

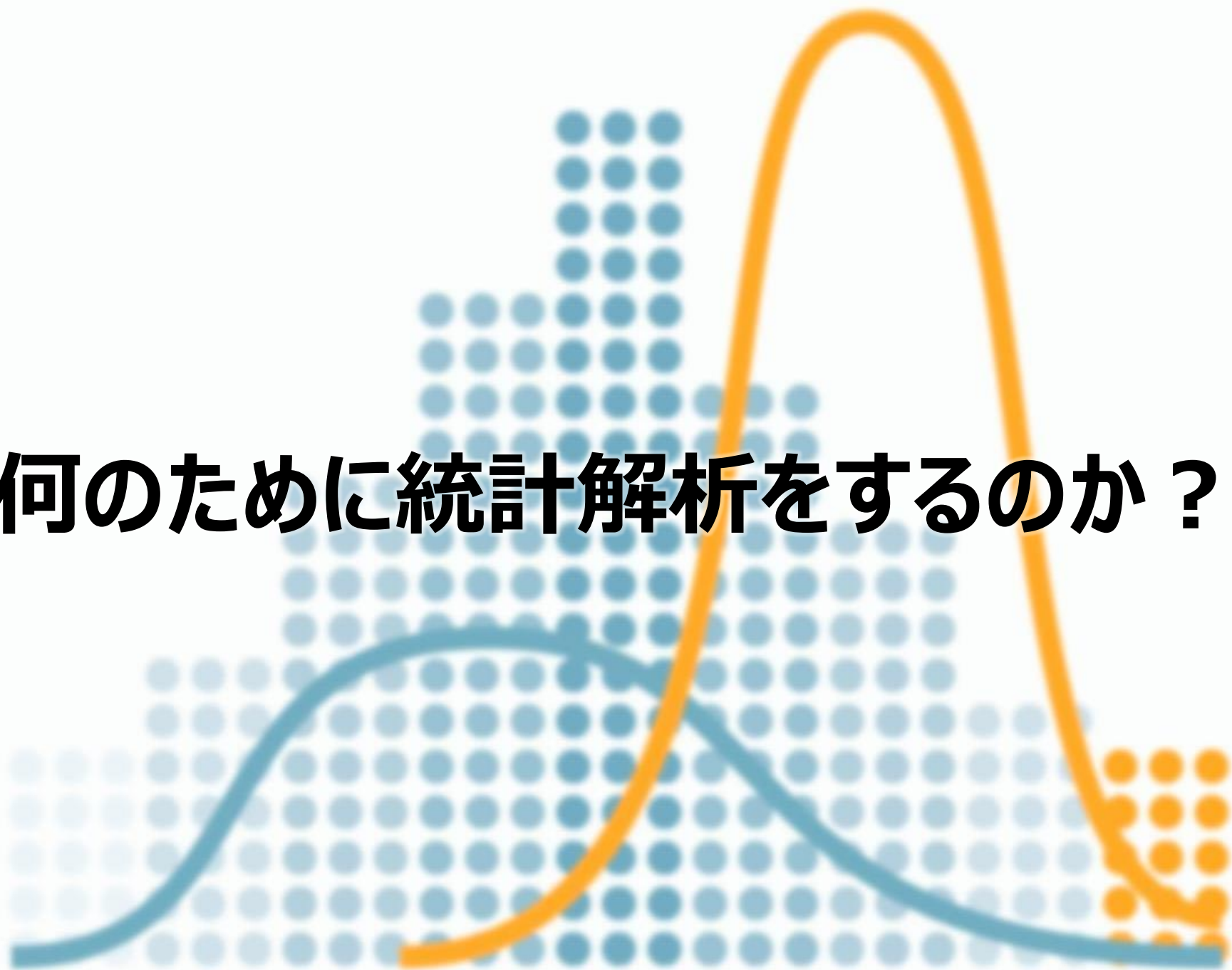
統計演習 (9)

Intermission

高谷 智英
t-takaya.net

Dec 5, 2018

何のために統計解析をするのか？





統計的有意 ≠ 生物学的に意味があること

山中：ほかに何か，統計学の専門家からみて，みのがしている誤解などあれば教えてください。

佐藤：よくあるのは検定の偏重です．有意差がでたらよかった，というのがあって，とくに生物医学分野での共通の誤解といったら変ですけども，検定の結果って，多くの対象数を集めたら絶対に有意になるんですね．身長が1ミリ違うというのも，1000人ずつくらい対象者を集めてきたら絶対に有意差がでますから．**統計的に非常に小さいp値になったということと，臨床的もしくは生物学的にそれが意味のあることだというのは別なんです．**



何のために統計解析をするのか？

統計解析は、実験結果（事実; facts）をどのように解釈するか、その示唆を提供する技術である。

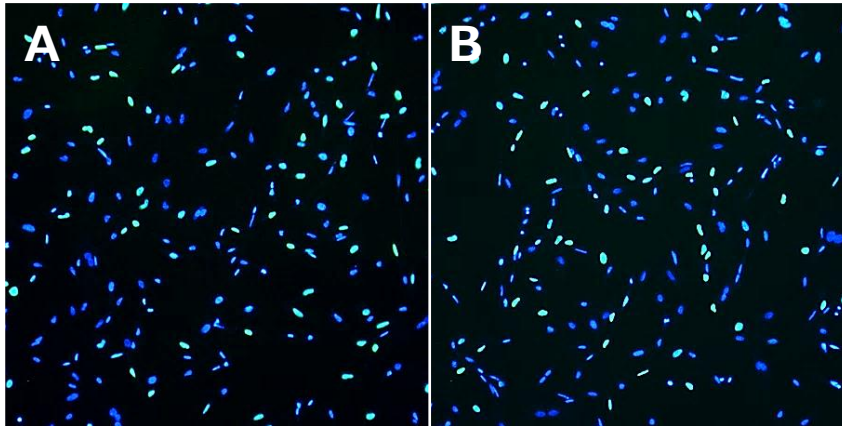
データ・情報の解釈に、客観的な判断基準を提供することで、他者との議論が可能になる。

ただし、統計解析は、真実（truth）を判定するものではない。

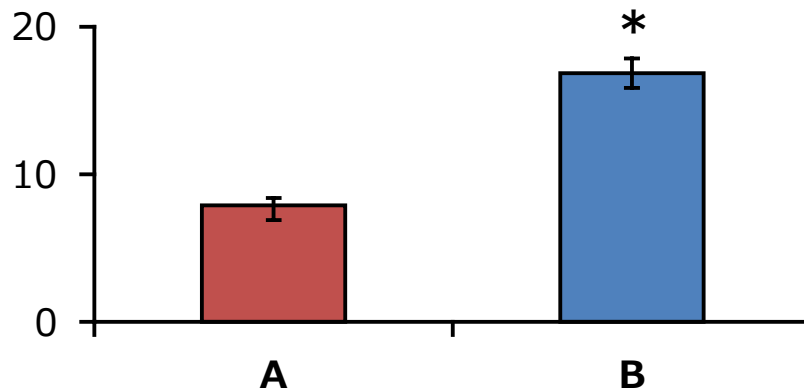


何のために統計解析をするのか？

実験結果



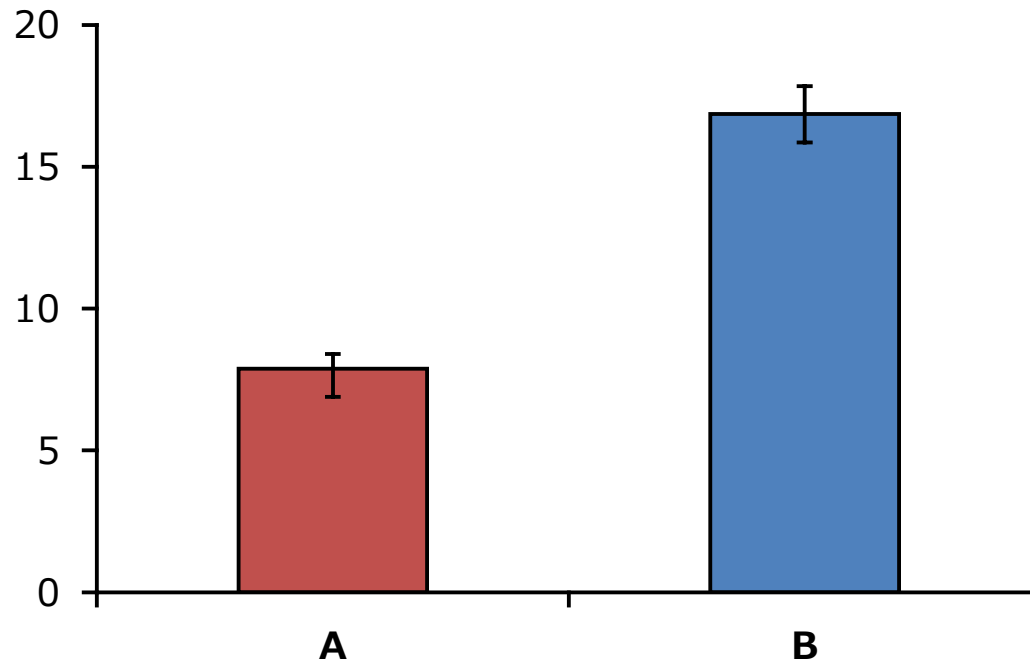
統計解析 (*, $p < 0.05$)





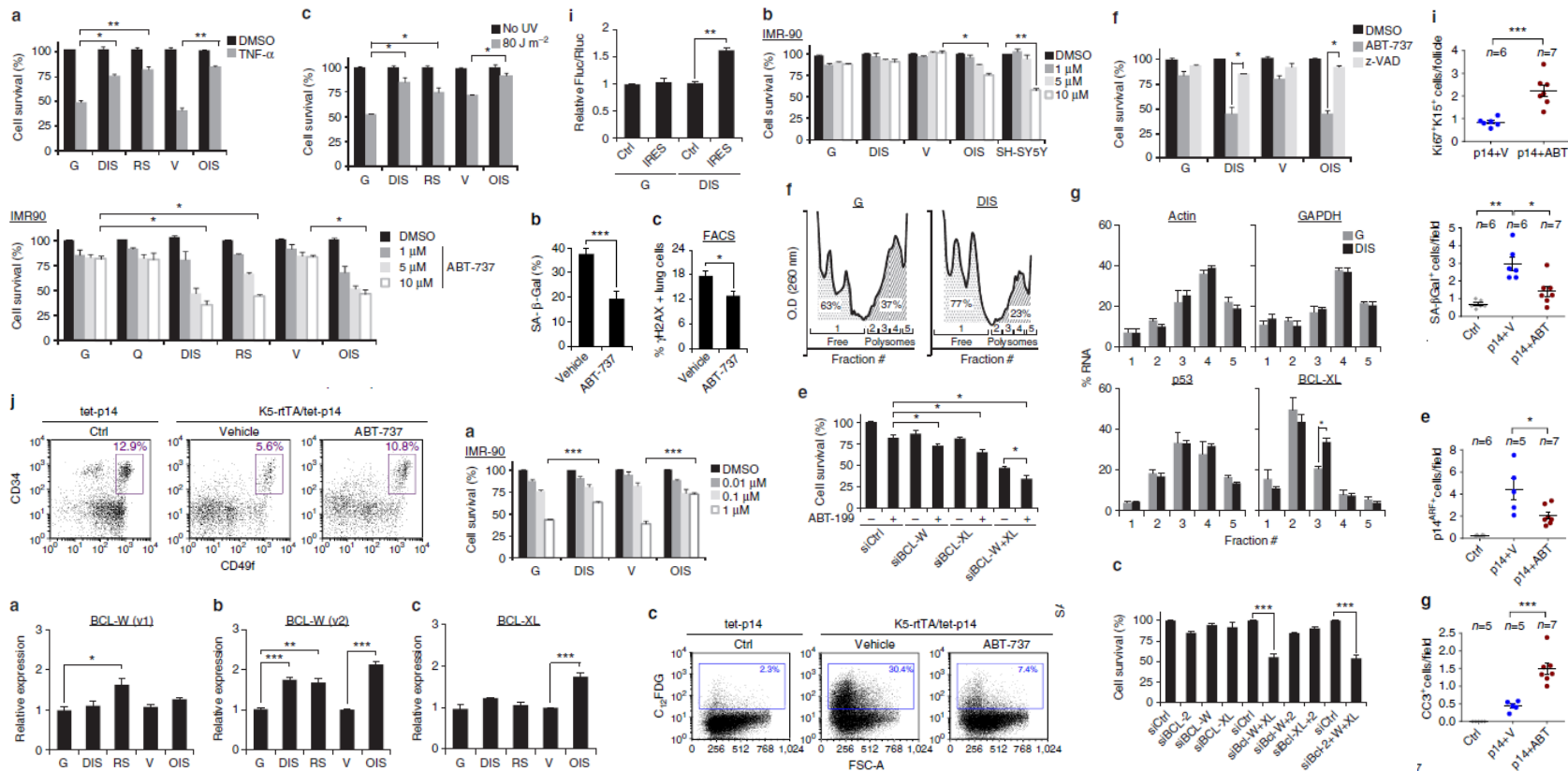
有意であるとは何か？

$p = 0.05$ とは？





現代の論文は膨大な量のデータから構成される



有意差あり、と明示されているだけでも40か所！



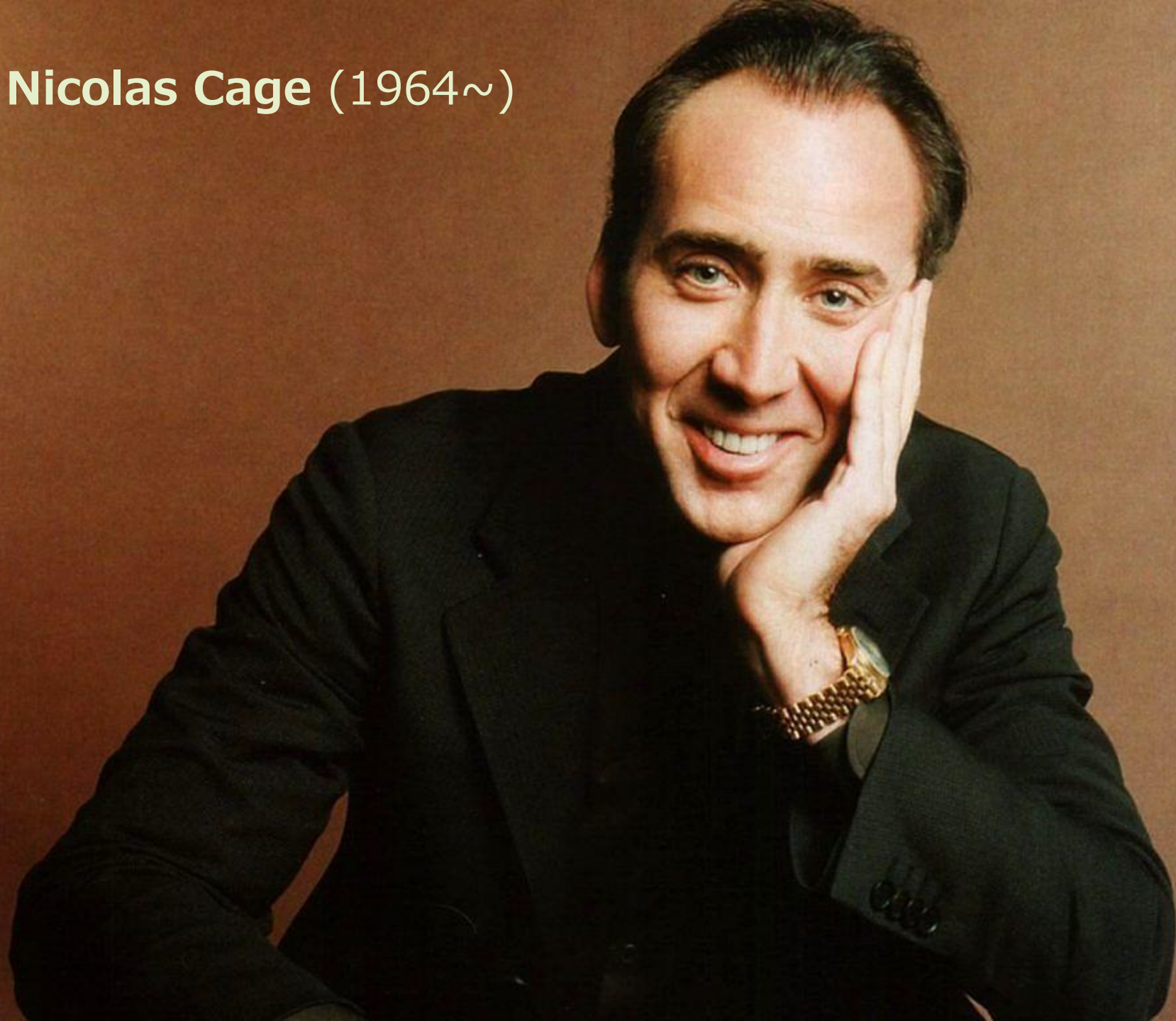
第1問

論文中に「有意差あり」とされるデータが多いほど、その論文は、「たまたまそうなった」データを含んでいる可能性が高くなる。

「たまたまそうなった」データを含む論文の結論は、正しいか否か。あるいはどのように考えるのか。自分の考えと、その理由を書きなさい。



Nicolas Cage (1964~)

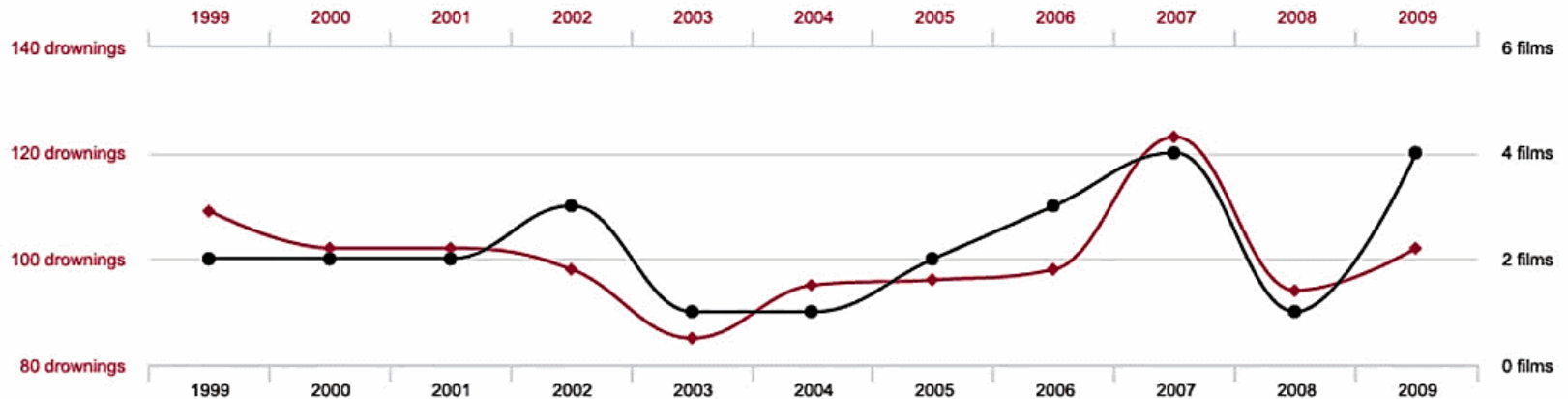




プールで溺死する人数/年と、ニコラス・ケイジの映画出演数/年



66.6%
CORRELATION



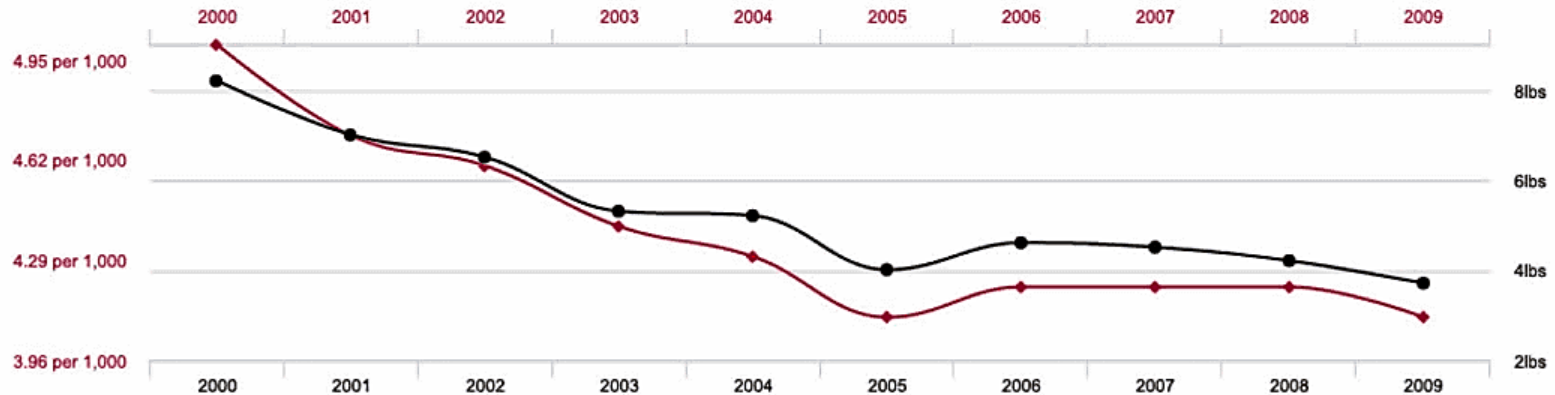
Source: tylervigen.com/spurious-correlations



メイン州の離婚率と、マーガリン消費量/人/年



99.26%
CORRELATION



Source: tylervigen.com/spurious-correlations



宇宙科学技術費と、自殺率

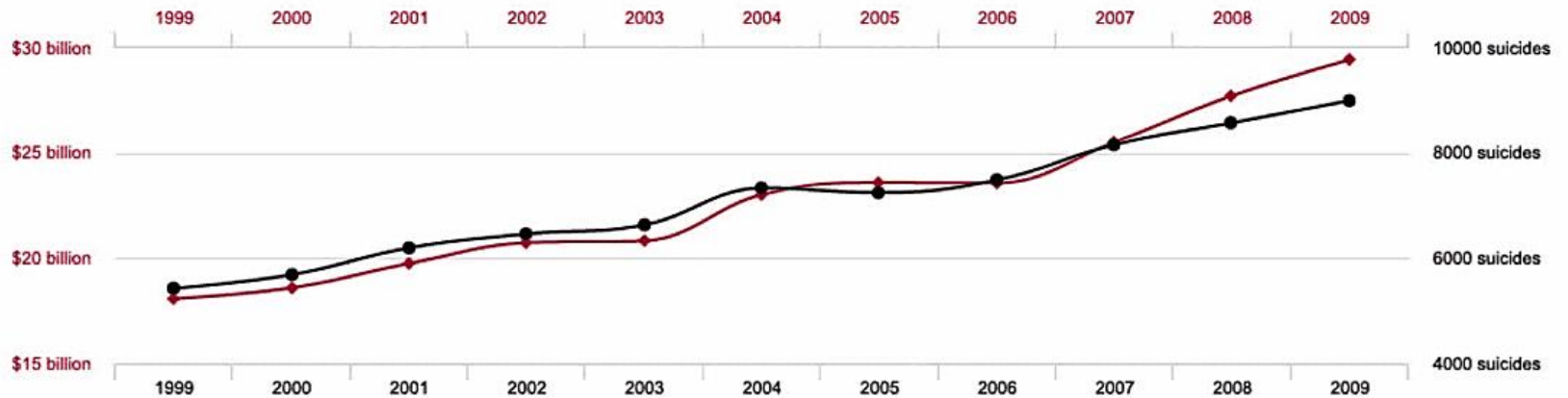


SPENDING
ON
STEM

99.79%
CORRELATION!



SUICIDE



Source: tylervigen.com/spurious-correlations



第2問

プールでの溺死数とニコラス・ケイジの映画数 (相関: 66.6%)

メイン州の離婚率とマーガリン消費量 (相関: 99.26%)

宇宙科学技術費と自殺率 (相関: 99.79%)

の間には、統計学的に極めて有意な相関が存在した。

両者に因果関係 (プールで溺れる者が増えるとニコラス・ケイジの出演が増える。あるいはその逆) は存在するか？

自分の考えと、その理由を書きなさい。