

Distr.
GENERAL

ECE/CES/SEM.54/10
18 May 2006

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Joint UNECE/Eurostat/OECD Seminar on the Management of Statistical Information Systems (MSIS)
Sofia, Bulgaria, 21-23 June 2006

Topic (i): Changes in statistical processes

**SURVEY ON IT ORGANIZATION AND STANDARDS IN NATIONAL STATISTICAL
INSTITUTES**

Supporting Paper prepared by Maura Giacommo and Carlo Vaccari, Istat, Italy

SURVEY ON IT ORGANIZATION AND STANDARDS IN NATIONAL STATISTICAL INSTITUTES

Supporting Paper prepared by Maura Giacommo and Carlo Vaccari, Istat, Italy
maura.giacummo@istat.it, carlo.vaccari@istat.it

1. Introduction

2. Data Collection

Data collection has been done using Data Capturing via web, considering the kind of interviewees and the survey's object.

Compared to a classic postal survey, an electronic on-line questionnaire has the following advantages:

- survey techniques is very cheap;
- data collection and entry are united;
- a constant survey and answering monitoring is possible.

Besides, most of times interviewee's subjects are National Statistic Institutes' s IT manager, so strongly oriented to use the possibilities of data transmission offered by Internet.

Electronic survey has improved the opportunity to access questionnaires giving better times of response.

We decided to use an electronic survey with controlled access managed with token; this permitted to overcome the bound about interviewee's identity and answer's duplication; the token, besides, allowed to find in a unique way the answering and to stop the access after data has been sent .

An e-mail is sent to explain the survey's purposes and to point out the survey's ISTAT responsible to contact for any questions or technical problems.

A great support to the survey has been given by International Relations and Cooperation Office (RIN) that supplied an update and exhaustive list of IT head's mail address.

From a technical point of view is important to underline that this survey has been totally managed with open source tools : the questionnaire has been realized with PhpSurveyor while data are collected in a MySql Data Base.

3.Questionnaire description

The questionnaire proposed to National Statistic Institutes is divided into five parts with 17 questions in total , ten of these need a mandatory closed answer.

Following the sections in which the survey is divided are described:

A) Human resources

In the first section is required to describe which are the human and technological resources available in National Statistic Institute. These permit to find out informations both on Institute' s size and importance and on how many different professionalism it has.

Information on the amount of installed clients and servers and on the number of employees, of software developers and engineers is required in the survey. .

The mandatory closed answers has been oriented to catch the variable's wideness.

B) IT Organization

the focus of this section is on IT organization, moving the attention from the whole Institute to the single IT structure. Information about software development method (centralised, decentralised or both) and allocation of people who developed it (temporary to projects, fixed to a particular survey, mixed) have been collected.

The last question of this section requires information on the presence of IT unit dedicated to some of the survey steps (e.g Data Capturing, Check, Dissemination, ...)

In this part we try to understand IT work's organization and the interaction among IT and other units, the attention is both on software development and on the whole National Statistic Institute survey 's management. A different work organization can affect the steps of statistical data production, both on data quality and on timeliness in delivery. The three questions of this sections are mandatory and closed, there is however the possibility to add some clarifications about the answers.

C) IT Standard

The third section collect informations on on IT standard applied on Operational System Management, Data Base Management, Programming Languages and Statistical Packages. Every question has a list of the software technology mainly used in ISTAT, for each element of the list it is necessary to show the usage level in the National Statistic Institute (standard, widely Used, Seldom used, not used).

The focus is on understanding the other National Statistic Institute 's real utilization of the software technologies used in Istat.

D) Generalized software

A special section is dedicated to generalized software used in the different steps of statistical data production: for each step National Statistic Institute could fill the generalized software used. The answer to these questions is not mandatory because the National Statistic Institute could decide not to use generalized software for all steps of a statistical survey

E) Miscellaneous

This session try to find out the areas of interest of each institute and the need of creating communication tools for exchanging of ideas. Some communication tools are : mailing list, periodic seminar, Web site with Forum, a specific network in the field.

4. Interviewees Analysis

International Relations and Cooperation Office (RIN), as outlined in the previous paragraph, has

given the National Statistic Institute list.

This survey has been led on thirty six National statistics institute, twenty eight of these have taken part in the survey. We have decided to contact mainly European National statistic institute and four extra-European countries. These are: Australia Canada Japan ad United States, they represent the most progressive countries from a statistical point of view and the most present in the international relationship.

Canada is the only one that didn't answered to the survey.

For Bosnia-Erzegovina and Serbia_Montenegro the questionnaire has been sent both to federal statistic office and to federal republic offices.

Istat has been added to the sampling, this has permitted to compare the technological choices done in our country with those of other Institutes , other comparative analysis are also possible.

The survey Interviewees in most of case have been the Information Technology heads , for the few countries that don't have a specific Information Technology unit, the Institute responsible have been contacted.

The survey doesn't have an expires date, this has allowed to send some reminder emails and to find the right interviewee. This choice has improved the survey from a qualitative and a quantitative point of view. Reminder emails increased the number of answers by five units.

The answer rate has been almost seventy eight per cent, this rate has been reached using two different reminder procedure: an email and a direct phone call. The first one has been done twice, the first time after about forty days from the email notification, while the second reminder email has been sent after few days from the first one. This choice has been done because the answer to the survey has been sent few days after the arrival of questionnaire while according longer time for return didn't give important increases in the answers.

Reminder email has been written in a more confidential way than the notification email, being sent directly to the referent, besides it underlined the importance given to the survey from Istat. In particular cases RIN has directly phoned some interviewees who didn't return the questionnaire.

The notification email contains an address to write for assistance in questionnaire compilation. Some interviewees have contacted both to ask some clarifications about survey purpose and to have suggests about compilation procedure.

The countries which have answered the questionnaire are: Albania, **Australia**, Austria, Czech Republic, Cyprus, Denmark , Estonia, Finland, France, Germany, Japan, Greece, Ireland, Italy, Latvia, Montenegro, Norway, Netherlands, Portugal, United Kingdom, Srpska Republic, Serbia and Montenegro, Serbia, Slovakia, Slovenia, Spain, United States of America, Sweden, while Canada, Lithuania, Luxembourg, Poland, Turkey, Hungarian, Bosnia -- Erzegovina and Bosnia-Erzegovina federation didn't return the questionnaire.

5. Data Analysis

The appendix includes both simple frequency distribution and cross class distribution by the Institutes' size. National Statistic Institute have been divided in three different classes:

- small: eight¹ Institute, from 1 to 499 number of clients;

¹ Albania, Cyprus , Estonia , Finland , Montenegro, Serbia and Montenegro , Slovenia, Srpska ,

- medium: eleven Institute², from 500 to 1500 number of clients;
- big: nine Institute³, more than 1500 number of clients.

The variable “number of clients” is a good indicator for measuring the Institutes' size. Other variables such as “Number of employees” gave similar results but with a major data variability.

5.1 Server Operating System

Every National Statistic Institute uses Windows as Operating System, see appendix tables 6.2.1.4 6.2.1.5 6.2.1.6 (only fourteen per cent of them uses it rarely), this trend is completely independent from the Institute size.

other operating systems have a very similar utilization rate: Unix and Linux have a 60% cumulative percentage, while Mainframe has 46.4 %.

A dimensional analysis, appendix tables 6.2.1.11 6.2.1.12 6.2.1.13, shows a greater utilization rate of Unix and Linux as the number of clients increases, with a light prevalence of Unix.

It's interesting to note that the server number it's strongly dependent from the choice of the standard operating system: if the standard is Windows the average servers number is 150 while this number decreases to 51 if the standard is Unix. This difference could be explained both from the greater power of Linux server and because they, probably, have been installed more recently and on more powerful hardware.

The use of Linux and Unix in medium and big institutes could underline a propensity to use these operating systems on servers and Windows only on clients.

The Mainframe utilization's rate is about 36 percent in small medium Institutes (see appendix table 6.2.2.1.1), they use this Operating System very frequently, in fact the item “Seldom Used” is equal to zero.

Mainframe is more present in big Institutes though one third of the Institute uses it quite seldom: this percentage could be explained by a migration towards other operating systems in big Institutes and by a standard utilization in small a medium Institutes that use Mainframe.

Following the Operating System utilization trend would be very interesting: our data permit to assume a migration process towards Linux and Windows in Institutes that used Mainframe or Unix.

Medium Institutes seem to apply conservatory politics, being still linked to previous generation systems and languages.

5.2 SERVER Data Base Management Systems

The DBMS diversification, appendix tables 6.2.1.8-6.2.1.12, is coherent with the operating systems' one; Ms SQL is the most used DBMS with a cumulative percentage equal to 82, table 6.2.1.10 , and it runs on Windows platform. The others DBMS, except Oracle, are rare in the National Statistic Institutes, their cumulative utilization rate is between 39 per cent (MySQL) and 21 percent (IBM DB2/Informix), tables 6.2.1.9 and 6.2.1.12.

DBMS's choice is highly correlated to the Institutes' size, table shows a standard utilization of MS SQL Server in small Institutes. With the Increasing of Institutes' size we note a shift towards Oracle: all big Institutes managed it and seventy-seven percent uses it in a standard way, while MS SQL Server utilization's rate decreases.

MySQL has been chosen in big Institutes, only two out of nine don't manage it, even if it seems an experimental usage as four users out of seven use it seldom.

5.3 Programming languages

² Austria, Danmark, Greece, Ireland, Japan, Latvia, Norway, Portugal, Serbia, Slovakia, Sweden

³ Australia, Czech, France, Germany, Italy, Netherlands, Spain, United Kindom, United States

Frequencies tables underline a wide use of all the programming languages included in the survey; utilization's rate of each of these is always over 60 percent, Perl and Python have a light decrease, in fact their utilization's rate is 46 per cent.

The 3GL Languages, such as Fortran and Cobol, are still used even if 39 per cent of the Institutes uses them seldom, .

Java and Visual Basic have been used in production ("widely used" and "standard") with a utilization's rate respectively of 64 and 60 per cent, tables 6.2.1.14 and 6.2.1.15.

Visual Basic is used independently of the Institutes' size, while other languages have a different rate dependent on the size: C and C++, for example, are widely used in big Institutes while medium Institutes prefer 3GL languages. This last result was expected in fact it's strictly linked to the choice of Mainframe as Operating System. Java (64 per cent) is used more than PHP (28.6 per cent) to develop web application, even if there's a great difference between big and small Institutes: the first have Java and PHP utilization's rate respectively of 100 and 80 per cent, while small ones use both languages rarerly.

Data show a use of the programming languages strictly related to the type of application to develop (web/Client-server/traditional) and to the Operating System.

5.4 Statistical Packages

SAS and SPSS are the statistical packages more used, their cumulative utilization's percent is respectively of 85 and 75 per cent, tables 6.2.1.19 6.2.1.20, versus R cumulative percent of 17 per cent for R and cumulative percent of 46 for Stata.

Dimensional classes analysis show a major use of SPSS in medium and small Institutes, where SAS, when present, is widely or standard used. This choice is due to SAS's higher cost, consequently this statistical package is used as standard or it's not used at all.

While R and Stata are seldom used and they appear only in big Institutes.

5.5 Generalized Software

Questions on generalized software, for different steps of statistical data production, have required an open and optional answer. This choice has been done to capture all type of information even if this caused more problems in answer understanding. Answers have been classified in four different types: In house development software, none software, mainly generalized software, others. Data show that the 42 percent of Institutes don't use any software for Survey design, this percentage decreases in the other steps. In most cases there is a preference towards commercial or free software, in fact this percentage, in the different steps, is always over 60 (sum of mainly software and others) .

Ad hoc development rate is about 15-17 per cent, depending on the steps, with a light increase in the check and correction step.

Blaise is mainly used in data capturing and survey design step while SAS manages sample definition and check and correction steps, Argus is used in statistical disclosure.

5.6 Italian environment

Istat, with 2500 clients, has been classified as a big institute. From a first comparison with other big Institutes has emerged that Istat number of servers (60) is lower then big Institute average (269) on the other hand every employee has a client.

Information technology unit has an employee number and consequently a weight greater than other big Institutes: in fact the "system engineers"(90) and "software developers"(250) are greater then big Institutes average (73 and 152) , even if the number of total employee is lower than the average.

Software development is both centralized and decentralized, this is quite different from the choice done by other big Institutes, where centralized software development has been chosen from 88 per cent of them.

Software development resources allocation is mixed, and there are some unit, inside Information Technologic, dedicated to different step of data production such as sample design, survey and questionnaire design and data capturing .

Linux has been chosen as standard operating system, this choice has given advantages both on lower product cost, since it's an open source system, and for the possibility to overcome proprietary platform restrictions.

Windows and Unix are also present in Istat but they are seldom used; Windows is used as server platform only in particular application, if a Linux corresponding toll doesn't exist, while Unix server (IBM/AIX) will be completely migrated versus Linux within one year.

In the last years Istat is pushing the use of DBMS: Oracle has been affirmed and MySql is being used . The first one has been chosen as standard, this choice has been made also from

other big Institutes, in fact Oracle is used in the National Statistic Institutes in a standard way only.

MySql has a different situation, in fact it 's widely used in Istat while other Institutes use it seldom: this difference is explained from the use of MySql in web applications and it's the consequence of a campaign toward open source technology.

All the programming languages are used in Istat even if Java and PHP has a greater diffusion , also statistical packages are used in some way even if SAS is the standard while the others are seldom used.

“Licensing“ policies and high cost of SAS brought Istat to try other statistic software (such as R and Splus), the scope is to replace SAS at least partially.

5.7 Open Source

The survey contains, between the different software technologies, open source options: this choice has been made to understand the will of the National Statistic Institute to adopt open source politics.

Istat, in the last years, has supported a campaign to promote open source tools, this has been done for different reasons such as: budget, free of research, project diffusion. To sponsor this new vision Istat has created working team that has the purpose to examine and disclose this topic.

Migration from proprietary system as Unix to Linux, the use of My Sql as DBMS, web application development with PHP and JAVA are some results of this politics.

There are many advantages coming from this choice: cost decrease, coming from supplier independence and costs reduction, and high quality of work, due to greater cooperation software reutilization.

This politics have some costs also, related to education and above all changing the standard way of working

This process has been supported by the central structures, particularly by CNIPA (National Center for Information Technology in Pubblc Administration)

6. Appendix

6.1 The questionnaire

6.2 Tables

6.2.1 Frequency Tables

Tab 6.2.1.1: How is the software development organised?

		Frequency	Percent	Cumulative Percent
Valid	Centralized in IT unit	19	67,9	67,9
	Decentralized in statistical unit	1	3,6	71,4
	Both centralized and decentralized	8	28,6	100,0
	Total	28	100	

Tab 6.2.1.2: Which kind of allocation is used for software development resources?

		Frequency	Percent	Cumulative Percent
Valid	Temporary to projects	2	7,1	7,1
	Fixed to a particular survey (or area)	2	7,1	14,3
	Mixed	24	85,7	100,0
	Total	28	100	

*Tab 6.2.1.3: Do you have any IT unit dedicated to some of the survey steps?
(eg. Data Capturing, Check, Dissemination, ...)*

		Frequency	Percent	Cumulative Percent
Valid	Yes	15	53,6	53,6
	No	13	46,4	100,0
	Total	28	100,0	

Tab 6.2.1.4: Server operating system: linux

		Frequency	Percent	Cumulative Percent
Valid	Standard	5	17,9	17,9
	Widely used	4	14,3	32,2
	Seldom used	8	28,6	60,7
	NOT used	11	39,3	100,0
	Total	28	100,0	

Tab 6.2.1.5: Server operating system: windows

		Frequency	Percent	Cumulative Percent
Valid	Standard	19	67,9	67,9
	Widely used	5	17,9	85,7
	Seldom used	4	14,3	100,0
	NOT used	0	0,0	100,0
	Total	28	100	

Tab 6.2.1.6: Server operating system: unix

		Frequency	Percent	Cumulative Percent
Valid	Standard	8	28,6	28,6
	Widely used	2	7,1	35,7
	Seldom used	7	25,0	60,7
	NOT used	11	39,3	100,0
	Total	28	100	

Tab 6.2.1.7: Server operating system: Z/OS, OS/390, other mainframes

		Frequency	Percent	Cumulative Percent
Valid	Standard	5	17,9	17,9
	Widely used	5	17,9	35,7
	Seldom used	3	10,7	46,4
	NOT used	15	53,6	100,0
	Total	28	100,0	

Tab 6.2.1.8: SERVER Data Base Management Systems: Oracle

		Frequency	Percent	Cumulative Percent
Valid	Standard	14	50,0	50,0
	Widely used	1	3,6	53,6
	Seldom used	3	10,7	64,3
	NOT used	10	35,7	100,0
	Total	28	100,0	

Tab 6.2.1.9: SERVER Data Base Management Systems: IBM DB2 Informix

		Frequency	Percent	Cumulative Percent
Valid	Standard	3	10,7	10,7
	Widely used	0	0,0	10,7
	Seldom used	3	10,7	21,4
	NOT used	22	78,6	100,0
	Total	28	100,0	

Tab 6.2.1.10: SERVER Data Base Management Systems: MS SQL Server

		Frequency	Percent	Cumulative Percent
Valid	Standard	12	42,9	42,9
	Widely used	2	7,1	50,0
	Seldom used	9	32,1	82,1
	NOT used	5	17,9	100,0
	Total	28	100,0	

Tab 6.2.1.11: SERVER Data Base Management Systems: Sybase

		Frequency	Percent	Cumulative Percent
Valid	Standard	4	14,3	14,3
	Widely used	0	0,0	14,3
	Seldom used	3	10,7	25,0
	NOT used	21	75,0	100,0
	Total	24	100,0	

Tab 6.2.1.12: SERVER Data Base Management Systems: MYSQL

		Frequency	Percent	Cumulative Percent
Valid	Standard	1	3,6	3,6

Widely used	2	7,1	10,7
Seldom used	8	28,6	39,3
NOT used	17	60,7	100,0
Total	28	100,0	

Tab 6.2.1.13: Programming languages: COBOL, FortTran, Other3GL

		Frequency	Percent	Cumulative Percent
Valid	Standard	2	7,1	7,1
	Widely used	4	14,3	21,4
	Seldom used	11	39,3	60,7
	NOT used	11	39,3	100,0
	Total	28	100	

Tab 6.2.1.14: Programming languages: Visual Basic,.NET

		Frequency	Percent	Cumulative Percent
Valid	Standard	10	35,7	35,7
	Widely used	7	25,0	60,7
	Seldom used	9	32,1	92,9
	NOT used	2	7,1	100,0
	Total	28	100,0	

Tab 6.2.1.15: Programming languages: Java (applet, JSP, EJB, etc.)

		Frequency	Percent	Cumulative Percent
Valid	Standard	10	35,7	35,7
	Widely used	8	28,6	64,3
	Seldom used	4	14,3	78,6
	NOT used	6	21,4	100,0
	Total	28	100,0	

Tab 6.2.1.16: Programming languages: PHP

	Frequency	Percent	Cumulative Percent
--	------------------	----------------	---------------------------

Valid	Standard	2	7,1	7,1
	Widely used	6	21,4	28,6
	Seldom used	9	32,1	60,7
	NOT used	11	39,3	100,0
	Total	28	100	

Tab 6.2.1.17: Programming languages: C, C++

		Frequency	Percent	Cumulative Percent
Valid	Standard	2	7,1	7,1
	Widely used	5	17,9	25,0
	Seldom used	12	42,9	67,8
	NOT used	9	32,1	100,0
	Total	28	100	

Tab 6.2.1.18: Programming languages: PERL, Phytion

		Frequency	Percent	Cumulative Percent
Valid	Standard	1	3,6	3,6
	Widely used	1	3,6	7,1
	Seldom used	11	39,3	46,4
	NOT used	15	53,6	100,0
	Total	28	100	

Tab 6.2.1.19: Statistical Packages: SAS

		Frequency	Percent	Cumulative Percent
Valid	Standard	6	21,4	21,5
	Widely used	8	28,6	50,0
	Seldom used	7	25,0	75,0
	NOT used	7	25,0	100,0
	Total	28	100	

Tab 6.2.1.20: Statistical Packages: SPSS

	Frequency	Percent	Cumulative Percent
--	------------------	----------------	---------------------------

Valid	Standard	6	21,4	21,5
	Widely used	8	28,6	50,0
	Seldom used	7	25,0	75,0
	NOT used	7	25,0	100,0
	Total	28	100	

Tab 6.2.1.21: Statistical Packages: R

		Frequency	Percent	Cumulative Percent
Valid	Standard	0	0,0	0,0
	Widely used	0	0,0	0,0
	Seldom used	5	17,9	17,9
	NOT used	23	82,1	100,0
	Total	28	100	

Tab 6.2.1.22: Statistical Packages: STATA

		Frequency	Percent	Cumulative Percent
Valid	Standard	0	0,0	0,0
	Widely used	0	0,0	0,0
	Seldom used	13	46,4	46,4
	NOT used	15	53,6	100,0
	Total	28	100	

Tab 6.2.1.23: Generalized Software: Survey Design

		Frequency	Percent	Cumulative Percent
Valid	None	12	42,9	42,9
	In house development software	4	14,3	57,1
	Blaise	3	10,7	67,9
	Others	9	32,1	100,0
	Total	28	100	

Tab 6.2.1.24: Generalized Software: Sample Definition

	Frequency	Percent	Cumulative Percent
--	-----------	---------	--------------------

Valid	None	5	17,9	17,9
	In house development software	5	17,9	35,7
	SAS	9	32,1	67,9
	Others	9	32,1	100,0
	Total	28	100	

Tab 6.2.1.25: Generalized Software: Check Correction

		Frequency	Percent	Cumulative Percent
Valid	No Software	3	10,7	10,7
	In house development software	7	25,0	35,7
	SAS	4	14,3	50,0
	Others	14	50,0	100,0
	Total	28	100	

Tab 6.2.1.26: Generalized Software: Statistical Disclosure

		Frequency	Percent	Cumulative Percent
Valid	No Software	6	21,4	21,4
	In house development software	5	17,9	39,3
	Argus	9	32,1	71,4
	Others	8	28,6	100,0
	Total	28	100	

6.2.2 Cross Tables:

6.2.2.1 Server Operating System Number of client

Tab 6.2.2.1.1: SERVER Operating System small Institute

Number of clients		Linux	Windows	Unix	Mainframe	
from 1 to 499	Standard	COUNT	2	5	1	2
		PCT	25	62,5	12,5	25
	widely used	COUNT	0	2	0	1
		PCT	0,00	25	0	12,5
	seldom used	COUNT	2	1	2	0
		PCT	25,00	12,5	25	0
	not used	COUNT	4	0	5	5
		PCT	50,00	0	62,5	62,5
	Total	COUNT	8	8	8	8
		PCT	100	100	100	100

Tab 6.2.2.1.2: SERVER Operating System Medium Institute

Number of clients		Linux	Windows	Unix	Mainframe	
from 500 to 1500	Standard	COUNT	1	8	3	2
		PCT	9,09	72,73	27,27	18,18
	widely used	COUNT	2	2	0	2
		PCT	18,18	18,18	0	18,18
	seldom used	COUNT	3	1	3	0
		PCT	27,27	9,09	27,27	0
	not used	COUNT	5	0	5	7
		PCT	45,45	0	45,45	63,64
Total	COUNT	11	11	11	11	
	PCT	100	100	100	100	

Tab 6.2.2.1.3: SERVER Operating System Big Institute

Number of clients		Linux	Windows	Unix	Mainframe	
more than 1500	Standard	COUNT	2	6	4	1
		PCT	22,22	66,67	44,44	11,11
	widely used	COUNT	2	1	2	2
		PCT	22,22	11,11	22,22	22,22
	seldom used	COUNT	3	2	2	3
		PCT	33,33	22,22	22,22	33,33
	not used	COUNT	2	0	1	3
		PCT	22,22	0	11,11	33,33
	Total	COUNT	9	9	9	9
		PCT	100	100	100	100

6.2.2.2 DBMS Number of client

Tab 6.2.2.2.1: DBMS Small Institute

Number of clients		DB2					
		Oracle	Informix	Ms SQL Server	Sybase	MySql	
from 1 to 499	Standard	COUNT	2	1	5	2	0
		PCT	25	12,5	62,5	25	0
	widely used	COUNT	0	0	1	0	0
		PCT	0,00	0	12,5	0	0
	seldom used	COUNT	1	0	1	0	2
		PCT	12,50	0	12,5	0	25
	not used	COUNT	5	7	1	6	6
		PCT	62,50	87,5	12,5	75	75
	Total	COUNT	8	8	8	8	8
		PCT	100	100	100	100	100

Tab 6.2.2.2.2: DBMS Medium Institute

Number of clients		DB2					
		Oracle	Informix	Ms SQL Server	Sybase	MySql	
from 500 to 1500	Standard	COUNT	5	2	5	1	0
		PCT	45,45	18,18	45,45	9,09	0
	widely used	COUNT	0	0	1	0	0
		PCT	0,00	0	9,09	0	0
	seldom used	COUNT	1	0	2	1	2
		PCT	9,09	0	18,18	9,09	18,18
	not used	COUNT	5	9	3	9	9
		PCT	45,45	81,82	27,27	81,82	81,82
	Total	COUNT	11	11	11	11	11
		PCT	100	100	100	100	100

Tab 6.2.2.2.3: DBMS Big Institute

Number of clients		DB2					
		Oracle	Informix	Ms SQL Server	Sybase	MySql	
from 1500	Standard	COUNT	7	0	2	1	1
		PCT	77,78	0	22,22	11,11	11,11
	widely used	COUNT	1	0	0	0	2
		PCT	11,11	0	0	0	22,22
	seldom used	COUNT	1	3	6	2	4
		PCT	11,11	33,33	66,67	22,22	44,44
	not used	COUNT	0	6	1	6	2
		PCT	0,00	66,67	11,11	66,67	22,22
	Total	COUNT	9	9	9	9	9
		PCT	100	100	100	100	100

6.2.2.3 Programming Languages Number of client

Tab 6.2.2.3.1: Programming Languages Small Institute

Number of clients		Cobol Fortran other 3GL	Visual Basic	Java	PHP	C ,C++	Perl Python
from 1 to 499	COUNT	0	4	2	0	0	0
	Standard PCT	0	50	25	0	0	0
	COUNT	1	1	0	0	0	0
	widely used PCT	12,50	12,5	0	0	0	0
	COUNT	2	2	1	4	2	3
	seldom used PCT	25,00	25	12,5	50	25	37,5
	COUNT	5	1	5	4	6	5
	not used PCT	62,50	12,5	62,5	50	75	62,5
	Total COUNT	8	8	8	8	8	8
	PCT	100	100	100	100	100	100

Tab 6.2.2.3.2: Programming Languages Medium Institute

Number of clients		Cobol Fortran other 3GL	Visual Basic	Java	PHP	C ,C++	Perl Python
from 500 to 1500	COUNT	2	4	3	0	0	0
	Standard PCT	18,18	36,36	27,27	0	0	0
	COUNT	2	3	5	3	2	1
	widely used PCT	18,18	27,27	45,45	27,27	18,18	9,09
	COUNT	3	3	2	3	6	3
	seldom used PCT	27,27	27,27	18,18	27,27	54,55	27,27
	COUNT	4	1	1	5	3	7
	not used PCT	36,36	9,09	9,09	45,45	27,27	63,64
	Total COUNT	11	11	11	11	11	11
	PCT	100	100	100	100	100	100

Tab 6.2.2.3.3: Programming Languages Big Institute

Number of clients		Cobol Fortran other 3GL	Visual Basic	Java	PHP	C ,C++	Perl Python
from 1500	COUNT	0	2	5	2	2	1
	Standard PCT	0	22,22	55,56	22,22	22,22	11,11
	COUNT	1	3	3	3	3	0
	widely used PCT	11,11	33,33	33,33	33,33	33,33	0
	COUNT	6	4	1	2	4	5
	seldom used PCT	66,67	44,44	11,11	22,22	44,44	55,56
	COUNT	2	0	0	2	0	3
	not used PCT	22,22	0	0	22,22	0	33,33
	Total COUNT	9	9	9	9	9	9
	PCT	100	100	100	100	100	100

6.2.2.4 Statistical Packages Number of client

Tab 6.2.2.4.1: Statistical Packages Small Institute

Number of clients		SAS	SPSS	R	STATA	
from1 to 499	Standard	COUNT	3	3	0	0
		PCT	37,5	37,5	0	0
	widely used	COUNT	2	3	0	0
		PCT	25,00	37,5	0	0
	seldom used	COUNT	0	1	0	3
		PCT	0,00	12,5	0	37,5
	not used	COUNT	3	1	8	5
		PCT	37,50	12,5	100	62,5
Total	COUNT	8	8	8	8	
	PCT	100	100	100	100	

Tab 6.2.2.4.2: Statistical Packages Medium Institute

Number of clients		SAS	SPSS	R	STATA	
from 500 to 1500	Standard	COUNT	5	1	0	0
		PCT	45,45	9,09	0	0
	widely used	COUNT	3	3	0	0
		PCT	27,27	27,27	0	0
	seldom used	COUNT	2	3	2	3
		PCT	18,18	27,27	18,18	27,27
	not used	COUNT	1	4	9	8
		PCT	9,09	36,36	81,82	72,73
Total	COUNT	11	11	11	11	
	PCT	100	100	100	100	

Tab 6.2.2.4.3: Statistical Packages Big Institute

Number of clients		SAS	SPSS	R	STATA	
from 1500	Standard	COUNT	8	2	0	0
		PCT	88,89	22,22	0	0
	widely used	COUNT	0	2	0	0
		PCT	0,00	22,22	0	0
	seldom used	COUNT	1	3	3	7
		PCT	11,11	33,33	33,33	77,78
	not used	COUNT	0	2	6	2
		PCT	0,00	22,22	66,67	22,22
	Total	COUNT	9	9	9	9
		PCT	100	100	100	100

6.2.3 Means

6.2.3.1 Statistics

		How many employees work in your NSI? (all locations)	How many software developers?	How many system engineers?	How many clients?	How many servers?
N	Valid	28,00	28,00	28,00	28,00	28,00
	Missing	0,00	0,00	0,00	0,00	0,00
Mean		1.479,14	68,26	31,11	1.568,82	115,29
Std. Deviation		1.539,80	74,18	47,61	1.830,41	160,75
Variance		2.370.977,98	5.502,20	2.266,87	3.350.394,30	25.839,77
Minimum		35,00	2,00	0,00	10,00	2,00
Maximum		6.500,00	250,00	200,00	8.400,00	555,00
Percentiles	25	462,00	9,00	3,00	350,00	20,00
	Median	950,00	40,00	11,00	861,00	60,00
	75	2.425,00	92,00	35,00	2.600,00	130,00

6.2.3.2 Statistics- Number of clients

How many employees work in your NSI? (all locations) How many software developers? How many system engineers? How many clients? How many servers? * Number of clients (recoded)

Number of clients (recoded)		How many employees work in your NSI? (all locations)	How many software developers?	How many system engineers?	How many clients?	How many servers?
from 1 to 499	Mean	295,0000	16,0000	5,0000	185,0000	18,0000
	N	8	8	8	8	8
	Std. Deviation	336,81990	26,78900	7,14000	129,92000	21,91800
from 500 to 1500	Mean	858,0000	45,0000	19,0000	839,0000	60,0000
	N	11	11	11	11	11
	Std. Deviation	301,04700	32,67410	33,90000	253,15200	43,01860
more than 1500	Mean	3.290,0000	152,0000	73,0000	3.690,0000	269,2220
	N	9	9	9	9	9
	Std. Deviation	1.472,00000	78,92390	60,67700	1.871,00000	210,17600
Total	Mean	1.479,14	68,26	31,11	1.568,82	115,29
	N	28	28	28	28	28
	Std. Deviation	1.539,80	74,18	47,61	1.830,41	160,75