فصل چهارم

پردازش داده‌های خام

محاسبه پارامترهای آماری داده‌های خام

اولین مرحلهٔ پردازش داده‌های ژئوشیمیایی، بررسی پارامترهای آماری مربوط به تک‌تک عناصر جهت شناخت ماهیت توزیع هریک از آنها می‌باشد که با محاسبهٔ پارامترهای آماری از قبیل میانگین، انحراف معیار، چولگی، کشیدگی، واریانس و . . . می‌توان به این موضوع دست‌یافت. در این قسمت برای هر عنصر به عنوان یک متغیر آماری در یک جدول، تعداد نمونه‌ها، حداقل و حداکثر عیار، میانگین، میانه، انحراف معیار، چولگی و کشیدگی و نمودارهای هیستوگرام توزیع فراوانی محاسبه و ترسیم شده‌اند.

بررسی مقادیر خارج از رده : (Outliers)

هنگام بررسی مقادیر داده‌های خام به نمونه‌هایی برخورد می‌شود که در آستانه‌های بالا و پایین جامعه داده‌ها قرار گرفته و از جامعه اصلی جدا افتاده‌اند. اگر نمودار جعبه‌ای (**Boxplot**) آنها ترسیم شود این نمونه‌ها به نحو بارزی خودشان را از بقیه جدا می‌کنند. مقادیر خارج از رده به سه حالت مختلف زیر ممکن است بوجود آیند:

حالت اول) از یک خطای سیستماتیک به هنگام نمونه‌برداری، آماده‌سازی یا تجزیه شیمیایی نمونه‌ها ناشی شده باشند که باید از مرحله پردازش حذف یا اصلاح شوند.

حالت دوم) مشاهداتی که به صورت یک پدیده فوق‌العاده نمود پیدا می‌کنند که باید پس از بررسی اعتبار آنها در مورد حفظ یا حذف آنها تصمیم گرفت.

حالت سوم) مشاهدات فوق‌العاده‌ای که هیچگونه توضیح مناسبی برای آنها وجود ندارد و کارشناس اگر احساس کند که آنها به عنوان گوشه‌ای از جامعه مورد بررسی هستند می‌تواند آنها را حفظ کند.

وجود مقادیر خارج از رده در جامعه نمونه‌ها موجب افزایش واریانس جامعه و نیز همبستگی بین متغیرها وهمچنین افزایش چولگی در نمودار توزیع عناصر می‌شود. برای کاهش این تاثیر راههای مختلفی نظیر محاسبه ضریب همبستگی با استفاده از روشهای ناپارامتری مانند روش اسپیرمن **(Spearman)** ، حذف یا جایگزین نمودن مقادیر استفاده می‌شود دراین گزارش از روش جایگزین نمودن مقادیر خارج از رده استفاده شده است. جدول (۴-۱) نمونه‌های دارای مقادیر خارج از رده را نشان می‌دهد.

نرمال سازی داده‌های خام :

استفاده از برخی روشهای آماری منوط به نرمال بودن تابع توزیع متغیرهای مورد مطالعه است در حالیکه توابع توزیع از نوع لاگ نرمال است، به همین علت قبل از استفاده از این روشها داده‌های خام باید نرمال شوند. دراین بخش از نوعی تبدیلات جهت نرمال کردن تابع توزیع داده‌های خام استفاده شده است . این کار شرط لازم کاربرد برخی روشهای آماری مانند تعیین نمونه‌های آنومالی با استفاده از اضافه کردن ضرایبی از انحراف معیار به حد آستانه‌ای و یا محاسبه ضرایب همبستگی پیرسون می‌باشد. روش لاگ نرمال به صورت یک روش توصیفی برای نرمال کردن تابع توزیع جوامعی که دارای چولگی در نمودار خود هستند به کار می رود.

Table (4-1) : Outlier Samples For Normal Raw Data

	Sample Number	
Elements	**Outlier (+)**	**Outlier (-)**
Au	BE-53 , BS-219 , BB-268	
W		
Mo		
Be		
Cr		
Co		
Ni		
Cu		
Zn	BB-408 , BE-006 , BB-380	
As		
Ag	BE-102 , BE-103 , BE-104 , BB-311 , BS-231	
Sn	BB-477 , BR-202 , BB-260 , BB-268	
Sb	BB-417 , BB-418 , BB-419 , BB-423 , BB-424	
B	BB-292	
Ba		
Pb		
Bi	BB-445 , BB-446 , BB-447	
Hg		
Ti	BB-452	
Mn		

در اینجا از لگاریتم طبیعی مقادیر داده‌های خام به اضافه یا منهای یک مقدار ثابت λ مطابق رابطه تبدیلی زیر استفاده شده است.

$$Z = \mathrm{Ln}\,(AE \pm \lambda)$$

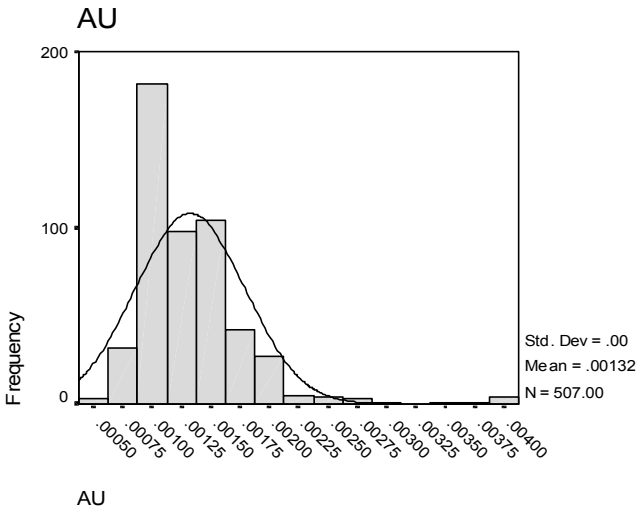
در این رابطه AE آنالیز نمونه برای هر عنصر است.

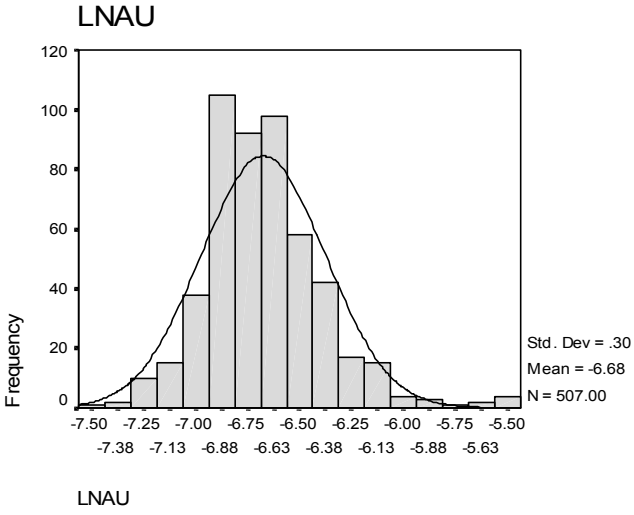
برای هر عنصر مقدار λ به گونه‌ای انتخاب می‌شود که پس از انتخاب داده‌ها به یک مقدار بهینه از چولگی و کشیدگی در منحنی توزیع نرمال دست یافته شود. پارامترهای آماری و هیستوگرام‌های ترسیم شده برای داده‌های نرمال در شکل (۴-۱) تا (۴-۷) آورده شده است. با توجه به این پارامترهای آماری می‌توان دریافت که مقادیر چولگی و کشیدگی متغیرها در مقایسه با مقادیر متناظر مربوط به داده‌های خام نرمال نشده تا چه اندازه کاهش یافته و منحنی توزیع تجمعی آنها به صورت یک خط راست که بیانگر توزیع نرمال می‌باشد، ظاهر شده است. هیستوگرام مقادیر نرمال شده نسبت به هیستوگرام مقادیر نرمال نشده نیز بیانگر مطلب فوق می‌باشد.

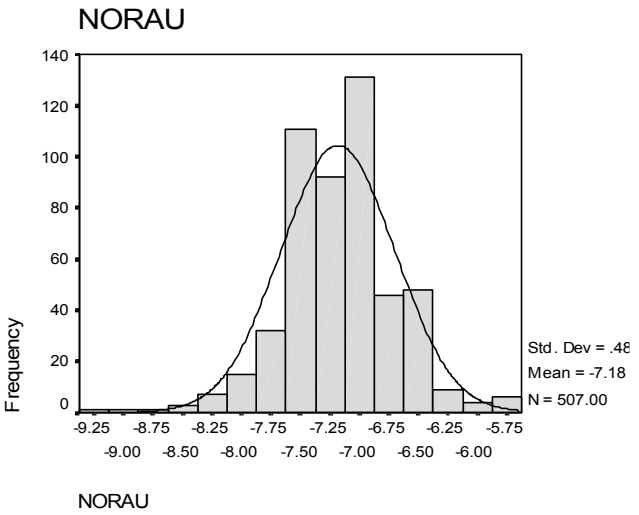
Fig (4-1) : Statiatical Parameters For Raw Data

Statistics

		AU	LNAU	NORAU
N	Valid	507	507	507
	Missing	0	0	0
Mean		1.324E-03	-6.6753	-7.1836
Std. Error of Mean		2.072E-05	1.327E-02	2.145E-02
Median		1.200E-03	-6.7254	-7.2221
Mode		.0010	-6.91	-7.54
Std. Deviation		4.665E-04	.2987	.4830
Variance		2.176E-07	8.924E-02	.2333
Skewness		2.361	.775	.000
Std. Error of Skewness		.108	.108	.108
Kurtosis		9.100	1.385	1.267
Std. Error of Kurtosis		.217	.217	.217
Range		.0033	1.94	3.64
Minimum		.0006	-7.49	-9.31
Maximum		.0039	-5.55	-5.68
Sum		.6715	-3384.35	-3642.09







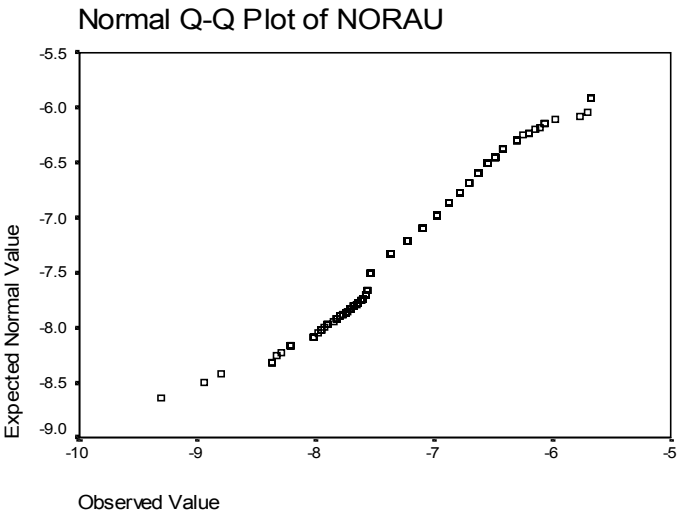
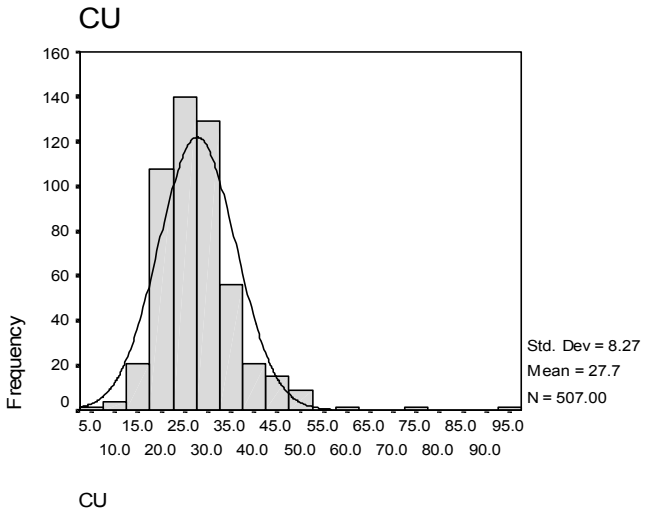
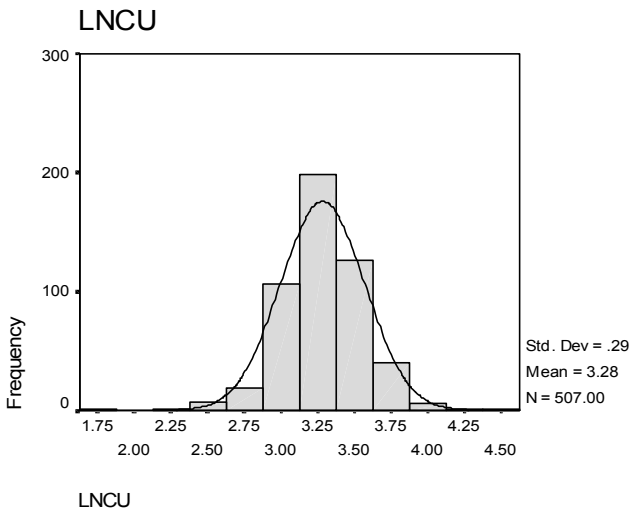


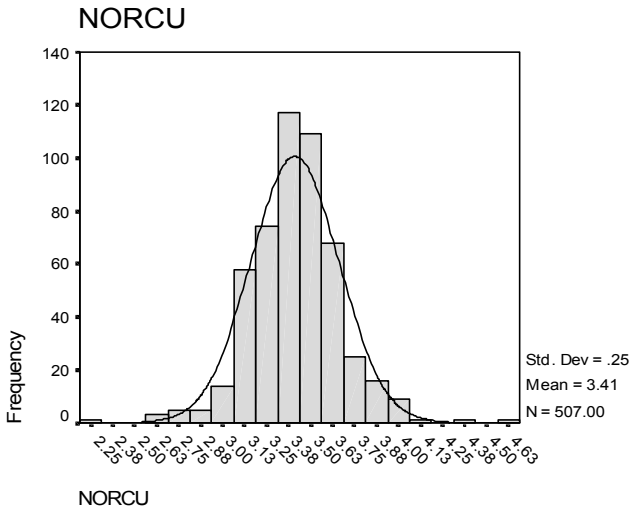
Fig (4-2): Statiatical Parameters For Raw Data

Statistics

		CU	LNCU	NORCU
N	Valid	507	507	507
	Missing	0	0	0
Mean		27.713	3.2811	3.4101
Std. Error of Mean		.367	1.275E-02	1.114E-02
Median		27.000	3.2958	3.4187
Mode		29.0	3.37	3.48
Std. Deviation		8.273	.2870	.2507
Variance		68.450	8.239E-02	6.287E-02
Skewness		1.756	-.257	.000
Std. Error of Skewness		.108	.108	.108
Kurtosis		9.175	2.727	2.117
Std. Error of Kurtosis		.217	.217	.217
Range		87.6	2.85	2.38
Minimum		5.4	1.69	2.19
Maximum		93.0	4.53	4.57
Sum		14050.6	1663.53	1728.92







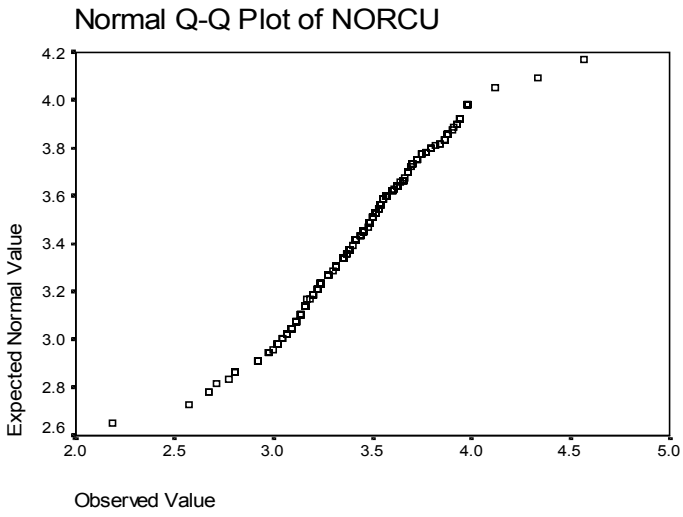
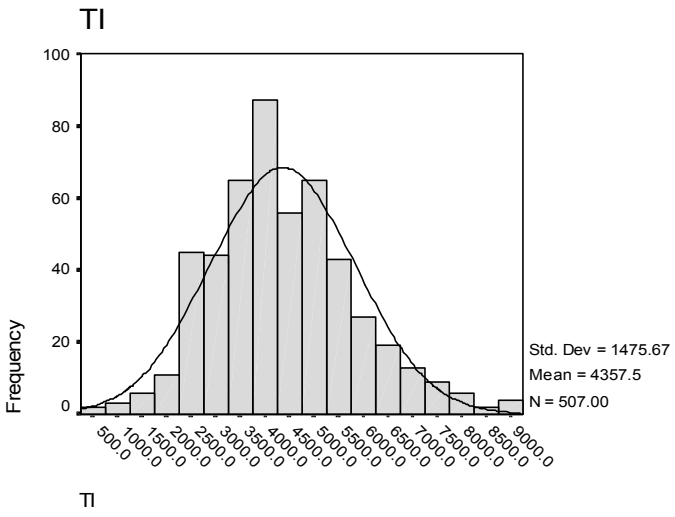
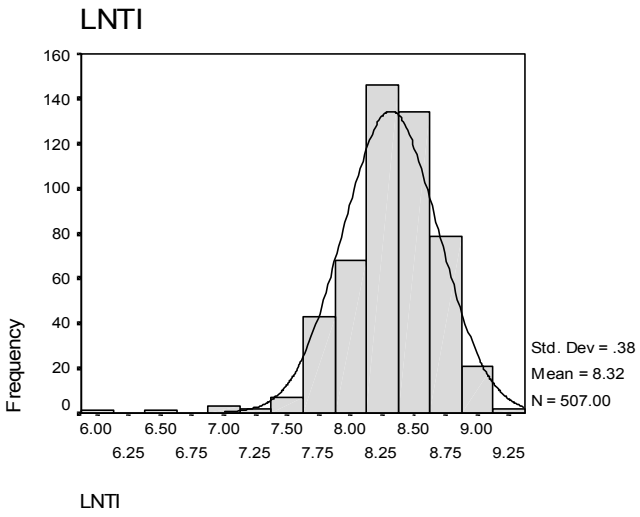


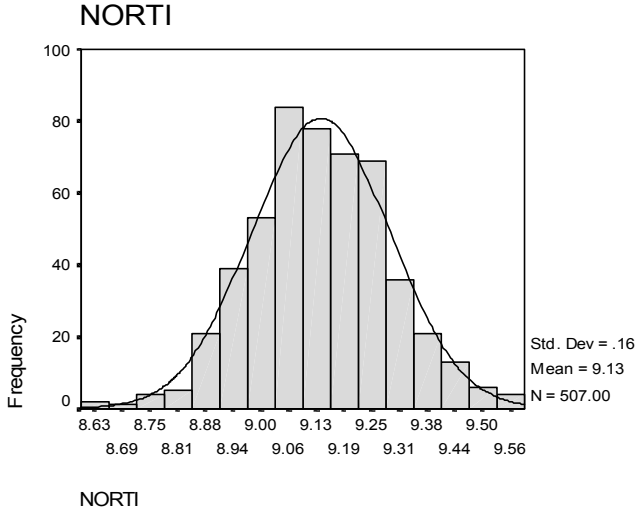
Fig (4-3) :Stastistical Parameters For Raw Data

Statistics

		TI	LNTI	NORTI
N	Valid	507	507	507
	Missing	0	0	0
Mean		4357.495	8.3166	9.1344
Std. Error of Mean		65.537	1.667E-02	6.936E-03
Median		4200.000	8.3428	9.1297
Mode		3800.0	8.24	9.09
Std. Deviation		1475.670	.3754	.1562
Variance		2177601.8	.1409	2.439E-02
Skewness		.498	-1.110	.000
Std. Error of Skewness		.108	.108	.108
Kurtosis		.396	4.132	.192
Std. Error of Kurtosis		.217	.217	.217
Range		8800.0	3.14	.96
Minimum		400.0	5.99	8.60
Maximum		9200.0	9.13	9.56
Sum		2209250.0	4216.51	4631.15







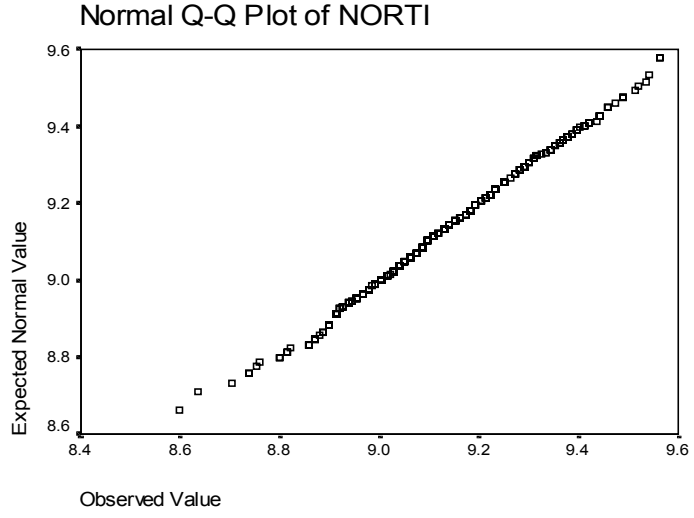
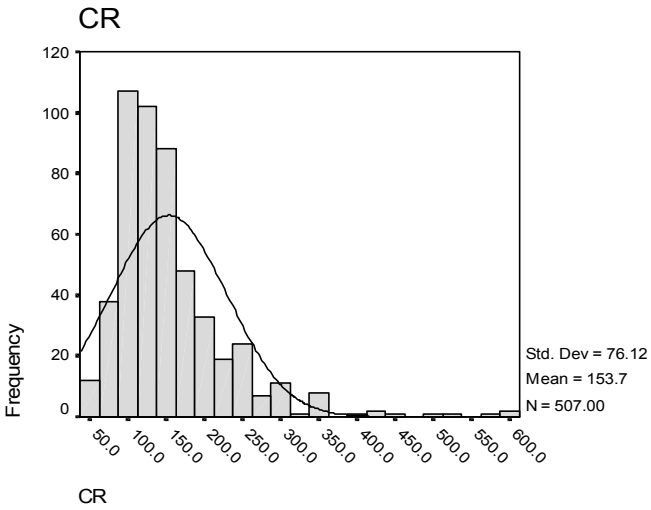
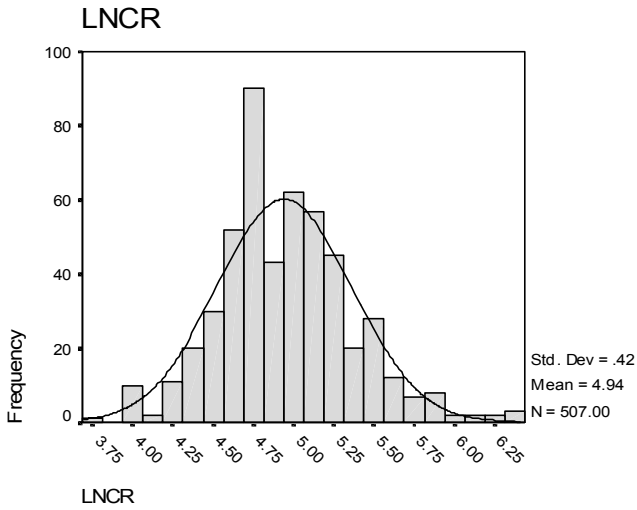


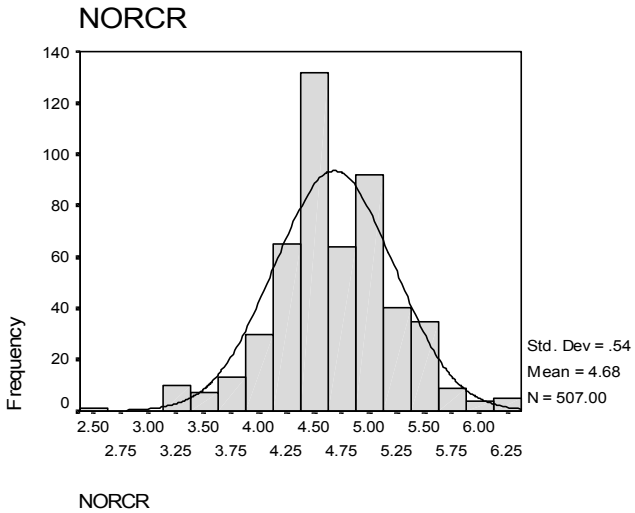
Fig (4-4): Statiatical Parameters For Raw Data

Statistics

		CR	LNCR	NORCR
N	Valid	507	507	507
	Missing	0	0	0
Mean		153.738	4.9413	4.6795
Std. Error of Mean		3.381	1.860E-02	2.396E-02
Median		130.000	4.8675	4.6161
Mode		120.0	4.79	4.51
Std. Deviation		76.118	.4188	.5394
Variance		5793.945	.1754	.2910
Skewness		2.332	.460	.000
Std. Error of Skewness		.108	.108	.108
Kurtosis		8.438	.693	.786
Std. Error of Kurtosis		.217	.217	.217
Range		558.0	2.66	3.77
Minimum		42.0	3.74	2.57
Maximum		600.0	6.40	6.35
Sum		77945.0	2505.23	2372.50







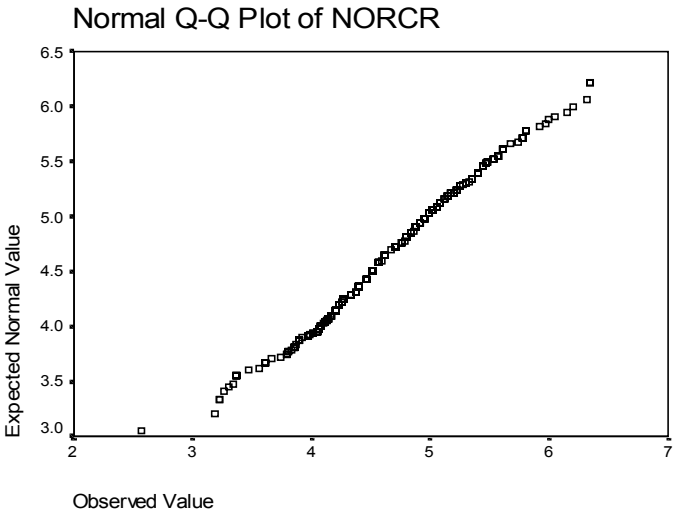
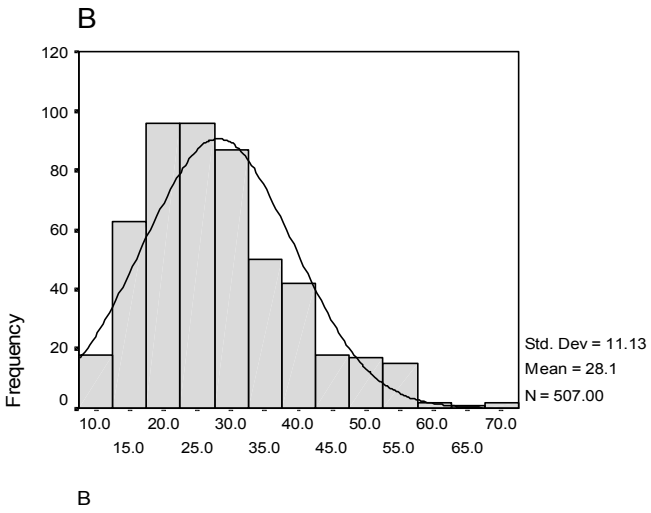
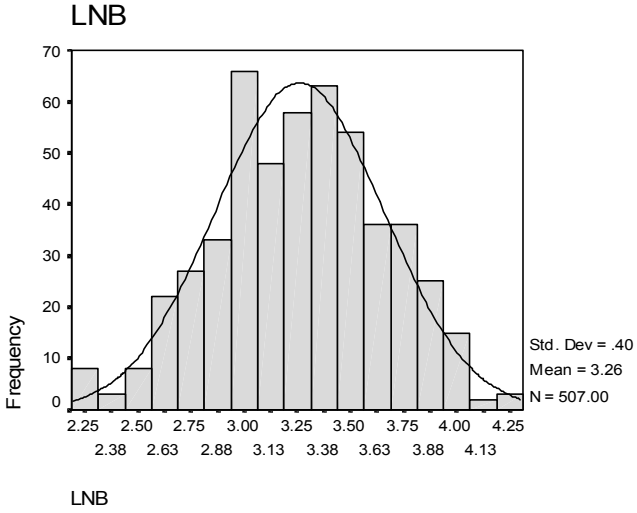


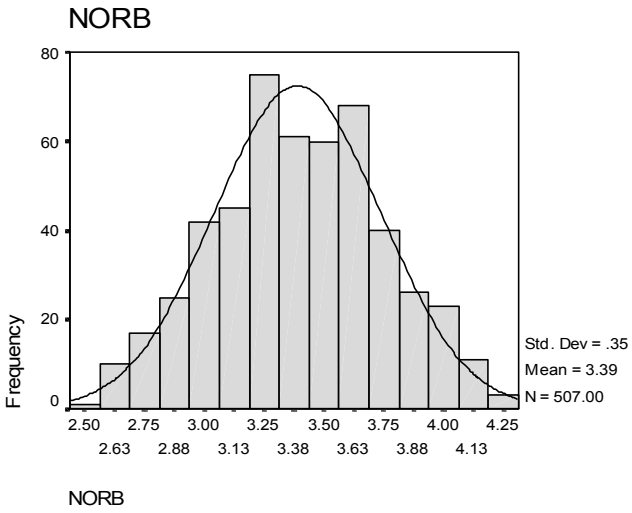
Fig (4-5) : Statiatical Parameters For Raw Data

Statistics

		B	LNB	NORB
N	Valid	507	507	507
	Missing	0	0	0
Mean		28.122	3.2597	3.3888
Std. Error of Mean		.494	1.760E-02	1.547E-02
Median		26.000	3.2581	3.3793
Mode		21.0	3.04	3.19
Std. Deviation		11.132	.3963	.3483
Variance		123.918	.1571	.1213
Skewness		.848	-.115	.000
Std. Error of Skewness		.108	.108	.108
Kurtosis		.625	-.351	-.404
Std. Error of Kurtosis		.217	.217	.217
Range		60.6	2.01	1.75
Minimum		9.4	2.24	2.55
Maximum		70.0	4.25	4.30
Sum		14258.0	1652.67	1718.14







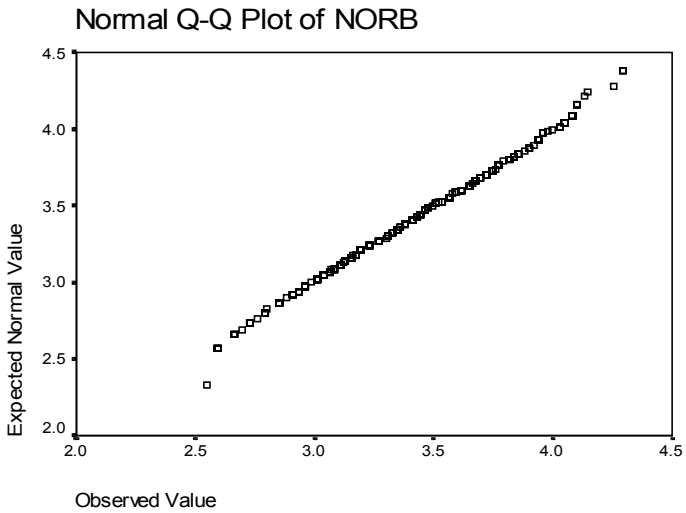
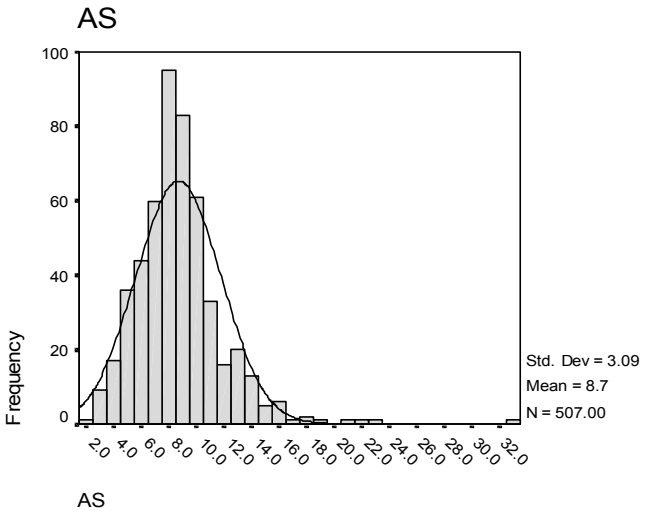
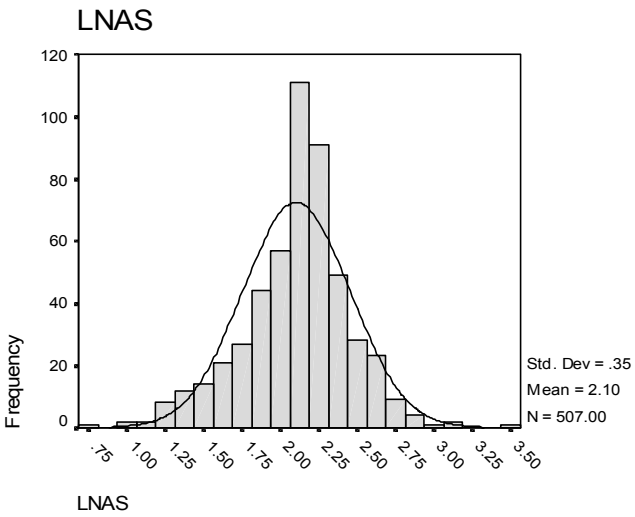


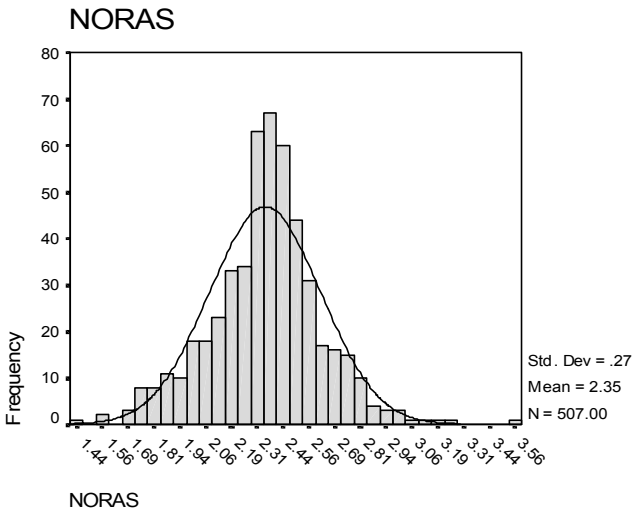
Fig (4-6) : Statiatical Prameters For Raw Data

Statistics

		AS	LNAS	NORAS
N	Valid	507	507	507
	Missing	0	0	0
Mean		8.701	2.1045	2.3504
Std. Error of Mean		.137	1.548E-02	1.197E-02
Median		8.440	2.1330	2.3627
Mode		10.5	2.35	2.54
Std. Deviation		3.091	.3486	.2696
Variance		9.556	.1215	7.270E-02
Skewness		1.744	-.360	.000
Std. Error of Skewness		.108	.108	.108
Kurtosis		8.962	1.359	1.278
Std. Error of Kurtosis		.217	.217	.217
Range		30.8	2.77	2.11
Minimum		2.1	.72	1.44
Maximum		32.9	3.49	3.56
Sum		4411.4	1066.96	1191.66







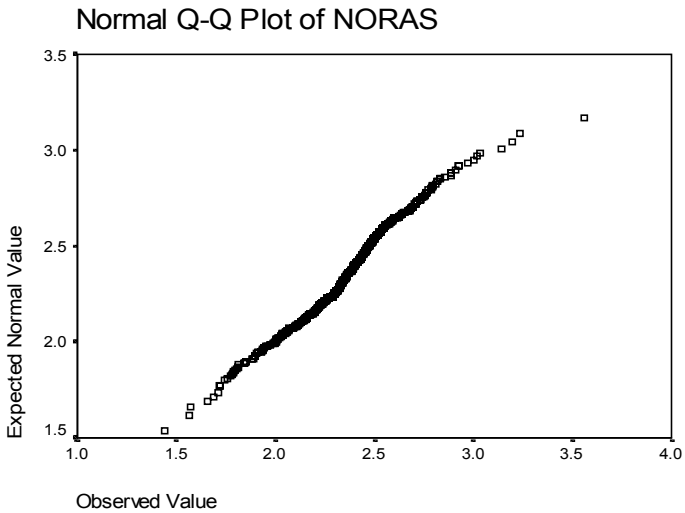
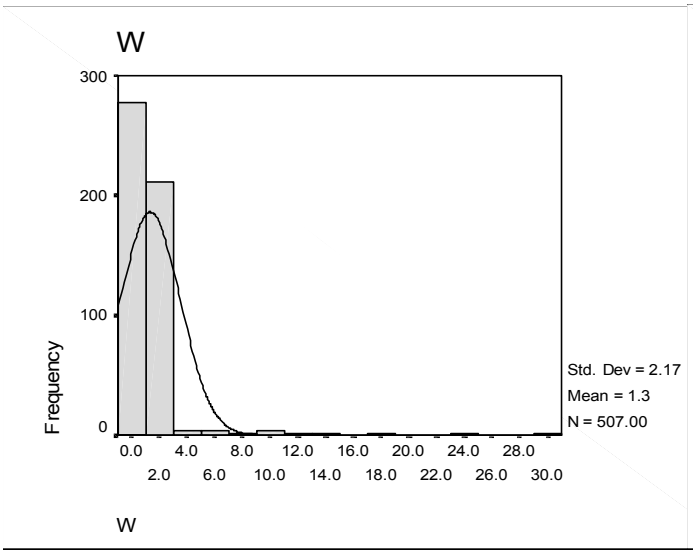
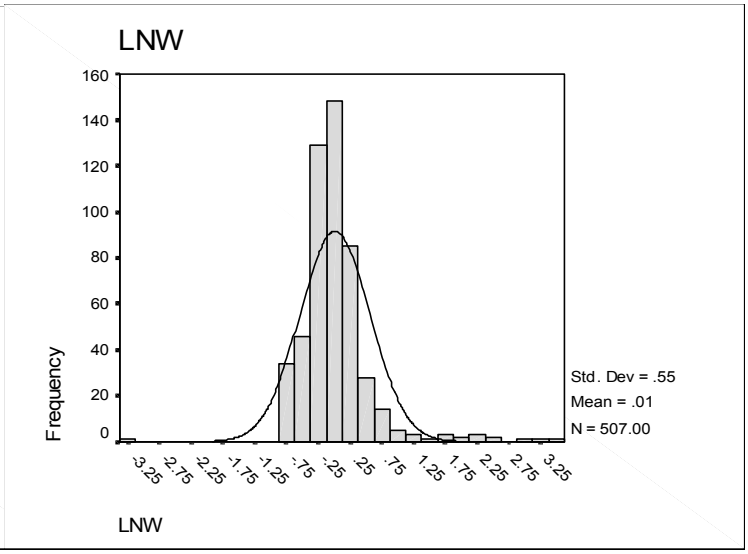


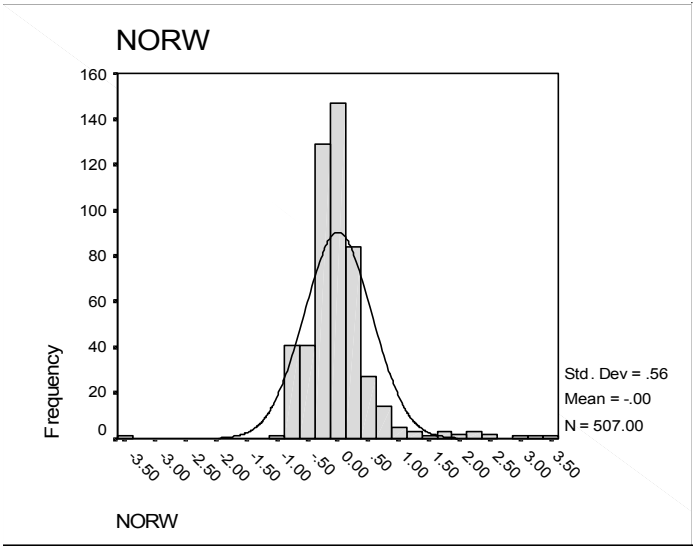
Fig (4-7) :Statistical Parameters Raw Data

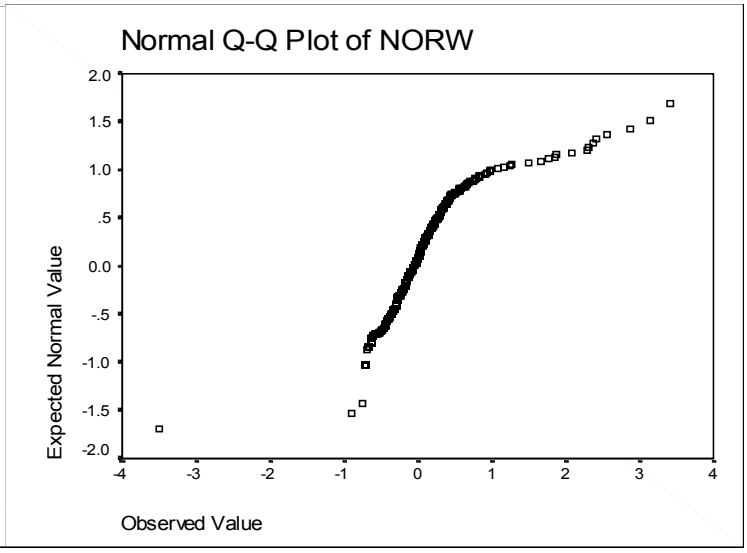
Statistics

		W	LNW	NORW
N	Valid	507	507	507
	Missing	0	0	0
Mean		1.315	1.075E-02	-7.51E-04
Std. Error of Mean		9.631E-02	2.450E-02	2.480E-02
Median		.950	-5.13E-02	-6.19E-02
Mode		.5	-.69	-.71
Std. Deviation		2.169	.5516	.5585
Variance		4.703	.3042	.3119
Skewness		8.977	1.922	1.771
Std. Error of Skewness		.108	.108	.108
Kurtosis		97.171	11.410	11.777
Std. Error of Kurtosis		.217	.217	.217
Range		30.6	6.64	6.93
Minimum		.0	-3.22	-3.51
Maximum		30.7	3.42	3.42
Sum		666.5	5.45	-.38









تعیین ضریب همبستگی :

برای تعیین اینکه آیا ارتباط معنی‌داری میان تغییرات متغیرهای آماری وجود دارد، ضرایب همبستگی میان آنها محاسبه می‌شود. این عمل به دو منظور کشف همبستگی بین متغیرها و تخمین مقدار یک یا چند متغیر دیگر صورت می‌گیرد. برای بررسی، دو نوع ضریب همبستگی پیرسون و اسپیرمن به صورت ماتریس ضرایب همبستگی محاسبه شده‌اند که در جداول (۴-۲) و (۴-۳) آمده است شرط محاسبه ضریب همبستگی پیرسون، نرمال بودن تابع توزیع متغیرها می‌باشد. دراین جداول، **Sig(2-Tailed)** میزان معنی‌دار بودن ضرایب همبستگی طبق آزمون فرض مساوی صفر بودن ضریب همبستگی می‌باشد.

برای محاسبه ضریب همبستگی پیرسون به علت تاثیرپذیری این پارامتر از آستانه‌های بالا و پایین حتماً باید داده‌های خام نرمال شوند تا ضریب همبستگی محاسبه شوند. جدول (۴-۲) مقادیر این ضرایب را نشان می‌دهد.

بر پایه جدول ضریب همبستگی پیرسون بین جفت متغیرهای **Sn,Ag(0.338)؛ Ni,Co(0.475) Mn,Be(0.348)؛Cr,Co(0.386)؛Mn,Co(0.393)؛ Ba,As(0.365)؛ As,W(0.312)؛ Zn,Sn(0.492)؛ B,Ba(0.328)؛** در سطح اعتماد مطلوب ۹۹٪ می‌باشد که بیشترین ارتباط همبستگی بین عناصر **Zn,Sn(0.492)** وجود دارد. این ضرایب بیانگر ارتباط پاراژنزی بین عناصر می‌باشند.

برای محاسبه ضریب همبستگی اسپیرمن از داده‌های خام استفاده شده است و همانطور که مشاهده می‌شود، در بعضی مواقع وضعیت متفاوتی نسبت به ضریب همبستگی پیرسون دارد. این اختلاف بیشتر زمانی بروز می‌کند که مقدار داده‌های خارج از رده زیاد باشد. اما مقایسه دقیق آنها، این نکته را بیان می‌کند که اختلاف این دو

ضریب همبستگی خیلی زیاد نیست ، این امر نشان دهنده تاثیرپذیری کم داده‌ها از مقادیر خارج از رده است. جدول (۴-۳) مقادیر این ضرایب را نشان می‌دهد.

Table(4-2) :Pearson Correlation for Normal Raw Data in Bazman 1:100000 Sheet

		NOR Au	NOR B	NOR Cu	NOR Pb	NOR Ag	NOR Sn	NOR Zn	NOR Mo	NOR W	NOR As	NOR Sb	NOR Bi	NOR Co	NOR Ni	NOR Be	NOR Ti	NOR Ba	NOR Mn	NOR Cr
NOR Au	Pearson Correlation	1.000	-0.025	0.029	-0.068	-0.120	0.019	0.038	0.108	0.219	0.186	0.048	-0.002	0.084	0.019	0.104	0.003	-0.011	0.103	-0.021
	Sig. (2-tailed)	.	0.580	0.521	0.127	0.007	0.677	0.398	0.015	0.000	0.000	0.278	0.970	0.058	0.672	0.019	0.942	0.800	0.021	0.632
NOR B	Pearson Correlation	-0.025	1.000	-0.049	0.151	0.099	0.005	0.129	-0.207	-0.143	0.143	-0.024	0.098	-0.321	-0.116	-0.026	-0.054	0.328	-0.052	-0.164
	Sig. (2-tailed)	0.580	.	0.267	0.001	0.025	0.904	0.004	0.000	0.001	0.001	0.588	0.027	0.000	0.009	0.566	0.226	0.000	0.241	0.000
NOR Cu	Pearson Correlation	0.029	-0.049	1.000	-0.153	-0.009	-0.005	0.083	0.206	0.096	0.114	0.047	0.277	0.253	0.161	0.043	0.057	-0.095	-0.031	0.160
	Sig. (2-tailed)	0.521	0.267	.	0.001	0.847	0.915	0.062	0.000	0.032	0.010	0.293	0.000	0.000	0.000	0.333	0.201	0.033	0.483	0.000
NOR Pb	Pearson Correlation	-0.068	0.151	-0.153	1.000	0.153	0.251	0.126	0.020	0.033	0.098	0.030	0.008	-0.171	0.023	0.004	0.039	0.263	-0.074	-0.093
	Sig. (2-tailed)	0.127	0.001	0.001	.	0.001	0.000	0.004	0.659	0.457	0.027	0.502	0.854	0.000	0.608	0.926	0.378	0.000	0.096	0.036
NOR Ag	Pearson Correlation	-0.120	0.099	-0.009	0.153	1.000	0.338	0.192	0.051	0.037	-0.021	-0.027	-0.027	-0.116	-0.118	-0.082	-0.045	-0.019	-0.199	0.052
	Sig. (2-tailed)	0.007	0.025	0.847	0.001	.	0.000	0.000	0.253	0.404	0.637	0.551	0.541	0.009	0.008	0.065	0.307	0.675	0.000	0.245
NOR Sn	Pearson Correlation	0.019	0.005	-0.005	0.251	0.338	1.000	0.492	0.208	0.175	-0.033	-0.014	-0.062	0.076	0.032	-0.021	0.082	-0.050	0.025	0.127
	Sig. (2-tailed)	0.677	0.904	0.915	0.000	0.000	.	0.000	0.000	0.000	0.456	0.758	0.160	0.086	0.473	0.638	0.064	0.262	0.567	0.004
NOR Zn	Pearson Correlation	0.038	0.129	0.083	0.126	0.192	0.492	1.000	0.274	0.207	0.065	0.024	0.014	0.098	0.047	0.020	0.098	0.015	-0.003	0.049
	Sig. (2-tailed)	0.398	0.004	0.062	0.004	0.000	0.000	.	0.000	0.000	0.144	0.590	0.755	0.027	0.288	0.653	0.027	0.734	0.954	0.268
NOR Mo	Pearson Correlation	0.108	-0.207	0.206	0.020	0.051	0.208	0.274	1.000	0.223	0.134	0.045	0.022	0.165	-0.008	-0.030	0.018	0.050	0.063	0.039
	Sig. (2-tailed)	0.015	0.000	0.000	0.659	0.253	0.000	0.000	.	0.000	0.003	0.308	0.618	0.000	0.853	0.501	0.681	0.259	0.154	0.382
NOR W	Pearson Correlation	0.219	-0.143	0.096	0.033	0.037	0.175	0.207	0.223	1.000	0.312	0.044	0.064	0.007	-0.001	0.108	0.073	0.011	-0.126	0.067
	Sig. (2-tailed)	0.000	0.001	0.032	0.457	0.404	0.000	0.000	0.000	.	0.000	0.318	0.151	0.869	0.987	0.015	0.103	0.811	0.005	0.129
NOR As	Pearson Correlation	0.186	0.143	0.114	0.098	-0.021	-0.033	0.065	0.134	0.312	1.000	0.229	0.183	-0.227	-0.115	0.009	0.000	0.365	-0.234	-0.154
	Sig. (2-tailed)	0.000	0.001	0.010	0.027	0.637	0.456	0.144	0.003	0.000	.	0.000	0.000	0.000	0.010	0.846	0.993	0.000	0.000	0.000
NOR Sb	Pearson Correlation	0.048	-0.024	0.047	0.030	-0.027	-0.014	0.024	0.045	0.044	0.229	1.000	0.120	0.031	0.060	0.115	0.165	-0.050	0.002	0.043
	Sig. (2-tailed)	0.278	0.588	0.293	0.502	0.551	0.758	0.590	0.308	0.318	0.000	.	0.007	0.490	0.176	0.010	0.000	0.260	0.973	0.328
NOR Bi	Pearson Correlation	-0.002	0.098	0.277	0.008	-0.027	-0.062	0.014	0.022	0.064	0.183	0.120	1.000	-0.073	0.002	0.135	-0.006	0.005	0.012	0.125
	Sig. (2-tailed)	0.970	0.027	0.000	0.854	0.541	0.160	0.755	0.618	0.151	0.000	0.007	.	0.100	0.967	0.002	0.901	0.910	0.789	0.005
NOR Co	Pearson Correlation	0.084	-0.321	0.253	-0.171	-0.116	0.076	0.098	0.165	0.007	-0.227	0.031	-0.073	1.000	0.475	0.246	0.253	-0.360	0.393	0.386
	Sig. (2-tailed)	0.058	0.000	0.000	0.000	0.009	0.086	0.027	0.000	0.869	0.000	0.490	0.100	.	0.000	0.000	0.000	0.000	0.000	0.000
NOR Ni	Pearson Correlation	0.019	-0.116	0.161	0.023	-0.118	0.032	0.047	-0.008	-0.001	-0.115	0.060	0.002	0.475	1.000	0.183	0.173	-0.188	0.234	0.244
	Sig. (2-tailed)	0.672	0.009	0.000	0.608	0.008	0.473	0.288	0.853	0.987	0.010	0.176	0.967	0.000	.	0.000	0.000	0.000	0.000	0.000
NOR Be	Pearson Correlation	0.104	-0.026	0.043	0.004	-0.082	-0.021	0.020	-0.030	0.108	0.009	0.115	0.135	0.246	0.183	1.000	0.102	-0.042	0.348	0.159
	Sig. (2-tailed)	0.019	0.566	0.333	0.926	0.065	0.638	0.653	0.501	0.015	0.846	0.010	0.002	0.000	0.000	.	0.022	0.347	0.000	0.000
NOR Ti	Pearson Correlation	0.003	-0.054	0.057	0.039	-0.045	0.082	0.098	0.018	0.073	0.000	0.165	-0.006	0.253	0.173	0.102	1.000	-0.017	0.139	0.083
	Sig. (2-tailed)	0.942	0.226	0.201	0.378	0.307	0.064	0.027	0.681	0.103	0.993	0.000	0.901	0.000	0.000	0.022	.	0.696	0.002	0.061
NOR Ba	Pearson Correlation	-0.011	0.328	-0.095	0.263	-0.019	-0.050	0.015	0.050	0.011	0.365	-0.050	0.005	-0.360	-0.188	-0.042	-0.017	1.000	-0.169	-0.263
	Sig. (2-tailed)	0.800	0.000	0.033	0.000	0.675	0.262	0.734	0.259	0.811	0.000	0.260	0.910	0.000	0.000	0.347	0.696	.	0.000	0.000
NOR Mn	Pearson Correlation	0.103	-0.052	-0.031	-0.074	-0.199	0.025	-0.003	0.063	-0.126	-0.234	0.002	0.012	0.393	0.234	0.348	0.139	-0.169	1.000	0.274
	Sig. (2-tailed)	0.021	0.241	0.483	0.096	0.000	0.567	0.954	0.154	0.005	0.000	0.973	0.789	0.000	0.000	0.000	0.002	0.000	.	0.000
NOR Cr	Pearson Correlation	-0.021	-0.164	0.160	-0.093	0.052	0.127	0.049	0.039	0.067	-0.154	0.043	0.125	0.386	0.244	0.159	0.083	-0.263	0.274	1.000
	Sig. (2-tailed)	0.632	0.000	0.000	0.036	0.245	0.004	0.268	0.382	0.129	0.000	0.328	0.005	0.000	0.000	0.000	0.061	0.000	0.000	.

Listwise N = 507

Table (4-3) :Spearman Correlation for Raw Data in Bazman 1:100000 Sheet

		Au	B	Cu	Pb	Ag	Sn	Zn	Mo	W	As	Sb	Bi	Co	Ni	Be	Ti	Ba	Mn	Cr
Au	Correlation Coefficient	1	-0.028	0.018	-0.085	-0.133	-0.007	0.03	0.106	0.146	0.19	0.101	-0.005	0.092	-0.005	0.107	0.028	-0.005	0.086	-0.034
	Sig. (2-tailed)	.	0.536	0.681	0.055	0.003	0.884	0.496	0.017	0.001	0	0.023	0.908	0.039	0.903	0.016	0.528	0.912	0.052	0.444
B	Correlation Coefficient	-0.028	1	-0.029	0.165	0.132	-0.012	0.143	-0.127	-0.109	0.126	0.054	0.082	-0.327	-0.101	-0.03	-0.065	0.356	-0.065	-0.153
	Sig. (2-tailed)	0.536	.	0.517	0	0.003	0.79	0.001	0.004	0.014	0.005	0.224	0.064	0	0.023	0.503	0.144	0	0.147	0.001
Cu	Correlation Coefficient	0.018	-0.029	1	-0.128	-0.001	0.002	0.065	0.162	0.099	0.069	0.014	0.353	0.23	0.132	0.047	0.031	-0.144	-0.018	0.192
	Sig. (2-tailed)	0.681	0.517	.	0.004	0.975	0.957	0.146	0	0.026	0.121	0.75	0	0	0.003	0.288	0.485	0.001	0.681	0
Pb	Correlation Coefficient	-0.085	0.165	-0.128	1	0.163	0.237	0.14	0.099	0.144	0.084	0.054	0.045	-0.196	0.045	-0.043	0.044	0.303	-0.09	-0.097
	Sig. (2-tailed)	0.055	0	0.004	.	0	0	0.002	0.025	0.001	0.058	0.227	0.314	0	0.307	0.337	0.328	0	0.043	0.029
Ag	Correlation Coefficient	-0.133	0.132	-0.001	0.163	1	0.381	0.179	0.075	0.097	-0.04	-0.004	-0.018	-0.17	-0.146	-0.069	-0.109	0.02	-0.205	-0.003
	Sig. (2-tailed)	0.003	0.003	0.975	0	.	0	0	0.092	0.03	0.369	0.925	0.692	0	0.001	0.121	0.014	0.654	0	0.95
Sn	Correlation Coefficient	-0.007	-0.012	0.002	0.237	0.381	1	0.533	0.246	0.245	0.007	-0.013	-0.033	0.042	0.022	-0.006	0.089	-0.06	0.022	0.127
	Sig. (2-tailed)	0.884	0.79	0.957	0	0	.	0	0	0	0.877	0.778	0.461	0.343	0.619	0.89	0.045	0.177	0.619	0.004
Zn	Correlation Coefficient	0.03	0.143	0.065	0.14	0.179	0.533	1	0.287	0.257	0.075	-0.018	0.013	0.087	0.077	0.045	0.097	0	0.029	0.05
	Sig. (2-tailed)	0.496	0.001	0.146	0.002	0	0	.	0	0	0.091	0.683	0.769	0.051	0.084	0.315	0.029	0.993	0.519	0.261
Mo	Correlation Coefficient	0.106	-0.127	0.162	0.099	0.075	0.246	0.287	1	0.194	0.127	0.061	0.037	0.135	-0.009	-0.073	0.025	0.058	0.042	0.007
	Sig. (2-tailed)	0.017	0.004	0	0.025	0.092	0	0	.	0	0.004	0.172	0.406	0.002	0.837	0.1	0.578	0.194	0.35	0.875
W	Correlation Coefficient	0.146	-0.109	0.099	0.144	0.097	0.245	0.257	0.194	1	0.332	0.142	0.135	-0.028	0.052	0.199	0.098	0.055	-0.141	0.083
	Sig. (2-tailed)	0.001	0.014	0.026	0.001	0.03	0	0	0	.	0	0.001	0.002	0.536	0.246	0	0.027	0.213	0.001	0.061
As	Correlation Coefficient	0.19	0.126	0.069	0.084	-0.04	0.007	0.075	0.127	0.332	1	0.316	0.159	-0.18	-0.116	-0.004	0.086	0.311	-0.215	-0.153
	Sig. (2-tailed)	0	0.005	0.121	0.058	0.369	0.877	0.091	0.004	0	.	0	0	0	0.009	0.936	0.052	0	0	0.001
Sb	Correlation Coefficient	0.101	0.054	0.014	0.054	-0.004	-0.013	-0.018	0.061	0.142	0.316	1	0.247	-0.064	-0.081	0.103	0.175	0.06	-0.038	-0.027
	Sig. (2-tailed)	0.023	0.224	0.75	0.227	0.925	0.778	0.683	0.172	0.001	0	.	0	0.148	0.069	0.021	0	0.177	0.389	0.543
Bi	Correlation Coefficient	-0.005	0.082	0.353	0.045	-0.018	-0.033	0.013	0.037	0.135	0.159	0.247	1	-0.069	0.029	0.136	0	-0.019	0.013	0.149
	Sig. (2-tailed)	0.908	0.064	0	0.314	0.692	0.461	0.769	0.406	0.002	0	0	.	0.122	0.508	0.002	0.999	0.674	0.764	0.001
Co	Correlation Coefficient	0.092	-0.327	0.23	-0.196	-0.17	0.042	0.087	0.135	-0.028	-0.18	-0.064	-0.069	1	0.451	0.218	0.274	-0.386	0.376	0.35
	Sig. (2-tailed)	0.039	0	0	0	0	0.343	0.051	0.002	0.536	0	0.148	0.122	.	0	0	0	0	0	0
Ni	Correlation Coefficient	-0.005	-0.101	0.132	0.045	-0.146	0.022	0.077	-0.009	0.052	-0.116	-0.081	0.029	0.451	1	0.167	0.186	-0.16	0.248	0.206
	Sig. (2-tailed)	0.903	0.023	0.003	0.307	0.001	0.619	0.084	0.837	0.246	0.009	0.069	0.508	0	.	0	0	0	0	0
Be	Correlation Coefficient	0.107	-0.03	0.047	-0.043	-0.069	-0.006	0.045	-0.073	0.199	-0.004	0.103	0.136	0.218	0.167	1	0.105	-0.058	0.342	0.169
	Sig. (2-tailed)	0.016	0.503	0.288	0.337	0.121	0.89	0.315	0.1	0	0.936	0.021	0.002	0	0	.	0.018	0.19	0	0
Ti	Correlation Coefficient	0.028	-0.065	0.031	0.044	-0.109	0.089	0.097	0.025	0.098	0.086	0.175	0	0.274	0.186	0.105	1	-0.034	0.122	0.121
	Sig. (2-tailed)	0.528	0.144	0.485	0.328	0.014	0.045	0.029	0.578	0.027	0.052	0	0.999	0	0	0.018	.	0.439	0.006	0.006
Ba	Correlation Coefficient	-0.005	0.356	-0.144	0.303	0.02	-0.06	0	0.058	0.055	0.311	0.06	-0.019	-0.386	-0.16	-0.058	-0.034	1	-0.139	-0.281
	Sig. (2-tailed)	0.912	0	0.001	0	0.654	0.177	0.993	0.194	0.213	0	0.177	0.674	0	0	0.19	0.439	.	0.002	0
Mn	Correlation Coefficient	0.086	-0.065	-0.018	-0.09	-0.205	0.022	0.029	0.042	-0.141	-0.215	-0.038	0.013	0.376	0.248	0.342	0.122	-0.139	1	0.310
	Sig. (2-tailed)	0.052	0.147	0.681	0.043	0	0.619	0.519	0.35	0.001	0	0.389	0.764	0	0	0	0.006	0.002	.	0
Cr	Correlation Coefficient	-0.034	-0.153	0.192	-0.097	-0.003	0.127	0.05	0.007	0.083	-0.153	-0.027	0.149	0.350	0.206	0.169	0.121	-0.281	0.310	1
	Sig. (2-tailed)	0.444	0.001	0	0.029	0.95	0.004	0.261	0.875	0.061	0.001	0.543	0.001	0	0	0	0.006	0	0	.

Listwise N = 507

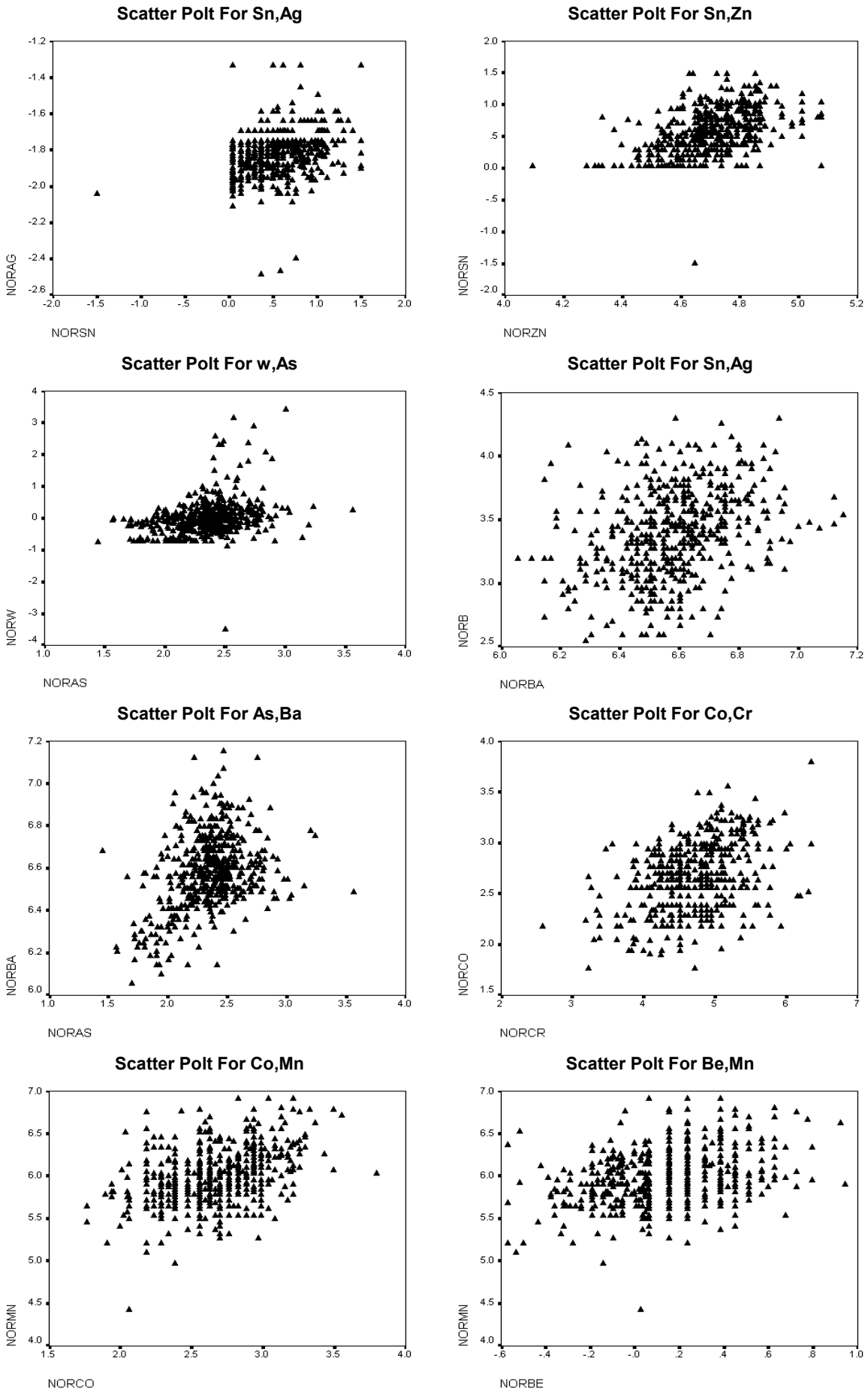
بـر پایـه ایـن جـدول ضریـب همبسـتگی مشـاهده شـده بیـن عنـاصر **Bi,Cu(0.353)** ؛ **Mn,Co(0.376)**؛ **Sb,As(0.316)**؛ **As,W(0.332)**؛ **Zn,Sn(0.533)**؛ **Ag,Sn(0.381)** ؛**Cr,Mn(0.310)**؛ **Co,Ni(0.451)**؛**Cr,Co(0.350)** در سـطح اعتمـاد ۹۹٪ می‌باشـد کـه بیشـترین ارتبـاط همبسـتگی بیـن عنـاصر **Zn,Sn(0.533)** وجـود دارد. ضریـب همبسـتگی بیـن جفـت متغیرهـا بـه روش پیرسـون و اسـپیرمن بیـانگر اختـلاف تقریبـا کـم بیـن ضرایـب همبستگی عنـاصر متنـاظر می‌باشـد کـه حکـایت از توزیـع نسـبتاً نرمـال عنـاصر و همیـن طـور عدم تـاثیر نمونـه‌های دور افتـاده دارد.

یکـی دیگـر از راهـهای بررسـی ارتبـاط تغیـیرات عنـاصر بـا یکدیگـر، رسـم نمـودار پراکنـش (**Scatter Plot**) می‌باشـد. زوج مرتب‌هـایی از مقـادیر دو متغـیر کـه دارای توزیـع دو متغـیره یکسـان باشـند بـر روی نمـودار دو بعـدی ترسـیم مـی‌گردنـد. هـر چـه پراکنـدگـی نقـاط در نمودارهای پراکنش بیشـتر باشـد پیونـد بیـن متغیرهـا ضعیف‌تـر اسـت. شـکل (۴-۸) پراکنـش مقـادیر داده‌هـای خـام نرمـال شـده بـرای چنـد زوج عنصـری اسـت کـه بیشـترین ارتبـاط را نشـان می‌دهـد. دراین نمودارهـا زوج عنصـر (**Zn,Sn**) بیشـترین همبسـتگی را بـا یکدیگـر نشان می‌دهـد.

بررسی‌های آماری چند متغیره:

هر تجزیـه و تحلیـل چنـد متغـیره کـه بـر روی بیـش از دو متغـیر انجـام گـیرد، می‌توانـد در قـالب آنالیزهـای چنـد متغـیره بیـان شـود. غـالب تکنیکـهای چنـد متغـیره در اصـل بسـط و توسـعه آنالیزهـای تـک متغـیره می‌باشـند و البتـه بعضـی از روشـهای چنـد متغـیره تنـها بـرای

Fig (4-8) : Scatter Plot for differnet Elements of Bazman (Normal Raw Data)



پاسخگویی به مقاصد چند متغیره طراحی شده‌اند که از جمله این روشها می‌توان به آنالیز فاکتوری اشاره کرد.

تجربه نشان داده است که چنانچه ترکیبی از متغیرها به جای یک متغیر به کار گرفته شوند و از نتایج ترکیبی آنها استفاده شود امکان تشخیص هاله‌های مرکب ژئوشیمیایی در اطراف توده‌های کانساری به مراتب افزایش می‌یابد. و از طرفی اثرات خطاهای تصادفی در بکارگیری ترکیبی متغیرها نسبتاً کاهش می‌یابد. از دیگر مزایای استفاده از روشهای چند متغیره، کاهش تعداد متغیرها در مباحث داده‌پردازی و درنتیجه کاستن از تعداد نقشه‌هاست. با استفاده از این روشها امکان مقایسه متغیرها و کسب نتایج راحت‌تر خواهد بود. البته استفاده بهینه از روشهای چند متغیره در حالتی صادق خواهد بود که در پردازش داده‌ها با تعداد زیادی متغیر روبرو باشیم و تا حدودی امکان اخذ نتیجه از متغیرها به گونه منفرد غیر ممکن و یا توام با خطای زیاد باشد. دراین گزارش از روشهای چند متغیره مانند روشهای آنالیز خوشه‌ای و آنالیز فاکتوری و ... استفاده شده است.

آنالیز خوشه‌ای و تفسیر آن:

به دلیل اینکه هر گروه از عناصر نسبت به یکسری از شرایط محیطی کم و بیش به طور مشابه حساسیت نشان می‌دهند، شناخت ارتباط و همبستگی ژنتیکی متقابل بین عناصر مختلف می‌تواند در شناخت دقیق‌تر تغییرات موجود در محیطهای ژئوشیمیایی به کار گرفته شود. ضمناً تجمع ژنتیکی بعضی از عناصر ممکن است به عنوان راهنمای مستقیم در تفسیر نوع نهشته‌ای که احتمالاً در ناحیه وجود دارد، به کار رود. در کل

شناخت همبستگی ژنتیکی که در بین عناصر وجود دارد اطلاعات لازم را برای تفسیر هر چه صحیح‌تر داده‌های ژئوشیمیایی در اختیار می‌گذارد.

آنالیز خوشه‌ای یک روش آماری چند متغیره است که عناصر را بر اساس شباهت تغییرپذیری بین آنها در قالب دسته‌ها یا گروههایی طبقه‌بندی می‌کند. دلایل زیادی برای ارزشمند بودن آنالیز خوشه‌ای وجود دارد، از جمله اینکه آنالیز خوشه‌ای می‌تواند در یافتن گروههای واقعی کمک کند و همچنین باعث کاهش تراکم داده‌ها شود. البته باید توجه داشت که آنالیز خوشه‌ای می‌تواند گروه‌های غیر قابل انتظاری را نیز ایجاد نماید که بیانگر روابط جدیدی خواهند بود و باید مورد بررسی قرار گیرند. در روش آنالیز خوشه‌ای از داده‌های خام نرمال شده استفاده شده است تا اثر مقادیر غیر همساز از جامعه اصلی و نیز اثر تغییر مقیاس داده‌ها از میان برود. نتایج حاصل از آنالیز خوشه‌ای عناصر مورد مطالعه در شکل (۴-۹) آورده شده است. با توجه به شکل می‌توان سه گروه اصلی را جدا نمود که بیانگر ارتباط پاراژنزی بین متغیرها باشد.

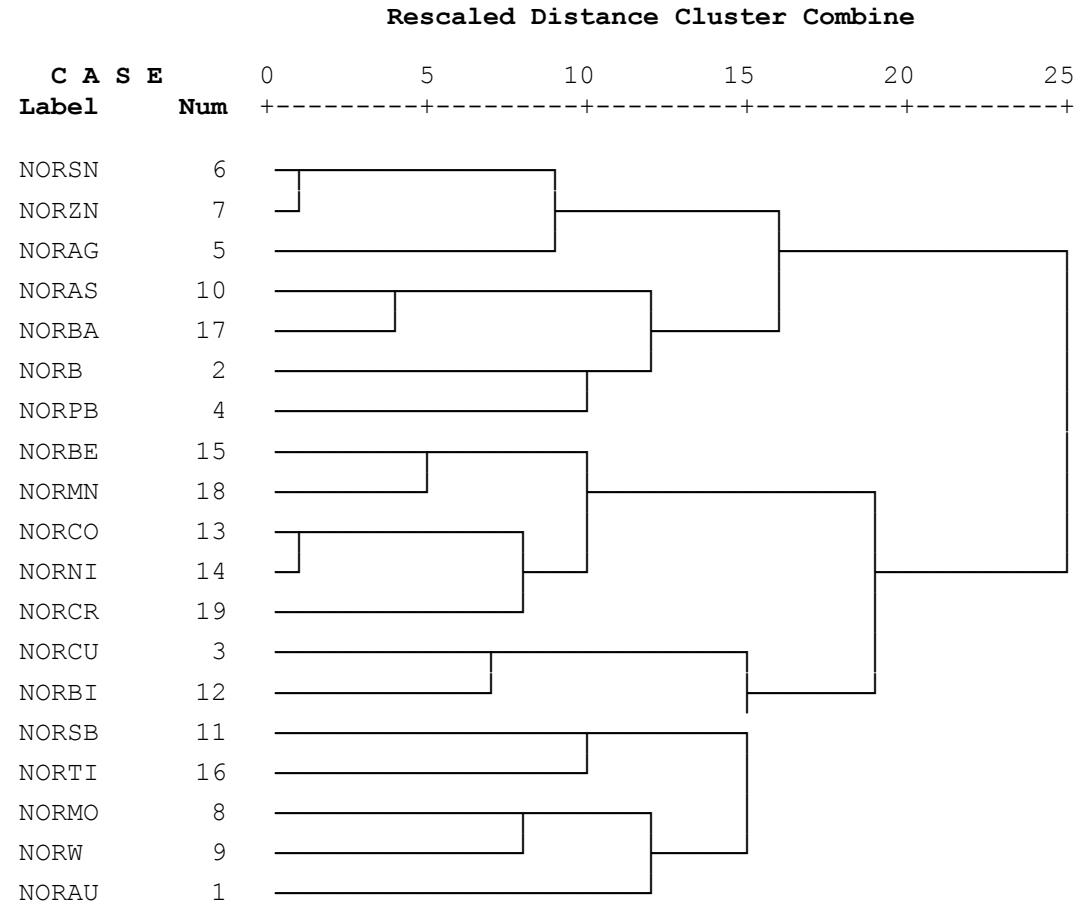
گروه اول: شامل عناصر **Sn,Zn,Ag,As,Ba,B,Pb**

گروه دوم: شامل عناصر **Be,Mn,Co,Ni,Cr**

گروه سوم: شامل عناصر **Cu,Bi,Sb,Ti,Mo,W,Au**

Fig (4-9) : Cluster Analyse for Normal Raw Data

Dendrogram using Complete Linkage



محاسبه پارامترهای آماری داده‌های خام

اولین مرحلهٔ پردازش داده‌های ژئوشیمیایی، بررسی پارامترهای آماری مربوط به تک‌تک عناصر جهت شناخت ماهیت توزیع هریک از آنها می‌باشد که با محاسبهٔ پارامترهای آماری از قبیل میانگین، انحراف معیار، چولگی، کشیدگی، واریانس و . . . می‌توان به این موضوع دست‌یافت. در این قسمت برای هر عنصر به عنوان یک متغیر آماری در یک جدول، تعداد نمونه‌ها، حداقل و حداکثر عیار، میانگین، میانه، انحراف معیار، چولگی و کشیدگی و نمودارهای هیستوگرام توزیع فراوانی محاسبه و ترسیم شده‌اند.

بررسی مقادیر خارج از رده : (Outliers)

هنگام بررسی مقادیر داده‌های خام به نمونه‌هایی برخورد می‌شود که در آستانه‌های بالا و پایین جامعه داده‌ها قرار گرفته و از جامعه اصلی جدا افتاده‌اند. اگر نمودار جعبه‌ای (**Boxplot**) آنها ترسیم شود این نمونه‌ها به نحو بارزی خودشان را از بقیه جدا می‌کنند. مقادیر خارج از رده به سه حالت مختلف زیر ممکن است بوجود آیند:

حالت اول) از یک خطای سیستماتیک به هنگام نمونه‌برداری، آماده‌سازی یا تجزیه شیمیایی نمونه‌ها ناشی شده باشند که باید از مرحله پردازش حذف یا اصلاح شوند.

حالت دوم) مشاهداتی که به صورت یک پدیده فوق‌العاده نمود پیدا می‌کنند که باید پس از بررسی اعتبار آنها در مورد حفظ یا حذف آنها تصمیم گرفت.

حالت سوم) مشاهدات فوق‌العاده‌ای که هیچگونه توضیح مناسبی برای آنها وجود ندارد و کارشناس اگر احساس کند که آنها به عنوان گوشه‌ای از جامعه مورد بررسی هستند می‌تواند آنها را حفظ کند.

وجود مقادیر خارج از رده در جامعه نمونه‌ها موجب افزایش واریانس جامعه و نیز همبستگی بین متغیرها وهمچنین افزایش چولگی در نمودار توزیع عناصر می‌شود. برای کاهش این تاثیر راههای مختلفی نظیر محاسبه ضریب همبستگی با استفاده از روشهای ناپارامتری مانند روش اسپیرمن **(Spearman)** ، حذف یا جایگزین نمودن مقادیر استفاده می‌شود دراین گزارش از روش جایگزین نمودن مقادیر خارج از رده استفاده شده است. جدول (۴-۱) نمونه‌های دارای مقادیر خارج از رده را نشان می‌دهد.

نرمال سازی داده‌های خام :

استفاده از برخی روشهای آماری منوط به نرمال بودن تابع توزیع متغیرهای مورد مطالعه است در حالیکه توابع توزیع از نوع لاگ نرمال است، به همین علت قبل از استفاده از این روشها داده‌های خام باید نرمال شوند. دراین بخش از نوعی تبدیلات جهت نرمال کردن تابع توزیع داده‌های خام استفاده شده است . این کار شرط لازم کاربرد برخی روشهای آماری مانند تعیین نمونه‌های آنومالی با استفاده از اضافه کردن ضرایبی از انحراف معیار به حد آستانه‌ای و یا محاسبه ضرایب همبستگی پیرسون می‌باشد. روش لاگ نرمال به صورت یک روش توصیفی برای نرمال کردن تابع توزیع جوامعی که دارای چولگی در نمودار خود هستند به کار می رود.

در اینجا از لگاریتم طبیعی مقادیر داده‌های خام به اضافه یا منهای یک مقدار ثابت λ مطابق رابطه تبدیلی زیر استفاده شده است.

$$Z = \mathrm{Ln}\,(AE \pm \lambda)$$

در این رابطه AE آنالیز نمونه برای هر عنصر است.

برای هر عنصر مقدار λ به گونه‌ای انتخاب می‌شود که پس از انتخاب داده‌ها به یک مقدار بهینه از چولگی و کشیدگی در منحنی توزیع نرمال دست یافته شود. پارامترهای آماری و هیستوگرام‌های ترسیم شده برای داده‌های نرمال در شکل (۴-۱) تا (۴-۷) آورده شده است. با توجه به این پارامترهای آماری می‌توان دریافت که مقادیر چولگی و کشیدگی متغیرها در مقایسه با مقادیر متناظر مربوط به داده‌های خام نرمال نشده تا چه اندازه کاهش یافته و منحنی توزیع تجمعی آنها به صورت یک خط راست که بیانگر توزیع نرمال می‌باشد، ظاهر شده است. هیستوگرام مقادیر نرمال شده نسبت به هیستوگرام مقادیر نرمال نشده نیز بیانگر مطلب فوق می‌باشد.

تعیین ضریب همبستگی :

برای تعیین اینکه آیا ارتباط معنی‌داری میان تغییرات متغیرهای آماری وجود دارد، ضرایب همبستگی میان آنها محاسبه می‌شود. این عمل به دو منظور کشف همبستگی بین متغیرها و تخمین مقدار یک یا چند متغیر دیگر صورت می‌گیرد. برای بررسی، دو نوع ضریب همبستگی پیرسون و اسپیرمن به صورت ماتریس ضرایب همبستگی محاسبه شده‌اند که در جداول (۴-۲) و (۴-۳) آمده است شرط محاسبه ضریب همبستگی پیرسون، نرمال بودن تابع توزیع متغیرها می‌باشد. دراین جداول، **Sig(2-Tailed)** میزان معنی‌دار بودن ضرایب همبستگی طبق آزمون فرض مساوی صفر بودن ضریب همبستگی می‌باشد.

برای محاسبه ضریب همبستگی پیرسون به علت تاثیرپذیری این پارامتر از آستانه‌های بالا و پایین حتماً باید داده‌های خام نرمال شوند تا ضریب همبستگی محاسبه شوند. جدول (۴-۲) مقادیر این ضرایب را نشان می‌دهد.

بر پایه جدول ضریب همبستگی پیرسون بین جفت متغیرهای **Sn,Ag(0.338)؛ Ni,Co(0.475) Mn,Be(0.348)؛Cr,Co(0.386)؛Mn,Co(0.393)؛ Ba,As(0.365)؛ As,W(0.312)؛ Zn,Sn(0.492)؛ B,Ba(0.328)؛** در سطح اعتماد مطلوب ۹۹٪ می‌باشد که بیشترین ارتباط همبستگی بین عناصر **Zn,Sn(0.492)** وجود دارد. این ضرایب بیانگر ارتباط پاراژنزی بین عناصر می‌باشند.

برای محاسبه ضریب همبستگی اسپیرمن از داده‌های خام استفاده شده است و همانطور که مشاهده می‌شود، در بعضی مواقع وضعیت متفاوتی نسبت به ضریب همبستگی پیرسون دارد. این اختلاف بیشتر زمانی بروز می‌کند که مقدار داده‌های خارج از رده زیاد باشد. اما مقایسه دقیق آنها، این نکته را بیان می‌کند که اختلاف این دو

ضریب همبستگی خیلی زیاد نیست ، این امر نشان دهنده تاثیرپذیری کم داده‌ها از مقادیر خارج از رده است. جدول (۴-۳) مقادیر این ضرایب را نشان می‌دهد.

بـر پایـه ایـن جـدول ضریـب همبسـتگی مشـاهده شـده بیـن عنـاصر **Bi,Cu(0.353)** ؛ **Mn,Co(0.376)**؛ **Sb,As(0.316)**؛ **As,W(0.332)**؛ **Zn,Sn(0.533)**؛ **Ag,Sn(0.381)** ؛**Cr,Mn(0.310)**؛ **Co,Ni(0.451)**؛**Cr,Co(0.350)** در سـطح اعتمـاد ۹۹٪ می‌باشـد کـه بیشـترین ارتبـاط همبسـتگی بیـن عنـاصر **Zn,Sn(0.533)** وجـود دارد. ضریـب همبسـتگی بیـن جفـت متغیرهـا بـه روش پیرسـون و اسـپیرمن بیـانگر اختـلاف تقریبـا کـم بیـن ضرایـب همبستگی عنـاصر متنـاظر می‌باشـد کـه حکـایت از توزیـع نسـبتاً نرمـال عنـاصر و همیـن طـور عدم تـاثیر نمونـه‌های دور افتـاده دارد.

یکـی دیگـر از راهـهای بررسـی ارتبـاط تغیـیرات عنـاصر بـا یکدیگـر، رسـم نمـودار پراکنـش (**Scatter Plot**) می‌باشـد. زوج مرتب‌هـایی از مقـادیر دو متغـیر کـه دارای توزیـع دو متغـیره یکسـان باشـند بـر روی نمـودار دو بعـدی ترسـیم مـی‌گردنـد. هـر چـه پراکنـدگـی نقـاط در نمودارهای پراکنش بیشـتر باشـد پیونـد بیـن متغیرهـا ضعیف‌تـر اسـت. شـکل (۴-۸) پراکنـش مقـادیر داده‌هـای خـام نرمـال شـده بـرای چنـد زوج عنصـری اسـت کـه بیشـترین ارتبـاط را نشـان می‌دهـد. دراین نمودارهـا زوج عنصـر (**Zn,Sn**) بیشـترین همبسـتگی را بـا یکدیگـر نشان می‌دهـد.

بررسی‌های آماری چند متغیره:

هر تجزیـه و تحلیـل چنـد متغـیره کـه بـر روی بیـش از دو متغـیر انجـام گـیرد، می‌توانـد در قـالب آنالیزهـای چنـد متغـیره بیـان شـود. غـالب تکنیکـهای چنـد متغـیره در اصـل بسـط و توسـعه آنالیزهـای تـک متغـیره می‌باشـند و البتـه بعضـی از روشـهای چنـد متغـیره تنـها بـرای

پاسخگویی به مقاصد چند متغیره طراحی شده‌اند که از جمله این روشها می‌توان به آنالیز فاکتوری اشاره کرد.

تجربه نشان داده است که چنانچه ترکیبی از متغیرها به جای یک متغیر به کار گرفته شوند و از نتایج ترکیبی آنها استفاده شود امکان تشخیص هاله‌های مرکب ژئوشیمیایی در اطراف توده‌های کانساری به مراتب افزایش می‌یابد. و از طرفی اثرات خطاهای تصادفی در بکارگیری ترکیبی متغیرها نسبتاً کاهش می‌یابد. از دیگر مزایای استفاده از روشهای چند متغیره، کاهش تعداد متغیرها در مباحث داده‌پردازی و درنتیجه کاستن از تعداد نقشه‌هاست. با استفاده از این روشها امکان مقایسه متغیرها و کسب نتایج راحت‌تر خواهد بود. البته استفاده بهینه از روشهای چند متغیره در حالتی صادق خواهد بود که در پردازش داده‌ها با تعداد زیادی متغیر روبرو باشیم و تا حدودی امکان اخذ نتیجه از متغیرها به گونه منفرد غیر ممکن و یا توام با خطای زیاد باشد. دراین گزارش از روشهای چند متغیره مانند روشهای آنالیز خوشه‌ای و آنالیز فاکتوری و ... استفاده شده است.

آنالیز خوشه‌ای و تفسیر آن:

به دلیل اینکه هر گروه از عناصر نسبت به یکسری از شرایط محیطی کم و بیش به طور مشابه حساسیت نشان می‌دهند، شناخت ارتباط و همبستگی ژنتیکی متقابل بین عناصر مختلف می‌تواند در شناخت دقیق‌تر تغییرات موجود در محیطهای ژئوشیمیایی به کار گرفته شود. ضمناً تجمع ژنتیکی بعضی از عناصر ممکن است به عنوان راهنمای مستقیم در تفسیر نوع نهشته‌ای که احتمالاً در ناحیه وجود دارد، به کار رود. در کل

شناخت همبستگی ژنتیکی که در بین عناصر وجود دارد اطلاعات لازم را برای تفسیر هر چه صحیح‌تر داده‌های ژئوشیمیایی در اختیار می‌گذارد.

آنالیز خوشه‌ای یک روش آماری چند متغیره است که عناصر را بر اساس شباهت تغییرپذیری بین آنها در قالب دسته‌ها یا گروههایی طبقه‌بندی می‌کند. دلایل زیادی برای ارزشمند بودن آنالیز خوشه‌ای وجود دارد، از جمله اینکه آنالیز خوشه‌ای می‌تواند در یافتن گروههای واقعی کمک کند و همچنین باعث کاهش تراکم داده‌ها شود. البته باید توجه داشت که آنالیز خوشه‌ای می‌تواند گروه‌های غیر قابل انتظاری را نیز ایجاد نماید که بیانگر روابط جدیدی خواهند بود و باید مورد بررسی قرار گیرند. در روش آنالیز خوشه‌ای از داده‌های خام نرمال شده استفاده شده است تا اثر مقادیر غیر همساز از جامعه اصلی و نیز اثر تغییر مقیاس داده‌ها از میان برود. نتایج حاصل از آنالیز خوشه‌ای عناصر مورد مطالعه در شکل (۴-۹) آورده شده است. با توجه به شکل می‌توان سه گروه اصلی را جدا نمود که بیانگر ارتباط پاراژنزی بین متغیرها باشد.

گروه اول: شامل عناصر **Sn,Zn,Ag,As,Ba,B,Pb**

گروه دوم: شامل عناصر **Be,Mn,Co,Ni,Cr**

گروه سوم: شامل عناصر **Cu,Bi,Sb,Ti,Mo,W,Au**