データの集計と
行動モデルの基礎

Data Aggregation and the Basis of Behavior Model

M2
大山雄己

No.
0

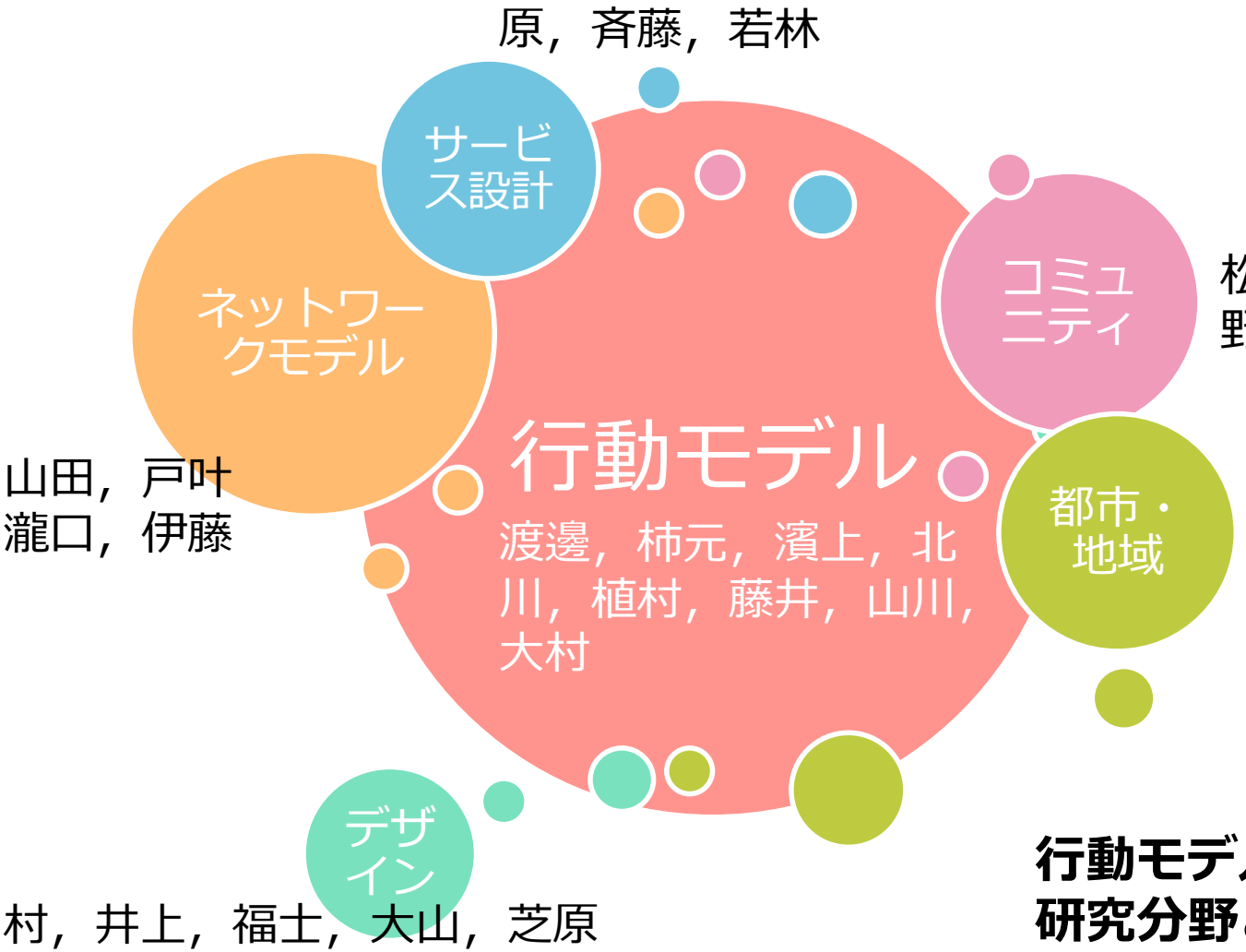
Flow of presentation

ゼミの流れ

0．はじめに
1．データの種類と説明
2．基礎集計の方法
3．行動モデルの基礎
4．課題

はじめに

■生活研ってどんな研究をしているんですか？



行動モデルはさまざまな研究分野と関係性が強い

はじめに

■主な研究テーマ

主な学位論文研究テーマ

１）街路特性	→渡辺（2008），柿元（2012）
２）経路選択モデル	→山川（2009）
３）歩行者	→濱上（2008），北川（2009），植村（2010）
４）ソーシャルネットワーク	→浦田（2009）
５）コミュニティと発話	→松村（2009），亀田（2010），野末（まち大）
６）ハバナ	→樋口（まちづくり大学院）
７）広場	→福山（まちづくり大学院）
８）都市類型	→中埜（まちづくり大学院）
９）アクティビティモデル	→藤井（2009），山田（2010）
１０）地域での活動	→斎藤有（2010）
１１）交通管制	→戸叶（2012），瀧口（2010）
１２）デザイン	→中村（2007），井上（2008），福士（2011）
１３）新交通サービス設計	→原（2012），斉藤（2012），若林（2013）
１４）移動データの精緻化	→大村（2012），今泉（2013）
１５）住宅	→國分（まち大）
１６）回遊シミュレーション	→伊藤（2012）
１７）行動圏域	→池田（2012）

はじめに

■では，行動モデルとは？

さまざまな**人の選択行動＝意思決定**を表現するモデル.

仮定1：効用最大化理論

各選択肢の望ましさを表す「効用」を考え，個人は最も「効用」が大きい行動選択肢を選択する.

$$U_{in} = V_{in} + \varepsilon_{in}$$

仮定2：確率的意思決定

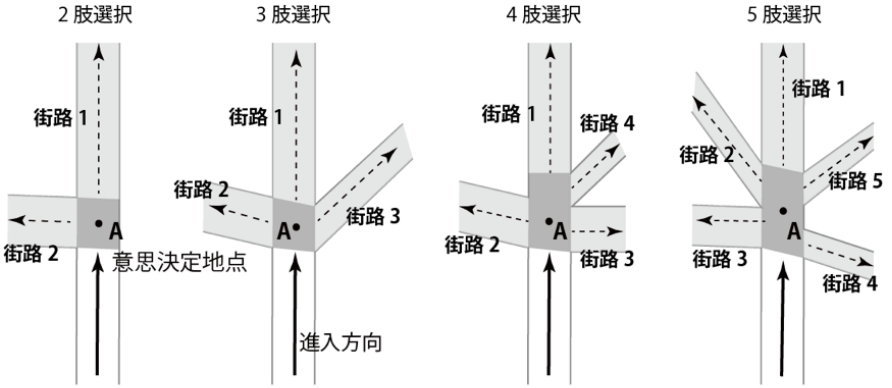
効用の要因である変数すべてを観測することはできない．非観測要因によって「効用」の大きさは確率的に変動する.

$$P_{in} = \frac{\exp(\mu V_{in})}{\sum_{j \in C} \exp(\mu V_{ij})}$$

はじめに

■どのような研究があるか

各交差点を意思決定地点として、
2～5の接続する街路を逐次選択
するモデル



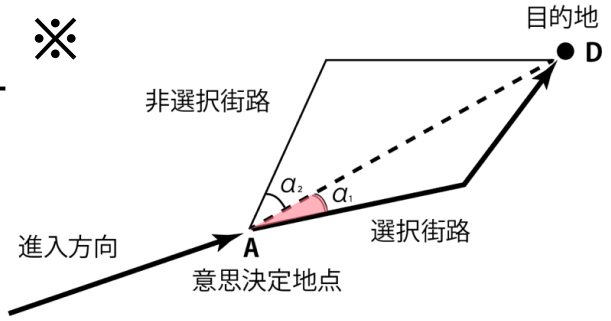
$$V_n = \beta_1 X_{destination} + \beta_2 X_{shortest} + \beta_3 X_{straight} + \sum_i \beta_i X_i$$

(1) (2)

(1)空間定位に関する説明変数

目的地方向角度 α（※），最短経路ダミー，直進ダミー

※



(2)空間特性に関する説明変数

A：街路空間の形態と機能を表す説明変数
B：街路景観パタンダミー（13パタン）

1：空間定位，2：空間定位＋構成要素（A），3：空間定位＋景観パタン
の3通りをモデルで推定

はじめに

■モデル推定結果

	説明変数	空間定位 推定値	空間定位 t値	空間定位＋空間構成要素 推定値	空間定位＋空間構成要素 t値	空間定位＋街路景観パタン 推定値	空間定位＋街路景観パタン t値
空間定位	空間特性の重み	0.724	7.72**	-0.0[illegible]	[illegible]	[illegible]	[illegible]**
	直進ダミー	0.720	3.61**	0.294	1.16	0.550	2.53*
	最短経路方向ダミー	0.876	6.00**	0.656	4.07**	0.721	4.56**
構成要素	歩道幅員			0.102	2.72**		
	平均間口長			-0.061	-3.20**		
	リンク幅			0.033	1.48		
街路景観パタン	街区内大通り（①）パタンダミー					1.251	3.52**
	センター街（⑦）パタンダミー					0.958	3.46**
	裏道店舗無（①）パタンダミー					-1.272	-1.58
サンプル数			306		306		306
初期尤度			-266.0		-266.0		-266.0
最終尤度			-193.0		-171.9		-177.6
尤度比			0.275		0.354		0.332
修正済尤度比			0.263		0.331		0.310

＊：5%有意 ＊＊：1%有意

空間特性を考慮したモデルでは，直進志向性が弱まる

構成要素や街路景観パタンが経路選択行動に有意な影響を与えている

一般的な分析の流れ

■どのように分析していくか.

（ものすごく簡単な例）

このあたりまで紹介します.

・データの特性を知る.

↓

分析を行なう前に，データに馴染む必要があります．どのような情報が得られているのか，データから何がわかるのかを把握しておきましょう.

・クロス集計を行い，傾向を探る.

↓

基礎的な集計として，様々な属性を掛けあわせて何と何に相関があるのか，行動の要因になっているものは何か，分析します.

・モデルを構築し，推定を行なう.

クロス集計結果から仮説が立ったら，モデルを構築して因果関係を定量的に分析します．ある選択に対して何が効いているのかを把握します.

データの種類と説明

No. 6

データの種類と説明

■どんなデータがあるか？

1）行動データ

他にも：検知器（断面），利用ログ（PASMO等），道路交通センサス（統計）

Bcals：加速度，歩数，運動負荷などの詳細な移動文脈情報

2）質的データ

- アンケートデータ（Web Diary），RP/SP調査
- ヒアリング（音声）

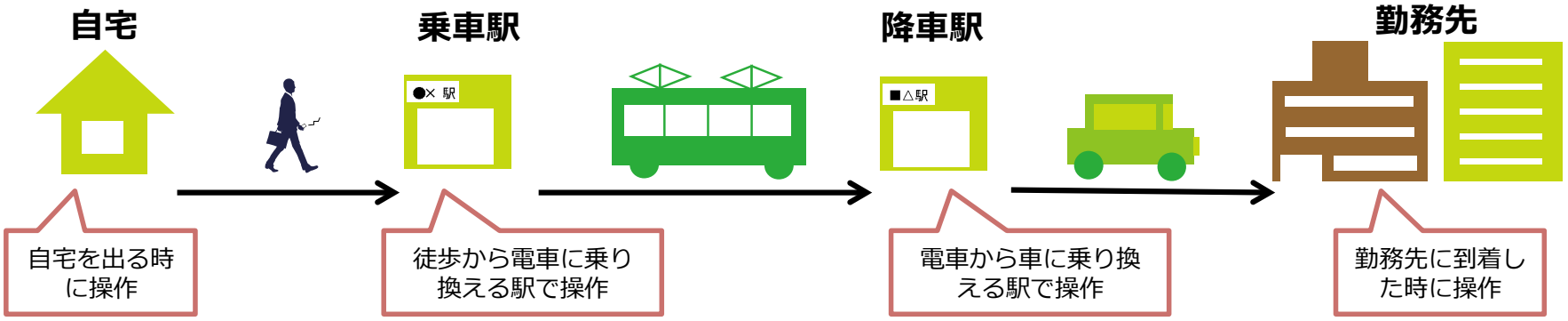
データの種類と説明

■PP（プローブパーソン）データとは？

GPS機能を搭載した携帯電話と移動通信機器と連動したWebダイアリーを用いてモニタの移動活動記録と数秒間隔の位置情報を取得できる

- **大量かつ詳細な移動データ**
- **day-to-dayの行動記録**
  （同一個人の複数日に渡る行動履歴）

移動手段
移動目的
個人属性
平均速度
トリップ長
トリップ時間
…



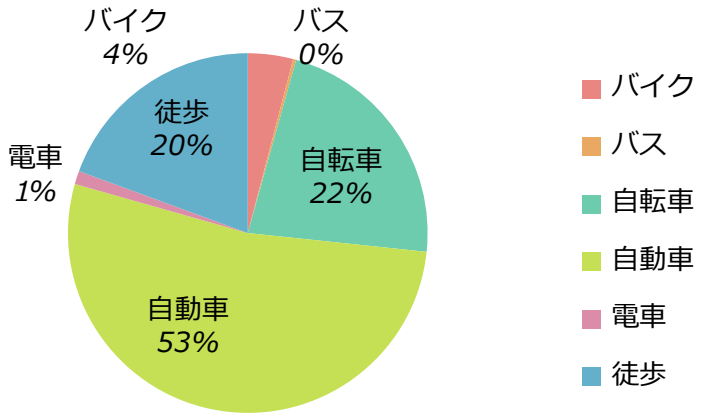
大きく，**location data**，**trip data** の2種類がある.

No. 8

データの種類と説明

■trip data（トリップごと）

tripID, userID, 移動目的・手段,
出発・到着時刻，出発地・到着地位置情報



個人情報保護のため除いています.

データの種類と説明

■location data（5～10秒間隔）

tripID, locationID, userID, 移動手段,
時刻，位置座標，測位モード

個人情報保護のため除いています.

基礎集計の方法

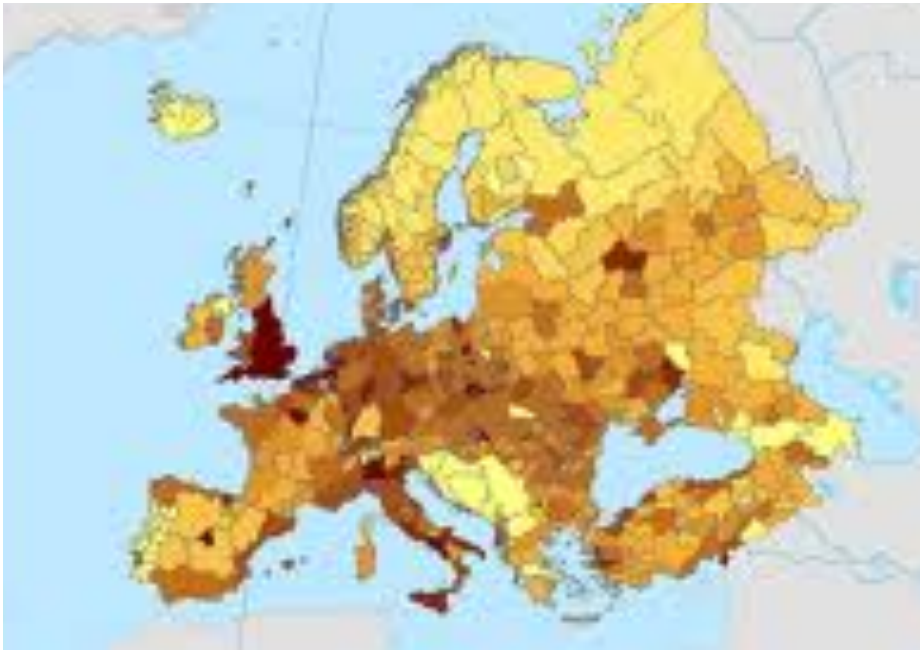
No. 10

基礎集計の方法

■集計に使うソフトウェア

1）データ整理/正規化
2）データ集計　　**R/Java/Excel…**

3）可視化　　**R/GIS/Excel/Google Earth…**



No. 11

基礎集計の方法

■データ整理/正規化

いくらExcelでも，ボタンひとつではグラフも描けません.
自分の目的に合わせて，まずデータを整理する必要があります.

データクリーニング／補正／マーケット・セグメンテーション…

・トリップデータ（1回の移動＝1行）

80010	80010	自転車	2007/12/12 07:55:37	2007/12/12 08:09:13	816
80568	80568	バイク	2007/12/12 20:15:15	2007/12/12 20:25:38	623
80696	80696	バイク	2007/12/13 07:59:09	2007/12/13 08:11:28	739
81160	81160	バイク	2007/12/13 19:51:42	2007/12/13 20:13:03	1281
81292	81292	バイク	2007/12/14 07:55:45	2007/12/14 08:06:26	641
81696	81696	タクシー	2007/12/14 17:54:21	2007/12/14 18:12:16	1075
81847	81847	タクシー	2007/12/14 23:37:43	2007/12/14 23:49:45	722
81938	81938	徒歩	2007/12/15 08:51:16	2007/12/15 09:11:08	1192
181939	181939	電車	2007/12/15 09:11:08	2007/12/15 09:19:36	508
181940	181940	徒歩	2007/12/15 09:19:36	2007/12/15 09:26:27	411

・ツアーデータ（1日の行動＝1行）

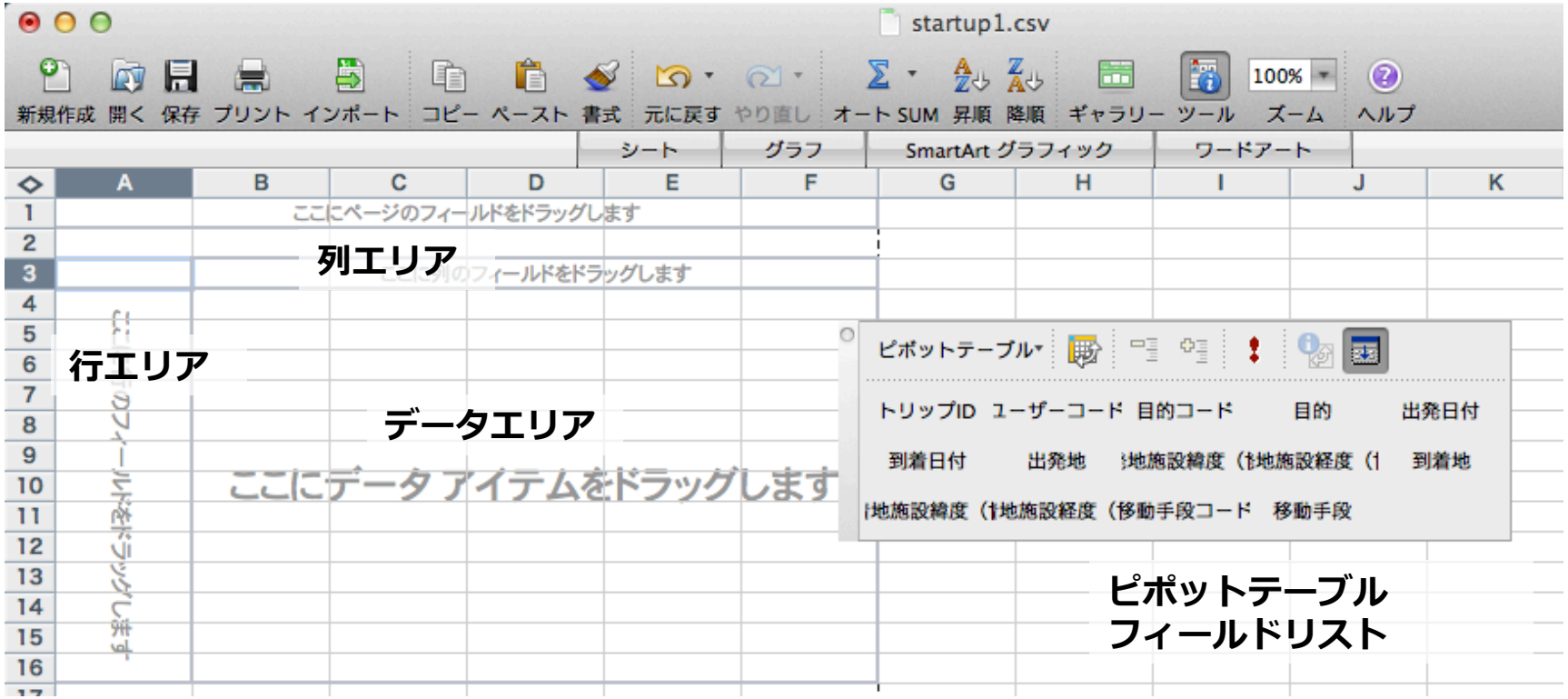
1707		HWH	2	2007/12/13 08:11:28	2007/12/13 19:51:42	42014	バイク	バイク	出勤・登校
1708		HWOH	3	2007/12/14 08:06:26	2007/12/14 17:54:21	35275	バイク	タクシー	出勤・登校
1709		HOOOOHOH	7	2007/12/15 09:19:36	2007/12/15 11:11:24	6708	電車	バイク	null
1710		HOHOH	4	2007/12/16 16:24:51	2007/12/16 18:45:34	8443	自動車	自動車	買い物
1726		HWH	2	2007/12/11 08:21:54	2007/12/11 19:17:49	39355	自転車	自転車	出勤・登校
1727		HWH	2	2007/12/12 08:20:27	2007/12/12 17:46:39	33972	自転車	自転車	出勤・登校
1728		HWH	2	2007/12/13 08:19:20	2007/12/13 11:57:01	13061	自転車	自転車	出勤・登校
1729		HWH	2	2007/12/14 08:20:48	2007/12/14 21:59:30	49122	自転車	自転車	出勤・登校

No. 12

基礎集計の方法

■ Excelでのクロス集計：ピポットテーブル

▶「データ」→「ピボットテーブルレポート」



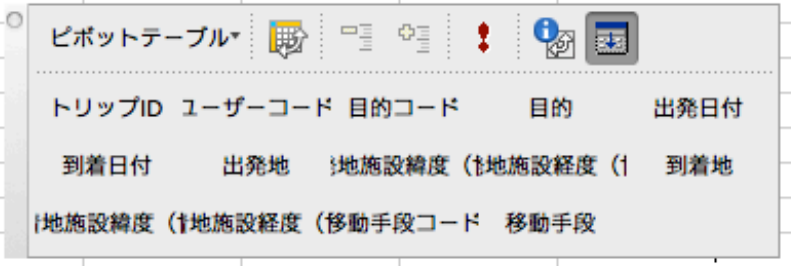
例えば，移動目的と移動手段の関係性が知りたい．

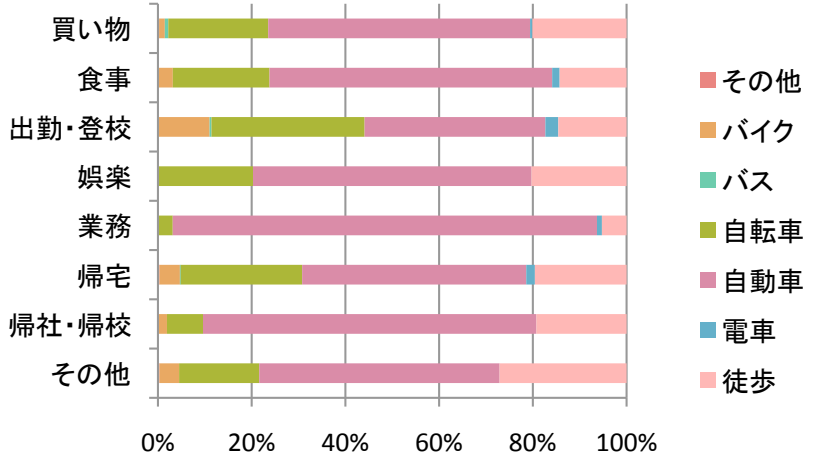
No. 13 基礎集計の方法

■ Excelでのクロス集計：ピポットテーブル

行エリアに目的，列エリア・データエリアに移動手段を入れれば，
目的別移動手段分担率が出る．

データの個数：移動手段	移動手段							
目的	その他	バイク	バス	自転車	自動車	電車	徒歩	総計
その他	1	13		53	159		84	310
帰社・帰校		1		4	37		10	52
帰宅	2	26	1	155	286	11	117	598
業務				3	86	1	5	95
娯楽				14	41		14	69
出勤・登校		28	1	83	98	7	37	254
食事		2		13	38	1	9	63
買い物	1	5	3	87	228	2	82	408
総計	4	75	5	412	973	22	358	1849





グラフにすれば，傾向がよりわかりやすい．（業務は自動車分担率が高く徒歩が少ないなど…）

（グラフ→100%積み上げ横棒）

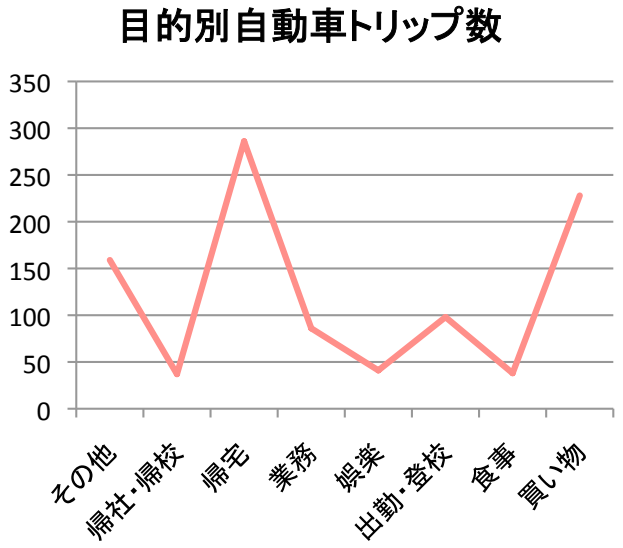
No. 14 基礎集計の方法

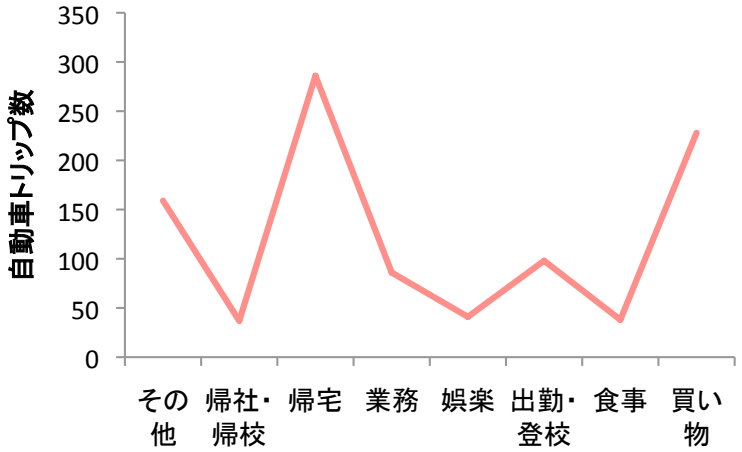
■集計結果の可視化

集計した結果はわかりやすいように
グラフ，表としてまとめましょう．

注意点

●縦軸，横軸には項目名・単位を記入
[レイアウト]→[軸ラベル]
●有効数字を揃える
[軸の書式設定]→[表示形式]→[数値]
●軸ラベルは斜めにしない
[軸の書式設定]→[配置]
●グラフ名、グラフエリア外枠を消す
[グラフエリアの書式設定]
●グラフの張り付けは拡張メタファイル形式
[貼り付け]→[形式を選択して貼り付け]
→[図（拡張メタファイル）]





行動モデルの基礎

行動モデルの基礎

■行動モデルとは？

さまざまな**人の選択行動＝意思決定**を表現するモデル.

「**離散選択モデル**」を指すことが多い？

仮定1：効用最大化理論

各選択肢の望ましさを表す「効用」を考え，個人は最も「効用」が大きい行動選択肢を選択する.

仮定2：確率的意思決定

効用の要因である変数すべてを観測することはできない．非観測要因によって「効用」の大きさは確率的に変動する.

行動モデルの基礎

■効用をどう記述するか.

▶効用関数

$$U_{in} = V_{in} + \varepsilon_{in}$$

U_{in} ：選択肢 i の効用　　V_{in} ：効用の確定項　　ε_{in} ：効用の確率項

▶確定項？…観測要因で記述できる効用

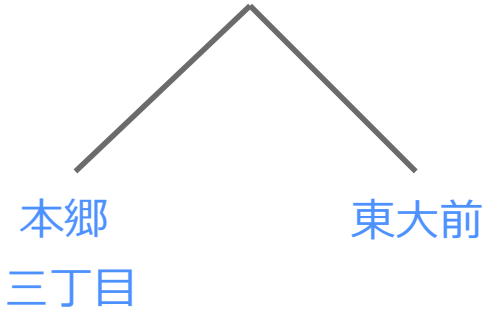
$$V_{in} = \sum_{k} \beta_k X_{ink}$$
$$= \beta_1 X_{in1} + \beta_2 X_{in2} + \ldots + \beta_K X_{inK}$$

X_{ink} ：説明変数（効用を変化させる要因）　　β_k ：パラメータ（どのくらい効いてるか）

行動モデルの基礎

■例：路線・駅の選択

後楽園駅からメトロに乗って学校に向かうとき，
「本郷三丁目」と「東大前」のどちらを利用するか？



▶説明変数は？

駅から学校（建物）までの距離，駅周りの店の数，所要時間，運行頻度，
改札階との高低差（後楽園）…

$$V_H = \beta_1 X_{dist,H} + \beta_2 X_{shop,H} + \beta_3 X_{time,H} + \beta_4 X_{freqency,H} + \beta_H$$

$$V_T = \beta_1 X_{dist,T} + \beta_2 X_{shop,T} + \beta_3 X_{time,T} + \beta_4 X_{freqency,T}$$

他にも，ダミー変数（個人属性などの定量化できない変数）などを説明変数として用いることがあるが，定数項やダミー変数はどちらかの選択肢にのみ入れる．

行動モデルの基礎

■例：路線・駅の選択

▶選択確率

本郷三丁目（H）が選択される確率は，

$$P_{iH} = \Pr[U_{iH} \geq U_{iT}]$$

誤差項にガンベル分布（Closed form）を仮定する．

$$P_{in} = \frac{1}{1+\exp(-\mu(V_{in} - V_{im}))} = \frac{\exp(\mu V_{in})}{\exp(\mu V_{in}) + \exp(\mu V_{im})}$$

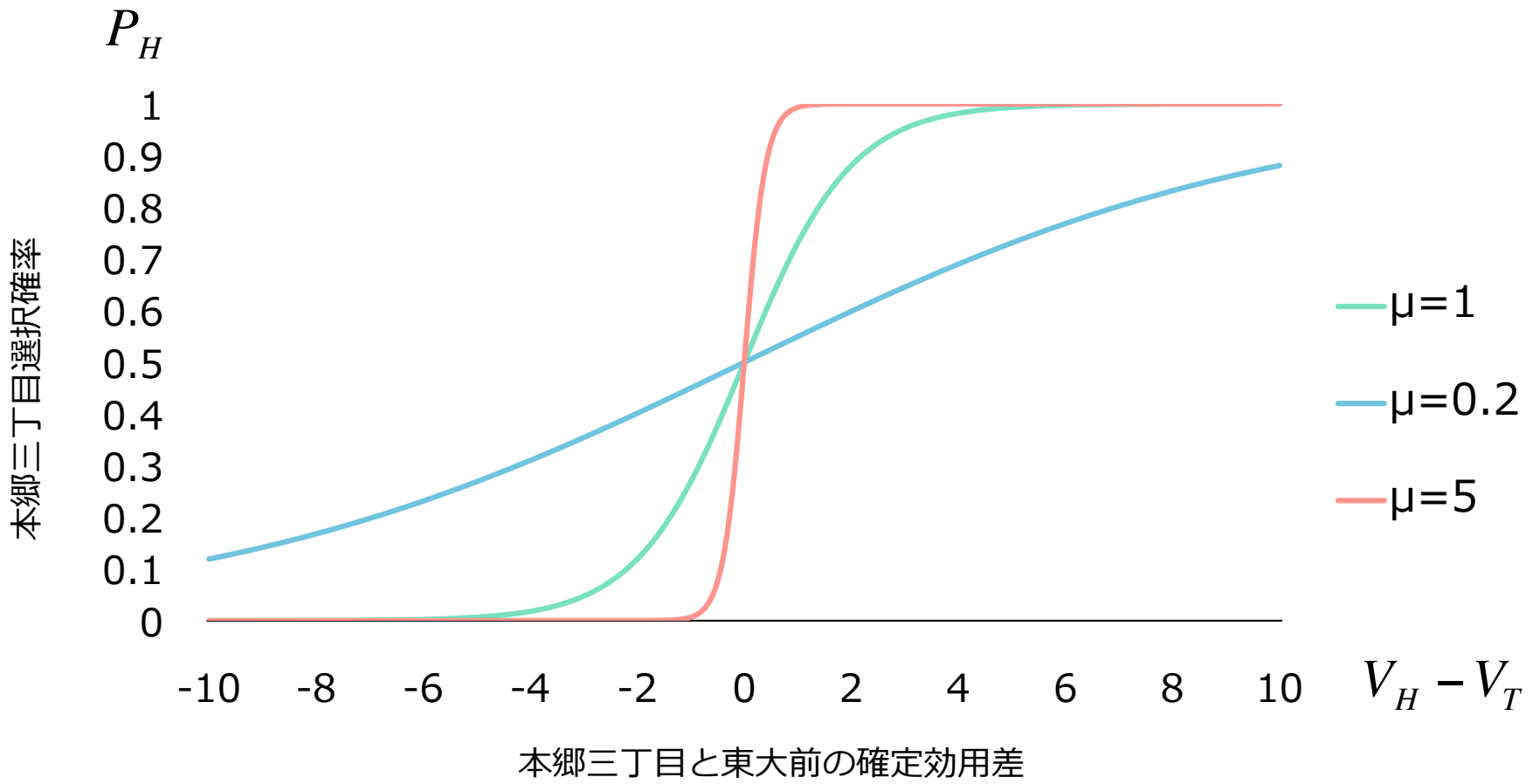
選択確率は最終的に，効用の確定項（**確定効用の差**）のみを用いて表される．

※ガンベル分布

$$f(\varepsilon) = e^{-\varepsilon} e^{-e^{-\varepsilon}}$$

行動モデルの基礎

■スケールパラメータ μ



行動モデルの基礎

■どのようにパラメータは決定するのか？

▶尤度（もっともらしさ）を最大化

$$L = \prod_i \prod_n P_{in}^{\ d_{in}}$$

d_{in} は個人iが選択肢nを選択したとき1，それ以外0.

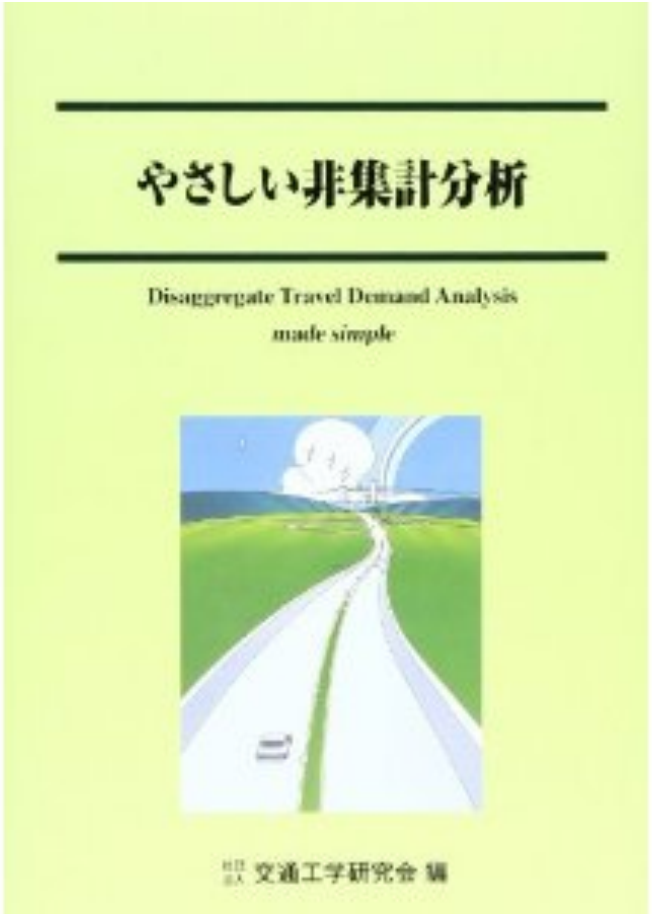
実際の推定では，計算を簡便化させるために以下の対数尤度を用いる.

$$\ln L = \sum_i \sum_n d_{in} \ln P_{in}$$

Newton-Raphson法などを用いて対数尤度を最大化させるパラメータを求める.

宿題

■6/26（水）まで



・担当を決め，「やさしい非集計分析」
　1章～3章を各自まとめてくる.
・ppt：10枚程度
・もう1人は四段階推定法について.