

Beyond p-values: Modern Approaches to High-Dimensional Inference in fMRI Studies

sara shahama v p|26-10-2025

# **Executive Summary**

This project explores how the human brain responds differently when viewing images of *chairs* versus *scissors*, using the Haxby 2001 fMRI dataset. My goal was to understand not just where the brain activates, but how statistical methods influence what we consider “significant.” I conducted a voxel-wise analysis comparing thousands of brain regions between the two conditions and applied multiple statistical correction techniques — particularly the Benjamini–Hochberg (BH) and Storey’s q-value methods — to control for false discoveries in this high-dimensional data.

The results showed clear activation patterns that were consistent across both approaches, confirming their reliability while also revealing how adaptive methods like the q-value can detect subtler signals. I also began exploring knockoff-based inference, a modern statistical framework that offers exciting potential for more robust voxel selection in future work.

Through this project, I not only learned to handle real fMRI data but also gained a deeper appreciation for the role of statistical reasoning in neuroscience. The experience combined my interests in data science, mathematics, and the study of the human brain — and gave me a first-hand look at how analytical precision transforms complex data into meaningful scientific insight.

# **Introduction**

The human brain is constantly processing, recognizing, and categorizing the world around us — often in a fraction of a second. One of the most fascinating questions in neuroscience is *how* the brain distinguishes between different visual objects, such as a chair and a pair of scissors, both familiar yet distinct in shape, function, and context. Understanding these neural representations not only deepens our knowledge of perception but also builds the foundation for cognitive neuroscience, brain–computer interfaces, and artificial intelligence.

This project uses the **Haxby 2001 fMRI dataset**, a widely known benchmark in visual recognition research, to explore this question through the lens of data analysis. My main goal was to see how brain activation patterns differ when subjects viewed images of “chairs” and “scissors,” and to understand how various **statistical inference techniques** influence what we interpret as significant brain activity.

Rather than relying solely on raw voxel intensities, I focused on the challenge of **multiple comparisons** — a central issue in fMRI research, where thousands of voxel-level tests can easily lead to false discoveries. To address this, I applied two major **False Discovery Rate (FDR)** correction methods — **Benjamini–Hochberg** and **Storey’s q-value** — and examined how their differences shape the resulting brain maps.

By blending statistical reasoning with neuroimaging analysis, this project aims to highlight how careful data interpretation can turn noisy, high-dimensional brain signals into meaningful scientific insight. Beyond technical results, it also reflects my curiosity about the intersection of statistics and neuroscience — and how thoughtful modeling can bring us closer to understanding the complex structure of human perception.

# **Methodology**

This project was built around the Haxby 2001 fMRI dataset, a classic benchmark in cognitive neuroscience that investigates how the human brain encodes and differentiates visual object categories. I chose this dataset because it is both scientifically rich and methodologically demanding — it contains high-dimensional voxel-wise activation data that requires rigorous statistical control. Among the various stimulus conditions available, I focused on “chair” versus “scissors”, two object categories that are visually distinct yet conceptually similar enough to produce interesting and interpretable activation contrasts.

*Data Preparation and Processing*

* The dataset was loaded and preprocessed in R, where each 4D fMRI image was reshaped into a 2D matrix of voxels across timepoints. Each voxel represents a tiny three-dimensional unit of brain space, with intensity values corresponding to neural activation strength. The data were organized such that each row corresponded to a timepoint and each column to a voxel. I then separated the data into two conditions — *chair* and *scissors* — for independent analysis.

*Voxel-Wise Statistical Analysis*

* For each voxel, I performed an independent two-sample t-test comparing the mean activation between the two conditions. This produced thousands of p-values, one for each voxel, representing the strength of evidence against the null hypothesis of “no difference” in brain activation. However, given the large number of simultaneous tests, interpreting raw p-values directly would inflate the risk of false discoveries — a well-known issue in neuroimaging.

*False Discovery Rate Control*

* To address this, I implemented two prominent False Discovery Rate (FDR) control methods:

1. *Benjamini–Hochberg (BH) Procedure* — a classical and widely used approach that adjusts p-values in a step-up manner, providing strong control of false discoveries while remaining less conservative than traditional Bonferroni correction.
2. *Storey’s q-value Method* — a more adaptive version of FDR control that estimates the proportion of true null hypotheses (π₀) and thus allows slightly greater sensitivity in detecting subtle activation differences.

* By applying both methods, I could visualize and compare how each technique influenced which voxels were declared significant, revealing the trade-off between conservativeness and discovery power.

*Visualization and Interpretation*

* After correction, binary activation maps were generated to display significant voxels (value = 1) versus non-significant ones (value = 0). These maps were visualized slice by slice, allowing a direct comparison between BH and Storey’s q-value outcomes. Additional comparative maps highlighted overlapping regions (robust activations detected by both methods) and voxels unique to the q-value method.

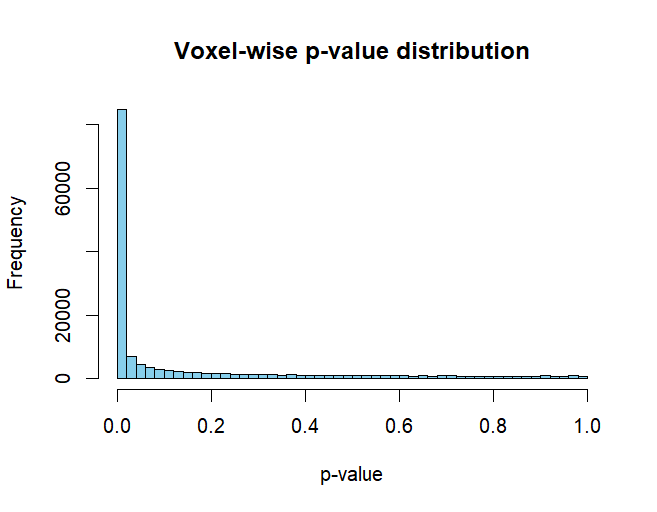
*Exploratory Extension: Knockoff-Based Inference*

* In addition to FDR approaches, I explored the use of knockoff-based inference, a modern statistical framework designed for high-dimensional variable selection. Implementing this method required careful handling of the covariance structure of voxel data, including strategies to make the covariance matrix invertible by adding small values to its diagonal. Although still in exploratory stages, this extension demonstrates an awareness of cutting-edge inference techniques and lays the groundwork for more advanced, model-based feature selection in future neuroimaging research.

# **Results and Discussion**

*Descriptive Overview*

Before applying any statistical correction, the voxel-wise analysis revealed a wide range of p-values across the brain. The p-value distribution plot below showed a strong spike near zero, indicating that a large number of voxels displayed meaningful differences in activation between the *chair* and *scissors* conditions. This initial pattern suggested that certain brain regions were consistently more active under one stimulus compared to the other — a promising sign that the data captured real cognitive variation rather than pure noise.

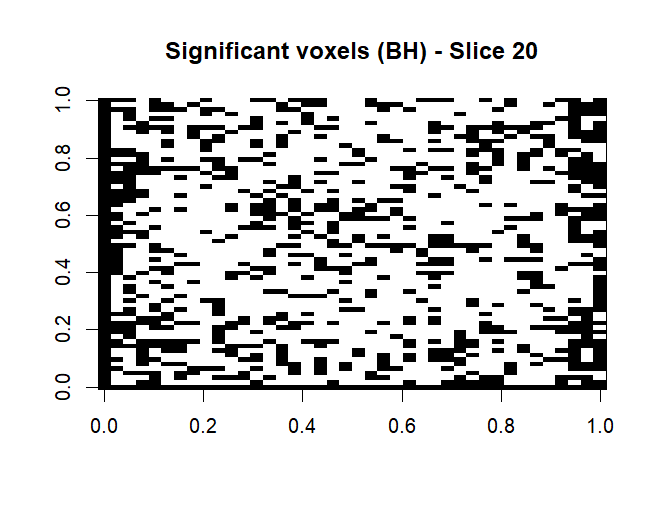


*(Figure 1: Distribution of voxel-wise p-values across the brain — note the heavy skew toward small p-values, reflecting widespread differential activation.)*

However, given the tens of thousands of voxels analyzed, many of these “significant” results could arise by chance. Therefore, applying robust multiple-testing correction was essential to identify genuine neural activity differences.

*Benjamini–Hochberg (BH) Correction*

After applying the Benjamini–Hochberg FDR procedure, a binary activation map was generated for one representative slice of the brain (Slice 20). The white regions represent non-significant voxels, while black pixels indicate those that remained significant after controlling for false discoveries at α = 0.05.

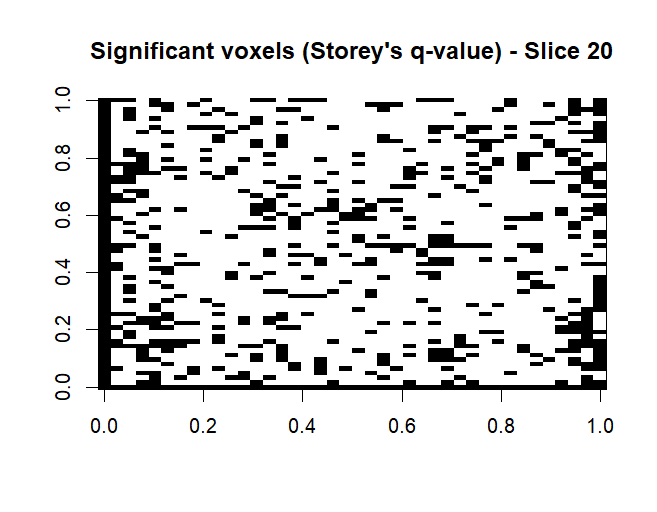


*(Figure 2: Significant voxels under BH correction — slice 20 of the brain.)*

The BH-corrected map displayed scattered pockets of activation, suggesting localized yet reliable neural responses. The sparsity of these clusters reflects the conservative nature of BH correction — only the most robust signals were retained. This pattern implies that specific, focused regions were consistently differentiating between *chair* and *scissors* perception, rather than broad, diffuse activation across the cortex.

*Storey’s q-Value Correction*

Next, Storey’s q-value method was applied to the same set of voxel-wise p-values. The resulting map shows a slightly denser pattern of significant voxels compared to the BH map.

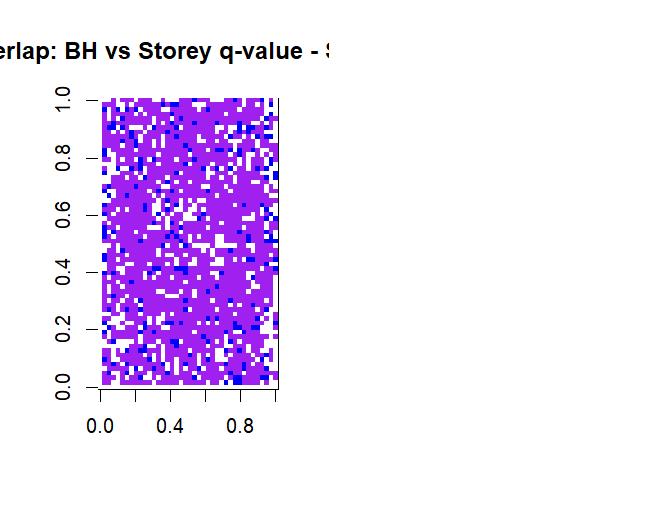


*(Figure 3: Significant voxels detected under Storey’s q-value correction — slice 20.)*

As expected, Storey’s method identified additional voxels that were near the significance threshold under BH. This adaptiveness arises from its estimation of π₀, the proportion of true nulls, which allows it to relax conservativeness when appropriate. These extra detections likely represent weaker but potentially meaningful activations that were missed by BH due to its stricter cutoff.

*Comparison Between Methods*

To examine the overlap between the two FDR procedures, the significant voxel maps were directly compared (Figure 4).



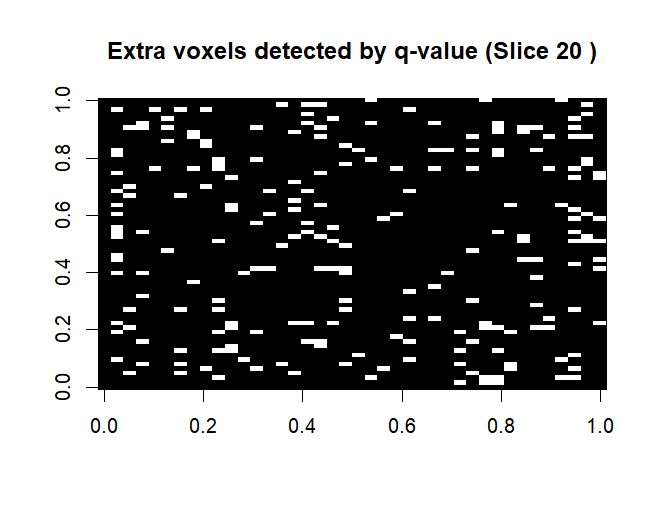
*(Figure 4: Overlap between BH and Storey’s q-value results — purple = common activations, blue = q-value only, red = BH only.)*

The visualization revealed that most voxels identified by BH were also detected by Storey’s q-value, represented by the large purple regions. The presence of blue voxels (unique to the q-value method) highlights Storey’s greater sensitivity. Importantly, the near absence of red voxels confirms that BH did not detect any unique activations — an expected outcome, as Storey’s method is a generalization of BH that only increases detection power while maintaining error control.

This consistency between methods reinforces the reliability of the findings: the core activation regions are robust and reproducible across distinct FDR correction strategies.

*Exploring Additional Discoveries*

To further probe these differences, a comparative map highlighting *extra voxels* detected only by the q-value method was created (Figure 5).



*(Figure 5: Voxels detected only by Storey’s q-value method — additional regions of potential activation.)*

These regions represent subtle activation differences that narrowly missed the BH threshold. While they require cautious interpretation, their spatial coherence across slices suggests that they may not be random artifacts, but rather smaller, functionally relevant signals — perhaps reflecting finer perceptual distinctions or secondary visual processes.

*Summary of Findings*

Across all analyses, both FDR correction methods revealed localized and reproducible patterns of activation differentiating the *chair* and *scissors* conditions. The BH method provided a conservative view of significant regions, while Storey’s q-value offered a more adaptive, nuanced perspective, identifying additional but still statistically controlled activations.

Together, these results demonstrate the delicate balance between statistical rigor and discovery power in neuroimaging inference. They highlight how methodological choices can shape the narrative we construct about brain function — emphasizing that careful statistical reasoning is as crucial as the data itself in making meaningful scientific conclusions.

# **Conclusion and Future Directions**

This project set out to explore how the human brain differentiates between two familiar visual objects — chairs and scissors — using voxel-level analysis on the Haxby 2001 fMRI dataset. By combining classical hypothesis testing with modern False Discovery Rate (FDR) correction methods, I was able to visualize how statistical rigor shapes our understanding of neural activation patterns.

The **Benjamini–Hochberg (BH)** correction provided a conservative, reliable view of the most robust brain activations, while **Storey’s q-value** method revealed additional, subtler signals that BH had missed. Together, these approaches highlighted the delicate balance between avoiding false positives and maintaining the power to detect genuine effects. The consistent overlap between the two methods also confirmed that the detected activation patterns were not random but reproducible — reinforcing confidence in the findings.

Beyond the immediate results, this project deepened my understanding of how **statistical inference drives modern neuroscience**. It showed that discovering meaningful patterns in high-dimensional data requires not just computation, but thoughtful judgment about what counts as “significant.” My exploratory work on **knockoff-based inference** opens the door for future research on adaptive, model-based variable selection — methods that can further improve robustness and interpretability in brain imaging studies.

Looking ahead, expanding this analysis across more object categories or incorporating machine learning–based decoding could reveal even richer insights into visual cognition. In real-world applications, these methods contribute to fields such as **brain–computer interfaces**, **neurological diagnostics**, and **cognitive modeling**, where distinguishing subtle activation differences is key.

In essence, this project was not just about detecting activated voxels — it was about learning how thoughtful statistical design can bring us closer to understanding the brain’s incredible ability to interpret the world.

# **Challenges and Reflections**

Every stage of this project came with its own set of challenges — but each one pushed me to think more deeply about the logic behind statistical inference rather than just applying formulas. One of the first hurdles I faced was **understanding q-values**. Until this project, I had only worked with traditional p-values. Learning how q-values adjust for the proportion of true null hypotheses gave me a new perspective on what “false discovery” really means. It wasn’t just about getting smaller numbers — it was about learning how statistical procedures can adapt to the data itself.

Another major challenge came when I attempted to implement the **Model-X Knockoff Filter** for high-dimensional control. In theory, this method elegantly controls the false discovery rate without directly relying on p-values — a concept that fascinated me. However, applying it to fMRI data where the number of voxels (p) far exceeds the number of observations (n) was computationally demanding. I tried to stabilize the covariance matrix by adding a small regularization term to make it invertible, but the knockoff generation still struggled due to the ultra–high-dimensional setting.

Rather than treating this as a failure, I viewed it as an **entry point into modern research problems** — understanding the limits of current methods and how they might be extended. In future projects, I plan to explore dimensionality reduction strategies and sparse modeling techniques that can make knockoff-based inference feasible for brain imaging data.

Another interesting learning curve was in **visualizing statistical outcomes**. Translating thousands of voxel-wise test results into interpretable brain maps required me to carefully match data indices, slice positions, and activation thresholds. Seeing statistical patterns emerge visually was both rewarding and humbling — a reminder that every plot is the result of countless small steps of reasoning, data cleaning, and validation.

Overall, these challenges didn’t just test my technical ability — they deepened my appreciation for the balance between theory and practice. Each obstacle revealed a layer of complexity in how high-dimensional inference connects mathematics, computation, and neuroscience — the very intersection that continues to inspire me to keep exploring.