

論文：

NDBオープンデータを利用するための教材作成の試み

畔津 忠博¹⁾, 増成 直美²⁾, 水津 久美子²⁾

An Attempt of Preparation of Teaching Materials to Utilize NDB Open Data

AZETSU Tadahiro¹⁾, MASUNARI Naomi²⁾, SUIZU Kumiko²⁾

要旨：本稿では、NDBオープンデータを効率的に処理するための教材作成の試みについて報告する。教材は、NDBオープンデータの説明と表計算ソフトのExcelで利用できるプログラミング言語であるVisual Basic for Applications(VBA)を用いたデータ処理の解説から構成される。また、作成した教材では、Excelのワークシート上での作業と比較しながらVBAを用いる利点について理解してもらうように配慮した。最後にVBAに関連した問題を解いてもらい、教材の理解度を確認した。

Abstract: In this article, we describe an attempt to prepare teaching materials to efficiently process NDB open data. The teaching materials consist of explanations of NDB open data and instructions of data processing using Visual Basic for Applications (VBA), a programming language that can be used with the spreadsheet software Excel. In addition, the teaching materials allow students to understand the advantages of using VBA by comparing it to working with Excel worksheets. Finally, we asked them to solve VBA-related problems and confirmed their understanding of the teaching materials.

Key words: NDB open data, teaching material preparation, Visual Basic for Applications
キーワード: NDBオープンデータ, 教材作成, VBA

1 はじめに

NDBオープンデータとは、厚生労働省が公開しているレセプト情報、特定健診情報のことである^[1]。今回は、第5回公表データのうち、特定健診情報（平成29年度分）を題材とする。ここでは、「BMI」、「AST」、「ALT」、「HbA1c」、「HDLコレステロール」、「LDLコレステロール」、「 γ -GT」、「ヘモグロビン」、「拡張期血圧」、「眼底検査」、「空腹時血糖」、「収縮期血圧」、「中性脂肪」、「尿蛋白」、「尿糖」、「腹囲」が公開されている。

本稿は、このデータを効率的に処理するための

教材作成の試みについて報告する。NDBオープンデータは表計算ソフトのExcel形式のデータとして公開されているので、データ処理はExcelを利用することが直接的で有効である。また、Excelでは、プログラミング言語の1つであるVisual Basicが利用可能である。Excel内で利用できるVisual Basicは、アプリケーションソフトウェア用に改良されたもので、Visual Basic for Applications(VBA)と呼ばれている^[2,3]。このVBAを利用すると、さらに効率的な処理が期待される。

VBAを短時間で学ぶことが教材の主目的である

1) 国際文化学部文化創造学科

2) 看護栄養学部栄養学科

が、単にプログラミング言語を学ぶだけではなく、NDBオープンデータの分析¹⁴⁾を行うという学習目標を明確にすることで、学習意欲も増すことが期待される。また、教材という形にすれば、多くの人が利用できるオープンデータの趣旨にも合うと考えられる。

2 作成した教材

NDBオープンデータの特典健診情報に関しては、一定のフォーマットに従ってデータが公開されている。そのため、VBAの基本的事項を理解するだけでも、多くのデータを短時間で有効に活用することができる。教材では、変数、配列、演算子、繰り返し処理、オブジェクトの操作といった今回のNDBオープンデータの処理に関連する内容を取り上げ、Excelのワークシートでの処理と比較することで、VBAの利点が理解できるようにした。

2.1 基本的事項の整理

まず、画面に文字を表示させるプログラムを作成することを目標として、エディタの使い方、プログラムの実行と修正、セキュリティレベルの設定などを説明する。

次に、データの記憶領域として変数、配列の説明を行う。このとき、表計算ソフトの学習の一環として、セルとの類似性を指摘する。また、値を変数に代入する方法として代入演算子について触れる。

さらに、配列を効果的に扱う方法と関連付けて、

プログラムの順番を制御する方法であるFor文またはWhile文を用いた繰り返し処理について説明する。また、制御命令のもう1つの重要な命令であるIf文またはSelect文を用いた条件分岐についても触れる。

2.2 オブジェクトの操作

他のプログラミング言語に対するVBAを用いる利点としては、Excelの様々な機能が利用できることが挙げられる。例えば、ブック、ワークシート、セルなどのオブジェクトを操作するための様々な命令があるが、特に、セルを参照するCells命令と、セル範囲を参照するRange命令を重点的に説明する。

また、合計を求めるSUM関数や平均を求めるAVERAGE関数など、多くのワークシート関数がVBAにおいても利用可能であることを説明し、さらに、関数では多くの場合、セル範囲を引数として指定する必要があるが、上記のRange命令とCells命令で指定することが可能であることを理解してもらう。また、ワークシート関数の機能にないような処理を行う方法として、ユーザー定義関数についても言及する。

2.3 NDBオープンデータへの応用

以上で学習したことを基に、VBAをNDBオープンデータに適用する方法について例を挙げる。図1は特典健診情報に関するNDBオープンデータの1つであり、都道府県、年齢ごとにクロス集計されたBMIデータである。年齢の範囲を40歳から74歳まで

A		B		C		D		E		F		G		H		I		J		K		L		M		N		O		P		Q		R	
1		特定健診(BMI):H29年度 ※集計結果が10未満の場合は「J」で表示(10未満の箇所が1箇所の場合は10以上の最小値を全て「J」で表示)																																	
3		検査値範囲 (kg/m ²)																																	
4		都道府県名																																	
5																																			
6																																			
7		300以上																																	
8		25以上30未満																																	
9		20以上25未満																																	
10		18.5以上20未満																																	
11		18.5未満																																	
12		300以上																																	
13		25以上30未満																																	
14		20以上25未満																																	
15		18.5以上20未満																																	
16		18.5未満																																	
17		300以上																																	
18		25以上30未満																																	
19		20以上25未満																																	
20		18.5以上20未満																																	
21		18.5未満																																	

図1. NDBオープンデータ(BMI)

A		B		C		D		E		F		G		H		I		J		K		L		M		N		O		P		Q		R	
1		特定健診(収縮期血圧):H29年度 ※集計結果が10未満の場合は「J」で表示(10未満の箇所が1箇所の場合は10以上の最小値を全て「J」で表示)																																	
3		検査値範囲 (mmHg)																																	
4		都道府県名																																	
5																																			
6																																			
7		180以上																																	
8		140以上180未満																																	
9		130以上140未満																																	
10		120以上130未満																																	
11		120未満																																	
12		180以上																																	
13		160以上180未満																																	
14		140以上160未満																																	
15		130以上140未満																																	
16		120以上130未満																																	
17		120未満																																	
18		180以上																																	
19		160以上180未満																																	
20		140以上160未満																																	
21		130以上140未満																																	
22		120以上130未満																																	
23		120未満																																	

図2. NDBオープンデータ(収縮期血圧)

とし男女別に5歳刻みで7階層に分けており、検査値であるBMIを5階層に分けている。他の収縮期血圧データ（図2）や空腹時血糖などでもフォーマットはほぼ一緒で、検査値である収縮期血圧や空腹時血糖の階層数が異なるだけである。そのため、BMIデータを処理するプログラムを作成すれば、他のデータにもほぼそのまま適用できる。

まず、今回のデータ処理の目的である年齢調整の流れを説明する。最初に、BMIデータにおいて、健康寿命に悪影響を及ぼすと考えられる検査値の範囲を設定し、その人数をリスク人口として算出する。例えば、BMIが25以上をリスクと考えたとき、北海道の男性の40～44歳を対象とすると、リスク人口は、 $28755+8934=37689$ となる。また、リスク割合をリスク人口/全体の人口とすると、 $(28755+8934)/(2533+6156+52188+28755+8934)=0.382373232\cdots$ となる。

次に、NDBオープンデータが採用している年齢区分における日本の人口を図3に示す。このデータでは、40-44歳の人口は4914018となっているが、これにさきほどのリスク割合をかけたものを調整リスク人口とすると、 $4914018 \times 0.382373232\cdots = 1878988.94\cdots$ となる。これを全年齢区分に適用した結果が図4となる。

	A	B	C	D	E	F	G
1		BMI			収縮期血圧		
2		男	女	計	男	女	計
3	北海道	0.382017	0.221907	0.300733	0.226146	0.159827	0.192478
4	青森県	0.371692	0.252394	0.311128	0.20406	0.144158	0.17365
5	岩手県	0.371198	0.250283	0.309813	0.223364	0.167043	0.194771
6	宮城県	0.369857	0.232416	0.300082	0.215368	0.154752	0.184595
7	秋田県	0.363783	0.245596	0.303782	0.255082	0.179567	0.216745
8	山形県	0.342412	0.232926	0.286829	0.231462	0.166328	0.198395
9	福島県	0.370905	0.252422	0.310754	0.211106	0.156514	0.183391
10	茨城県	0.355955	0.225372	0.289662	0.198933	0.147842	0.173488
11	栃木県	0.346791	0.226218	0.285579	0.219492	0.162115	0.190363
12	群馬県	0.330579	0.213484	0.271133	0.219932	0.168714	0.19393
13	埼玉県	0.339293	0.20493	0.271108	0.207666	0.156416	0.181648
14	千葉県	0.351573	0.205232	0.277279	0.194674	0.14887	0.17142
15	東京都	0.333197	0.182379	0.256631	0.181471	0.129362	0.155016
16	神奈川県	0.330205	0.181614	0.25477	0.18297	0.133549	0.15788
17	新潟県	0.302395	0.191435	0.246063	0.203687	0.138851	0.170771
18	富山県	0.328332	0.194207	0.26024	0.197809	0.145402	0.171204
19	石川県	0.324268	0.192147	0.257194	0.213053	0.151387	0.181747
20	福井県	0.325296	0.192762	0.258012	0.225602	0.162873	0.193756

図5. VBAを用いた結果

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1		男																	
2			40～44歳	45～49歳	50～54歳	55～59歳	60～64歳	65～69歳	70～74歳	中計	40～44歳	45～49歳	50～54歳	55～59歳	60～64歳	65～69歳	70～74歳	中計	
3		日本人人口	4914018	4354877	3968311	3729523	3729523	4151119	4659662	3582440	29359950	4818200	4307927	3961985	3785723	4303891	4984205	4113371	30275302

図3. 日本の基準人口

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	
1	特定健診2019～2022年度 ※集計結果が10未満の場合は「-」で表示(10未満の箇所が1箇所の場合は10以上の最小値を全て「-」で表示)		全体																				
2			男																	女			
3			男																	女			
4	都道府県名		40～44歳	45～49歳	50～54歳	55～59歳	60～64歳	65～69歳	70～74歳	中計	40～44歳	45～49歳	50～54歳	55～59歳	60～64歳	65～69歳	70～74歳	中計	全調整リスク人口				
5			人数	人数	人数	人数	人数	人数	人数	人数	人数	人数	人数	人数	人数	人数	人数	人数	男	女	計		
6	リスク割合		0.382373	0.406477	0.409343	0.391628	0.384557	0.363385	0.336613	0.386266	0.193412	0.212953	0.213234	0.217192	0.225652	0.239693	0.252007	0.220741					
7	日本人人口		4914018	4354877	3968311	3729523	4151119	4659662	3582440	29359950	4818200	4307927	3961985	3785723	4303891	4984205	4113371	30275302	28059950	30275302	59635252		
8	調整リスク人口		1878989	1770159	1610909	1469595	1596349	1693113	1205997	11340743	9318963	917386	8448307	8222306	9711907	1194179	1036500	6682992	11215596	6718027	17934290		
9																		0.3820169	0.221907	0.3007332	全調整リスク割合		
10																							
11	リスク割合		0.402314	0.386988	0.384724	0.370148	0.360127	0.342874	0.32414	0.374374	0.222441	0.240282	0.244788	0.244567	0.259723	0.270085	0.285591	0.25319					
12	日本人人口		4914018	4354877	3968311	3729523	4151119	4659662	3582440	29359950	4818200	4307927	3961985	3785723	4303891	4984205	4113371	30275302	28059950	30275302	59635252		
13	調整リスク人口		1976939	1737544	1566307	1300476	1484920	1595345	1161212	10991602	1071764	1035117	9606465	9259152	1117019	1346150	1174741	7695404	10912072	76413068	18554179		
14																		0.3716925	0.2523941	0.3111277	全調整リスク割合		
15																							

図4. 表計算ソフトによるNDBオープンデータの処理の例

すべての年齢区分で調整リスク人口を合計したものを全調整リスク人口とし、全日本人人口も同様にすべての年齢区分で日本の人口を合計したものと定義する。また、全調整リスク割合を下記3つのもので定義する。

- 全調整リスク割合(男) = 全調整リスク人口(男) / 全日本人人口(男)
- 全調整リスク割合(女) = 全調整リスク人口(女) / 全日本人人口(女)
- 全調整リスク割合(男女計) = 全調整リスク人口(男女計) / 全日本人人口(男女計)

この3つを計算したものを図4の右側に示す。

このような計算はExcelの数式で簡単に行うことができるが、都道府県ごとに同じことを何度も繰り返す必要がある。実際にExcelのワークシート上でこの作業を行ってもらい作業量を確認した上で、この操作をVBAにより記述することで負担が軽減されることを理解してもらおうとする。教材では、全調整リスク割合を都道府県ごとに求めるプログラムの作成を教材の最終目標としている。プログラムから得られる結果を図5に示す。

2.4 VBAとLibreOffice Basicとの比較

Excelを含むMS-Officeは、日常的な業務でよく用いられる文書作成ソフトや表計算ソフトなど複数のアプリケーションソフトが統合されたものであり、オフィススイートと呼ばれている。これと類似した機能をもつソフトとして、オープンソースのLibreOfficeがある^[7]。無料で使用できるため、遠隔授業等で個人のパソコンしか使用できない状況でも、インストール方法さえ指示すれば全員がソフトウェアを利用することができる。

また、LibreOfficeで利用できるプログラミング言語としてLibreOffice Basicがあり、LibreOfficeの表計算ソフトであるCalcにおいてVBAと同様の使い方ができる。本稿の2.1の基本事項の整理の内容は、そのままLibreOffice Basicでも当てはまり、そこで

問. NDB オープンデータの BMI の結果が示された下図において、北海道の 40～44 歳の BMI が 30 以上の人数 8934 が格納されているセルの場所を、Cells 命令の Cells(1,1)や Cells(1,2)などの表現を用いて、下記の選択肢から答えること。

	A	B	C	D	E
1	特定保健調査(平成19年度) ※集計結果が未済の場合は「」で表す(0未満の箇所)				
2					
3	都道府県名	検査値範囲 (kg/m ²)	40～44歳	45～49歳	50～54歳
4					
5			A数	A数	A数
6	北海道	30以上	8934	9188	6887
7		25.0以上30.0未満	98755	81189	97649
8		20.0以上25.0未満	52188	51186	47547
9		18.5以上20.0未満	6136	6481	7516
10		18未満	2533	2122	1111
11	青森県	30以上	9441	9068	8779
12		25.0以上30.0未満	7190	7570	7287
13		20.0以上25.0未満	2322	2783	11793

1つ選択

Cells(6,3) Cells(5,2) Cells(2,5) Cells(7,3) Cells(3,6)

図6. 問題A

問. BMI_3 のプログラムでは、[14]の For で都道府県の繰り返し処理を行っているが、これに対応する Next はどれか。下記の選択肢から行頭の番号を選択すること。

```

Sub BMI_3
[1] Dim i As Integer, j As Integer
[2] Dim total As Double, risk As Double
[3] Dim sum As Double, sum_m As Double, sum_f As Double
[4] Dim sum_risk As Double, sum_risk_m As Double, sum_risk_f As Double
[5] Dim risk_p(15) As Double, population(15) As Double
[6] Dim ws1 As Worksheet, ws2 As Worksheet
[7] Set ws1 = ActiveWorkbook.Sheets("BMI")
[8] Set ws2 = ActiveWorkbook.Sheets("results")
[9] ws2.Select
[10] ws2.Cells(1, 2) = "BMI"
[11] ws2.Cells(2, 2) = "男"
[12] ws2.Cells(2, 3) = "女"
[13] ws2.Cells(2, 4) = "計"
[14] For j = 1 To 47
[15] ws2.Cells(j + 2, 1) = ws1.Cells(5 * (j - 1) + 6, 1)
[16] For i = 1 To 16
[17] total = WorksheetFunction.Sum(Range(ws1.Cells(5*(j-1)+6, i+2), ws1.Cells(5*(j-1)+6+4, i+2)))
[18] risk = WorksheetFunction.Sum(Range(ws1.Cells(5*(j-1)+6, i+2), ws1.Cells(5*(j-1)+6+1, i+2)))
[19] population(i - 1) = Worksheets("basic_data").Cells(3, i + 2)
[20] risk_p(i - 1) = risk / total * population(i - 1)
[21] Next
[22] sum = 0: sum_m = 0: sum_f = 0
[23] sum_risk = 0: sum_risk_m = 0: sum_risk_f = 0
[24] For i = 0 To 6
[25] sum = sum + population(i) + population(i + 8)
[26] sum_risk = sum_risk + risk_p(i) + risk_p(i + 8)
[27] sum_m = sum_m + population(i)
[28] sum_risk_m = sum_risk_m + risk_p(i)
[29] sum_f = sum_f + population(i + 8)
[30] sum_risk_f = sum_risk_f + risk_p(i + 8)
[31] Next
[32] ws2.Cells(j + 2, 2) = sum_risk_m / sum_m
[33] ws2.Cells(j + 2, 3) = sum_risk_f / sum_f
[34] ws2.Cells(j + 2, 4) = sum_risk / sum
[35] Next
End Sub
    
```

1つ選択

[21] [31] [35]

図7. 問題B

問. BMI_3 のプログラムで、リスク人口を計算しているところは[18]の `risk = WorksheetFunction.Sum(Range(ws1.Cells(5*(j-1)+6, i+2), ws1.Cells(5*(j-1)+6+1, i+2)))` である。ここでは Sum 関数の引数として Range 命令により BMI25 以上のセル範囲を指定している。これを BMI20 以上のセル範囲にしたときの命令に修正したものを下記の選択肢から 1 つ選ぶこと。

1 つ選択

<code>risk = WorksheetFunction.Sum(Range(ws1.Cells(5*(j-1)+6, i+2), ws1.Cells(5*(j-1)+6, i+2)))</code>
<code>risk = WorksheetFunction.Sum(Range(ws1.Cells(5*(j-1)+6, i+2), ws1.Cells(5*(j-1)+6+2, i+2)))</code>
<code>risk = WorksheetFunction.Sum(Range(ws1.Cells(5*(j-1)+6, i+2), ws1.Cells(5*(j-1)+6+3, i+2)))</code>
<code>risk = WorksheetFunction.Sum(Range(ws1.Cells(5*(j-1)+6, i+2), ws1.Cells(5*(j-1)+6+4, i+2)))</code>

図8. 問題C

問. BMI_3 のプログラムで、上記と同様に[18]のリスク人口を計算する命令を、今度は痩身傾向を調べるため、リスク人口を BMI20 未満に設定した場合に修正したものを下記の選択肢から 1 つ選ぶこと。

1 つ選択

<code>risk = WorksheetFunction.Sum(Range(ws1.Cells(5*(j-1)+6+2, i+2), ws1.Cells(5*(j-1)+6+3, i+2)))</code>
<code>risk = WorksheetFunction.Sum(Range(ws1.Cells(5*(j-1)+6+2, i+2), ws1.Cells(5*(j-1)+6+4, i+2)))</code>
<code>risk = WorksheetFunction.Sum(Range(ws1.Cells(5*(j-1)+6+3, i+2), ws1.Cells(5*(j-1)+6+4, i+2)))</code>
<code>risk = WorksheetFunction.Sum(Range(ws1.Cells(5*(j-1)+6+3, i+2), ws1.Cells(5*(j-1)+6+5, i+2)))</code>

図9. 問題D

作成したプログラムは、LibreOffice Basicにおいても変更せずにそのまま実行できる。ただし、2.2のオブジェクトの操作の内容は変更が必要であり、特にセルを参照する命令が異なる点に注意が必要である⁸⁾。

今回の教材はLibreOfficeへの対応は行っていないが、オープンデータとの親和性を考慮した場合にはLibreOfficeなどのオープンソースのソフトウェアの使用を前提に考えることも重要であるため、今後の課題として検討したい。

3 教材を用いた授業の実施

教材を用いて1コマ90分の授業を遠隔で行った。教材の理解度を確保するため、授業の終わりの15分程度で、VBAに関連した問題を解いてもらった。問題は10問用意したが、その中でNDBオープンデータの処理に直接関連のある4つの問題を図6から図9に示す。

問題AはCells命令によるセルの参照、問題BはFor文による2重ループ、問題Cと問題DはRange命令によるセル範囲の指定について、それぞれの理解度を確保するものである。

問題の回答者数は24人で、正解率は、問題Aで45.8%、問題Bで37.5%、問題Cで75.0%、問題Dで33.3%であった。問題の中で特に重要なものが問題Cであり、これが理解できればリスクとして扱う検査値の範囲を自由に変更できる。Range命令に関しては、授業で重点的に説明したので、正解率が高くなったと思われる。ただし、より応用的な問題である問題Dに関しては正解率が低く、改善が必要であ

ることもわかった。

また、授業後に感想を記入してもらった。その中で肯定的な意見として、

- BMIを25以上を選択するのも、20未満を選択するのもプログラムのセル選択の式を少し変更するだけでできてしまうというのが楽だなと体感することができた。

- Rangeは実際に利用してみても便利だと感じたので、1つひとつのデータを手入力するような場面があったら、プログラムを作成し一括することに挑戦してみたいと思った。

否定的な意見として、

- Excelのプログラムについて学習したが、自分が今まで見たことない数式が出てきてよく理解できない部分があった。もう少しゆっくり説明してほしい。

などがあった。また、教材についての感想もあり、その中の肯定的な意見として、

- 授業資料として配付された一連の解説付きのマニュアルが手元にあることで、いつでも内容を見返すことができるのでとても助かると思った。

- 授業中わからない点があったが、PDF資料のおかげで理解することができた。

教材についての否定的な意見として、

- はじめにExcelのプログラムを表示する部分で手間取ってしまったので、NDBデータの授業資料のように、写真などわかりやすい形でプログラムを表示するところまで説明して欲しい。

などがあった。今後の改善点として、取り入れていきたい。

4 まとめ

本稿では、NDBオープンデータを効率的に処理するためにVBAを利用する方法について学習する教材作成の試みについて報告した。教材を用いて授業を行った結果、短時間でVBAの基本的事項を説明することは一定程度可能であったが、1コマの授業ですべてを説明することの困難さも確認された。事前、事後学習などと組み合わせて、学習プランを設計する必要があることが示唆された。

参考文献

- [1] 厚生労働省 NDBオープンデータ, <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000177182.html> (最終閲覧日 2022年1月8日)
- [2] 加藤潔, 「Excel環境におけるVisual Basicプログラミング」, 共立出版, 1999年.
- [3] 縄田和満, 「Excel VBAによる統計データ解析入門」, 朝倉書店, 2000年.
- [4] 末永瑤葉, 周田紗里奈, 照屋裕菜, 中島海桜, 増成直美, 「NDBオープンデータを活用した平均寿命と健康寿命に関連する要因の解析」, 山口県立大学看護栄養学部紀要14号, pp.9-23, 2021年.
- [5] 兼重美沙季, 畔津忠博, 水津久美子, 寺田亜希, 増成直美, 「死亡率に関する要因についての検討—NDBオープンデータを用いた解析—」, 山口県立大学看護栄養学部紀要15号, 2022年.
- [6] 濱口裕太, 畔津忠博, 寺田亜希, 水津久美子, 増成直美, 「リアルワールドデータを活用した健康寿命に関連する要因の解析」, 山口県立大学看護栄養学部紀要15号, 2022年.
- [7] LibreOffice The Document Foundation, <https://ja.libreoffice.org/> (最終閲覧日 2022年1月8日)
- [8] 畔津忠博, 吉永敦征, 永崎研宣, 「LibreOfficeで学ぶ情報リテラシー」, 東京電機大学出版局, 2016年.