

情報科学演習
第 11 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 OpenOffice Calc に取り込む	3
2.1.2	成績表の作成	4
2.2	練習問題 2	4
2.3	発展問題	5

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 OpenOffice で読む方法について知る.

前回に引き続き, 表計算ソフトの使用法を学びます. 今回は成績処理を例に統計の基本用語と, その定義について学習します. 既に Excel 等を使用したことのある人とそうでない人で進度に違いがあるので, 授業では練習 1 までを扱います. 余裕のある人は練習 2, 発展問題へと進んで下さい.

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.

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

偏差値は, 素点を x とすると, 標準偏差 σ と平均 $\bar{x}$ を使って次の式で定義されます.

$$10 \times \frac{x - \bar{x}}{\sigma} + 50$$

標準偏差は, 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, $x_1, x_2, \dots, x_n$ とし, 平均を $\bar{x}$, 分散を V , 標準偏差を σ とすると,

$$\begin{aligned}\bar{x} &= \frac{x_1 + \dots + x_n}{n} \\ V &= \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} = \frac{x_1^2 + \dots + x_n^2}{n} - \left(\frac{x_1 + \dots + x_n}{n} \right)^2 \\ \sigma &= \sqrt{V}\end{aligned}$$

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 (純粋に日本語で, 例えば英語に対応する言葉はありません. deviation value と直訳すると違う意味になります.) 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 B8 に 6 人の成績の平均値を計算する式を書きます。
3. C2 に「としひこ」の得点 (B2) と 6 人の平均点 (B8) の差を式で入力します。

この時、「=B2-B8」と入力してしまうと、C3 にこの式をコピーした時に「=B3-B9」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 (B8) 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、\$を入れます。例えば、セル B8 を絶対参照するためには、\$B\$8 とします。したがって、C2 に入れる式は、B2-\$B\$8 となります。

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。

絶対参照とコピー&ペーストを利用して、表を完成させます。

1. C2 をコピーしてから C3 ~ C7 にペーストします
2. D2 ~ D7 には C 列の 2 乗を式で入力します。
3. E 列には、B 列の 2 乗が入るように式を入力します。
4. E8 には「各人の得点の 2 乗」の平均を入力します。(E2 ~ E7 の平均を計算する式を入力。)
5. B9 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、E8 から B8 の 2 乗を引いた式を書くことになります。)
6. D8 に D2 ~ D7 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。B9 の値と一致することを確認して下さい。)
7. B10 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. F2 ~ F7 に各人の偏差値を計算する式を入力します。
9. C8 に C2 ~ C7 の平均を計算する式を書きます。(この値は、理論上 0 と表示されるものですが、 $x.xxxxE-10$ 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。)

注意

1. 分散や標準偏差は標準的な統計関数ですから、当然それを求めるための関数が備わっています。ただし、それを Help で正確に探すのは、難しいです。理由は、サンプリング調査をして検定、推定をする場合の不変分散と言う概念と、母集団の分散と言う概念があり、これらの正確な説明が Help に書かれていないからです(このへんの事情は Excel でも同じ)。詳しくは、統計と社会や、3 年次の確率統計学の授業で勉強して下さい。
2. 上の成績を後述の 5 段階相対評価で評価すると、2 が 2 人、3 が 3 人、5 が 1 人となります。

2.1 練習問題 1

例題の方法を参考に図のような表を作成してみましょう。

	A	B	C	D	E	F
1		全体	99年度生	00年度生	01年度生	
2	平均	59.56	65.67	60.53	52.47	
3	分散	294.16	279.82	199.58	314.52	
4	標準偏差	17.15	16.73	14.13	17.73	
5						
6	No.	学籍番号	文理	得点	得点の2乗	偏差値
7	1	993101	理	86	7396	65.42
8	2	993102	文	65	4225	53.17
9	3	993103	文	46	2116	42.10
10	4	993104	文	51	2601	45.01

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、元となるデータはこの講義のページにありますからファイルの取り寄せます。

1. firefox で <http://www.math.u-ryukyu.ac.jp/~suga/joho/sampleddata.txt> を表示します
2. 「ファイルメニュー → 名前を付けてページを保存」としてホームフォルダかデスクトップに保存します。

2.1.1 テキストファイルを OpenOffice Calc に取り込む

取り寄せたファイルをダブルクリックして見て下さい。これは、各項目がタブと改行で区切られたテキストデータです。このファイルを OpenOffice Calc で読み込みます。

1. OpenOffice Calc の「挿入 (I)」メニューから「表」を選びます。
2. 「表の挿入」のウィンドウが現れます。「ファイルから作成 (F)」を選びます。
3. 「検索 (B)」をクリックします。
4. 先程保存した sampledata.txt を探して開くを選びます。
5. 「テキストのインポート」のウィンドウで「区切りオプション」の所が「区切る」の所に印がつき、その下の「タブ」にチェックが入っている事を確認する。
6. 同じウィンドウの、1 番下にプレビュー画面が現れますが、学籍番号の上の標準と書いてある文字を右クリックしてテキストに変更します。
7. 右上の OK を押し、「表の挿入」のウィンドウに戻るのもう一度右上の OK を押す。

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 4 回繰り返し、1 行から 5 行まで空の行を作ります。
4. 図に従って、セル A2, A3, A4, B1, D1, D1, E1 の項目をタイプします。
5. セル E6 に「得点の 2 乗」とタイプし、改行キーを押します。
6. セル E7 に式「=D7*D7」を入力します。
7. セル E7 をセル E8 からセル E51 にコピーします。
8. セル B2,B3,B4 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. 同様にして、C2, C3, C4, D2, D3, D4, E2, E3, E4 にも式をタイプします。
10. セル F6 に「偏差値」とタイプします。
11. セル F7 から F51 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. セル G6 に「合否」と入力します。
2. セル G7 に次の式タイプします。

=IF(D7>=60;"○";"×")

3. セル G7 をセル G8 からセル G51 にコピーします。

次に隣の列に成績 (優, 良, 可, 不可) を書き込みます。

1. セル H6 に「絶対評価」と書き込みます。
2. セル H7 に次の式をタイプします。

=IF(D7>=80;"優"; IF(D7>=70; "良"; IF(D7>=60; "可"; "不可")))

3. セル H7 をセル H8 からセル H51 にコピーします。
4. 終わった人は発展問題に進んで下さい。

2.3 発展問題

相対評価と最後のページにあるような統計表を作成します。

- 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.
- 図のような成績分布の表を作って下さい。例えばセル H7 から H53 の中にある優の数を数えるには、COUNTIF(H7:H53;"優") と入力します (コロン : と セミコロン ; に注意)。
- 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) を作って下さい。

52					
53			成績分布		
54		絶対評価	人数	相対評価	人数
55		優	5	A	2
56		良	7	B	11
57		可	14	C	19
58		不可	19	D	8
59				E	5

