Population Distribution of Cities in Khyber Pakhtunkhwa and ZIPF`s Law

Kifayat Ullah[1], Nisar Ahmed[2], Daniyal Waseem[3] and Muhammad Ismail[4]

https://doi.org/10.62345/jads.2025.14.3.40

Abstract

In this study, we investigate Zipf's Law, along with Pareto and Power Law distributions, which are often observed in situations where a small number of events dominate while the majority occur less frequently, resulting in a right-skewed distribution. Human population distribution across cities frequently follows this pattern. Using data from the 2017 Population Census, we analyze the population distribution of cities in Khyber Pakhtunkhwa, Pakistan, to test the validity of Zipf's Law. The results reveal that the population distribution of cities in the province closely adheres to Zipf's Law. We attribute this conformity to the cohesive nature of the urban system in Khyber Pakhtunkhwa, where common cultural background, shared language, and consistent administrative structures contribute to the observed population hierarchy. These findings highlight the importance of Zipf's Law as a useful framework for understanding regional urbanization patterns. Moreover, the study provides a foundation for future research on city-size distributions and urban planning policies in Pakistan.

Keyword: Zipf's Law, Population Distribution, Urban System, City-Size Distribution.

Introduction

Zipf`s law was first introduced in quantitative linguistics. This law describes the distribution of words in a corpus. It is a famous law, named after George Kingsley Zipf (Zipf, 1949), who derived it. Zipf ranked the words based on their frequency of occurrence in a written text and studied the relationship between frequency and rank. Rank 1 is assigned to the word with the largest frequency, rank 2 is assigned to the word with the second largest frequency, and so on. Zipf found a connection between the frequency and rank of a word. He discovered that the frequency of a word has an inverse relationship with its rank i.e. $f_r \propto \frac{1}{r}$, where f_r is the frequency of the word with rank r. Thus, the generalized form of the equation is given as:

$$f_r \propto \frac{A}{r^{-\alpha}} \quad (1)$$

Where A and α are positive constants to be estimated from given data. Thus, this law states that, the occurrence of the most frequent word will be almost twice the occurrence of the second most frequent word, three times the occurrence of the third most frequent word and so on. Zipf's Law, along with Pareto and Power Law distributions, describes situations where a few events dominate and most occur infrequently, producing a right-skewed distribution. City population distribution

[1]*Department of Statistics, Quaid-i-Azam University, Islamabad. Email: kifayatshangra@gmail.com*
[2]*Department of Statistics, Quaid-i-Azam University, Islamabad. Email: nkkakar37@gmail.com*
[3]*Department of Statistics, Quaid-i-Azam University, Islamabad. Email: daniyalkhan99dani@gmail.com*
[4]*Department of Statistics, University of Peshawar. Email: ismailfromdir@gmail.com*

often follows this pattern. First noted by Felix Auerbach (1913), Zipf's Law has since been widely applied in urban studies, suggesting that the population of the largest city is roughly double that of the second, triple the third, and so forth. While alternative models such as primacy measures, q-exponential, and log-normal distributions exist, Zipf's Law remains the most widely used due to its simplicity and ability to capture the structure of entire urban systems. The literature, however, presents mixed evidence. Some researchers argue that Zipf's Law universally applies to city populations (Krugman, 1996; Ezzahid et al., 2015), while others find deviations or inconsistencies over time (Eechkout, 2004; Soo, 2007; Arshad et al., 2018). This ongoing debate motivates further empirical investigation. Previous studies have tested Zipf's Law across multiple countries and provinces, including Pakistan, where results showed weak adherence at the national level but stronger conformity within provinces due to cultural and administrative cohesion. The analyze the urban population distribution of Khyber Pakhtunkhwa (KP), Pakistan, using the 2017 census. By applying the rank-population model to all urban areas in the province, we assess whether KP's urban system follows Zipf's Law and explore the factors underlying the observed patterns. Auerbach (1913), Singer (1936) and Zipf (1949) identified that the cities population distribution could be expressed as a pareto distribution.

$$R = AP^{-\alpha} \quad (2)$$

OR $\log(R) = \log(A) - \alpha \log(P)$ (3)

Zipf's Law relates city population P and rank R through a power-law distribution, where A is the population of the largest city and α is the Zipf exponent. If α=1, the distribution strictly follows Zipf's Law, producing a straight line with slope −1 in a log-log plot of rank versus population. The value of α reflects the hierarchy of the urban system. Traditionally, α is estimated using OLS regression of log (R) on log log(P). However, OLS introduces downward bias and underestimates standard errors due to serial correlation from ranking. To address this, Gabaix and Ibragimov (2011) proposed a simple correction by replacing log(R) with log(R−1/2), improving the reliability of Zipf exponent estimation.

$$\log\left(R - \frac{1}{2}\right) = \log(A) - \alpha \log(P) \quad (4)$$

In this thesis, we use the recommendations of Gabaix and Ibragimov to find the value of α from equation (1.3) by using $\log(R - \frac{1}{2})$ as response variable and $\log(P)$ as explanatory variable. Then, we apply the t-test to determine if the predicted value of α is equal to 1 i.e. $H_o: \alpha = 1$ in equation (3). If the calculated value of t-test fall in the acceptance region then we will accept our null hypothesis i.e. we will accept that the estimated value of α is equal to one and conclude that the population distribution of cities in Khyber Pakhtunkhwa follows Zipf`s law otherwise, it will not follow. The normalcy assumption is commonly used for the standard error computed from equation (4) when using the t-test, but this can be limited in some conditions. The variable which is corrected through taking its log produces a distribution which is close to normality. For large samples (N > 20) normality has little effects on results (E. Uriel, 2013).

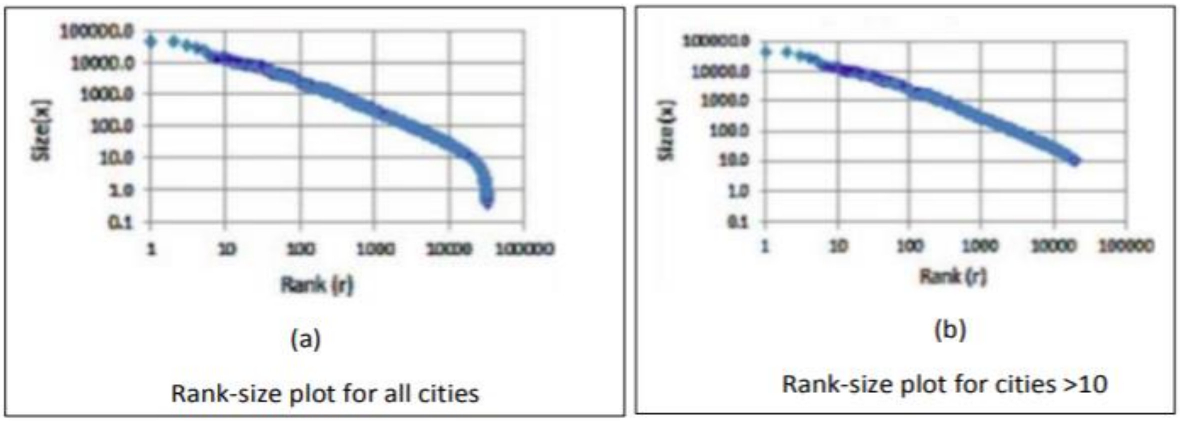
Literature Review

Zipf's Law, initially introduced by George Zipf in linguistics and first observed in city population distribution by Felix Auerbach (1913), has since been widely applied in urban studies. Numerous researchers have found that it often provides a good fit for city-size distributions across regions and time periods.

Jiang et al. (2015) conducted a comprehensive study on Zipf's Law for natural cities worldwide. Using a methodology to identify natural cities across five continents and over 230 countries, they tested the law at global, continental, and national levels. Their results showed that Zipf's Law holds strongly at the global scale, particularly for cities larger than 10 km², with rank-population plots confirming a good fit. At the continental level, Zipf's Law generally holds, though the Zipf exponent varies between 1.9 and 2.1. However, at the national level, the law is less consistent and not always valid across all time periods.

The study concluded that while Zipf's Law demonstrates universality at broader scales, its accuracy weakens at smaller scales and over time, highlighting the need for careful contextual analysis when applying it to urban systems.

Figure 1: Rank-population plots for all natural cities and for cities > 10 km^2



Jiang et al. (2015) observed that Zipf's exponent for countries is approximately 1 but not exact, with the law applying only in certain times and places. Their findings suggest that Zipf's law is more applicable to aggregated city systems rather than random sets of cities, and it also holds in the distribution of city numbers across regions.

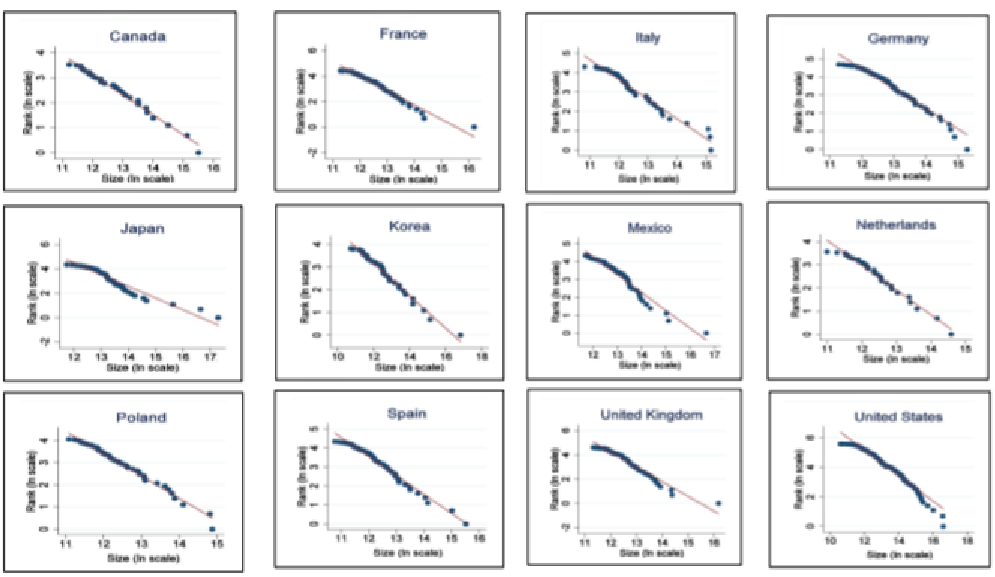
Similarly, Veneri (2013) examined OECD countries, analyzing city populations using two methods: the simple rank-population rule and the Gabix–Ibragimov (2011) rank-½ method. Results showed that the estimated exponent (α) was consistently close to 1. However, the hypothesis $\alpha = 1$ was accepted in only 3 out of 13 cases using the simple method, while the Gabix–Ibragimov method confirmed the hypothesis for all countries. High R^2 values and Zipf plots further demonstrated that Zipf's law reliably describes urban population distributions across OECD nations

Figure 2: Results of OLS regresion for equation (1) and equation (2)

	rank-size			(rank-1/2)-size		
	α coeff.			α coeff.		
Country	Eq (2)	R. sq	t-test α=-1	Eq (3)	R. sq	t-test α=-1
Canada	-0.798	0.99	169.12***	-0.887	0.97	0.541
Chile	-0.832	0.94	14.55***	-0.962	0.96	0.144
France	-1.128	0.97	33.47***	-1.209	0.97	1.114
Germany	-1.106	0.96	26.59***	-1.161	0.95	1.024
Italy	-1.019	0.97	0.71	-1.092	0.96	0.511
Japan	-0.957	0.93	1.9	-1.035	0.94	0.210
Korea	-0.715	0.99	467.37***	-0.788	0.99	1.273
Mexico	-0.999	0.95	0	-1.072	0.95	0.418
Netherlands	-1.067	0.98	5.46**	-1.187	0.96	0.660
Poland	-1.008	0.99	0.26	-1.087	0.97	0.433
Spain	-0.997	0.98	0.03	-1.068	0.97	0.391
United Kingdom	-1.215	0.96	73.66***	-1.292	0.97	1.606
United States	-0.864	0.97	183.6***	-0.888	0.96	1.449

Source: Author's elaborations on OECD data.

Figure 3: Zipf`s plots for OECD countries



That paper also provides Zipf`s law for a combination of countries. particularly, OECD countries are combined by continents and by the overall urban system of the OECD in order to see whether the Zipf`s fits the data.

Figure 4: Zipf`s plot for Asia, Europe, America and OECD

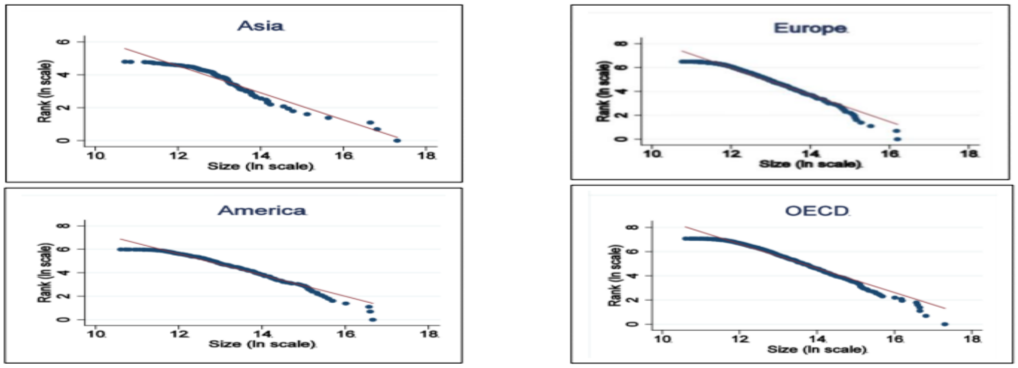


Figure (4-5) shows us the Zipf`s plot for Asia, Europe, America and all OECD countries. Included Asia (Japan and Korea), all graphs explain almost completely linear relation in the majority of the distribution. The predicted values are is given in figure, which shows us that, despite the countries being grouped by continent, Zipf`s law fits well. The coefficients for the city-size ranges from 0.82 to 1.13 for Asia and Europe respectively. Surprisingly, Zipf`s law is best fits for city population at the OECD level i.e. the value of $\alpha$ is equal to 1. On the use of equation (4), the Zipf`s law is accepted at America, Asia and OECD besides Europe, while on the use of equation (3) the Zipf`s law is accepted only at the level of OECD, as shown in figure.

Figure 5: Results for equation (3) and equation (4) by continent

	rank-size			(rank-1/2)-size		
	α coeff. Eq(2)	R. sq	t-test α=-1	α coeff. Eq(3)	R. sq	t-test α=-1
America	-0.903	0.97	129.98***	-0.921	0.96	1.21
Asia	-0.820	0.93	75.06***	-0.864	0.93	1.22
Europe	-1.127	0.97	274.63***	-1.143	0.97	2.27**
OECD (29)	-1.005	0.97	1.07	-1.015	0.97	0.35

Source: Author's elaborations on OECD data.

Arshad et al. (2019) studied the Zipf`s law for the cities population distribution of Pakistan utilizing the data from 1951 to 1998 censuses. They examined that Zipf`s law violates at national level in all of five censuses. Next, they applied Zipf`s law at provincial level and observed that Zipf`s law holds for the cities population distribution of cities at provincial level. They termed these findings as a cohesion quality of the urban system because in the provinces the urban system is more cohered than national level in Pakistan. For empirical analysis, they first estimated Zipf's exponent at the federal level by taking the urban population of the whole country to explore whether Zipf's law holds at the federal level. After that they estimated α for provinces to determine the recognition of the Zipf`s law at the provincial level. Later they drew the Zipf`s curve for visual depiction. The points on curve confirms a line with slope of -1 that show us an exact shape of Zipf`s law. First, they drew a single Zipf`s plot for all censuses at country level to examine the general pattern of the distribution as displayed in figure (6-7). For 1951, the Zipf`s curve show us a discontinuous and evidential deviations from linear line. Similarly, for 1998, the Zipf`s curve show us that urban structure approaches to a further linear distribution but it is not yet perfectly linear with a little difference.

Figure (6-7) shows us a separate fitted line for each of the five census years. In all five census years the actual data points have deviated from the fitted line. This plot leads us to the conclusion that the exact shape of the Zipf`s law for Pakistani cities is not being confirmed.

Figure 6: Zipf's plot for all cities

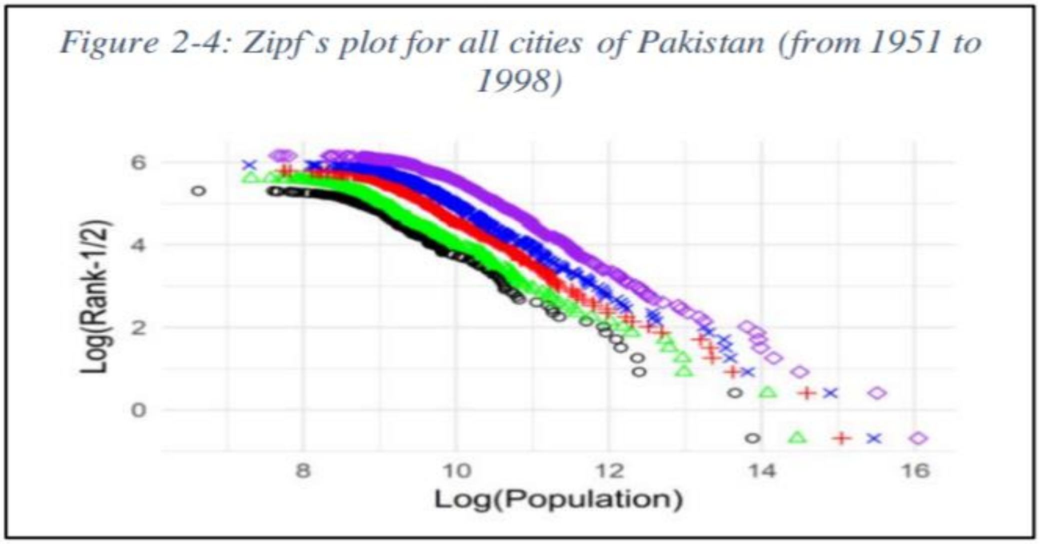
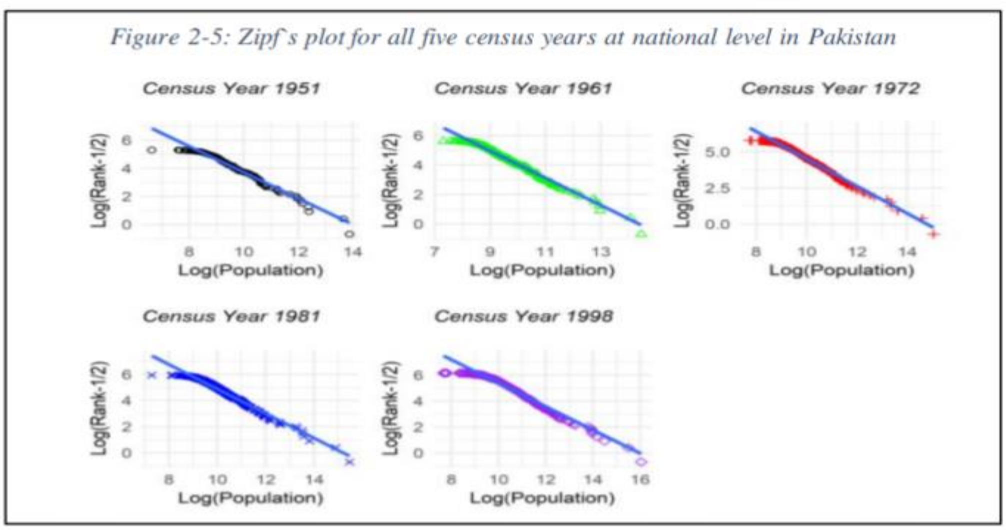


Figure 7: Zips plot for all five census



Next, they estimate the value of α using the equation (4) which is given in figure (7). The predicted values of α for all censuses are smaller than 1, indicating that the cities population distribution is more unbalanced. The predicted value of α has increased from 0.91 to 0.94 and then decreased again to 0.91. The hypothesis that $H_0: \alpha = 1$ is rejected for all censuses. Therefore, they conclude that the cities population distribution defies Zipf`s law in Pakistan at federal level.

Figure 8: Equation results

Table 2-3: Results of equation (1.4) at country level using the population of all cities of Pakistan for all five census years

Census Year	Number of cities	Log (A)	α	t-test $\alpha=1$	P-value of t-test
1951	205	12.988	0.927	Rejected	0.000
1961	276	13.236	0.919	Rejected	0.000
1972	332	13.968	0.944	Rejected	0.000
1981	383	14.241	0.936	Rejected	0.000
1998	474	14.382	0.900	Rejected	0.000

Next, they have checked the population distribution of cities in provinces. They have find the estimated value of α for provinces for all censuses. The predicted value of α for every province in all censuses is approximately equivalent to 1. Therefore, the Zipf`s law holds at provincial level. Figure below show us curves and addresses the results of equation (4) for the Punjab province of Pakistan. Figure (9-10) displays that Zipf`s curves for Punjab are strongly linear then the curves for the entire country in Figure above. Furthermore, the predicted values of α for Punjab are approximately 1 in 1961, 1972 and 1981 as shown in figure (9-10).

Figure 9: Zipf's plot for all census

Figure 2-6: Zipf`s plots for all five census years for Punjab province

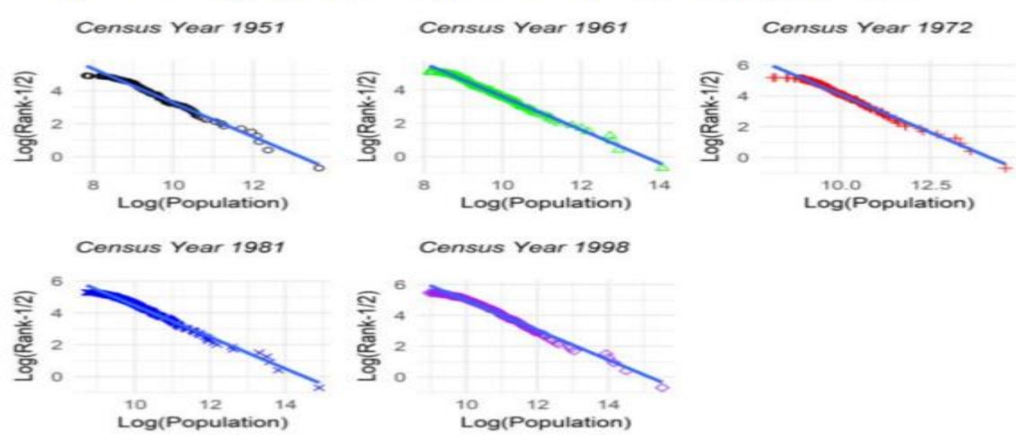


Figure (10) displays the Zipf`s curves and figures below addresses the results of equation (4). The estimated value of α for Sindh does not equivalent to 1 in any of five censuses and the distribution of cities is uneven as shown in figure 10-11. The reasons behind this problem is the city of Karachi. By dropping the city of Karachi from sample, the population distribution of cities in Sindh is more probably to follow Zipf`s law.

Figure 10: Zipfs plot for Sindh

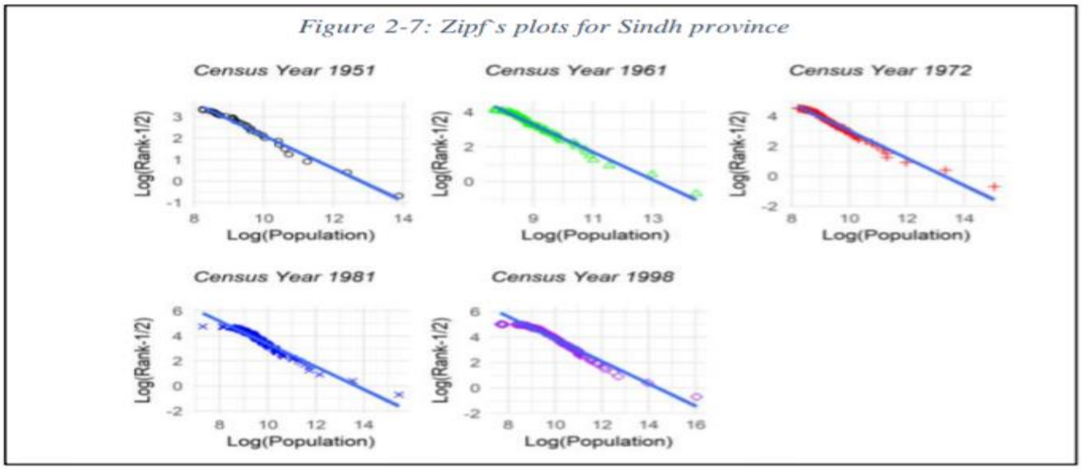


Figure 11: Results for Sindh

Table 2-5: Results of equation (1.4) for Sindh province

	Include Karachi					*Exclude Karachi*				
Census Year	Number of cities	*Log (A)*	*α*	t-test *α=1*	P-value of t-test	Number of cities	*Log (A)*	*α*	t-test *α=1*	P-value of t-test
1951	29	9.738	0.762	Rejected	0.000	28	11.632	0.977	Do not reject	0.456
1961	64	10.438	0.795	Rejected	0.000	63	11.718	0.945	Rejected	0.001
1972	94	12.118	0.910	Rejected	0.000	93	13.770	1.093	Rejected	0.000
1981	118	12.450	0.909	Rejected	0.000	117	13.837	1.061	Rejected	0.004

Figure 12 and 13 below displays the Zipf`s plots addresses the results of equation (4) for Khyber Pakhtunkhwa. As given in figure, the predicted values of $\alpha$ are equivalent to 1 for Khyber Pakhtunkhwa (KP). These findings conclude that the population distribution of cities in KP accepts Zipf`s law in all censuses.

Figure 12: Zipf's plot for KPK

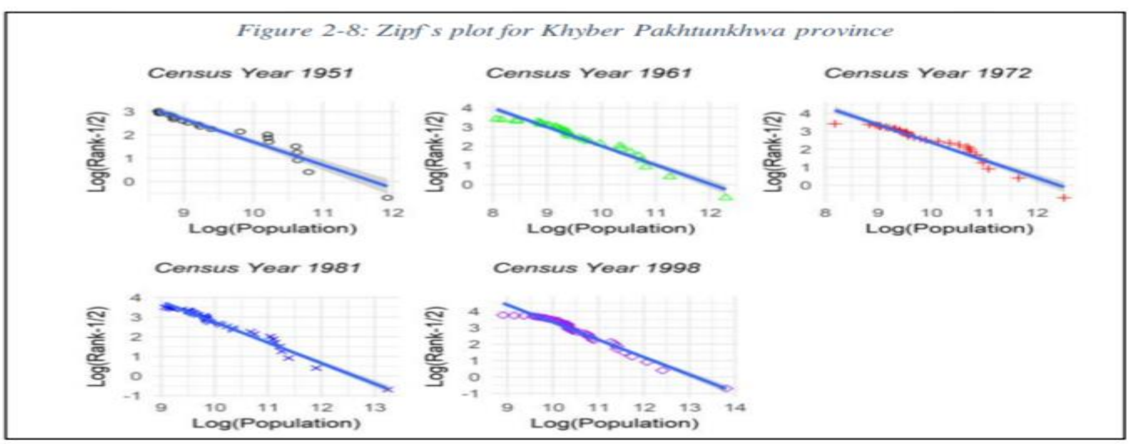
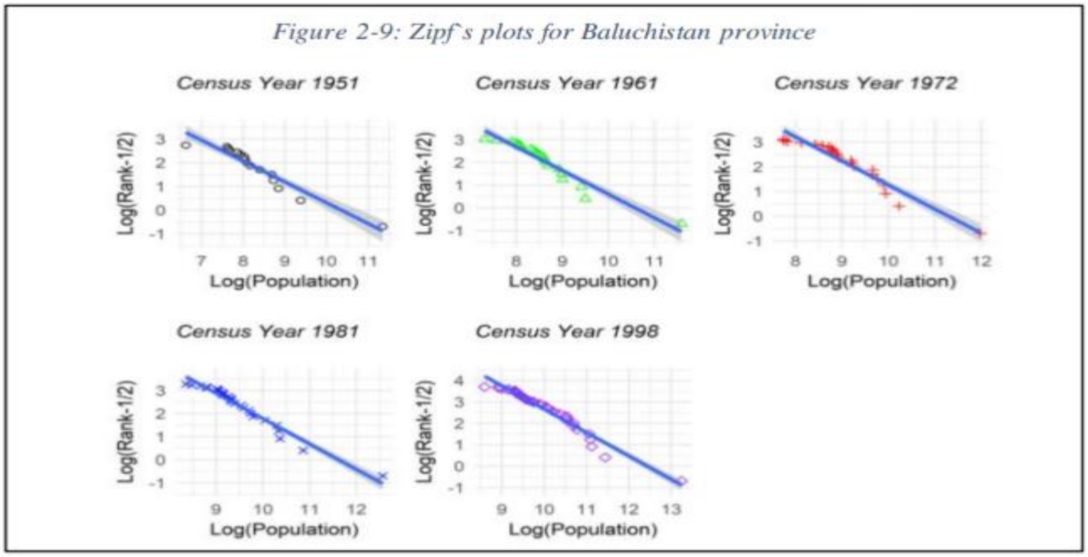


Figure 13: Results for KPK



In this chapter, we applied Zipf`s law for the cities population distribution of KP province of Pakistan using the data of 2017 census year. We have fitted the rank-population model to the cities of Khyber Pakhtunkhwa. Based on t-test, we have proved that the cities population distribution in Khyber Pakhtunkhwa follows Zipf`s law.

Khyber Pakhtunkhwa

Khyber Pakhtunkhwa (abbreviated as KP) is a province of Pakistan. It is placed in the northwest area of Pakistan, with a total area of 101,741 km^2. The capital city of Khyber Pakhtunkhwa is Peshawar. It is the third biggest province of Pakistan by population. It is also the third biggest province of Pakistan by its economy but in terms of area, it is the lowest province in the country. According to 2017 census, the total population of Khyber Pakhtunkhwa is 35.53 million, in which 5,687,637 and 24,821,283 is the urban and rural population respectively. It is host to 17.9% of Pakistan's overall population, with the majority of ethnic Pashtuns and Hindko speakers. In addition, the province is home to 1.5 million Afghan refugees. Pashto is the most widely spoken language in the province, with 78.89 percent of the population speaking it. Hindko (9.87 percent) and Saraiki are two more languages having considerable speakers (3.21 percent). The vast majority of Khyber Pakhtunkhwa citizens follow and believe in Sunni Islam, although a small number of Shia Muslims live among the Isma'ilis in the Chitral district. There are 7 divisions, 35 districts and 46 cities in Khyber Pakhtunkhwa, according to 2017 census. Figure (3.1) shows the map of Khyber Pakhtunkhwa and its location in Pakistan.

Figure 14: Map of Khyber Pakhtunkhwa and its location in Pakistan

Definition of a City in Zipf`s law

The definition of a city plays an important role in Zipf`s law and cities population distribution. The definition of a city has an impact on Zipf's law's application, according to the recent literature. In general, literature has established three different basic sorts of city definitions: cities with administrative boundaries, cities that are functionally defined and natural cities. In this thesis, we draw our analysis on the administrative definition of cities used in census, which states that if a settlement has regional administrative authority, it is designated an urban area.

Data Collection

The information about population of cities in Khyber Pakhtunkhwa is gathered from the website of Pakistan Bureau of statistics (PBS) [1]. Metropolitan corporations, municipal corporations, municipal committees, town committees and cantonments are the four levels of administrative units used by PBS to measure urban localities. Table (1) shows the total population, rural and urban population and percentage of urban population for Khyber Pakhtunkhwa in 2017 census. Similarly, Table (2) reports the total population as well as total number of districts and total number of cities in each division of Khyber Pakhtunkhwa in 2017 census year.

Table 1: Total, Urban and Rural population of Khyber Pakhtunkhwa in 2017 census year

Census year	Total population	Urban population	Rural population	Percentage of urban population
2017	30,508,920	5,687,637	24,821,283	18.64

Table 2: Divisions, Total population, number of districts and number of cities in each division in 2017 census years

Census years	Divisions	Total population	Number of Districts	Number of Cities
2017	Peshawar	8,922,465	5	13
	Mardan	3,997,677	2	6
	Malakand	8,608,378	9	10
	Kohat	3,211,458	5	5
	Hazara	5,325,121	8	6
	Dera Ismail Khan	2,803,147	3	4
	Bannu	2,656,801	3	2

Data Analysis and Results

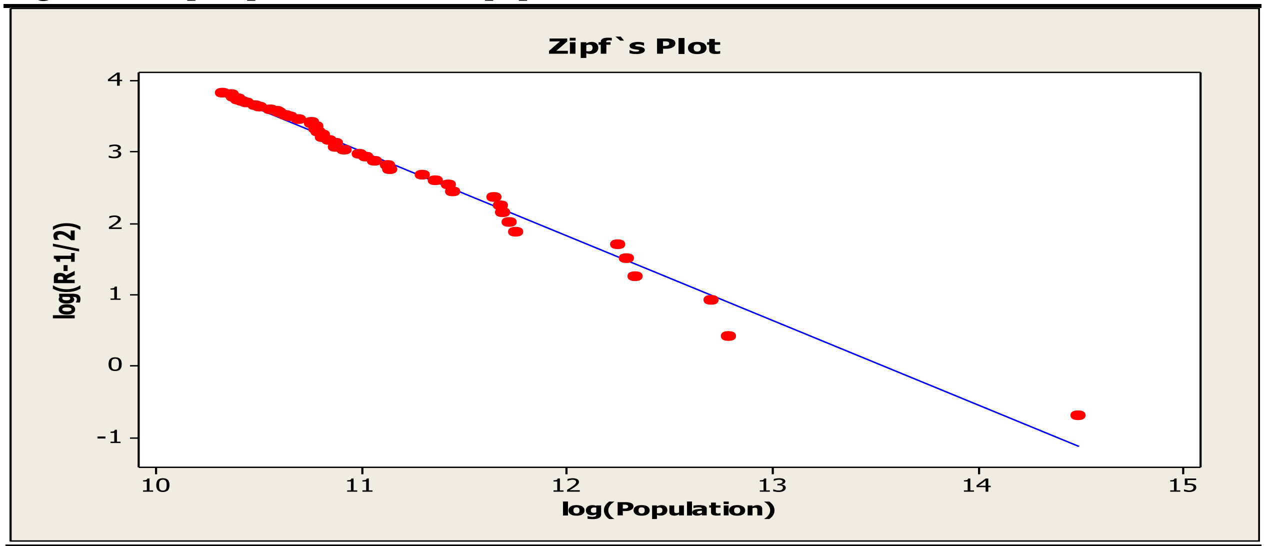
This section comprises empirical evidence. In order to do empirical investigation, we estimate the value of ‘α’ for cities of Khyber Pakhtunkhwa. We use OLS method to estimate the value of ‘α’ in equation (4). As we know that the equation (4) is a regression model with $\log(R-\frac{1}{2})$ as a dependent variable and $\log(Population)$ as an independent variable and in which ‘α’ plays the role of slope coefficient and $\log(A)$ is a constant. We also know that if the estimated value of slope coefficient i.e. ‘α’ in this given regression model is statistically equal to 1, then we say that we are under the Zipf`s law. For this purpose, we apply student t-tes to find out that the predicted value of ‘α’ in the given regression equation (i.e. in equation (1.4)) is equal to 1, i.e. $H_o: \alpha = 1$. If null hypothesis is accepted by t-test at 95% confidence, then we conclude from this that we are under the Zipf`s law and the cities population distribution in KP follows Zipf`s law. Later we draw a scatter plot (Usually called Zipf`s plot or Zipf`s curve), in which we take $\log(Rank-\frac{1}{2})$ on y-axis and $\log(Population)$ on x-axis, for visual representation of our results. To create the perfect form of Zipf's law, the dots on the curve must fit to a line with a slope of -1. For the cities population distribution of Khyber Pakhtunkhwa first we estimated the value of ‘α’ by OLS method using data of 2017 census year. Then we have applied the t-test and identified the critical points for it and also computed the p-value for t-test as they all are given in Table (3.1).

Table 3: Estimated value of α, log(A), t-test computed and tabulated values and p-value

Estimated value of α	$\log(A)$	Computed value of t-test for $H_o: \alpha = 1$	Tabulated value of t-test for $H_o: \alpha = 1$	P-value of t-test
1.19	14.4905	0.63	2.015	0.531953

The computed value of t-test is less than its tabulated value as given in Table (3.3) which shows that it does not fall in critical region i.e. it falls in acceptance region. Therefore, we decide that the null hypothesis is accepted. It is also evident from p-value, because a p-value greater than 0.05 lead to acceptance of null hypothesis. We conclude from these facts that the population pattern of cities in Khyber Pakhtunkhwa follows Zipf`s law. Next, we have drawn a scatter plot which is usually called Zipf`s curve or Zipf`s plot which is given in Figure (3.2). This plot shows a visual representation of our results. In this plot we have taken the dependent variable ($\log(Rank - \frac{1}{2})$) on y-axis and the independent ($\log(Population)$) on x-axis. When we have joined the points on it by a line, then we get a linear line whose slope is close to -1. Therefore, it clearly shows that the population distribution of cities in Khyber Pakhtunkhwa follows Zipf`s law.

Figure 15: Zipf`s plot for the cities population distribution of KP in 2017 census



Since empirical evidence show us that the cities population distribution of KP accepts Zipf's law for 2017 census. Now we come to the part where we talk about the reason behind this. However, the majority of the population in Khyber Pakhtunkhwa consists of pathans. Therefore culture, language and rules are common here because of this there is a coherent urban system in Khyber Pakhtunkhwa. In term of cohesion quality of urban system, the urban population here is in a pattern due to which the Zipf`s law well holds for popualtion distribution of cities in Khyber Pakhtunkhwa.

Conclusion

In this paper, we comprehensively investigate the validity of Zipf's law for the population distribution of cities in Khyber Pakhtunkhwa using the official 2017 census data. The estimated value of α for the city populations was calculated, and the t-test confirmed the hypothesis $\boldsymbol{\alpha = 1}$ at the 95% confidence level, indicating strong statistical support for Zipf's law. This finding was further reinforced through the Zipf plot on a logarithmic scale, where the data points closely aligned with a linear straight line, exhibiting a slope approximately equal to **–1**. Taken together, these results provide compelling evidence that Zipf's law holds remarkably well for the city population distribution of Khyber Pakhtunkhwa. We attribute this consistent pattern to the cohesion and stability of the urban system, which is shaped and reinforced by shared cultural values, a common language, and unified institutional and regulatory frameworks across the province.

References

- Anderson, G., & Ge, Y. (2005). The size distribution of Chinese cities. *Regional Science and Urban Economics*, *35*(6), 756-776.
- Arshad, S., Hu, S., & Ashraf, B. N. (2018). Zipf's law and city size distribution: A survey of the literature and future research agenda. *Physica A: Statistical mechanics and its applications*, *492*, 75-92.
- Arshad, S., Hu, S., & Ashraf, B. N. (2019). Zipf's law, the coherence of the urban system and city size distribution: Evidence from Pakistan. *Physica A: Statistical Mechanics and its Applications*, *513*, 87-103.
- Auerbach, F. (1913). Das gesetz der bevölkerungskonzentration. *Petermanns Geographische Mitteilungen*, *59*, 74-76.
- Bajracharya, P., & Sultana, S. (2020). Rank-size Distribution of Cities and Municipalities in Bangladesh. *Sustainability*, *12*(11), 4643.
- Bee, M., Riccaboni, M., & Schiavo, S. (2013). The size distribution of US cities: Not Pareto, even in the tail. *Economics Letters*, *120*(2), 232-237.
- Berry, B. J., & Okulicz-Kozaryn, A. (2012). The city size distribution debate: Resolution for US urban regions and megalopolitan areas. *Cities*, *29*, S17-S23.
- Dimitrova, Z. I., & Ausloos, M. (2015). Primacy analysis in the system of Bulgarian cities. *Open Physics*, *13*(1).
- Eaton, J., & Eckstein, Z. (1997). Cities and growth: Theory and evidence from France and Japan. *Regional science and urban Economics*, *27*(4-5), 443-474.
- Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, *94*(5), 1429-1451.
- Elazzouzi-Hafraoui, S., Nishiyama, Y., Putaux, J. L., Heux, L., Dubreuil, F., & Rochas, C. (2008). The shape and size distribution of crystalline nanoparticles prepared by acid hydrolysis of native cellulose. *Biomacromolecules*, *9*(1), 57-65.
- Ezzahid, E., & ElHamdani, O. (2015). Zipf's Law in the Case of Moroccan Cities. *Review of Urban & Regional Development Studies*, *27*(2), 118-133.
- Ezzahid, E., & Maouhoub, B9. (2015). Capital account liberalization and Moroccan macroeconomic performances.
- Gabaix, X., & Ibragimov, R. (2011): A Simple Way to Improve the OLS Estimation of Tail Exponents. *Journal of Business & Economic Statistics*, *29*(1), 24-39.
- Gabaix, X., & Ioannides, Y. M. (2004). The evolution of city size distributions. *In Handbook of regional and urban economics* (Vol. 4, pp. 2341-2378). Elsevier.
- Ghosh, A., & Basu, B. (2019). Universal City-size distributions through rank ordering. *Physica A: Statistical Mechanics and its Applications*, *528*, 121094.
- Giesen, K., & Südekum, J. (2011). Zipf's law for cities in the regions and the country. *Journal of economic geography*, *11*(4), 667-686.
- Henderson, V. (2002). Urbanization in developing countries. *The World Bank Research Observer*, *17*(1), 89-112.Jefferson, M. (1939). Why geography? The law of the primate city. *Geographical review*, *79*(2), 226-232.
- Jiang, B., Yin, J., & Liu, Q. (2015). Zipf's law for all the natural cities around the world. *International Journal of Geographical Information Science*, *29*(3), 498-522.
- Konishi, Y., & Nishiyama, Y. (2009). Hypothesis testing in rank-size rule regression. *Mathematics and Computers in Simulation*, *79*(9), 2869-2878.

- Krugman, P. (1996). Confronting the mystery of urban hierarchy. *Journal of the Japanese and International economies*, *10*(4), 399-418.
- Kundak, S., & Dökmeci, V. (2018). A Rank-Size Rule Analysis of The City System at The Country and Province Level in Turkey. *ICONARP International Journal of Architecture and Planning*, *6*(1), 77-98.
- Lalanne, A. (2014). Zipf's law and Canadian urban growth. *Urban Studies*, *51*(8), 1725-1740.
- Luckstead, J., & Devadoss, S. (2014). A comparison of city size distributions for China and India from 1950 to 2010. *Economics Letters*, *124*(2), 290-295.
- Luckstead, J., Devadoss, S., & Danforth, D. (2017). The size distributions of all Indian cities. *Physica A: Statistical Mechanics and its Applications*, *474*, 237-249.
- Malacarne, L. C., Mendes, R. S., & Lenzi, E. K. (2001). q-Exponential distribution in urban agglomeration. *Physical Review E*, *65*(1), 017106.
- Margonari, C., Freitas, C. R., Ribeiro, R. C., Moura, A. C. M., Timbó, M., Gripp, A. H., & Dias, E. S. (2006). Epidemiology of visceral leishmaniasis through spatial analysis, in Belo Horizonte municipality, state of Minas Gerais, Brazil. *Memórias do Instituto Oswaldo Cruz*, *101*(1), 31-38.
- Matlaba, V. J., Holmes, M. J., McCann, P., & Poot, J. (2013, November). A Century of the Evolution of the Urban System in Brazil. In *Review of Urban & Regional Development Studies: Journal of the Applied Regional Science Conference* (Vol. 25, No. 3, pp. 129-151).
- Peng, J. F., Hu, M., Wang, Z. B., Huang, X. F., Kumar, P., Wu, Z. J., & He, L. Y. (2014). Submicron aerosols at thirteen diversified sites in China: size distribution, new particle formation and corresponding contribution to cloud condensation nuclei production. *Atmospheric Chemistry and Physics*, *14*(18), 10249-10265.
- Pérez-Campuzano, E., Guzmán-Vargas, L., & Angulo-Brown, F. (2015). Distributions of city sizes in Mexico during the 20th century. *Chaos, Solitons & Fractals*, *73*, 64-70.
- Pumain, D., Swerts, E., Cottineau, C., Vacchiani-Marcuzzo, C., Ignazzi, C. A., Bretagnolle, A., & Baffi, S. (2015). Multilevel comparison of large urban systems. *Cybergeo: European Journal of Geography*.
- Singer, H. W. (1936). The" courbe des populations." A parallel to Pareto's Law. *The Economic Journal*, *46*(182), 254-263.
- Soo, K. T. (2007). Zipf's Law and urban growth in Malaysia. *Urban Studies*, *44*(1), 1-14.
- Soo, K. T. (2014). Zipf, G ibrat and geography: Evidence from C hina, I ndia and B razil. *Papers in Regional Science*, *93*(1), 159-181.
- Uriel, E. (2013). Hypothesis testing in the multiple regression model. *Universidad de Valencia: Department of economics*.
- Veneri, P. (2013). On city size distribution: evidence from OECD functional urban areas.
- Zipf, G. K. (1949). Human behavior and the principle of least effort: an introd. to human ecology.