

§1 データリテラシー (1) 演習問題

📖 問題の難易度の目安【易】☆☆☆ 【基礎】★★☆ 【標準】★★★

1 (☆☆☆)(定量的データと定性的データ)

(1) 次の語群を，定量的データと定性的データに分類せよ：

5段階アンケート調査， 不快感指数， 新生児指数， 気温， 国籍， 距離，
職種， 価格， 視聴率， 摂取カロリー， 血液型， 好きな食べもの。

(2) (1) で分類した各データはさらに細かく分類することができる。実際に，定量的データを **間隔尺度データ** (観測値間の差に意味があるもの)， **比率尺度データ** (観測値間の比率に意味があるもの) に分類せよ。また，定性的データを **名義尺度データ** (他と区別し分するもの)， **順序尺度データ** (順序には意味があるが間隔には意味がないもの) に分類せよ。

2 (☆☆☆)(定性的データに対する度数分布表)

次の表はある町に住む成人の母集団から，無作為に選んだ 500 人の血液型のデータである：

血液型	人数	相対度数 [%]
A 型	163	
B 型	146	
O 型	137	
AB 型	54	
計	500	100

このとき各血液型の相対度数を求め，データの中で最も現れる値 (**最頻値**) を求めよ。

3 (☆☆☆)(定量的データに対する度数分布表)

次の表は，高速道路のある地点における，走行中の自動車 500 台の時速 km/h を計測したものである。

時速	27.5 ~ 32.5	32.5 ~ 37.5	37.5 ~ 42.5	42.5 ~ 47.5	47.5 ~ 52.5	52.5 ~ 57.5	57.5 ~ 62.5
台数	8	32	66	89	119	141	45

このとき，度数分布表とヒストグラムを作成し，それからわかることを述べよ。

4 (☆☆☆)(無作為抽出)

ある母集団から無作為にサンプルが抽出された標本のことを **無作為標本** という。この無作為標本によって帰納的にその母集団を推察することができる。この無作為抽出について，次の問いに答えよ：

ある地域に住んでいる 100,000 人の中から 5,000 人を無作為に抽出して、ある感染症に罹患しているか否か検査を行うこととなった。この無作為に抽出する手順について、簡単に述べよ。

5 (★★★)(平均値と中央値の違いの例)

プロ野球のある球団における選手の年棒データに対して、平均値は球団側にとって有益であるが、選手側にとってあまり有益ではなく、反対に中央値は球団側にとってはあまり有益ではないが、選手側にとっては有益である。この理由を述べよ。

6 (★★★)(分散に関する等式)

N 個の実数値のデータ x_1, x_2, \dots, x_n に対して、 \bar{x} をその平均とする。このとき分散 s^2 について次の等式が成り立つことを示せ：

$$s^2 = \overline{x^2} - \bar{x}^2.$$

ただし、 $\overline{x^2} := \frac{1}{N} \sum_{i=1}^N x_i^2$.

7 (★★★)(Cauchy-Schwarz の不等式)

a_i, b_i ($i = 1, \dots, n$) を実数とする。任意の実数 t に対して $\sum_{i=1}^n (a_i t + b_i)^2 \geq 0$ が成り立つことから、次の Cauchy-Schwarz の不等式を導け：

$$\sum_{i=1}^n a_i b_i \leq \left(\sum_{i=1}^n a_i^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^n b_i^2 \right)^{\frac{1}{2}}$$

8 (★★★)(相関係数の範囲)

s_x, s_y をそれぞれ変数 x , 変数 y のデータに関する標準偏差とする。このとき、相関係数 $r_{xy} := \frac{s_{xy}}{s_x s_y}$ は $-1 \leq r_{xy} \leq 1$ を満たすことを、Cauchy-Schwarz の不等式を用いることで示せ。

9 (★★★)(相関の調査)

圧延加工したアルミニウム板の高度と抗張力を測定して以下の表のデータを得た。硬度と抗張力の相関を調べよ。ただし、計算にあたり電卓、スマホ、PC を用いてもよいとする。

硬度 x (H_g)	25	26	25	25	26	27	25	25	28	27
抗張力 y (kg/m^2)	14.2	15.9	13.7	14.3	15.1	17.6	15.9	16.0	18.0	17.8