

Zero effects



Non-zero effects



False discoveries



True discoveries

We have p hypotheses to test, one for each variable/feature.

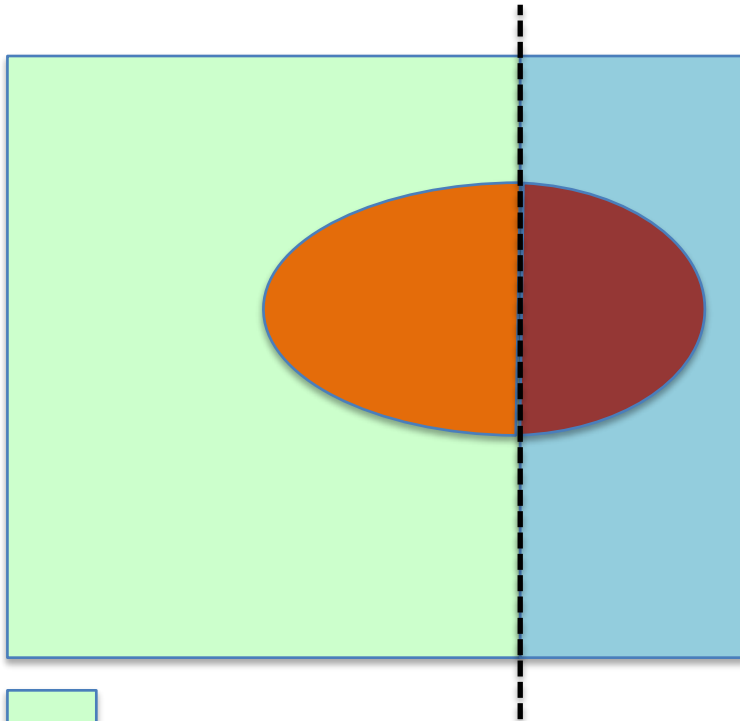
Each null hypothesis states that the effect is zero.

Hypotheses are represented by the square that is divided into two parts corresponding to truly zero and non-zero effects.

A statistical procedure is used to label each variable as a discovery or non-discovery. (“Discovery” = “significant variable” = “rejected null hypothesis”)

Some discoveries are true (non-zero effects) and some are false (zero effects).

Controlling false positive rate at level α



Zero effects



Non-zero effects



False discoveries



True discoveries

Proportion of false discoveries out of all zero effects $\leq \alpha$

Nothing is said about true discoveries.

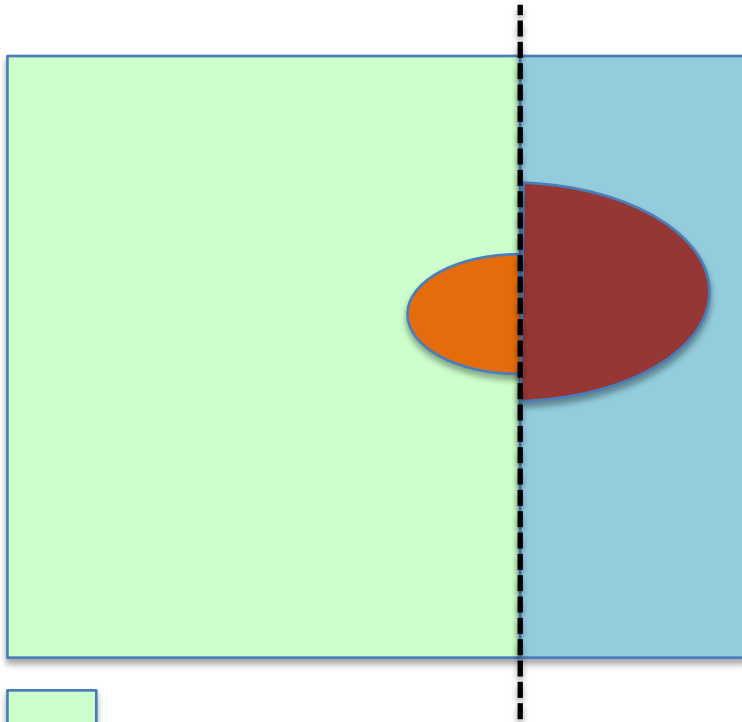
Can be done by thresholding P-value at $\leq \alpha$





This is the standard way of doing hypothesis testing in statistics.

Empirical estimate from Figure:

$$\frac{\text{False discoveries}}{\text{Zero effects}} \leq \alpha$$

Controlling false discovery rate at level α



-  Zero effects
-  Non-zero effects
-  False discoveries
-  True discoveries

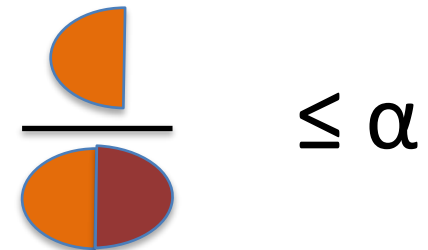
Proportion of false discoveries out of all discoveries $\leq \alpha$

Can be done by

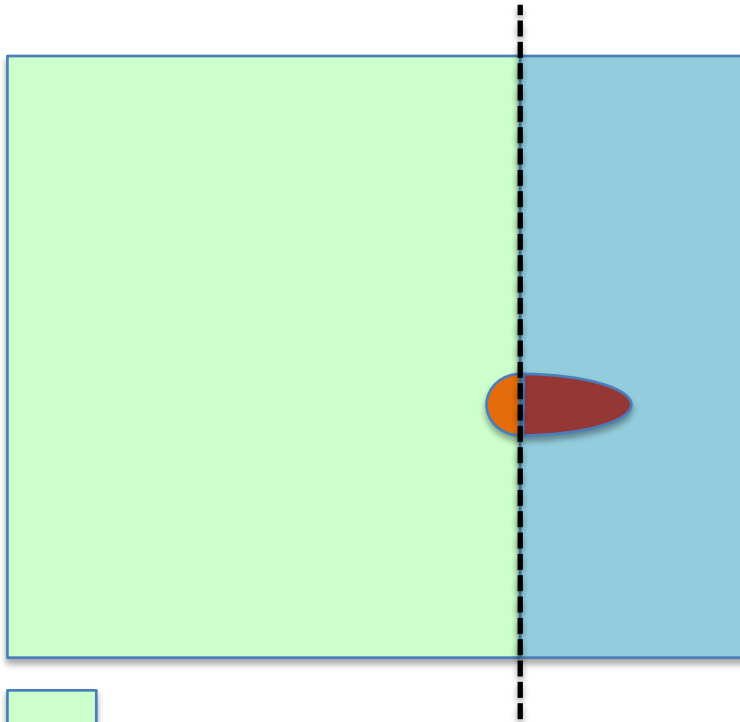
1. Benjamini-Hochberg (independence)
2. Benjamini-Yekutieli (any dependence)

This approach correctly discovers many non-zero effects and keeps the proportion of zero effects low among the discoveries

Empirical estimate from Figure:

$$\frac{\text{False discoveries}}{\text{False discoveries} + \text{True discoveries}} \leq \alpha$$


Controlling family-wise error rate at level α



Zero effects



Non-zero effects



False discoveries



True discoveries

Probability of at least one false discovery $\leq \alpha$

Nothing is said about true discoveries.

Can be done by

1. thresholding P-value at $\leq \alpha/p$ (Bonferroni)
2. Holm method

Very stringent requirement and discovers usually only at most a few non-zero effects.

Empirical estimate from Figure:

Whether there is any element in

