

2015 年 RSS/JSS 試験 (Higher Certificate)

HIGHER CERTIFICATE IN STATISTICS, 2015

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. (i) 線形単回帰モデル

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i=1, 2, \dots, n$$

をデータにあてはめる. 誤差項(ε_i)に課される通常の仮定は何か答えなさい.

(2)

- (ii) 適当な関数を最小化することによって, β_0 および β_1 の最小 2 乗推定値 $\hat{\beta}_0$ および $\hat{\beta}_1$ が

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

で与えられることを示しなさい. ただし, $S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y})$ および $S_{xx} = \sum(x_i - \bar{x})^2$ とする.

(8)

臨床家がある種の疾患にかかった 20 人の患者の年齢(x) と血中コレステロール値(y) を記録した. データの要約統計量は $\sum x_i = 809$, $\sum y_i = 68.3$, $S_{xx} = 3630.95$, $S_{xy} = 201.665$, $S_{yy} = 12.9455$ である.

- (iii) 推定された線形単回帰モデルの式を答えなさい.

(3)

- (iv) 分散分析表を求めて, $\beta_1 = 0$ の帰無仮説を検定しなさい.

(7)

2. 4つのカーペットの試作品の耐久性を比較することが求められている。各製品の4つのサンプルが作られている。つまり、全てで16のサンプルである。これらのサンプルの一つずつが完全無作為化法によって16の家庭の各々に割り当てられている。60日後の各サンプルの耐久性の程度 y は以下の表の通りである。

	カーペット製品			
	1	2	3	4
1	18.95	10.06	10.92	10.46
2	12.62	7.19	13.28	21.40
3	11.94	7.03	14.52	18.10
4	14.42	14.66	12.51	22.50

- (i) 一元配置分散分析に対する通常モデルは

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

である。誤差項(ε_{ij})に関する仮定を述べなさい。パラメータが推定できるようにするためのパラメータに対する適切な制約条件を述べなさい。
(3)

- (ii) $\sum \sum y_{ij}^2 = 3350.27$ であることを使って、これらのデータの分散分析を行いなさい。そして、平均の耐久性がカーペットの異なる製品の間で等しいとする帰無仮説を有意水準5%で検定しなさい。
(11)

- (iii) 今、カーペットの製品2は合成繊維でできており、製品4は羊毛だけでできていることを知らされたとする。カーペットの二種類の製品の間には耐久性の違いがあるかを調べるために、0.5%の有意水準で適切な t 検定を行いなさい。
(6)

3. (i) 2変量データ $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ が与えられている. 標本積率相関係数 (pmcc) を定義しなさい. (3)
- (ii) 新たな変数 z が, a, b を定数として $z_i = ax_i + b, i = 1, 2, \dots, n$, と定義されている. $a > 0$ の時, z と y の積率相関係数は x と y の積率相関係数と等しいことを示しなさい. また, $a < 0$ の時, z と y の積率相関係数はどうなるか答えなさい. (6)

コオロギの毎秒の鳴き声の数 (音の振動数) がいくつかの気温で計測されている. コオロギは冷血の生物であるため, 鳴き声の数のようなコオロギの挙動に気温が影響を及ぼすことには根拠がある. データは以下の表に与えられていて, それはコオロギの母集団からの無作為標本とみなせるものである.

毎秒の鳴き声数 y	気温 (華氏 °F) x
14.4	76.3
14.7	69.7
15.0	79.6
15.4	69.4
15.5	75.2
15.7	71.5
16.0	71.6
16.1	80.5
16.3	83.3
17.0	83.5
17.1	80.6
17.2	82.6
18.4	84.3
19.8	93.3
20.0	88.6

$$\sum x_i = 1190.0, \sum y_i = 248.6, \sum x_i^2 = 95\,098, \sum y_i^2 = 4161.1, \sum x_i y_i = 19\,862.6$$

が与えられている.

- (iii) これらのデータの積率相関係数を計算して, 母集団での鳴き声の数と気温が無相関であるとの帰無仮説を正の相関があるという対立仮説に対して, 有意水準 1% で検定しなさい. (6)
- (iv) これらのデータからスピアマンの順位相関係数を計算しなさい (y は順番に並んでいることに注意せよ). これを使って, 鳴き声の数と気温が無相関であるとの帰無仮説を, 正の相関があるという対立仮説に対して, 有意水準 1% で検定しなさい. (5)

4. 応答変数 Y と説明変数 x_1, x_2, \dots, x_p に対する線形重回帰モデルを書いて，必要な仮定を述べなさい. (4)

生態研究において，35 匹の熊を無作為標本として選ぶ．そしてそれぞれの熊を，一時的に麻酔にかけて，その体を測定し，体重を計測する．この研究のひとつの目的は，熊の体重を他の測定結果に基づいて推定することにある．この結果は，森林においては熊の体重よりも長さを計るほうがやさしいことから，利用されるだろう．

以下の変数が記録されている: 月齢; 頭 (長さ, インチ); 首 (首周り, インチ); 胸囲 (インチ); 体重 (ポンド) .

研究者たちは体重を応答変数とし，残りの 4 つの変数を説明変数とする線形重回帰モデルを作ることを目指している．このモデルの作成から得られた結果は以下の通りである．

- (i) 分散分析表を用いて何の仮説が検定されるか述べなさい．5%有意水準を用いてこの検定から得られる結論は何かを述べなさい．また，'決定係数 = 96.3%'の意味を説明しなさい. (6)
- (ii) 回帰において，頭の変数について帰無仮説を述べて，有意水準 5% の検定を行いなさい．また，胸囲の係数の 90% 信頼区間を計算しなさい. (8)
- (iii) 月齢 71, 頭 7, 首 27 および胸囲 44 の熊の体重の予測値を，回帰式を使って計算しなさい．なお，これらのデータはモデルが推定されるために用いられた観測値の範囲内に十分入っていると仮定してよい. (2)

回帰分析: 体重と月齢, 頭, 首, 胸囲

回帰式

$$\text{体重} = -203 + 0.655 \text{ 月齢} - 9.01 \text{ 頭} + 11.8 \text{ 首} + 6.26 \text{ 胸囲}$$

(予測量)	推定値	標準誤差
定数	-202.52	35.05
月齢	0.6555	0.1905
頭	-9.013	4.693
首	11.815	3.559
胸囲	6.255	1.677

$$\text{標準誤差} = 25.7088 \quad \text{決定係数} = 96.3\% \quad \text{自由度修正済み決定係数} = 95.8\%$$

分散分析

変動要因	自由度	平方和	平均平方	F 値
回帰	4	513741	128435	194.3
残差	30	19828	661	
合計	34	533570		

BLANK PAGE

BLANK PAGE

BLANK PAGE