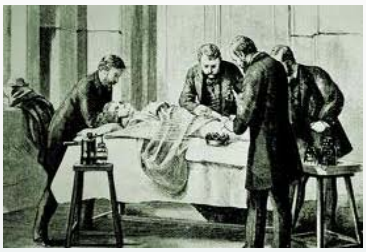
Historical Data Use in Software Estimation – Caveat Emptor!

David DeWitt
Senior Consultant
Galorath Incorporated
PSM Conference July 29th, 2010

Why do we love historical data?

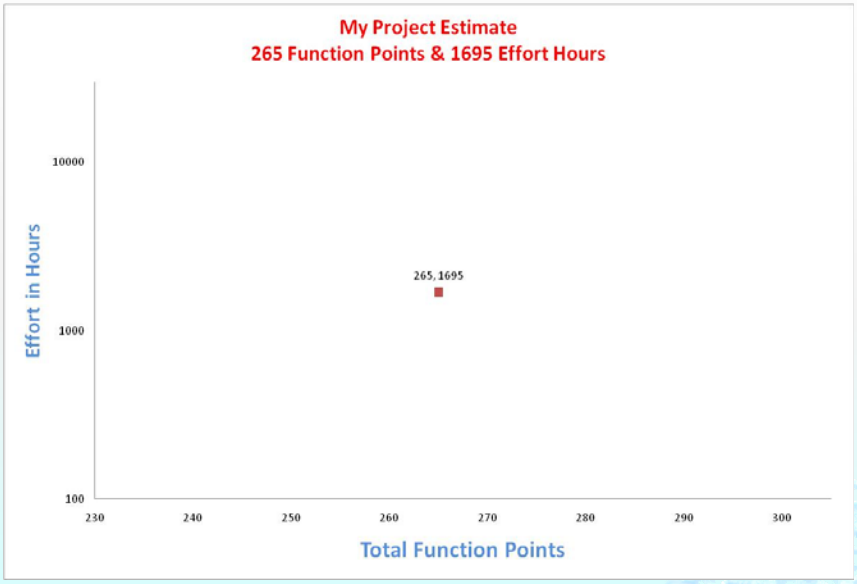
Either we enjoy doing forensics…



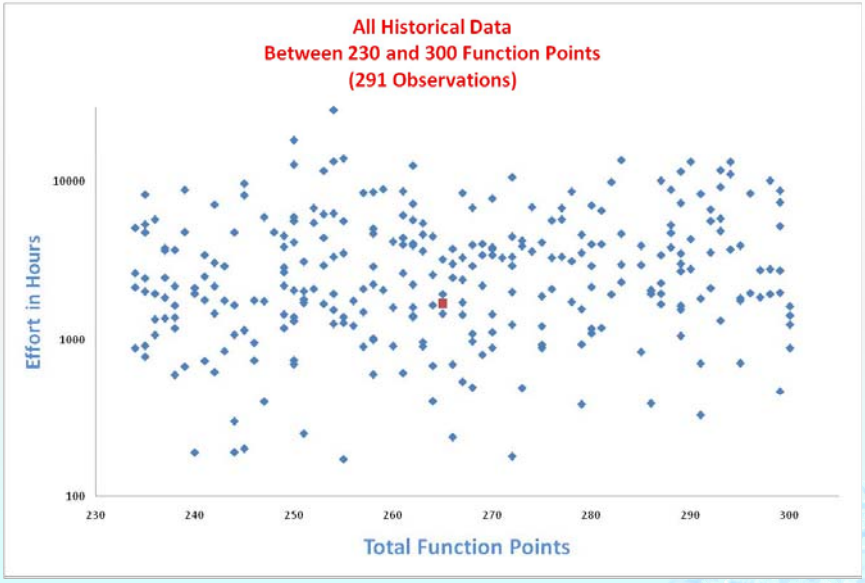
… or crave Nostalgia!

In the sunset of dissolution, everything is illuminated by the aura of nostalgia, even the guillotine.— Milan Kundera (The Unbearable Lightness of Being)

You have an estimate – Now what?



Use Historical Data to evaluate your estimate!



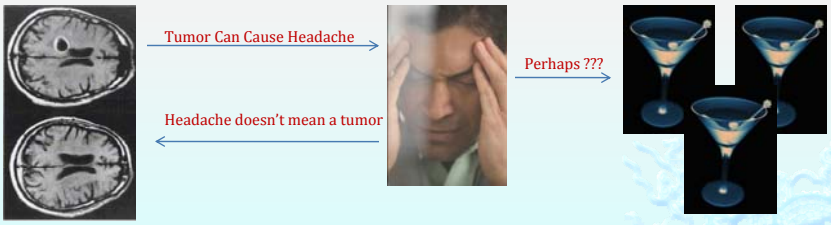
It's easy to dig deeper and deeper to justify an estimate!

Parsing historical data looks logical …
… but there are many potential/likely flaws

- The Error of Casual Analysis (False Association) - accumulating facts with minimal generalizations
- Narrative fallacy - when the set of connected and disconnected facts are picked to fit a story
- Fallacy of silent evidence - seeing only what has been recorded and remaining ignorant of the missing evidence
- Ludic Fallacy - assuming the data to be statistically analyzed is; complete, unaffected by small variations, and not intentionally corrupted

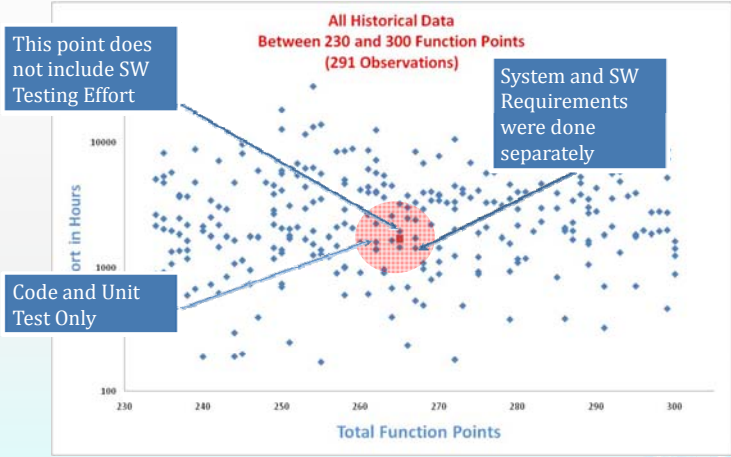
The Error of Causal Analysis
Creating a False Association

- Correlation does not imply causation
  - Just because two data points may sit side by side doesn't mean they are the same or will have the same outcome
- Casual analysis is a recognized error in Medicine.



The Error of Causal Analysis

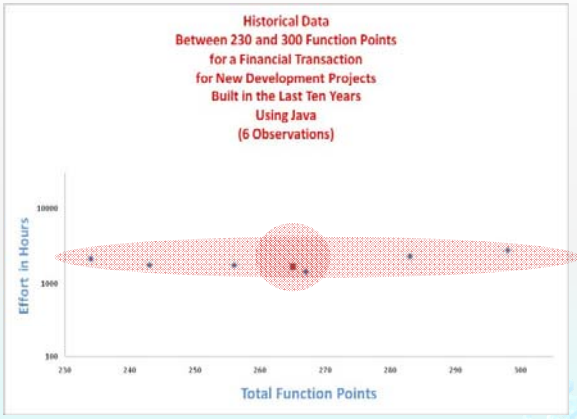
But is often encountered in data analysis?



There is no correlation that estimate of size 'X' is an equivalent of historical projects of about 'X'.

Narrative fallacy

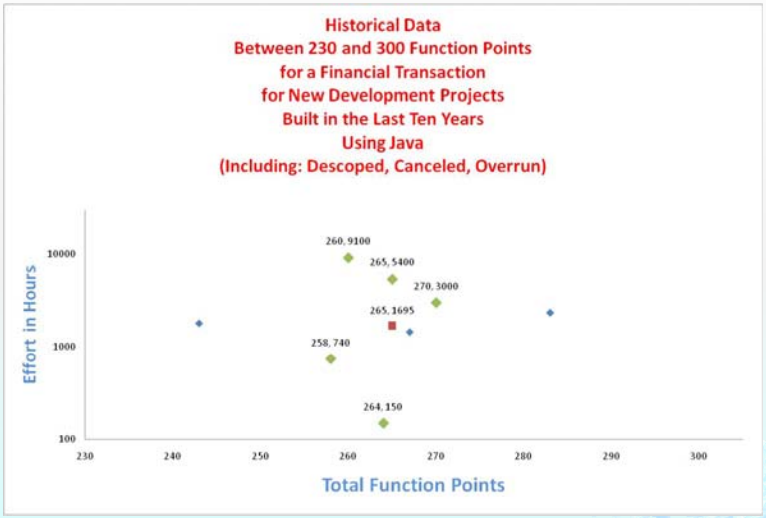
As our data becomes less cluttered…



It's almost automatic – and natural - to look for clumping and create relationships that likely have no historical correlation to one another.

Fallacy of Silent Evidence

What about what we don't know?



How confident would you feel if the Silent Evidence was visible?

Fallacy of Silent Evidence

Drives the conversation about data quality

- There are fundamental problems with historical data:
  - It's costly to obtain
  - Very difficult to catalog and store
  - Hard to manipulate, retrieve and review
  - Can often be wrong
  - Can be intentionally laced with inaccuracies
  - is incomplete

… Care to add to the list?

Note: Silent Evidence is different from Errors of Omission – Silent Evidence includes data that was intentionally omitted or not properly represented.

Ludic Fallacy

- "Ludic" is from the Latin *ludus*, meaning "play, game, sport, pastime"
- Are you now or has someone previously been playing games with historical data
- In this context, Ludic Fallacy is wrongly assuming historical data is complete, unaffected by small variations, and not intentionally corrupted

Ludic Fallacy

- Most common omissions from historical data ranked in order of significance

Sources of Cost Errors

1) Unpaid overtime by exempt staff
2) Charging time to the wrong project
3) User effort on software projects
4) Management effort on software projects
5) Specialist effort on software projects
   - Human factors specialists
   - Data base administration specialists
   - Integration specialists
   - Quality assurance specialists
   - Technical writing specialists
   - Education specialists
   - Hardware or engineering specialists
   - Marketing specialists
   - Metrics and function point specialists
6) Effort spent prior to cost tracking start up
7) Inclusion/exclusion of non-project tasks
   - Departmental meetings
   - Courses and education
   - Travel

Source: Capers Jones, *Errors And Omissions In Software Historical Data: Separating Fact From Fiction,* August 17, 2009

Ludic Fallacy

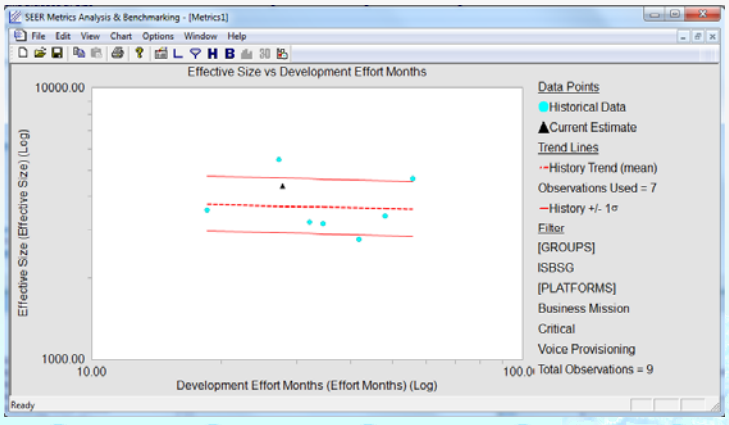
- Typical results reviewing customer historical data

Activities Performed	Completeness of historical data
01 Requirements	Missing or Incomplete
02 Prototyping	Missing or Incomplete
03 Architecture	Missing or Incomplete
04 Project planning	Missing or Incomplete
05 Initial analysis and design	Missing or Incomplete
06 Detail design	Incomplete
07 Design reviews	Missing or Incomplete
08 Coding	Complete
09 Reusable code acquisition	Missing or Incomplete
10 Purchased package acquisition	Missing or Incomplete
11 Code inspections	Missing or Incomplete
12 Independent verification and validation	Complete
13 Configuration management	Missing or Incomplete
14 Integration	Missing or Incomplete
15 User documentation	Missing or Incomplete
16 Unit testing	Incomplete
17 Function testing	Incomplete
18 Integration testing	Incomplete
19 System testing	Incomplete
20 Field testing	Missing or Incomplete
21 Acceptance testing	Missing or Incomplete
22 Independent testing	Complete
23 Quality assurance	Missing or Incomplete
24 Installation and training	Missing or Incomplete
25 Project management	Missing or Incomplete
26 Total project resources, costs	Incomplete

Source: Capers Jones, *Errors And Omissions In Software Historical Data: Separating Fact From Fiction*, August 17, 2009

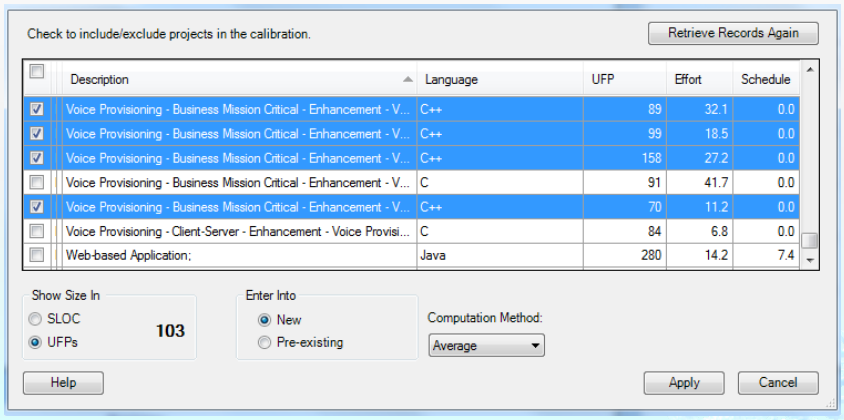
Some Safe Ways to Use Historical Data

- Given a set of similar projects that you're familiar with… 'Does my estimate look reasonable? '
- That is not the same as "Is it correct!"



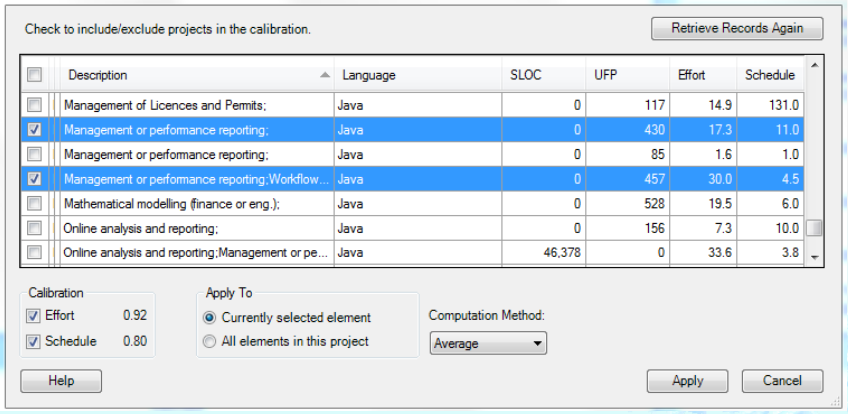
Some Safe Ways to Use Historical Data

- Use it to provide an analogy for potential size range



Some Safe Ways to Use Historical Data

- Use it to apply analogous calibration factors.



Some Basic Rules

1. Know from where the data comes
2. Know the age of the data
3. Know what is included (Phases, People, Processes, …)
4. Know the different sizing schemas and how applied
5. Know what was intentionally left out (and why)
6. Look for internal score cards
   - ISBSG uses A-D ratings for data quality
7. Avoid Nostalgia!

Something to take away…

Given a puddle of water, one cannot know the shape, size, or even quantity of ice cubes that formed the puddle.

Historical databases are much the same…

Caveat Emptor!

Thank You!
Questions?

Galorath Incorporated
David DeWitt
(321) 848-3410
ddewitt@galorath.com
www.galorath.com

SEER *Designed to Meet the Challenges of Today's IT Projects*

IBM Business Partner

Microsoft CERTIFIED Partner