

# Competitive Analytics - IT Parallels

September, 2011

## International Software Measurement & Analysis Conference

Richmond, VA.

**Joe Schofield**

joescho@joejr.com

Sandia National Laboratories

# A Quick Look Back & Updates on ISMA Presentations

|      |   |
|------|---|
| 2011 | <i>Competitive Analytics – IT Implications</i>  |
| 2010 | <i>Using Function Point Analysis for Software Measurement and Estimating</i> ; International Software Measurement & Analysis Conference; Sao Paulo, Brazil; September 14, 2010<br><i>Repeatable and Relevant Functional Software Measurement using Function Point Analysis</i> ; Lockheed Martin Measurement Workshop; Colorado Springs, CO.; October 13, 2010<br><i>Certified Function Point Specialist Examination Guide</i> ; Garmus, Russac, Edward; 2010 (editor)  |
| 2009 | <i>Leaning Lean Six Sigma for Results</i> ; ISMA; September, 2009<br><i>When Did Six Sigma Stop Being a Statistical Measure?</i> ; CrossTalk, April 2006<br><i>Lean Six Sigma - Real Stories from Real Practitioners</i> ; Albuquerque, N.M.; N.M. SPIN; August 2005<br><i>Six Sigma &amp; Software Engineering: Complement or Collision</i> ; Albuquerque, N.M.; N.M. SPIN; August, 2004   |
| 2008 | <i>Estimating Latent Defects Using Capture-Recapture: Lessons from Biology</i> ; Arlington, VA.; 2008 International Software Measurement and Analysis (ISMA) Conference; September 18, 2008<br><i>Economics of Software Quality</i> ; Capers Jones; July, 2011 (editor)<br><i>Beyond Defect Removal: Latent Defect Estimation with Capture Recapture Method</i> ; CrossTalk, August 2007 (reprinted in IFPUG's MetricViews, Winter 2008)<br><i>Latent Defect Estimation - Maturing Beyond Defect Removal using Capture-Recapture Method</i> ; QAI QAAM Conference; September 10, 2008   |
| 2007 | <i>'Manda, Panda, and the CMMI(R)</i> ; Las Vegas, NV.; 2007; ISMA Conference; September 14, 2007   |
| 2006 | <i>Defect Collection &amp; Analysis – The Basis of Software Quality Improvement</i> ; ISMA Conference, September, 2006<br><i>Defects: Perspective on Prevalence and Prevention</i> , chapter in forthcoming IFPUG book<br><i>Defect Management through the Personal Software Process<sup>SM</sup></i> ; CrossTalk, September 2003<br><i>The Team Software Process<sup>SM</sup> – Experiences from the Front Line</i> ; Software Quality Forum; Arlington, Virginia, March; 2003<br><i>Measuring Software Process Improvement - How to Avoid the Orange Barrels</i> ; System Development, December 2001<br><i>Usable Metrics for Software Improvement within the CMM</i> ; Software Quality Forum 2000; Santa Fe, N.M.; April, 2000  |
| 2004 | <i>Applying Lean Six Sigma to Software Engineering</i> ; IFPUG Conference; September, 2004  |
| 2003 | <i>Amplified Lessons from the Ant Hill – What Ants and Software Engineers Have in Common</i> ; IFPUG Conference, Sept., 2003<br><i>Lessons from the Ant Hill - What Ants and Software Engineers Have in Common</i> ; Information Systems Management, Winter 2003  |
| 2002 | <i>Counting KLOCs – Software Measurement's Ultimate Futility (I can't do this anymore, or who am I fooling?, or why not count ants?)</i> ; IFPUG Conference; September, 2002<br><i>Counting Lines of Code: Virtually Worthless for Estimating and Software Sizing</i> , IT Metrics and Productivity Journal; December, 2009<br><i>Is There a Weakest Link After All?</i> , IT Metrics and Productivity Journal; December, 2009<br><i>Is There Value to using Lines of Code for Measuring People After All?</i> , IT Metrics and Productivity Journal; December, 2009<br><i>Lines of Code - Statistically Unreliable for Software Sizing?</i> ; Computer Aid, Inc.; Webinar; October 14, 2008<br><i>The Statistical Case Against the Case for using Lines of Code in Software Estimation</i> ; 4th World Congress on Software Quality; Bethesda, MD.; September 17, 2008<br><i>The Statistically Unreliable Nature of Lines of Code</i> ; CrossTalk, April 2005 (Reprinted at least twice, cited by NIST <i>Metrics and Measures</i> <a href="http://samate.nist.gov/index.php/Metrics_and_Measures">http://samate.nist.gov/index.php/Metrics_and_Measures</a> )<br><i>A Practical, Statistical, and Criminal Look at the Use of Lines of Code as a Software Sizing Measure</i> ; N.M. SPIN; March, 2004 |

**(Conduct) Survey Here!**

# A Survey Related to Decision-Making<sup>1</sup>

1. How many countries have at least one McDonald's?
2. What is the range of a Minuteman (III) Missile?
3. How long (minutes & seconds) was the song "Stop in the Name of Love" recorded by the Supremes?
4. If the air temperature (F) is 5 degrees below zero and the wind speed is 15 mph, what would be the wind chill?
5. How many sovereign rulers has England had in the last 1000 years?
6. What is the average cost of testing in software development relative to total cost?
7. How many meters high is the Sears Tower?
8. The Airbus A380 has 525 seats when configured for three classes. How many seats would it hold if all the seats were economy class?
9. How many inches does the hair on a human head grow in a year?
10. On average, a software development project projected to take 17 months actually takes how long?

<sup>1</sup>These questions were posed to ISMA Cinco! attendees by Dr. Ricardo Valerdi, MIT, as part of his keynote presentation. Used with permission 8-16-2011.

See also: The Human Element of Decision Making in Systems Engineers: A Focus on Optimism; Ricardo Valerdi & Craig Blackburn, INCOSE

# Response Range Considerations

| Q# | General Knowledge  | Bounds   |
|----|--|--|
| 1  | Been in or at or seen on TV  | Less than number of countries                          |
| 2  | Been in AF or a defense “expert”   | Less than $\frac{1}{2}$ the circumference of the earth |
| 3  | Radio listener in or since 1960s   | Less than 4 minutes                                    |
| 4  | Live in a cold climate   | Likely not a positive number                           |
| 5  | Knowledgeable about English history (monarchs rule longer than presidents) | $> 1, < 100$   |
| 6  | Software engineer or project manager                                       | $>1, < 100$  |
| 7  | Been to Chicago; an architect  | $> 1, < \sim 1000$                                     |
| 8  | Aviation buff; a pilot   | $> 525, < \sim 1000$                                   |
| 9  | Have hair; cut hair  | $> 3, < 10$  |
| 10 | Software engineer or project manager                                       | $>17, < 51$  |

# The Questions, Results, and Observations . . .

|    | Questions  | R1R><br>R2R | R2R ><br>R1R | Tied |
|----|--|-------------|--------------|------|
| 1  | How many countries have at least one McDonald's?   | 31          | 20           | 4    |
| 2  | What is the range of a Minuteman III Missile?  | 20          | 28           | 7    |
| 3  | How long (minutes & seconds) was the song "Stop in the Name of Love" recorded by the Supremes?                                     | 24          | 22           | 9    |
| 4  | If the air temperature (F) is 5 degrees below zero and the wind speed is 15 mph, what would be the wind chill?                     | 25          | 20           | 10   |
| 5  | How many sovereign rulers has England had in the last 1000 years?  | 31          | 15           | 9    |
| 6  | What is the average cost of testing in software development relative to total cost?  | 16          | 26           | 13   |
| 7  | How many meters high is the Sears Tower?   | 32          | 18           | 5    |
| 8  | The Airbus A380 has 525 seats when configured for three classes. How many seats would it hold if all the seats were economy class? | 21          | 26           | 8    |
| 9  | How many inches does the hair on a human head grow in a year?  | 26          | 16           | 13   |
| 10 | On average, a software development project projected to take 17 months actually takes how long?                                    | 19          | 30           | 6    |

Thanks to Jacqueline Dominguez who contributed to the development of the formats and underlying formulae used above.

R1R = Round 1 responses

R2R = Round 2 responses

- (R1R) Unsatisfied with allowing respondents to cavalierly answer questions with a wide, but *meaningless* range, I attempted to alter the "experiment" by offering a prize for *closeness*. One might think of this as the "horse shoe game" twist—closer matters.
- (R2R) On four questions (2, 6, 8, 10) when respondents were placed under pressure to be "better than their peers" their answers actually got worse. (Daniel Pink talks about this in his book *Drive* as it relates to cognitive tasks.)
- More alarming was that two of the four questions (6, 10) in which answers got worse, the questions were related to software project management.
- The answers to those two questions **ONLY** are refutable; that is, there are similar studies in which answers would be distinguishably different.
- Questions 2 & 8 are the other two questions in which response grew worse. Since these questions were posed (by me) to folks who work in an area related to #2, this is troubling.

# The Questions, Results, and Observations (cont'd) . . .

|    | Questions  | R1R ><br>R2R | R2R ><br>R1R | Tied |
|----|--|--------------|--------------|------|
| 1  | How many countries have at least one McDonald's?   | 31           | 20           | 4    |
| 2  | What is the range of a Minuteman III Missile?  | 20           | 28           | 7    |
| 3  | How long (minutes & seconds) was the song "Stop in the Name of Love" recorded by the Supremes?                                     | 24           | 22           | 9    |
| 4  | If the air temperature (F) is 5 degrees below zero and the wind speed is 15 mph, what would be the wind chill?                     | 25           | 20           | 10   |
| 5  | How many sovereign rulers has England had in the last 1000 years?  | 31           | 15           | 9    |
| 6  | What is the average cost of testing in software development relative to total cost?  | 16           | 26           | 13   |
| 7  | How many meters high is the Sears Tower?   | 32           | 18           | 5    |
| 8  | The Airbus A380 has 525 seats when configured for three classes. How many seats would it hold if all the seats were economy class? | 21           | 26           | 8    |
| 9  | How many inches does the hair on a human head grow in a year?  | 26           | 16           | 13   |
| 10 | On average, a software development project projected to take 17 months actually takes how long?                                    | 19           | 30           | 6    |

- The sum of all variances for R1R was 51,032,348 and for R2R 870,167 – a decrease of 58.6 times the variation of R1R.
- Excluding one set of (outlier) answers from both question sets, the sum of the variances for R1R was 1,009,619 and for R2R 868,643 or just 1.16 times the variance of R1R.
- Excluding the next two largest outliers from both question sets, the sum of the variances for R1R was 453,277 and for R2R 864,146. This exclusion may mislead the casual observer into believing that the sum of the variances – when respondents are asked to “compete” – actually increases the variances significantly. This appears to be a distortion of the data and what the data is telling us.
- Excluding the next largest “outlier” from both questions sets, the sum of the variances for R1R was 288,836 and R2R 251,859.

# New questions were developed because:

1. Questions needed answers that were repeatable when searched and researched.
2. Questions needed answers that were current in the literature; that is, less subject to change over time.
3. Increased liberty was desired to analyze the data based on unforeseen inquiry.
4. Questions needed to have less global sensitivity.
5. Questions needed to have less industry specific sensitivity.
6. Questions could be added to trigger desired specific interest in respondents (latent defect estimation as an example below)

Three samples of fish are taken from a lake. 70 fish were found in the largest sample, a total of 90 fish in the other two samples. 50 fish were common to both the sample of 70 and the sample of 90 fish. What's the predicted number of fish in the lake that were not captured in the samples?  $((70 * 90) / 50) - (70 + 90 - 50) = 130 - 110 = 20$

Ref: *Beyond Defect Removal: Latent Defect Estimation with Capture Recapture Method*; CrossTalk, August 2007



# New questions include:

1. On average, how far is the sun from Neptune when compared to the distance of the sun to Earth? (or distance from Earth to sun)
2. How many tenths of an inch do fingernails grow within a year?
3. What is the flight distance in kilometers from New York City to Mumbai India?
4. How many feet above sea level is Mt. Kilimanjaro?
5. How long is the song “Hey Jude”, originally recorded by the Beatles in 1968?
6. The Oasis of the Seas is listed as the world’s largest cruise ship (circa 2011). What is the maximum passenger capacity listed for this vessel?
7. If it’s 80 degrees Fahrenheit, what’s the temperature in Celsius?
8. What is the estimated maximum number of military deaths that resulted from WWII expressed in millions?
9. What is the number of gallons in a US barrel of oil?
10. The gestation period of an elephant is how many months?

# Response Range Considerations

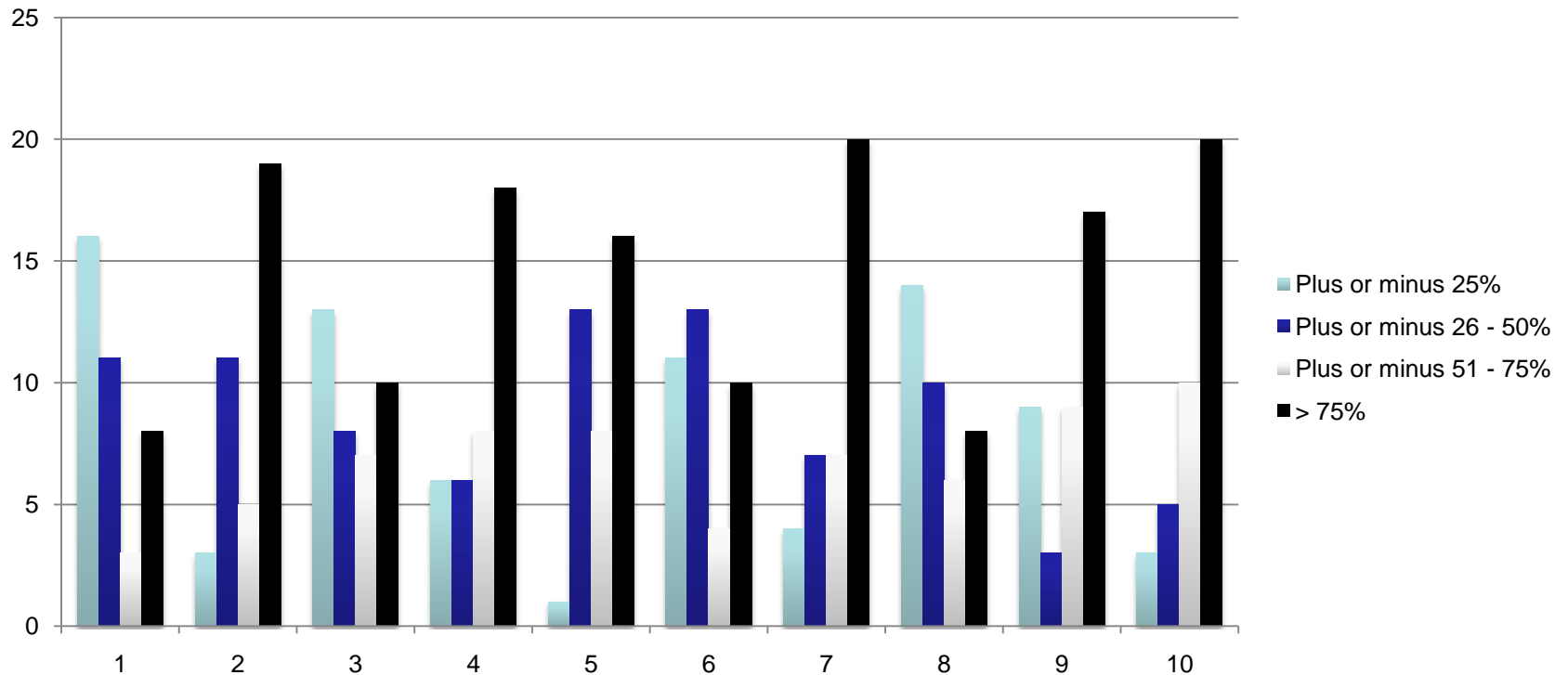
| Q# | General Knowledge                           | “Logical” Bounds                             |
|----|---|--|
| 1  | Astronomy interest                          | Uncertain                                    |
| 2  | Have or cut “nails”                         | < six inches                                 |
| 3  | Traveler, global geography                  | < ½ the global circumference times about .6? |
| 4  | African, African traveler, mountain climber | > 10,000, < ~30,000 ft.                      |
| 5  | Beatles fan, music enthusiast from 60s      | > 3, < 10                                    |
| 6  | Cruise traveler, trivia expert              | > 3000, < 10,000                             |
| 7  | Celsius familiar, meteorologist             | > 1, < 80                                    |
| 8  | War, history buff                           | > 5,000,000 < 100,000,000                    |
| 9  | Oil person, savvy consumer                  | > 30, < 60                                   |
| 10 | Elephantologist, veterinarian               | > 3, < 20                                    |

# New questions & results . . .

| Q# | Question   | R1R > R2R | R2R > R1R | Tied |
|----|--|-----------|-----------|------|
| 1  | On average, how far is the sun from Neptune when compared to the distance of the sun to Earth?                                     | 46%       | 34%       | 20%  |
| 2  | How many tenths of an inch do fingernails grow within a year?  | 51%       | 31%       | 17%  |
| 3  | What is the flight distance in kilometers from New York City to Mumbai India?  | 49%       | 49%       | 3%   |
| 4  | How many feet above sea level is Mt. Kilimanjaro?  | 49%       | 31%       | 20%  |
| 5  | How long in seconds is the song "Hey Jude" originally recorded by the Beatles in 1968?   | 26%       | 57%       | 17%  |
| 6  | The Oasis of the Seas is listed as the world's largest cruise ship. What is the maximum passenger capacity listed for this vessel? | 54%       | 34%       | 11%  |
| 7  | If it's 80 degrees Fahrenheit, what's the temperature in Celsius?  | 43%       | 34%       | 23%  |
| 8  | What is the estimated maximum number of military deaths that resulted from WWII in millions?                                       | 40%       | 46%       | 14%  |
| 9  | What is the number of gallons in a US barrel of oil?   | 57%       | 34%       | 9%   |
| 10 | The gestation period of an elephant is how many months?  | 40%       | 37%       | 23%  |

**Note that ties are typically from respondents who enter the same answer for both sets of responses.**

# Responses within ranges. . .



## Note:

- The values represented as greater than 75 percent can exceed 100 percent; a number of them exceed 10,000 percent, a small number exceed 60,000 percent
- Before considering the excessive percentages outliers or irrelevant, review the values for the Taurus project

# Project and product performance values:

> 75 percent industry variation

|   | % Variation |
|---|-------------|
| UK's National Health Care Service, budgeted at \$12B, closer to \$24B in 2007 <sup>1</sup>          | 100         |
| London Stock Exchange's Taurus project: estimated 6 million pounds, actual 800 million <sup>1</sup> | 13,200      |

**Note:** The *mode* of the responses to 6 (of the 10 questions) fell in the range > 75 percent; questions 2, 4, 5, 7, 9, and 10 (fingernails, Kilimanjaro, "Hey Jude", Fahrenheit to Celsius, gallons to barrel, gestation)

<sup>1</sup>MIS Quarterly Executive Vol. 6 No. 2 / June 2007; University of Minnesota

# Project and product performance values: 50 - 75 percent industry variation

|   | % Variation |
|---|-------------|
| 70% of large IT programs don't reach their goals in the allotted time and budget <sup>1</sup>         | 70          |
| Rework consumes almost 50 percent of resources for lower quality large software projects <sup>2</sup> | 50          |

## Note:

- While about 10 percent of responses fell in this range, the range was not the statistical *mode* for any question
- the largest 500 U.S. companies lose in excess of \$14 billion a year because of failed technology projects<sup>1</sup>

<sup>1</sup>Taming Information Technology Risk A New Framework for Boards of Directors, National Association of Corporate Directors, 2011

<sup>2</sup>Software Project Failure Costs Billions. Better Estimation & Planning Can Help; Dan Galorath on Estimating:, June 7, 2008

# Project and product performance values: 0 - 50 percent industry variation

|  | % Variation |
|--|-------------|
| Formal inspections find twice as many defects at 1/5 the cost of testing. <sup>1</sup>   | 20          |
| Defect detection and removal account for at least 40 percent of total software costs, exceeding the cost of its development. <sup>1</sup>  | 40          |
| Thirty percent of project effort can be traced to rework <sup>2</sup>  | 30          |
| Rework consumes almost 50 percent of resources for lower quality large software projects <sup>3</sup>  | 50          |
| Reworking defective requirements, design, and code typically consumes 40 to 50 percent or more of the total cost of most software projects and is the single largest cost driver. <sup>4</sup> | 40 – 50     |

## Note:

- **3 questions' responses *mode* fell in the range of 0 – 25 percent (astronomical units, flight kilometers, WWII deaths)**

<sup>1</sup>MIS Quarterly Executive Vol. 6 No. 2 / June 2007; University of Minnesota

<sup>2</sup>*Dr. Dobb's Report*; informationweek; July 12, 2010; study by Dean Lefingwell, 1997

<sup>3</sup>*Software Project Failure Costs Billions. Better Estimation & Planning Can Help*; Dan Galorath on Estimating.; June 7, 2008

<sup>4</sup>Jones, Capers. *Estimating Software Costs*, New York: McGraw-Hill, 1998

# Takeaways

*"The quality of our decision-making is deteriorating across the board. Not because the people in charge are stupid. But because they're all running too fast, making too many decisions, too fast, about too many things they know too little about."* - Alvin Toffler, futurist

- **Eagerness to do well may result in worse estimation results (not including the influence of optimism)**
- **Cost & schedule are easier to track, so we do (PMPs, Earned Value, actual vs. estimated, Scorecards, budget reports, budget analysts, forecasts, schedulers)**
- **Limited use of or availability of objective, relevant, and quantitative “actual” data during estimating**
- **Changing and evolving technology becomes an excuse for under performance because we don’t manage and understand the changes well:**
  - **Why are we adopting this tool, method, approach?**
  - **How does the change impact our defect rate?**
  - **How does it transform our business processes and culture towards intended future states?**



# Are these comparisons reasonable?

- The data seems intuitively correct versus the following Department of Education example: Roughly 90 million Americans over age 16—almost half that category's total population—are, as far as most workplaces are concerned, basically unfit for employment; yet, 71 percent of these felt like they read well or very well.
- Both survey and project results were measured, not confabulated
- Both sets of results are drawn from populations of software engineers and project managers
- The comparison of results suggests that the differences aren't that significant; that is, they are well within the range of credibility
- Comparisons can trigger introspection
  - What are our numbers?
  - How do we compare to the broader population?
  - How do we use these numbers to improve in targeted areas?
- This is my data from which I've drawn insights and value. Where's the data to support your assertions?