712CD
75TH MORSS CD Cover Page

Name of Principal Author and all other author(s): Juan P. Perez	
Principal Author's Organization and address: 59 Test and Evaluation Squadron 4414 Tyndall Ave Nellis AFB, NV 89191	Phone: 702-384-2700
	Fax: 702-652-4324
	Email: juanpablo.perez@nellis.af.mil

Original title on 712 A/B: **Resampling Statistics for the F-22A Lot 5 Suitability Analysis**			
PRESENTED IN: 75the MORSS			
WORKING GROUP:	25	**DEMONSTRATION:**	
COMPOSITE GROUP:	1	**POSTER:**	**PPT Presentation**

This presentation is believed to be: *Unclassified, approved for public release, distribution unlimited, and is exempt from U.S. export licensing and other export approvals including the International Traffic in Arms Regulations (22CFR120 et seq.)*

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 JUN 2007	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Resampling Statistics for the F-22A Lot 5 Suitability Analysis	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **59 Test and Evaluation Squadron 4414 Tyndall Ave Nellis AFB, NV 89191**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM202526. Military Operations Research Society Symposium (75th) Held in Annapolis, Maryland on June 12-14, 2007, The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**31**	

59th Test and Evaluation Squadron

Resampling Statistics for the F-22A Lot 5 Suitability Analysis

Mr. Juan P Perez
59 TES/EAA

UNCLASSIFIED

59th TES Operations

- 59th TES in conjunction with 422nd TES
  - Perform Operational Test & Evaluation (OT&E)
  - Tactics Development
  - Fielding Recommendations
- Weapon Systems
  - A-10, F-15C, F-15E, F-16, F-22A, HH-60

Purpose

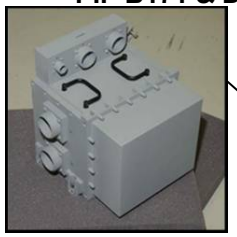
- Explain the problem
- Demonstrate the Resampling Technique
  - Construct the Cumulative Distribution Function (CDF)
  - Build a Confidence Interval
  - Calculate Power/Consumer Risk
- Demonstrate application tool used for the suitability analysis

Overview

- Test Problem
- Background
- Test Objective
- Test Methodology
- Reporting
- Conclusion

Test Problem: Lot 5 System Description

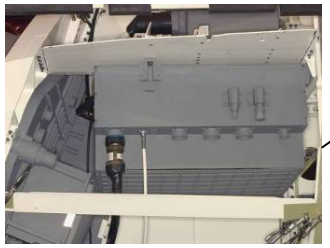
Advanced Radar Program
PIP B171 & DO-03



IPFA

Radar Antenna

New S/W for Lot 5 & Sp 2



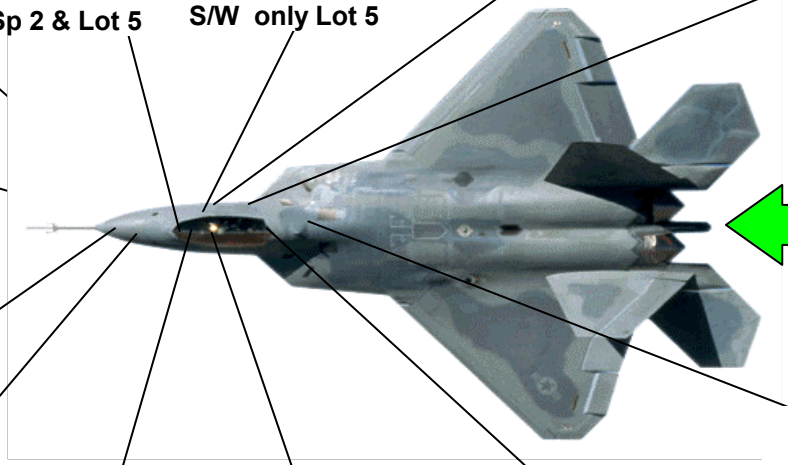
Radar Receiver Exciter

Radar Processor

HUD
DMS & PIP
New S/W for
Sp 2 & Lot 5

Aux Com/ACP
PIP211 (Lot 6 H/W)
S/W only Lot 5

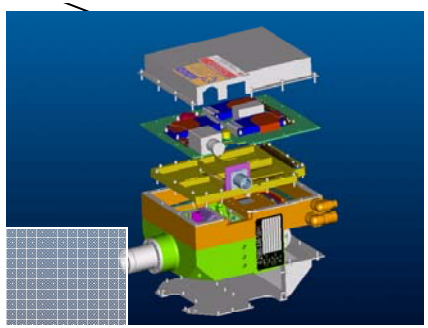
IFDL Controller
PIP212



DEW
PIP209A
New L5 S/W

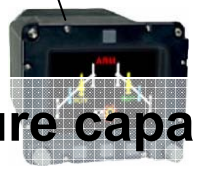
OFP 3.1.3 L5 Functionality Equivalent to Lot 3 Production Aircraft

MLD Sensor
PIP238FGA

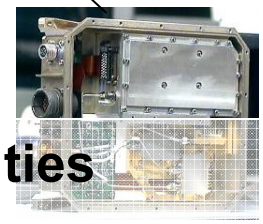


DMVR
DMS

UFD
DMS & PIP
New L5 S/W



IFDL RT
PIP212



- Enhance radar's future capabilities
- New contractor building "same" parts

Test Problem: F-22A Lot 5 Suitability Analysis

- **Test's Purpose: To compare currently fielded aircraft suitability results to Lot 5 aircraft.**
  - Operational Utility Evaluation (OUE): compare effectiveness and suitability between F-22A Lot 5 aircraft (new data set) and currently fielded F-22A aircraft (baseline data set)
  - Suitability goal: evaluate Lot 5 hardware updates only
  - Suitability structure: compare data using Reliability, Maintainability and Availability (RM&A).

Test Problem: F-22A Lot 5 Suitability Analysis

- **Analyst's Challenge: How to compare current suitability results to new Lot 5 suitability measures?**
  - Baseline data (currently fielded aircraft) includes 758 flight hours. Lot 5 test has shorter timeline requirements.
  - The test requires a minimum 70% power / 30% consumer risk in detecting at least a doubling (twice as bad) in any suitability measure.
  - The new test consists of 200 flight hours

Overview

- Test Problem
- Background
- Test Objective
- Test Methodology
- Reporting
- Conclusion

Background: Suitability

- Operational suitability: the degree to which a system can be satisfactorily placed into fielded use
  - reliability, maintainability, availability (RM&A)
- Availability: affected by reliability and maintainability
- Other considerations: compatibility, transportability, wartime usage rates, safety, human factors, manpower supportability, logistics supportability, documentation, training requirements, and natural environmental effects and impacts

Background: Suitability

- Reliability: the probability of a system to perform its function [3]
  - Time-continuous: operate the system until it fails, fix it and continue to operate. Process repeats until enough information is collected
  - Success/Fail: test the system and record successes and failures
- Maintainability: the ability of an item to be retained in, or restored to, a specified conditions when maintenance is performed (AFI 10-602)
  - Average time between maintenance activity
  - Cumulative maintenance time divided by flying hours
- Availability: the probability that a system will be in an operable state at a random point in time [3]
  - Operational time divided by the total time

Background: Suitability Analysis

- **Reliability, Maintainability and Availability Evaluation**
  - **New Lot 5 hardware data collection only**
  - **Test scoring data Measures of Performance (MOP):**
    - ✓ Break rate
    - ✓ Mean time between critical failure (MTBCF)
    - ✓ Could not duplicate rate
    - ✓ Maintenance man hours per flying hour
    - ✓ 2/4/8 hour fix rate
    - ✓ Abort rate
    - ✓ Mean time between maintenance
    - ✓ Weapons system reliability
    - ✓ Integrated diagnostics accuracy
    - ✓ Mean down time
    - ✓ Mean repair time

$$\text{MTBCF} = \frac{\text{Total Flying Hours}}{\text{Number of Critical Failures}}$$

Background: Resampling

What is Resampling?

- Mechanism used to produce a hypothetical distribution by randomly taking samples from an observed distribution or baseline distribution
- Essentially: Monte Carlo simulation of statistical results
- Basic Rules:

1. Specify the universe to sample from
   - ✓ observed data set or baseline data set
2. Specify the sampling procedure
   - ✓ number of samples
   - ✓ sizes of samples
   - ✓ sampling with or without replacement
3. Specify the statistic you wish to keep track of
   - ✓ ratio, mean, variance, etc.

Background: Resampling

- Advantages
  - Simple to use and teach
  - Avoids using the wrong method
  - Knowledge of distribution not needed
  - Free of mathematical formulas and restrictive assumptions

- Disadvantages
  - Requires empirical data
  - Sample size can be a problem if the baseline data is limited

Overview

- Test
- Background
- Test Objective
- Test Methodology
- Reporting
- Conclusion

Test Objective: F-22A Lot 5 Suitability Analysis

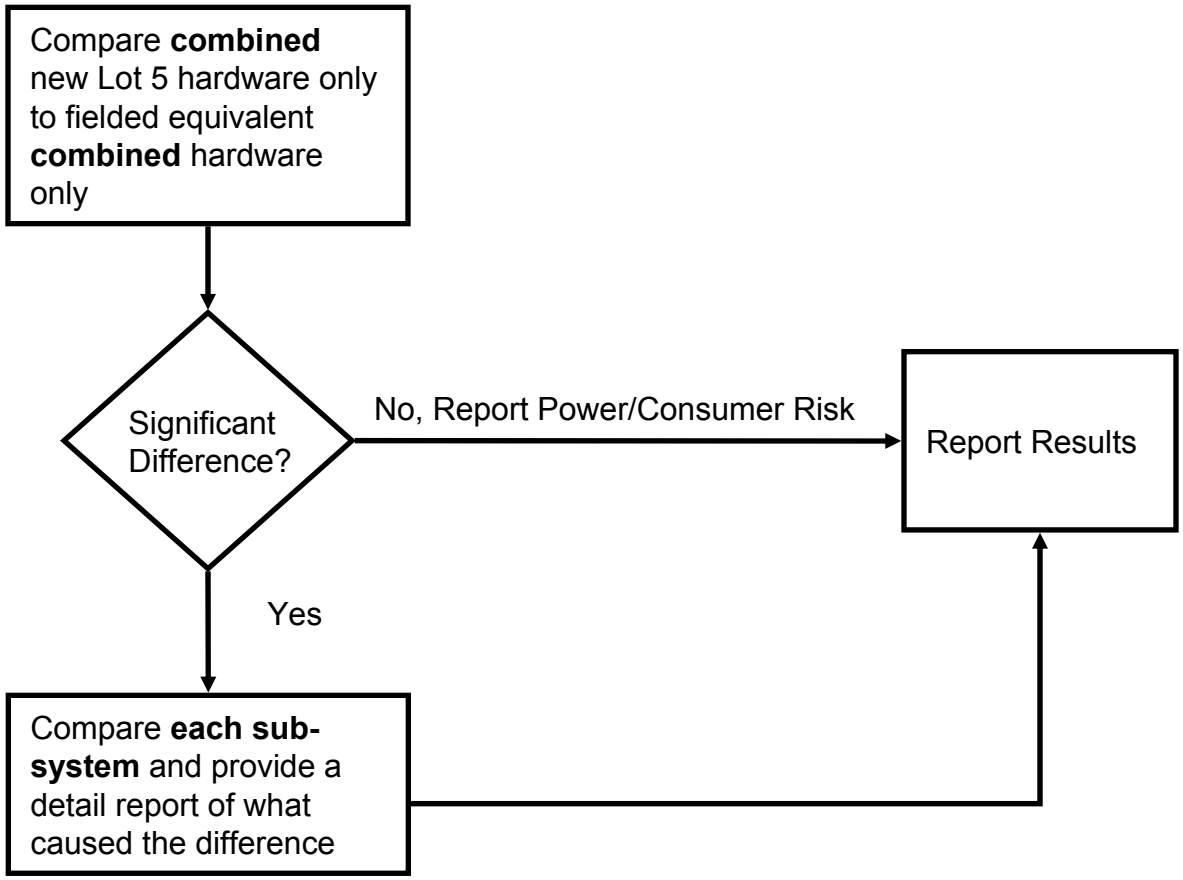
- Assess operational effectiveness of new Lot 5 hardware
- Compare Lot 5 hardware and equivalent hardware for currently fielded aircraft suitability
  - Report results outside a 90% confidence bound as significantly different
  - If not significantly different, report power/consumer risk
  - The formal hypothesis for each MOP for this approach is as follows:

$$H_o : \text{Lot 5 MOP} \geq \text{FDE MOP}$$

$$H_1 : \text{Lot 5 MOP} < \text{FDE MOP} \ (\alpha = .1)$$

** Note: hypothesis assumes that a larger value is better

Test Objective: Algorithm



Repeat flow for each suitability MOP

Overview

- Test
- Background
- Test Objective
- Test Methodology
- Reporting
- Conclusion

Test Methodology: Example

Mean Time Between Critical Failure (MTBCF)

$$\text{MTBCF} = \frac{\text{Total Flight Hours}}{\text{Number of Critical Failures}}$$

- The suitability study for currently fielded aircraft recorded 18 critical failures (CF) during 758 hours of operation for a MTBCF of 42.11
- Create a representative population using 758 binary values
  - Assume 1 sample per flying hour
  - Assign a '1' to CFs (18 occurrences)
  - Assign a '0' to no CFs (740 occurrences)
- Use resampling to produce a 200 hour representative (hypothetical) population
- Construct a CDF of the 200 hour MTBCF
- Use the CDF of the 200 hour data to find the lower 90% confidence bound

Test Methodology: Test Model

- Cumulative Distribution Function of the [illegible]
- [illegible]
- [illegible]
- [illegible]
- Now we have a hypothetical population that represents the observed population

Compare the new set of performance data (Lot 5) to the established baseline (currently fielded aircraft) with a 90% lower confidence bound

ample from

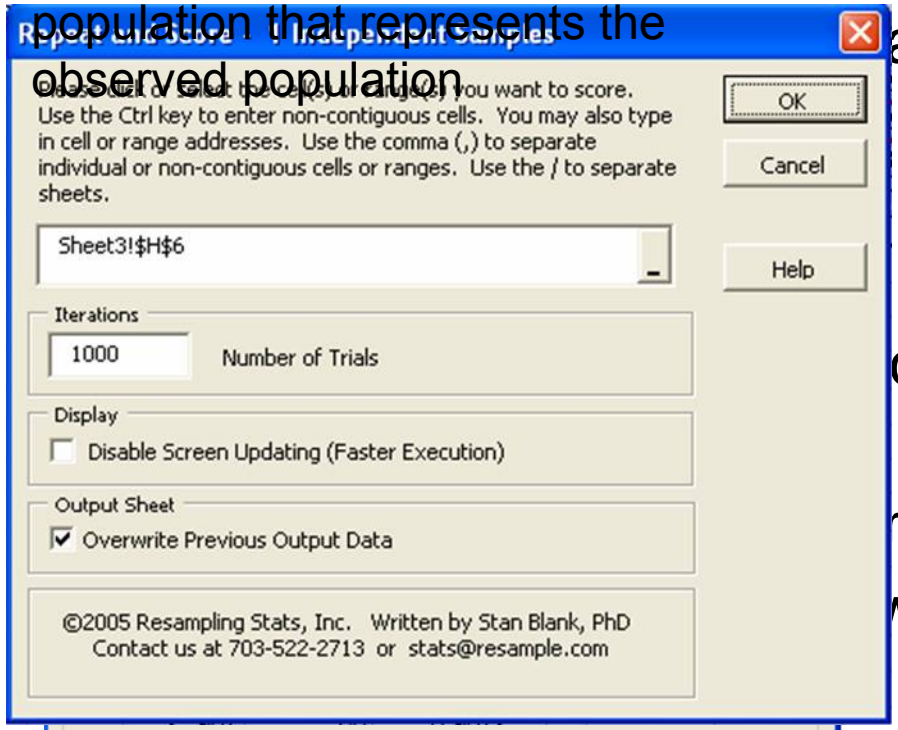
FH)

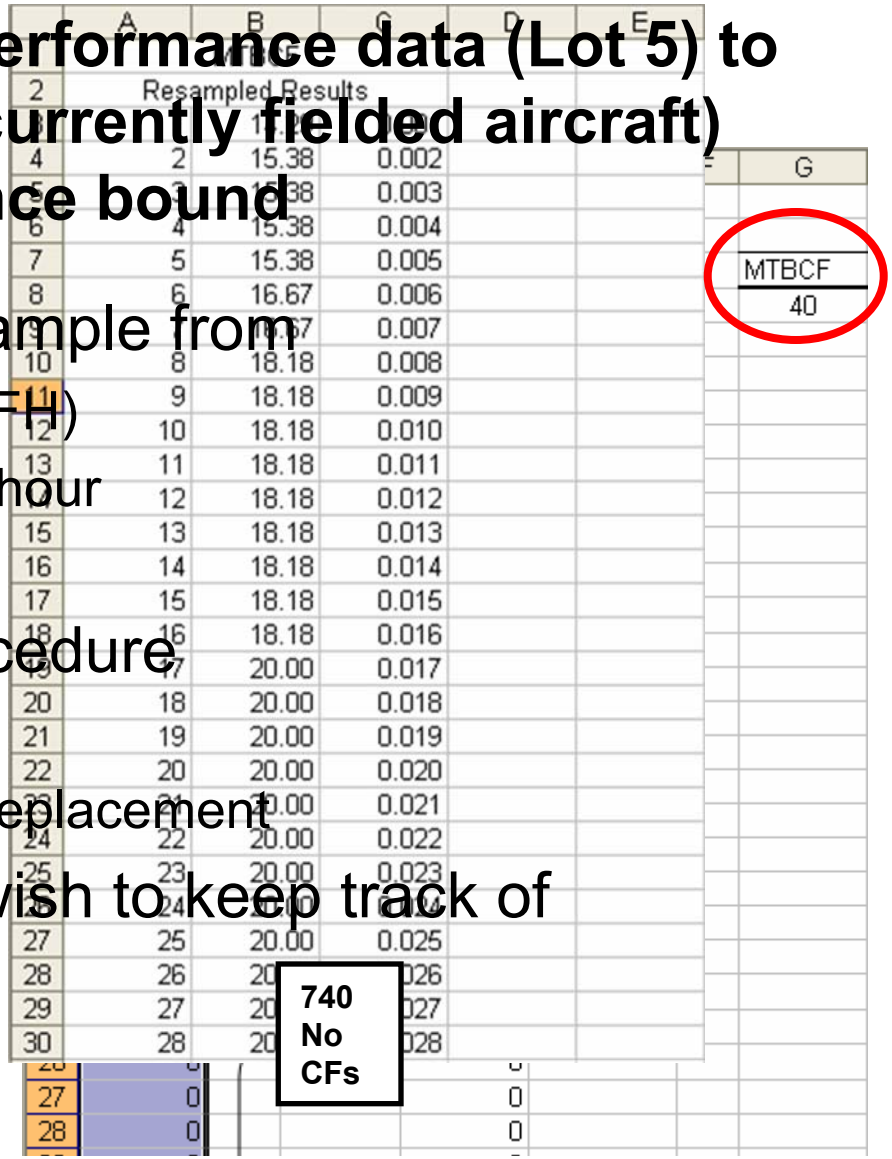
hour

cedure

replacement

wish to keep track of





Test Methodology: Confidence Interval

Compare **combined** new Lot 5 hardware only to fielded equivalent **combined** hardware only

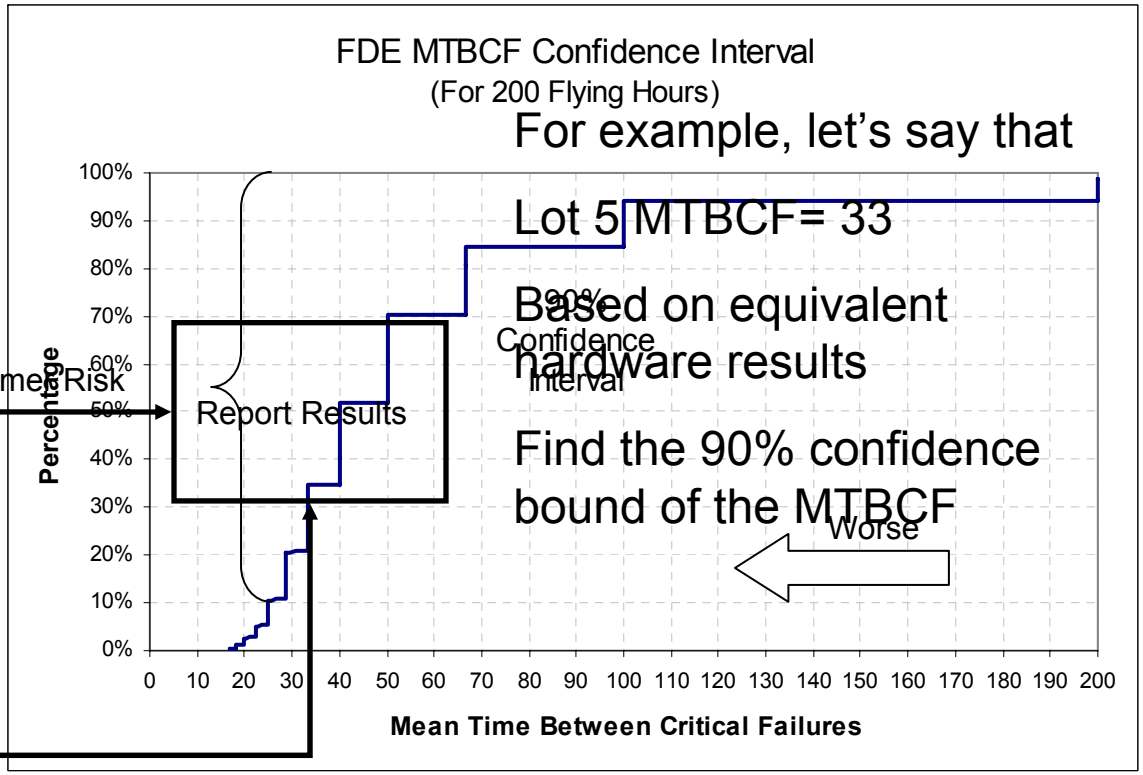
Significant Difference?

No, Report Power/Consumer Risk

Report Results

Yes

Compare **each sub-system** and provide a detail report of what caused the difference



The 90% confidence bound is 25

33 is in the 90% of the data, therefore there is no significant difference between Lot 5 and the currently fielded hardware

Test Methodology: MTBCF Power

- CDF constructed using resampling statistics
- Use resampled results to calculate power
- Power indicates the confidence level the test design provides to detect a particular level of increase in the MTBCF based on possible results

	Null hypothesis is True	Null hypothesis is false
Reject the null hypothesis	**Type I error (rejecting a true null hypothesis) α**	**Correct decision**
Fail to reject the null hypothesis	**Correct decision**	**Type II error (failing to reject a false null hypothesis) β**

Test Methodology: Power (Classical Method)

- Classical method: theory based
- Resampling method: observation based
- An example to demonstrate the classical method
  - Measuring successes and failures, therefore use the binomial distribution
  - Probability Mass Function: PMF

The binomdist excel function provides the CDF/PMF of the binomial distribution

Function → BINOMDIST(# success, trials, probability of success, cumulative)

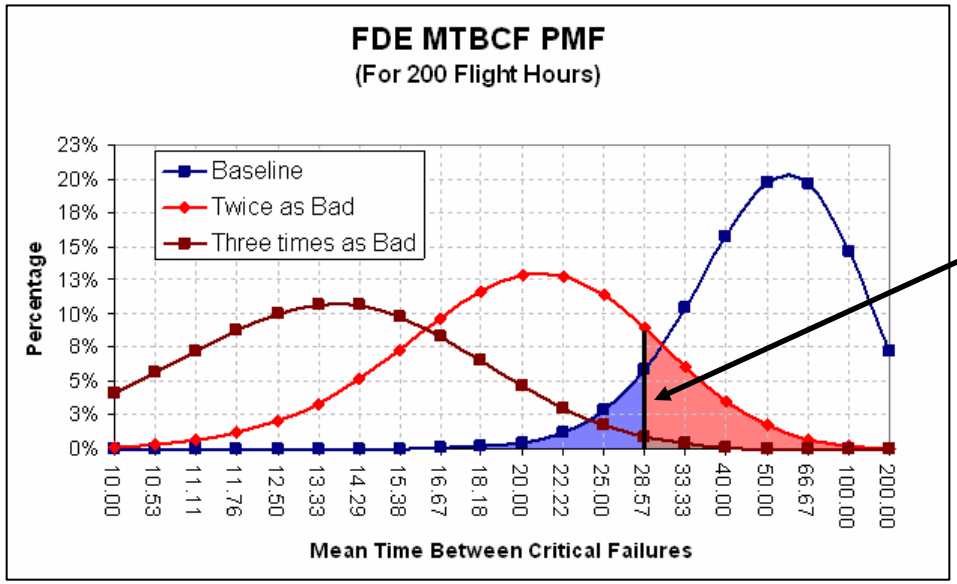
Number of success → (Trials-CF)

Trials → Total Flight Hours

Probability → (Trials-CF)/Trials

Cumulative → TRUE provides the CDF; FALSE provides the PMF

Test Methodology: Power (Classical Method)



90% confidence bound

- The PMF (blue) was generated based on the known baseline(758FH with 18 CFs)
- The red PMF – Distribution is twice as bad as the main baseline
- The blue PMF – Distribution is three times a bad as the main baseline
- Any results outside the 90% (lower than 28.57 MTBCF) will be considered significantly different.

Test Methodology: MTBCF Power (Resampling)

	A	B	C	D	E	F	G	H
1	CF	5	8	**10**	12	15		
2	MTBCF	40	25	**20**	16.67	13.33		
3								
4	Unknown Distribution							
5	200 FH / 10 CFs				200 Hrs Resampled			MTBCF
6	1				0			25.00
7	1				0			
8	1				0			
9	1				0			
10	1				0			
11	1				0			
12	1				0			
13	1				0			
14	1				0			
15	1				0			
16	0				0			
17	0				0			
18	0				1			
19	0				0			
20	0				0			
21	0				0			
22	0				0			
23	0				0			
24	0				0			
25	0				0			
26	0				0			
27	0				0			
28	0				0			
29	0				0			

10 CFs

190 Non CFs

	J	K	L
	Resampled Results		
1	10.00	0.001	
2	10.00	0.002	
3	10.53	0.003	
4	10.53	0.004	
5	11.11	0.005	
6	11.11	0.006	
7	11.11	0.007	
⋮	⋮	⋮	
774	25	0.774	
775	25	0.775	
776	28.57143	0.776	←
777	28.57143	0.777	
⋮	⋮	⋮	
863	28.57143	0.863	
864	33.33333	0.864	←
865	33.33333	0.865	
⋮	⋮	⋮	
933	33.33333	0.933	
934	40	0.934	←
935	40	0.935	
⋮	⋮	⋮	
986	50	0.986	
987	50	0.987	
988	66.66667	0.988	←
⋮	⋮	⋮	
997	66.66667	0.997	
998	100	0.998	←

MTBCF Power for 200 hrs

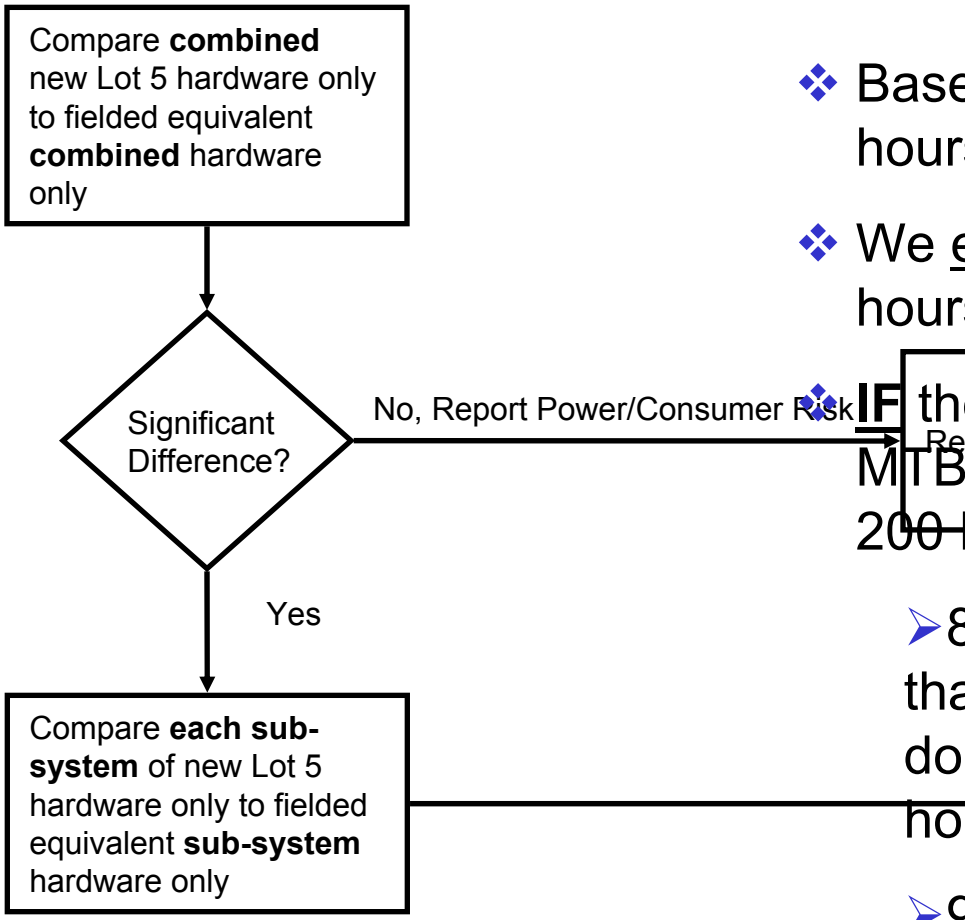
MTBCF Results	Actual (Unknown) MTBCF 25	20	16.7	13.33
100	98.7%	99.7%	99.9%	99.9%
67	96.7%	98.7%	99.9%	99.9%
40	80.8%	93.3%	98.2%	99.8%
33	66.7%	86.3%	95.9%	99.1%
28.57	40.2%	77.5%	84.8%	97.3%

Sheet2 | **Sheet3** | Sheet4 | Sheet5 | Sheet6 | Sheet7 | MTBM | MTBR | BR | Results

Overview

- Test
- Background
- Test Objective
- Test Methodology
- Reporting
- Conclusion

Report Results



- Baseline MTBCF of 42.11 FH (758 hours over **18 critical failures**)
- We expect **5 critical failures** over 200 hours
- **IF** the new Lot 5 hardware provided a MTBCF of 33 FH (**6 critical failures** 200 hours)
  - 86% power (14% consumer risk) that the true Lot 5 MTBCF did not double (**10 critical failures** over 200 hours)
  - 99% power (1% consumer risk) that the true Lot 5 MTBCF did not triple (**15 critical failures** over 200 hours)

Overview

- Test
- Background
- Test Design
- Test Methodology
- Reporting
- Conclusion

Conclusion

- Resampling Statistics makes possible:
  - Construction of CDFs
  - Building Confidence Intervals
  - Calculating Power/Consumer Risk
- Resampling Statistics with F-22A Lot 5 suitability:
  - Facilitates comparison between large baseline sample size and small test sample size
  - Provides method to compute power
  - Evaluates data without knowledge of the distribution
  - Eliminates mathematical formulas and restrictive assumptions

References

1. Simon, J. L. (1997): Resampling: The New Statistics
2. DOD Directive 5000.3
3. Sushinski, T.C. (1989): Operational Suitability Guide, 57th Fighter Weapons Wing
4. Simon, J.L.: The Philosophy and Practice of Resampling Statistics
5. Resampling Stats Add-in for excel: http://www.resample.com/

Questions?