HƯỚNG DẪN THỰC HÀNH MÔN XÁC SUẤT THỐNG KÊ

Tài liệu này hướng dẫn cho sinh viên Trường Đại học Xây dựng sử dụng phần mềm Mathematica hỗ trợ việc học tập và thi môn Xác suất thống kê.

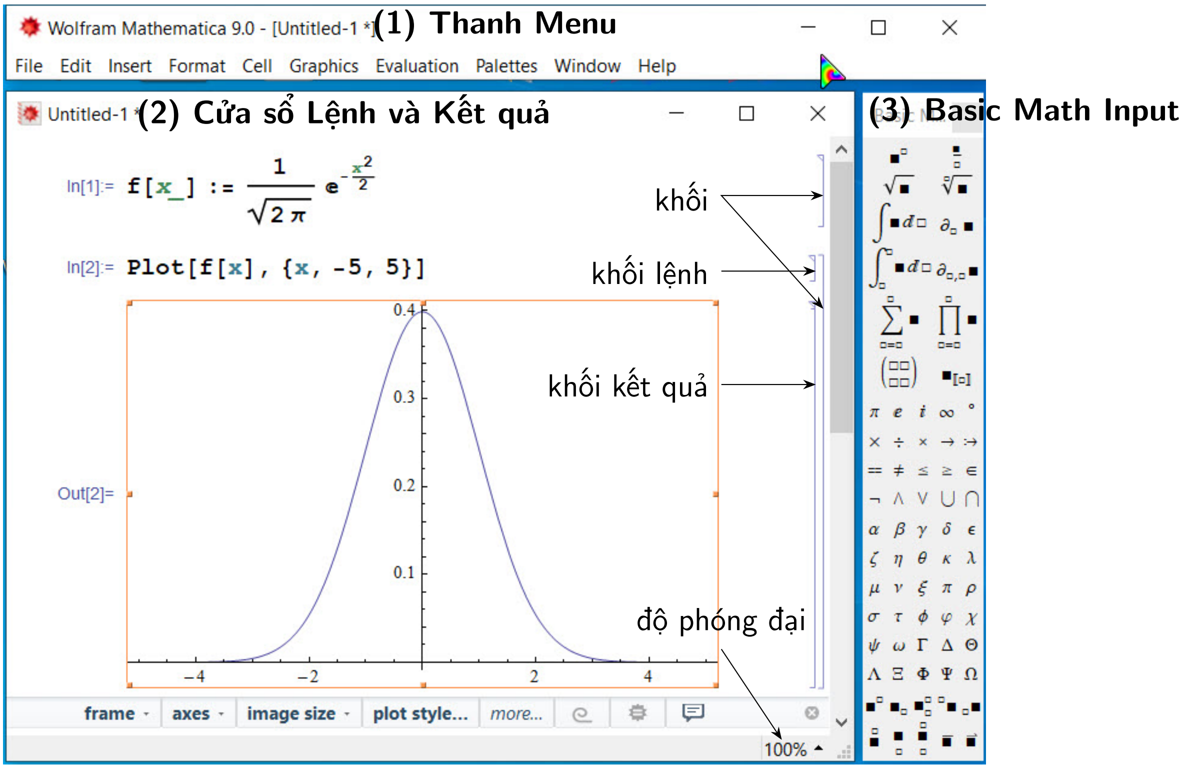
Yêu cầu: tự học 10 tiết, tổng thời gian thực hành làm mẫu trên lớp 3 tiết, thực hành trên phòng máy 3 tiết.

1 Mathematica

Link tải phần mềm, tài liệu, video hướng dẫn thực hành,... tại:

https://github.com/ThinhND-HUCE/xstk2

Giao diện:



Cách sử dụng Mathematica:

* *Enter*: xuống dòng để nhập lệnh tiếp theo trong cùng khối lệnh.

* *Shift* + *Enter*: chạy cả khối lệnh chứa con trỏ hiện thời, xem kết quả, soát và sửa lỗi.

* Lập biểu thức toán học bằng các phép toán $+$, $-$, $*$, $/$ (hoặc *Ctrl* + $/$), lũy thừa ^ (hoặc *Ctrl* + *6*), nhóm biểu thức trong dấu ( ) và các hàm

Hàm lý thuyết	$\sin x$	$\cos x$	$\tan x$	$\cot x$	$\ln x$
Nhập trong Math	`Sin[x]`	`Cos[x]`	`Tan[x]`	`Cot[x]`	`Log[x]`
	$\arcsin x$	$\arccos x$	$\arctan x$	$\mathrm{arccot} x$	$\log_a x$
	`ArcSin[x]`	`ArcCos[x]`	`AcrTan[x]`	`ArcCot[x]`	`Log[a,x]`

với x là biến hoặc biểu thức.

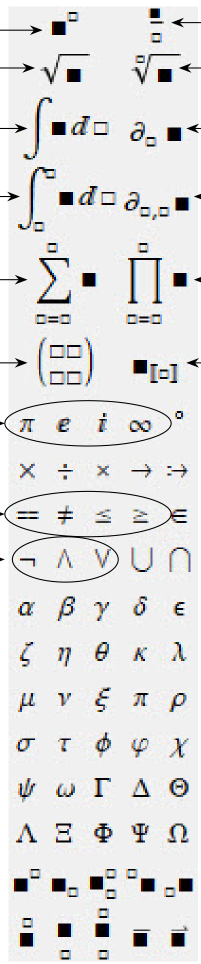
* Mathematica phân biệt **CHỮ HOA** và **chữ thường**. Các lệnh (hàm) có sẵn của Mathematica có chữ cái đầu mỗi từ viết hoa (lệnh có thể là từ ghép), các tham số của lệnh được đặt trong dấu `[ ]` và ngăn cách bởi dấu `,`

`TênLệnh[ tham số 1 , tham số 2 , ... ]`

* Vào `[Menu]Help/Help Browser` để tìm hiểu các lệnh.

* Thanh công cụ **Basic Math Input** hỗ trợ nhập nhanh các kí hiệu và biểu thức toán học theo mẫu, gọi ra bằng thao tác $\boxed{\texttt{[Menu]Palettes} \to \texttt{Other}}$

lũy thừa — phân số
căn bậc 2 — căn bậc n
nguyên hàm — đạo hàm riêng
tích phân xác định — đạo hàm riêng theo nhiều biến
tổng của dãy — tích của dãy
ma trận cấp 2 — phần tử của dãy, ma trận...
π, e, i, ∞
phép toán so sánh
phép toán logic



`x^y`
`Sqrt[x]`
`Integrate[f[x], x]`
`Integrate[f[x], {x, a, b}]`
`Sum[a_i, {i, m, n}]`
`Pi, E, I, Infinity`
`==, !=, <=, >=`
`!, &&, ||`

`x/y`
`x^(1/n)`
`D[expr, x]`
`D[expr, x, y]`
`Product[a_i, {i, m, n}]`
`X[[i]], A[[i, j]]`

Các lệnh trong môn học:

	Lệnh	**Kết quả**
1	`Binomial[n, k]`	C_n^k
2	`var = expr`	gán biểu thức `expr` cho biến `var`
3	`f[x_] := expr` `f[x_, y_] := expr`	khai báo hàm số 1 hay nhiều biến
4	`f[x]` `f'''[x]`	$f(x)$ $f'''(x)$

	`Integrate[f[x], x]`	$\int f(x)\,dx$
	`Integrate[f[x], {x, a, b}]`	$\int_a^b f(x)\,dx$
	`Integrate[f[x, y], {x, a, b}], {y, c, d}]`	$\int_a^b dx \int_c^d f(x,y)\,dy$
$\boxed{5}$	`Plot[f[x], {x, a, b}]`	đồ thị hàm số $y = f(x)$ trên (a, b)
6	`expr // N`	đưa các số trong biểu thức về số thập phân
7	`expr /. x -> a` `expr /. {x -> a, y -> b}`	giá trị của biểu thức khi thay x bởi $a, \ldots$
$\boxed{8}$	`Simplify[expr]`	rút gọn biểu thức
9	`If[điều kiện, kq1, kq2]`	$\begin{cases} \text{kq1} & \text{nếu điều kiện đúng} \\ \text{kq2} & \text{ngược lại} \end{cases}$
10	Thao tác *[RClick]* $\rightarrow$ *Insert Table/Matrix...*	mẫu điền ma trận

11	`Table[expr, {n}]`
	`Table[expr, {i, a, b}]`
	`Table[expr, {i, a, b}, {j, c, d}]`
		tạo véctơ gồm n phần tử `expr`
		tạo véctơ, ma trận theo công thức

12	`Join[X, Y]`	nối hai véctơ X và Y ra véctơ mới
13	`Total[X]` `Total[A, 2]`	tổng các phần tử của véctơ X tổng các phần tử của ma trận A
14	`Transpose[A]`	A^T
15	`A[[i, All]]` `A[[All, j]]`	véctơ gồm các phần tử hàng i của A véctơ gồm các phần tử cột j của A

2 Xác suất

VD1:

Kết quả	**Lệnh** *(bằng mẫu hoặc chữ)*
$C_{10}^4 = 210$	`Binomial[10, 4]`
$\sum_{k=0}^{10} C_{800}^k 0.005^k \cdot 0.995^{800-k} = 0.9972$	$\sum_{\text{k}=0}^{10} \text{Binomial}[800,\text{k}] * 0.005^{\text{k}} * 0.995^{800-\text{k}}$ `Sum[Binomial[800, k] * 0.005^k * 0.995^(800 - k), {k, 0, 10}]`
$\sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda$	$\sum_{\text{k}=0}^{\infty} \text{k} \frac{\lambda^{\text{k}}\,\text{E}^{-\lambda}}{\text{k}!}$ `Sum[k * l^k * E^-l / k!, {k, 0, Infinity}]`

VD2: Đại lượng ngẫu nhiên X có hàm mật độ $\boxed{f(x) = ae^{4x-x^2}}$

a) Tìm a. b) Tìm EX, DX. c) Tìm $P(1 < X < 3.5)$.

HD	**Lệnh**
	$\mathtt{f[x\_] := a * E^{4x-x^2}}$ `f[x_] := a*E^(4 x - x^2)`
a) $\boxed{\int_{-\infty}^{\infty} f(x)\,dx} = 1 \Rightarrow ae^4\sqrt{\pi} = 1$	$\int_{-\infty}^{\infty} \mathtt{f[x]}\,dx$ `Integrate[f[x], {x, -Infinity, Infinity}]`
$\Rightarrow \boxed{a = \dfrac{1}{e^4\sqrt{\pi}}} = \boxed{0.01033}$	$\mathtt{a} = \dfrac{1}{\mathtt{E}^4\sqrt{\pi}}$ hoặc `a = 1 / E^4 / Sqrt[Pi]` `a // N`
b) $\boxed{EX = \int_{-\infty}^{\infty} xf(x)\,dx} = 2$	$\mathtt{ex} = \int_{-\infty}^{\infty} \mathtt{x * f[x]}\,dx$ `ex = Integrate[x * f[x], {x, -Infinity, Infinity}]`
$\boxed{E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx} = \dfrac{9}{2}$	$\mathtt{ex2} = \int_{-\infty}^{\infty} \mathtt{x^2 * f[x]}\,dx$ `ex2 = Integrate[x^2 * f[x], {x, -Infinity, Infinity}]`
$DX = \boxed{E(X^2) - (EX)^2} = \dfrac{1}{2}$	$\mathtt{ex2} - \mathtt{ex}^2$ hoặc `ex2 - ex^2`
c) $P(1 < X < 3) = \boxed{\int_1^3 f(x)\,dx} =$ 0.8427	$\int_{1.0}^{3} \mathtt{f[x]}\,dx$ hoặc `Integrate[f[x], {x, 1.0, 3}]`

VD3: Đại lượng ngẫu nhiên X có hàm mật độ $\boxed{f(x) = \begin{cases} kx, & x \in [0,2] \\ 0, & x \notin [0,2] \end{cases}}$

a) Xác định k. b) Tìm hàm phân bố $F(x)$. c) Tính $P(0 < X < 1)$.

HD	**Lệnh**
	`f[x_] := If[0 <= x <= 2, k*x, 0]`
a) $\boxed{\int_{-\infty}^{\infty} f(x)\,dx} = 1 \Rightarrow 2k = 1$	$\int_{-\infty}^{\infty} \mathtt{f[x]}\,dx$ `Integrate[f[x], {x, -Infinity, Infinity}]`
$\Rightarrow \boxed{k = \dfrac{1}{2}}$	$\mathtt{k} = \dfrac{1}{2}$ hoặc `k = 1/2`
b) $F(x) = \boxed{\int_{-\infty}^{x} f(t)\,dt} = \begin{cases} 0 & \text{nếu } \boxed{x \le 0} \\ \dfrac{x^2}{4} & \text{nếu } 0 < x \le 2 \\ 1 & \text{nếu } x > 2. \end{cases}$	`g[t_] := If[t < x, f[t], 0]` $\int_{-\infty}^{\infty} \mathtt{g[t]}\,dt$

(trường hợp $x \le 0$ trong Mathematica 9 hiển thị là `True`)	`Integrate[g[t], {t, -Infinity, Infinity}]`
c) $P(0 < X < 1) = \boxed{\int_0^1 f(x)\,dx} = \dfrac{1}{4}$	$\int_0^1$ f[x]dx `Integrate[f[x], {x, 0, 1}]`

Chú ý: Mathematica có thể tính nhanh $I = \int_D f(x)\,dx$, khi $f(x)$ xác định phân nhánh, $D \subset \mathbb{R}^n$ là miền phức tạp hoặc phụ thuộc tham số, qua hai bước:

$$g(x) = \begin{cases} f(x), & \text{nếu } x \in D \\ 0 & \text{nếu } x \notin D \end{cases} \Rightarrow I = \int_{\mathbb{R}^n} g(x)\,dx.$$

VD4: Véctơ ngẫu nhiên (X, Y) có hàm mật độ xác suất đồng thời $\boxed{f(x,y) = \begin{cases} A(2x^2 + xy + y^2) & \text{nếu } (x,y) \in [0,1] \times [0,1] \\ 0 & \text{nếu } (x,y) \notin [0,1] \times [0,1] \end{cases}}$. Tìm

a) Tìm A. b) Tính EX. c) Tính $P(X < 0.5 \mid Y > 0.5)$.

HD	**Lệnh**
	f[x_, y_] := If[0 ≤ x ≤ 1&&0 ≤ y ≤ 1, A(2x² + x ∗ y + y²), 0] `f[x_, y_] := If[0 <= x <= 1 && 0 <= y <= 1, A (2 x^2 + x*y + y^2), 0]`
a) $\boxed{\iint_{\mathbb{R}^2} f(x,y)\,dxdy} = 1 \Rightarrow \dfrac{5A}{4} = 1$	$\int_{-\infty}^{\infty}(\int_{-\infty}^{\infty}$ f[x, y]dy)dx `Integrate[f[x, y], {x, -Infinity, Infinity}, {y, -Infinity, Infinity}]`
$\Rightarrow \boxed{A = \dfrac{4}{5}}$	A $= \dfrac{4}{5}$ hoặc `A = 4/5`
b) $EX = \boxed{\iint_{\mathbb{R}^2} xf(x,y)\,dxdy} = \dfrac{2}{3}$	$\int_{-\infty}^{\infty}(\int_{-\infty}^{\infty}$ x ∗ f[x, y]dy)dx `Integrate[x * f[x, y], {x, -Infinity, Infinity}, {y, -Infinity, Infinity}]`
c) $p = P(X < 0.5 \mid Y > 0.5) = \dfrac{P(X < 0.5, Y > 0.5)}{P(Y > 0.5)} = \dfrac{t}{m}$. $\boxed{t = \iint_{x<0.5, y>0.5} f(x,y)\,dxdy} = 0.1875$	t $= \int_{-\infty}^{0.5}\left(\int_{0.5}^{\infty}\right.$ f[x, y]dy$\left.\right)$ dx `t = Integrate[f[x, y], {x, -Infinity, 0.5}, {y, 0.5, Infinity}]`

$\boxed{m = \iint\limits_{y>0.5} f(x,y)\,dxdy} = 0.65$	$\mathtt{m} = \int_{-\infty}^{\infty}\left(\int_{0.5}^{\infty} \mathtt{f}[\mathtt{x},\mathtt{y}]\,d\mathtt{y}\right) d\mathtt{x}$ `m = Integrate[f[x, y], {x, -Infinity, Infinity}, {y, 0.5, Infinity}]`
$\Rightarrow p = 0.2885$	$\frac{\mathtt{t}}{\mathtt{m}}$ hoặc `t/m`

3 Thống kê

	Giá trị	Tham số	Lệnh	Ví dụ & chú thích
16	$\Phi(x)$	$x \in \mathbb{R}$	`CDF[NormalDistribution[0, 1], x]`	$\Phi(2) = 0.9772$
17	z_0 s/c: $\Phi(z_0) = p$	$0 < p < 1$	`Quantile[NormalDistribution[0, 1], p]`	$\Phi(z_0) = 0.95$ $\Rightarrow z_0 = 1.6449$
18	t_p^k	$k \in \mathbb{N}^*$	`-Quantile[StudentTDistribution[k], p/2]`	$t_{0.1}^{29} = 1.6991$
19	$\chi^2(p,k)$		`Quantile[ChiSquareDistribution[k], 1-p]`	$\chi^2(0.05, 25) = 37.6525$
20	n	$x_1, x_2, \ldots, x_n$	`Length[X]`	$\mathtt{X} = \{\mathtt{x_1}, \mathtt{x_2}, \ldots, \mathtt{x_n}\}$
21	$\overline{x}$		`Mean[X]`	
22	s'^2		`Variance[X]`	
23	s'		`StandardDeviation[X]` hoặc `Sqrt[`s'^2`]`, $\sqrt{\mathtt{s}'^2}$	
	s^2		$\frac{n-1}{n}s'^2$	
	s		$\sqrt{\frac{n-1}{n}}s'$	
24	r	$x_1, x_2, \ldots, x_n$ $y_1, y_2, \ldots, y_n$	`Correlation[X, Y]`	$\mathtt{X} = \{\mathtt{x_1}, \mathtt{x_2}, \ldots, \mathtt{x_n}\}$ $\mathtt{Y} = \{\mathtt{y_1}, \mathtt{y_2}, \ldots, \mathtt{y_n}\}$

25	a		`data = Transpose[{X, Y}]`
			 `model = LinearModelFit[data, x, x]`
			 `model["BestFitParameters"][[2]]`	

	b		`model["BestFitParameters"][[1]]`	
26	$ax+b$		`model["BestFit"]`	

VD5: Mẫu cỡ 50 từ $X \in N(a, \sigma^2)$ với $\sigma = 2$ cho số liệu theo bảng sau:

x_i	10 – 12	12 – 14	14 – 16	16 – 18
n_i	9	18	17	6

a) Tìm ƯLKC của a.

b) Tìm KTC của a với độ tin cậy 97%.

c) Kiểm định ở mức ý nghĩa 7% xem $EX = 13$ hay $EX > 13$.

d) Kiểm định ở mức ý nghĩa 6% xem có phải $EX = 13$ hay không.

HD.

```
X = Join[ Table[11, {9}], Table[13, {18}], Table[15, {17}], Table[17, {6}] ]
```

a) $X \in N\left(a, \sigma^2\right) \Rightarrow a = EX$ có ƯLKC $\boxed{\overline{x} = 13.8}$

```
xNg = Mean[X] // N
```

b) $\sigma = 2 \Rightarrow$ KTC của a: $\left(\overline{x} - z_0\frac{\sigma}{\sqrt{n}},\ \overline{x} + z_0\frac{\sigma}{\sqrt{n}}\right)$.

$\Phi\left(z_0\right) = \frac{1+\gamma}{2} = \frac{1+0.97}{2} = 0.985 \Rightarrow \boxed{z_0 = 2.1701}$

KTC: $(13.1862,\ 14.4138)$.

```
n = Length[X]
z0 = Quantile[NormalDistribution[0, 1], 0.985]
xNg - z0*2/Sqrt[n]
xNg + z0*2/Sqrt[n]
```

c) $z_{qs} = \frac{\overline{x} - a_0}{\sigma}\sqrt{n} = 2.8284.$

$\Phi\left(z_0\right) = 1 - \alpha = 0.93 \Rightarrow \boxed{z_0 = 1.4758}$

$z_{qs} > z_0 \Rightarrow$ bác bỏ $EX = 13$ (chấp nhận $EX > 13$).

```
zQs = (xNg - 13)/2*Sqrt[n]
z0 = Quantile[NormalDistribution[0, 1], 0.93]
```

d) $H_0 : EX = 13,\ H_1 : EX \neq 13,\ \alpha = 0.06.$

$z_{qs} = 2.8284$ (như ý (c)).

$\Phi\left(z_0\right) = 1 - \frac{\alpha}{2} = 0.97 \Rightarrow \boxed{z_0 = 1.8808}$

$|z_{qs}| > z_0 \Rightarrow$ bác bỏ H_0 (chấp nhận $EX \neq 13$).

```
z0 = Quantile[NormalDistribution[0, 1], 0.97]
```

□

VD6: Mẫu cỡ $n = 31$ từ $X \in N\left(a, \sigma^2\right)$ cho số liệu theo bảng sau

x_i	58 – 60	60 – 62	62 – 64	64 – 66
n_i	3	12	13	3

a) Tìm ƯLKC của a và σ^2.

b) Tìm KTC của a với độ tin cậy 92%.

c) Kiểm định ở mức ý nghĩa 4% xem $EX = 64$ hay $EX < 64$.

HD.

```
X = Join[ Table[59, {3}], Table[61, {12}], Table[63, {13}], Table[65, {3}] ]
```

a) $X \in N(a, \sigma^2) \Rightarrow EX = a, DX = \sigma^2$. ƯLKC của a là $\boxed{\overline{x} = 62.0322}$, của σ^2 là $\boxed{s'^2 = 2.6323}$

```
xNg = Mean[X] // N
sDc2 = Variance[X] // N
```

b) KTC của a: $\left(\overline{x} - t_0\frac{s'}{\sqrt{n}},\ \overline{x} + t_0\frac{s'}{\sqrt{n}}\right)$.

$\boxed{s' = 1.6224}$

$t_0 = t_{1-\gamma}^{n-1} = \boxed{t_{1-0.92}^{31-1} = 1.8121}$

KTC: $(61.5042,\ 62.5603)$.

```
n = Length[X]
sDc = StandardDeviation[X] // N
t0 = -Quantile[StudentTDistribution[n - 1], 0.08/2]
xNg - t0*sDc/Sqrt[n]
xNg + t0*sDc/Sqrt[n]
```

c) $t_{qs} = \dfrac{\overline{x} - a_0}{s'}\sqrt{n} = -6.7528.$

$t_0 = t_{2\alpha}^{n-1} = t_{0.08}^{30} = 1.812.$

$t_{qs} < -t_0 \Rightarrow$ bác bỏ $EX = 64$ (chấp nhận $EX < 64$).

```
tQs = (xNg - 64)/sDc*Sqrt[n]
t0 = -Quantile[StudentTDistribution[n - 1], 0.08/2]
```

□

VD7: Phương pháp thứ nhất cho tỷ lệ sản phẩm tốt là 85%. Kiểm tra 300 sản phẩm sản xuất theo phương pháp thứ hai thì thấy có 30 phế phẩm. Hãy kiểm định ở mức ý nghĩa 8% xem có phải phương pháp thứ hai tốt hơn phương pháp thứ nhất không.

HD. $p =$ tỷ lệ sản phẩm tốt sản xuất theo phương pháp thứ hai. Bài toán

$$H_0 : p = 0.85,\ H_1 : p > 0.85,\ \alpha = 0.08.$$

$m =$ số sản phẩm tốt sản xuất theo phương pháp thứ hai $= 270$.

$$z_{qs} = \frac{\dfrac{m}{n} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \boxed{\frac{\dfrac{270}{300} - 0.85}{\sqrt{0.85 \cdot 0.15}}\sqrt{300}} = 2.4254.$$

$\Phi(z_0) = 1 - \alpha = 0.92 \Rightarrow \boxed{z_0 = 1.4051}$

$z_{qs} > z_0 \Rightarrow$ bác bỏ H_0 tức là phương pháp thứ hai tốt hơn.

```
n = 300
m = n - 30
p0 = 0.85
zQs = (m/n - p0)/Sqrt[p0 (1 - p0)]*Sqrt[n]
z0 = Quantile[NormalDistribution[0, 1], 0.92]
```

□

VD8: Với mức ý nghĩa 0.06 hãy kiểm định $H_0 : EX = EY$ với đối thuyết $H_1 : EX > EY$ trong đó X, Y là 2 đại lượng ngẫu nhiên có phân bố chuẩn. Biết rằng 2 mẫu độc lập cỡ $n = 17$ và từ X và $m = 13$ từ Y cho ta số liệu sau:

x_i	21.7	21.9	22.1	22.3	22.5
n_i	1	4	6	5	1

y_i	21.8	22	22.2
m_i	5	7	1

Cho biết $DX = 0.03, DY = 0.02$.

HD. $\boxed{\overline{x} = 22.1118}$, $\boxed{\overline{y} = 21.9385} \Rightarrow z_{qs} = \boxed{\dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{DX}{n} + \dfrac{DY}{m}}}} = 3.01537.$

$\varPhi\left(z_0\right) = 1 - \alpha = 0.94 \Rightarrow \boxed{z_0 = 1.5548}$

$z_{qs} > z_0 \Rightarrow$ bác bỏ H_0.

```
X = Join[Table[21.7, {1}], Table[21.9, {4}], Table[22.1, {6}],
         Table[22.3, {5}], Table[22.5, {1}]
    ]
n = Length[X]
xNg = Mean[X]

Y = Join[Table[21.8, {5}], Table[22, {7}], Table[22.2, {1}]]
m = Length[Y]
yNg = Mean[Y]

dx = 0.03
dy = 0.02

zQs = (xNg - yNg)/Sqrt[dx/n + dy/m]
z0 = Quantile[NormalDistribution[0, 1], 0.94]
```

□

VD9: Khảo sát thu nhập (triệu đồng) trong 1 tháng của 10 công nhân ngành A và 15 công nhân ngành B ta thu được số liệu sau:

Ngành A	Ngành B
1, 1.2, 1.3, 1.3, 1.2, 1.3, 1.4, 1.2, 1.3, 1.4	1.1, 1.3, 1.3, 1.3, 1.4, 1.4, 1.4, 1.3, 1.4, 1.2, 1.5, 1.5, 1.5, 1.2, 1.2

Giả sử thu nhập trong 1 tháng của 1 công nhân ngành A và B là những đại lượng ngẫu nhiên X, Y có phân bố chuẩn với phương sai bằng nhau. Hãy kiểm định ở mức ý nghĩa 3% giả thuyết nói rằng thu nhập trung bình của công nhân 2 ngành trên như nhau với đối thuyết cho rằng thu nhập trung bình của công nhân ngành B cao hơn ngành A.

HD. $H_0 : EX = EY,\ H_1 : EX < EY,\ \alpha = 0.03.$

$\boxed{\overline{x} = 1.26}$, $\boxed{\overline{y} = 1.3333}$, $\boxed{s_X'^2 = 0.01378}$, $\boxed{s_Y'^2 = 0.01524}$

$$t_{qs} = \frac{\overline{x} - \overline{y}}{\sqrt{ns_X^2 + ms_Y^2}} \sqrt{\frac{nm\left(n+m-2\right)}{n+m}} = \boxed{\frac{\overline{x} - \overline{y}}{\sqrt{\left(n-1\right) s_X'^2 + \left(m-1\right) s_Y'^2}} \sqrt{\frac{nm\left(n+m-2\right)}{n+m}}} = -1.4832.$$

$t_0 = t_{2\alpha}^{n+m-2} = \boxed{t_{0.06}^{23} = 1.9782}$

$t_{qs} > -t_0 \Rightarrow$ chấp nhận H_0, tức là thu nhập trung bình của công nhân 2 ngành trên như nhau.

```
X = {1, 1.2, 1.3, 1.3, 1.2, 1.3, 1.4, 1.2, 1.3, 1.4}
n = Length[X]
xNg = Mean[X]
sXDc2 = Variance[X]

Y = {1.1, 1.3, 1.3, 1.3, 1.4, 1.4, 1.4, 1.3, 1.4, 1.2, 1.5, 1.5, 1.5, 1.2, 1.2}
m = Length[Y]
yNg = Mean[Y]
sYDc2 = Variance[Y]

tQs = (xNg - yNg) / Sqrt[(n - 1) sXDc2 + (m - 1) sYDc2] * Sqrt[n*m (n + m - 2)
   /(n + m)]
t0 = -Quantile[StudentTDistribution[n + m - 2], 0.06/2]
```

□

VD10: Mẫu cỡ $n = 100$ từ đại lượng ngẫu nhiên X cho ta số liệu sau:

x_i	1 – 2	2 – 3	3 – 4	4 – 5	5 – 6	6 – 7	7 – 8
n_i	4	12	27	30	20	5	2

Hãy kiểm định ở mức ý nghĩa 7% xem có phải X có phân bố chuẩn $N\left(4, 1.3^2\right)$.

HD. $H_0 : X \in N\left(4, 1.3^2\right),\ H_1 : X \notin N\left(4, 1.3^2\right),\ \alpha = 0.07$

S_i	n_i	p_{i0}	e_i	$\frac{\left(n_i - e_i\right)^2}{e_i}$
$(-\infty, 2]$	4	0.06197	6.1968	0.7788
$(2, 3]$	12	0.1589	15.891	0.9527
$(3, 4]$	27	0.2791	27.9122	0.02981

$(4, 5]$	30	0.2791	27.9122	0.1561
$(5, 6]$	20	0.1589	15.891	1.0625
$(6, 7]$	5	0.05146	5.146	0.004141
$(7, \infty)$	2	0.01051	1.0508	0.8574
			χ^2_{qs}	3.8415

$\chi^2_0 = \chi^2(0.07, 7-1) = 11.6599$. $\chi^2_{qs} < \chi^2_0 \Rightarrow$ chấp nhận H_0.

```
f[x_] := 1/1.3/Sqrt[2 Pi]*E^(-(x - 4)^2/2/1.3^2)   (* hoặc f[x_] := 1/(1.3√(2π)) E^(-(x-4)^2/(2*1.3^2)) *)
s = {-Infinity, 2, 3, 4, 5, 6, 7, Infinity}  (*cận trên của khoảng này là cận dưới của khoảng sau*)
n = {4, 12, 27, 30, 20, 5, 2}

h = Length[n]
size = Total[n]

p0 = Table[Integrate[f[x], {x, s[[i]], s[[i + 1]]}], {i, 1, h}]
p0 = Re[p0]

e = size*p0
t = (n - e)^2/e   (*tính theo vị trí tương ứng*)
chiQs = Total[t]
chi0 = Quantile[ChiSquareDistribution[h - 1], 1 - 0.07]
```

□

VD11: Mẫu cỡ $n = 60$ từ đại lượng ngẫu nhiên X cho ta số liệu dưới đây:

x_i	10 – 11	11 – 12	12 – 13	13 – 14	14 – 15	15 – 16	16 – 18
n_i	9	6	7	8	6	7	17

Kiểm định ở mức ý nghĩa 8% xem X có phân bố đều không?

HD. * $H_0 : X \in U[a, b]$, $H_1 : X \in U[a, b]$, $\alpha = 0.08$.

$\overline{x} = 14.0583$, $s = 2.3648$. Ước lượng theo phương pháp bình phương tối thiểu của a là $a^* = \overline{x} - s\sqrt{3} = 9.9623$, của b là $b^* = \overline{x} + s\sqrt{3} = 18.1543$ (số tham số chưa biết $r = 2$).

```
X = Join[Table[10.5, {9}], Table[11.5, {6}], Table[12.5, {7}],
         Table[13.5, {8}], Table[14.5, {6}], Table[15.5, {7}], Table[17, {17}]]
xNg = Mean[X]
s = Sqrt[(n - 1)/n]*StandardDeviation[X]
a = xNg - s*Sqrt[3]
b = xNg + s*Sqrt[3]
```

* $H_0' : X \in U[a^*, b^*]$, $H_1' : X \notin U[a^*, b^*]$, $\alpha = 0.08$.

S_i	n_i	p_{i0}	e_i	$\dfrac{(n_i - e_i)^2}{e_i}$
$(-\infty, 11]$	9	0.1267	7.6001	0.2578
$(11, 12]$	6	0.1221	7.3242	0.2394
$(12, 13]$	7	0.1221	7.3242	0.01435
$(13, 14]$	8	0.1221	7.3242	0.06236
$(14, 15]$	6	0.1221	7.3242	0.2394
$(15, 16]$	7	0.1221	7.3242	0.01435
$(16, \infty)$	17	0.263	15.7788	0.09451
			χ^2_{qs}	0.9222

$\chi^2_0 = \chi^2(\alpha, h - r - 1) = \chi^2(0.08, 7 - 2 - 1) = 8.3365.$

$\chi^2_{qs} < \chi^2_0 \Rightarrow$ chấp nhận H (X có phân bố đều).

```
f[x_] := If[a <= x <= b, 1/(b - a), 0]
s = {-Infinity, 11, 12, 13, 14, 15, 16, Infinity}
n = {9, 6, 7, 8, 6, 7, 17}

h = Length[n]
size = Total[n]

p0 = Table[Integrate[f[x], {x, s[[i]], s[[i + 1]]}], {i, 1, h}]
e = size*p0
t = (n - e)^2/e
chiQs = Total[t]
chi0 = Quantile[ChiSquareDistribution[h - 2 - 1], 1 - 0.08]
```

$\square$

VD12: Mẫu cỡ 200 từ VTNN (X, Y) cho ta số liệu sau:

$X \backslash Y$	T	N	M
C	26	48	24
K	51	43	8

Kiểm định ở mức ý nghĩa 5% xem X và Y độc lập nhau không?

HD. * Bảng n_{ij}, n_{i*}, n_{*j}:

$X \backslash Y$	T	N	M	$\sum$
C	26	48	24	98
K	51	43	8	102
$\sum$	77	91	32	200

```
n = {{26, 48, 24}, {51, 43, 8}}

{h, k} = Dimensions[n]
nx = Table[Total[n[[i, All]]], {i, 1, h}]
ny = Table[Total[n[[All, j]]], {j, 1, k}]
```

* Bảng e_{ij}:

$X \backslash Y$	T	N	M
C	37.73	44.59	15.68
K	39.27	46.41	16.32

```
size = Total[n, 2]
e = Table[nx[[i]]*ny[[j]]/size // N, {i, 1, h},
    {j, 1, k}]
```

* Bảng $\dfrac{(n_{ij} - e_{ij})^2}{e_{ij}}$:

$X \backslash Y$	T	N	M
C	3.6468	0.2608	4.4147
K	3.5038	0.2506	4.2416

$$\Rightarrow \chi^2_{qs} = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = 16.3181$$

```
t = (n - e)^2/e
chiQs = Total[t, 2]
chi0 = Quantile[ChiSquareDistribution[(h -
    1) (k - 1)], 1 - 0.05]
```

* $\chi^2_0 = \chi^2\left[\alpha, (h-1)(k-1)\right] = \chi^2\left[0.05, (2-1)(3-1)\right] = \boxed{5.9915}$

$\chi^2_{qs} > \chi^2_0 \Rightarrow X, Y$ không độc lập. □

VD13: Mẫu cỡ $n = 12$ từ véctơ ngẫu nhiên (X, Y) cho ta số liệu sau:

x_i	3	3.5	2.5	3.5	3	3.1	2	4	4.5	3	3.5	3.1
y_i	13	14	10	14.5	14	13.5	9.5	16	18	12.5	15	14.5

a) Tìm hệ số tương quan mẫu giữa X và Y. Có thể dùng hồi quy bình phương trung bình tuyến tính của Y đối với X để dự báo giá trị của Y được không, vì sao?

b) Tìm hàm hồi quy bình phương trung bình tuyến tính thực nghiệm của Y đối với X và ước lượng sai số bình phương trung bình.

c) Dự báo giá trị của Y khi biết $X = 2.3$.

HD. a) $\boxed{r = 0.957}$. $|r|$ khá lớn (≥ 0.8) $\Rightarrow$ có thể dùng hồi quy bình phương trung bình tuyến tính của Y theo X để dự báo giá trị của Y.

```
X = {3, 3.5, 2.5, 3.5, 3, 3.1, 2, 4, 4.5, 3, 3.5, 3.1}
Y = {13, 14, 10, 14.5, 14, 13.5, 9.5, 16, 18, 12.5, 15, 14.5}
r = Correlation[X, Y]
```

b) Hàm hồi quy $Y = 3.4397X + 2.6154$

```
data = Transpose[{X, Y}]
model = LinearModelFit[data, x, x]
y = model["BestFit"]
```

$\boxed{s^2_Y = 5.0191}$ $\Rightarrow$ sai số $s^2_{Y/X} = s^2_Y\left(1 - r^2\right) = 0.4221.$

```
n = Length[Y]
sY2 = (n - 1)/n*Variance[Y]
sY2*(1 - r^2)
```

c) Khi $X = 2.3$ ta dự báo $Y = aX + b = 10.5266$.

```
y /. x -> 2.3
```

□

Thống kê thời gian thực hành:

VD	1	2	3	4	5	6	7	8	9	10	11	12	13	$\sum$
Time	1'2"	1'21"	1'1"	1'37"	3'9"	2'33"	47"	2'5"	2'24"	3'9"	4'26"	1'55"	2'45"	31'56"