

***Comments on
“BIG Annual Report ‘15 –
Wastewater Treatment Facilities, Section 1.1,
“What is a Geomean?””***

*26 January, 2016
R.D. Pavlovich
Dsk RDP 742 HD 71(15-4057)*

What is a Geomean?

Bacteria data are often summarized using a geometric mean. ~~Outliers and~~ Extreme values are common in such data, and the geometric mean (or geomean) is not as sensitive to them as an arithmetic mean. ~~H-GAC calculates the mean of the natural logarithms of each bacteria value and then converts the logarithm back into a number by exponentiation.~~ Geomean is defined as The n th root of the product of n numbers.

Concerns:

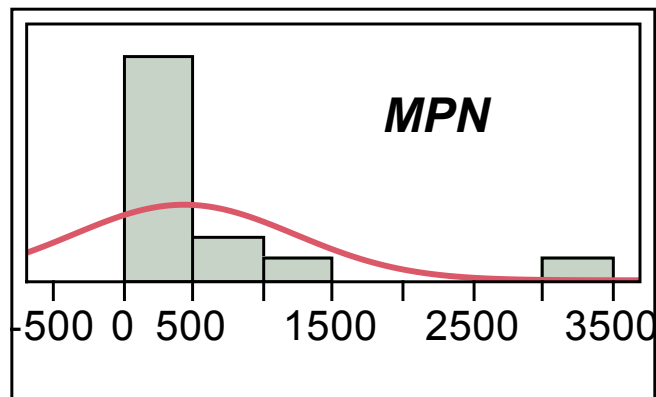
- 1. In order to conduct statistical procedures data must be more or less “normal” (Gaussian). Bacteria data are usually exponential or “log-normal” in nature and, therefore, must be transformed by using natural logarithms.**
- 2. Geomean does not address outliers.**

Discussion follows

From: Section 9-11. Table 90920:VIII, "Standard Methods."

Table 1

Data Point	MPN
1	11
2	27
3	36
4	48
5	80
6	85
7	120
8	130
9	136
10	161
11	317
12	601
13	760
14	1020
15	3100



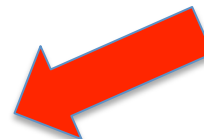
Summary Statistics

Mean	442.13333
Std Dev	795.13044
Upper 95% Mean	882.46189
Lower 95% Mean	1.804775
N	15
Geometric Mean	152.21841

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.568391	<.0001*

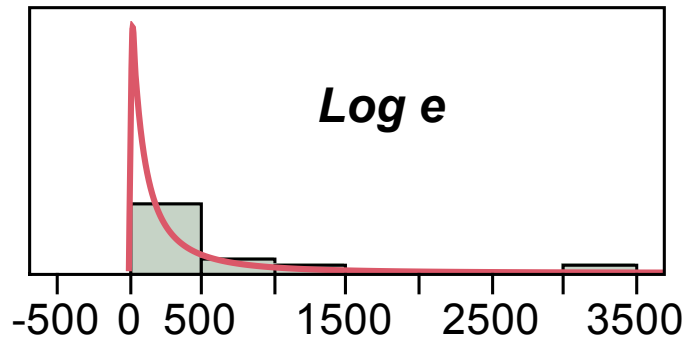


Note: H_0 = The data is from the Normal distribution.
Small p-values reject H_0 .

Raw Data Not Normally Distributed

Table 1A

Data Point	MPN	Log e
1	11	2.40
2	27	3.30
3	36	3.58
4	48	3.87
5	80	4.38
6	85	4.44
7	120	4.79
8	130	4.87
9	136	4.91
10	161	5.08
11	317	5.76
12	601	6.40
13	760	6.63
14	1020	6.93
15	3100	8.04
$\Sigma \rightarrow$		75.38
Arithmeric Mean, $\bar{x}_A =$		442
Geometric Mean, $\bar{x}_G =$		152



Summary Statistics

Mean	442.13333
Std Dev	795.13044
Upper 95% Mean	882.46189
Lower 95% Mean	1.804775
N	15
Geometric Mean	152.21841

Goodness-of-Fit Test

Kolmogorov's D

D	Prob>D
0.151319	> 0.1500

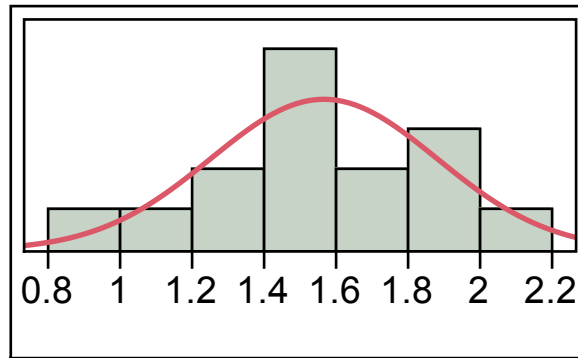


Note: H_0 = The data is from the LogNormal distribution.
Small p-values reject H_0 .

Log e Transformation Valid

Table 1A

Data Point	MPN	Log e
1	11	2.40
2	27	3.30
3	36	3.58
4	48	3.87
5	80	4.38
6	85	4.44
7	120	4.79
8	130	4.87
9	136	4.91
10	161	5.08
11	317	5.76
12	601	6.40
13	760	6.63
14	1020	6.93
15	3100	8.04
Σ →		75.38
Arithmeric Mean, $\bar{x}_A =$		442
Geometric Mean, $\bar{x}_G =$		152



Summary Statistics

Mean	1.570046
Std Dev	0.315543
Upper 95% Mean	1.7447879
Lower 95% Mean	1.3953041
N	15

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.976155	0.9365

Note: H_0 = The data is from the Normal distribution.
Small p-values reject H_0 .

$$\text{Mean} = \frac{\log x_1 + \log x_2 + \dots + \log x_n}{n}$$

From the Identity:

$$x_G = \text{Geo Mean} = [(x_1)(x_2) \dots (x_n)]^{\frac{1}{n}} = \sqrt[n]{(x_1)(x_2) \dots (x_n)}$$