# Advances in statistical analysis from the ISBSG benchmarking database

GUFPI-ISMA SBC (Software Benchmarking Committee)
Participating authors: Luca Santillo, Stefania Lombardi, Domenico Natale

## Abstract

*This work presents statistical analyses of adequate sub-samples of development and enhancement software projects extracted from the ISBSG Benchmark 8 (International Software Benchmarking Standards Group, 2003). This research is an incremental process based on the voluntary participation of the members of the Software Benchmarking Committee of the Italian Software Metrics Association (GUFPI-ISMA SBC). The research mainly focuses on the distribution of the project in the ISBSG sample with regard to: functional size method, project size, completion date, development platform, work effort & productivity, primary programming language, and project solar duration. Specific data selection, filtering and transformation criteria are explained and applied. Correlation analysis is proposed – wherever significant – in order to suggest possible utilizations in the field of software estimation. Some suggestions arise from this research in order to achieve an effective data collection by any organization within its own benchmarking database.*

## 1. Introduction

GUFPI-ISMA is a non-profit organization, whose mission is to promote and encourage the use of software measurement methods in Italy [1]. The GUFPI-ISMA Software Benchmarking Committee (SBC), under the guidance of Domenico Natale and Luca Santillo, is aimed to study methods and techniques to analyse and compare software performances, with special attention to software productivity and cost [2]. In the second half of 2003, the GUFPI-ISMA SBC started a series of analysis on the ISBSG Benchmark (Release 8, February 2003) – a database of over 2,000 software development and enhancement projects collected by the ISBSG [5]. Similar analyses have been already performed by ISBSG or others – the SBC's aim is to diffuse, extend, validate and enhance such kind of analysis.

In the current work phase, a subset of main variables was extracted and analysed; the chosen variables are: *measurement method*, *project type*, *project size*, *development platform*, *completion date*, *work effort*, *project delivery rate*, and *programming language*. Further analyses will take into account more variables and eventually more possible correlations and regression will be investigated and reported in future publications by the SBC. Although this is just the first step of the SBC's analysis plan, some suggestions already arise in order to improve and enhance the collection and presentation of the benchmarking data, in order to provide more effective and complete analysis results, in terms of both quality and quantity.

*GUFPI-ISMA SBC analyzes ISBSG and other benchmarking data with the intent of better understanding their meaning, usefulness and consistency. The results of such analysis are not to be considered as a valid reference for any possible official, commercial or legal utilization. Neither GUFPI-ISMA, nor the authors can be hold responsible for errors or damages coming form external utilization of their analysis results.*

## 2. Demographic overview of the ISBSG Benchmark 8

The SBC's analyses are performed on specific data subsets, selected by means of filters to represent significant information. Throughout the paper, graphs and tables are presented – with the number of projects involved – in order to report some key statistics of selected variables. Variables not explicitly involved in the analyses are not described in further details. Table 1 below describes a reduced set of variables from the ISBSG Benchmark 8 database (the "value range" column is the number of different value instances, not the list of such values); a complete table with all the 66 ISBSG variables and a discussion of their interpretation and completeness, along with suggestions for their collection improvement can be found in the work describing the previous, first step of the SBC analysis [3].

The projects origin is not reported by ISBSG, for anonymity reasons. According to ISBSG, major contributors are: Australia, Japan, the United States, the Netherlands, Canada, and United Kingdom; among smaller contributors: India, France, Brazil, and others [5].

***Table 1.*** *Relevant variables in the ISBSG benchmarking sample (extract).*

| Variable | ISBSG Name | N | % | Type | Range | Multiple | Calc |
|---|---|---|---|---|---|---|---|
| DQR | Data Quality Rating | 2,027 | 100.0% | Ord. | 4 | | |
| MM | Count Approach | 2,024 | 99.9% | Text | 14 | | |
| FP_STD_PRIMARY | FP Standard | 1,938 | 95.6% | Text | 25 | | |
| WE_TOT | Summary Work Effort | 2,025 | 99.9% | Num. | - | | |
| RL | Resource Level | 2,027 | 100.0% | Ord. | 4 | | |
| MTS | Max Team Size | 1,015 | 50.1% | Num. | - | | |
| PRJ_TYPE | Development Type | 2,027 | 100.0% | Text | 5 | | |
| PLATFORM | Development Platform | 1,418 | 70.0% | Text | 3 | | |
| LANG | Programming Language | 1,691 | 83.4% | Text | 122 | Yes | |
| DT | Development Techniques | 1,025 | 50.6% | Text | 227 | Yes | |
| PRJ_TIME | Project Elapsed Time | 1,639 | 80.9% | Num. | - | | |
| PRJ_ITIME | Project Inactive Time | 701 | 34.6% | Num. | - | | |
| IMPL_DATE | Implementation Date | 1,802 | 88.9% | Date | - | | |
| PACKAGE | Package Customisation | 1,322 | 65.2% | Y/N | 3 | | |
| PRJ_SCOPE | Project Scope | 1,275 | 62.9% | Text | 26 | Yes | |
| WE1P | WE Plan | 957 | 47.2% | Num. | - | | |
| WE2S | WE Spec | 1,180 | 58.2% | Num. | - | | |
| WE4B | WE Build | 1,291 | 63.7% | Num. | - | | |
| WE5T | WE Test | 1,245 | 61.4% | Num. | - | | |
| WE6I | WE Impl | 846 | 41.7% | Num. | - | | |
| FP_EI/EO/EQ | I/O/Inquiry count (x3) | 2,027 | 100.0% | Num. | - | | |
| FP_ILF/EIF | Int./Ext. Files count (x2) | 2,027 | 100.0% | Num. | - | | |
| FP_ADD/CHG/DEL | Adds/Changes/Del's (x3) | 2,027 | 100.0% | Num. | - | | |
| WE_NORM | Normalised Work Effort | 2,024 | 99.9% | Num. | - | | Yes |
| UFP | Unadjusted Size (FP) | 1,568 | 77.4% | Num. | - | | Part. |
| UFP_RAT | Size Rating | 2,027 | 100.0% | Ord. | 4 | | Yes |
| PDR | Project Delivery Rate | 1,569 | 77.4% | Num. | - | | Yes |
| PDR_NORM | Normalised PDR | 1,569 | 77.4% | Num. | - | | Yes |

## 2.1. Analysis of variables collection

In the previous step of the analysis [3], each of the 66 variables - collected by means of questionnaires by ISBSG – have been analysed. A list of comments on observed data was achieved, in order to improve the collection quality and usefulness. Mostly, comments refer to missing values, ambiguous values and misuse of textual values. Some data manipulations were performed to resolve some of the critical aspects:

- *dichotomization* (from multiple combinations of values to multi-column binarization);
- *nomenclature fixing* (for textual variables).

## 2.2. Overall values distributions for selected variables

The selected variables for the analyses by SBC are: size (UFP), work effort (WE_TOT), project delivery rate (PDR), platform (PLATFORM), primary programming language (LANG), implementation date (IMPL_DATE) and project solar duration (PRJ_TIME).

Distribution analyses are reported for selected variables over the whole ISBSG sample. The following header is common to the following tables:

| N | % | Min | P10 | P25 | Median | P75 | P90 | Max | Mean | Std Dev |
|---|---|---|---|---|---|---|---|---|---|---|

- **N** is the number of cases or data instances in the sample.
- **%** is the percent amount with respect to the sample.
- **Min** and **Max** are, respectively, the minimum and the maximum values in the sample.
- **Pxx** is the $xx^{th}$ percentile (it is that value which is greater than the values of *xx* percent of the members of the sample); P25 is also known as the first quartile, P75 as the third quartile; the $50^{th}$ percentile, or **Median**, divide the sample in two equal parts..
- **Mean** and **Std Dev** are, respectively, the arithmetic mean and the standard deviation.

Since many distributions are skewed towards low values – and the data contains outliers – the median is a more useful measure, with respect to the mean. The maximum value of N is 2,027, since this is the total number of project instances in the Benchmark 8 – but this amount is rarely reached, due to void, unknown, or unclear values. Next, basic distributions are reported, but no diagram is plotted, since no differentiation is made among distinct measurement methods and project types.

***UFP (unadjusted function point size – measurement unit: UFP)***

| N | % | Min | P10 | P25 | Median | P75 | P90 | Max | Mean | Std Dev |
|---|---|---|---|---|---|---|---|---|---|---|
| 1,568 | 77.4 | 6.0 | 63.0 | 109.8 | 224.0 | 476.0 | 1,182.2 | 19,050.0 | 514.2 | 1,087.3 |

***WE_TOT (summary work effort – measurement unit: ph)***

| N | % | Min | P10 | P25 | Median | P75 | P90 | Max | Mean | Std Dev |
|---|---|---|---|---|---|---|---|---|---|---|
| 2,025 | 99.9 | 5.0 | 419.4 | 888.0 | 2,200.0 | 5,307.0 | 13,737.6 | 645,694.0 | 6,883.4 | 26,160.8 |

***PDR (project delivery rate – measurement unit: ph/UFP)***

| N | % | Min | P10 | P25 | Median | P75 | P90 | Max | Mean | Std Dev |
|---|---|---|---|---|---|---|---|---|---|---|
| 1,569 | 77.4 | 0.02 | 2.0 | 4.2 | 9.0 | 18.0 | 33.7 | 640.0 | 15.6 | 26.7 |

***PLATFORM (development platform – textual variable)***

| N | % | MF | %_N | MR | %_N | PC | %_N |
|---|---|---|---|---|---|---|---|
| 1,418 | 70.0 | 844 | 59.5 | 252 | 17.8 | 322 | 22.7 |

*Captions: MF = Mainframe, MR = Midrange, PC = Personal Computer*

*LANG (primary programming language – textual variable)*

| N | % | Cobol* | %_N | C | %_N | VB | %_N | C++ | %_N | Oracle | %_N | Rest | %_N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1,691 | 83.4 | 451 | 26.7% | 153 | 9.0% | 115 | 6.8% | 114 | 6.7% | 108 | 6.4% | 751 | 44.4% |

*IMPL_DATE (implementation date – date field)*

| N | % | 2002 | %_N | 2001 | %_N | 2000 | %_N | 1999 | %_N | 1998 | %_N | Rest | %_N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1,802 | 88.9 | 147 | 8.2% | 75 | 4.2% | 387 | 21.5% | 282 | 15.6% | 236 | 13.1% | 675 | 37.5% |

*PRJ_TIME (project solar duration – measurement unit: month)*

| N | % | Min | P10 | P25 | Median | P75 | P90 | Max | Mean | Std Dev |
|---|---|---|---|---|---|---|---|---|---|---|
| 1,639 | 80,9 | 0,5 | 2,0 | 4,0 | 7,0 | 11,0 | 17,0 | 84 | 8,6 | 7,4 |

## 3. Subsets selection for analyses

In order to analyse the selected variables, some filtering and transformation actions had to be performed on the original ISBSG data sample, leading to two sub-samples, denoted as sample A (or "soft filter" sub-sample) and sample B (or "severe filter" sub-sample) [REF]; the criteria are briefly reported in Table 2. The current work reports the analysis results from the "severe filter" sample, only. The "project type" attribute has been kept in the sub-samples to differentiate the analysis results by project type: "Development" or "Enhancement".

Note that filtering out records with PACKAGE = "Y" left records with both explicit "N" values and void values, so that a remaining impact of undocumented package customisation for some projects could still affect the analysis results.

***Table 2.*** *Applied filters; starting N is 2,027.*

| Step | Filtering Variable | Filtering Criteria | Excluded Records | Residual | Res. % |
|---|---|---|---|---|---|
| 1 | PRJ_TYPE | = "New Dev." Or "Enh." | 57 (various) | 1,970 | 97.2% |
| 2 | MM | = "IFPUG" | 195 not "IFPUG" | 1,775 | 87.6% |
| 3 | DQR | = "A" Or "B" | 113 "C" or "D" | 1,662 | 82.0% |
| 4 | FP_STD_PRIMARY | = "IFPUG *" | 336 not "IFPUG *" | 1,326 | 65.4% |
| **Ä** | **Sample A ("soft filter" sub-sample; 1,326 records)** | | | | |
| 5 | PACKAGE | ≠ "Y" | 68 "Y" | 1,258 | 62.1% |
| 6 | UFP_RAT | = "A" Or "B" | 185 "C" or "D" | 1073 | 52.9% |
| 7 | FP_STD_PRIMARY | = "IFPUG 4.*" | 159 unspecified or "< 4" | 914 | 45.1% |
| 8 | RL | = "1" Or "2" | 140 "3" or "4" | 774 | 38.2% |
| **Ä** | **Sample B ("severe filter" sub-sample; 774 records)** | | | | |

### 3.1. Data categorization

Due to the wide variety of value instances, two variables have been "translated" into new category variables (or "classes"):

- IMPL_DATE (Implementation Date), re-aggregated into IMPL_PERIOD (Implementation Period); selected periods are non-overlapping: 1989-1990, 1991-1992, 1993-1994, …, 1999-2000, 2001-2002. IMPL_PERIOD is an indicator of "when the project was completed", not of "when the project was executed".
- PRIM_PROG_LANG (Primary Programming Language), re-aggregated into LANG_LEV (Language Level), and further re-aggregated into LL_CAT (Language Level Category);

Capers Jones' well-known programming languages table was taken as a reference for such classification (Table 3).

***Table 3.** Programming languages (examples), level ranges and categories.*

| LL_CAT | Range | Examples |
|---|---|---|
| LL_CAT 1 | 1-3 | ASSEMBLER, C, COBOL, COBOL 2, MVS COBOL, FORTRAN, PASCAL. |
| LL_CAT 2 | 4-8 | 3rd Gen. Lang, PL/I, LISP, C++, JAVA, ADA, CICS, ORACLE, MS ACCESS. |
| LL_CAT 3 | 9-15 | VISUAL BASIC, DELPHI, LOTUS NOTES, UNIX SHELL SCRIPT […]. |
| LL_CAT 4 | 16-23 | 4th Gen. Lang., CLIPPER, POWERBUILDER, TELON, SAP ABAP, HTML, ASP. |
| LL_CAT 5 | 24-55 | SQL, EASYTRIEVE, PL/SQL, SQL WINDOWS, Spreadsheets. |
| LL_CAT 6 | >55 | 5th Generation Languages. |

## 3.2. Data transformation

The only relevant data transformation taken by the SBC was:

- UFP size attribute equalized to the (adjusted) FP values from the ISBSG database for those projects, where only "FP", with no "VAF" and no "function breakdown detail", was provided; analysing further data only by means of size ranges – see next section 4 – can smooth the risk carried by this hypothesis;
- PDR re-calculated by SBC, including previously void values were UFP is achieved (previous item).

## 3.3. Values distributions in final sub-samples

Sample A ("soft filter" sub-sample) and sample B ("severe filter" sub-sample) were obtained through the previously depicted filtering, categorization and transformation actions; they contain, respectively, 65.4% and 38.2% of the original ISBSG database recordset. The sub-sample B characteristics are reported in the following Table 4, where:

- "DEV" and "ENH" stand, respectively, for "(new) development" and "enhancement" (project type) - "Aggr." stands for "Aggregated";
- "PRJ_ID" is omitted, but still present for sake of traceability of records;
- percentages are referred to the sub-sample, by column, – not to the overall database;
- aspects that remained critical are highlighted in grey;
- aspects that were improved are highlighted with bold, italic text style.

Note that no variable in this sub-sample has multiple values.

***Table 4.** Characteristics of selected variables in sample B ("severe filter").*

| Variable | N | % | $N_{DEV}$ | $\%_{DEV}$ | $N_{ENH}$ | $\%_{ENH}$ | Type | Range | Calc |
|---|---|---|---|---|---|---|---|---|---|
| UFP | 774 | [100%] | 299 | [100%] | 475 | [100%] | Num. | - | Part. |
| PLATFORM | 386 | 49.9 | 168 | 56.2% | 217 | 45.7% | Text | 3 | |
| LANG_LEV | 587 | 75.8 | 214 | 71.6% | 373 | 78.5% | ***Ord.*** | ***29*** | Aggr. |
| LANG_CAT | 587 | 75.8 | 214 | 71.6% | 373 | 78.5% | ***Ord.*** | ***6*** | Aggr. |
| IMPL_PERIOD | 719 | ***92.9*** | 270 | 90.3% | 449 | 94.5% | ***Ord.*** | ***6*** | Aggr. |
| WE_TOT | 774 | ***100.0%*** | 299 | 100.0% | 475 | 100.0% | Num. | - | |
| PDR | 774 | ***100.0%*** | 299 | 100.0% | 475 | 100.0% | Num. | - | Yes |

## 4. Distribution analyses

***UFP (unadjusted function point size – measurement unit: UFP) – Sample B.***

| | N | % | Min | P10 | P25 | Median | P75 | P90 | Max | Mean | Std Dev |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DEV | 299 | 38.6% | 25.0 | 112.0 | 174.5 | ***334.0*** | 676.0 | 1,438.4 | 16,148.0 | 666.5 | ***1,227.8*** |
| ENH | 475 | 61.4% | 6.0 | 56.0 | 88.0 | ***153.0*** | 281.0 | 604.4 | 7,134.0 | 282.2 | ***470.4*** |
| **TOT** | **774** | **100%** | **6.0** | **63.0** | **108.3** | ***201.5*** | **429.8** | **928.3** | **16,148.0** | **430.6** | ***867.1*** |

From the analysis of percentiles on the logarithmic distribution for development and enhancement projects (separately), a limited set of project size classes is obtained (Table 5). Such classes are used for categorization of subsequent two-variable analyses; they are proposed for standardized use in agreement definitions and software estimation approaches.

***Table 5.*** *Development and enhancement projects size classes.*

| SIZE_CLASS | Dev Code | DEV UFP Range | ENH Code | ENH UFP Range |
|---|---|---|---|---|
| Very Small | $DEV_{XS}$ | 0-150 | $ENH_{XS}$ | 0-60 |
| Small | $DEV_{S}$ | 150-300 | $ENH_{S}$ | 60-120 |
| Medium | $DEV_{M}$ | 300-600 | $ENH_{M}$ | 120-240 |
| Large | $DEV_{L}$ | 600-*1,200* | $ENH_{L}$ | 240-*480* |
| Very Large | $DEV_{XL}$ | *1,200*-5,000 | $ENH_{XL}$ | *480*-2,000 |
| Extremely Large | $DEV_{XXL}$ | > 5,000 | $ENH_{XXL}$ | > 2,000 |

***WE_TOT (summary work effort – measurement unit: ph) – Sample B.***

| | N | % | Min | P10 | P25 | Median | P75 | P90 | Max | Mean | Std Dev |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DEV | 299 | 38.6% | 50.0 | 597.6 | 1,057.5 | 2,540.0 | 5,924.0 | 14,911.8 | 73,920.0 | 6,232.1 | ***10,593.9*** |
| ENH | 475 | 61.4% | 90.0 | 426.0 | 762.5 | 1,642.0 | 3,911.0 | 8,245.4 | 53,830.0 | 3,560.2 | ***5,607.1*** |
| **TOT** | **774** | **100%** | **50.0** | **455.5** | **867.5** | **1,913.5** | **4,713.5** | **10,023.3** | **73,920.0** | **4,592.3** | ***8,014.9*** |

No specific consideration are reported for effort distributions. More significant information is expected from the project delivery rate (i.e. summary work effort by size – next).

***PDR (project delivery rate – measurement unit: ph/UFP) – Sample B.***

| | N | % | Min | P10 | P25 | Median | P75 | P90 | Max | Mean | Std Dev |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DEV | 299 | 38.6% | 0.1 | 2.1 | 3.9 | 8.0 | 16.2 | 23.1 | 300.3 | 12.8 | ***21.5*** |
| ENH | 475 | 61.4% | 0.3 | 2.5 | 4.8 | 11.0 | 23.0 | 44.6 | 327.4 | 19.8 | 29.1 |
| **TOT** | **774** | **100%** | **0.1** | **2.3** | **4.3** | **9.5** | **19.3** | **36.0** | **327.4** | **17.1** | ***26.6*** |

As for the size and effort distribution, a skewed distribution is observed for project delivery rate. A log-normal distribution should be considered.

***PLATFORM (development platform – textual variable) – Sample B.***

| | N | % | MF | %_N | MR | %_N | PC | %_N | Check |
|---|---|---|---|---|---|---|---|---|---|
| DEV | 168 | 43.6% | 98 | 58.3% | 34 | 20.3% | 36 | 21.4% | 100% |
| ENH | 217 | 56.4% | 196 | 90.3% | 9 | 4.2% | 12 | 5.5% | 100% |
| **TOT** | **385** | **100%** | **294** | **76.3%** | **43** | **11.2%** | **48** | **12.5%** | **100%** |

Although the platform distributions are clear enough, some difficulty is found in interpreting the assignment of such values (e.g. “midrange” versus “personal computer” for some application types).

***LL_CAT (language level category – ordinal variable) – Sample B.***

| | N | % | LLC1 | %_N | LLC2 | %_N | LLC3 | %_N | LLC4 | %_N | LLC5 | %_N | LLC6 | %_N | Check |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DEV | 214 | 36.5% | 98 | 45.8% | 38 | 17.8% | 28 | 13.1% | 28 | 13.1% | 22 | 10.3 | 0 | 0% | 100% |
| ENH | 373 | 63.5% | 195 | 52.3% | 102 | 27.3% | 34 | 9.1% | 27 | 7.2% | 14 | 3.8% | 1 | 0.3% | 100% |
| **TOT** | **587** | **100%** | **293** | **49.9%** | **140** | **23.9%** | **62** | **10.6%** | **55** | **9.4%** | **36** | **6.1%** | **1** | **0.2%** | **100%** |

The Language Level Category distribution has its maximum at LLC1 (Language Level 1-3), followed by LLC 2 (Language Level 4-8). A more specific analysis (not reported here) shows a peak for Level 3 languages (mostly COBOL). Many projects are present, where the Primary Programming Language is not provided.

***IMPL_PERIOD (implementation period – ordinal variable) – Sample B.***

| | N | % | 1989-90 | 1991-92 | 1993-94 | 1995-96 | 1997-98 | 1999-2000 | 2001-02 |
|---|---|---|---|---|---|---|---|---|---|
| DEV | 270 | 37.6% | 0 | 2 | 42 | 41 | 106 | 73 | 6 |
| ENH | 449 | 62.4% | 0 | 0 | 10 | 24 | 163 | 95 | 157 |
| **TOT** | **719** | **100%** | **0** | **2** | **52** | **65** | **269** | **168** | **163** |

No specific distribution is expected for implementation period..

***PRJ_TIME (project solar duration – measurement unit: month) – Sample B.***

| | N | % | Min | P10 | P25 | Median | P75 | P90 | Max | Mean | Std Dev |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DEV | 264 | | 0,5 | 3,0 | 5,0 | 7,0 | 13,0 | 20,0 | 44,0 | 9,7 | 6,8 |
| ENH | 309 | | 1,0 | 2,0 | 4,0 | 6,0 | 9,0 | 13,0 | 27,0 | 6,9 | 4,4 |
| **TOT** | **573** | | **0,5** | **3,0** | **4,0** | **7,0** | **10,0** | **16,0** | **44,0** | **8,2** | **5,8** |

As for the size, effort, and productivity distribution, a skewed distribution is observed for project solar duration. A log-normal distribution should be considered. Further analysis should consider the project *inactive* time.

## 5. Correlation analysis

A Pearson correlation table is reported for selected numerical variables [8].

| | *DEV (N = 264)* | | | | *ENH (N = 309)* | | | |
|---|---|---|---|---|---|---|---|---|
| | *UFP* | *WE_TOT* | *PDR* | *PRJ_TIME* | *UFP* | *WE_TOT* | *PDR* | *PRJ_TIME* |
| *UFP* | 1,00 | 0,71 | -0,07 | 0,36 | 1,00 | 0,40 | -0,14 | 0,21 |
| *WE_TOT* | 0,71 | 1,00 | 0,32 | 0,65 | 0,40 | 1,00 | 0,30 | 0,21 |
| *PDR* | -0,07 | 0,32 | 1,00 | 0,28 | -0,14 | 0,30 | 1,00 | 0,10 |
| *PRJ_TIME* | 0,36 | 0,65 | 0,28 | 1,00 | 0,21 | 0,21 | 0,10 | 1,00 |

We can argue a moderate correlation between size and effort for new developments, while the same does not hold true for enhancements. One possible explanation is that the IFPUG size for enhancements includes the entire functions impacted, not only their modified portions. Also, for developments, a linear regression cannot be considered ($r^2 = 0,5$), but a significance test yields positive for the correlation.

This initial correlation analysis confirms that the (functional) size is a primary cost driver for the project, but not the only one. Since the best fitting relationships among such variables have already been proved to be non-linear, further analysis is required, replacing linear correlation with other indicators and/or looking for multivariate relations.

## 6. Two-variables analyses

Two-variables analysis is introduced to investigate the dependency of PDR against other variables in a visual approach. Numerical regression analysis is omitted, by now, to avoid a potential misuse of the analysis results. Only enhancement cases are reported.

### 6.1. PDR vs. SIZE_CLASS

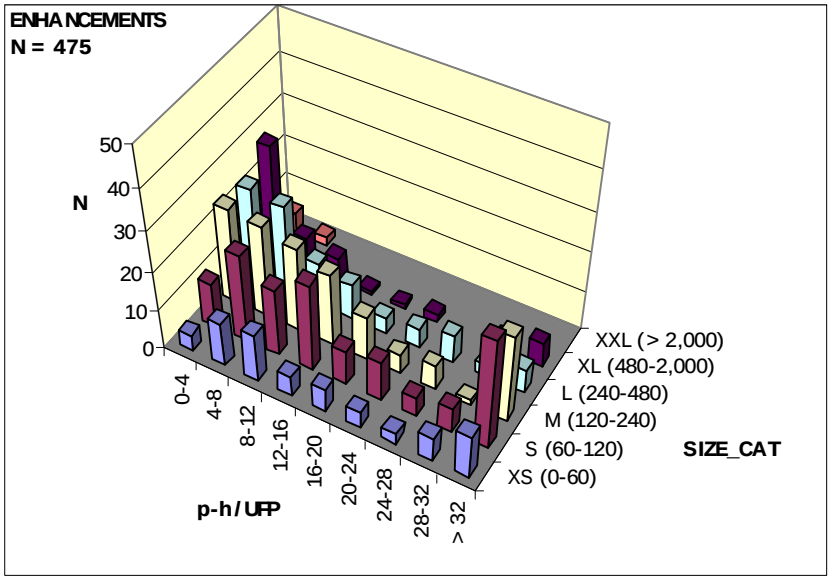


***Figure 1.*** *PDR distributions against Size Class (Sample B). Although some log-normal trends could be perceived, more data could provide more regular distributions. Peak frequency values are between 4 and 12 person-hours per UFP (8 p-h » 1 person-day).*

### 6.2. PDR vs. LL_CAT

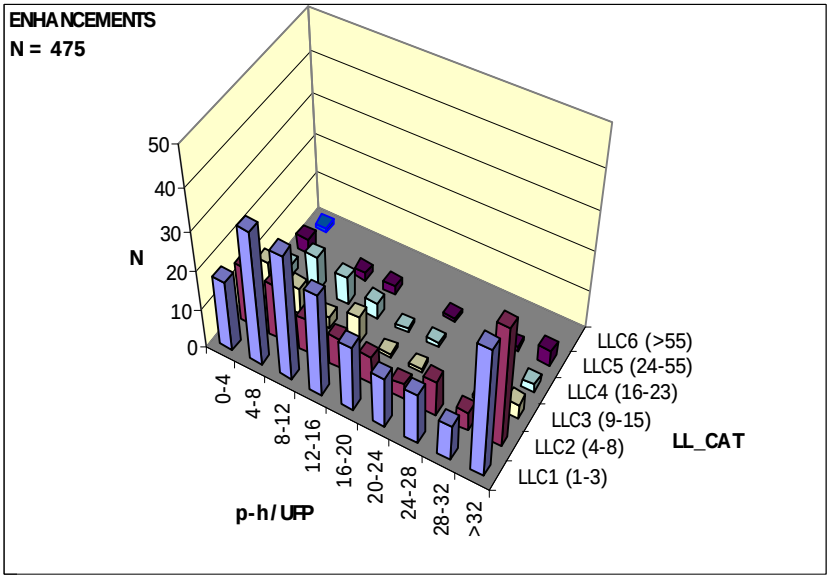


***Figure 2.*** *PDR distributions against Language Level Category (Sample B). As easily argued, most enhancement projects are implemented with low level programming languages. The peak for PDRs' > 32 p-h/UFP is due to a tail containing high values.*

## 6.3. PDR vs. IMPL_PERIOD

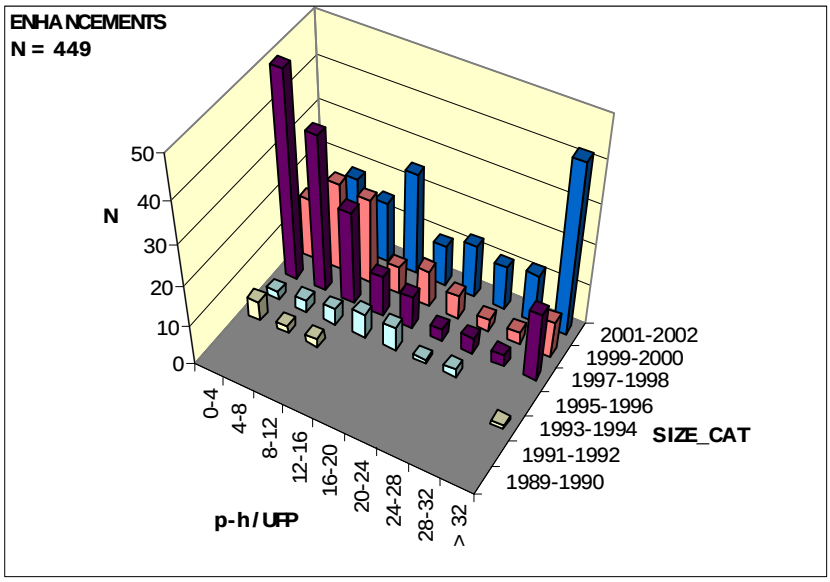


***Figure 3.*** *PDR distributions against Implementation Period (Sample B). The PDR distribution in recent years (2001-2002) has a peak for PDRs' > 32 p-h/UFP, due to a tail containing several extremely high values.*

## 6.4. PDR vs. PRJ_TIME

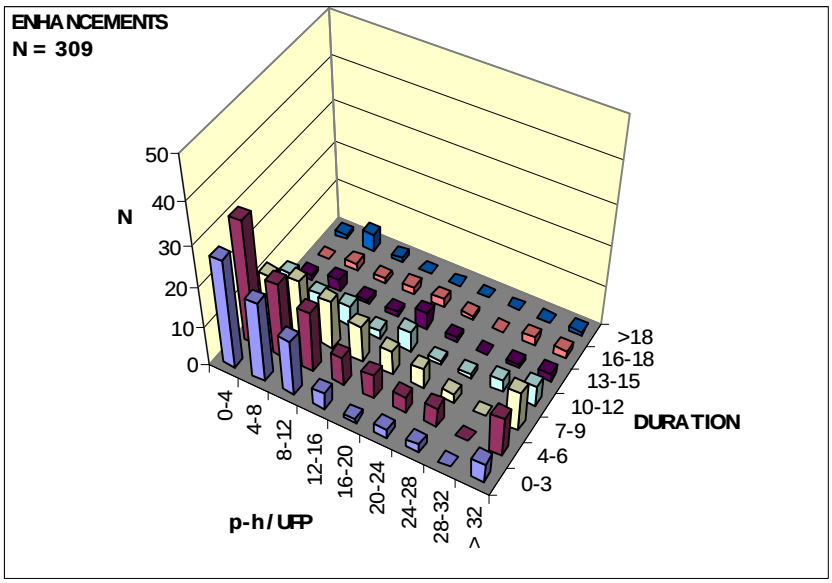


***Figure 3.*** *PDR distributions against Project Solar Duration (Sample B). As expected, most of the enhancement projects has a limited duration (time-to-market reasons). A log-normal distribution could be argued; more data, or different categorization, is required.*

## 7. Conclusions

Several suggestions had been highlighted throughout the previous step of benchmarking analysis by the GUFPI-ISMA SBC:

- dichotomization (to avoid multiple values per record);
- nomenclature (to avoid distinct values for identical instances);
- taxonomy (to avoid open ranges – including a single "other" exception);
- completeness (to avoid excessive sample filtering, or interpretation of void values);
- variable categorization (to avoid excessive variety of instances).

While these suggestions may be considered for future improvements of the ISBSG collection process, as well as for the implementation of any local benchmarking database within an organization, some useful hints can be obtained from the analysis of the current data. The ongoing analysis should provide a double-check, by means of distinct methods, of statistical relationships that can be found in literature or from the ISBSG publications.

Although some projects are present in the overall ISBSG sample with measurement method alternative to IFPUG, as COSMIC and MkII approaches, such sample were found too sparse to permit significant analyses. The next benchmarking database release by the ISBSG, the Benchmark 9 issued at the end of 2004 with over 3'000 projects, will open new views on different measurement methods and their application in the statistical analysis. Therefore, further research developments are:

- extension over larger samples, including new sizing methods;
- outliers analysis and possible deletion;
- extension over more variables (e.g. functional breakdown by function type and by enhancement operation type, work effort phase breakdown, differentiation by methodology, by software domain, etc.);
- two-variables and N-variables regression analysis;
- factor and principal component analysis.

The GUFPI-ISMA Software Benchmarking Committee wishes to thank all its members for providing useful hints and collaboration on the benchmarking analysis (previous step) and to encourage further research on this subject.

## References

[1] GUFPI-ISMA website, Gruppo Utenti Function Point Italia – Italian Software Metrics Association, http://www.gufpi.org.

[2] GUFPI-ISMA SBC webpage, Software Benchmarking Committee, http://www.gufpi.org/sbc.

[3] GUFPI-ISMA SBC, "Proposals for project collection and classification from the analysis of the ISBSG Benchmark 8", in IWSM 2004 – International Workshop on Software Measurement proceedings, Berlin, 3-5 November 2004.

[4] ISBSG, "ISBSG Shared Benchmarking Repository Report, Release 5", Australia, March 1998.

[5] ISBSG, "Estimating, Benchmarking & Research Suite (incorporating the data disk), CD, Release 8", Australia, February 2003.

[6] ISBSG website, International Software Benchmarking Standards Group, http://www.isbsg.org.

[7] Jones, C., "Programming Languages Table, Release 8.2", SPR (USA), March 1996.

[8] Levine, D.M., Krehbiel, T.C., Berenson, M.L., "Business Statistics: A First Course, 2nd ed.", Prentice-Hall, 2000.