

Std.Normal
UNI\4SN36_V1.MTB

Sampling Distribution Summary
1/11/95

Milo Schield

APPROACH

1. Select a population where the shape is an analytic function
Select number of points in the finite sampled population.
These points are the mid-points of equal areas
Obtain the values of these points using INVCDF.
 2. Obtain random samples of predetermined size from this population
This was done using the SAMPLE command in Minitab
This sample is uniformly random from the list of the population.
The BASE seed value was set to a constant (for repeatability).
 3. For each sample, the following statistics are available:
 - C1: Sample Mean
 - C2: Sample Median
 - C3: Sample Std. Deviation
 - C4: Sample Interquartile Range
 - C5: Sample Mid-IQR
 - C6: Sample Approx. Skewness [3*(Mean-Median)/Std.Deviation]
 - C7: Sample Exact Skewness [(Mean(Score-Mean)^3)/(Std.Dev)^3]
 - C8: Sample Kurtosis [(Mean(Score-Mean)^4)/(Std.Dev)^4]
 - C9: Sample Correlation [Sample items and NSCORES of same]
 - C10: Sample Skew Difference [C6-C7]
 - C25: Boxplot Outlier 1=Yes, 0=No
 - C26: |Max| category 1=1.8, 2=2.2, 3=2.5
 - C27: Median>0.4 1=Yes, 0=No
 - C28: Sample Minimum
 - C29: Sample Maximum
- =====

OUTPUT SECTION: Summarize results

For each univariate variable (sampling distribution)

1. Display graphs (Dotplot and Boxplot)
2. Present summary statistics

For all bivariate variables

1. Calculate correlation coefficients (Pearson & Spearman)
2. Select those that are statistically significant
3. Show in XY plot
4. Step-wise regression involving one of the variables

ASSUMPTIONS FOR SAMPLING

Data Display

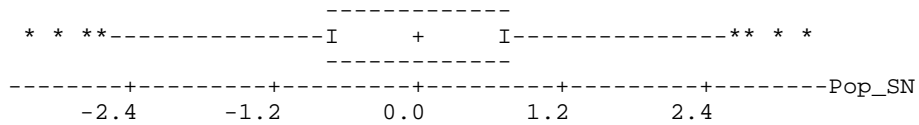
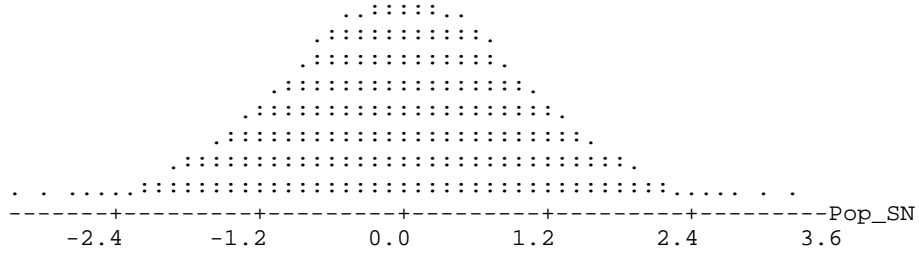
Seed Value = 1. Number of Samples = 200.
Size of each sample = 36. Number in population = 999.

Sampling Distribution of Statistics

Std.Normal

Sampling Distribution Summary
POPULATION: STD.NORMAL DISTRIBUTION

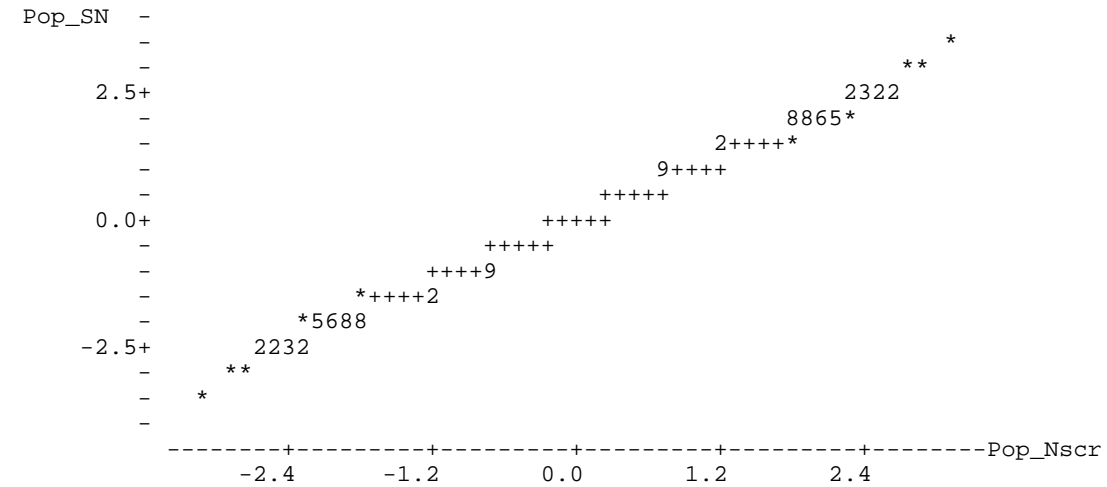
Each dot represents 3 points



Variable	N	Mean	Median	TrMean	StDev	SEMean
Pop_SN	999	-0.0000	0.0000	-0.0000	0.9998	0.0316

Variable	Min	Max	Q1	Q3
Pop_SN	-3.2902	3.2902	-0.6753	0.6753

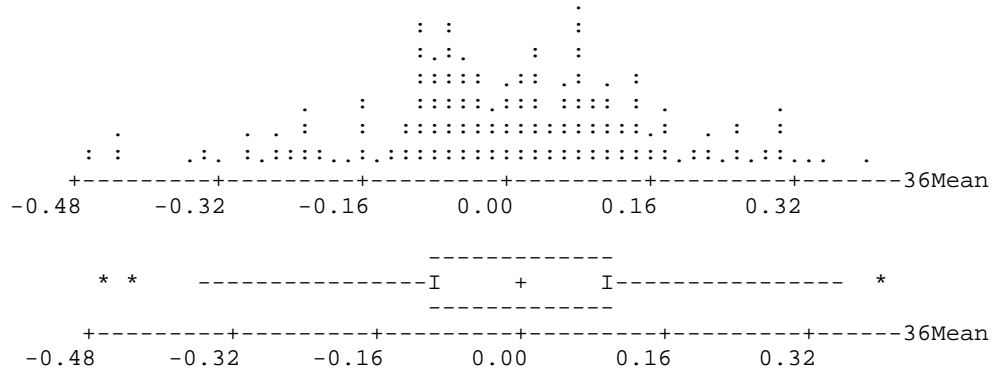
SKEWNESS: Exact: -0.00, Approximate: -0.00, Approx. (% Max): -0%
 PEAKEDNESS: Kurtosis: 2.97, 75% of IQR: 1.00, StdDev / 75%IQR: 100%
 Correlation w Nscores: 2E+02



Sampling Distribution
of Statistics

Std.Normal

Sampling Distribution Summary
SAMPLE MEANS (C1)

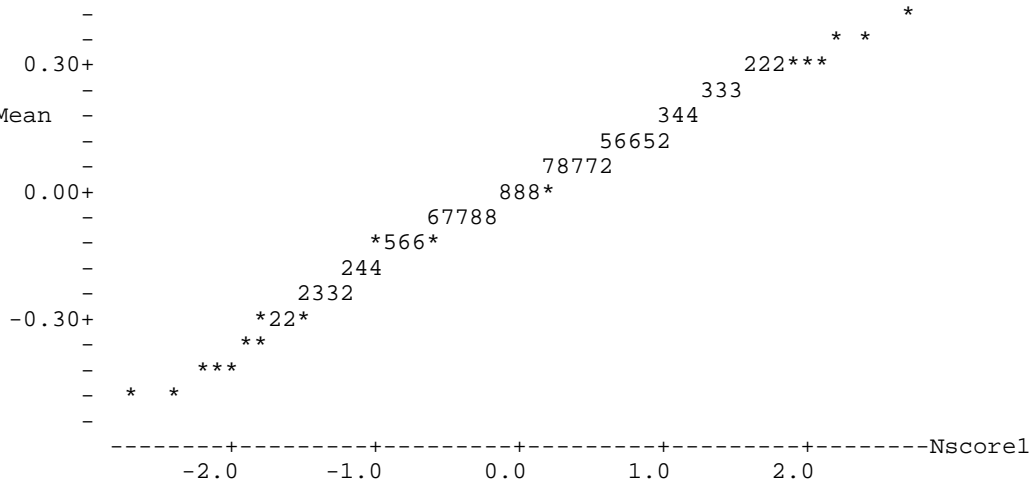


SUMMARY STATISTICS: (Population Mean = -0.00, Std.Dev = 1.00)

Variable	N	Mean	Median	TrMean	StDev	SEMean
36Mean	200	-0.0012	0.0016	0.0023	0.1665	0.0118

Variable	Min	Max	Q1	Q3
36Mean	-0.4692	0.4023	-0.0913	0.1044

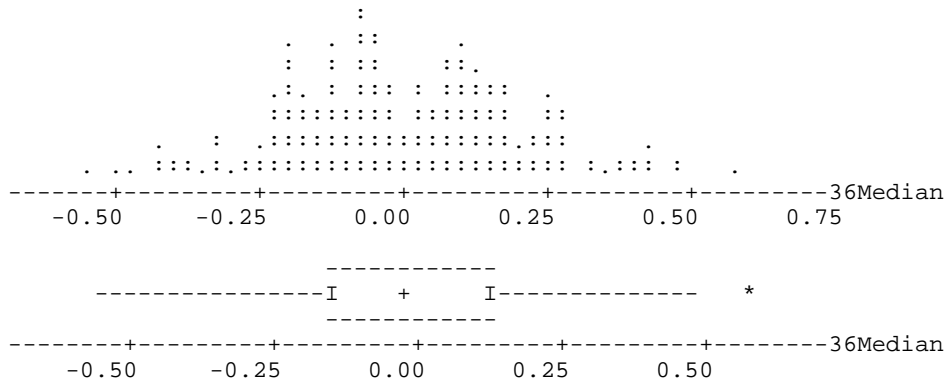
SKEWNESS: Exact: -0.29, Approximate: -0.05, Approx.(% Max): -2%
 PEAKEDNESS: Kurtosis: 3.21, 75% of IQR: 0.15, StdDev / 75%IQR: 113%
 Correlation w Nscores: 0.994, % Outliers: 0.22



Sampling Distribution of Statistics

Std.Normal

Sampling Distribution Summary SAMPLE MEDIANS (C2)

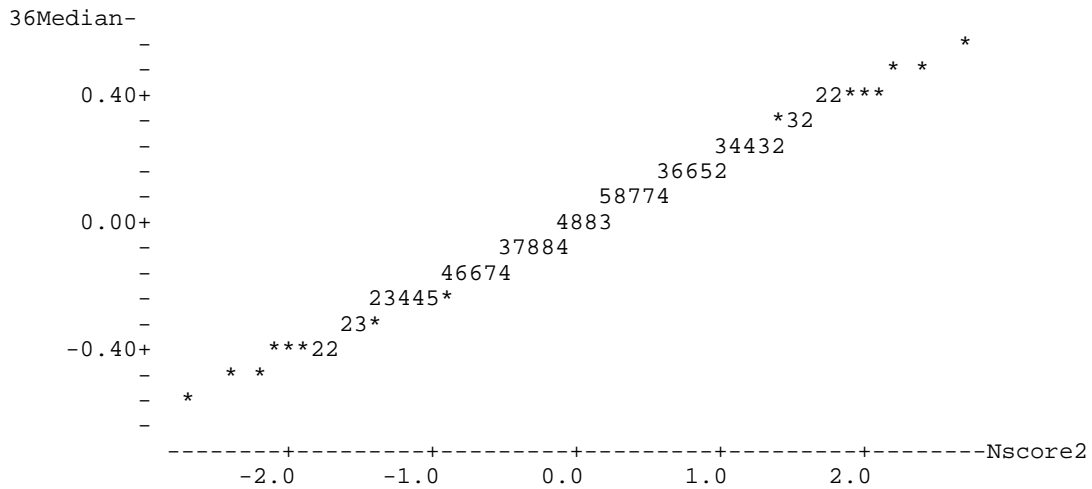


SUMMARY STATISTICS: (Population Median = 0.00, Std.Dev = 1.00)

Variable	N	Mean	Median	TrMean	StDev	SEMean
36Median	200	-0.0065	-0.0251	-0.0071	0.2080	0.0147

Variable	Min	Max	Q1	Q3
36Median	-0.5540	0.5808	-0.1467	0.1302

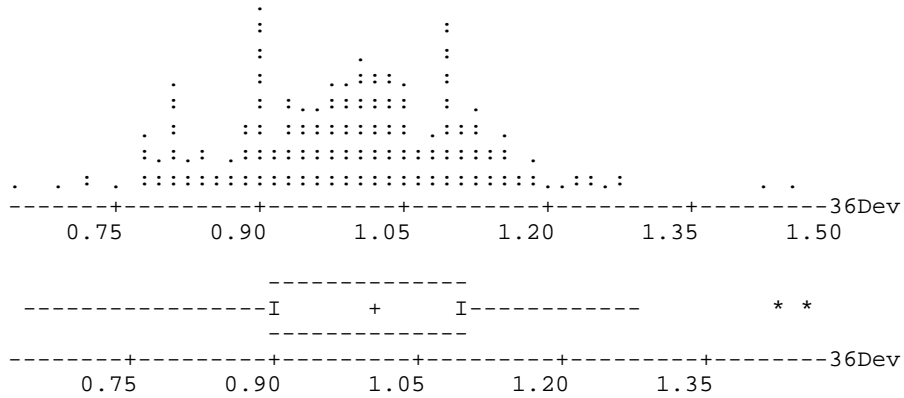
SKEWNESS: Exact: 0.06, Approximate: 0.27, Approx. (% Max): 9%
 PEAKEDNESS: Kurtosis: 2.88, 75% of IQR: 0.21, StdDev / 75%IQR: 100%
 Correlation w Nscores: 0.999



Sampling Distribution of Statistics

Std.Normal

Sampling Distribution Summary SAMPLE STANDARD DEVIATIONS (C3)

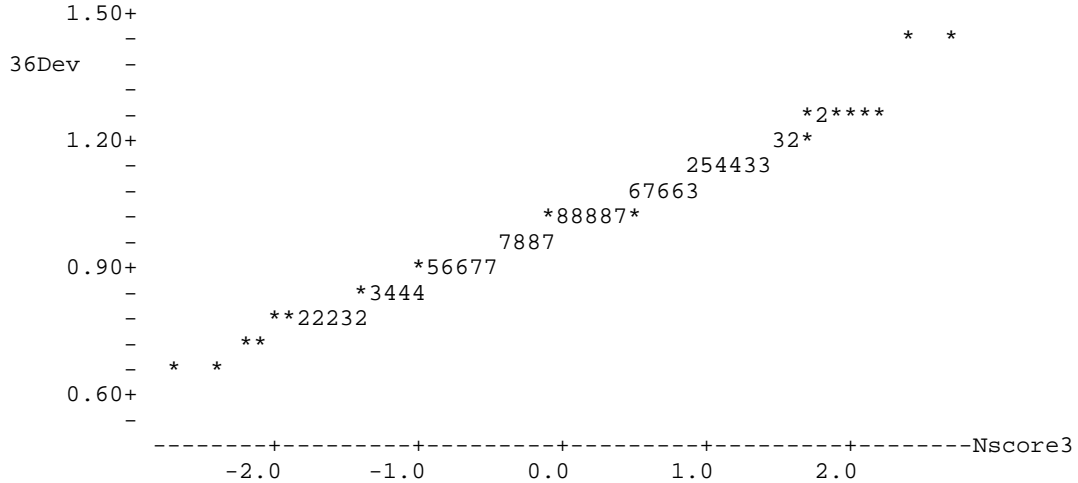


SUMMARY STATISTICS: (Population Std.Dev = 1.00)

Variable	N	Mean	Median	TrMean	StDev	SEMean
36Dev	200	0.99488	0.99909	0.99259	0.13128	0.00928

Variable	Min	Max	Q1	Q3
36Dev	0.63878	1.45143	0.90154	1.08825

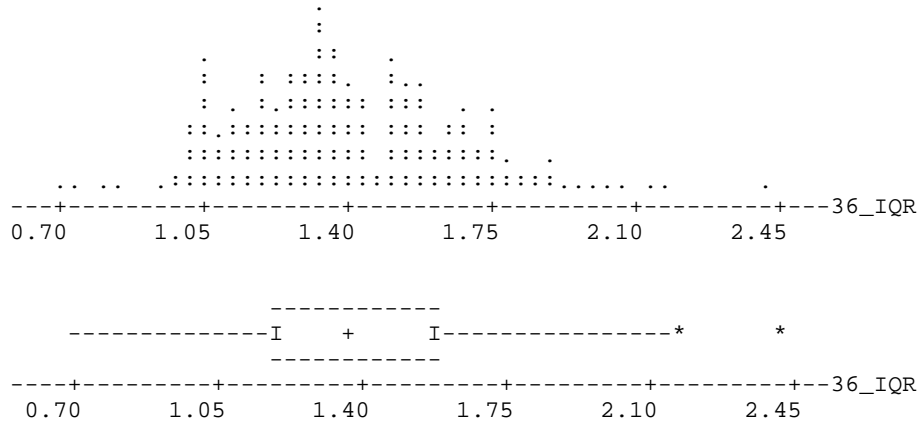
SKEWNESS: Exact: 0.22, Approximate: -0.10, Approx. (% Max): -3%
 PEAKEDNESS: Kurtosis: 3.41, 75% of IQR: 0.14, StdDev / 75%IQR: 94%
 Correlation w Nscores: 0.995



Sampling Distribution of Statistics

Std.Normal

Sampling Distribution Summary
SAMPLE INTERQUARTILE RANGES (C4)

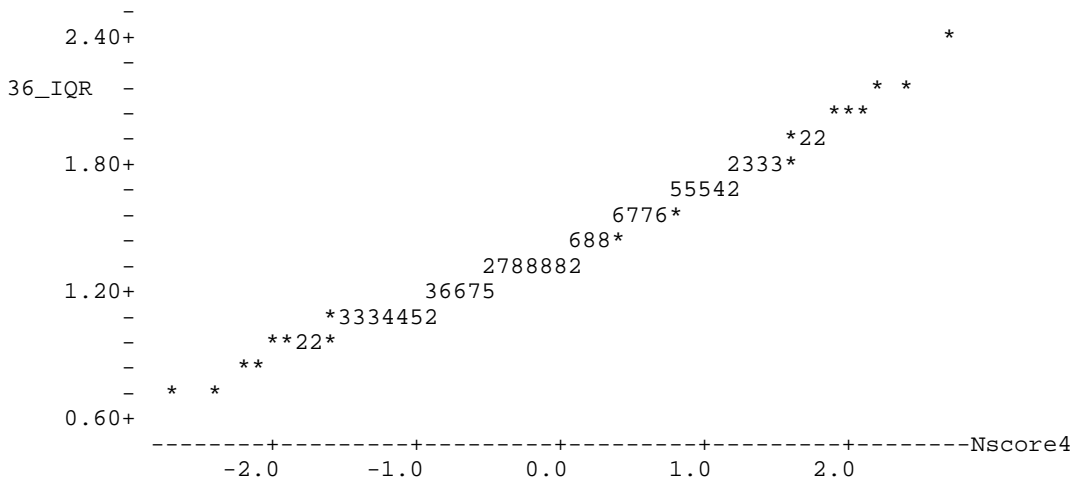


SUMMARY STATISTICS: (Population IQR = 1.33)

Variable	N	Mean	Median	TrMean	StDev	SEMean
36_IQR	200	1.3993	1.3615	1.3912	0.2792	0.0197

Variable	Min	Max	Q1	Q3
36_IQR	0.7159	2.4063	1.1982	1.5737

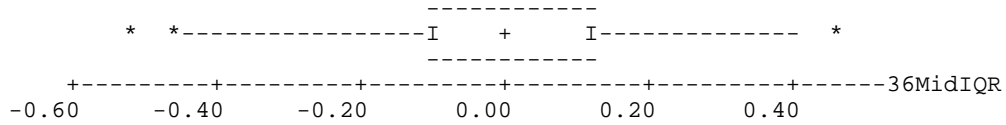
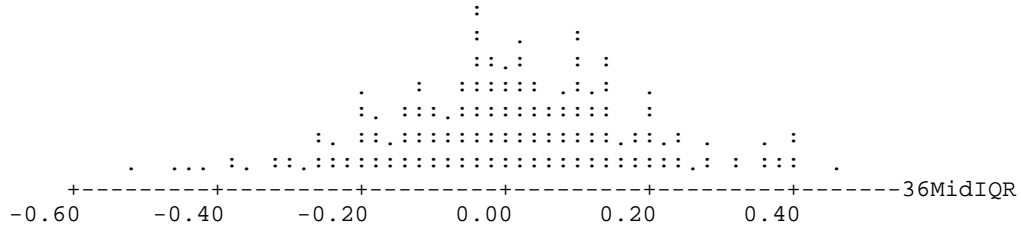
SKEWNESS: Exact: 0.44, Approximate: 0.41, Approx. (% Max): 14%
 PEAKEDNESS: Kurtosis: 3.36, 75% of IQR: 0.28, StdDev / 75%IQR: 99%
 Correlation w Nscores: 0.993



Sampling Distribution of Statistics

Std.Normal

Sampling Distribution Summary SAMPLE MID-INTERQUARTILE RANGES (C5)

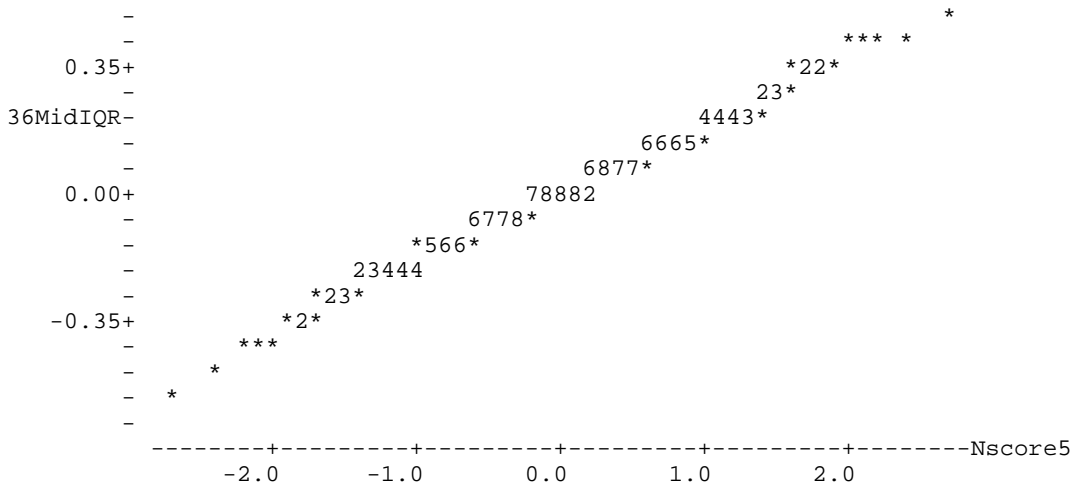


SUMMARY STATISTICS: (Population MID-IQR = 0.00)

Variable	N	Mean	Median	TrMean	StDev	SEMean
36MidIQR	200	0.0061	0.0041	0.0069	0.1819	0.0129

Variable	Min	Max	Q1	Q3
36MidIQR	-0.5256	0.4691	-0.1103	0.1206

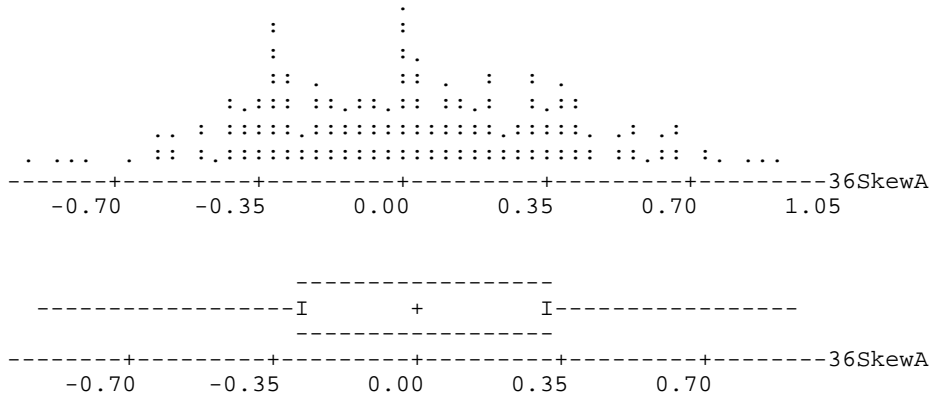
SKEWNESS: Exact: -0.09, Approximate: 0.03, Approx.(% Max): 1%
 PEAKEDNESS: Kurtosis: 3.11, 75% of IQR: 0.17, StdDev / 75%IQR: 105%
 Correlation w Nscores: 0.998



Sampling Distribution of Statistics

Std.Normal

Sampling Distribution Summary SAMPLE SKEWNESSES - APPROXIMATE (C6)

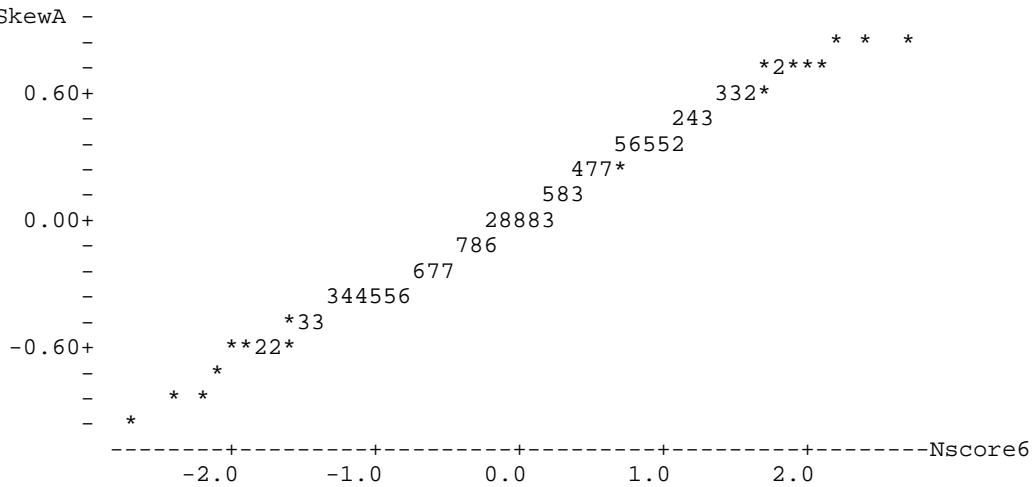


SUMMARY STATISTICS: (Population Skewness(approx) =-0.00)

Variable	N	Mean	Median	TrMean	StDev	SEMean
36SkewA	200	0.0177	0.0082	0.0162	0.3643	0.0258

Variable	Min	Max	Q1	Q3
36SkewA	-0.9054	0.8989	-0.2847	0.3009

SKEWNESS: Exact: 0.10, Approximate: 0.08, Approx. (% Max): 3%
 PEAKEDNESS: Kurtosis: 2.56, 75% of IQR: 0.17, StdDev / 75%IQR: 210%
 Correlation w Nscores: 0.997

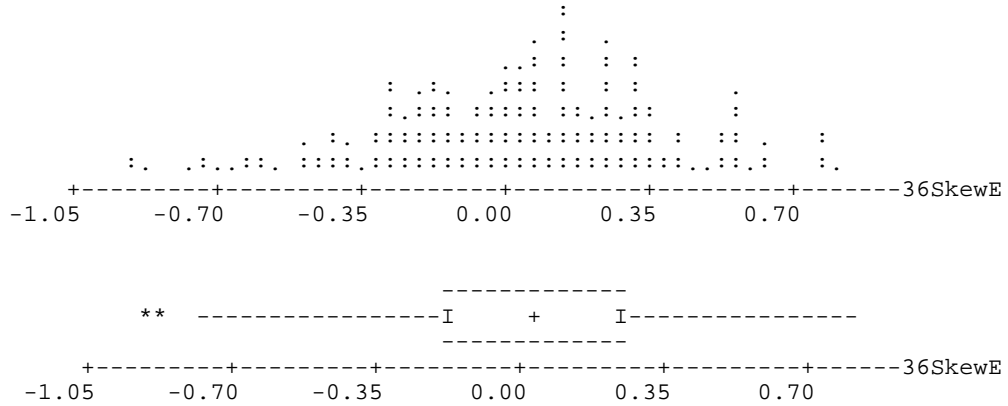


Sampling Distribution of Statistics

Std. Normal
Population

Std.Normal

Sampling Distribution Summary
SAMPLE SKEWNESSES - EXACT (C7)

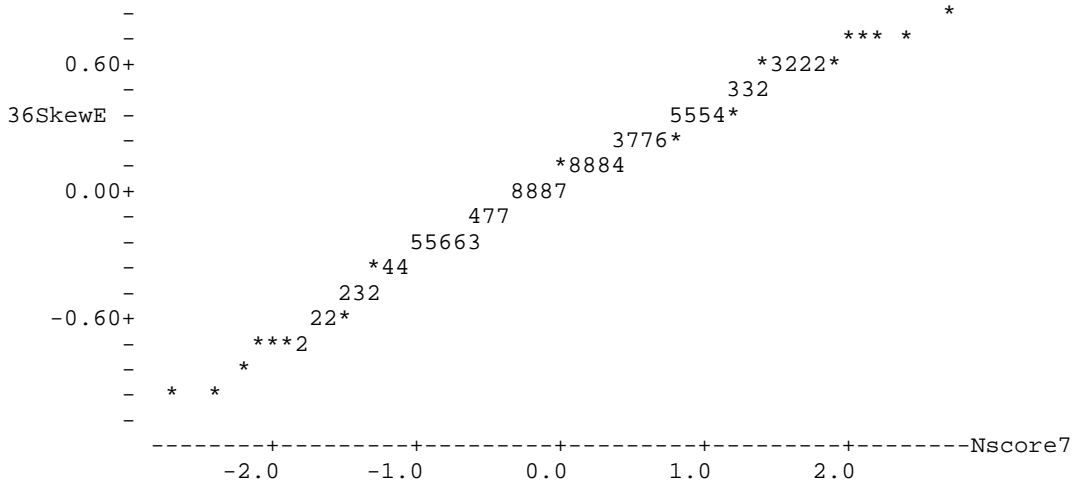


SUMMARY STATISTICS: (Population Skewness(exact) =-0.00)

Variable	N	Mean	Median	TrMean	StDev	SEMean
36SkewE	200	0.0315	0.0518	0.0384	0.3467	0.0245

Variable	Min	Max	Q1	Q3
36SkewE	-0.9118	0.8104	-0.1925	0.2587

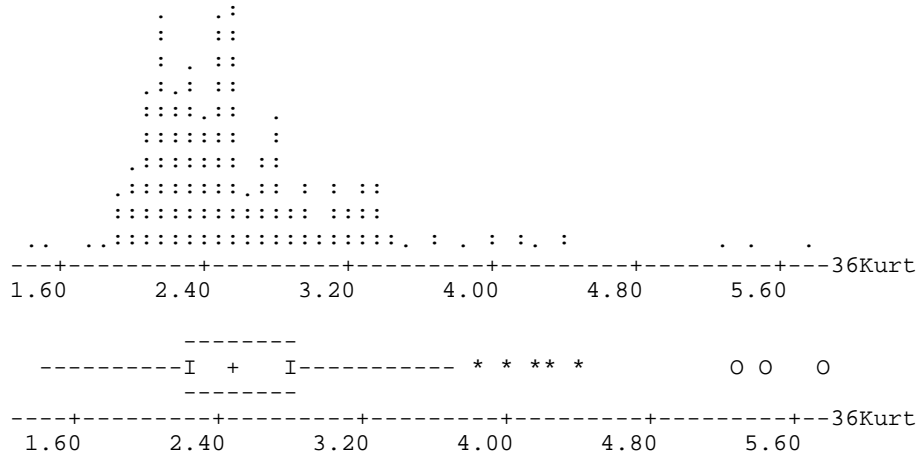
SKEWNESS: Exact: -0.26, Approximate: -0.17, Approx. (% Max): -6%
 PEAKEDNESS: Kurtosis: 2.98, 75% of IQR: 0.34, StdDev / 75%IQR: 102%
 Correlation w Nscores: 0.996



Sampling Distribution of Statistics

Std.Normal

Sampling Distribution Summary
SAMPLE KURTOSIS (C8)

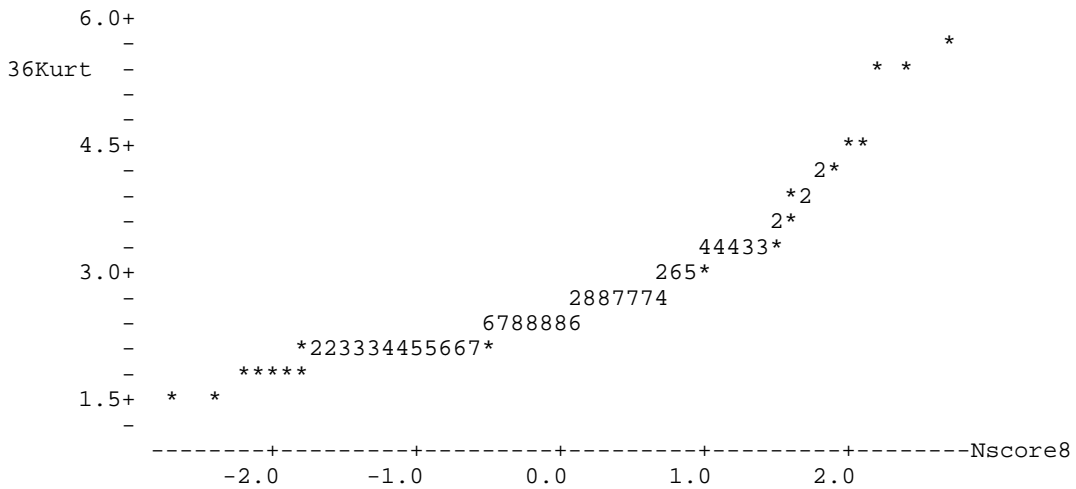


SUMMARY STATISTICS: (Population Kurtosis = 2.97)

Variable	N	Mean	Median	TrMean	StDev	SEMean
36Kurt	200	2.6325	2.4983	2.5681	0.6320	0.0447

Variable	Min	Max	Q1	Q3
36Kurt	1.4666	5.7253	2.2058	2.8380

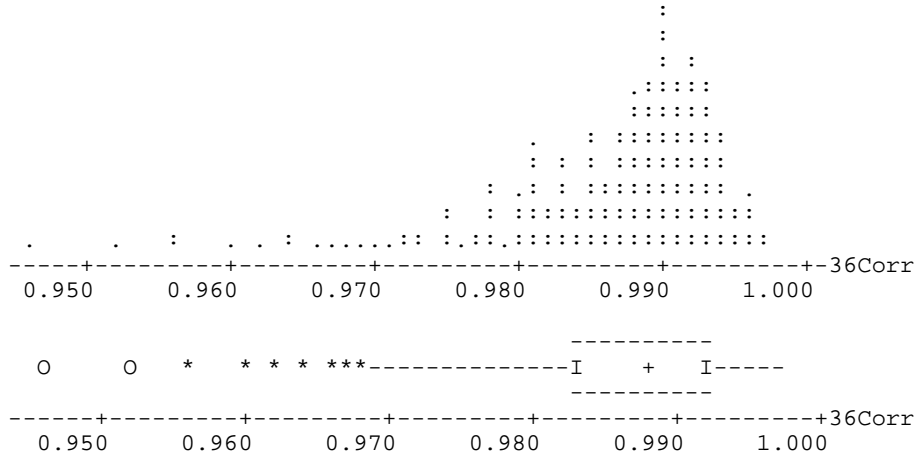
SKEWNESS: Exact: 1.91, Approximate: 0.64, Approx.(% Max): 21%
 PEAKEDNESS: Kurtosis: 8.36, 75% of IQR: 0.47, StdDev / 75%IQR: 133%
 Correlation w Nscores: 0.917



Sampling Distribution of Statistics

Std.Normal

Sampling Distribution Summary
SAMPLE CORRELATIONS: Data & Nscores (C9)

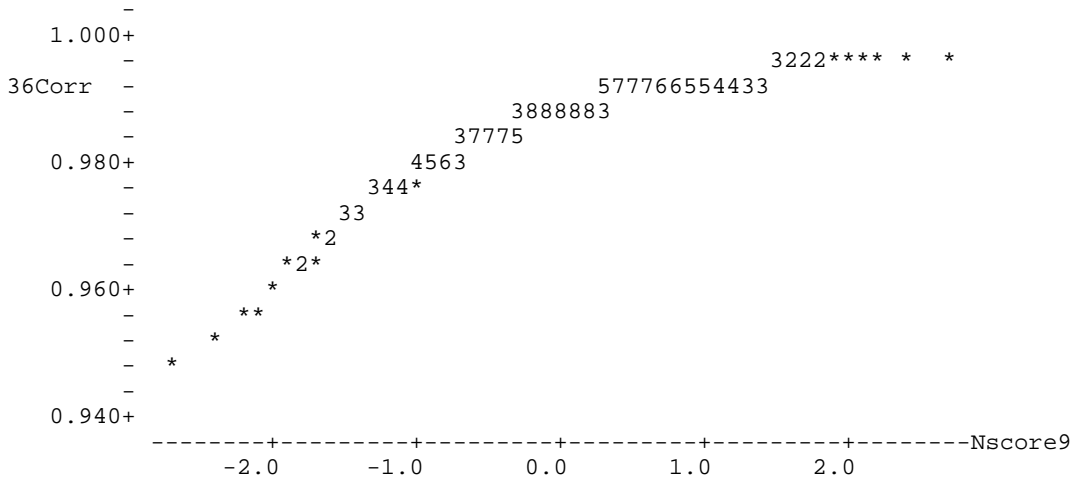


SUMMARY STATISTICS:

Variable	N	Mean	Median	TrMean	StDev	SEMean
36Corr	200	0.98582	0.98844	0.98674	0.00865	0.00061

Variable	Min	Max	Q1	Q3
36Corr	0.94626	0.99731	0.98251	0.99152

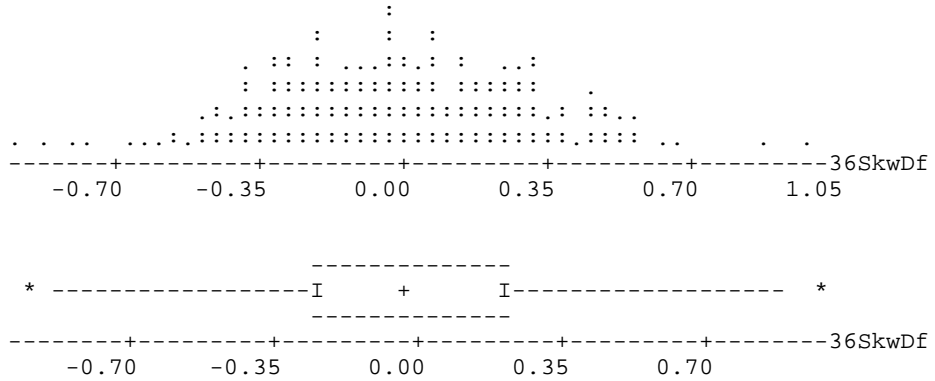
SKEWNESS: Exact: -1.77, Approximate: -0.91, Approx. (% Max): -30%
 PEAKEDNESS: Kurtosis: 6.76, 75% of IQR: 0.01, StdDev / 75%IQR: 128%
 Correlation w Nscores: 0.916



Sampling Distribution of Statistics

Std.Normal

Sampling Distribution Summary
SAMPLE SKEW DIFF: APPROX.- EXACT (C10)

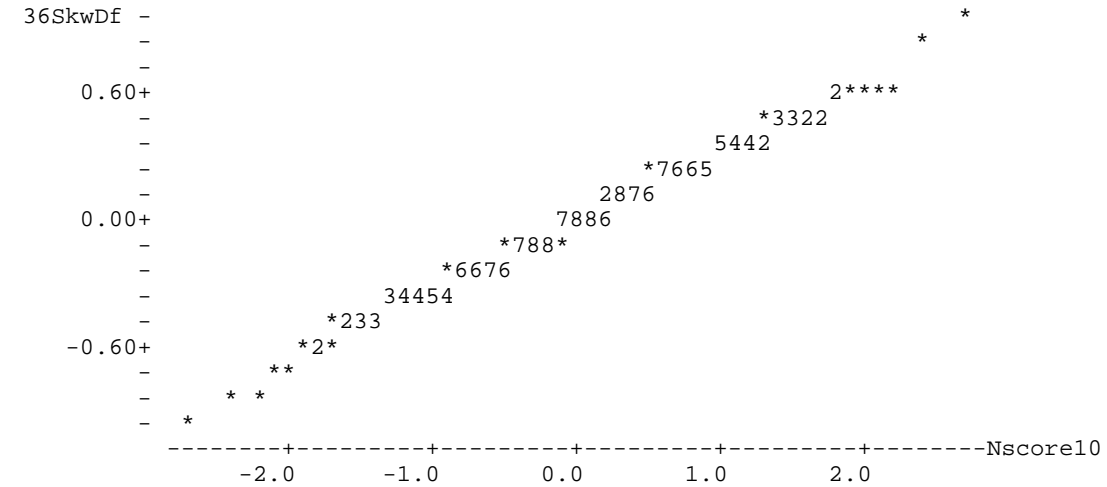


SUMMARY STATISTICS: (Population Skewness: Approx - Exact = -0.00)

Variable	N	Mean	Median	TrMean	StDev	SEMean
36SkwDf	200	-0.0139	-0.0288	-0.0125	0.3251	0.0230

Variable	Min	Max	Q1	Q3
36SkwDf	-0.9491	0.9660	-0.2437	0.2189

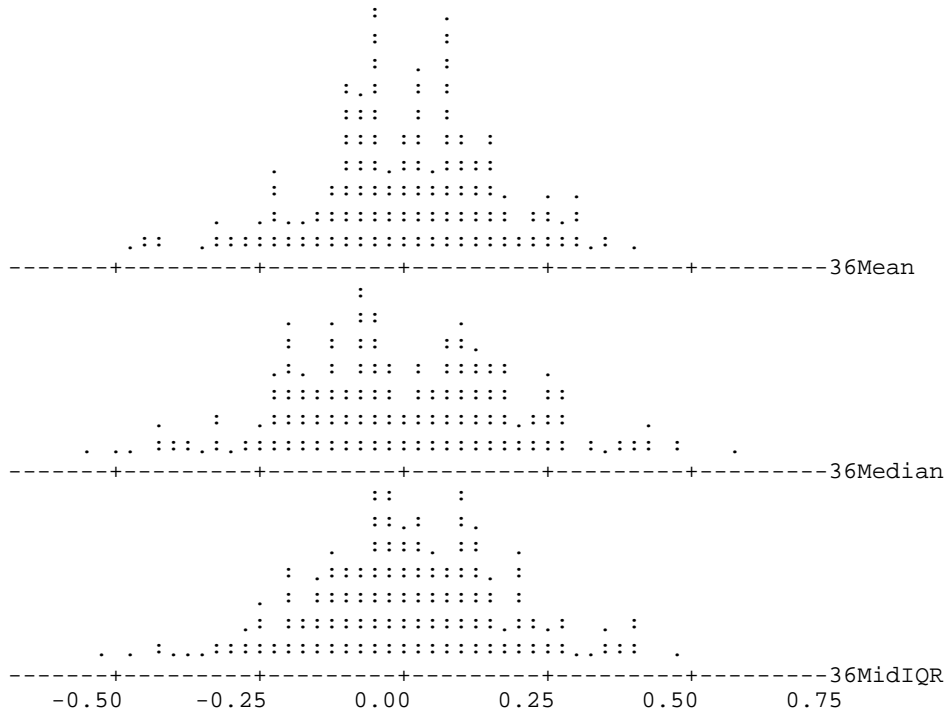
SKEWNESS: Exact: 0.01, Approximate: 0.14, Approx. (% Max): 5%
 PEAKEDNESS: Kurtosis: 3.04, 75% of IQR: 0.35, StdDev / 75%IQR: 94%
 Correlation w Nscores: 0.998



Sampling Distribution of Statistics

Std.Normal

Sampling Distribution Summary UNIVARIATE SUMMARY CENTRAL TENDANCY



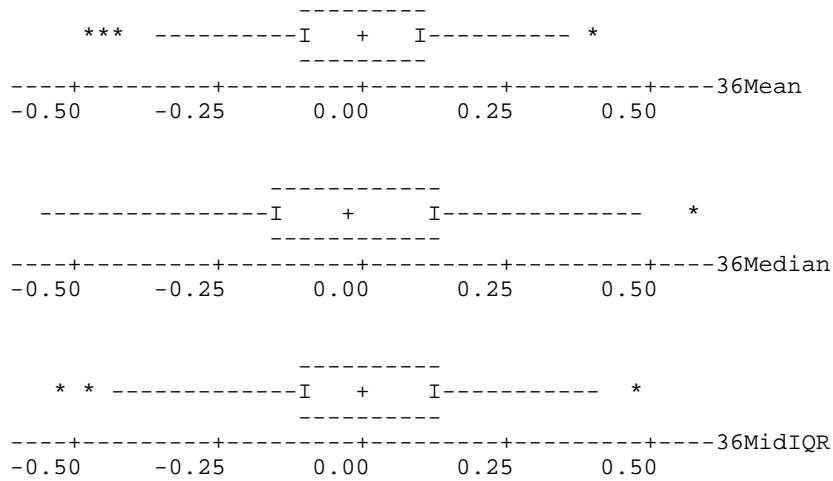
Variable	N	Mean	Median	TrMean	StDev	SEMean
36Mean	200	-0.0012	0.0016	0.0023	0.1665	0.0118
36Median	200	-0.0065	-0.0251	-0.0071	0.2080	0.0147
36MidIQR	200	0.0061	0.0041	0.0069	0.1819	0.0129

Variable	Min	Max	Q1	Q3
36Mean	-0.4692	0.4023	-0.0913	0.1044
36Median	-0.5540	0.5808	-0.1467	0.1302
36MidIQR	-0.5256	0.4691	-0.1103	0.1206

Sampling Distribution of Statistics

Std.Normal

Sampling Distribution Summary UNIVARIATE SUMMARY CENTRAL TENDANCY (continued)



The centers of these 3 distributions are all non-zero. But, as the number of samples increases, we expect the centers to approach zero (symmetry). There is no reason for bias.

As the number of samples increases, we expect the spreads to stabilize at some non-zero value. The statistic whose sampling distribution has the smallest standard deviation is said to be the most efficient statistic.

The most efficient statistic will tend to have the smallest center (for a given number of samples).

Based on this simulation, it appears that the mean is the most efficient statistic.

The apparant order is Mean, Mid-IQR and Median

=====
How can the sample median be further from the population center than is the sample mean (when the population is Normal)?

It can't be explained by the normal "severity-outliers"! The median is sensitive to "frequency-outliers".

Despite its overt sensitivity to severity outliers, the non-democratic mean is a better predictor than is the mean or mid-IQR with their covert sensitivity to democratic fluctuations (frequency outliers) within the "Normal" range.

Sampling Distribution of Statistics

Std.Normal Sampling Distribution Summary UNIVARIATE SUMMARY SPREAD

In this section, IQR is multiplied by 3/4 for better comparison.

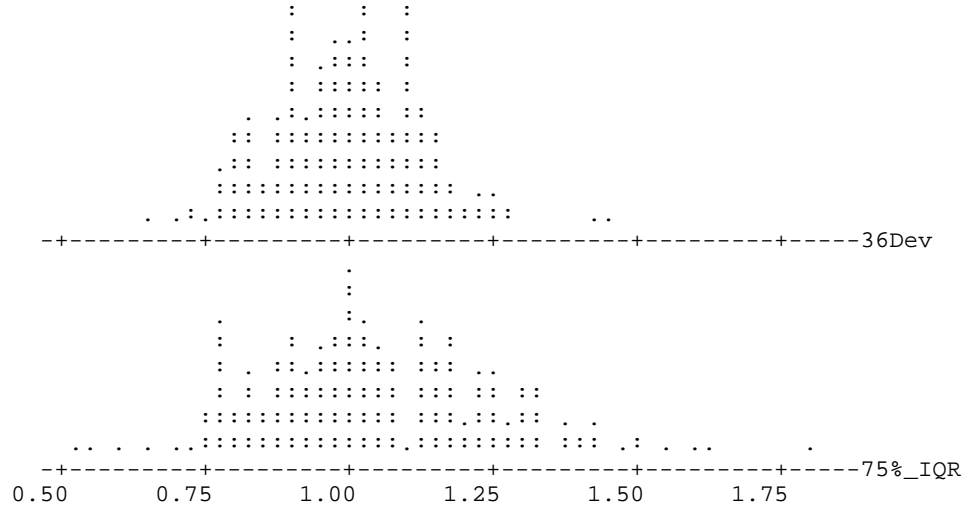
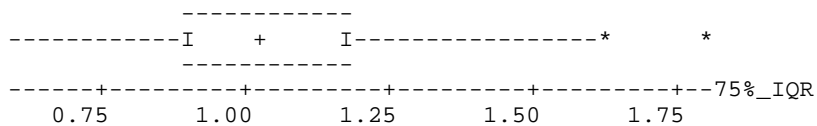
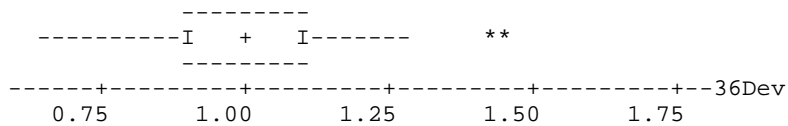


Table with statistical data for 36Dev and 75%_IQR, including columns for Variable, N, Mean, Median, TrMean, StDev, SEMean, Min, Max, Q1, and Q3.



Std.Normal Sampling Distribution Summary
 UNIVARIATE SUMMARY
 SPREAD (Continued)

After multiplying the IQR by 3/4ths, the centers approach one.
As the number of samples increases, we expect the centers to equal 1.

As the number of samples increases, the standard deviations of
these sampling distributions approaches some non-zero values.

The statistic whose sampling distribution has the smallest std.
deviation is considered the most efficient statistic.

In this simulation, the standard deviation is more efficient than
the 75% of the Inter-Quartile Range.

The apparant order from best to worst is Std.Dev and then IQR.
=====

Why does the 75%IQR vary more than the standard deviation?

- The standard deviation is extremely sensitive to outliers.
Q. How can it vary less than the 75% IQR?
A. The 75% IQR is bigger than the std.deviation in some samples.

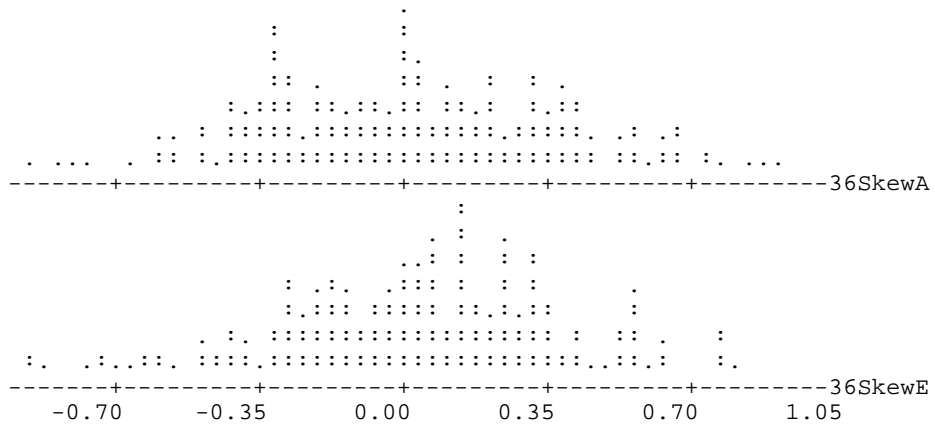
- Q. How can 75%_IQR be greater than the standard deviation?
A. For the same reason the median can be bigger than the mean.

The democratic IQR is easily swayed by frequency fluctuations
of points within the "Normal" range. The authoritarian Std.Dev.
is sensitive to severity outliers. For this Normal population,
frequency outliers exceed severity outliers -- or else frequency
outliers are just more common the severity outliers.

Sampling Distribution of Statistics

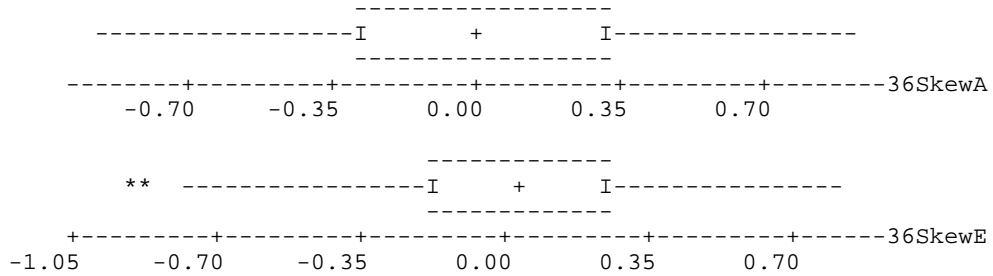
Std.Normal

Sampling Distribution Summary UNIVARIATE SUMMARY SKEWNESS



Variable	N	Mean	Median	TrMean	StDev	SEMean
36SkewA	200	0.0177	0.0082	0.0162	0.3643	0.0258
36SkewE	200	0.0315	0.0518	0.0384	0.3467	0.0245

Variable	Min	Max	Q1	Q3
36SkewA	-0.9054	0.8989	-0.2847	0.3009
36SkewE	-0.9118	0.8104	-0.1925	0.2587



Although the Approx.skewness has a smaller center this is not considered fundamental. Both centers should go to zero as more samples are aquired.

As the number of samples increases, the standard deviations should stabilize. The statistic with the smaller standard deviation is considered the more "efficient".

Based on this simulation, it appears that the "Exact" form is somewhat more efficient than the "Approximate" form.

Sampling Distribution
of Statistics

Std.Normal Sampling Distribution Summary
CORRELATIONS

Correlations involving differences in skewness (C10) are ignored.

Pearson Correlation Coefficients (C1-C9)

	36Mean	36Median	36Dev	36_IQR	36MidIQR	36SkewA	36SkewE	36Kurt
36Median	0.815							
36Dev	0.061	0.070						
36_IQR	0.064	0.071	0.758					
36MidIQR	0.912	0.724	0.073	0.099				
36SkewA	-0.023	-0.593	-0.034	-0.040	0.013			
36SkewE	-0.066	-0.389	-0.006	-0.108	-0.268	0.583		
36Kurt	-0.125	-0.056	-0.074	-0.491	-0.145	-0.072	0.057	
36Corr	0.078	0.062	0.086	0.237	0.060	-0.002	-0.004	-0.629

=====

Spearman Rank Correlation (C31-C39)

	RkMean	RkMedian	RkStDev	RkIQR	RkMdIQR	RkSkAp	RkSkEx	RkKurt
RkMedian	0.817							
RkStDev	0.062	0.045						
RkIQR	0.048	0.054	0.743					
RkMdIQR	0.907	0.729	0.052	0.079				
RkSkAp	-0.050	-0.571	-0.054	-0.058	-0.007			
RkSkEx	-0.114	-0.411	-0.011	-0.130	-0.299	0.598		
RkKurt	-0.116	-0.069	-0.084	-0.484	-0.163	-0.055	0.109	
RkCorr	0.026	0.052	0.108	0.206	0.027	-0.041	-0.100	-0.335

=====

The following are the nine most significant coefficients.
Spearman Rank correlation is shown in parenthesis.

	Pearson			Spearman	STEPWISE
1.	+.905	Mean	Mid-IQR	(+.895)	C1 All
2.	+.815	Mean	Median	(+.817)	C1 All but C5
3.	+.739	Median	Mid-IQR	(+.742)	C2 All
4.	+.730	Std.Dev.	IQuartileRg.	(+.709)	C3 All
5.	-.629	Kurtosis	Correlation	(-.335)	C8 All
6.	-.593	Median	Skew-Approx	(-.571)	C2 All but C5
7.	+.583	Skew-approx	Skew-exact	(+.598)	C6 All
8.	-.483	IQR	Kurtosis	(-.471)	Not Shown
9.	-.389	Skew_exact	Median	(-.411)	Not Shown

Shouldn't KURTOSIS be symmetric with respect to SKEW?
Perhaps KURTOSIS is related to |SKEW2| or (SKEW2)^2 (Tom Morgan)