

看護・介護研究を始めよう
—統計の基礎と質問紙調査のやりかた(第1回)—

平均値と標準偏差

医療法人社団ときわ練馬在宅クリニック
カ石辰也



TOKIWA GROUP
医療法人社団ときわ

はじめにお断りしておきますが・・・

- 力石は統計学の専門家ではありません。
- 必要に迫られてほぼ独学で統計学をほんの少しだけかじりました。
- 数学は苦手です。
- 内容に間違いがある場合はご容赦ください。
- 質問や誤りの指摘を歓迎しますが、対応できないこともたくさんあると思います。
- エクセル・スプレッドシート・EZRのすべてを理解しているわけではありません。

このセミナーの対象者と目的

- 対象は統計は全く初めてという方です。
- 目的は難しい数式や高価な統計ソフトを用いずに、基本的な統計解析と看護・介護研究ができるようになることです。



TOKIWA GROUP

医療法人社団ときわ

商品のお客様満足度95%！！（当社アンケート調査による）

- よく見るこんなCMをみて皆さんはどう思いますか？
 - とっても良い商品だからすぐかわなくちゃ。
 - そんなの嘘に決まってる。
 - 数字は本当かもしれないけど、満足した人だけがアンケートに答えたんじゃないの？？
- アンケートの回答率が示されなければ全く信用できません。
- 統計学を勉強するとこんなCMに騙されにくくなります。



TOKIWA GROUP

医療法人社団ときわ

では早速はじめてみましょう。
今日は平均値と標準偏差について学びます。

- おかあさん、僕のテストの点は平均点より上だったよ！
 - あなたならどんな返事・判断をしますか？
 - あらそう、よかったわね。
 - 平均点より上なら安心かしら。

 - しか～し！そもそも平均値ってなんでしょう？



TOKIWA GROUP

医療法人社団ときわ

平均値とは集団を代表する数値のひとつです

- 平均値（Average または Mean）は集団の中心を見る数値の一つです。
 - 10人のクラスのテストの平均値の計算は、全員の点数を足し算して10で割ります。
 - 100人のクラスなら全員の点数を足して100で割ります。
 - ここまではそんなに難しくありませんね？
- 平均値の弱点
 - 集団の中で大きくはずれた値があると、平均値はそれに引っ張られてしまいます。

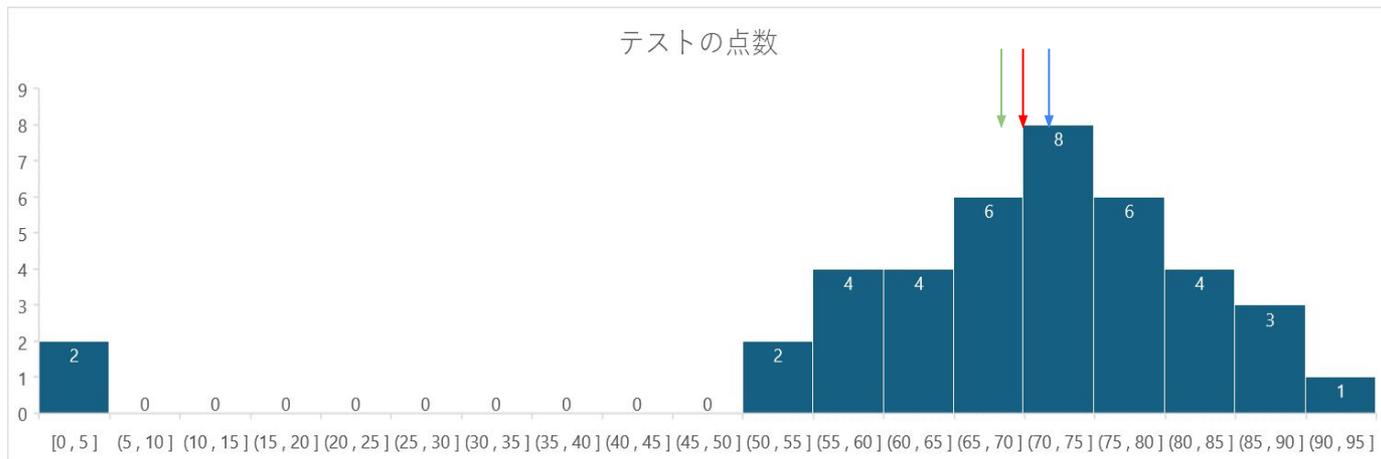


TOKIWA GROUP

医療法人社団ときわ

平均値ははずれ値の影響を受けやすい

- 平均値は集団の中心を見る数値です。
 - 例えば、40人のクラスで平均点が68点(緑の矢印)のテストがあったとします。
 - 正孝君は70点(赤の矢印)だったので、平均点より少し上だと思いました。
 - しかし、このクラスには0点の生徒が2人いました。
 - この2人を除いた38人の平均点は72点(青の矢印)になります。
 - 実質的には正孝君は平均よりやや下だということになります。



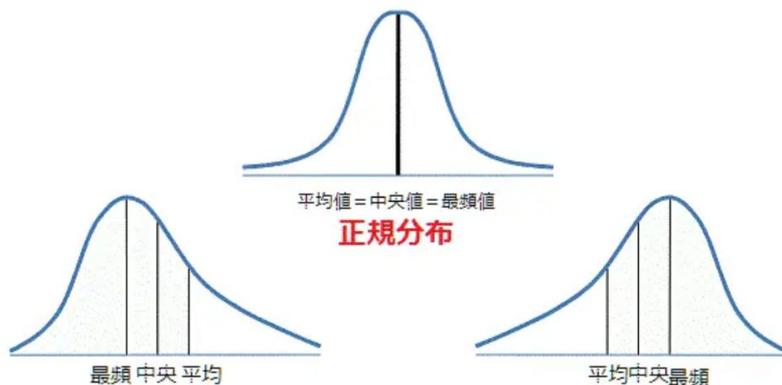
中央値ははずれ値の影響を受けにくい

- 中央値も集団の中心を見る数値です。
 - データを値の小さいほうから順にならべたときにちょうど半分にデータを分ける値です。
 - データの数が奇数の場合は、ちょうど中央にくるデータの値、データの数が偶数の場合は、前半の最大値と後半の最小値の真ん中の値になります。
 - 例えば「2, 3, 5, 40, 100」のデータでは、真ん中の「5」が中央値です。
 - 同じデータで、平均値は「30」になります。
 - 中央値ははずれ値の影響を受けにくいのが特徴です。



集団を代表する値には最頻値もあります。

- 最頻値は最も多く出現する値です。
 - 「5, 2, 3, 5, 7, 9, 5」のデータでは「2」「3」「7」「9」はそれぞれ1個しかありませんが、「5」は3つあります。よって「5, 2, 3, 5, 7, 9, 5」の最頻値は「5」です。
 - 最頻値ははずれ値の影響を受けにくく、中央値と異なり、分布も考慮されるので、分布に偏りがある集団で用いられます。
- 富士山のような形の正規分布では平均値・中央値・最頻値は一致します。



<https://avilen.co.jp/personal/knowledge-article/normal-distribution/> から引用

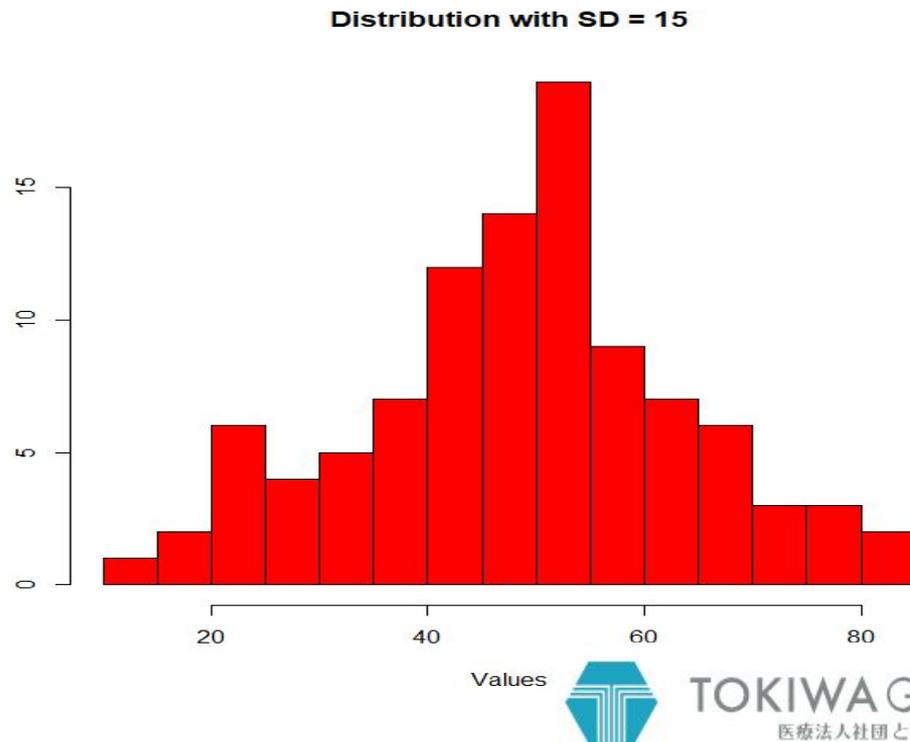
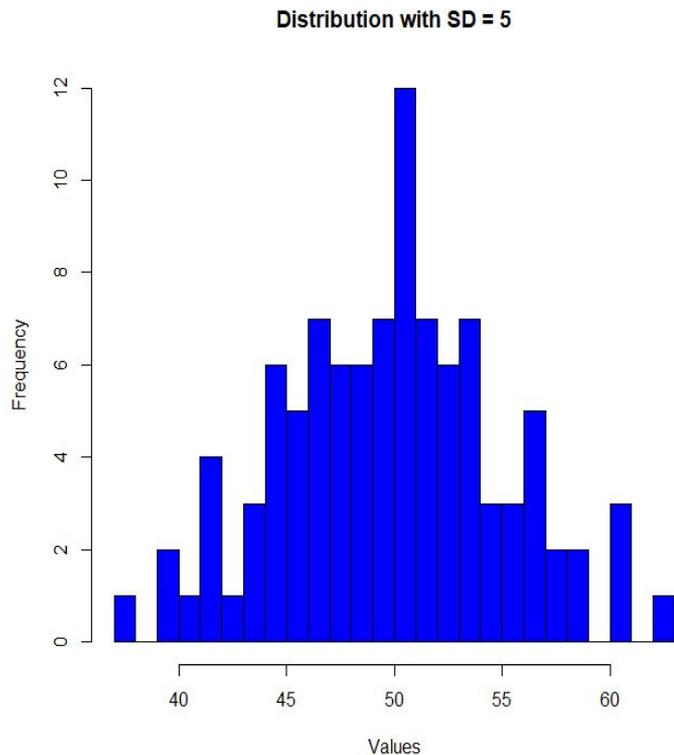


集団の中央をみるだけで全体を推測するには無理がある

- 集団全体を把握するにはばらつきの程度を見る必要があります。
- そこで標準偏差（Standard Deviation）の登場です。
- 標準偏差をどのように計算するかは、興味のある方は調べてみて下さい。
- とりあえず、計算式がわからなくても以下のことを理解できればOKです。
- 標準偏差とは、集団のばらつきを表す指標です。
 - 標準偏差が小さい＝全体のバラツキが小さい＝測定値の分布が平均値の周りに集まっている
 - 標準偏差が大きい＝全体のばらつきが大きい＝平均値から遠く離れている測定値が多くある



同じ平均値でも標準偏差が大きくなると裾野が広がります。



Values



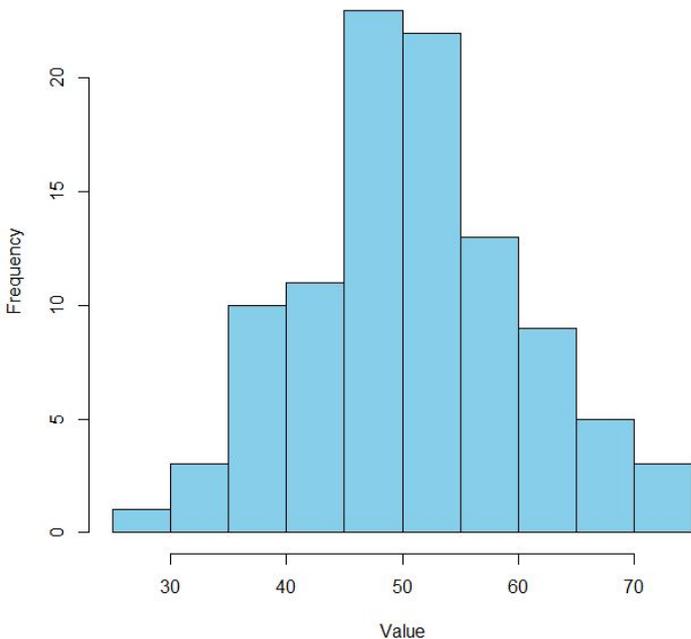
TOKIWA GROUP

医療法人社団ときわ

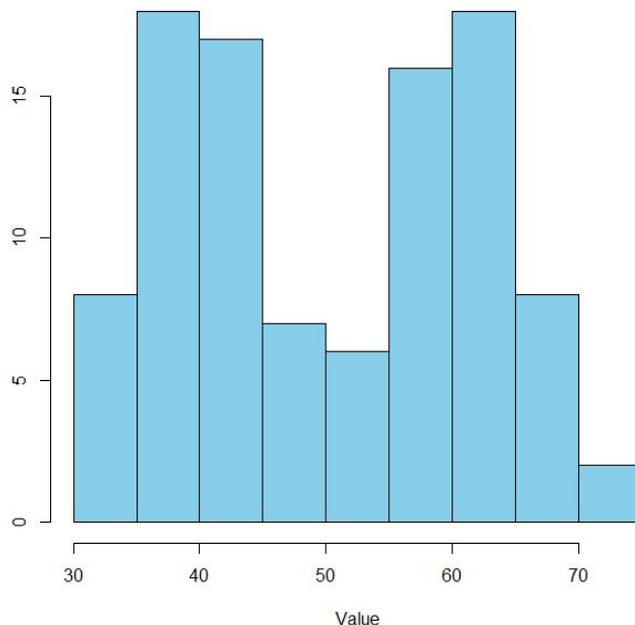
平均値と標準偏差だけで全体を推測するにも限界がある

- 全体を把握するにはヒストグラムが有用です。どちらも平均50の分布です。

Normal Distribution (Mean=50, SD=10)



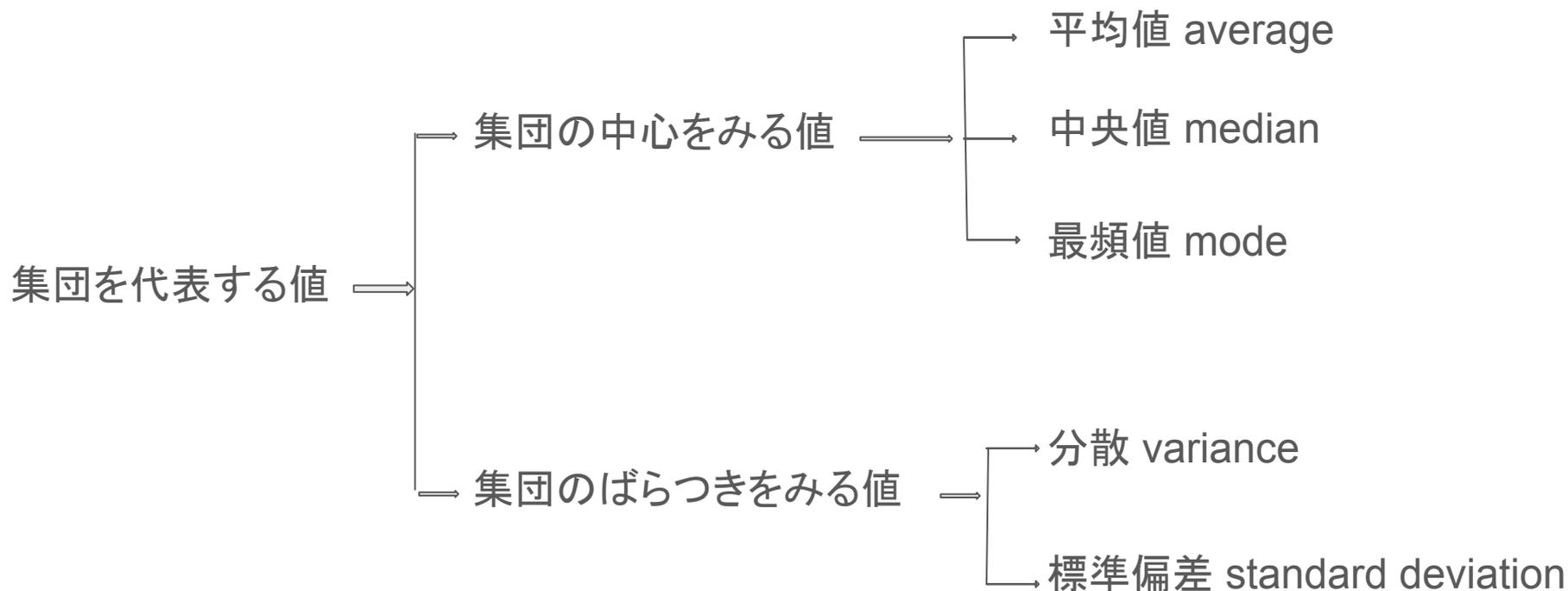
Bimodal Distribution (Mean=50)



左は正規分布、
右は二峰性分布
と呼ばれています。



集団を代表する数値：



ここまでのまとめ

- 平均値は集団の中心を見る値です。
- 標準偏差は集団のばらつきを見る値です。
- 平均値と標準偏差だけでは集団を把握できないことがあります。
- 集団を代表する値には、平均値のほかに中央値・最頻値があります。

- それでは実際にデータを入力して平均値などを求める練習をしましょう。



最初に Google の account を作っておきましょう

- スプレッドシートやフォームが使えるようになります。
 - 無料です。多分 Google や CIAあたりにデータを抜かれて解析されていますがタダなので仕方ありません。
 - 看護・介護研究で使う主なアプリケーションは以下の通りです。

Google	Microsoft
スプレッドシート	⇔ エクセル
ドキュメント	⇔ ワード
プレゼンテーション	⇔ パワーポイント
 - フォーム(質問紙調査をする時に非常に強力なツールです)



次に無料統計ソフト EZR をダウンロードします。

- 無料で公開されている統計ソフト「R」を日本語で動かせるようにしたプログラムです。
 - <https://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmed.html> からDLしてください。
 - ↑ google で EZR と検索するとトップで表示されます。
 - 作成した自治医科大学血液内科の神田先生に感謝。
 - SPSSに匹敵する強力な統計機能があります。
 - ファイル名やデータに日本語(全角文字)が入っていると動作しないので注意が必要です。
 - EZRを使った統計解析を含む論文は、それを理由に雑誌に採用を拒否されることはありません。

これで準備が整いました。



TOKIWA GROUP

医療法人社団ときわ

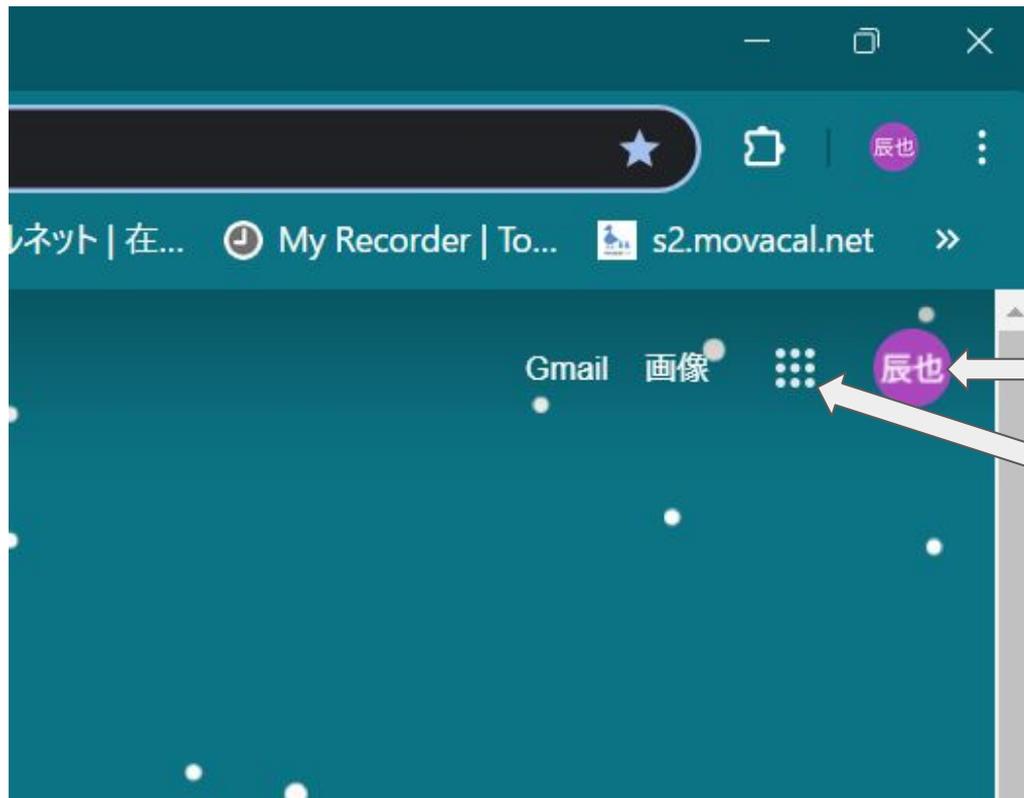
データの入力や編集はスプレッドシート、解析はEZRで

- データの入力や編集はスプレッドシート(または Excel)がやりやすい
 - EZRでもデータ入力や編集もできますが、上記のほうがはるかにやりやすいです。
 - スプレッドシートやエクセルのデータをクリップボードにコピーして、それを EZRで読み込むのが一番楽です。
 - 他にも色々なやり方がありますが、まずはこの方法をお勧めします。
 - EZRはファイル名やデータに日本語(全角文字)が入っていると正しく動作しないことがあるので注意が必要です。

では早速データ入力・解析を行ってみましょう。



ご自分のアカウントで google に login し、
新しいスプレッドシートを開きます。



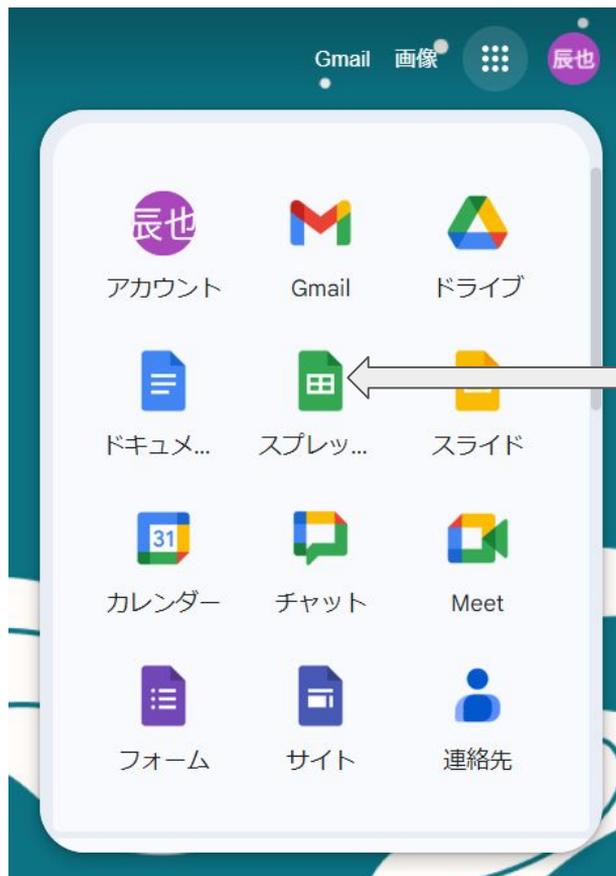
自分のアカウントで
ログイン

ここをクリック



TOKIWA GROUP
医療法人社団ときわ

スプレッドシートをクリックして開きます



ここをクリック



TOKIWA GROUP
医療法人社団ときわ

空白のスプレッドシートを開きます

新しいスプレッドシートを作成



空白のスプレッドシート



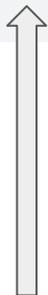
ガントチャート
作成者: Smartsheet



ToDo



Google Finance 投資ト...



ここをクリック



空白のスプレッドシートを開きます

無題のスプレッドシート ☆

ファイル 編集 表示 挿入 表示形式 データ ツール 拡張機能 ヘルプ

100% | ¥ % .0 .00 123 | デフォ...

A1 | fx

	A	B	C	D	E
1	ユーザーのスマートチップを挿入するには「@」の後に名前を入力します				
2					
3					
4					
5					
6					
7					
8					
9					

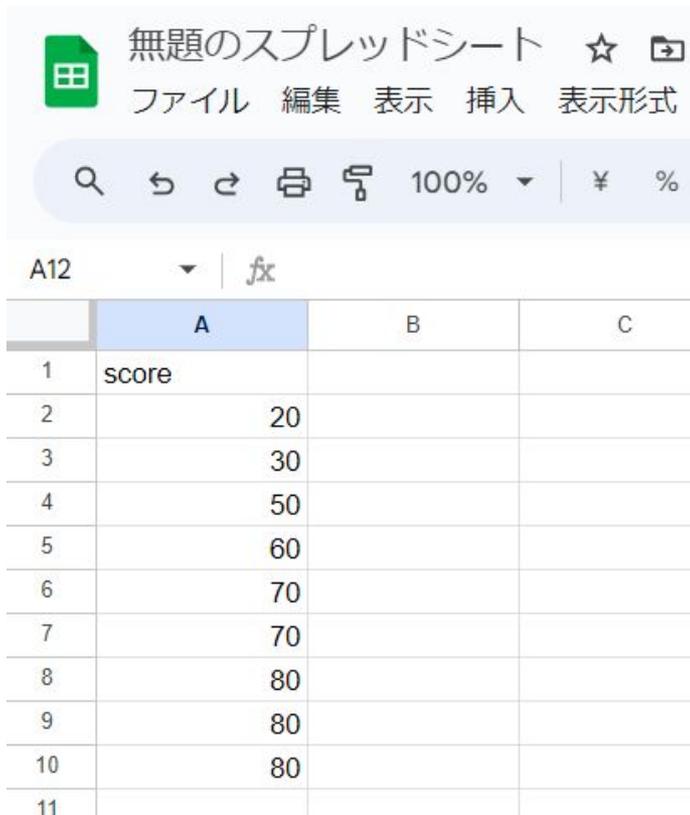
A1 のセルに文字が入っている場合は削除します。



TOKIWA GROUP

医療法人社団ときわ

9人の生徒のテストの点数を入力してみましょう



無題のスプレッドシート ☆

ファイル 編集 表示 挿入 表示形式

100% | ¥ %

	A	B	C
1	score		
2	20		
3	30		
4	50		
5	60		
6	70		
7	70		
8	80		
9	80		
10	80		
11			

- 9人の点数は、それぞれ20、30、50、60、70、70、80、80、80点としてみましょう。
こんな感じになります。
- 1行目には項目名を入れます。
今回はscore としました。
- スプレッドシートだけならA1に「点数」と入れても問題ありませんが、EZRを使うことを念頭に入れて、すべて半角文字で入力しましょう。もちろんtennsuu でもOKです。



9人の生徒のテストの点数の平均値をスプレッドシートで求めてみましょう(1)

The screenshot shows the Google Sheets interface with the 'Insert' menu open. The '関数' (Functions) option is selected, and the 'AVERAGE' function is highlighted in the list. The spreadsheet data is as follows:

	A	B
1	score	
2		20
3		30
4		50
5		60
6		70
7		70
8		80
9		80
10		80
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		

- 今回はA11のセルに平均値を計算します。
- A11のセルをアクティブにして、挿入→関数→すべて→averageを選びます。



9人の生徒のテストの点数の平均値をスプレッドシートで求めてみましょう(2)

無題のスプレッドシート ☆
ファイル 編集 表示 挿入 表示形式

100%

A11 | fx =AVERAGE()

	A	B	C
1	score		
2	20		
3	30		
4	50		
5	60		
6	70		
7	70		
8	80		
9	80		
10	80		
11	=AVERAGE()		
12			

無題のスプレッドシート ☆
ファイル 編集 表示 挿入 表示形式

100%

A11 | fx =AVERAGE(A2:A10)

	A	B	C
1	score		
2	20		
3	30		
4	50		
5	60		
6	70		
7	70		
8	80		
9	80		
10	80		
11	=AVERAGE(A2:A10)		
12			

- このようになり
ます。
- 平均値を求めたいのはA2からA10までなのでこの範囲を選択します。
- enter キーを押
します。



9人の生徒のテストの点数の平均値をスプレッドシートで求めてみましょう(3)

無題のスプレッドシート ☆ 📄
ファイル 編集 表示 挿入 表示形式

🔍 ↶ ↷ 🖨 📄 100% ▾ | ¥ %

A11 | fx =AVERAGE(A2:A10)

	A	B	C
1	score		
2	20		
3	30		
4	50		
5	60		
6	70		
7	70		
8	80		
9	80		
10	80		
11	60		
12			

- このようになります。
- 平均点は60点です。
- A11セルをアクティブにすると、
`fx=average(A2:A10)`
と表示されています。A2からA10までの平均を表示しているという意味です。



同様に9人の生徒のテストの点数の中央値 (median)・最頻値 (mode) を求めてみましょう (4)

	A	B	C
1	score		
2	20		
3	30		
4	50		
5	60		
6	70		
7	70		
8	80		
9	80		
10	80		
11	60		
12	70		
13	80		
14			
15			

- 平均値をA11に表示させたので、中央値をその下のA12に表示させてみます。
- 最頻値をA13に表示させます。
- この様になっていれば正しく入力できています。
- 中央値は70、最頻値は80です。
- 正しい値が表示されない場合は、それぞれの関数が正しい範囲を参照しているかどうかを確認して下さい。



同じ計算をEZRにやらせてみましょう(1)

	A	B	C
1	score		
2	20		
3	30		
4	50		
5	60		
6	70		
7	70		
8	80		
9	80		
10	80		
11	60		
12	70		
13	80		
14			

- まず、ワークシートのデータをクリップボードにコピーします。
- 項目を表す1行目も忘れずにコピー範囲に入れて下さい。
- 先ほど計算した平均値や中央値を含めてコピーすると間違えてしまいます。
- 選択範囲に注意して下さい。



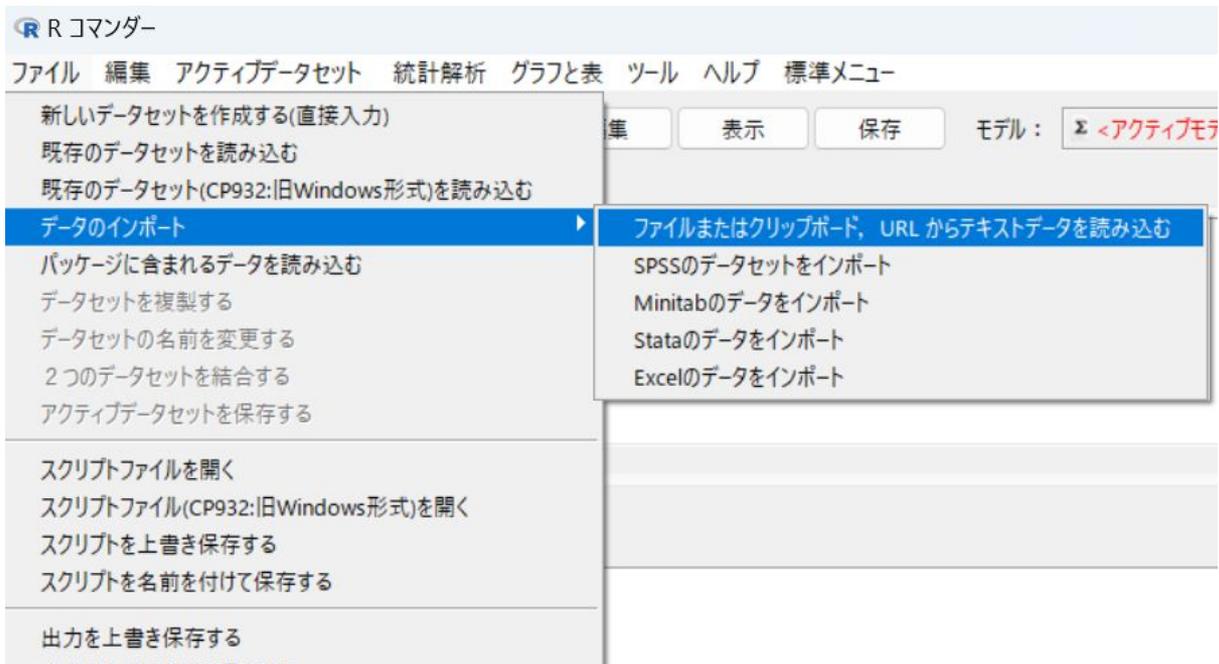
同じ計算をEZRにやらせてみましょう(2)



- EZRを起動するとこんな感じの2つのウィンドウが開きます。
- 主に使うのはRコマンドーの方です。



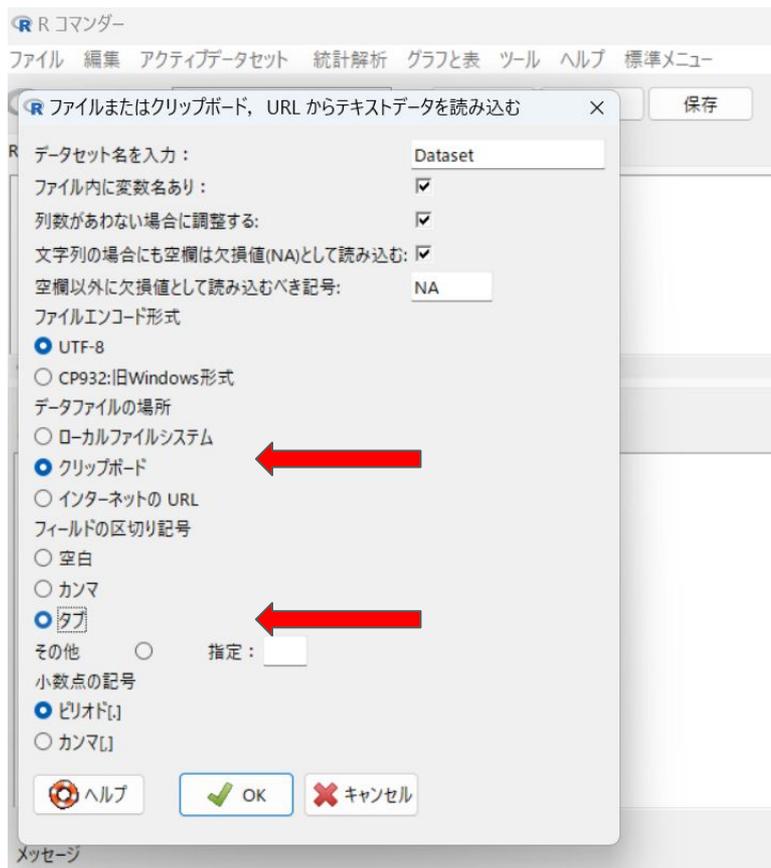
同じ計算をEZRにやらせてみましょう(3)



- ファイル→データのインポート→ファイルまたはクリップボード、URLからテキストデータを読み込むの順に進みます。



同じ計算をEZRにやらせてみましょう(4)



- データファイルの場所をクリップボード、フィールドの区切り記号をタブにします。(エクセル・スプレッドシートのデータの場合)
- その他はいじる必要はありません。
- OKボタンを押します。
- 何も変わらないように見えて不安になりますが、大丈夫です。
- よく見ると一番下に
メモ: データセット Dataset には 9行、1 列あります。
と表示されています。



同じ計算をEZRにやらせてみましょう(5)



R コマンドー

ファイル 編集 アクティブデータセット 統計解析 グラフと表 ツール ヘルプ 標準メニュー

データセット: Dataset 編集 表示 保存 モデル: Σ <アクティブモデルなし>

Rスクリプト Rマークダウン

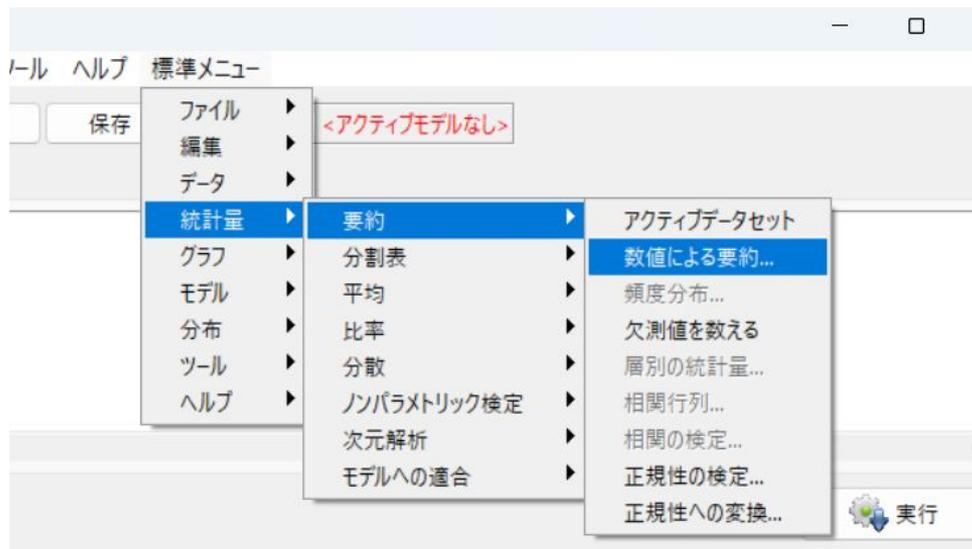
```
#####テキストファイルからデータを読み込む#####  
Dataset <- read.table("clipboard", header=TRUE, sep="¥t", na.strings=c("", "NA"), dec=".", fill=TRUE,  
  quote="¥", comment.char="", strip.white=TRUE, fileEncoding="UTF-8")
```

score	
1	20
2	30
3	50
4	60
5	70
6	70
7	80
8	80
9	80

表示ボタンをクリックするとインポートしたデータが表示されます。



同じ計算をEZRにやらせてみましょう(6)



表示メニュー



統計量



要約



数値による要約
の順に進みます。



TOKIWA GROUP

医療法人社団ときわ

同じ計算をEZRにやらせてみましょう(7)

The screenshot shows the EZR R Commander interface. The main window displays the R script editor with the following code:

```
library(abind, pos=18)
library(e1071, pos=19)
```

The output window shows the execution results:

```
> library(abind, pos=18)
> library(e1071, pos=19)
```

A dialog box titled "R 数値による要約" (R Summary by Numerical Values) is open. It has tabs for "データ" (Data) and "統計量" (Statistics). Under "変数 (1つ以上選択)" (Variables (select 1 or more)), the variable "score" is selected. There is a "層別して要約..." (Summary by layer...) button. At the bottom of the dialog, there are buttons for "ヘルプ" (Help), "リセット" (Reset), "OK", "キャンセル" (Cancel), and "適用" (Apply).

表示メニュー

↓
統計量

↓
要約

↓
数値による要約
の順に進みます。

このままOKを押します。



同じ計算をEZRにやらせてみましょう(8)



The screenshot shows the R Commander window with the following content:

メニュー: ファイル 編集 アクティブデータセット 統計解析 グラフと表 ツール ヘルプ 標準メニュー

データセット: Dataset [編集] [表示] [保存] モデル: <アクティブモデルなし>

Rスクリプト Rマークダウン

```
library(abind, pos=18)
library(e1071, pos=19)
numSummary(Dataset[, "score", drop=FALSE], statistics=c("mean", "sd", "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))
```

実行

```
> library(abind, pos=18)
> library(e1071, pos=19)
> numSummary(Dataset[, "score", drop=FALSE], statistics=c("mean", "sd", "IQR", "quantiles"), quantiles=c(0,.25,.5,.75,1))
mean      sd IQR 0% 25% 50% 75% 100% n
 60 22.36088 30 20 50 70 80 80 9
```

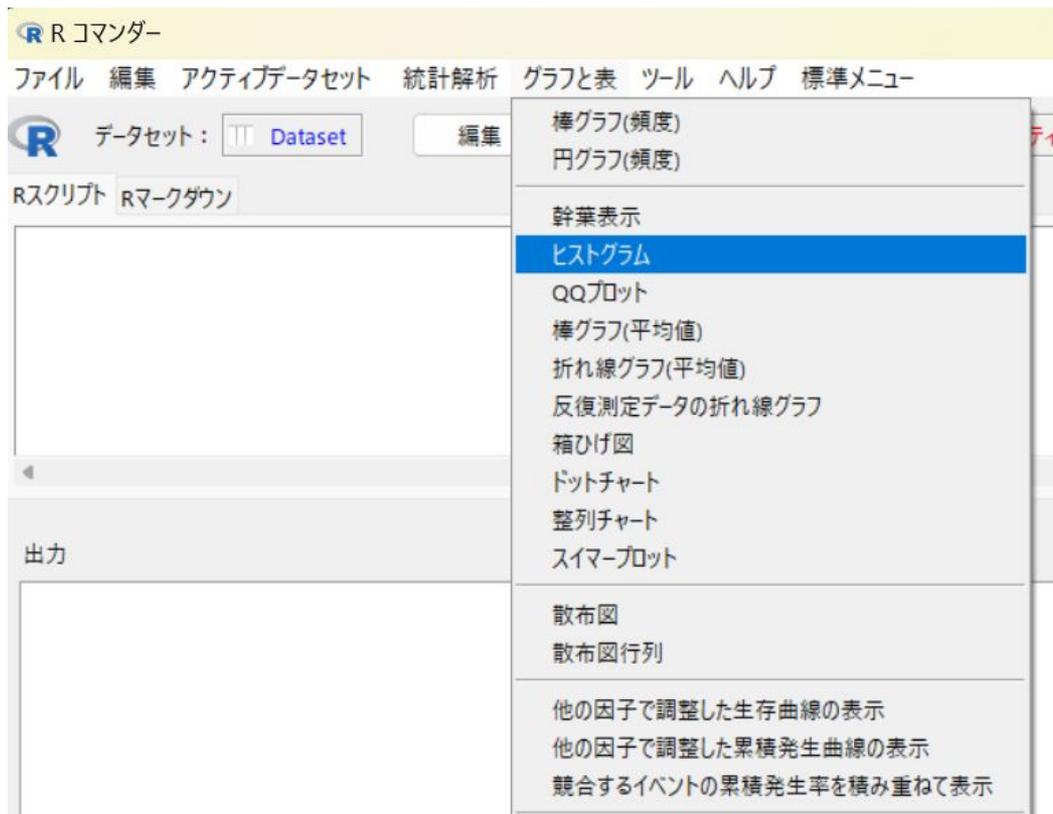
結果は少し見にくい感じがしますが、タダなので我慢しましょう。

平均値60、
標準偏差22.36、
中央値70、
標本数9
と表示されています。

最頻値を表示させる機能はありません。



EZRにヒストグラムを描かせてみましょう(1)



Rコマンドーで
グラフと表



ヒストグラム
を選びます。



TOKIWA GROUP

医療法人社団ときわ

EZRにヒストグラムを描かせてみましょう(2)



変数(1つ選択)
でscore を選びます。

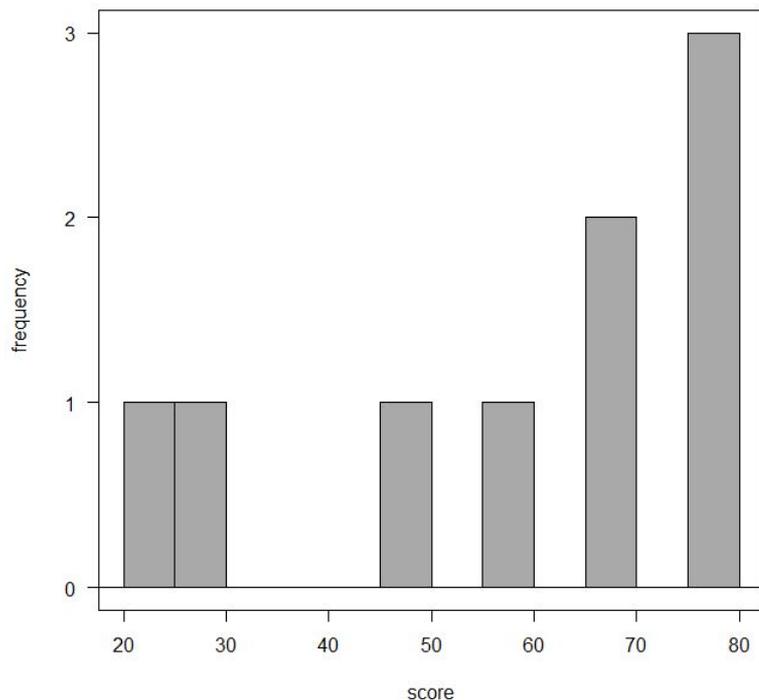
群別する変数(0-1つ選択)
でscore が青くなっている場合は、
ctrl キーを押しながらクリックして選
択を解除します。

区間の数(群別しない場合)に
<auto>を消して10と入力します。

これでOKボタンを押します。



EZRにヒストグラムを描かせてみましょう(3)



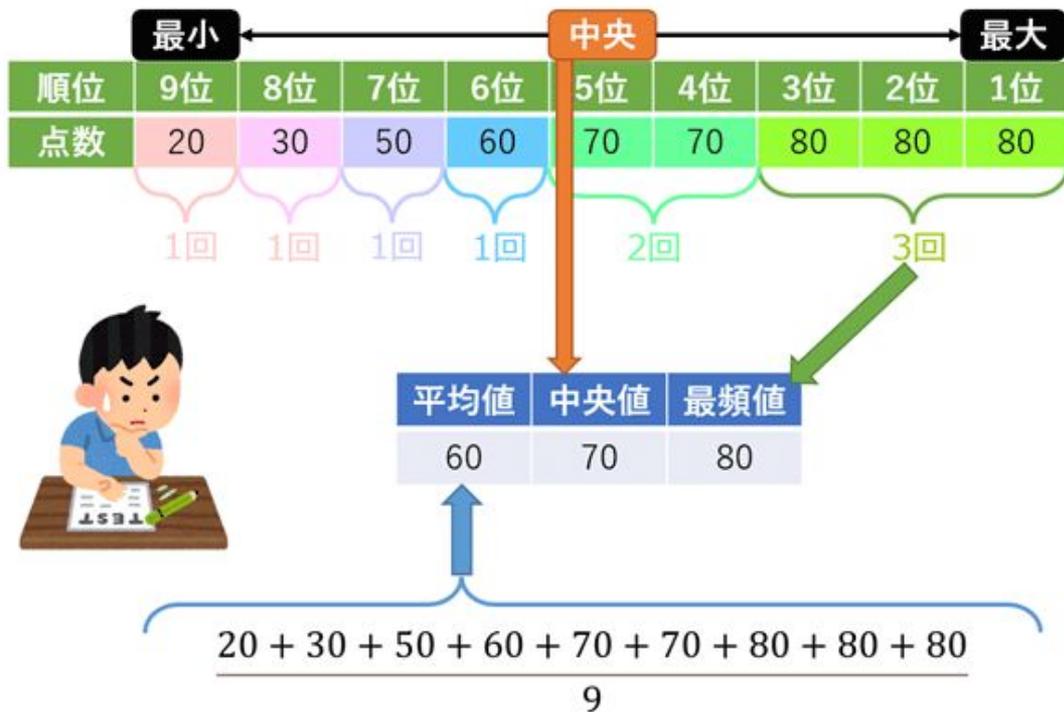
score (得点)の分布が示されました。

これで80点の生徒が3人いて、80点が最頻値だということが視覚的にわかります。

ヒストグラムはスプレッドシートでも描けます。
興味のある方は挑戦してみてください。



今回の勉強を図にまとめたのがこちらです。



<https://atmarkit.itmedia.co.jp/ait/articles/2109/15/news033.html> から引用



TOKIWA GROUP
医療法人社団ときわ

EZRの終了



右上のxをクリックします。

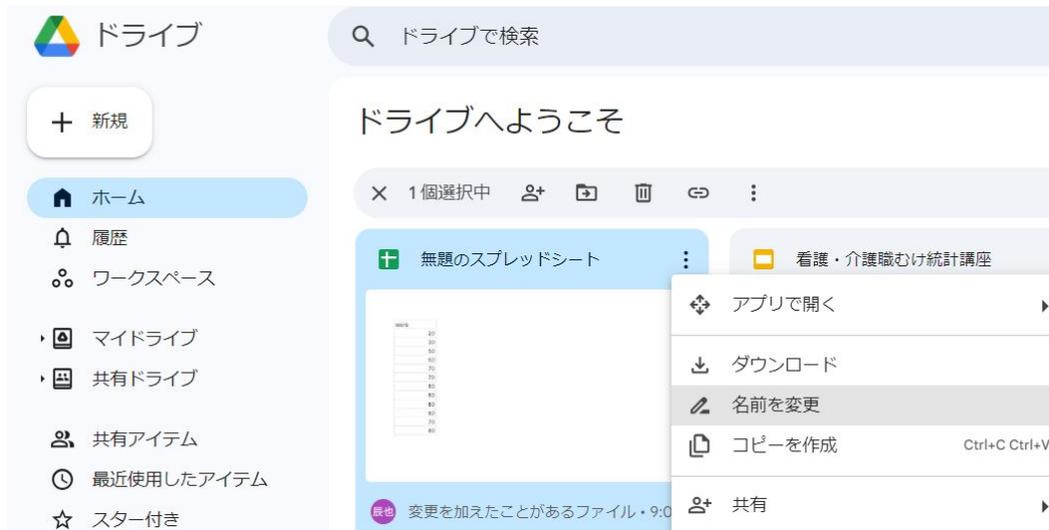
OKボタンを押し、その後のメッセージは全ていいえで終了します。



スプレッドシートの終了



	A	B	C	D
1	score			
2	20			
3	30			
4	50			
5	60			
6	70			
7	70			
8	80			
9	80			
10	80			
11	60			
12	70			
13	80			



スプレッドシートは常に自動的に保存されているので、保存して終了という概念がありません。

chrome のタブを閉じて終了します。

ドライブのホーム画面で必要に応じて名前を変更できます。



今回はこれでおしまいです。

- 第1回の統計学講座はいかがでしたか？
 - 質問・間違いの指摘・感想を歓迎します。
 - 次回はアンケート調査の作り方・やり方を勉強する予定です。
-
- この勉強会が統計は全く初めてという方に少しでも役に立てば幸いです。



TOKIWA GROUP

医療法人社団ときわ