

データリテラシー

第4回 度数分布表とヒストグラム

メディア情報コース
平居 悠（ひらい ゆたか）

到達目標

- データサイエンス・AIが社会でどのように活用されているのかを理解する。
- データを適切に読み解き理解する力をつける。
- データを適切な可視化手法で他者に説明できるようになる。
- 表計算ソフトを使ってデータを集計・加工できるようになる。

前回の目標

表計算ソフトでデータを集計
できるようになる。

前回学んだこと

1. セルの参照
2. 数式
3. 関数

セルの参照

セルの場所を列番号と行番号で指定してそのセルの値を参照すること

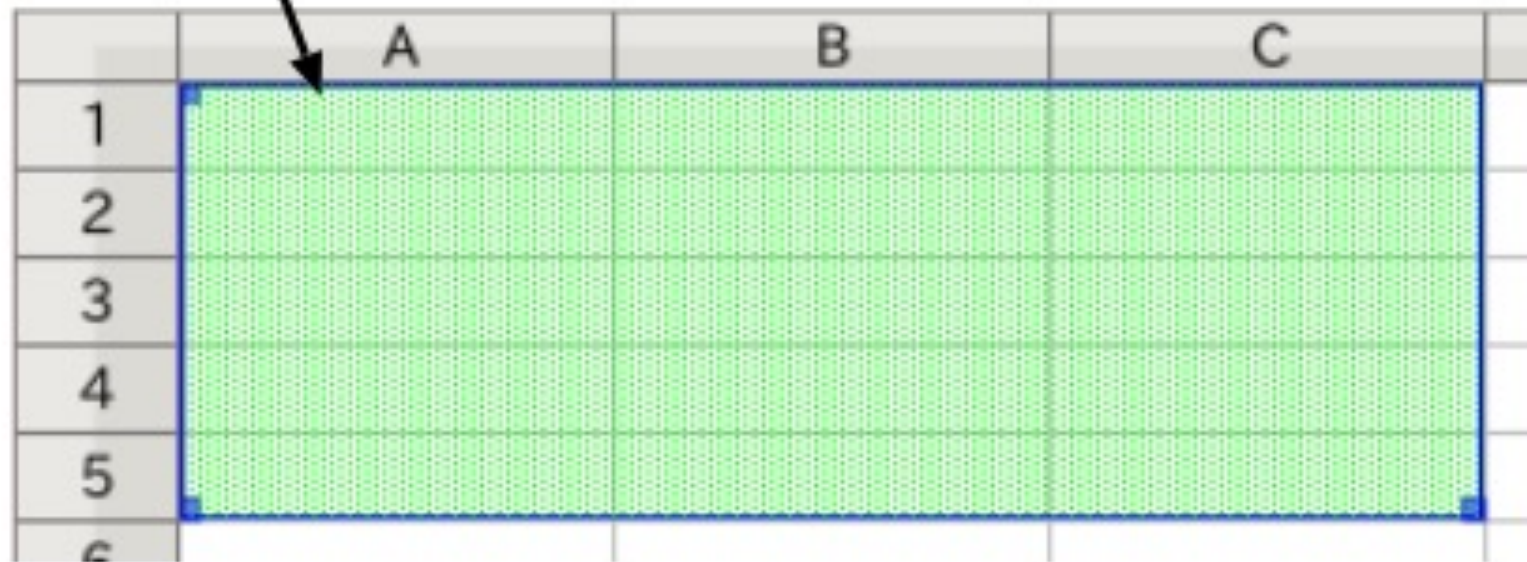
相対参照と絶対参照

セルの参照方法	列番号	行番号
B4	相対参照	相対参照
\$B4	絶対参照	相対参照
B\$4	相対参照	絶対参照
\$B\$4	絶対参照	絶対参照

セル範囲の参照

「セル1：セル2」のように「：」（コロン）を用いることで、連続する複数セルから成る範囲をまとめて参照できる。マウスでドラッグしても参照できる。

「A1:C5」



	A	B	C
1			
2			
3			
4			
5			
6			

数式

セルに「**=**」で入力を始めると
数式として認識される。

数式の中で文字列は「**“”**」（ダブルクォーテーション）で囲む必要がある。

算術演算子

算術演算子	意味	優先順位
+	加算（足し算）	低
-	減算（引き算）	低
*	乗算（かけ算）	中
/	除算（わり算）	中
^	べき乗	高

比較演算子

IF関数を使用して「真 (TRUE)」または「偽 (FALSE)」の答えを返す数式に使用する。

比較演算子	意味
=	左辺と右辺が等しいか？
<>	左辺と右辺が等しくないか？
<	左辺が右辺より小さいか？
<=	左辺が右辺以下か？ ($\leq$)
>	左辺が右辺より大きいのか？
>=	左辺が右辺以上か？ ($\geq$)

関数

ある入力に対して計算や処理を行ってその結果を出力するものを**関数**という。

関数

関数は以下の形をしている。

関数名 (引数)

引数（ひきすう）：関数に渡す入力

SUM関数

引数の合計を計算する。

SUM (引数)

- 引数には数値・セル・セル範囲を記述できる。
- 引数は「,」で区切って複数記述できる。

IF(論理式, 真の場合の値, 偽の場合の値)

論理式が「真」か「偽」かに応じて値を返す。

前回の問い

- セルの参照にはどのような方法があるか？
 - 列番号か行番号の前に「\$」を付けると絶対参照、何も付けないと相対参照になる。
- 数式はどのように入力すればよいか？
 - セルに「=」で入力を始めると数式として認識される。
- 表計算ソフトの関数とは何か？
 - ある入力に対して計算や処理を行ってその結果を出力するもの。

タイピング練習スケジュール

第1回	ホームポジション
第2回	ホームポジション
第3回	ローマ字
第4回	英語初級
第5回	日本国憲法（trr試験）
第6回	ホームポジション
第7回	ローマ字
第8回	英語初級
第9回	日本国憲法（trr試験）
第10回	ホームポジション
第11回	ローマ字
第12回	英語初級
第13回	日本国憲法（trr試験）

タイピングの練習 (英語初級)

1. ブラウザを起動し、<https://www.koeki-prj.org/trr/>に繋ぐ。
2. 学籍番号（Cは大文字、省略なし8桁）を入力する。
3. Koeki MAILに届いたパスコードをPasscode: 欄に入力する。

ホームポジション

左手でタイプするキー

右手でタイプするキー



左手の人差指から小指までの
ホームポジション

両手の親指の
ホームポジション

右手の人差指から小指までの
ホームポジション

今回

第1回	社会におけるデータ・AI利活用
第2回	表計算ソフトの基本操作
第3回	セルの参照と集計
第4回	度数分布表とヒストグラム
第5回	代表値と散らばりの尺度
第6回	データのグラフによる表現とデータ比較
第7回	データの分布と標本抽出
第8回	推定
第9回	分割表とクロス集計
第10回	2変量データと相関
第11回	時系列データ
第12回	仮説検定
第13回	回帰分析

今回の目標

度数分布表を用いてヒストグラムを描けるようになる。

今回の問い

- 度数分布表とは何か？
- ヒストグラムを描くと何がわかるか？

今回学ぶこと

1. データとは
2. 度数分布表
3. ヒストグラム

今回学ぶこと

1. データとは
2. 度数分布表
3. ヒストグラム

データとは

何らかの実験・観測・調査
を行って得られる観測値の
集合

生データ

実験を行って記録したデータやアンケート調査で得られたデータで、集計などの加工を行う前のデータ

完全データと不完全データ

- **完全データ**：全ての観測値がそろっているデータ
- **不完全データ**：所々の観測値が欠けているデータ
- **欠損値（欠測値）**：観測値が欠けているところ

表による表現

例 1 : ある集団の性別、年齢、身長の実データ

番号	性別	年齢[歳]	身長[cm]
1	女	22	160.6
2	男	19	171.6
3	男	21	173.2
⋮	⋮	⋮	⋮

変数 (変量)

表による表現

変数は x, y, z などの記号で表し、それぞれの変数の値は $x_1, x_2, x_3, \dots$ のように番号を下付きの添え字で区別することがある。

番号 i	性別 x	年齢[歳] y	身長[cm] z
1	x_1	y_1	z_1
2	x_2	y_2	z_2
3	x_3	y_3	z_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$

表による表現

例 2：例 1 の表の行と列を入れ替えたもの

番号	1	2	3	...
性別	女	男	男	...
年齢[歳]	22	19	21	...
身長[cm]	160.6	171.6	173.2	...

これでも問題はないが、例 1 の形式の方が表計算ソフトでは扱いやすい。

表による表現（度数分布表）

例3：例1の生データから「年齢」を10代、20代、30代でそれぞれ集計して表にしたデータ

年齢	度数
10代	341
20代	156
30代	3
合計	500

表による表現（分割表、クロス集計表）

例4：例1の生データで「年齢」と「性別」で人数を集計して表にしたデータ

		年齢			
		10代	20代	30代	合計
性別	男	173	80	0	253
	女	168	76	3	247
	合計	341	156	3	500

分布

データがどういう値でどのような散らばっているかを**分布**という。

分布を読み取る方法

1. グラフを描いて視覚的に読み取る
2. 代表値と散らばりの尺度で分布の特徴を数量的に読み取る

データの分布をグラフとして表現するために、度数分布表を作成してグラフ（**ヒストグラム**）に描くことがよく行われる。

今回学ぶこと

1. データとは
- 2. 度数分布表**
3. ヒストグラム

度数分布表とは

生データをいくつかの**階級**というグループに分け、それぞれの階級に入る観測値がいくつか数えて**度数**（**頻度**）で表した表

度数分布表の例

階級を区別するための番号。元データの個体を区別するための番号とは異なる。

階級番号 i	下限値 l_i	上限値 u_i	階級値 v_i	度数 f_i	累積度数 F_i	相対度数 f_i/n
1	140	145	142.5	2	2	0.004
2	145	150	147.5	12	14	0.024
3	150	155	152.5	52	66	0.104
4	155	160	157.5	91	157	0.182
5	160	165	162.5	111	268	0.222
6	165	170	167.5	100	368	0.200
7	170	175	172.5	84	452	0.168
8	175	180	177.5	34	486	0.068
9	180	185	182.5	13	499	0.026
10	185	190	187.5	1	500	0.002
合計				500		1.000

度数分布表の例

階級の下限值

階級番号 i	下限値 l_i	上限値 u_i	階級値 v_i	度数 f_i	累積度数 F_i	相対度数 f_i/n
1	140	145	142.5	2	2	0.004
2	145	150	147.5	12	14	0.024
3	150	155	152.5	52	66	0.104
4	155	160	157.5	91	157	0.182
5	160	165	162.5	111	268	0.222
6	165	170	167.5	100	368	0.200
7	170	175	172.5	84	452	0.168
8	175	180	177.5	34	486	0.068
9	180	185	182.5	13	499	0.026
10	185	190	187.5	1	500	0.002
合計				500		1.000

度数分布表の例

階級の上限値

階級番号 i	下限値 l_i	上限値 u_i	階級値 v_i	度数 f_i	累積度数 F_i	相対度数 f_i/n
1	140	145	142.5	2	2	0.004
2	145	150	147.5	12	14	0.024
3	150	155	152.5	52	66	0.104
4	155	160	157.5	91	157	0.182
5	160	165	162.5	111	268	0.222
6	165	170	167.5	100	368	0.200
7	170	175	172.5	84	452	0.168
8	175	180	177.5	34	486	0.068
9	180	185	182.5	13	499	0.026
10	185	190	187.5	1	500	0.002
合計				500		1.000

度数分布表の例

階級を代表する
値で、通常は下
限值と上限値の
中央の値をとる。

階級番号 i	下限值 l_i	上限値 u_i	階級値 v_i	度数 f_i	累積度数 F_i	相対度数 f_i/n
1	140	145	142.5	2	2	0.004
2	145	150	147.5	12	14	0.024
3	150	155	152.5	52	66	0.104
4	155	160	157.5	91	157	0.182
5	160	165	162.5	111	268	0.222
6	165	170	167.5	100	368	0.200
7	170	175	172.5	84	452	0.168
8	175	180	177.5	34	486	0.068
9	180	185	182.5	13	499	0.026
10	185	190	187.5	1	500	0.002
合計				500		1.000

度数分布表の例

階級に属する観測値の個数。度数の合計が元データの総数 n となる。

階級番号 i	下限値 l_i	上限値 u_i	階級値 v_i	度数 f_i	累積度数 F_i	相対度数 f_i/n
1	140	145	142.5	2	2	0.004
2	145	150	147.5	12	14	0.024
3	150	155	152.5	52	66	0.104
4	155	160	157.5	91	157	0.182
5	160	165	162.5	111	268	0.222
6	165	170	167.5	100	368	0.200
7	170	175	172.5	84	452	0.168
8	175	180	177.5	34	486	0.068
9	180	185	182.5	13	499	0.026
10	185	190	187.5	1	500	0.002
合計				500		1.000

度数分布表の例

最初の階級からその階級までの度数の和。最後の階級の累積度数は総数 n に一致する。

階級番号 i	下限値 l_i	上限値 u_i	階級値 v_i	度数 f_i	累積度数 F_i	相対度数 f_i/n
1	140	145	142.5	2	2	0.004
2	145	150	147.5	12	14	0.024
3	150	155	152.5	52	66	0.104
4	155	160	157.5	91	157	0.182
5	160	165	162.5	111	268	0.222
6	165	170	167.5	100	368	0.200
7	170	175	172.5	84	452	0.168
8	175	180	177.5	34	486	0.068
9	180	185	182.5	13	499	0.026
10	185	190	187.5	1	500	0.002
合計				500		1.000

度数分布表の例

度数 f_i を総数 n で割った割合。相対度数の合計は1になる。

階級番号 i	下限値 l_i	上限値 u_i	階級値 v_i	度数 f_i	累積度数 F_i	相対度数 f_i/n
1	140	145	142.5	2	2	0.004
2	145	150	147.5	12	14	0.024
3	150	155	152.5	52	66	0.104
4	155	160	157.5	91	157	0.182
5	160	165	162.5	111	268	0.222
6	165	170	167.5	100	368	0.200
7	170	175	172.5	84	452	0.168
8	175	180	177.5	34	486	0.068
9	180	185	182.5	13	499	0.026
10	185	190	187.5	1	500	0.002
合計				500		1.000

度数分布表の作成方法

観測値の総数 n の
生データがある
とする。

番号	身長[cm]
1	160.6
2	171.6
3	173.2
⋮	⋮
n	163.9

度数分布表の作成方法

①全体の範囲を決める

データの最小値と最大値から全体の範囲を決める。階級幅と合わせてきりがよくなるように決めた方が分かりやすい。

度数分布表の作成方法

②階級数と階級幅を決める

それぞれの階級の**階級幅**は上限値から下限値を引いたものに等しい。普通は階級幅を等幅にとるため、全体の範囲を階級数で割ったものが階級幅である。

階級数の決め方：スタージェスの公式

$$k \simeq 1 + \log_2 n$$

k: 階級数

n: データの総数

右辺が整数にならない場合は、切り上げて整数にする。

階級幅の決め方

スタージェスの公式の階級数をもとに、階級幅が5や10など、きりのよい数字になるように階級数を決めると分かりやすい。

度数分布表の作成方法

③各階級の下限值と上限値を設定する

階級数と階級幅をもとに各階級の下限值と上限値を設定する。 i 番目の階級を代表する値を階級値 v_i という。階級値 v_i は普通下限値 l_i と上限値 u_i の中央の値にとる。

$$v_i = \frac{l_i + u_i}{2}$$

度数分布表の作成方法

④観測値がどの階級に属するか調べて度数として数える

i 番目の階級にいくつの観測値が入るかを数えて度数 f_i とする。通常**下限値以上、上限値未満**の観測値を数える。

$$(\text{下限値}) \leq (\text{観測値}) < (\text{上限値})$$

度数分布表の作成方法

⑤ 度数の合計が観測値の総数と一致することを確認する

観測値はいずれかの階級に属するため、度数 f_i の合計は観測値の総数 n に一致する。

$$n = f_1 + f_2 + \cdots + f_k = \sum_{i=1}^k f_i$$

度数分布表の作成方法

⑥累積度数や相対度数を求める

必要なら度数から累積度数や相対度数を求める。

今回の問い

- 度数分布表とは何か？
 - データをいくつかの階級に分け、それぞれの階級に入る観測値がいくつか数えて度数で表した表。
- ヒストグラムを描くと何がわかるか？

演習：度数分布表の作成

「身長」のデータ (dataset1.csv) から度数分布表を作成する。

演習：度数分布表の作成

データの横のセルG1からM2に以下のような度数分布表のラベルを用意する。

	G	H	I	J	K	L	M
1	身長の度数分布表						
2	階級番号	下限値	上限値	階級値	度数	累積度数	相対度数

演習：度数分布表の作成

①全体の階級を決める

データの最小値と最大値を求めてデータの範囲を決める。

1. 最小値は空のセルに「**=MIN(D2:D501)**」と入力して求める。
2. 最大値は空のセルに「**=MAX(D2:D501)**」と入力して求める。

演習：度数分布表の作成

②階級数と階級幅を決める

1. スタージェスの公式は空のセルに「 $=1+\text{LOG}(500, 2)$ 」と入力して計算する。
2. $(190-140)/10 = 5$ なので、階級数を10にとると階級幅は5となっていてきりが良い。
3. セルG3に1を入力し、セルG12まで1ずつ増やした数を入力する。

演習：度数分布表の作成

③各階級の下限值と上限値を設定する

1. 下限値はセルH3に140を入力し、セルH12まで階級幅5ずつ増やしていく。
2. 上限値はセルI3に145を入力し、セルI12まで階級幅5ずつ増やしていく。
3. 各階級の階級値は、セルJ3に「 $=(H3+I3)/2$ 」と入力し、セルJ3をコピーしてセルJ4からJ12に貼り付ける。

演習：度数分布表の作成

	G	H	I	J	K	L	M
1	身長の度数分布表						
2	階級番号	下限値	上限値	階級値	度数	累積度数	相対度数
3	1	140	145	142.5			
4	2	145	150	147.5			
⋮	⋮	⋮	⋮	⋮			
12	10	185	190	187.5			

演習：度数分布表の作成

④観測値がどの階級に属するか調べて度数として数える

COUNTIFS関数を使って度数を数える

COUNTIFS関数

COUNTIFS(範囲1, 条件1, 範囲2, 条件2)

範囲1の中で条件1に合うセル、かつ、範囲2の中で条件2に合うセルの個数を数える。範囲と条件のセットは3セット以上でも良い。

演習：度数分布表の作成

④観測値がどの階級に属するか調べて度数として数える

1. セルK3に「**=COUNTIFS(D\$2:D\$501, “>=”&H3, D\$2:D\$501, “<”&I3)**」と入力する。
2. セルK3をコピーしてK4からK12に貼り付ける。
「“>=”&H3」は「セルH3の値以上」という条件、
“<”&I3」は「セルI3の値未満」という条件。
「&」はテキストを結合する演算子

演習：度数分布表の作成

④観測値がどの階級に属するか調べて度数として数える

この条件によって各階級で数えられる度数は以下の観測値に合うものになる。

$$(\text{下限値}) \leq (\text{観測値}) < (\text{上限値})$$

参考：FREQUENCY関数

FREQUENCY(データ範囲,区間)

データ範囲の観測値を区間に従って階級に分けて度数を求める。

参考：FREQUENCY関数を使う方法

1. セルK3に「**=FREQUENCY(D2:D501, I3:I11)**」と入力する。
2. 入力したら**Shift+Ctrl+Enter**を押す。
3. 自動的に度数が求められる。

参考：FREQUENCY関数を使う方法

FREQUENCY関数で指定する区間は上限値である。ただし、各階級で数えられる度数は以下の観測値に合うものになる。

$$(\text{下限値}) < (\text{観測値}) \leq (\text{上限値})$$

演習：度数分布表の作成

⑤度数の合計が観測値の総数と一致することを確認する

1. セルK13に「**=SUM(K3:K12)**」と入力する。
2. 合計が500になることを確認する。

演習：度数分布表の作成

⑥ 累積度数や相対度数を求める

累積度数を求める。

1. セルL3に「**=K3**」と入力する。
2. セルL4に「**=L3+K4**」と入力する。
3. セルL4をコピーしてセルL5からL12に貼り付ける。

演習：度数分布表の作成

⑥ 累積度数や相対度数を求める

相対度数を求める。

1. セルM3に「**=K3/500**」と入力する。
2. セルM3をコピーしてセルM4からM12に貼り付ける。
3. セルM13に「**=SUM(M3:M12)**」と入力して相対度数の合計を求める。
4. 合計が1になることを確認する。

今回学ぶこと

1. データとは
2. 度数分布表
3. ヒストグラム

ヒストグラムとは

度数分布表を柱状グラフで
表したもの

ヒストグラムとは

横軸に階級値を取り、横幅を階級幅として長方形を隙間なく並べた柱状グラフ。

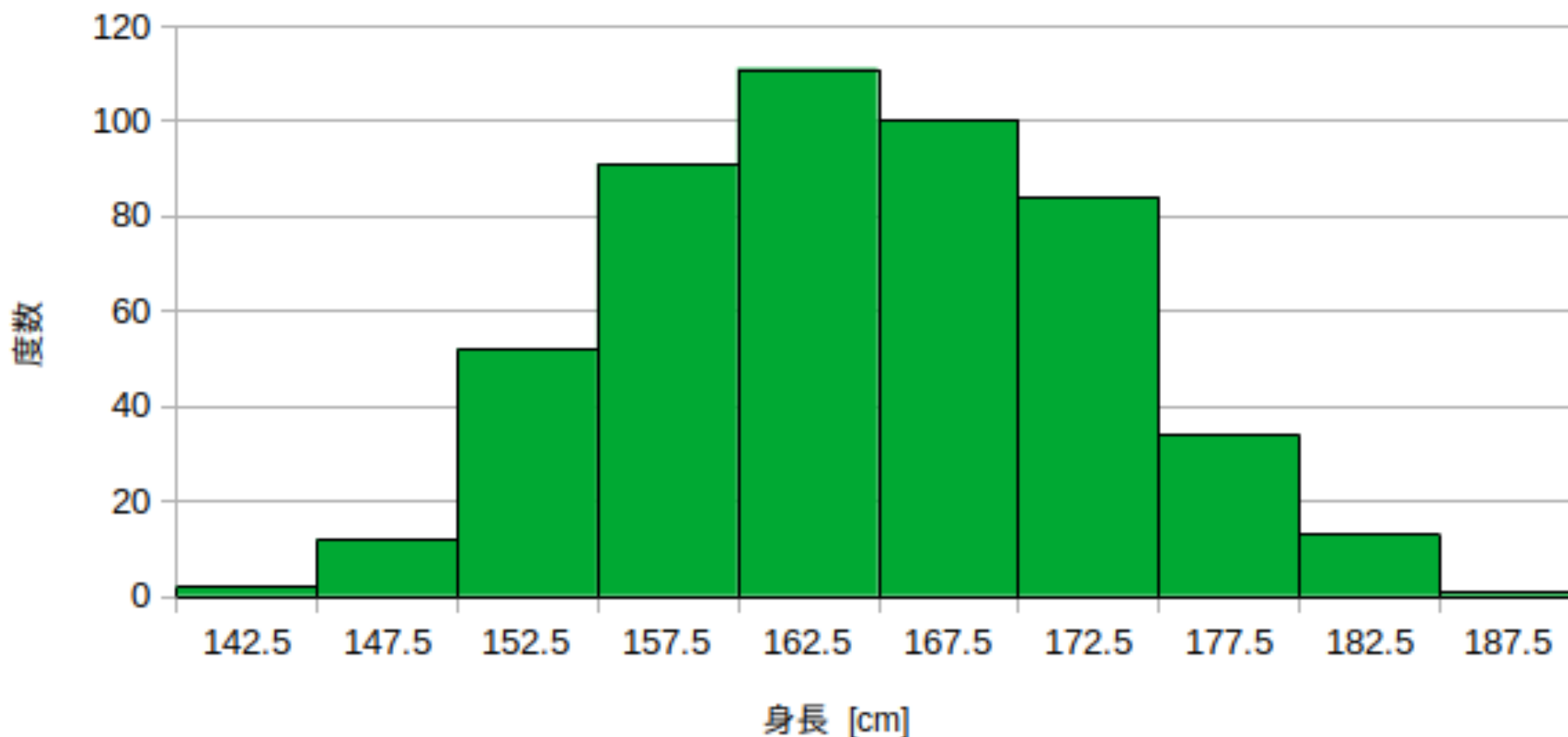
ヒストグラムとは

普通、階級幅は全ての階級で同じなので、長方形の高さは度数にとる。階級幅が階級ごとに異なる場合は、面積が度数に比例するように高さを調整する。

ヒストグラムの例

dataset1.csvの身長の高さのヒストグラム。スタージェスの公式を使って階級幅は5 cmにしてある。

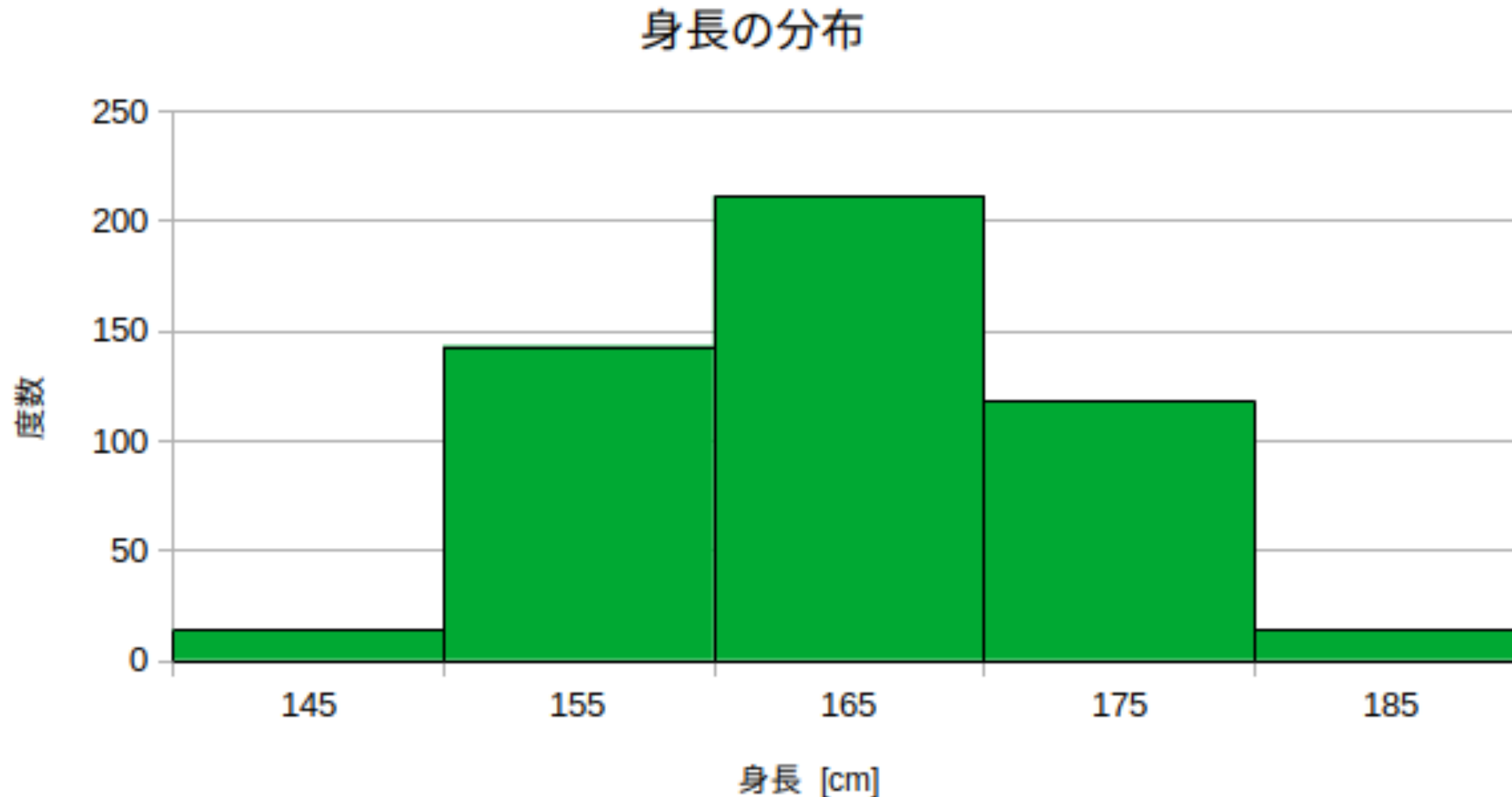
身長分布



ヒストグラム

ヒストグラムを描くことで、分布の中心位置や散らばり具合などの特徴を知ることができる。

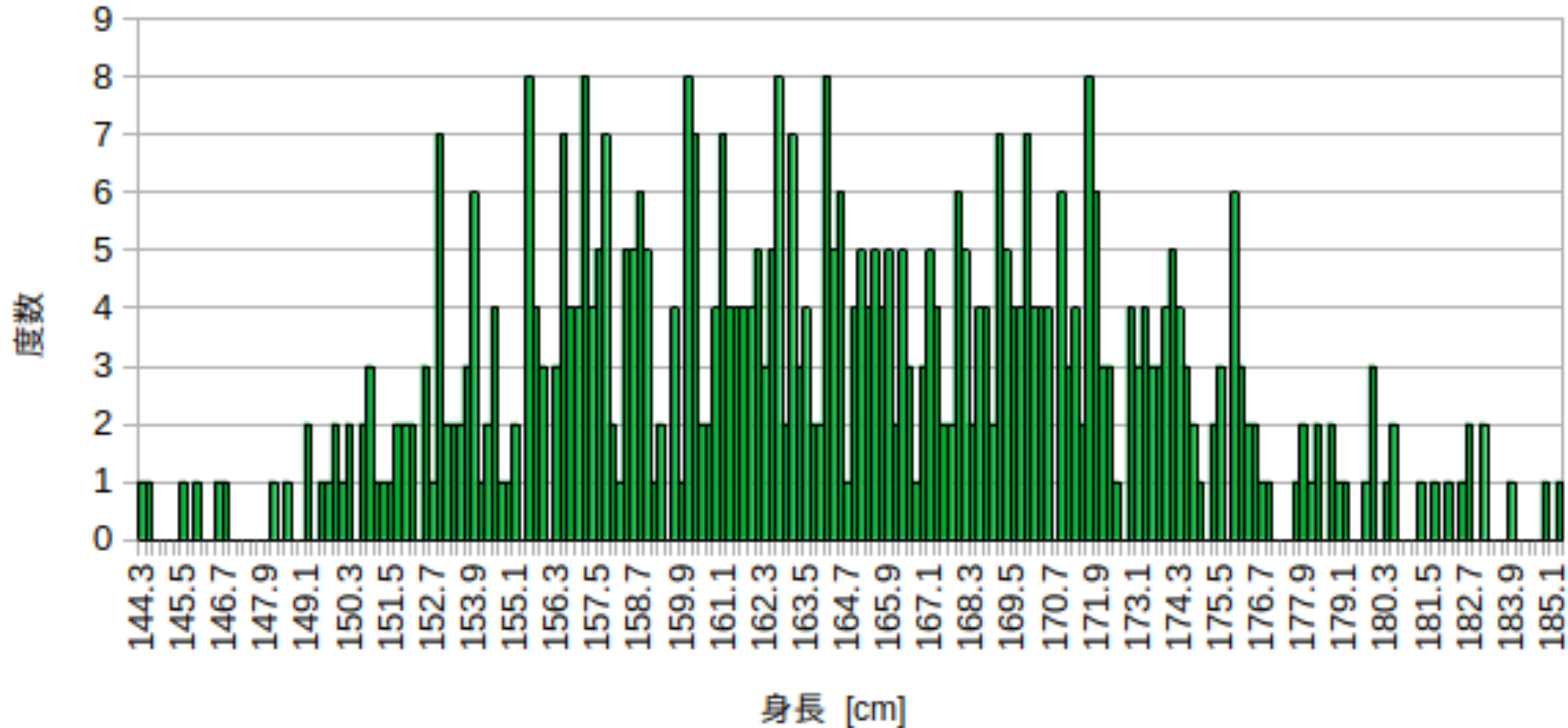
階級幅10 cmにした場合



分布が大雑把になって細かな分布が見えなくなる。 75

階級幅0.2 cmにした場合

身長の分布



細かいところが見えすぎて凸凹が激しく、全体的な分布の傾向がつかみにくくなる。

演習：ヒストグラムの作成

① グラフの作成

1. 度数のセルK3からK12を選択する。
2. メニューバーの[挿入]→[グラフ]を選択してグラフウィザードのダイアログを表示する。
3. 「**グラフの種類**」で[次へ]をクリックする（デフォルトだとグラフの種類は縦棒）。
4. 「**データの範囲**」で[次へ]をクリックする。
5. 「**データの系列**」で[次へ]をクリックする。
6. 「**グラフの要素**」で次のように入力して[完了]をクリックする。
 1. タイトル：「身長分布」
 2. X軸：「身長 [cm]」
 3. Y軸：「度数」

演習：ヒストグラムの作成

②グラフの移動

グラフが度数分布表の上に重なって邪魔な場合は以下の手順で移動させる。

1. グラフ以外のセルを一度クリックする。
2. グラフをシングルクリックし、ドラッグで度数分布表の下に空いているセルの上に移動させる。

演習：ヒストグラムの作成

③階級値の追加

1. グラフ以外のセルを一度クリックする。
2. グラフをダブルクリックする。
3. グラフ上で右クリックして[**データ範囲**]を選択して「データ範囲」のダイアログを表示する。
4. [**データ系列**]タブをクリックし、右下の「範囲」の入力フォームの右側のボタンをクリックして、マウスでセルJ3からJ12を選択する。
5. [**OK**]をクリックする。

演習：ヒストグラムの作成

④棒グラフをヒストグラムにする

1. グラフ以外のセルを一度クリックする。
2. グラフをダブルクリックする。
3. 系列の棒をダブルクリックして「データ系列」のダイアログを表示する。
4. [オプション]タブをクリックし、「間隔」を0%にする。
5. [線]タブ（もしくは[外枠]タブ）をクリックし、「スタイル」で[実線]を選択する。
6. [OK]をクリックする。

今回の問い

- 度数分布表とは何か？
 - データをいくつかの階級に分け、それぞれの階級に入る観測値がいくつか数えて度数で表した表。
- ヒストグラムを描くと何がわかるか？
 - 分布の中心位置や散らばり具合などの特徴を知ることができる。

今回学んだこと

1. データとは
2. 度数分布表
3. ヒストグラム

今回の目標

度数分布表を用いてヒストグラムを描けるようになる。

課題

dataset1.csvの「体重」のデータから度数分布表を作成し、ヒストグラムを描きなさい。作成したファイルはODF表計算ドキュメント（拡張子.ods）で提出すること。

提出先：yutaka.hirai@koeki-u.ac.jp

件名：データリテラシー第4回課題 [学籍番号]

締め切り：10月30日 (金)

次回

第1回	社会におけるデータ・AI利活用
第2回	表計算ソフトの基本操作
第3回	セルの参照と集計
第4回	度数分布表とヒストグラム
第5回	代表値と散らばりの尺度
第6回	データのグラフによる表現とデータ比較
第7回	データの分布と標本抽出
第8回	推定
第9回	分割表とクロス集計
第10回	2変量データと相関
第11回	時系列データ
第12回	仮説検定
第13回	回帰分析