1 Hãy cho biết có bao nhiêu loại xu hướng tính dục trong tập dữ liệu 5

2Hãy cho biết phân phối của biến thời gian học hằng ngày lệch phải hay trái lech trai

3Dựa trên biểu đồ Pie. Hãy cho biết tỉ lệ phần trăm số lượng sinh viên trên nhóm tính dục mà đạt tuyển dụng

a.

[T,G,B,L,Q] = [20.2,13.9,14.5,8.6,42.7]

b.

[T,Q, B,L,G] = [20.2,13.9,14.5,8.6,42.7]

c.

[T,Q, L,B,G] = [20.2,13.9,14.5,8.6,42.7]

d.

[B,G,L,T,Q] = [20.2,13.9,14.5,8.6,42.7]

e.

[G,L,B,T,Q] = [20.2,13.9,14.5,8.6,42.7]

f.

[L,G,B,T,Q] = [20.2,13.9,14.5,8.6,42.7]

g.

[Q,G,B,T,L] = [20.2,13.9,14.5,8.6,42.7]

h.

[T,Q, L,G,B] = [20.2,13.9,14.5,8.6,42.7]

4Hãy cho biết kết quả phỏng vấn có phụ thuộc vào nhóm xu hướng tính dục hay không

**5**Giá trị lớn nhất trong bảng Crosstab giữa xu hướng tính dục và điểm phỏng vấn là bao nhiêu

a.

200

b.

5

c.

122

d.

120

e.

84

f.

126

g.

24

h.

64

**6**Dựa vào biểu đồ Box-plot cho biết dữ liệu điểm đánh giá năng lực có xuất hiện giá trị nghi ngờ là bất thường hay không

7 Căn cứ vào biểu đồ boxplot nếu cần sử dụng giá trị thay thế cho giá trị điểm đánh giá năng lực bị thiếu (dựa trên các nhóm xu hướng tính dục) thì có thể dùng phương án thay bằng giá trị trung vị toàn cục hay không?

a.

Thay thế bằng hằng số

b.

Remove missing data

c.

Dùng giá trị trung vị toàn cục thay thế được

d.

Không thể dùng giá trị trung vị toàn cục thay thế được

8Dựa vào ma trận hiệp phương sai, hãy cho biết tương quan giữa các cặp thời gian học, thời gian chơi và điểm đánh giá năng lực thì cặp nào tương quan nào cao nhất

**TimeToPlay ,** ProfessionalScore

9Hãy cho biết phân nửa sinh viên có thời gian dành cho việc học hằng ngày lớn hơn bao nhiêu giờ

a.

2.900000

b.

5.064000

c.

**5.200000**

d.

10.000000

e.

5.077300

f.

2.843235

g.

7.400000

10Hãy cho biết điểm điểm đánh giá năng lực có phụ thuộc vào nhóm xu hướng tính dục hay không

a.

Phụ thuộc

b.

Độc lập

c.

Phụ thuộc yếu

d.

Rất phụ thuộc

11Dựa trên mô hình hồi quy tuyến tính phân tích ảnh hưởng của thời gian nghỉ và thời gian tự học tác động đến điểm đánh giá năng lực. Hãy cho biết nếu thời gian tự học là 5.0 và thời gian nghỉ là 7.0 thì điểm đánh giá năng lực được dự đoán là bao nhiêu

a.

1.7752999999999998

b.

2.9752999999999998

c.

3.9752999999999998

d.

1.9752999999999998

12Dựa trên biểu đồ Pie. Hãy cho biết tỉ lệ phần trăm các nhóm giới tính trong dữ liệu

a.

[L,G,B,T,Q] = [20.2,14.7,13.3,10.5,41.3]

b.

[B,G,L,T,Q] = [20.2,14.7,13.3,10.5,41.3]

c.

[G,L,B,T,Q] = [20.2,14.7,13.3,10.5,41.3]

d.

[L,G,T,B,Q] = [20.2,14.7,13.3,10.5,41.3]

e.

[Q,G,B,T,L] = [20.2,14.7,13.3,10.5,41.3]

13Dựa vào bảng thống kê mô tả hãy cho biết cận trên (upper bound) của điểm đánh giá năng lực là bao nhiêu

a.

2.9

b.

9.05

c.

1.0

d.

8.05

e.

10.05

f.

5.75

g.

7.05

14Dựa vào phân tích hồi quy tuyến tính OLS phân tích ảnh hưởng của thời gian dành cho viêc riêng, thời gian dành cho việc học đến điểm đánh giá năng lực. Hãy giải thích ý nghĩa thống kê của các tham số trong mô hình

a.

Biến TimeToPlay có ý nghĩa thống kê

b.

Các tham số đều không có ý nghĩa thống kê trong mô hình

c.

Biến TimeToStudy có ý nghĩa thống kê

d.

Các tham số đều có ý nghĩa thống kê trong mô hình

15Dựa vào phân tích hồi quy tuyến tính OLS phân tích ảnh hưởng của thời gian dành cho viêc riêng, thời gian dành cho việc học đến điểm đánh giá năng lực. Hãy giải thích ý nghĩa thống kê của các tham số trong mô hình

a.

Biến TimeToPlay có ý nghĩa thống kê

b.

Các tham số đều không có ý nghĩa thống kê trong mô hình

c.

Biến TimeToStudy có ý nghĩa thống kê

d.

Các tham số đều có ý nghĩa thống kê trong mô hình

16Giữa hai biến số thời gian dành cho việc học và thời gian dành cho việc riêng để so sánh sự ổn định ta dùng đại lượng nào?

a.

variance

b.

standard deviation

c.

correlation

d.

coefficient of variation

17Hãy cho biết có bao nhiêu sinh viên đạt điểm đánh giá năng lực cao nhất

a.

16

b.

14

c.

18

d.

25

e.

17

f.

15

g.

20

h.

19

18Giá trị yếu vị trong bảng Crosstab giữa xu hướng tính dục và điểm phỏng vấn là bao nhiêu

a.

28

b.

29

c.

31

d.

32

e.

30

19Dựa vào đồ thị phân phối dữ liệu của thời gian dành cho việc học với số lượng bin là 5. Hãy cho biết khoảng thời gian nào dành cho việc học có xác suất xảy ra nhiều nhất

a.

[2,4)

b.

[0,2)

c.

[4,6)

d.

[6,8)

e.

[8,10]

20Cho kết quả của pivot table cho trung bình thời gian học theo nhóm xu hướng tính dục căn cứ vào kết quả phỏng vấn. Hãy cho biết giá trị 4.75 là của nhóm nào

a.

G, YES

b.

T, NO

c.

G, NO

d.

L, YES

e.

T, YES

f.

Q, YES

g.

B, NO

h.

B, YES

i.

L, NO

j.

[Q, B], YES

k.

Q, NO

21Hãy cho biết phân nửa sinh viên có thời gian dành cho việc học hằng ngày lớn hơn bao nhiêu giờ

a.

2.900000

b.

5.064000

c.

5.200000

d.

10.000000

e.

5.077300

f.

2.843235

g.

7.400000

22ãy cho biết kết quả phỏng vấn có phụ thuộc vào nhóm xu hướng tính dục hay không

a.

Không phụ thuộc

b.

Có phụ thuộc

c.

Hoàn toàn phụ thuộc

d.

Có phụ thuộc tương đối cao

23Trong mô hình hồi quy tuyến tính OLS phân tích ảnh hưởng của thời gian dành cho việc riêng, thời gian dành cho việc học đến điểm đánh giá năng lực khóa học. Hãy cho biết các biến độc lập tác động đến bao nhiêu phần trăm sự biến thiến của điểm đánh giá năng lực

a.

Xấp xỉ 0%

b.

Xấp xỉ 10%

c.

Tất cả đều sai

d.

Xấp xỉ 100%

24Có đủ dữ kiện để nói rằng trung bình điểm đánh giá năng lực là 2.4575 hay không với mức tin cậy là 95%

a.

Không đủ dữ kiện để kết luận trung bình điểm đánh giá năng lực bằng 2.4575

b.

Có thể kết luận là trung bình điểm đánh giá năng lực nhỏ hơn 2.4575

c.

Đủ dữ kiện để kết luận trung bình điểm đánh giá năng lực bằng 2.4575

d.

Có thể kết luận là trung bình điểm đánh giá năng lực lớn hơn 2.4575

e.

Kết quả giá trị bị lỗi

25Dựa vào biểu đồ cột, hãy cho biết trung bình điểm đánh giá năng lực của nhóm đạt tuyển dụng trên nhóm tính dục nào có khác biệt lớn khi so với nhóm không đạt tuyển dụng

Gợi ý: sns.barplot(x='........', y='..........', hue='Result', data=df)

a.

G

b.

T

c.

Q

d.

B

e.

L