

TED ANKARA COLLEGE FOUNDATION HIGH SCHOOL

IB EXTENDED ESSAY

EFFECTS OF STATISTICS ON PROBABILITY

Candidate name: Yusuf Can OKŞAR

Candidate number: 1129-0068

Supervisor's name: M. Levend DEMİRBAŞ

ABSTRACT

My work about statistics and probability. I realize that in some cases we can guess events by looking statistics. But our guess is a probability which tells how will be an events. This shows statistical results can also effect probibility. I started my study firstly giving information about statistics and probability. To give these informations firstly I made a research from books and internet. Working on statistics and probability could reveal their relation easily. I saw that in some situations probability is directly propotional to the statistics. For example we can guess a football match's result with looking early matches' results. And I found a table which shows result of challenges between Fenerbahçe and Galatasaray. These two teams' new match can be guess by looking this table. But I agree with that there are infinite number of factors that can effect the result of match. So we can not find the absolute value of winning chance of a team. But as we increase the factors we thought we get close the real probability. This study also shows every event is not dependent to statistical results. Sometimes events can be independent from statistics and we can not guess their probability by looking statistics. Becuse these events has constant probabilities. For example throwing coin, tossing die... etc.

CONTENTS

PREFACE	2
I. STATISTICS	
A. Definition and Types of Statistics	2
1. Definition of Statistics	
2. Types of Statistics	
a. Mathematical Statistics	3
b. Practical Statistics	3
B. Data Collection and Processing	4
1. Data Collection	4
2. Organizing the Data	5
3. Presentation of Data	5
II. PROBABILITY	8
A. Permutation and Combination	9
B. Random Variables	14
III. EFFECTS OF STATISTICS ON PROBABILITY	16
A. Relation Between Statistics and Probability	16
CONCLUSION	18
BIBLIOGRAPHY	19

PREFACE

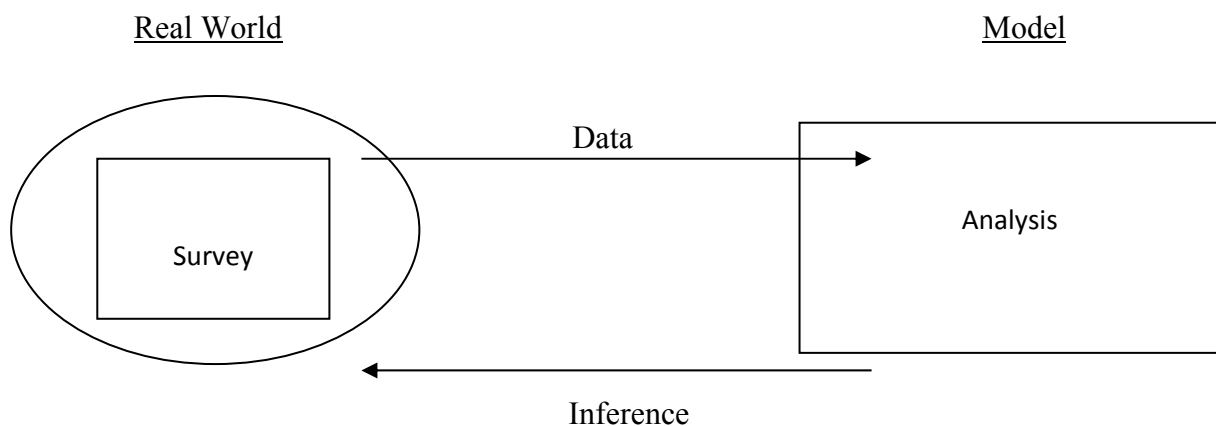
When I was watching a football match I thought about how can I guess this match's result. Then I thought that there can be three different result at the end of the match. Then I said that the winning probability for each team and also draw is $1/3$. Then I thought early matches I saw. My thinking way was very wrong I think. Because early matches' statistics should directly effect this match's result. But if I were throwing a coin this thinking way will work. This encourage me to work on this subject. Firstly I will get informations about statistics and probability to understand their relation more. I bought few books about probability and statistics. Also I search about this subject on internet. In my work firstly I give informations and examples about statistics and probability. Then I will examine the effect of statistics on probability and in what cases statistics can effect probability.

STATISTICS

Definition and Types of Statistics

Statistics is the study of the collection, organization, analysis, interpretation and presentation of data. It deals with all aspects of data, including the planning of data collection in terms of the design of surveys and experiments.¹In statistics datas can be quantity or qualification.

Statistics establishes a model about random events, process and systems.



¹ Dodge, Y. (2006) *The Oxford Dictionary of Statistical Terms*, OUP.

There are few types of statistics ;

- Mathematical Statistics
- Practical Statistics

Mathematical Statistics

Mathematical statistics is the study of statistics from a mathematical standpoint, using probability theory as well as other branches of mathematics such as linear algebra and analysis.

Mathematical statistics is related to statistical theory which includes study design and data analysis.

Practical Statistics

- **Descriptive Statistics**

If there is huge amount of data, there should be used graph, table ... (etc.) to deduce from datas. This method named as descriptive statistics. For example students' marks from a lesson are our datas. Each student is an element of datas where as each mark is an observation. If there is a lot of student in class then there will be many datas. It is difficult to deduce from these datas. So we can make a graph or table in order to comment datas easily.

- **Inferential Statistics**

In statistics, group of all elements specify a stack. Datas which selected random from this stack are samples. Sample is a subset of stack.

In inferential statistics, we look at samples and make comments about all datas (stack). Deducing and decisions create a important part of statistics. For example randomly selected 2000 student from few university generate sample. We can make comments about university students' life by looking only this sample.

If sample selection is random this means all datas have same chance in choice. But this selection can be in two type. We can put again a selected data or we do not put.

Data Collection and Processing

Data Collection

Data is a result that observed by a observer. Data can be numeral or not.

Data can be observed by;

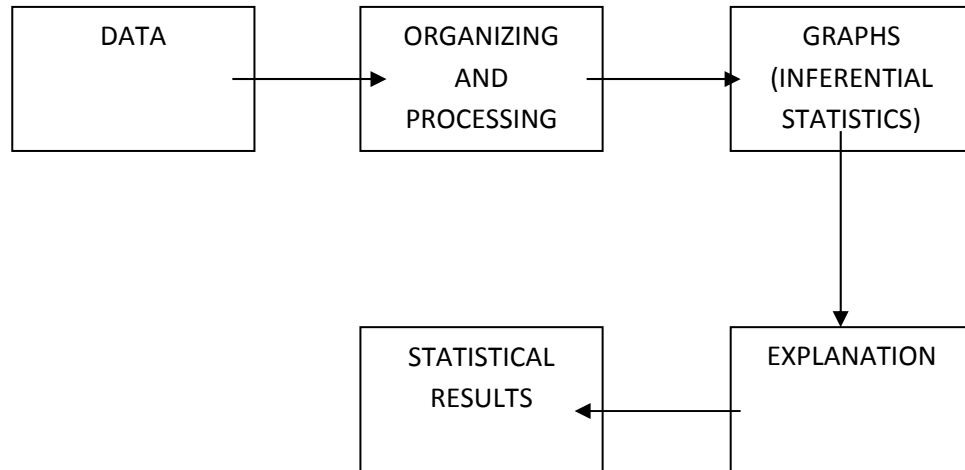
- Published sources
- A designed trial
- Survey results
- Observation results

In statistics observation of data named as data collection. For example if we learn about students' size in a specific class at a specific time this will be a data collection.

Data collection is the one of the most important parts of statistics. So it is very important that choosing the most appropriate data collection method (counting all datas or selection of sample).

Organizing the Data

After data collection next step in statistical process is organizing the data. Tables, graphs or lines can be used for this. Organizing data makes easier using the data.²



PRESENTATION OF DATA

Data can be presented by table or graph. Graph is a better way to present data because it is more clear than table.

There is few types of graphs to present data;

- Histogram
- Column graph
- Line graph
- Pie chart

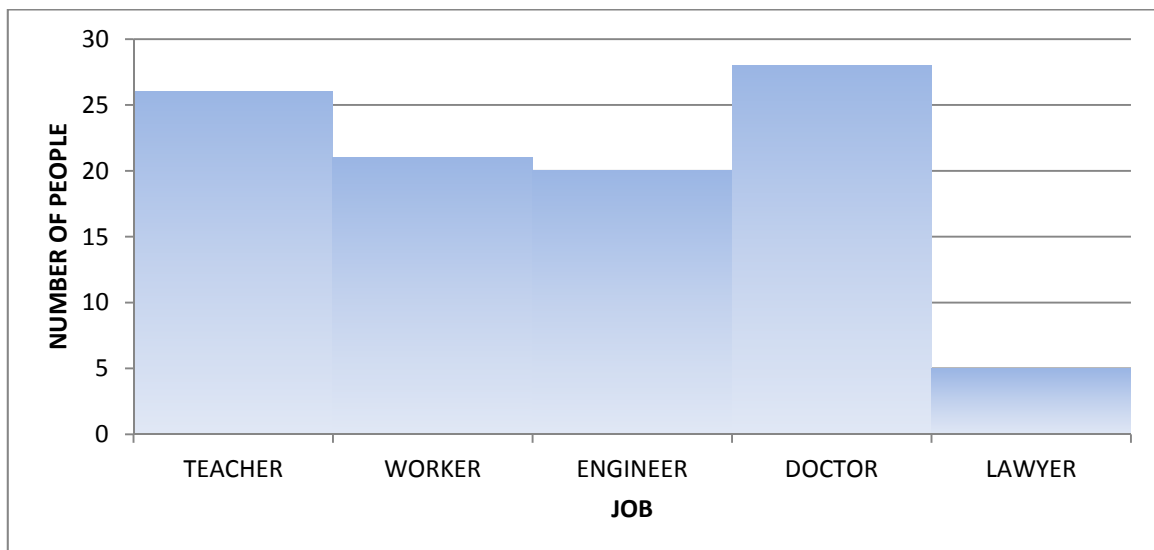
² AKDENİZ, Fikri (2012), *Olasılık ve İstatistik*, 17.B., Nobel Kitabevi y., Adana, 2012.

EXAMPLE OF GRAPHS

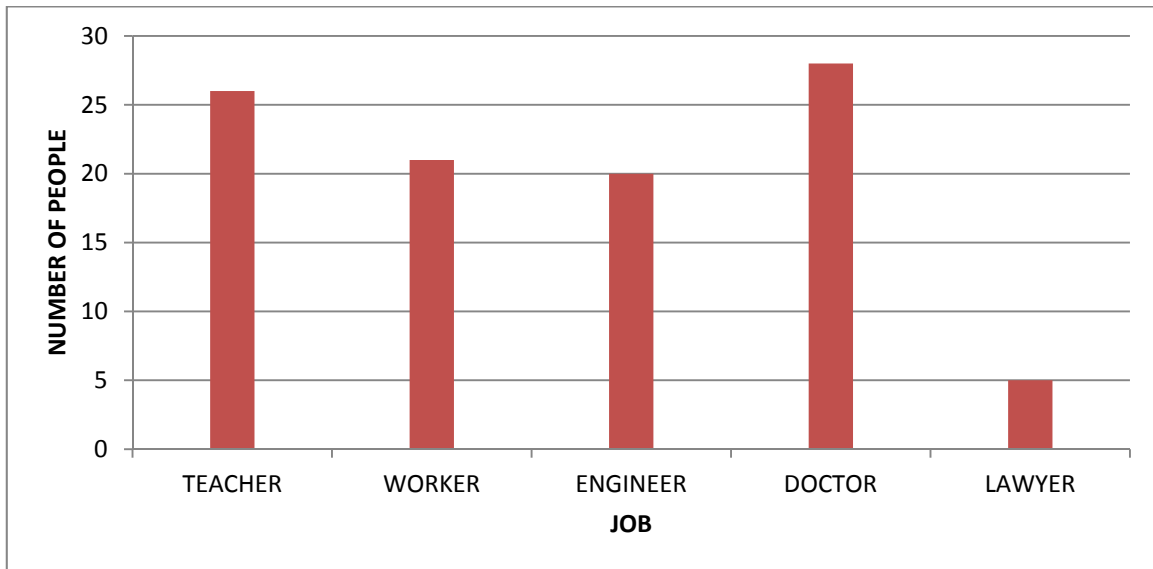
JOB	NUMBER OF PEOPLE
Teacher	26
Worker	21
Engineer	20
Doctor	28
Lawyer	5

Table1. Distribution of 100 people according to their jobs.

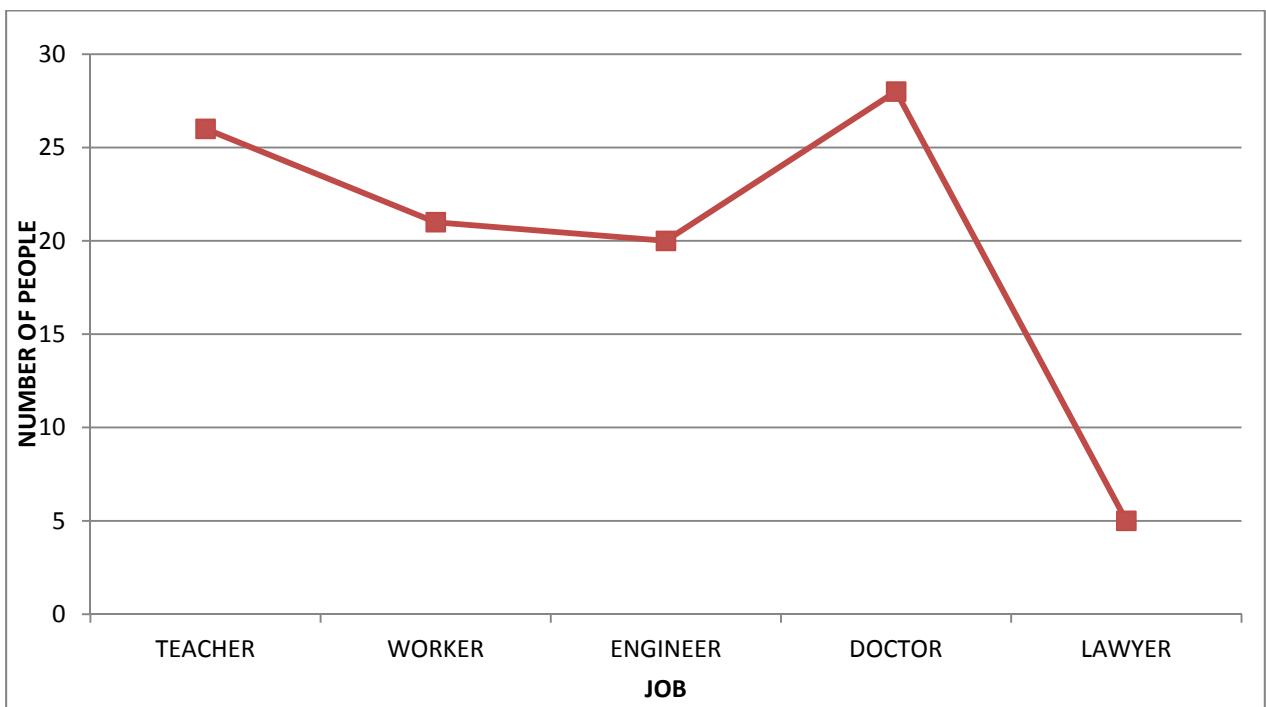
HISTOGRAM



COLUMN GRAPH



LINE GRAPH



Event: Subset of sample space

For example⁴; One die has tossed, set “S” represents sample space of this event.

$$S = \{1,2,3,4,5,6\}$$

There is 2^6 subset (event) of this sample space.

A coin tossed three times, “S” represents sample space of this experiment.

$$S = \{HHH,HHT,HTH,HTT,THH,THT,TTH,TTT\}$$

Where as; H = head

T = tail

To calculate probability of an event we look how many times there are this event in sample space. This means number of intended subset divided to number of elements in sample space.

For example at previous example if we look at possibility of at least one tail; set “A” represents our intended event (subsets),

$A = \{HHT,HTH,HTT,THH,THT,TTH,TTT\}$ then probability of this event is “**7/9**”

PERMUTATION and COMBINATION

Product of positive integers from 1 to n is named as n-factorial (n!).

$$n! = 1.2.3. \dots .(n-1).n = (n-1)! .n$$

$$0! = 1 \text{ and } 1! = 1^5$$

⁴ ÖZTÜRK, Fikri (2011), *Olasılık ve İstatistiğe Giriş I*, 1.B., Gazi Kitabevi y., Ankara, 2011.

With increasing value of n calculation becomes harder. In these cases we use Stirling formule in order to calculate an approximate value.

$$\text{With increasing } n \rightarrow n! = \sqrt{2 \cdot \pi \cdot n} \cdot n^n \cdot e^{-n} \text{ }^6$$

Lining up a part of object or all of it is called as permutation.

For example: In how many ways three book can be put in a bookshelf?

$$3 \quad \times \quad 2 \quad \times \quad 1$$

For first place we have three options in order to replace a book. After we put a book there are two options left. For last place we have only one chance. With multiplication of these options we find our answer to this question. In other words since there are 3 book we can calculate our permutation as 3!.

⁵ AKDENİZ, Fikri (2012), *Olasılık ve İstatistik*, 17.B., Nobel Kitabevi y., Adana, 2012.

⁶ ERDEM, İsmail (2012), *Matematiksel İstatistik- Olasılık- Beklenen Değer- Parametre Tahmini*, 1.B., Seçkin y., Ankara, 2012.

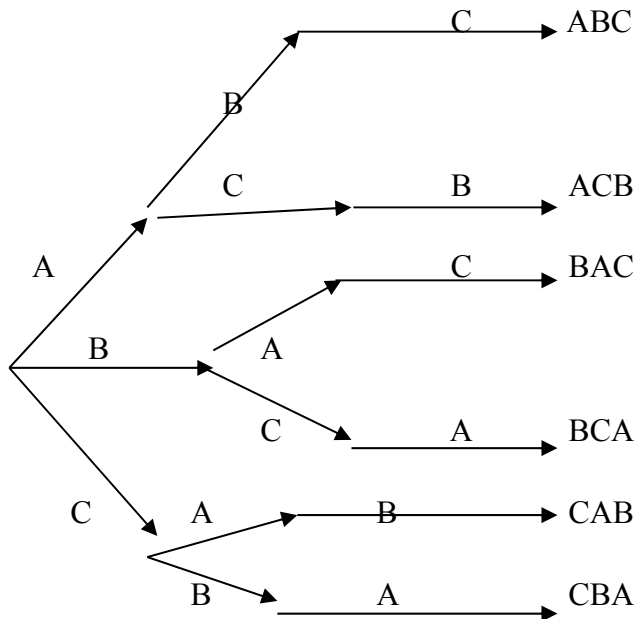


Figure1.⁷

Probability not always current for single events. Sometimes we use permutation and combination to calculate probability of dependent events. These events can be two or more. In order to make easy these calculations we use “tree diagram” as shown as in the figure1.

While we are using all element we are calculating $n!$. This means $P(n,n)$. But in some cases we should not only set these element but also select elements from sample space.

For example : From ten books in how many ways we can put six books in a bookshelf?

First place can be put by ten books. There will be nine choice left for next place. This will be continue like previous example till we put six books to bookshelf.

$$P(10,6) = 10 \times 9 \times 8 \times 7 \times 6 \times 5 = 10! / 4!$$

⁷ AKDENİZ, Fikri (2012), *Olasılık ve İstatistik*, 17.B., Nobel Kitabevi y., Adana, 2012.

Number of permutations where setting of “r” number element from “n” number set is ;

$$P(n,r) = n! / (n-r)!^8$$

Combination include only selection of elements. It does not deal with setting them.

From n element how many different r number element we can chose;

$$C(n,r) \text{ or } \binom{n}{r} = \frac{n!}{r!.(n-r)!}$$

$$C(n,r) = C(n,n-r) = \frac{n!}{r!.(n-r)!}$$

Example⁹: From nine women and six men a group of five people will be selected. What is the probabily of this group will contain three men and two women?

If there is a random choice from these fifteen people we can select C(15,5) different groups. So our sample space is “C(15,5)”. But our intended event is three men and two women. ;

- There are C(6,3) possible men.
- There are C(9,2) possible women.

$$\text{So; } \frac{C(6,3).C(9,2)}{C(15,5)} = \frac{240}{1001}$$

⁸ AKDENİZ, Fikri (2012), *Olasılık ve İstatistik*, 17.B., Nobel Kitabevi y., Adana, 2012.

⁹ ROSS, Sheldon M. (2012), *Olasılık ve İstatistiğe Giriş*, (Çev. Ed.: ÇELEBİOĞLU, Salih-KASAP, Reşat), 4.B., Nobel y., Ankara, 2012.



Picture1.¹⁰

Combinations and permutations are in our lives. For example we use combinations to select which numbers can be used in a lock. There are ten different figure that we can use but we should select 5 number. These numbers can be same too. So in this example we can choose 10^5 different combinations.

COMBINATIONS	PERMUTATIONS					
ABC	ABC	ACB	BAC	BCA	CAB	CBA
ABD	ABD	ADB	BAD	BDA	DAB	DBA
ACD	ACD	ADC	CAD	CDA	DAC	DCA
BCD	BCD	BDC	CBD	CDB	DBC	DCB

Table2¹¹. An example of comparison between combination and permutation

Example: $A = \{1,2,3,4,5\}$;

a-) What is the number of subsets set “A”?

b-) How many subset of “A” contains 1 and 2 ?

¹⁰ http://www.123rf.com/photo_11862583_vector-illustration-of-a-combination-lock-set-with-all-ten-numbers.html

¹¹ AKDENİZ, Fikri (2012), *Olasılık ve İstatistik*, 17.B., Nobel Kitabevi y., Adana, 2012

c-) How many subset of “A” contains 1 or 2 ?

Answer:

a-) subsets of A= $C(5,0) + C(5,1) + C(5,2) + C(5,3) + C(5,4) + C(5,5) = 2^5 = 32$

b-) 1 and 2 will be %100 in these sets. So we can look other elements situation;

$$C(3,0) + C(3,1) + C(3,2) + C(3,3) = 2^3 = 8$$

c-) There are two situation;

$$1-) \text{ There is 1 and there is not 2} = 2^3 = 8$$

$$2-) \text{ There is 2 and there is not 1} = 2^3 = 8$$

So there is $8+8 = 16$ subsets which include 1 or 2.

RANDOM VARIABLES

In probability and statistics, a random variable is a variable whose value is subject to variations due to chance.¹²

¹² Yates, Daniel S.; Moore, David S; Starnes, Daren S. (2003). *The Practice of Statistics* (2nd ed.). New York: Freeman.

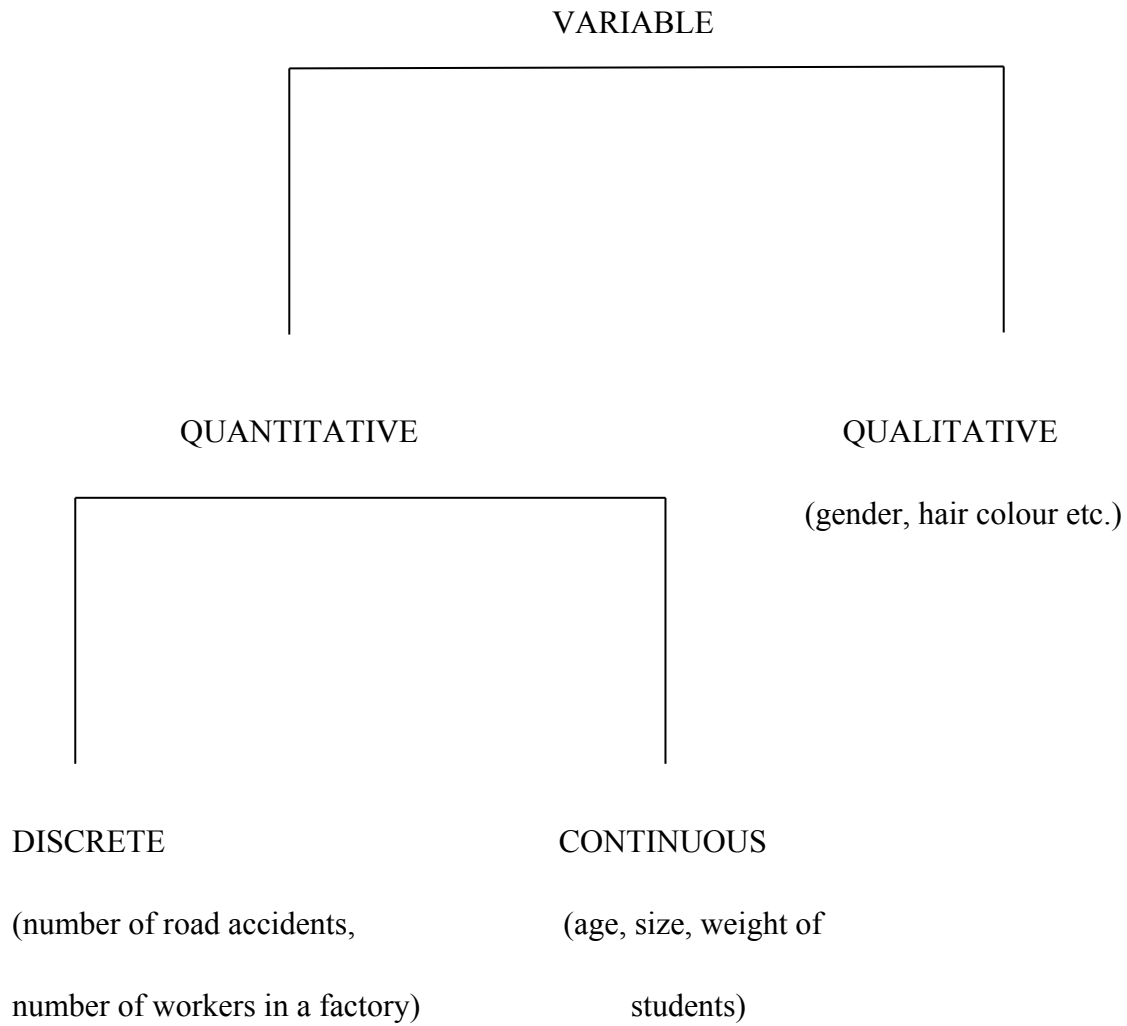


Figure2.¹³

Example¹⁴: While X is a random variable which represents sum of two die;

$$P(X=2) = P\{(1,1)\} = 1/36$$

$$P(X=3) = P\{(1,2), (2,1)\} = 2/36$$

$$P(X=4) = P\{(1,3), (2,2), (3,1)\} = 3/36$$

¹³ ERBAŞ, Semra Oral (2008), *Olasılık ve İstatistik*, 2.B., Gazi Kitabevi y., Ankara, 2008.

¹⁴ ROSS, Sheldon M. (2012), *Olasılık ve İstatistiğe Giriş*, (Çev. Ed.: ÇELEBİOĞLU, Salih-KASAP, Reşat), 4.B., Nobel y., Ankara, 2012.

$$P(X=5) = P\{(1,4), (2,3), (3,2), (4,1)\} = 4/36$$

$$P(X=6) = P\{(1,5), (2,4), (3,3), (4,2), (5,1)\} = 5/36$$

$$P(X=7) = P\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\} = 6/36$$

$$P(X=8) = P\{(2,6), (3,5), (4,4), (5,3), (6,2)\} = 5/36$$

$$P(X=9) = P\{(3,6), (4,5), (5,4), (6,3)\} = 4/36$$

$$P(X=10) = P\{(4,6), (5,5), (6,4)\} = 3/36$$

$$P(X=11) = P\{(5,6), (6,5)\} = 2/36$$

$$P(X=12) = P\{(6,6)\} = 1/36$$

EFFECTS OF STATISTICS ON PROBABILITY

RELATION BETWEEN STATISTICS AND PROBABILITY

Probability, is the measure of how can an event or events can be. We use this measurement in our lifes mostly. For example, “ there will be rain tomorrow” or “you can live more if you do not smoke.”¹⁵. These statements can change according to staitictics of course.

Probability and statistics are in a relation. We can use probability to decide uncertain situations. But while using probability, we take advantage of simple events which form statistics.

For example we know that people who smoke die earlier. Because we can see it in past. These statistics which formed by early people shows us that people who do not smoke live more generally. So we can make a guess like if you do not smoke you can live more.

For example we can think about a football match result. All possible events can effect this match’s result. There are a table that shows the win ratio between football teams Fenerbahçe and Galatasaray.

¹⁵ ERBAŞ, Semra Oral (2008), *Olasılık ve İstatistik*, 2.B., Gazi Kitabevi y., Ankara, 2008.

PLACE	Played	FB (win)	Draw	GS (win)
Fenerbahçe Şükrü Saraçoğlu (1931 – 2008)	81	39	23	19
Ali Sami Yen (1966 – 2008)	45	11	15	19

Table3.¹⁶ (this table shows only matches in these teams' stadiums)

Firstly if we look at simple, there is three possible result in a football match. One Fenerbahçe wins, one Galatasaray wins and other is draw. In this way we can say that all results' probability is 1/3. This is not wrong technically. But if we think about other effective points about match result we will see that actually it is harder to calculate which team will win.

We can say that Fenerbahçe won 50 matches out of all 126 matches. Then there are 38 matches end up with draw and there are 38 matches that Galatasaray won. When we look at at this point next game with 50/126 chance Fenerbahçe will win this match. With 38/126 chance result will be draw and with same probability Galatasaray will win.

If we think about Fenerbahçe and Galatasaray matches this thinking way is correct. But we can increase number of factors that can effect result. With increasing factors probability we will find will be more correct. We know that next match of these two teams in Galatasaray's house. So if we look at only matches in Galatasaray's stadium most probably Galatasaray will win this match. We see that it is not same with we found before. Like this we also add other factors which like weather conditions, match time, etc. All of these factors can effect and change match score.

This shows us the relation between probability and statistics. We need statistics of early matches to guess the result of next match. So as it can be seen statistics can directly effect the probability. This is also valid for weather. To guess how weather will be tomorrow we need the statistics of how weather was in these time early years.

But this is not valid in all cases. I prepared an experiment to show that statistics are not always effect probability. In my experiment I throw a coin fifty times and recorded the results. (head or tail)

¹⁶ <http://www.turkfutbolu.net/fenerbahce/fbgsmac.htm>

NUMBER OF THROWS	HEAD	TAIL
50	39	11

Table⁴

In my experiment as shown in the table I saw “head” part of coin 39 times over 50 throw. If I try to use this statistics to guess the next throw’s result I will find that result of my next throw will be head with 78% chance. But it does not represent the truth. Because while I was throwing coin every trial were independent from early ones. This means if I throw a coin one more time my probability to get head will be $1/2$ not the $39/50$. This is also current for tail. In every throw I have a chance to get head or tail %50.

CONCLUSION

In this study I tried to show the effect of statistics on probability. Firstly I talked about statistics and probability. To connect them each other I should know what they are. When I talked about their properties it was clear that they are in a relation. The first two part of this study contain information about statistics and probability. Then in the third part is the best section to see effect of statistics on probability. In my experiment about Fenerbahçe and Galatasaray it can be shown that match result directly dependent to the early statistics. Because match result can change according to people (football players). So it will be too simple to say that chance of a team’s winning is $1/3$. Thinking about every factor gives us beter result. But there is impossible to find absolute value of probability. But adding every factor bring close us real probability. Then as shown as in my experiment some times probability is independent from statistics. Because the table I did for coins can not help me to guess the result of next throwing. Every trial in this experiment is independent from each other so every chance (head, tail) has a %50 probability. So we can state that the events which dependent human(sports, race), periods or time(weather) can be guess by statistics. But independent events like throwing coin or die have a constant probability which does not depend on statistics.

BIBLIOGRAPHY

- AKDENİZ, Fikri (2012), *Olasılık ve İstatistik*, 17.B., Nobel Kitabevi y., Adana, 2012.
- AYDIN, Hüseyin- SELBES, Hilmi- ÖZER, M. Emin (2012), *Mathematics-11*, 3.B., Turkish Education Association Publications, Ankara, 2012.
- Dodge, Y. (2006) *The Oxford Dictionary of Statistical Terms*, OUP
- ERBAŞ, Semra Oral (2008), *Olasılık ve İstatistik*, 2.B., Gazi Kitabevi y., Ankara, 2008.
- ERDEM, İsmail (2012), *Matematiksel İstatistik- Olasılık- Beklenen Değer- Parametre Tahmini*, 1.B., Seçkin y., Ankara, 2012.
- Feller, W. (1968), *An Introduction to Probability Theory and its Applications* (Volume 1)
- FREUND, John E. (2007), *Matematiksel İstatistik*, (Çev.: ŞENESEN, Ümit), 6.B., Literatür y., İstanbul, 2007.
- "Illustration - Vector Illustration of a Combination Lock Set with All Ten Numbers." *123RF Stock Photos*. N.p., n.d. Web. 24 Feb. 2014. <http://www.123rf.com/photo_11862583_vector-illustration-of-a-combination-lock-set-with-all-ten-numbers.html>.
- ÖZTÜRK, Fikri (2011), *Olasılık ve İstatistiğe Giriş I*, 1.B., Gazi Kitabevi y., Ankara, 2011.
- ROSS, Sheldon M. (2012), *Olasılık ve İstatistiğe Giriş*, (Çev. Ed.: ÇELEBİOĞLU, Salih-KASAP, Reşat), 4.B., Nobel y., Ankara, 2012.
- Yates, Daniel S.; Moore, David S; Starnes, Daren S. (2003). *The Practice of Statistics* (2nd ed.). New York: Freeman.
- YILDIZ, Ekrem (2012), *İstatistik- Eğilim ve Dağılım Ölçüleri- İndeksler- Korelasyon*, 3.B., Seçkin y., Ankara, 2012.

