Analysis of Variance

Martin G. Larson

Circulation 2008;117;115-121
DOI: 10.1161/CIRCULATIONAHA.107.654335
Circulation is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 72514
 Print ISSN: 0009-7322. Online ISSN: 1524-4539

Statistical Primer for Cardiovascular Research

Analysis of Variance

Martin G. Larson, SD

Analysis of variance (ANOVA) is a statistical technique to analyze variation in a response variable (continuous random variable) measured under conditions defined by discrete factors (classification variables, often with nominal levels). Frequently, we use ANOVA to test equality among several means by comparing variance among groups relative to variance within groups (random error).

Sir Ronald Fisher pioneered the development of ANOVA for analyzing results of agricultural experiments.[1] Today, ANOVA is included in almost every statistical package, which makes it accessible to investigators in all experimental sciences. It is easy to input a data set and run a simple ANOVA, but it is challenging to choose the appropriate ANOVA for different experimental designs, to examine whether data adhere to the modeling assumptions, and to interpret the results correctly. The purpose of this report, together with the next 2 articles in the Statistical Primer for Cardiovascular Research series, is to enhance understanding of ANVOA and to promote its successful use in experimental cardiovascular research. My colleagues and I attempt to accomplish those goals through examples and explanation, while keeping within reason the burden of notation, technical jargon, and mathematical equations.

Here, I introduce the ANOVA concept and provide details for 2 common models. The first model, 1-way fixed-effects ANOVA, is an extension of the Student 2-independent-samples *t* test that lets us simultaneously compare means among several independent samples. The second model, 2-way fixed-effects ANOVA, has 2 factors, A and B, and each level of factor A appears in combination with each level of factor B. This model lets us compare means among levels of factor A and among levels of factor B; furthermore, we may examine whether combined factors induce interaction effects (synergistic or antagonistic) on the response.

In the second ANOVA article, the author reviews several multiple-comparisons procedures for analysis of differences among means, including comparisons between pairs of group means and more general contrasts among group means. Usually, multiple-comparisons procedures are used to control type I error rate across numerous hypothesis tests. In the third ANOVA report, the author introduces repeated-measures ANOVA for use when each experimental unit contributes response data at each level of a fixed factor (eg, different treatment doses). Statistical textbooks[2,3] and online documents[4,5] provide readers with more technical detail for similar material or with broader coverage of topics beyond the scope of these articles.

Background

In this section, I briefly review key terminology for defining experimental design and ANOVA. An "experimental unit" is the smallest unit of experimental material to which a factor or combination of factors may be applied. Typically, each experimental unit is a whole organism (eg, human, mouse, or rat), but it may be at the suborganism level (eg, individual myocytes) or supraorganism level (eg, an institution). To determine the appropriate ANOVA model, we must know the relations between factors and experimental units.

Statisticians distinguish 2 types of factors in experimental design and ANOVA: "fixed factors" and "random factors." A "fixed factor" is one for which the specific levels are of interest. An investigator could repeat the entire experiment using identical factor levels both times. Conceptually, each level of a fixed factor represents a distinct population with a unique response mean. When an investigator deliberately arranges or modifies the levels of a fixed factor, we call those levels treatments. The primary ANOVA objective is to test whether response means are identical across factor levels. In contrast to a fixed factor, the levels of a "random factor" represent a random sample from a potentially infinite number of levels. Different factor levels would be chosen randomly if the experiment were redone. With random factors, the ANOVA objective is to make an inference about random variation within a population.

When a factor level is applied to 2 or more independent experimental units, it is "replicated." If replicates are equal in number for each factor level, the experimental design is "balanced." These concepts generalize to combinations of factor levels.

An experiment may contain 2 or more factors combined in 2 different ways, either "crossed" or "nested." With crossed factors, each level of factor A is present in combination with each level of factor B. For instance, each of 2 different medications (factor A levels are drugs X and Y) could be administered at either of 2 doses (factor B levels are low or high), with each experimental unit receiving 1 drug at 1 dose. In contrast to the situation with crossed factors, each level of

From the Department of Mathematics and Statistics, Boston University, Boston, Mass, and the Framingham Heart Study of the National Heart, Lung, and Blood Institute, Framingham, Mass.

Correspondence to Martin Larson, SD, Framingham Heart Study, 73 Mount Wayte Ave, Framingham, MA 01702. E-mail mlarson@bu.edu

(*Circulation*. 2008;117:115-121.)

Circulation is available at http://circ.ahajournals.org

DOI: 10.1161/CIRCULATIONAHA.107.654335

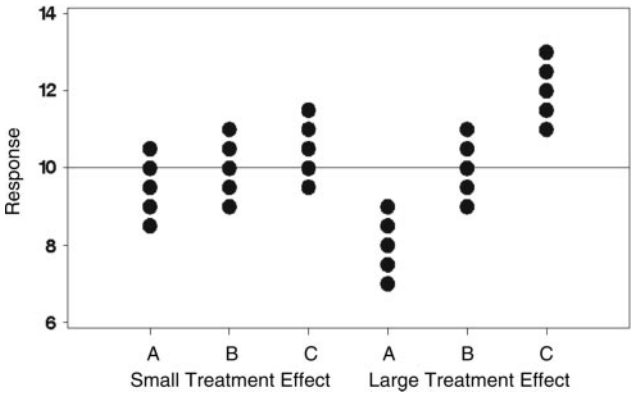

Figure 1. Illustration of treatment effects. Left, Small treatment differences relative to error variance. Right, Large treatment differences relative to error variance.

a nested factor occurs in just 1 level of the factor within which it is nested. In a study to compare for-profit versus nonprofit institutions with respect to patients' length of stay after coronary artery bypass, institutional status is a fixed-factor status (for profit/nonprofit), "hospital" is a random factor (specific hospitals are its levels) nested within the fixed-factor levels, and individual patients are the experimental units.

We usually refer to ANOVA models using the terms "fixed effects" or "random effects." This should not cause confusion, because fixed factors correspond with fixed effects among factor levels (that is, between-population mean differences), and random factors correspond with random effects among levels (that is, within-population random differences). If the experimental design includes fixed and random effects, then we use a "mixed-effects" ANOVA model.

One-Way Fixed-Effects ANOVA

Consider an experiment that has 2 or more treatments and multiple replicates of each treatment. We use a 1-way fixed-effects ANOVA model to test the null hypothesis that all treatments have the same population mean. The alternative hypothesis is that at least 1 population mean differs from the others. We assess whether variability among sample means is sufficiently large, relative to random error variance, that we should reject the null hypothesis and conclude that true differences exist among population means. The dot plot in Figure 1 illustrates hypothetical data in which variation among treatment means is small, consistent with identical population means (left-most 3 groups), or large, favoring unequal population means (right-most 3 groups).

Assumptions

When we model data using 1-way fixed-effects ANOVA, we make 4 assumptions: (1) individual observations are mutually independent; (2) the data adhere to an additive statistical model comprising fixed effects and random errors; (3) the random errors are normally distributed; and (4) the random errors have homogenous variance. Violations of these assumptions may compromise or invalidate the ANOVA results, so let us examine each individually.

Independence

The value of 1 observation must not influence the value of other observations. All experimental units must be independent, and each experimental unit must contribute only 1 response value.

Additivity

We can represent the data using a statistical model with additive components. The model for 1-way fixed-effects ANOVA may be written as follows: individual response=(grand mean)+(treatment effect)+(random error).

Normality

We assume that the random errors within each treatment group, the deviations from each group mean, have a normal, or gaussian, probability distribution.

Homogeneous Variance

Finally, we assume that the within-group random errors have identical variance across all treatment groups, represented by the parameter σ^2.

Together, assumptions of independence, homogeneous variances, and normality imply that residual errors are a sample of independently and identically distributed normal deviates.

ANOVA Calculations

Without going into mathematical details, the calculations proceed as follows. For each observation, we write: deviation from overall mean=individual value−overall mean. Squaring each deviation and summing over all observations yields the "total sum of squares" (SST). SST represents total variability of observations from their overall mean, quantified by the sum of their squared differences. An individual deviation also can be written as: deviation from overall mean=(treatment mean−overall mean)+(individual value−treatment mean). With some algebra found in statistical textbooks, SST partitions into 2 independent parts. The 2 parts are (1) "sum of squares between treatments" (SSA), which is obtained by summing the terms (treatment mean−overall mean)^2, and (2) "sum of squares within treatments" (SSE), which is obtained by summing the terms (individual value−treatment mean)^2. SSA represents variability among group means, and SSE represents within-group residual variability. Each sum of squares has its corresponding "degrees of freedom" (abbreviated *df*), which is the effective number of independent observations used in forming that sum of squares. With N observations, the total sum of squares, SST, has $N-1$ *df*; with $a \geq 2$ treatment groups, the "between-treatments" sum of squares, SSA, has $(a-1)$ *df*; finally, the "within-treatments" residual sum of squares has $(N-1)-(a-1)=(N-a)$ *df*. Interested readers can find rules for determining *df* in standard statistical texts or online.[2–4]

Dividing each sum of squares by its *df* yields a quantity called a "mean square." The residual mean square, MSE=SSE/$(N-a)$, estimates the error variance, σ^2. If the "null hypothesis" is correct, such that all treatments have the same population mean, then the between-treatments mean square, MSA=SSA/$(a-1)$, also estimates σ^2. In that situation, the ratio of the 2 variance estimates, denoted by

Table 1. Display of Results for 1-Way Fixed-Effects ANOVA

Source of Variation	*df*	Sums of Squares	Mean Square	F Statistic	*P*
Treatments	$a-1$	SSA (among treatments)	MSA=SSA/($a-1$)	MSA/MSE	*P*
Error	$N-a$	SSE (within treatments)	MSE=SSE/($N-a$)	...	...
Total	$N-1$	SST	...	...	...

Note that *P* is the probability that an $\mathcal{F}$ random variable with *df* ($a-1$) and ($N-a$) exceeds the observed F statistic.

F=MSA/MSE, has the statistical distribution called the $\mathcal{F}$ distribution, with ($a-1$) and ($N-a$) *df*. (The $\mathcal{F}$ distribution was named in honor of Fisher.) Large values of the F ratio provide evidence against the null hypothesis of equal treatment population means. The probability value is the probability that a random variable selected from an $\mathcal{F}$ distribution with ($a-1$) and ($N-a$) *df* will exceed the observed F value.

Table 1 displays calculations for the 1-factor fixed-effects model. Scientific journals usually do not publish the full ANOVA table due to limited space; some journals report the F statistic, its *df*, and probability value, whereas others report only the probability value. Subsequent to a "statistically significant" result (that is, obtaining $P<\alpha$, where α is the prespecified type I error rate), one may explore differences in treatment means using multiple-comparisons methods covered in the next article in the present series on statistics.

Example 1: One-Way ANOVA

To illustrate 1-way ANOVA, let us explore data on levels of soluble leptin receptor (sOB-R; ng/mL) according to categories of body mass index (BMI; kg/m^2) for 188 men in the Framingham Third Generation Cohort.[6] sOB-R was measured on a 10% random sample drawn from the full cohort. For convenience, I analyzed men only and classified them into 4 BMI categories (20 to 24, 25 to 29, 30 to 34, and ≥35 kg/m^2). Additionally, I used natural-logarithm transformation to normalize the distribution of response values. Table 2 and Figure 2 display descriptive statistics for log(sOB-R) in each BMI group. Note that sample sizes are moderate to large (n=26 to 62), data distributions are approximately symmetrical, and measures of spread (SDs and interquartile ranges) are similar across groups. The box-plot graph (Figure 2) contains substantially more information about the distribution of values than does a bar chart with error bars.

Table 3 displays calculations from 1-way ANOVA (SAS procedure ANOVA).[7] With N=188 men in 4 BMI categories, there are (4−1)=3 *df* among groups and (188−4)=184 *df* within groups. The sum of squares among BMI groups (SSA=4.24) is 19.1% of the total sum of squares (SST=22.17), and the ratio of mean squares is highly statistically significant (F=14.50, *df*=3 and 184, $P<0.0001$).

Table 2. Descriptive Statistics for Log(sOB-R) by BMI Category

BMI Group, kg/m^2	Sample Size	Mean	SD
20–24	60	3.81	0.32
25–29	62	3.52	0.31
30–34	40	3.56	0.32
≥35	26	3.39	0.29

Units for sOB-R are ng/mL.

These data provide strong evidence against the null hypothesis that the BMI groups have the same population mean level of log(sOB-R). Inspection of Table 2 or Figure 2 suggests an inverse association, with decreasing log(sOB-R) as BMI increases.

Checking Assumptions

If the design is balanced and the sample is large, ANOVA is robust with regard to moderate deviations from assumptions of homogenous variances and normal error. The calculated F statistic still has approximately an $\mathcal{F}$ distribution. In contrast, fixed-effects 1-way ANOVA is invalid if the observations are not independent.[3] It is important to check and report whether one's data adhere to the assumptions and to perform supplementary analyses if serious violations exist.

Independence

Independence of observations is the most critical among the 4 assumptions. To check this assumption, we must examine the research design. If the protocol stipulates random selection of experimental units from a defined population and random assignment of treatments to experimental units, and if the analysis uses a single response value for each experimental unit, then observations might be independent. Some sources of nonindependence are obvious: multiple values recorded over time for each experimental unit, or observations on multiple members of the same family. Matching or blocking in the experimental design is not as obvious but is a source of nonindependence. If the data contain correlated observations, we must use a more complex model instead of

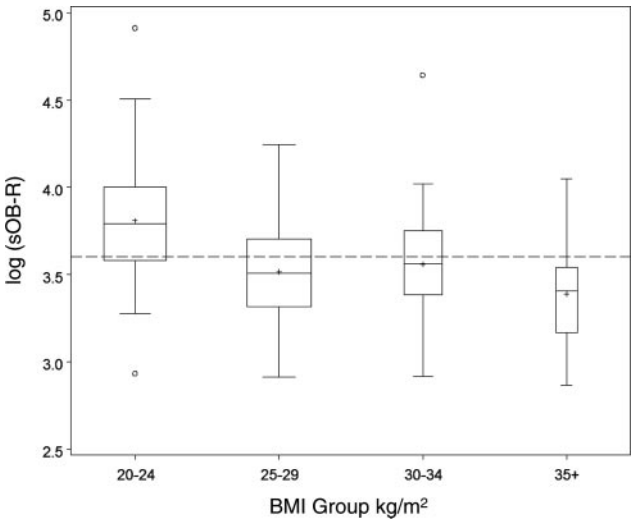

Figure 2. Box plots of log(sOB-R) levels by BMI group for 188 men in the Framingham Third Generation Cohort. Box width is proportional to sample size (Table 2). Units are ng/mL for sOB-R and kg/m^2 for BMI.

Table 3. Results of 1-Way ANOVA for Log(sOB-R) by BMI Category

Source of Variation	*df*	Sums of Squares	Mean Square	F Statistic	*P*
BMI category	3	4.24	1.413	14.50	<0.0001
Error	184	17.93	0.097	...	...
Total	187	22.17	...	...	...

Units for sOB-R are ng/mL.

1-way fixed-effects ANOVA. One approach to analyze correlated observations, repeated-measures ANOVA, appears later in the present series on statistics. In this example, it is reasonable to assume independent observations, because this is a random sample from a large cohort, and 1 response measurement per person is present.

Additivity

In the 1-way ANOVA model, failure to satisfy the additive assumption often leads to nonhomogeneous variances, which are covered next.

Homogeneous Variance

Levene's test[8] is widely used to test the null hypothesis that variances are homogeneous. An alternative procedure, Bartlett's test, performs poorly with nonnormal data[3] and should not be used unless normality has been validated. Visual inspection of Table 2 and Figure 2 suggests that the spread of log(sOB-R) is similar in all BMI categories, and this is confirmed by Levene's test ($P=0.94$), so we conclude that variances are not heterogeneous.

Normality

The Shapiro-Wilk procedure[9] may be used to test normality in samples with fewer than 2000 observations. In this example, log-transformed sOB-R data are approximately normally distributed in each BMI category (Shapiro-Wilk test $P=0.11$, 0.52, 0.17, and 0.91, respectively). The raw data deviate severely from normality (at $P<0.001$ in 3 BMI categories) with right skewness and/or high kurtosis, and this justifies application of the normalizing logarithmic transformation.

The ANOVA model is just an approximation for the data, and ANOVA assumptions may not be satisfied completely. With normal data but heterogeneous variances, ANOVA is robust for balanced or nearly balanced designs but not for highly unbalanced designs.[3] In the setting of normal data, heterogeneous variances, and an unbalanced design, one might use Welch's ANOVA to accommodate unequal variances.[10] With homogeneous variances but nonnormal data, ANOVA is robust for balanced designs with large samples but not for unbalanced design or small samples ($n<5$ per group). In the setting of nonnormal data, homogeneous variances, and a small sample or highly unbalanced design, a nonparametric procedure such as the Kruskal-Wallis test[11] may be preferred over 1-way ANOVA. If the data are not normally distributed and variances are heterogeneous, a transformation may be necessary. At the research design stage, an investigator must realize the importance of a balanced design and large sample.

Confidence Intervals

After 1-way ANOVA, one may wish to estimate a confidence interval (CI) for a population mean or for the difference between 2 population means. The form of the CI is (sample estimate)±(confidence coefficient)×(standard error of sample estimate). To construct a $100(1-\alpha)\%$ CI for the i-th population mean, we proceed as follows. For the first quantity, substitute the sample mean of group i. For the standard error of the sample mean, use $(MSE/n_i)^{1/2}$, where n_i is the sample size for the i-th group, and MSE is the mean squared error from the ANOVA model. Finally, for the confidence coefficient, use the $(1-\alpha/2)$ quantile of a t distribution with *df* equal to "error *df*" in the ANOVA model. MSE appears in the standard error calculation (not the individual group variance estimator, s_i^2), because MSE is the ANOVA estimate of the homogeneous within-population variance. Also, the ANOVA "error *df*" is the *df* for the t distribution (not n_i-1, the *df* for s_i^2), because it is the *df* associated with MSE.

To construct a $100(1-\alpha)\%$ CI for the difference between means of populations i and j, the sample estimate is (sample mean for group i−sample mean for group j), the standard error is $[MSE\ (1/n_i+1/n_j)]^{[1/2]}$, and the confidence coefficient is as defined above. In constructing both types of CIs, for 1 population mean or for the difference between 2 population means, we gain precision by using the ANOVA variance estimate, MSE, instead of group-specific variances; consequently, the average length of these CIs is shorter than CIs based on group-specific variances. The next report in the present series on statistics offers detailed discussion of analyses after the initial F test, specifically, the use of multiple-comparisons procedures.

Two-Way Fixed-Effects ANOVA

In a factorial experimental design, each factor is crossed with the other factors. Consider 2 fixed factors, A and B, with a levels for factor A, b levels for factor B, and ab levels formed by combinations of A and B. Individual factors are associated with "main" effects, whereas crossed factors create "interaction" effects. If replicates exist for all ab levels, it is a "complete" factorial design; otherwise, it is an "incomplete" factorial design. For the following discussion, I assume that the design is complete.

The factorial design enables one to examine individual factors and their interactions; furthermore, the design provides natural replications that result from crossed factors. Tests of main effects are tests of 1 factor averaged over levels of the other factors. Absence of interaction between 2 factors implies that the additive effect of 1 factor is identical across all levels of the other factor. In that situation, tests and interpretation of main factors are straightforward. If interactions exist, one must interpret main effects cautiously, because relations among mean levels of 1 factor differ according to levels of the second factor. See Figure 3 for an example without interaction (top panel) and with interaction (bottom panel).

Formal definition of the factorial 2-way fixed-effects ANOVA model requires statistical notation to identify specific levels of A and B and of their combination, as well as to

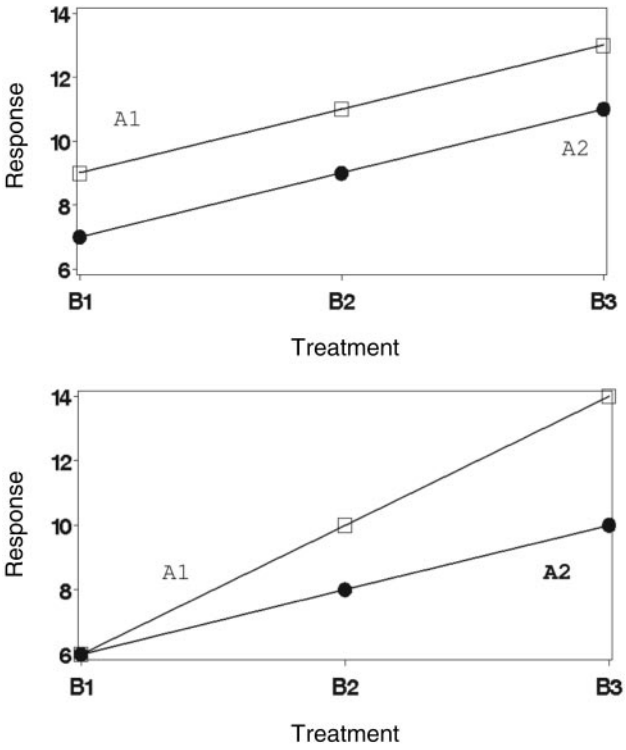

Figure 3. Illustration of interaction effects. Top, No interaction between factors A and B. Bottom, Interaction (synergistic) between factors A and B. Open squares (factor A, level 1) and solid circles (factor A, level 2) represent population mean values at 3 levels of factor B.

denote each replicate within each combination. Conceptually, the model for each observation is as follows:

(individual response)=(grand mean)
+(additive effect for the level of factor A)
+(additive effect for the level of factor B)
+(interaction effect for the combination of levels of A and B)
+(random error).

As with 1-way ANOVA, deviations from the grand mean when expanded algebraically, squared, and summed across levels of both factors produce sums of squares associated with main effects for factor A, main effects for factor B, interaction effects due to combinations of A and B, and random error. Corresponding *df*, mean squares, and F ratios and probability values from hypothesis tests are displayed in Table 4.

Table 5. Descriptive Statistics for Log(sOB-R) by BMI and HDL Cholesterol Categories

	HDL Cholesterol ≤40 mg/dL			HDL Cholesterol >40 mg/dL		
BMI Group, kg/m^2	Sample Size	Mean	SD	Sample Size	Mean	SD
20–24	14	3.61	0.21	46	3.87	0.33
25–29	24	3.45	0.28	38	3.56	0.32
30–34	19	3.45	0.33	21	3.66	0.29
≥35	9	3.27	0.20	17	3.45	0.31

Units for sOB-R are ng/mL.

Some computational algorithms for 2-way ANOVA use formulas that are valid only for complete, balanced factorial designs. In practice, it is common to have unequal numbers in each group, either because the study does not control the numbers of observations or because some response data are missing. When confronted with data from incomplete or unbalanced factorial designs, an investigator must choose a statistical software package that correctly handles the calculations.

Example 2: Two-Way ANOVA

Here, I use the data set from the prior example with men classified by BMI category and by high-density lipoprotein (HDL) cholesterol category (low=HDL ≤40 mg/dL, high=HDL >40 mg/dL). See Table 5 for descriptive statistics; sample sizes vary from n=9 to n=46, SDs vary from 0.20 and 0.33, and the means vary from 3.27 (men with low HDL, very obese) to 3.87 (men with high HDL, normal BMI). Box plots (Figure 4) show that the data distributions are reasonably symmetrical and that interquartile ranges are roughly equal across BMI×HDL groups. Furthermore, variances are homogeneous (Levene's test, $P=0.82$), and the data are approximately normal (Shapiro-Wilk test, $P=0.0025$ in men with low HDL and BMI 30 to 34 kg/m^2, but $P=0.07$ to 0.89 in other groups).

Table 6 shows results from the 2-way ANOVA model with interaction that was fitted with the SAS GLM (general linear model) procedure.[12] Because of the highly unbalanced design, typical ANOVA calculations (eg, SAS ANOVA procedure[7]) would produce incorrect results. Table 6 displays type III sums of squares and F tests. Type III sums of squares are

Table 4. Display of Results for 2-Way Fixed-Effects ANOVA

Source	*df*	Sums of Squares	Mean Square	F Statistic	*P*
Factor A	$(a-1)$	SSA	SSA/$(a-1)$	MSA/MSE	P_A
Factor B	$(b-1)$	SSB	SSB/$(b-1)$	MSB/MSE	P_B
A*B interaction	$(a-1)(b-1)$	SSAB	SSAB/$[(a-1)(b-1)]$	MSAB/MSE	P_{AB}
Error	$N-ab$	SSE	SSE/$(N-ab)$	...	...
Total	$N-1$	SST	...	...	...

Note that P_A, P_B, and P_{AB} are the respective probabilities that an $\mathcal{F}$ random variable with appropriate *df* (source *df*, error *df*) exceeds the observed F statistic.

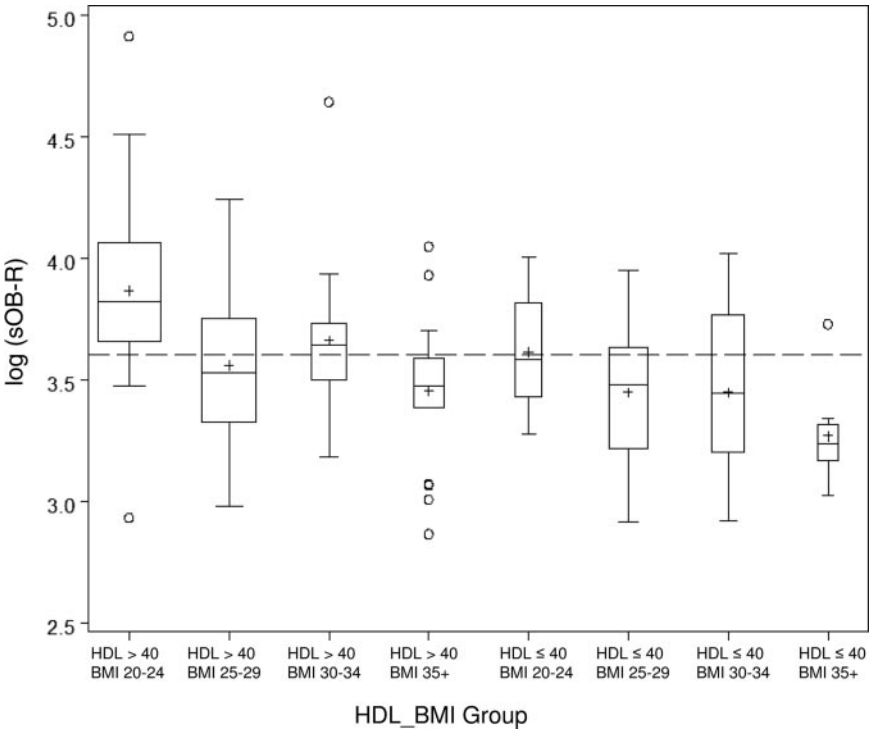

Figure 4. Box plots of log(sOB-R) levels by HDL cholesterol group and BMI group for 188 men in the Framingham Third Generation Cohort. Box width is proportional to sample size (Table 5). Units are ng/mL for sOB-R, mg/dL for HDL, and kg/m^2 for BMI.

preferred in analyses of unbalanced designs, because these statistics are calculated for each factor or interaction after adjustment for all other effects in the model; they do not depend on the ordering of variables. In this example, the main effects are highly statistically significant with regard to both the BMI group (F=9.18, 3 and 180 *df*, $P<0.0001$) and the HDL group (F=14.67, 1 and 180 *df*, $P=0.0002$), but the BMI×HDL interaction is not significant (F=0.51, 3 and 180 *df*, $P=0.67$). When the interaction is not statistically significant, it is common to refit the model with the exclusion of the interaction term to simplify the interpretation of main effects. Here, one concludes that levels of log(sOB-R) are lower in men with low HDL than in men who have higher HDL, that levels of log(sOB-R) tend to decrease across BMI groups, and that the pattern of decrease in log(sOB-R) across BMI groups is similar in both HDL groups.

Study Design, Effect Size, Sample Size, and Statistical Power

Principles that guide the design of randomized, controlled trials include a clear statement of study objective, choice of experimental design, selection of treatments, randomization of subjects to treatments, and a priori determination of sample size to achieve adequate statistical power.[13] Here, I illustrate the interplay of treatment effect size, sample size, and statistical power. Effect size is a measure of scaled differences among population means, and power is the probability of detecting a nonzero effect if one exists.

Table 6. Results of 2-Way ANOVA for Log(sOB-R) by BMI and HDL Cholesterol Categories

Source	*df*	Sums of Squares, Type III	Mean Square	F Statistic	*P*
BMI group	3	2.51	0.838	9.18	<0.0001
HDL group	1	1.34	1.338	14.67	0.0002
BMI*HDL	3	0.14	0.047	0.51	0.67
Error	180	16.42	0.091	...	...
Total	187	22.17	...	...	...

In 1-way ANOVA, power depends on the number of treatments, the sample size distribution among groups, the true effect size, the error variance, and the statistical significance level for the hypothesis test. I consider a simple case of a balanced design having $a=3$ groups with n observations per group and $\alpha=0.05$ significance level. Furthermore, I adopt a common convention that defines effect size by $\delta=$(maximum population mean−minimum population mean)/σ, where σ is the within-population SD. By defining effect size relative to σ, we eliminate σ from subsequent calculations. This convention also sets the intermediate population mean exactly halfway between the smallest and largest means, such that rescaled population means may be represented with values $-\delta/2$, 0, and $+\delta/2$. Once all required design features have been specified, statistical power may be calculated with formulas and charts from textbooks,[3] special statistical software,[14] or online power calculators.[15]

Figure 5 displays power for selected sample sizes from n=5 to n=50 per group and effect sizes from $\delta=0.20$ to $\delta=1.20$ for conditions just described. Increased sample size or effect size results in higher power. Sample size n=50 per group provides good power (say, 0.80) if true effect size is $\delta=0.63$, but a study with n=15 per group has power 0.80 only if effect size is $\delta=1.18$, and a study with n=5 per group has power 0.80 only if effect size is very large, $\delta=2.24$ (not shown on graph). Also, power is higher for balanced designs than for unbalanced and with few rather than many treatment

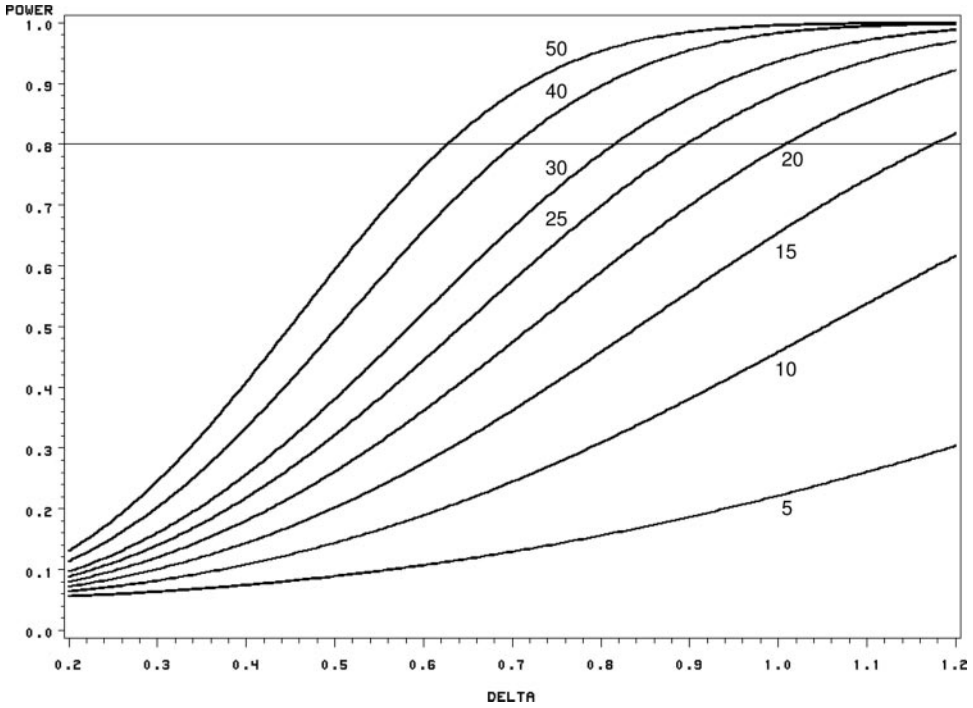

Figure 5. Power in 1-way ANOVA as a function of sample size (n per group) and effect size (δ). Significance level is 0.05; population means are $-\delta/2$, 0, and $\delta/2$; and $\sigma=1$.

groups. Experiments should be designed to have reasonable power (typically set at 0.80) to detect realistic treatment differences, because inadequately powered experiments usually yield inconclusive results.

Acknowledgments

Data on sOB-R levels were kindly provided by Dr Vasan S. Ramachandran.

Sources of Funding

Salary support and examination data were provided by contract NO1 HC 25195 (Principal Investigator P.A. Wolf) from the National Heart, Lung, and Blood Institute, National Institutes of Health. sOB-R levels were measured with support from grant K24 HL04334 (Principal Investigator V.R. Ramachandran), National Heart, Lung, and Blood Institute, National Institutes of Health.

Disclosures

None.

References

1. Fisher RA. *Statistical Methods for Research Workers.* Edinburgh, United Kingdom: Oliver & Boyd; 1925.
2. Kleinbaum DG, Kupper LL, Muller KE. *Applied Regression Analysis and Other Multivariable Methods.* 2nd ed. Boston, Mass: PWS-Kent Publishing; 1988.
3. Zar JH. *Biostatistical Analysis.* Upper Saddle River, NJ: Prentice Hall; 1999.
4. Sit V. *Analyzing ANOVA Designs: Biometrics Information Handbook No. 5.* Province of British Columbia, Ministry of Forests Research Program. Working paper 07/1995. Available at: http://www.for.gov.bc.ca/hfd/pubs/docs/Wp/Wp07.pdf. Accessed July 25, 2007.
5. National Institute of Standards and Technology, Information Technology Library. *NIST/SEMATECH e-Handbook of Statistical Methods*. Available at: http://www.itl.nist.gov/div898/handbook/prc/section4/prc43.htm. Accessed July 25, 2007.
6. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB Sr, Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol*. 2007;165:1328–1335.
7. SAS Institute Inc. *SAS/STAT User's Guide, Version 8.* Cary, NC: SAS Institute; 1999:337–392.
8. Levene H. Robust tests for the equality of variance. In: Olkin I, ed. *Contributions to Probability and Statistics.* Palo Alto, Calif: Stanford University Press; 1960:278–292.
9. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. 1965;52:591–611.
10. Welch BL. On the comparison of several mean values: an alternative approach. *Biometrika*. 1951;38:330–336.
11. Kruskal WH, Wallis WA. Use of ranks in one-criterion analysis of variance. *J Am Stat Assoc*. 1952;47:583–621.
12. SAS Institute Inc. *SAS/STAT User's Guide, Version 8.* Cary, NC: SAS Institute; 1999:1465–1636.
13. Stanley K. Design of randomized controlled trials. *Circulation*. 2007;115:1164–1169.
14. Friendly M. *Power Computations for ANOVA Designs* [computer software]. Version 1.2. Toronto, Canada: York University; 2006. Available at: http://www.math.yorku.ca/SCS/sasmac/fpower.html. Accessed July 26, 2007.
15. Lenth RV. *Java Applets for Power and Sample Size* [computer software]. Iowa City, Iowa: University of Iowa; 2006. Available at: http://www.stat.uiowa.edu/rlenth/Power. Accessed July 26, 2007.

KEY WORDS: ANOVA ■ epidemiology ■ statistics