

応用統計学 2017 第1回 記述統計の基礎

2017年9月27日 (水)

清 智也 sei@mist.i.u-tokyo.ac.jp

<http://www.stat.t.u-tokyo.ac.jp/~sei/lec-j.html>

データを代表値 (summary) や図表によって分かりやすい形に整理することを、一般に記述統計 (descriptive statistics) という。

- 1次元：平均，分散，標準偏差，度数分布表，経験分布関数，分位点関数，中央値，四分位点，箱ひげ図，ヒストグラム¹。
- 多次元：共分散，相関係数，散布図，分割表，層別，シンプソンのパラドックス²。

演習問題

問題 1-1. 次のデータは，あるクラスの23人の学生に1円玉の所持枚数を聞いた結果である。枚数の平均，中央値，度数分布表を求めよ。また箱ひげ図を描け。

3, 3, 2, 5, 0, 6, 3, 0, 4, 7, 2, 0, 1, 33, 0, 2, 3, 13, 2, 14, 4, 13, 7

問題 1-2 (Anscombe の例). 次の4組の2変量データについて，それぞれ相関係数を求めよ。また散布図を描け。なお，CSV ファイルを講義ページからダウンロードすることができる。また統計ソフトウェアのRでは，デフォルトで `anscombe` というオブジェクトが用意されている。

x	y	x	y	x	y	x	y
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

¹mean, variance, standard deviation, frequency table, empirical distribution function, quantile function, median, quartile, box plot, histogram.

²covariance, correlation coefficient, scatter plot, contingency table, stratification, Simpson's paradox.

問題 1-3. 次の 3 元分割表に対し、ベクトル $(a_1, b_1), (c_1, d_1), (a_2, b_2), (c_2, d_2)$ を 2 次元平面に描き、シンプソンのパラドックスが生ずる条件を幾何学的に説明せよ³。

n_{ij1}	1	2	n_{ij2}	1	2
1	a_1	b_1	1	a_2	b_2
2	c_1	d_1	2	c_2	d_2

問題 1-4. n を奇数とする。 $x_1, \dots, x_n \in \mathbb{R}$ の中央値を $f(x_1, \dots, x_n)$ と表すとき、以下の 3 つの性質が満たされることを示せ。また、 n が偶数のとき、これらの性質を満たすような関数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ は存在しないことを示せ。

- (i) 任意の置換 (i_1, \dots, i_n) に対して $f(x_{i_1}, \dots, x_{i_n}) = f(x_1, \dots, x_n)$.
- (ii) 任意の単調増加関数 $g: \mathbb{R} \rightarrow \mathbb{R}$ に対して $g(f(x_1, \dots, x_n)) = f(g(x_1), \dots, g(x_n))$.
- (iii) $f(-x_1, \dots, -x_n) = -f(x_1, \dots, x_n)$.

宿題

問題 1-5. 講義の web ページ

<http://www.stat.t.u-tokyo.ac.jp/~sei/lec-j.html>

にアクセスし、アンケートに答えよ。

³宮川「統計技法」7章。