

総合演習 (1) 問題

【問題 1】 次の各項は、統計学の観点からみて正しい（あるいは適切である）かどうかを答えよ。正しくない（適切でない）ときは、どういう点がどのように正しくない（適切でない）かを説明せよ。

1. 高校のクラスの卒業生の、現在の平均年収を調査するため、クラス会幹事が、クラス会のパーティの案内の返信はがきに年収を書く欄をつくり、回答を求めた。
2. 第 2 次大戦時、イギリス軍の爆撃機は、ドイツ軍の迎撃戦闘機の攻撃により大きな損害を受けた。そこで、イギリス軍では、帰還した爆撃機乗務員に「機体のどちらの方向から攻撃を受けたか」を聴取し、その方向の防備を厚くするなどの対策を練った。
3. 全国の食料品店から、いくつかの店を無作為抽出して調査したところ、全品目のうち 3% について、内容量が表示されている量に足りない、という不正が発見された。したがって、消費者は平均して 3% 損をしていることになる。
4. 各県について婚姻数と死亡数を調べると、これらには負の相関があることがわかった。したがって、結婚と死亡にはなんらかの直接の医学的関係があると考えられる。

【問題 2】 次の「平均」に関する記述について、正しい（あるいは適切である）かどうかを答えよ。正しくない（適切でない）ときは、どういう点がどのように正しくない（適切でない）かを説明せよ。

1. 気象用語でいう「雲量」とは、快晴を 0、本曇りを 10 として空のうち雲の占める割合を表現する。平均は 5 程度である。したがって、一年間のうち多少雲がある曇りの日が一番多い。
2. 力士の能力は「心・技・体」の 3 項目で計られる。力士「甲の海」の評価は、10 点満点で心 6 点、技 7 点、体 7 点で平均 6.7 点である。力士「乙の花」の評価は、心 0 点、技 10 点、体 10 点で平均 6.7 点である。したがって、両力士の能力は同じである。
3. 「人生 50 年」という言葉があったように、100 年前の日本人の平均寿命は 50 歳程度であった。当時の女性は 5 人 6 人の子供を生むことは当たり前であったので、子育てが終わった後の人生はたった数年しかなかった。
4. 現在、東京証券取引市場の日経平均株価は 9,500 円程度である。市場には 1 株あたり 9,500 円程度で取引されている銘柄がたくさんある。

【問題 3】 5 人の生徒に英語と数学の試験を行なった。各生徒の得点は、(英語の点数, 数学の点数) の形で表すと、それぞれ (50, 60), (55, 55), (70, 75), (75, 90), (80, 80) であった。

1. 回帰直線を求め、各生徒の得点とともに散布図に表わせ。
2. 決定係数を求めよ。さらに、決定係数と回帰直線の関係を、本問を例にとって説明せよ。

【問題 4】 5 人の学生が、数学・英語の両科目の試験を受けたところ、その成績は下の表の通りであった。

1. これらの5人について、両科目の総合得点による順位をつけたい。そこで、5人の点の分散がもっとも大きくなるような総合得点を求め、5人の学生に順位をつけよ。
2. 5人の両科目の成績を、数学の点を横軸、英語の点を縦軸とした散布図上に図示せよ。さらに、各人を表す点の傍に(1)で求めた総合得点を書き込め。

学生番号	数学の点数	英語の点数
1	40	40
2	60	40
3	20	30
4	30	60
5	70	70

表 1: 問題 4 ・ 表

解答例

【問題1】

1. 年収の少ない人は、平均年収を書かないと思われるし、人は見栄をはるから正確なことを書かないおそれも高い。そもそも、同窓会の返信はがきで、年収のようなプライベートなことを尋ねてはいけない。
2. 無事帰還した乗務員は攻撃を免れたのであり、攻撃により損害をうけた爆撃機の大半は帰還しなかったであろう。だから、帰還した乗務員だけから攻撃について聴取しても、適切な対応はとれないと思われる（これは実話です）。
3. 不正が発見されたのは「全品目のうち3%の品目」についてであり、不足の量がいくらだったかは調査されていない。もし、残りの97%の品目の内容量が表示よりもずっと多いのなら、消費者は平均して損などしていない、ということになる。
4. 若い世代の人口が多いほど、死亡数が少なく結婚数は多い。つまり、死亡数と若年人口の間、結婚数と若年人口の間に直接関連があるのであって、死亡数と結婚数の間にあるのは見かけ上の相関である。

【問題2】

1. 雲量5は平均であって、「もっとも頻繁に現れる雲量」（モードという）と混同してはいけない。実は、雲量5の日数は一番少ない。つまり、晴れまたは雨（当然雲量10）の日は多いが、中途半端な曇りの日は少ない。このように、ヒストグラムのピークがひとつの分布（単峰性分布という）ではない場合、平均値とモードとは大きく異なることもある。
2. 「総合能力」を、いつも算術平均で表してよいとは限らない。伝統的な相撲観では「心・技・体」はいずれも欠けてはいけないもので、その場合乙の花はまったくだめということになる。このような時に用いる平均として「幾何平均」というものがある。これは、項目が3つなら3つの項目の点数を掛け合わせて3乗根を計算したものである。
あるいは、「心などどうでもよくて、技、体がよければ強いはず」という価値観もあるかもしれない。この場合、「心に1倍、技、体に5倍」という「重み」を定めて、 $(1 \times \text{心} + 5 \times \text{技} + 5 \times \text{体}) / (1 + 5 + 5)$ という平均を求める方法もある。これを重みつき平均という。
3. 平均寿命とは、「現在ある年齢の人が来年まで生きられる確率」がどの年齢層でも将来にわたって不変であるとしたときの、現在0歳の人が将来生きる人生の長さの平均を意味する。したがって、乳児死亡率が高いと平均寿命は短くなる。子育てが終わって現在40歳になっている人が、平均してあと何年生きるかを表す「40歳平均余命」は、乳児死亡率は関係ないので、 $(\text{平均寿命} - 40)$ とはだいぶ異なった値になる。
4. 東証日経平均株価は、東京証券取引所に上場されている銘柄のうち代表的な225銘柄の株価から算出される。しかし、その値は「225銘柄（額面50円換算の）株価の合計を225で割ったものではなく、24.656で割ったものである。したがって、9,500円前後の株価の銘柄はほとんどなく、ずっと安い株価の銘柄がほとんどである。

除数が225でないのは、「株式分割」というものがあるからである。株式分割とは、企業が株主に対して現在の保有株数に応じて無償で追加の株を割り当てることである。このとき、各株主の保

有する株数は増えるが、資産総額は増えないので、見かけ上株価が安くなる。例えば、株価 500 円の企業が、1 株に対して 1 株を追加して 2 株を割り当てる株式分割をおこなった場合、分割後の株価は半分の 250 円となる。しかし、このように株価が安くなったのは見かけ上のもので、会社の価値が下がったわけではないので、この分割によって平均株価が変化しないようにするため、除数を小さくして対応する。

【問題 3】

1. 英語の点数を x 、数学の点数を y で表し、生徒数を n とする。表 (左) の通り計算すると、 $n = 5$ 、 $\bar{x} = 66$ 、 $\bar{y} = 72$ 、 $\sum x_i^2 = 22450$ 、 $\sum x_i y_i = 24425$ であるから、回帰方程式を $y = a + bx$ とすると

$$b = \frac{24425 - 5 \cdot 66 \cdot 72}{22450 - 5 \cdot (66)^2} = 0.993, a = 72 - 0.993 \cdot 66 = 6.49 \text{ となる。 (散布図は、当日黒板に描いた通りで、略)}$$

2. 表 (右) のように計算すると、相関係数 r は $r = \frac{665}{\sqrt{670} \sqrt{830}} = 0.892$ となり、決定係数 $r^2 = 0.795$ となる。このことは、数学の点数が分布している理由を、「英語の点数と線形の関係にある」と説明することで、もとの数学の点数の分散の 79.5% が説明できることを示している。

生徒	英語 x	数学 y	x^2	xy
1	50	60	2500	3000
2	55	55	3025	3025
3	70	75	4900	5250
4	75	90	5625	6750
5	80	80	6400	6400
	平均 \bar{x} 66	平均 \bar{y} 72	$\sum x^2$ 22450	$\sum xy$ 24425

生徒	英語 x	数学 y	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
1	50	60	256	144	192
2	55	55	121	289	187
3	70	75	16	9	12
4	75	90	81	324	162
5	80	80	196	64	112
	\bar{x} 66	\bar{y} 72	計 670	計 830	計 665

【問題 4】

1. 求める総合得点は分散が最大になる総合得点で、これはすなわち第 1 主成分である。そこで、主成分分析を行う。問題のデータから、数学の平均=44、英語の平均=48、数学の分散=344、英語の分散=216、共分散=148 となる。これらからなる分散共分散行列の固有値をプリントの通り求めると 441.25 と 118.75 となる。大きいほうの固有値に対応する固有ベクトルは (0.836, 0.549) であり、よって第 1 主成分は「数学 $\times 0.836$ + 英語 $\times 0.549$ 」で求められる。番号 15 の学生について第 1 主成分を求めると、順に 55.4, 72.1, 33.2, 58.0, 97.0 となる。したがって、成績順位は 5, 2, 4, 1, 3 の順となる。
2. 下図の通り。

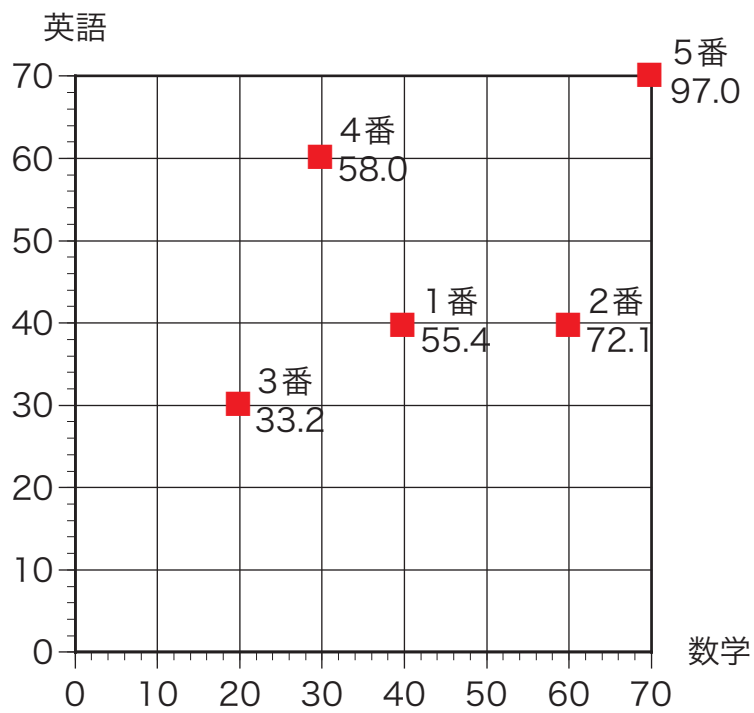


図 1: 問題 4