

統計レポート課題 1 : 解答例

1. Excel を用いて、指定されたデータの平均、分散、標準偏差、共分散、相関係数などを求める

結果

	1 回目	2 回目	
データ数	122	122	
平均	78.84	76.41	
分散	89.86	195.08	
標準偏差	9.48	13.97	
共分散			54.07
相関係数			0.41

Excel のワークシートの抜粋

◇	A	B	C	D	E	F	G	H
1		1回目	2回目	1回目の偏差	2回目の偏差	1回目の偏差の2乗	偏差の積	2回目の偏差の2乗
2	1	55	61	-23.83606557	-15.40983607	568.158022	367.3098629	237.4630476
3	1	59	64	-19.83606557	-12.40983607	393.4694974	246.162322	154.0040312
4	1	59	75	-19.83606557	-1.409836066	393.4694974	27.96560064	1.987637732
5	1	60	65	-18.83606557	-11.40983607	354.7973663	214.9164203	130.184359
120	1	97	111	18.16393443	34.59016393	329.9285138	628.2934695	1196.479441
121	1	98	76	19.16393443	-0.409836066	367.2563827	-7.854071486	0.167965601
122	1	100	102	21.16393443	25.59016393	447.9121204	541.5885515	654.8564902
123	1	103	104	24.16393443	27.59016393	583.895727	666.6869121	761.2171459
124	122	9618	9322			10962.72131	6596.196721	23799.5082
125	平均	78.83607	76.40984		分散	89.85837141		195.077936
126					標準偏差	9.479365559		13.96703032
127						共分散	54.06718624	
128						相関係数	0.408366794	

データ数を求める

- 1 回目のデータは B 列 [B2:B123]、2 回目のデータは C 列 [C2:C123] にそれぞれ入力する。
- A 列にすべて 1 を入れる [A2:A123]。
- A 列の値を合計したものがデータ数となる [A124=SUM(A2:A123)]。ここではデータ数 (n) は 122。

(注) こうしておけば、データを挿入したり削除した場合でも自動的に再計算されるハズ。

平均を求める

定義式は以下のとおり。データの総和をデータ数で割る。

$$\mu_x = \frac{1}{n} \sum_{k=1}^n x_k$$

1 回目のデータの平均を求める [B125]

- SUM() 関数を利用して、データの総和を求める [B124=SUM(B2:B123)]。
- 求めた総和をデータ数で割る [B125=B124/A124]。
- 値は 78.84(小数点以下第 3 位四捨五入)。

2 回目のデータの平均を求める [C125]

- SUM() 関数を利用して、データの総和を求める [C124=SUM(C2:C123)]。
- 求めた総和をデータ数で割る [C125=C124/A124]。
- 値は 76.41(小数点以下第 3 位四捨五入)。

分散を求める

定義式は以下のとおり。偏差の 2 乗の総和をデータ数で割る。

$$\sigma^2(x) = \frac{1}{n} \sum_{k=1}^n (x_k - \mu_x)^2$$

1 回目のデータの分散を求める [F125]

- D 列 2 行に 1 件目のデータと平均との偏差を求める [D2=B2-\$B\$125]。\$マークは絶対参照を意味する(あるいは平均の列をつくる)。
- 2 件目以降も同様に偏差を求める [D3=B3-\$B\$125 … D123=B123-\$B\$125]。
- F 列に偏差の 2 乗を求める [F2=D2*D2 … F123=D123*D123]。
- SUM() 関数を利用して、偏差の 2 乗の総和を求める [F124=SUM(F2:F123)]。
- 偏差の 2 乗の総和をデータ数で割る [F125=F124/A124]。
- 値は 89.86(小数点以下第 3 位四捨五入)。

2 回目のデータの分散を求める [H125]

- E 列 2 行に 1 件目のデータと平均との偏差を求める [E2=C2-\$C\$125]。
- 2 件目以降も同様に偏差を求める [E3=C3-\$C\$125 … E123=C123-\$C\$125]。
- H 列に偏差の 2 乗を求める [H2=E2*E2 … H123=E123*E123]。
- SUM() 関数を利用して、偏差の 2 乗の総和を求める [H124=SUM(H2:H123)]。
- 偏差の 2 乗の総和をデータ数で割る [H125=H124/A124]。
- 値は 195.08(小数点以下第 3 位四捨五入)。

標準偏差を求める

定義式は以下のとおり。分散の平方根。

$$\sigma(x) = \sqrt{\sigma^2(x)}$$

1 回目のデータの標準偏差を求める [F126]

- SQRT() 関数を用いて、分散の平方根を求める [F126=SQRT(F125)]。
- 値は 9.48(小数点以下第 3 位四捨五入)。

2 回目のデータの標準偏差を求める [H126]

- SQRT() 関数を用いて、分散の平方根を求める [H126=SQRT(H125)]。
- 値は 13.97(小数点以下第 3 位四捨五入)。

共分散を求める

定義式は以下のとおり。1 回目と 2 回目の偏差の積の総和をデータ数で割る。

$$cov(x, y) = \frac{1}{n} \sum_{k=1}^n (x_k - \mu_x)(y_k - \mu_y)$$

1 回目と 2 回目の共分散を求める [G127]

- G 列に 1 回目と 2 回目の偏差の積を求める [G2=F2*H2 … G123=F123*H123]。
- SUM() 関数を利用して、偏差の積の総和を求める [G124=SUM(G2:G123)]
- 偏差の積の総和をデータ数で割る [G127=G124/A124]。
- 値は 54.07(小数点以下第 3 位四捨五入)。

相関係数を求める

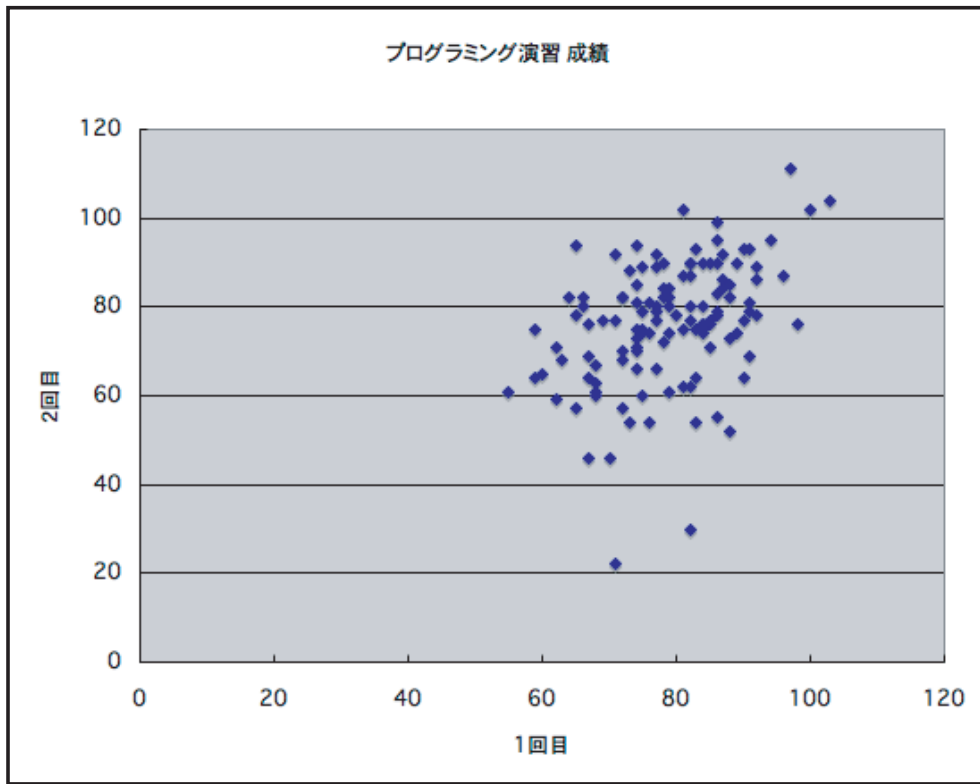
定義式は以下のとおり。共分散を 1 回目と 2 回目の標準偏差の積で割る。

$$r(x, y) = \frac{cov(x, y)}{\sigma(x)\sigma(y)}$$

1 回目と 2 回目の相関係数を求める [G128]

- 共分散を 1 回目と 2 回目の標準偏差の積で割る [G128=G127/(F126*H126)]。
- 値は 0.41(小数点以下第 3 位四捨五入)。

データの散布図



2. 統計の間違った使い方を指摘する

以下に示す2つの図は、「NPO 法人日本むし歯予防フッ素推進会議. ”新潟県はなぜ日本で一番むし歯が少ないの?”」という記事からの抜粋である。

図1において、新潟県では12歳の子供の虫歯が最も少ない理由としてフッ化物洗口の普及を挙げている。しかし、図2を見ると、新潟県に続いて虫歯が少ない広島県でのフッ化物洗口実施率は47都道府県のうち34位であり決して高いとは言えない。つまり、虫歯が少ないこととフッ化物洗口の実施率には、おそらく強い相関関係はない。従って、虫歯が最も少ない理由としてフッ化物洗口の普及があるという結論をここから導くことは誤りである。

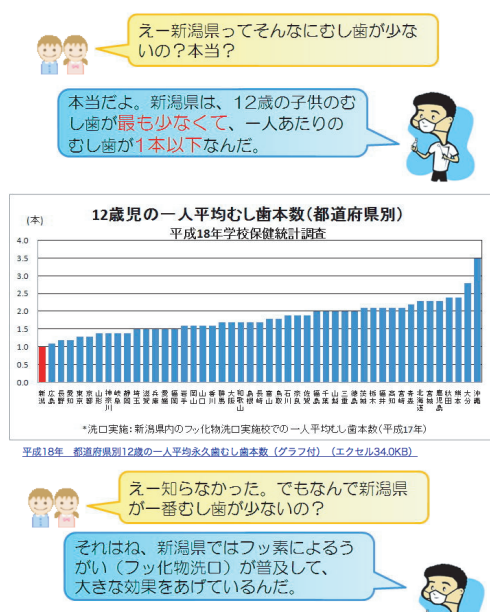


図1: 12歳児の一人平均虫歯本数(都道府県別)

出典

NPO 法人日本むし歯予防フッ素推進会議. ”新潟県はなぜ日本で一番むし歯が少ないの?”. NPOnitif. 2007-08-01. <http://www.nponitif.jp/newpage57.html>. (参照 2008-01-08).

わが国の集団フッ化物洗口実施状況(2006年3月末調査)

全国47都道府県 5,133施設491,334人

都道府県	施設数	人数	都道府県	施設数	人数
北海道	137	7,840	滋賀	71	9,368
青森	38	5,175	京都	124	23,323
岩手	81	2,953	大阪	2	246
宮城	69	5,650	兵庫	203	8,361
秋田	77	3,147	奈良	14	1,110
山形	104	10,565	和歌山	65	5,538
福島	68	8,542	鳥取	8	179
茨城	2	119	島根	108	6,084
栃木	32	3,227	岡山	10	2,141
群馬	74	4,786	広島	46	1,841
埼玉	76	8,207	山口	179	21,552
千葉	14	1,217	徳島	5	113
東京	3	235	香川	78	12,911
神奈川	14	1,186	愛媛	120	18,789
新潟	760	82,814	高知	37	1,303
富山	192	24,355	福岡	24	2,080
石川	42	2,007	佐賀	359	34,764
福井	35	1,435	長崎	158	9,621
長野	121	22,322	熊本	193	7,023
山梨	19	769	大分	31	1,071
岐阜	64	9,514	宮崎	163	5,355
静岡	456	36,478	鹿児島	118	4,689
愛知	496	68,803	沖縄	31	2,029
三重	12	297	Total	5,133	491,334

広島県は34位

2006年NPO法人日本むし歯予防フッ素推進会議調査

図2: フッ素化洗口実施状況