



# Air Force Flight Test Center



*War-Winning Capabilities ... On Time, On Cost*

## Monte Carlo

## Techniques for

## Estimating Power in

## Aircraft T&E Tests

## July 2011



Todd Remund  
& William Kitto

412 TW

661-277-6384

Approved for public release; distribution is unlimited.

AFFTC-PA No.: 11244

*Integrity - Service - Excellence*



# Power—what is it?

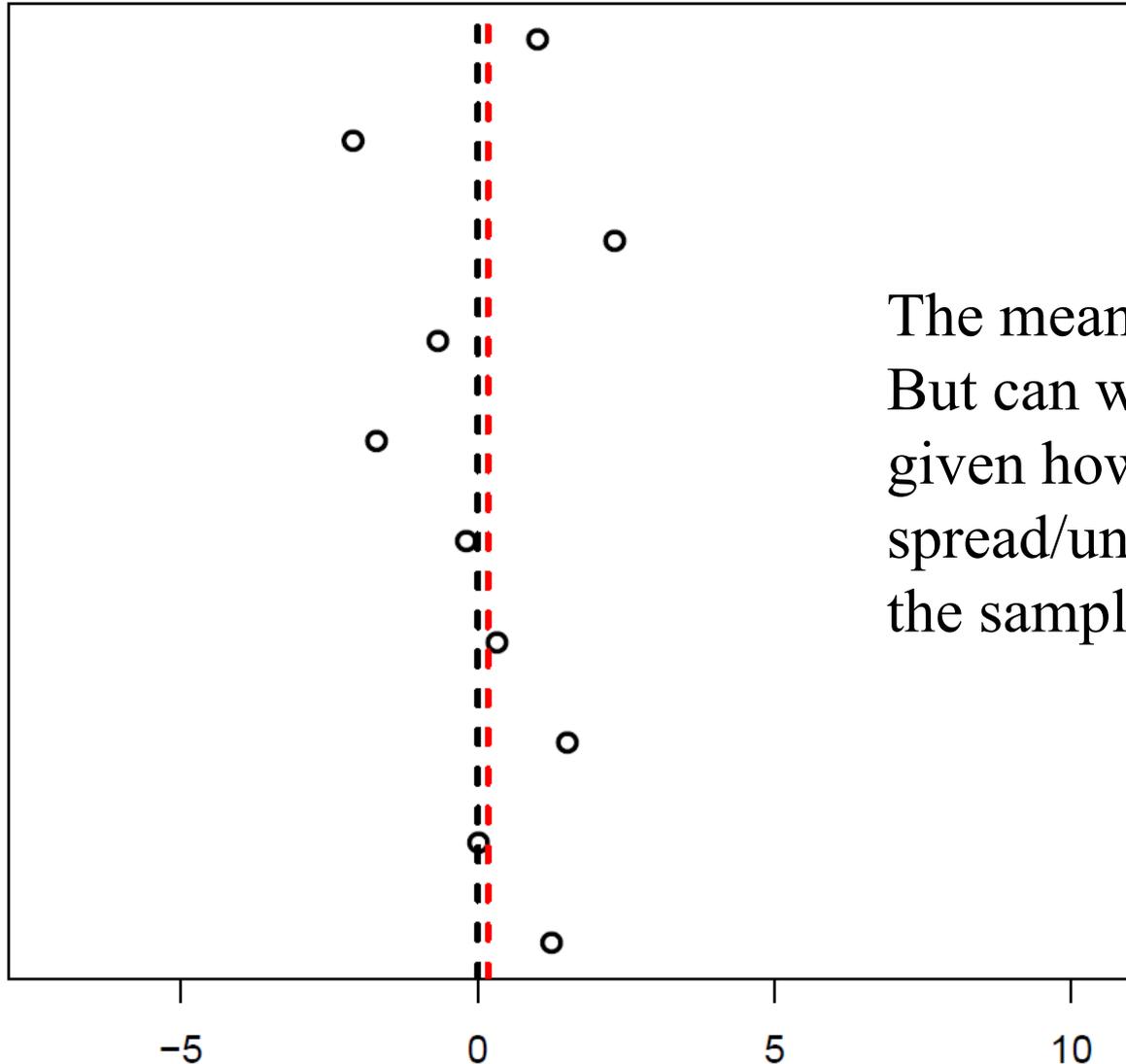


Truth	Decision	
	$H_0$ : No Change	$H_A$ : Change
$H_0$ : No Change	Correct Negative	False Positive
$H_A$ : Change	False Negative	Correct Positive

- **Power =  $\Pr(H_A | H_A \text{ is true})$** 
  - Choose sample size,  $n$ , to get this
  - Also need to decide what you want to see...stay tuned...
- **$\alpha = \Pr(H_A | H_0 \text{ is true})$** 
  - Choose this number directly
  - Normally 0.05 or 0.1



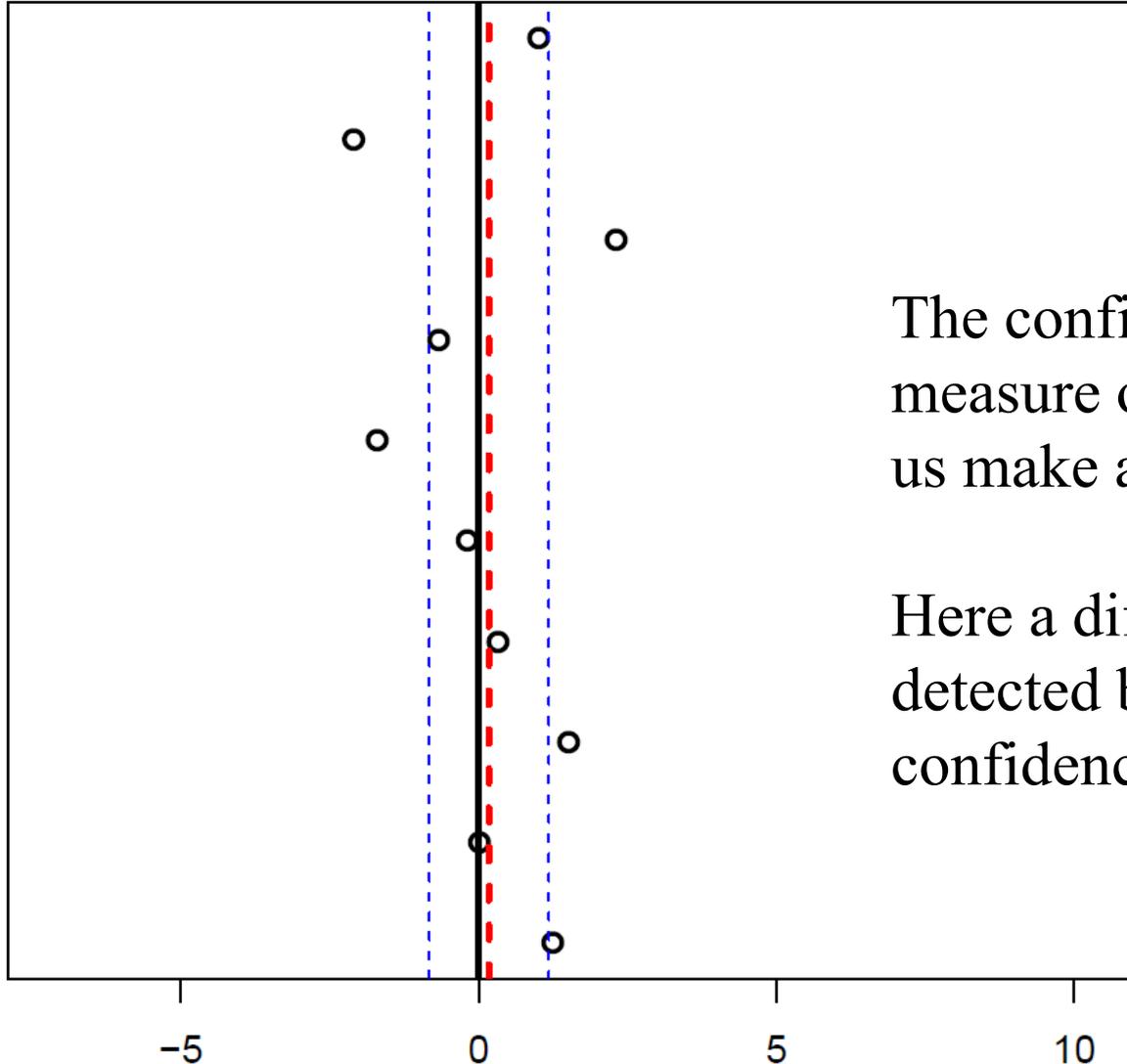
# Uncertainty and its vicissitudes



The mean is never equal to 0.  
But can we see the difference  
given how much  
spread/uncertainty there is in  
the sample?



# Uncertainty and its vicissitudes



The confidence intervals put a measure on uncertainty to help us make a decision.

Here a difference is not detected because the confidence intervals cross zero.



# Again...Power—what is it?



- **Power is the proportion of times, in the long run, that our test (t-test, CI) identifies a difference, when it really exists.**
- **WRT the mean, we need to decide how big of a difference from zero do we care about.**
  - **This is the effect size, called  $\delta$ .**
- **If we choose enough samples the CI shrinks, and it is easier to see a difference.**
- **But how large of a sample do we need to get to see the  $\delta$  we want?**



# Power—via Monte Carlo



- For many applications, such as the one given, power calculations are closed form.
- For other difficult applications, the calculation doesn't exist.
- **Generate Alternate Population**
  - With the desired effect
  - With the distribution characteristics needed
- **Sample from it repeatedly**
  - Each time analyzing the sample and record significance
- **Compute the proportion of significant outcomes**
  - This is power via Monte Carlo

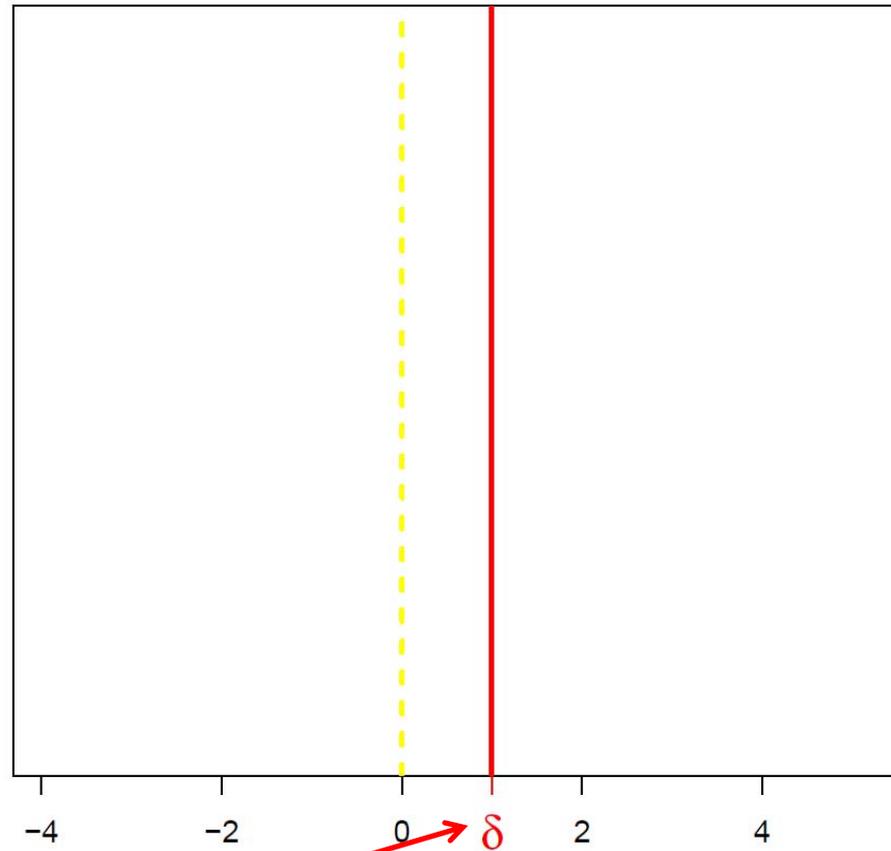


# Setting Up the 'Minimal' Alternative (1-sample t-test)



Effect Size  $\delta$

Couch the example in a one-sample t-test for simplicity's sake.



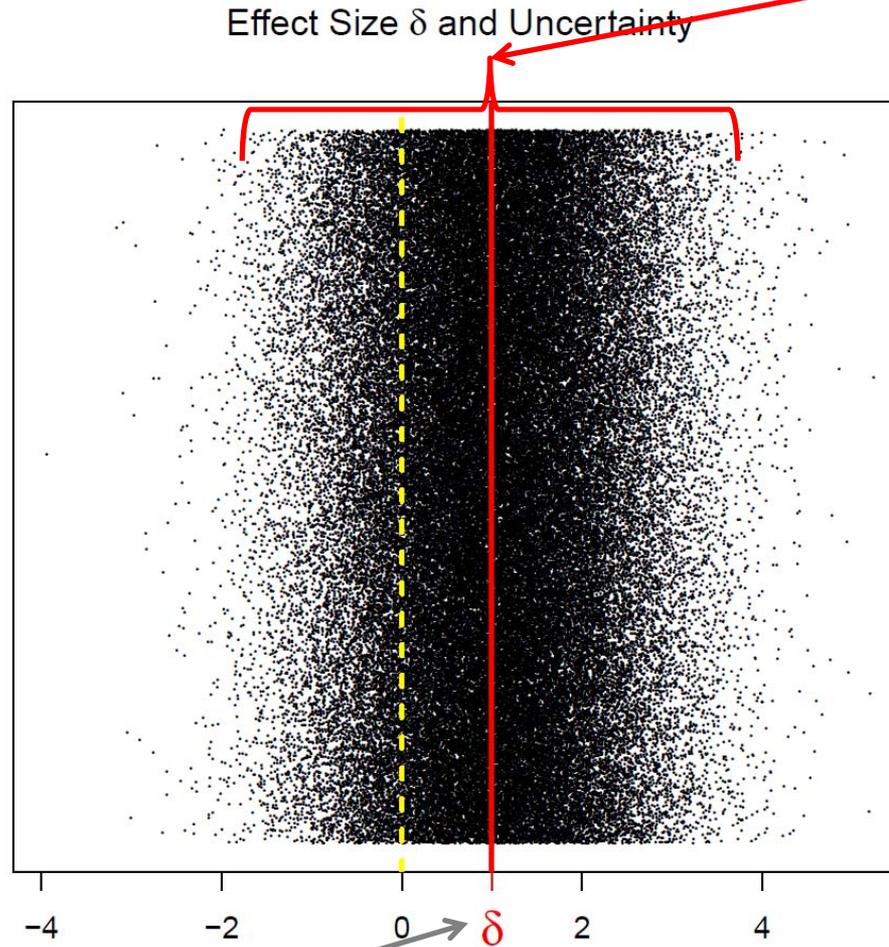
We want to see at least a  $\delta$  effect size 80% of the time...



# Setting Up the 'Minimal' Alternative (1-sample t-test)



Couch the example in a one-sample t-test for simplicity's sake.



With this much uncertainty (characteristics of population).

We want to see at least a  $\delta$  effect size 80% of the time...

Minimal Alternate Hypothetical Population



# Using the Alternate World

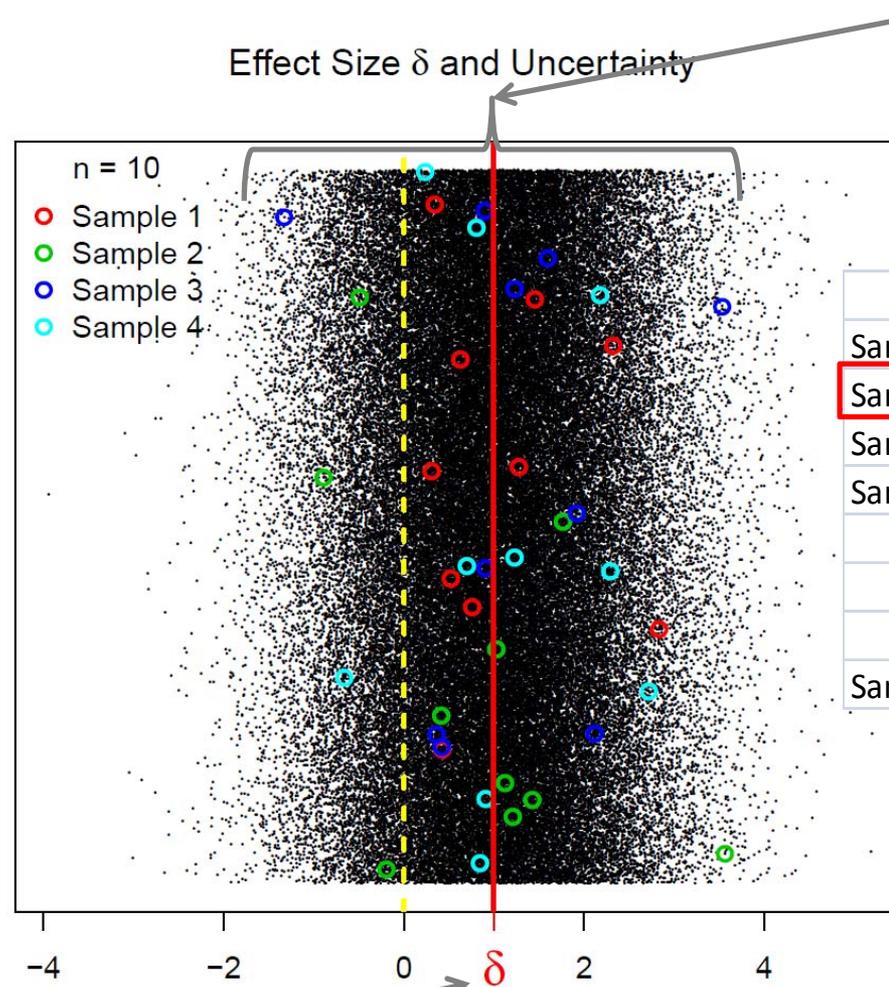


Couch the example in a one-sample t-test for simplicity's sake.

Now repeatedly sample from the population and run t-tests each time.

Each sample will have a different estimate for mean and standard deviation.

We want to see a  $\delta$  effect size 80% of the time...



With this much uncertainty (characteristics of population).

	95% CI	
Sample 1	0.45	1.72
Sample 2	-0.02	1.81
Sample 3	0.25	2.08
Sample 4	0.39	1.85
.	.	.
.	.	.
.	.	.
Sample 1000	0.41	1.59

Minimal Alternate Hypothetical Population



# Power Estimate



- With  $n=10$ ,  $st.dev=1$ ,  $\delta=1$  for Normal distribution we get—
- Using the Monte Carlo method to calculate power for the one-sample t-test:
  - Power = 80.2%
  - This method has a little variation in the estimate because it is a simulation approach.
- Using conventional methods:
  - Power = 80.31%



Click through PDF  
“SerialMeans.pdf” File

# SERIAL CORRELATION

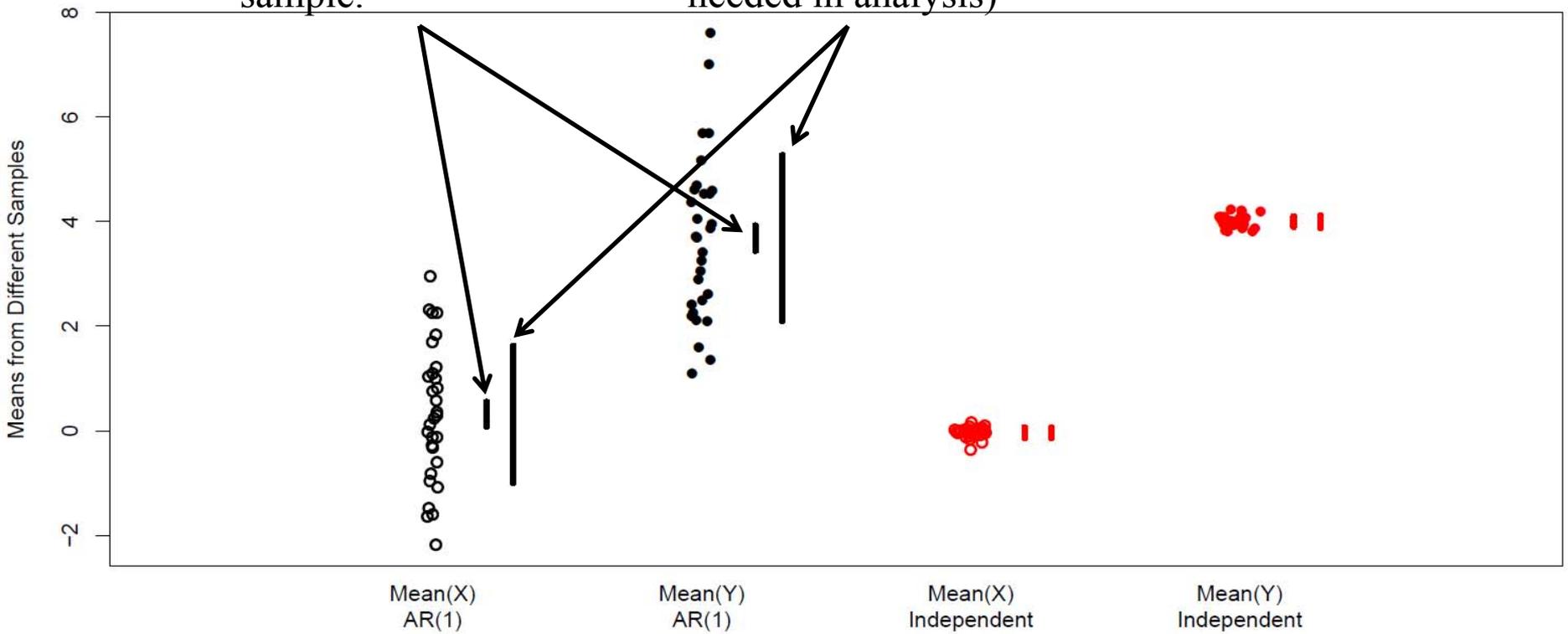


# 2-Sample t-test w/ Serial Corr.



Incorrect SE of mean  
from an individual  
sample.

Actual SE of mean.  
(An adjustment is  
needed in analysis)





# Need a different version of test.



- The difference in a regular 2-sample t-test, and an adjusted test is,
  - Estimate the autocorrelation,  $r$
  - Adjust the SE of the test statistic:

$$SE(\bar{x} - \bar{y})_{adjusted} = SE(\bar{x} - \bar{y}) * \sqrt{\frac{1+r}{1-r}}$$

$$CI = \bar{x} - \bar{y} \pm z_{1-\alpha/2} SE(\bar{x} - \bar{y})_{adjusted}$$

- What is the conventional method of computing power for this?



# How do we do it?



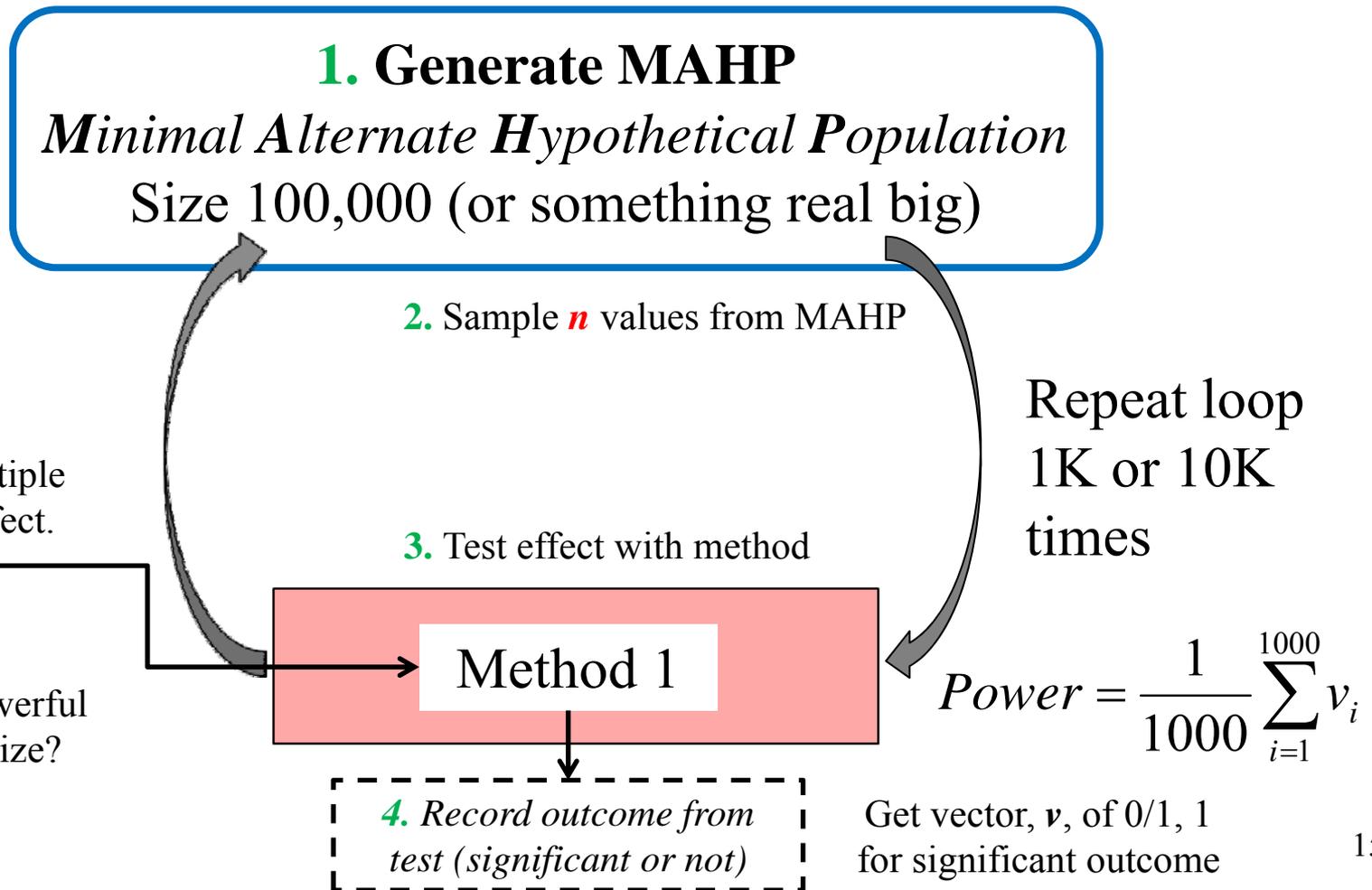
1. Create the *minimal alternate hypothetical population MAHP*
2. Take sample of size  $n$  from the MAHP
3. Test to see significance with chosen test, (here we're using the adjusted CI previous page).
4. Repeat steps 2 and 3 1000, 10000, or more times while recording how many are significant.
5. Find proportion that are significant out of number of repeated loops.



# Method: Plug and Play



Question: What  $n$  do we use to detect  $\delta$  with chosen power?





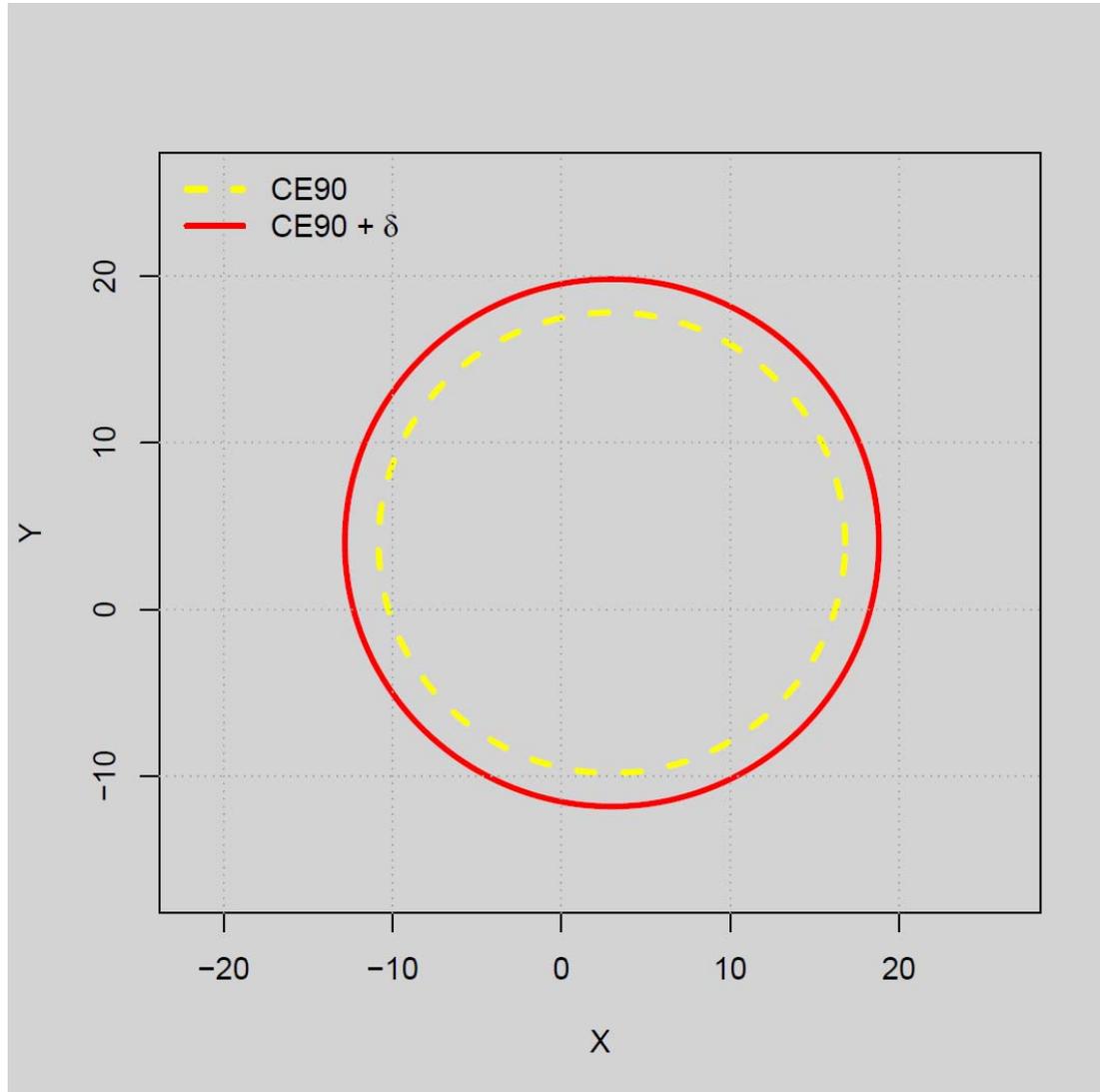
# CE90 Power



- **We want to know how many runs we need to prove CE90 is meeting spec for a targeting device.**
  - **How close to spec do you want to be before you are willing to concede that you are no different from spec?**
  - **We want to have proof of meeting spec if it is at least 2 feet beyond CE90.**

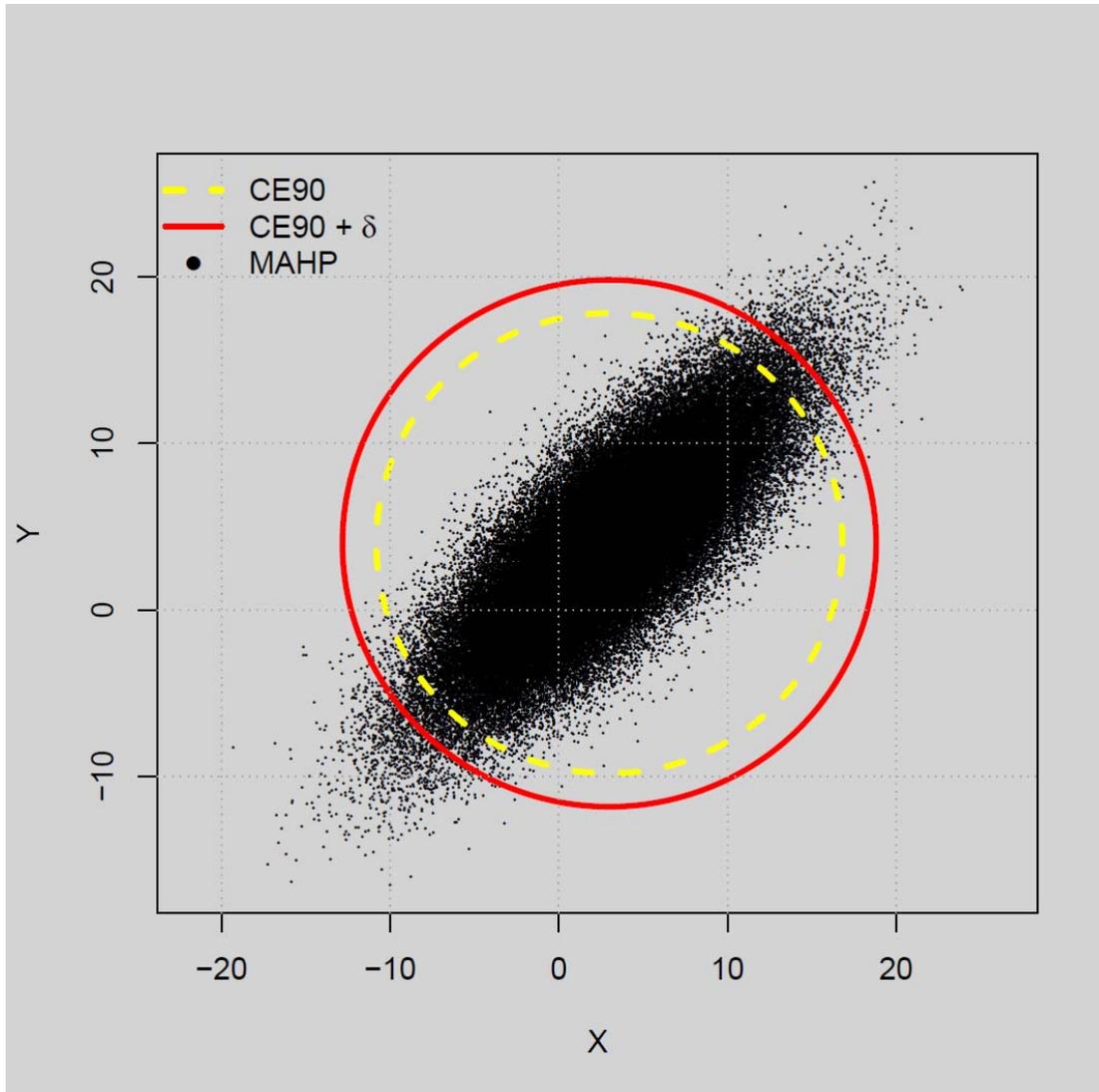


# CE90 MAHP



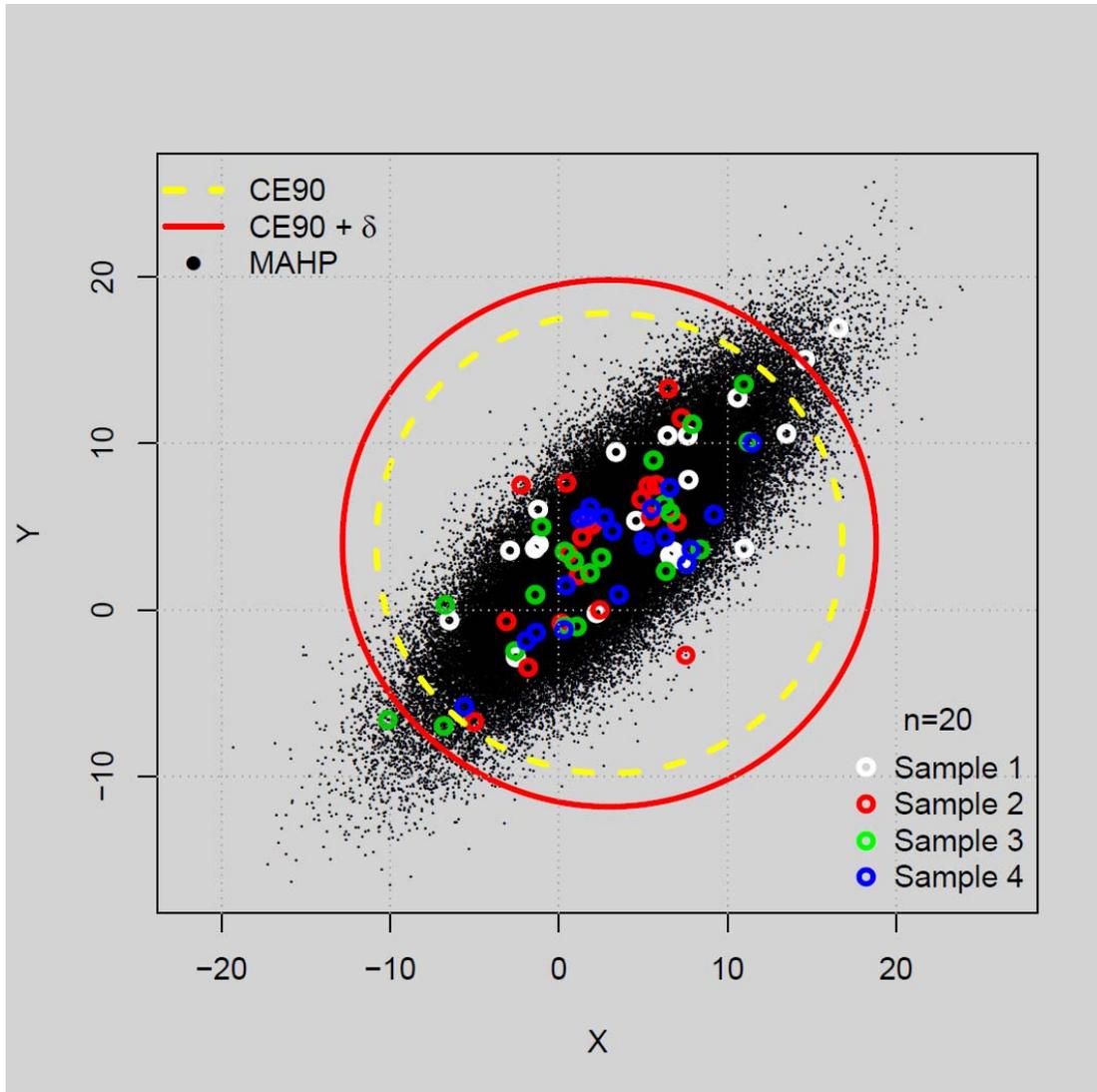


# CE90 MAHP





# CE90 MAHP





# CE90 Power



- **With  $\delta=2$  and sample size of 60,**
  - **Power = 70%**
- **This power calculation is only for the specific situations similar to the MAHP. Any different pattern in CE will require a separate power analysis.**



# Summary



- **Monte Carlo power estimation is versatile and can handle most situations.**
- **It is difficult sometimes to create the MAHP.**
- **A statistician is likely needed to aid in the process.**