

表計算 (Excel) とデータ処理 (8)  
二変数の統計量

1 資料用ファイルのダウンロード

ダウンロード済みでない場合は、<http://isl.sss.fukushima-u.ac.jp/> から前回の資料をダウンロード。

2 2つの変数の記述統計

人口が多ければ診療所数が多いことが想定できる。ワークシート「都道府県別の診療所数と人口・面積」のデータについて、このことを確かめてみよう。

2つの変数（この場合は一般診療所数と人口）同士の関係のことを相関という。ただし、変数が間隔尺度以上の場合に相関といい、「好き／嫌い」や出身県（「福島／山形／…」）など名義尺度、順序尺度の場合は連関という。

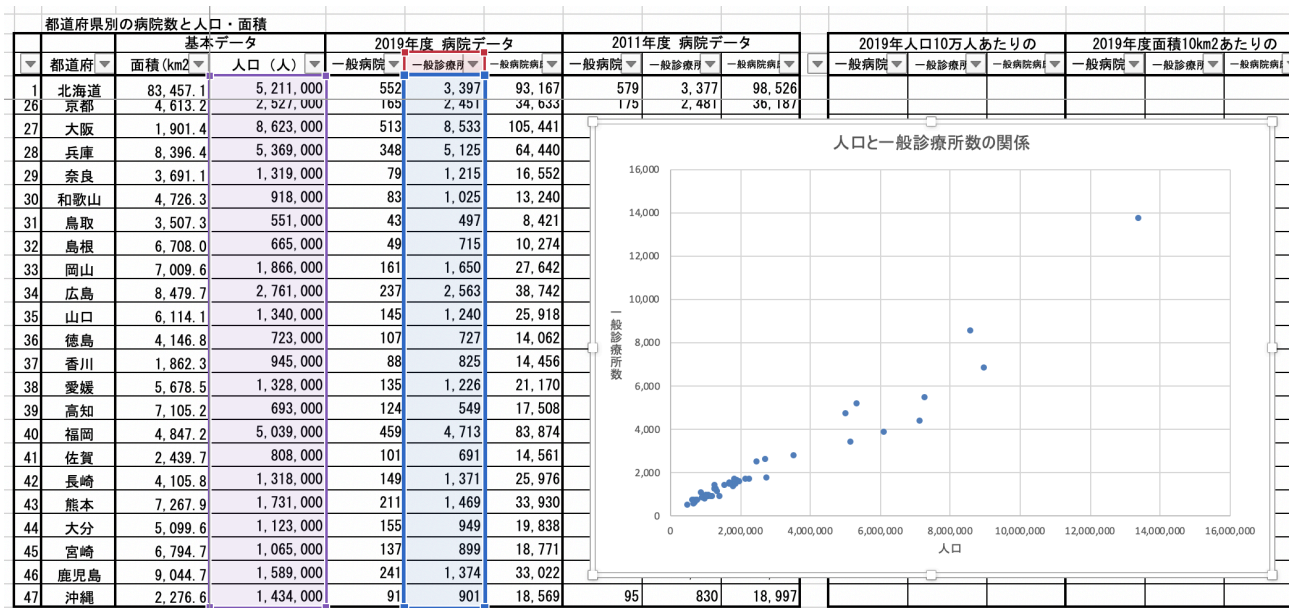
2.1 散布図 (scattergram) と相関 (correlation)

散布図とは、2つの変数について、一方を横軸（x軸）の値、もう一方を縦軸（y軸）の値としてグラフ化した図である。

（練習）ワークシート「都道府県別の診療所数と人口・面積」で人口を横軸、一般診療所数を縦軸にした散布図を書いてみよう。

- 1) 「人口（人）」と「2019年度一般診療所数」のデータを項目名も含めて選択する。
- 2) 「挿入」－「グラフ：散布図」を選択し散布図を書く。

※横軸（x軸）が診療所数、縦軸（y軸）が人口になってしまう場合は、グラフを右クリックして、「データの選択」「編集」で、系列X, Yの値をそれぞれ人口と診療所数にする。



（練習）同様に、「都道府県別の診療所数と人口・面積」で面積を横軸、一般診療所数を縦軸にした散布図を書いてみよう。人口と面積でどう違うか。

グラフを見ると、人口が多いと診療所数もおおむね多いことがわかる。このように、一方の変数の値が大きくなると、もう一方の値も大きくなることを「正の相関」という。逆に、小さくなる関係を「負の相関」と言う。また、このような関係が見えないことを「無相関」という。

### 2.2 共分散 (covariance)

共分散とは、各変数のデータについて、平均との差（平均からの偏差）の積和の平均のことで、以下のよう表せる。変数  $x$  と変数  $y$  の共分散を  $s_{xy}$  とすると、

$$s_{xy} = \left( \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n} \right)$$

シグマ記号を用いて書き直すと、

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

共分散の正負は、相関の正負と同じになるが、相関の「程度」については、共分散の値だけではわからないので、次項の「相関係数」の計算が必要になる。

共分散は、エクセルでは、`COVAR(変数xのデータ, 変数yのデータ)` という関数が用意されている。

### 2.3 相関係数 (correlation coefficient)

相関係数とは、2つの変数の相関の度合いを示す値で、-1 から +1 までの値になる。+1 に近づくほど正の相関の度合いが強くなり、-1 に近づくほど負の相関の度合いが強くなる。相関係数が 0 に近いほど無相関になる。

相関係数  $r_{xy}$  は、以下の式で計算できる。 $s_{xy}$  は変数  $x$  と  $y$  の共分散、 $s_x, s_y$  はそれぞれ変数  $x, y$  の標準偏差である。

$$r_{xy} = \frac{s_{xy}}{s_x \times s_y}$$

相関係数は、エクセルでは、`CORREL(変数xのデータ, 変数yのデータ)` という関数が用意されている。

(練習) 診療所数と人口の相関係数を計算させてみよう。

手順) セルG64に、一般診療所数と人口のデータの相関を計算する関数を書く。

`=CORREL($E5:$E51, G5:G51)`

|          |  |      |  |      |      |      |      |
|----------|--|------|--|------|------|------|------|
| 面積との相関係数 |  | 0.33 | 0.00                                     | 0.26 | 0.35 | 0.01 | 0.28 |
| 人口との相関係数 |  | 0.89 | <code>=CORREL(\$E5:\$E51, G5:G51)</code> |      |      | 0.97 | 0.93 |

相関の強さは相関係数の値でおおむね以下のように評価することが多い。

| 相関係数 $r$                               | 評価        |
|--|-----------|
| $-0.2 \leq r \leq 0.2$                 | ほとんど相関がない |
| $-0.4 \leq r < -0.2, 0.2 \leq r < 0.4$ | 弱い相関      |
| $-0.7 \leq r < -0.4, 0.4 \leq r < 0.7$ | 中程度の相関    |
| $-1.0 \leq r < -0.7, 0.7 \leq r < 1.0$ | 強い相関      |

(問) 一般診療所数との相関の強さはどの程度と評価できるか？

記入する式：

|       |                            |
|-------|----------------------------|
| 合計    | =SUM (D5:D51)              |
| 最小値   | =MIN (D5:D51)              |
| 第1四分位 | =QUARTILE. INC (D5:D51, 1) |
| 中央値   | =MEDIAN (D5:D51)           |
| 第3四分位 | =QUARTILE. INC (D5:D51, 3) |
| 最大値   | =MAX (D5:D51)              |
| 平均    | =AVERAGE (D5:D51)          |
| 分散    | =VARP (D5:D51)             |
| 標準偏差  | =STDEVP (D5:D51)           |

面積との相関係数

| 2019年人口10万人あたりの |                 |                 |
|-----------------|-----------------|-----------------|
| 一般病院数           | 一般診療所数          | 一般病院病床数         |
| =F5*100000/\$E5 | =G5*100000/\$E5 | =H5*100000/\$E5 |

| 2019年度面積10km2あたりの |             |             |
|-------------------|-------------|-------------|
| 一般病院数             | 一般診療所数      | 一般病院病床数     |
| =F5*10/\$D5       | =G5*10/\$D5 | =H5*10/\$D5 |