

行政データにおける機械可読性に関するルール（案）

令和 8 年 3 月 31 日
各府省庁DX推進連絡会議・
デジタル社会推進会議幹事会決定

データは経済成長や社会課題解決の基盤資源であり、特に近年ではAIやデジタル技術の進展を背景として、新たな価値創出の原動力として不可欠な存在となっている。

行政機関が保有する行政データについても、「電子行政オープンデータ戦略」（平成24年7月4日高度情報通信ネットワーク社会推進戦略本部決定）において、機械可読可能な形式で公開し、営利・非営利を問わず自由に利用できることを基本原則として掲げて以降、官民データ活用推進基本法（平成28年法律第103号）や「オープンデータ基本指針」（平成29年5月30日高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議決定。令和6年7月5日デジタル社会推進会議幹事会改正）等により、基本的考え方を示し、その利活用を推進してきた。

本ルールは、これらの取組を一層推進するため、「データ利活用の在り方に関する基本方針」（令和7年6月13日閣議決定）等を踏まえ、各府省が保有するデータの品質向上に取り組む際に目安となる具体的な基準を定めるものである。

1. 機械可読性に関するルールの意義

「機械可読」とは、コンピュータプログラムが自動的にデータを加工、編集等できることを指す¹。

行政データの中には、人間が目で見えて理解し易いよう、表のレイアウトや文字を工夫して視認性を高めたものが多くあり、機械可読性が高いとされるテーブルデータ²にその傾向が強い。視認性を高める作業は、情報を受け取る側に配慮したものと評価できる一方、機械可読性が損なわれ、データ利活用の場面で支障を来す一因となっている。

機械可読性が確保されていないデータがあると正しい処理ができないため、当該データを利用する際に、データ品質の改善作業（データクレンジング³）が必要となる。データ品質に責任を持つデータ提供者において、適切なデータクレンジングが行われることで、行政におけるデータの転記ミスや手作業の削減、BIツールでの即座の可視化、府省間でのデータ連携、AIを活用した業務効率化・政策分析等が可能となり、EBPM（根拠に基づく政策

¹ 機械可読性を確保する上では、単にCSVやExcel形式で提供されていればよいというものではなく、コンピュータが正しく構造を解釈できるように整理されていることが重要。そのためには、セル結合されていない、各セルに1つの値のみが入力されている（1セル1データ）、データ中に空白が挿入されていない等の構造上の原則が守られている必要がある。

² 行と列で構成される表形式のデータのこと。ExcelやCSVファイルのように、同じ構造を持つ情報を規則的に並べた構造化データがこれに当たる。

³ データ分析や利活用を行う前に、データの誤り・欠損・重複・不整合などを検出し、正確で一貫性のある状態に整える作業のこと。信頼できる分析結果等を得るための基礎的な工程と位置づけられる。

立案)の推進や行政の効率化・高度化にも資する。また、行政が内部で分析等にデータを利用する場合はもとより、公開されたデータを利用する事業者や研究者等も、データ整形や確認作業に膨大な時間を費やす必要がなくなるため、社会全体のデータクレンジングコストが縮減される⁴。今後、AIやデジタル技術の活用が一層進むであろうことに鑑みれば、データの機械可読性を確保することは喫緊の課題である。

2. ルールの適用とレベルの概要

国の行政機関が作成するテーブルデータのうち、オープンなデータを対象として、ルールを適用し機械可読性の向上を図る。その際、各府省の担当者が段階的に取り組めるよう、データ品質のレベルを3段階に分けて整理することとする。

このうち、レベル1は遵守すべき基本的なルールとして位置づけ、レベル2とレベル3はより高度な機械可読性を実現するための技術的な道筋を示すものである。AI活用やデータ連携を実現するためには、少なくともレベル1の基本的な機械可読性を確保することが不可欠であり、その上でレベル2・3へと段階的に取り組むことが实际的である。

今後、オープンなデータとして作成するものについては、少なくともレベル1の水準は充たすよう順次取り組むこととする⁵。

なお、より厳しい水準で機械可読性を求める「統計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」のルールが適用される基幹統計等については、引き続き既往のルールの下で取組を推進する。

レベル1：閲覧・転記可能 (Viewable & Transferable)

- 定義：表形式データとして最低限の機械処理が可能な状態
- 概要：行と列で分けられた表形式で、文字や数値がコンピュータで認識可能な形式で保存されており、人が意図した通りに機械が読み取れる状態。
- 効果：表計算ソフト等での基本操作（コピーペースト等）が可能となる。複雑な装飾や結合セルによる「神エクセル」からの脱却により、データの再利用性が向上し、手作業での転記ミスや作業時間が大幅に削減される。

レベル2：集計・分析可能 (Aggregatable & Analyzable)

- 定義：統計的な集計・分析作業が前処理なしで即座に実施可能な状態
- 概要：列単位でルールが一貫しており、行をサンプル、列を変数として扱える統計データ形式。データ型（定量・定性）が明確で、欠損値の扱いが統一されている状態。

⁴ データ品質に責任を持つデータ提供者（各府省）ではなく、データ利用者がその作業負担を強いられているという問題がある。データ利用者の一人一人が同様のデータクレンジングを行う必要があるため、社会全体でコストが発生しているとの指摘がある。

⁵ 令和8年4月以降作成されるデータから順次取り組む。利用頻度が高いものは、ルール適用による効果も大きいと優先的に取り組む。請負契約等によりデータ整備を外部の事業者が行う場合には、あらかじめ仕様書等により、ルールに適合した形式で成果物が納品されるようにする。

- 効果：ピボットテーブルやBI ツール（Tableau、PowerBI 等）での自動集計・可視化が前処理なしで実行可能となる。データクレンジング作業が不要になり、定期レポートの自動生成やリアルタイムでのダッシュボード更新が実現される。

レベル3：連携・自動化可能（Linkable & Automatable）

- 定義：異なるデータセット間での結合・比較分析、プログラムによる自動処理が可能な状態
- 概要：標準化されたコード体系とメタデータ（単位、定義、更新履歴等）が整備され、時系列または他の統計調査と横断的に比較可能な状態。
- 効果：プログラミング言語（Python、R、SQL 等）での自動データ取得・結合が可能となる。省庁横断的なデータ統合、経年変化の自動追跡、AI モデルへの即座のデータ投入などが実現される。

3. 具体的なルール

レベル1：閲覧・転記可能

【共通⁶】ファイル形式は機械が直接読み取れる Excel や CSV 等となっているか

- PDF や Word といった人による閲覧や印刷を目的とした形式では、手作業による再利用のほか、ツールやプログラムによる自動的なデータ取得も困難であり、機械による自動処理が困難で分析前に多大なデータ整形作業が必要となるため、円滑なデータ利活用ができない。このため、データ形式は、機械が直接読み取れる Excel (.xlsx) や CSV、XML、Open Document Format (.ods)、JSON 等とする（データベース形式として公表している場合はこの限りではない）。

⁶ 【共通】は Excel と CSV 共通に適用されるルール

【共通】 1シート（ファイル）に複数の表が掲載されていないか

- 1シート（CSV は1ファイル）に複数の表が掲載されていると、正確な計算や並べ替えといった手作業だけでなく、機械による自動処理も正しく実行できない場合があるため、1つのシートには1つの表のみを掲載すること。
- 例1

修正前		修正後																																																																																																															
1シートに複数の表が掲載されている		1つの表を1シートに分割した状態																																																																																																															
<div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;">Excel シート 1</div> <table border="1"> <thead> <tr> <th>都道府県</th> <th>市区町村</th> <th>合計</th> <th>男</th> <th>女</th> </tr> </thead> <tbody> <tr> <td>東京都</td> <td>千代田区</td> <td>20000</td> <td>10000</td> <td>10000</td> </tr> <tr> <td>東京都</td> <td>中央区</td> <td>6000</td> <td>2000</td> <td>4000</td> </tr> <tr> <td>東京都</td> <td>港区</td> <td>8000</td> <td>3000</td> <td>5000</td> </tr> <tr> <td>東京都</td> <td>新宿区</td> <td>5000</td> <td>4000</td> <td>1000</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>都道府県</th> <th>出荷本数</th> <th>在庫本数</th> </tr> </thead> <tbody> <tr> <td>北海道</td> <td>429</td> <td>141183</td> </tr> <tr> <td>青森</td> <td>321</td> <td>243283</td> </tr> <tr> <td>岩手</td> <td>384</td> <td>333560</td> </tr> <tr> <td>宮城</td> <td>408</td> <td>219724</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>西暦（暦年）</th> <th>出荷本数</th> <th>在庫</th> </tr> </thead> <tbody> <tr> <td>2017年</td> <td>429</td> <td>756</td> </tr> <tr> <td>2018年</td> <td>321</td> <td>648</td> </tr> <tr> <td>2019年</td> <td>384</td> <td>438</td> </tr> <tr> <td>2020年</td> <td>408</td> <td>775</td> </tr> </tbody> </table>		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	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	西暦（暦年）	出荷本数	在庫	2017年	429	756	2018年	321	648	2019年	384	438	2020年	408	775	<div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;">Excel シート 1</div> <table border="1"> <thead> <tr> <th>都道府県</th> <th>市区町村</th> <th>合計</th> <th>男</th> <th>女</th> </tr> </thead> <tbody> <tr> <td>東京都</td> <td>千代田区</td> <td>20000</td> <td>10000</td> <td>10000</td> </tr> <tr> <td>東京都</td> <td>中央区</td> <td>6000</td> <td>2000</td> <td>4000</td> </tr> <tr> <td>東京都</td> <td>港区</td> <td>8000</td> <td>3000</td> <td>5000</td> </tr> <tr> <td>東京都</td> <td>新宿区</td> <td>5000</td> <td>4000</td> <td>1000</td> </tr> </tbody> </table> <div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;">Excel シート 2</div> <table border="1"> <thead> <tr> <th>都道府県</th> <th>出荷本数</th> <th>在庫本数</th> </tr> </thead> <tbody> <tr> <td>北海道</td> <td>429</td> <td>141183</td> </tr> <tr> <td>青森</td> <td>321</td> <td>243283</td> </tr> <tr> <td>岩手</td> <td>384</td> <td>333560</td> </tr> <tr> <td>宮城</td> <td>408</td> <td>219724</td> </tr> </tbody> </table> <div style="border: 1px solid black; padding: 5px;">Excel シート 3</div> <table border="1"> <thead> <tr> <th>西暦（暦年）</th> <th>出荷本数</th> <th>在庫</th> </tr> </thead> <tbody> <tr> <td>2017年</td> <td>429</td> <td>756</td> </tr> <tr> <td>2018年</td> <td>321</td> <td>648</td> </tr> <tr> <td>2019年</td> <td>384</td> <td>438</td> </tr> <tr> <td>2020年</td> <td>408</td> <td>775</td> </tr> </tbody> </table>		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	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	西暦（暦年）	出荷本数	在庫	2017年	429	756	2018年	321	648	2019年	384	438	2020年	408	775
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																																																																																																													
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																																																																																																															
西暦（暦年）	出荷本数	在庫																																																																																																															
2017年	429	756																																																																																																															
2018年	321	648																																																																																																															
2019年	384	438																																																																																																															
2020年	408	775																																																																																																															
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																																																																																																													
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																																																																																																															
西暦（暦年）	出荷本数	在庫																																																																																																															
2017年	429	756																																																																																																															
2018年	321	648																																																																																																															
2019年	384	438																																																																																																															
2020年	408	775																																																																																																															

※修正前後の対照表は各ルールの適用前後の状態を表しており、修正後欄は当該部分のルールを適用した場合の例示である（すべてのルールを満たした場合の例示ではない）。※修正前後の対照表は各ルールの適用前後の状態を表しており、修正後欄は当該部分のルールを適用した場合の例示である（すべてのルールを満たした場合の例示ではない）。以下同様。

例2 CSV ファイルの場合

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	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
<div style="border: 1px solid black; padding: 5px;"> 都道府県, 年度, 人口 (人) 東京都, 2005, 12576601 東京都, 2010, 13159388 神奈川県, 2005, 8791597 神奈川県, 2010, 9048331 </div> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> 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 </div>	<div style="border: 1px solid black; padding: 5px;"> 都道府県, 年度, 人口 (人) 東京都, 2005, 12576601 東京都, 2010, 13159388 神奈川県, 2005, 8791597 神奈川県, 2010, 9048331 </div> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> 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 </div>

【共通】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、機械可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、表の印刷や整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	市区町村	台	台		台	円	円
		導入台数	増減数		材料	仕入単価	仕入額
東京都	千代田区	58406	11291		429	28	12012
東京都	中央区	141183	18421		321	35	11235
東京都	港区	243283	38152		384	42	16128

修正後							
空白列を削除した状態							
都道府県	市区町村	台	台	台	円	円	
		導入台数	増減数	材料	仕入単価	仕入額	
東京都	千代田区	58406	11291	429	28	12012	
東京都	中央区	141183	18421	321	35	11235	
東京都	港区	243283	38152	384	42	16128	

• 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			
		人	人			人	人
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	66680	8274	東京都	千代田区	66680	8274
東京都	中央区	169179	27996	東京都	中央区	169179	27996
東京都	港区	260486	17203	東京都	港区	260486	17203
§				§			
東京都	品川区	422488	35633	東京都	新宿区	349385	15825
東京都	目黒区	288088	10466	東京都	文京区	240069	20345
東京都	大田区	748081	30999	東京都	台東区	211444	13371
§				§			
都道府県	市区町村	人口	増減数	東京都	墨田区	272085	15811
東京都	世田谷区	943664	40318	東京都	江東区	524310	26201
東京都	渋谷区	243883	19350	§			
東京都	中野区	344880	16665	東京都	杉並区	591108	27111
§				東京都	豊島区	301599	10432
東京都	足立区	695043	24921	東京都	北区	355213	14137
東京都	葛飾区	453093	10180	東京都	荒川区	217475	5211
東京都	江戸川区	697932	16634	東京都	板橋区	584483	22567
				東京都	練馬区	752608	30886
				東京都	足立区	695043	24921
				東京都	葛飾区	453093	10180
				東京都	江戸川区	697932	16634

• 例 3 CSV ファイルの場合

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, 導入台数, 増減数	都道府県, 市区町村, 導入台数 (台), 増減数 (台)
東京都, 千代田区, 56406, 11291	東京都, 千代田区, 56406, 11291
東京都, 中央区, 141183, 18421	東京都, 中央区, 141183, 18421
東京都, 港区, 243288, 38152	東京都, 港区, 243288, 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

【共通】 データ本体と無関係な情報がないか

- 表のデータ範囲の外側にメモ等の無関係な情報が残されていると、並べ替えといった手作業だけでなく、機械がデータ範囲を誤認識して自動処理も正しく実行できない場合があるため、データ本体と無関係な情報は完全に削除すること。
- 例

修正前				
商品名	商品コード	個	円	入荷日
ノートパソコン	PC001	25	85,000	2024-03-15
ワイヤレスマウス	MS002	120	2,500	2024-04-01
USBメモリ	USB003	300	1,200	2024-03-28
モニター	MN004	45	22,000	2024-04-05

新モデル、在庫少なめ
人気商品、週末セール対象
大量在庫あり、販促中
一部不良品あり、検品中

修正後				
商品名	商品コード	個	円	入荷日
ノートパソコン	PC001	25	85,000	2024-03-15
ワイヤレスマウス	MS002	120	2,500	2024-04-01
USBメモリ	USB003	300	1,200	2024-03-28
モニター	MN004	45	22,000	2024-04-05

【共通】 すべての列に意味が推測できる項目名が存在するか

- 項目名が存在しない、あるいは内容を推測できない符号等で表記されている場合、人間による内容の理解を妨げるだけでなく、機械が列を正しく識別して処理することもできなくなる。そのため、すべての列には、その内容を明確に示す具体的な項目名を付与すること。
- 例

修正前				修正後			
	項目1	項目2	項目3		度	%	hPa
10月1日	24	45	1,012	10月1日	気温	湿度	気圧
10月2日	24	50	1,010	10月2日	24	45	1,012
10月3日	23	48	1,013	10月3日	24	50	1,010
10月4日	23	47	1,011	10月4日	23	48	1,013
				10月4日	23	47	1,011

【共通】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			修正後			
スペースで体裁を整えている			地域コードを併記した状態			
	本	本		本	本	
地域名	出荷本数	在庫本数	地域コード	地域名	出荷本数	在庫本数
津市	429	756	24201	津市	429	756
四日市市	321	648	24202	四日市市	321	648
伊勢市	384	438	24203	伊勢市	384	438
松阪市	408	775	24204	松阪市	408	775
			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			
	本	本		本	本	
地域名	出荷本数	在庫本数		出荷本数	在庫本数	
三重県津市	429	756		429	756	
三重県四日市市	321	648		321	648	
三重県伊勢市	384	438		384	438	
三重県松阪市	408	775		408	775	

- 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			
	団体	団体	団体		団体	団体	団体
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

- 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	個 出荷 本数	個 在庫	薬剤名	個 出荷 本数	個 在庫
鎮痛剤 A-1	429	756	鎮痛剤A-1	429	756
鎮痛剤 A-2	321	648	鎮痛剤A-2	321	648
鎮痛剤 A-3	384	438	鎮痛剤A-3	384	438
鎮痛剤 A-4	408	775	鎮痛剤A-4	408	775

- 例4 CSV ファイルの場合

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 (本), 在庫本数 (本) 津 <input type="text"/> 市, 429, 756 四日市市, 321, 648 伊 <input type="text"/> 勢 <input type="text"/> 市, 384, 438 松 <input type="text"/> 阪 <input type="text"/> 市, 408, 775	不要なスペースのない状態 地域名, 出荷本数 (本), 在庫本数 (本) 津市, 429, 756 四日市市, 321, 648 伊勢市, 384, 438 松阪市, 408, 775

【共通】 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。そのため、1セル1データの入力とすること。
- 例1

修正前	修正後																								
1セルに複数のデータが入力されている <table border="1"> <thead> <tr> <th></th> <th>全国</th> </tr> </thead> <tbody> <tr> <td>仕入額</td> <td>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)</td> </tr> <tr> <td>出荷額</td> <td>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)</td> </tr> </tbody> </table>		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)	1セル1データとして入力した状態 <table border="1"> <thead> <tr> <th></th> <th>全国 仕入額</th> <th>全国 出荷額</th> </tr> </thead> <tbody> <tr> <td>平成27年度</td> <td>373</td> <td>973</td> </tr> <tr> <td>平成28年度</td> <td>434</td> <td>1234</td> </tr> <tr> <td>平成29年度</td> <td>549</td> <td>1449</td> </tr> <tr> <td>平成30年度</td> <td>638</td> <td>1738</td> </tr> <tr> <td>平成31年度</td> <td>741</td> <td>1841</td> </tr> </tbody> </table>		全国 仕入額	全国 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
	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)																								
	全国 仕入額	全国 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																							

- 例2

修正前	修正後
1セルに複数のデータが入力されている	1セル1データとして入力した状態

	社	千円	
	企業等数	売上金額	(費用総額)
総計	1188389	391445	(355943)
A	50384	69565	(63683)
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)

	社	千円	千円
	企業等数	売上金額	費用総額
総計	1188389	391445	355943
A	50384	69565	63683
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

• 例3

修正前					修正後				
表の左上セルに、セル内改行を用いて表頭と表側の分類を記入している。(斜線は罫線)					セル内改行されていたセルを修正した。				
商品		商品A	商品B	商品C	月	商品A	商品B	商品C	
月					1月	84	11	92	
1月		84	11	92	2月	82	90	42	
2月		82	90	42	3月	27	92	51	
3月		27	92	51					

• 例4 CSV ファイルの場合

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		1フィールド1データとして入力した状態	
区分, 企業等数, 売上金額 (費用総額) 総計, 1188389, 391445 (355943) A, 50384, 69565 (63683) 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)		区分, 企業等数(企業), 売上金額(百万円), 費用総額(百万円) 総計, 1188389, 391445, 355943 A, 50384, 69565, 63683 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	

【共通】機種依存文字を使用していないか

- 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ほか、検索やデータ照合の際に同一の文字として認識されないなど、機械による自動処理の阻害要因となる。そのため、原則として機種依存文字は使用しないこと。

※固有名詞の正確な記述を期する等、やむを得ない事由がある場合はこの限りではない。

- 例 1

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			
	①	②	②／①		1	2	2／1
	千円	千円			千円	千円	
地域	収納額	未収納額	未収納割合	地域	収納額	未収納額	未収納割合
地域A	1000	100	0.1	地域A	1000	100	0.1
地域B	2000	200	0.1	地域B	2000	200	0.1
地域C	3000	300	0.1	地域C	3000	300	0.1
地域D	4000	400	0.1	地域D	4000	400	0.1

- 例 2 CSV ファイルの場合

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	
種類, 収納済額 (①), 収納未済額 (②), 収納未済割合 (②／①)		種類, 収納済額 (1), 収納未済額 (2), 収納未済割合 (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	

【共通】空白とゼロが明確に区別されているか

- データに欠損値（空欄）とゼロ（0）が混在していると、機械は両者を区別できずに処理してしまうことがある。これにより、集計や統計分析において、意図しない数値（たとえば、未回答を0として計算）が混入し、誤った結果を招く可能性。そのため、データ作成者は、値が存在しない場合（欠損値）と、値がゼロである場合を明確に区別して入力すること。

- 例

修正前	修正後
欠損値のセルに0を入れた結果、平均値が正しく計算できていない	欠損値のセルを空白にした結果、平均値が正しく計算できている

	2022	2023	2024	2025	年平均
A市	4	5	6	2	4.25
B市	0	1	0	2	0.75
C市	5	2	7	8	5.50
D市	0	2	0	2	1.00
E町	10	8	9	3	7.50
都市平均	3.8	3.6	4.4	3.4	

	2022	2023	2024	2025	年平均
A市	4	5	6	2	4.25
B市	0	1		2	1.00
C市	5	2	7	8	5.50
D市	0	2	0	2	1.00
E町	10	8	9	3	7.50
都市平均	3.8	3.6	5.5	3.4	

【Excel】オブジェクトを使用していないか

- 画像、図形やテキストボックスといったオブジェクトは、セルのグリッドとは別のレイヤーに存在するため、プログラムはこれを一切認識できない。結果として、オブジェクト内に記載された注釈等の重要な情報が欠落したままデータが利用されることになり、分析の信頼性を著しく損なう。そのため、データの解釈に必要なすべての情報は、オブジェクトとして配置するのではなく、セル内に直接入力すること。
- 例1

修正前				修正後			
オブジェクトを使用している				オブジェクトを削除した状態			
	名称	内容	数量		名称	内容	数量
商品A		サンプル1	100	商品A		サンプル1	100
		サンプル2	200	商品A		サンプル2	200
		サンプル3	300	商品A		サンプル3	300

- 例2

修正前					修正後				
表の左上セルに、セル内改行を用いて表頭と表側の分類を記入している。(斜線はオブジェクト)					セル内改行されていたセルを修正し、オブジェクトを削除。				
	商品	商品A	商品B	商品C		商品A	商品B	商品C	
月					1月	84	11	92	
1月		84	11	92	2月	82	90	42	
2月		82	90	42	3月	27	92	51	
3月		27	92	51					

【Excel】書式でデータの違いを表現していないか

- 太字やセルの色、網掛け、フォントサイズといった書式は、セルの値とは別に管理される視覚情報のため、機械はこれを原則として認識できない。そのため、データの状態や区別といった情報は書式で表現するのではなく、文字情報としてテーブルに直接入力すること。

• 例 1

修正前		修正後		
セルの色で分類の違いを表している		分類を表す列を追加		
商品名	売上 (千円)	商品名	売上	分類
りんご	120,000	りんご	120,000	果物
みかん	85,000	みかん	85,000	果物
大根	72,000	大根	72,000	野菜
きゅうり	50,000	きゅうり	50,000	野菜
キャベツ	110,000	キャベツ	110,000	野菜

• 例 2

修正前		修正後		
予測値を斜体で示している		実績と予測の別を表す列を追加		
年	売上	年	売上	実績と予測の別
2024年	40,970	2024年	40,970	実績値
2025年	65,808	2025年	65,808	実績値
2026年	<i>78,358</i>	2026年	78,358	予測値
2027年	<i>87,604</i>	2027年	87,604	予測値

【Excel】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可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）する必要があります。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を行わず、同一内容を各セルに入力すること。

• 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
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

• 例 2

修正前		修正後		
セルが結合（又は分離）されている		セルの結合を解除した状態		
市区町村	生産本数	市区町村	ふりがな	生産本数
ちよだく	58406	千代田区	ちよだく	58406
千代田区		中央区	ちゅうおうく	141183
ちゅうおうく	141183	港区	みなとく	243283
中央区		港区		
みなとく	243283			
港区				

• 例 3

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			
		導入台数	増減数			導入台数	増減数
都道府県	市区町村	導入台数	増減数	都道府県	市区町村	導入台数	増減数
	千代田区	58406	11291	東京都	千代田区	58406	11291
	中央区	141183	18421	東京都	中央区	141183	18421
	港区	243288	38152	東京都	港区	243288	38152
	}			}			
東京都	葛飾区	442913	327	東京都	葛飾区	442913	327
	江戸川区	681298	2331	東京都	江戸川区	681298	2331
	八王子市	577513	-2540	東京都	八王子市	577513	-2540
	立川市	176295	-3879	東京都	立川市	176295	-3879
	武蔵野市	144730	5986	東京都	武蔵野市	144730	5986
	三鷹市	186936	853	東京都	三鷹市	186936	853
	青梅市	137381	-1958	東京都	青梅市	137381	-1958
	府中市	206274	4766	東京都	府中市	206274	4766
}			}				
	青ヶ島村	—		東京都	青ヶ島村	—	—
	小笠原村	—		東京都	小笠原村	—	—

【Excel】 不要な行や列が非表示のまま残されていないか

- 非表示に設定された行や列は、人間の目には見えないが、機械からは通常の商品データとして読み込まれてしまう可能性がある。これにより、作成者が意図しないデータが混入し、集計結果の誤りやデータ構造の誤解釈を招くなど、分析の信頼性を損なう直接的な原因となる。そのため、公開するデータに不要な行や列は、グループ化による折りたたみも含め、非表示にするのではなく、完全に削除すること。
- 例

修正前		修正後																																																		
グループ化による折りたたみ		折りたたみ、非表示された行を削除																																																		
<table border="1"> <thead> <tr> <th>1</th> <th>2</th> <th>A</th> <th>B</th> </tr> </thead> <tbody> <tr> <td>1</td> <td></td> <td></td> <td>個</td> </tr> <tr> <td>2</td> <td></td> <td>商品名</td> <td>売上個数</td> </tr> <tr> <td>3</td> <td></td> <td>おにぎり</td> <td>120</td> </tr> <tr> <td>4</td> <td></td> <td>サンドイッチ</td> <td>85</td> </tr> <tr> <td>11</td> <td></td> <td>プリン</td> <td>55</td> </tr> <tr> <td>12</td> <td></td> <td>雑誌</td> <td>30</td> </tr> </tbody> </table>		1	2	A	B	1			個	2		商品名	売上個数	3		おにぎり	120	4		サンドイッチ	85	11		プリン	55	12		雑誌	30	<table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> </tr> </thead> <tbody> <tr> <td>1</td> <td></td> <td>個</td> </tr> <tr> <td>2</td> <td>商品名</td> <td>売上個数</td> </tr> <tr> <td>3</td> <td>おにぎり</td> <td>120</td> </tr> <tr> <td>4</td> <td>サンドイッチ</td> <td>85</td> </tr> <tr> <td>5</td> <td>プリン</td> <td>55</td> </tr> <tr> <td>6</td> <td>雑誌</td> <td>30</td> </tr> </tbody> </table>			A	B	1		個	2	商品名	売上個数	3	おにぎり	120	4	サンドイッチ	85	5	プリン	55	6	雑誌	30
1	2	A	B																																																	
1			個																																																	
2		商品名	売上個数																																																	
3		おにぎり	120																																																	
4		サンドイッチ	85																																																	
11		プリン	55																																																	
12		雑誌	30																																																	
	A	B																																																		
1		個																																																		
2	商品名	売上個数																																																		
3	おにぎり	120																																																		
4	サンドイッチ	85																																																		
5	プリン	55																																																		
6	雑誌	30																																																		
<p>行の非表示</p> <table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> </tr> </thead> <tbody> <tr> <td>1</td> <td></td> <td>個</td> </tr> <tr> <td>2</td> <td>商品名</td> <td>売上個数</td> </tr> <tr> <td>3</td> <td>おにぎり</td> <td>120</td> </tr> <tr> <td>4</td> <td>サンドイッチ</td> <td>85</td> </tr> <tr> <td>11</td> <td>プリン</td> <td>55</td> </tr> <tr> <td>12</td> <td>雑誌</td> <td>30</td> </tr> </tbody> </table>			A	B	1		個	2	商品名	売上個数	3	おにぎり	120	4	サンドイッチ	85	11	プリン	55	12	雑誌	30																														
	A	B																																																		
1		個																																																		
2	商品名	売上個数																																																		
3	おにぎり	120																																																		
4	サンドイッチ	85																																																		
11	プリン	55																																																		
12	雑誌	30																																																		

【CSV】 1行1データで表現されているか

- CSV形式において、改行コードはレコードの区切りを示す絶対的な記号であるため、一つのデータ（レコード）の内部に改行が含まれていると、プログラムはそこをレコードの終端と誤認識し、データ構造を正しく読み取れなくなる。そのため、1つのデータは必ず1行で完結させること。
- 例

修正前	修正後
-----	-----

データが複数の行にまたいでいる	データが 1 行に収まった状態
薬剤名, 出荷本数 (本), 単価 (円) 鎮痛剤 A-1, 429, 756 鎮痛剤 A-2, 321, 648	薬剤名, 出荷本数 (本), 単価 (円) 鎮痛剤 A-1, 429, 756 鎮痛剤 A-2, 321, 648

【CSV】 文字列にカンマが含まれているフィールドの値は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 (人) 15 歳以下, 5621 16, 17 歳, 12145 18, 19 歳, 15892 20~25 歳, 19224	文字列にカンマが含まれる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 (人) 15 歳以下, 5621 "16, 17 歳", 2145 "18, 19 歳", 5892 20~25 歳, 19224

レベル 2 : 集計・分析可能

【共通】 数値データは数値属性とし、文字列を含まないこと

- 数値データに単位記号 (円、%、人等)、区切り文字 (カンマ)、脚注記号 (※1、*等) が混在していると、機械は文字列として認識し、計算処理や統計分析が実行できない。また、ピボットテーブルでの集計や BI ツールでの自動グラフ化においても、数値として扱われないため正しい分析結果が得られない。そのため、数値データには数値のみを入力し、単位や注釈は別列に分離すること。
 ※セルの値を数値として保持し、セルの表示形式により記号 (▲等) を表示する場合は、数値データの要件を満たすものとする。
- 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

	円	円	台
	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
SUM関数	0	630	0
+(加算演算)	#VALUE!	#VALUE!	121000

↑黄色の行は関数で合計を表示した例

	円	円	台
	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	8020	-350	37000
サンプル4	7500	500	43000
SUM関数	34650	80	121000
+(加算演算)	34650	80	121000

↑黄色の行は関数で合計を表示した例

- 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		
	台	台		台	台
	仕入台数	在庫台数		仕入台数	在庫台数
サンプル1	5 000	1300	サンプル1	5000	1300
サンプル2	4 000	1800	サンプル2	4000	1800
サンプル3	3 000	1000	サンプル3	3000	1000
サンプル4	2 000	1600	サンプル4	2000	1600

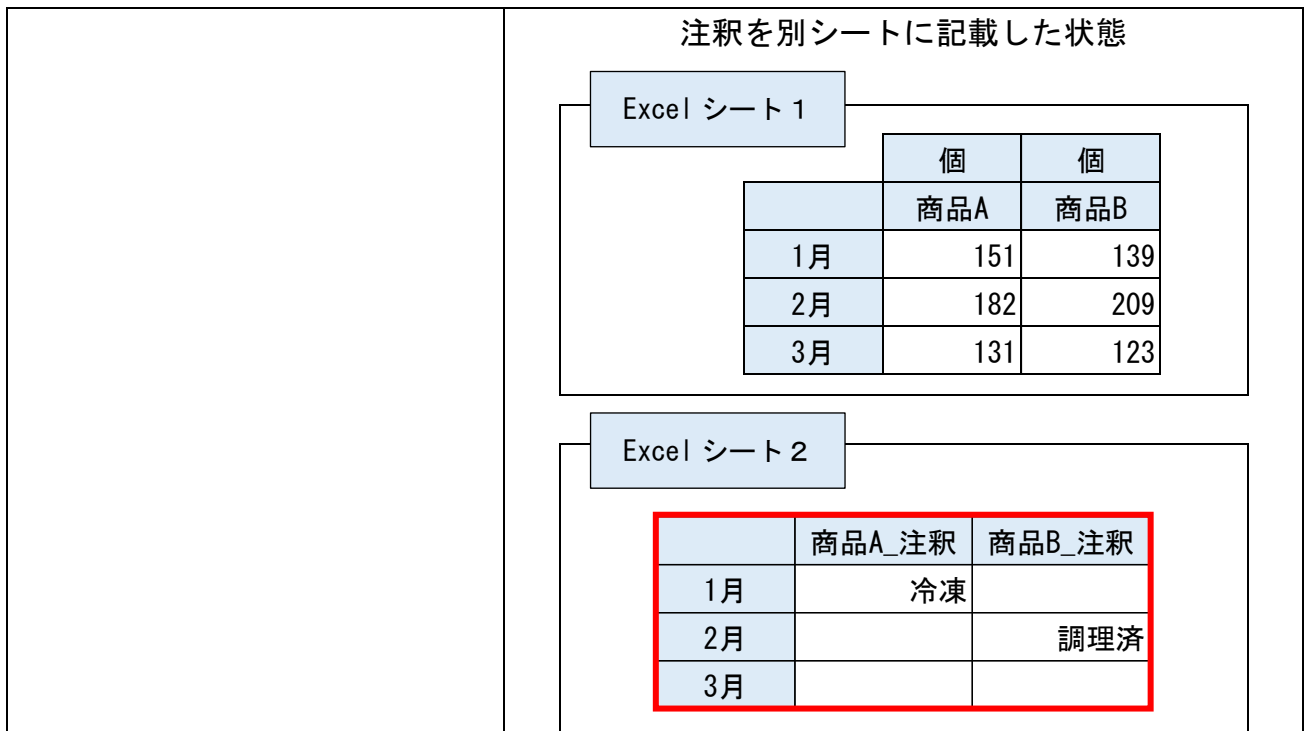
- 例3 数値データ内に注釈・脚注が含まれる場合

修正前			修正後			
セルに注釈が含まれている			注釈を分離した状態			
	個	個	個	個	個	個
	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	151	冷凍	139	
2月	182	209 b)	182		209	調理済
3月	131	123	131		123	

注釈を欄外に記載した状態

	個	個
	商品A	商品B
1月	151	139
2月	182	209
3月	131	123

1月の商品Aは冷凍
2月の商品Bは調理済み



- 例4 数値データ内に文字列が含まれる場合（CSV ファイルの場合）

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
種類, "単価（円）", 前回差分（円）, "生産台数（台）" 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"	種類, 単価（円）, 前回差分（円）, 生産台数（台） 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

- 例5 数値データ内に空白がある場合（CSV ファイルの場合）

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
種類, 仕入台数（台）, 在庫台数（台） 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	種類, 仕入台数（台）, 在庫台数（台） 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6 数値データ内に注釈・脚注が含まれる場合（CSV ファイルの場合）

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	注釈や脚注を除いた状態
仕入月, 商品 A (円), 商品 B (円) 1 月, 151 a) , 139 2 月, 182, 201 b) 3 月, 131, 123	仕入月, 商品 A (円), 商品 A_注釈 , 商品 B (円), 商品 B_注釈 1 月, 151, 冷凍 , 139, 2 月, 182, , 201, 調理済 3 月, 131, , 123,

【共通】データ内での項目名等の省略をしていないか

- 前行と同じ値の場合に項目名や地域名等を省略していると、機械は各行を独立したレコードとして処理するため、省略された項目の値を NULL や空白として扱い、正しいグループ化や集計ができなくなる。また、データの並べ替えを行った際に、省略された情報と元の値との関連性が失われ、データの整合性が保てなくなる。そのため、すべての行において必要な項目値を省略なく入力すること。
- 例 1

修正前	修正後																														
同じ名称を空白で省略している <table border="1"> <thead> <tr> <th>薬剤名</th> <th>出荷本数</th> <th>在庫</th> </tr> </thead> <tbody> <tr> <td>鎮痛剤A-1</td> <td>429</td> <td>756</td> </tr> <tr> <td>2</td> <td>321</td> <td>648</td> </tr> <tr> <td>3</td> <td>384</td> <td>438</td> </tr> <tr> <td>4</td> <td>408</td> <td>775</td> </tr> </tbody> </table>	薬剤名	出荷本数	在庫	鎮痛剤A-1	429	756	2	321	648	3	384	438	4	408	775	省略せずに入力した状態 <table border="1"> <thead> <tr> <th>薬剤名</th> <th>出荷本数</th> <th>在庫</th> </tr> </thead> <tbody> <tr> <td>鎮痛剤A-1</td> <td>429</td> <td>756</td> </tr> <tr> <td>鎮痛剤A-2</td> <td>321</td> <td>648</td> </tr> <tr> <td>鎮痛剤A-3</td> <td>384</td> <td>438</td> </tr> <tr> <td>鎮痛剤A-4</td> <td>408</td> <td>775</td> </tr> </tbody> </table>	薬剤名	出荷本数	在庫	鎮痛剤A-1	429	756	鎮痛剤A-2	321	648	鎮痛剤A-3	384	438	鎮痛剤A-4	408	775
薬剤名	出荷本数	在庫																													
鎮痛剤A-1	429	756																													
2	321	648																													
3	384	438																													
4	408	775																													
薬剤名	出荷本数	在庫																													
鎮痛剤A-1	429	756																													
鎮痛剤A-2	321	648																													
鎮痛剤A-3	384	438																													
鎮痛剤A-4	408	775																													

- 例 2 CSV ファイルの場合

修正前	修正後
同じ名称を空白で省略している 薬剤名, 出荷本数 (本), 単価 (円) 鎮痛剤 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 (本), 単価 (円) 鎮痛剤 A-1, 429, 756 鎮痛剤 A-2 321, 648 鎮痛剤 A-3 384, 438 鎮痛剤 A-4 408, 775

【共通】各列が一意に識別可能な項目名を持っているか

- 同じ項目名が複数の列に存在する場合、機械は列を正しく区別できず、データの集計や結合処理において誤った結果を生成する可能性がある。特に時系列データや階層構造を持つ項目名（年度別売上、地域別人口等）において、複数行にわたって項目名が分割されている場合、各列を一意に識別するためには、それらの行を組み合わせた完全な項目名が必要となる。そのため、機械が単一行の読み取りで各列を明確に識別できるよう、

項目名は一意かつ完結した形で表現すること。

- 例

修正前			
	財務情報	財務情報	財務情報
	売上高 (千円)	営業利益 (千円)	経常利益 (千円)
会社A	120,000	15,000	14,000
会社B	130,000	16,000	15,000
会社C	125,000	15,500	14,500

修正後			
	千円	千円	千円
	財務情報 売上高	財務情報 営業利益	財務情報 経常利益
会社A	120,000	15,000	14,000
会社B	130,000	16,000	15,000
会社C	125,000	15,500	14,500

【共通】選択肢回答が標準化されているか

- 選択肢回答が表記揺れのある文字列で記載されていると、異なるデータセット間での結合や比較分析において正確なマッチングができない。また、従来のプログラム処理では文字列の完全一致判定が困難となる。そのため、選択肢回答は数値コード(「1」、「2」等)または表記揺れのない統一された文字列(「取組済」、「検討中」等)のいずれかで一貫して表記し、データ連携と自動処理を可能にすること。

- 例

修正前		修正後	
表記ゆれのある文字列		表記ゆれのない文字列	
部署	回答	部署	回答
A	取組済	A	取組済
B	取り組み済み	B	取組済
C	取組み済	C	取組済
D	検討中	D	検討中
		数値コードによる表記	
部署	回答	部署	回答
A	1	A	1
B	1	B	1
C	1	C	1
D	2	D	2

【共通】選択肢列と「その他」の詳細記入が分離されているか

- 選択式の項目において「その他」選択肢の詳細内容が同一列に混在していると、機械は定型的な選択肢と自由記述を区別できず、カテゴリカルデータとしての統計処理が正しく実行できない。また、「その他(具体的に:〇〇〇)」といった記載では、機械による自動分類や頻度集計が困難となる。そのため、選択肢は標準化された値のみを格納する

列と、その他の詳細内容を記録する列を分離し、それぞれを独立した変数として扱えるようにすること。

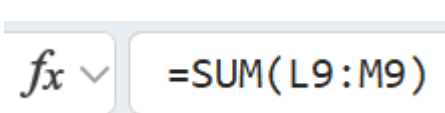
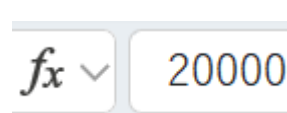
• 例

修正前		修正後		
その他と自由記入の内容が同一セルに記入されている		その他と自由記入の内容のセルが分離されている		
No.	改善点	No.	改善点	その他の内容
1	操作が複雑	1	操作が複雑	
2	重くて持ち運びにくい	2	重くて持ち運びにくい	
3	その他：保証期間が短い	3	その他	保証期間が短い
4	その他：問合せ対応が遅い	4	その他	問合せ対応が遅い

【Excel】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

- セルに数式が残されていると、他の環境でファイルを開いた際に参照エラーや計算エラーが発生し、元のデータと異なる値が表示される可能性がある。また、機械による自動処理において、数式の依存関係を正しく解釈できずにデータ読み込みが失敗したり、意図しない再計算が実行されたりする。そのため、数式を使用して算出した値については、「値の貼り付け」機能を使用して数値データに変換し、数式への依存関係を完全に除去すること。

• 例

修正前					修正後				
合計が数式で入力されている					数値データに修正した状態				
都道府県	市区町村	合計	男	女	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	千代田区	20000	10000	10000
東京都	中央区	6000	2000	4000	東京都	中央区	6000	2000	4000
東京都	港区	8000	3000	5000	東京都	港区	8000	3000	5000
東京都	新宿区	5000	4000	1000	東京都	新宿区	5000	4000	1000
東京都	文京区	7000	5000	2000	東京都	文京区	7000	5000	2000
									

レベル3：連携・自動化可能

【共通】項目名行から始まり、次行からデータ入力がされているか

- ファイルの先頭にタイトルや空白行が存在すると、機械がデータ構造を正しく解釈できず、変数行をデータの一部と誤認識したり、読み込み自体に失敗したりする。そのため、

ファイルは必ず変数行から開始し、2行目からデータを入力すること。

例 1

修正前	修正後
1 行目に項目名が記載されていない	1 行目に項目名が記載された状態
<div style="border: 1px solid black; padding: 5px;"> 東京都, 2005, 12576601 東京都, 2010, 13159388 神奈川県, 2005, 8791597 神奈川県, 2010, 9048331 </div>	<div style="border: 1px solid black; padding: 5px;"> 都道府県, 年度, 人口 (人) 東京都, 2005, 12576601 東京都, 2010, 13159388 神奈川県, 2005, 8791597 神奈川県, 2010, 9048331 </div>

例 2

修正前
項目名が 2 行にまたがっている
<div style="border: 1px solid black; padding: 5px;"> 総数 (人), 男 (人), 女 (人), 総数 (人), 男 (人), 女 (人) 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, 434704, 9048331, 4544545, 4503786 </div>
修正後
項目名が 1 行に記載された状態
<div style="border: 1px solid black; padding: 5px;"> 都道府県, 年度, 総数 (人), 男 (人), 女 (人) 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 </div>

【共通】回答のコード表が別添されているか

- コード表記された選択肢の意味が不明確な場合、データを利用する他の機関や研究者が正しい解釈を行えず、誤った分析結果を生成する可能性がある。また、時系列分析や横断的比較において、コードの定義が統一されていないと、一貫性のある分析が実行できない。そのため、すべてのコード値について、その意味を明確に定義したコード表を別添し、データの正確な解釈と自動処理を可能にすること。
- 例

コード	回答内容	備考
1	徒歩	徒歩のみで通勤
2	自転車	電動自転車含む
3	自家用車	自動車、バイクなど
4	公共交通機関	電車、バス、地下鉄など
5	会社の送迎車	社用車による送迎
6	テレワーク	自宅等での勤務
7	その他	上記以外（船、タクシー等）
9	無回答・不明	回答拒否、記入漏れなど

【共通】数値データの同一列内に特殊記号（秘匿等）を使用する場合は、その定義を明記しているか

- 数値データ列に秘匿記号や調査対象外記号等が混在している場合、その意味が明確でないと、機械は適切な処理を行えず、統計分析や他データセットとの結合において誤った結果を生成する可能性がある。また、省庁横断的なデータ統合において、特殊記号の定義が統一されていないと、データの整合性が保てない。そのため、特殊記号を使用する場合は、その定義と処理方法を明確に文書化し、自動処理プログラムが適切に対応できるようにすること。

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 1

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	***	***

- 例 2 CSV ファイルの場合

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

【共通】データの単位を記載しているか

- データの単位が明記されていない場合、異なるデータセット間での結合や比較分析において、スケールの違いを正しく認識できず、誤った統合や計算が実行される可能性がある。また、AIモデルへのデータ投入において、単位の不整合により学習精度が著しく低下する。そのため、すべての数値データについて、その単位を明確に記載し、プログラムによる自動的なデータ変換と統合を可能にすること。

例1

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

例2 CSVファイルの場合

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

【共通】時間軸の表記が標準化されているか

- 時間軸の表記が統一されていない場合、時系列データの結合や経年変化の自動追跡において、正確な時間順序での処理ができない。また、異なる府省庁のデータを統合する際に、時間軸の不整合により分析の信頼性が損なわれる。そのため、e-Statの時間軸コード表記に準拠するか、西暦表記または和暦に西暦を併記し、プログラムによる自動的な時系列処理を可能にすること。

例1

修正前	修正後
和暦のみで表記している	時間軸コードに和暦を併記した状態

	本	本
年次（暦年）	出荷本数	在庫
平成29年	429	756
平成30年	321	648
平成31年	384	438
令和2年	408	775

		本	本
時間軸コード	和暦（暦年）	出荷本数	在庫
2017000000	平成29年	429	756
2018000000	平成30年	321	648
2019000000	平成31年	384	438
2020000000	令和2年	408	775

西暦で表記した状態

	本	本
西暦（暦年）	出荷本数	在庫
2017年	429	756
2018年	321	648
2019年	384	438
2020年	408	775

西暦に和暦を併記した状態

		本	本
西暦（暦年）	和暦（暦年）	出荷本数	在庫
2017年	平成29年	429	756
2018年	平成30年	321	648
2019年	平成31年	384	438
2020年	令和2年	408	775

• 例2 CSV ファイルの場合

修正前	修正後
和暦のみで表記している	時間軸コードに和暦を併記した状態
年次（暦年），出荷本数（本），在庫本数（本） 平成 29 年 429, 756 平成 30 年 321, 648 平成 31 年 384, 438 令和 2 年, 408, 775	時間軸コード, 和暦（暦年），出荷本数（本），在庫本数（本） 2017000000, 平成 29 年, 429, 756 2018000000, 平成 30 年, 321, 648 2019000000, 平成 31 年, 384, 438 2020000000, 令和 2 年, 408, 775
	西暦で表記した状態
	西暦（暦年），出荷本数（本），在庫本数（本） 2017, 429, 756 2018, 321, 648 2019, 384, 438 2020, 408, 775
	西暦に和暦を併記した状態
	西暦（暦年），和暦（暦年），出荷本数（本），在庫（本） 2017 年, 平成 29 年, 429, 756 2018 年, 平成 30 年, 321, 648 2019 年, 平成 31 年, 384, 438 2020 年, 令和 2 年, 408, 775

【共通】地域コード又は正式な地域名称が表記されているか

- 地域を示すデータにおいて、標準化された地域コードや正式な地域名称が記載されていない場合、地理的データとの結合や地域横断的な比較分析が困難となる。また、県名の省略や順不同の記載により、同一名称の地域で混乱が生じ、GIS システムとの連携や地図データとの自動マッピングが実行できない。そのため、地域データについては、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）を可能な限り使用し、これによらない場合は同コードに定められた正式な地域名称を記載すること。既存の調査独自で地域コードがある場合は当該コードを記載（維持）し、標準地域コードとの重複を避けること。

• 例 1

修正前	修正後
都道府県名を略称している	地域コードを併記した状態

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

同コードで定められた名称を表記した状態

	本	本
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

• 例2 CSV ファイルの場合

修正前	修正後
都道府県名を略称している	地域コードを併記した状態
都道府県, 出荷本数 (本), 在庫本数 (本) 北海道, 429, 141183 青森 321, 243283 岩手 384, 333560 宮城 408, 219724	地域コード, 都道府県, 出荷本数 (本), 在庫本数 (本) "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 (本), 在庫本数 (本) 北海道, 429, 141183 青森県 321, 243283 岩手県 384, 333560 宮城県 408, 219724

【共通】年によってフォーマットが著しく異なっていないか

- 同一の調査や統計において、年度ごとに項目名や構造が大幅に変更されていると、時系列分析や経年変化の自動追跡が困難となる。また、過去データとの結合処理において、プログラムが構造の違いを適切に処理できず、データ統合に失敗する可能性がある。そのため、継続的な調査については、可能な限り項目名もしくは構造の一貫性を保持し、必要な変更がある場合は変更履歴を明記して、自動的な時系列処理を可能にすること。

【共通、CSV】データの定義や更新履歴が記載されているか

- データの項目定義や算出方法、更新頻度、最終更新日等の情報が不明確な場合、異なる時点や組織のデータを結合する際に、データの整合性や比較可能性を正しく判断できない。また、データの信頼性や適用可能性を評価することができず、誤った分析や政策判断につながる可能性がある。そのため、各データについて、項目の定義、算出方法、データソース、更新履歴、留意事項等を明記し、省庁横断的なデータ統合と長期的なデータ管理を可能にすること。また、スクリプト処理を前提とした場合、フィールド物理名は英数字で定義されていること。
- CSV ファイルを介したデータ連携においては、文字コードの不一致は文字化けやシステムエラーの直接的な要因となる。そのため、文字コードの規格及びBOM(Byte Order Mark)の有無を明確に定義すること。なお、選定にあたっては以下の特性を考慮し、連携対象のシステム環境に基づき決定すること。
 - 業務データベース等から出力された UTF-8 形式のデータは、分析や自動処理の分野で広く使われている文字コードであるものの、Windows 標準の環境下において、意図しない表示形式（文字化け）となる恐れがあること。
 - Windows 環境の標準的な表計算ソフトウェアにおいて、既定の設定ではファイル作成・出力時に Shift-JIS が適用される仕様が一般的であること。

【共通】データが縦持ち形式になっているか

- 横持ち形式（年度や項目が列方向に展開された形式）のデータは、新しい時点や項目が追加される度に列構造が変更され、プログラムによる自動処理が困難となる。また、リレーショナルデータベースでの効率的な処理や、統計分析ツールでの標準的な処理ができない。そのため、時系列データや繰り返し項目については、各観測値を独立した行として記録する縦持ち形式とし、データベースでの効率的な処理と自動分析を可能にすること。

修正前				修正後		
横持ち形式				縦持ち形式		
回答者ID	カスタマーサポート	配送速度	Webサイトの使いやすさ	購入手続きの簡便さ	メール通知の頻度	SNSでの情報発信
R001	1	2	4	1	2	4
R002	1	1	1	3	5	4
R003	3	5	1	4	3	2
R004	2	1	1	3	2	3
R005	2	5	2	1	4	2
回答者ID	項目	満足度				
R001	カスタマーサポート	1				
R001	配送速度	2				
R001	ウェブサイトの使いやすさ	4				
R001	製品品質	2				
R001	返品手続き	2				
R001	モバイルアプリの使い勝手	5				
R001	支払い方法の選択肢	3				
R001	在庫の充実度	1				
R001	価格の妥当性	3				
R001	キャンペーンの魅力	5				
R001	問い合わせ対応	3				
R001	レビューの信頼性	1				
R001	商品説明の分かりやすさ	2				
R001	会員制度の特典	3				
R001	店舗スタッフの対応	3				
R001	配送日時の正確さ	2				
R001	梱包の丁寧さ	5				
R001	購入手続きの簡便さ	1				
R001	メール通知の頻度	2				
R001	SNSでの情報発信	4				
R002	カスタマーサポート	1				
R002	配送速度	1				