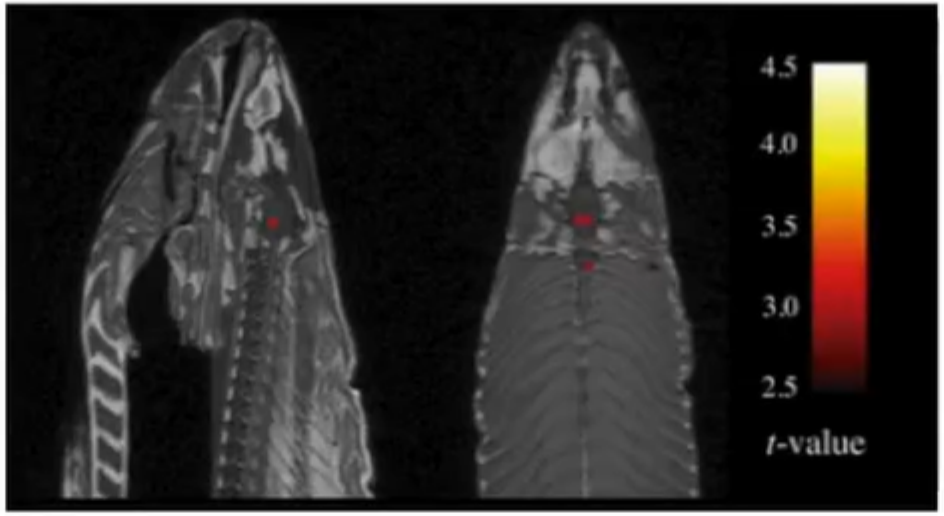
The dead salmon experiment:results



- The salmon brain was split into $N = 8,064$ voxels
- In each voxel a statistical test at level $\alpha = .1\%$ was performed to see if activity was statistically significant
- 8 voxels* were found to be.
- Conclusion: these voxels contain salmon neurons involved in social perception. NO !

Statistical analysis

- Statistical model for a fixed voxel: observe (normalized) signal $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $i = 1, \ldots, n$, $n = 15$.
- Hypothesis testing problem:

$$H_0 : \mu = 0$$
$$H_1 : \mu \neq 0$$

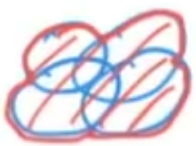
- Apply the T-test
- Recall the interpretation of level $\alpha = .1\%$ of a test:

  If we repeat the experiment many times, at most .1% *of the tests will make an error of type* 1.

- $.1\% \cdot 8{,}064 \simeq 8$

Multiple testing

- We cannot make conclusions about "all the voxels" at once: we are bound to make mistakes N voxels
- Two solutions:
  - Control Family Wise Error Rate (FWER): Find $C_1, \ldots, C_N$ such that



$$\mathbb{P}_{\mu_i=0}\Big(\bigcup_{i=1}^{N}\{|T_i| > C_i\}\Big) \leq \alpha$$

  - Control False Discovery Rate (FDR): Find $C_1, \ldots, C_N$ such that

$$\text{FDR} = \mathbb{E}\Big[\frac{\#\{i\,:\,|T_i| > C_i \;\&\; \mu_i = 0\}}{\#\{i\,:\,|T_i| > C_i\}}\Big] = \mathbb{E}_{\mu=0}\Big[\frac{\#\text{ of False discoveries}}{\#\text{ of discoveries}}\Big] \leq \alpha$$

- In both cases, it is easier to work with the p-values

$$P_i = \mathbb{P}_{\mu=0}(|T| > |t_i^{\text{obs}}|)$$

where t_i^{obs} is the observed value of the test statistic for the ith test and $T \sim t_{n-1}$

The Bonferroni method

- To control FWER, we use use the *Bonferroni correction*.
- Rather than rejecting each test at level α, we use the (much smaller) level $\frac{\alpha}{N}$.

- In other words:

$$\text{Reject } i\text{th test} \quad \Longleftrightarrow \quad P_i < \frac{\alpha}{N}$$

- In the salmon example this means that each test is performed at level $0.001/8{,}064 = 1.24 \cdot 10^{-7}$
- Often this is way to *conservative* (no discoveries).

Note that, with the Bonferroni method: $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$

$$\text{FWER} = \mathbb{P}_{\mu_i=0}\Big(\bigcup_{i=1}^{N}\{P_i < \frac{\alpha}{N}\}\Big) = \mathbb{P}_{\mu_i=0}\Big(\bigcup_{i=1}^{N}\{|T_i| > q^{t_{n-1}}_{\frac{\alpha}{2N}}\}\Big)$$

$$\leq \sum_{i=1}^{N} \mathbb{P}_{\mu_i=0}\Big(|T_i| > q^{t_{n-1}}_{\frac{\alpha}{2N}}\Big) = \sum_{i=1}^{N} \frac{\alpha}{N} = N \cdot \frac{\alpha}{N} = \alpha$$

The Benjamini-Hochberg method

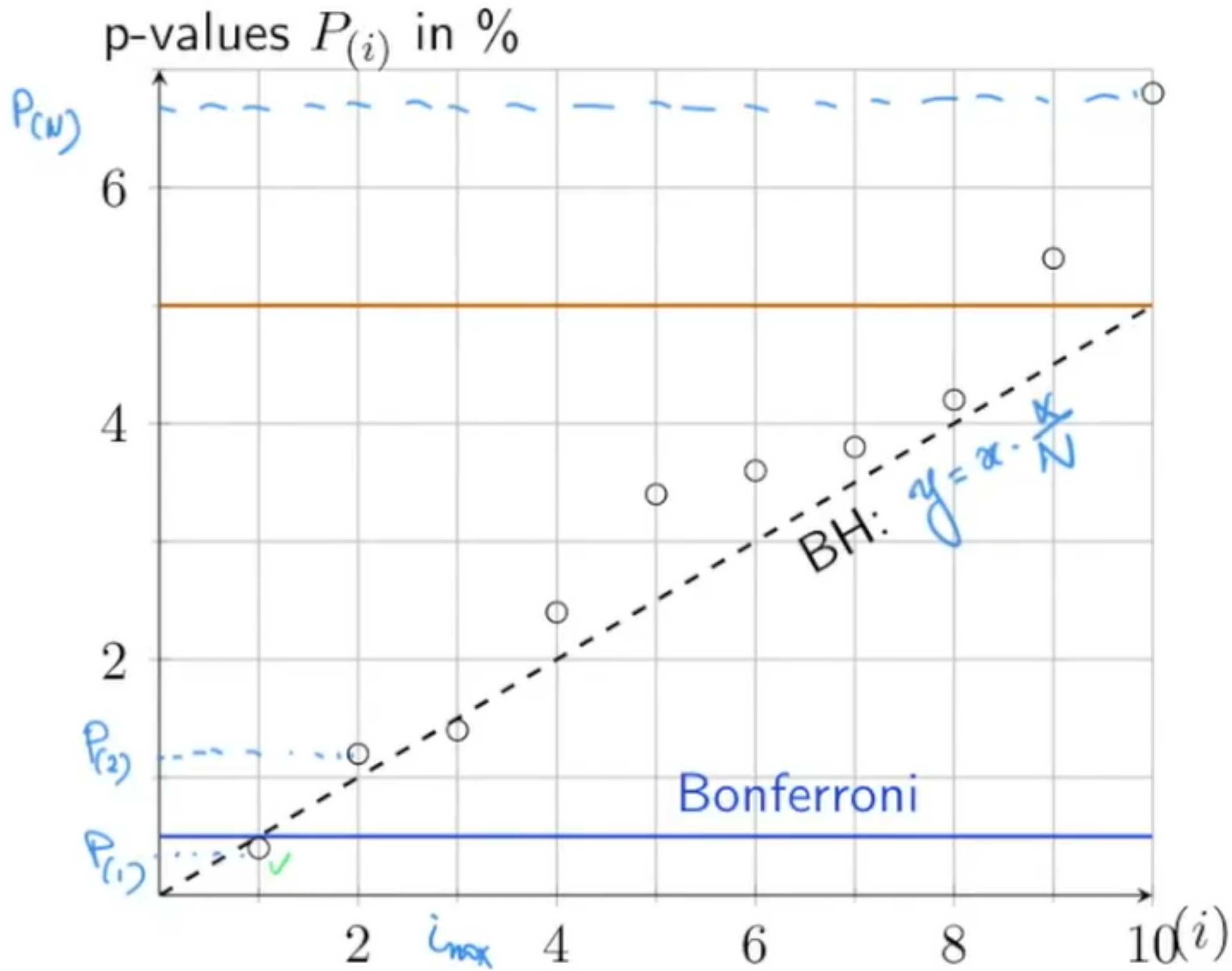
- To control FDR, we use use the *Benjamini-Hochberg (BH)* method.
- Intuitively, we should reject tests with the smallest p-values
- Order p-values: $P_{(1)} < P_{(2)} < P_{(3)} < P_{(4)} < \cdots < P_{(N)}$ and call "the (i)th test" the test with p-value $P_{(i)}$.
- Idea: reject all tests (i) such that $i \leq i_{\max}$
- Rule

$$i_{\max} := \max\left\{i \,:\, P_{(i)} < i \cdot \frac{\alpha}{N}\right\}$$

- Benjamini and Hochberg (1995, 68K citations) have shown that with this procedure:

$$\text{FDR} \leq \alpha$$

- There are *many* variations of the BH procedure, in particular to account for correlations between p-values.



$P_{(i)} < i \frac{\alpha}{N}$

In this figure, $N = 10, \alpha = 5\%$.

- Bonferonni: $\alpha/N = .005$. Only test (1) is rejected.
- BH: $i_{\max} = 3$. Tests $(1), (2)$ and (3) are rejected.