

Tên học phần:	Thực hành Xác suất Thống kê	Mã HP:	MTH00085
Thời gian làm bài:	60 phút	Ngày thi:	
Ghi chú: Sinh viên <input type="checkbox"/> không được phép sử dụng tài liệu khi làm bài.			

Chú thích: Đáp án phần điền khuyết chữ màu đỏ, đáp án phần trắc nghiệm tô vàng, những phần giải thích thêm cho sinh viên, hoặc nhắc nhớ nội dung của lý thuyết liên quan đến câu hỏi thì chữ màu xanh dương.

Câu 1–3 sử dụng thông tin trong đoạn code sau

```
# 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) ## Tạo ra một biến list là vecto gồm n số 0
  for(k in 1 : n){
    list[k] = shtq(k) } # gán giá trị shtq(k) vào phần tử thứ k của biến list
  return(list) }
z = day(8)
## z: 1 1 4 7 19 40 97 217
```

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

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

Câu 2. Kết quả của câu lệnh `rep(z[5],z[3])`: 19 19 19 19

####Hướng dẫn: (z[5]=19, z[3]=4, vậy lặp lại 4 lần số 19)

và `rep(c(z[2],z[5]),c(z[3],z[4]))`: 1 1 1 1 19 19 19 19 19 19 19 19

####Hướng dẫn: (z[2]=1, z[3]=4, z[4]=7, z[5]=19, vậy lặp lại 4 lần số 1 và lặp lại 7 lần số 19)

###Xem lại : `rep(x,n)` trong đó x, n là các vecto có thể có 1 phần tử (trường hợp `rep(z[5],z[3])`) hoặc có nhiều phần tử.

Câu 3. `mean(z)+ max(z) -sd(z)` = 48.25+217-75.5=189.75 ## có thể dùng máy tính casio

Dạng câu hỏi tương tự : `median(z)+min(z)-length(z)`

Câu 4- Câu 7 sử dụng thông tin trong đoạn code sau

```
x <- seq(10, 22, 2) ## tạo ra một vecto gồm các số trong đoạn từ 10 đến 22, cách nhau 2 đơn vị
####x
##10 12 14 16 18 20 22
y <- x - 3 ## lấy từng phần tử của x trừ cho 3
### 7 9 11 13 15 17 19
z <- x + 3 ## lấy từng phần tử của x cộng cho 3
##13 15 17 19 21 23 25
df = data.frame(x, y, z)
u = (length(df)*max(df)) %% abs(ncol(df) - nrow(df))
```

Câu 4. Kết quả của x là **10 12 14 16 18 20 22**

Câu 5. Kết quả của df

```
  x y z
10 7 13
12 9 15
14 11 17
16 13 19
18 15 21
20 17 23
22 19 25
```

Câu 6. Để trích ra một vector (ví dụ cột y) trong df, ta có thể thực hiện bởi câu lệnh **df\$y** và nếu muốn in một dataframe con gồm các cột x,z ở dòng 2,5,6 thì ta sẽ thực hiện câu lệnh **df[c(2,5,6),c(1,3)]**

Cần xem lại lý thuyết về nhập dữ liệu (vecto, ma trận, dataframe....), chiết dữ liệu (vecto, ma trận, dataframe....),,, ghép dữ liệu (vecto, ma trận, dataframe....),

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

A. 3

B. 4

C. 2

D. Tất cả đều sai

##Hướng dẫn: length(df)= số biến trong dataframe, max(df)= giá trị lớn nhất trong dataframe, nrow(df),ncol(df) lần lượt là số dòng, số cột trong dataframe. abs() phép toán trị tuyệt đối, a%%b phép toán lấy phần dư trong phép chia a:b

###Sinh viên cần ôn tập lại các phép toán trong R

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

Sv cần xem lại bài 1 giới thiệu về R, file bài 1 Lý thuyết Full. Địa chỉ làm việc, đổi địa chỉ làm việc, Liệt kê các file trong thư mục làm việc, Lưu Workspace đang làm việc, lưu biến, xoá biến, khôi phục biến,....

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

```
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)))
curve(dnorm(x, 0, 1), from = -2, to = 2)
```

A. 1

B. 2

C. 3

D. 4

Câu 10- Câu 11 sử dụng giả thiết sau (Sinh viên cần xem lại bài Một số phân phối xác suất thường dùng)

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.

X có phân phối Poisson với $\lambda = 3$

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

câu hỏi $P(2 \leq X \leq 8) = P(1 < X \leq 8) = F_X(8) - F_X(1)$

A. `dpois(8,3) - dpois(2,3)`

B. `ppois(8,3) - ppois(2,3)`

C. `dpois(8,3) - dpois(1,3)`

D. `ppois(8,3) - ppois(1,3)`

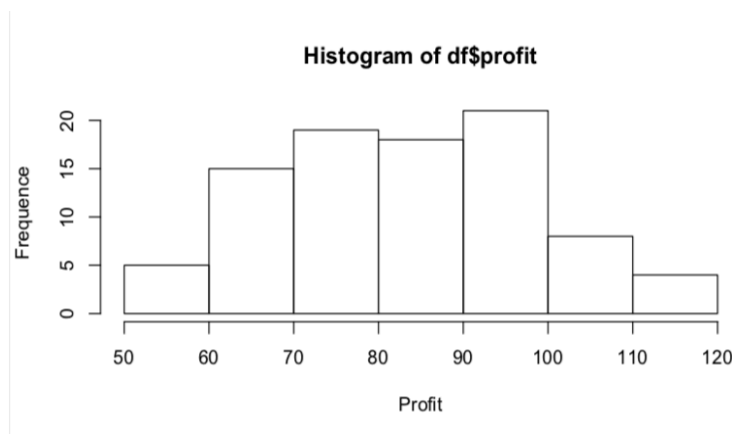
Câu 11. Ta xét khoảng 100 khoảng thời gian một phút liên tiếp và gọi U là số khoảng thời gian một phút không nhận được cuộc gọi nào. Viết câu lệnh tính $P(U \leq 1)$.

U có phân phối nhị thức $U \sim B(100, p)$ trong đó $p = P(X=0)$

..## Tính $P(U \leq 1)$ là hàm phân phối của biến ngẫu nhiên có phân phối nhị thức tại giá trị 1 nên.....
`pbinom(1,100,dpois(0,3))`

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

Sinh viên cần xem lại các loại đồ thị thống kê, đặt tên đồ thị, các trục....



A. `hist(df$profit, xlab="Profit", ylab="Frequency")`

B. `plot(df$profit)`

C. `hist(df$profit)`

D. `plot(df$profit, type='h')`

Câu 13. Cho $X \sim B(n, p)$, giả sử rằng n đủ lớn, p đủ nhỏ sao cho $np \geq 5, n(1-p) \geq 5$ 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)))`

Sinh viên xem lại định lý giới hạn trung tâm

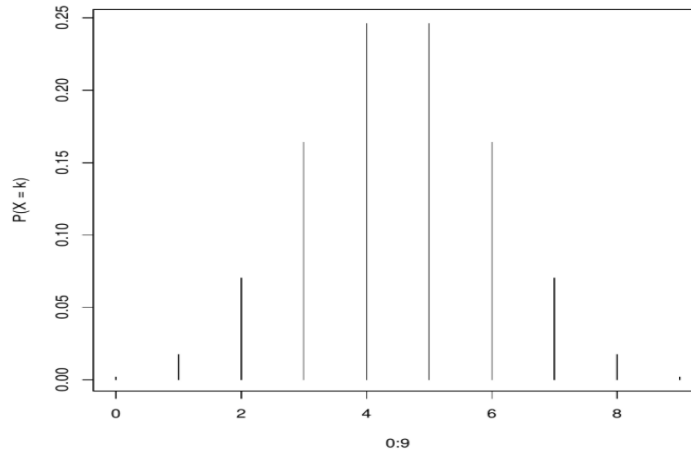
Câu 14. Giá trị của `ppois(x0, lambda)` bằng với

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI MẪU 02
Học kỳ 1 – Năm học 2019-2020

- A. Giá trị của hàm phân phối (tích lũy) của biến ngẫu nhiên tuân theo phân phối Poisson, $P(\lambda)$, tại x_0 .
 B. $\sum_{k \in \mathbb{Z}, 0 \leq k \leq x_0} e^{-\lambda} \frac{\lambda^k}{k!}$
 C. $P(X \leq x_0)$ trong đó $X \sim P(\lambda)$.
 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.

Quan sát biểu đồ, nhận diện đồ thị thống kê sau đó chọn câu lệnh phù hợp với loại đồ thị đó



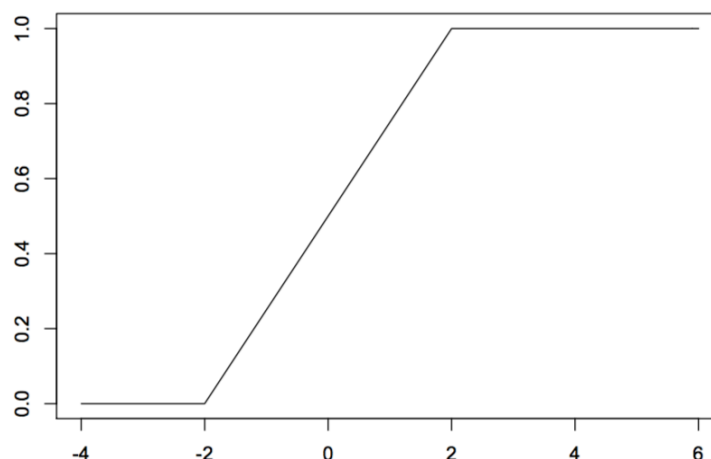
- A. `curve(dbinom(x, 9, 0.5), from = 0, to = 10)`
 B. `curve(dnorm(x, 2, 1), from = -1, to = 5)`
 C. `hist(c(0:9), dbinom(0:9, 9, 0.5))`
 D. `plot(0:9, dbinom(0:9, 9, 0.5), type='h', ylab = "P(X = k)")`

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 `rhyper(10,25,75,15)`.

Câu 18. Cho biết đồ thị hàm phân phối (tích lũy), như hình bên dưới, là của phân phối nào




```
else if (n >= 30)
  eps = sd*tailp/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 (# 1)

data: y out of n

X-squared = 2.2222, df = 1, **p-value** = 0.06802 (#3)

alternative hypothesis: less (#2)

95 percent confidence interval: (# Khoảng ước lượng với độ tin cậy 95%)

-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 **tỷ lệ 2 mẫu** (HD: xem #1), trong đó đối thuyết của kiểm định là $H_1: p_1 - p_2 < 0$ (HD: Xem #2), cùng với p-giá trị bằng 0.06802 (HD: Xem #3), ta có thể kết luận rằng chưa đủ cơ sở bác bỏ H_0 nghĩa là **tỷ lệ mẫu 1 không nhỏ hơn tỷ lệ mẫu 2** ($p_1 - p_2 \geq 0$) với mức ý nghĩa 5%.

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.

Elevation<-c(600,1000,1250,1600,1800,2100,2500,2900)

Temperature<-c(56,54,56,50,47,49,47,45)

plot(Elevation, Temperature)

abline(lm(Temperature~Elevation))

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) ## tỷ lệ mẫu
  eps = qnorm(1 - alpha/2)*sqrt(phat*(1-phat)/n)
  print('KTC cho tỷ lệ là')
  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 alpha.
- C. A, B đều đúng.
- D. A, B đều sai.

Câu 26-27 sử dụng thông tin trong 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. t.test(Age, alternative = "two.sided", mu = mu0, 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. 1.9799 C. 19 D. 0.6238

Câu 28. Xét dữ liệu trong file **house.price.csv** với các tên biến như đoạn lệnh bên dưới, hãy

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

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 **trung bình 2 mẫu**, trong đó đối thuyết của kiểm định là $H_1: \mu_{Tax} - \mu_{Sales} > 0$, cùng với mức ý nghĩa **5%**, ta có thể kết luận rằng **chưa đủ cơ sở bác bỏ H_0 vì $p - value = 1 > 0.05$, nghĩa là $\mu_{Tax} - \mu_{Sales} \leq 0$**

Câu 29.

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

Call:

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

Coefficients:

```
(Intercept)      x
  0.4705    20.5673
```

Đoạn lệnh trên cho biết

- 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 $\hat{\beta}_1 = 0.4705$.
- 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 $\hat{\beta}_1 = 20.5673$ và $\hat{\beta}_0 = 0.4705$.
- 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 $\hat{\beta}_1 = 0.4705$.
- 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 $\hat{\beta}_1 = 20.5673$ và $\hat{\beta}_0 = 0.4705$.

Câu 30. Để tìm tính khoảng tin cậy 99% cho β_1 , 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 $MSE = \frac{SSE}{n-2}$; $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$; $Sxx = \sum_{i=1}^n (x_i - \bar{x})^2$.

```
beta1_CI <- function(x, y, alpha){
```

```
  n = length(x)
```

```
  result = lm(y~x) # kết quả của mô hình hồi quy tuyến tính y theo x được gán vào biến result
```

```
  res = resid(result) ## Tính các giá trị thặng dư (yi - ŷi)
```

```
  beta1.hat = (coef(result))[[2]] # coef(result) trả ra các hệ số của mô hình hồi quy # (coef(result))[[2]]: Hệ số beta 1, (coef(result))[[1]] hệ số beta0.
```

```
  MSE = sum( res^2 )/(n-2) ##SSE= sum( res^2 )
```

```
  Sxx = sum((x - mean(x))^2)
```

```
  eps = qt(1 - alpha/2, df=n-2)*sqrt(MSE/Sxx)
```

```
  print('KTC cho beta 1')
```

```
  return(c(beta1.hat - eps, beta1.hat + eps))
```

```
}
```

```
beta1_CI(x, y, 0.01)
```




TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI MẪU 02
Học kỳ 1 – Năm học 2019-2020

(Đề thi gồm 9 trang)

Họ tên người ra đề/MSCB: Nguyễn Thị Hồng Nhung Chữ ký: [Trang 9/9]
Họ tên người duyệt đề: Chữ ký: