Statistical Analysis of the Count of the Prime Numbers at Certain Intervals

King Saud University, Teachers College, Department of Mathematics
Prof. Dr. Hamid Al-Oklah **Department of Mathematics**
Dr. Mahmod Rahal **Department of Computer**
Dr. Ahmad Jello **Department of Mathematics**

Abstract

This research study aims at analyzing the count of prime numbers at regular intervals of 10000. The values referring to the count of prime numbers were generated at these intervals using a special program of generating prime numbers and counting them at each of 100000 intervals. The statistical analysis of these data addresses the general trend of these data after dividing them into five sets. It has been noted that there is a general declining trend, and because of the massive amount of data, we divided the100000 data into 200 sequentially groups (or sets), and every group contains 500 sequentially data. Then the researchers had to take a uniform random sample of the sets above-mentioned in order to render a more detailed analysis of the data of these sets. So, the descriptive presentation of the most important numerical features of the data was made and then the validity intervals and the general trend of each of these sets were assigned. Finally, the probability distributions that are consistent with the data of these sets were identified. As a result, it was too difficult to find a general and comprehensive formula of test to assign a count of prime numbers at intervals of specific length because of the variety of probability distributions that are consistent with the data of these sets as well as the change in the general trend of the sets under consideration.

Key Words: Prime numbers, count of prime numbers**,** statistical analysis**, random processes, Intervals confidence of mean and median.**

Preface:

Areas and Castor's paper (Ref. [1]) show the randomness of the prime numbers. Van der Galien (2002), (Ref. [11]), asks: "Are the prime numbers randomly distributed?", and Vaughan (Ref. [12]) investigates the mean value theorem in the prime numbers. Liu and Jia investigate the prim numbers in short intervals (Ref. [3], [7]).

In this research, we investigate a problem in the prime numbers theory. This problem says: how is the behavior of the count of the prime numbers in certain long regular intervals ?

Our aim in this research is the statistical analyzing of the count of the prime numbers at long regular intervals (intervals of long 10000). The values referring to the count of prime numbers were generated at these intervals using a special program of generating prime numbers and counting them in each of 100000 intervals (Ref. [9]).

Research Strategies

It is well-known that statistical studies on any subject offer practical perspectives on the subject under consideration and provide an idea about the development of this subject in the future. In what follows, the researchers will carry out a statistical analysis of the count of the prime numbers that fall between 1 and 10^9 divided by certain intervals and occurring in uniform intervals or divided at regular intervals. Because of the massive amount of data and the nature of these data, the researchers followed a specific approach to analyzing these data:

1. The researchers took intervals of equal length each of which was 10000. In other words, the first interval starts with the number (1) and ends with the number (10000). Then the second interval starts with the number (10001) and ends with the number (20000) and so on and so forth until the last interval starts with the number (999,990,001) and ends with 1.000.000.000.
2. The researchers assigned the count of prime numbers at each interval using an invented program for this purpose (Ref. [9]). For example, the count of prime numbers at the first interval is 1229 prime numbers and at the second interval is 1033 prime numbers and so on and so forth for the rest of the intervals. The values of 1229, 1033, etc.. . represent the data to be analyzed, and these data can be looked at as random values because they assume our knowledge of prime numbers at a certain interval. In fact, our knowledge of the count of numbers at the next interval cannot be pre-assumed. Besides, the values we get from these intervals match with the results of the random experiment.

3. The researchers analyzed and clarified the general trend of all these data after dividing them into five sets because of the massive amount of data.

4. Because of the massive amount of data in most above-mentioned sets, quite many features of these data might not be clear. Therefore, the researchers distributed these data over two sets each of which has 500 values of successive intervals. Each of these sets was regarded as a sample unit. For example, in the first set, there are 500 values (each value represents a count of the prime numbers at intervals of 10000 as mentioned previously). The second set includes the next 500 and so on and so forth so as to have 200 sets.

5. The researchers will study the first and the last sets separately because the first group behaves completely differently from the rest of the sets. However, the last set is important and necessary as it indicates the end of the values under consideration.

6. A uniform random sample will be taken including 28 sets out of the remaining 198 sets to be examined in order to avoid studying many successive probable sets in the case of simple random samples (we know that simple random samples are characterized by the probability of selecting any other random sample of the same size and can be formed from the statistical body itself) which might result in ignoring the study of a big number of other successive sets and the induction of path of the distinctive values of the data will be clearer. Thus, 30 sets will have to be analyzed (each of which includes 500 values).

Last, the researchers would like to say that various statistical programs will be used for the mathematical operations and the necessary data presentations. These programs include new editions of Minitab, SPSS, EasyFit depending on the need for each of them.

Discussion and Findings

1. Analysis of the general trend of the data:

Data (100000 values) have been divided into five sets as follows:

The first set includes the first 500 values, the second set includes the next 24500 values and the remaining three sets each include 25000 values respectively.

By using Minitab program and MSD (mean square deviation) accuracy measures in selecting the best analysis, the following analyses and presentations of the five sets have been found. The first set (contain the first 500 data) has the following presentations (figure (1)) for the trend:

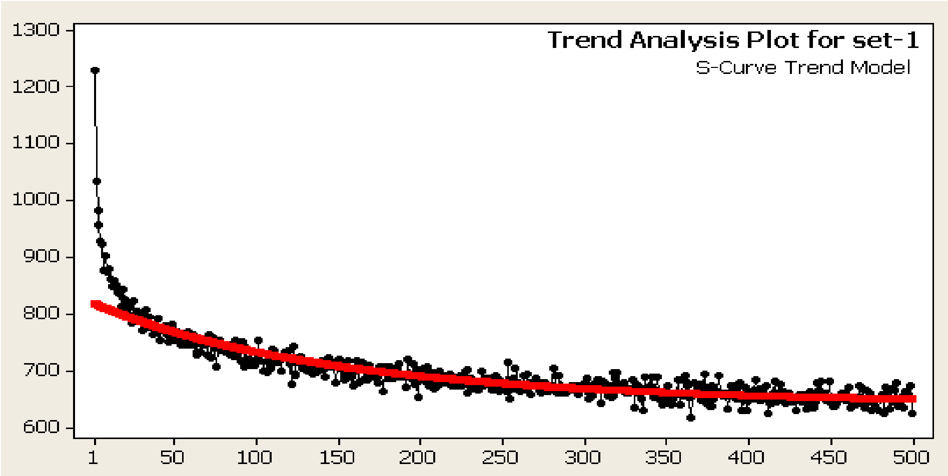


Figure (1)

Here we find that the proper curve of the general trend of the data is a S-curve and its fitted trend equation is as follows:

$$Y(t) = \frac{10^4}{15.5962 - 3.40488(0.907283^t)}$$

The second set (contain the data from 501 until 25000) has the following presentations (figure (2)):

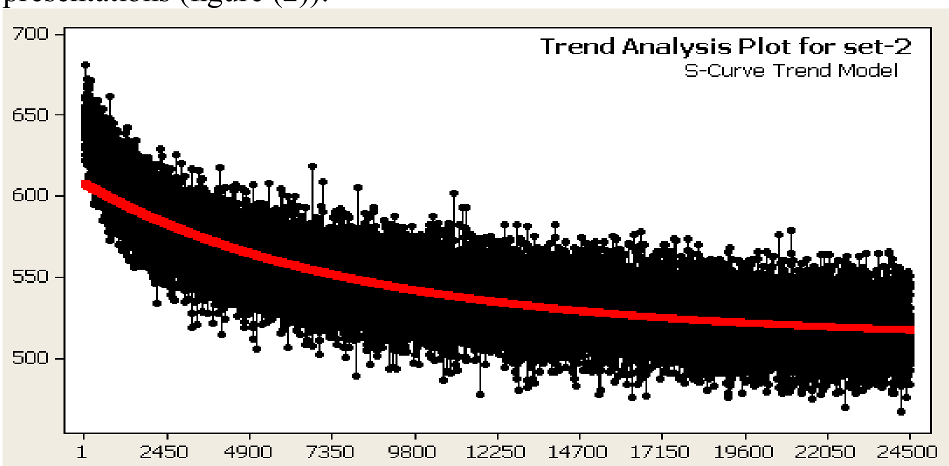


Figure (2)

Here we find that the proper curve of the general trend of the data is a S-curve and its fitted trend equation is as follows:

$$Y(t) = \frac{10^4}{19.5388 - 3.09022(0.999895^t)}$$

The third set (contain the data from 25001 until 50000) has the following presentations (figure (3)):

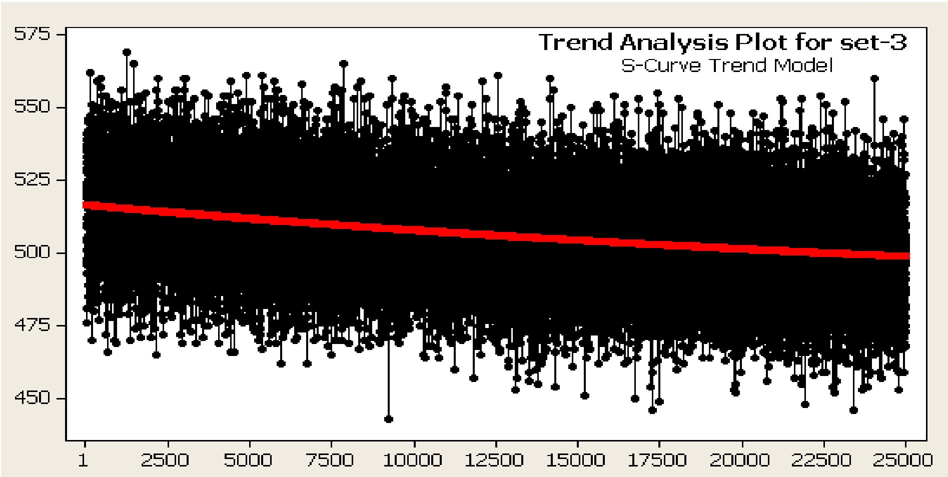


Figure (3)

Here we find that the proper curve of the general trend of the data is a S-curve and its fitted trend equation is as follows:

$$Y(t) = \frac{10^4}{20.7848 - 1.42349(0.999974^t)}$$

The fourth set (contain the data from 50001 until 75000) has the following presentations (figure (4)):

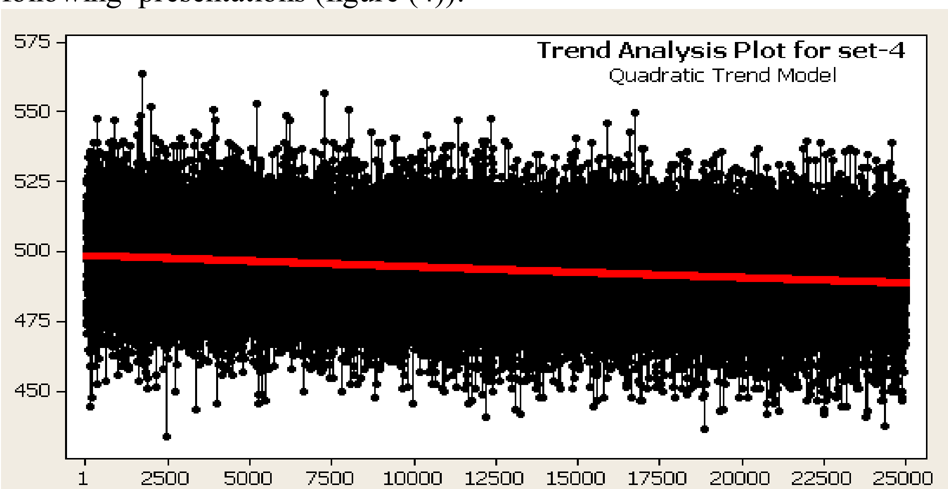


Figure (4)

Here we find that the proper curve of the general trend of the data is a **quadratic polynomial** curve and its fitted trend equation is as follows:

$$Y(t) = 498.936 - 0.000424\ t + 0.000000\ t^2$$

The coefficient of t^2 are smaller than 10^{-6}. Given the fact that the coefficient of t express the general trend, then t^2 here expresses the regression in this trend. Since the value of the coefficient of t^2 is too small, this means that the following linear equation can be accepted to represent the general trend:

$Y = f(t) = 498.936 - 0.000424\ t$

Since the value of the coefficient of t is too small (with a minus sign), this means that on average there will be an increase in the count of the prime numbers at intervals whose length is 10000 and whose position after the value 5×10^8 will be slow and with a linear curve too. In fact, if we look at the general linear trend of these data, we will find the following equation for it:

$Y(t) = 498.806 - 0.000393\ t$

It is a very close formula to the previous one.

The fifth set (contain the data from 75001 until 100000) has the following presentations (figure (5)):

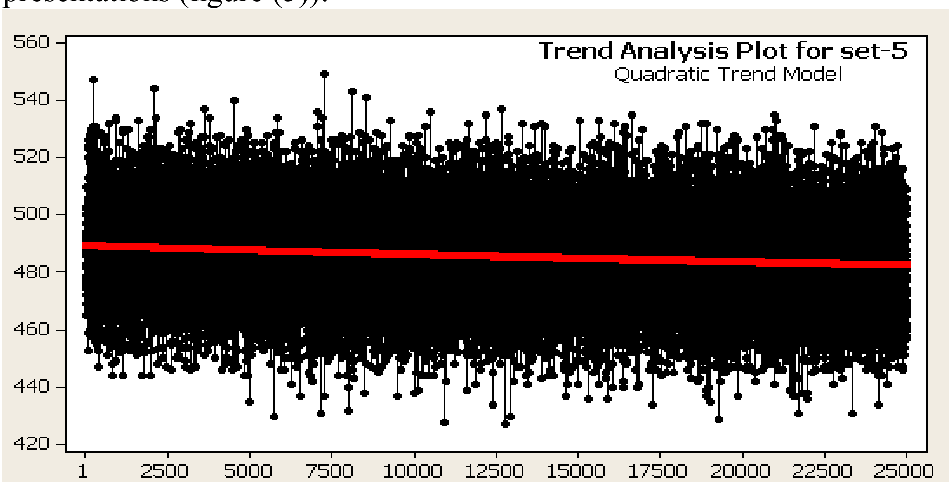


Figure (5)

Here we find that the proper curve of the general trend of these data is a quadratic polynomial curve whose mathematical equation is as follows:

$Y(t) = 489.400 - 0.000329\ t + 0.000000\ t^2$

In this equation the coefficient of t^2 are smaller than 10^{-6} too. Then the following linear equation can be accepted:

$Y(t) = 489.400 - 0.000329\ t$

which represents the general trend. Since the values of the coefficients of t are very small which means that on average there will be an increase in the count of the prime numbers at intervals whose length is 10000 and whose position after the value 75×10^7 will be slow and with a linear curve too, but even slower than before because the absolute value of the coefficients of t here are smaller. In fact, if we look at the general linear trend of these data, we will find the following equation for it:

$Y(t) = 489.156 - 0.000271\ t$

It is a very close formula to the previous one.

Now we will carry out a statistical study of a random sample of the sets of these data in order know more closely the behavior of these data so that each set includes 500 consecutive values. We will depend on a uniform random sample so that the analysis of these data will be as objective as possible because bias may result from the succession of many sets when simple samples are used.

It is well-known that a uniform random sample is a kind of random sample based on the following principles (Ref. [5]):

1. All members of the population will be given consecutive numbers starting with number one (1).
2. The value of a whole number will be s and called " pull interval " and given in the following relation $s := \left\lfloor \frac{N}{n} \right\rfloor$ where N and n represent the population and sample respectively and $\left\lfloor \frac{N}{n} \right\rfloor$ refers to the biggest whole number smaller than or equal to $\frac{N}{n}$ (it is possible $\frac{N}{n}$ is not a whole number) and on condition that the population is limited (because if it is not limited, the last relation is meaningless).
3. A member of the population will be randomly selected provided that its number falls between (1) and s and this member is called the initiation element. If we assume that this element is u then $1 \le u \le s$.
4. The rest of the sample will be taken from the population accompanied by $u + k \cdot s$ with $k = 1, 2, \ldots\ldots$ until all elements of the population are analyzed.

To know the groups which will be studied according to the uniform random sample we have $N = 198$ and $n = 28$, then the value of the pull interval will be $s = \left\lfloor \frac{198}{28} \right\rfloor = \lfloor 7.07 \rfloor = 7$, and when we choose a random number that falls between 1 and 7 using the generator of random numbers

in calculators (according to uniform distribution (Ref. [5], [6])), we have number 4. Since we will study the first group and the last group separately, the groups that will be studied are the ones that have the following numbers:
5 , 12 , 19 , 26 , 33 , 40 , 47 , 54 , 61 , 68 , 75, 82 , 89 , 96 , 103 , 110 , 117 , 124 , 131 ,138 , 145 , 152 , 159 , 166 , 173 , 180 , 187 , 194 .
It is well-known that when carrying out a statistical analysis it is better to conduct a descriptive study of the data because this kind of study gives a preliminary conception of the nature and probable distribution of the data in addition to acquainting us with some distinctive numerical values of these data. But, before we look at the path of the numerical values of the prime numbers using the time series, we shall assume that each interval (which is 10000 long) is a time unit.
The first group looks graphically as in figure (1). Whereas we find that the other groups have similar representations (Figure (6-a until 6-e)) that look like the following ones which are related to some groups (because displaying the 29 representations requires a big number of pages, and this is not acceptable at least here in this research. Those who would like to learn about all these representations can contact the researchers to get more details).

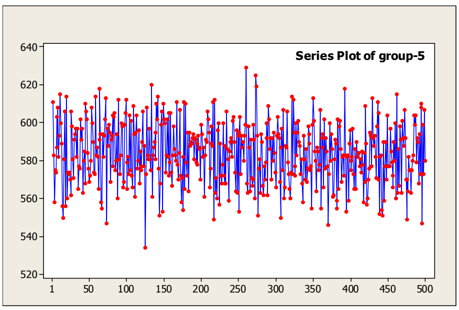


Figure (6-a)

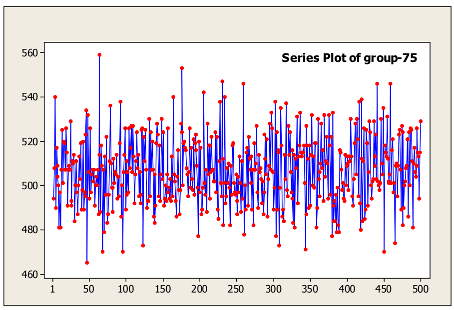


Figure (6-b)

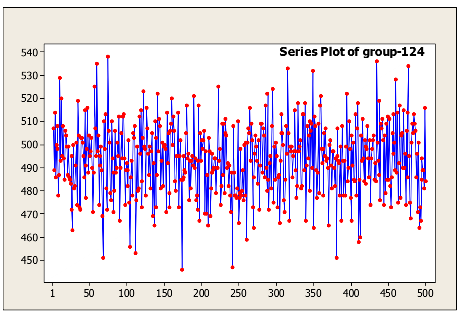


Figure (6-c)

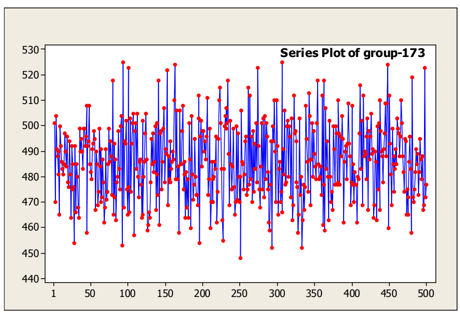
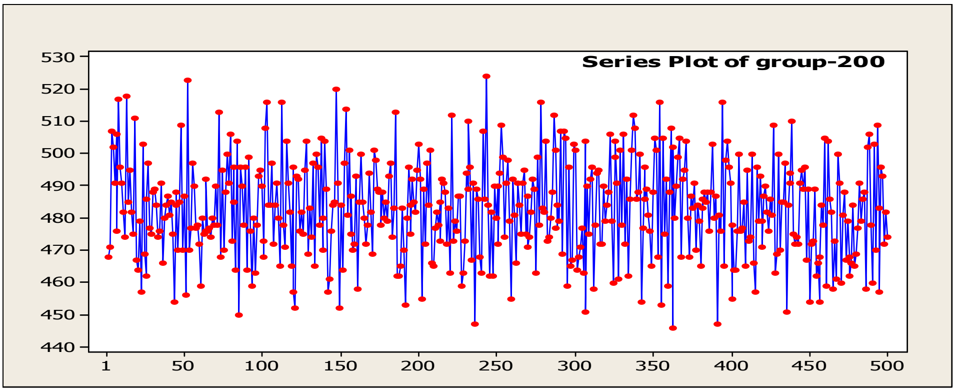


Figure (6-d)



Figures (6-e)

We find that the first group is completely different in its behavior from the other following groups. For this reason, the first group will be exclusively examined here in this study.

Statistical Analysis of the groups

First, some basic analysis of the groups which will be examined beginning with the first group will be presented.

1- Statistical Analysis of the first Element of the Sample (Group-1):

We have noticed a great regression in the beginning of the path of Group-1. The extreme decrease in the first thirty values is noticed, then the next seventy values regress less. For this reason, this group is divided into three sectors to know how the path of this group changes. These sectors are:

The first sector is from 1 to 30. In the second sector, the values from 31 to 100 are used. The rest of the data is put in the last sector (from 101 to 500). Then we have the following representations and equations of the paths of these sectors represented in the analysis of their general trend using Minitab program and the Mean squared deviation (MSD) for accuracy measures in the test of the best analysis:

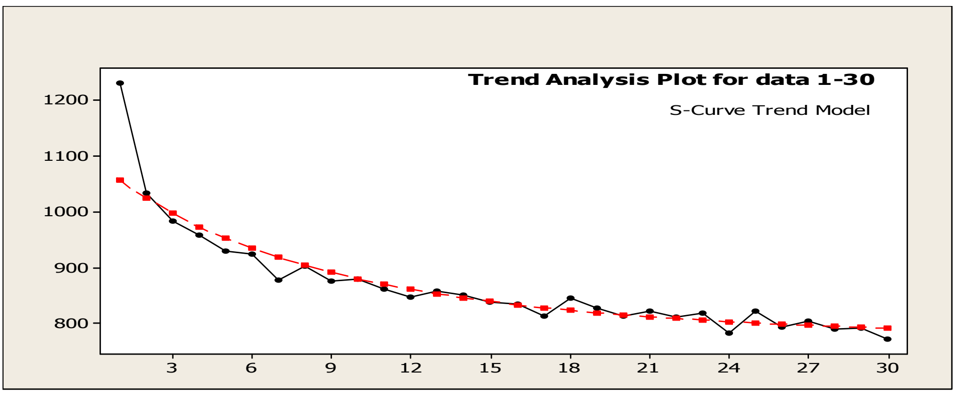


Figure (7-a)

The fitted trend equation is $Y(t) = \dfrac{10^4}{12.8700 - 3.72328(0.913655^t)}$.

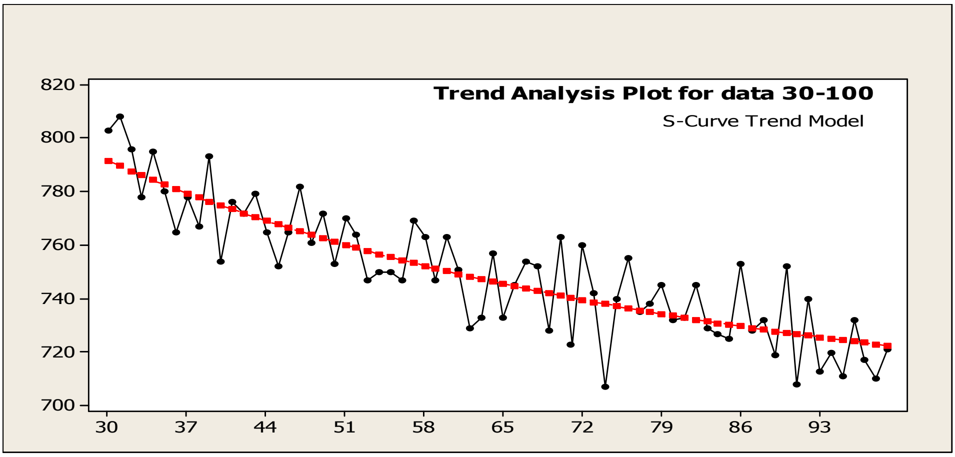


Figure (7-b)

The fitted trend equation is $Y(t) = \dfrac{10^4}{14.4052 - 1.79735(0.983609^t)}$

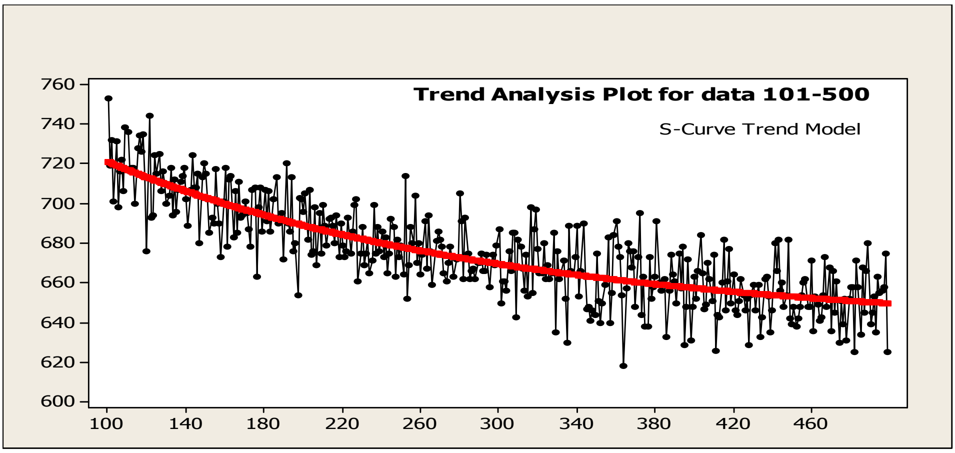


Figure (7-c)

The fitted trend equation is $Y(t) = \dfrac{10^4}{15.7319 - 1.86823(0.995763^t)}$.

We notice here that the general trend of the three sectors matches the S-curve but with different coefficients. The most important numerical properties of these three sectors are presented in the following table:

Table (1)

sector	Sample Size	Min	Max	Mean	Median	Std. Deviation
1	30	773	1229	866.57	841.5	91.962
2	70	707	808	750.01	750.5	23.813
3	400	618	753	675.02	673	24.593

The statistical account above shows a decrease in the median and mean values, and it is crystal clear the standard deviation for the first sector is big compared to the values of the standard deviation of the next two sectors. This is a normal result because of the great fall in the data of the first sector.
When a test was conducted to see if the data of these sectors were normally distributed (at a level of significance $\alpha = 0.05$ (Ref. [5])) or not, the normal distribution as a proper one for the data of these sectors which the following test results show was not acceptable:

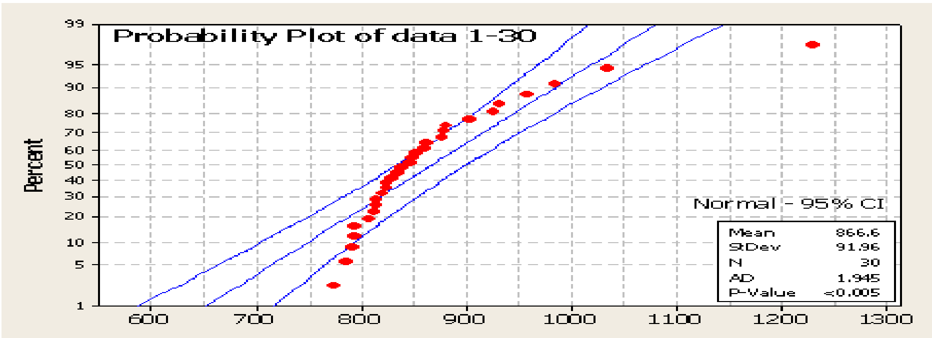


Figure (8-a) the normal distribution isn't acceptable

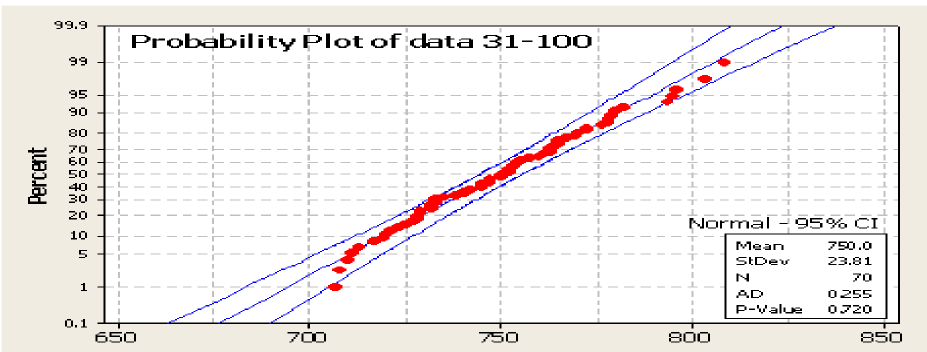


Figure (8-b) the normal distribution is acceptable

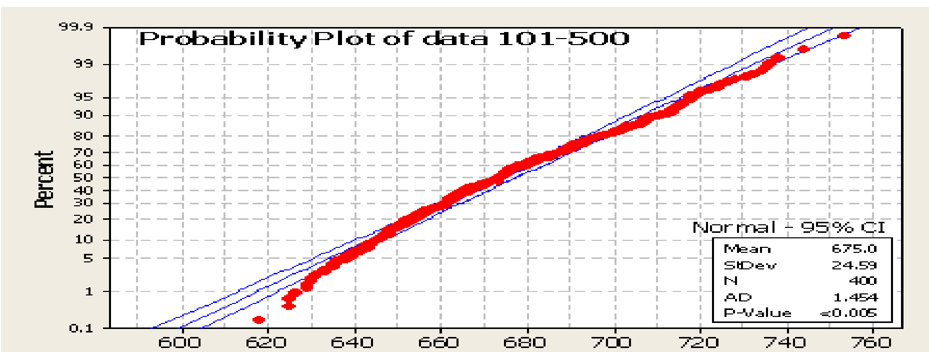


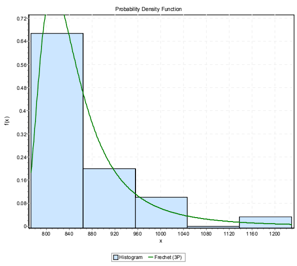
Figure (8-c) the normal distribution isn't acceptable

By using EasyFit program, we find that the best distribution of the data of these sectors according to Kolmogorov-Smirnov Test (Ref. [5]) at a level of significance $\alpha = 0.05$ is respectively as follows:

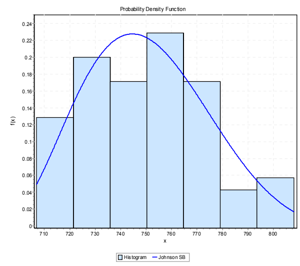
Frechet (3P) Distribution with parameters $\alpha = 3.11, \beta = 133.38$, $\gamma = 690.15$

Johnson-SB Distribution with parameters $\gamma = 0.70$, $\delta = 1.63$, $\lambda = 174.05$, $\xi = 679.89$

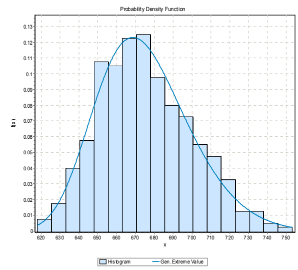
General Extreme Value with parameters $k =$ -0.15, $\sigma = 22.70$, $\mu = 664.91$ which are presented as follows:



Frechet (3P) Distribution
Figure (9-a)



Johnson SB Distribution
Figure (9-b)



Gen. Extr. Value Distribution
Figure (9-c)

It is clear that there is a big difference in the probability distributions of these three sectors.

Now we will carry out a detailed analysis of one of these groups (the second element will be from the sample (group 5)) in order to present a statistical analysis for the rest of the groups. Then we will summarize the results of these groups because of the too many pages this will take and the similar statistical behavior of these groups.

2- Statistical Analysis of the Second Element of the Sample (Group5):

When analyzing the general trend of the data of this group, we find the following representation:

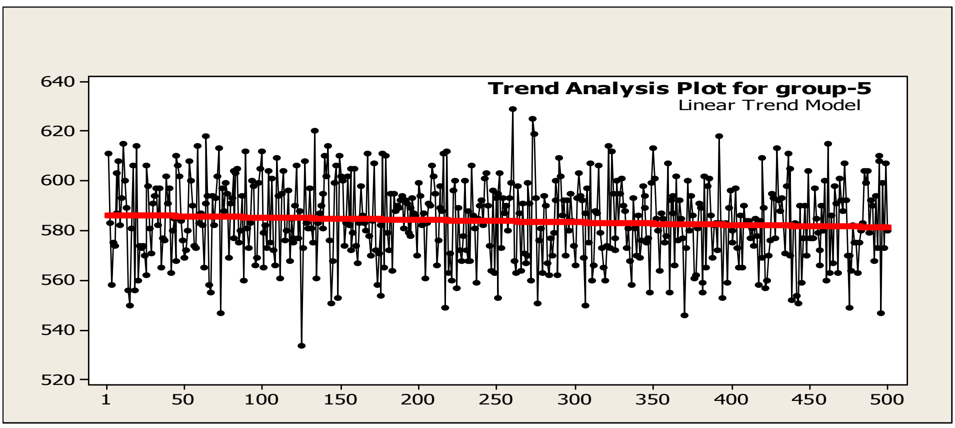


Figure (10)

We notice that the general trend of the data of this group has a general linear equation which is $Y(t) = 586.32 - 0.0099\,t$ with a negative inclination and a very small value equal to 0.0099. If we assume that the data have generative random processes, we can look at the mean function of these random processes as constantly significant because of the small value of the trend (regardless of its significance). We also notice that the significance of the variance function is (almost) constant too, and this gives us an impression that the generative random processes of these data are stationary random processes (Ref. [4], [6]], [10]).
Now when we examining the autocorrelation function (Ref. [9]) of the data of this group (at a level of significance $\alpha = 0.05$) we find the following representation:

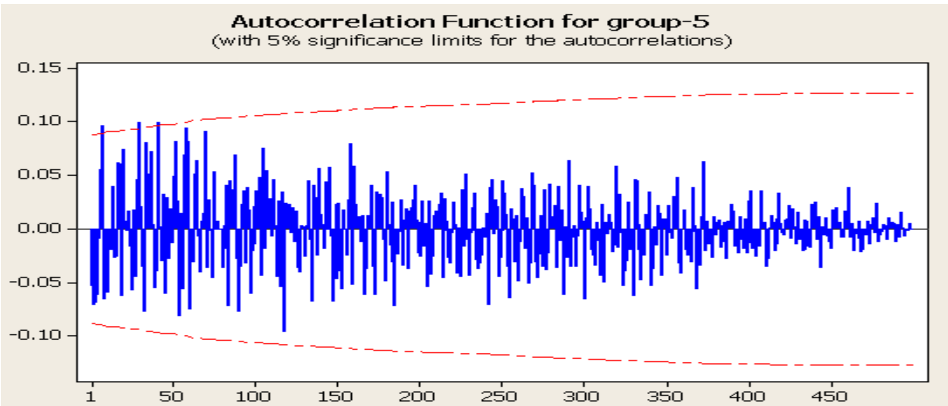


Figure (11)

Here we notice that the autocorrelation function has a decreasing formula with the increasing lag between the values. This proves that the generative random processes of the data of this group behave like stationary random processes.
Incoming diagrams show the histogram which is close to the normal distribution (Figure (12-a)) and the best probability distribution function (Figure (12-b)) of the data of this group according to Kolmogorov-Smirnov Test at a significance level $\alpha = 0.05$.

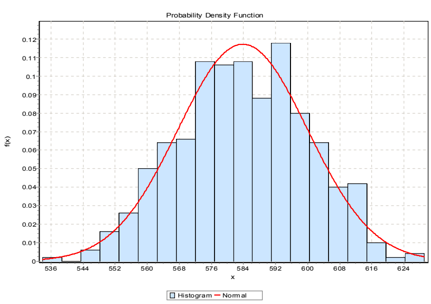


Figure (12-a) normal distribution

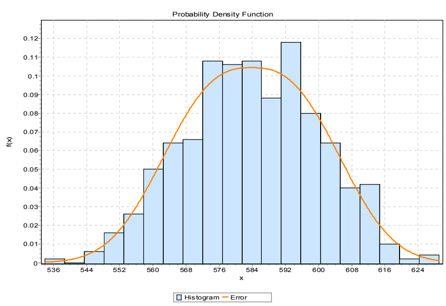


Figure (12-b) error distribution

The incoming tables show some characteristic numerical values (table (2-a)) and the confidence intervals of the mean and median at a significance level $\alpha = 0.05$ and the fitted trend equation Table (2-b) of data of this group.

Table (2-a)

Min	Max	Median	Mean	Std.Dev	Coef. Of Variation
534	629	583	583.83	16.163	0.02768

Table (2-b)

95% C.I for Median	95% C.I for Mean	Fitted Trend Equation
582.41 , 585.25	58200 , 586.00	Y(t) = 586.32 - 0.009941*t

We also notice a slight decrease in the value of the standard of deviation of the data compared to the value that preceded it.

Some numerical measures of all groups

According to a similar approach to the previous one, we find the following results for the groups under investigation:

Table (3)

S.N.	Group	Min	Max	Median	Mean	Std.Dev	Coef. Of Variation
1	1	618	1229	680	697.02	19.577	0.10147
2	5	534	629	583	583.83	16.163	0.02768
3	12	506	608	560	559.85	14.979	0.02676
4	19	502	593	546	545.61	15.214	0.02788
5	26	495	583	536	536.1	15.087	0.02814
6	33	486	576	528	529.02	15.222	0.02877
7	40	476	569	523	522.72	14.979	0.02866
8	47	470	563	518	518.87	15.335	0.02955
9	54	471	555	516	515.21	15.74	0.03055
10	61	467	561	512	512.22	15.307	0.02988
11	68	468	554	509	508.91	14.56	0.02861
12	75	465	559	506	506.71	15.412	0.03042
13	82	462	551	505	504.29	14.921	0.02959
14	89	463	551	503	502.95	15.493	0.03081
15	96	452	551	501	501.17	16.158	0.03224
16	103	458	540	498	497.76	15.364	0.03087
17	110	457	537	496.5	497.39	15.245	0.03065

18	117	453	540	495	495.36	15.373	0.03103
19	124	446	538	494	493.44	15.387	0.03118
20	131	445	539	493	492.79	14.809	0.03005
21	138	437	532	492	492.03	15.047	0.03058
22	145	447	539	491	490.91	14.68	0.0299
23	152	444	534	489	489.24	14.774	0.0302
24	159	446	528	487.5	487.74	15.066	0.03089
25	166	444	528	486.5	487.03	15.6	0.03203
26	173	448	525	486	485.99	15.045	0.03096
27	180	437	523	485	484.5	14.936	0.03083
28	187	443	531	485	484.46	15.497	0.03199
29	194	431	522	483	483.28	15.465	0.032
30	200	446	524	483	482.86	15.129	0.03133

Confidence intervals and trend of all groups

The confidence intervals of mean and median value and the fitted trend equations of these groups(including the first one) presented in the following table:

Table (4)

S.N.	Group	95% C.I for Median	95% C.I for Mean	Fitted Trend Equation Y(t)=
1	1	691.77 , 702.26	676.00 , 685.39	(10^4)/(15.5962-3.40488(0.994474^t))
2	5	582.41 , 585.25	58200 , 586.00	586.32 - 0.009941*t
3	12	558.53 , 561.17	558.00 , 561.00	562.06 - 0.008821*t
4	19	544.27 , 546.94	544.00 , 547.00	546.43 - 0.003264*t
5	26	534.77 , 537.42	534.00 , 537.00	536.79 - 0.002773*t
6	33	527.69 , 530.36	527.00 , 531.00	529.38 - 0.001439*t
7	40	521.40 , 524.03	522.00 , 525.00	522.12 + 0.00239*t
8	47	517.52 , 520.21	517.00 , 520.00	519.65 - 0.003123*t
9	54	513.82 , 516.59	513.61 , 51800	516.63 - 0.005694*t
10	61	510.87 , 513.56	511.00 , 514.00	512.26 - 0.000163*t
11	68	507.63 , 510.19	508.00 , 511.00	509.11 - 0.000813*t
12	75	505.35 , 508.06	505.00 , 507.00	504.91 + 0.00718*t
13	82	502.97 , 505.60	503.61 , 506.39	504.96 - 0.002700*t
14	89	501.59 , 504.31	501.00 , 505.00	504.07 - 0.004478*t
15	96	499.75 , 502.59	499.00 , 502.39	500.32 + 0.00341*t
16	103	496.41 , 499.11	495.00 , 500.00	497.95 - 0.000741*t

17	110	496.05 , 498.73	495.00 , 499.00	497.29 + 0.000396*t
18	117	494.01 , 496.71	494.00 , 497.00	496.00 - 0.002525*t
19	124	492.09 , 494.80	493.00 , 495.00	492.93 + 0.00204*t
20	131	491.49 , 494.10	491.00 , 495.00	493.74 - 0.003794*t
21	138	490.71 , 493.35	490.00 , 493.00	492.33 - 0.001207*t
22	145	489.62 , 492.20	489.00 , 492.00	491.18 - 0.001082*t
23	152	487.94 , 490.53	488.00 , 491.00	490.97 - 0.006903*t
24	159	486.42 , 489.07	486.00 , 489.00	488.86 - 0.004447*t
25	166	485.66 , 488.40	485.00 , 488.00	487.44 - 0.001640*t
26	173	484.66 , 487.31	484.00 , 487.00	485.03 + 0.00381*t
27	180	483.18 , 485.81	483.00 , 486.00	484.70 - 0.000797*t
28	187	483.10 , 485.82	483.00 , 486.00	484.61 - 0.000615*t
29	194	481.93 , 484.64	481.00 , 485.00	482.97 + 0.00126*t
30	200	481.53 , 484.19	481.00 , 484.39	484.25 - 0.005577*t

Probability distributions of all groups

The following table shows the probability distributions of all groups including the first one.

Table (5)

S.N.	Group	Distribution	Parameters
1	1	Burr	k=0.16815 α=83.24 β=648.78
2	5	error	$k = 2.65$ $\sigma = 16.16$ $\mu = 583.83$
3	12	Lognormal	σ=0.02676 μ=6.3273
4	19	Log-Logistic (3P)	α=232.41 β=2006.5 γ=-1461.0
5	26	Gamma (3P)	α=114.56 β=1.4102 γ=374.54
6	33	Dagum (4P)	k=0.86434 α=17.875 β=148.44 γ=381.92
7	40	Burr	k=1.4847 α=54.763 β=528.18
8	47	Burr	k=1.0019 α=59.974 β=518.61
9	54	Weibull (3P)	α=4.0509 β=62.397 γ=458.62
10	61	Gen. Extreme Value	k=-0.29522 σ=15.32 μ=506.94
11	68	Kumaraswamy	α_1=3.8215 α_2=214.6 a=458.77 b=684.99
12	75	Gen. Extreme Value	k=-0.22097 σ=14.863 μ=500.85
13	82	Burr	k=1.3372 α=54.447 β=508.18
14	89	Error	k=2.5517 σ=15.493 μ=502.95
15	96	Erlang	m=962 β=0.52096
16	103	Pert	m=497.15 a=458.0 b=540.0
17	110	Gen. Extreme Value	k=-0.23548 σ=14.921 μ=491.66

18	117	Burr	k=1.1206 α=54.642 β=496.7
19	124	Weibull (3P)	α=4.2035 β=63.803 γ=435.38
20	131	Error	k=2.1614 σ=14.809 μ=492.79
21	138	Lognormal (3P)	σ=0.02986 μ=6.2184 γ=-10.079
22	145	Johnson SU	γ=0.41613 δ=4.0581 λ=57.461 ξ=496.99
23	152	Normal	σ=14.774 μ=489.24
24	159	Gen. Extreme Value	k=-0.26209 σ=14.925 μ=482.27
25	166	Gen. Extreme Value	k=-0.27991 σ=15.57 μ=481.51
26	173	Pert	m=485.73 a=448.0 b=525.0
27	180	Inv. Gaussian	λ=5.0980E+5 μ=484.5
28	187	Nakagami	m=244.37 Ω=2.3494E+5
29	194	Inv. Gaussian (3P)	λ=8.9582E+7 μ=2787.7 γ=-2304.4
30	200	Nakagami	m=254.49 Ω=2.3338E+5

Here we notice sixteen different probability distributions, and when the probability distribution is repeated, the parameters are different.

When we examining the autocorrelation function of the data of the groups 2 until 30 at a level of significance $\alpha = 0.05$ we find a similar representation as Figure (11), and this proves that the generative random processes of the data of these groups behave like also stationary random processes.

Diagrams for the min., max., mean and median of all groups

The following diagram shows the minimum, maximum, mean and median values of these groups:

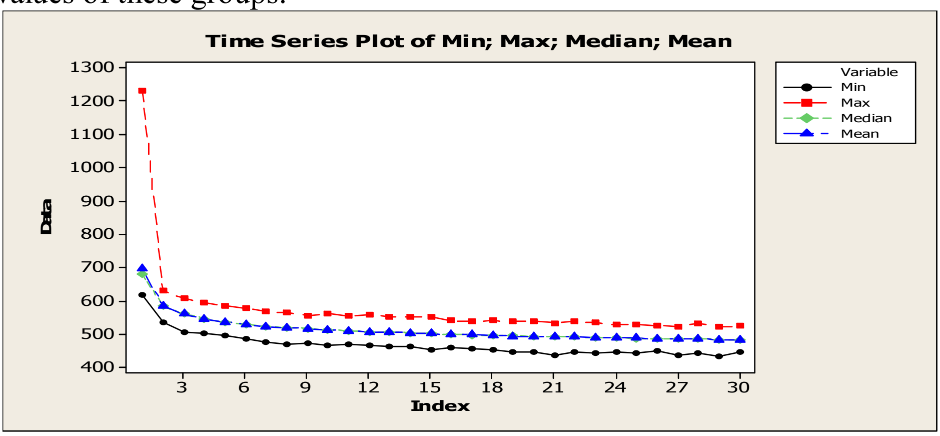


Figure (15)

Without the first group we find the following:

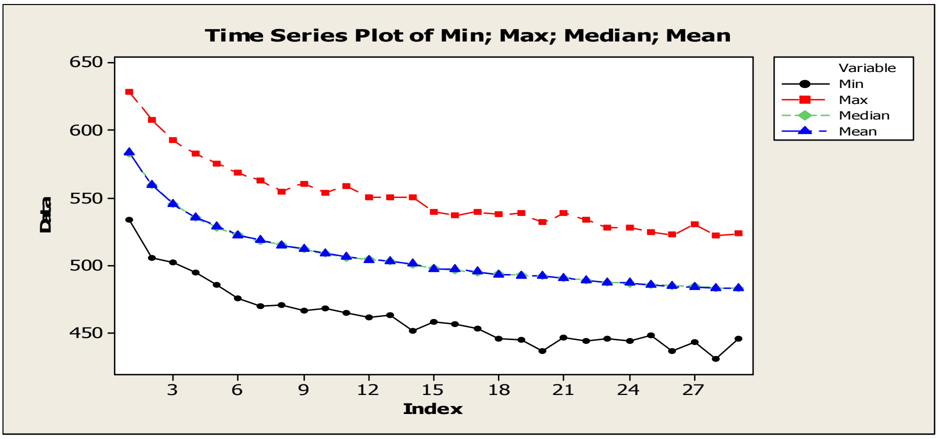


Figure (16)

We notice that the median values tend to have a constant linear path which is almost nonexistent. This is consistent with the results arrived at previously that came to the conclusion that the generative random processes of these data are constantly significant.

By using the regression analysis to show the relationship between the median and mean values, we find that the regression equation is as follows:

$\textbf{median} = 30.9 + 0.939\ \textbf{mean}$ for all groups

$\textbf{median} = 2.04 + 0.996\ \textbf{mean}$ for all without the first group

Conclusion

What has been previously stated, it can be concluded that it is very hard to find a general test formula to specify a count of prime numbers at intervals of the same length because there is a variety of probability distributions that are consistent with these groups and a change in the in the path of the trend of the groups under investigation.

King Saud University, Teachers College, Department of Mathematics

Acknowledgment.

The authors extend their appreciation to the Research center of Teacher college, King Seud University for funding this work through the research group project No RGP-TCR-30 .

References

1	Ares S., Castro M.; Hidden structure in the randomness of the prime number sequence , (2005),arXiv:cond-mat/0310148.
2	James D. Hamilton; Time Series Analysis, Published by Princeton University Press. United Kingdom. ; 1994 .
3	Jia C. H.; Almost all short intervals containing prime numbers, Acta Arith. 76 (1996), 21–84.
4	Krishnan V.; Probability and random processes Published by John Wiley & Sons, Inc., Hoboken, New Jersey 2006.
5	LARSEN R.J., MARX M.L; An introduction to Mathematical Statistics and its Applications– Prentice-Hall International, Inc. USA. 1993.
6	Leonid B. Yakov K., Sinai G.; Theory of Probability and Random Processes - Second Edition .Springer-Verlag Berlin Heidelberg 2007.
7	Liu H. Q.; Almost primes in short intervals, J. Number Theory 57 (1996), 303–322.
8	Rahal M.; A logarithms for finding prime numbers by smaller reduplication method, Aleppo university journal for the basic sciences series (2007), vol. 56 . In Arabic
9	Robert H. David S., Stoffer S.; Time Series Analysis -and Its Applications Springer Science and Business Media - LLC, 2006.
10	Scott L. Donland M., Childers G.; Probability and random processes Published in the United States by Elsevier Academic Press. 2004.
11	Van der Galiën J.G.; Are The Prime Numbers Randomly Distributed ? OPAS Journal of Mathematics, 1.2. (2002).
12	Vaughan R. C.; Mean value theorems in prime number theory, J. London Math. Soc. (2) 10 (1975), 153–162.