
**FOXCONN**
鴻海科技集團

相關性和迴歸分析


phân tích tương quan và hồi quy

Six Sigma

1

本單元綜覽 tổng quan về bài này


- 介紹相關性 giới thiệu tính tương quan.
- 介紹一元迴歸分析 giới thiệu hồi quy đơn.
- 討論簡單的線性迴歸 thảo luận hồi quy tuyến tính đơn.
- 執行殘值的診斷分析 tiến hành phân tích chẩn đoán sai số.
- 介紹高階數學模式 giới thiệu mô hình toán học bậc cao.
- 介紹多元迴歸分析 giới thiệu hồi quy bội.

Six Sigma

2

流程改善路徑圖 sơ đồ lưu trình cải tiến


定義 định nghĩa	測量 đo lường	分析 phân tích	改善 cải thiện	控制 không chế
定義專案的範圍和目標 định nghĩa phạm vi và mục tiêu của chuyên án 指定黑或綠帶 (Black or Green Belt) chỉ định đai đen hoặc đai xanh 組成團隊，並建立專案章程 thành lập nhóm và lập chương trình của chuyên án	繪製流程圖 và biểu đồ lưu trình phân tích hệ thống đo lường 分析測量系統 評估流程控制和能力 đánh giá lưu trình không chế và năng lực	排定要改善的優先次序 sắp xếp ưu tiên theo thứ tự 啟動流程 FMEA khởi động FMEA của lưu trình 進行流程觀察研究 nghiệm cứu quan sát lưu trình 確認流程關鍵因素 xác nhận biến số quan trọng trong lưu trình	行動計畫 kế hoạch hành động 設計和執行流程改善 thí nghiệm 解決方案 驗證 nghiệm chứng nghiệm chứng nghiệm 解決方 案 驗證 nghiệm chứng nghiệm chứng nghiệm phương án giải quyết	確定 FMEA 和最終控制計畫 xác định FMEA và kế hoạch không chế cuối cùng 實施流程控制機制 (移交和培訓) thi hành kế hoạch không chế lưu trình 按專案目的長期能力 xác nhận năng lực lâu dài của chuyên án

Six Sigma

3

相關性

tính tương quan

Six Sigma

11-1

一元回歸分析

phân tích hồi quy đơn

Six Sigma

分析路徑圖：單個 Y 和單個 X
sơ đồ phân tích: một X và một Y

Y 的種類 loại hình Y

連續型 liên tục

連續型 liên tục

離散型 rời rạc

X 的種類 loại hình X

連續型 liên tục

連續型 liên tục

離散型 rời rạc

回歸 hồi quy

邏輯 hồi quy logic

1. T 檢驗證檢 định T

2. 方差分析 ANOVA

1. 卡方檢驗證檢 định chi-square

2. 邏輯回歸 hồi quy logic

Six Sigma

14

範例 thí dụ

制程工程師想要知道膠化時間和催化劑用量是否相關
biết thời gian gần và liều lượng chất xúc tác có phải tương quan

使用什麼工具? Sử dụng phương pháp nào?

Phân tích hồi quy

Y 的種類 loại hình Y

連續型 liên tục

連續型 liên tục

離散型 rời rạc

X 的種類 loại hình X

連續型 liên tục

連續型 liên tục

離散型 rời rạc

回歸 hồi quy

邏輯 hồi quy logic

1. T 檢驗證檢 định T

2. 方差分析 ANOVA

1. 卡方檢驗證檢 định chi-square

2. 邏輯回歸 hồi quy logic

Six Sigma

15

分析路徑圖 trình phân tích

迴歸 hồi quy

做分析計畫 lập kế hoạch phân tích

收集資料 thu tập dữ liệu

在 Minitab 建立模擬曲線圖 ở Minitab lập đồ thị đường cong Stat>Regression>Fitted Line Plot

評估 R² 和 P 值的顯著性 đánh giá R² và trị số P có phải rõ ràng

評估殘值項 đánh giá sai số Stat>Regression> Fitted Line Plot>Graphs> Residual Plots

做決策 làm quyết sách

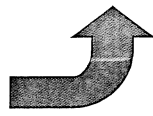
Six Sigma

16

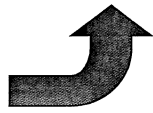
11-4

大體上的方法 cách tiếp cận chung

實際問題 vấn đề thực tế



數學模式 mô hình toán học



實際的解決方案 phương
án giải quyết thực tế

Six
Sigma

17

建立一個模式 lập một cái mô hình

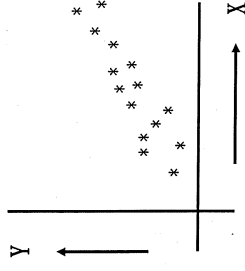
■技術上的 trên kỹ thuật:

--建立一個可描述資料的線性模式。Lập một mô hình có thể miêu tả mô hình tuyến tính của tư liệu

■該模式幫助我們理解資料必須有 mô hình này giúp chúng ta hiểu biết tư liệu cần phải có:

--輸入變數 X (自變量) biến số đưa vào X (biến độc lập)

--輸出變數 Y (因變數) biến số đưa ra Y (biến phụ thuộc)



Six
Sigma

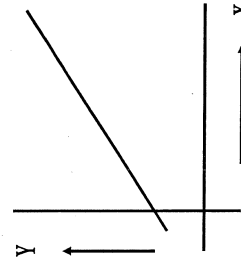
18

建立一個模式 lập một mô hình

$$\mathbf{y = mx + b}$$

$b = y$ 截距 chân của y

$m =$ 直線的斜率 độ lệch của đường thẳng



$$\mathbf{y = \beta_0 + \beta_1 x_1}$$

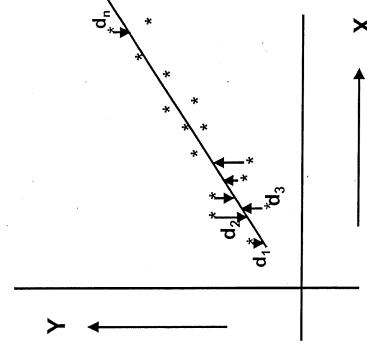
$\beta_0 =$ 常數 số thường

$\beta_1 = x_1$ 的係數 hệ số của x_1

Six
Sigma

19

配適成線性模式 phối thành mô hình tuyến tính



人的手與眼睛可以畫出一條相當接近資料中心的線。Tay và mắt của con người có thể vẽ ra một đường tương đương tiếp gần trung tâm tư liệu

電腦比較不靈活，必須靠數據計算才能畫出。Máy tính không linh hoạt, cần phải dựa số liệu tính ra mới có thể vẽ ra

加總 $d_1^2 + d_2^2 + d_3^2 + \dots + d_n^2$ cộng $d_1^2 + d_2^2 + d_3^2 + \dots + d_n^2$

向周圍移動直線，直到總平方和的值最小 (最小平方拟合法) đi chuyển đường thẳng, đến lúc số tổng bình phương nhỏ nhất (phương pháp mô phỏng bình phương nhỏ nhất)

Six
Sigma

20

迴歸: 模式檢定 hồi quy: Kiểm định mô hình

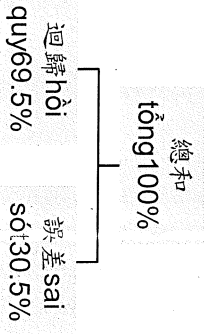
■數學模式應該 mô hình toán học nên :

- 是顯著的: 可依迴歸分析的 F 檢定值判定 là rõ ràng: có thể theo trị số kiểm định F của phân tích hồi quy phân định
- 對預測目的很有用: R²大於 70% với mục đích dự báo rất có ích: R² lớn hơn 70%
- 殘值分析無異常 phân tích sai số không có bất thường

Six Sigma

25

迴歸樹狀圖 biểu đồ hình cây hồi quy



這是呈現資料的常用方法。Đây là phương pháp thường dùng thể hiện tư liệu

Six Sigma

27

迴歸: 結果 hồi quy: kết quả

The regression equation is
CustomerData = 182.8 + 0.4763 SupplierData

Model Summary

S	R-sq	R-sq(adj)
13.5571	69.50%	67.89%

R-sq代表數學模型能解釋的變異, R-sq đại diện tỷ lệ biến thể có thể được mô hình toán học giải thích.

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	7955.9	7955.91	43.29	0.000
Error	19	3492.1	183.79		
Total	20	11448.0			

模型是顯著的 mô hình là rõ ràng

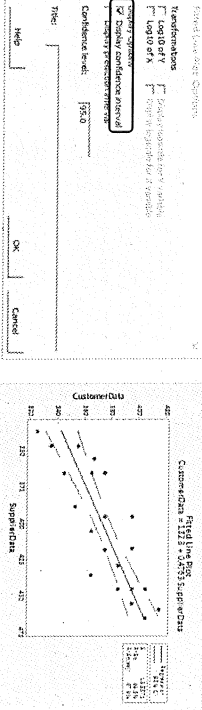
$$R^2 = \frac{SS_{regression}}{SS_{total}} = (7955.9 / 11448.0) = 0.695$$

Six Sigma

26

信賴區間 Khoảng tin cậy

■再作一次迴歸，在選項中選擇“顯示信賴區間帶” làm lại một lần hồi quy, trong hàng mục chọn “hiện thị Khoảng tin cậy”



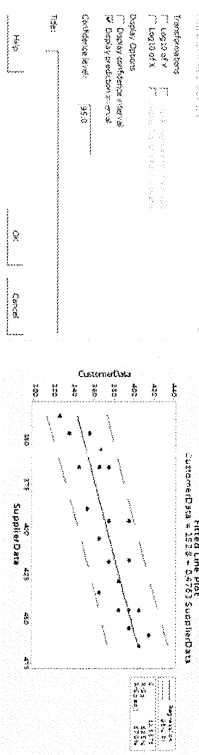
在迴歸線(紅色)上, 下方的綠線定義了你可以有 95% 的信心度會包含客戶資料的群體平均值的範圍。比如, 迴歸方程式告訴你相對於供應商的觀測值 400, 預測用戶端的值會是 373.32, 然而實際的群體平均值是不大可能剛好等於這個值的, 信賴區間帶指出你可以有 95% 的信心度認為群體平均值大概會落在 367 到 379 之間。 ở trong phạm vi đường xanh quanh đường hồi quy(màu đỏ) định nghĩa bạn có thể có 95% lòng tin cho rằng sẽ gồm phạm vi của giá trị bình quân quần thể của tư liệu khách hàng. ví dụ, phương trình hồi quy báo cho bạn tương đối tin số quan trắc 400 của nhà cung ứng, dự báo trị của khách hàng là 373.32. Nhưng mà giá trị bình quân quần thể thực thể là không khả năng vừa đúng trị này, không tin cậy chỉ ra bạn có thể có 95% lòng tin cho rằng là giá trị bình quân quần thể khoảng ở giữa 367 đến 379

Six Sigma

28

預測區間 khoảng cự bảo

■ 再作一次迴歸，在選項中選擇 “顯示預測區間帶” làm lại một lần hồi quy, trong hàng mục chọn “hiển thị khoảng tin cậy”

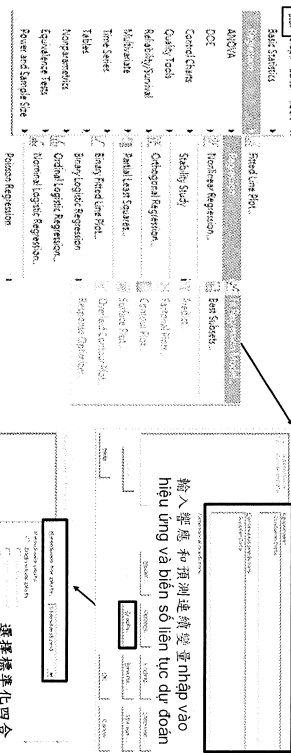


預測區間帶代表一系列的從將某觀測到的供應商數值範圍的預測區間。你可以有 95% 的信心是認為新的觀測值將落在顏色後的範圍內。(但是，這只對在分析中所涉及的范围內的供應商數據是有效的。) khoảng dự báo biểu thị khoảng dự báo do một loại dữ liệu nhà cung ứng quan trắc được ra b. n có thể có 95% long tin cho rằng trị mới quan trắc sẽ nằm trong phạm vi đường màu xanh.(nhưng phải chú ý đây chỉ là dữ liệu nhà cung ứng nằm trong phạm vi phân tích mới có hiệu quả) 相對於供應商觀測密度值 400，預測用戶端的數值會是 373。然而用戶端的數值是不大可能剛好就是 373。預測區間帶指出，你可以有 95% 的信心是認為實際值會落在 344 到 402 之間。ví dụ: phương tính hỏi quý báo cho bạn tương đối trị số quan trắc 400 của nhà cung ứng， dự đoán trị của khách hàng là 373. Nhưng mà giá trị bình quân quan thể thực thể là không khả năng và đúng trị 373, phạm vi dự đoán chỉ ra bạn có thể có 95% long tin cho rằng là giá trị bình quân thể khoảng ở giữa 344 đến 402

29

回歸分析：擬合回歸模型 phân tích hồi quy:Fit Regression Model

■ Stat>Regression>Regression>Fit Regression Model



輸入響應和預測連續變量輸入 vào hiệu ứng và biến số liên tục dự đoán

30

回歸分析結果結果 quả phân tích hồi quy

Regression Equation
CustomerData = 182.8 - 0.4763 SupplierData

Coefficients						
Term	Coef	SE Coef	T-Value	P-Value	VIF	
Constant	182.8	28.4	6.23	0.000		
SupplierData	-0.4763	0.0724	-6.58	0.000	1.00	

Model Summary

S	R-sq	R-sq(Adj)	R-sq(Predict)
1	49.50%	48.53%	48.53%

R-sq代表數學模型能解釋的變異，R-sq đại diện bao nhiêu biến thể được mô hình toán học giải thích.

Analysis of Variance						
Source	DF	Adj SS	Adj MS	F-Value	P-Value	
Regression	1	7955	7955.3	43.28	0.000	
SupplierData	1	7956	7955.9	43.28	0.000	
Error	19	2492	130.8			
Lack-of-Fit	13	2423	186.6	1.05	0.508	
Pure Error	6	1067	177.8			
Total	20	11448				

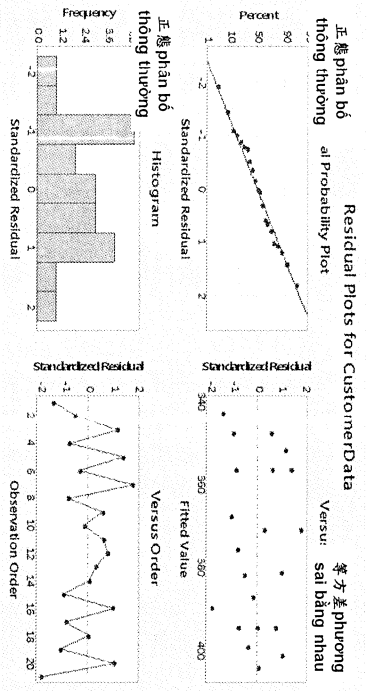
回歸P值=0.000，回歸模型是顯著的 số P hồi quy=0.000, mô hình hồi quy rõ ràng

Lack-of-fit P值=0.508，模型無失擬 số P Lack-of-fit P=0.508, mô hình phù hợp

32

標準化殘差圖 biểu đồ sai số tiêu chuẩn hóa

殘差 = 實際值 - 擬合值 sai số dư = số thực tế - số fitted



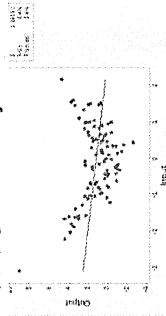
Six Sigma

31

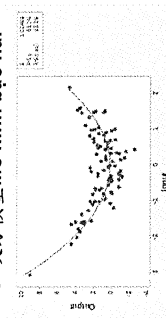
現在你認為如何？

Bây giờ bạn cho rằng thế nào?

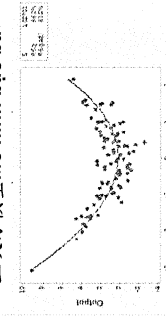
線性模型 Mô hình tuyến tính



二次方模型 Mô hình bậc hai



三次方模型 Mô hình bậc ba



哪個是“最佳”數學模式？
Cái nào là mô hình toán học “tốt nhất”？

練習題：預測Output bài tập:dự báo Output

根據你所得到的最佳數學模型進行預測：căn cứ mô hình toán học tốt nhất của bạn tiến hành dự báo:

1. 在95%的信心水平下，當Input=2，寫出Output的預測區間 dưới 95% lòng tin, khi Input=2, viết ra khoảng dự báo của Output

2. 在95%的信心水平下，當Input=3，寫出Output的預測區間 dưới 95% lòng tin, khi Input=3, viết ra khoảng dự báo của Output

開啟檔案 mở dữ liệu: Mystery.mpi

宣稱有因果關係前要小心 trước khi tuyên bố có quan hệ nhân quả phải cẩn thận

■ 有因果關係存在的地方，也存在著相關性。 ở chỗ có quan hệ nhân quả tồn tại, cũng tồn tại tính tương quan.

– 注意相關性可能不是線性的！ Chú ý tính tương quan có khả năng không phải là tuyến tính

■ 不要只因為有相關性，就認定有因果關係。 Không phải chỉ vì có tính tương quan, thì cho rằng có quan hệ nhân quả.

– 腳的大小與數學演算能力有高相關性。 Kích thước chân và năng lực tính toán toán học có tính tương quan cao

– 嬰兒的腳很小，他們甚至連數位都不認識。 chân của em bé rất nhỏ, họ thậm chí còn số đều không nhận biết.

– 4歲幼兒的腳稍微大些，通常可以數數了。 Chân của trẻ con 4 tuổi lớn hơn một tí, thông thường có thể đếm số.

– 10歲兒童的腳更大些，通常還可以做加減法。 Chân của trẻ con 10 tuổi càng lớn hơn một tí, thông thường còn có thể làm phép cộng trừ

小結 sơ kết

■ 迴歸分析可以用於找出兩個變數之間是否存在相關性。 Phân tích hồi quy có thể dùng để tìm ra tính tương quan có phải tồn tại giữa hai cái biến số

■ 迴歸分析將建立一個可用作預測的線性數學模式。 Phân tích hồi quy sẽ lập một cái mô hình toán học tuyến tính dụng để làm dự báo

■ 相關性分析可用來找出在多個變數中哪些是線性相關的。 Phân tích tính tương quan có thể dùng để tìm ra trong nhiều cái biến số những cái nào có tương quan tuyến tính

■ 宣稱有因果關係前要小心。 Trước cho rằng có tính tương quan phải cẩn thận

■ 觀察離群值—它們的影響很大。 Quan sát rời rạc- chúng ảnh hưởng rất lớn

■ 利用殘值檢查數學模式的好壞。 sử dụng sai số kiểm tra mô hình là tốt hay xấu.

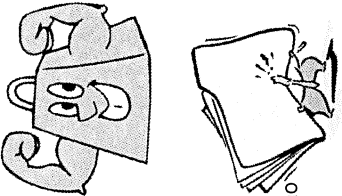
多元線性回歸模型 quy tuyến tính bội

優點 ưu điểm

- 不需要平衡設計 (Balance Design) không cần thiết kế cân bằng
- 在模式運用上有良好的功效 khi ứng dụng mô hình có hiệu quả tốt

缺點 khuyết điểm

- 一般而言無法發現交互作用 phổ biến không thể phát hiện tác động tương tác
- 技巧不純熟時容易被誤導 kỹ xảo không thành thạo dễ bị hướng dẫn sai



Six Sigma

45

多元線性回歸模型 quy tuyến tính bội

經驗值 số kinh nghiệm

觀察資料組的數量應至少達到輸入變數數目的15倍，如果有3個輸入變數，可能需至少45組觀察值。Số lượng nhóm của dữ liệu quan sát ít nhất nên đạt đến 15 cấp số lượng của biến đầu vào, nếu có ba cái biến đầu vào, khả năng ít nhất cần 45 nhóm số quan sát

Six Sigma

46

分析路徑圖 lưu trình phân tích

多元回歸模型 quy bội

計畫研究 kế hoạch nghiên cứu

收集數據 thu tập dữ liệu

實際地檢查資料 thực tế kiểm tra dữ liệu

使用下列工具分析： sử dụng phương pháp sau đây phân tích
逐步回歸 (Stepwise)

執行診斷分析 phân tích chẩn đoán

更換模式重複分析 thay đổi mô hình lặp lại phân tích

制訂決策 chế định quyết sách

Six Sigma

47

回顧這個練習 nhìn về bài tập này

輸出變數被一組輸入變數所 “驅動” biến đầu ra bị một nhóm biến đầu vào thúc đẩy

$$y = f(x_1, x_2, \dots, x_n)$$

資料依出現的次序來記錄 dữ liệu ghi chép theo thứ tự xuất hiện

輸出變數 (Y) : 腳踏車騎乘時間 biến đầu ra (Y) : thời gian đi xe đạp

可能的輸入變數有 : biến số đầu vào khả năng có :
天氣 thời tiết : Weather (X₁)
座椅高度 độ cao ghế : Seat Height (X₂)
溫度 nhiệt độ : Temperature (X₃)
起始時間 thời gian bắt đầu : Start Time (X₄)
輪胎胎壓 : Tire Pressure (X₅)
路線 đường : Route (X₆)
風速 tốc độ gió : Wind Speed (X₇)

目的 mục đích :

通過對一組輸入變數進行分析，創建出一個數學模式以預測輸出變數 (騎乘時間)。Thông qua biến đầu vào của một nhóm phân tích, lập ra một mô hình toán học để dự báo biến đầu ra (thời gian đi xe)

開啟檔案 mở dữ liệu :
Bike Trip2.MPJ

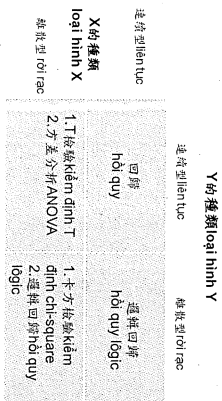
Six Sigma

48

11-12

還記得嗎？圖形分析法 còn nhớ nữa không? Phương pháp thị đồ

- 確定輸入因數的種類(離散或連續) xác định từng loại biến đầu vào(rời rạc hoặc liên tục).
 - 找出有影響的變異組成(active components of variation, Xs) tìm ra phần hợp thành của biến thể có ảnh hưởng.
 - 移除沒有影響的因數 từ bỏ biến số không có ảnh hưởng.
 - 決定有影響因數的效應大小 quyết định hiệu ứng lớn hoặc nhỏ cho biến số có ảnh hưởng.
 - 選擇能改善輸出結果的輸入因數的組合: chọn tổ hợp của biến số đầu vào có thể cải tiến kết quả đầu ra.
- 最小化耗車時間 thời gian đi xe đáp ngắn nhất.



不同類型的模式 các loại mô hình

- (a) 線性模式(經常使用) mô hình tuyến tính(thường dùng)
 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \text{error}$
- (b) 多項式模式(經常使用) mô hình nhiều bậc(cường dùng)
 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \text{error}$
- (c) 交互作用模式(較少用，收集的數據可能無法建立交互作用模型) mô hình tác động tương tác(ít khi dùng, dữ liệu thu tập không thể lập mô hình tác động tương tác)
 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \text{error}$
- (d) 二次方模式(較少用，原因與交互作用模型相同) mô hình bậc hai (ít khi dùng, nguyên nhân với mô hình tác động tương tác giống nhau)
 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \text{error}$

本單元主要在介紹線性模型 bài này chủ yếu giới thiệu mô hình tuyến tính

建立數學模型 lập mô hình toán học

回歸分析方法- 線性模式 phương pháp phân tích hồi quy- mô hình tuyến tính

回歸分析的步驟 các bước phân tích hồi quy

■ 篩選因子 lực chọn biến số :

— 逐步 (stepwise) : 此程序可幫助對所有的輸入變量進行篩選，以產生一“最佳”模型。 **stepwise** :B-rúc này để tiến hành lựa chọn cho tất cả biến số, đề sinh ra một mô hình “tốt nhất”.

■ 決定回歸模型 : quyết định mô hình hồi quy

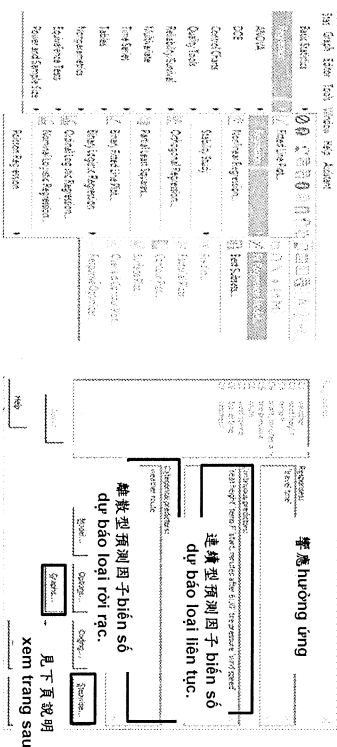
— 回歸 (Regression) : 一旦 “最佳” 模型確定，回歸分析可以用來針對該模型做更詳細的診斷。 **hồi quy**: một khi xác định mô hình “tốt nhất”， phân tích hồi quy có thể dùng để chẩn đoán kỹ cho mô hình này.

Six Sigma



Minitab : 擬合回歸模型 lập mô hình hồi quy

◆ Stat>Regression>Regression



Six Sigma

55

回歸分析的步驟 (續) trình tự phân tích hồi quy(tiếp)

■ 用逐步回歸 (Stepwise Regression)找出有影響的因數，移除沒有影響的因數。 Dùng hồi quy từng bước tìm ra biến số có ảnh hưởng, từ bỏ biến số không có ảnh hưởng

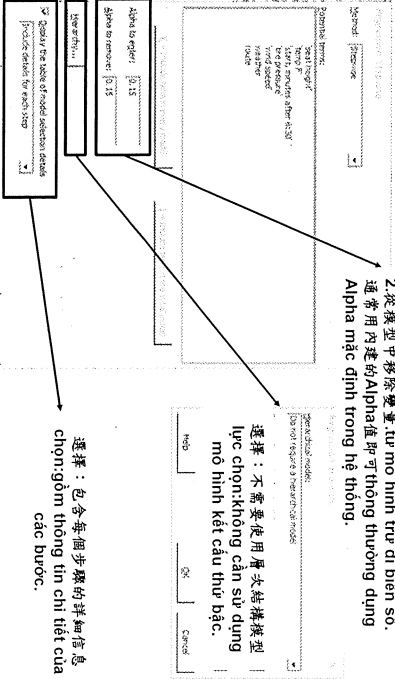
■ 對有影響的因數進行回歸分析 (Regression)並執行診斷檢查： với biến số có ảnh hưởng tiến hành phân tích hồi quy và kiểm tra chẩn đoán:

- 模式是否顯著？ Mô hình có phải rõ ràng
- 對預測是否有用？ Với dự báo có phải có ích？
- 殘值是否正當？ Sai số có phải bình thường？
- 是否有任何證據顯示擬合不佳 (lack of fit)？ Có phải có bất cứ chứng cứ nào hiện thị không phù hợp (lack of fit)？

Six Sigma

54

回歸：逐步 hồi quy: từng bước



1.增加變量到模型 thêm biến số vào mô hình.
2.從模型中移除變量, từ mô hình trừ đi biến số.
Alpha mặc định trong hệ thống.

選擇：不需要使用層次結構模型
lựa chọn: không cần sử dụng
mô hình kết cấu thứ bậc.

選擇：包含每個步驟的詳細信息
chọn: gồm thông tin chi tiết của
các bước.

Six Sigma

56

逐步回歸：原理từng bước:nguyên lý

■ 逐步 (Stepwise) 通過在回歸模型中減少及增加變量來找到一個有效的預測因子組 Stepwise là thông qua từ mô hình từ đi hoặc thêm biến số để tìm được một nhóm biến số dự dự bảo có hiệu quả.

■ Minitab提供三種最常用的程序：Minitab có ba phương pháp thường dụng

逐步 (增加及減少變量) Stepwise(thêm và từ biến số)
前進法 (增加變量) Forward selection(thêm biến số)
後退法 (減少變量) Backward elimination(từ biến số)

Six Sigma

逐步回歸：結果hỏi quy từng bước:kết quả

Stepwise Selection of Terms									
C:\Program Files\MiniTab\MiniTab.exe - Project: D:\Data\Stepwise.mtw									
Worksheet: Step1									
Step 1									
Term	S	SE	T	P	Step 2	Step 3	Step 4	Step 5	Step 6
Constant	52.574	0.203	258.9	0.000	0.000	0.000	0.000	0.000	0.000
Seat height	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Route	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Weather	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R-sq	0.000				0.000	0.000	0.000	0.000	0.000
R-sq (adj)	0.000				0.000	0.000	0.000	0.000	0.000
Adjusted	0.000				0.000	0.000	0.000	0.000	0.000
Model	0.000				0.000	0.000	0.000	0.000	0.000
Sum of Squares	0.000				0.000	0.000	0.000	0.000	0.000
Mean Square	0.000				0.000	0.000	0.000	0.000	0.000
F-Value	0.000				0.000	0.000	0.000	0.000	0.000
P-Value	0.000				0.000	0.000	0.000	0.000	0.000

Six Sigma
項的逐步選擇只到步驟4，意味着再多增加變量，模型不會變得更好R-sq代表結果的變異有多少百分比可由該模型解釋。 Sự lựa chọn mục chỉ đến bước 4, có nghĩa là mức độ

逐步：原理từng bước:nguyên lý

■ 入選用a值 số a về thêm
此數值用以決定任何一個不在目前模式中的因數，應否增加到該模式中來。 Số này dùng để quyết định bất cứ một cái biến số nào bầy giờ chưa có trong mô hình, nên thêm vào mô hình hay không?

– 將此a水準與每個不在該模式中的因數的p值較比較，若某因數的p值較該a值小，則該因數應考慮增加入模式中。 該a值介於0跟1之間。 So sánh số a này với số P của mỗi biến số không trong mô hình này, nếu số P của biến số nhỏ hơn số a, thì nên suy nghĩ thêm biến số vào mô hình. Số a này giữa 0 và 1

資料來源： Minitab線上輔助 nguồn gốc dữ liệu từ giúp trên Minitab

■ 剔除用a值 số a về trừ
此數值用以決定任何一個已在目前模式中的因數，應否自該模式中剔除出去。 Số này dùng để quyết định bất cứ một cái biến số nào bầy giờ đã có trong mô hình, nên suy nghĩ từ mô hình trừ đi hay không

– 將此a值與每個在該模式中的因數的p值相比較，若某一因數的p值較該a大，則該因數考慮自該模式中剔除。 該a值介於0跟1之間。 So sánh số a này với số P của mỗi biến số không trong mô hình này, nếu số P của biến số lớn hơn số a, thì nên suy nghĩ từ mô hình trừ đi biến số. Số a này giữa 0 và 1

資料來源： Minitab線上輔助 nguồn gốc dữ liệu từ giúp trên Minitab

Six Sigma

逐步回歸：推論 hỏi quy từng bước:suy luận

■ 使用 R-square比較具有相同數量預測因數的模式 sử dụng R-square so sánh mô hình có biến số dự tính số lượng bằng nhau.

■ 使用 R-sq (預測) 及 Mallow的 Cp值比較具有不同數量預測因子的模式 sử dụng R-sq(Pred) và số Cp của Mallow để so sánh mô hình có số lượng biến số dự đoán không giống nhau.

• 尋找 R-sq (預測) 較高， Mallow的 Cp值較低 (與模型中預測因子數目相近) 之模型。 Tìm mô hình có R-sq (Pred) cao mà số Cp của Mallow thấp (số lượng tiếp cần số lượng biến số dự đoán)

• 較低之 Mallow的 Cp值意味著該模型在估計真實回歸係數及預估未來觀察值時有較小的變異 số Cp thấp thì có ý nghĩa là mô hình này khi dự tính hệ số hồi quy thực tế và số quan trắc thường lại có biến thể nhỏ hơn

• 擬合不佳的模型其 Mallow的 Cp值大於預期因子數目 mô hình không phù hợp thì số Cp của Mallow lớn hơn số lượng biến số dự đoán.

Six Sigma

逐步回歸：推論 hồi quy từng bước:suy luận

■ AICc 和 BIC：Minitab20 新增了这两个指标，当您显示逐步方法每个步骤的详细信息时或者显示分析的扩展结果时，Minitab 会多显示这两个统计量。AICc và BIC：Minitab20 mới thêm hai chỉ số này, khi bạn chọn hiện thị thông tin chi tiết của từng bước về phương pháp Stepwise, Minitab sẽ hiện thị thêm hai số liệu thống kê này.

■ AICc 是更正的 Akaike 信息标准，BIC 是 Bayesian 信息标准，使用这些统计量可以比较不同的模型，较小的值比较合适。AICc là chỉnh xác tiêu chuẩn thông tin Akaike, BIC là tiêu chuẩn thông tin Bayesian, sử dụng số liệu thông kê này có thể so sánh những mô hình khác nhau, giá trị càng nhỏ thì phù hợp hơn.

Six Sigma

回歸分析：模式檢定
phân tích hồi quy: kiểm định mô hình

- 對於高品質的預測與優化，數學模式應該。Với dự báo và ưu hóa có chất lượng cao, mô hình toán học nên
 - 顯著：檢視回歸模型的 P 值 rõ ràng: kiểm tra số P của mô hình hồi quy
 - 能有效預測：R2 大於 70% có thể dự đoán có hiệu quả: R2 lớn hơn 70%
 - 有正常的殘值 sai số bình thường
 - 輸入變量之間沒有共線性 giữa biến số không có tính cùng tuyến

Six Sigma

一個最佳模型應該盡可能簡單一直到不能再簡單為止
một mô hình tốt nhất nên có càng đơn giản trực đến không đơn giản được nữa

- The best modle is the one that is as simple as possible, but not simpler
- Albert Einstein

“最佳”模型包含： mô hình “tốt nhất” gồm

- Start time
- Seat height
- Rout
- Weather

Six Sigma

回歸：模型 hồi quy: mô hình

最佳回歸方程的系数 hệ số của mô hình tối ưu

Coefficients					
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	38.814	0.706	54.78	0.000	
Seat height	-2.052	0.120	-17.42	0.000	1.04
Start minutes after 0:30	0.2908	0.0159	20.98	0.000	1.07
Weather	-3.905	0.405	-9.65	0.000	1.04
Route	4.001	0.493	8.11	0.000	1.38
City route	-2.824	0.503	-5.62	0.000	1.40
Country route					
Regression Statistics					
Term	Adjusted R-Square	F-Value	Prob > F	Adjusted T-Value	Prob > T
Constant					
Seat height	0.9230	10.00	0.000	-17.42	0.000
Start minutes after 0:30	0.9230	10.00	0.000	20.98	0.000
Weather	0.9230	10.00	0.000	-9.65	0.000
Route	0.9230	10.00	0.000	8.11	0.000
City route	0.9230	10.00	0.000	-5.62	0.000
Country route	0.9230	10.00	0.000		

逐步選擇後，找到的最佳回歸方程，因為 Weather và route 是類別預測變量，展開後有六個方程式 sau lựa chọn, tìm ra phương trình tối ưu, vì Weather và route là biến loại rời rạc, nên phát triển ra 6 phương trình.

Six Sigma

圖形：殘差圖 biểu đồ biểu đồ sai số

Stat>Regression>Regression>Fit Regression Model

Figure 1: Regression Model

Fit

Save

Standardized

Residuals For Plots

Residuals plots

Input residuals

Plotting residuals

Standardized residuals

Residuals versus the variables

Four in one

OK

Cancel

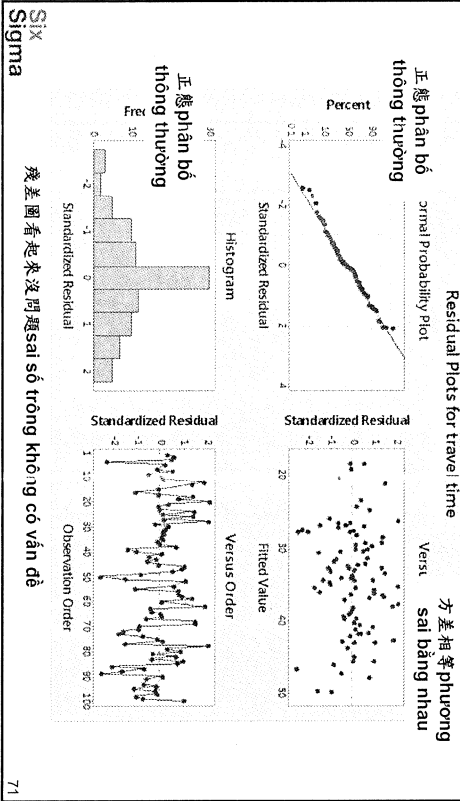
選擇 Pareto 效應圖，可顯示模型中各預測變量對變異的影響大小，chọn Effecta Plots-Pareto, có thể hiển thị các biến ảnh hưởng tới biến thể như thế nào

選擇標準化、四合一圖 chọn Standardized、Four in one

Six Sigma

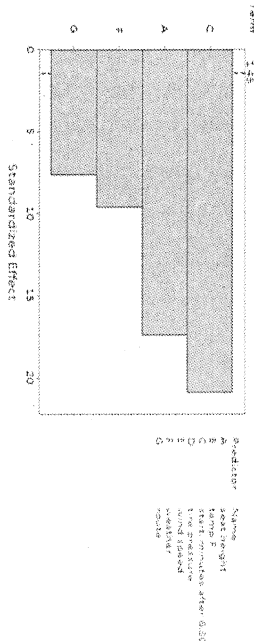
69

標準化四合一殘差圖 biểu đồ Four in one



圖形：效應 Pareto 圖 thị đồ: thị đồ Pareto của hiệu ứng

Pareto Chart of the Standardized Effects (Response is travel time (s) 0.15)



■ 预测变量对响应的影响从大到小排序 thứ từ lớn đến nhỏ về sự ảnh hưởng của các biến: Start, minutes after 6:30 > seat height > weather > route

輸入變數 Xs 的共線性 (使用變異膨脹係數) tính cùng tuyến của các biến đầu vào Xs (sử dụng VIF)

- 在有超過 1 個以上自變數 (X) 的情況下，如果 Xs 間的相關性太高，則會使得回歸分析的結果變成無效。dưới tình hình có biến số (X) nhiều hơn một cái, nếu tình tương quan giữa Xs quá cao, thì sẽ làm cho kết quả phân tích hồi quy là vô hiệu.
- 逐步回歸可用來減輕這類問題。Hồi quy từng bước có thể dùng để giảm nhẹ vấn đề này
- 另一方法是使用變異膨脹係數 (Variance Inflation Factor, VIF) 來檢查高度相關的預測因數。Phương pháp khác là sử dụng Variance Inflation Factor, VIF để kiểm tra biến dự báo có tính tương quan mức độ cao.

共線性：Variance Inflation Factor, VIF
cùng tuyến tính: Variance Inflation Factor, VIF

變異膨脹係數 (VIF) 用來檢定一預測因數是否與其他因數有強烈線性相關。Variance Inflation Factor, VIF dùng để kiểm định một biến số dự báo nào đó có phải có tính tương quan mạnh với biến số khác.

VIF 衡量出如果預測因數間相關的話，回歸係數會增大多少。VIF sẽ đánh giá nếu giữa biến số dự báo có tương quan thì hệ số hồi quy sẽ tăng bao nhiêu

VIF = 1 表示不相關；VIF > 1 則表示相關。VIF = 1 chứng tỏ không tương quan; VIF > 1 thì chứng tỏ liên quan

當 VIF 大於 5-10 時，回歸係數的計算是很有問題的。Khi VIF lớn hơn 5-10, sự tính toán của hệ số hồi quy sẽ rất có vấn đề.

您應該考慮分開共線性的作法，收集更多資料，刪除因數，使用不同因數，或用最小平方回歸法。Bạn nên suy nghĩ phương pháp tách ra cùng tuyến tính, thu nhiều dữ liệu, xóa biến số, sử dụng biến số không giống nhau, hoặc dùng phương pháp hồi quy bình phương nhỏ nhất.

資料來源：Minitab 線上輔助 nguồn gốc dữ liệu: từ giúp trên Minitab

本例模式中的 VIF 值為多少？Số VIF trong mô hình này là bao nhiêu

Six Sigma

73

回歸分析：結果 phân tích hồi quy: kết quả

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	35.814	0.706	50.73	0.000	
seat height	-2.092	0.120	-17.42	0.000	1.04
start, minutes after 6:30	0.2908	0.0139	20.93	0.000	1.07
weather					
sunny	-3.905	0.405	-9.65	0.000	1.04
route					
city route	4.001	0.493	8.11	0.000	1.33
country route	-2.824	0.502	-5.62	0.000	1.40

VIF 皆為 1 左右，Xs 間無相關性，正是我們希望的。VIF đều khoảng là 1, giữa Xs không có tính tương quan, chính là chúng ta hy vọng

Six Sigma

74

診斷分析：回顧 phân tích chẩn đoán: nhìn về

- 模式是顯著的， mô hình là rõ ràng
- 模式能有效預測 mô hình có thể dự đoán có hiệu quả
- 沒有顯著的擬合不佳問題， không có vấn đề về không phù hợp rõ ràng
 - 殘值正常 sai số bình thường
 - 是常態的。 Là thái thường
 - 沒有型態，即隨機分佈。 Không có hình thái cố định, tức là phân bố tùy cơ
- VIF 顯示沒有共線性的問題， VIF hiện thị không có vấn đề cùng tuyến tính

Six Sigma

75

運用模式做預測
áp dụng mô hình dự báo

Six Sigma

11-19

運用“最佳”模式：áp dụng mô hình “tốt nhất”

對下列情景進行 travel time 的預測： dự báo travel time với tình hình sau đây

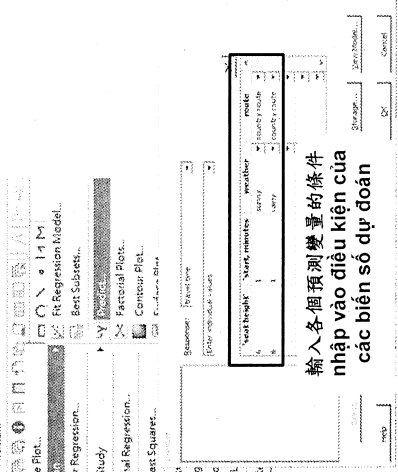
- 情景 1 tình hình 1
95%的信心水平下，晴天最短騎乘時間（travel time）會落在哪個區間範圍？ Dưới 95% lòng tin, khi trời nắng thời gian đi xe ngắn nhất sẽ trong khoảng thời gian nào?
- 情景 2 tình hình 2
95%的信心水平下，雨天最短騎乘時間（travel time）會落在哪個區間範圍？ Dưới 95% lòng tin, khi trời mưa thời gian đi xe ngắn nhất sẽ trong khoảng thời gian nào?

Six Sigma

77

回歸分析 phân tích hồi quy

Stat>Regression>Regression> Predict



輸入各個預測變量的條件
輸入 vào điều kiện của
các biến số dự đoán

Six Sigma

78

進行預測：結果 tiến hành cứ đoán:kết quả

晴天 trời nắng

Variable	Setting
feet height	6
start minutes after 6:30	1
weather	sunny
route	country route

Prediction
Est. SE Est. 95% CI 95% PI
16.8265 0.52375 (15.7814, 17.8716) (11.3395, 20.9164)

雨天 trời mưa

Variable	Setting
feet height	6
start minutes after 6:30	1
weather	rainy
route	country route

Prediction
Est. SE Est. 95% CI 95% PI
20.7316 0.57824 (19.5825, 21.8807) (16.7181, 24.7500)

擬合值 số fitted=16.8265
95%預測區間 (12.84 , 20.82) khoảng dự báo 95%: (12.84 , 20.82)

擬合值 số fitted=20.7316
95%預測區間 (16.71 , 24.75) khoảng dự báo 95%: (16.71 , 24.75)

Six Sigma

79

練習 bài tập

Q1:打開檔案 Hald.MTW，寫下最佳回歸方程式.mở ra dữ liệu Hald.MTW, ghi xuống mô hình hồi quy tốt nhất


Q2:打開檔案 Collinearity.MTW，回答下面問題： mở ra dữ liệu Collinearity.MTW， trả lời vấn đề sau:

- 對Yield有影響的是哪（幾）個因子.những biến số nào ảnh hưởng Yield.
- 對Yield有影響的因子該設在什麼條件才能使Yield最高 biến số ảnh hưởng Yield nên đặt ở điều kiện nào mới làm cho Yield cao nhất.
- 95%信心水平下，最高Yield的預測區間為何？ Dưới 95% lòng tin, khoảng dự báo của Yield cao nhất là bao nhiêu?

Six Sigma

80

11-20

<div data-bbox="172 1211 220 1256">  </div> <div data-bbox="196 1529 233 1688"> <p>小結 sơ kết</p> </div>	<div data-bbox="260 1249 520 1986"> <ul style="list-style-type: none"> ■ 介紹多元回歸分析的觀念 giới thiệu khái niệm về phân tích hồi quy bội. ■ 多元回歸分析技巧回顧 nhìn về kỹ xảo phân tích hồi quy bội <ul style="list-style-type: none"> ✓ 逐步回歸 hồi quy từng bước ✓ 最終回歸分析 phân tích hồi quy cuối cùng ■ 討論如何處理殘值異常和共線性的問題 thảo luận xử lý sai số bất thường và tính cùng tuyến như thế nào. </div> <div data-bbox="675 1930 722 2009"> <p>Six Sigma</p> </div> <div data-bbox="707 1211 722 1238"> <p>81</p> </div>
--	--

