A Comparison of Effect Size Statistics

by Karen

Share / Save

If you're in a field that uses Analysis of Variance, you have surely heard that p-values alone don't indicate the size of an effect. You also need to give some sort of effect size measure.

Why? Because with a big enough sample size, any difference in means, no matter how small, can be statistically significant. P-values are designed to tell you if your result is a fluke, not if it's big.

Truly the simplest and most straightforward effect size measure is the difference between two means. And you're probably already reporting that. But the limitation of this measure as an effect size is not inaccuracy. It's just hard to evaluate.

If you're familiar with an area of research and the variables used in that area, you should know if a 3-point difference is big or small, although your readers may not. And if you're evaluating a new type of variable, it can be hard to tell.

Standardized effect sizes are designed for easier evaluation. They remove the units of measurement, so you don't have to be familiar with the scaling of the variables.

Cohen's d is a good example of a standardized effect size measurement. It's equivalent in many ways to a standardized regression coefficient (labeled beta in some software). Both are standardized measures-they divide the size of the effect by the relevant standard deviations. So instead of being in terms of the original units of X and Y, both Cohen's d and standardized regression coefficients are in terms of standard deviations.

There are some nice properties of standardized effect size measures. The foremost is you can compare them across variables. And in many situations, seeing differences in terms of number of standard deviations is very helpful.

But they're most useful if you can also recognize their limitations. Unlike correlation coefficients, both Cohen's d and beta can be greater than one. So while you can compare them to each other, you can't just look at one and tell right away what is big or small. You're just looking at the effect of the independent variable in terms of standard deviations.

This is especially important to note for Cohen's d, because in his original book, he specified certain d values as indicating small, medium, and large effects in behavioral research. While the statistic itself is a good one, you should take these size recommendations with a grain of salt (or maybe a very large bowl of salt). What is a large or small effect is highly dependent on your specific field of study, and even a small effect can be theoretically meaningful.

Another set of effect size measures for categorical independent variables have a more intuitive interpretation, and are easier to evaluate. They include Eta Squared, Partial Eta Squared, and Omega Squared. Like the R Squared statistic, they all have the intuitive interpretation of the proportion of the variance accounted for.

Eta Squared is calculated the same way as R Squared, and has the most equivalent interpretation: out of the total variation in Y, the proportion that can be attributed to a specific X.

Eta Squared, however, is used specifically in ANOVA models. Each categorical effect in the model has its own Eta Squared, so you get a specific, intuitive measure of the effect of that variable.

Eta Squared has two drawbacks, however. One is that as you add more variables to the model, the proportion explained by any one variable will automatically decrease. This makes it hard to compare the effect of a single variable in different studies.

Partial Eta Squared solves this problem, but has a less intuitive interpretation. There, the denominator is not the total variation in Y, but the unexplained variation in Y plus the variation explained just by that X. So any variation explained by other Xs is removed from the denominator. This allows a researcher to compare the effect of the same variable in two different studies, which contain different covariates or other factors.

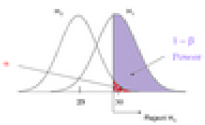
In a one-way ANOVA, Eta Squared and Partial Eta Squared will be equal, but this isn't true in models with more than one independent variable.

The drawback for Eta Squared is that it is a biased measure of population variance explained (although it is accurate for the sample). It always overestimates it.

This bias gets very small as sample size increases, but for small samples an unbiased effect size measure is Omega Squared. Omega Squared has the same basic interpretation, but uses unbiased measures of the variance components. Because it is an unbiased estimate of population variances, Omega Squared is always smaller than Eta Squared.

Other recent posts contain equations of all these effect size measures and a list of great references for further reading on effect sizes.

SHARE



Want to get up to speed on the meaning and logic of power, sample size, and how to calculate estimates? Check out our new On Demand online workshop called Calculating Power and Sample size.

Like this post? Enter your email address to have posts delivered

Subscribe

Tagged as: ANOVA, Cohen's d, effect size, Eta Squared, Omega Squared, Partial Eta Squared

{ 13 comments… read them below or add one }

Jim February 25, 2016 at 1:32 pm

I am currently taking a statistics for doctoral learners course working on effect size correlations. I am confused on the r-squared and Cohen's d (formula which uses the t value and square root of n). Working a problem with one study using 10 subjects having a t=1.0 and comparing to another study with 100 subject also with a t=1.9. In computing the r-squared and Cohen's d it appears as the sample size increases the effect size is less? For the above sample with 9(df) the r-squared is .535 and d=1.267, with 99(df) the r-squared is .187 and d=.3819. I understand that another important factor is testing for power, but looking at this correlation both the r-squared and Cohen's d appear to show less effect as sample size increases?

Reply

sarah June 30, 2015 at 2:36 pm

Hi Karen,

What does it mean if the associated p-value for R2 is not significant. For example:
R-SQUARE

Observed Two-Tailed
Variable Estimate S.E. Est./S.E. P-Value

Variable1 0.138 0.074 1.860 0.063 (or higher)

Does this invalidate the results of the regression model, even if some of my IVs are significant?
Thank you!

Reply

Usha May 27, 2015 at 6:43 am

Hi,
How to interpret the effect size of eta-squarred as small, medium, large? What reference can be cited for the same?

Reply

Karen June 3, 2015 at 9:13 am

The usual reference is Cohen (1988). But I suggest you don't do it.

Reply

Bryna Chrismas November 25, 2014 at 8:24 am

Good afternoon,

Am I correct in thinking that you cannot calculate cohens d for linear mixed models?

Do you have a reference/paper that discusses the use of effect sizes for mixed models?

Is the reason why you cannot use cohen's d due to the way a LMM works ie maximum likelihood?

Many thanks

Kind regards

Bryna

Reply

Susan March 22, 2014 at 3:14 pm

Thought you might be interested in this article: Bias and precision of some classical ANOVA effect sizes when assumptions are violated. It was published in Behavior Research Methods (doi: 10.3758/s13428-012-0257-2). There is a free spreadsheet available @ http://www.shsu.edu/~sts008/ to calculate eta squared, epsilon squared, and omega squared.

Reply

Richard May 16, 2013 at 9:13 am

Hi Karen,

I am wondering if there's a way to obtain measures of effect size when using Stata survey commands in generalized linear models. Based on your post and on my stats textbook, I think that omega squared would be the most appropriate given that it's a population-based survey.

Thanks in advance,

Richard

Reply

Karen February 24, 2012 at 8:10 pm

Hi Jae,

Almost. The partial eta-squared itself won't follow an F. The partial eta-squared is SS(effect)/SS(Error-other effects).

To get an F distribution, each SS has to be divided by its degrees of freedom.

Karen

Reply

Dan January 26, 2011 at 7:57 pm

Hello, Karen.
I wonder if there is a way two compare two partial eta measures between themselves in order to claim that one effect is stronger than another?

Thanks,
Dan

Reply

Karen February 25, 2011 at 12:58 pm

Hi Dan,

I don't know of a statistic that directly tests the partial eta-squared values. But it doesn't seem too hard to construct something at least close.

So partial eta-squared is the ratio of two Sums of Squares. Any ratio of two Mean Squares (which is just Sum of Squares/df) follows an F distribution. So you could create an F-test on your own, if you included the appropriate df.

It's really a matter of using a different denominator for the values of your F statistic. This isn't unheard off–it's done in simple effects testing.

Karen

Reply

Jae February 14, 2012 at 12:57 pm

Hi Karen,

I also had the same question.

Do you mean that the ratio of partial eta-squared, which follows an F distribution, also follows an F distribution?

Thank you,
Jae

Reply

Patrick January 15, 2011 at 1:39 pm

The linl to "equations of all these effect size measures" does not work

Reply

Karen January 20, 2011 at 10:44 am

Thanks, Patrick. I fixed it.

Reply

Leave a Comment

Name *

E-mail *

Website

Please note that Karen receives hundreds of comments at The Analysis Factor website each month. Since Karen is also busy teaching workshops, consulting with clients, and running a membership program, she seldom has time to respond to these comments anymore. If you have a question to which you need a timely response, please check out our low-cost monthly membership program, or sign-up for a quick question consultation.

Submit

{ 1 trackback }

- Science-Palooza Helps Students' Academic and Professional Development :: Oglethorpe University Stories

Previous post: How to Calculate Effect Size Statistics

Next post: Random Intercept and Random Slope Models Webinar

- **Join over 16,500 Subscribers**

Get our Monthly eNewsletter StatWise
for Tips, Resources, and News

First Name

Email

required

Sign Me Up!

We hate spam too. We will *never* share your email address w/anyone.

- **Upcoming Workshops**
  - Structural Equation Modeling

- **Data Analysis Brown Bag Online Membership Monthly Topic**
  - March 2016 Topic Webinar: Inter-rater Reliability

- **Get Blog Updates**

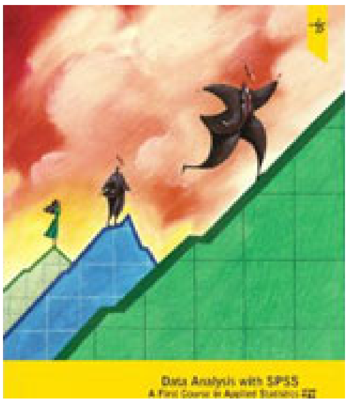
your email Subscribe

Follow @analysis_factor

- **Search**

To search, type and hit enter

- **Read Our Book**



Data Analysis with SPSS (4th Edition)
by Stephen Sweet and
Karen Grace-Martin

- **Statistical Resources by Topic**
  - Analysis of Variance and Covariance
  - Books
  - Complex Surveys & Sampling
  - Count Regression Models
  - Effect Size Statistics, Power, and Sample Size Calculations
  - Linear Regression Resources
  - Logistic Regression
  - Missing Data
  - Mixed and Multilevel Models
  - R
  - SPSS
  - Stata

Copyright © 2008-2016 The Analysis Factor. All rights reserved.
877-272-8096
Contact Us

WordPress Admin

:)