# 統計学実習資料

2018/5/7

## ＜記述統計の復習(度数分布表とヒストグラム、散布図)＞

### 1. 準備

Excelでさまざまな統計分析をおこなう際、最初に使用するデータを準備する必要がある。本来であれば、データを自分で入力したり、ダウンロードしたファイルを整形したりする必要があるが、今回はあらかじめ河田が作成したファイルを用いることにする。

【 課題 1 】 統計学の講義用 HP から、都道府県別球場数のデータと、打者成績 2017 パのファイルをダウンロードしてみよう。

📖 手順

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」をクリックし、「第6回 4月26日(水)」の配布資料にある、「都道府県別球場数」をクリックし、自分の使いやすい場所に保存する。「打者成績2017パ」のファイルも同様である。

### 2. 度数分布表の作成

連続変量や離散変量のうち取りうる値が多いものは幅のある階級を作成し、それぞれの階級に含まれる度数を求める必要がある。

【 課題 2 】 都道府県別球場数のデータを度数分布表にまとめてみよう。

| | D | E | F | G | H |
|---|---|---|---|---|---|
| 3 | 度数分布表 | | | | |
| 4 | 階 | | 級 | 階級値 | 度数 |
| 5 | 0 | − | 49 | =(D5+F5)/2 | |
| 6 | 50 | − | 99 | | |
| 7 | 100 | − | 149 | | |
| 8 | 150 | − | 199 | | |
| 9 | 200 | − | 249 | | |
| 10 | 250 | − | 299 | | |
| 11 | 300 | − | 349 | | |
| 12 | 350 | − | 399 | | |
| 13 | 400 | − | 449 | | |
| 14 | 450 | − | 499 | | |
| 15 | 500 | − | 549 | | |
| 16 | 550 | − | 599 | | |
| 17 | 600 | | | | |

最初に、右図のように、階級の上限と下限、および表頭部を記入する。階級の上限と下限を入力するときは、**連続データの作成**を利用するとよい。

※ 連続データの作成

この場合なら、

i D5とD6に0,50と入力した後で、ドラッグして範囲指定をおこなう。

ii 左ボタンを離した後、カーソルを反転している長方形の右下隅に移動すると、+ の形状のカーソルに変化する。

iii この状態でボタンを押して下方向にドラッグし、最後のセル(この場合はD17)で左ボタンを開放すれば 100, 150, 200 … と等間隔の数字が入力される。F列の場合も同様である。

次に、各階級の階級値として、上限と下限を加えて2で割った値を用いる。G5に **=(D5+F5)/2** と記入し、それをG6:G16にコピーする。

度数を求めるには、FREQUENCY関数を用いる。

この FREQUENCY 関数は**配列関数**[1]である。Excel で関数は、1 つの数値を返すものであるが、配列関数は複数の配列を返すものである。この場合、配列を記入する範囲を指定し、関数を入力した上で、Enter キーの代わりに、Ctrl + Shift + Enter キーを入力する。

ここではデータ範囲、区間範囲にそれぞれ名前をつけて、それらをFREQUENCY関数の中で用いることにする。

📖 手順

① 範囲B4:B50を選択する。数式のタブをクリックし、リボンの中から「名前の定義」をクリックし、「名前」として、 ballpark と入力する。同様に、F4:**F16**の部分に、 class と名前を付ける。

② 範囲H5:H17を選択する。

③ その状態でセルH5に式 =FREQUENCY(ballpark,class) と入力する。

④ Ctrl + Shift + Enter とする。

以上で、都道府県別球場数の度数分布表が作成された。

## 3. ヒストグラムの作成

【 課題 3 】都道府県別球場数の度数分布表をヒストグラムに表してみよう。

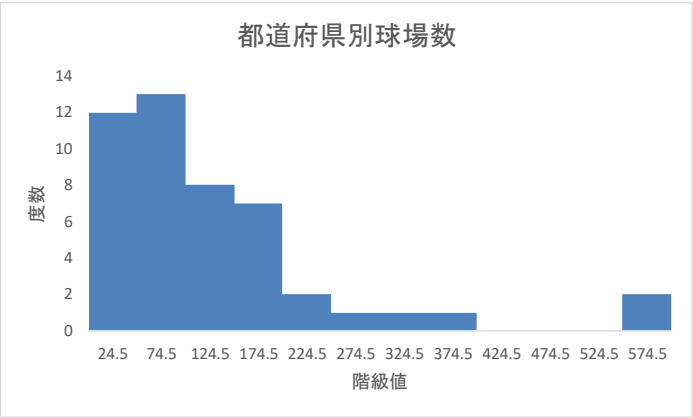
📖 手順

① 最初にグラフに描く範囲を範囲指定する。ここでは、H5:H16(最後の0は除く)を範囲指定する。

② グラフを作成するには、**挿入タブ**をクリックすることで、リボン内にグラフのグループが表示される。ヒストグラムは縦棒グラフの１種なので、 縦棒 のボタンをクリックする。

③ すると縦棒グラフのフォーマット（型式）メニューが出るので、集合縦棒（2-D縦棒の中の左端）をクリックする。

④ この時点でグラフのサンプルが自動的に描かれている。これを修正していく。
まず、横軸ラベルに階級値を用いる。リボンの中の「データの選択」ボタン（「データ」のグループにある）をクリックし、横(項目)軸ラベル の 「編集」ボタンをクリックし、G5:G16を範囲指定し、OKボタンを押す。




⑤ 次に、レイアウトを変更する。リボンの中の「グラフのレイアウト」を展開し、レイアウト8（3段目の真ん中）をクリックし、タイトルや軸ラベルなどが書き込めるようにする。そして、

タイトル：「都道府県別球場数」と記入する。
X/項目軸：「階級値」と記入する。
Y/数値軸：「度数」と記入する。

⑥ 以上で、左図のようなヒストグラムが描けたはずである。グラフの移動はグラフの白い部分をドラッグすることで、大きさの変更はグラフの 4 隅または 4 辺の真ん中にマウスを合わせ、ドラッグすることでできる。各自こころみよ。

---

[1] このような配列関数には、行列の積や逆行列を求めるものなどが他にある。

## 4. 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のようになる。

**一般形：＝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：関数 AVERAGE はこれまでのように、引数に範囲または名前をとる。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(_X1) ― 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( ) などがある。

【 課題 4 】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。

| | 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 | | |

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

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 5 】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。

## 5. 散布図

【 課題 6 】打者成績 2017 パのファイルについて、本塁打数と三振数の相関をみるために、散布図を描いてみよう。

📖 手順

① 最初にグラフに描く範囲を範囲指定する。E2:F28をドラッグし、範囲指定する。

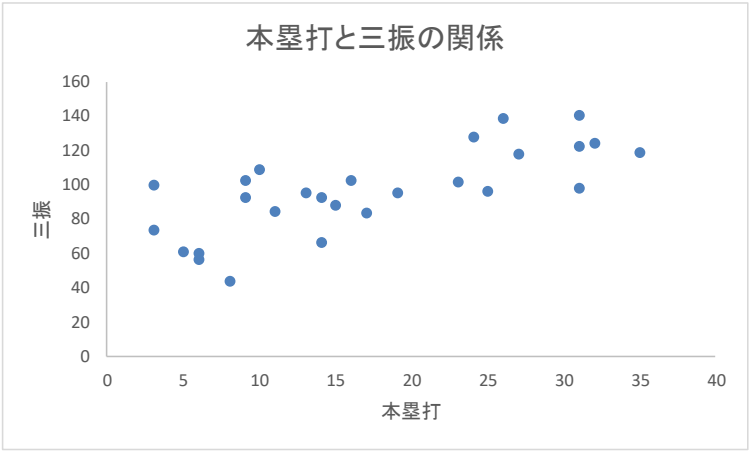
② グラフを作成するには、**挿入タブ**をクリックすることで、リボン内にグラフのグループが表示される。ここでは、散布図 のボタンをクリックする。

③ すると散布図グラフのフォーマット（型式）メニューが出るので、左上の散布図（マーカーのみ）をクリックする。

④ この時点でグラフのサンプルが自動的に描かれている。これを修正していく。まずレイアウトを変更する。リボンの中の「グラフのレイアウト」を展開し、レイアウト1（左上）をクリックし、タイトルや軸ラベルなどが書き込めるようにする。そして、

タイトル：「本塁打と三振の関係」と記入する。
X 軸：「本塁打」と記入する。
Y 軸：「三振」と記入する。

⑤ さらにいくつかの細かい修正を加えたものが右図である。右図のようにするには、
ⅰ．凡例の消去
ⅱ．「軸の書式設定」において、最大値、最小値の変更
ⅲ．目盛線の消去
ⅵ．グラフの大きさ変更
をおこなっている。




⌦ 演習問題

他の指標のくみあわせについても、いろいろ散布図を描いてみよう。隣接しない 2 変量は、1 変量をドラッグした後、Ctrl キーを押しながらもう 1 つをドラッグすることで範囲指定できる。

また、複数の変量間の相関係数をいっぺんに計算するには、統計分析を行うための分析ツールを用いればよい。

分析ツールを最初に使用する場合には、アドイン(有効にすること)しなくてはならない。分析ツールのアドインは次のようにおこなう。

① 「ファイル」のタブをクリックし、下にある「オプション」のボタンをクリックする。

② 「Excel のオプション」のウインドウが開くので、左側の「アドイン」をクリックする。

③ 一番下に表示される「Excel アドイン」の右の設定ボタンを押す。

④ 「分析ツール」にチェックをつけ、OK ボタンをクリックする。

すると、データタブの中に「データ分析」のボタンが出てくるので、下のほうにある、「相関」を選び、ウイザードの要求に従ってデータ範囲を指定すれば、相関係数行列が計算できる。

⌦ 本日実習した 2 つのファイルは、E-mail に添付ファイル、または webclass 経由で河田(アドレスは kawada@tokuyama-u.ac.jp)まで提出すること。

締め切りは **5 月 14 日(月)9:15** とする。

ファイル名に「都道府県別球場数 E47-〇〇〇」のように、**学籍番号をつける**こと。