

確認試験問題

データの分析

(配点 100点)

(日付) _____ 年 _____ 月 _____ 日

(開始)

--	--	--	--	--

 :

--	--	--	--

 ~ (90分) ~

--	--	--	--	--

 (終了)

注意事項

1. 上の日付の欄に、試験を行う日付を記入しなさい。
2. 上の開始と終了の欄に、試験開始予定の時刻とその90分後の時刻を、それぞれ記入しなさい。試験はその時間内に行われます。
3. 試験に関係の無い物の持ち込みは、原則として認められません。
4. 試験時間中は、アラーム機能以外での電子機器の使用は認められません。
5. 試験開始の時刻になるまで、この試験の問題を見てはいけません。
6. この問題冊子は全部で6ページあります。落丁、乱丁または印刷不鮮明の箇所があれば、監督者に知らせなさい。
7. この試験は5問で構成されています。解答用紙は第1問から第5問までに対応するものを用意しなさい。
8. 解答には、黒色鉛筆か、または黒色シャープペンシルを使用しなさい。
9. 解答用紙の指定欄に、試験名、氏名、学生番号を記入しなさい。
10. 問題ごとに、解答欄が指定されています。解答は、必ず解答用紙の指定された箇所に記入しなさい。
11. 解答用紙の解答欄に、関係のない文字、記号、符号などを記入してはいけません。また、解答用紙の欄外の余白には、何も書いてはいけません。
12. この問題冊子の余白は、書き込みに関しては自由に使用してもよいが、どのページも破棄してはいけません。
13. 試験時間中は、やむを得ない場合を除き、退場してはいけません。
14. 試験後は、問題用紙と解答用紙を自由に活用しなさい。

第 1 問

(配点 20)

ある学級の生徒 20 人が 10 点満点の試験を受けたところ、次のような結果であった。

表 1.1: 生徒と試験の得点

生徒	A	B	C	D	E	F	G	H	I	J
得点	1	2	7	8	6	4	4	6	7	8
生徒	K	L	M	N	O	P	Q	R	S	T
得点	9	7	3	6	7	5	6	8	10	6

得点のデータに関して、以下の問いに答えよ。必要ならば、 $2.2 < \sqrt{5} < 2.3$ であることを用いてよい。

- (1) このデータの平均値 μ を求めよ。
- (2) このデータの中央値 m を求めよ。
- (3) このデータの最頻値 M を求めよ。
- (4) このデータの分散 V を求めよ。
- (5) このデータの標準偏差 s を求めよ。結論には根号を用いてよい。
- (6) 得点が x である生徒の偏差値得点 z を

$$z = \frac{x - \mu}{s} \cdot 10 + 50$$

で定める。このとき、偏差値得点が 60 以上である生徒の人数を求めよ。

第 2 問

(配点 24)

図 2.1 は、ある場所の 16 日間にわたる各日の気温のデータをヒストグラムに、図 2.2 は、ある学校に属する生徒 12 人の体力テストの成績を箱ひげ図に、図 2.3 は、ある 13 人の集団の身長と体重を測定した結果を散布図にして表したものである。集計したデータは、いずれも整数値であったとする。また、図 2.1 のヒストグラムについて、各階級は 15°C 以上 18°C 未満、 18°C 以上 21°C 未満、..., 30°C 以上 33°C 未満のように区切っている。

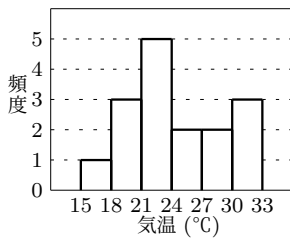


図 2.1

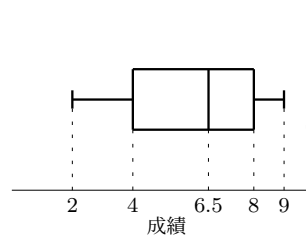


図 2.2

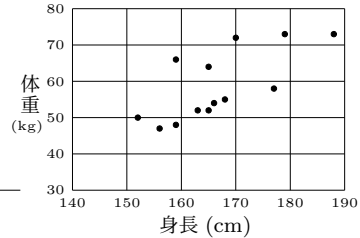


図 2.3

以下の (1) ~ (8) の記述に対して、それが正しいかどうかを理由をつけて判定せよ。

- (1) 図 2.1 から、気温 ($^{\circ}\text{C}$) のデータの範囲は 18 と確定する。
- (2) 図 2.1 から、気温 ($^{\circ}\text{C}$) のデータの中央値は必ず 21 以上 24 未満の範囲内にあるといえる。
- (3) 図 2.1 から、気温 ($^{\circ}\text{C}$) のデータの平均値は必ず 21 以上 24 未満の範囲内にあるといえる。
- (4) 図 2.2 から、成績のデータの範囲は 7 と確定する。
- (5) 図 2.2 から、成績のデータの平均値は必ず 6 以上であるといえる。
- (6) 図 2.2 から、成績が 7 の生徒が必ず存在するといえる。
- (7) 図 2.3 から、身長が 160 cm 以上、体重が 60 kg 以上の生徒が 3 人以上いることが分かる。
- (8) 図 2.3 に対応する身長 (cm) と体重 (kg) のデータの共分散を計算し、小数第二位を四捨五入すると 63.2 であった。このとき、身長 (cm) と体重 (kg) の相関係数は正であるといえる。

第 3 問

(配点 20)

問い A, B に答えよ。

A

- (1) 変数 x についての n 個のデータの値において, x のデータの平均値を \bar{x} , x のデータの各値をそれぞれ 2 乗したものの平均値を $\overline{x^2}$ とおくと, x の分散 v に関して,

$$v = \overline{x^2} - (\bar{x})^2$$

が成り立つ。 $n = 4$ の場合について, この等式が成り立つことを計算によって確かめよ。例えば, 4 つの値 x_1, x_2, x_3, x_4 について, その分散 v と $\overline{x^2} - (\bar{x})^2$ をそれぞれ計算し, それらから同じ式を得られることを確かめるとよい。

- (2) 2 つの集団 A, B があり, それぞれが果樹園で果物の収穫を行った。集団の人数と, 収穫した果物の個数について, 集団内の平均値と分散を調べたものを次の表に示した。

表 3.1: 集団と収穫した果物の個数

集団	人数	平均値	分散
A	36	29	31
B	54	24	36

このとき, A と B を合わせた集団全体について, 収穫した果物の個数の平均値と分散を求めよ。

B

- (3) 変数 x のデータの大きさが n であるとする。 a, b を実数の定数とし, 変数 x について, 変数 y を $y = ax + b$ と定める。このとき, 変数 x のデータの平均値を \bar{x} , 変数 y のデータの平均値を \bar{y} , 変数 x のデータの分散を v_x , 変数 y のデータの分散を v_y とおくと,

$$\bar{y} = a\bar{x} + b, \quad v_y = a^2 v_x$$

が成り立つ。 $n = 4$ の場合について, この等式が成り立つことを計算によって確かめよ。例えば, 変数 x についての 4 つの値 x_1, x_2, x_3, x_4 について, 対応する変数 y の値を y_1, y_2, y_3, y_4 とおき, それぞれデータの平均値と分散を求め, それらから関係式を得られることを確かめるとよい。

- (4) 次の 6 つの値からなるデータの平均値と分散を求めよ。

$$314.15926535, \quad 514.15926535, \quad 1014.15926535, \\ 764.15926535, \quad -285.84073465, \quad -135.84073465$$

第 4 問

(配点 20)

あなたは、シミュレーションソフトを使って仮想的な国家を構築できるとし、構築した国家では自動的に人々が経済活動を行うものとする。

いま、1つ国家を構築したとする。この国家には首都近辺に地方が8つあり、各地方の首都からの距離 (km) と、その地方で生活する人々の収入 (Mvy) が次の表に示した通りであったとする。ただし、収入の単位は仮想国家独自のものである。

表 4.1: 各地方の首都からの距離と収入

地方	A	B	C	D	E	F	G	H
距離 (km)	17	23	26	27	38	43	50	56
収入 (Mvy)	19	14	11	12	15	13	12	8

以下の問いに答えよ。

- (1) 距離 (km) の平均値, 分散をそれぞれ求めよ。
- (2) 収入 (Mvy) の平均値, 分散をそれぞれ求めよ。
- (3) 距離 (km) と収入 (Mvy) の共分散を求めよ。
- (4) 距離 (km) と収入 (Mvy) の相関係数を求めよ。
- (5) 仮に, B の収入 (Mvy) が 14 から 15 に増加し, E の収入が 15 から 14 に減少した場合, 距離 (km) と収入 (Mvy) の相関係数は (4) で求めた値より大きくなるか, 小さくなるか, 変化しないか, 理由をつけて判定せよ。

第 5 問

(配点 16)

問い A, B に答えよ。

A

「当たり」または「はずれ」のどちらかが出るくじを 10 回引いたところ、「当たり」が 8 回出た。このくじは「当たり」が「はずれ」より出やすいといえるか。仮説検定の考え方をを用い、有意水準 0.05 で考察せよ。ただし、各回のくじの結果は、別の回のくじの結果に影響を及ぼさないとする。また、表と裏の出る確率がどちらも $\frac{1}{2}$ である硬貨を 10 回投げたとき、ちょうど n 回表が出る確率 $P(n)$ を表 5.1 に示したため、必要ならばこれを用いてもよい。

表 5.1: n 回表が出る確率 $P(n)$

n	0	1	2	3	4	5
$P(n)$	$\frac{1}{1024}$	$\frac{10}{1024}$	$\frac{45}{1024}$	$\frac{120}{1024}$	$\frac{210}{1024}$	$\frac{252}{1024}$
n	6	7	8	9	10	
$P(n)$	$\frac{210}{1024}$	$\frac{120}{1024}$	$\frac{45}{1024}$	$\frac{10}{1024}$	$\frac{1}{1024}$	

B

ある企業が、年に 1 度行われる総会の後、従業員全員に対して、業務時間の快適さに関するアンケートを取った。「快適である」、「快適でない」の 2 つの選択肢を提示して回答を得たところ、「快適である」と回答したのは全体の $\frac{1}{3}$ であった。後日、その企業がデジタル化の一環で事務作業の一部を自動化するサービスを契約し、しばらくして従業員 8 人に同様のアンケートを取ったところ、「快適である」と回答したのは 6 人であった。この結果から、業務について「快適である」と回答した人は増えたといえるか。仮説検定の考え方をを用い、有意水準 7% で考察せよ。ただし、1, 2, 3, 4, 5, 6 の目が等確率 $\frac{1}{6}$ で出る立方体のさいころを 8 回投げたとき、3 以上の目がちょうど n 回出る確率 $P(n)$ を表 5.2 に示したため、必要ならばこれを用いてもよい。

表 5.2: 3 以上の目が n 回出る確率 $P(n)$

n	0	1	2	3	4
$P(n)$	$\frac{1}{6561}$	$\frac{16}{6561}$	$\frac{112}{6561}$	$\frac{448}{6561}$	$\frac{1120}{6561}$
n	5	6	7	8	
$P(n)$	$\frac{1792}{6561}$	$\frac{1792}{6561}$	$\frac{1024}{6561}$	$\frac{256}{6561}$	