

RSS Higher Certificate in Statistics, 2010

Module 4 : Linear models

1. 以下の各問に答えよ.

- (a) (i) 完全無作為化実験計画法に関連して, (b)の応用とは異なる例を挙げながら, 無作為化と反復の原理について簡潔に説明せよ.
- (ii) 完全無作為化法において, 全ての処理グループで繰り返し数が等しいときの等式モデルを述べよ. なお用いた各文字の定義を示せ.
- (b) (i) 合成繊維の張力(TS)が綿の割合に依存するかどうかを判断するために観測を行った. 5つの繊維のサンプルについて, それぞれ4種類の綿の割合で調べたら, 以下のようになった.

<i>% cotton</i>	<i>TS (grams per mm²)</i>					<i>Row Totals</i>	<i>Row Sums of Squares</i>
15	7	7	15	11	10	50	544
20	10	17	12	18	18	75	1181
25	14	20	13	16	22	85	1505
30	19	25	16	19	21	100	2044

これらのデータを分散分析し, 非科学者でも理解できるような言葉で結果を報告せよ.

- (ii) ある助手は, データの回帰分析が役に立つのではと示唆している. この示唆の理由を述べよ. 更に, 分析の進行方法とそこから得られる情報を, 更なる計算をせずに説明せよ.
2. ある金属は, 空気にさらされると変色する. 変色から保護するために, 金属は化学物質で被覆されている. この金属の標準サンプルに様々な厚さ(x mm) の化学物質をコーティングし, 変色するまでの時間(t 時間)を記録する実験を行った. 結果は以下のようであった.

x	1.8	3.0	4.0	5.7	7.2	8.4	10.3
t	3.4	5.9	7.0	8.7	9.5	10.4	11.1

- (i) これらのデータの最小二乗回帰直線は

$$t = 3.027 + 0.8617x$$

と与えられている. データの散布図を描き, この回帰直線を加えよ. また, t の x への単純線形回帰モデルの妥当性について解説せよ.

- (ii) ある研究者は, t と x との理論的な関係は, A, B を定数として

$$\exp(t) = Ax^B$$

という形をしていると考えている. この関係は

$$t = a + b \log x$$

と表せることを示し, a, b をそれぞれ A, B の関数形で明示せよ.

(iii) $\sum \log x = 11.2476$, $\sum t = 56$, $\sum (\log x)^2 = 20.3687$, $\sum t \log x = 100.101$ が与えられている.

これらを用いて t の $\log x$ への最小二乗回帰直線を計算し, 横軸に $\log x$ の値をとった散布図にこの回帰直線をプロットせよ.

(iv) 優れた方のモデルを選びその理由を述べよ. また 2 つのモデルそれぞれについて, $x=6$ のときの t の予測値を計算し, 簡単にコメントせよ.

3. 以下の各問に答えよ.

(a) 2 変量データの組 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ が与えられていて, これらは 2 変量母集団からの無作為抽出であると仮定する. 標本の積率相関係数, および同様の形をしたスピアマンの (標本) 順位相関係数の計算式を定義せよ. また, それぞれの相関係数について有用である状況を説明せよ.

(b) 下の表 (2008 年 5 月付) で与えられる欧州 10 ヶ国では, 無鉛ガソリン (unleaded) とディーゼル (diesel) に対して消費税 (1 ℓ あたりのペニー) が課税される.

Country	Unleaded (x)	Diesel (y)
Austria	36	28
Denmark	42	29
Estonia	23	19
Germany	52	37
Greece	26	22
Hungary	30	25
Italy	45	34
Poland	33	23
Spain	31	24
United Kingdom	57	57

$\sum x = 375$, $\sum y = 298$, $\sum x^2 = 15193$, $\sum y^2 = 9974$, $\sum xy = 12191$ が与えられている.

- (i) x に対する y の散布図を作成し, y と x との関係があればコメントせよ.
- (ii) 上記のデータに対する積率相関係数を計算し, x と y は相関がないという帰無仮説を有意水準 1% で検定せよ.
- (iii) 上記のデータに対するスピアマンの順位相関係数を計算せよ. これを用いて, 基本的な母集団における x と y には相関がないという帰無仮説を有意水準 1% で検定せよ.
- (iv) (ii) と (iii) の結果について, ここではどちらの分析が優れているか, それはなぜか. また, この分析において考えられる注意点を述べよ.

4. 以下の各問に答えよ.

(i) 線形重回帰モデル

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i, \quad i = 1, \dots, n$$

において, 残差 (誤差) の文字 e_1, \dots, e_n について通常どんな前提条件が必要か.

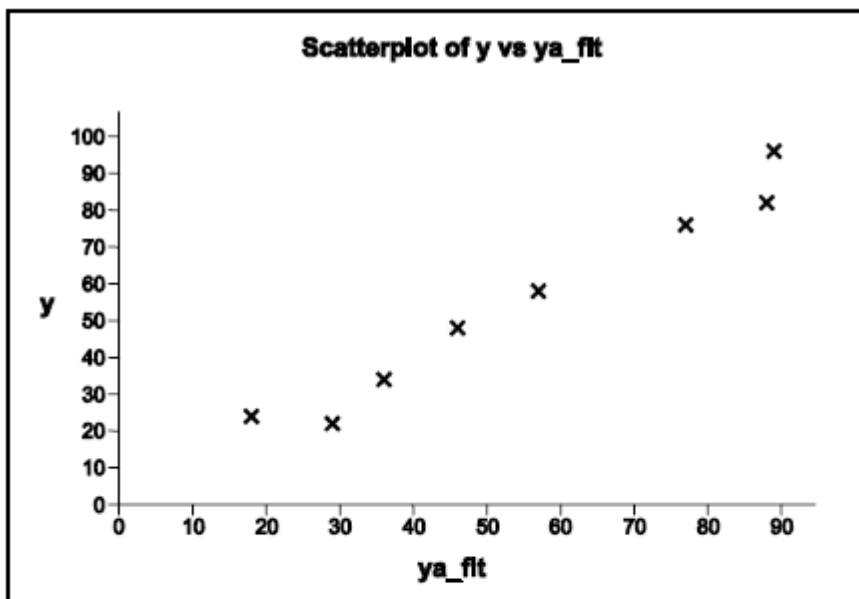
(ii) 以下の 3 ページにわたって, コンピュータに対する適性スコア(y) を, 言語能力(x_1)及び演算能力(x_2)の一方もしくは両方に回帰した 3 つの回帰分析 (モデル A,B,C) の出力を示している.

<i>Student</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>Mean</i>
<i>y</i>	22	24	34	48	58	76	82	96	55
<i>x</i> ₁	53	52	58	60	57	61	66	65	59
<i>x</i> ₂	50	45	51	55	62	70	73	74	60

出力を用いて以下の各問に答えよ.

- (a) 3 つのモデルについて, 平均二乗誤差の平方根 S の各値と関連させながら, 横軸を y のあてはめ値, 縦軸を y の観測値とした散布図について簡単にコメントせよ.
- (b) モデル B と C において, x_1 と x_2 の係数それぞれの有意性を検定せよ.
- (c) モデル A において, x_2 の存在下での x_1 の有意性, および x_1 の存在下での x_2 の有意性を, 5%水準で検定せよ. さらに, x_1 と x_2 の回帰の全体的な有意性を検定せよ. “ $R^2 = 96.7\%$ ” という式を説明し, この値を計算せよ.
- (d) 3 つのモデルのうち, どれがデータを記述するのに最適か. その理由も示せ.

モデル A



Regression Analysis: y versus x1, x2

The regression equation is $y = -124 + 1.00 x1 + 2.00 x2$

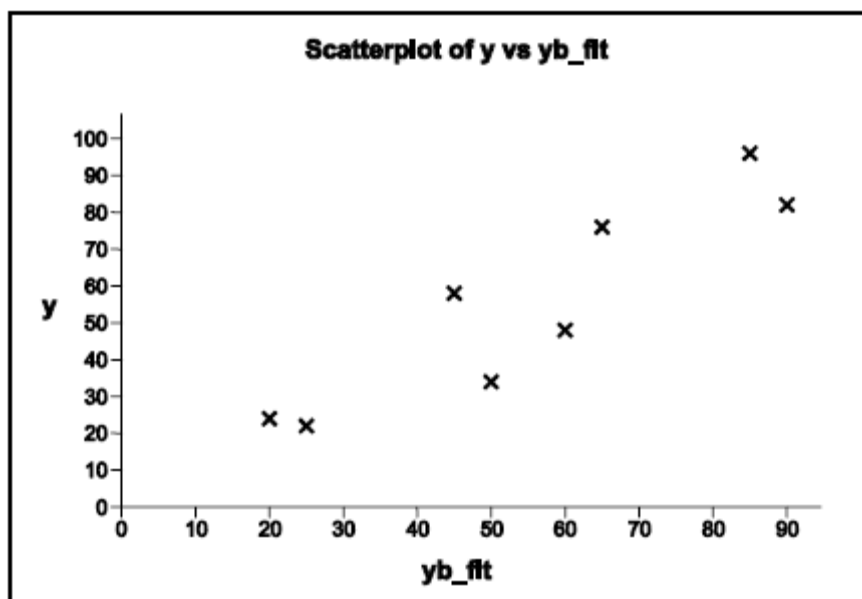
Predictor	Coef	SE Coef
Constant	-124.00	37.06
x1	1.0000	1.0000
x2	2.0000	0.4472

$\sqrt{\text{(mean square error)}}$ is $S = 6.0000$; $R\text{-Sq} = 96.7\%$

Analysis of Variance

Source	DF	SS	MS
Regression	2	5220.0	2610.0
Residual Error	5	180.0	36.0
Total	7	5400.0	

モデル B



Regression Analysis: y versus x1

The regression equation is $y = -240 + 5.00 x1$

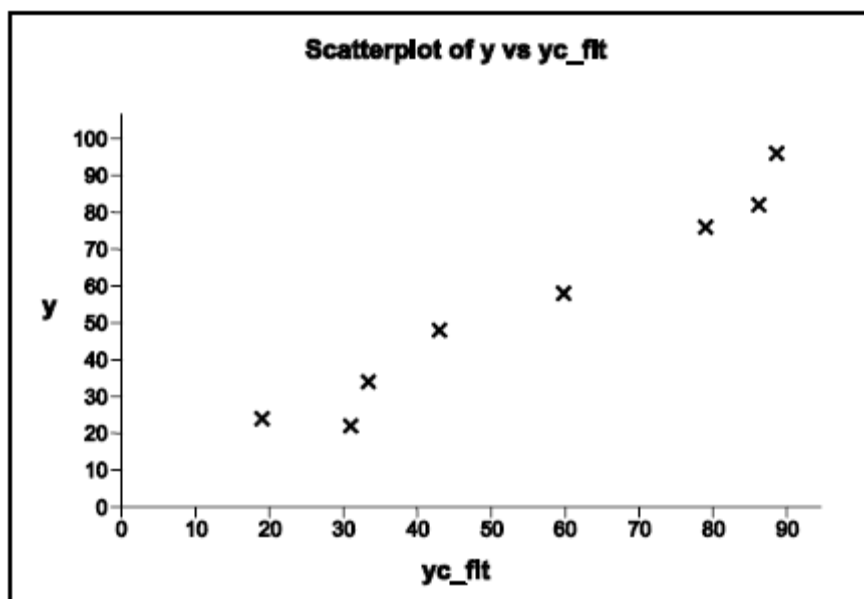
Predictor	Coef	SE Coef
Constant	-240.00	54.03
x1	5.0000	0.9129

√(mean square error) is $S = 12.2474$; $R-Sq = 83.3\%$

Analysis of Variance

Source	DF	SS	MS
Regression	1	4500.0	4500.0
Residual Error	6	900.0	150.0
Total	7	5400.0	

モデル C



Regression Analysis: y versus x2

The regression equation is $y = -89.0 + 2.40 x2$

Predictor	Coef	SE Coef
Constant	-89.00	12.19
x2	2.4000	0.2000

√(mean square error) is $S = 6.0000$; $R-Sq = 96.0\%$

Analysis of Variance

Source	DF	SS	MS
Regression	1	5184.0	5184.0
Residual Error	6	216.0	36.0
Total	7	5400.0	