

解答例

1.

- (a) 「『公共事業へのこれ以上の投資は、財政への影響が深刻である』と思いますか」という質問と、「『公共事業へのこれ以上の投資は取りやめるべきである』と思いますか」という質問の、2つの質問がひとつに交じり合っている。いわゆるダブルバレル質問である。問題のような質問では、「公共事業へのこれ以上の投資は財政への影響が深刻であるが、それでも実施すべきである」「公共事業へのこれ以上の投資は、財政への影響はあまりないが、それでも実施すべきではない」という意見の持ち主は、回答することができない。
- (b) 問題が二者択一で答えられるものならば、「10%しか当たらない占い師」が告げたほうとは逆の選択肢を選べば90%当たっていることになり、「70%当たる占い師」が告げた選択肢よりも当たっている確率は大きい。
- (c) 葉や豆を直接食べるのではなく、淹れて飲むわけだから、同量の葉・豆を比較しても意味はなく、淹れた後の同量の液を比べなければならない。通常、紅茶の液はコーヒーの液よりもずっと薄い。
- (d) 電球の用途が、例えば家庭の電灯のように「切れたらその時点で取り換える」ようなものなら、平均寿命が長いほうの電球を選べばよい。しかし、交通信号機のように、常時使用し続けなければならない、故障の許されない装置の場合は、故障するよりも前に、ある期間ごとにいっせいに電球を交換するほうが安全である。この場合は、平均寿命よりも「最短保証寿命」が長い電球を選ぶほうが有利である。もし、A社の電球の寿命の分散が大きく、B社の電球の寿命の分散が小さいならば、最短保証寿命はB社の電球のほうが長いこともありうる。
- (e) 「死亡率と婚姻率の相関関係」と「死亡と婚姻の因果関係」は別のものである。また、この相関関係は「みかけ上の相関関係」であり、実際には、死亡率の高い県は高齢化が進み若年人口が少ないので婚姻率が低い、と考えられる。

2. (a) 「日本で肺がんで死亡する人が、1年間に何人いるか」のデータが必要である。もし、日本で1年間に肺がんで死亡する人が「たばこを吸わない人1人、たばこを1日に20本吸う人10人」なのであれば、肺がんによる死亡はきわめて珍しいということになり、他の病気への対策をとるほうが、国民の健康の向上に効果的である。もし、日本で1年間に肺がんで死亡する人が「たばこを吸わない人10万人、たばこを1日に20本吸う人100万人」なのであれば、肺がんによる死亡はきわめて重大な問題であり、たばこを禁止すれば死亡者数を大きく減らすことができる。
- (b) 変数 X と変数 Y の相関を考えると、 X が大きい個体では Y も大きいという傾向があることを「正の相関」といい、 X が大きい個体では Y は小さいという傾向があることを「負の相関」という。また、その傾向が明確で、散布図上で個体を表す点が一直線に近い形状に分布している場合は「強い相関」といい、その傾向が明確でなく、散布図上で個体を表す点が一直線状ではなく広がりをもって分布している場合は「弱い相関」という。
 - (c) 変数 Y から変数 X への回帰を考えると、 X と Y の決定係数は、もとの Y の分散に比べて回帰直線に対しての Y のばらつきがどれだけ減少しているか、を表している。例えば、決定係数が0.7であれば、もとの Y の分散のうち70%は、線形単回帰モデルをあてはめることで「ただ単にばらついているのではなく、回帰直線上に並んでいる」という説明ができる、という意味になる。

3.
 - 1) グラフで示されている「0.5万円程度の増加」は、有効な増加なのか、よくわからない。それにもかかわらず、縦軸の下端をゼロにしないこともあわせて、変化が大きく誇張されている。
 - 2) 時間に伴う変化を示すときは、折れ線グラフの方が、変化をたどることを表現できるので適切である。
4.
 - 1) 満足度が高いほど数字が大きいことから、グラフでは「5: 大変満足」を右端、「1: 大変不満」を左端に配置するほうが、直観的に読み取れる。
 - 2) 色の使い方が満足度と関連していない。満足度が上がるにつれて寒色から暖色になっていく、のような工夫が必要である。
 - 3) 人によっては色による区別がしにくい場合もあるので、棒グラフの棒の下に満足度の数字を書いておくほうがのぞましい。
5.
 - (a) 感度は「本当に病気の人のうち、検査で陽性となった人の割合」で、 $9000/(9000+1000) = 0.9$ である。また、特異度は「本当は病気でない人のうち、検査で陰性となった人の割合」で、 $19000/(1000+19000) = 0.95$ である。
 - (b) 感度だけを高くすればよいのであれば、検査をせずにすべての被験者を陽性とすれば、感度は1すなわち100%となる。しかし、そのような検査には意味はなく、「本当は病気でない人」を正しく陰性と判定する（誤って陽性としない）ことも、検査の重要な能力である。したがって、すぐれた検査であるためには、感度だけでなく特異度も高くなくてはならない。
6.
 - (a) 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。 i 番目の受験者の点数を (x_i, y_i) とし、 x, y の平均をそれぞれ \bar{x}, \bar{y} 、 x の分散を σ_x^2 、 x, y の共分散を σ_{xy} とし、表 A1 の通り計算すると、 $n = 5$ 、 $\bar{x} = 66$ 、 $\bar{y} = 72$ 、 $\sigma_{xy} = 665/5 = 133$ 、 $\sigma_x^2 = 670/5 = 134$ であるから、回帰方程式を $y = a + bx$ とすると

$$b = \frac{133}{134} = 0.993, a = 72 - 0.993 \cdot 66 = 6.46$$
 となる。散布図と回帰直線は図 A2 の通りである。

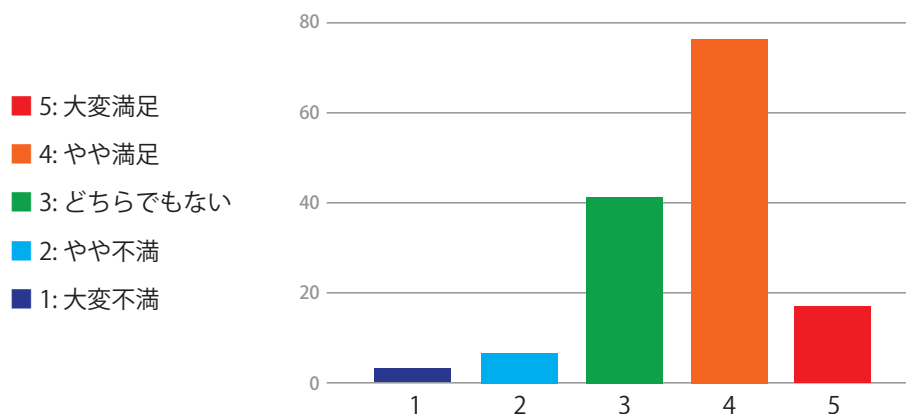


図 A1: 問題 4 の図

(解説) 回帰直線のグラフを描くためには、直線がどこを通るかを確かめてみます。 $x = 50$ を回帰方程式 $y = 6.46 + 0.993x$ に代入すると、 $y = 56.11$ となりますから、回帰直線は $(50, 56.11)$ を通ります。

先に散布図を描いておいて、回帰直線がだいたいどんな感じかの見当をつけておきましょう。計算で回帰直線を求めた後で、その計算が間違っていないかどうか確かめることができます。

- (b) 表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 $r^2 = 0.796$ となる。このことは、数学の点数の分布を、「数学の点数は英語の点数から回帰方程式によって推定できる」と表すことで、もとの数学の点数の分散の 79.6% が説明できることを示している。

表 A1: 問題 6 の表

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}	\bar{y}	計	計	計
	66	72	670	830	665

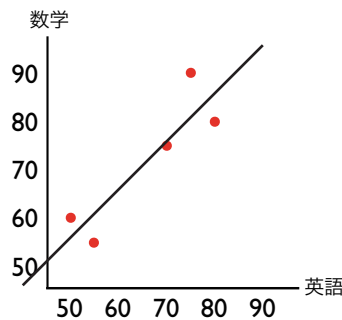


図 A2: 問題 6 : 散布図と回帰直線