

第 13 回大都市交通センサス
一件明細調査
基礎集計報告書

令和 5 年 3 月

目次

1.	はじめに	5
1.1	用語	5
1.2	注意点	5
2.	調査概要	7
3.	0次OD	7
3.1	作成データ	7
3.1.1	元データ	7
3.2	一次処理（クレンジング）	7
3.2.1	目的	7
3.2.2	入力	7
3.2.3	出力	7
3.2.4	処理方法	7
3.3	二次処理（0次ODデータ作成）	8
3.3.1	目的	8
3.3.2	入力	8
3.3.3	出力	8
3.3.4	処理方法	8
3.4	最終処理	12
3.4.1	駅名によるエリア分け	12
3.4.2	データ付与と削除	12
4.	1次OD	14
4.1	作成データ	14
4.2	元データ	14
4.3	一次処理（クレンジング）	14
4.3.1	目的	14
4.3.2	入力	14
4.3.3	出力	14
4.3.4	処理方法	14
4.4	二次処理（1次ODデータ作成）	15
4.4.1	目的	15
4.4.2	入力	15
4.4.3	出力	15
4.4.4	処理方法	15
4.5	最終処理	18
4.5.1	駅名によるエリア分け	18
4.5.2	データ付与と削除	18
5.	2次OD	19
5.1	作成データ	19
5.2	元データ	19
5.3	一次処理（クレンジング）	19

5.3.1.	目的	19
5.3.2.	入力	19
5.3.3.	出力	19
5.3.4.	処理方法	19
5.4	二次処理（2次ODデータ作成）	20
5.4.1.	目的	20
5.4.2.	入力	20
5.4.3.	出力	20
5.4.4.	処理方法	20
5.5	最終処理	24
5.5.1.	駅名によるエリア分け	24
5.5.2.	データ付与と削除	25
6.	3次OD	26
6.1	作成データ	26
6.2	元データ	26
6.3	一次処理（クレンジング）	26
6.3.1.	目的	26
6.3.2.	入力	26
6.3.3.	出力	26
6.3.4.	処理方法	26
6.4	二次処理（2次ODデータ作成）	27
6.4.1.	目的	27
6.4.2.	入力	27
6.4.3.	出力	27
6.4.4.	処理方法	27
6.5	三次処理（3次ODデータ作成）	31
6.5.1.	目的	31
6.5.2.	入力	31
6.5.3.	出力	31
6.5.4.	処理方法	32
6.6	最終処理	34
6.6.1.	駅名によるエリア分け	34
6.6.2.	データ付与と削除	34
7.	集計表	35
7.1	一次処理集計件数	35
7.1.1.	一次処理集計件数	35
7.2	二次処理集計件数	35
7.2.1.	0次OD二次処理集計件数	35
7.2.2.	1次OD二次処理集計件数	36
7.2.3.	2次OD・3次OD二次処理集計件数	36
7.3	三次処理集計件数	37
7.3.1.	3次OD三次処理集計件数	37
7.4	最終処理集計件数	38
7.4.1.	0次OD最終処理集計件数	38
7.4.2.	1次OD最終処理集計件数	39
7.4.3.	2次OD最終処理集計件数	39
7.4.4.	3次OD最終処理集計件数	40
8.	補足説明	41

8.1	所要時間の考え方.....	41
-----	---------------	----

1. はじめに

1.1 用語

本書では、以下の用語、略称を使用する。（順不同）

用語・略称	説明
入場	改札機を通過して駅構内に入ること。
出場	改札機を通過して駅構内を出ること。
入場データ	入場時に生成される改札機の通過データ。
出場データ	出場時に生成される改札機の通過データ。
入場時刻	入場駅の改札機を通過する時刻。
出場時刻	出場駅の改札機を通過する時刻。
ID	個々の IC 券が保有している識別番号。
ハッシュ化	別の値にデータを変換する方法。基本的に不可逆。
レコード	各データにおける一行分のデータ。
トリップ	ある駅を入場して、ある駅を出場するまでの移動。
データ種別	カード事業者を表す。
処理種別	処理の種類。入場/出場を表す。
処理日時	入場時/出場時に生成される改札機の通過時刻。
事業者名	鉄道事業者の名称。
駅名	駅の名前。

1.2 注意点

- ・本書は、データ処理の概要について解説用に整理したものであり、全ての処理を記載しているわけではない。
- ・取りまとめられたデータはある仮想的な1日のものとして作成した統計値であり、実際のデータとは異なる。
- ・今回、調査手法を大きく変更しているので、前回までとの比較には注意が必要。
- ・集計結果は統計として前処理、修正処理等を経て作成されたものであり、実際の記録とは必ずしも一致しない。
- ・ある一日を対象としていることから、特異値が含まれる可能性があるため、詳細な分析を行った場合、詳細であればあるほど、それを一般的な分析の成果として扱うことは推奨されない。（分析の粒度によってその妥当性が大きく異なることに留意する必要がある。）
- ・調査日は、2021年12月の平日であり、日付は非公開としている。
- ・ODには、窓口で出場処理をしたもの等が含まれる（不可能なODや同一駅のODが含まれる）。
- ・1次OD：同一駅で2回のタッチを行う構造の駅においては、同一駅のODとして記録されることがある。
- ・事業者間の境界駅等では、入出場が複数事業者に分散して記録されていることがある。
- ・1次OD：定期券は含まれない。定期券との組み合わせ利用については、別紙のルールで集計している。
- ・2次OD：1次ODでの同一駅ODを除いたもので生成しており、2次ODで出現する同一駅ODは、規定の改札外滞在時間の範囲内での乗り継ぎを経て同一駅で出場したものである。
- ・2次OD, 3次OD：個人の流動ではない（ICカード単位の流動である）ことに留意する必要がある。
- ・駅乗降者数のそれぞれごとに10人に達しない駅については除外している。

- ・本集計は、三大都市圏の交通系 IC カードを対象としている。
- ・圏域を跨ぐ OD は含まれていない。
- ・データは時刻等について複数の処理等がなされたものであり、現実の記録とは一致しない。
- ・基本ゾーンは市区町村と同一としている。

2. 調査概要

複数の IC カード媒体毎の 1 日の改札機通過データを元に、0 次 OD データ、1 次 OD データ、2 次 OD データ、3 次 OD データと呼ばれるトリップデータを首都圏、中京圏、近畿圏別で作成を行った。今回対象となる IC カード媒体は、PiTaPa, ICOCA, TOICA, モノレール Suica, りんかい Suica, PASMO, manaca, Suica であり、2021 年 12 月某日の 2 日間のデータを用いた。

3. 0 次 OD

3.1 作成データ

0 次 OD とは、複数の IC カード媒体毎の 1 日の改札機通過データを元に、個人が駅を出場した後に駅に入場するという、出場/入場の組み合わせを一組にしたトリップデータである。これを首都圏、中京圏、近畿圏別で作成し、これらを圏域 OD データとする。

3.1.1. 元データ

元データとは、鉄道事業者様から提供いただいた IC カード媒体毎のデータのことである。

元データの各レコードには以下の項目が記載されている。

- ・ ID (ハッシュ化済み)
 - ・ 処理種別 (入場 or 出場)
 - ・ 処理日時
 - ・ 入場情報 (事業者名と駅名) or 出場情報 (事業者名と駅名)
- ※入場情報、出場情報の事業者名と駅名については各カード媒体毎で記載方法が異なる

3.2 一次処理 (クレンジング)

3.2.1. 目的

- ・ ID の再ハッシュ化を行うこと。
- ・ 事業者名と駅名の記載を統一すること。

3.2.2. 入力

入力データは、元データである。

3.2.3. 出力

一次処理後のデータの各レコードは以下の項目が記載されている。

- ・ ID (再ハッシュ化済み)
- ・ 処理種別 (入場 or 出場)
- ・ 処理日時
- ・ 事業者名 (事業者名記載統一済)
- ・ 駅名 (駅名記載統一済)

3.2.4. 処理方法

- ① ID の再ハッシュ化を行う。
- ② 駅情報を統一の事業者名・駅名に変換する。
- ③ 入場の処理日時が不明な場合は、対応する出場の処理日時を使用する。具体的な説明

は 8.1 にて記載している。また、この先の二次処理で連続したトリップの時間差を用いるため、（入場時間）－（直前のトリップの出場時間）が負の値となるデータは、差が 3 分以内であれば同じ時間に補正する。

3.3 二次処理（0 次 OD データ作成）

3.3.1. 目的

・各 ID の持つデータに関して、駅を出場した後に駅から入場するという、出場/入場の組み合わせを一組にしたトリップデータを作成すること。

・

3.3.2. 入力

入力データは一次処理後の、対象日付 2 日間のデータである。ただし、対象 2 日間のデータを入力としたもの、更に仮想的に 2 日間のデータとするため、対象となる 1 日目のデータに日付を 1 日追加したデータを追加したもの、対象となるの 2 日目のデータに日付を 2 日追加したデータを追加したものの、3 種類のデータを用意した。

3.3.3. 出力

二次処理後のデータの各レコードは以下の項目が記載されている。

- ・ ID（再ハッシュ化済み）
- ・ 入場情報（事業者名、駅名、入場時間）
- ・ 出場時間（事業者名、駅名、出場時間）

3.3.4. 処理方法

本処理は、一次処理で出力した全てのトリップデータから、出場情報/入場情報を一組にしたトリップデータを出力する処理である。すなわち、図 1 0 次 OD 作成における一次処理後の出力データのようなトリップデータからは、図 2 0 次 OD 作成における二次処理で生成した出力データのようなトリップデータを出力する。

ID	処理種別	処理日時	事業者名	駅名
XXXX	入場	8:25:40	aaa 事業者	A 駅
XXXX	出場	8:35:30	aaa 事業者	B 駅
XXXX	入場	9:02:15	aaa 事業者	C 駅
XXXX	出場	9:31:10	aaa 事業者	D 駅

図 1 0 次 OD 作成における一次処理後の出力データ

ID	入場情報			出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	A 駅	8:25:40			
XXXX	aaa 事業者	C 駅	9:02:15	aaa 事業者	B 駅	8:35:30
XXXX				aaa 事業者	D 駅	9:31:10

図 2 0 次 OD 作成における二次処理で生成した出力データ

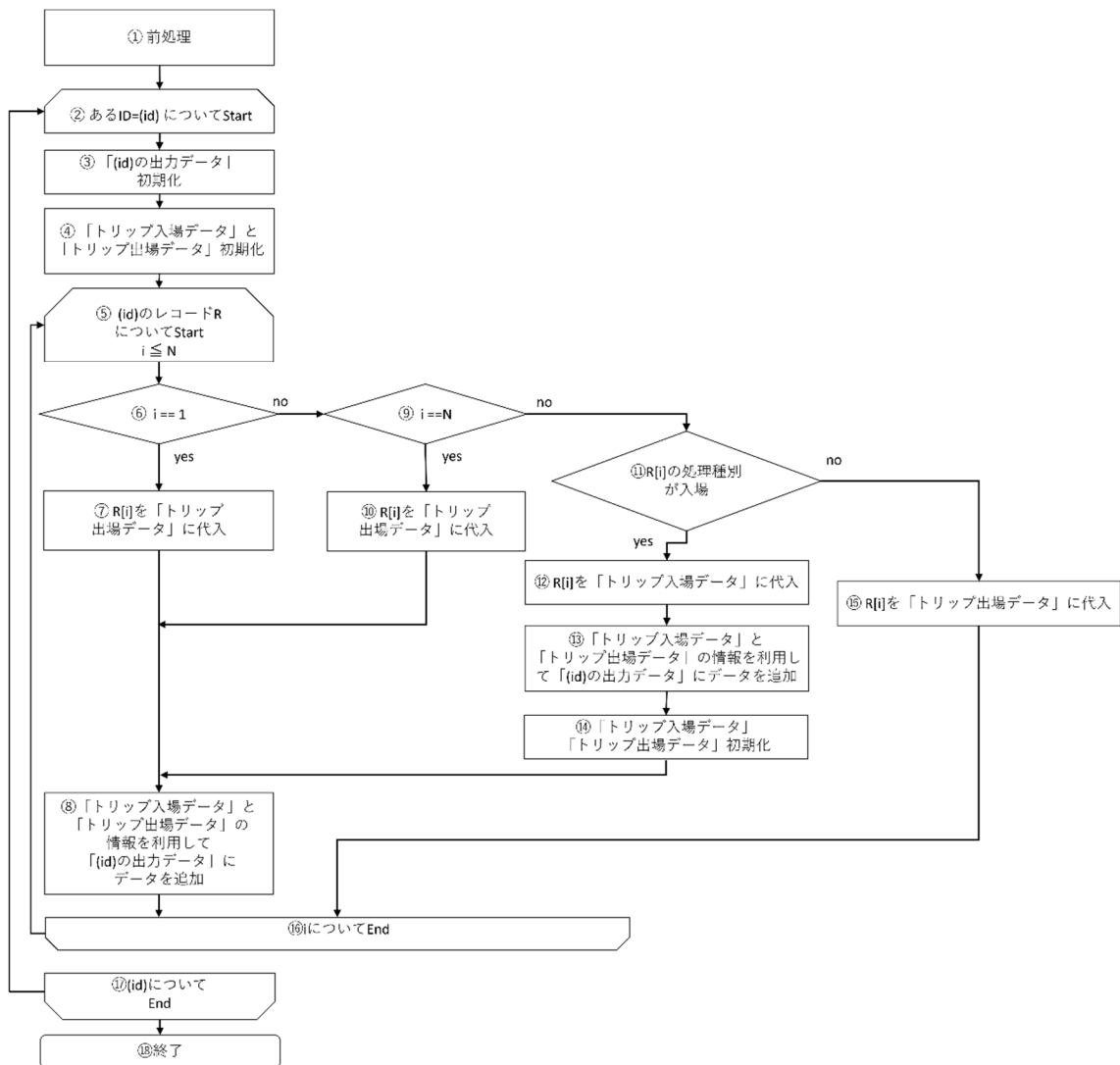


図 3 0次 OD 作成における二次処理の処理手順

- 図 3 0次 OD 作成における二次処理の処理手順の「① 前処理」では以下の処理を行う。
 - 全てのデータ種別のデータを一つのデータにする。
 - 同一駅で入出場しているトリップデータを削除する。
 - 圏域外のデータを削除する。
- 「② ある ID=(id) について Start」では、ある ID についての処理のループを開始する。
- 「③ 「(id)の出力データ」 初期化」では、(id)に関する二次処理出力データを格納するための変数を空で初期化する。
- 「④ 「トリップ入場データ」と「トリップ出場データ」 初期化」では、あるトリップの入場データと出場データを格納するための変数を空で初期化する。
- 「⑤ (id)のレコード R について Start」では、ID が(id)であるレコード R についてのループ処理を開始する。i は何回目のループであるかを表し、以降 R[i]は R の i 番

目のレコードを示す。(ある ID のレコード数) = N とし、 $i \leq N$ を満たさなくなるまでループ処理を行う。

- 「⑥ $i == 1$ 」では、(id)に関する一回目の処理、レコード R[1]についての処理であるかを判別する。
- 「⑦ R[i]を「トリップ入場データ」に代入」では、「トリップ入場データ」に R[i]を代入する。

	ID	処理種別	処理日時	事業者名	駅名
R[i]	XXXX	入場	8:25:40	aaa 事業者	A 駅

図 4 ⑦における R[i]

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ					

図 5 ⑦における「トリップ入場データ」

- 「⑧「トリップ入場データ」と「トリップ出場データ」の情報を利用して「(id)の出力データ」にデータを追加」では、「トリップ入場データ」と空である「トリップ出場データ」の情報における、二次処理出力データの項目を記載したデータを「(id)の出力データ」に追加する。

ID	入場情報			出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	A 駅	8:25:40			

図 6 ⑧における (id) の出力データに追加するデータ

- 「⑨ $i == N$ 」では、R[i]がある ID の最後の出場データであるかを判別する。
- 「⑩ R[i]を「トリップ出場データ」に代入」では、「トリップ出場データ」に R[i]を代入する。この場合、最後のデータなので、⑧で空である「トリップ入場データ」と「トリップ出場データ」の情報における、二次処理出力データの項目を記載したデータを「(id)の出力データ」に追加する。

	ID	処理種別	処理日時	事業者名	駅名
	XXXX	入場	9:02:15	aaa 事業者	C 駅
R[i]	XXXX	出場	9:31:10	aaa 事業者	D 駅

図 7 ⑨における R[i]

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ					
トリップ出場データ	XXXX	出場	9:31:10	aaa 事業者	D 駅

図 8 ⑩における「トリップ出場データ」

ID	入場情報			出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時

XXXX				aaa 事業者	D 駅	9:31:10
------	--	--	--	---------	-----	---------

図 9 ⑩における(id)の出力データに追加するデータ

- 「⑪ R[i]の処理種別が入場」では、R[i]の処理種別が入場であるかを判別する。
- 「⑫ R[i]を「トリップ入場データ」に代入」では、「トリップ入場データ」に R[i]を代入する。
-

R[i]	ID	処理種別	処理日時	事業者名	駅名
	XXXX	入場	8:25:40	aaa 事業者	A 駅
	XXXX	出場	8:35:30	aaa 事業者	B 駅
	XXXX	入場	9:02:15	aaa 事業者	C 駅
	XXXX	出場	9:31:10	aaa 事業者	D 駅

図 10 ⑪における R[i]

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	9:02:15	aaa 事業者	C 駅
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 11 ⑫における「トリップ入場データ」と「トリップ出場データ」

- 「⑬ 「トリップ入場データ」と「トリップ出場データ」の情報を利用して「(id)の出力データ」にデータを追加」では、「トリップ入場データ」と「トリップ出場データ」の情報における、二次処理出力データの項目を記載したデータを「(id)の出力データ」に追加する。

ID	入場情報			出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	C 駅	9:02:15	aaa 事業者	B 駅	8:35:30

図 12 ⑬における(id)の出力データに追加するデータ

- 「⑭ 「トリップ入場データ」と「トリップ出場データ」初期化」では、あるトリップの入場データと出場データを格納するための変数を空で初期化する。
- 「⑮ R[i]を「トリップ出場データ」に代入」では、「トリップ出場データ」に R[i]を代入する。

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ					
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 13 ⑮における「トリップ入場データ」と「トリップ出場データ」

- 「⑯i について End」は、⑤の処理ループの終了を表す。
- 「⑰ (id)について End」は、②の処理ループの終了を表す。
- 「⑱ 終了」で処理を終了する。
-

ID	入場情報			出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	A 駅	8:25:40			

XXXX	aaa 事業者	C 駅	9:02:15	aaa 事業者	B 駅	8:35:30
XXXX				aaa 事業者	D 駅	9:31:10

図 14 0 次 OD 作成における最終的な「(id)の出力データ」

本操作を行うことで、0 次 OD データが得られる。

3.4 最終処理

3.4.1. 駅名によるエリア分け

第 12 回大都市交通センサス調査結果集計表の WEB ページより各エリアの鉄道駅コードをダウンロードする。

参照: [公共交通政策：第 12 回大都市交通センサス調査結果集計表 - 国土交通省 \(mlit.go.jp\)](http://mlit.go.jp)

【鉄道調査】コードデータファイル(Excel形式)

1	ゾーンコード表	首都圏	中京圏	近畿圏
2	鉄道事業者コード	首都圏	中京圏	近畿圏
3	鉄道路線コード	首都圏	中京圏	近畿圏
4	鉄道駅コード	首都圏	中京圏	近畿圏

この鉄道駅コード表が作成されたのが平成 28 年のため、それ以降に発生した以下に對してのデータ修正・追加を行う。詳細内容については表 1 駅名変更、表 2 新駅追加参照。

(1) 駅名変更

(2) 新駅追加

上記を修正・追加したコード表の「事業者名」「駅名」によりエリア分けを行う。

3.4.2. データ付与と削除

- ・ 2 次処理で出力された結果と、上記 1 項のコード表の「事業者名」「駅名」をキーにして入場駅、出場駅に結合し該当する「エリア」情報を各レコードに付加する。
- ・ 「ID」をデータから削除する。
- ・ 「入場時間」から時間のみを抽出したデータに変換する。
- ・ 「出場時間」を削除し、「入場時間」から「出場時間」までの「所要時間」を 5 分単位で付与する。

集計の際に、乗降客数がそれぞれ 10 人以下の駅については、集計結果から除外している。

表 1 駅名変更

エリア	駅コード	事業者名	路線名	旧駅名	新駅名
首都圏	00612	東日本旅客鉄道	常磐線快速	佐貫	龍ヶ崎市
首都圏	20019	京浜急行電鉄	京浜急行本線	花月園前	花月総持寺
首都圏	20024	京浜急行電鉄	京浜急行本線	中木戸	京急東神奈川
首都圏	20104	京浜急行電鉄	逗子線	新逗子	逗子・葉山
首都圏	20306	京浜急行電鉄	空港線	羽田空港国際線ターミナル	羽田空港第 3 ターミナル
首都圏	20307	京浜急行電鉄	空港線	羽田空港国内線ターミナル	羽田空港第 1・第 2 ターミナル
首都圏	20406	京浜急行電鉄	大師線	産業道路	大師橋
首都圏	35525	東急電鉄	田園都市線	南町田	南町田グランベリーパーク
首都圏	40607	西武鉄道	多摩湖線	西武遊園地	多摩湖
首都圏	41002	西武鉄道	山口線	遊園地西	西武園ゆうえんち
首都圏	45016	東武鉄道	伊勢崎線（1）	松原団地	獨協大学前
首都圏	88008	ゆりかもめ	東京臨海新交通臨海線	船の科学館	東京国際クルーズターミナル
首都圏	88011	ゆりかもめ	東京臨海新交通臨海線	国際展示場正門	東京ビッグサイト
首都圏	96001	東京モノレール	東京モノレール羽田空港線	羽田空港第 2 ビル	羽田空港第 2 ターミナル
首都圏	96002	東京モノレール	東京モノレール羽田空港線	羽田空港第 1 ビル	羽田空港第 1 ターミナル
首都圏	96004	東京モノレール	東京モノレール羽田空港線	羽田空港国際線ビル	羽田空港第 3 ターミナル
近畿圏	01715	西日本旅客鉄道	片町線（学研都市線）	東寝屋川	寝屋川公園
近畿圏	07101	阪神電気鉄道	本線	梅田	大阪梅田
近畿圏	07113	阪神電気鉄道	本線	鳴尾	鳴尾・武庫川女子大前
近畿圏	08101	阪急電鉄	神戸本線	梅田	大阪梅田
近畿圏	08211	阪急電鉄	宝塚本線	石橋	石橋阪大前
近畿圏	08328	阪急電鉄	京都本線	河原町	京都河原町
近畿圏	10126	京阪電気鉄道	京阪本線	八幡市	石清水八幡宮
近畿圏	10133	京阪電気鉄道	京阪本線	深草	龍谷大前深草
近畿圏	10407	京阪電気鉄道	京津線	浜大津	びわ湖浜大津
近畿圏	10514	京阪電気鉄道	石山坂本線	別所	大津市役所前
近畿圏	10515	京阪電気鉄道	石山坂本線	皇子山	京阪大津京
近畿圏	10521	京阪電気鉄道	石山坂本線	坂本	坂本比叡山口
近畿圏	18408	京福電気鉄道	北野線	等持院	等持院・立命館大学衣笠キャンパス前
近畿圏	26108	神戸新交通	ポートアイランド線	京コンピュータ前（神戸どうぶつ王国）	計算科学センター
近畿圏	29103	大阪モノレール	大阪モノレール線	柴原	柴原阪大前

表 2 新駅追加

エリア	駅コード	事業者名	路線名	旧駅名	新駅名
首都圏		東日本旅客鉄道	山手線		高輪ゲートウェイ
首都圏		東日本旅客鉄道	常磐線		J ワレージ
首都圏		東日本旅客鉄道	南武線		小田栄
首都圏		東京地下鉄	日比谷線		虎ノ門ヒルズ
首都圏		東武鉄道	東上線		みなみ寄居
首都圏		秩父鉄道	秩父本線		ふかや花園
首都圏		秩父鉄道	秩父本線		ソシオ流通センター
首都圏		相模鉄道	相鉄・JR直通線		羽沢横浜国大
近畿圏		西日本旅客鉄道	おおさか東線		南吹田
近畿圏		西日本旅客鉄道	おおさか東線		J R 淡路
近畿圏		西日本旅客鉄道	おおさか東線		城北公園通
近畿圏		西日本旅客鉄道	おおさか東線		J R 野江
近畿圏		西日本旅客鉄道	山陰本線		梅小路京都西
近畿圏		西日本旅客鉄道	おおさか東線		衣摺加美北
近畿圏		西日本旅客鉄道	東海道本線		JR総持寺
近畿圏		西日本旅客鉄道	東海道本線		摩耶
近畿圏		西日本旅客鉄道	山陽本線		東姫路
近畿圏		京福電気鉄道	嵐山本線		撮影所前
近畿圏		伊賀鉄道	伊賀線		四十九

4. 1 次 OD

4.1 作成データ

1 次 OD を用いた、圏域 OD データを作成する。1 次 OD とは複数の IC カード媒体毎の 1 日の改札機通過データを元に、入場/出場を一組にしたトリップデータである。これを首都圏、中京圏、近畿圏別で作成し、これらを圏域 OD データとする。

4.2 元データ

元データとは、鉄道事業者様から提供いただいた IC カード媒体毎のデータのことである。

元データの各レコードには以下の項目が記載されている。

- ・ ID (ハッシュ化済み)
 - ・ 処理種別 (入場 or 出場)
 - ・ 処理日時
 - ・ 入場情報 (事業者名と駅名) or 出場情報 (事業者名と駅名)
- ※入場情報、出場情報の事業者名と駅名については各カード媒体毎で記載方法が異なる

4.3 一次処理 (クレンジング)

4.3.1. 目的

- ・ ID の再ハッシュ化を行うこと。
- ・ 事業者名と駅名の記載を統一すること。

4.3.2. 入力

入力データは、元データである。

4.3.3. 出力

一次処理後のデータの各レコードは以下の項目が記載されている。

- ・ ID (再ハッシュ化済み)
- ・ 処理種別 (入場 or 出場)
- ・ 処理日時
- ・ 事業者名 (事業者名記載統一済)
- ・ 駅名 (駅名記載統一済)

4.3.4. 処理方法

- ④ ID の再ハッシュ化を行う。
- ⑤ 駅情報を統一の事業者名・駅名に変換する。
- ⑥ 入場の処理日時が不明な場合は、対応する出場の処理日時を使用する。具体的な説明は 8. 1 にて記載している。また、この先の二次処理で連続したトリップの時間差を用いるため、(入場時間) - (直前のトリップの出場時間) が負の値となるデータは、差が 3 分以内であれば同じ時間に補正する。

4.4 二次処理（1次ODデータ作成）

4.4.1. 目的

- ・各IDの持つデータに関して、入場/出場を一組にしたトリップデータを作成すること。
- ・

4.4.2. 入力

入力データは一次処理後のデータである。

4.4.3. 出力

二次処理後のデータの各レコードは以下の項目が記載されている。

- ・ID（再ハッシュ化済み）
- ・入場情報（事業者名、駅名、入場時間）
- ・出場時間（事業者名、駅名、出場時間）

4.4.4. 処理方法

本処理は、一次処理で出力した全てのトリップデータから、入場情報/出場情報を一組にしたトリップデータを出力する処理である。すなわち、図15 1次OD作成における一次処理後の出力データのようなトリップデータからは、図16 1次OD作成における二次処理で生成した出力データのようなトリップデータを出力する。

ID	処理種別	処理日時	事業者名	駅名
XXXX	入場	8:25:40	aaa 事業者	A 駅
XXXX	出場	8:35:30	aaa 事業者	B 駅
XXXX	入場	9:02:15	aaa 事業者	C 駅
XXXX	出場	9:31:10	aaa 事業者	D 駅

図 15 1次OD作成における一次処理後の出力データ

ID	入場情報			出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	A 駅	8:25:40	aaa 事業者	B 駅	8:35:30
XXXX	aaa 事業者	C 駅	9:02:15	aaa 事業者	D 駅	9:31:10

図 16 1次OD作成における二次処理で生成した出力データ

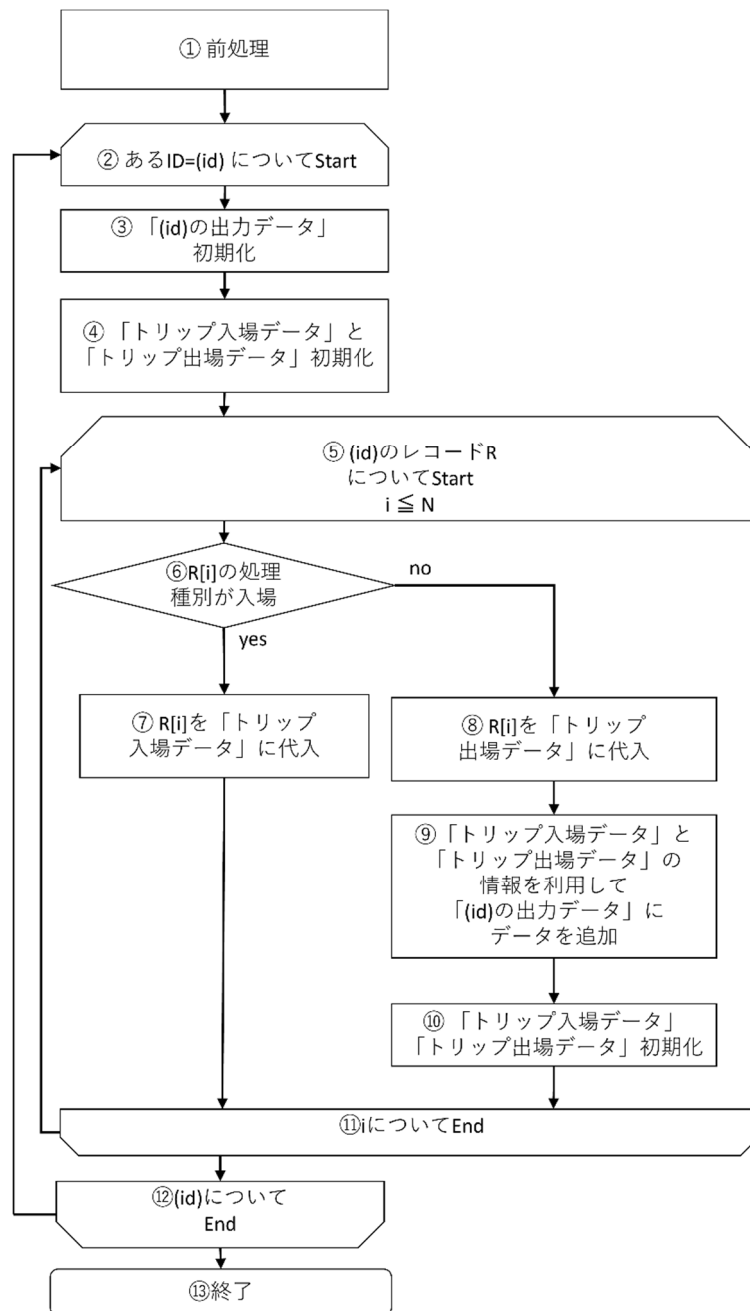


図 17 1次 OD 作成における二次処理の処理手順

- 図 17 1次 OD 作成における二次処理の処理手順の「① 前処理」では以下の処理を行う。
 - 全てのデータ種別のデータを一つのデータにする。
 - 同一駅で入出場しているトリップデータを削除する。
 - 圏域外のデータを削除する。
- 「② ある ID=(id) について Start」では、ある ID についての処理のループを開始する。
- 「③ 「(id) の出力データ」 初期化」では、(id)に関する二次処理出力データを格納

するための変数を空で初期化する。

- 「④ 「トリップ入場データ」と「トリップ出場データ」初期化」では、あるトリップの入場データと出場データを格納するための変数を空で初期化する。
- 「⑤ (id)の記録RについてStart」では、ID が(id)である記録Rについてのループ処理を開始する。iは何回目のループであるかを表し、以降R[i]はRのi番目の記録を示す。(あるIDの記録数)=Nとし、 $i \leq N$ を満たさなくなるまでループ処理を行う。
- 「⑥ R[i]の処理種別が入場」では、R[i]の処理種別が入場であるかを判別する。
- 「⑦ R[i]を「トリップ入場データ」に代入」では、「トリップ入場データ」にR[i]を代入する。

	ID	処理種別	処理日時	事業者名	駅名
R[i]	XXXX	入場	8:25:40	aaa 事業者	A 駅

図 18 ⑦におけるR[i]

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ					

図 19 ⑦における「トリップ入場データ」

- 「⑧ R[i]を「トリップ出場データ」に代入」では、「トリップ出場データ」にR[i]を代入する。

	ID	処理種別	処理日時	事業者名	駅名
	XXXX	入場	8:25:40	aaa 事業者	A 駅
R[i]	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 20 ⑧におけるR[i]

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 21 ⑧における「トリップ入場データ」と「トリップ出場データ」

- 「⑨「トリップ入場データ」と「トリップ出場データ」の情報を利用して「(id)の出力データ」にデータを追加」では、「トリップ入場データ」と「トリップ出場データ」の情報における、二次処理出力データの項目を記載したデータを「(id)の出力データ」に追加する。

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 22 ⑨における「トリップ入場データ」と「トリップ出場データ」

ID	入場情報			出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	A 駅	8:25:40	aaa 事業者	B 駅	8:35:30

図 23 ⑨における「(id)の出力データ」

- 「⑩「トリップ入場データ」と「トリップ出場データ」初期化」では、あるトリップの入場データと出場データを格納するための変数を空で初期化する。
- 「⑪i について End」は、③の処理ループの終了を表す。
- 「⑫(id)について End」は、②の処理ループの終了を表す。
- 「⑬ 終了」で処理を終了する。

本操作を行うことで、全データの 1 次 OD データが得られる。

4.5 最終処理

4.5.1. 駅名によるエリア分け

第 12 回大都市交通センサス調査結果集計表の WEB ページより各エリアの鉄道駅コードをダウンロードする。

参照: [公共交通政策：第 12 回大都市交通センサス調査結果集計表 - 国土交通省 \(mlit.go.jp\)](http://mlit.go.jp)

【鉄道調査】コードデータファイル(Excel形式)

1	ゾーンコード表	首都圏	中京圏	近畿圏
2	鉄道事業者コード	首都圏	中京圏	近畿圏
3	鉄道路線コード	首都圏	中京圏	近畿圏
4	鉄道駅コード	首都圏	中京圏	近畿圏

この鉄道駅コード表が作成されたのが平成 28 年のため、それ以降に発生した以下に對してのデータ修正・追加を行う。詳細内容については表 1 駅名変更、表 2 新駅追加参照。

(1) 駅名変更

(2) 新駅追加

上記を修正・追加したコード表の「事業者名」「駅名」によりエリア分けを行う。

4.5.2. データ付与と削除

- ・ 2 次処理で出力された結果と、上記 1 項のコード表の「事業者名」「駅名」をキーにして入場駅、出場駅に結合し該当する「エリア」情報を各レコードに付加する。
- ・ 「ID」をデータから削除する。
- ・ 「入場時間」から時間のみを抽出したデータに変換する。
- ・ 「出場時間」を削除し、「入場時間」から「出場時間」までの「所要時間」を 5 分単位で付与する。

集計の際に、乗降客数がそれぞれ 10 人以下の駅については、集計結果から除外している。

5. 2次OD

5.1 作成データ

2次ODを用いた、圏域ODデータを作成する。2次ODとは、複数のICカード媒体毎の1日の改札機通過データを元に、入場/出場を一組にしたトリップデータ（ただし、次のトリップまでの時間がある閾値時間以内の場合は同一トリップとみなす）である。これを首都圏、中京圏、近畿圏別で作成し、これらを圏域ODデータとする。

5.2 元データ

元データとは、鉄道事業者様から提供いただいたICカード媒体毎のデータのことである。

元データの各レコードには以下の項目が記載されている。

- ・ID（ハッシュ化済み）
 - ・処理種別（入場 or 出場）
 - ・処理日時
 - ・入場情報（事業者名と駅名）or 出場情報（事業者名と駅名）
- ※入場情報、出場情報の事業者名と駅名については各カード媒体毎で記載方法が異なる

5.3 一次処理（クレンジング）

5.3.1. 目的

- ・IDの再ハッシュ化を行うこと。
- ・事業者名と駅名の記載を統一すること。

5.3.2. 入力

入力データは、元データである。

5.3.3. 出力

一次処理後のデータの各レコードは以下の項目が記載されている。

- ・ID（再ハッシュ化済み）
- ・処理種別（入場 or 出場）
- ・処理日時
- ・事業者名（事業者名記載統一済）
- ・駅名（駅名記載統一済）

5.3.4. 処理方法

- ⑦ IDの再ハッシュ化を行う。
- ⑧ 駅情報を統一の事業者名・駅名に変換する。
- ⑨ 入場の処理日時が不明な場合は、対応する出場の処理日時を使用する。具体的な説明は8.1にて記載している。また、この先の二次処理で連続したトリップの時間差を用いるため、（入場時間）－（直前のトリップの出場時間）が負の値となるデータは、差が3分以内であれば同じ時間に補正する。

5.4 二次処理（2次ODデータ作成）

5.4.1. 目的

- ・各IDの持つデータに関して、入場/出場を一組にしたトリップデータ（ただし、次のトリップまでの時間がある閾値時間未満の場合は同一トリップとみなす）を作成すること。

5.4.2. 入力

入力データは一次処理後のデータである。

5.4.3. 出力

二次処理後のデータの各レコードは以下の項目が記載されている。

- ・ID（再ハッシュ化済み）
- ・入場情報（事業者名、駅名、入場時間）
- ・出場時間（事業者名、駅名、出場時間）

5.4.4. 処理方法

本処理は、一次処理で出力した全てのトリップデータから、次のトリップまでの時間がある閾値時間以上であるような、入場情報/出場情報を一組にしたトリップデータを出力する処理である。すなわち、図24 2次OD作成における一次処理後の出力データのようなトリップデータからは、閾値時間が60分の場合、図25 2次OD作成における二次処理で生成した出力データのようなトリップデータを出力する。

ID	処理種別	処理日時	事業者名	駅名
XXXX	入場	8:25:40	aaa 事業者	A 駅
XXXX	出場	8:35:30	aaa 事業者	B 駅
XXXX	入場	9:02:15	aaa 事業者	C 駅
XXXX	出場	9:31:10	aaa 事業者	D 駅

図 24 2次OD作成における一次処理後の出力データ

ID	入場情報			出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	A 駅	8:25:40	aaa 事業者	D 駅	9:31:10

図 25 2次OD作成における二次処理で生成した出力データ

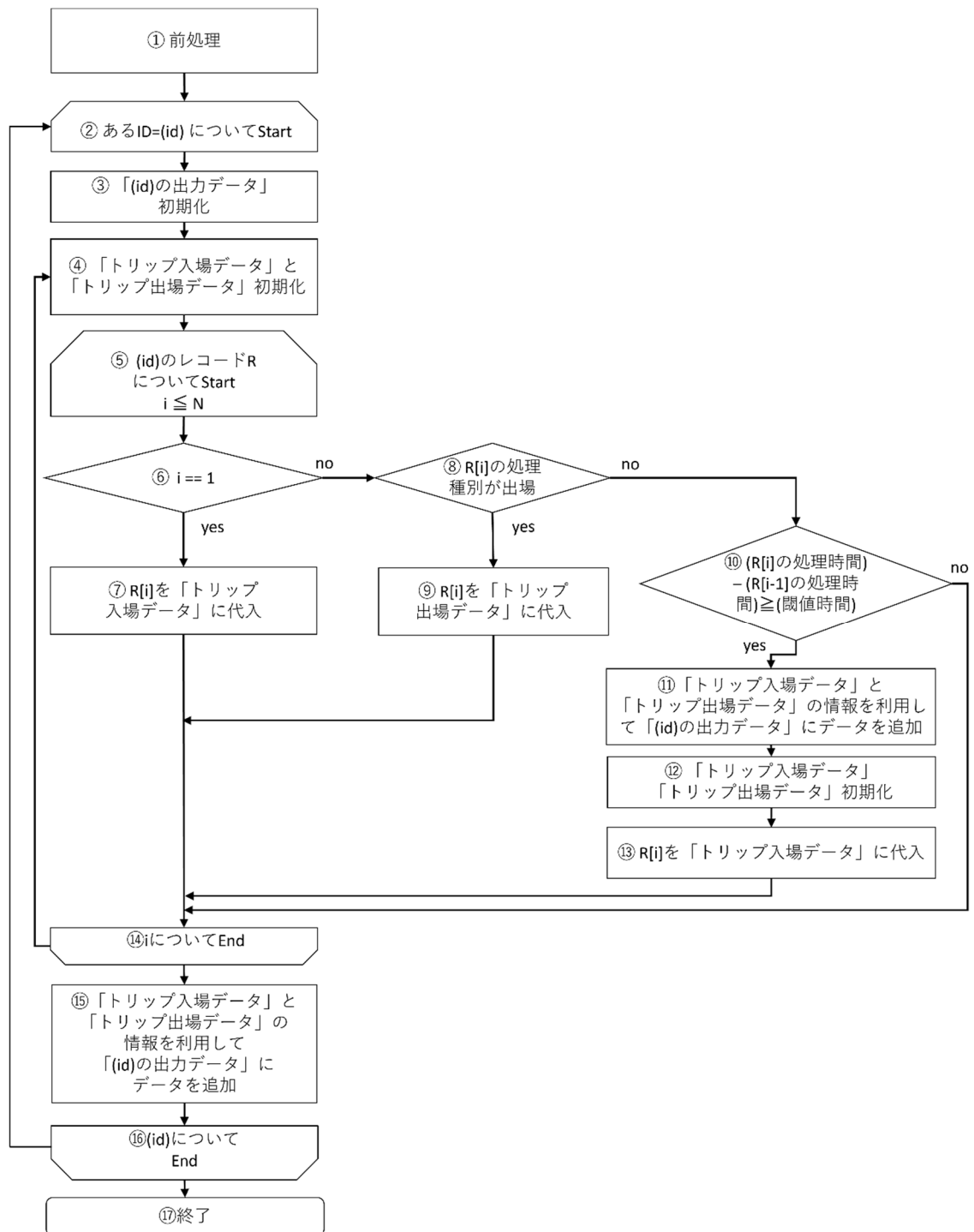


図 26 2次 OD 作成における二次処理の処理手順

- 図 26 2次 OD 作成における二次処理の処理手順の「① 前処理」では以下の処理を行う。
 - 全てのデータ種別のデータを一つのデータにする。
- 「② ある ID=(id) について Start」では、ある ID についての処理のループを開始する。
- 「③ 「(id)の出力データ」 初期化」では、(id)に関する二次処理出力データを格納

するための変数を空で初期化する。

- 「④ 「トリップ入場データ」と「トリップ出場データ」初期化」では、あるトリップの入場データと出場データを格納するための変数を空で初期化する。
- 「⑤ (id)の記録Rについて Start」では、ID が(id)である記録Rについてのループ処理を開始する。iは何回目のループであるかを表し、以降R[i]はRのi番目の記録を示す。(あるIDの記録数)=Nとし、 $i \leq N$ を満たさなくなるまでループ処理を行う。
- 「⑥ $i == 1$ 」では、(id)に関する一回目の処理、記録R[1]についての処理であるかを判別する。
- 「⑦ R[i]を「トリップ入場データ」に代入」では、「トリップ入場データ」にR[i]を代入する。

	ID	処理種別	処理日時	事業者名	駅名
R[i]	XXXX	入場	8:25:40	aaa 事業者	A 駅

図 27 ⑦におけるR[i]

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ					

図 28 ⑦における「トリップ入場データ」

- 「⑧ R[i]の処理種別が出場」では、R[i]の処理種別が出場であるかを判別する。
- 「⑨ R[i]を「トリップ出場データ」に代入」では、「トリップ出場データ」にR[i]を代入する。(※1 既に「トリップ出場データ」にデータが入っている場合は、R[i]で置き換える)

	ID	処理種別	処理日時	事業者名	駅名
	XXXX	入場	8:25:40	aaa 事業者	A 駅
R[i]	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 29 ⑨におけるR[i]

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 30 ⑨における「トリップ入場データ」と「トリップ出場データ」

※1

	ID	処理種別	処理日時	事業者名	駅名
	XXXX	入場	8:25:40	aaa 事業者	A 駅
	XXXX	出場	8:35:30	aaa 事業者	B 駅
	XXXX	入場	9:02:15	aaa 事業者	C 駅
R[i]	XXXX	出場	9:31:10	aaa 事業者	D 駅

図 31 ⑨における R[i] (トリップ出場データを置き換える場合)

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 32 ⑨における「トリップ入場データ」と「トリップ出場データ」(置き換え前)

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	9:31:10	aaa 事業者	D 駅

図 33 ⑨における「トリップ入場データ」と「トリップ出場データ」(置き換え後)

- 「⑩ (R[i]の処理時間)-(R[i-1]の処理時間)≥ある閾値時間」では、次のトリップまでの時間がある閾値時間以上であるかを判別する。これを満たす場合は、既に格納されている「トリップ入場データ」「トリップ出場データ」を一つのトリップとして確定する。満たさない場合は、入場データ R[i]を構成するトリップも「トリップ入場データ」「トリップ出場データ」のトリップと同一トリップとみなす。

	ID	処理種別	処理日時	事業者名	駅名
R[i]	XXXX	入場	8:25:40	aaa 事業者	A 駅
	XXXX	出場	8:35:30	aaa 事業者	B 駅
	XXXX	入場	10:02:15	aaa 事業者	C 駅
	XXXX	出場	10:31:10	aaa 事業者	D 駅

図 34 ⑩における R[i] (⑩を満たす場合)

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 35 ⑩における「トリップ入場データ」と「トリップ出場データ」(確定)

	ID	処理種別	処理日時	事業者名	駅名
R[i]	XXXX	入場	8:25:40	aaa 事業者	A 駅
	XXXX	出場	8:35:30	aaa 事業者	B 駅
	XXXX	入場	9:02:15	aaa 事業者	C 駅
	XXXX	出場	9:31:10	aaa 事業者	D 駅

図 36 ⑩における R[i] (⑩を満たさない場合)

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 37 ⑩における「トリップ入場データ」と「トリップ出場データ」
(確定せず、処理継続)

- 「⑪「トリップ入場データ」と「トリップ出場データ」の情報を利用して「(id)の出力データ」にデータを追加」では、「トリップ入場データ」と「トリップ出場データ」の情報における、二次処理出力データの項目を記載したデータを「(id)の出力データ」に追加する。

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 38 ⑪における「トリップ入場データ」と「トリップ出場データ」（確定済み）

ID	入場情報			出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	A 駅	8:25:40	aaa 事業者	B 駅	8:35:30

図 39 ⑪における「(id)の出力データ」

- 「⑫「トリップ入場データ」と「トリップ出場データ」初期化」では、あるトリップの入場データと出場データを格納するための変数を空で初期化する。
- 「⑬ R[i]を「トリップ入場データ」に代入」では、「トリップ入場データ」に R[i]を代入する。
- 「⑭ i について End」は、③の処理ループの終了を表す。
- 「⑮「トリップ入場データ」と「トリップ出場データ」の情報を利用して「(id)の出力データ」にデータを追加」では、「トリップ入場データ」と「トリップ出場データ」の情報における、二次処理出力データの項目を記載したデータを「(id)の出力データ」に追加する。
- 「⑯ (id)について End」は、②の処理ループの終了を表す。
- 「⑰ 終了」で処理を終了する。

本操作を行うことで、全データの 2 次 OD データが得られる。

5.5 最終処理

5.5.1. 駅名によるエリア分け

第 12 回大都市交通センサス調査結果集計表の WEB ページより各エリアの鉄道駅コードをダウンロードする。

参照：[公共交通政策：第 12 回大都市交通センサス調査結果集計表 - 国土交通省 \(mlit.go.jp\)](http://mlit.go.jp)

【鉄道調査】コードデータファイル(Excel形式)

1	ゾーンコード表	首都圏	中京圏	近畿圏
2	鉄道事業者コード	首都圏	中京圏	近畿圏
3	鉄道路線コード	首都圏	中京圏	近畿圏
4	鉄道駅コード	首都圏	中京圏	近畿圏

この鉄道駅コード表が作成されたのが平成 28 年のため、それ以降に発生した以下に對してのデータ修正・追加を行う。詳細内容については表 1 駅名変更、表 2 新駅追加参照。

- (1) 駅名変更
- (2) 新駅追加

上記を修正・追加したコード表の「事業者名」「駅名」によりエリア分けを行う。

5.5.2. データ付与と削除

- ・2次処理で出力された結果と、上記1項のコード表の「事業者名」「駅名」をキーにして入場駅、出場駅に結合し該当する「エリア」情報を各レコードに付加する。
- ・「ID」をデータから削除する。
- ・「入場時間」から時間のみを抽出したデータに変換する。
- ・「出場時間」を削除し、「入場時間」から「出場時間」までの「所要時間」を5分単位で付与する。

集計の際に、乗降客数がそれぞれ10人以下の駅については、集計結果から除外している。

6. 3 次 OD

6.1 作成データ

3 次 OD を用いた、圏域 OD データを作成する。3 次 OD とは、個人の一日の動きを 1 レコードとして出力したデータである。複数の IC カード媒体毎の 1 日の改札機通過データから作成した、2 次 OD データ（入場/出場を一組にしたトリップデータであり、次のトリップまでの時間がある閾値時間以内の場合は同一トリップとみなしている）を元に、各 ID の入場/出場の組を一日分並べたトリップデータを作成する。これを首都圏、中京圏、近畿圏別で作成し、これらを圏域 OD データとする。

6.2 元データ

元データとは、鉄道事業者様から提供いただいた IC カード媒体毎のデータのことである。

元データの各レコードには以下の項目が記載されている。

- ・ ID（ハッシュ化済み）
- ・ 処理種別（入場 or 出場）
- ・ 処理日時
- ・ 入場情報（事業者名と駅名） or 出場情報（事業者名と駅名）

※入場情報、出場情報の事業者名と駅名については各カード媒体毎で記載方法が異なる

6.3 一次処理（クレンジング）

6.3.1. 目的

- ・ ID の再ハッシュ化を行うこと。
- ・ 事業者名と駅名の記載を統一すること。

6.3.2. 入力

入力データは、元データである。

6.3.3. 出力

一次処理後のデータの各レコードは以下の項目が記載されている。

- ・ ID（再ハッシュ化済み）
- ・ 処理種別（入場 or 出場）
- ・ 処理日時
- ・ 事業者名（事業者名記載統一済）
- ・ 駅名（駅名記載統一済）

6.3.4. 処理方法

- ⑩ ID の再ハッシュ化を行う。
- ⑪ 駅情報を統一の事業者名・駅名に変換する。
- ⑫ 入場の処理日時が不明な場合は、対応する出場の処理日時を使用する。具体的な説明は 8.1 にて記載している。また、この先の二次処理で連続したトリップの時間差を用いるため、（入場時間）－（直前のトリップの出場時間）が負の値となるデータは、差が 3 分以内であれば同じ時間に補正する。

6.4 二次処理（2次ODデータ作成）

6.4.1. 目的

- ・各IDの持つデータに関して、入場/出場を一組にしたトリップデータ（ただし、次のトリップまでの時間がある閾値時間未満の場合は同一トリップとみなす）を作成すること。

6.4.2. 入力

入力データは一次処理後のデータである。

6.4.3. 出力

二次処理後のデータの各レコードは以下の項目が記載されている。

- ・ID（再ハッシュ化済み）
- ・入場情報（事業者名、駅名、入場時間）
- ・出場時間（事業者名、駅名、出場時間）

6.4.4. 処理方法

本処理は、一次処理で出力した全てのトリップデータから、次のトリップまでの時間がある閾値時間以上であるような、入場情報/出場情報を一組にしたトリップデータを出力する処理である。すなわち、図40 2次OD作成における一次処理後の出力データのようなトリップデータからは、閾値時間が60分の場合、図41 2次OD作成におけるから二次処理で生成した出力データのようなトリップデータを出力する。

ID	処理種別	処理日時	事業者名	駅名
XXXX	入場	8:25:40	aaa 事業者	A 駅
XXXX	出場	8:35:30	aaa 事業者	B 駅
XXXX	入場	9:02:15	aaa 事業者	C 駅
XXXX	出場	9:31:10	aaa 事業者	D 駅

図 40 2次OD作成における一次処理後の出力データ

ID	入場情報			出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	A 駅	8:25:40	aaa 事業者	D 駅	9:31:10

図 41 2次OD作成におけるから二次処理で生成した出力データ

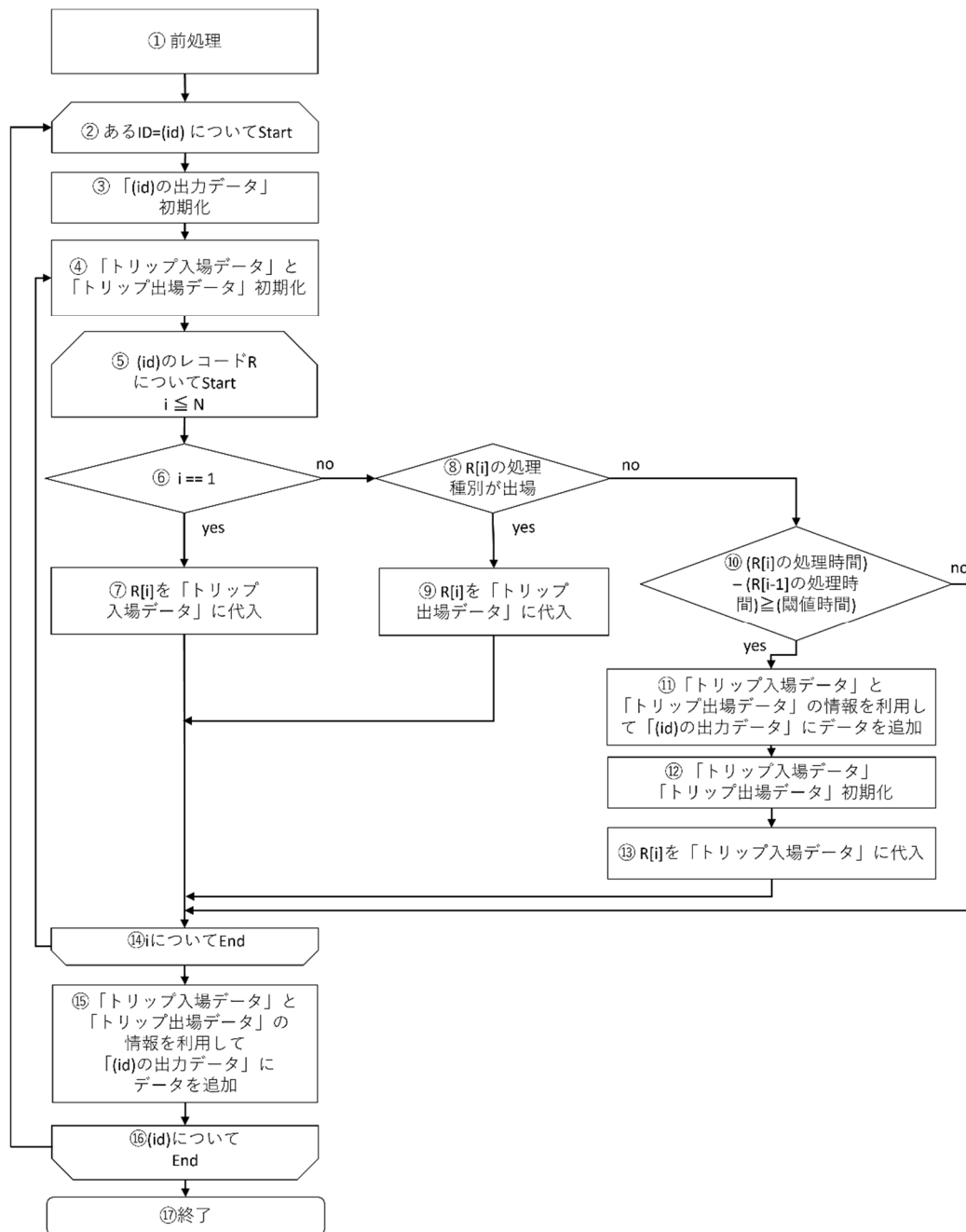


図 42 2次 OD 作成における二次処理の処理手順

- 図 42 2次 OD 作成における二次処理の処理手順の「① 前処理」では以下の処理を行う。
 - 全てのデータ種別のデータを一つのデータにする。
 - 同一駅で入出場しているトリップデータを削除する。
 - 圏域外のデータを削除する。
- 「② ある ID=(id) について Start」では、ある ID についての処理のループを開始する。
- 「③ 「(id)の出力データ」 初期化」では、(id)に関する二次処理出力データを格納するための変数を空で初期化する。
- 「④ 「トリップ入場データ」と「トリップ出場データ」 初期化」では、あるトリッ

プの入場データと出場データを格納するための変数を空で初期化する。

- 「⑤ (id)の記録Rについて Start」では、ID が(id)である記録Rについてのループ処理を開始する。iは何回目のループであるかを表し、以降R[i]はRのi番目の記録を示す。(あるIDの記録数)=Nとし、 $i \leq N$ を満たさなくなるまでループ処理を行う。
- 「⑥ $i == 1$ 」では、(id)に関する一回目の処理、記録R[1]についての処理であるかを判別する。
- 「⑦ R[i]を「トリップ入場データ」に代入」では、「トリップ入場データ」にR[i]を代入する。

	ID	処理種別	処理日時	事業者名	駅名
R[i]	XXXX	入場	8:25:40	aaa 事業者	A 駅

図 43 ⑦におけるR[i]

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ					

図 44 ⑦における「トリップ入場データ」

- 「⑧ R[i]の処理種別が出場」では、R[i]の処理種別が出場であるかを判別する。
- 「⑨ R[i]を「トリップ出場データ」に代入」では、「トリップ出場データ」にR[i]を代入する。(※1 既に「トリップ出場データ」にデータが入っている場合は、R[i]で置き換える)

	ID	処理種別	処理日時	事業者名	駅名
	XXXX	入場	8:25:40	aaa 事業者	A 駅
R[i]	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 45 ⑨におけるR[i]

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 46 ⑨における「トリップ入場データ」と「トリップ出場データ」

※1

	ID	処理種別	処理日時	事業者名	駅名
	XXXX	入場	8:25:40	aaa 事業者	A 駅
	XXXX	出場	8:35:30	aaa 事業者	B 駅
	XXXX	入場	9:02:15	aaa 事業者	C 駅
R[i]	XXXX	出場	9:31:10	aaa 事業者	D 駅

図 47 ⑨におけるR[i] (トリップ出場データを置き換える場合)

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 48 ⑨における「トリップ入場データ」と「トリップ出場データ」（置き換え前）

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	9:31:10	aaa 事業者	D 駅

図 49 ⑨における「トリップ入場データ」と「トリップ出場データ」（置き換え後）

- 「⑩ (R[i]の処理時間)-(R[i-1]の処理時間) ≥ ある閾値時間」では、次のトリップまでの時間がある閾値時間以上であるかを判別する。これを満たす場合は、既に格納されている「トリップ入場データ」「トリップ出場データ」を一つのトリップとして確定する。満たさない場合は、入場データ R[i]を構成するトリップも「トリップ入場データ」「トリップ出場データ」のトリップと同一トリップとみなす。

R[i]	ID	処理種別	処理日時	事業者名	駅名
	XXXX	入場	8:25:40	aaa 事業者	A 駅
	XXXX	出場	8:35:30	aaa 事業者	B 駅
	XXXX	入場	10:02:15	aaa 事業者	C 駅
	XXXX	出場	10:31:10	aaa 事業者	D 駅

図 50 ⑩における R[i]（⑩を満たす場合）

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 51 ⑩における「トリップ入場データ」と「トリップ出場データ」（確定）

R[i]	ID	処理種別	処理日時	事業者名	駅名
	XXXX	入場	8:25:40	aaa 事業者	A 駅
	XXXX	出場	8:35:30	aaa 事業者	B 駅
	XXXX	入場	9:02:15	aaa 事業者	C 駅
	XXXX	出場	9:31:10	aaa 事業者	D 駅

図 52 ⑩における R[i]（⑩を満たさない場合）

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 53 ⑩における「トリップ入場データ」と「トリップ出場データ」（確定せず、処理継続）

- 「⑩「トリップ入場データ」と「トリップ出場データ」の情報を利用して「(id)の出力データ」にデータを追加」では、「トリップ入場データ」と「トリップ出場データ」の情報における、二次処理出力データの項目を記載したデータを「(id)の出力データ」に追加する。

	ID	処理種別	処理日時	事業者名	駅名
トリップ入場データ	XXXX	入場	8:25:40	aaa 事業者	A 駅
トリップ出場データ	XXXX	出場	8:35:30	aaa 事業者	B 駅

図 54 ⑩における「トリップ入場データ」と「トリップ出場データ」（確定済み）

ID	入場情報			出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	A 駅	8:25:40	aaa 事業者	B 駅	8:35:30

図 55 ⑩における「(id)の出力データ」

- 「⑫「トリップ入場データ」と「トリップ出場データ」初期化」では、あるトリップの入場データと出場データを格納するための変数を空で初期化する。
- 「⑬ R[i]を「トリップ入場データ」に代入」では、「トリップ入場データ」に R[i]を代入する。
- 「⑭ i について End」は、③の処理ループの終了を表す。
- 「⑮「トリップ入場データ」と「トリップ出場データ」の情報を利用して「(id)の出力データ」にデータを追加」では、「トリップ入場データ」と「トリップ出場データ」の情報における、二次処理出力データの項目を記載したデータを「(id)の出力データ」に追加する。
- 「⑯ (id)について End」は、②の処理ループの終了を表す。
- 「⑰ 終了」で処理を終了する。

本操作を行うことで、全データの 2 次 OD データが得られる。

6.5 三次処理（3 次 OD データ作成）

6.5.1. 目的

・各 ID の持つ 2 次 OD データ（入場/出場を一組にしたトリップデータであり、次のトリップまでの時間がある閾値時間以内の場合は同一トリップとみなしている）を元に、一日分の入場/出場の組を並べたトリップデータを作成すること。

・

6.5.2. 入力

入力データは二次処理後のデータである。

6.5.3. 出力

二次処理後のデータの各レコードは以下の項目が記載されている。

- ・ ID（再ハッシュ化済み）
- ・ 入場情報（事業者名、駅名、入場時間）
- ・ 出場時間（事業者名、駅名、出場時間）

6.5.4. 処理方法

本処理は、二次処理で出力した全てのトリップデータから、各 ID の一日分の入場情報/出場情報を並べたトリップデータを出力する処理である。すなわち、図 56 二次処理後の出力データのようなトリップデータからは、図 57 3 次 OD 作成における三次処理で生成した出力データのようなトリップデータを出力する。

ID	入場情報			出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	A 駅	8:25:40	aaa 事業者	D 駅	9:31:10
XXXX	aaa 事業者	E 駅	10:15:30	aaa 事業者	F 駅	11:24:20
XXXX	aaa 事業者	G 駅	12:32:50	aaa 事業者	H 駅	12:45:30

図 56 二次処理後の出力データ

ID	第 1 入場情報			第 1 出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	A 駅	8:25:40	aaa 事業者	D 駅	9:31:10

第 2 入場情報			第 2 出場情報		
事業者名	駅名	日時	事業者名	駅名	日時
aaa 事業者	A 駅	8:25:40	aaa 事業者	D 駅	9:31:10

第 3 入場情報			第 3 出場情報		
事業者名	駅名	日時	事業者名	駅名	日時
aaa 事業者	A 駅	8:25:40	aaa 事業者	D 駅	9:31:10

図 57 3 次 OD 作成における三次処理で生成した出力データ

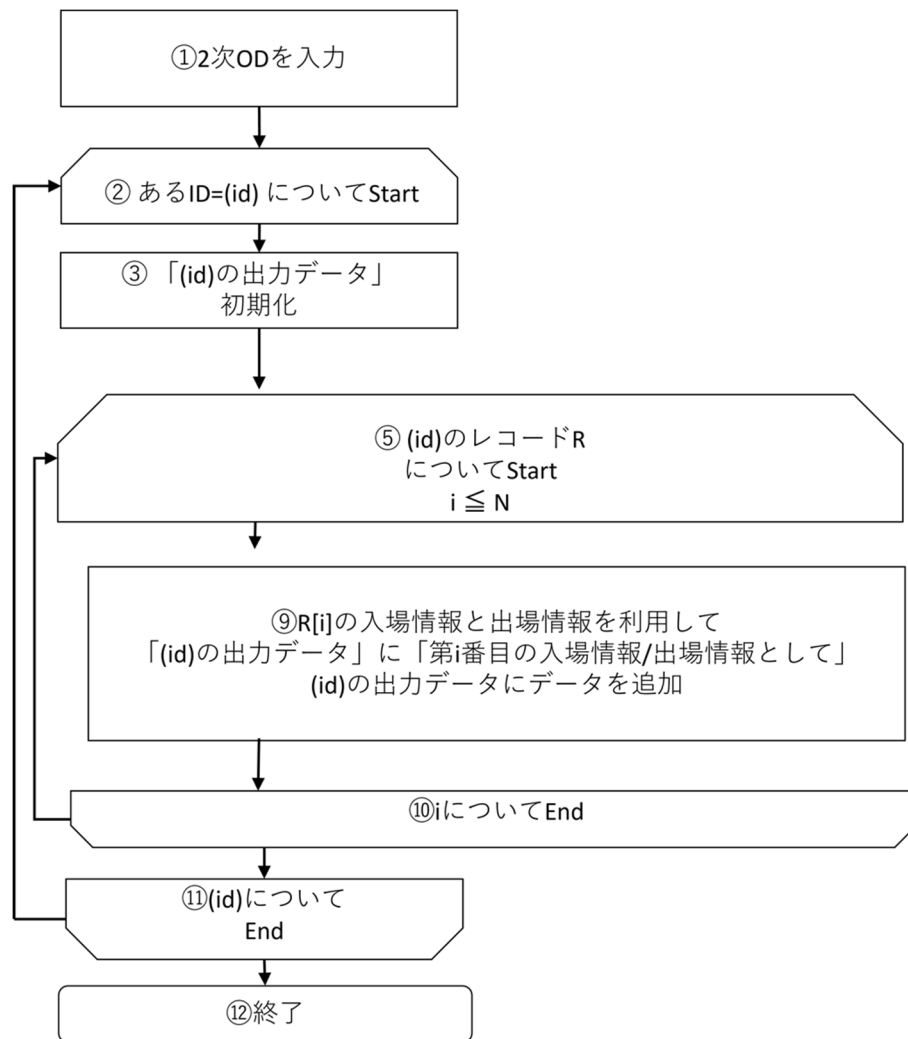


図 58 三次処理の処理手順

- 図 58 三次処理の処理手順の「① 2 次 OD を入力」では入力として、二次処理で作成した 2 次 OD を用いる。
- 「② ある ID=(id) について Start」では、ある ID についての処理のループを開始する。
- 「③ 「(id) の出力データ」 初期化」では、(id) に関する二次処理出力データを格納するための変数を空で初期化する。
- 「④ (id) のレコード R について Start」では、ID が (id) であるレコード R についてのループ処理を開始する。i は何回目のループであるかを表し、以降 R[i] は R の i 番目のレコードを示す。(ある ID のレコード数) = N とし、 $i \leq N$ を満たさなくなるまでループ処理を行う。
- 「⑨ R[i] の入場情報と出場情報を利用して 「(id) の出力データ」 に「第 i 番目の入場情報/出場情報」として (id) の出力データにデータを追加」では、R[i] の入場情報と出場情報を、それぞれ第 i 入場情報と第 i の出場情報として (id) の出力データに追加する。

ID	入場情報			出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	A 駅	8:25:40	aaa 事業者	D 駅	9:31:10

図 59 ⑨における R[i]

ID	第 1 入場情報			第 1 出場情報		
	事業者名	駅名	日時	事業者名	駅名	日時
XXXX	aaa 事業者	A 駅	8:25:40	aaa 事業者	D 駅	9:31:10

図 60 ⑨における (id) の出力データ

- 「⑩i について End」 は、③の処理ループの終了を表す。
- 「⑪ (id) について End」 は、②の処理ループの終了を表す。
- 「⑫ 終了」 で処理を終了する。

本操作を行うことで、全データの 3 次 OD データが得られる。

6.6 最終処理

6.6.1. 駅名によるエリア分け

第 12 回大都市交通センサス調査結果集計表の WEB ページより各エリアの鉄道駅コードをダウンロードする。

参照: [公共交通政策：第 12 回大都市交通センサス調査結果集計表 - 国土交通省 \(mlit.go.jp\)](http://mlit.go.jp)

【鉄道調査】コードデータファイル(Excel形式)

1	ゾーンコード表	首都圏	中京圏	近畿圏
2	鉄道事業者コード	首都圏	中京圏	近畿圏
3	鉄道路線コード	首都圏	中京圏	近畿圏
4	鉄道駅コード	首都圏	中京圏	近畿圏

この鉄道駅コード表が作成されたのが平成 28 年のため、それ以降に発生した以下に對してのデータ修正・追加を行う。詳細内容については表 1 駅名変更、表 2 新駅追加参照。

(1) 駅名変更

(2) 新駅追加

上記を修正・追加したコード表の「事業者名」「駅名」によりエリア分けを行う。

6.6.2. データ付与と削除

- ・ 3 次処理で出力された結果と、上記 1 項のコード表の「事業者名」「駅名」をキーにして入場駅、出場駅に結合し該当する「エリア」情報を各レコードに付加する。
- ・ 「ID」をデータから削除する。
- ・ 「入場時間」から時間のみを抽出したデータに変換する。
- ・ 「出場時間」を削除し、「入場時間」から「出場時間」までの「所要時間」を 5 分単位で付与する。

集計の際に、乗降客数がそれぞれ 10 人以下の駅については、集計結果から除外している。

7. 集計表

以下に、各 OD データ処理ごとの集計結果を示す。

7.1 一次処理集計件数

7.1.1. 一次処理集計件数

0 次 OD データ、1 次 OD データ、2 次 OD データ、3 次 OD データの一次処理（クレンジング）における集計結果は、表 3 2021 年 12 月某日（1 日目）の各 IC カード媒体の一次処理データ件数、表 4 2021 年 12 月某日（2 日目）の各 IC カード媒体の一次処理データ件数の通りである。

No	カード種別	件数	除外件数	有効数	除外率
1	PiTaPa	3,672,964	11,328	3,661,636	0.3%
2	ICOCA	4,462,778	48,462	4,414,316	1.1%
3	TOICA	220,014	3,938	216,076	1.8%
4	モノレール Suica	48,384	300	48,084	0.6%
5	りんかい Suica	31,291	143	31,148	0.5%
6	PASMO	11,900,896	145,980	11,754,916	1.2%
7	manaca	1,143,003	23,291	1,119,712	2.0%
8	Suica	15,503,681	11,449	15,492,232	0.1%
	合計	36,983,011	244,891	36,738,120	0.7%

表 3 2021 年 12 月某日（1 日目）の各 IC カード媒体の一次処理データ件数

No	カード種別	件数	除外件数	有効数	除外率
1	PiTaPa	3,737,177	12,855	3,724,322	0.3%
2	ICOCA	4,731,271	49,883	4,681,388	1.1%
3	TOICA	235,459	3,979	231,480	1.7%
4	モノレール Suica	51,638	328	51,310	0.6%
5	りんかい Suica	33,778	118	33,660	0.3%
6	PASMO	12,904,693	156,159	12,748,534	1.2%
7	manaca	1,180,269	20,957	1,159,312	1.8%
8	Suica	16,783,008	12,848	16,770,160	0.1%
	合計	39,657,293	257,127	39,400,166	0.6%

表 4 2021 年 12 月某日（2 日目）の各 IC カード媒体の一次処理データ件数

7.2 二次処理集計件数

7.2.1. 0 次 OD 二次処理集計件数

0 次 OD データの二次処理（0 次 OD データ作成）における集計結果は、表 5 2021 年 12 月某日 2 日間の 0 次データ二次処理データ件数、表 6 2021 年 12 月某日（1 日目）×2 の 0 次データ二次処理データ件数、表 7 2021 年 12 月某日（2 日目）×2 の 0 次データ二次処理データ件数の通りである。

ただし、対象となる 1 日目のデータに日付を 1 日追加したデータを追加したものを「2021

年 12 月某日（1 日目）×2」、対象となるの 2 日目のデータに日付を 2 日追加したデータを追加したものを「2021 年 12 月某日（2 日目）×2」と表記している。

※1 0 次 OD データ特性上、有効件数に ID 数の 2 倍を加えたものとなる。

No	件数	除外件数	有効件数	レコード数(CSV) ※1
1	38,024,326	23,267,288	14,757,038	49,369,505

表 5 2021 年 12 月某日 2 日間の 0 次データ二次処理データ件数

No	件数	除外件数	有効件数	レコード数(CSV) ※1
1	36,695,370	15,000,208	21,695,162	43,919,552

表 6 2021 年 12 月某日（1 日目）×2 の 0 次データ二次処理データ件数

No	件数	除外件数	有効件数	レコード数(CSV) ※1
1	39,353,282	16,003,350	23,349,932	47,053,949

表 7 2021 年 12 月某日（2 日目）×2 の 0 次データ二次処理データ件数

7.2.2. 1 次 OD 二次処理集計件数

1 次 OD データの二次処理（1 次 OD データ作成）における集計結果は、表 8 各日付データの 1 次 OD データ件数の通りである。

※2 入出場で 1 件なので有効数の 1/2 となる。

No	日付	件数	レコード数(CSV) ※2
1	2021 年 12 月某日 (1 日目)	36,738,120	18,369,060
2	2021 年 12 月某日 (2 日目)	39,400,166	19,700,083

表 8 各日付データの 1 次 OD データ件数

7.2.3. 2 次 OD・3 次 OD 二次処理集計件数

2 次 OD データの二次処理（2 次 OD データ作成）、3 次 OD データの二次処理（2 次 OD データ作成）における集計結果は、表 9 各日付データの 2 次 OD データ件数（トリップ接続閾値 60 分）、表 10 各日付データの 2 次 OD データ件数（トリップ接続閾値 30 分）、表 11 各日付データの 2 次 OD データ件数（トリップ接続閾値 15 分）の通りである。

※3 入出場で 1 件なので有効数の 1/2 となる。

No	日付	件数	除外件数	接続除外件数	有効数	レコード数 (CSV)※3
1	2021 年 12 月某日 (1 日目)	36,738,120	1,327,730	8,703,764	26,706,626	13,353,313
2	2021 年 12 月某日 (2 日目)	39,400,166	1,405,474	9,357,074	28,637,618	14,318,809

表 9 各日付データの 2 次 OD データ件数（トリップ接続閾値 60 分）

No	日付	件数	除外件数	接続除外件数	有効数	レコード数 (CSV)※3
1	2021 年 12 月某日 (1 日目)	36,738,120	1,327,730	7,678,106	27,732,284	13,866,142
2	2021 年 12 月某日 (2 日目)	39,400,166	1,405,474	8,184,922	29,809,770	14,904,885

表 10 各日付データの 2 次 OD データ件数（トリップ接続閾値 30 分）

No	日付	件数	除外件数	接続除外件数	有効数	レコード数 (CSV)※3
1	2021 年 12 月某日 (1 日目)	36,738,120	1,327,730	7,104,138	28,306,252	14,153,126
2	2021 年 12 月某日 (2 日目)	39,400,166	1,405,474	7,524,970	30,469,722	15,234,861

表 11 各日付データの 2 次 OD データ件数（トリップ接続閾値 15 分）

7.3 三次処理集計件数

7.3.1. 3 次 OD 三次処理集計件数

3 次 OD データの三次処理（3 次 OD データ作成）における集計結果は、表 12 各日付データの 3 次 OD データ件数（トリップ接続閾値 60 分）、表 13 各日付データの 3 次 OD データ件数（トリップ接続閾値 30 分）、表 14 各日付データの 3 次 OD データ件数（トリップ接続閾値 15 分）の通りである。

No	日付	件数	除外件数	接続除外件数	有効数
1	2021 年 12 月某日 (1 日目)	13,353,313	2,839	6,216,879	7,133,595
2	2021 年 12 月某日 (2 日目)	14,318,809	2,954	6,702,311	7,613,544

表 12 各日付データの 3 次 OD データ件数（トリップ接続閾値 60 分）

No	日付	件数	除外件数	接続除外件数	有効数
1	2021 年 12 月某日 (1 日目)	13, 866, 142	1, 886	6, 730, 291	7, 133, 965
2	2021 年 12 月某日 (2 日目)	14, 904, 885	2, 090	7, 288, 939	7, 613, 856

表 13 各日付データの 3 次 OD データ件数（トリップ接続閾値 30 分）

No	日付	件数	除外件数	接続除外件数	有効数
1	2021 年 12 月某日 (1 日目)	14, 153, 126	1, 892	7, 017, 238	7, 133, 996
2	2021 年 12 月某日 (2 日目)	15, 234, 861	2, 094	7, 618, 880	7, 613, 887

表 14 各日付データの 3 次 OD データ件数（トリップ接続閾値 15 分）

7.4 最終処理集計件数

各 OD データの最終処理件数は、表 15 各日付データの最終処理後圏域別 0 次 OD 件数、表 16 各日付データの最終処理後圏域別 1 次 OD 件数、表 17 各日付データの最終処理後圏域別 2 次 OD 件数（トリップ接続閾値 60 分）、表 18 各日付データの最終処理後圏域別 2 次 OD 件数（トリップ接続閾値 30 分）、表 19 各日付データの最終処理後圏域別 2 次 OD 件数（トリップ接続閾値 15 分）、表 20 各日付データの最終処理後圏域別 3 次 OD 件数の通りである。

7.4.1. 0 次 OD 最終処理集計件数

No	ファイル名	日付	首都圏	中京圏	近畿圏
1	0ji_0_1.csv ～ 0ji_0_9.csv	2021 年 12 月某日 (2 日間)	18, 736, 825	949, 330	5, 802, 058
2	0ji_1_1.csv ～ 0ji_1_11.csv	2021 年 12 月某日 (1 日目×2)	20, 463, 181	1, 114, 610	6, 529, 307
3	0ji_2_1.csv ～ 0ji_2_12.csv	2021 年 12 月某日 (2 日目×2)	22, 263, 694	1, 158, 891	6, 781, 166

表 15 各日付データの最終処理後圏域別 0 次 OD 件数

7.4.2. 1次OD最終処理集計件数

No	ファイル名	日付	首都圏	中京圏	近畿圏
1	1ji_1_1.csv ～ 1ji_1_6.csv	2021年 12月某日 (1日目)	12,834,456	756,950	4,195,731
2	1ji_2_1.csv ～ 1ji_2_6.csv	2021年 12月某日 (2日目)	13,944,849	785,415	4,357,067

表 16 各日付データの最終処理後圏域別1次OD件数

7.4.3. 2次OD最終処理集計件数

No	ファイル名	日付	首都圏	中京圏	近畿圏
1	2ji_60_1_1.csv ～ 2ji_60_1_7.csv	2021年 12月某日 (1日目)	9,378,546	642,137	3,329,791
2	2ji_60_2_1.csv ～ 2ji_60_2_7.csv	2021年 12月某日 (2日目)	10,188,469	667,340	3,460,046

表 17 各日付データの最終処理後圏域別2次OD件数（トリップ接続閾値 60 分）

No	ファイル名	日付	首都圏	中京圏	近畿圏
1	2ji_30_1_1.csv ～ 2ji_30_1_7.csv	2021年 12月某日 (1日目)	9,756,561	659,587	3,448,108
2	2ji_30_2_1.csv ～ 2ji_30_2_7.csv	2021年 12月某日 (2日目)	10,630,599	685,920	3,586,276

表 18 各日付データの最終処理後圏域別2次OD件数（トリップ接続閾値 30 分）

No	ファイル名	日付	首都圏	中京圏	近畿圏
1	2ji_15_1_1.csv ～ 2ji_15_1_7.csv	2021年 12月某日 (1日目)	9,968,824	668,538	3,513,872
2	2ji_15_2_1.csv ～ 2ji_15_2_7.csv	2021年 12月某日 (2日目)	10,880,277	695,370	3,657,120

表 19 各日付データの最終処理後圏域別2次OD件数（トリップ接続閾値 15 分）

7.4.4. 3次OD最終処理集計件数

No	ファイル名	日付	接続閾値	件数
1	3ji_60_1_1.csv ～ 3ji_60_1_6.csv	2021年 12月某日 (1日目)	60分	7,133,595
2	3ji_60_2_1.csv ～ 3ji_60_2_7.csv	2021年 12月某日 (2日目)	60分	7,613,544
3	3ji_30_1_1.csv ～ 3ji_30_1_6.csv	2021年 12月某日 (1日目)	30分	7,133,965
4	3ji_30_2_1.csv ～ 3ji_30_2_7.csv	2021年 12月某日 (2日目)	30分	7,613,856
5	3ji_15_1_1.csv ～ 3ji_15_1_6.csv	2021年 12月某日 (1日目)	15分	7,133,996
6	3ji_15_2_1.csv ～ 3ji_15_2_7.csv 3ji_15_2c.csv	2021年 12月某日 (2日目)	15分	7,613,887

表 20 各日付データの最終処理後圏域別3次OD件数

※なお、公表データについては、さらに統計上の処理等を行っており、上記の数値と異なる。

8. 補足説明

8.1 所要時間の考え方

あるトリップに関する所要時間は、利用した区間によって以下のように異なる。

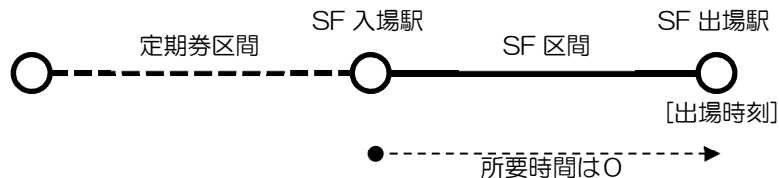
- ・ IC カード利用

入場時刻から出場時刻までに要する時間とする。



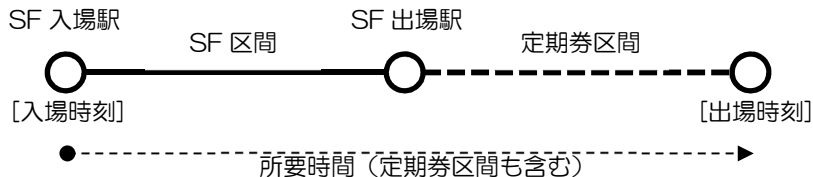
- ・ IC カードの定期券区間内から乗車し、定期券区間外で下車

定期券区間外の入場時刻が不明なため入場時刻を出場時刻とし所要時間は0とする。



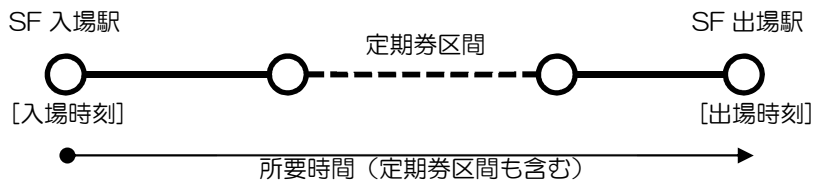
- ・ IC カードの定期券区間外から乗車し、定期券区間内で下車

定期券区間内の出場時刻が不明なため入場時刻から定期券区間内の出場時刻までに要する時間とする。



- ・ IC カードの定期区間外から乗車し、定期券区間を使用し、定期区間外で下車

入場時刻から出場時刻までに要する時間とする。



2023 年 3 月 17 日版