

Free Anote

統計データ解析 過去問

<2010 年度>試験問題

1. 4名の生徒が2つの異なるテストAとテストBを受けた。それぞれのテストに対する正解数 x_A と x_B [問/10問中]を表1に示す。2つのテスト問題で難易度に違いがあるかどうかを、t検定を用いて検定したい。以下の問いに答えよ。ただし必要であれば $\sqrt{2}=1.414$ 、 $\sqrt{3}=1.732$ 、 $\sqrt{5}=2.236$ を用いよ。

表1 正解数[問/10問中]

生徒	x_A	x_B	$d = x_A - x_B$
1	10	9	1
2	7	7	0
3	8	6	2
4	9	8	1
平均	8.50	7.50	
標準偏差	1.29	1.29	

- (1) 正解数の差 d の平均値 \bar{d} を求め、小数点以下1桁で記せ。
- (2) 正解数の差 d の標準偏差 s を求め、小数点以下1桁で記せ。このときの標準偏差は母集団の普遍推定値として計算すること。つまり表を母集団の標本とみなして、母集団に対する標準偏差 σ の推定値 s を求めること。
- (3) 正解数の差 d の標準偏差を求め、小数点以下1桁で記せ。標準偏差とは、同じ標本数の標本抽出を繰り返したときの標本の平均値の分布(標本分布)における標準偏差のことである。
- (4) 自由度 df を答えよ。
- (5) 以上から $t=2.449$ を得た。また、前問の自由度 df に対し有意水準 $\alpha=0.05$ (5%)として、t分布表より、 $t(\text{片側})=2.353$ 、 $t(\text{両側})=3.182$ を得た。2つのテストの難易度に有意な差があるかどうか判定した。検定結果として適切な選択肢を○で囲め。

[$t(\text{片側})=2.353 < 2.449$ 、 $2.449 < 3.182 = t(\text{両側})$]であり、
難易度には違いが[ある、ない]

(6)テスト B の方がテスト A よりも難易度が高いと言えるだろうか。前問と同様に、有意水準 $\alpha=0.05$ (5%)として判別した。検定結果として適切な選択肢を○で囲め。

[t (片側) $=2.353 < 2.449$ 、 $2.449 < 3.182 = t$ (両側)]であり、
テスト A よりもテスト B の難易度が[高いと言える、高いとは言えない]

2.下記設問に対し文章形式(必要ならば図、表、数式を用いる)で答えよ。

- (1)標本の偏りにはどのようなものがあるか。また偏りが出ないようにするための対策について簡潔に記せ。
- (2)重回帰分析とは何か説明せよ。目的変数 y (結果)とそれに影響を与える説明変数 x_1, x_2, x_3 (原因)と誤差 ε を用いて式も記せ。
- (3)主成分分析とは何か説明せよ。要因(x_1, x_2, x_3)と第1主成分 z_1 を用いて z_1 の式も記せ。
- (4)因子分析とは何か説明せよ。要因が3個(x_1, x_2, x_3)、共通因子が2個(f_1, f_2)の場合を例にとってパス図で示せ。
- (5)判別分析の2種類の手法(線形判別関数、マハラノビスの距離)について図を用いて説明せよ。

<2010 年度> 解答例

1.

- (1) 1.0
- (2) 0.8
- (3) 0.4
- (4) 3
- (5) $2.449 < 3.182 = t(\text{両側})$ 、ある
- (6) $t(\text{片側}) = 2.353 < 2.449$ 、高いとは言えない

2.

(1)

偏り	対策
調査上の偏り	基本特性に差がないか、層別化して分析、実験的に確認 予備調査、計画的な実験 何を調べるのか、調査対象は正しいか
実験上の偏り	
欠落値と偏り	

(2)

重回帰分析: いくつかの原因と結果をつなぐもの

$$y_1 = a_1 x_{11} + a_2 x_{21} + a_0 + \varepsilon_1$$

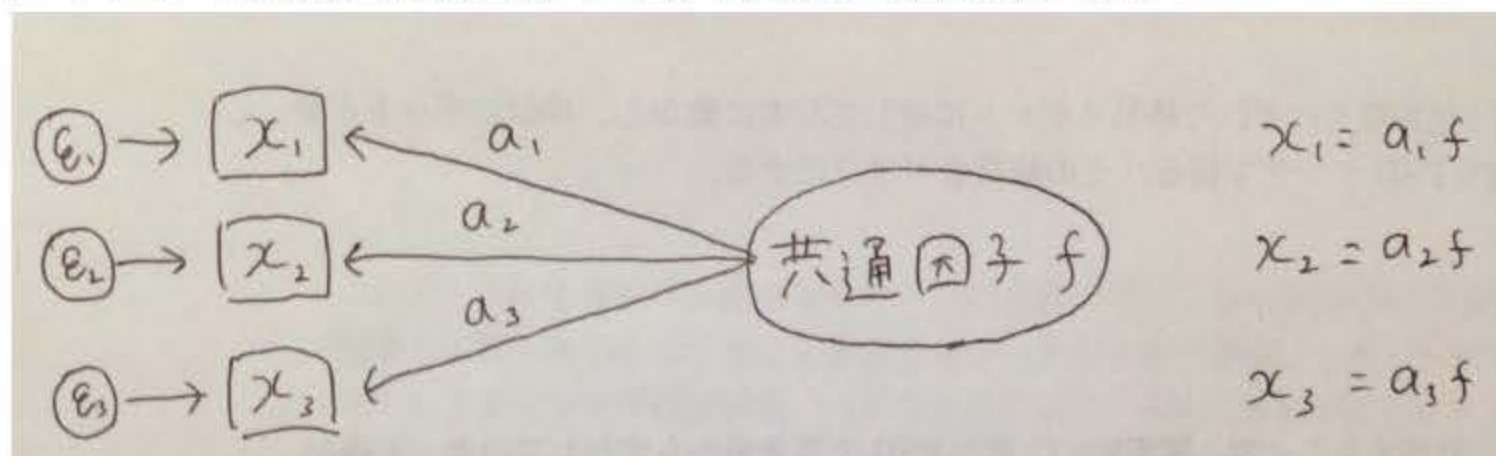
(3)

多くの変数の値をできるだけ情報の損失なしに互いに独立な総合的指標で表すもの

$$Z_1 = a_{11} x_1 + a_{12} x_2 + a_{13} x_3$$

(4)

いくつかの原因の共通因子の抽出をしたもの



(5)

