

「ちょっと問題」の解答例

1. 相関関係は、「因果関係」については何も述べていません。所得が増えたから酒をたくさん飲むのか、酒をたくさん飲むから所得が増えるのか、この問題の場合は常識的に前者とわかりますが、相関関係自体は前者であるとも後者であるとも言っていません。
2. 「年数と台数の相関関係が1である」とは、年数と台数の増減に直線的な関係があることを意味しています。この問題の場合は、年数と台数の間に指数関数の関係があるので、相関係数はほぼ1にはなりません。このような場合、台数のほうを対数関数で変換すると、年数と「台数の対数」の増減に直線的な関係ができ、相関関係として取り扱うことができます（このような変換については、この講義では扱いません）。

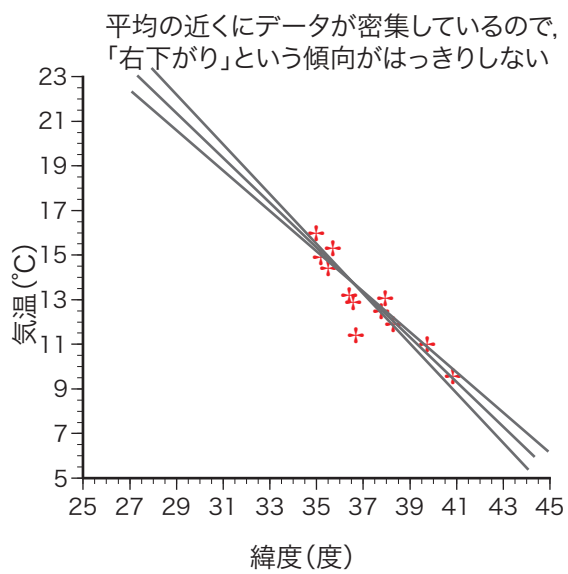
「今日の演習」の解答例

下の表の通りです。「緯度（気温）の分散 = [緯度（気温）の偏差]²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。

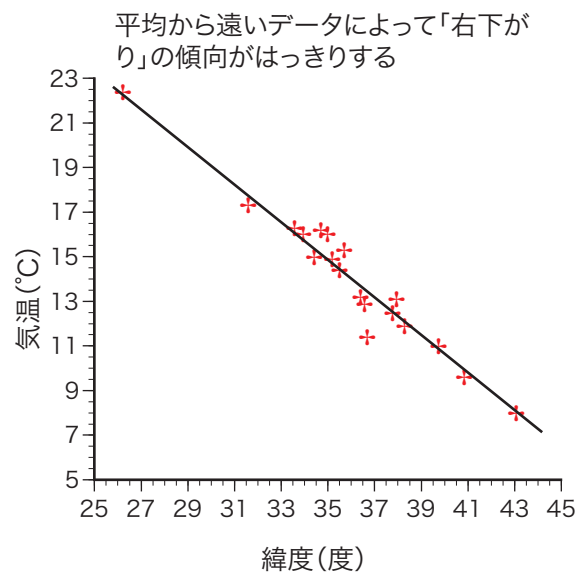
地名	緯度 (度)	気温 (°C)	緯度の偏差	左の2乗	気温の偏差	左の2乗	両偏差の積
長野	36.67	11.4	2.18	4.752	-3.878	15.037	-8.454
静岡	34.97	16.0	0.48	0.230	0.722	0.522	0.347
名古屋	35.17	14.9	0.68	0.462	-0.378	0.143	-0.257
大阪	34.68	16.2	0.19	0.036	0.922	0.850	0.175
鳥取	35.48	14.4	0.99	0.980	-0.878	0.770	-0.869
広島	34.40	15.0	-0.09	0.008	-0.278	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 A1: 相関係数を求める

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。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ので、「右下がり」の傾向はありますがそれがはっきりしません。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、相関係数が大きくなります。これについては、第7回の回帰分析についての講義で説明します。



(a) 長野～鹿児島データのデータを使った場合



(b) 札幌～那覇データ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