Urban Informatics

Fall 2018

dr. federica bianco fbianco@nyu.edu

@fedhere

CUSP

Summary:

- **Epistemological concepts:**
  falsifiability, law of parsimony,
- **Good scientific practice:**
  reproducibility of research
- **Gathering parsing data, API:**
  data munging or wrangling, data jujitsu

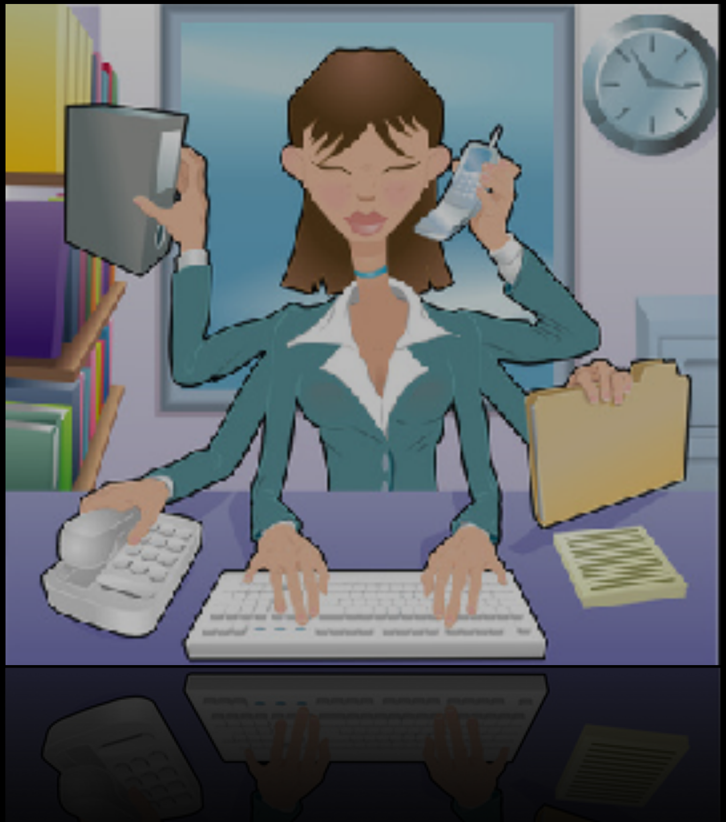
Summary:

- **Epistemological concepts:**
  falsifiability, law of parsimony,
- **Good scientific practice:**
  reproducibility of research
- **Gathering parsing data, API:**
  data munging or wrangling, data jujitsu
- **Formulating and testing a scientific Hypothesis**
  Basic statistics: distributions and their moments
  Hypothesis testing: p-value, statistical significance

Summary:

- **Epistemological concepts:**
  falsifiability, law of parsimony,
- **Good scientific practice:**
  reproducibility of research
- **Gathering parsing data, API:**
  data munging or wrangling, data jujitsu
- **Formulating and testing a scientific Hypothesis**
  1. How to go from idea to Null and Alternate Hypothesis
  2. Establishing the significant of a result through p-vale
  3. Statistical Distributions

- IDEA
- dataset
    - define ideal data
    - figure out best data available
    - figure out if you can get new data
    - obtain data (including policy issues + technical issues)
- data handling
    - joining databases
    - formatting data
- exploratory data analysis
    - machine learning (clustering? dimensionality reduction?)
- statistics
    - models (regression)
    - prediction
    - validation (simulations)
- interpretation
- presentation
    - visualization
    - write a paper!



• IDEA **FORMULATING HYPOTHESIS**

• dataset

  • define ideal data
  • figure out best data available
  • figure out if you can get new data
  • obtain data (including policy issues + technical issues)

• data handling

  • joining databases
  • formatting data

DATA WRANGLING

• exploratory data analysis

  • machine learning (clustering? dimensionality reduction?)

• statistics

  • models (regression)
  • prediction
  • validation (simulations)

HYPOTHESIS TESTING

• interpretation

• presentation

  • visualization
  • write a paper!

CUSP

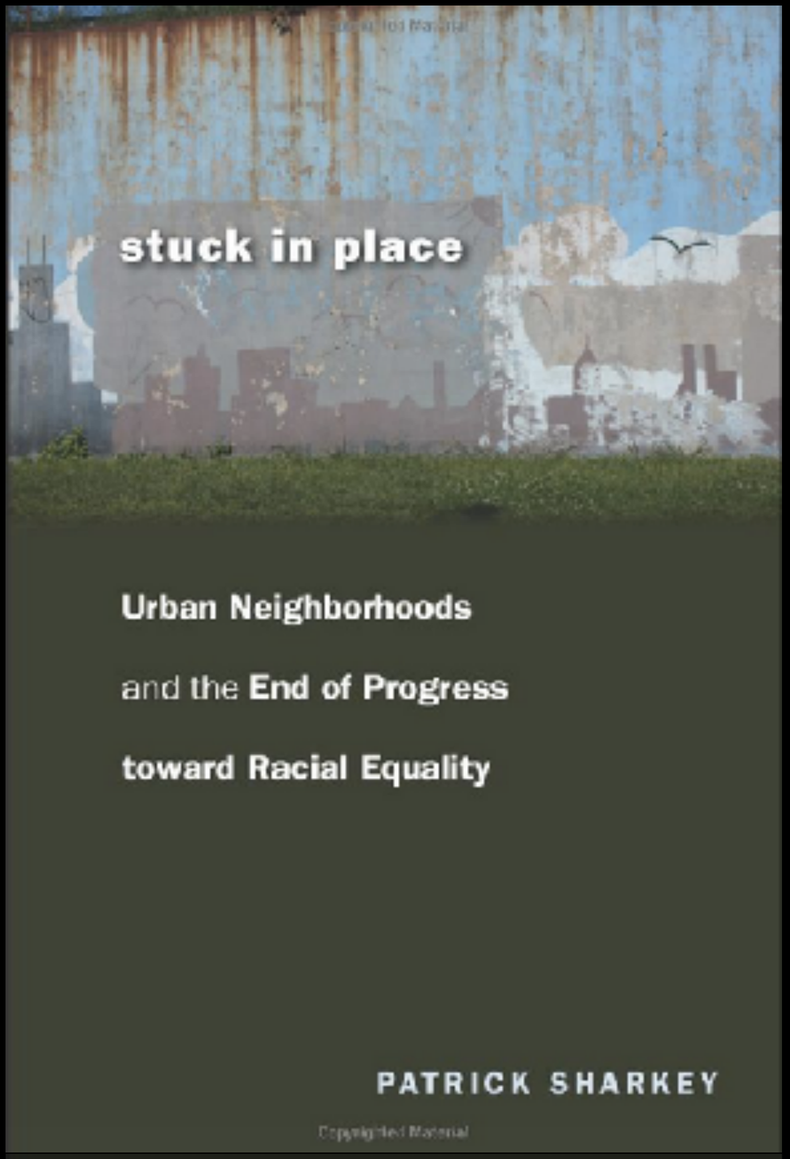
Part I

FORMULATING HYPOTHESIS

idea → *Null Hypothesis*

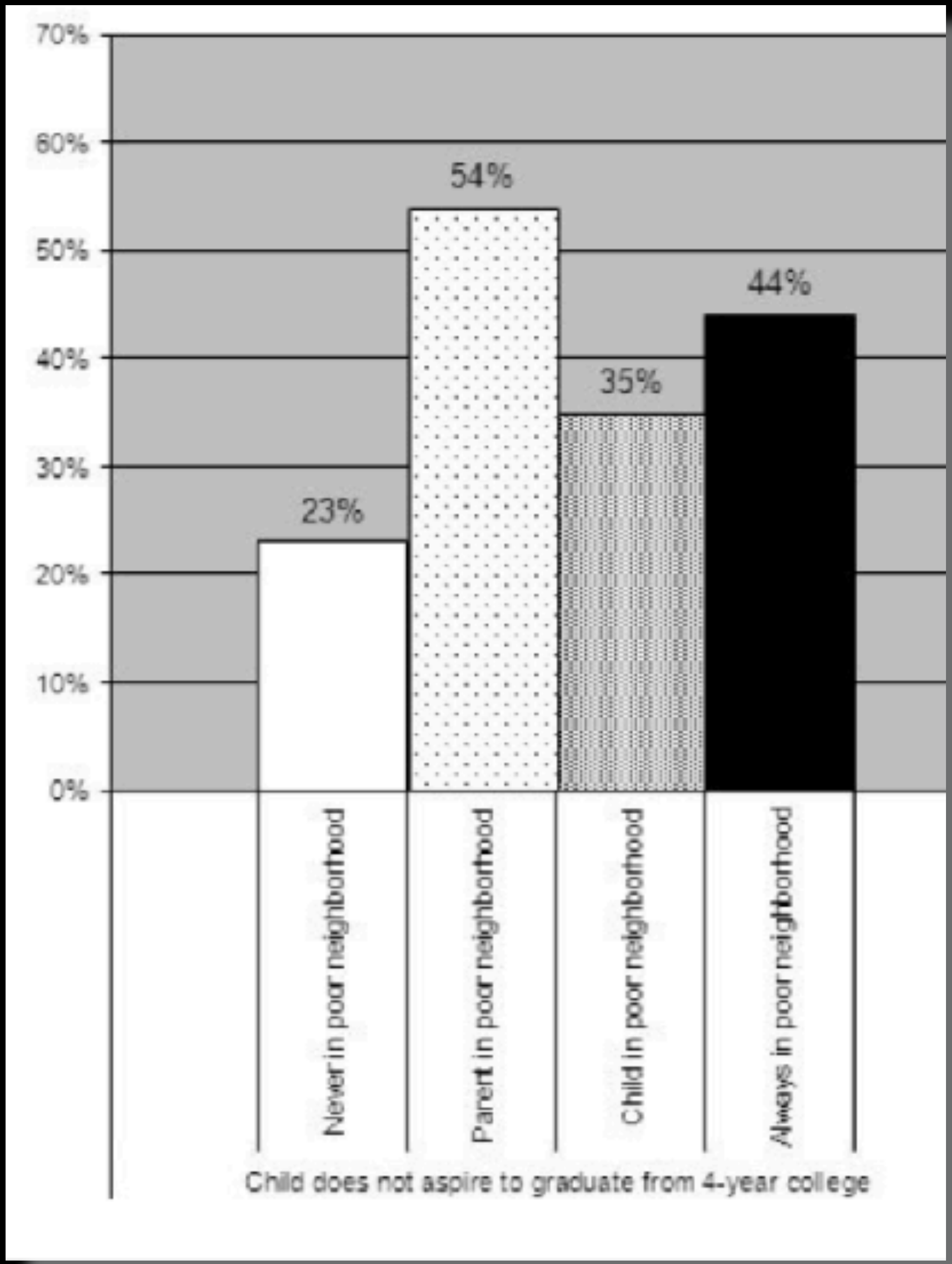
CUSP seminar
Friday 9/18/15

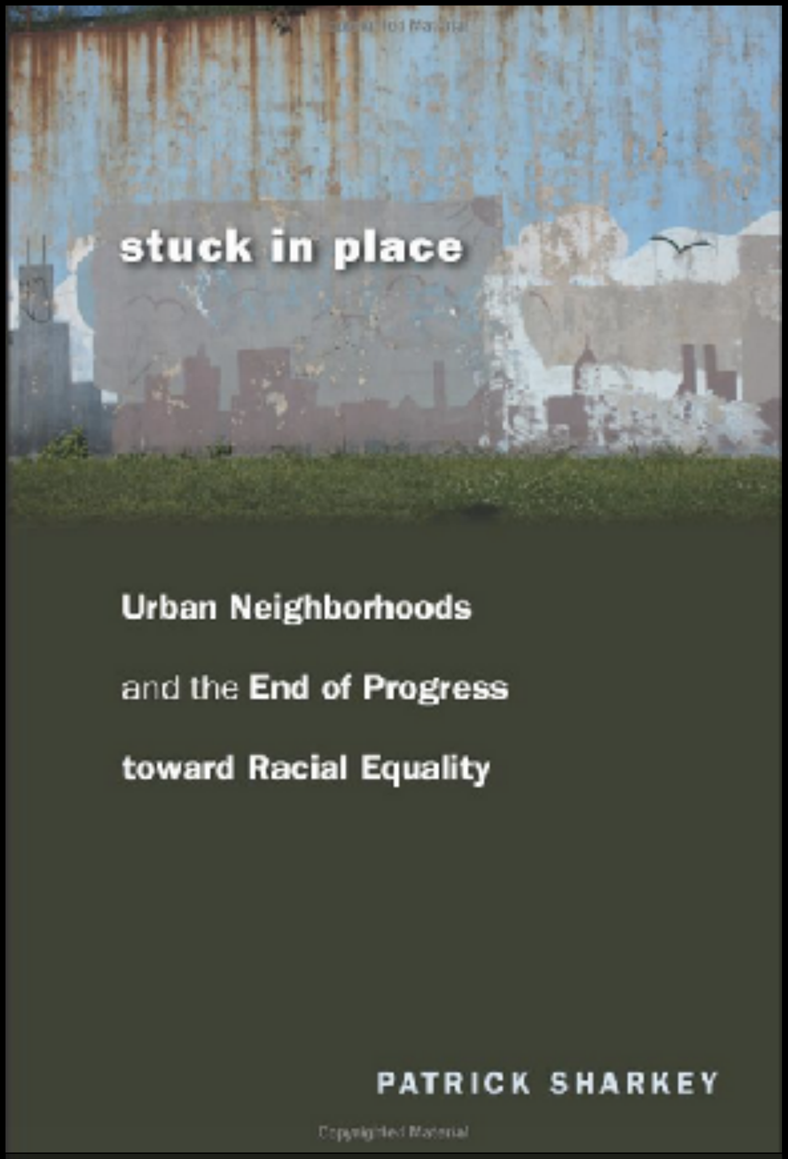
Stuck in Place
Urban Neighborhoods and the end of Progress Toward Racial Equality



III: Introduction to statistics

CUSP seminar
Friday 9/18/15

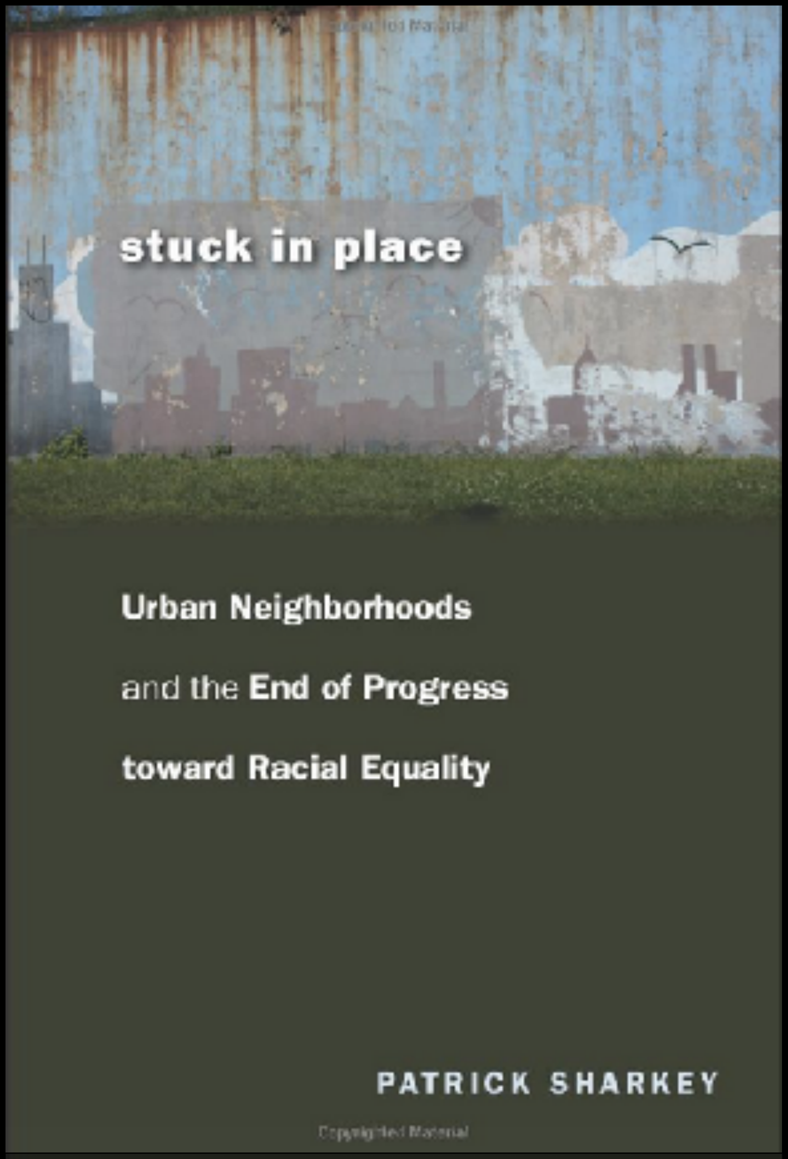




III: Introduction to statistics

1. develop a hypothesis that can be tested mathematically & state the *Null hypothesis* and alternative hypothesis

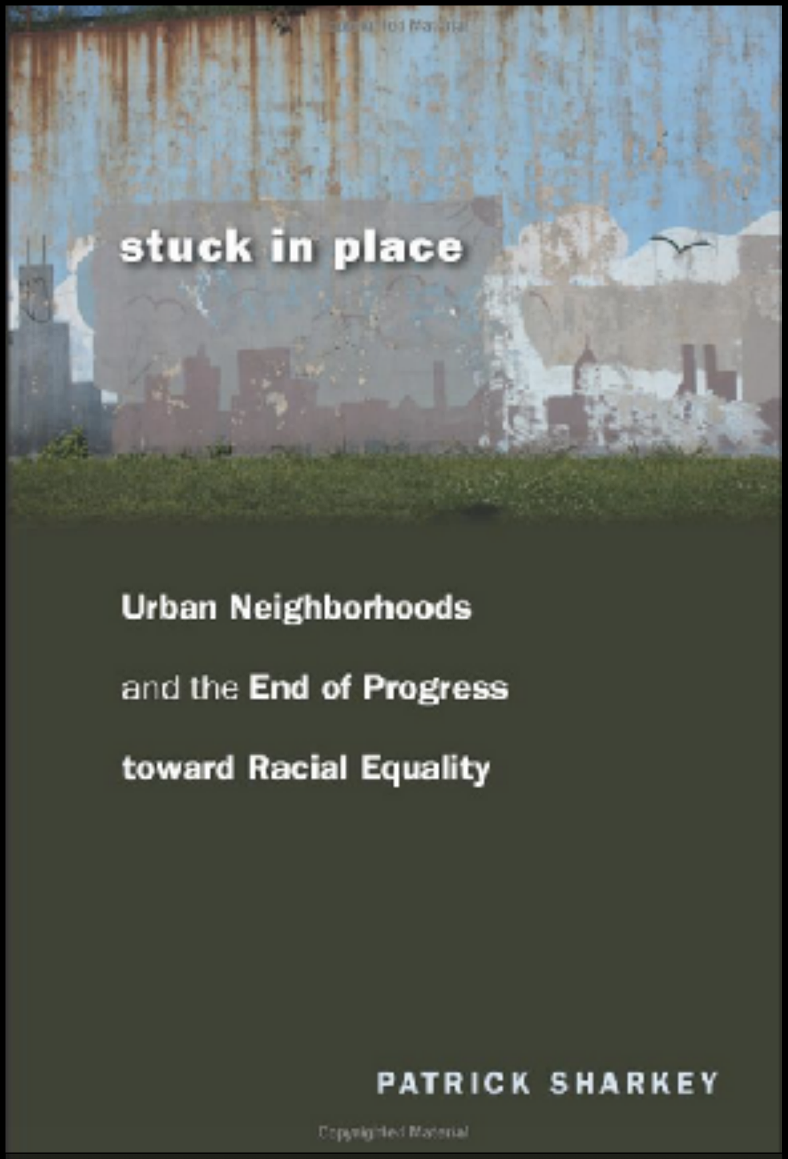
CUSP seminar Friday 9/18/15



1. develop a hypothesis that can be tested mathematically & state the *Null hypothesis* and alternative hypothesis

e.g.:

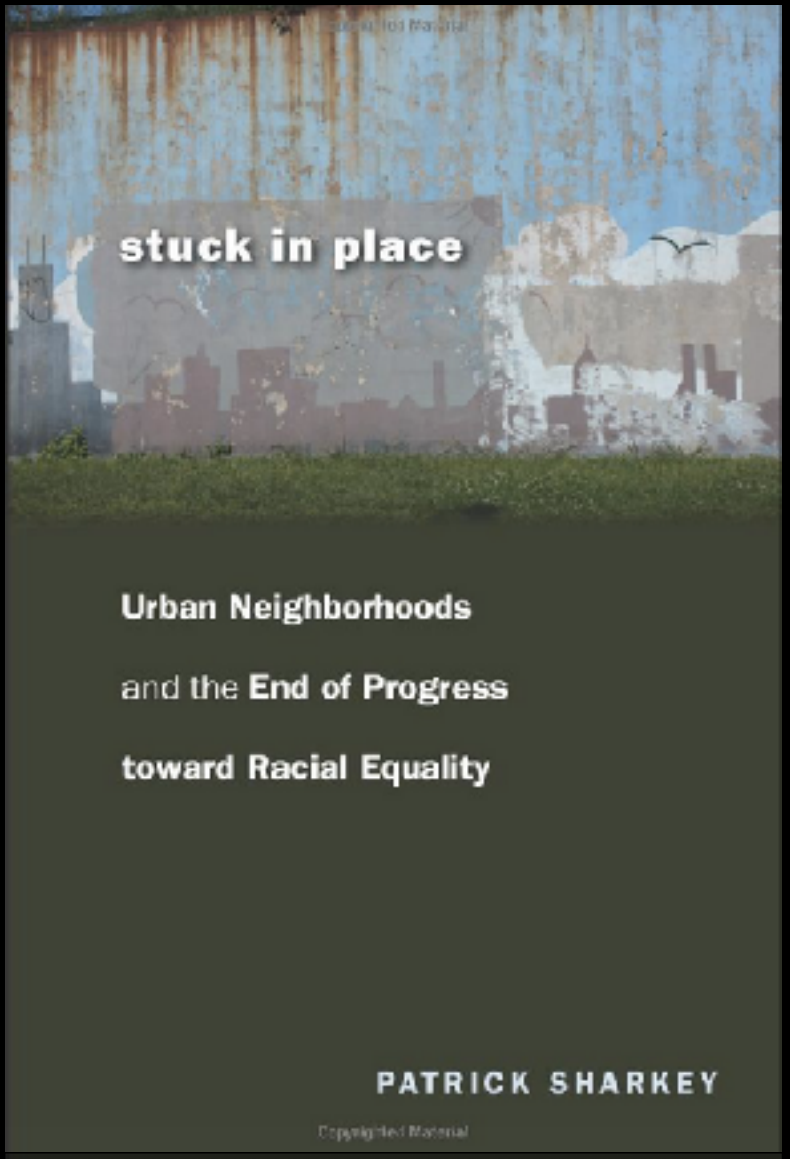
QUESTION: does proximity to violence affect children's development ?



1. develop a hypothesis that can be tested mathematically & state the *Null hypothesis* and alternative hypothesis

QUESTION: does proximity to violence affect children's development?

HYPOTHESIS: the reading test score of children who live near the site of a violent crime is lower after the crime occurred

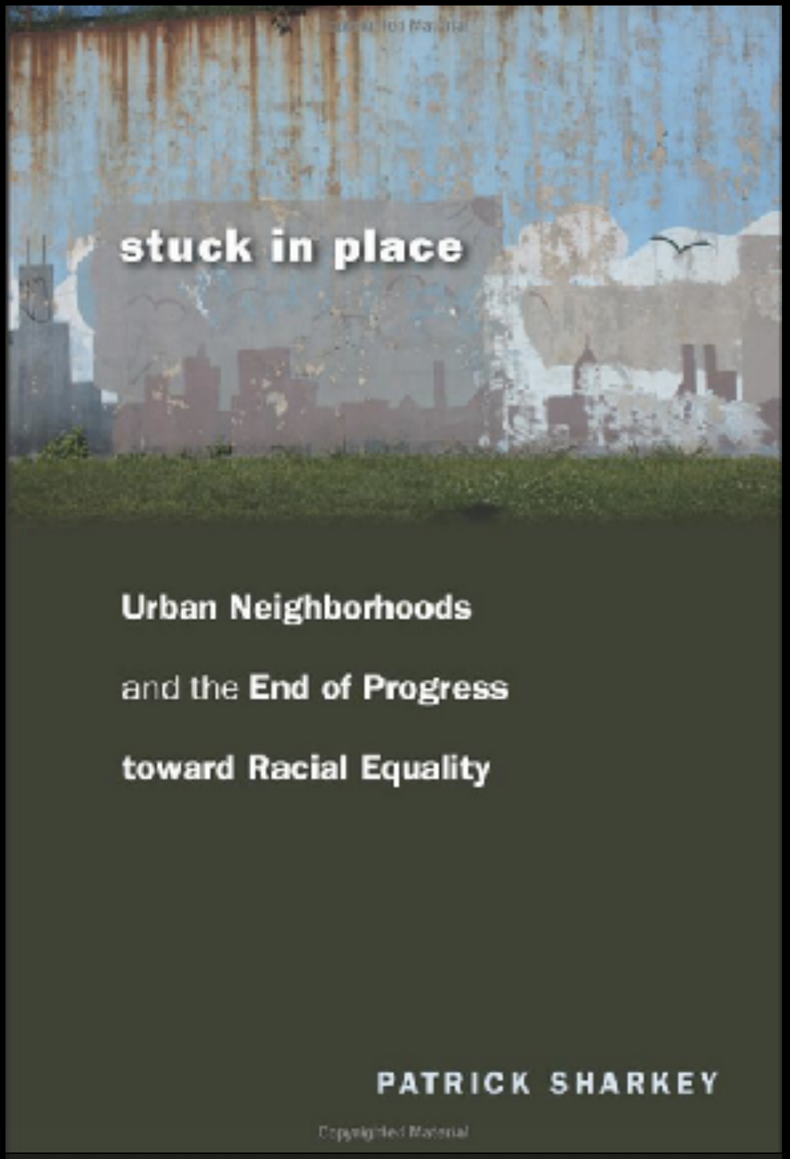


1. develop a hypothesis that can be tested mathematically & state the *Null hypothesis* and alternative hypothesis

QUESTION: does proximity to violence affect children's development?

HYPOTHESIS: the reading test score of children who live near the site of a violent crime is lower after the crime occurred

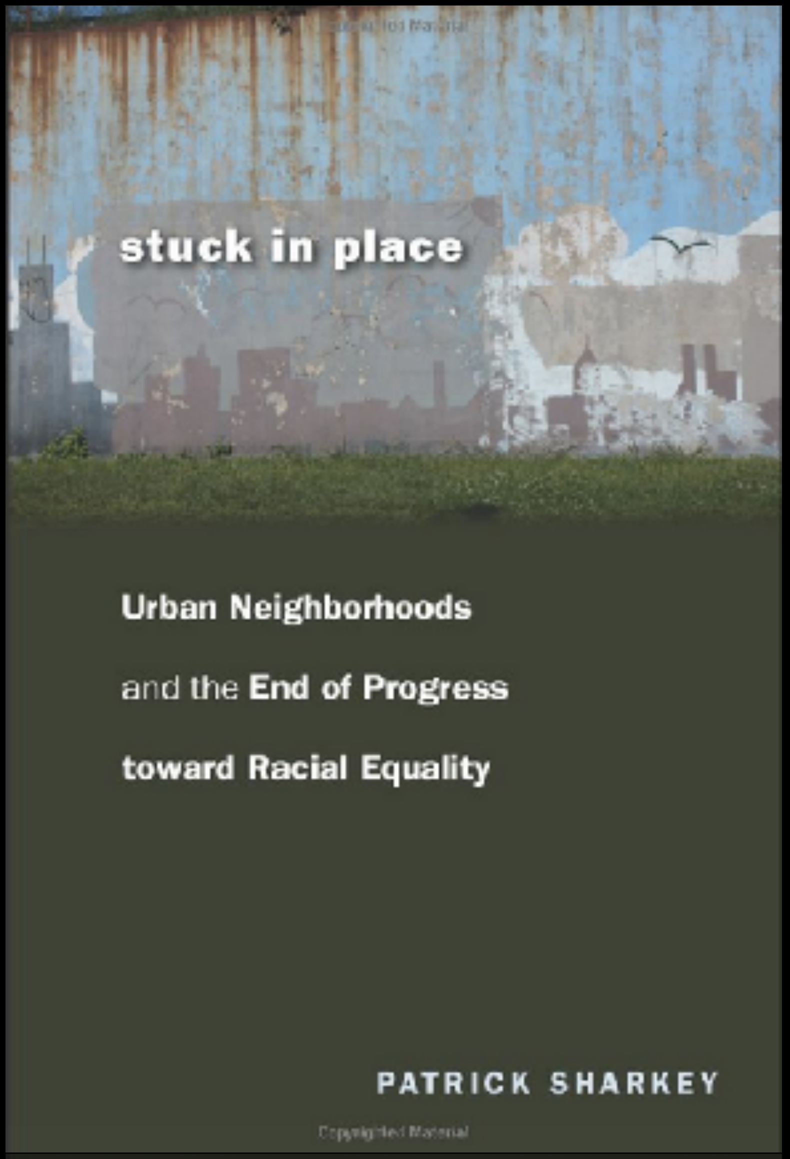
***TESTABLE* HYPOTHESIS**: the *average* test score of children who live *within a block of* the site of a violent crime is *significantly* lower in the days following the crime



1. develop a hypothesis that can be tested mathematically & state the *Null hypothesis* and alternative hypothesis

***NULL* HYPOTHESIS**: the average reading test score of children who live within a block of the site of a violent crime is *the same or higher* than the average score for the control group in the days following the crime, *significance level $p=0.05$*

***ALTERNATIVE* HYPOTHESIS**: the *average* test score of children who live *within a block of* the site of a violent crime is *significantly* lower in the days following the crime



1. develop a hypothesis that can be tested mathematically & state the *Null hypothesis* and alternative hypothesis

IDEA: does the NYC Post-Prison Employment Programs increase employment?

What Strategies Work for the Hard-to-Employ?

Final Results of the Hard-to-Employ Demonstration and Evaluation Project and Selected Sites from the Employment Retention and Advancement Project

OPRE Report 2012-08

March 2012

goo.gl/h2RfkS

1. develop a hypothesis that can be tested mathematically & state the *Null hypothesis* and alternative hypothesis

QUESTION: does the NYC Post-Prison Employment Programs increase employment?

HYPOTHESIS: the number of former prisoners employed 3 years after release is higher for candidates who participated in the program

1. develop a hypothesis that can be tested mathematically & state the *Null hypothesis* and alternative hypothesis

QUESTION: does the NYC Post-Prison Employment Programs increase employment?

HYPOTHESIS: the number of former prisoners employed 3 years after release is higher for candidates who participated in the program

***TESTABLE* HYPOTHESIS**: the % of former prisoners employed 3 years after release is *significantly* higher for candidates who participated in the program

1. develop a hypothesis that can be tested mathematically & state the *Null hypothesis* and alternative hypothesis

***NULL* HYPOTHESIS**: the % of former prisoners employed 3 years after release is *the same or lower* for candidates who participated in the program as for the control group, *significance level p=0.05*

***ALTERNATIVE* HYPOTHESIS**: the % of former prisoners employed 3 years after release is *significantly* higher for candidates who participated

The Hard-to-Employ evaluation was a 10-year study that used a rigorous random assignment research design in four sites to evaluate innovative strategies aimed at improving employment and other outcomes for groups who face serious barriers to employment. The

The Programs in the Hard-to-Employ Evaluation

Following discussions with HHS and extensive research about the implications of different targeting strategies, program models, and best practices for the evaluation design, the MDRC team recruited four sites to participate in the Hard-to-Employ study. Three of the four participating sites targeted discrete hard-to-employ populations, while the fourth (Kansas and Missouri Enhanced Early Head Start) served low-income parents with very young children, a population with more general barriers to finding and keeping jobs:

- **Center for Employment Opportunities, New York City.** Parolees were placed in temporary paid jobs at work sites around the city for several months and received a variety of other supports, along with job placement assistance.

Study Design

For each program presented in this report, the research teams studied the implementation of the programs and the programs' impacts. Additionally, the study of the Center for Employment Opportunities included a benefit-cost analysis, and the studies of the other three HtE programs included estimates of their financial costs.

Study participants at each site were assigned at random to either a program group, which had access to the program's services, or to a control group, which was not permitted to receive program services but could receive any public services that were normally available. The two research groups together make up the "research sample" or "study sample." A random assignment (experimental) design ensures that there are no systematic differences between the members of the two groups when they enter the study, so that any significant differences (that is, differences that are unlikely to arise by chance alone) that emerge over time between the groups can be reliably attributed to the fact that one group was exposed to the experimental program and the other was not. Such differences are known as impacts, or effects, of the program.

hypothesis testing

One vs Two tailed tests

null hypothesis: one-tail-

the phenomenon measures more (less) for one group than for the other

if you have a test control sample: test sample is the same or better than control sample $P_0>=P_1$ or $P_0<=P_1$

falsify the null hypothesis: do you see an improvement (worstening)?

is your test sample better/worse?

hypothesis testing

One vs Two tailed tests

null hypothesis: two-tails -

no relationship between two measured phenomena,
or no difference among groups
if you have a test control sample: test sample and control sample are the same - no effect $P_0=P_1$

falsify the null hypothesis: do you see an effect?
do you see a difference b/w samples?

***NULL* HYPOTHESIS**: the % of former prisoners employed 3 years after release is *the same or lower* for candidates who participated in the program as for the control group,

significance level p=0.05

What Strategies Work for the Hard-to-Employ?

Final Results of the Hard-to-Employ Demonstration and Evaluation Project and Selected Sites from the Employment Retention and Advancement Project

OPRE Report 2012-08

March 2012

http://www.mdrc.org/sites/default/files/What%20Strategies%20Work%20for%20the%20Hard%20FR.pdf

***NULL* HYPOTHESIS**: the % of former prisoners employed 3 years after release is *the same or lower* for candidates who participated in the program as for the control group,

significance level p=0.05

The Enhanced Services for the Hard-to-Employ Demonstration and Evaluation Project

Table 2.1

Summary of Impacts, New York City Center for Employment Opportunities

Outcome	Program Group	Control Group	Difference (Impact)	P-Value
Employment (Years 1-3) (%)				
Ever employed	83.8	70.4	13.4 ***	0.000
Ever employed in a CEO transitional job[a]	70.1	3.5	66.6 ***	0.000
Ever employed in an unsubsidized job	63.7	69.0	-5.3 *	0.078
Postprogram unsubsidized employment (Years 2-3)				
Ever employed in an unsubsidized job (%)	53.3	52.1	1.2	0.713
Employed in an unsubsidized job, average per quarter (%)	28.2	27.2	1.1	0.618
Employed for six or more consecutive quarters (%)	14.7	11.9	2.8	0.195
Total UI-covered earnings[b] ($)	10,435	9,846	589	0.658
Sample size (total = 973)[c]	564	409		

SOURCES: MDRC earnings calculations from the National Directory of New Hires (NDNH) database and employment calculations from the unemployment insurance (UI) wage records from New York State, MDRC calculations using data from the New York State Division of Criminal Justice Services (DCJS) and the New York City Department of Correction (DOC).

NOTES: Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.
The p-value indicates the likelihood that the difference between the program and control groups arose by chance.

http://www.mdrc.org/sites/default/files/What%20Strategies%20Work%20for%20the%20Hard%20FR.pdf

***NULL* HYPOTHESIS**: the % of former prisoners employed 3 years after release is *the same or lower* for candidates who participated in the program as for the control group,

significance level p=0.05

The Enhanced Services for the Hard-to-Employ Demonstration and Evaluation Project

Table 2.1

Summary of Impacts, New York City Center for Employment Opportunities

Outcome	Program Group	Control Group	Difference (Impact)	P-Value
Employment (Years 1-3) (%)	P_1	P_0		
Ever employed	83.8	70.4	13.4 ***	0.000
Ever employed in a CEO transitional job[a]	70.1	3.5	66.6 ***	0.000
Ever employed in an unsubsidized job	63.7	69.0	-5.3 *	0.078
Postprogram unsubsidized employment (Years 2-3)				
Ever employed in an unsubsidized job (%)	53.3	52.1	1.2	0.713
Employed in an unsubsidized job, average per quarter (%)	28.2	27.2	1.1	0.618
Employed for six or more consecutive quarters (%)	14.7	11.9	2.8	0.195
Total UI-covered earnings[b] ($)	10,435	9,846	589	0.658
Sample size (total = 973)[c]	564	409		

SOURCES: MDRC earnings calculations from the National Directory of New Hires (NDNH) database and employment calculations from the unemployment insurance (UI) wage records from New York State, MDRC calculations using data from the New York State Division of Criminal Justice Services (DCJS) and the New York City Department of Correction (DOC).

NOTES: Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.
The p-value indicates the likelihood that the difference between the program and control groups arose by chance.

$$H_0: P_0 - P_1 >= 0$$

$$H_a: P_0 - P_1 < 0$$

$$\alpha = 0.05$$

http://www.mdrc.org/sites/default/files/What%20Strategies%20Work%20for%20the%20Hard%20FR.pdf

***NULL* HYPOTHESIS**: the % of former prisoners employed 3 years after release is *the same or lower* for candidates who participated in the program as for the control group,

significance level p=0.05

The Enhanced Services for the Hard-to-Employ Demonstration and Evaluation Project

Table 2.1

Summary of Impacts, New York City Center for Employment Opportunities

Outcome	Program Group	Control Group	Difference (Impact)	P-Value
Employment (Years 1-3) (%)	P_1	P_0		
Ever employed	83.8	70.4	13.4 ***	0.000
Ever employed in a CEO transitional job[a]	70.1	3.5	66.6 ***	0.000
Ever employed in an unsubsidized job	63.7	69.0	-5.3 *	0.078
Postprogram unsubsidized employment (Years 2-3)				
Ever employed in an unsubsidized job (%)	53.3	52.1	1.2	0.713
Employed in an unsubsidized job, average per quarter (%)	28.2	27.2	1.1	0.618
Employed for six or more consecutive quarters (%)	14.7	11.9	2.8	0.195
Total UI-covered earnings[b] ($)	10,435	9,846	589	0.658
Sample size (total = 973)[c]	564	409		

SOURCES: MDRC earnings calculations from the National Directory of New Hires (NDNH) database and employment calculations from the unemployment insurance (UI) wage records from New York State, MDRC calculations using data from the New York State Division of Criminal Justice Services (DCJS) and the New York City Department of Correction (DOC).

NOTES: Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.
The p-value indicates the likelihood that the difference between the program and control groups arose by chance.

H_0: $P_0 - P_1 >= 0$

H_a: $P_0 - P_1 < 0$

$\alpha = 0.05$

http://www.mdrc.org/sites/default/files/What%20Strategies%20Work%20for%20the%20Hard%20FR.pdf

hypothesis testing

null hypothesis: no relationship between two measured phenomena,
or no difference among groups
if you have a test control sample: test sample and
control sample are the same - no effect

falsify the null hypothesis: do you see an effect?
do you see a difference b/w samples?

hypothesis testing and *significance*

<u>what is the probability that we would have gotten the same result out of just chance?</u>

null hypothesis: no relationship between two measured phenomena,
or no difference among groups
if you have a test control sample: test sample and control sample are the same - no effect

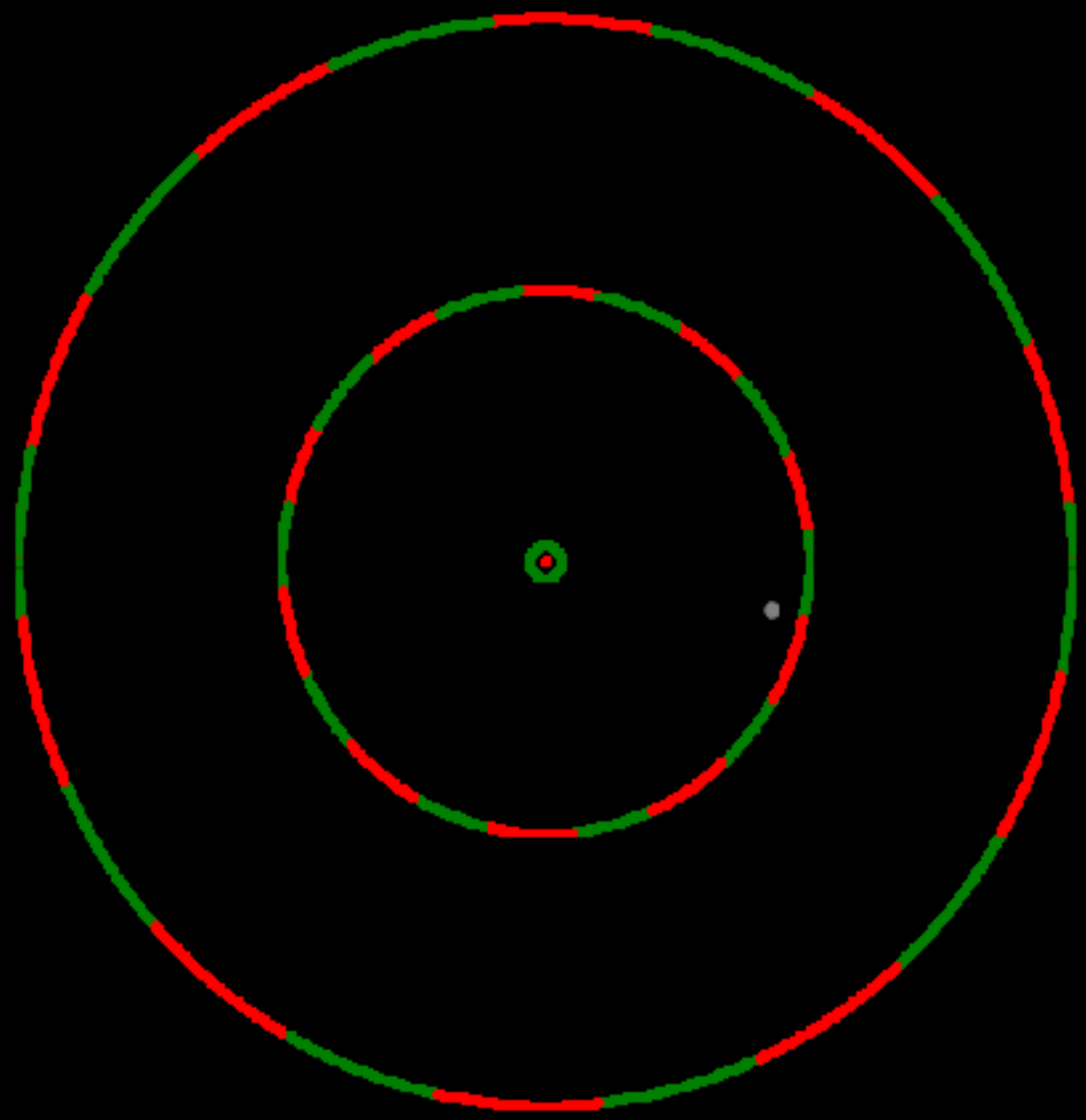
falsify the null hypothesis: do you see an effect?
do you see a difference b/w samples?
<u>is this effect larger than the just by chance?</u>

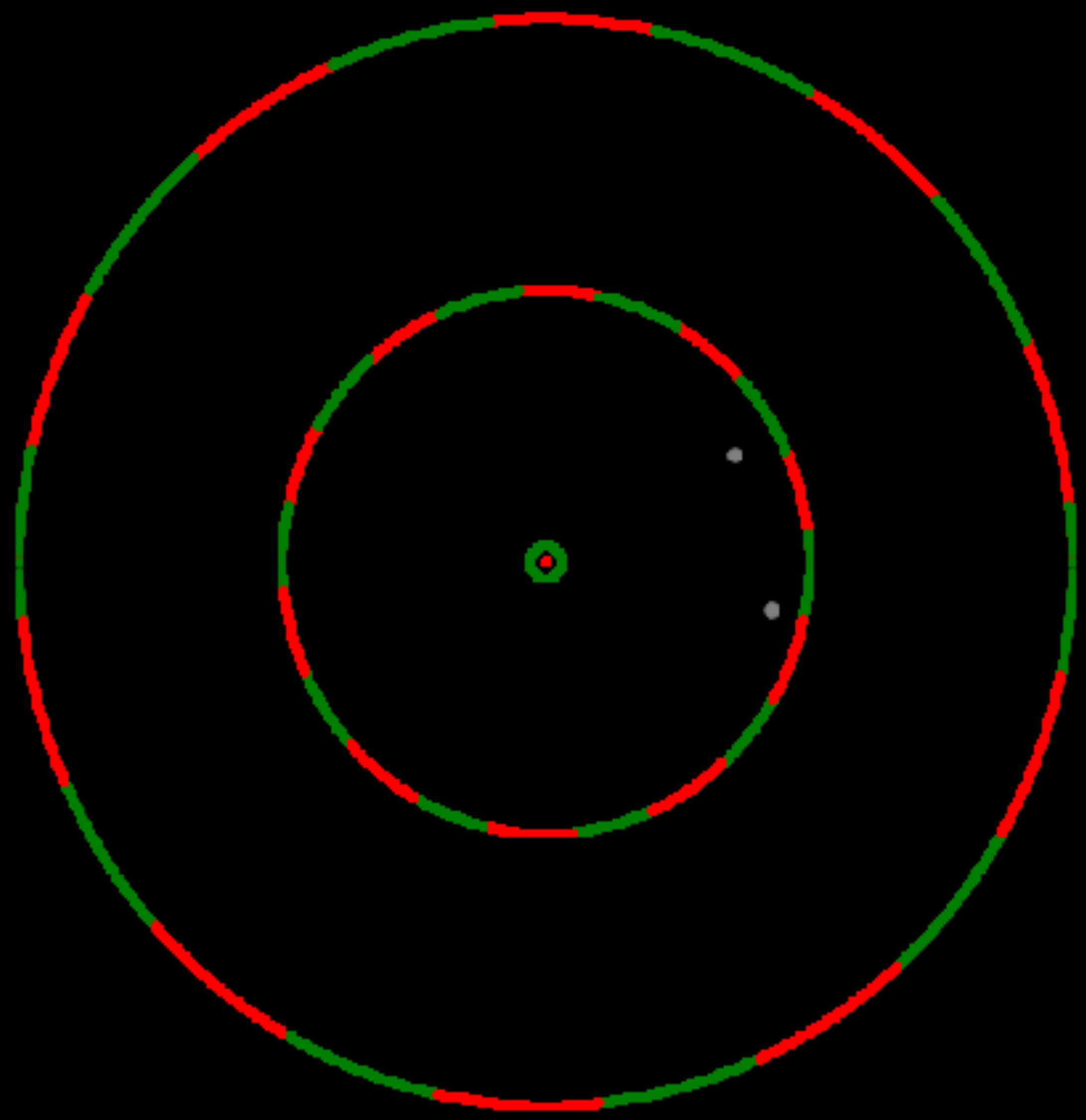
CUSP

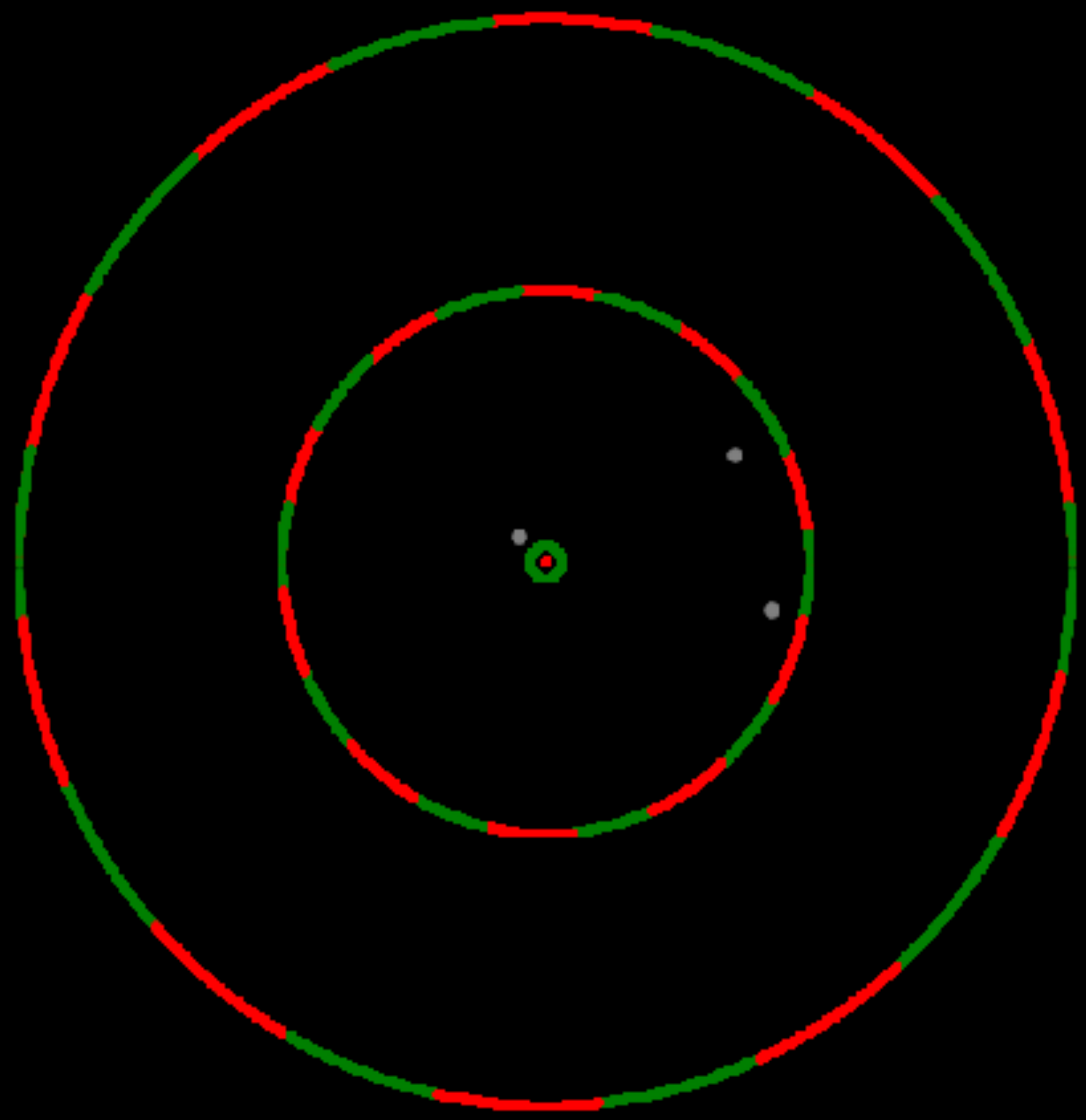
HYPOTHESIS TESTING

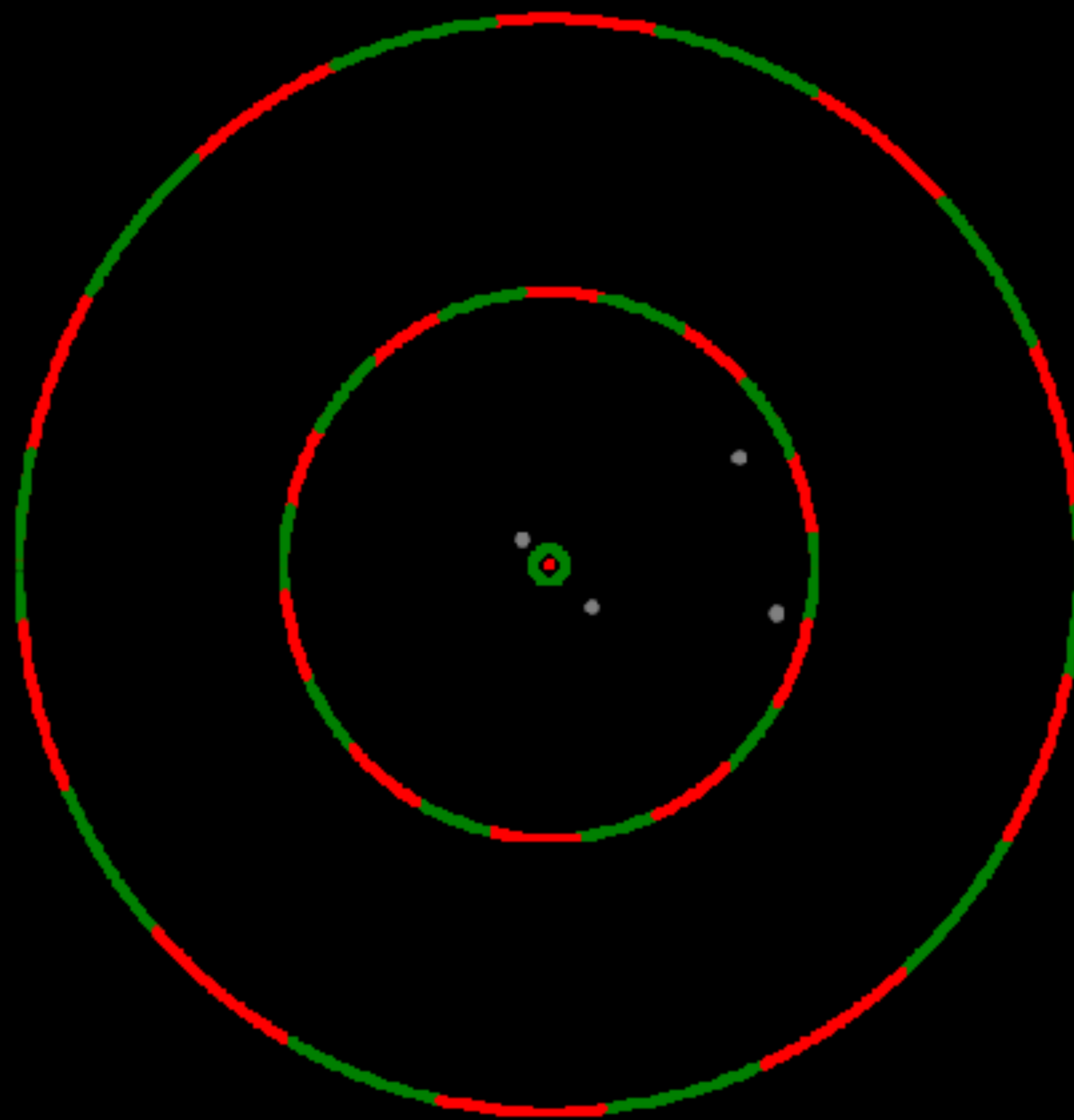
Hypothesis testing ⟺ Statistical Analysis is this effect larger than the just by chance?

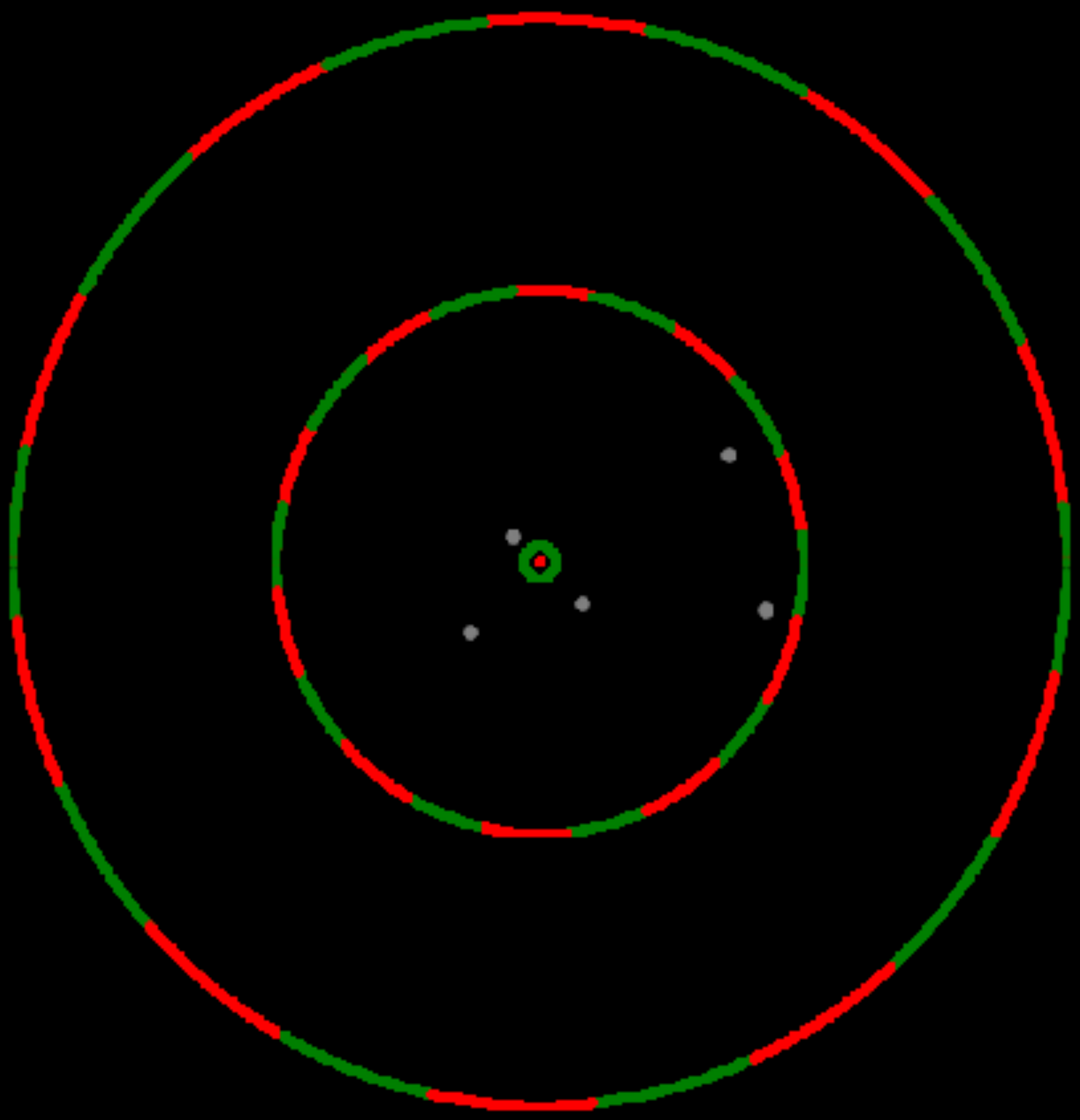
distributions

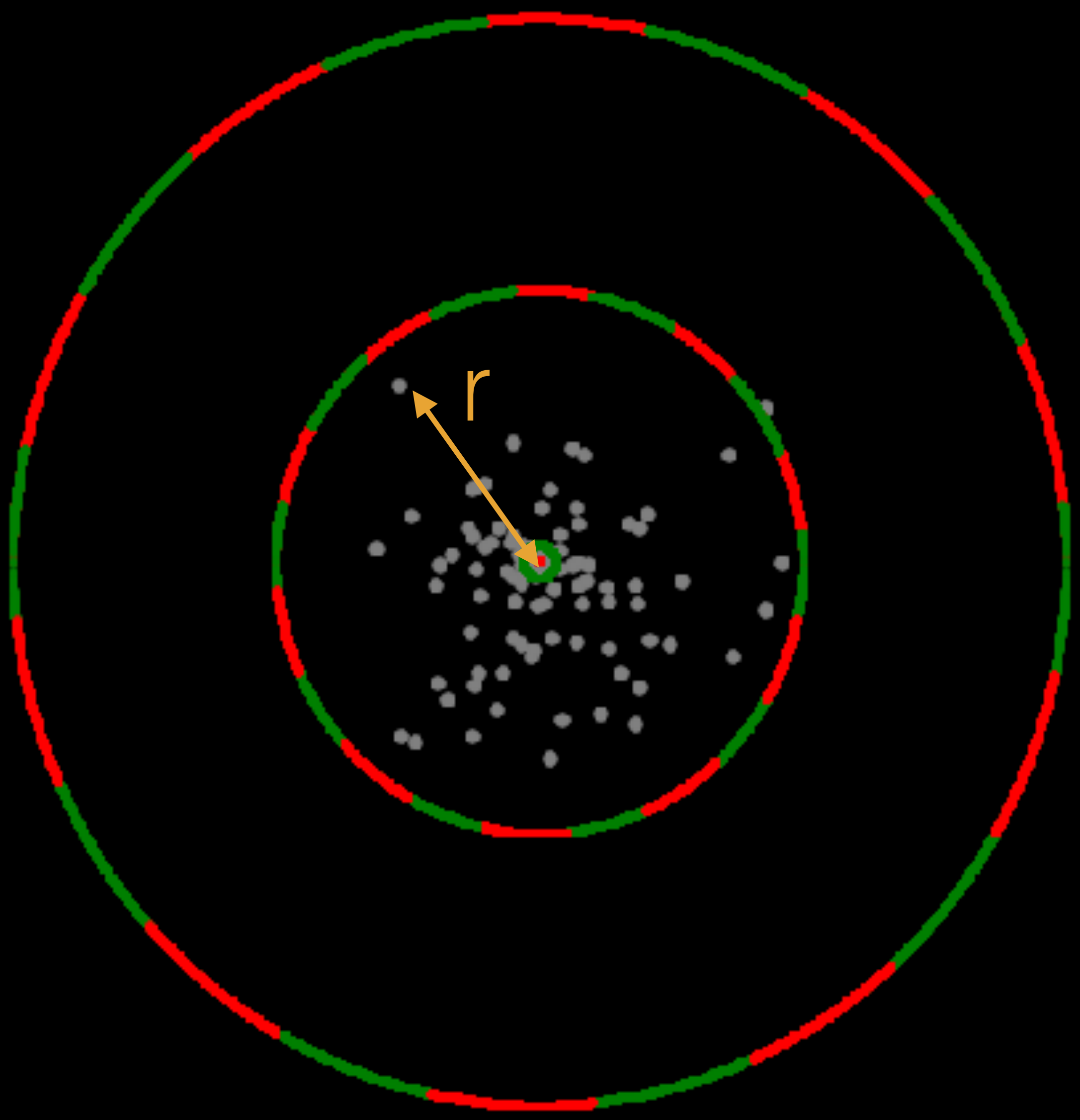


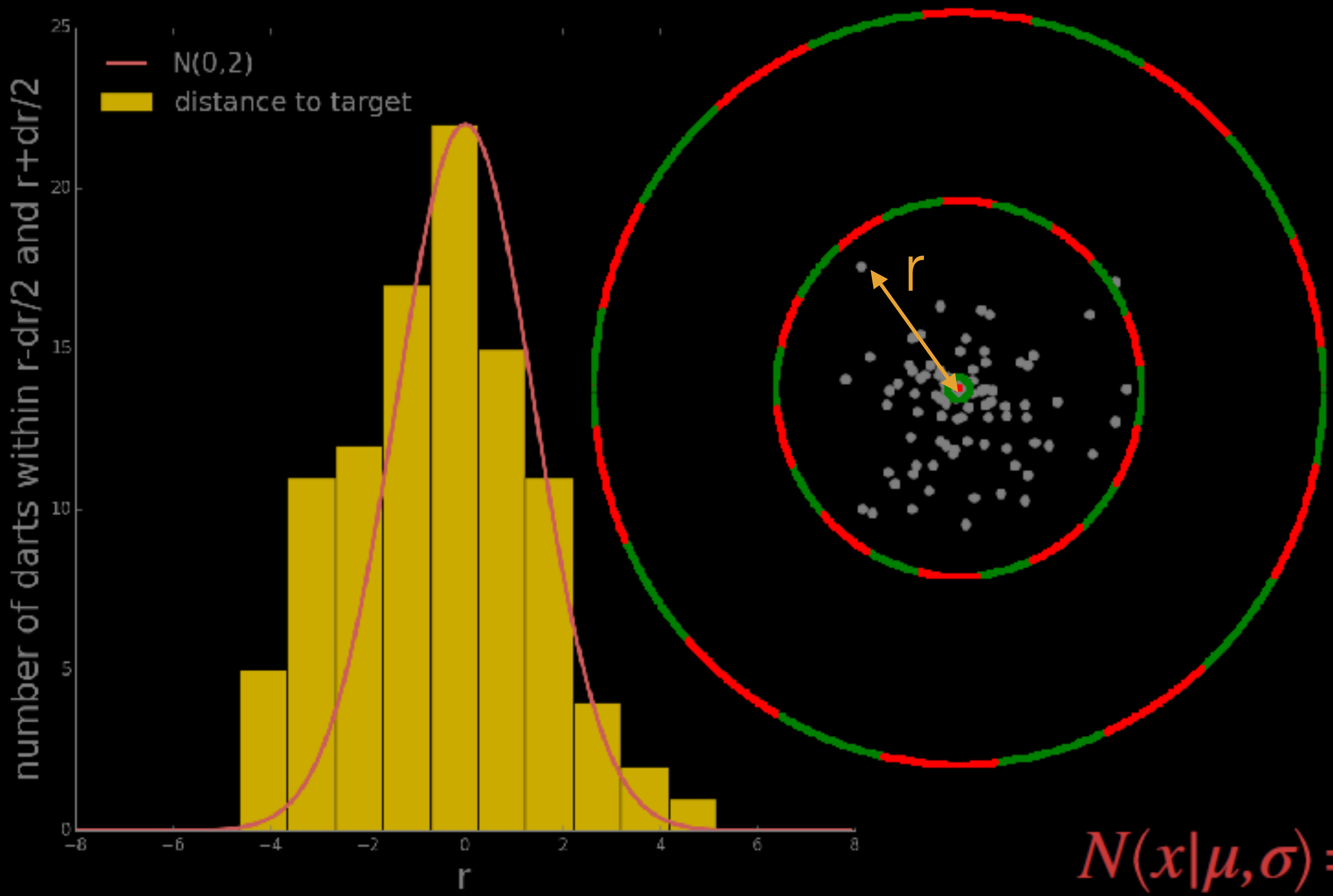









r



$$N(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

III: Introduction to statistics

Random sampling (numpy.random)

Simple random data

rand(d0, d1, ..., dn)	Random values in a given shape.
randn(d0, d1, ..., dn)	Return a sample (or samples) from the "standard normal" distribution.
randint(low[, high, size])	Return random integers from *low* (inclusive) to *high* (exclusive).
random_integers(low[, high, size])	Return random integers between *low* and *high*, inclusive.
random_sample([size])	Return random floats in the half-open interval [0.0, 1.0).
random([size])	Return random floats in the half-open interval [0.0, 1.0).
ranf([size])	Return random floats in the half-open interval [0.0, 1.0).
sample([size])	Return random floats in the half-open interval [0.0, 1.0).
choice(a[, size, replace, p])	Generates a random sample from a given 1-D array
bytes(length)	Return random bytes.

Permutations

Save the figure

shuffle(x)	Modify a sequence in-place by shuffling its contents.
permutation(x)	Randomly permute a sequence, or return a permuted range.

Distributions

beta(a, b[, size])	The Beta distribution over `[0, 1]`.
binomial(n, p[, size])	Draw samples from a binomial distribution.
chisquare(df[, size])	Draw samples from a chi-square distribution.
dirichlet(alpha[, size])	Draw samples from the Dirichlet distribution.
exponential([scale, size])	Exponential distribution.

Table Of Contents

Previous topic

numpy.RankWarning

Next topic

numpy.random.rand

http://docs.scipy.org/doc/numpy/reference/routines.random.html

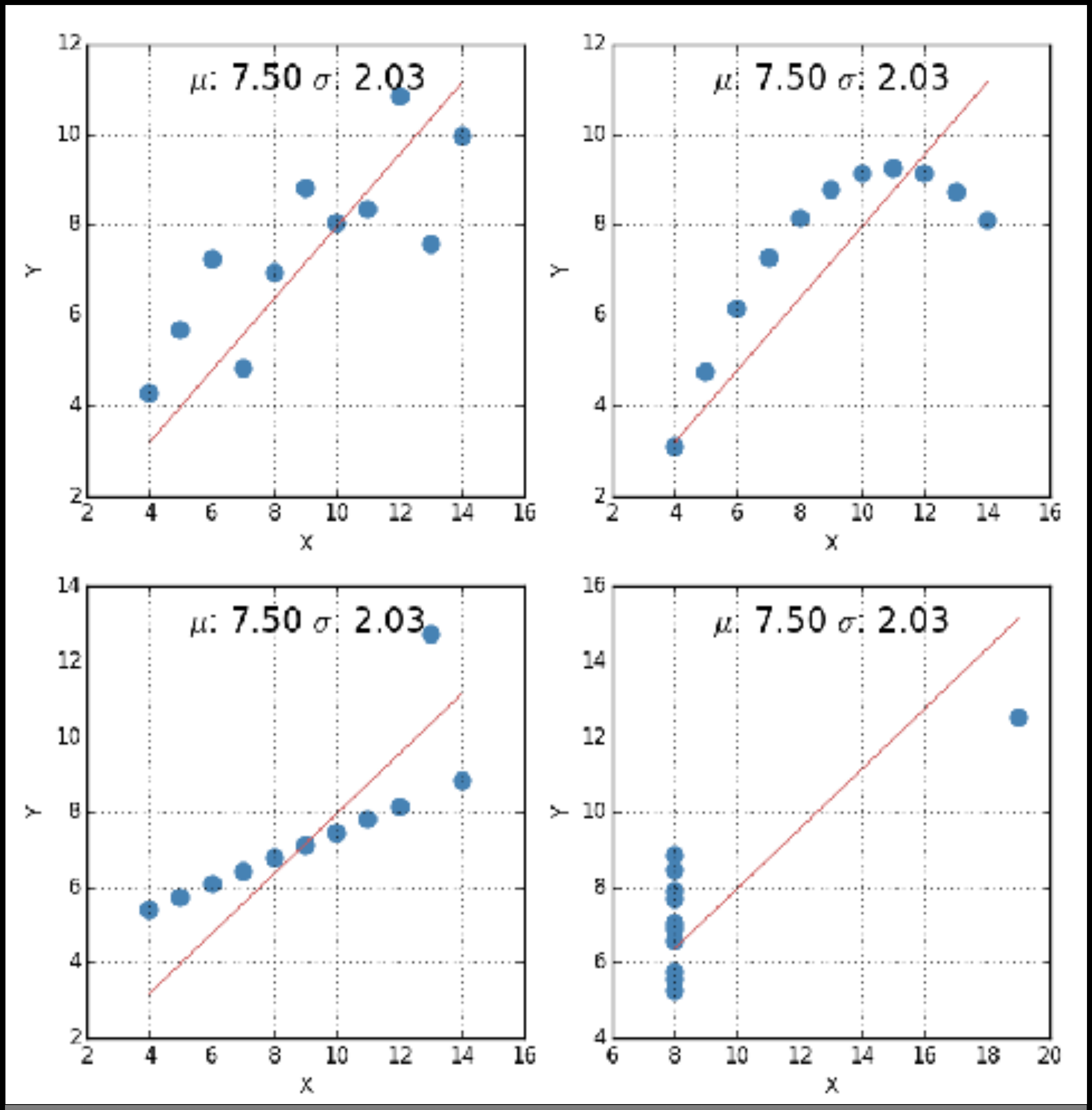
distribution moments

a distribution's moments summarize its properties:

$$m_n = \int_{-\infty}^{\infty} (x-c)^n f(x)\,dx.$$

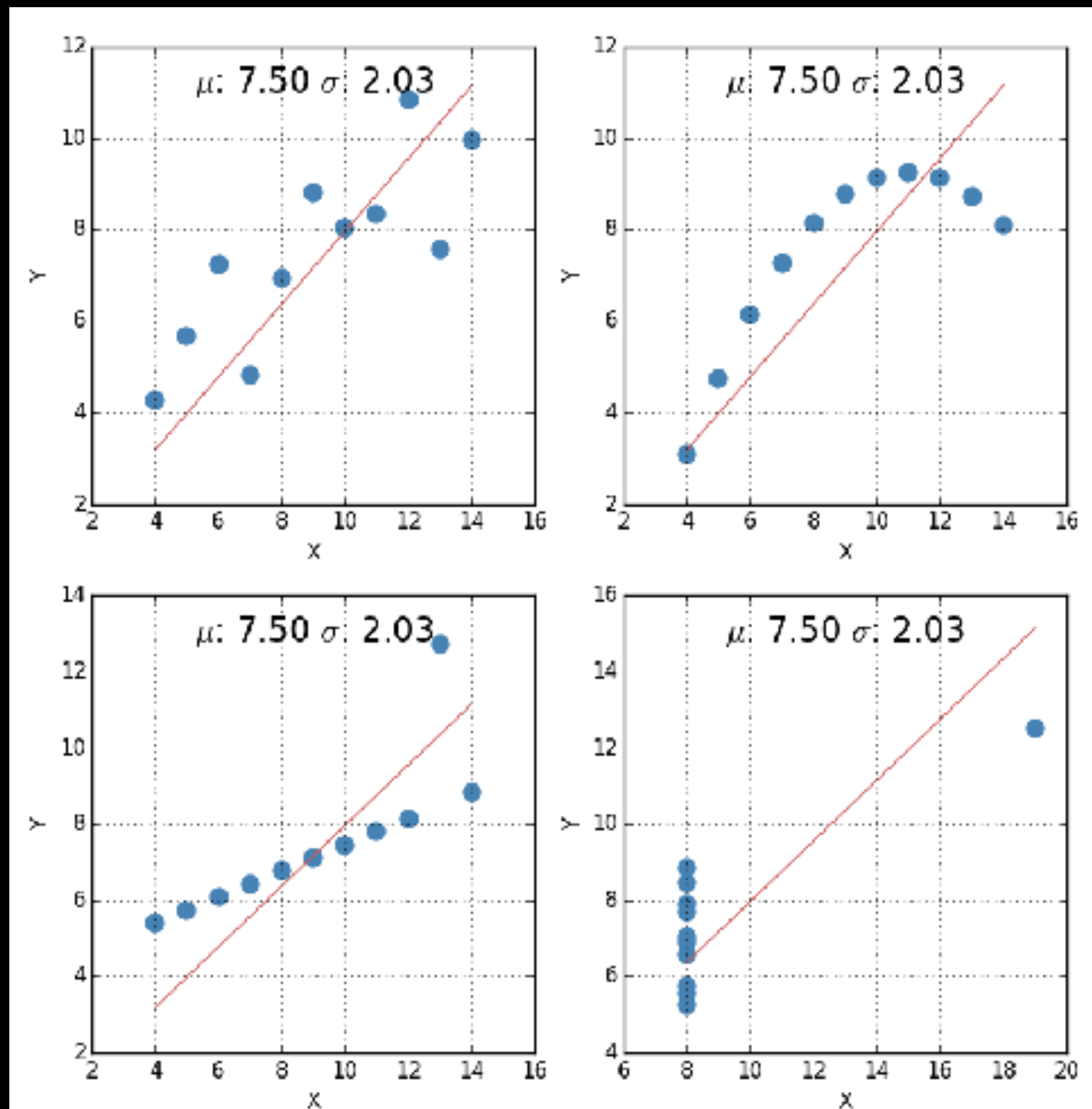
central tendency: mean (n=1), median, mode
spread: standard deviation/variance (n=2), quartiles range
symmetry: skewness (n=3)
CUSPiness: kurtosis (n=4)
...

distribution moments



https://github.com/fedhere/UInotebooks/blob/master/Anscombe's%20Quartet.ipynb

distribution moments



- read in a csv as a pandas DF
- extract statistics from distributions
- various plot styles
- inspection of residuals

https://github.com/fedhere/UInotebooks/blob/master/Anscombe's%20Quartet.ipynb

distributions: Central Limit Theorem

Laplace (1700s)
but also: Poisson, Bessel, Dirichlet, Cauchy, Ellis

Let $X_1 \ldots X_N$ be an N-elements sample from a population whose distribution has **mean μ and standard deviation σ**
In the limit of N -> infinity
the sample mean m approaches a Normal (Gaussian) distribution with mean μ and standard deviation σ
regardless of the distribution of X

$$\bar{x} \sim N\left(\mu, \sigma/\sqrt{N}\right)$$

distributions: Central Limit Theorem

At the root is the fact that a sample drawn from a parent distribution will look increasingly more like the parent distribution as the size of the sample increases.

<u>More formally</u>: The distribution of the means of N samples generated from the same parent distribution will

I. be normally distributed (i.e. will be a Gaussian)

II. have *mean* equal to the *mean of the parent distribution*, and

III. have *standard deviation* equal to the *parent population standard deviation divided by the square root of the sample size*

$$\bar{x} \sim N\ (\mu, \sigma/\sqrt{N}\ )$$

distributions: Central Limit Theorem

In practice:

it is telling us that if we have enough observations in our sample we can assume that the population parameters are equal to the sample statistics within a known uncertainty

$$\bar{x} \sim N\ (\mu, \sigma/\sqrt{N}\ )$$

CUSP

HOMEWORK 1 :

1. GENERATE 100 samples of different sizes N (N>10 & N<2000) from each of 5 different distributions (500 samples in total), all with the same *population* mean. Include a Normal, a Poisson, a Binomial, a Chi-Squared distribution, and 1 more of your choice.
2. For each sample plot the sample mean (dependent var.) against the sample size N (independent var.) (if you want you can do it with the sample standard deviation as well). Describe the behavior you see in the plots in terms of the law of large numbers.
3. PLOT the distributions of all sample means (together for all distributions). Mandatory: as a histogram, optional: in any other way you think is convincing
4. EC: FIT a gaussian to the distribution of means

e.g. how to fit function to data in numpy:

http://glowingpython.blogspot.com/2012/07/distribution-fitting-with-scipy.html

http://stackoverflow.com/questions/7805552/fitting-a-histogram-with-python

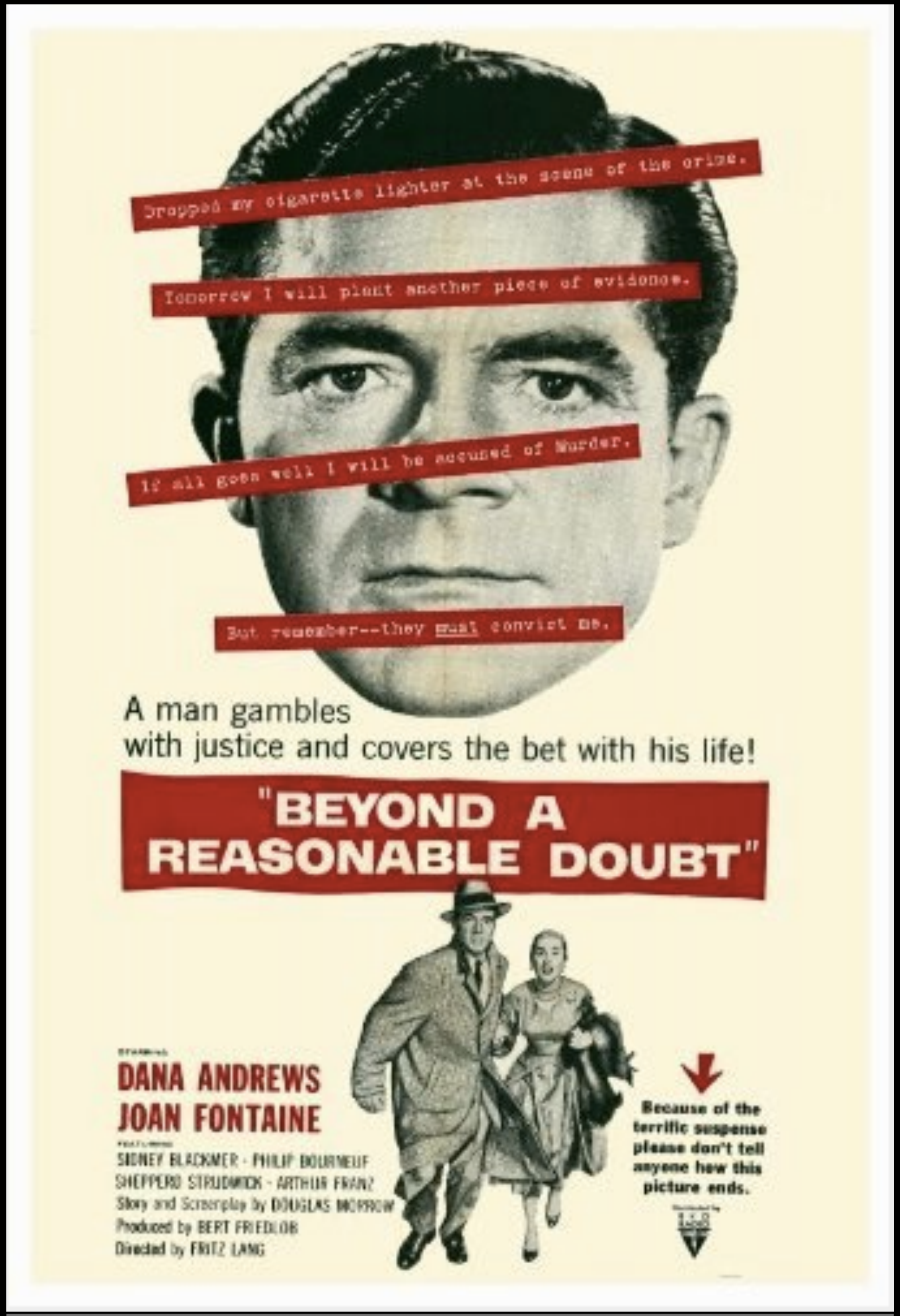
CUSP

distributions, moments, and Central Limit Theorem

HOMEWORK 1 :

- create a new Github directory called HW4_<netID> inside of your PUI2018_<netID> repo.
- upload an ipython notebook, with the rendered plots.
- include in the readme README.md, who you worked with, what you are doing, why, and, if appropriate, how to run the notebook (e.g. global variables that need to be setup?).
- the grade will be based on the rendered notebook. 75% of the grade will be awarded based on the rendered version, *the remaining 25% will be awarded if the TA can download and run the notebook in ADRF*. **When you are finished coding and the notebook looks as you want it to remember to rerun it** for example by clicking on Kernel->restart&run all on Jupyter Hub

rejecting the Null beyond any reasonable doubt



CUSP

Rejecting the Null Hypothesis: what is the *p* value?

is the probability of getting a result <u>at least as extreme as</u> the one you observed just by chance <u>lower than the significance level you established?</u>

Rejecting the Null Hypothesis: what is the *p* **value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

- decide what the significance threshold is: typically 5%

  $\alpha=0.05$

Rejecting the Null Hypothesis: what is the *p* **value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

- decide what the significance threshold is: typically 5% $\alpha=0.05$
- choose a statistical test (T-test, *Z*-test, bayesian analysis…)

Rejecting the Null Hypothesis: what is the *p* **value**?

is the probability of getting a result <u>at least as extreme as</u> the one you observed just by chance <u>lower than the significance level you established?</u>

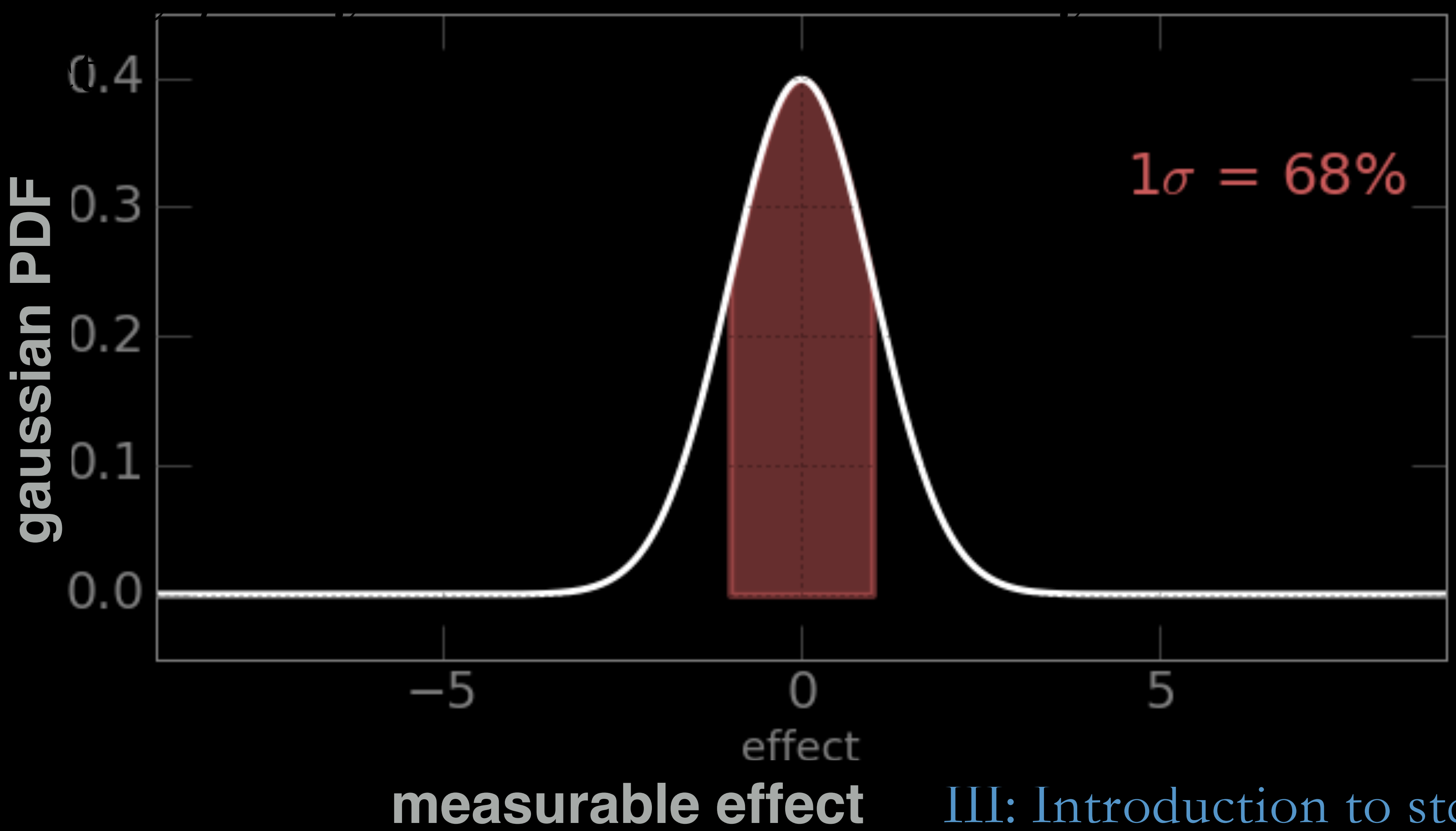
- decide what the significance threshold is: typically 5% α=0.05
- choose a statistical test (T-test, *Z*-test, bayesian analysis…)
- find the probability *p* of your measurement for your test H_a <u>in absence of effect</u>

Rejecting the Null Hypothesis: what is the ***p* value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

All statistical tests rely on assuming the possible results of your measurements (or some quantities derived from it) follow a certain distribution e.g. Gaussian, Poisson ….

T-test, *Z*-test, bayesian analysis…

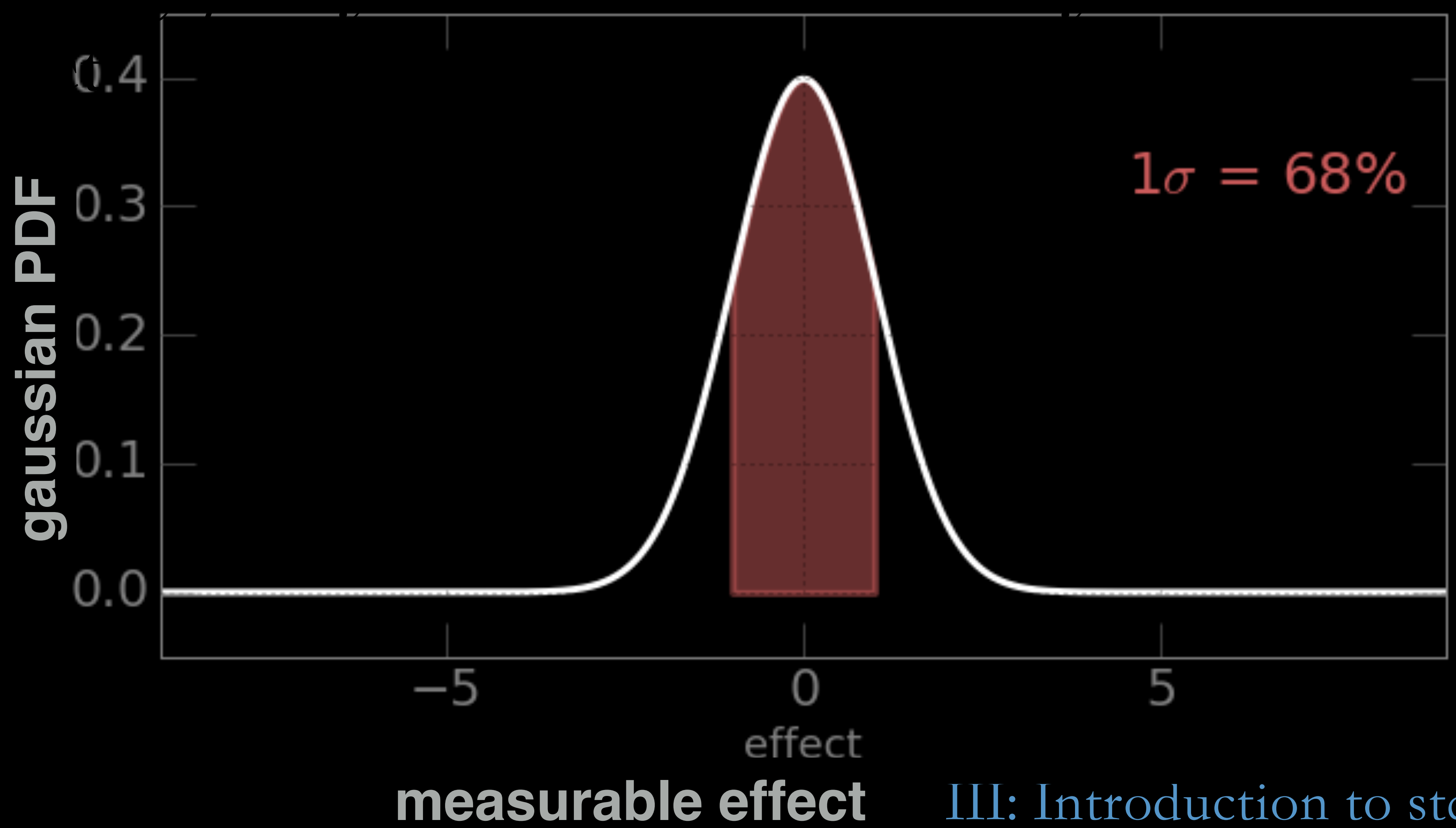


Rejecting the Null Hypothesis: what is the *p* **value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

So you can answer the question:
"I find my measurement fall where the probabiliy of the appropriate distribution is p_{meas}.
Is p_{meas} < p-value?

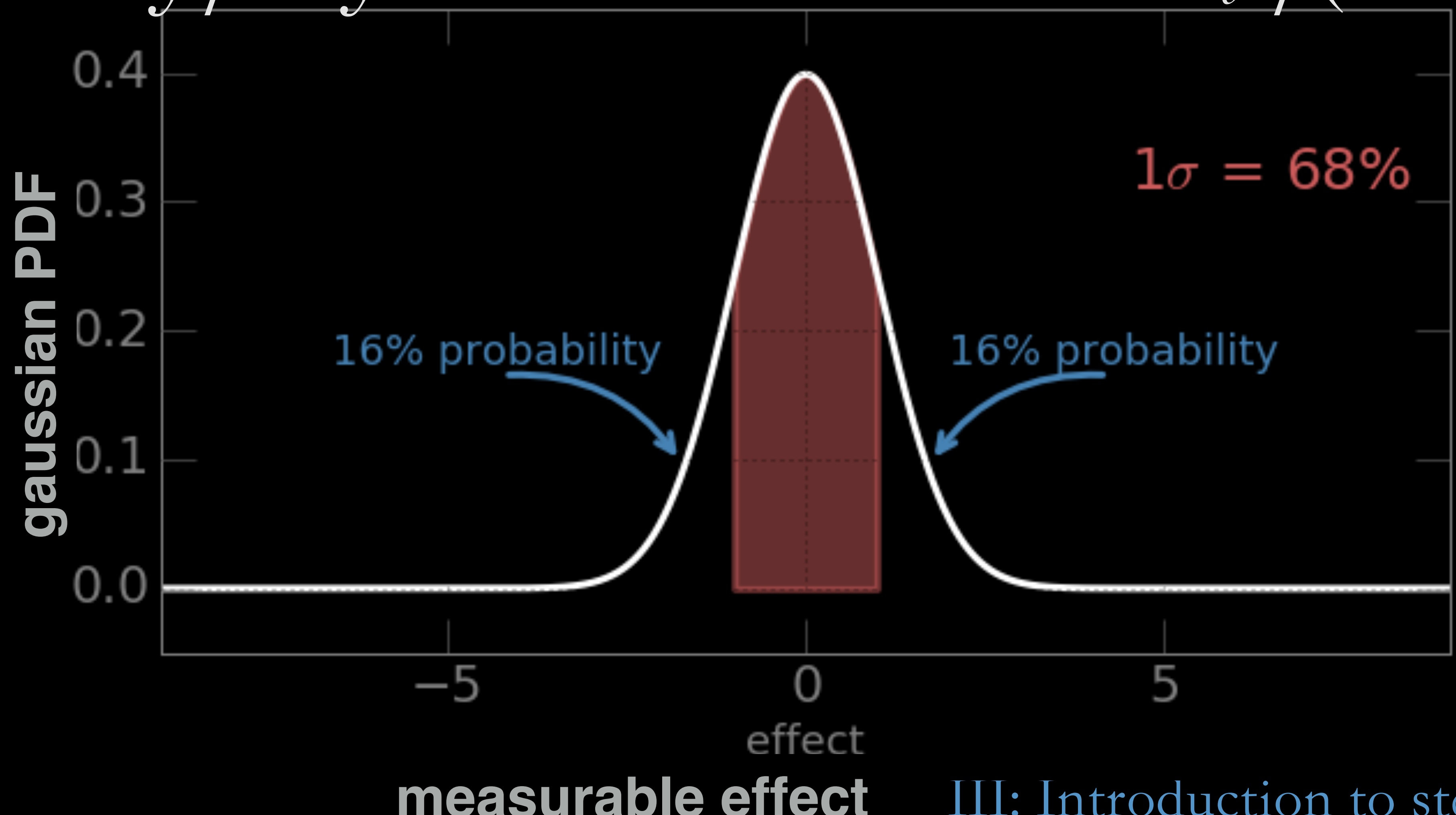
T-test, *Z*-test, bayesian analysis…



Rejecting the Null Hypothesis: what is the *p* **value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

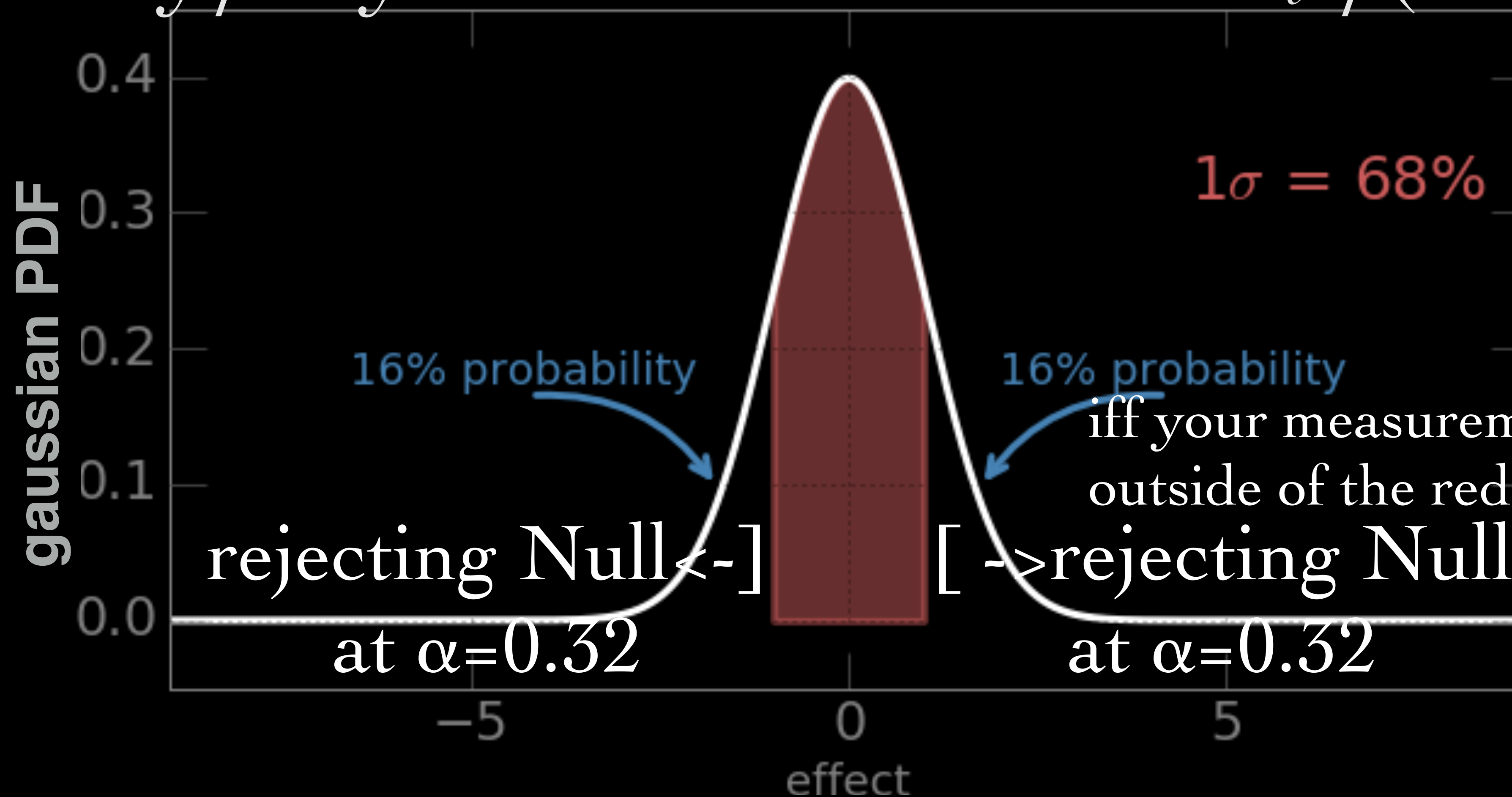
- decide what the significance threshold is: typically 5% $\alpha=0.05$
- choose a statistical test (T-test, *Z*-test, bayesian analysis…)
- find the probability *p* of your measurement under $H_0 : p(\mathrm{m} \mid H_0)$



Rejecting the Null Hypothesis: what is the ***p* value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

- decide what the significance threshold is: typically 5% $\alpha=0.05$
- choose a statistical test (T-test, *Z*-test, bayesian analysis…)
- find the probability *p* of your measurement under H_0: $p(\mathrm{m} \mid H_0)$

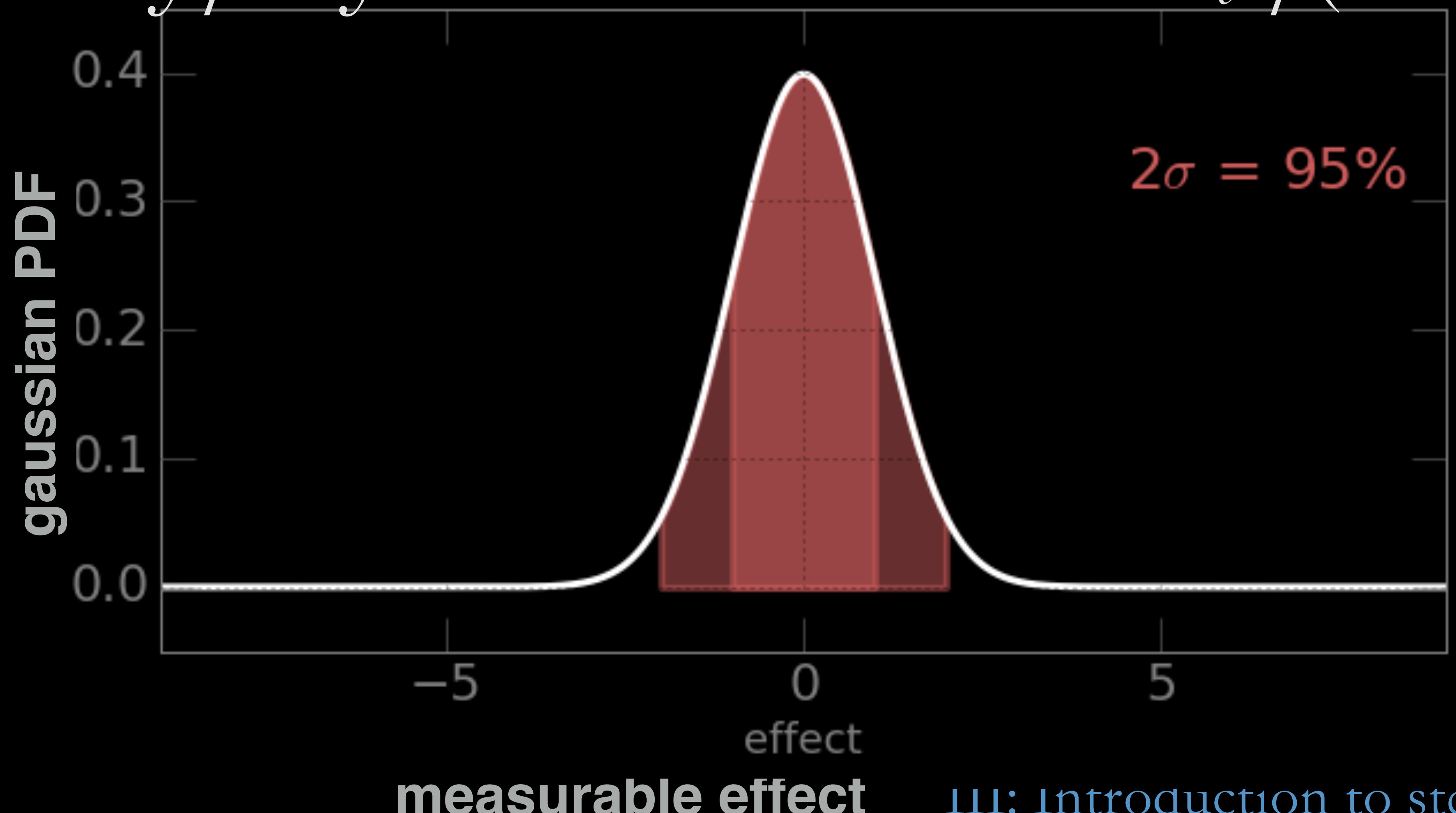


measurable effect

Rejecting the Null Hypothesis: what is the ***p*** **value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

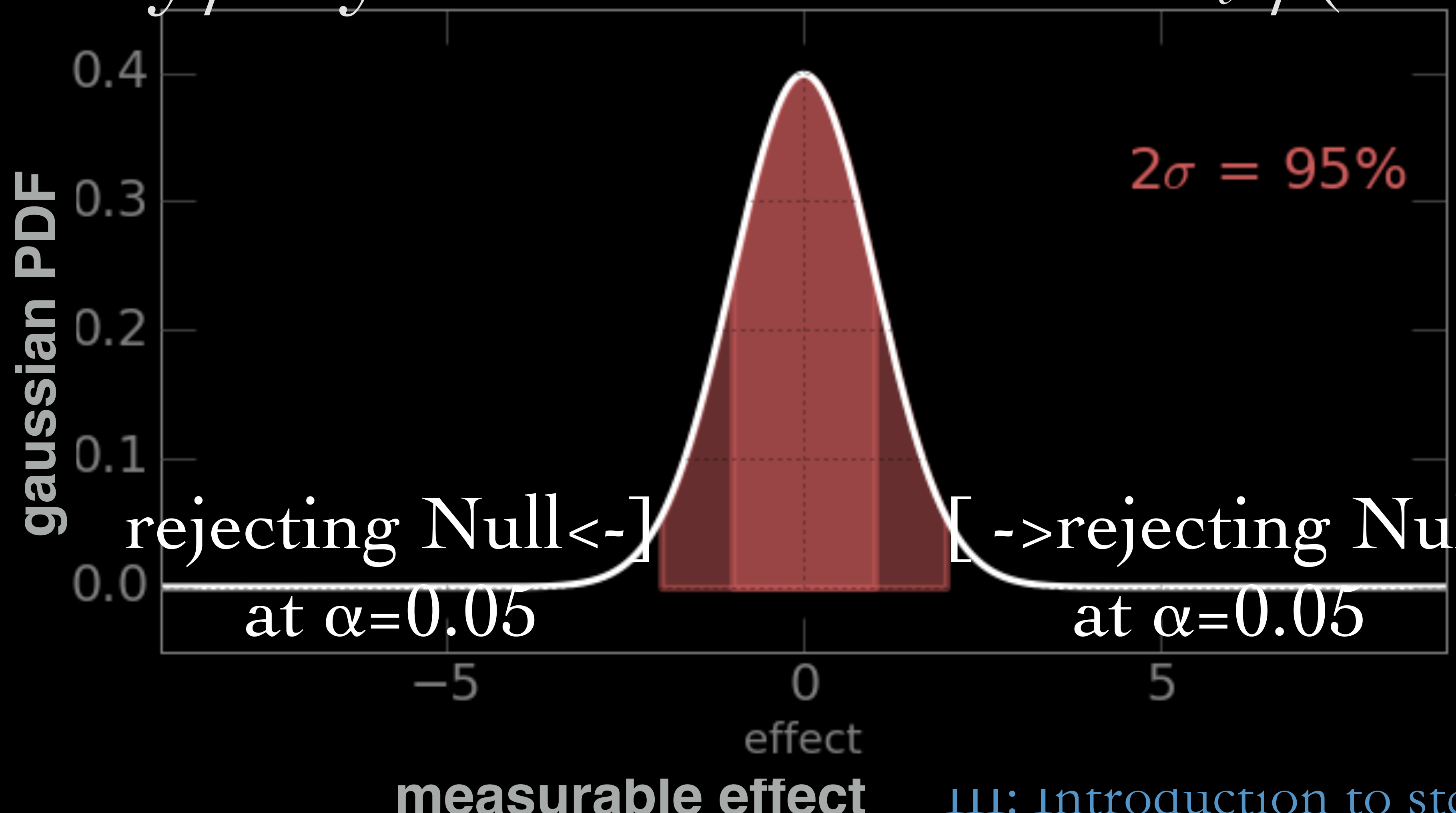
- decide what the significance threshold is: typically 5% α=0.05
- choose a statistical test (T-test, *Z*-test, bayesian analysis…)
- find the probability *p* of your measurement under H_0: $p(\mathrm{m} \mid H_0)$



Rejecting the Null Hypothesis: what is the *p* **value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

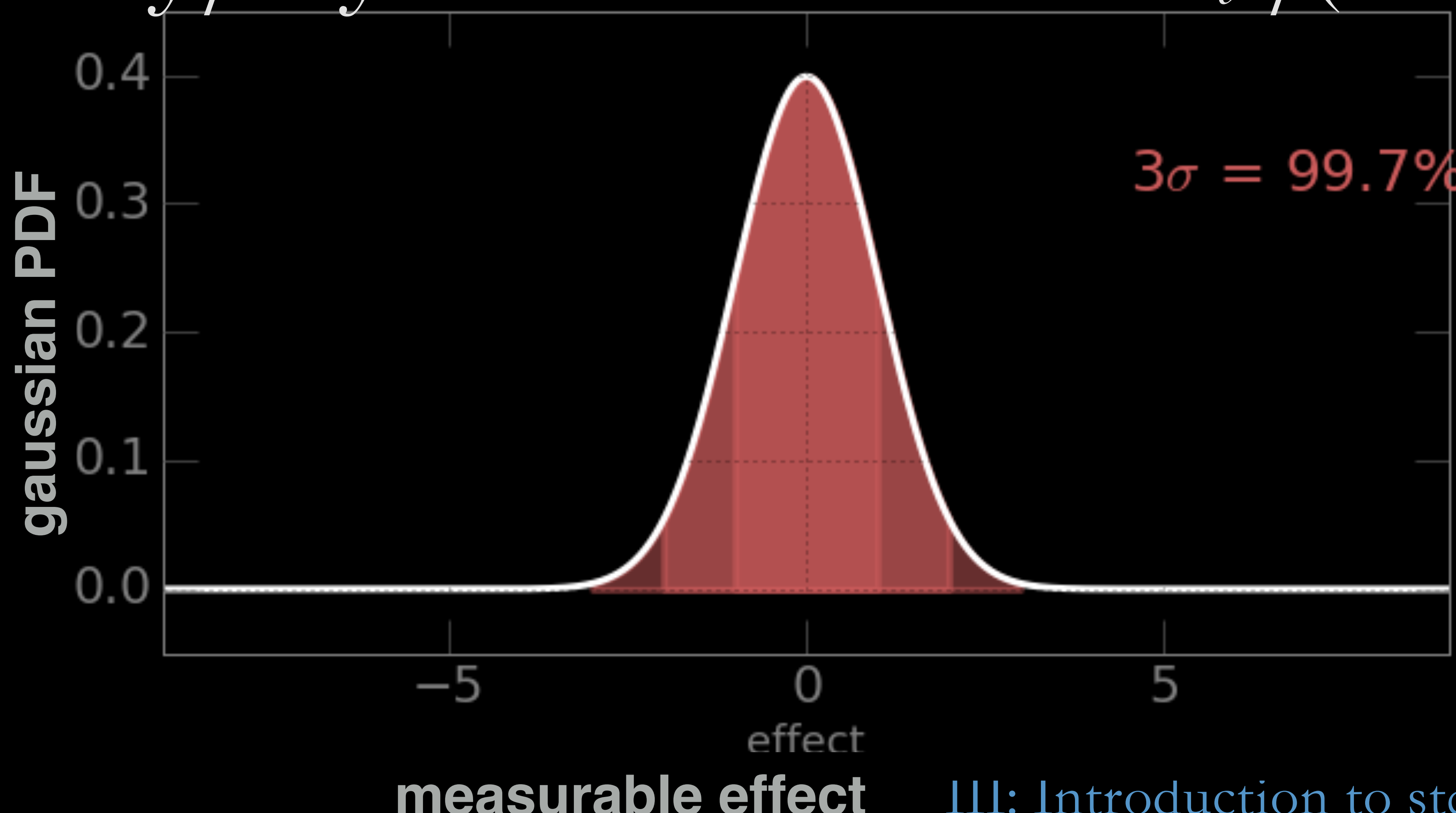
- decide what the significance threshold is: typically 5% $\alpha=0.05$
- choose a statistical test (T-test, *Z*-test, bayesian analysis…)
- find the probability p of your measurement under $H_0: p(m \mid H_0)$



Rejecting the Null Hypothesis: what is the *p* **value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

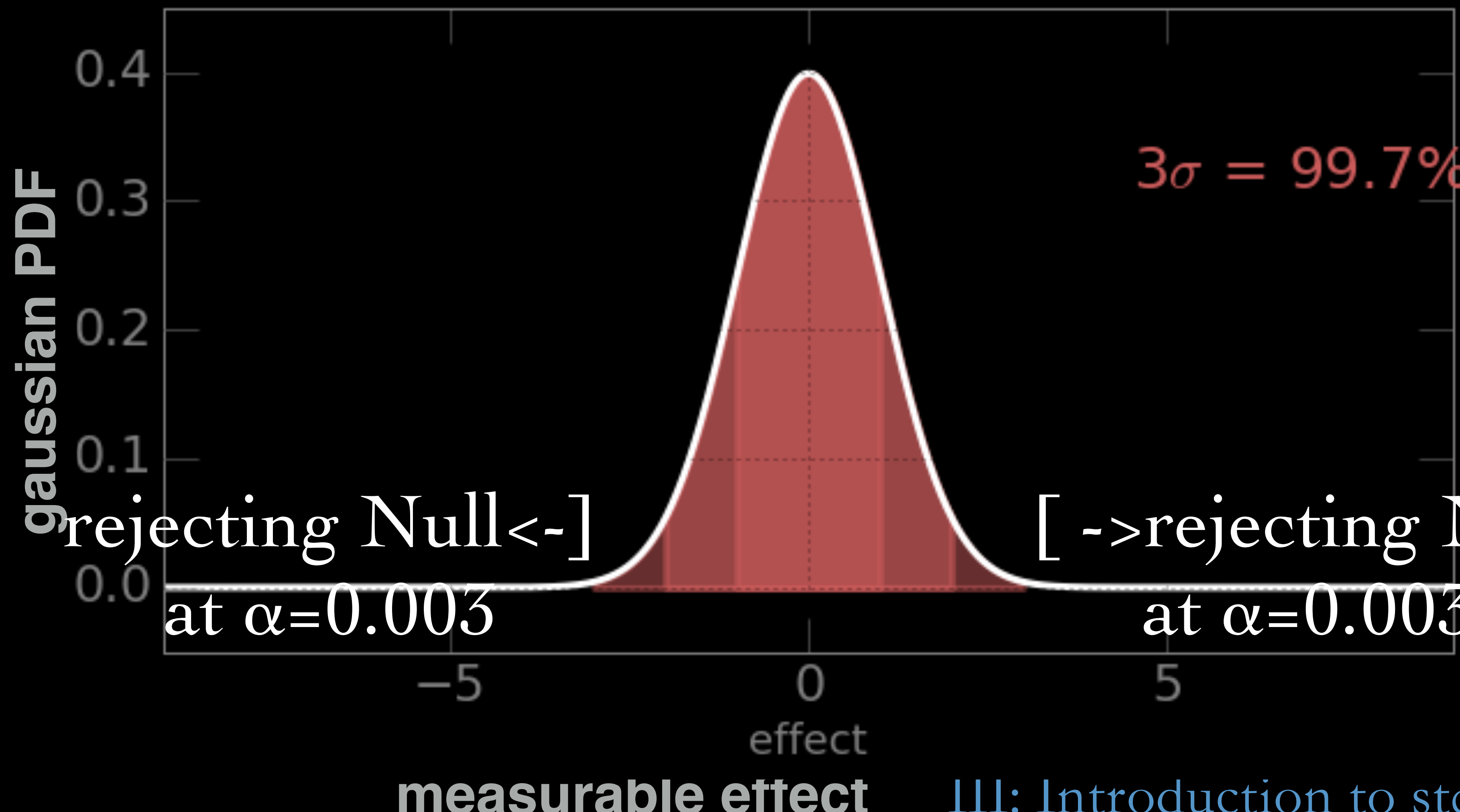
- decide what the significance threshold is: typically 5% $\alpha=0.05$
- choose a statistical test (T-test, *Z*-test, bayesian analysis…)
- find the probability p of your measurement under H_0: $p(\mathrm{m} \mid H_0)$



Rejecting the Null Hypothesis: what is the *p* **value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

- decide what the significance threshold is: typically 5% α=0.05
- choose a statistical test (T-test, *Z*-test, bayesian analysis…)
- find the probability *p* of your measurement under H_0: $p(\mathrm{m} \mid H_0)$



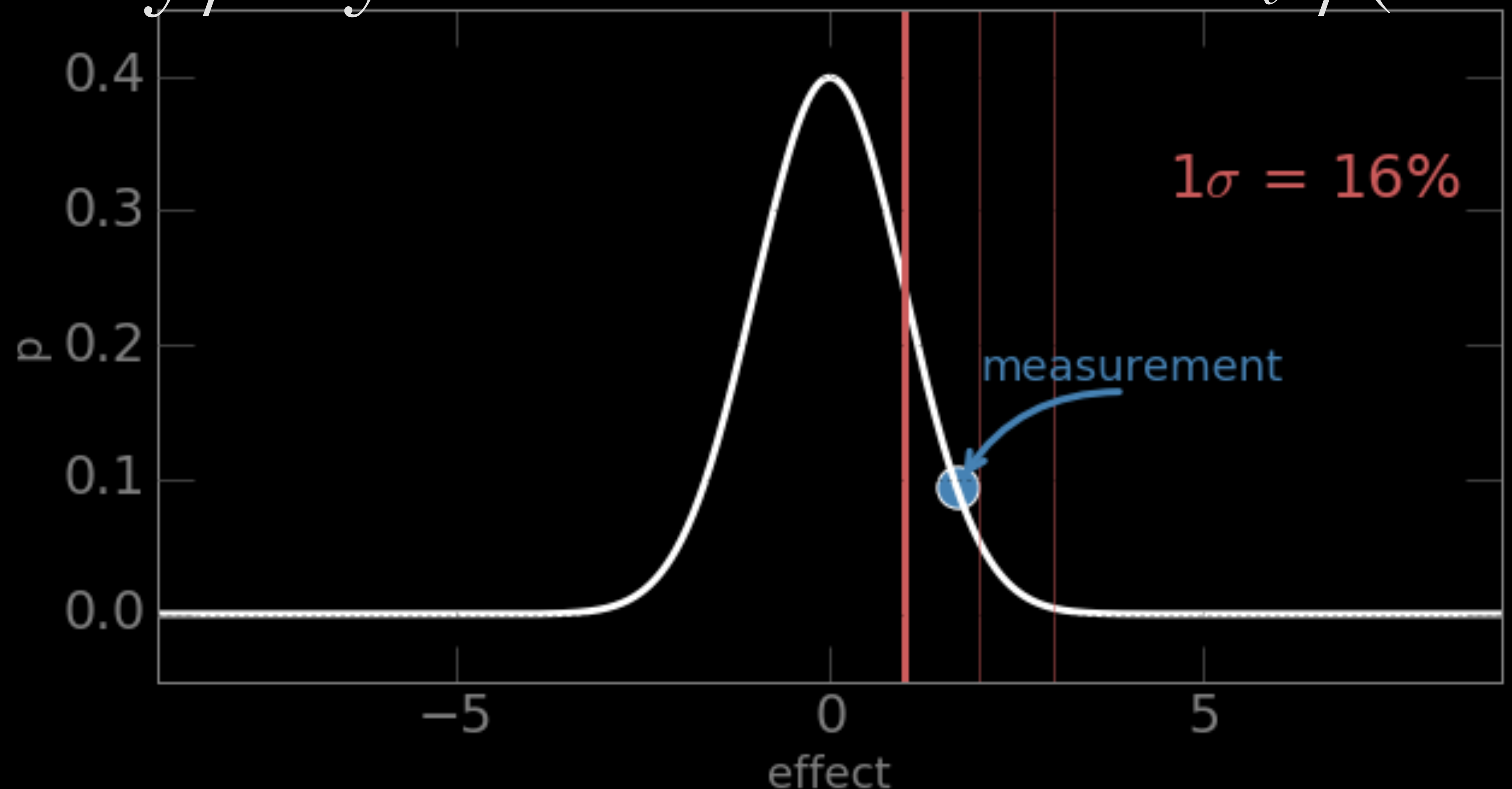
rejecting Null<-] at α=0.003

[->rejecting Null at α=0.003

Rejecting the Null Hypothesis: what is the *p* **value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

- decide what the significance threshold is: typically 5% $\alpha=0.05$
- choose a statistical test (T-test, *Z*-test, bayesian analysis…)
- find the probability p of your measurement under H_0: $p(\mathrm{m} \mid H_0)$

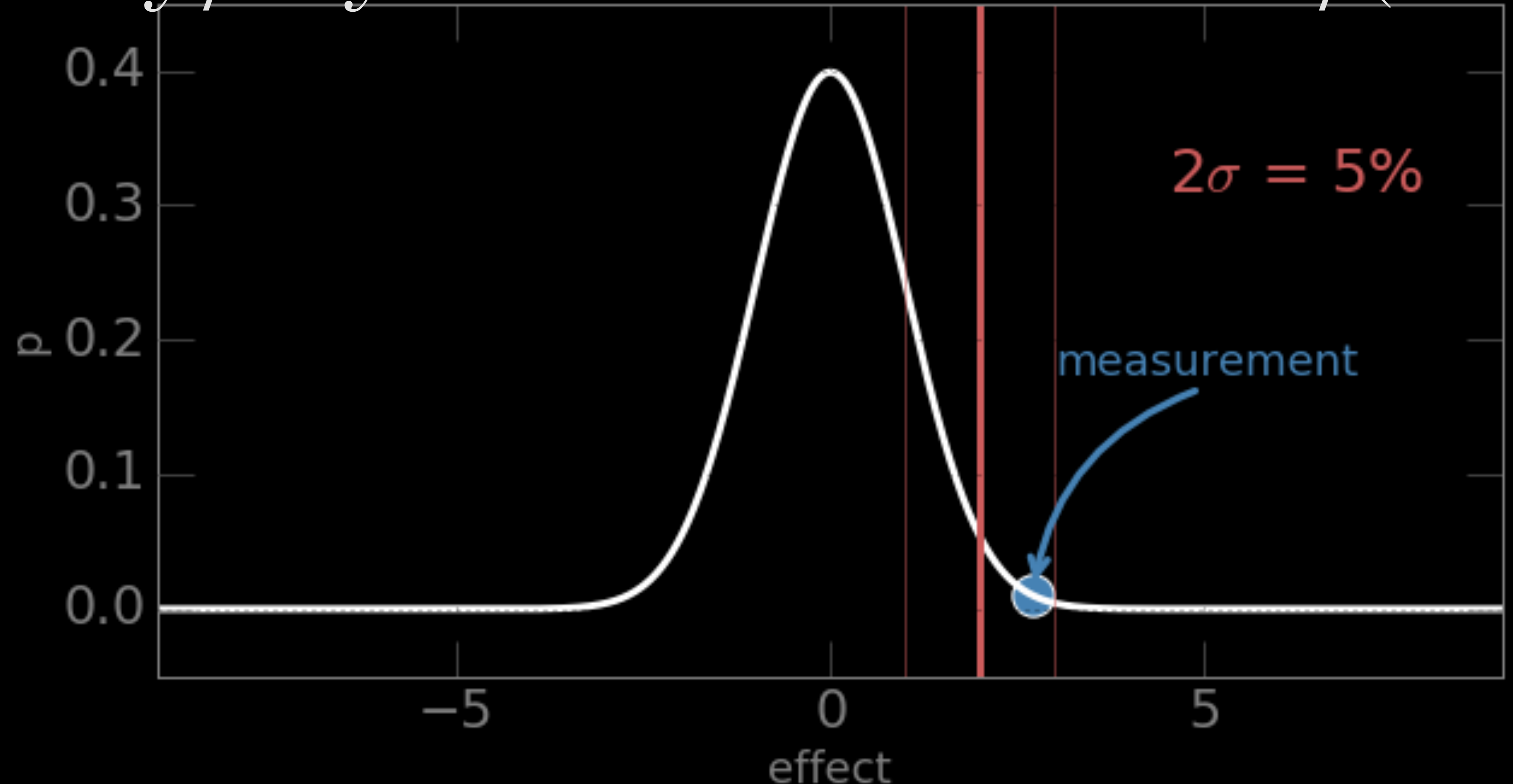


measurable effect

Rejecting the Null Hypothesis: what is the *p* **value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

- decide what the significance threshold is: typically 5% $\alpha=0.05$
- choose a statistical test (T-test, *Z*-test, bayesian analysis…)
- find the probability p of your measurement under H_0 : $p(\mathrm{m} \mid H_0)$

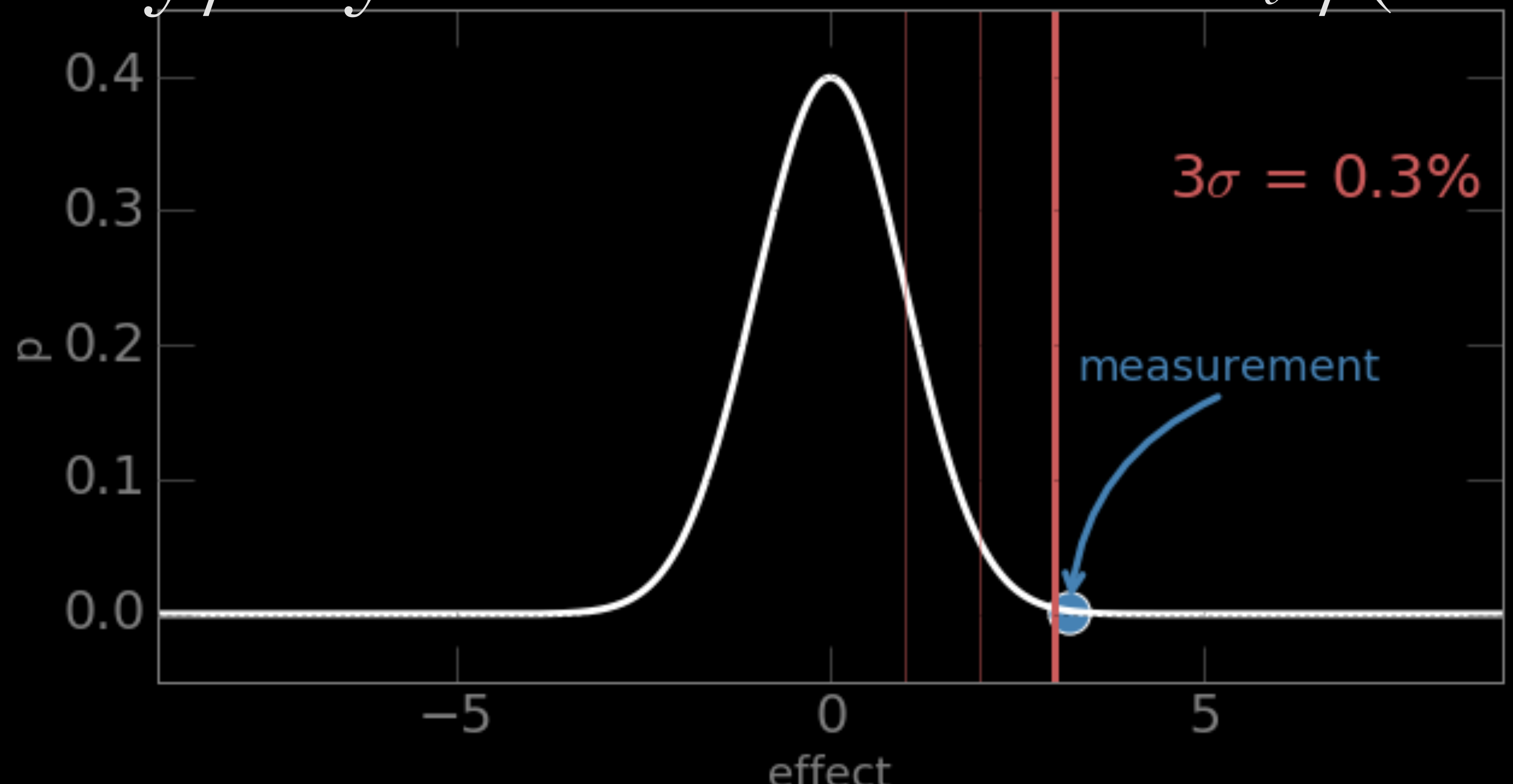


measurable effect

Rejecting the Null Hypothesis: what is the *p* **value**?

is the probability of getting a result <u>at least as extreme as</u> the one you observed just by chance <u>lower than the significance level you established?</u>

- decide what the significance threshold is: typically 5% α=0.05
- choose a statistical test (T-test, *Z*-test, bayesian analysis…)
- find the probability *p* of your measurement under H_0: $p(\mathrm{m} \mid H_0)$



measurable effect

Rejecting the Null Hypothesis: what is the *p* **value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

- decide what the significance threshold is: typically 5% α=0.05
- choose a statistical test (T-test, *Z*-test, bayesian analysis…)
- find the probability p of your measurement under H_0: $p(\mathrm{m} \mid H_0)$
- if $p(H_a)-p(H_0) > \alpha$ the null hypothesis H_0 is falsified at the 1-α confidence level

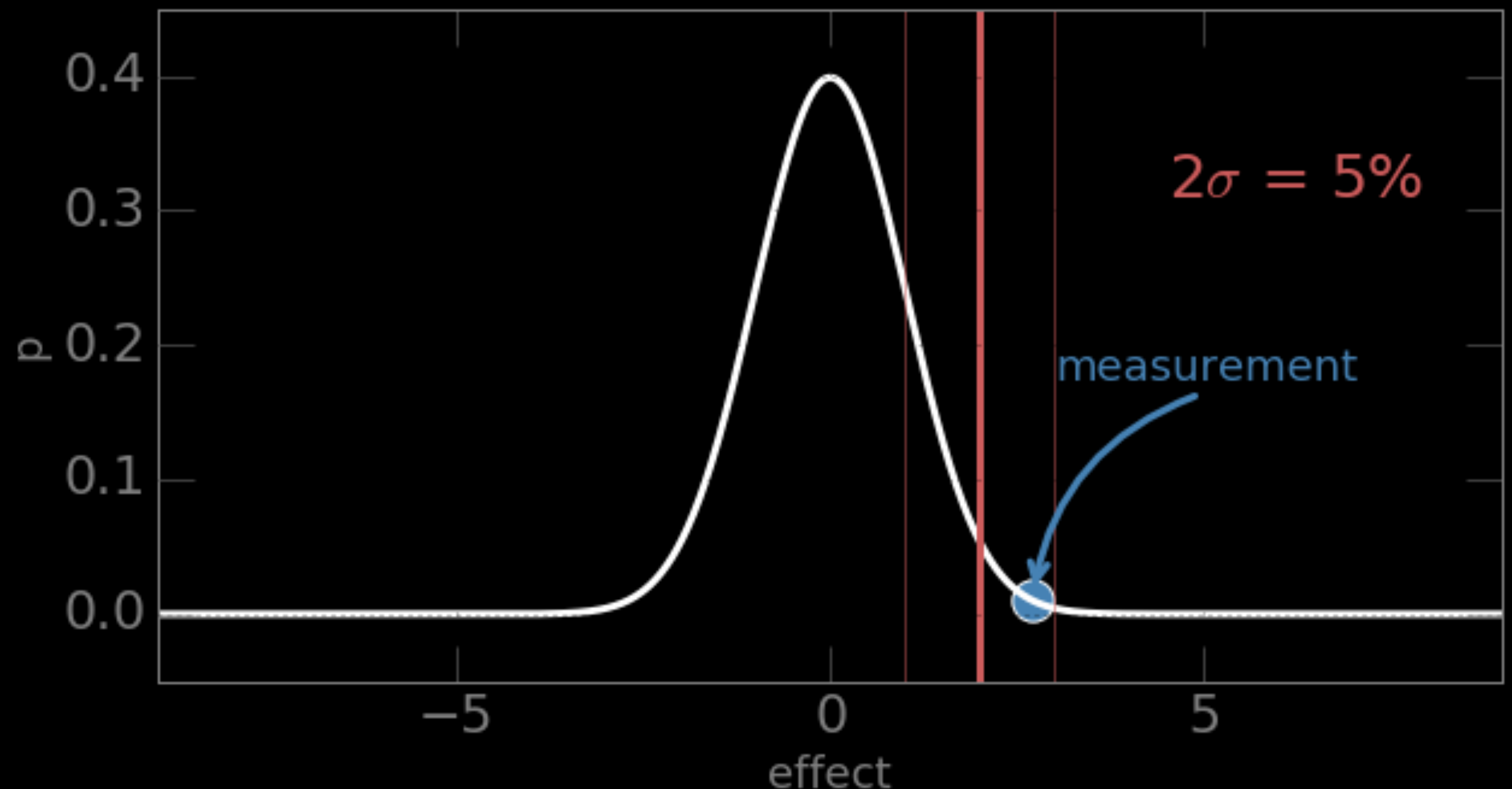
Rejecting the Null Hypothesis: what is the *p* **value**?

is the probability of getting a result at least as extreme as the one you observed just by chance lower than the significance level you established?

- decide what the significance threshold is: typically 5% α=0.05
- choose a statistical test (T-test, *Z*-test, bayesian analysis…)
- find the probability p of your measurement under H_0: $p(\mathrm{m} \mid H_0)$
- if $p(H_a)$-$p(H_0)$ > α the null hypothesis H_0 is falsified at the 1-α confidence level

The P-value is the probability that a test statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true.

Z-score: how many standard deviations away from mean



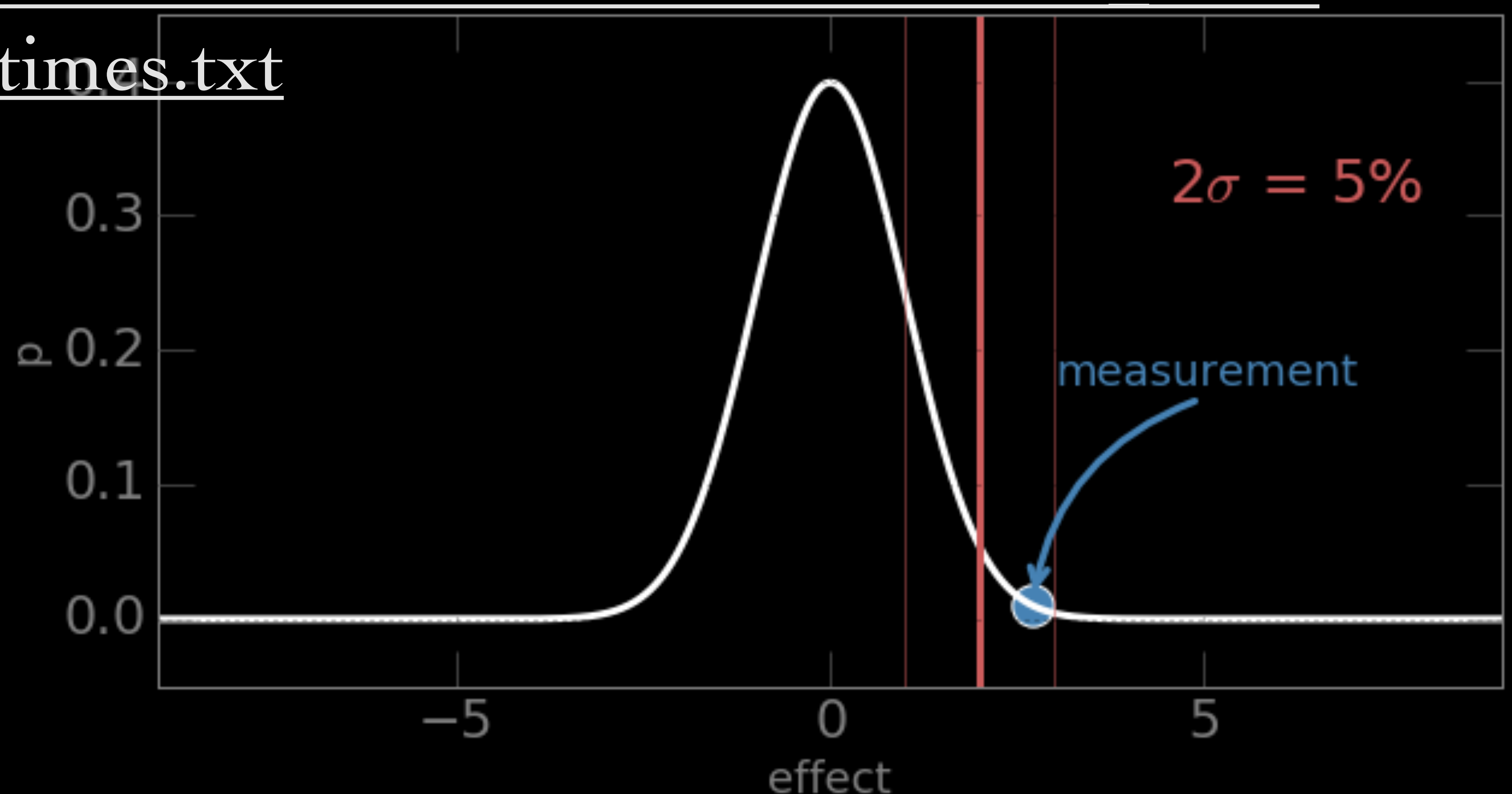
Z-score: how many standard deviations away from mean

Question: is the new Bus route improving commute?

A new bus route for line X8 is implemented. MTA wants to know if it improves commute time (travel time at peak hours).

They know what the mean travel time used to be, and measure the new travel time 100 times. The data is in https://raw.githubusercontent.com/fedhere/PUI2018_fb55/master/Lab3_fb55/times.txt

Told ~ $N(\mu=36, \sigma=6)$

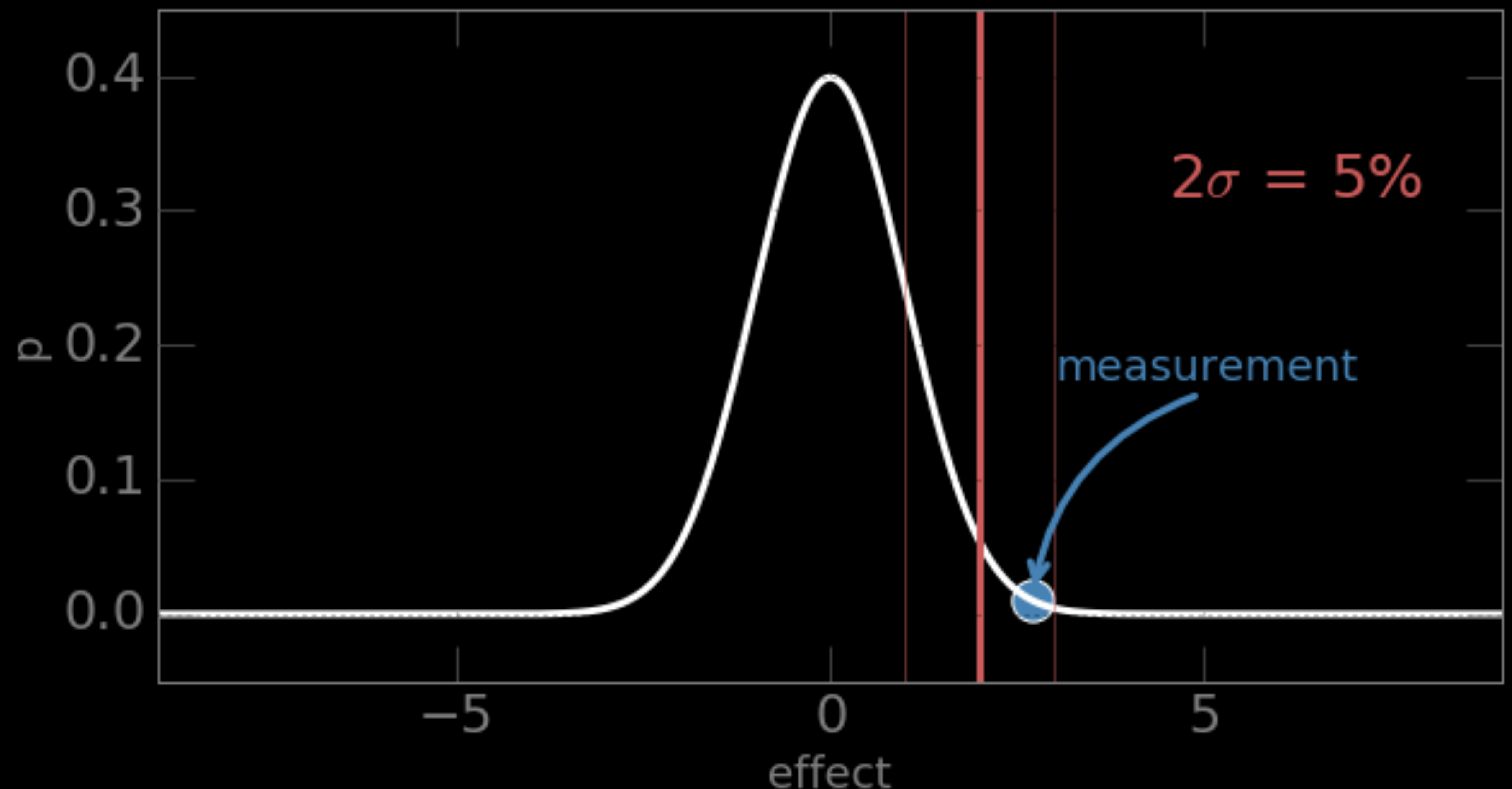


Z-score: how many standard deviations away from mean

Question: is the new Bus route imprving commute?

- H_0: the commute time is the same or longer with the new bus route as it was before

Told ~ N(μ=36, σ=6)

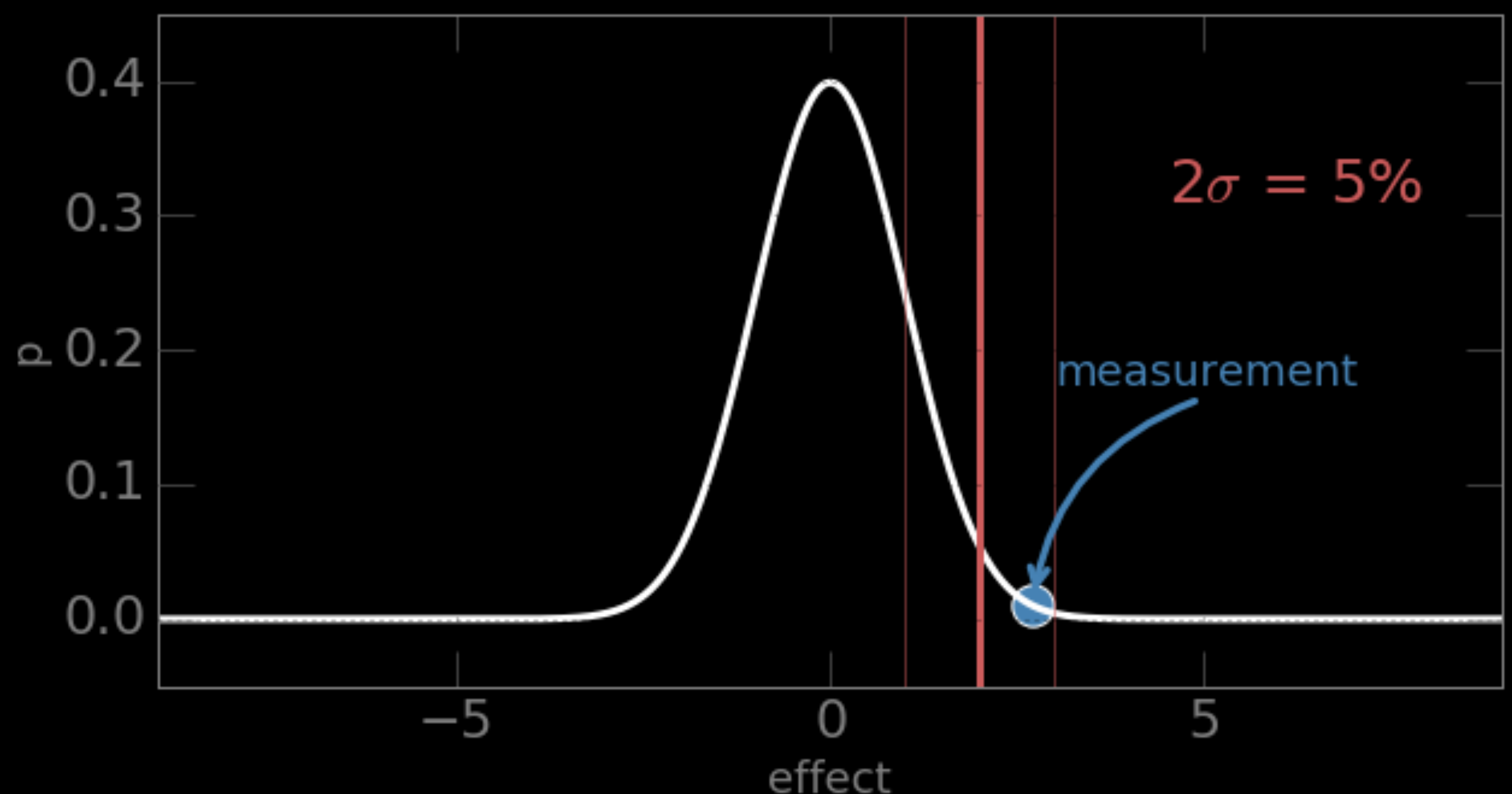


Z-score: how many standard deviations away from mean

Question: is the new Bus route imprving commute?

- H0: the commute time is on average the same or longer with the new bus route as it was before: Tnew >= Told

Told ~ $\mathcal{N}$(μ=36, σ=6)



Z-score: how many standard deviations away from mean

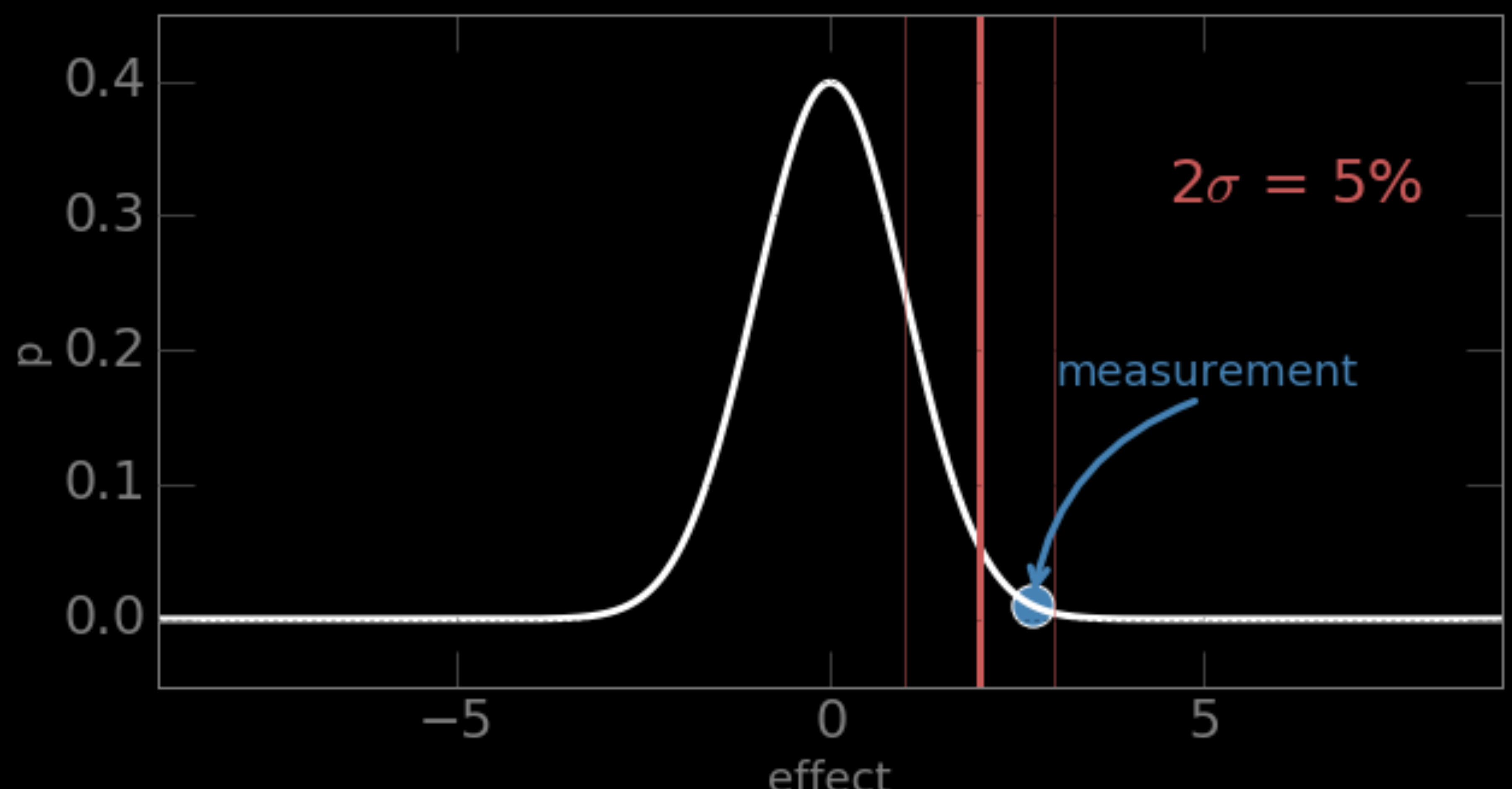
Question: is the new Bus route imprving commute?

- H0: the commute time is the same or longer with the new bus route as it was before: Tnew >= Told

- Ha: the commute time is shorter with the new bus route as it was before: Tnew < Told

$\alpha = 0.05$

Told $\sim N(\mu=36, \sigma=6)$



Z-score: how many standard deviations away from mean

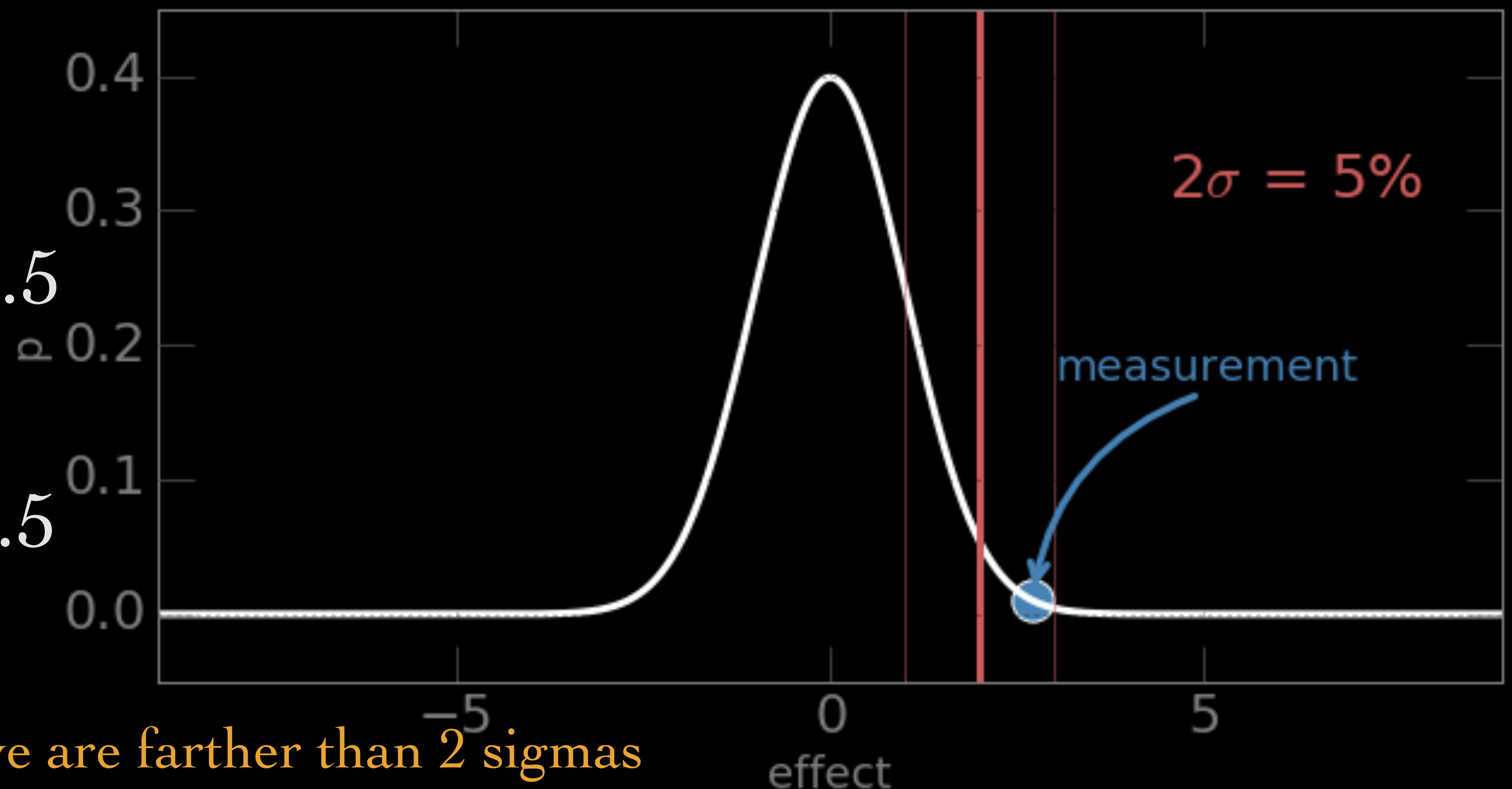
Question: is the new Bus route imprving commute?

- H0: the commute time is the same or longer with the new bus route as it was before: Tnew >= Told
- Ha: the commute time is shorter with the new bus route as it was before: Tnew < Told

Told ~ $N(\mu=36, \sigma=6)$

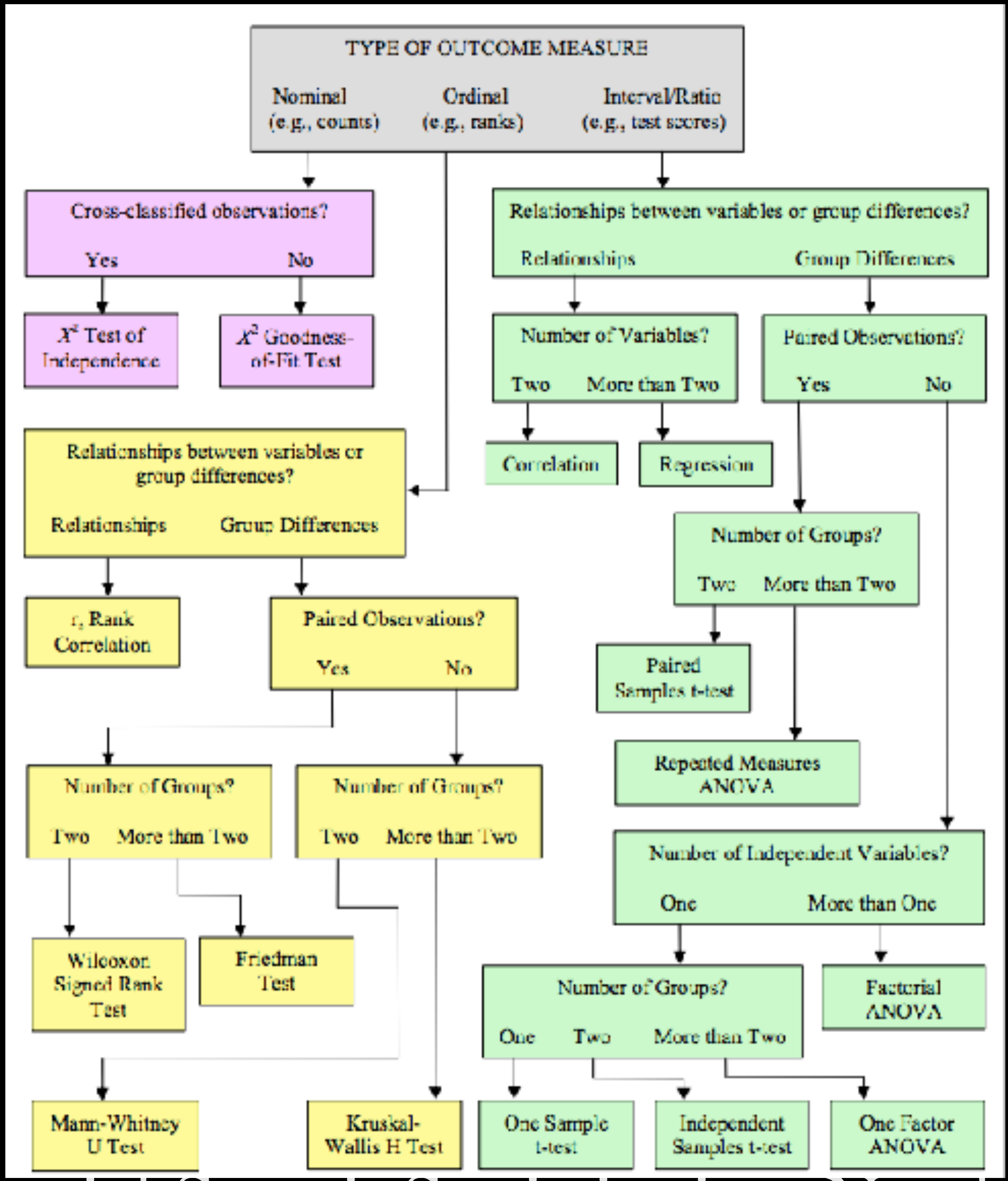
$\bar{x}$ = mean(Tnew) = 34.5

$$Z = \frac{\mu_{pop} - \mu_{sample}}{\sigma / \sqrt{N}} = 2.5$$



2.5 > 2 => we are farther than 2 sigmas away from the old mean => reject NULL (this is a Z-test)

which is the correct statistical test?? it depends on your data!



more on this next week

Chapter on Inferential Stats inStatistics in a Nutshell: http://dbmanagement.info/Books/Others/Statistics_in_a_Nutshell_A_Desktop_Quick_Reference

HYPOTHESIS TESTING & EXPERIMENT DESIGN HOMEWORK:

Assignment 2: work on CitiBikes data to assess a proportion or a mean problem. I prepared an example here: https://github.com/fedhere/UInotebooks/blob/master/citibikes_1950s.ipynb
you can test any breakdown, by age (older vs younger then), by gender, tourists vs locals…

- describe your idea
- state your Null and alternative hypothesis
- choose a confidence level
- mangle your data
- choose a statistical test. Use z-score if the sample is small, while the chi square statistics if the sample is better if the sample is large.
- assess whether you can reject the Null Hypothesis

CUSP

best to work in groups! but 5 ppl max

MUST KNOWS:

- Null Hypothesis
- Normal (Gaussian)
- Poisson, Chi-Squared, Binomial distributions
- Central Limit Theorem
- Moments of a distribution
- *p*-value

Resources:

Jacob Cohen, 1994

The earth is round (p=0.05)

http://ist-socrates.berkeley.edu/~maccoun/PP279_Cohen1.pdf

Sunil Ray , 2015

Your Guide to Master Hypothesis Testing in Statistics

https://www.analyticsvidhya.com/blog/2015/09/hypothesis-testing-explained/

Sarah Boslaugh, Dr. Paul Andrew Watters, 2008 (in the CUSP library)

Statistics in a Nutshell (Chapters Inferential Stats 7/3 depend on version)

https://books.google.com/books/about/Statistics_in_a_Nutshell.html?id=ZnhgO65Pyl4C

David M. Lane et al. (free online)

Introduction to Statistics (Chapter 3, 7, 11)

http://onlinestatbook.com/Online_Statistics_Education.epub

http://onlinestatbook.com/2/index.html

Max Mether

The history of the central limit theorem

http://salserver.org.aalto.fi/vanhat_sivut/Opinnot/Mat-2.4108/pdf-files/emet03.pdf

William Chen & Joe Blitzstein

Probability Cheatsheet v2.0

http://alturl.com/b22bs