Investigation of the Effort Data Consistency in the ISBSG Repository

David Déry
École de Technologie Supérieure
david.dery.1@ens.etsmtl.ca

Alain Abran
École de Technologie Supérieure
aabran@ele.etsmtl.ca

Abstract

To develop adequate software project estimation models using statistical techniques, the consistency of historical data is important. This paper investigates this issue by looking into the consistency of the information contained in one of the most important fields in the International Software Benchmarking Standards Group (ISBSG) repository, that is, the project effort data field. This paper also presents an example of how effort data from projects that include a large number of project phases can be used for extrapolation, through a normalization process, to projects with fewer phases. The ISBSG organization has attempted to tackle this issue on the variability of phases included in the project effort field by deriving a normalized work effort field. This paper investigates this problem and reports on a number of related issues.

1. Introduction

In project management, the importance of planning, estimation and control is well known, and the problem of producing adequate estimation for software development projects has often been discussed in the literature [1-11]. This problem is not due to a lack of estimation process alternatives: expert-based, analogy-based, price to win, available resources-based and parametric models-based. Most estimation processes have defined steps such as the estimation of the size of the software (in lines of code, function points, etc.), which is then taken as input to estimation models based on the analysis of effort of past, completed projects.

For benchmarking, as well as for estimation models and tools that require historical data, at least two prerequisites must be met: there must be enough historical data to ensure statistical validity, and the data must be homogeneous enough to provide meaningful interpretations.

The International Software Benchmarking Standards Group (ISBSG), a not-for-profit organization, was established in the late 1990s [12] to improve software benchmarking and estimation by setting up a publicly available data repository. The 2005 ISBSG repository, release 9 (R9), contains information on 100 data fields for 3024 projects; however, these projects might not contain information about all 100 fields that can be collected in the ISBSG repository, and only a subset of these are mandatory fields. This ISBSG repository contains data originating from organizations across the world with projects from different industries which have used different methodologies, phases and techniques. Of course, the ISBSG captures information about the project process, the technology, the people, the effort and the product of the project, and has defined each of the fields to be collected in its repository with great care.

Two of the key fields in such repositories are, of course, the size and effort fields. The effort data originate from organizations' time reporting systems, which, in practice, vary considerably across organizations: one system may include effort from initial planning to full deployment, while another will only report effort for the programming and testing phases. To adequately record the required background information on the total project effort reported in its repository, the ISBSG asks data collectors to map their own life cycle to a standardized ISBSG life cycle with six phases: Planning, Specification, Design, Build, Test and Implement.

In the ISBSG repository, total project effort is a mandatory field, while the effort per project phase is an optional one. This heterogeneity in project phases included in the effort data collected means, then, that the total project effort [1] recorded in the ISBSG repository has to be handled with care (e.g. the life cycle coverage is not identical across projects). A key challenge in data analysis using the ISBSG repository, as with many multi-organizational repositories where there exists the possibility of variability in life cycle coverage for the effort data is therefore to first assess the consistency of the data collected.

The ISBSG has recognized this problem, and, in late 2001, developed a procedure to partially tackle it by devising a procedure to derive a new field, referred to as 'normalized work effort' for the purpose of partially

[1] The ISBSG refers to this total project work effort recorded as the 'summary work effort'

tackling the disparity of project phase profiles [14]. The purpose of this added field is to provide an indication of what the total effort of the project could be if the effort from the missing phase or phases had actually been reported and added to the total effort recorded in the ISBSG repository. The ISBSG normalization procedure has been described in internal correspondence [14] and has since been carried out manually by the ISBSG data administrator. It requires a number of steps and the use of some data fields that are not included in the releases made available to purchasers, in the industry, of a license to the ISBSG dataset.

This paper presents an independent investigation of the consistency of the total project effort data originating from projects having different sets of phases, thereby revealing variation in the meaning and coverage of the effort field. This paper also compares some results from the analysis of the 2005 ISBSG repository with data reported by the ISBSG on the initial 2001 normalization procedure.

This paper is organized as follows. Section 2 presents the data preparation procedures and an overview of the variability in the effort data field in the 2005 ISBSG repository, and section 3 presents an analysis of the ISBSG-derived normalized project effort field. Some summary remarks are presented in section 4.

2. ISBSG variability in the effort data

2.1. Initial data preparation

Before analyzing the ISBSG data, it is important to understand how fields are defined, used and recorded, as recommended in [13]. In data set preparation, two verification steps must be carried out: data quality verification and data completeness verification. The verification of effort data quality analysis is easy: the ISBSG data repository manager himself carries out such a data quality analysis right at collection time, and records his judgment in a project field, and his rating will vary from very good (A) to unreliable (D). To reduce the risk of poor quality data and to improve the validity of the analysis reported here, projects with a data quality rating of C and D were removed prior to the analysis.

In the ISBSG repository, not all projects have been sized according to the same functional sizing method. The IFPUG, COSMIC-FFP and Mark II methods are used, as well as a few others. In the ISBSG 2005 release, the sizing method applied to the greatest number of projects is the IFPUG method, with 2,562 projects. For the analysis reported here, only projects sized with this method have been retained.

To identify the presence of variability in the total project effort field, it is important to understand that:
A- not all projects provide information on which phases are included in the total project effort reported;
B- when project phases are reported:

- total project effort reported will cover a variety of combinations of phases;
- not all phases have effort assigned to them;
- for some projects, the sum of each individual effort phase might not be equal to the total project work effort, or, as the ISBSG labels it, summary work effort .

In the repository, there may be any combination of the 6 phases defined by the ISBSG. The variability in the effort field is illustrated in Table 1 for the 2,562 projects measured with the IFPUG method and for which data quality is considered to be high. In this table, the columns indicate the project phases, while the rows indicate the combination of phases identified in the repository, i.e. there are 31 different combinations of project phases present in the repository. For example:
- rows 1, 9 10 and 27 indicate projects with a single-phase effort;
- rows 2, 3, 11, 15, 16, 25 and 28 indicate projects with a two-phase effort; and
- row 7[2] and 22 indicate projects with the full, six-phase effort coverage.

These different combinations of project phases will be referred to as project phase profiles.

Row 32 indicates that 1,006 projects have no any indication of projects phases. Furthermore, since the ISBSG allows, on practical grounds, for a "partial life cycle" and for the "I don't know" option, there are another 29 projects in row 6 without adequate information on the phase coverage for the effort data. It can be observed that there are only three phase profiles with over 100 projects in each, as follows:
- profile 19 with 405 projects includes 4 phases (Plan, Specification, Build and Test);
- profile 20 with 350 projects includes 5 phases (Plan, Specification, Build, Test and Implement);

[2] Profile 7 is labeled as a "full life cycle" and profile 22 contains all 6 phases: there is no explanation as to why these two profiles are different.

- profile 30 with 349 projects includes another combination of 4 phases (Specification, Build, Test and Implement)

From Table 1, it can be seen that there are a number of projects where industry has indicated that the design phase in their development effort is too small for any statistical analysis (that is, profiles 5, 7, 21 and 22, for a total of 11 projects). By contrast, for a large number of projects, there is an indication that the effort includes both the previous and later phases (that is, Specifications and Build). Since it would be surprising if recorded effort for such projects excluded the intermediate design phase, it can be postulated that, in industry, the time recording system does not segregate the design phase for practical reasons; that is, design-related effort could be included either in the specification phase or in the build phase. However, there is not enough documentation to reasonably postulate in which of these two later phases industry records design effort.

Table 1: Projects and their phases in the ISBSG repository

Profile #	# projects	Planning	Specification	Design	Build	Test	Implement	I don't know	Full Life Cycle	Blank	Total
1	11				v						
2	2				v		v				
3	19				v	v					
4	13				v	v	v				
5	1			v	v	v					
6	29							v			
7	5								v		
8	1			H/L	v	v	v				
9	3						v				
10	5	v									
11	20	v			v						
12	7	v			v		v				
13	64	v			v	v					
14	37	v			v	v	v				
15	3	v					v				
16	9	v	v								
17	68	v	v		v						
18	39	v	v		v		v				
19	**405**	v	v		v	v					
20	**350**	v	v		v	v	v				
21	1	v	v	v	v	v					
22	4	v	v	v	v	v	v				
23	5	v	v			v					
24	1	v	v			v	v				
25	1	v				v					
26	1	v				v	v				
27	0		v								
28	8		v		v						
29	3		v		v		v				
30	349		**v**		**v**	**v**					
31	**92**		v		v	v	v				
32	1006									v	
Total	2562										v

With so much diversity in the actual content (and scope) of the collected effort field, and correspondingly so many possible combinations of phases recorded, how can meaningful benchmarking comparisons be carried out and adequate estimation models be built? Of course, performing productivity benchmarking and building estimation models based on projects where effort data do not share the same definition in terms of phases included is challenging. However, not taking such disparities into account would be hazardous: indeed, calculating average

effort, for instance, can be done meaningfully only within a specific profile, and each profile should be dealt with distinctly; unless, of course, some of them could be normalized.

2.2. Second level of data preparation

Can some profiles be merged under reasonable conditions? One strategy could be to identify the various profiles and determine whether or not the information found in the profile with the greater number of phases could be used to extrapolate information to the adjacent profiles (that is, where only a single adjacent phase is missing). If so, such adjacent profiles could be merged under some reasonable assumptions to provide a new sample with a greater number of projects. In the initial phase of this research, the profiles with only one missing phase from a reference profile are investigated.

For ease of reading, each phase effort profile is labeled here with an acronym representing the phases included (1st letters). For instance, profile 3 (Build-Test) is labeled BT; profile 4 (Build-Test-Implement) is labeled BTI, etc.

In Table 1, the profile with the most inclusive number of phases is PSBTI (profile 20). The other profiles, which are adjacent, are: PSBT (profile 19) and SBTI (profile 31). These three profiles are listed next in Table 2; they correspond to 3 of the 4 profiles with the largest number of projects in their profiles (ranging from 92 to 405 projects per profile).

However, not all projects with phase tags also have concurrently detailed effort by project phase. Since only projects with effort data recorded by project phase can be analyzed, the sample size is reduced significantly: for instance, for the PSBTI profile, out of the 350 projects in this profile, only 113 have detailed effort data by phase (Table 2, column 3).

In addition, after verifying for the consistency of the detailed effort by phase with the total project effort recorded, 37 projects had to be dropped from further analysis, thereby leading to a sample of only 76 projects which meet this consistency criterion for our analytical purposes (Table 2, column 4). Profile 19 ends up with 100 projects and profile 31 with only 3 projects.

Table 2: PSBT-related profiles: with phase tags and effort data by phase

	Number of projects		
Project Profile	With phase tags	With detailed effort by phase	Phase effort consistent with Summary Effort
20: PSBTI	**350**	**113**	**76**
19: PSBT	**405**	**200**	**100**
31: SBTI	**92**	**12**	**3**

2.3. Third level of data preparation

Before calculating the average effort of each phase of the 76 projects in the PSBTI profile, it is important to verify whether or not there are obvious outliers or unusual patterns in terms of effort recorded in each project phase. An obvious outlier in the effort and planning fields was identified and removed[3]. The size of the sample then dropped to 75 projects. Furthermore, another unusual effort pattern was identified: 34 projects had, on average, 98% of the effort recorded in the specification phase, and less than 1% in each of the other 4 phases (Table 3). Of course, for the purposes of our analysis, these projects must also be discarded. In summary, out of the 350 projects with a PSBTI profile, only 41 projects have detailed and credible effort data recorded by phase (e.g. 12%).

Table 3: Average effort distribution for 34 projects for with very high Specification Effort

Phase	P	S	B	T	I
34 Projects	0,1%	98,5%	0,7%	0,5%	0,2%

Table 4 displays next the average distribution of effort for the 41 projects with the PSBTI profile, 62 projects with the PSBTI profile (when outliers and unusual patterns[4] are removed) and only 3 projects for the PSBT profile.

[3] The project with the greatest amount of effort had no size in function points assigned to it. This project is therefore of no use either for benchmarking or for estimation purposes..

[4] For the PSBT profile, 48 projects had 94% or more of the effort, in the specification phase alone.

Table 4: Average effort distribution for the PSBTI (excluding outliers and unusual distributions)

Profile	No. of Projects	Project Phases - % Effort P	S	B	T	I
PSBTI	41	9,1	24,7	39,1	19,7	7,3
PSBT	62	11,2	18,3	34,6	35,9	0
SBTI	3	0	27,6	49,0	15,3	8,1

3. Analysis of the ISBSG Normalized Effort

3.1. ISBSG normalization extent

It is reasonable to verify how information from R9 compares to the data used in 2001 by the ISBSG to derive their normalization procedure, and to ask how extensively has the normalization of phase effort been performed in ISBSG R9. Table 5 presents the 2001 data provided by the ISBSG [14] for the same three profiles (PSBTI, PSBT and SBTI)[5]. In the PSBTI profile, the 2001 effort for the planning and implementation phases are respectively 9.3% and 7.3%.

An important assumption made in the ISBSG -generated normalization process is that all three profiles, PSBTI, PSBT and SBTI, would have the same effort distribution if they had the same number of phases. This means, for instance, that the 9.3% effort in the planning phase of the PSBTI profile is used by the ISBSG to extrapolate the missing planning phase in the SBTI profile; put differently, this means that the effort recorded in the PSBT profile would correspond to 92.7% of the effort in a PSBTI profile. Similarly, the 7.3% effort in the implementation phase of the PSBTI profile is used by the ISBSG to extrapolate the missing effort data in the implementation phase of the PSBT profile. This illustrates the ISBSG -generated effort normalization process to recalculate what could have been the full phase cycle effort of the project, and from there the ISBSG -simulated normalized work effort was added as a field.

Table 5: Profile percentage from internal ISBSG correspondence (referring to 2001 data)

	P	S	B	T	I	# Projects
PSBTI	9.3	19.4	47.4	16.5	7.3	47
PSBT	11.3	22.0	39.1	20.3	(7.3)	11
SBTI	(9.3)	15.1	46.6	23.2	5.7	9

Comparison of data for the PSBTI profile in Table 4 (verified detailed data by project phase – 2005) and Table 5 (ISBSG -generated effort normalization approach – 2001) indicates that the planning and implementation phases have very similar ratios of effort (9.1 and 93 for Planning, and 7.3 for Implementation (for the other phases, the differences across the two datasets are much larger per phase)[6].

This suggests that the basis for extrapolation of the missing information for these two phases is reasonably consistent for these adjacent profiles across both the 2001 and 2005 releases.

However, it is not sufficient to find a way to compensate for missing phases, as the adjustment factor must not introduce any significant bias. In statistics, a population is characterized by its average and its mean value. Table 6 presents the average and means of the three profiles studied to date, including the subsamples (1,3,5) with only the phase indicators, and the R9 samples (2,4,6) with detailed effort data by project phase. In addition, line 7 corresponds to the concatenating of the projects with detailed effort data by project phase (that is, lines 2, 4 and 6 together).

[5] The ISBSG normalization process starts with an analysis of the Effort level data field [14]. This field is not made available in the ISBSG Releases. There was therefore no attempt to replicate as is the ISBSG normalization process for the effort data field.

[6] Considering that the ISBSG normalization includes an additional preliminary step – Footnote 5 – this introduces some extraneous factors in the comparability of tables 4 and 5; for instance, it could partly explain why there is higher Build percentage in the ISBSG normalized data for the PSBTI profile. Of course, this would need to be verified, should the full detailed data set be available.

To ensure that the aggregate profile of row 7 has no significant bias, it is compared with the population of row 1, which is the point of reference in terms of profile.

Table 6 : Statistical characteristics of profile samples

#	Profile	No. projects	Avg (Effort)	σ (effort)	Avg (size)	σ (size)
1	PSBTI	350	6379	12969	473	841
2	PSBTI	76	4417	5889	435	623
3	PSBT	405	4802	7557	544	981
4	PSBT	100	5311	6613	553	573
5	SBTI	92	7132	13793	352	539
6	SBTI	3	35042	50903	1121	191
7	**Cont2-4-6**	**179**	**5705**	**10189**	**512**	**604**

It is found that the average of the population on row 1 is not significantly different from that in row 7 (i.e. the hypothesis that both averages are the same is accepted if $Z_{calculated} = 0{,}65 < Z_{0.05} = 1{,}6$). However, the same cannot be said about the variance of the two populations; the normalization process changed the variance of the population (i.e. $F = s_1^2/s_7^2 = 12969^2/10189^2 = 1{,}62 > F_{0,95}(349, 178) = 1{,}24$).

Where $Z_{calculated} = \dfrac{6379 - 4417}{\sqrt{\dfrac{12969^2}{350} + \dfrac{10189^2}{179}}}$

3.2. Extensivity and range of the ISBSG normalization

To investigate the question of how extensively this effort normalization has been performed by the ISBSG for this field, the full data set was looked at. It was observed that normalization had been carried out for only 43% of the dataset: for 1,465 projects out of the 2,562 projects initially qualified for our analysis (e.g. 57%), the value of the added normalized effort field was the same as that of the summary work effort: that is, the ISBSG has considered that no data normalization was required or possible. Indeed, an analysis of the data set indicates that most of these projects are tagged to indicate that the effort recorded was for all project phases, or that the scope (number of phases) is unknown and that normalization is not possible.

To verify the range of normalization, the ratio between the summary work effort and normalized work effort has been calculated for each project within a specific profile.

Analysis of the ratios obtained using the above definitions indicates that they are not the same within profiles. For example, for PSBTI (profile 20), the calculated ratio between the summary work effort and the normalized work effort varies from 0.56 to 1. For instance, in the PSBTI profile, the project with a normalization ratio of 0.56 has effort for only one phase, which means that the ISBSG has extrapolated data for the other 4 phases all together. Similarly, even if the presence of effort determines the profile, some projects with the same profile may have slightly different normalization factors: it is not clear whether this is due to rounding discrepancies or to other factors that only have a very small impact on the normalization factor. In addition, for some projects the summary work effort is not normalized, even though some projects with the same profile (the same set of phases) have been normalized. This needs further investigation.

Of course, the ISBSG normalization process is based on more information on the nature of the projects (such as effort by developers, development team support, computer operations involvement and end-users) [14]; however, the normalization procedure is not transparent, making independent verification challenging.

4. Concluding remarks

Data provided by different organizations can generate a very heterogeneous database, as observed in section 2. To develop adequate software project estimation models using statistical techniques, the consistency of historical data is important. This paper has investigated this issue by looking into the consistency of the information of one of the most

important fields in the International Software Benchmarking Standards Group (ISBSG) repository, that is, the project effort data field. This paper has also presented an example of how effort data from projects that include a large number of project phases can be used for extrapolation, through a normalization process, to projects with fewer phases.

To analyze such data adequately, the ISBSG has a number of additional descriptive fields which can be used in modeling relationships across fields. However, these additional descriptive fields are a mixed blessing: on the one hand, they allow better characterization of projects with similar characteristics.

On the other hand, this leads to two major issues in data collection and data analysis:

- the missing data in many of those complementary explanatory fields lead to much smaller usable samples with less statistical scope for analysis and a corresponding challenge when extrapolation is desirable;
- with more than one field to indicate specific information, fields may contradict one another, leading to inconsistencies – data analysts must then either make an assumption on which field is the correct one or drop the projects containing contradictory information.

The ISBSG has attempted to tackle this issue on the variability of phases included in the project effort field by deriving a normalized work effort field. This paper has investigated this problem and reported on a number of related issues.

References

[1] Institute, P. M. (2000). *A guide to the project management body of knowledge (PMBOK guide)* (2000 ed.). Newton Square, Pa.: Project Management Institute.

[2] Angelis, L., Stamelos, I., & Morisio, M. (2001). *Building a software cost estimation model based on categorical data.* Paper presented at the Software Metrics Symposium, 2001. METRICS 2001. Proceedings. Seventh International.

[3] Boehm, B., Winsor Brown, A., Madachy, R., & Yang, Y. (2004). *A software product line life cycle cost estimation model.* Paper presented at the Empirical Software Engineering, 2004. ISESE '04. Proceedings. 2004 International Symposium on.

[4] Chulani, S., Boehm, B., & Steece, B. (1999). Bayesian analysis of empirical software engineering cost models. *Software Engineering, IEEE Transactions on, 25*(4), 573-583.

[5] Ferens, D. V. (1999). The conundrum of software estimation models. *Aerospace and Electronic Systems Magazine, IEEE, 14*(3), 23-29.

[6] Finnie, G. R., & Wittig, G. E. (1996). *AI tools for software development effort estimation.* Paper presented at the Software Engineering: Education and Practice, 1996. Proceedings. International Conference.

[7] Heemstra, F. J. (1990). *Software cost estimation models.* Paper presented at the Information Technology, 1990. 'Next Decade in Information Technology', Proceedings of the 5th Jerusalem Conference on (Cat. No. 90TH0326-9).

[8] Molokken, K., & Jorgensen, M. (2003). *A review of software surveys on software effort estimation.* Paper presented at the Empirical Software Engineering, 2003. ISESE 2003. Proceedings. 2003 International Symposium on.

[9] Levy, Z. (1990). *The software estimation dilemma, a small step forward.* Paper presented at the CompEuro '90. Proceedings of the 1990 IEEE International Conference on Computer Systems and Software Engineering.

[10] Paynter, J. (1996). *Project estimation using Screenflow Engineering.* Paper presented at the Software Engineering: Education and Practice, 1996. Proceedings. International Conference.

[11] Abran, A. N., I. Bourque, P. (2003). Contribution of Software Size in Effort Estimation.

[12] ISBSG. (2005). *Worldwide Software Development - the Benchmark*. from www.isbsg.org

[13] Maxwell, K. D. (2002). *Applied statistics for software managers*. Upper Saddle River, N.J.: Prentice Hall PTR.

[14] Lokan, C., (2005) Internal correspondence to the authors - Feb. 25, 2005.