



2016 年 RSS/JSS 試験 (Higher Certificate)

HIGHER CERTIFICATE IN STATISTICS, 2016

モジュール 4：線型モデル

制限時間：90 分

4 問中 3 問を選択の上 解答のこと。

各問は合計 20 点である。小問の配点は括弧の中に記されている。

グラフ用紙と統計数値表は配布する。

解答にあたっては電卓を使用してよい。
ただし、一般財団法人統計質保証推進協会による「受験要領」に記された範囲で使用する。

数学記号 \log は e を底とする自然対数を表す。
その他の底をもつ対数は、例えば \log_{10} のように底を明示する。

また、 $\binom{n}{r}$ は ${}_nC_r$ と同じ意味とする。

問題用紙は 8 頁からなり、それぞれの頁は片面にのみ印刷されている。

この表紙が 1 頁目である。
第 1 問は 2 頁目から始まる。

問題は全部で 4 問である。

1. ある石油精製所の管理者が、異なる 7 回にわたり、比重 x をもつ石油に対する石油スピリット(石油系溶剤の一種)の収量のパーセンテージ y を測定する。 x の値が小さい方から順に並べた結果は以下の通りである。

x	30.2	32.8	32.9	35.1	42.3	45.5	46.0
y	6.8	10.1	14.3	19.3	10.2	20.0	23.7

- (i) データの散布図を描き、このデータに対して単回帰を行うことの是非について短くコメントせよ。(5)
- (ii) このデータに単回帰モデル $E(Y) = \alpha + \beta x$ を当てはめよ。計算の詳細も示せ。(6)
- (iii) 必要な仮定を述べた上で、計算の詳細を示しつつ、 β の 95%信頼区間を求めよ。(7)
- (iv) $x = 40$ のとき石油スピリットの収量のパーセンテージの点予測値を与えよ。(2)

2. (a) (i) 2 変量のデータ $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ がある. 標本積率相関係数を定義せよ. (3)

(ii) あるチョコレート製造業者が, 売り上げにおける広告の効果に興味がある. 彼は無作為に 8 種の商品を選ぶ. 各商品に費やされた広告費 X (単位 1000 英ポンド) とある期間における各商品の売り上げ額 Y (単位 10000 英ポンド) とが記録される. 以下の情報を用いて, 「売り上げ」と「広告費」の間の標本積率相関係数を計算せよ.

$$\sum_{i=1}^8 x_i = 386 \quad \sum_{i=1}^8 y_i = 460 \quad \sum_{i=1}^8 x_i^2 = 25\,426$$

$$\sum_{i=1}^8 y_i^2 = 28\,867 \quad \sum_{i=1}^8 x_i y_i = 26\,161$$

(4)

(iii) 「売り上げ」と「広告費」の間に, 正の相関を持つ根拠があるか否かについて有意水準 1% で検定せよ. 臨界値と結論も述べよ. (4)

(b) キャットショー(猫の品評会)において, A から H にラベル付けされた 8 匹のシャム猫について, 2 人の審判員が最高のもので最低のものへと以下のように順位を付けた.

順位	1 位	2 位	3 位	4 位	5 位	6 位	7 位	8 位
審判 1	B	C	E	A	D	F	G	H
審判 2	C	B	E	D	F	A	G	H

(i) 各審判員に対する 8 匹のシャム猫のランクを示す表を作れ. (2)

(ii) この順位に対するスピアマンの順位相関係数を計算し, 2 人の審判員によるシャム猫のランクの間には正の相関があるか否か有意水準 1% で検定せよ. (4)

(iii) スピアマンの順位相関係数を使う検定は分布の仮定に依存するか? 積率相関係数と比較してこの係数の長所と短所を一つずつ述べよ. (3)

3. (i) 完全無作為化実験計画法において、無作為化と反復を用いる理由を短く述べよ。ただし、以下で与えられる応用とは異なる例を用いて答えよ。(4)

- (ii) すべての処理群において等しい反復数をもつ完全無作為化実験計画法に対して、用いる記号をすべて定義してモデルの式を書け。(4)

4 種の異なる船舶用塗料について、航海の環境において船を守る能力を比較した。それぞれが 4 種類の塗料のうち 1 種を処理され、計 16 隻の船が使われた。それぞれの船は 6 ヶ月間配備され、船の帰還の際にキズ、剥離、残っている塗料の平均濃度に従ったスコアが記録された。より高いスコアはより良い状態を表している。スコアは以下の表のようになった。

塗料 1	80	73	72	90
塗料 2	81	82	88	84
塗料 3	93	80	80	97
塗料 4	89	86	96	99

- (iii) 以下は、因子「塗料」をもつ一元配置分散分析の分散分析表の一部である。

変動要因	自由度	平方和	平均平方	F
塗料				
誤差		577.5		
全体		983.8		

表を完成させ、4 種の塗料の間に差があるかどうか検定せよ。

(8)

- (iv) この実験計画法において 2 つの問題点と 2 つの改良点を述べよ。

(4)

4. (i) 2つの説明変数 X_1 , X_2 が従属変数 Y を予測するために用いられるとする. Y , X_1 , X_2 を含むデータを上記のように分析する時の重回帰分析モデルを書け. その際に, モデルの各項の意味と特性を説明せよ.

(5)

セメントが固化する間に熱がどのように放出されるかを実験調査する. なおその際に放出される熱量をセメントの化学的組成の関数と考える. 記録されたデータは, 180日固化した後に計測された, セメント1グラム当たりの熱量(Y), アルミン酸三カルシウムのパーセント割合(X_1), ケイ酸三カルシウムのパーセント割合(X_2)である.

データは, 統計パッケージによって分析される. 統計パッケージによる出力結果は **問題の末尾**に掲載されている. この出力結果を用いて以下の問いに答えよ.

- (ii) 1%の有意水準で回帰全体の有意性を検定しなさい. ただし, その際に統計学を専門としないものにも理解できる用語によって, 検定の結果を説明せよ.

(4)

- (iii) 上記のように定義された時, Y を X_1 と X_2 に回帰した時の回帰式を書け. また, この推定された回帰式を用いて, $X_1 = 8$ と $X_2 = 35$ のセメントが固化する間に放出される熱量を予測せよ.

(5)

- (iv) 推定されたモデルの出力結果において, 回帰パラメータの t 検定の一部の p 値が欠けている. これらの p 値が欠けているパラメータを有意水準0.1%で両側検定せよ. なお, その際には棄却域の臨界値も答えよ. また, 固化する間に放出される熱量に対してアルミン酸三カルシウムとケイ酸三カルシウムが与える影響について, 検定結果は何を意味するか.

(4)

- (v) R^2 決定係数の値を用いて, このモデルの全体的な当てはまりについてコメントせよ.

(2)

Regression Analysis: y versus x1 and x2

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Regression	2	2657.9	1328.95	229.53	4.404e-09
Residuals	10	57.9	5.79		
Total	12	2715.8			

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	52.57735	2.28617	23.00	5.46e-10
x1	1.46831	0.12130	12.11	
x2	0.66225	0.04585	14.44	

Residual standard error: 2.406 on 10 degrees of freedom
Multiple R-squared: 0.9787

BLANK PAGE

BLANK PAGE

BLANK PAGE