

Entropy Nucleus and Use in Waste Disposal Policies

Ramalingam Shanmugam

School of Health Administration, Texas State University, San Marcos, TX78666, USA

ABSTRACT

The central theme of this article is that the usual Shannon's entropy [1] is not sufficient to address the unknown Gaussian population average. A remedy is necessary. By peeling away entropy junkies, a refined version is introduced and it is named nucleus entropy in this article. Statistical properties and advantages of the Gaussian nucleus entropy are derived and utilized to interpret 2005 and 2007 waste disposals (in 1,000 tons) by fifty-one states (including the District of Columbia) in USA. Each state generates its own, imports from a state for a revenue, and exports to another state with a payment waste disposal [2]. Nucleus entropy is large when the population average is large and/or when the population variance is lesser. Nucleus entropy advocates the significance of the waste policies under four scenarios: (1) keep only generated, (2) keep generated with receiving in and shipping out, (3) without receiving in, and (4) without shipping out. In the end, a few recommendations are suggested for the waste management policy makers.

KEYWORDS

Shannon's entropy, p-value, statistical power, hypothesis testing

1. MOTIVATION TO REFINE SHANNON'S ENTROPY

What is an entropy? In essence, an *entropy* or its equivalent nomenclature *information* is a knowledge basis, which helps to change one's opinion. Fisher [3] initiated the very thought of information in statistical contexts. In 1948, Shannon promoted logarithmic measure of the information to quantify signals in communication disciplines. *Shannon* named his seminal idea to measure information as entropy. Why did he do so? Entropy is an elusive but a useful concept.

An entropy should be quantifiable, partially orderable, additive, storable and transmittable. Claude Shannon himself mentioned: "My greatest concern was what to call it. I thought of calling it *information*, but the word was overly used, as I decided to call it *uncertainty*. When I discussed it with John von Neumann, he had a better idea. Von told me; you should call it *entropy* for two reasons. In the first place, your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, more important, *nobody knows what entropy really is*, so in a debate you will always have the advantage" (Tribus and McIrvine [4]).

Does entropy refer *uncertainty* or *dis (order)*? Shannon interpreted data information as a positive entropy. If so, it creates conflicts. The greater information ought to imply smaller entropy. The information might be lost in a process of transmission while the entropy might increase. This conflicting thought of *entropy* originated in quantum physics (Jaynes [5]). However, the entropy concept is utilized in economics and statistics among other disciplines with a contextual interpretation opposite to what Shannon intended (Lippman and McCall [6]). The entropy of a continuous random variable (RV) may be negative and of a discrete RV may even be infinite. Such controversies lead to give up on entropy as stated in Ben-Naim [7].

Shannon's entropy possesses several useful properties but not the much-needed additive property, which is a requirement in data analysis. When an additional observation becomes available, the expected entropy ought to increase. The Shannon's entropy does not do so and hence, it needs a modernization. This article modernizes Shannon's entropy by peeling away unnecessary entropy junkies in it and names the new version *nucleus entropy*. In particular, the properties of the Gaussian nucleus entropy are derived and illustrated using the log-transformation of the generated, shipped, and received waste disposals (in 1,000 tons) among the fifty-one (including the Washington, District of Columbia) states in USA in [2].

2. NUCLEUS ENTROPY DEFINED WITH PROPERTIES

What is Gaussian population? A brilliant French mathematician with the name *Abraham de Moivre* [8] wrote a self-published seven pages description of a bell shaped curve. Only 71 years later, a German astronomer Johann Carl Frederich Gauss [9] utilized de Movire's idea to model the errors between the actual and projected position of the celestial bodies of our universe. As a misnomer, the bell shaped curve is recognized as *Gaussian* not *de Movire's* population frequency curve once a random sample is drawn from it.

Before discussing further, consider the Gaussian population frequency curve

$$f(y|\mu, \sigma^2) = e^{-(y^2 - 2y\mu + \mu^2)/2\sigma^2} / \sqrt{2\pi\sigma^2}; -\infty < \mu < \infty; \sigma^2 > 0; -\infty < y < \infty \quad (1)$$

where μ is an unknown natural parameter and σ^2 is a known shape parameter. The Shannon's entropy $H(\cdot)$ is then

$$H_{Gauss}(y|\mu, \sigma^2) = -\int f(y|\mu, \sigma^2) \ln f(y|\mu, \sigma^2) dy = -\frac{1}{2} \ln \sqrt{2\pi e\sigma^2}. \quad (2)$$

Notice that its natural parameter μ is not even a part of the entropy. Furthermore, the Shannon's entropy echoes a conflict. To see it, suppose that a random sample y_1, y_2, \dots, y_n is drawn from a Gaussian population $f(y|\mu, \sigma^2)$. It is known (Mood, Graybill, and Boes [10]) that a sum $s = y_1 + y_2 + \dots + y_n$ of n independent and identically normally distributed outcomes follows a *normal probability structure* $f(s|n\mu, n\sigma^2)$. In which case, the Shannon entropy of the sum $s = y_1 + y_2 + \dots + y_n$ ought to be n times $H_{Gauss}(y|\mu, \sigma^2)$. But, it did not happen so. The Shannon's entropy of the sum is not additive of the individual entropies. That is, $H_{Gauss}(s|n\mu, n\sigma^2) \neq nH_{Gauss}(y|\mu, \sigma^2)$. For entropy practitioners, this causes confusion, as the Shannon's entropy is not adding up as a new Gaussian observation becomes available. The existence of such a deficiency in Shannon's entropy is a sufficient reason to modernize the entropy idea in an alternative way and it is what exactly done in this article.

In other words, this article introduces a new and novel approach based on *Gaussian nucleus entropy*. That is,

Definition 1. A *nucleus entropy*, N_y^{μ, σ^2} resides in the Gaussian population frequency curve once it is written as $f(y|\mu, \sigma^2) = A(\mu, \sigma^2)B(y, \sigma^2)N_y^{\mu, \sigma^2}$, with an *observation* y , a *natural* parameter

μ and an *entropy accumulator* parameter σ^2 . In other words, the Gaussian nucleus entropy is $N_{Gauss:Y}^{\mu, \sigma^2} = e^{\frac{y\mu}{\sigma^2}}$.

The function $A(\mu, \sigma^2)$ is insulated from observation y . The function $B(y, \sigma^2)$ does not connect to the unknown natural parameter μ . In a sense, both the functions (that is, $A(\mu, \sigma^2)$ and $B(y, \sigma^2)$) are *entropy junkies*. Without losing any generality, the nucleus entropy could be expressed in a logarithmic scale just for the sake of a comparison with the Shannon's entropy, which is in a logarithmic scale. Notice that the nucleus entropy involves both the unknown *natural* parameter μ and an *entropy accumulator* parameter σ^2 . The Gaussian nucleus entropy is more appropriate, appealing, and meaningful than the Shannon's entropy. The expected *nucleus entropy* is $S_{Gauss:Y}^{\mu, \sigma^2} = E_f \{ \ln(N_{Gauss:Y}^{\mu, \sigma^2}) \}$ which simplifies to

$$S_{Gauss:Y}^{\mu, \sigma^2} = \int [A(\mu, \sigma^2)B(y, \sigma^2)N_Y^{\mu, \sigma^2}] \ln N_Y^{\mu, \sigma^2} = \frac{\mu^2}{\sigma^2}, \quad (3)$$

The sample counterpart of (3) is named *observable Gaussian nucleus entropy* and it is $O_{Gauss}^{\bar{y}, n} = \frac{\bar{y}^2}{\sigma^2}$. The nucleus entropy happens to be the squared inverse of the coefficient of variation (CV). Mood et al. [10] for the definition of CV. The entropy increases with an increase of the average and/or with a decrease of the variance. The sample mean \bar{y} is the maximum likelihood estimator (MLE) of the unknown natural parameter μ . The MLE is invariant (Mood et al. [10]). That is, the MLE of a function of the parameter is simply the MLE of a function of the parameter.

Practitioners would wonder: Is an observed Gaussian nucleus entropy statistically significant? An answer depends on the outcome of testing the null hypothesis $H_0 : S_{Gauss:Y}^{\mu, \sigma^2} = 0$ against an alternative hypothesis $H_1 : S_{Gauss:Y}^{\mu, \sigma^2} \neq 0$. For this purpose, we proceed as follows. The statistic $O_{Gauss}^{\bar{y}, n}$ asymptotically follows a normal distribution with an expected value $E(O_{Gauss}^{\bar{y}, n}) \approx S_{Gauss:Y}^{\mu, n}$ and variance $\text{var}(O_{Gauss}^{\bar{y}, n}) \approx [\partial_{\bar{y}=\mu} O_{Gauss}^{\bar{y}, n}]^2 \sigma^2 / n$ where the notations $\partial_{\bar{y}=\mu}$ and σ^2 denote respectively the derivative with respect to \bar{y} evaluated at μ and the $\text{var}(Y)$. Note $\hat{\text{var}}(\bar{y}) = \frac{\sigma^2}{n}$ and $|\hat{\partial}_{\bar{y}} O_{Gauss}^{\bar{y}, n}| = \frac{2\bar{y}}{\sigma^2}$. Hence, the score is $|Z_{Gauss}| = \frac{\bar{y}}{2} \sqrt{\frac{n}{\sigma^2}}$ to test whether the null hypothesis $H_0 : S_{Gauss:Y}^{\mu, \sigma^2} = 0$. Hence, the null hypothesis is rejected in favor of the alternative hypothesis $H_1 : S_{Gauss:Y}^{\mu, \sigma^2} \neq 0$ with a p-value.:

$$p\text{-value} \approx 2 \Pr(Z > \left| \frac{\bar{y}}{2} \sqrt{\frac{n}{\sigma^2}} \right|) \quad (4)$$

where Z is the standardized Gaussian random variable.

The (statistical) power of accepting a given true specific alternate value $S_{Gauss:Y}^{\mu, \sigma^2} = S_{Gauss:Y}^{\mu_1, \sigma^2}$ is

$$power \approx \Pr\left(Z < \left| \frac{z_{\alpha/2} \mu_1}{\bar{y}} \right| \right) \quad (5)$$

where $0 < \alpha < 1$ is a chosen significance level.

The above results are illustrated in the next section using logarithm of waste disposals (in 1,000 tons) generated, shipped out, and received in fifty-one (including Washington, D.C.) states of USA in [2].

2.1. Illustration with Waste Disposals in Fifty-One US States

Archeologists (William and Murphy [11]) prove that even in 6,500 B. C., the North American communities generated as much as 5.3 pounds of waste per day. In the current modern age of 21st century with a high quality life (which requires a large amount of material consumptions), the waste disposals are extremely large and become a challenge to environmentalists. The health perils persist when the waste disposals are not properly processed. Otherwise, the trashes might explode and/or contaminate the *land, air, and water* sources. The sanitation is ruined and it causes known and unknown viruses and later on illnesses. The management of waste disposals (including the health hazardous medical wastes) is known to be infectious and causing chronic diseases (Reinhardt and Gordon [12]). WHO reports that in 2002 alone, about 25.9% of all 14.7 million deaths worldwide are due to infection and it could have been averted with a hygienic living environment. A lesson is that the wastes must be properly collected, managed and disposed to maintain a hygienic living environment. Otherwise, the residents and visitors might undergo a health risk. Often, killer *methane gas* is released from the improperly maintained waste filled land sites. Many health hazards like *fire due to ignitability, bad smell due to corrosiveness, radioactivity in surrounding water sources, toxicity etc.* exist in such sites. Remedial actions include *recycling, neutralizing, incineration, destruction, and conversion to energy*. Effective on 5th May, 1992, the 172 countries worldwide started implementing the Basel Convention’s agreement to practice waste management using technologies [13]. In USA, the fifty-one (including Washington, D.C.) states practice shipping out and receiving in to their own generated waste. This article investigates the policy scenarios (with respect to shipping out and/or receiving in waste) using significant changes in the Gaussian nucleus entropies.

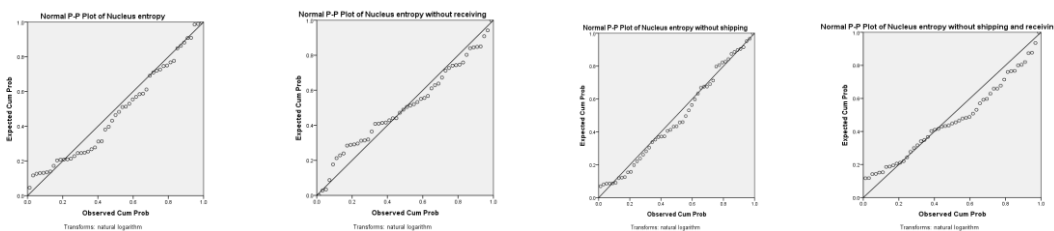


Figure 1. As it is Figure 2. No receiving Figure 3. No shipping Figure 4. Only generated

To be specific, the generated, Y_g , shipped out, Y_s and received in, Y_r waste disposals by the US states as displayed in [2] for the years 2005 and 2007 are considered and analyzed. The amounts $\ln(Y_g + Y_r - Y_s)$, $\ln(Y_g - Y_s)$, $\ln(Y_g + Y_r)$, and $\ln(Y_g)$ respectively represent the waste under current policy with shipping out and receiving, under a policy of cancelling receiving, under a policy of stopping shipping out, and under a policy of just doing without shipping out and receiving in. The amounts follow Gaussian frequency curve (because dots are closer to diagonal line in P-P plots in Figures 1 through 4) with averages μ_{grs} , μ_{gs} , μ_{gr} and μ_g and variances σ_{grs}^2 , σ_{gs}^2 , σ_{gr}^2 and σ_g^2 .

The observed Gaussian nucleus entropy $O_{Gauss}^{\bar{y}, n=2}$ is calculated and displayed in Table 1 for the fifty-one states along with their p-values, according to (4). When the p-value is smaller (0.05 or less), the alternative $H_1 : S_{Gauss, Y}^{\mu, \sigma^2} \neq 0$ is acceptable meaning that nucleus entropy about the natural parameter μ is significant. We notice the following. The generated waste is negligible only in states: *District of Columbia, Idaho, and South Dakota*, according to the nucleus entropy. The nucleus entropy of the current practice of receiving in and shipping out along with the generated waste in all the fifty-one states including *District of Columbia, Idaho, and South Dakota* is significant validating the current practice. If receiving in the waste is discontinued, the nucleus entropy becomes negligible in states: *Alabama, Arkansas, Arizona, California, District of Columbia, Hawaii, Idaho, Kentucky, Maine, Maryland, Michigan, Minnesota, Missouri, Montana, Nebraska, New Hampshire, New Jersey, North Carolina, Pennsylvania, Utah, Virginia, Washington, Wisconsin, and Wyoming* meaning that these states could consider cancelling the policy of receiving in the waste to their states. When the nucleus entropy remains negligible under a cancellation of shipping out waste, then those states may not consider shipping out and such states are: *Delaware, Hawaii, Maine, and Vermont*.

How sound is the methodology based on the nucleus entropy? To answer it, the statistical power is calculated using the national average for μ_1 in $H_1 : S_{Gauss, Y}^{\mu, \sigma^2} \neq 0$, under each of all four scenarios, according to (5). The minimum and maximum of the statistical powers across all four scenarios for each state are slated in Table 1. No minimum is lesser than 0.21 and the maximum is mostly 0.99 implying that the methodology is powerful enough.

3. COMMENTS AND RECOMMENDATIONS

Four scenarios to deal with the waste disposals are evaluated using nucleus entropies. They are: (1) keep only the generated waste, (2) receiving in waste in addition to generated waste without sending out any, (3) sending out a part of generated waste without receiving in additional waste, and (4) the current policy of receiving in and sending out waste in addition to generated waste. A large entropy is indicative of significant waste. One scenario does not fit all fifty-one states, according to the nucleus entropies. In an overall sense, some states perform similarly as a cluster (Figure 5).

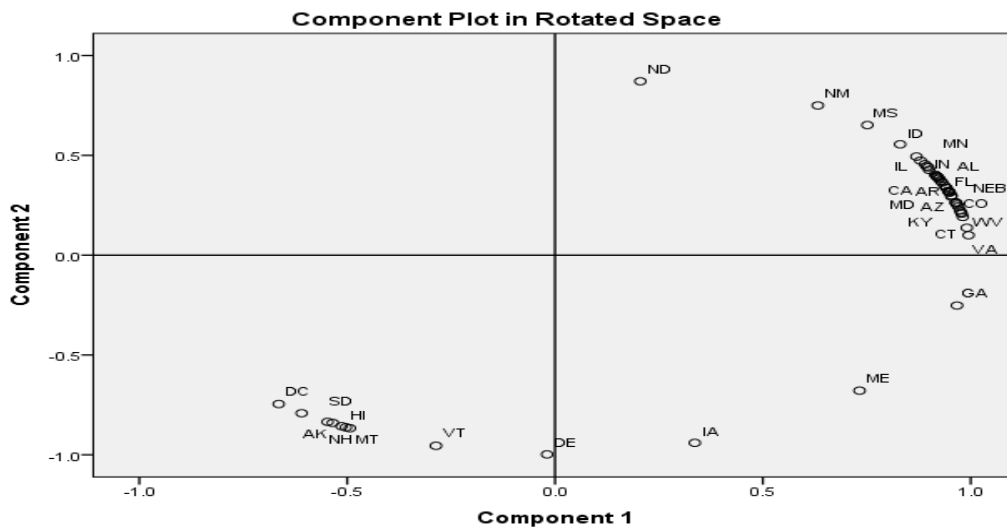


Figure 5. Clusters of states in the current practice of waste disposals

ACKNOWLEDGEMENTS

The author thanks Texas State University for a travel support to present this research work in the *International Conference on Recent Innovations in Engineering & Technology – 13-14 February 2015, Chinna Salem, Tamil Nadu, India.*

REFERENCES

- [1] C. E. A. Shannon (1948). Mathematical theory of communication. *Bell Sys. Tech. J.*, 1948, 27, 323-332; 379-423.
- [2] www.epa.gov/epawaste/inforesources/data/biennialreport/index.htm
- [3] R. A. Fisher (1925). Theory of Statistical Estimation. *Proceedings of the Cambridge Philosophical Society* **22** (5): 700–725.
- [4] M. Tribus and E. C. McIrvine (1971). Energy and information, *Scientific American* **224**:179–186.
- [5] E. T. Jaynes (1957). "Information theory and statistical mechanics" (PDF). *Physical Review* **106** (4): 620–630.
- [6] S. S. Lippman and J.J. McCall (2001). "Information, Economics of", *International Encyclopedia of the Social & Behavioral Sciences*, pp. 7480–7486.
- [7] A. Ben-Naim (2011). *A Farewell to Entropy: Statistical Thermodynamics Based on Information*. Singapore: World Scientific Press.
- [8] A. de Moivre (1738). *The Doctrine of Chances*. ISBN 0-8218-2103-2.
- [9] C. F. Gauss (1809). *Theoria motvs corporvm coelestivm in sectionibvs conicis Solem ambientivm* [*Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections*] (in Latin). English translation.
- [10] A. M. Mood, F. A. Graybill and D. C. Boes (1974). *Introduction to the Theory of Statistics*, New York: McGraw Hill Press.
- [11] R. William and C. Murphy (1992). *Rubbish! The Archaeology of Garbage*, New York, NY, Harper Collin Publishers.
- [12] P. A. Reinhardt, and J. G. Gordon (1991). *Infectious and medical waste management*. Chelsea, Michigan, Lewis Publishers.
- [13] <http://www.basel.int/Countries/StatusofRatifications/BanAmendment/tabid/1344/Default.aspx>.

TATBLE 1. Gauss nucleus entropy $O_{Gauss}^{\bar{y}, n=2}$ under four policies:

Sta tes	Waste generated		Waste with generating, receiving in, and sending out		Waste without receiving in		Waste without sending out		Power	
	Nucleus entropy	p- value	Nucleus entropy	P value	Nucleus entropy	P value	Nucleus entropy	P value	Mini mum	Maxi mum
AL	149.33	0.001	784.6822	0.001	1.903	0.14	666.57	0.001	0.76	0.87
AK	307.27	0.001	27.6477	2E-04	0.241	0.4	27.439	0.001	0.99	0.99
AZ	36.354	2E-05	676.829	0.001	0.251	0.88	86.178	0.001	0.94	0.99
AR	6157.2	0.001	4846.406	0.001	876.6	0.001	1937.6	0.001	0.74	0.99
CA	2017.1	0.001	194.1282	0.001	0.549	0.98	161.1	0.001	0.69	0.99
CO	119.95	1E-14	6389.492	0.001	3.16	0.08	4227.3	0.001	0.94	0.99

CT	287.25	0.001	203.9754	0.001	382.6	0.001	993.5	0.001	0.96	0.99
DE	163.37	0.001	12.10809	0.014	59.11	0.001	2.774	0.24	0.99	0.99
DC	1.178	0.4428	219615.9	0.001	0.001	0.62	117.5	0.001	0.69	0.99
FL	284.56	0.001	573.037	0.001	3.189	0.08	4929.3	0.001	0.75	0.92
GA	24.516	0.0005	13920.87	0.001	13.24	0.04	34.568	0.001	0.91	0.99
HI	5.4806	0.0978	37.59688	1E-05	4.626	0.31	3.0336	0.22	0.99	0.99
ID	5.3376	0.1023	245.1826	0.001	4.017	0.36	1297.8	0.001	0.81	0.99
IL	76420	0.001	1954.163	0.001	18.64	0.01	59727	0.001	0.66	0.8
IN	26276	0.001	7308.858	0.001	24919	0.001	4139	0.001	0.67	0.94
IA	5842.8	0.001	75.07698	9E-10	26769	0.001	218.32	0.001	0.98	0.99
KA	1031.3	0.001	1358.054	0.001	14.07	0.03	1676.9	0.001	0.75	0.98
KY	16.168	0.0045	50.84616	5E-07	0.018	0.69	42.437	0.001	0.82	0.97
LA	146.36	0.001	929.7007	0.001	33.33	0.001	417.7	0.001	0.4	0.68
ME	76.006	7E-10	0.212375	0.745	0.187	0.42	4.0711	0.15	0.99	0.99
MD	3182.6	0.001	226.9416	0.001	0.101	0.47	132.61	0.001	0.91	0.99
MA	396.63	0.001	268.1957	0.001	55.6	0.01	286.69	0.001	0.71	0.89
MI	20.727	0.0013	83.65424	1E-10	1.882	0.64	76.6	0.001	0.67	0.88
MN	63.819	2E-08	338.6517	0.001	0.974	0.23	191.25	0.001	0.71	0.93
MS	1004.6	0.001	1843.475	0.001	133.1	0.01	2620.8	0.001	0.29	0.79
MO	56.763	1E-07	389.6764	0.001	3.135	0.45	323.07	0.001	0.76	0.97
MT	7.2453	0.057	335.9534	0.001	2.834	0.49	18.012	0.001	0.74	0.99
NE	493.58	0.001	34399.47	0.001	0.575	0.3	7299.5	0.001	0.96	0.99
NV	183.35	0.001	582.0498	0.001	20.26	0.01	797.73	0.001	0.91	0.99
NH	424.86	0.001	4851403 1	0.001	0.31	0.37	3676.9	0.001	0.69	0.99
NJ	339.14	0.001	381.7374	0.001	2E-06	0.62	5532.2	0.001	0.77	0.99
NM	3E+10	0.001	843.8355	0.001	28360	0.001	381.36	0.001	0.25	0.9
NY	6960.8	0.001	1231.097	0.001	96.95	0.001	5758.5	0.001	0.67	0.8
NC	28.76	0.0001	17.34944	0.003	0.005	0.65	18.219	0.001	0.89	0.99
ND	191564	0.001	885.0282	0.001	1706	0.001	116.55	0.001	0.21	0.99
OH	1364.2	0.001	105938.4	0.001	35826	0.001	3359.4	0.001	0.65	0.95
OK	253.79	0.001	6984.599	0.001	8.924	0.01	19407	0.001	0.74	0.92
OR	83.575	1E-10	856.3681	0.001	9.004	0.01	2180.9	0.001	0.88	0.99
PA	12606	0.001	19056.32	0.001	7.274	0.16	77976	0.001	0.73	0.99
RI	60.517	4E-08	9.295961	0.031	22.24	0.001	8.9891	0.03	0.99	0.99
SC	2028.2	0.001	103306	0.001	583.6	0.001	19951	0.001	0.84	0.99
SD	0.5591	0.597	1785.35	0.001	70.25	0.001	49.721	0.001	0.99	0.99
TN	856.17	0.001	478.5049	0.001	68.01	0.001	569.76	0.001	0.45	0.85
TX	9714	0.001	11243.15	0.001	6749	0.001	4555.9	0.001	0.44	0.66
UT	11144	0.001	4567.066	0.001	1.789	0.66	29631	0.001	0.84	0.99
VT	109.93	1E-13	162.5951	0.001	22.96	0.001	0.7019	0.55	0.99	0.99
VA	368.36	0.001	87.10448	4E-11	0.24	0.4	111.95	0.001	0.93	0.99
WA	36424	0.001	185.0243	0.001	3.132	0.45	2505.4	0.001	0.91	0.99
WV	13086	0.001	4901.977	0.001	3450	0.001	2766.3	0.001	0.96	0.99
WI	49.045	7E-07	267.538	0.001	1.843	0.65	144.88	0.001	0.88	0.99
WY	43.738	3E-06	5085.835	0.001	0.006	0.58	887.56	0.001	0.7	0.99

Author

Dr. Ram Shanmugam is currently a professor in the School of Health Administration in Texas State University, San Marcos, USA. He obtained a Ph. D, degree in applied statistics from the School of Business Administration, Temple University, Philadelphia. He has published more than 120 research articles in refereed national and international journals. He is an elected fellow of the prestigious International Statistical Institute. He is the



Editor-in-Chief of the international journals: [1] *Advances in Life Sciences and Health (ALSH)*.

In addition [2] *Global Journal of Research and Review*.