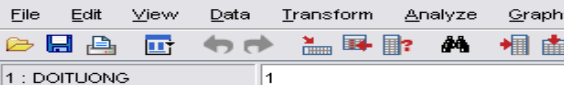


Nhận biết phân phối chuẩn trong SPSS

Có nhiều cách để nhận biết một phân phối chuẩn trong SPSS.

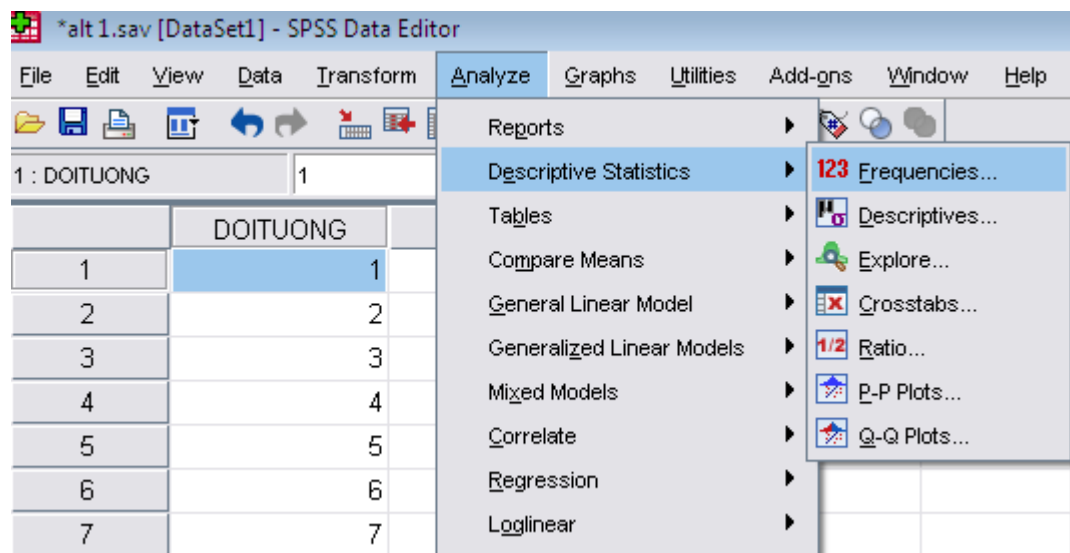
- (1) Đơn giản nhất là xem biểu đồ với đường cong chuẩn (Histograms with normal curve) với dạng hình chuông đối xứng với tần số cao nhất nằm ngay giữa và các tần số thấp dần nằm ở 2 bên. Trị trung bình (mean) và trung vị (median) gần bằng nhau và độ xiên (skewness) gần bằng zero.
- (2) Vẽ biểu đồ xác suất chuẩn (normal Q-Q plot). Phân phối chuẩn khi biểu đồ xác suất này có quan hệ tuyến tính (đường thẳng)
- (3) Dùng phép kiểm định Kolmogorov-Smirnov khi cỡ mẫu lớn hơn 50 hoặc phép kiểm Shapiro-Wilk khi cỡ mẫu nhỏ hơn 50. Được coi là có phân phối chuẩn khi mức ý nghĩa (Sig.) lớn hơn 0,05.

Ví dụ 1. Khảo sát men ALT (ug/l) trên 30 người bình thường, kết quả được nhập vào SPSS như sau:

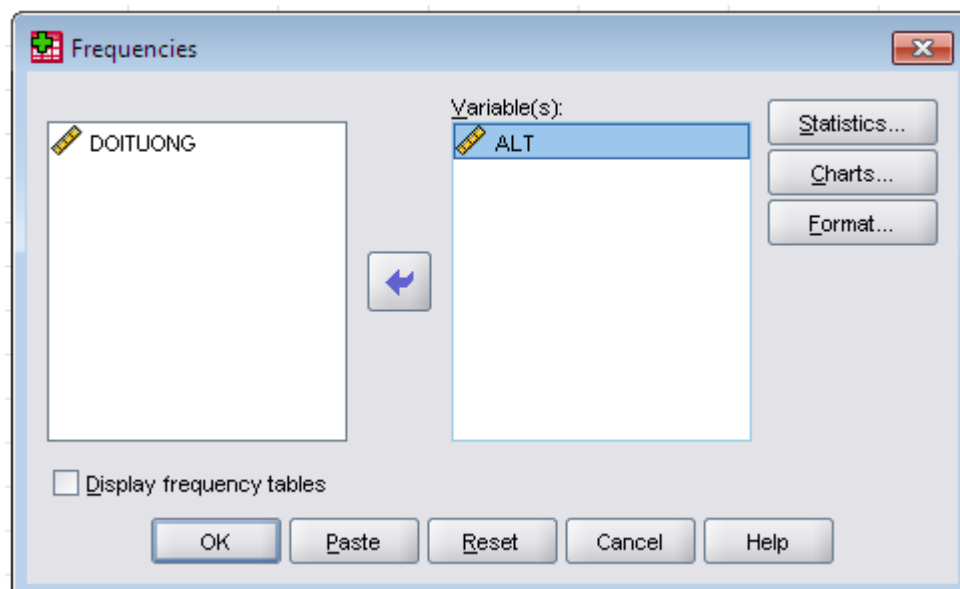


	DOITUONG	ALT
1	1	12
2	2	13
3	3	14
4	4	15
5	5	16
6	6	17
7	7	18
8	8	19
9	9	20
10	10	21
11	11	22
12	12	23
13	13	24
14	14	25
15	15	26
16	16	27
17	17	28
18	18	29
19	19	30
20	20	31
21	21	32
22	22	33
23	23	34
24	24	26
25	25	22
26	26	23
27	27	24
28	28	25
29	29	26
30	30	44

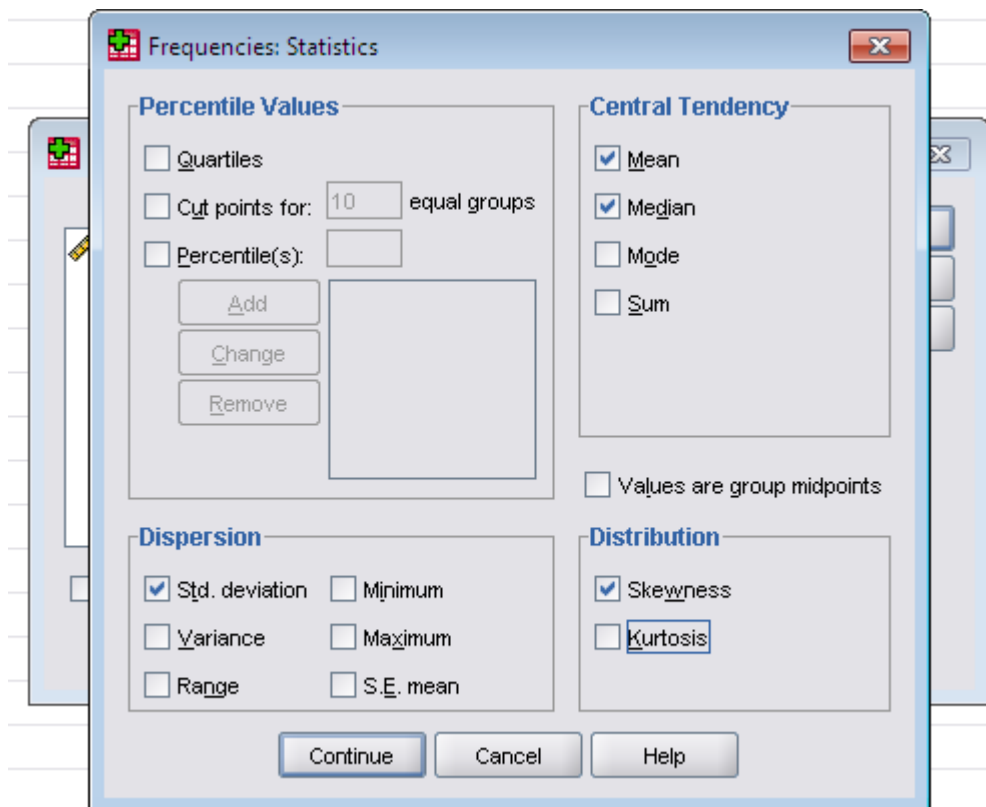
Vào thực đơn Analyze > Descriptive Statistics> Frequencies...



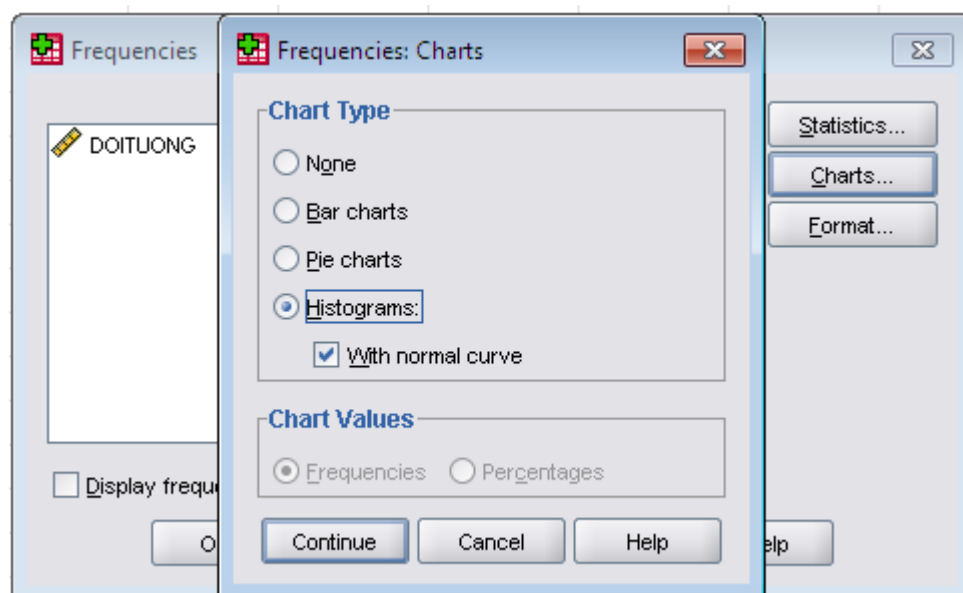
Mở hộp thoại Frequencies. Nhấp chuyển biến ALT từ ô bên trái vào ô Variable(s).



Nhấp hộp thoại Statistics... Vào màn hình Statistics, đánh dấu nhảy vào 4 ô: Mean, Median, Std. deviation và Skewness và nhấp Continue



Nhấn tiếp hộp thoại Charts.. Đánh dấu vào ô tròn Histograms: và đánh dấu nhảy vào ô With Normal curve, nhấn Continue. Nhấn OK sẽ cho kết quả sau.



➔ Frequencies

[DataSet1] F:\BAIVIET\BAITAP\alt 1.sav

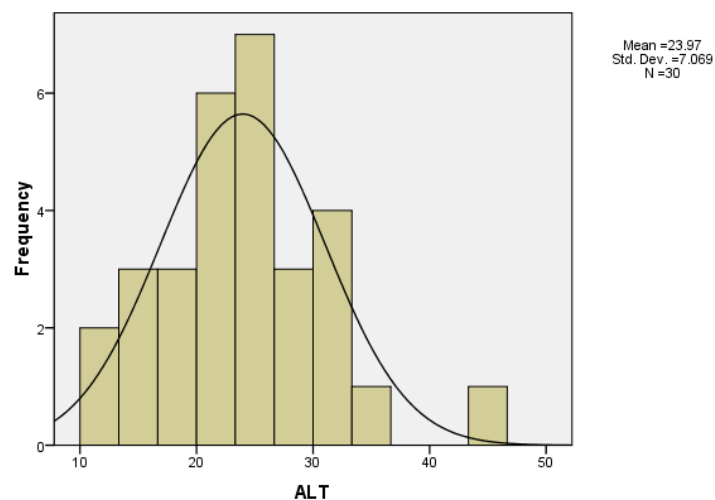
Statistics

ALT		
N	Valid	30
	Missing	0
Mean		23.97
Median		24.00
Std. Deviation		7.069
Skewness		.533
Std. Error of Skewness		.427
Kurtosis		.856
Std. Error of Kurtosis		.833

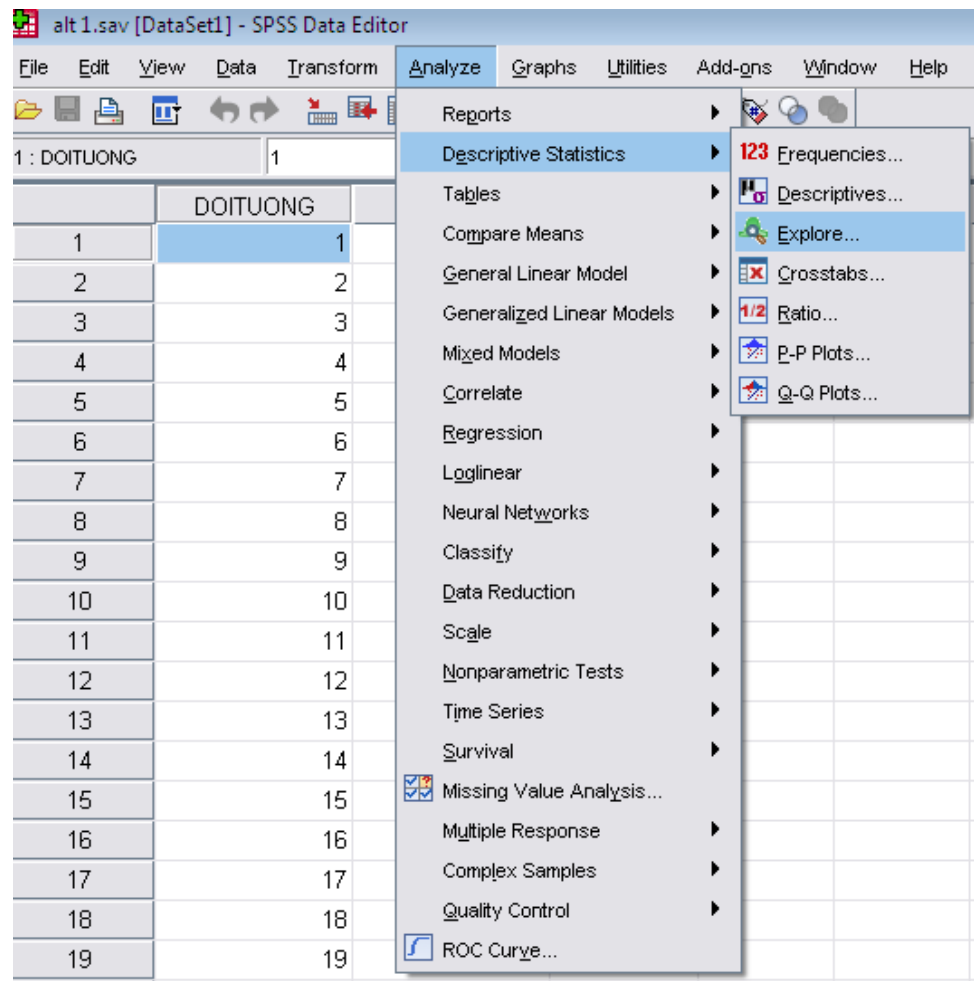
Trị trung bình (mean)= 23,97, trung vị (median)=24,00 và độ xiên (skewness)=0,533

Trong phân phối này, trị số trung bình và trung vị gần bằng nhau và độ xiên dao động từ -1 đến +1, vì vậy được coi như có phân phối chuẩn. Thật vậy xem biểu đồ phân phối với đường cong chuẩn có dạng hình chuông, có trị trung bình là 23,97 và số liệu phân phối khá đều 2 bên.

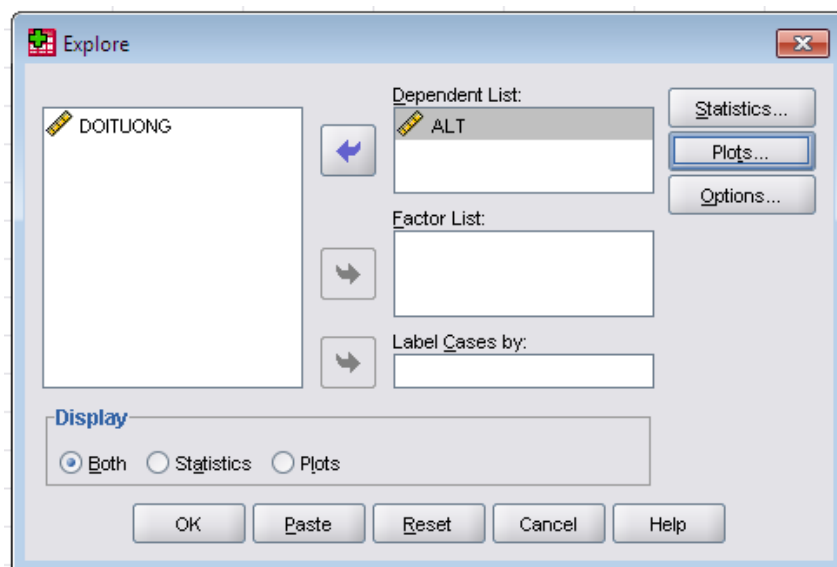
Histogram



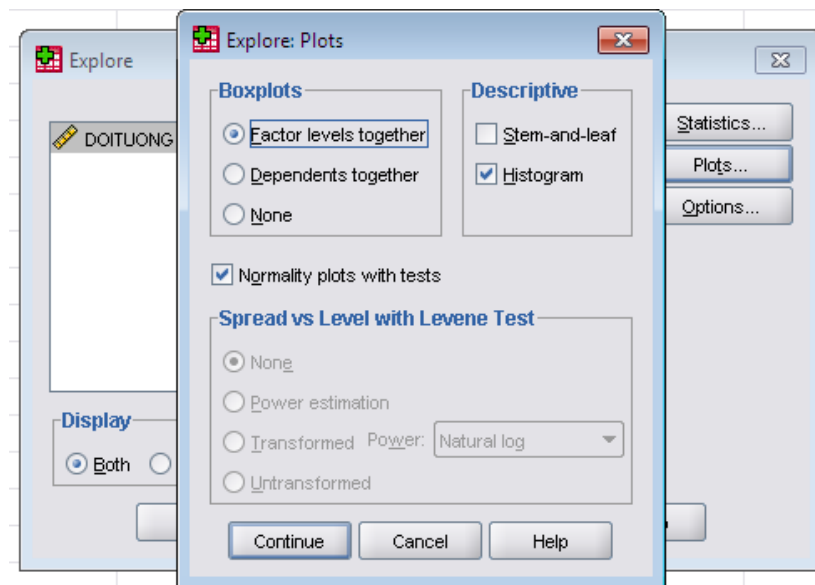
Để kiểm định Kolmogorov-Smirnov hoặc vẽ biểu đồ xác suất chuẩn Q-Q (Normal Q-Q plot) vào thực đơn: **Analyze > Descriptive Statistics > Explore...**



Khi xuất hiện màn hình Explore, chuyển ALT từ ô bên trái vào ô Dependent List:



Nhấn vào nút Plots. Đánh dấu nháy vào ô Histogram và ô Normality plots with tests.
Nhấp Continue và nhấp OK



Kết quả kiểm định phân phối chuẩn như sau:

Descriptives				Statistic	Std. Error
ALT	Mean			23.97	1.291
	95% Confidence Interval for Mean	Lower Bound		21.33	
		Upper Bound		26.61	
	5% Trimmed Mean			23.69	
	Median			24.00	
	Variance			49.964	
	Std. Deviation			7.069	
	Minimum			12	
	Maximum			44	
	Range			32	
	Interquartile Range			10	
	Skewness			.533	.427
	Kurtosis			.856	.833

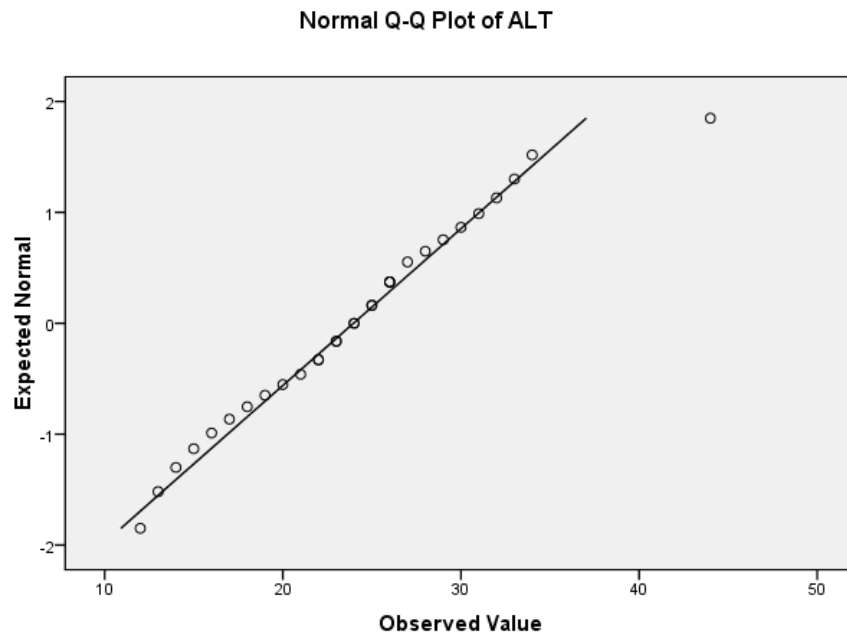
Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ALT	.087	30	.200 [*]	.971	30	.571

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Vì cỡ mẫu 30 (nhỏ hơn 50), dùng kiểm định Shapiro-Wilk với $Sig.=0,571$ (lớn hơn 0,05). Chứng tỏ phân phối này là phân phối chuẩn.

Xem biểu đồ Normal Q-Q Plot bên dưới, các trị số quan sát và trị số mong đợi đều nằm gần trên đường chéo.



Ví dụ 2. Khảo sát men ALT (ug/l) trên 30 người mắc viêm gan siêu vi B mãn tính

alt 2.sav [DataSet2] - SPSS Data Editor

	DOITUONG	ALT_VIEMGAN
1	1	20
2	2	35
3	3	40
4	4	30
5	5	40
6	6	60
7	7	80
8	8	90
9	9	100
10	10	110
11	11	130
12	12	140
13	13	150
14	14	180
15	15	190
16	16	190
17	17	200
18	18	210
19	19	220
20	20	270
21	21	280
22	22	290
23	23	300
24	24	320
25	25	380
26	26	400
27	27	900
28	28	1000
29	29	1200
30	30	1500

Vào Analyze> Descriptive > Frequencies như phần trên
Kết quả

➔ **Frequencies**

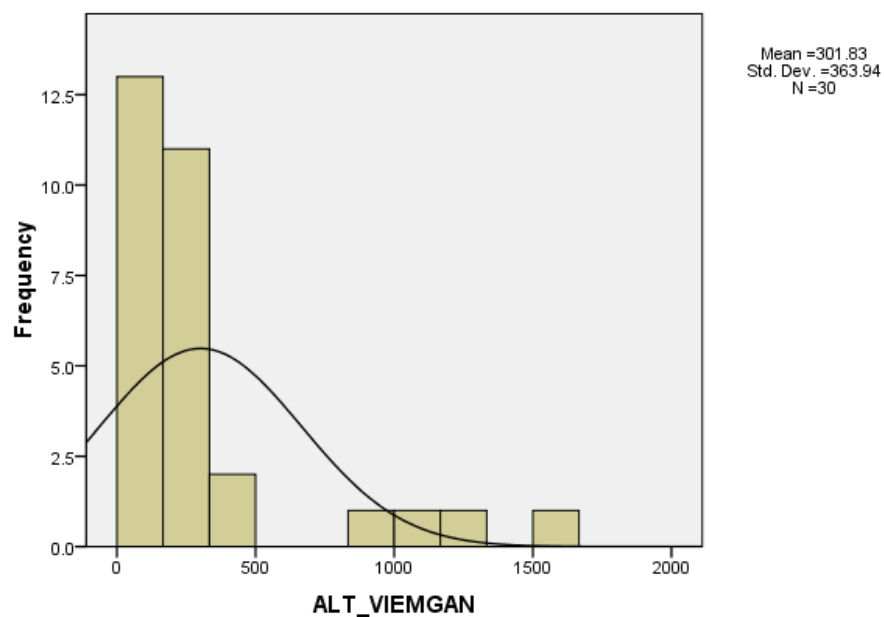
[DataSet2] F:\BAIVIET\BAITAP\alt 2.sav

Statistics

ALT_VIEMGAN		
N	Valid	30
	Missing	0
Mean		301.83
Median		190.00
Std. Deviation		363.940
Skewness		2.172
Std. Error of Skewness		.427

Phân phối này có trung bình (301,82) và trung vị (190,00) khá khác biệt. Hơn nữa độ xiên (2,17) lớn hơn +1, vì vậy có thể không phải là phân phối chuẩn. Thật vậy, xem biểu đồ với đường cong chuẩn cho thấy số liệu không phân phối đều, tập trung nhiều về phía đuôi trái và ít về phía đuôi phải (xệch phải)

Histogram



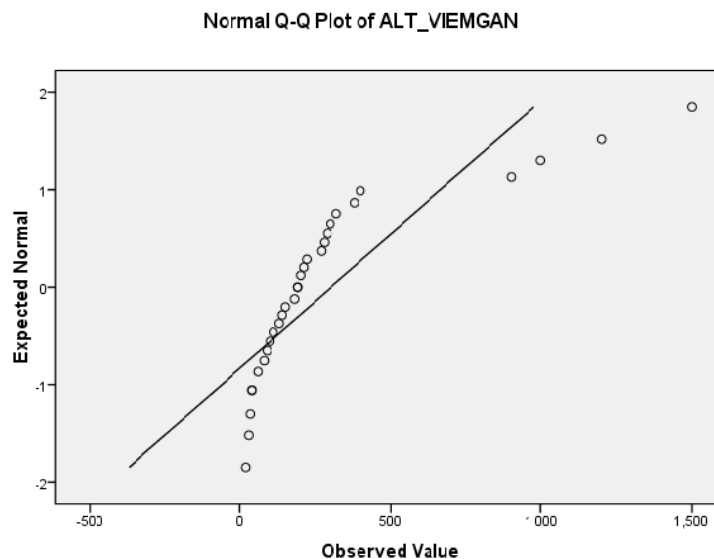
Để rõ hơn vào Analyze>Descriptives> Explore để xem kiểm định Kolmogorov-Smirnov và biểu đồ Normal Q-Q plot

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ALT_VIEMGAN	.280	30	.000	.690	30	.000

a. Lilliefors Significance Correction

Kiểm định Shapiro-Wilk với Sig=0,000 (nhỏ hơn 0,05 → bác bỏ giả thuyết không).

Như vậy phân phối này không phải là phân phối chuẩn. Trên biểu đồ Q-Q plot ta thấy sự liên quan giữa trị số quan sát và trị số mong đợi không nằm trên đường thẳng chéo.



TS Nguyễn Ngọc Rạng, Email:rangbvag@yahoo.com, website:bvag.com.vn