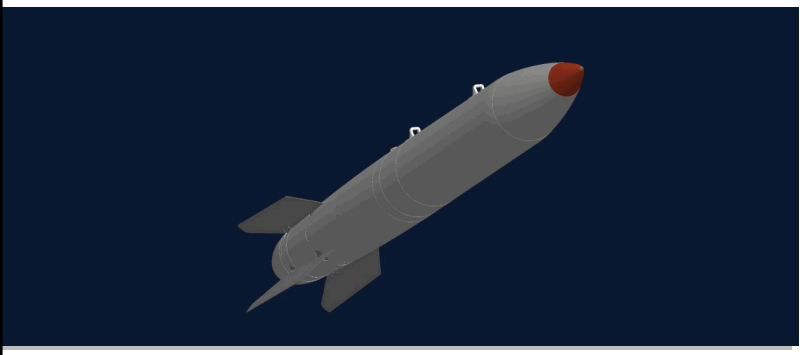
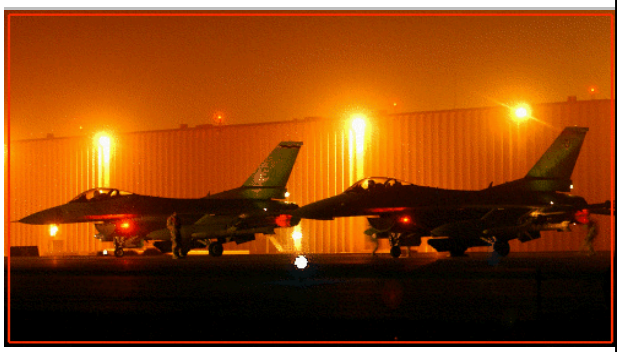
Exceptional service in the national interest

SAND2012-1049P

Sandia National Laboratories





Data Torturing

And The Misuse of Statistical Tools

Marcey Hoover, Ph.D.

Senior Manager, Surety Engineering

U.S. DEPARTMENT OF ENERGY | NNSA

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Statistics Quotes …

- **Facts are stubborn, but statistics are more pliable.**
  - *Mark Twain*
- **In ancient times they had no statistics so they had to fall back on lies.**
  - *Stephen Leacock, Canadian economist*
- **Like dreams, statistics are a form of wish fulfillment.**
  - *Jean Baudrillard, French sociologist*
- **Statistics are no substitute for judgment.**
  - *Henry Clay*
- **There are lies, damned lies and statistics.**
  - *Mark Twain*

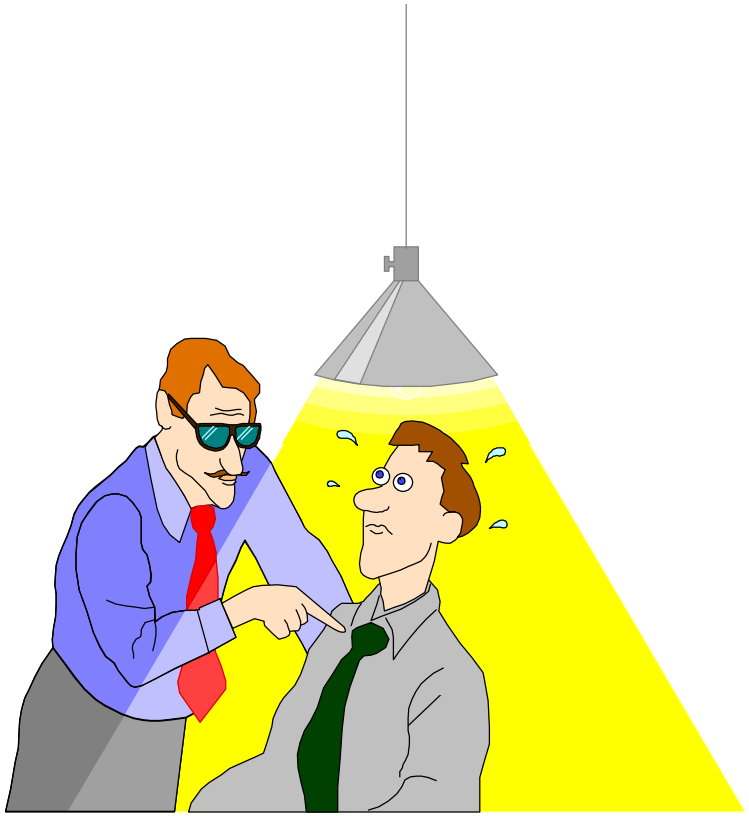
Objective of This Talk

Sandia National Laboratories

- Demonstrate the pitfalls of common statistical tools
- Give you three questions that can be used to prevent data torturing …

Data Torturing Definition

- When the data analysis process goes beyond reasonable interpretation of the facts, it becomes data torturing

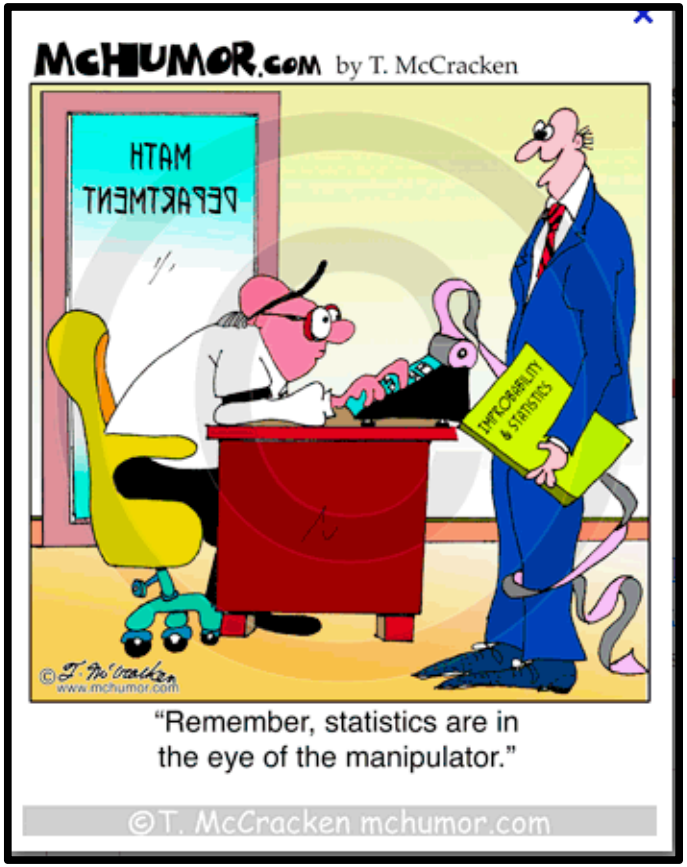


Remarks on Data Torturing

Torture numbers, and they'll confess to anything. (G. Easterbrook)

It is impossible to tell how widespread data torturing is. Like other forms of torture, it leaves no incriminating marks when done skillfully. And like other forms of torture, it may be difficult to prove even when there is incriminating evidence.

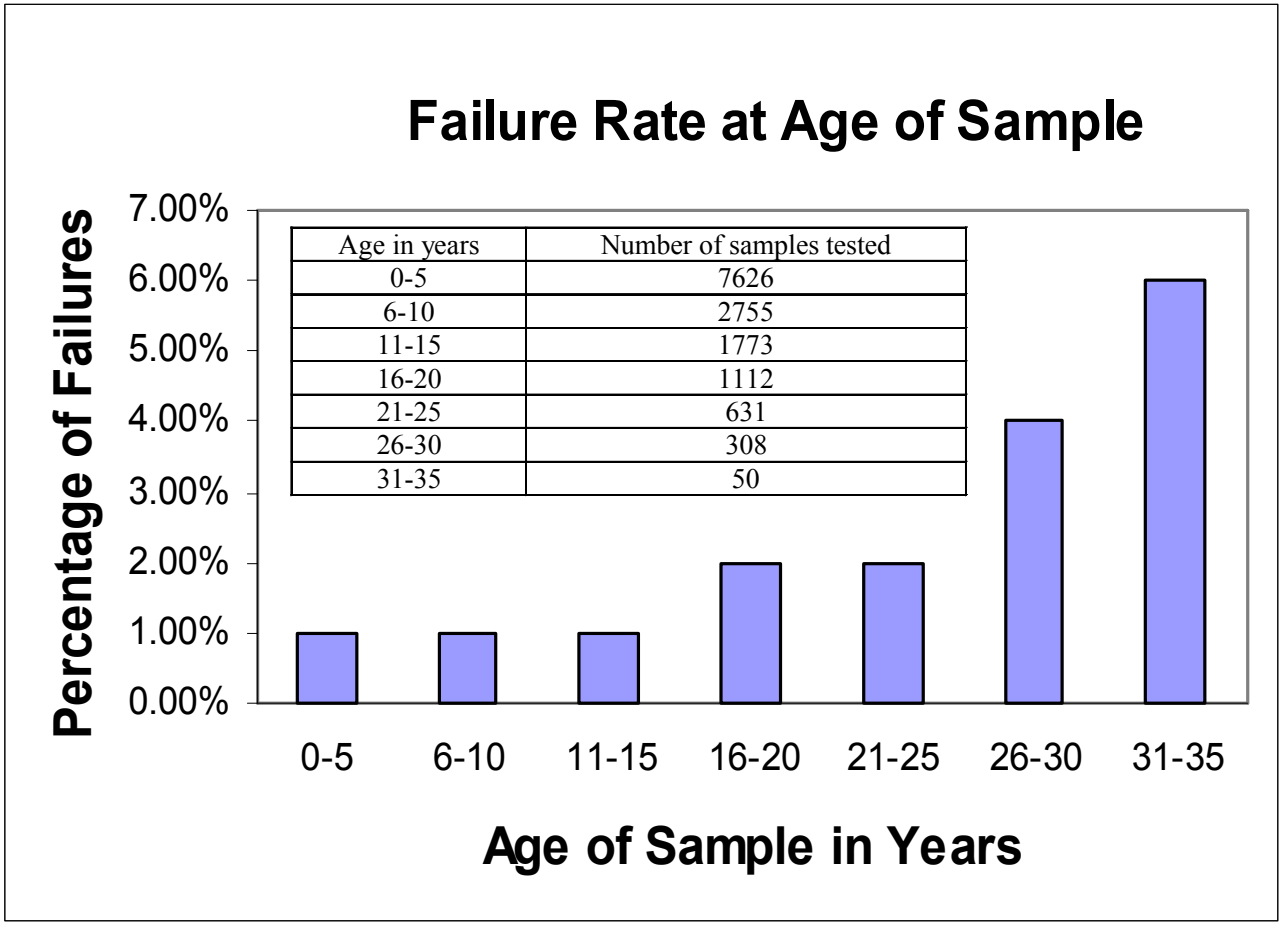
(J. Mills)



Data Torturing
via the “Misuse of Statistical Tools”

- Percentages
- Tabulations
- Comparisons to Averages
- Predictions and Extrapolations

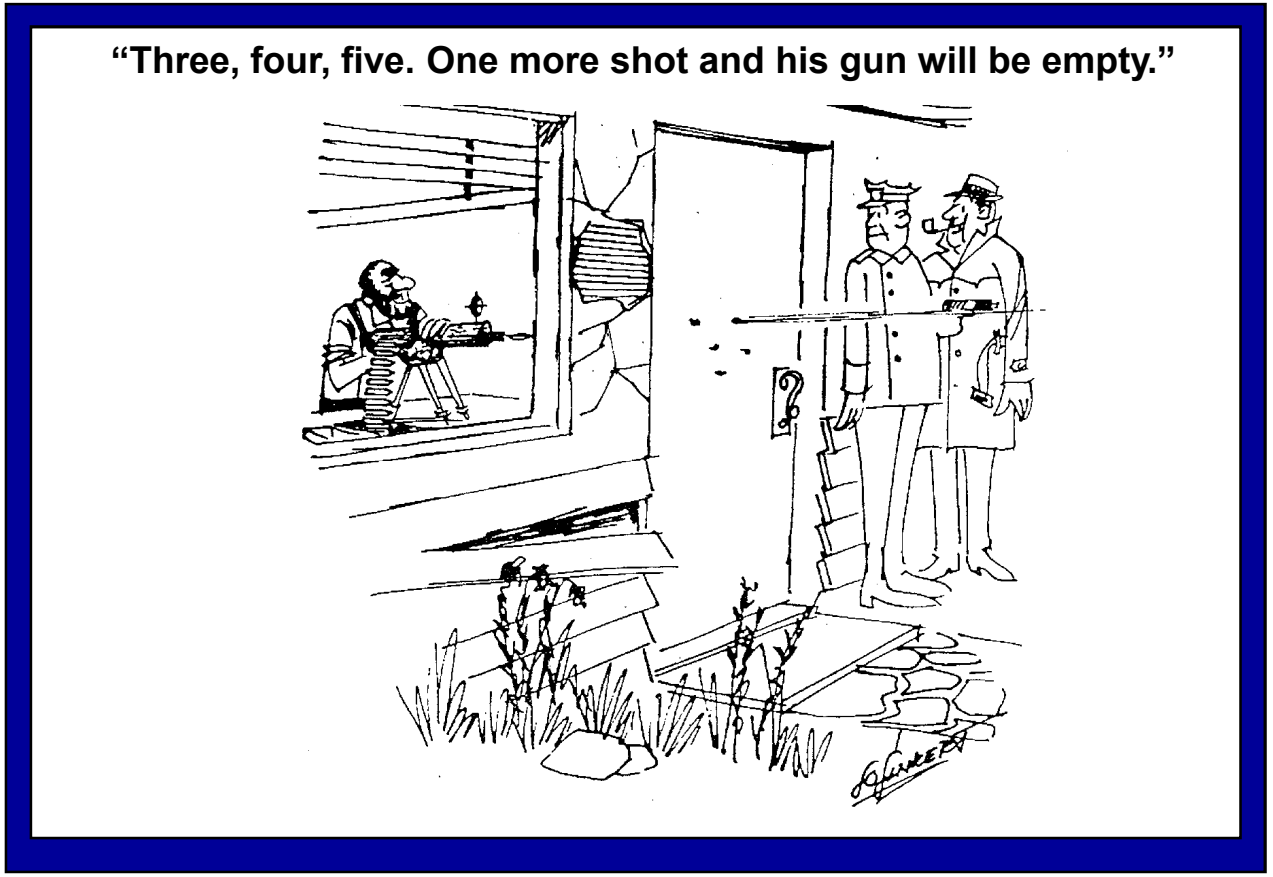
Percentages



Age in years	Number of samples tested
0-5	7626
6-10	2755
11-15	1773
16-20	1112
21-25	631
26-30	308
31-35	50

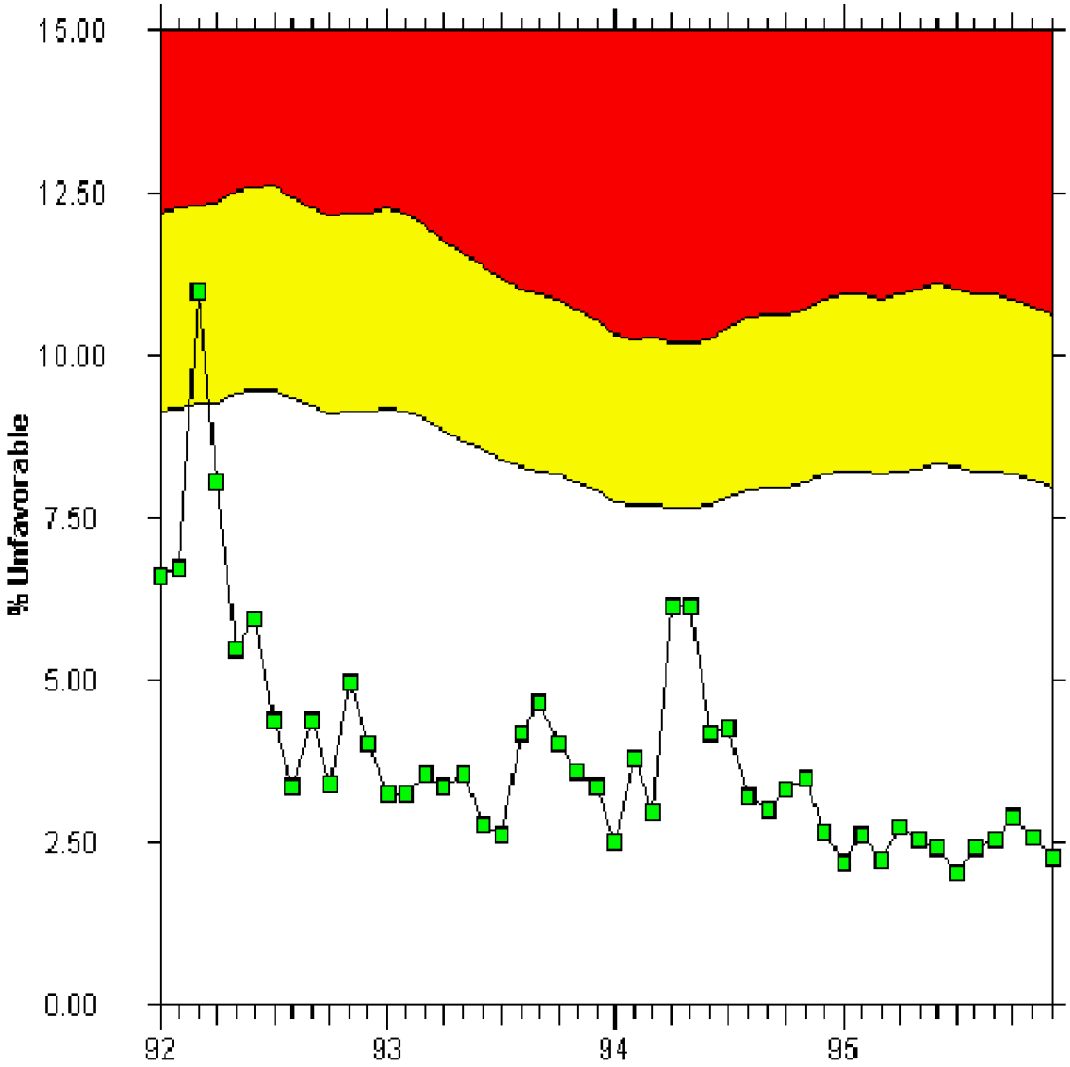
Percentages

USA Today has come out with a new survey – apparently, three out of every four people make up 75% of the population.
(David Letterman)



Q1: What is the sample size?

Comparisons to Averages



Comparisons to Averages

- Then there is the man who drowned crossing a stream with an average depth of six inches. (W.I.E. Gates)

- You know, Shaquille O'Neal and I have an average height of 6 feet. (U.S. Labor Secretary Robert Reich)

Q1: What is the sample size?
Q2: What do the raw data show?

Tabulations (& Lurking Variables)

Sandia National Laboratories

	Airline N	Airline S
Pass	25	15
Fail	35	35
Total	60	50

Pass rate for Airline N = 25/60 = 42%

Pass rate for Airline S = 15/50 = 30%

Summer

	Airline N	Airline S
Pass	20	5
Fail	20	5

Winter

	Airline N	Airline S
Pass	5	10
Fail	15	30

Tabulations (& Lurking Variables)

Sandia National Laboratories

- Statistics have shown that mortality increases perceptibly in the military during wartime. (Robert Boynton)

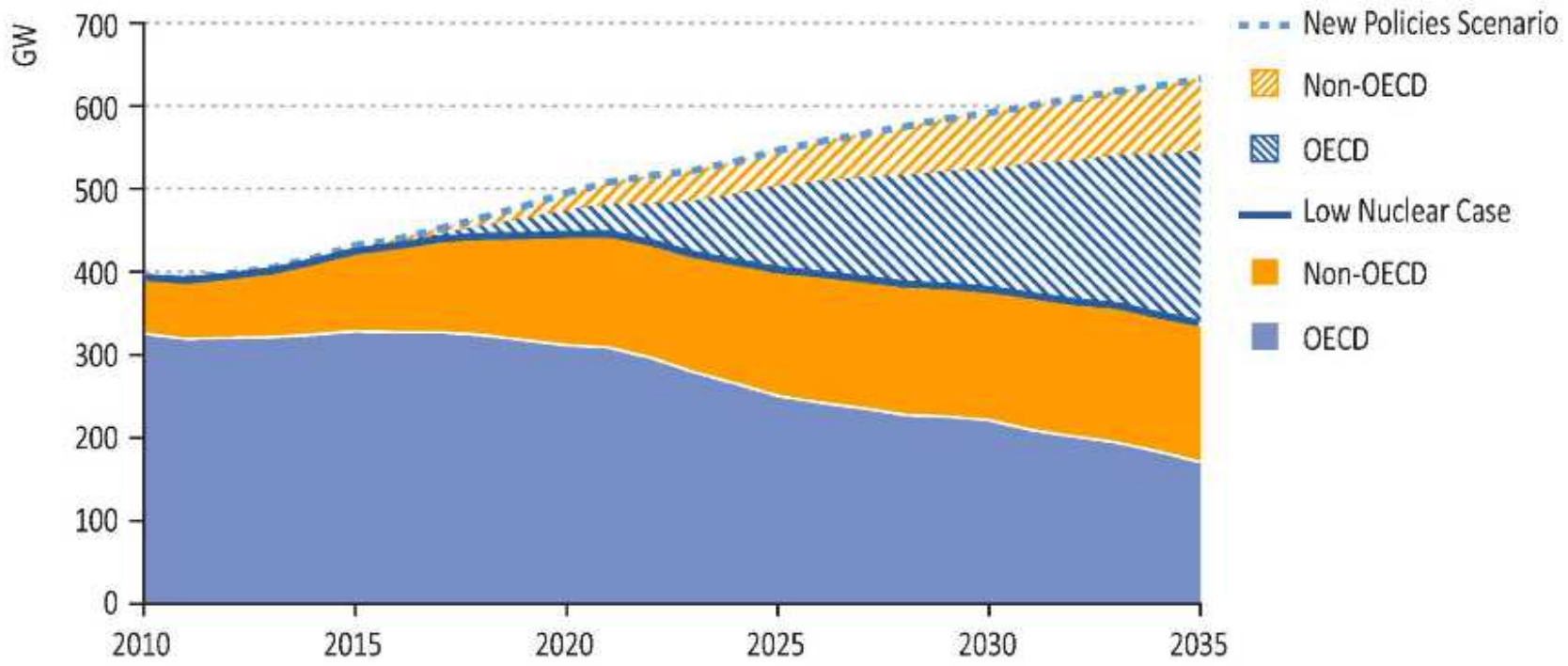
- The statistics on sanity are that one out of every four Americans is suffering from some form of mental illness. Think of your three best friends. If they're okay, then it's you. (Rita Mae Brown)

Q1: What is the sample size?
Q3: What are the important factors?

Predictions/Extrapolations

Sandia National Laboratories

Figure 12.3: Nuclear power capacity in the Low Nuclear Case



Nuclear power capacity drops by 15% between 2010 & 2035 as a result of a bigger wave of retirements outweighing a slower rate of new construction than in the New Policies Scenario

Source: IEA Energy Outlook 2011

Predictions/Extrapolations

Once upon a time, there was a sensible straight line who was hopelessly in love with a dot.
(Norton Juster)

If she loves you more each and every day,
by linear regression she hated you before you met.

Q1: What is the sample size?
Q2: What do the raw data show?
Q3: What are the important factors?

Recommendations

Sandia National Laboratories

- Consider the process creating the data and the variability in the process
  - Plot the raw data (over time if appropriate)
  - Include sample size information
  - Be wary of conclusions drawn from few data points
  - Employ subject matter expertise in understanding data
  - Take care in extrapolating data

- Ask the following questions:
  - **Q1: What is the sample size?**
  - **Q2: What do the raw data show?**
  - **Q3: What are the important factors?**

Scenario

- A series of tests was conducted on a product
- Under some test conditions, the product was damaged
  - A “severity of damage” index was assigned with values ranging from 0 to 12
- You need to decide under what conditions the product should be used

Scenario

	% Tests resulting in damage
Development Tests	0%
Non-Development Tests	40%

Q1: What is the sample size?

	% Tests resulting in damage	Number of tests
Development Tests	0%	4
Non-Development Tests	40%	23

Scenario

	% Tests resulting in damage	Number of tests	Average Damage Index
Development Tests	0%	4	0
Non-Development Tests	40%	23	1.4

Q2: What do the raw data show?

Scenario

	Damage Severity
Dev Test 1	0
Dev Test 2	0
Dev Test 3	0
Dev Test 4	0
Test 1	11
Test 2	4
Test 3	4
Test 4	4
Test 5	2
Test 6	0
Test 7	0
Test 8	0
Test 9	0

	Damage Severity
Test 10	0
Test 11	0
Test 12	0
Test 13	0
Test 14	0
Test 15	0
Test 16	0
Test 17	0
Test 18	0

Q3: What are the important factors?

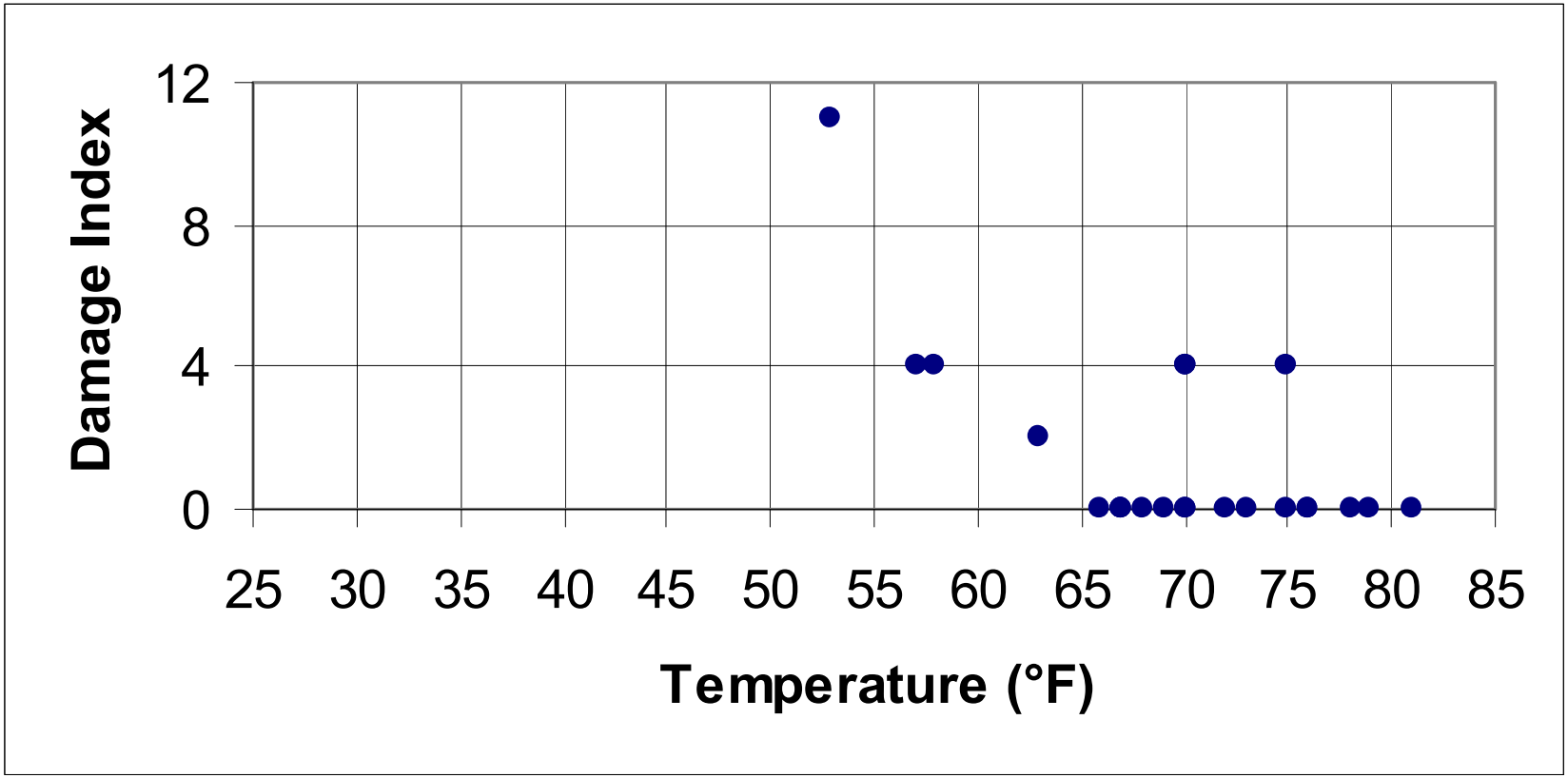
A: Development tests had special configurations

Scenario

	Temp (deg F)	Damage Severity
Test 1	53	11
Test 2	57	4
Test 3	58	4
Test 4	63	2
Test 5	70	4
Test 6	70	4
Test 7	75	4
Test 8	66	0
Test 9	67	0
Test 10	67	0
Test 11	67	0
Test 12	68	0

	Temp (deg F)	Damage Severity
Test 13	69	0
Test 14	70	0
Test 15	70	0
Test 16	72	0
Test 17	73	0
Test 18	75	0
Test 19	76	0
Test 20	76	0
Test 21	78	0
Test 22	79	0
Test 23	81	0

Scenario: sample size, raw date, important factors included



You need to decide under what conditions the product should be used….

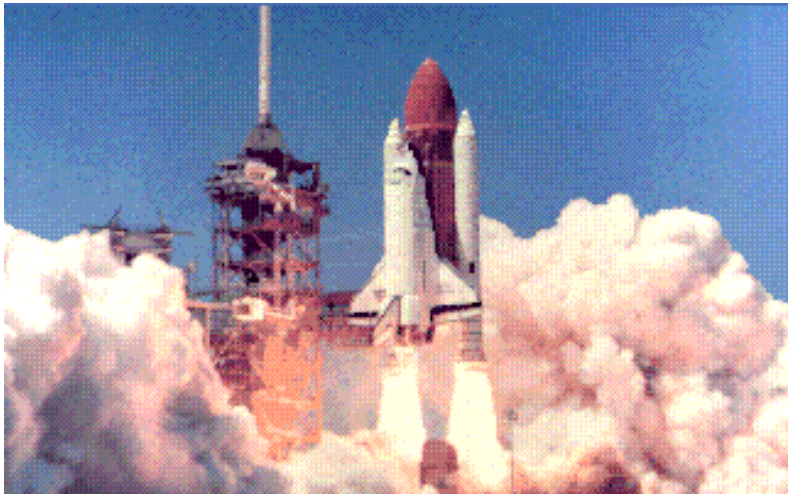
Scenario

- This was essentially the same information that was supplied to NASA by Morton Thiokol the night before the Challenger launch!
  - *Would you have detected a relationship between temperature and (O-ring) product damage?*
  - *Would you have launched the Challenger when the predicted temperature was 26-29 degrees?*

Source: Tufte (1997)

Challenger Accident Testimony

- *It was a matter of assembling that data and looking at it in the proper fashion. Had we done that, the data just jumps off the page at you.*



Summary

Sandia National Laboratories

- Improper use of common statistical methods may contribute to incorrect interpretations of variation, data torturing, and faulty decisions

- Always ask the following questions:
  - Q1: What is the sample size?
  - Q2: What do the raw data show?
  - Q3: What are the important factors?