZ
97
98 NB. 1.6.4. 国語の点数テストデータ 太郎 ST、次郎 SZ に代入
99     ST=:78 79 82 81 80
100    SZ=:70 69 71 68 72
101
102 NB. 1.6.5. 太郎、次郎の数学テスト得点の相関比を求める
103     ]corro ST;SZ
104
105 NB. 1.7. 移動平均と指数平滑化は気温などのデータを眺めるのに適している
106 NB. 1.7.1. 移動平均を
107     NB.移動平均を求める関数の定義
108     mav=:+∧%[
109
110 NB. 1.7.2. 東京都の最低気温の 5 項移動平均を求める
111     5 mav Q
112
113 NB. 1.7.3. 指数平滑化を求める関数
114     NB.指数平滑化の関数定義
115     smexp=:3 : 0
116         :
117         s=:{.Z=:y
118         while. 1<#Z do.
119             s=:s,(x*{s)+(1-x)*{Z=:}.Z
120         end.
121     )
122
123 NB.1.7.4. ウェイトの値を「0.6」として、最低気温に指数平滑化に適用
124     NB.指数平滑化を計算
125     ]0j2":0.6 smexp Q
126 NB.1.8. 相関係数を計算する
127 NB.1.8.1. テストデータの前期を x に、後期を y に代入する
128     x=:8 8 6 6 6 4 4 3 3 2
129     y=:9 5 9 7 5 7 4 7 4 3
130
131 NB.1.8.3. 前期、後期の点数を標準化する
132     NB.点数の標準化関数の定義
133     stand=(]-mean@)]%sd@]
134

```

135 NB.1.8.2. x と Y の相関係数を計算する関数を定義
136 cor=:+/@[(*&stand)]%#@]
137
138 NB.1.8.4. この関数を使用して前期、後期を標準化の値を観察してみる
139
140 stand x
141
142 stand y
143
144 NB.1.8.5. cor には、標準化の機能も含まれている。相関係数を計算
145]x cor y
146
147 NB.1.8.6. テストデータ 5 都市の緯度 N、最低気温(平年値)M を代入する
148 N=:43.05 38.25 35.68 34.68 31.57
149 M=:_8.9 _4.2 0.5 2.2 2.4
150
151 NB.1.8.7. 5 都市の緯度 N と最低気温 M との相関係数を計算する
152]N cor M
153
154 NB.1.9. 回帰直線：5 年間の年次と家電販売店の
155 NB.販売実績データとの回帰関係を考える
156 NB.1.9.1. 直線回帰分析のアルゴリズムで、回帰式を作る
157 NB.回帰式関数の定義
158 reg=:[%.1&(",0)@]
159
160 NB.1.9.2. この式に、年次と販売実績の直線回帰で切片と勾配を求めている
161 NB.
162]6 9 10 9 reg 3 4 5 6 7
163
164 NB.1.9.3. 5 都市の緯度 N と最低気温 M のデータを当てはめると
165]M reg N
166
167 NB.1.10. 分類と集計：統計分析の目的は、データだけでは判らない詳細を
168 NB.1.10. 分類や集計してわかりやすくするのが狙いである。
169 NB.1.10.1 データの分類関数
170 NB.データ分類関数の定義
171 cly=:>.@(]&]
172
173 NB.1.10.2 先に使用した、最低気温データを分類してみる
174]T1=:2 cly Q
175
176 NB.1.10.3 分類されたデータ T1 を、先に使用した「order」に当てはめてみよう
177]T2=:order T1
178
179 NB.1.10.4 ここでは、昇順に並んだデータの度数を数える関数を定義する
180 NB.度数カウント関数の

```
181     acum=:~.,"0(+/"1@=)
182
183 NB.1.10.5. 度数を数えてみよう
184     ]acum T2
185
186 NB.1.10.6. 上の「order」と「acum」を組み合わせると、同時に集計が可能だ
187     NB.三つの関数をミックスした関数の定義
188     table=:acum@(order@([cly]))
189
190 NB.1.10.7. 簡単なグラフを書く関数も作っておこう
191     NB.グラフ描画の関数を定義
192     grf=:3 : 0
193         NB.Original (.,~.z){&' *'z>:/>:i.>./z=:+/"1@=y
194         NB.2019/04/26_Modify_by_rToribe
195         (.,>:i.#z){&' *'z>:/>:i.>./z=:+/"1@=y
196     )
197
198 NB.1.10.8. グラフ作成の関数を実行してみよう
199     grf T2
200
201 NB.第 1 章終わり 2019-04-26
```


1.2 実行結果

行番号	第1章.統計データの性格と縮約値.ijs 実行結果
1	NB.第1章.統計データの性格と縮約値.ijs
2	NB. japla
3	NB. 第1章 統計データの縮約
4	NB. 1.2 最大値と最小値
5	NB. 1.2.1_max
6	max=:3 : 0 NB.右片側:y List内の の最大値を求める
7	NB.両項型:x、y の大きい方を求める
8	>./y
9	:
10	x>y
11)
12	
13	NB. 1.2.1_min
14	min=:3 : 0 NB.右片側:y List内の の最少値を求める
15	NB.両項型:x、y の小さい方を求める
16	<./y
17	:
18	x<.y
19)
20	
21	NB. 1.2.3_order NB.List y を昇順に並べ替える
22	NB.並べ替え関数を定義する
23	order=:/:~
24	
25	NB. 1.2.4_T NB.変数 T=: への「東京地区での最低気温のデータ」の代入
26	data1=:3.4 5.5 7.7 10.2 8.5
27	data2=:7.6 10.1 11.5 6.9 2
28	data3=:2.3 5.9 7.6 1.8 3.3
29	data4=:1.7 2.6 6.6 10.3 9.4
30	data5=:7.2 8.8 10.2 10.7 10.8
31	
32]Q=:data1,data2,data3,data4,data5
33	3.4 5.5 7.7 10.2 8.5 7.6 10.1 11.5 6.9 2 2.3 5.9 7.6 1.8 3.3 1.7 2.6 6.6 10.3 9.4 7.2 8.8 10.2 10.7 10.8
34	
35]T=:order Q
36	1.7 1.8 2 2.3 2.6 3.3 3.4 5.5 5.9 6.6 6.9 7.2 7.6 7.6 7.7 8.5 8.8 9.4 10.1 10.2 10.2 10.3 10.7 10.8 11.5
37	
38	NB. 1.3.1_rng NB.List y の Data Range(範囲)
39	NB.範囲関数を定義
40	rng=:>./-<./
41	
42	NB. 1.3.2.rngq NB.List y の四分位偏差を求める

```

43 NB.四分位偏差関数を定義
44 rngq=:3 : 0
45 r=:a,b,b+a=:<.-:b=:<.0.5+-: #y
46 ({:-{.)-:+"/"1(r{z),.(<:r){z=:/:~y
47 )
48
49 NB.四分位偏差の計算
50 ]rngq Q
51 6.8
52
53 NB. 1.4.1. 3種類の平均値
54 NB.テストデータ
55 ]DATA=:>:i.5
56 1 2 3 4 5
57 NB.3つの平均値間の定義
58 mean=:+/%# NB.算術平均関数定義
59 ]mean DATA
60 3
61
62 meang=:#%:* / NB.幾何平均関数定義
63 ]meang DATA
64 2.60517
65
66 NB.meanh=:mean&.% NB.調和平均関数定義
67 meanh=:mean&.(%"_)
68 ]meanh DATA
69 2.18978
70
71
72 NB. 1.5.1. 偏差平方和から標準偏差を求める
73 NB. 標準偏差等の関数を定義
74 ss=:+/%*:&(-mean) NB.偏差平方和
75 ]ss DATA
76 4 1 0 1 4
77
78 sd=:+/%:&(ss%#) NB.標準偏差
79 ]sd DATA
80 1.89443 2.44721 3 4.44721 5.89443
81
82 NB. 1.5.2. 平均と標準偏差を同時に求める
83 NB.平均と標準偏差が同時に計算する関数の定義
84 mesd=:mean, sd
85
86 NB.平均と標準偏差を同時に求める
87 ]mesd DATA
88 3 1.89443 2.44721 3 4.44721 5.89443

```

```

89
90 NB. 1.6.1. 相関比を求める
91 NB. 相関比関数を定義
92 corro=:3 : 0
93 m=:+/(k=(+/%#)&>y)*w=:n%/n=:#&>y
94 v=:(((+/"1*:(>y)-k)%n),*:k-m)+/ . * w
95 w=: 'within variance :',8j2":{.v
96 b=: 'between variance :',8j2":bv=:}.v
97 t=: 'total variance :',8j2":v=:+/%v
98 w,b,t,: 'correration ratio :',8j2":bv%v
99 )
100
101 NB. 1.6.2. 国語の点数テストデータ 太郎 KT、次郎 KZ に代入
102 KT=:80 70 60 90 100
103 KZ=:50 60 70 80 90
104
105 NB. 1.6.3. 太郎、次郎の国語テスト得点の相関比を求める
106 ]corro KT;KZ
107 within variance : 200.00
108 between variance : 25.00
109 total variance : 225.00
110 correration ratio : 0.11
111
112 NB. 1.6.4. 国語の点数テストデータ 太郎 ST、次郎 SZ に代入
113 ST=:78 79 82 81 80
114 SZ=:70 69 71 68 72
115
116 NB. 1.6.5. 太郎、次郎の数学テスト得点の相関比を求める
117 ]corro ST;SZ
118 within variance : 2.00
119 between variance : 25.00
120 total variance : 27.00
121 correration ratio : 0.93
122
123 NB. 1.7. 移動平均と指数平滑化は気温などのデータを眺めるのに適している
124 NB. 1.7.1. 移動平均を
125 NB. 移動平均を求める関数の定義
126 mav=:+/%#[
127
128 NB. 1.7.2. 東京都の最低気温の5項移動平均を求める
129 5 mav Q
130 7.06 7.9 8.82 9.58 8.92 7.62 6.56 5.72 4.94 3.92 4.18 4.06 3.4 3.2 4.9
131 6.12 7.22 8.46 9.18 9.26 9.54
132
133 NB. 1.7.3. 指数平滑化を求める関数
134 NB. 指数平滑化の関数定義

```

```

134     smexp=:3 : 0
135     :
136     s=: {.Z=:y
137     while. 1<#Z do.
138         s=:s,(x*{:s)+(1-x)*{:Z=:}.Z
139     end.
140 )
141
142 NB.1.7.4. ウエイトの値を「0.6」として、最低気温に指数平滑化に適用
143 NB.指数平滑化を計算
144 ]0j2":0.6 smexp Q
145 3.40 4.24 5.62 7.45 7.87 7.76 8.70 9.82 8.65 5.99 4.51 5.07 6.08 4.37
146 3.94 3.04
147 2.87 4.36 6.74 7.80 7.56 8.06 8.91 9.63 10.10
148 NB.1.8. 相関係数を計算する
149 NB.1.8.1. テストデータの前期を x に、後期を y に代入する
150
151 x=:8 8 6 6 6 4 4 3 3 2
152 y=:9 5 9 7 5 7 4 7 4 3
153
154 NB.1.8.3. 前期、後期の点数を標準化する
155 NB.点数の標準化関数の定義
156 stand=:([ -mean@])%sd@]
157
158 NB.1.8.2. x と Y の相関係数を計算する関数を定義
159 cor=:+/@([*&stand])%#@]
160
161 NB.1.8.4. この関数を使用して前期、後期を標準化の値を観察してみる
162
163 stand x
164 0.335245 0.335245 0.158322 0.158322 0.158322 _0.231684 _0.231684
165 _0.550592 _0.550592 _1.0174
166
167 stand y
168 0.301547 _0.188103 0.301547 0.136682 _0.188103 0.136682 _0.431736
169 0.136682 _0.431736 _0.759747
170
171 NB.1.8.5. cor には、標準化の機能も含まれている。相関係数を計算
172 ]x cor y
173 0.108141
174
175 NB.1.8.6. テストデータ 5 都市の緯度 N、最低気温(平年値)M を代入する
176 N=:43.05 38.25 35.68 34.68 31.57
177 M=:_8.9 _4.2 0.5 2.2 2.4
178
179 NB.1.8.7. 5 都市の緯度 N と最低気温 M との相関係数を計算する
180 ]N cor M
181 _0.00404653

```

```

177
178 NB.1.9.  回帰直線：5年間の年次と家電販売店の
179 NB.販売実績データとの回帰関係を考える
180 NB.1.9.1. 直線回帰分析のアルゴリズムで、回帰式を作る
181     NB.回帰式関数の定義
182     reg=:[%.1&(,"0)@]
183
184 NB.1.9.2. この式に、年次と販売実績の直線回帰で切片と勾配を求めている
185     NB.
186     ]6 6 9 10 9 reg 3 4 5 6 7
187 3 1
188
189 NB.1.9.3. 5都市の緯度 N と最低気温 M のデータを当てはめると
190     ]M reg N
191 38.2955 _1.08867
192
193 NB.1.10. 分類と集計：統計分析の目的は、データだけでは判らない詳細を
194 NB.1.10. 分類や集計してわかりやすくするのが狙いである。
195 NB.1.10.1 データの分類関数
196     NB.データ分類関数の定義
197     cly=:>.@(]&[)
198
199 NB.1.10.2 先に使用した、最低気温データを分類してみる
200     ]T1=:2 cly Q
201 4 6 8 11 9 8 11 12 7 2 3 6 8 2 4 2 3 7 11 10 8 9 11 11 11
202
203 NB.1.10.3 分類されたデータ T1 を、先に使用した「order」に当てはめてみよう
204     ]T2=:order T1
205 2 2 2 3 3 4 4 6 6 7 7 8 8 8 8 9 9 10 11 11 11 11 11 11 12
206
207 NB.1.10.4 ここでは、昇順に並んだデータの度数を数える関数を定義する
208     NB.度数カウント関数の
209     acum=:~., "0(+/"1@=)
210
211 NB.1.10.5. 度数を数えてみよう
212     ]acum T2
213 2 3
214 3 2
215 4 2
216 6 2
217 7 2
218 8 4
219 9 2
220 10 1
221 11 6
222 12 1

```

```

223
224 NB.1.10.6. 上の「order」と「acum」を組み合わせると、同時に集計が可能だ
225 NB.三つの関数をミックスした関数の定義
226 table=:acum@(order@([cly]))
227
228 NB.1.10.7. 簡単なグラフを書く関数も作っておこう
229 NB.グラフ描画の関数を定義
230 grf=:3 : 0
231 NB.Original (,~.z);{&' *'z>:/>:i.>./z=:+/"1@=y
232 NB.2019/04/26_Modify_by_rToribe
233 (,.>:i.#z);{&' *'z>:/>:i.>./z=:+/"1@=y
234 )
235
236 NB.1.10.8. グラフ作成の関数を実行してみよう
237 grf T2
238
239 1 ***
240 2 **
241 3 **
242 4 **
243 5 **
244 6 ****
245 7 **
246 8 *
247 9 *****
248 10 *
249
250
251 NB.第1章終わり 2019-04-26

```

作業内容	書籍解析	J-Script 化
【第1章統計データの縮約】	2018/12/20 完了	2019/05/19 完了
【第2章確率と確率分布】	2019/03/09 完了	進行中
【第3章推定・検定の考え方】	現在進行中	
【第4章情報量基準とモデル選択】	予定	
【第5章分散分析】	予定	
【第6章回帰分析】	予定	
【第7章主成分分析】	予定	

次回は、第2章.確率と確率分 2019/06/29 後樂園の予定