

情報科学演習
第12回
表計算ソフトを用いた統計処理

目次

1	本日の目標	1
2	本日の実習	1
2.1	練習問題 1	3
2.1.1	テキストファイルを Excel のデータに取り込む	3
2.1.2	成績表の作成	4
2.2	練習問題 2	5
2.3	ファイルの保存	5
2.4	発展問題	6
3	資格試験の案内	6
4	付録：FFFTP の設定の確認方法	7

1 本日の目標

- 分散, 標準偏差など統計の基本用語の定義と意味を知る.
- テキストファイルで書かれた表計算のデータを Excel で読む方法について知る.
- 表計算のデータをテキストファイルで保存する方法について知る.

前回に引き続き, 表計算ソフトの使用法を学びます. 今回は成績処理に関するデータの扱いを例に統計の基本用語と, その定義について学習します.

実習に入る前に Excel を起動して下さい.

2 本日の実習

としひこ, せいこ, まさひこ, なおこ, いよ, よしおの6人がある試験でそれぞれ, 3点, 4点, 8点, 10点, 7点, 5点を取ったとします. これらのデータをもとに, 平均点と各人の偏差値を計算します. 偏差値は標準偏差と平均を使って次の式で定義されます.

	A	B	C	D	E	F	G
1	7月21日	成績表					
2							
3		名前	得点	平均との差	その2乗	得点の2乗	偏差値
4		としひこ	3				
5		せいこ	4				
6		まさひこ	8				
7		なおこ	10				
8		いよ	7				
9		よしお	5				
10		平均					
11		分散					
12		標準偏差					

図 1:

$$\frac{\text{素点} - \text{平均}}{\text{標準偏差}} \times 10 + 50$$

標準偏差は字から推測されるように, 偏差 (平均値からの偏り) の平均です. 正確には次のように, 分散の平方根として定義されます: n 人の人の点数が, $x_1, x_2, \dots, x_n$, その平均が $\bar{x}$ とすると,

$$\begin{aligned} \text{平均: } \bar{x} &= \frac{x_1 + \dots + x_n}{n} \\ \text{分散} &= \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} = \frac{x_1^2 + \dots + x_n^2}{n} - \left(\frac{x_1 + \dots + x_n}{n} \right)^2 \\ \text{標準偏差} &= \sqrt{\text{分散}} \end{aligned}$$

となります。分散の式の 2 番目の等式は簡単に証明できるので、証明してみてください。

なお、偏差値は受験用語で数学用語ではありませんが、分散、標準偏差は、医学や工学など実験系はもちろん、経済学、社会学、教育学などデータ分析をするときには必要とされる基本用語です。定義も簡単なので、この機会に覚えて下さい。

次の指示に従い、図 1 の表を完成させます。

1. 図 1 にあるデータおよび項目名を入力して下さい。
2. 次に、関数 AVERAGE を使用して、セル C10 に 6 人の成績の平均値を計算する式を書きます。
3. D4 に「としひこ」の得点 (C4) と 6 人の平均点 (C10) の差を式で入力します。

この時、「=C4-C10」と入力してしまうと、D5 にこの式をコピーした時に「=C5-C11」が入力され、本来求める値とは異なる計算結果になります。(このようなセルの参照を「相対参照」といいます。)

これに対して、平均点の記述されたセル (C10) のようにどのセルからもそのセルの値を共通に利用したい場合、「絶対参照」という方法を用います。絶対参照をするためには、セルの行番号と列のアルファベットの前に、ドル記号 (\$) を付けます。例えば、セル C10 を絶対参照するためには、\$C\$10 とします。他にも「複合参照」という方法がありますが、こちらは自習して下さい。

絶対参照を利用して、表を完成させます。

1. D4～D9 に各人の得点 (C 列) と平均の差を式で入力します。
2. E4～E9 には D 列の 2 乗を式で入力します。
3. F4～F9 には C 列の 2 乗を式で入力します。
4. F10 には「各人の得点の 2 乗」の平均を入力します。(F4～F9 の平均を計算する式を入力。)
5. C11 に 6 人の成績の分散を入力します。(分散は「2 乗の平均-平均の 2 乗」ですから、F10 から C10 の 2 乗を引いた式を書くことになります。)
6. E10 に E4～E9 の平均を計算する式を入力します。(この値は「各人の得点から平均点を引いたもの」の 2 乗ですから、分散の定義式です。C11 の値と一致することを確認して下さい。)
7. C12 に標準偏差を入力します。平方根を求めるには、SQRT という関数を利用します。)
8. G4～G9 に各人の偏差値を計算する式を入力します。
9. D10 に D4～D9 の平均を計算する式を書きます。(この値は、理論上 0 と表示される筈ですが、 $x.xxxx E-10$ のように表示されることがあります。これは、 $x.xxxx \times 10^{-10}$ の意味で 0 に非常に近い値です。小数の計算においては、計算機は無限小数や非常に小さい数を途中で値を四捨五入するため理論値との誤差を生じることがあります。)

終わったらファイルを閉じます。(後で使用することはないので、保存しなくても構いません。ひき続いて Excel を使うので、Excel は終了しないで下さい。)

	A	B	C	D	E	F
1	7月21日					
2						
3		全体	99年度生	00年度生	01年度生	
4	平均	59.555556				
5	分散	294.158				
6	標準偏差	17.1510				
7						
8	No.	学籍番号	文理	得点	得点の2乗	偏差値
9	1	993101	理	86	7396	65.42
10	2	993102	文	65	4225	53.17
11	3	993103	文	46	2116	42.10

図 2:

2.1 練習問題 1

例題の方法を参考に図 2 のような表を作成してみましょう。

図では省略されていますが、これは 45 人の成績からなるデータで、元となるデータは数理科学科の FTP サーバにありますから、ファイルの取り寄せます。FFFTP を起動して、最後の節に書かれた「FFFTP の設定の確認方法」を参照して転送の設定を確認して下さい。

確認が済んだら、以下のファイルを取り寄せます。

ディレクトリ: /pub/joho/

ファイル名: seiseki-dos.txt

取り寄せたファイルは、通常、「マイ ドキュメント」に保存されます。取り寄せた筈なのに、「マイ ドキュメント」にファイルが見つからない人は、検索を行なって下さい。

2.1.1 テキストファイルを Excel のデータに取り込む

取り寄せたファイルを note pad で見て下さい。このファイルを Excel のデータに変換します。

- Excel の「ファイル」メニューから「開く」を選びます。
- 「ファイルを開く」のウィンドウが現れます。「ファイルの種類」が「すべてのファイル (*.*)」となっていることを確認します。(そうでない場合は、この欄を変更して「すべてのファイル (*.*)」となるように変更します。)
- 「マイ ドキュメント」から seiseki-dos.txt を選び、「開く」をクリックします。
- 「テキストファイルウィザード」というウィンドウがポップアップするので、その中の「元のデータの形式」という枠内の項目のうち、「カンマやタブで区切られた文字によってフィールドごとに区切られたデータ」がチェックされているかどうかを確認する。(他の項目がチェックされている場合には、この項目をチェックするように変更する。)

5. 取込み開始行が「1」であることを確認します。
6. 「元のファイル」の欄を「932:Japanese (Shift-JIS)」に変更する。(右側の下向き三角をクリックし、さらに表示されるメニューの右側の下向き三角をクリックすると、ABC 順にコードの形式が表示されるので、その中から選ぶ)
7. 「データのプレビュー」が、メモ帳で見たものと同じ表示になっているのを確認して、「次へ」ボタンをクリック
8. 「フィールドと区切りの文字の指定」において、「タブ」がチェックされていることを確認する。(他の項目がチェックされている場合には、「タブ」をチェックするように変更する。)
9. 「次へ」ボタンをクリック。
10. 「データのプレビュー」枠に現れる表の2列目(「学籍番号」から始まる列)をクリックし、この列の表示の白黒を反転させる。(この作業と次の作業で、学籍番号を数字でなく、文字で認識するように Excel に教えます。学籍番号が数字で認識されると、0 で始まる学籍番号の最初の0が省略されるなど、いろいろ面倒なことが起こります。)
11. 「列のデータ形式」枠の「文字列」横のラジオボタンをクリックする。
12. 「完了」ボタンをクリックする。
13. A 列 1 行から、D 列 46 行からなる表がワークシート「seiseki-dos.txt」に書き込まれたことを確認して下さい。

2.1.2 成績表の作成

次にこれらのデータから、学年別の平均点と全体のデータにおける各人の偏差値を計算し、もとの表に加えます。(平均、分散、標準偏差を求める式は、前に書いてあるものを参考にして下さい。)

1. 1 行目の行番号をクリックし、1 行目をハイライト表示にします。
2. 「挿入」メニューから「行」を選びます。(全体のデータが1行繰り下げられます。)
3. 上の操作をあと6回繰り返し、1 行から7 行まで空の行を作ります。
4. 図2に従って、セル A1, A4, A5, A6, B3, C3, D3, E3 の項目をタイプします。
5. セル E8 に「得点の2乗」とタイプし、改行キーを押します。
6. セル E9 に式「=D9*D9」を入力します。
7. セル E9 をセル E10 からセル E53 にコピーします。
8. セル B4,B5,B6 に必要な値が得られるように計算式をタイプして下さい。
9. 同様に、C4, C5, C6, D4, D5, D6, E4, E5, E6 にも式をタイプします。
10. セル F8 に「偏差値」とタイプします。
11. セル F9 から F53 に全体のデータにおける各人の偏差値が入るように式を入れて下さい。
12. 練習問題2に進む。

2.2 練習問題 2

練習問題 1 のデータを使い成績評価をします。まず得点の隣の列に合否 (○, ×) を書き込みます。

1. E 列の列番号「E」をクリックし、「挿入」メニューから、「列」を選び 1 列挿入します。この際 E3, E4, E5, E6 に書かれた項目が F 列に移る。この欄に書かれた式中の参照セルが適切に記述されていれば, F 列に移った後も表示される値は変わりません。(そうでない人は前に戻って式を書き直して下さい。)
2. セル E8 に「合否」と書き込みます。
3. セル E9 に次の式タイプします。

=IF(D9>=60,"○","×")

4. セル E9 をセル E10 からセル E53 にコピーします。

次に得点の隣の列に成績 (優, 良, 可, 不可) を書き込みます。

1. E 列の列番号「E」をクリックし、「挿入」メニューから、「列」を選び 1 列挿入します。
2. セル E8 に「絶対評価」と書き込みます。
3. セル E9 に次の式をタイプします。

=IF(D9>=80,"優", IF(D9>=70, "良", IF(D9>=60, "可", "不可")))

4. セル E9 をセル E10 からセル E53 にコピーします。
5. 3 行から 6 行のデータで列の挿入によって, ずれた表示を「カット&ペースト」で直します。
6. 終わった人は発展問題に進んで下さい。

2.3 ファイルの保存

Excel のブックを保存する場合は, 標準の形式である「Excel ブック」(拡張子 xls) の他に様々な形式があります。作成したデータを自分自身が同じような環境ですぐ利用する場合は, 標準の形式で保存するのが一番便利です。

データを何年も保存したり, 他の環境でもデータを利用する場合 (Excel で編集したデータを年賀状の住所録に保存したい場合など) は, テキスト形式で保存すると良いでしょう。ただし, この場合は, 表の罫線や飾り, 数式などは保存されず, データ (計算された結果) のみが保存されます。

練習問題 2 で作成したファイルを以下の 3 通りの形式で保存して下さい。

Excel 形式 Excel 形式でファイルを保存する場合は, 「ファイル」メニューから「名前を付けて保存」を選びます。適当な名前を付けてから, 「ファイルの種類」を「Microsoft Excel ブック (*.xls)」に変更して下さい。テキストデータのまま保存すると, セルに書き込んだ式ではなく, 計算された数値が保存され, 文字飾りや罫線などは保存されません。

テキスト形式 テキスト形式で保存する方法の 1 つは「タブ区切り」と呼ばれるもので, 表のデータを保存する際, 列の区切りをタブコード, 行の区切りを改行に代えて保存します。「ファイル」メニューから「名前を付けて保存」を選択したあと適当な名前を付けてから, 「ファイルの種類」を「テキスト (*.txt)」に変更してから保存します。

CSV 形式 CSV 形式もテキスト形式の 1 つです。データを保存する際、列の区切りをカンマ、行の区切りを改行にして保存します。「ファイル」メニューから「名前を付けて保存」を選択したあと適当な名前を付けて、「ファイルの種類」を「CSV 形式 (*.csv)」に変更して保存します。

ファイルの中身を確認するため、メモ帳を開き、上記のファイルを開いてみて下さい。

来年の 4 月に、センター機器の入れ替えがありますが、その際に Excel の利用できる機械を減らすというのが、総合情報処理センターの意向です。コンピュータの環境が変わっても、対応出来るような知識を身につけるようにして下さい。

2.4 発展問題

51	43	013113	理	30	900	32.77
52	44	013114	理	69	4761	55.51
53	45	013115	文	52	2704	45.59
54						
55			成績分布			
56		絶対評価	人数	相対評価	人数	
57		優		A		
58		良		B		
59		可		C		
60		不可		D		
61				E		

図 3:

- I 列に相対評価を記述します。偏差値をもとに、I 列に A,B,C,D,E からなる相対評価を入れて下さい。相対評価の基準は次のようにします。偏差値 65 以上 A, 偏差値 55 以上 65 未満 B, 偏差値 45 以上 55 未満 C, 偏差値 35 以上 45 未満 D, 偏差値 35 未満 E。
- 図 3 のような成績分布の表を作して下さい。(ヒント: COUNTIF という関数を使うと良い。)
- 成績分布の表をもとに、成績分布のヒストグラム (柱状グラフ) を作して下さい。

3 資格試験の案内

4 年次の就職活動に備えて「情報処理技術者」の資格を取る事を強く勧めます。詳しくは独立法人情報処理推進機構 (<http://www.jitec.jp/>) を参照して下さい。試験はいくつかのレベルがありますが、まず、「基本情報技術者」の資格を目指して下さい。試験は年 2 回あります。これより上級の資格として「ソフトウェア開発技術者」の資格もあります。こちらも年 2 回 (1 月申し込み 4 月受験, 7 月申し込み 10 月受験) です。大学在学中に十分取得可能です。一番簡単な初級システムアドミニストレータは、大卒の資格としては意味をなしませんので、注意して下さい。

4 付録：FFFTP の設定の確認方法

1. 「ホストの一覧」 window が出ているかどうか確認して下さい。出ていない場合は、FFFTP の window のメニュー「接続」から「ホストの設定」を選びます。
2. 「ホストの一覧」から「数理」または「ftp.math.u-ryukyu.ac.jp」を選び「設定変更」をシングルクリックで選びます。(間違ってダブルクリックしてしまった人は、「接続」メニューから「切断」を選び、1. からやり直して下さい。)
3. 「ホストの設定」 window が現れます。その中の「基本」タブの2つ隣に「文字コード」というタブがあります。それをクリックします。
4. 「ホストの漢字コード」が「無変換 (SHIFT-JIS)」, 「ファイル名の漢字コード」が「SHIFT-JIS」となっていることを確認します。異なる設定の場合は、この設定に直して下さい。
5. 「OK」ボタンをクリックします。
6. 「ホストの一覧」 window に戻るので、今度は「接続」ボタンをクリックします。(この際、パスワードを聞かれたら、パスワードではなく、自分のメールアドレス (**e06****@eve.u-ryukyu.ac.jp**) を入力します。)
7. 「FFFTP(*)」の window のメニュー「オプション」をクリックし、「環境設定」を選びます。
8. 「転送 1」というタブをクリックし、その中の「転送モード」を「常にバイナリモード」になっているかを確認し、「OK」ボタンをクリックします。(そうでない場合は「常にバイナリモード」を選択し、「OK」ボタンをクリックします。)