

<特性値>

1. 準備

今回の実習では、10月15日に実習した都道府県別球場数のファイルを用いる。欠席などによって、その部分の実習が終わっていないものは、講義用HPから、「都道府県別球場数(ヒストグラムあり)」のファイルをダウンロードする。

📖 手順 (前回の実習が終わっていない者)

- ① 検索エンジンで、「河田研究室」と入力し検索すると、「河田研究室」のページにジャンプする。(ここまでの手順は、<http://www2.tokuyama-u.ac.jp/kawada> とアドレスを直接入力してもよい)
- ② 「統計学基礎」をクリックし、「第10回 10月29日(月)」の配布資料にある、「都道府県別球場数(ヒストグラムあり)」をクリックし、自分の使いやすい場所に保存する。

2. 相対度数・累積相対度数の導出

【課題 9】 前回実習した度数分布表に、相対度数・累積相対度数をつけ加えよう。

📖 手順

- ① まず、H17セルに度数の合計を求めよう。=SUM(H5:H17) で求めることができる。
- ② I列に相対度数を求める。I5セルに、=H5/H17 と入力し、**F4**キーを押すと、=H5/\$H\$17 となる。これを I6:I16 セルにコピーする。I17 セルは、H17 セルをコピーする。I6:I17セルを範囲指定し、**%**ボタンを押し、小数点以下第1位まで表示するようにする。
- ③ J列に累積相対度数を求める。J5セルには =I5 と入力し、J6セルに =J5+I6 と入力する。これをJ7:J16セルにコピーすれば累積相対度数が求まる。

3. 統計関数を利用した特性値の算出

Excel で算術平均、メディアンなどのデータ集団の特性値を求める場合、統計関数の利用が便利である。Excel における関数を一般的にあらわすと以下ようになる。

=関数名 (引数 1, 引数 2, 引数 3, …, 引数 k)

引数の数は関数によって異なり、0個のものもあるが、その場合でも()は必要である。

例1：算術平均を求める関数 AVERAGE は、引数に算術平均を計算するデータの範囲をとる。あるいは、③のように数値を直接書き込むこともできる。

- | | | | |
|---|----------------------------------|---|---------|
| ① | =AVERAGE(A1:A20) | — | 引数の数は1個 |
| ② | =AVERAGE(gym) | — | 引数の数は1個 |
| ③ | =AVERAGE(5, 3, 6, 8, 9, 5, 8, 9) | — | 引数の数は8個 |

AVERAGE() と同様の引数をとる統計関数に、MAX(), MIN(), COUNT(), MODE(), MEDIAN(), STDEVP(), VARP(), MAX(), MIN()などがある。

【課題 10】都道府県別球場数のデータについて、データ数(COUNT)、算術平均(AVERAGE)、メディアン(MEDIAN)、モード(MODE)、分散(VARP)、標準偏差(STDEVP)を求めてみよう。

📖 手順

- ① Sheet1のあいている部分(ここでは、F20セルから始める)に、特性値を求めていく。最初に右図のF列のように、求める特性値を記入する。
- ② データ数ならCOUNT関数(関数名は小文字でも良い)を用いて求める。データ範囲は、**ballpark** と名前をつけているので、それを用いる。以下同様に最小値まで求める。

例2：関数 QUARTILE(引数 1, 引数 2) は四分位数を求める関数である。引数 1 は範囲、引数 2 は 0 から 4 までの数値をとり、以下に示すようなデータを戻り値として与える。

- | | |
|---|---|
| 0 | データの最小値 |
| 1 | 下位 4 分の 1 (25%) に相当するデータ (第 1 四分位数 : q ₁) |
| 2 | データの中央値 (50%) (第 2 四分位数 : q ₂) |
| 3 | 上位 4 分の 1 (75%) に相当するデータ (第 3 四分位数 : q ₃) |
| 4 | データの最大値 |

	F	G	H
20	特性値		
21	データ数	=COUNT(ballpark)	
22	算術平均		
23	メディアン		
24	モード		
25	分散		
26	標準偏差		
27			
28	最大値		
29	第3四分位数		
30	第2四分位数		
31	第1四分位数		
32	最小値		
33	レンジ		
34	四分位範囲		

第2引数に 0, 2, 4 のいずれかの数値を指定すると、QUARTILE 関数の戻り値は、それぞれ MIN 関数、MEDIAN 関数、MAX 関数の戻り値に等しくなる。

【課題 11】都道府県別球場数のデータについて、5 数要約をおこなってみよう。最大値(MAX)、最小値(MIN)は AVERAGE 関数と同様の方法で、四分位数は、関数 QUARTILE を用いて求める。レンジ、四分位範囲は、計算によって導出できる。

4. ランキングの表示

特定のデータ(指定数値)が全体で何番目の大きさになるかを調べるには、関数 RANK を用いる。=RANK(引数 1, 引数 2, 引数 3)において、引数 1 は指定数値、引数 2 は範囲を指定する。引数 3 は昇順(小さい順)か降順(大きい順)を決めるためのオプションで、1 のとき昇順、0 のとき降順となる。省略すると、0 を指定したことになる。

【課題 12】関数 RANK を用いて、球場数の多い順に都道府県に順位を付けてみよう。

📖 手順

- ① C4セルに =RANK(B4, ballpark) と記述する。(B4セルがballparkの中で何番目に大きいのか)
- ② C5:C50セルに、C4セルをコピーすると、全都道府県の順位が求められる。

※ C列とD列の間に1列挿入しよう。「D」と書いてある所にカーソルをおき、「挿入」ボタンをクリックする。

5. 箱ひげ図の表示

Excel2016以降、箱ひげ図を描く機能が標準装備された。しかし、この機能で描く箱ひげ図は、授業内で説明した箱ひげ図とは少し異なり、「外れ値」を考慮している。

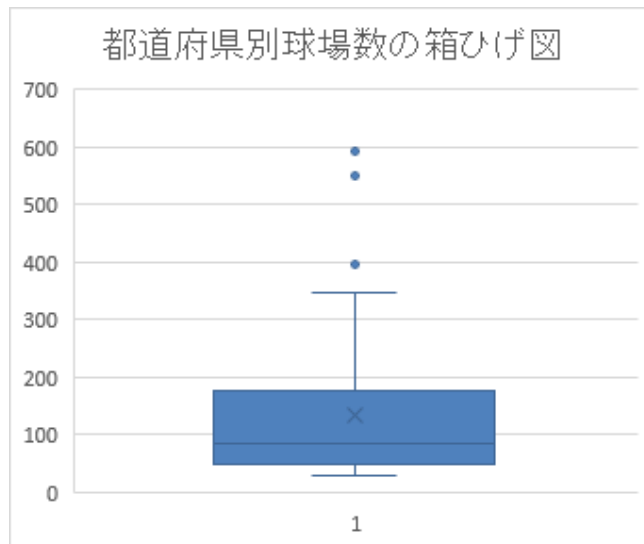
この箱ひげ図では、ひげの長さは箱の長さの1.5倍までとなっており、その範囲を超えた値は、「外れ値」として扱われ、丸印であらわされる。

また、箱の内部にある×印は平均を表す。Excelではデフォルトで表示されてしまうようである。

【課題13】Excel2016以降に標準増備されている、箱ひげ図の描画機能を使って、箱ひげ図を描いてみよう。

📖 手順

- ① B4:B50を範囲指定し、メニューの「挿入」－「統計グラフ」－「箱ひげ図」を選択する。
- ② グラフタイトルを「都道府県別球場数の箱ひげ図」とし、グラフのサイズを変更すると右のようになる。



6. 罫線・フォントの変更

罫線は必ずしも必要なものではないが、引いておくと表を見やすくすることができる。罫線を引くときは、あらかじめ、罫線を引きたい範囲を選択しておいてから、リボン内のフォントのグループ（ホームタブをクリックすることで表示される。）にある、罫線設定のボタンを展開し、罫線を引く場所を選択する。罫線について細かく指定したい場合は、最下部にある「その他の罫線」を選択すれば、指定することができる。この方法はExcel2003まででおこなわれていた方法（セルの書式設定）である。

フォントの変更も表を見やすくすることができる。また、文字の大きさを変えたり、太字や斜体字を用いることもできる。各自試してみよ。

☒ 本日実習したファイルは、河田まで提出すること。

1. Webclass 経由
2. E-mail に添付ファイルとして(送付先アドレスは kawada@tokuyama-u.ac.jp)のいずれでもかまわない。

締め切りは 11月12日(月)9:15 とする。

なお、ファイル名は「都道府県別球場数 E48-○○○」のように、学籍番号をつけること。

☒ 任意課題（加点対象）

自分の関心のある都道府県別データを探し、それについて、都道府県別球場数の課題と同じように、

1. 度数分布表を作成し、
2. ヒストグラムを描き、
3. 特性値を求め、
4. 都道府県に順位をつけ、
5. 罫線を引いたり、フォントを変更して、

A4判1枚に収まるように課題を完成しよう。

☞ 課題作成のヒント

- 都道府県別のデータは、総務省統計局『社会生活統計指標 -都道府県の指標-2018』において、多く見つけることができる。

<http://www.stat.go.jp/data/shihyou/> において、「目次」の中の「本書の内容」から、「Ⅱ 基礎データ」へとリンクをたどればデータにたどりつく。

また、総務省統計局『日本統計年鑑』(<http://www.stat.go.jp/data/nenkan/index1.htm>)の中にも、多くの都道府県別データが含まれており、使用することが可能である。

課題の余白(下の方がよいであろう)に、**必ずデータ出典を記入**すること。

『社会生活統計指標 -都道府県の指標-2018』の「Ⅱ 基礎データ」の「A 人口・世帯」の表から人口のデータを利用したのであれば

データ出典： 「社会生活統計指標」A表

のようにあらわす。自分で探したデータであれば、出典のアドレスも明記すること。

「どのデータを用いたのか」を、後から検証できるようにすることが大事である。

- 度数分布表の階級幅は、そのままではうまくいかないことが多い。というのも、都道府県別のデータといっても、ケタ数が大きく異なるからである。たとえば、人口についておこなう場合には、最大値の東京が10000を超える数値になる(千人単位のデータの場合)。このとき、階級幅が50の度数分布表では、階級の数が莫大になってしまう。最大値と最小値の情報をうまく使いながら、適切な階級幅を自分で設定してみよう。

¶ 提出について

提出は他のファイルと同様に、Webclass 経由か、E-mail の添付ファイルで提出すること。

締め切りは **11月27日(月)9:15** とする。

なお、ファイル名は「統計学基礎任意課題 E48-○○○」のように、**学籍番号をつける**こと。