AGAINST ALL ODDS
EPISODE 30 – "INFERENCE FOR REGRESSION"
TRANSCRIPT

FUNDER CREDITS

Funding for this program is provided by Annenberg Learner.

INTRO

Pardis Sabeti
Hello, I'm Pardis Sabeti and this is *Against All Odds*, where we make statistics count.

In our module on Fitting Lines to Data, we examined the relationship between winter snowpack and spring runoff. Colorado resource managers make predictions about the seasonal water supply using a least squares regression line through the scatterplot of their measurement data.

But would we really see a linear relationship between snowpack and runoff if we had all the possible data, or might the pattern we see in the sample data scatterplot occur just by chance? We'd like to know whether the positive association we see between snowpack and runoff in the sample is strong enough to conclude that the same relationship holds for the whole population.

Statisticians rely on inference to determine whether the relationship observed between two variables in a single scatterplot is valid for some larger population.

Inference is a powerful tool – powerful enough, in fact, to help bring an entire bird species back from the brink of extinction.

After World War II, the agrichemical industry began mass-producing chemicals to control pests.

> **1940's Announcer**
> With the possibility of a serious infantile paralysis epidemic, health authorities of the city of San Antonio, Texas attacked the germ carriers throughout the city. With the war-discovered DDT and special sprayers, sections of the city are blanketed with the insecticide…

Pardis Sabeti
Unfortunately, there weren't many safeguards or screening processes in place, and the damaging environmental effects of these compounds weren't taken into account.

Eventually, changes in the natural environment due to chemical pesticides became apparent. One species that was severely affected was the peregrine falcon.

> **Carl Thelander**
> In Great Britain in particular, Derek Ratcliffe, in the 1950s noticed that peregrine falcons were declining at the nesting sites and they were unable to hatch their eggs. The decline of peregrines was eventually

demonstrated to be a worldwide phenomenon. They nest virtually around the world.

Pardis Sabeti
Researchers determined that the reason peregrine falcons weren't successfully hatching their eggs was due to eggshell thinning – a very serious problem, since the weaker shells were prone to breaking before the baby birds were ready to hatch.

Carl Thelander
It is a gradual build-up. They get a percentage of eggshell thinning until they reach a critical percentage when the egg literally collapses under the incubating female. And the end result of that is much depressed reproduction, and if you do that for an extended period of time it actually causes a population decline or extinction.

Pardis Sabeti
After looking into possible causes for this eggshell thinning, scientists began to zero in on a likely culprit: DDT and its breakdown product, DDE.

Linnea Hall
There were a couple of reasons why scientists believe there was a relationship between eggshell thinning and DDT or DDE. And, some of those came from just the fact that they were collecting the eggs in the field and they were seeing that they were extremely thin-shelled. And then, when they, they took the yolk contents or whatever other pieces of the egg they could and analyzed them, they found very high residues or prominent residues of DDE, DDT that had not been seen in the historic samples.

Pardis Sabeti
The falcons were ingesting DDT through their prey – birds they ate had small concentrations of the chemical in their flesh. Over time the DDT built up in the Peregrines' own bodies and started to affect the females' ability to lay healthy eggs.

Even though scientists had a pretty strong hunch that DDT was the cause of peregrine falcon eggshell thinning, they couldn't rely on their scientific instincts alone. So researchers turned to statistics as a way to validate their analysis.

They needed data about hundreds of bird eggs collected over time. Luckily, this information was just waiting to be accessed at the Western Foundation of Vertebrate Zoology, which houses the world's largest egg collection.

Linnea Hall

> In regards to peregrine eggs, in addition to the whole eggs that we have that collectors collected in the 1960s, 1970s, as well as earlier by some of the natural—naturalist collectors, we also have eggs that were collected specifically that were addled or abandoned or were broken in the nests. And those actually form, in large part, the collection of, of pesticide samples that we have here at the Foundation.

Pardis Sabeti
Researchers had a wealth of information to work with as they began to explore the relationship between eggshell thickness and DDE.

We can follow in their footsteps by taking a look at a data set compromised of 68 peregrine falcon eggs from Alaska and Northern Canada.

The two variables we'll be looking at are eggshell thickness and the log-concentration of DDE. Here's our scatterplot. The eggshell thickness will be our response variable on the *y*-axis. Our explanatory variable on the *x*-axis will be the log-concentration of DDE measured in the 68 eggs in our sample.

Here's the least squares regression line. Remember it's described by the equation "*y*-hat" = $a + bx$. The data show a negative, linear relationship between the two variables.

Using the equation, we can predict eggshell thickness for any measurement of DDE. The slope *b* and intercept *a* are statistics, meaning we calculated them from our sample data. But if we repeated the study with a different sample of eggs, the statistics would take on somewhat different values. So now what we want to know is whether there really is a negative linear relationship between these variables for the entire population of all peregrine eggs, beyond just the eggs that happen to be in our sample. Or might the pattern we see in our scatterplot simply be due to chance variation?

Data of the entire peregrine egg population might look like this. For any pesticide level, many different eggshell thicknesses may be observed. If the scatterplot is correct, the mean eggshell thickness, *y*, does have a linear relationship with log-concentration of DDE, *x*. This line is called the population regression line and it's estimated from the sample data since we never know all the population data.

Several conditions must be met in order to move forward with a regression inference, but for now we can assume that we're good to go.

The population regression model is written mu-*y*, equals alpha plus beta times *x*. Mu-*y* represents the true population mean of the response variable, *y,* for the given value of *x*. Alpha is the population *y*-intercept, or the true mean value of *y* when *x* is zero. Beta is the population slope, which quantifies the true relationship between *x* and *y* in this linear model.

Let's look back at our least squares regression line, based on the sample of 68 bird eggs. The sample intercept *a* is 2.146 and it's an estimate for the population intercept, alpha. And our sample slope *b*, -.3191, is an estimate of the population slope, beta.

Of course, we've learned by now that other samples from the same population will give us different data, resulting in different parameter estimates based on the different values for *a* and *b*.

So where do we go from here? Well, in repeated sampling, the value of these statistics *a* and *b* form sampling distributions, which provide the basis for statistical inference. In particular, we want to infer from the sampling distribution for our statistic *b* whether the sample data provide sufficiently strong evidence that higher levels of DDE are related to eggshell thinning in the population. To answer this question, we set up our null and alternative hypotheses.

In this case, the null hypothesis is that the amount of DDE and eggshell thickness have no linear relationship—that is, DDE has no effect on eggshell thickness.

In terms of our population parameters, that's the same as saying that beta, which quantifies how changes in *x* are related to changes in *y*, equals zero.

The alternative hypothesis is that the amount of DDE and eggshell thickness have a negative linear relationship. Remember, a negative association is when one variable increases as another decreases. In this case, that would mean that as the amount of DDE increases, the eggshell thickness would decrease.

We can also restate this in terms of our population parameters, as beta is less than zero.

We'll use this equation for our test statistic, *t*. If the null hypothesis is true, the test statistic follows a pattern similar to the one we saw in the module on Small Sample Inference for One Mean. Start with the point estimate for the population slope, beta, which is *b*. In our sample, *b* equaled -.3191. Under our null hypothesis H_0, beta is zero so we can plug that in to the equation for beta sub-zero. Then divide the result by the standard error of *b*, which software tells us is .0255. The resulting *t*-value is negative 12.5. That's an extreme result. The *p*-value for this test is essentially zero, meaning we have strong evidence to reject the null hypothesis.

By rejecting the null hypothesis, we can confirm what scientists already suspected—that there is a connection between peregrine falcon eggshell thickness and the presence of DDE. More precisely, there is a statistically

significant, negative linear relationship between the log-concentration of DDE and thickness of peregrine eggshells.

Before researchers could present this finding to the public, however, they had to quantify this relationship. That means computing a confidence interval for the population slope.

To do this, we take the sample slope, b, plus or minus the margin of error, t^*s_b, where the degrees of freedom for t^* equals n-2. In this case that's 68 eggs minus 2, or 66.

We can then calculate this using software to find a 95% confidence interval of negative .3191, plus or minus .0509…or from negative .3700 to negative .2681.

Based on our sample of 68 peregrine falcon eggs, we are 95% confident that a one-unit increase in the log-concentration of DDE is associated with a decrease in true average eggshell thickness between .27 and .37 on Ratcliffe's Index.

Armed with this information, scientists were able to make a strong argument against the use of DDT because of its dangerous impact on peregrines and the environment as a whole.

> **Linnea Hall**
> After the correlation was established between DDT or DDE and eggshell thinning, basically the whole biological world and, and legal world got kind of turned on its head. Especially in this country, there was a very, very prolonged legal battle with people on both sides presenting evidence.
>
> **Carl Thelander**
> The American Chemical Society and the wildlife societies were in debate over the cause and effect, was not really demonstrated ultimately that DDT was causing the eggshell thinning. But the inference was so huge, the correlation was so huge, and so well supported by scientific evidence that the body of information overwhelmed the critics and DDT was identified as the problem. But statistics allowed us to make the case, a very strong case and an, an accumulation of a large geographic area and a large number of studies, to build a case that was defensible and indisputable, that statistically these were valid results.

Pardis Sabeti
Due to scientific and statistical evidence, the United States and many Western European countries banned DDT production. Since then, the peregrine falcon population has rebounded significantly—so much so that they are now off the

endangered species list. Thanks to statistics—and some dedicated bird-lovers—this environmental detective story has a happy ending for the peregrine falcons.

I’m Pardis Sabeti for *Against All Odds*. See you next time!

PRODUCTION CREDITS

Host – Dr. Pardis Sabeti

Writer/Producer/Director – Maggie Villiger

Associate Producer – Katharine Duffy

Editor – Seth Bender

Director of Photography – Dan Lyons

Sound Mix – Richard Bock
Animation – Jason Tierney

Title Animation – Jeremy Angier

Web + Interactive Developer – Matt Denault / Azility, Inc.

Website Designer – Dana Busch

Teleprompter – Sue Willard-Kiess

Hair/Makeup - Amber Voner

Additional Footage & Stills

- Coast Learning Systems
- Macaulay Library
- BBC Motion Gallery
- Western Foundation of Vertebrate Zoology
- Nick Dunlop?!

Music
DeWolfe Music Library

Based on the original Annenberg/CPB series *Against All Odds*,
Executive Producer Joe Blatt

Annenberg Learner Program Officer – Michele McLeod

Project Manager – Dr. Sol Garfunkel

Chief Content Advisor – Dr. Marsha Davis

Executive Producer – Graham Chedd

FUNDER CREDITS

Funding for this program is provided by Annenberg Learner.

For information about this, and other Annenberg Learner programs, call 1-800-LEARNER, and visit us at www.learner.org.