2025/02/19

#### 藪友良

# 付録 B

# エクセルの使い方

統計学を勉強しても、やはり実際に自分で使ってみないと理解は十分ではあ りません。ここでは、実際に統計分析を使う方法のひとつとして、エクセルの 使い方を解説します。本稿は、藪友良『入門 実践する統計学』(2012 年、東洋 経済新報社)の付録 B をアップデートしたものです。データは、本書ウェブサイ トからファイル appendix\_b\_data.xlsx をダウンロードしてください。

# B.1 分析ツール

エクセルについている分析ツールという機能を使えば、さまざまな統計分析 が可能です。まず、この機能を使えるように設定します。もし[データ]タブに [データ分析]という項目があれば、本節は無視してください。[データ分析]がな ければ、以下の手順に従って設定します<sup>1</sup>。

エクセル画面左上の[ファイル]タブをクリックします(下画面参照)。



そうすると、[Excel のオプション]ウィンドウ画面がでますから、下画面から [アドイン]を選んで、下画面下にある[設定]をクリックします。

リボンのユーザー設定 クイック アクセス ツール バー アドイン トラスト センター	EVIews Add In Microsoft Actions Pane 3 Microsoft Power Map for Excel Microsoft Power Pivot for Excel VID/モーアドイン Eldf (XML) 分析(サール - VBA	C:¥aming¥Microsoft¥AddIns¥EViews Add In.xla C:¥r Map Excel Add-in¥EXCELPLUGINSHELL.DLL C:¥Excel Add-in¥PowerPivotExcelClientAdIn.dli C:¥excel Add-in¥AdHocReportingExcelClient.dli C:¥iexcel Add-in¥AdHocReportingExcelClient.dli C:¥iexf0ffce16¥Library¥SOLYER¥SOLYERX.LAM C:¥iles¥Microsoft Shared¥Smart Tag¥MOFL.DLL C:¥Office16¥Library¥Analysis¥ATPVBAEN.XLAM	Excel アドイン XML 拡張パック COM アドイン COM アドイン COM アドイン Excel アドイン 操作 Excel アドイン
	ドキュメント関連アドインはありません		-
	アドイン: EViews Add-In for Excel 2010 第行者: くなレ> 互換性: 互換性に関する情報はありません 場所: file:///C:/Program Files/EVier	0 ws 11/Excel2010AddIn/EViews_Excel2010_40_AddIn.vsto vstok	ocal
	説明: EViews Add-In for Excel 2016	0 (and later)	
	管理(A): Excel アドイン	(G)	
			OK キャンセル

<sup>&</sup>lt;sup>1</sup> ここではエクセル 2007 に基づいて説明をしています。本節のやり方で設定できないときは、ヘルプ機能で分析ツールを検索し、そこで示されている方法に従って設定してください。

それから下画面で[分析ツール]をチェックして OK とします。

アドイン		? ×	
有効なアドイン( <u>A</u> ):			
Euro Currency Tools	^	ОК	
□ <del>ソルバ</del> ー アドイン ✓ 分析ツール		キャンセル	
□分析ツ <sup>−</sup> ル - VBA		参照( <u>B</u> )	
		オートメーション( <u>U</u> )	

これで[データ]タブの一番右に[データ分析]という項目が加わります。

## B.2 無作為抽出

1章では無作為抽出の方法として、10面体のサイコロを用いる方法を説明しました。しかし、エクセルを使えば無作為抽出を容易に行うことができます(エクセルのシート「ID」参照)。

たとえば、1000人の学生から 10人を無作為抽出するときは、まず全員に 1 から 1000までの番号(ID)を付けます。次に、[データ]タブの[データ分析]を クリックします。そして、[サンプリング]を選んでから OK をクリックします。

	५ - ५ - ।	ñ · 🗸	• •							appendix_	_b_da	ata - Excel					<u>.</u>	YT YT [
ファイル		挿入	ページ レイアウト	数式	データ	<b></b>		アドイン	ヘルプ	♀ 何ē	します;	か						
データの 取得・	♪ テキストま ♪ Web から … テーブルま デ	たは CSV が う たは範囲か ータの取得。	から	·ソース 売	し すべて 更新・日 の	クエリと 指 プロパテ・ 、 リンクの 約 「リンクの 約 「リン 接続	<b>登続</b> 〈 福集	2↓ <u>₹ 2</u> <sub>₹↓</sub> 並べ替え 並	フィルター マイルター	<ul> <li>∑ クリア     <li>万 再適用     <li>▼ 詳細設     <li>ルター     </li> </li></li></li></ul>	定	■ 区切り位置 データッール	What-I	<b>?</b> 分析 予測 シート 予測	1個 グ 1個 グ 1回 小	ループ化 ・ ループ解除 い計 アウトライン	+	データ分析
08	٣	×	√ fx															
1 ID 2 3 4 5 6	A 1 2 3 4 5	В	C [	デー: 分: フレ	E タ分析 析ツール(A) ーリエ解析 ストグラム	F	C	3	H	I	J	К ? X ОК	L	M	N	0	Ρ	Q
7 8 9 10 11 12 13	6 7 8 9 10 11 12			移乱 順 世 t t t	<ul> <li>動平均</li> <li>動平均</li> <li>調査と百分位</li> <li>調査の析</li> <li>ンプリング</li> <li>検定: 等分帯</li> <li>検検定: 分帯</li> </ul>	数 の標本に。 )散を仮定し が等しくな	tる平均の した 2 標ス いと仮定し	検定 本による検定 た 2 標本に	よる検定		v	ヘルプ(日)						

すると、左下の画面が出てきます。ここで□に情報を入力します。A 列の 2 行から 1001 行まで番号が記録されていますから、入力範囲として、 \$A\$2:\$A\$1001 を入力します。入力が面倒であれば、▲ボタンを押して、デー タの入力範囲を画面上で範囲指定することもできます。また、10 人を無作為 抽出しますから、データの個数に 10 を入力します。そして OK を押すと右下 の画面となり、乱数が抽出されます。この場合、選ばれた乱数は 949、853、 517、284、956、628、396、67、739、3 です。この番号に対応した 10 人の 学生を選べば無作為抽出は完了です<sup>2</sup>。無作為抽出ですから、みなさんは異な る番号が抽出されたと思います。

サンプリング		?	×		A	В
入力元		ОК		1	949	
入力範囲( <u>I</u> ):	\$A\$2:\$A\$1001	キャンセリ	L.	2	853	
□ ラベル( <u>L</u> )				3	517	
標本の採取方法 ○ 周期変化(E)		ヘルフ(日	[)	4	284	
周期:				5	956	
ランダム( <u>R</u> )				6	628	
データの個数:	10			7	396	
出力オプション				8	67	
〇 出力先( <u>Q</u> ):	Ť			9	739	
● 新規ワークシート( <u>P</u> ):				10	3	
○ 新規ブック( <u>W</u> )				4.4		

#### B.3 特性値の計算方法

エクセルを使った特性値の計算方法を説明します。エクセルにはさまざまな 特性値を計算するため、多くの関数が定義されています。たとえば、平均であ れば=AVERAGE(配列)とすれば、配列指定したデータを使って、平均を計算して くれます(配列はデータ入力範囲)。代表的関数には

平均=AVERAGE(配列)、中央値=MEDIAN(配列)、最頻値=MODE(配列)、標本分散=VAR(配列)、標本標準偏差=STDEV(配列)、標本共分散 =COVAR(配列1,配列2)、標本相関係数=CORREL(配列1,配列2) などがあります。標本共分散と標本相関係数は2変数の関係をとらえる指標で すから、配列を2つ指定します。

下画面は GDP を 1991~2004 年まで記録したものです(エクセルのシート 「GDP」参照)。A 列には時間(年)、B 列には GDP が記録され、C 列では

<sup>&</sup>lt;sup>2</sup>無作為抽出すると、同じ番号が重複することがあります。番号の重複を認める方法を復元抽出、 認めない方法を非復元抽出といいます。たとえば、番号 1~100 から 3 つを抽出したとき、乱数と して 91、7、91 が抽出されたとしましょう。復元抽出なら、選ばれた乱数は 91、7、91 ですが、 非復元抽出なら(もう一度抽出したところ乱数 80 が得られたとする)、選ばれた乱数は 91、7、 80 です。非復元抽出は同一母集団からの抽出とならず、i.i.d.の仮定は満たされません。ただし、 母集団規模が十分に大きければどちらでも変わりません。

GDP 成長率(変化率)を計算しています。たとえば、C列の3行目は0.0245 となっていますが、これは=(B3-B2)/B2として計算できます<sup>3</sup>。そして、この セル(C3)をコピーして、C列の4~15行までに貼り付ければ変化率を全て 計算できます。ここで GDP の平均は=AVERAGE(B2:B15)、GDP と成長率の相関 は=CORREL(B3:B15, C3:C15)として計算できます(計算すると、それぞれ 4934272と0.1009)。

SUM		• : ×	✓ fx	=AVERAGE	E(B2:B15)	
	А	В	С	D	Е	F
1		GDP(名目)	成長率			
2	1991	4722614			AVERAGE	(B2:B15)
3	1992	4838375	0.02451206			
4	1993	4806615	-0.006564187			
5	1994	4869469	0.013076562			
6	1995	4957357	0.018048785			
7	1996	5064795	0.021672436			
8	1997	5104658	0.007870605			
9	1998	5013835	-0.017792181			
10	1999	4966058	-0.009529033			
11	2000	5027831	0.012439041			
12	2001	4923467	-0.020757261			
13	2002	4887244	-0.007357214			
14	2003	4935532	0.009880415			
15	2004	4961970	0.005356667			
10			T			

# B.4. 図の描き方

ここでは図の書き方を説明します。下画面では、1991~2004 年までの GDP が記録されています。図を描くためには、まず、下画面のようにデータ 範囲を指定します。指定が終わってから、[挿入]タブをクリックし、折れ線マ ーク<sup>が、</sup>をクリックします(折れ線以外に、散布図、縦棒、横棒、円グラフな どもあります)。

<sup>&</sup>lt;sup>3</sup> 1992年の変化率は、1991~92年の変化分(B3-B2)を1991年の値(B2)で割ったものです。

E	<b>5</b> • d	» - 🛍 - 🗶	* ÷						
ファ	イル ホー	ム挿入	ページ レイアウト		数式		校閲	表示	アドイン
ピポテー	シー おすう ブル ピボットラ テーン	<b>?</b> すめ テーブル モーブル ガル		∃ア ●個	ドインを入手  人用アドイ アドイン		おすす グラフ		■ ・ ペ ・ ・ ・ グラフ
B1	5	• : ×	$\checkmark f_X$	49	961970				
	А	B	C		D	F		F	G
1		GDP(名目)	成長率		0	_		•	ų
2	1991	4722614							
3	1992	4838375	0.02451206						
4	1993	4806615	-0.006564187	7					
5	1994	4869469	0.013076562						
6	1995	4957357	0.018048785						
7	1996	5064795	0.021672436						
8	1997	5104658	0.007870605						
9	1998	5013835	-0.017792181						
10	1999	4966058	-0.009529033	3					
11	2000	5027831	0.012439041						
12	2001	4923467	-0.020757261						
13	2002	4887244	-0.007357214	ŧ.					
14	2003	4935532	0.009880415						
15	2004	4961970	0.005356667						
16									

そして折れ線の中で、自分が描きたい図の種類を選択します。ここでは、左 上の図を選びましょう。



そうすると、下図が出てきます。横軸は年、縦軸は GDP の規模となってい ます。このままではあまりきれいな図ではありませんから、微調整して図を見 やすくする必要があります。作成された図をクリックして、[デザイン]タブを 選ぶと、さまざまな微調整ができるようになります。また、図の軸をクリッ クし、書式設定を選ぶことでも調整ができます。



# B.5 標本相関係数の計算

エクセルを使って標本相関係数を計算しましょう。5章ではトヨタ、ホン ダ、ユニクロの株価の動きを紹介しました。ここでは、これら3系列の標本相 関係数を計算します(エクセルのシート「株価」参照)。まず、[データ]タブの [データ分析]をクリックします。そして、[相関]をクリックして OK します。

6	5-∂-	🛍 * 🗶 *																	үт ਆ
	イル ホーム	挿入 ベ	ミージ レイアウ	ト数式	<del>7</del> -9	交閲	表示	アドイン	ヘルプ	Q 何を	しますか								
<b>デ</b> ー 取行	□ テキスト □ テキスト 900 □ Web か 日 テーブル	または CSV から ら または範囲から データの取得と愛	5 🔂 最近使 最近使 の の の の の の の の の の の	ったソース 接続		クエリと接 プロパティ ,リンクの編 と接続	続 会↓ 該集	並べ替え	フィルター 替えとフィリ	<ul> <li>ヘリア</li> <li>● 再適用</li> <li>● 詳細設</li> </ul>	e E	● ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■	- 0	What-	? If 分析予述 * シー		グループ化 ・ グループ解除 小計 アウトライン		■データ分析 分析
GL	1 -	: ×	∠ fx																
GI	+		- J.																
	А	В	С	D	Е	F	G	E F	4	I	J	к		L	м	N	0	Р	Q
1		F38 7	トンダ :	ユニクロ				_		_			_						
2	2007年1月	7950	4750	9530	データ分析							?	×						
3	2007年2月	8020	4420	9680	A15W 11/	A.)							_						
4	2007年3月	7550	4110	9150	75479-74	8)						OK							
5	2007年4月	7320	4130	8250	分散分析:	一元配置	+ 7	-			^		-						
0	2007年5月	7300	4280	9430	プ配分析: 分散分析:	線り返しの	ゆる二兀配百 かい二元配	旦 満				キャンセル	·						
8	2007年0月	7800	4300	6750	相関	AN	//or v · / LBU	paar.											
a	2007年8日	6760	3820	6790	共分散							ヘルプ(日	)						
10	2007年9月	6780	3860	6630	基本統計量	ł													
11	2007年10月	6570	4300	6620	指数平滑	机电子放散	-t-/\#baik												
12	2007年11月	6240	3760	7300	► 使定: 2 7—UT解析	伝承を使	つに対戦の棋												
13	2007年12月	6040	3750	7980	ヒストグラム						~								

すると、左下の画面が出てきます。ここで□に情報を入力します。B~D列の 1~31行にデータが記録されていますから、入力範囲として、\$B\$1:\$D\$31を代 入します。ただし、最初のセルはデータ名ですから、[先頭行をラベルとして使 用]をチェックしてください。そうすると、最初のセルをラベルとして認識して くれます。そして OK すると、右下の画面が出てきます。同じ系列同士では相 関係数は1となりますから、対角要素は全て1です。トヨタとホンダの相関係 数は0.960、トヨタとユニクロの相関係数は-0.697、ホンダとユニクロの相関 係数は-0.629となっています。



#### **B.6** 確率の計算

ここでは確率変数の確率を計算する方法を説明します。

*Z*は標準正規確率変数とします。確率 *P*{*Z*<*z*}を求めたい場合、=NORMSDIST(z) と入力します。たとえば、*P*{*Z*<1.96}を求めるには=NORMSDIST(1.96)と入力す れば 0.9750 が得られます(エクセルのシート「確率、臨界値の計算」参照)。

Wは自由度 nの  $\chi^2$  確率変数とします。このとき、 $P\{\chi^2_{n,\alpha} < W\} = \alpha$ となる  $\chi^2_{n,\alpha}$ の 値を計算するには、=CHIINV( $\alpha$ , n)と入力します。たとえば、 $\chi^2_{99,0.025}$ を求める ときは、=CHIINV(0.025,99)と入力すれば 128.42 が得られます。また、 $\chi^2_{99,0.975}$ を求めたい場合は、=CHIINV(0.975,99)と入力すれば 73.36 が得られます。

Uは自由度 n の t 確率変数とします。このとき、 $P\{t_{n,\alpha} < |U|\} = \alpha$  が成立する  $t_{n,\alpha}$ を計算するには、=TINV( $\alpha$ , n)と入力します。たとえば、 $t_{1,0,1}$ は、=TINV(0.1, 1)と入力すると、6.314 が得られます。

Vは第1自由度 m、第2自由度 n の F 確率変数とします。P{F<sub>m,n</sub><V}=αとなる F<sub>m,n</sub>の値を計算するには=FINV(α,m,n)と入力します。α=0.05、m=n=2499 であれば、=FINV(0.05, 2499, 2499)として 1.068 が得られます。

#### <u>B.7. 回帰分析</u>

回帰分析の仕方とその解釈を説明します。2010/1/1~2010/12/31の円ドルレートを使って、為替レートが予測できるのかを調べます(エクセルのシート「円ドル」参照)。t日の為替レートを St とし、t日の変化率を

$$dS_t = \frac{S_t - S_{t-1}}{S_{t-1}}$$

と表します。

ここでは被説明変数を t 日の変化率 $(dS_t)$ 、説明変数を t-1 日の変化率 $(dS_{t-1})$  とします。

$$dS_t = \alpha + \beta dS_{t-1} + u_t$$

βが0から有意に異ならなければ為替の予測はできませんし、βが有意に0か ら異なれば為替の予測が可能だといえます。これは自己回帰(AR)モデルとい われ、時系列分析で重要なモデルのひとつです(詳しくは、参考文献で紹介さ れている本を読んでください)。 [データ]タブの[データ分析]をクリックします。そして、[回帰分析]を選択して OK をクリックします。

1	A	В	С	D	E	F	G	Н	I	J	к	L
1		円ドル(17:00)	t日の変化率	t-1日の変化	七率							
2	2010/1/6	92.12	0.002448447	-0.01231		-					2	~
3	2010/1/7	92.79	0.007273122	0.002448	アーダ分析	Л					?	×
4	2010/1/8	93.28	0.005280741	0.007273	分析ツー	-JL(A)						
5	2010/1/12	91.885	-0.014954974	0.005281	L7L/	r=1.					OK	
6	2010/1/13	91.265	-0.006747565	-0.01495	北部					^	100 million (10	
7	2010/1/14	91.85	0.006409905	-0.00675	乱数争	5年					キャンレ	
8	2010/1/15	90.955	-0.009744148	0.00641	順位と	百分位数						
9	2010/1/18	90.99	0.000384806	-0.00974	回帰分	浙					~)[)(	H)
10	2010/1/19	90.62	-0.004066381	0.000385	サンプ.	ルング						
11	2010/1/20	90.905	0.003145001	-0.00407	t 検定	:一対の標本	による平均の	検定				
12	2010/1/21	91.52	0.006765304	0.003145	t 検定	: 等分散を仮	定した 2 標本	による検定	7+4-			
13	2010/1/22	90.43	-0.011909965	0.006765	て 使正	:	ふいと仮定し	に Z 信本によ	21快止	~		
14	2010/1/25	90,225	-0.002266947	-0.01191	2 192	· 2154166	1-10/19/2					

すると下画面が表示されます。ここで被説明変数(Y) は C の列(t 日の変化 率、つまり $dS_t$ )、説明変数(X) は D の列(t-1 日の変化率、つまり $dS_{t-1}$ )となり ます。したがって、入力 Y範囲は C の列C1:C244 とし、入力 X範囲は D の例D1:D244 と範囲指定します<sup>4</sup>。 Y と X の入力範囲は 1 行目から始めて いますが、1 行目はデータの名前(ラベル)で、データそのものではありませ ん。このため、ラベルにチェックを入れて、データではないことを明示してお きます。また、ここでは有意水準にチェックし、横の空欄に 90 という値を入れ ておきます。こうすることで信頼区間 90%を計算してくれます。



最後に OK をクリックすると、以下の画面が表示されます。ここでは重要な 情報とその意味を紹介していきます。

<sup>&</sup>lt;sup>4</sup>入力 X範囲は1変数ではなく、複数の変数を指定することもできます。その場合は多重回帰分析となります。

回帰	統計							
重相関 R	0.111537							
重決定 R2	0.012441							
補正 R2	0.008343							
標準誤差	0.006376							
観測数	243							
分散分析表								
	自由度	変動	分散	川された分背	有意 F			
回帰	1	0.000123	0.000123	3.035945	0.082715			
残差	241	0.009796	4.06E-05					
合計	242	0.009919						
	係数	標準誤差	t	P−值	下限 95%	上 <mark>限</mark> 95%	下限 90.0%	上限 90.0%
切片	-0.00053	0.00041	-1.28724	0.199245	-0.00134	0.00028	-0.00121	0.000149
+-1日の恋	-0 11007	0.06360	-1 7424	0.082715	-0 23643	0.014487	-0.21614	-0.00581

最初のブロックには、当てはまりの尺度である決定係数などがまとめられて います。重決定 R2 は決定係数であり、0.0124 とあまり当てはまりはよくあり ません。また、補正 R2 は自由度調整済み決定係数であり、0.0083 と低い値で すから、やはり当てはまりはよくありません。観測数はサンプルサイズであり、 この場合は 243 となります。

最後のブロック(分散分析表)に、母数( $\alpha$ 、 $\beta$ )の推定値、標準誤差、95% と 90%信頼区間がまとめられています(有意水準をチェックし、90 と入力し たため、90%信頼区間も表示)。たとえば、 $\alpha$ は-0.00053、 $\beta$ は-0.11097と推 定されています。それぞれの標準誤差は 0.00041 と 0.06369 です。t 値は推定 値を標準誤差で割ったもので、それぞれ-1.28724(=-0.00053/0.00041)、-1.7424(=-0.11097/0.06369)となります。この場合、 $\alpha$ に関するt 値は 0 に近 いため、帰無仮説( $H_0$ :  $\alpha$ =0)は採択されます。これに対し、 $\beta$ に関するt 値 は-1.7424 と 0 から乖離し、有意水準 10%で帰無仮説( $H_0$ :  $\beta$ =0)は棄却さ れ、対立仮説が支持されます。90%信頼区間は、母数が 90%の確率でその範囲 に含まれることを意味します。たとえば、 $\beta$ は 90%の確率で-0.21614 から-0.00581の範囲に収まり、信頼区間内に 0 を含みませんから有意な結果といえ ます。

最後に、p値は「 $H_0$  が正しいにもかかわらず、t統計量の絶対値がt値 ( $t^*$ )の絶対値より大きな値をとる確率( $P\{|t_{\hat{\beta}}|>|t^*|\}$ )」です。たとえば、こ の表から、 $\beta$ に関する t値は  $t^*=-1.7424$  で、p値は 0.082715 となっていま す。p値は、 $H_0$ が正しいもとで、t統計量の絶対値( $|t_{\hat{\beta}}|$ )がt値の絶対値 ( $|t^*|=1.7424$ )より大きくなる確率ですから、 $P\{|t_{\hat{\beta}}|>1.7424\}=0.082715$ となって いるのです。p値を見れば何%の有意水準で帰無仮説を棄却できるかが分かり ます。かりにp値が 1%を下回っていれば、1%の有意水準でも帰無仮説を棄却 できます。この場合、 $\beta$ のp値 0.082715 は 0.1 を下回りますから、有意水準 10%で帰無仮説は棄却されます。