

Tên học phần: Thực hành Xác suất Thống kê

Mã HP:

Thời gian làm bài:

Ngày thi:

Ghi chú: Sinh viên không được sử dụng tài liệu. Đề thi gồm ?? trang.

Họ tên sinh viên: _____ MSSV: _____ STT: _____

Cho đoạn code sau, trả lời các câu hỏi 1-3.

```
# So hang tong quat
shtq <- function(n) {
  if (n < 2)
    n
  else
    shtq(n - 1) + 3*shtq(n - 2)
}
## Day so
day <- function(n){
  list = numeric(n)
  for(k in 1 : n){
    list[k] = shtq(k) }
  return(list)
}
z = day(8)
```

Câu 1. Trong các đáp án sau; đáp án nào mô tả đúng nhất về số hạng tổng quát trong câu lệnh trên

A.
$$\begin{cases} a_0 = 0 \\ a_1 = 1 \\ a_n = a_{n-1} + 3a_{n-2}; n \geq 2. \end{cases}$$

B.
$$\begin{cases} a_0 = 1 \\ a_1 = 1 \\ a_n = a_{n-1} + 3a_{n-2}; n \geq 2. \end{cases}$$

C.
$$a_n = \begin{cases} n & ; n < 2 \\ a_{n-1} + 3a_{n-2} & n \geq 2. \end{cases}$$

D.
$$\begin{cases} a_0 = 0 \\ a_1 = 1 \\ a_n = \sqrt{n-1} + 3\sqrt{n-2}; n \geq 2. \end{cases}$$

Câu 2. Kết quả của câu lệnh `rep(z[5], z[3])` và `rep(c(z[2], z[5]), c(z[3], z[4]))` lần lượt là
và

Câu 3. Tính: `mean(z) + max(z) - sd(z)` =

Cho đoạn code sau, trả lời các câu hỏi 4-7.

```
x <- seq(10, 22, 2)
y <- x - 3
z <- x + 3
df = data.frame(x, y, z)
u = (length(df)*max(df)) %% abs(ncol(df) - nrow(df))
```

Câu 4. Kết quả của x là

Người ra đề/MSCB: Nguyễn Thị Hồng Nhung/ .. Người duyệt đề: Trang 1/??

Chữ ký: Chữ ký:

Câu 5. Kết quả của df là

Câu 6. Để trích ra một vector (ví dụ cột y) trong df, ta có thể thực hiện bởi câu lệnh và nếu muốn in một dataframe con gồm các cột x,z ở dòng 2,5,6 thì ta sẽ thực hiện câu lệnh

Câu 7. Trong các đáp án sau, đâu là kết quả của u

- A. 3. B. 4.
C. 2. D. Tất cả đều sai.

Câu 8. Để hiển thị địa chỉ hiện hành trong R, ta sẽ dùng lệnh và nếu muốn thay đổi địa chỉ làm việc trong R ta dùng lệnh

Câu 9. Trong các lệnh sau, có bao nhiêu câu lệnh dùng để vẽ hàm mật độ của phân phối chuẩn $\mathcal{N}(0,1)$.

```
curve(dnorm(x, 0, 1), from = -2, to = 2)
curve(pnorm(x, 0, 1), from = -2, to = 2)
line(seq(-2,2,0.01),dnorm(seq(-2,2,0.01)))
plot(seq(-2,2,0.01),pnorm(seq(-2,2,0.01)))
```

- A. 1. B. 2.
C. 3. D. 4.

Sử dụng giả thiết sau cho câu hỏi 10-11

Trung bình mỗi phút có 3 cuộc gọi đến tổng đài của trung tâm đặt vé máy bay. Gọi X là số cuộc gọi đến tổng đài đó trong 1 phút.

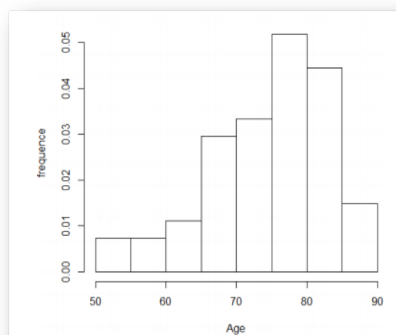
Câu 10. Người ta muốn biết xác suất có từ 2 đến 8 cuộc gọi trong 1 phút thì có thể dùng câu lệnh nào

- A. $\text{dpois}(8,1) - \text{dpois}(2,1)$. B. $\text{ppois}(8,1) - \text{ppois}(2,1)$.
C. $\text{dpois}(8,1) - \text{dpois}(1,1)$. D. $\text{ppois}(8,1) - \text{ppois}(1,1)$.

Câu 11. Ta xét khoảng 100 khoảng thời gian một phút liên tiếp và gọi U là số khoảng thời gian một phút không nhận được cuộc gọi nào. Viết câu lệnh tính $\mathbb{P}(U \leq 1)$

.....
.....
.....
.....

Câu 12. Ta có thể dùng lệnh nào trong các lệnh sau để được biểu đồ dưới đây?



- A. `hist(df$Age, xlab='Age', ylab='frequece')` .
B. `plot(df$Age)` .
C. `hist(df$Age)` .
D. `plot(df$Age, type='h')` .

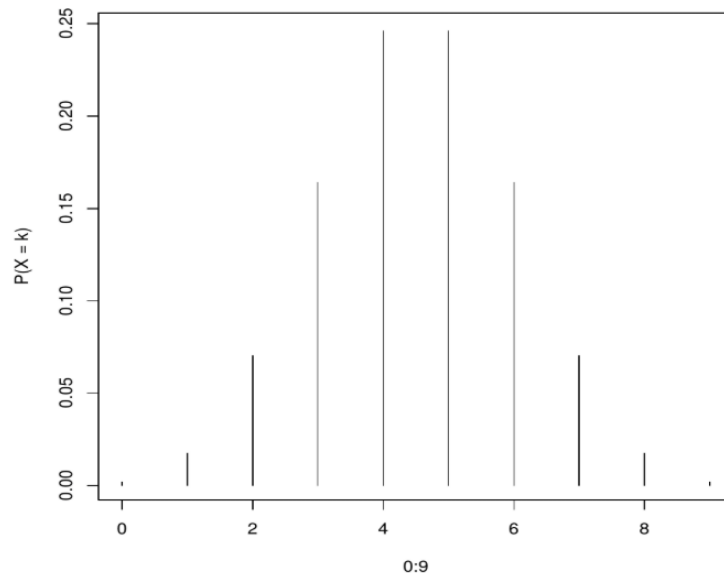
Câu 13. Cho X là BNN có phân phối nhị thức $B(n, p)$; giả sử rằng n đủ lớn, p đủ nhỏ sao cho $np \geq 5$ và $np(1-p) \geq 5$ thì ta có thể xấp xỉ xác suất của nó tại k với $k = \{0, 1, \dots, n\}$ bởi lệnh nào trong các lệnh dưới đây?

- A. `dnorm(k, n*p, n*p*(1-p))`.
- B. `dnorm(k, n*p, sqrt(n*p*(1-p)))`.
- C. `pnorm(k, n*p, n*p*(1-p))`.
- D. `pnorm(k, n*p, sqrt(n*p*(1-p)))`.

Câu 14. Giá trị của $ppois(x_0, \lambda)$ bằng với

- A. Giá trị của hàm phân phối (tích lũy) của biến ngẫu nhiên phân phối Poisson $\mathcal{P}(\lambda)$ tại x_0 .
- B. $\sum_{k \in \mathbb{Z}: 0 \leq k \leq x_0} \frac{e^{-\lambda} \lambda^k}{k!}$.
- C. $\mathbb{P}(X \leq x_0)$ trong đó $X \sim \mathcal{P}(\lambda)$.
- D. Tất cả các giá trị liệt kê ở trên.

Câu 15. Trong các lệnh sau, lệnh nào có thể vẽ được hình bên dưới.



Hình 1: Biểu đồ cột

- A. `curve(dbinom(x, 9, 0.5), from = 0, to = 10)`.
- B. `curve(dnorm(x, 2, 1), from = -1, to = 5)`.
- C. `hist(c(0:9), dbinom(0:9, 9, 0.5))`.
- D. `plot(0:9, dbinom(0:9, 9, 0.5), type='h', ylab = "P(X = k)")`.

Người ra đề/MSCB: Nguyễn Thị Hồng Nhung/ .. Người duyệt đề: Trang 3/??

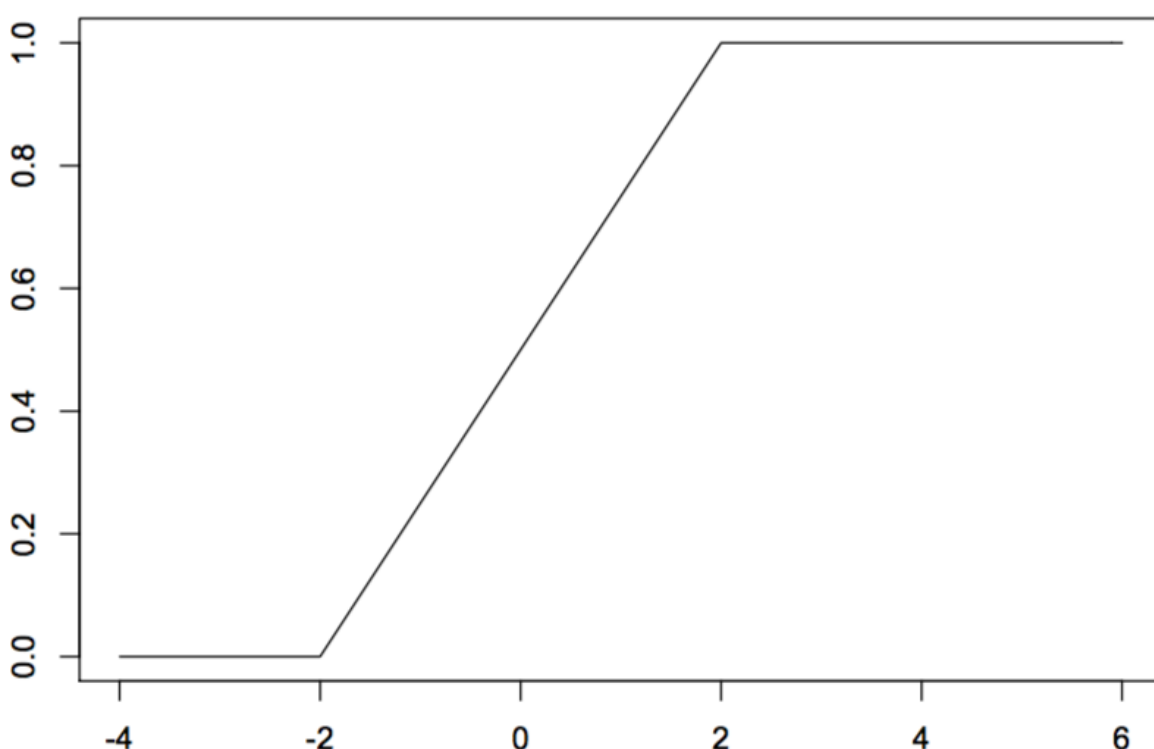
Chữ ký: Chữ ký:

Câu 16. Giá trị của `dchisq(3, 5)` cho ta biết

- A. Giá trị của hàm mật độ biến ngẫu nhiên có phân phối Chi bình phương (5 bậc tự do) tại $x = 3$.
- B. Giá trị của hàm phân phối biến ngẫu nhiên có phân phối Chi bình phương (5 bậc tự do) tại $x = 3$.
- C. Giá trị của hàm mật độ biến ngẫu nhiên Chi bình phương (3 bậc tự do) tại $x = 5$.
- D. A và B đều đúng.

Câu 17. Để phát sinh một mẫu 10 phần tử có phân phối siêu bội với $N = 100$, $M = 25$ và cỡ mẫu $n = 15$; dùng một hàm có sẵn trong R, hãy viết một đoạn lệnh mô phỏng điều đó.....

Câu 18. Cho biết đồ thị hàm phân phối (tích lũy), như Hình 2, là của phân phối nào



Hình 2: Hàm phân phối tích lũy

- A. Phân phối chuẩn $\mathcal{N}(1, 1)$.
- B. Phân phối đều $U([-2, 2])$.
- C. Phân phối mũ $\text{Exp}(2)$.
- D. Phân phối Student(10).

Câu 19. Đề thi cuối học kỳ môn toán của học sinh lớp 12 có 50 câu hỏi dạng trắc nghiệm, mỗi câu 5 đáp án trong đó chỉ có 1 đáp án đúng. Một học sinh không học bài, khi đi thi sinh viên làm bài bằng cách chọn ngẫu nhiên một đáp án. Tính xác suất sinh viên đó trả lời đúng ít nhất 25 câu hỏi.

- A. `pnorm(15/sqrt(8))`.
- B. `pnorm(15/8)`.
- C. `1-pnorm(15/sqrt(8))`.
- D. `dnorm(0.5)`.

Câu 20. Để kiểm định trung bình của hai mẫu X,Y độc lập với đối thuyết $\mu_X > \mu_Y$ và độ tin cậy $1 - \alpha = 0.95$, hãy viết một đoạn code thực hiện điều đó

Câu 21. Hãy nối 2 bảng sau để hoàn thành chức năng của các câu lệnh

A. Phát sinh 1000 số ngẫu nhiên có phân phối đều $U(-1, 1)$	1. <code>rnorm(100)</code>
B. Tính giá trị hàm mật độ phân phối chuẩn $N(0, 1)$ tại $x = 2$	2. <code>dpois(3, 1)</code>
C. Tính xác suất tại $x = 3$ của biến ngẫu nhiên có pp Poisson ($\mu = 1$)	3. <code>runif(1000, -1, 1)</code>
D. Phát sinh 100 số ngẫu nhiên có phân phối chuẩn $N(0, 1)$	4. <code>dnorm(2)</code>

Trả lời: 1-; 2 -; 3-; 4-

Câu 22. Cho

$$z_{1-\alpha/2} \triangleq \text{qnorm}(1 - \alpha/2) \text{ và } t_{1-\alpha/2}^{n-1} \triangleq \text{qt}(1 - \alpha/2, \text{df} = n-1)$$

hoàn thành các chỗ trống trong đoạn code sau

```
path = 'D://Works'
setwd(path)
dtf = read.csv('data01.csv', header=TRUE)
Age = dtf$Age
KTC_mean <- function(data, alpha, sig = 'None'){
  n = length(data)
  m = mean(data)
  sd = sd(data)
  zalp = qnorm(1 - ...../2)
  talp = qt(1 - alpha/2, ..... )
  if(sig != 'None')
    eps = sig*...../sqrt(n)
  else if(sig == 'None')
    if( n < 30)
      eps = sd*talp/sqrt(n)
    else if (n >= ..... )
      eps = sd*...../sqrt(n)
  return(c(m - eps, m + eps))
}
KTC_mean(Age, 0.05)
```

Câu 23. Cho kết quả của kiểm định sau

```
test2
2-sample test for equality of proportions with continuity correction

data: y out of n
X-squared = 2.2222, df = 1, p-value = 0.06802
alternative hypothesis: less
95 percent confidence interval:
 -1.00000000 0.01374729
sample estimates:
prop 1 prop 2
0.25 0.50
```

Hãy cho biết kết quả trên nói về kiểm định của
, trong đó đối thuyết của kiểm định là, cùng với
 p -giá trị bằng, ta có thể kết luận rằng

Câu 24. Một nhóm sinh viên đo nhiệt độ ở những độ cao khác nhau và thu được bảng số liệu sau.

Elevation(ft)	600	1000	1250	1600	1800	2100	2500	2900
Temperature(F)	56	54	56	50	47	49	47	45

Sử dụng các câu lệnh trong R để vẽ đồ thị phân tán và đường hồi quy nhiệt độ theo độ cao cùng hệ trục tọa độ.

Câu 25.

```
path = 'D://Works'
setwd(path)
dtf = read.csv('data01.csv', header=TRUE)
Age = dtf$Age
U70 = Age[Age > 70]
KTC_prop <- function(data.p, data, alpha){
  phat = length(data.p)/ length(data) ## ty le mau
  eps = qnorm(1 - alpha/2)*sqrt(phat*(1-phat)/n)
  print('KTC cho ty le la')
  return(c(phat - eps, phat + eps))
}
KTC_prop(U70, Age, 0.05)
```

Hàm KTC_prop cho biết

- A. Input các tham số dữ liệu mẫu (data), dữ liệu thỏa tính chất nào đó để truy xuất tỷ lệ mẫu (data.p) và mức ý nghĩa (alpha) .
- B. Output là khoảng tin cậy cho tỷ lệ p với độ tin cậy α ..
- C. A, B đều đúng .
- D. A, B đều sai.

Xem đoạn code và kết quả sau

```
path = 'D://Works'
setwd(path)
data = read.csv('rocket.motor.csv', header=TRUE)
SK = data$streng; mu_0 = 2000
test = t.test(SK, alternative = "two.sided", mu = mu_0, conf.level = 0.95)
```

One Sample t -test

```
data: SK
t = 1.9799, df = 19, p-value = 0.06238
alternative hypothesis: true mean is not equal to 2000
95 percent confidence interval:
 1992.438 2272.377
sample estimates:
mean of x
 2132.407
```

Câu 26. Hàm `t.test(Age, alternative = "two.sided", mu = mu_0, conf.level = 0.95)` dùng để :

- A. Khoảng tin cậy cho trung bình của mẫu với đối thuyết $\mu \neq 2000$ và độ tin cậy $1 - \alpha = 95\%$.
- B. Khoảng tin cậy cho trung bình của mẫu với đối thuyết $\mu < 2000$ và độ tin cậy $1 - \alpha = 95\%$.
- C. Khoảng tin cậy cho trung bình của mẫu với đối thuyết $\mu \neq 2000$ và mức ý nghĩa $1 - \alpha = 0.05$.
- D. Khoảng tin cậy cho trung bình của mẫu với đối thuyết $\mu > 2000$ và độ tin cậy $1 - \alpha = 0.95$.

Câu 27. Kết quả của lệnh `test$statistic` bằng

- A. 0.9218 .
- B. 1.979949.
- C. 19 .
- D. 0.06238 .

Xét dữ liệu trong file `house.price.csv` với các tên biến như đoạn lệnh bên dưới, hãy trả lời các câu hỏi 23 và 24.

```
dtf = read.csv('house.price.csv', header=TRUE)
Tax = dtf$taxes
Sales = dtf$sale.price
Tax2 = Tax[Tax > 8]
Sale2 = Sales[Sales > 35]
```

Câu 28. Cho kết quả của kiểm định sau

```
test1
      Welch Two Sample t-test

data:  Tax and sales
t = -22.2571, df = 26.179, p-value = 1
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -30.36865      Inf
sample estimates:
mean of x mean of y
 6.404917 34.612500
```

Hãy cho biết kết quả trên nói về kiểm định của
....., trong đó đối thuyết của kiểm định là, cùng với
mức ý nghĩa, ta có thể kết luận rằng

```
df = read.csv('chloride.csv', header=TRUE)
y = df$y ## Nồng độ cloride
x = df$x ## tỷ lệ phân trăm
lm(y ~ x)
```

```
Call:
lm(formula = y ~ x)
```

```
Coefficients:
(Intercept)          x
    0.4705         20.5673
```

Câu 29. Đoạn lệnh trên cho biết

- A. Kết quả mô hình hồi quy $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, cho y- nồng độ clorua(đv: mg/l) theo x- diện tích ở đầu nguồn x(đv:%) với các hệ số hồi quy $\hat{\beta}_0 = 20.5673$ và $\hat{\beta}_1 = 0.4705$.
- B. Kết quả mô hình hồi quy $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, cho y- nồng độ clorua(đv: mg/l) theo x- diện tích ở đầu nguồn x(đv:%) với các hệ số hồi quy $\hat{\beta}_0 = 0.4705$ và $\hat{\beta}_1 = 20.5673$.
- C. Kết quả mô hình hồi quy $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, cho x- diện tích ở đầu nguồn x(đv:%) theo y- nồng độ clorua(đv: mg/l) với các hệ số hồi quy $\hat{\beta}_0 = 20.5673$ và $\hat{\beta}_1 = 0.4705$.
- D. Kết quả mô hình hồi quy $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, cho x- diện tích ở đầu nguồn x(đv:%) theo y- nồng độ clorua(đv: mg/l) với các hệ số hồi quy $\hat{\beta}_0 = 0.4705$ và $\hat{\beta}_1 = 20.5673$.

Câu 30. Để tìm tính khoảng tin cậy 99% cho β_0 , hãy hoàn chỉnh vào đoạn lệnh sau , biết rằng $\beta_1 \in \left[\hat{\beta}_1 - t_{1-\alpha/2}^{n-2} \sqrt{\frac{MSE}{Sxx}}; \hat{\beta}_1 + t_{1-\alpha/2}^{n-2} \sqrt{\frac{MSE}{Sxx}} \right]$ trong đó $\hat{\beta}_1$ là hệ số góc trong mô hình hồi quy và MSE, SSE, Sxx thoả các công thức sau

$$MSE = \frac{SSE}{n-2}, \quad SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad Sxx = \sum_{i=1}^n (x_i - \bar{x})^2.$$

```
beta1_CI <- function(x, y, alpha){
  n = length(x)
  result = lm(.....)
  res = resid(result)
  beta1.hat = (coef(.....))[[2]]
  MSE = sum( res^2 )/(n-2)
  Sxx = sum((x - mean(.....))^2)
  eps = qt(1 - alpha/2, df=n-2)*sqrt(...../.....)
  print('KTC cho beta 1')
  return(c(..... - eps, beta1.hat + .....))
}
beta1_CI(x, y, .....
```

HẾT.