

Excel × 亥年現象

データ整形と分析 事始め

このスライドの最新版とデータは
github.com/nishioWU/JNPC1906
にあります

データ分析の3ステップ

- ① データを入手する
- ② データを整形する
- ③ 図に描いてみる

①は今回はオープンデータで。③は前回のおさらいも含めて。

②が時間としては全体の7割ぐらい、というのが偽らざる実感。複数の表を付き合わせ処理したり、表記の揺れを直したり、結構大変。本日は②を中心に、「選挙」のお題でやってみたい。

亥年現象、選挙疲れは本当か

- データが取れるのは、前回の亥年＝2007年。その1回前の2004年の参院選と比べて、変動を見る
- 07年のほうが投票率は上がっている！
- とはいえ、亥年でなかったら、もっと上がったのかも
- 「選挙疲れ」説は本当なのか、統一地方選の有無で比較できないか
- 都道府県議選や知事選の有無で分けてみたが、いまひとつ
- では、どうするか……

全国の推移の折れ線グラフと、07年の散布図：[!\[\]\(c507f772dba2b921f86777f01218e570_img.jpg\)](#)

県議の「無投票当選」に注目

- 道府県議選の選挙区単位で見れば、無投票当選が結構ある
- 亥年の統一選で県議選があった特定の県に着目して、県議選が無投票だった市区町村と、投票が行われた自治体とで、その夏の参院選の投票率（3年前と比べた高低）に差があるか、は調べられる
- 2019年も、岐阜などで無投票当選が多かった。今夏の国政選の結果が出たときに、同じ手法を適用したら、何か見えてくるかも
- ただし、統計学的には問題あり。県議選が無投票になるかどうか、は国政選の投票率と無関係に決まっている（ランダムにコントロールされている）とは考えにくいため

サイトでデータを見つけたら

- ダウンロードできるものは、する
- サイトに載っているだけで、ダウンロードできないときは、コピー & ペーストで、Excelのシートに貼り付ける
- Chromeを使っているなら、「Copytables」を入れておくと便利。Altキーを押しながら、マウスで範囲指定すると、さほど困らない。簡易集計もできる
- 貼り付けるときは、Ctrl + Vではなく、Alt + Ctrl + Vで。「テキスト」を選ぶと、リンクも書式もない状態に変換して貼り付けることができる

実際にやってみる@大阪府

- ① Excelのファイルを新規作成。空っぽのシートを用意しておく
- ② ブラウザで2007年の大阪府議選のデータを開く。今回知りたいのは、投票があったのか無投票当選だったのか、なので「投票速報」をクリック。ダウンロード用のデータは用意されていないので、サイトの表をコピーして貼り付けるしかない
- ③ Excel のシートへのペーストは、Alt + Ctrl + Vで。「テキスト」を選ぶこと
- ④ 表が複数に分かれているときは、これを繰り返す
- ⑤ 表の範囲外でも必要な事項はコピー＆ペーストしておく

実際にやってみる@大阪府 2

- ⑥ Excelの別のシートに、07年と04年の府議選の投票率データも持ってくる。空っぽのシートを新規作成
- ⑦ ブラウザで07年の参院選大阪選挙区のデータを開く。「大阪府選出選挙」の「投票結果」をクリック。前ページの方法で、シートに貼り付けていく
- ⑧ ブラウザで04年の参院選大阪選挙区のデータを開く。やはりExcelのシートに貼り付ける
- ⑨ シート名は何でもよいが、ハンズオンの足並みを揃えるため「san2007」「san2004」「fugi2007」としておく

ここまで済んだ状態：[osaka-01.xlsx](#)

実際にやってみる@大阪府 3

- ⑩ 3種のデータを付き合わせるための、作業用シートを作りたい。
「form」フォルダの「form.xlsx」を開き、シートをコピーする。
シート名のタブをつかみ、Ctrl + ドラッグすると早い
- ⑪ この練習で必要なのは大阪府の分だけ。都道府県の列で「並び替えとフィルター」を使う。漏斗のマークを押し、下の窓で「大阪府」と「都道府県」だけに✓を入れると、絞り込まれる。左端の行番号が青くなるのが、フィルターが効いている目印
- ⑫ Ctrl + A で全範囲指定。次にAlt + ; で、いま見えている列しか対象にしないようにして、Ctrl + C でコピー
- ⑬ 新しいシートに、Ctrl + V で貼り付ける

実際にやってみる@大阪府 4

- ⑭ このシートを、先ほどの先ほどのExcelのブック（参院選と府議選のデータを貼り付けたもの）に、ドラッグする。シートがそのままコピーされた

※今回はこれでOKだが、別シートを参照する数式が入っていた場合は、ほかのブックに移すと、正しく動かない。その場合は、「値を貼り付け直す」作業をしてから、ブック間のドラッグをすること

Ctrl + A （全範囲選択）

Alt + Ctrl + V（オプションを指定する貼り付け）

V （数式を消し、値のみのテキストに）

ここまで済んだ状態：[osaka-02.xlsx](#)

ところで、整形に進む前に

- Excelの関数を使っている？
- VLOOKUP関数も使ったことがある？
- 絶対参照と相対参照を使い分けている？
- 表を使っている？
- 散布図を描いたことがある？
- ■MENTI■
- かけ算の表

Excelの勘どころ

▶行と列

横を行、縦を列とかカラムと呼んで区別している。

▶式は小文字で

Excelは、大文字でも小文字でも命令を聞いてくれる。なので、関数は小文字で入力するとよい。正しく認識されれば大文字に変換される。小文字のまま残ったら、打ち間違いだと分かる。

▶絶対参照

式をコピーすると、気を利かして、計算対象の行や列をずらしてくれる。それが便利だからだが、困る場合もある。そのときは、ずらされては困るものに「\$」マークをつけると、コピー先でもずれない。これが絶対参照。式の入力中に「F4」キーを押すと、行と列の両方またはどちらかに、\$がついたり消えたりして切り替えられる

能率が上がるショートカット

▶ シート名のタブをつかんで、Ctrl + ドラッグ

そのシートのコピーを作成

▶ ブック間で、シートをドラッグ

シート丸ごと移動できる。ただし、数式は消し値だけにしてから

▶ Shift + ドラッグ

行や列を選択し、その境目をつかみながらだと、並び替え

▶ Ctrl + 1

セルの書式設定。エルではなくて数字の一（テンキーの1はダメ）

▶ Ctrl + A

シート全体/表全体を選択（選択中のセルが表の外/内の場合）

▶ Ctrl + C

コピー

ショートカット 2

▶ Ctrl + V

通常の貼り付け。セル幅以外すべて引き継ぐ。もう一度押すと、貼り付けの形式を選ぶ

▶ Alt + Ctrl + V

形式を選択して貼り付け。関数を使って整形をした後、貼り直して「値だけ」に固定するのに便利（Vかテキストを選ぶ）

▶ Alt + ;

絞り込み時に表示されているセルだけをコピー元にする。重宝する

▶ Ctrl + S

ファイルを上書き保存

▶ 「F12」

ファイルを別名で保存

ショートカット 3

- ▶ Ctrl + Z

直前の変更を元に戻す

- ▶ 「F2」

セルの編集

- ▶ Ctrl + F

検索

- ▶ Ctrl + H

置換

- ▶ Ctrl + カーソルキー

空白セルは飛ばし、その次にデータの入っているセルにジャンプ

- ▶ Ctrl + Home

A1セル（左上）にジャンプ

ショートカット 4完

▶ Ctrl + End

データの入っている最終セルにジャンプ

▶ Ctrl + ;

きょうの日付を入力。便利

▶ Ctrl + :

現在の時刻を入力。便利

▶ Ctrl + *

データが入っている範囲を選択。離れ小島は選択されない

▶ Ctrl + T

テーブルにする

▶ Ctrl + Enter

複数のセルに同じデータを入れる。一括して修正するときに便利

▶ Alt + 下矢印

そのカラムに入力済みのデータのリストから選ぶ

自治体ごとに結合する

- 自治体名をキーにして、集めたデータを結合していく。せっかくの自治体コードは、今回は役に立たないようだ
- 並び順がまちまちなので、コピペではうまく行かない。VLOOKUPを使う。その他、出番のありそうな関数の一覧は緑色のページに
- キー列が一意でないとNG。同名の項目の有無は「条件付き書式」を使ってチェックできる
- 探しに行く先の「範囲」の左端がキー列でないと動かない

結合時の注意点

- 並び順が同じとは限らない
- 同名の自治体はないか
- 合併や合区、区割り変更、政令指定都市への移行もある
- 表記に揺れがある。「関ヶ原」か「関ヶ原」か、など
- 空白の有無でマッチングしないことがある。途中だけでなく、前後に入っていることも。後ろに入っていると、とくに気づきにくい
- 市部計とか、郡部計などダブって集計されている項目をどうするか

VLOOKUP(ア, イ, ウ, エ)

ア：ここに入っているのと同じものを

イ：この範囲の左端の列で探して、一致するものがあつたら

ウ：その右にこれだけ行ったところのデータを取ってくる

エ：FALSE

※アを探しに行く、イのキー列は範囲の左端でないとダメ

※イのキーは一意でないといけない。同名のものが複数回出てくるのはNG

※イは範囲に「名前をつける」のも一案

※エは「FALSE」で決め打ち。省略すると一致検索ではなくなる

SUBSTITUTE(ア, "イ", "ウ")

ア：このセルのうち

イ："ここに書いた文字（または文字列）"を

ウ："こちらの文字（または文字列）"に置き換える

※ウで、引用符の中に何も入れない（引用符を続けて2つ打つ）ようにすれば、イを見つけたら削除する、という動作になる

IF(ア, イ, ウ)

ア：この条件が成立しているか、真偽を調べて

イ：成立している（つまり「TRUE」のとき）なら、セルの中身をこれに

ウ：成立していない（つまり「FALSE」のとき）なら、これに

※アの部分には、AND関数やOR関数を使った、複合条件を入れることもできる

※ウは省略可能

これらの関数も便利

ASC	全角文字を半角に変換。反対はJIS
COUNTA	空白以外のセルの数をカウント。COUNTなら数値の入ったセルのみ
EXACT	2つの文字列が等しいかどうか判定
FIND	ある文字列が他の文字列の中にあるか検索。大文字と小文字を区別。ワイルドカードは使えない
JIS	半角文字を全角に変換。反対はASC
LEFT	文字列の左から、指定された字数を取り出す
LEN	文字列の字数が分かる
MID	文字列の途中の指定の位置から、指定の字数を取り出す
RIGHT	文字列の右から、指定された字数を取り出す
SEARCH	ある文字列が他の文字列の中にあるか検索。FINDと違い、大文字と小文字は区別せず。ワイルドカードが使える
TRIM	前後の余分な空白を取り除く。途中の空白は1つだけ残す

絶対参照と\$と「F4」

セルの式をコピー&ペーストやフラッシュフィルで増殖させると、気を利かして、計算対象の行や列をずらしてくれる。それが便利だからで、このデフォルトの動作を「相対参照」という。

困る場合もある。そのときは、ずらされては困るものに「\$」マークをつけると、コピー先でもずれなくなる。これが「絶対参照」。式の入力中に「F4」キーを押すと、行と列の両方またはどちらかに、\$がついたり消えたりして切り替えられる。

VLOOKUPの2番目の引き数である「探す範囲」は、絶対参照でないと、だんだんずれていってしまって困るはず。

または、「範囲に名前を付ける」という方法でもよい。名前をつけた場合は、ずれずに固定される

結合の手順 ハンズオンで

- ① VLOOKUP
- ② 絶対参照、INDIRECT、名前をつける
- ③ SUBSTITUTEなど
- ④ 縦横変換

CSVデータを読み込む

- 拡張子「.csv」はコンマ区切りのテキストのこと
- テキストエディタで中身を確認しておくとい
- 文字コード違いで化ける。UTF-8か、それともS-JIS（CP932）か
- 先頭の0を削るなど、Excelが気を利かせて勝手に変換するのが困りもの。日付も要注意。年がない場合は今年にされてしまう

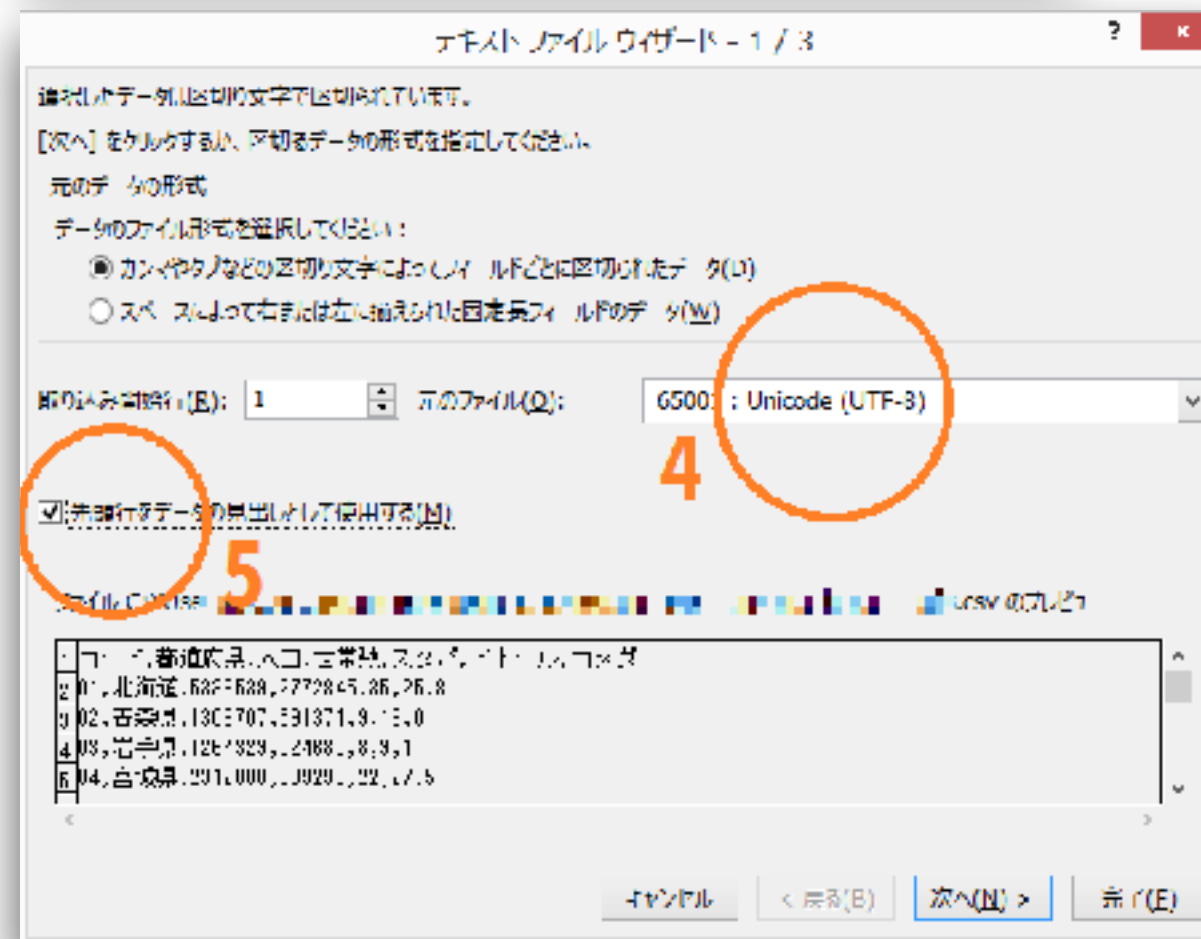
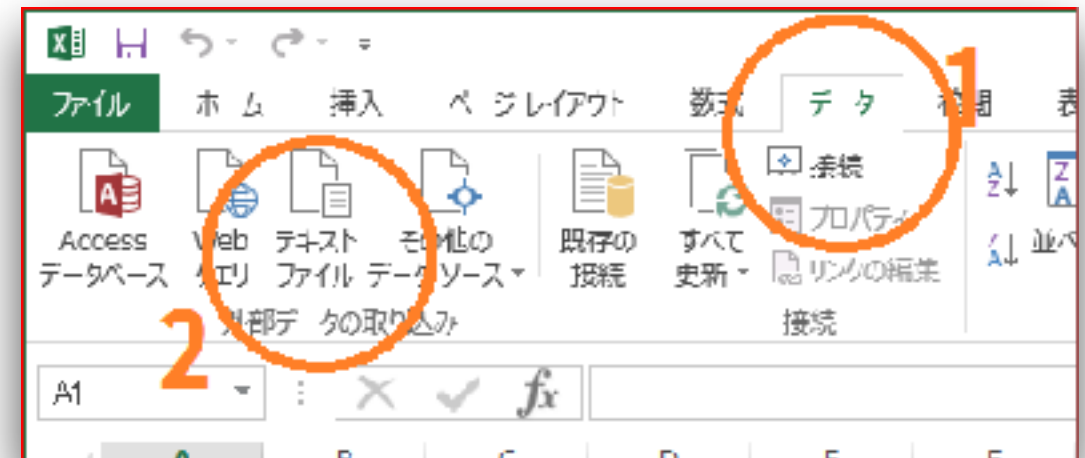
なので、CSVファイルのダブルクリックでExcelを開くのはNG！

Excelを先に起動しておき、そこにデータを取り込む

整形後のデータの保存形式としてはオススメ。受け渡ししやすい

CSVの読み込み手順 1

- ① Excel左上の「データ」タブで
- ② 「テキストファイル」を選択
- ③ CSVファイルを選ぶ。今回は、「data」「san_csv_UTF8」フォルダにある「san_UTF8.csv」。「開く」が「インポート」に変わるので、そのボタンをクリック
- ④ 「Unicode (UTF-8)」を選ぶ
- ⑤ 取り込み開始行を2に
- ⑥ 「先頭行をデータの見出しとして使用」に✓



CSVの読み込み手順 2

- ⑧ 「区切り文字」を変更。「カンマ」に✓を
- ⑨ プレビュー画面を確認して「次へ」
- ⑩ 列のデータ形式を指定。「コード」の列を選択してから、上のラジオボタン●で「文字列」にする。0で始まる場合があるため。「標準」だと数値に変換され、0が消えてしまう
- ⑪ 年の入っていない日付も、勝手に今年にされてしまう。もしあれば、やはり「文字列」が安全
- ⑫ 今回は「コード」列以外はデフォルトで、「完了」
- ⑬ データの貼り付け先を適宜指定し、「OK」

CSVの読み込み手順 3

テキスト ファイル ウィザード - 2 / 3

フィールドの区切り文字を指定してください。[データのプレビュー] ボックスには区切り位置が表示されます。

区切り文字

- ☐ タブ(T)
- ☐ セミicolon(M)
- ☒ カンマ(C)
- ☐ スペース(S)
- ☐ その他(O):

連続した区切り文字は 1

文字列の引用符(Q): "

データのプレビュー(P)

コード	都道府県	人口	世帯数	スタバ	ドトール
01	北海道	5339539	2772845	35	25
02	青森県	1308707	591371	9	13
03	岩手県	1264329	524685	8	9
04	宮城県	2312080	989296	22	27

テキスト ファイル ウィザード - 3 / 3

区切ったあとの列のデータ形式を選択してください。

列のデータ形式

- ☐ G/標準(G)
- ☒ 文字列(I)
- ☐ 日付(D): YMD
- ☐ 削除する(I)

[G/標準] を選択すると、数字は数値に、日付は日付形式の値に、その他の値は文字列に変換されます。

詳細(A)...

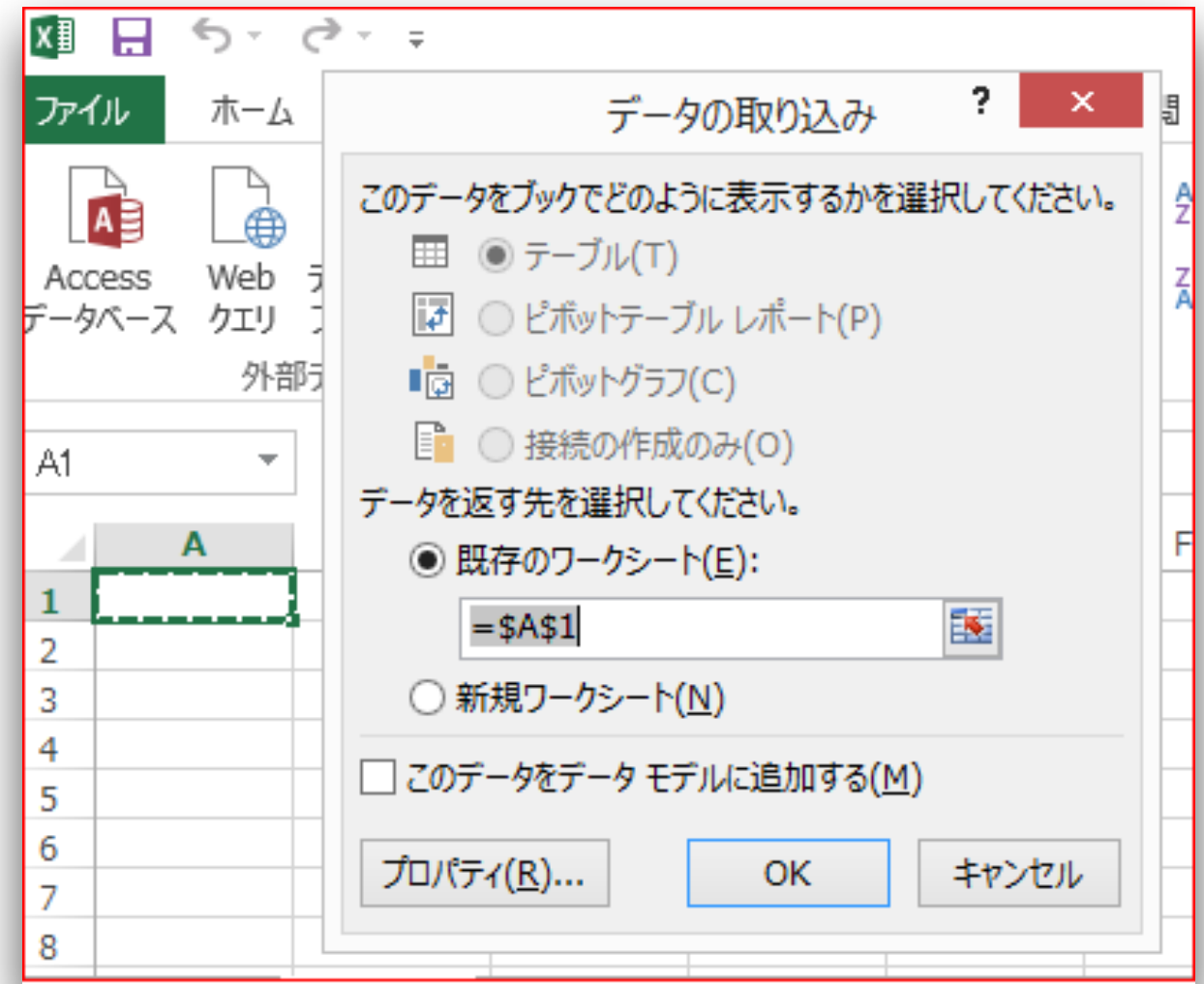
データのプレビュー(P)

文字列	G/標準	G/標準	G/標準	G/標準	G/標準	G/標準
コード	都道府県	人口	世帯数	スタバ	ドトール	コメント
01	北海道	5339539	2772845	35	25	8
02	青森県	1308707	591371	9	13	0
03	岩手県	1264329	524685	8	9	1
04	宮城県	2312080	989296	22	27	5

キャンセル < 戻る(B) 次へ(N) > 完了(E)

CSVの読み込み完了

- OKを押せば、シートの左上隅を起点に貼りつく
- 左上隅以外や、新しくシートを増やして貼る場合には、その旨指定を



ここまで済んだ状態：[osaka-02.xlsx](#)

文字化けをIEで直す方法

CSVかTXT形式のデータが文字化けしているときは、エンコードの違いが原因。UTF-8かSHIFT-JISを試してみる。実はInternet Exploreでコードを変換して保存し直す手がある。知られていないが、覚えておくと役に立つ

- ① 拡張子が「.csv」の場合は「.txt」に変えて保存。ピリオドまで消さないように注意
- ② IEでファイルを開く。化けていれば、画面を右クリックしてエンコードを直す。たいてい、自動認識してくれる
- ③ 保存したい形式（上記2通りのどちらか）を選び、別名で保存。別名にしないと、元が消えてしまう
- ④ 拡張子を「.csv」に戻す

図を描く前に表にする

- ① Excelのシートを複写しておく。失敗した場合のリカバリー用。
簡単なやり方は、Ctrl + ドラッグ
- ② データに通し番号を打っておく。ソートで順番を変えても、元に戻せるようにしておく
- ③ 表にすると便利。Ctrl + A または Ctrl + * で範囲指定してCtrl + Tが早い。左上隅をクリックしても全範囲指定できる。これで、絞り込みやソートが可能になった
- ④ この表で絞り込んだ状態が、グラフにも反映されるので便利

グループ分けした散布図

- ① 都道府県議選だけがあった、知事選だけがあった、両方があった、の3通りにグループ分けして、散布図にする。「層別散布図」という
- ② 横軸を04年の投票率、縦軸を07年の投票率とする
- ③ 07年のほうを、グループごとに新しく作る計3列に分ける
- ④ IF関数を使って、その列の条件に当てはまるものは07年の数値を入れ、当てはまらなければ空欄にする。「=if(and(○○=TRUE, △△=FALSE), □□, "")」という書式で、考えてみて
- ⑤ 04年と、3列に分けた07年の計4列を選択して、散布図にする

散布図を描く お任せの方法

- ① 表にある列を最低2つ選ぶ。
2列目が離れているなら、1列目を選んだ後、Ctrl + クリックで追加
- ② 「挿入」タブから「グラフ」の「散布図」を選ぶ
- ③ 3列以上選んだ場合は、左端の列がX座標、残りはY座標になる

散布図 手動で描くなら

- ① 今あるグラフを手直しするか、列を選ばずに空っぽのままの散布図を挿入。グラフエリア右の漏斗のアイコンから、「データの選択」に進み、左側の「凡例項目」の窓で指示する
- ② 「追加」で開く「系列の編集」パネルで、上から順に項目名（じか打ちしても、入っているセルを指定しても可）、X座標のデータ範囲、Y座標のデータ範囲、を指定。XとYは先頭のイコールを残しておく。「OK」で完了
- ③ やり直す場合は「編集」。不要な列があれば「削除」

参考) 散布図をひとひねり

- ① グループごとに大まかな分布を見る「箱ひげ図」の代用めいた図にもできる。横軸は地方選の有無、縦軸は07年と04年の投票率の差にする。
- ② X軸は新たに1列作る。IF関数を使って、「県議選あり」「知事選あり」「両方あり」のいずれかが入るようにする
- ③ Y軸も新たに1列作る。07年の投票率から04年の投票率を引いたものを入れる
- ④ 04年と、07年の3列の計4列を選択して、散布図にする
- ⑤ 点の重なり（オーバープロット）が発生するので、透明度を上げて薄くしておく

参考になるサイト

- ・「エクセルによる層別散布図の作り方」

<https://hitorimarketing.net/tools/stratified-scatterplot.html>

- ・「エクセル 散布図グラフの作り方」

<https://www.tipsfound.com/excel/05036>

- ・「統計グラフ Excelで簡単に箱ひげ図を作る方法」

<https://did2memo.net/2017/02/12/excel-easy-box-and-whisker-plot/>

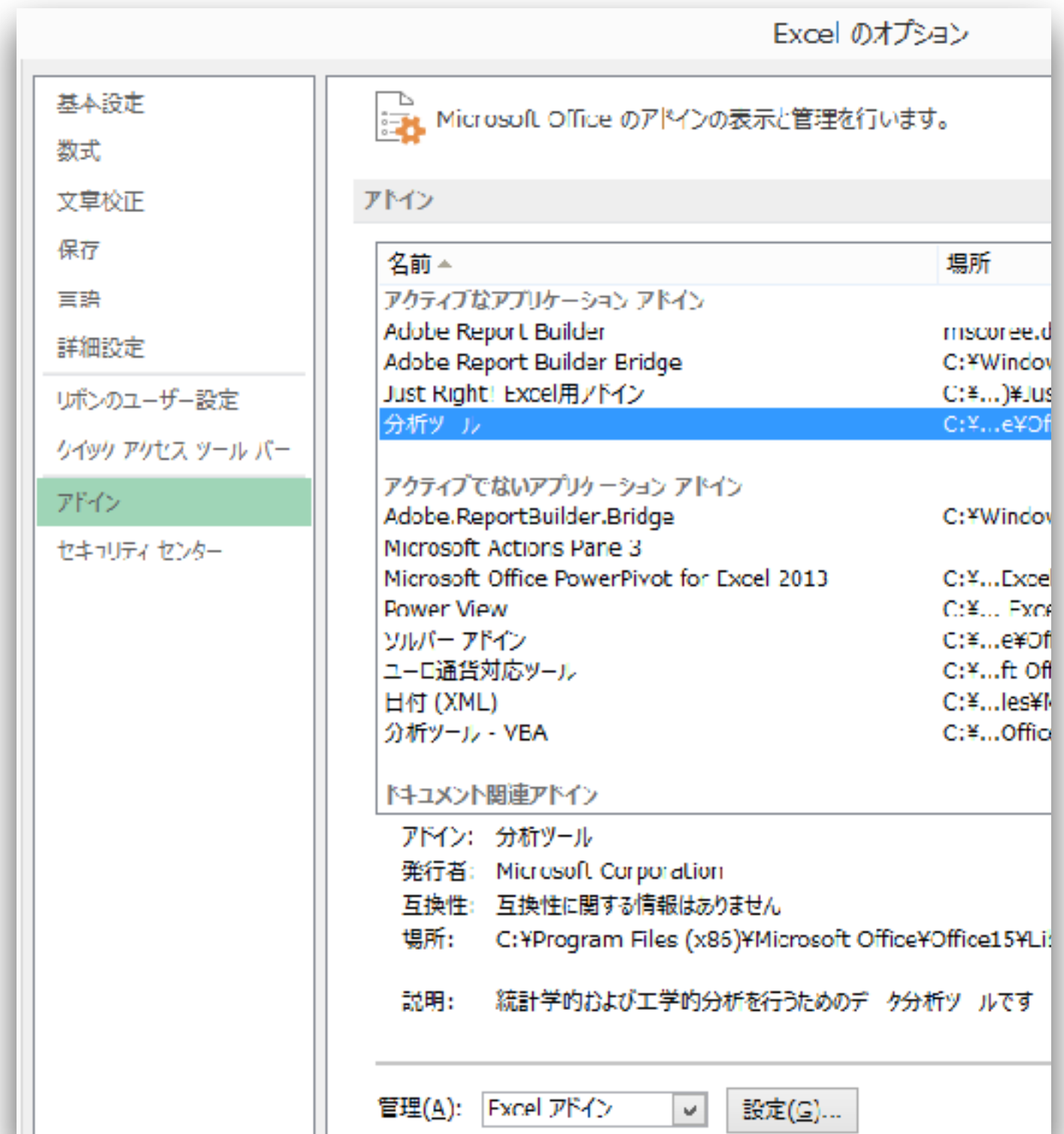
2016以降では、箱ひげ図がデフォルトで用意されている

参考) 分析ツールのアドイン

デフォルトではツールが使えないので、使えるようにする（一度やればOK）。

左上の「ファイル」タブから「オプション」→「アドイン」→「Excel アドイン」と進み、「分析ツール」を有効にするよう設定

2群に分けて差を見るには、「t 検定：分散が等しくないと仮定した2標本の検定」が使える



お疲れさまでした！

受講ありがとうございます

- ・ 質問はみんなの宝物です。ご遠慮なく質問を。分からないのは、説明が行き届かないから。たぶん、ほかの方も困っています
- ・ 内容改善につなげるために、ぜひフィードバックをお願いします
- ・ そのほか、お問い合わせは
yoshito.nishio+JNPC@gmail.com へ

ご協力いただいたアドバイザーのみなさま、いつも助けていただいている事務局のみなさま、そして受講いただいたみなさまに深謝します。