Biến cố & Xác suất

QUAN HỆ
relations

- Kéo theo $A \subset B$
- Đối $\bar{A}$
- Hợp $A \cup B$
- giao $A \cap B$

$P(\cup A_i) = \sum P(A_i)$ (xung khắc)

KHÁI NIỆM
definition

- $P(A) = \frac{|A|}{|\Omega|}$ (cổ điển)
- $P(A) = \lim_{n \to \infty} \frac{k(A)}{n}$ (tần suất)

QUI TẮC
basic laws

- ✚ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- ✖ $\underset{\text{độc lập}}{P(AB)} = P(A) \cdot P(B)$

Bernuli : $P_k(n, p) = C_n^k p^k (1-p)^{n-k}$

XÁC SUẤT CÓ ĐIỀU KIỆN
conditional probability

$P(\underset{x}{A} \mid \underset{p}{B}) = \frac{x}{p}$

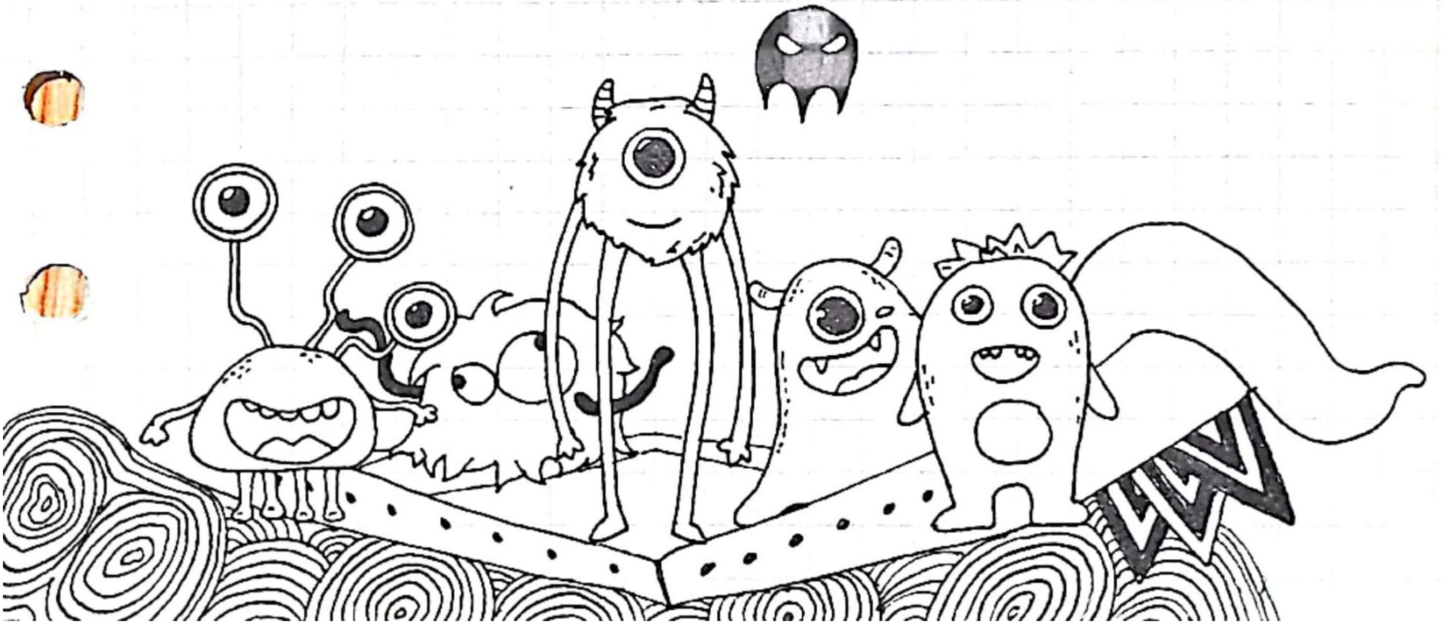
$P(AB) = P(A|B) \cdot P(B)$

$P(A|B) = \frac{P(AB)}{P(B)}$

XÁC SUẤT ĐẦY ĐỦ
total probability theorem

$P(A) = \sum_{i=1}^{N} P(A|B_i) \cdot P(B_i)$

BAYES $\underset{\text{(hậu nghiệm)}}{P(B_k|A)} = \frac{P(AB_k)}{P(A)} = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i=1}^{N} P(A|B_i) \cdot P(B_i)}$



Khảo sát 1 vùng dân cư, ta có:
15% nghiện thuốc, ung thư
25% nghiện thuốc, ko ung thư
50% ko nghiện thuốc, ko ung thư
10% ko nghiện thuốc, ung thư

Biến cố A: người đó nghiện thuốc, B: người đó ung thư.

$$P(\text{ung thư} \mid \text{nghiện thuốc}) = P(B \mid A) = \frac{P(AB)}{P(A)} = \frac{0.15}{0.15 + 0.25}$$

Gieo đồng thời 2 ⚀ cân đối. Tính XS tổng số nốt xuất hiện ≥ 10 biết rằng ít nhất 1 con đã ra 5.

$$P((A_1 + A_2) \geq 10 \mid (A_1 = 5 \| A_2 = 5)) = \frac{P(A_1 + A_2 \geq 10 \ \&\& \ (A_1 = 5 \| A_2 = 5))}{P(A_1 = 5 \| A_2 = 5)}$$

$$= \frac{3}{11}$$

Tỉ lệ nữ hoàng mang gene dễ xuất huyết là 0.5. Nếu có mang, mỗi hoàng tử có 50% nguy cơ dễ xuất huyết. Việc các hoàng tử mang bệnh là độc lập với nhau. Nếu nữ hoàng không mang gene này, hoàng tử không bị dễ xuất huyết.
a) Nữ hoàng có 3 hoàng tử không bị bệnh, tính XS nữ hoàng mang gene dễ xuất huyết.
b) Nếu có hoàng tử thứ tư, XS hoàng tử bị bệnh?

Biến cố A: nữ hoàng mang gene dễ xuất huyết: $P_A = 0.5$

B_i: hoàng tử thứ i bị bệnh: $P(B_i \mid A) = 0.5$, $P(B_i \mid \bar{A}) = 0$

a) $$P(A \mid \bar{B_1}.\bar{B_2}.\bar{B_3}) = \frac{P(A\bar{B_1}\bar{B_2}\bar{B_3})}{P(\bar{B_1}\bar{B_2}\bar{B_3})} = \frac{P(A\bar{B_1}\bar{B_2}\bar{B_3})}{P(A\bar{B_1}\bar{B_2}\bar{B_3}) + P(\bar{A}\bar{B_1}\bar{B_2}\bar{B_3})}$$

$$= \frac{(1/2).(1/2)^3}{(1/2).(1/2)^3 + (1/2).1} = \frac{1}{9}$$

b) XS hoàng tử thứ tư bị bệnh. $P(A \mid \bar{B_1}\bar{B_2}\bar{B_3}).P(B_4 \mid A) = \frac{1}{9} \cdot \frac{1}{2} = \frac{1}{18}$

DLNN Rời Rạc

Hữu hạn
Vô hạn đếm được

PHÂN BỐ XÁC SUẤT
probability mass function

X	x_1	x_2	∘∘∘	x_n
P	p_1	p_2	∘∘∘	p_n

Kỳ vọng : $\underset{\mu}{E(X)} = \sum x_i p_i$

PHÂN BỐ NHỊ THỨC
binomial distribution

$P(X = k) = C_n^k p^k (1-p)^{n-k}$
$X \sim B(n, p)$
Kỳ vọng : $E_X = np$
Phương sai: $D_X = np(1-p)$

PHÂN BỐ TÍCH LŨY
cumulative distribution function

$F(X) = P(X < x)$

Phương sai : $D(X) = E(X^2) - \mu^2$
$\sigma_x = \sqrt{D(X)}$

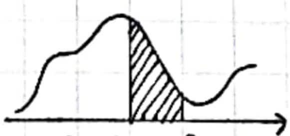
PHÂN BỐ POISSON
poisson distribution

$P(X = k) = \dfrac{e^{-\mu} \mu^k}{k!}$

PHÂN BỐ ĐỒNG THỜI
joint probability distribution

$P(X = x_i, Y = y_i) = P(X = x_i) . P(Y = y_i)$
độc lập

DLNN Liên Tục



Một khoảng
Cả trục số

PHÂN BỐ XÁC SUẤT
probability density function

$f(x) \geq 0 \quad \forall X$

$\int_{-\infty}^{\infty} f(x)\,dx = 1$

$P(a < x < b) = \int_a^b f(x)\,dx$

Kỳ vọng: $\boxed{E(X) = \int_{-\infty}^{\infty} x f(x)\,dx}$

PHÂN BỐ TÍCH LŨY
cumulative distribution functio

$0 \leq F(X) = P(X \leq x) \leq$

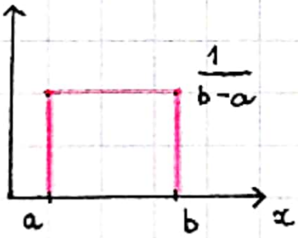
$P(a \leq X \leq b) = F(b) - F(a$

$\rightsquigarrow F'(X) = f(x)$

Phương sai

$D(X)$
$Var(X) = \int_{-\infty} x^2 f(x)\,dx - \mu$

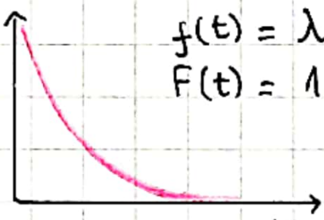
PHÂN PHỐI ĐỀU
continuous uniform distribution



$E(X) = \frac{a+b}{2}$

$D(X) = \frac{(a-b)^2}{12}$

PHÂN BỐ MŨ
exponential distribution



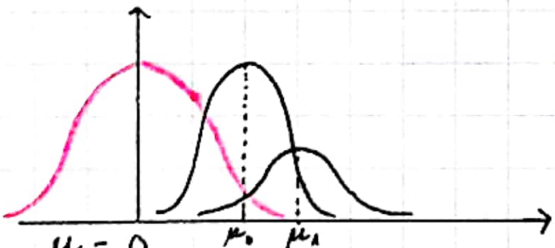
$f(t) = \lambda e^{-\lambda t} \quad (t > 0)$

$F(t) = 1 - e^{-\lambda t}$

$X \sim Poison(\lambda = ?)$

Số event / time

PHÂN BỐ CHUẨN
normal / gaussian distribution



$\mu = 0$
$\sigma = 1$

$X \sim N(\mu, \sigma^2)$

$Z = \frac{X - \mu}{\sigma}$

fx-580VN

MENU

Giả sử X nằm trong đoạn $[0, 3]$ và hàm mật độ XS $f(x) = cx^3$.

a) Tính c   c) $D(X)$, σ_x

b) $E(X)$ d) Median m_x : $P(X < m) = P(X > m)$; $F(m) = \frac{1}{2}$

a) $\int_{-\infty}^{\infty} f(x)\,dx = \int_0^3 cx^3\,dx = \frac{cx^4}{4}\Big|_0^3 = \frac{3^4 c}{4} = 1$

$\Leftrightarrow \quad c = \frac{4}{81}$

b) $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^3 \frac{4x^4}{81}\,dx = \frac{4}{81} \cdot \frac{x^5}{5}\Big|_0^3 = \frac{12}{5}$

c) $D(X) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2 = \int_0^3 \frac{4}{81} x^{6-1}\,dx - E(X)^2$

Var(X)

$= \frac{4}{81} \frac{x^6}{6}\Big|_0^3 - \left(\frac{12}{5}\right)^2 = 0.24$

d) $F(m) = \int_{-\infty}^{m} \frac{4}{81} x^3\,dx = \frac{x^4}{81}\Big|_0^m = \frac{m^4}{81} = \frac{1}{2} \Leftrightarrow m = \frac{3}{\sqrt[4]{2}}$

Luật Số Lớn & ĐL GH Trung Tâm

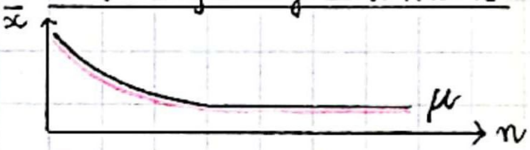
ĐL GIỚI HẠN TRUNG TÂM
central limit theorem

$$S = \{X_1, X_2, \ldots, X_n\}_{ĐLNN, \mu, \sigma^2}$$

$$S_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \quad \text{pb chuẩn}$$

$$E(S_n) = \mu$$
$$D(S_n) = \frac{\sigma^2}{n}$$

LUẬT SỐ LỚN
law of large numbers



$$T_n = X_1 + X_2 + \ldots + X_n = nS_n$$

$$E(T_n) = n\mu$$
$$D(T_n) = n\sigma^2$$

Khối lượng người VN có pb chuẩn, trung bình 65 kg với độ lệch chuẩn là 5 kg. 1 thang máy cho phép đi tối đa 10 người, và có khối lượng không quá 700 kg.
Tính XS 10 ng bất kỳ đi vào thang máy không bị quá tải.

Gọi X_i là khối lượng của người thứ i với $i = 1, \dots, 10$.
Các X_i là dãy biến ngẫu nhiên độc lập có cùng phân bố:

$$E(X_i) = 65 \qquad D(X_i) = 5^2 = 25$$

Theo ĐL giới hạn trung tâm, trung bình cân nặng 10 ng:

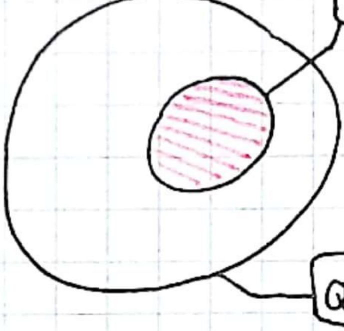
$$S_{10} = \frac{X_1 + X_2 + \dots + X_n}{10} \sim N\left(65, \frac{25}{10}\right)$$

$$P(T_n \leqslant 700) = P(S_{10} \leqslant 70) = P\left(\frac{S_{10} - 65}{\sqrt{\frac{25}{10}}} \leqslant \frac{70 - 65}{\sqrt{\frac{25}{10}}}\right)$$

$$= P(Z \leqslant 3,162)$$

$$= \phi(3,162) \approx 0,9992$$

Thống kê



MẪU sample

Kỳ vọng BNN trung bình mẫu : $\mu_{\bar{x}}$

Phương sai BNN trung bình mẫu : $\sigma_{\bar{x}}^2 = \dfrac{S^2}{n}$

Trung bình mẫu : $\bar{x}$

Phương sai mẫu : $S^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

QUẦN THỂ population

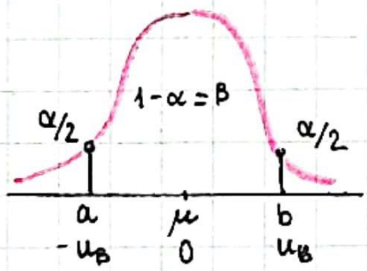
Kỳ vọng : μ

Phương sai : σ^2

Ước lượng kì vọng : $\bar{x}$

Ước lượng phương sai : $S^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

KHOẢNG TIN CẬY CHO KỲ VỌNG
confidence interval for μ



$[\bar{x} - u_\beta \sigma_{\bar{x}},\ \bar{x} + u_\beta \sigma_{\bar{x}}]$

PHÂN PHỐI STUDENT
student's t-distribution

$n < 30$: $\bar{X}$ có phân bố student với bậc $(n-1)$

KÍCH THƯỚC MẪU

$n \geq \left(\dfrac{\sigma u_\beta}{\varepsilon}\right)^2$

KHOẢNG TIN CẬY CHO TỈ LỆ
confidence interval for proportion

$[f - u_\beta \sigma_F,\ f + u_\beta \sigma_F]$

p : tỉ lệ cá thể có A trong quần thể

$f = k/n$: tần suất A trong mẫu

$F \sim N\left(p, \dfrac{p(1-p)}{n}\right)$ Xấp xỉ $D(F) = \dfrac{f(1-f)}{n}$

ĐK: $\begin{cases} np > 5 \\ n(1-p) > 5 \end{cases}$ ĐK: $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$

Biết phương sai tổng thể σ^2	$\sigma_{\bar{x}}^2 = \dfrac{\sigma^2}{n}$; $[\bar{x} \pm u_\beta \sigma_{\bar{x}}]$ Bảng Z
Không biết σ^2 $n \geq 30$	$\sigma_{\bar{x}}^2 = \dfrac{S^2}{n}$; $[\bar{x} \pm u_\beta S/\sqrt{n}]$ Bảng z
Không biết σ^2 $n < 30$	S ; $[\bar{x} \pm u_\beta S/\sqrt{n}]$ Bảng Student Bậc tự do $n-1$

Chiều cao TB của 50 SV ĐHCN là 160 cm. Độ lệch chuẩn chiều cao người lớn là 5cm. Tính khoảng tin cậy chiều cao TBSVĐHCN với:
a) Độ tin cậy 90%, 95%   b) Giả vẫn $n = 30$

Theo đề bài: $\bar{x} = 160$ cm

$\sigma = 5$ cm

$n = 50$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{50}}$$

a) Độ tin cậy 90%: $u_\beta = z_{0.95} = 1.645$

Độ tin cậy 95%: $u_\beta = z_{0.975} = 1.96$

$$[\bar{x} - u_\beta \sigma_{\bar{x}}, \bar{x} + u_\beta \sigma_{\bar{x}}]$$

b) Thay $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{30}}$

Một p² điều trị mới tính hiệu quả dựa trên số ngày điều trị đến khi khỏi. 11 bệnh nhân ngẫu nhiên điều trị trong:
4, 4, 3, 8, 5, 6, 7, 12, 5, 3, 8 (ngày)
$\bar{x} = (x_1 + .. + x_n)/n$ $\sigma^2 \cong s^2 = \sum (x_i - \bar{x})^2/(n-1)$
Tính khoảng tin cậy cho μ với độ tin cậy $\beta = 99\%$

(STUDENT)

Ta có: $\bar{x} =$ $n = 11$

$\sigma^2 =$ $\sigma^2_{\bar{x}} =$

$\bar{x}$ có phân bố student bậc 10: $t_{.005} = 3.169$

Khoảng tin cậy: $[\bar{x} \mp 3.169 . \sigma_{\bar{x}}]$

Một mẫu ngẫu nhiên 100 người dùng xe máy có 30 người dùng xe Honda. Tìm khoảng tin cậy cho tỉ lệ người dùng xe Honda với độ tin cậy 96%.

ĐK: $nf = 100.\frac{30}{100} > 10$; $n(1-f) = 70 > 10$

$$DF = \frac{f(1-f)}{n} = \frac{0,3 . 0.7}{100} \Rightarrow \sigma_F = \frac{0.1\sqrt{3.7}}{10} = \frac{\sqrt{21}}{100}$$

$n \geq 30$, ta có $u_\beta = z_{0.98} = 2.05$

Khoảng tin cậy $[f \mp u_\beta \sigma_F]$

Công ty coca cải tiến công thức: CT cũ 130/500 người dùng thích, CT mới 300/1000. KĐ: CT mới có tăng ng thích. 1%

$$H_0: p_1 = p_2,\ H_A: p_1 < p_2,\ z = \frac{(13/50 - 3/10)}{\sqrt{\frac{130+300}{1500}.(1-\hat{p}).\left(\frac{1}{500} + \frac{1}{1000}\right)}} = -1,615 > z(0.01)$$

Kiểm định Giả thuyết

ĐỘNG CƠ

H_0 null hypothesis	H_A alternative hypothesis

- ✓ TRUE — ✗ rej — TYPE I ERROR
- ✗ WRONG — ✓ app — TYPE II ERROR

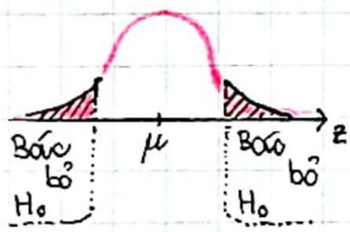
P-value: α_{min} : bỏ H_0

1. H_0, H_A
2. $\alpha (0.05)$
3. Chọn kiểm định
4. GT kiểm định
5. QĐ H_0
6. KL

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

KĐGT về giá trị trung bình $\boxed{\bar{x}}$

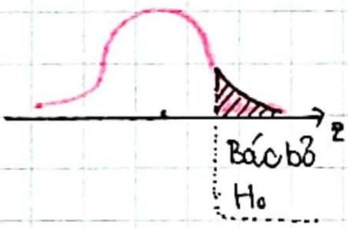
có thay đổi



$H_A : \mu \neq \mu_0$

P-val: $2P(\bar{X} \geq \bar{x})$ $(\bar{x} > \mu)$
$2P(\bar{X} \leq \bar{x})$ $(\bar{x} < \mu)$

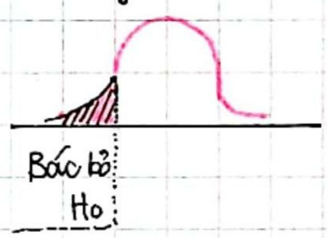
thay đổi lớn hơn



$H_A : \mu > \mu_0$

P-val: $P(\bar{X} \geq \bar{x})$

thay đổi nhỏ hơn



$H_A : \mu < \mu_0$

P-val: $P(\bar{X} \leq \bar{x})$

KĐGT về $\boxed{P}$ giá trị xác suất

$$DF = \frac{p_0(1-p_0)}{n} \quad ; \quad z = \frac{f - p_0}{\sqrt{DF}}$$

$(np_0 > 5, \ nq_0 > 5)$

nhiều xác suất

$$T = \sum_{1}^{k} \frac{(n_i - \hat{n}_i)^2}{\underset{>5}{\hat{n}_i}} \sim \chi^2 (k-1)$$

HAI MẪU 2-sample

$\boxed{\bar{x}}$

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2}$$

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

$\boxed{P}$

$$\mu = \hat{p}_1 - \hat{p}_2$$

$$\sigma^2 = \hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

$\hat{p} = (k_1 + k_2)/(n_1 + n_2)$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_{H_0}}{\sigma_{(\hat{p}_1 - \hat{p}_2)}}$$

Sau khi thay giám đốc, nhà máy SX thép ghi nhận sản lượng 100 ngày, có trung bình và độ lệch chuẩn của mẫu là 980t và 50t. KĐGT: sản lượng bình quân hàng ngày tăng so với TB 892t ghi nhận 1 năm trước với $\alpha = 0.05$.

Gọi μ là sản lượng bình quân hàng ngày sau khi thay giám đ[ốc]
Theo đề bài: $n = 100$, $\mu_0 = 892$, $\bar{x} = 980$, $s = 50$,

1	$H_0: \mu = 892$ $\qquad$ $H_A: \mu > 892$
2	$\alpha = 0.05$
3	Do $n \geq 30$ và kiểm định thay đổi lớn hơn nên ta chọn kđt bên phải đồ thị z
4	$P\text{-value} = P(\bar{X} \geq 980) = P\left(Z \geq \dfrac{980-892}{\frac{50}{\sqrt{100}}}\right) = 1 - \phi(17.6) \approx 0$
5	Do $\alpha >$ P-value nên ta bác bỏ H_0.
6	Ta có đủ cơ sở KL rằng...

Một đảng chính trị tuyên bố 45% cử tri sẽ bỏ phiếu cho ứng cử viên A của họ. Chọn ngẫu nhiên 200 cử tri, 80 trong số đó bỏ phiếu cho A. Kiểm thử dự đoán trên với $\alpha = 0.02$.

Gọi F là tỉ lệ người bỏ phiếu cho A.

Do $\begin{cases} n \cdot p_0 = 90 > 5 \\ n(1-p_0) = 110 > 5 \end{cases}$ nên F có pb xấp xỉ $F' \sim N(np, npq)$

1	$H_0: f = 0.45$ $\qquad$ $H_A: f \neq 0.45$
2	$\alpha = 0.02$
3	
4	$P\text{-value} = 2P\left(F > \frac{80}{200}\right) = 2P(Z \leq -1,4434)$ $= 2(0 + \varphi(-1,4434)) \approx$
5	
6	

PT Tương quan & Hồi quy

HỆ SỐ TƯƠNG QUAN
correlation coefficient

$$\rho = \frac{E(X-\mu_X)(Y-\mu_Y)}{\sigma_X \sigma_Y}$$

$$r = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_1^n (x_i - \bar{x})^2 \sum_1^n (y_i - \bar{y})^2}} = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2} \cdot \sqrt{n\sum y^2 - (\sum y)^2}}$$

KIỂM ĐỊNH
testing

$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$	$\rho \neq 0$
$T = \frac{u-m}{\sigma}$ $u = \frac{1}{2}\ln\frac{1+r}{1-r}$; $\sigma = \frac{1}{\sqrt{n-3}}$ $m = \frac{1}{2}\ln\frac{1+\rho_0}{1-\rho_0}$	$\rho \neq \rho_0$

Thu thập mẫu 8 quốc gia để nghiên cứu mối liên hệ giữa số ca mắc covid-19 (Y: số ca / 10M dân) và mật độ dân số (X người / km²) Tính hệ số tương quan giữa X và Y.

X	153	464	36	151	287	25	226	9
Y	1	51	2656	8	16	155	39	1879

$n = 8$

Từ bảng số liệu, ta có: $r = \dfrac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}} = -0,575$

MENU $\boxed{6}$ → 2 : $y = ax + b$ $\boxed{AC}$

OPTN $\boxed{3}$ Tính hồi quy r, a, b

fx-580VN