

Use of the *p-value* as a size-dependent function: model and applications

Estibaliz Gómez de Mariscal

esgomez@pa.uc3m.es

Confronting *p-hacking*: addressing *p-value* dependence on sample size

E. Gómez-de-Mariscal, A. Sneider, H. Jayatilaka, J. M. Phillip, D. Wirtz, A. Muñoz-Barrutia

bioRxiv 2019; doi: <https://doi.org/10.1101/2019.12.17.878405>

Impact of Criticism of Null-Hypothesis Significance Testing on Statistical Reporting Practices in Conservation Biology

FIONA FIDLER,^{*§} MARK A. BURGMAN,^{*} GEOFF CUMMING,[†] ROBERT BUTTROSE,^{*} AND NEIL THOMASON[#]

^{*}School of Botany, University of Melbourne, Victoria 3010, Australia

[†]School of Psychological Science, La Trobe University, Victoria 3086, Australia

[#]Department of History and Philosophy of Science, University of Melbourne, Victoria 3010, Australia

STATISTICAL ERRORS

P values, the ‘gold standard’ of statistical validity, are not as reliable as many scientists assume.

BY REGINA NUZZO

Retire statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • <https://doi.org/10.1371/journal.pmed.0020124>

POINTS OF SIGNIFICANCE

Error bars

The meaning of error bars is often misinterpreted, as is the statistical significance of their overlap.

POINTS OF SIGNIFICANCE

Significance, P values and t-tests

The P value reported by tests is a probabilistic significance, not a biological one.

BIOINFORMATICS



Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values

Stan Pounds^{1,*} and Stephan W. Morris²

¹Department of Biostatistics and ²Departments of Pathology and Hematology/Oncology, St. Jude Children's Research Hospital, 332 N. Lauderdale St., Memphis, TN 38105-2794, USA

Received on October 18, 2002; revised on January 10, 2003; accepted on January 22, 2003

An estimate of the science-wise false discovery rate and application to the top medical literature

LEAH R. JAGER

Department of Mathematics, United States Naval Academy, Annapolis, MD 21402, USA

JEFFREY T. LEEK*

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

jleek@jhsph.edu

I. Description of the problem

- *p-values* change with the size of the sample

II. The p-value can be modeled as a continuous exponential function

III. Assessment of the null hypothesis (H_0)

IV. Use cases: real examples

I. Evaluation of drug effects

- Cellular morphology
- Flow cytometry data

II. Evaluation of features

- Cellular aging effects in motility and morphology

V. Take home messages & Discussion

VI. Try it yourself

Why are we here?

What we write

“... extremists ... see the world in black and white...

*political moderates saw shades of grey more
accurately than did either left-wing or right-wing
extremists...*

Nosek, B. A., Spies, J. R. & Motyl, M. “Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability” *Perspect. Psychol. Sci.*, 2012

Nuzzo, R., “Scientific method: Statistical errors”, *Nature*, 2014

Why are we here?

What we write

“... extremists ... see the world in black and white...

*political moderates saw shades of grey more
accurately than did either left-wing or right-wing
extremists...*

*“... our method improves 9% the accuracy of the state-
of-the-art method...*

... it has an accuracy of 85% against 78%...”

Nosek, B. A., Spies, J. R. & Motyl, M. “Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability” *Perspect. Psychol. Sci.*, 2012

Nuzzo, R., “Scientific method: Statistical errors”, *Nature*, 2014

Why are we here?

What we write

“... extremists ... see the world in black and white...

political moderates saw shades of grey more

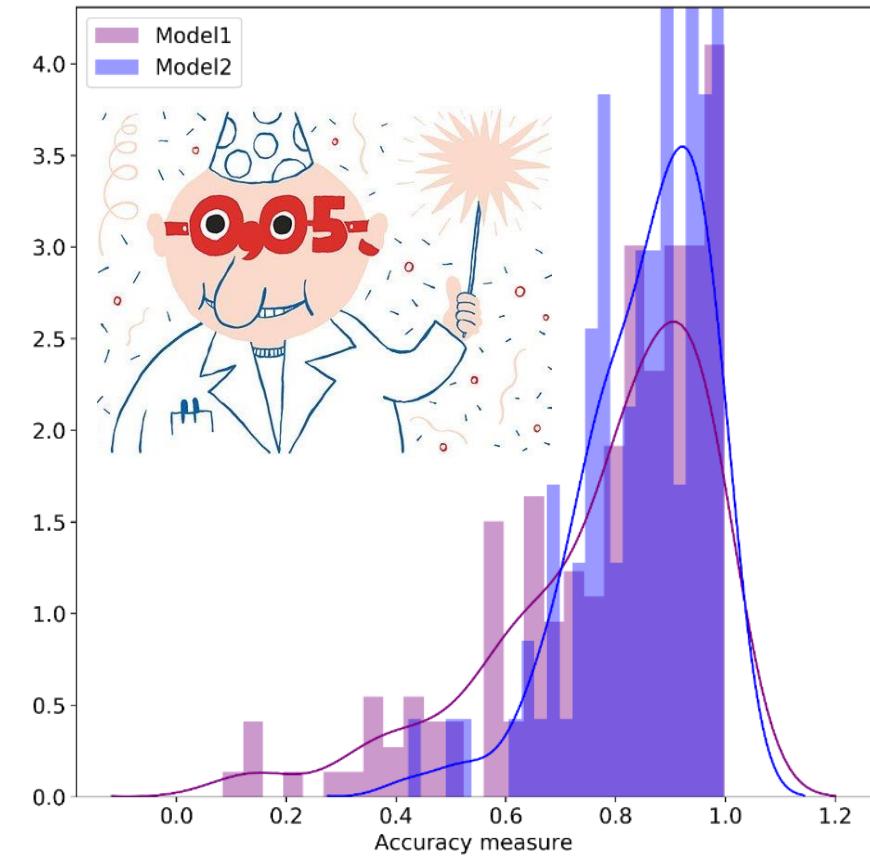
*accurately than did either left-wing or right-wing
extremists...*

“... our method improves 9% the accuracy of the state-of-the-art method...

... it has an accuracy of 85% against 78%...”

Our data

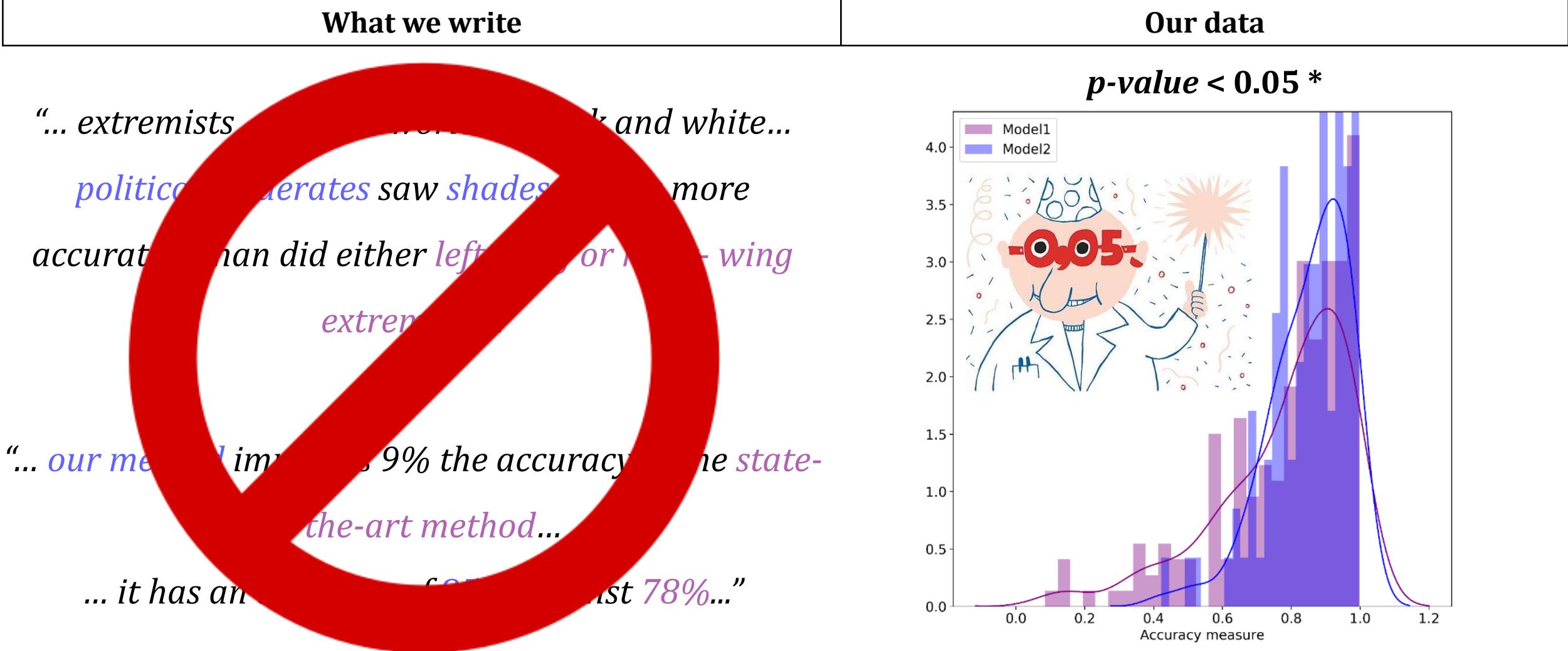
*p-value < 0.05 **



Nosek, B. A., Spies, J. R. & Motyl, M. “Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability” *Perspect. Psychol. Sci.*, 2012

Nuzzo, R., “Scientific method: Statistical errors”, *Nature*, 2014

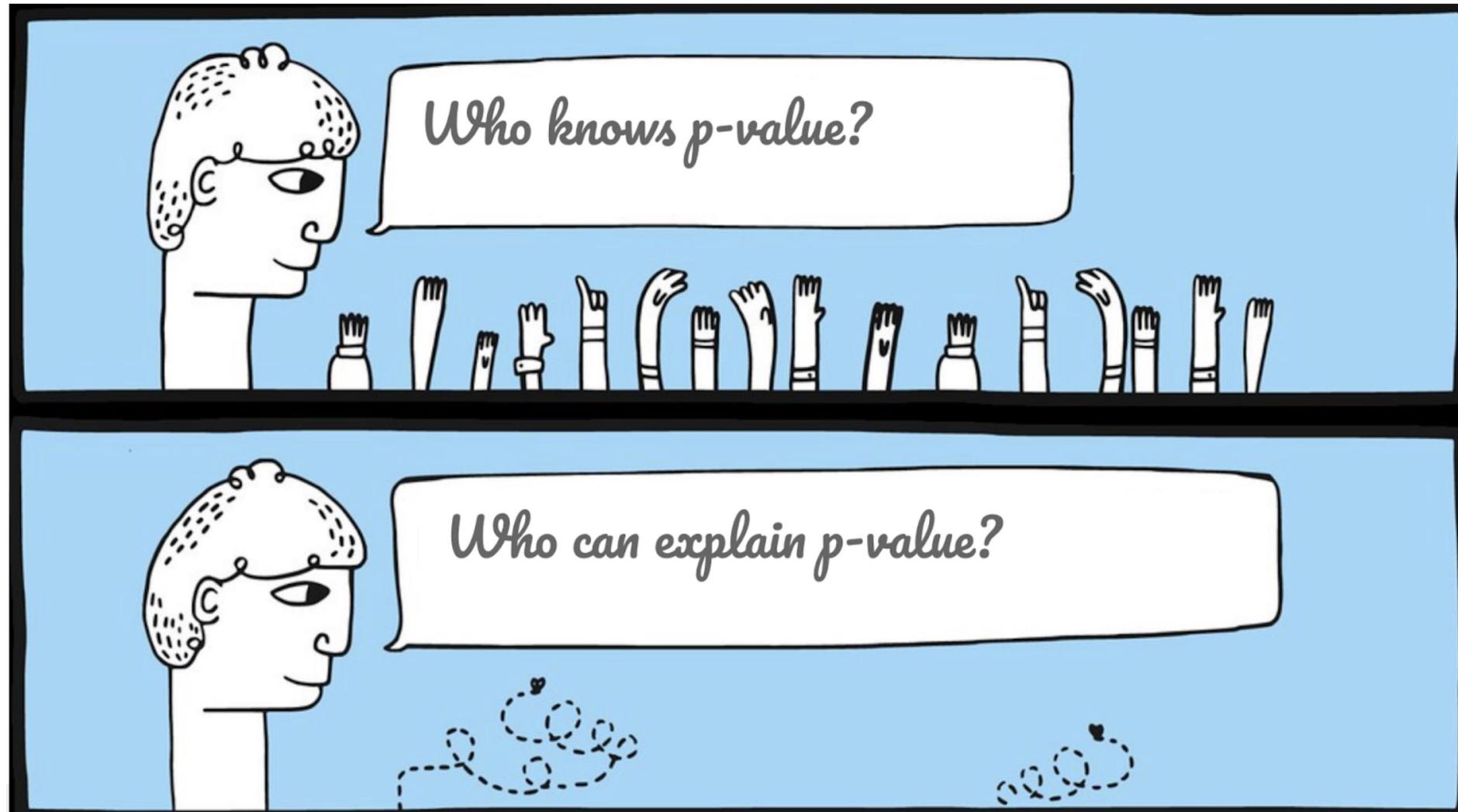
Why are we here?



Nosek, B. A., Spies, J. R. & Motyl, M. “Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability” *Perspect. Psychol. Sci.*, 2012

Nuzzo, R., “Scientific method: Statistical errors”, *Nature*, 2014

Going back to statistics class

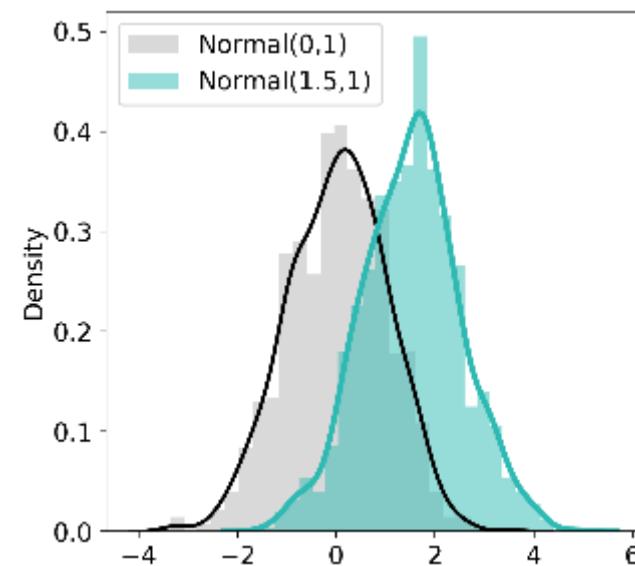


Going back to statistics class

Null hypothesis statistical tests

Given two normal datasets, we run a Student's t-test:

Null hypothesis H0: the means of both distributions are equal



T-statistic for these datasets is calculated:

$$t = \frac{\mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

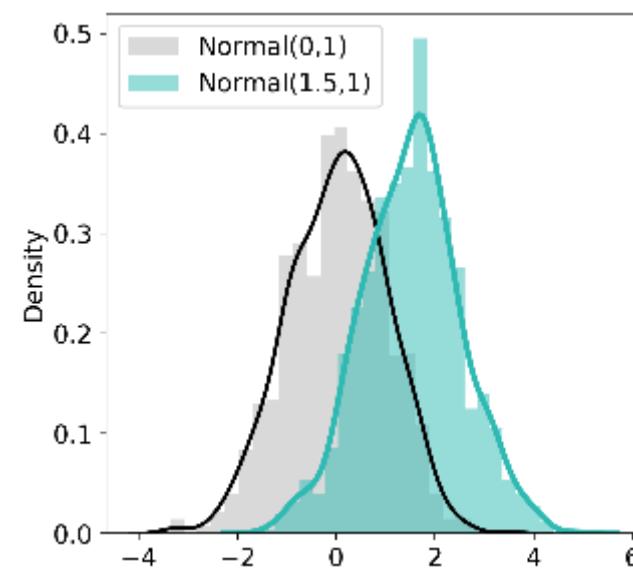
Going back to statistics class

Null hypothesis statistical tests

Distribution of t under the null hypothesis

Given two normal datasets, we run a Student's t-test:

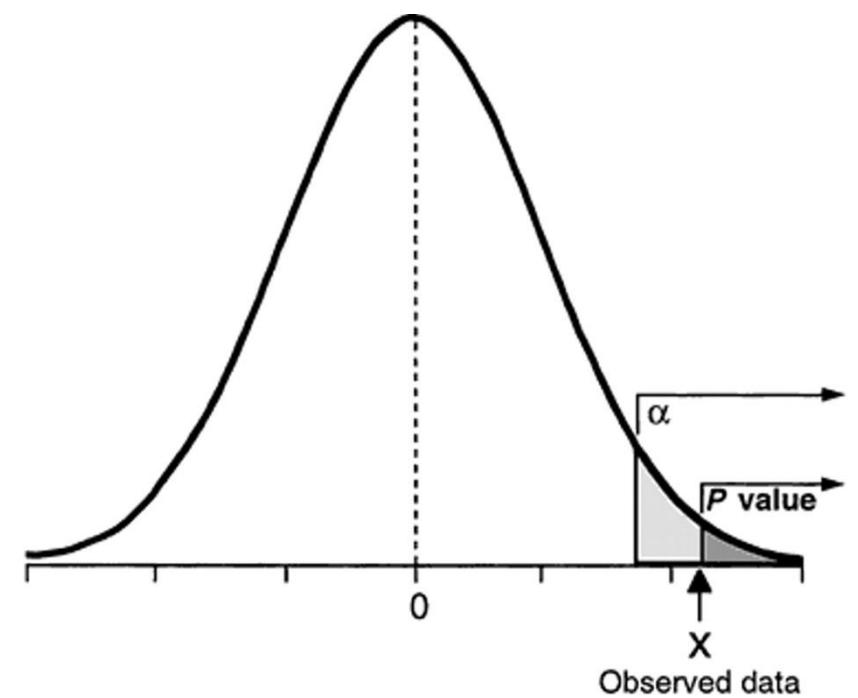
Null hypothesis H₀: the means of both distributions are equal



T-statistic for these datasets is calculated:

$$t = \frac{\mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

P-value: the probability that we would observe a result as extreme or more than our result **IF H₀ were true.**



Going back to statistics class

Null hypothesis statistical tests

Distribution of t under the null hypothesis

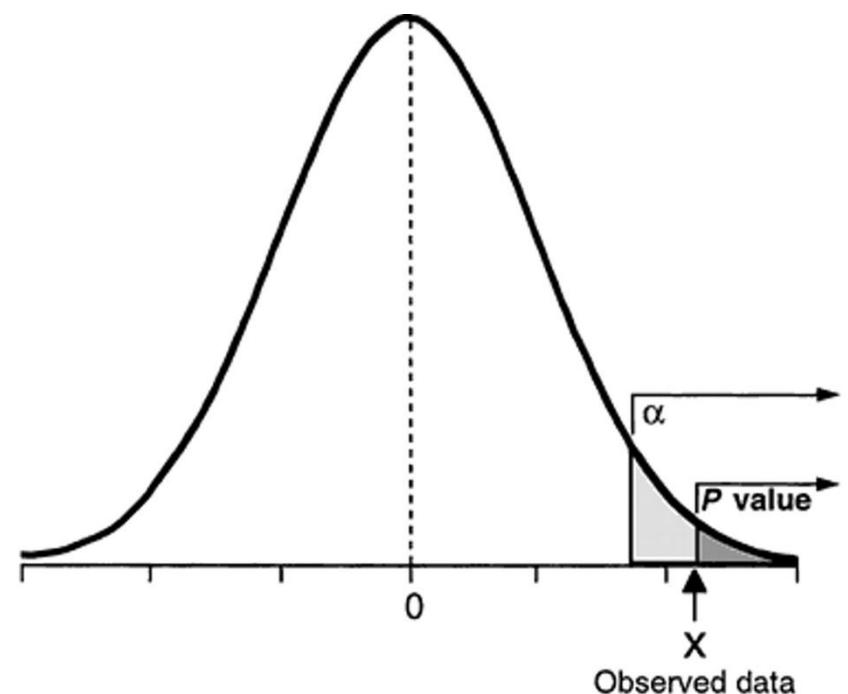
Given two normal datasets, we run a Student's t-test:

Null hypothesis H_0 : the means of both distributions are equal

Small p-value → the result is very unlikely under H_0 , then there are two chances:

- I. We observed a low probability event.
- II. H_0 is not true.

P-value: the probability that we would observe a result as extreme or more than our result **IF H_0 were true.**

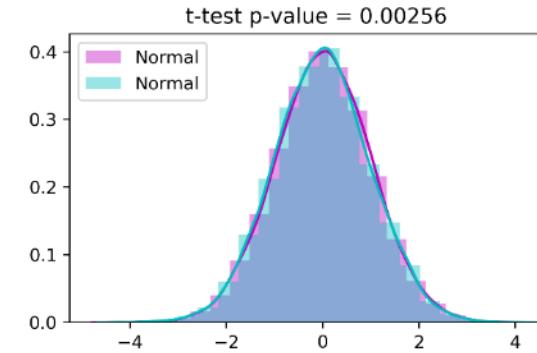
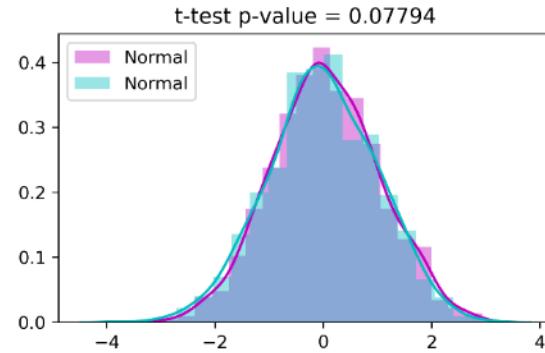
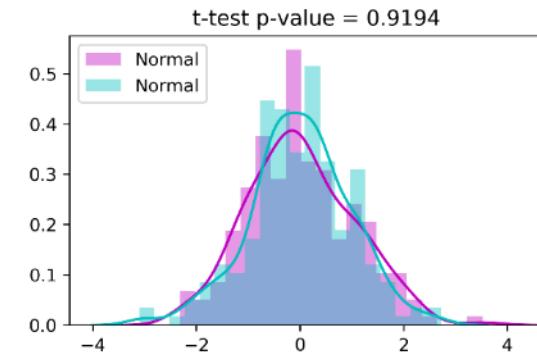
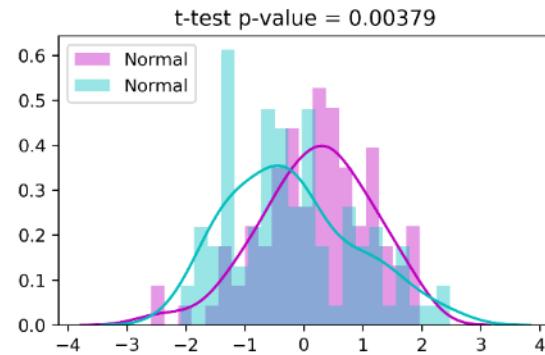


Description of the current problem

- I. Let's simulate two normal distributions $N(0.01, 1)$ and $N(0, 1)$.
- II. Compare them using Student's t -test. H_0 : the means of both distributions are equal.

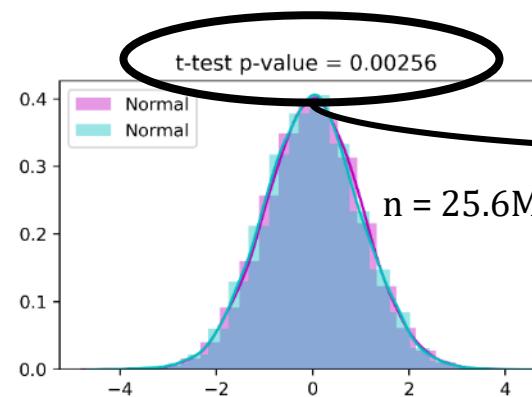
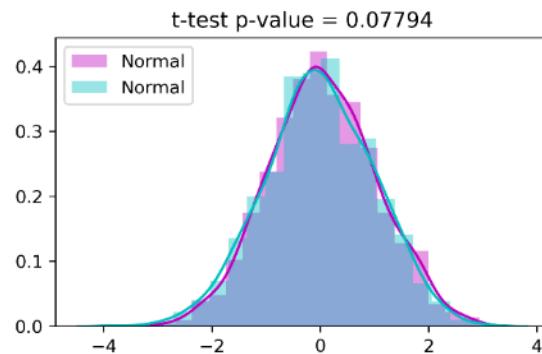
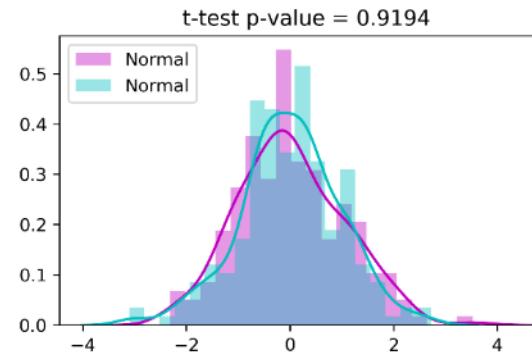
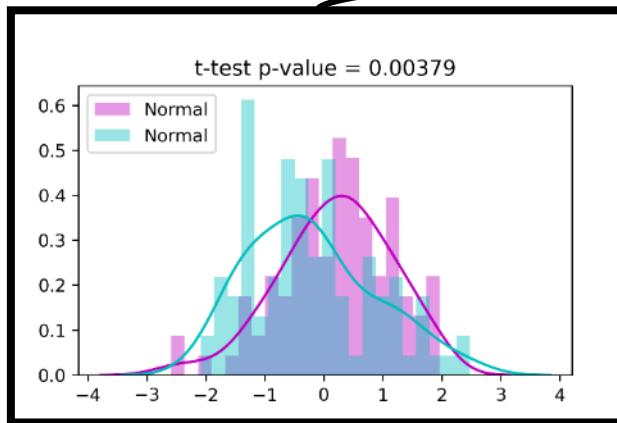
Description of the current problem

- I. Let's simulate two normal distributions $N(0.01, 1)$ and $N(0, 1)$.
- II. Compare them using Student's t -test. H_0 : the means of both distributions are equal.



Description of the current problem

- I. Let's simulate two normal distributions $N(0.01, 1)$ and $N(0, 1)$.
- II. Compare them using Student's t -test. H_0 : the means of both distributions are equal.

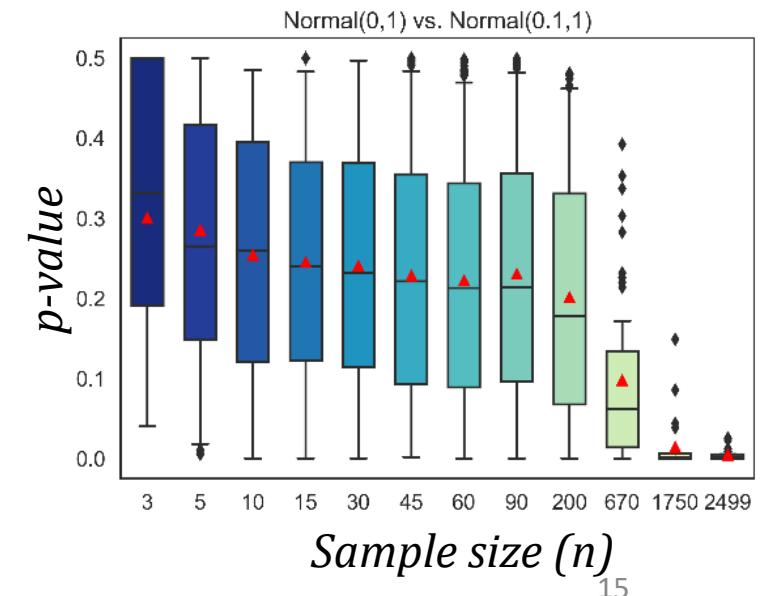
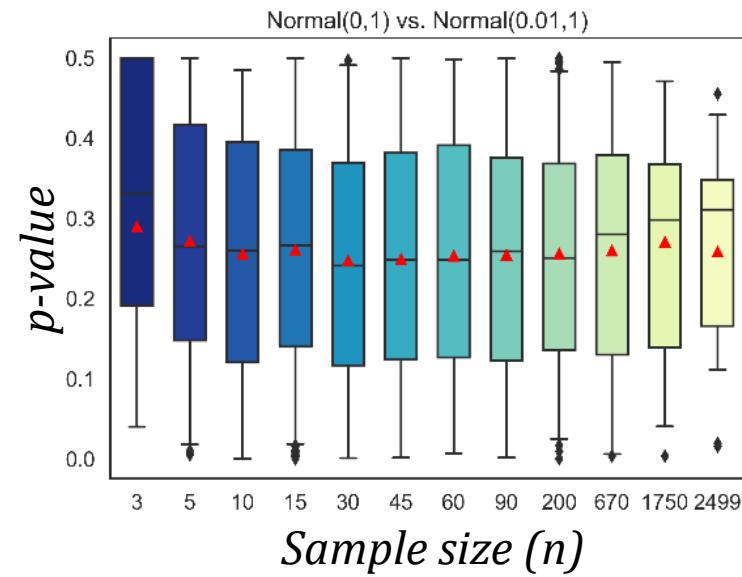
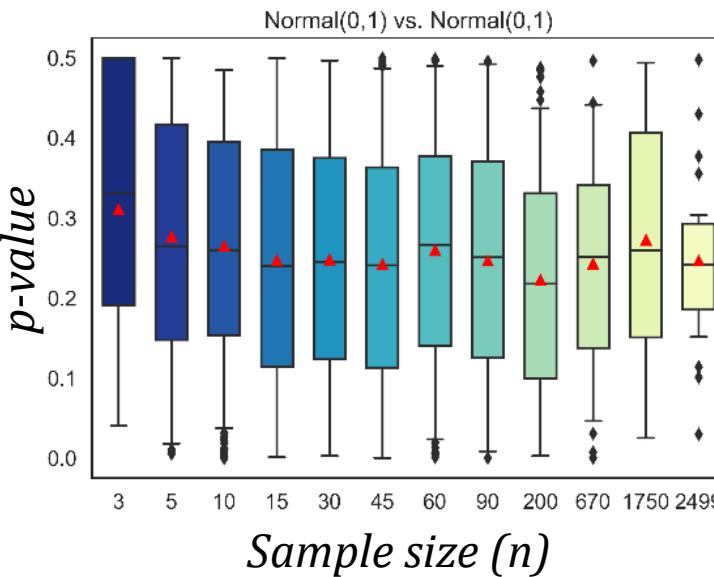


$p\text{-value} < 0.05$ while both distributions are almost the same.

Description of the current problem

- I. Let's simulate two normal distributions $N(0.01, 1)$ and $N(0, 1)$.
- II. Compare them using Student's t -test. H_0 : the means of both distributions are equal.
- III. Let's do it for different normal distributions and using different sample sizes

Fact: p -values follow a distribution

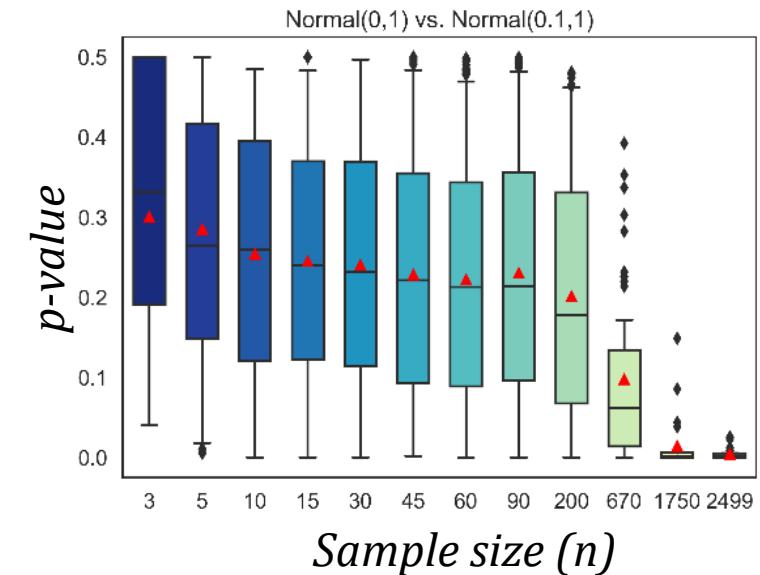
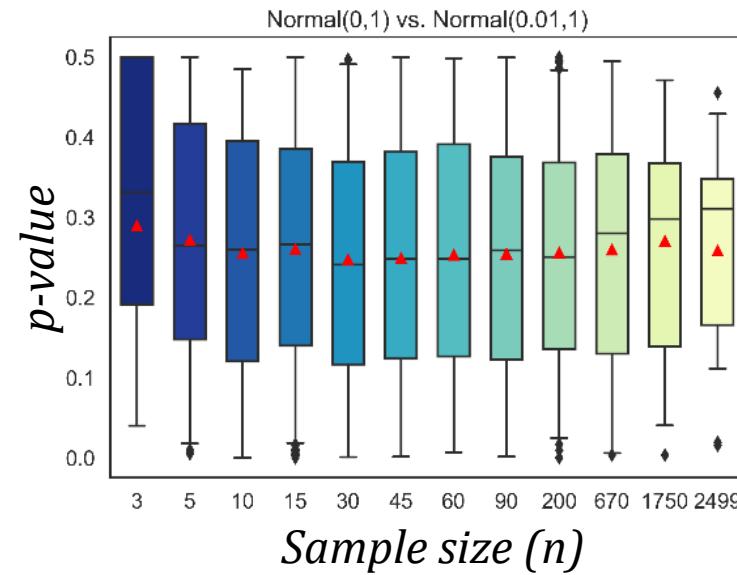
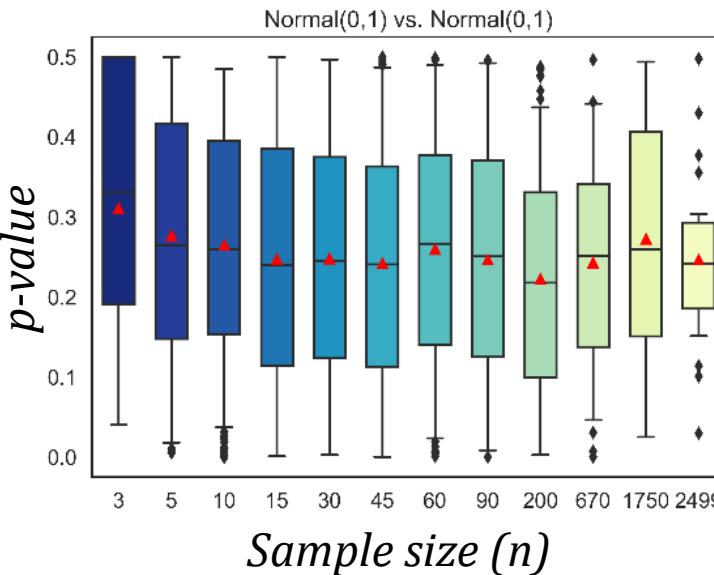


Description of the current problem

- I. Let's simulate two normal distributions $N(0.01, 1)$ and $N(0, 1)$.
- II. Compare them using Student's t -test. H_0 : the means of both distributions are equal.
- III. Let's do it for different normal distributions and using different sample sizes

Fact: p -values follow a distribution

Problem I: In similar datasets, the distribution is uniform but not every p -value is always > 0.05

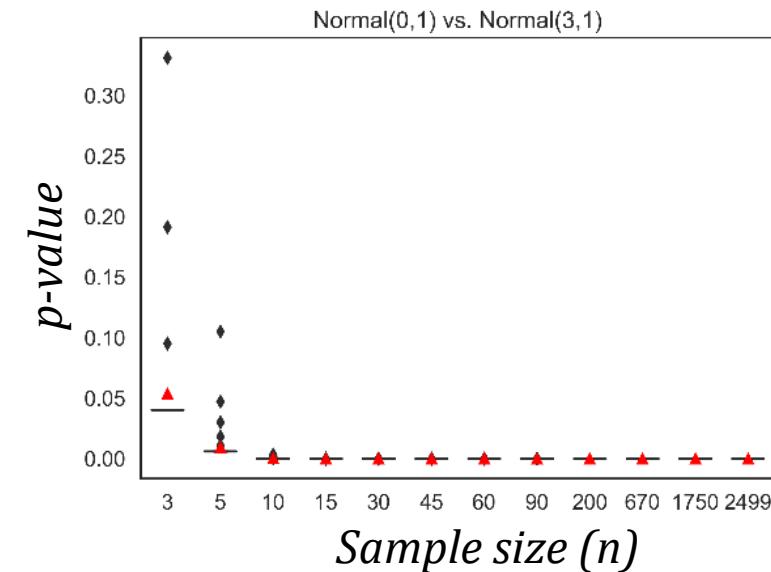
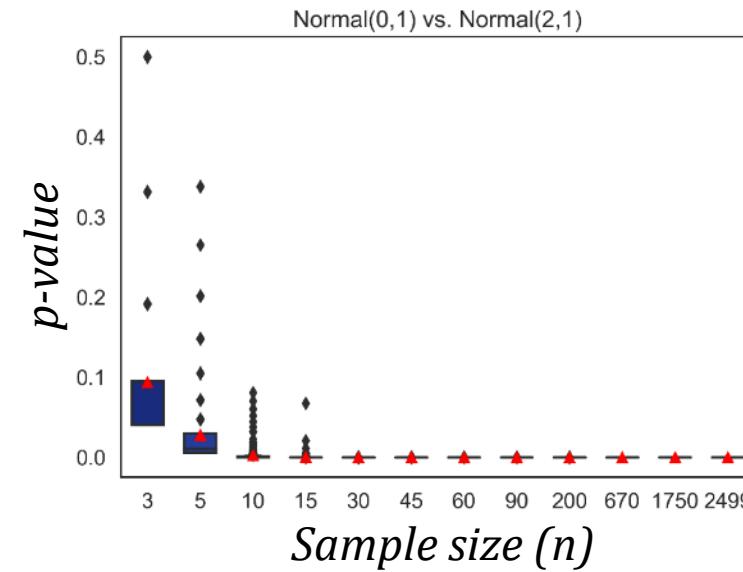
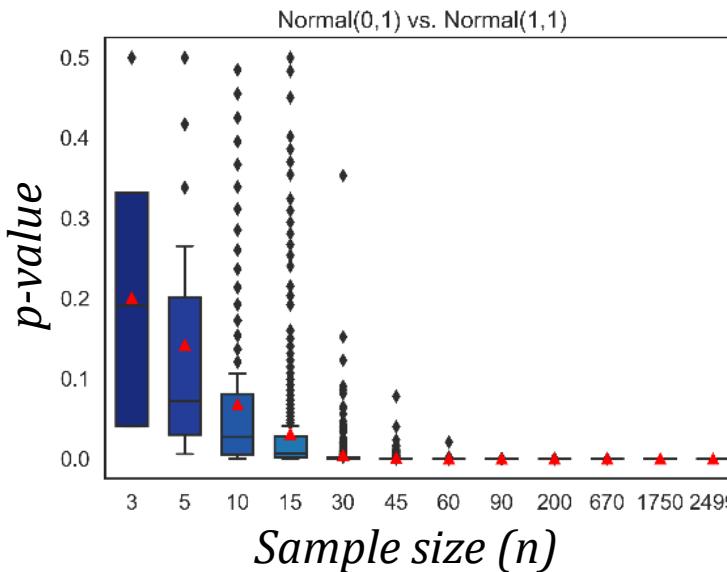


Description of the current problem

- I. Let's simulate two normal distributions $N(0.01, 1)$ and $N(0, 1)$.
- II. Compare them using Student's t -test. H_0 : the means of both distributions are equal.
- III. Let's do it for different normal distributions and using different sample sizes

Fact: p -values follow a distribution

Problem I: In similar datasets, the distribution is uniform but not every p -value is always > 0.05



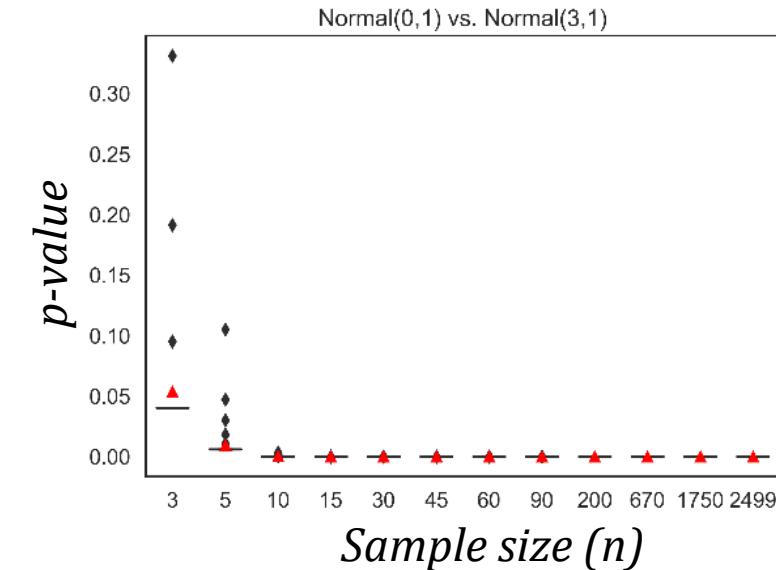
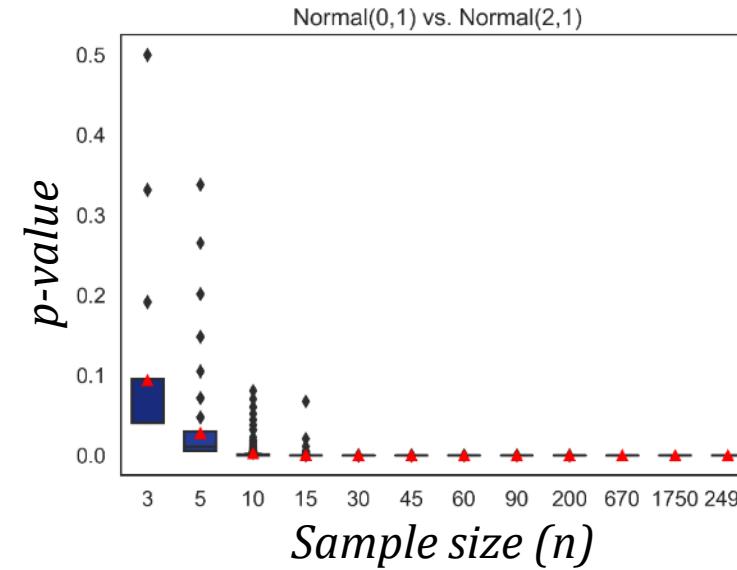
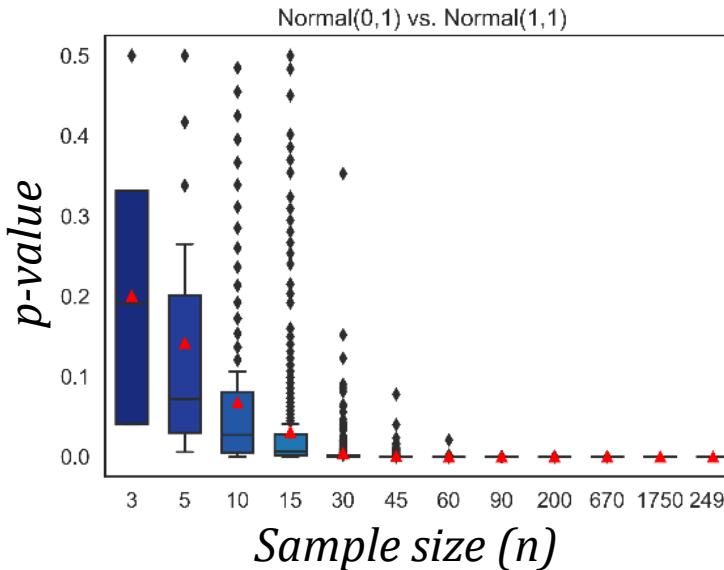
Description of the current problem

- I. Let's simulate two normal distributions $N(0.01, 1)$ and $N(0, 1)$.
- II. Compare them using Student's t -test. H_0 : the means of both distributions are equal.
- III. Let's do it for different normal distributions and using different sample sizes

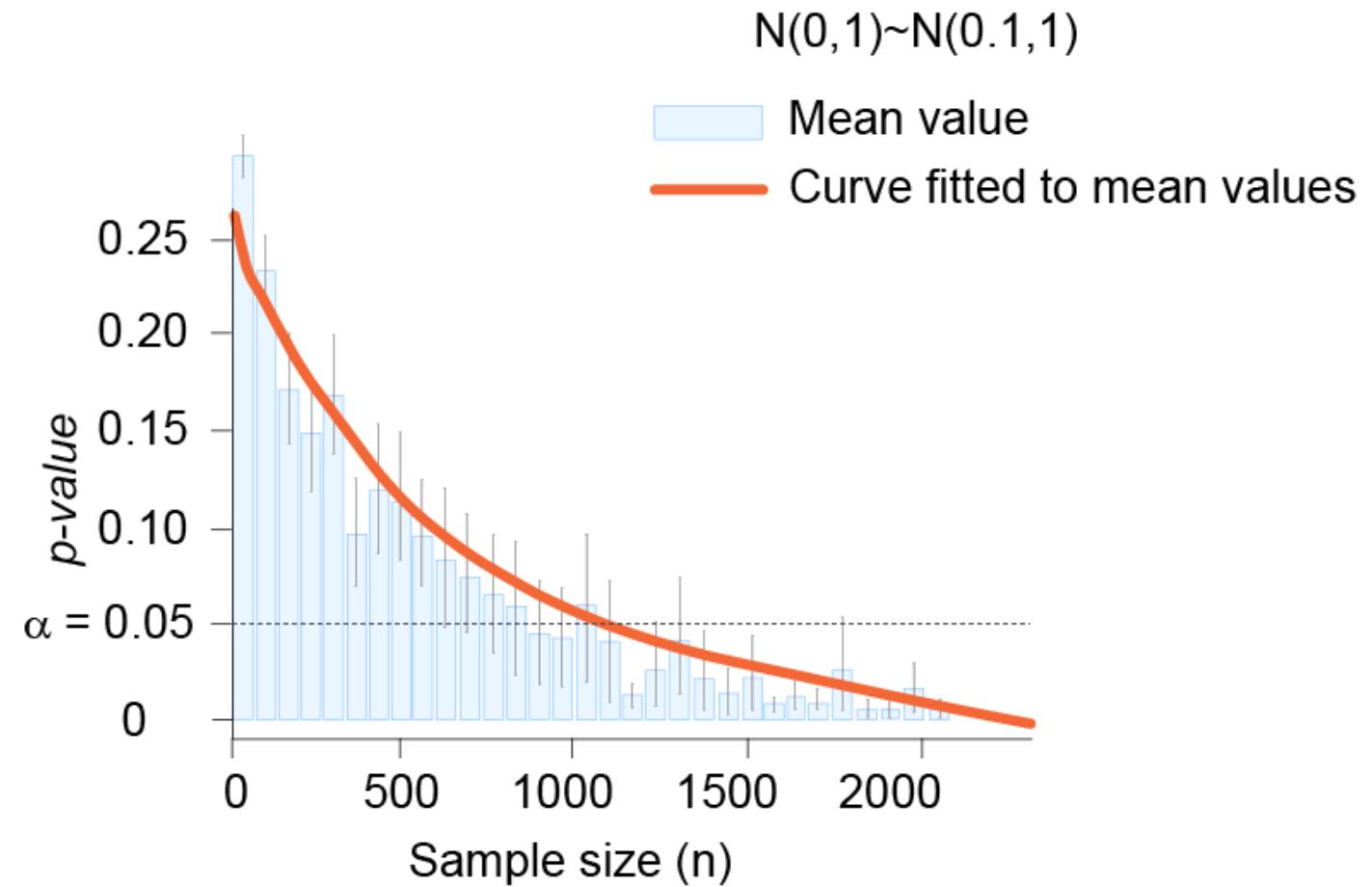
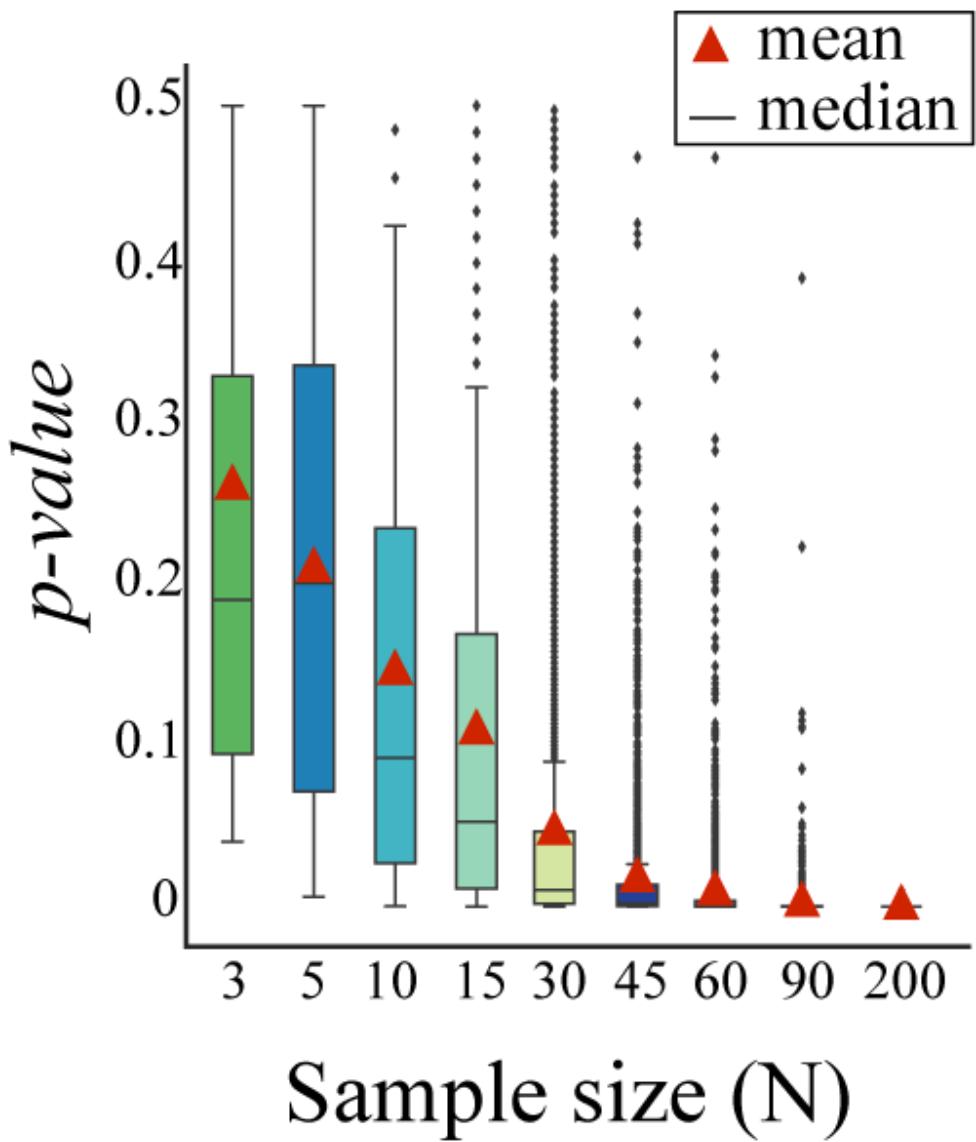
Fact: p -values follow a distribution

Problem I: In similar datasets, the distribution is uniform but not every p -value is always > 0.05

Problem II: If datasets are NOT EXACTLY the same, the p -value behaves as a function that depends on n .

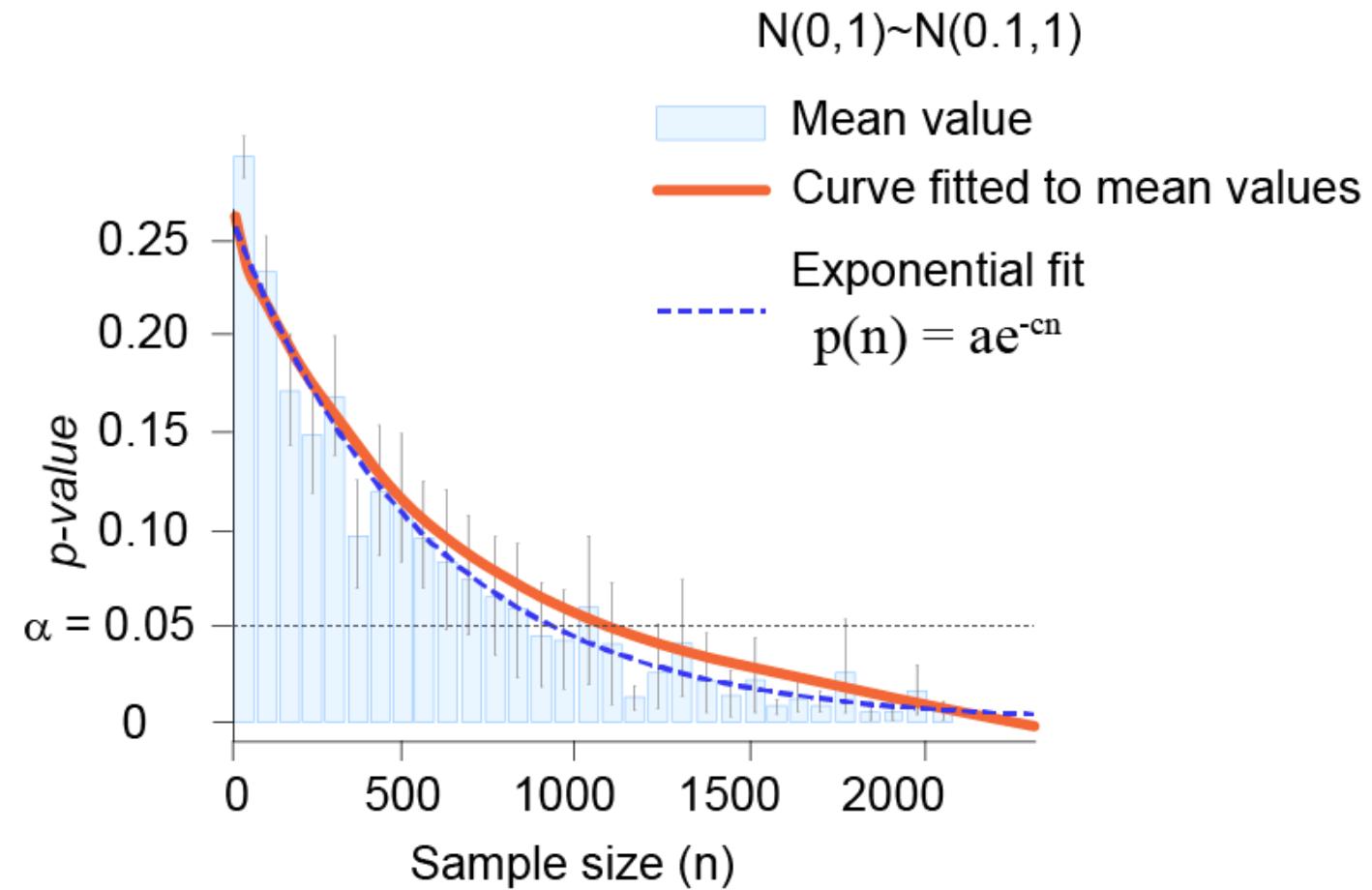
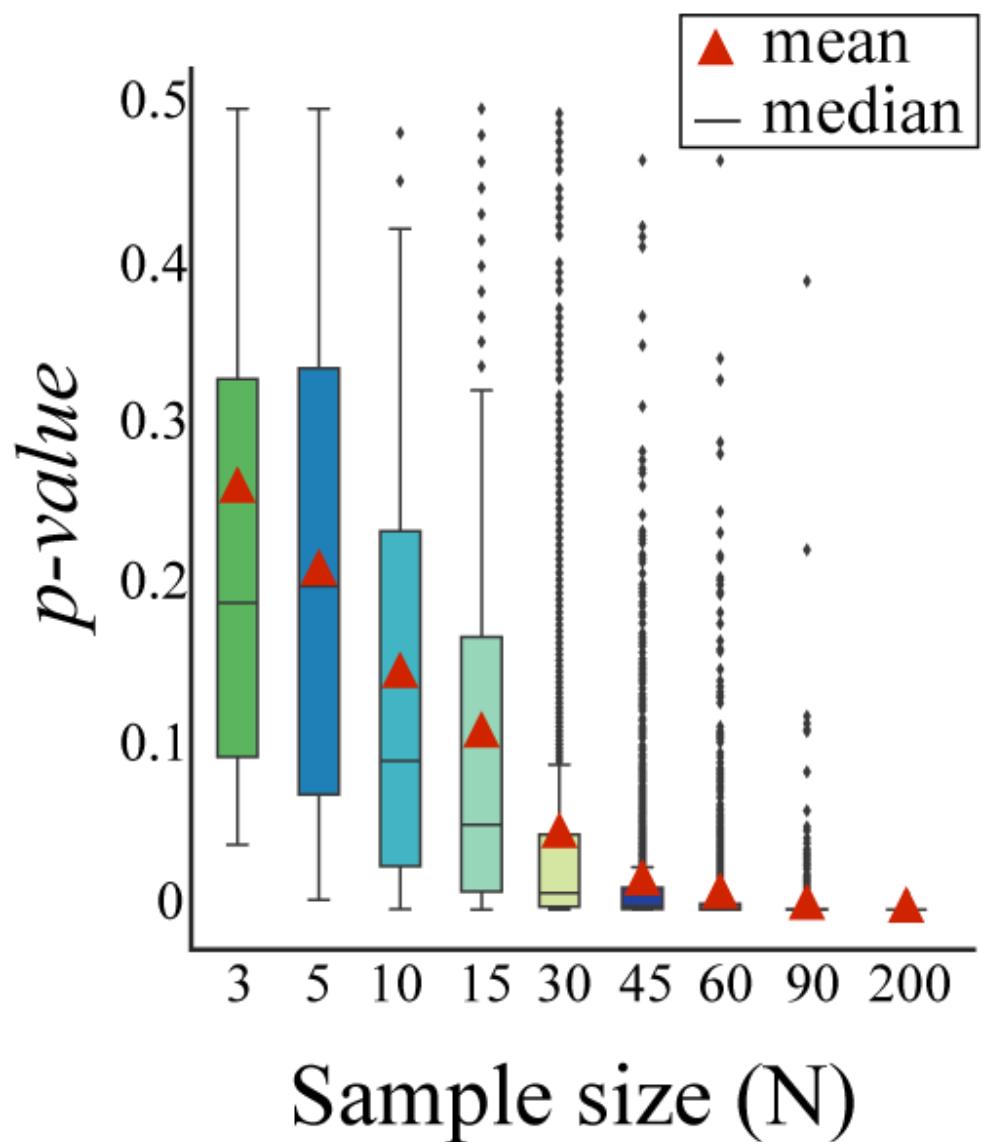


p-values change with the size of the sample



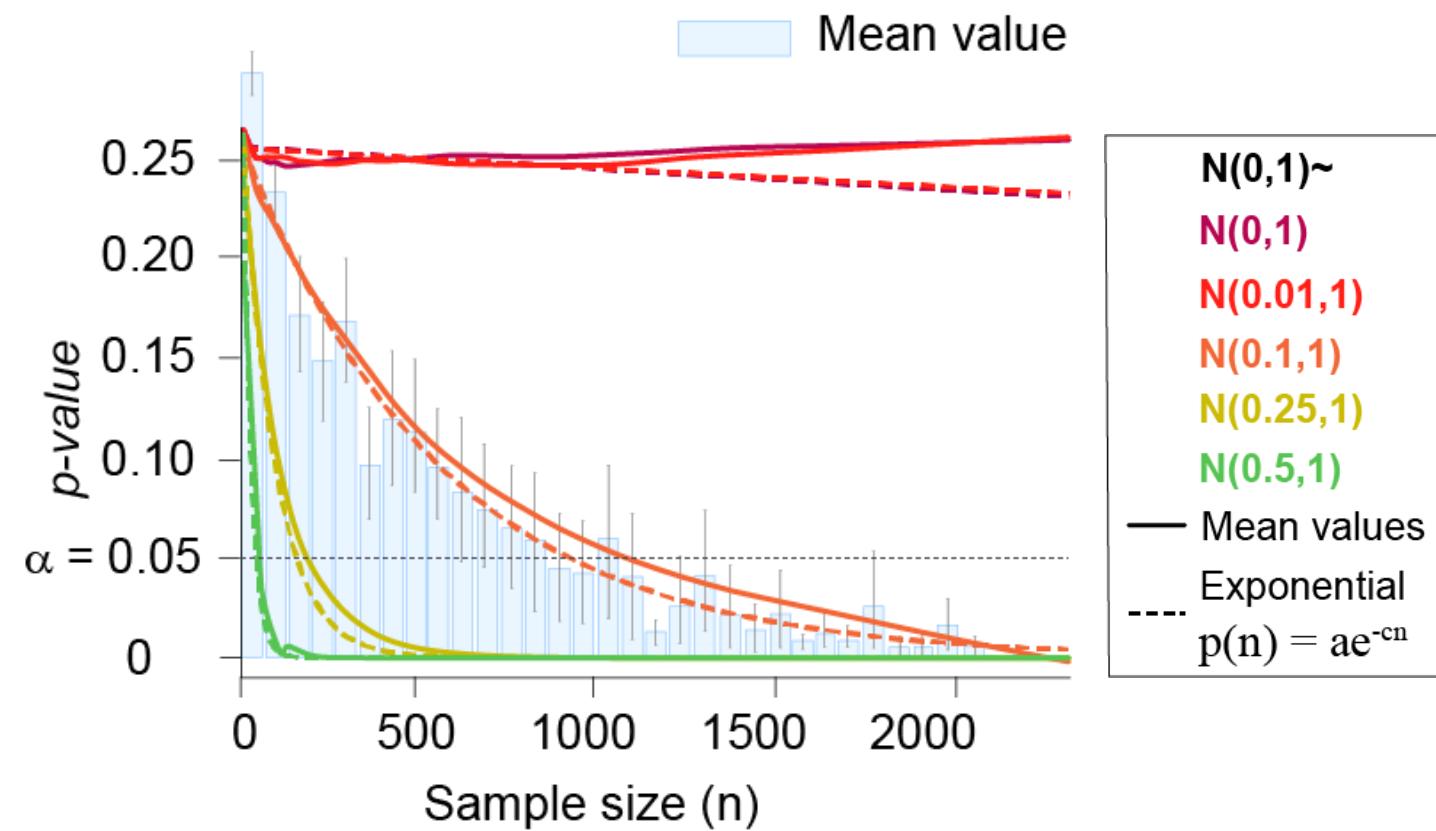
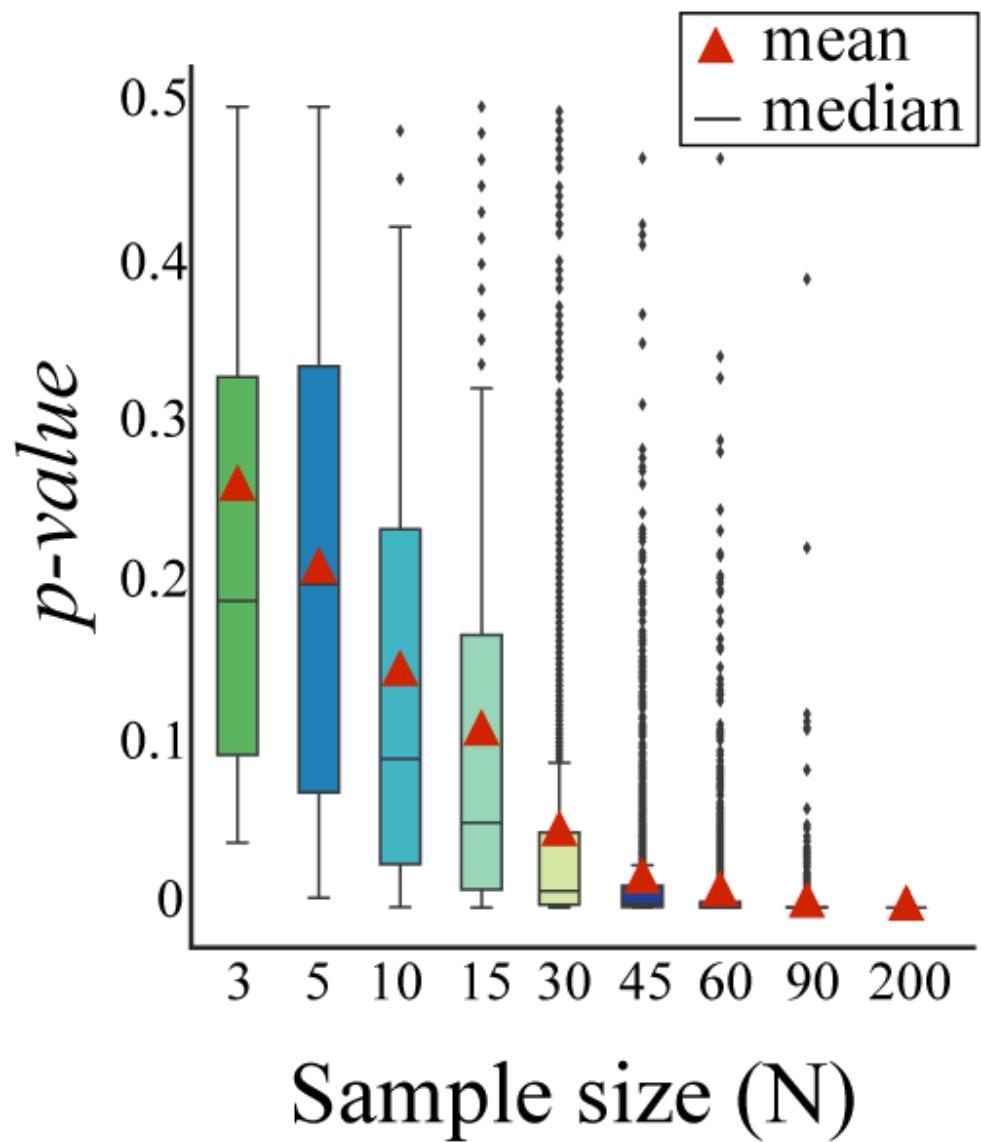
Take the mean values and adjust a smooth curve

p-values change with the size of the sample



Take the mean values and adjust a smooth curve

p-values change with the size of the sample

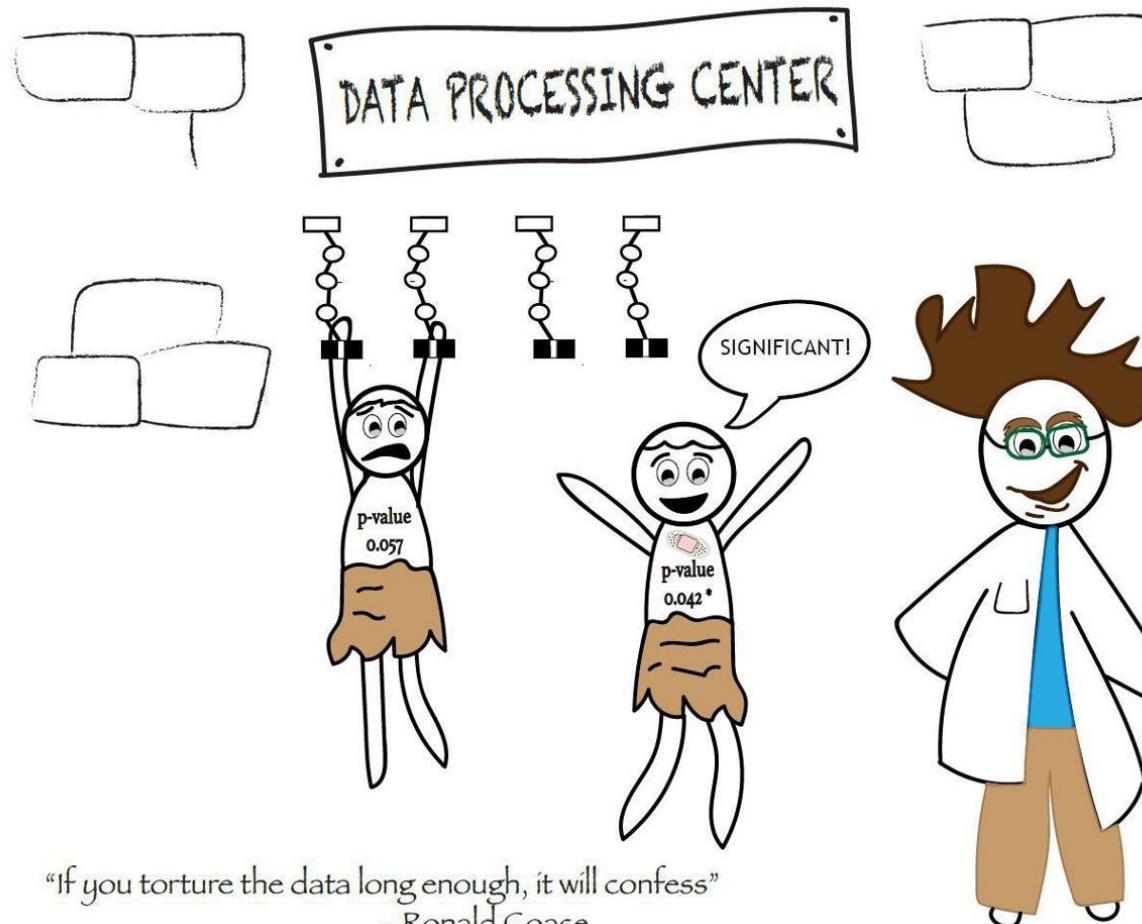


Take the mean values and adjust a smooth curve

The p-value can be modeled as a continuous exponential function

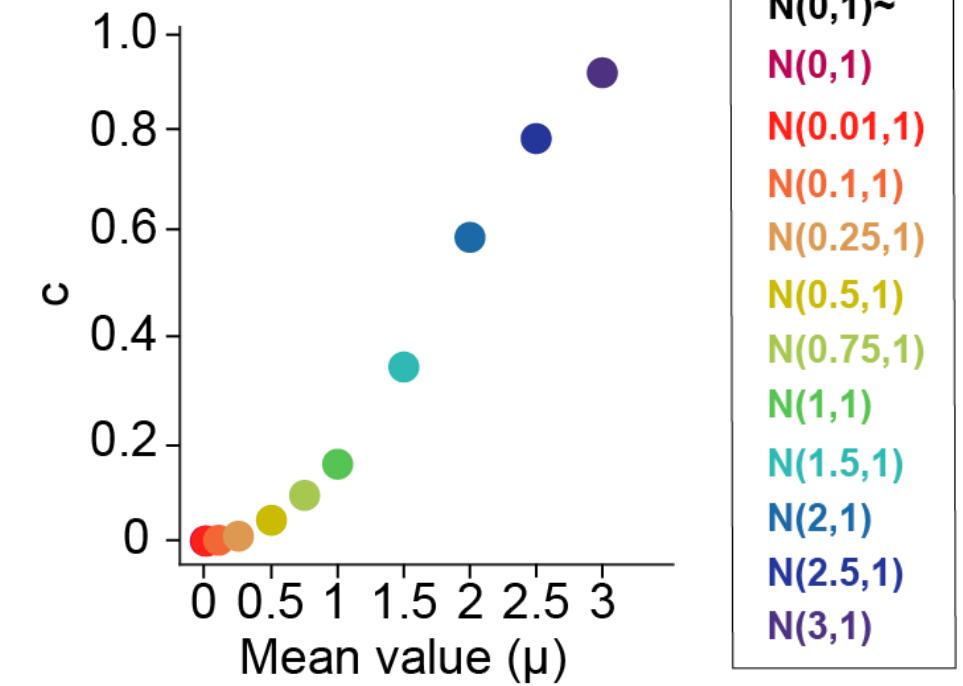
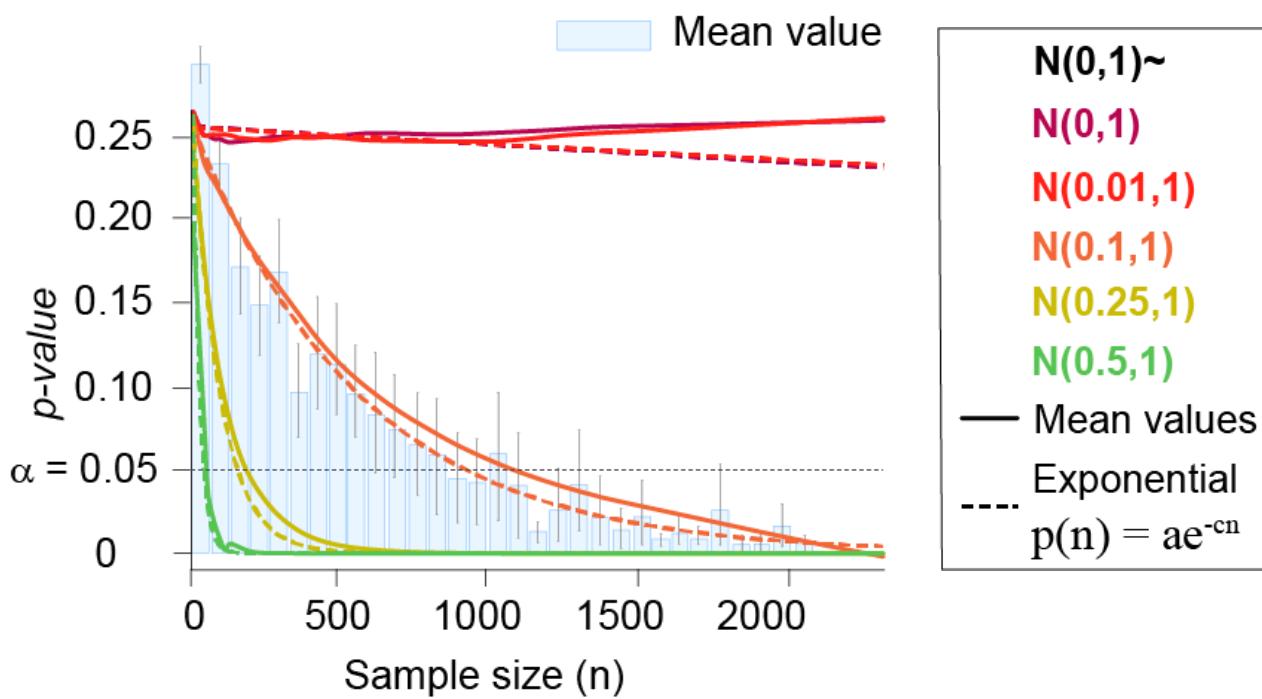
Under the same conditions, the *p-value* is different depending on the size of the sample.

→ It is possible to get the **DESIRED** p-value with a large enough dataset: **p-HACKING**.



The p-value can be modeled as a continuous exponential function

New model: $p(n) \cong a \cdot e^{-cn}$ where $a, c \in \mathbb{R}^+$

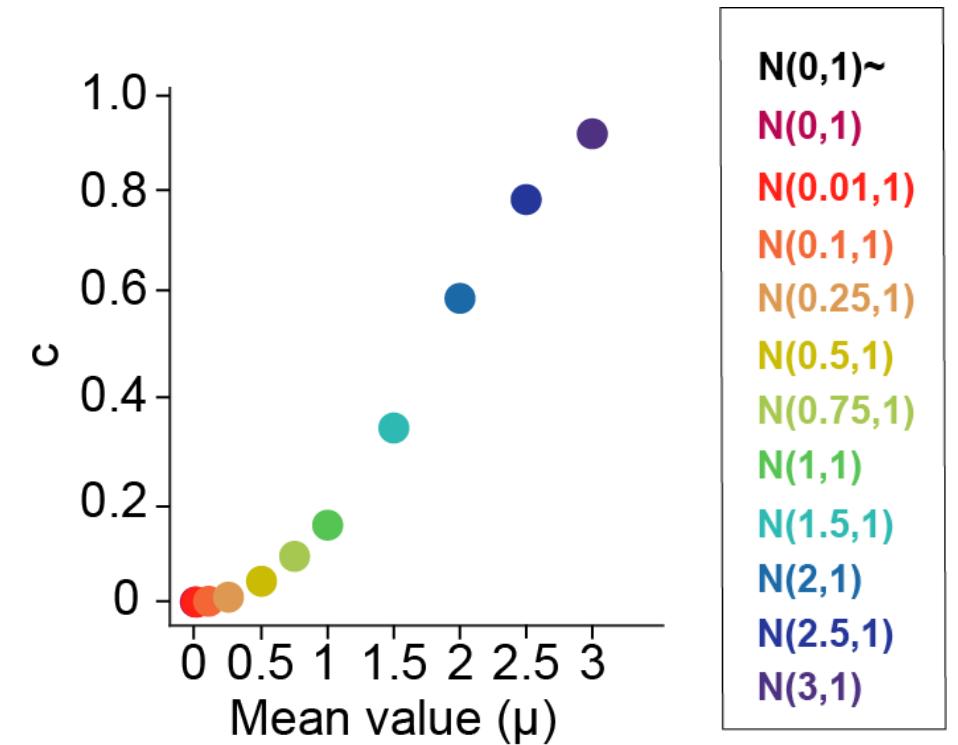


The p-value can be modeled as a continuous exponential function

New model: $p(n) \cong a \cdot e^{-cn}$ where $a, c \in \mathbb{R}^+$

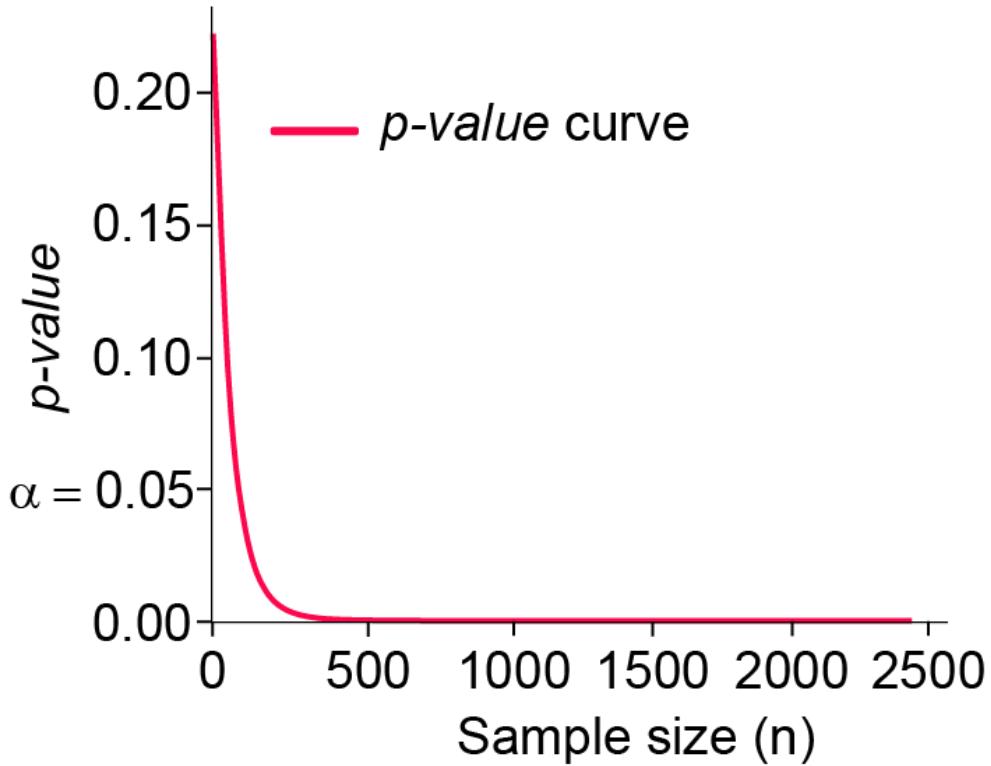
The larger the difference between the compared groups, the larger the decay

Statistical significance \leftrightarrow decay (c)



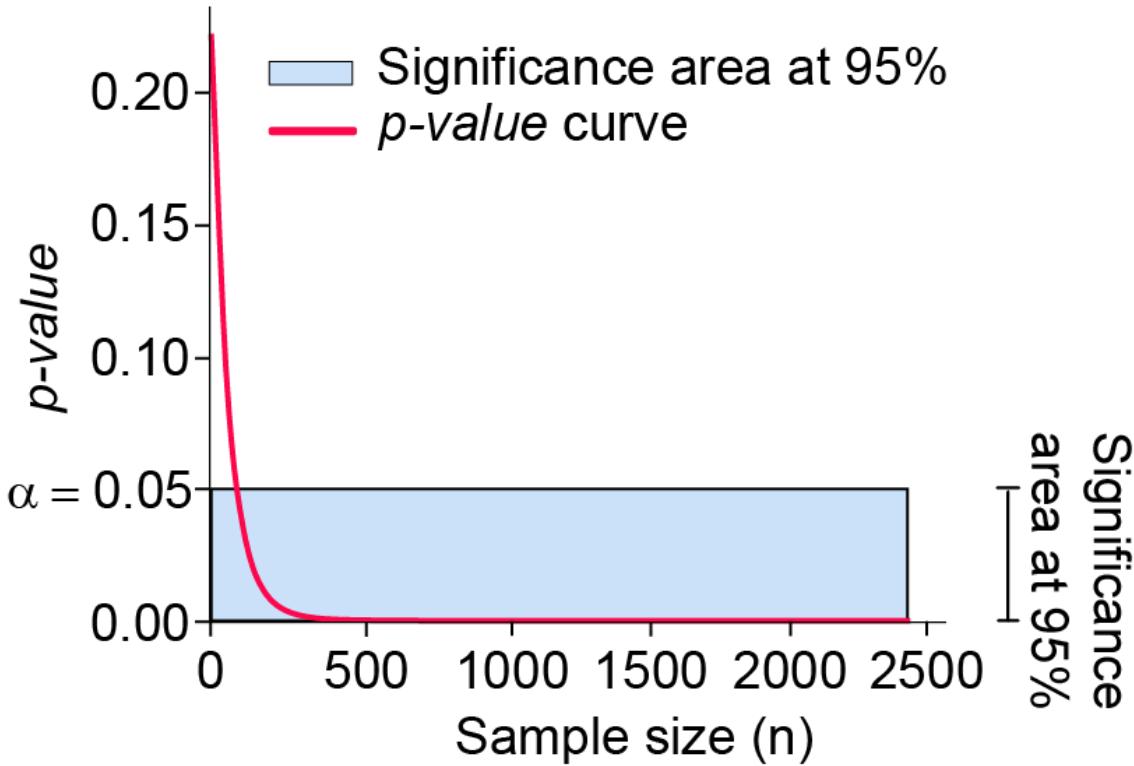
Assessment of the null hypothesis (H_0)

$$p(n) \approx a \cdot e^{-cn} \text{ where } a, c \in \mathbb{R}^+$$



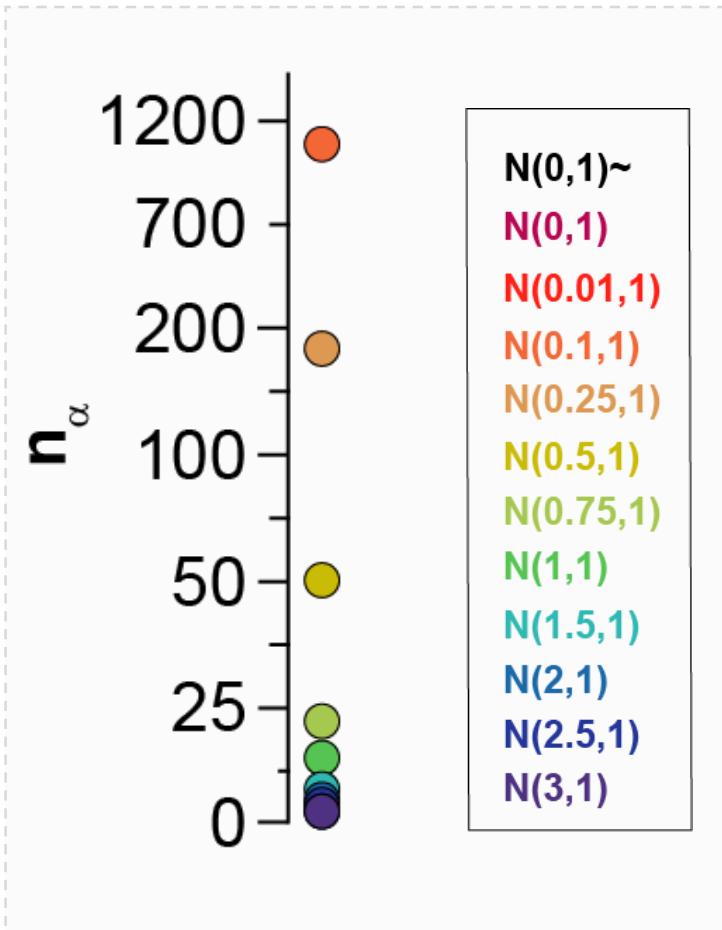
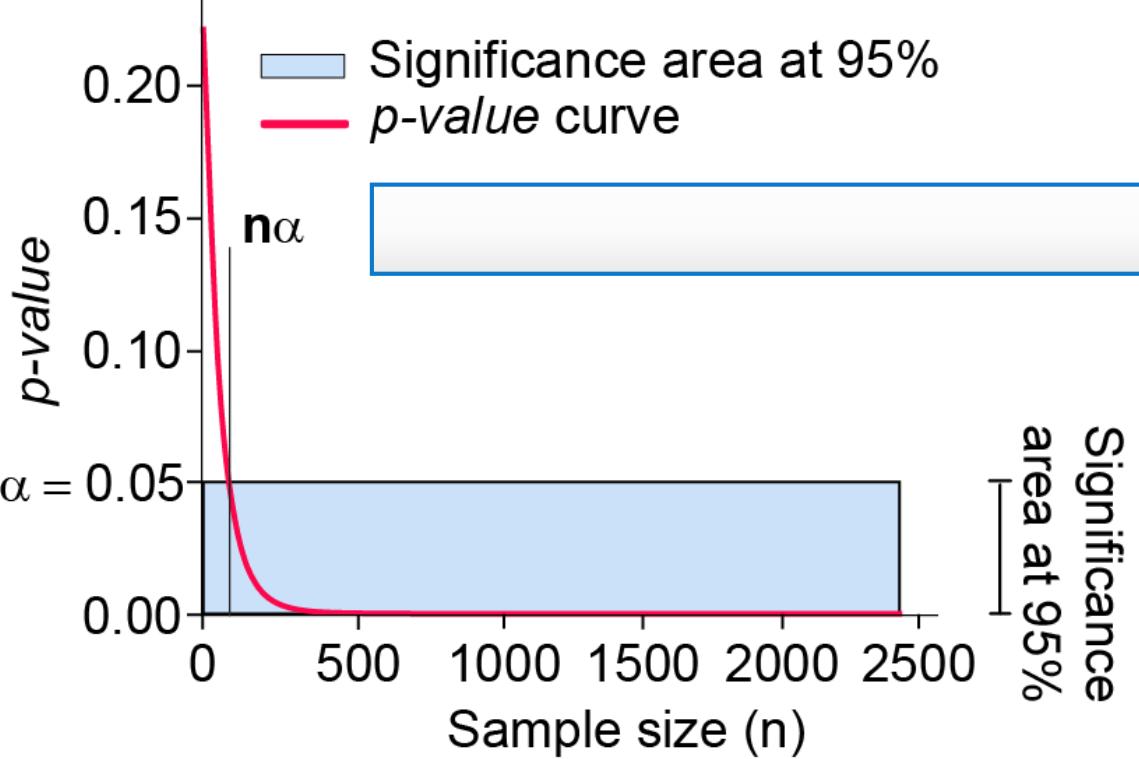
Assessment of the null hypothesis (H_0)

$$p(n) \approx a \cdot e^{-cn} \text{ where } a, c \in \mathbb{R}^+$$



Assessment of the null hypothesis (H_0)

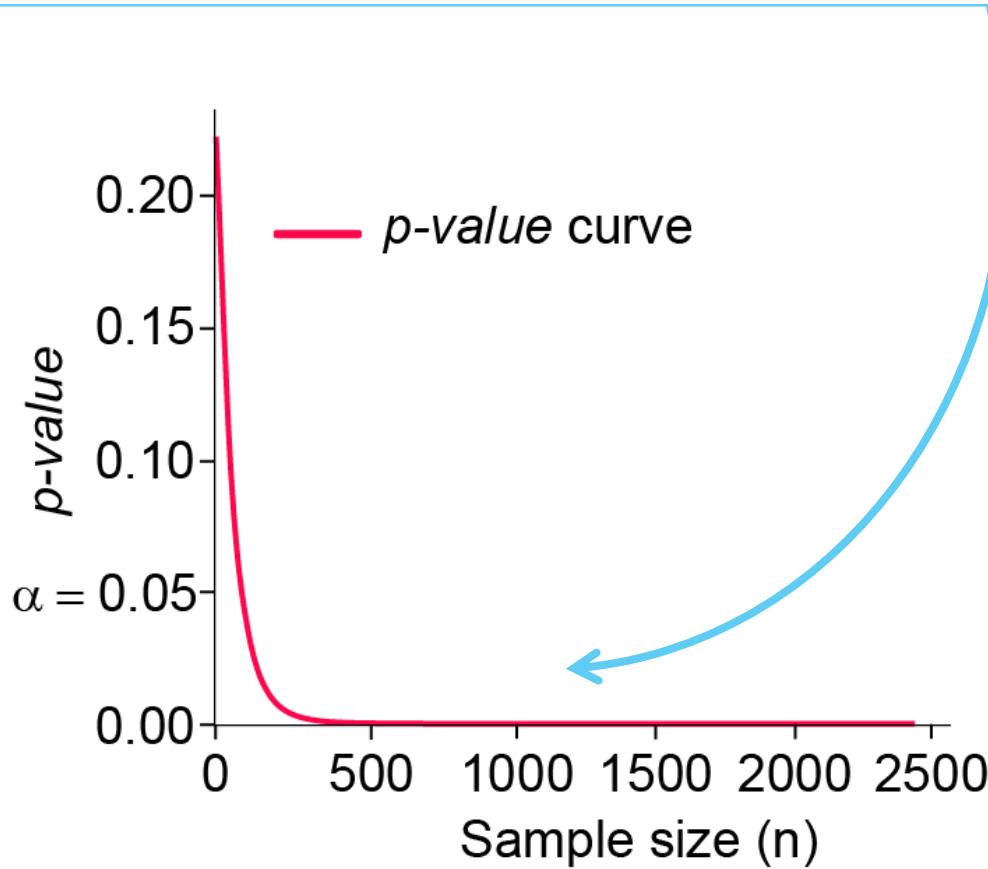
$$p(n) \approx a \cdot e^{-cn} \text{ where } a, c \in \mathbb{R}^+$$



Assessment of the null hypothesis (H_0)

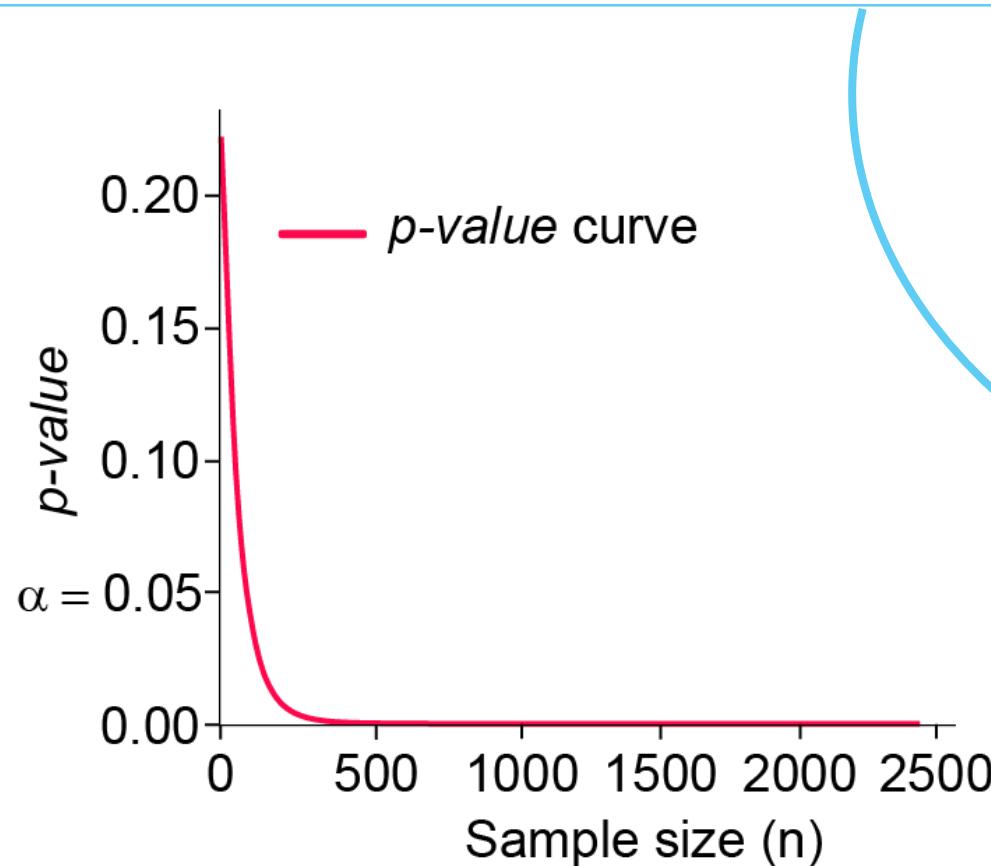
If our data holds with H_0 , most of the *p-values* will be smaller than 0.05

$$p(n) \cong a \cdot e^{-cn} \text{ where } a, c \in \mathbb{R}^+$$

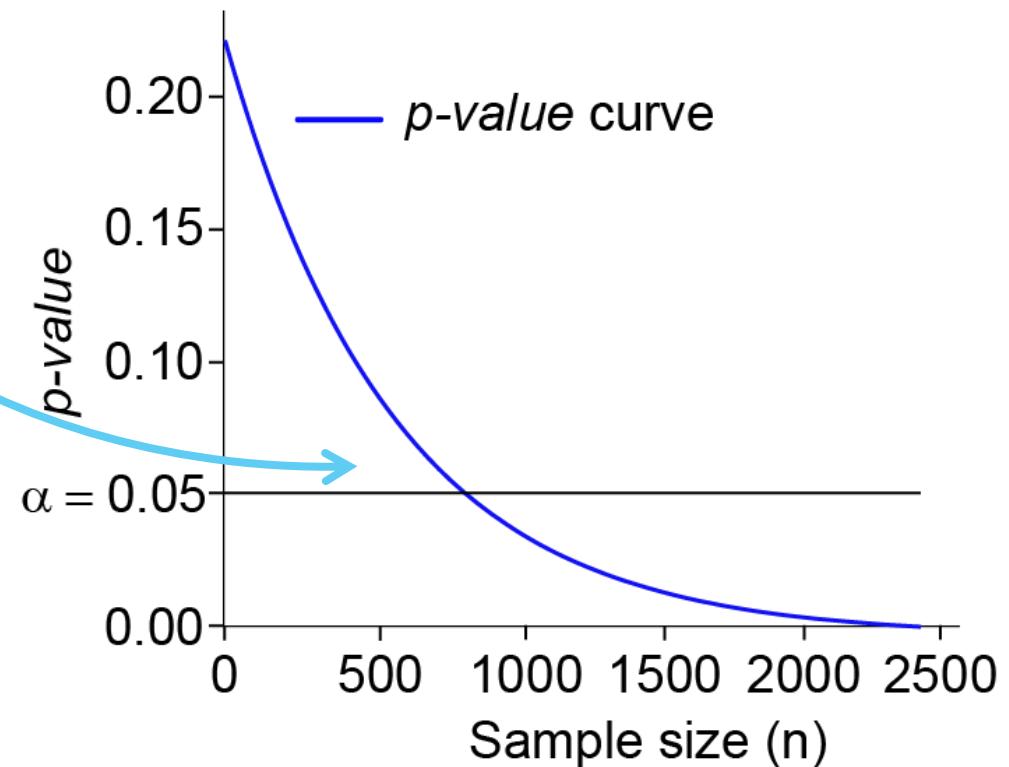


Assessment of the null hypothesis (H_0)

If our data DOES NOT hold with H_0 , most of the *p-values* will be larger than 0.05

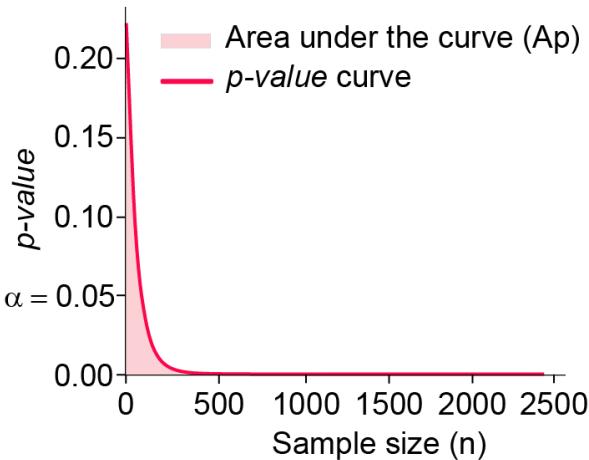
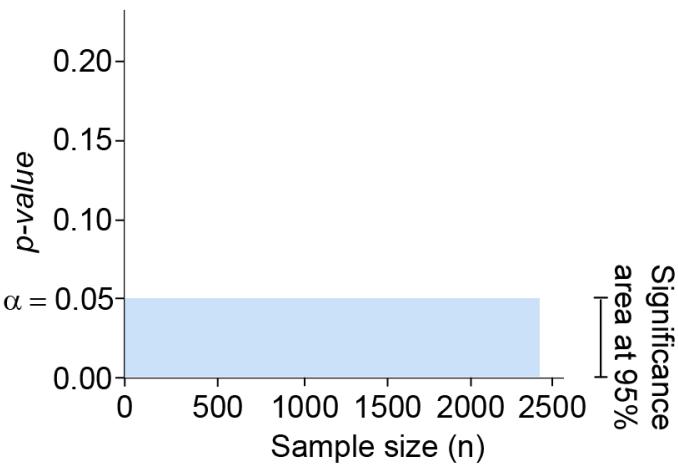


$$p(n) \approx a \cdot e^{-cn} \text{ where } a, c \in \mathbb{R}^+$$



Assessment of the null hypothesis (H_0)

$$p(n) \approx a \cdot e^{-cn} \text{ where } a, c \in \mathbb{R}^+$$



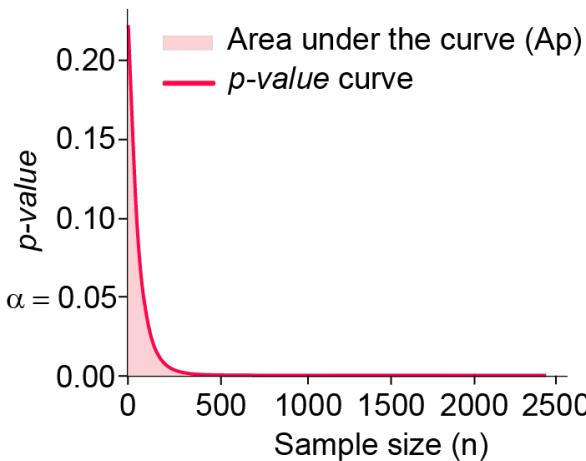
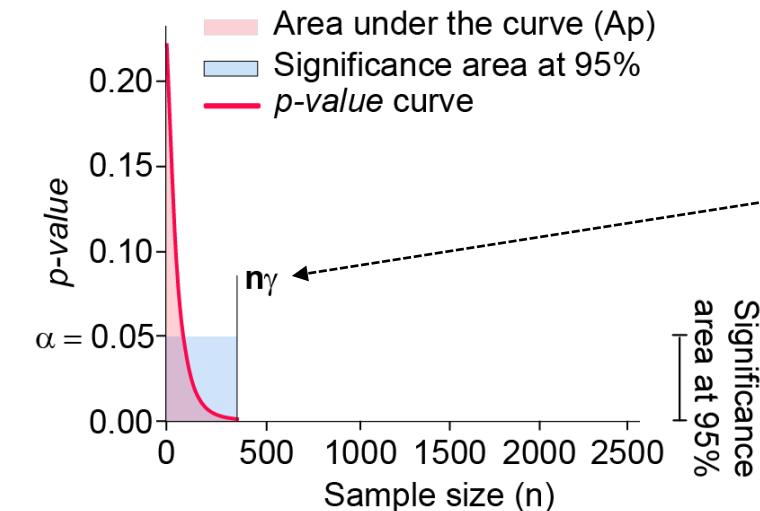
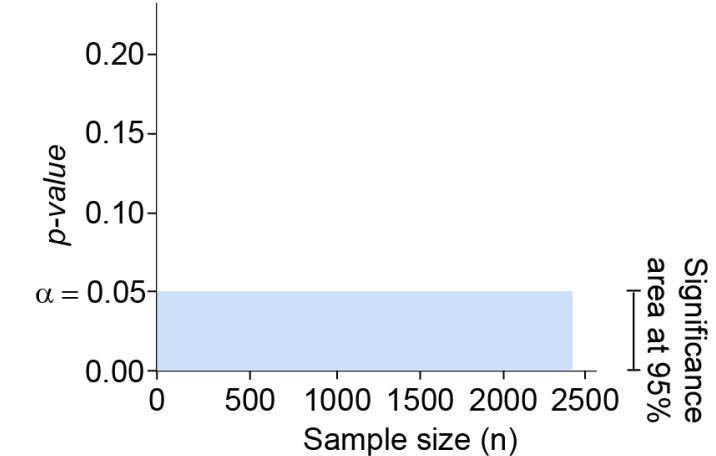
Compare $p(n)$ with α for each value of n

i.e. compare the area under the curve (Ap)
with the significance area ($A\alpha$):

$$\delta = A\alpha - Ap$$



Assessment of the null hypothesis (H_0)



$$n_\gamma = \operatorname{argmin}_n \left\{ \left| \frac{\partial p(n)}{\partial n} \right| < \gamma \right\}$$

and γ is constrained:

- $\gamma < 0.1$
- $\frac{\gamma}{c} < \alpha$

$$p(n) \approx a \cdot e^{-cn} \text{ where } a, c \in \mathbb{R}^+$$

Compare $p(n)$ with α for each value of n

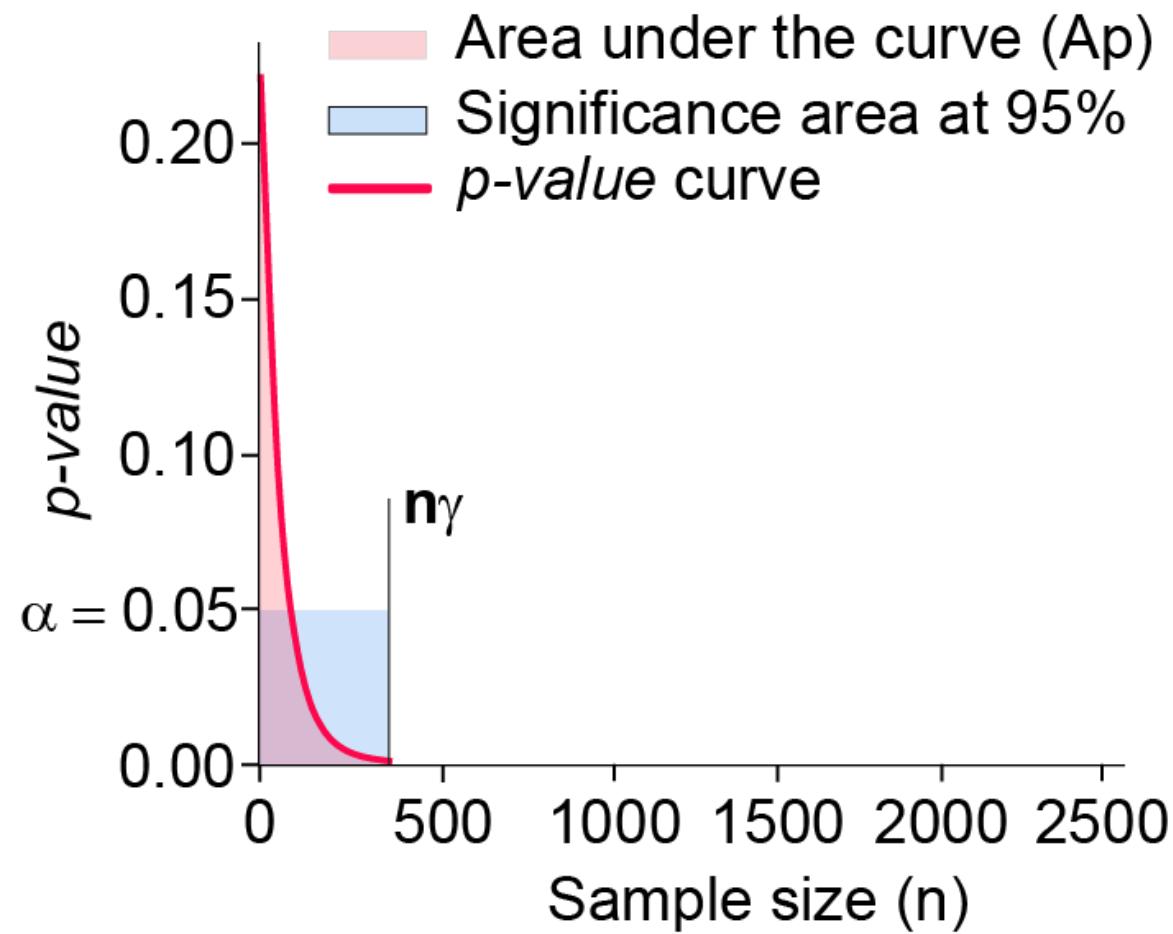


i.e. compare the area under the curve (A_p)

with the significance area (A_α):

$$\delta = A_\alpha - A_p$$

Assessment of the null hypothesis (H_0)



$$p(n) \approx a \cdot e^{-cn} \text{ where } a, c \in \mathbb{R}^+$$

Compare $p(n)$ with α for each value of n

i.e. compare the area under the curve (A_p)
with the significance area (A_α):

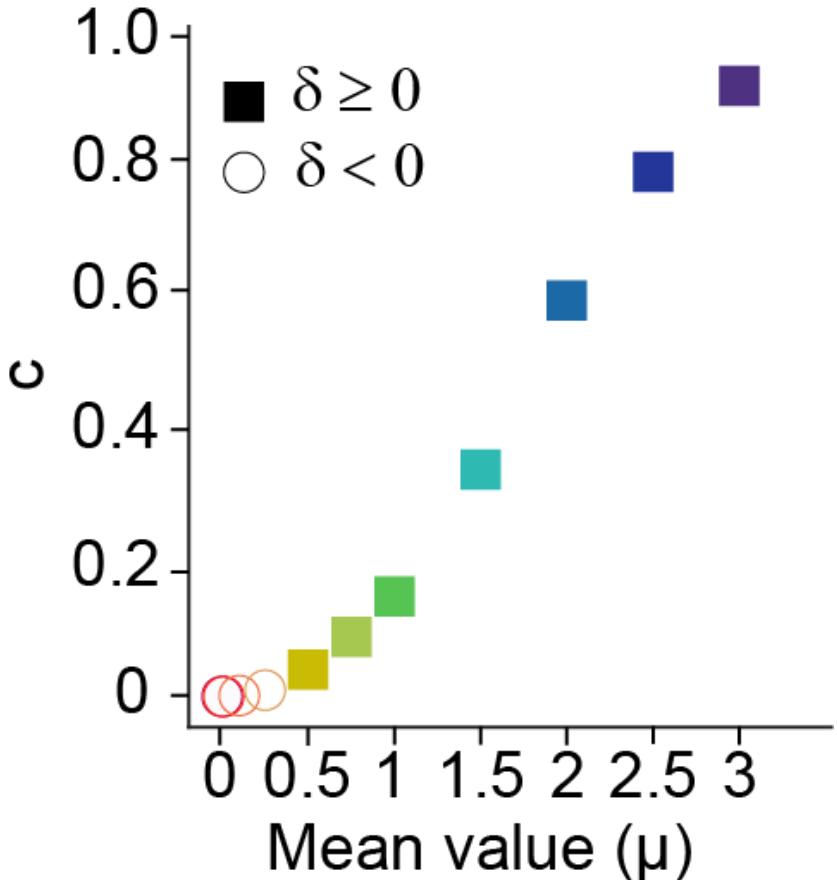
$$\delta = A_\alpha - A_p$$

Use δ to evaluate H_0 :
 $\delta > 0 \rightarrow \text{Reject } H_0$

Assessment of the null hypothesis (H_0)

H_0 : two samples come from the same distribution

$N(0,1) \sim$
 $N(0,1)$
 $N(0.01,1)$
 $N(0.1,1)$
 $N(0.25,1)$
 $N(0.5,1)$
 $N(0.75,1)$
 $N(1,1)$
 $N(1.5,1)$
 $N(2,1)$
 $N(2.5,1)$
 $N(3,1)$



$$p(n) \cong a \cdot e^{-cn} \text{ where } a, c \in \mathbb{R}^+$$

Compare $p(n)$ with α for each value of n

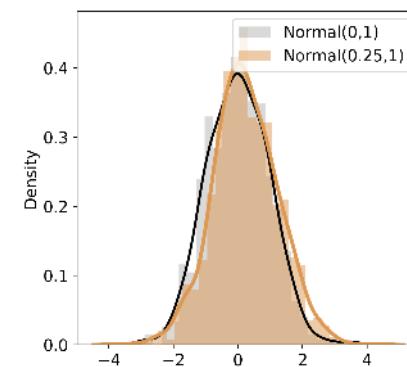
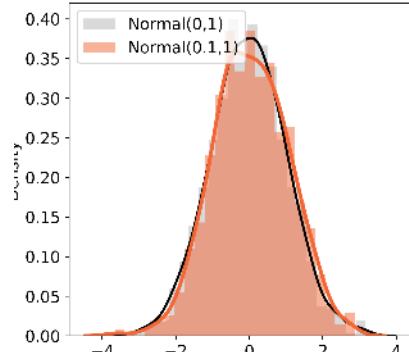
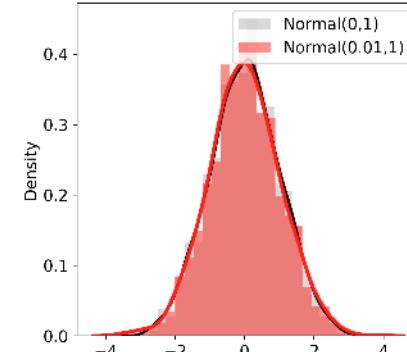
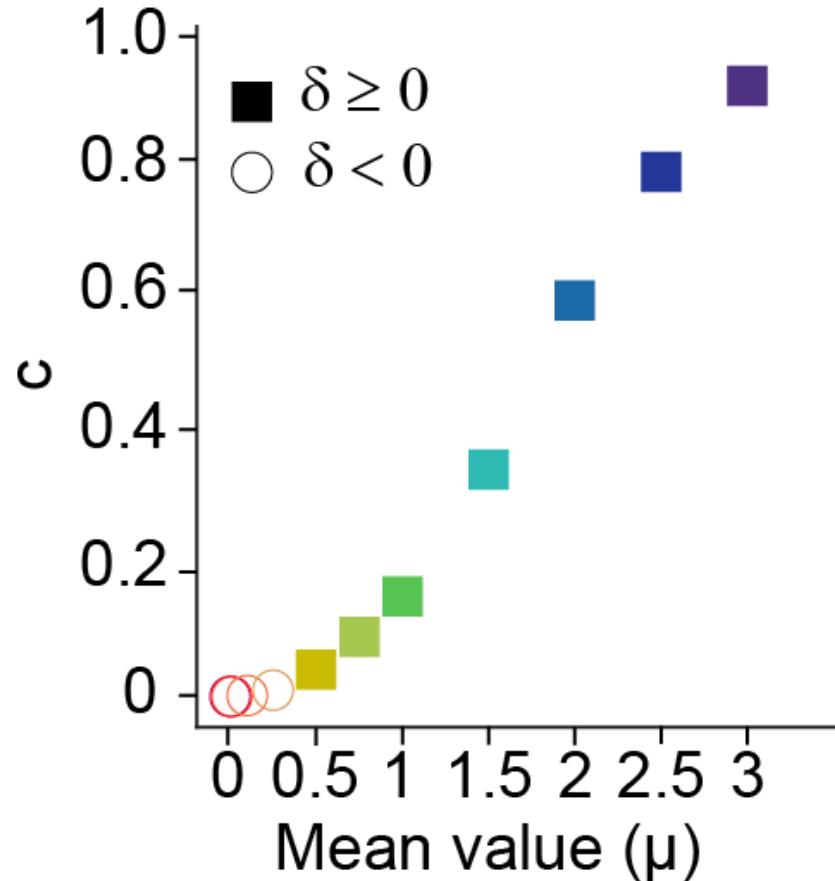
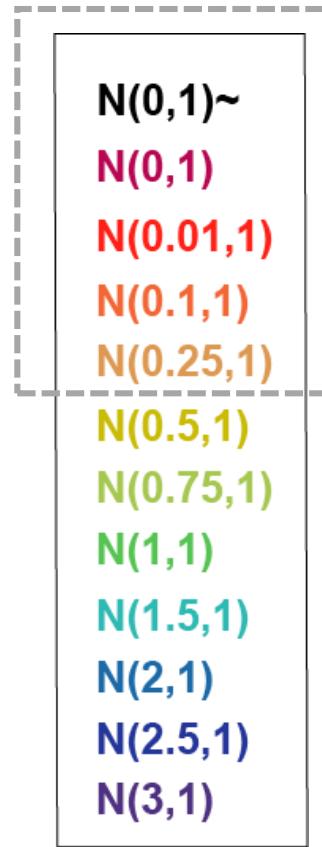
i.e. compare the area under the curve (A_p)
with the significance area (A_α):

$$\delta = A_\alpha - A_p$$

Use δ to evaluate H_0 :
 $\delta > 0 \rightarrow \text{Reject } H_0$

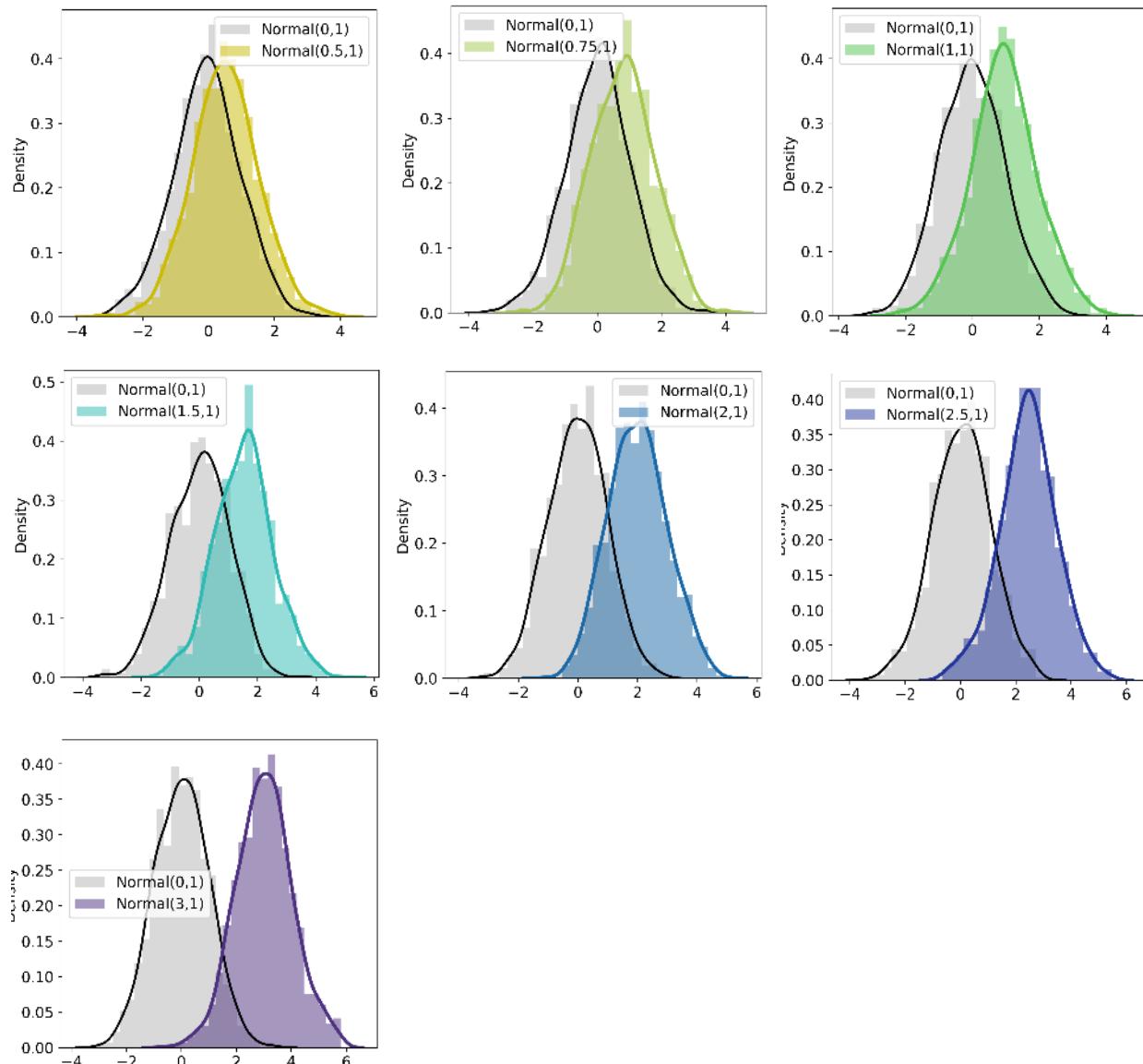
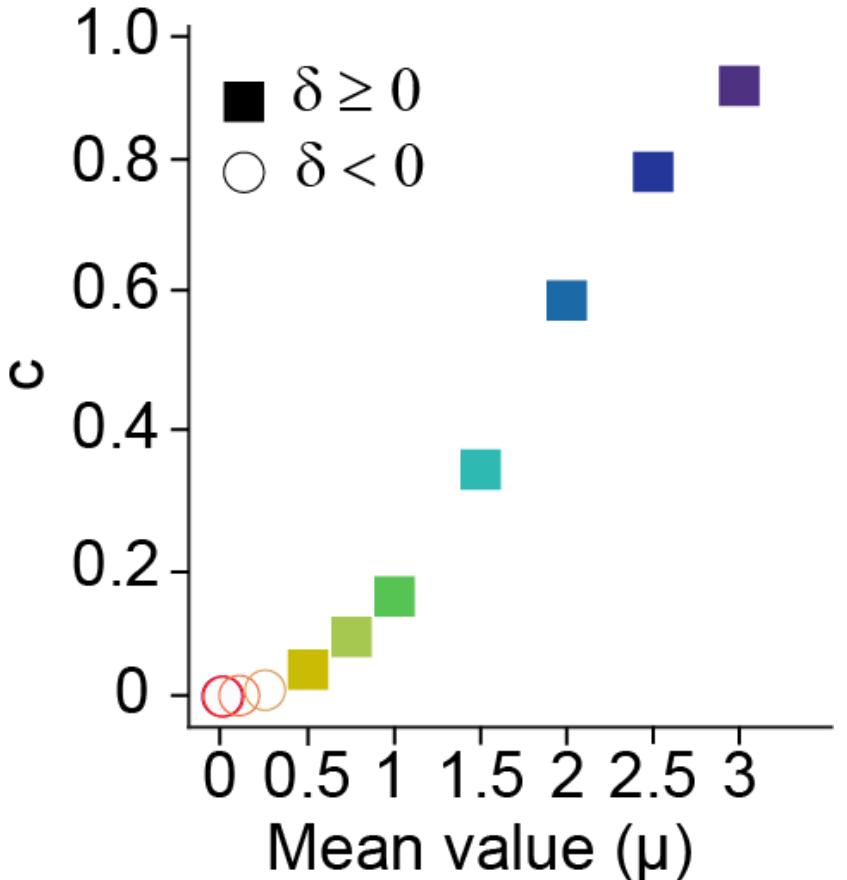
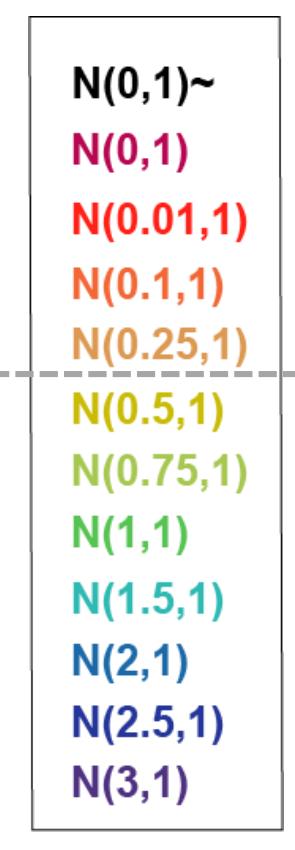
Assessment of the null hypothesis (H_0)

H_0 : two samples come from the same distribution



Assessment of the null hypothesis (H_0)

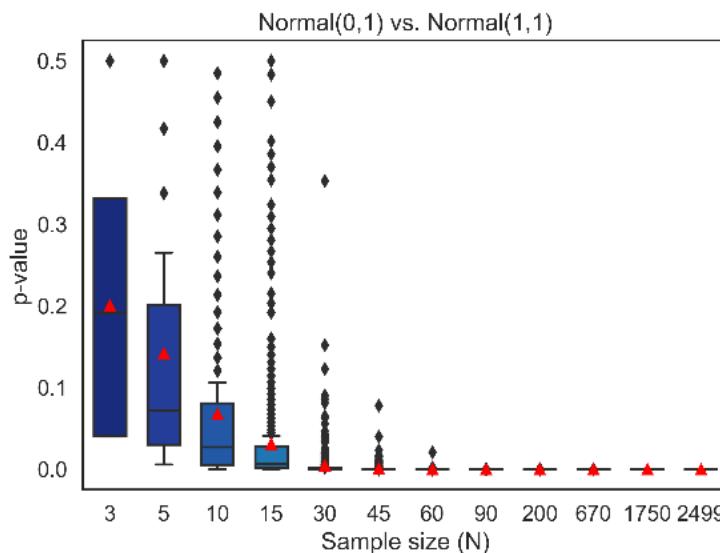
H_0 : two samples come from the same distribution



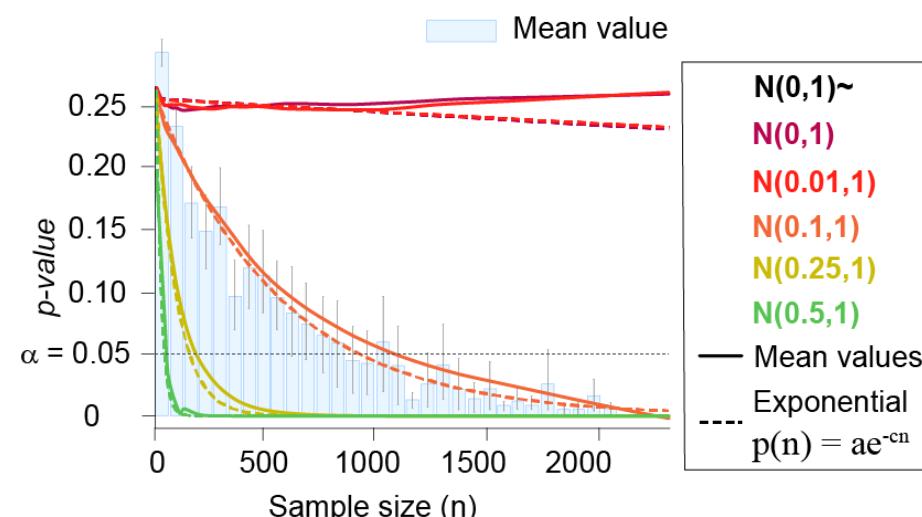
Use cases

Use cases: Workflow for real data analysis

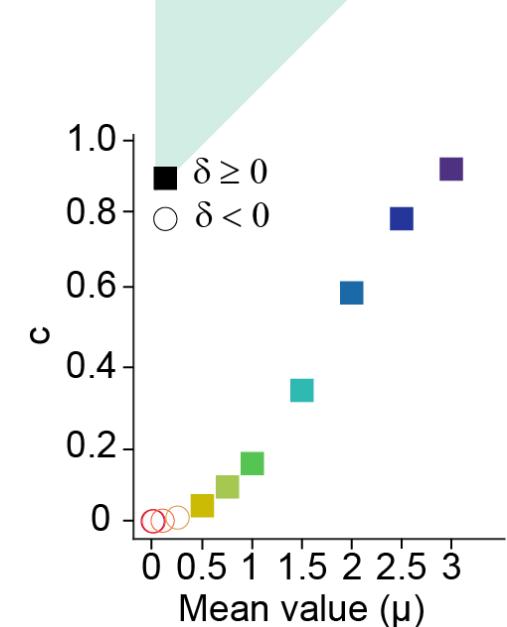
Estimate the distribution of
p-values
Monte Carlo cross validation



Fit the exponential function
 $p(n) \cong a \cdot e^{-cn}$

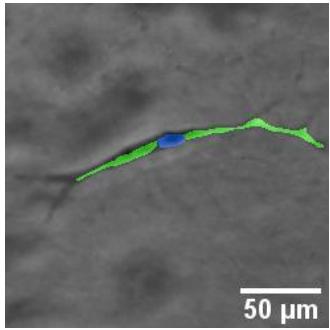


Evaluate H_0 with δ

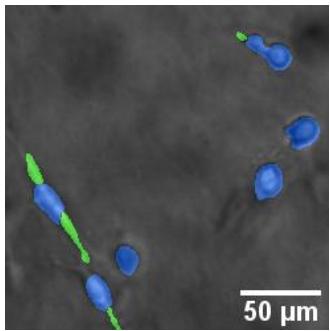


Use cases: I. Evaluation of drug effects

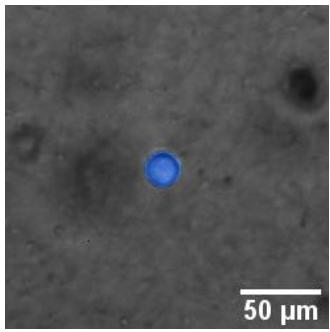
Question: Does Taxol have any effect in the cellular morphology?



Control



1nM Taxol

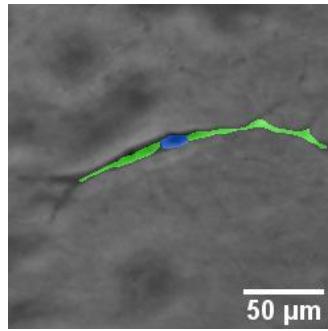


50nM Taxol

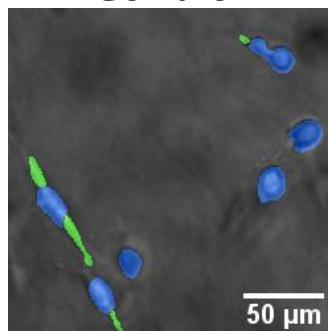
Microscopy images of cancer cells (MDA-MB-231) embedded in a 3D collagen Type I matrix.

Use cases: I. Evaluation of drug effects

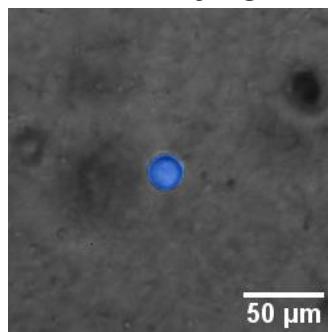
Question: Does Taxol have any effect in the cellular morphology?



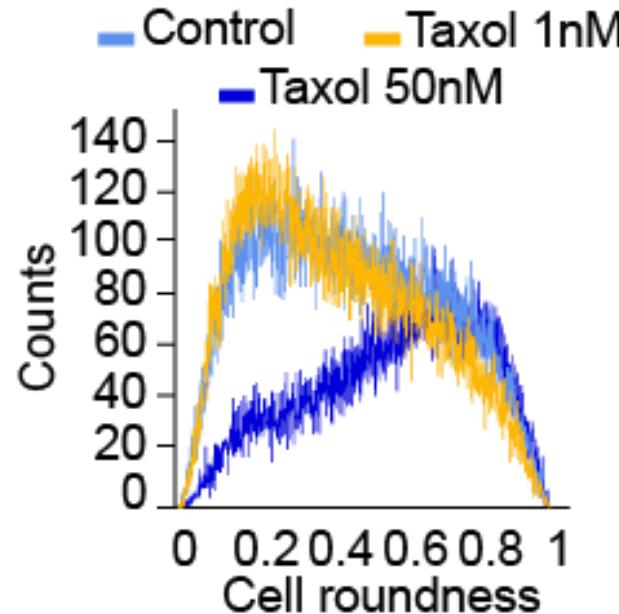
Control



1nM Taxol



50nM Taxol

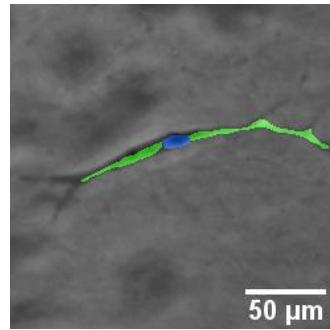


50K data points per group,
so all p-values are much
smaller than 0.05.

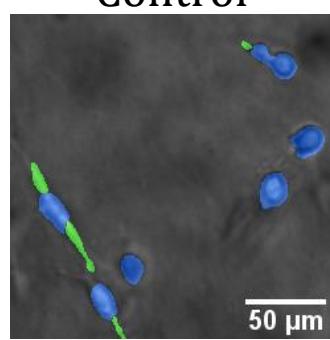
Microscopy images of cancer cells
(MDA-MB-231) embedded in a 3D
collagen Type I matrix.

Use cases: I. Evaluation of drug effects

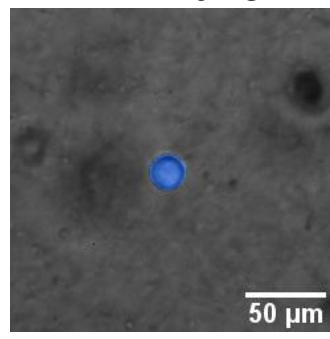
Question: Does Taxol have any effect in the cellular morphology?



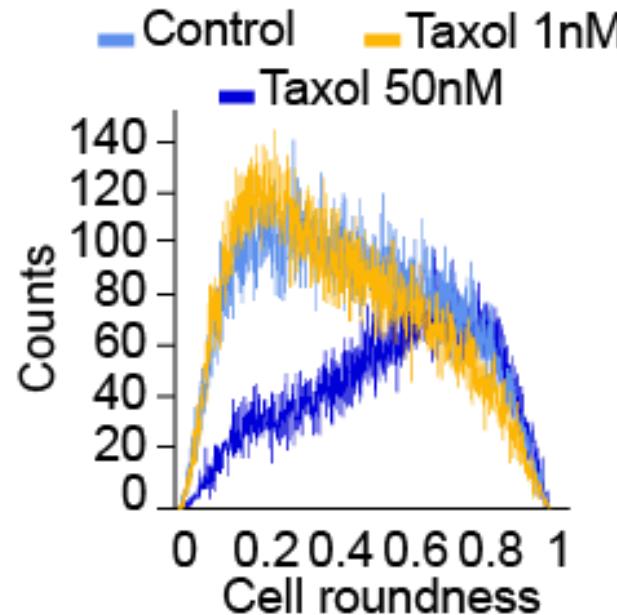
Control



1nM Taxol

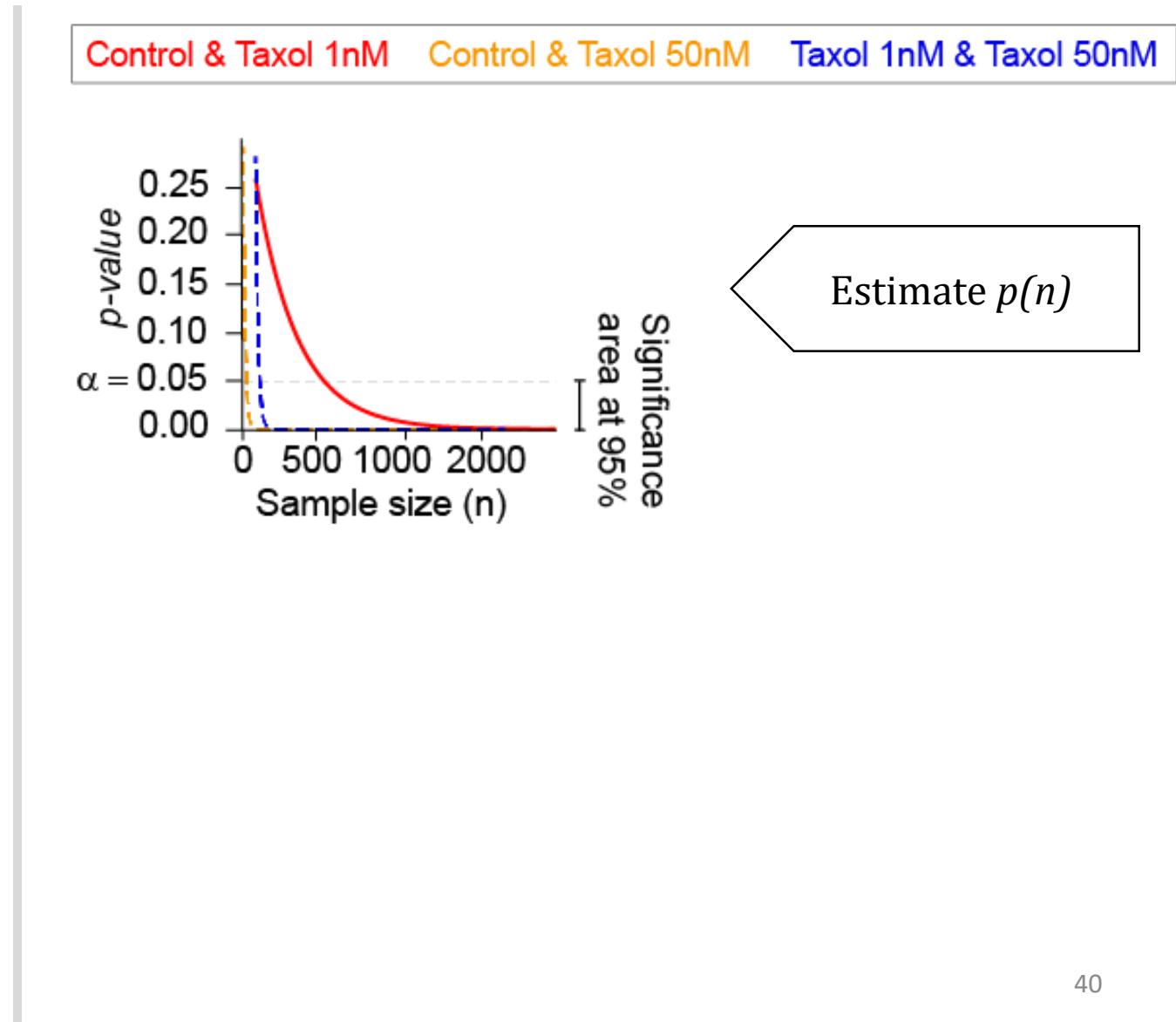


50nM Taxol



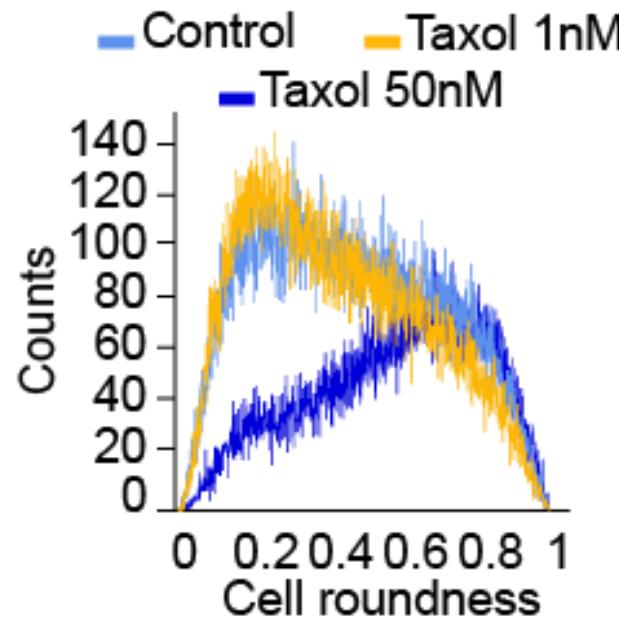
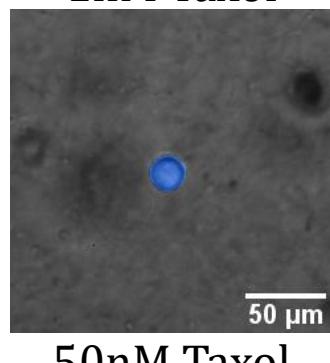
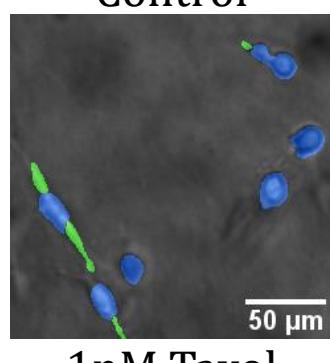
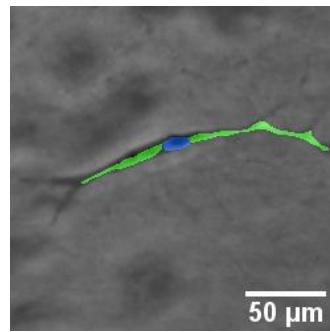
50K data points per group,
so all p-values are much
smaller than 0.05.

Microscopy images of cancer cells
(MDA-MB-231) embedded in a 3D
collagen Type I matrix.



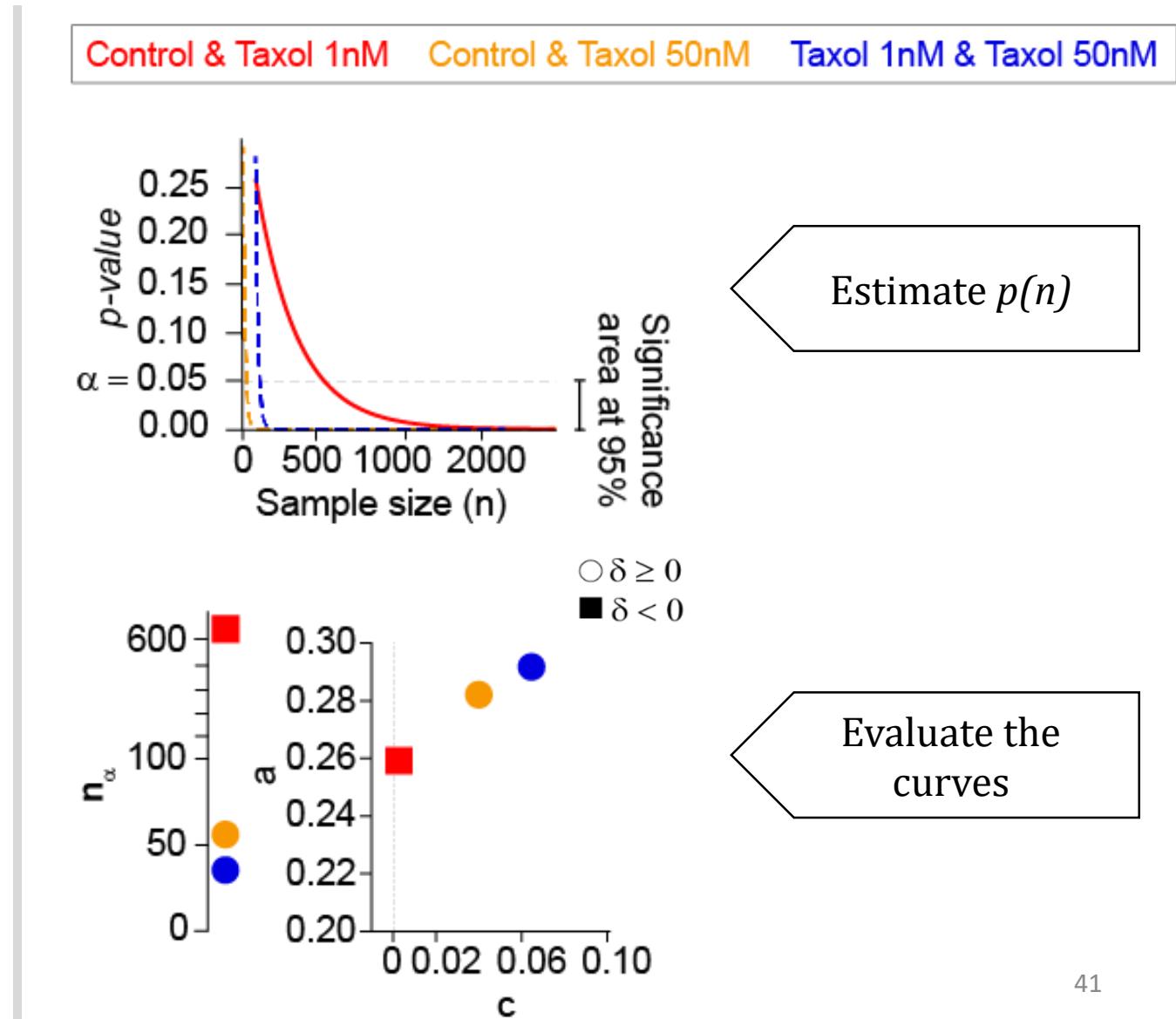
Use cases: I. Evaluation of drug effects

Question: Does Taxol have any effect in the cellular morphology?



50K data points per group,
so all p-values are much
smaller than 0.05.

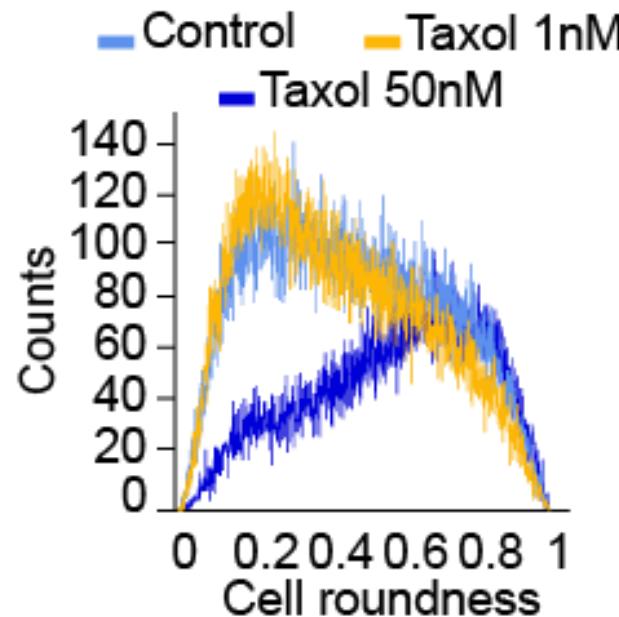
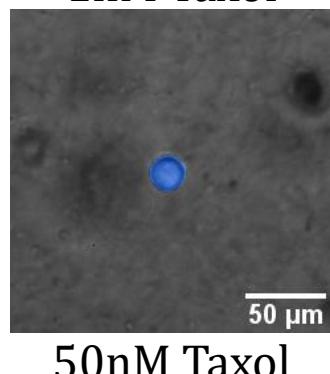
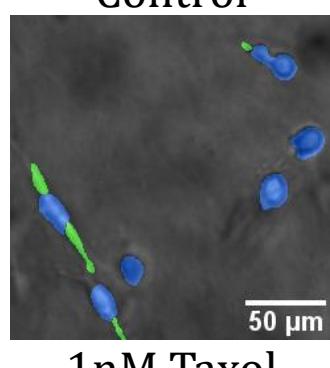
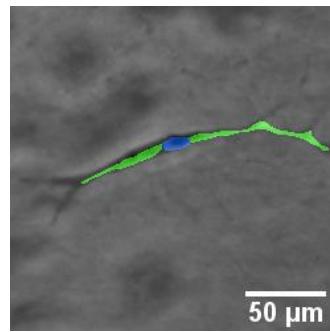
Microscopy images of cancer cells
(MDA-MB-231) embedded in a 3D
collagen Type I matrix.



Evaluate the curves

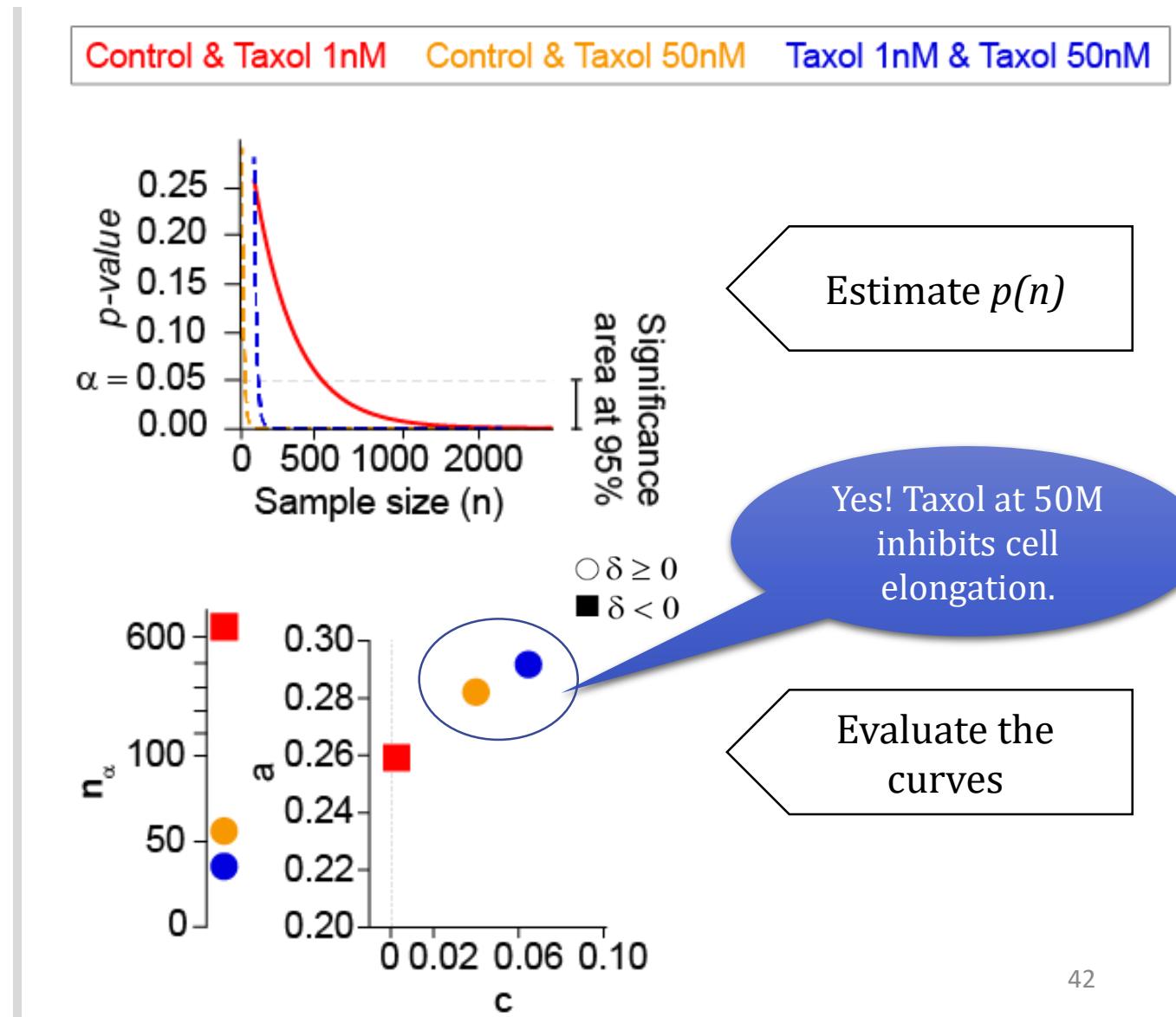
Use cases: I. Evaluation of drug effects

Question: Does Taxol have any effect in the cellular morphology?



50K data points per group,
so all p-values are much
smaller than 0.05.

Microscopy images of cancer cells
(MDA-MB-231) embedded in a 3D
collagen Type I matrix.



Estimate $p(n)$

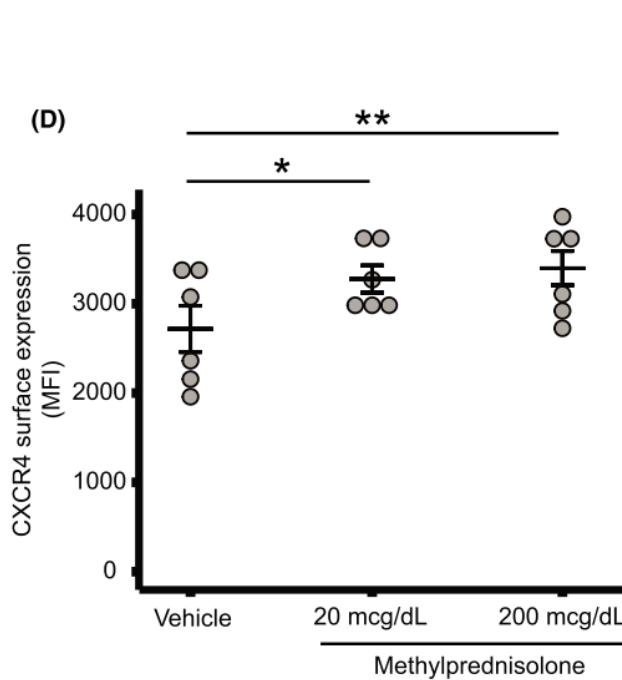
Yes! Taxol at 50nM
inhibits cell
elongation.

Evaluate the
curves

Use cases: II. Evaluation of drug effects

Task: Determine the transcriptional changes induced by the *in vivo* exposure of human eosinophils to glucocorticoids.

Flow cytometry data: >>>10K data points, so all p-values are much smaller than 0.05.

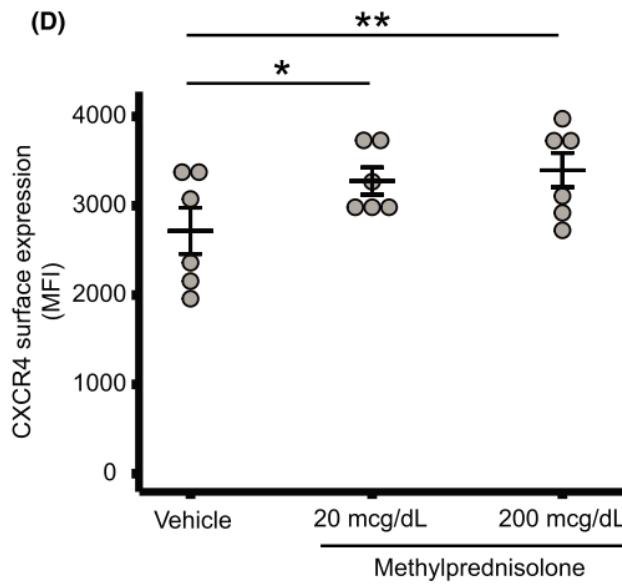


Human eosinophils exposed to Methylprednisolone. Flow cytometry data [1]. MFI: median of fluorescence intensity.

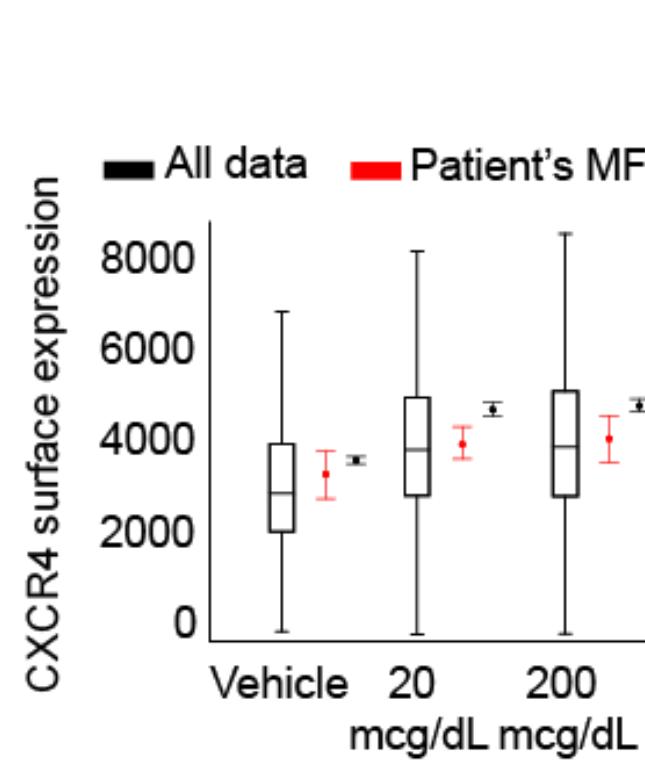
Use cases: II. Evaluation of drug effects

Task: Determine the transcriptional changes induced by the *in vivo* exposure of human eosinophils to glucocorticoids.

Flow cytometry data: >>>10K data points, so all p-values are much smaller than 0.05.



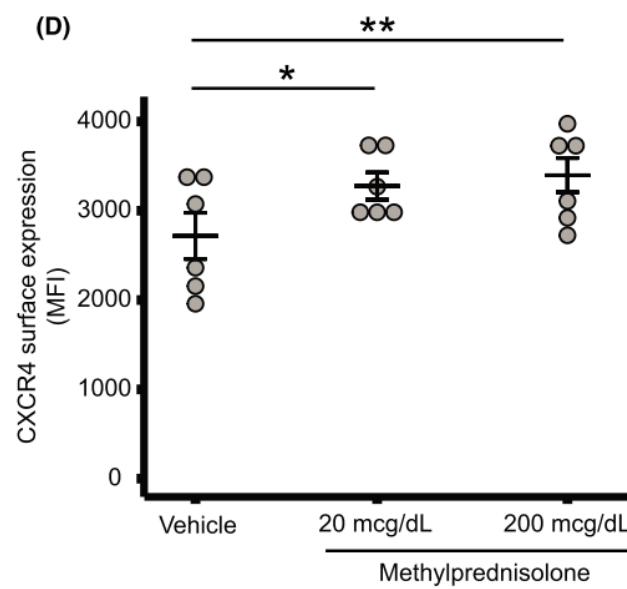
Human eosinophils exposed to Methylprednisolone. Flow cytometry data [1]. MFI: median of fluorescence intensity.



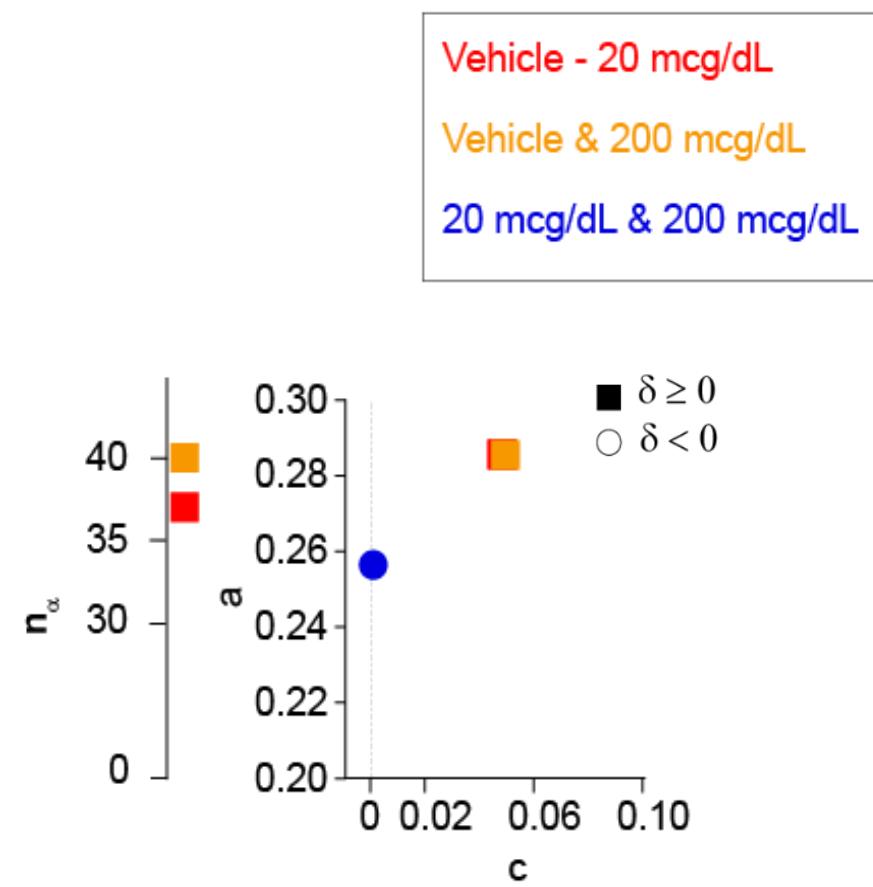
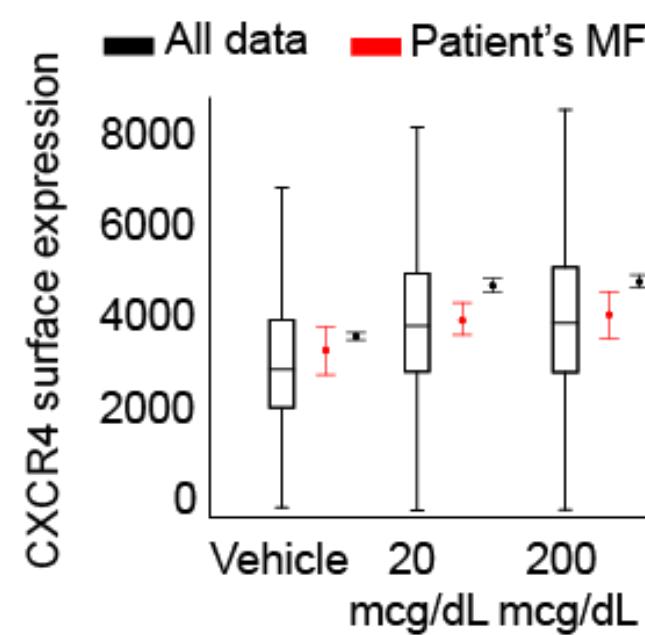
Use cases: II. Evaluation of drug effects

Task: Determine the transcriptional changes induced by the in vivo exposure of human eosinophils to glucocorticoids.

Flow cytometry data: >>>10K data points, so all p-values are much smaller than 0.05.



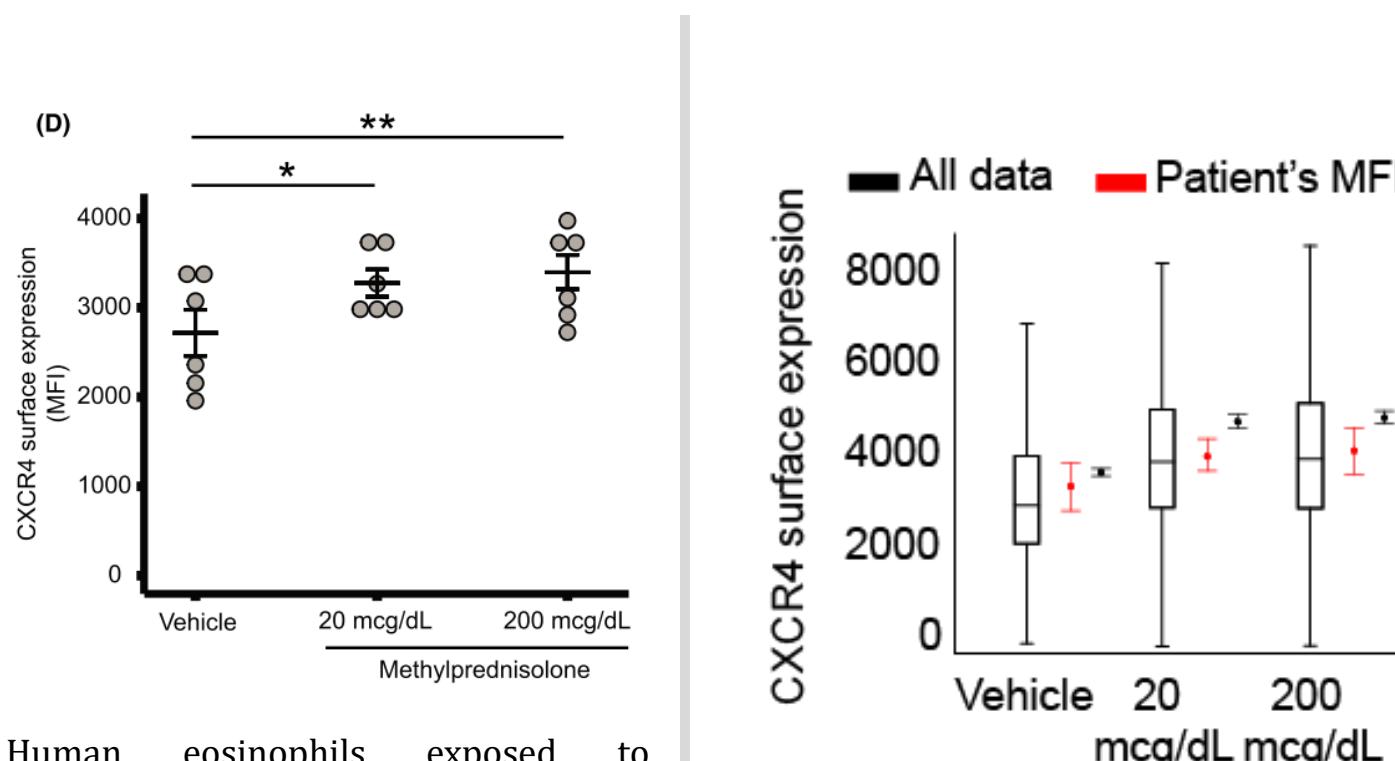
Human eosinophils exposed to Methylprednisolone. Flow cytometry data [1]. MFI: median of fluorescence intensity.



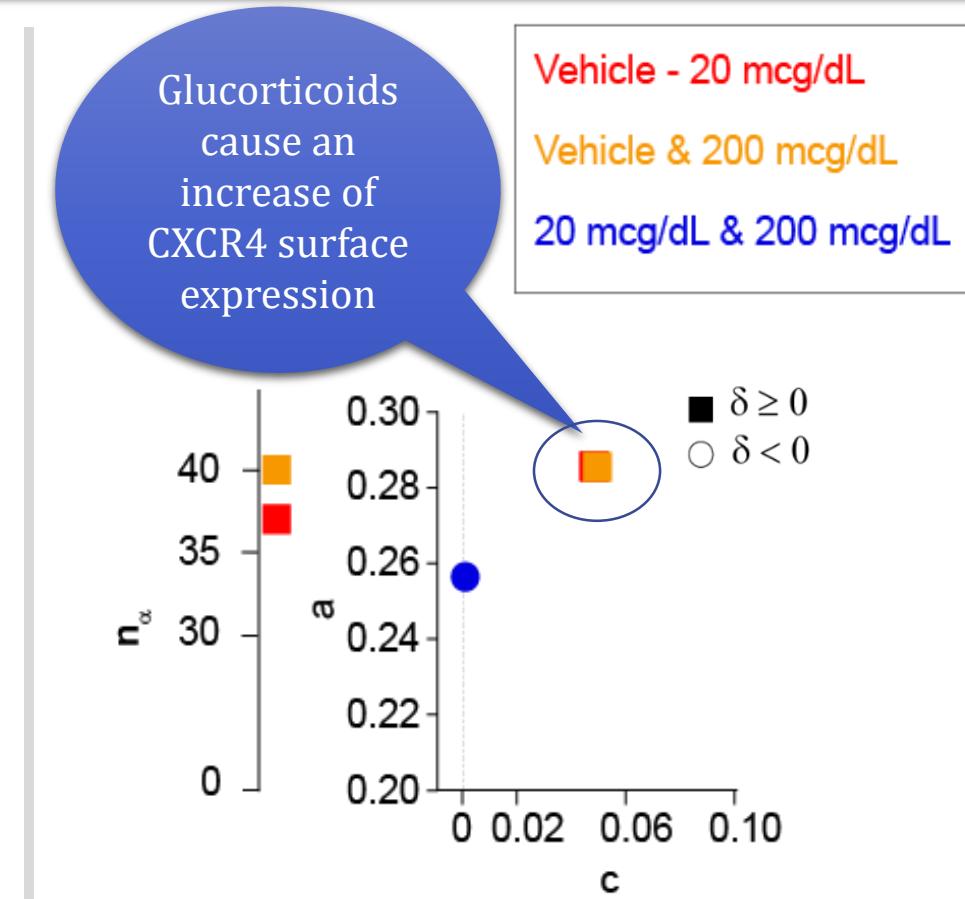
Use cases: II. Evaluation of drug effects

Task: Determine the transcriptional changes induced by the in vivo exposure of human eosinophils to glucocorticoids.

Flow cytometry data: >>>10K data points, so all p-values are much smaller than 0.05.



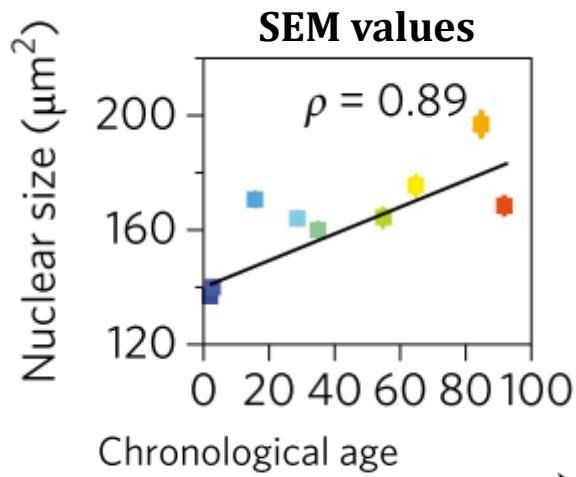
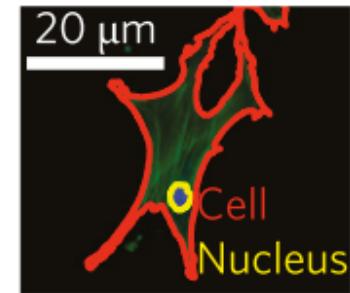
Human eosinophils exposed to Methylprednisolone. Flow cytometry data [1]. MFI: median of fluorescence intensity.



Use cases: III. Cellular aging effects in motility and morphology

Task: Determine which features characterize the age of healthy human cells.

Morphology



A02 A03 A16 A29 A35 A55 A65 A85 A92

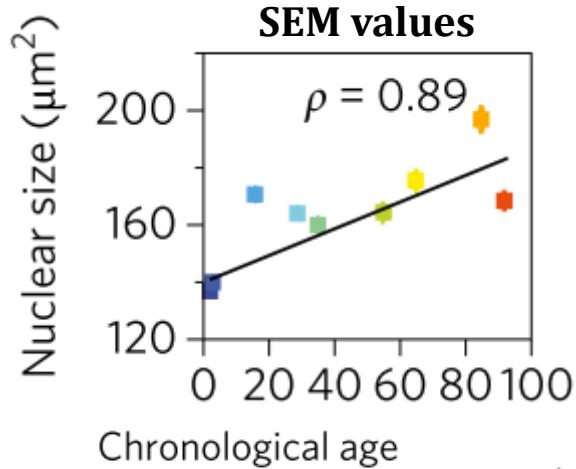
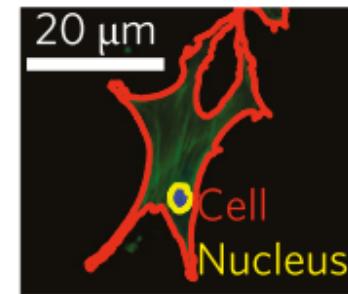
Human primary dermal fibroblasts extracted from individuals between 2 and 96 years old [1].

- The estimation of the p-value function determines which features contain enough information about the aging of the patients.

Use cases: III. Cellular aging effects in motility and morphology

Task: Determine which features characterize the age of healthy human cells.

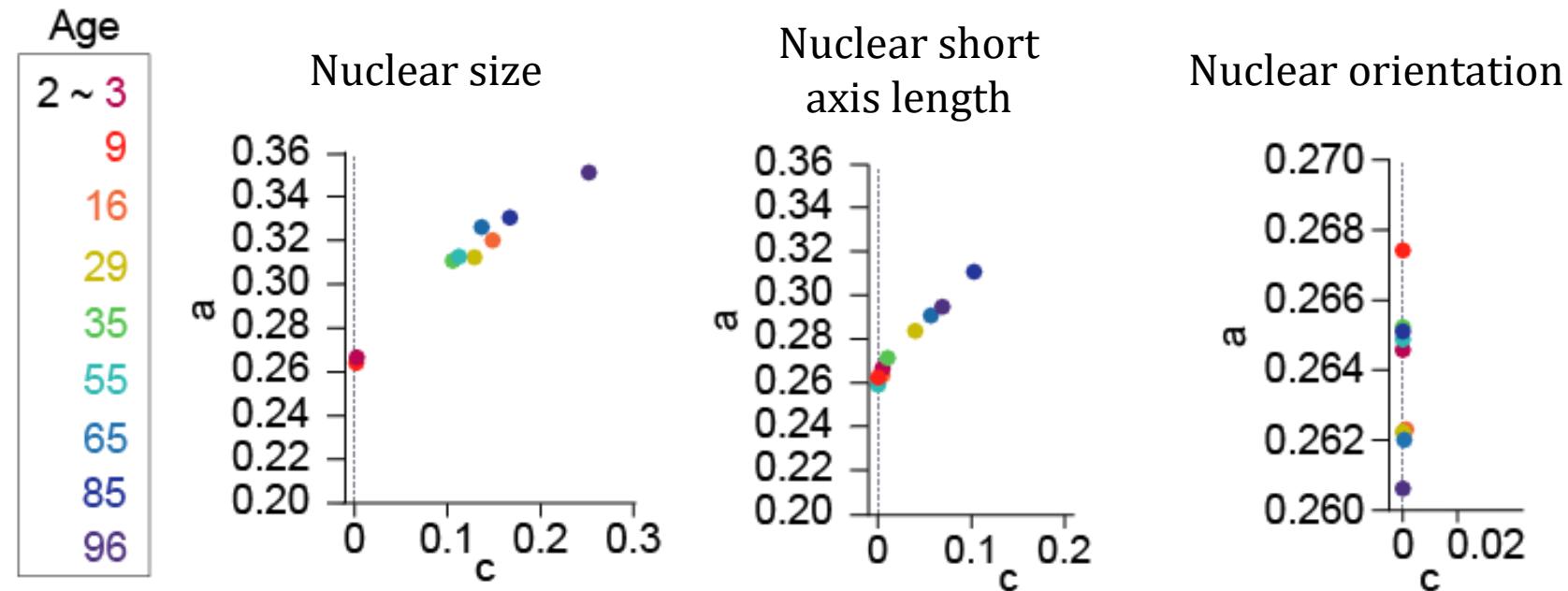
Morphology



A02A03A16 A29A35 A55A65A85A92

Human primary dermal fibroblasts extracted from individuals between 2 and 96 years old [1].

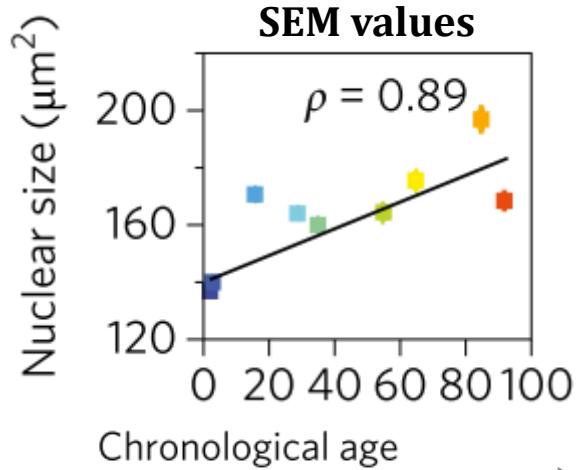
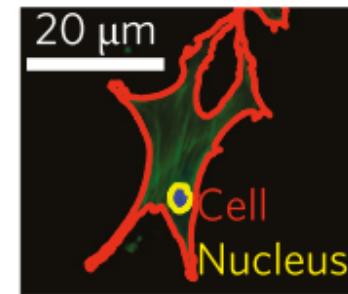
- The estimation of the p-value function determines which features contain enough information about the aging of the patients.



Use cases: III. Cellular aging effects in motility and morphology

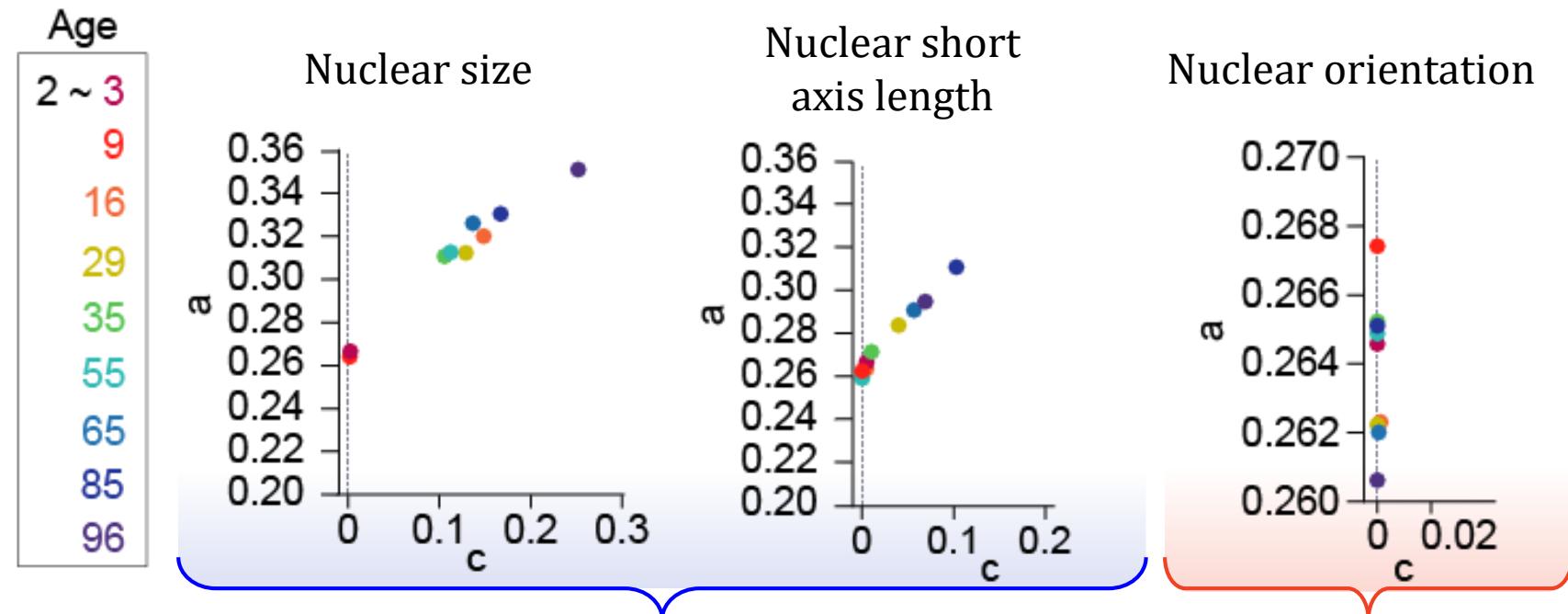
Task: Determine which features characterize the age of healthy human cells.

Morphology



A02A03A16 A29A35 A55A65A85A92
Human primary dermal fibroblasts extracted from individuals between 2 and 96 years old [1].

- The estimation of the p-value function determines which features contain enough information about the aging of the patients.



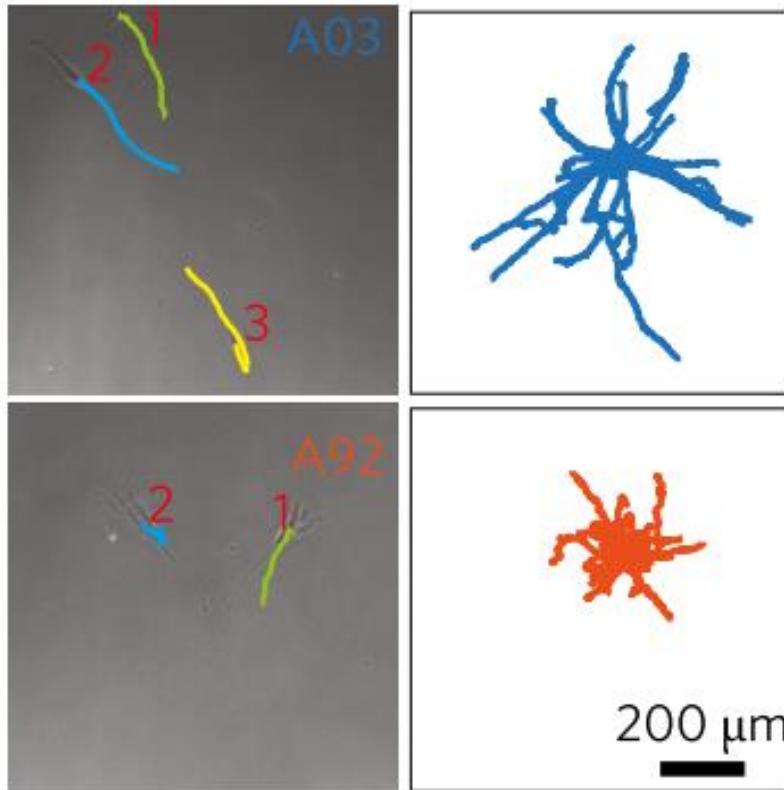
The parameters of the exponential function $a \cdot e^{-cn}$ increase proportionally to the age.
Valid features

The decay of $p(n)$ is zero always
Non-informative feature

Use cases: III. Cellular aging effects in motility and morphology

Task: Determine which features characterize the age of healthy human cells.

Motility



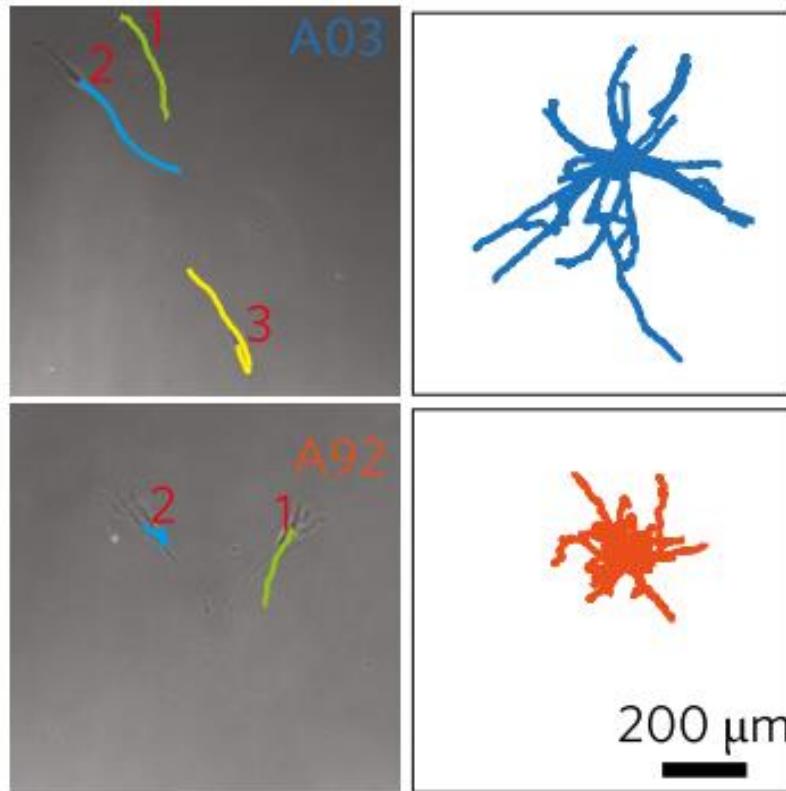
- The estimation of $p(n)$ allows one to decide whether it is valuable to collect new data.

Human primary dermal fibroblasts extracted from individuals between 2 and 96 years old [1].

Use cases: III. Cellular aging effects in motility and morphology

Task: Determine which features characterize the age of healthy human cells.

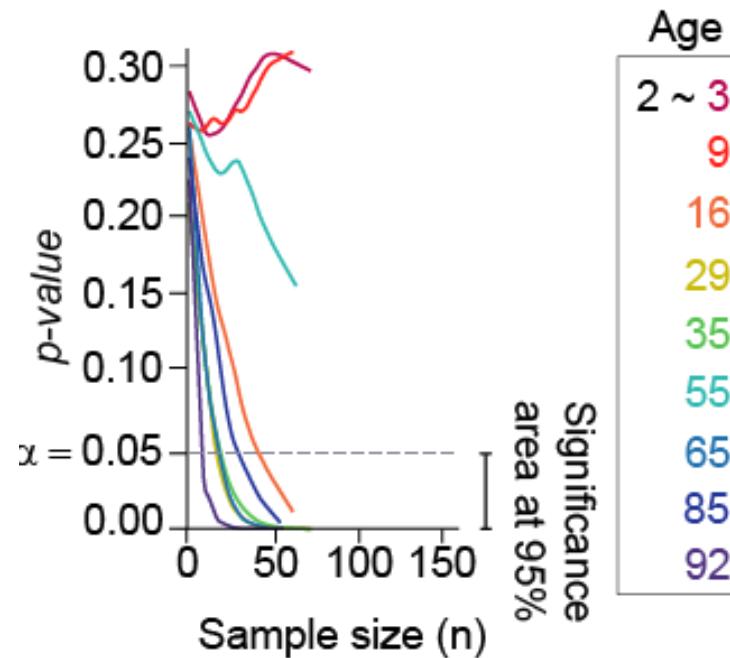
Motility



Human primary dermal fibroblasts extracted from individuals between 2 and 96 years old [1].

- The estimation of $p(n)$ allows one to decide whether it is valuable to collect new data.

Total diffusivity



Even if small, the data is enough to determine that the total diffusivity can characterize the cellular aging in humans.

Take home messages & Discussion

Take home messages & Discussion

$$p(n) \cong a \cdot e^{-cn} \text{ where } a, c \in \mathbb{R}^+$$

$$\delta = A\alpha - Ap$$

$\delta > 0 \rightarrow \text{Reject H0}$

- The *p-values* can be estimated with an exponential function:
 - Easy to integrate and evaluate.
 - Its values are completely controlled.
- Compared to the classical *p-value*, δ is mathematically constrained and stable.
- This new approach:
 - Enables good praxis in the same context as statistical hypothesis testing.
 - Provides evidence about the existence of statistical significance in preliminary analysis, with few data points.

Take home messages & Discussion

You can use this work or forget it ...

... but two things are true:

Take home messages & Discussion

You can use this work or forget it ...

... but two things are true:

- I. The *p-value* is a changeable measure that **depends on the sample size**.
- II. If you use single *p-values* as a MEASURE OF STATISTICAL SIGNIFICANCE, your conclusion will be **UNRELIABLE, UNREPEATABLE** and most probably, **NOT COMPLETELY TRUE.**

Try it yourself





<https://github.com/BIIG-UC3M/pMoSS>

BIIG-UC3M / pMoSS

Watch 3 Star 1 Fork 0

Code Issues 0 Pull requests 0 Actions Projects 0 Wiki Security Insights Settings

Software to model the p-value as a function of the sample size, and confront p-hacking

Edit Manage topics

3 commits 1 branch 0 packages 1 release 1 contributor BSD-3-Clause

Branch: master New pull request Create new file Upload files Find file Clone or download

esgomez Update README.md
data Upload the code
examples Upload the code
images Upload the code
pMoSS Remove double files
LICENSE Upload the code
README.md Update README.md
requirements.txt Upload the code

Clone with HTTPS Use SSH
Use Git or checkout with SVN using the web URL.
<https://github.com/BIIG-UC3M/pMoSS.git>

Open in Desktop Download ZIP

README.md

pMoSS: *p-value* Model using the Sample Size

Python>= 3.6 License BSD 3-Clause-Clear

pMoSS (*p-value* Model using the Sample Size) is a Python code to model the *p-value* as an n-dependent function using Monte Carlo cross-validation.

How to use it?

1.- Google Colab.

2.- Install Python and use Jupyter notebooks.

All the information is in the GitHub web page.



<https://github.com/BIIG-UC3M/pMoSS>

<https://www.youtube.com/watch?v=sT1eHp8ZVUk>

I. Use pMoSS in Google Colab

<https://github.com/BIIG-UC3M/pMoSS>



<https://github.com/BIIG-UC3M/pMoSS>

<https://www.youtube.com/watch?v=2iNj96r0N7U>

II. Analyze new data with pMoSS

<https://github.com/BIIG-UC3M/pMoSS>

Arrate
Muñoz-Barrutia



uc3m | Universidad Carlos III de Madrid

Estibaliz
Gómez-de-Mariscal



Hasini Jayatilaka



Jude M. Philip



Alexandra Sneider



Denis Wirtz



Use of the *p-value* as a size-dependent function: model and applications

Estibaliz Gómez de Mariscal

esgomez@pa.uc3m.es

Confronting *p*-hacking: addressing *p*-value dependence on sample size

E. Gómez-de-Mariscal, A. Sneider, H. Jayatilaka, J. M. Phillip, D. Wirtz, A. Muñoz-Barrutia
bioRxiv 2019; doi: <https://doi.org/10.1101/2019.12.17.878405>

Acknowledgments: This work was produced with the support of the Spanish Ministry of Economy and Competitiveness (TEC2015-73064-EXP, TEC2016-78052-RTC2017-6600-1, RTC-2017-6600-1), a 2017 Leonardo Grant for Researchers and Cultural Creators, BBVA Foundation, and grants from the US National Institutes of Health (UO1AG060903 and U54CA143868). We also want to acknowledge the support of NVIDIA Corporation with the donation of the Titan X (Pascal) GPU used for this research. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1746891.

Fundación **BBVA**



Instituto de Investigación Sanitaria Gregorio Marañón



uc3m | Universidad Carlos III de Madrid



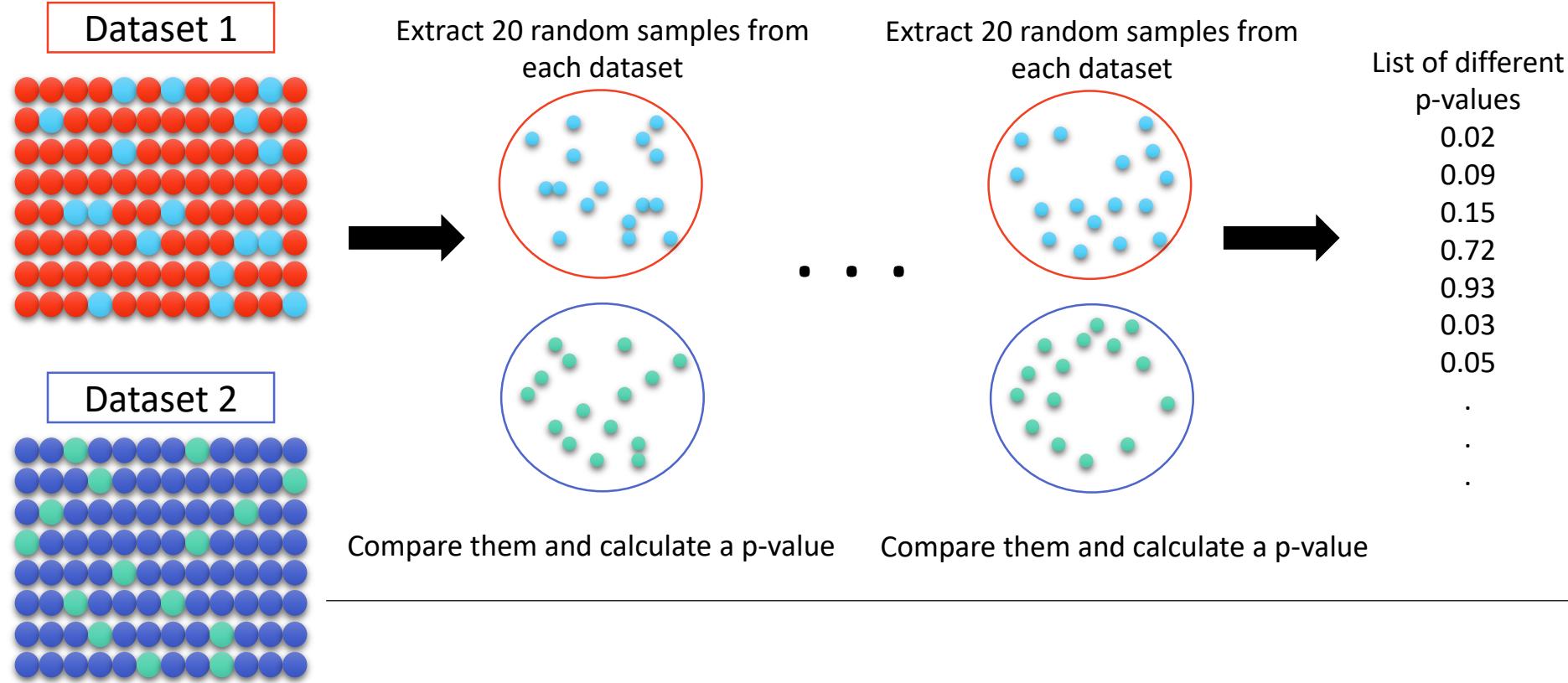
Stanford Children's Health

Bass Center
for Childhood Cancer
and Blood Diseases

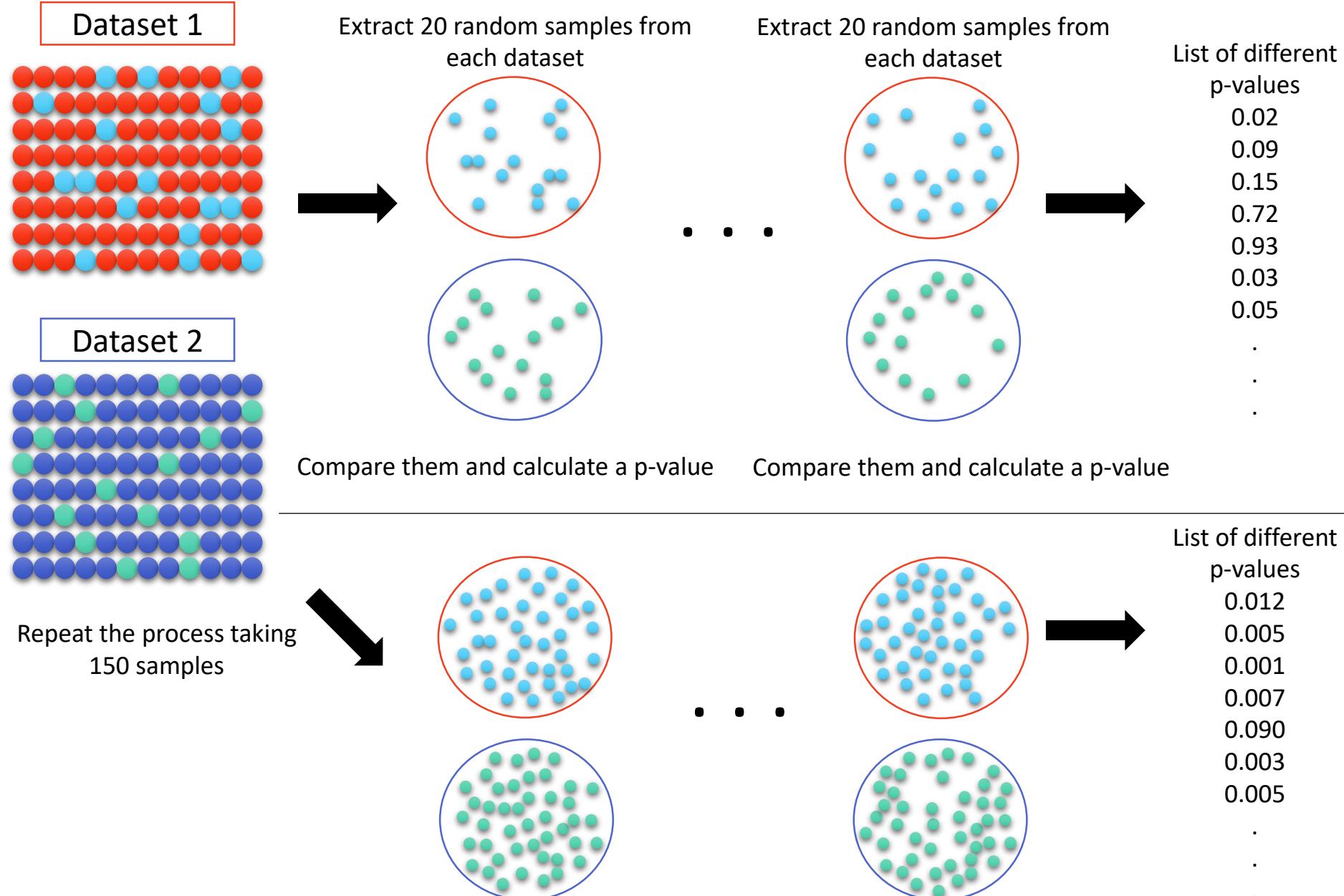

JOHNS HOPKINS
UNIVERSITY



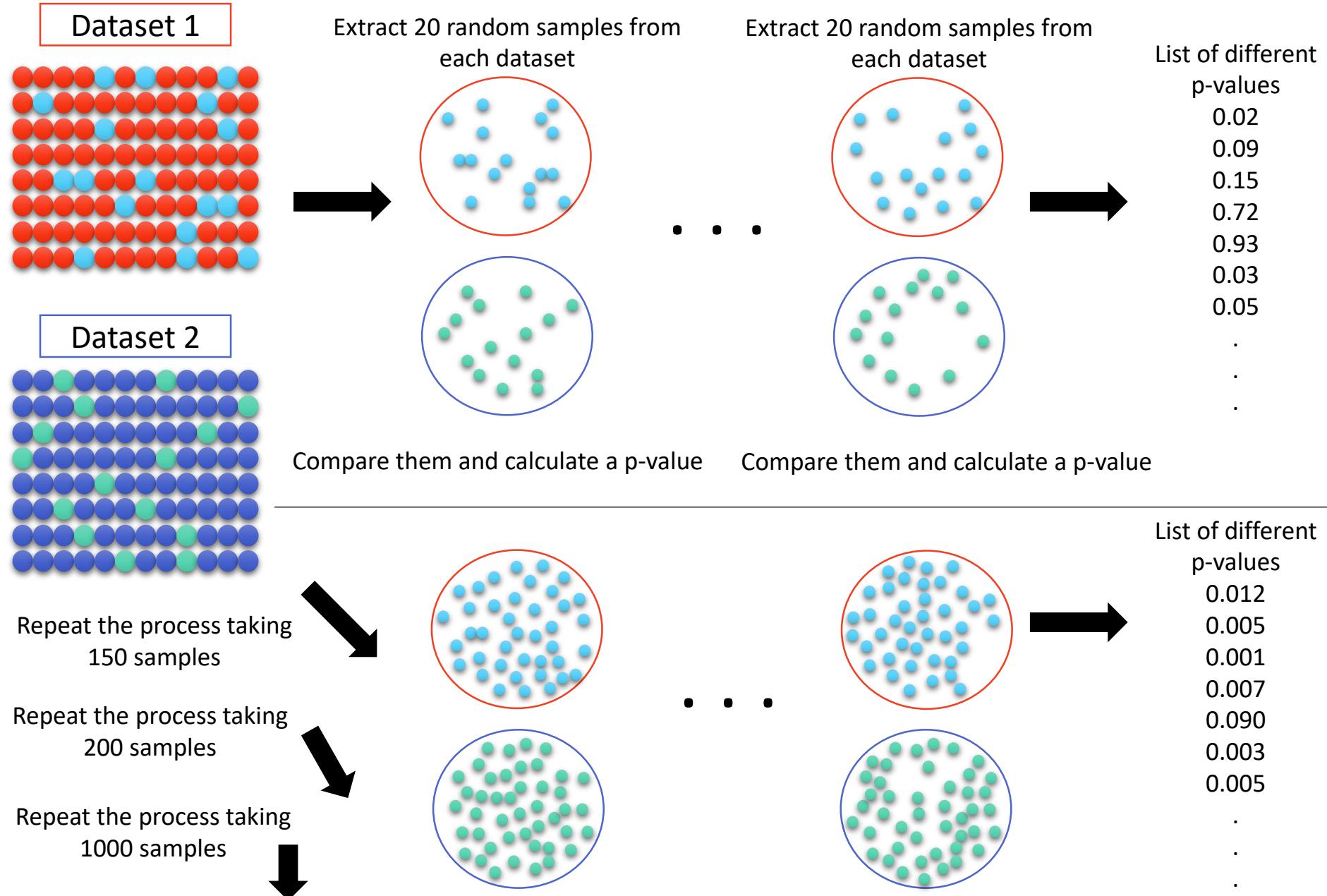
To obtain unbiased empirical measures we use cross validation



To obtain unbiased empirical measures we use cross validation



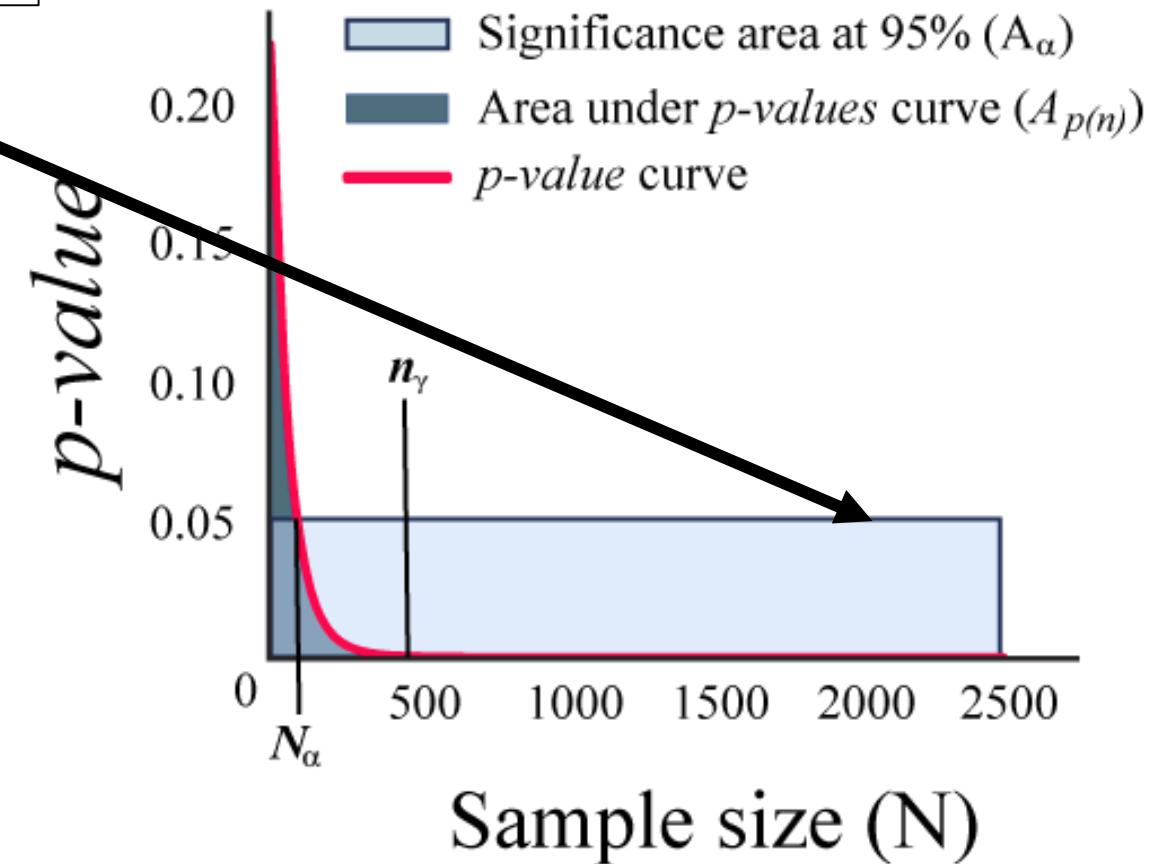
To obtain unbiased empirical measures we use cross validation



$$p(n) \cong a \cdot e^{-cn} \text{ where } a, c \in \mathbb{R}^+$$

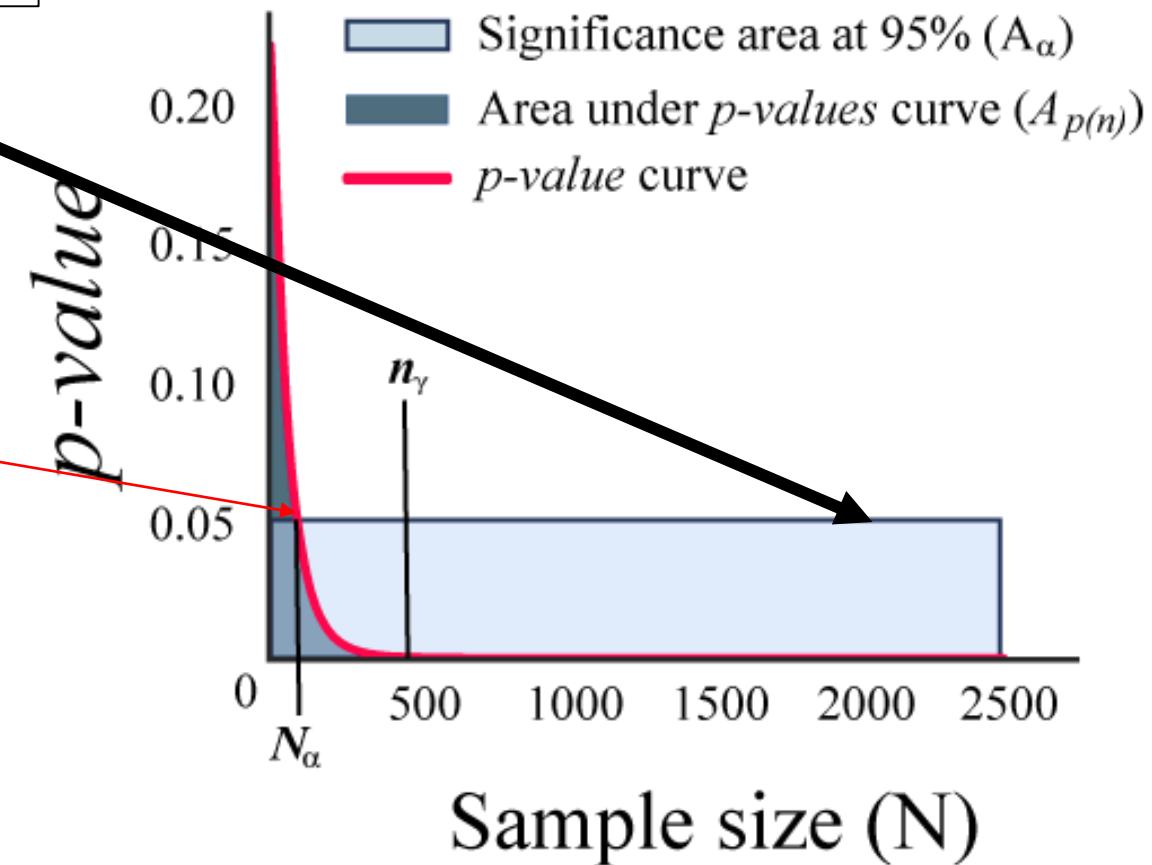
- The obtaining of $p(n)$ let us studying its behavior.
- When there are many observations of a variable, we can estimate the values of $p(n)$ using cross validation.
- Then, an exponential curve can be fitted very easily.
- The exponential function has some nice features:
 - Easy to integrate and evaluate.
 - Its values are completely controlled.

Whenever the p-value is significant,
it will always be smaller than 0.05



Whenever the p-value is significant, it will always be smaller than 0.05

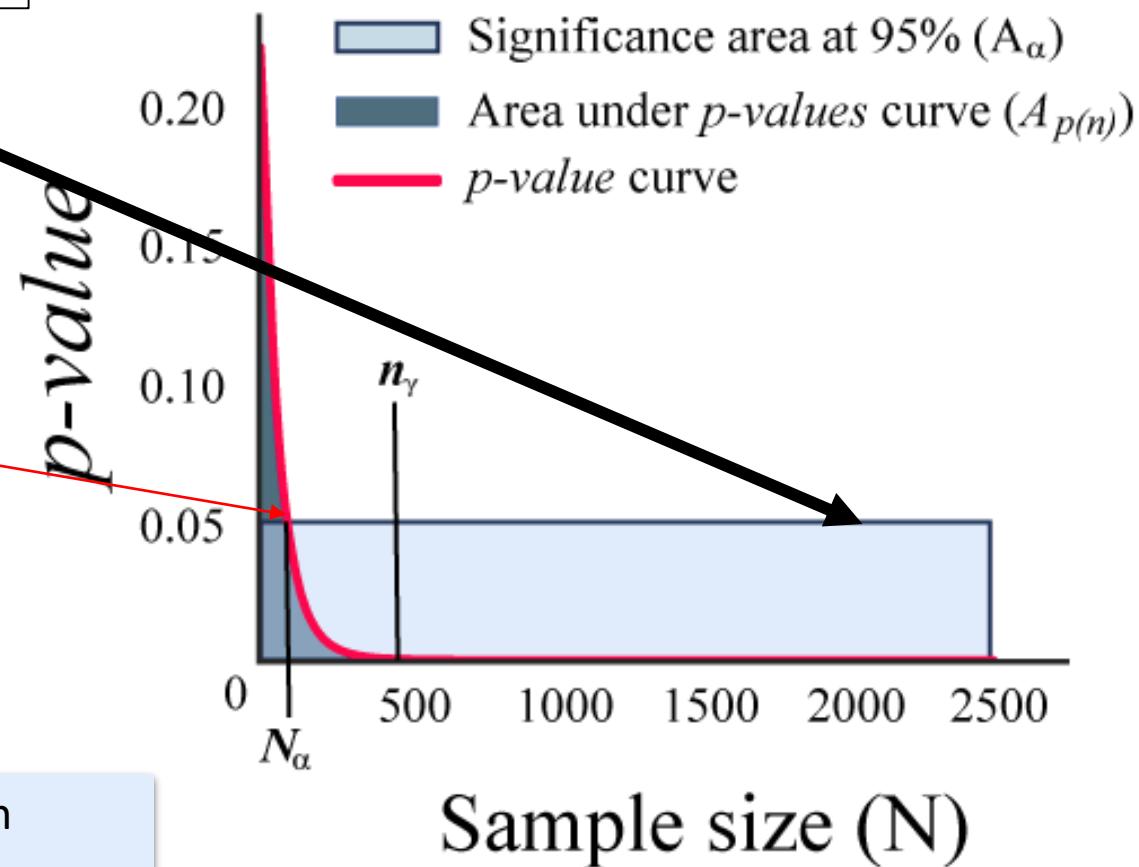
It is possible to fit a function to the p-value



Whenever the p-value is significant, it will always be smaller than 0.05

It is possible to fit a function to the p-value

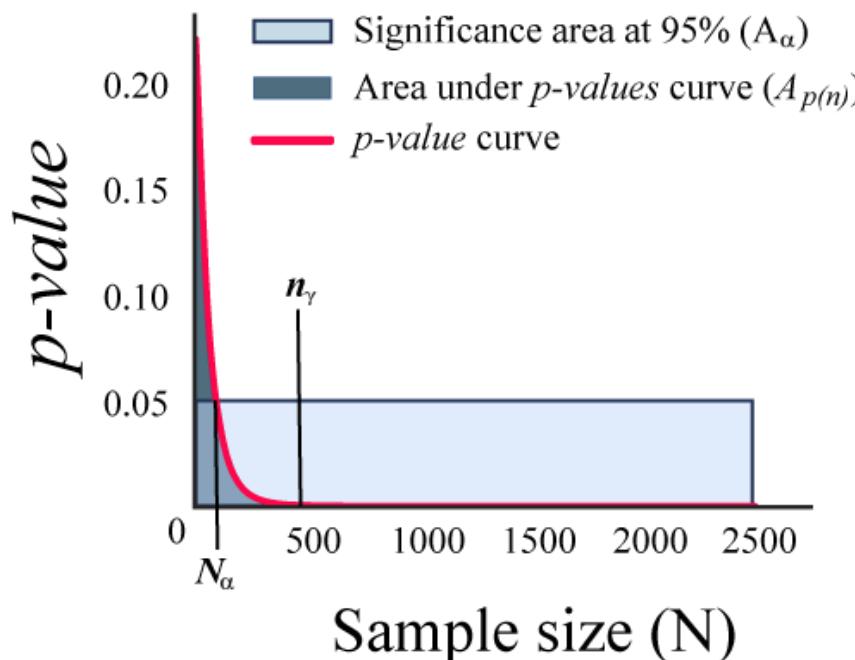
$$\delta_\alpha(n) = \int (\alpha - p(n)) dn = A_{\alpha(n)} - A_{p(n)}$$



Compare the calculated function of the p-value with 0.05:

Compare the calculated function of the p-value with 0.05:

$$\delta_\alpha(n) = \int (\alpha - p(n)) dn = A_{\alpha(n)} - A_{p(n)}$$



If we compare $A_{\alpha(n)}$ and $A_{p(n)}$ when n is too big, then the area $A_{\alpha(n)}$ will always be much larger.

→ $p(n)$ converges to zero, so n_γ indicates this point.

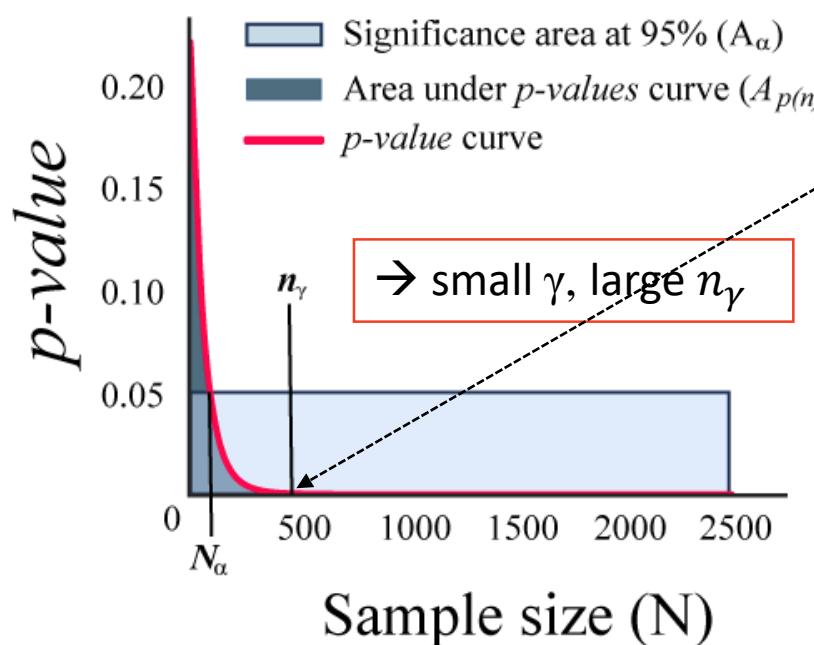
→ We need the threshold γ to decide when $p(n)$ has really converged.

$$\theta_{\alpha,\gamma} = \begin{cases} 1, & \delta_{\alpha,y} \geq 0 \\ 0, & otherwise \end{cases}$$

Method: The only parameter to think about is γ . How this affects the study?

Find a nice n_γ to study the decrease of $p(n)$.

How to use γ



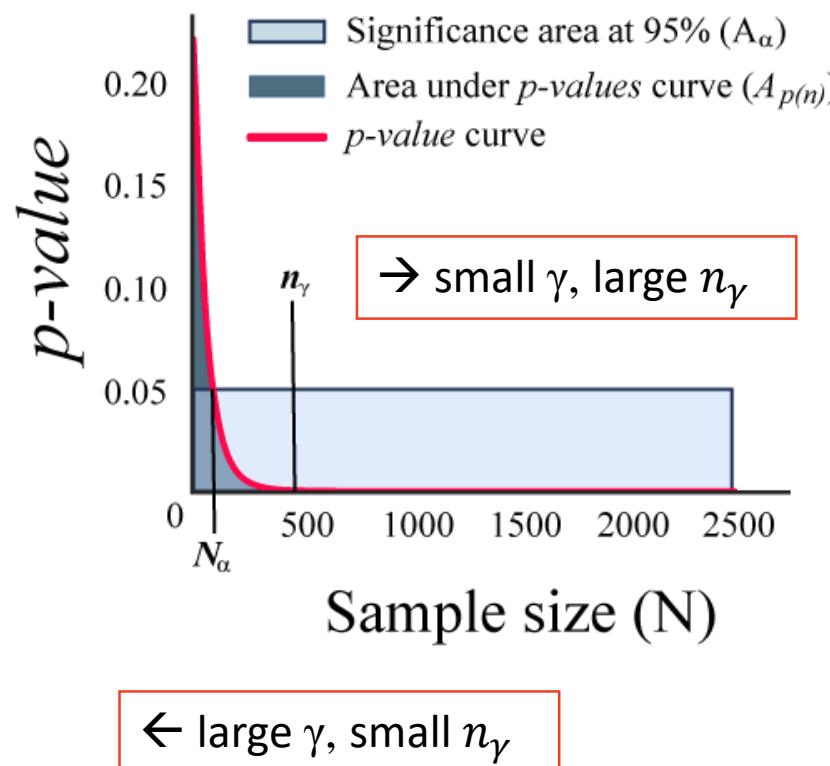
$$n_\gamma = \operatorname{argmin}_n \left\{ \left| \frac{\partial p(n)}{\partial n} \right| < \gamma \right\}$$

$$\delta_{\alpha,\gamma} = A_{\alpha_\gamma} - A_{p(n_\gamma)}$$

$$\theta_{\alpha,\gamma} = \begin{cases} 1, & \delta_{\alpha,\gamma} \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

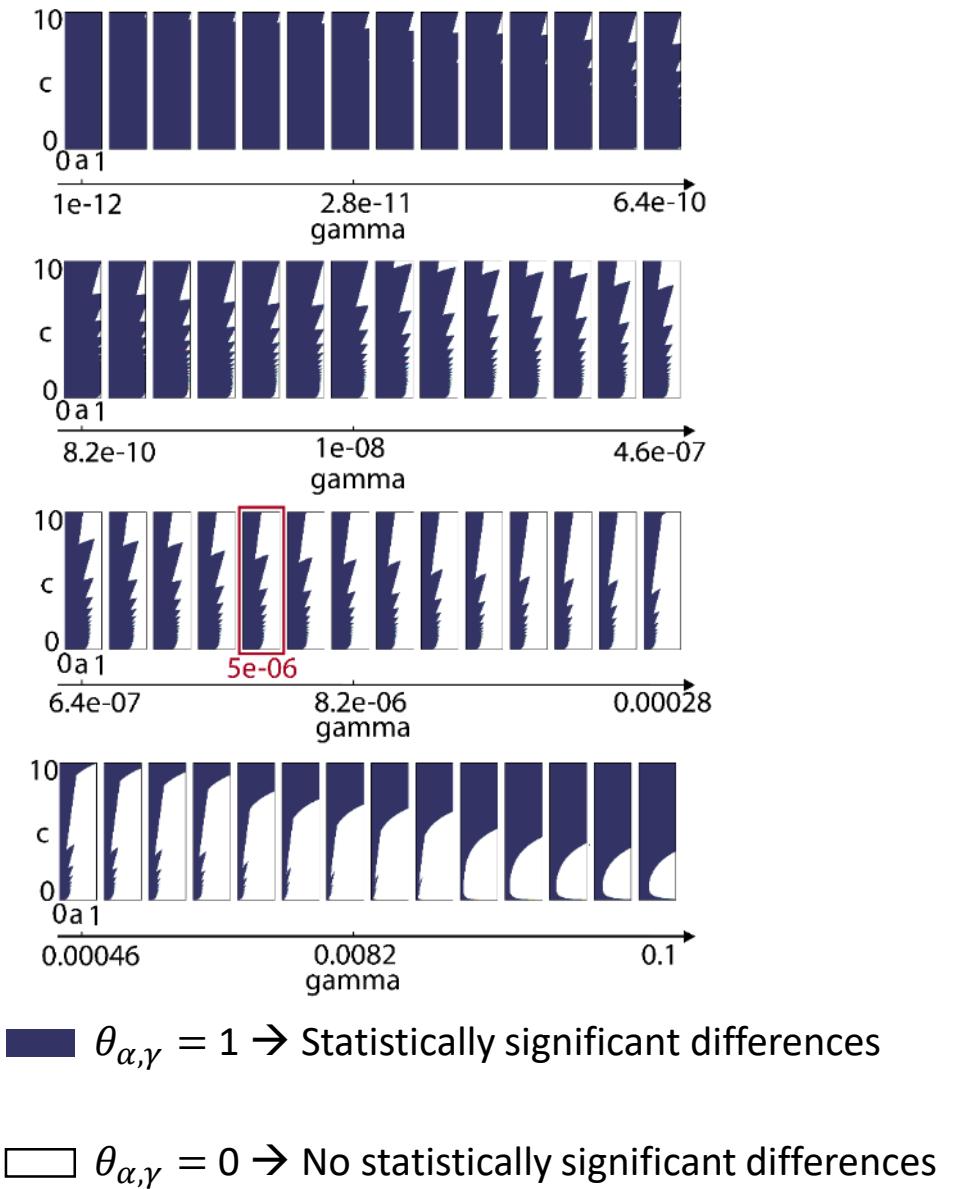
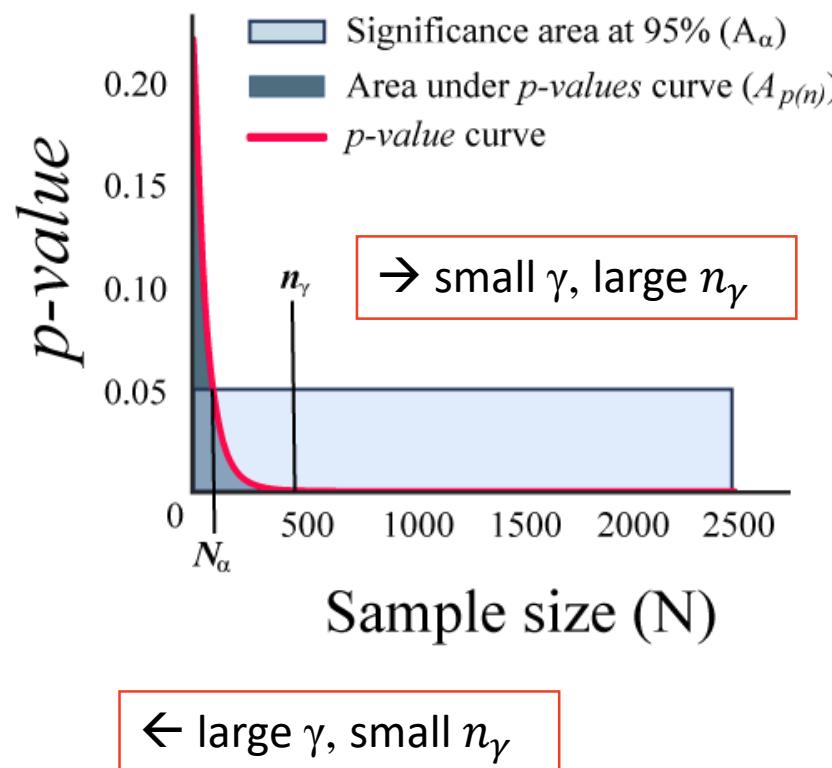
Method: The only parameter to think about is γ . How this affects the study?

If we are too restrictive with γ , we will
always say that there are never
differences.



Method: The only parameter to think about is γ . How this affects the study?

If we are too restrictive with γ , we will always say that there are never differences.



1.- Using normal distributions, which are the results according to γ ?

γ	$2.5 \cdot 10^{-6}$		$5 \cdot 10^{-6}$		$5 \cdot 10^{-5}$		$5 \cdot 10^{-4}$		
Comparison	$\Theta_{\alpha,\gamma}$	$\delta_{\alpha,\gamma}$	$\Theta_{\alpha,\gamma}$	$\delta_{\alpha,\gamma}$	$\Theta_{\alpha,\gamma}$	$\delta_{\alpha,\gamma}$	$\Theta_{\alpha,\gamma}$	$\delta_{\alpha,\gamma}$	n_α
$\mathcal{N}(0,1) - \mathcal{N}(0,1)$	0	-490.25	0	-490.25	0	-490.25	0	-490.25	0
$\mathcal{N}(0,1) - \mathcal{N}(0.01,1)$	0	-3576	0	-488.25	0	-488.25	0	-488.25	0
$\mathcal{N}(0,1) - \mathcal{N}(0.1,1)$	1	0.782	0	-21.922	0	-79.938	0	-42.375	1237
$\mathcal{N}(0,1) - \mathcal{N}(0.25,1)$	1	9.523	1	5.848	0	-5.840	0	-12.945	186
$\mathcal{N}(0,1) - \mathcal{N}(0.5,1)$	1	0	1	0	1	0	1	0	45
$\mathcal{N}(0,1) - \mathcal{N}(0.75,1)$	1	0	1	0	1	0	1	0	22
$\mathcal{N}(0,1) - \mathcal{N}(1,1)$	1	0	1	0	1	0	1	0	0
$\mathcal{N}(0,1) - \mathcal{N}(1.5,1)$	1	0	1	0	1	0	1	0	0
$\mathcal{N}(0,1) - \mathcal{N}(2,1)$	1	0	1	0	1	0	1	0	0
$\mathcal{N}(0,1) - \mathcal{N}(2.5,1)$	1	0	1	0	1	0	1	0	0
$\mathcal{N}(0,1) - \mathcal{N}(3,1)$	1	0	1	0	1	0	1	0	0

Table S1. Table of decision index $\Theta_{\alpha,\gamma}$, difference $\delta_{\alpha,\gamma} = A_{\alpha\gamma} - A_{p(n_\gamma)}$ and n_α for $\alpha = 0.05$ and $\gamma = 2.5 \cdot 10^{-6}, 5 \cdot 10^{-6}, 5 \cdot 10^{-5}, 5 \cdot 10^{-4}$.

2.- The method is calculated with a computer and randomness is intrinsic (because of cross validation):

- Decide how many iterations to do.
- Test whether it affects to the final result and how.

→ All computations resulted in the same result as in Table S1 ✓