

# 線形回帰モデル 追加レポート

## 共分散について

参考サイト：

- [共分散の意味と求め方、共分散公式の使い方](https://sci-pursuit.com/math/statistics/covariance.html) (https://sci-pursuit.com/math/statistics/covariance.html)
- [【Pythonで学ぶ】絶対にわかる共分散【データサイエンス:統計編⑩】](https://datawokagaku.com/covariance/) (https://datawokagaku.com/covariance/)

共分散とは、2つのデータの関係を示す指標で、求め方は、2変数の偏差の積の平均となる。

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

ここで、 $\bar{x}, \bar{y}$  は平均を表す。

## 例題

次に示す英語と数学の共分散を求める。

英語	数学
50	40
60	70
70	90
80	60
90	100

In [1]:

```
import numpy as np
```

In [2]:

```
# 初期値を与える。英語をx、数学をyとする。
x = np.array([50, 60, 70, 80, 90])
y = np.array([40, 70, 90, 60, 100])
```

In [3]:

```
# 英語と数学の平均を求める
x_mean = np.mean(x)
y_mean = np.mean(y)
print(f'{x_mean}, {y_mean}')
```

70.0, 72.0

In [4]:

```
# 偏差を求める
x_dev = x - x_mean
y_dev = y - y_mean
```

In [5]:

```
print(x_dev)
print(y_dev)
```

```
[-20. -10.  0. 10. 20.]
[-32. -2. 18. -12. 28.]
```

In [6]:

```
# 共分散を求める
cov = (x_dev * y_dev / 5).sum()
```

In [7]:

```
print(f'共分散 : {cov}')
```

共分散 : 220.0

In [8]:

```
# NumPyの関数で共分散を求めたい
np.cov(np.array([[50, 60, 70, 80, 90], [40, 70, 90, 60, 100]]))
```

Out[8]:

```
array([[250., 275.],
       [275., 570.]])
```

上記の結果は、不偏分散と不偏共分散が出力されている。これは、偏差の合計をかけ合わせた後  $n - 1$  で割ったときの値となる。

共分散を出力したいときは、`np.cov` 関数のパラメータに `bias=True` を追加する。

In [9]:

```
np.cov(np.array([[50, 60, 70, 80, 90], [40, 70, 90, 60, 100]]), bias=True)
```

Out[9]:

```
array([[200., 220.],
       [220., 456.]])
```

上記結果の(0, 1)要素と(1, 0)要素が共分散となる。

(0, 0)要素はxの分散、(1, 1)要素はyの分散である。

In [10]:

```
x_var = np.sum((x - x_mean)**2) / 5
y_var = np.sum((y - y_mean)**2) / 5
```

In [11]:

```
print(f'Xの分散 : {x_var}, Yの分散 : {y_var}')
print(' - np.covの出力結果と等しいことを確認する')
```

Xの分散 : 200.0, Yの分散 : 456.0  
- np.covの出力結果と等しいことを確認する