

# 集計プロセスの適正化【統計の精度向上】 について〈特別テーマ〉

---

平成30年11月14日  
統計委員会担当室

# 集計プロセスの適正化 【統計の精度向上】 <特別テーマ>

個票段階及び集計段階において、一般的に行われることの多いエラーチェックの手法について、実施状況を点検

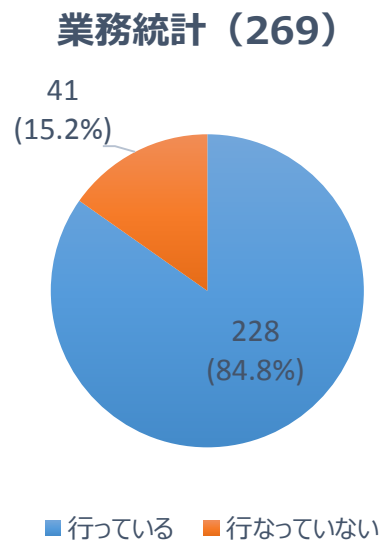
## (1) 個票データ等の入力後チェック<sup>(注)</sup>の実施状況

(注) 回答の有無(記入漏れ)、回答方法の誤り(単独チェック)、回答内容の矛盾(関連チェック)等

### ① 実施の有無

- 一般統計は、対象統計(41※)の**全てで「行なっている」と回答**
- 業務統計は、対象統計(269※)のうち、「**行なっていない」との回答が41統計(15.2%)**

	行なっている	行なっていない
一般統計	41	0
N=41	100	0
業務統計	228	41
N=269	84.8	15.2
合計	269	41
N=310	86.8	13.2



※今回の棚卸し対象統計は、一般統計37、業務統計244であるが、1つの統計でも、年報・月報等で業務プロセスの異なる統計は、それぞれの業務プロセスごとに確認、集計を行なったため、「N」は棚卸し対象の統計数と一致しない

個票データ等の入力後チェックを行っていない主な理由は以下の通り

## ○ 他の機関等で作業を行なっているため、把握していない等（15統計）

### ・民間委託しているもの

- 「民間業者へ委託しているため」
- 「集計者によりデータの集計を行うため」
- 「仕様書に記載がないため不明」

### ・他の行政機関等で作業を行なっているもの

- 「市区町村の業務であり詳細については把握していないため」
- 「個票データ等の入力業務は〇〇が行い、集計結果を送付してもらっている。個票データ等は国に送付されず、入力後のチェックは行っていない」



何らかのチェックが実施されている可能性はあるが、統計作成者が実態を把握しておらず、適切に実施されているとはいえない。

- ・民間業者の判断で実施している場合は、業者が変更した場合は行われなくなる
- ・市町村や地方支分部局ごとの判断で行っている場合は、地域によるバラつきが発生


チェックが安定的に行われるよう、作業実施機関に対して仕様書や手順書等により行うべき作業を示す必要がある。

## ○ 業務処理過程で入力されたデータから統計を作成しているため（23統計）

「業務システムへ入力された申告書等の個票データを基に集計を行っており、業務システムへの入力及び入力後のチェックについて地方支分部局における業務処理の過程で行われることから、統計作成部署では行っていないため」

「業務情報システム上で一元管理しているものであり、入力データは自動集計しているため」

「報告書作成業務でシステムに入力したデータを抽出しており、改めて確認する必要がないため」

 業務処理過程で仕様書や手順書等により適切なチェックが行われているか確認し、行われていない場合は、統計の精度確保のため、適切なチェックを実施する必要がある。

## ② 個票データ等の入力後チェックの実施状況（スコアリング）

- 一般的にデータの誤りを除去するために行なうとされる以下のチェックの実施状況を  
書面調査にて確認した（複数回答可）
  - ①回答の有無（記入漏れ）
  - ②回答方法の誤り（単独チェック）
  - ③回答内容の矛盾（関連チェック）
  - ④①～③以外のチェック（回答重複確認、他の機関等での事前確認など）
- 入力後チェックの実施状況を整理するため、以下の基準に当てはめて、スコア化した（※）
  - スコア 0 入力後のチェックは**未実施**
  - スコア 1 ①～④までのチェック項目の実施が**1項目**
  - スコア 2 ①～④までのチェック項目の実施が**2項目**
  - スコア 3 ①～④までのチェック項目の実施が**3項目以上**

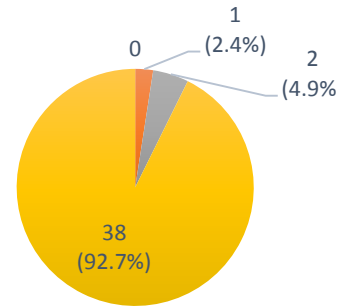
※スコア化にあたっては、書面調査の回答票の結果から作成したものであり、個々の統計の特性や業務事情は考慮していない

# 集計プロセスの適正化 【統計の精度向上】 <特別テーマ>

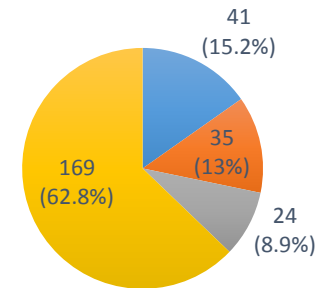
- 一般統計は、今回の棚卸し対象統計(41※)のうち、**スコア3が38統計 (92.7%)**
- 業務統計でも、今回の棚卸し対象統計(269※)のうち、**スコア3が169統計 (62.8%)**  
と最も多いが、**スコア0とスコア1の合計も76統計 (28.2%)**

	スコア0	スコア1	スコア2	スコア3
一般統計	0	1	2	38
N=41	0	2.4	4.9	92.7
業務統計	41	35	24	169
N=269	15.2	13.0	8.9	62.8
合計	41	36	26	207
N=310	13.2	11.6	8.4	66.5

一般統計 (41)



業務統計 (269)



■スコア0 ■スコア1 ■スコア2 ■スコア3 ■スコア0 ■スコア1 ■スコア2 ■スコア3

※今回の棚卸し対象統計は、一般統計37、業務統計244であるが、1つの統計でも、年報・月報等で業務プロセスの異なる統計はそれぞれの業務プロセスごとに確認、集計を行なったため、「N」は棚卸し対象の統計数と一致しない

## ③ 項目別の実施状況

- 一般統計での入力後チェックの実施上位は、③回答内容の矛盾が41統計（100%）、②回答方法の誤りが40統計（97.6%）
- 業務統計での入力後チェックの実施上位は、①回答の有無が206統計（76.6%）、③回答内容の矛盾が192統計（71.4%）

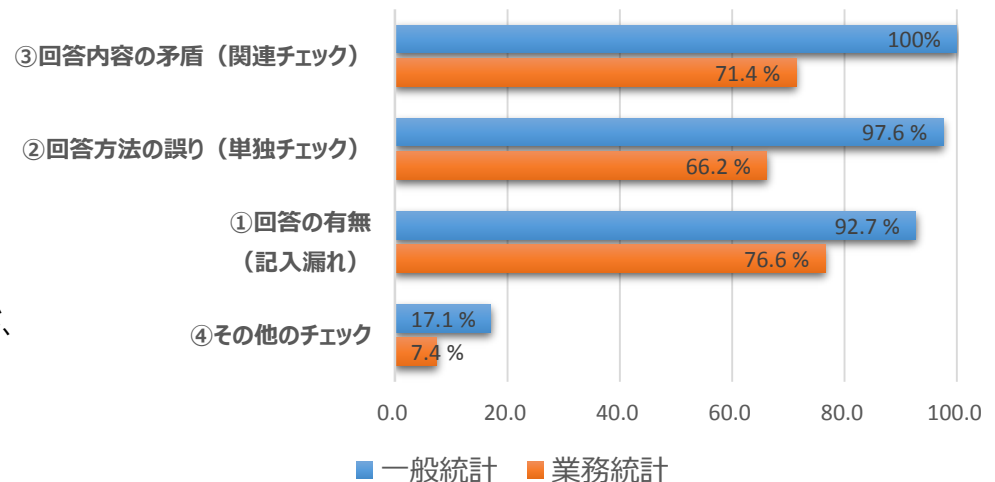
	①回答の有無	②回答方法の誤り	③回答内容の矛盾	④その他のチェック
一般統計	38	40	41	7
N=41	92.7	97.6	100	17.1
業務統計	206	178	192	20
N=269	76.6	66.2	71.4	7.4
合計	244	218	233	27
N=310	78.7	70.3	75.2	8.7

※今回の棚卸し対象統計は、一般統計37、業務統計244であるが、1つの統計でも、年報・月報等で業務プロセスの異なる統計はそれぞれの業務プロセスごとに確認、集計を行なったため、「N」は棚卸し対象の統計数と一致しない

※①から④は、複数回答可としているため、下部の数字を合計しても100%とはならない



統計種別ごとの実施状況

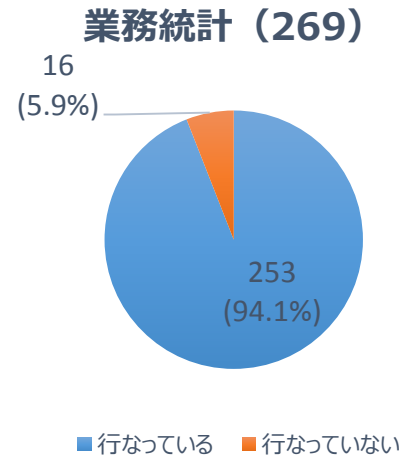
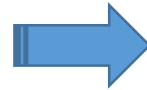


## (2) 集計された統計表（又は統計データ）の内容チェックの実施状況

### ①実施の有無

- 一般統計は、今回棚卸しの対象とした統計(41※)の**全てで集計された統計表（又は統計データ）の内容チェックを行なっていると回答**
- 業務統計は、今回棚卸しの対象とした統計(269※)のうち、**集計された統計表（又は統計データ）の内容のチェックを行っていないと回答したものが16統計（5.9%）**

	行なっている	行なっていない
一般統計	41	0
N=41	100	0
業務統計	253	16
N=269	94.1	5.9
合計	294	16
N=310	94.8	5.2



行なっていない理由については確認中

※今回の棚卸し対象統計は、一般統計37、業務統計244であるが、1つの統計でも、年報・月報等で業務プロセスの異なる統計はそれぞれの業務プロセスごとに確認、集計を行なったため、「N」は棚卸し対象の統計数と一致しない



## ② スコアリング

○一般的に集計された統計表の正確性を高めるために行なうとされる以下のチェックの実施状況を書面調査にて確認した（複数回答可）

- ①表内検算
- ②表間照合（表間で論理矛盾等がないか）
- ③様式チェック（結果表が所定の様式どおりになっているか）
- ④時系列チェック（過去の当該統計との比較）
- ⑤関連統計との比較（社会情勢を反映したのものになっているか）
- ⑥その他のチェック
- ⑦行なっていない

○内容チェックの実施状況を整理するため、以下の基準に当てはめて、スコア化した（※）

スコア 0 集計された統計表（又は統計データ）の内容チェックは**未実施**

スコア 1 ①～⑥までのチェック項目の実施が**1項目**

スコア 2 ①～⑥までのチェック項目の実施が**2～3項目**

スコア 3 ①～⑥までのチェック項目の実施が**4項目以上**

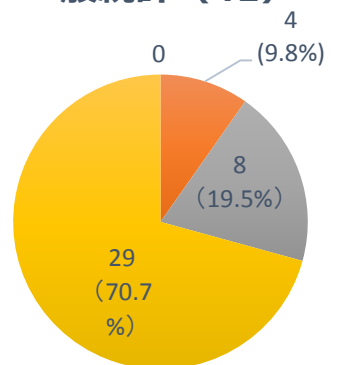
※スコア化にあたっては、書面調査の回答票の結果から作成したものであり、個々の統計の特性や業務事情は考慮していない

# 集計プロセスの適正化 【統計の精度向上】 <特別テーマ>

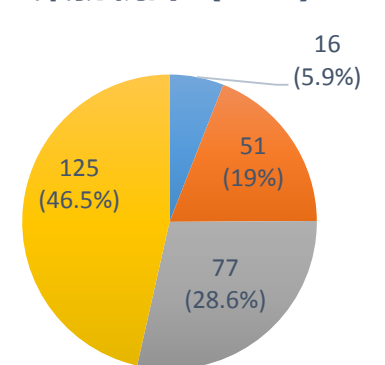
- 一般統計は、今回棚卸しの対象とした統計のうち、**スコア3が29統計（70.7%）**
- 業務統計は、今回棚卸しの対象とした統計のうち、**スコア3が125統計（46.5%）**、**スコア2が77統計（28.6%）**

	スコア0	スコア1	スコア2	スコア3
一般統計	0	4	8	29
N=41	0	9.8	19.5	70.7
業務統計	16	51	77	125
N=269	5.9	19.0	28.6	46.5
合計	16	55	85	154
N=310	5.2	17.7	27.4	49.7

一般統計（41）



業務統計（269）



■スコア0 ■スコア1 ■スコア2 ■スコア3

■スコア0 ■スコア1 ■スコア2 ■スコア3

※今回の棚卸し対象統計は、一般統計37、業務統計244であるが、1つの統計でも、年報・月報等で業務プロセスの異なる統計はそれぞれの業務プロセスごとに確認、集計を行なったため、「N」は棚卸し対象の統計数と一致しない

- 個票データの入力後チェック、統計表の内容チェックのいずれについても、業務上のルールとしてマニュアル等への記載状況を追加で確認中（安定的・継続的に行われているものか、現在の担当者や受託業者の判断で行われているものか）

# 集計プロセスの適正化 【統計の精度向上】 <特別テーマ>

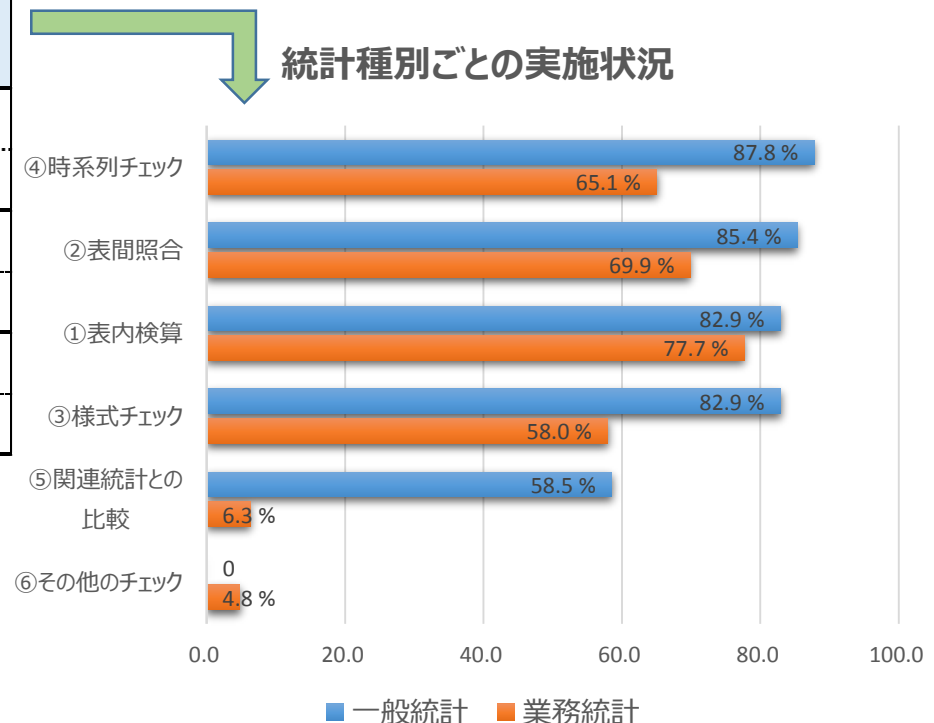
## ③ 項目別の実施状況

- 一般統計での内容チェックの実施上位は、④時系列チェック36統計（87.8%）、②表間照合（85.4%）となっている
- 業務統計での内容チェックの実施上位は、①表内検算209統計（77.7%）、②表間照合（69.9%）となっている

	①表内検算	②表間照合	③様式チェック	④時系列チェック	⑤関連統計との比較	⑥その他のチェック
一般統計	34	35	34	36	24	0
N=41	82.9	85.4	82.9	87.8	58.5	0
業務統計	209	188	156	175	17	13
N=269	77.7	69.9	58.0	65.1	6.3	4.8
合計	243	223	190	211	41	13
N=310	78.4	71.9	61.3	68.1	13.2	4.2

※今回の棚卸し対象統計は、一般統計37、業務統計244であるが、1つの統計でも、年報・月報等で業務プロセスの異なる統計はそれぞれの業務プロセスごとに確認、集計を行なったため、「N」は棚卸し対象の統計数と一致しない

※①から⑥は、複数回答可としているため、下部の数字を合計しても100%とはならない



## 【参考資料】個票データ等の入力後チェック方法の基本

チェックの方法	チェックの目的
回答の有無（記入漏れ）	個々の調査事項について記入の有無を点検すること
回答方法の誤り（単独チェック）	調査事項についての正否を判定し、補正すること (例) ・単一選択法で複数回答となっていないか ・定められた符号以外の記入がないか →規定コード以外 3（本来、男性1、女性2） ・数値が許容範囲内（想定した上限と下限の間）であるか
回答内容の矛盾（関連チェック）	調査事項の相互の関連からみて、矛盾した記入がないかどうか、その妥当性を点検すること (関連する事項間での矛盾や不合理の関係を審査) (例) ・配偶者が未婚 ・会社組織で家族従事者あり ・居住室数が3室で畳数の合計が6畳未満

©公益財団法人 統計情報研究開発センター作成 「統計実務基礎知識」より引用

# 集計プロセスの適正化 【統計の精度向上】 <特別テーマ>

## 【参考資料】集計された統計表（又は統計データ）の内容チェック方法の基本

チェックの方法	チェックの目的
表内検算	集計や製表過程において生じた誤りがないかどうかを確認するため、検算を行なうこと
表間照合 (表間で論理矛盾等がないか)	集計や製表過程において生じた誤りがないかどうかを確認するため、表間における結果数値の監査を行なうこと  (例) ・表間で相互の数を確認 (A表総数 = B表総数 = C表総数) ・表間で組み合わせた項目に矛盾がないか (会社役員 ≠ 雇用者) ・表間で項目相互の数が一定の相対的關係にあるか (夫婦のみの世帯 ≠ 世帯人員3名)
様式チェック (結果表が所定の様式どおりになっているか)	所定の様式に見合う結果表となっているかを確認すること  (例) ・結果表の枚数確認 ・表題、表頭、表側、脚注の表現の確認 ・行ずれ、欄ずれはないか、字間隔は正しいかの確認
時系列チェック (過去の当該統計との比較)	出てきた結果数値の持つ意味について理論的な立場から価値判断を行なう、あるいはその信頼性を検討するため、統計数値を時間的経過 (年・月) に従って比較し、趨勢的变化を審査すること
関連統計との比較 (社会情勢を反映したものとなっているか)	出てきた結果数値の持つ意味について理論的な立場から価値判断を行なう、あるいはその信頼性を検討するため、単一の項目だけでなく、複数項目を組み合わせた内容について、同種統計や関連統計と比較検討すること