

機械学習・ディープラーニングのための
基礎数学講座 確率・統計

SkillUP AI

データ分析の基礎 問題

問題 1：統計学の基礎

次のデータはそれぞれ質的データ、量的データのどちらに分類されるか？

1. 年収
2. 住宅のタイプ
3. 人の名前
4. マラソンの順位
5. 物体の密度

問題 2 : データの可視化

データセット: 3, 5, 7, 7, 13, 15, 19, 19, 20, 21, 25
に対して箱ひげ図を描け

問題 3 : 量的データの中心

データセット : 3 , 1 , 5 , 0 , 1 , 2 に対して

(1) 標本平均を求めよ

(2) 中央値を求めよ

(3) 最頻値を求めよ

問題 4 : 量的データのばらつき・広がり

データセット(3,1,5,0,1)に対して

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad s = \sqrt{s^2} \quad (\bar{x} = \text{平均})$$

を用いて、不偏分散 s^2 と不偏標準偏差 s を求めよ。

問題 5 : 二変数間の関連の強さ

データセット : $(0, 30), (2, 50), (7, 10)$ に対して

$$S_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

(\bar{x} と \bar{y} はそれぞれ x と y の標本平均)

を用いて不偏共分散 S_{xy} を求めよ。

問題 6 : 正規化

データ	x_1	x_2	x_3
値	0	3	6
偏差	-3	0	3
偏差 ²	9	0	9

平均 : 3

不偏分散 : $\frac{1}{3-1} (9 + 0 + 9) = 9$

不偏標準偏差 : 3

- (1) 各 x を標準化して、 z_1, z_2, z_3 を得よ
- (2) z の平均を求めよ
- (3) z の分散を求めよ

演習

演習 1：総合問題

データ：(6, 5), (6, 7), (7, 9), (7, 9), (9, 10)に対して第一成分、第二成分をそれぞれ x, y とする

- (1) x の標本平均を求めよ
- (2) y の標本平均を求めよ
- (3) x の不偏分散・不偏標準偏差を求めよ
- (4) y の不偏分散・不偏標準偏差を求めよ
- (5) x, y の不偏共分散を求めよ
- (6) x, y の相関係数を求め、相関関係の強弱について述べよ
- (7) x を標準化し、新たに得られたデータ z_x の平均を求めよ
- (8) データ z_x の分散を求めよ

必要ならば以下を使い

$$\sqrt{2} = 1.414$$

$$\sqrt{3} = 1.732$$

宿題

宿題 1 : 総合問題

データ : $(2, 20), (6, 20), (6, 15), (10, 25)$ に対して第一成分、第二成分をそれぞれ x, y とする

- (1) x の標本平均を求めよ
- (2) y の標本平均を求めよ
- (3) x の不偏分散・不偏標準偏差を求めよ
- (4) y の不偏分散・不偏標準偏差を求めよ
- (5) x, y の不偏共分散を求めよ
- (6) x, y の相関係数を求め、相関関係の強弱について述べよ
- (7) x を標準化し、新たに得られたデータ z_x の平均を求めよ
- (8) データ z_x の分散を求めよ