

# データ解析 第1回, オリエンテーション

北九州市立大学経済学部

齋藤 朗宏

# “データ解析”について

- 統計学をベースとし，実データを分析するための技術習得を目的とする.
- 分析するための技術を習得することは，メディア等で報じられる調査結果をどう読むかにも繋がる.
- 何故世論調査は3000件でいいのか？
- 何故天気や気温から売り上げが予測できるのか？

# 統計学

- 統計学：科学的推論のための道具。  
観察を重ねることによる法則性の発見。
- メンデルの遺伝の法則
  - ◆ <http://www.tmd.ac.jp/artsci/biol/textbook/genetics.htm>
- ゴールトンの親子の身長分析
  - ◆ <http://www.keio-up.co.jp/kup/sp/statistics/data/data107.xls>

# 2009年セ・リーグ順位表

	勝率	打率	得点	本塁打	盗塁	失点	被本塁打
巨人	0.659	0.275	650	182	84	493	123
中日	0.566	0.258	605	136	81	508	102
ヤクルト	0.497	0.259	548	116	106	606	140
阪神	0.479	0.255	548	106	79	534	102
広島	0.464	0.245	528	101	77	575	117
横浜	0.354	0.239	497	128	51	685	164

プロ野球では、勝率が最も高いチームが優勝となる。つまり、最終目標は、高い勝率を獲得する事である。そのためには、どんな要因が必要となるだろうか？

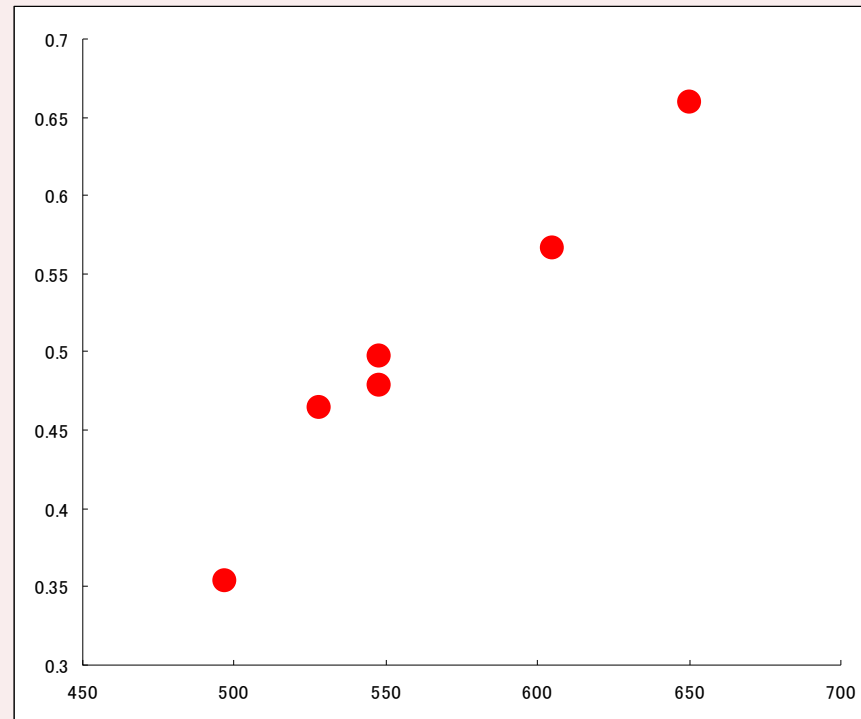
<http://bis.npb.or.jp/2009/stats/> より作成.

# 2009年セ・リーグ順位表

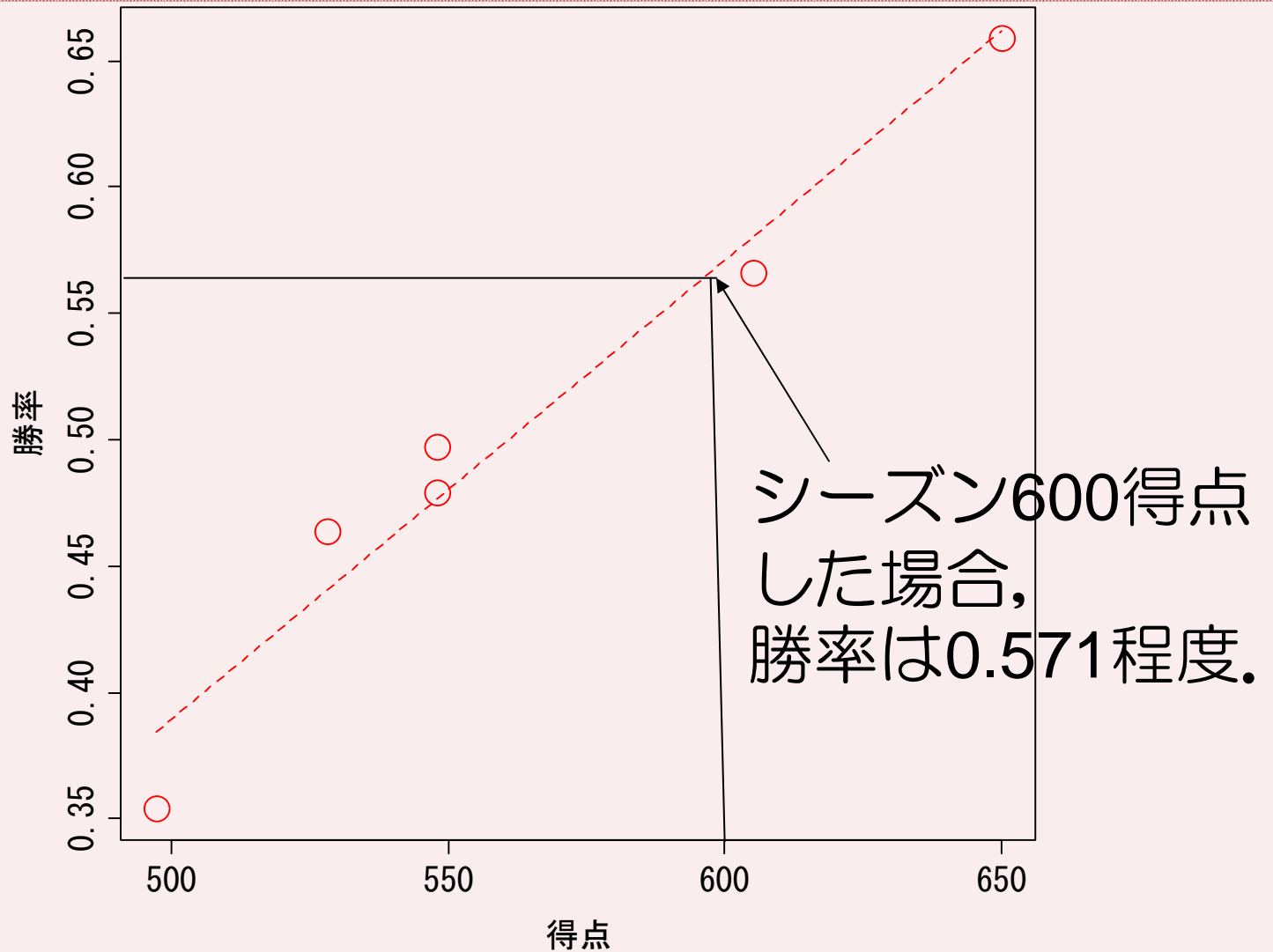
	勝率	打率	得点	本塁打	盗塁	失点	被本塁打
巨人	0.659	0.275	650	182	84	493	123
中日	0.566	0.258	605	136	81	508	102
ヤクルト	0.497	0.259	548	116	106	606	140
阪神	0.479	0.255	548	106	79	534	102
広島	0.464	0.245	528	101	77	575	117
横浜	0.354	0.239	497	128	51	685	164

- 優勝の巨人は、打率・得点・本塁打が最多、失点が最少だが被本塁打は少なくない。
- 最下位の横浜は、本塁打以外のすべての指標で最悪。
- ヤクルトは盗塁数がずば抜けて多いが、順位は3位。
- 阪神は被本塁打で中日と並んで最少だが、順位は4位。

# 得点と勝率の散布図



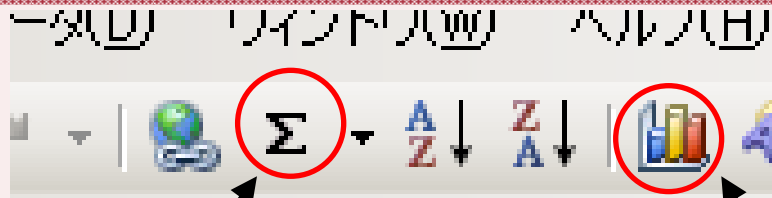
# 勝率の予測



# 基本テーマ

- 単変量データに関する現象の記述
  - 多変量データに関する現象の記述
  - グラフによる可視化
  - 確率
  - 平均値の比較
  - 予測，説明モデル
- 
- データを扱うものに関しては，すべてExcelを利用する予定.

# Microsoft Excel



計算

グラフ作成

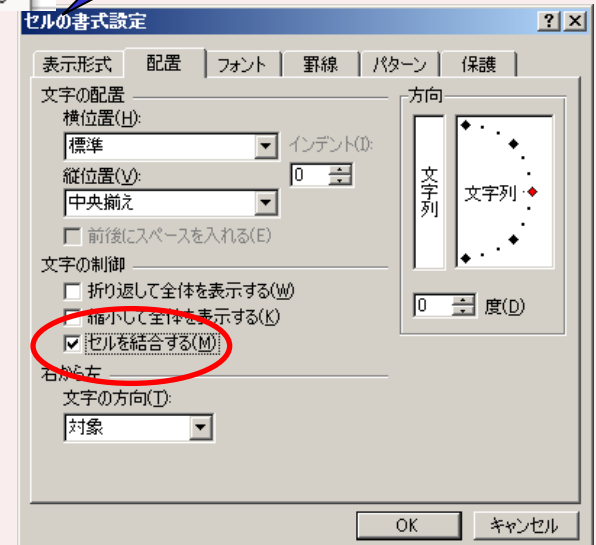
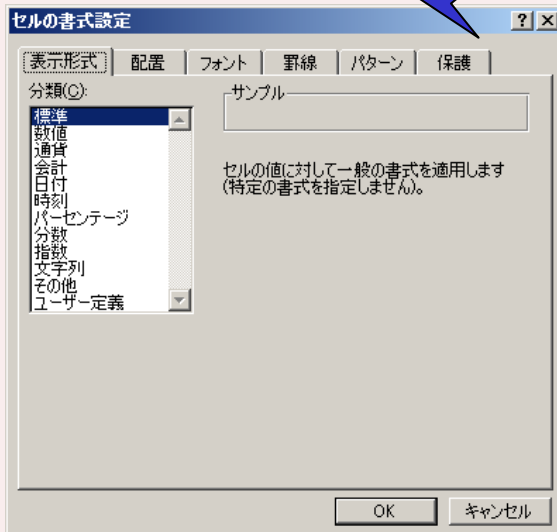
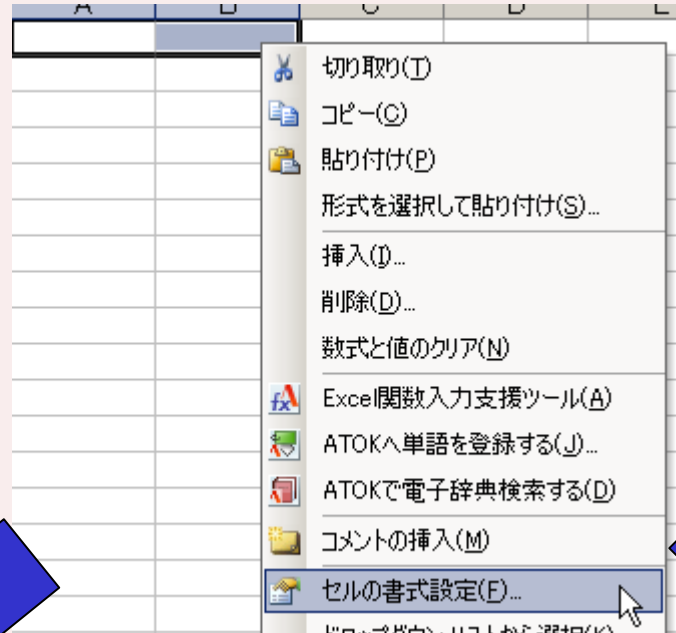
列→

セル

行↓

	A	B	C	D
1				
2				
3				
4				
5				
6				
7				

# セルの書式設定



# オートフィル

p.22

	A	B
1	1	
2	2	
3		
4		
5		

数字を2つ入力後、  
セルの右下をドラッグ

	A	B
1	1	
2		
3		
4		
5		

数字を1つ入力後、  
Ctrlキーを押しながら  
セルの右下をドラッグ

# オートフィル

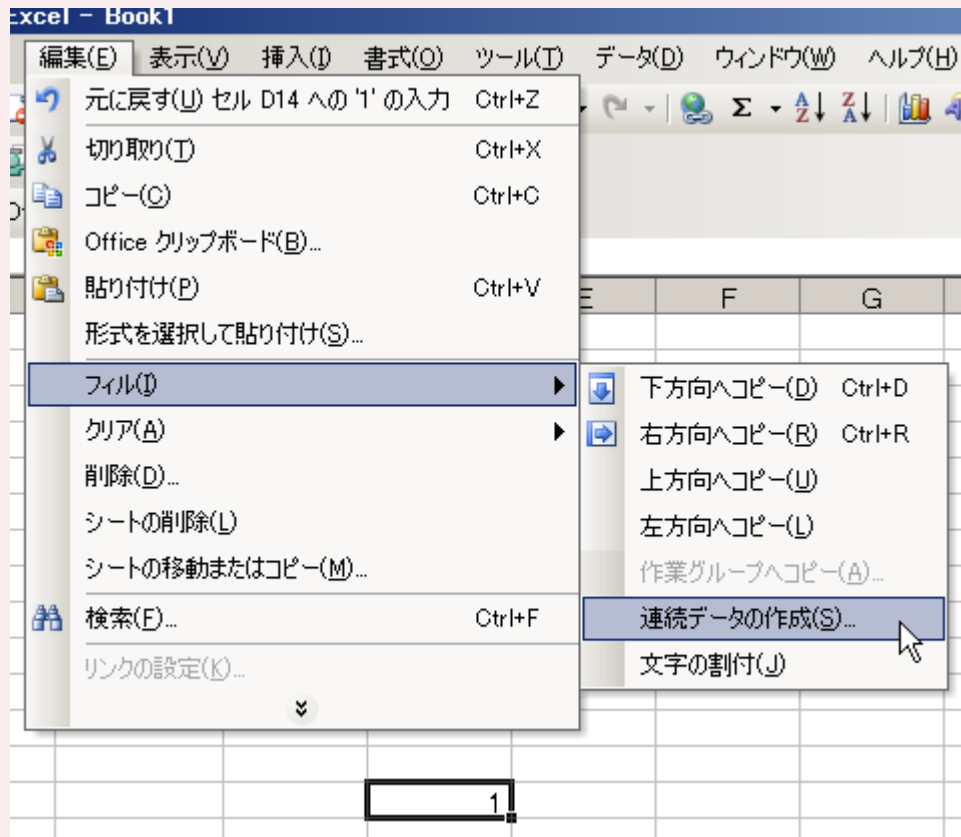
	A	B
1	月	
2		
3		
4		
5		
6	+	金

曜日も可能

	A	B	
1	月	2009/4/27	
2	火	2009/4/28	
3	水	2009/4/29	
4	木	2009/4/30	
5	金	2009/5/1	
6			+
7			土

日付, また, 二つのセル  
同時にも可能

# オートフィル



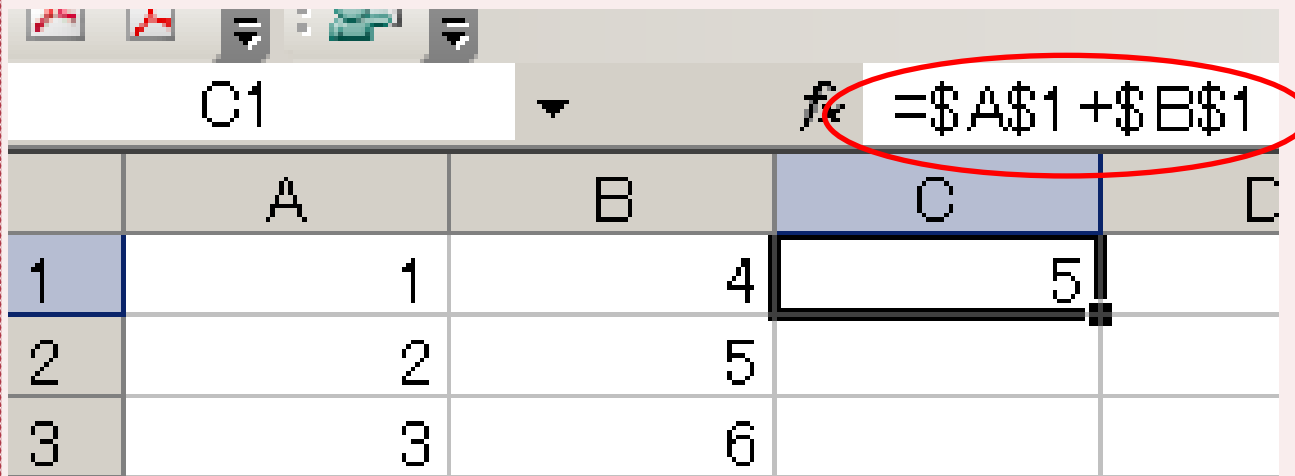
# 相対参照

	A	B	C
1	1	4	=A1+B1
2	2	5	
3	3	6	
4			

	A	B	C	D
1	1	4	5	
2	2	5		
3	3	6		
4				

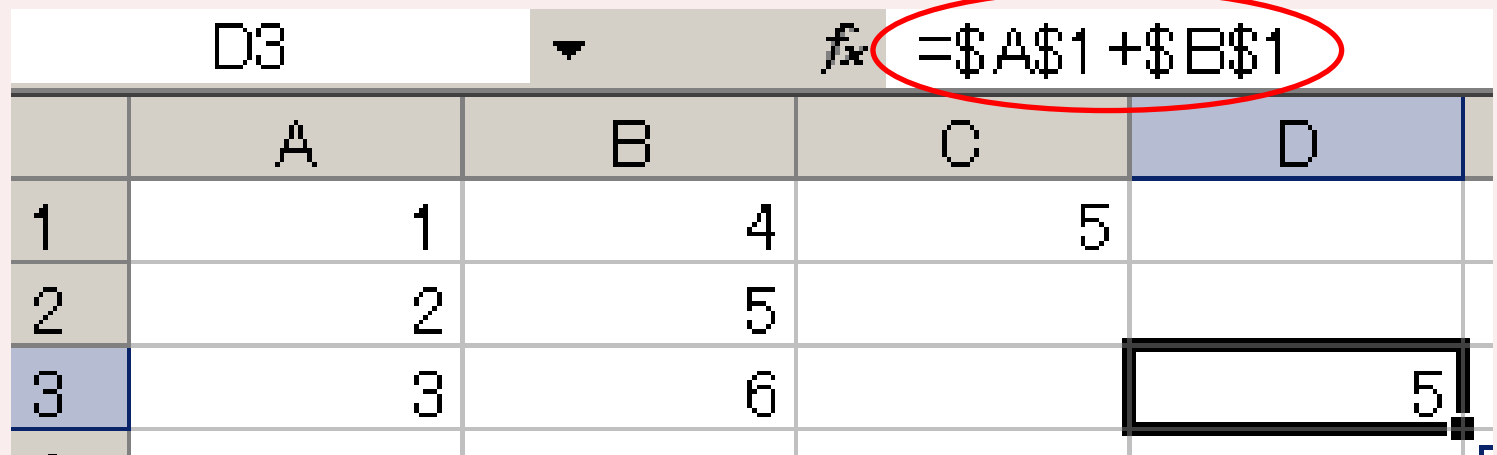
➤ 1列右, 2行下にずれているので, A1はB3.

# 絶対参照



A screenshot of an Excel spreadsheet. The formula bar at the top shows the formula  $=\$A\$1 + \$B\$1$  in cell C1, which is circled in red. The spreadsheet has columns A, B, and C. Row 1 contains values 1, 4, and 5. Row 2 contains values 2, 5, and an empty cell. Row 3 contains values 3, 6, and an empty cell.

	A	B	C
1	1	4	5
2	2	5	
3	3	6	



A screenshot of an Excel spreadsheet showing the same formula  $=\$A\$1 + \$B\$1$  copied to cell D3, also circled in red. The spreadsheet has columns A, B, C, and D. Row 1 contains values 1, 4, 5, and an empty cell. Row 2 contains values 2, 5, and an empty cell. Row 3 contains values 3, 6, and 5.

	A	B	C	D
1	1	4	5	
2	2	5		
3	3	6		5

➤ 別のセルにコピーしても、参照が維持される。

# 複合参照（絶対行参照）

C1		fx =A\$1 +B\$1		
	A	B	C	
1	1	4	5	
2	2	5		
3	3	6		

D3		fx =B\$1 +C\$1			
	A	B	C	D	
1	1	4	5		
2	2	5			
3	3	6		9	

- 1行目は固定されるが、A列という情報は固定されない。

# 複合参照（絶対列参照）

	C1			$=\$A1+\$B1$
	A	B	C	
1	1	4	5	
2	2	5		
3	3	6		

	D3				$=\$A3+\$B3$
	A	B	C	D	
1	1	4	5		
2	2	5			
3	3	6			9

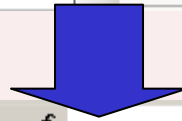
- A列，B列という部分は固定されるが，行番号は固定されない．F4キーで切り替え可能．

# (通常の) 貼り付け

C1      fx =A1+B1

	A	B	C	D	E
1		4	5		
2		5			
3		6			

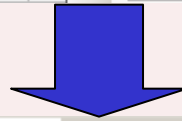
Context menu options:  
切り取り(T)  
コピー(C) (highlighted)  
貼り付け(P)  
形式を選択して貼り付け(S)...



C2      fx

	A	B	C	D	E
1	1	4	5		
2	2	5			
3	3	6			
4					
5					
6					
7					

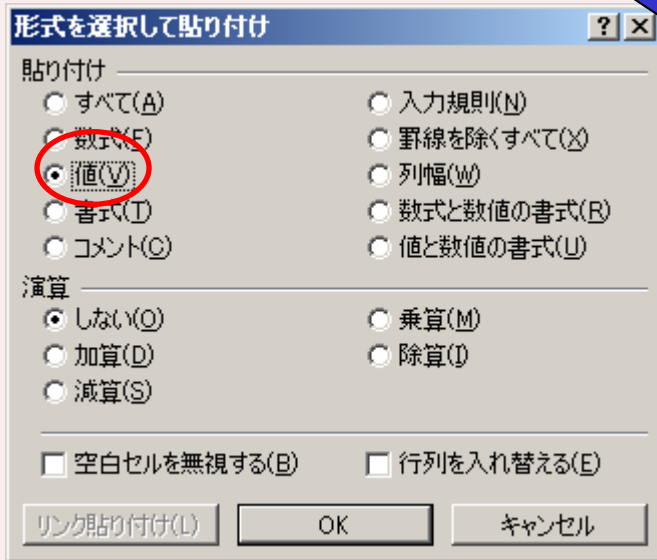
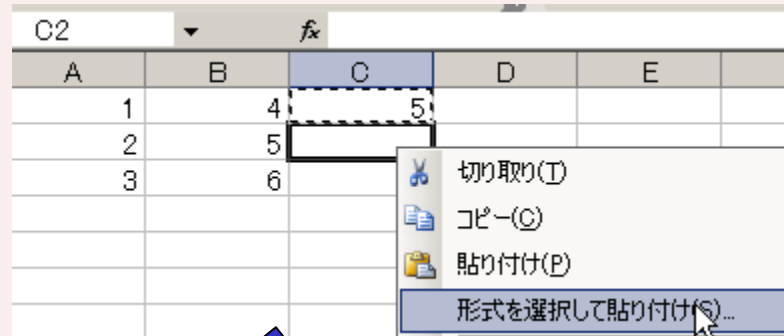
Context menu options:  
切り取り(T)  
コピー(C)  
貼り付け(P) (highlighted)  
形式を選択して貼り付け(S)...



C2      fx =A2+B2

	A	B	C
1	1	4	5
2	2	5	7
3	3	6	
4			

# 形式を選択して貼り付け



	A	B	C
1	1	4	5
2	2	5	5
3	3	6	

式ではなく値を貼り付けることができる。

# 形式を選択して貼り付け

	A	B	C	D	E
1	1	4			
2	2	5			
3	3				
4					
5					
6					
7					

メニュー:  
切り取り(T)  
コピー(C)  
貼り付け(P)  
形式を選択して貼り付け(S)...

形式を選択して貼り付け

貼り付け

- すべて(A)
- 数式(F)
- 値(V)
- 書式(T)
- コメント(O)
- 入力規則(N)
- 罫線を除くすべて(X)
- 列幅(W)
- 数式と数値の書式(B)
- 値と数値の書式(L)

演算

- しない(O)
- 加算(D)
- 減算(S)
- 乗算(M)
- 除算(I)

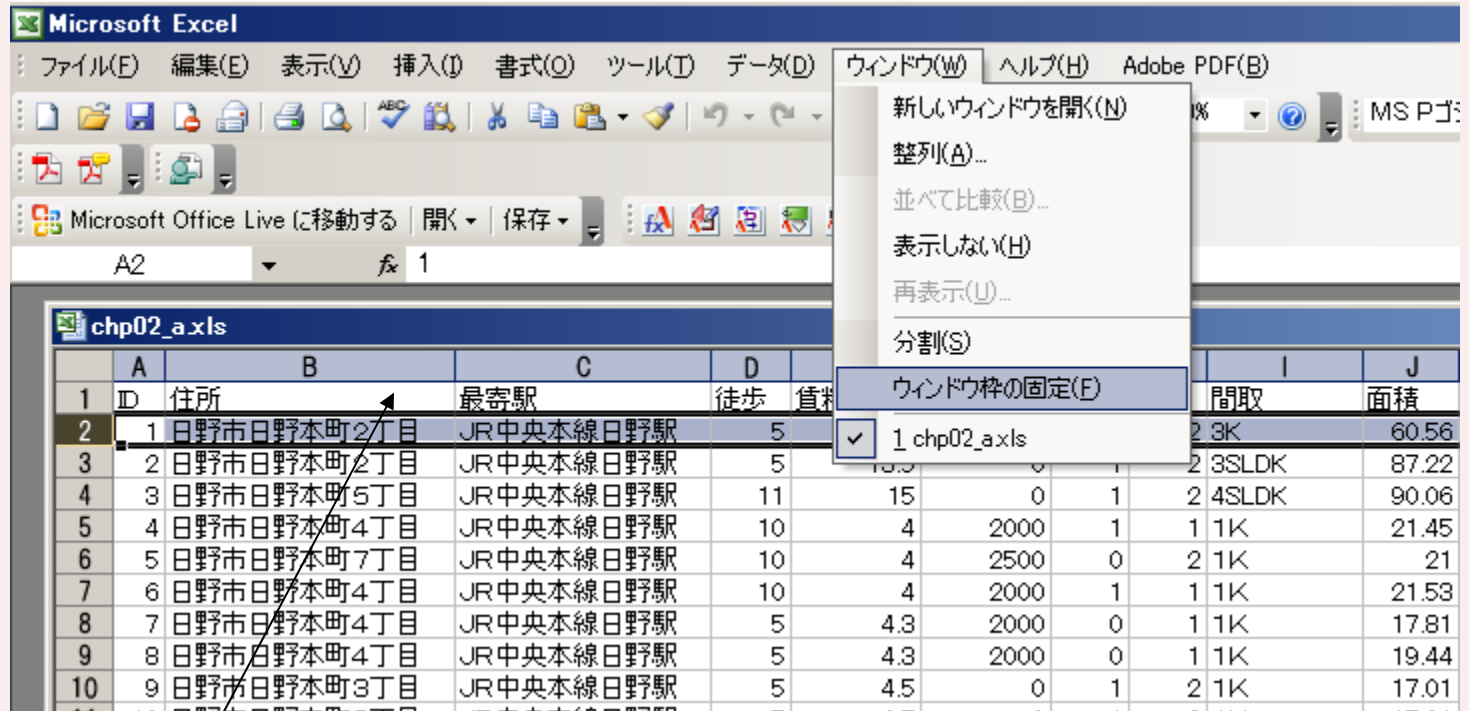
空白セルを無視する(B)  行列を入れ替える(E)

リンク貼り付け(L) OK キャンセル

A	B	C	D	E	F
1	4				
2	5		1	2	3
3	6		4	5	6

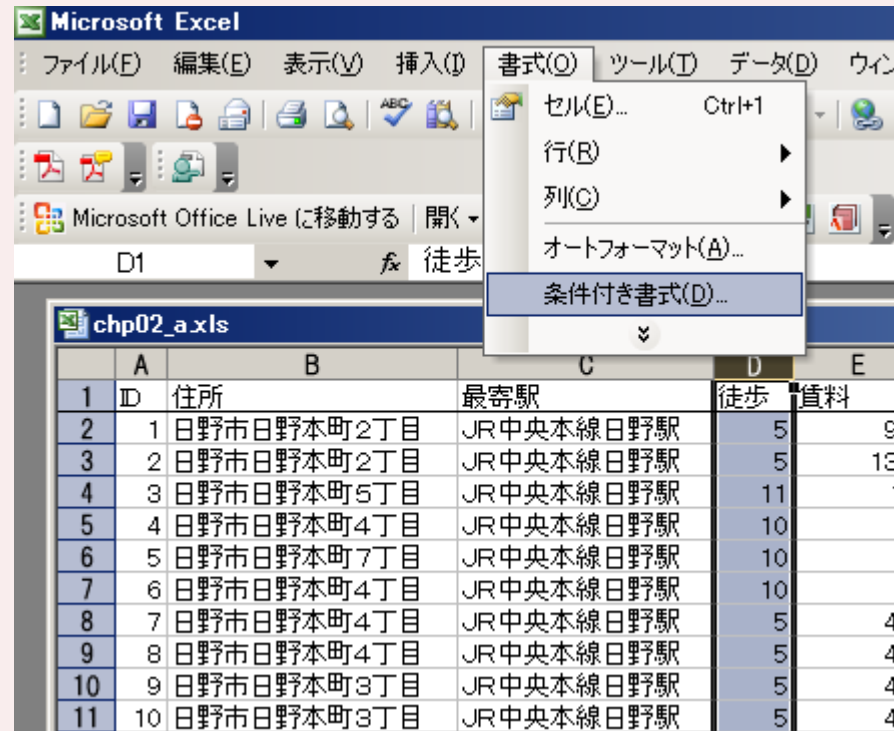
縦横を置き換えて貼り付けられる。

# 画面分割



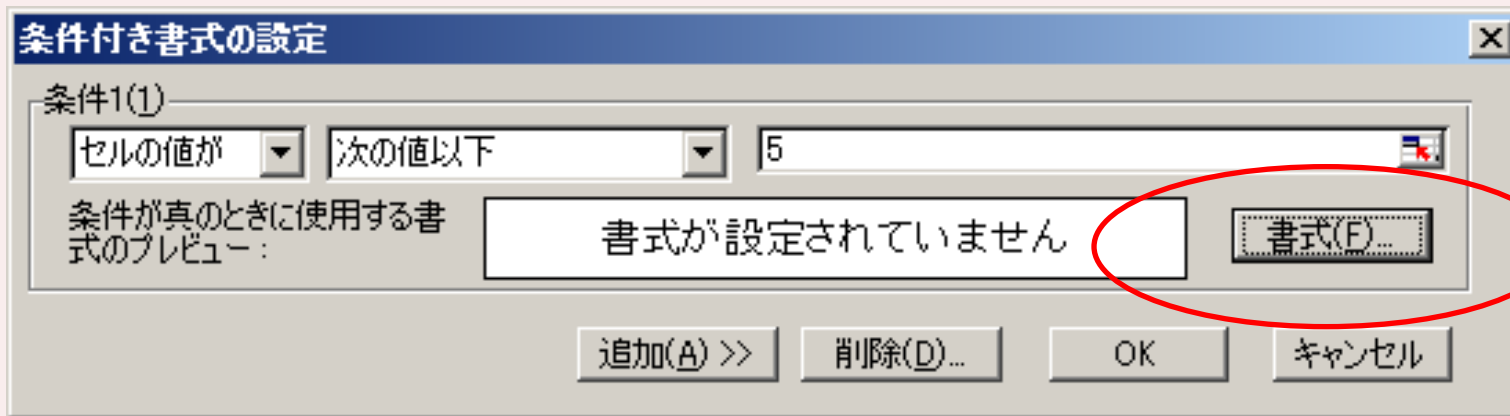
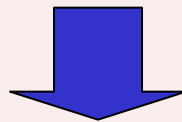
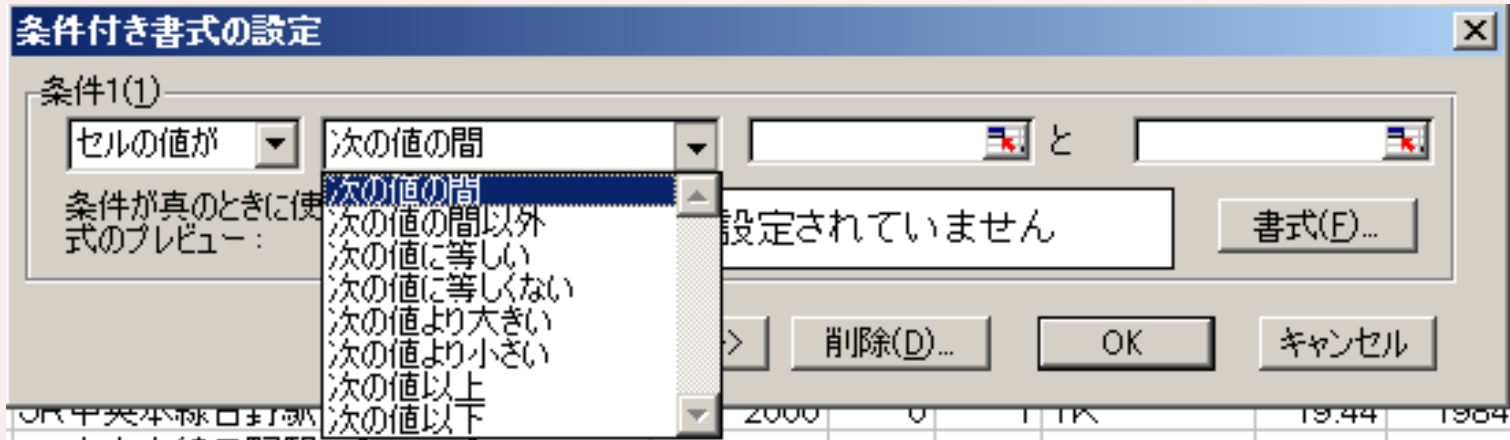
この行を常時表示する

# 条件付き書式



書式を設定したいエリアを選択した上で、  
書式→条件付き書式

# 条件付き書式





# 条件付き書式

条件付き書式の設定

条件1(1)  
セルの値が 次 の値以下 5

条件が真のときに使用する書式のプレビュー: AaBbCcYyZz

追加(A) >> 削除(D)... OK キャンセル

条件付き書式の設定

条件1(1)  
セルの値が 次 の値以下 5

条件が真のときに使用する書式のプレビュー: AaBbCcYyZz

条件2(2)  
セルの値が 次 の値の間 6 と 10

条件が真のときに使用する書式のプレビュー: AaBbCcYyZz

条件3(3)  
セルの値が 次 の値以上 11

条件が真のときに使用する書式のプレビュー: AaBbCcYyZz

追加(A) >> 削除(D)... OK キャンセル

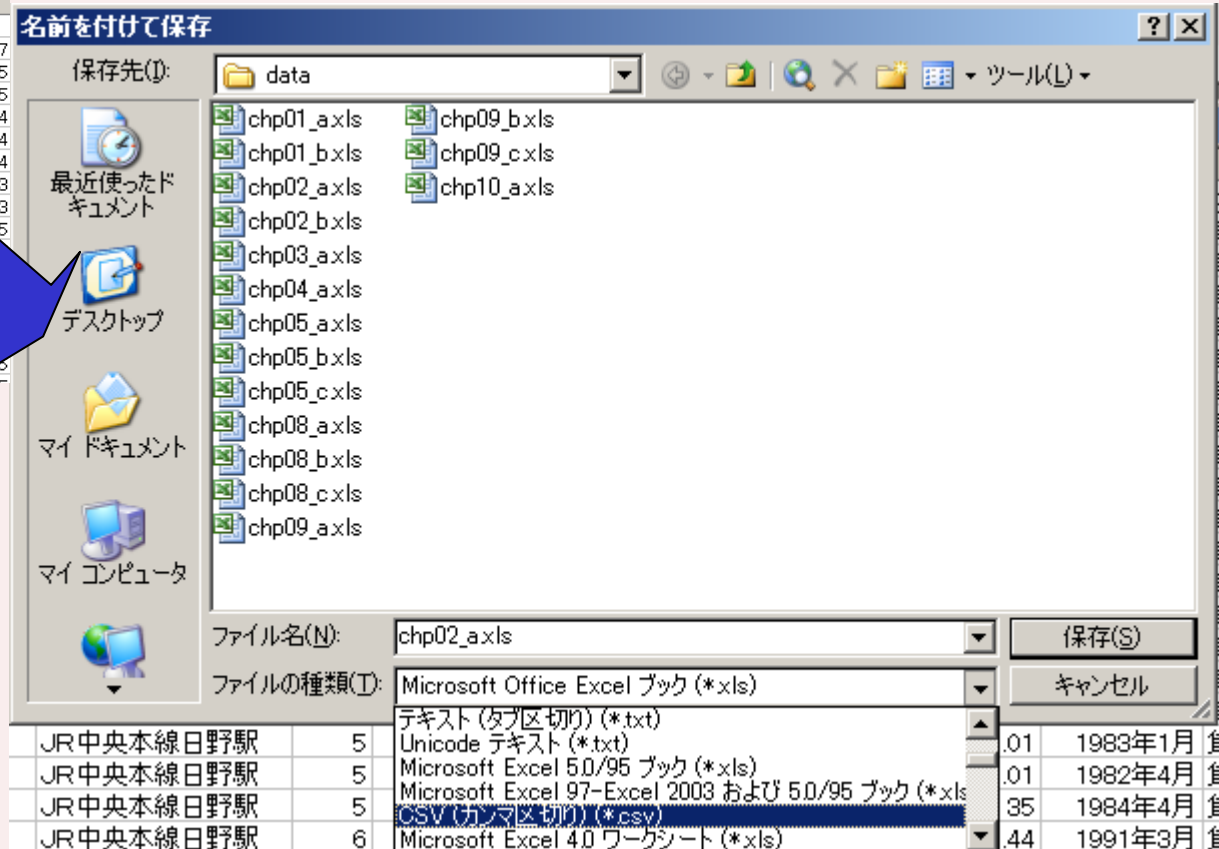
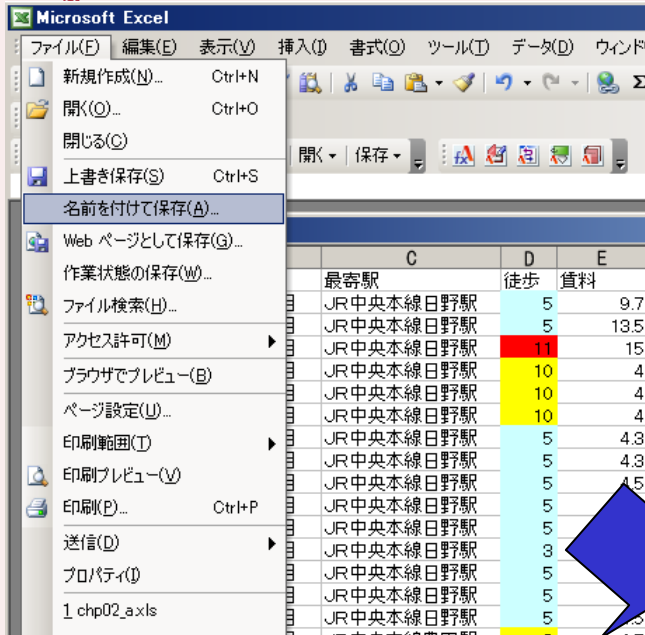
# 条件付き書式

C	D	E	F	G	H	I
最寄駅	徒歩	賃料	管理費等	礼金	敷金	間取
JR中央本線日野駅	5	9.7	3000	1	2	3K
JR中央本線日野駅	5	13.5	0	1	2	3SLDK
JR中央本線日野駅	11	15	0	1	2	4SLDK
JR中央本線日野駅	10	4	2000	1	1	1K
JR中央本線日野駅	10	4	2500	0	2	1K
JR中央本線日野駅	10	4	2000	1	1	1K
JR中央本線日野駅	5	4.3	2000	0	1	1K
JR中央本線日野駅	5	4.3	2000	0	1	1K
JR中央本線日野駅	5	4.5	0	1	2	1K
JR中央本線日野駅	5	4.5	0	1	2	1K
JR中央本線日野駅	5	4.5	0	1	1	1K
JR中央本線日野駅	3	4.5		1	1	ワンルーム
JR中央本線日野駅	5	4.5	0	1	1	1K
JR中央本線日野駅	5	4.5	0	1	1	ワンルーム
JR中央本線日野駅	5	4.5	0	1	1	ワンルーム
JR中央本線日野駅	5	4.5	0	1	1	ワンルーム
JR中央本線日野駅	5	4.5	0	1	1	1K
JR中央本線日野駅	5	4.5	0	1	1	1K
JR中央本線日野駅	5	4.9	0	1	1	2K
JR中央本線日野駅	6	4.9	2000	0.5	0.5	1K
JR中央本線日野駅	5	5	2000	1	1	2DK
JR中央本線日野駅	15	5	2000	1	1	ワンルーム
JR中央本線日野駅	11	5	2000	0	1	ワンルーム
JR中央本線日野駅	6	5	2000	0.5	0.5	1K

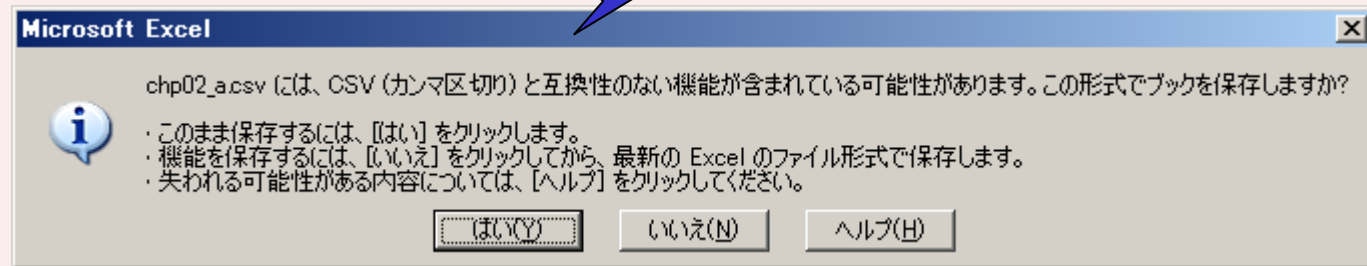
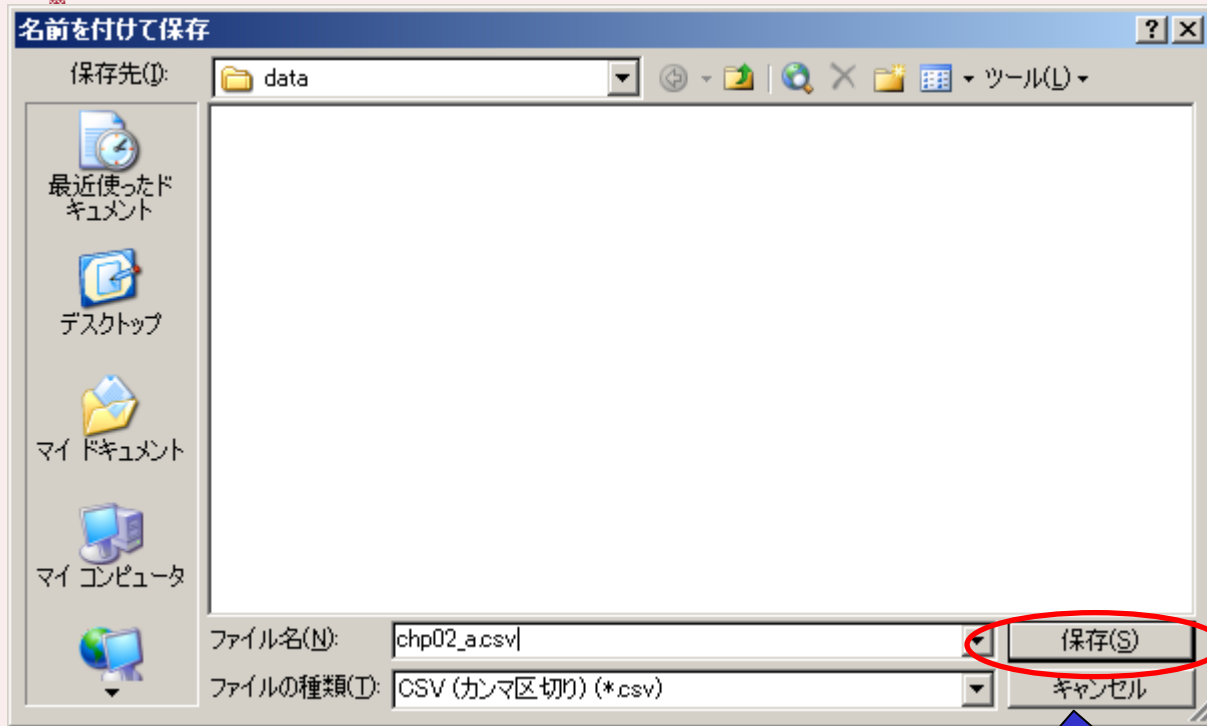


# 保存形式の選択

pp.36-39



# 保存形式の選択

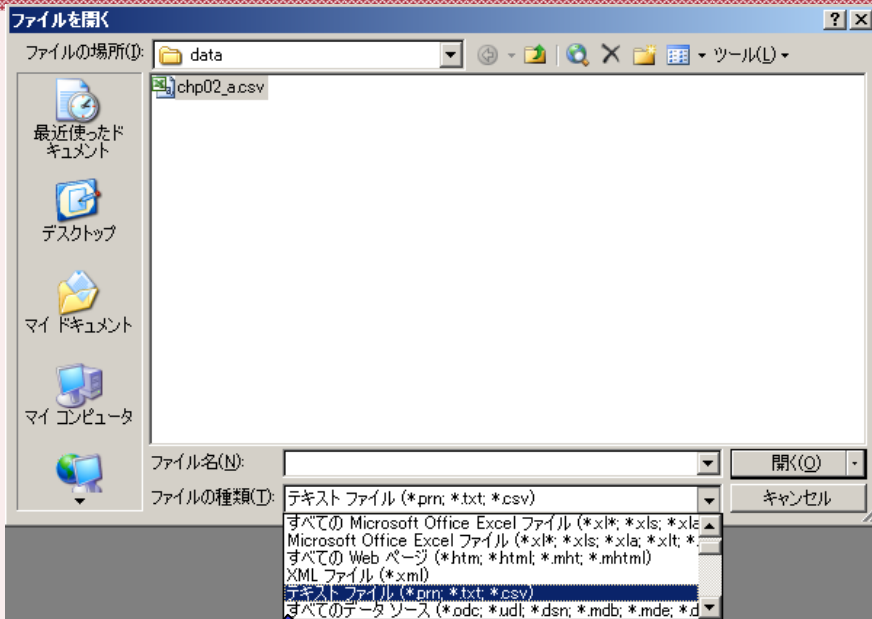


# 保存形式の選択

chp02\_acsv

	0	10	20	30	40	50	60	70
1	ID,住所,最寄駅,徒歩,賃料,管理費等,礼金,敷金,間取,面積,築年数,物件種目↓							
2	1,日野市日野本町2丁目,JR中央本線日野駅,5,9.7,3000,1,2,3K,60.56,1990年3月,賃貸 テラスハウス↓							
3	2,日野市日野本町2丁目,JR中央本線日野駅,5,13.5,0,1,2,3SLDK,87.22,1974年3月,賃 貸一戸建て住宅↓							
4	3,日野市日野本町5丁目,JR中央本線日野駅,11,15,0,1,2,4SLDK,90.06,1985年5月,賃貸 一戸建て住宅↓							
5	4,日野市日野本町4丁目,JR中央本線日野駅,10,4,2000,1,1,1K,21.45,1987年3月, 賃貸アパート↓							
6	5,日野市日野本町7丁目,JR中央本線日野駅,10,4,2500,0,2,1K,21,1985年3月,賃貸 アパート↓							
7	6,日野市日野本町4丁目,JR中央本線日野駅,10,4,2000,1,1,1K,21.53,1987年3月, 賃貸アパート↓							
8	7,日野市日野本町4丁目,JR中央本線日野駅,5,4.3,2000,0,1,1K,17.81,1984年4月, 賃貸アパート↓							
9	8,日野市日野本町4丁目,JR中央本線日野駅,5,4.3,2000,0,1,1K,19.44,1984年4月, 賃貸アパート↓							
10	9,日野市日野本町3丁目,JR中央本線日野駅,5,4.5,0,1,2,1K,17.01,1982年4月,賃 貸アパート↓							
11	10,日野市日野本町3丁目,JR中央本線日野駅,5,4.5,0,1,2,1K,17.01,1982年4月,賃 貸アパート↓							
12	11,日野市日野本町3丁目,JR中央本線日野駅,5,4.5,0,1,1,1K,17.01,1982年4月,賃 貸アパート↓							

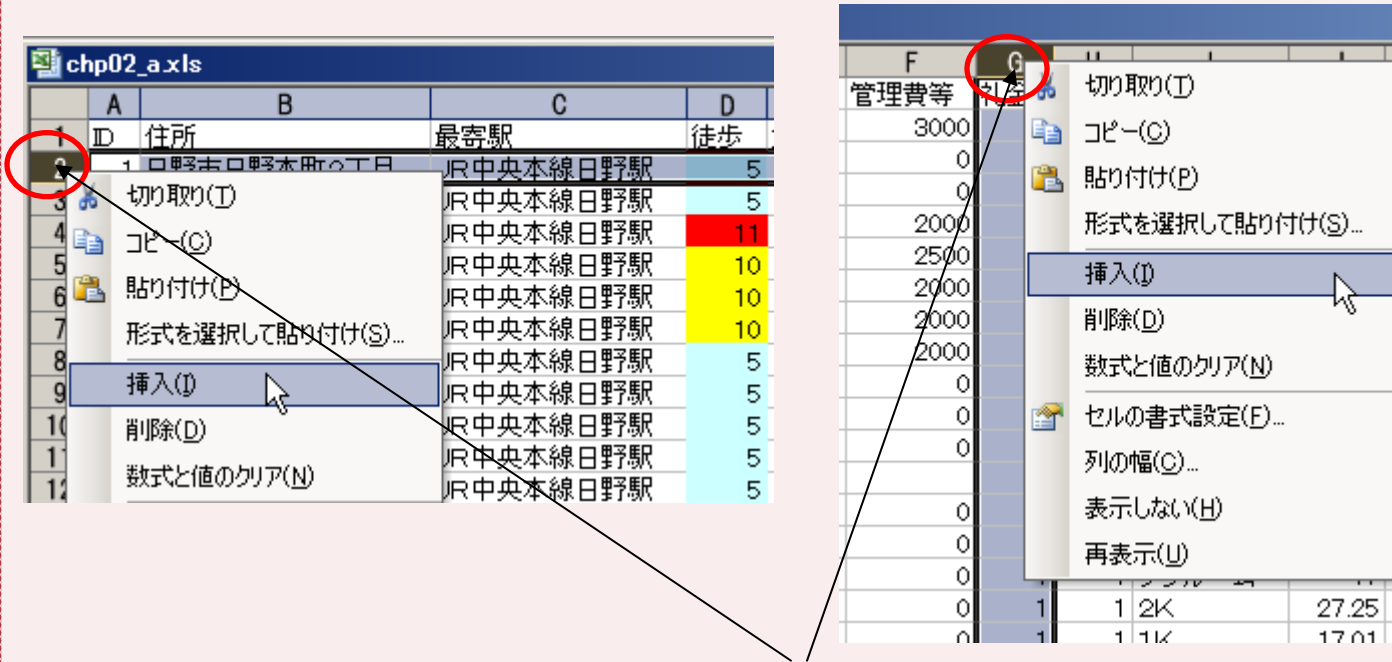
# 保存形式の選択



	A	B	C	D	E	F	G	H	I
1	ID	住所	最寄駅	徒歩	賃料	管理費等	礼金	敷金	間取
2	1	日野市日野	JR中央本線	5	9.7	3000	1	23K	
3	2	日野市日野	JR中央本線	5	13.5	0	1	23SLDK	
4	3	日野市日野	JR中央本線	11	15	0	1	24SLDK	
5	4	日野市日野	JR中央本線	10	4	2000	1	11K	
6	5	日野市日野	JR中央本線	10	4	2500	0	21K	
7	6	日野市日野	JR中央本線	10	4	2000	1	11K	
8	7	日野市日野	JR中央本線	5	4.3	2000	0	11K	

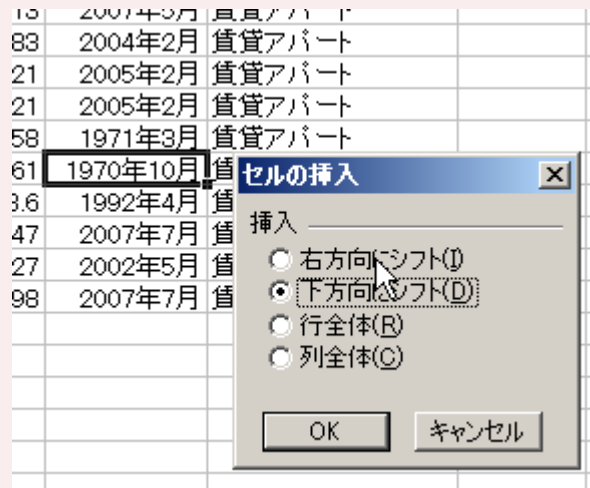
書式情報（ここでは，条件付き書式など）が消えているのに注意。

# セルの挿入（削除）



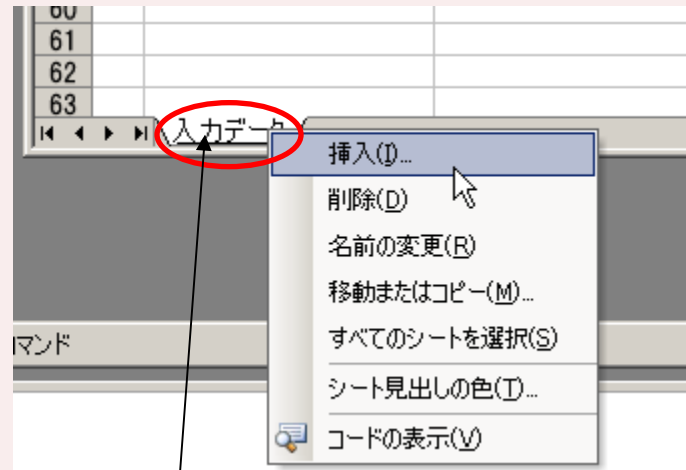
ここで右クリックしてメニュー選択すると、行単位、列単位で挿入、削除などが出来る。挿入の場合、その行の上(1行目と2行目の間)、その列の左(F列とG列の間)に行、列が挿入される。

# セルの挿入（削除）



- **右方向にシフト**：選択したセルと、それより右にあるセルを右に移動させてセルを挿入する。
- **下方方向にシフト**：選択したセルと、それより下にあるセルを右に移動させてセルを挿入する。
- **行全体**：選択したセルのある行の上に1行挿入する。
- **列全体**：選択したセルのある列の左に1列挿入する。
- 削除の場合は「右方向」が「左方向」，「下方方向」が「上方方向」に変わる。

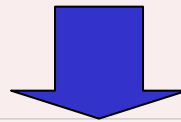
# ワークシートの挿入 (削除)



ここを右クリックすることで、ワークシートもセル同様の操作ができる。

# 関数の利用

	C	D	E
関数の挿入	本線日野駅	2	6.5
目	JR中央本線日野駅	8	7
目	JR中央本線日野駅	7	7
目	JR中央本線日野駅	8	7.3
目	JR中央本線日野駅	7	7.3
目	JR中央本線日野駅	8	7.4



**関数の挿入** [?] [X]

関数の検索(S):  
何がしたいかを簡単に入力して、[検索開始] をクリックしてください。

関数の分類(C):  ▼

関数名(N):

- CONVERT
- CORREL
- COS
- COSH
- COUNT**
- COUNTA
- COUNTBLANK

COUNT(値1,値2,...)  
引数リストに含まれる数値の個数を返します。

[この関数のヘルプ](#)

# 関数の利用

=COUNT(E2:E54)

C	D	E	F	G	H	I	J	K	L
JR中央本線日野駅	3	5.2	0	1	2 1K	23.77	1985年2月	賃貸アパート	
JR中央本線日野駅	5	5.3	2000	1.5	1 1K	19.4	1993年3月	借管アパート	
JR中央本線日野駅	5	5.3							
JR中央本線日野駅	9	5.3							
JR中央本線日野駅	5	5.3							
JR中央本線日野駅	11	5.4							
JR中央本線日野駅	9	5.5							
JR中央本線日野駅	9	5.5							
JR中央本線日野駅	9	5.5							
JR中央本線日野駅	5	5.5							
JR中央本線日野駅	9	5.5							
JR中央本線日野駅	9	5.5							
JR中央本線日野駅	5	5.5							
JR中央本線日野駅	9	5.5							
JR中央本線日野駅	6	5.6							
JR中央本線日野駅	8	5.8							
JR中央本線日野駅	4	5.9							
JR中央本線日野駅	9	6							
JR中央本線日野駅	4	6	3000	2	2 1K	20.13	2007年5月	賃貸アパート	
JR中央本線日野駅	8	6.1	2000	1	2 1K	20.83	2004年2月	賃貸アパート	
JR中央本線日野駅	8	6.3	0	1	2 1K	21	2005年2月	賃貸アパート	
JR中央本線日野駅	8	6.3	0	1	2 1K	21	2005年2月	賃貸アパート	
JR中央本線日野駅	2	6.5	0	1	1 2DK	43.58	1971年3月	賃貸アパート	
JR中央本線日野駅	8	7	0	1	2 2DK	37.61	1970年10月	賃貸アパート	
JR中央本線日野駅	7	7	2000	1	2 1LDK	33.6	1992年4月	賃貸アパート	
JR中央本線日野駅	8	7.3	2000	1	2 1LDK	37.47	2007年7月	賃貸アパート	
JR中央本線日野駅	7	7.3	4000	1	2 1K	27	2002年5月	賃貸アパート	
JR中央本線日野駅	8	7.4	2000	1	2 1LDK	34.98	2007年7月	賃貸アパート	

関数の引数

COUNT

値1 E2:E54 = {9;7;13;5;15;4;4;4;3;4}

値2 = 数値

= 53

引数リストに含まれる数値の個数を返します。

値1: 値1,値2,... (にはデータまたはデータが入力したセルの参照を 1 から 30 個まで指定します。数値データだけがカウントされます。)

数式の結果 = 53

[この関数のヘルプ\(H\)](#) OK キャンセル

ここをクリックし、関数を適用する範囲をマウス等で選択する。

# 関数の利用

	C	D	E
	JR中央本線日野駅	8	7
	JR中央本線日野駅	7	7
	JR中央本線日野駅	8	7.3
	JR中央本線日野駅	7	7.3
	JR中央本線日野駅	8	7.4
			53

建物ID1から53までの物件が，欠番なしであるので，合計53軒である。

セルをクリックし，**=COUNT(件数を数えた  
い範囲)**，すなわち **=COUNT(E2:E54)** と入力しても同じ結果が得られる。

# 分析ツールの利用

- “ツール”メニューの中に“分析ツール”があるか確認。なければ“ツール”から“アドイン”で“分析ツール”にチェック。

