



FPT EDUCATION GLOBAL

MAS291 COMPUTER PROJECT

Giảng viên: Nguyễn Việt Anh

Phân tích và dự đoán về các cơn đau tim thông qua tuổi, giới tính, huyết áp, lượng cholesterol.

Thành viên

Nghiêm Xuân Trình – HE151020

Đỗ Hoàng Hiệp – HE151113

Cao Đức Mạnh - HE150993

Mục Lục

A. Giới thiệu chung	2
B. Thu thập dữ liệu và xử lý	2
1 . Số lượng cholestoral (mg/dl) của tất cả mọi người trên thế giới có đúng là 245 hay không ?	3
2. Ít nhất 60% người có huyết áp trong khoảng từ 110 đến 130 (mm Hg) có khả năng cao bị đau tim.....	4
3. Sự khác nhau của lượng cholesterol trong máu của người có độ tuổi từ 20 đến 59 là 16 (mg/dl) ở nam giới nhiều hơn so với nữ giới với level of significance ($\alpha = 0.05$) ?	5
4. Sự khác biệt về tỉ lệ huyết áp ở người khi so sánh giữa nam và nữ trong độ tuổi từ 20 đến 59 là khoảng 2.8% với level of significance ($\alpha = 0.05$) ?	6
5 Số tuổi có ảnh hưởng gì đến số huyết áp của người đó hay không ? Nếu có thì dựa đoán xem người đang 20 tuổi và người 50 tuổi có huyết áp là bao nhiêu	8
C. Kết luận	11

A. Giới thiệu chung

Các bệnh tim mạch là một trong những căn bệnh nguy hiểm chết người nhưng lại là kẻ giết người thầm lặng khiến tỷ lệ người mắc phải tử vong mỗi năm ngày càng tăng. Từ những dữ liệu có được cộng với việc ứng dụng những kiến thức được học chúng em muốn tìm hiểu xem những yếu tố nào tác động đến những cơn đau tim .

B. Thu thập dữ liệu và xử lý

Dữ liệu được thu thập từ trang <https://www.kaggle.com/> và được lưu trữ trong file excel.

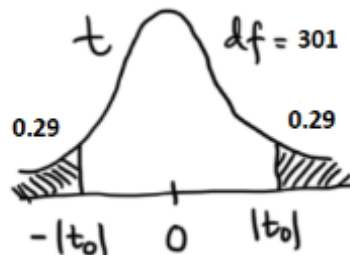
1 . Số lượng cholestoral (mg/dl) của tất cả mọi người trên thế giới có đúng là 245 hay không ?

Population: Tất cả người dân trên giới

Sample: Dữ liệu bên bảng excel

$$\alpha = 0.05$$

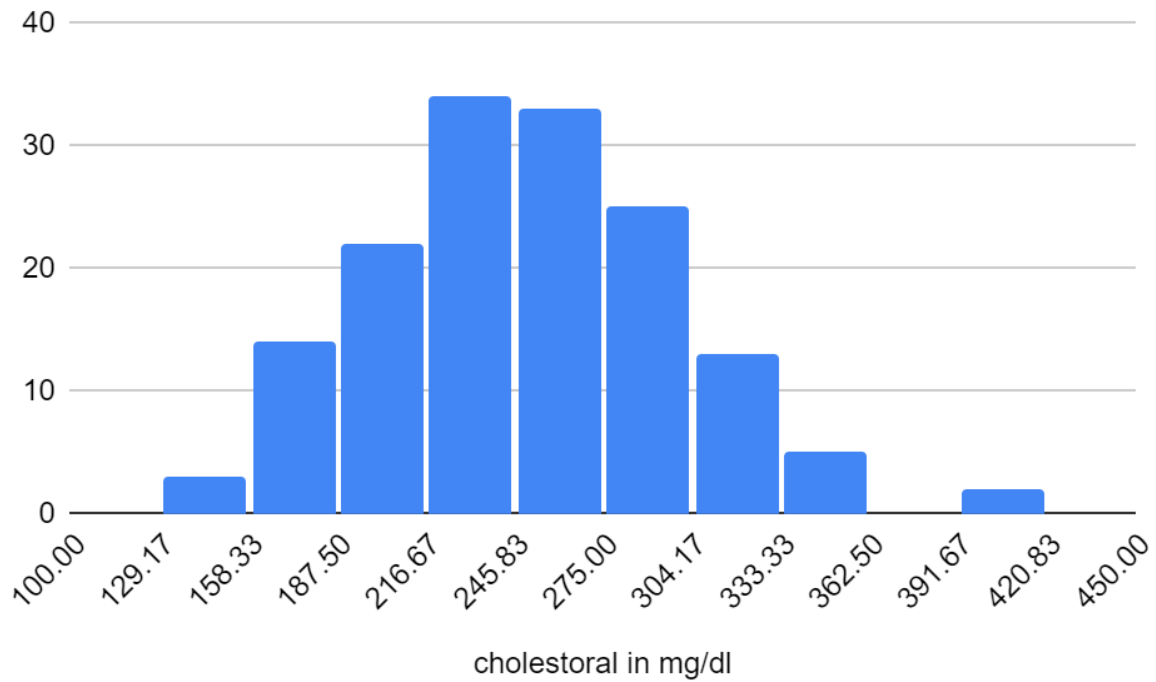
$$\begin{cases} H_0: \mu = 245 \\ H_1: \mu \neq 245 \end{cases}$$



Hypothesis Testing	
alpha	5%
x ngang	246.65
n	302
s	51.49
mu0	245
df	301
Test statistic t0	0.56
P_value	0.58
P = 0.58 > alpha = 0.05	

=> Confidence interval (252.48;240.82)

=> Đủ căn cứ để kết luận rằng số lượng cholestoral (mg/dl) của tất cả mọi người trên thế giới là 245 .



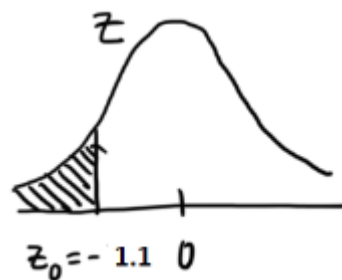
2. Ít nhất 60% người có huyết áp trong khoảng từ 110 đến 130 (mm Hg) có khả năng cao bị đau tim

Population: Tất cả người dân trên giới

Sample: Dữ liệu bên bảng excel

$$\alpha = 0.05$$

$$\begin{cases} H_0: p = 0.6 \\ H_1: p < 0.6 \end{cases}$$



Hypothesis Test	
alpha	5%
Pmu	0.56
P0	0.6
n	151
test Statistic Z0	-1.10
P_value	0.14
P-value = 0.14 > alpha 0.05	

=> Confidence interval left : 0.60

=> Đủ căn cứ để kết luận rằng Claim: Ít nhất 60% người có huyết áp trong khoảng từ 110 đến 130 (mm Hg) có khả năng cao bị đau tim

3. Sự khác nhau của lượng cholesterol trong máu của người có độ tuổi từ 20 đến 59 là 16 (mg/dl) ở nam giới nhiều hơn so với nữ giới với level of significance ($\alpha = 0.05$) ?

Population 1 : Tất cả nam giới trên thế giới có độ tuổi từ 20 đến 59

Sample 1: Tất cả nam giới trong bảng excel có độ tuổi từ 20 đến 59

Population 2: Tất cả nữ giới trên thế giới có độ tuổi từ 20 đến 59

Sample 2: Tất cả nữ giới trong bảng excel có độ tuổi từ 20 đến 59

μ_1 : Lượng cholesterol trung bình trong máu ở nam giới

μ_2 : Lượng cholesterol trung bình trong máu ở nữ giới

$$\begin{cases} H_0: \mu_1 - \mu_2 = 16 \\ H_1: \mu_1 - \mu_2 \neq 16 \end{cases}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 1}$$

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

delta 0	16
alpha	5%
n1	102
n2	54
x1 ngang	238.3627451
x2 ngang	253.6296296
s1	47.78516055
s2	53.91268143
pool variance	1516.123056
Test Statistic	-4.771468383
right t-value	1.975488058
left t-value	-1.975488058

=> Không đủ bằng chứng để kết luận điều trên.

4. Sự khác biệt về tỉ lệ huyết áp ở người khi so sánh giữa nam và nữ trong độ tuổi từ 20 đến 59 là khoảng 2.8% với level of significance ($\alpha = 0.05$) ?

Population 1: Tất cả nam giới trên thế giới có độ tuổi từ 20 đến 59

Sample 1 ;Tất cả nam giới trong bảng excel có độ tuổi từ 20 đến 59

Population 2: Tất cả nữ giới trên thế giới có độ tuổi từ 20 đến 59

Sample 2 : Tất cả nữ giới trong bảng excel có độ tuổi từ 20 đến 59

p1: Tỷ lệ nam giới trong độ tuổi 20 đến 59 có huyết áp cao hơn 120

p2: Tỷ lệ nữ giới trong độ tuổi 20 đến 59 có huyết áp cao hơn 120

$$\hat{P} = \frac{X_1 + X_2}{n_1 + n_2}$$

Null hypothesis: $H_0: p_1 = p_2$

Test statistic:
$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Delta 0	0.028
Alpha	0.05
n1	40
n2	26
x1	28
x2	17
p^1	0.7
p^2	0.6538461538
pooled p^	1.123595506
z test num	0.04615384615
z test denom	0.04323681396
z test	1.0674664
z critical value	-1.959963986
z critical value	1.959963986
p value	0.2857612783

5 Số tuổi có ảnh hưởng gì đến số huyết áp của người đó hay không ? Nếu có thì dựa đoán xem người đang 20 tuổi và người 50 tuổi có huyết áp là bao nhiêu .

Regression analysis	
Random variables X	Age
Random variables Y	resting blood pressure (in mm Hg)

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$SS_T = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$\widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{X}$$

Estimated regression line	
n	303
$\sum (X_i^2)$	920487
$(\sum X_i)^2$	271359729
$\sum (Y_i^2)$	5342310
$(\sum Y_i)^2$	1590573924
$\sum (X_i * Y_i)$	2181676
$\sum X_i * \sum Y_i$	656976186
S_{xx}	24910.33663
S_{xy}	13437.76238
SS_t	92891.10891
beta1 hat	0.5394452341
beta0 hat	102.2961012

Test the significance of regression	
alpha	5%
$(\sigma \text{ hat})^2$	284.5254885
$Se(\beta_1)$	0.1068736866
t test	5.047502817
t value (right)	1.967876531
t value (left)	-1.967876531

=>Reject H_0

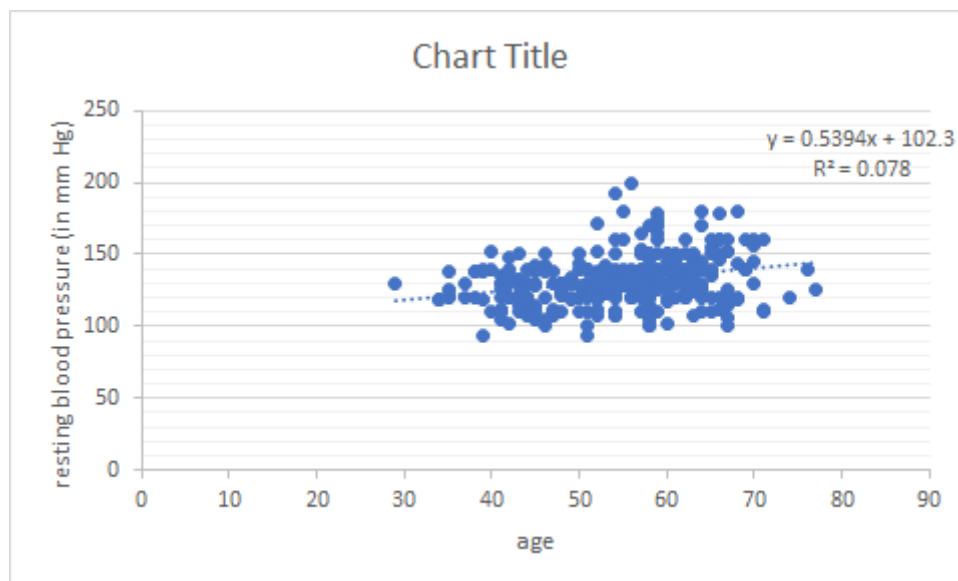
$$\hat{\sigma}^2 = \frac{SS_T - \widehat{\beta}_1 \cdot S_{xy}}{n - 2}$$

$$S_e(\beta_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

$$\text{Test statistic } t_0 = \frac{\widehat{\beta}_1 - \beta_{1,0}}{S_e(\widehat{\beta}_1)}$$

Predict a future value for Y	
assume x = 20	113.0850059
assume x = 50	129.2683629

Biểu đồ thể hiện sự phân tán giữa tuổi và huyết áp



C.Kết luận

Các yếu tố làm tăng khả năng bị đau tim là hút thuốc, lười tập thể dục, huyết áp cao, cholesterol cao, chế độ ăn uống không lành mạnh, sử dụng rượu có hại và lượng đường cao .

Số lượng cholestoral (mg/dl) của tất cả mọi người trên thế giới là 245 ,điều này cho thấy rằng tại sao có rất nhiều người bị các vấn đề về tim mạch , đặc biệt là các cơn đau tim.

Ít nhất 60% người có huyết áp trong khoảng từ 110 đến 130 (mm Hg) có khả năng cao bị đau tim .

Số tuổi cũng là một trong những yếu tố quan trọng tác động đến cơn đau tim , từ biểu đồ trên ta thấy được những cơn đau tim thường xuất hiện trong độ tuổi từ 40 đến 75 ,