Scientific Research Publishing

Power Law for the Rates of Different Numbers of Chronic Diseases among Elderly Chinese People

Handong Li, Ping Gao

School of Systems Science, Beijing Normal University, Beijing, China
Email: lhd@bnu.edu.cn, gaoping4069111@126.com

Abstract

Chronic disease is an important factor that affect the health of elderly people. We analyzed the 2006 and 2010 data from the Chinese Urban and Rural Elderly Population Surveys, which are nationally representative surveys of elderly people aged 60 years and above. We found that there existed a typical power-law distribution for the rates of different numbers of chronic diseases among elderly Chinese people. A Kolmogorov-Smirnov test indicated that the result was robust, and the power exponents were approximately −2.5. In addition, a paired t-test was conducted, which demonstrated that the rates of different numbers of chronic diseases did not have significant urban-rural differences, time differences or gender differences.

Keywords

Chronic Disease, Power Law Distribution, Elderly People, Kolmogorov-Smirnov Test (K-S Test), Paired T-Test

1. Introduction

Chronic diseases are chronic noncommunicable diseases, and they include cardiovascular diseases, diabetes, chronic obstructive pulmonary disease, and chronic kidney disease. Chronic diseases not only seriously affect the quality of life of the elderly and their families but also cause an economic burden on the family and society. Many studies have postulated that chronic diseases are significant risk factors for the development of disabilities. Over the past 20 years, there has been an increase in the prevalence of chronic disease [1], and the majority of elderly people aged 65 years and older suffer from multiple chronic diseases [2] [3] [4] [5] [6].

China has become an aging society, and chronic diseases have become important factors that affect the health of elderly Chinese people [7] [8]. The most common cause of death among Chinese residents is chronic disease rather than infectious disease [9]

[10]. In 2010, the number of patients with chronic diseases reached 300 million, and 75% of the deaths were caused by chronic diseases [11].

Power laws in general have and continue to attract considerable attention in a wide variety of disciplines—from astronomy to demographics to software structure to economics to finance to zoology, and even to warfare [12] [13] [14]. Typically, an analyst must work with integer-valued random variable n, its observables (numbers of objects, people, cities, words, animals, corpses) are positive inters with $n \in \{1, 2, 3, \cdots\}$. Sometimes, the range of values is allowed to be infinite (at least in principle), and sometimes a hard upper bound N is fixed (e.g., the total population if one is interested in subdividing a fixed population into sub-classes). Probability distributions that are especially interesting arise from probability laws of the following form: $p_n \propto 1/n^z$.

Power law distributions describe common features of many complex systems. The power-law scaling observed in the primary statistical analysis is an important feature but is by far not the only feature that characterizes experimental data. It provides us with important information about a system's stability and evolution.

Although a power-law phenomenon usually exists in economic and social systems, few studies focus in the distributed characteristics of the number of chronic diseases in a special subpopulation such as elderly people. We used 2006 and 2010 data from the Chinese Urban and Rural Elderly People Survey published in 2012 to find the distribution of the rates of different numbers of chronic diseases.

The remainder of this paper is structured as follows. Section 2 introduces the data and methods, including the paired t-test, power-law distribution and Kolmogorov-Smirnov test. Section 3 provides an empirical study of the 2006 and 2010 data, and Section 4 presents our conclusions.

2. Methods

2.1. Data

The data used in this study were obtained from the 2006 and 2010 Chinese Urban and Rural Elderly Population Surveys, conducted by the China Research Center on Aging of the National Committee on Aging[1]. These data cover the following 20 provinces in China: North China—Beijing, Hebei, and Shanxi; Northeast China—Liaoning and Heilongjiang; East China—Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, and Shandong; Mid-South China—Henan, Hubei, Hunan, Guangdong, and Guangxi; Southwest China—Sichuan and Yunnan; and Northwest China—Shanxi and Xinjiang. The data sampling method was the same as that for the Fifth Population Census, and it was based on the distribution of the population 60 years and older; a quota from each of these six regions could be determined. Then, stratified sampling was used to confirm that the survey results represented the total elderly population in China. The main study cohort in 2010 was the same elderly population that was investigated in 2006; these two surveys obtained samples of 19,947 responses and 19,986 responses, respectively. The study

---

[1]This project has been performed three times in 2002, 2006 and 2010, respectively. Because there is a flaw in 2002 data, we only use the data from 2006 and 2010.

subjects included elderly individuals who ranged from 60 to 102 years of age.

The appropriate processing of the data from the two surveys was a key aspect of the analysis, and the data were selected as follows:

1) All of the indicators that we used here were default-free. Data is from household tracking, including many indicators such as gender, age, census register etc. We give up some data that indicators are not complete to ensure the reliability of data.

2) We discarded samples that reported inconsistent chronic conditions. Specifically, we excluded cases in which suffering from chronic diseases was indicated while not specifying any chronic disease.

3) In general, women cannot have prostatitis, and men cannot have gynecological disease; thus, we discarded data samples that reported impossible chronic diseases.

4) To facilitate the distinction between urban and rural, we used the agricultural and nonagricultural census registers (e.g., an individual who transferred from an agricultural to a nonagricultural census register was considered to be a nonagricultural census register).

5) There were 25 types of chronic diseases in the survey; the 25th was "other chronic disease", which was not a specific disease. One could choose this option if one is suffering from a type or more than one type of chronic disease that was not among the 24 types of chronic diseases in the survey. We hypothesize that one chose the option of "other chronic disease" was just suffering from a single type of chronic disease. Thus, the minimum number of chronic diseases that one person could suffer from was 0, and the maximum number, in theory, was 25.

Using the processing methods described above, the final number of available samples was 19,691 in 2006 and 19,841 in 2010.

2.2. Paired T-Test

Because most of chronic diseases are reversible, the morbidity of chronic diseases among elderly people of various ages does not have a cumulative effect. Thus, we could not use the distribution fitting method to conduct a variance analysis. Meanwhile, when analyzing the influence of some social-economic factor on chronic disease, we usually presume the influence of the factor is same on different elderly cohorts. Therefore, in this study, we used the paired t-test to perform the analysis.

The paired t-test can verify whether the effect of a factor is significant, For example, we can use paired t test to found out whether two weighing machines are the same. The objects to be weighted can be very different, it can be light, such as bag, shoe, cat, book and food, if possible, it can also be very heavy, such as refrigerator and elephant. We can get the weight of the objects by using the two machines, and then proceed paired t test to verify our hypothesis. If the machines have no system error, the mean of the weight differences should be close to 0, if else, we can infer that the two machines are different.

The theory of paired t-test is as follows:

Consider n pairs of independent observation data $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$. Let

$D_1 = X_1 - Y_1$, $D_2 = X_2 - Y_2$, ..., $D_n = X_n - Y_n$; thus, $D_1, D_2, \cdots, D_n$ are independent and identically distributed, in which the differences result from the same factor (for example, gender, time or census registry). For the hypothesis that $D_i \sim N\left(\mu_D, \sigma_D^2\right), i = 1, 2, \cdots, n$, μ_D and σ_D^2 are unknown, and then, the hypothesis is tested as follows: H_0: $\mu_D = 0$; H_1 $\mu_D \neq 0$. The t-test, statistics and rejection regions used in this study are obtained as follows:

$$|t| = \left| \frac{\bar{D}}{S_D / \sqrt{n}} \right| \geq t_{a/2}(n-1) \tag{1}$$

where $\bar{D}$, S_D and a are the sample mean, sample standard deviation and significance level, respectively.

2.3. Power-Law Distribution

In mathematics, if the density function of a random variable X is the following:

$$f(x) = Cx^{-a} \ (x > 0) \tag{2}$$

then we can say that x follows a power distribution, in which a is a constant that is called the power exponent. The power distribution showed very strong heterogeneity, and the power distribution is the only distribution that has the property of "no scaling".

When the density function is $f(x) = Cx^{-a}$,

$$f(bx) = C(bx)^{-a} = b^{-a}f(x) \propto f(x) \tag{3}$$

The meaning of the function is that when the unit or the "scale" of the variable x changes a constant number of times, the form of the distribution $f(x)$ remains unchanged. Taking logarithm on both sides of Equation (2), we get

$$\log(f(x)) = \log C - a\log(x) \tag{4}$$

From (4), we can see that a is the slope that reflects the rate of change between logarithmic probability density and logarithmic random variable.

2.4. Kolmogorov-Smirnov Test

In statistics, the Kolmogorov-Smirnov test (K-S test) is a form of minimum distance estimation that is used as a nonparametric test of equality of one-dimensional probability distributions, and it is used to compare a sample with a reference probability distribution (one-sample K-S test) or to compare two samples (two-sample K-S test).

Hypothesis test problems are as follows:

H0: the samples are drawn from the same distribution (in the two-sample case) or the sample is drawn from the reference distribution (in the one-sample case).

H1: the samples are not drawn from the same distribution (in the two-sample case) or the sample is not drawn from the reference distribution (in the one-sample case).

The Kolmogorov-Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution or between the empirical distribution functions of two samples.

The empirical distribution function F_n for n i.i.d. observations X_i is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i < x} \tag{5}$$

where $I_{X_i < x}$ is the indicator function, which is equal to 1 if $X_i < x$ and equal to 0 otherwise.

The Kolmogorov-Smirnov statistic for a given cumulative distribution function $F(x)$ is

$$D_n = \sup_x \left| F_n(x) - F(x) \right| \tag{6}$$

When $D_n > D(n, \alpha)$, we can reject the null hypothesis that the samples are drawn from the same distribution (in the two-sample case) or that the sample is drawn from the reference distribution (in the one-sample case); otherwise, we should accept the null hypothesis. Here, $D(n, \alpha)$ is the critical value at a significance level of a, and the sample size is n.

3. Data Analysis Results

3.1. Power-Low for the Rates of Different Numbers of Chronic Diseases by Gender and Census Registry

The fitting information on the samples by gender and census registry in 2006 and 2010 was shown in the following sections. The dependent variable was the rates of the different numbers of chronic diseases, and the independent variable was the number of chronic diseases.

The power-law fitting results for the rates of different numbers of chronic diseases by gender and census registry in 2006 and 2010 were shown in Table 1. For example, the power-law function of the urban samples in 2006 was $f(x) = 160.10 \times x^{-2.18}$. The 95% confidence intervals of the first coefficient were [112.40, 207.80], and the 95% confidence intervals of the second coefficient were [−2.41, −1.95]. The adjusted R-square of the fit was 0.9846, the SSE was 3.1630, and the RMSE was 0.5134.

Next, two-sample Kolmogorov-Smirnov tests were performed to verify that the distribution of the rates of different numbers of chronic diseases was a power-law distribution. The two-sample K-S test returns a test decision for the null hypothesis that the data in vectors ×1 and ×2 are from the same continuous distribution, and the alternative hypothesis is that ×1 and ×2 are from different continuous distributions. The result (h) is 1 if the test rejects the null hypothesis at the 5% significance level and 0 otherwise.

The two-sample K-S test results (h) of the fitting in 2006 and 2010 were all 0, and the asymptotic p-values were all greater than 0.05, which means that the K-S test did not reject the null hypothesis at the 5% significance level; in other words, the rates of different numbers of chronic diseases in 2006 and 2010 obeyed a power-law distribution.

Table 1 shown that the power exponents of male elderly and female elderly in 2006 and 2010 were close to a mean −2.5; the power exponents of the urban elderly in 2006 and 2010 were under −2.5, while the power exponents of the rural elderly in 2006 and 2010 were under −2.5.

Table 1. Power-law fitting results and K-S test.

	Fitting result ($f(x) = a \times x^b$)						K-S (95%)	
	a	b	SSE	R-square	Adjusted R-square	RMSE	h	p
2006								
Male	191.40	−2.56	0.5426	0.9960	0.9956	0.2126	0	0.2672
95% CI	(154.50, 228.40)	(−2.71, −2.40)						
Female	214.80	−2.46	2.2380	0.9896	0.9887	0.4319	0	0.1106
95% CI	(150.90, 278.60)	(−2.70, −2.22)						
Urban	160.10	−2.18	3.1630	0.9858	0.9846	0.5134	0	0.2672
95% CI	(112.40, 207.80)	(−2.41, −1.95)						
Rural	348.80	−3.12	0.5460	0.9955	0.9951	0.2337	0	0.1862
95% CI	(245.70, 451.90)	(−3.37, −2.87)						
2010								
male	225.20	−2.49	2.1500	0.9902	0.9894	0.4233	0	0.2672
95% CI	(159.20, 291.20)	(−2.73, −2.25)						
female	248.00	−2.45	3.3610	0.9886	0.9876	0.5292	0	0.2672
95% CI	(171.20, 324.90)	(−2.70, −2.20)						
urban	191.90	−2.23	3.8680	0.9865	0.9854	0.5678	0	0.2672
95% CI	(134.40, 249.30)	(−2.46, −1.99)						
rural	320.40	−2.80	2.4520	0.9886	0.9875	0.4951	0	0.0656
95% CI	(189.90, 451.00)	(−3.13, −2.46)						

3.2. Paired T-Test of the Census Registry and Gender and Power-Law Fit

This paper used the paired t-test to verify the census registry and the effects of gender on the rates of chronic diseases. The original data for the paired t-test were the rates of different numbers of chronic diseases. The minimum number in the survey was 0, which represents that one person was not suffering from any of the diseases, and the maximum number in the survey was 16, which represents that one person was suffering from 16 diseases at the same time.

3.2.1. Paired T-Test on the Census Registry

In this section, we provided the test results for the sub-groups by the census registry from two surveys, and the corresponding results were shown in Table 2.

As shown in Table 2, all of the paired t-test p-values were close to 1, which indicates that the rates of the different numbers of chronic diseases did not exist urban-rural differences.

3.2.2. Paired T-Test of Gender

In this section, we given the test results on the male elderly and female elderly groups, and the results were shown in Table 3.

As shown in Table 3, all of the paired t-test p-values were close to 1, which indicates that the rates of different numbers of chronic diseases did not exist gender differences.

Table 2. Paired t-test of the census registry.

| | Paired Differences | | | | | | |
| | Mean | Std. | 95% CI of the Difference | | t | df | Sig. (2-tailed) |
			Lower	Upper			
2006							
Urban male-rural male	−0.0012	3.8117	−1.9610	1.9586	−0.0013	16	0.9990
Urban female-rural female	0.0006	3.5912	−1.8458	1.8470	0.0007	16	0.9995
2010							
Urban male-rural male	0.0000	3.1917	−1.6410	1.6410	0.0000	16	1.0000
Urban female-rural female	−0.0012	2.5006	−1.2869	1.2845	−0.0019	16	0.9985

Table 3. Paired t-test of gender.

| | Paired Differences | | | | | | |
| | Mean | Std. | 95% CI of the Difference | | t | df | Sig. (2-tailed) |
			Lower	Upper			
2006							
Urban male-urban female	−0.0006	1.6671	−0.8577	0.8565	−0.0015	16	0.9989
Rural male-rural female	0.0012	1.9924	−1.0232	1.0256	0.0024	16	0.9981
2010							
Urban male-urban female	0.0000	1.3464	−0.6923	0.6923	0.0000	16	1.0000
Rural male-rural female	−0.0012	1.6839	−0.8669	0.8646	−0.0029	16	0.9977

3.2.3. Power Law of the Rates of Different Numbers of Chronic Diseases in 2006 and 2010

The paired t-test results showed that the rates of different numbers of chronic diseases did not have any urban-rural differences and gender differences. The results mean that the samples of urban male, urban female, rural male and rural female can be combined together to obtain the total number of samples in 2006 and 2010. Fitting information on the samples in 2006 and 2010 were shown in the following sections.

We further performed a two-sample K-S test to verify that the distribution of the rates of different numbers of chronic diseases was a power-law distribution. The test results (h) in 2006 and 2010 were all 0, and the asymptotic p-values were all 0.2672 > 0.05, which mean that the K-S test did not reject the null hypothesis at the 5% significance level. In other words, the rates of different numbers of chronic diseases in 2006 and 2010 obeyed a power-law distribution.

The power-law fitting results for 2006 and 2010 were shown in **Figure 1** and **Figure 2**, respectively.

The power-law function of the samples in 2006 was $f(x) = 201.7 \times x^{-2.50}$; the 95% confidence bounds of the first coefficient were [153.30, 250.10], and the 95% confidence bounds of the second coefficient were [−2.70, −2.31]. The adjusted R-square of the fit was 0.9929, the SSE was 1.112, and the RMSE was 0.3044.

The power-law function of the samples in 2010 was $f(x) = 236.00 \times x^{-2.47}$; the 95%
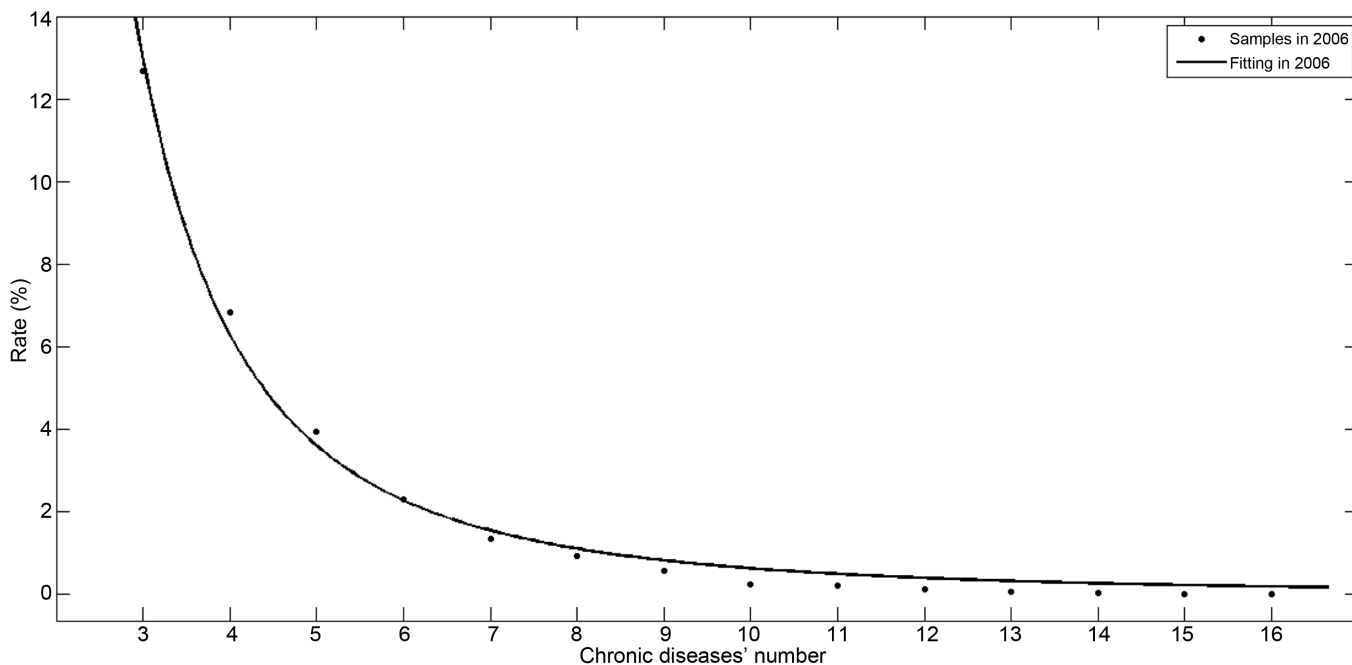
Figure 1. Power law for the rates of different numbers of chronic diseases in 2006.
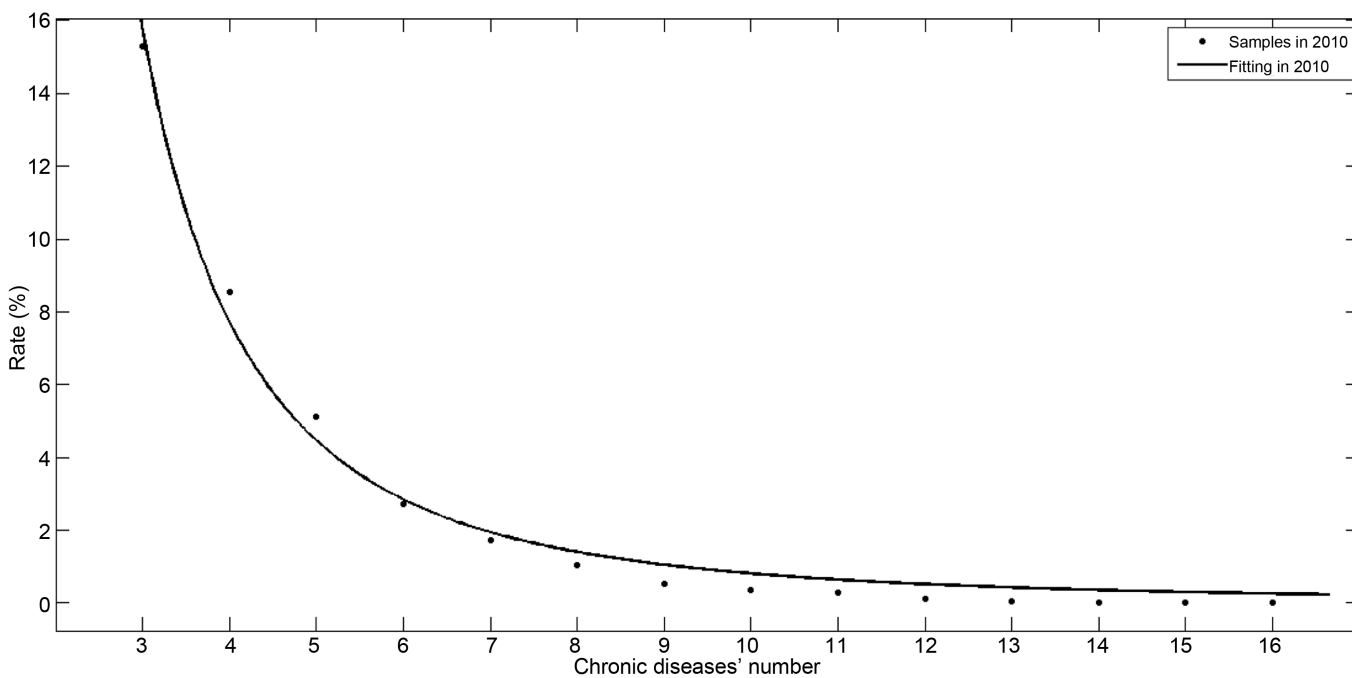
Figure 2. Power law for the rates of different numbers of chronic diseases in 2010.

confidence bounds of the first coefficient were [165.50, 306.50], and the 95% confidence bounds of the second coefficient were [−2.71, −2.23]. The adjusted R-square of the fit was 0.9887, the SSE was 2.6440, and the RMSE was 0.4694.

All of the analysis above showed that the rates of different numbers of chronic diseases had a power-law distribution. The power exponent of the samples in 2006 was −2.50,

and the power exponent of the samples in 2010 was −2.47, and they were approximately equal to −2.5.

3.3. Paired T-Test of the Time and Power-Law Fitting

3.3.1. Paired T-Test of Time

In this section, we given the test results on the time, and these test results can reveal the difference between 2006 and 2010. The results were shown in Table 4.

As shown in Table 4, all of the paired t-test p-values were close to 1, which indicates that the rates of different numbers of chronic diseases did not have time differences.

3.3.2. Power-Law Fitting of the Total Samples

The paired t-test results showed that the rates of different numbers of chronic diseases did not exist time differences. The results mean that the samples in 2006 and the samples in 2010 can be combined together to obtain the power-law fit for the rates of different numbers of chronic diseases (see Figure 3).

Table 4. Paired t-test of time.

	Paired Differences						
	Mean	Std.	95% CI of the Difference		t	df	Sig. (2-tailed)
			Lower	Upper			
Urban male in 2006-Urban male in 2010	−0.0006	1.1533	−0.5935	0.5924	−0.0021	16	0.9983
Urban female in 2006-Urban female in 2010	0.0006	2.3363	−1.2006	1.2018	0.0010	16	0.9992
Rural male in 2006-Rural male in 2010	0.0000	0.9137	−0.4698	0.4698	0.0000	16	1.0000
Rural female in 2006-Rural female in 2010	−0.0018	2.1160	−1.0897	1.0862	−0.0034	16	0.9973
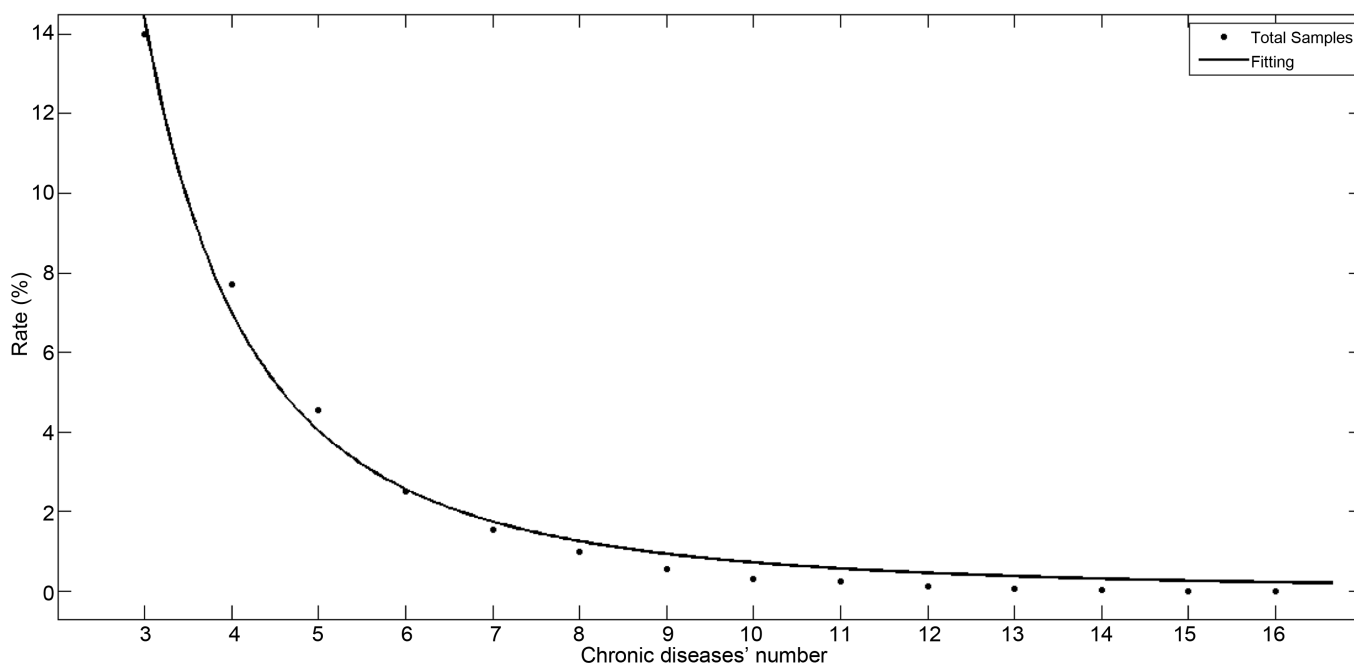
Figure 3. Power law for the rates of different numbers of chronic diseases.

The power-law function of the total samples was $f(x) = 218.80 \times x^{-2.48}$. The 95% confidence bounds of the first coefficient were [159.50, 278.10], and the 95% confidence bounds of the second coefficient were [−2.70, −2.26]. The adjusted R-square of the fit was 0.9908, the SSE was 1.7800, and the RMSE was 0.3851.

The K-S test result (h) was 0, and the asymptotic p-values were all 0.2672 > 0.05, which mean that the K-S test did not reject the null hypothesis at the 5% significance level. In other words, the rates of different numbers of chronic diseases followed a power-law distribution. Furthermore, the power exponent of the samples was −2.48, which was very close to mean −2.5. This indicates that the logarithmic number of chronic illness that an elderly had declines at a certain ratio as logarithmic age increases.

4. Conclusions

We used the data that were obtained from the 2006 and 2010 Chinese Urban and Rural Elderly Population Surveys to examine whether there was a power-law distribution of the number of chronic diseases among elderly Chinese people. The paired t-test method was used to analyze the urban-rural differences, time differences and gender differences, and then, we proceeded with power-low fitting of the data. The main conclusions were as follows: 1) All of the paired t-test p-values of the urban males and rural males in 2006 and 2010, and the urban females and rural females in 2006 and 2010 were close to 1, which mean that the rates of different numbers of chronic diseases did not exist urban-rural differences, and thus, the samples of the urban elderly and rural elderly can be combined together. 2) All of the paired t-test p-values of the urban males and urban females in 2006 and 2010, and the rural males and rural females in 2006 and 2010 were close to 1, which mean that the rates of different numbers of chronic disease did not exist gender differences. The samples of male elderly and female elderly can be combined together. 3) All of the paired t-test p-values of urban males in 2006, urban males in 2010, urban females in 2006, urban females in 2010, rural males in 2006, rural males in 2010, rural females in 2006 and rural females in 2010 were close to 1, which mean that the rates of different numbers of chronic diseases did not exist time differences, and thus, the samples from 2006 and 2010 can be combined together. 4) There was a power-law distribution of the rates of different numbers of chronic diseases; the power-law distribution was $f(x) = 218.80 \times x^{-2.48}$, x = 3, 4, …, 16. 5). The power exponents were approximately −2.5.

In the future, we wish to explore whether the distribution of chronic diseases has differences among the different types of chronic diseases, such as fatal chronic disease and non-fatal chronic disease, and in addition, we want to explore the intrinsic mechanism of the distribution of the number of chronic diseases.

Acknowledgements

References

[1] Freedman, V.A. and Martin, L.G. (2000) Contribution of Chronic Conditions to Aggregate Changes in Old-Age Functioning. *American Journal of Public Health*, **90**, 1755-1760. https://doi.org/10.2105/AJPH.90.11.1755

[2] Thorpe, K.E. and Howard, D.H. (2006) The Rise in Spending among Medicare Beneficiaries: The Role of Chronic Disease Prevalence and Changes in Treatment Intensity. *Health Affairs*, **25**, 378-388. https://doi.org/10.1377/hlthaff.25.w378

[3] Wolff, J.L., Starfield, B. and Anderson, G. (2002) Prevalence, Expenditures and Complications of Multiple Chronic Conditions in the Elderly. *Archives of Internal Medicine*, **162**, 2269-2276. https://doi.org/10.1001/archinte.162.20.2269

[4] Vogeli, C., Shields, A.E., Lee, T.A., Gibson, T.B., Marder, W.D., Weiss, K.B. and Blumenthal, D. (2007) Multiple Chronic Conditions: Prevalence, Health Consequences, and Implications for Quality, Care Management and Costs. *JGIM*, **22**, 391-395. https://doi.org/10.1007/s11606-007-0322-1

[5] Vassilev, I., Rogers, A., Sanders, C., Cheraghi-Sohi, S., Blickem, C., Brooks, H., Kapadia, D., Reeves, D., Doran, T. and Kennedy, A. (2014) Social Status and Living with a Chronic Illness: An Exploration of Assessment and Meaning Attributed to Work and Employment. *Chronic Illness*, **10**, 273-290. https://doi.org/10.1177/1742395314521641

[6] Zhang, W.J. and Wei, M. (2015) Disability Level of the Chinese Elderly: Comparison from Multiple Data Sources. *Population Research*, **39**, 34-48.

[7] Li, S.X., Chen, C.X., Zhao, Y.N. and Ma, S.H. (2013) The Urban Elderly's in Common Chronic Diseases Life Quality Survey. *Modern Preventive Medicine*, **40**, 3435-3437.

[8] Shao, J.J. Li, D., Fan, L.J. and Shen, J.L. (2013) Chronic Mental Adaptation: Concepts, Models and Influencing Factors. *Journal of Southwest University* (*Social Sciences Edition*), **2**, 83-90.

[9] Wu, D.W., Fan, H. and Xiao, X.Y. (2005) Domestic Present Situation, Trend, Coping Strategies of Chronic Diseases. *Chinese Journal of Clinical Rehabilitation*, **9**, 126-128.

[10] Zhong, G.F. and Xiang, Y.Z. (2010) Chinese Hypertension Epidemiology and Its Influencing Factors. *Chinese Journal of Public Health*, **26**, 301-302.

[11] Hu, J.F., Luo, F.J. and Kong, H.N. (2013) Chronic Disease Health Management Research Progress. *Chronic Pathematology Journal*, **14**, 526-529.

[12] Visser, M. (2013) Zipf's Law, Power Laws, and Maximum Entropy. *New Journal of Physics*, **15**, 43021-43033. https://doi.org/10.1088/1367-2630/15/4/043021

[13] Newman, M.E.J. (2005) Power Laws, Pareto Distributions and Zipf's Law. *Contemporary Physics*, **46**, 323-351. https://doi.org/10.1080/00107510500052444

[14] Clauset, A., Shalizi, C.R. and Newman, M.E.J. (2009) Power-Law Distributions in Empirical Data. *SIAM Review*, **51**, 661-703. https://doi.org/10.1137/070710111