Posted In

Fun Statistics
Statistics Help
Statistics in the News

7 Deadly Statistical Sins Even the Experts Make

Do you know how to avoid them?

Get the facts >

The Stats Cat on Sample Size, Statistical Power, and the Revenge of the Zombie Salmon

Stats Cat · 19 October, 2012

[Twitter] 15   [Facebook] 16   [LinkedIn] 0   [Comment] 0

Marlowe the Stats Cat here. That guy I share my house with left his laptop unattended again, and I spent the evening searching the web for news about one of my favorite subjects: *salmon*. Yum. But I wound up getting more than a collection of cool salmon pictures...I also got a better understanding of the role the size of a dataset plays when you're doing a hypothesis test.

You see, my search led me to this paper that summarized a 2009 analysis of neuroimaging data collected from a frozen salmon. Yes, you did read that correctly: some people with Ph.D.'s actually ran an MRI on a dead fish! Here's how the study's authors describe their subject:

> The salmon measured approximately 18 inches long, weighed 3.8 pounds, and was not alive at the time of scanning. It is not known if the salmon was male or female, but given the post-mortem state of the subject this was not thought to be a critical variable.

The researchers put the dead salmon in a functional magnetic resonance imaging (fMRI) machine and showed it a series of emotionally charged photographs, in the spirit of experiments commonly done with human subjects. Then they used statistical analysis methods to explore the fMRI data they collected.

Their analysis showed evidence of brain activity. Got that? Brain activity...in the *dead, frozen salmon*.

Uh-oh, I thought. If dead salmon *can think*, could they also become reanimated? The prospect of zombie salmon horrified me...for a panicky moment, I felt like I was living in a lunatic mashup of *Night of the Living Dead* and *Finding Nemo*.

But as I continued reading, I realized the subject of this study wasn't *really* the dead salmon, but statistical analysis itself. The researchers cleverly demonstrated the importance of carefully applying the proper statistical methods, especially given the sheer volume of data available to, well, *everyone* today.

Essentially, in this study the amount of data being analyzed made the detection of "brain activity" in the dead salmon almost inevitable, unless the analysis included appropriate corrections. With corrections, the "brain activity" would be rightly filtered out as noise in the data.

Without the corrections, it would be all too easy to conclude that salmon were coming back from the grave.

How Sample Size Affects Statistical Significance

Being a cat, of course I found it very easy to grasp everything the paper was saying. But in an interview with BoingBoing, Dr. Abigail Baird broke it down so that even that human who shares my house could get it:

> ...there's a certain cutoff level where we consider the things we've found significant or not. The gold standard is .01, less than a 1% chance that you're seeing something just by accident. Or a 99% chance that it's an actual difference. But, still, 1 out of 100 times you'd get that exact same result just by chance. ...we'd never had any tools that produced the magnitude of data that fMRI has. Instead of making comparisons between two groups of 40 people, you're making comparisons between 100,000 points in the brain and that .01 no longer says as much because you have so much more information to work with.

Now, their study of dead salmon dealt with specialized statistical issues involved in analyzing fMRI results, but whether you're a research scientist, a quality improvement engineer, or just a cat like me who likes to make decisions using data, the take-home lessons are the same:

1. Make sure you trust your data.

2. Remember that an overabundance of data can lead to statistically significant results that don't have any practical significance.

3. Be sure you're analyzing *your* data—and interpreting that analysis—appropriately.

An Example of Irrelevant Statistical Significance

Let's apply this to an everyday situation. I like to entertain, so I usually keep 15 containers each of KittyLuv salmon-flavored cat treats and GroovyCat tuna-flavored treats on hand. Each container is supposed to weigh 5 oz. I weighed the ones in my pantry now and recorded the data in Minitab, then ran a 2-sample t-test to see if there's a significant difference in the average weight between the salmon and the tuna treats. Here's the data I used, and here are the results of the t-test:

```
Two-sample T for KittyLuv vs GroovyCat

          N   Mean  StDev  SE Mean
KittyLuv  15  4.966  0.693    0.18
GroovyCat 15  5.004  0.709    0.18


Difference = mu (KittyLuv) - mu (GroovyCat)
Estimate for difference:  -0.038
95% CI for difference:  (-0.564, 0.487)
T-Test of difference = 0.009 (vs not =): T-Value = -0.18  P-Value = 0.855  DF =
       27
```

The means and standard deviations for both brands are nearly the same, with an estimated difference of about .004. Nothing worth mewing about, right? And if we look at the p-value of .855, we can reject the null hypothesis that there's a significant difference between the weights of these two types of treats.

But what if I have *200,000* packages of each, sampled from that same distribution of cat treats? Let's run the analysis again...

```
Two-sample T for Kitty Luv vs Groovy Cat

           N      Mean  StDev  SE Mean
Kitty Luv  200000  5.001  0.700  0.0016
Groovy Cat 200000  4.999  0.650  0.0015


Difference = mu (Kitty Luv) - mu (Groovy Cat)
Estimate for difference:  0.00217
95% CI for difference:  (-0.00202, 0.00636)
T-Test of difference = 0.009 (vs not =): T-Value = -3.19  P-Value = 0.001  DF =
       397827
```

The means and standard deviations are virtually identical to our smaller samples, and the estimated difference is even less than that predicted from our previous test at .002. But take a gander at that p-value: at 0.001, it's telling us that there *is* a statistically significant difference in the average weight of KittyLuv and GroovyCat treats.

And there's the lesson of the zombie salmon in a nutshell. Even though we can detect a statistically significant difference in a sample of 200,000 packages, there's no excuse to call up the GroovyCat company and throw a hissy fit. I mean, I have a pretty discriminating palette when it comes to kitty treats, but even *I'm* not going to notice a .002 oz difference in size.

So take it from a dead salmon, the Ph.D.'s who scanned its brain, and me: statistical significance does not equal practical significance.

You Might Also Like:

Everyone's Talking "Big Data"...But Size Isn't What Matters
A Look at Pet Ownership Numbers with the Stats Cat
Gummi Bear DOE: Replicates and Center Points, Part 2
10 Statistical Terms Designed to Confuse Non-Statisticians

Comments

Authors

Carly Barry
Patrick Runkel
Kevin Rudy
Jim Frost
Greg Fox
Eric Heckman
Dawn Keller
Eston Martz
Bruno Scibilia
Eduardo Santiago
Cody Steele