

# CSE 599K

## Empirical Research Methods

Winter 2025

SE meets Science

**Analysis validity**

### Today

- Recap: analysis validity
- SE principles for rigorous Science
- Two example studies

### Analysis validity: open questions

#### **External validity**

- Does the experiment generalize (to larger population, other subjects, etc.)?
- How representative is the sample?

#### **Internal validity**

- Does the experiment isolate the variable(s) of interest?
- Does the experiment control for confounders and unwanted effects?

#### **Construct validity**

- Does the experiment measure what it claims to measure?
- Do the proxy measures and tools adequately measure the concept of interest?

#### **(Statistical) conclusion validity**

- Are the conclusions valid based on the chosen statistical test and sample size?
- Are the conclusions valid based on the observed significance (p value)?

## SE principles for rigorous science

Science to practice is not a one-way street!



Let's improve **scientific rigor** with  
**SE principles** and **best practices!**

### Design reviews



Design reviews are common in practice.



Embrace and value **pre-registrations**.



RFCs and public discussions (e.g., GH) provide valuable context.



**Public (open) reviews** should be a no-brainer!

*Hark no more: On the preregistration of chi experiments, Cockburn et al., CHI 2018*  
<https://openreview.net/>

### Quality assurance



Modern code review is incremental (not holistic).



Move to **pre-acceptance artifact evaluations**.



Software testing is the most common QA approach in practice.



Require **evidence for artifact testing**.

*Expectations, outcomes, and challenges of modern code review, Bacchelli and Bird, ICSE 2013*  
*Modern code review: a case study at Google, Sadowski et al., ICSE 2018*

## Process



Merge conflicts (branches) are resolved by branch authors.



Expect **resolution** (knowledge) of **conflicting results**.



Don't expect others to resolve your merge conflict!

## Process



Merge conflicts (branches) are resolved by branch authors.



Expect **resolution** (knowledge) of **conflicting results**.



No premature optimizations.



**Focus on design validity** before scrutinizing artifacts.

## Science is a collaborative effort!



Software Engineering is a **collaborative effort**.  
We should view science the same way!

## Science as Amateur Software Development



1. How can software engineering principles improve the rigor of data analyses?
2. Are these principles equally applicable to computational notebooks?
3. Describe three specific quality control mechanisms.
4. McElreath attributes a significant number of incorrect (scientific) studies to “sloth”. What are the specific issues he is calling out, and what solutions does he propose?
5. Provide an argument for why or why not general-purpose programming languages such as Python are an adequate choice for data analysis.

## Two example studies

### An example study: design

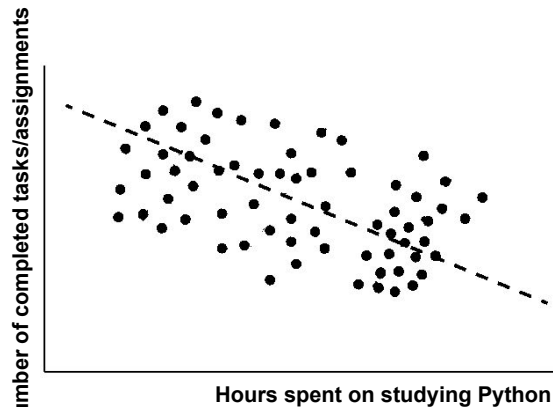
#### Goal:

Studying the **relationship** between **time spent on studying** Python and **success rate** in completing coding assignments.

#### Methodology:

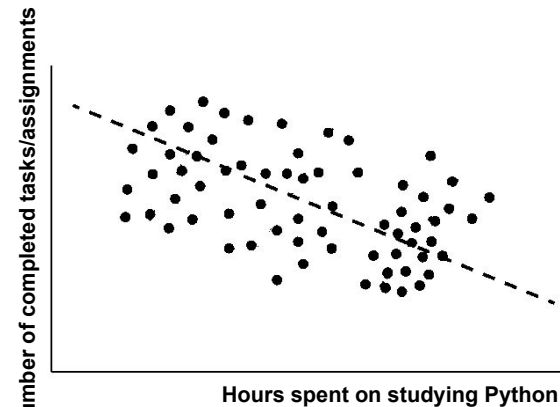
- ~100 participants are randomly selected in front of CSE.
- Each participant is given a high-level overview of the study.
- Each participant decides on how long to study before attempting to solve any coding assignment.
- Each participant solves as many coding assignments as possible in one hour (after studying).

### An example study: conclusions



**Conclusion: Spending more time on learning Python makes you a worse Python programmer.**

### An example study: conclusions

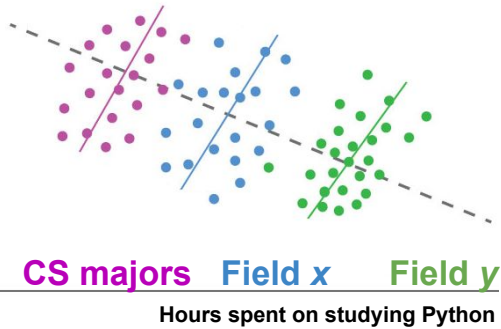


**What may cause this result?**

**Conclusion: Spending more time on learning Python makes you a worse Python programmer.**

# An example study: Simpson's paradox

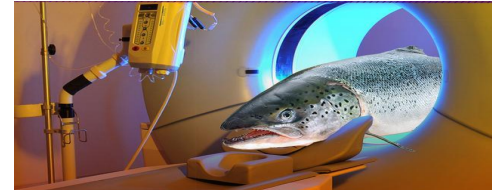
Number of completed tasks/assignments



Where did this study fail?

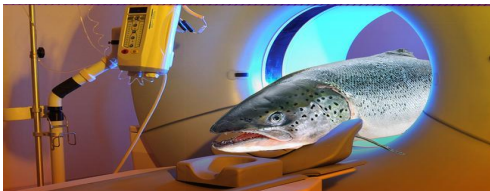
This phenomenon is called: **Simpson's paradox.**

# Another example study



<http://www.prefrontal.org/files/posters/Bennett-Salmon-2009.pdf>

# Another example study: design



**Neural correlates of interspersed perspective taking in the post-mortem Atlantic Salmon: An argument for multiple-comparisons correction**  
 Craig M. Bennett<sup>1</sup>, Aigal A. Bais<sup>1</sup>, Michael B. Miller<sup>1</sup>, and George L. Wolford<sup>1</sup>  
<sup>1</sup>Department of Psychology, University of Regina, Saskatchewan, Canada; <sup>2</sup>Department of Psychology, University of Regina, Saskatchewan, Canada

**INTRODUCTION**  
 With the recent demonstration of functional neuroimaging data values corresponding to brain activation, it is now possible to examine brain activity in response to stimuli. However, the analysis of such data is complex, and multiple comparisons should be considered with these data, but it is often difficult to know how to proceed. To illustrate the magnitude of this problem, we used an fMRI dataset that demonstrates the danger of not controlling for multiple comparisons.

**METHODS**  
 Subjects, the seven Atlantic Salmon, were scanned in an fMRI unit. The subjects were approximately 18 inches long, weighed 3.8 lbs, and were approximately 18 months old. The subjects were scanned in an fMRI unit. The subjects were approximately 18 inches long, weighed 3.8 lbs, and were approximately 18 months old. The subjects were scanned in an fMRI unit. The subjects were approximately 18 inches long, weighed 3.8 lbs, and were approximately 18 months old.

**RESULTS**  
 A cluster was found in the anterior cingulate cortex, an area associated with social cognition. This cluster was found in the anterior cingulate cortex, an area associated with social cognition. This cluster was found in the anterior cingulate cortex, an area associated with social cognition.

**VOXELWISE VARIABILITY**  
 The observed variability of signal intensity within the group was examined. The observed variability of signal intensity within the group was examined. The observed variability of signal intensity within the group was examined.

**DISCUSSION**  
 Our results show that the data do not support the hypothesis that the anterior cingulate cortex is involved in social cognition. Our results show that the data do not support the hypothesis that the anterior cingulate cortex is involved in social cognition.

**REFERENCES**  
 Bennett, C.M., Bais, A.A., Miller, M.B., and Wolford, G.L. (2009). Neural correlates of interspersed perspective taking in the post-mortem Atlantic Salmon: An argument for multiple-comparisons correction. *Journal of Cognitive Neuroscience*, 21(12), 2100-2110.

# Another example study: design

**Subject:** One mature Atlantic Salmon (*Salmo salar*) participated in the **fMRI study**. The salmon was approximately 18 inches long, weighed 3.8 lbs, and was **not alive at the time of scanning**.

**Neural correlates of interspersed perspective taking in the post-mortem Atlantic Salmon: An argument for multiple-comparisons correction**  
 Craig M. Bennett<sup>1</sup>, Aigal A. Bais<sup>1</sup>, Michael B. Miller<sup>1</sup>, and George L. Wolford<sup>1</sup>  
<sup>1</sup>Department of Psychology, University of Regina, Saskatchewan, Canada; <sup>2</sup>Department of Psychology, University of Regina, Saskatchewan, Canada

**INTRODUCTION**  
 With the recent demonstration of functional neuroimaging data values corresponding to brain activation, it is now possible to examine brain activity in response to stimuli. However, the analysis of such data is complex, and multiple comparisons should be considered with these data, but it is often difficult to know how to proceed. To illustrate the magnitude of this problem, we used an fMRI dataset that demonstrates the danger of not controlling for multiple comparisons.

**METHODS**  
 Subjects, the seven Atlantic Salmon, were scanned in an fMRI unit. The subjects were approximately 18 inches long, weighed 3.8 lbs, and were approximately 18 months old. The subjects were scanned in an fMRI unit. The subjects were approximately 18 inches long, weighed 3.8 lbs, and were approximately 18 months old.

**RESULTS**  
 A cluster was found in the anterior cingulate cortex, an area associated with social cognition. This cluster was found in the anterior cingulate cortex, an area associated with social cognition. This cluster was found in the anterior cingulate cortex, an area associated with social cognition.

**VOXELWISE VARIABILITY**  
 The observed variability of signal intensity within the group was examined. The observed variability of signal intensity within the group was examined. The observed variability of signal intensity within the group was examined.

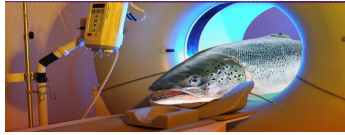
**DISCUSSION**  
 Our results show that the data do not support the hypothesis that the anterior cingulate cortex is involved in social cognition. Our results show that the data do not support the hypothesis that the anterior cingulate cortex is involved in social cognition.

**REFERENCES**  
 Bennett, C.M., Bais, A.A., Miller, M.B., and Wolford, G.L. (2009). Neural correlates of interspersed perspective taking in the post-mortem Atlantic Salmon: An argument for multiple-comparisons correction. *Journal of Cognitive Neuroscience*, 21(12), 2100-2110.

# Another example study: design

**Subject:** One mature **Atlantic Salmon** (*Salmo salar*) participated in the **fMRI study**. The salmon was approximately 18 inches long, weighed 3.8 lbs, and was **not alive at the time of scanning**.

**Task:** [...] open-ended mentalizing task. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.



**Neural correlates of interspersed perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction**  
Craig M. Berrnet<sup>1</sup>, Algalp A. Bano<sup>1</sup>, Michael B. Miller<sup>1</sup>, and George L. Wulfeff<sup>1</sup>  
<sup>1</sup>Department of Psychology & Behavioral Science, Simon Fraser University, Burnaby, BC, Canada

**INTRODUCTION**  
With the recent demonstration of functional neuroimaging data across various tasks in the post-mortem Atlantic Salmon, we have begun to explore the possibility of a fish possessing a visual cortex. However, the multiple comparisons correction used in the original analysis of the data was not a valid correction that demonstrates the degree of our confidence in the data.

**METHODS**  
Salmon, the mature Atlantic Salmon (Salmo salar) participated in an fMRI study. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.

**RESULTS**  
A cluster was used to test for regions with significant BOLD signal change during the fish's viewing of the photos. The parameters for the analysis were:  $T(10) = 3.15$ ,  $p < 0.001$ ,  $k = 10$ ,  $T(10) = 3.15$ ,  $p < 0.001$ ,  $k = 10$ .

**DISCUSSION**  
Our results show that the fish, when shown a series of photographs depicting human individuals in social situations with a specified emotional valence, showed significant BOLD signal change in the visual cortex. This result is consistent with the idea that the fish has a visual cortex.

**REFERENCES**  
Berrnet, C. M., Bano, A. A., Miller, M. B., & Wulfeff, G. L. (2018). Neural correlates of interspersed perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. *Journal of Experimental Psychology: Applied*, 24(1), 1-10.

# Another example study: conclusions

**Subject:** One mature **Atlantic Salmon** (*Salmo salar*) participated in the **fMRI study**. The salmon was approximately 18 inches long, weighed 3.8 lbs, and was **not alive at the time of scanning**.

**Task:** [...] open-ended mentalizing task. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.

**Results:** Several active voxels were discovered [...] Out of a search volume of 8064 voxels a total of **16 voxels** were significant.

**Neural correlates of interspersed perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction**  
Craig M. Berrnet<sup>1</sup>, Algalp A. Bano<sup>1</sup>, Michael B. Miller<sup>1</sup>, and George L. Wulfeff<sup>1</sup>  
<sup>1</sup>Department of Psychology & Behavioral Science, Simon Fraser University, Burnaby, BC, Canada

**INTRODUCTION**  
With the recent demonstration of functional neuroimaging data across various tasks in the post-mortem Atlantic Salmon, we have begun to explore the possibility of a fish possessing a visual cortex. However, the multiple comparisons correction used in the original analysis of the data was not a valid correction that demonstrates the degree of our confidence in the data.

**METHODS**  
Salmon, the mature Atlantic Salmon (Salmo salar) participated in an fMRI study. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.

**RESULTS**  
A cluster was used to test for regions with significant BOLD signal change during the fish's viewing of the photos. The parameters for the analysis were:  $T(10) = 3.15$ ,  $p < 0.001$ ,  $k = 10$ ,  $T(10) = 3.15$ ,  $p < 0.001$ ,  $k = 10$ .

**DISCUSSION**  
Our results show that the fish, when shown a series of photographs depicting human individuals in social situations with a specified emotional valence, showed significant BOLD signal change in the visual cortex. This result is consistent with the idea that the fish has a visual cortex.

**REFERENCES**  
Berrnet, C. M., Bano, A. A., Miller, M. B., & Wulfeff, G. L. (2018). Neural correlates of interspersed perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. *Journal of Experimental Psychology: Applied*, 24(1), 1-10.

# Another example study: conclusions

## Interpretation of pure noise

- Noisy data source
- Multiple hypotheses tested on the same data
- An argument for multiple comparisons correction

- Analysis grounded in a **conceptual model?**
- Clear **operationalization (implementation)?**
- Implementation consistent with the model?**
- Proper use of statistical methods?**
- Data interpreted in **context of prior knowledge?**
- Explored and validated **alternative hypotheses?**



**Neural correlates of interspersed perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction**  
Craig M. Berrnet<sup>1</sup>, Algalp A. Bano<sup>1</sup>, Michael B. Miller<sup>1</sup>, and George L. Wulfeff<sup>1</sup>  
<sup>1</sup>Department of Psychology & Behavioral Science, Simon Fraser University, Burnaby, BC, Canada

**INTRODUCTION**  
With the recent demonstration of functional neuroimaging data across various tasks in the post-mortem Atlantic Salmon, we have begun to explore the possibility of a fish possessing a visual cortex. However, the multiple comparisons correction used in the original analysis of the data was not a valid correction that demonstrates the degree of our confidence in the data.

**METHODS**  
Salmon, the mature Atlantic Salmon (Salmo salar) participated in an fMRI study. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.

**RESULTS**  
A cluster was used to test for regions with significant BOLD signal change during the fish's viewing of the photos. The parameters for the analysis were:  $T(10) = 3.15$ ,  $p < 0.001$ ,  $k = 10$ ,  $T(10) = 3.15$ ,  $p < 0.001$ ,  $k = 10$ .

**DISCUSSION**  
Our results show that the fish, when shown a series of photographs depicting human individuals in social situations with a specified emotional valence, showed significant BOLD signal change in the visual cortex. This result is consistent with the idea that the fish has a visual cortex.

**REFERENCES**  
Berrnet, C. M., Bano, A. A., Miller, M. B., & Wulfeff, G. L. (2018). Neural correlates of interspersed perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. *Journal of Experimental Psychology: Applied*, 24(1), 1-10.

## Interpretation of pure noise

- Noisy data source
- Multiple hypotheses tested on the same data
- An argument for multiple comparisons correction

- Analysis grounded in a **conceptual model?**
- Clear **operationalization (implementation)?**
- Implementation consistent with the model?**
- Proper use of statistical methods?**
- Data interpreted in **context of prior knowledge?**
- Explored and validated **alternative hypotheses?**



**Neural correlates of interspersed perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction**  
Craig M. Berrnet<sup>1</sup>, Algalp A. Bano<sup>1</sup>, Michael B. Miller<sup>1</sup>, and George L. Wulfeff<sup>1</sup>  
<sup>1</sup>Department of Psychology & Behavioral Science, Simon Fraser University, Burnaby, BC, Canada

**INTRODUCTION**  
With the recent demonstration of functional neuroimaging data across various tasks in the post-mortem Atlantic Salmon, we have begun to explore the possibility of a fish possessing a visual cortex. However, the multiple comparisons correction used in the original analysis of the data was not a valid correction that demonstrates the degree of our confidence in the data.

**METHODS**  
Salmon, the mature Atlantic Salmon (Salmo salar) participated in an fMRI study. The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.

**RESULTS**  
A cluster was used to test for regions with significant BOLD signal change during the fish's viewing of the photos. The parameters for the analysis were:  $T(10) = 3.15$ ,  $p < 0.001$ ,  $k = 10$ ,  $T(10) = 3.15$ ,  $p < 0.001$ ,  $k = 10$ .

**DISCUSSION**  
Our results show that the fish, when shown a series of photographs depicting human individuals in social situations with a specified emotional valence, showed significant BOLD signal change in the visual cortex. This result is consistent with the idea that the fish has a visual cortex.

**REFERENCES**  
Berrnet, C. M., Bano, A. A., Miller, M. B., & Wulfeff, G. L. (2018). Neural correlates of interspersed perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. *Journal of Experimental Psychology: Applied*, 24(1), 1-10.

Where did this study fail (on purpose)?

Valid data analysis goes well beyond implementation correctness.