BÀI THỰC HÀNH R PHẦN KIỂM ĐỊNH

1. Các ký hiệu chú ý :
   1. alternative : kiểu character hoặc string. Một trong các giá trị “greater”, “less” hoặc “two.side”. Chỉ định vùng cần kiểm tra của giá trị ‘*mu*’. (mặc định “two.side”)

ví dụ :

- … hãy cho biết sáng kiến có thực sự làm giảm mức tiêu hao nhiên liệu hay không?.

=> alternative = “less”

- có thể nói thời gian đông cứng của vật liệu đã thấp hơn 72 giờ hay chưa. (72 giờ đóng vai trò ‘*mu*’)

=> alternative = “less”

- Kiểm định tính chính xác của của 1 giá trị trung bình cho trước

=> alternative = “two.side”

* 1. conf.level : Độ tin cậy ( 0 <= conf.level <= 1) (conf.level = 1 – <mức độ ý nghĩa>)
  2. mu : giá trị trung bình của x hoặc giá trị cần kiểm định từ alternative
  3. sigma.x : Độ lệch chuẩn của x
  4. n.x : kích thước mẫu x
  5. s.x : Độ lệch chuẩn của mẫu x
  6. mean.x : Giá trị trung bình của mẫu x
  7. var.equal : Giá trị TRUE or FALSE. Nên có khi so sánh 2 mẫu x, y. (FALSE)
  8. n : số lượng kiểm tra (bỏ qua khi x là trong 1 ma trận)
  9. p : xác suất thành công ( xác suất cần kiểm tra) ( 0 <= p <= 1)
  10. correct : (FALSE)

1. Hàm sử dụng :

|  |  |
| --- | --- |
| z.test | x, alternative, mu, sigma.x, conf.level |
| t.test | x, alternative, mu, conf.level |
| tsum.test | mean.x, s.x, n.x, mu, alternative, var.equal, conf.level |
| prop.test | <số lượng thành công>, n, p, alternative, conf.level, correct = FALSE |

TƯƠNG QUAN VÀ HỒI QUY

1. Các ký hiệu :

- x, y : vector

-

1. Hệ số tương quan : (coeficient of correlation)

- Hàm : cor(x, y)

- công thức :

2. Ước lượng các hệ số hồi quy : (linear model) (hàm xu thế tuyến tính)

- Hàm : lm (x ~ y)

- Công thức :

+ Với (β\_0 ) , (β\_1) là hai hệ số hồi quy thực nghiệm được ước lượng theo công thức :

1. Vẽ biểu đồ miêu tả dữ liệu (biểu đồ phân tán) :

- plot(x, y)

4. Tạo object chứa các thông tin về hồi quy :

- reg <- lm(y ~ x)

5. Vẽ đường hồi quy thực nghiệm :

- abline(reg)

6. Tính sai số của dữ liệu được cung cấp và hàm hổi quy thực nghiệm tại các điểm quan sát :

- Hàm sai số : residuals(reg) hoặc resid(reg)

7. Dự báo giá trị và tìm khoảng tin cậy :

7.1. Dùng predict :

- predict(reg, newdata=data.frame(y = <giá trị cần dự đoán>), interval=’prediction’, level = 0.95)

|  |
| --- |
| fit lwr upr  1 8477.327 8459.016 8495.638 |

- Với fit : tâm của khoảng dự báo và là ước lượng điểm của y0 của giá trị dự báo y0

- lwr : cận dưới của khoảng ước lượng

- upr : cận trên của khoảng ước lượng

7.2. Sử dụng lý thuyết :

- Đưa ra công thức khoảng tin cậy :

🡪 Những giá trị cần tìm : y0mu, s2, xtb, Sxx, t(n-2.a/2)

- Nhập x0 và tính y0mu : ( Đề bài yêu cầu dự báo khi x = x0 thì y0mu bằng bao nhiêu )

|  |
| --- |
| > x0 <-2017  > beta0mu <- coef(reg)[1]  > beta1mu <- coef(reg)[2]  > y0mu <- beta0mu+beta1mu\*x0 |

- Tính s2 (sbp) :

|  |
| --- |
| > n <- length(tgian)  > sbp <- sum(resid(reg)^2)/(n-2) |

- Tính xtb, Sxx :

|  |
| --- |
| > xtb <- mean(tgian)  > Sxx <-sum(tgian^2)-n\*xtb^2 |

- Định nghĩa biến “phân vị” để ghi giá trị t(n-2.a/2) :

|  |
| --- |
| > phanvi <- qt(0.99,4) |

🡪 Áp dụng công thức khoảng tin cậy tính bán kính khoảng tin cậy :

|  |
| --- |
| > bkinh <- phanvi\*sqrt(sbp\*(1/n + (x0 – xtb)^2/Sxx)) |

🡪 Tính khoảng ước lượng :

|  |
| --- |
| > y0mu+c(-1,1)\*bkinh |

8. Tổng kết :

|  |  |
| --- | --- |
| lm(x~y) | Hàm hồi quy tuyến tính x theo y |
| plot(y, x)  reg <- lm(x~y)  abline(reg) | Vẽ biểu đồ mô tả giữ liệu (phân tán) và đồ thị hàm hồi quy tuyến tính |
| resid(reg) | Sai số của dữ liệu và hàm hồi quy thực nghiệm tại cảm điểm |
| predict(reg, newdata=data.frame(y = y0), interval = ‘prediction’, level = 0.95) | Dự báo khi y = y0 thì x thuộc khoảng nào  fit : tâm của khoảng dự báo  lwr : cận dưới  upr : cận trên |