

Sử dụng thông tin bên dưới để trả lời câu hỏi 1.

Đề thi cuối kỳ môn ABC gồm 2 phần. Phần 1 gồm 15 câu trắc nghiệm, mỗi câu có 4 phương án lựa chọn trong đó chỉ có 1 lựa chọn đúng. Phần 2 gồm 15 câu điền khuyết, xác suất sinh viên trả lời đúng 1 câu điền khuyết là 70%. Đặt X, Y lần lượt là số câu trả lời đúng phần trắc nghiệm và phần tự luận của 1 sinh viên.

Câu 1. X và Y có phân phối tương ứng là

- ☒ A. Nhị thức $B(15;0.25)$ và $B(15,0.7)$
- ☐ B. Nhị thức $B(15,0.25)$ và $B(15,0.34)$
- ☐ C. Nhị thức $B(15,0.75)$ và $B(15,0.3)$
- ☐ D. Nhị thức $B(15,0.75)$ và $B(15,0.7)$

Câu 2. Trong các lệnh sau đây, lệnh nào dùng để vẽ hàm mật độ xác suất của biến ngẫu nhiên X biết $X \sim U([-5;5])$.

- ☒ A. `curve(dunif(x, -5, 5), from = -10, to = 10)`
- ☐ B. `line(seq(-5,5,0.01), dunif(seq(-5,5,0.01)))`
- ☐ C. `curve(punif(x, -5, 5), from = -10, to = 10)`
- ☐ D. `plot(seq(-5,5,0.01), punif(seq(-5,5,0.01)))`

Câu 3. Trung bình mỗi phút có 10 cuộc gọi đến tổng đài của trung tâm đặt ngoại ngữ. Gọi X là số cuộc gọi đến tổng đài đó trong 1 phút. Người ta muốn biết xác suất có ít nhất 1 cuộc gọi và không quá 15 cuộc gọi trong 1 phút, câu lệnh nào dưới đây có thể mô tả bài toán đó.

- ☒ A. `ppois(15,10) - ppois(0,10)`
- ☐ B. `dpois(15,10) - dpois(0,10)`
- ☐ C. `dpois(15,10) - dpois(1,10)`
- ☐ D. `ppois(15,10) - ppois(1,10)`

Sử dụng thông tin sau để trả lời câu 4-5

Doanh thu bán hàng của một công ty được lưu trong biến profit. Một ngày có doanh thu trên 75 triệu đồng được xem là ngày có doanh thu cao. Có ý kiến cho rằng tỷ lệ ngày có doanh thu cao không quá 40%. Với độ tin cậy 97%, hãy cho nhận xét về ý kiến đó.

Câu 4. Đối thuyết cần kiểm định của bài toán trên là:

- ☐ A. $H_1: \mu > 90$
- ☒ B. $H_1: p < 0.4$
- ☐ C. $H_1: p > 0.4$
- ☐ D. $H_1: \mu < 90$

Câu 5. Với đoạn lệnh sau, có bao nhiêu câu lệnh có thể dùng để làm bài toán kiểm định trên:

```
x=length(subset(profit,profit>75))
n=length(profit)
binom.test(x,n,0.4, alternative="greater", conf.level=0.97)
binom.test(x,n,0.4, alternative="less", conf.level=0.97)
prop.test(x,n,0.4, alternative="greater", conf.level=0.97)
prop.test(x,n,0.4, alternative="less", conf.level=0.97) ✓
t.test(profit,75,alternative="greater", conf.level=0.97)
t.test(profit,75,alternative="less",conf.level=0.97)
```

- ☒ A. 1
- ☐ B. 2
- ☐ C. 3
- ☐ D. 4

Câu 6. Hàm `t.test(profit, mu=75)` dùng để:

- ☐ A. Trả ra kết quả điểm định cho trung bình của mẫu với đối thuyết là $\mu \neq 75$ với độ tin cậy $1 - \alpha = 0.95$
- ☒ B. Trả ra kết quả điểm định cho trung bình của mẫu với đối thuyết là $\mu = 75$ với độ tin cậy $1 - \alpha = 0.95$
- ☐ C. Trả ra kết quả điểm định cho trung bình của mẫu với đối thuyết là $\mu \neq 75$ với độ tin cậy $1 - \alpha = 0.05$
- ☐ D. Trả ra kết quả điểm định cho trung bình của mẫu với đối thuyết là $\mu = 75$ với độ tin cậy $1 - \alpha = 0.05$

Câu 7- 9 sử dụng thông tin trong đoạn code và kết quả sau,

```
setwd("D://Works//Data")
```

```
df = read.csv("insurance.csv")
x = df$bmi
test = t.test(x, alternative = "two.sided", mu = 29, conf.level = 0.95)
test
```

One Sample t-test

```
data: x
t = 9.9775, df = 1337, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 29
95 percent confidence interval:
 30.33635 30.99045
sample estimates:
mean of x
 30.6634
```

Câu 7. Để trích ra p -giá trị trong kiểm định trên, ta có thể trích nó bằng lệnh

- A. test\$p.value B. p.vlaue C. t.test\$p.valu D. p_value

Câu 8. Kết quả của lệnh `test$statistic` bằng

- A. 9.9775 B. 2.2e-16 C. 1337 D. 29

Câu 9. Chọn phát biểu sai trong các phát biểu sau về đoạn code trên

- A. Đoạn code trên dùng để kiểm định hai phía cho tham số trung bình $H_1: \mu \neq 29$ ✓
 B. Khoảng tin cậy 95% cho trung bình là [30.33635 30.99045] ✓
 C. BMI trung bình khác 29 vì $p.value < 0.05$.
 D. Bác bỏ giả thuyết không, nghĩa là BMI trung bình là 29.

Câu 10. Kết quả của câu lệnh `x%*%x` biết x được tạo bởi: `x=matrix(c(3,-2,1,-1),nrow=2,byrow=T);`

- A. $\begin{bmatrix} 9 & -4 \\ 1 & -1 \end{bmatrix}$ B. $\begin{bmatrix} 7 & -4 \\ 2 & -1 \end{bmatrix}$ C. $\begin{bmatrix} 9 & 1 \\ -4 & -1 \end{bmatrix}$ D. $\begin{bmatrix} 7 & 2 \\ -4 & -1 \end{bmatrix}$

Sử dụng thông tin trong đoạn code sau để trả lời câu 11-12

```
binom.test(c(200,250),c(400,500),alternative="greater",conf.level=0.99)
```

Câu 11. Cỡ mẫu là

- A. $n_1 = 400; n_2 = 500$ B. $n_1 = 200; n_2 = 250$
 B. $n_1 = 200; n_2 = 400$ D. Chưa biết cỡ mẫu.

Câu 12. Tìm phát biểu đúng

- A. Câu lệnh trên dùng để kiểm định một phía trên trái cho hai tỷ lệ của hai tổng thể độc lập.
 B. Câu lệnh trên dùng để kiểm định một phía trên phải cho hai tỷ lệ của hai tổng thể ~~không~~ độc lập.
 C. Câu lệnh trên dùng để kiểm định một phía trên phải cho hai tỷ lệ với mức ý nghĩa 0.99
 D. Câu lệnh trên dùng để kiểm định một phía trên phải cho hai tỷ lệ của hai tổng thể độc lập.

Câu 13. Cho đoạn code sau

```
KTC_mean <- function(data, alpha, sig = 'None'){
  n = length(data)
  m = mean(data)
  std = sd(data)
  zalp = qnorm(1 - alpha/2)
  talp = qt(1 - alpha/2, n-1)
  if(sig != 'None') eps = sig*zalp/sqrt(n)
  else if(sig == 'None')
    eps = std*talp/sqrt(n)
  return(c(m+eps, m - eps))
}
```

Hàm KTC_mean cho biết

- A. Input các tham số dữ liệu mẫu (data), mức ý nghĩa (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. ✓

- C. A, B đều sai.
☒ D. A, B đều đúng.

Lưu ý: Sinh viên thay giá trị tham số m, n, o tương ứng.

m : là chữ số cuối cùng của MSSV; n : là tháng sinh; t : là ngày sinh.

Nếu chữ số cuối là 0 thì $m=5$.

Câu 14. Viết câu lệnh tạo ra một dãy số ngẫu nhiên gồm 11 chữ số có giá trị từ 0 đến 20.

Cho đoạn code sau

```
x <- c(rep(1,2),rep(2,3),rep(3,4),rep(4,5),rep(5,6))
```

```
A <- length(x) 20
```

```
B <- mean(x) 3,5
```

```
C <- median(x) 4
```

Sử dụng đoạn code trên trả lời các câu hỏi 15-16

Câu 15. Giá trị của $A*B-C$ là 65

Câu 16. Giá trị của $A\%m$ là chia dư

Câu 17. Để phát sinh mẫu ngẫu nhiên gồm 100 phần tử từ phân phối mũ có trung bình bằng 2 là

$rnorm(100, \lambda=2)$

Sử dụng thông tin sau để trả lời câu 18-19:

Thời gian hoàn thành bài thi XSTK của sinh viên là biến ngẫu nhiên có phân phối chuẩn với trung bình là 45 phút và độ lệch chuẩn là 5 phút.

Câu 18. Câu lệnh $pnorm(50, 45, 5) - pnorm(40, 45, 5)$ là xác suất sinh viên hoàn thành bài thi trong khoảng thời gian từ 40 đến 50 phút.

Câu 19. Chọn ngẫu nhiên 25 sinh viên. Câu lệnh $binom(n, 25, p)$ dùng để tính xác suất có nhiều nhất n sinh viên hoàn thành bài thi trong khoảng thời gian từ 40 đến 50 phút.

Sử dụng thông tin sau trả lời câu 20-21

Cho biến ngẫu nhiên liên tục X có hàm mật độ xác suất như sau:

$$f(x) = \begin{cases} \frac{1}{2t}, & x \in [-t; t] \\ 0, & x \notin [-t; t] \end{cases}$$

Câu 20. Câu lệnh $curve()$ để vẽ hàm mật độ xác suất của X .

Câu 21. Phân vị 25% của X :

Câu 22. Viết lệnh xuất ra 1 data frame gồm cột chiều cao, bán kính và thể tích của khối nón. Biết rằng, chiều cao gấp đôi bán kính và bán kính $R=c(3,4,5)$.

$h = 2R$
 $V = \pi R^2 h$
 $df = data.frame(h, R, V)$

Ka hie

Câu 23. Câu lệnh `gl(3, 5, 20)` trả ra kết quả là

Sử dụng đoạn code sau để trả lời các câu hỏi 24-26

```
x <- data.frame(letter=LETTERS[2:5])
x <- data.frame(x, thutu=2:5)
y <- data.frame(letter=LETTERS[4:8])
y <- data.frame(y, thutu=4:8)
> union(x$letter, y$letter)
> is.element(c("B", "D", "F" ), y$letter)
> is.element(x$thutu, y$thutu)
```

Câu 24. Kết quả của câu lệnh `union(x$letter, y$letter)` là

Câu 25. Kết quả của câu lệnh `is.element(c("B", "D", "F"), y$letter)` là

Câu 26. Kết quả của câu lệnh `is.element(x$thutu, y$thutu)` là

Câu 27. Giá một căn nhà (đv: 1000 USD) phụ thuộc vào số phòng ngủ trong căn nhà đó. Giả sử rằng dữ liệu sau được ghi lại cho các căn nhà ở một thành phố

Price	300	250	400	550	317	389	425	289	389	559
No.bedrooms	3	3	4	5	4	3	6	3	4	5

Vẽ đồ thị hàm phân tán và đường hồi quy lên cùng một hệ trục tọa độ.

.....

.....

.....

.....

.....

Câu 28.

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

Coefficients:

(Intercept) x

210.0485 -0.7977

Với kết quả trên, độ dốc của đường hồi quy là và tung độ góc là

Câu 29. File `Diemthi.xlsx` chứa cột `DiemGK` và cột `DiemCK`. Có ý kiến cho rằng `DiemCK` thấp hơn `DiemGK` với mức ý nghĩa 1%.Viết đoạn lệnh xuất ra kết quả ý kiến trên có đáng tin hay không. (Sử dụng hàm có sẵn trong R.)

.....

.....

.....

Câu 30. File `diesel_engine.dat` chứa số liệu về hoạt động của động cơ chạy bằng dầu diesel. Viết câu lệnh để xác xác định có bao nhiêu dữ liệu khuyết trong `diesel-engine`.

.....

.....