Null Hypothesis Testing

Klinkenberg

14 sep 2017

TOC

- null hypothesis significance testing
  - h_0 and h_a
  - frequentist probability
  - standard error

decision table

  - alpha alpha
  - power
  - post hoc power
  - effect size
  - alpha
  - beta
  - p value

reasoning scheme

Neyman-Pearson Paradigm

Null Hypothesis Significance Testing

H_0 and H_A

H_0

- Skeptical point of view
- No effect
- No preference
- No Correlation
- No difference

H_A

- Refute Skepticism
- Effect
- Preference
- Correlation
- Difference

Frequentist probability

- Objective Probability
- Relative frequency in the long run

Standard Error

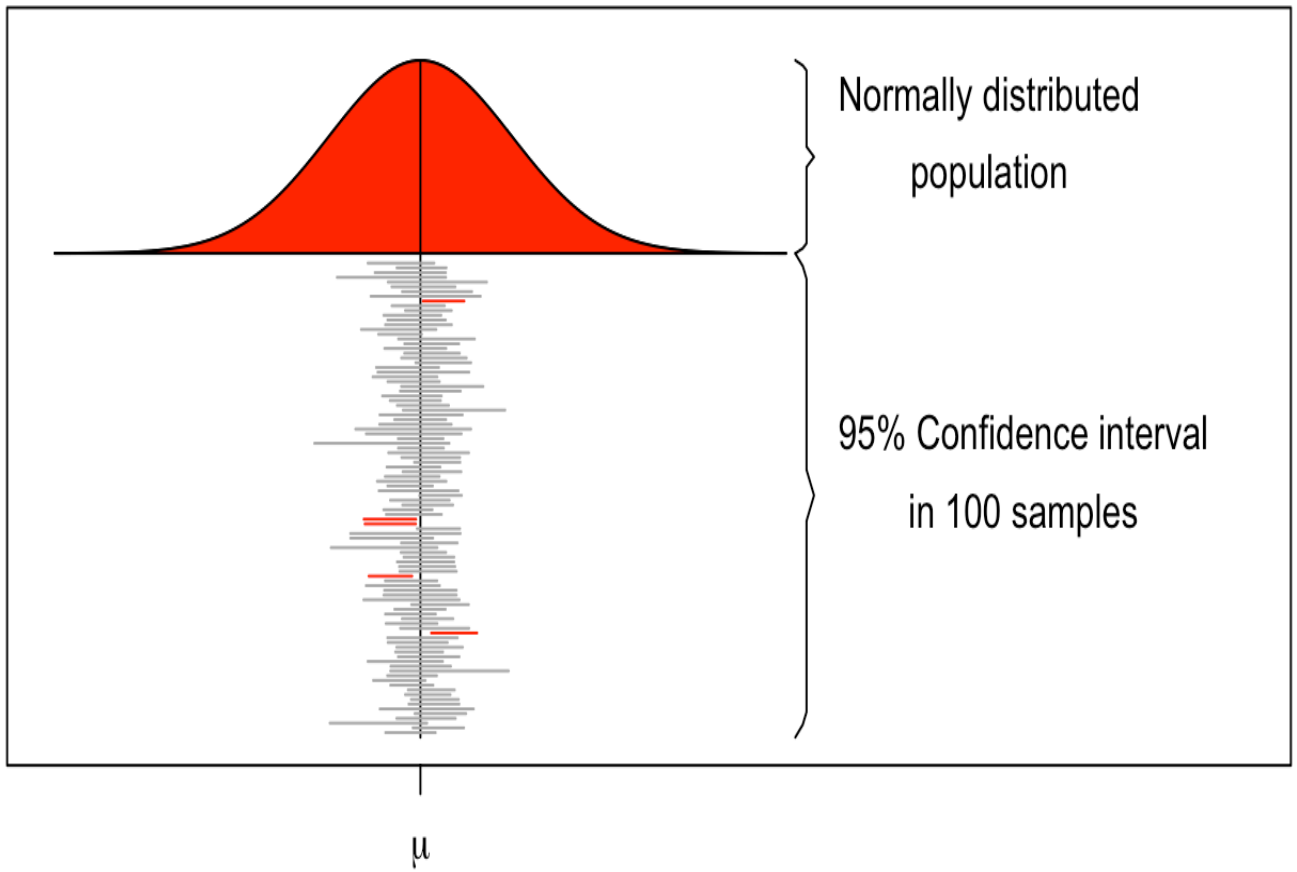
95% confidence interval

$$SE = \frac{\text{Standard deviation}}{\text{Square root of sample size}} = \frac{s}{\sqrt{n}}$$

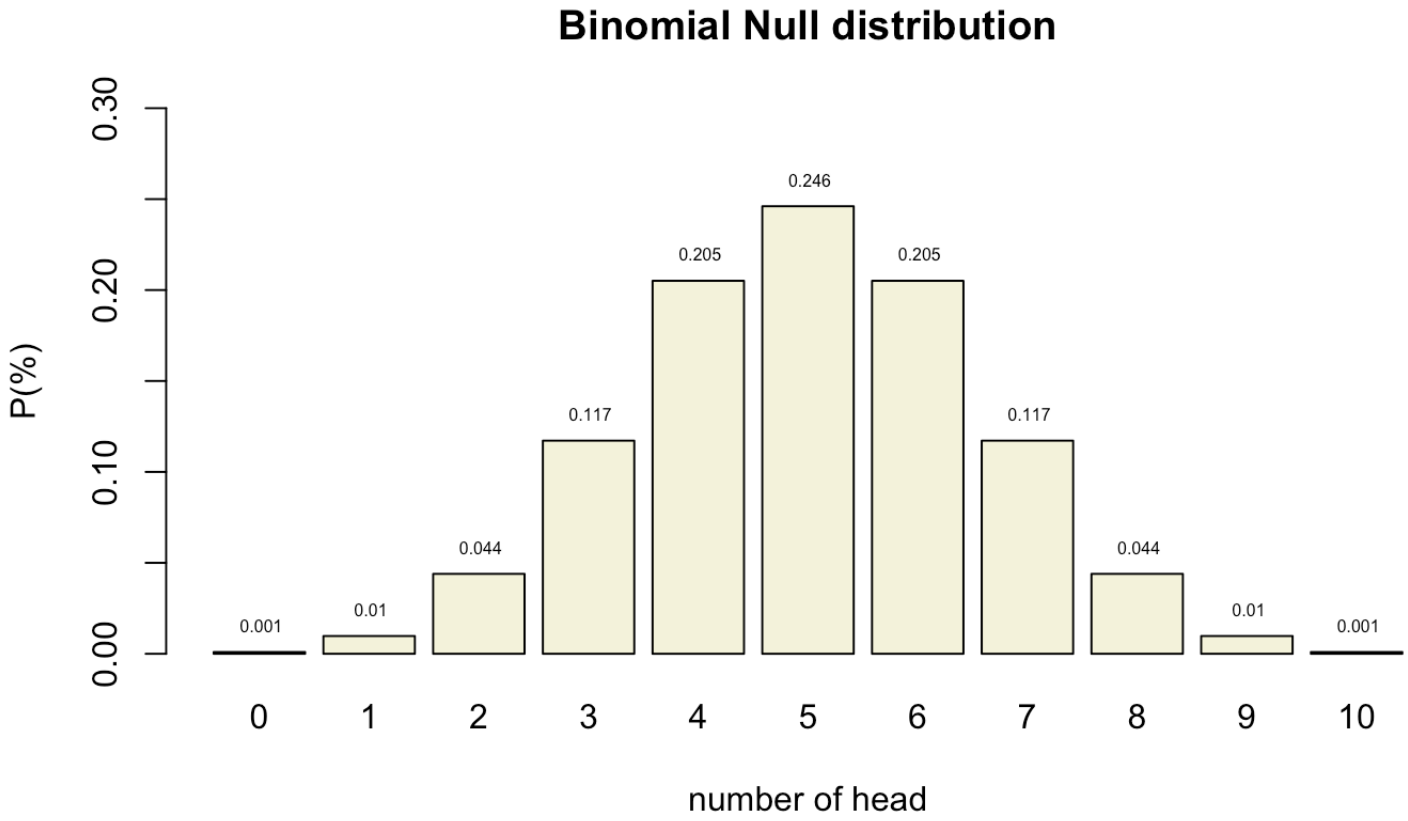
- Lowerbound = $\bar{x} - 1.96 \times SE$
- Upperbound = $\bar{x} + 1.96 \times SE$

SAMPLES FROM NORMAL DISTRIBUTION

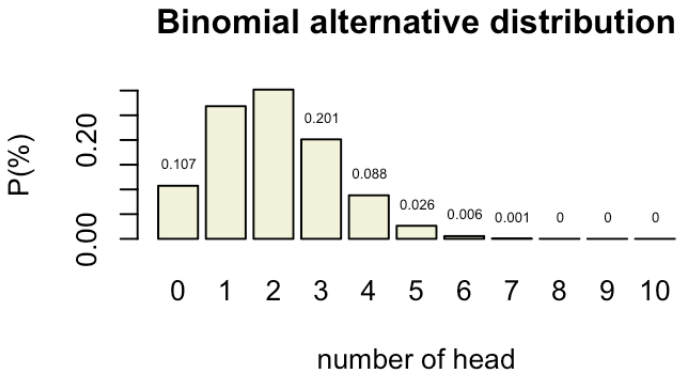
μ falls within the CI in ~95% of the samples

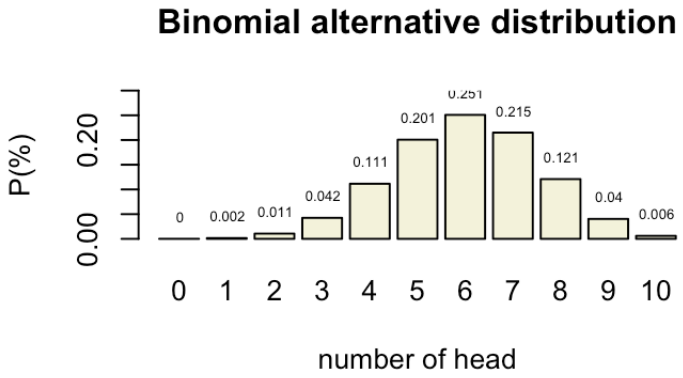


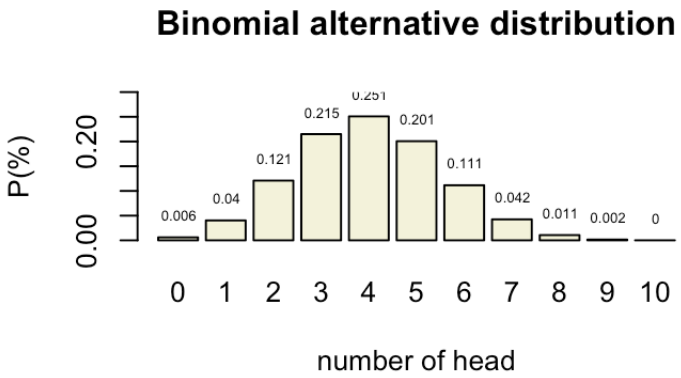
Binomial H_0 distribution

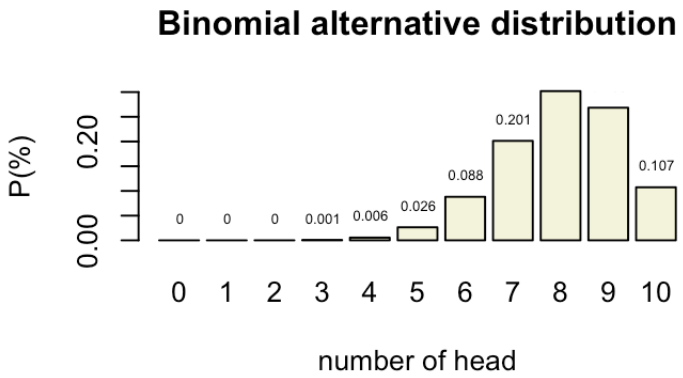

Binomial Null distribution
P(%)
0.00
0.10
0.20
0.30
0.001
0.01
0.044
0.117
0.205
0.246
0.205
0.117
0.044
0.01
0.001
0
1
2
3
4
5
6
7
8
9
10
number of head

Binomial H_A distribution


Binomial alternative distribution
P(%)
0.00
0.20
0.107
0.201
0.088
0.026
0.006
0.001
0
0
0
0 1 2 3 4 5 6 7 8 9 10
number of head


Binomial alternative distribution
P(%)
0.00
0.20
0
0.002
0.011
0.042
0.111
0.201
0.251
0.215
0.121
0.04
0.006
0 1 2 3 4 5 6 7 8 9 10
number of head


Binomial alternative distribution
P(%)
0.00
0.20
0.006
0.04
0.121
0.215
0.251
0.201
0.111
0.042
0.011
0.002
0
0 1 2 3 4 5 6 7 8 9 10
number of head

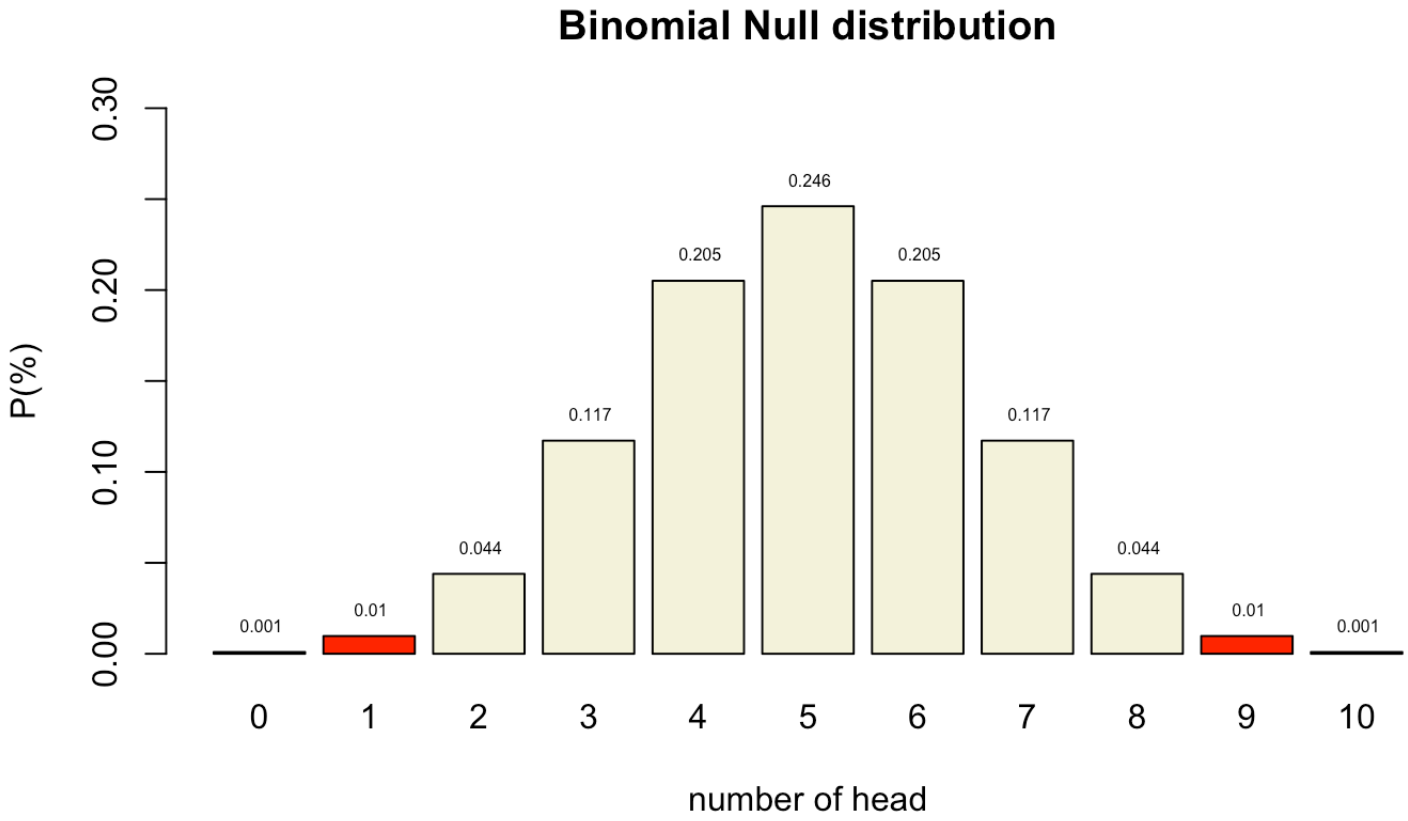

Binomial alternative distribution
P(%)
0.00
0.20
0
0
0
0.001
0.006
0.026
0.088
0.201
0.107
0 1 2 3 4 5 6 7 8 9 10
number of head

Decision table

	H_0 = True	H_0 = False
Decide to reject H_0	Type I error Alpha α	Correct True positive = Power
Decide not to reject H_0	Correct True negative	Type II error Beta β

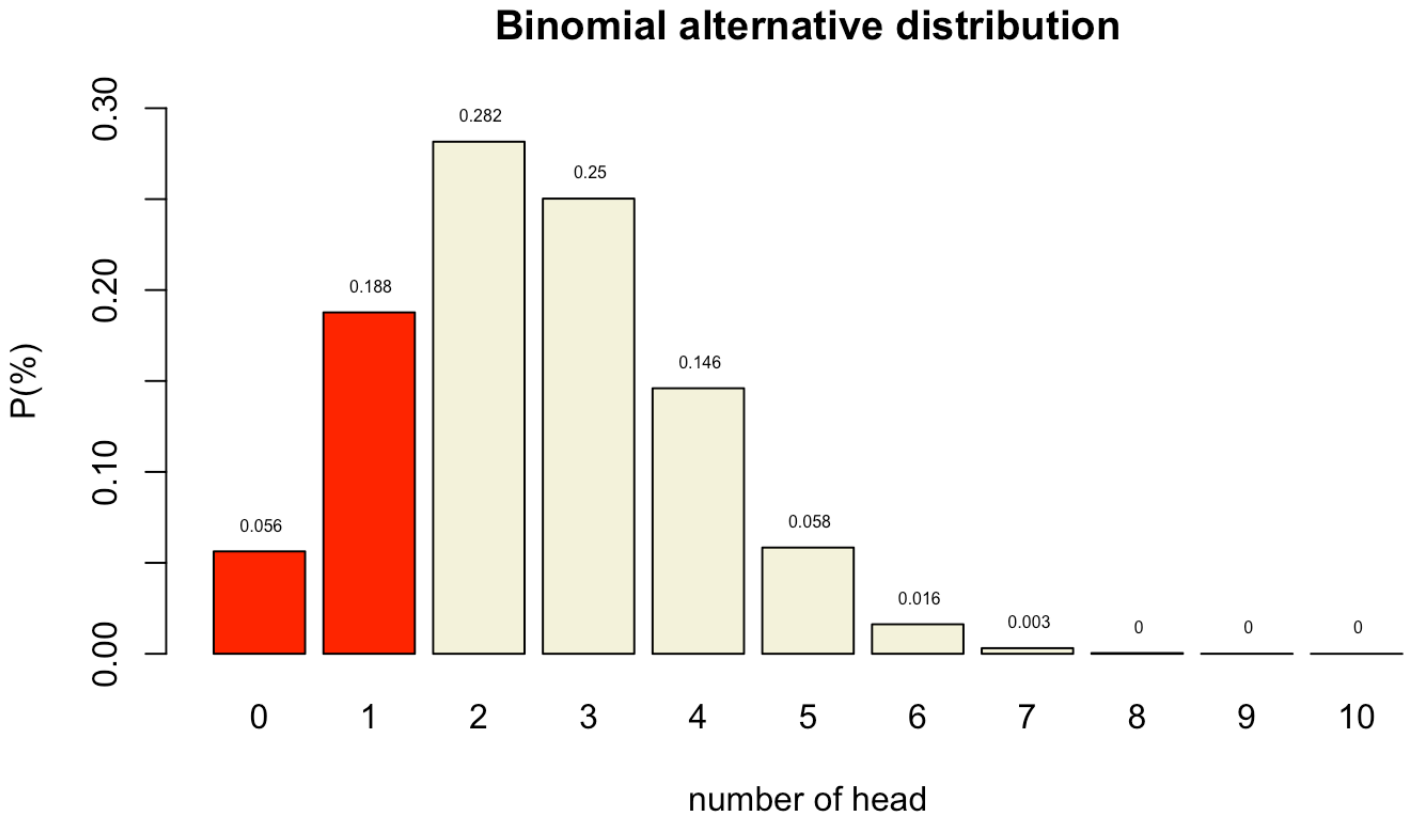
Alpha α

- Incorrectly reject H_0
- Type I error
- False Positive
- Criteria often 5%
- Distribution depends on sample size


Binomial Null distribution
P(%)
0.00
0.10
0.20
0.30
0.001
0.01
0.044
0.117
0.205
0.246
0.205
0.117
0.044
0.01
0.001
0
1
2
3
4
5
6
7
8
9
10
number of head

Power

- Correctly reject H_0
- True positive
- Power equal to: 1 - Beta
  - Beta is Type II error
- Criteria often 80%
- Depends on sample size


Binomial alternative distribution
P(%)
0.00
0.10
0.20
0.30
0.056
0.188
0.282
0.25
0.146
0.058
0.016
0.003
0
0
0
0
1
2
3
4
5
6
7
8
9
10
number of head

Post-Hoc Power

- Also known as: observed, retrospective, achieved, prospective and a priori power
- Specificly meaning:

The power of a test assuming a population effect size equal to the observed effect size in the current sample.

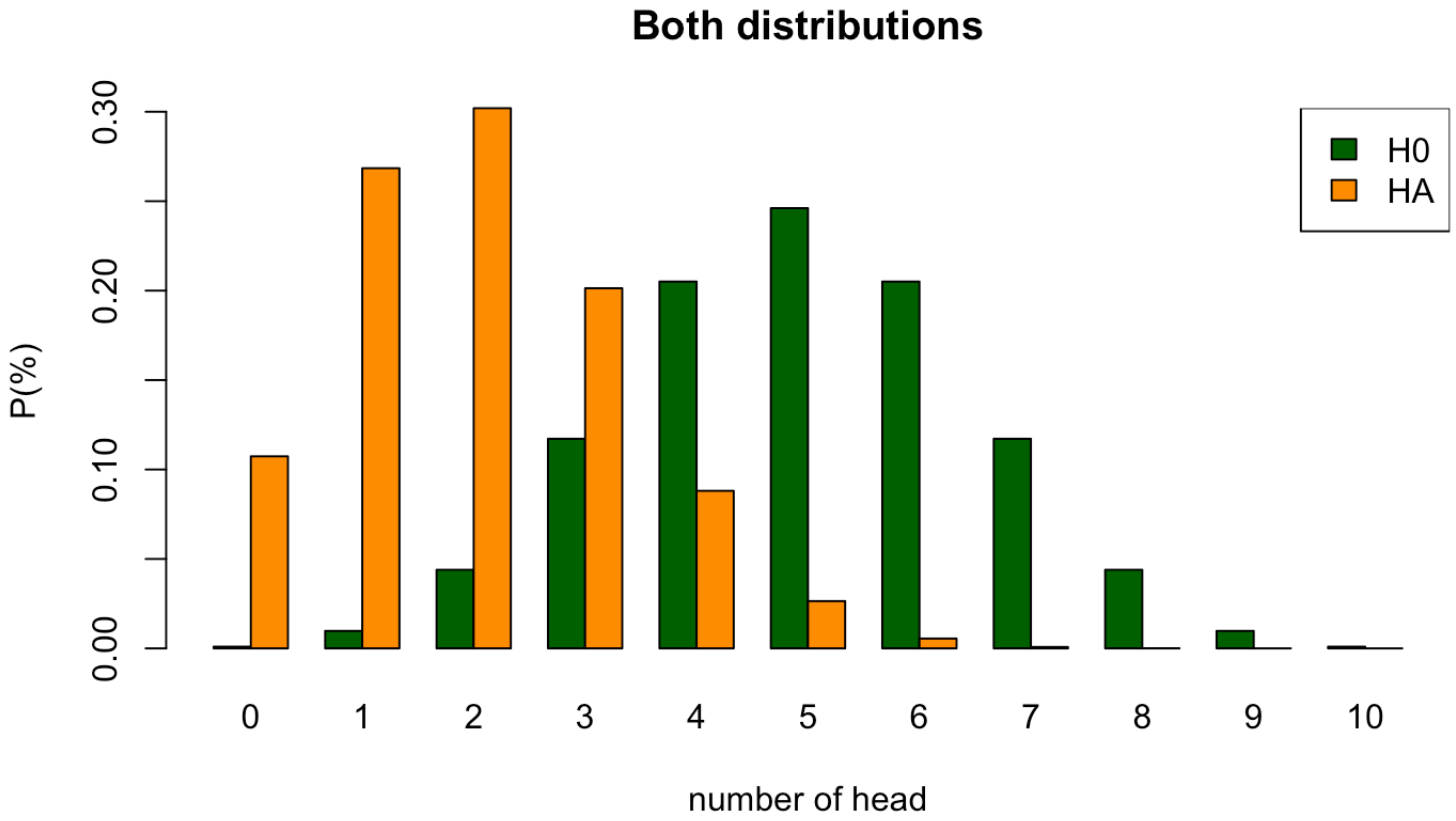
Source: O'Keefe (2007)

Effect size

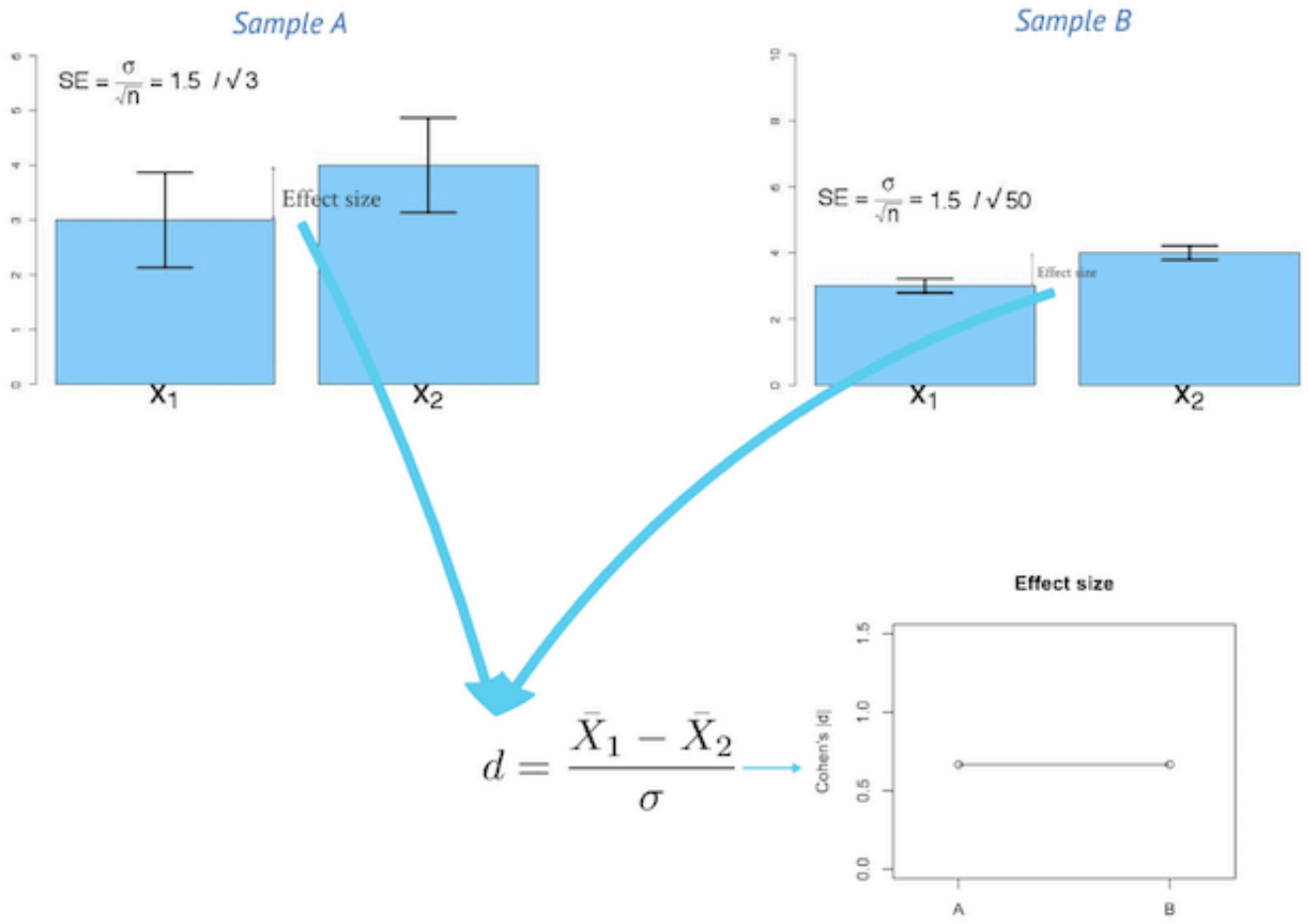
In statistics, an effect size is a quantitative measure of the strength of a phenomenon. Examples of effect sizes are the correlation between two variables, the regression coefficient in a regression, the mean difference and standardised differences.

For each type of effect size, a larger absolute value always indicates a stronger effect. Effect sizes complement statistical hypothesis testing, and play an important role in power analyses, sample size planning, and in meta-analyses.

Source: WIKIPEDIA

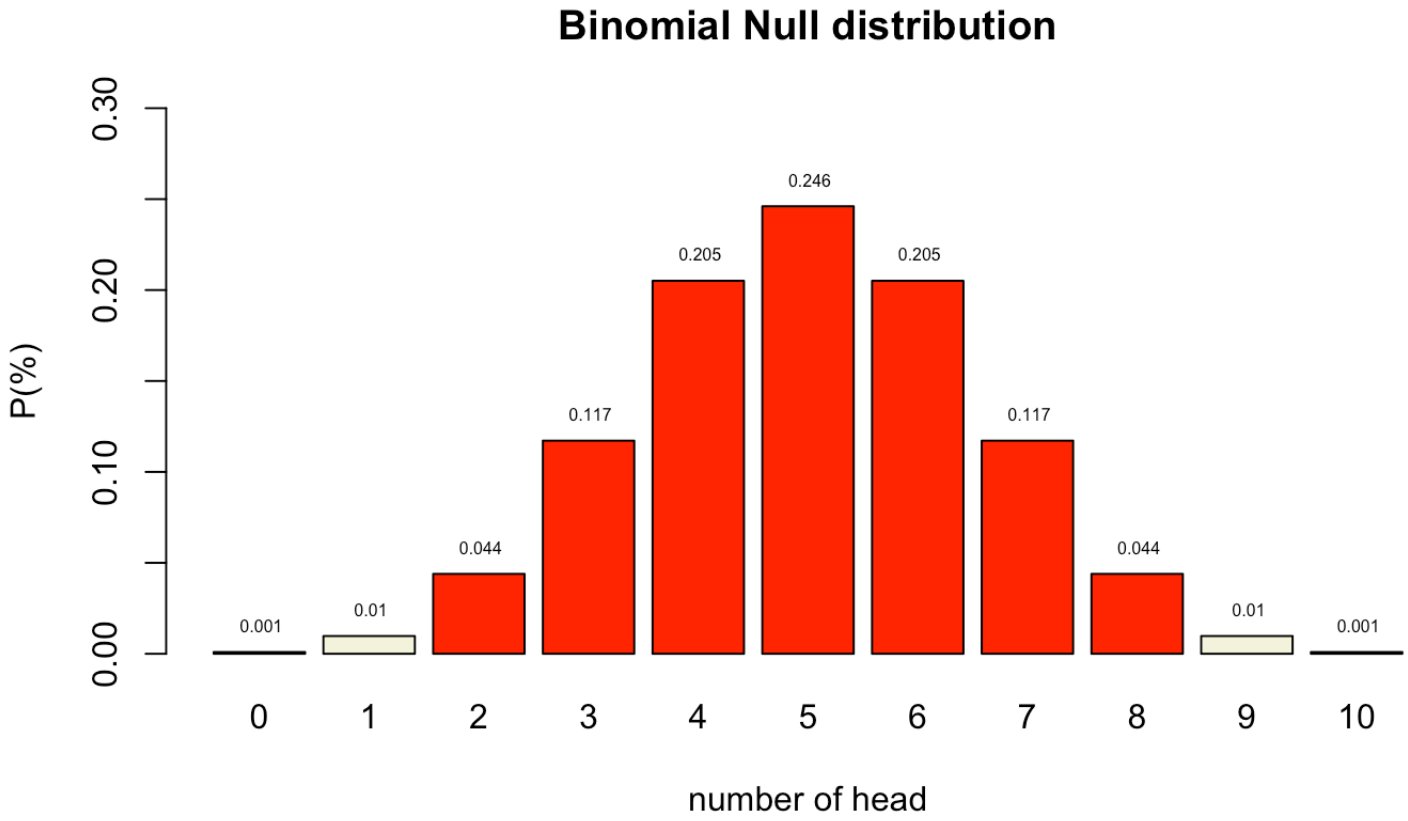

Both distributions
P(%)
0.00
0.10
0.20
0.30
H0
HA
0
1
2
3
4
5
6
7
8
9
10
number of head

Effect size


Sample A
SE = σ/√n = 1.5 /√3
Effect size
X1
X2
Sample B
SE = σ/√n = 1.5 /√50
X1
X2
$d = \frac{\bar{X}_1 - \bar{X}_2}{\sigma}$
Effect size
Cohen's |d|
0.0
0.5
1.0
1.5
A
B

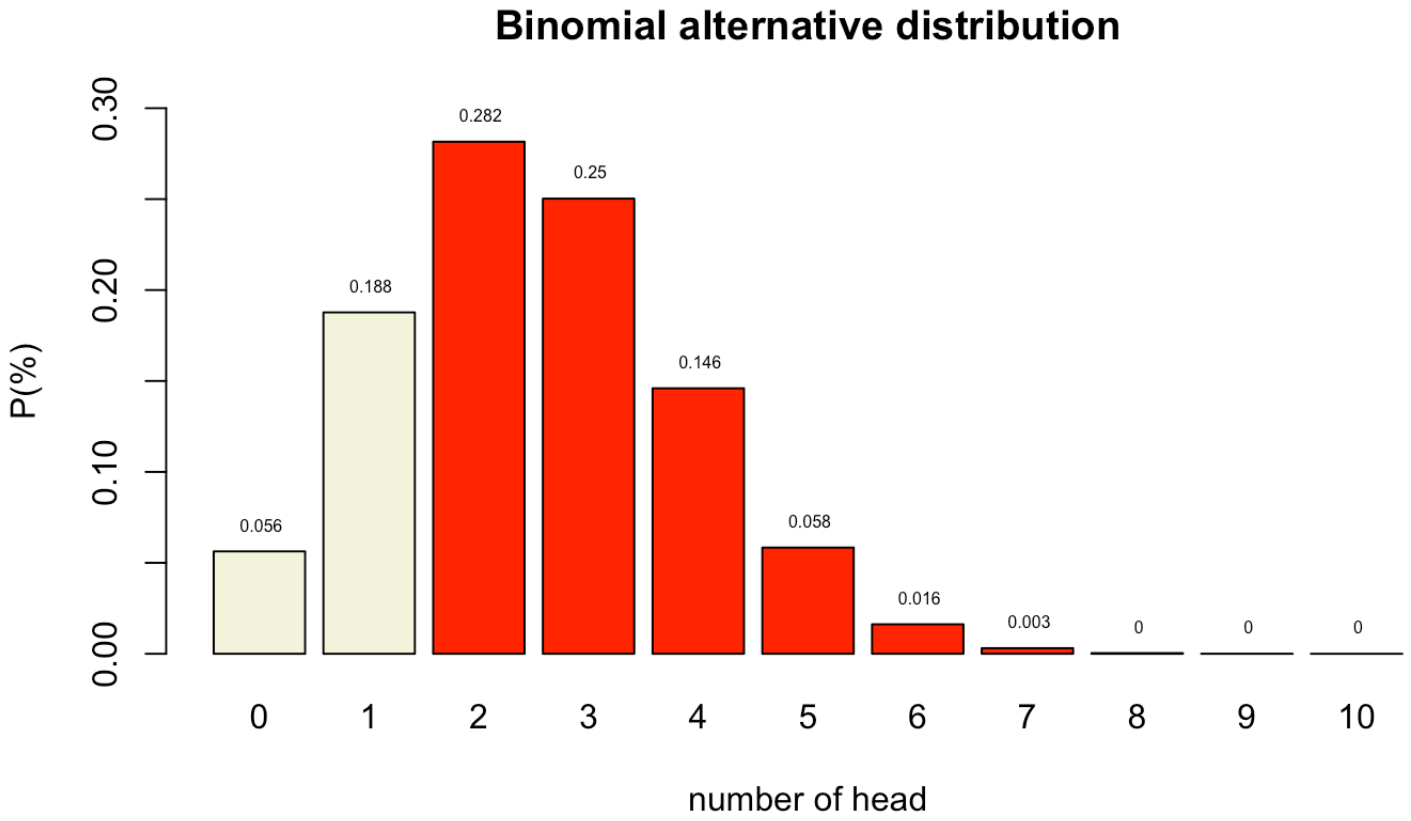
1 - alpha

- Correctly accept H_0
- True negative


Binomial Null distribution
P(%)
0.00
0.10
0.20
0.30
0.001
0.01
0.044
0.117
0.205
0.246
0.205
0.117
0.044
0.01
0.001
0
1
2
3
4
5
6
7
8
9
10
number of head

Beta

- Incorrectly accept H_0
- Type II error
- False Negative
- Criteria often 20%
- Distribution depends on sample size

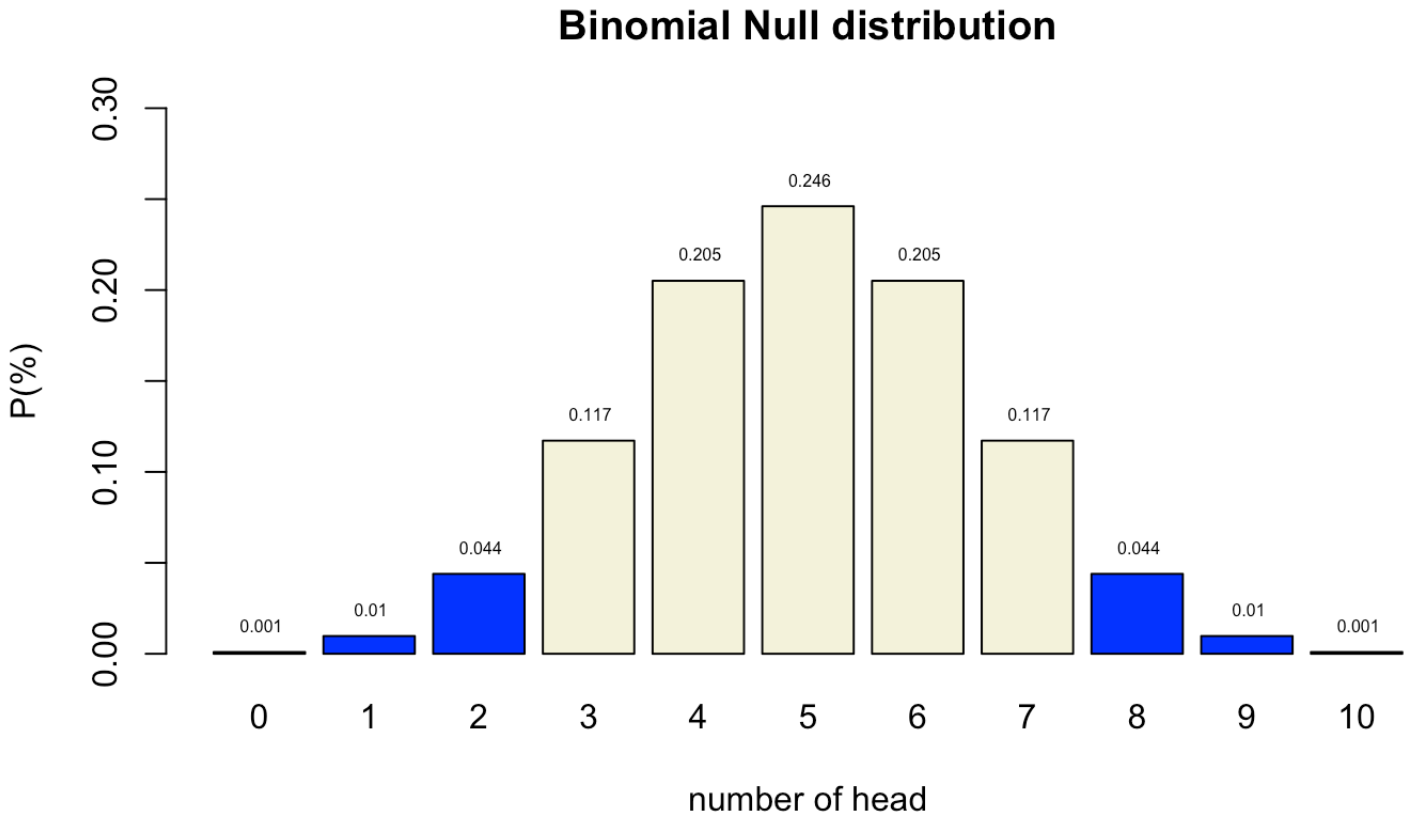

Binomial alternative distribution
P(%)
0.00
0.10
0.20
0.30
0.056
0.188
0.282
0.25
0.146
0.058
0.016
0.003
0
0
0
0
1
2
3
4
5
6
7
8
9
10
number of head

P-value

Conditional probability of the found test statistic or more extreme assuming the null hypothesis is true.

Reject H_0 when:

- p-value <= $alpha$
- test statistic more extreme than your critical value


Binomial Null distribution
P(%)
0.00
0.10
0.20
0.30
0.001
0.01
0.044
0.117
0.205
0.246
0.205
0.117
0.044
0.01
0.001
0
1
2
3
4
5
6
7
8
9
10
number of head

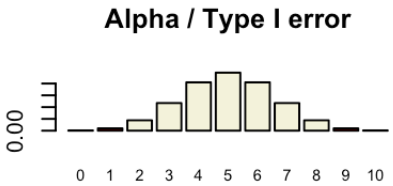
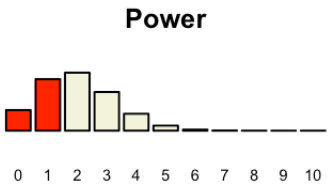
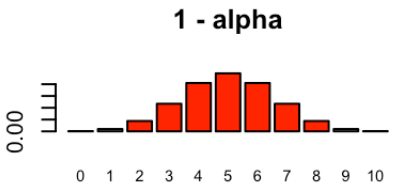
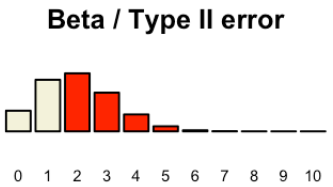
Test statistics

Some common test statistics

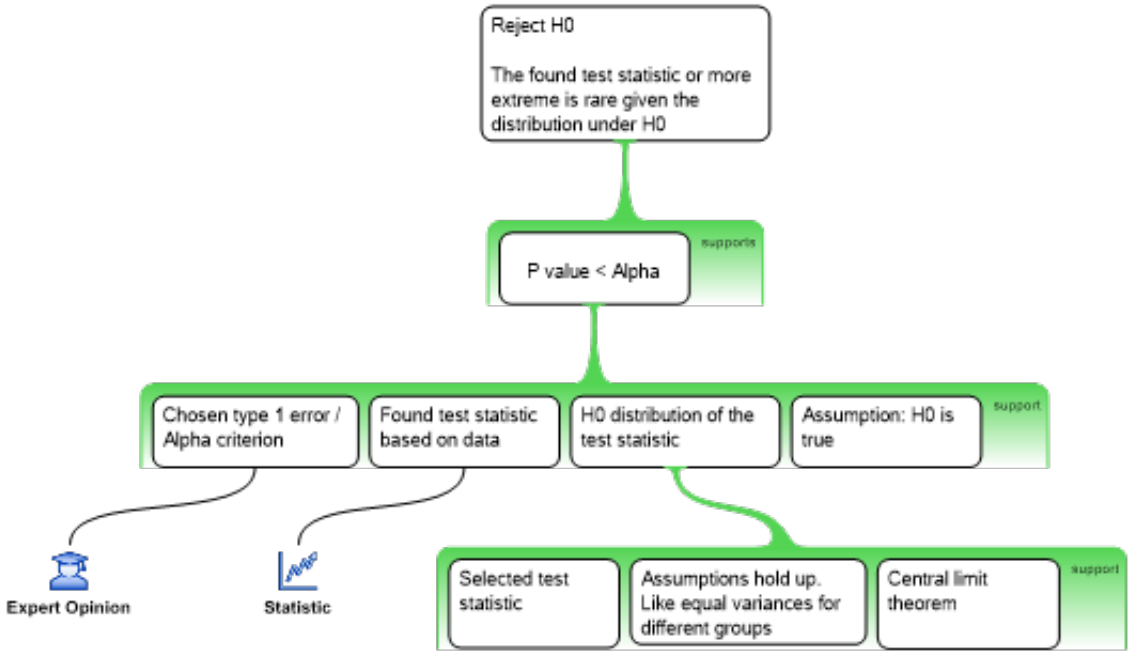
- Number of heads
- Sum of dice
- Difference
- t-statistic
- F-statistic
- χ^2-statistic
- etc…

Decision Table

```
N     = 10  # Sample size
H0    = .5  # Probability of head under H0 50/50
HA    = .2  # Aternative expected value
alpha = .05 # Selected type I error

# Color areas red for selected alpha
area <- dbinom(0:N, N, H0) < alpha/2

col = rep("beige", N+1)
col[area] = "red"

col2 = rep("red", N+1)
col2[area] = "beige"

# Delete # to not color the plots
# col = col2  = "beige"

layout(matrix(1:9,3,3, byrow=T))

plot.new()
text(0.5,0.5,"Binomial Distribution",cex=1.5)
```

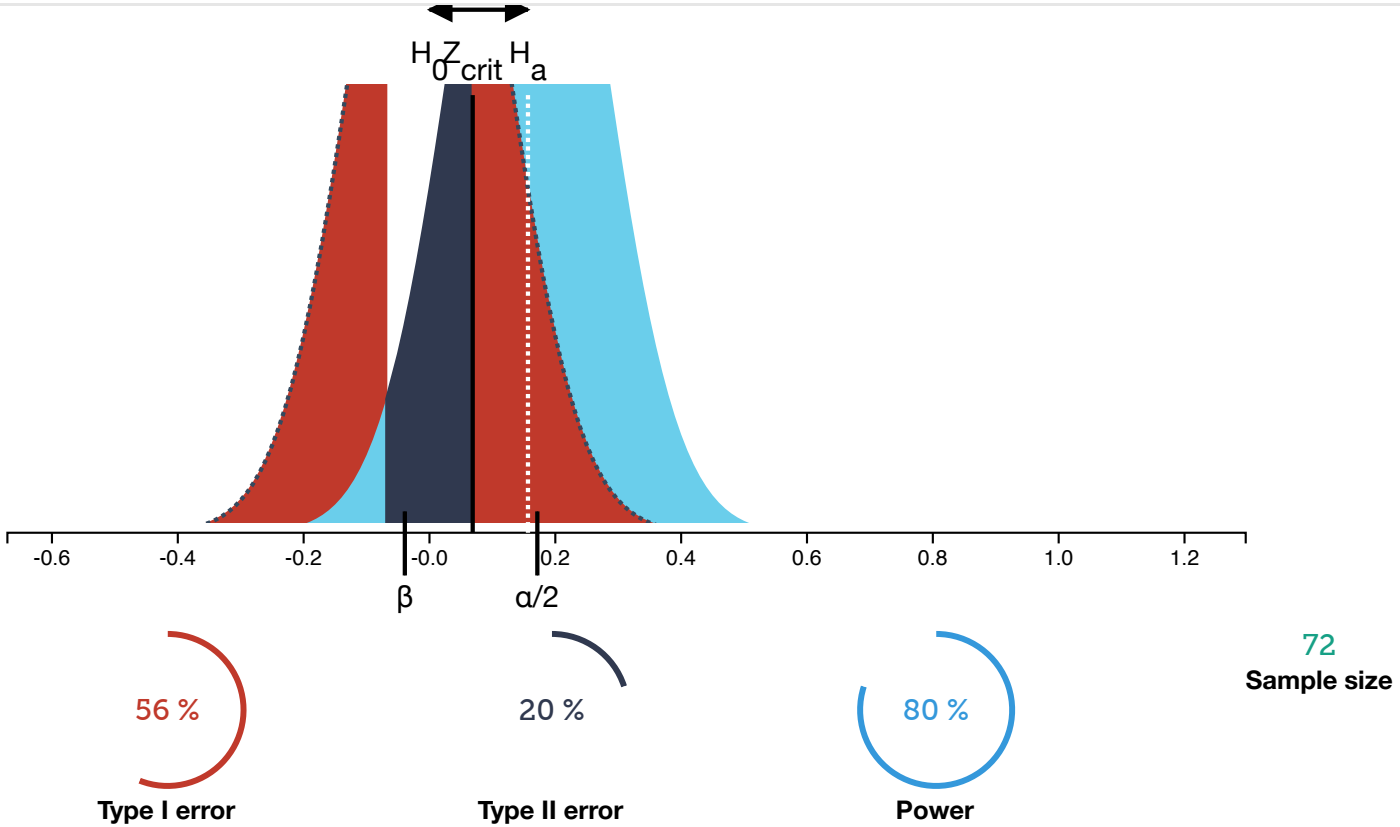
Binomial Distribution	H0 True	H0 False
Reject H0		
Accept H0		

Reasoning Scheme



NHST Reasoning Scheme

R<-PSYCHOLOGIST



Clarification on power ("-") when the effect is 0

The visualization will show that "power" and "Type II error" is "-" when *d* is set to zero. However, the Type I error

Source: Kristoffer Magnusson

- @shklinkenberg
- Klinkenberg
- S.Klinkenberg@UvA.nl
- ShKlinkenberg

END