

結果表における機械判読可能な データ作成に関する表記方法

Ver.1.2

目次

第1章 総則.....	3 -
第1節 目的	3 -
第2節 適用範囲	3 -
第3節 適用時期	3 -
第4節 適用内容	3 -
第2章 機械判読可能なデータの作成	4 -
第1節 データ形式の留意点	4 -
□チェック項目1 ファイル形式は Excel か CSV となっているか	4 -
第2節 Excel ファイルによる結果表のレイアウト	4 -
第1項 データ・項目の取扱い	4 -
□チェック項目2-1 1セル1データとなっているか	4 -
□チェック項目2-2 数値データは数値属性とし、文字列を含まないこと	6 -
□チェック項目2-3 セルの結合をしていないか	9 -
□チェック項目2-4 スペースや改行等で体裁を整えていないか	11 -
□チェック項目2-5 項目名等を省略していないか	13 -
□チェック項目2-6 数式を使用している場合は、数値データに修正しているか	13 -
□チェック項目2-7 オブジェクトを使用していないか	14 -
□チェック項目2-8 データの単位を記載しているか	14 -
□チェック項目2-9 機種依存文字を使用していないか	15 -
□チェック項目2-10 e-Stat の時間軸コードの表記、西暦表記又は和暦に西暦の併記がされてい るか	16 -
□チェック項目2-11 地域コード又は地域名称が表記されているか	18 -
□チェック項目2-12 数値データの同一列内に特殊記号(秘匿等)が含まれる場合	19 -
第2項 表の構成の取扱い	20 -
□チェック項目3-1 データが分断されていないか	20 -
□チェック項目3-2 1シートに複数の表が掲載されていないか	23 -
第3節 CSV ファイルによる結果表のレイアウト	24 -
第1項 データ・項目の取扱い	24 -
□チェック項目4-1 変数行から始まり、次行からデータ入力されているか	24 -
□チェック項目4-2 1フィールド1データとなっているか	25 -
□チェック項目4-3 数値データは数値属性とし、文字列を含まないこと	26 -
□チェック項目4-4 スペースを使っていないか	28 -
□チェック項目4-5 1行1データで表現されているか	28 -
□チェック項目4-6 項目名等を省略していないか	28 -
□チェック項目4-7 データの単位を記載しているか	29 -
□チェック項目4-8 機種依存文字を使用していないか	30 -

□チェック項目4-9 e-Stat の時間軸コードの表記、西暦表記又は和暦に西暦の併記がされているか	- 30 -
□チェック項目4-10 地域コード又は地域名称が表記されているか	- 33 -
□チェック項目4-11 数値データの同一列内に特殊記号(秘匿等)が含まれる場合	- 34 -
□チェック項目4-12 各フィールドの値をダブルコーテーション(“)で囲んでいるか	- 35 -
□チェック項目4-13 データが分断されていないか	- 35 -
□チェック項目4-14 1ファイル内に変数とデータのセットが複数掲載されていないか	- 36 -

第1章 総則

第1節 目的

結果表における機械判読可能なデータ作成に関する表記方法(以下「本表記方法」という。)は、『機械判読可能な統計データを作成し、提供する』ために必要な考え方や実施方法を統一的な観点で明確化・具体化し、各府省が確実かつ円滑に実施できるよう策定するものである。

これにより、各府省による利便性の高いスプレッドシート・csv の作成・提供が促進され、統計利用者のデータの利活用、ひいては EBPM の推進に寄与することを目的とする。

第2節 適用範囲

本表記方法は、各府省が政府統計の総合窓口(e-Stat)に掲載する結果表(既にデータベース形式として対応している結果表を除く。)について適用するものとする。ただし、各府省において、必要に応じ、本表記方法によらない表記方法のものを並行して公表することを妨げるものではない。

第3節 適用時期

本表記方法は、令和3年1月1日以降に公表する結果表について適用するものとする。ただし、各府省のシステム更改、外注等により即時に対応できないものについては、準備でき次第対応するものとする。

第4節 適用内容

本表記方法では、政府統計の総合窓口(e-Stat)に統計データをスプレッドシート及び csv で掲載する場合の表記方法について、第2章以降で示している。

各府省は、本表記方法を基に機械判読可能な統計データの作成・提供に取り組むものとする。

第2章 機械判読可能なデータの作成

第1節 データ形式の留意点

□チェック項目1 ファイル形式はExcelかCSVとなっているか

統計データのファイル形式は、仕様が公開・標準化されているもの又は国際的に標準化されているフォーマットであること(データベース形式として公表している場合はその限りではない。)。具体的には、Excel(.xlsx、.xls)やCSV、XML、Open Document Format(.ods)が挙げられる。

また、結果表の解説や注釈、図が入った報道発表資料や報告書ベースのファイル(PDF や印刷することを目的として作成した表形式データ)で掲載する場合であっても、それとは別途、その表の基データを登録する。

第2節 Excel ファイルによる結果表のレイアウト

本節では、ファイル形式による結果表における留意点を以下に示す。統計データを作成する際に、本節で示すチェック項目を1つずつ踏まえ改善していくことで、PDF 形式による結果表や、印刷して報告書として提供することを主眼にした表形式データを、機械判読可能なデータとして利用できる結果表に改善することが可能となる。

第1項 データ・項目の取扱い

□チェック項目2-1 1セル1データとなっているか

1セルに複数のデータが入力されていると、計算や昇順・降順の並べ替え、コピーペーストやグラフ化等加工編集する場合に多くの手作業やプログラムの作成が必要となり、すぐにデータとして利用できないため、1セル1データの入力とすること。

例1

修正前

1セルに複数のデータが入力されている

	全国
仕入額	373（平成27年度）、434（平成28年度）、549（平成29年度）、638（平成30年度）、741（平成31年度）
出荷額	973（平成27年度）、1234（平成28年度）、1449（平成29年度）、1738（平成30年度）、1841（平成31年度）

修正後

1セル1データとして入力した状態

	全国	全国
	仕入額	出荷額
平成27年度	373	973
平成28年度	434	1234
平成29年度	549	1449
平成30年度	638	1738
平成31年度	741	1841

仕入れ額と出荷額のセルに複数のデータが入力されている。このような場合は、年度ごとに列・行を分け、各セルにデータを分離して入力する。

例2

修正前

1セルに複数のデータが入力されている

	企業等数	売上金額（費用総額）	
総計	1188389	391445	(355943)
A	50384	69565	(63883)
B	154138	50468	(45332)
C	86522	44098	(33210)
D	85983	22258	(20237)
E	580003	37652	(35016)
F	27456	15288	(14633)
G	119085	115499	(110735)
H	84818	36617	(32897)

修正後

1セル1データとして入力した状態

	企業等数	売上金額	費用総額
総計	1188389	391445	355943
A	50384	69565	63883
B	154138	50468	45332
C	86522	44098	33210
D	85983	22258	20237
E	580003	37652	35016
F	27456	15288	14633
G	119085	115499	110735
H	84818	36617	32897

売上金額と費用総額が1セルに入力されているため、例えば、売上金額に占める費用総額の割合を計算したい場合、修正前の場合、「（）」内の数字を分離してから、割り算を行う必要が生じる。

一方、修正後の場合では、費用総額列の値を売上金額列の値で割るだけでよく、単純な処理で計算が可能となる。

そのため、1セル1データの入力すること。

□チェック項目2-2 数値データは数値属性とし、文字列を含まないこと

数値データに、円、¥、kg、トなどやマイナス記号(▲)を文字列として入力すると、Excel では数値ではなく文字列として扱われてしまうため、関数等で計算ができなくなる(エラーとなる)ほか、昇順・降順等の並べ替えも正確にできない場合がある。

また、ヒトが見やすくするための工夫として、千円単位を示す「, (カンマ)」を文字列として直接入力している場合、関数によっては正確に計算できない場合があるほか、当該 Excel ファイルを csv 化すると思わぬ挙動を示す場合がある(csv はカンマでデータを区切るため)。

そのため、数値データは数値属性とし、文字列を含まないこと。

なお、数値データの同一列内に特殊記号(秘匿等)が含まれる場合は、例外処理を認めているため、[チェック項目2-12](#)を参照すること。

補足: Excel の場合「書式設定」により体裁を整えることは可能であるため、データは数値として扱いつつ、見た目の体裁を整えることができる。

ただし、書式設定を施した Excel を csv 化すると、思わぬ挙動を示す場合があるため、印刷を意識した報告書ベースの Excel にのみ使用する等、使い分ける必要がある。

例1 数値データ内に文字列が含まれる場合

修正前				修正後			
「円」、「▲(マイナス表記)」、「, (カンマ)」が文字列として入力されている				数値データを数値属性として入力した状態			
	単価	前回差分	生産台数		単価	前回差分	生産台数
サンプル1	10,000円	130	12,000	サンプル1	10030	130	12000
サンプル2	9,100円	▲200	29,000	サンプル2	9100	-200	29000
サンプル3	8,020円	▲350	37,000	サンプル3	8020	-350	37000
サンプル4	7,500円	500	43,000	サンプル4	7500	500	43000
SUM関数	0	630	0	SUM関数	34650	80	121000
+(加算演算)	#VALUE!	#VALUE!	121000	+(加算演算)	34650	80	121000
↑黄色の行は関数で合計を表示した例				↑黄色の行は関数で合計を表示した例			

「円」、「▲(マイナス表記)」、「, (カンマ)」が文字列として入力されているため、関数計算を行うとエラーとなる。

数値データは数値属性として入力することで、関数計算が可能となる。

例2 数値データ内に空白がある場合

修正前			修正後		
桁区切りのために空白が含まれている			空白を除いた状態		
	仕入台数	在庫台数		仕入台数	在庫台数
サンプル1	5 000	1 300	サンプル1	5000	1300
サンプル2	4 000	1 800	サンプル2	4000	1800
サンプル3	3 000	1 000	サンプル3	3000	1000
サンプル4	2 000	1 600	サンプル4	2000	1600

数値データに、空白が含まれる場合、数値ではなく文字列として扱われ、関数計算を行うとエラーとなる。また、昇順・降順等の並べ替えも正確にできない。

そのため、数値データは数値属性とし、文字列を含まないこと。

例3 数値データ内に注釈・脚注が含まれる場合

修正前			修正後				
セルに注釈が含まれている			注釈を分離した状態				
	商品 A	商品 B		商品 A	商品 A_注釈	商品 B	商品 B_注釈
1月	151 a)	139	1月	151	冷凍	139	
2月	182	201 b)	2月	182		201	調理済
3月	131	123	3月	131		123	
			注釈を欄外に記載した状態				
	商品 A	商品 B					
1月	151	139					
2月	182	201					
3月	131	123					
			1月の商品 Aは冷凍 2月の商品 Bは調理済				
			注釈を別シートに記載した状態				
	商品 A	商品 B					
1月	151	139					
2月	182	201					
3月	131	123					
	商品 A	商品 B					
1月	冷凍						
2月		調理済					
3月							

数値データに、注釈・それを表す記号が含まれる場合、数値ではなく文字列として扱われるため、同じ列について数値だけのデータと、文字列を含めたデータが混在することとなる。

また、Python 等でデータを読み込む際は、1つの列は単一のデータの型(整数、小数、文字列といったもの)であることを前提としているため、注釈・脚注等の文字列が混在すると、その列は数値を含めて、全て文字列として扱われてしまい、その後の処理で思わぬ挙動を示すことがある。

そのため、表内に注釈・脚注がある場合は、「別途列を設けて注釈を分離させる」、「表の欄外に記載する」又は「注釈を別シートに記載する」のいずれかの方法により記載すること。

□チェック項目2-3 セルの結合をしていないか

表形式のデータのうち、特にレコード形式のデータを機械判読可能なデータとして利用するためには、1件のデータを、横1行で入力(レコード)又は縦1列で入力(フィールド)する必要がある。

レコード及びフィールドはそれぞれ独立しており、レコードの場合は上下の並べ替えをしてもデータの意味が変わらず、フィールドの場合は左右で入れ替えても意味が変わらないようにセルの結合又は不必要な分離を行わないこと。

例1

修正前

セルが結合(又は分離)されている

	管理職	従業員数（上段は正社員、下段はパート）	
第一営業所	3	15	2

市区町村	生産本数
ちよだく	58406
千代田区	
ちゅうおうく	141183
中央区	
みなとく	243283
港区	

修正後

セルの結合を解除した状態

	管理職	従業員数（正社員）	従業員数（パート）
第一営業所	3	15	2

市区町村	ふりがな	生産本数
千代田区	ちよだく	58406
中央区	ちゅうおうく	141183
港区	みなとく	243283

セルが結合(又は分離)されており、機械判読に適していない。

そのため、1件のデータは、横1行で表記し、セルの結合又は不必要な分離を行わないこと。

例2

修正前					修正後				
セルが結合されている					セルの結合を解除した状態				
都道府県	エリア	市区町村	導入台数	増減数	都道府県	エリア	市区町村	導入台数	増減数
東京都	特別区部	千代田区	58406	11291	東京都	特別区部	千代田区	58406	11291
	特別区部	中央区	141183	18421	東京都	特別区部	中央区	141183	18421
	特別区部	港区	243283	38152	東京都	特別区部	港区	243283	38152
	Σ				Σ				
	特別区部	葛飾区	442913	327	東京都	特別区部	葛飾区	442913	327
	特別区部	江戸川区	681298	2331	東京都	特別区部	江戸川区	681298	2331
	市町村	八王子市	577513	-2540	東京都	市町村	八王子市	577513	-2540
	市町村	立川市	176295	-3373	東京都	市町村	立川市	176295	-3373
	市町村	武蔵野市	144730	5996	東京都	市町村	武蔵野市	144730	5996
	市町村	三鷹市	186936	853	東京都	市町村	三鷹市	186936	853
	市町村	青梅市	137381	-1958	東京都	市町村	青梅市	137381	-1958
	市町村	府中市	260274	4768	東京都	市町村	府中市	260274	4768
Σ					Σ				
	市町村	青ヶ島村	-	-	東京都	市町村	青ヶ島村	-	-
	市町村	小笠原村	-	-	東京都	市町村	小笠原村	-	-

セルを結合した場合、並べ替えができない(エラーとなる)、グラフ化ができないほか、次の例のとおり、範囲選択しにくい、コピーペーストできないなどが発生する。

例：市区町村と「導入台数」をコピー
すべく、範囲選択しても…

市区町村	導入台数	増減数
西東京市	200012	3501
瑞穂町	33445	-52
日の出町	17446	796
檜原村	2209	-349
奥多摩町	5234	-811
青ヶ島村	-	-
小笠原村	-	-

セルが結合されている青ヶ島村以降は
不要な「増減数」まで選択されてしまう

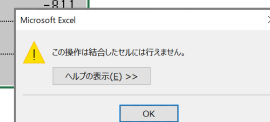
市区町村	導入台数	増減数
西東京市	200012	3501
瑞穂町	33445	-52
日の出町	17446	796
檜原村	2209	-349
奥多摩町	5234	-811
青ヶ島村	-	-
小笠原村	-	-

例：別の表からコピーして貼付け
ようとしても…

檜原村	2209	-349
奥多摩町	5234	-811
青ヶ島村	2749	-134
小笠原村	1891	2

エラーになり貼付けできない

檜原村	2209	-349
奥多摩町	5234	-811
青ヶ島村	-	-
小笠原村	-	-



□チェック項目2-4 スペースや改行等で体裁を整えていないか

スペースや改行等で体裁を整えた場合、データの検索性が低下するほか、複数の表を横断的に利用する場合においても支障が生じる可能性がある。

そのため、体裁を整えるためのスペースや改行等は削除すること。

例1

修正前

スペースで体裁を整えている

地域名	出荷本数	在庫本数
津□□市	429	756
四日市市	321	648
伊勢市	384	438
松阪市	408	775

修正後

地域コードを併記した状態

地域コード	地域名	出荷本数	在庫本数
24201	津□□市	429	141183
24202	四日市市	321	243283
24203	伊勢市	384	333560
24204	松阪市	408	219724

スペースによる整形を解除しつつ、地域コードの地域名を表記した状態

地域名	出荷本数	在庫本数
三重県津市	429	141183
三重県四日市市	321	243283
三重県伊勢市	384	333560
三重県松阪市	408	219724

[チェック項目2-11](#)の例を参考に、地域コードを併記するか、スペースによる整形を解除しつつ、地域コードの地域名を表記した状態とすること。

例2

修正前				修正後			
スペースで体裁を整えている				スペースを解除した状態			
分類	総数	事業所数	企業数	分類	総数	事業所数	企業数
合計	900	450	450	合計	900	450	450
A	200	100	100	A	200	100	100
B	300	150	150	B	300	150	150
C	400	200	200	C	400	200	200

分類の「A」以降の項目が総計の内数であることを示すために、「□A」とスペースを挿入することで体裁を整えており、本来の情報とは無関係な情報が入力されているため、他の統計データと結合しようとしても（関数の VLOOKUP 等）同じ分類がマッチせずに結合ができないなどの可能性がある。

そのため、体裁のためのスペースによる整形をしないこと。

例3

修正前			修正後		
改行で体裁を整えている			改行を解除した状態		
薬剤名	出荷 本数	単価	薬剤名	出荷本数	在庫本数
鎮静剤 A - 1	429	758	鎮静剤 A - 1	429	758
鎮静剤 A - 2	321	648	鎮静剤 A - 2	321	648
鎮静剤 A - 3	384	438	鎮静剤 A - 3	384	438
鎮静剤 A - 4	408	775	鎮静剤 A - 4	408	775

項目の体裁を整えるため改行が使用されているが、その改行に意味があるのか機械は判別が出来ない。
そのため、体裁のための改行による整形をしないこと。

□チェック項目2-5 項目名等を省略していないか

ヒトであれば省略されている部分の意味を判断できるが、ソフトウェア等のプログラムでは判断ができない。

そのため、項目名等を省略しないこと。

例

修正前

修正後

同じ名称を空白で省略している

省略せずに入力した状態

薬剤名	出荷本数	在庫本数
鎮静剤 A - 1	429	756
2	321	648
3	384	438
4	408	775

薬剤名	出荷本数	在庫本数
鎮静剤 A - 1	429	756
鎮静剤 A - 2	321	648
鎮静剤 A - 3	384	438
鎮静剤 A - 4	408	775

項目名が何を意味するのか分からない構造になっている(レコードとして成り立っていない)ため、省略せずに入力すること

□チェック項目2-6 数式を使用している場合は、数値データに修正しているか

例

修正前

合計が数式で入力されている

都道府県	市区町村	合計	男	女
東京都	千代田区	20000	10000	10000
東京都	中央区	6000	2000	4000
東京都	港区	8000	3000	5000
東京都	新宿区	5000	4000	1000
東京都	文京区	7000	5000	2000

fx

=SUM(Q100:R100)

修正後

数値データに修正した状態

都道府県	市区町村	合計	男	女
東京都	千代田区	20000	10000	10000
東京都	中央区	6000	2000	4000
東京都	港区	8000	3000	5000
東京都	新宿区	5000	4000	1000
東京都	文京区	7000	5000	2000

fx

20000

数式を使用してセルの値を入力している場合、並べ替え等を行った場合、正確な値が表示されなくなる可能性がある。

そのため、セルのデータは値のみとすること。

□チェック項目2-7 オブジェクトを使用していないか

例

修正前

オブジェクトを使用している

名称	内容	数量
商品 A	サンプル 1	100
	サンプル 2	200
	サンプル 3	300

修正後

オブジェクトを削除した状態

名称	内容	数量
商品 A	サンプル 1	100
商品 A	サンプル 2	200
商品 A	サンプル 3	300

オブジェクトを使用して体裁を整えているため、機械判読に適していない。
オブジェクトを削除した上で、それぞれのセルにデータを入力すること。

□チェック項目2-8 データの単位を記載しているか

例

修正前

単位が表記されていない

薬剤名	出荷本数	単価
鎮静剤 A - 1	429	756
鎮静剤 A - 2	321	648
鎮静剤 A - 3	384	438
鎮静剤 A - 4	408	775

修正後

単位を入力した状態

	出荷本数	単価
薬剤名	本	円
鎮静剤 A - 1	429	756
鎮静剤 A - 2	321	648
鎮静剤 A - 3	384	438
鎮静剤 A - 4	408	775

データの単位(物理単位、貨幣単位)は、データ処理に必須である。
そのため、単位が含まれる項目については、別セルにその項目の単位を入力すること。

□チェック項目2-9 機種依存文字を使用していないか。

例

修正前				修正後			
機種依存文字を使用している				機種依存文字を削除した状態			
	収納済額	収納未済額	収納未済額割合		収納済額	収納未済額	収納未済額割合
	①	②	②／①		1	2	2／1
サンプルA	1000	100	0.1	サンプルA	1000	100	0.1
サンプルB	2000	200	0.1	サンプルB	2000	200	0.1
サンプルC	3000	300	0.1	サンプルC	3000	300	0.1
サンプルD	4000	400	0.1	サンプルD	4000	400	0.1

機種依存文字は利用者の環境によっては正しく表示されない等の可能性がある。

そのため、機種依存文字は使用しないこと。

□チェック項目2-10 e-Stat の時間軸コードの表記、西暦表記又は和暦に西暦の併記がされているか

例

修正前

和暦のみで表記している

年次（暦年）	出荷本数	在庫本数
平成29年	429	141183
平成30年	321	243283
平成31年	384	333560
令和2年	408	219724

修正後

時間軸コードに和暦を併記した状態

時間軸コード	和暦（暦年）	出荷本数	在庫本数
2017000000	平成29年	429	141183
2018000000	平成30年	321	243283
2019000000	平成31年	384	333560
2020000000	令和2年	408	219724

西暦で表記した状態

西暦（暦年）	出荷本数	在庫本数
2017年	429	141183
2018年	321	243283
2019年	384	333560
2020年	408	219724

西暦に和暦を併記した状態

西暦（暦年）	和暦（暦年）	出荷本数	在庫本数
2017年	平成29年	429	141183
2018年	平成30年	321	243283
2019年	平成31年	384	333560
2020年	令和2年	408	219724

時間軸を示すデータについて、ソフトウェア等のプログラムは、年の値の大小により認識することが多いため、和暦表示のみでは、元号が切り替わる際に手作業で西暦等に変換する必要がある。

そこで、既に e-Stat で用いている時間軸コードを活用することで、府省間の共通化を図ることともに、データ収集や横断的利用の自動化が容易となる。

そのため、時間軸は、可能な限り e-Stat で用いられる時間軸コード（次頁「e-Stat で用いられる時間軸コード体系」参照）を記載すること。

また、e-Stat で用いられる時間軸コード以外で、既に調査独自で時間軸コードが設定されている場合は、原則当該コードを記載（維持）すること。ただし、この場合、同一表内で e-Stat に用いられる時間軸コードと重複させないこと。

さらに、時間軸コードのセルには、時間軸コード以外の情報を記載しないこと。これにより難しい場合は、その旨を明記すること。

なお、時間軸コードを用いない場合は、単調増加する西暦での表記又は和暦に西暦を併記すること。

e-Stat で用いられる時間軸コード体系

コード名称	説明
時間軸(年)コード	[年(4桁)][種別(1桁)][上/下期(1桁)][期(4桁)] ・[年] ----- 西暦 4 桁 ・[種別] ----- 0(年を意味する)固定 ・[上期/下期] --- 0=指定なし, 1=1～6 月期, 2=7～12 月期 ・[期] ----- 0000=年, 0103= 1～ 3 月期, 0101= 1 月, 0202= 2 月, 0303= 3 月, 0406= 4～ 6 月期, 0404= 4 月, 0505= 5 月, 0606= 6 月, 0709= 7～ 9 月期, 0707= 7 月, 0808= 8 月, 0909= 9 月, 1012=10～12 月期, 1010=10 月, 1111=11 月, 1212=12 月
時間軸(年度)コード	[年(4桁)][種別(1桁)][上/下期(1桁)][期(4桁)] ・[年] ----- 西暦 4 桁 ・[種別] ----- 1(年度を意味する)固定 ・[上期/下期] --- 0=指定なし, 1=4～9 月期, 2=10～3 月期 ・[期] ----- 0000=年度, 0103= 1～ 3 月期, 0101= 1 月, 0202= 2 月, 0303= 3 月, 0406= 4～ 6 月期, 0404= 4 月, 0505= 5 月, 0606= 6 月, 0709= 7～ 9 月期, 0707= 7 月, 0808= 8 月, 0909= 9 月, 1012=10～12 月期, 1010=10 月, 1111=11 月, 1212=12 月

例

コード	説明
2006000000	2006 年
2006010000	2006 年 1～6 月期
2006100000	2006 年度
2006110000	2006 年度 4～9 月期
2006000103	2006 年 1～3 月期
2006000101	2006 年 1 月

□チェック項目2-11 地域コード又は地域名称が表記されているか

例

修正前

都道府県名を略称している

都道府県	出荷本数	在庫本数
北海道	429	141183
青森	321	243283
岩手	384	333560
宮城	408	219724

修正後

地域コードを併記した状態

地域コード	地域名	出荷本数	在庫本数
01	北海道	429	141183
02	青森	321	243283
03	岩手	384	333560
04	宮城	408	219724

同コードで定められた名称を表記した状態

都道府県	出荷本数	在庫本数
北海道	429	141183
青森県	321	243283
岩手県	384	333560
宮城県	408	219724

地域の表記については、県名の省略や順不同の記載、同一名称の場合に利用者が混乱する場合がある。

そこで、都道府県及び市区町村の区域を示す統計情報の表章及び当該情報の相互利用のための基準である「統計に用いる標準地域コード(昭和45年4月統計審議会答申決定)(以下、「標準地域コード」という。)」を可能な限り記載すること。これによらない場合は同コードに定められた地域の名称を記載すること。

また、標準地域コード以外で、既に調査独自で地域コードが設定されている場合は、原則当該コードを記載(維持)すること。ただし、この場合、同一表内で標準地域コードと重複させないこと。

さらに、地域コードのセルには、地域コード以外の情報を記載しないこと。これにより難しい場合は、その旨を明記すること。

□チェック項目2-12 数値データの同一列内に特殊記号(秘匿等)が含まれる場合

数値データの同一列内に秘匿等の特殊処理を行っていることを示す記号を記載している場合、数値ではなく文字列として扱われるが、同一セルには数値が入っておらず、Excel 関数で計算した場合、当該セルは無視されて計算される。

そのため、特殊記号は例外として、数値データと同一列に記載することを許容する。ただし、同一セル内の数値と文字の混在はさせないこと。

なお、特殊文字は原則以下の記号に統一する。これにより難しい場合は、使用した記号の意味を明記すること。

特殊処理	記 号
集計した結果がゼロ、表章桁未満	0
集計に必要なデータがない	***
秘匿データ	X

例

修正前			修正後		
都道府県	サンプル 1	サンプル 2	都道府県	サンプル 1	サンプル 2
北海道	52954	44940	北海道	52954	44940
青森県			青森県	***	***
岩手県	6566	5933	岩手県	6566	5933
宮城県			宮城県	***	***

第2項 表の構成の取扱い

□チェック項目3-1 データが分断されていないか

空白列などを追加したり、Excelの1シートに複数の表頭・表側を設定した場合、データが分断されてしまい、機械判読に支障をきたす可能性がある。

そのため、不必要な表の分離は行わないこと。

また、表頭、表側、欄外等に、表の印刷や成形を意識した不必要な空白行(列)を追加しないこと。

例1

修正前

データが分断されている

都道府県	エリア	市区町村	導入台数 台	増減数		材料 台	仕入単価 円	仕入額 円
東京都	特別区部	千代田区	58406	11291		429	28	12012
東京都	特別区部	中央区	141183	18421		321	35	11235
東京都	特別区部	港区	243283	38152		384	42	16128

修正後

空白列を削除した状態

都道府県	エリア	市区町村	導入台数	増減数	材料	仕入単価	仕入額
			台		台	円	円
東京都	特別区部	千代田区	58406	11291	429	28	12012
東京都	特別区部	中央区	141183	18421	321	35	11235
東京都	特別区部	港区	243283	38152	384	42	16128



データが分断されているため、空白列を削除し、データが分断されないようにすること。

例2

修正前				修正後			
データが分断されている				空白行を削除した状態			
都道府県	市区町村	人口	増減数	都道府県	市区町村	人口	増減数
東京都	千代田区	58406	11291	東京都	千代田区	58406	11291
東京都	中央区	141183	18421	東京都	中央区	141183	18421
東京都	港区	243283	38152	東京都	港区	243283	38152
}				東京都	新宿区	333560	7251
東京都	中野区	328215	13465	東京都	中野区	328215	13465
東京都	杉並区	563997	14428	東京都	杉並区	563997	14428
東京都	豊島区	291167	6489	東京都	豊島区	291167	6489
				東京都	北区	341076	5532
都道府県	市区町村	人口	増減数	東京都	荒川区	212264	8968
東京都	北区	341076	5532	東京都	板橋区	561916	26092
東京都	荒川区	212264	8968	東京都	練馬区	721722	5598
東京都	板橋区	561916	26092	東京都	足立区	670122	-13304
}				東京都	葛飾区	442913	327
東京都	府中市	260274	4768	東京都	調布市	229061	5468
東京都	昭島市	111539	-758	東京都	町田市	432348	5332
東京都	調布市	229061	5468	東京都	小金井市	121396	2544

空白行でデータが分断されているため、空白行を削除し、データが分断されないようにすること。

例3

修正前

紙面に収めるために表を折り返している

		給 与 所 得 者 数							給 与 所 得 者 数				
		3 月 末	6 月 末	9 月 末	12 月 末	年 間 月 平 均			3 月 末	6 月 末	9 月 末	12 月 末	年 間 月 平 均
		千 人	千 人	千 人	千 人	千 人			千 人	千 人	千 人	千 人	千 人
		千 人	千 人	千 人	千 人	千 人			千 人	千 人	千 人	千 人	千 人
平成7年分	55982	55935	55673	55354	55736	平成13年分	64379	64325	64024	63657	64096		
平成8年分	55987	56809	56038		56189	平成14年分	64385	65330	64444	64311	64617		
平成9年分	56850			56689		平成15年分	65378			65192			
平成10年分	56948			56958		平成16年分	67790			67112			
平成11年分	59264		58300	58108	58761	平成17年分	68154	68280	67045	66824	67575		
平成12年分	61001	60901	59276	59114	60151	平成18年分	70513	70036	68167	67981	69174		
		収録範囲 1							収録範囲 2				

		給 与 所 得 者 数							給 与 所 得 者 数				
		3 月 末	6 月 末	9 月 末	12 月 末	年 間 月 平 均			3 月 末	6 月 末	9 月 末	12 月 末	年 間 月 平 均
		千 人	千 人	千 人	千 人	千 人			千 人	千 人	千 人	千 人	千 人
		千 人	千 人	千 人	千 人	千 人			千 人	千 人	千 人	千 人	千 人
平成19年分	61580	61529	61240	60889	61310	平成25年分	67178	67122	66808	66425	66883		
平成20年分	61586	62490	61642	61008		平成26年分	67184	68171	67246	67108	67427		
平成21年分	62535			62358		平成27年分	68220			68027			
平成22年分	64843			64194		平成28年分	70738			70030			
平成23年分	65190	64111	64130	63919	64637	平成29年分	71117	71249	69960	69730	70513		
平成24年分	66991	66991	65204	65025	66166	平成30年分	73579	73081	71131	70937	72181		
		収録範囲 3							収録範囲 4				

修正後

収録範囲を1つにした状態

		給 与 所 得 者 数				
		3 月 末	6 月 末	9 月 末	12 月 末	年 間 月 平 均
		千 人	千 人	千 人	千 人	千 人
1995年	平成7年分	55982	55935	55673	55354	55736
1996年	平成8年分	55987	56809	56038	55923	56189
}						
2017年	平成29年分	71117	71249	69960	69730	70513
2018年	平成30年分	73579	73081	71131	70937	72181

紙面に収めるために表を折り返しているため、正確な計算や並べ替えなどができる場合がある。
収録範囲を1つにし、データが分断されないようにすること。

□チェック項目3-2 1シートに複数の表が掲載されていないか

例

修正前

1シートに複数の表が掲載されている

都道府県	市区町村	合計	男	女
東京都	千代田区	20000	10000	10000
東京都	中央区	6000	2000	4000
東京都	港区	8000	3000	5000
東京都	新宿区	5000	4000	1000
東京都	文京区	7000	5000	2000

都道府県	出荷本数	在庫本数
北海道	429	141183
青森県	321	243283
岩手県	384	333560
宮城県	408	219724

西暦（暦年）	出荷本数	在庫本数
2017年	429	141183
2018年	321	243283
2019年	384	333560
2020年	408	219724

修正後

1つの表を1シートに分割した状態

都道府県	市区町村	合計	男	女
東京都	千代田区	20000	10000	10000
東京都	中央区	6000	2000	4000
東京都	港区	8000	3000	5000
東京都	新宿区	5000	4000	1000
東京都	文京区	7000	5000	2000

西暦（暦年）	出荷本数	在庫本数
2017年	429	141183
2018年	321	243283
2019年	384	333560
2020年	408	219724

都道府県	出荷本数	在庫本数
北海道	429	141183
青森県	321	243283
岩手県	384	333560
宮城県	408	219724

1シートに複数の表が掲載されているため、正確な計算や昇順・降順の並べ替えなどできない場合がある。

1つの表を1シートに分割して収録すること。

第3節 CSV ファイルによる結果表のレイアウト

本節では、CSV ファイル形式による結果表における留意点を以下に示す。統計データを作成する際に、本節で示すチェック項目を1つずつ踏まえ改善していくことで、PDF 形式による結果表や、印刷して報告書として提供することを主眼にした表形式データを、機械判読可能なデータとして利用できる結果表に改善することが可能となる。

第 3 節に出てくる主な用語の定義は以下の通り。

用語	説明
ファイル	CSV ファイル。
レコード	テキストの行。
フィールド	カンマで区切られた範囲の数字または文字列。
変数	コンピューターのプログラム中で記憶場所の名称として用いられる文字列。

第1項 データ・項目の取扱い

□チェック項目4-1 変数行から始まり、次行からデータ入力がされているか

1 行目に変数を入力し、2 行目以降にデータを入力することで、データが属している変数が明確になる。また、活用を考えた場合、Excel で閲覧したときにピボットグラフを使用してグラフ表示に切り替えられる、データベースの作成において特定のデータを抜き出しやすくなるなどの利点がある。

例1

修正前	修正後
1 行目に変数が記載されていない 東京都, 2005, 12576601 東京都, 2010, 13159388 神奈川県, 2005, 8791597 神奈川県, 2010, 9048331	1 行目に変数が記載された状態 都道府県, 年度, 人口 東京都, 2005, 12576601 東京都, 2010, 13159388 神奈川県, 2005, 8791597 神奈川県, 2010, 9048331

例2

修正前
変数が 2 行にまたがっている , 総数 , 男 , 女 , 総数 , 男 , 女 都道府県, 2005 年 , 2005 年 , 2005 年 , 2010 年 , 2010 年 , 2010 年 東京都 , 12576601, 6264895, 6311706, 13159388, 6512110, 6647278 神奈川県, 8791597 , 4444555, 4347042, 9048331 , 4544545, 4503786
修正後
変数が 1 行に記載された状態 都道府県, 年度, 総数, 男, 女 東京都, 2005, 12576601, 6264895, 6311706 東京都, 2010, 13159388, 6512110, 6647278 神奈川県, 2005, 8791597, 4444555, 4347042 神奈川県, 2010, 9048331, 4544545, 4503786

□チェック項目4-2 1フィールド1データとなっているか

1フィールドに複数のデータが入力されていると、計算や昇順・降順の並べ替え、コピーペーストやグラフ化等加工編集する場合に多くの手作業やプログラムの作成が必要となり、すぐにデータとして利用できないため、1フィールド1データの入力とすること。

例

修正前	修正後
1フィールドに複数のデータが入力されている 区分, 企業数等, 売上金額 (費用総額) 総計, 1188389, 391445 (355943) A, 50384, 69565 (63883) B, 154138, 50468 (45332) C, 86522, 44098 (33210) D, 85983, 22258 (20237) E, 580003, 37652 (35016) F, 24756, 15288 (14633) G, 119085, 115499 (110735) H, 84818, 36617 (32897)	1フィールド1データとして入力した状態 区分, 企業数等, 売上金額, 費用総額 総計, 1188389, 391445, 355943 A, 50384, 69565, 63883 B, 154138, 50468, 45332 C, 86522, 44098, 33210 D, 85983, 22258, 20237 E, 580003, 37652, 35016 F, 24756, 15288, 14633 G, 119085, 115499, 110735 H, 84818, 36617, 32897

売上金額と費用総額が1フィールドに入力されているため、例えば、売上金額に占める費用総額の割合を計算したい場合、修正前の場合、「()」内の数字を分離してから、割り算を行う必要が生じる。

一方、修正後の場合では、売上金額と費用総額が別々のフィールドになっているため、単純な処理で計算が可能となる。

そのため、1フィールド1データの入力すること。

□チェック項目4-3 数値データは数値属性とし、文字列を含まないこと

数値データに、円、¥、kg、トやマイナス記号(▲)などの文字列を含めると、数値データとして扱うことができない。

また、ヒトが見やすくするための工夫として、千円単位を示す「,(カンマ)」を文字列として直接入力している場合、機械はそのカンマでフィールドが区切れると判別し、思わぬ挙動を示す場合がある。

そのため、数値データは数値属性とし、文字列を含まないこと。

なお、数値データの同一列内に特殊記号(秘匿等)が含まれる場合は、例外処理を認めているため、[チェック項目 4-11](#)を参照すること。

例1 数値データ内に文字列が含まれる場合

修正前	修正後
「円」、「▲(マイナス表記)」、「,(カンマ)」が文字列として入力されている。 種類, "単価", 前回差分, "生産台数" サンプル 1, "10, 030 円", 130, "12, 000" サンプル 2, "9, 100 円", ▲200, "29, 000" サンプル 3, "8, 000 円", ▲350, "37, 000" サンプル 4, "7, 500 円", 500, "43, 000"	数値データを数値属性として入力した状態 種類, 単価, 前回差分, 生産台数 サンプル 1, 10030, 130, 12000 サンプル 2, 9100, 200, 29000 サンプル 3, 8000, -350, 37000 サンプル 4, 7500, 500, 43000

上記例の場合、「円」、「▲(マイナス表記)」、「,(カンマ)」が文字列として入力されているため、数値データとして扱うことができない。

そのため、数値データは数値属性として入力すること。

例2 数値データ内に空白がある場合

修正前	修正後
桁区切りのために空白が含まれている 種類, 仕入台数, 在庫台数 サンプル 1, 5 000, 1 300 サンプル 2, 4 000, 1 800 サンプル 3, 3 000, 1 000 サンプル 4, 2 000, 1 600	空白を除いた状態 種類, 仕入台数, 在庫台数 サンプル 1, 5000, 1300 サンプル 2, 4000, 1800 サンプル 3, 3000, 1000 サンプル 4, 2000, 1600

例3 数値データ内に注釈・脚注が含まれる場合

修正前	修正後
数値データ内に注釈や脚注が含まれている 仕入月, 商品 A, 商品 B 1 月, 151 a) 139 2 月, 182, 201 b) 3 月, 131, 123	注釈や脚注を除いた状態 仕入月, 商品 A, 商品 A_注釈, 商品 B, 商品 B_注釈 1 月, 151, 冷凍 139, 2 月, 182, , 201, 調理済 3 月, 131, , 123,

また、Python 等でデータを読み込む際は、1つの列は単一のデータの型(整数、小数、文字列といったもの)であることを前提としているため、文字列が混在すると、その列は数値を含めて、全て文字列として扱われてしまい、その後の処理で思わぬ挙動を示すことがある。

そのため、数字と文字列は別のフィールドとして扱うこと。

□チェック項目4-4 スペースを使っていないか

体裁を整えるために単語の文字間にスペースを挿入すると、単語を検索した場合にヒットしない。そのため、不要なスペースを挿入しないこと。

例

修正前	修正後
単語の文字間にスペースが挿入されている 地域名, 出荷本数, 在庫本数 津 市, 429, 756 四日市市, 321, 648 伊 勢 市, 384, 438 松 阪 市, 408, 775	不要なスペースのない状態 地域名, 出荷本数, 在庫本数 津市, 429, 756 四日市市, 321, 648 伊勢市, 384, 438 松阪市, 408, 775

□チェック項目4-5 1行1データで表現されているか

1 データが複数の行にまたいで入力されていると、機械が判読できない。そのため、1 データは 1 行に収めること。

例

修正前	修正後
1 データが複数の行にまたいでいる 薬剤名, 出荷本数, 単価 鎮静剤 A-1, 429, 756 鎮静剤 A-2, 321, 648	1 データが 1 行に収まった状態 薬剤名, 出荷本数, 単価 鎮静剤 A-1, 429, 756 鎮静剤 A-2, 321, 648

□チェック項目4-6 項目名等を省略していないか

ヒトであれば省略されている部分の意味を判断できるが、ソフトウェア等のプログラムでは判断ができない。

そのため、項目名等を省略しないこと。

例

修正前	修正後
同じ名称を省略している 薬剤名, 出荷本数, 単価 鎮静剤 A-1, 429, 756 2, 321, 648 3, 384, 438 4, 408, 775	省略せずに入力した状態 薬剤名, 出荷本数, 単価 鎮静剤 A-1, 429, 756 鎮静剤 A-2, 321, 648 鎮静剤 A-3, 384, 438 鎮静剤 A-4, 408, 775

上記の例では、項目名が何を意味するのか分からない構造になっている(レコードとして成り立っていない)ため、省略せずに入力すること。

□チェック項目4-7 データの単位を記載しているか

例

修正前	修正後
単位が表記されていない 薬剤名, 出荷本数, 単価 鎮静剤 A-1, 429, 756 鎮静剤 A-2, 321, 648 鎮静剤 A-3, 384, 438 鎮静剤 A-4, 408, 775	単位を入力した状態 薬剤名, 出荷本数(本), 単価(円) 鎮静剤 A-1, 429, 756 鎮静剤 A-2, 321, 648 鎮静剤 A-3, 384, 438 鎮静剤 A-4, 408, 775

データの単位(物理単位、貨幣単位)は、データ処理に必須である。

そのため、単位が含まれる項目については、変数と同じフィールドに単位を記載すること。

□チェック項目4-8 機種依存文字を使用していないか

例

修正前	修正後
機種依存文字を使用している 種類, 収納済額 (①), 収納未済額 (②), 収納未済額割合 ((②/①)) サンプル A, 1000, 100, 0.1 サンプル B, 2000, 200, 0.1 サンプル C, 3000, 300, 0.1 サンプル D, 4000, 400, 0.1	機種依存文字を削除した状態 種類, 収納済額 (1), 収納未済額 (2), 収納未済額割合 (2/1) サンプル A, 1000, 100, 0.1 サンプル B, 2000, 200, 0.1 サンプル C, 3000, 300, 0.1 サンプル D, 4000, 400, 0.1

機種依存文字は利用者の環境によっては正しく表示されない等の可能性がある。

そのため、機種依存文字は使用しないこと。

□チェック項目4-9 e-Stat の時間軸コードの表記、西暦表記又は和暦に西暦の併記がされているか

例

修正前	修正後
和暦のみで表記している 年次（暦年），出荷本数, 在庫本数 平成 29 年 429, 141183 平成 30 年 321, 243283 平成 31 年 384, 333560 令和 2 年, 408, 219724	時間軸コードに和暦を併記した状態 時間軸コード, 和暦（暦年），出荷本数, 在庫本数 2017000000, 平成 29 年 429, 141183 2018000000, 平成 30 年 321, 243283 2019000000, 平成 31 年 384, 333560 2020000000, 令和 2 年, 408, 219724 西暦で表記した状態 西暦（暦年），出荷本数, 在庫本数 2017, 429, 141183 2018, 321, 243283 2019, 384, 333560 2020, 408, 219724

	西暦に和暦を併記した状態 <table><tr><th>西暦（暦年）</th><th>和暦（暦年）</th><th>出荷本数</th><th>在庫本数</th></tr><tr><td>2017</td><td>平成 29 年</td><td>429</td><td>141183</td></tr><tr><td>2018</td><td>平成 30 年</td><td>321</td><td>243283</td></tr><tr><td>2019</td><td>平成 31 年</td><td>384</td><td>333560</td></tr><tr><td>2020</td><td>令和 2 年</td><td>408</td><td>219724</td></tr></table>	西暦（暦年）	和暦（暦年）	出荷本数	在庫本数	2017	平成 29 年	429	141183	2018	平成 30 年	321	243283	2019	平成 31 年	384	333560	2020	令和 2 年	408	219724
西暦（暦年）	和暦（暦年）	出荷本数	在庫本数																		
2017	平成 29 年	429	141183																		
2018	平成 30 年	321	243283																		
2019	平成 31 年	384	333560																		
2020	令和 2 年	408	219724																		

時間軸を示すデータについて、ソフトウェア等のプログラムは、年の値の大小により認識することが多いため、和暦表示のみでは、元号が切り替わる際に手作業で西暦等に変換する必要がある。

そこで、既に e-Stat で用いている時間軸コードを活用することで、府省間の共通化を図ることともに、データ収集や横断的利用の自動化が容易となる。

そのため、時間軸は、可能な限り e-Stat で用いられる時間軸コード(次頁「e-Stat で用いられる時間軸コード体系」参照)を記載すること。

また、e-Stat で用いられる時間軸コード以外で、既に調査独自で時間軸コードが設定されている場合は、原則当該コードを記載(維持)すること。ただし、この場合、同一表内で e-Stat に用いられる時間軸コードと重複させないこと。

さらに、時間軸コードのフィールドには、時間軸コード以外の情報を記載しないこと。これにより難い場合は、その旨を明記すること。

なお、時間軸コードを用いない場合は、単調増加する西暦での表記又は和暦に西暦を併記すること。

e-Stat で用いられる時間軸コード体系

コード名称	説明
時間軸(年)コード	[年(4桁)][種別(1桁)][上/下期(1桁)][期(4桁)] ・[年] ----- 西暦 4 桁 ・[種別] ----- 0(年を意味する)固定 ・[上期/下期] --- 0=指定なし, 1=1～6 月期, 2=7～12 月期 ・[期] ----- 0000=年, 0103= 1～ 3 月期, 0101= 1 月, 0202= 2 月, 0303= 3 月, 0406= 4～ 6 月期, 0404= 4 月, 0505= 5 月, 0606= 6 月, 0709= 7～ 9 月期, 0707= 7 月, 0808= 8 月, 0909= 9 月, 1012=10～12 月期, 1010=10 月, 1111=11 月, 1212=12 月
時間軸(年度)コード	[年(4桁)][種別(1桁)][上/下期(1桁)][期(4桁)] ・[年] ----- 西暦 4 桁 ・[種別] ----- 1(年度を意味する)固定 ・[上期/下期] --- 0=指定なし, 1=4～9 月期, 2=10～3 月期 ・[期] ----- 0000=年度, 0103= 1～ 3 月期, 0101= 1 月, 0202= 2 月, 0303= 3 月, 0406= 4～ 6 月期, 0404= 4 月, 0505= 5 月, 0606= 6 月, 0709= 7～ 9 月期, 0707= 7 月, 0808= 8 月, 0909= 9 月, 1012=10～12 月期, 1010=10 月, 1111=11 月, 1212=12 月

例

コード	説明
2006000000	2006 年
2006010000	2006 年 1～6 月期
2006100000	2006 年度
2006110000	2006 年度 4～9 月期
2006000103	2006 年 1～3 月期
2006000101	2006 年 1 月

□チェック項目4-10 地域コード又は地域名称が表記されているか

例

修正前	修正後
都道府県名を略称している 都道府県, 出荷本数, 在庫本数 北海道, 429, 141183 青森, 321, 243283 岩手, 384, 333560 宮城, 408, 219724	地域コードを併記した状態 地域コード, 都道府県, 出荷本数, 在庫本数 01, 北海道, 429, 141183 02, 青森, 321, 243283 03, 岩手, 384, 333560 04, 宮城, 408, 219724 同コードで定められた名称を表記した状態 都道府県, 出荷本数, 在庫本数 北海道, 429, 141183 青森県, 321, 243283 岩手県, 384, 333560 宮城県, 408, 219724

地域の表記については、県名の省略や順不同の記載、同一名称の場合に利用者が混乱する場合がある。

そこで、都道府県及び市区町村の区域を示す統計情報の表章及び当該情報の相互利用のための基準である「統計に用いる標準地域コード(昭和 45 年4月統計審議会答申決定)(以下、「標準地域コード」という。)」を可能な限り記載すること。これによらない場合は同コードに定められた地域の名称を記載すること。

また、標準地域コード以外で、既に調査独自で地域コードが設定されている場合は、原則当該コードを記載(維持)すること。ただし、この場合、同一表内で標準地域コードと重複させないこと。

さらに、地域コードのフィールドには、地域コード以外の情報を記載しないこと。これにより難しい場合は、その旨を明記すること。

□チェック項目4-11 数値データの同一列内に特殊記号(秘匿等)が含まれる場合

特殊記号は例外として、数値データと同一列に記載することを許容する。ただし、同一フィールド内の数値と文字の混在はさせないこと。

なお、特殊文字は原則以下の記号に統一する。これにより難しい場合は、使用した記号の意味を明記すること。

特殊処理	記 号
集計した結果がゼロ、表章桁未満	0
集計に必要なデータがない	***
秘匿データ	X

例

修正前	修正後
都道府県, サンプル 1, サンプル 2 北海道, 52954, 44940 青森, , 岩手, 6566, 5933 宮城, ,	都道府県, サンプル 1, サンプル 2 北海道, 52954, 44940 青森, ***, ***, 岩手, 6566, 5933 宮城, ***, ***,

上記例の場合、変数サンプル 1 に対するフィールドが空欄となっているが、空欄の意図を機械で判読することができない。そのため、特殊処理に該当する記号を活用すること。

□チェック項目4-12 各フィールドの値をダブルコーテーション(“)で囲んでいるか

フィールドの文字列にカンマが含まれている場合、機械はそのカンマでフィールドが区切れると判別し、以降のフィールドを列がずれた状態で読み取る。フィールドの値をダブルコーテーションで囲むと、ダブルコーテーションに囲まれた文字列を1つの値として読み取るため、誤った読み取りを防ぐことができる。

例

修正前	修正後
誤読の可能性がある状態 年齢階級, 人口 15 歳以下, 5621 16, 17 歳, 12145 18, 19 歳, 15892 20～25 歳, 19224	フィールドの値をダブルコーテーションで囲んだ状態 “年齢階級”, 人口 “15 歳以下”, 5621 “16, 17 歳”, 12145 “18, 19 歳”, 15892 “20～25 歳”, 19224

□チェック項目4-13 データが分断されていないか

機械読み取りでは、空白行列をデータの終わりとみなすため、空白行があるとそれ以降のデータを読み取ることができない。

そのため、レコードの分離は行わないこと。

例

修正前	修正後
データが分断されている 都道府県, 市区町村, 人口, 増減数 東京都, 千代田区, 58406, 11291 東京都, 中央区, 141183, 18421 東京都, 港区, 243283, 38152 東京都, 新宿区, 333560, 7251 東京都, 中野区, 328215, 13465 東京都, 杉並区, 563997, 14428	空白行を削除した状態 都道府県, 市区町村, 人口, 増減数 東京都, 千代田区, 58406, 11291 東京都, 中央区, 141183, 18421 東京都, 港区, 243283, 38152 東京都, 新宿区, 333560, 7251 東京都, 中野区, 328215, 13465 東京都, 杉並区, 563997, 14428

データが分断されているため、空白行を削除し、データが分断されないようにすること。

□チェック項目4-14 1ファイル内に変数とデータのセットが複数掲載されていないか

例

修正前	修正後
1ファイルに 2 セットの変数とデータが掲載されている 都道府県, 年度, 人口 東京都, 2005, 12576601 東京都, 2010, 13159388 神奈川県, 2005, 8791597 神奈川県, 2010, 9048331 都道府県, 面積 (km2), 人口密度 (1km2 当たり) 東京都, 2190. 93, 6168. 7 神奈川県, 1905. 14, 4639. 8	1ファイルにつき1セットの変数とデータを掲載した状態 都道府県, 年度, 人口 東京都, 2005, 12576601 東京都, 2010, 13159388 神奈川県, 2005, 8791597 神奈川県, 2010, 9048331 都道府県, 面積 (km2), 人口密度 (1km2 当たり) 東京都, 2190. 93, 6168. 7 神奈川県, 1905. 14, 4639. 8

1 ファイル内に変数とデータのセットが複数掲載されているため、正確な計算や昇順・降順の並べ替えなどができない場合がある。

そのため、1 ファイル内に変数とデータのセットが複数ある場合は、1ファイルにつき1セットに分割して収録すること。

改訂履歴

Ver.	改訂内容
1.0	総務省 報道資料 統計表における機械判読可能なデータの表記方法の統一ルールの策定 https://www.soumu.go.jp/menu_news/s-news/01toukatsu01_02000186.html
1.1	・「第1章 総則」を追加 ・「第3章 CSV ファイルによる結果表のレイアウト」を追加
1.2	・P6 「チェック項目1-13」を「チェック項目2-12」に修正 □チェック項目2-2 数値データは数値属性とし、文字列を含まないこと - なお、数値データの…、<u>チェック項目1-13</u>を参照すること。 ・P11 「チェック項目 1-12」を「チェック項目 2-11」に修正 □チェック項目2-4 スペースや改行等で体裁を整えていないか <u>チェック項目1-12</u>の例を参考に、…すること。