The Descriptive statistic

1)The raw data analysis:

The sample data: $X_1,...,X_n$ and $X_{(1)},X_{(2)},...,X_{(n)}$ is after ascending sorted.

1.1)Sample mean:

$$X_1+...+X_n=X_{(1)}+X_{(2)}+...+X_{(n)},\overline{X}=\frac{X_{(1)}+X_{(2)}+...+X_{(n)}}{n}.$$

1.2) Sample Geometric mean:

$$\prod_{i=1}^{n}X_i=\prod_{i=1}^{n}X_{(i)},G=\left(\prod_{i=1}^{n}X_{(i)}\right)^{\frac{1}{n}}.$$

1.3) Sample Harmonic mean:

$$\sum_{i=1}^{n}\frac{1}{X_i}=\sum_{i=1}^{n}\frac{1}{X_{(i)}},H=\frac{n}{\sum_{i=1}^{n}\frac{1}{X_{(i)}}}.$$

1.4)Sample median:

Sample mean= $X_{\left(\frac{n+1}{2}\right)}$, n is odd number,

Sample mean= $\frac{x_{\left(\frac{n}{2}\right)}+x_{\left(\frac{n+2}{2}\right)}}{2}$, n is even number.

1.5)Sample quartile:

$$Q_1=X_{\left(\frac{n}{4}+\frac{1}{2}\right)},Q_2=X_{\left(\frac{2n}{4}+\frac{1}{2}\right)},Q_3=X_{\left(\frac{3n}{4}+\frac{1}{2}\right)}.$$

1.6)Sample Midrange:

$$Midrange=\frac{X_{(1)}+X_{(n)}}{2}.$$

1.7)Sample Variance:

$$S^2=\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-1}.$$

1.8)Sample Standard Deviation:

$$S=\sqrt{\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-1}}.$$

1.9)Sample MAD:

$$MAD=\frac{\sum_{i=1}^{n}\left|X_i-\overline{X}\right|}{n}.$$

1.10)Sample Range:

$$Range=X_{(n)}-X_{(1)}.$$

1.11)Sample coefficient of skewed:

$$\frac{\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{S}\right)^3}{n}.$$

1.12)Sample coefficient of kurtosis:

$$\frac{\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{S}\right)^4}{n}.$$

2)The frequency distribtuion:

2.1)The process:

step 1：sample dat $X_1,...,X_n$ and $X_{(1)},X_{(2)},...,X_{(n)}$ is after ascending sorted.

step 2： =R= $Max(X_1,...,X_n)-Min(X_1,...,X_n)=X_{(n)}-X_{(1)}$,

step 3：number of class=k= $\log_2 n+1$,

step 4：class width=c= R/k ,

step 5：class limit.

step 6：the times of each class。

2.2)The table:

class	Class low limit ~Class upper limit	Class midpoint	frequency	Relative frequency	Cumulative relative frequency
1	$C_0 \sim C_1$	$m_1=\frac{C_0+C_1}{2}$	f_1	f_1/n	f_1/n
2	$C_1 \sim C_2$	$m_2=\frac{C_1+C_2}{2}$	f_2	f_2/n	$(f_1+f_2)/n$
…	…	…	…	…	…
k	$C_{k-1} \sim C_k$	$m_k=\frac{C_{k-1}+C_k}{2}$	f_k	f_k/n	1

2.3)coefficient：

sample mean： $\overline{X}=\frac{\sum_{i=1}^{k} m_i \times f_i}{n}$,

sample variance： $S^2=\frac{\sum_{i=1}^{k} m_i^2 \times f_i}{n}-\overline{X}^2=\frac{\sum_{i=1}^{k} m_i^2 \times f_i-\left(\sum_{i=1}^{k} m_i \times f_i\right)^2 \Big/ n}{n}$,

3)X1~Normal(mu=10,sigma*sigma=25),

X1
7.7201682102
-2.2894789917
4.0536320802
7.2119291036
19.5167656440
10.0788292012
17.1165443004
12.9333560933
3.9731342997
9.2312149618
3.9021027245
11.9064074021
15.3635895673
9.6402589370
4.2291059631
10.0825246374
-2.8015363963
17.6000152676
8.0492817656
2.9747309312
4.4795657566
15.7269991956
6.0978695650
7.4711976174
11.5224422489
14.0607663424
1.0811032008
-0.2337120745
4.7063227582
9.7465338352
7.5477445702
11.0206913959
12.8942708088
10.4459062337
8.7462432235
0.1816695265
12.0706359217
12.9596133696
3.9285453351
8.1220133674
4.2926775046
16.5786905196
10.2114330986
9.0152326204
4.7691006482
16.5533720126
3.2332007419
2.4168761454
4.4168062304
6.1156229436
22.2421828344
10.2894725615
3.6152019322
16.5351071844
9.5307881990
10.9228748573
-1.3503730666
14.6878354104
3.8047138217
11.3653277906
11.2513175878
-0.3223120820
9.3756760099
19.5565389599
9.9847419996
14.5191075152
12.0443732821
20.9921734918
4.7966027963
11.2505580734
14.5865189766
13.1853083353
12.7582212676

6.9453765282
11.2771214919
13.3188089263
7.7250651883
1.0256642807
6.2106671259
12.3660483105
6.7408731923
8.0829208609
9.2577693010
13.5628624652
11.1965103144
3.1182088230
8.0214910304
2.3787230835
5.5636009868
17.0492155487
10.5481247237
13.7773725944
8.3302576499
8.4419273511
12.7238465371
18.1992372610
4.5932064719
7.1398807863
14.3356323131
4.1553942982

X1 is Normal(mu=10.000000,sigma*sigma=25.000000),

X1 is mean= 9.1237977555, s.d.= 5.3422143406, variance= 28.5392540614,
skewed coefficient= 0.0576100743, kurtosis coefficient= 2.5718373720, MAD= 4.3102810367,
Q1= 4.6497646150, median= 9.3167226555, Q3= 12.8942708088,
MIN= -2.8015363963, MAX= 22.2421828344, Range= 25.0437192307,
Mid-Range= 9.7203232190, C.V.= 0.5855252915, sample size=100

after storing the sample data is below

	X1
1	-2.8015363963
2	-2.2894789917
3	-1.3503730666
4	-0.3223120820
5	-0.2337120745
6	0.1816695265
7	1.0256642807
8	1.0811032008
9	2.3787230835
10	2.4168761454
11	2.9747309312
12	3.1182088230
13	3.2332007419
14	3.6152019322
15	3.8047138217
16	3.9021027245
17	3.9285453351
18	3.9731342997
19	4.0536320802
20	4.1553942982
21	4.2291059631
22	4.2926775046
23	4.4168062304
24	4.4795657566
25	4.5932064719
26	4.7063227582
27	4.7691006482
28	4.7966027963
29	5.5636009868
30	6.0978695650
31	6.1156229436
32	6.2106671259
33	6.7408731923
34	6.9453765282
35	7.1398807863
36	7.2119291036
37	7.4711976174

38	7.5477445702
39	7.7201682102
40	7.7250651883
41	8.0214910304
42	8.0492817656
43	8.0829208609
44	8.1220133674
45	8.3302576499
46	8.4419273511
47	8.7462432235
48	9.0152326204
49	9.2312149618
50	9.2577693010
51	9.3756760099
52	9.5307881990
53	9.6402589370
54	9.7465338352
55	9.9847419996
56	10.0788292012
57	10.0825246374
58	10.2114330986
59	10.2894725615
60	10.4459062337
61	10.5481247237
62	10.9228748573
63	11.0206913959
64	11.1965103144
65	11.2505580734
66	11.2513175878
67	11.2771214919
68	11.3653277906
69	11.5224422489
70	11.9064074021
71	12.0443732821
72	12.0706359217
73	12.3660483105
74	12.7238465371
75	12.7582212676
76	12.8942708088
77	12.9333560933
78	12.9596133696
79	13.1853083353
80	13.3188089263
81	13.5628624652
82	13.7773725944
83	14.0607663424
84	14.3356323131
85	14.5191075152
86	14.5865189766
87	14.6878354104
88	15.3635895673
89	15.7269991956
90	16.5351071844
91	16.5533720126
92	16.5786905196
93	17.0492155487
94	17.1165443004
95	17.6000152676
96	18.1992372610
97	19.5167656440
98	19.5565389599
99	20.9921734918
100	22.2421828344

==========

The sample data rank is below

	X1	rank(X1)
1	7.7201682102	39.000
2	-2.2894789917	2.000
3	4.0536320802	19.000
4	7.2119291036	36.000
5	19.5167656440	97.000
6	10.0788292012	56.000
7	17.1165443004	94.000
8	12.9333560933	77.000
9	3.9731342997	18.000

10	9.2312149618	49.000
11	3.9021027245	16.000
12	11.9064074021	70.000
13	15.3635895673	88.000
14	9.6402589370	53.000
15	4.2291059631	21.000
16	10.0825246374	57.000
17	-2.8015363963	1.000
18	17.6000152676	95.000
19	8.0492817656	42.000
20	2.9747309312	11.000
21	4.4795657566	24.000
22	15.7269991956	89.000
23	6.0978695650	30.000
24	7.4711976174	37.000
25	11.5224422489	69.000
26	14.0607663424	83.000
27	1.0811032008	8.000
28	-0.2337120745	5.000
29	4.7063227582	26.000
30	9.7465338352	54.000
31	7.5477445702	38.000
32	11.0206913959	63.000
33	12.8942708088	76.000
34	10.4459062337	60.000
35	8.7462432235	47.000
36	0.1816695265	6.000
37	12.0706359217	72.000
38	12.9596133696	78.000
39	3.9285453351	17.000
40	8.1220133674	44.000
41	4.2926775046	22.000
42	16.5786905196	92.000
43	10.2114330986	58.000
44	9.0152326204	48.000
45	4.7691006482	27.000
46	16.5533720126	91.000
47	3.2332007419	13.000
48	2.4168761454	10.000
49	4.4168062304	23.000
50	6.1156229436	31.000
51	22.2421828344	100.000
52	10.2894725615	59.000
53	3.6152019322	14.000
54	16.5351071844	90.000
55	9.5307881990	52.000
56	10.9228748573	62.000
57	-1.3503730666	3.000
58	14.6878354104	87.000
59	3.8047138217	15.000
60	11.3653277906	68.000
61	11.2513175878	66.000
62	-0.3223120820	4.000
63	9.3756760099	51.000
64	19.5565389599	98.000
65	9.9847419996	55.000
66	14.5191075152	85.000
67	12.0443732821	71.000
68	20.9921734918	99.000
69	4.7966027963	28.000
70	11.2505580734	65.000
71	14.5865189766	86.000
72	13.1853083353	79.000
73	12.7582212676	75.000
74	6.9453765282	34.000
75	11.2771214919	67.000
76	13.3188089263	80.000
77	7.7250651883	40.000
78	1.0256642807	7.000
79	6.2106671259	32.000
80	12.3660483105	73.000
81	6.7408731923	33.000
82	8.0829208609	43.000
83	9.2577693010	50.000
84	13.5628624652	81.000

85	11.1965103144	64.000
86	3.1182088230	12.000
87	8.0214910304	41.000
88	2.3787230835	9.000
89	5.5636009868	29.000
90	17.0492155487	93.000
91	10.5481247237	61.000
92	13.7773725944	82.000
93	8.3302576499	45.000
94	8.4419273511	46.000
95	12.7238465371	74.000
96	18.1992372610	96.000
97	4.5932064719	25.000
98	7.1398807863	35.000
99	14.3356323131	84.000
100	4.1553942982	20.000

---------- inference statistiscs ----------

* Suppose the population distribution is the normal distribution.

1. one population mean test and mu confidence interval when population sigma is unknown

H0: mu=0 , mu is population mean

t(df=99)=17.078682 which formula is t=(X1 sample mean-0)/standard error

the standard error =sample stand deviation/(n-1)^0.5, n is sample size=100

left tail test p-value= 1.0000

right tail test p-value= 0.0000

two tailes test p-value= 0.0000

90% confidence interval for mu

[8.236937 , 10.010658]

95% confidence interval for mu

[8.064183 , 10.183413]

99% confidence interval for mu

[7.720836 , 10.526759]

2. one population sigma confidence interval when population mean is unknown

90% confidence interval for population variance

[22.928648 , 36.673049]

90% confidence interval for population standard deviation

[4.788387 , 6.055828]

95% confidence interval for population variance

[22.001469 , 38.513809]

95% confidence interval for population standard deviation

[4.690572 , 6.205950]

99% confidence interval for population variance

[20.332385 , 42.490018]

99% confidence interval for population standard deviation

[4.509145 , 6.518437]

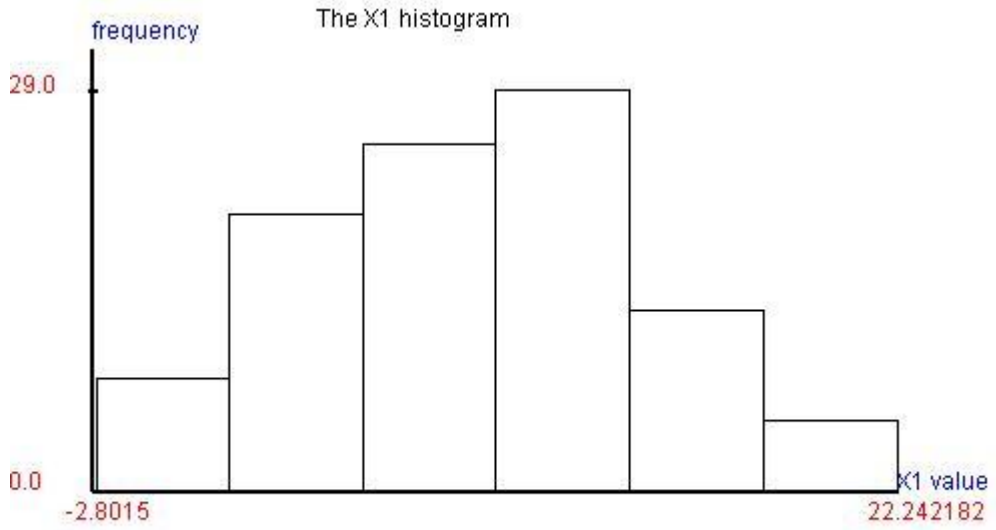
random variable X1

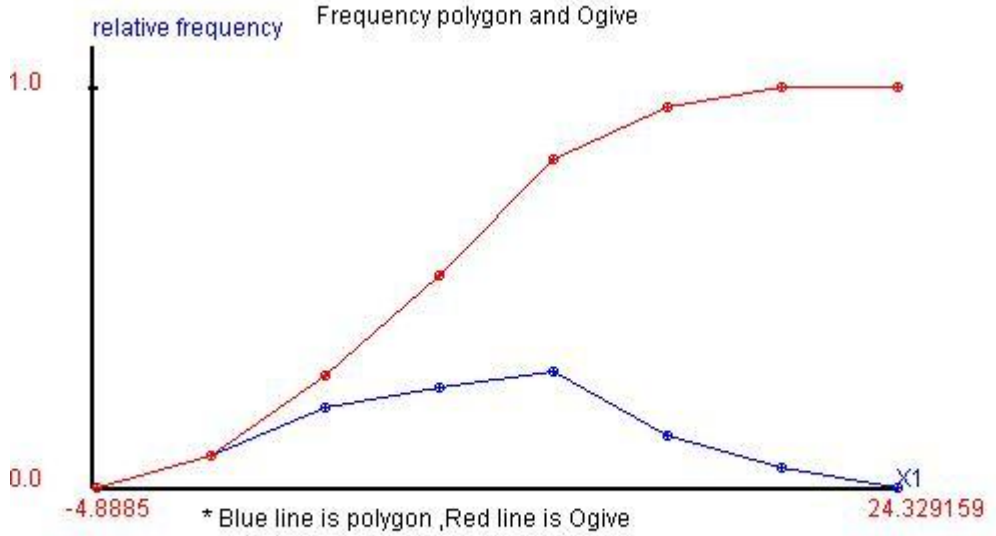
class	class limit	class midpoint	frequency	relative frequency	cumulative frequency
[1]	-2.80154~ 1.37242	-0.71456	8.00000	0.0800000	0.0800000
[2]	1.37242~ 5.54637	3.45939	20.00000	0.2000000	0.2800000
[3]	5.54637~ 9.72032	7.63335	25.00000	0.2500000	0.5300000
[4]	9.72032~ 13.89428	11.80730	29.00000	0.2900000	0.8200000
[5]	13.89428~ 18.06823	15.98125	13.00000	0.1300000	0.9500000
[6]	18.06823~ 22.24218	20.15521	5.00000	0.0500000	1.0000000

frequency distribution: sample mean=9.052491 , sample variance=28.996986 , sample sd=5.384885

The histogram is stored in c:\book_01\histogramX_image.txt

The polygon and ogive is stored in c:\book_01\polygonX_image.txt


The X1 histogram
frequency
29.0
0.0
X1 value
-2.8015
22.242182


Frequency polygon and Ogive
relative frequency
1.0
0.0
X1
-4.8885
24.329159
* Blue line is polygon ,Red line is Ogive