Uludağ Üniversitesi
İktisadi ve İdari Bilimler Fakültesi Dergisi
Uludağ Journal of Economy and Society
Cilt/Vol. XXXIII, Sayı/No. 1, 2014, pp. 119-136

TESTING THE PROBABILITY DISTRIBUTION MODELS FOR THE PATIENTS' LENGTHS OF STAY IN HOSPITAL

Necmi GÜRSAKAL [1]
Selim TÜZÜNTÜRK [2]
Zehra Berna AYDIN [3]

Abstract

The Probability Distribution of Lengths of Stay in Hospital is a matter of debate. Weibull, Gamma and Lognormal distributions are commonly used distributions for the Lengths of Stay in hospital data in the related Literature. In recent years, researchers found evidence that Power Law Probability Distribution fits well to this data.

This study focused on the investigation of whether the distribution of the data follows Power Law, Weibull, Gamma and Lognormal or not. For this purpose, a sample of a Turkish Hospital data was used and tested. Results show that the data follows a Lognormal Probability Distribution for the Turkish Case.

Key Words: *Probability Distributions, Power Law Probability Distribution, Lengths of Stay in Hospital.*

Hastanede Yatma Süreleri İçin Olasılık Dağılımı Modellerinin Test Edilmesi

Özet

Hastanede Yatma Sürelerinin Olasılık Dağılımı tartışma konusudur. Literatürde, hastanede yatma süreleri verileri için Weibull, Gamma ve Lognormal dağılımları çoğunlukla kullanılan dağılımlardır. Son yıllarda, araştırmacılar Kuvvet Yasası Olasılık Dağılımının bu veriye iyi uyduğuna yönelik kanıtlar bulmuştur.

---

[1] Prof. Dr. Necmi GÜRSAKAL; Uludağ Üniversitesi, İ.İ.B.F., Ekonometri Bölümü.
[2] Yrd. Doç. Dr. Selim Tüzüntürk; Uludağ Üniversitesi, İ.İ.B.F., Ekonometri Bölümü.
[3] Öğr. Gör. Dr. Zehra Berna Aydın; Uludağ Üniversitesi, İ.İ.B.F., Ekonometri Bölümü.

Bu çalışma, verilerin Kuvvet Yasası, Weibull, Gamma ve Lognormal dağılımına sahip olup olmadığının araştırılmasına odaklanmıştır. Bu amaçla, bir Türk hastanesi örneklemi kullanılmış ve test edilmiştir. Sonuçlar, Türkiye örneği için verilerin Lognormal Olasılık Dağılımına sahip olduğunu göstermektedir.

Anahtar Kelimeler: Olasılık Dağılımları, Kuvvet Yasası Olasılık Dağılımı, Hastanede Yatma Süresi.

1. LITERATURE

The length of stay (LOS) in hospital is defined as the number of days of a patient's stay in hospital to get treatment in a certain period (Esatoğlu and Bozat (2002) and number of days a patient remains in the hospital from admission to discharge (Chassin, 1983). This concept was examined from different perspectives in academic studies. These studies generally focused on management of hospital care, quality control, appropriateness of hospital use and hospital planning (Marazzi, Paccaud and Ruffieux, 1998), health planning and formation of payment policy (Xiao, Lee and Vemuri (1999), and performance indicator for an acute care hospital (Keefler, Duder and Lechman (2001). Especially, LOS is a basic decision making factor (Faddy, Graves and Pettitt, 2009) for evaluating the quality and effectiveness of the medical care in hospital (Esatoğlu and Bozat, 2002), for measuring the hospital efficiency (Rafiei, Ayatollahi and Behboodian, 2007), for resource allocation process in the hospital (Hellervik and Rodgers, 2007). Lim and Tongkumchum (2009) expressed that the length of hospital stay is a common parameter used to indicate health resource utilization, health care cost and severity of disease. Lee and others (2010) declared that prolonged inpatient care, acute care and rehabilitation therapy are the main causes of hospital expenditure inflation. Bai and others (2014) considered the LOS in hospital as a reasonable estimate of resource use inpatient care as well as treatment outcomes.

It is well known that the empirical distribution of LOS is positively skewed (Xiao, Lee and Vemuri, 1999). The best way to model LOS is to use logarithmic transformation of the variable and use it with ordinary least square (OLS) regression (Faddy, Graves and Pettitt, 2009). The lying factor behind the logarithmic transformation is to attain normality so that multiple regression and associated tests can be applied (Xiao, Lee and Vemuri, 1999). Because of the empirical distribution of LOS is very asymmetrical with a broad tail, it makes many statistical estimates less robust (Hellervik and Rodgers, 2007). The determination of the distribution of LOS comes into prominence within this scope. Marazzi, Paccaud and Ruffieux (1998) represent the adequacy of three commonly used models for describing the distributions of LOS. These are: Weibull, Gamma and Lognormal probability distributions. Besides these widely used models, Hellervik and Rodgers (2007) showed that the distribution of LOS in hospitals is found to be well approximated by power law distribution.

In the context above, the aim of this paper is to investigate whether the distribution of the Lengths of Stay in Hospital data (a sample of a Turkish Hospital in Edirne) follows a Power Law, Weibull, Gamma or Lognormal probability distribution or not. To do so, in Section 2 and Section 3 the probability distributions that are mentioned above are explained theoretically. Research process and results were given Section 4. In this section, our approach will be to test the recent evidence that the Lengths of Stay data follows a Power Law probability distribution at first. Then, we will test whether the data follows one of the commonly used distributions. And the final section covers the summary and conclusion.

2. POWER LAW PROBABILITY DISTRIBUTION

Mathematically a random variable x is said to have a Power Law Distribution if

$$p(x) = C\,x^{-\alpha} \quad (1)$$

for constants $C > 0$ and $\alpha > 0$ (Mitzenmacher, 2003). α is known as the exponent or scaling parameter (Clauset, Shalizi and Newman, 2009: 2).

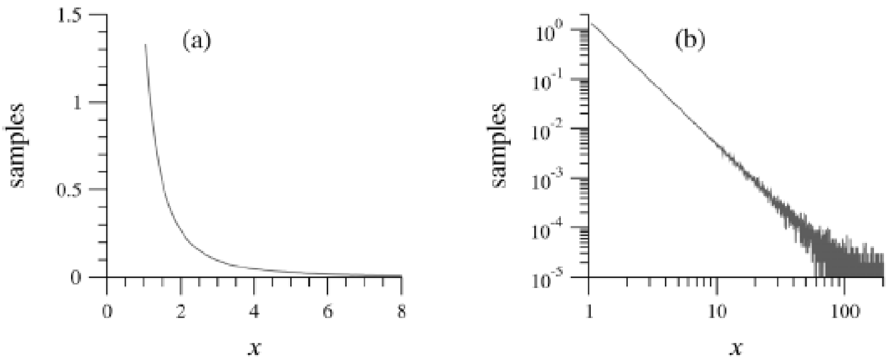


Figure 1.
Visual View of the Power Law Distribution (Newman, 2005: 326)

Figure 1(a) presents the histogram of the set of random numbers which have a power-law distribution with exponent $\alpha = 2.5$ and Figure 1(b) shows the same histogram on logarithmic scales. Most power-law distributions occurring in the nature have $2 \leq \alpha \leq 3$ (Newman, 2005: 326).

There are occasional exceptions that the scaling parameter ranges between $1.5 \leq \alpha \leq 3$.

There are two main ways of the determination of whether the empirical data follow Power Law distribution or not. These are graphical and formal methods.

2.1. Graphical Method

In graphical method, a histogram and the same histogram on logarithmic scales of the empirical data is drawn. Then, the shapes of these two graphs are evaluated with naked eye. A histogram of a quantity with a power-law distribution appears as is seen in Figure 1(a) and appears as is seen in Figure 1(b) as a straight line when plotted on logarithmic scales.

2.2. Formal Methods

In formal methods, there are a variety of ways to detect whether the empirical data follow a power-law distribution or not. These are value of scaling parameter, Transformation Method and Kolmogorov Smirnov (KS) Test.

2.2.1. Value of Scaling Parameter

One way is to take logarithms of both sides of the equation 1 as below:

$$\log p(x) = \log C - \alpha \log x \tag{2}$$

The parameters of this equation are estimated with Ordinary Least Squares (OLS) Estimation Method. If the estimated value of the scaling parameter ($\hat{\alpha}$) ranges between 1.5 and 3 then it can be said that the data follow Power Law probability distribution. However, the estimated value of the scaling parameter is biased (Newman, 2005: 327; Clauset, Shalizi and Newman, 2009: 665). In this situation, Maximum Likelihood Estimates (MLE) is proposed by Clauset, Shalizi and Newman (2009). When the variable is continuous then the parameters and standard deviation of the scaling parameter is estimated with the following formulas:

$$\hat{C} = (\alpha - 1)x_{min}^{\alpha-1} \tag{3}$$

$$\hat{\alpha} = 1 + n\left[\sum_{i=1}^{n} ln\frac{x_i}{x_{min}}\right]^{-1} \tag{4}$$

$$\hat{\sigma}_{\hat{\alpha}} = \frac{\hat{\alpha}-1}{\sqrt{n}} + O\left(\frac{1}{n}\right) \tag{5}$$

On the other hand, when the variable is discrete then the parameter estimates turn into following formulas:

$$\hat{C} = \frac{1}{\sum_{n=0}^{\infty}(n+x_{min})^{-\alpha}} \tag{6}$$

$$\hat{\alpha} \cong 1 + n\left[\sum_{i=1}^{n} ln\frac{x_i}{x_{min}-\frac{1}{2}}\right]^{-1} \tag{7}$$

Researchers indicated that when the x_{min} value is taken equal to or greater than six ($x_{min} \geq 6$), the data fits well to the power-law distribution. In the end, the data should be reduced. After the MLE estimation, if the estimated value of the scaling parameter ($\hat{\alpha}$) ranges between 1.5 and 3 then it can be said that the data follow Power Law probability distribution.

In some cases the behavior of power-law probability distribution breaks down. For instance; the noise is seen in the tail of the histogram on the logarithmic scales as it is in lower side of Figure 1(b). In such case, the estimation of x_{min} becomes important in order to take its proposed value 6. Because, x_{min} represents the starting point of the behavior of power-law probability distribution. By estimating the true value of x_{min}more reliable estimate of the scaling parameter can be obtained with the MLE method above.

2.2.2. Transformation Method

One of the ways of testing the data in a formal way is to use the transformation method which was proposed by Clauset, Shalizi and Newman (2009). 1000 uniformly distributed random real numbers that are between zero and one $(0 \leq r \leq 1)$ reproduced with this approach. $\hat{\alpha}$ parameter that was estimated with MLE method, $x_{min} = 6$ value and reproduced random numbers (r) are used in the following transformation formula to produce random numbers that follow Power Law probability distribution.

$$x = x_{min}(1-r)^{-1/\alpha-1} \tag{8}$$

After this simulation, 95 th percentile of the reduced data and 95 th Percentiles of 100 samples for n=1000 are compared. When 95 th Percentiles of 100 samples for n=1000 is higher than 95 th percentile of the reduced data, then it is said that the data follows a Power Law distribution. The weakness of this approach is the usage of $x_{min} = 6$, not estimating it .

2.2.3. Kolmogorov Smirnov Test

More reliable formal method is goodness of fit test named Kolmogorov Smirnov Test. KS statistic (D) is simply the maximum distance between cumulative distribution function $S(x)$ of the data for the observations with value at least x_{min} and cumulative distribution function for the power-law model that fits the data in the region $x \geq x_{min}$.

$$D = \max_{x \geq x_{min}} |S(x) - P(x)| \tag{9}$$

The usage of estimation of x_{min} minimizes D (Clauset, Shalizi and Newman, 2009: 672). The hypotheses are as follows:

H_0: The data follow Power Law Distribution

H_1: The data do not follow Power Law Distribution

The p-value of D statistic that is close to “1” indicates that the data follows a Power Law probability distribution. On the other hand, Goldstein and others (2004) derived following KS Test Table for Power Law Distribution which considers different n:

Table 1. KS Test Table for Power Law Distribution (Goldstein and others, 2004:3)

	Quantile			
n	0.9	0.95	0.99	0.999
10	0.1765	0.2103	0.2835	0.3874
20	0.1257	0.1486	0.2003	0.2696
30	0.1048	0.1239	0.1627	0.2127
40	0.0920	0.1075	0.1439	0.1857
50	0.0826	0.0979	0.1281	0.1719
100	0.0580	0.0692	0.0922	0.1164
500	0.0258	0.0307	0.0412	0.0550
1000	0.0186	0.0216	0.0283	0.0358
2000	0.0129	0.0151	0.0197	0.0246
3000	0.0102	0.0118	0.0155	0.0202
4000	0.0087	0.0101	0.0131	0.0172
5000	0.0073	0.0086	0.0113	0.0147
10000	0.0059	0.0069	0.0089	0.0117
50000	0.0025	0.0034	0.0061	0.0077

If the D statistic is lower than the Table value, then then it is said that the data follows a Power Law distribution.

3. WEIBULL, GAMMA AND LOGNORMAL PROBABILITY DISTRIBUTIONS

A random variable x is said to have a **Weibull Distribution** with parameters α and β if the probability density function of x is

$$f(x) = \frac{\alpha}{\beta^{\alpha}} x^{\alpha-1} e^{-(\frac{x}{\beta})^{\alpha}} \tag{10}$$

Here, $\alpha > 0$ and $\beta > 0$.

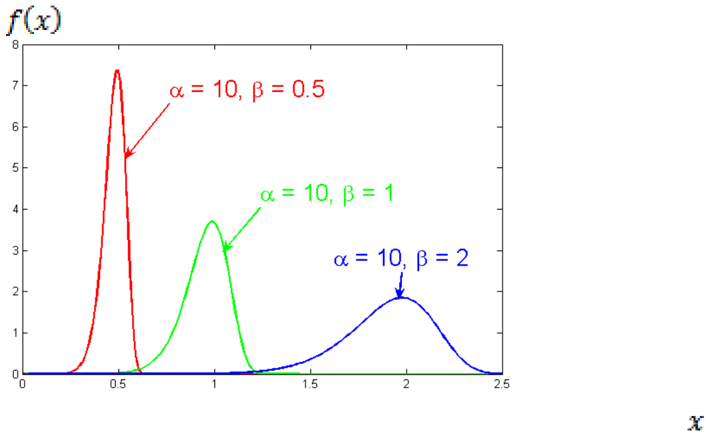


Figure 2.
Visual View of the Weibull Distribution

The figure above shows that the Weibull distribution's shape changes depending on the different values of α and β parameters.

If x is a continuous random variable then is said to have a **Gamma Distribution** if the probability density function of x is

$$f(x) = \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$$

Here $x \geq 0$, $\alpha > 0$ and $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\, dx$ (Aytaç, 2004: 308). When $\beta = 1$ then we have the standard Gamma Distribution. α parameter governs the shape of the gamma density and β parameter is a scale parameter. Gamma distribution shape changes depending on the different values of α parameter as is seen below.

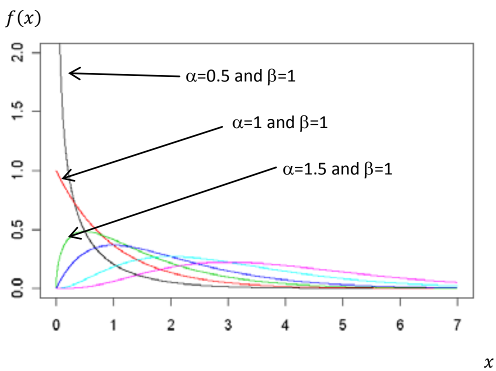
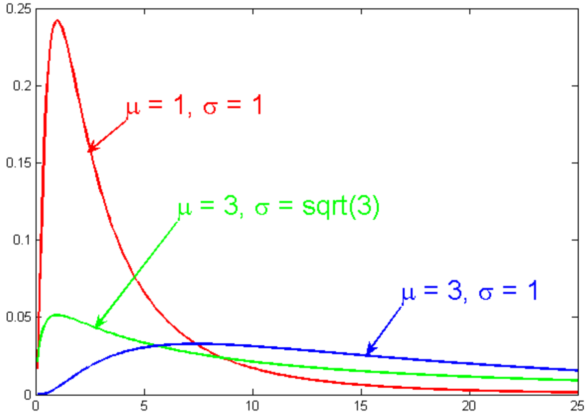


Figure 3.
Visual View of the Gamma Distribution

A random variable x is said to have a **Lognormal Distribution** if the random variable $y = \ln(x)$ has a normal distribution. The resulting probability density function of a lognormal random variable when $\ln(x)$ is normally distributed with parameters μ and σ is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-[\ln(x)-\mu]^2/(2\sigma^2)}$$

$f(x)$



x

Figure 4.
Visual View of the Lognormal Distribution

There are two main ways of the determination of whether the empirical data follow one of the common distributions or not. These are graphical and formal methods.

3.1. Graphical Method

Probability plot (pp-plot) is a graphical method for assessing whether or not a data set follows a given distribution such as Normal, Weibull, Gamma or Lognormal and etc. The values of the given variable which was arranged in ascending order (vertical axis) are plotted against the percentiles of a given distribution (horizontal axis).

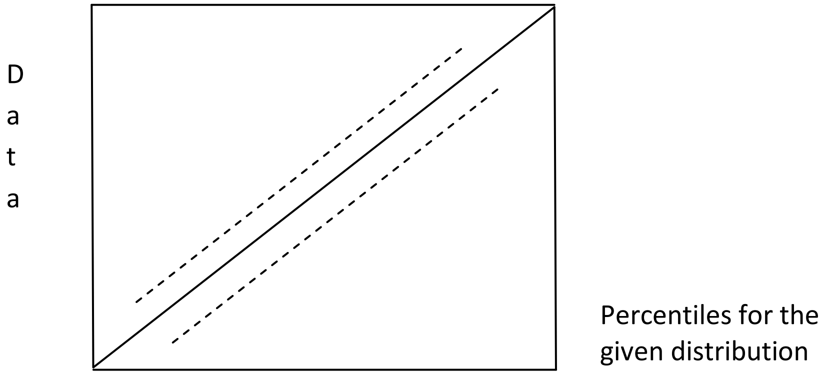


Figure 5.
Probability Plot

The data are plotted against a theoretical distribution in such a way that the points should form approximately a straight line. Departures from this straight line and upper and lower bounds (the dashed lines) indicate departures from the specific distribution.

pp-plot provide the information about the outliers of the given data and the skewness of the data graphically (Ravi and Butar, 2010: 2). If the points are above the straight line, then this means that the data is left skewed data. If the points are below the straight line, then this means that the data is right skewed data.

3.2. Formal Methods

The formal methods are **Kolmogorov Smirnov (KS) Test** and its modification named **Anderson Darling (AD) Test** for assessing whether or not a data set follows a given distribution such as Normal, Weibull, Gamma or Lognormal and etc.

The Anderson Darling Test is used to test if a sample of data came from a population with a specific distribution. It is the modification of the KS Test and gives more weight to tails than does KS test. The KS test is distribution free in the sense that the critical values do not depend on the specific distribution being tested. On the other hand, the Anderson Darling test makes use of the specific distribution in calculating critical values. This has the advantage of allowing a more sensitive test. Hypotheses of this test are as follows:

H_0: The data follow a specified Distribution

H_1: The data do not follow a specified Distribution

4. RESEARCH AND RESULTS

In this section, whether the Lengths of Stay in Hospital data (a sample of a Turkish Hospital in Edirne) follows a Power Law probability distribution or not will be tested at first. Then, the commonly used distributions Weibull, Gamma or Lognormal will be tested.

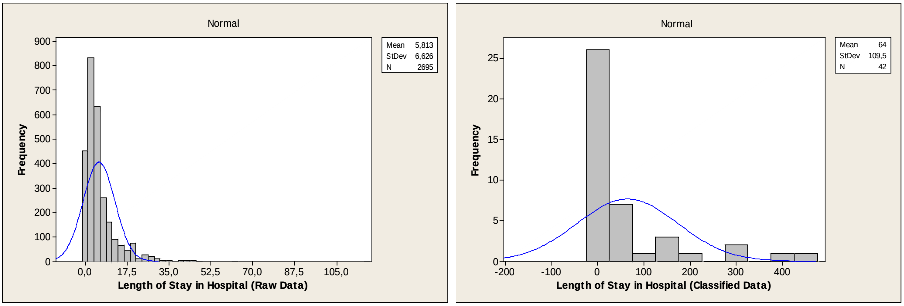


Figure 6.
Histograms of the Length of Stay in Hospital

In the figures above the histograms of the data of Lengths of Stay in Hospital are seen (See Appendix 1 for the data). The distributions of both data (raw data and classified data) have very asymmetrical distributions. The right tail of the both distribution is longer which means that the distribution is right skewed or right-tailed. Mode, median and mean of the raw data are respectively in ascending order Mo=2, Me=4 and $\bar{x} = 5{,}813$. The data gives signals that it follows a Power Law distribution.

The log-log regression estimation results are shown in the table below:

Table 2. Log-log Regression Estimation Results

Models / Estimates	Log-log model with Raw Data	Log-log model with Classified Data
$\hat{C}$	6.940	8.250
p-value	0,000	0,000
$\hat{\alpha}$	-0.805	-1.970
p-value	0.000	0.000
r^2	0.85	0.88
$F\ Statistic$	14363.51	287.2
p-value	0.000	0.000

In both models, $\hat{C}$ and $\hat{\alpha}$ parameters were found statistically significant (all p-values < α=0.05). Namely, the coefficient of determinations and F Statistics are fine, as well. However, main interest is $\hat{\alpha}$ parameter's value that may give clue whether the data follows a Power Law probability distribution or not. The expectation of its value is only satisfied with classified data model which is between $1.5 \leq \hat{\alpha} \leq 3$. The data gives signals that it follows a Power Law distribution. However, as is mentioned before this estimation is biased. This value can't be used as an indicator for the behavior of power-law. More appropriate method MLE used to solve this problem. To do so the reduced data ($x_{min} \geq 6$) used (See Appendix 2 for the reduced data) as is proposed. The MLE estimation results are shown below:

$$p(x) = 28.63\ x^{-2.65}$$

$\hat{C}$ and $\hat{\alpha}$ parameters found respectively 28.63 and -2.65. In addition to this, standard error of the scaling parameter ($\hat{\sigma}_{\hat{\alpha}}$) found 0.275. Although, the $\hat{\alpha}$ value (2.65) is between $1.5 \leq \hat{\alpha} \leq 3$, we can't say that the data follows a Power Law distribution. The data gives signals that it follows a Power Law distribution, again.

One of the ways of testing the data in a formal way is to use the transformation method which was proposed by Clauset, Shalizi and Newman (2009). 1000 uniformly distributed random real numbers that are between zero and one $(0 \leq r \leq 1)$ reproduced with this approach. $\hat{\alpha}$ parameter that was estimated with MLE method, $x_{min} = 6$ value and reproduced random numbers (r) are used in the following transformation formula to produce random numbers that follow Power Law probability distribution.

$$x = x_{min}(1-r)^{-1/\alpha-1}$$

After this simulation, 95 th Percentiles of 100 samples for n=1000 are given in Appendix 3. Note that 95 th percentile of the reduced data is 27 days. When 95 th percentile of the reduced data (27) is compared with the 95 th Percentiles of 100 samples for n=1000 are given in Appendix 3, it is determined that all percentile values are higher than this value. This indicates that the data of lengths of stay in hospital follow a Power Law distribution.

In the above analysis, some signals observed that the data of lengths of stay in hospital may follow a Power Law distribution. However, unless we estimate x_{min} value and perform KS test we cannot still say that the data exactly follow a Power Law probability distribution. Because more reliable C and α parameter MLE estimates can be obtained by estimating x_{min} value. Following two tables present the mentioned MLE estimates, KS test statistics and the comparisons.

Table 3. MLE Estimates and KS Test Statistics

Estimates / Data	$\hat{x}_{min}$	$\hat{C}$	$\hat{\alpha}$	D	p-values
Classified Data	7	-147.0075	1.52	0.1599	0.0590
Raw Data	6	-2696.6	2.56	0.0641	0

p-values that are close to "1" indicates that the data follows a Power Law probability distribution. In the Table above, none of the p-values are close to 1. Results are exactly opposite that they are respectively zero and close to zero. Moreover, comparisons of KS test values with Goldstein and others (2004) KS table values give same results in the table below.

Table 4. Comparisons of KS Test Values with Goldstein and other's (2004) KS Table Values

	n	D	0.95	0.99	0.999
Classified Data	42	0.1599	>0.0979 (doesn't fit)	>0.1281 (doesn't fit)	<0.1719(fits)
Raw Data	927	0.0641	>0.0216 (doesn't fit)	>0.0283 (doesn't fit)	>0.0358 (doesn't fit)

As a result, KS test results show that the data of lengths of stay in hospital don't fit Power Law probability distribution.

Second, a formal test called Anderson Darling (AD) test and a visual decision tool that gives upper and lower bounds called Probability Plot used to decide whether the data in Appendix 1 follow one of the commonly used distributions (Weibull, Gamma and Lognormal).

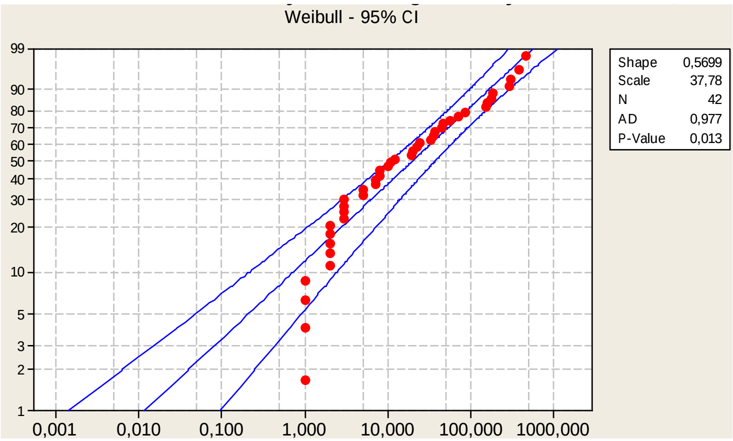


Figure 7.
PP Plot of Length of Stay for the Weibull Distribution and AD Test Statistic

AD test statistics is found 0.977. When p-value of AD statistic is compared with significance level (p=0.013<α=0.05) null hypothesis rejected which means that the data doesn't follow Weibull probability distribution.

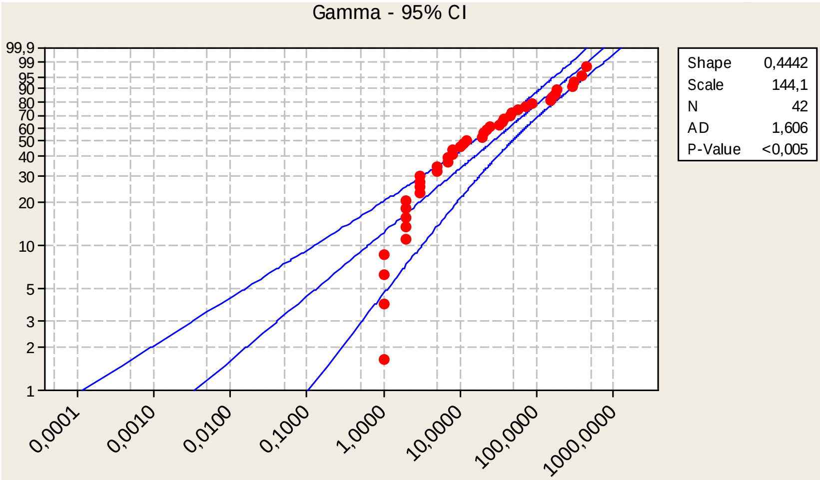


Figure 8.
PP Plot of Length of Stay for the Gamma Distribution and AD Test Statistic

AD test statistics is found 1.606. When p-value of AD statistic is compared with significance level (p=0.005<α=0.05) null hypothesis rejected which means that the data doesn't follow Gamma probability distribution.

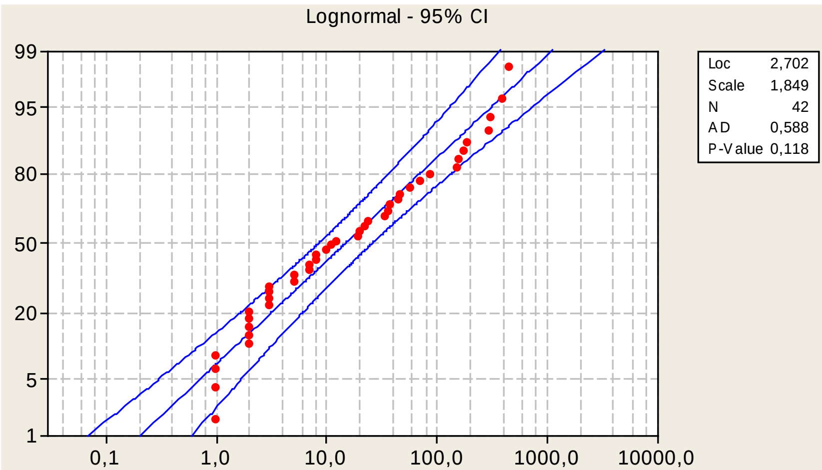


Figure 9.
PP Plot of Length of Stay for the Lognormal Distribution and AD Test Statistic

AD test statistics is found 0.588. When p-value of AD statistic is compared with significance level ($p=0.118>\alpha=0.05$) null hypothesis didn't reject which means that the data follow Lognormal probability distribution.

5. CONCLUSION

The LOS in hospital has great importance for hospital management. It is used for various purposes such as management of hospital care, quality control, appropriateness of hospital use, hospital planning, formation of payment policy, measure of hospital efficiency and etc. In the related literature, it is well known that the empirical distribution of LOS is positively skewed. Because of the empirical distribution of LOS is very asymmetrical with a broad tail, it makes many statistical estimates less robust. The determination of the distribution of LOS comes into prominence within this scope.

Besides three commonly or widely used models for describing the distributions (Weibull, Gamma and Lognormal) of LOS indicated by Marazzi, Paccaud and Ruffieux (1998), Hellervik and Rodgers (2007) showed that the power law probability distribution model can also be used for describing the distribution of LOS. In this study, we examined Paccaud and Ruffieux's (1998) and Hellervik and Rodgers's (2007) mentioned models. Namely, we focused on the investigation of whether the distribution of the data follows Power Law, Weibull, Gamma and Lognormal or not. For this purpose, a sample of a Turkish Hospital data was used and tested. Results show that the data follows a Lognormal Probability Distribution for the Turkish Case. To conclude, the best way to model LOS for the sample used in this study is to use logarithmic transformation of the variable and use it with ordinary least square (OLS) regression and apply associated tests.

The results of this research have some limitations. First, these results are valid for the Edirne Hospital. More studies for the other Turkish hospitals are needed to be performed. Second, the sample that was used in this study covers a specific period of time. Sample can be extended for more periods. Finally, the population of the data assumed to be homogeneous and the sample tested for a specific probability distribution. However, there exists a literature that states the population of the data can be heterogeneous which means that the data can be characterized more than one distribution. For the further studies the data should be tested whether it can be modeled with the mixture probability models.

REFERENCES

Aytaç, M. (2004), ***Matematiksel İstatistik***, Ezgi Kitabevi, Bursa.

Bai, X., Kwok, T. C. Y., Ip, I. N., Woo, j., Chui, M. Y. P. ve Ho, F. K. Y. (2014), "Physical restraint use and older patients' length of hospital stay", ***Health Psychology & Behavioural Medicine,*** 2(1), s.160-170.

Chassin, M. R. (1983), ***Variations in Hospital Length of Stay: Their Relationship to Health Outcomes***, Health Technology Case Study: 24, Ota Publishing.

Clauset, A.,Shalizi, C. H. ve Newman, M. E. J. (2009), "Power-law distributions in empirical data", 51(4), s.661-703.

Esatoğlu, A. E. ve Bozat, S. (2002), "Survey on the Length of Stay for the Patients with Chronic Obstructive Pulmonary Disease: An Application on Atatürk Chest Disease Hospital", ***Journal of Ankara Medical School,*** 24(4), s.165-176.

Faddy, M., Graves, N. ve Pettitt, A. (2009), "Modeling Length of Stay in Hospital and Other Right Skewed Data: Comparison of Phase-Type, Gamma and Log-Normal Distributions", ***Value in Health,*** 12(2), s. 309-314.

Goldstein, M. L., Morris, S. A. & Yen, G. G. (2004), "Problems with Fitting to Power-law Distribution", ***European Physical Journal B,*** 41(2), s.1-4.

Hellervik, A. & Rodgers, G. J. (2007) "A power law distribution in patients lengths of stay in hospital", ***Physica A,*** 379, s. 235-240.

Keefler, J., Duder, S. ve Lechman, C. (2001), "Predicting Length of Stay in an Acute Care Hospital: The Role of Psychosocial Problems***", Social Work in Health Care,*** 33(2), s.1-16.

Lee, W., Chiu, W., Tsai, S., Lin, M., Chu, S. ve Lam, C. (2010), "Comparison of Length of Hospital Stay Between Taipei City and Hualien Country in Spinal Trauma Patients", ***J Exp Clin Med***, 2(5), s.239-244.

Lim, A. & Tongkumchum, P. (2009), "Methods for Analyzing Hospital Length of Stay with Application to Inpatients Dying in Southern Thailand", ***Global Journal of Health Science,*** 1(1), s.27-38.

Marazzi, A., Paccaud, F. & Ruffieux, C. (1998), "Fitting the distributions of length of stay by parametric models", ***Medical Care***, 36(6), s.915-927.

Mitzenmacher, M. (2003), "A Brief History of Generative Models for Power Law and Lognormal Distributions", ***Internet Mathematics***, 1(2), s.226-251.

Newman, M. E. J. (2005), "Power Laws, Pareto Distributions and Zipf's Law", ***Contemporary Physics,*** 46(5), s.323-351.

Rafiei, M., Ayatollahi, S.M.T. & Behboodian, J. (2007), "Length of Hospital Stay at Arak Maternity Clinics Using Proposed Zero-Inflated Negative Binomial Modeling", ***Pakistan Journal of Biological Sciences***, 10(15), s. 2510-2516.

Ravi, A. ve Butar, F. B. (2010), "An Insight to Heavy-Tailed Distribution", ***Journal of Mathematical Sciences & Mathematics Education,*** 5(1), s.1-15.

Xiao, J., Lee, A. H. & Vemuri, S. R. (1999), "Mixture distribution analysis of length of hospital stay for efficient funding", ***Socio-Economic Planning Sciences***, 33, s. 39-59.

Appendix 1. Data

Lengths of Stay in Hospital (Days)	Frequency
0	158
1	294
2	451
3	382
4	299
5	184
6	153
7	174
8	86
9	70
10	46
11	44
12	57
13	33
14	37
15	22
16	5
17	20
18	24
19	12
20	19
21	36
22	7
23	2
24	10
25	7
26	8
27	8
28	11
29	3
30	2
31	5
32	1
33	3
36	2
39	3
42	2
43	2
45	3
48	1
62	1
117	1

Appendix 2. Reduced Data ($x \geq 6$)

Lengths of Stay in Hospital (Days)	Frequency
6	153
7	174
8	86
9	70
10	46
11	44
12	57
13	33
14	37
15	22
16	5
17	20
18	24
19	12
20	19
21	36
22	7
23	2
24	10
25	7
26	8
27	8
28	11
29	3
30	2
31	5
32	1
33	3
36	2
39	3
42	2
43	2
45	3
48	1
62	1
117	1

Appendix 3. 95 th Percentiles of 100 samples for n=1000

37,000	41,552	32,874	38,380	34,380	32,849	35,040	38,420	33,030	39,400
34.030	38,364	35,010	33,746	35,481	36,897	33,573	39,530	30,314	36,104
36,530	39,571	35,028	36,231	34,480	35,090	37,265	31,929	32,590	35,409
39,609	36,590	35,620	33,646	37,552	33,190	35,416	39,100	35,916	37,630
40,304	34,330	37,223	32,936	36,599	36,733	36,218	38,021	37,573	37,908
35,111	33,616	36,323	30,796	32,296	35,607	39,435	32,569	38,580	30,834
38,940	34,578	35,424	32,410	34,546	37,689	38,475	39,501	38,581	41,980
39,007	37,360	37,215	31,734	35,009	37,750	39,511	32,170	40,160	32,708
36,754	34,907	34,663	42,173	36,522	39,106	37,873	37,743	42,483	40,454
41,401	42,974	40,313	40,887	41,400	36,373	34,593	38,018	36,382	39,621