

19.4: Proportion of Variance Explained

Learning Objectives

- State the difference in bias between η^2 and ω^2
- Compute η^2 Compute ω^2
- Distinguish between ω^2 and partial ω^2
- State the bias in R^2 and what can be done to reduce it

Effect sizes are often measured in terms of the proportion of variance explained by a variable. In this section, we discuss this way to measure effect size in both ANOVA designs and in correlational studies.

ANOVA Designs

Responses of subjects will vary in just about every experiment. Consider, for example, the "Smiles and Leniency" case study. A histogram of the dependent variable "leniency" is shown in Figure 19.4.1. It is clear that the leniency scores vary considerably. There are many reasons why the scores differ. One, of course, is that subjects were assigned to four different smile conditions and the condition they were in may have affected their leniency score. In addition, it is likely that some subjects are generally more lenient than others, thus contributing to the differences among scores. There are many other possible sources of differences in leniency ratings including, perhaps, that some subjects were in better moods than other subjects and/or that some subjects reacted more negatively than others to the looks or mannerisms of the stimulus person. You can imagine that there are innumerable other reasons why the scores of the subjects could differ.

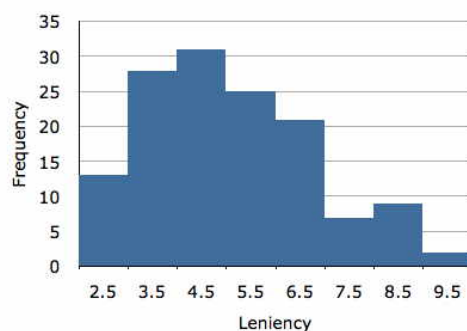


Figure 19.4.1: Distribution of leniency scores

One way to measure the effect of conditions is to determine the proportion of the variance among subjects' scores that is attributable to conditions. In this example, the variance of scores is 2.794. The question is how this variance compares with what the variance would have been if every subject had been in the same treatment condition. We estimate this by computing the variance within each of the treatment conditions and taking the mean of these variances. For this example, the mean of the variances is 2.649. Since the mean variance within the smile conditions is not that much less than the variance ignoring conditions, it is clear that "Smile Condition" is not responsible for a high percentage of the variance of the scores. The most convenient way to compute the proportion explained is in terms of the sum of squares "conditions" and the sum of squares total. The computations for these sums of squares are shown in the chapter on ANOVA. For the present data, the sum of squares for "Smile Condition" is 27.535 and the sum of squares total is 377.189. Therefore, the proportion explained by "Smile Condition" is:

$$\frac{27.535}{377.189} = 0.073 \quad (19.4.1)$$

Thus, 0.073 or 7.3% of the variance is explained by "Smile Condition."

An alternative way to look at the variance explained is as the proportion reduction in error. The sum of squares total (377.189) represents the variation when "Smile Condition" is ignored and the sum of squares error ($377.189 - 27.535 = 349.654$) is the variation left over when "Smile Condition" is accounted for. The difference between 377.189 and 349.654 is 27.535. This reduction in error of 27.535 represents a proportional reduction of $27.535/377.189 = 0.073$ the same value as computed in terms of proportion of variance explained.

This measure of effect size, whether computed in terms of variance explained or in terms of percent reduction in error, is called η^2 where η is the Greek letter eta. Unfortunately, η^2 tends to overestimate the variance explained and is therefore a biased estimate of the proportion of variance explained. As such, it is not recommended (despite the fact that it is reported by a leading statistics package).

An alternative measure, ω^2 (omega squared), is unbiased and can be computed from

$$\omega^2 = \frac{SSQ_{condition} - (k-1)MSE}{SSQ_{total} + MSE} \quad (19.4.2)$$

where MSE is the mean square error and k is the number of conditions. For this example, $k = 4$ and $\omega^2 = 0.052$.

It is important to be aware that both the variability of the population sampled and the specific levels of the independent variable are important determinants of the proportion of variance explained. Consider two possible designs of an experiment investigating the effect of alcohol consumption on driving ability. As can be seen in Table 19.4.1, Design 1 has a smaller range of doses and a more diverse population than Design 2. What are the implications for the proportion of variance explained by Dose? Variation due to Dose would be greater in Design 2 than Design 1 since alcohol is manipulated more strongly than in Design 1. However, the variance in the population should be greater in Design 1 since it includes a more diverse set of drivers. Since with Design 1 the variance due to Dose would be smaller and the total variance would be larger, the proportion of variance explained by Dose would be much less using Design 1 than using Design 2. Thus, the proportion of variance explained is not a general characteristic of the independent variable. Instead, it is dependent on the specific levels of the independent variable used in the experiment and the variability of the population sampled.

Table 19.4.1: Design Parameters

Design	Dose	Population
1	0.00	All Drivers between 16 and 80 Years of Age
	0.30	
	0.60	
2	0.00	Experienced Drivers between 25 and 30 Years of Age
	0.50	
	1.00	

Factorial Designs

In one-factor designs, the sum of squares total is the sum of squares condition plus the sum of squares error. The proportion of variance explained is defined relative to sum of squares total. In an $A \times B$ design, there are three sources of variation (A , B , $A \times B$) in addition to error. The proportion of variance explained for a variable (A , for example) could be defined relative to the sum of squares total ($SSQ_A + SSQ_B + SSQ_{A \times B} + SSQ_{error}$) or relative to $SSQ_A + SSQ_{error}$.

To illustrate with an example, consider a hypothetical experiment on the effects of age (6 and 12 years) and of methods for teaching reading (experimental and control conditions). The means are shown in Table 19.4.2 The standard deviation of each of the four cells ($Age \times Treatment$ combinations) is 5. (Naturally, for real data, the standard deviations would not be exactly equal and the means would not be whole numbers.) Finally, there were 10 subjects per cell resulting in a total of 40 subjects.

Table 19.4.2: Condition Means

Age	Treatment	
	Experimental	Control
6	40	42
12	50	56

The sources of variation, degrees of freedom, and sums of squares from the analysis of variance summary table as well as four measures of effect size are shown in Table 19.4.3. Note that the sum of squares for age is very large relative to the other two effects. This is what would be expected since the difference in reading ability between 6- and 12-year-olds is very large relative to the effect of condition.

Table 19.4.3: ANOVA Summary Table

Source	df	SSQ	η^2	partial η^2	ω^2	partial ω^2
Age	1	1440	0.567	0.615	0.552	0.586
Condition	1	160	0.063	0.151	0.053	0.119
A x C	1	40	0.016	0.043	0.006	0.015
Error	36	900				
Total	39	2540				

First, we consider the two methods of computing η^2 , labeled η^2 and partial η^2 . The value of η^2 for an effect is simply the sum of squares for this effect divided by the sum of squares total. For example, the η^2 for Age is $1440/2540 = 0.567$. As in a one-factor design, η^2 is the proportion of the total variation explained by a variable. Partial η^2 for Age is SSQ_{Age} divided by ($SSQ_{Age} + SSQ_{error}$), which is $1440/2340 = 0.615$.

As you can see, the partial η^2 is larger than η^2 . This is because the denominator is smaller for the partial η^2 . The difference between η^2 and partial η^2 is even larger for the effect of condition. This is because SSQ_{Age} is large and it makes a big difference whether or not it is included in the denominator.

As noted previously, it is better to use ω^2 than η^2 because η^2 has a positive bias. You can see that the values for ω^2 are smaller than for η^2 . The calculations for ω^2 are shown below:

$$\omega^2 = \frac{SSQ_{effect} - df_{effect}MS_{error}}{SSQ_{total} + MS_{error}} \quad (19.4.3)$$

$$\omega^2_{partial} = \frac{SSQ_{effect} - df_{effect}MS_{error}}{SSQ_{effect} + (N - df_{effect})MS_{error}} \quad (19.4.4)$$

where N is the total number of observations.

The choice of whether to use ω^2 or the partial ω^2 is subjective; neither one is correct or incorrect. However, it is important to understand the difference and, if you are using computer software, to know which version is being computed. (Beware, at least one software package labels the statistics incorrectly).

Correlational Studies

In the section "Partitioning the Sums of Squares" in the Regression chapter, we saw that the sum of squares for Y (the criterion variable) can be partitioned into the sum of squares explained and the sum of squares error. The proportion of variance explained in

multiple regression is therefore:

$$SSQ_{explained} / SSQ_{total} \quad (19.4.5)$$

In simple regression, the proportion of variance explained is equal to r^2 ; in multiple regression, it is equal to R^2 .

In general, R^2 is analogous to η^2 and is a biased estimate of the variance explained. The following formula for adjusted R^2 is analogous to ω^2 and is less biased (although not completely unbiased):

$$R^2_{adjusted} = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1} \quad (19.4.6)$$

where N is the total number of observations and p is the number of predictor variables.

This page titled [19.4: Proportion of Variance Explained](#) is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by [David Lane](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.