

1.Cho tập dữ liệu về tuổi của người lao động tại một công ty như sau:

50, 30, 35, 31, 21, 28, 34, 43, 36, 45, 41, 33, 47, 47, 32, 47, 44, 45, 32, 46, 47, 23, 30, 23, 49, 20, 24, 20, 40, 55.

Lập bảng tần số, khi đó tần số tại giá trị 47 là

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

ANSWER: D

2.Cho tập dữ liệu về tuổi của người lao động tại một công ty như sau:

50, 30, 35, 31, 21, 28, 34, 43, 36, 45, 41, 33, 47, 47, 32, 47, 44, 45, 32, 46, 47, 23, 30, 23, 49, 20, 24, 20, 40, 55.

Lập bảng tần suất, khi đó tần suất tại giá trị 32 là

- A. 0.06666666
- B. 0.06666667
- C. 0.06666668
- D. 0.06666669

ANSWER: B

3.Cho tập dữ liệu về tuổi của người lao động tại một công ty như sau:

50, 30, 35, 31, 21, 28, 34, 43, 36, 45, 41, 33, 47, 47, 32, 47, 44, 45, 32, 46, 47, 23, 30, 23, 49, 20, 24, 20, 40, 55.

Phân dữ liệu thành 5 tổ và lập bảng tần số phân tổ, khi đó tần số của (27,34] là

- A. 6
- B. 7
- C. 8
- D. 9

ANSWER: C

1.

```
> x=c(50, 30, 35, 31, 21, 28, 34, 43, 36, 45, 41, 33, 47, 47, 32, 47, 44, 45,
32, 46, 47, 23, 30, 23, 49, 20, 24, 20, 40, 55)
> table(x)
x
20 21 23 24 28 30 31 32 33 34 35 36 40 41 43 44 45 46 47 49 50 55
 2  1  2  1  1  2  1  2  1  1  1  1  1  1  1  1  2  1  4  1  1  1
```

2.

```
> prop.table(table(x))
x
      20      21      23      24      28      30      31
32 0.06666667 0.03333333 0.06666667 0.03333333 0.03333333 0.06666667 0.03333333
0.06666667 0.03333333
      34      35      36      40      41      43      44
45 0.03333333 0.03333333 0.03333333 0.03333333 0.03333333 0.03333333 0.03333333
0.06666667 0.03333333
      47      49      50      55
0.13333333 0.03333333 0.03333333 0.03333333
```

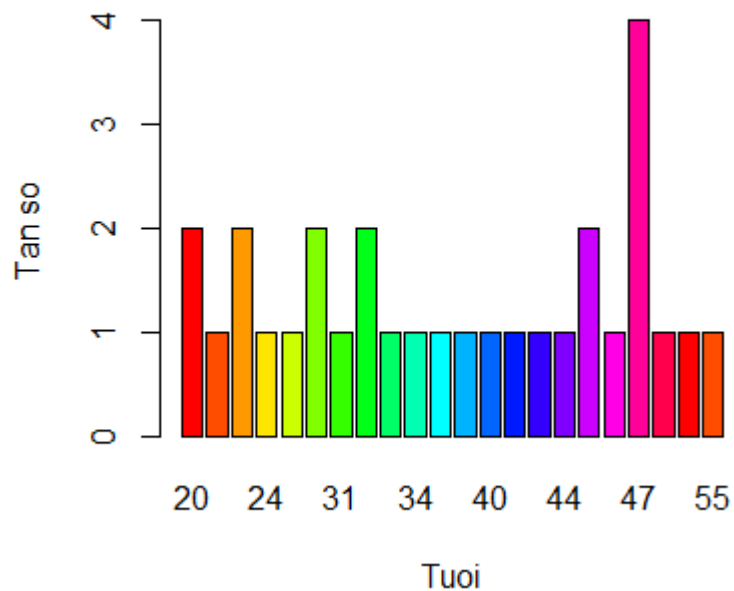
3.

```
> table(cut(x, breaks = 5))
(20, 27] (27, 34] (34, 41] (41, 48] (48, 55]
      6      8      4      9      3
> barplot(table(x), beside=TRUE, col=rainbow(20), xlab="Tuoi ", ylab="Tan so")
```

4. Cho tập dữ liệu về tuổi của người lao động tại một công ty như sau:

50, 30, 35, 31, 21, 28, 34, 43, 36, 45, 41, 33, 47, 47, 32, 47, 44, 45, 32, 46, 47,
23, 30, 23, 49, 20, 24, 20, 40, 55.

Biểu đồ tần số của tập dữ liệu trên là

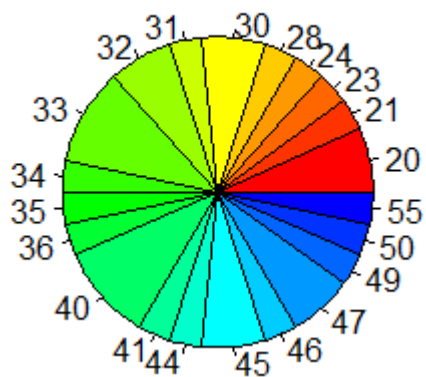


5. Cho tập dữ liệu về tuổi của người lao động tại một công ty như sau:

50, 30, 35, 31, 21, 28, 34, 43, 36, 45, 41, 33, 47, 47, 32, 47, 44, 45, 32, 46, 47, 23, 30, 23, 49, 20, 24, 20, 40, 55.

Biểu đồ tần suất của tập dữ liệu trên là

> `pie(table(x), col = rainbow(30))`



Câu 6. Cho file dữ liệu ChiTieu2010.csv. Phân dữ liệu số người trong hộ thành 6 tổ và lập bảng tần số phân tổ, khi đó số gia đình có số người trong hộ thuộc (5.67,8] là

- A. 1202
- B. 3493
- C. 4610
- D. 2150

ANSWER: A

```
> DL=read.csv("Chi Ti eu2010. csv")
> attach(DL)
```

```
> table(cut(SoNguoi TrongHo, breaks = 6))
```

```
(0. 986, 3. 33]   (3. 33, 5. 67]   (5. 67, 8]   (8, 10. 3]   (10. 3, 12. 7]   (12. 7, 15]
      3493           4619           1202           67           14           3
```

Câu 7. Cho bảng điểm thi THPT Quốc gia môn Vật lý năm học 2017-2018 (xem file DuLieuMau.doc). Tổng số thí sinh và số thí sinh có điểm từ 5 trở lên là

- A. 367864 và 198085
- B. 367865 và 198086
- C. 367866 và 198087
- D. 367867 và 198088

Dữ liệu giả lập

Điểm	2.5	3	5	5.25
Số TS	23	11	3	56

```
> Di em=scan()
1: 2.5 3 5 5.25
5:
```

Read 4 i t e m s

```
> TanSo=scan()
```

```
1: 23 11 3 56
5:
```

Read 4 i t e m s

```
> x=rep(Di em, TanSo)
```

```
> l e n g t h ( x )
```

```
[1] 93
```

```
> l e n g t h ( x [ x >= 5 ] )
```

```
[1] 59
```

Câu 8. Cho bảng điểm thi THPT Quốc gia môn Vật lý năm học 2017-2018 (xem file DuLieuMau.doc). Tỷ phân vị của tập dữ liệu điểm thi THPT Quốc gia môn Vật lý là

- A. 25% 50% 75%

4.00 5.00 6.25
 B. 25% 50% 75%
 4.25 5.00 6.25
 C. 25% 50% 75%
 4.00 5.25 6.25
 D. 25% 50% 75%
 5.00 5.00 6.25

Dữ liệu giả lập(Câu 4)

```
> quantile(x, probs = c(0.25, 0.50, 0.75))
 25%  50%  75%
3.00  5.25  5.25
```

CÂU 9. Cho file Dulieudiemthi.doc (trong thư mục Sinhvien), file này mô tả điểm thi môn Vật lý của kì thi THPT quốc gia năm 2017-2018. Tỷ lệ thí sinh có điểm thi dưới trung bình

Dữ liệu giả lập(Câu 4)

```
> length(x[x<5])
[1] 34
> length(x)
[1] 93
#Ta tìm được x=34, n=93
#Tỷ lệ = x/n
> 34/93
[1] 0.3655914
```

CÂU 10. Khoảng tin cậy 95% cho tỷ lệ điểm dưới trung bình

Dữ liệu giả lập(Câu 4)

```
> prop.test(34, 93, alternative = "t", conf.level = 0.95, correct = FALSE)
```

1-sample proportions test without continuity correction

```
data: 34 out of 93, null probability 0.5
X-squared = 6.7204, df = 1, p-value = 0.009531
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.2748570 0.4669891
sample estimates:
p
0.3655914
```

CÂU 11. Kiểm định gt tỷ lệ điểm dưới trung bình lớn hơn 0.35, mức ý nghĩa 0.05

Dữ liệu giả lập(Câu 4)

```
> prop.test(34, 93, p=0.35, alternative = "g", conf.level = 0.95, correct = FALSE)
```

1-sample proportions test without continuity correction

```
data: 34 out of 93, null probability 0.35
```

```
X-squared = 0.099374, df = 1, p-value = 0.3763
alternative hypothesis: true p is greater than 0.35
95 percent confidence interval:
 0.2883288 1.0000000
sample estimates:
      p
0.3655914
```

CÂU 12. Cho file Chitieu2010.csv (trong thư mục Sinhvien), khoảng tin cậy 99% cho chi phí y tế trung bình của các hộ gia đình ở khu vực 1 là

```
> DL=read.csv("Chi Ti eu2010.csv")
> attach(DL)
> x=Chi Ti euYTe[KhuVuc==1]
> t.test(x, alternative="t", conf.level = 0.99)
```

One Sample t-test

```
data: x
t = 25.024, df = 2646, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 268.8778 330.6326
sample estimates:
mean of x
299.7552
```

CÂU 13. Cho file Chitieu2010.csv (trong thư mục Sinhvien), kiểm định gt chi tiêu y tế trung bình của tổng thể các hộ là nhỏ hơn 250, mức ý nghĩa 0.05

```
> y=Chi Ti euYTe
> t.test(y, mu=250, alternative="l", conf.level = 0.95)
```

One Sample t-test

```
data: y
t = -1.3225, df = 9397, p-value = 0.09301
alternative hypothesis: true mean is less than 250
95 percent confidence interval:
 -Inf 252.0287
sample estimates:
mean of x
241.6802
```

CÂU 14.

Dữ liệu giả lập

Điểm	2.5	3	5	5.25
Số TS	23	11	3	56

Kiểm định gt điểm trung bình là 4, mức ý nghĩa 0.05

```
> Di em=scan()
1: 2.5 3 5 5.25
5:
Read 4 items
> TanSo=scan()
1: 23 11 3 56
5:
Read 4 items
```

```
> x=rep(Diem, TanSo)
> t.test(x, mu=4, alternative="t", conf.level = 0.95)
```

One Sample t-test

```
data: x
t = 2.2705, df = 92, p-value = 0.02551
alternative hypothesis: true mean is not equal to 4
95 percent confidence interval:
 4.037042 4.554356
sample estimates:
mean of x
 4.295699
```

Bài tập bổ sung: Cho file điểm thi KT1.csv

+ Thống kê số điểm và tính tỷ lệ điểm A, B, C, D của điểm DQT, THI, TKHP và vẽ các biểu đồ trên

+ Tìm khoảng tin cậy 96% cho điểm trung bình của DQT, THI, TKHP, DQT-THI, THI-TKHP, cho hiệu 2 tỷ lệ điểm trên trung bình giữa DQT và THI.

+ Gọi x là điểm THI, y là điểm TKHP; trong mô hình HQT của y với x, trả lời tất cả các câu hỏi như trong bài tập thực hành hồi quy tuyến tính đơn.

+ Gọi x1 là DQT, x2 là THI, y là TKHP; trong mô hình HQT của y với x1 và x2, trả lời tất cả các câu hỏi như trong bài tập thực hành HQT đa biến.

+ Với mức ý nghĩa 95% có thể xem trung bình điểm THI lớn hơn 6 hay không? Trung bình điểm THI lớn hơn TKHP hay không?

HD:

#Cách 1: Đọc dữ liệu từ file excel

(Mở file excel KT1.xls, chọn "Save As", "Save as type" chọn lưu file định dạng đuôi .CSV(comma delimited)).

Copy file KT1.csv vào thư mục R đang làm việc (mặc định là Documents)

```
> Diem=read.csv("KT1.csv")
> attach(Diem)
> x=DQT
> c=cut(x, breaks = c(0, 3.9, 5.4, 6.9, 8.4, 10), include.lowest = T, labels =
c("F", "D", "C", "B", "A"))
> table(c)
C
 F  D  C  B  A
 5 12 37 59 30
Số SV đạt điểm A là 30,...

> prop.table(table(c))
C
```

```

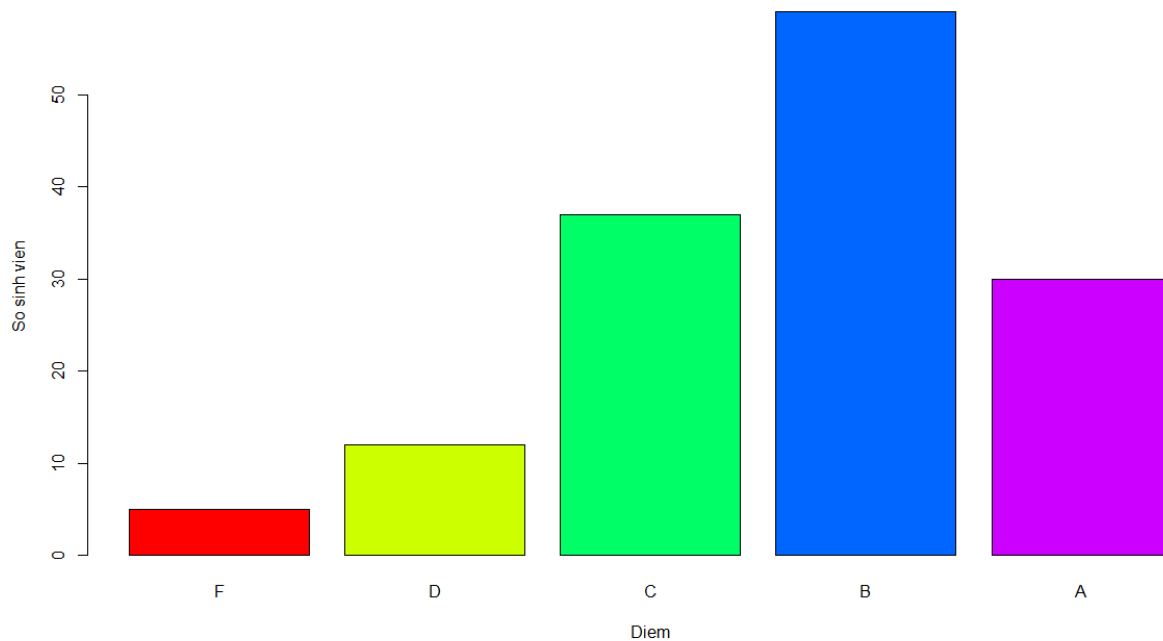
          F          D          C          B          A
0. 03496503 0. 08391608 0. 25874126 0. 41258741 0. 20979021
Tỷ lệ điểm A là 0. 20979021, . . .

```

```
#####
```

Biểu đồ cột phổ điểm

```
> barplot(table(c), beside=TRUE, col =rainbow(5), xlab="Điểm", ylab="Số sinh viên")
```



Biểu đồ tròn tần suất

```
> pie(table(c), col =rainbow(5))
```