An Animated Guide©:

The Logic of Statistical Power and *SAS® POWER PROCEDURES 13.2 PROC Power & PROC GLM Power*

Russell Lavery

How to calculate

How many subjects do I need in my study?

The Plan/Design is Very Important!

"To consult a statistician after an experiment is finished is often merely to ask him to conduct a post-mortem examination. He can perhaps say what the experiment died of." **R. A. Fisher (1938)**

Presentation Structure

- Intro to power
- The Central Limit theorem
- The Geometry of power
- Why do power & sample size calculations
- Components of a Power Analysis
- What tests must be covered
- Power Analysis in SAS

Hints for Delivering a Good Talk

- Social contract
- Know your audience

- Talk should have a Purpose
- Explicit Attendee benefits

- My Problems with this talk

Make you, and your boss, happy you attended BASUG

Email Slides & help you do a lunch and learn
-- Justify another day at BASUG

Part 1: Overview of statistical power

Part2: More examples than in any other paper

Save you some web searching

See some chart options

See command interactions (some of them)

See some “new” examples (of interest?)

Save you some typing (cut & paste)as you play

Examples of simulations and use of PROC Mixed

Importance of Power

When N and Power are Too small

- Fail to answer question
- Fail to detect associations
- Inconvenience subjects for no reason
- Misdirect future research!
  - Waste resources that could be better used elsewhere
  - Greatly delays application discovery/treatment

When N & Power are Too Large

- Study is more difficult
  - Study is more costly
  - Denies budget to other studies
- Inconveniences more subjects than is required
- Delays application of discovery/treatment

Importance of Power

If power is high and the test is non-significant, the result is useful.

Implies that the effect of treatment, if any, is small.
Non-result studies will be informative if power is high.

High powered studies give investigators more confidence in interpreting results – when Ho rejected or not

How many subjects have people needed in the past?

Polio Study:
The average incidence during the 1950s was about 50 per 100,000 or 0.0005.
Diagnosis was difficult, particularly in mild cases.
Mild polio symptoms are fever and weakness and are symptoms of many diseases

1,829,916 subjects

Lack of Power Analysis has lead to problems

In Prestigious medical journals between 1975 and 1990 more than 80% of randomized controlled trials that reported negative results didn't collect enough data to detect a 25% difference in primary outcome.

And nearly 2/3 did not have the power to detect a 50% difference.

From: Statistics done wrong by Alex Reinhart

Lack of Power Analysis has lead to problems

From 1959 to 1988 the result of a drug investigation was unclear

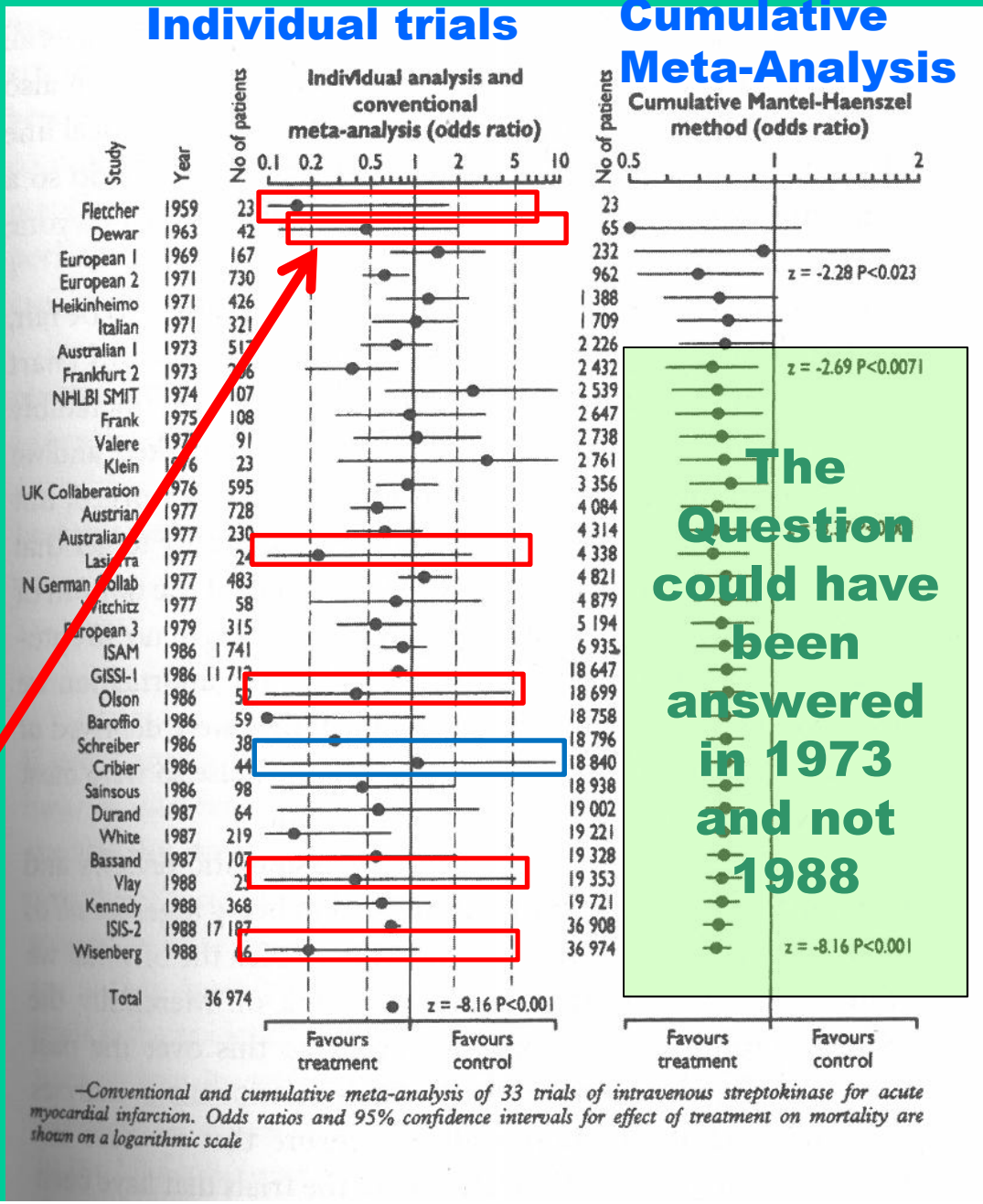
I suggest when you see a large effect and a CI that includes Ho you are underpowered

Power Analysis is useful and we should do it.

From a book: "Bad Pharma"

Individual trials

Cumulative Meta-Analysis



–Conventional and cumulative meta-analysis of 33 trials of intravenous streptokinase for acute myocardial infarction. Odds ratios and 95% confidence intervals for effect of treatment on mortality are shown on a logarithmic scale

The Question could have been answered in 1973 and not 1988

Power Analysis is considered difficult

Simple problems have simple formulas/ pictures

You need to know the stat to do a power analysis

A complete review of POWER is a review of stat.

Subject of ongoing research

Simulation might be the answer:

- Data Step programming
- IML can be convenient, if you have it

Presentation Structure

- Preliminary Information
- The Four Conditions that affect power
  - The **Degree of Shift** or Effect size
  - The effect of **Alpha**
  - The effect of **Variability**
  - Combining variability and degree of shift
  - **The number of obs** & Central Limit Theorem.
  - Integrating Variability, shift and N
- Review of power calculations balance cost and chance of two errors

Preliminary Info

- Generally considered a difficult topic
  One article advises not to attempt to explain it
- Aimed at non-statisticians
  Will provide background
- Learning theory recommends a “sandwich”
  Will contain repetition
- To avoid formulas we’ll use pictures
- Each statistical test has it’s own picture
  We’ll use a simple picture - a change in mean

Preliminary Info

The Logic of Power Calculations

- Power is the "Power of a test to detect an effect"
  (in this talk -a shift in the mean or average)

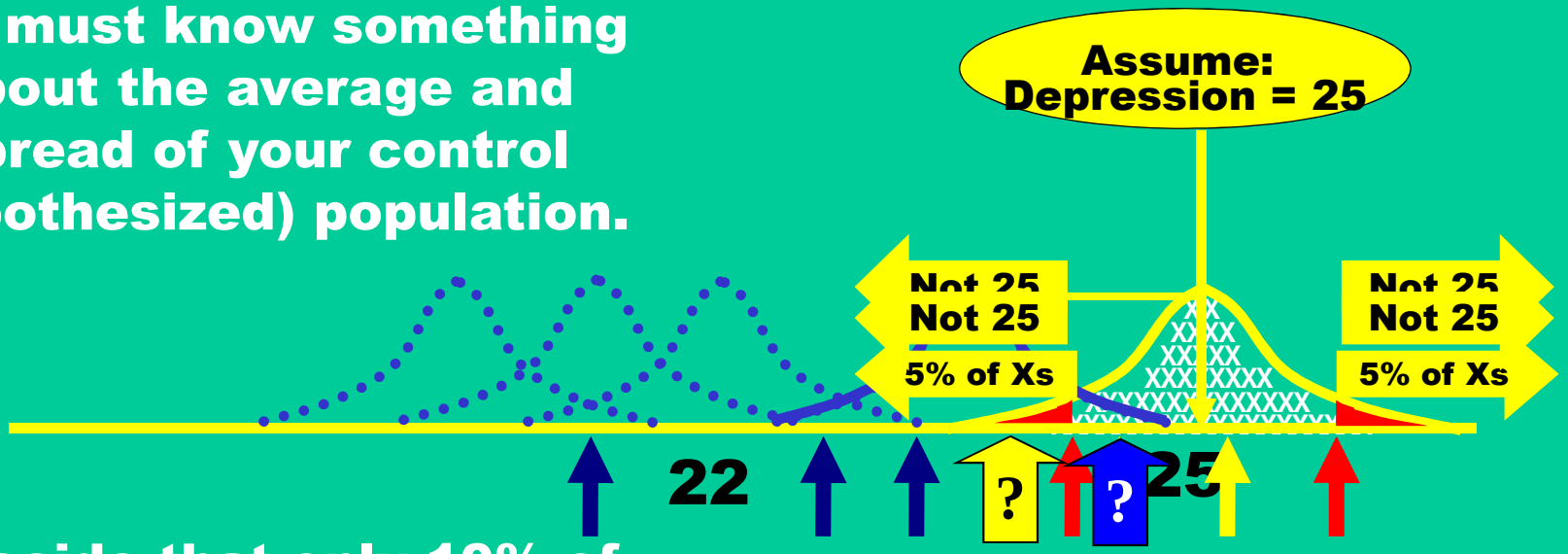
- Power analysis is balancing two statistical errors
  -Alpha Error and Beta Error.

- You can only talk about the power of a "situation"
  -when a set of conditions hold.

The Two Types of Mistakes

•Preliminary Information

You must know something about the average and spread of your control (hypothesized) population.



You decide that only 10% of the time you should see observations outside a certain range- alpha.

You expect your drug treatment to shift the mean.

But you do not know how much.

If you make an observation and it is outside the range, you conclude the depression score has changed

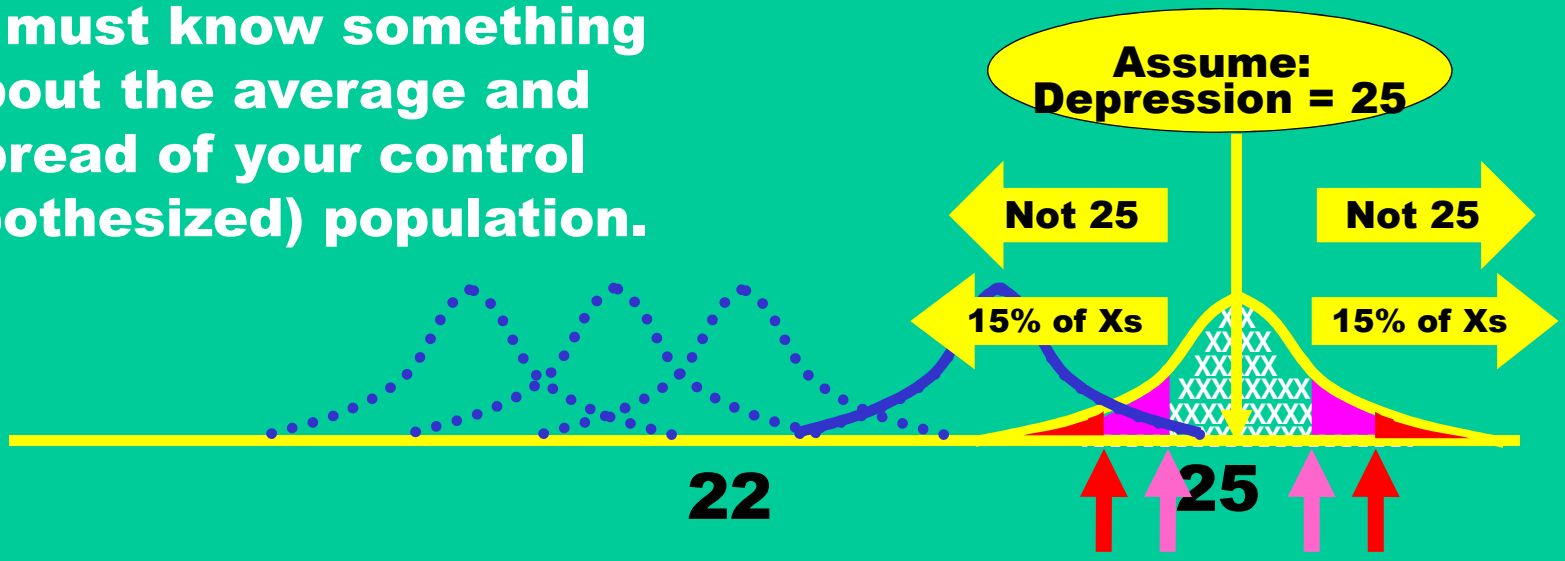
Some decisions are not easy to make.

You can make mistakes. The BIG yellow arrow could come from the distribution centered at 25 but the test would have you say the drug is effective in changing depression score

The Two Types of Mistakes

•Preliminary Information

You must know something about the average and spread of your control (hypothesized) population.



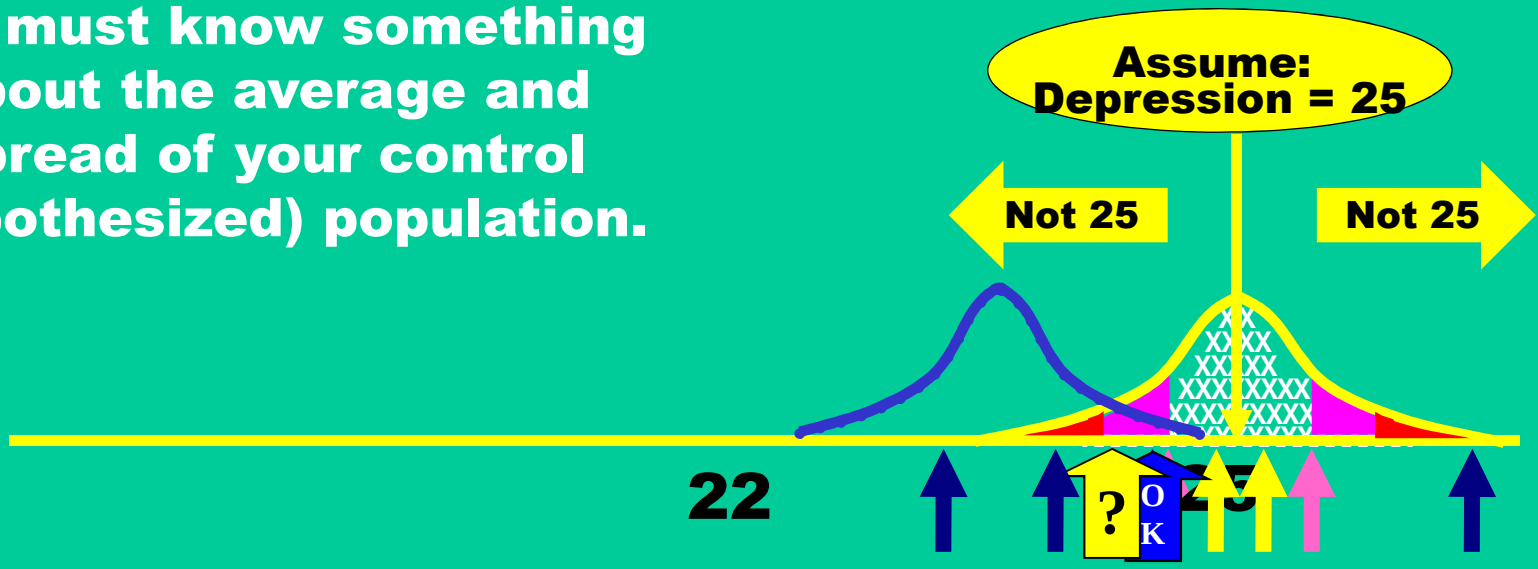
If you decide on a bigger alpha (say 30%)

The colored tails of the distributions must move in so that they contain 30% of the observations (X values)

The Two Types of Mistakes

•Preliminary Information

You must know something about the average and spread of your control (hypothesized) population.



If you decide on a bigger alpha (say 30%)

The colored tails of the distributions must move in so that they contain 30% of the observations (X values)

If you make alpha larger, you are more likely to detect that something has changed (more likely to Claim to have found an effective drug).

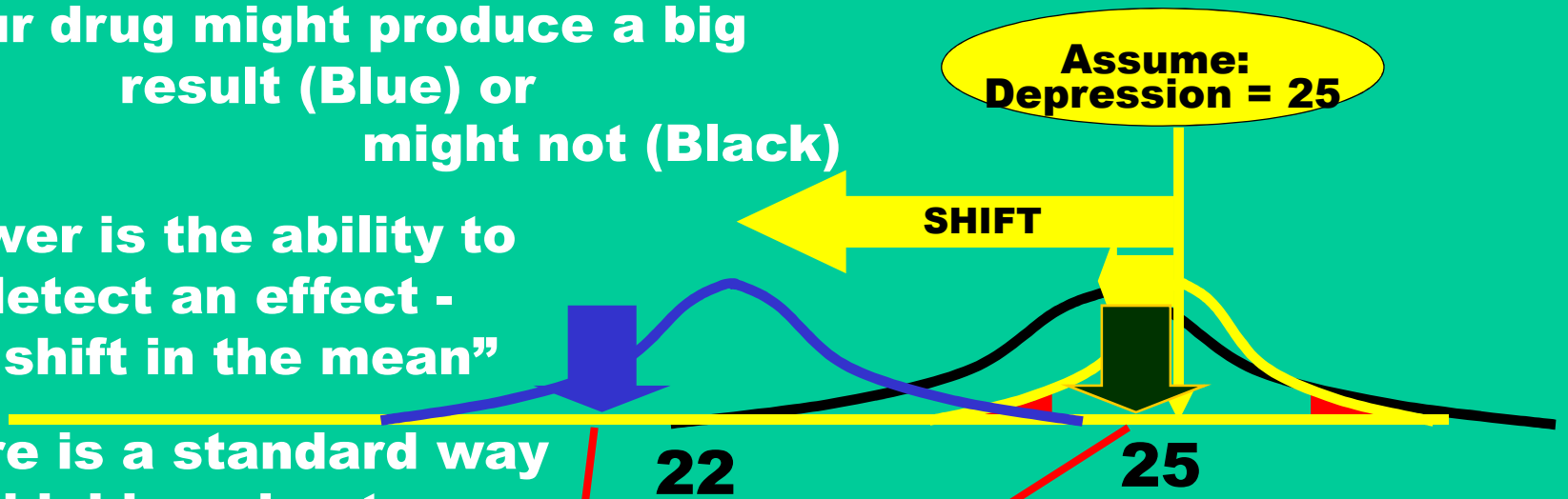
But you are also more likely to make a mistake by saying something has changed (claim an effective drug) when nothing has changed.

The Two Types of Mistakes

•Preliminary Information

Your drug might produce a big result (Blue) or might not (Black)

Power is the ability to detect an effect - "a shift in the mean"



There is a standard way of thinking about your decisions and possible mistakes.

Power **Analysis** is the **Balancing** of the two types of statistical mistakes.

Assume drug is ineffective

Assume Ho Depression=25

	Ho=True	Ho=False
Assume HO & kill the Drug	OK Decision	Beta Error
Reject Ho & market the Drug	Alpha Error	OK Decision

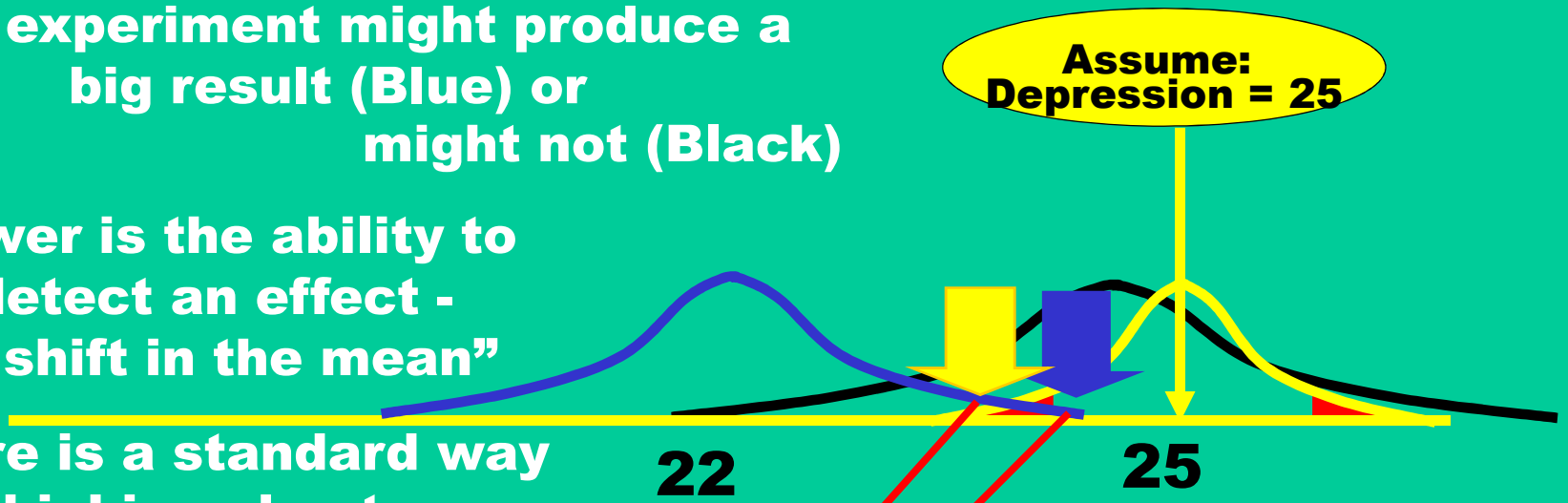
The Two Types of Mistakes

•Preliminary Information

Your experiment might produce a big result (Blue) or might not (Black)

Power is the ability to detect an effect - "a shift in the mean"

There is a standard way of thinking about your decisions and possible mistakes.



Power **Analysis** is the **Balancing** of the two types of statistical mistakes.

	Assume Ho Depression=25 Ho=True	Ho=False
Assume HO & kill the Drug	OK Decision	Beta Error
Reject Ho & market the Drug	Alpha Error	OK Decision

You can fail to detect when the mean has shifted (False Negative).

You can say the mean has shifted when it has not (False Positive).

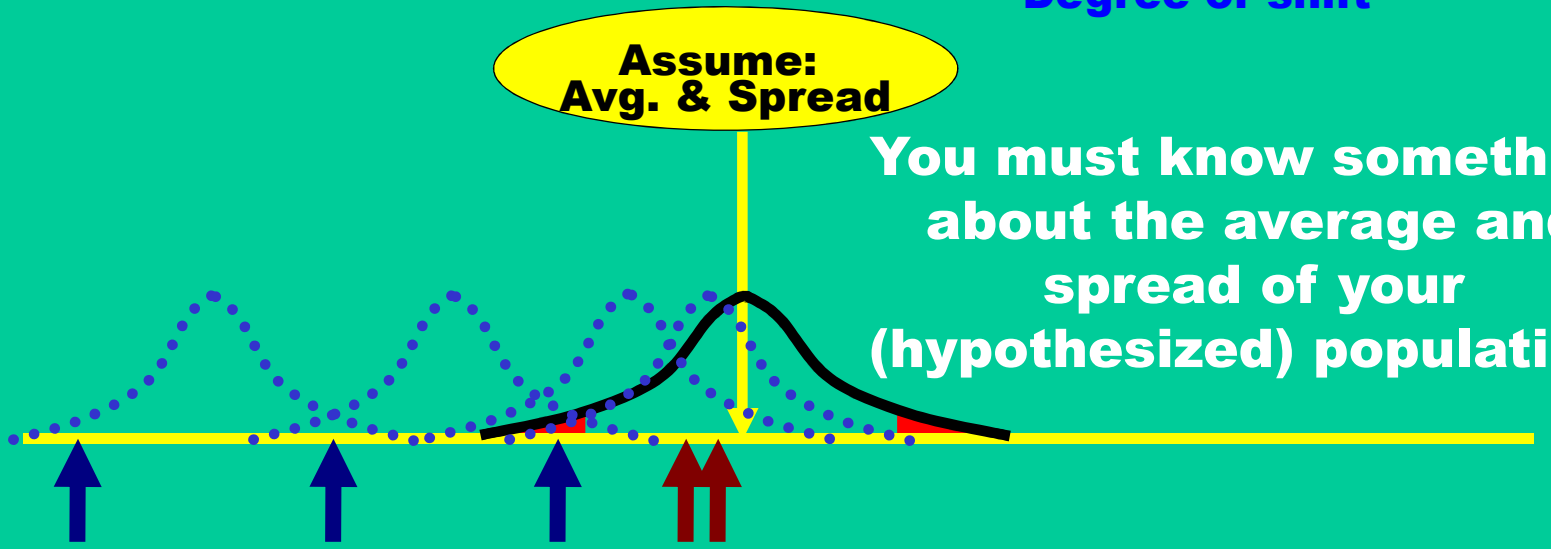
What is Statistical Power?

The Conditions of Power Calculations

- Power depends on the **degree of shift**
  (large shifts are easy to detect)
- Power depends on the **alpha error you select**
- Power depends on the **variability** in the process
  (lots of variability means low power)
- Power depends on the **number of observations**
  (more observations help detect shifts)
- Power depends on the **test** used
  (Some tests are more powerful than others)

What is Statistical Power?

- Intro to power: Conditions
- Degree of shift



You must know something about the average and spread of your (hypothesized) population.

You decide that, without **effective** treatment, 90% of the time you should see observations in a certain range.

You expect your drug to shift the mean of the distribution.

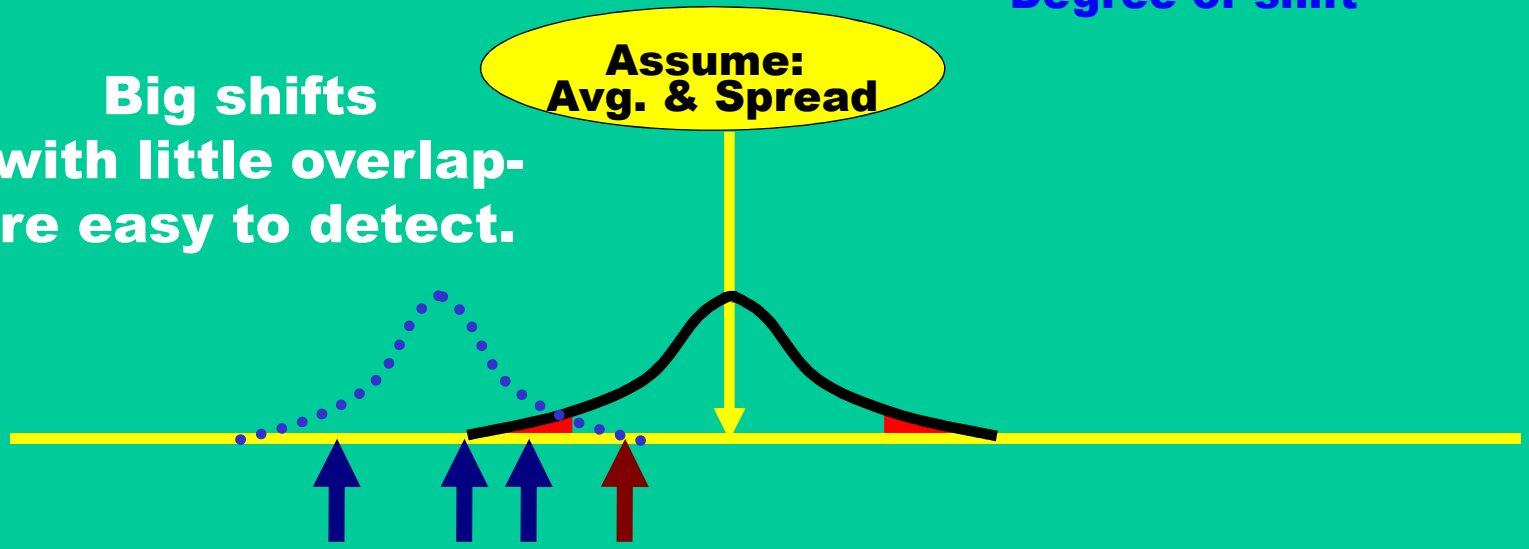
But you do not know how much.

If, after the experiment, you make an observation and it is between the red "tails", you assume the process has not changed ... and the drug is ineffective

What is Statistical Power?

- Intro to power: Conditions
- Degree of shift

Big shifts
- with little overlap-
are easy to detect.



Power is the percent of the area under the blue curve
NOT between the red triangles
The Issue is..

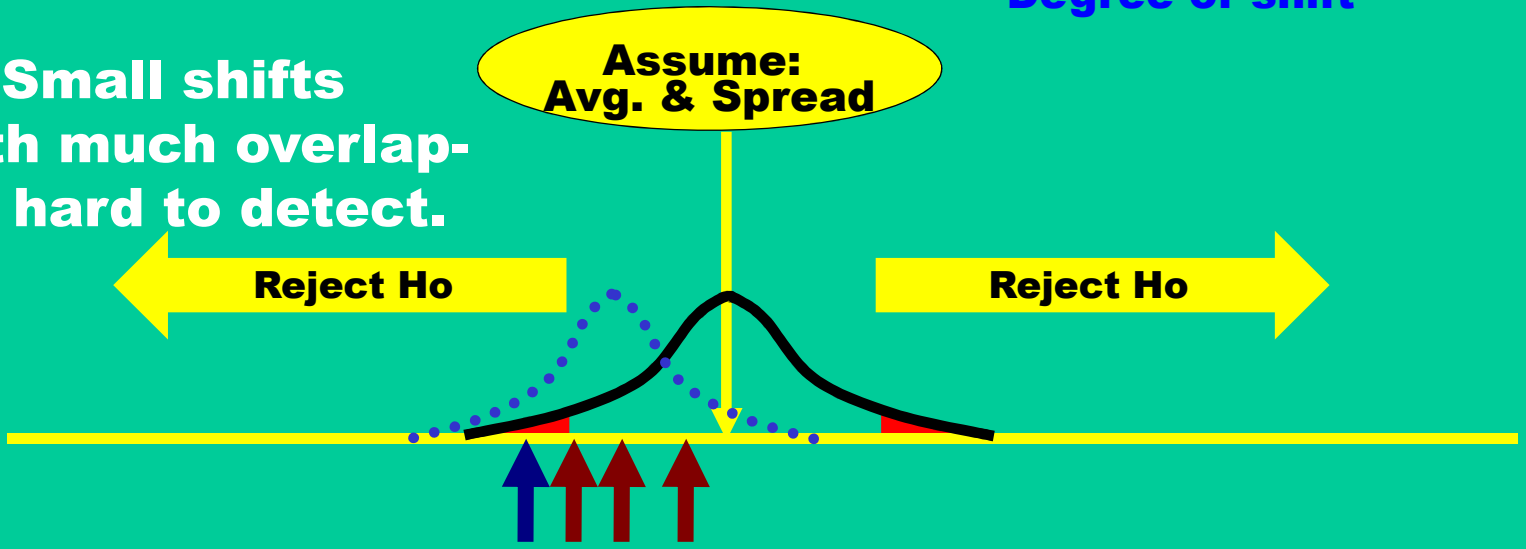
under the conditions pictured above, how likely
are you to get an observation
outside the red triangles.

With this blue distribution, what is the probability
of detecting that the average has changed.

What is Statistical Power?

- Intro to power: Conditions
- Degree of shift

Small shifts
- with much overlap-
are hard to detect.



Power is the percent of the area under the blue curve
NOT between the red triangles
The Issue is..

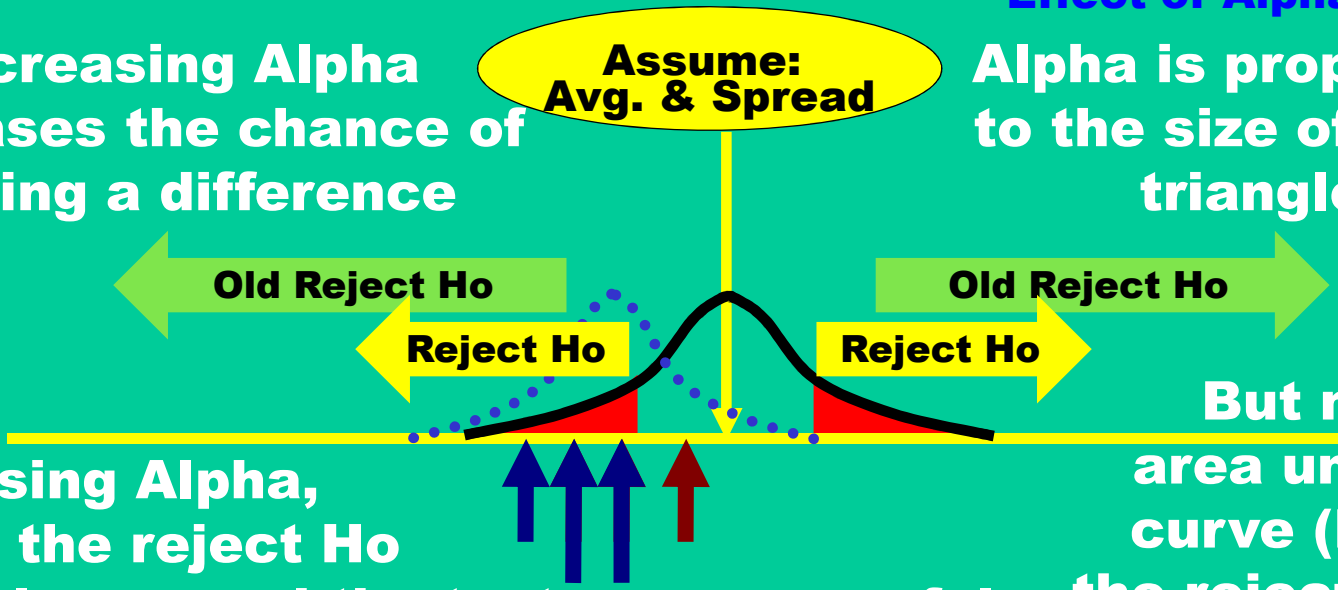
under the conditions pictured above, how likely
are you to get an observation
outside the red triangles.

With this blue distribution, what is the probability
of detecting that the average has changed.

What is Statistical Power?

- Intro to power: Conditions
- Effect of Alpha

Increasing Alpha Increases the chance of finding a difference



Alpha is proportional to the size of the red triangles.

Increasing Alpha, makes the reject Ho region larger and the test more powerful.

But more of the area under the Ho curve (Black) is in the reject Ho region

Power is the percent of the area under the blue curve ***NOT*** between the red triangles

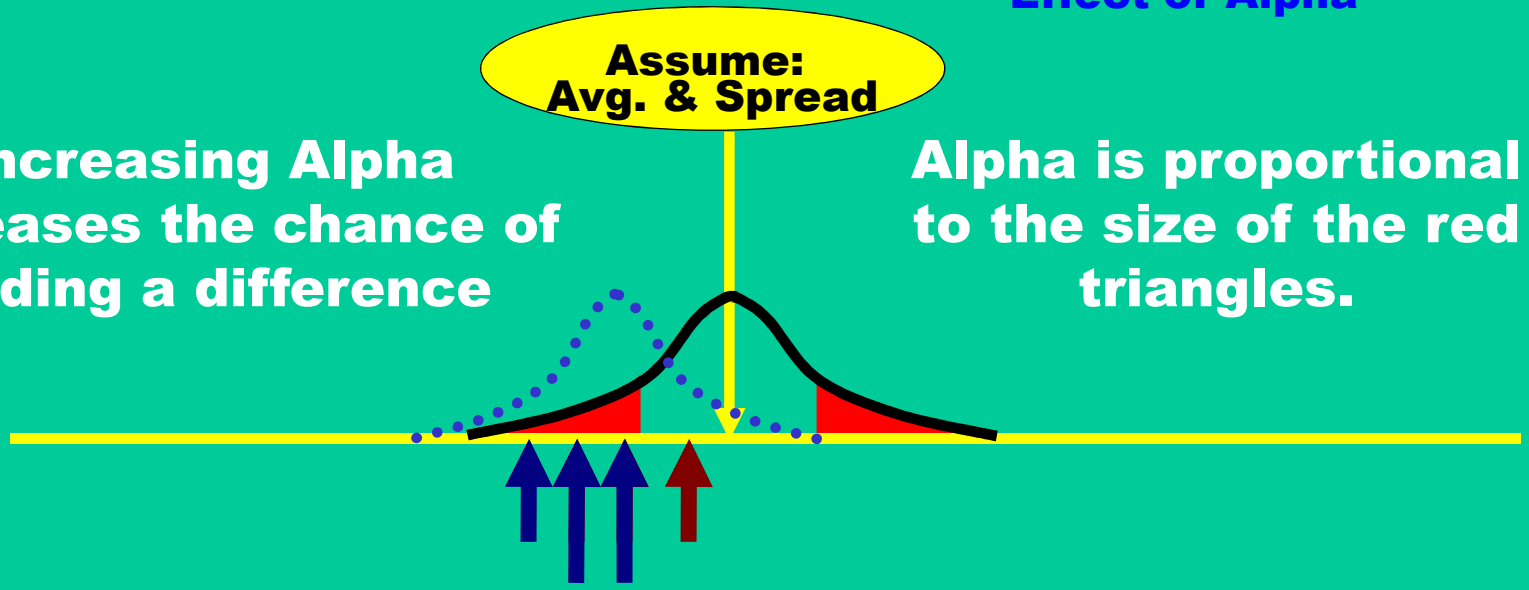
The Issue is..

under the conditions pictured above, how likely are you to get an observation outside the red triangles.

With this blue distribution, what is the probability of detecting that the average has changed.

What is Statistical Power?

- Intro to power: Conditions
- Effect of Alpha



Increasing Alpha Increases the chance of finding a difference

Alpha is proportional to the size of the red triangles.

Power is the percent of the area under the blue curve *NOT* between the red triangles

The Issue is..

under the conditions pictured above, how likely are you to get an observation outside the red triangles.

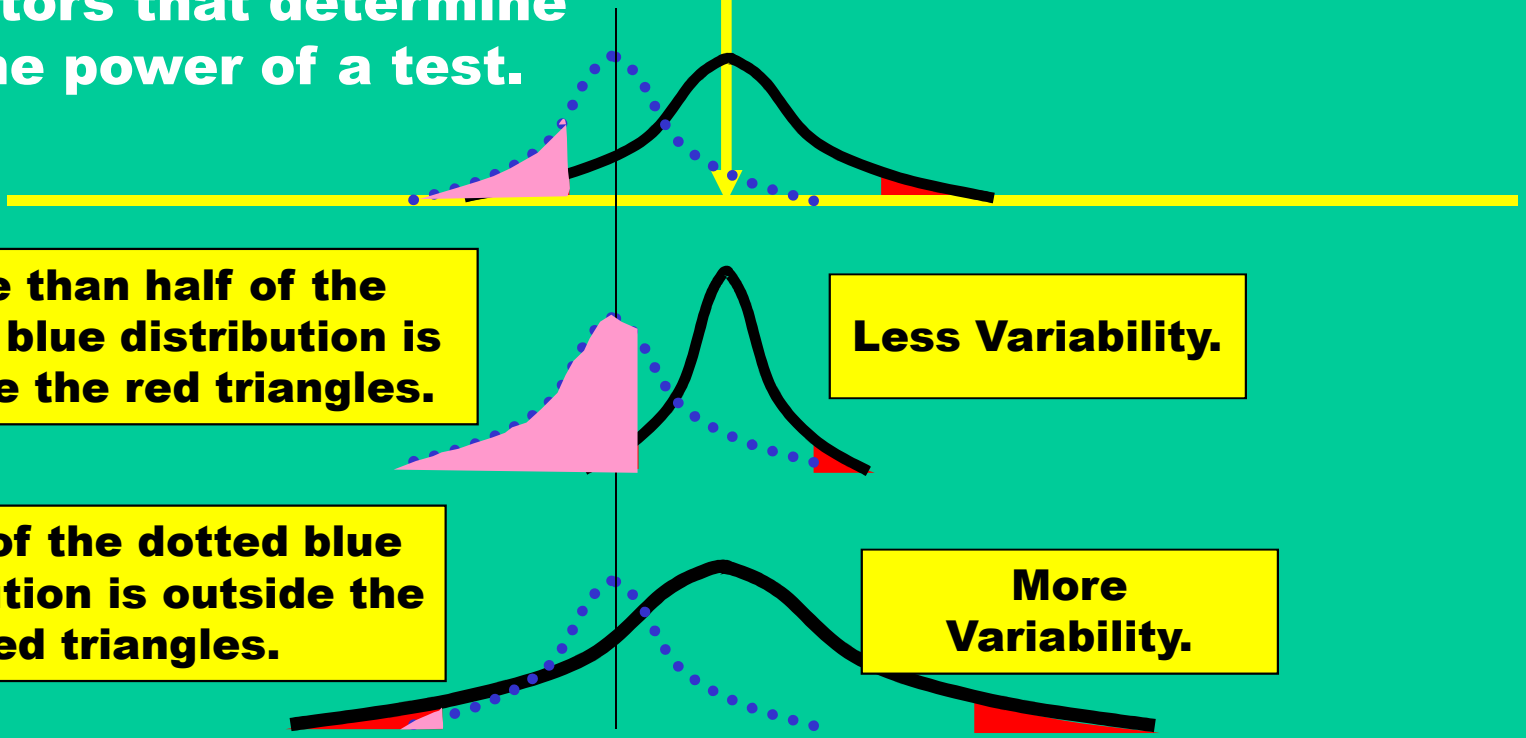
With this blue distribution, what is the probability of detecting that the average has changed.

What is Statistical Power?

- Intro to power: Conditions
- Variability

Assume:
Avg. & Spread

Variability is one of the factors that determine the power of a test.



More than half of the dotted blue distribution is outside the red triangles.

Less Variability.

Little of the dotted blue distribution is outside the red triangles.

More Variability.

The more variability in your data,
the more overlap you will have between distributions,
....the farther out the red triangles are
.....and the lower the power.

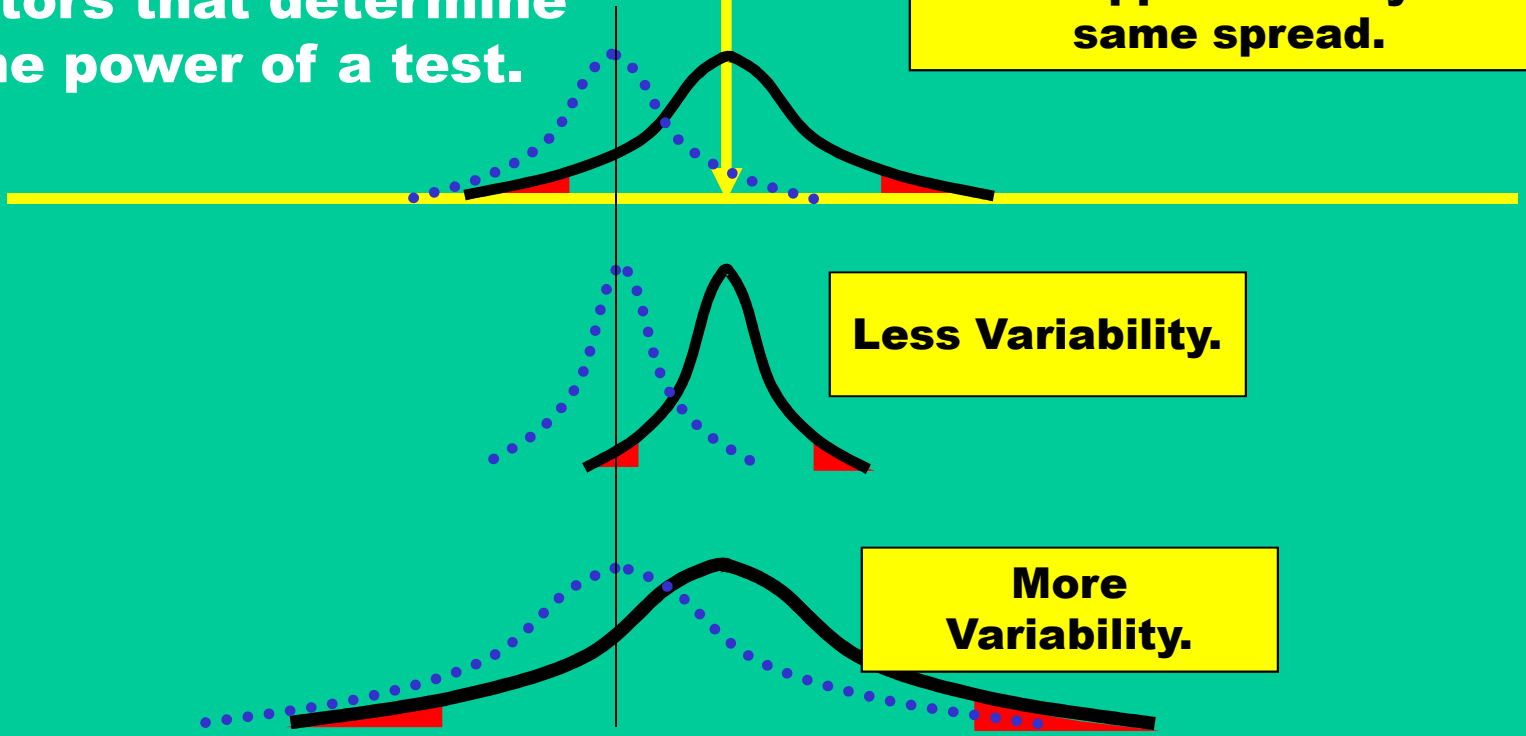
What is Statistical Power?

- Intro to power: Conditions
- Variability

Assume: Avg. & Spread

Variability is one of the factors that determine the power of a test.

Often, both distributions have approximately the same spread.



Less Variability.

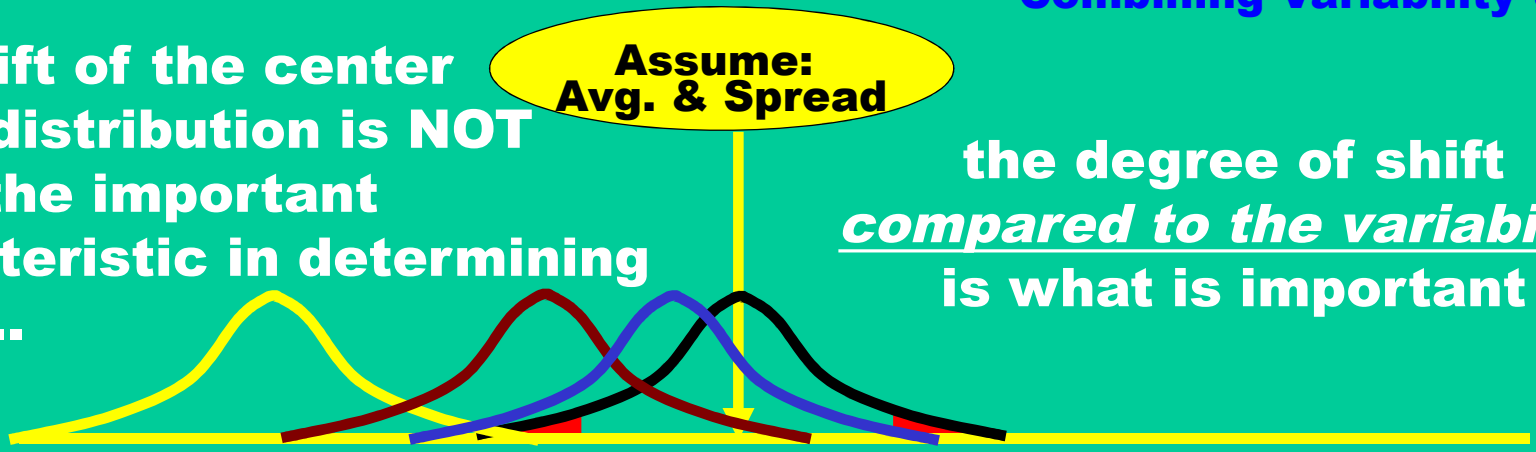
More Variability.

The more variability in your data,
the more overlap you will have between distributions,
....the farther out the red triangles are
.....and the lower the power.

What is Statistical Power?

- Intro to power: Conditions
- Combining Variability &shift

The shift of the center of the distribution is NOT really the important characteristic in determining power...



the degree of shift *compared to the variability* is what is important

If the degree of shift is large enough so that the distributions do not overlap, ...that shift would be easy to detect.

The blue distribution illustrates a small shift (lots of overlap) with the black (Control Distribution) and would be difficult to detect.

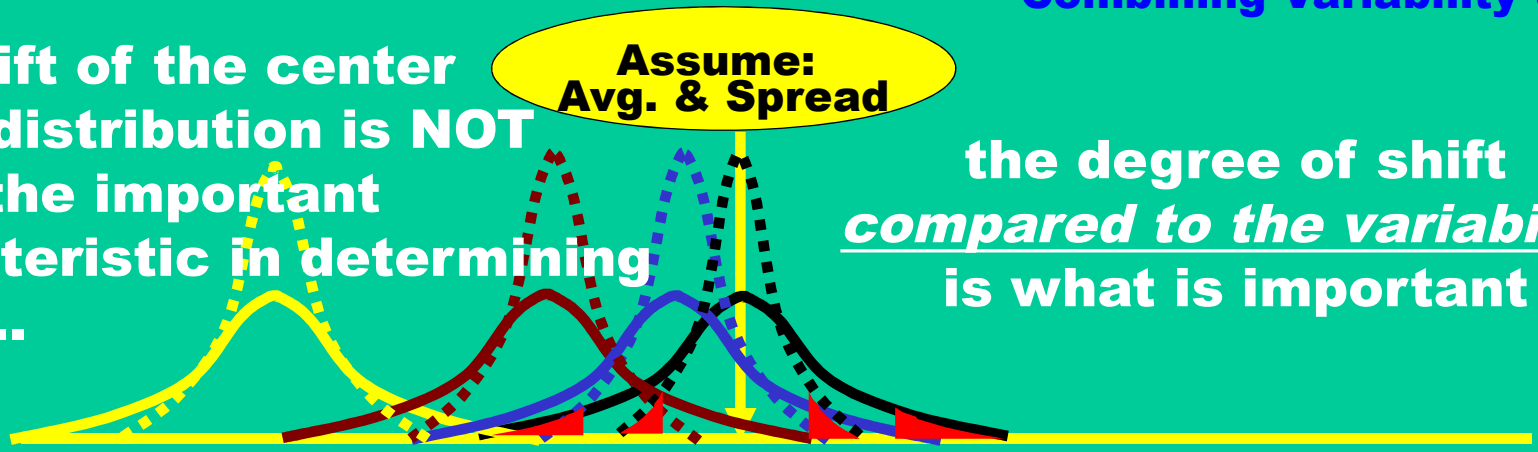
The gold distribution has little overlap with the black (Control Distribution) and would be easy to detect.

The red distribution illustrates a moderate effect and is where we want to concentrate.

What is Statistical Power?

- Intro to power: Conditions
- Combining Variability &shift

The shift of the center of the distribution is NOT really the important characteristic in determining power...



the degree of shift
compared to the variability
is what is important

If we could reduce the overlap, we could have a more powerful test.

A larger shift would reduce the overlap.

But the strength of the drug determines the degree of shift. We can only change shift by inventing a new drug.

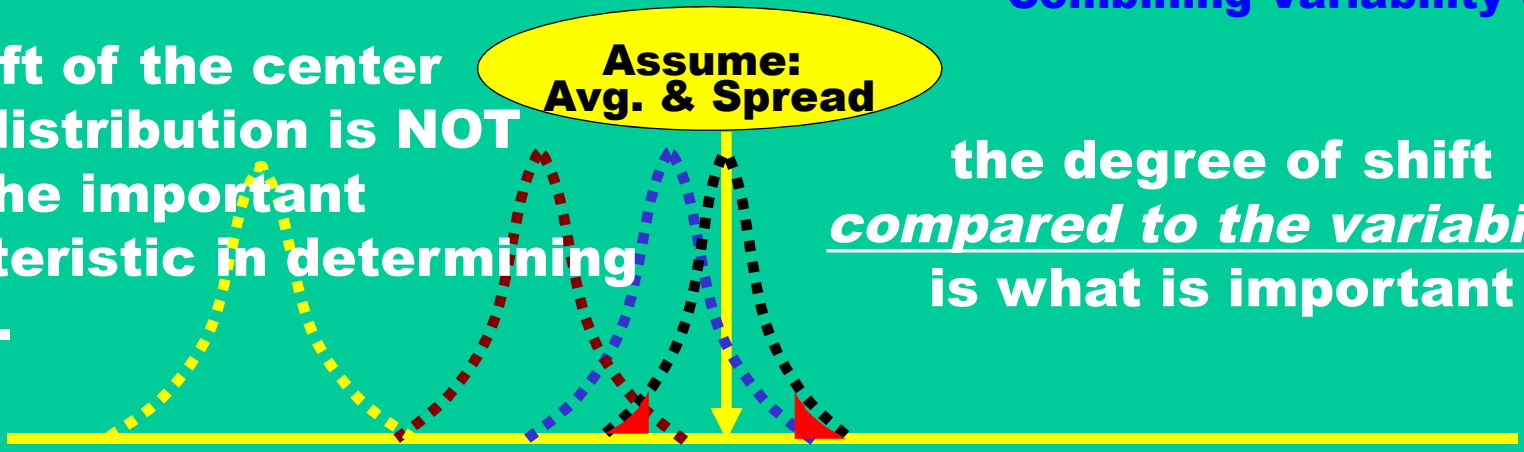
Lets clean things up a bit.

If we could make the distributions skinnier, we could reduce the overlap.

What is Statistical Power?

- Intro to power: Conditions
- Combining Variability &shift

The shift of the center of the distribution is NOT really the important characteristic in determining power...



the degree of shift *compared to the variability* is what is important

If we could reduce the overlap, we could have a more powerful test.

A larger shift would reduce the overlap.

But the strength of the drug determines the degree of shift. We can only change shift by inventing a new drug.

Lets clean things up a bit.

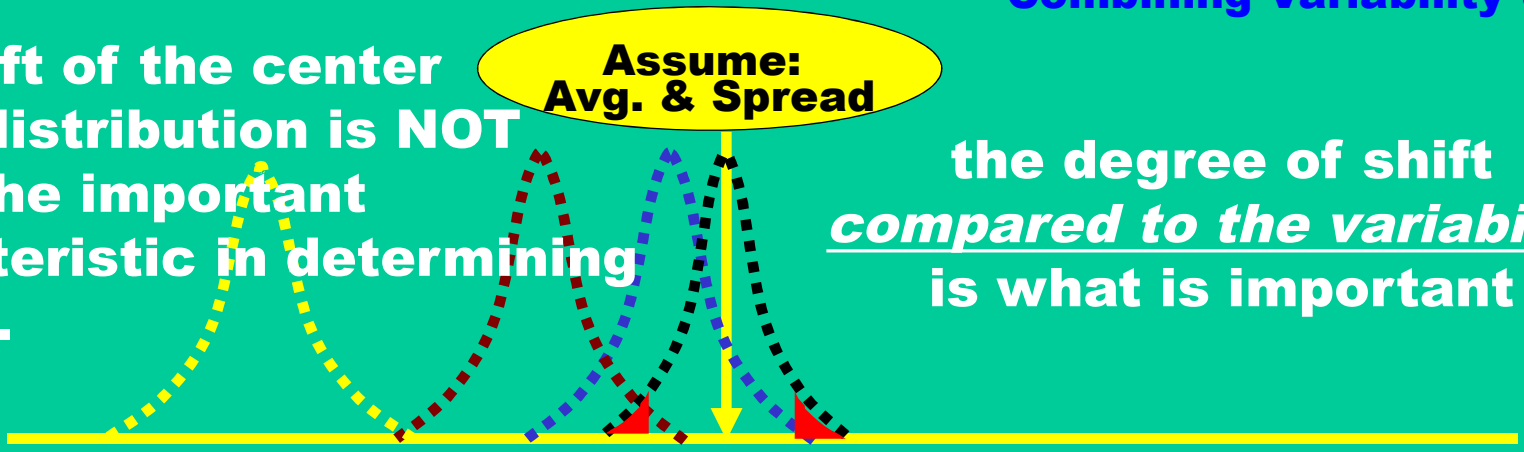
If we could make the distributions skinnier, we could reduce the overlap.

The red curve now has little overlap with the Black

What is Statistical Power?

- Intro to power: Conditions
- Combining Variability &shift

The shift of the center of the distribution is NOT really the important characteristic in determining power...



the degree of shift
compared to the variability
is what is important

If we could reduce the overlap, we could have a more powerful test.

A larger shift would reduce the overlap.

But the strength of the drug determines the degree of shift. We can only change shift by inventing a new drug.

Lets clean things up a bit.

If we could make the distributions skinnier, we could reduce the overlap.

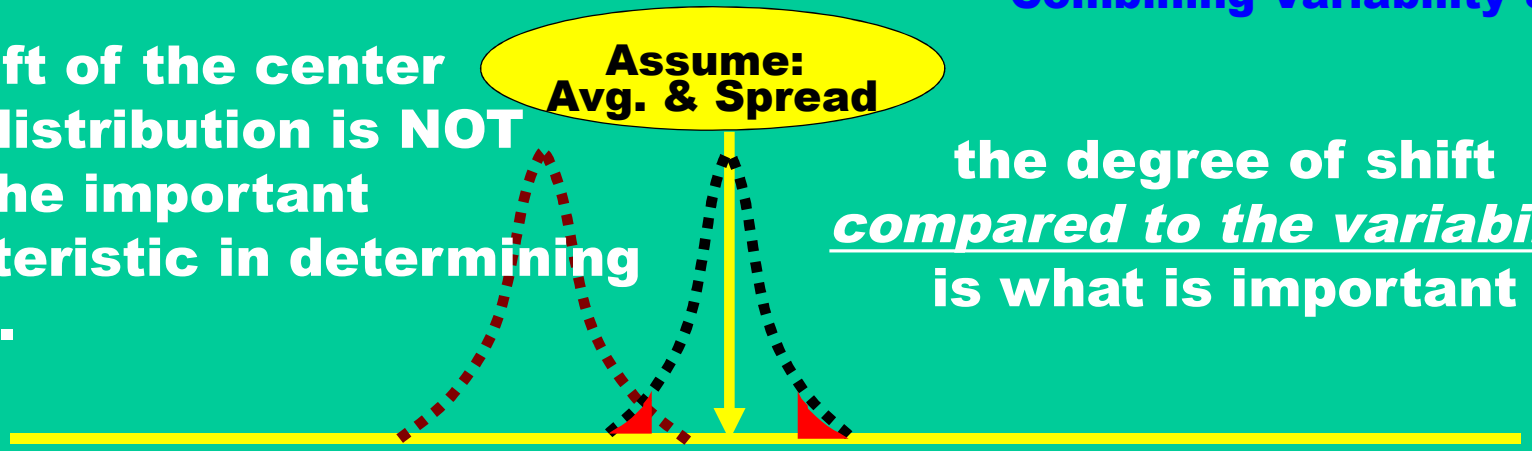
The red curve now has little overlap with the Black

And we CAN make the distributions skinnier by increasing the sample size.

What is Statistical Power?

- Intro to power: Conditions
- Combining Variability &shift

The shift of the center of the distribution is NOT really the important characteristic in determining power...



the degree of shift
compared to the variability
is what is important

If we could reduce the overlap, we could have a more powerful test.

A larger shift would reduce the overlap.

But the strength of the drug determines the degree of shift. We can only change shift by inventing a new drug.

Sweet.

If we could make the distributions skinnier, we could reduce the overlap.

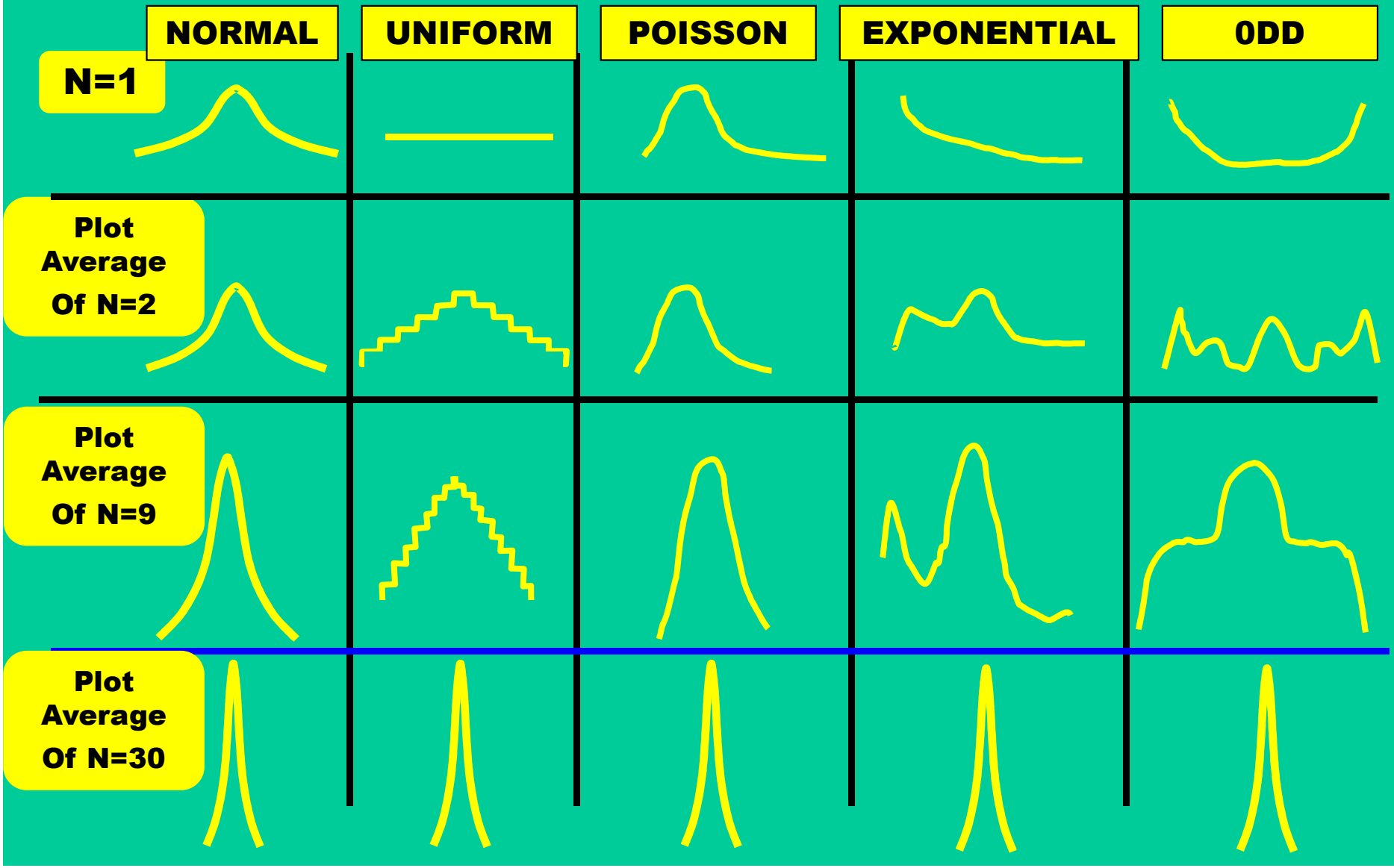
The red curve now has little overlap with the Black

And we CAN make the distributions skinnier by increasing the sample size.

What is Statistical Power?

- Intro to power: Conditions
- The central limit theorem

Increasing N does not change the average and makes distributions get narrower and mound shaped



What is Statistical Power?

The Conditions of Power Calculations

- Power depends on the **degree of shift**
  (large shifts are easy to detect)
- Power depends on the **alpha error you select**
- Power depends on the **variability** in the process
  (lots of variability means low power)
- Power depends on the **number of observations**
  (more observations help detect shifts)
- Power depends on the **test** used
  (Some tests are more powerful than others)

What is Statistical Power?

•The central limit theorem

Why increasing N makes dis et narrow:

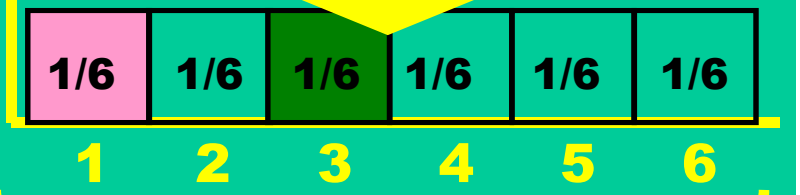
If n=,1 the distribution of probabilities is flat at p= .166

1/6 or .166

Prob of the event

Average=

(1+2+3+4+5+6)/6
=21/6 = 3.5



1/6 * 6 = 1.00

Probability
,for equally likely outcomes,
is the number of ways an event could occur
divided by the number of possible events.

The sum of the probability of all possible events is 1.00.

What is Statistical Power?

•The central limit theorem

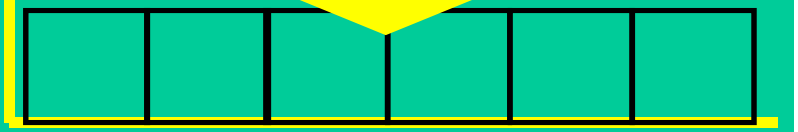
Why increasing N makes dis… narrow:

Average=

(1+2+3+4+5+6)/6
=21/6 = 3.5

1/6 or .166

Prob of the event



Prob. distrib. is flat at N=1
Mound & "squshed" at N=2

	1+1= 2	1+2= 3	1+3= 4	1+4= 5	1+5= 6	1+6= 7
	2+1= 3	2+2= 4	2+3= 5	2+4= 6	2+5= 7	2+6= 8
	3+1= 4	3+2= 5	3+3= 6	3+4= 7	3+5= 8	3+6= 9
	4+1= 5	4+2= 6	4+3= 7	4+4= 8	4+5= 9	4+6= 10
	5+1= 6	5+2= 7	5+3= 8	5+4= 9	5+5= 10	5+6= 11
	6+1= 7	6+2= 8	6+3= 9	6+4= 10	6+5= 11	6+6= 12

Doubling the size and analyzing the average (N=2)makes the distribution narrower, but doubles cost.

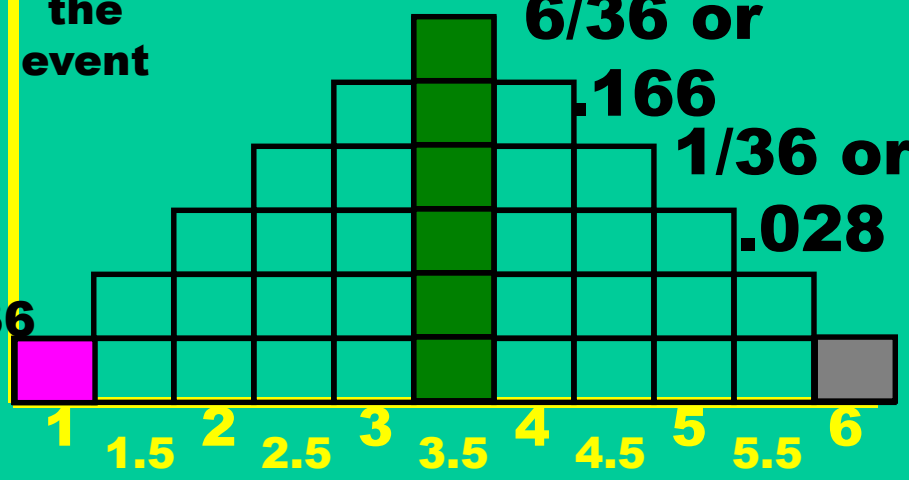
Prob of the event

Average= 3.5

6/36 or .166

1/36 or .028

1/36



What is Statistical Power?

Why increasing N makes dis et narrow:

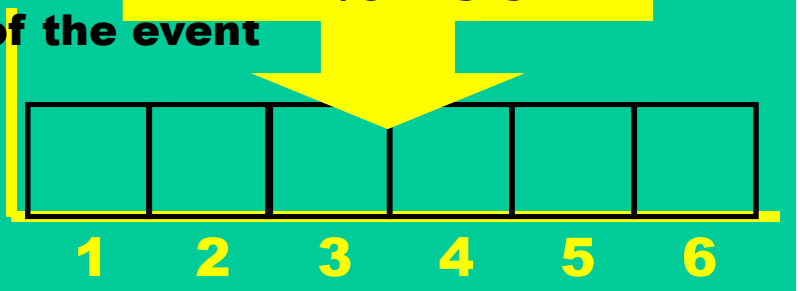
- The central limit theorem

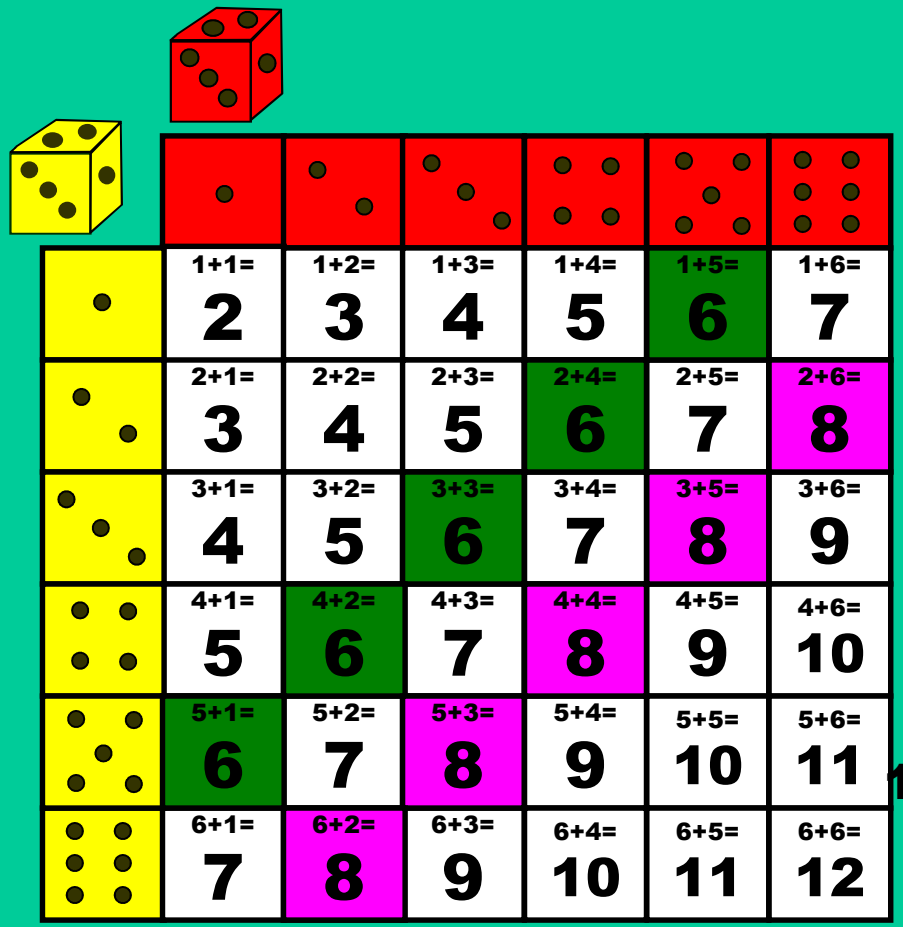
Average=

(1+2+3+4+5+6)/6
=21/6 = 3.5

1/6 or .166

Prob of the event





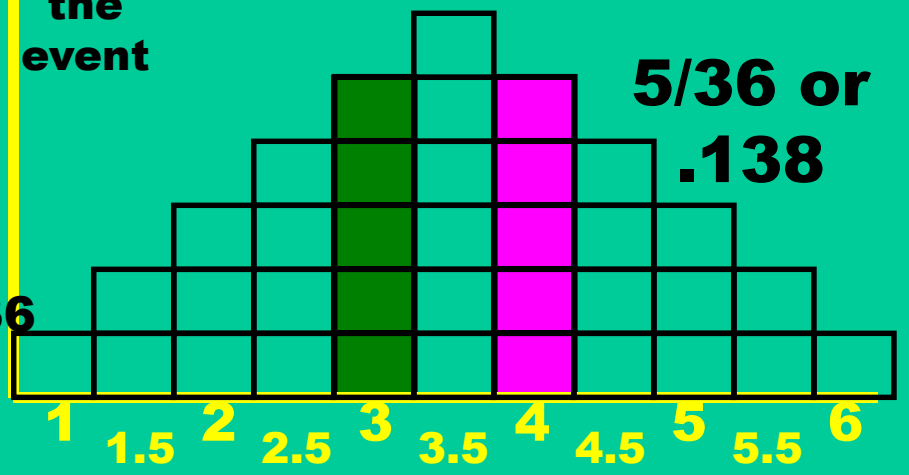
1+1= 2	1+2= 3	1+3= 4	1+4= 5	1+5= 6	1+6= 7
2+1= 3	2+2= 4	2+3= 5	2+4= 6	2+5= 7	2+6= 8
3+1= 4	3+2= 5	3+3= 6	3+4= 7	3+5= 8	3+6= 9
4+1= 5	4+2= 6	4+3= 7	4+4= 8	4+5= 9	4+6= 10
5+1= 6	5+2= 7	5+3= 8	5+4= 9	5+5= 10	5+6= 11
6+1= 7	6+2= 8	6+3= 9	6+4= 10	6+5= 11	6+6= 12

Prob of the event

Average= 3.5

5/36 or .138

1/36



What is Statistical Power?

Why increasing N makes dis… narrow:

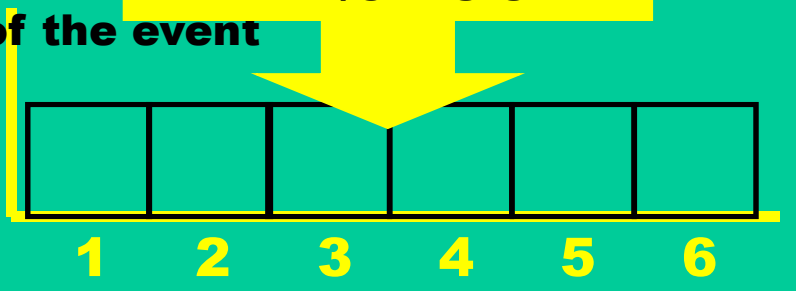
- The central limit theorem

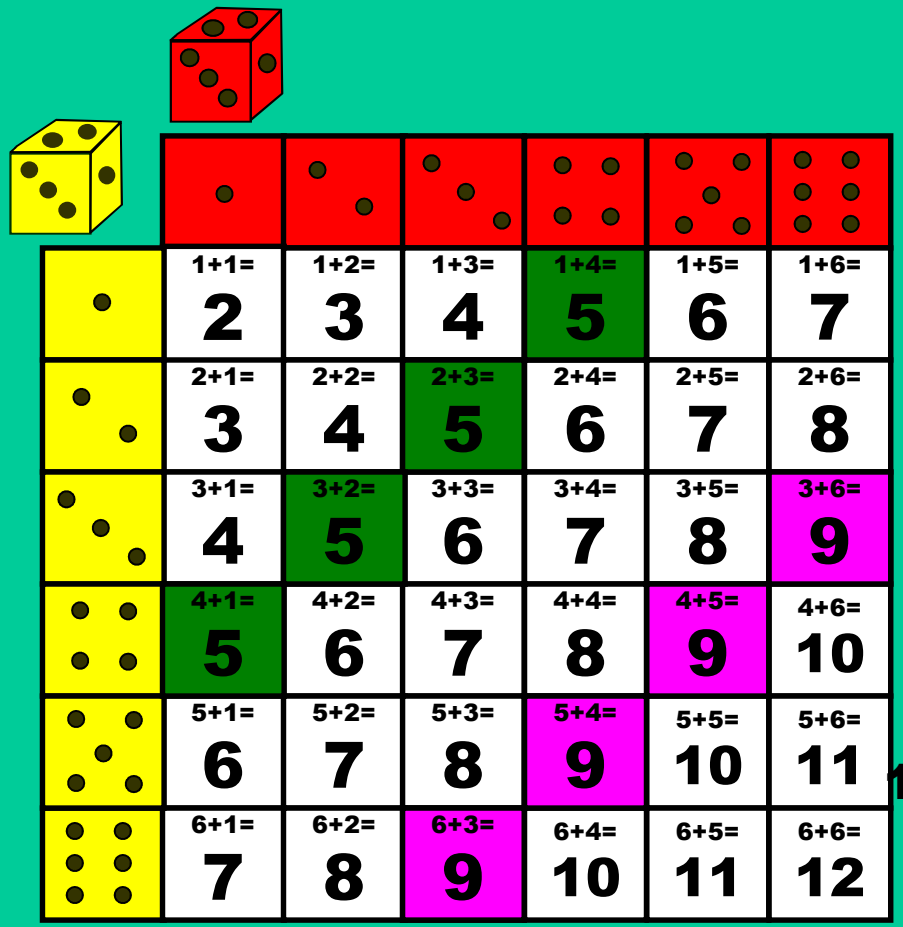
Average=

(1+2+3+4+5+6)/6
=21/6 = 3.5

1/6 or .166

Prob of the event





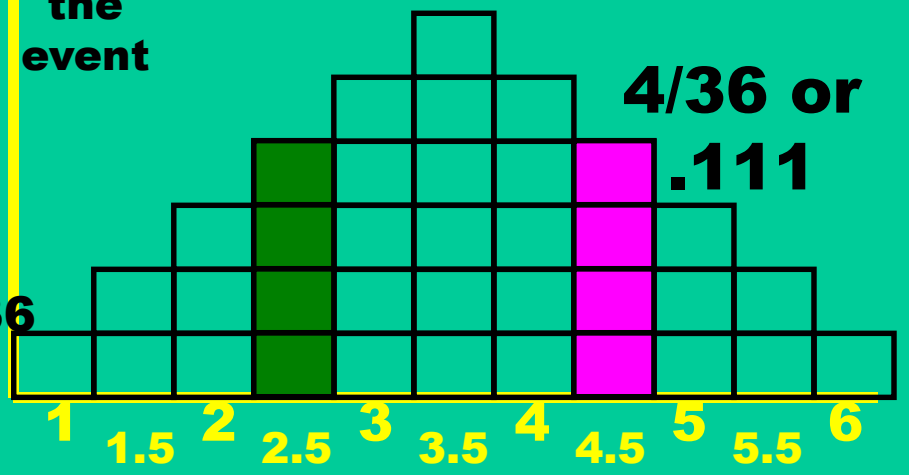
1+1= 2	1+2= 3	1+3= 4	1+4= 5	1+5= 6	1+6= 7
2+1= 3	2+2= 4	2+3= 5	2+4= 6	2+5= 7	2+6= 8
3+1= 4	3+2= 5	3+3= 6	3+4= 7	3+5= 8	3+6= 9
4+1= 5	4+2= 6	4+3= 7	4+4= 8	4+5= 9	4+6= 10
5+1= 6	5+2= 7	5+3= 8	5+4= 9	5+5= 10	5+6= 11
6+1= 7	6+2= 8	6+3= 9	6+4= 10	6+5= 11	6+6= 12

Prob of the event

Average= 3.5

4/36 or .111

1/36

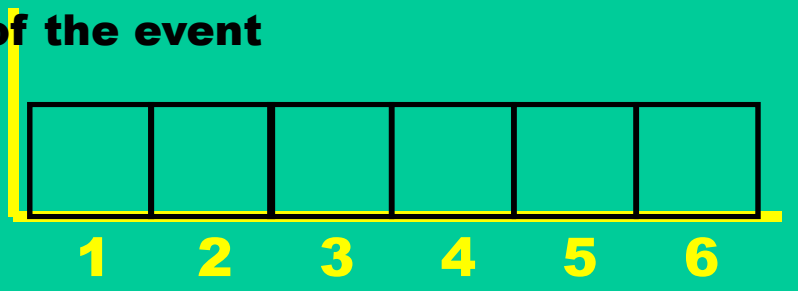


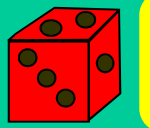
What is Statistical Power?

•The central limit theorem

Why increasing N makes distributions get narrow:

1/6 or .166

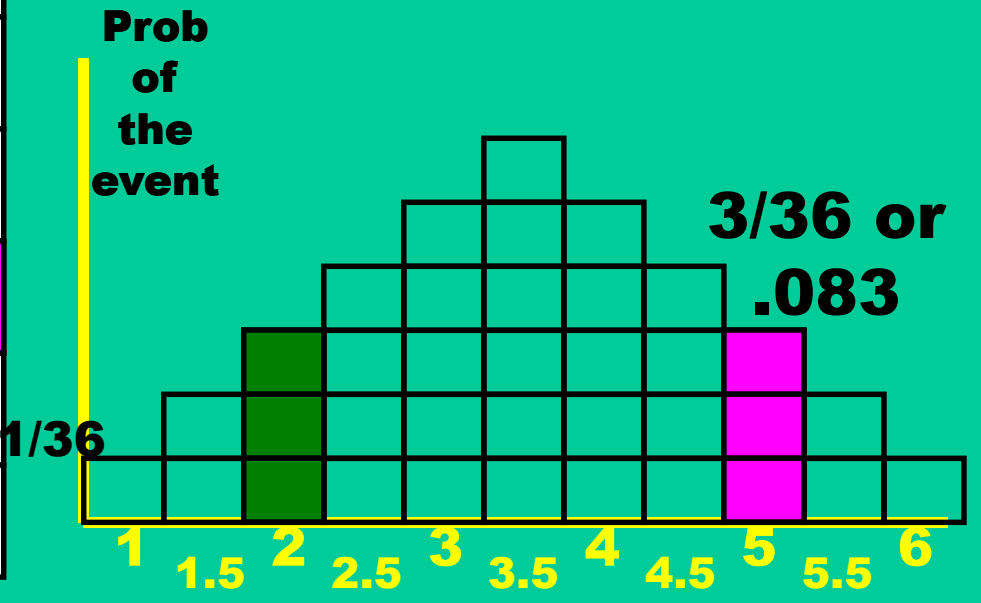




Same average, but narrower and more mound shaped

	1+1= 2	1+2= 3	1+3= 4	1+4= 5	1+5= 6	1+6= 7
	2+1= 3	2+2= 4	2+3= 5	2+4= 6	2+5= 7	2+6= 8
	3+1= 4	3+2= 5	3+3= 6	3+4= 7	3+5= 8	3+6= 9
	4+1= 5	4+2= 6	4+3= 7	4+4= 8	4+5= 9	4+6= 10
	5+1= 6	5+2= 7	5+3= 8	5+4= 9	5+5= 10	5+6= 11
	6+1= 7	6+2= 8	6+3= 9	6+4= 10	6+5= 11	6+6= 12

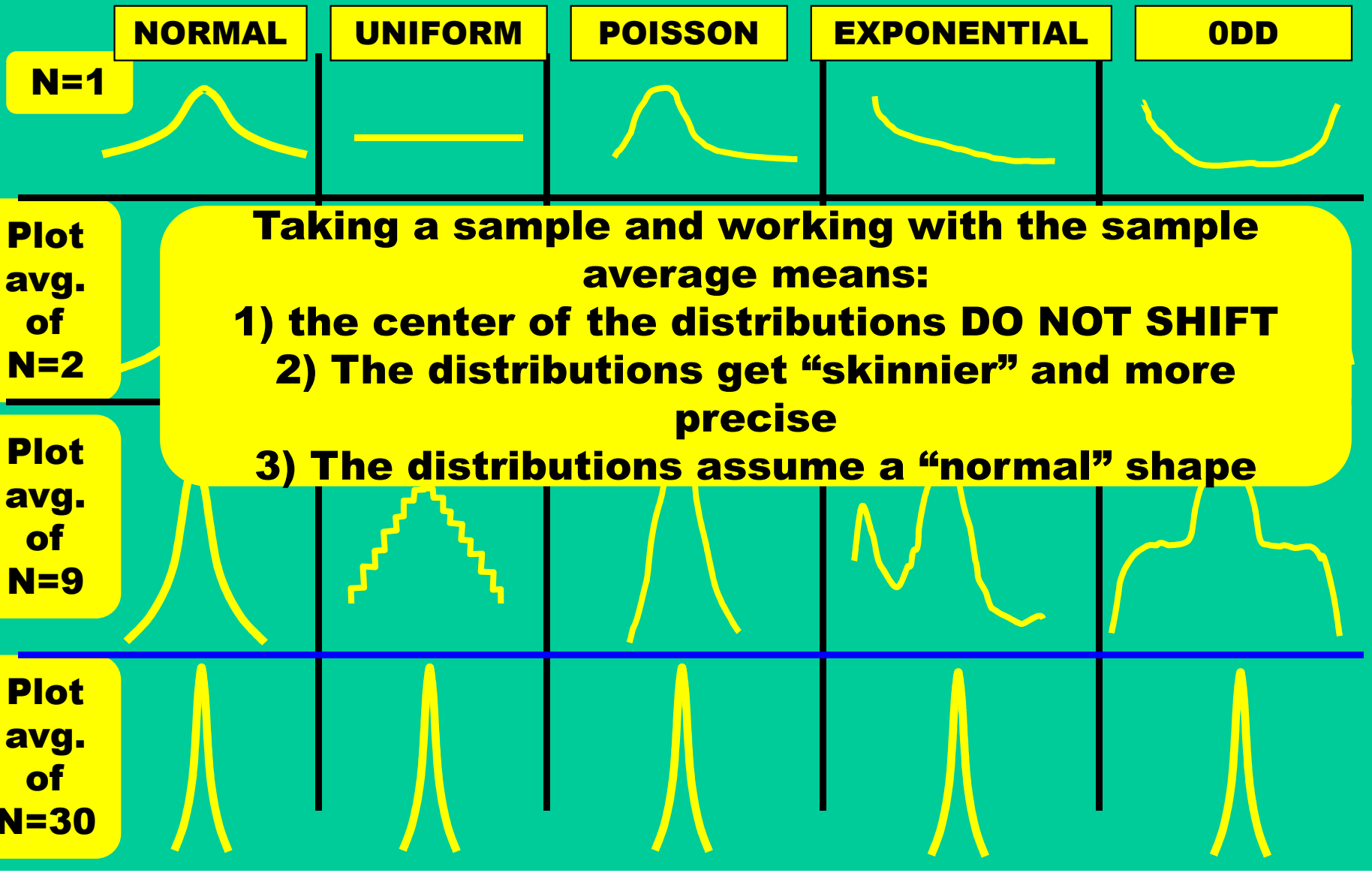
Prob. distrib. is flat at N=1
Mound & "squshed" at N=2



What is Statistical Power?

•The central limit theorem

Increasing N does not change the average and makes distributions get narrower and mound shaped



Taking a sample and working with the sample average means:

1) the center of the distributions DO NOT SHIFT

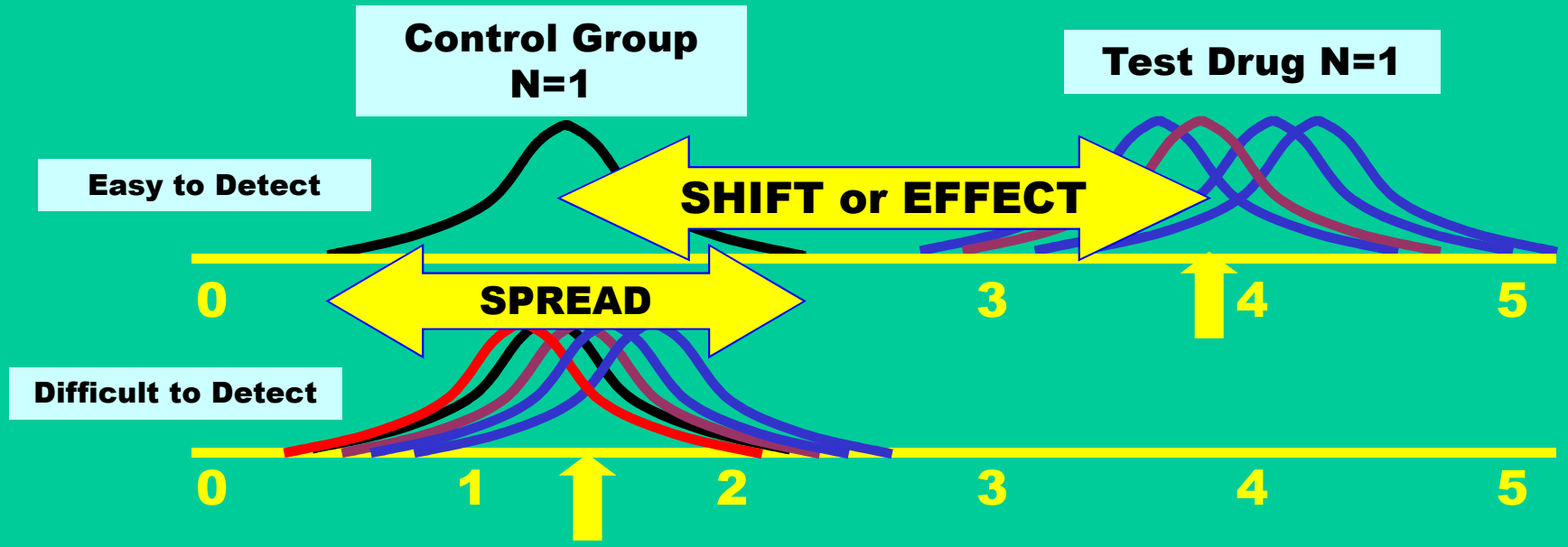
2) The distributions get "skinnier" and more precise

3) The distributions assume a "normal" shape

What is Statistical Power?

•Integrating variability, shift & N

- Power is a shorthand for the phrase
  "Power of a test to detect an effect/shift"

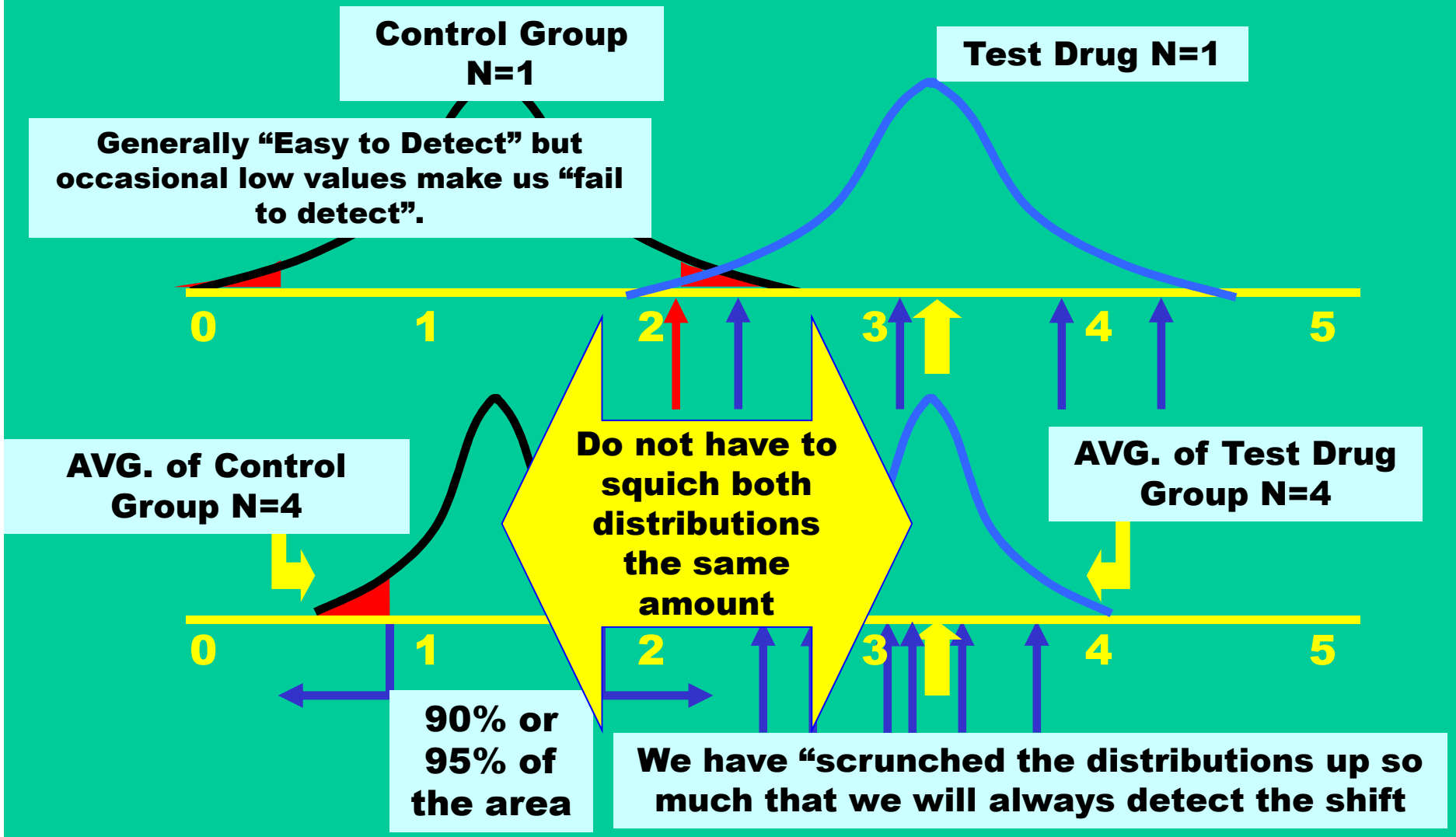


- The ability of a test to detect an effect depends on:
  - The statistical test itself *(not covered)*
  - The number of subjects used (increasing sample size squishes)
  - The original spread in the data (the variance)
  - The size of the effect (difference in means or drug strength)
  - How unusual an event must be before "concluding a change"
    The Alpha Level

What is Statistical Power?

•Integrating variability, shift & N

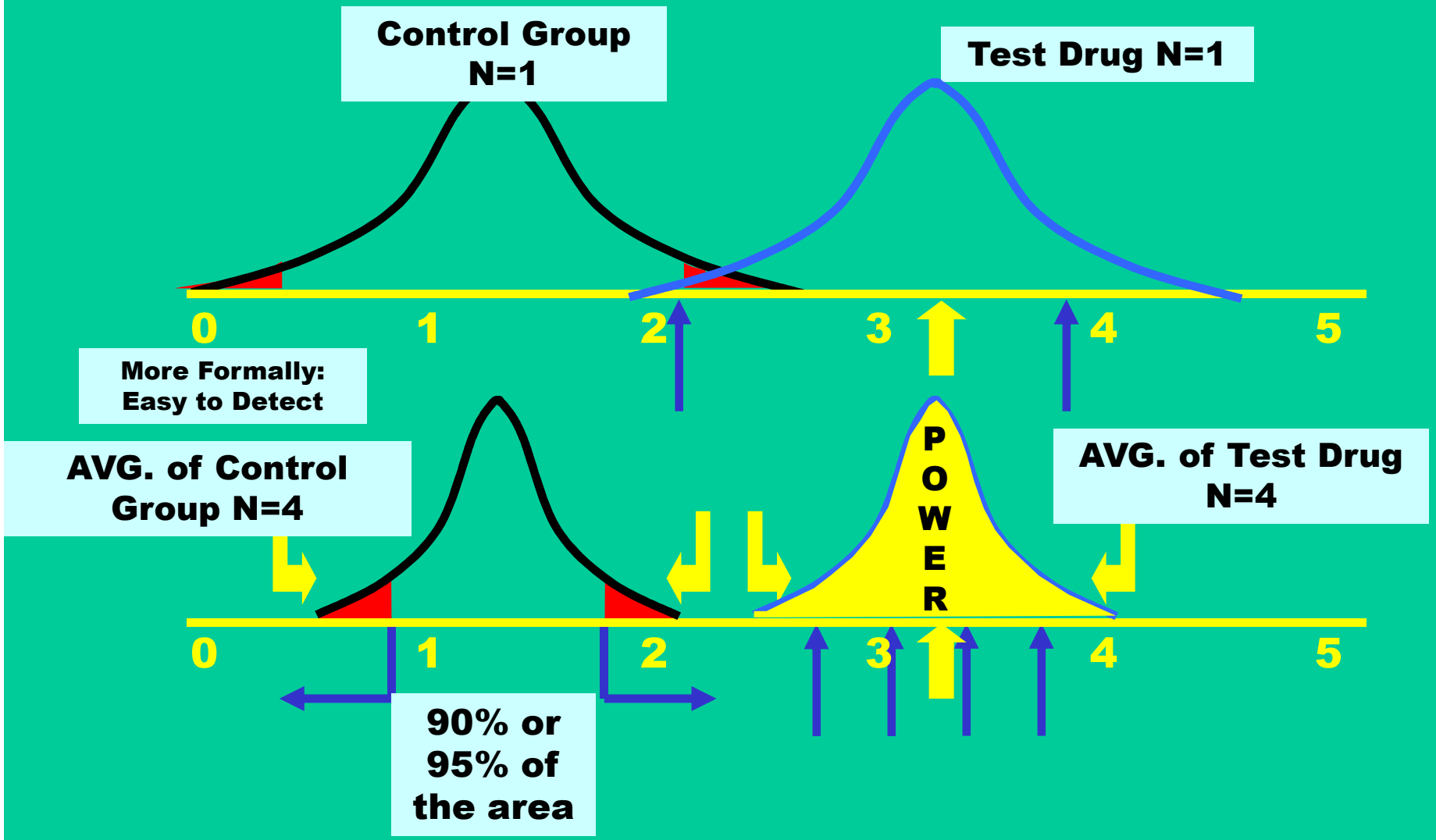
•Example 1: increasing N to squich



What is Statistical Power?

•Integrating variability, shift & N

•Example 1: increasing N to squich

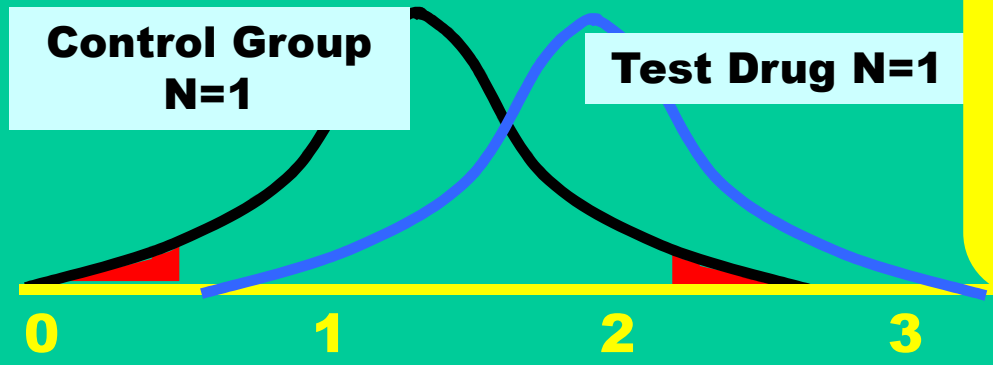


What is Statistical Power?

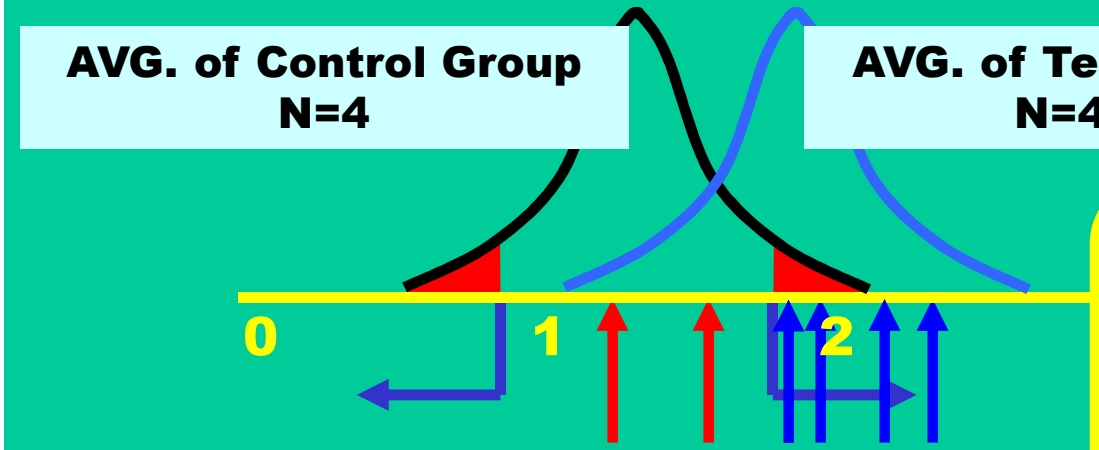
•Integrating variability, shift & N

- **Example 2: increasing N to squich**

  Lets look at a less effective drug- less shift



With n=1, too much of the blue distribution is in the red region. We will rarely reject Ho! Try n=4



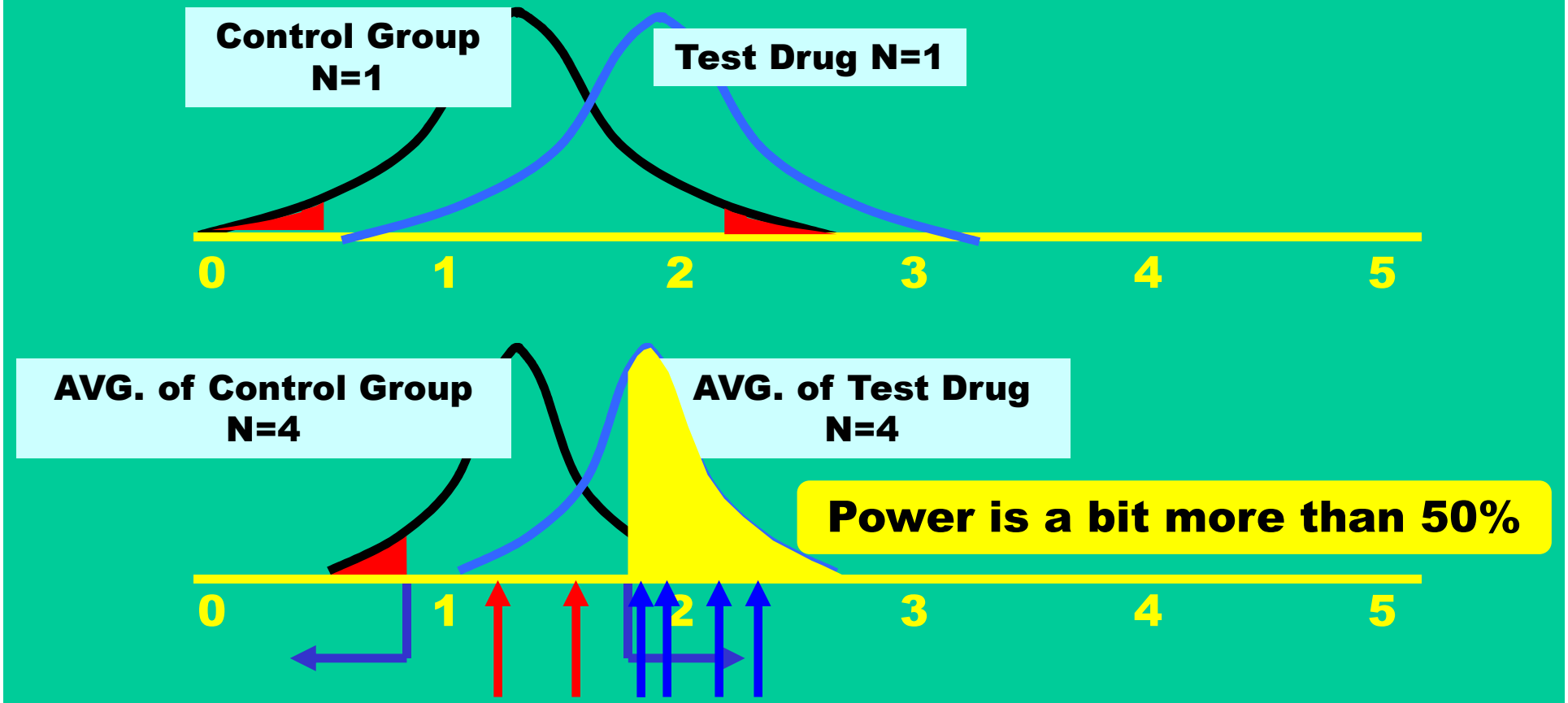
With (N=4) MANY of the sample averages are inside the CI. We see that small shifts are hard to detect even though increasing N narrowed the distribution

What is Statistical Power?

•Integrating variability, shift & N

- Example 2: increasing N to squich

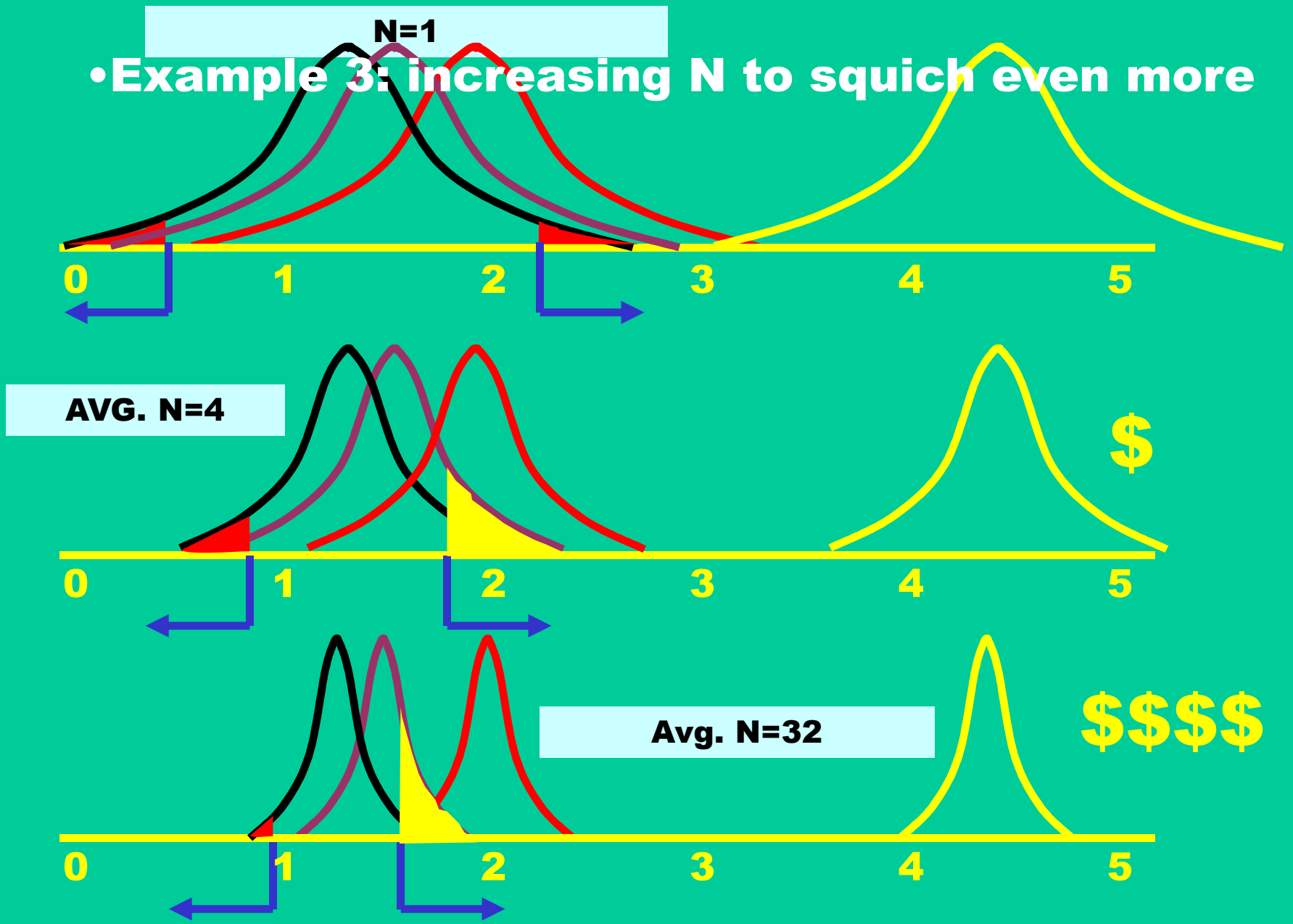
  Lets look at a less effective drug- less shift



What is Statistical Power?

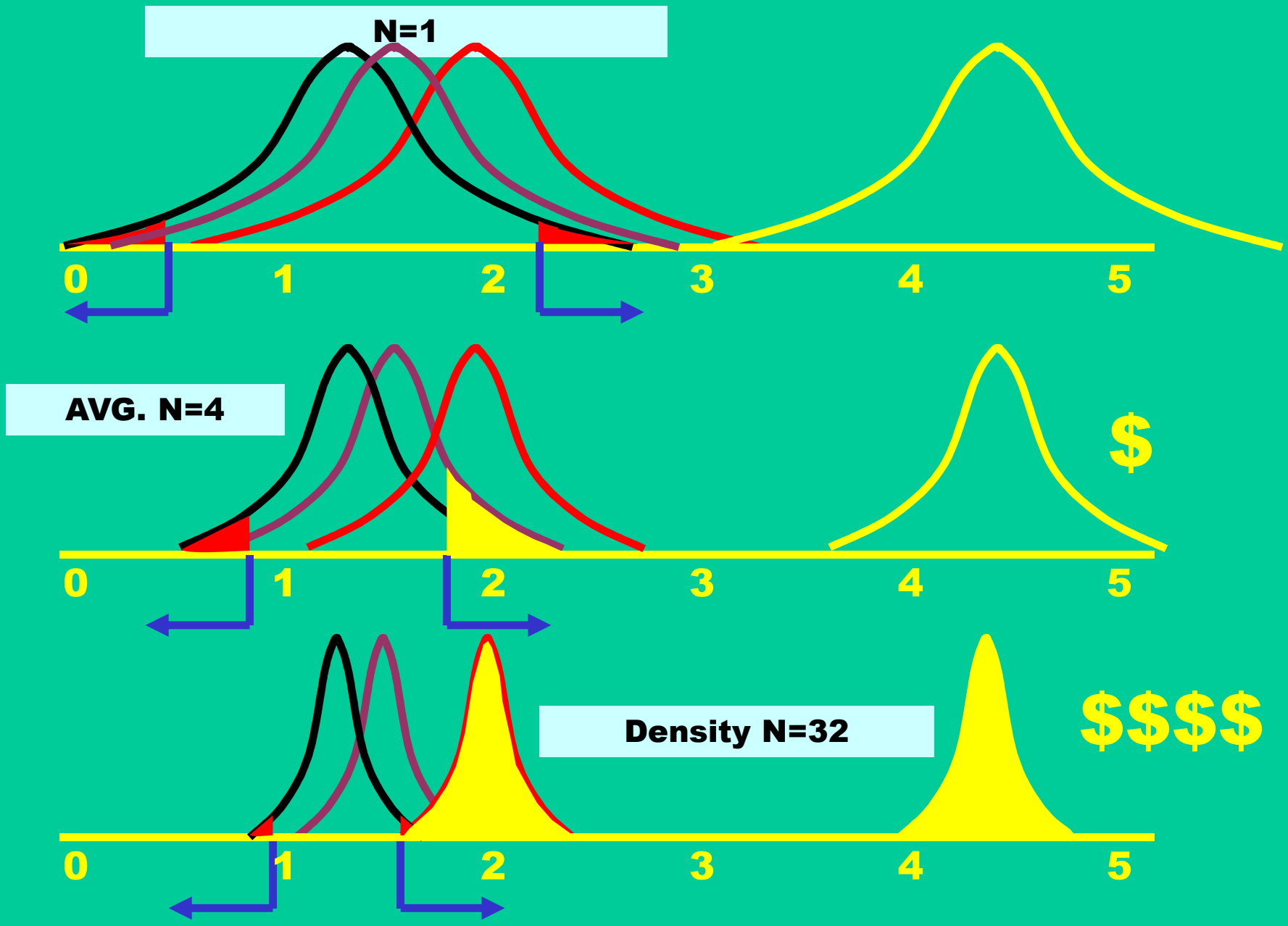
•Integrating variability, shift & N

•Example 3: increasing N to squich even more



What is Statistical Power?

•Integrating variability, shift & N

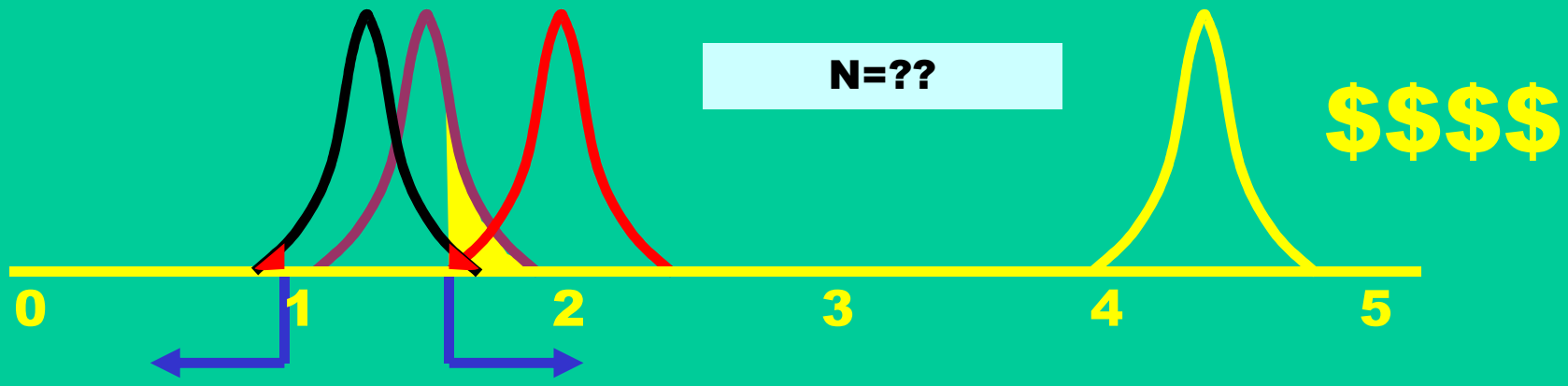


What is Statistical Power?

•Integrating variability, shift & N

- **The gold and red experiments have too much power**
  - Waste money
    - Inconvenience more subjects than is required
      - Delay introduction of drug to sufferers / market

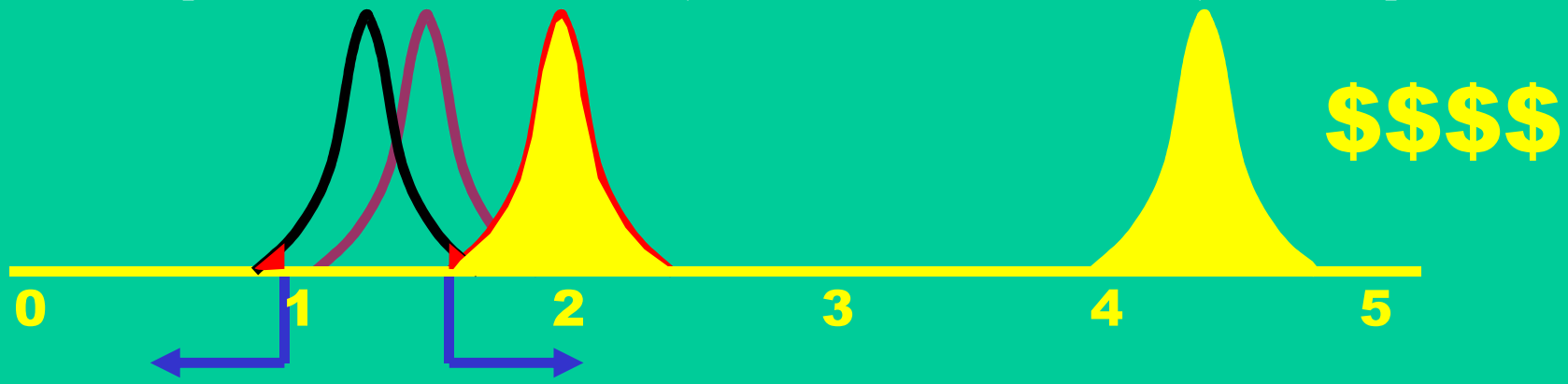
- **The purple experiment is under-powered.**
  - There is an effect, but it will likely NOT be found
    - A negative result will disrupt the research stream
      - Delay introduction of drug to sufferers / market
        - Inconvenience subjects for no good reason



What is Statistical Power?

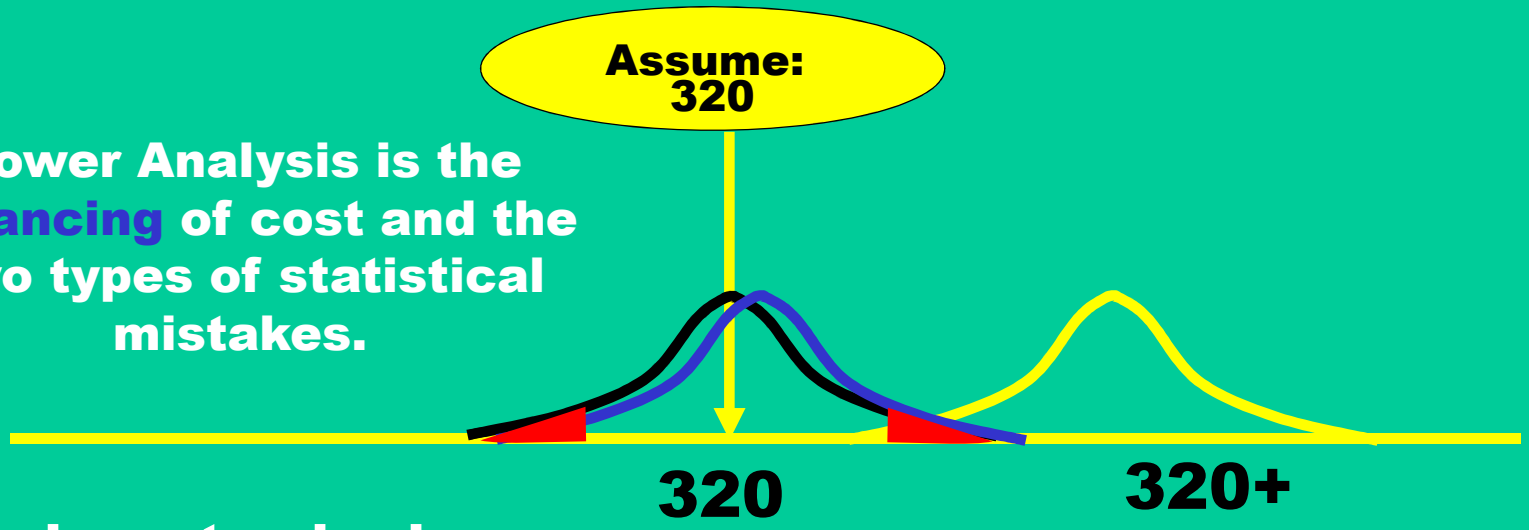
•Integrating variability, shift & N

- We usually can not change the effect size
  and usually only have an **estimate** of effect size
- We usually can not change the spread
  and usually only have an **estimate** of spread
- For many reasons, subjects leave the study
  and we can only estimate end-of-study N
- The above points force us to do repeated
  "what if calculations" to calculate power
  Formulas are complex.
- Computer assistance, like SAS PROCS, is required



Review of Logic:

•Balancing cost and errors



Power Analysis is the **Balancing** of cost and the two types of statistical mistakes.

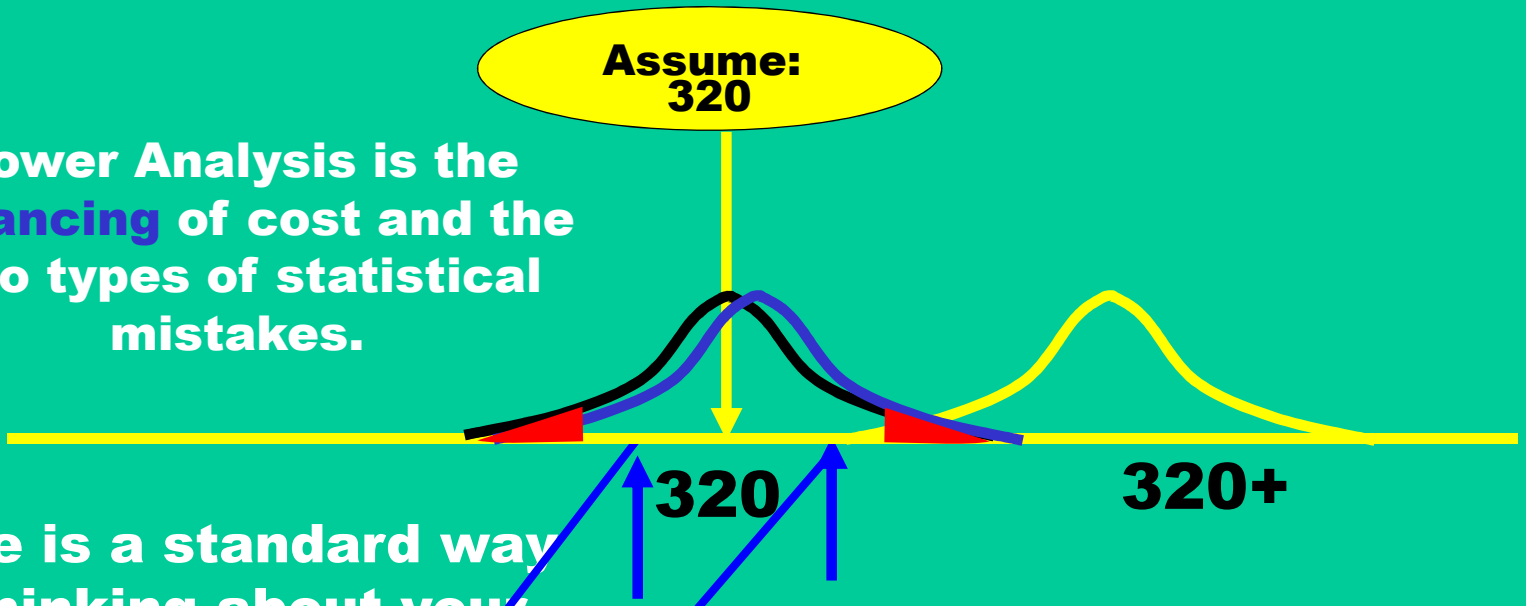
There is a standard way of thinking about your decision.

Assume Ho
Mean = 320

	Ho=True	Ho=False
Assume HO & kill the Drug	OK Decision	Beta Error
Reject Ho & market the Drug	Alpha Error	OK Decision

Review of Logic:

•Balancing cost and errors



Power Analysis is the **Balancing** of cost and the two types of statistical mistakes.

There is a standard way of thinking about your decision.

Assume Ho
Mean = 320

	Ho=True	Ho=False
Assume HO & kill the Drug	OK Decision	Beta Error
Reject Ho & market the Drug	Alpha Error	OK Decision

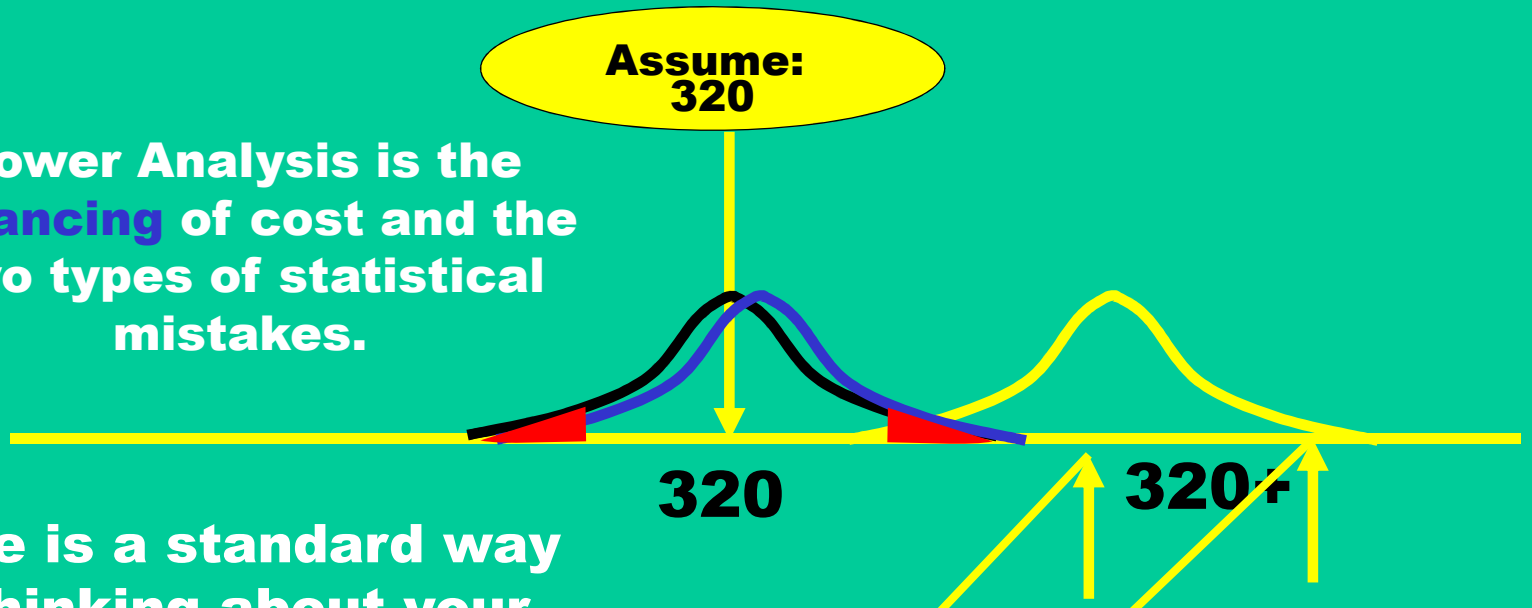
Samples come from the center of the Blue distribution.

We will call ths “blue” shift “unimportant” – it is close to zero. The drug is Ineffective!

These are is not mistakes! Good Decisions!

Review of Logic:

•Balancing cost and errors



Power Analysis is the **Balancing** of cost and the two types of statistical mistakes.

There is a standard way of thinking about your decision.

Assume Ho
Mean = 320

	Ho=True	Ho=False
Assume HO & kill the Drug	OK Decision	Beta Error
Reject Ho & market the Drug	Alpha Error	OK Decision

Samples come from the Center/Right of the Gold distribution.

The drug is Effective!

These are is not mistakes! Good Decisions!

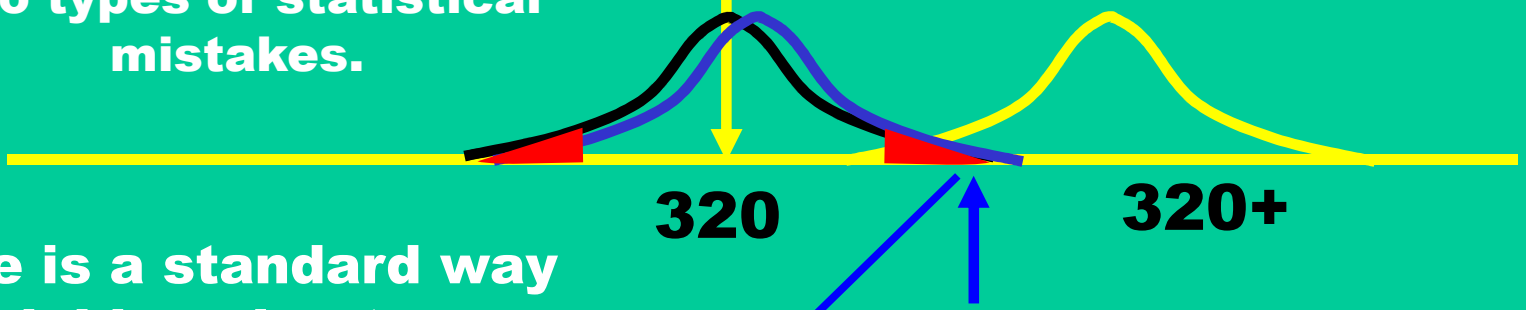
Review of Logic:

•Balancing cost and errors

Assume: 320

Rejecting a true Ho is an alpha error

Power Analysis is the **Balancing** of cost and the two types of statistical mistakes.



There is a standard way of thinking about your decision.

Assume Ho
Mean = 320

	Ho=True	Ho=False
Assume HO & kill the Drug	OK Decision	Beta Error
Reject Ho & market the Drug	Alpha Error	OK Decision

Samples come from the Right end of the Blue distribution.

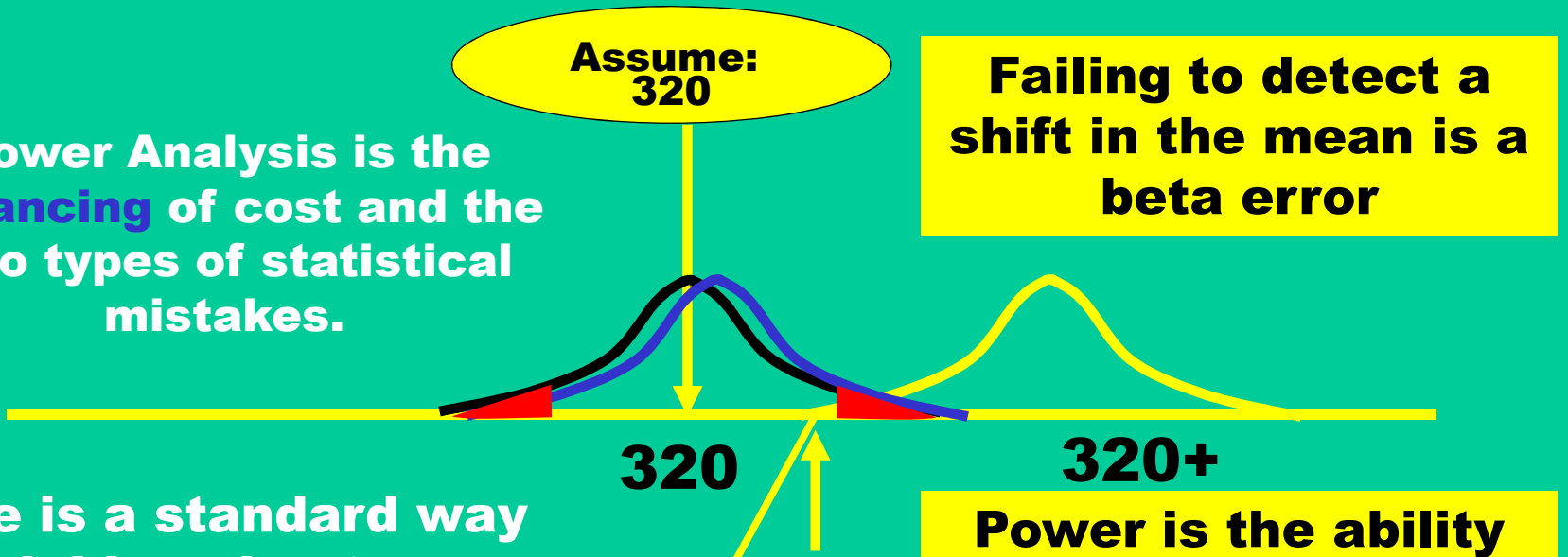
The drug is INeffective & we do mistakenly conclude it is!

This is a mistake!

Review of Logic:

•Balancing cost and errors

Power Analysis is the **Balancing** of cost and the two types of statistical mistakes.



Failing to detect a shift in the mean is a beta error

Power is the ability of a test to detect a shift in mean

There is a standard way of thinking about your decision.

Assume Ho
Mean = 320

	Ho=True	Ho=False
Assume HO & kill the Drug	OK Decision	Beta Error
Reject Ho & market the Drug	Alpha Error	OK Decision

Samples come from the Left end of the Gold distribution.

The drug is Effective and we do not detect that!

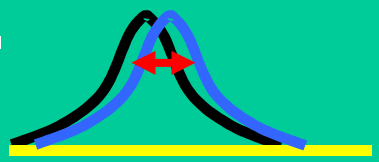
This is a mistake!

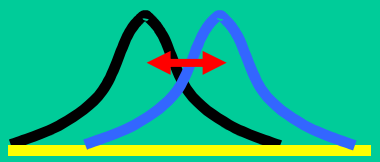
What is Statistical Power?

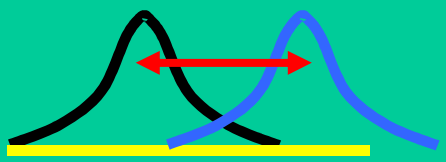
•Why computers are needed

Parts of a Power analysis:

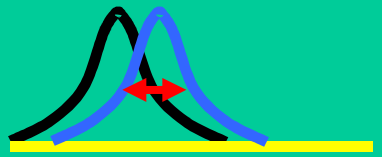
- A Lit search and research review to estimate effect size.

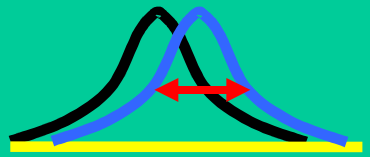


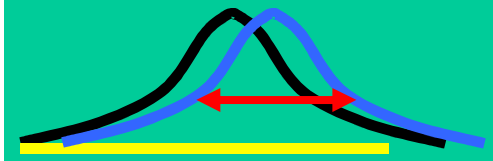




- And estimate Variance







And estimate complete/dropout rates

95 % 90 % 85 %	95 % 90 % 85 %	95 % 90 % 85 %
95 % 90 % 85 %	95 % 90 % 85 %	95 % 90 % 85 %
95 % 90 % 85 %	95 % 90 % 85 %	95 % 90 % 85 %

- Complex formulas make this is a good process to computerize
- SAS syntax makes it easy

The End

Russ.Lavery@verizon.net

Thanks To:
Dr. Bajgier