

Bài tập về nhà

Hồi quy tuyến tính

Bài 1

Một công ty tư vấn nguồn nhân lực tiến hành một nghiên cứu trên một mẫu 22 giám đốc điều hành công ty trong thời gian hai tuần. Họ muốn kiểm tra xem liệu các yếu tố sau có thể ảnh hưởng đến mức độ lo lắng:

- X_1 : áp lực công việc
- X_2 : kỹ năng quản lý
- X_3 : mức độ hài lòng với chức vụ của mình

Có bảng ANOVA sau:

Nguồn gốc của sự biến thiên	Tổng bình phương	Bậc tự do
Hồi quy trên X_1	981.326	1
Hồi quy trên $X_2 \mid X_1$	190.232	1
Hồi quy trên $X_3 \mid X_1, X_2$	129.431	1
Sai số	442.292	18
Tổng quát	1743.281	21

1. Tính tổng bình phương hồi quy trên X_1 , X_2 và X_3 ?
2. Xác định tỷ lệ phần trăm sự biến thiên của mức độ lo lắng được giải thích bởi các biến độc lập.
3. Có thể kết luận rằng trong tất cả ba biến giải thích đều có ảnh hưởng đáng kể đến mức độ lo lắng hay không? Chỉ rõ kiểm định nào được dùng.
4. Nếu chúng ta chỉ xét biến giải thích X_1 , hãy lập bảng ANOVA ?
5. Kiểm định giả thiết sau với mức ý nghĩa 5%
 - a. $H_0 : \beta_1 = 0$ cho mô hình $Y = \beta_0 + \beta_1 X_1 + \epsilon$
 - b. $H_0 : \beta_2 = 0$ cho mô hình $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$
 - c. $H_0 : \beta_3 = 0$ cho mô hình $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$
6. Xác định hệ số xác định cho mỗi mô hình trong câu 5.
7. Trong các mô hình trên, mô hình nào thích hợp nhất để giải thích sự biến động mức độ lo lắng của các giám đốc ?

Bài 2

Hình 1: Số liệu bài 5

Essai numéro	Résistance à la rupture Y_i	Épaisseur du matériau X_{i_1}	Densité X_{i_2}
1	37,8	4	4,0
2	22,5	4	3,6
3	17,1	3	3,1
4	10,8	2	3,2
5	7,2	1	3,0
6	42,3	6	3,8
7	30,2	4	3,8
8	19,4	4	2,9
9	14,8	1	3,8
10	9,5	1	2,8
11	32,4	3	3,4
12	21,6	4	2,8

Công ty A sản xuất vật liệu nhựa được sử dụng trong sản xuất đồ chơi. Bộ phận kiểm soát chất lượng của công ty đã tiến hành một nghiên cứu nhằm xác định mức độ bền dẻo của nhựa (Y) có thể bị ảnh hưởng bởi độ dày của vật liệu (X_1) và mật độ của vật liệu (X_2). Mười hai thử nghiệm đã được tiến hành và kết quả được trình bày trong bảng trên đây.

1. Tìm 2 phương trình đường thẳng hồi quy và 1 phương trình siêu phẳng (nếu có) ?
2. Xác định tỷ lệ phần trăm sự biến thiên của biến phụ thuộc cho từng mô hình có thể có trên.
3. Nếu chúng ta chỉ quan tâm đến cả 2 biến giải thích, hãy lập bảng ANOVA ?
4. Kiểm định giả thiết sau với mức ý nghĩa 5%,

$$H_0 : \beta_1 = \beta_2 = 0$$

5. Xác định khoảng tin cậy với mức ý nghĩa 5% cho β_1 trong trường hợp mô hình chỉ có biến độc lập là độ dày của vật liệu.
6. Với khoảng tin cậy vừa tìm được ở câu 5, chúng ta có thể khẳng định rằng hồi quy tuyến tính là có ý nghĩa giữa mức độ bền dẻo của nhựa và độ dày của vật liệu và mật độ của vật liệu không? Chứng minh điều khẳng định của bạn.

Bài 3 Sử dụng bảng số liệu cho như sau:

1. Viết các mô hình tuyến tính với 2 biến độc lập (có thể).
2. Ước lượng các hệ số hồi quy trong từng mô hình tuyến tính ở câu 1.
3. Với độ tin cậy 95%, tìm khoảng tin cậy cho các tham số trong mô hình với 2 biến độc lập

Hình 2: Số liệu bài 6

y	x_1	x_2	x_3
12	2	45	121
14	1	43	132
10	3	43	154
16	6	47	145
14	7	42	129
19	8	41	156
21	8	32	132
19	5	33	147
21	5	41	128
16	8	38	163
19	4	32	161
21	9	31	172
25	12	35	174
21	7	29	180

x_1 và x_2 .

4. Xác định hệ số xác định cho mỗi mô hình trong câu 1.
5. Trong các mô hình trên, mô hình nào thích hợp nhất để giải thích sự biến thiên của Y ?
6. Viết mô hình tuyến tính dưới dạng ma trận với số biến độc lập nhiều nhất có thể, và xác định kích thước của ma trận.
7. Ước lượng các hệ số hồi quy trong mô hình tuyến tính ở câu 6.
8. Trong mô hình tuyến tính ở câu 6, tính ước lượng của $V(\epsilon)$ và $V(\hat{\beta})$.
9. Với độ tin cậy 95%, tìm khoảng tin cậy cho $V(\epsilon)$.
10. Khi thêm 2 biến độc lập x_3 và x_2 vào mô hình chỉ với 1 biến độc lập x_1 thì làm cho chất lượng ước lượng cao hơn không?