

第3章 練習問題

本問において databank とは<http://www.qmss.jp/databank/> をさす。

[3.1] (データ出所についての考慮) ある初級者の風景画につき、同じ町内に住む6人の画家が10点満点で次の評点を付けた。

1, 6, 5, 5.5, 10, 5

この結果から、この絵に対する平均的評価を判断する適切・公平な方法を、自ら状況を想定しながら複数提案しなさい。

[3.2] (1次元データの基本計算) データ{5, 3, 1, 4, 2}の分散、標準偏差および平均偏差を筆算で計算し、また各値の偏差値を算出しなさい。

[3.3] (同, エクセルによる) databank の「計量マス」データに対し、つぎの統計量を求めなさい。

- a) 平均
- b) 中央値 (中位数) $med.$
- c) レンジ (範囲) R
- d) 平均偏差 d
- e) 分散および標準偏差
- d) 最大値および最小値それぞれの偏差値

[3.4] (X^2 統計量) 表3-2に対して X^2 の統計量を計算しなさい。肥沃度と利用形態の間には関連はあるか。なお関連がない(独立している)と想定した場合のクロス表の数字もあわせて与えること。

[ヒント] $f_{i,j}/n$

[3.5] (回帰分析への理解) databank の「私鉄データ」のデータに対して、独立変数と従属変数を入れ替えることで、もう1通り回帰分析を行いなさい。

[3.6] (オーバー・フィッティング) メンデルのデータで適合度の X^2 等計量の値が極めて小さいことの問題点を述べなさい。これに関して、適合度の検定を両側検定とすることもできる。このことの意味を述べなさい。

[3.7] (研究) 右のデータは、()フィリップ曲線といわれる。インフレ率と失業率の関係、()1人当りGNP水準とその成長率(経済成長率)のデータである。2つのデータをエクセルで入力し、図示してその関係を各200字前後で述べなさい[研究]。

[3.8] (多重共線性, multicollinearity) 相関の高い2つの独立変数 x_1, x_2 によって重回帰分析をおこなうとき、標準誤差が極めて大きくなり、 t 値が著しく低下することを、databank 2-9dによって確認しなさい[研究]。

フィリップス曲線

年	失業	インフレ
1961	6.70	1.10
62	5.60	2.10
63	5.70	1.20
64	5.20	1.70
65	4.50	2.70
66	3.80	3.50
67	3.85	3.00
68	3.50	4.75
69	3.40	4.90
70	5.00	5.10
71	6.00	5.50
72	5.70	4.60
73	4.85	6.00
74	5.70	8.40
75	8.50	9.20
76	7.70	6.10
77	7.10	6.60
78	6.10	7.60
79	5.80	8.40
80	7.20	9.00
81	7.60	9.50
82	9.70	6.00
83	9.60	4.00
84	7.50	4.40
85	7.20	3.60
86	7.00	2.70
87	6.20	3.00
88	5.50	3.80
89	5.30	4.50
90	5.60	4.20
91	6.70	3.80
92	7.40	2.80
93	6.80	2.20
94	6.10	2.10

1人当たり GNP とその成長率

国名	1870 年値	1979 vs 1870
	Y/N	d (Y/N)
日本	5.84	2.87
フィンランド	6.27	2.40
スウェーデン	6.36	2.45
ノルウェー	6.53	2.27
ドイツ	6.64	2.20
イタリア	6.65	1.78
オーストラリア	6.66	2.02
フランス	6.78	2.04
カナダ	6.82	2.10
デンマーク	6.83	2.00
アメリカ	6.97	2.04
オランダ	7.03	1.64
スイス	7.05	1.74
ベルギー	7.07	1.65
イギリス	7.13	1.42
オーストラリア	7.58	1.15

自然対数値 .なお、図からの読取値で誤差を含む(同).

図からの読取値で誤差を含む(ロー
マー).