Research Methodology
Statistics 2

Maha Omair

Assistant Professor

Department of Statistics, College of Science

King Saud University

Why we need statistical data analysis?

Investigations in diverse fields like agriculture, medicine, physics, biology, chemistry etc. require collection of “observations”. Observations are almost always subject to random error. Hence statistical methods have to be employed to collect as well as to analyze the data.

Statistical data analysis

Studying a problem through the use of statistical data analysis usually involves four basic steps:

1. Defining the problem.
2. Collecting the data.
3. Analyzing the data.
4. Conclusions and recommendations.

Defining the problem

An exact definition of the problem is imperative in order to obtain accurate data about it. It is extremely difficult to gather data without a clear definition of the problem.

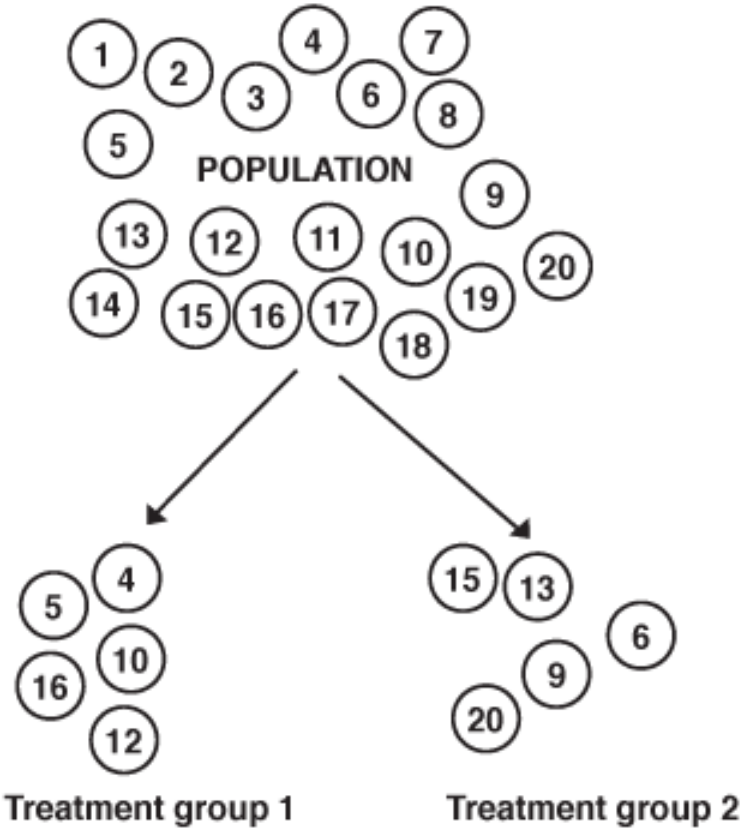
Collecting the data

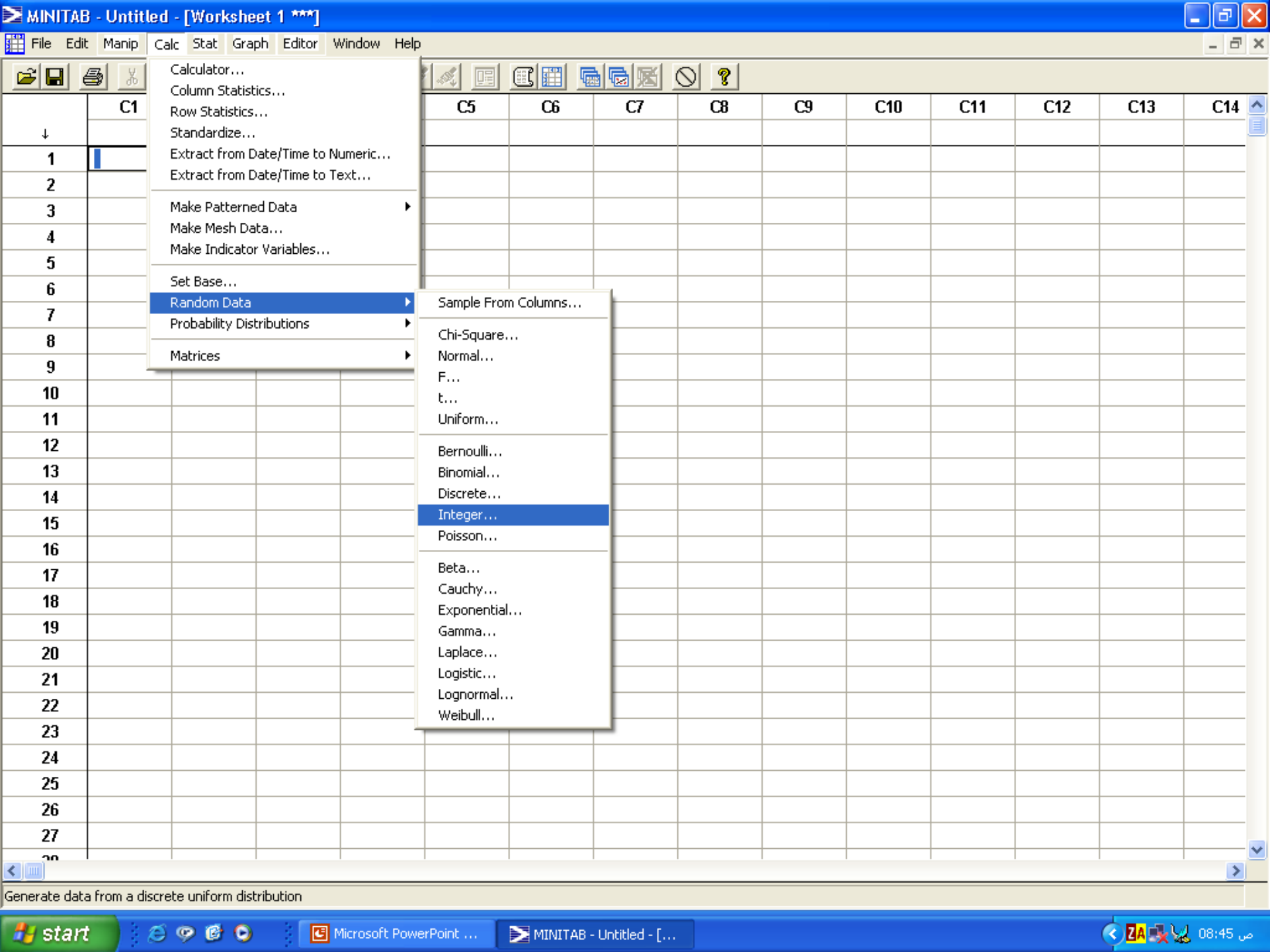
The three basic principles of experimental design are:

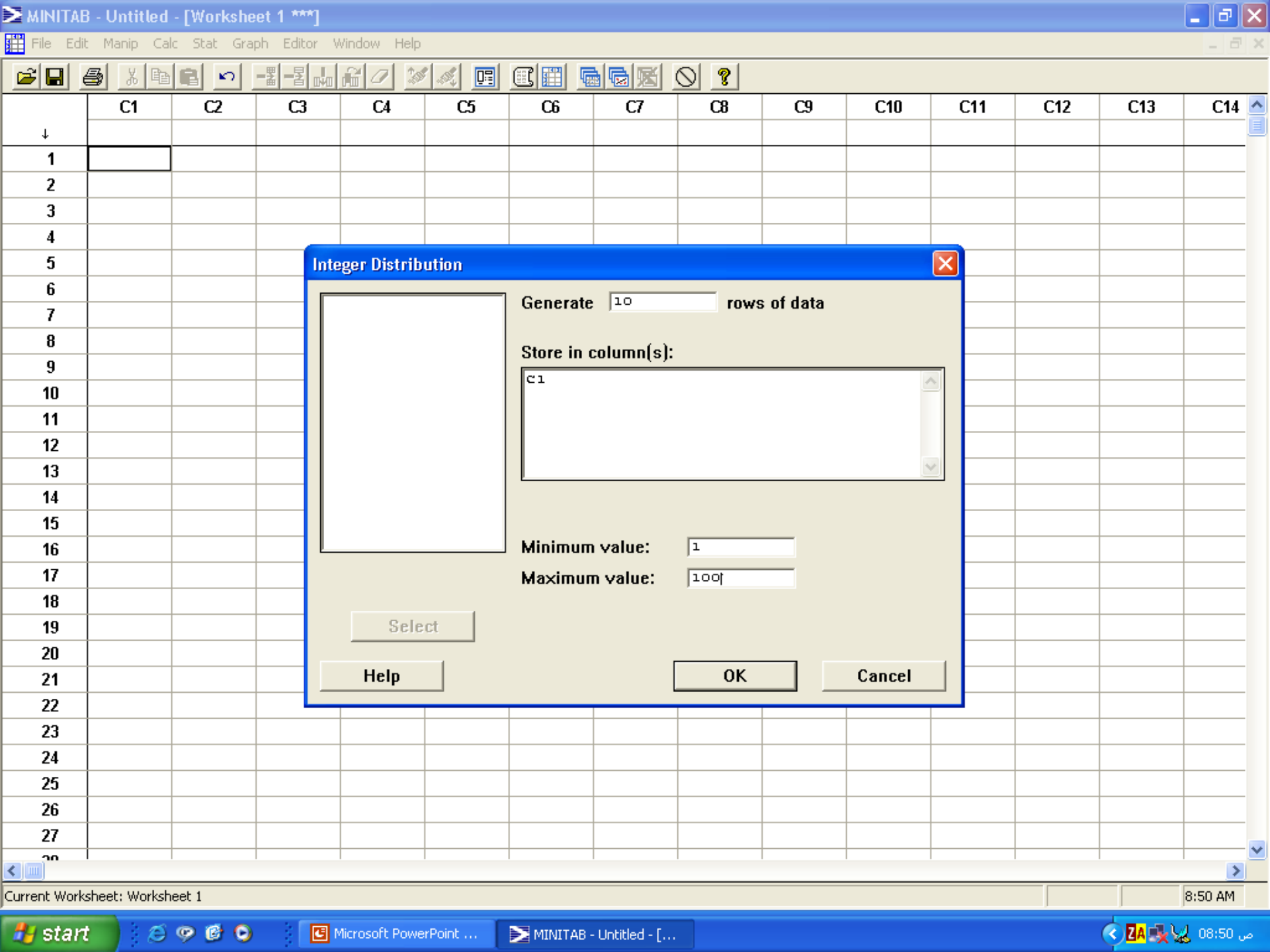
1. Randomization.
2. Replication.
3. Blocking.

Randomization

Randomization is the cornerstone underlying the use of statistical methods in experimental design. By randomization we mean that both the allocation of the experimental material and the order in which individual runs or trials of the experiment are to be performed are randomly determined.




MINITAB - Untitled - [Worksheet 1 ***]
File Edit Manip Calc Stat Graph Editor Window Help
Calculator...
Column Statistics...
Row Statistics...
Standardize...
Extract from Date/Time to Numeric...
Extract from Date/Time to Text...
Make Patterned Data
Make Mesh Data...
Make Indicator Variables...
Set Base...
Random Data
Probability Distributions
Matrices
Sample From Columns...
Chi-Square...
Normal...
F...
t...
Uniform...
Bernoulli...
Binomial...
Discrete...
Integer...
Poisson...
Beta...
Cauchy...
Exponential...
Gamma...
Laplace...
Logistic...
Lognormal...
Weibull...
C1 C5 C6 C7 C8 C9 C10 C11 C12 C13 C14
Generate data from a discrete uniform distribution
start
Microsoft PowerPoint ...
MINITAB - Untitled - [...
08:45 ص


MINITAB - Untitled - [Worksheet 1 ***]
File Edit Manip Calc Stat Graph Editor Window Help
C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 C12 C13 C14
Integer Distribution
Generate 10 rows of data
Store in column(s):
C1
Minimum value: 1
Maximum value: 100
Select
Help
OK
Cancel
Current Worksheet: Worksheet 1
8:50 AM
start
Microsoft PowerPoint ...
MINITAB - Untitled - [...
08:50 ص

MINITAB - Untitled - [Worksheet 1 ***]

File Edit Manip Calc Stat Graph Editor Window Help

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
↓														
1	89													
2	31													
3	93													
4	67													
5	22													
6	45													
7	52													
8	17													
9	94													
10	55													
11														
12														
13														
14														
15														
16														
17														
18														
19														
20														
21														
22														
23														
24														
25														
26														
27														

Current Worksheet: Worksheet 1

8:54 AM

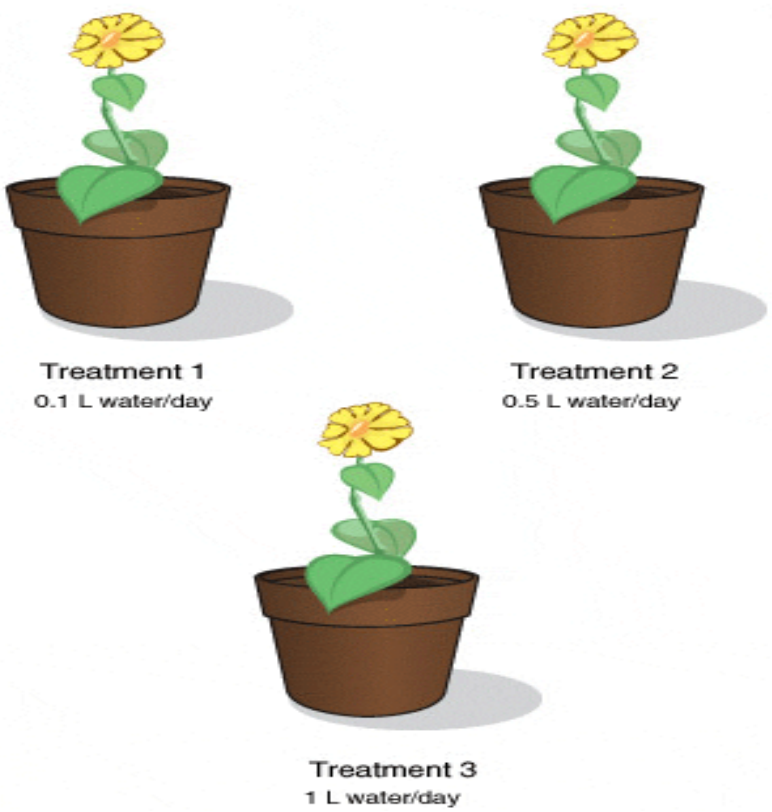
start Microsoft PowerPoint ... MINITAB - Untitled - [... 08:54 ص

Replication

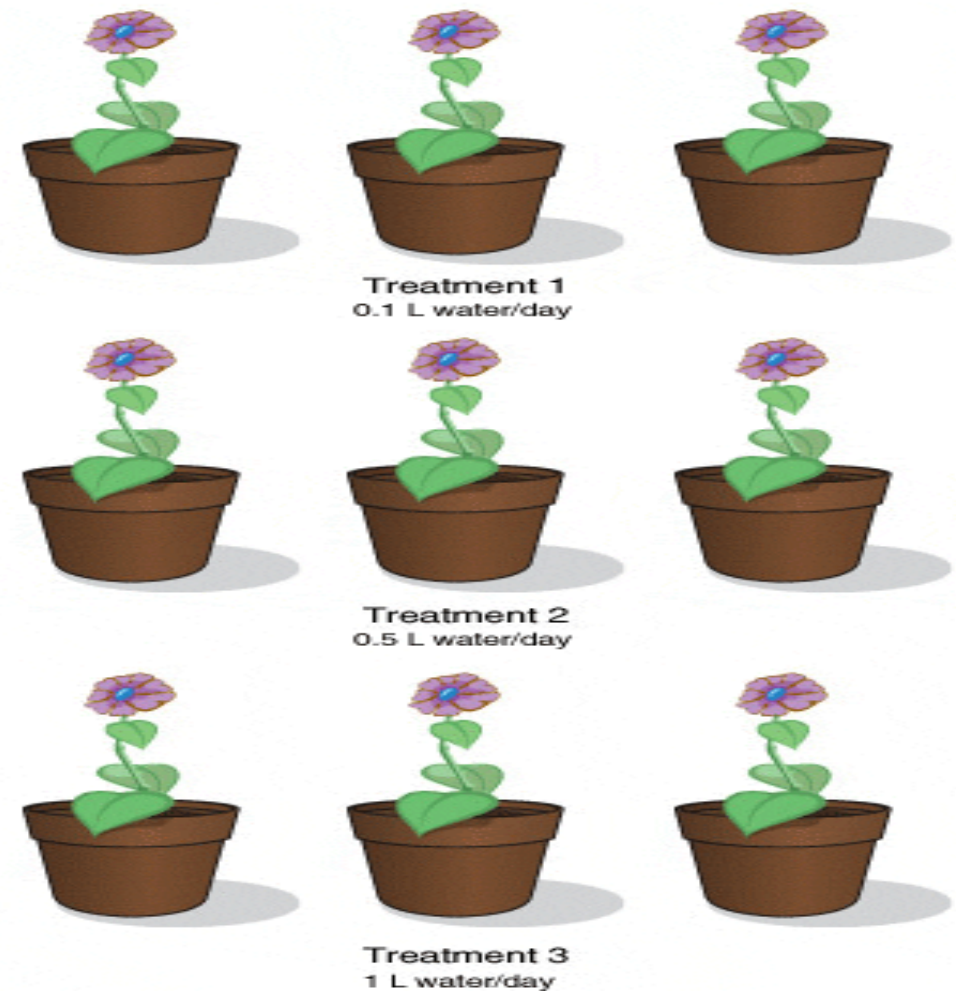
By replication we mean a repetition of the basic experiment. Replication has two important properties:

1. It allows the experimenter to obtain an estimate of the experimental error.
2. If the sample mean is used to estimate the effect of a factor in the experiment, then replication permits the experimenter to obtain a more precise estimate of this effect.

Without replication



With replication



Choice of sample size

Why would we want to plan?

1. The larger the sample sizes are, the easier it is to detect or find differences in the means.
2. The larger the sample size is, the higher the “cost” and the more likely that practically unimportant differences are to be found statistically significant.

Basic Statistics Terms

In order to determine the needed sample size we first must define some basic statistics terms.

Null hypothesis Ho is a hypothesis that is presumed true until statistical evidence in the form of a hypothesis test indicates otherwise.

In formulating a particular null hypothesis, we are always also formulating an **alternative hypothesis Ha**, which we will accept if the observed data values are sufficiently improbable under the null hypothesis .

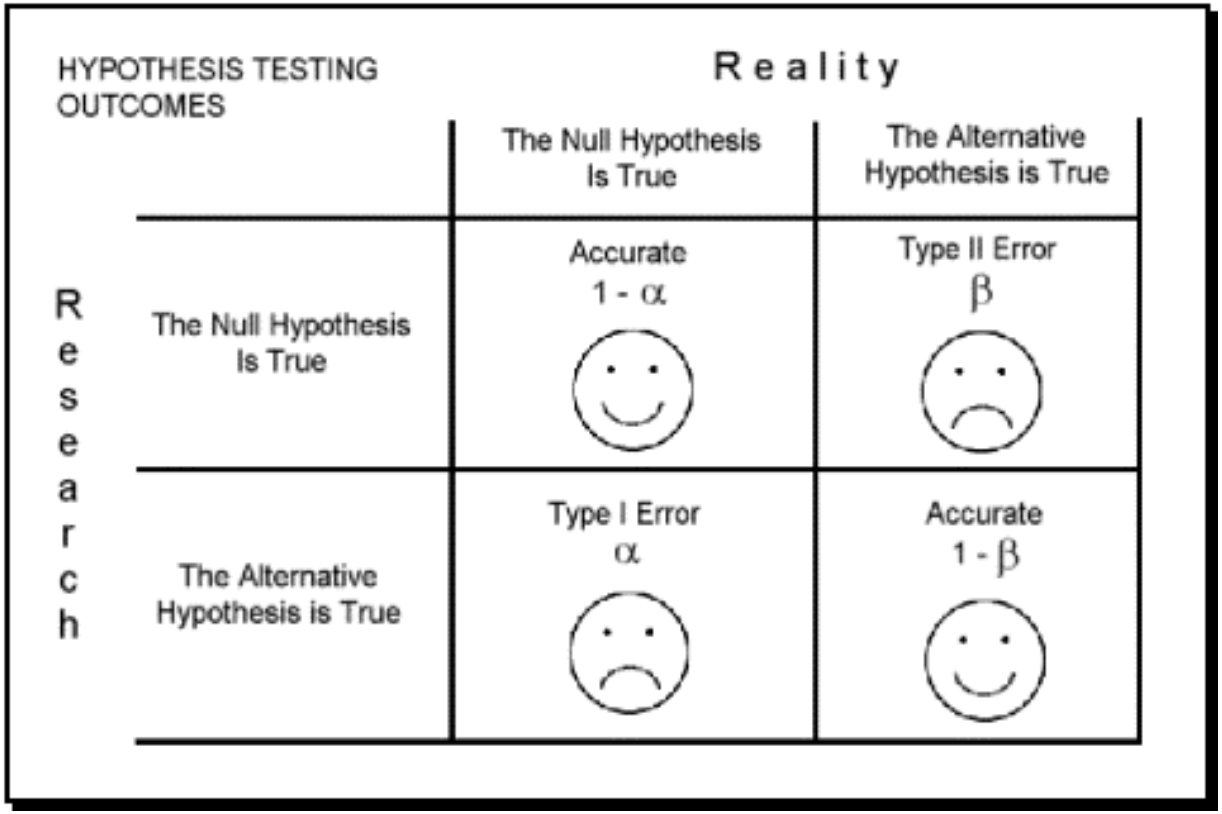
Definition of Type I and Type II errors

Sometimes our decisions will be correct and sometimes not. There are two possible errors, which we will call Type I and Type II errors, respectively.

A *Type I error* is the error of rejecting the null hypothesis when it is true. The probability of committing a Type I error is usually denoted by α.

A *Type II error* is the error of accepting the null hypothesis when it is false. The probability of making a Type II error is usually denoted by β.

Type I and Type II errors



Planning to detect any important difference

Let Δ= smallest difference range considered important by the researcher.

Specify Δ, β, α, σ and r use table A.10 (Applied linear statistical models by Neter, Wasserman and Kunter) to determine the needed sample size n ($=n_1=n_2=\ldots=n_r$).

Planning to detect any important difference

Example:

Let $\Delta=3$, $\beta=0.1$, $\alpha=0.05$, $\sigma=2$ and $r=4$

$\Delta/\sigma=1.5$, Power=$1-\beta=0.9$

→ Need n=14 observations at each factor level.

→ Need 14*4=56 homogeneous units.

Planning sample size to find the best treatment

Let λ= important difference between any two adjacent means.

r=number of factor level.

σ= standard deviation.

Specify λ, α, σ and r use table A.11 (Applied linear statistical models by Neter, Wasserman and Kunter) to get $d = \lambda\sqrt{n} / \sigma$ and solve for n.

Planning sample size to find the best treatment

Example:

Let $\lambda=2$, $\alpha=0.05$, $\sigma=3$ and r=5

1- $\alpha=0.95$ → d=3.0552

$n= (3.0552*3/2)^2= 21.002 \approx 21$

We need 21 observations at each of 5 levels

→ we need 105 experimental units.

Blocking

If the experimental units are not homogeneous, considerable improvement can be achieved by blocking (grouping) together units that are homogeneous.

Example: Rats coming from the same litter.

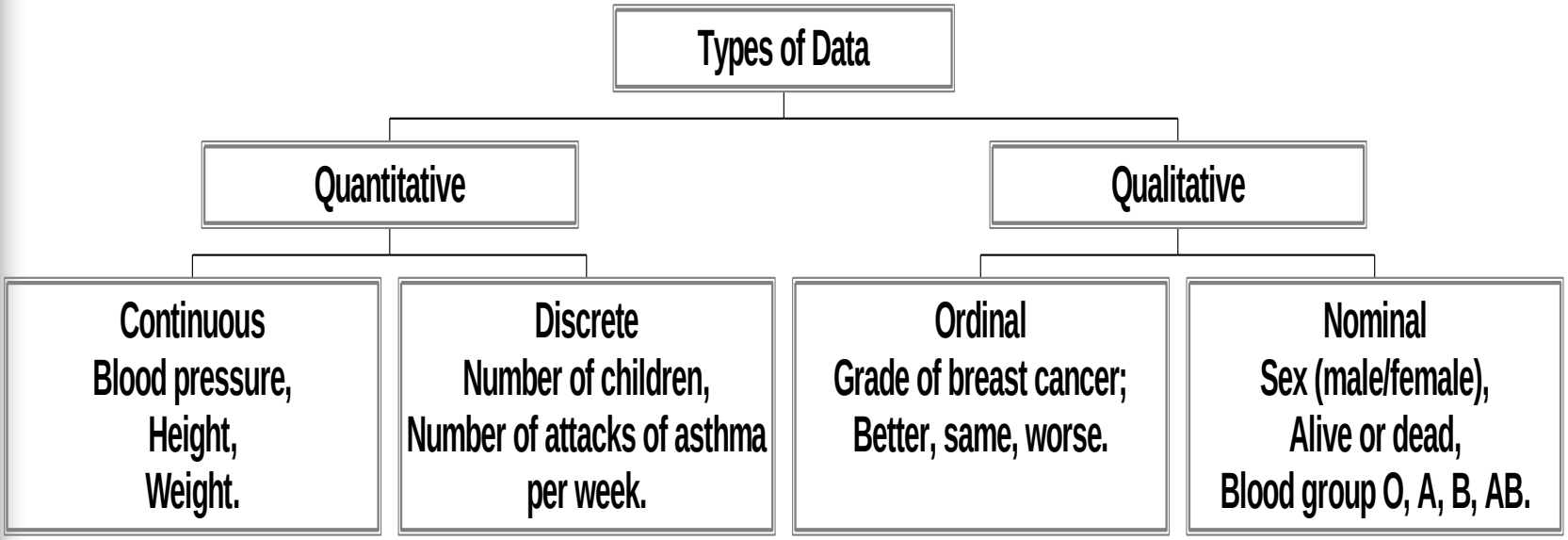
Graph

The purpose of a graph is to present data in a pictorial format that is easy to understand. A graph should make sense without any additional explanation needed from the body of a report. Every part of a graph should be absolutely necessary. Keep it simple.

Types of Graphs

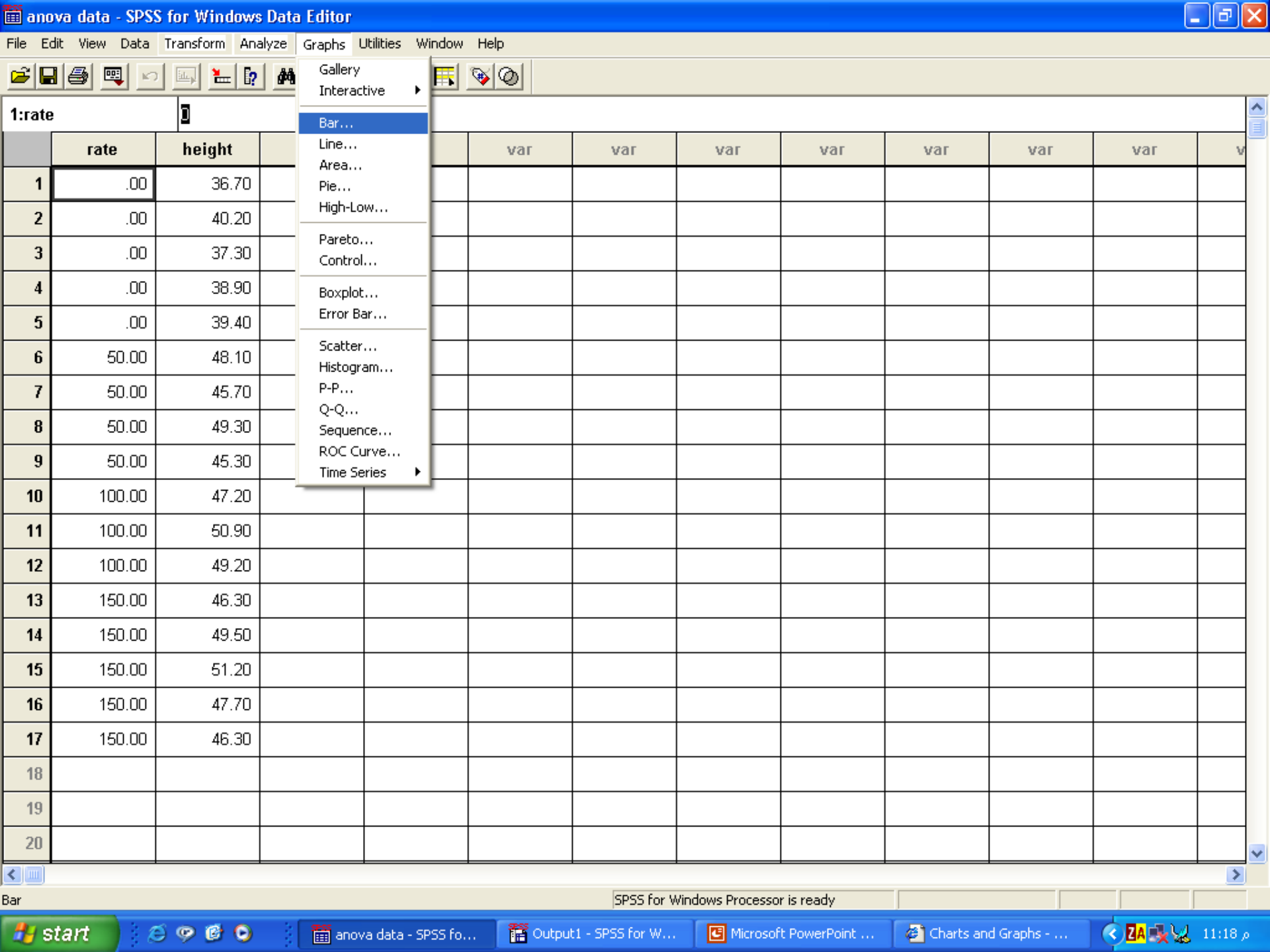
The type of graph one uses depends on the type of data collected and the point one is trying to make. In determining what type of graph to make, it is often useful to sketch out a graph to see whether it makes sense or is expressing the idea you wish to convey. Four of the most common types of graphs are discussed here.

Types of Data



Bar Graphs

Bar graphs are often used when comparing values from two or more groups or categories. It's a fast way to show big differences. Notice how easy it is to see what was done in the experiment below with cereal plant growth and different rates of nitrogen fertilizer.


anova data - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Gallery
Interactive
Bar...
Line...
Area...
Pie...
High-Low...
Pareto...
Control...
Boxplot...
Error Bar...
Scatter...
Histogram...
P-P...
Q-Q...
Sequence...
ROC Curve...
Time Series
1:rate
rate height var var var var var var var var
1 .00 36.70
2 .00 40.20
3 .00 37.30
4 .00 38.90
5 .00 39.40
6 50.00 48.10
7 50.00 45.70
8 50.00 49.30
9 50.00 45.30
10 100.00 47.20
11 100.00 50.90
12 100.00 49.20
13 150.00 46.30
14 150.00 49.50
15 150.00 51.20
16 150.00 47.70
17 150.00 46.30
18
19
20
Bar
SPSS for Windows Processor is ready
start
anova data - SPSS fo...
Output1 - SPSS for W...
Microsoft PowerPoint ...
Charts and Graphs - ...
11:18

anova data - SPSS for Windows Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1:rate

	rate	height	var	var	var	var	var	var	var	var	var	var
1	.00	36.70										
2	.00	40.20										
3	.00	37.30										
4	.00	38.90										
5	.00	39.40										
6	50.00	48.10										
7	50.00	45.70										
8	50.00	49.30										
9	50.00	45.30										
10	100.00	47.20										
11	100.00	50.90										
12	100.00	49.20										
13	150.00	46.30										
14	150.00	49.50										
15	150.00	51.20										
16	150.00	47.70										
17	150.00	46.30										
18												
19												
20												

Bar Charts

Simple
Clustered
Stacked

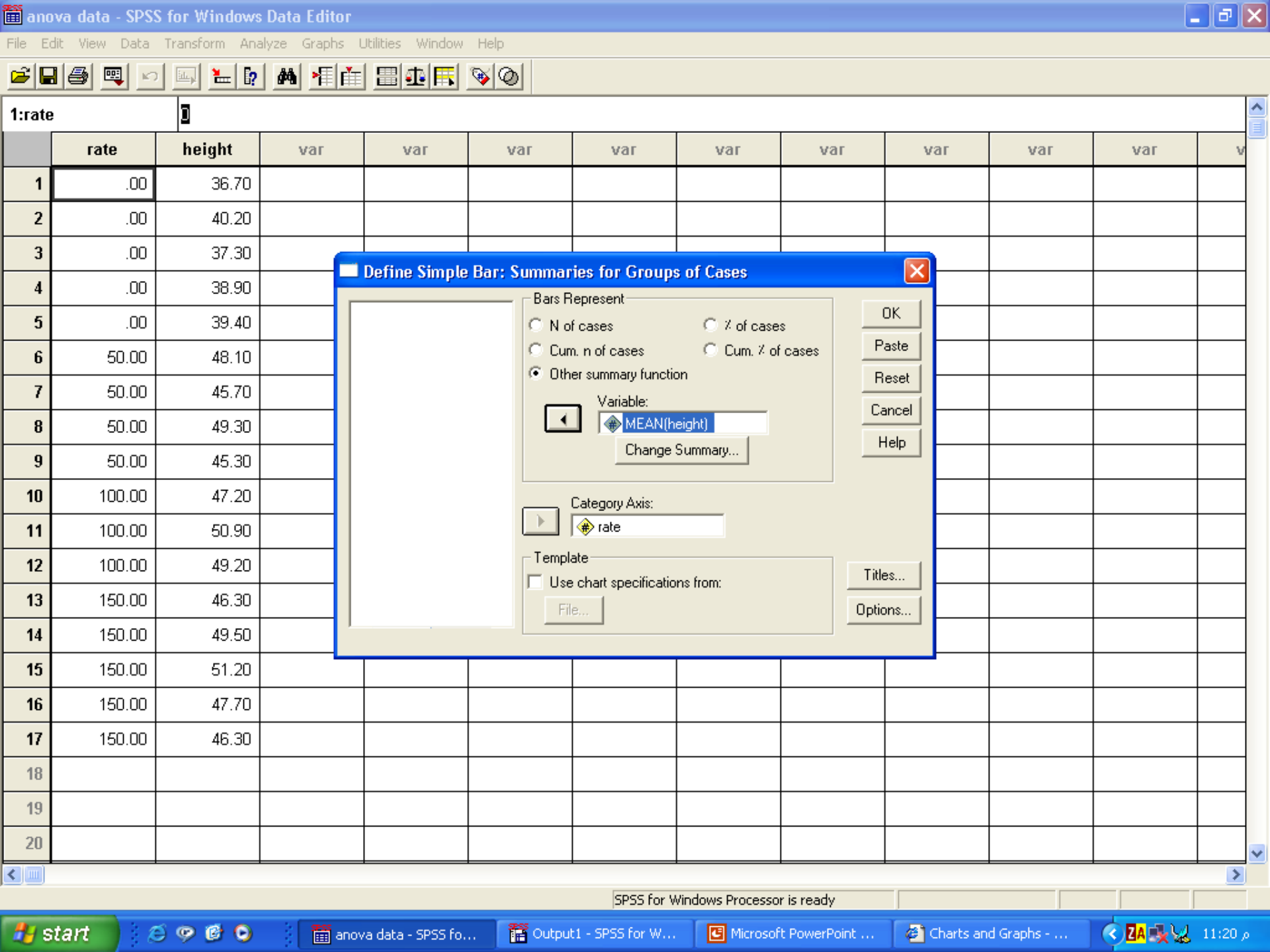
Define
Cancel
Help

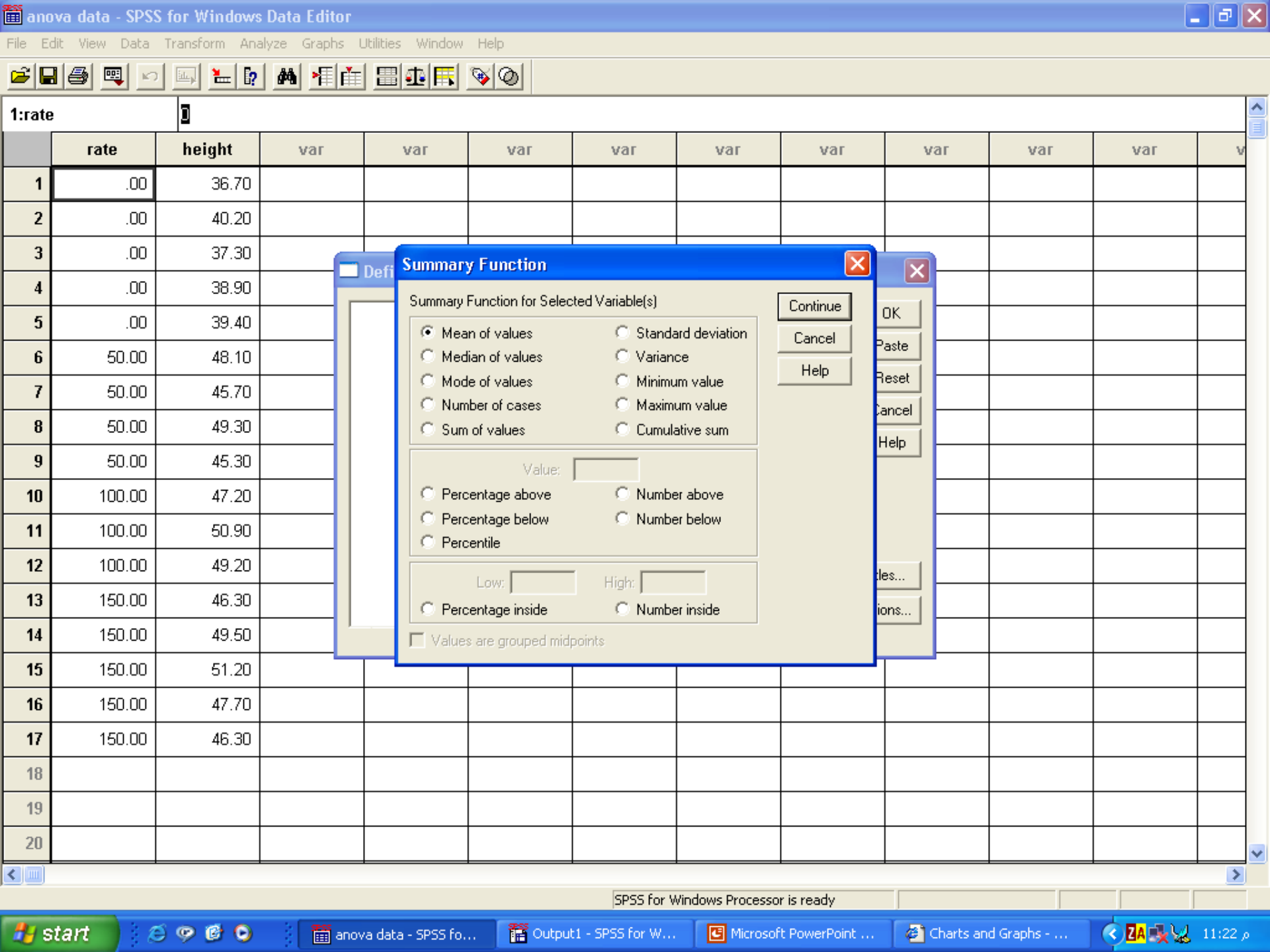
Data in Chart Are

- (•) Summaries for groups of cases
- ( ) Summaries of separate variables
- ( ) Values of individual cases

SPSS for Windows Processor is ready

start | anova data - SPSS fo... | Output1 - SPSS for W... | Microsoft PowerPoint ... | Charts and Graphs - ... | 11:19


anova data - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1:rate
rate height var var var var var var var var var var
1 .00 36.70
2 .00 40.20
3 .00 37.30
4 .00 38.90
5 .00 39.40
6 50.00 48.10
7 50.00 45.70
8 50.00 49.30
9 50.00 45.30
10 100.00 47.20
11 100.00 50.90
12 100.00 49.20
13 150.00 46.30
14 150.00 49.50
15 150.00 51.20
16 150.00 47.70
17 150.00 46.30
18
19
20
Define Simple Bar: Summaries for Groups of Cases
Bars Represent
N of cases
% of cases
Cum. n of cases
Cum. % of cases
Other summary function
Variable:
MEAN(height)
Change Summary...
OK
Paste
Reset
Cancel
Help
Category Axis:
rate
Template
Use chart specifications from:
File...
Titles...
Options...
SPSS for Windows Processor is ready
start
anova data - SPSS fo...
Output1 - SPSS for W...
Microsoft PowerPoint ...
Charts and Graphs - ...
11:20

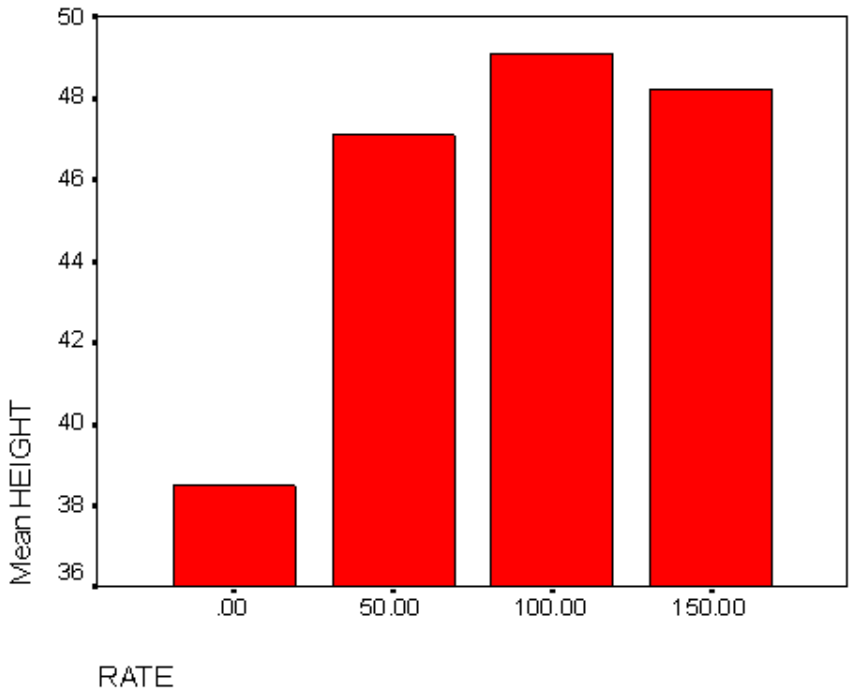

anova data - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1:rate
rate height var var var var var var var var var var
1 .00 36.70
2 .00 40.20
3 .00 37.30
4 .00 38.90
5 .00 39.40
6 50.00 48.10
7 50.00 45.70
8 50.00 49.30
9 50.00 45.30
10 100.00 47.20
11 100.00 50.90
12 100.00 49.20
13 150.00 46.30
14 150.00 49.50
15 150.00 51.20
16 150.00 47.70
17 150.00 46.30
18
19
20
Summary Function
Summary Function for Selected Variable(s)
Mean of values
Median of values
Mode of values
Number of cases
Sum of values
Standard deviation
Variance
Minimum value
Maximum value
Cumulative sum
Value:
Percentage above
Percentage below
Percentile
Number above
Number below
Low:
High:
Percentage inside
Number inside
Values are grouped midpoints
Continue
Cancel
Help
OK
Paste
Reset
Cancel
Help
SPSS for Windows Processor is ready
start
anova data - SPSS fo...
Output1 - SPSS for W...
Microsoft PowerPoint ...
Charts and Graphs - ...
11:22

Output1 - SPSS for Windows Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

- Output
  - Graph
    - Notes
  - Graph
    - Title
    - Notes
    - Bar of mean(height) by r

Graph



Double click to edit Title

SPSS for Windows Processor is ready

H: 17 , W: 720

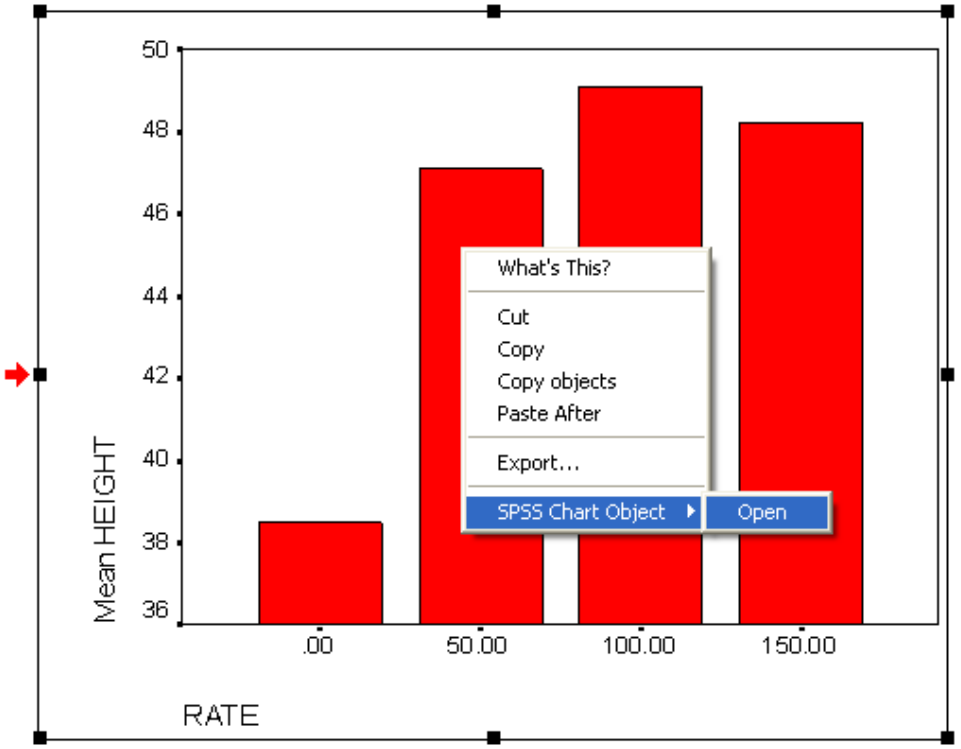
start | anova data - SPSS fo... | Output1 - SPSS for W... | Microsoft PowerPoint ... | Charts and Graphs - ... | 11:23

Output1 - SPSS for Windows Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

Output
Graph
Notes
Graph
Title
Notes
Bar of mean(height) by r

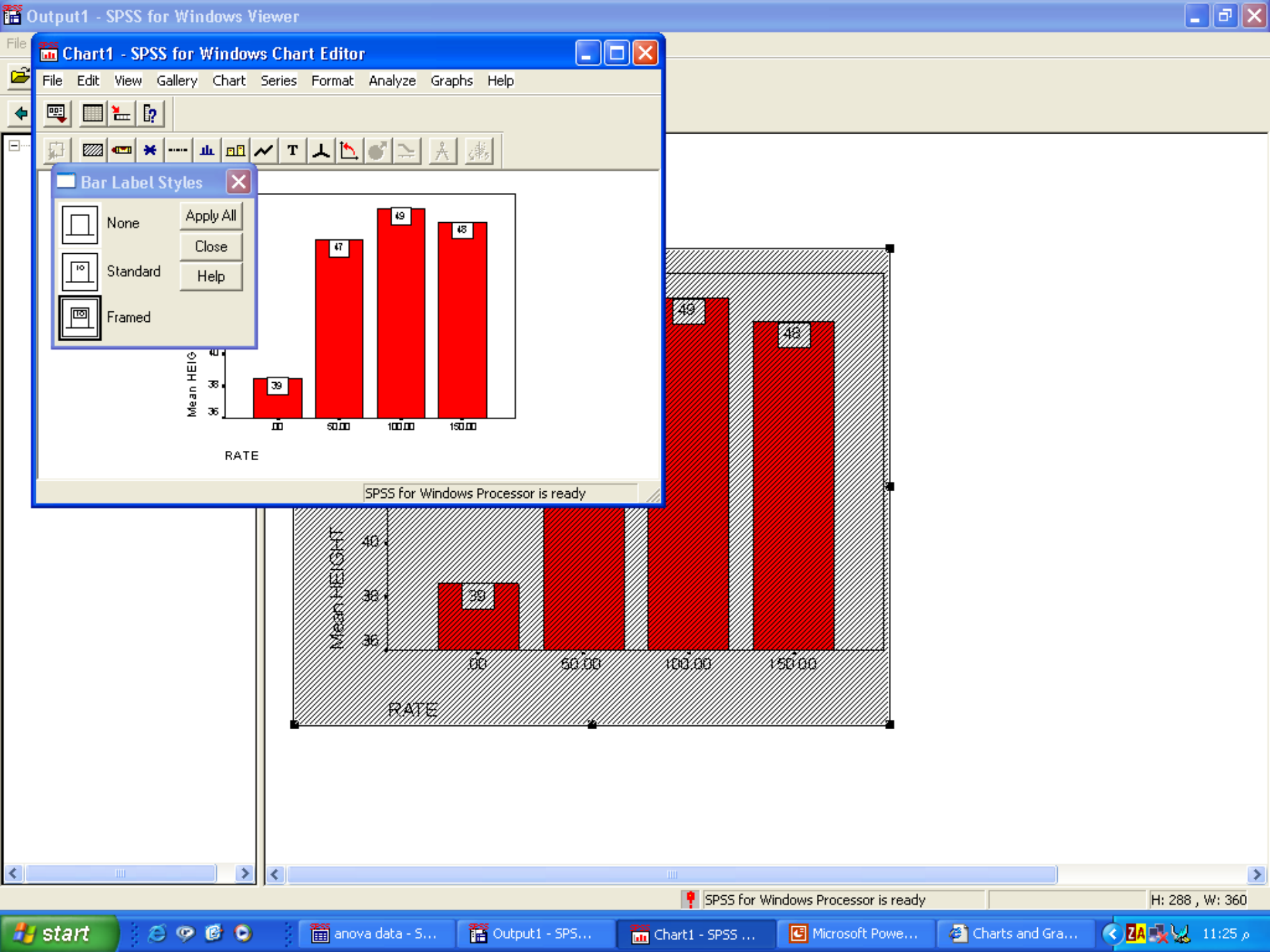
Graph



SPSS for Windows Processor is ready

H: 288 , W: 360

start anova data - SPSS fo... Output1 - SPSS for W... Microsoft PowerPoint ... Charts and Graphs - ... 11:23 م

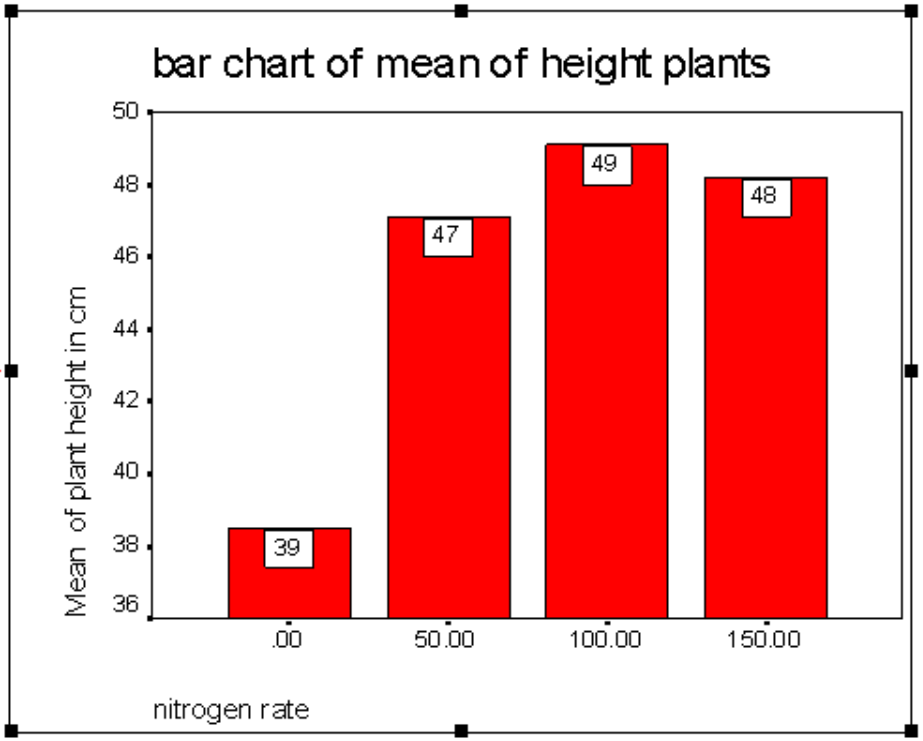

Output1 - SPSS for Windows Viewer
Chart1 - SPSS for Windows Chart Editor
File Edit View Gallery Chart Series Format Analyze Graphs Help
Bar Label Styles
None
Standard
Framed
Apply All
Close
Help
Mean HEIGHT
40
38
36
39
47
49
48
00
50.00
100.00
150.00
RATE
SPSS for Windows Processor is ready
SPSS for Windows Processor is ready
H: 288 , W: 360
start
anova data - S...
Output1 - SPS...
Chart1 - SPSS ...
Microsoft Powe...
Charts and Gra...
11:25

Output1 - SPSS for Windows Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

Output
Graph
Title
Notes
Bar of mean(height) by n

Graph



1 items selected (0 hidden/collapsed)

SPSS for Windows Processor is ready

H: 288 , W: 360

start

Microsoft PowerPoint ...

anova data - SPSS fo...

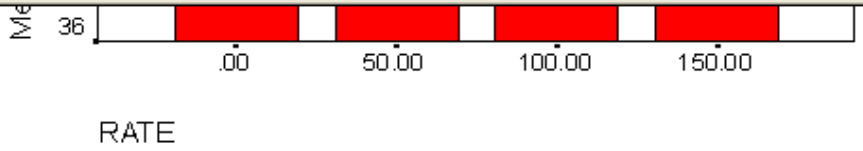
Output1 - SPSS for W...

02:20

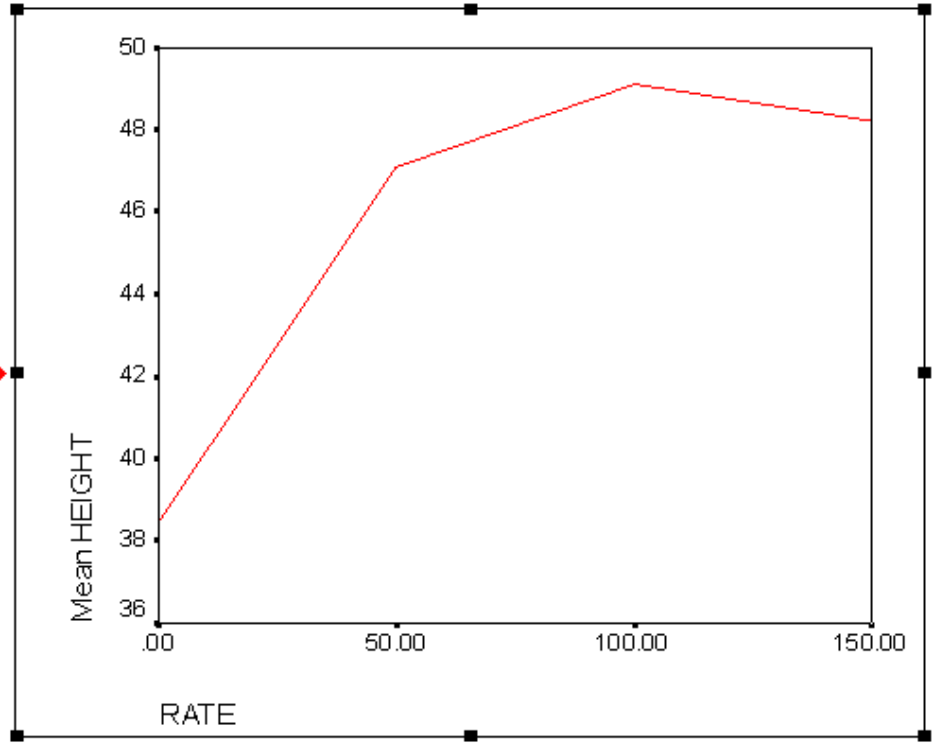
Output1 - SPSS for Windows Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

- Output
  - Graph
    - Notes
  - Graph
    - Title
    - Notes
    - Bar of mean(height) by r
  - Graph
    - Title
    - Notes
    - Line of mean(height) by



Graph



ouble click to edit Chart

SPSS for Windows Processor is ready

H: 288 , W: 360

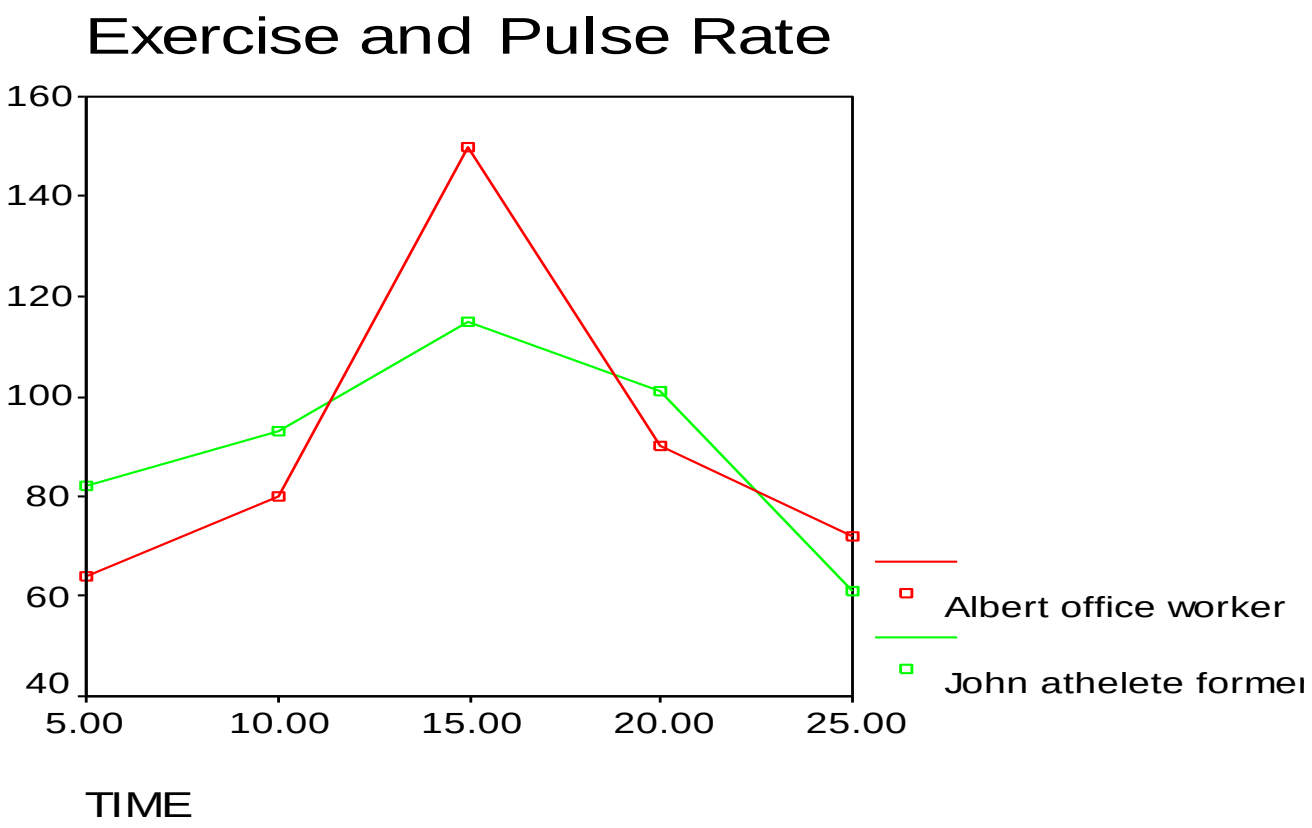
start anova data - SPSS fo... Output1 - SPSS for W... Microsoft PowerPoint ... 11:32 م

Line Graphs

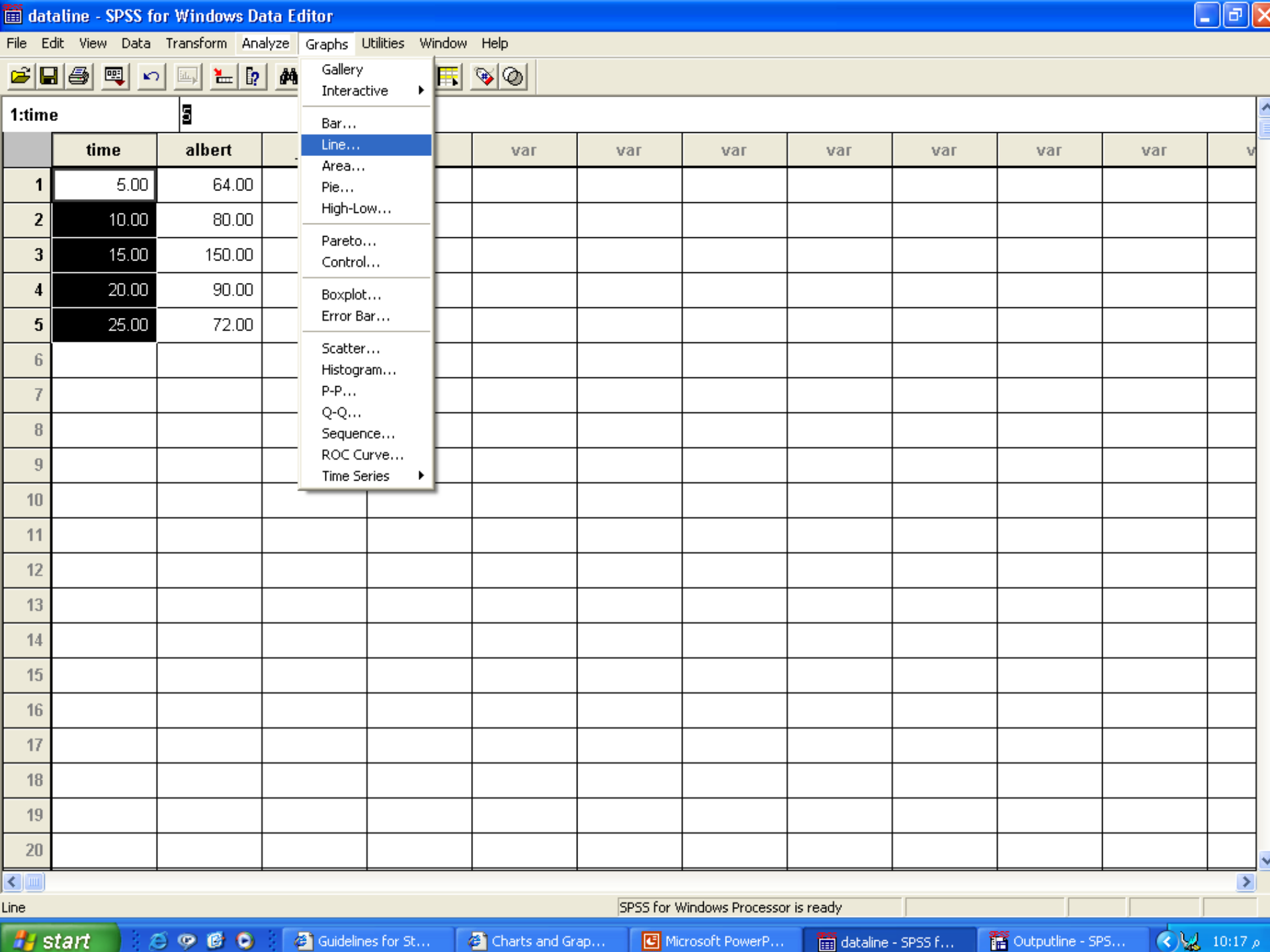
A line graph is used to show continuing data; how one thing is affected by another. It's clear to see how things are going by the rises and falls a line graph shows. This kind of graph is needed to show the effect of an independent variable on a dependent variable. In the sample below, the pulse rate of a person is shown to change over time. As time continues, the pulse rate changes.

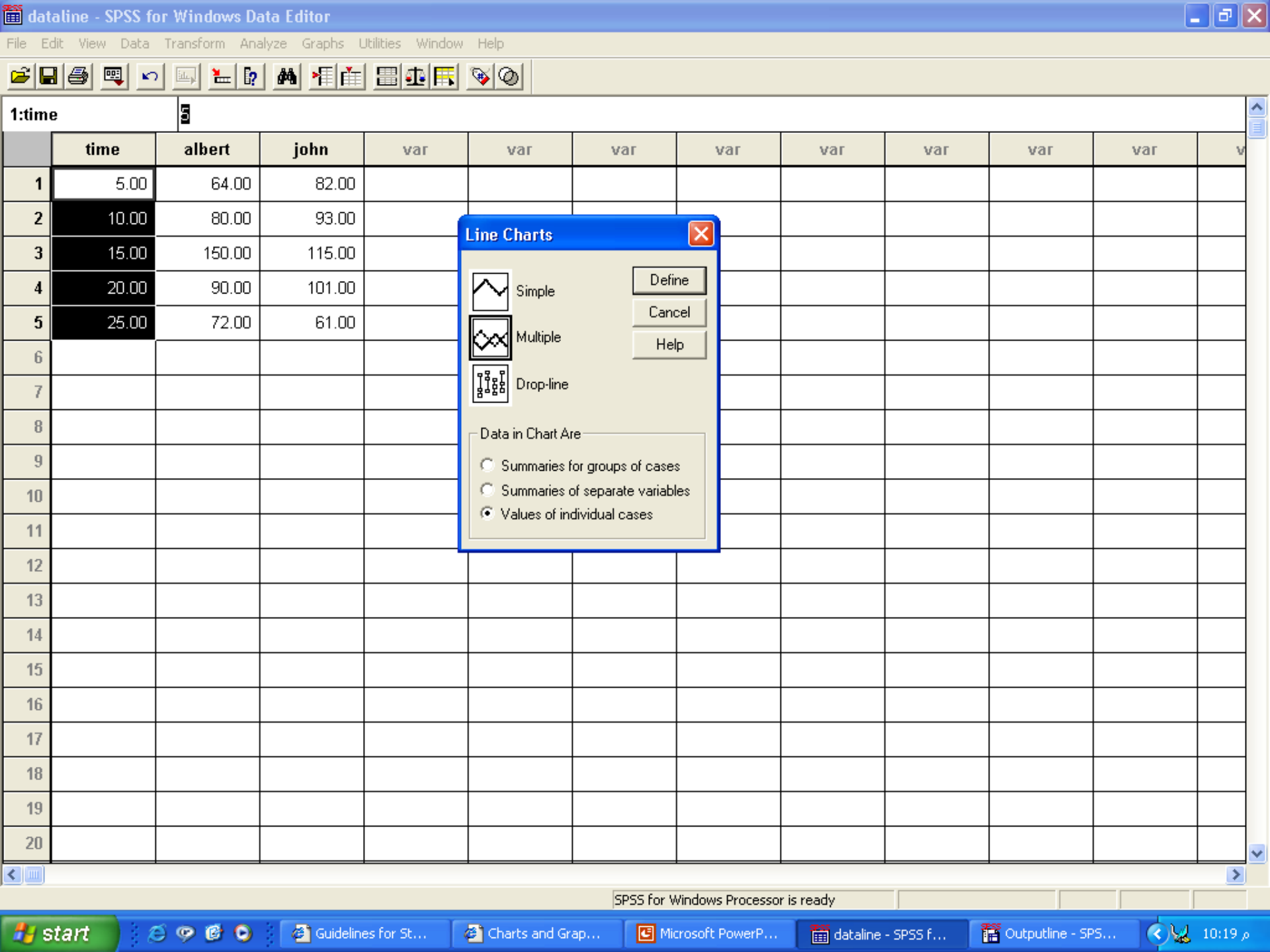
A typical chart for this graph might look like this:

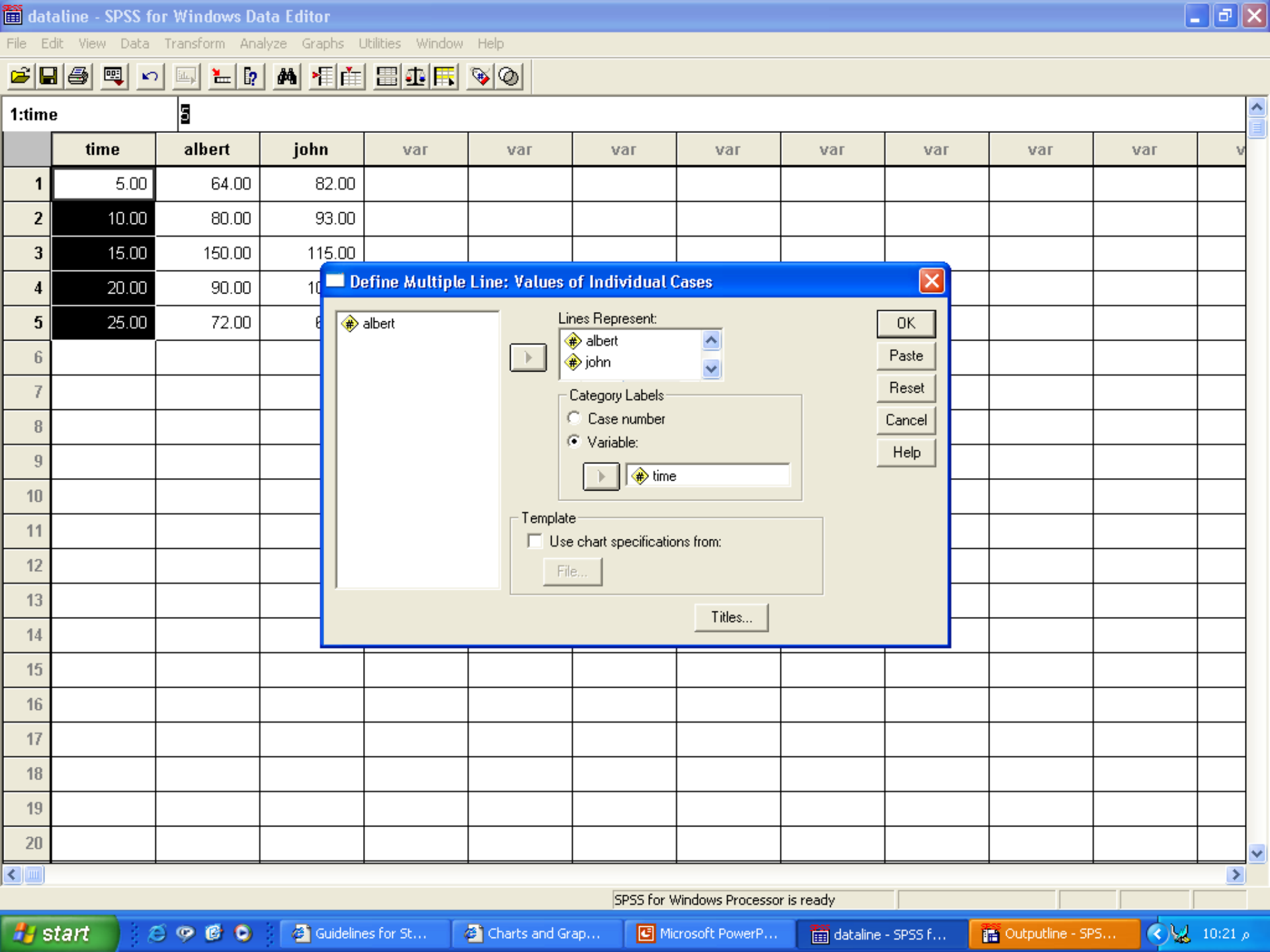
Line Graph


Exercise and Pulse Rate
160
140
120
100
80
60
40
5.00
10.00
15.00
20.00
25.00
TIME
Albert office worker
John athelete formei


dataline - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1:time
time albert john var var var var var var var var var
1 5.00 64.00 82.00
2 10.00 80.00 93.00
3 15.00 150.00 115.00
4 20.00 90.00 101.00
5 25.00 72.00 61.00
6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
SPSS for Windows Processor is ready
start
Guidelines for St...
Charts and Grap...
Microsoft PowerP...
dataline - SPSS f...
Outputline - SPS...
10:16


dataline - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Gallery
Interactive
Bar...
Line...
Area...
Pie...
High-Low...
Pareto...
Control...
Boxplot...
Error Bar...
Scatter...
Histogram...
P-P...
Q-Q...
Sequence...
ROC Curve...
Time Series
1:time
time albert var var var var var var var
1 5.00 64.00
2 10.00 80.00
3 15.00 150.00
4 20.00 90.00
5 25.00 72.00
Line
SPSS for Windows Processor is ready
start
Guidelines for St...
Charts and Grap...
Microsoft PowerP...
dataline - SPSS f...
Outputline - SPS...
10:17


dataline - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1:time
time albert john var var var var var var var var var
1 5.00 64.00 82.00
2 10.00 80.00 93.00
3 15.00 150.00 115.00
4 20.00 90.00 101.00
5 25.00 72.00 61.00
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Line Charts
Simple
Multiple
Drop-line
Define
Cancel
Help
Data in Chart Are
Summaries for groups of cases
Summaries of separate variables
Values of individual cases
SPSS for Windows Processor is ready
start
Guidelines for St...
Charts and Grap...
Microsoft PowerP...
dataline - SPSS f...
Outputline - SPS...
10:19

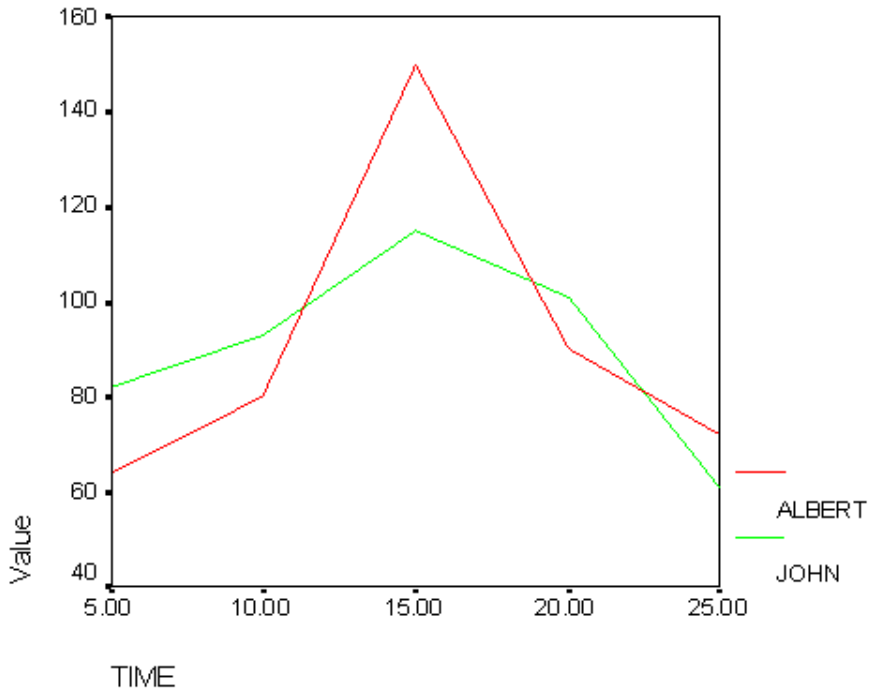

dataline - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1:time
time albert john var var var var var var var var var
1 5.00 64.00 82.00
2 10.00 80.00 93.00
3 15.00 150.00 115.00
4 20.00 90.00 10
5 25.00 72.00 8
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Define Multiple Line: Values of Individual Cases
albert
Lines Represent:
albert
john
OK
Paste
Reset
Cancel
Help
Category Labels
Case number
Variable:
time
Template
Use chart specifications from:
File...
Titles...
SPSS for Windows Processor is ready
start
Guidelines for St...
Charts and Grap...
Microsoft PowerP...
dataline - SPSS f...
Outputline - SPS...
10:21

Outputline - SPSS for Windows Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

- Output
  - Graph
    - Notes
  - Graph
    - Notes
  - Graph
    - Notes
  - Graph
    - Notes
  - Graph
    - Title
    - Notes
    - Line of value(albert john)

Graph

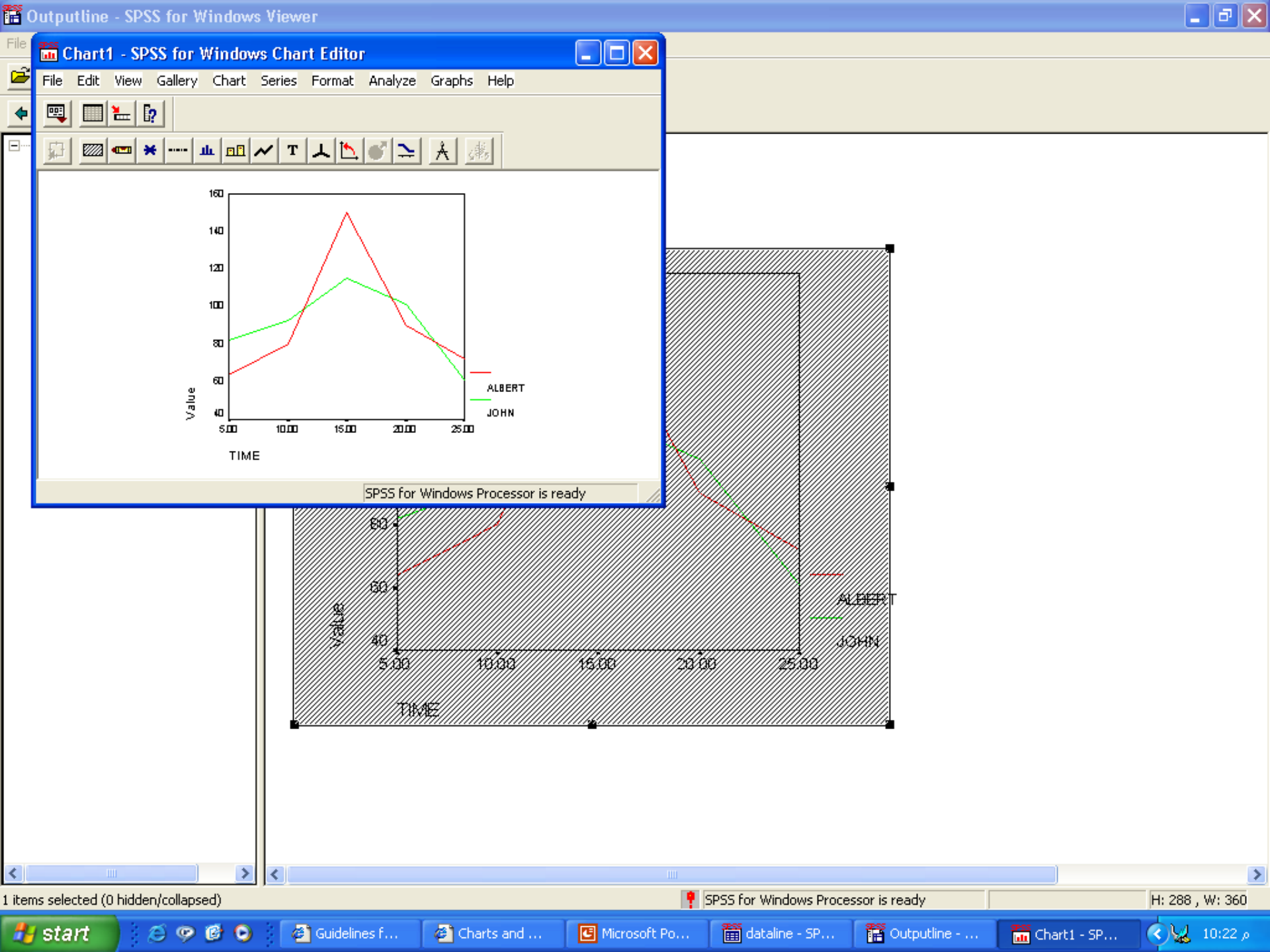


Double click to edit Title

SPSS for Windows Processor is ready

H: 17 , W: 720

start | Guidelines for St... | Charts and Grap... | Microsoft PowerP... | dataline - SPSS f... | Outputline - SPS... | 10:22

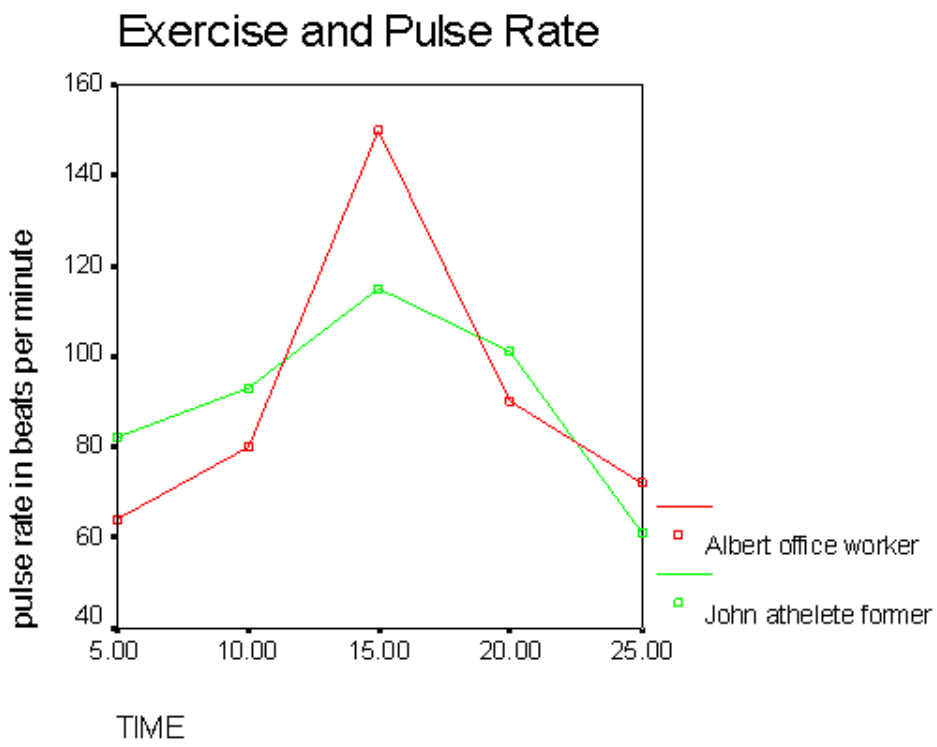

Outputline - SPSS for Windows Viewer
Chart1 - SPSS for Windows Chart Editor
File Edit View Gallery Chart Series Format Analyze Graphs Help
160
140
120
100
80
60
40
Value
5.00
10.00
15.00
20.00
25.00
TIME
ALBERT
JOHN
SPSS for Windows Processor is ready
80
60
40
Value
5.00
10.00
15.00
20.00
25.00
TIME
ALBERT
JOHN
1 items selected (0 hidden/collapsed)
SPSS for Windows Processor is ready
H: 288 , W: 360
start
Guidelines f...
Charts and ...
Microsoft Po...
dataline - SP...
Outputline - ...
Chart1 - SP...
10:22

Outputline - SPSS for Windows Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

- Output
  - Graph
    - Notes
  - Graph
    - Notes
  - Graph
    - Title
    - Notes
    - Line of value(mabel albe

Graph



SPSS for Windows Processor is ready

start | Guidelines f... | Charts and ... | Microsoft Po... | dataline - SP... | Output2 - S... | Outputline - ... | 10:26

Pie Graph

A circle graph is used to show how a part of something relates to the whole. This kind of graph is needed to show percentages effectively.

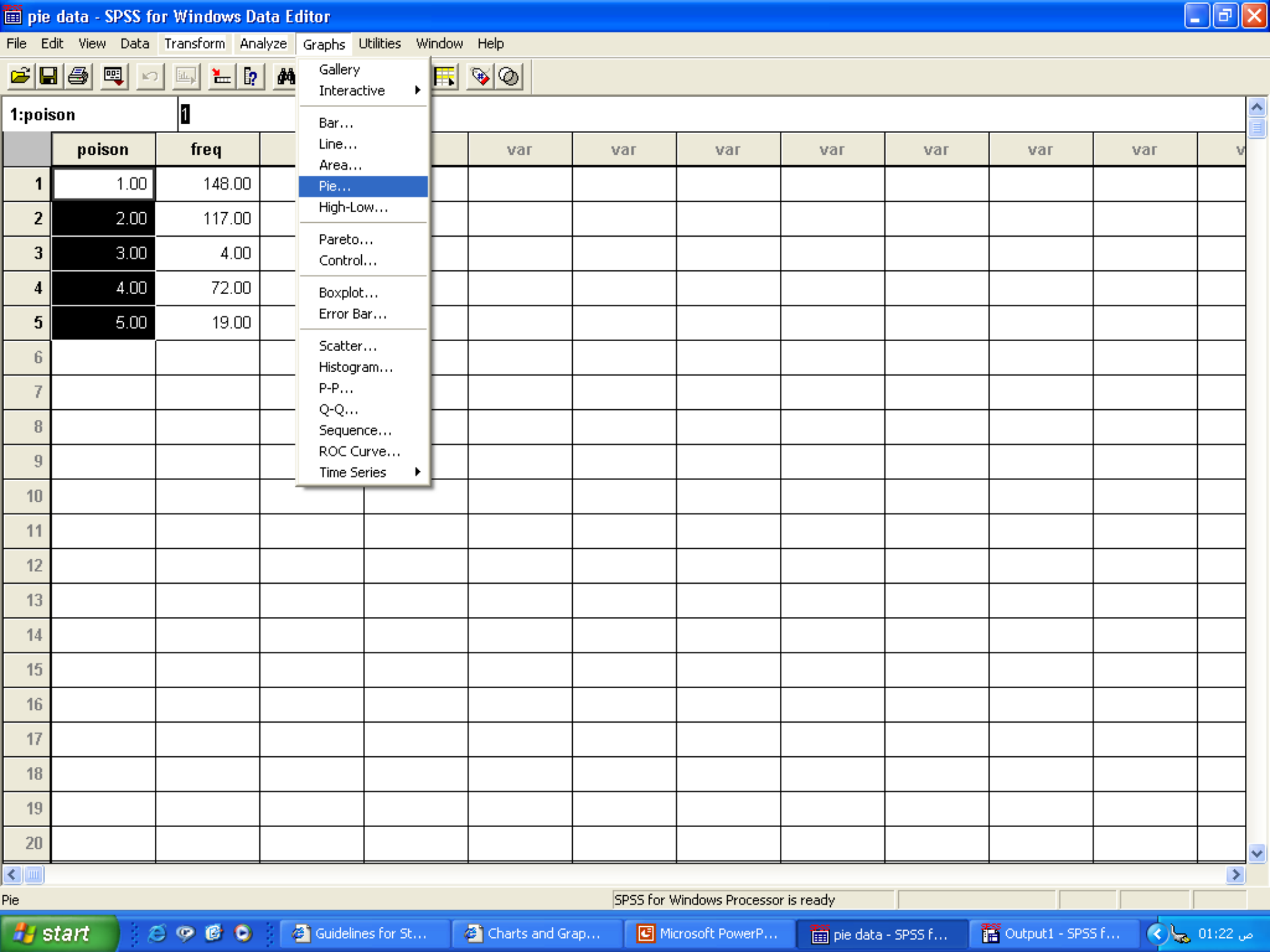
A typical chart for this graph might look like this:

Pie Graph

Example:

Cases of accidental poisoning seen at a particular Saudi hospital in a particular year were classified by the type of ingested poison [El Mouzan et al. (1986)].

Type of poison	Frequency
Kerosene or gasoline	148
Clorox (bleach)	117
Soap or shampoo	4
Drugs (medications)	72
Chemicals	19


pie data - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Gallery
Interactive
Bar...
Line...
Area...
Pie...
High-Low...
Pareto...
Control...
Boxplot...
Error Bar...
Scatter...
Histogram...
P-P...
Q-Q...
Sequence...
ROC Curve...
Time Series
1:poison
poison freq
1.00 148.00
2.00 117.00
3.00 4.00
4.00 72.00
5.00 19.00
Pie
SPSS for Windows Processor is ready
start
Guidelines for St...
Charts and Grap...
Microsoft PowerP...
pie data - SPSS f...
Output1 - SPSS f...
01:22 ص

pie data - SPSS for Windows Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1:poison 1

	poison	freq	var	var	var	var	var	var	var	var	var	var
1	1.00	148.00										
2	2.00	117.00										
3	3.00	4.00										
4	4.00	72.00										
5	5.00	19.00										
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												

Pie Charts

Data in Chart Are

- Summaries for groups of cases
- Summaries of separate variables
- Values of individual cases

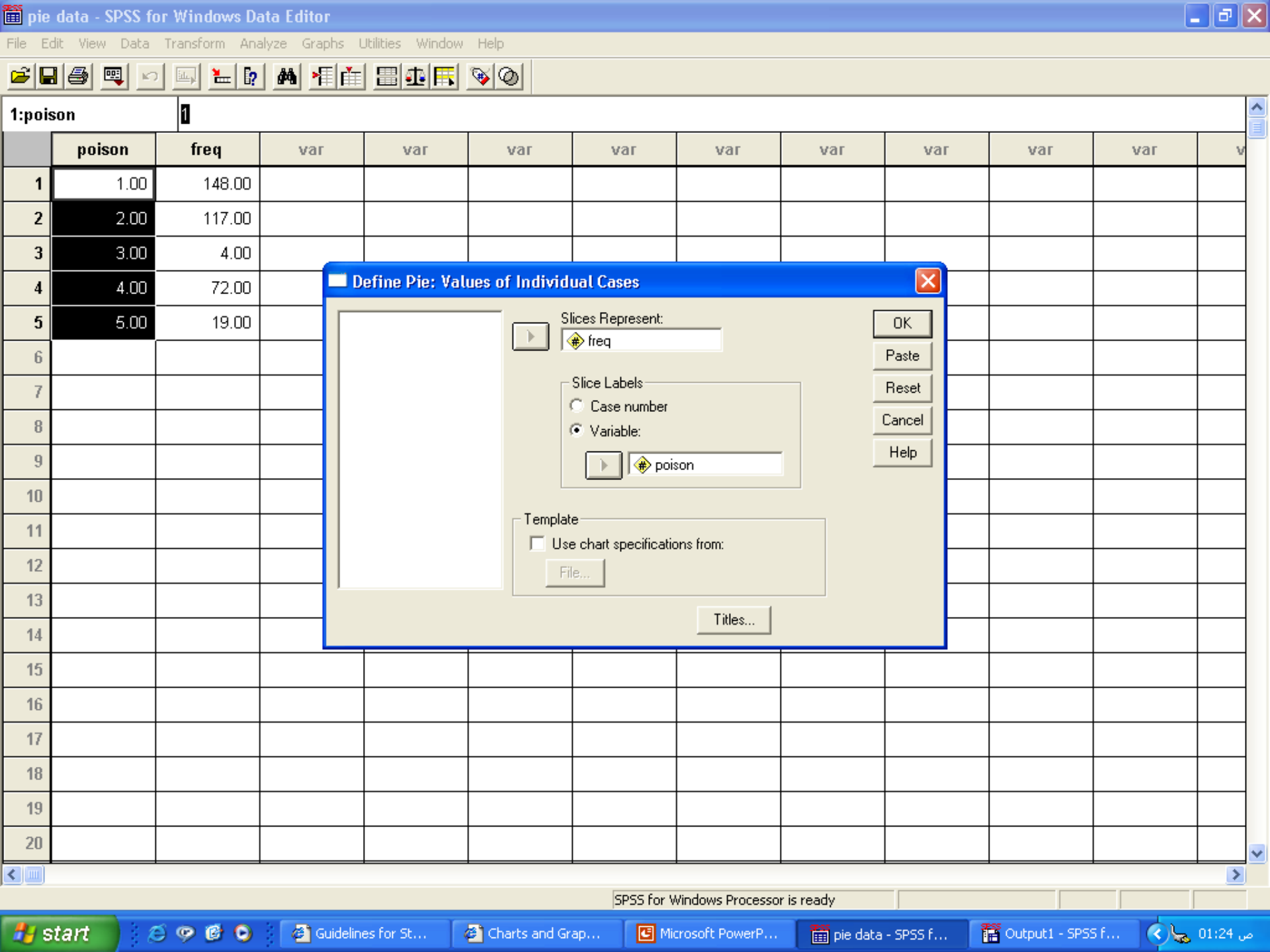
Define

Cancel

Help

SPSS for Windows Processor is ready

start Guidelines for St... Charts and Grap... Microsoft PowerP... pie data - SPSS f... Output1 - SPSS f... 01:23 ص

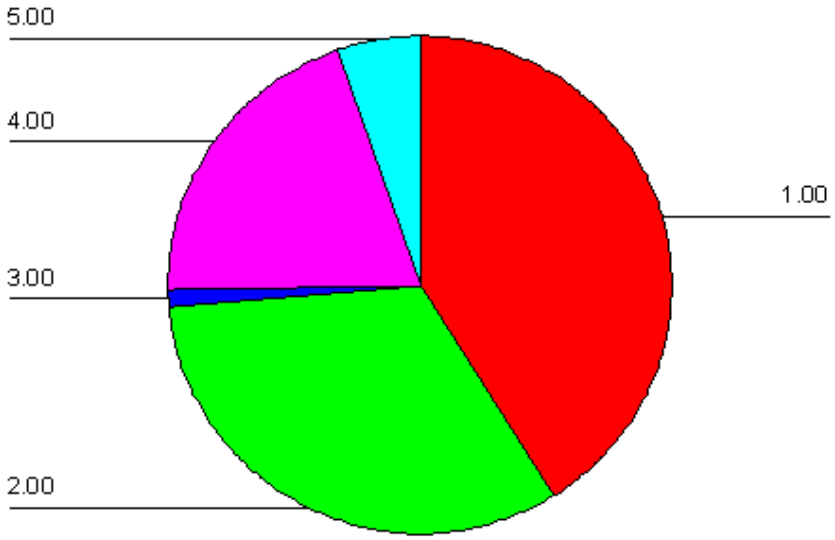

pie data - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1:poison
poison freq var var var var var var var var var var
1 1.00 148.00
2 2.00 117.00
3 3.00 4.00
4 4.00 72.00
5 5.00 19.00
Define Pie: Values of Individual Cases
Slices Represent:
freq
Slice Labels
Case number
Variable:
poison
Template
Use chart specifications from:
File...
Titles...
OK
Paste
Reset
Cancel
Help
SPSS for Windows Processor is ready
start
Guidelines for St...
Charts and Grap...
Microsoft PowerP...
pie data - SPSS f...
Output1 - SPSS f...
01:24 ص

Output1 - SPSS for Windows Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

- Output
  - Graph
    - Notes
  - Graph
    - Title
    - Notes
    - Pie of value(freq) by poi

Graph

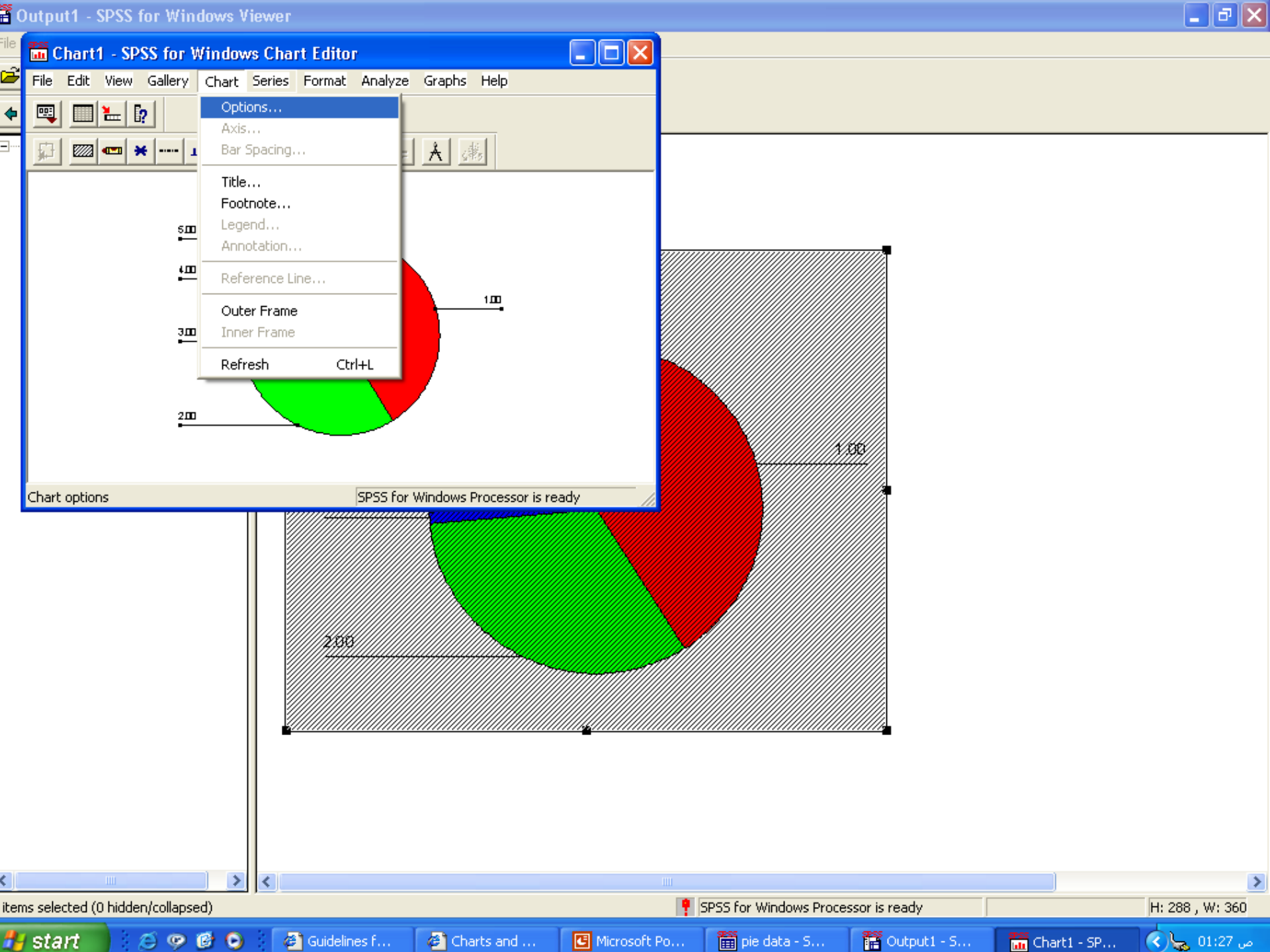


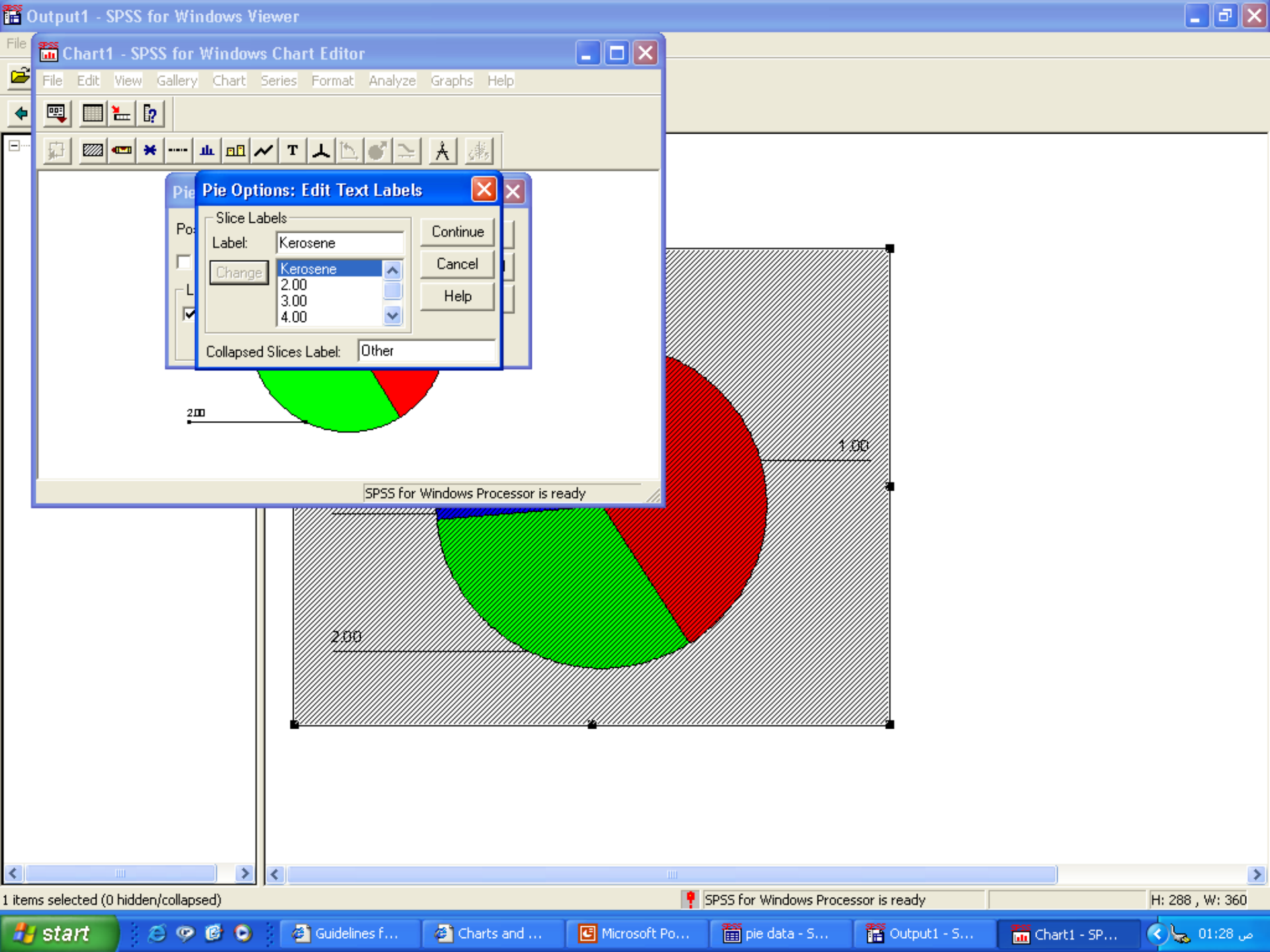
ouble click to edit Title

SPSS for Windows Processor is ready

H: 17 , W: 720

start | Guidelines for St... | Charts and Grap... | Microsoft PowerP... | pie data - SPSS f... | Output1 - SPSS f... | 01:25 ص


Output1 - SPSS for Windows Viewer
Chart1 - SPSS for Windows Chart Editor
File Edit View Gallery Chart Series Format Analyze Graphs Help
Options...
Axis...
Bar Spacing...
Title...
Footnote...
Legend...
Annotation...
Reference Line...
Outer Frame
Inner Frame
Refresh Ctrl+L
5.00
4.00
3.00
2.00
1.00
Chart options
SPSS for Windows Processor is ready
1.00
2.00
items selected (0 hidden/collapsed)
SPSS for Windows Processor is ready
H: 288 , W: 360
start
Guidelines f...
Charts and ...
Microsoft Po...
pie data - S...
Output1 - S...
Chart1 - SP...
01:27 ص

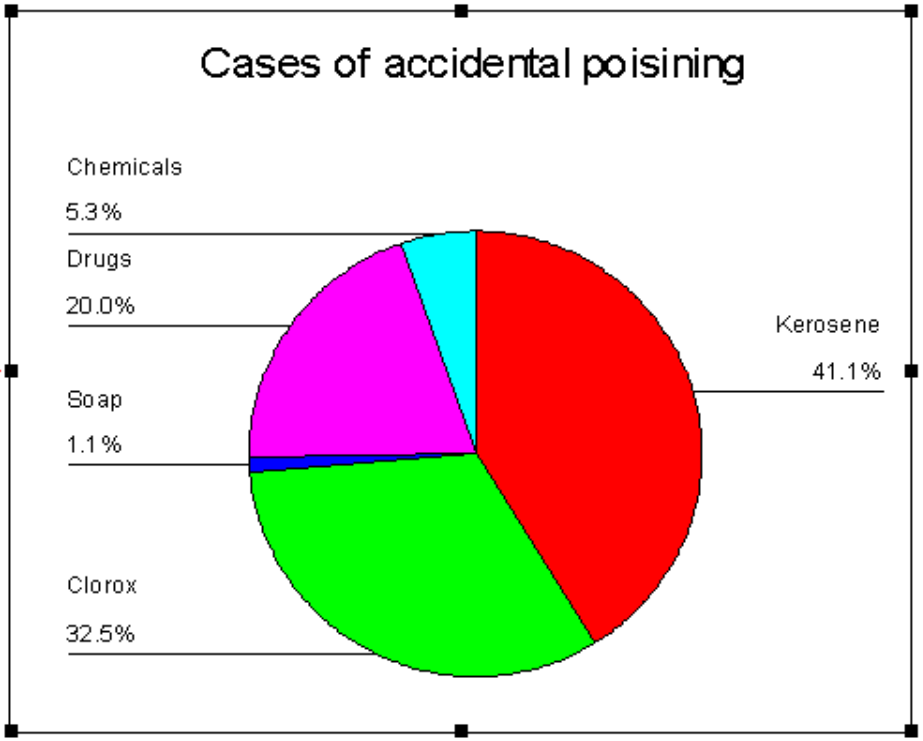

Output1 - SPSS for Windows Viewer
Chart1 - SPSS for Windows Chart Editor
File Edit View Gallery Chart Series Format Analyze Graphs Help
Pie Options: Edit Text Labels
Slice Labels
Label: Kerosene
Change
Kerosene
2.00
3.00
4.00
Continue
Cancel
Help
Collapsed Slices Label: Other
2.00
SPSS for Windows Processor is ready
1.00
2.00
1 items selected (0 hidden/collapsed)
SPSS for Windows Processor is ready
H: 288 , W: 360
start
Guidelines f...
Charts and ...
Microsoft Po...
pie data - S...
Output1 - S...
Chart1 - SP...
01:28 ص

Output1 - SPSS for Windows Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

- Output
  - Graph
    - Title
    - Notes
    - Pie of value(freq) by poi

Graph



SPSS for Windows Processor is ready

H: 288 , W: 360

start | Guidelines for St... | Charts and Grap... | Microsoft PowerP... | Untitled - SPSS f... | Output1 - SPSS f... | 01:21 ص

Scatter Plots

Scatter plots are often used to visualize a correlation between two parameters. For example, if one wished to see whether there was a relationship between the daily average percent relative humidity and the daily average temperature (in c) we could construct a scatter diagram.

for a sample of days taken in the Qassim region [Moustafa et al. (1978)]

Untitled - SPSS for Windows Data Editor

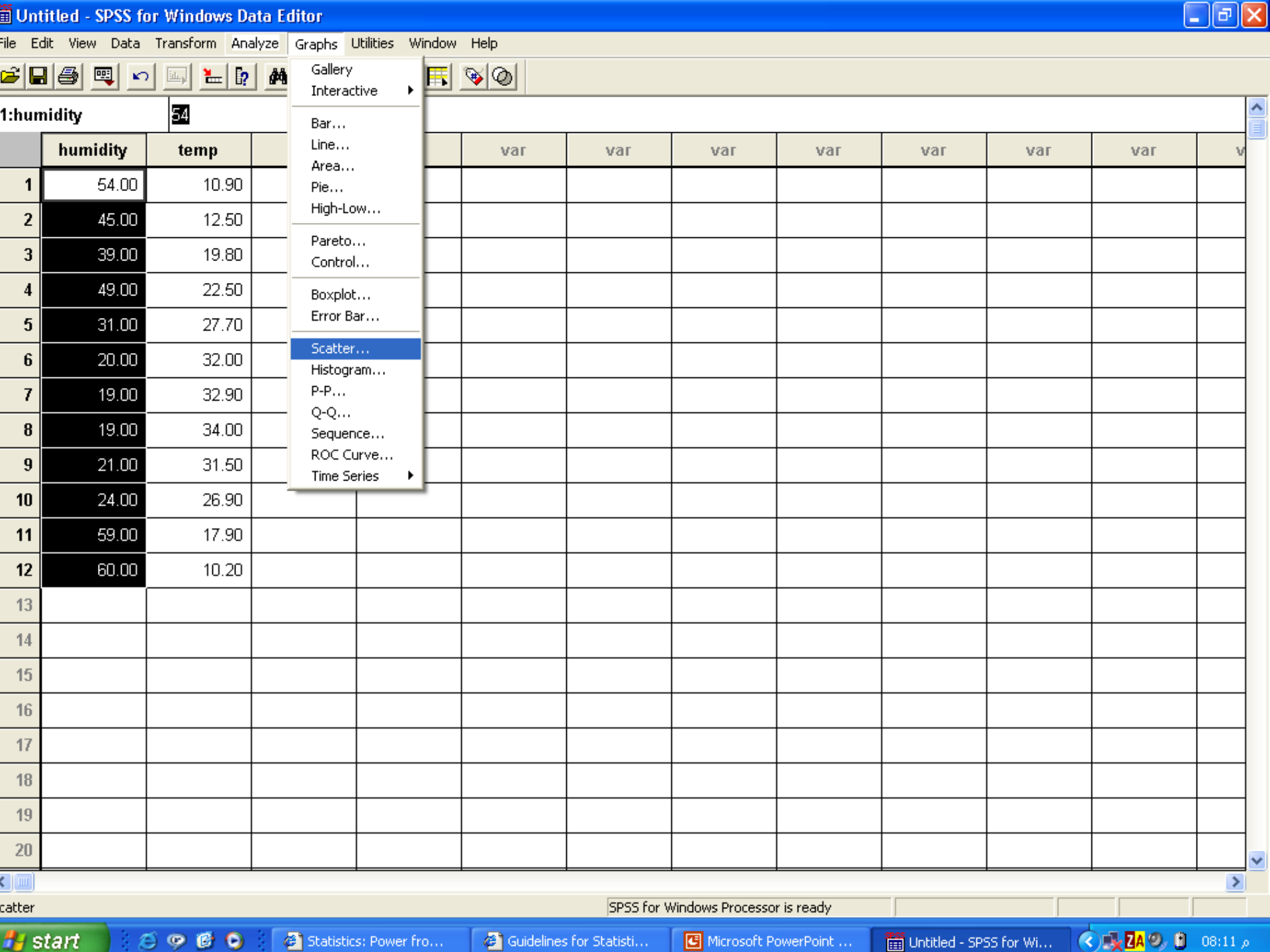
File Edit View Data Transform Analyze Graphs Utilities Window Help

1:humidity 54

	humidity	temp	var	var	var	var	var	var	var	var	var	var
1	54.00	10.90										
2	45.00	12.50										
3	39.00	19.80										
4	49.00	22.50										
5	31.00	27.70										
6	20.00	32.00										
7	19.00	32.90										
8	19.00	34.00										
9	21.00	31.50										
10	24.00	26.90										
11	59.00	17.90										
12	60.00	10.20										
13												
14												
15												
16												
17												
18												
19												
20												

SPSS for Windows Processor is ready

start Statistics: Power fro... Guidelines for Statisti... Microsoft PowerPoint ... Untitled - SPSS for Wi... 08:10


Untitled - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Gallery
Interactive
Bar...
Line...
Area...
Pie...
High-Low...
Pareto...
Control...
Boxplot...
Error Bar...
Scatter...
Histogram...
P-P...
Q-Q...
Sequence...
ROC Curve...
Time Series
1:humidity
54
humidity
temp
var
1 54.00 10.90
2 45.00 12.50
3 39.00 19.80
4 49.00 22.50
5 31.00 27.70
6 20.00 32.00
7 19.00 32.90
8 19.00 34.00
9 21.00 31.50
10 24.00 26.90
11 59.00 17.90
12 60.00 10.20
SPSS for Windows Processor is ready
start
Statistics: Power fro...
Guidelines for Statisti...
Microsoft PowerPoint ...
Untitled - SPSS for Wi...
08:11

Untitled - SPSS for Windows Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1:humidity 54

	humidity	temp	var	var	var	var	var	var	var	var	var	var
1	54.00	10.90										
2	45.00	12.50										
3	39.00	19.80										
4	49.00	22.50										
5	31.00	27.70										
6	20.00	32.00										
7	19.00	32.90										
8	19.00	34.00										
9	21.00	31.50										
10	24.00	26.90										
11	59.00	17.90										
12	60.00	10.20										
13												
14												
15												
16												
17												
18												
19												
20												

Scatterplot

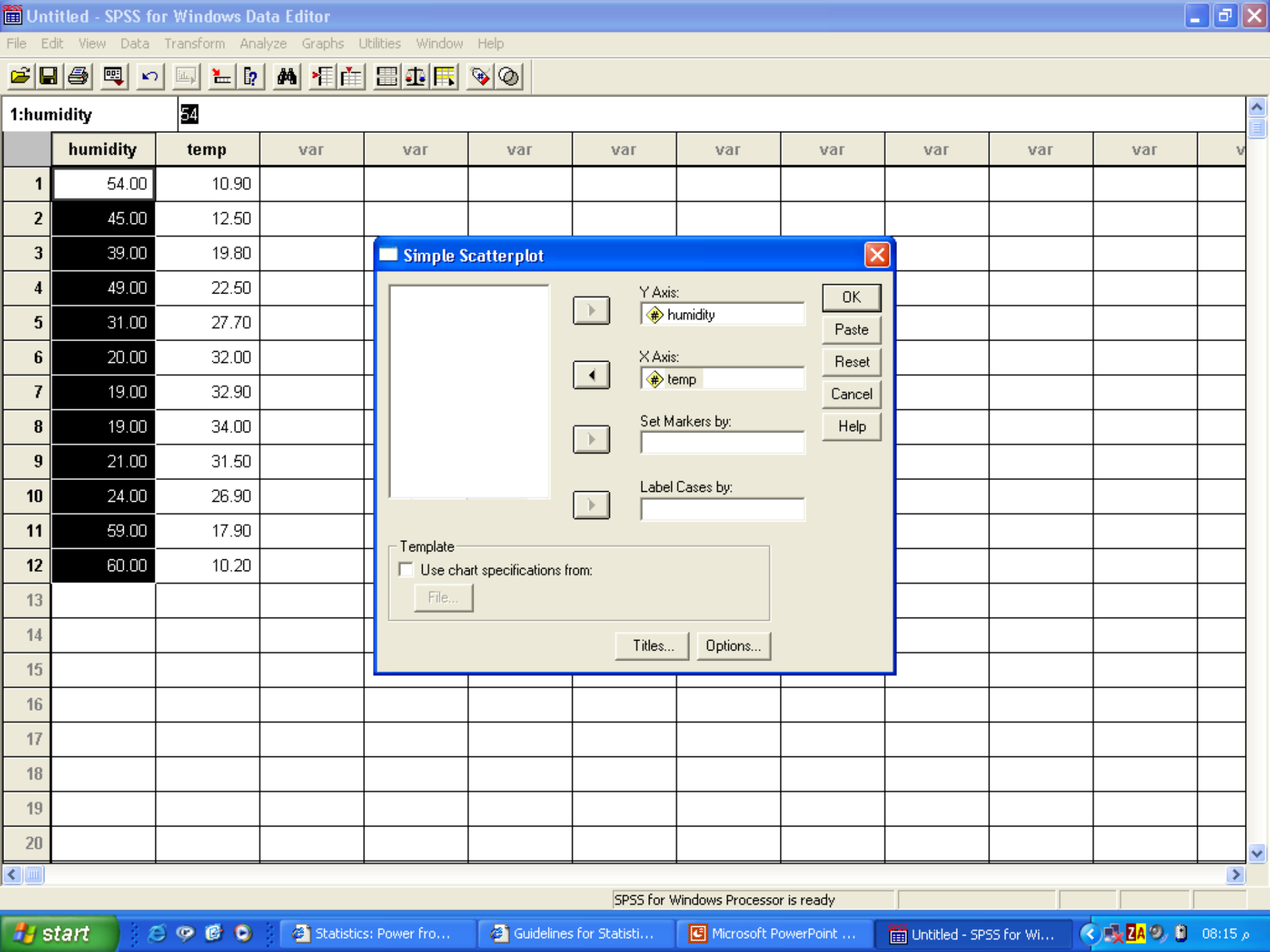
Simple Matrix

Overlay 3-D

Define Cancel Help

SPSS for Windows Processor is ready

start Statistics: Power fro... Guidelines for Statisti... Microsoft PowerPoint ... Untitled - SPSS for Wi... 08:11

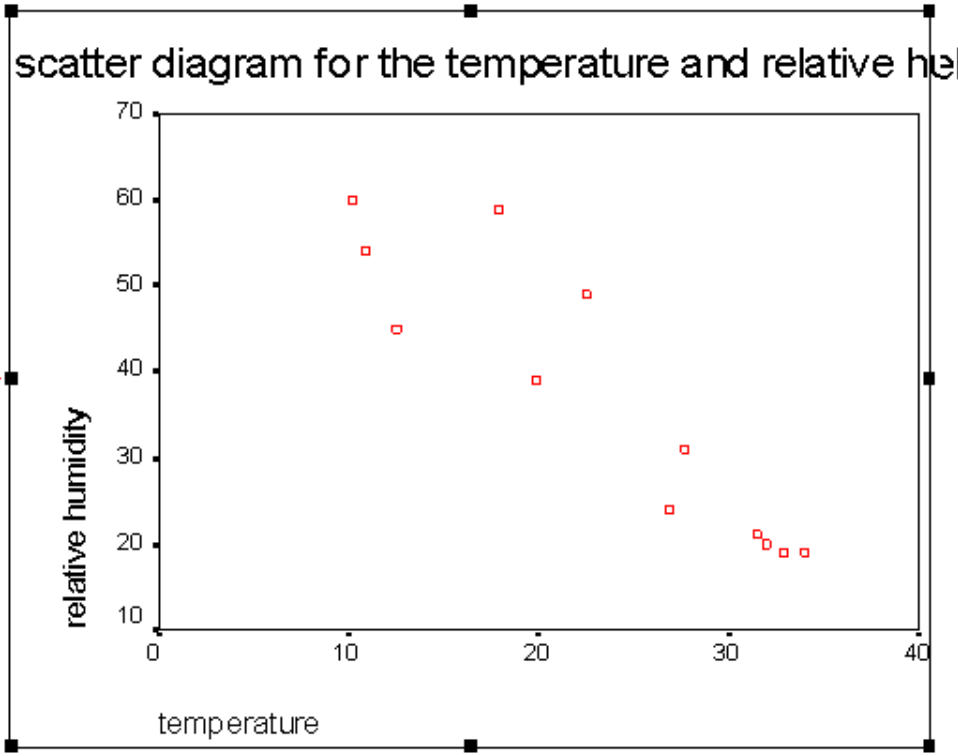

Untitled - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1:humidity
54
humidity temp var var var var var var var var var var
1 54.00 10.90
2 45.00 12.50
3 39.00 19.80
4 49.00 22.50
5 31.00 27.70
6 20.00 32.00
7 19.00 32.90
8 19.00 34.00
9 21.00 31.50
10 24.00 26.90
11 59.00 17.90
12 60.00 10.20
13
14
15
16
17
18
19
20
Simple Scatterplot
Y Axis:
humidity
X Axis:
temp
Set Markers by:
Label Cases by:
OK
Paste
Reset
Cancel
Help
Template
Use chart specifications from:
File...
Titles...
Options...
SPSS for Windows Processor is ready
start
Statistics: Power fro...
Guidelines for Statisti...
Microsoft PowerPoint ...
Untitled - SPSS for Wi...
08:15

Output1 - SPSS for Windows Viewer

File Edit View Insert Format Analyze Graphs Utilities Window Help

- Output
  - Graph
    - Title
    - Notes
    - Scatter of scatter diagra

Graph



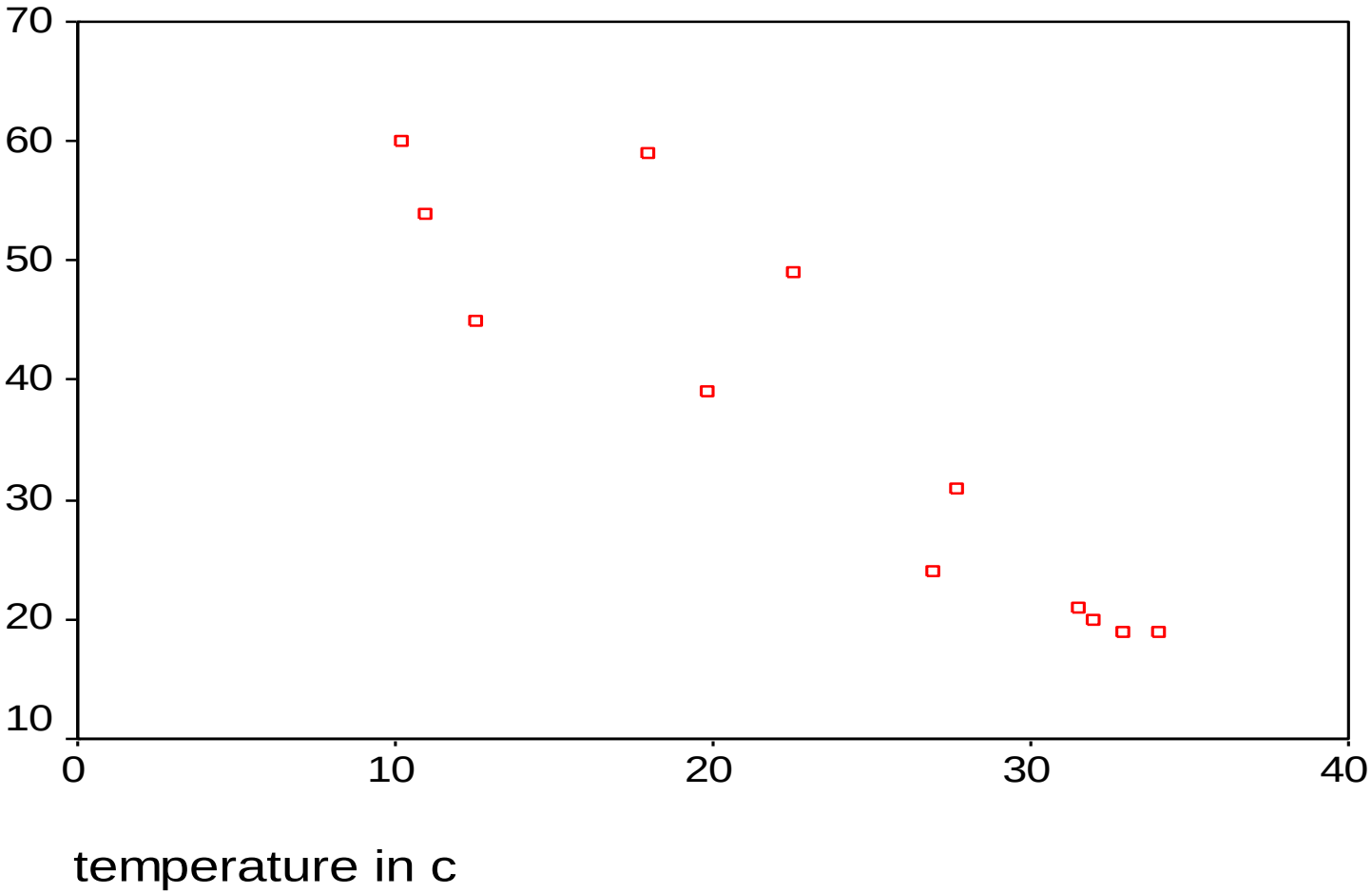
1 items selected (0 hidden/collapsed)

SPSS for Windows Processor is ready

H: 294 , W: 368

start | 2 Internet Explorer | Microsoft PowerPoint ... | Untitled - SPSS for Wi... | Output1 - SPSS for W... | 08:18

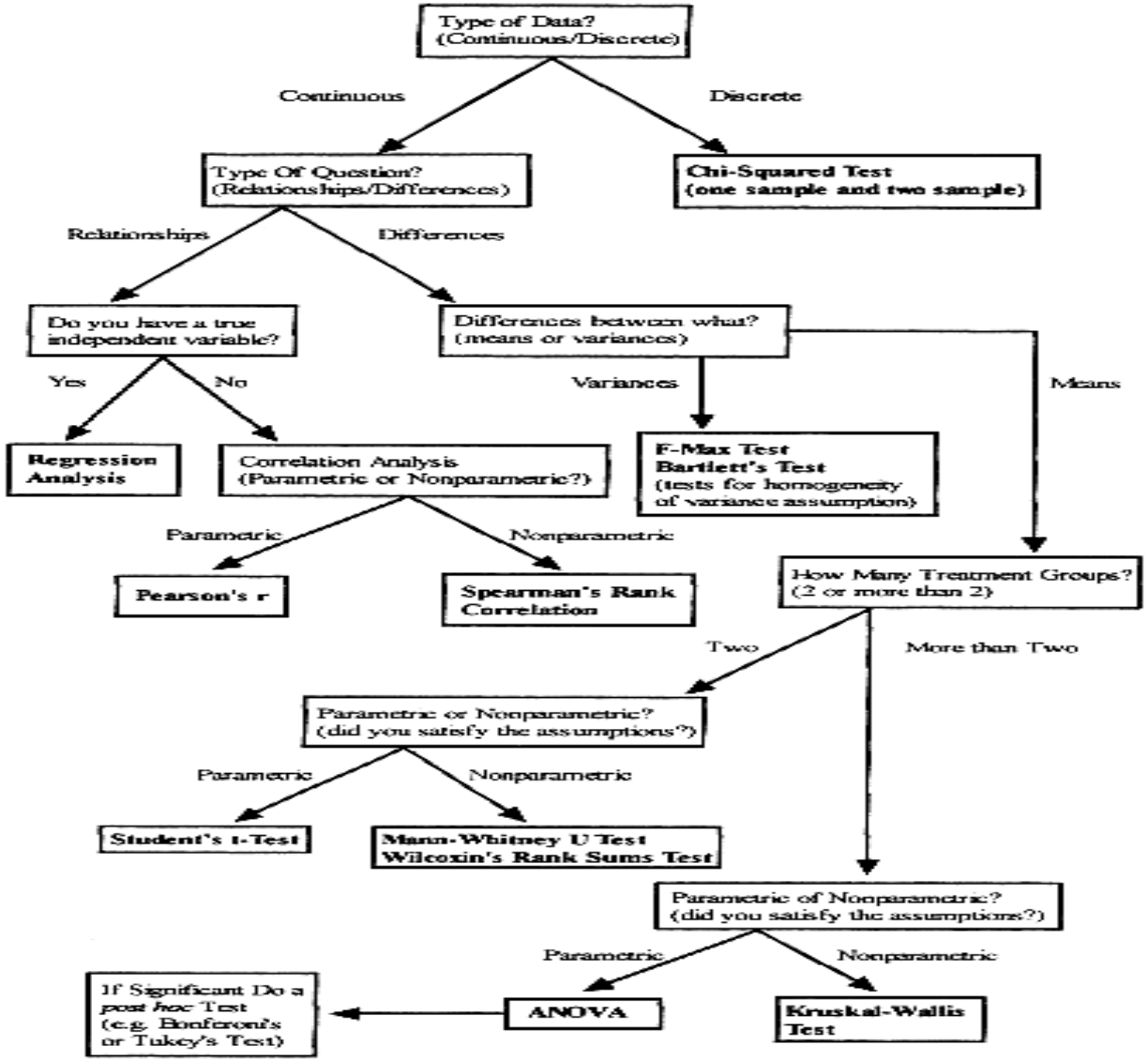
scatter diagram for the temperature and relative



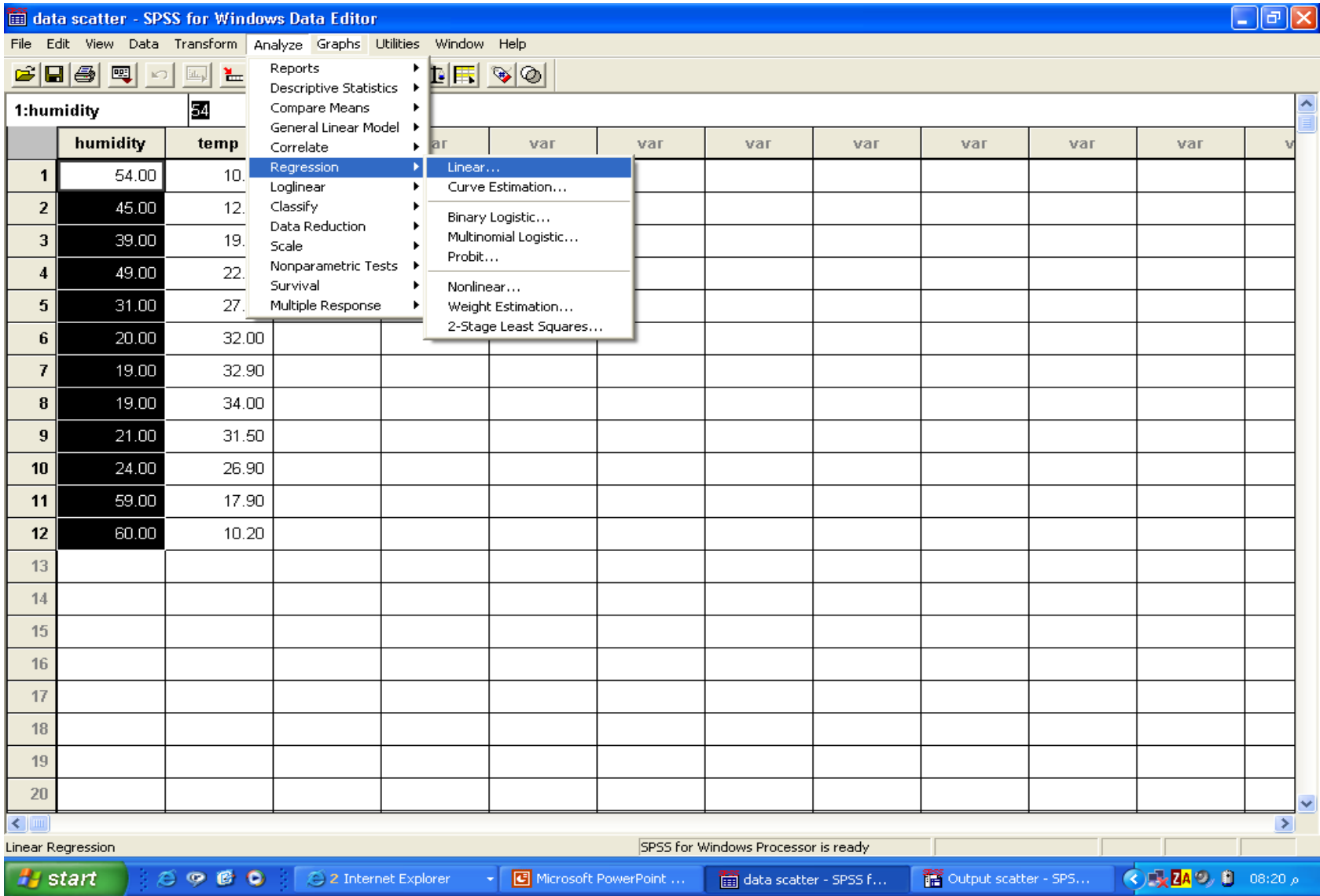
Analyzing the data

Your choice of statistical analysis should be made prior to conducting an experiment. There is little sense in collecting data that you can't analyze properly. Use the following flow chart to help you decide which statistic to use.

Hypothesis Test Decision Making Flow Chart



Linear Regression



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.915[a]	.836	.820	6.9016

a. Predictors: (Constant), TEMP

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2434.346	1	2434.346	51.107	.000[a]
	Residual	476.320	10	47.632		
	Total	2910.667	11			

a. Predictors: (Constant), TEMP

b. Dependent Variable: HUMIDITY

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	75.622	5.802		13.034	.000
	TEMP	-1.677	.235	-.915	-7.149	.000

a. Dependent Variable: HUMIDITY

Linear Regression

We can predict now the percent relative humidity for any given temperature using the following equation:

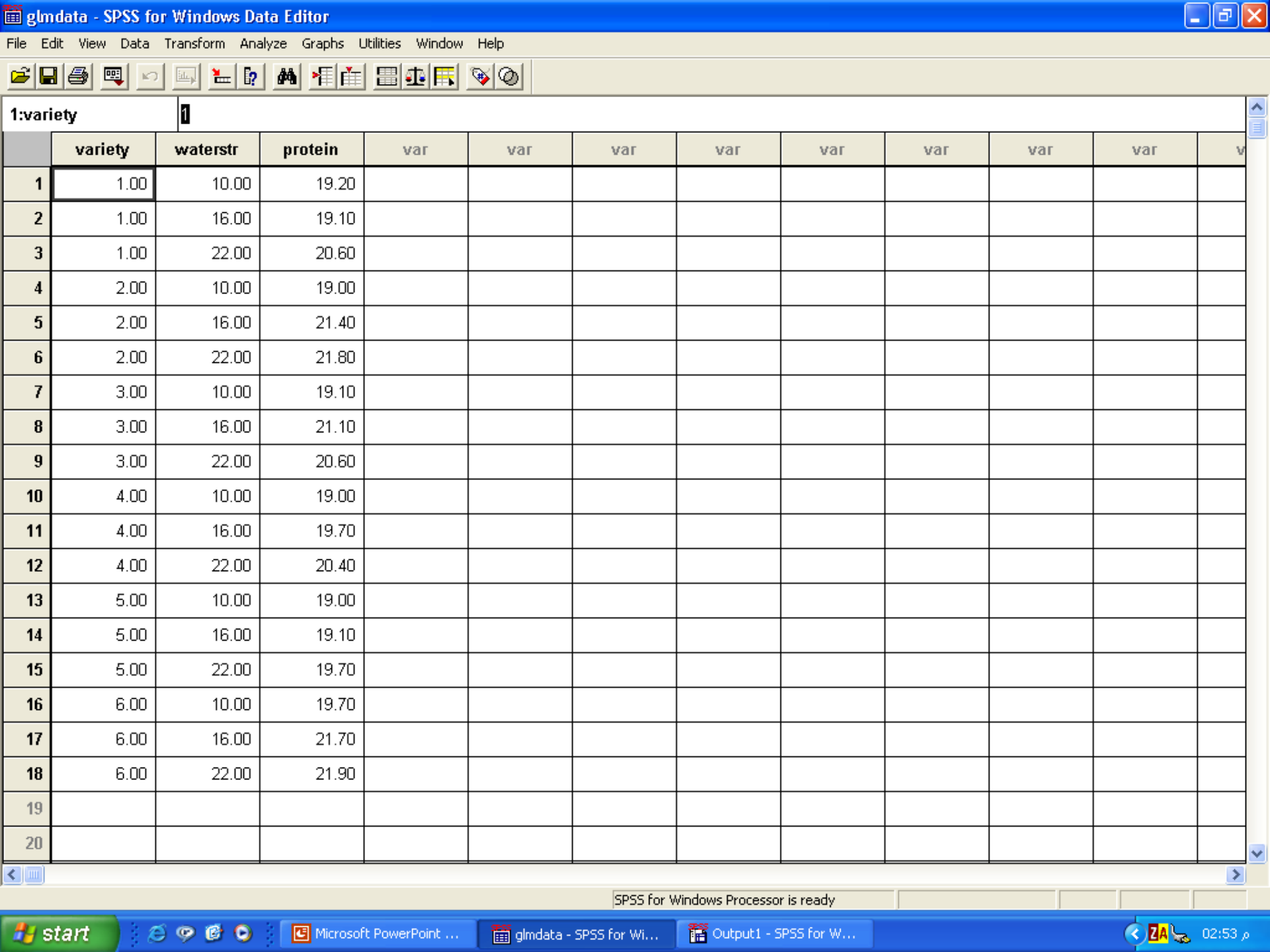
$Y = 75.62 - 1.68 * X$

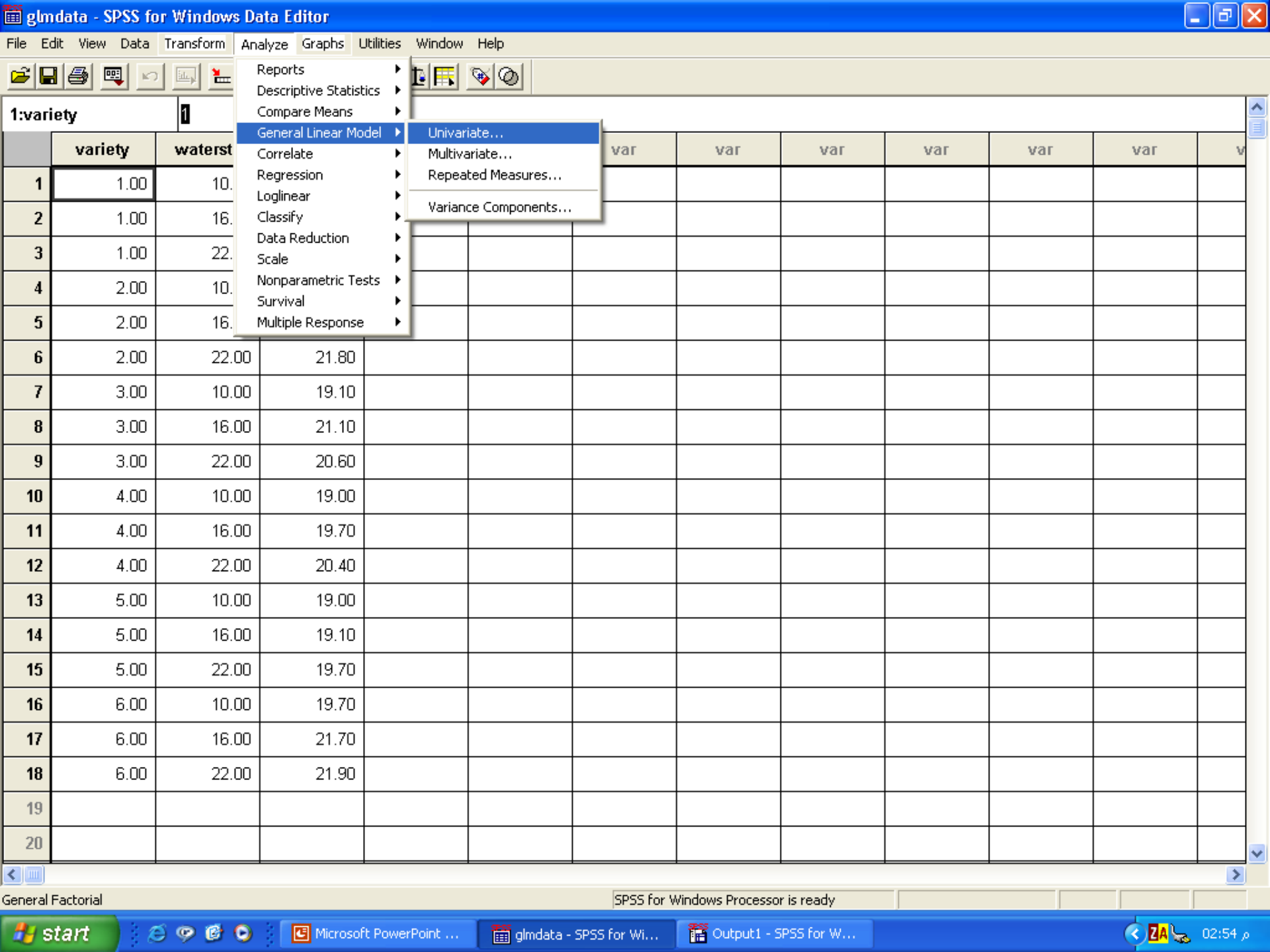
Where X=temperature,Y=relative humidity.

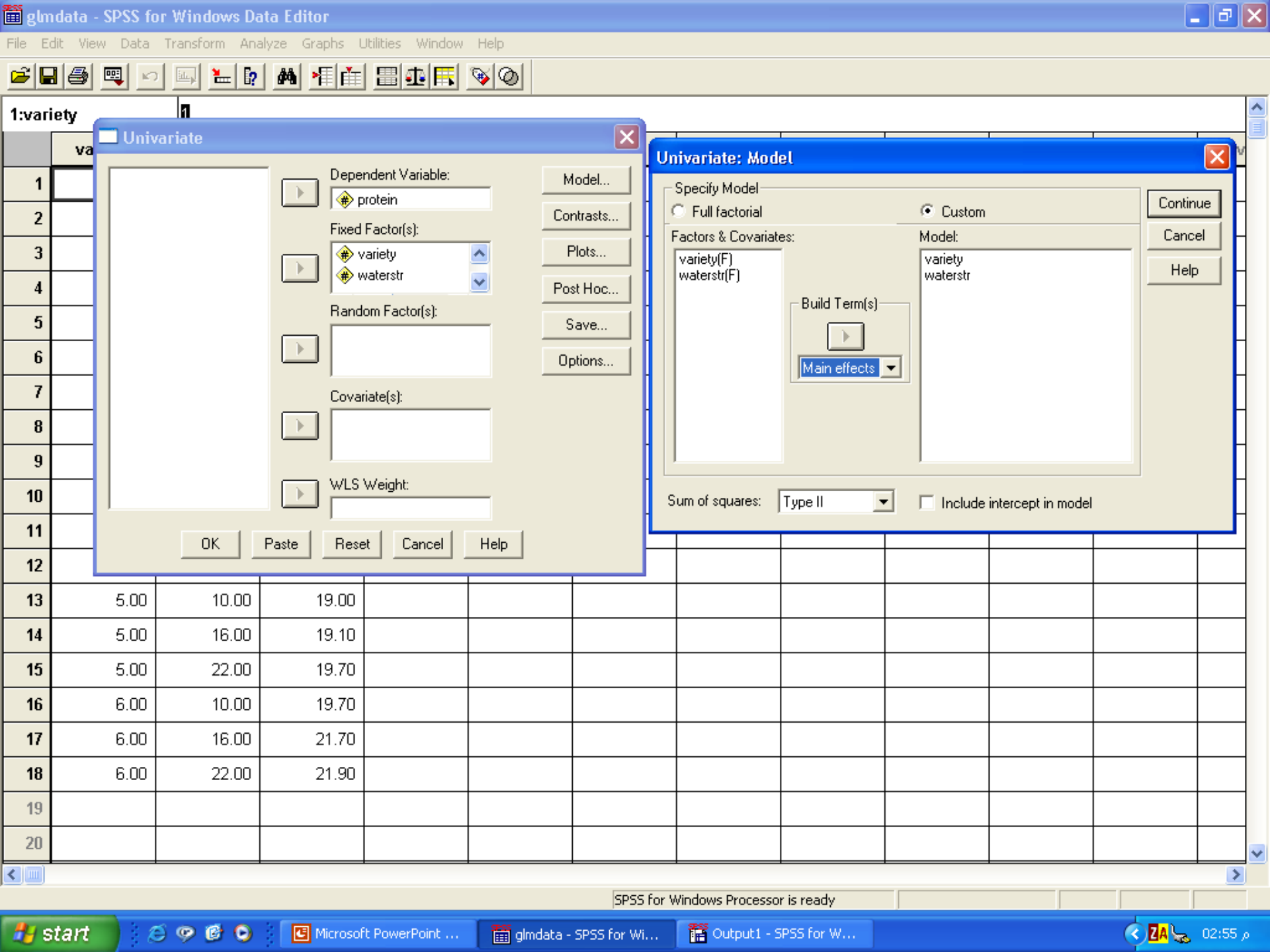
Extra Example on Blocking

In a study it was desired to know the effect of water stress on the protein content of wheat.

Because the protein content of wheat is known to differ from one variety to another, six local varieties of Saudi wheat were chosen for the experiment and it was assumed that there is no interaction between the wheat varieties and the water stress levels on the protein content. Therefore, three plots of each type of wheat were chosen and randomly assigned to the three levels of water stress, namely three watering intervals of every 10, 16, and 22 days. After harvest, the wheat from each plot was separately ground into flour, and the protein content (as a percent of the dry weight) was measured [Based on Basahy (1990)].


glmdata - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1:variety
1
variety waterstr protein var var var var var var var var var
1 1.00 10.00 19.20
2 1.00 16.00 19.10
3 1.00 22.00 20.60
4 2.00 10.00 19.00
5 2.00 16.00 21.40
6 2.00 22.00 21.80
7 3.00 10.00 19.10
8 3.00 16.00 21.10
9 3.00 22.00 20.60
10 4.00 10.00 19.00
11 4.00 16.00 19.70
12 4.00 22.00 20.40
13 5.00 10.00 19.00
14 5.00 16.00 19.10
15 5.00 22.00 19.70
16 6.00 10.00 19.70
17 6.00 16.00 21.70
18 6.00 22.00 21.90
19
20
SPSS for Windows Processor is ready
start
Microsoft PowerPoint ...
glmdata - SPSS for Wi...
Output1 - SPSS for W...
02:53


glmdata - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Reports
Descriptive Statistics
Compare Means
General Linear Model
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Survival
Multiple Response
Univariate...
Multivariate...
Repeated Measures...
Variance Components...
1:variety
variety waterst
var
1.00 10.
1.00 16.
1.00 22.
2.00 10.
2.00 16.
2.00 22.00 21.80
3.00 10.00 19.10
3.00 16.00 21.10
3.00 22.00 20.60
4.00 10.00 19.00
4.00 16.00 19.70
4.00 22.00 20.40
5.00 10.00 19.00
5.00 16.00 19.10
5.00 22.00 19.70
6.00 10.00 19.70
6.00 16.00 21.70
6.00 22.00 21.90
General Factorial
SPSS for Windows Processor is ready
start
Microsoft PowerPoint ...
glmdata - SPSS for Wi...
Output1 - SPSS for W...
02:54


glmdata - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1:variety
Univariate
Dependent Variable:
protein
Fixed Factor(s):
variety
waterstr
Random Factor(s):
Covariate(s):
WLS Weight:
Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...
OK
Paste
Reset
Cancel
Help
Univariate: Model
Specify Model
Full factorial
Custom
Factors & Covariates:
variety(F)
waterstr(F)
Build Term(s)
Main effects
Model:
variety
waterstr
Continue
Cancel
Help
Sum of squares:
Type II
Include intercept in model
5.00 10.00 19.00
5.00 16.00 19.10
5.00 22.00 19.70
6.00 10.00 19.70
6.00 16.00 21.70
6.00 22.00 21.90
SPSS for Windows Processor is ready
start
Microsoft PowerPoint ...
glmdata - SPSS for Wi...
Output1 - SPSS for W...
02:55

Output
Univariate Analysis of Variar
Title
Notes
Between-Subjects Facto
Tests of Between-Subje

Univariate Analysis of Variance

Between-Subjects Factors

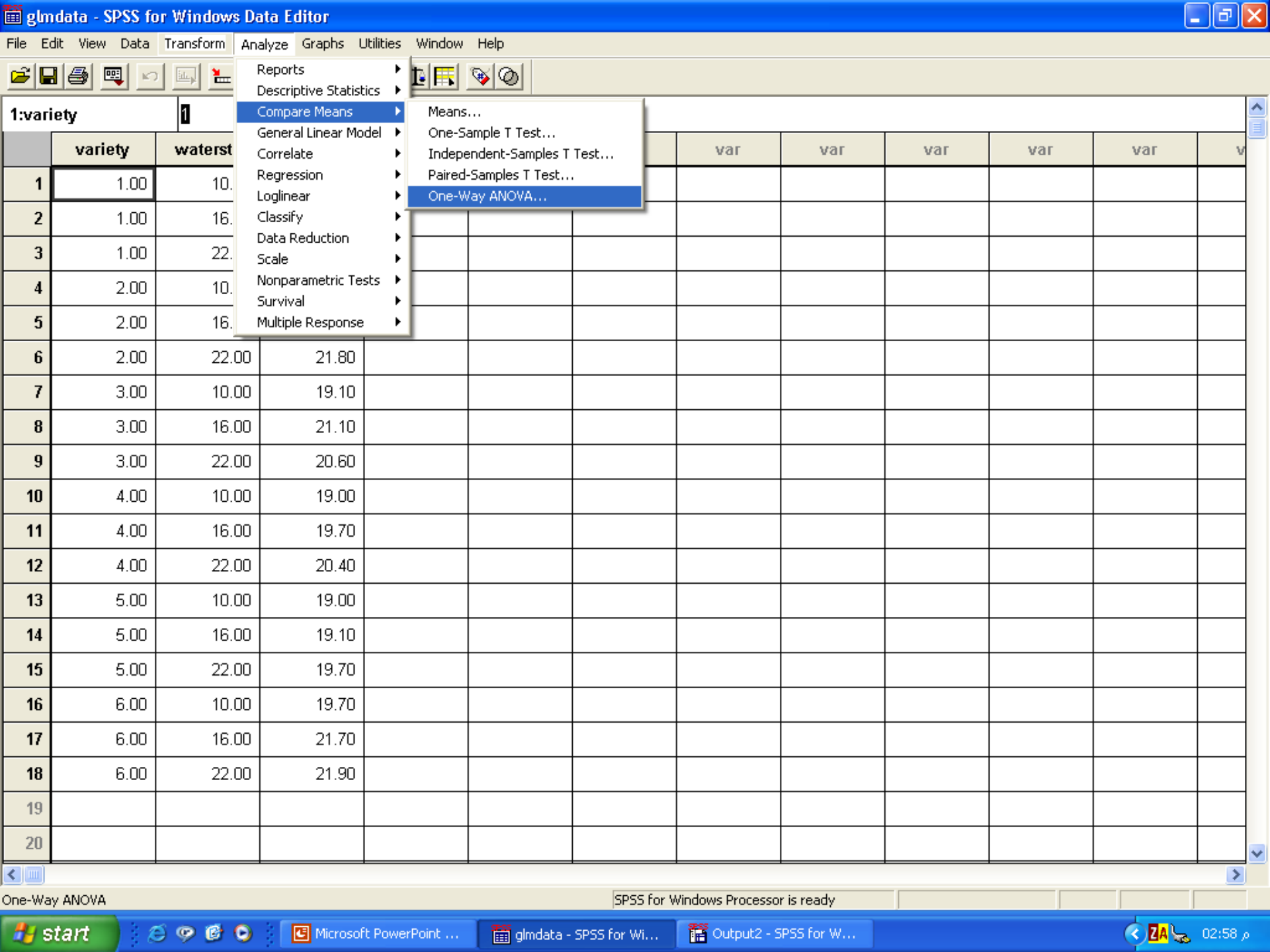
		N
VARIETY	1.00	3
	2.00	3
	3.00	3
	4.00	3
	5.00	3
	6.00	3
WATERSTR	10.00	6
	16.00	6
	22.00	6

Tests of Between-Subjects Effects

Dependent Variable: PROTEIN

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Model	7300.567[a]	8	912.571	2561.003	.000
VARIETY	7.498	5	1.500	4.209	.025
WATERSTR	8.823	2	4.412	12.381	.002
Error	3.563	10	.356		
Total	7304.130	18			

a. R Squared = 1.000 (Adjusted R Squared = .999)


glmdata - SPSS for Windows Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Reports
Descriptive Statistics
Compare Means
General Linear Model
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Survival
Multiple Response
Means...
One-Sample T Test...
Independent-Samples T Test...
Paired-Samples T Test...
One-Way ANOVA...
1:variety
variety waterst
var var var var var
1 1.00 10.
2 1.00 16.
3 1.00 22.
4 2.00 10.
5 2.00 16.
6 2.00 22.00 21.80
7 3.00 10.00 19.10
8 3.00 16.00 21.10
9 3.00 22.00 20.60
10 4.00 10.00 19.00
11 4.00 16.00 19.70
12 4.00 22.00 20.40
13 5.00 10.00 19.00
14 5.00 16.00 19.10
15 5.00 22.00 19.70
16 6.00 10.00 19.70
17 6.00 16.00 21.70
18 6.00 22.00 21.90
19
20
One-Way ANOVA
SPSS for Windows Processor is ready
start
Microsoft PowerPoint ...
glmdata - SPSS for Wi...
Output2 - SPSS for W...
02:58

Results

Using the randomized block design there is a significant difference in the protein content using different levels of water stress while using the simple one way anova there is no significant difference at α=0.01.

Tests of Between-Subjects Effects

Dependent Variable: PROTEIN

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Model	7300.567[a]	8	912.571	2561.003	.000
VARIETY	7.498	5	1.500	4.209	.025
WATERSTR	8.823	2	4.412	12.381	.002
Error	3.563	10	.356		
Total	7304.130	18			

a. R Squared = 1.000 (Adjusted R Squared = .999)

ANOVA

PROTEIN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.823	2	4.412	5.982	.012
Within Groups	11.062	15	.737		
Total	19.885	17			