Statistical Fallacies in Neuroscience

Mark Reimers, PhD
Neuroscience Program,
Biomedical Engineering
Michigan State University

Cold Spring Harbor Course on Neural Data Science 2019

What is a Statistical Fallacy?

Common practice or thinking that leads to false conclusions in some real situations

What is a Statistical Fallacy?

Common practice or thinking that leads to false conclusions in some real situations:

- "I just tested 500 mouse V2 neurons for orientation tuning and found statistically significant ($p<0.05$) results for 50 of them. So the proportion of V2 neurons showing orientation tuning is about 10%"

What is a Statistical Fallacy?

Common practice or thinking that leads to false conclusions in some real situations:

- "The new *Nature* paper finds an association between a SNP in CACNA1 and schizophrenia risk at $p<10^{-10}$. Wow. I'm going to do my thesis on exploring the role of voltage-gated calcium in mental illness."

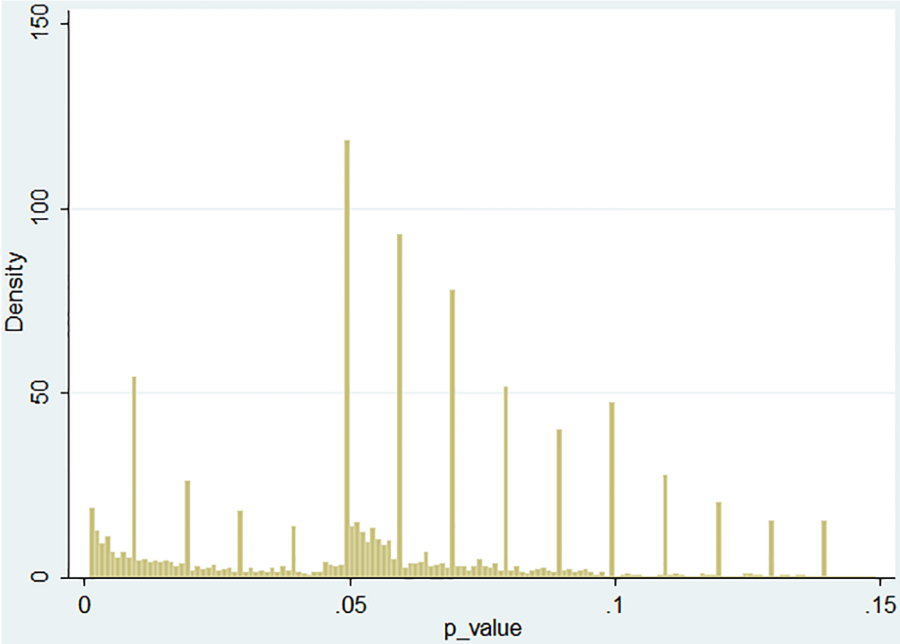
What is a Statistical Fallacy?

Common practice or thinking that leads to false conclusions in realistic situations:

- "We were struggling to show the statistical significance of our results, but when we dropped the outliers, eliminated the samples taken during the heat wave and used the Spearman correlation rather than the Pearson correlation we finally got $p < .02$, well into significance, and we were able to publish."

P-Hacking (Data Dredging)

- For studies which are under-powered (almost all social science studies), the temptation to adjust things 'just a bit' to publish is overwhelming



Krawczyk (2015) gathered p-values from published papers in experimental psychology.

Here the distribution is weighted so all papers contribute equally

Correlation Fallacies

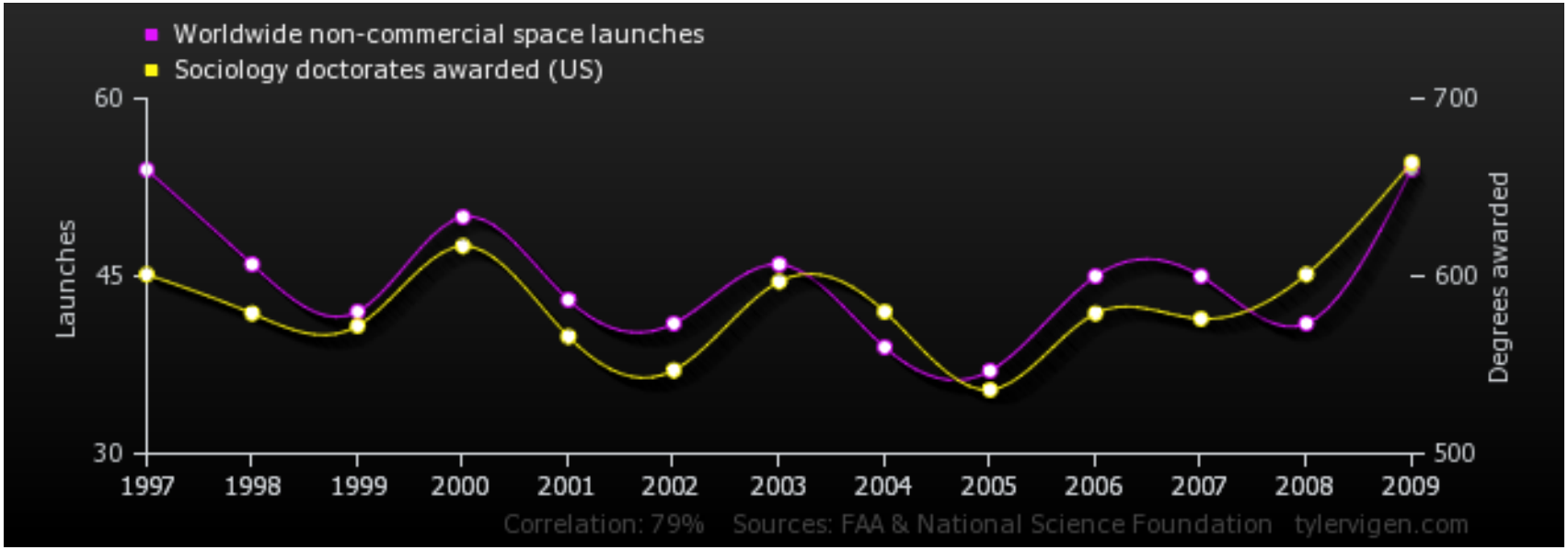
Correlation summarizes most information about the relationship between measures

- “If two variables are highly correlated they must be related somehow”

Correlation Fallacies

Correlation summarizes most information about the relationship between measures

- "If two variables are highly correlated they must be related somehow, even if through mediators"



Correlation Fallacies

Correlation summarizes most information about the relationship between measures

- "We found a strong association (r = .5) between spikes of a key neuron in CA1 and performance on this memory task"

Correlation Fallacies

Correlation summarizes most information about the relationship between measures

- "Correlation is not causation – but the correlation tells me most of what I need to know about the descriptive relation between variables"

Principal Components Analysis Fallacy

- “We identified a strong correlation between PCs 10, and 13 of neural population activity with the complexity of the stimulus”

Principal Components Analysis Fallacy

- “We identified a strong correlation between PCs 10, and 13 of neural population activity with the complexity of the stimulus”

- PCA simply can’t estimate accurately the true directions of distinct components further out than the ‘knee’ of the scree plot.

Randomization Fallacies

"When trying to assess the significance of a complex statistical procedure, where there's no analytic solution, repeating the procedure with randomized data is a tedious but always reliable way to get a comparison (null distribution)."

Randomization Fallacies

"When trying to assess the significance of a complex statistical procedure, where there's no analytic solution, repeating the procedure with randomized data is a tedious but always reliable way to get a comparison (null distribution)."

Yes, randomization often gives you the complete Null (no relationships), but does not give you a comparison data set with all relationships except the one you try to show.

The problem of finding an appropriate null for complex statistical procedures is very hard

Structure in neural population recordings: an expected byproduct of simpler phenomena?
Elsayed & Cunningham (2017)