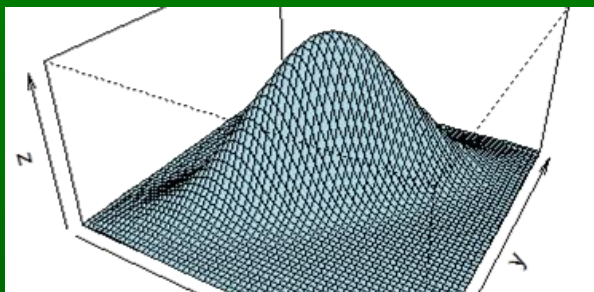


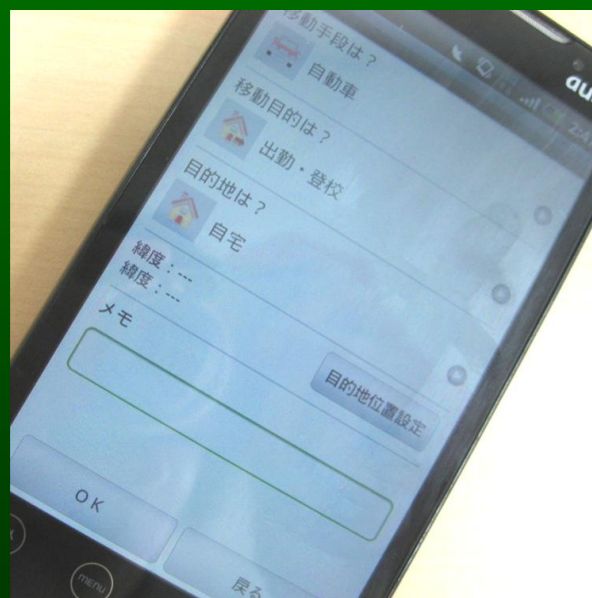
スタートアップゼミ2012

第1回 データ基礎、クロス集計、行動モデル基礎



$$P_1 = \frac{\exp(\mu V_1)}{\exp(\mu V_1) + \exp(\mu V_2)}$$

A hand is shown writing the equation on a chalkboard. An arrow points to the denominator of the equation.



目次

・データ概要

PP調査データ、PT調査データの仕様

・おもなソフトの紹介

R、Eclipse、プロアトラス等の概要とインストール

・Excel

よく使う操作と、クロス分析について

・行動モデル基礎

行動モデルの基礎の理論、次回以降の推定につなげる

課題

第2回へ

取り扱う行動データ

- ・プローブパーソン調査 (PP) データ
- ・パーソントリップ調査 (PT) データ
- ・映像データ

- ・アンケートデータ



(道後での広場の人の流れの調査)

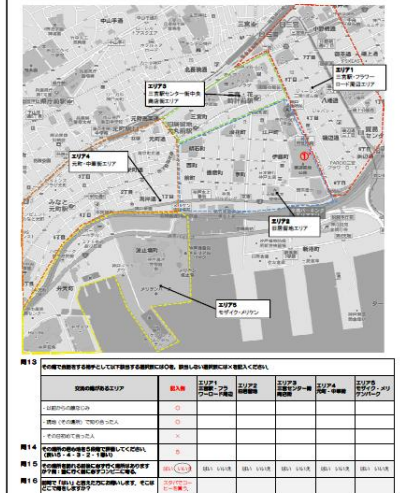
人の移動実態に関する詳細調査 追加アンケート

【調査の目的】 本調査は、都市部における人の移動実態を把握し、都市計画や交通政策の策定に資することを目的とする。調査対象は、都市部における人の移動実態を把握し、都市計画や交通政策の策定に資することを目的とする。

1. 日常生活における自発的トリップ (行きつけの交通の調査) の調査

※1 調査対象は、日常生活における自発的トリップ (行きつけの交通) の調査に限定する。業務上や娯楽上の移動は含まれない。また、調査対象は、日常生活における自発的トリップ (行きつけの交通) の調査に限定する。

調査対象の自発的トリップ	徒歩	自転車	バス	地下鉄	タクシー	その他
※2 目的地が自宅であるトリップ (通勤・通学を除く)	○	○	○	○	○	○
※3 目的地が自宅以外であるトリップ (通勤・通学を除く)	○	○	○	○	○	○
※4 目的地が自宅であるトリップ (通勤・通学を含む)	○	○	○	○	○	○
※5 目的地が自宅以外であるトリップ (通勤・通学を含む)	○	○	○	○	○	○
※6 目的地が自宅であるトリップ (通勤・通学を除く)	○	○	○	○	○	○
※7 目的地が自宅以外であるトリップ (通勤・通学を除く)	○	○	○	○	○	○
※8 目的地が自宅であるトリップ (通勤・通学を含む)	○	○	○	○	○	○
※9 目的地が自宅以外であるトリップ (通勤・通学を含む)	○	○	○	○	○	○
※10 目的地が自宅であるトリップ (通勤・通学を除く)	○	○	○	○	○	○
※11 目的地が自宅以外であるトリップ (通勤・通学を除く)	○	○	○	○	○	○
※12 目的地が自宅であるトリップ (通勤・通学を含む)	○	○	○	○	○	○
※13 目的地が自宅以外であるトリップ (通勤・通学を含む)	○	○	○	○	○	○



(神戸PP調査に付帯した追加アンケート)

プローブパーソンデータ

・プローブパーソン調査

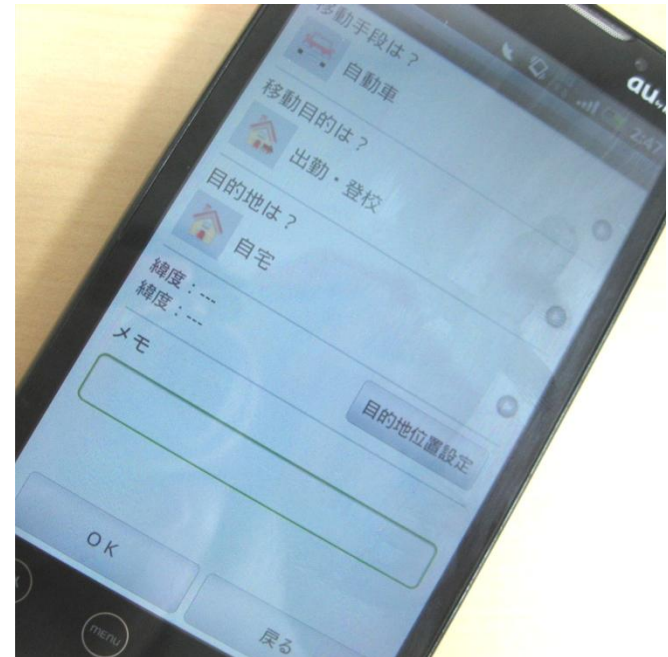
被験者に移動体をとりにつけて、移動状態を詳細に把握
マイクロレベルの移動軌跡、継続的データを収集可能

調査：数日～数十日間の行動

対象：個人（数人～数十人）

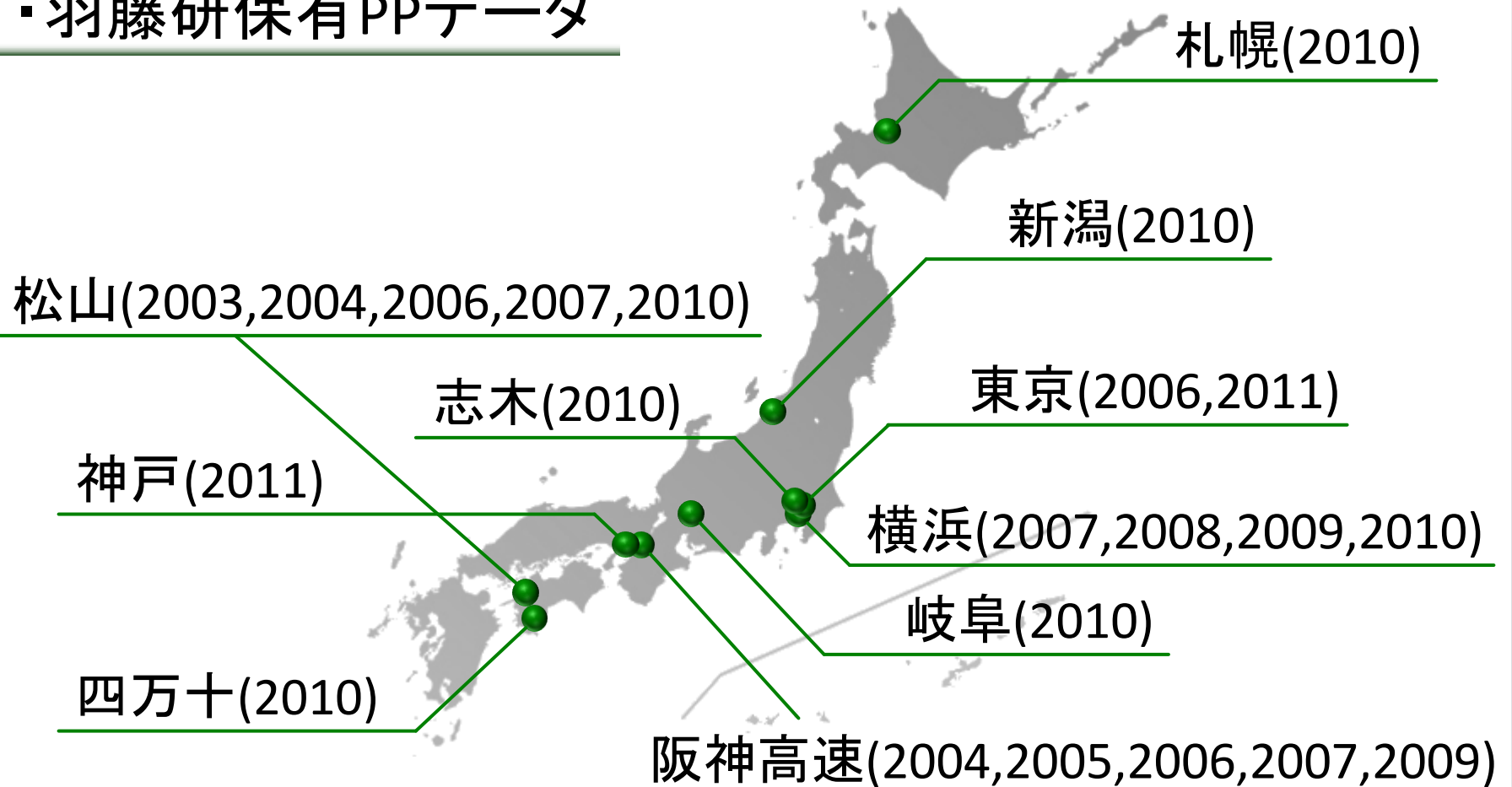
調査項目：

個人属性・目的・交通手段・
出発地・目的地・位置データ・
（加速度・気圧・磁気データ）



プローブパーソンデータ

・羽藤研保有PPデータ



プロジェクトと連動、調査ごとにテーマがある

プローブパーソンデータ

・ロケーションデータ

数秒ごとに位置データを記録

BCALs・スマートフォン端末の調査では
加速度データがある場合もある



例:

LocID	移動手段	モニターID	取得日時	緯度 (世界測地系)	経度 (世界測地系)	測位 モード	有効性
29393656	徒歩	a0102	2012/05/03 08:43:00	35.658610	139.745447	0	1
29393657	徒歩	a0102	2012/05/03 08:43:05	35.658628	139.745449	0	1
29393658	徒歩	a0102	2012/05/03 08:43:09	35.658623	139.745442	0	1
29393659	徒歩	a0102	2012/05/03 08:43:13	35.658625	139.745434	0	1
29393660	徒歩	a0102	2012/05/03 08:43:19			6	1
29393661	徒歩	a0102	2012/05/03 08:43:27	35.658633	139.745434	0	1
29393662	徒歩	a0102	2012/05/03 08:43:32	35.658639	139.745430	0	1
:	:	:	:	:	:	:	:

プローブパーソンデータ

・トリップデータ

トリップの出発・到着地の情報、出発・到着時刻や移動手段についての情報

→目的地選択、出発時刻選択、交通手段選択など

例:

TripID	モニターID	目的コード	目的	出発日時	到着日時	出発地	属性コード	出発地属性	出発地(世界)
0001	a0101	100	通学	2012/05/03 05:16:21	2012/05/03 06:49:50	自宅	110	自宅	35.658
0002	a0102	400	買い物	2012/05/03 12:22:05	2012/05/03 12:24:15	東大工14号館	120	勤務・通学先	35.658
0003	a0102	800	帰校	2012/05/03 12:27:09	2012/05/03 08:29:42	ファミマ東大正門前	210	コンビニ・スーパー	35.658
0004	a0102	600	娯楽	2012/05/03 16:43:13	2012/05/03 17:04:23	東大工14号館	120	勤務・通学先	35.658
0005	a0103	400	買い物	2012/05/03 10:22:14	2012/05/03 10:49:19	自宅	110	自宅	
0006	a0103	600	娯楽	2012/05/03 14:01:09	2012/05/03 14:14:27	ららぽーと	240	商業施設	35.658
0007	a0103	200	帰宅	2012/05/03 15:20:00	2012/05/03 15:43:32	図書館	330	公共施設	35.658
:	:			:					:

パーソントリップ調査

・調査の概要

調査票による調査、郵送・訪問により回収

調査: ある平日の1日の行動

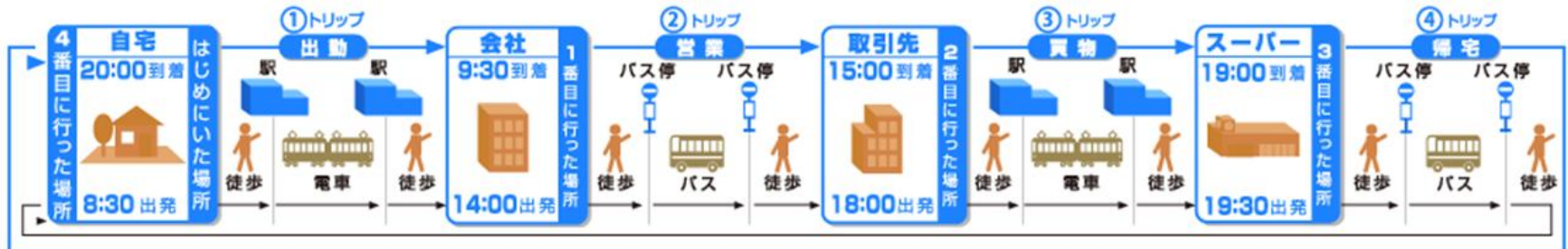
対象:

満5歳以上の対象地域居住者(数千～数万人)

調査項目:

個人属性・目的・交通手段・出発地・目的地など

集計はゾーン単位



パーソントリップ調査

・羽藤研保有PTデータ



おもな使うソフト

・R



統計処理に特化したインタープリター型言語

モデルの推定、グラフ描画、データ処理、ベクトル計算などに使う

```
RGui
ファイル 編集 閲覧 その他 パッケージ ウィンドウ ヘルプ
R Console
R version 2.11.1 (2010-05-31)
Copyright (C) 2010 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

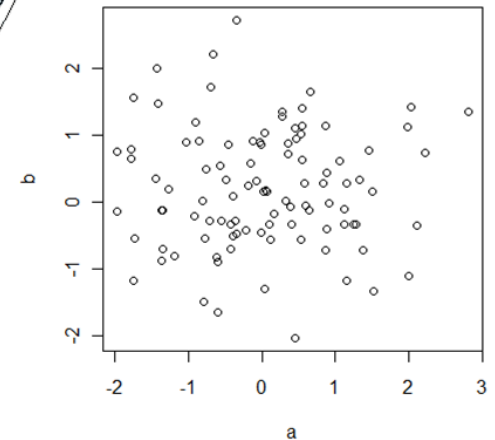
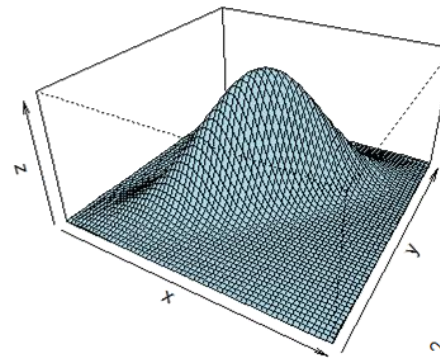
Rは、自由なソフトウェアであり、「完全に無保証」です。
一定の条件に従えば、自由にこれを再配布することができます。
配布条件の詳細に関しては、'license()'あるいは'licence()'を入力してください。

Rは多くの貢献者による共同プロジェクトです。
詳しくは'contributors()'を入力してください。
また、RやRのパッケージを出版物で引用する際の形式については
'citation()'を入力してください。

'demo()'を入力すればデモをみることができます。
'help()'とすればオンラインヘルプが出ます。
'help.start()'でHTMLブラウザによるヘルプがみられます。
'q()'を入力すればRを終了します。

[以前にセーブされたワークスペースを復帰します]

> ### Logit model estimation
> zm(list=ls())
> options(max.print=999999)
>
>
> Data <- read.csv("FIn1.csv",header=T)
以下にエラー file(file, "rt") : コネクションを開くことができません
```



おもな使うソフト

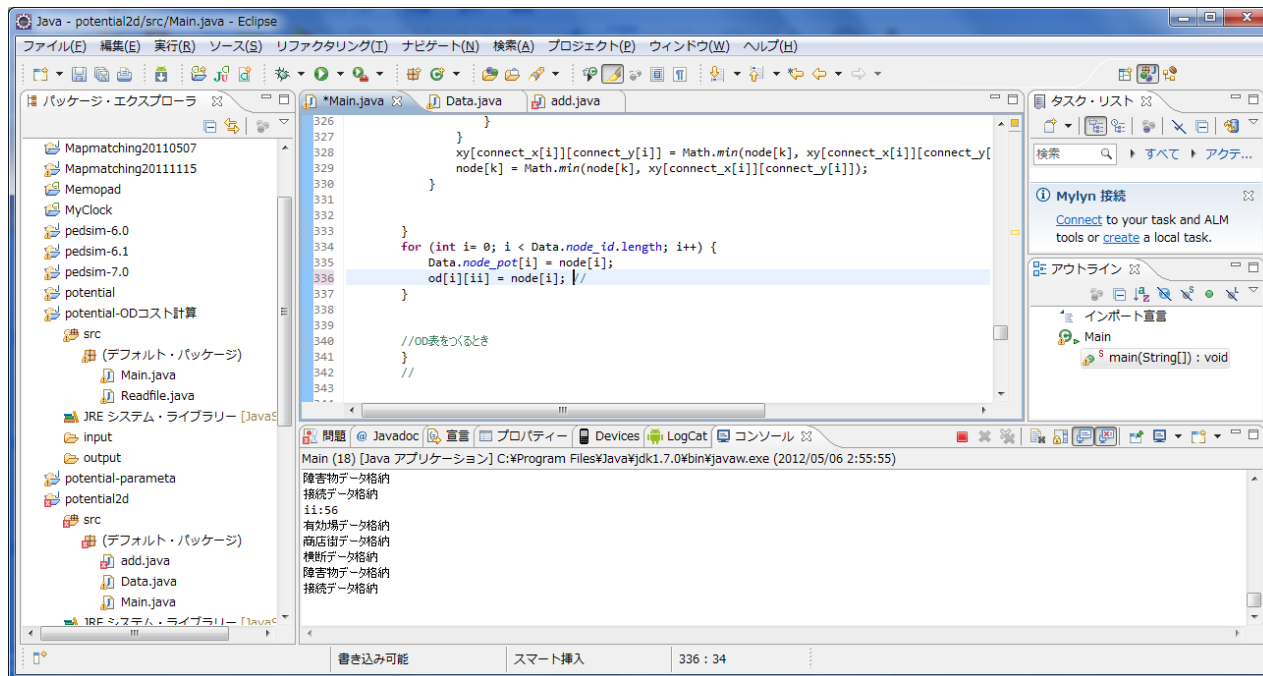
▪ Eclipse



Javaの統合開発環境

研究室保有プログラムはJavaが多い

(他のエディタやコンパイラでもいいのだけれど・・・)



おもな使うソフト

・プログラムで何ができるのか

データの操作(集計・トリップ分割など)

マップマッチング

シミュレーション(計算でいろいろな状況が発生させる)

データ形式の変換・加工(データの可視化など)

・プログラムの利点

エクセルで開けないような大容量データも扱える

コードがあれば、いじることで他のデータにも適用できる

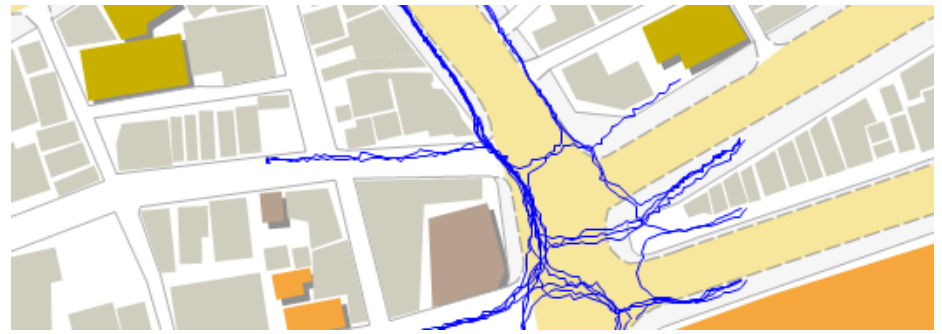
その他自分なりの工夫

おもな使うソフト

・プロアトラス

電子地図ソフト

Google マップなどの台頭により必要性は薄れているのかも・・・
移動軌跡、位置データの分布などをプロットできる



Excel ショートカットキー

・基礎編

Ctrl + A 全選択

Ctrl + C コピー

Ctrl + F 検索

Ctrl + H 置換

Ctrl + S 保存

Ctrl + V 貼り付け

Ctrl + ↓ データ末端まで移動

Ctrl + D 1つ上のセルからコピー

Ctrl + R 1つ左のセルからコピー

Ctrl + X 切り取り

Ctrl + Y 操作の繰り返し

Ctrl + Z 操作を元に戻す



Excel ショートカットキー

・応用編

Ctrl + Space

その列を選択

Shift + Space

その行を選択

F4

数式入力中に絶対/相対参照を切り替え

Alt + Shift + =

SUM関数を自動挿入

Ctrl + 1

書式設定ダイアログの表示

Ctrl + Shift + *

データが入力されている範囲を選択



Excel ピポットテーブル

・クロス集計

項目を掛け合わせて集計する方法
属性別の傾向などを分析したい

例)

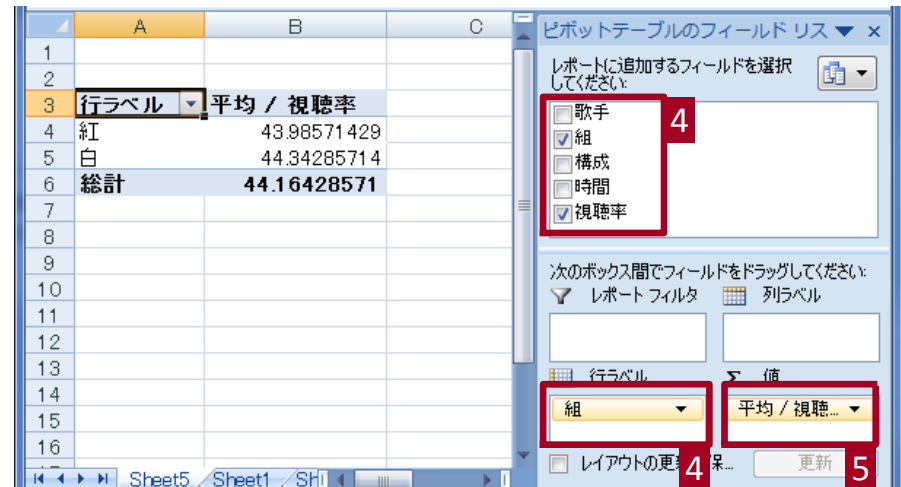
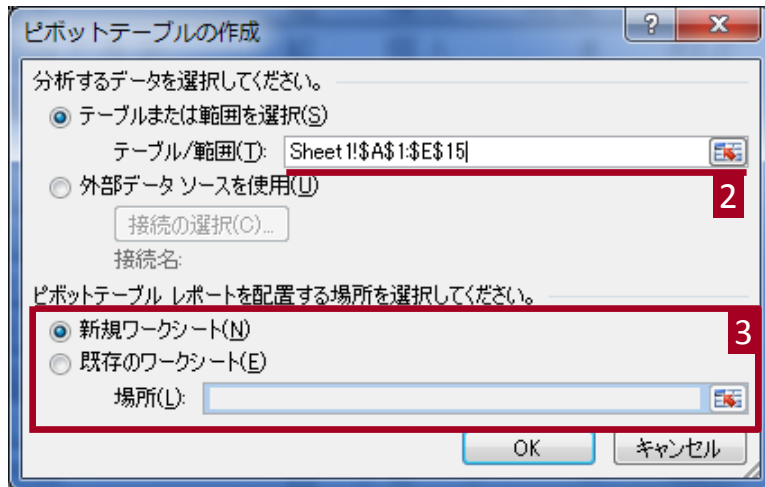
- ・どんな属性のアーティストが何組出場している？
- ・どういう属性のアーティストの視聴率が高いのか？
- ・演奏時間と視聴率の相関はありそうだろうか？

歌手	組	構成	時間	視聴率
和田アキ子	紅	個人	4	43.2
嵐	白	グループ	6	44.6
いきものがかり	紅	グループ	5	45.2
五木ひろし	白	個人	6	41.6
松田聖子	紅	個人	6	43.4
氷川きよし	白	個人	4	43.1
坂本冬実	紅	個人	6	42.1
福山雅治	白	個人	6	43.6
松任谷由実	紅	個人	6	44.2
EXILE	白	グループ	4	45.7
天童よしみ	紅	個人	5	45.1
北島三郎	白	個人	4	43.6
石川さゆり	紅	個人	5	44.7
SMAP	白	グループ	7	48.2

Excel ピボットテーブル

・ピボットテーブルの使い方

1. [挿入]-[ピボットテーブル]をクリック
2. [テーブル/範囲の選択]で集計したい範囲を選択 → [OK]
3. ピボットテーブルを配置する場所を指定
4. 行ラベル, 列ラベルを指定.
5. 集計したい列名・集計項目(合計、平均など)を指定



Excel ピボットテーブル

・クロス集計例

組・構成別歌手数

	A	B	C	D
1	データの個数 / 視聴率	列ラベル		
2	行ラベル	グループ	個人	総計
3	紅	1	6	7
4	白	3	4	7
5	総計	4	10	14

構成別平均演奏時間・視聴率

	A	B	C
1		値	
2	行ラベル	平均 / 時間	平均 / 視聴率
3	グループ	5.50	45.93
4	個人	5.20	43.46
5	総計	5.29	44.16

組・構成別平均視聴率

	A	B	C	D
1	平均 / 視聴率	列ラベル		
2	行ラベル	グループ	個人	総計
3	紅	45.20	43.78	43.99
4	白	46.17	42.98	44.34
5	総計	45.93	43.46	44.16

組・構成別視聴率標準偏差

	A	B	C	D	E
1	標準偏差 / 視聴率	列ラベル			
2	行ラベル	グループ	個人	総計	
3	紅	0.00	1.01	1.05	
4	白	1.51	0.82	1.96	
5	総計	1.37	1.02	1.59	

Excel VLOOKUP関数

・つかうとき

データの連結や抽出に使える
条件に一致するセルを検索して取り出す

Lunch

DataID
PersonID
名前
日付
昼食

Person

PersonID
名前
性別
実家通学
学年
通学手段

.....●

Lunchデータ

	A	B	C	D	E
1	DataID	PersonID	名前	日付	昼食
2	1	1	Eいぼよ	2012/5/3	手弁当
3	2	3	Oやま	2012/5/3	ファミマ
4	3	4	Yいぼし	2012/5/3	手弁当
5	4	1	Eいぼよ	2012/5/4	サブウェイ
6	5	2	Tぎー	2012/5/4	手弁当
7	6	4	Yいぼし	2012/5/4	100ローソン
8	7	4	Yいぼし	2012/5/5	手弁当
9	8	1	Eいぼよ	2012/5/5	ほっともっと
10	9	2	Tぎー	2012/5/5	手弁当
11	10	4	Yいぼし	2012/5/6	ファミマ
12	11	2	Tぎー	2012/5/6	手弁当
13	12	1	Eいぼよ	2012/5/6	ほっともっと
14	13	3	Oやま	2012/5/7	手弁当
15	14	4	Yいぼし	2012/5/7	ファミマ

Personデータ

	A	B	C	D	E	F
1	PersonID	名前	性別	実家通学	学年	通学手段
2	1	Eいぼよ	女	0	2	徒歩
3	2	Tぎー	男	0	2	自転車
4	3	Oやま	男	0	1	自転車
5	4	Yいぼし	男	1	1	鉄道

Excel VLOOKUP関数

・関数について

指定した列の中で、検索したセルと同じ行にある値を返す

= VLOOKUP (検索値 , 範囲 , 列番号 , 検索方法)

VLOOKUP `=VLOOKUP(B2,H2:M5,3,FALSE)`

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	DataID	PersonID	名前	性別	日付	昼食		PersonID	名前	性別	実家通学	学年	通学手段
2	1	1	Eいぼよ	女	2012/5/3	手弁当		1	Eいぼよ	女	0	2	徒歩
3	2	3	Oやま	男	2012/5/3	ファミマ		2	Tぎー	男	0	2	自転車
4	3	4	Yいし	男	2012/5/3	手弁当		3	Oやま	男	0	1	自転車
5	4	1	Eいぼよ	女	2012/5/4	サブウェイ		4	Yいし	男	1	1	鉄道
6	5	2	Tぎー	男	2012/5/4	手弁当							
7	6	4	Yいし	男	2012/5/4	100ローソン							
8	7	4	Yいし	男	2012/5/5	手弁当							
9	8	1	Eいぼよ	女	2012/5/5	ほっともっと							
10	9	2	Tぎー	男	2012/5/5	手弁当							
11	10	4	Yいし	男	2012/5/6	ファミマ							
12	11	2	Tぎー	男	2012/5/6	手弁当							
13	12	1	Eいぼよ	女	2012/5/6	ほっともっと							
14	13	3	Oやま	男	2012/5/7	手弁当							
15	14	4	Yいし	男	2012/5/7	ファミマ							
16													

3列目

※検索方法
TRUE: 一致またはそれ未満の最大値
(昇順並べ替えが必要)
FALSE: 完全一致

Excel データ分割

・区切り位置分割

カンマ・スペースなどで区切られたデータを扱いたいことがある例)

カンマ区切り・・・csvファイル等

3243, 2012/05/06 12:00:00, 2, 20, 本郷・・・

スペース区切り・・・datファイル等

0A 00 00 00 64 00 00 00・・・

タブ区切り

1.2534 1.4533 2.6788 4.2222・・・

	A	B	
1	1.25	1.50	1.75
2	2.20	2.95	2.78
3	3.45	1.23	4.32
4			

..... こうしたい →

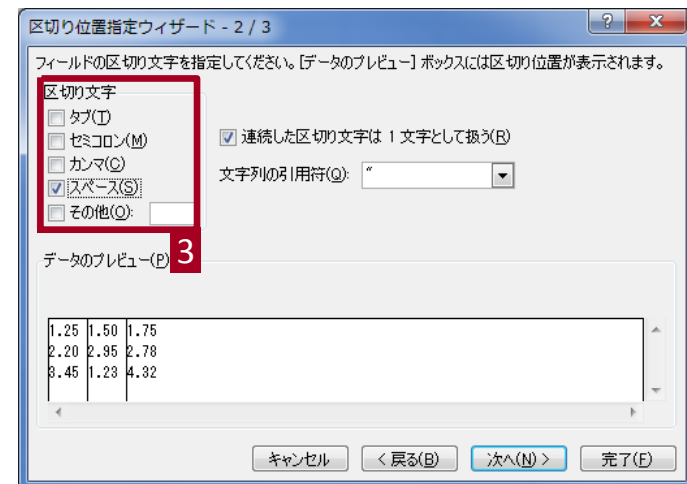
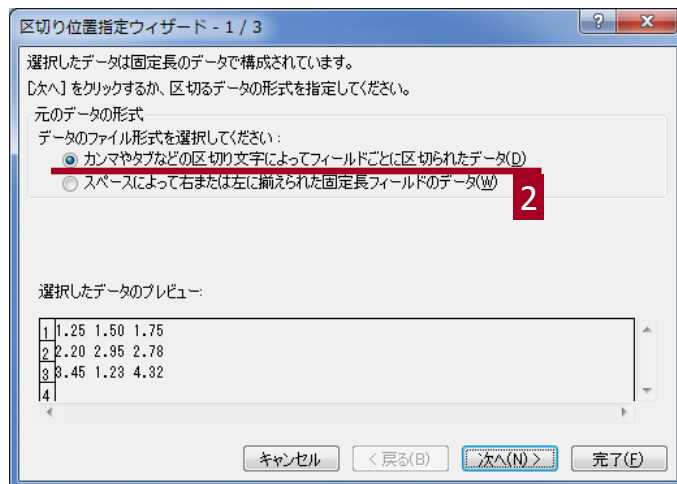
	A	B	C
1	1.25	1.5	1.75
2	2.2	2.95	2.78
3	3.45	1.23	4.32
4			

Excel データ分割

・区切り位置分割

カンマ・スペース・タブなどで区切られたデータを扱いたいとき

1. データ選択 → [データ]-[区切り位置]をクリック
2. [カンマやタブなどの区切り文字によって...]を選択 → [次へ]
3. 区切る文字を指定 → [次へ]
4. 表示形式・表示先を指定・確認 → [完了]



Excel グラフの作り方

・注意事項

【極めて重要】

●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

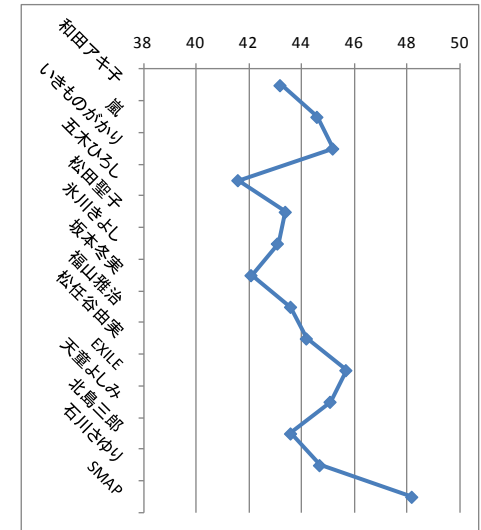
[グラフエリアの書式設定]

●グラフの張り付けは拡張メタファイル形式

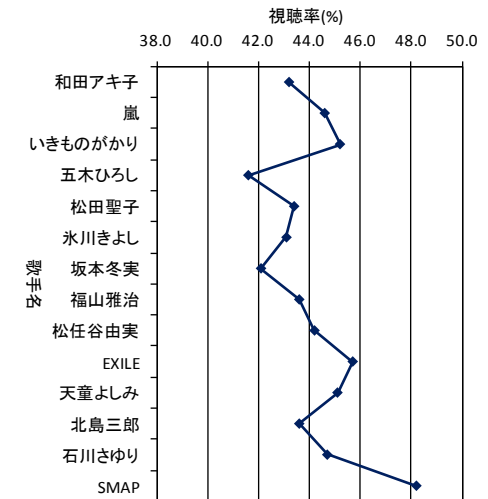
[貼り付け]→[形式を選択して貼り付け]

→[図(拡張メタファイル)]

NG



OK



Excel 時刻データの扱い方

・注意事項

【極めて重要】

時刻データを扱うとき、表示形式に気をつける

実際のデータ

2012/05/15 08:40:44

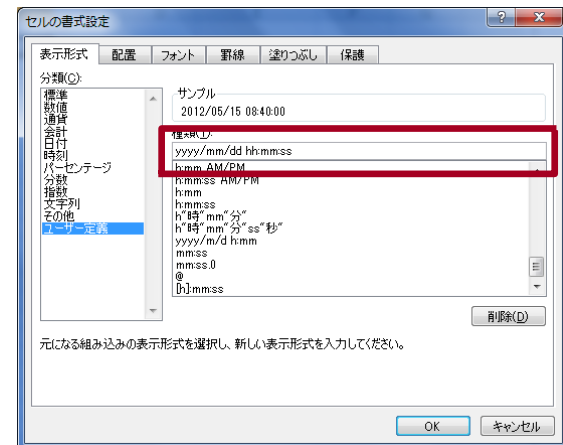
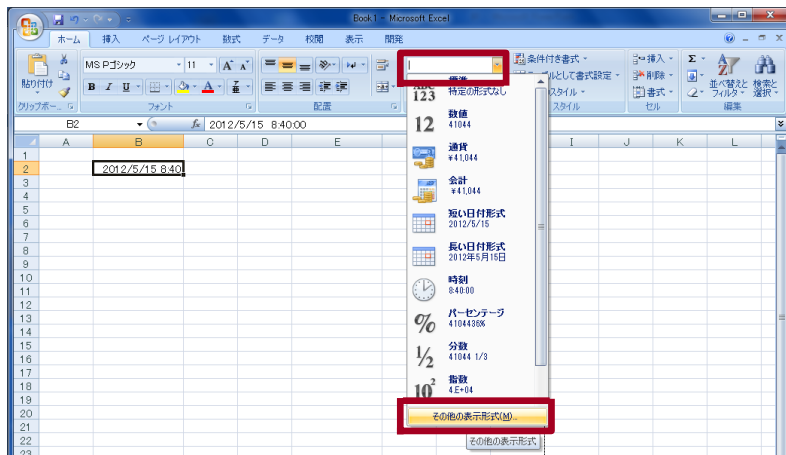


Excel

	A	B
1		
2		2012/5/15 8:40

このまま保存すると「秒」データが失われてしまう！！

[その他の表示形式]-[ユーザー定義]-"yyyy/mm/dd hh:mm:ss"



行動モデル基礎

・離散選択モデル

個人の意思決定を分析したい

個人の行動は何らかの「**選択**」である

離散的選択

<交通手段選択>

自動車 鉄道 自転車 徒歩

<歩行速度加減速選択>

加速 そのまま 減速

連続的選択

<施設滞在時間選択>

5分 27分 2時間

<歩行距離選択>

50m 650m 950m

行動モデル基礎

・確率的選択モデル

ある選択肢の効用は確率分布を持っている
意思決定は確率的に起こる

確率的選択

自動車の効用： 0.5

鉄道の効用： 1.0

自転車の効用： 0.3



自動車
29%

鉄道
48%

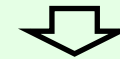
自転車
24%

確定的選択

自動車の効用： 0.5

鉄道の効用： 1.0

自転車の効用： 0.3



鉄道
100%

行動モデル基礎

・効用の定式化

合理的な個人は「**効用**」が最大になる選択肢を選択する

効用 U_{ni} : 個人 n が選択肢 i に感じている望ましさの程度

例: 交通手段二項選択(自動車 vs 鉄道)

$U_{train} > U_{car} \rightarrow$ 鉄道を選択



効用の説明変数は何か?? (仮説をたてる)

\rightarrow **所要時間**、**費用**、**性別**...

行動モデル基礎

・効用の定式化

個人 n の効用関数(多項式)

$$\begin{aligned} U_{train,n} &= \beta_1 x_{price,train,n} + \beta_2 x_{time,train,n} + \beta_3 x_{female,n} + \beta_4 + \varepsilon_{train} \\ U_{car,n} &= \beta_1 x_{price,car,n} + \beta_2 x_{time,car,n} + \varepsilon_{car} \end{aligned}$$

費用

所要時間

女性ダミー

定数項

誤差項

確定項 V

確定項: 観測可能な部分

誤差項: 以下の項目を表す

- (1) 確定項変数以外の要因
- (3) 説明変数の測定誤差

- (2) 線形和とした関数形の誤差
- (4) 属性の重みを個人間で均一とした誤差

行動モデル基礎

・ロジットモデルでの選択確率式

ランダム効用モデル (McFadden提案、2000年ノーベル経済学賞)

選択肢*i*を選ぶ選択確率

$$P_{train} = \frac{\exp(\mu V_{train})}{\exp(\mu V_{train}) + \exp(\mu V_{car})} = \frac{1}{1 + \exp(-\mu (V_{train} - V_{car}))}$$

μ : 誤差項のばらつきを表すスケールパラメータ

ロジットモデルでは誤差項分布に**ガンベル分布**を仮定

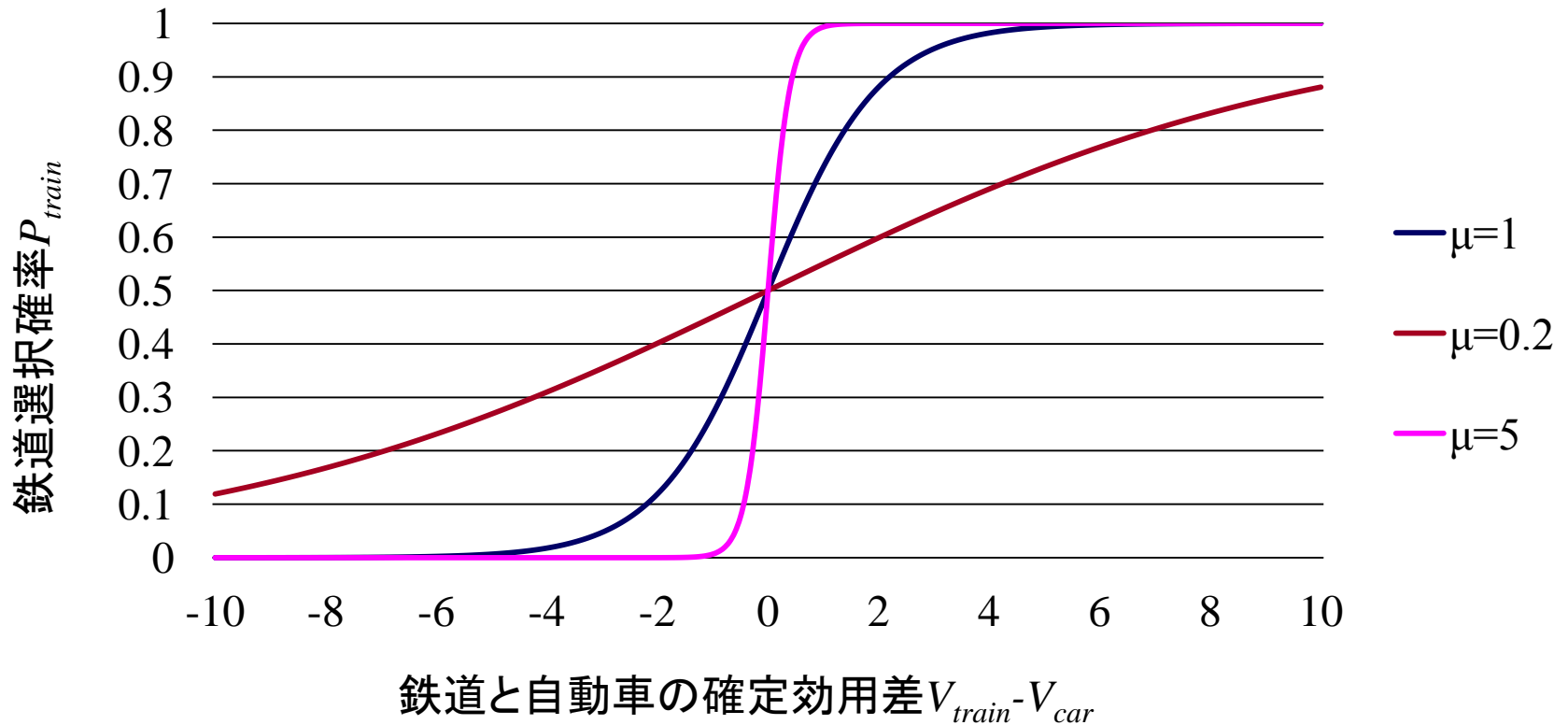
選択確率は、選択肢間の**確定効用の「差」**で表される

→全ての選択肢に同一ダミー変数や、定数項を入れることは意味がない(3肢以上の選択でも同様)

行動モデル基礎

・選択確率とスケールパラメータの関係

選択肢*i*を選ぶ選択確率の変化



次回までの課題

課題 1

以下のソフトをインストール・設定する

- ・R
- ・Tinn-R (R用エディタ)
- ・Eclipse (他の使えるJava環境があればそれでも良い)
- ・プロアトラス (研究室大部屋にインストールCD有)

課題 2

1. サンプルのPPTリップデータ・モニターデータを結合
2. サンプルデータを使った基礎分析

分析内容は何でもよい、次回までにパワーポイントでまとめる。
(グラフ作成・表など)、次回1~2分くらいで説明

次回までの課題

・課題データについて

Webページでダウンロードしてください

<http://bin.t.u-tokyo.ac.jp/startup2012/>

([研究室HP] - [講座] - [2012年] - [スタートアップゼミ])

The screenshot shows the homepage of the Behavior in Networks Studies Unit. The page title is "Behavior in Networks Studies Unit" with the subtitle "東京大学大学院工学系研究科 都市生活学・ネットワーク行動学研究室". A navigation menu includes "表紙", "紹介", "論文", "研究", "論文", "講座", "行事", "仲間", "リンク集", and "所在&連絡先". A "お知らせ / 更新履歴" section contains several tweets from the account "hatolabo". A world map highlights various international locations with labels: SIENA, STOCKHOLM, NANTES, PARIS, STRASBOURG, BARCELONA, BILBAO, MALTA, VENICE, BOLOGNA, NEPAL, MAKAO, BOSTON, HIRAYAMA, KANAZAWA, SAPPORO, MATSUE, SHIMANAMI, DOGO, MATSUYAMA, SHIBUYA, MIHARA, OHTSUKI, KASHIWA, YOKOHAMA, SHIMANTO, OWASE, and OSAKA. The "bin" logo and "Behavior in Networks Research Unit" are also visible.

The screenshot shows the "スタートアップゼミ2012" (Startup Seminar 2012) page. The main heading is "スタートアップゼミ2012" with a background image of a hand holding a whiteboard marker. Below the heading is a link to "研究室HPへ". A "概要" (Overview) section states the seminar's purpose and provides details for the first session: "第1回 データ仕様概説、クロス集計、行動モデル基礎(2012年5月15日(火) 08:40~10:20)". It includes a note about preparation and links to "詳細用データ1(トリップデータ)" and "詳細用データ2(モニターデータ)". A "予定表" (Schedule) table is provided at the bottom.

日程	担当	内容
5月15日(火)	伊藤	第1回 データ概説、クロス集計、行動モデル基礎(1)
5月18日(金)	穂村	第2回 行動モデル基礎(2)
5月22日(火)	瀬口	第3回 モデルの推定、Rの操作