

MR. HUYNH NAM

EMAIL: giangdayit@gmail.com

FB: www.facebook.com/mrhuynhnam

Website: www.huynhnam.com || www.giangdayit.com

LAB THỰC HÀNH LẬP TRÌNH PHÂN TÍCH DỮ LIỆU

Nội dung: Phân tích dữ liệu `dulieuxettuyendaihoc.csv`

Mục tiêu: Sinh viên đạt được kiến thức

1. Nhập liệu từ file trong pandas
2. Thao tác cơ bản với pandas
3. Thao tác gom nhóm và sử dụng các hàm tổng hợp trong pandas
4. Mô tả thống kê cho dữ liệu định tính
5. Mô tả thống kê cho dữ liệu định lượng

Mô tả dữ liệu

Dữ liệu lưu trữ điểm trung bình môn, khu vực, khối thi và điểm thi đại học của 100 học sinh.

- T1, L1, H1, S1, V1, X1, D1, N1 lần lượt là điểm trung bình các môn Toán, Lý ,Hóa, Sinh, Văn, Sử, Địa, Ngoại ngữ tháng thứ nhất
- T2, L2, H2, S2, V2, X2, D2, N2 lần lượt là điểm trung bình các môn Toán, Lý ,Hóa, Sinh, Văn, Sử, Địa, Ngoại ngữ tháng thứ hai....
- T6, L6, H6, S6, V6, X6, D6, N6 lần lượt là điểm trung bình các môn Toán, Lý ,Hóa, Sinh, Văn, Sử, Địa, Ngoại ngữ tháng thứ sáu
- GT: Giới tính
- DT:
- KV, KT lần lượt là khu vực thi và khối thi
- DH1, DH2, DH3 lần lượt là điểm thi đại học môn 1, môn 2, môn 3

Hướng dẫn

1. Tiến hành phân tích dữ liệu định tính: Xác định dữ liệu định tính (KT, KV....)
 - a. Lập bảng tần số, tần suất, tần số tích lũy và tần suất tích lũy cho biến KT

Kết quả

KT				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	49	49.0	49.0
	A1	6	6.0	55.0
	B	9	9.0	64.0
	C	14	14.0	78.0
	D1	22	22.0	100.0
	Total	100	100.0	100.0

MR. HUYNH NAM

EMAIL: giangdayit@gmail.com

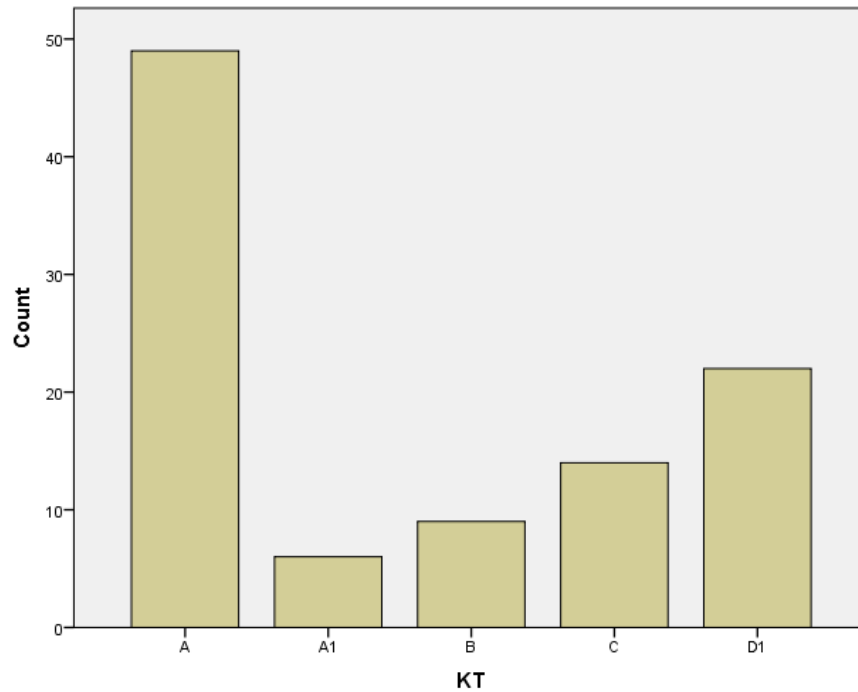
FB: www.facebook.com/mrhuynhnham

Website: www.huynhnham.com || www.giangdayit.com

b. Vẽ biểu đồ cột

i. Vẽ biểu đồ thể hiện tần số của biến KT

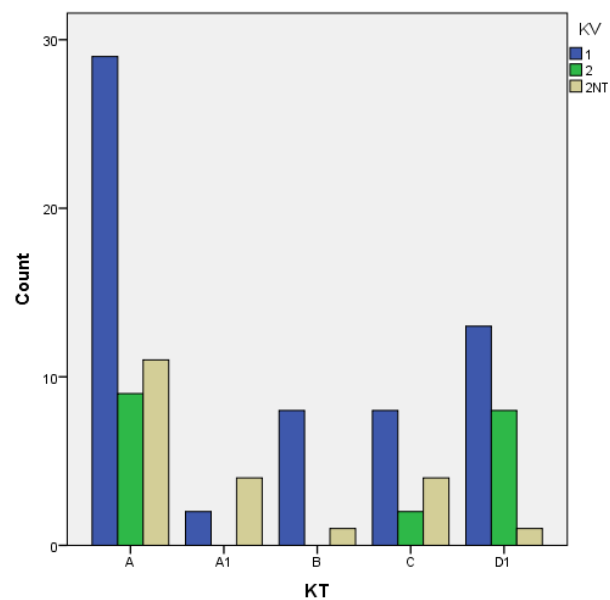
Kết quả



ii. Vẽ biểu đồ cột cho biến KT chia nhóm bởi biến KV

1. Biểu đồ dạng Clustered

Kết quả



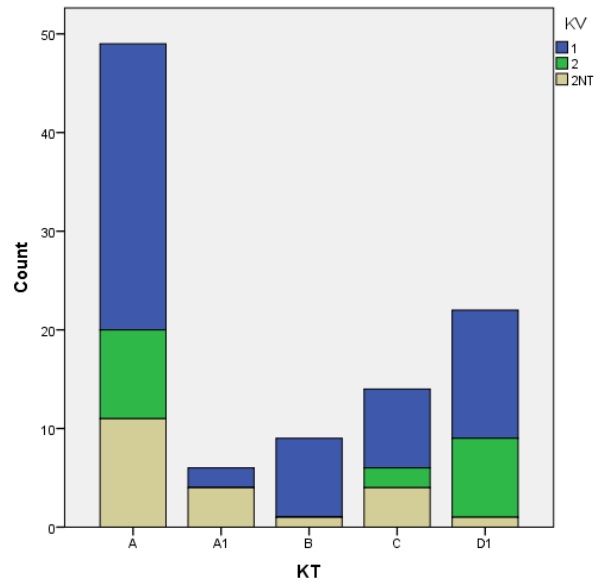
MR. HUYNH NAM

EMAIL: giangdayit@gmail.com

FB: www.facebook.com/mrhuynhnam

Website: www.huynhnam.com || www.giangdayit.com

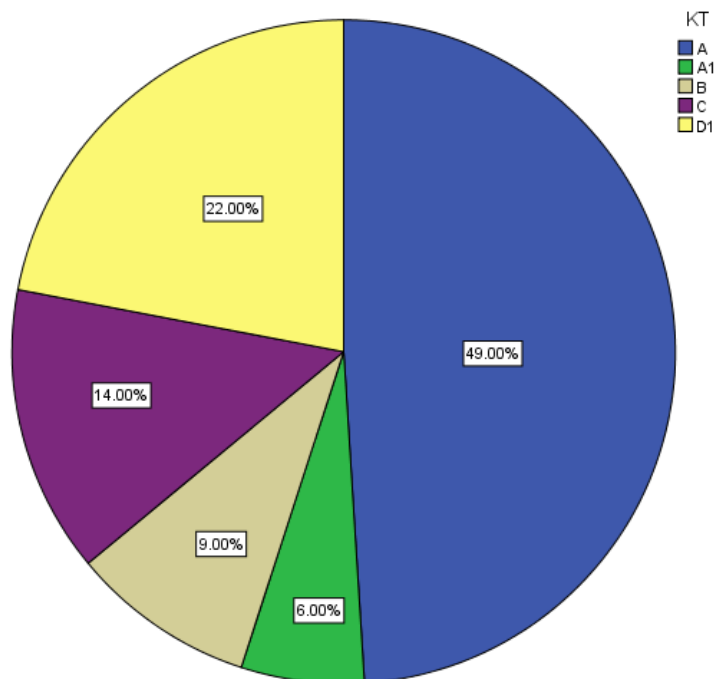
2. Sinh viên tự làm: Biểu đồ dạng Stacked
Kết quả



c. Vẽ biểu đồ tròn

i. Vẽ biểu đồ tròn thể hiện tần suất của biến KT

Kết quả



MR. HUYNH NAM

EMAIL: giangdayit@gmail.com

FB: www.facebook.com/mrhuynhnam

Website: www.huynhnam.com || www.giangdayit.com

2. Tiến hành phân tích dữ liệu định lượng

- a. Tạo phân lớp cho dữ liệu định lượng T1 để phân tích tần số và tần suất. Sau đó, hãy vẽ các biểu đồ cột và biểu đồ tròn tương ứng. Biết rằng phân lớp

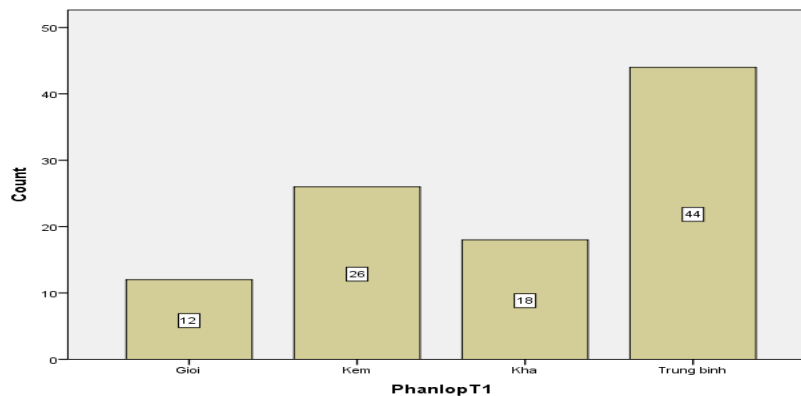
(PhanlopT1) của dữ liệu T1 là:

- i. 0 đến dưới 5 = Kém
- ii. 5 đến dưới 7 = Trung bình
- iii. 7 đến dưới 8 = Khá
- iv. từ 8 trở đi = Giỏi

Kết quả

Statistics		
PhanlopT1		
N	Valid	100
	Missing	0

PhanlopT1					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Gioi	12	12.0	12.0	12.0
	Kem	26	26.0	26.0	38.0
	Kha	18	18.0	18.0	56.0
	Trung binh	44	44.0	44.0	100.0
	Total	100	100.0	100.0	

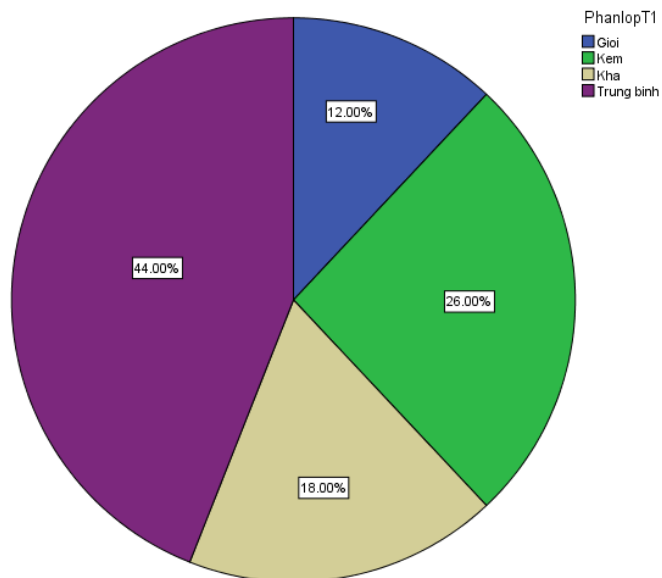


MR. HUYNH NAM

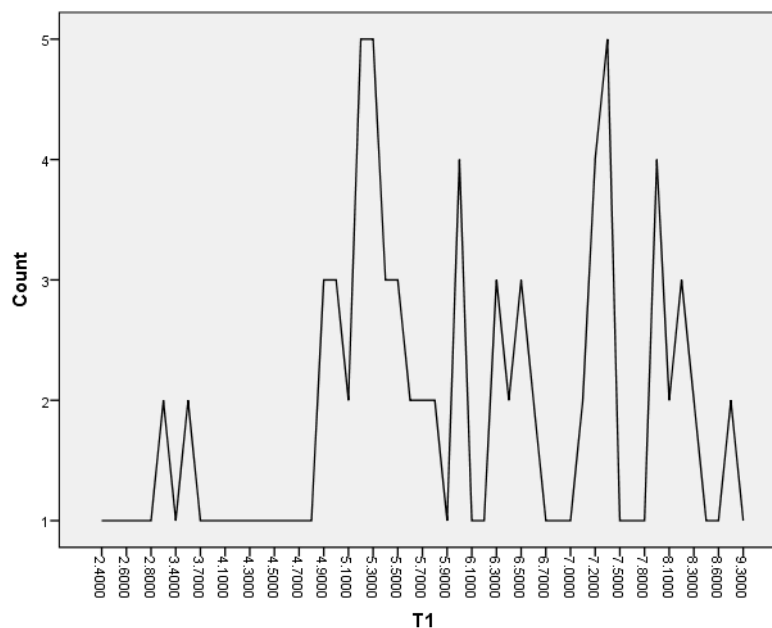
EMAIL: giangdayit@gmail.com

FB: www.facebook.com/mrhuynhnam

Website: www.huynhnam.com || www.giangdayit.com



- b. Vẽ biểu đồ đường cho dữ liệu định lượng T1 thể hiện số lượng từng điểm số
Kết quả



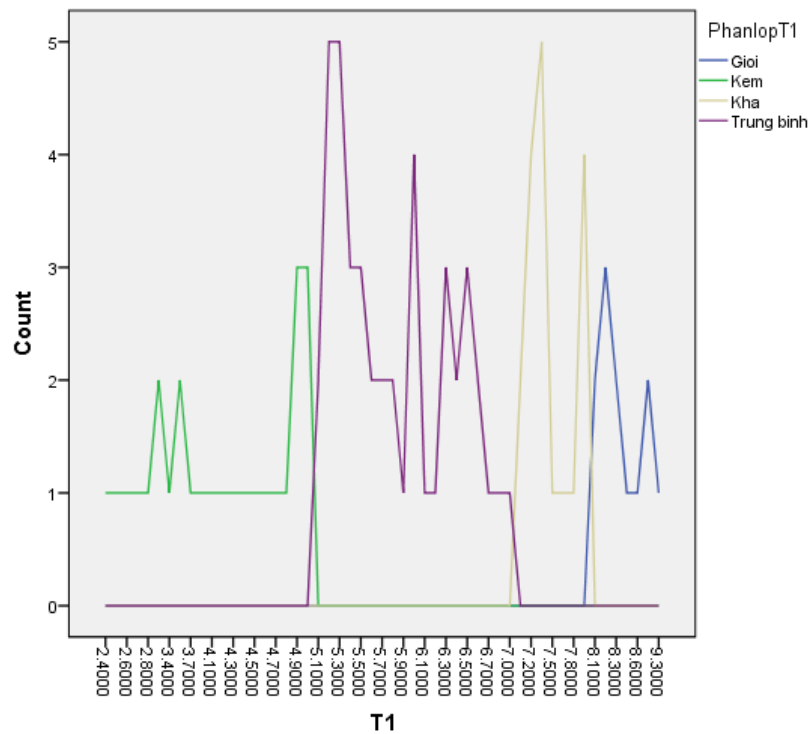
- c. Vẽ biểu đồ đường cho biến T1 thể hiện số lượng từng điểm số mà được phân loại bởi biến PhanlopT1
Kết quả

MR. HUYNH NAM

EMAIL: giangdayit@gmail.com

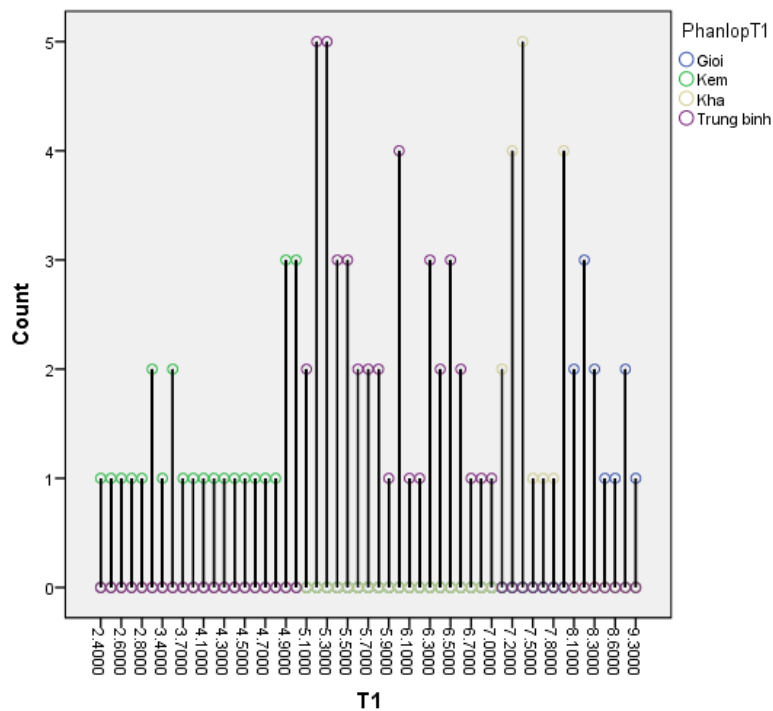
FB: www.facebook.com/mrhuynhnam

Website: www.huynhnam.com || www.giangdayit.com



- d. Vẽ biểu đồ đường Drop-line cho biến T1 thể hiện số lượng từng điểm số mà được phân loại bởi biến PhanlopT1

Kết quả



MR. HUYNH NAM

EMAIL: giangdayit@gmail.com

FB: www.facebook.com/mrhuynhnam

Website: www.huynhnam.com || www.giangdayit.com

- e. Vẽ đồ thị stem-and-leaf và boxplot cho biến T1
Kết quả

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
T1	100	100.0%	0	0.0%	100	100.0%

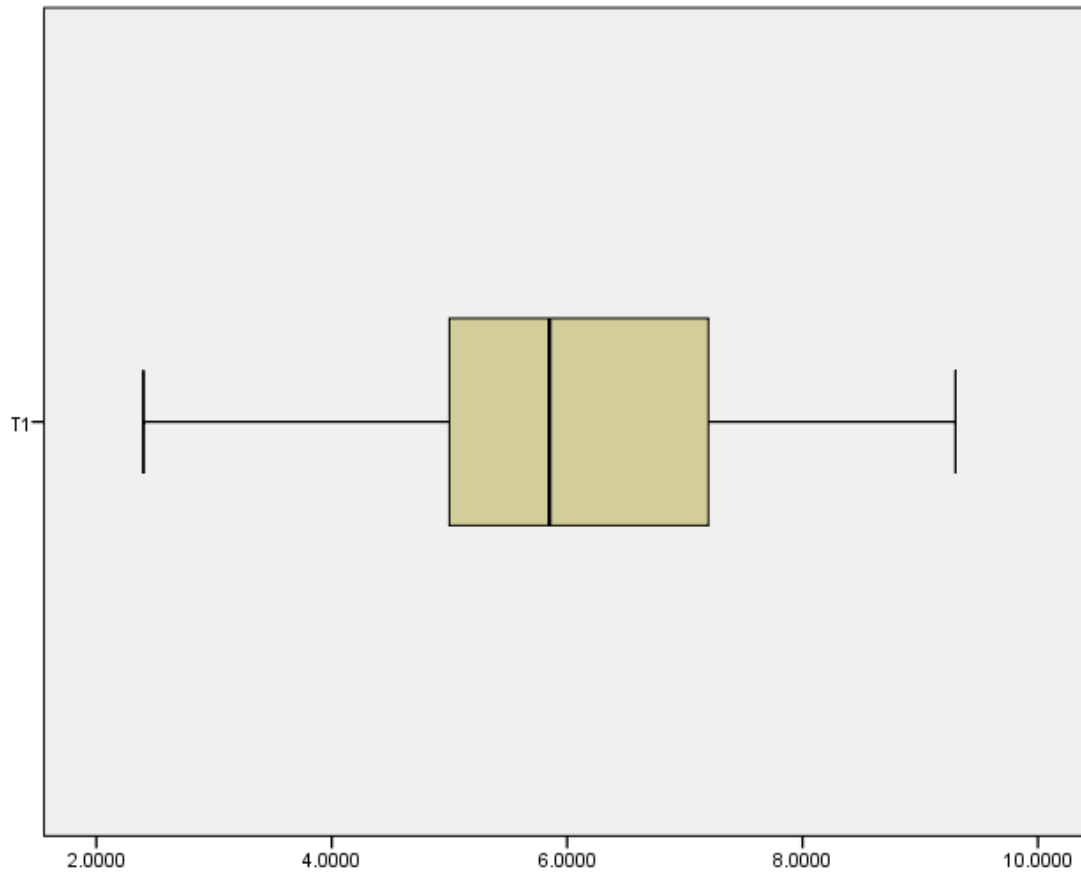
T1 Stem-and-Leaf Plot		
Frequency	Stem	Leaf
1.00	2	4
4.00	2	5 6 7 8
3.00	3	0 0 4
4.00	3	6 6 7 9
4.00	4	1 2 3 4
7.00	4	5 6 7 8 9 9 9
18.00	5	0 0 1 1 2 2 2 2 2 3 3 3 3 4 4 4
10.00	5	5 5 5 6 6 7 7 8 8 9
11.00	6	0 0 0 1 2 3 3 3 4 4
7.00	6	5 5 5 6 6 7 9
12.00	7	0 1 1 2 2 2 3 3 3 3 3
7.00	7	5 6 8 9 9 9 9
8.00	8	1 1 2 2 2 3 3 4
3.00	8	6 8 8
1.00	9	3
Stem width:	1.0000	
Each leaf:	1 case(s)	

MR. HUYNH NAM

EMAIL: giangdayit@gmail.com

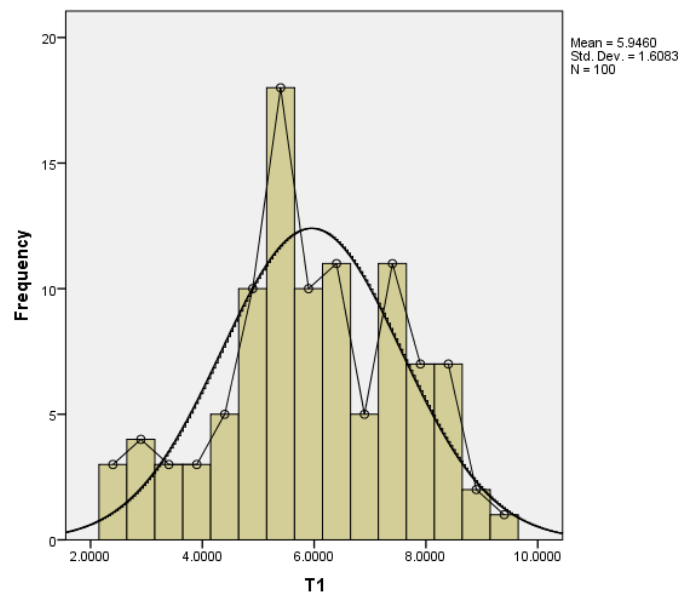
FB: www.facebook.com/mrhuynhnam

Website: www.huynhnam.com || www.giangdayit.com



f. Vẽ histogram cho biến T1

Kết quả



MR. HUYNH NAM

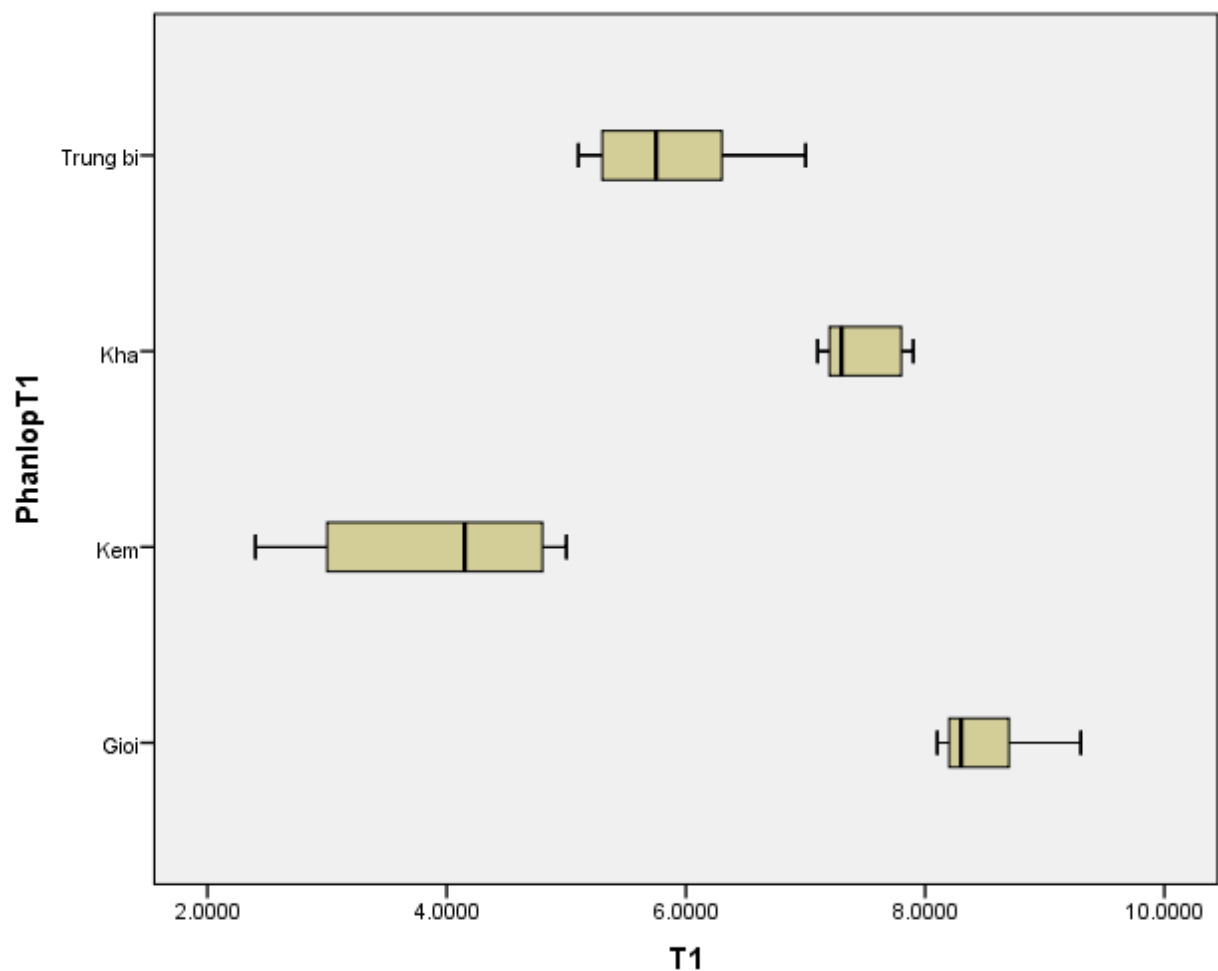
EMAIL: giangdayit@gmail.com

FB: www.facebook.com/mrhuynhnam

Website: www.huynhnam.com || www.giangdayit.com

g. Vẽ Boxplot cho biến T1 theo PhanlopT1
Kết quả

Case Processing Summary							
	PhanlopT1	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
T1	Gioi	12	100.0%	0	0.0%	12	100.0%
	Kem	26	100.0%	0	0.0%	26	100.0%
	Kha	18	100.0%	0	0.0%	18	100.0%
	Trung bi	44	100.0%	0	0.0%	44	100.0%



MR. HUYNH NAM

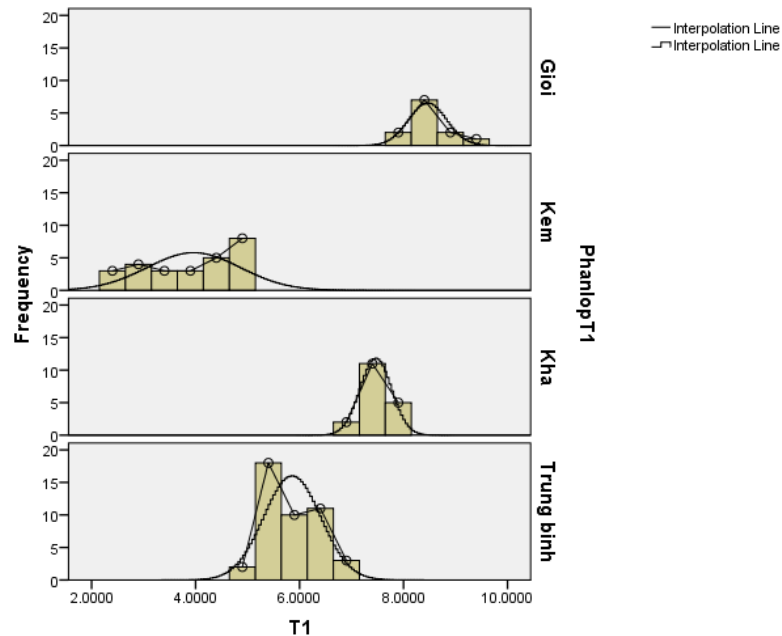
EMAIL: giangdayit@gmail.com

FB: www.facebook.com/mrhuynhnam

Website: www.huynhnam.com || www.giangdayit.com

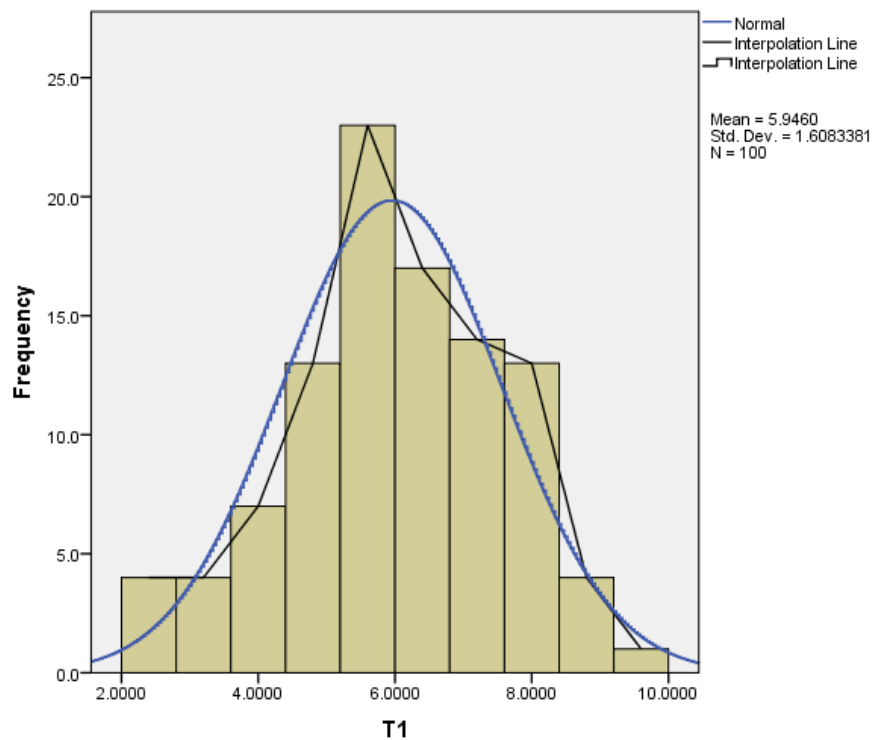
h. Vẽ histogram của T1 theo biến PhanlopT1

Kết quả



i. Vẽ histogram cho biến T1 với số khoảng là 10

Kết quả



MR. HUYNH NAM

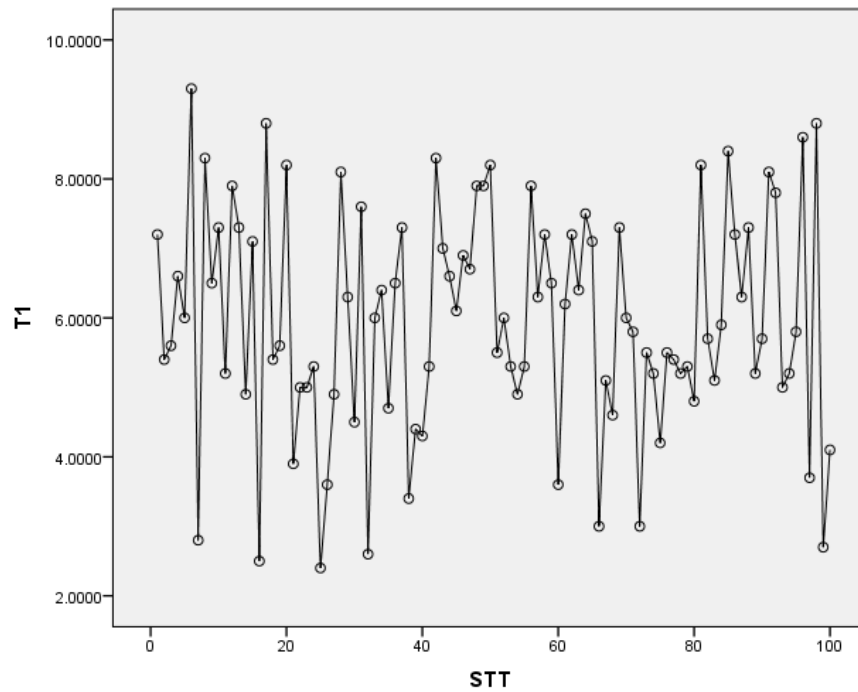
EMAIL: giangdayit@gmail.com

FB: www.facebook.com/mrhuynhnam

Website: www.huynhnam.com || www.giangdayit.com

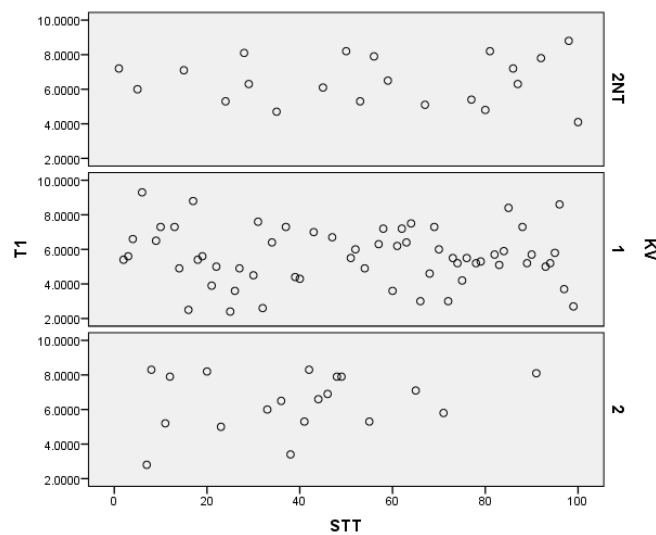
j. Vẽ đồ thị phân tán của biến T1

Kết quả



k. Vẽ đồ thị phân tán của biến T1 phân biệt theo biến KV

Kết quả



MR. HUYNH NAM

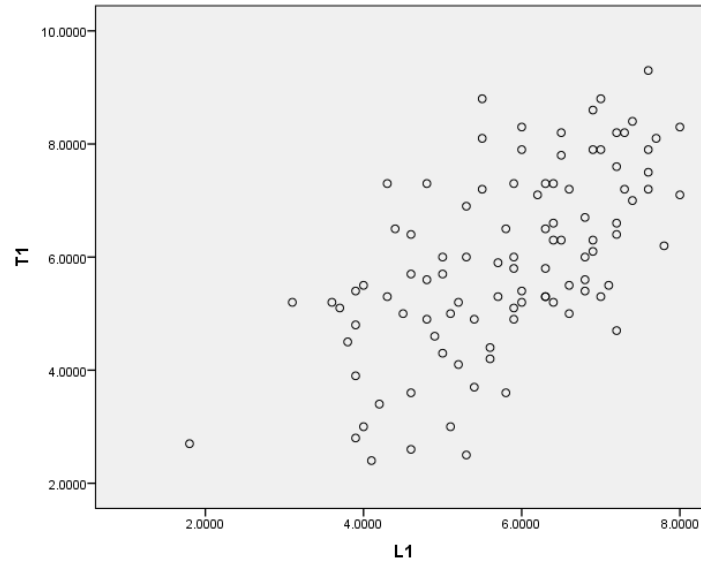
EMAIL: giangdayit@gmail.com

FB: www.facebook.com/mrhuynhnam

Website: www.huynhnam.com || www.giangdayit.com

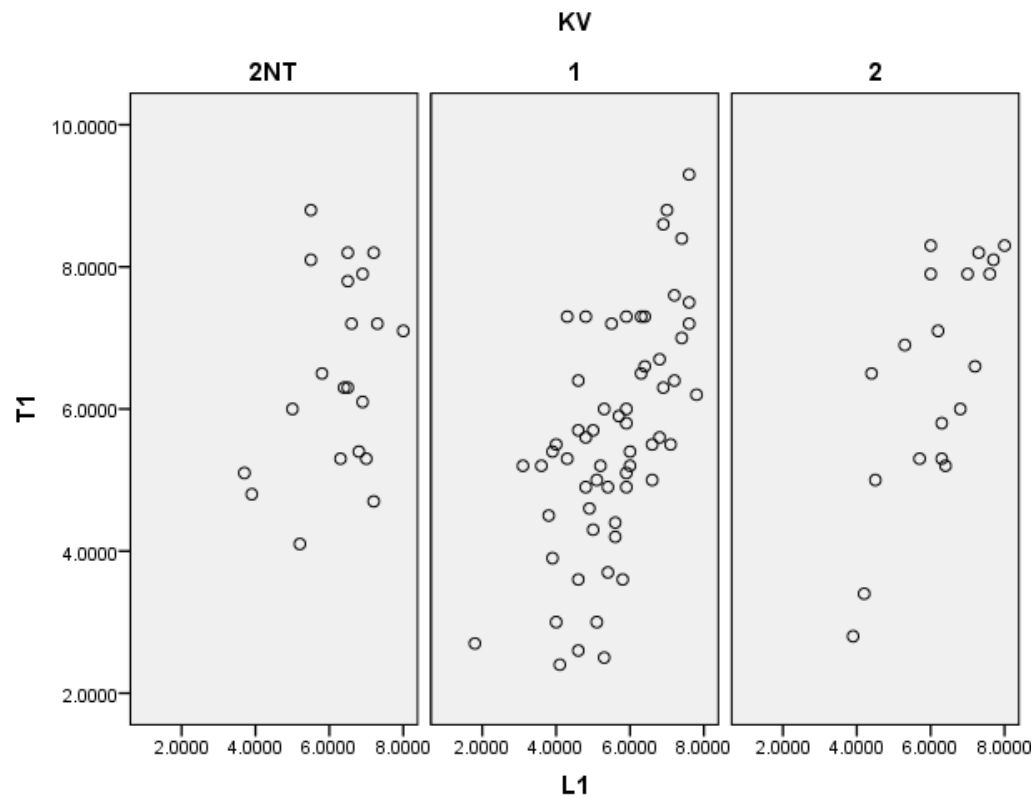
- l. Vẽ đồ thị phân tán của biến T1 theo L1

Kết quả



- m. Vẽ đồ thị phân tán của biến T1 theo L1 và có phân biệt theo biến KV

Kết quả



MR. HUYNH NAM

EMAIL: giangdayit@gmail.com

FB: www.facebook.com/mrhuynhnam

Website: www.huynhnam.com || www.giangdayit.com

n. Tính các đại lượng thống kê mô tả cho biến T1

Kết quả

Descriptive Statistics												
	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
T1	100	6.9000	2.4000	9.3000	5.946000	.1608338	1.6083381	2.587	-.178	.241	-.480	.478
Valid N (listwise)	100											

o. Tính tứ phân vị và phân vị thứ 20%, 30% cho biến T1

Kết quả

Statistics		
T1		
N	Valid	100
	Missing	0
Percentiles	20	4.820000
	25	5.000000
	30	5.200000
	50	5.850000
	75	7.200000

p. Hãy khảo sát hình dạng phân phối của dữ liệu T1

Hướng dẫn:

1. *Hiện thị số liệu thống kê: mean, median, mode, kurtosis, skewness, tứ phân vị, standard deviation*
2. *Hiện thị biểu đồ Stem leaf, box plot và histogram*
3. *Số sánh Mean, Median, Mode*
4. *Kết luận hình dạng phân phối*

Kết quả: T1 phân phối lệch trái

Bài tập tự làm

Sinh viên thực hiện tương tự các bước thống kê cho biến L1