

## 2010 年度前期 応用統計学 第 2 回演習の解答例

データ数を  $n$  とします。

(1) 表 A1 の通り計算します。

$$n = 9, \bar{x} = 34.49, \bar{y} = 15.28$$

$$\sum x_i^2 = 10722.25, \sum x_i y_i = 4726.07$$

ですから、回帰方程式を  $y = a + bx$  とすると

$$b = \frac{4726.07 - 9 \cdot 34.49 \cdot 15.28}{10722.25 - 9 \cdot (34.49)^2} = -1.048$$

$$a = 15.28 - (-1.048) \cdot 34.49 = 51.43$$

となります。

地名	緯度 (x)	気温 (y)	$x^2$	xy
長野	36.67	11.4	1344.69	418.04
静岡	34.97	16.0	1222.90	559.52
名古屋	35.17	14.9	1236.93	524.03
大阪	34.68	16.2	1202.70	561.82
鳥取	35.48	14.4	1258.83	510.91
広島	34.40	15.0	1183.36	516.00
高知	33.55	16.3	1125.60	546.87
福岡	33.92	16.0	1150.57	542.72
鹿児島	31.57	17.3	996.66	546.16
	平均	平均	$\sum x^2$	$\sum xy$
	34.49	15.28	10722.25	4726.07

表 A1: 長野～鹿児島のデータを使った場合

(2) 同様に、表 A2 の通り計算します。

$$n = 19, \bar{x} = 35.93, \bar{y} = 14.07$$

$$\sum x_i^2 = 24765.99, \sum x_i y_i = 9407.79$$

ですから、回帰方程式を  $y = a + bx$  とすると

$$b = \frac{9407.79 - 19 \cdot 35.93 \cdot 14.07}{24765.99 - 19 \cdot (35.93)^2} = -0.831$$

$$a = 14.07 - (-0.831) \cdot 35.93 = 43.93$$

となります。

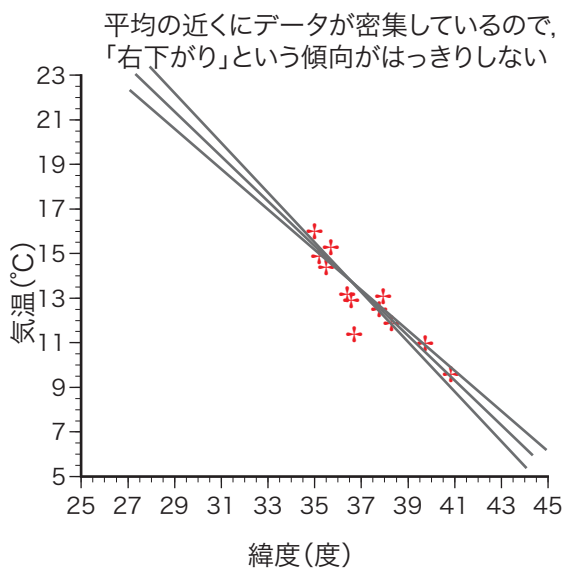
地名	緯度 (x)	気温 (y)	$x^2$	xy
札幌	43.05	8.0	1853.30	344.40
青森	40.82	9.6	1666.27	391.87
秋田	39.72	11.0	1577.68	436.92
仙台	38.27	11.9	1464.59	455.41
福島	37.75	12.5	1425.06	471.88
宇都宮	36.55	12.9	1335.90	471.50
水戸	36.38	13.2	1323.50	480.22
東京	35.68	15.3	1273.06	545.90
新潟	37.92	13.1	1437.93	496.75
長野	36.67	11.4	1344.69	418.04
静岡	34.97	16.0	1222.90	559.52
名古屋	35.17	14.9	1236.93	524.03
大阪	34.68	16.2	1202.70	561.82
鳥取	35.48	14.4	1258.83	510.91
広島	34.40	15.0	1183.36	516.00
高知	33.55	16.3	1125.60	546.87
福岡	33.92	16.0	1150.57	542.72
鹿児島	31.57	17.3	996.66	546.16
那覇	26.20	22.4	686.44	586.88
	平均	平均	$\sum x^2$	$\sum xy$
	35.93	14.07	24765.99	9407.79

表 A2: 札幌～那覇の全データを使った場合

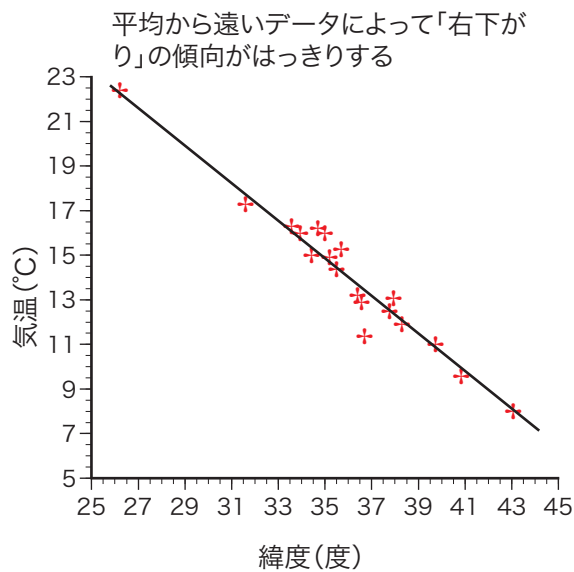
ところで、長野～鹿児島間のデータのみを用いたときの相関係数は、表 A3 のように計算されます。「緯度 (気温) の分散 = [緯度 (気温) の偏差]<sup>2</sup> の和 / データ数」, 「共分散 = [緯度の偏差 × 気温の偏差] の和 / データ数」, 「相関係数 = 共分散 / [√{緯度の分散} × √{気温の分散}]」です。

同様に、札幌～那覇の全データを使って相関係数を計算すると  $-0.974$  となり、長野～鹿児島の場合よりも大きくなります。また、前者の決定係数は  $0.712$ 、後者の決定係数は  $0.949$  となります。

データの大半は共通なのにこのような違いがあるのは、後者のほうには、札幌・那覇という平均から離れたデータがあるために、「散布図のうえで右下がり」という傾向がよりはっきりしていることにより、長野～鹿児島のデータでは、各データが平均付近に密集しているため、「右下がり」の傾向はありません。ところがこれに札幌や那覇などの平均から遠いデータが加わると、「右下がり」の傾向がよりはっきりとするため、決定係数が大きくなります。



(a) 長野～鹿児島までのデータを使った場合



(b) 札幌～那覇までのデータを使った場合

図 A1: (a) 平均に近いデータばかりの場合, (b) 平均から遠いデータがある場合

地名	緯度 (度)	気温 (°C)	緯度の偏差	左の2乗	気温の偏差	左の2乗	両偏差の積
長野	36.67	11.4	2.18	4.752	-3.878	15.037	-8.454
静岡	34.97	16.0	0.48	0.230	0.722	0.522	0.347
名古屋	35.17	14.9	0.68	0.462	-0.378	0.143	-0.257
大阪	34.68	16.2	0.19	0.036	0.922	0.850	0.175
鳥取	35.48	14.4	0.99	0.980	-0.878	0.770	-0.869
広島	34.40	15.0	-0.09	0.008	-0.278	0.077	0.025
高知	33.55	16.3	-0.94	0.884	1.022	1.045	-0.961
福岡	33.92	16.0	-0.57	0.325	0.722	0.522	-0.412
鹿児島	31.57	17.3	-2.92	8.526	2.022	4.089	-5.905
	緯度の平均 = 34.49	気温の平均 = 15.278		緯度の分散 = 1.800		気温の分散 = -2.562	共分散 = -1.812 相関係数 = -0.844

表 A3: 相関係数を求める