

統計学で使う表計算ソフトの解説

	目次	
1	表計算ソフトの概念	1 頁
2	度数分布表	2 頁
	2.1 度数分布表-countif(2 頁) 2.2 正の字で集計(7 頁) 2.3 度数分布表- frequency(7 頁)	
3	エクセル 2 0 0 3 [データ分析]	8 頁
4	記述統計	8 頁
	4.1 基本統計量(8 頁) 4.2 ヒストグラム(11 頁) 4.3 相関(12 頁) 4.4 共分散(13 頁) 4.5 順位と百分位数(14 頁)	
5	シミュレーション	14 頁
	5.1 擬似乱数の生成(14 頁) 5.2 サンプリング(15 頁) 5.3 順位と百分位数(16 頁)	
6	検定	17 頁
	6.1 F 検定(17 頁) 6.2 t 検定(17 頁) 6.3 z 検定(20 頁)	
7	分散分析	21 頁
	7.1 分散分析: 一元配置(21 頁) 7.2 分散分析: 繰り返しのある二元配置(21 頁)	
8	予測のための統計	23 頁
	8.1 回帰分析(23 頁) 8.2 移動平均(24 頁) 8.3 指数平滑(24 頁)	

1 表計算ソフトの概念

よく用いられる表計算ソフトには、マイクロソフト (Microsoft) 社のエクセル (excel)  や、サン (Sun Microsystems) 社のオープンオフィス=カルク (open-office calc)  MySQL  があります。前者は購入する必要がありますが、後者はフリーウェアですから、規約を守ることですべて自由に使えます。

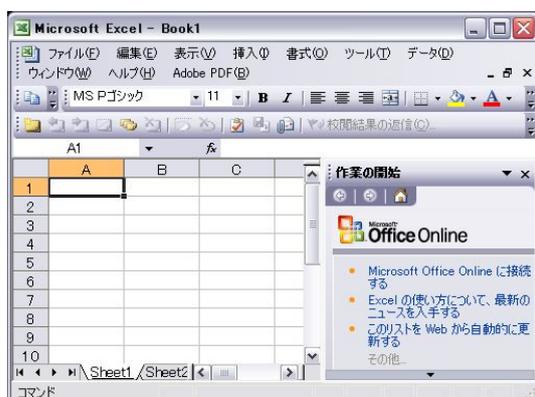


図 1 エクセルの立ち上がり画面

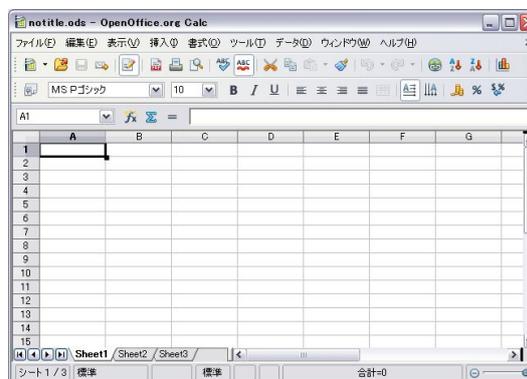
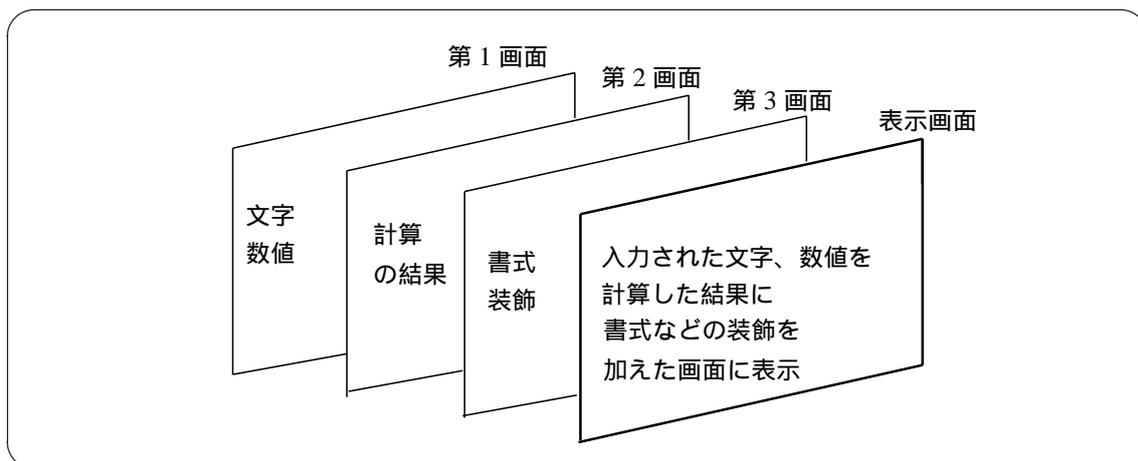


図 2 オープンオフィスのカルク

表計算ソフトの画面には、3 つの階層がまとめられて表示されると考えれば理解し易いでしょう。その 3 つの層とは、つぎのものです。表計算ソフトの 1 枚のシートには、3 層と表示領域から構成される。多くの機能を使いこなすことは、データの整理やプレゼンテーションにはたいへん役に立ちます。

- 第 1 層には文字，数値，数式が値の規則として入力
- 第 2 層には数式による計算結果が値として入力
- 第 3 層には表示形式，書式など，さらに表示文字色などで修飾した結果が入力
- 最終の表示領域にはこれらの結果が表示される



以下では、統計ではよく用いられるデータ集計の整理のために、度数分布表とそのグラフ化（ヒストグラム）をつくる方法の説明を述べます。

2 度数分布表

2.1 度数分布表-countif

Excel のグラフを用いてヒストグラムを作成する方法を紹介します。

ヒストグラムとは、データの度数分布をグラフにしたものです。データがどのような分布をしているかを知るためのもっとも簡便でわかりやすいグラフです。これから紹介する方法を用いて 22 図のようなヒストグラムを作成します。

参考：統計WEB <http://software.ssri.co.jp/statweb2/> - アマチュアのための統計学講座 -

集計データの作成

1. ある統計学試験を 111 人に行った結果として得られたデータ結果：図 3 から、得点の階級別の集計とそのヒストグラムを作成します。
2. 図 4 のように度数分布表の枠組みをまず準備作成します。

	A	B	C	D
1				
2		統計学試験の得点		
3		得点		
4		87		
5		87		
6		31		
7		20		
8		66		
9		41		
10		100		
11		67		
12		92		
13		96		
14		00		

図 3 入力されたデータ

	A	B	C	D	E	F	G
1							
2		統計学試験の得点					
3		得点	階級値	階級名	度数		
4		87		0			
5		87		1			
6		31		20			
7		20		40			
8		66		60			
9		41		80			
10		100		100			
11		67			合計		
12		92					

図 4 度数分布表の枠組み

3. 「階級名」のセル範囲「E4:E10」の書式設定を行います。セル範囲「E4:E10」を選択後、右クリックメニューから「セルの書式設定」ウィンドウを開きます。[表示形式] タブで「分類」を[文字列]とし、[OK]をクリックします。文字列として設定している理由はセルの中に用いるハイフ

	A	B	C	D	E	F
1						
2		統計学試験の得点				
3		得点	階級値	階級名	度数	
4		87	0			
5		87	1			
6		31	20			
7		20	40			
8		66	60			
9		41	80			
10		100	100			
11		67		合計		
12		92				

図 5 階級を「文字」として設定



図 6 書式設定で文字列を選択

ン「-」を数式のマイナス「-」とは異なった意味で用いるためです。

4. 「階級名」のセル範囲「E4:E10」の書式設定を行います。セル範囲「E4:E10」を選択後、右クリックメニューから「セルの書式設定」ウィンドウを開きます。[表示形式]タブで「分類」を[文字列]とし、[OK]をクリックします。文字列として設定している理由はセルの中に用いるハイフン「-」を数式のマイナス「-」とは異なった意味で用いるためです。
5. 階級名を入力します。前の手順で表示形式を設定したので、「1-19」のような階級名が日付や計算などに自動変換されることはありません。
6. 各階級の度数を Excel の COUNTIF 関数を用います。COUNTIF 関数の引数は、「データ範囲」と「データを数えるための条件」です。「100」に等しいというデータの件数を数える場合、=COUNTIF(データ範囲,100)とします。「20以上」のデータの件数を数える場合、=COUNTIF(データ範囲,">= 20")とします。「20以上40未満」のデータの件数を数える場合、=COUNTIF(データ範囲,">= 20") - COUNTIF(データ範囲,">= 40")とします。このように論理式が成り立つものの個数を計算します。なお、度数の合計を「合計」の欄に計算し、元データの件数と同じ値であることを確認しておきます。この計算方法は、ある区間における度数の計算は2つの裾部分での差で実現しています。

	A	B	C	D	E	F	G
1							
2		統計学試験の得点					
3		得点	階級値	階級名	度数		
4		87	0	0			
5		87	1	1-19			
6		31	20	20-39			
7		20	40	40-59			
8		66	60	60-79			
9		41	80	80-99			
10		100	100	100			
11		67		合計			
12		92					

図 7 階級名が入りました

	A	B	C	D	E	F	G
1							
2		統計学試験の得点					
3		得点	階級値	階級名	度数		
4		87	0	0	1		
5		87	1	1-19	22		
6		31	20	20-39	13		
7		20	40	40-59	24		
8		66	60	60-79	28		
9		41	80	80-99	21		
10		100	100	100	2		
11		67		合計	111		
12		92					

図 8 度数の集計結果

つぎに縦棒グラフとして作成

7. セル範囲「E3:F10」を選択後、「グラフウィザード」ボタンをクリックします。左側の「グラフの種類」の欄で[縦棒]を選択し、右側の「形式」の欄で[集合縦棒]を選択して、[次へ]をクリックします。(9図参照)
8. [グラフの元データ]ウィンドウではこのまま[次へ]をクリックします。(10図参照)
9. [タイトルとラベル]タブで「グラフタイトル」を入力します。(図参照)



図 9 まず棒グラフを選択



図 10 ウィザードの段階



図 11 グラフオプション (凡例あり)



図 12 オプション「凡例なし」に設定

10. [凡例] タブで [凡例を表示する] のチェックを外し、[完了] をクリックします。(図参照)
この時点でグラフは図のようになります。

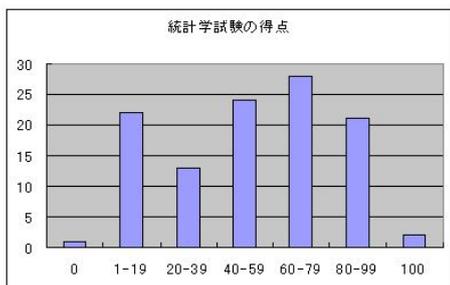


図 13 一時的なヒストグラム



図 14 棒の間隔をゼロに設定

棒の間隔を調整する

ここからは「グラフ」ツールバーを用いてグラフの設定を行うと便利です。グラフを選択しても「グラフ」ツールバーが表示されない場合、グラフを選択した状態で、メニューより [表示] [ツールバー] [グラフ] を選択してください。

11. 系列「度数」の書式設定ウィンドウを開き、[オプション] タブで「棒の間隔」に「0」を入力します。間隔調整する理由はヒストグラムの幅に意味をもたせるためです。オープンオフィスでは、棒に合わせて右クリック、オブジェクトの属性に「データ系列」「オプション」「設定」「間隔(S)」「0%」でOKをクリック

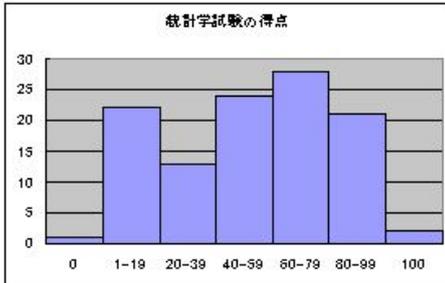


図 15 ヒストグラムの完成

A	B	C	D	E	F	G	H	I
1								
2	統計学試験の得点							
3	得点	階級値	階級名	度数	相対度数	累積相対度数		
4	87	0		1	0.009	0.009		
5	87	1	1-19	22	0.198	0.207		
6	31	20	20-39	13	0.117	0.324		
7	20	40	40-59	24	0.216	0.541		
8	66	60	60-79	28	0.252	0.793		
9	41	80	80-99	21	0.189	0.982		
10	100	100	100	2	0.018	1.000		
11	67		合計	111				
12	92							

図 16 累積度数の計算

累積相対度数を追加する方法

12. 度数分布表の右側に相対度数と累積相対度数を求めます。相対度数は各階級の度数を合計で割った値、累積相対度数は相対度数を最初の階級から順に足し上げていった値です。対応する値より小さい部分が全体のうちのどの位を占めているかを表します。
13. 累積相対度数のセル範囲「H3:H10」をコピーします。グラフの余白で右クリックし、[貼り付け] を選択します。
14. 系列「累積相対度数」の書式設定ウィンドウを開きます。[軸] タブで「表示する軸」として[第 2 軸] を選択し、[OK] をクリックします。
15. 累積相対度数の棒グラフの上で右クリックし、[グラフの種類] を選択します。左側の「グラフの種類」の欄で [折れ線] を選択し、右側の「形式」の欄で「データにマーカーが付けられた折れ線グラフ」を選択して、[OK] をクリックします。
- ヒストグラムと相対度数曲線（多角形）の完成
16. 適宜グラフの書式設定を行い、ヒストグラムの完成です。

最後にヒストグラムは階級値の値を変えるとグラフの形が変わります。階級の区間幅、個数は最大値、最小値、データの数から考えます。データ数と階級の個数の全体像をバランスよく把握することは

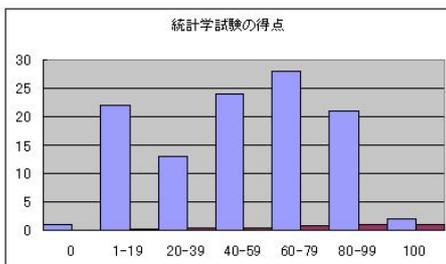


図 17 最初にグラフを貼り付け

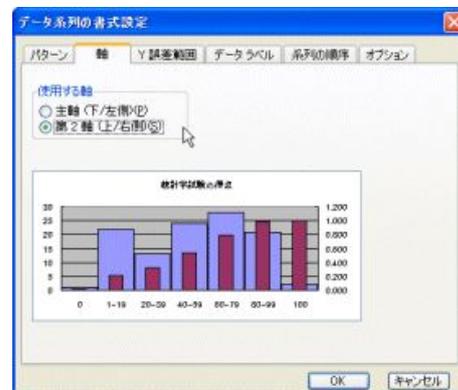


図 18 グラフを追加すると

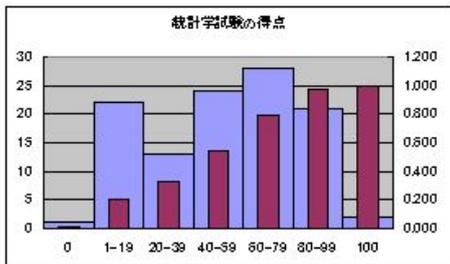


図 19 表示の方法を変えていく



図 20 折れ線グラフに変更

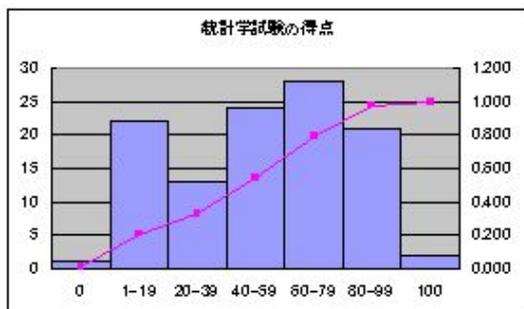


図 21 ヒストグラムと累積相対度数曲線

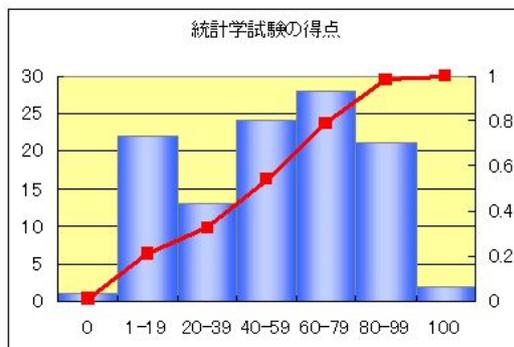


図 22 オプションで見栄えよく

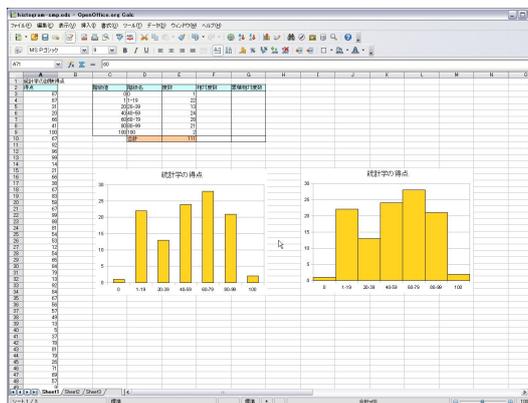


図 23 オフィスカルクでの操作 (間隔を変更)

探索的なデータ解析の基礎で、試行錯誤しながら、またまたスタージェスの公式 $k = 1 + \log_2 n$ なども考えられています。例えば、観測値の数 n を 100 とすると、スタージェスの公式では $k = 5.6$ となります。多くの場合、階級の個数 $k = 6$ くらいが見た目がよいという目安です。今回は省いていますが、本来はもっとも適切な階級値の幅とその個数を探る必要があるでしょう。

2.2 正の字で集計

【例題】観測データを整理する方法ではつぎの表も使われます。計数には項目別に集計し、正マークやタリーマーク (tally mark) が使われます。

項目	正マーク	tally mark	値
A	┆		2
B	正正下		13
C	正正		9
D	正正正正下		22
E	下		3
F	—		1
計			50

2.3 度数分布表— frequency

組み込み関数 frequency 関数をもちいる方法を説明します。ウィザードに従えばできますが、最後の OK をクリックするとき、数式配列の入力に注意します。

1. 予めデータの最大値 (max) と最小値 (min) をもとめて、データの個数から階級の幅と各階級の始まりと終わりを求める。
2. これを D13 から D17 までに入力 (連続データの作成)
3. 度数 (集計) を記入する E13 から E17 までのセルをアクティブに
4. 数式バーに=frequency と書き
5. 組み込み関数 FREQUENCY(データ配列, 区間配列) については、データ配列が A2 から A10 とし、区間配列には D13 から D17 までとする
6. ここで単に OK をクリックするのではなく、数式配列の入力 CTRL+SHIFT を押しながら OK をクリックする。

いづれの方法にしても答えは次の通りです。

	総得点の階級分け	集計	度数	比率
(1)	<200	—	1	0.11
(2)	200-299	┆	2	0.22
(3)	300-399	正正	4	0.44
(4)	400-500	┆	2	0.22
	計		9	1.00

ここで度数から比率の計算をするためには、度数の合計を”=sum(A2:A9)”で求め、このセル値を「絶対参照」してそれぞれの度数を割っていきます。これは各度数が「相対参照」され、分母が絶対参照ですから、コピーとペーストでドラックすれば計算できます。

【演習のための課題】2変数データの場合について、組み込み関数 countif と論理関数 AND を組み合わせると、2次元の度数分布表が求められます。これを考えてみて下さい。

つぎの2つ図は、「エクセル」と「オープンオフィスのカルク」でのグラフを作るそれぞれのウィザードを示しています。ほぼ同じようにして作成することができます。

カーソル, セルの移動	矢印キー、Ctrl, Shift の組み合わせ	shift+ ctrl+ arrow key
セルの選択	active, nonactive	
データの並び替え		
データの抽出		

表 1 命令



図 24 エクセルのグラフを作るウィザード



図 25 オープンオフィスのカルクでグラフを作る

グラフ作成の基本手順、項目名の入れ方、凡例の配置、グラフの配色、グラフエリアの仕上げ

3 エクセル 2003 [データ分析]

エクセル 2003 [データ分析] にはいろいろな機能が備わっています。用語が統計の教科書とは異なる部分もありますから、そのヘルプ命令からの説明に、統計理論をすこし補足しながら解説をします。

Excel のヘルプを開くと、

「分析ツールについて」Microsoft Excel には、分析ツールと呼ばれるデータ分析ツールが用意されています。これらのツールを使うと、複雑な統計や技術計算に基づく分析を簡単に実行できます。これらのツールを使う場合は、データとパラメータを指定します。各ツールでは、統計または技術計算用のマクロ関数を使ってデータを分析し、その結果を出力テーブルに表示します。出力テーブルだけでなく、グラフを出力するツールもあります。

「データ分析ツールへのアクセス」[ツール] メニューに [分析ツール] コマンドが表示されない場合は、分析ツール アドイン プログラムを読み込む必要があります。「ツール」 -> 「アドイン」 -> 「分析ツール」にチェックとすれば、コマンドに表示されます。

4 記述統計

4.1 基本統計量

基本統計量分析ツールは、入力範囲のデータを対象に、一変量による基本統計量を調べ、対象となるデータの主要な傾向と変動に関する情報を出力します。

グラフ化	範囲の選択 グラフの選択 パラメータオプション
棒グラフ	一定間隔の数値の変化や傾向を調べ、項目間を比較する。加算に意味をもつデータに適用
折れ線グラフ	データの変化の状況を表し、いくつかの傾向を同時に表示できる。とくに時間変化の傾向を調べたりする。
円グラフ	それぞれの項目間の値に対して、全体に対する割合を比較する。面積比で数値の比較
散布図 (相関図)	2 変量を縦軸と横軸にして、収集したデータをカテゴリー化しないで座標軸にプロットする。もしデータがいくつかの群に分けられているときには、グラフ内の点の記号を変えてプロットする。複数の変数観の散布図をまとめて表示するれば、相関の強弱が視覚的に捉えられる。
レーダーチャート	複数のデータ系列の構成成分を比較表示する。各系列間のバランスや描かれた多角形の面積の数値により、多次元 (多変量) データ間の比較ができる
その他のグラフ	層グラフ (連続するデータの合計と内訳の推移を表す)

表 2 グラフのいろいろ

[基本統計量] ダイアログ ボックス

入力範囲: 分析するデータ範囲のセル参照を入力します。入力範囲には、列または行単位に配置された、複数の隣接しているデータ範囲を指定する必要があります。

グループ化のキー: 入力範囲のデータが配置されている方向を指定するには、[行] または [列] をクリックします。

先頭行をラベルとして使用/先頭列をラベルとして使用 入力範囲の先頭行にラベルが入力されている場合は、[先頭行をラベルとして使用] チェック ボックスをオンにします。入力範囲の先頭列にラベルが入力されている場合は、[先頭列をラベルとして使用] チェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、チェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

平均の信頼区間の出力: 出力テーブルの行に平均の信頼区間を出力するには、このチェック ボックスをオンにし、テキスト ボックスに使用する信頼区間を入力します。たとえば、「95%」と入力すると、平均の信頼区間が有意水準 5% で計算されます。

K 番目に大きな値: 各データ グループの k 番目に大きい値を出力テーブルの行に出力するには、このチェック ボックスをオンにし、テキスト ボックスに k に相当する数値を入力します。k が 1 の場合は、この行にデータ グループの最大値が出力されます。

K 番目に小さな値: 各データ グループの k 番目に小さい値を出力テーブルの行に入力するには、このチェック ボックスをオンにし、テキスト ボックスに k に相当する数値を入力します。k が 1 の場合は、この行にデータ グループの最小値が出力されます。

結果の出力範囲を指定: 出力テーブルの左上隅のセル参照を入力します。このツールを使うと、1 つのデータ グループにつき 2 列の情報が出力されます。左の列には統計量の見出しが出力され、右の列には統計量が出力されます。選択した [グループ化のキー] オプションによって、入力範囲の各行または各列に対して、それぞれ 2 列の統計情報が出力されます。

新規のワークシート: 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を

表計算ソフトの命令		例
四則演算記号 (加減乗除)	+ , - , * , /	
比較演算記号 (大小、等値)	= , < , > , <= , >= ,	
セル番地	セルの行 (数字) と列 (英字) で指定	セルは列をアルファベットの並び順, 行を番号付けした座標形式で表す
セル範囲 (矩形領域の表現)	開始セル番地 ~ 終了セル番地	
絶対参照	列あるいは行を表す英数字の前に記号 \$ をつけて使う	相対参照は数式を作成するセルを基点として座標を指定する
複写 (コピー)	セルやセル範囲の参照を含む計算式を複写した場合, 相対的な位置関係が保たれるように参照する列、行が変更される	
ワークシート参照	別のワークシートでのセルやセル範囲を参照するには, ワークシート名と記号 ! をつける	=関数名 (引数 1, 引数 2,, 引数 n)
数式 (関数の構成)	等号で始めて, セル間の演算や組み込み関数を用いた式	
式 (論理式, 条件式)	数値、セル間の関係や数式の命題、条件を表すもの	
I F (条件式, 式 1, 式 2)	条件式が成り立つ (真) とき, 式 1 のの値を返し, そうでないとき (偽) には式 2 の値を返す	条件式は論理式ともいいかえられる
A N D (条件式 1, 条件式 2,, 条件 n)	条件式 1, 条件式 2,, 条件 n がすべて成り立つとき, 真を返し, そうでないとき偽を返す	
O R (条件式 1, 条件式 2,, 条件 n)	条件式 1, 条件式 2,, 条件 n のうち一つでも成り立つとき, 真を返し, そうでないとき偽を返す	
組み込み関数	あらかじめプログラムで登録定義されている関数	関数を利用するには, 関数名の前に等号 (=) をつけるが, 引数として使うときには等号はつけない
真 (true)	条件の命題が満たされる, あるいは成り立つとき。	
偽 (false)	条件の命題が満たされない, あるいは成り立たないとき	
等号 (=)	条件命題では等しい、同じことを意味する。さらに値や式の結果値を代入する、割り当てる場合にももちいる。	

表 3 命令の説明

起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック: 新規ブックを作成し、その新規ワークシートに結果を出力します。

[統計サマリー] 出力テーブルに、平均、(平均の) 標準誤差、メジアン、モード、標準偏差、分散、尖度、歪度、範囲、最小、最大、合計、データ数、最大値、最小値、および信頼区間の統計量をそれぞれ出力するには、このチェック ボックスをオンにします。

4.2 ヒストグラム

ヒストグラム分析ツールは、対象となるデータおよびデータ区間のセル範囲の個別頻度および累積頻度を計算します。この分析ツールを使って、データ グループに含まれる特定の値の出現頻度を求めます。

たとえば、20 人の学生を対象に、テストの点数をアルファベットで段階に分けて評価し、成績の度数分布を調べるとします。このとき、ヒストグラム テーブルには、各段階の境界値と、それぞれの段階に属する点数の頻度が表示されます。最も出現頻度の高い点数がデータのモード (最頻値) です。

[ヒストグラム] ダイアログ ボックス

入力範囲: 分析するデータ範囲のセル参照を入力します。

データ区間 (optional): データ区間を定義する境界値が入力された範囲のセル参照を入力します。境界値は、昇順で入力します。データ区間を指定した場合は、あるデータ区間と次のデータ区間の間のデータ要素の個数が数えられます。たとえば、データ区間が 2、3 となっている場合、データ 2.9999 はデータ区間 3 の要素として見なされ、データ 3.9999 はデータ区間 4 の要素として見なされます。最初のデータ区間より小さい値のデータは最初のデータ区間の要素として数えられ、最後のデータ区間より大きい値のデータは最後のデータ区間の要素として数えられます。

データ区間を省略した場合は、データの最小値と最大値の間を均等に区切ったデータ区間が自動的に作成されます。

より高度な関数	形式	用例など
I F 関数	= I F (論理式 , 値 1 , 値 2)	論理式が真の場合は値 1 を偽の場合には値 2 を返す
A N D 関数	= A N D (論理式 1 , 論理式 2 , ...)	= I F (A N D (論理式 1 , 論理式 2 , ...) , 真の場合 , 偽の場合)
O R 関数	= O R (論理式 1 , 論理式 2 , ...)	= I F (O R (論理式 1 , 論理式 2 , ...) , 真の場合 , 偽の場合)
S U M 関数	= S U M (引数 1 , 引数 2 ,)	引数で定められる数値や範囲について , 数値の合計を返す
S U M I F 関数	= S U M I F (範囲 , 検索条件 , 合計範囲)	範囲 : 検索条件で評価される対象となるセル 範囲検索条件 : 計算の対象となるセルを定義する条件 合計範囲 : 実際に計算の対象となるセルの範囲
C O U N T 関数	= C O U N T (値 1 , 値 2 ,)	
C O U N T I F 関数	= C O U N T I F (範囲 , 検索条件)	

表 4 基本的な命令 (論理と集計)

ラベル: 入力範囲の先頭行または先頭列にラベルが入力されている場合は、このチェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、このチェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

結果の出力範囲を指定: 出力テーブルの左上隅のセル参照を入力します。既存のデータが出力テーブルの内容に置き換えられる場合は、出力範囲のサイズが自動的に設定され、メッセージが表示されます。

新規のワークシート: 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック: 新規ブックを作成し、その新規ワークシートに結果を出力します。

パレート図: 頻度の高いデータ区間から順に出力テーブルに表示するには、このチェック ボックスをオンにします。チェック ボックスをオフにすると、昇順に配置された各データ区間の頻度が表示され、頻度順の表示は省略されます。

累積度数分布の表示: 出力テーブルに累積度数分布を示す列を作成し、ヒストグラムのグラフに累積度数分布を示す折れ線を追加するには、このチェック ボックスをオンにします。チェック ボックスをオフにすると、累積度数分布を示す列と折れ線は出力されません。

グラフ作成: 出力テーブルに、埋め込みのヒストグラム グラフを作成するには、このチェック ボックスをオンにします。

4.3 相関

N 個の対象物それぞれに対して各変数の測定を行う場合、CORREL ワークシート関数と PEARSON ワークシート関数は共に 2 つの測定変数間の相関係数を計算します。いずれかの対象物に対する観察が行われないと、分析時にその対象物が無視されます。相関分析ツールは、N 個の対象物それぞれに対して 3 つ以上の測定変数がある場合に特に役立ちます。この分析を行うと、測定変数の可能な組み合わせそれぞれに対して適用された CORREL (または PEARSON) 関数の値を示した相関マトリクスが、出力テーブルとして得られます。

共分散と同じように、相関係数は 2 つの測定変数と一緒に変化する範囲で測定します。共分散とは異なり、相関係数は 2 つの測定変数を表現する単位とは関係なくその値の基準が決められます。たとえば、2 つの測定変数が重量と高さの場合、重量がポンドからキログラムに変更されても相関係数の値は変わりません。相関係数のすべての値は、-1 から +1 までの範囲に収まる必要があります。

相関分析ツールを使うと、測定変数の組み合わせそれぞれについて 2 つの測定変数と一緒に変化する傾向があるかどうかを調べることができます。一方の変数の大きな値がもう一方の変数の大きな値と関連する傾向があるか (正の相関)、一方の変数の小さな値がもう一方の変数の大きな値と関連する傾向があるか (負の相関)、両方の変数の値が関連しない傾向があるか (0 に近い相関) などを調べることができます。

[相関] ダイアログ ボックス

入力範囲: 分析するデータ範囲のセル参照を入力します。入力範囲には、列または行単位に配置された、複数の隣接しているデータ範囲を指定する必要があります。

グループ化のキー: 入力範囲のデータが配置されている方向を指定するには、[行] または [列] をクリックします。

先頭行をラベルとして使用/先頭列をラベルとして使用 入力範囲の先頭行にラベルが入力されている場合は、[先頭行をラベルとして使用] チェック ボックスをオンにします。入力範囲の先頭列にラベルが入力されている場合は、[先頭列をラベルとして使用] チェック ボックスをオンにします。入力範囲にラ

ベルが入力されていない場合は、チェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

結果の出力範囲を指定: 出力テーブルの左上隅のセル参照を入力します。2 つのデータ グループの相関は処理順序とは無関係であるため、テーブルの半分だけが入力されます。また、各データ セットは相関しているため、出力テーブルの行番号および列番号が共に一致するセルには 1 が入力されます。

新規のワークシート: 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック: 新規ブックを作成し、その新規ワークシートに結果を出力します。

4.4 共分散

一連の個別の対象物に対して測定される N 個の異なる測定変数がある場合、相関分析ツールと共分散分析ツールは同じ設定で使うことができます。相関分析ツールと共分散分析ツールでは、測定変数の各組み合わせ間の相関係数または共分散をそれぞれ示すマトリクスが、出力テーブルとして得られます。相関係数が -1 から +1 までの範囲に収まるのに対し、対応する共分散はこの範囲に収まらない点があります。相関係数と共分散は共に、2 つの変数と一緒に変化する範囲で測定されます。

共分散分析ツールは測定変数のそれぞれの組み合わせについて COVAR ワークシート関数の値を計算します。たとえば N=2 の 2 つの測定変数のみの場合は、共分散分析ツールではなく COVAR 関数を直接使用する方法が適しています。共分散分析ツールの出力テーブルで対角線上の i 行と i 列の値は、それ自身の i 番目の測定変数の共分散を表します。これは、VARP ワークシート関数で計算されるその変数に対する母集団の分散の値と同じです。

共分散分析ツールを使うと、測定変数の組み合わせそれぞれについて 2 つの測定変数と一緒に変化する傾向があるかどうかを調べることができます。一方の変数の大きな値がもう一方の変数の大きな値と関連する傾向があるか (正の共分散)、一方の変数の小さな値がもう一方の変数の大きな値と関連する傾向があるか (負の共分散)、両方の変数の値が関連しない傾向があるか (0 に近い共分散) などを調べることができます。

[共分散] ダイアログ ボックス入力範囲 分析するデータ範囲のセル参照を入力します。入力範囲には、列または行単位に配置された、複数の隣接しているデータ範囲を指定する必要があります。

グループ化のキー 入力範囲のデータが配置されている方向を指定するには、[行] または [列] をクリックします。

先頭行をラベルとして使用/先頭列をラベルとして使用 入力範囲の先頭行にラベルが入力されている場合は、[先頭行をラベルとして使用] チェック ボックスをオンにします。入力範囲の先頭列にラベルが入力されている場合は、[先頭列をラベルとして使用] チェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、チェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

結果の出力範囲を指定してください 出力テーブルの左上隅のセル参照を入力します。2 つのデータグループの共分散は処理順序とは無関係であるため、テーブルの半分だけが入力されます。出力テーブルの対角線上にあるセルには、各グループの分散が入力されます。

新規のワークシート 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック 新規ブックを作成し、その新規ワークシートに結果を出力します。

4.5 順位と百分位数

順位と百分位数分析ツールは、対象となるデータ グループに含まれる各値に対して、その順位と百分位を含む出力テーブルを作成します。このツールを使って、対象となるデータ グループの中での、各値の相対的な位置を分析します。このツールは RANK ワークシート関数と PERCENTRANK ワークシート関数を使います。RANK 関数は結合値を扱いません。結合値を扱う場合は、RANK のヘルプ ファイルで説明されている補正要因と共に RANK ワークシート関数を使用します。

[順位と百分位数] ダイアログ ボックス入力範囲 分析するデータ範囲のセル参照を入力します。

グループ化のキー 入力範囲のデータが配置されている方向を指定するには、[行] または [列] をクリックします。

先頭行をラベルとして使用/先頭列をラベルとして使用 [列] をクリックし、入力範囲の先頭列にラベルが入力されている場合は、[先頭行をラベルとして使用] チェック ボックスをオンにします。[行] をクリックし、入力範囲の先頭行にラベルが入力されている場合は、[先頭行をラベルとして使用] チェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、このチェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルに行または列のラベルが自動的に作成されます。

結果の出力範囲を指定してください 出力テーブルの左上隅のセル参照を入力します。入力範囲のデータ グループ 1 つにつき、1 つの出力テーブルが作成されます。各出力テーブルは 4 つの列で構成され、データ要素の番号、値、順位、および百分位が、順位の高い順に並べられます。

新規のワークシート 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック 新規ブックを作成し、その新規ワークシートに結果を出力します。

5 シミュレーション

5.1 擬似乱数の生成

乱数発生分析ツールは、複数の分布の 1 つを基に発生させた、独立した乱数を範囲に入力します。一定の確率分布を使って母集団を特徴付けるには、この分析ツールを使います。たとえば、個人の身長の実験結果を特徴付けるには正規分布を使います。また、2 とおりの結果しか得られないコイン投げなどの結果を特徴付けるにはベルヌーイ分布を使います。

[乱数発生] ダイアログ ボックスについて

[変数の数] ボックス出力テーブルに表示される乱数の列数を入力します。入力を省略すると、指定した出力範囲のすべての列に乱数が出力されます。

[乱数の数] ボックス各変数に対して生成される乱数の個数を入力します。生成された乱数は、変数ごとに出力テーブルの 1 行に出力されます。入力を省略すると、指定した出力範囲のすべての行に乱数が出力されます。

[分布] ボックス確率変数を作成するときに使う分布方法をクリックします。

[均一] (一様分布、矩形分布ともいう) 下限と上限により特徴付けられます。範囲内のすべての値から、等しい確率で値が抽出されます。通常、0 ~ 1 の範囲の均一分布を使います。

[正規] (正規分布、平均と分散がパラメータ) 平均と標準偏差 (分散) により特徴付けられます。通常、平均が 0、標準偏差が 1 の標準正規分布を使います。

[ベルヌーイ] (2 項分布で繰り返し数が 1) 試行における成功率 (p 値) により特徴付けられます。ベ

ルヌーイ確率変数の値は、0 または 1 です。たとえば、0 ~ 1 の範囲で均一な確率変数を抽出できます。乱数の値が成功率以下の場合は、ベルヌーイ確率変数には 1 が割り当てられ、それ以外の場合は、0 が割り当てられます。

[二項] (2 項分布、繰り返し数と生起確率) 試行回数に対する成功率 (p 値) によって特徴付けられます。たとえば、試行回数分のベルヌーイ確率変数を作成すると、その合計が二項確率変数となります。

[ポワソン] (ポアソン分布、正值のパラメータ) 1/平均 (平均の逆数) に等しいパラメータ値 によって特徴付けられます。ポアソン分布は、単位時間あたりに発生する事象の数を表すためによく使われます。たとえば、一定時間内に料金所に到着する車の平均的割合はポアソン分布で表すことができます。

[パターン] (繰り返しを行なう場合) 下限と上限、ステップ数、数値の繰り返し回数、および数列の繰り返し回数によって特徴付けられます。

[離散] (一般の離散型分布、とり得る値とその確率) 値とそれに関連する確率の入力範囲によって特徴付けられます。入力範囲は、2 列で構成し、左の列には値を、右の列には同じ行の値に関連する確率を入力します。右の列の確率の合計は 1 になります。[値と確率の入力範囲] ボックス選択した分布を特徴付けるために使う値を入力します。

[ランダム シード] ボックス乱数を発生させるときの識別番号を入力します。同じ乱数を発生するようにするには、この番号をもう一度使います。[出力先] オプション ボタン、ボックス出力テーブルの左上隅のセル参照を入力します。既存のデータが出力テーブルの内容に置き換えられる場合は、出力範囲のサイズが自動的に設定され、メッセージが表示されます。

[新規又は次のワークシート] オプション ボタン、ボックス作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、右のテキスト ボックスにワークシート名を入力します。

[新規ブック] オプション ボタン新規ブックを作成し、そのブックの新規ワークシートに結果を出力します。

変数の数: 出力テーブルに表示される乱数の列数を入力します。入力を省略すると、パラメータ 選択した分布を特徴付けるために使う値を入力します。

ランダム シード: シードとは種 (seed) 乱数を発生させるときの識別番号を入力します。同じ乱数を発生するようにするには、この番号をもう一度使います。

結果の出力範囲を指定してください: 出力テーブルの左上隅のセル参照を入力します。既存のデータが出力テーブルの内容に置き換えられる場合は、出力範囲のサイズが自動的に設定され、メッセージが表示されます。

新規のワークシート: 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック 新規ブックを作成し、その新規ワークシートに結果を出力します。

5.2 サンプリング

サンプリング分析ツールは、入力範囲のデータを母集団と見なし、母集団から標本を抽出します。母集団の規模が大きすぎて統計処理やグラフ化が困難な場合は、母集団全体ではなく標本を対象に分析すると便利です。また、入力データにある種の周期性が認められる場合は、周期の特定部分から一定数の標本を抽出することもできます。たとえば、入力範囲に四半期ごとの売上高が入力されている場合、周期を 4 に設定すると、同じ四半期の売上高だけを出力テーブルに抽出できます。

[サンプリング] ダイアログ ボックス入力範囲: サンプリングの対象となる母集団が入力されている

データ範囲のセル参照を指定します。第 1 列、第 2 列の順に、標本が抽出されます。

ラベル: 入力範囲の先頭行または先頭列にラベルが入力されている場合は、このチェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、このチェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

標本の採取方法: [周期変化] または [ランダム] をクリックし、標本を抽出する方法を指定します。

期間: 標本を抽出する周期を入力します。入力範囲に含まれるデータのうち、ここで指定した周期に当たるデータが出力テーブルの 1 列にコピーされます。入力範囲の末尾に達すると標本の抽出は終了します。

サンプル数: 出力テーブルの列に抽出する値の個数を入力します。標本は、入力範囲の任意の位置から抽出されます。抽出された数値は、次に抽出を行うときも対象となります。

結果の出力範囲を指定: 出力テーブルの左上隅のセル参照を入力します。抽出された標本は、そのセルの 1 列下に出力されます。[周期変化] をクリックすると、出力テーブルに抽出される値の個数は、入力範囲のデータの個数を周期で割った値に等しくなります。また、[ランダム] をクリックすると、出力テーブルに抽出される値の個数は、標本数に等しくなります。

新規のワークシート: 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック: 新規ブックを作成し、その新規ワークシートに結果を出力します。

5.3 順位と百分位数

順位は、データを大きさの順に並べます。昇順と降順があります。また同数値に対して、順位を定めるために、中間順位 (mid rank) という修正を施す場合もあります。

[順位と百分位数] ダイアログ ボックスについて [入力範囲] ボックス分析するデータ範囲のセル参照を入力します。

[データ方向] グループ入力範囲のデータが配置されている方向を指定するには、[列] または [行] をクリックします。

[先頭行をラベルとして使用]、[先頭列をラベルとして使用] チェック ボックス入力範囲の先頭行にラベルが入力されている場合は、[列] をクリックし、[先頭行をラベルとして使用] チェック ボックスをオンにします。入力範囲の先頭列にラベルが入力されている場合は、[行] をクリックし、[先頭列をラベルとして使用] チェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、チェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

[出力先] オプション ボタン、ボックス出力テーブルの左上隅のセル参照を入力します。入力範囲のデータ グループ 1 つにつき、1 つの出力テーブルが作成されます。各出力テーブルは 4 つの列で構成され、データ要素の番号、値、順位、および百分位が、順位の高い順に並べられます。

[新規又は次のワークシート] オプション ボタン、ボックス作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、右のテキスト ボックスにワークシート名を入力します。

[新規ブック] オプション ボタン新規ブックを作成し、そのブックの新規ワークシートに結果を出力します。

6 検定

統計学の推定や検定はもっとも重要な働きをするものです。ここでは3つの検定を述べますが、名前の由来は、検定にもちいる統計量の分布からきています。いろいろ場合について同じ検定統計量を用いることがあります。

6.1 F 検定

F 検定：2 標本を使った分散の検定分析ツールは、2 つの標本を使った F 検定を行って、2 つの母集団の分散を比較します。

たとえば、2 つのチームそれぞれの水泳大会でのタイムの標本について F 検定ツールを使用することができます。このツールを使用すると、分散が基となる分布に等しくないという対立仮説に対して等しい分散の分布からこれら 2 つの標本が得られるという帰無仮説に対する検定の結果を求めることができます。

このツールは F 統計 (または F 比率) の値を計算します。 f の値が 1 に近い場合は基になる母集団の分散が等しいことを示します。出力テーブルでは、 $f < 1$ の場合、“ $P(F \leq f)$ 片側” で、母集団の分散が等しいときに f より小さい F 統計の値が測定される確率が求められ、“F 境界値 片側” で、選択した基準値のレベル α に対して 1 未満の境界値が求められます。 $f > 1$ の場合は、“ $P(F \leq f)$ 片側” で、母集団の分散が等しいときに f より大きい F 統計の値が測定される確率が求められ、“F 境界値 片側” で、選択したレベル α に対して 1 を超える境界値が求められます。

[F 検定: 2 標本を使った分散の検定] ダイアログ ボックス変数 1 の入力範囲 分析するデータ範囲の 1 番目の列または行の参照を入力します。

変数 2 の入力範囲 分析するデータ範囲の 2 番目の列または行の参照を入力します。

ラベル 入力範囲の先頭行または先頭列にラベルが入力されている場合は、このチェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、このチェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

検定の有意水準 α を指定：0 ~ 1 の範囲の値で入力します。 レベルは、真の仮説を棄却するという I タイプの誤差を犯す確率にもので、これを有意水準といいます。 $\alpha = 0.01, 0.05, 0.10$ などとします。

結果の出力範囲を指定：出力テーブルの左上隅のセル参照を入力します。既存のデータが出力テーブルの内容に置き換えられる場合は、出力範囲のサイズが自動的に設定され、メッセージが表示されます。

新規のワークシート：作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック：結果を新たなブックに書き込みたい場合にチェックします。新規ブックを作成し、その新規ワークシートに結果を出力します。

6.2 t 検定

2 標本による t 検定分析ツールは、各標本の基になる母集団の平均の等しさを検定します。母集団の分散が等しいという仮定、母集団の分散が等しくないという仮定、および 2 つの標本が同じ対象物についての処理の前と後の測定を表すという仮定、という 3 つの異なる仮定に 3 つのツールを使用します。

以下の 3 つのツールすべてにおいて、t 統計値 t が計算されて出力テーブルに “t Stat” として示されます。データに応じてこの t という値は負の値にも正の値にもなります。基になる母集団の平均値が等しいとする仮定では、 $t < 0$ の場合、“ $P(T \leq t)$ 片側” で、 t より小さい t 統計の値が測定される確率が求められます。 $t \geq 0$ の場合は、“ $P(T \leq t)$ 片側” で、 t より大きい t 統計の値が測定される確率が求め

られます。"t 境界値 片側" で、"t 境界値 片側" より大きいか等しい t 統計の値が測定される確率が α になるような区分値が求められます。

"P(T ≤ t) 両側" で、絶対値が t よりも大きい t 統計の値が測定される確率が求められます。"P 境界値 両側" で、絶対値が "P 境界値 両側" より大きい t 統計の値が測定される確率が α になるような区分値が求められます。

t 検定: 一对の標本による平均の検定 1 つの標本グループをある実験の前後で 2 度検定する場合のように、2 つの観測値に自然な対の関係がある場合は、この形式の t 検定を使います。この分析ツールとその数式は、一对の標本を使ったスチューデントの t 検定を行い、処理の前に行われた測定と処理の後に行われた測定が、母集団の等しい平均値の分布から求められたかどうかを調べます。この形式の t 検定では、データ グループを抽出した 2 つの母集団の分散が等しいことを前提としません。

メモ この分析ツールを使って出力する結果には、次の数式で求められる、プールされた分散 (結果に関するデータの分散がプールされた量) が現れることもあります。

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \quad (6.1)$$

[t 検定: 一对の標本による平均の検定] ダイアログ ボックス変数 1 の入力範囲 分析する 1 番目のデータ範囲のセル参照を入力します。入力範囲には 1 列または 1 行を指定し、また、2 番目の範囲と同数のデータ要素を入力する必要があります。

変数 2 の入力範囲 分析する 2 番目のデータ範囲のセル参照を入力します。入力範囲には 1 列または 1 行を指定し、また、1 番目の範囲と同数のデータ要素を入力する必要があります。

二標本の平均値の差 標本の平均値の変動として使う数値を入力します。0 (ゼロ) は、標本の仮説平均が等しいことを示します。

ラベル 入力範囲の先頭行または先頭列にラベルが入力されている場合は、このチェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、このチェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

検定の有意水準を、0 ~ 1 の範囲で入力します。 レベルは、真の仮説を拒絶するという I タイプのエラーの発生率に関連する有意水準です。

結果の出力範囲を指定してください 出力テーブルの左上隅のセル参照を入力します。既存のデータが出力テーブルの内容に置き換えられる場合は、出力範囲のサイズが自動的に設定され、メッセージが表示されます。

新規のワークシート 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック 新規ブックを作成し、その新規ワークシートに結果を出力します。

t 検定: 等分散を仮定した 2 標本による検定 2 つの標本を使ったスチューデントの t 検定を行います。この形式の t 検定では、2 つのデータ セットが同じ分散の分布から求められると仮定するため、等分散 t 検定とも呼ばれます。2 つの標本が母集団の等しい平均値の分布から求められるかどうかを調べるためにこの t 検定を使います。

[t 検定: 等分散を仮定した 2 標本による検定] ダイアログ ボックス変数 1 の入力範囲 分析する 1 番目のデータ範囲のセル参照を入力します。入力範囲には、データが入力されているワークシートの 1 列または 1 行を指定する必要があります。

変数 2 の入力範囲 分析する 2 番目のデータ範囲のセル参照を入力します。入力範囲には、データが入力されているワークシートの 1 列または 1 行を指定する必要があります。

二標本の平均値の差 標本の平均値の変動として使う数値を入力します。0 (ゼロ) は、標本の仮説平均が等しいことを示します。

ラベル 入力範囲の先頭行または先頭列にラベルが入力されている場合は、このチェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、このチェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

検定の有意水準を、0 ~ 1 の範囲で入力します。 レベルは、真の仮説を拒絶するという I タイプのエラーの発生率に関連する有意水準です。

結果の出力範囲を指定してください 出力テーブルの左上隅のセル参照を入力します。既存のデータが出力テーブルの内容に置き換えられる場合は、出力範囲のサイズが自動的に設定され、メッセージが表示されます。

新規のワークシート 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック 新規ブックを作成し、その新規ワークシートに結果を出力します。

t 検定: 分散が等しくないと仮定した 2 標本による検定 2 つの標本を使ったスチューデントの t 検定を行います。この形式の t 検定では、2 つのデータ グループが、等しくない分散の分布から求められると仮定するため、異分散 t 検定とも呼ばれます。前述の等分散の場合と同じように、この t 検定は、2 つの標本が母集団の等しい平均値の分布から求められるかどうかを調べるために使用することができます。2 つの標本に他とははっきり異なる対象物がある場合は、こちらの検定方法を使います。単一の連続の対象物と各対象物の処理の前と後の測定を表す 2 つの標本がある場合は、次の例で説明する一対の標本による平均の検定を使います。

統計定数 t を決めるには、次の数式を使います。

$$t' = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}} \quad (6.2)$$

自由度 df を計算するには、次の数式を使用します。通常は計算結果が整数ではないため、t テーブルからの境界値を求める際に、df の値が最も近い整数に四捨五入されます。Excel のワークシート関数 TTEST は、整数以外の df 値で TTEST の値を計算できるため、計算された df 値が四捨五入されずそのまま使われます。自由度を決める方法のこの違いのために、異分散における TTEST とこの t 検定ツールの結果は異なります。

$$df = \frac{\left(\frac{S_1^2}{m} + \frac{S_2^2}{n}\right)^2}{\frac{(S_1^2/m)^2}{m-1} + \frac{(S_2^2/n)^2}{n-1}} \quad (6.3)$$

[t 検定: 分散が等しくないと仮定した 2 標本による検定] ダイアログ ボックス変数 1 の入力範囲 分析する 1 番目のデータ範囲のセル参照を入力します。入力範囲には、データが入力されているワークシートの 1 列または 1 行を指定する必要があります。

変数 2 の入力範囲 分析する 2 番目のデータ範囲のセル参照を入力します。入力範囲には、データが入力されているワークシートの 1 列または 1 行を指定する必要があります。

二標本の平均値の差 標本の平均値の変動として使う数値を入力します。0 (ゼロ) は、標本の仮説平均が等しいことを示します。

ラベル 入力範囲の先頭行または先頭列にラベルが入力されている場合は、このチェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、このチェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

検定の有意水準を、0 ~ 1 の範囲で入力します。 レベルは、真の仮説を拒絶するという I タイプのエラーの発生率に関連する有意水準です。

結果の出力範囲を指定してください 出力テーブルの左上隅のセル参照を入力します。既存のデータが出力テーブルの内容に置き換えられる場合は、出力範囲のサイズが自動的に設定され、メッセージが表示されます。

新規のワークシート 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック 新規ブックを作成し、その新規ワークシートに結果を出力します。

6.3 z 検定

z 検定: 2 標本による平均の検定分析ツールは、既知の分散を使って、2 つの標本を対象に平均値の z 検定を行います。この分析ツールを使って、片面または両面の対立仮説いずれかに対して 2 つの母集団の平均値の間に有意な差がないという帰無仮説を検定します。分散が未知の場合は、代わりに ZTEST ワークシート関数を使う必要があります。

z 検定分析ツールを使う場合は、出力を詳しく調べる必要があります。" $P(Z \leq z)$ 片側" は実際には $P(Z \geq ABS(z))$ を表し、母集団の平均値に差がない場合、z 値の確率は測定される z 値と同じ方向に 0 から遠い値になります。" $P(Z \leq z)$ 両側" は実際には $P(Z \geq ABS(z) \text{ or } Z \leq -ABS(z))$ を表し、母集団の平均値に差がない場合、z 値の確率は測定される z 値のいずれかの方向に 0 から遠い値になります。両側の結果は、片側のちょうど 2 倍の値になります。z 検定分析ツールは、2 つの母集団の平均値の差が 0 以外の特定の値であるという仮説が帰無仮説となる場合に使用することもできます。たとえば、z 検定を行うと、2 種類の車の性能の差異を検定できます。

[z 検定: 2 標本による平均の検定] ダイアログ ボックス変数 1 の入力範囲 分析する 1 番目のデータ範囲のセル参照を入力します。入力範囲には、データが入力されているワークシートの 1 列または 1 行を指定する必要があります。

変数 2 の入力範囲 分析する 2 番目のデータ範囲のセル参照を入力します。入力範囲には、データが入力されているワークシートの 1 列または 1 行を指定する必要があります。

二標本の平均値の差 標本の平均値の変動として使う数値を入力します。0 (ゼロ) は、標本の仮説平均が等しいことを示します。

変数 1 の分散 (既知) 変数 1 の入力範囲の母集団の、既知の分散を入力します。

変数 2 の分散 (既知) 変数 2 の入力範囲の母集団の、既知の分散を入力します。

ラベル 入力範囲の先頭行または先頭列にラベルが入力されている場合は、このチェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、このチェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

検定の有意水準を、0 ~ 1 の範囲で入力します。 レベルは、真の仮説を拒絶するという I タイプのエラーの発生率に関連する有意水準です。

結果の出力範囲を指定してください 出力テーブルの左上隅のセル参照を入力します。既存のデータが出力テーブルの内容に置き換えられる場合は、出力範囲のサイズが自動的に設定され、メッセージが表示されます。

新規のワークシート 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック 新規ブックを作成し、その新規ワークシートに結果を出力します。

7 分散分析

分散分析ツールでは、さまざまな種類の分散分析 (Analysis Variance) を行います。使用するツールは、検定する母集団の要因数と標本数によって選びます。

7.1 分散分析: 一元配置

このツールは、複数の標本のデータについて分散の簡単な分析を行います。この分析を行うと、すべての標本に関して基になる確率分布が同じではないという対立仮説に対して同じ確率分布からそれぞれの標本が得られるという仮説を検定することができます。標本が 2 つのみの場合は、TTEST ワークシート関数を使用できます。3 つ以上の標本については TTEST 関数は利用できませんが、代わりに一元配置分散分析モデルを使うことができます。

[分散分析: 一元配置] ダイアログ ボックス

入力範囲: 分析するデータ範囲のセル参照を入力します。入力範囲には、列または行単位に配置された、複数の隣接しているデータ範囲を指定する必要があります。

グループ化のキー: 入力範囲のデータが配置されている方向を指定するには、[行] または [列] をクリックします。

先頭行をラベルとして使用/先頭列をラベルとして使用 入力範囲の先頭行にラベルが入力されている場合は、[先頭行をラベルとして使用] チェック ボックスをオンにします。入力範囲の先頭列にラベルが入力されている場合は、[先頭列をラベルとして使用] チェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、チェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

α : F 統計量の臨界値を評価するための水準で、0 と 1 の間の数です。 α レベルは、真の仮説を棄却するという I タイプの誤り確率に関連する有意水準です。

結果の出力範囲を指定してください

出力テーブルの左上隅のセル参照を入力します。既存のデータが出力テーブルの内容に置き換えられる場合、または出力テーブルがワークシートの範囲より大きい場合は、出力範囲のサイズが自動的に設定され、メッセージが表示されます。

新規のワークシート 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック 新規ブックを作成し、その新規ワークシートに結果を出力します。

7.2 分散分析: 繰り返しのある二元配置

この分析ツールは、データを 2 つの異なる次元に従って分類できる場合に役立ちます。たとえば植物の高さを計測する実験では、その植物に A、B、C など異なる肥料を与え、低温、高温などの異なる温度環境に置く場合があります。肥料、温度の 6 種類の組み合わせそれぞれについて、植物の高さを同じ回数観察します。この分散分析ツールを使うと、以下の検定を行うことができます。

同じ母集団に異なる種類の肥料を与えた場合の植物の高さを測定します。この分析方法では温度環境を無視します。同じ母集団を異なる温度環境に置いた場合の植物の高さを測定します。この分析方法で

は肥料の種類を無視します。箇条書きの最初の肥料の種類と 2 つ目の温度環境の違いによる影響について考慮したことにより、同じ母集団から 肥料、温度 値のすべての組み合わせに対応する 6 つの標本が得られます。この場合の対立仮説は、肥料の種類単独または温度環境単独における違いに対して、肥料、温度 の特定の組み合わせによる影響があるということです。

	グループ 1	グループ 2
試験 1	75	58
	68	56
	71	61
	75	60
試験 2	66	62
	70	60
	68	59
	68	68

[分散分析 繰り返しのある一元配置] ダイアログ ボックス

入力範囲: 分析するデータ範囲のセル参照を入力します。入力範囲には、列または行単位に配置された、複数の隣接しているデータ範囲を指定する必要があります。

1 標本あたりの行数: 1 標本あたりの行数を入力します。各行はデータの繰り返しを表すので、標本の行数にはすべて同じ数値を入力します。

: F 統計量の臨界値を評価するための水準を入力します。レベルは、真の仮説を拒絶するという I タイプのエラーの発生率に関連する有意水準です。

結果の出力範囲を指定: 出力テーブルの左上隅のセル参照を入力します。既存のデータが出力テーブルの内容に置き換えられる場合、または出力テーブルがワークシートの範囲より大きい場合は、出力範囲のサイズが自動的に設定され、メッセージが表示されます。

新規のワークシート: 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック: 新規ブックを作成し、その新規ワークシートに結果を出力します。

[分散分析 繰り返しのない二元配置] ダイアログ ボックス

この分析ツールは、繰り返しのある二元配置の場合と同じように、データを 2 つの異なる次元で分類する場合に役立ちます。ただしこのツールでは、1 つの組み合わせ (たとえば、前の例の場合は、肥料、温度 の組み合わせの 1 つ) について、1 つの観察のみを行うと仮定します。

[分散分析: 繰り返しのない一元配置] ダイアログ ボックス入力範囲 分析するデータ範囲のセル参照を入力します。入力範囲には、列または行単位に配置された、複数の隣接しているデータ範囲を指定する必要があります。

ラベル入力範囲にラベルが入力されていない場合は、このチェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

F 統計量の臨界値を評価するための水準を入力します。レベルは、真の仮説を拒絶するという I タイプのエラーの発生率に関連する有意水準です。

結果の出力範囲を指定してください 出力テーブルの左上隅のセル参照を入力します。既存のデータが出力テーブルの内容に置き換えられる場合、または出力テーブルがワークシートの範囲より大きい場合は、出力範囲のサイズが自動的に設定され、メッセージが表示されます。

新規のワークシート 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック 新規ブックを作成し、その新規ワークシートに結果を出力します。

8 予測のための統計

8.1 回帰分析

回帰分析ツールは、線形回帰分析を行います。回帰分析では、R-2 乗値を使って、観測値のデータが最適な直線に当てはめられます。このツールを使って、複数の独立変数が 1 つの従属変数に与える影響を分析することができます。たとえば、スポーツ選手の年齢、身長、体重などの要素が成績に与える影響を分析できます。成績データに基づいて、これらの要素それぞれが成績に影響した比率を割り当てたり、回帰分析の結果を使って、他のスポーツ選手の成績を予測することもできます。

回帰分析ツールは LINEST ワークシート関数を使用します。

[R 回帰] ダイアログ ボックス

入力 Y 範囲: 分析するデータ範囲のうち、従属変数のデータ範囲のセル参照を入力します。従属変数のデータ範囲には、データが入力されている 1 列を指定する必要があります。

入力 X 範囲: 分析するデータ範囲のうち、独立変数のデータ範囲のセル参照を入力します。このデータ範囲の独立変数は、左から右へ昇順で並べられます。指定できる独立変数の最大個数は 16 です。

ラベル: 入力範囲の先頭行または先頭列にラベルが入力されている場合は、このチェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、このチェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

信頼度: 一覧の出力テーブルに有意水準を出力するには、このチェック ボックスをオンにします。既定の 95% 以外の値を使うには、設定する有意水準を右のテキスト ボックスに入力します。

定数に 0 を使用: 原点を通る回帰直線を作成するには、このチェック ボックスをオンにします。

結果の出力範囲を指定: 出力テーブルの左上隅のセル参照を入力します。一覧の出力テーブルには、分散分析テーブル、係数、y の予測値の標準誤差、R-2 乗値、観測値の個数、および係数の標準誤差が表示されるので、少なくとも 7 列必要です。

新規のワークシート: 作業中のブックに新規ワークシートを挿入し、新規ワークシートのセル A1 を起点として結果を出力します。新規ワークシートに名前を付けるには、テキスト ボックスにワークシート名を入力します。

新規ブック: 新規ブックを作成し、その新規ワークシートに結果を出力します。

残差: 残差の出力テーブルに残差を表示するには、このチェック ボックスをオンにします。

標準化された残差: 残差の出力テーブルに標準化された残差を表示するには、このチェック ボックスをオンにします。

残差グラフの作成: 残差に対する独立変数のグラフを作成するには、このチェック ボックスをオンにします。

観測値グラフの作成: 観測値に対する予測値のグラフを作成するには、このチェック ボックスをオンにします。

正規確率グラフの作成: 正規確率のグラフを作成するには、このチェック ボックスをオンにします。

8.2 移動平均

移動平均分析ツールは、一定数の過去の観測値の平均値に基づき、将来の値を予測します。移動平均を使うと、過去のデータを平均するだけではわからない傾向についての情報を予測できます。売り上げや棚卸しなどの傾向を予測するには、この分析ツールを使います。予測値は、次の数式を基に計算されます。

$$F_{(t+1)} = \frac{1}{N} \sum_{j=1}^N NA_{t-j+1} \quad (8.1)$$

ここで N は、移動平均の計算に使う過去の観測値の個数です。 A_j は、 j 回目の観測値です。 F_j は、 j 回目の予測値です。

[移動平均] ダイアログ ボックス

入力範囲：分析するデータ範囲のセル参照を入力します。入力範囲には、データを入力したセルが 4 つ以上ある列を 1 つ指定します。

先頭行をラベルとして使用：入力範囲の先頭行にラベルが入力されている場合は、このチェック ボックスをオンにします。入力範囲にラベルが入力されていない場合は、このチェック ボックスをオフにします。チェック ボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

間隔：移動平均に含まれる区間数を入力します。既定の区間数は 3 です。

結果の出力範囲を指定してください 出力テーブルの左上隅のセル参照を入力します。[標準誤差] チェック ボックスをオンにすると、2 列の出力テーブルが作成され、右の列に標準誤差の値が出力されます。過去のデータが不足しているために、データの予測や標準誤差の計算ができない場合は、エラー値 #N/A が返されます。

出力範囲は、入力範囲に使用するデータと同じワークシートに指定します。このため、[新規又は次のワークシート] および [新規ブック] オプションは使用できません。

グラフ作成：出力テーブルに、埋め込みのヒストグラム グラフを作成するには、このチェック ボックスをオンにします。

標準誤差：出力テーブルの列に標準誤差の値を出力するには、このチェック ボックスをオンにします。標準誤差の値を出力しないで 1 列の出力テーブルを出力するには、このチェック ボックスをオフにします。

8.3 指数平滑

指数平滑分析ツールは、前回の観測値と予測値との誤差を調整することにより、前回の予測値に基づいてデータを予測します。平滑化定数 a を使って計算します。この定数は、前回の予測値の誤差に対する予測値の反応の強さを決める値です。

メモ：平滑化定数に指定する値は 0.2 ~ 0.3 が適しています。この範囲の値を使うと、現在の予測が前回の予測の 20% ~ 30% の誤差に調整されます。定数に大きい値を指定すると、予測値の計算に必要な時間は短くなりますが、予測の誤差は大きくなります。定数に小さい値を指定すると、予測値の計算に必要な時間が長くなります。

[指数平滑] ダイアログ ボックス

入力範囲：分析するデータ範囲のセル参照を入力します。入力範囲には、データが入力されている 4 つ以上のセルを含む 1 列、または 1 行を指定する必要があります。

減衰率：指数平滑化定数として使用する減衰率を入力します。減衰率は、母集団から採取したデータの不安定性を最小限にする補正係数です。既定の減衰率は 0.3 です。

メモ：平滑化定数に指定する値は 0.2 ~ 0.3 が適しています。この範囲の値を使うと、現在の予測が前回の予測の 20% ~ 30% の誤差に調整されます。定数に大きい値を指定すると、予測値の計算に必要な時間は短くなりますが、予測の誤差は大きくなります。定数に小さい値を指定すると、予測値の計算に必要な時間が長くなります。

ラベル：入力範囲の先頭行または先頭列にラベルが入力されている場合は、このチェックボックスをオンにします。入力範囲にラベルが入力されていない場合は、このチェックボックスをオフにします。チェックボックスをオフにすると、出力テーブルにラベルが自動的に作成されます。

結果の出力範囲を指定：出力テーブルの左上隅のセル参照を入力します。[標準誤差] チェックボックスをオンにすると、2 列の出力テーブルが作成され、右の列に標準誤差の値が出力されます。過去のデータが不足しているために、データの予測や標準誤差の計算ができない場合は、エラー値 #N/A が返されます。

メモ：出力範囲は、入力範囲に使用するデータと同じワークシートに指定します。このため、[新規又は次のワークシート] および [新規ブック] オプションは使用できません。

グラフ作成：観測値と予測値を使って、出力テーブルに埋め込みグラフを作成するには、このチェックボックスをオンにします。

標準誤差：出力テーブルの列に標準誤差の値を出力するには、このチェックボックスをオンにします。標準誤差の値を出力しないで 1 列の出力テーブルを出力するには、このチェックボックスをオフにします。