

CẤU TRÚC RỜI RẠC CHO KHMT (CO1007)

ĐẠI HỌC BÁCH KHOA THÀNH PHỐ HỒ CHÍ MINH
KHOA KHOA HỌC VÀ KỸ THUẬT MÁY TÍNH

Ứng dụng thống kê phân tích kết quả việc thực
tập của sinh viên

Nhóm 8

Ngày 11 tháng 12 năm 2020

Nội dung báo cáo

- 1 Lí thuyết
- 2 Công thức
- 3 Code R và Kết quả
- 4 Tài liệu tham khảo

- Tính tổng điểm, trung bình cộng, tìm số lớn nhất, tìm số nhỏ nhất.
- Tính tần số tương đối tích lũy.

Tính trung bình cộng các tiêu chí sinh viên

Bước 1: Tổng điểm của tiêu chí.

$$\text{Sum TC1} = \text{ScoreSV1.TC1} + \text{ScoreSv2.TC1} + \dots + \text{ScoreSVn.TC1}$$

$$\text{Sum TCn} = \text{ScoreSV1.TCn} + \text{ScoreSv2.TCn} + \dots + \text{ScoreSVn.TCn}$$

Bước 2: Tổng số sinh viên trên tiêu chí.

$$\text{Sum Amount} = n.$$

Bước 3: Trung bình cộng:

$$\text{TC1} = \text{Sum TC1} / \text{Sum Amount}$$

$$\text{TCn} = \text{Sum TCn} / \text{Sum Amount}$$

Tính điểm trung bình cho mỗi sinh viên

Bước 1: Tổng điểm của mỗi sinh viên.

$$\text{Sum SV1} = \text{Score Sv1} \cdot \text{TC1} + \text{Score Sv1} \cdot \text{TC2} + \dots + \text{Score SV1} \cdot \text{TCn}$$

$$\text{Sum SVM} = \text{Score SVM} \cdot \text{TC1} + \text{Score SVM} \cdot \text{TC2} + \dots + \text{Score SVM} \cdot \text{TCn}$$

Bước 2: Tổng số tiêu chí.

$$\text{Sum TC} = n.$$

Bước 3: Tính điểm trung bình cho mỗi sinh viên:

$$\text{TBC SV1} = \text{Sum SV1} / \text{sum TC}$$

$$\text{TBC SVM} = \text{SUM SVM} / \text{Sum TC}$$

Tính tần số tích lũy theo tiêu chí

- Bước 1: Xác định tổng số lần lặp.
- Bước 2: Tính tần số của từng điểm.
- Bước 3: Tần số tích lũy tương đối.

Điểm	Tần số	Tần số tích lũy tương đối
0	$T0 = \text{Số lần điểm 0} / \text{Tổng số lần lặp}$	$T0$
1	$T1 = \text{Số lần điểm 1} / \text{Tổng số lần lặp}$	$T1 + T0$
2	$T2 = \text{Số lần điểm 2} / \text{Tổng số lần lặp}$	$T2 + T1 + T0$
3	$T3 = \text{Số lần điểm 3} / \text{Tổng số lần lặp}$	$T3 + T2 + T1 + T0$
4	$T4 = \text{Số lần điểm 4} / \text{Tổng số lần lặp}$	$T4 + T3 + T2 + T1 + T0$
5	$T5 = \text{Số lần điểm 5} / \text{Tổng số lần lặp}$	$T5 + T4 + T3 + T2 + T1 + T0 = 100\%$

Tính outlier cho đồ thị dotplot

Để tìm outlier cho đồ thị dotplot ta dùng công thức

$$Lowfence = Q1 - IQR$$

$$Highfence = Q3 - IQR$$

$$IQR = Q3 - Q1$$

Giá trị ngoài lowfence và highfence của đồ thị là outlier.
IQR (interquartile range) gọi là độ trải giữa.




Tính trung bình cộng các tiêu chí sinh viên.

```
>data2<-read_excel("C:/BTL/file/BTLCTRR.xlsx",sheet =  
"TONG HOP")  
i<-1:196  
k<-2:27  
cau2<-data.frame(data2[i,k])  
a2=array(dim=26)  
i<-1:196  
for(m in 1:26) a2[m]<-signif(sum(cau2[i,m])/196,digits = 3)  
>q2<-data.frame("TBC"=a2)  
>q2
```


	TBC
1	2.82
2	2.88
3	2.6
4	2.28
5	2.37
6	2.68
7	2.68
8	2.62
9	2.43

Tính trung bình cộng điểm đánh giá sinh viên theo từng tiêu chí.

```
cau3<-read_excel("C:/BTL/file/BTLCTRR.xlsx",sheet = "TONG  
HOP")  
dulieu <- data.frame(cau3[1:196,2:27])  
sum= c(rowSums(dulieu))  
SV = data.frame(cau3[,1])  
colnames(SV) <- c("SV")  
Trung_Binh_SV = data.frame(SV, GTTB = round(sum/26,digits  
= 2))  
view(Trung_Binh_SV)
```

	SV 	GTTB 
1	P1.1	2.15
2	P1.2	2.50
3	P1.3	2.38
4	P1.4	2.58
5	P1.5	2.54
6	P1.6	2.69
7	P1.7	2.73
8	P1.8	2.00
9	P1.9	2.31

Danh sách tiêu chí mà tổng điểm sinh viên là nhỏ nhất

```
cau6<-data.frame(data6[i,k])
b6=array(dim=26)
for(m in 1:26) b6[m]<-sum(cau6[i,m])
min=600;
for(i in 1:26) if(b6[i]>min){min<-min}else{min<-b6[i]}
min
for(i in 1:26) if(b6[i]==min){print(i)}
> min
[1] 398
> for(i in 1:26) if(b6[i]==min){print(i)}
[1] 11
```

Liệt kê tổng điểm các tiêu chí và số lượng sinh viên mà tổng điểm các tiêu chí là bằng nhau

	tongdiem_tc	soluong_sv
1	56	1
2	65	10
3	62	12
4	67	12
5	66	7
6	70	5
7	71	7

Tính minimum, maximum, median, tứ phân vị (quartile) thứ nhất (Q1) và thứ ba (Q3) cho 2 tiêu chí trên của mẫu.

```
ctr_r_c9 <- read.xlsx("C:/BTL/file/BTLCTRR.xlsx", sheetName  
="TONG HOP")  
min(ctr_r_c9[,7])  
max(ctr_r_c9[,7])  
median(ctr_r_c9[,7])  
quantile(ctr_r_c9[,7])  
min(ctr_r_c9[,17])  
max(ctr_r_c9[,17])  
median(ctr_r_c9[,17])  
quantile(ctr_r_c9[,17])
```

```
> min(ctr[,7])  
[1] 0  
> max(ctr[,7])  
[1] 5  
> median(ctr[,7])  
[1] 3  
> quantile(ctr[,7])  
0%    25%    50%    75%   100%  
0      1      3      4      5
```

Liệt kê danh sách sinh viên mà có tổng điểm các tiêu chí về tính kỷ luật là nhỏ nhất

```
kyluat_c17 <- read.xlsx("C:/BTL/file/BTLCTRR.xlsx",  
sheetName = "KL")  
tong_kyluat_c17 <- rowSums(kyluat_c17[,2:4])  
sv_c17 <- kyluat_c17[,1]  
thongke_kyluat_c17 <- data.frame(sv_c17,tong_kyluat_c17)  
tk_nho_c17 <- thongke_kyluat_c17  
[order(thongke_kyluat_c17$tong_kyluat_c17),]
```

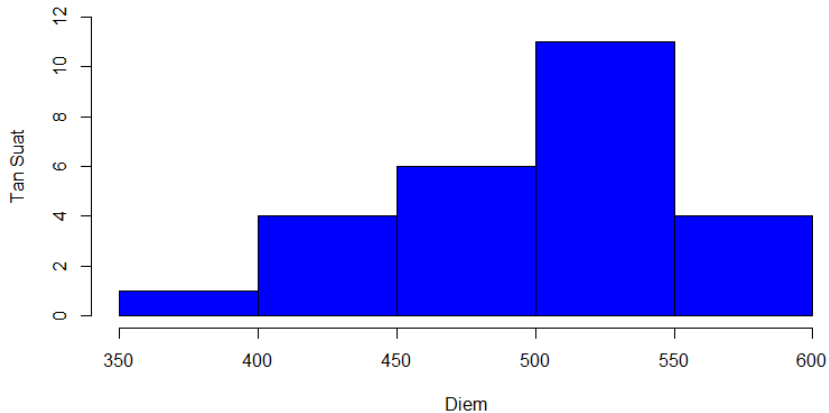

	sv	tong_ky luat
83	P5.27	0
17	P1.17	1
98	P6.7	1
167	P8.9	1
12	P1.12	2
77	P5.21	2
90	P5.34	2
93	P6.2	2
129	P7.11	2
157	P7.39	2
28	P1.28	3

Kiểm tra các cặp tiêu chí sau xem liệu các công ty đã đánh giá tương thích chưa

	V1	V2	V3	V4	V5	V6	V7
1	1	2	3	1	0	0	1
2	2	1	1	1	0	0	1
3	0	1	1	1	2	3	2
4	4	5	1	1	2	1	1
5	5	0	3	2	1	5	0
6	3	0	3	4	0	3	1
7	4	1	2	1	0	2	3
8	1	1	1	2	1	2	2
9	1	2	3	1	1	4	2

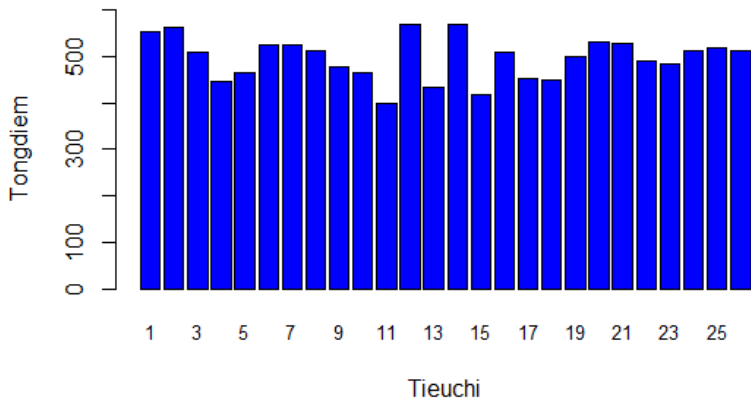
Vẽ biểu đồ hiển thị phân phối tổng điểm các tiêu chí của sinh viên.

```
cau20<-read_excel("C:/BTL/file/BTLCTRR.xlsx",sheet =  
"TONG HOP")  
dulieu <- data.frame(cau20[1:196,2:27])  
diemTongTC = c()  
for (i in 1:26)  
values= sum(dulieu[1:196,i])  
diemTongTC[i] <- values  
hist(diemTongTC,xlab="Diem",ylab="Tan Suat",main="Bieu do  
hien thi phan phoi tong diem cac tieu chi cua  
SV",ylim=c(0,12),col="blue")
```



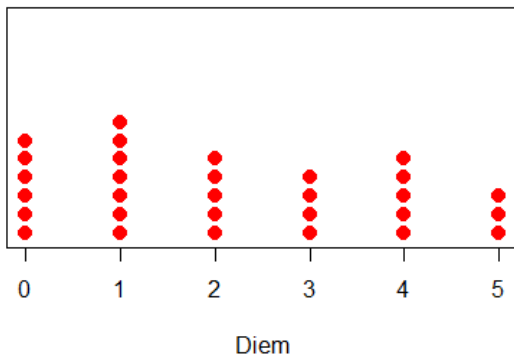
Vẽ phổ tổng điểm sinh viên theo tiêu chí

```
cau22<-data.frame(data22[i,k])  
c=array(dim=26)  
for(m in 1:26) c[m]<-sum(cau22[i,m])  
k<-c(1:26)  
barplot(c,names.arg=k,xlab="Tieuchi",ylab="Tongdiem",ylim=c(0,600),  
col="blue",cex.names = 0.8)
```



Hãy vẽ đồ thị dotplot cho tiêu chí '01'

Sheet E1 - TC01



So sánh phân phối của 2 biến tiêu chí bằng dotplot.

Vẽ đồ thì với Sheet E1.

#Tiêu chí 6

```
> cau24e1<- read.xlsx("C:/BTL/file/BTLCTRR.xlsx",sheet = 1,  
colNames = TRUE)
```

```
> stripchart(cau24e1[,7], method = "stack",  
+ pch = 20, cex = 2 , las=1,at= c(0), offset = 1/5,  
+ main= "SHEET E1 TC6", xlab = "Diem")
```

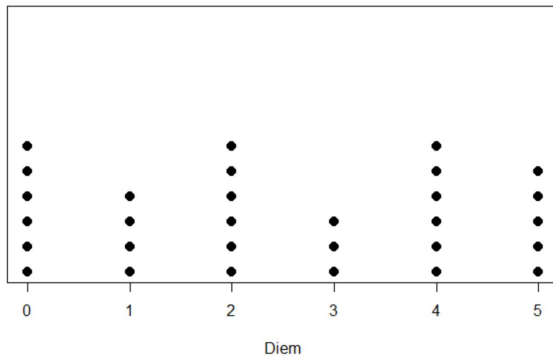

So sánh phân phối của 2 biến tiêu chí bằng dotplot.

Tính giá trị biên trên và dưới của đồ thị để xác định outlier

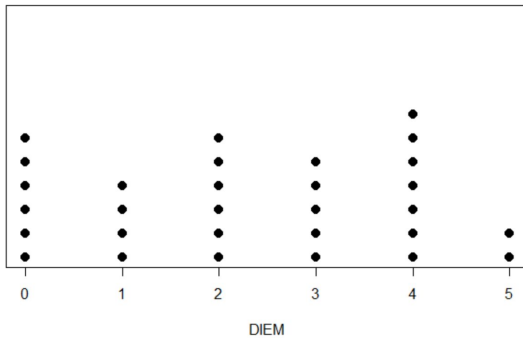
$$lf6e1 = \text{quantile}(\text{cau24e1}[, 7], \text{probs} = 0.25) - 1.5 * \text{IQR}(\text{cau24e1}[, 7])$$

$$hf6e1 = \text{quantile}(\text{cau24e1}[, 7], \text{probs} = 0.75) + 1.5 * \text{IQR}(\text{cau24e1}[, 7])$$

SHEET E1 TC6

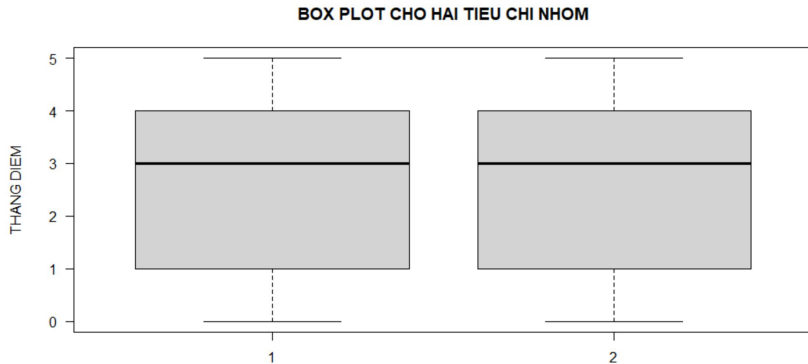


SHEET E1 TC16



Vẽ box-and-whisker cho 2 tiêu chí của nhóm

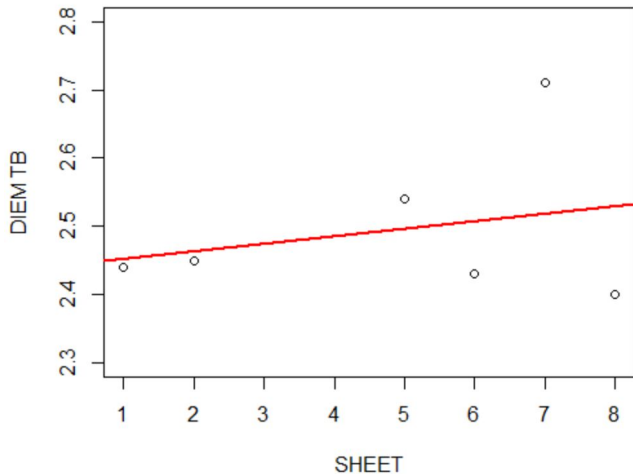
```
> cau25tcn <- read.xlsx("C:/BTL/file/BTLCTRR.xlsx",  
> boxplot(cau25tcn$TC6, cau25tcn$TC16,  
+ main="BOX PLOT CHO HAI TIEU CHI NHOM",  
+ ylab="THANG DIEM", ylim=c(0,5),  
+ las=1, names = TRUE)
```



Vẽ scatterplot cho tổng điểm trung bình gồm các sheets của tập các tiêu chí về kỹ luật

```
> cau28e1 <- read.xlsx("C:/BTL/file/BTLCTRR.xlsx",  
  +sheet = "E1",cols = c(8,16,21), colNames = TRUE)  
> s28e1 = sum(cau28e1)  
> tb28e1 = signif(s28e1/(30*3),digits = 3)  
>tb28=c(tb28e1, tb28e2, tb28e5, tb28e6, tb28e7, tb28e8)  
>tt28=c(1,2,5,6,7,8)  
> plot(x=tt28, y=tb28,ylab = "DIEM TB",  
  + ylim = c(2.3, 2.8),xlab = "SHEET",  
  + main = "SCATTERPLOT TIEU CHI KI LUAT CHO  
TUNG SHEET")
```

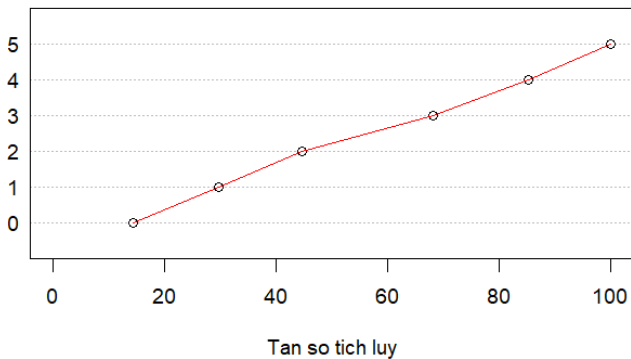
SCATTERPLOT TIÊU CHÍ KI LUAT CHO TUNG SHEET



Vẽ đồ thị tần số tích lũy tương đối trên tập các tiêu chí về kỹ năng mềm.

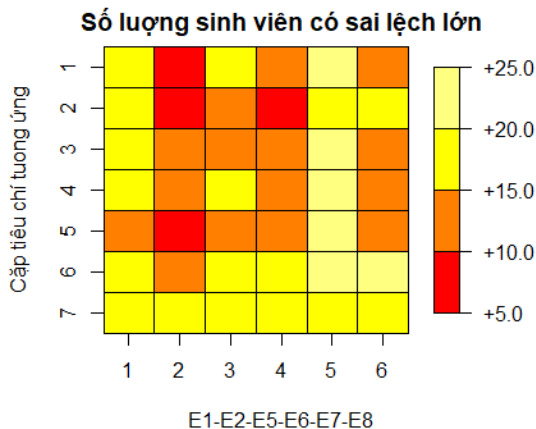
```
>data30<-read_excel("C:/BTL/file/BTLCTRR.xlsx",sheet =
"TONG HOP")
>i<-{1:196}
>k<-{2:27}
>cau30<-data.frame(data30[i,k])
dem0=dem1=dem2=dem3=dem4=dem5=0;
>tc<-c(cau30[,6],cau30[,8],cau30[,9],cau30[,10],cau30[,12],...)
>for (i in 1:588) if(tc[i]==0){dem0=dem0+1}else
if(tc[i]==1){dem1=dem1+1}else if(tc[i]==2){dem2=dem2+1}...
>k<-c(dem0/588*100,(dem0+dem1)/588*100,...)
>dotchart(x=k,labels = 0:5,xlab = "Tanso",cex=1.5,main =
"Bieu do tan so tích lũy theo tiêu chí về kỹ năng mềm")
>lines(x=k,y=1:6,col="red")
```


Biểu đồ tan so tích lũy theo tiêu chí về kỹ năng mềm



Vẽ biểu đồ số lượng sinh viên bị sai lệnh lớn trong đánh giá các cặp tiêu chí tương thích

```
library(plot.matrix)
m_c32 <- matrix(c(slv_e1_c19,slv_e2_c19,slv_e5_c19,slv_e6_c19,slv_e7_c19,
ncol = 6)
plot(m_c32, xlab="E1-E2-E5-E6-E7-E8",ylab="Cặp tiêu chí
tương ứng",main ="Số lượng sinh viên có sai lệch lớn")
```



- [Dal] Dalgaard, P. Introductory Statistics with R. Springer 2008.
- [K-Z] Kenett, R. S. and Zacks, S. Modern Industrial Statistics: with applications in R, MINITAB and JMP, 2nd ed., John Wiley and Sons, 2014.
- [Ker] Kerns, G. J. Introduction to Probability and Statistics Using R, 2nd ed., CRC 2015.

Chân thành cảm ơn mọi người đã lắng nghe

