

データの分析・SQL

I. データの分析

数学Iで履修した「データの分析」と同じ内容で、さらに簡単です。

Excelの関数を使って標準偏差や相関係数が計算できれば大丈夫です。

右図の**10**個のデータの標準偏差を求めましょう。公式の通りに計算していきます。

	A	B	C
1	データ	偏差	偏差の二乗
2	59	7.5	56.25
3	55	3.5	12.25
4	64	12.5	156.25
5	42	-9.5	90.25
6	38	-13.5	182.25
7	46	-5.5	30.25
8	37	-14.5	210.25
9	42	-9.5	90.25
10	57	5.5	30.25
11	75	23.5	552.25
12			
13	平均値	51.5	
14	偏差の二乗の平均	141.05	←分散
15	↑の平方根	11.8764	←標準偏差

1. 平均値を求める

B13: =AVERAGE (A2:A11)

2. 偏差を求める

B2のセルに平均値との偏差を代入します。

B2: =A2-B13

B3:B11のセルも同様にします。

3. 偏差を二乗する

C2のセルに二乗した値を代入します。2乗の演算は **^** で行います。

C2: =B2^2

C3:C11のセルも同様にします。

4. 3の平均を求める

分散が求められます。

B14: =AVERAGE (C2:C11)

5. 4の平方根を求める

平方根は**SQRT**関数で求められます。Square RooTの意味。

B15: SQRT (B14)

共分散と相関係数について

残念ながら筆者の課題の進捗が芳しくないため、割愛します。しかし、標準偏差の求め方を理解し、関数で表現することのできたあなたなら大丈夫です。

なお、分散、標準偏差、相関係数はそれぞれ、

VAR. P () : **VAR**iance **P**opulation

STDEV. P () : **ST**andard **DEV**iation based on the entire **P**opulation

CORREL () : **CORREL**ation coefficient

で求められます。

II. SQL

DEPTNO	DNAME	LOC
10	ACCOUNTING	NEW YORK
20	RESEARCH	DALLAS
30	SALES	CHICAGO
40	OPERATIONS	BOSTON

↑ **DEPT**テーブル

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
7369	SMITH	CLERK	7902	1980-12-17	800	300	20
7499	ALLEN	SALESMAN	7698	1981-02-20	1600	300	30
7521	WARD	SALESMAN	7698	1981-02-22	1250	500	30
7566	JONES	MANAGER	7839	1981-04-02	2975		20
7654	MARTIN	SALESMAN	7698	1981-09-28	1250	1400	30
7698	BLAKE	MANAGER	7839	1981-05-01	2850		30
7782	CLARK	MANAGER	7839	1981-06-09	2450		10
7839	KING	PRESIDENT		1981-11-17	5000		10
7844	TURNER	SALESMAN	7698	1981-09-08	1500	0	30
7900	JAMES	CLERK	7698	1981-12-03	950		30
7902	FORD	ANALYST	7566	1981-12-03	3000		20
7934	MILLER	CLERK	7782	1982-01-23	1300		10

↑ **EMP**テーブル

SQLについては、学校の「SQL演習プリント」がよくまとまっているので、こちら（二次元コード、URL）を見て、

<https://sql.main.jp/sql01.html>

演習することをお勧めします。プリントにはデータベースが全て載っていなかった
ので、上に載せておきます。

