CS240 Exploratory Data Analysis Project

Name : Muhammed Can Özdemir

Student Number : 211590856

In this Project i analysing that after the year of 2000, is number of Error decreasing in year by year ?

To start with, i look at the data which after year of 2000, i saw that between 2006-2009, number of Error average decreasing year on year.

In year of **2006**, mean is **1.74**

In year of **2007**, mean is **1.67**

In year of **2008**, mean is **1.66**

In year of **2009**, mean is **1.62**

Also, The total error mean of **after year of 2000 is 1.70.**

We can see that, **in 2009**, error mean is the reach the least with **1.62**.

On the other hand, one year after 2010, error mean is dramatically increasing.

The error mean of **2010 is 1.73. So, after the 2000, some years, there are some important decreasing of error mean such as between 2006 and 2009, but it is not always happening. So, we can't say that** every following years, error mean is decreasing. We can see the 2010 example…

In the following pictures, we can see that mean of errors

(year of after 2000, 2006,2007,2008,2009,2010)

```
yearof2007.describe()
```

```
count    1782.000000
mean        1.677890
std         3.396427
min         0.000000
25%         0.000000
50%         0.000000
75%         2.000000
max        26.000000
Name: E, dtype: float64
```

```
yearof2006.describe()
```

```
count    1756.000000
mean        1.746583
std         3.522838
min         0.000000
25%         0.000000
50%         0.000000
75%         2.000000
max        28.000000
Name: E, dtype: float64
```

```
yearof2009.describe()
```

```
count    1763.000000
mean        1.620533
std         3.049640
min         0.000000
25%         0.000000
50%         0.000000
75%         2.000000
max        23.000000
Name: E, dtype: float64
```

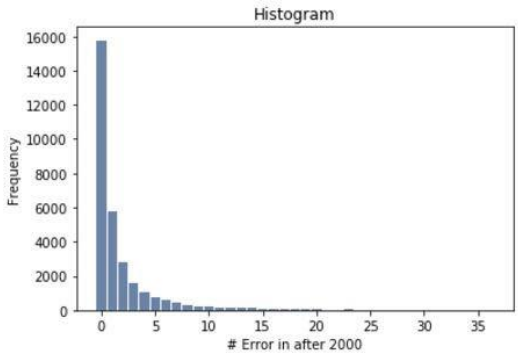
```
yearof2008.describe()
```

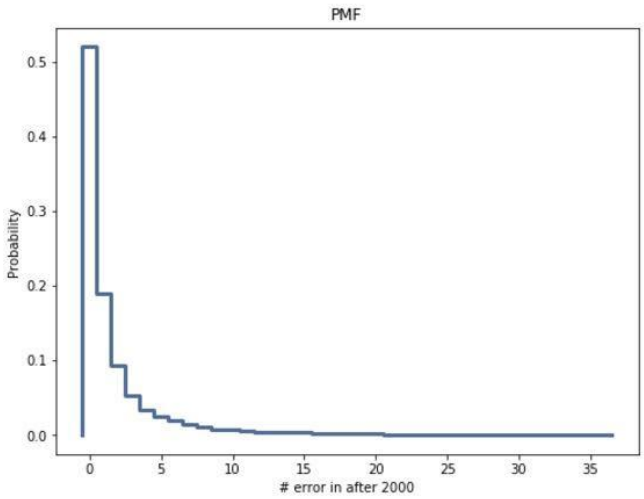
```
count    1776.000000
mean        1.669482
std         3.190919
min         0.000000
25%         0.000000
50%         0.000000
75%         2.000000
max        34.000000
Name: E, dtype: float64
```

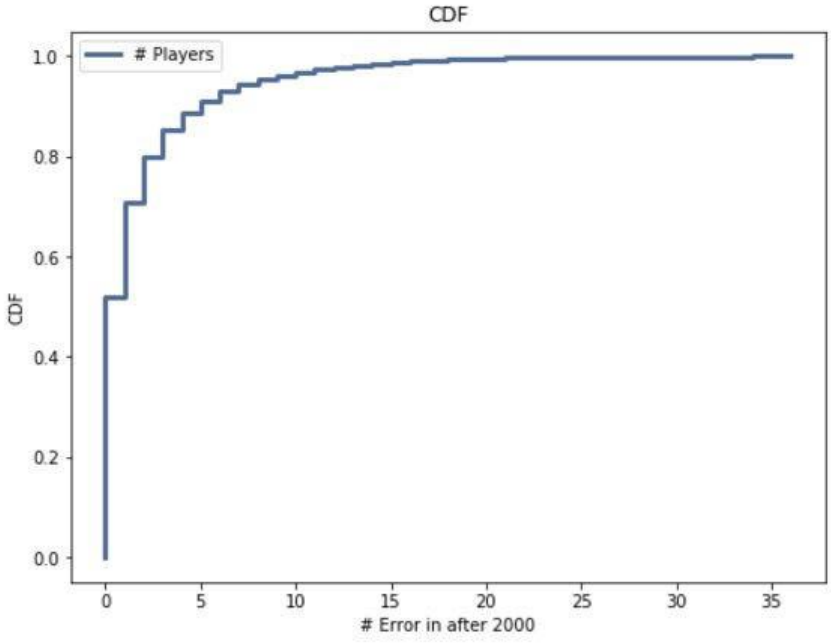
```
yearof2010.describe()
```

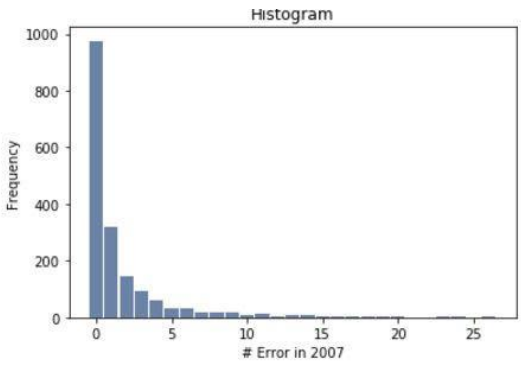
```
count    1751.000000
mean        1.730440
std         3.319791
min         0.000000
25%         0.000000
50%         0.000000
75%         2.000000
max        34.000000
Name: E, dtype: float64
```

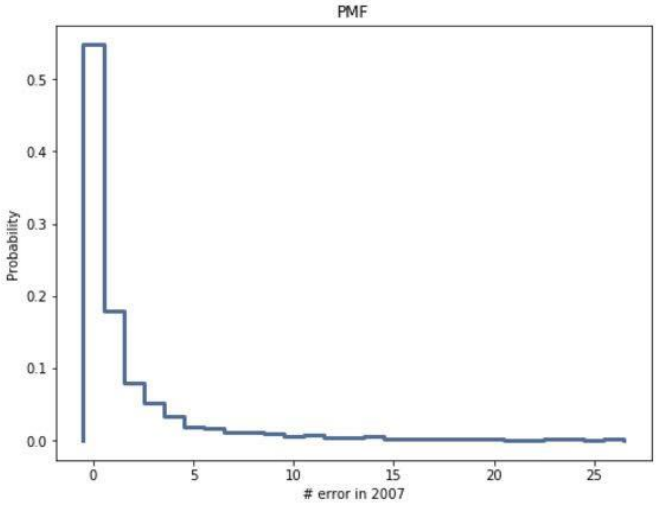
Analysis of Histogram, Pmf, Cdf

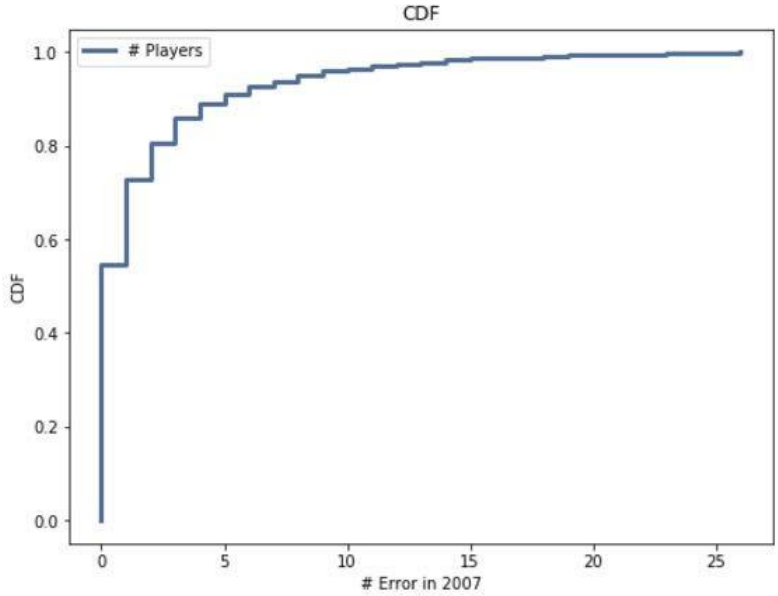


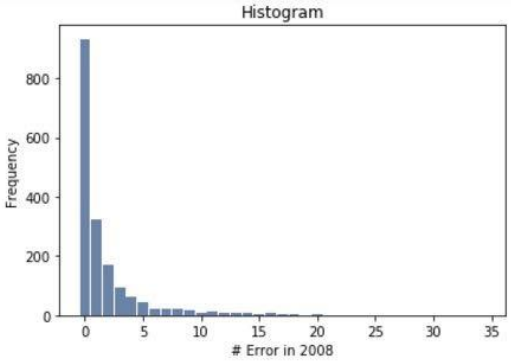


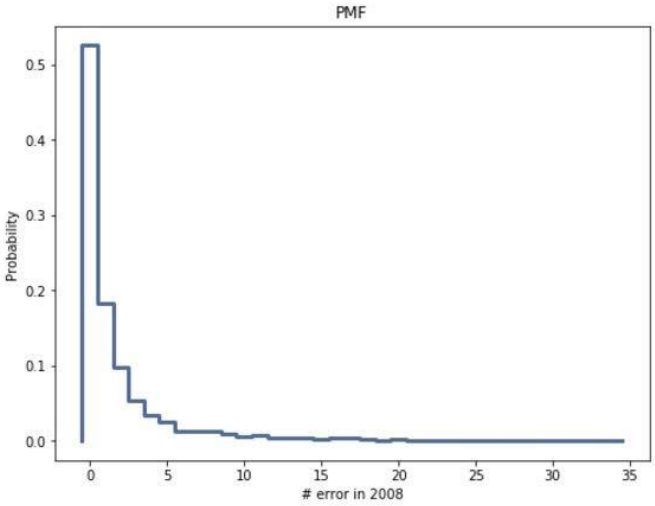


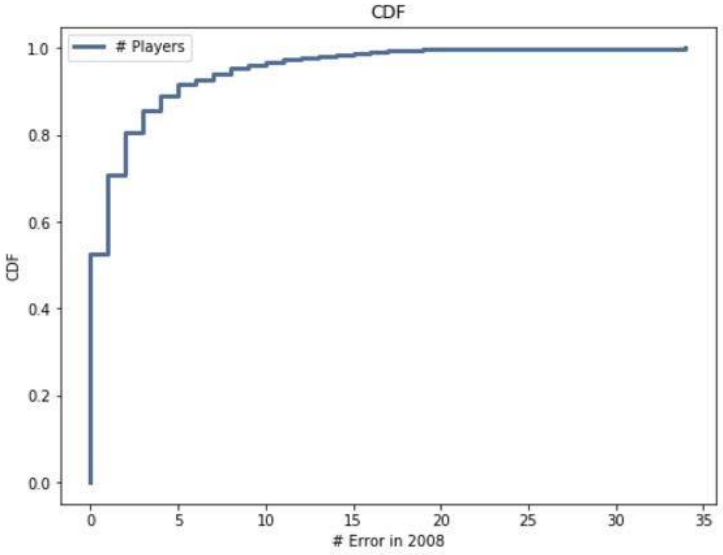

Histogram
Frequency
Error in 2007

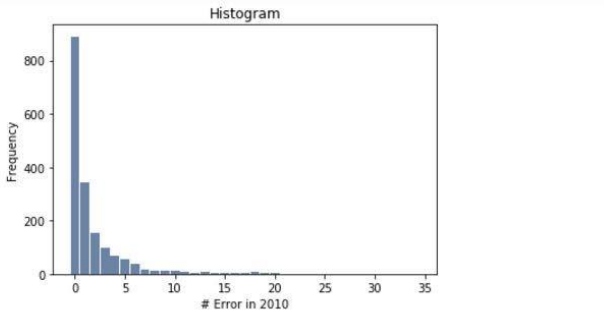

PMF
Probability
error in 2007

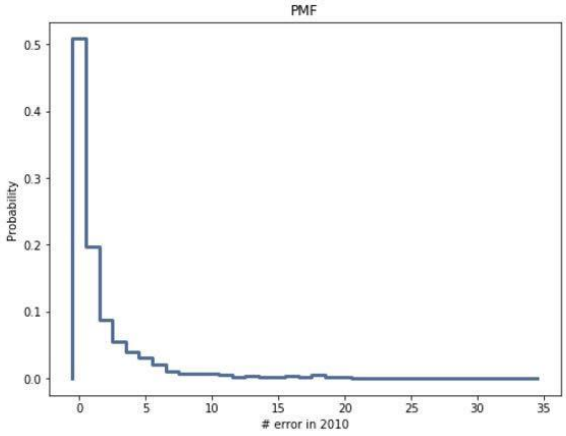

CDF
Players
CDF
Error in 2007

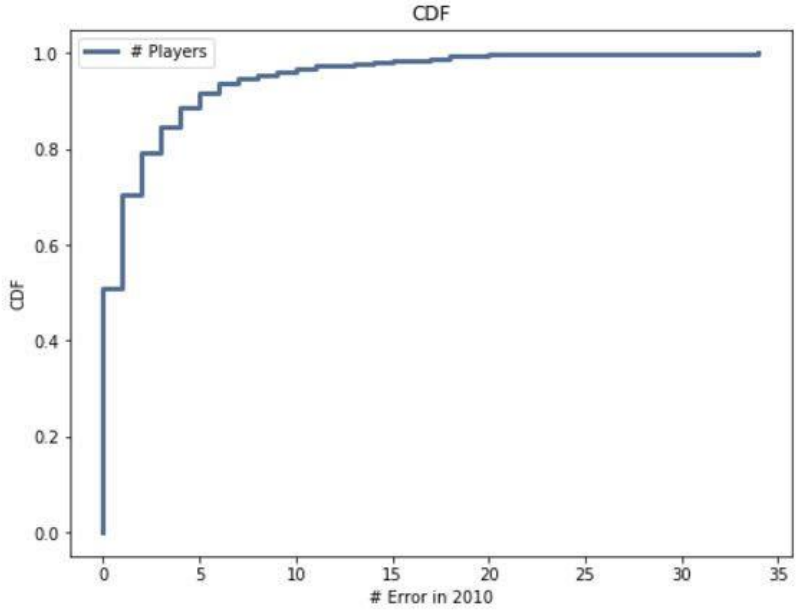

Histogram
Frequency
800
600
400
200
0
0
5
10
15
20
25
30
35
Error in 2008


PMF
Probability
0.5
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
25
30
35
error in 2008


CDF
Players
CDF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
35
Error in 2008

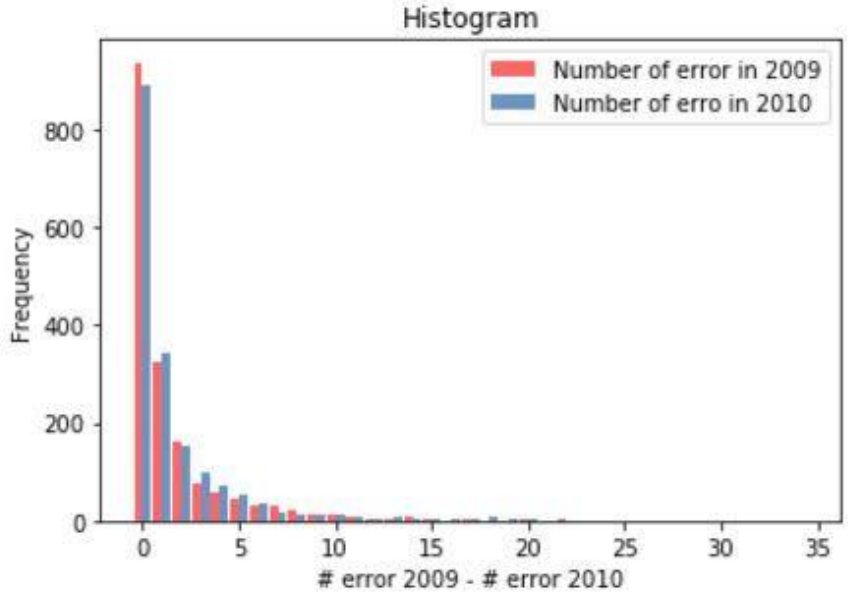

Histogram
Frequency
Error in 2010


PMF
Probability
error in 2010


CDF
Players
CDF
Error in 2010

Firstly, i analysing that all years after the years pmf, cdf and histogram.
When look at the histogram, pmf and cdf, i can see that %55-%60 play wiyh 0 errors. And, between 0-5 errors is about %10-20.
Moreover, Between, 5-35 erros count have almost same percentage with apporximately %10. There may lots of reasong why more than %50 player have 0 erros. For example, may did not play all season.

When we are look at years of 2006-2007-2008-2009,
We can see that frequency, probability and cdf rates not dramatically diffrent. It just small changes between these years.
For example, when i compare error frequency of 2006 to 2007, in 2007 frequency of zero error is increasing.
Thats why, error mean is decreasing. Because when zero error count is increasing, mean of error is decreasing.
Also, when i look at the 2010, frequency of zero error and probability of the zero error, i can see that it was dramatically decreasing. Thats why, in 2010, mean of error is increasing.
So, i can also make this argument, if zero error count is increasing, mean of error is decreasing for each year.



Correlation

In this Project, i analyse that relationship between years and error mean.

So, i look at the correlation between years and error, it gives me **minus value** that mean so i understand that mostly there is **a inverse proportion** between years and erros.

Hypothesis

In hyptohesis test, test statistic is after year of 2000, number of error is decreasing in year by year.

My null hyptohesis is number of error is increasing in year by year.

Conclusion

In this Project, i compare that after year of 2000, number of error is decreasing or increasing in year by year. I found out that In the some period such as between 2006-2009, number of error is decreasing in year by year. The most important reason fort hat there are lots of player who have zero error, when zero error count is increase, number of error is decreasing.

Also, some years, this trend (number of error increasing) is broken. For example, we can see that in 2000, number of error increase.

So, after year of 2000, number of error is not always decreasing in year by year.