

# 厚生労働省で提供する オーダーメイド集計及び 匿名データについて

平成23年度

公的統計のミクロデータの利用に関する研究集会

統計数理研究所 3階セミナー室2 (D304号室)

平成23年11月11日 (金)

久住和弘 (厚生労働省 大臣官房統計情報部)

# 統計法の改正経緯

- 1945年 終戦
- 1947年 (旧) 統計法の施行
- 1952年 統計報告調整法の施行

～ 日本における統計を巡る環境変化 ～  
(旧統計法から60年)

⇒ **統計法の全部改正へ**

- 2007年 5月 新統計法の公布
- 2009年 4月 新統計法の全面施行

# 旧統計法の課題

- ① 産業構造の変化（経済活動における第三次産業のウエイトの上昇等）への対応
- ② 調査環境の変化（調査対象者のプライバシー意識の高まり等）への対応
- ③ 情報処理技術の発達による統計情報への新たなニーズへの対応

⇒ 二次的利用制度の創設等

# 調査票情報の提供と二次的利用

1 法第32条 (二次利用)

⇒ 省内における調査票情報の統計利用

2 法第33条 (調査票情報の提供)

⇒ 高度な公益性が認められる学術研究等

(例：科学研究費補助金) での調査票情報の統計利用

3 法第34条 (委託による統計の作成)

．．．．． オーダーメイド集計

4 法第35条、第36条

(匿名データの作成、提供)

(新設規定)  
**二次的利用**

# 二次的利用制度の拡大

【旧法下】

- 一般的学術研究
  - ・ 研究論文
  - ・ 学位論文 等
- 高等教育
  - ・ 講義、実習 等

- 高度な公益性
    - ・ 行政目的
    - ・ 公益性の高い学術研究
- ＜調査票情報の提供＞

- 営利目的
  - ・ 商業利用
  - ・ 商品開発

調査票情報の利用が認められなかった一般的な学術研究等にも統計利用の裾野を拡大

(変更なし)

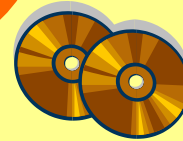
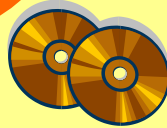
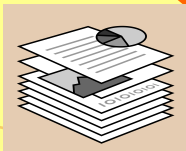
(変更なし)

【新法下】

オーダーメイド集計 (法第34条)

匿名データ (法第35, 36条)

(法第33条)



適正管理義務、守秘義務、提供を受けた目的以外の利用・  
第三者提供の禁止、罰則の新設

# 匿名データとは

【 統計法による定義 】（第2条第12項）

「一般の利用に供することを目的として調査票情報を**特定の個人又は法人その他の団体の識別**（他の情報との照合による識別を含む。）**ができないように加工したもの**をいう。」

<よくある誤解>

① 「名前と住所だけ消せば匿名データ」ではない！

⇒ 外観から識別できる情報（世帯人員、世帯の構成、階数）や一般に入手可能な情報（資本金、従業者規模、売上高）等の組み合わせにより被調査者等が識別されないように秘匿する必要があるため、匿名データの作成は容易ではない。  
また、秘匿は確実に行う一方、秘匿しすぎると有用性が低減してしまう。

② 「匿名データ」 ≠ 「調査票情報」

⇒ いずれも Micro data だが、利用手続、提供基準、情報量、秘匿度、有用性、守秘義務規定、罰則等が大きく異なる。

③ 「匿名データ」 ≠ 「Public Use File」（アメリカのセンサス局等が提供）

# 平成16年国民生活基礎調査

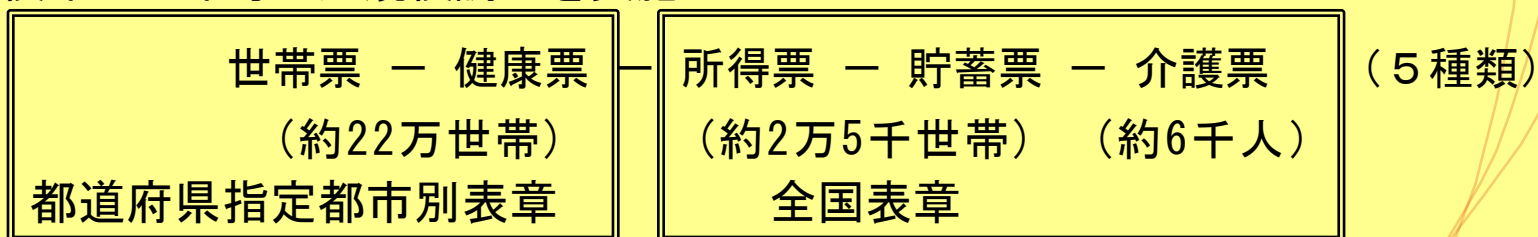
## 1 調査の目的

保健、医療、福祉、年金、所得等国民生活の基礎的事項を調査し、厚生労働省の所掌事務に関する政策の企画及び立案に必要な基礎資料を得る。

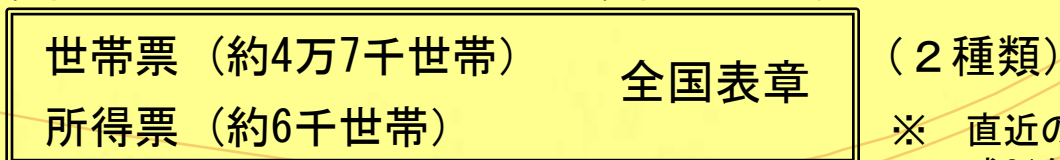
## 2 調査の周期

昭和61年以降毎年実施

(1) 大規模年：3年毎に大規模調査を実施 ⇒ 匿名データ化



(2) 中間年：大規模調査の中間の2年間に小規模調査を実施



※ 直近の中間年は  
平成21年

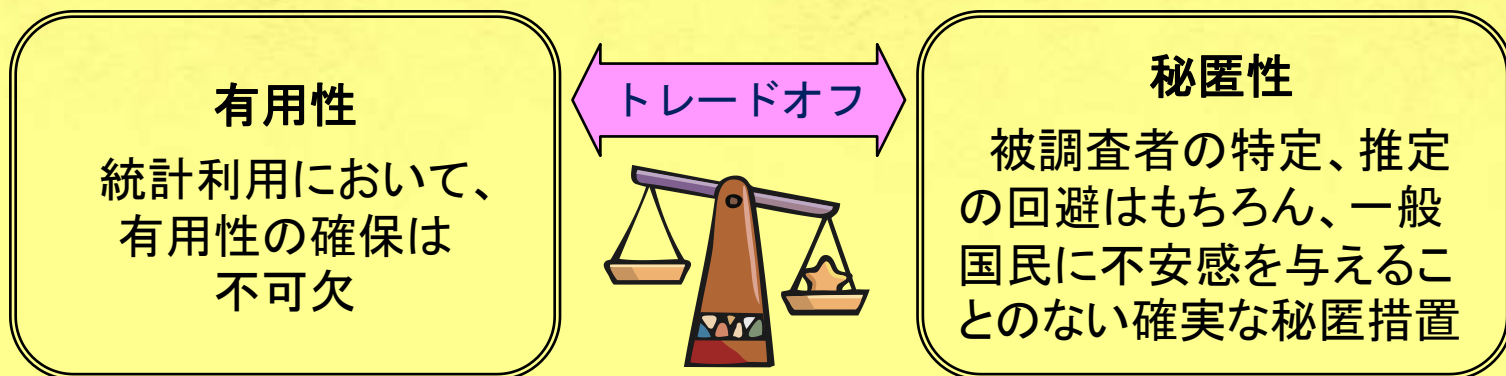
# 調査項目

- 世帯票：** 性、出生年月、世帯主との続柄、配偶者の有無、公的年金・恩給の受給状況、住居の状況、家計支出額 等
- 健康票：** 傷病の状況、通院・通所の状況、悩みやストレスの状況、健康診断受診状況 等
- 所得票：** 所得の種類別金額、課税等の状況、生活意識の状況 等
- 貯蓄票：** 貯蓄現在高、貯蓄の増減の状況、借入金残高
- 介護票：** 要介護度の状況、介護が必要となった原因、居宅サービスの利用状況、介護者の状況 等

# 国民生活基礎調査の匿名データ化

本調査は、**統計法**で定める**基幹統計調査**として、学術研究や高等教育においても利用ニーズの高い世帯調査。

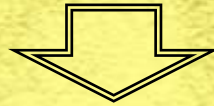
集落抽出である特性、**複数の調査票**により**多様な項目**を把握しており、それらを**接続して匿名データ化**すること等から生じる**個体識別リスク**を踏まえた、総合的な秘匿措置の確保が必要。



- 1 本調査の匿名データ化は、初回であることを踏まえ、データの有用性を考慮しつつ、**相対的には秘匿性を優先して作成**
- 2 **大規模調査年である平成16年調査**について、平成23年9月から**提供を開始**し、他の大規模調査年について、順次提供年次を拡大予定
- 3 有用性を勘案して、**データをリンケージした2種類のデータセットを提供**
- 4 秘匿性及び有用性の検討の結果、**サンプルサイズは概ね中間年並みを確保**

# 調査の対象と匿名データ

平成12年国勢調査の調査地区 (約94万地区)



後置番号「1」と「8」のみ抽出

国民生活基礎調査の調査地区を抽出するための調査地区



(約90万地区)  
層化無作為抽出

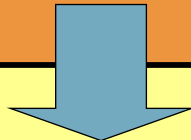
平成16年国民生活基礎調査

世帯票・健康票 (5,280地区、約10,000単位区)

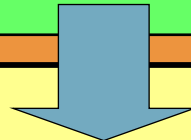


後置番号「1」のみ抽出

所得票・貯蓄票  
(2,000単位区)



匿名  
データ  
A



匿名  
データ  
B

調査の対象：

調査区内の全ての世帯及び世帯員

※ 単位区は、一つの国勢調査区(50～60世帯)を概ね20～30世帯毎に地理的に分割したもの。

# 提供する匿名データの構成概要

## (平成16年国民生活基礎調査)

匿名データの種類	匿名データの構成 (接続する調査票)	リサンプリング率 ( $\frac{\text{提供レコード件数}}{\text{調査本体のサンプルサイズ}}$ )	想定される利用	ウェイト
A	世帯票、健康票	約2割 ( $\frac{\text{約4万世帯}}{\text{約22万世帯}}$ )	人口、社会統計分野での <b>世帯数の推計分析等</b> を中心とした利用	全国推計世帯数の算出が可能となるよう、全国一律のウェイトを付与
B	世帯票、健康票、所得票、貯蓄票	約2割 ( $\frac{\text{約6千世帯}}{\text{約2万5千世帯}}$ )	世帯の <b>所得及び貯蓄に関する分析等</b> を中心とした利用 (推計世帯数の算出は前提としていない)	ウェイトは付与しない (ただし、各レコードのウェイトは一律となるようリサンプリング)

※1 秘匿性を確保するため、地域区分は全国のみとする。

※2 介護票はサンプルサイズが小さいため匿名データ化しない。

# 各提供データの構成

○ 1レコード1世帯員の構成だが、「世帯」への統合が可能。

データA

世帯番号	世帯員番号	項目				
001	01	001世帯共通 (世帯票)	...	01世帯員	...	ウエイ
001	02	"	...	02世帯員	...	ウエイ
001	03	"	...	03世帯員	...	ウエイ
001	04	"	...	04世帯員	...	ウエイ
002	01	002世帯共通 (世帯票)	...	01世帯員	...	ウエイ
003	01	003世帯共通 (世帯票)	...	01世帯員	...	ウエイ
003	02	"	...	02世帯員	...	ウエイ
003	03	"	...	03世帯員	...	ウエイ
003	04	"	...	04世帯員	...	ウエイ
003	05	"	...	05世帯員	...	ウエイ
003	06	"	...	06世帯員	...	ウエイ
004	01	004世帯共通 (世帯票)	...	01世帯員	...	ウエイ
004	02	"	...	02世帯員	...	ウエイ
004	03	"	...	03世帯員	...	ウエイ
004	04	"	...	04世帯員	...	ウエイ
004	05	"	...	05世帯員	...	ウエイ

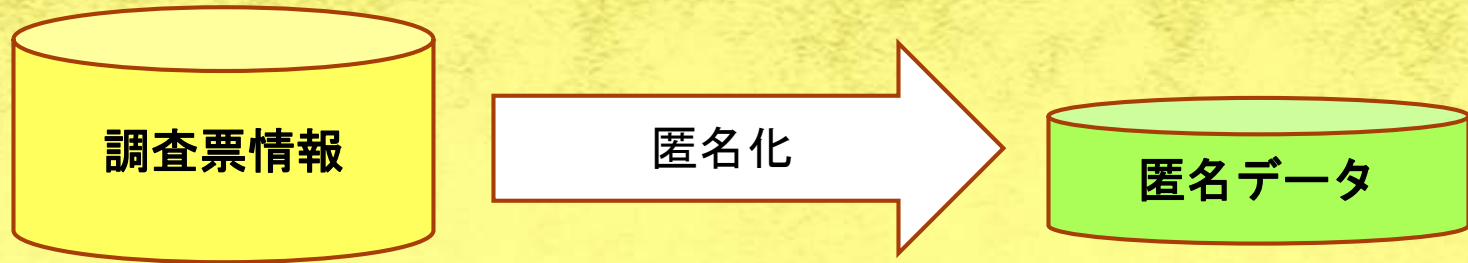
データB

世帯番号	世帯員番号	項目					
001	01	001世帯共通 (世帯票)	...	001世帯共通 (所得票、貯蓄票)	...	01世帯員	...
001	02	"	...	"	...	02世帯員	...
002	01	002世帯共通 (世帯票)	...	002世帯共通 (所得票、貯蓄票)	...	01世帯員	...
002	02	"	...	"	...	02世帯員	...
002	03	"	...	"	...	03世帯員	...
002	04	"	...	"	...	04世帯員	...
002	05	"	...	"	...	05世帯員	...
002	06	"	...	"	...	06世帯員	...
003	01	003世帯共通 (世帯票)	...	003世帯共通 (所得票、貯蓄票)	...	01世帯員	...
003	02	"	...	"	...	02世帯員	...
003	03	"	...	"	...	03世帯員	...
003	04	"	...	"	...	04世帯員	...
003	05	"	...	"	...	05世帯員	...
004	01	004世帯共通 (世帯票)	...	004世帯共通 (所得票、貯蓄票)	...	01世帯員	...

※ 世帯員の項目は、世帯票及び健康票のみ

# 匿名化技法の概要

単に氏名・住所を削除しただけでは、年齢、家族構成、職業分類、住居形態等の調査項目と、**外観から確認できる情報や一般に入手可能な情報の組み合わせ等により、被調査者が特定されるリスクがある**ため、以下の匿名化技法により、被調査者の情報を確実に秘匿する。



## 1 情報の削除

- (1) データの再抽出 (リサンプリング) ..... 元の統計調査のデータすべてを匿名データに用いるのではなく、そのうちの何割かを再抽出したデータを用いる。
- (2) 識別情報の削除等 ..... 識別情報は、データから全面的に削除する(例: 氏名、地域等を削除)。また、データの配列順により特定されないように、無作為に並べ替えを行う。
- (3) 裾切りによるデータ削除 ..... 特徴的で、出現率が低い値があるデータは、削除する(例: 8人以上世帯等を削除)

## 2 識別情報の階級区分統合

- (1) 上限(下限)階級区分の統合(トップ(ボトム)コーディング) ..... 極端に大きな(小さな)値は、上限(下限)値を設けて統合する。(例: 85歳以上を統合等)
- (2) 再コード化(リコーディング) ..... 分類事項の程度は、詳細なものではなく、粗いものとする(例: 各歳階級を5歳階級化、小分類を中分類化)

# 匿名データ利用手続の流れ

【 申出者 】



個票を利用して、報告書にない分析をやりたい。

利用方法、申出書の記載内容等について相談

提供  
依頼  
申出書

依頼書



集計・分析

成果の  
公表

適正に  
管理

匿名データ、中間生成物を消去

利用  
実績  
報告書

データ  
消去等  
報告書

提出、返却

【 厚生労働省 】



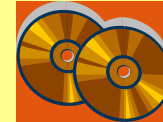
受付、審査

諾否の通知

受理、確認

提出、納付

匿名データの提供（貸与）



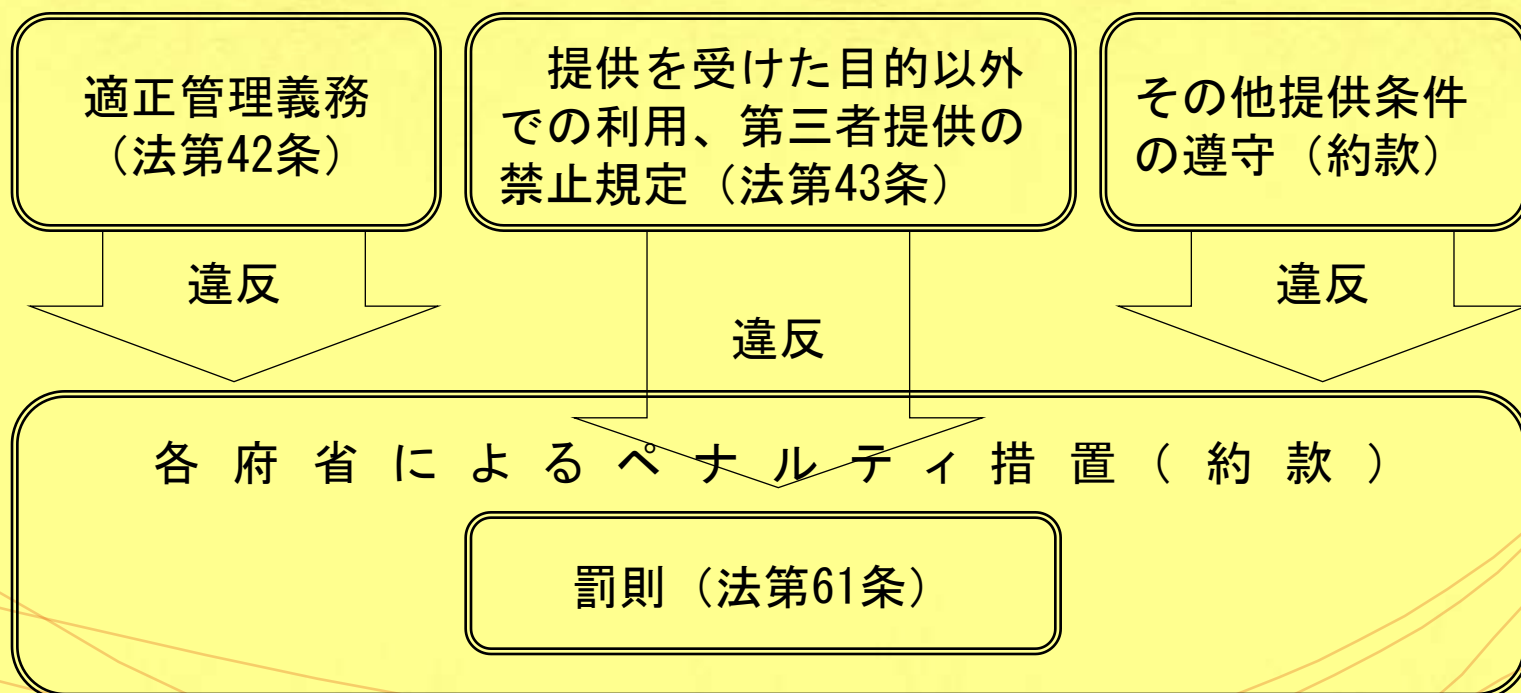
受理、確認

# 匿名データ利用上の注意

- 1 **「貸与」**なので、返却いただく必要があります。
- 2 特に高等教育目的等、調査票情報等の個票を扱ったことがない利用者に対しては**統計倫理的な教育**を行ってください。
- 3 レコード件数のみならず変数も多い大量データであるため、**データ処理**に**習熟**している**必要**があります。
- 4 秘匿措置を施しているため、**集計結果は必ずしも公表されている当該統計又は報告書とは一致しません。**  
秘匿措置等を十分理解した上で利用してください。
- 5 **研究成果**は**公表**してください。  
(成果が社会に還元されることを前提に提供する制度です。)

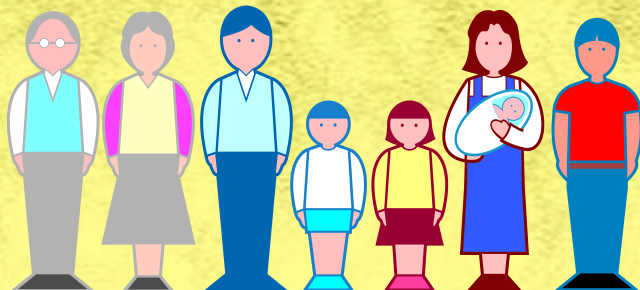
# 匿名データの適正管理義務、 第三者提供の禁止と罰則、ペナルティ

匿名データの提供を受けた方、業務の委託を受けた方等には



# 制度の原点.....国民の信頼確保

基幹統計調査には、  
報告義務と罰則規定



- ・ 個人情報保護意識の高まり
  - ・ 調査環境の悪化等
- ⇒ 回収率は年々低下傾向

信頼・協力

- ・ 二次的利用制度の普及、利用拡大

調査票（個票）情報

- ・ 適正管理義務
- ・ 守秘義務規定
- ・ 罰則規定



【 万が一、秘匿が破られると... 】

- ・ 国民の政府統計に対する信頼が損なわれ、
- ⇒ 回収率の低下  
⇒ 更には「根拠に基づく政策」の推進に影響が生じる。

- ・ 社会の発展に寄与することを通じて国民に還元

有用性

トレード  
オフ

秘匿性



- ・ 国民の信頼を維持できる確実な秘匿措置、利用者における適切な利用・管理が不可欠

【 統計制度、二次的利用制度に対する国民の理解の広がり 】

# オーダーメイド集計

統計表作成を希望する**申出者が**、

- 1 対象調査について、厚生労働省が提示する分類項目の中から**利用したい項目を選択**
- 2 選択した項目の組み合わせによる統計表の作成を厚生労働省へ**委託**
- 3 作成された統計表の**提供を受ける**。

## 【 現在提供を行っている調査 】

- ・ 平成19、20年人口動態調査（出生票、死亡票）
- ・ 平成20年医療施設（静態）調査
- ・ 平成20年患者調査
- ・ 平成21、22年毎月勤労統計調査（特別調査票）
- ・ 平成18、19年賃金構造基本統計調査（個人票）

## 【 平成23年度中に提供開始予定の調査 】

- ・ 平成20年賃金構造基本統計調査（個人票）

# オーダーメイド集計の留意事項

- 1 「オーダーメイド」といっても**全ての分類項目を使い自由に作表出来るわけではありません**。対象調査ごとに使用出来る分類項目が指定されており、この中からの組み合わせで統計表作成を委託する事が出来ます。
- 2 要望する統計表によっては、**結果精度が十分でない場合や、必要な秘匿措置を講じるため、必ずしも申出者が期待する結果が得られないことがあります**。
- 3 各調査の「集計の仕様」にある提供可能な**次元数**については、例えば**合計3次元**までと記載している場合、表頭：性別（1次元）、表側：年齢階級（2次元）、表別：都道府県（3次元）となります。

# 秘匿措置された統計表の例示

第1表 産業、性、通勤・住込別年齢、勤続年数、月間出勤日数、通常日1日の実労働時間数、  
Table1 Age, duration of service, days worked per month, hours worked in an ordinary day,  
employees by industry, sex, living out or in (事業所規模1~4人及び企業規模1~4人)

産業、規模、性、通勤・住込 Industry, size, sex, living out or in	年齢 Age	勤続年数 Duration of service	出勤日数 Days worked	通常日1日の実労働時間数 Hours worked	きまって支給する現金給与額 Contractual cash earnings	常用労働者数 Number of regular employees	うち勤続1年以上 Service one year or more	過去1年間特別に支払われた現金給与額 (勤続1年以上) Special cash earnings in the past one year (Service one year or more)
	歳 (years old)	年 (years)	日 (days)	時間 (hours)	円 (yen)	人 (persons)	人 (persons)	円 (yen)
<b>F 電気・ガス・熱供給・水道業 事業所規模 Establishments 1~4人</b>								
全労働者 Total	48.6	9.9	20.4	8.1	273,945	1,446	1,267	751,998
通勤 Living out	48.6	9.9	20.4	8.1	273,945	1,446	1,267	751,998
住込 Living in	-	-	-	-	-	-	-	-
男性労働者 Male	49.2	10.0	21.1	8.2	306,456	1,118	1,012	834,252
通勤 Living out	49.2	10.0	21.1	8.2	306,456	1,118	1,012	834,252
住込 Living in	-	-	-	-	-	-	-	-
女性労働者 Female						328	x	x
通勤 Living out	—				—	328	x	x
住込 Living in							-	-
	X				X			
	X				X			
	—				—			

(毎月勤労統計調査 特別調査)

# オーダーメイド集計手続の流れ

【 申出者 】



利用したい統計表が



にない。。。。

【 厚生労働省 】



希望する統計表について利用相談

分類一覧から、作成を希望する統計表の仕様を決定

委託  
申出書

提出

受付、審査

依頼書



諾否、手数料額の通知

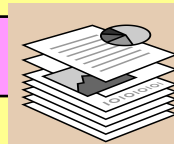
提出、納付

受理、確認

分析

統計表の提供

統計表の作成



成果の  
公表



利用  
実績  
報告書

提出

受理、確認



# ご静聴ありがとうございました。

## 【 照会先 】

厚生労働省 大臣官房統計情報部 審査解析室

nijitekiryoyou@mhlw.go.jp

委託統計係（オーダーメイド集計）

匿名データ提供係

<http://www.mhlw.go.jp/toukei/itaku/>

ただし、賃金構造基本統計調査のオーダーメイド集計については、

独立行政法人統計センター 製表部 統計作成支援課

nijiriyoyou@nstac.go.jp

利用審査担当