

Verification of Productivity Models

(Chapter 6 – Software Project Estimation)

Alain Abran

(Tutorial Contribution: Dr. Monica Villavicencio)

Topics covered

1. Introduction
2. Criteria describing the relationship across variables
3. Verification of the model assumptions
4. Evaluation of models by their builders
5. Models already built?
6. Lessons learned

6.1 Introduction

In this chapter, you will see

- **Criteria to provide information on the quality of a productivity model**
 - The ranges of estimates
 - Uncertainties
- **Examples:**
 - Simple linear regression models
 - Single independent variable

6.2 Relationship across variables

Productivity models

Productivity models represent the relationships of a dependent variable with respect to:

- a single independent variable
 - ex. Effort or duration)
- a number of independent variables
 - Ex. Size, programming languages, etc,

Criteria to analyze relationship across variables

- Coefficient of determination (R^2)
- Error of an estimate (E)
- Relative error (RE)
- Magnitude of the relative error (MRE)
- Mean magnitude of the relative error (MMRE)
- Prediction level of the model (Pred(x))

Coefficient of determination (R^2)

- It describes the percentage of variability explained by the independent variable(s).
- Value between 0 and 1:
 - R^2 close to 1 \rightarrow strong relationship between the independent and dependent variables.
 - R^2 close to 0 \rightarrow there is a no relationship between the independent and dependent variables.

Error and Relative Error

$$\text{Error} = E = \text{Actual} - \text{Estimated}$$

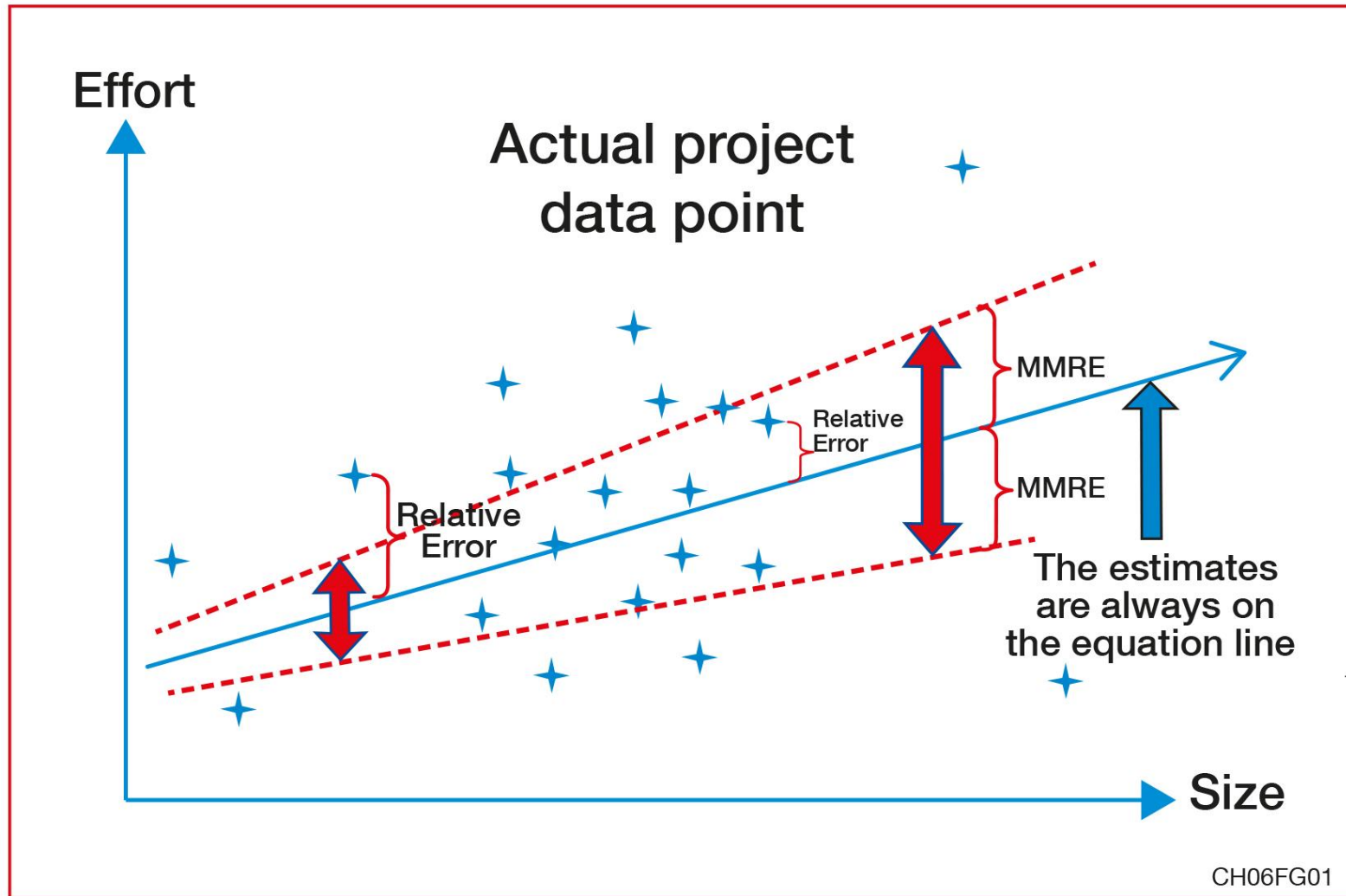
$$\text{Relative error} = RE = \left| \frac{\text{Actual} - \text{Estimate}}{\text{Actual}} \right|$$

Magnitude of Relative Error

Magnitude of the relative error: $MRE = |RE| = \left| \frac{Actual - Estimate}{Actual} \right|$

Mean magnitude of the relative error: $MMRE = \overline{MRE} = \frac{1}{n} \sum_{i=1}^n MRE_i$

Mean Magnitude of Relative Error – MMRE



Prediction level of the model

$$PRED(l) = \frac{k}{n}$$

where k is the number of projects in a specific sample of size n for which $MRE \leq l$.

Good models:



- Are based on high-quality data.
- The assumptions of the statistical models have been proved.
- The outcomes of the model adequately describe reality.

Example 1:

- Productivity model based on data from an organization where:
 - CMMI levels 4 or 5
 - Project data from a homogeneous environment
 - Strong expertise in development process
 - Model built with 1 or 2 independent variables (including software size)
- Expected values reached for 2 criteria of the model:
 - $R^2 \geq 0.8$
 - $\text{PRED}(0.20) = 0.80$

Example 2

- Productivity model based on data from an organization where:
 - CMMI levels 4 or 5
 - Dataset of projects with brand new technologies and a variety of development environments
- Expected values reached for 2 criteria of the model:
 - $R^2 \geq 0.5$
 - $\text{PRED}(0.20) = 0.50$

Example 3

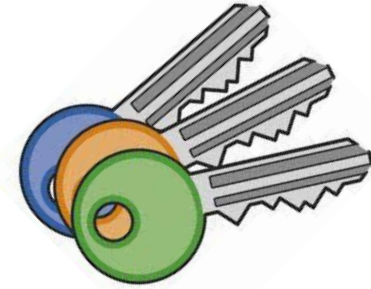
- Productivity model based on data from an organization where:
 - CMMI level 1: Ad hoc development process
 - Development environment is typically risky and unpredictable.
 - Model built with 1 or 2 independent variables
- Expected values reached by 1 criteria of the model:
 - $R^2 < 0.5$

More advanced criteria

- The root of the mean square error
- The relative root of the mean square error
- P-value of the statistical parameters
- Testing the significance of the model
- Analysis of the residuals

6.3 Verification of the Assumptions

Key conditions



- Productivity models built using simple parametric regression techniques require:
 - A normal distribution for both the dependent and independent variables.
 - No outlier that influences the model unduly.
 - A large enough dataset (typically, 30 data points for each independent parameter).

The sample size: ISBSG recommendations (1)

- *“For regression to make sense, you need quite a lot of data. A sample size of 30 or more is desirable, although a sample size of 20 or more should provide reasonable results.”*
- *“Not much can be concluded from a sample size of less than 10 projects.”*

ISBSG: International Software Benchmarking Standards Group – www.isbsg.org see also Chapter 8 of this book

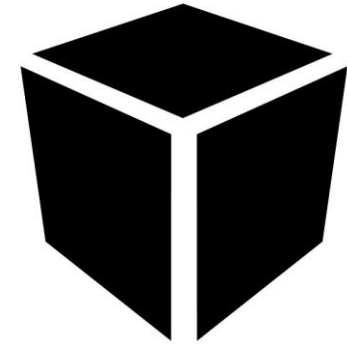
The sample size: ISBSG recommendations (2)

- *“There are several other restrictions on when regression can be used. For example, the data:*
 - *should be ‘normally distributed’ (this is not the case in the common situation where there are many small values and only a few large ones); and*
 - *should not show a fan-shaped pattern (small spread for small values, and a large spread for large values).*
 - *These restrictions are often violated in software engineering datasets.”*

“Before performing a regression analysis, you need to study your data carefully to make sure they are appropriate. For more information, consult any standard textbook on statistics.”

6.4 Evaluation of models by their builders

Evaluation by builders



<http://myqaengineering.blogspot.ca/>

- **Model builders should document:**
 - The information on the variation ranges
 - Levels of uncertainty

- **White-box models:**
 - the details of such models are documented (structure, equations, parameters).

- **Black box models:**
 - Their data, the models themselves, & their performance on specific datasets are hidden.

6.5 Models already built

Evaluation of already built models

Estimation models should be evaluated:

- using different datasets from the ones that were used to build them.
 - How do the models perform in other environments?
- by independent researchers using empirical datasets
- by practitioners using historical data within their own organizations.

Evaluation of already built models

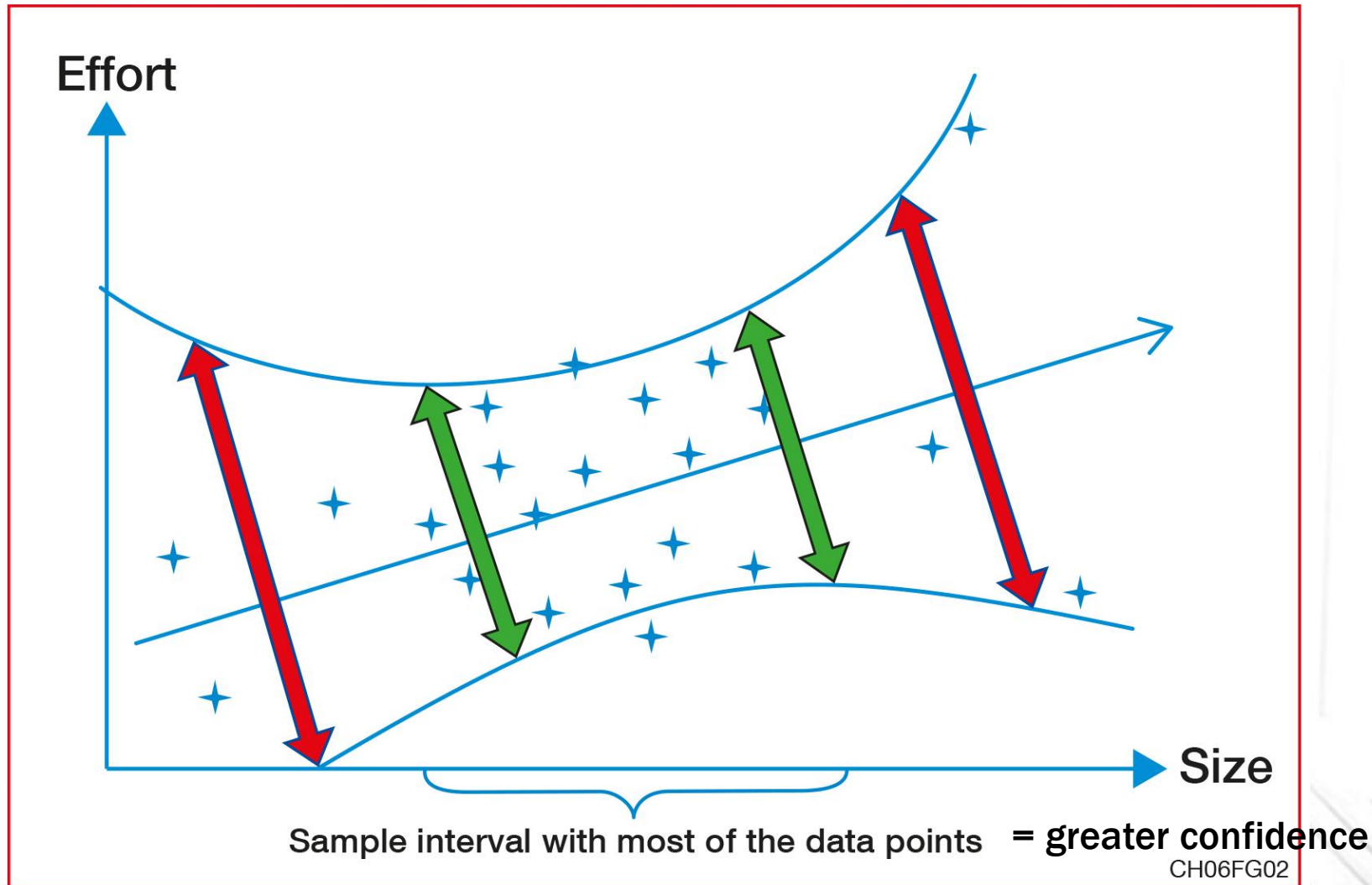
- **Small-scale replication studies**
 - Small sample
- **Large-scale replication studies**
 - Larger sample
- **Results of studies:**
 - Estimation errors in white-box models are much lower than black-box models.

6.6 Lessons learned

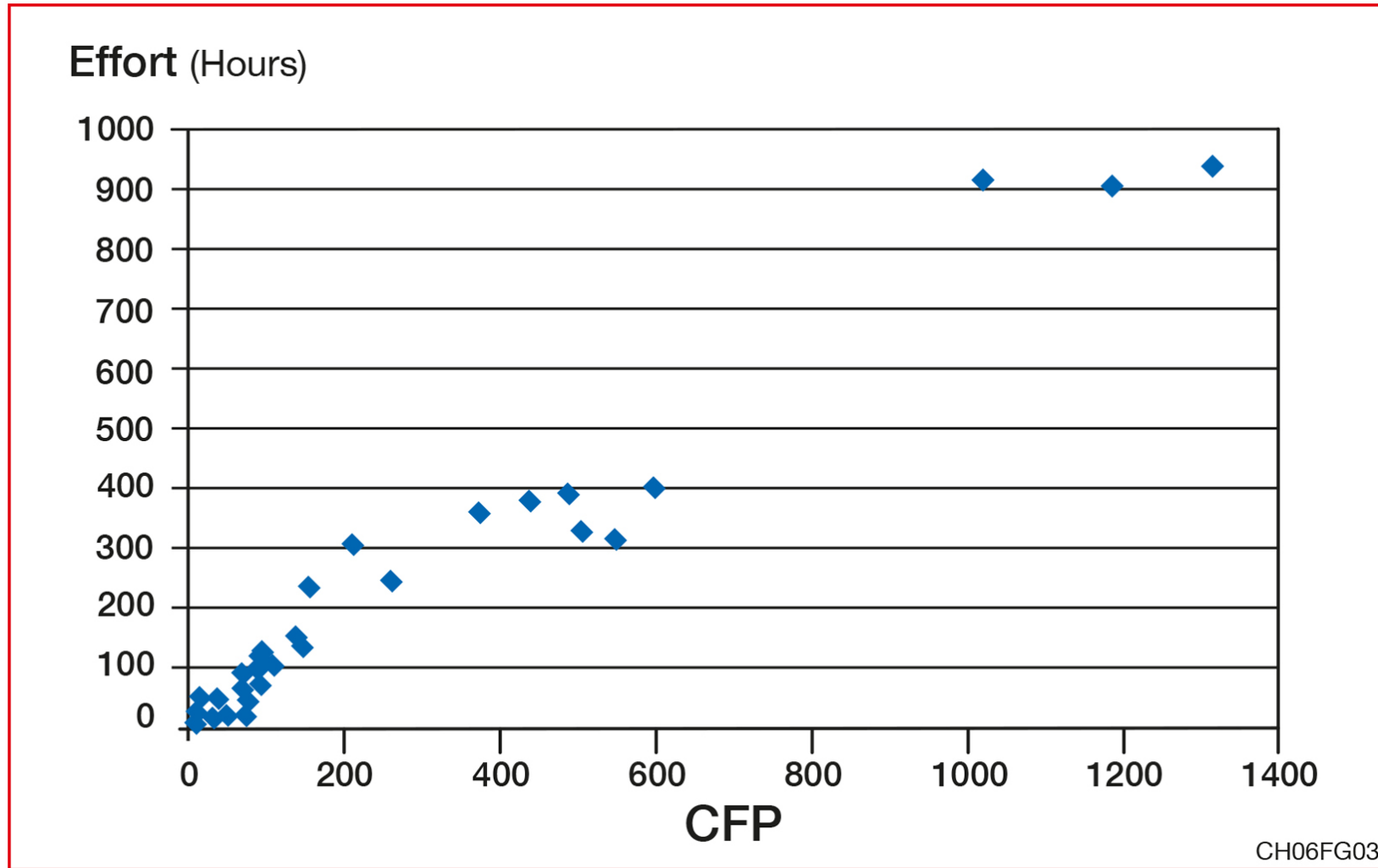
Distinct models by size range

- **General practice in software engineering:**
 - To build a single model with a large number of cost drivers (independent variables) but without taking into account the size range of the independent variable.
- **Good practices in statistical analysis:**
 - The assumptions of a model have to be verified (Normal-Gaussian distribution of the independent and dependent variables)
 - The results have to be interpreted within the range of size and effort for which there are enough data points.

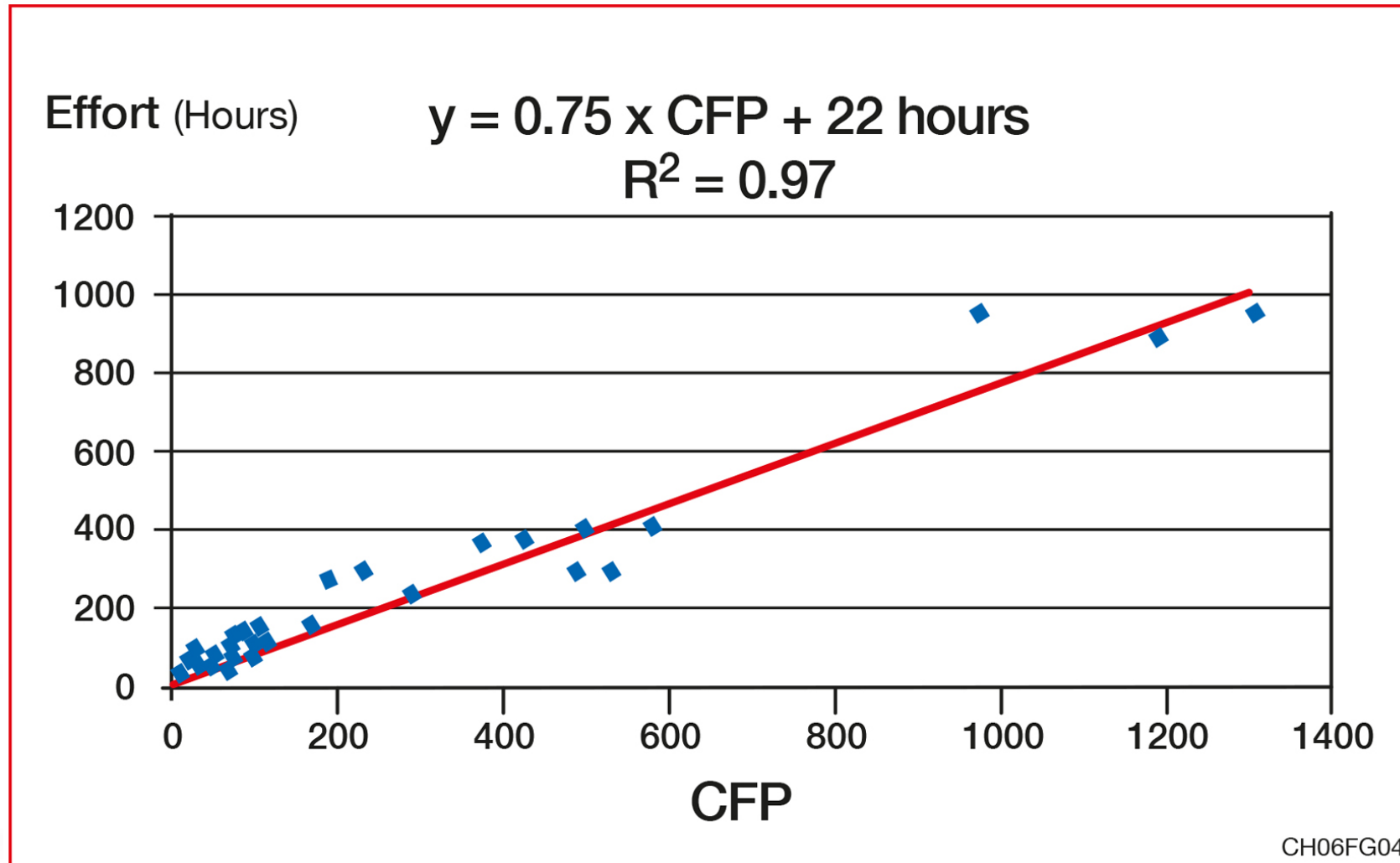
Confidence intervals and sample intervals



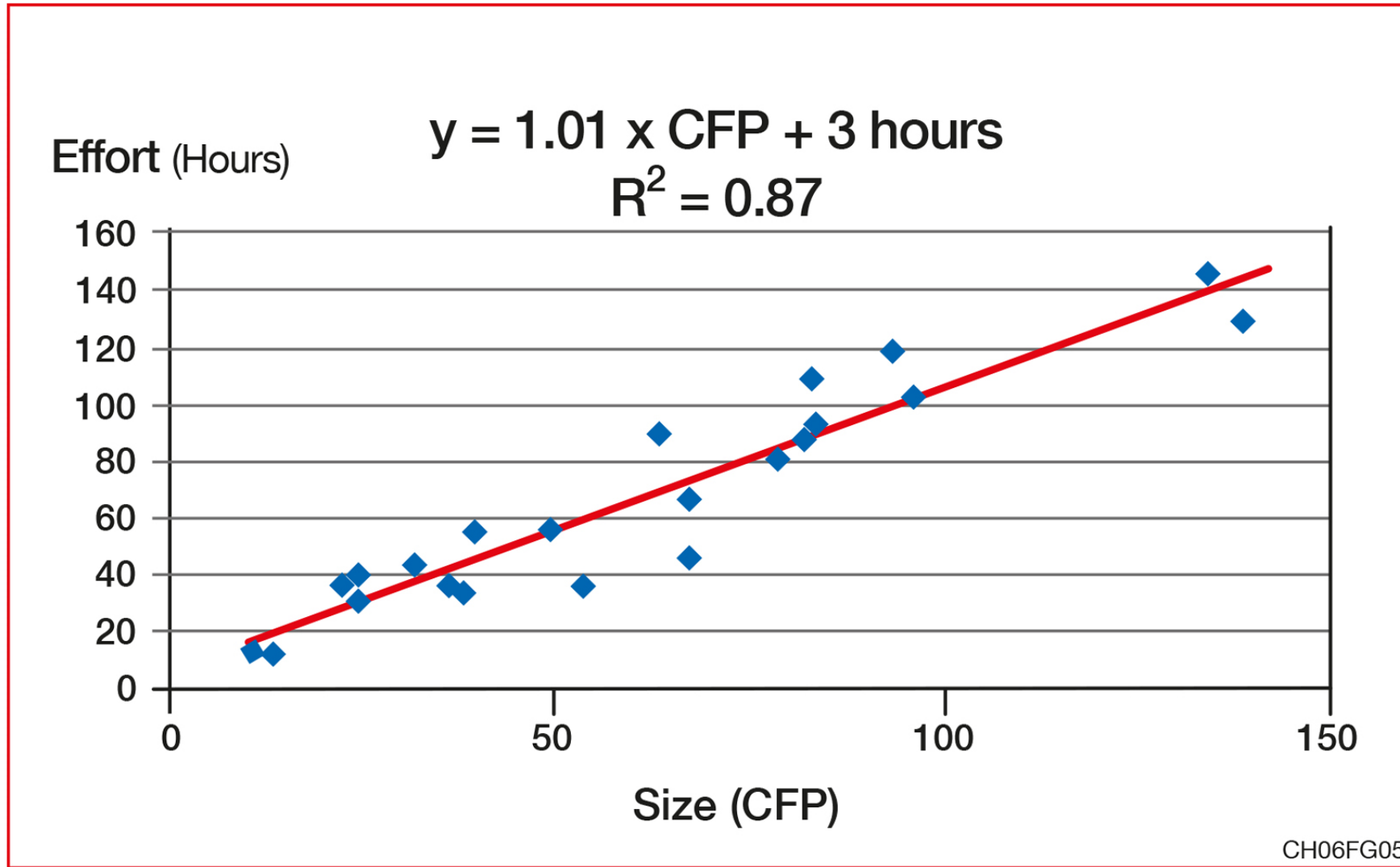
2D representation of 34 projects



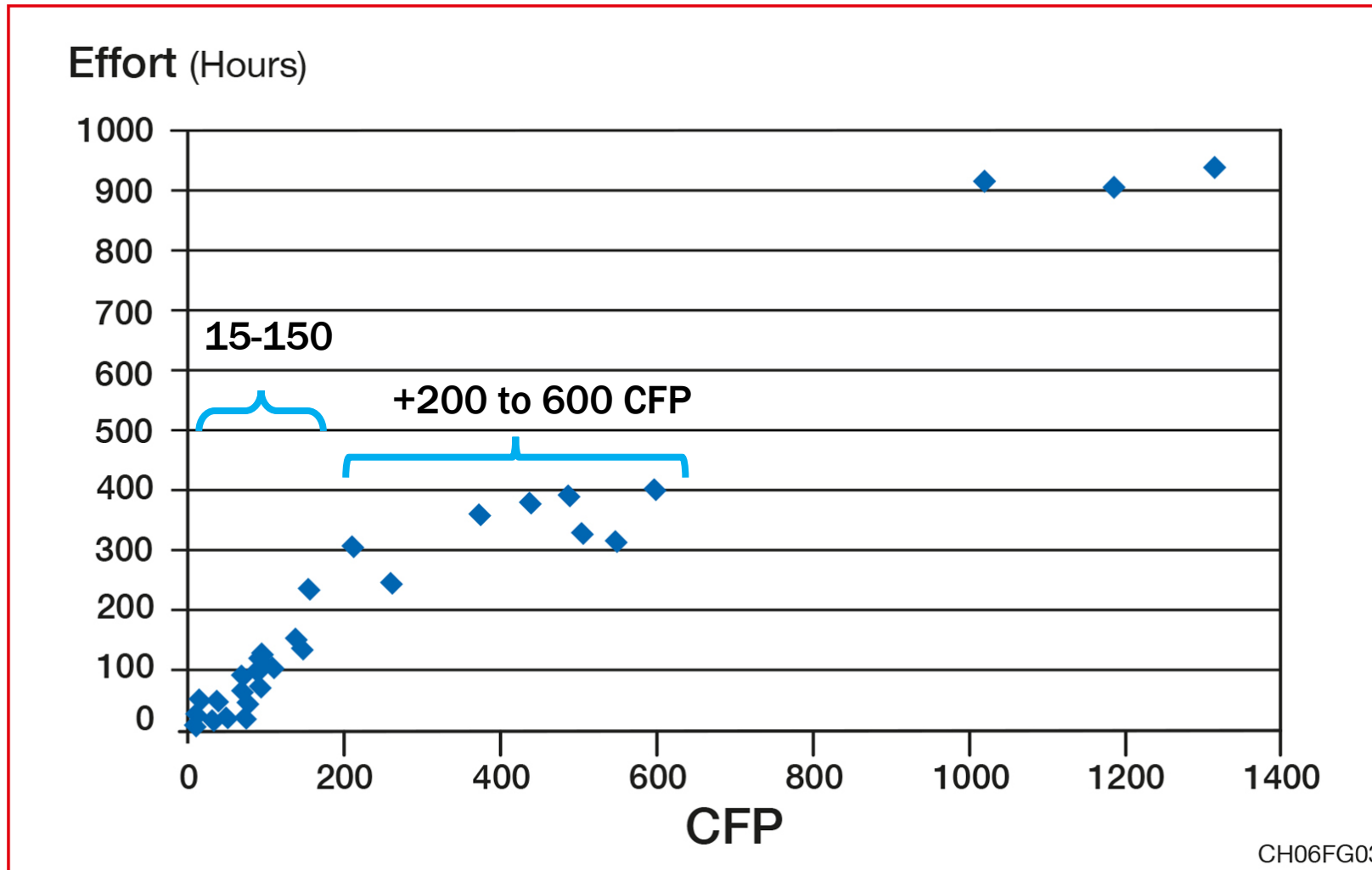
Single linear regression model



Regression model for the low range interval (15 to 150 FP and N=22)

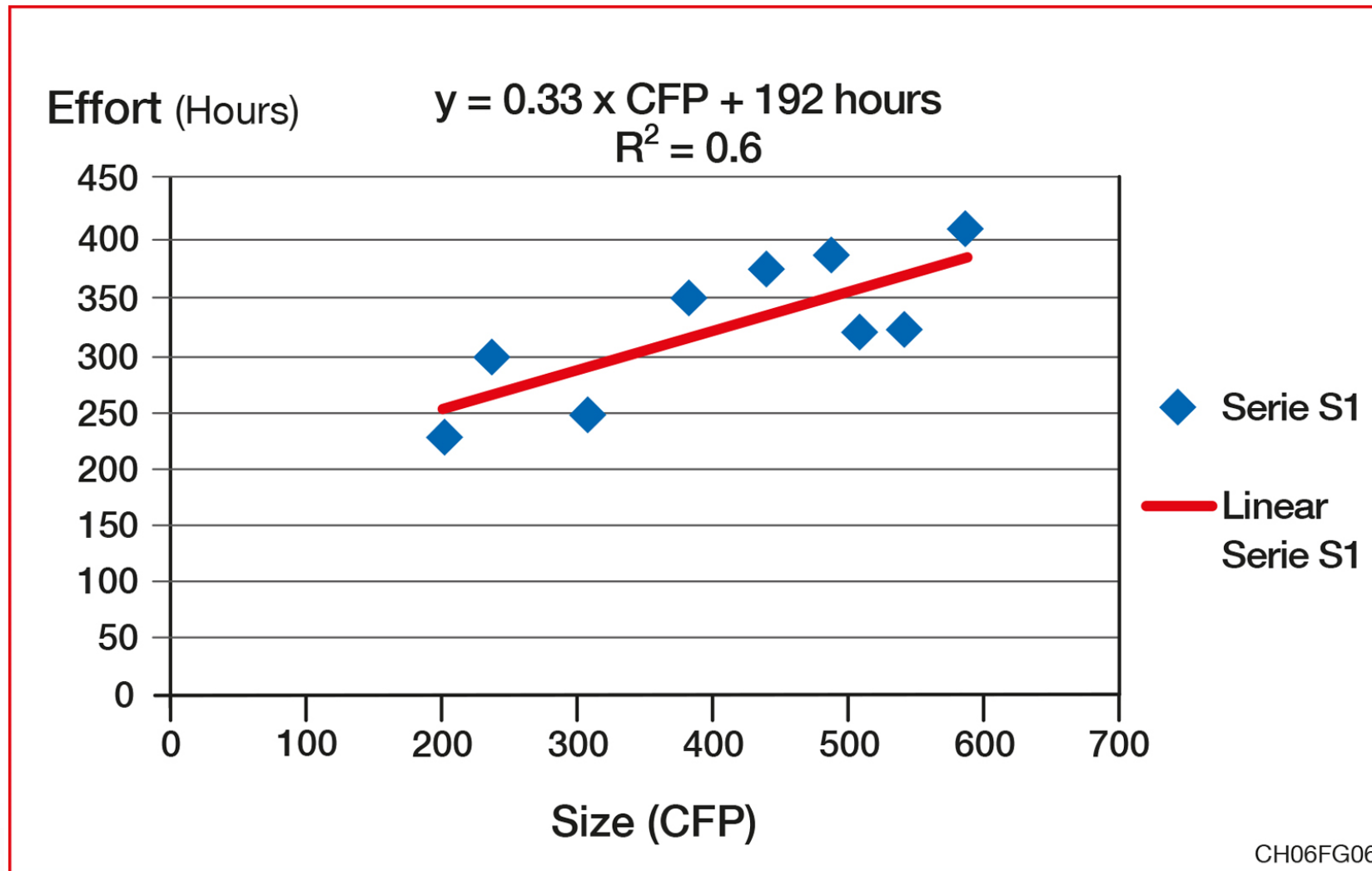


34 projects: 2 intervals of distinct density of data points



CH06FG03

Regression model for the mid-range interval (200 to 600 CFP and N=9)



Lesson learned

The model with the highest R^2 (0.97) is not the best model to use.

- it is poorly built, and uninformed users would be overconfident in its performance.

Exercises

1. List five criteria for verifying the performance of a model.
2. What does an R^2 of 1.0 mean?
3. What does an MMRE represent?
4. What does $PRED(10) = 70\%$ mean?
5. What are the three conditions that must be met to ensure the proper use of the statistical regression technique?
6. What is the recommended size of a dataset sample for a simple regression technique?
7. Does this mean that building a productivity model with your own productivity data for a small sample size is useless? Discuss.
8. Take the average productivity ratio of the software projects in your organization's repository of projects. What is the error range on this average productivity ratio? What criteria should you be using?
9. Look at whether or not authors and vendors have documented the quality of the productivity model they are proposing to practitioners. Comment on your findings. What are the short-term consequences for your organization? What are the long-term consequences for your profession?

Exercises

10. What is a replication study? How do you carry out a replication study?
11. Table 6.2 presents a comparison of the results of two estimations models. For a project using Visual Basic, which model would you trust more? Up to what point?
12. If you are using (or plan to use) a statistically based productivity model, what should you be verifying?
13. If the data points for one of the cost factors in your model have a non normal distribution, how do you take that into account in analyzing the quality of the productivity model built with such data?
14. Describe to your project management, in layman's terms, the practical meaning of a coefficient of regression – R^2 – and its relevance to them when selecting a specific budget for a project on the basis of the productivity model (built from your organizational dataset, or from an outside repository).
15. Take the data from Table 6.3 and Figure 6.3. Explain to your manager why it is preferable for him to have two or three productivity models rather than a single one.

Term Assignments

1. Take project 10 in Table 6.3, and multiply the effort by 3. What impact does this have on the regression model? Take project 31, and multiply the effort by 3. What impact does that have on the regression model? Explain the impact of both on the quality of the regression model.
2. Select one of the estimation models available free of charge on the Web, and test it using any available dataset (from your own organization, from the literature, or from the ISBSG repository). Document its quality using the criteria listed in section 4.1.
3. Typically, you need 20 to 30 data points for each variable in a statistical study. Look at a study on software productivity or productivity models, and discuss the significance of the statistical findings based on the number of data points available and the number of cost factors taken into account in these models.