８日目：項目のチェック（２）

昨日は，平均値などの基礎統計量を計算する試行錯誤へご招待しましたが，今日は簡単にやってみます。そのためには，psych というパッケージが必要となりますので，R を起動したら，まずこれをとってきてください。やり方は，web で検索してみてください。

以下の説明は，psych パッケージの導入が済み，いつもの練習用のファイルを読み込んでいるというところから始めます。

せっかくパッケージを導入してもらったのですが，先に psych パッケージを使わない，summary というコマンドを使ってみます。まずは以下のように入力し，実行してみてください。

```
summary(x)
```

ずらずらと，基礎統計量が出てきます。何が算出されているかをチェックすると，最小値(Min.)，第１四分位 (1st Qu.)，中央値 (Median)，平均値 (Mean)，第３四分位(3rd Qu.)，最大値(Max.)，そして，「NA」がある場合は，その数(NA's)です。

summaryは，(x)と指定しても警告は出てこないし，最小値，最大値も変数ごとにやってくれるし，「NA」も自動的に省いてくれるし，その数も出してくれる…と，いいことが多いのですが，問題は標準偏差を計算してくれないところ…。

そこで，psychパッケージに登場してもらいます。

パッケージを使うには，基本的には，最初にそれを呼び出す必要があります。Rを起動しただけでは，パッケージは読み込んでくれません。

```
library(psych)
```

と，まずは入力します。これを実行しても，Rコンソールには何の変化もありません。次に，

```
describe(x)
```

と入力して実行します。すると，欲しかった数値が！　欠損値を含むb2は，nが19になっているように省いて計算されています。

解説するまでもないでしょうが，左から変数名，列番号(var)，ケース数(n)，平均値(mean)，標準偏差(sd)，中央値(median)，トリムのある平均値(trimmed)，中央値絶対偏差：median absolute deviation (mad)，最小値(min)，最大値(max)，レンジ(range)，歪度(skew)，尖度(kurtosis)，標準誤差(se)です。ちなみに，「中央値絶対偏差，トリムとはなんぞや？」と思う人は，統計の本を読むなり，ググるなりしてください。

```
> library(psych)
> describe(x)
     vars  n    mean   sd median trimmed  mad  min  max range  skew kurtosis   se
no      1 20 1010.50 5.92 1010.5 1010.50 7.41 1001 1020    19  0.00    -1.38 1.32
sex     2 20    1.70 0.47    2.0    1.75 0.00    1    2     1 -0.81    -1.41 0.11
age     3 20   19.35 0.59   19.0   19.25 0.00   19   21     2  1.30     0.58 0.13
b1      4 20    2.95 1.28    3.0    2.94 1.48    1    5     4  0.23    -1.26 0.29
b2      5 19    3.95 0.85    4.0    3.94 1.48    3    5     2  0.09    -1.68 0.19
b3      6 20    4.45 0.69    5.0    4.56 0.00    3    5     2 -0.76    -0.72 0.15
b4      7 20    4.00 0.73    4.0    4.00 0.74    3    5     2  0.00    -1.19 0.16
b5      8 20    3.25 0.97    3.0    3.19 1.48    2    5     3  0.19    -1.12 0.22
```

さて，今回は，ここでちょっとRの基本操作のお勉強をしましょう。

まずひとつめに，パッケージの読み込みについてです。先にパッケージを使うには，最初にlibrary(psych)で呼び出す作業が必要なことをお伝えしました。Mac版のRには，もう一つ別のやり方があります。

まず，メニューバーの「パッケージとデータ」をクリックし，「パッケージマネージャ」を選択します。すると，すでに自分のPCに取ってきてあるパッケージのリストが出てきます。もしこのリストにpsychがなければ，まだとってきていないということです。

この一覧で，使いたいパッケージの先頭にある□をクリックすると，自動的にロードしてくれ，使える状態になります。library(psych)を入力するのとどちらが簡単かというところは，判断の分かれるところでしょうが…

もう一つは，Ｒの命令の中身を見たり，ヘルプを見たりする方法です。Rコンソールの方で良いので，以下だけ（変数指定をしない）を入力して実行してください。

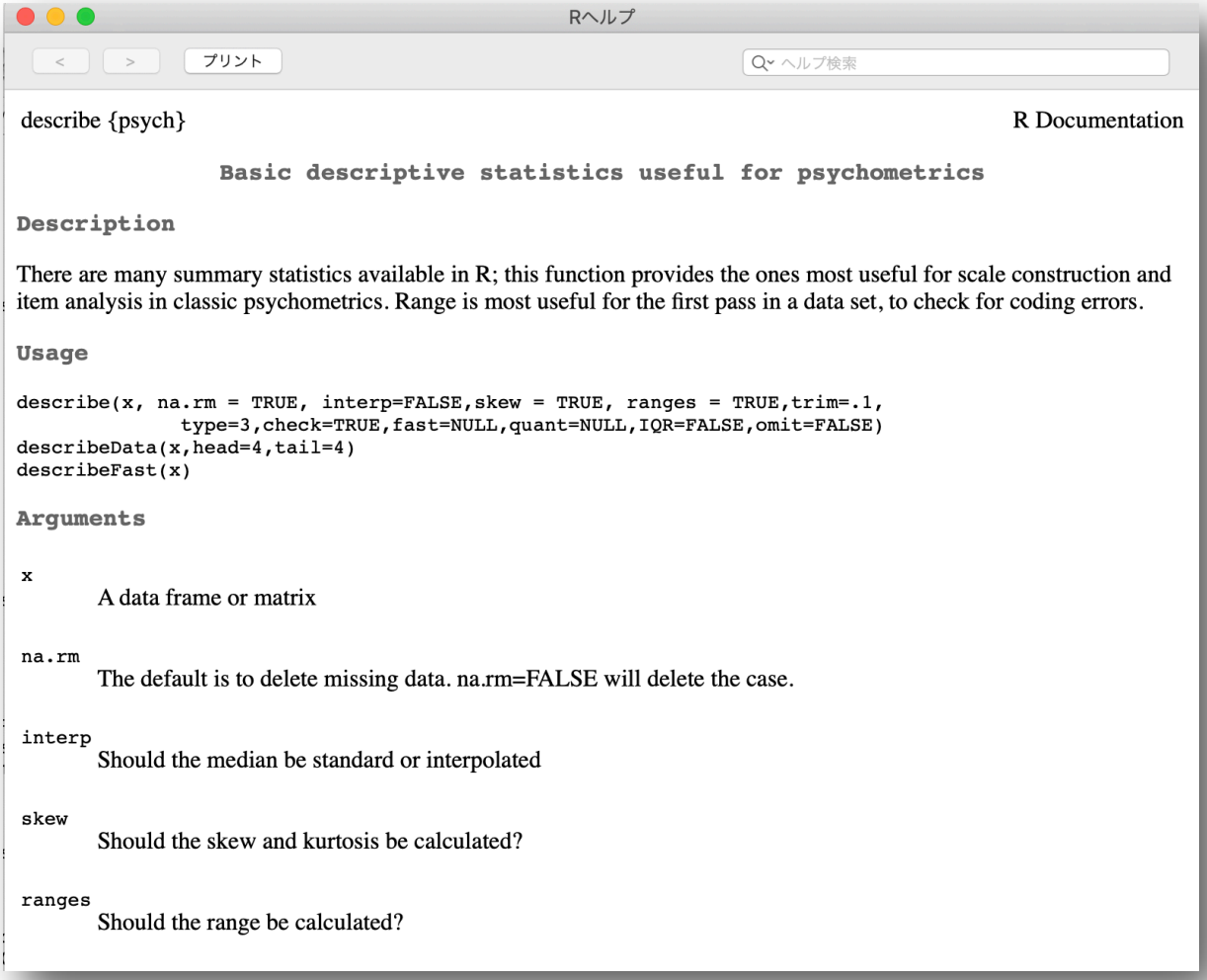
```
describe
```

すると，一見でコンピュータのプログラムらしきものが表示されると思います。その通りで，これが describe の中身（プログラム）なのです。このようにすれば，中身を見ることができます（できないものも結構あります）。

次には，以下のように入力し，実行してください。

```
?describe
```

こちらは新しいウインドが開きます。これはRのヘルプ画面です。英語ですが，嫌がらずに眺めてみてください。まず Description で，概要の説明がされています。



Usageは，コマンドの詳しい説明です。そこには，describe(x, na.rm = TRUE, interp=FALSE, skew = TRUE, ranges = TRUE, trim=.1, type=3, check=TRUE, fast=NULL, quant=NULL, IQR=FALSE, omit=FALSE)と記載されています。na.rm = TRUE以下はデフォルトの設定であり，何も指定しなければこの通りに実行されます。試しに，describe(x)の結果と，describe(x, na.rm = TRUE, interp=FALSE, skew = TRUE, ranges = TRUE, trim=.1, type=3, check=TRUE, fast=NULL, quant=NULL, IQR=FALSE, omit=FALSE)の結果を比べてみてください。同じ出力結果になります。

デフォルトを変更する場合は，変更する部分のみを書き換えて加えればよいです。たとえばskewだけを変更したければ，describe(x, skew = FALSE)とすればOKです。

さらに下の方には，Examples もあります。このヘルプにはいろんな情報がありますので，積極的に見るようにしておくと，いろいろな発見があると思います。

ちなみに，ヘルプを参照するには?describe以外にも方法はあります。ひとつは，Ｒコンソールや Ｒエディタで describe という文字列を選択しておいて，「コントロールキー　＋　H」というショートカットで見る方法。これに似ていますが，トラックパッド

があるなら，describeの部分を二本指でタップし，「現在位置の関数のヘルプを表示」を選ぶ（「コントロールキー　＋　H」というショートカットもOK）やり方もあります。

では，話を戻して，次に男女別に基礎統計量を求めることをやってみます。コマンドは describeBy です。まずはヘルプ探して見ることで，これの使い方を試行錯誤してみてください。

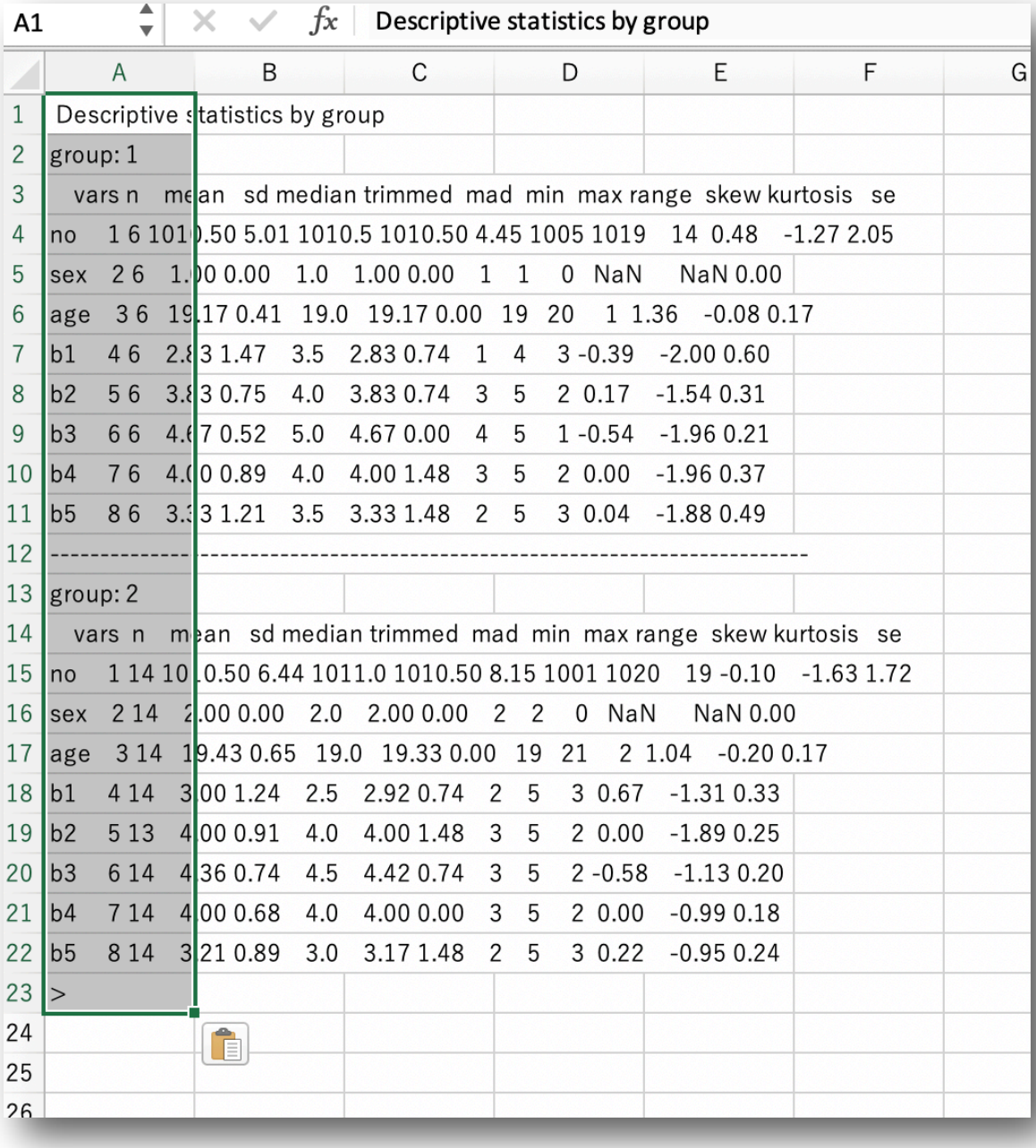
Examples もありますが，簡単な設定は以下のようでしょう。これで性別に計算をしてくれます。

```
describeBy(x, x$sex)
```

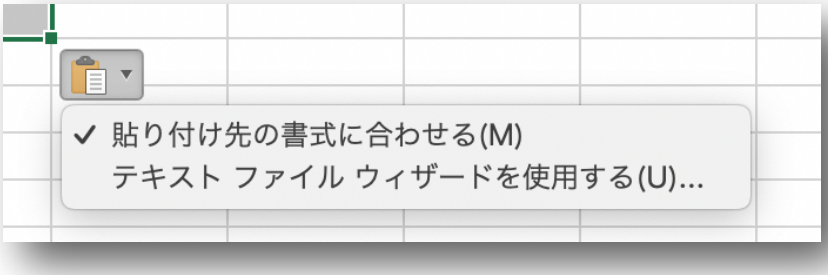
さて今日の最後に，この R での計算結果をエクセルに移すことをやってみます。R の出力のままでは論文の表としては使えません。何とかして右のような表に仕上げる必要があるでしょう。エクセルに結果を移すのはファイルを介してもできますが，簡単なのはコピペです。

		人数	平均値	標準偏差	最小値	最大値
男性	age	6	19.17	0.41	19	20
	b1	6	2.83	1.47	1	4
	b2	6	3.83	0.75	3	5
	b3	6	4.67	0.52	4	5
	b4	6	4	0.89	3	5
	b5	6	3.33	1.21	2	5
女性	age	14	19.43	0.65	19	21
	b1	14	3	1.24	2	5
	b2	13	4	0.91	3	5
	b3	14	4.36	0.74	3	5
	b4	14	4	0.68	3	5
	b5	14	3.21	0.89	2	5

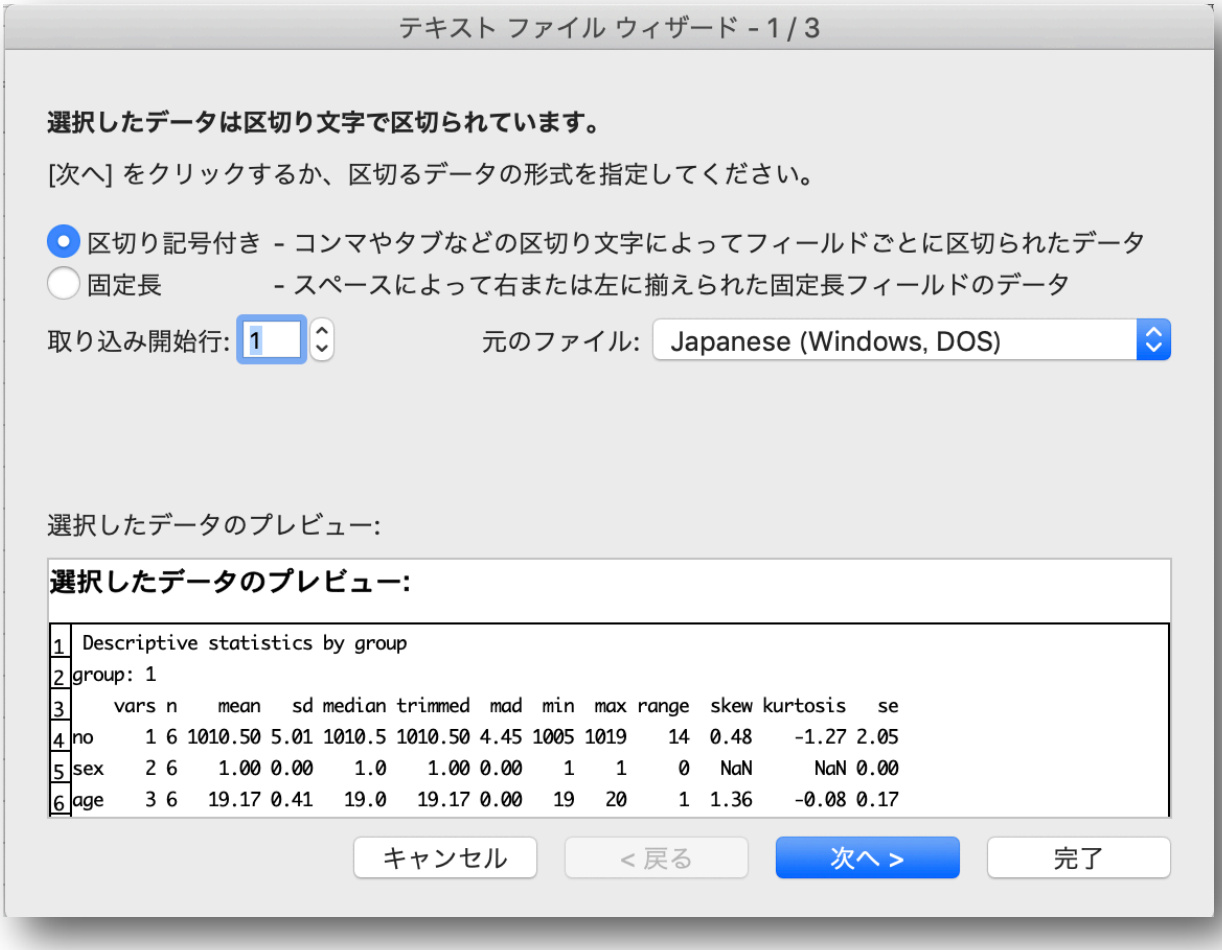
まず，R コンソールの describeBy の結果部分をコピーします。そしてエクセルのシートにペーストします。すると以下の図のようになると思います。



次にこの図の中にもあるクリップボードのようなアイコンをクリックします。すると以下のようなメニューが出てきます。

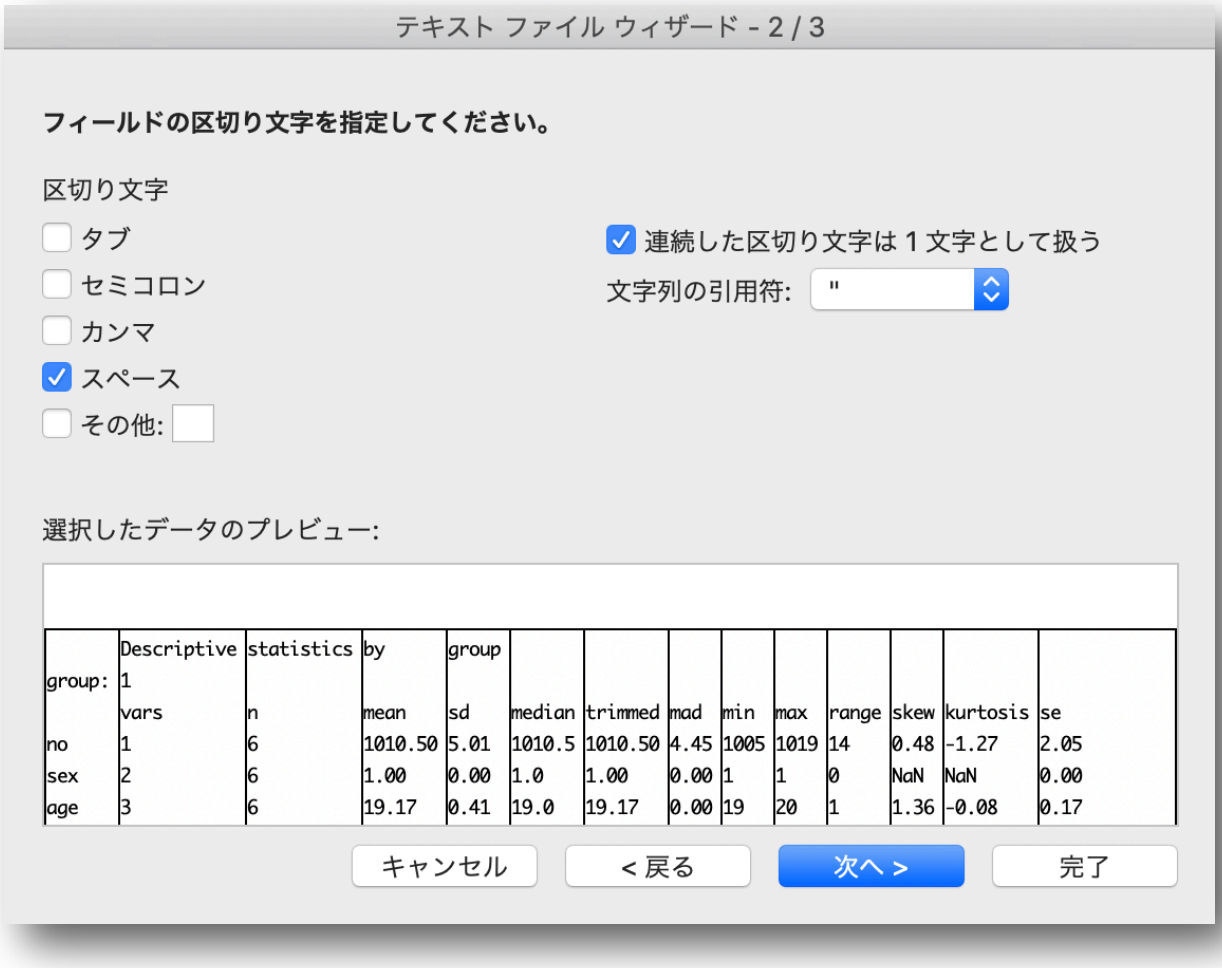


このメニューのうち下側の「テキスト ファイル ウィザードを使用する」を選択します。すると，次の図のようなウインドが開きます。



この画面ではさわるところはありません。R からコピーしてきたデータは，スペース（空白）によって整形されています。しかし，それは「固定長」ではないので，「データのファイル形式」は「区切り記号付き」のままで OK です。

「次へ」をクリックします。

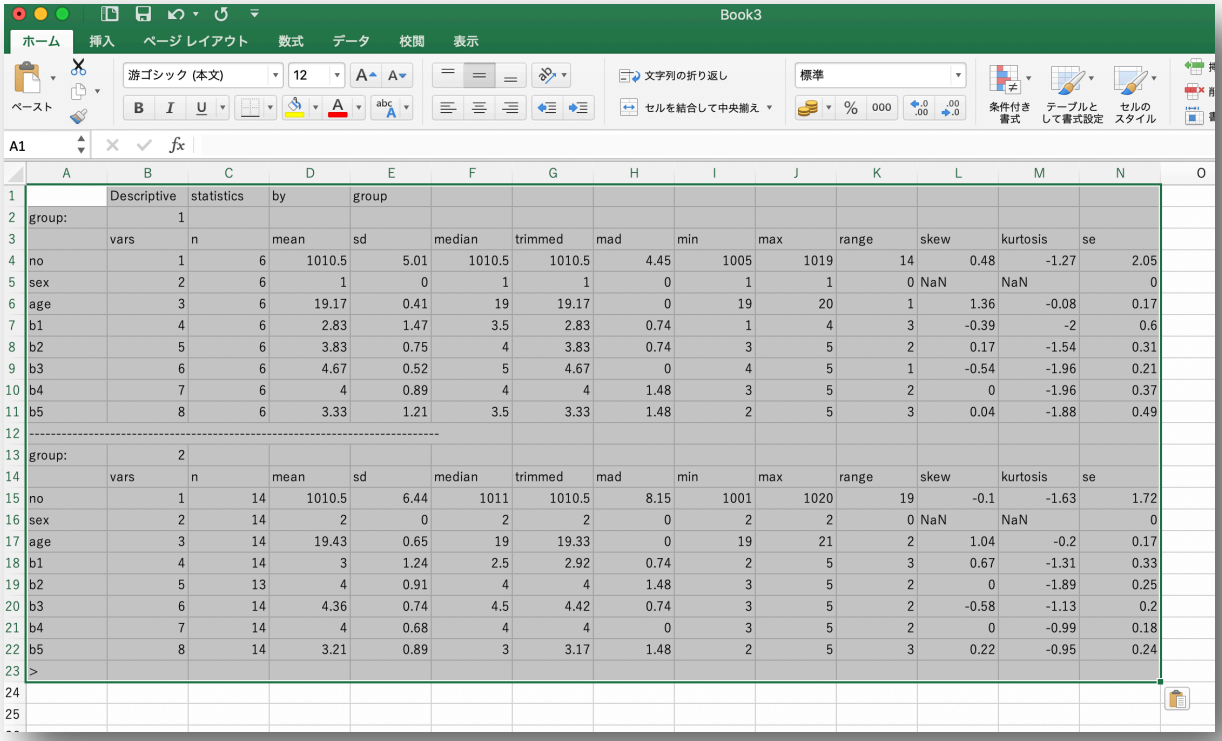


このウィザードは，結構うまく区切りをつけてくれます。「区切り文字」で「スペース」を指定しなくても，たいていは「スペース」にチェックが入っていると思います。

また「データのプレビュー」には，区切りの部分に縦線が入っています。このまま続けると，この線の部分でデータを区切ってくれます。

これ以上特に触る部分もないので，「完了」をクリックします。

すると以下のように数値がセルに分けられていると思います。



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		Descriptive	statistics	by	group										
2	group:	1													
3		vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se	
4	no	1	6	1010.5	5.01	1010.5	1010.5	4.45	1005	1019	14	0.48	-1.27	2.05	
5	sex	2	6	1	0	1	1	0	1	1	0	NaN	NaN	0	
6	age	3	6	19.17	0.41	19	19.17	0	19	20	1	1.36	-0.08	0.17	
7	b1	4	6	2.83	1.47	3.5	2.83	0.74	1	4	3	-0.39	-2	0.6	
8	b2	5	6	3.83	0.75	4	3.83	0.74	3	5	2	0.17	-1.54	0.31	
9	b3	6	6	4.67	0.52	5	4.67	0	4	5	1	-0.54	-1.96	0.21	
10	b4	7	6	4	0.89	4	4	1.48	3	5	2	0	-1.96	0.37	
11	b5	8	6	3.33	1.21	3.5	3.33	1.48	2	5	3	0.04	-1.88	0.49	
12	----------														
13	group:	2													
14		vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se	
15	no	1	14	1010.5	6.44	1011	1010.5	8.15	1001	1020	19	-0.1	-1.63	1.72	
16	sex	2	14	2	0	2	2	0	2	2	0	NaN	NaN	0	
17	age	3	14	19.43	0.65	19	19.33	0	19	21	2	1.04	-0.2	0.17	
18	b1	4	14	3	1.24	2.5	2.92	0.74	2	5	3	0.67	-1.31	0.33	
19	b2	5	13	4	0.91	4	4	1.48	3	5	2	0	-1.89	0.25	
20	b3	6	14	4.36	0.74	4.5	4.42	0.74	3	5	2	-0.58	-1.13	0.2	
21	b4	7	14	4	0.68	4	4	0	3	5	2	0	-0.99	0.18	
22	b5	8	14	3.21	0.89	3	3.17	1.48	2	5	3	0.22	-0.95	0.24	
23	>														
24															
25															

ここまできたら，後はエクセルで整形するだけですから，先のような表に仕上げるのはすぐでしょう。Descriptive statistics by group という部分もセルに分割されますが，これは仕方ないとしましょう。

なお，ヘルプを参照すると，以下でも同じような結果を得ることができることがわかります。このあたりは好き好きかも。

```
describeBy(x, x$sex, mat=TRUE)
```

これで８日目は終了です。明日は度数分布表を作ってみます。