岐阜数学教育研究
2016, Vol. 15, 1-10

視覚的に確率密度関数の理解を促す授業実践

赤堀　克己[1]

高等学校の数学 B の「確率分布と統計的な推測」の章において，離散型から連続型へ確率変数の定義を拡張する際に，「確率密度関数はヒストグラムの階級の幅をどんどん狭くしていくとき，ヒストグラムの形が近づいていく曲線」と説明している。しかし，この記述は極限の概念を含んでおり，学習者が頭の中でイメージするのは非常に困難である。そこで，我々は Excel を用いて学習者自身がヒストグラムを描き，視覚的に理解を促す授業実践を試みた。本論文では，大学１年生を対象とした視覚的に確率密度関数の理解を促す教材とそれを用いた授業実践の結果を報告する。

＜キーワード＞　確率密度関数，ヒストグラム，連続型確率変数，離散型確率変数

１．はじめに

統計学の学習過程において，数学の知識や概念を利用する場面は多い。導入として，敢えて数式の利用を避けて統計学の内容を紹介するのは有益であろうが，学習者が深い理解を得る為には，数学的素養は不可欠であると筆者は感じている。たとえば，現行の高等学校の「数学Ｂ」の教科書において，「中心極限定理」や「大数の法則」は，次の様に記述されている。

[中心極限定理(標本平均の分布)]

母平均m，母標準偏差σの母集団から大きさｎの無作為標本を抽出するとき，標本平均Ｘは，ｎが十分大きいとき，近似的に正規分布Ｎ(m，σ^2/n)に従う。

[大数の法則]

母平均ｍの母集団から大きさｎの無作為標本を抽出するとき，その標本平均Ｘは，ｎが大きくなるに従って，母平均ｍに近づく。

これらの内容を十分理解するためには，「無限大(∞)」や「極限(n → ∞)」といった概念に慣れ親しんでいることが必要である。

しかしながら，現行の高等学校の学習過程において，「無限大(∞)」や「極限(n → ∞)」を学習するのは「数学Ⅲ」であり，統計的な内容(「確率分布と統計的な推測」)を含む「数学Ｂ」のあとに配置されている。上記の要約で述べたように，本実践のテーマである「確率密度関数」の説明においても，「極限」の概念の理解が必要である。では，数学的素養が十分でない場合，統計学をどのように展開するのがよいだろうか？我々は一つの試みとして，Excel で作成したグラフ(今回はヒストグラム)を用いた視覚的に理解を促す教材を作成し，大学１年生に対して実践を行った。以下，実践内容と受講者に対して行ったアンケート結果を報告する。

[1] 岐阜薬科大学

２．研究のねらい

２．１．高等学校での履修状況

実社会において, 統計学の素養が求められる機会が多いにも関わらず, 高等学校での統計的内容の履修率は高くない。今年度の受講者に対して行った別の回のアンケートの結果([1])においても, 新課程履修者の数学 B の「確率分布と統計的な推測」の履修率は２割程度であった。

２．２．「連続型確率変数」と「確率密度関数」の重要性

「確率密度関数はヒストグラムの階級の幅をどんどん狭くしていくとき, ヒストグラムの形が近づいていく曲線」という説明において, ヒストグラムの極限はヒストグラムではなく, 「曲線」であることの把握は非常に重要である。

高等学校「数学Ｉ」の２次関数の学習の際にグラフの移動：

「関数 y=f(x)のグラフＦを

ｘ軸方向にｐ， ｙ軸方向にｑ

だけ平行移動して得られる曲線をＧとするとき， Ｇの方程式は

$$y-q=f(x-p)$$

である。」

を学ぶ。ここで, グラフの移動と関数の対応を学習する。このことをよく理解できている学習者は, 統計においても, 変数変換と確率密度関数のグラフの移動の対応も容易に理解できる。しかし, ヒストグラムでは, 変数変換が一般には平行移動に対応しない。例えば, テストの点を確率変数Ｘとして１６人分のデータが

62,76,42,55,96,88,35,63,82,48,51,58,61,84,69,76

であったとする。このデータを素点としてテストの点を１０点かさ上げした点を確率変数Ｙ＝Ｘ＋１０とすると, Ｙのデータは,

72,86,52,65,106,98,45,73,92,58,61,68,71,94,79,86

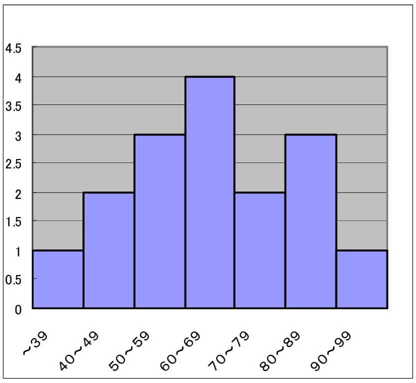
となる。同様に, テストの点を５点かさ上げした場合を考える。Ｚ＝Ｘ＋５とすると, Ｚのデータは,

67,81,47,60,101,93,40,68,87,53,56,63,66,89,74,81

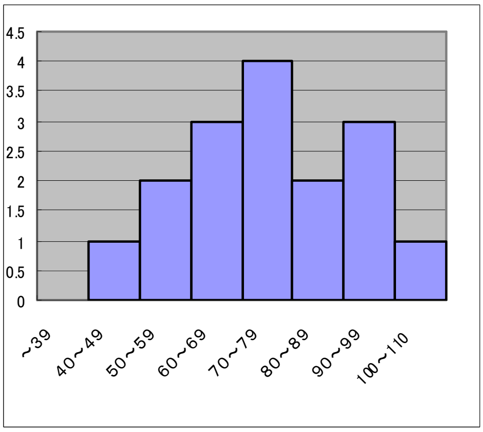
となる。Ｘ，Ｙ，Ｚの度数分布表は[表１]であり，Ｘ，Ｙ，Ｚのヒストグラムはそれぞれ[図１]，[図２]，[図３]である。

階級	Ｘの度数	Ｙの度数	Ｚの度数
～39	1	0	0
40～49	2	1	2
50～59	3	2	2
60～69	4	3	5
70～79	2	4	1
80～89	3	2	4
90～99	1	3	1
100～109		1	1

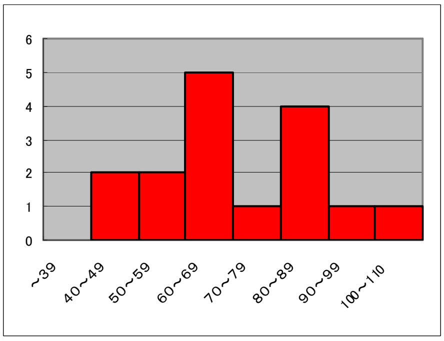
[表１]



[図１]



[図２]



[図３]

Ｙのヒストグラムは，Ｘのヒストグラムを横軸方向に１０だけ平行移動したものになっているが，Ｚのヒストグラムは，Ｘのヒストグラムを横軸方向に５だけ平行移動したものになっていない。

高等学校「数学Ｂ」の統計の学習過程をみると，

「連続型確率変数と確率密度関数」

↓①

「平均がｍで分散がσ^2の正規分布Ｎ(ｍ，σ^2)の説明」

↓②

「標準正規分布Ｎ(０，１)の説明」

と展開していく。②において，変数変換Ｚ＝(Ｘ－ｍ)/σを行う。ＸがＮ(ｍ，σ^2)に従うとき，ＺがＮ(０，１)に従うことの厳密な証明は置換積分によるが，正規分布のグラフのピークの移動については，先に述べたグラフの移動と関数の対応を理解していれば容易であり，標準正規分布のグラフの理解の強力な助けとなる。標準正規分布のグラフの理解が不十分な場合，それ以降の統計学の学習に大きな支障を来たしてしまう。

２．３．学習者の理解を困難にする要因

高等学校「数学Ｂ」において，確率変数Ｘの例として，サイコロを n 回振ったときに１の目が出た回数など「離散型」を始めに学習する。「離散型」の場合の確率分布（表）は，Ｘ＝k となる確率Ｐ(X=k)をとりうる全ての k の値に調べて表にすればよいだけなので，学習者は違和感をいだかない。しかし，確率変数が「連続型」の場合，「離散型」の場合と同様に，Ｘ＝k となる確率Ｐ(X=k)を求めようとしても，Ｐ(X=k)＝１/∞≒０となってしまい，意味を成さない。よって，学習者は，１点での確率を考えるのではなく，区間での確率Ｐ(a≦x≦b)を考えるという発想の転換を求

められる。このことが，学習者の理解に困難をもたらす一つの要因と考えられる。

また，「確率密度関数はヒストグラムの階級の幅をどんどん狭くしていくとき，ヒストグラムの形が近づいていく曲線」という説明における極限の操作は，高等学校「数学Ⅲ」の「区分求積法」の極限の操作とも異なり，階級の幅を変化させるごとにヒストグラムの形が変化していくので非常にイメージがし難い(「区分求積法」では，考えている関数のグラフは変化しない)。つまり，同様の極限の操作は，数学の学習過程では出会わないのである。ゆえに，学習者の理解が容易でないのは，至極当然である。よって，「確率密度関数」は，丁寧に扱うべき項目であると筆者は感じている。

２．４．学習効果の評価

「データの大きさを増し，階級の幅も狭くしていくと，ヒストグラムの形は次第に１つの曲線に近づいていく。」という表現の理解度と「確率密度関数」の理解の深まりについて，アンケートを実施し，学習効果を数値化して評価を試みた。

３．実践内容

講義名：「情報処理科学」

実践日：平成 27 年 10 月 26 日(月)

場所:岐阜薬科大学三田洞学舎村山情報処理センター

対象：大学１年生（有効回答数５５名）

３．１．受講者について

今回の受講者のほとんどが高等学校の「情報」の授業で Excel を経験しており，さらに前期の「情報基礎実習」でも Excel の操作方法を復習済みである。また，前回までに，Excel の FREQUENCY 関数の操作方法も学習済みである。

３．２．時計を題材にした導入と Excel を用いたヒストグラムの作成

まず，時計を題材にした教材を用いて，確率変数Ｘが連続型の場合の導入を試みた。[2]の p.40 において，時計を題材にして確率密度関数を説明してあるが，離散型と連続型の対比をデジタル時計とアナログ時計という身近な例を用いているため，受講者は大変受け入れやすい。本実践では，その内容を設問形式にアレンジした教材を用いて，段階的な違和感のない確率変数Ｘの概念の拡張を促した。

次に，受講者に Excel でヒストグラムを３つ以上作成してもらい，確率密度関数のイメージを視覚的に確認してもらった。以下は，受講者に配布した教材である。

[時計を題材にした確率密度関数の導入]

(デジタル時計とアナログ時計)

(Ⅰ) 時刻Ｘ＝０，１，２，…，１１を表示するデジタル時計(確率変数Ｘが離散型)

問　確率変数Ｘについて，Ｘ＝ｘとなる確率をＰ(x)とするとき，Ｐ(x)を求めよ。

(Ⅱ) 短針のみのアナログ時計(確率変数Ｘが連続型)

(ただし，短針はスムーズに動くものとする。)

問１　短針の指す値を確率変数Ｘとすると，Ｘの取り得る値は何通りあるか？

問２　Ｘ＝１となる確率を求めよ。

問３　０≦Ｘ≦６となる確率を求めよ。

問４　６≦Ｘ≦１０となる確率を求めよ。

問５　次の空欄を埋めよ。

一般に，連続型の確率は（　①　）で求める。すなわち，a≦Ｘ≦b となる確率をＰ(a≦Ｘ≦ｂ)とすると，

$$P(a \leqq X \leqq b) = \int_a^b f(x)dx$$

であり，この式の f(x)を（　　②　　）という。

アナログ時計の短針は一定速度で進むので，f(x) = k(一定)とおけ，

$$\int_0^{12} f(x)dx = (\ ③\ )\text{より}\quad k = (\ ④\ )$$

よって，$f(x) = \begin{cases} (\ ⑤\ )\ (0 \leqq x < 12\text{のとき}) \\ (\ ⑥\ )\ (0 < x\text{ または }x \geqq 12\text{のとき}) \end{cases}$

(解答例)

(Ⅰ)　$P(x) = \frac{1}{12}$　(x = 0, 1, 2,⋯, 11 のとき)

(Ⅱ)　問１　∞通り　問２　$\frac{1}{\infty} \fallingdotseq 0$　問３　$\frac{1}{2}$　　問４　$\frac{1}{3}$

問５　①面積　②確率密度関数　③１　④$\frac{1}{12}$　⑤$\frac{1}{12}$　⑥０

[EXCEL による実習]

（操作手順）

① A1 に「データ配列」と入力し，A 列に平均０，分散１のある分布に従うデータを１００００個用意する（A2 に＝NORMSINV(RAND（）)と入力してオートフィルする)。

② B１に「区間配列（階級)」と入力し、B 列に－2.5 から１きざみで 3.5 まで値を入力する(オートファイルを利用)。

③ C1 に「度数」と入力する。

④ C２にカーソルを移動し,「数式」→「その他の関数」→「統計」から「FREQUENCY」を選択し，データ配列，区間配列を入力し「OK」をクリック。

⑤ C２のセルをコピーし，C８までドラッグして貼り付ける。

⑥ 数式バーの表示の最後の「)」の右をクリックして，カーソルを移動させる。

⑦ 「Ctrl」+ 「shift 」を押しながら「Enter 」を押す。

⑧ C 列に度数が表示される。

⑨ C9 に度数の合計を求める。

⑩ D1 に「ヒストグラム１」と入力し，D 列で相対度数を求める。

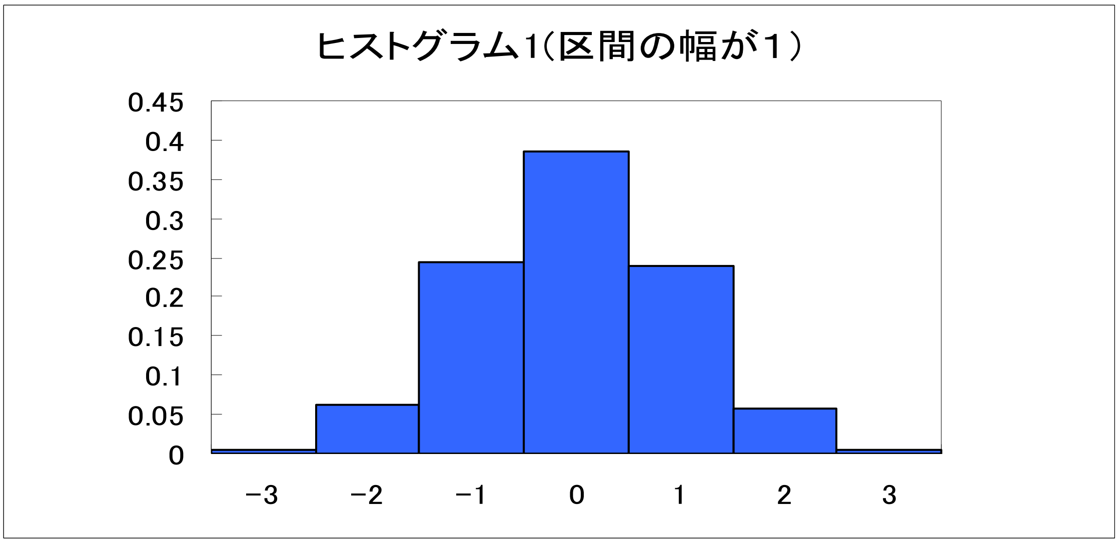
⑪ D9 で相対度数の合計を求める（１になるか確認！)。

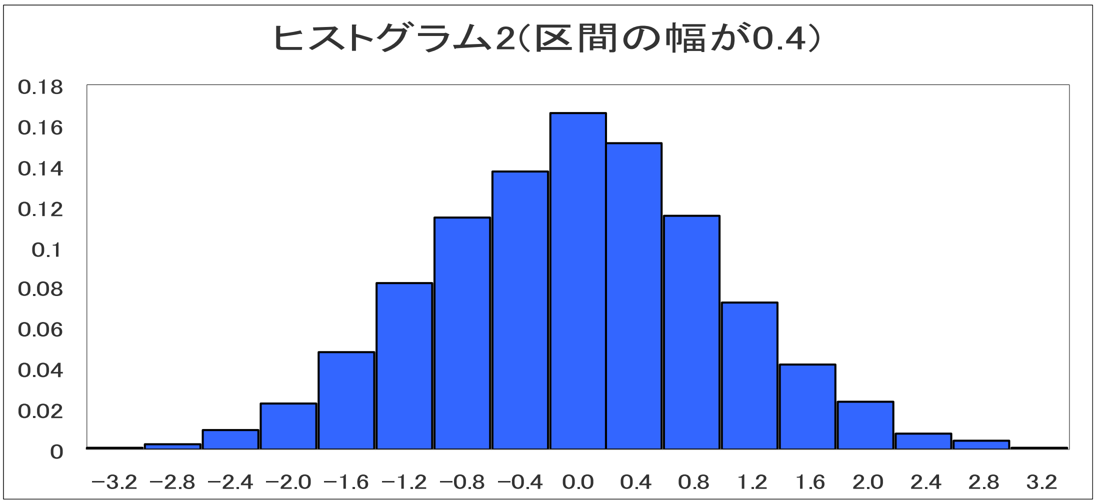
⑫ F1 に「階級値」と入力する。

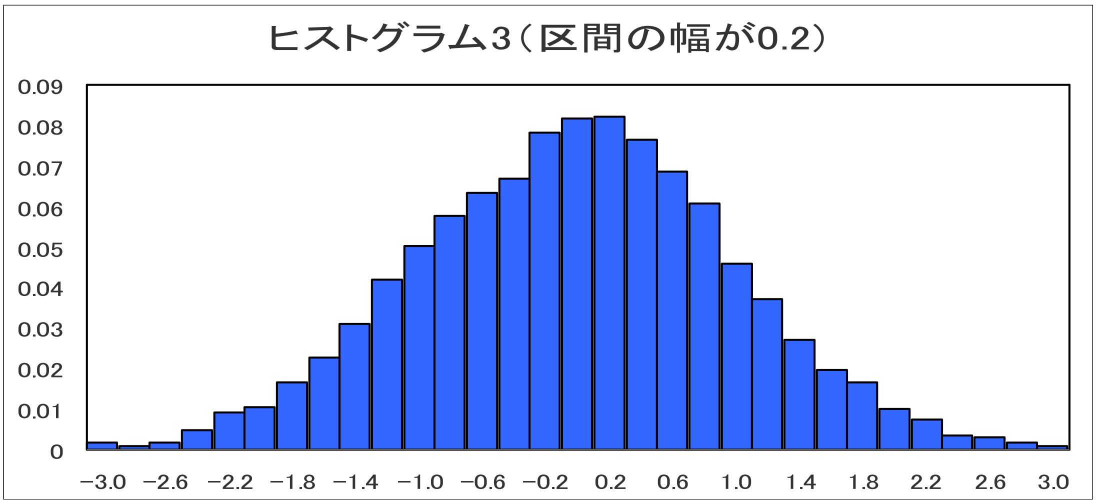
⑬　F2に＝FIXED(B2－0.5,0,TRUE) と入力してF8までオートフィルする。
注))FIXEDは数値を文字列に変換する関数。今，階級の幅が1なので（各階級の最大値）－0.5 で階級値（各階級の中央値）を求めている。2 番目の引数が「１」だと小数点1桁，「２」と小数点2桁でそれぞれ出力される。
⑭　G1に「ヒストグラム１」と入力し，G列に相対度数の値を入れて，度数分布表を完成させる（G2に＝D2と入力して、D8までオートフィルする)。
⑮　度数分布表を範囲指定して，縦棒でグラフを描く。
⑯　カーソルを棒グラフの棒にあわせて「右クリック」する。
⑰　「データ系列の書式」を選択する。
⑱　系列のオプションで「要素の間隔」を「０」にする。
⑲　区間の幅を狭くして，同様にヒストグラムを描く。
⑳　さらに区間の幅を狭くして同様にヒストグラムを描き，３つのヒストグラムの変化の様子を観察する。

(出力結果)

階級値	ヒストグラム 1
−3	0.00590059
−2	0.062406241
−1	0.238723872
0	0.380338034
1	0.245924592
2	0.062006201
3	0.00470047
計	1







３．３．教材における留意点

（時計の例題）

問１でまず, 円周上には無限個の実数が存在し, 短針が指す値は無限通りあることを確認させる。

問２において，P(X=１)＝１/∞≒０であり，同様に任意のｋについて，P(X=k)≒０となってしまい, 意味を成さないことを感じさせる。

問３では, 問５の連続型の場合の確率の定義を天下り的に説明するのではなく，区間を設定したら, 確率が自然と考えられることを学習者に感じさせる。また，区間を半円周に対応するように設定し, 確率の計算での煩雑さを感じないようにした。

問４は，問３で体感した，区間で確率を考えるという発想を, 半円周以外の別の区間でも試してみるための練習問題である。

問５は，実際に確率密度関数を求めて，離散型との違いを受講者が確認できるようにした。

（Excel での実習）

配布教材の(操作手順)①で，標準正規分布に従うデータを出力させている。しかし,

一連の授業の中で, 正規分布はこの時点では未習であるため, 「ある分布」とだけ記載し, 詳細な解説は敢えて省いた。

４．実践における受講者の活動の様子

４．１．時計の例題

各自解答する時間を設けた後, １問１問受講者に質問しながら解説をした。受講者の集中度は非常に高く, 理解が深まった様子であった。

４．２．Excel による実習

１つ目のヒストグラムを描くまでの過程では, Excel の操作方法に関する質問はほとんど出なかった。しかし, ２つ目以降の作成過程で, 区間の幅を極端に狭くするとデータ数が十分でなくなるため, ヒストグラムの形がギザギザになり, 戸惑っている受講者がいた。

５．アンケート結果と考察

授業の前後で「確率密度関数」の理解度を確認するために以下のアンケートを実施した。

Ｑ１．今現在(講義開始時), 「確率密度関数」について

１．全然理解できていない（理解度１０％未満)。
２．少ししか理解出来ていない（理解度１０％以上４０％未満)。
３．半分ぐらいは理解している（理解度４０％以上６０％未満)。
４．大体理解している（理解度６０％以上９０％未満)。
５．完璧に理解している（理解度９０％以上)。

Ｑ２．プリントの「データの大きさを増し, 階級の幅も狭くしていくと, ヒストグラムの形は次第に１つの曲線に近づいていく。」という表現は

１．全然イメージ出来ない。
２．あまりイメージ出来ない。
３．わかったような, わからないような感じ。
４．まあまあイメージ出来る。
５．完璧にイメージ出来る。

Ｑ３．Excel の実習でグラフを見た後, プリントの「データの大きさを増し, 階級の幅も狭くしていくと, ヒストグラムの形は次第に１つの曲線に近づいていく。」という表現は

注)）Ｑ２で選んだ回答番号以上の番号を選んでください。

１．全然イメージ出来ない。
２．あまりイメージ出来ない。
３．わかったような, わからないような感じ。
４．まあまあイメージ出来る。
５．完璧にイメージ出来る。

Ｑ４．今現在(講義終了時), 「確率密度関数」について

１．全然理解できていない（理解度１０％未満)。
２．少ししか理解出来ていない（理解度１０％以上４０％未満)。
３．半分ぐらいは理解している（理解度４０％以上６０％未満)。
４．大体理解している（理解度６０％

以上９０％未満)。

５．完璧に理解している（理解度９０％以上)。

Ｑ５. 今日の講義の感想を何か書いてください。

Ｑ１からＱ４の各問に対する回答の集計結果が[表１]である。

回答番号	Ｑ１	Ｑ２	Ｑ３	Ｑ４
1	22 人	9 人	1 人	1 人
2	12 人	12 人	4 人	4 人
3	16 人	15 人	11 人	27 人
4	4 人	16 人	27 人	16 人
5	1 人	3 人	12 人	7 人
合計	55 人	55 人	55 人	55 人

[表１]

さらに，Excel で作成したグラフによる視覚的効果と「確率密度関数」の理解の深まりを見る指標として，回答番号の差（Ｑ３－Ｑ２）と（Ｑ４－Ｑ１）についても調べた。[表２]がその集計結果である。

回答番号の差	Ｑ3－Ｑ2	Ｑ4－Ｑ1
0 point	21 人	15 人
1 point	22 人	17 人
2 point	6 人	15 人
3 point	5 人	5 人
4 point	1 人	3 人
合計	55 人	55 人
効果あり	34 人(62%)	40 人(73%)

[表２]

[表２]より，Excel で作成したグラフの視覚的効果は６２％の受講者において確認できた。また,「確率密度関数」の理解についても７３％の受講者が, 理解が深まったと回答している。

次に, Ｑ５の感想の代表的なものを紹介する。

①アナログ時計とデジタル時計の例はとてもわかりやすかったです。感動しました。
②確率を表すのに積分の考え方が使われていることが分かり，驚きました。
③グラフを自分でつくったので, ヒストグラムが曲線に近づいていく様子を視覚的に捉えることができ, イメージしやすかったです。
④最後のグラフができたときは, 少し感動しました。
⑤曲線が見えました!!
⑥階級の幅をかなり小さくしたら, グラフがギザギザになった。なめらかにするには標本の数を多くしなければならない。
⑦ヒストグラムをつくるのは楽しかったので，何かのデータを考察するときに，また自分でやってみたい。
⑧グラフを３個描く(同じ操作)が大変でした。

６．今後の課題

６．１．ヒストグラムの作成方法

Excel には「分析ツール」が用意されている。それを利用するとヒストグラムの作成自体は大変容易になるが, ヒストグラムの作成の過程が全く見えなくなる。今回の実践では，作業過程を確認できるよう **FREQUENCY** 関数を用いて各階級の度数を求め，相対度数を棒グラフで描き，棒の間隔を０にしてヒストグラムを作成し

た。どちらの方が，より学習効果が高いかは今後の課題と捉えている。

６．２．実習時間

本実践の授業時間は９０分間であり，時間的余裕があったので，３つのヒストグラム全てを受講者各自で作成してもらった。時間が十分とれない場合は，クラスを３つのグループに分けて作業した後，指導者がファイルをまとめてパワーポイントのスライドショーで全員に見せる方法なども考えられる。

引用・参考文献

[1]赤堀克己，2015，統計学における手計算の学習効果について，岐阜数学教育研究，Vol. 14, 7-13

[2]馬場敬之・久池井茂，2003，確率統計，マセマ出版社