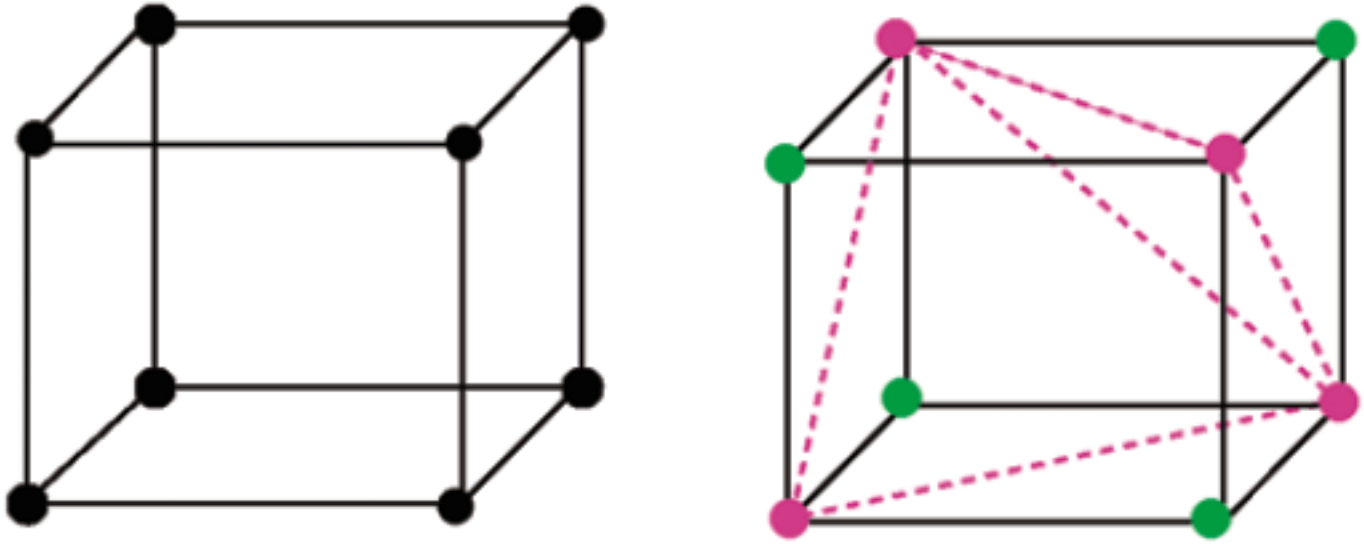


Figure 1. Full factorial and one-half factorial in three dimensions.

THE UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING

War of the Benchmark Means: Time for a Truce

John Mashey

Is GM meaningful metric for in any situation?

- We examined AM and HM earlier.
- We had examples showing GM is consistent but consistently wrong.
- Is GM meaningful for anything?

Workload Characterization Analysis

- Gather info at various levels about existing workloads
- Compile list of programs known to be important
- Frequencies of execution
- Fraction of total time in the environment
- Historical trends
- Long-term effort
- Can be done by owners of large computers or
- large number of computers

SERPOP

Sample Estimation of Relative Performance of Programs

Multi-element Benchmark Suite that is a sample of some population of programs

An appropriate model of the populations distribution

Mean, indices of dispersion, confidence levels, goodness of fit

Identify specific codes, describe their nature

SERPOP analysis admits to uncertainty and preferably quantifies

Enable users to selectively use relevant benchmarks only

Eg: LFK, NAS kernels, DR CPU2, SPEC CPU suite

SERPOP

SERPOP analysis admits to uncertainty and

preferably quantifies it

"If a man will begin in certainties, he shall end in doubts, but if he will be content to begin with doubts, he shall end with certainties."

– Francis Bacon, in Savage [1].

VAX 8700/8800 paper from McInnis et al

3.0–3.4	4.0%	FFFD
3.4–3.8	6.1%	FFFFFF
3.8–4.2	7.1%	IFFFFDL
4.2–4.6	19.2%	IIFFFFFFFFFDDDDDDLL
4.6–5.0	11.1%	CIIIFFLLLLL
5.0–5.4	15.2%	FFFFDDDDDDLLLLL
5.4–5.8	23.2%	CIFFFDDDDDDDDDDDDLLLLLL
5.8–6.2	6.1%	CCFDLL
6.2–6.6	3.0%	DDL
6.6–7.0	2.0%	DD
7.0–7.4	3.0%	DLL

Table 1 - Example from Digital Equipment [19]

VAX 8700/8800 paper from McInnis et al

Description	#	Metric	Value
Sample HM	**99**	**HM**	**4.84**
Sample GM	**99**	**GM**	**4.92**
Sample AM	**99**	**AM**	**5.01**
Sample Std Dev	**99**	**STDV**	**.88**
Sample Median	**99**	**Mdian**	**5.0**
Subsample			**Estimated**
C: Cobol	4	GM	5.6
I: Fortran Integer	7	GM	4.3
F: Fortran Single	32	GM	4.7
D: Fortran Double	32	GM	5.3
L: LISP	24	GM	5.3

Table 2 – Statistics for Table 1

VAX 8700/8800 paper from McInnis et al

3.0–3.4	4.0%	FFFD
3.4–3.8	6.1%	FFFFFF
3.8–4.2	7.1%	IFFFFDL
4.2–4.6	19.2%	IIFFFFFFFFFDDDDDDLL
4.6–5.0	11.1%	CIIIFFLLLLL
5.0–5.4	15.2%	FFFFDDDDDDLLLLL
5.4–5.8	23.2%	CIFFFDDDDDDDDDDDLLLLLL
5.8–6.2	6.1%	CCFDLL
6.2–6.6	3.0%	DDL
6.6–7.0	2.0%	DD
7.0–7.4	3.0%	DLL

Table 1 - Example from Digital Equipment [19]

Description	#	Metric	Value
Sample HM	**99**	**HM**	**4.84**
Sample GM	**99**	**GM**	**4.92**
Sample AM	**99**	**AM**	**5.01**
Sample Std Dev	**99**	**STDV**	**.88**
Sample Median	**99**	**Mdian**	**5.0**
Subsample			**Estimated**
C: Cobol	4	GM	5.6
I: Fortran Integer	7	GM	4.3
F: Fortran Single	32	GM	4.7
D: Fortran Double	32	GM	5.3
L: LISP	24	GM	5.3

Table 2 – Statistics for Table 1

Usage

Near zero knowledge of workload

Some knowledge of workload

Good knowledge of workload

Near-perfect knowledge of workload

Assumptions for SERPOP

Programs and inputs should be Representative

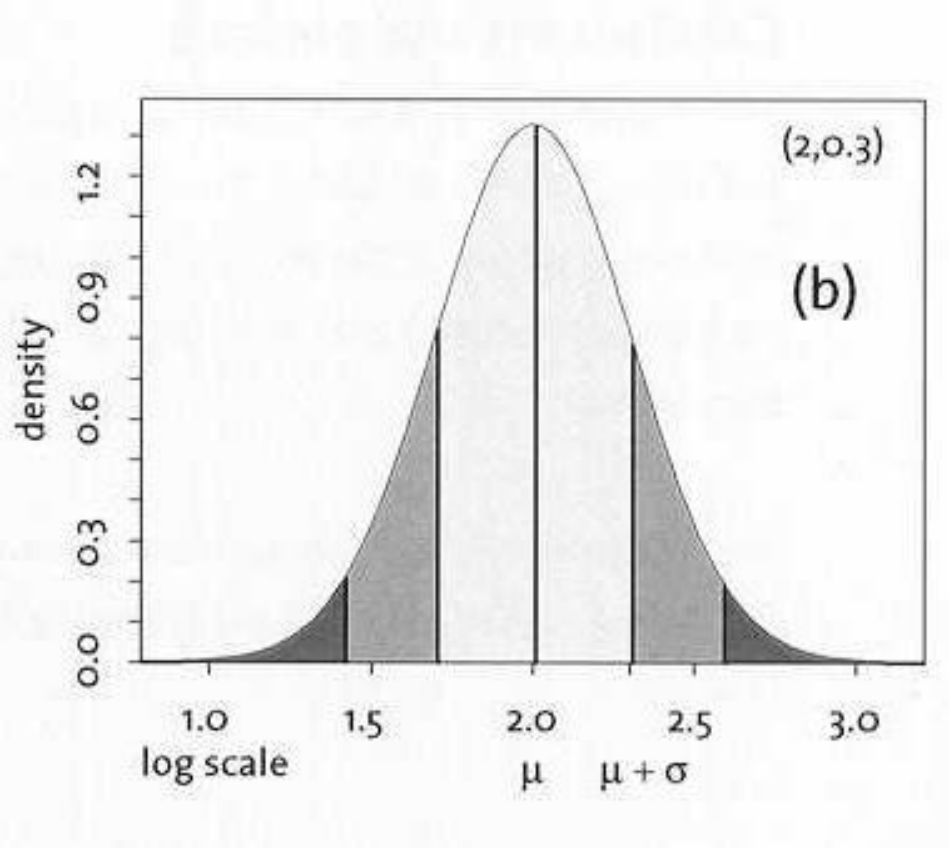
Small sample sizes should be avoided

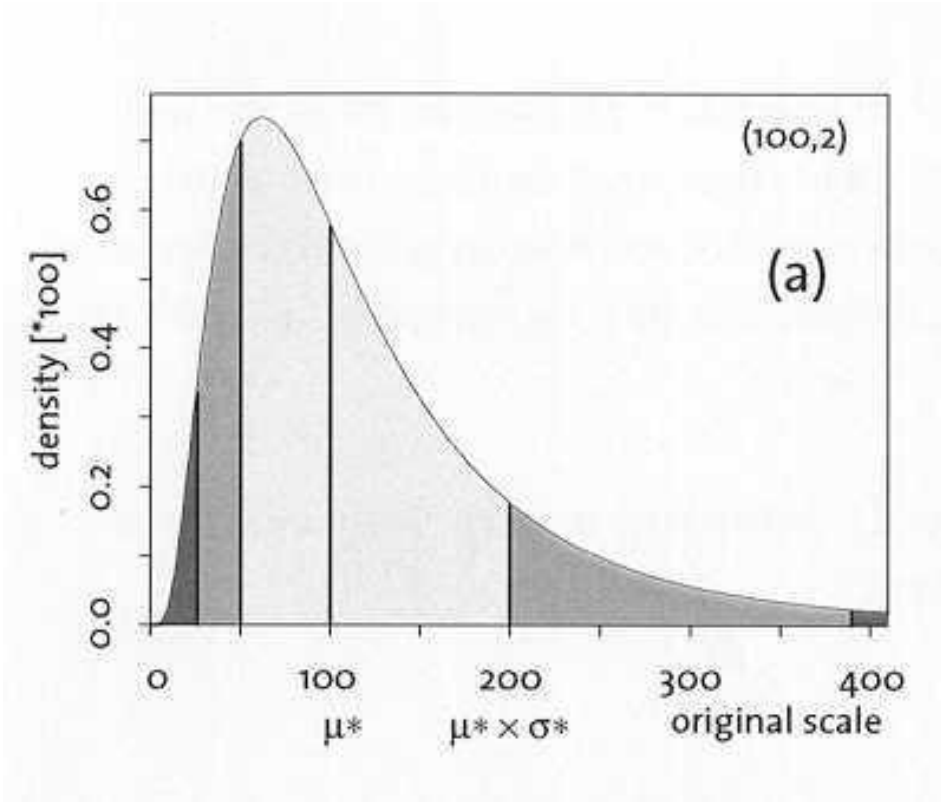
Measurements should be repeatable

Performance ratios must vary little across inputs

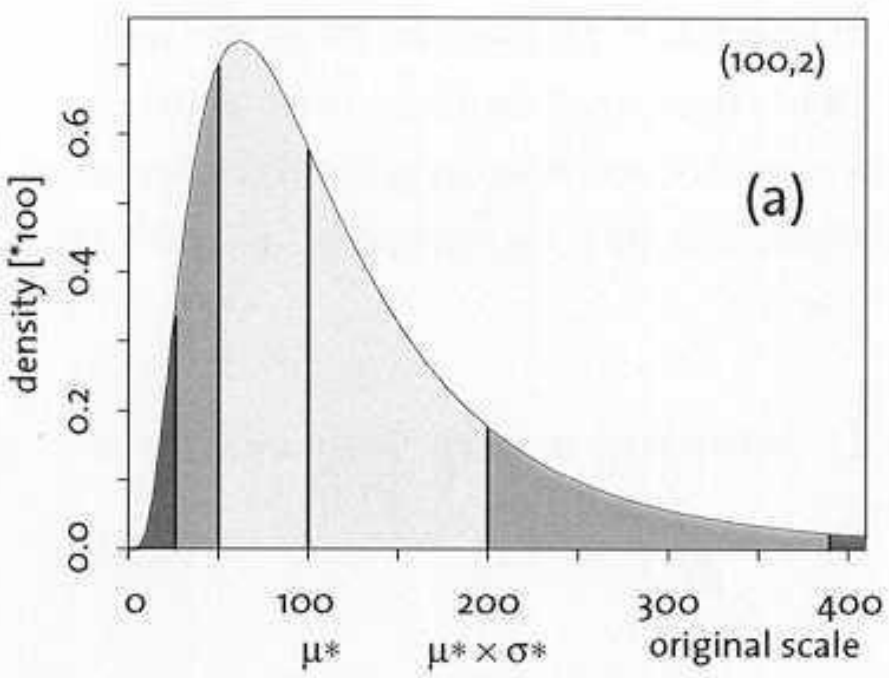
Programs should be recognizable

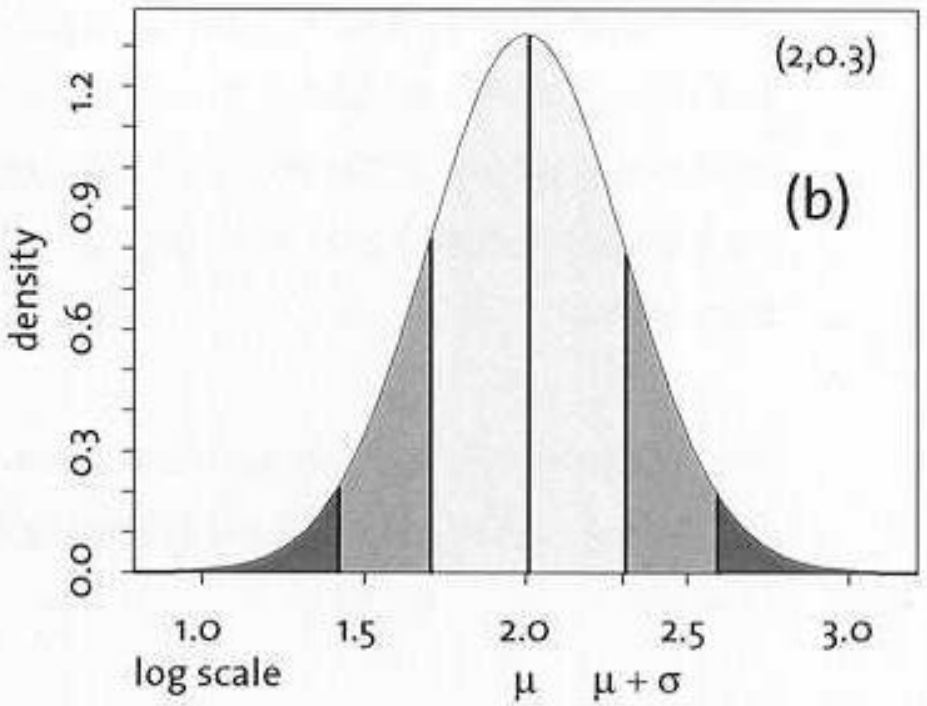
Normal and Log-Normal Distributions





Given a Log-Normal Distribution, do log transformation to convert it into Normal



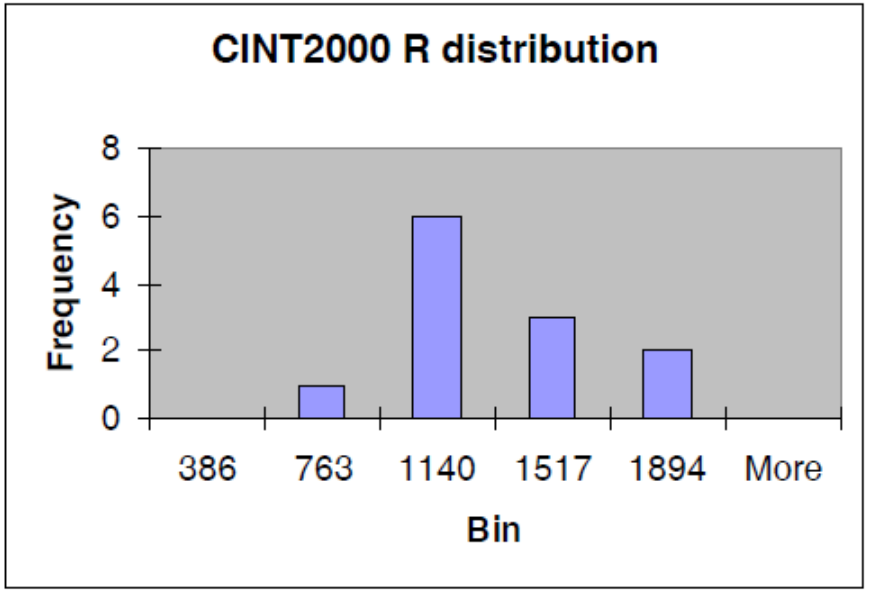


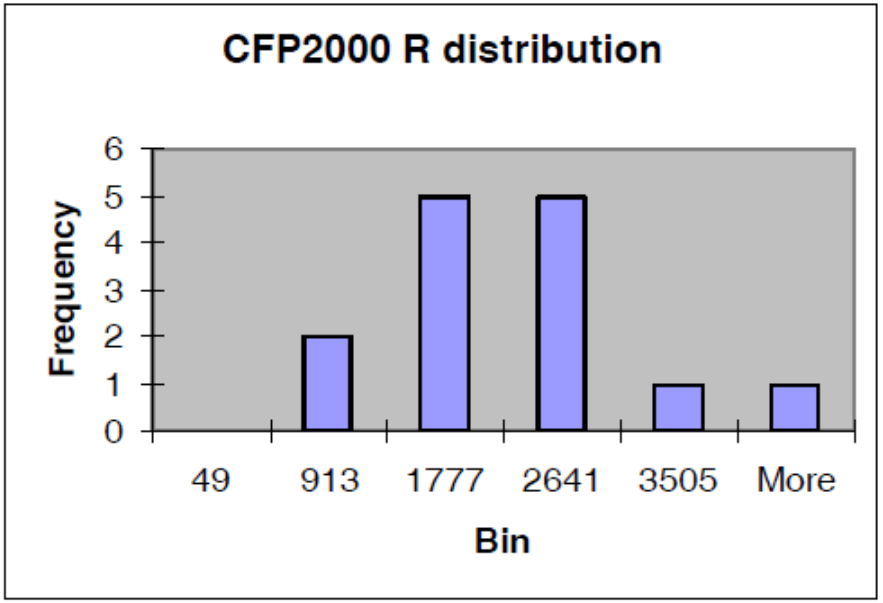
$$\overline{x_A} = \frac{1}{n}\sum_{i=1}^{n} \log_{10} x_i$$

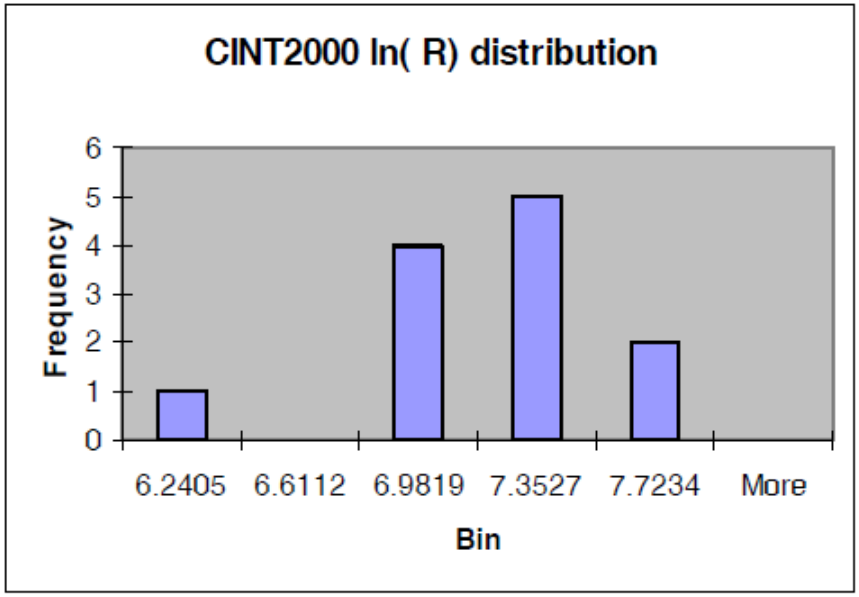
$$Mean = \exp(\overline{x_A}) = GM$$

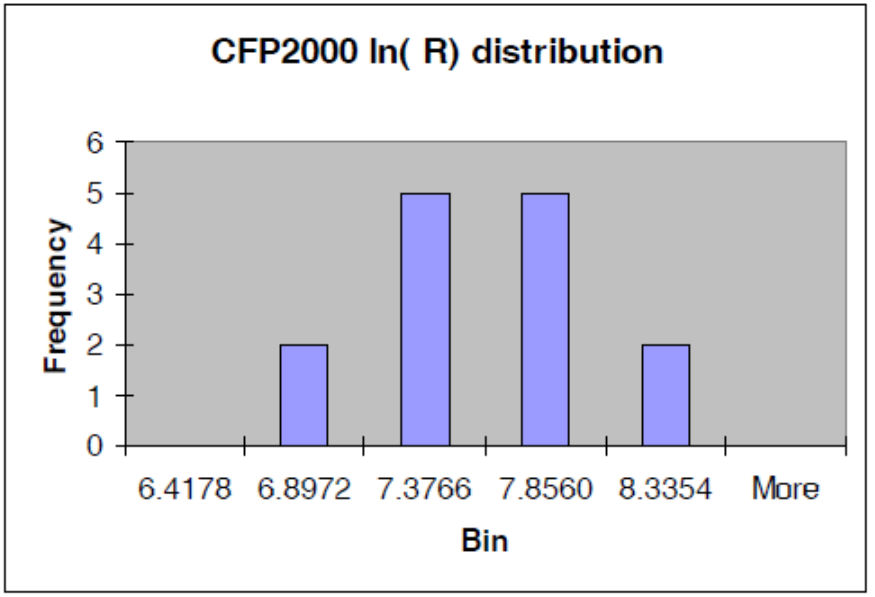
Long-tailed distributions

- A type of heavy-tailed distribution
- A probability distribution is said to have a long tail if a larger share of population rests within its tail than would under a normal distribution.
- A long-tail distribution will arise with the inclusion of many values unusually far from the mean, which increase the magnitude of the skewness of the distribution


CINT2000 R distribution
Frequency
0
2
4
6
8
386
763
1140
1517
1894
More
Bin


CFP2000 R distribution
Frequency
0
1
2
3
4
5
6
49
913
1777
2641
3505
More
Bin


CINT2000 ln(R) distribution
Frequency
0
1
2
3
4
5
6
6.2405
6.6112
6.9819
7.3527
7.7234
More
Bin


CFP2000 ln(R) distribution
Frequency
0
1
2
3
4
5
6
6.4178
6.8972
7.3766
7.8560
8.3354
More
Bin

	S1 Ti1	S2 Ti2	S2 Ri21*	S2 ln(R..)	S2 z(ln..)
197.parser	1800	398.0	452	*6.11*	*-2.34*
164.gzip	1400	175.0	800	*6.68*	*-0.80*
253.perlbmk	1800	217.0	829	*6.72*	*-0.70*
186.crafty	1000	106.0	943	*6.85*	*-0.36*
175.vpr	1400	133.0	1053	*6.96*	*-0.06*
254.gap	1100	98.6	1116	*7.02*	*0.10*
256.bzip2	1500	133.0	1128	*7.03*	*0.12*
176.gcc	1100	96.4	1141	*7.04*	*0.16*
252.eon	1300	96.7	1344	*7.20*	*0.60*
300.twolf	3000	213.0	1408	*7.25*	*0.72*
255.vortex	1900	114.0	1667	*7.42*	*1.18*
181.mcf	1800	99.8	1804	*7.50*	*1.39*
MEDIAN	1450	124	1122	*7.02*	
HM	1462	131	1005	*6.96*	
GM	1521	141	1077	*6.97*	
AM	1592	157	1140	*6.98*	
STDEV	540	88	377	*0.37*	
SKEW	1.64	2.16	0.12	*-0.96*	
KURTOSIS	3.74	5.19	0.07	*1.76*	**EXP**
95% lo	1252	102	903	*6.75*	853
95% hi	1591	155	1140	*7.22*	1360

	S1 Ti1	S2 Ti2	S2 Ri21*	S2 ln(R..)	S2 z(ln..)
200.sixtrack	1100	152.0	724	6.58	-1.65
177.mesa	1400	164.0	854	6.75	-1.31
172.mgrid	1800	173.0	1040	6.95	-0.90
188.ammp	2200	210.0	1048	6.95	-0.88
191.fma3d	2100	163.0	1288	*7.16*	-0.45
301.apsi	2600	193.0	1347	*7.21*	-0.36
173.applu	2100	151.0	1391	*7.24*	-0.29
189.lucas	2000	109.0	1835	*7.51*	0.29
187.facerec	1900	98.5	1929	*7.56*	0.39
171.swim	3100	145.0	2138	*7.67*	0.61
168.wupwise	1600	72.1	2219	*7.70*	0.68
179.art	2600	112.0	2321	*7.75*	0.78
183.quake	1300	44.1	2948	*7.99*	1.28
178.galgel	2900	76.4	3796	*8.24*	1.80
MEDIAN	2050	148.0	1613	*7.38*	
HM	1881	111	1440	*7.35*	
GM	1970	123	1598	*7.36*	
AM	2050	133	1777	*7.38*	
STDEV	597	49	864	*0.48*	
SKEW	0.19	-0.29	0.99	*0.07*	
KURTOSIS	-0.69	-0.77	0.83	*-0.69*	**EXP**

MEDIAN	1450	124	1122	*7.02*	
HM	1462	131	1005	*6.96*	
GM	1521	141	1077	*6.97*	
AM	1592	157	1140	*6.98*	
STDEV	540	88	377	*0.37*	
SKEW	1.64	2.16	0.12	*-0.96*	
KURTOSIS	3.74	5.19	0.07	*1.76*	**EXP**
95% lo	1252	102	903	*6.75*	853
95% hi	1591	155	1140	*7.22*	1360
Bin<m-2s	511	-19	386	*6.24*	513
Bin<m-s	1052	69	763	*6.61*	743
Bin<m	1592	157	1140	*6.98*	1077
Bin<m+s	2132	244	1517	*7.35*	1560
Bin<m+2s	2672	332	1894	*7.72*	2261
COEDET			0.940	0.88	

Histogram	**R..**		**Hist**	**ln(R..)**
Bin	**Freq**		**Bin**	**Freq**
386	0		6.24	1
763	1		6.61	0
1140	6		6.98	4
1517	3		7.35	5
1894	2		7.72	2
More	0		More	0

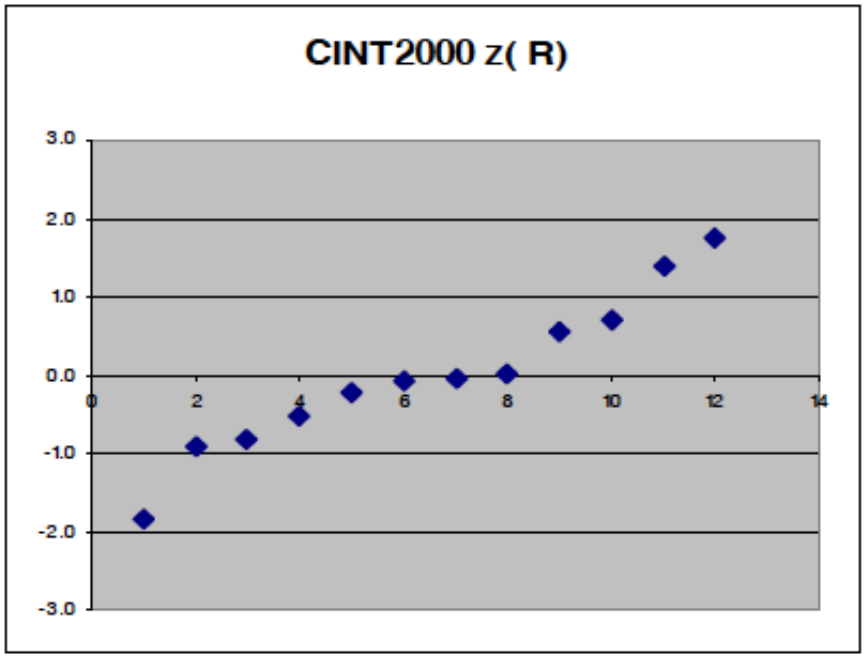
Table 8 - SPEC CINT2000 SERPOP Analysis

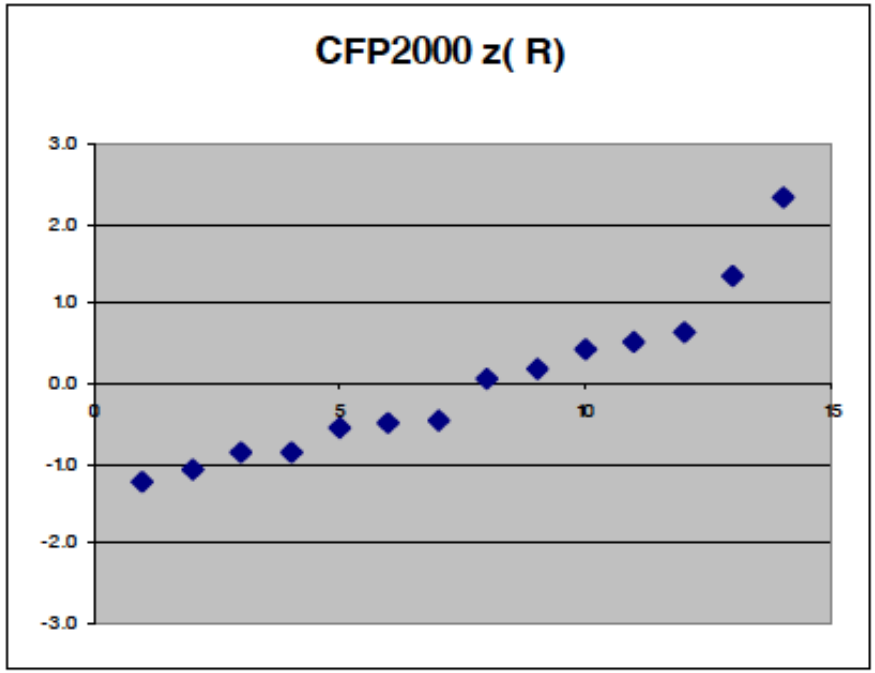
On the next page are plotted distributions using the bins above, and z-plots that should approximate straight lines

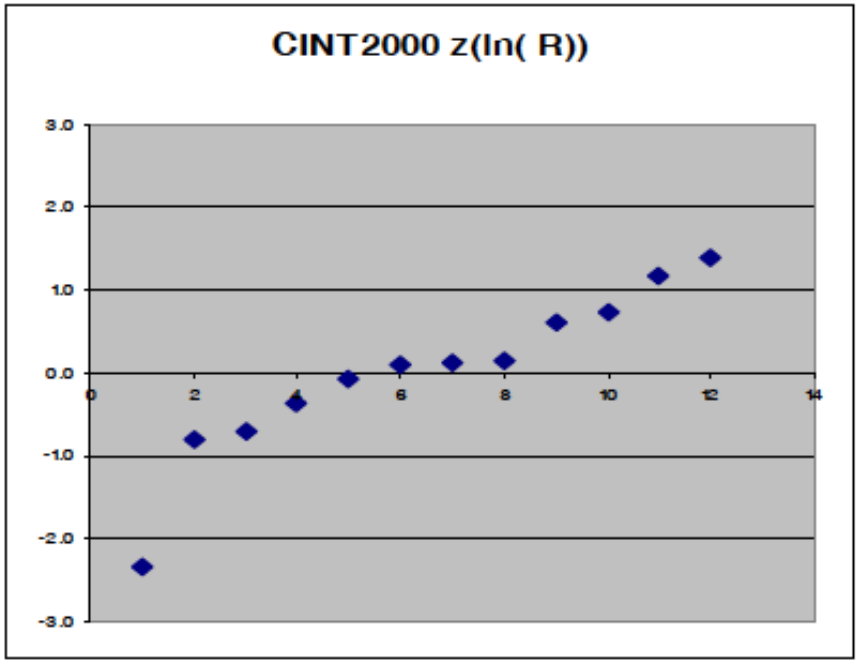
183.quake	1300	44.1	2948	*7.99*	1.28
178.galgel	2900	76.4	3796	*8.24*	1.80
MEDIAN	2050	148.0	1613	*7.38*	
HM	1881	111	1440	*7.35*	
GM	1970	123	1598	*7.36*	
AM	2050	133	1777	*7.38*	
STDEV	597	49	864	*0.48*	
SKEW	0.19	-0.29	0.99	*0.07*	
KURTOSIS	-0.69	-0.77	0.83	*-0.69*	**EXP**
95% lo	1708	105	1282	*7.10*	1214
95% hi	2392	161	2272	*7.65*	2104
Bin<m-2s	856	36	49	*6.42*	613
Bin<m-s	1453	84	913	*6.90*	989
Bin<m	2050	133	1777	*7.38*	1598
Bin<m+s	2647	182	2641	*7.86*	2581
Bin<m+2s	3244	230	3505	*8.34*	4169
COEDET			0.900	0.977	

Histogram	**R..**		**Hist**	**ln(R..)**
Bin	Freq		*Bin*	Freq
49	0		*6.418*	0
913	2		*6.897*	2
1777	5		*7.377*	5
2641	5		*7.856*	5
3505	1		*8.335*	2
More	1		*More*	0

Table 9 - SPEC CFP2000 SERPOP Analysis







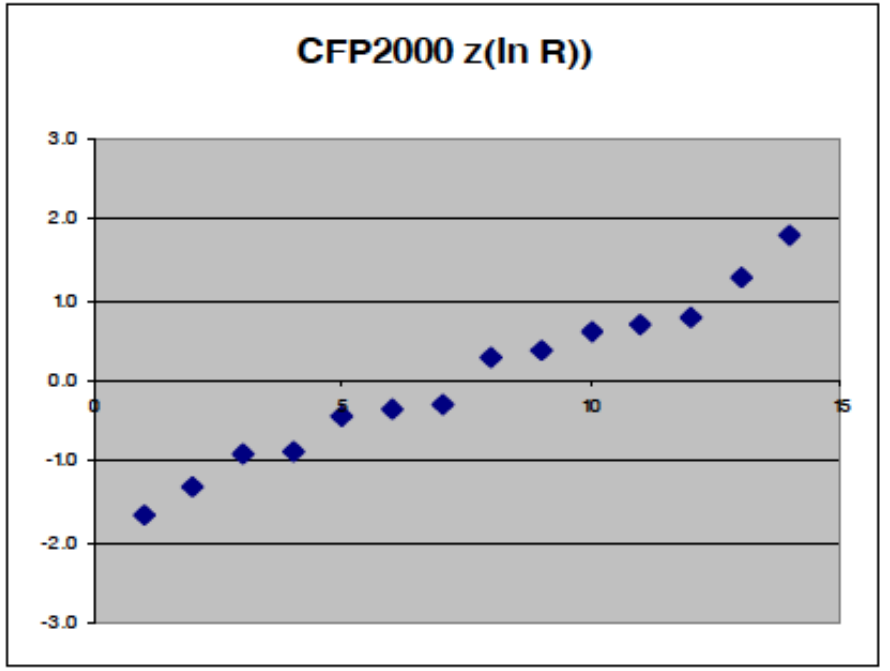
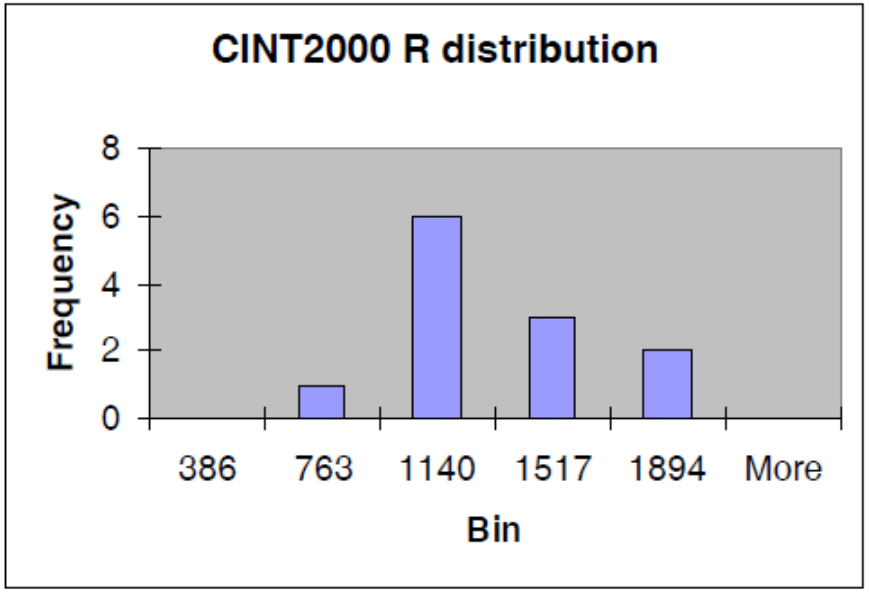
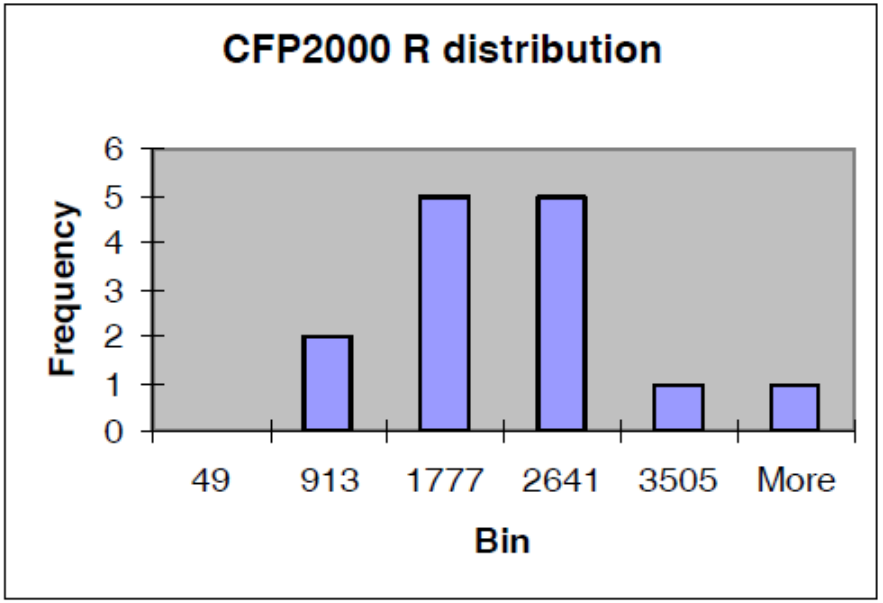
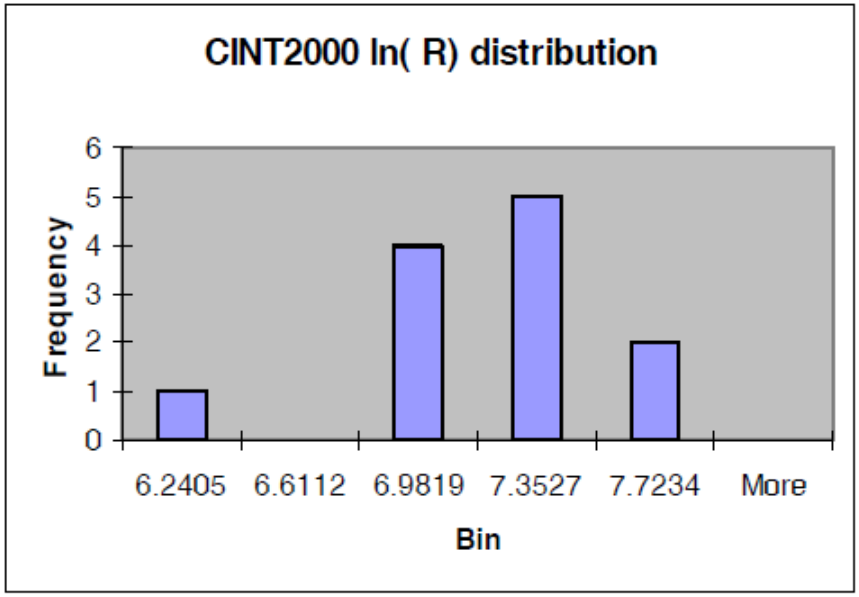
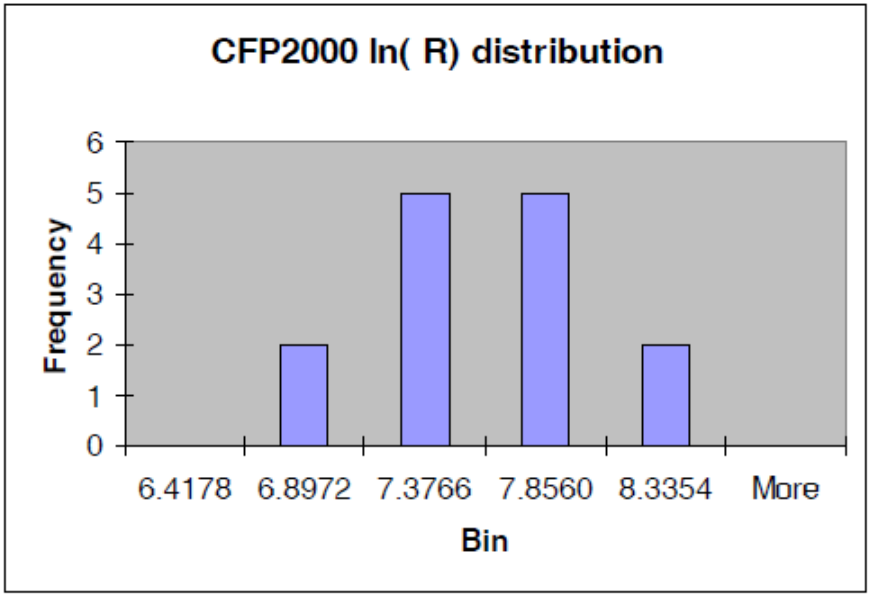


Table 10 - Distributions and z-scores


CINT2000 R distribution
Frequency
0
2
4
6
8
386
763
1140
1517
1894
More
Bin


CFP2000 R distribution
Frequency
0
1
2
3
4
5
6
49
913
1777
2641
3505
More
Bin


CINT2000 ln(R) distribution
Frequency
0
1
2
3
4
5
6
6.2405
6.6112
6.9819
7.3527
7.7234
More
Bin


CFP2000 ln(R) distribution
Frequency
0
1
2
3
4
5
6
6.4178
6.8972
7.3766
7.8560
8.3354
More
Bin