TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI

**KHOA CÔNG NGHỆ THÔNG TIN**



BÀI BÁO CÁO

Xác Suất Thống Kê

Giảng viên hướng dẫn : Trần Văn Long

**Sinh viên thực hiện** : Trịnh Thành Nam

Hoàng Trung Nguyên

Vũ Thế Nguyên

**Lớp** : Công nghệ thông tin 1

**Khoá**  : 62

**Hà Nội, 2022**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ho va ten | Ma sinh vien | KhoaDangTheoHoc | DiemTB | ThoiGianTuHoc |
| Dang Tran Trung Dung | 201214499 | CNTT | 8.00 | 1 |
| Tran Van Thang | 201203116 | CNTT | 5.75 | 35 |
| Nguyen Duc Dung | 201200917 | CNTT | 6.00 | 0.5 |
| Pham Thi Ha | 201240940 | CNTT | 7.00 | 21 |
| Dang Thai Ha | 201200574 | CNTT | 6.00 | 7 |
| Tran Cong Thanh | 201240331 | CNTT | 9.65 | 7 |
| Nguyen Ha Phuong | 201206094 | CNTT | 9.19 | 7 |
| Hoang Huy Hieu | 201243214 | CNTT | 7.00 | 7 |
| Luu Hong Quan | 201210020 | CNTT | 7.00 | 14 |
| Khuat Dinh Quang | 201200865 | CNTT | 7.1 | 4 |
| Tran Trung Hieu | 201114556 | CNTT | 3.42 | 22 |
| Trinh Thanh Nam | 201201058 | CNTT | 8.4 | 13 |
| Nguyen Dinh The | 201203812 | CNTT | 9.3 | 10 |
| Vu The Nguyen | 201201974 | CNTT | 7.3 | 16 |
| Nguyen Tien Anh | 201256738 | CNTT | 8.45 | 15 |
| Dang Ba Kien | 206373377 | CNTT | 8.3 | 20 |
| Vu Van Nam | 201238419 | CNTT | 7.45 | 19 |
| Hoang Thi Hien | 204598333 | CNTT | 6.45 | 5 |
| Tran Huu Cong | 201713324 | CNTT | 6.1 | 12 |
| Vu Van Duc | 203746949 | CNTT | 8.56 | 40 |
| Vu Bao Toan | 204838334 | CNTT | 8.9 | 25 |
| Phung Phi Van | 203467892 | CNTT | 7.13 | 14 |
| Tran Van Cuong | 204373795 | CNTT | 7.6 | 21 |
| Dinh Khac Tuan | 208447383 | CNTT | 9.1 | 26 |
| Vu Manh Hung Quan | 200333859 | CNTT | 8.13 | 48 |
| Tran Khac Tung | 207373322 | CNTT | 5.3 | 10 |
| Ha Duc Chinh | 202382414 | CNTT | 7.32 | 32 |
| Tran Duc Tam | 209976541 | CNTT | 6.4 | 12 |
| Tran Cong Chien | 202008866 | CNTT | 7.9 | 12 |
| Nguyen Tran An Giang | 204811779 | CNTT | 8.1 | 15 |
| Tran Ha An | 208191788 | CNTT | 8.3 | 19 |
| Nguyen Thanh Tam | 208737228 | CNTT | 7.5 | 12 |
| Phung Huy Doan | 201239399 | CNTT | 6.78 | 15 |
| Ho Duc Hieu | 208899121 | CNTT | 8.3 | 13 |
| Hoang Thi My Tam | 203928244 | CNTT | 8.64 | 22 |
| Vu Minh Thinh | 203838281 | CNTT | 9.13 | 48 |
| Le Thi Chau Giang | 203727864 | CNTT | 8.6 | 36 |
| Ho Hai Anh | 203281256 | CNTT | 8.15 | 38 |
| Tran Hai An | 207747372 | CNTT | 9.3 | 18 |
| Ha Thi Thao Mai | 202883822 | CNTT | 7.34 | 16 |
| Le Thi Thu Hien | 208473953 | CNTT | 7.3 | 12 |
| Phung Chi Thien | 202728835 | CNTT | 7.45 | 36 |
| Vu Minh Duc Hoang | 201835321 | CNTT | 7.83 | 34 |
| Phan Cong Chien | 204966745 | CNTT | 8.3 | 15 |
| Dang Thu Ha | 203937379 | CNTT | 7.9 | 13 |
| Vu Thi Hanh Nhan | 201828292 | CNTT | 7.93 | 19 |
| Ngo Van Cu | 201336236 | CNTT | 7.88 | 17 |
| Nguyen Trung Kien | 203677675 | CNTT | 7.2 | 15 |
| Vu Minh Ha | 203728182 | CNTT | 9.3 | 47 |
| Vu Trong Nghia | 208484386 | CNTT | 6.1 | 10 |
| Hoang Vu Minh Anh | 201828285 | CNTT | 8.15 | 48 |
| Ho Sy Han | 207716638 | Công trình | 7.3 | 23 |
| Vu Van Binh | 204589375 | Công trình | 7.8 | 48 |
| Tran Thanh Tung | 208373829 | Công trình | 8.13 | 14 |
| Trinh Dinh Quang | 201838292 | Công trình | 7.33 | 23 |
| Hoang Thi Cam Van | 200134526 | Công trình | 7.82 | 24 |
| Ly Thi Quynh Giao | 204993766 | Cơ điện tử | 8.2 | 21 |
| Vu Hoang Hai Linh | 209333472 | Cơ điện tử | 7.61 | 23 |
| Hoang Le Anh Dung | 201300192 | Cơ khí | 4.60 | 2 |
| Tran Bao Anh | 204838392 | Cơ khí | 6.2 | 16 |
| Vu Hoang Mai | 201364759 | Cơ khí | 7.8 | 30 |
| Trinh My Ha | 201238489 | Cơ khí | 6.52 | 7 |
| Nguyen Hai Dang | 202046898 | Kế toán | 5.6 | 8 |
| Mac Phuong Ha | 202043029 | Kế toán | 7.73 | 12 |
| Le Thi Nguyet | 202059421 | Kinh tế vận tải | 7.00 | 14 |
| Nguyen Thi Thuy Trang | 202011947 | Kinh tế vận tải | 6.5 | 10 |
| Dinh Gia Bao | 202003452 | Kinh tế vận tải | 5.8 | 7 |
| Phung Van Tung | 204578976 | Kinh tế vận tải | 8.9 | 60 |
| Dang Thi Nhai | 203213689 | Kinh tế Vận tải | 7.8 | 46 |
| Le duc trung | 202030334 | Kinh tế vận tải | 7.4 | 10 |
| Do Gia Han | 202205456 | Kinh tế vận tải du lịch | 8.5 | 15 |
| Nguyen Minh Tu | 202030499 | Kinh tế vận tải du lịch | 6.52 | 8 |
| Doan Trong Hieu | 202055496 | Kinh tế vận tải ô tô | 7.4 | 10 |
| Nguyen Hong Quan | 202001485 | Kinh tế vận tải thủy bộ | 6.35 | 9 |
| Nguyen Thi Ha Giang | 202030494 | Kinh tế vận tải thủy bộ | 8.45 | 13 |
| Nguyen Thi Thu Huong | 202034024 | Kinh tế xây dựng Công trình giao thông | 8.32 | 12 |
| Tran Hoang Diep | 201895452 | Kỹ thuật cơ khí động lực | 5.5 | 10 |
| Nguyen Van Cuong | 201503731 | Kỹ thuật điện | 9.2 | 10 |
| Duong Minh Chien | 201503789 | Kỹ thuật điện | 7.4 | 8 |
| Nguyen Tung Lam | 201503820 | Kỹ thuật điện | 7.9 | 9 |
| Le Dinh Ninh | 201405728 | Kỹ thuật điện tử - viễn thông | 6.8 | 9 |
| Nguyen Van Hiep | 201413971 | Kỹ thuật điện tử - viễn thông | 6.7 | 5 |
| Nguyen Gia Huy | 201404033 | Kỹ thuật điện tử - viễn thông | 6.3 | 10 |
| Ngo Tri Nam | 201604367 | Kỹ thuật điều khiển | 8.5 | 12 |
| Nguyen Duc Trong | 201604243 | Kỹ thuật điều khiển & TĐH | 7.6 | 11 |
| Tran Duc Hoat | 201604364 | Kỹ thuật điều khiển & TĐH | 5.3 | 8 |
| Tran Van Tuan | 202039482 | Kỹ thuật Giao thông đường bộ | 7.3 | 9 |
| Bui Minh Ngoc | 202016448 | Kỹ thuật Ô tô | 6.5 | 8 |
| Tran Dinh Quan | 202034955 | Kỹ thuật xây dựng Cầu hầm | 6.5 | 13 |
| Nguyen Hoang Viet | 204678999 | Logistics | 9.2 | 65 |
| Tran Dinh Phong | 202034029 | Logistics | 8.63 | 14 |
| Dinh Cong Son | 202034991 | Logistics | 8.91 | 15 |
| Nguyen Son Tung | 204304748 | Môi trường | 9.00 | 35 |
| Ha Thi Cau | 202005642 | Quản lý xây dựng | 8.2 | 15 |
| Bui Thanh Cong | 202003746 | Quản trị kinh doanh | 7.5 | 10 |
| Hoang Hong Diep | 200394233 | Quản trị kinh doanh | 6.3 | 9 |
| Bui Thi Bich Mai | 203040230 | Quản trị kinh doanh | 6.7 | 5 |
| Vu Dinh Sang | 200474838 | Toán Tin | 8.25 | 46 |
| Vu Van Duc | 202839444 | Toán Tin | 8.91 | 39 |
| Phi Thi Thuy Kieu | 200254878 | Toán ứng dụng | 8.5 | 13 |

Đề bài

Lấy một mẫu gồm 100 sinh viên năm thứ hai trường Đại học Giao Thông Vận Tải và hỏi về điểm trung bình học kỳ I năm thứ nhất và thời gian tự học hàng tuần

Câu 1: Trình bày bảng phân phối tần số và tần suất về thời gian tự học hàng tuần của sinh viên

Câu 2: Tính trung bình và độ lệch tiêu chuẩn về thời gian tự học hàng tuần của sinh viên

Câu 3: Với độ tin cậy 90%, tìm khoảng tin cậy cho thời gian tự học trunh bình hàng tuần của một sinh viên

Câu 4: Với mức ý nghĩa 5%, kiểm định ý kiến thời gian tự học trung bình hàng tuần của sinh viên là 8 giờ

Câu 5:

a, Tính hệ số tương quan giữa thời gian tự học hàng tuần và điểm trung bình học kỳ

b, Tìm hàm hồi quy tuyến tính của điểm trung bình học kỳ theo thời gian tự học hàng tuần

C, Với độ tin cậy 95%, tìm khoảng tin cậy cho điểm trung bình học kỳ của sinh viên với thời gian tự học là 5 giờ

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| > install.packages("BSDA")  >install.packages(“readxl”)  > require("BSDA")  > library(readxl)  > data<-read\_excel("data.xlsx")  **câu 1 Bảng phân phối tần số tần suất về thời gian tự học hàng tuần của sinh viên.**   |  |  | | --- | --- | | Dùng hàm table để tính tần số của các giá trị | > table(data$ThoiGianTuHoc)  0.5   1   2   4   5   7   8   9  10  11  12  13  14  15  16  17  18  19  20  21    1   1   1   1   3   6   5   5  10   1   8   6   5   8   3   1   1   3   1   3   22  23  24  25  26  30  32  34  35  36  38  39  40  46  47  48  60  65    2   3   1   1   1   1   1   1   2   2   1   1   1   2   1   4   1   1 | | Dùng hàm prop.table để tính tần suất của các giá trị | > prop.table(table(data$ThoiGianTuHoc))   0.5    1    2    4    5    7    8    9   10   11   12   13   14   15   16   17  0.01 0.01 0.01 0.01 0.03 0.06 0.05 0.05 0.10 0.01 0.08 0.06 0.05 0.08 0.03 0.01    18   19   20   21   22   23   24   25   26   30   32   34   35   36   38   39  0.01 0.03 0.01 0.03 0.02 0.03 0.01 0.01 0.01 0.01 0.01 0.01 0.02 0.02 0.01 0.01    40   46   47   48   60   65  0.01 0.02 0.01 0.04 0.01 0.01 |   **câu 2 Trung bình và độ lệch tiêu chuẩn về thời gian tự học hàng tuần của sinh viên**   |  |  | | --- | --- | | trung bình | > mean(data$ThoiGianTuHoc)  [1] 18.465 | | độ lệch chuẩn | > sd(data$ThoiGianTuHoc)  [1] 13.43696 | |
| **câu 3, Tìm khoảng tin cậy cho thời gian học tập trung bình của 1 sinh viên, với độ tin cậy 90%**   |  | | --- | | Ta dùng hàm t.test với các tham số cần thiết như sau:  data$ThoiGianTuHoc là vector dữ liệu về thời gian tự học của sinh viên  Độ tin cậy là 90% nên conf.level = 0.90 |   Thực hiện trên R  > t.test(data$ThoiGianTuHoc,conf.level=0.9)          One Sample t-test  data:  data$ThoiGianTuHoc  t = 13.742, df = 99, p-value < 2.2e-16  alternative hypothesis: true mean is not equal to 0  90 percent confidence interval:   16.23394 20.69606  sample estimates:  mean of x     18.465  Theo kết quả đưa ra, ta có khoảng tin cậy 90% cho Thời Gian Tự Học hàng tuần của một sinh viên là [16.23394, 20.69606]  **câu 4, Kiểm định ý kiến thời gian tự học trung bình hàng tuần của sinh viên là 8 giờ, với mức ý nghĩa 5%**  bước 1:tóm tắt bài toán   |  |  | | --- | --- | | xác định tham số | Ta quan tâm tới biến ngẫu nhiên X là thời gian truy gian tự học hàng tuần của sinh viên. Theo giả thiết  2 X N( , ) μ σ với 2 σ chưa biết. Ta kiểm định giá trị của thời gian tự học trung bình hàng tuần của sinh viên | | Phát biểu giả thuyết | H0:μ = 8 giờ                                H1:μ ≠ 8 giờ | | Mức ý nghĩa | α  = 0,05. |   bước 2: Xác định hàm kiểm định và các tham số trong R:   |  |  | | --- | --- | | x | véc tơ dữ liệu mẫu là dữ liệu về thời gian tự học hàng tuần | | y | = null (do ở đây là 1 mẫu) | | alternative | = “two.sided” hoặc “t” | | mu | = 8 | | paired | bỏ qua, mặc định là FALSE | | var.equal | bỏ qua do ở đây là 1 mẫu | | conf.level | 1-α = 0.95 |   bước 3: Thực hiện kiểm định trên R  > t.test(data$ThoiGianTuHoc,alternative = "two.sided",mu=8,conf.level=0.95)  bước 4: phân tích kết quả và kết luận  Sau khi thực hiện các lệnh ở bước 3, ta thu được kết quả sau:          One Sample t-test  data:  data$ThoiGianTuHoc  t = 7.7882, df = 99, p-value = 6.802e-12  alternative hypothesis: true mean is not equal to 8  95 percent confidence interval:   15.79882 21.13118  sample estimates:  mean of x     18.465  kết quả trên cho ta một số thông tin sau:   |  |  | | --- | --- | | +Giá trị thống kê | t = x-0s/n= 7.7882 | | +Bậc tự  do (df: degree freedom) | df =  n - 1 = 99 | | Trị số -p của bài toán là | p-value = 6.802e-12 | | + Thời gian tự học trung bình | x =  18.465 |   Kết luận: Vì p-value < α nên ta  bác bỏ giả thuyết gốc H0 . Do đó với mức ý nghĩa 5%, ta nói rằng thời gian tự học  trung bình của sinh viên không phải là 8 giờ. |
| **Câu 5**  a)  1. Nhập biến Điểm Trung bình  > x<-data$DiemTB  > x    [1] 8.00 5.75 6.00 7.00 6.00 9.65 9.19 7.00 7.00 7.10 3.42 8.40 9.30 7.30 8.45 8.30 7.45 6.45 6.10 8.56 8.90 7.13 7.60 9.10 8.13 5.30 7.32 6.40 7.90 8.10 8.30 7.50 6.78 8.30 8.64 9.13   [37] 8.60 8.15 9.30 7.34 7.30 7.45 7.83 8.30 7.90 7.93 7.88 7.20 9.30 6.10 8.15 7.30 7.80 8.13 7.33 7.82 8.20 7.61 4.60 6.20 7.80 6.52 5.60 7.73 7.00 6.50 5.80 8.90 7.80 7.40 8.50 6.52   [73] 7.40 6.35 8.45 8.32 5.50 9.20 7.40 7.90 6.80 6.70 6.30 8.50 7.60 5.30 7.30 6.50 6.50 9.20 8.63 8.91 9.00 8.20 7.50 6.30 6.70 8.25 8.91 8.50  2. Nhập biến Thời gian tự học  > y<-data$ThoiGianTuHoc  > y    [1]  1.0 35.0  0.5 21.0  7.0  7.0  7.0  7.0 14.0  4.0 22.0 13.0 10.0 16.0 15.0 20.0 19.0  5.0 12.0 40.0 25.0 14.0 21.0 26.0 48.0 10.0 32.0 12.0 12.0 15.0 19.0 12.0 15.0 13.0 22.0 48.0   [37] 36.0 38.0 18.0 16.0 12.0 36.0 34.0 15.0 13.0 19.0 17.0 15.0 47.0 10.0 48.0 23.0 48.0 14.0 23.0 24.0 21.0 23.0  2.0 16.0 30.0  7.0  8.0 12.0 14.0 10.0  7.0 60.0 46.0 10.0 15.0  8.0   [73] 10.0  9.0 13.0 12.0 10.0 10.0  8.0  9.0  9.0  5.0 10.0 12.0 11.0  8.0  9.0  8.0 13.0 65.0 14.0 15.0 35.0 15.0 10.0  9.0  5.0 46.0 39.0 13.0  3. Hệ số tương quan giữa Thời gian tự học hàng tuần và Điểm trung bình học kỳ là:  > cor(y,x)  [1] 0.4103219  4. Ước lượng các hệ số hồi quy theo cú pháp:  > lm(x ~ y)  *Xác nhận kết quả sau trên màn hình*  Call:  lm(formula = x ~ y)  Coefficients:  (Intercept)            y      6.88001      0.03513  5. Tạo object chứa các thông tin về hồi quy trong R theo lệnh  > reg <- lm (x~y)  6. Tính sai số của dữ liệu được cung cấp và hàm hồi quy thực nghiệm tại các điểm quan  sát theo lệnh  > residuals (reg)  *Xác nhận kết quả R đưa ra*             1            2            3            4            5            6            7            8            9           10           11           12           13           14   1.084861847 -2.359395399 -0.897575547 -0.617642415 -1.125889432  2.524110568  2.064110568 -0.125889432 -0.371765924  0.079486207 -4.232767628  1.063359290  2.068734929 -0.142016350            15           16           17           18           19           20           21           22           23           24           25           26           27           28   1.043108863  0.717482798 -0.097391989 -0.605639006 -1.201515497  0.274978536  1.141856732 -0.241765924 -0.017642415  1.306731519 -0.436023169 -1.931265071 -0.684019759 -0.901515497            29           30           31           32           33           34           35           36           37           38           39           40           41           42   0.598484503  0.693108863  0.752608011  0.198484503 -0.626891137  0.963359290  0.987232372  0.563976831  0.455479388 -0.064771038  1.787733224 -0.102016350 -0.001515497 -0.694520612            43           44           45           46           47           48           49           50           51           52           53           54           55           56  -0.244270185  0.893108863  0.563359290  0.382608011  0.402858437 -0.206891137  0.769102044 -1.131265071 -0.416023169 -0.387892841 -0.766023169  0.758234076 -0.357892841  0.096981945            57           58           59           60           61           62           63           64           65           66           67           68           69           70   0.582357585 -0.077892841 -2.350263366 -1.242016350 -0.133769333 -0.605889432 -1.561014645  0.428484503 -0.371765924 -0.731265071 -1.325889432 -0.087525726 -0.695772743  0.168734929            71           72           73           74           75           76           77           78           79           80           81           82           83           84   1.093108863 -0.641014645  0.168734929 -0.846139858  1.113359290  1.018484503 -1.731265071  1.968734929  0.238985355  0.703860142 -0.396139858 -0.355639006 -0.931265071  1.198484503            85           86           87           88           89           90           91           92           93           94           95           96           97           98   0.333609716 -1.861014645  0.103860142 -0.661014645 -0.836640710  0.036848208  1.258234076  1.503108863  0.890604601  0.793108863  0.268734929 -0.896139858 -0.355639006 -0.245772743            99          100   0.660103749  1.163359290  7. Đưa ra công thức khoảng tin cậy sau để thực hiện tính toán theo yêu cầu d  https://lh5.googleusercontent.com/kEriAcd-d59wPyA4qXEiyChmFaZ_wz6trtrvsM7ubaRDFVZGVvYtuBTVDP-gYJItZZTf8-DpIfBsVGj5sfTmL0WbvBmfA6msMAVuYQENtwXUIZh0ZCmpYvnXoi_PSP1ctmgOZzsWSfHZYgGikA  8. Nhập giá trị và tính theo các câu lệnh  > x0 <- 5  > beta0mu <- coef(reg)[1]  > beta1mu <- coef(reg)[2]  > y0mu <- beta0mu+beta1mu\*x0  9. Đọc giá trị của từ R theo câu lệnh  > y0mu  *Xác nhận kết quả được R đưa ra*  (Intercept)     7.055639  10. Tính giá trị S2  theo các câu lệnh  > n <- length(x)  > sbp <- sum(resid(reg)^2)/(n-2)  11. Đọc giá trị của S2 từ R theo câu lệnh  > sbp  *Xác nhận kết quả được R đưa ra*  [1] 1.11156  12. Tính x và Sxx theo các câu lệnh sau:  > xtb <- mean(x)  > Sxx <-sum(x^2)-n\*xtb^2  13. Đọc các giá trị x và Sxx và nhận kết quả từ R:  > xtb  [1] 7.5286  > Sxx  [1] 130.9862  14. Định nghĩa biến “phân vị” để ghi lại giá trị t(n - 2,/2)  > phanvi <- qt(0.975,4)  15. Đọc giá trị phân vị và xác nhận các kết quả từ R  > phanvi  [1] 2.776445  16. Tính bán kính khoảng ước lượng theo các câu lệnh  bkinh <- phanvi\*sqrt(sbp\*(1/n+(x0-xtb)^2/Sxx))  *và xác nhận kết quả từ R*  > bkinh  [1] 0.7098912  17. Tính khoảng ước lượng theo câu lệnh  > y0mu+c(-1,1)\*bkinh  *và xác nhận kết quả từ R*  [1] 6.345748 7.765530 |