演習問題

復習問題

1 高校の数学の復習

次の足し算の内容を、シグマ記号を用いて書きなさい。

(1) $x_1 + x_2 + x_3 + x_4$
(2) $y_1 + y_2 + \ \dots \ + y_n$
(3) $x_1 y_2 + x_2 y_3 + x_3 y_4 + x_4 y_5$
(4) $a y_8 + a y_9 + a y_{10} + a y_{11}$ $= a\,(y_8 + y_9 + y_{10} + y_{11})$
(5) $a_1 x_1 + a_2 x_2 + a_3 x_3 + \cdots + a_n x_n$

1 (1) $\sum_{i=1}^{4} x_i$
(2) $\sum_{i=1}^{n} y_i$
(3) $\sum_{i=1}^{4} x_i y_{i+1}$
(4) $\sum_{i=8}^{11} a y_i = a \sum_{i=8}^{11} y_i$
(5) $\sum_{i=1}^{n} a_i x_i$

基礎問題

2 記法

問１ 次の記号が何を表すか、数式を使わず答えなさい。

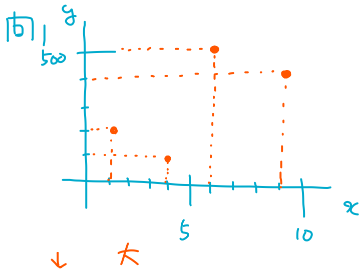
(1) $x_{(1)}$ ： x の最小値（小さい順で最小）
(2) x_1 ： x の最初のデータ
(3) $\bar{x}$ ： x の平均
(4) $Q_3 - Q_1$ ： 四分位範囲
(5) s_x ： x の標準偏差
(6) s_x^2 ： x の分散
(7) s_{ab} ： a と b の共分散
(8) $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ ： x と y の共分散
(9) $\frac{1}{n}\sum_{i=1}^{n}|u_i - \bar{u}|$ ： u の平均偏差

問２ 問１の記法を参考に、次の概念を数学記号や数式で表現しなさい。

(1) 偏差 ： $x_i - \bar{x}$ → 絶対偏差 $|x_i - \bar{x}|$
(2) （算術）平均 ： $\frac{1}{n}\sum_{i=1}^{n} x_i$
(3) 分散 ： $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
(4) 範囲 ： $x_{(n)} - x_{(1)}$
(5) 標準偏差 ： $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
(6) 相関係数 ： $s_{xy} \,/\, (s_x s_y)$ ← $\frac{s_{xy}}{s_x s_y}$
(7) 中央値（サンプルサイズが奇数の時） ： $x_{(\frac{n+1}{2})}$
(8) 中央値（サンプルサイズが偶数の時） ： $\frac{1}{2}\left\{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}\right\}$
†(9) 単回帰における切片
†(10) 単回帰における係数

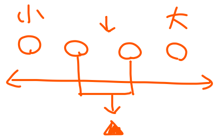
3 基本的な統計量

次のデータに基づき、以下の問いに答えなさい。



ID	X	Y
1	4	100
2	1	200
3	6	500
4	9	400

$\div \frac{1}{2} = 5$



問１　横軸を変数X、縦軸を変数Yに対応させた散布図を書き、各データをプロットしなさい。

問２　指示された統計量を計算しなさい。

(1) X の標本平均　$\frac{1}{4}\sum_{i=1}^{4} x_i = \frac{1}{4}(4+1+6+9) = 5$

(2) X の標本中央値　$\frac{1}{2}(4+6) = 5 = \frac{1}{2}(x_{(2)} + x_{(3)})$

(3) X の標本最頻値　ありません

(4) X の標本平均偏差　$\frac{1}{4}\sum_{i=1}^{4} |x_i - \bar{x}| = \frac{1}{4}(1+4+1+4) = 2.5$

(5) X の標本四分位範囲　$x_{(3)} - x_{(2)} = 6-4 = 2$

(6) X の標本範囲　$x_{(4)} - x_{(1)} = 9-1 = 8$

(7) X の標本分散　$\frac{1}{4}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{4}(1^2+4^2+1^2+4^2) = \frac{17}{2}$

(8) Y の標本範囲　$y_{(4)} - y_{(1)} = 500 - 100 = 400$

(9) X の標本標準偏差　$\sqrt{\frac{17}{2}} = \frac{\sqrt{34}}{2}$

(10) Y の標本標準偏差　$\sqrt{\frac{1}{4}\sum_{i=1}^{4} (y_i - \bar{y})^2} = \sqrt{\frac{1}{4}\binom{200^2+100^2+200^2+100^2}{(4+1+4+1)\times 10000}} = \frac{1}{2}\times 100\sqrt{10} = 50\sqrt{10}$

(11) X と Y の共分散　$\frac{1}{4}\sum_{i=1}^{4} (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{4}(\underbrace{1\times 200}_{200} + \underbrace{4\times 100}_{400} + \underbrace{1\times 200}_{200} + \underbrace{4\times 100}_{400}) = 300.$

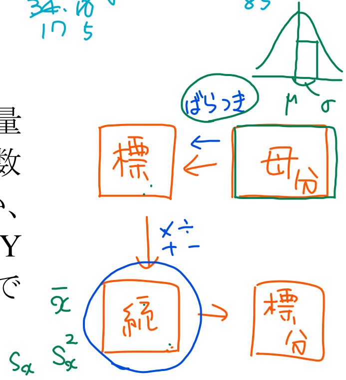
(12) X と Y の相関係数　$\frac{S_{xy}}{S_x S_y} = \frac{300}{\frac{\sqrt{34}}{2}\times 50\sqrt{10}} = \frac{12}{\sqrt{34}\sqrt{10}} = \frac{12}{34\cdot 10}\sqrt{34\times 10} = \frac{6}{85}\sqrt{85}$

4 100
1 200
1 500
9 400

4 100
1 200
6 500
9 400

4 統計量

次のデータに基づき、以下の問いに答えなさい。



問１　ID = *1* 番の人から ID = *n* 番の人に対して、X と Y という量を調査し、次のような表を得た。この時、次に与えられる数式で定義される量は統計量か、それとも統計量ではないか、答えなさい。ただし、x_i, y_iは観測された i 番目の人の X, Y に関する測定値を表し、$\bar{x}, \bar{y}$はそれぞれ観測値の標本平均である。また、(2)の a は(1)の a を示す。

ID	X	Y
1	x_1	y_1
2	x_2	y_2
⋮	⋮	⋮
n	x_n	y_n

$= \triangle_1$
$= \triangle_2$
$= \triangle_n$

$y_1 = f(y_1)$
$= y_1 \times 1$

(1) ○
(2) ○
(3) ○
(4) ○
(5) ○
(6) ○

共分散(統)　分散(統)　yの平均(統)　統　xの平均(統)

(1) $a = \frac{S_{xy}}{S_x^2}$ (2) $b = \bar{y} - a\bar{x}$ (3) y_1
(4) $\sum_{i=1}^{n}(y_i - x_i)$ (5) $\bar{y} - \bar{x}$ (6) $y_1 - x_1$

「自分の母親の年」.

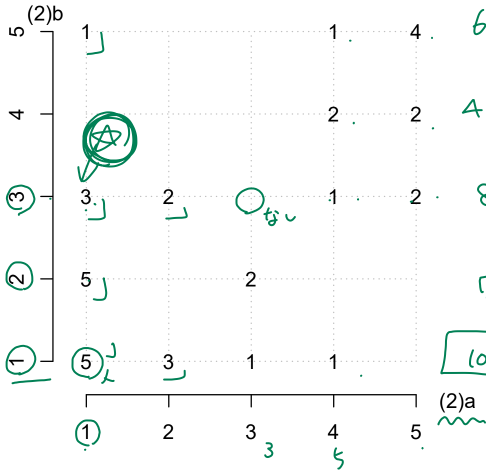
問２ 授業で扱ったもの以外に、統計量の定義に当てはまるものを三つ自分で考え、答えなさい。
問３ 統計量とは言えないものを三つ自分で考え、答えなさい。 例: π e 3.6 n. (3.14 ≒2.7.. 定数 ネイピア数 μ, σ^2)

○ $\bar{x}$: × μ
○ S_x : × σ_x

5 中心を表す統計量

普通、英語では(1) a, b が示すように縦続接続詞 that の使用は随意的である。だが、従属節内の主語を問う場合には、that を発音できないとされる。すなわち、(2)a とは異なり、(2)b は非文だとされる。

(1) a. You think [she came to the party].
b. You think [that she came to the party].
(2) a. Who do you think [came to the party]?
b. Who do you think [that came to the party]?



6人 4人 8人 7人 17人 10人 10人

この対比は、しかし日本の英語教育ではあまり強調されない。どのくらい日本人大学生がこの違いを知っているかを調べるために、東京都出身の大学生 35 人に、(2)a, b の文を、1（とても不自然）から 5（とても自然）という 5 段階評価をつけてもらった。次の散布図はこの結果を示し交点の数字はその組み合わせを答えた人の人数を表している（※これは架空のデータです）。 18番

1×14 + 2×5 + 3×3 + 4×5 + 5×8　14 10 9 20 40　93 ÷ 35 = $\frac{93}{35}$

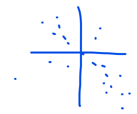
問１ この集団の文(2)a に対する容認度判断の算術平均を求めなさい。
問２ この集団の文(2)b に対する容認度判断の中央値を求めなさい。 $\tilde{x}_{(b)}$ = 3
問３ この集団の文(2)b に対する容認度判断のモードを求めなさい。 10人の"1".

6 二変量の関係を表す統計量

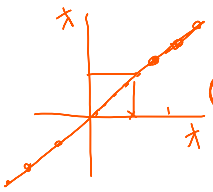
問１ 次の記述の中で、正しいものには○を、誤っているものには×をつけて正しく訂正しなさい。

(1) 標本分散は、各観測点から平均を引き二乗したものをすべて足し合せたもののことを指す。 ×) $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$

(2) 変数 x と変数 y とが負の相関関係にあるとき共分散は負の値を取る。 ○)



(3) 相関係数は、0 から 1 までの値を取り、1 に近づけば近づくほど変数間に強いつながりがあることを表す。 ×) −1



(4) 相関係数が同じなら、散布図も同じ値を示す。わけではない、 ×)

(5) x と y の共分散とは、$\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})$である。 ×)

(6) 共分散が最大になるのは、すべての観測データが右肩上がりの直線状に並ぶときである。 ○)

$-S_xS_y \le S_{xy} \le S_xS_y$

$\boxed{S_{xy} = S_xS_y}$

$\boxed{S_{xy} < S_xS_y}$

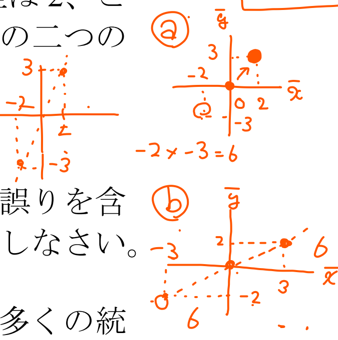
$S_{xy} = S_xS_y$

$S_{xx} = S_xS_x = S_x^2$

問２ 変数 x の標本標準偏差は 3、変数 y の標本標準偏差は 2、この二つの変数の標本共分散は 4 である。この時、この二つの変数の標本相関係数を求めなさい。

$S_x=3$　$S_y=2$　$S_{xy}=4$

$r = \frac{S_{xy}}{S_xS_y} = \frac{4}{3\times 2} = \frac{2}{3}$



ⓐ $2\times3=6$　$-2\times-3=6$　$\frac{1}{2}(6+6)$

ⓑ $\frac{1}{2}(4+8)$

7　正規分布の性質

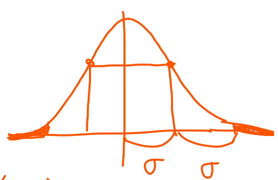
正規分布に関して、次の説明のうち正しいものには○、誤りを含むものには×を付け、不適切な部分を正しいものへ訂正しなさい。

× (1) 正規分布は、有標な分布と知られ、それゆえ多くの統計的な研究で母集団分布として想定される。 無標

○ (2) 正規分布は、左右が対称な釣鐘状の分布である。

○ (3) 正規分布は、その位置と広がりを規定するパラメータがある。

× (4) 正規分布に従うデータは、その平均を中心として 1 シグマ区間に、データのおよそ 95%が生じることが知られている。 2



(1.96)　σ　σ

8　中心極限定理

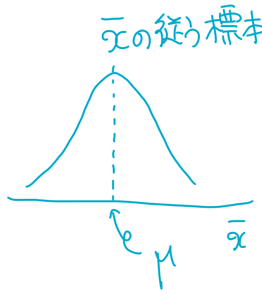
次の単語を使いながら、中心極限定理がどのようなものか、わかりやすく説明しなさい。

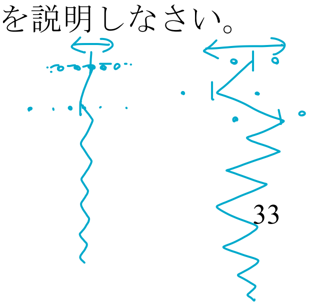
サンプルサイズ、母集団、標本、正規分布、母平均μ（母集団分布の平均）、母分散σ^2（母集団分布の分散）

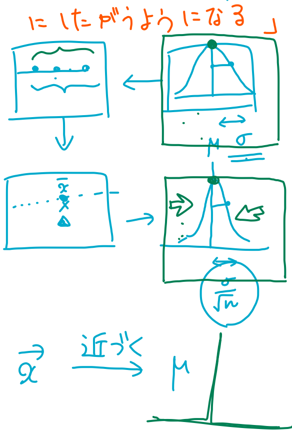
「サンプルサイズnが大きくなればなるほど、母集団分布の形状に関係なく、標本平均の標本分布は、正規分布 $N(\mu, \frac{\sigma^2}{n})$ にしたがうようになる」

9　不偏性

標本分布を例に、不偏性という統計量の性質を説明しなさい。







10 大数の法則

次の単語を使いながら、大数の法則がどのようなものか、わかりやすく説明しなさい。

サンプルサイズ、標本平均、正規分布、母平均μ（母集団分布の平均）、母分散σ^2（母集団分布の分散）

応用問題

11 重み付き平均

価格指数とは、複数の財の価格を加重平均して指数化したものであり、総合的な価格動向を把握するために利用されている。ラスパイレス価格指数は、個別価格指数を合成するときに、ウェイトとして基準時点の購入金額の割合を用いるものである。

次の表は、2015 年及び 2016 年における「牛肉」と「豚肉」の一世帯当たり（全国、二人以上の世帯）の年間購入数量（g）及び平均価格（円/100g）である。

	2015 年		2016 年	
	購入数量	平均価格	購入数量	平均価格
牛肉	6200	340.73	6,422	340.03
豚肉	19,865	149.57	20,418	144.30

資料：総務省「家計調査」

問１ 2015 年を基準年（指数を 100 とする）として、「牛肉」と「豚肉」の２種類の価格からラスパイレス価格指数を作成する場合、2016 年の指数はいくらか答えなさい。

問２ ラスパイレス価格指数の代表的な例として、消費者物価指数がある。消費者物価指数は、類・品目ごとにも作成されており、次の図は 1970 年から 2016 年までの魚介類の価格指数（2015 年を 100 とする）をプロットしたものである。魚介類の価格指数の変化率の図として、次の①から④の中から最も適切なものを一つ選び記号で答えなさい。

（2017 年度 11 月出題の「統計検定２級」より；一部改め）

数式問題

12　標本平均と式展開

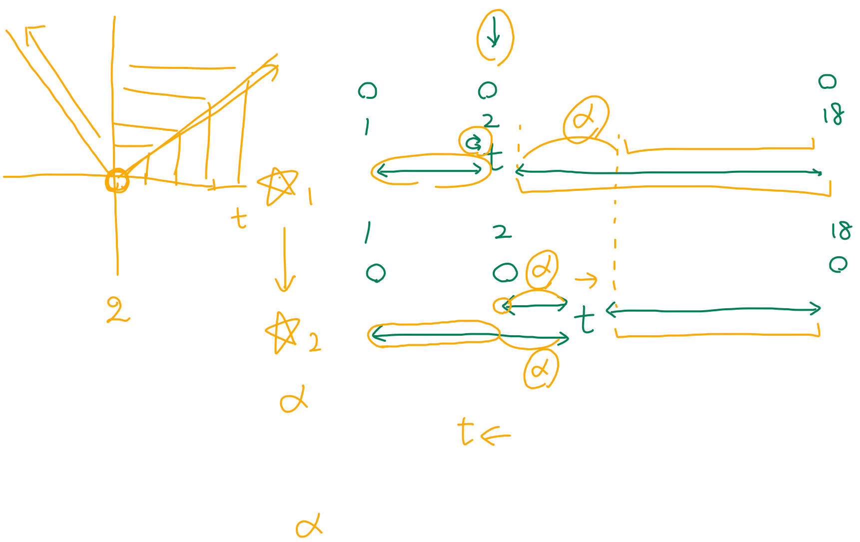
C を任意の定数とする。次の式が成り立つことを示しなさい。(ヒント：左の式を変形すると右の式になることを示せばよい)

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - C)^2 = (\bar{x} - C)^2 + \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

※この式変形は何回かこの先の授業で出てきます。

13　相関係数

$x_1, x_2, \dots, x_n$と$y_1, y_2, \dots, y_n$との相関係数をr_{xy}とする。$a, c > 0$とし、$z_i = ax_i + b$、$w_i = cy_i + d$とする。このとき、z と w の相関係数をr_{zw}とするとき、r_{zw}は、r_{xy}とどのような関係にあるかを説明しなさい。



$t=2$

$|1-t| + |2-t| + \cdots + |18-t|$

$= 1 + \cancel{2}0 + \cancel{18}16$

$= 17$

$t = 2 + \alpha$

$|1-(2+\alpha)| + |2-(2+\alpha)| + \cdots + |18-(2+\alpha)|$

$= 1+\alpha + \cancel{\alpha} + 16 - \cancel{\alpha}$

$= 17 + \alpha$

$$\boxed{\sum_{i=1}^{n} |x_i - t|}$$