

Multiple testing correction

Laurent Modolo

April 20 2017

Table of Contents

1 Hypothesis testing

2 Multiple hypotheses testing

Table of Contents

1 Hypothesis testing

2 Multiple hypotheses testing

Hypothesis testing

Rejection of a null hypothesis H_0

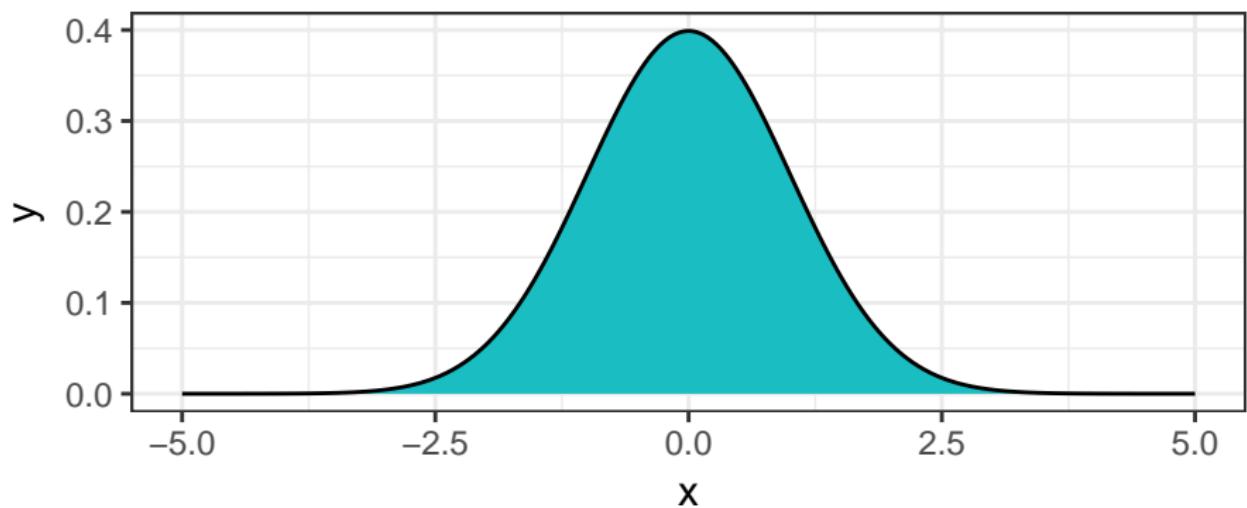
Given the null model of our data how likely are we to observe a value? We can compute this likelihood from the probability distribution of the null hypothesis.

We reject the hypothesis at risk α , the probability that the null hypothesis was true for the observed value.

p-value

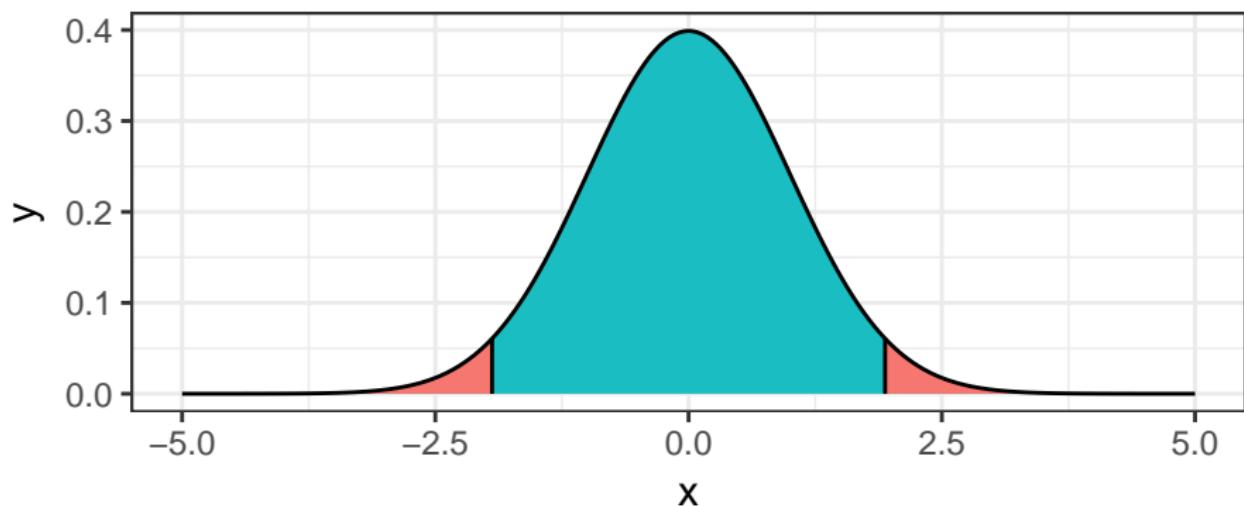
The *p*-value is the probability to observe a value as or more extreme under the null hypothesis model.

Hypothesis testing



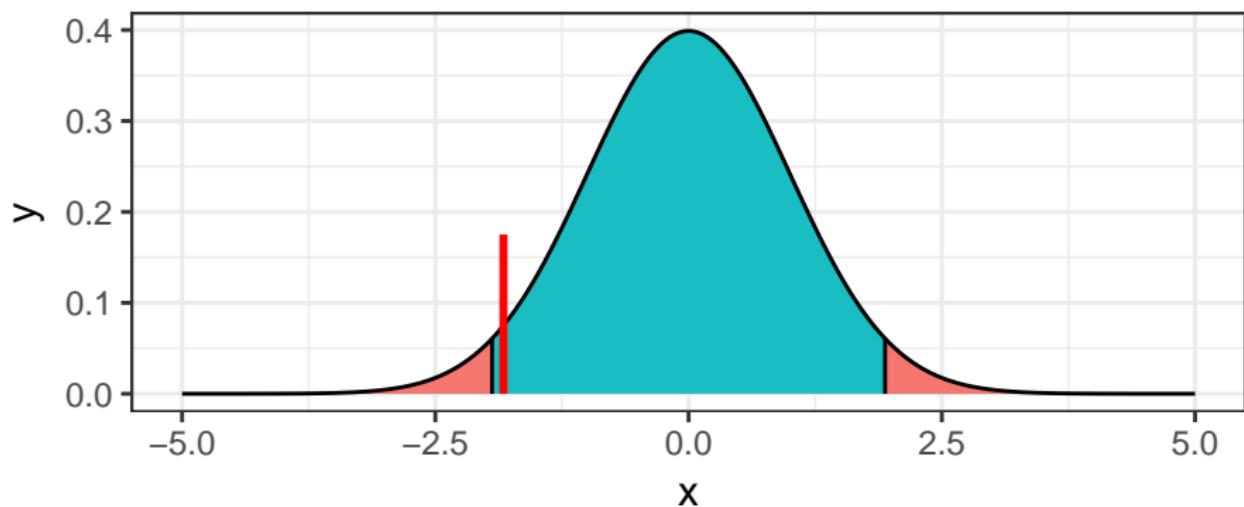
- distribution under the null hypothesis H_0

Hypothesis testing



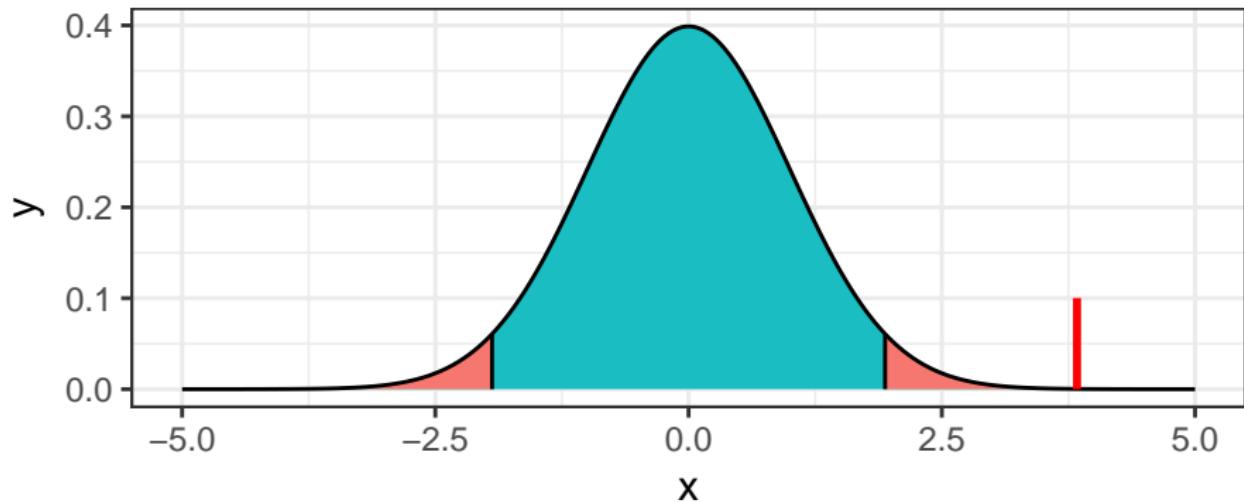
- distribution under the null hypothesis H_0
- rejection zone at a risk α

Hypothesis testing



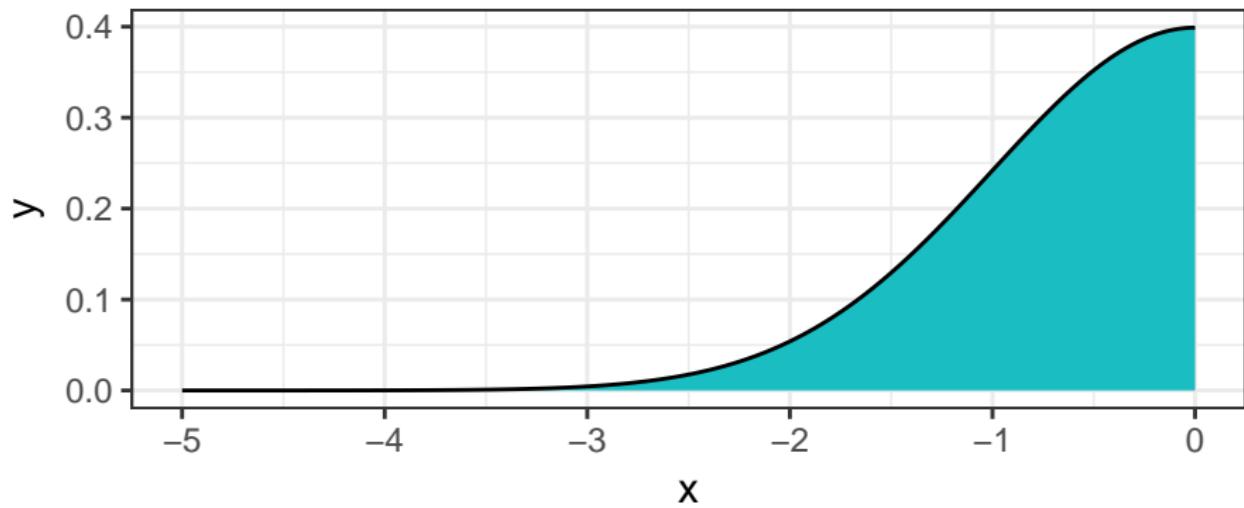
- distribution under the null hypothesis H_0
- rejection zone at a risk α

Hypothesis testing



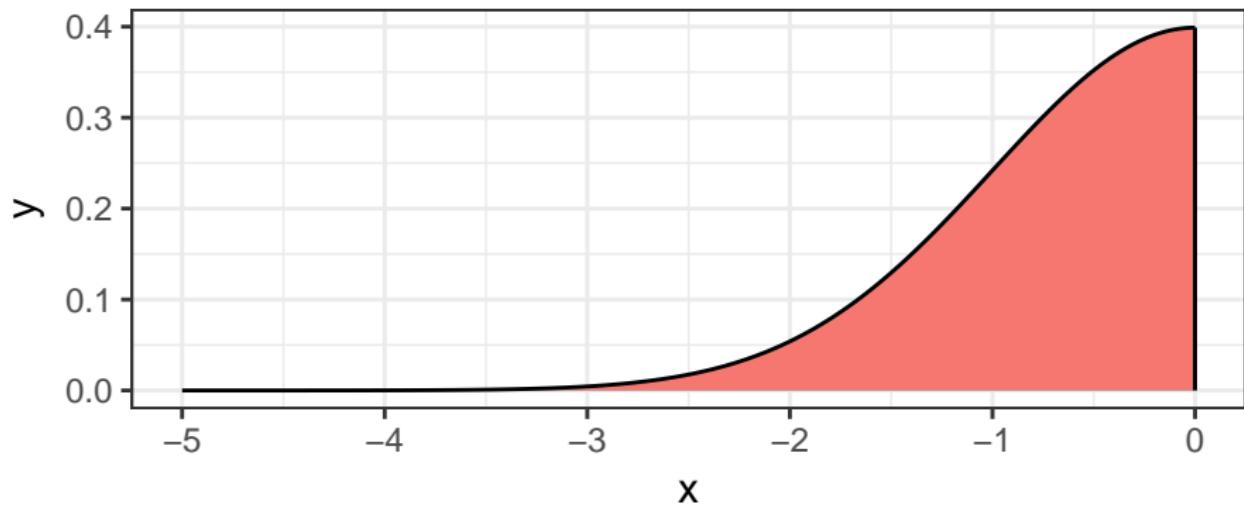
- distribution under the null hypothesis H_0
- rejection zone at a risk α

p-value construction



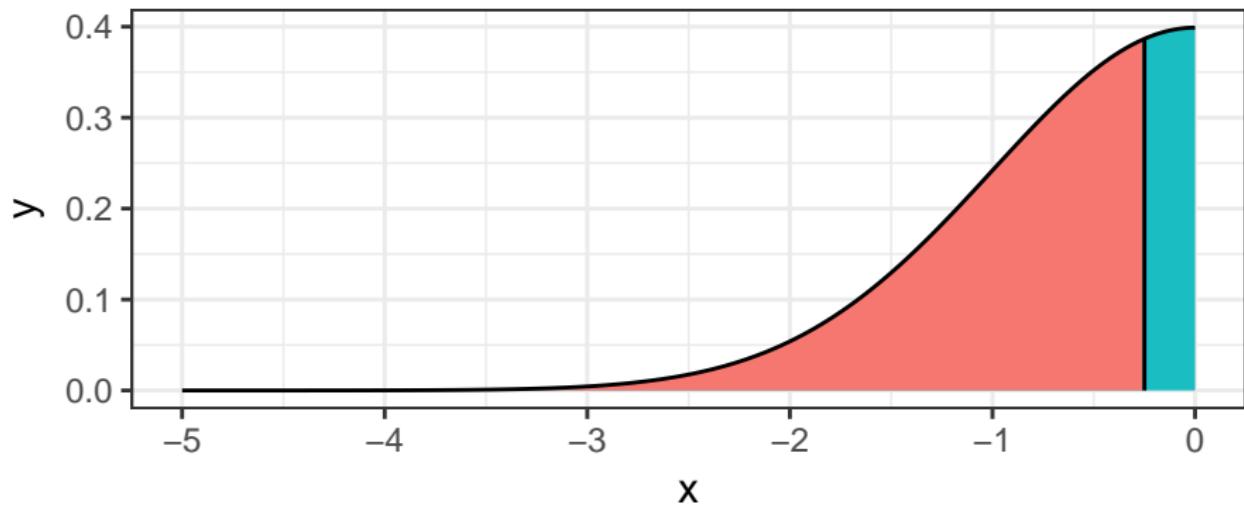
probability to observe a value as or more extreme under the null hypothesis model

p-value construction



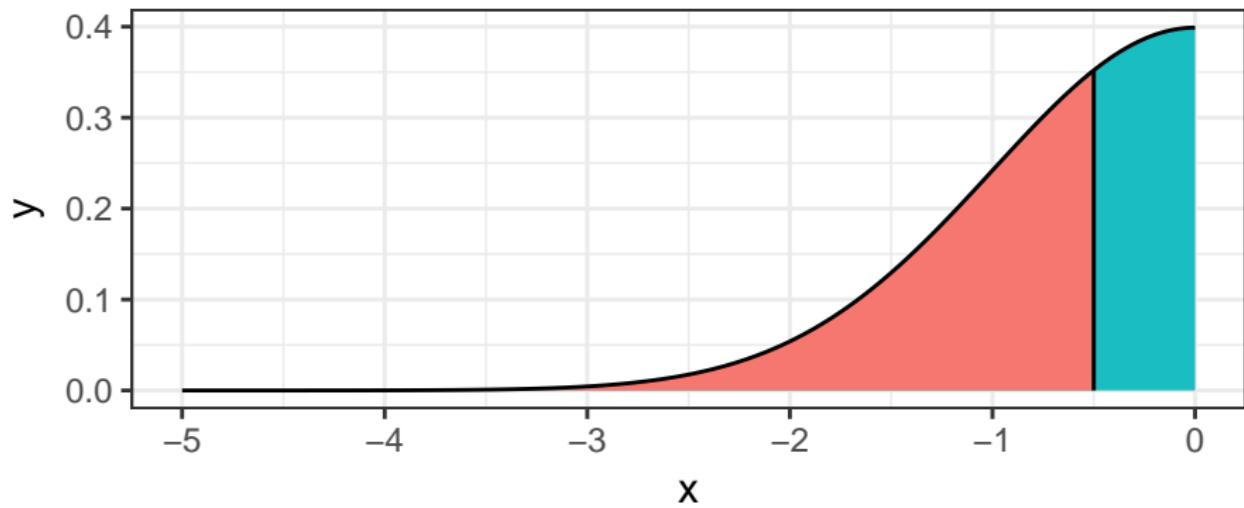
probability to observe a value as or more extreme under the null hypothesis model

p-value construction



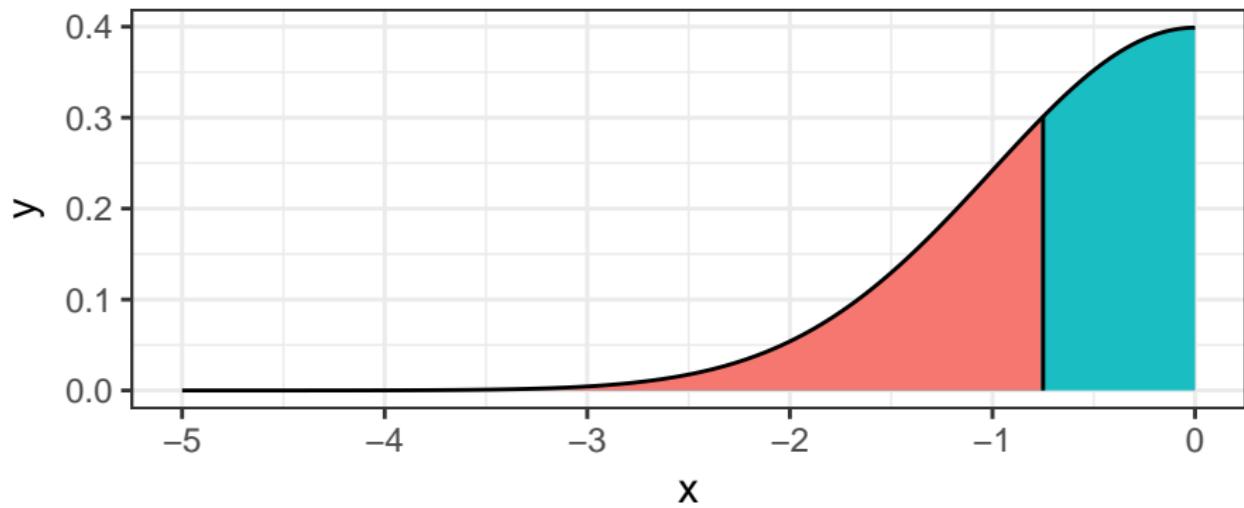
probability to observe a value as or more extreme under the null hypothesis model

p-value construction



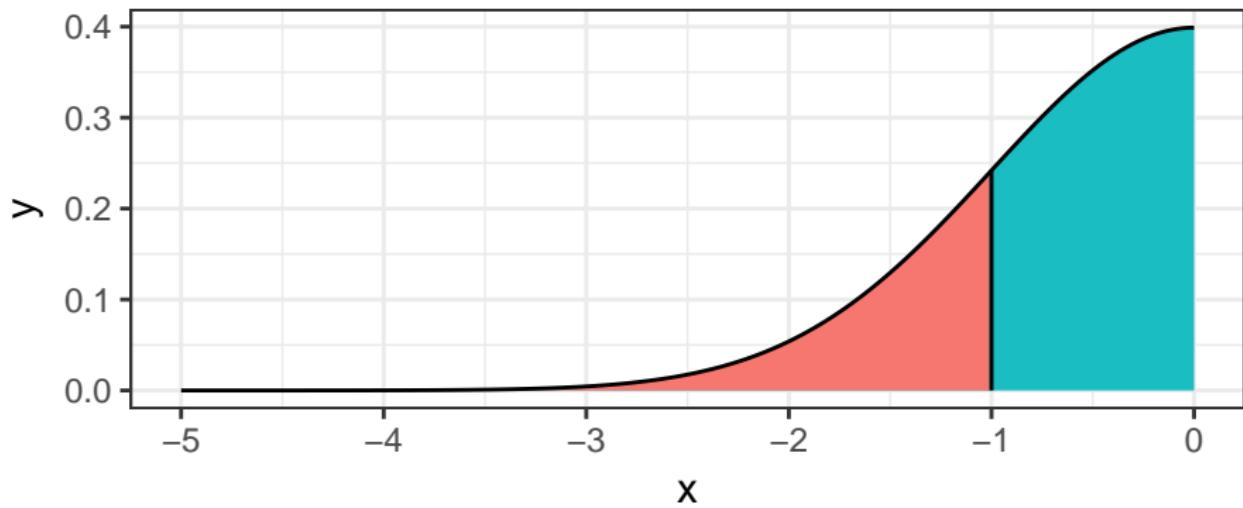
probability to observe a value as or more extreme under the null hypothesis model

p-value construction



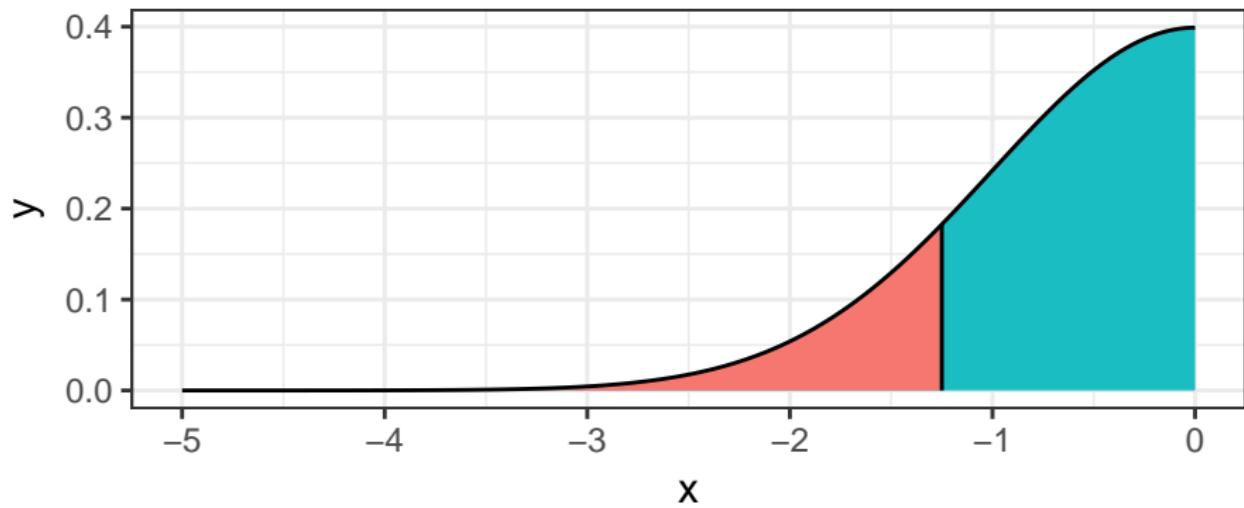
probability to observe a value as or more extreme under the null hypothesis model

p-value construction



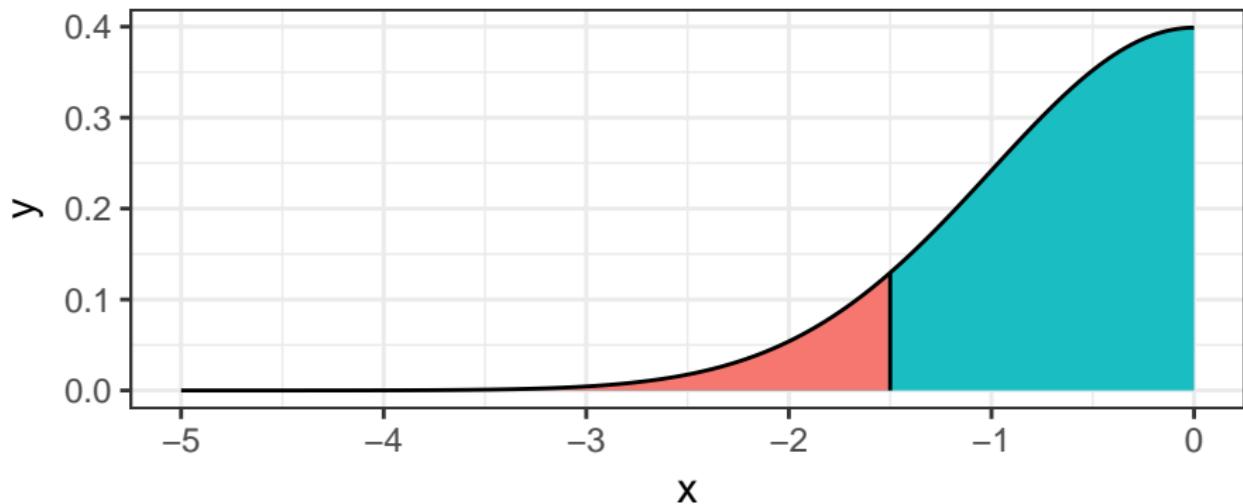
probability to observe a value as or more extreme under the null hypothesis model

p-value construction



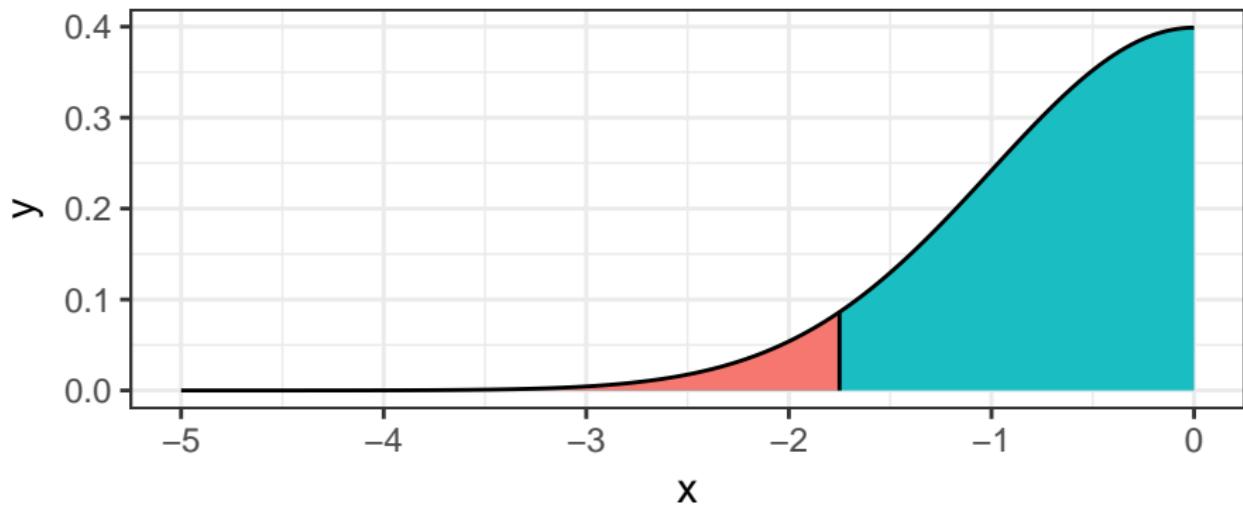
probability to observe a value as or more extreme under the null hypothesis model

p-value construction



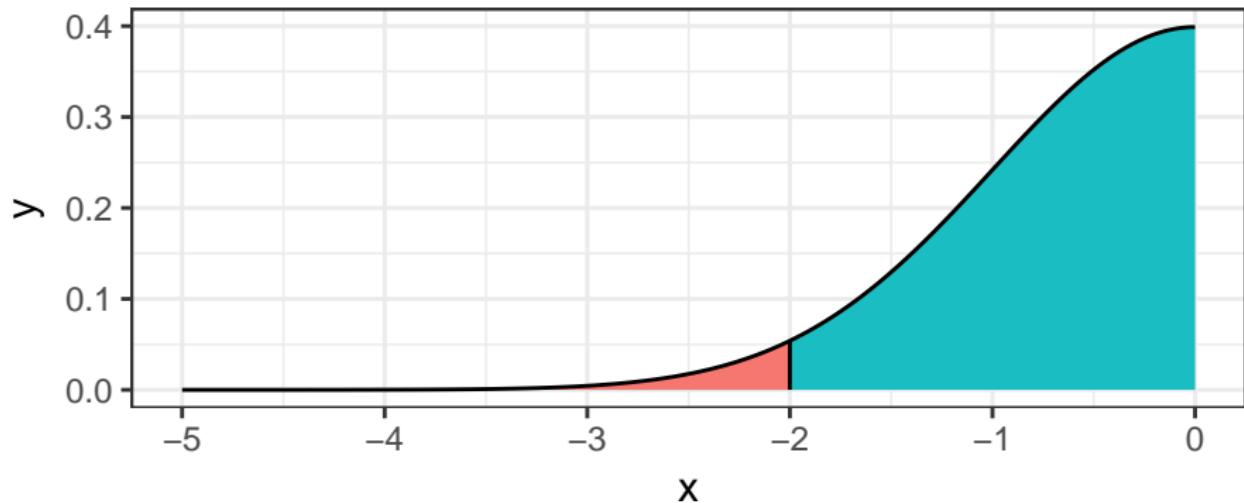
probability to observe a value as or more extreme under the null hypothesis model

p-value construction



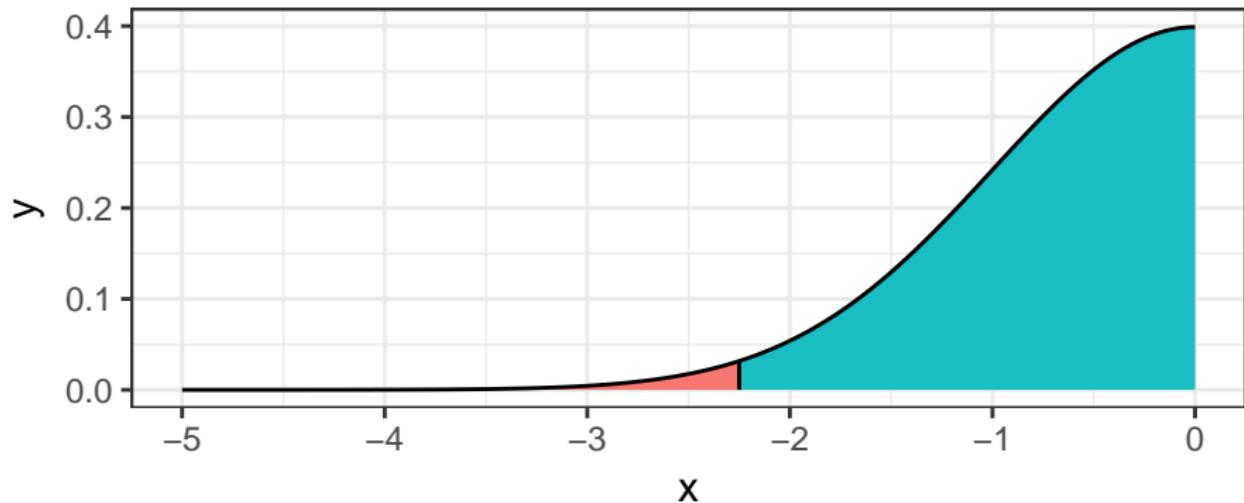
probability to observe a value as or more extreme under the null hypothesis model

p-value construction



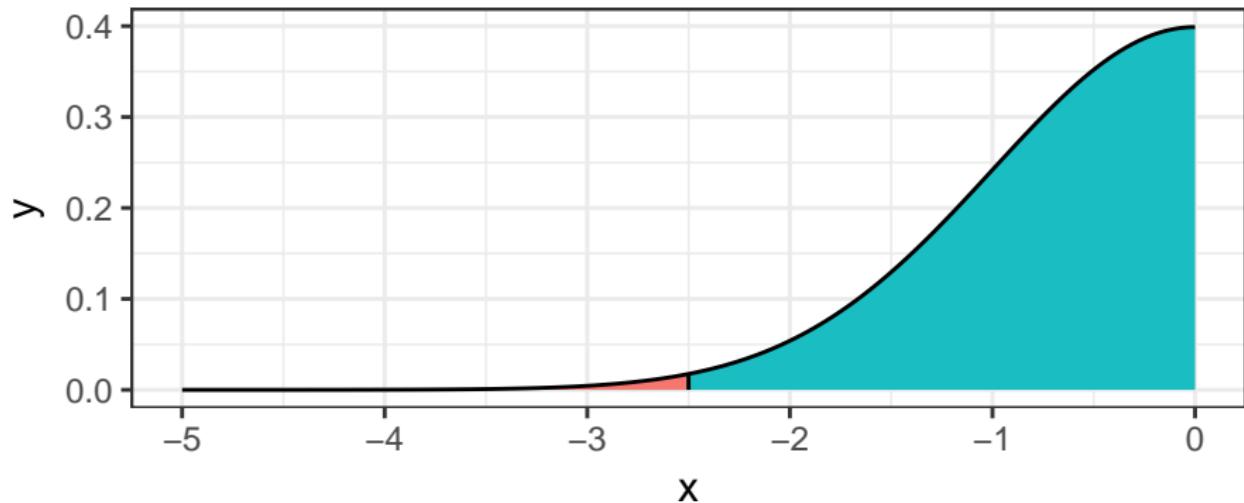
probability to observe a value as or more extreme under the null hypothesis model

p-value construction



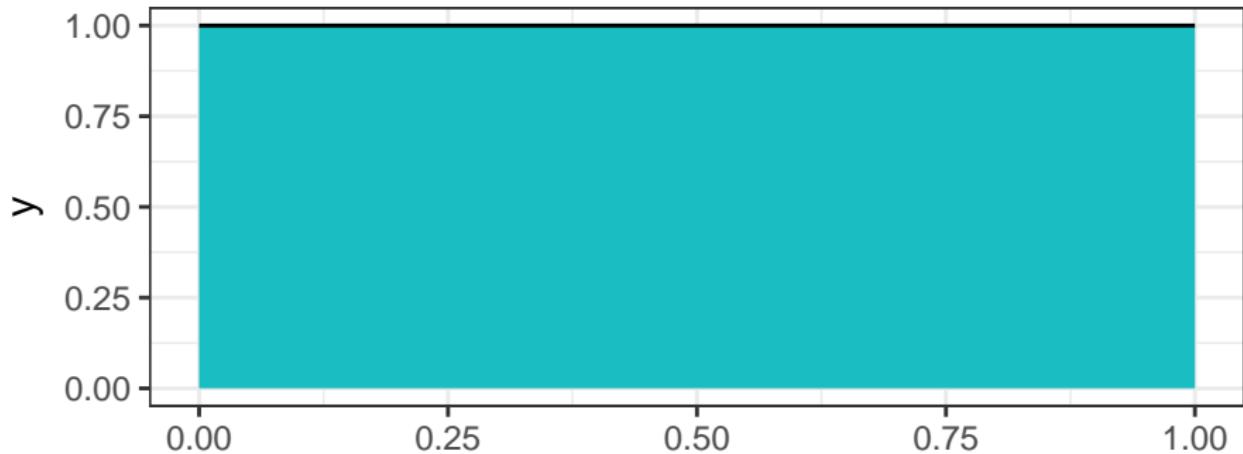
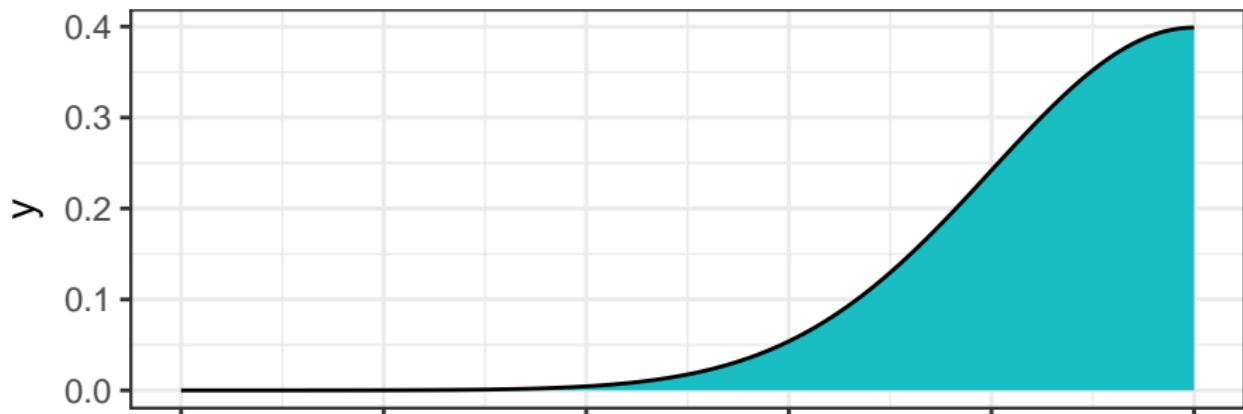
probability to observe a value as or more extreme under the null hypothesis model

p-value construction

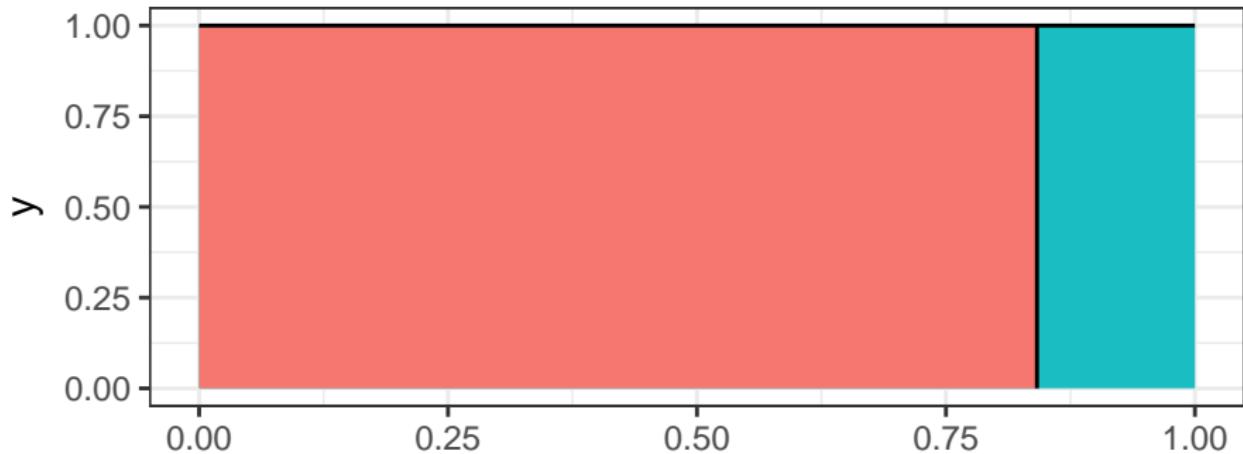
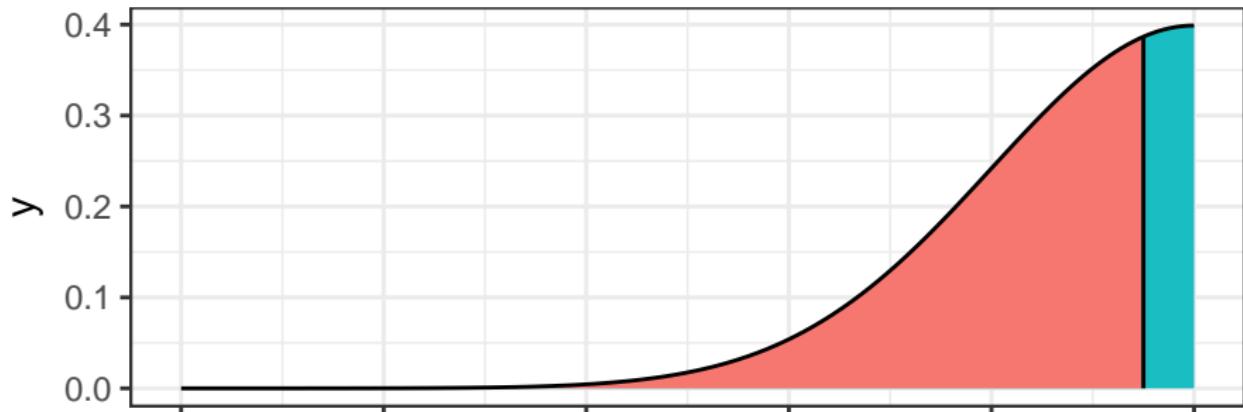


probability to observe a value as or more extreme under the null hypothesis model

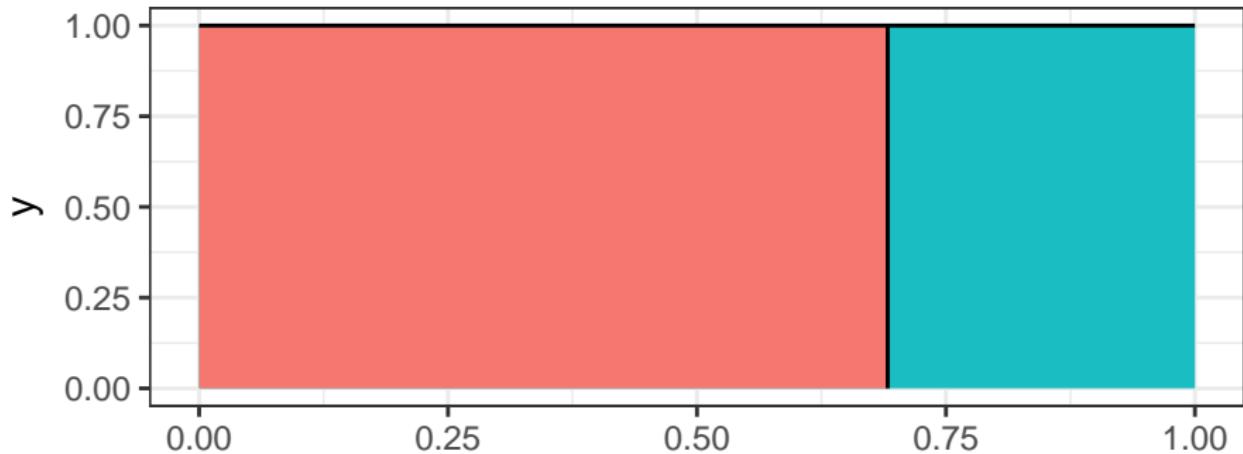
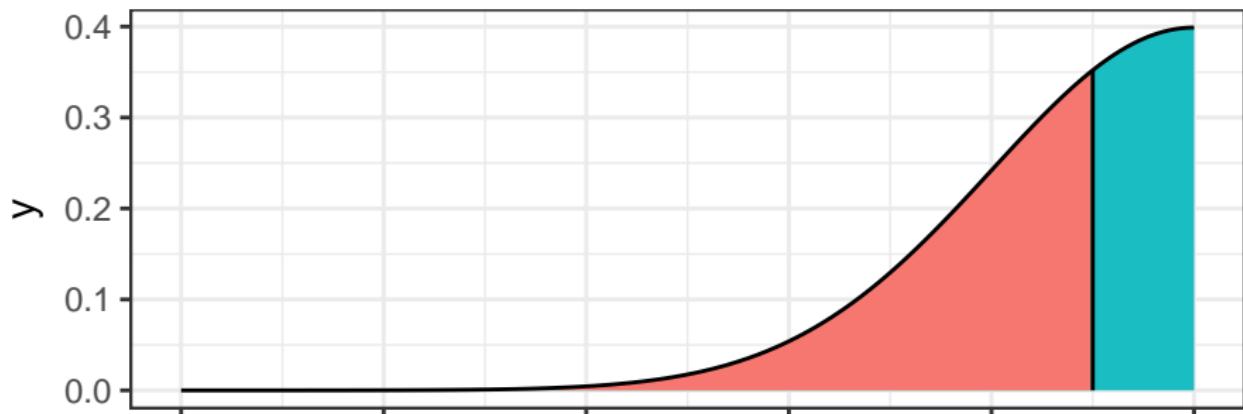
p-value construction



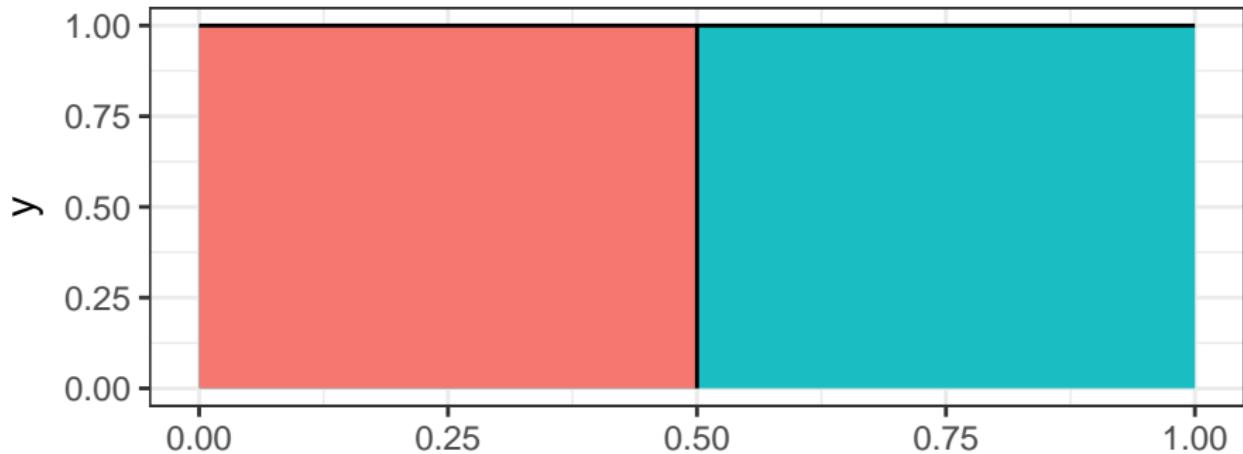
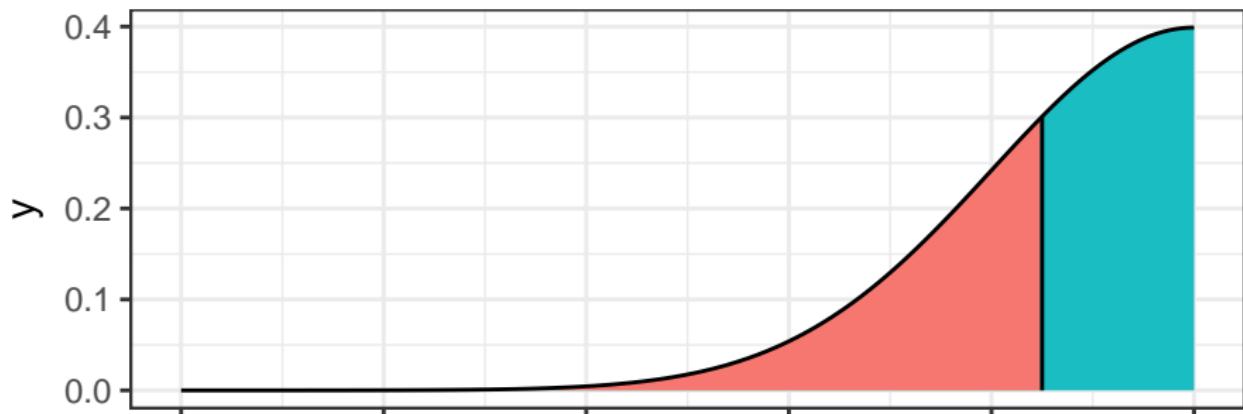
p-value construction



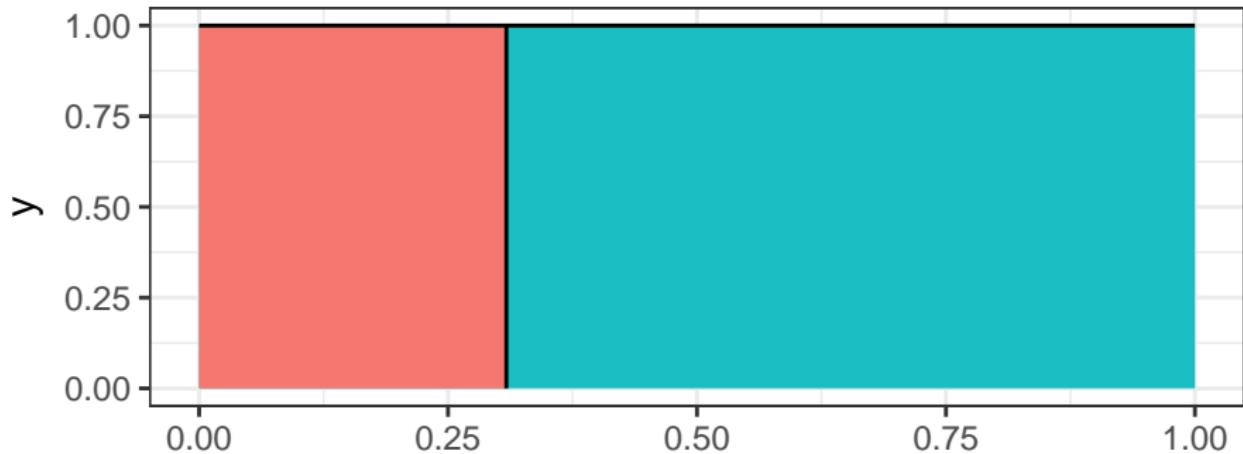
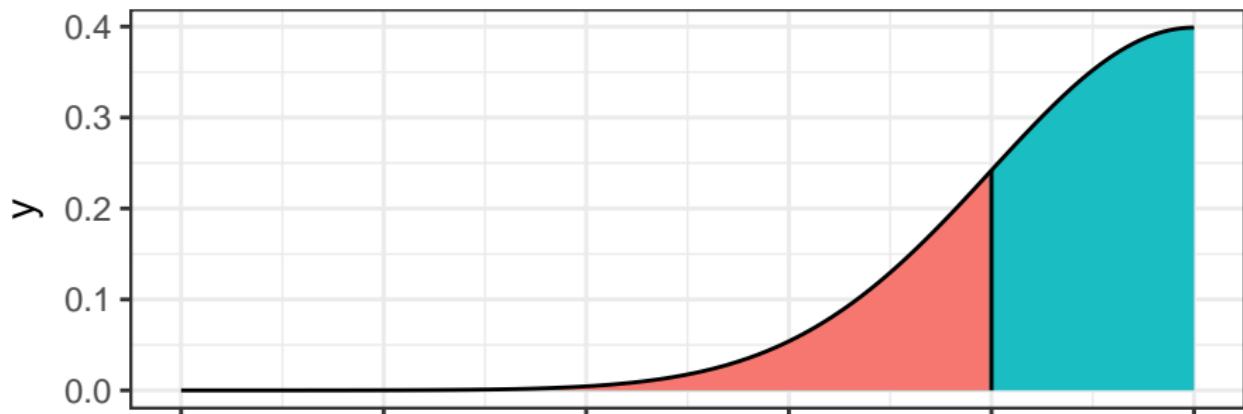
p-value construction



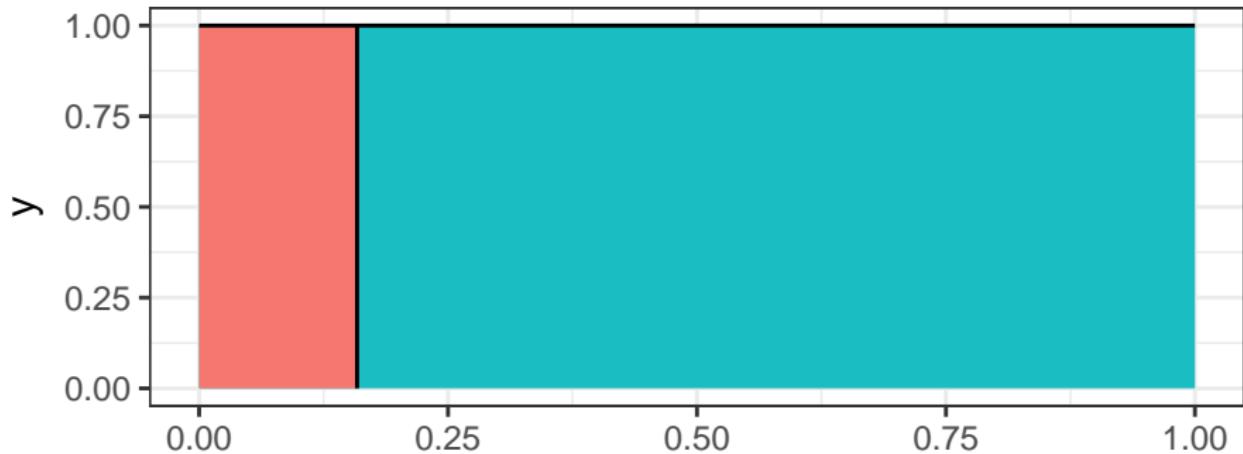
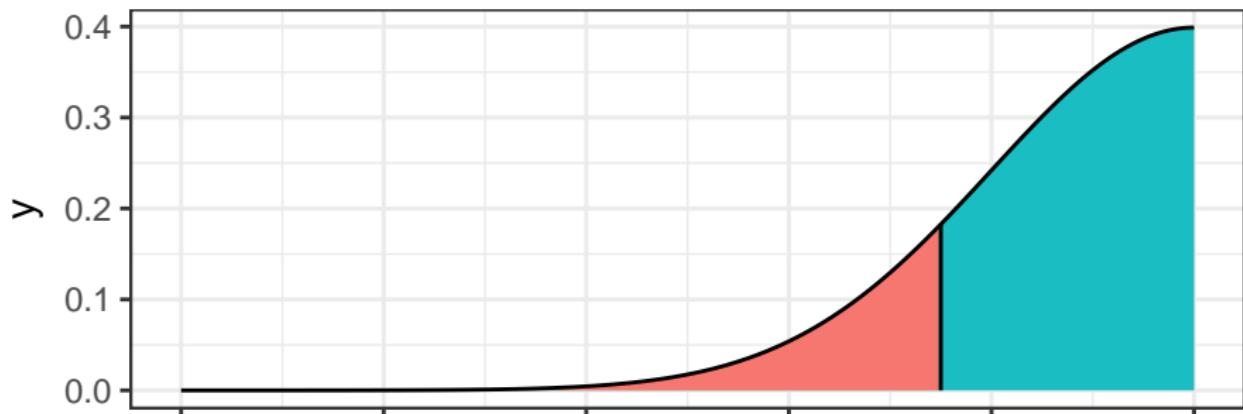
p-value construction



p-value construction



p-value construction



p-value construction

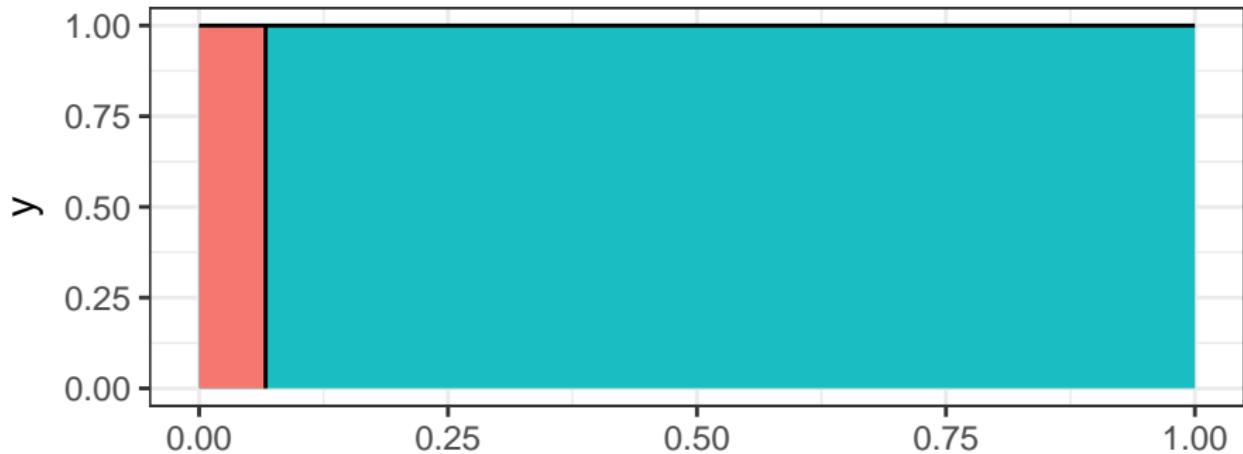
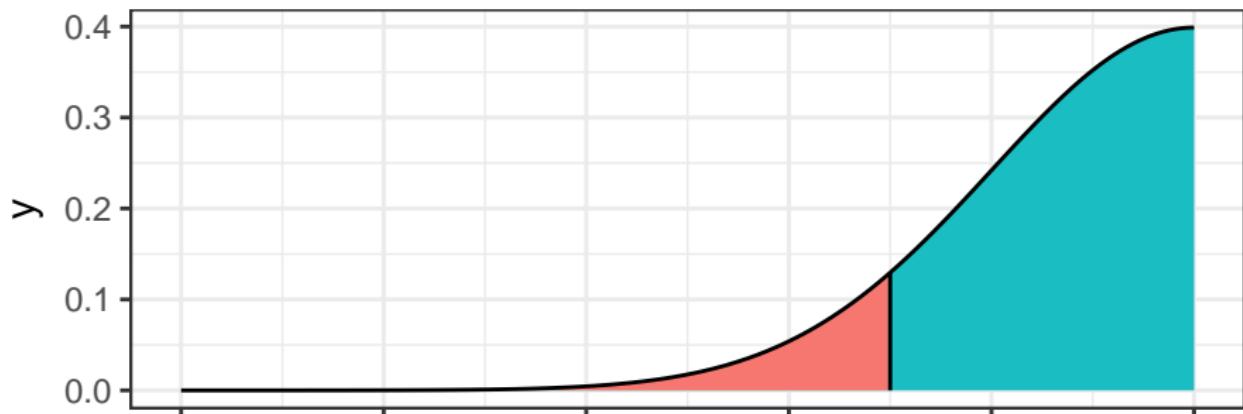
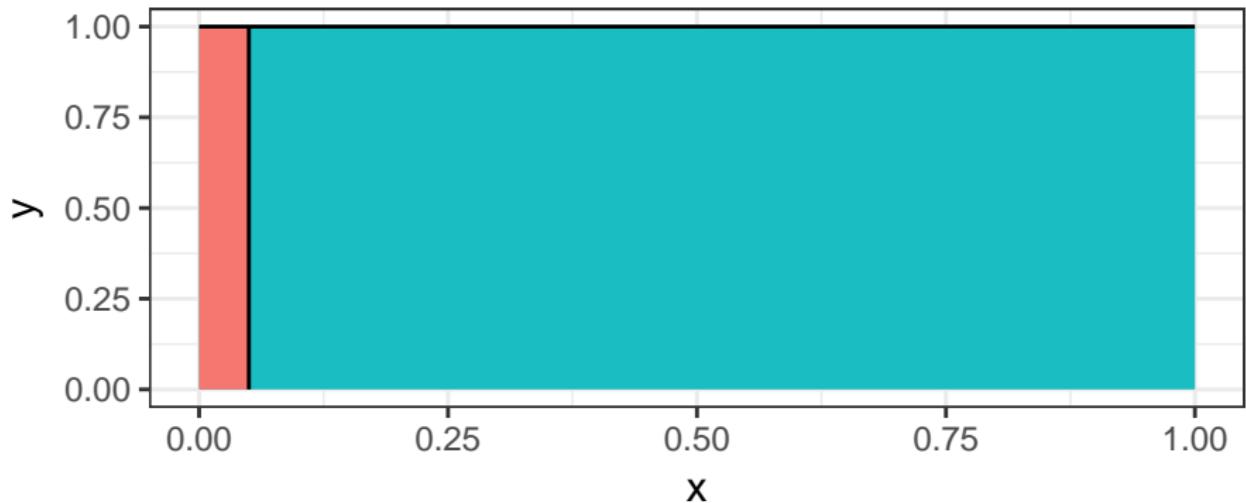


Table of Contents

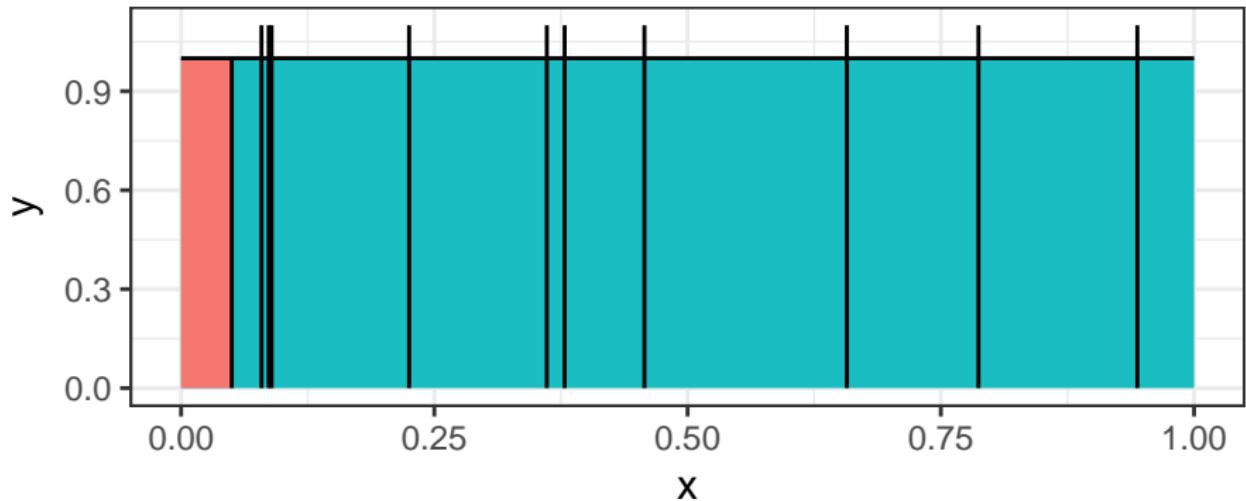
1 Hypothesis testing

2 Multiple hypotheses testing

Multiple hypotheses problem

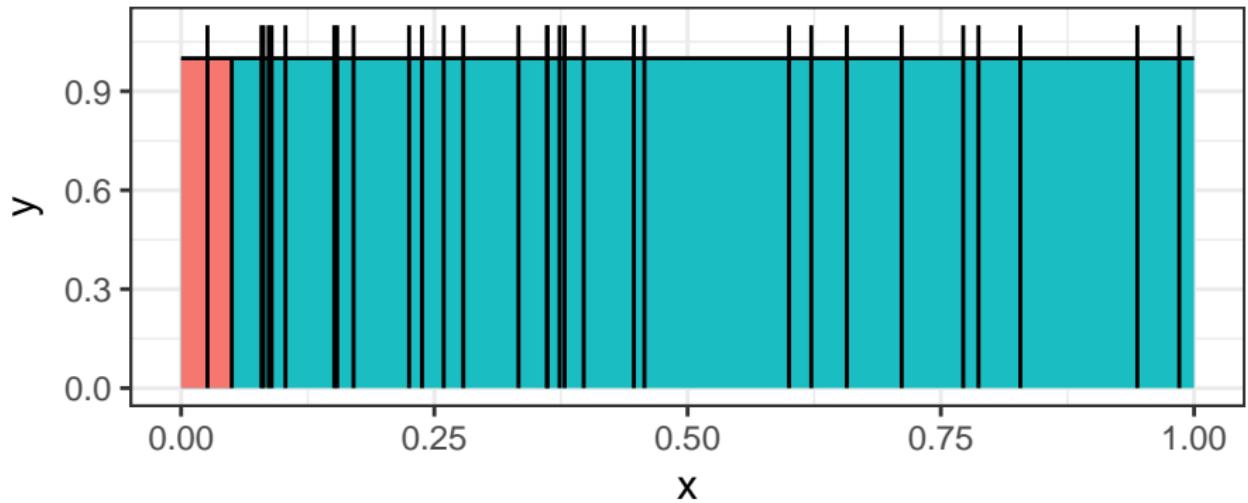


Multiple hypotheses problem



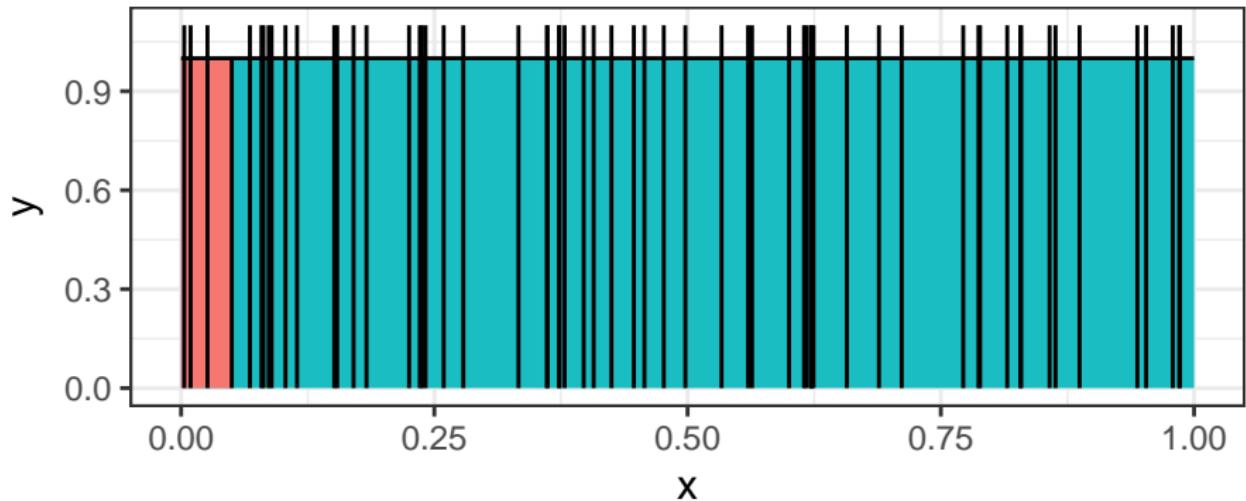
$$n = 10$$

Multiple hypotheses problem



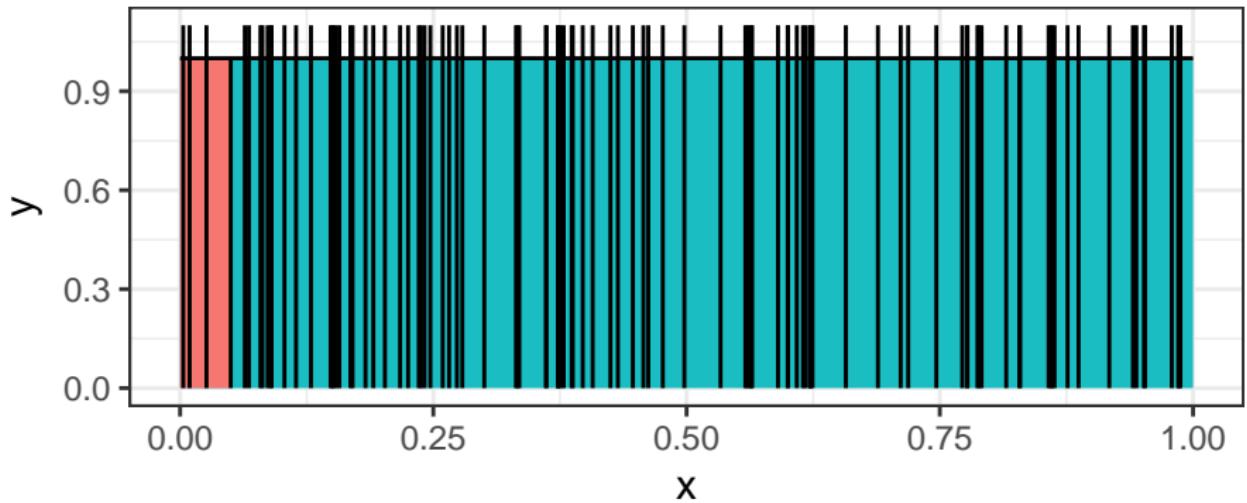
$$n = 30$$

Multiple hypotheses problem



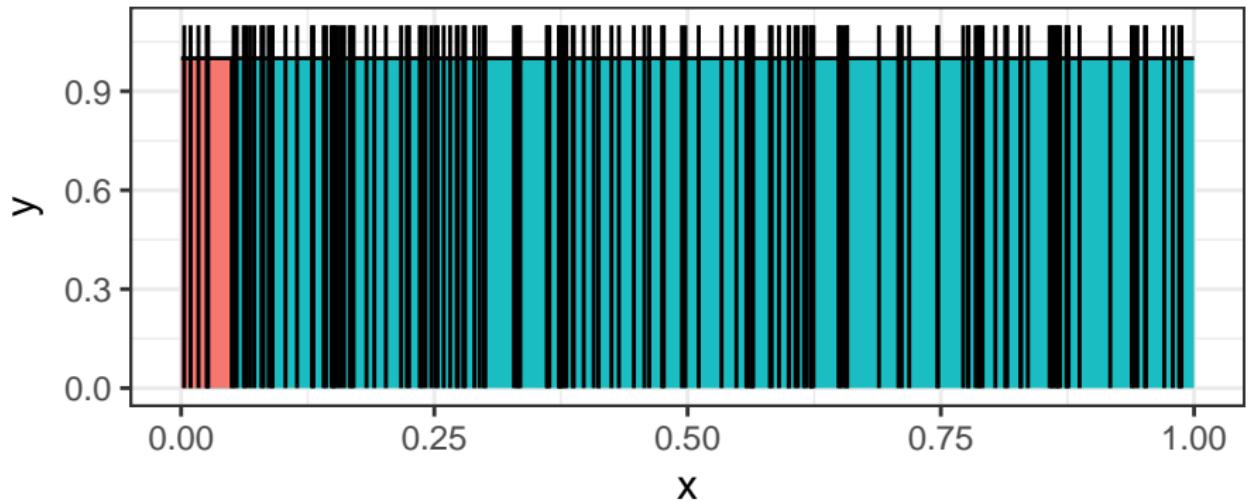
$$n = 60$$

Multiple hypotheses problem



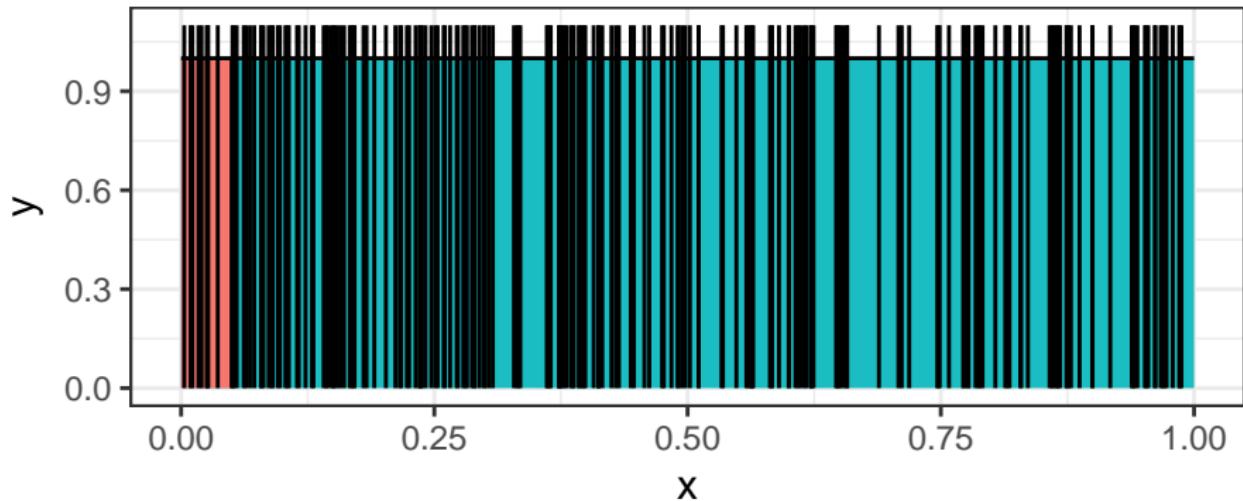
$$n = 100$$

Multiple hypotheses problem



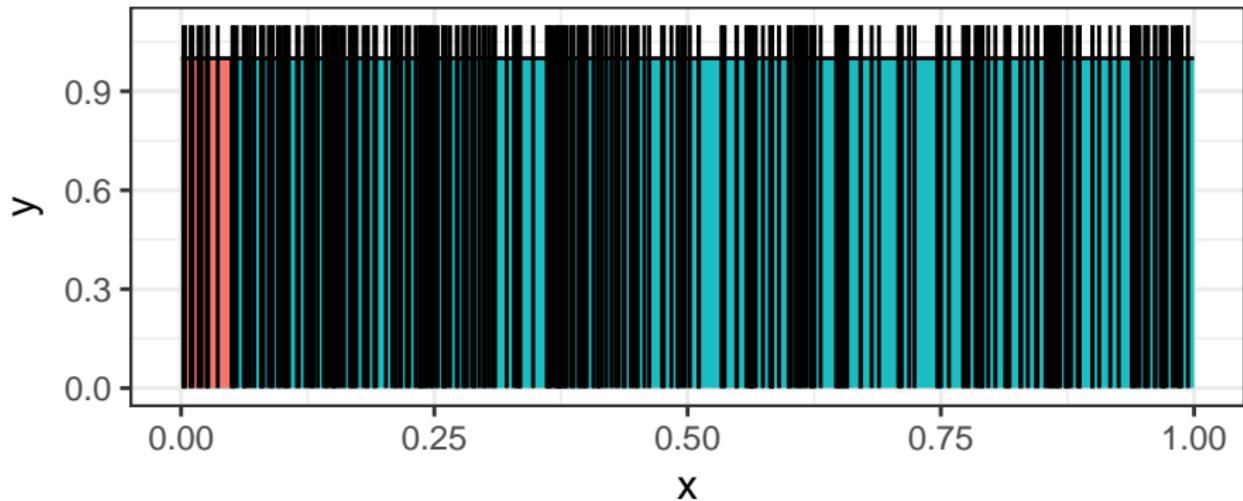
$$n = 150$$

Multiple hypotheses problem



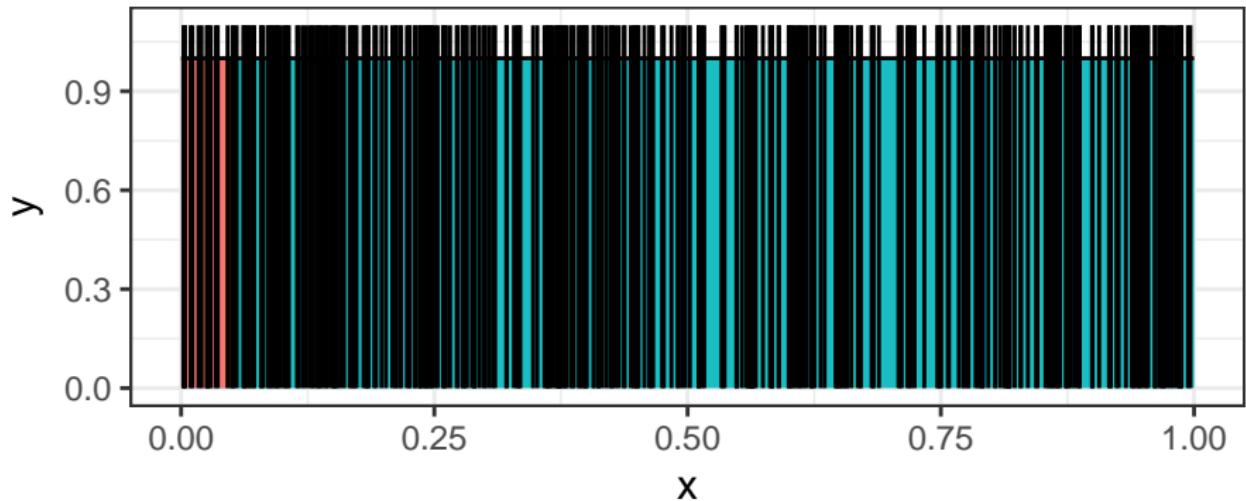
$$n = 210$$

Multiple hypotheses problem



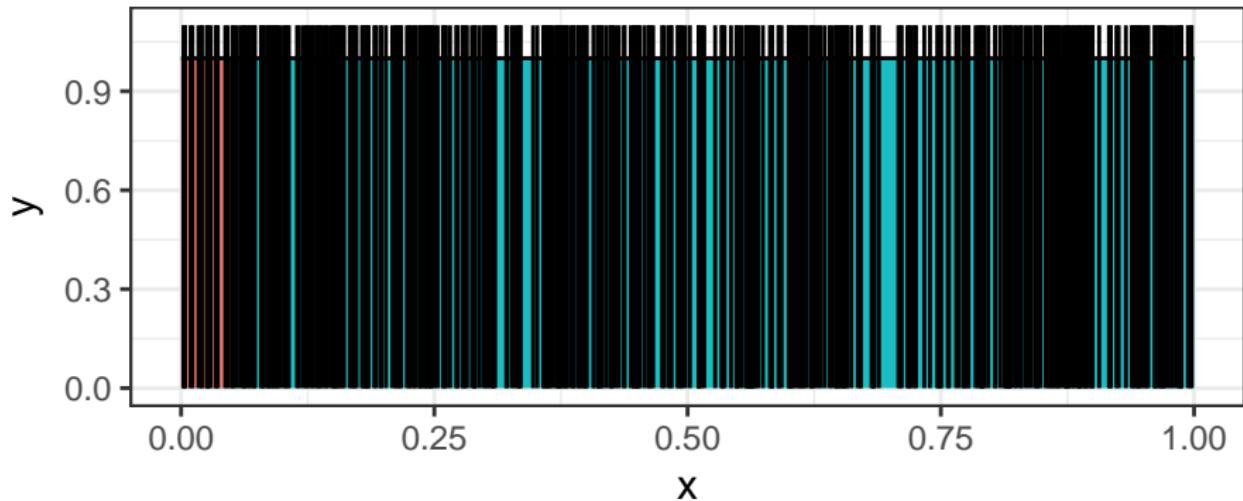
$$n = 280$$

Multiple hypotheses problem



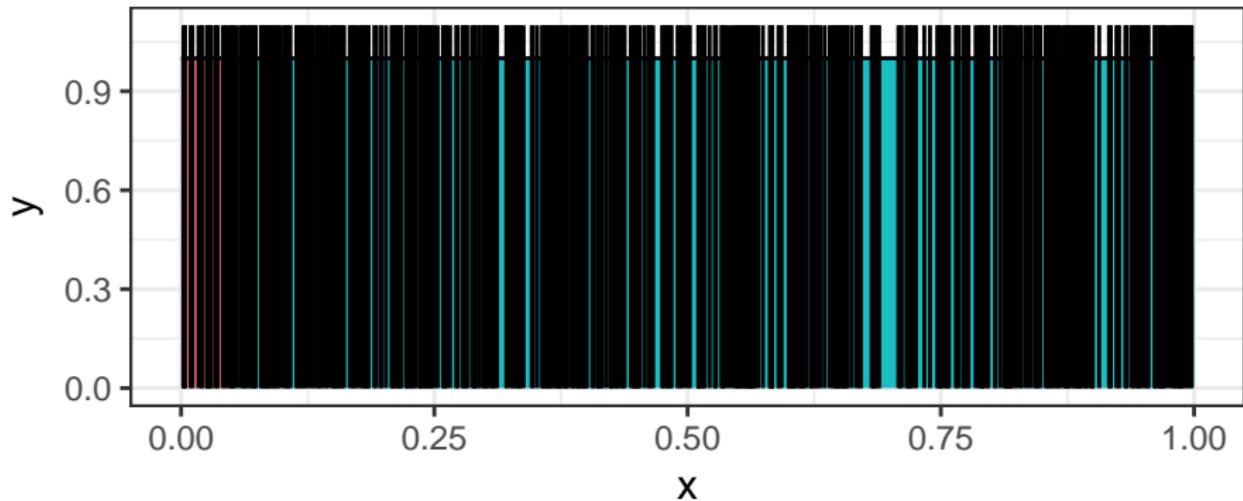
$$n = 360$$

Multiple hypotheses problem



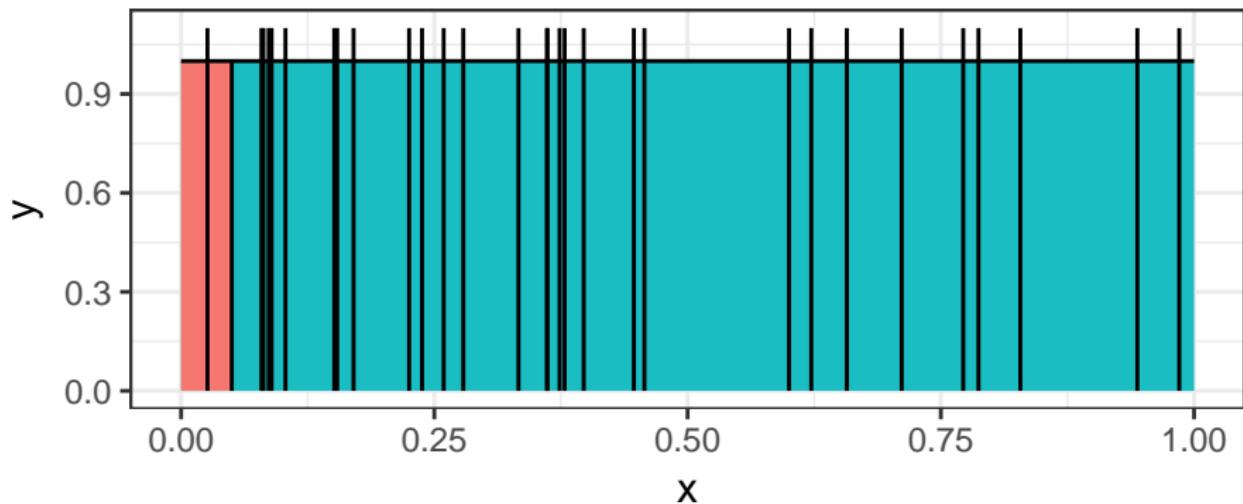
$$n = 450$$

Multiple hypotheses problem



$$n = 550$$

Multiple hypotheses problem



The threshold α don't control the risk of having fewer than one false positive.

Multiple hypotheses problem

| hypothesis | Claimed non-significant | Claimed significant | Total |
|------------|-------------------------|---------------------|-------|
| Null | TN | FP | m_0 |
| Non-null | FN | TP | m_1 |
| Total | S | R | m |

Multiple hypotheses solutions

Family Wise Error Rate (FWER)

- Bonferroni like procedure
- Control the risk of having fewer than one false positive
- $\Pr(FP < 1) < \alpha_{FWER}$
- $\alpha_{FWER} = \frac{\alpha}{m}$

Example

“We reject 14 hypothesis with a FWER of 0.05” “We reject 14 hypothesis at a level of 0.05 after Bonferoni correction”

Means: 14 hypotheses are not following the null distribution and we make this statement with a probability 0.05 of having fewer than one false positives in the 14 tests.

Multiple hypotheses solutions

False Discovery Rate (FDR)

- Benjamini-Hochberg like procedure
- Control the risk of having less than a proportion of false positive
- $\Pr \left(\mathbb{E} \left[\frac{FP}{R} \mid R > 0 \right] \right) \Pr (R > 0) < \alpha_{FDR}$
- adaptive procedure $\alpha_{FDR} \sim f_0$

| hypothesis | Claimed non-significant | Claimed significant | Total |
|------------|-------------------------|---------------------|-------|
| Null | TN | FP | m_0 |
| Non-null | FN | TP | m_1 |
| Total | S | R | m |

Multiple hypotheses solutions

False Discovery Rate (FDR)

- Benjamini-Hochberg like procedure
- Control the risk of having less than a proportion of false positive
- $\Pr \left(\mathbb{E} \left[\frac{FP}{R} \mid R > 0 \right] \right) \Pr (R > 0) < \alpha_{FDR}$
- adaptive procedure $\alpha_{FDR} \sim f_0$

Example

"We reject 254 hypothesis with a FDR of 0.05" "We reject 254 hypothesis with a level of 0.05 after BH correction"

Means: 254 hypotheses are not following the null distribution and we expect on average 5% or less of false positives in the 254.

Multiple hypotheses solutions

False Discovery Rate (FDR)

- Benjamini-Hochberg like procedure
- Control the risk of having less than a proportion of false positive
- $\Pr \left(\mathbb{E} \left[\frac{FP}{R} \mid R > 0 \right] \right) \Pr (R > 0) < \alpha_{FDR}$
- adaptive procedure $\alpha_{FDR} \sim f_0$

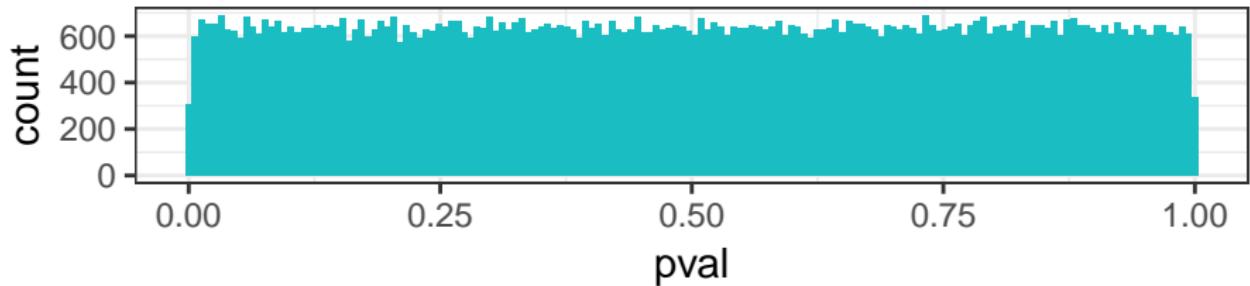
Example

"We reject 254 hypothesis with a FDR of 0.05" *"We reject 254 hypothesis with a level of 0.05 after BH correction"*

Means: 254 hypotheses are not following the null distribution and we expect on average 5% or less of false positives in the 254.

The number of FPs increases with the number of TPs

FWER versus FDR control



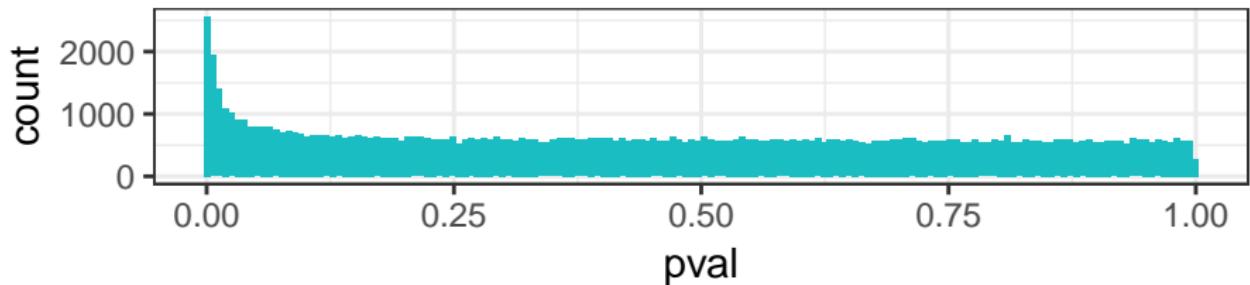
$$\Pr(FP < 1) < \alpha_{FWER}$$

$$\Pr\left(\mathbb{E}\left[\frac{FP}{R} \middle| R > 0\right]\right) \Pr(R > 0) < \alpha_{FDR}$$

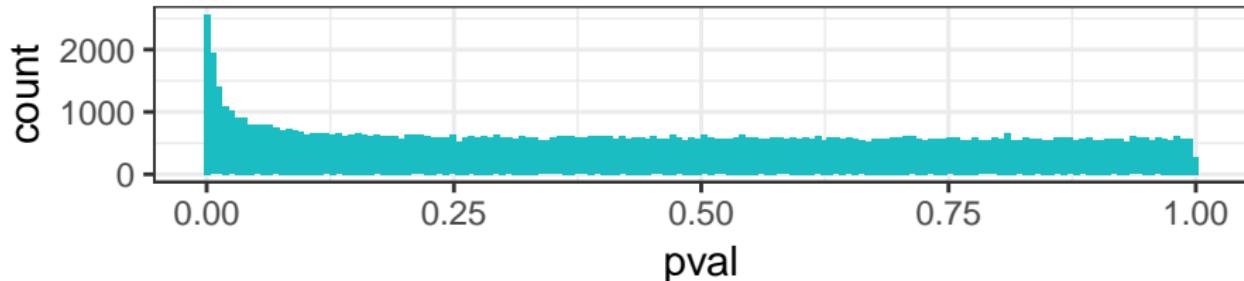
When $TP \leq 1$ FWER and FDR control are identical.

The difference increases with the number of TPs

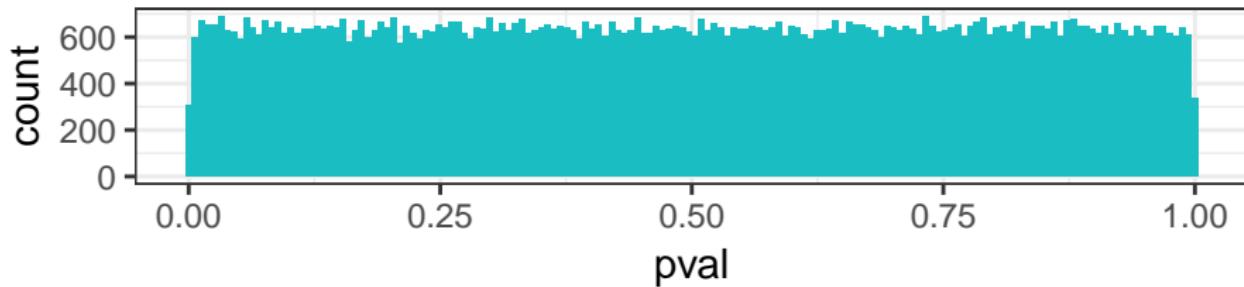
FDR control



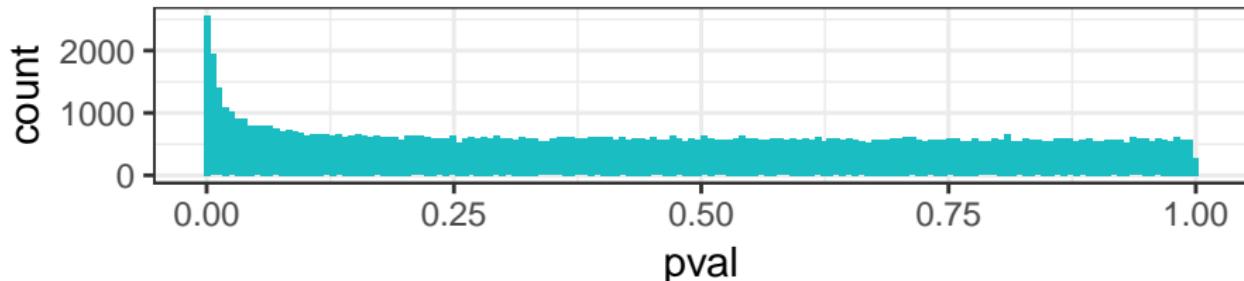
FDR control



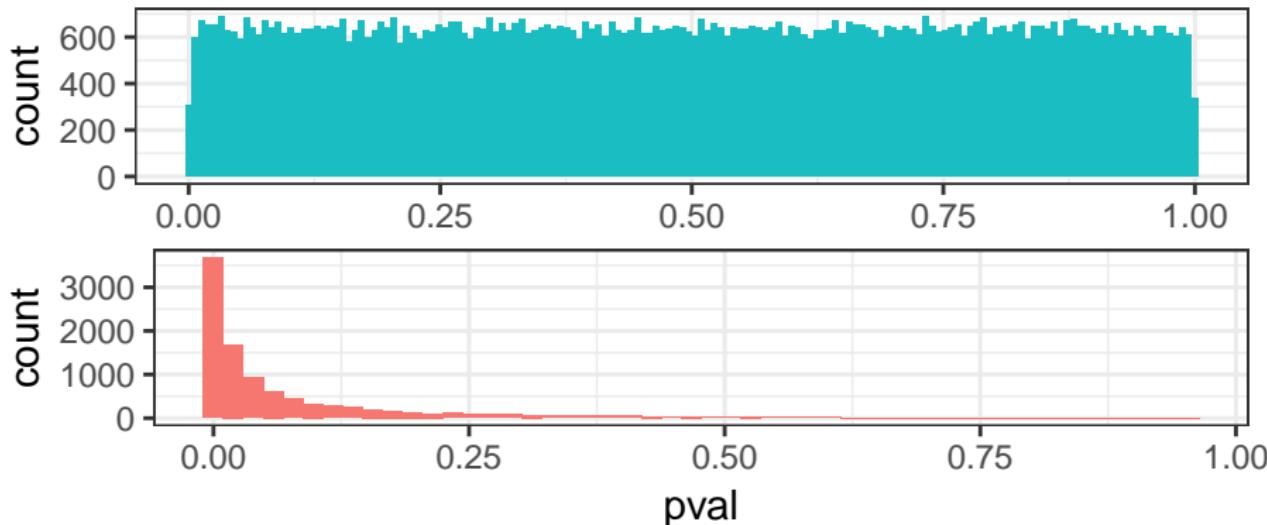
When we analyse data we hope to get a mixture between:



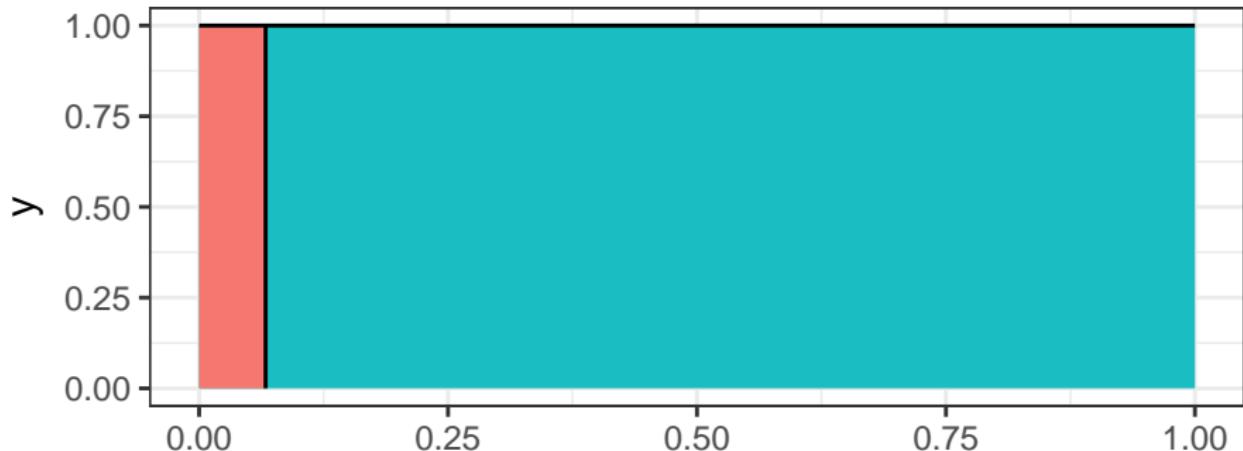
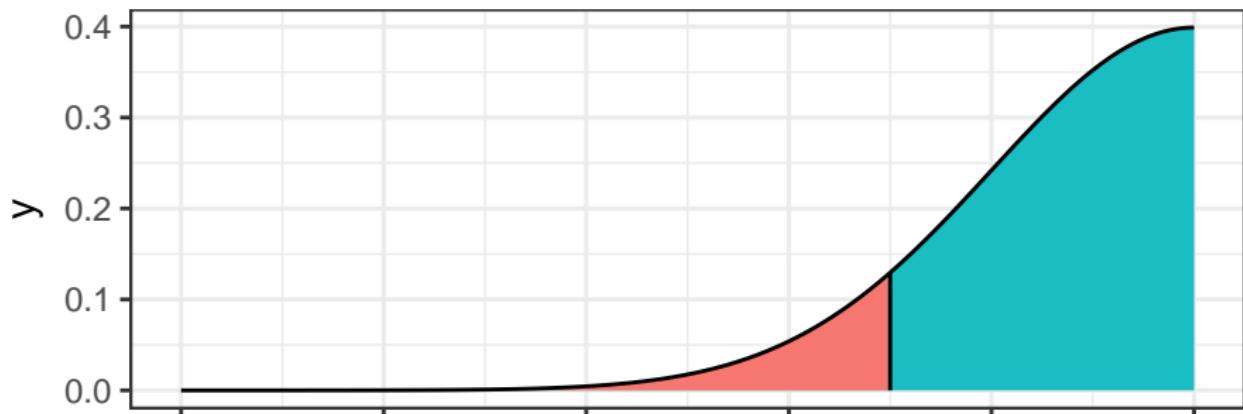
FDR control



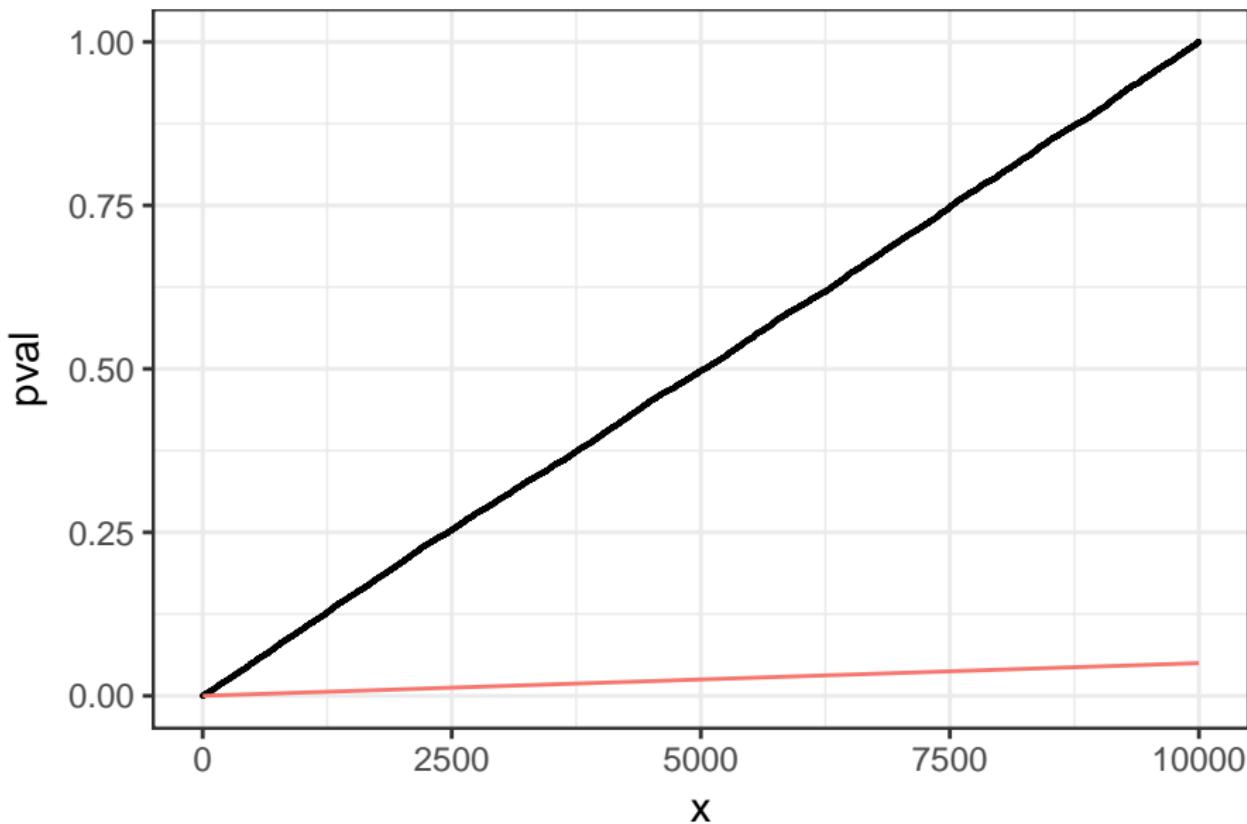
When we analyse data we hope to get a mixture between:



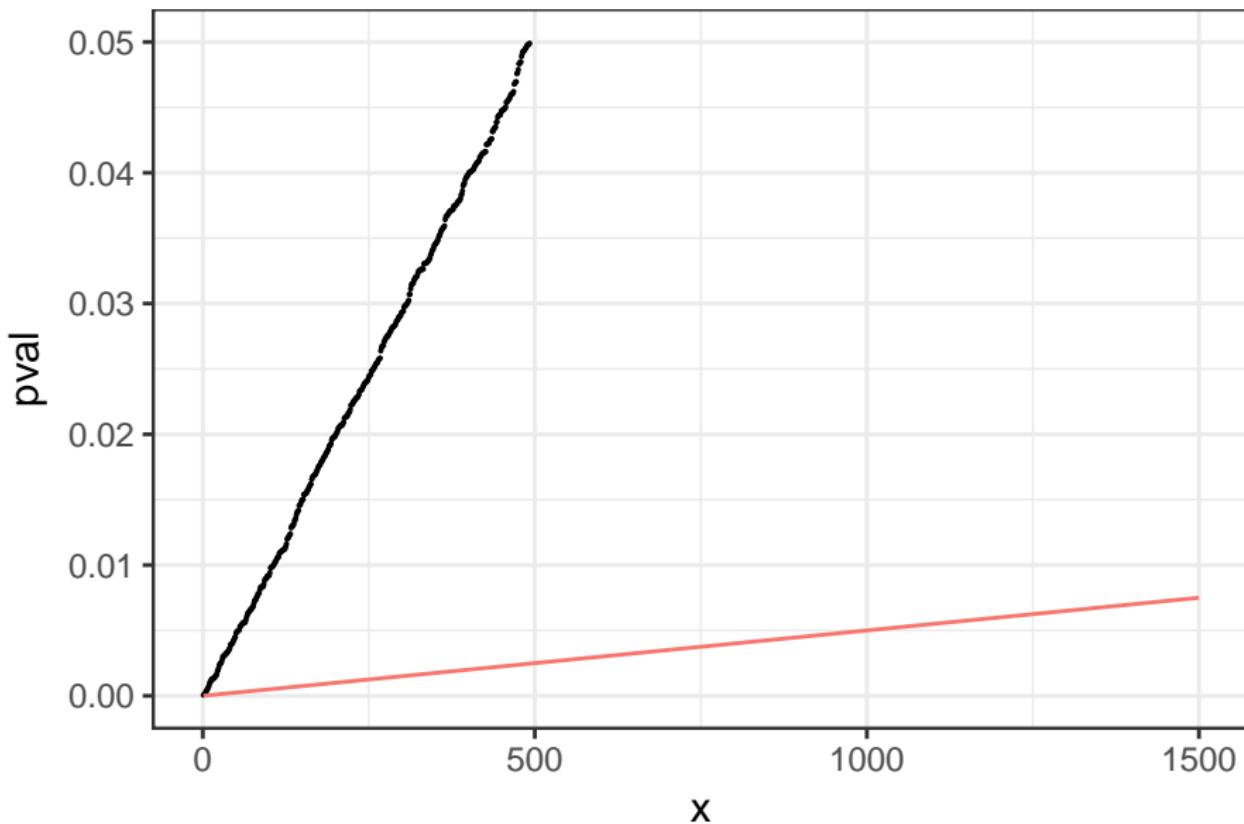
FDR control: p -value construction



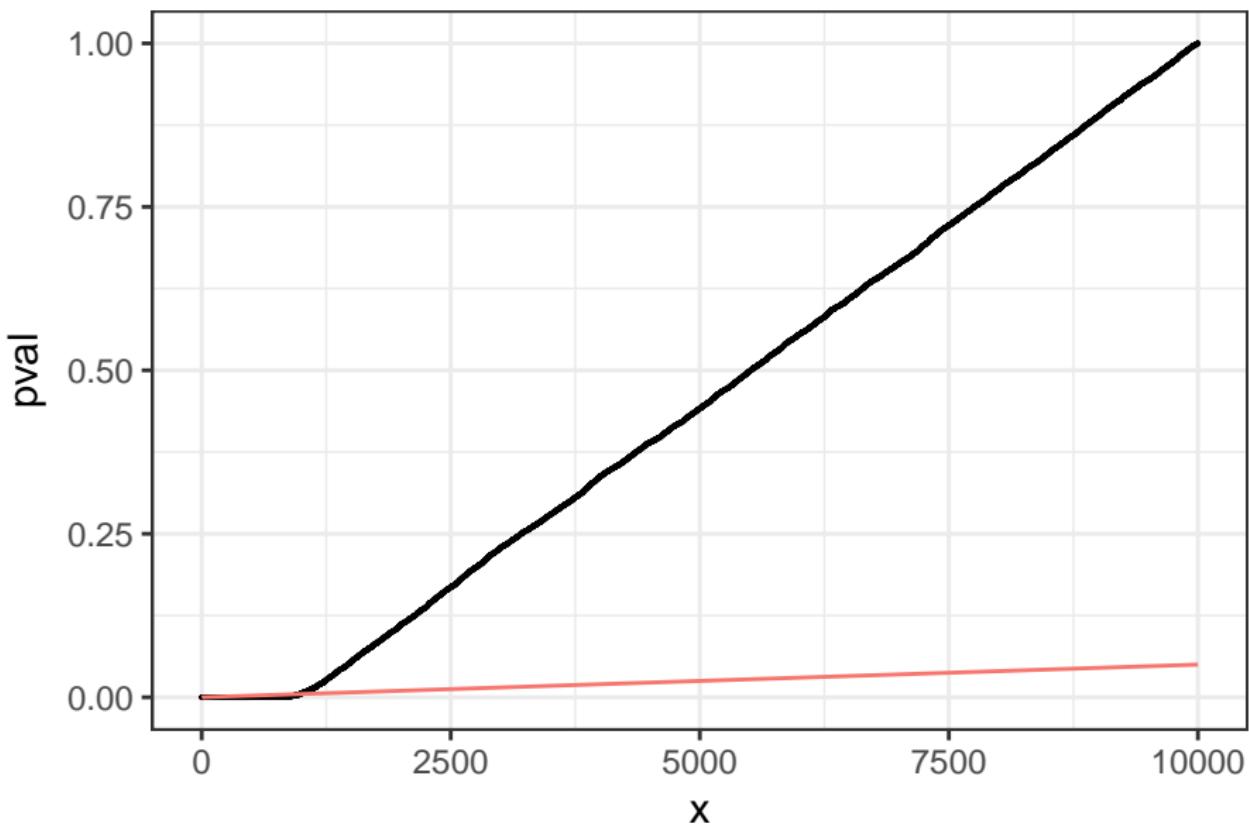
FDR control: Benjamini-Hochberg



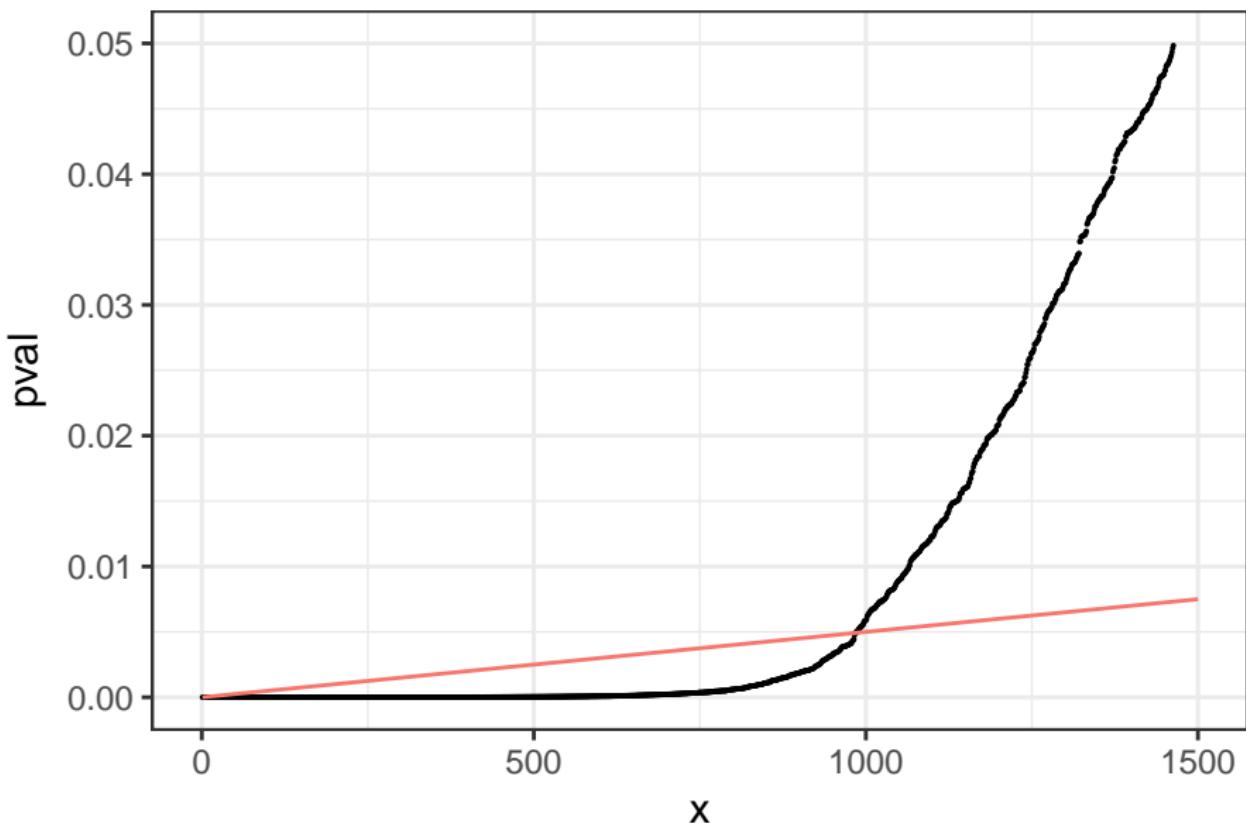
FDR control: Benjamini-Hochberg



FDR control: Benjamini-Hochberg



FDR control: Benjamini-Hochberg

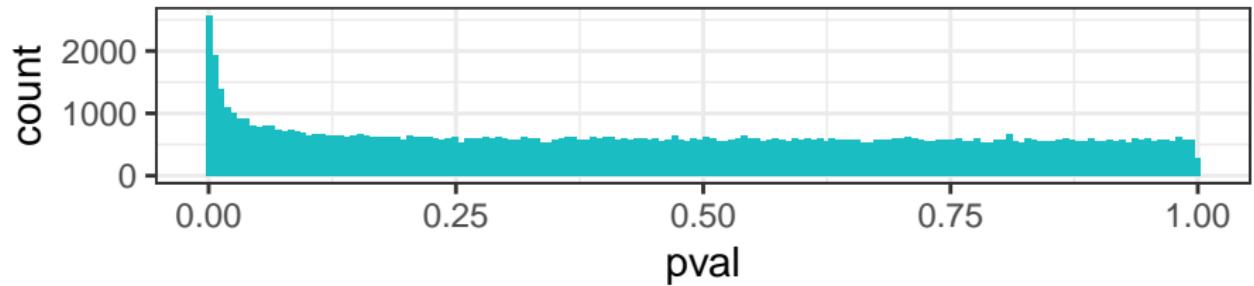


Simple multiple correction with R:

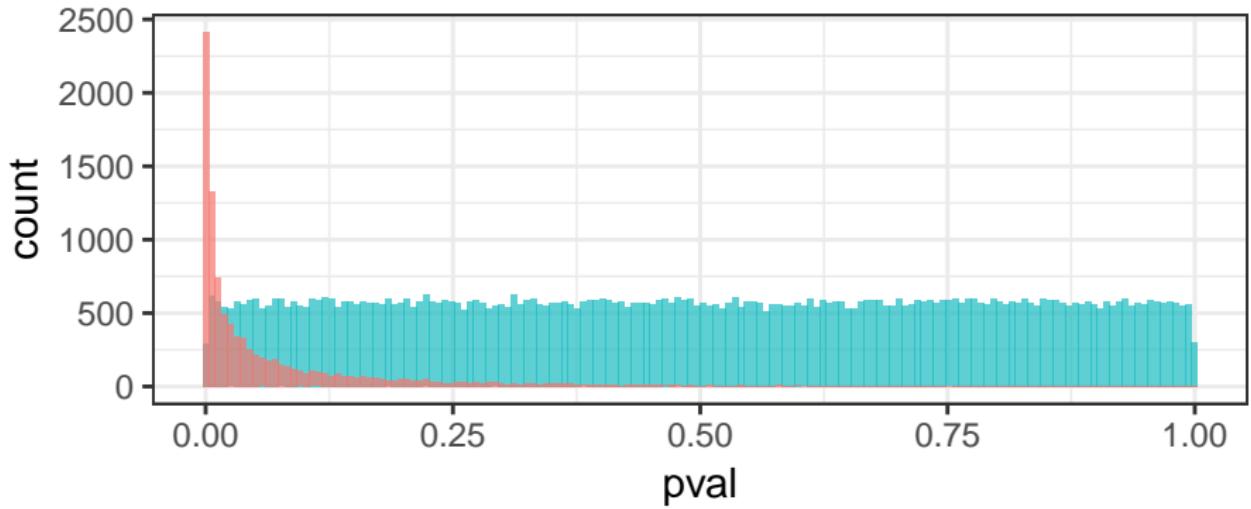
```
p.adjust(p, method = p.adjust.methods)
```

```
"holm", "hochberg", "hommel", "bonferroni", "BH", "BY",  
"fdr", "none"
```

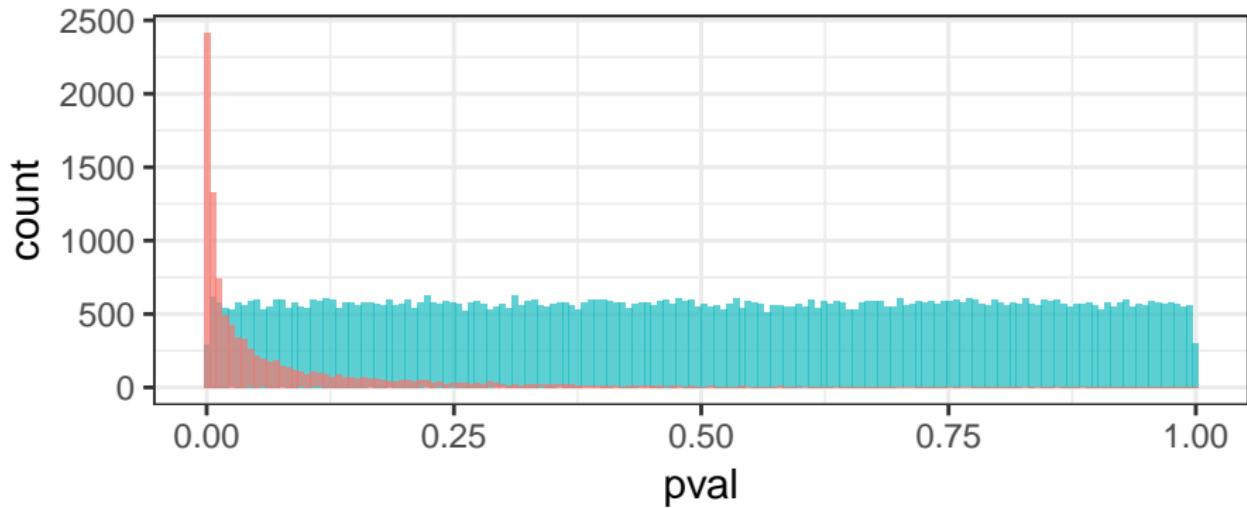
FDR control: local FDR (ℓFDR) of Efron



FDR control: local FDR (ℓFDR) of Efron

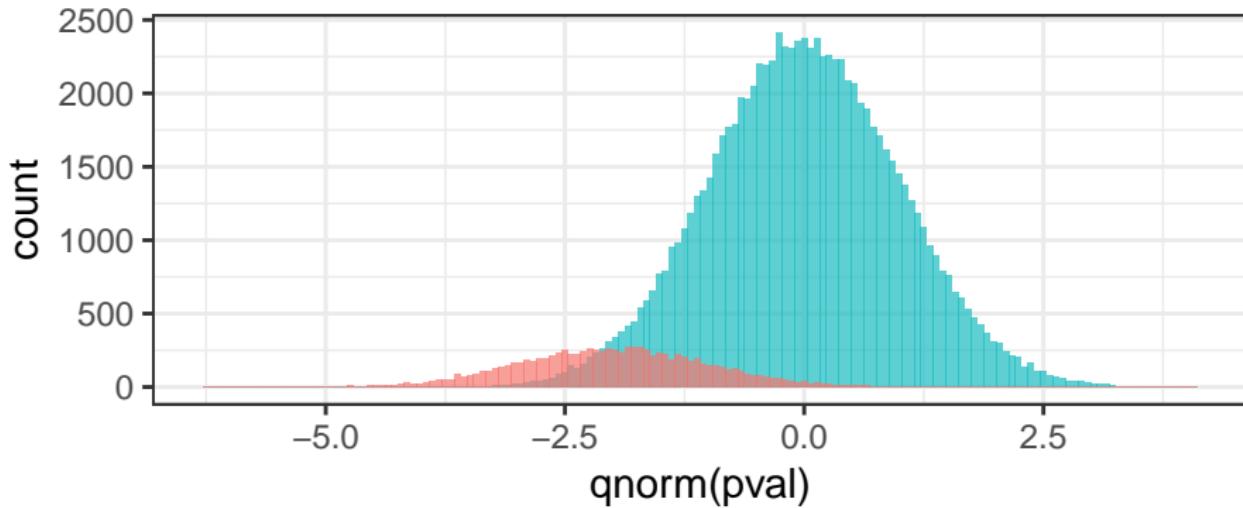


FDR control: local FDR (ℓFDR) of Efron



$$\ell FDR(x_i) = \frac{\Pr(x_i|H_i = 0)}{\Pr(x_i|H_i = 0)\Pr(x_i|H_i = 1)}$$

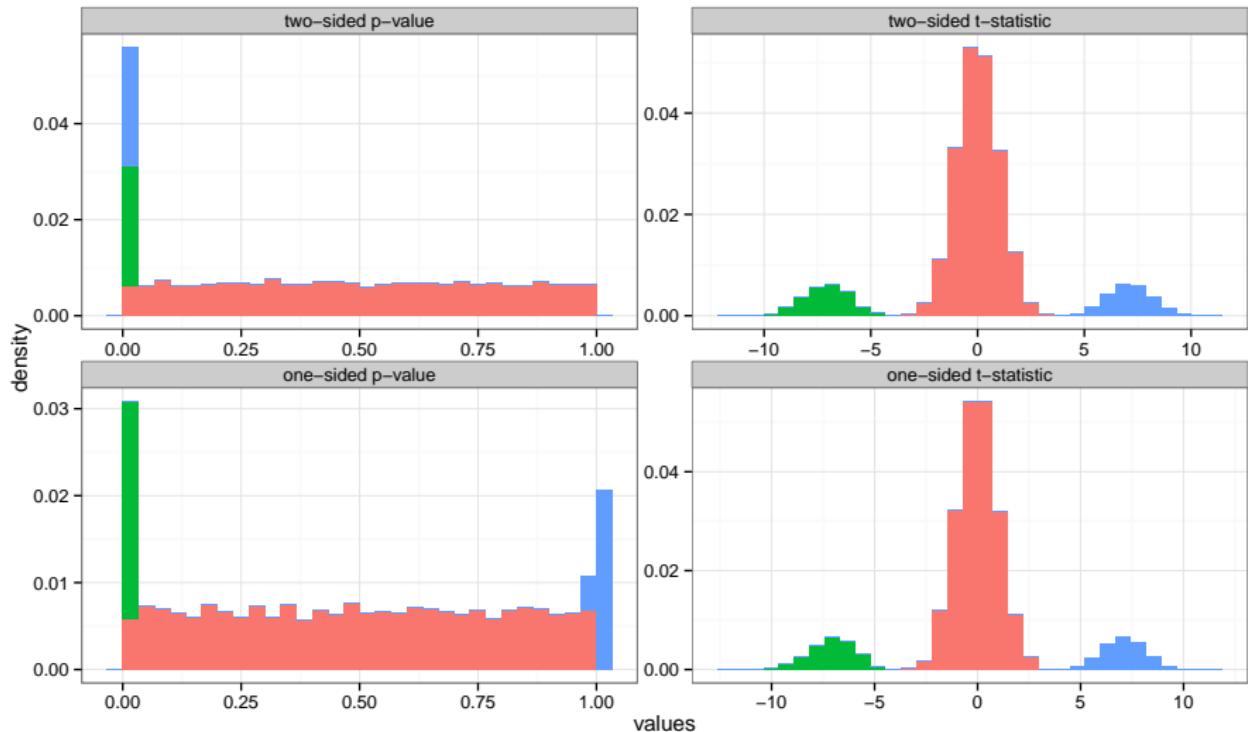
FDR control: local FDR (ℓFDR) of Efron



$$\ell FDR(x_i) = \frac{\Pr(x_i|H_i=0)}{\Pr(x_i|H_i=0)\Pr(x_i|H_i=1)}$$

work with z -values instead of p -values

Unilateral versus Bilateral tests



Take home message

- FWER vs FDR
- Under H_0 i.i.d. p -values follow an uniform distribution
- Unilateral versus bilateral test
- be careful about the correlation between tests



That's all Folks!

and thank you for your attention !