

<度数分布表、ヒストグラム、円グラフ>

1. 準備

今回も、実習にはあらかじめ河田が作成したファイルを用いる。

【課題 5】 統計学基礎の講義用HPから、誕生日血液型のデータと都道府県別球場数¹のファイルをダウンロードしてみよう。

📖 手順

- ① 徳山大学のトップページ (<http://www.tokuyama-u.ac.jp>) で、「在学生の皆様へ」のメニューから、「教員紹介」を選び、河田の教員紹介を開く。
- ② そのページの下の方にある、「個人サイト」のところをクリックすると、「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)
- ③ 「統計学基礎」をクリックし、「第6回 10月11日(木)」の配布資料にある、「誕生日血液型2012」をクリックし、自分の使いやすい場所に保存する。「都道府県別球場数」のファイルも同様である。

2. 円グラフの描画

【課題 6】 この講義の履修者の血液型の割合を、円グラフであらわしてみよう。(誕生日血液型 2012 のファイルを使用する。)

📖 手順

- ① 円グラフを描くとき、まず度数分布表を作成し、それをグラフに描くという手順をとる。この場合は血液型という質的変量であるので、それぞれの1つの階級となる。F8:F12に右図のように記入しよう。

	F	G
8	度数分布表	
9	1	
10	2	
11	3	
12	4	
13		

- ② G9:G12に対応する度数を求める。ここでは、FREQUENCY関数を用いる。

関数FREQUENCYは配列関数²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。具体的には、

- i. 範囲 G9:G12 を選択する。
- ii. その状態でセルG9 に式 =FREQUENCY(D2:D41,F9:F12)³ と入力する。
- iii. **Ctrl** + **Shift** + **Enter** とする。

- ③ 次に、グラフを描く範囲として、G9:G12を範囲指定して、円グラフのボタンをクリックする。(リボン内にグラフのグループが表示されない場合は、挿入タブをクリックことで表示される。)

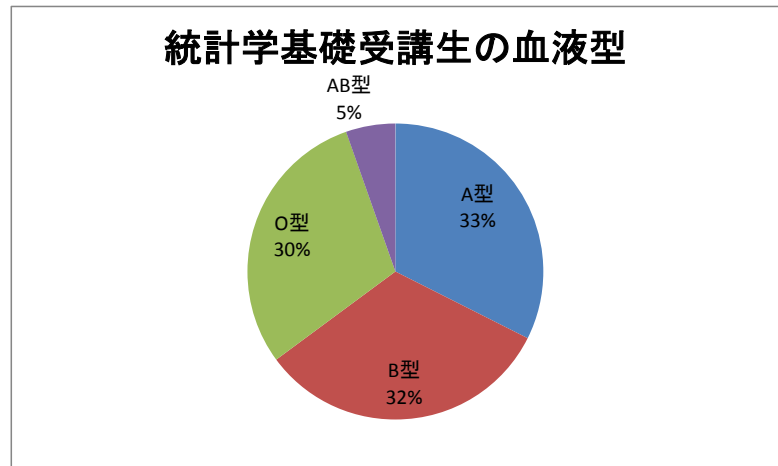
- ④ すると円グラフのフォーマット (型式) メニューが出るので、円 (2-D円の中の左上) をクリックする。

¹ 野球場、ソフトボール場をあわせたものを、ここでは「球場」とよぶことにする。

² 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。

³ FREQUENCY 関数は、=FREQUENCY(データ配列, 区間配列)と記述し、データ配列の中で区間配列の値を超えないものの数を数える。正確には最大の階級の値は区間配列として指定しなくてよい。=FREQUENCY(D2:D54,F9:F11)でよいのである。これは、最大の階級が、すぐ下の階級の値を超える数を数えるからである。

- ⑤ 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。
凡例を「1,2,3,4」ではなく、「A型,B型・・・」のようにしましょう。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、G3:G6を範囲指定し、OKボタンを押す。



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 (左下のもの) をクリックし、タイトルが書き込め、凡例とその割合が表示されるようにする。そして、

タイトル:「統計学基礎受講生の血液型」と記入する。

このグラフでは、AB型の割合が非常に小さく、引き出し線で割合を表示することになる。「AB型 5%」というデータラベル右クリックし、「データラベルの書式設定」を選び、「引き出し線を表示する」のチェックをはずすと、右上のようなグラフが描かれる。

3. 幅のある階級をもつ度数分布表の作成

質的変量や、離散変量のうち取りうる値が少ないものは、取りうる値それぞれに対応した度数を求め、度数分布表を作成していた。しかし、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 7】 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。

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、必要なところ(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。

	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				

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 $=(D5+E5)/2$ と記入し、それをG6:G16にコピーする。

度数を求める手順は、【課題 6】 とほぼ同様であるが、今回はデータ範囲、区間範囲にそれぞれ名前をつけて、それを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。

4. ヒストグラムの作成

【課題 8】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

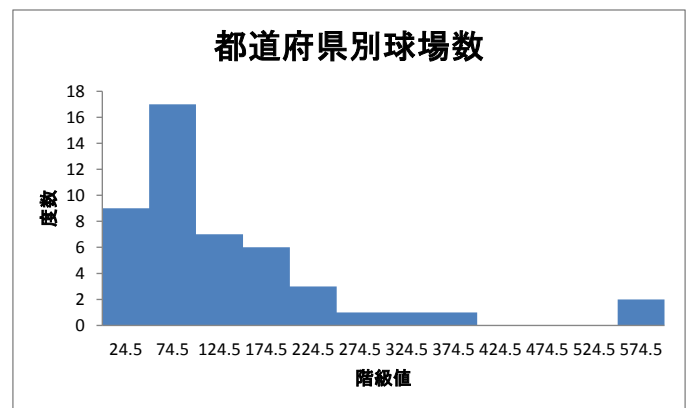
- ① 最初にグラフに描く範囲を範囲指定する。ここでは、H5:H16(最後の0は除く)を範囲指定する。
- ② グラフを作成するには、**挿入タブ**をクリックすることで、リボン内にグラフのグループが表示される。ヒストグラムは縦棒グラフの1種なので、**縦棒**のボタンをクリックする。
- ③ すると縦棒グラフのフォーマット(型式)メニューが出るので、**集合縦棒(2-D縦棒)**の中の**左端**をクリックする。
- ④ この時点でグラフのサンプルが自動的に描かれている。これを修正していく。まず、横軸ラベルに階級値を用いる。リボンの中の「データの選択」ボタン(「データ」のグループにある)をクリックし、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。



- ☒ 本日実習したファイルのうち、**誕生日血液型 2012**の分は、河田まで提出すること。なお、誕生日のデータに関して、適切なグラフを描いたものはその分加点する。

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 **10月18日(木)9:10** とする。

なお、ファイル名は「誕生日血液型 2012 E42-○○○」のように、**学籍番号**をつけること。