

# Is Heterogeneity Your Friend?

Using sensitivity analyses to design improved studies

**IDENTIFYING, ELIMINATING** or controlling heterogeneity is a fundamental principle in many statistical techniques. Heterogeneity and its opposite, homogeneity, refer to how consistent or stable a particular data set or variable relationship are. Having statistical heterogeneity is not a good or bad thing in and of itself for the analysis; however, it's useful to know to design, choose and interpret statistical analyses. Indeed, the comparison of heterogeneity often will be the outcome of interest, especially in quality fields.

There are pros and cons to having heterogeneity in a sample or analysis. Statisticians are trained to analyze variability (analysis of variance), decompose variability (variance components), control variability (modeling confounders and covariates) and to use other methods. In general, controlling for or lowering heterogeneity can remove potential confounders or noise to increase sensitivity for the measured outcomes.

High heterogeneity, in contrast, is often more realistic for modeling the messy real world and may give better results or identify subpopulations. In clinical trials of new therapeutics, for example, homogeneous patient groups are sought to clearly identify efficacy. In market research and polling, however, heterogenous samples are requested to get a full picture.

Traditionally, statisticians will aim for homogeneity of controlled experiments

because it allows for straightforward testing and conclusions. While reducing variability and aiming for homogeneity in testing is important, recognizing and examining the diversity should play a key role in any analyses. The reputation of heterogeneity has been increasing as evidenced in the meta-analysis field's phrase, "Yes, heterogeneity is your friend," commonly used when examining and performing meta-analyses. Including and even inviting variability in analyses ultimately can create stronger and more representative models and conclusions.

## Ignoring heterogeneity: Anscombe's quartet

### Basic Anscombe data analysis:

Anscombe introduces six data sets with two showing identical X summary statistics and four showing similar Y summary statistics. In examining these summary statistics, we assume we understand quite a bit about the distribution of the data set.

Table 1 gives the summary of four small data sets with similar means and standard deviations. The addition of the median (Table 2) provides more information showing some differences in the distributions and differences from the mean values indicating skewness in the data. Only after the raw data are examined do the differences become completely apparent, and when plotted against Y1 – Y4 do the differences in relationships between X and Y become completely transparent (Figure 1).

### Anscombe regression analysis:

Anscombe created this data set to illustrate linear regression similarities with disparate data sets. His famous set of linear regression equations illustrates the perils of ignoring the variability in the data. Creating almost identical regression equations, Anscombe shows that highly heterogeneous data can deceive the researcher. The four regressions shown in Table 3 are all almost identical to  $y = 0.5x + 3$  with a r-squared fit near 0.66.

Based on the equations alone, you might assume the data came from the same source on four different times and represents a consistent output. Can you reasonably assume, however, that the data that generated these models are similar? In fact, the data are quite heterogeneous as shown in Figure 1.

Equation one is a typical scatter plot for a linear regression, and equation two indicates clear nonlinearity. Most interesting are equations three and four because they show data with clear outliers. Assuming these are part of a clinical study, investigating these points can identify patients with important characteristics, or simply inaccurately entered data. An analysis strategy to eliminate these outlier data points, and analysis of equation two using a quadratic function will give different results than the initial four equations.

While Anscombe's quartet<sup>1</sup> is usually used to represent the limitations of linear regressions, the variability between and

## Summary statistics for 6 data sets / TABLE 1

	X1	X4	Y1	Y2	Y3	Y4
Mean	9.00	9.00	7.50	7.50	7.50	7.50
SD	3.317	3.317	2.032	2.032	2.030	2.031

SD = standard deviation

## Median for the 6 data sets / TABLE 2

	X1	X4	Y1	Y2	Y3	Y4
Median	9.00	8.00	7.58	8.14	7.11	7.04

within the data sets are a good study for heterogeneity. If equations one to three measured a different variable for a given  $X_1$ , this demonstrates heterogeneity in the relationships of the variables. Within each equation, there are examples of heterogeneity. Equations one and two are examples of heterogeneity in the residuals to the regression, known as heteroscedasticity; the pattern in the heteroscedasticity of equation two also would indicate the poor fit for linear regression.

For equations three and four, the outlier points may be interesting or mundane. It could simply (and often) be an entry mistake or calculation error. If it is a trustworthy data point, however, it is worthwhile to investigate. These points may hint at underlying subpopulations that have a distinctly different response (heterogeneity in the sample or independent variables), or the measured outcome is distinctly different at that measurement (heterogeneity in the function, output or dependent variable), or both.

While it is often frustrating to find these issues in an analysis, heterogeneity can lead to new questions and conclusions that create better models and discoveries. For multivariate linear models, diagnostic statistics such as Cook's D, leverage and residual analyses can point to issues with the relationship between independent and dependent variables.<sup>2</sup>

### Controlling heterogeneity: Using REMs in meta-analyses

The majority of meta-analyses are conducted using either fixed effects models (FEM) or random effects models (REM). The decision to switch from FEM to REM is often based solely on the values of Cochran's Q-statistic or the  $I^2$  statistic.

Cochran's Q statistic represents the total variance between the studies, while  $I^2$  is a measure of how much heterogeneity there is between the studies.<sup>3</sup> Researchers will use the REM if Cochran's Q is significant or if the  $I^2$  statistic is large (greater than 50% is

one suggested threshold). The FEM assumes that all the studies have a "true" effect size that is identical and the only variation in a study's results is sampling error. The REM assumes each study provides information about different effect sizes and controls for this between-study variability in its summary values.

When reporting results as an REM, researchers can assume they have controlled for heterogeneity between studies but have not explained why the studies vary. It is even possible that when switching from a FEM to a REM, the results will switch conclusions from favoring one treatment to favoring another.

In a meta-analysis comparing the persistence of asthma over and under age 12, seven studies were identified and a risk difference (RD) ( $> \text{age } 12 - \leq \text{age } 12$ ) was synthesized over these studies. The results (see Figure 2) show that with the FEM, the overall effect was significant ( $p < 0.001$ ) and was negative (significantly lower persistence over age 12). For the REM, the overall effect was inconclusive and not significant ( $p = 0.513$ ). Highly significant heterogeneity was seen with the variation in the RD attributable to heterogeneity = 76% (Cochran's Q statistic  $p < 0.0001$ ). The REM was chosen as the appropriate model and presented by one of the authors.

Selecting the appropriate model for reporting, however, should not be the final step in the process. Identifying possible causes of heterogeneity and analyzing the results—including important covariates in the model—should be part of the analysis plan.

## 4 linear regression equations with Anscombe's data / TABLE 3

Equation 1: $Y_1 = 3.00 + 0.500 X_1$				
Predictor	Coef	SE coef	T	P
Constant	3.000	1.125	2.67	0.026
$X_1$	0.5001	0.1179	4.24	0.002
S = 1.23660 R-sq = 66.7%				
Equation 2: $Y_2 = 3.00 + 0.500 X_1$				
Predictor	Coef	SE coef	T	P
Constant	3.001	1.125	2.67	0.026
$X_1$	0.5000	0.1180	4.24	0.002
S = 1.23721 R-sq = 66.6%				
Equation 3: $Y_3 = 3.15 + 0.486 X_1$				
Predictor	Coef	SE coef	T	P
Constant	3.152	1.136	2.77	0.022
$X_1$	0.4861	0.1191	4.08	0.003
S = 1.24948 R-sq = 64.9%				
Equation 4: $Y_4 = 3.07 + 0.496 X_2$				
Predictor	Coef	SE coef	T	P
Constant	3.071	1.082	2.84	0.019
$X_2$	0.4963	0.1134	4.38	0.002
S = 1.18948 R-sq = 68%				

Coef = coefficient  
P = p-value

SE = standard error  
T = test statistic

For this meta-analysis, several covariates were available: year of the study, location of the study, percentage of females in the study, lost to follow-up study and year of asthma onset. Because of the small number of studies, each was examined separately. Initially, a jackknife procedure omitting individual studies was performed, and the omission of one study (the only study from Iceland) reduced the heterogeneity to 51%.

After further examination, this study (with the largest positive effect size) had an imbalance of females and had many more patients lost to follow-up than the other six studies. The final decision of the authors was to include the REM as well as an extensive sensitivity analysis to identify the potential causes of heterogeneity.

### Improving future studies

These examples illustrate the problem that simply examining results of statistical

analyses—even those as simple as summary statistics—can hide heterogeneity in the data set. Several rules of thumb for statistical analyses can be suggested by these examples:

1. Always examine your raw data. If the data set is large, take a random sample and examine the distribution.
2. When fitting a model, ensure that the relationships between the independent and dependent variables are linear (for general linear models) and don't include outliers that are influencing the resulting regressions. Examine residuals and regression diagnostic measures.
3. Identify the possible causes of heterogeneity. Analyze your data with and without outliers, and attempt to identify why the data point is an outlier. Include covariates and confounders in your analyses to explain the heterogeneity in the model.

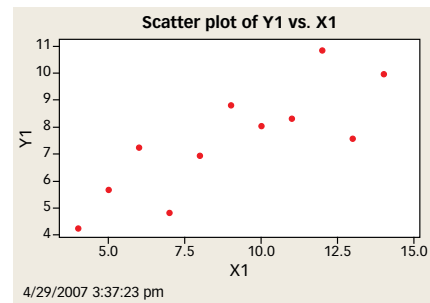
Use the results of these sensitivity analyses to design improved studies in the future. **QP**

#### REFERENCES

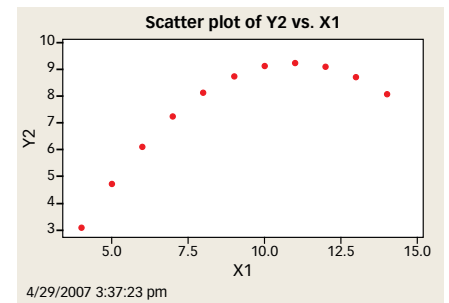
1. Wikipedia, "Anscombe's quartet," [https://en.wikipedia.org/wiki/Anscombe's\\_quartet](https://en.wikipedia.org/wiki/Anscombe's_quartet).
2. "Research Methods II: Multivariate Analysis," *Journal of Tropical Pediatrics*, Oxford Journals, chapter five, "Regression Diagnostics," [www.oxfordjournals.org/our\\_journals/tropej/online/ma\\_chap5.pdf](http://www.oxfordjournals.org/our_journals/tropej/online/ma_chap5.pdf).
3. Julian P.T. Higgins, Simon G. Thompson, Jonathan J. Deeks

## Plots of Anscombe data / FIGURE 1

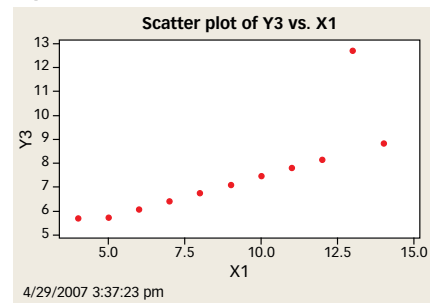
Equation 1



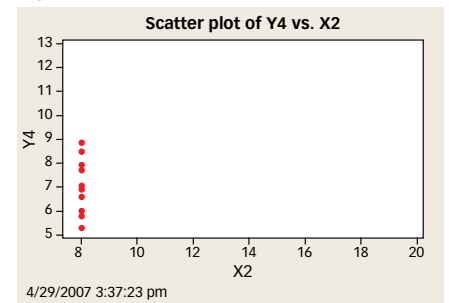
Equation 2



Equation 3



Equation 4



and Douglas G. Altman, "Measuring Inconsistency in Meta-Analyses," *British Medical Journal*, Sept. 6, 2003, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC192859>.



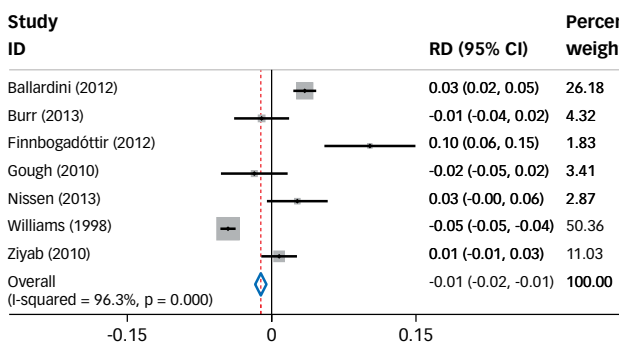
**I. ELAINE ALLEN** is professor of biostatistics at the University of California, San Francisco and emeritus professor of statistics at Babson College in Wellesley, MA. She also is director of the Babson Survey Research Group at Babson College. She earned a doctorate in statistics from Cornell University in Ithaca, NY. Allen is a member of ASQ.



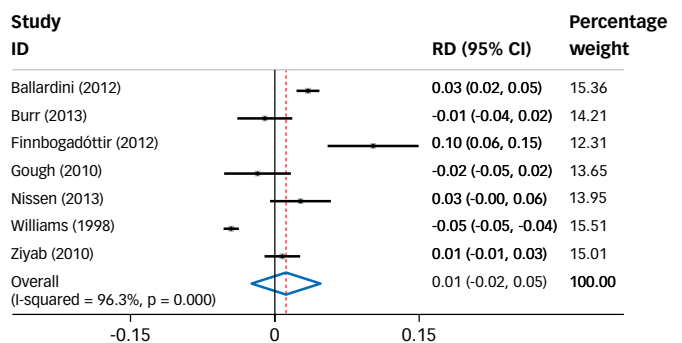
**JULIA E. SEAMAN** is a research scientist at the University of California, San Francisco, focused on pharmaceutical chemistry. She also is a statistical consultant for the Babson Survey Research Group. She earned her doctorate in pharmaceutical chemistry and pharmacogenomics from the University of California, San Francisco.

## Fixed and random effects meta-analysis on persistence of asthma / FIGURE 2

A: Fixed effects model



B: Random effects model



Note: Weights are from random effects analysis  
CI = confidence interval  
RD = risk difference