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

1章

データ分析の基礎

解答

演習問題

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：量的データ

2.住宅のタイプ：質的データ

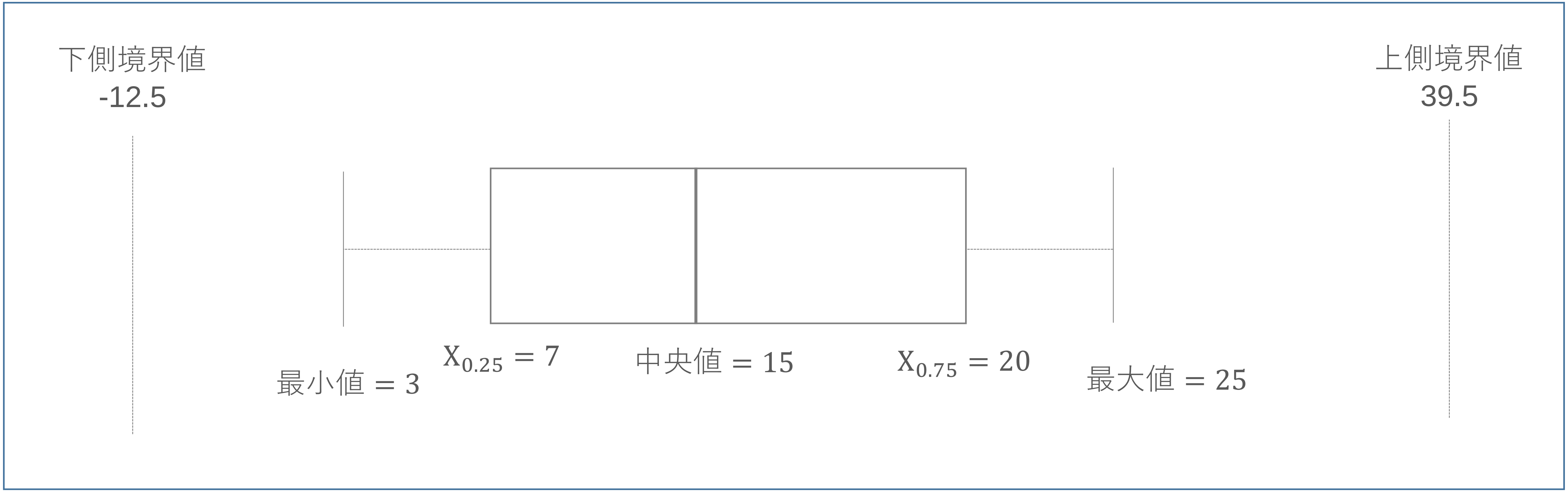
3.人の名前：質的データ

4.マラソンの順位：質的データ

5.物体の密度：量的データ

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

(1) 標本平均を求めよ

$$\frac{1}{n}\sum x_i = \frac{3+1+5+0+1+2}{6} = 2$$

(2)中央値を求めよ

小さい順に並べ替えたデータは (0 ,1 ,1 ,2 ,3, 5) で、その中央にある値は 1 と 2

中央値はその平均値である (1 + 2) / 2 = 1.5

(3)最頻値を求めよ

最も度数（頻度）の大きい値は 1 （度数＝ 2）

よって最頻値は 1

x_i	0	1	2	3	4	5
度数	1	2	1	1	0	1

問題４：量的データのばらつき・広がり

データの数nが 5でデータセットが（３,1, 5 ,0 ,1）の場合 ：

平均値は $\bar{x} = \frac{3+1+5+0+1}{5} = 2$ なので不偏分散 s^2と不偏標準偏差 s は

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

$$= \frac{(3-2)^2 + (1-2)^2 + (5-2)^2 + (0-2)^2 + (1-2)^2}{5-1}$$

$$= \frac{1^2 + (-1)^2 + 3^2 + (-2)^2 + (-1)^2}{5-1} = \frac{16}{4} = 4$$

$$s = \sqrt{s^2} = \sqrt{4} = 2$$

問題５：二変数間の関連の強さ

$$\bar{x} = \frac{\sum x_i}{3} = \frac{0+2+7}{3} = 3$$

$$\bar{y} = \frac{\sum y_i}{3} = \frac{30+50+10}{3} = 30$$

$$S_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{(0-3)(30-30)+(2-3)(50-30)+(7-3)(10-30)}{3-1}$$

$$= \frac{0-20-80}{2} = -50$$

問題６：正規化

(1) 各xを標準化して、z_1, z_2, z_3を得よ

$$z_1 = \frac{0-3}{3} = -1 \qquad z_2 = \frac{3-3}{3} = 0 \qquad z_3 = \frac{6-3}{3} = 1$$

(2) zの平均を求めよ

$$\mu_z = \frac{-1+0+1}{3} = 0$$

(3) zの分散を求めよ

$$\sigma_z = \frac{1^2+0^2+1^2}{3-1} = 1$$

標準化されたデータ	z_1	z_2	z_3
値	-1	0	1
偏差	-1	0	1
偏差^2	1	0	1

演習解答

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

$$\bar{x} = \frac{\sum x_i}{n} = \frac{6+6+7+7+9}{5} = 7$$

(2) yの標本平均を求めよ

$$\bar{y} = \frac{\sum y_i}{n} = \frac{5+7+9+9+10}{5} = 8$$

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

(3) xの不偏分散・不偏標準偏差を求めよ

$$s_x^2 = \frac{1}{5-1}\sum_{i=1}^{n}(x_i - 7)^2 = \frac{1}{4}\{(6-7)^2 + (6-7)^2 + (7-7)^2 + (7-7)^2 + (9-7)^2\} = \frac{6}{4}$$

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{6}{4}} = \frac{\sqrt{6}}{2}$$

(4) yの不偏分散・不偏標準偏差を求めよ

$$s_y^2 = \frac{1}{5-1}\sum_{i=1}^{n}(y_i - 8)^2 = \frac{1}{4}\{(5-8)^2 + (7-8)^2 + (9-8)^2 + (9-8)^2 + (10-8)^2\} = \frac{16}{4} = 4$$

$$s_y = \sqrt{s_y^2} = \sqrt{4} = 2$$

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

(5) x, yの不偏共分散を求めよ

$$S_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{5 - 1}$$

$$= \frac{1}{4}\{(6-7)(5-8) + (6-7)(7-8) + (7-7)(9-8) + (7-7)(9-8) + (9-7)(10-8)\}$$

$$= \frac{1}{4}(3 + 1 + 4) = 2$$

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

(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

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{2}{\frac{\sqrt{6}}{2} * 2} = \frac{2}{\sqrt{6}} = \frac{\sqrt{6}}{3} \approx 0.816$$

r_{xy}が0.816と1に近いので正の相関がある

$$\sqrt{2} = 1.414$$
$$\sqrt{3} = 1.732$$

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

(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

$$z=\frac{x-\mu}{\sigma}$$

μ： xの平均

σ: xの標準偏差

$$z_1=\frac{x_1-\bar{x}}{s_x}=\frac{6-7}{\frac{\sqrt{6}}{2}}=-\frac{\sqrt{6}}{3}$$

$$z_2=\frac{x_2-\bar{x}}{s_x}=\frac{6-7}{\frac{\sqrt{6}}{2}}=-\frac{\sqrt{6}}{3}$$

$$z_3=\frac{x_3-\bar{x}}{s_x}=\frac{7-7}{\frac{\sqrt{6}}{2}}=0$$

$$z_4=\frac{x_4-\bar{x}}{s_x}=\frac{7-7}{\frac{\sqrt{6}}{2}}=0$$

$$z_5=\frac{x_5-\bar{x}}{s_x}=\frac{9-7}{\frac{\sqrt{6}}{2}}=\frac{2\sqrt{6}}{3}$$

$$\bar{z}_x=\frac{\sum z_i}{n}=\frac{-\frac{\sqrt{6}}{3}-\frac{\sqrt{6}}{3}+0+0+\frac{2\sqrt{6}}{3}}{5}=0$$

標準化を行なっているので平均が0になるのは当然！

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

(8) データz_xの分散を求めよ

$$s_z^2 = \frac{1}{5-1}\sum_{i=1}^{n}(z_i - 0)^2 = \frac{1}{4}\left\{\left(\frac{-\sqrt{6}}{3}\right)^2 + \left(\frac{-\sqrt{6}}{3}\right)^2 + 0^2 + 0^2 + \left(\frac{2\sqrt{6}}{3}\right)^2\right\}$$

$$= \frac{1}{4}\left(\frac{6}{9} + \frac{6}{9} + 0 + 0 + \frac{24}{9}\right) = \frac{1}{4} * \frac{36}{9} = 1$$

標準化を行なっているので分散が1になるのは当然！

宿題解答

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

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2+6+6+10}{4} = \frac{24}{4} = 6$$

(2) yの標本平均を求めよ

$$\bar{y} = \frac{\sum y_i}{n} = \frac{20+20+15+25}{4} = \frac{80}{4} = 20$$

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

(3) xの不偏分散・不偏標準偏差を求めよ

$$s_x^2 = \frac{1}{4-1}\sum_{i=1}^{n}(x_i - 6)^2 = \frac{1}{3}\{(2-6)^2 + (6-6)^2 + (6-6)^2 + (10-6)^2\} = \frac{32}{3}$$

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{32}{3}} = \frac{4\sqrt{6}}{3}$$

(4) yの不偏分散・不偏標準偏差を求めよ

$$s_y^2 = \frac{1}{4-1}\sum_{i=1}^{n}(y_i - 20)^2 = \frac{1}{3}\{(20-20)^2 + (20-20)^2 + (15-20)^2 + (25-20)^2\} = \frac{50}{3}$$

$$s_y = \sqrt{s_y^2} = \sqrt{\frac{50}{3}} = \frac{5\sqrt{6}}{3}$$

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

(5) x, yの不偏共分散を求めよ

$$S_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{4 - 1}$$

$$= \frac{1}{3}\{(2 - 6)(20 - 20) + (6 - 6)(20 - 20) + (6 - 6)(15 - 20) + (10 - 6)(25 - 20)\}$$

$$= \frac{1}{3}(0 + 0 + 0 + 20) = \frac{20}{3}$$

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

(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

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\frac{20}{3}}{\frac{4\sqrt{6}}{3} * \frac{5\sqrt{6}}{3}} = \frac{\frac{20}{3}}{\frac{120}{9}} = \frac{1}{2} = 0.5$$

r_{xy}が0.5と1に近いので正の相関がある

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

(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

$$z = \frac{x - \mu}{\sigma}$$

μ： xの平均

σ:　xの標準偏差

$$z_1 = \frac{x_1 - \bar{x}}{s_x} = \frac{2 - 6}{\frac{4\sqrt{6}}{3}} = \frac{-\sqrt{6}}{2}$$

$$z_2 = \frac{x_2 - \bar{x}}{s_x} = \frac{6 - 6}{\frac{4\sqrt{6}}{3}} = 0$$

$$z_3 = \frac{x_3 - \bar{x}}{s_x} = \frac{6 - 6}{\frac{4\sqrt{6}}{3}} = 0$$

$$z_4 = \frac{x_4 - \bar{x}}{s_x} = \frac{10 - 6}{\frac{4\sqrt{6}}{3}} = \frac{\sqrt{6}}{2}$$

$$\bar{z}_x = \frac{\sum z_i}{n} = \frac{\frac{-\sqrt{6}}{2} + 0 + 0 + \frac{\sqrt{6}}{2}}{5} = 0$$

標準化を行なっているので平均が0になるのは当然！

宿題１：総合問題

データ：$(12, 30), (8, 20), (5, 20), (15, 50)$に対して第一成分、第二成分をそれぞれ$x, y$とする

(8) データz_xの分散を求めよ

$$s_z^2 = \frac{1}{4-1}\sum_{i=1}^{n}(z_i - 0)^2 = \frac{1}{3}\left\{\left(\frac{-\sqrt{6}}{2}\right)^2 + 0^2 + 0^2 + \left(\frac{\sqrt{6}}{2}\right)^2\right\}$$

$$= \frac{1}{3}\left(\frac{6}{4} + 0 + 0 + \frac{6}{4}\right) = \frac{1}{3} * \frac{12}{4} = 1$$

標準化を行なっているので分散が1になるのは当然！