Vũ Huy Quang - HE151077
Hà Duy Anh - HE151311
Nguyễn Ngọc Tuấn - HE151047

Project MAS SE1622

Chủ đề số liệu thống kê của Covid 19 tại Việt Nam từ ngày 1 tháng 9 năm 2021 đến ngày 7 tháng 11 năm 2021.

COVID-19 (bệnh vi-rút corona 2019) là một bệnh do vi-rút có tên SARS-CoV-2 gây ra và được phát hiện vào tháng 12 năm 2019 ở Vũ Hán, Trung Quốc. Căn bệnh này rất dễ lây lan và đã nhanh chóng lan ra khắp thế giới. COVID-19 thường gây ra các triệu chứng hô hấp, có thể cảm thấy giống như cảm lạnh, cúm hoặc viêm phổi

Kể từ khi dịch bùng phát từ đầu 2020 đến nay, Việt Nam đã ghi nhận 8.527.776 ca nhiễm, 4.826.024 người khỏi bệnh, 3.659.660 bệnh nhân đang điều trị và 42.148 ca tử vong.

7 ngày qua, tổng nhiễm trên cả nước giảm 479.026 (↓25%) so với cùng kỳ, tổng bệnh nhân tử vong giảm 64 (↓12%), số người khỏi bệnh tăng 362.413 (↑47%). Hôm nay nhóm em (nhóm 6) sẽ thống kê hai nước gồm Nhật Bản và Việt Nam cho thầy và các bạn cùng hiểu rõ.

Task 1:Giả thiết rằng 2.35% của tổng số ca mắc covid 19 ở Việt Nam là số ca chết. Liệu data này có support cho giả định trên ko?

Giả thiết: H0: p = 0.0235
H1: p != 0.0235

α =5%.
n=968684.
x=22531.
$\overline{p}$ =0.0233.
p_0=0.0235.

khoảng tin cậy bên phải : $\widehat{p} + \frac{z_\alpha}{2} * \frac{\sqrt{\widehat{p} * (1-\widehat{p})}}{n}$ =0.0236

khoảng tin cậy bên trái :$\widehat{p}$ + $\frac{z_\alpha}{2}$ *$\frac{\sqrt{\widehat{p}*(1-\widehat{p})}}{n}$ =0.0230

*kiểm tra thực tế

z_0 = $\frac{(\widehat{p}-p_0)}{\sqrt{\frac{p_0*(1-p_0)}{n}}}$ = -1.56

giá trị khoảng tin cậy bên phải = z_α=1.96

giá trị khoảng tin cậy bên trái = - z_α=-1.96

=> Do z0 nằm trong khoảng -1.96 đến 1.96 nên fail to reject H0

P-value= 2* normdist(-|z_0|) => do P-value lớn hơn alpha nên => fail to reject H0

Task 2: giả thiết rằng trung bình số ca mới tại việt nam nhiều hơn 5000 ca so với nhật. Liệu data này có support cho điều đó hay không?

Giả thiết:H0: μ1-μ2=5000 μ1= trung bình số ca mới tại việt nam

H1:μ1-μ2 != 5000 μ2= trung bình số ca mới tại nhật bản

$\delta_0 = 5000$ s_1= 3563.7

$\alpha = 5\%$ s_2= 4854.5

n_1 = 68 $\overline{x_1}$ = 7,449.8

n_2 = 68 $\overline{x_1}$ = 3372.2

(SP)phương sai gộp = $(n1-1)s1^2 + \frac{(n2-1)*s2^2)}{(n1+n2-2)}$=18132885.9

*kiểm tra thực tế Tính thống kê thử

$t_0 = \frac{(\overline{x_1} - \overline{x_2} - \delta_0)}{sp^{2/n_1} + sp^{2/n_2}}$ = -1.26

t value right = tinv($\frac{\alpha}{2}$, $n_1 + n_2 - 2$) =1.98

t value left = -1.98

P-value = tdist($|t_0|$, n_1+n_2-2, 2 i) > α $\rightarrow$ $Fail\ to\ reject\ H0$

CI: giá trị khoảng tin cậy bên phải

= $\overline{x_1} + \overline{x_2} + (\frac{t_\alpha}{2})^{n_1+n_2-2} * \sqrt{sp^{2/n_1} + sp^{2/n_2}}$ = 2,633.2

CI: giá trị khoảng tin cậy bên trái

=$\overline{x_1} - \overline{x_2} - (\frac{t_\alpha}{2})^{n_1+n_2-2} * \sqrt{sp^{2/n_1} + sp^{2/n_2}}$ =5,522.0

=>t_0 nằm trong khoảng, p-value lớn hơn alpha nên fail to reject H0 => dữ liệu support giả sử này

Task 3:Giả thiết rằng rằng phần trăm tổng số ca tử vong do covid 19 trên tổng số ca mắc tại việt nam nhỏ hơn hoặc bằng so với nhật bản.
Liệu rằng dữ liệu này có support giả thuyết trên ko? Bọn em đã tiến hành tính toán phía bên dưới

α=5%

n_1=968684

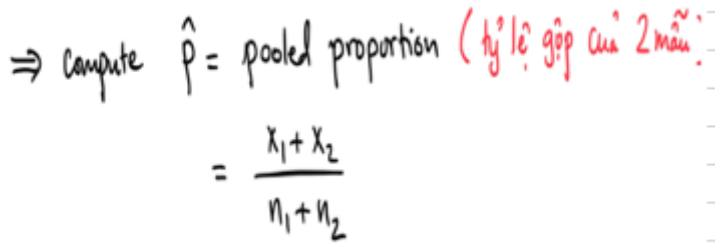
n_2=1723682

x_1=22531

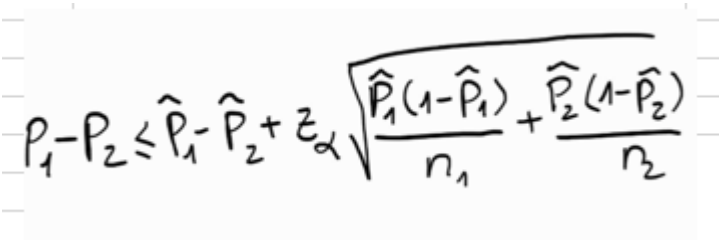
x_2=18306

$\widehat{p_1}$=2.33%(x1/n1)

$\widehat{p_2}$=1.06%(x2/n2)



p^=1.52%



Khoảng tin cậy bên phải=

z_0 = 81.4>value right→Reject H0

z=normsiv(1-α) =1.64

P=1-normdist(z_0)=0<alpha →reject H0

z_0 nằm ngoài khoảng, P value < alpha => reject H0 ⇒ dữ liệu này không hỗ trợ giả sử này.

Task 4 : Regression analysis total case in VietNam

Tại đây thì em đã lấy X hay còn gọi là biến số độc lập là ngày nhiễm

Y hay còn gọi là biến phụ thuộc là tổng các ca nhiễm tại VN

Đầu tiên nhìn vào hình này dựa theo 3 mô hình thầy từng dạy thì em có thể nhìn thấy X và Y có mối quan hệ tuyến tính với nhau và là mối quan hệ strong positive

Vậy có tuyến tính hay không thì bọn em sẽ bước vào tiến hành thực hiện tính toán

ở phần 4 thì nhóm em thực hiện việc phân tích hồi quy với dữ liệu là tổng số ca cô vít tại Việt Nam

n=68 n là cỡ mẫu

Sxx=21697

SSt=1.25E+12.

Sxy=15909813.5

Mean x=34.5

Mean y=773651.8

R=0.971069106

B1^=6714.9(Sxy/ Sxx)

B0^=541988.3

SSe=7.14E+10

y^=541988.3 +6714.9*x

Xuống dưới thì ta sẽ thực hiện test significance => để xem giữa X và Y có mối liên hệ tuyến tính hay không => có nên dùng mô hình này để dự đoán mối quan hệ giữa X và Y hay k
H0: B1 = 0
H1: B1 != 0
Significance level=5%
t value right=1.997
t value left=-1.997
(sigma hat)^2=1.08E+09
t test=3.30E+01
Giá trị t vượt ngoài khoảng cho nên là cta reject H0
Nghĩa là B1 khác 0 => có sự phụ thuộc tuyến tính giữa X và Y và có thể thực hiện tính X và Y dựa trên mô hình này.