

Tên học phần: 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: _____

1. (A) (B) (C) (D) 11. (A) (B) (C) (D) 21. (A) (B) (C) (D)
2. (A) (B) (C) (D) 12. (A) (B) (C) (D) 22. (A) (B) (C) (D)
3. (A) (B) (C) (D) 13. (A) (B) (C) (D) 23. (A) (B) (C) (D)
4. (A) (B) (C) (D) 14. (A) (B) (C) (D) 24. (A) (B) (C) (D)
5. (A) (B) (C) (D) 15. (A) (B) (C) (D) 25. (A) (B) (C) (D)
6. (A) (B) (C) (D) 16. (A) (B) (C) (D) 26. (A) (B) (C) (D)
7. (A) (B) (C) (D) 17. (A) (B) (C) (D) 27. (A) (B) (C) (D)
8. (A) (B) (C) (D) 18. (A) (B) (C) (D) 28. (A) (B) (C) (D)
9. (A) (B) (C) (D) 19. (A) (B) (C) (D) 29. (A) (B) (C) (D)
10. (A) (B) (C) (D) 20. (A) (B) (C) (D) 30. (A) (B) (C) (D)

Câu 1. Trong các đoạn code R sau, đoạn nào trả ra kết quả là TRUE .

- A. $(2 \neq 2) \ \&\& \ (3 \leq 3)$. B. $(3 \neq 4) \ || \ (4 \geq 3)$.
C. $(2 \leq 5) \ \&\& \ (4 \leq 3)$. D. `isTRUE(4 == 5)` .

Cho đoạn code sau, trả lời các câu hỏi 2-3.

```
perm <- function(n){  
  if(n==1){  
    return(matrix(1))  
  } else {  
    sp <- perm(n-1)  
    p <- nrow(sp)  
    A <- matrix(nrow=n*p,ncol=n)  
    for(i in 1:n){  
      A[(i-1)*p+1:p,] <- cbind(i,sp+(sp>=i))  
    }  
    return(A)  
  }  
}  
Z=perm(3);
```

Câu 2. Cho biết số dòng của ma trận Z

- A. 3. B. 9. C. 8. D. 6.

Câu 3. Tìm số tự nhiên n sao cho ma trận $perm(n)$ có 24 dòng.

- A. 3. B. 4. C. 12. D. 24.

Câu 4. Câu lệnh R nào dùng để mô phỏng việc tung đồng xu cân đối đồng chất 5 lần.

- A. `sample(c('H','T'),5)`.
B. `sample(c(0,1), 5, prob=c(1/2,1/2))`.
C. `sample(c(0,1), 5, replace= T)`.
D. `sample(5, c('H','T'), replace= T)`.

Cho đoạn code sau, trả lời các câu hỏi 5-6.

```
x <- c(rep(1, 3), rep(2, 4), rep(3,5), rep(4,4) ,rep(5,3))  
a = mean(x)  
b = length(x)  
c = median(x)
```

Câu 5. Kết quả của $a * b$ bằng và $b * a$ bằng

Câu 6. Giá trị của $a * b - c$ bằng

- A. 54. B. 24. C. 10. D. 60.

Câu 7. Câu lệnh R: `3*length(sample(1:6,5))+10` cho ra kết quả nào sau đây?

- A. 25. B. 28. C. 43. D. 13.

Cho đoạn code sau, trả lời các câu hỏi 2-3.

```
df <- data.frame(  
  STT = c(seq(1,6,1)),  
  Name= c("Pearson", "Neymann","Fisher","Gosset","Bayes","Poisson"),  
  Birthday = c(1857,1894,1890,1876,1702,1781),  
  Died=c(1936, 1981, 1962, 1937, 1761, 1840)  
)
```

Câu 8. Dựa vào dataframe df, hãy cho biết năm sinh và năm mất của cha để kiểm định t (Gosset, William Sealy Gosset).

- A. 1890-1962. B. 1936-1937. C. 1876-1937. D. 1890-1962.

Câu 9. Lệnh `subset (Name, Birthday%%25>13)` cho ra tên của hai nhà thống kê đã có quan điểm đối lập nhau trong vấn đề kiểm định:

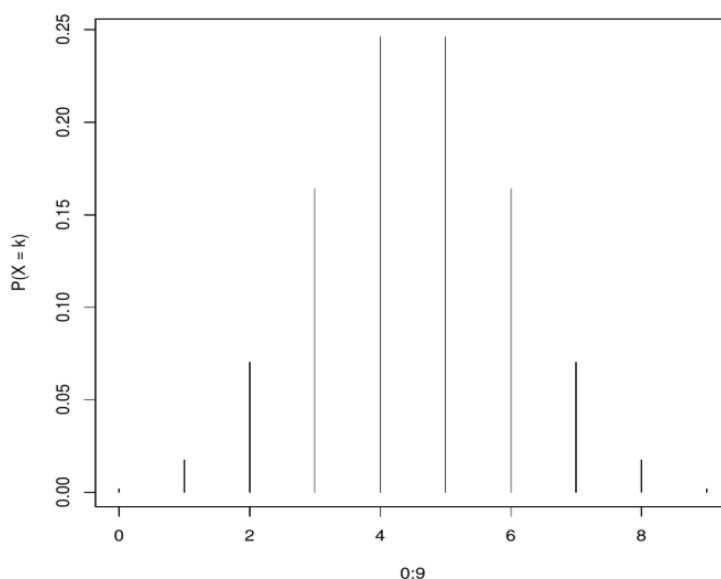
- A. [1] "Bayes" "Poisson". B. [1] "Pearson" "Neymann" .
C. [1] "Gosset" "Poisson". D. [1] "Neymann" "Fisher" .

Câu 10. Giá trị của `ppois (x0, 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.

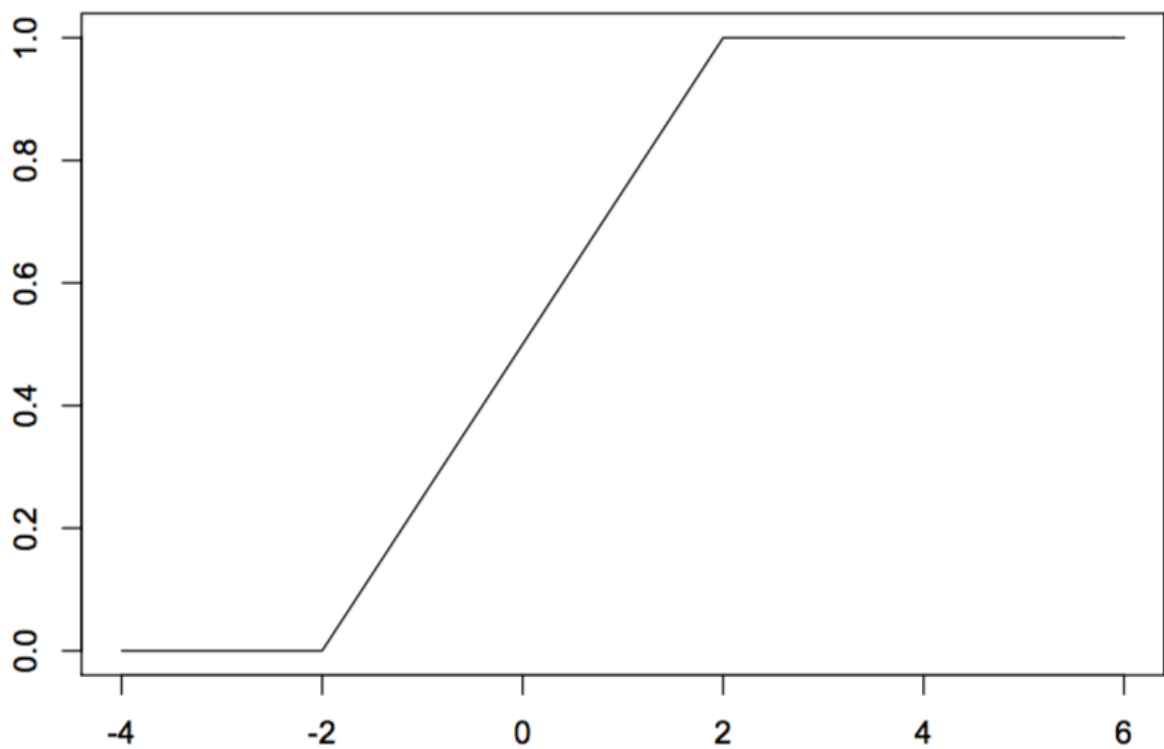
Câu 11. Trong các lệnh sau, lệnh nào có thể vẽ được hình bên dưới.



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)").`

Câu 12. 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. PP chuẩn $\mathcal{N}(1, 1)$.
C. PP mũ $\text{Exp}(2)$.

B. PP đều $U([-2, 2])$.
D. PP Student(10).

Câu 13. Để phát sinh hàm mật độ của biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu = 2, \sigma^2 = 1)$ và mẫu 1000 phần tử là biến ngẫu nhiên có phân phối Poisson $\mathcal{P}(\lambda = 2)$ lần lượt bằng các lệnh và

Đề thi cuối kỳ môn THỰC HÀNH XÁC SUẤT THỐNG KÊ dạng trắc nghiệm có 50 câu hỏi, mỗi câu 5 đáp án. Sinh viên A không học bài, khi đi thi thì xác suất trả lời đúng của mỗi câu là như nhau. Sinh viên B học khá trong lớp, cảm thấy xác suất để mình chọn đúng mỗi câu là 0.6. Đặt

$$X = \sum_{i=1}^{50} X_i \text{ và } Y = \sum_{i=1}^{50} Y_i$$

trong đó, X_i, Y_i lần lượt là các biến ngẫu nhiên phản ánh kết quả chọn câu thứ i là đúng của sinh viên A và sinh viên B (với $i = 1, \dots, 50$), trả lời các câu hỏi 13 và 14:

Câu 14. X và Y có phân phối gì?

- | | |
|---------------------------------------------------|---------------------------------------------------|
| A. Nhị thức $B(50, 0.2)$ và $B(50, 0.4)$. | B. Nhị thức $B(50, 0.2)$ và $B(50, 0.6)$. |
| C. Nhị thức $B(50, 0.8)$ và $B(50, 0.6)$. | D. Nhị thức $B(50, 0.8)$ và $B(50, 0.4)$. |

Câu 15. Sinh viên sẽ qua môn nếu trả lời đúng ít nhất 25 câu hỏi. Xác suất sinh viên A không qua môn; được xấp xỉ bằng câu lệnh nào sau đây,

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

Câu 16. Các cuộc gọi điện đến tổng đài tuân theo phân phối Poisson với mức λ trên mỗi phút. Từ kinh nghiệm có được trong quá khứ, ta biết rằng xác suất nhận được chính xác một cuộc gọi trong một phút bằng ba lần xác suất không nhận được cuộc gọi nào trong cùng thời gian. 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 17. 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 18. Hàm KTC_mean cho biết

- A. Input các tham số dữ liệu mẫu (**data**), độ tin cậy (**alpha**) và giả thiết về **sigma** .
- B. Output là khoảng tin cậy của trung bình trong các trường hợp biết phương sai, không biết phương sai và cỡ mẫu.
- C. A, B đều sai .
- D. A, B đều đúng.

Câu 19.

```
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 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 20. 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 21. Kết quả của lệnh `test$statistic` bằng

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

Câu 22. 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

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 23. 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

Câu 24. 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 25. Cho mẫu X , dùng những hàm có sẵn hãy viết các đoạn lệnh thực hiện *trung bình mẫu* :, *phương sai mẫu* :, *trung vị mẫu* và *độ lệch chuẩn (mẫu)*

Câu 26. Để tính p -giá trị của $Y = 18$ với $Y \sim \mathcal{B}(50, 0.5)$ với đối thuyết $H_1 : p \neq p_0$, hãy hoàn thành đoạn code sau

```
z_0 = .....
p.value = 2*min(pnorm(...), 1 - .....(z_0))
```


Xem đoạn lệnh sau và kết quả của nó để trả lời các câu hỏi 27 và 28.

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

Call:

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

Coefficients:

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

Câu 27. Đoạn lệnh trên cho biết giá trị các hệ số ước lượng $\hat{\beta}_0$ và $\hat{\beta}_1$ lần lượt bằng và

Câu 28. 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 29. Để tìm tính khoảng tin cậy 99% cho β_0 , hãy hoàn chỉnh vào đoạn lệnh sau , biết rằng $\beta_0 \in \left[\hat{\beta}_0 - t_{1-\alpha/2}^{n-2} \sqrt{MSE \left(\frac{1}{n} + \frac{\bar{x}}{Sxx} \right)}; \hat{\beta}_0 + t_{1-\alpha/2}^{n-2} \sqrt{MSE \left(\frac{1}{n} + \frac{\bar{x}}{Sxx} \right)} \right]$ trong đó $\hat{\beta}_0$ là hệ số góc trong mô hình hồi quy và MSE, SSE, Sxx thỏa các công thức sau

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

```
beta0_CI <- function(x, y, .....){
  n = length(x); x.mean = mean(x)
  result = lm(y~x)
  res = resid(result)
  beta0.hat = (coef(result))[[1]]
  MSE = .....
  Sxx = sum((x-x.mean)^...)
  eps = qt(1 - alpha/2, df=n-2)*sqrt(.....*(1/n + x.mean**2/.....))
  print('KTC cho beta 0')
  return(c(beta0.hat - eps, beta0.hat + eps))
}
beta0_CI(x, y, 0.01)
```

Câu 30. 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.

HẾT.