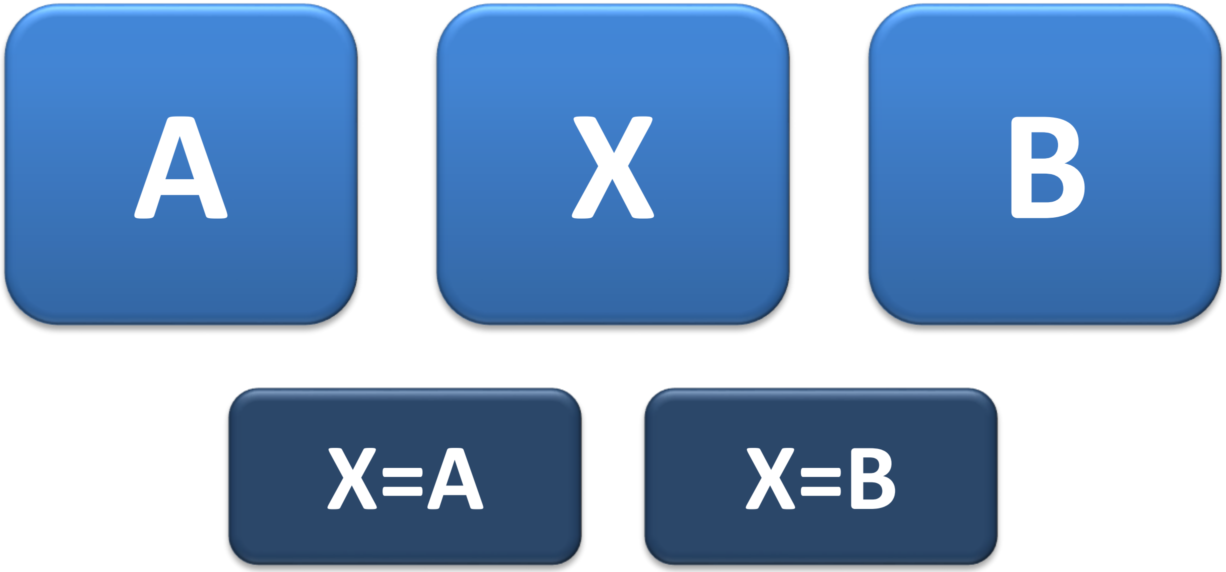
Statistical Analysis of ABX Results Using Signal Detection Theory

Jon Boley[1] and Michael Lester[2]

[1] LSB Audio, LLC

[2] Shure Inc

ABX



Topics

- Experimental Design
- Statistical Analysis
  - Binomial Distributions
  - Signal Detection Theory
- Reporting & Interpreting Results
- Examples

Experimental Design

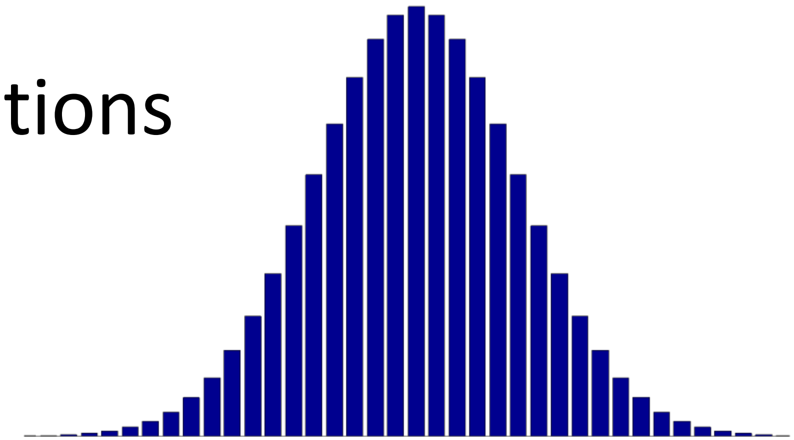
- Clearly define your question & hypothesis
- Choose a test methodology that fits your question
  - Note: ABX can tell you there is a difference, but no test can tell you they are perceptually the same!
- Keep all variables constant (room configuration, level, etc)
- Randomize anything that is not held constant

A Great Resource:

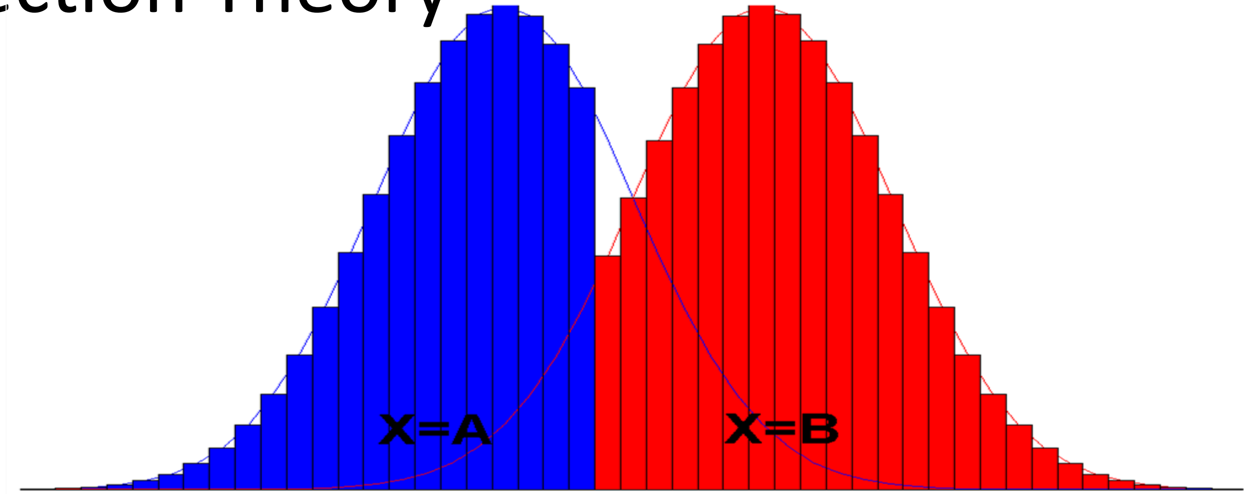
Bech, S., N. Zacharov. Perceptual Audio Evaluation – Theory, Method and Application. First Edition. West Sussex: John Wiley and Sons, Ltd. 2006.

Statistical Analysis

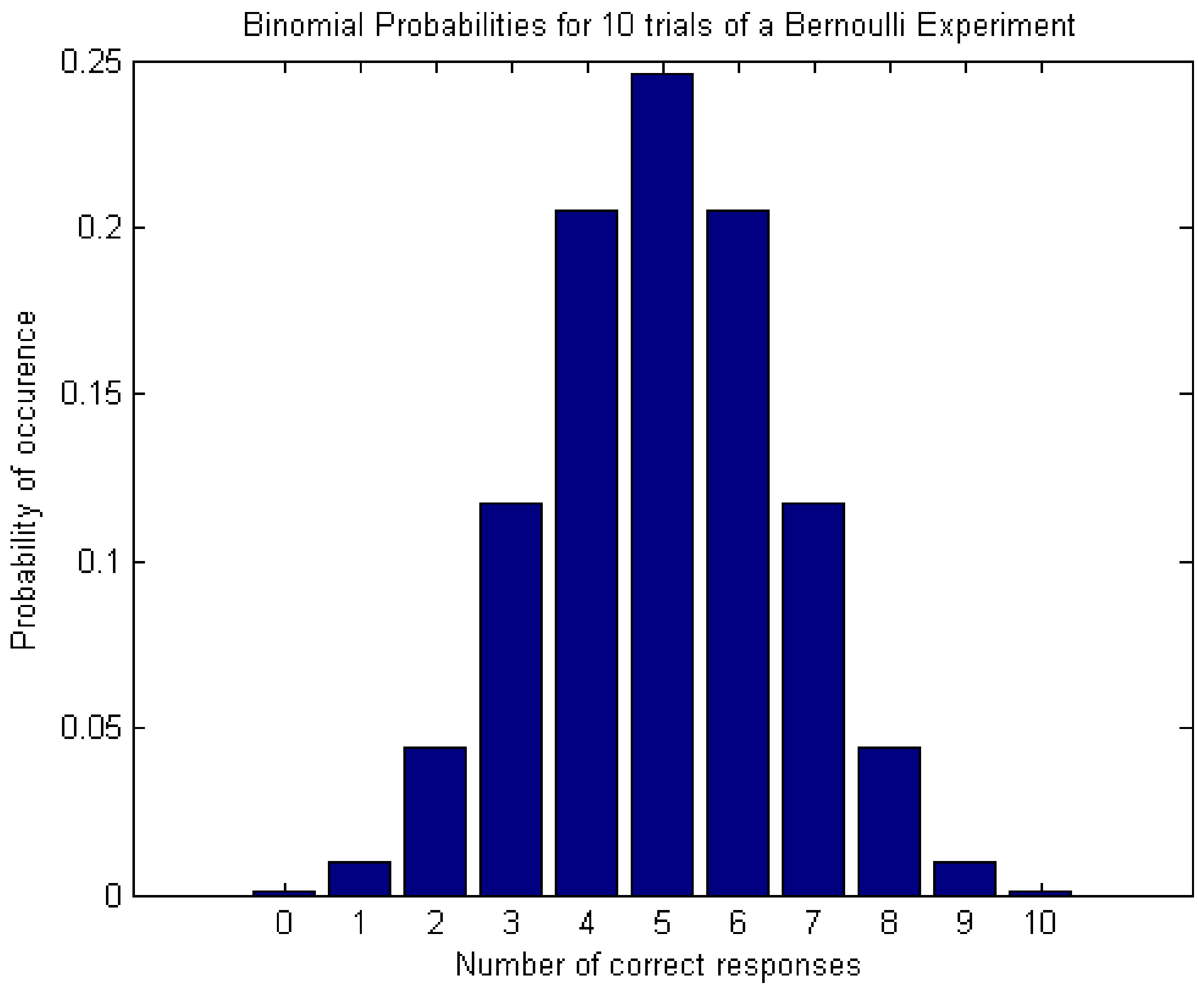
- Binomial Distributions



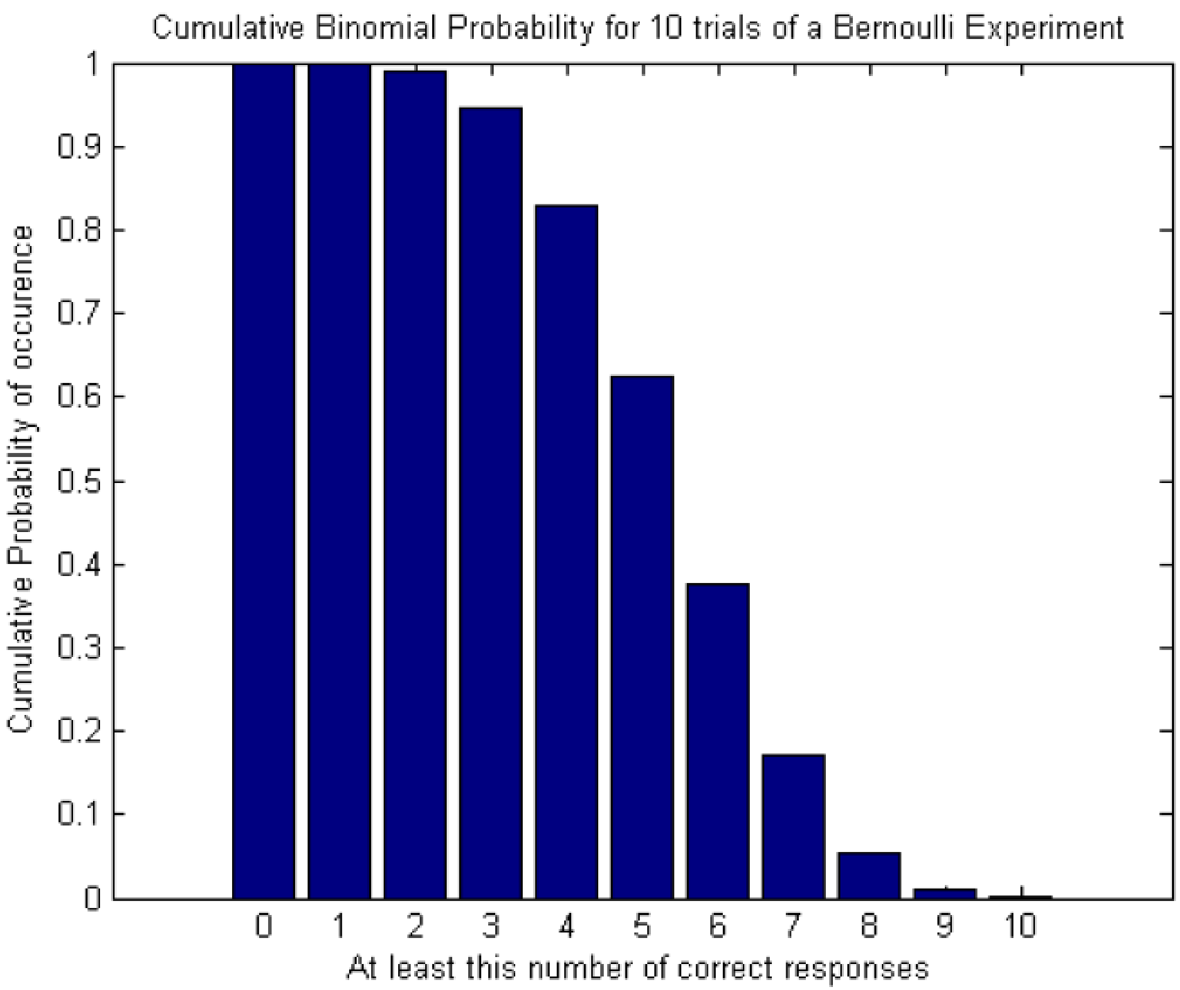
- Signal Detection Theory



Binomial Distribution



$$f(s) = \binom{n}{s} * p^s * (1-p)^{n-s}$$



Cumulative Binomial Probability for 10 trials of a Bernoulli Experiment
Cumulative Probability of occurence
At least this number of correct responses
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0 1 2 3 4 5 6 7 8 9 10

Conventional ABX

	Stimulus	**Response**	**Example Value**
1)	X=A or X=B	Correct (X=A or X=B)	70
2)	X=A or X=B	Incorrect (X=B or X=A)	30

Percent Correct:

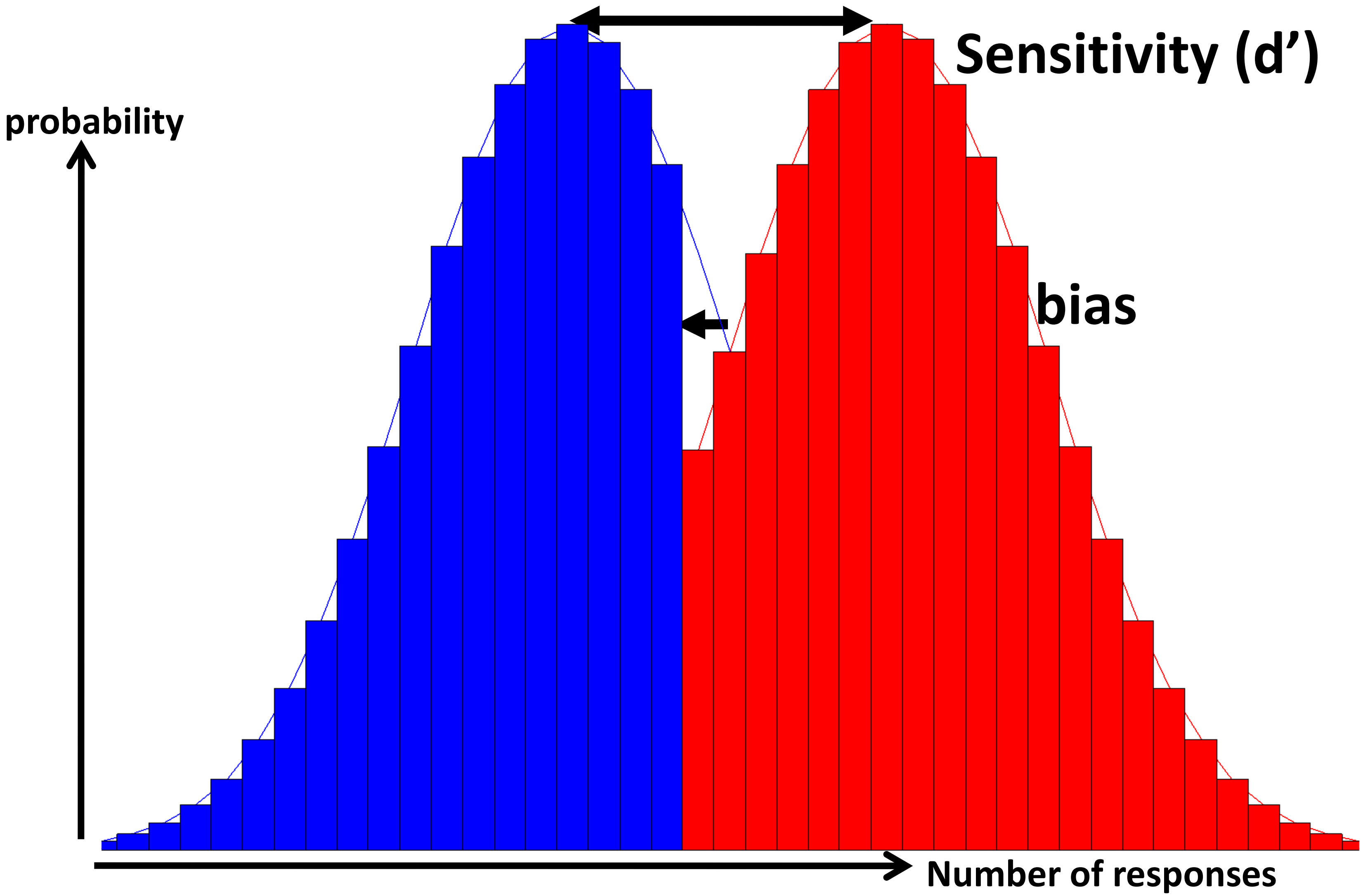
70%

Signal Detection Theory

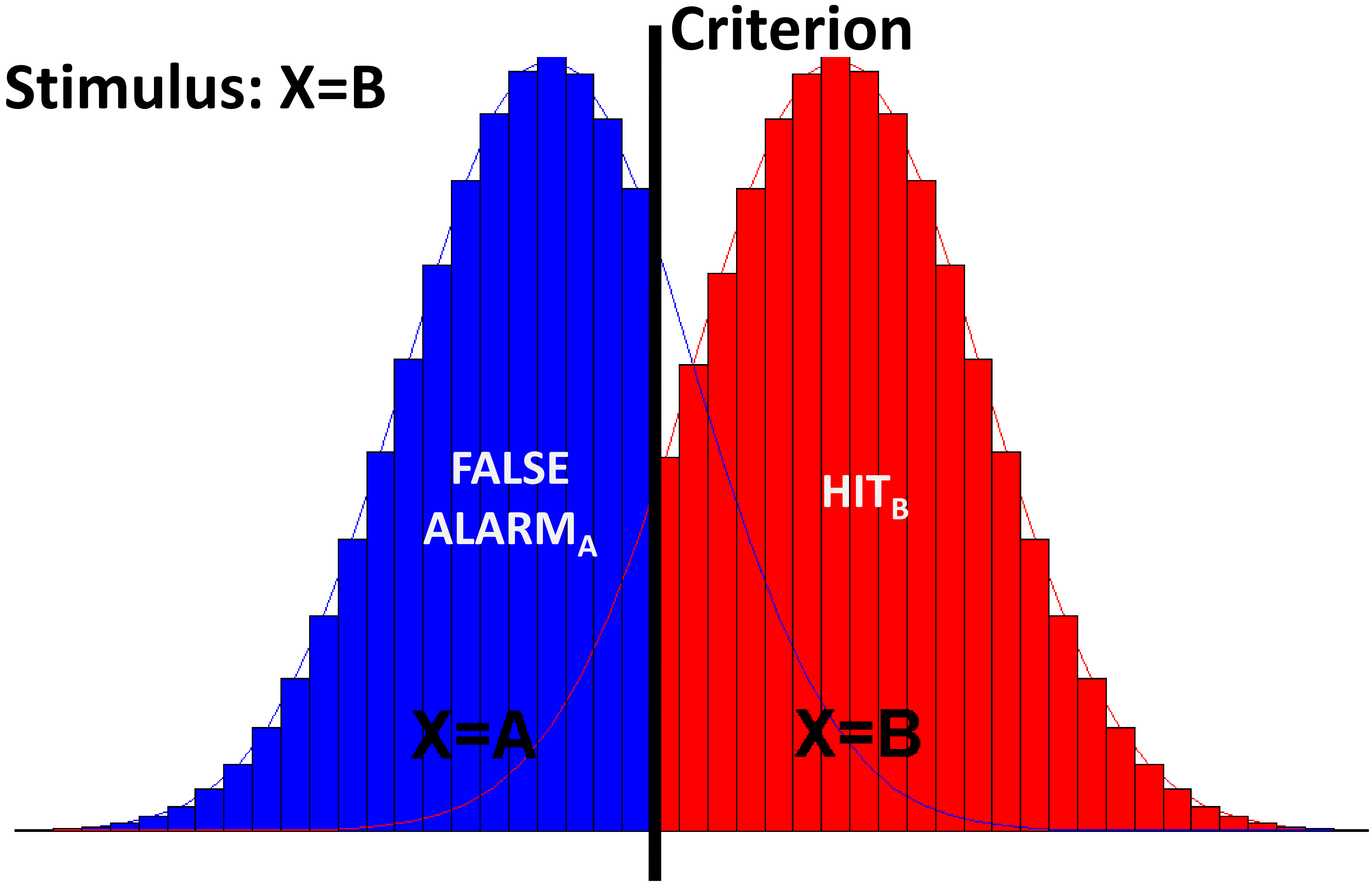
	Stimulus	**Response**	**Example Value**	**Hit / False Alarm**
1)	X=A	X=A	30	$H_A = 30/50 = 0.6$
2)	X=A	X=B	20	$F_B = 20/50 = 0.1$
3)	X=B	X=A	10	$F_A = 10/50 = 0.2$
4)	X=B	X=B	40	$H_B = 40/50 = 0.8$

Percent Correct = 70%

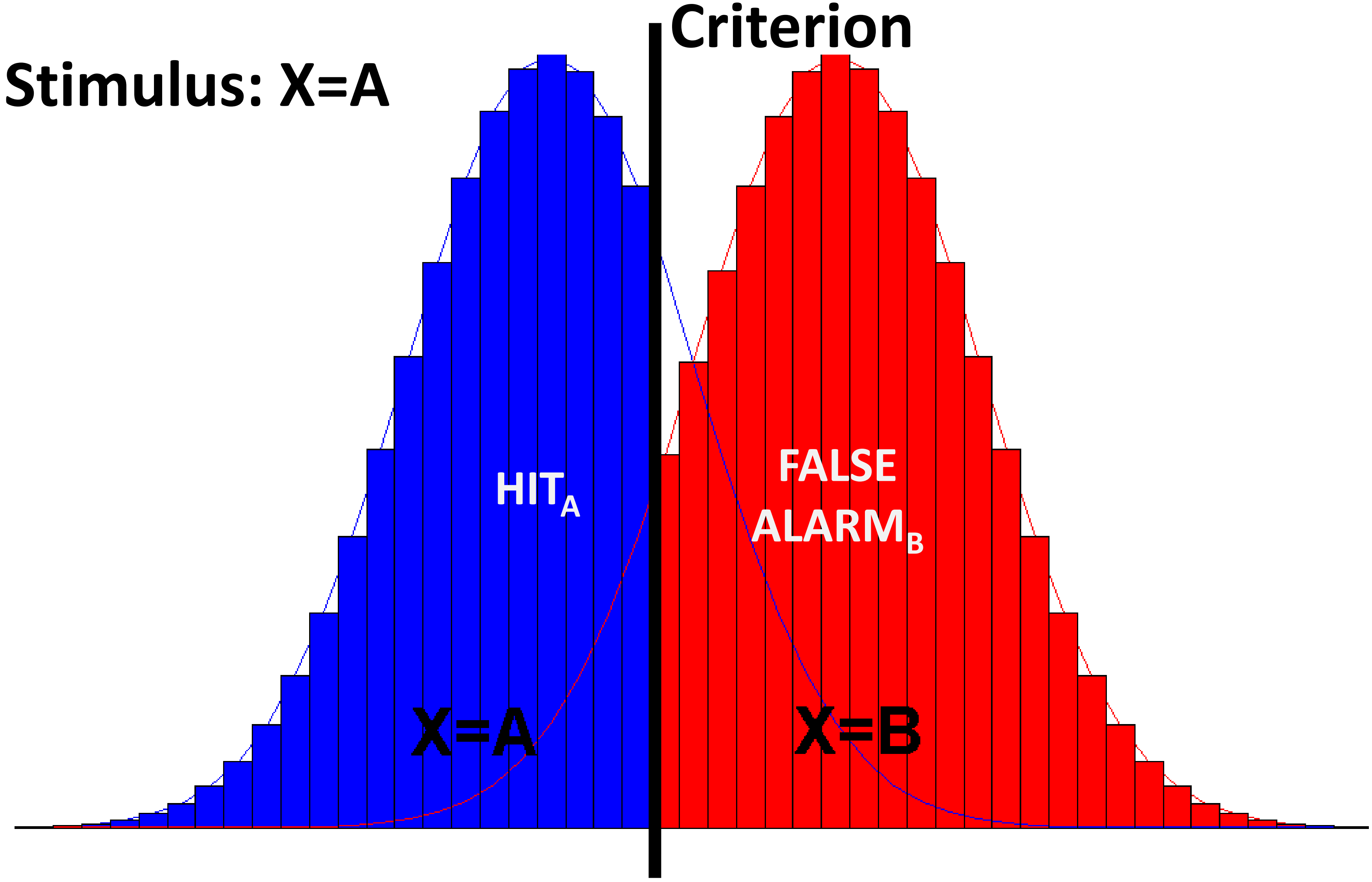
Signal Detection Theory



Signal Detection Theory

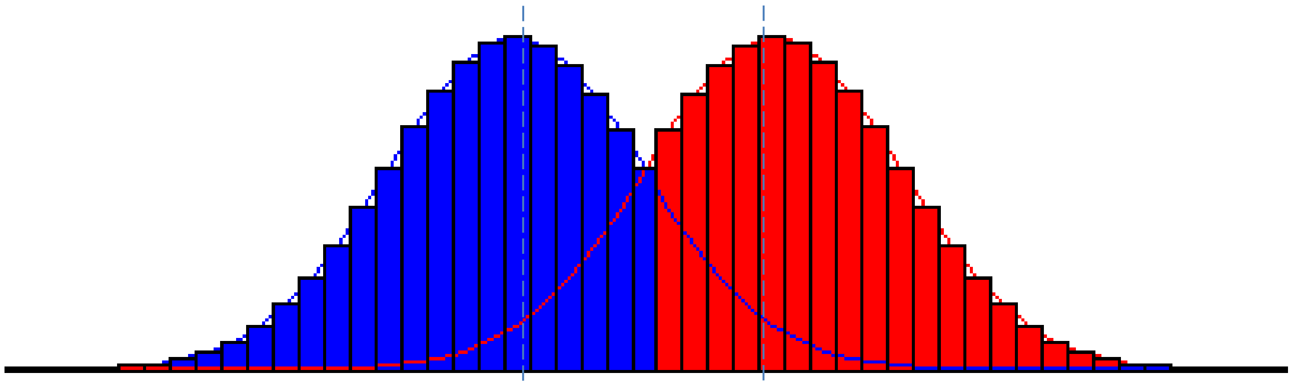


Signal Detection Theory

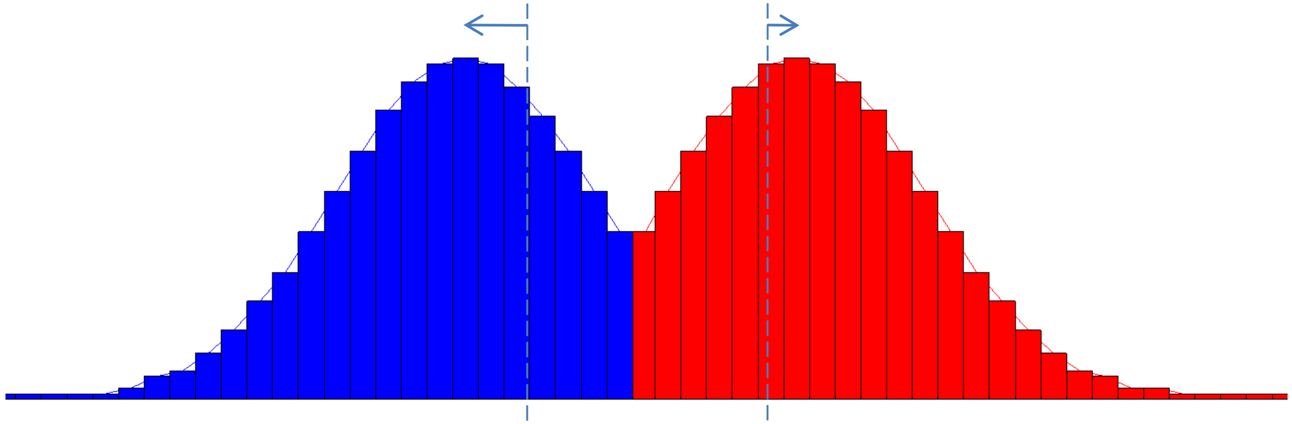


The Problem with Percent Correct

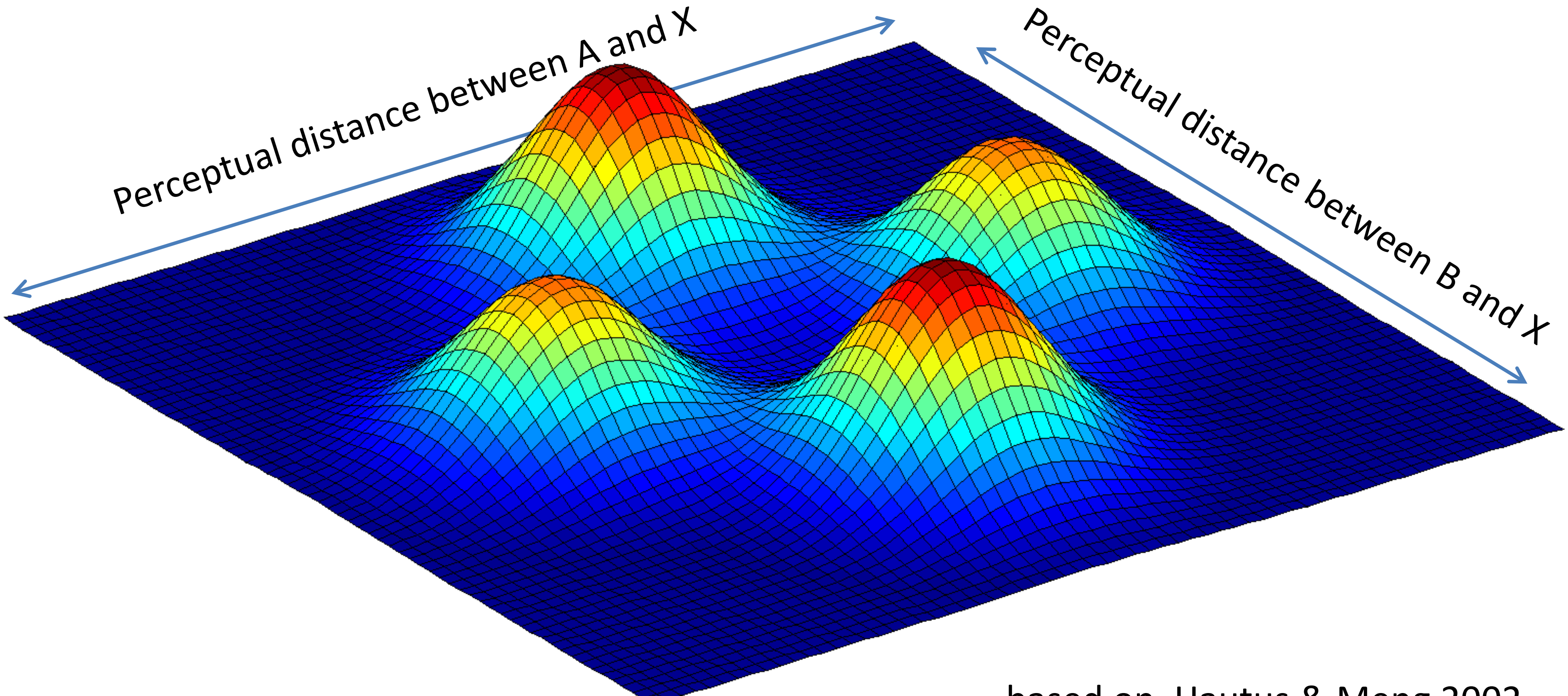
34% Hits_A + 34% Hits_B = 68% correct



40% Hits_A + 40% Hits_B = 80% correct



		Response	
		X=A	**X=B**
Stimulus	**X=A**	Hit_A	$FalseAlarm_B$
	X=B	$FalseAlarm_A$	Hit_B



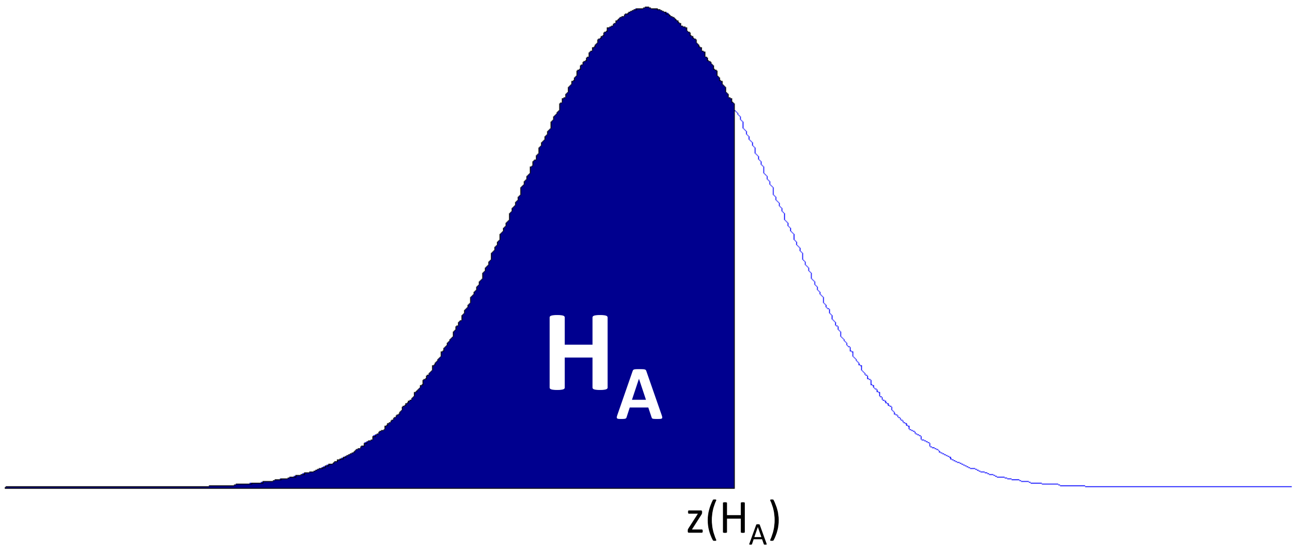
based on Hautus & Meng 2002

Response

		X=A	X=B
Stimulus	**X=A**	H_A	F_B
	X=B	F_A	H_B

$$z(H_A) - z(F_A) = z(H_B) - z(F_B)$$

where $p(S < z(H_A)) = H_A$ for $S \sim \text{Normal}(0,1)$



		Response	
		X=A	**X=B**
Stimulus	**X=A**	H_A	F_B
	X=B	F_A	H_B

$$z(H_A) - z(F_A) = z(H_B) - z(F_B)$$

TABLE A5.3 *Values of d′ for Same-Different and ABX Models (cont.)*

		d′		
		Same-Different	*ABX*	
z(H) – z(F)	*p(c)*$_{unb}$	*(Independent Observation)*	*Independent Observation*	*Differencing*
1.16	0.719	1.92	1.63	1.83
1.17	0.721	1.93	1.64	1.84
1.18	0.722	1.94	1.64	1.85
1.19	0.724	1.95	1.65	1.86
1.20	0.726	1.96	1.66	1.87

Macmillan, N.A. and C.D. Creelman, Detection Theory: A User's Guide 2005.

Sensitivity (d') Confidence Interval

- Calculate the variance of d'

$$var(\mathrm{d}') = \left(\frac{\mathrm{H_A}(1-\mathrm{H_A})}{\mathrm{N_A z^2(H_A)}}\right) + \left(\frac{\mathrm{F_A}(1-\mathrm{F_A})}{\mathrm{N_B z^2(F_A)}}\right)$$
$$= \left(\frac{\mathrm{H_B}(1-\mathrm{H_B})}{\mathrm{N_B z^2(H_B)}}\right) + \left(\frac{\mathrm{F_B}(1-\mathrm{F_B})}{\mathrm{N_A z^2(F_B)}}\right)$$

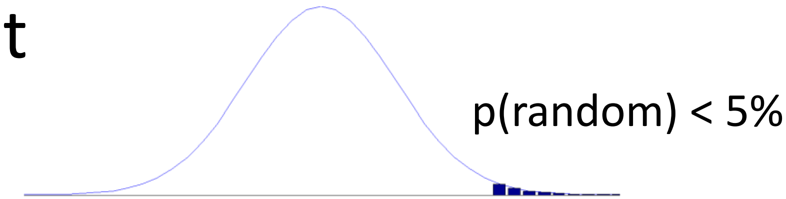
- For 95% confidence interval:
  - assuming normal distribution:

$$95\%\, CI = d' \pm\ 1.96 * \sqrt{var(d')}$$

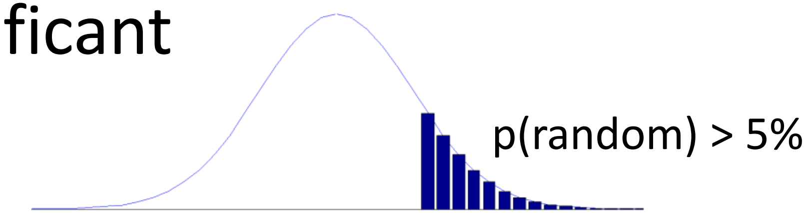
  - based on bootstrapped data analysis:
    - Chapter 4 of *Shao, J., & Tu, D. (1995). The jackknife and bootstrap. New York:Springer-Verlag.*

Reporting & Interpreting Results

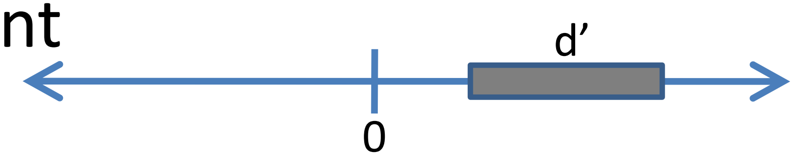
- Binomial Analysis
  - Significant



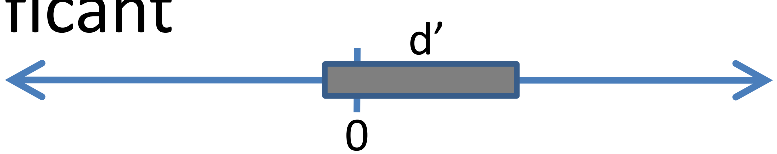
  - Not significant



- Signal Detection Theory Analysis
  - Significant



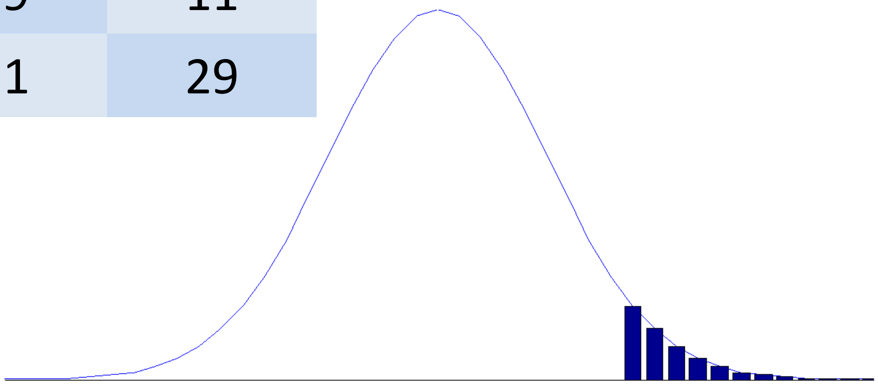
  - Not Significant



Example 1

N = 100 trials

		Responses	
		X=A	**X=B**
Stimuli	**X=A**	29	11
	X=B	31	29



<u>Simple Binomial Analysis:</u>

100 trials, 58 correct
–> lookup in binomial cdf table
-> 4.43% probability of getting the same result or larger by guessing

Since this is <5%, we would typically say that there is a statistical difference and the listener was likely able to differentiate between the sounds.

Example 1

N = 100 trials

		Responses		
		X=A	**X=B**	
Stimuli	**X=A**	29	11	40
	X=B	31	29	60
		60	40	

<u>Signal Detection Theory Analysis:</u>

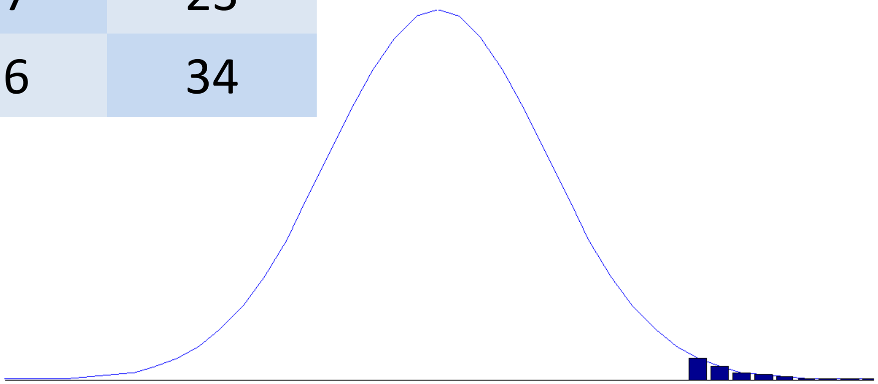
H_A = 29/(29+11) = 0.73;
H_B = 29/(31+29) = 0.48;
F_A = 31/(31+29) = 0.52;
F_B = 11/(29+11) = 0.28;

d' = 1.17
95% confidence interval = 1.17 ± 2.35 = (-1.18, 3.52) -> Because this confidence interval intersects with zero, the perceptual difference is **not** statistically significant.

Example 2

N = 100 trials

		Responses	
		X=A	**X=B**
Stimuli	**X=A**	27	23
	X=B	16	34



<u>Simple Binomial Analysis:</u>

100 trials, 61 correct
–> lookup in binomial cdf table
-> 1.05% probability of getting the same result or larger by guessing

Since this is <5%, we would typically say that there is a statistical difference and the listener was likely able to differentiate between the sounds.

Example 2

N = 100 trials

		Responses		
		X=A	**X=B**	
Stimuli	**X=A**	27	23	50
	X=B	16	34	50
		43	57	

<u>Signal Detection Theory Analysis:</u>

H_A = 27/(27+23) = 0.54;
H_B = 34/(16+34) = 0.68;
F_A = 16/(16+34) = 0.32;
F_B = 23/(27+23) = 0.46;

d' = 1.18
95% confidence interval = 1.18 ± 1.39 = (-0.21, 2.57) -> Because this confidence interval intersects with zero, the perceptual difference is **not** statistically significant.

Concluding Remarks

Remarks on the Determination of a Differential Threshold by the So-Called *ABX* Technique

J. Donald Harris
Medical Research Laboratory, Submarine Base, New London, Connecticut
(Received March 31, 1952)

The case seems to be that *ABX* is too complicated to yield the finest measures of sensitivity.

Of course, from one point of view it is quite justified to use almost any psychophysical procedure, provided others can repeat it exactly, but certainly for some specific purposes a preference can be clearly stated. If I want, for example, to find the finest discrimination of which the ear is capable, I do not want the data to be contaminated with the question how willing the subject is to report similarities and differences. In this connection one recalls

Concluding Remarks

- Signal Detection Theory (SDT) allows us to separate sensitivity and bias
  - If you can show that there is no bias, other methods are sufficient
- SDT requires that you keep track of what was presented, and how the listener responded
  - Most existing software packages do not do this

Sample Code

- An example Matlab script is posted at http://www.lsbaudio.com/ABX