

RSS Higher Certificate in Statistics, Specimen B

Module 4 : Linear Models

1. 10匹のスズキが水産研究者によって捕獲され、長さ x (mm) と重さ y (g) が計測された。データは以下のようである。

Length (x)	387	366	329	293	273	268	294	198	185	169
Weight (y)	720	680	480	330	270	220	380	108	89	68

- (i) 重さ y を縦軸に、長さ x を横軸にとってデータをプロットせよ。これらのデータへの直線の当てはめは適切であるか。
- (ii) (a) 回帰直線 $y = \beta_0 + \beta_1 x$ におけるパラメータ β_0, β_1 の最小2乗推定値を求めよ。ただし、以下の値を用いよ：

$$n = 10, \quad \sum_{i=1}^n x_i = 2762, \quad \sum_{i=1}^n y_i = 3345,$$

$$\sum_{i=1}^n x_i^2 = 812594, \quad \sum_{i=1}^n y_i^2 = 1610009, \quad \sum_{i=1}^n x_i y_i = 1075861.$$

- (b) スズキの長さとの重さの関係を表すのに上問 (a) で求められた回帰直線が適当であるかどうかをコメントせよ。
- (iii) 決定係数を計算し、解釈せよ。

2. 以下の各問に答えよ。

- (i) 一元配置分散分析に用いられる線形モデルを示し、モデルの中のどの項が何を表し、分析が妥当であるために置かれる仮定について説明せよ。
- (ii) ある心理学者は適度なレジャーはストレスの軽減に役立つとの仮説を持っている。彼女は遊びが重要で、特にそれが強制されたものではなく本人の自由な選択によるものであるほど効果があると考えている。

それを立証するため彼女は36人の大学生を無作為に3つのグループに振り分けた。第1のグループは何をやるか高度に強制された形での遊び (High) を行うとし、第2のグループは強制力の弱い遊び (Low) をするとした。第3のグループは遊びというより作業 (Work) に近いものを行わせた。

被験者全員にはまずストレスを与える作業、すなわちヘッドフォンで大きな音を聞きながら数学の問題を解くことを30分間行なわせた。次に被験者には各グループごとに10分間上述のような遊びをさせた。

最後に被験者全員に幾何的なパズルを2題解かせた。そのうちの1題は答がないものであったがそのことは被験者には知らせなかった。そしてその問題を解くのをあきらめるまでに時間 (単位: 分) を計測した。その時間によって各グループでの遊びの効果を知

ろうとするものであった。結果は以下のようである。

<i>High</i>	<i>Low</i>	<i>Work</i>
347	504	398
567	420	492
424	583	97
239	183	357
256	279	184
682	381	554
435	118	354
666	317	275
825	359	198
102	77	163
601	336	284
384	197	155

このデータに対し一元配置分散分析を行い、かつ3つの平均値間の差の吟味によって各グループ間の違いを論ぜよ。そして実験結果の解釈を研究者に報告するための簡潔なレポートを作成せよ。

3. 以下の各問に答えよ。

- (a) ある心理学者は知能と水平思考の能力との間の関連の強さを調べようとしている。無作為に選んだ12名の被験者に対し、知能検査と水平思考能力のテストが行われた。結果は以下の表のようであった。知能検査は150点満点、水平思考テストは10点満点である。

<i>Subject</i>	1	2	3	4	5	6	7	8	9	10	11	12
<i>Intelligence test score</i>	121	148	108	137	141	124	131	115	118	110	132	127
<i>Lateral thinking test score</i>	3	9	6	7	8	2	5	5	6	4	8	7

心理学者はこのデータをスピアマンの順位相関係数により分析することにした。

- (i) 水平思考テストでは同点者が何人かいる。このことを分析にどのように反映させるべきであろうか。
- (ii) スピアマンの順位相関係数を計算し、その統計的有意性を評価せよ。
- (iii) このデータから知能と水平思考の能力との間の関連性について何がいえるであろうか。
- (b) n 組のデータ $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ が与えられたとき、積率相関係数 r の計算式を示し、その値の意味するところを述べよ。また、以下のそれぞれに対応する散布図の例を描け。
- (i) 強い正の相関
 - (ii) 強い負の相関
 - (iii) 互いに独立
 - (iv) 無相関であるが独立ではない

4. 以下の各問に答えよ.

- (i) 実験計画法の2つの基本的な原則は無作為化と反復である. これらが何故重要であるのか, およびこれらが実験結果の解析を妥当なものとするためどのような手助けとなるのかを述べよ.
- (ii) 樹木の3つの種が実験林で生育された. 植えられたすべての苗が育つわけではないので, 反復数は種ごとに異なっている. 以下のデータは一定期間後の生育した木の高さ (単位: メートル) を表したものである.

<i>Species</i>	<i>Replicates</i>	<i>Observations</i>	<i>Total</i>
Pinus caribea	9	4.20 4.30 3.50 3.90 5.00 4.80 4.60 4.50 4.00	38.80
Pinus kesiya	12	3.95 3.85 4.25 4.70 4.15 3.30 3.65 3.70 3.95 4.00 3.70 4.30	47.50
Eucalyptus deglupta	8	7.95 8.10 8.30 6.60 7.50 7.70 7.25 8.00	61.40
Total	29		147.70

29本の木すべての高さの平方和は 831.8900 である. 種の間で有意な差があるかどうかを一元配置分散分析により評価せよ. 特に以下の間での差があるかどうかを調べよ:

(a) pinus と eucalyptus

(b) 2つの pinus 間

さらに各観測データの残差を求め, ドットプロットを作成し, 残差がどのような情報をもたらすかを述べよ.