War-Winning Capabilities … On Time, On Cost

Statistical Defense Course #1

May 2011

Arnon Hurwitz & Todd Remund
812 TSS/EN
Edwards AFB, CA 93524
arnon.hurwitz@edwards.af.mil

Approved for public release; distribution is unlimited.
AFFTC-PA No.: PA-10900

Statistical Defense!

Picture Courtesy of the U.S. Air Force Official Web Site
Screaming Eagle created by Ken Chandler

➢Introduction

- Observational Studies and Experimental Design (DOE)
- Statistical Modeling
- Bayesian Techniques

Introduction

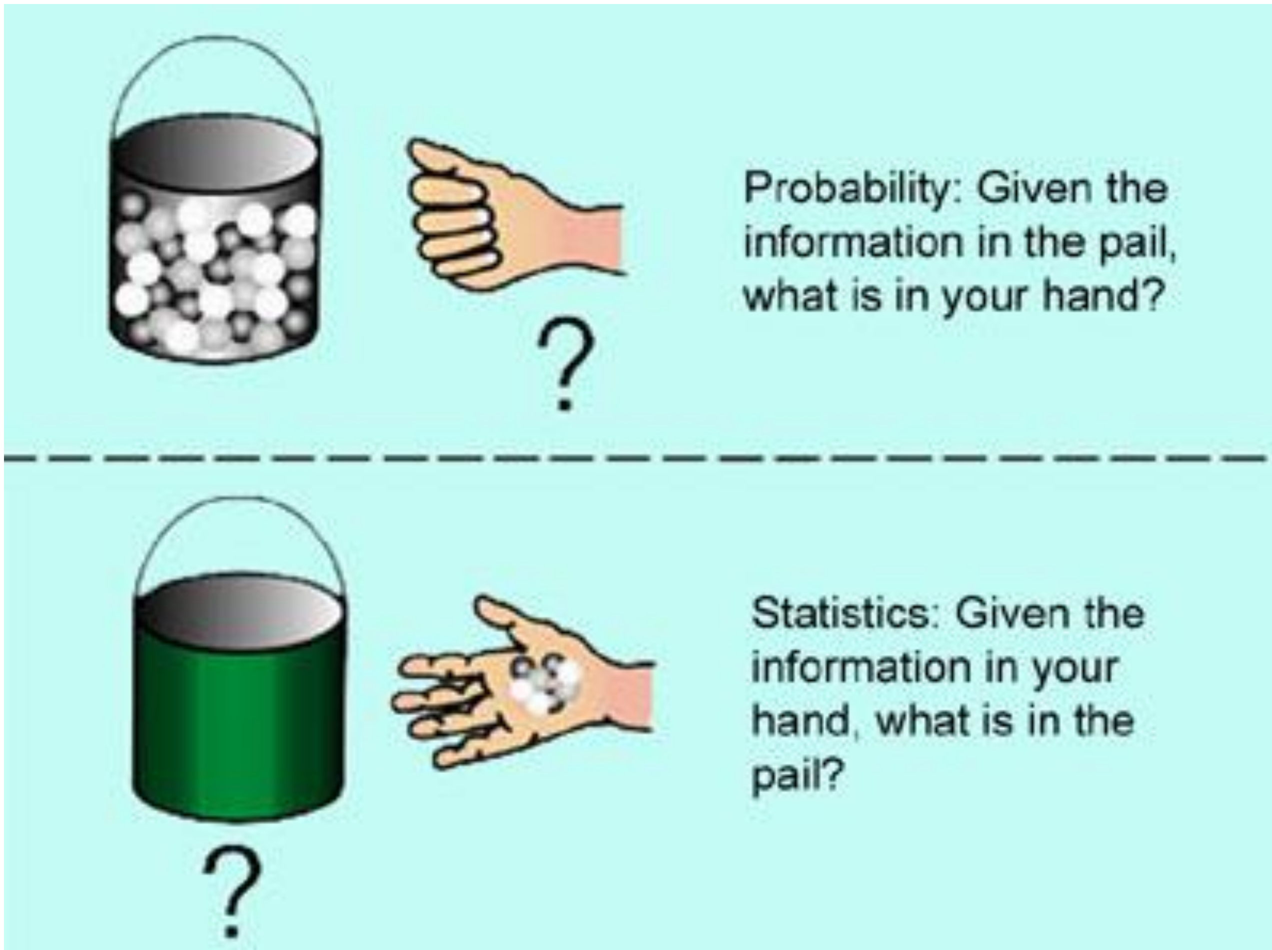
- Statistical Inference: what is it, what is "scope" of inference?
- Drawing Statistical Conclusions: what is inference, what is the "scope" of inference?
- Basic Tenets of Inferential Statistics: hypothesis tests, confidence intervals.
- Type I and Type II Errors: Confidence and power in a test.
- Difference between an observational study and an experiment.
- Central Limit Theorem, t-tests.

AFFTC

Statistical Inference

What is it?

Statistical Inference



Statistical Inference Definition

AFFTC

- An *inference* is a conclusion that patterns in the data are present in some broader context.

- A *statistical inference* is an inference justified by a probability model linking the data to the broader context.

Motivation

- **In many studies, we wish to compare two groups, each with a different ‘treatment,’ and answer the question: “Does one group differ from the other ?”**
- **We would also like to extend the inference based on our discoveries so as say something about the <u>population as a whole</u>.**
- **To test such a question we usually compare the average taken from one group with that of another group.**
- **IDEALLY, WE SHOULD BE ABLE TO DRAW A <u>RANDOM SAMPLE</u> FROM THE <u>ENTIRE POPULATION</u> OF INTEREST, THEN <u>RANDOMLY ASSIGN TREATMENTS</u> TO SUBSETS OF THAT SAMPLE. REALITY IS SELDOM THAT EASY.**
- **<u>ANY</u> RESTRICTION ON RANDOMIZATION DIRECTLY AFFECTS OUR SCOPE OF INFERENCE.**

Example: Gas Mileage
Does Gas Brand Make a Difference?

(Picture courtesy of Ron Proesl – His Harley Sportster)

The Experiment

- We want to compare the gas mileage between two popular brands of gasoline used in 2009 Harley Davidson Sportsters to determine if there is a difference in the two brands.

- Out of a large available group of Sportsters we drew a random sample of 34 “Sportsters,” and then randomly assigned two brands of gasoline to subsets of the samples.

The Hypothesis

- **Our hypothesis is that both brands of gasoline will yield the same gas mileage.**
- **The alternative hypothesis is that they will be different.**

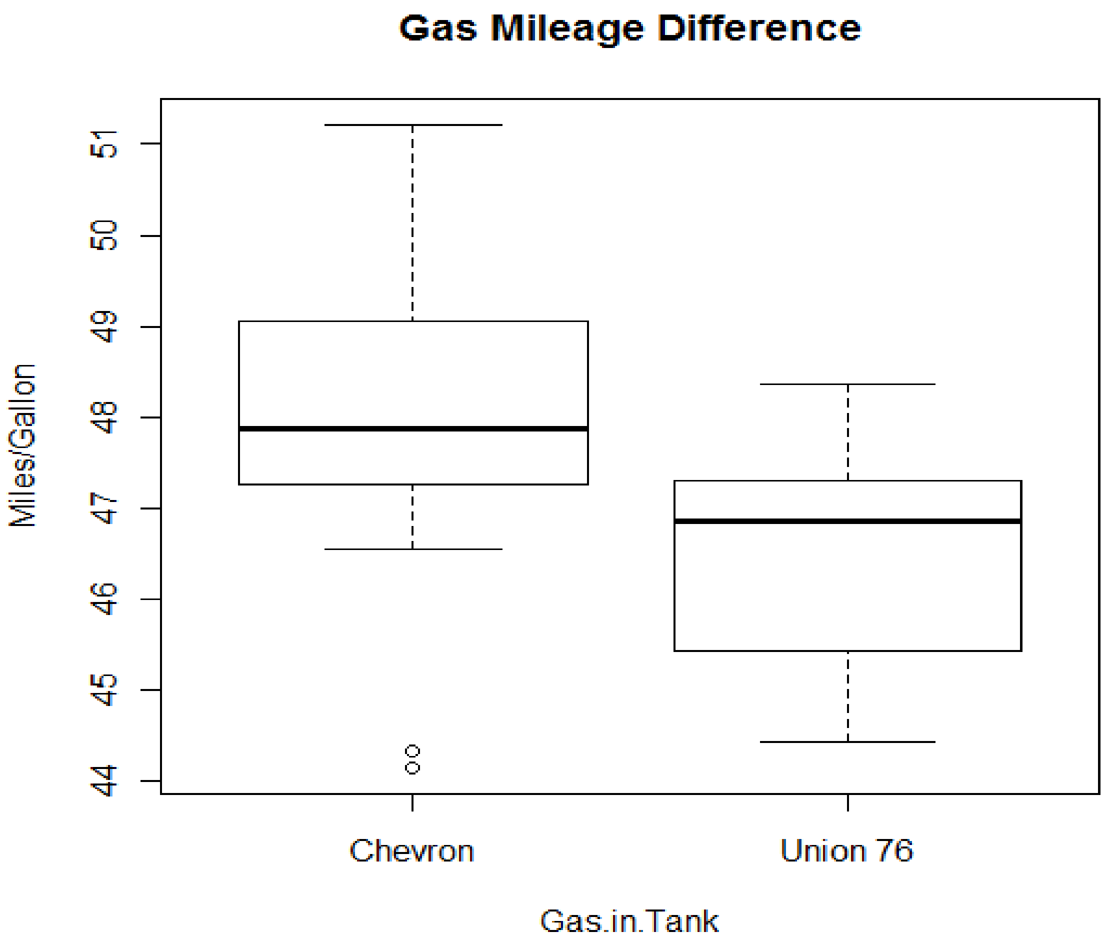
$$H_0: \mu_{\text{Union 76}} = \mu_{\text{Chevron}}$$

$$H_1: \mu_{\text{Union 76}} \neq \mu_{\text{Chevron}}$$

Data

Mileage	Gas in Tank	Mileage	Gas in Tank
45.913	Union 76	44.154	Chevron
45.576	Union 76	44.333	Chevron
44.433	Union 76	47.883	Chevron
46.608	Union 76	48.682	Chevron
47.138	Union 76	51.210	Chevron
48.357	Union 76	47.561	Chevron
47.302	Union 76	48.524	Chevron
45.389	Union 76	47.258	Chevron
44.725	Union 76	49.275	Chevron
47.048	Union 76	47.547	Chevron
47.879	Union 76	47.849	Chevron
46.854	Union 76	50.000	Chevron
45.439	Union 76	47.187	Chevron
45.365	Union 76	50.960	Chevron
48.334	Union 76	49.060	Chevron
47.286	Union 76	46.542	Chevron
48.019	Union 76	48.729	Chevron

Is this Apparent Difference Caused by Chance?



T-Test Results

T-Test: Two-Sample Assuming Unequal Variances

Mileage	*Union 76*	*Chevron*
Mean	46.56853266	48.0443621
Variance	1.582167504	3.69178852
Observations	17	17
Hypothesized Mean Difference	0	
df	28	
t Stat	-2.649673597	
P(T<=t) one-tail	0.006548866	
t Critical one-tail	1.701130908	
P(T<=t) two-tail	0.013097733	
t Critical two-tail	2.048407115	

Confidence Intervals

Union 76 Mileage	
Mean	46.56853266
Standard Error	0.305071593
Median	46.8537415
Mode	#N/A
Standard Deviation	1.2578424
Sample Variance	1.582167504
Kurtosis	1.212445586
	-
Skewness	0.159400967
Range	3.923817821
Minimum	44.43298969
Maximum	48.35680751
Sum	791.6650551
Count	17
Confidence Level(95.0%)	0.646722882

Chevron Mileage	
Mean	48.0443621
Standard Error	0.466008616
Median	47.88349515
Mode	#N/A
Standard Deviation	1.921402748
Sample Variance	3.69178852
Kurtosis	0.460084378
	-
Skewness	0.471202755
Range	7.056446961
Minimum	44.15390362
Maximum	51.21035058
Sum	816.7541556
Count	17
Confidence Level(95.0%)	0.987894129

$$45.922 \leq \mu \leq 47.215$$

$$47.056 \leq \mu \leq 49.032$$

Conclusion

- It appears that there is a significant statistical difference in our gas mileage results.
- Chevron gasoline seems to improve mileage by about 3% for 2009 Sportsters.

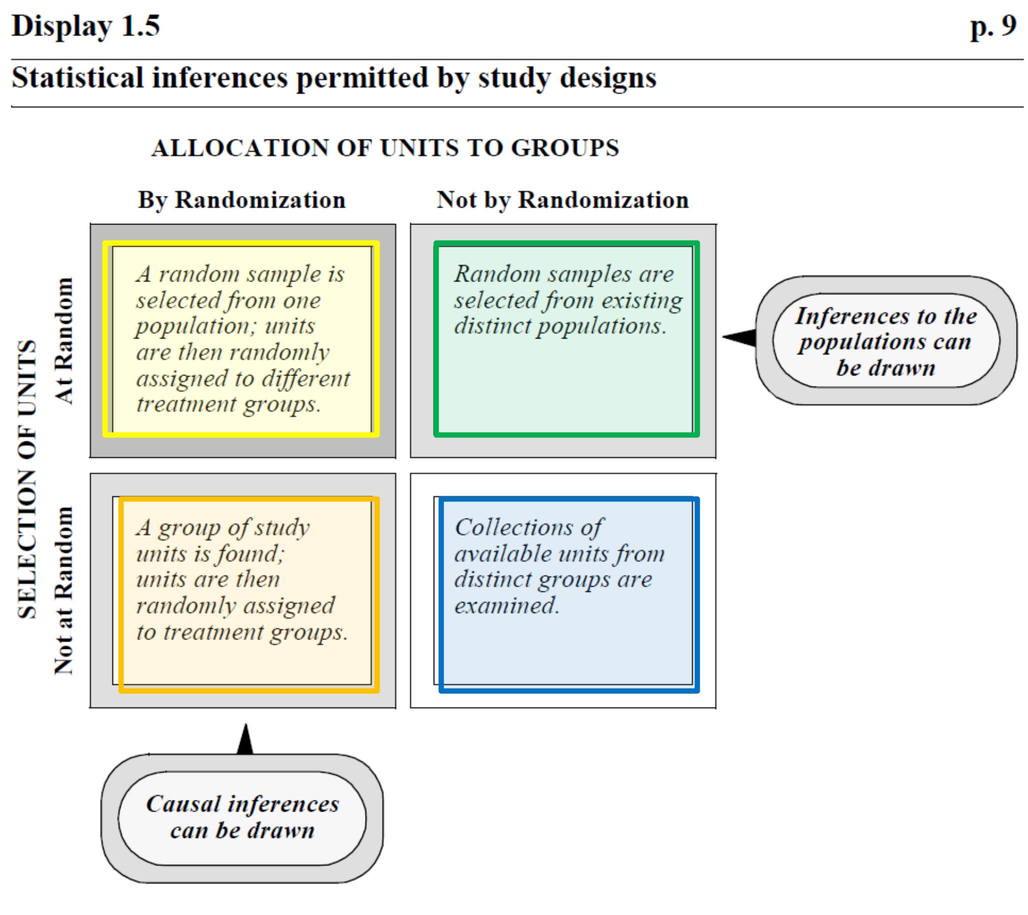
Scope of Inference

- **What can you say based on the data you gathered?**

Scope of Inference

- **We want two things (optimal states):**
  - **Establish Causality**
  - **Generalize Results to the Population of Interest**
  - **Sometimes we need to settle for less…**
    - **The situation will not allow us to gather the data in the optimal fashion.**

Scope of Inference



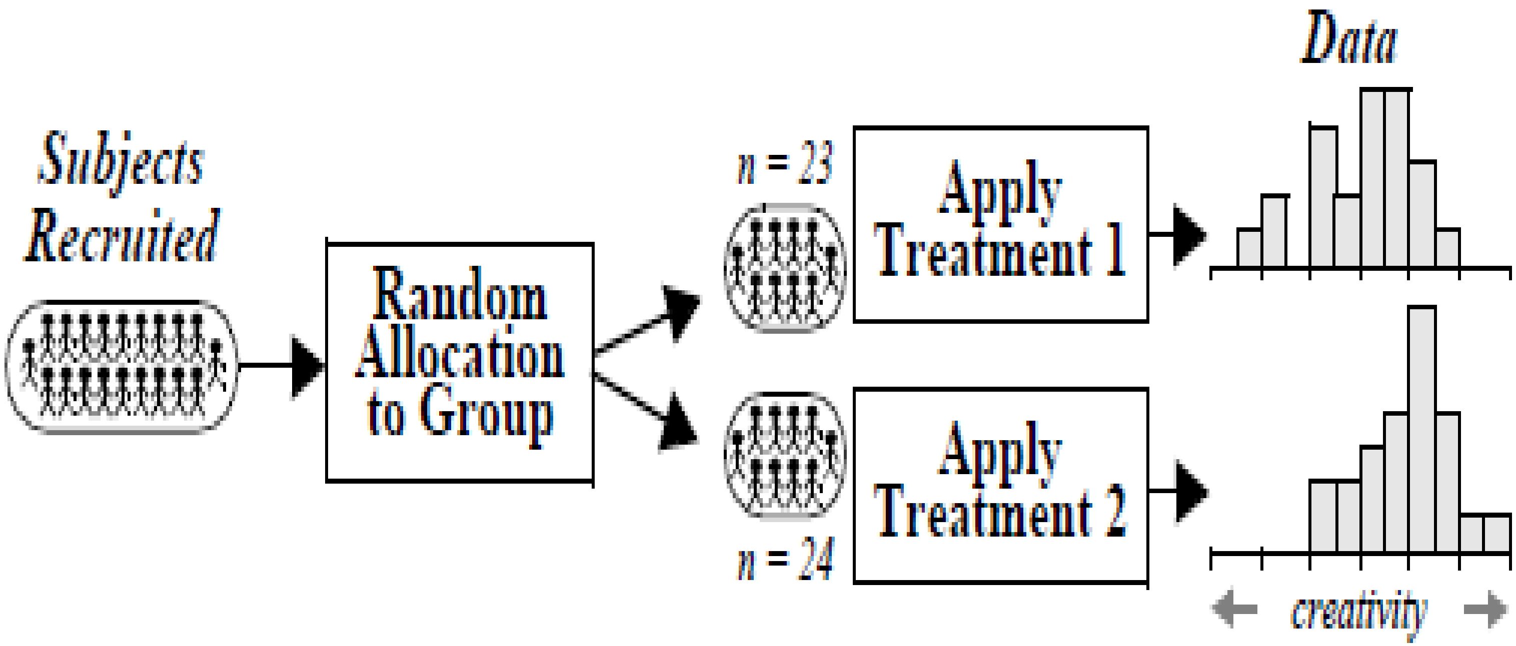
Ramsey, F., Schaffer, D., *The Statistical Sleuth*, Brooks-Cole, Belmont, California, page 9.

Scope of Inference

- The scope of inference from this test only applies to the large available group of 2009 Harley Davidson Sportsters.

- Normally testers would tend to try to expand the scope to all Sportsters but a careful Statistics Consultant would object.

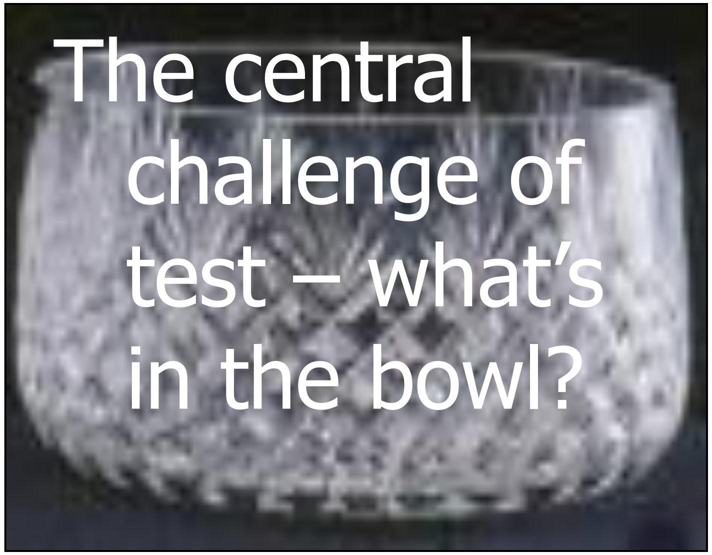
Illustration of a randomized experiment with two treatment groups



Ramsey, F., Schaffer, D., *The Statistical Sleuth*, Brooks-Cole, Belmont California, page 15.

The Central Test Challenge …

- **In all our testing – we reach into the bowl (reality) and draw a sample of operational performance.**



Basic Tenets of Inferential Statistics

- Hypothesis tests
- Confidence intervals

Hypothesis Testing Overview

- **Fundamentally, a *hypothesis* is a statement—the decision making procedure to prove or disprove the statement is called a *hypothesis test.***
- **A *statistical hypothesis* is a statement about the parameters of one or more populations.**
- **A *hypothesis test* relies on information from random samples from a population of interest.**
  - **If the information is consistent with the hypothesis then we will conclude it is true.**
  - **If the information is inconsistent then we will conclude it is probably false.**

The Null and the Alternative

Null Hypothesis: H_0

- Must contain condition of equality: =, $\geq$, or $\leq$
- Test the Null Hypothesis directly
- Reject H_0 or fail to reject H_0

Alternative Hypothesis: H_1

- Must be true if H_0 is false: $\neq$, <, >
- 'opposite' of Null or the "Alternative"

Always consider which is the most serious error in stating your claims. Design the test to guard against it. More to come.

Hypothesis Test: General Procedure

- Identify the parameter of interest.
- State the null hypothesis, $H_{0,}$ e.g. No difference in gas mileage using two brands of gasoline.
- State the alternative hypothesis, H_a , e.g. There is a difference in gas mileage.
- Choose a significance level, α.
- Determine the appropriate test statistic. For our gas mileage example we used the t-test.
- Compute sample quantities for the test statistic (s, y, etc.).
- Compute the p-value with software and data.
- Decide if H_0 should be rejected.

Hypothesis Test Picture

- A court of law is a type of hypothesis testing agency
- There are risks associated with the verdict, ***either*** conclusion has some probability of error
- Hypothesis test:

H_0: Defendant is Innocent

H_1: Defendant is Guilty

		True State of Nature	
		H_0 actually True	H_1 Actually True
Decision: Conclusion We Draw	H_0	Conclusion is Correct ($1-\alpha$) Confidence	Conclusion results in a Type II error (β)
	H_1	Conclusion results in a Type I error (α)	Conclusion is Correct ($1-\beta$) Power

Type I or II Error Occurs if Conclusion **Not** Correct

Central Limit Theorem

T-Tests

Central Limit Theorem: Means of Subsets are Normal!

This is both the most important and least understood result in statistics.

Given:

1. Random variable y distributed *arbitrarily* with population mean μ and standard deviation *s*.
2. Samples of size *n* are randomly drawn (*independently*!) from this population.

Then:

1. The distribution of sample means y will, as the sample size (n) increases, approach normal.
2. The mean of the sample means will approach the population mean μ.
3. The standard deviation of the sample means $\sigma_{\bar{y}}$ is called the ***standard error of the mean***

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem (continued)

- **Practical Rules of Thumb**
  - For samples of size $n > 30$, the distribution of the sample means can be approximated well by a normal distribution. The approximation gets better as the sample size n becomes larger.
  - If the original population is itself normally distributed, then the sample means will be normally distributed for *any* sample size n.
  - Even moderate departures from normality will prove to be inconsequential in subsequent analyses (stay tuned).

Example

- **A resistor is to be used in a satellite guidance package.**
- **The manufacturer reports that the mean resistance is 10 ohms and the standard deviation is 1 ohm.**
- **If we take a sample of 30, what *standard error of the mean* may be reported?**

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{30}} = 0.183 \text{ ohms}$$

- **We should develop this concept further and establish confidence intervals so we can report the mean +/- error.**

Difference Between Observational Studies and Experiments

- In a *randomized experiment*, the investigator controls the assignment of experimental units to groups and uses a chance mechanism to make the assignment.
- In an *observational study*, the group status of the subjects are established beyond the control of the investigator.
- Note that statistical inferences of cause-and-effect relationships can be drawn from *randomized experiments*, but not from *observational studies*.

SUMMARY

- Motivation
- Example: Two kinds of gasoline for motorcycle
- What is your question? The *Hypothesis*
- α-risk of a test: ‘Type I’ Error
- Inference, Causality, and Relevance
- Box-Plots

War-Winning Capabilities … On Time, On Cost

Statistical Defense Course #2

May 2011

Arnon Hurwitz & Todd Remund
812 TSS/EN
Edwards AFB, CA 93524
arnon.hurwitz@edwards.af.mil

Statistical Defense!

Picture Courtesy of the U.S. Air Force Official Web Site
Screaming Eagle created by Ken Chandler

Defensible Statistics

- Introduction
- ➢ Observational Studies and Experimental Design (DOE)
- Statistical Modeling
- Bayesian Techniques

Design OF Experiments (DOE)

Arnon Hurwitz, PhD. Statistical Methods Group, Edwards AFB, CA
arnon.hurwitz@edwards.af.mil

A DESIGNED EXPERIMENT : Purposefully change input levels

3 TYPES OF 'EXPERIMENTS': DISCOVERY, HYPOTHESIS, OBSERVATION

STATISTICS FOR ENGINEERS – WHEN SHOULD WE DO D.O.E. ?

EXPERIMENTAL STRATEGIES–OFAT, FULL, & FRACTIONAL FACTORIALS

EXAMPLE: AUTOMATIC TARGET RECOGNITION

EXAMPLE: VANE-CLEANING

MIXTURE AND OTHER DESIGNS

EXAMPLE: TIME TO REMOVE AEROSPACE COATINGS

POWER – TYPE I & II ERRORS – SOME USEFUL SOFTWARE

MANY STATS TESTS: A GUIDE FOR THE PERPLEXED (TABLE)

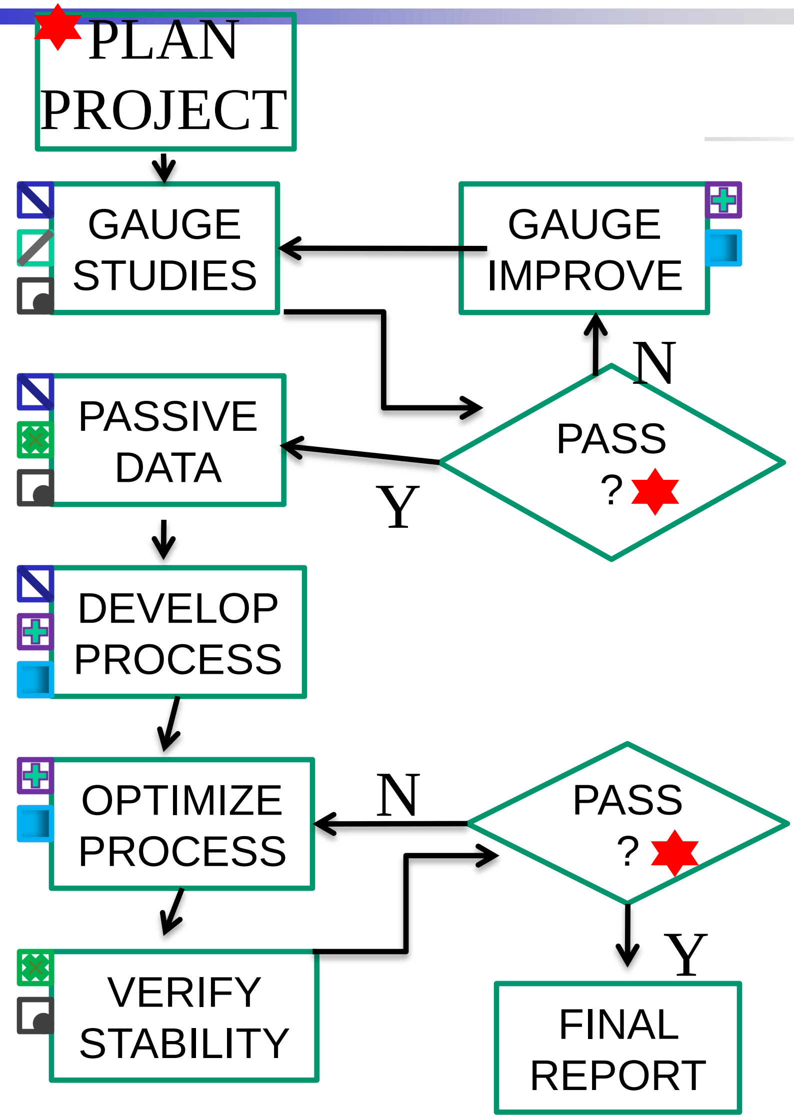
REFERENCES

Three Types of 'Experiments'

- DISCOVERY (Process Development)
  - Designed to generate new ideas or approaches
  - Usually involve 'hands-on' activities
  - May involve processes that are not well understood
  - Processes should be at least reasonably stable
- HYPOTHESIS (Refining a Process; Making decisions)
  - We are trying to 'prove' a theory
  - Accept or reject some pre-specified hypothesis
- OBSERVATION
  - 'Take what data we can get' & look for associations
  - Not discussed here
  - Lots of 'grey' areas
- STATS METHODS: a TABLE is given at the END

Ideal Engineering Environment

- MANAGEMENT
- EXPLORE THE DATA
- GAUGE STUDIES
- CHARACTERIZATION
- MODELING
- DOE ↔IMPROVE
- CONTROL/MONITOR
- Comparisons (Stat tests) occur at all stages except
- Note: Long-term trials (Reliability) not covered



Strategies: OFAT vs. FACTORIAL

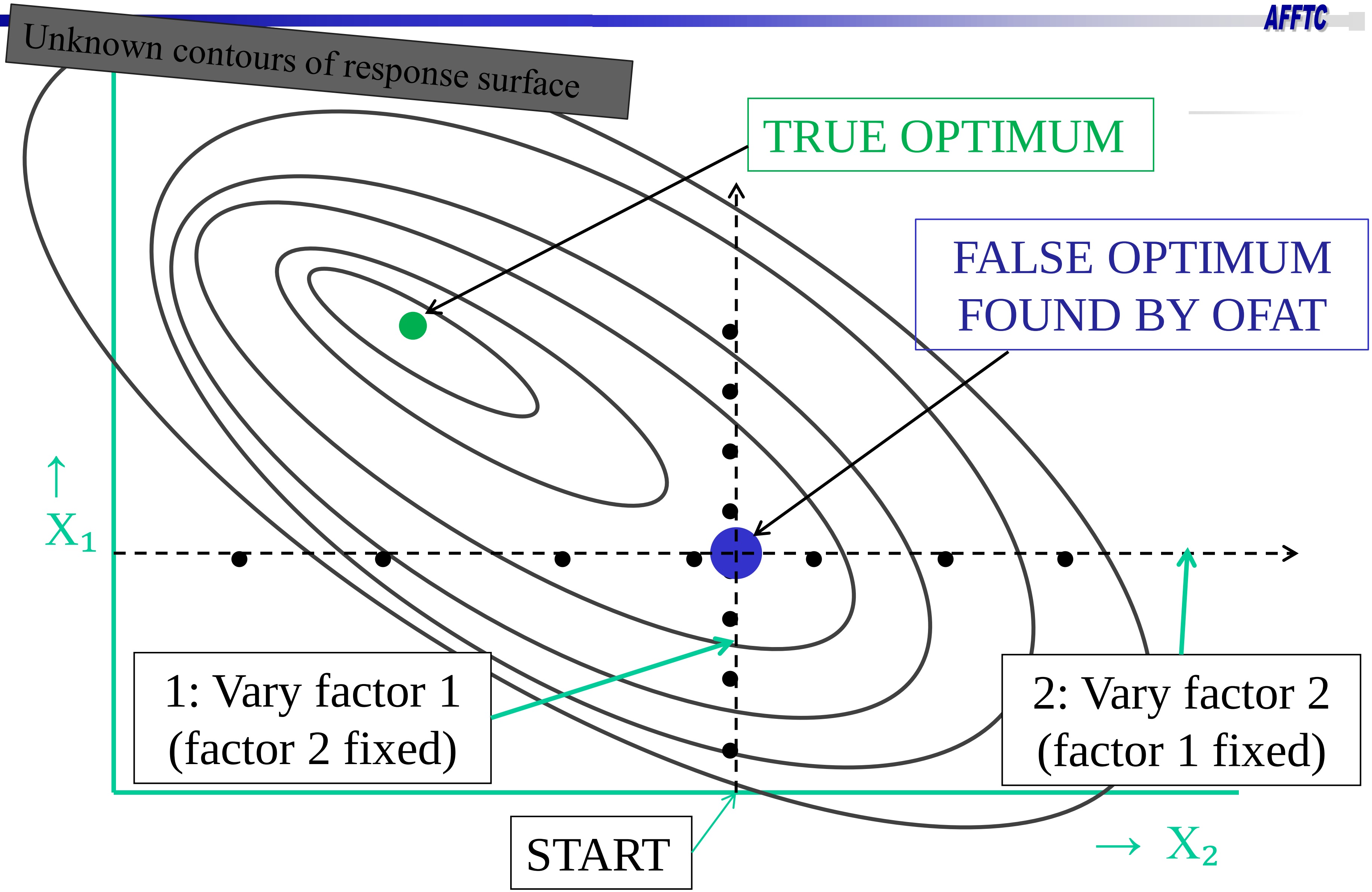
Experimental Strategies

- <u>One Factor at a Time (OFAT)</u>
– PRO: Straight-forward
– CONs: 'Over-collects' data
  – Can't find Interactions
- Not recommended.
- <u>Factorials-Full & Fractional</u>
– PROs: Individual & interaction effects estimated
- Most efficient designs for the sample size –> save $$
– CON: More complex

One-factor-at-a-time

- Start at a known point and vary one factor only until a 'best' value for the system is obtained.
- Fix that first factor at its new 'best' value and vary only the second factor. Find a new system 'best' value.
- Call the new point 'Optimal'

Why OFAT Often Fails to Optimize



OFAT vs. FULL FACTORIAL DESIGNS

AUTOMATIC TARGET RECOGNITION

(James M. Higdon, Air University, Wright-Patterson AFB, Ohio, 2001)

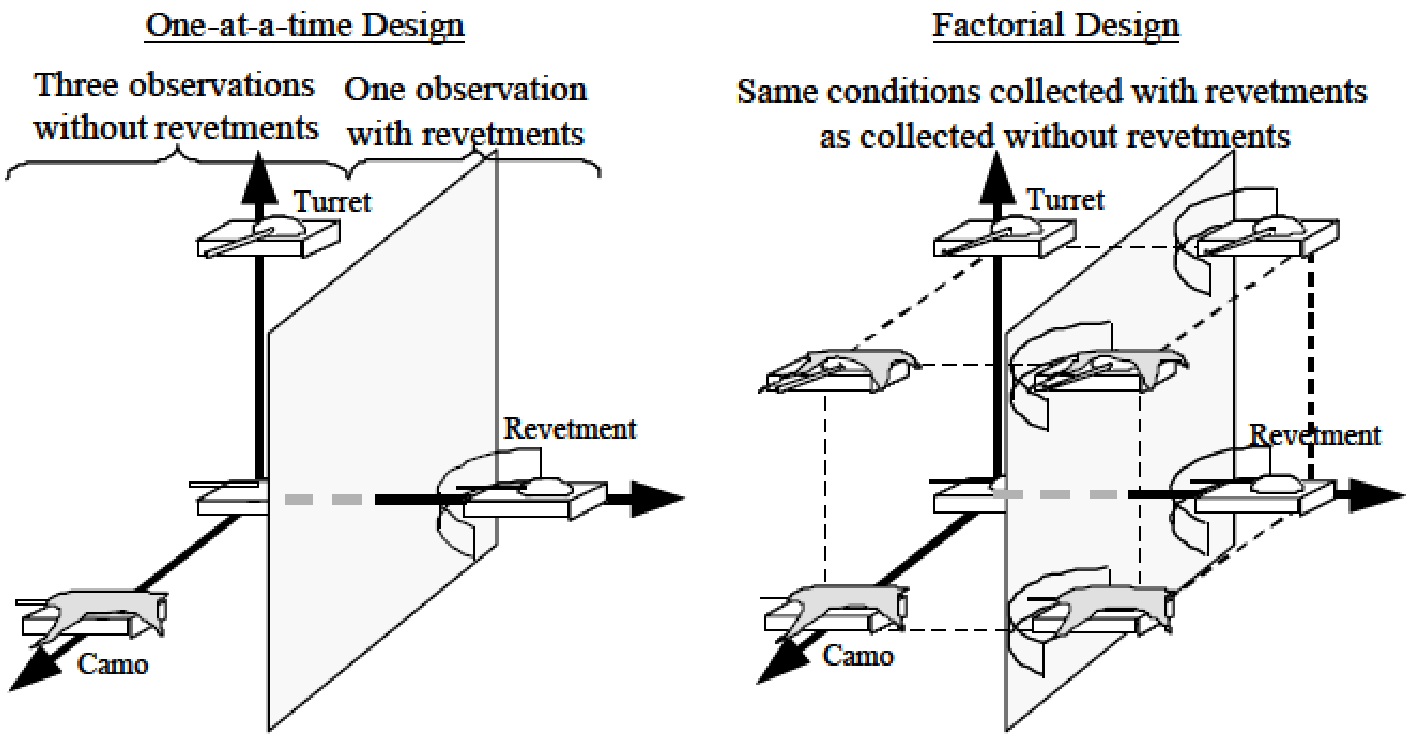


Figure 3.1 Comparison of One-at-a-time and Factorial Conditions in Three Factors

Design: A 2^3 Full Factorial
Vane Cleaning Example (John C. Sparks, Wright-Patterson AFB)

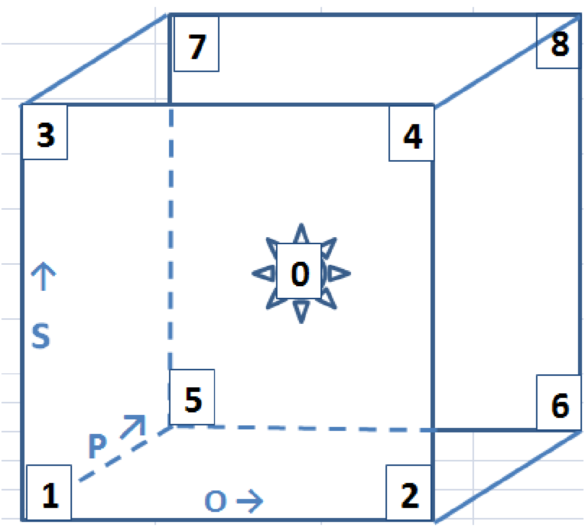
AFFTC

- **PROBLEM**: A gas-turbine vane becomes corroded in service & requires periodic cleaning with high-pressure water delivered via a tiny orifice. Want to minimize % corrosion after cleaning.
- **Brainstorm Factors**
  - O: Orifice size
  - S: Standoff
  - P: Pressure

FACTOR	-1 LEVEL	0 LEVEL	+1 LEVEL
O	0.007"	0.5035"	1.0"
S	0.5"	0.75"	1.0"
P	20KSI	27.5KSI	35KSI

BASIC DESIGN: $2^3 = 8$

RUN	O	S	P
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	-1
5	-1	-1	+1
6	+1	-1	+1
7	-1	+1	+1
8	+1	+1	+1



Randomization + Control Runs

Random run order

Random	BASIC DESIGN: $2^3 = 8$			
order	RUN	O	S	P
4	1	-1	-1	-1
5	2	+1	-1	-1
1	3	-1	+1	-1
3	4	+1	+1	-1
7	5	-1	-1	+1
2	6	+1	-1	+1
8	7	-1	+1	+1
6	8	+1	+1	+1

Control runs (0) added

Actual	Rand	BASIC DESIGN: $2^3 = 8$				Y
order	order	RUN	O	S	P	Output
1		control	0	0	0	y_1
2	1	3	-1	+1	-1	y_2
3	2	6	+1	-1	+1	y_3
4	3	4	+1	+1	-1	y_4
5	4	1	-1	-1	-1	y_5
6		control	0	0	0	y_6
7	5	2	+1	-1	-1	y_7
8	6	8	+1	+1	+1	y_8
9	7	5	-1	-1	+1	y_9
10	8	7	-1	+1	+1	y_{10}
11		control	0	0	0	y_{11}

AFTER RUNNING THE EXPERIMENT AS DESIGNED, PUT THE RESULTS INTO A REGRESSION.

Note that running control runs (0) achieves two things: You can see if the process has drifted during the experiment, and you get a better estimate of 'noise.'

Get The Data; Run Regressions

AFFTC

DOE outcomes: (simulated data)

O	S	P	Y
0.0	0.0	0.0	79.1
-1.0	-1.0	-1.0	77.1
1.0	-1.0	-1.0	89.4
-1.0	1.0	-1.0	78.0
1.0	1.0	-1.0	77.5
0.0	0.0	0.0	78.6
-1.0	-1.0	1.0	76.8
1.0	-1.0	1.0	90.2
-1.0	1.0	1.0	75.4
1.0	1.0	1.0	84.3
0.0	0.0	0.0	79.0

Regression 1 : Y = f(O,S,P)

	coef(b)	SE	t	P(>|t|)
(const)	80.4923	0.5677	141.790	7.74e-07
O	4.2558	0.6657	6.393	0.00775
S	-2.2815	0.6657	-3.427	0.04162
P	0.5771	0.6657	0.867	0.44977
O:S	-2.1659	0.6657	-3.254	0.04736
O:P	1.3115	0.6657	1.970	0.14342
S:P	0.4508	0.6657	0.677	0.54678
O:S:P	1.0353	0.6657	1.555	0.21774

- --- Adjusted R-squared: 0.8643

Regression 2: $Y = b_0 + b_1 O + b_2 S + b_{12} OS$

O	4.2558	0.8157	5.217	0.00123
S	-2.2815	0.8157	-2.797	0.02664
O:S	-2.1659	0.8157	-2.655	0.03269

- --- Adjusted R-squared: 0.7963

Full vs. Fractional Designs

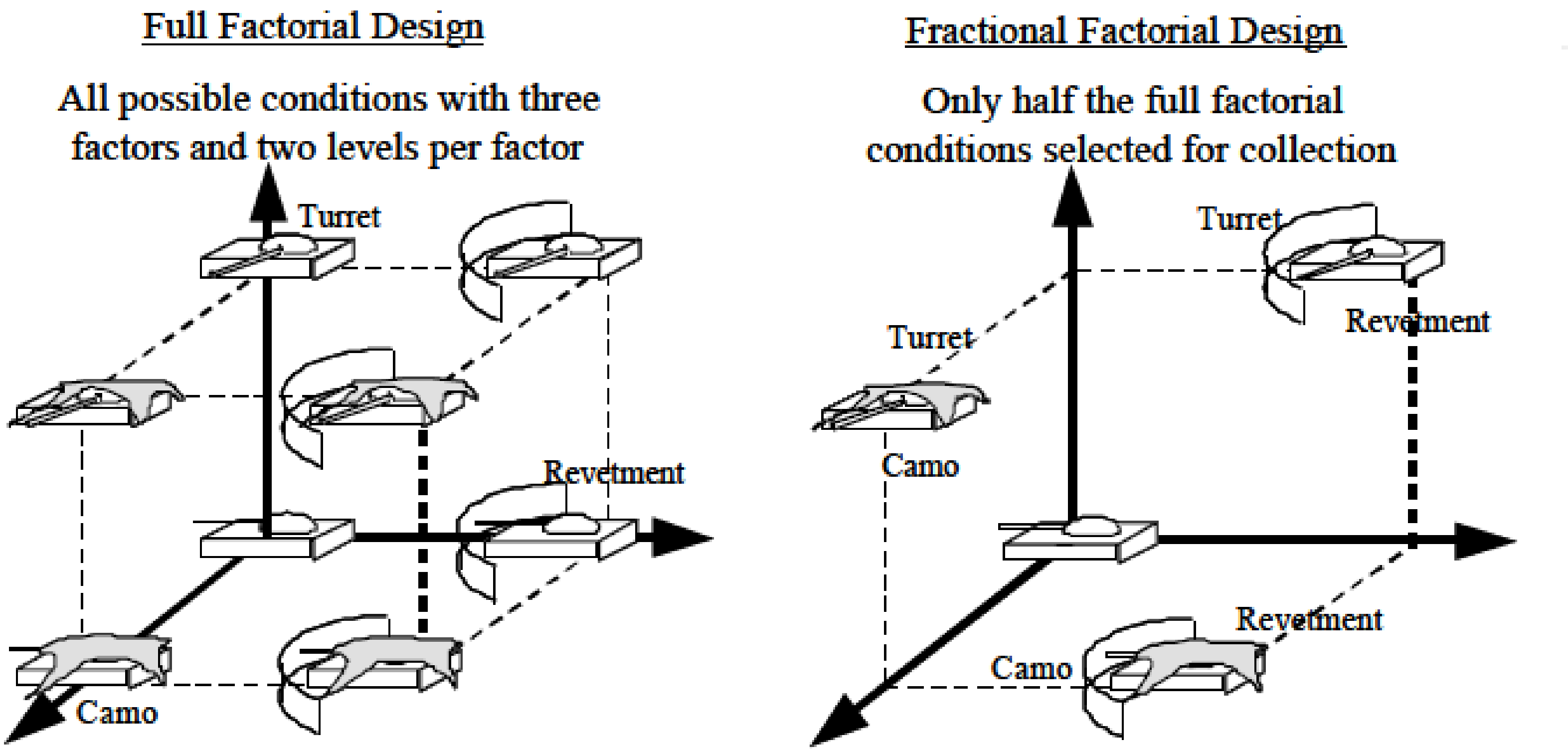
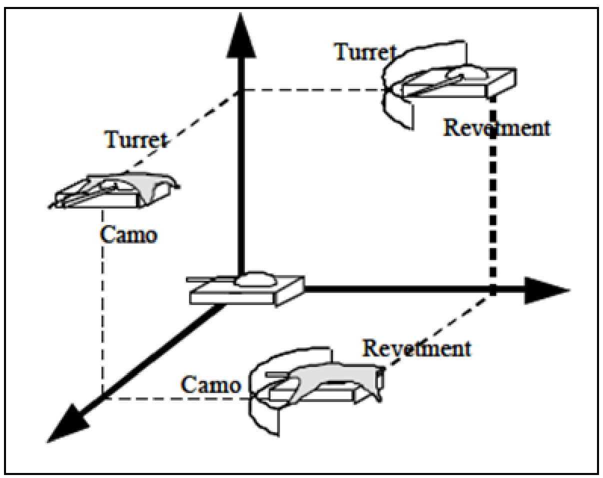
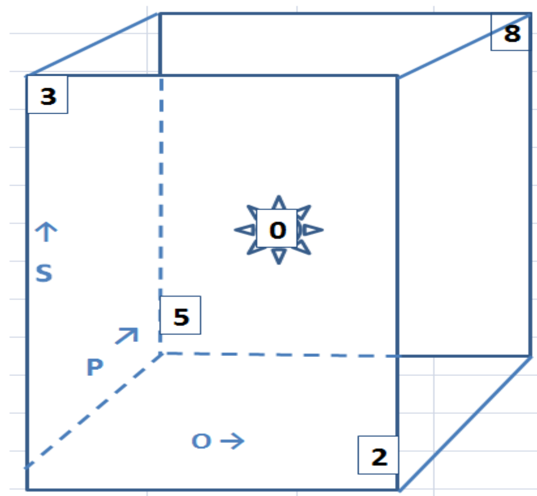


Figure 3.2 Comparison of Full and Fractional Factorial Conditions in Three Factors

2^{3-1} Fractional Factorial

- With the 2^3 design we ran <u>all possible</u> factor combinations
  - That is, all corners of the 'box'. 2^3 is called a <u>full</u> factorial
- We can save on number of runs by running only some corners of the box, say half of them. $2^{3-1} = 2^3 2^{-1} = 2^3 \times ½ = 4$ runs
  - That is, a half-fraction of 2^3 -- a <u>fractional</u> factorial design





BASIC DESIGN: $2^{3-1} = 4$

RUN	O	S	P
2	+1	-1	-1
3	-1	+1	-1
5	-1	-1	+1
8	+1	+1	+1

- Similarly, there we have 2^{4-1}, 2^{4-2}, 2^{9-3}, 2^{n-k}, 3^{n-k}, etc.
- 2^{n-k} designs are the simplest and most popular for DOE
- To understand them better, we need to understand <u>***INTERACTION***</u>

What is a (Two-Way) Interaction ?

AFFTC

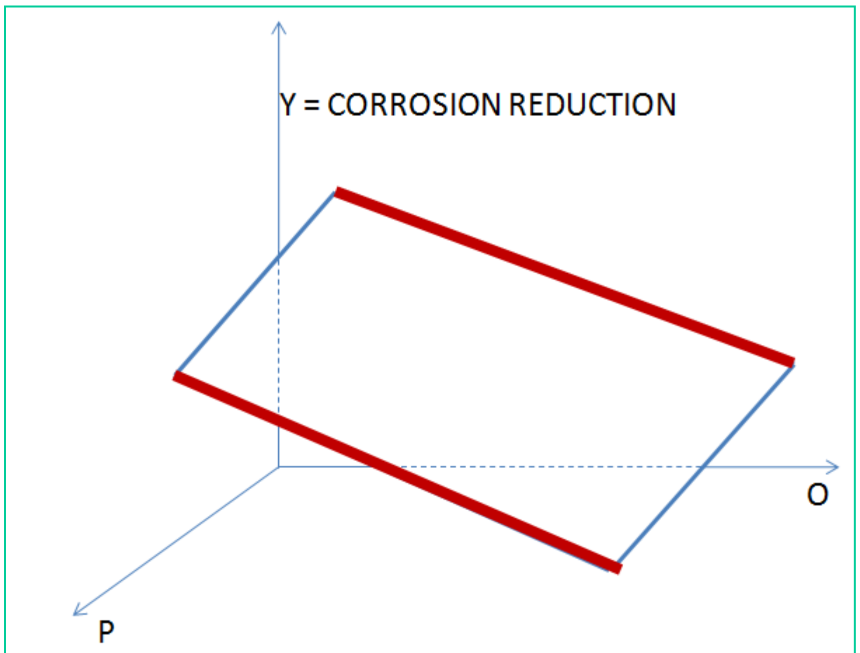
$$Y = b_0 + \underbrace{b_1 O + b_2 S + b_3 P}_{\text{main effects}} + \underbrace{b_{12} OS + b_{23} SP + b_{13} OP}_{\text{2-way interactions}}$$

constant

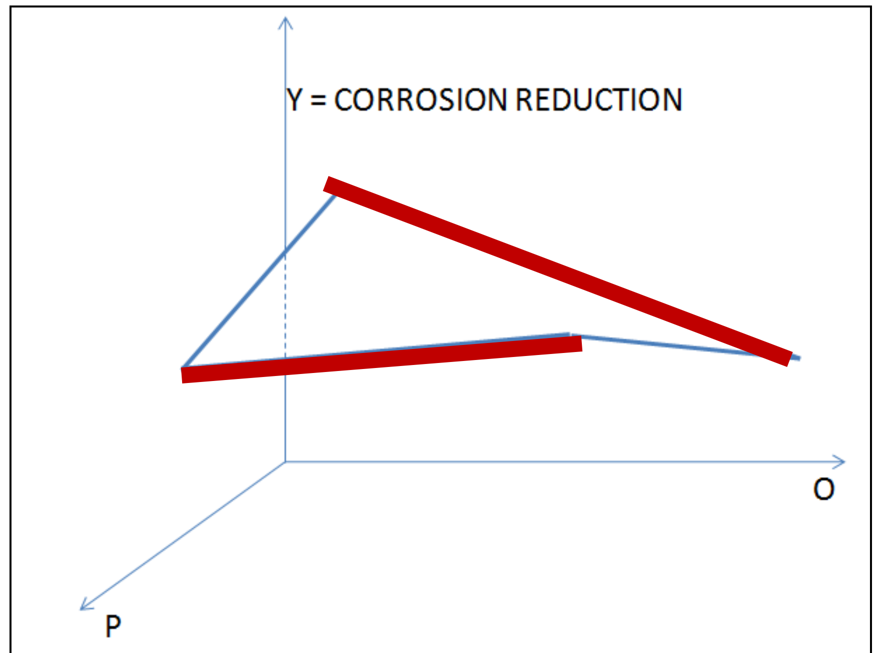
main effects

2-way interactions

NO OP-INTERACTION	**OP-INTERACTION**



The effect of O on Y does not depend on P



As P changes, the effect of O on Y changes

What We Lose With A Fractional Design

AFFTC

- With the full 2^3 factorial we can estimate all main and interaction effects free & clear.
- With the fraction 2^{3-1} we alias (confound) 2FI with main effects. The 3FI = the average.
- So we have lost the ability to estimate all effects free and clear.
- The 2^{3-1} give just main effects. Such designs are called *main effects*, *saturated*, or *screening.*
- Other fractionated designs have different aliasing patterns. E.g. Main and 3FI's may be aliased.

		MAIN EFFECTS			2-FCTR. INTERACTIONS			3-F-I
	RUN	O	S	P	OS	OP	SP	OSP
2^3	1	-1	-1	-1	1	1	1	-1
	2	1	-1	-1	-1	-1	1	1
	3	-1	1	-1	-1	1	-1	1
	4	1	1	-1	1	-1	-1	-1
	5	-1	-1	1	1	-1	-1	1
	6	1	-1	1	-1	1	-1	-1
	7	-1	1	1	-1	-1	1	-1
	8	1	1	1	1	1	1	1

		MAIN EFFECTS			2-FCTR. INTERACTIONS			3-F-I
	RUN	O	S	P	OS	OP	SP	OSP
2^{3-1}	2	1	-1	-1	-1	-1	1	1
	3	-1	1	-1	-1	1	-1	1
	5	-1	-1	1	1	-1	-1	1
	8	1	1	1	1	1	1	1
	ALIASED WITH ------>				P	S	O	AVG.

Replication

- **Running a design over one or more times, with different random order for the runs (<u>except</u> the control runs), is called 'replication'.**
- **This is good when we want a higher confidence/more power to our estimates of the factor effects. It also gives better estimates of the Noise in the Signal/Noise ratios.**
- **But there is magic—if you show that one or more factors are not significant, you <u>automatically get replication</u> !**

<u>Suppose P is unimportant</u> in 2^3

	MAIN EFFECTS		2FI
RUN	**O**	**S**	**OS**
1	-1	-1	1
2	1	-1	-1
3	-1	1	-1
4	1	1	1
5	-1	-1	1
6	1	-1	-1
7	-1	1	-1
8	1	1	1

Classic factorial designs have great advantages, some hidden. If you can do them, they yield many benefits including $ savings.

Blocking The 2^3

	MAIN EFFECTS			2-FCTR. INTERACTIONS			3-F-I
RUN	O	S	P	OS	OP	SP	OSP
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

BLOCK 1

BLOCK 2

- Say you are obliged to run an experiment on 2 different tails (aircraft).
- You know 'tails' has an effect—an effect you are <u>not</u> interested in.
- Split the runs by sign of the 3FI & run one set (or 'block') per tail.
- This effectively 'blocks out' the tail effect from the other factor effects: **Blocking !**

Randomization, Replication, Blocking

- Randomization
  - Avoids bias / interdependence of observations
  - Helps “average out” effects of unknown nuisance factors
  - Special designs when complete randomization not feasible
- Replication
  - Permits better estimation of experimental error
  - Permits more precise estimates of the factor effects
  - Do not confuse with measuring repeatedly !
- Blocking
  - Designed to reduce unwanted ‘noise’- better S/N ratios

Other Designs: e.g. Mixture DOE

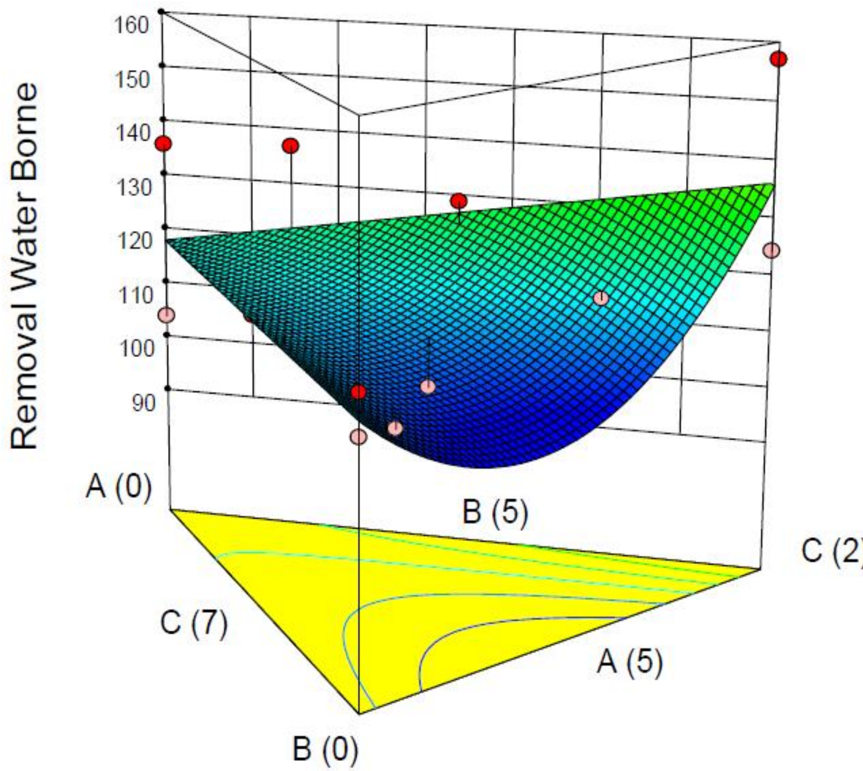
Time To Remove Aerospace Coating **(Chris Hensley, Aerochem, Inc.)**

AFFTC

- **An aircraft manufacturer needed to remove a chromated primer prior to applying harness hardware**
- **Existing mixes took 8 hours**
- **MSDS: Be in approved limits:**

$$A+B+C=12\% \quad A\leq 5 \quad B\leq 5 \quad 2\leq C\leq 7$$

- **The proportion of ingredient A was varied between 0 and 5%, ingredient B between 0 and 5%, and ingredient C between 2 and 7%.**
- **New formula took 35-40 min.**



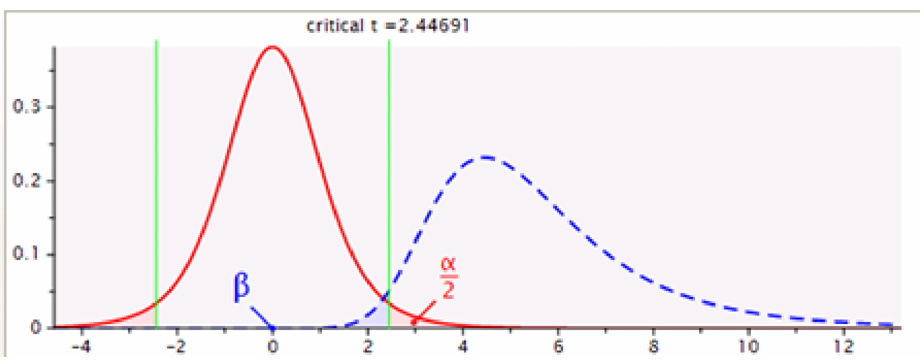
Some Other Designs

AFFTC

- SCREENING/SATURATED – MAIN-EFFECTS-ONLY DESIGNS
  - e.g., High-fraction factorials (2^{3-1}, 2^{7-4}, ..., Plackett-Burman)
- RSM-RESPONSE SURFACE METHODS
  - EG. Full and low-fraction factorials, Central composite, ...
- NESTED/HIERACHICAL – FACTORS ‘NESTED’ IN OTHERS
  - EG. tests nested in samples nested in batches nested in lots, ...
- SPLIT-PLOT –Randomization restricted in some way
- D-OPTIMAL – THE DESIGN SPACE IS IRREGULAR/Constrained
- SEQUENTIAL – NEXT RUN DEPENDS ON PREVIOUS RUN(s)

Power Assessment

	TRUE STATE OF NATURE	
DECISION	Ho true	Ha true
Accept Ho Reject Ha	No error Confidence = 1 - α	Type II error P(Type II) = β
Reject Ho Accept Ha	Type I error P(Type I) = α	No error Power = 1 - β



- In DOE we want to identify important effects with high Confidence (1-α) & high Power (1-β).
- This boils down to a t-test on coefficients.
- Useful free software is Lenth's power tool and G*Power.
- Commercial: PASS

Many Stat Tests: A Guide For The Perplexed

STILL NEED HELP? SPEAK TO YOUR LOCAL STATISTICIAN !

AFFTC

GOAL ↓ DATA →	Normal ~ (Gaussian)	Rank, Score, Non-Normal	Binomial (Two outcomes)
Describe one group	Mean, SD	Median, Interquartile range	Proportion
Compare one group to a constant	One-sample t-test	Wilcoxon's Signed-Rank test, Exact Permutation test, Sign test	Binomial or Chi-square
Compare two unpaired groups	Unpaired t-test	Exact (Permutation), Rank-Sum (aka Wilcoxon or Mann-Whitney), Welch's t-test	Fisher's Exact test, Chi-square (approximate)
Compare two paired groups	Paired t-test	Wilcoxon's Signed-Rank test, Exact Permutation test, Sign test	McNemar's test
Compare 3+ unmatched groups	1-way ANOVA	Kruskal-Wallis	Chi-square
Compare 3+ matched groups	Repeated measures ANOVA	Friedman test	Cochrane's Q
Quantify association between two variables	Pearson correlation	Spearman correlation	Contingency coefficients
Predict a value from one other variable	Simple linear/non-lin regression	Nonparametric regression	Simple logistic regression
Predict a value from several other variables	Multiple linear/non-lin regression		Multiple logistic regression

Adapted from: H. Motulsky 'Intuitive Biostatistics' © OUP, 1st ed. 1995

References

- **Hensley, Chris / Aerochem Inc.**
  - **Aerochem, Inc., 1017 Kingston Drive, Yukon, OK 73099. *Time to Remove Aerospace Coating*, Contact Chris Hensley, Ph: (Office) 888-241-5758 chensley@aerochem.net.**
- **Higdon, James M.**
  - ***Utility of Experimental Design in Automatic Target Recognition Performance Evaluation* by *James M. Higdon*. Air University, Wright-Patterson AFB, Ohio, 2001.**
- **Montgomery, Doug**
  - **Design & Analysis of Experiments. Wiley. 2005 – a popular text.**
- **SEMATECH/NIST - http://www.itl.nist.gov/div898/handbook/index.htm**
  - **A free, online statistics reference handbook.**
- **Sparks, John C.**
  - ***John C. Sparks* (AFRL/XPX). Wright-Patterson AFB, Ohio. *Process Control and Factorial Design of Experiments.***
- **Motulsky, Harvey**
  - **Intuitive Biostatistics, Oxford University Press, 1995.**
- **Lenth's Piface software tool**
  - **http://www.stat.uiowa.edu/~rlenth/Power/**
- **G*Power**
  - **http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/**