CHAPTER 11

Miscellaneous Topics

In this chapter we briefly cover several topics that have not been covered to any extent, if at all, in previous chapters.

11.1 CASE–CONTROL STUDIES

A case–control study is an analytical study in which a group of people with a certain disease are compared with a group that does not have the disease, with the comparison made on one or more characteristics. These are retrospective, non-randomized studies and because of this they do not have the value of clinical trials and are generally rated as low quality. See Ury (1975) for a discussion of the efficiency of case–control designs. One option for determining sample size for a matched case–control study is the user-written command `sampsi_mcc` for Stata, which implements the approach described in Dupont (1988).

Satten and Kupper (1990) presented the calculation of sample sizes for unmatched case–control (or cohort) studies in which the objective is interval estimation of the odds ratio. Taylor (1986) discussed determining the number of controls to use in a case–control study. Many other papers have been written on determining sample size for case–control studies, including M'Lan, Joseph, and Wolfson (2006), Hanley, Csizmadi, and Collet (2005), De Santis, Pacifico, and Sambucini (2004), Gauderman (2002), Qiu, Moeschberger, Cooke, and Goldschmidt-Clemont (2000), Garcia-Closas and Lubin (1999), Foppa and Spiegelman (1997), Nam (1992), Hwang, Beaty, Liang, Coresh, and Khoury (1994), and Lubin, Gail, and Ershow (1987).

Sample Size Determination and Power, First Edition. Thomas P. Ryan.

11.2 EPIDEMIOLOGY

Cheng, Branscum, and Stamey (2010b) examined the effect of measurement error and misclassification of a response variable and advised researchers to account for these (potential) problems in sample size determination and power calculations of epidemiologic data. See also Smith (1997), who considered sample size determination and power for epidemiologic surveillance. Luan, Wong, Day, and Wareham (2001) considered sample size determination for investigating potential gene–gene and gene–environment interactions. See also Kasiulevičius, Šapoka, and Filipaviciute (2006), which is a somewhat general paper.

11.3 LONGITUDINAL STUDIES

Longitudinal studies are studies of a group of subjects over time, with the group typically being small because of the cost of such studies, which can last for years and sometimes decades. Therefore, it is obviously important that sample sizes should only be as large as necessary for the stated objectives of the study.

Much of the work on sample size determination for longitudinal studies has been performed by Basagaña and Spiegelman (2010a,b, 2011a,b), with a User's Manual for an R program that is available at `http://hsph.harvard.edu/spiegelman/optitxs/optitxs_user_manual.pdf`. Shieh (2003) considered power and sample size calculations for repeated measures and longitudinal studies within the framework of a multivariate general linear model. See also sample size for longitudinal studies as discussed by Barrera-Gomez, Spiegelman, and Basagaña (2012), Donohue, Edland, and Gamst (2010), Lu, Mehrotra, and Liu (2009), Roy, Bhaumik, Aryal, and Gibbons (2007), Tu, Kowalski, Zhang, Lynch, and Crits-Christoph (2004), and Hedeker, Gibbons, and Waternaux (1999). Dawson (1998) illustrated sample size determination using summary statistics and how sample size is affected by summary statistic choice and other factors using an example with longitudinal data.

A case–cohort study is a form of a longitudinal study. Cai and Zeng (2004) considered sample size and power calculations for such studies. See also Kubota (2011). Falagas, Kouranos, Michalopoulos, Rodopoulou, Athanasoulia, and Karageorgopoulos (2010) reported that many comparative cohort studies have not had adequate power. Musonda, Farrington, and Whitaker (2006) derived several sample size formulas for the self-controlled case series method, which is a modified cohort method.

The LBPOWER module in Stata can be used to calculate either sample size or approximate power for longitudinal studies with a binary response and two treatment groups of equal size, provided that the log odds can be assumed to be linear and there is a constant correlation between repeated measurements.

11.4 MICROARRAY STUDIES

One strategy in determining sample size in microarray experiments is to look at each gene separately and determine the desired sample size for that gene, then combine the results to determine the overall sample size relative to the desired overall power. Mukherjee, Tamayo, Rogers, Rifkin, Engle, Campbell, Golub, and Mesirov (2004) considered estimating dataset size requirements for classifying microarray data using learning curves.

As far as multiple comparison tests are concerned, microarray experiments do present special problems because of the sheer volume of data that is involved. Müller, Parmigiani, Robert, and Rousseau (2004) considered the determination of optimal sample size for experiments that involve, in their words, "massive" multiple comparisons. Shao and Tseng (2007) also considered sample size determination for multiple comparisons, as did Jung, Bang, and Young (2005). Orr and Liu (2009) described a package in R, `ssize.fdr`, that can be used for determining sample size.

For additional information on sample sizes for microarray studies, see Liu and Hwang (2007), Tibshirani (2006), Dobbin and Simon (2005), Wei, Li, and Bumgartner (2004), Hwang, Schmitt, Ge. Stephanopoulos, and Gr. Stephanopoulos (2002), Lee and Whitmore (2002), and Chapter 12 of Chow, Shao, and Wang (2008b).

11.5 RECEIVER OPERATING CHARACTERISTIC ROC CURVES

Obuchowski and McClish (1997) considered sample size determination when ROC curves are used in diagnostic accuracy studies. See also Obuchowski (1998).

11.6 META-ANALYSES

Computation of power in a meta-analysis involves similar steps as in power computations for which a single sample is to be obtained (Hedges and Pigott, 2001, 2004). For example, we need the number of studies in the review (the "sample size"), the anticipated overall mean effect size, and the criterion for statistical significance. Meta-analysis in clinical trials is covered in Section 7.6 and power analysis for meta-analysis is covered in detail in Chapter 29 of Borenstein, Hedges, Higgins, and Rothstein (2009). See also Cohn and Becker (2003), who showed that fixed-effects meta-analysis does increase power.

11.7 SEQUENTIAL SAMPLE SIZES

In many statistical applications a single sample size is not determined. This was discussed in Chapter 6 for the use of a sequence of at least two experimental

designs. Sequential samples are also popular in clinical trial work, as group sequential trials were covered briefly in Section 7.1.12.

11.8 SAMPLE SURVEYS

Surveys are conducted every day, ranging from surveys to gauge consumer preferences to approval ratings for the President of the United States. Because survey results can be highly influential, it is important that they be well designed. Of course, sample size determination is very important in sample surveys, just as it is in sampling in general.

The following quote from Lohr (1999, p. 8) is important.

> But large unrepresentative samples can perform as badly as small unrepresentative samples. A large unrepresentative sample may do more harm than a small one because many people think that large samples are better than small ones. The design of the survey is far more important than the absolute size of the sample.

As discussed by Lohr (1999, p. 241), "design effects are extremely useful for estimating the sample size needed for a survey. The term "design effect" is due to Kish (1965) and is defined as the variance of the estimator (for whatever is being estimated, such as the population mean) under the sampling plan divided by what the variance of that estimator would have been if simple random sampling had been used with the same sample size. Thus, it is a measure of the efficiency of the sampling plan. Lohr (1999, p. 241) stated that if the design effect for a previous, similar survey is known, the sample size if a simple random sample were to be taken could be estimated and the design effect value then multiplied times that number to obtain the sample size necessary for the complex survey design to be used for the survey that is to be conducted.

Tian, Tang, Liu, Tan, and Tang (2011) considered sample size determination for surveys that contain sensitive questions and Bankier (1988) addressed the problem of determining sample sizes for subnational areas in surveys. Johnson, Su, Gardner, and Christensen (2004) developed a Bayesian approach to sample size computations for surveys that are designed to show evidence of freedom from a disease or infectious agent.

Sample size formulas for various types of sampling (ratio estimation, cluster sampling, etc.) are given by Scheaffer, Mendenhall, and Ott (1979). We discuss sample size determination for cluster sampling in Section 11.9. See also the discussion of sample size determination in survey research in Bartlett, Kotrik, and Higgins (2001).

11.8.1 Vegetation Surveys

Sparks, Mountford, Manchester, Rothery, and Treweek (1997) considered sample size determination for estimating species lists in vegetation surveys.

11.9 CLUSTER SAMPLING

When subjects/objects in a population have some common characteristics, data obtained using these subjects will likely be correlated. One consequence of this is that a sample of correlated observations provides less information than a sample obtained from a population of subjects that are relatively heterogeneous. Accordingly, it is desirable to form clusters of heterogeneous objects and to have clusters that are similar. Sampling would be performed by selecting a sample of clusters from the population of clusters that has been formed and then use every element in each sampled cluster.

■ EXAMPLE 11.1

Kerry and Bland (1998) gave an example with the objective to lower cholesterol readings in patients. A new approach was to be studied relative to the standard approach, with the new approach consisting of intensive dietary intervention by practice nurses using a behavioral approach and the standard approach being the usual practice care. The outcome measure was the average cholesterol reading among patients attending each practice after one year. The study participants will have different medical providers and the number of such providers to sample must be determined, as well as the number of patients per medical practice.

Kerry and Bland (1998) started with the sample size formula when two means are to be compared, which they gave as $n = 21s^2/d^2$, where s^2 denotes the variance of the outcome measure, d is the difference between the means that is to be detected, power $= .90$, and $\alpha = .05$. [*Note:* Although this formula has not been given in previous chapters, it results, as was explained in Section 3.6, from Eq. (3.6) if we set $\sigma_1^2 = \sigma_2^2 = s^2$ and use the Z-values for power $= .90$ and $\alpha = .05$; namely, 1.28155 and 1.95996, respectively. Doing so produces $n = 21.0148$, which of course is virtually equal to 21.]

When clusters are involved, there will be a variance between clusters and a variance within each cluster. Kerry and Bland (1998) let s_w^2 denote the within-cluster variance and s_c^2 the variance of a cluster mean from cluster to cluster. With m subjects in each cluster, the variance of a cluster mean is thus s_w^2/m. The total variance is thus $s_w^2/m + s_c^2$, which is the "within" variance plus the "between" variance, using analysis of variance terminology. They obtained estimates of s_w^2 and s_c^2 from a previous study so that $s^2 = s_w^2/m + s_c^2 = 0.0046 + 1.28/50 = 0.0302$. With this estimate, the required sample size is thus $n = 21(0.0302)/(0.1)^2 = 63$. Thus, 63 groups would be sampled with 50 practices in each group for a total of 3150 patients in each group. ■

See also Tokola, Larocque, Nevalainen, and Oja (2011), who addressed sample size determination, power, and sampling costs when there is clustered data and Kumar and Indrayan (2002), who considered determination of the number of clusters for single-stage cluster-sample surveys.

11.10 FACTOR ANALYSIS

Factor analysis is a statistical method for uncovering the latent structure in a set of variables and is essentially a dimension-reduction technique. There are two types of factor analysis: exploratory and confirmatory. Many different views have been expressed about a minimum sample size for factor analysis and these are summarized here: `http://www.encorewiki.org/display/~nzhao/The+Minimum+Sample+Size+in+Factor+Analysis`. Kline (2005) stated that sample size should depend on the number of parameters that are to be estimated. That recommendation was investigated by Jackson (2007). See also Marsh, Balla, and McDonald (1988), Marsh, Hau, Balla, and Grayson (1998), and Jackson (2001). Muthén and Muthén (2002) discussed how to use a Monte Carlo study to determine sample size, using confirmatory factor analysis and a growth model for illustration.

11.11 MULTIVARIATE ANALYSIS OF VARIANCE AND OTHER MULTIVARIATE METHODS

Multivariate methods are applied when there is more than one response variable. Hardly any sample size determination software can be used for determining sample size for multivariate methods, however, and D'Amico, Neilands, and Zambarano (2001) lamented this, while explaining how SPSS can be used to compute power. PASS goes further and can be used to determine both sample size and power for various multivariate methods. Specifically, PASS has capability for Hotelling's T^2 (either one sample or two samples, each with p variables) and MANOVA (Multivariate Analysis of Variance). Thus, there are several multivariate methods that are not included (e.g., cluster analysis and factor analysis) and "multivariate" is not a menu item and the multivariate methods are thus not easy to find, but at least some well-known methods are included. G*Power also has the capability for Hotelling's T^2 and MANOVA. Efird and Alimineti (2004) derived a generalized method for computing the exact power of Hotelling's T^2 test and provided SAS code for accomplishing it. Estimating the covariance matrix with good precision is of paramount important and Gupta and Gupta (1987) gave closed-form sample size formulas for the case of a diagonal matrix and gave an integral equation for use in the general case.

The one-sample T^2 test is the extension of the univariate t-test to more than two means, and in fact the PASS Hotelling's T^2 procedure for one sample could, if desired, be used to determine sample size for a univariate one-sample t-test since one response variable with one group is one of the options. This will be illustrated later.

The MANOVA procedure allows the user to choose one of the following test statistics: the Wilk's Lambda, the Pillai–Bartlett Trace, or the Hotelling–Lawley Trace.

The output of PASS for Hotelling's T^2 procedure includes the "effect size," which is defined, in their notation, as

$$\Delta = \sqrt{(\boldsymbol{\mu}_A - \boldsymbol{\mu}_0)' \sum^{-1} (\boldsymbol{\mu}_A - \boldsymbol{\mu}_0)}$$

with $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ being the multivariate mean under the null hypothesis, $\boldsymbol{\mu}_A$ denoting the multivariate mean that ones wishes to detect with a specified power, and $\sum$ being the assumed variance–covariance matrix, which the user can enter in the spreadsheet.

Power is computed using the noncentrality parameter, which is defined as $\lambda = N\Delta^2$, with power computed using the noncentral F-distribution.

To illustrate, consider the following example.

■ EXAMPLE 11.2

Assume that we wish to test the equality of three means —μ_1, μ_2 andμ_3—with the random variables Y_1, Y_2, and Y_3 assumed to be jointly distributed as a multivariate normal distribution with the following variance–covariance matrix:

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ & \sigma_2^2 & \sigma_{23} \\ \text{sym.} & & \sigma_3^2 \end{bmatrix} = \begin{bmatrix} 10 & 5 & 6 \\ 5 & 15 & 10 \\ 6 & 10 & 18 \end{bmatrix}$$

with "sym" denoting that the matrix is symmetric.

In PASS, Hotelling's T^2 would be accessed by selecting "Means," the "Multivariate Means," then "Hotelling's T-Squared." Use "1" for the number of groups, $\alpha = .05$, 3 for the number of response variables, then click the "Covariance" tab, then click "Spreadsheet," then enter the variance–covariance matrix given above, using the first 3 columns and first 3 rows of the spreadsheet. For "Mean Differences," enter " 2 2 2," signifying that the three means differ from the hypothesized means by two units each. For "Power," enter .80. The result is a sample size of 27 and a power of .8088. ■

Hotelling's T^2 is the multivariate generalization of the univariate one-sample t-test, and the former can be used in PASS to solve for the sample size of the univariate test. Of course, we wouldn't do that in practice, but it is instructive to see the relationship. For example, for Hotelling's T^2 procedure, specify "1" for both the number of groups and the number of response variables, and enter 2 for the mean differences. Use the same values for power and α as before and enter the number "10" in the first row and first column of the spreadsheet. The output shows $n = 22$ and power = .8075. Now select "Means" followed by "One Mean" from the menu and then select "Inequality Tests" and "Specify Using Differences." Enter $3.162(= \sqrt{10})$ for the standard deviation, .80 for power, $\alpha = .05$, 0 for the

hypothesized mean, 2 for the alternative mean, and indicate a two-sided test. The sample size is $n = 22$ and the power = .80752—in agreement with Hotelling's T^2 results.

Other work on sample size determination in a multivariate setting includes Liu and Liang (1997), who gave a multivariate extension of the work of Self and Mauritsen (1988) for generalized linear models that was cited in Chapter 5, and Wolynski (2005), who considered sample size determination for a Bayes classification procedure. Efird and Alimineti (2005) derived a method for computing power, based on Hotelling's T^2 statistic, for a multivariate repeated measurements design. See also Muller and Peterson (1984), who examined methods of computing power for testing the multivariate general linear hypothesis and Maccallum, Browne, and Sugawara (1996), who considered sample size determination for covariance structure modeling. Muller, LaVange, Ramey, and Ramey (1992) considered power calculations for general linear multivariate models. O'Brien and Shieh (1999) provided an algorithm for computing power probabilities for commonly used F-tests of the multivariate general linear hypothesis. The user-written Stata command `mvsampsi` can be used to compute power or sample size for multivariate F-tests that are derived from Wilks' lambda.

11.12 STRUCTURAL EQUATION MODELING

Structural equation modeling is essentially a combination of factor analysis and path analysis (see, e.g., Olobatuyi, 2006, for the latter). As discussed by Westland (2010), sample size determination for structural equation modeling has been quite a challenge, so it is not surprising that this paper discusses lower bounds on sample size rather than the determination of specific sample sizes. Enders and Peugh (2004) considered the choice of an adjusted sample size for estimating structural equation models with missing data through the use of an EM covariance matrix. Hancock and Freeman (2001) considered sample size for use with an approximation test of not close fit in structural equation modeling. Jackson (2003) investigated the suggestion of some authors that an adequate sample size is related to the number of parameters to be estimated and found only weak support for this contention.

Although structural equation modeling is relatively popular, it is nonetheless a specialized topic and thus it is not surprising that there is almost no available software for determining sample size for such models. Daniel Soper (`www.danielsoper.com`) has developed many statistical applets, including many for sample size determination. This includes one for structural equation modeling, which can be accessed from `www.danielsoper.com/statcalc3/calc.aspx?id=89`. The input includes the anticipated effect size, target power, significance level, number of latent variables, and number of observed variables. For example, if these are input as 0.9, 0.8, .05, 2, and 8,

respectively, the applet gives 2 as the minimum sample size to detect the effect, and 100 as both the minimum sample size for the model structure and the recommended minimum sample size.

11.13 MULTILEVEL MODELING

The term "multilevel," as in multilevel modeling, is used in different ways, including designating different levels of analysis, such as individuals that are nested within spatial units. Sample size and power issues in multilevel modeling are discussed here, `http://www.esourceresearch.org/eSourceBook/MultilevelModeling/17PowerandSampleSize/tabid/353/Default.aspx`, and in Maas and Hox (2005). See also Moerbeek and Wong (2008), Cohen (2005), Snijders (2005), and Snijders and Bosker (1993). MLPowSim is freeware that can be used for sample size and power calculations in multilevel modeling (see `http://www.bristol.ac.uk/cmm/software/mlpowsim`). A recommended book source for information on multilevel modeling is Goldstein (2011).

11.14 PREDICTION INTERVALS

Prediction intervals are frequently used since a predicted value is of limited use without knowing the uncertainty associated with the predicted value, and that uncertainty is built into a prediction interval, just as it is built into a confidence interval. As is true with a confidence interval, a wide prediction interval is of limited value.

Wallis (1980) quoted W. Edwards Deming (1900–1993) as stating: "The only useful function of a statistician is to make predictions, and thus to provide a basis for action." Many users of statistics know prediction intervals only as an interval about a predicted value in regression analysis but there are other types of prediction intervals.

Assume that observations have approximately a normal distribution and there is interest in obtaining a prediction interval for a single future observation. If σ is assumed unknown and to be estimated from a sample, the prediction interval has lower bound given by $\bar{x} - t_{\alpha,n-1}s\sqrt{(1+1/n)}$ and an upper bound obtained from $\bar{x} + t_{\alpha,n-1}s\sqrt{(1+1/n)}$. (This result is derived in the chapter Appendix.) The width of the interval is thus the upper bound minus the lower bound, with that expression being $2t_{\alpha/2,n-1}s\sqrt{(1+1/n)}$. Of course, the general idea is to select the sample size so that the result for whatever is being computed will be acceptable. Here we have the same type of problem that is encountered in trying to determine the sample size for hypothesis tests: the width depends on *s,* which would be computed *from* a sample, not *before* the sample is taken. This problem

might be addressed by estimating σ from the range of possible observations, as was discussed in Section 2.1. If that is done, $t_{\alpha/2,n-1}$ would be replaced by $z_{\alpha/2}$. The sample size expression would then be

$$n = \frac{4\, z_{\alpha/2}^2\, \widehat{\sigma}^2}{w^2 - 4\, z_{\alpha/2}^2\, \widehat{\sigma}^2}$$

with w denoting the width of the prediction interval.

11.15 MEASURES OF AGREEMENT

Sample size for testing a Pearson correlation coefficient was given in Section 5.7, with tests on the equality of two Pearson correlations given in Section 5.7.2 and for intraclass correlation in Section 5.7.1. The Spearman rank correlation coefficient and Kendall's tau were covered in Section 10.8. Correlation coefficients are in the category of *measures of agreement* and there are many such measures. Another such measure is the *kappa coefficient* for testing agreement between two raters. Donner and Rotondi (2010) considered sample size determination when the kappa statistic is used with multiple raters. Flack, Afifi, Lachenbruch, and Schouten (1988) gave a method for determining a sample size that will give a specified bound on the width of a confidence interval for kappa and also presented a table of sample sizes for a power of .80. Sample size determination for testing agreement is available in PASS, SiZ, and nQuery Advisor.

11.16 SPATIAL STATISTICS

Spatial statistics is a relatively new area of statistics. Before the issue of sample size determination can be addressed, the determination must first be made of the effective sample size in the presence of spatial correlations. There appears to be ongoing research on that problem (`http://www.am2v.cl/index.php?option=com_jresearch&view=researcharea&id=7&task=show&Itemid=154`).

11.17 AGRICULTURAL APPLICATIONS

Agricultural infestations in California, for example, can be a serious problem. Schwertman and Smith (1998) addressed the question of determining an optimal sample size for early detection of the problem. Yamamura and Ishimoto (2009)

considered the determination of sample size for estimating the proportion of pecky rice grains when there is composite sampling with subsampling.

11.18 ESTIMATING THE NUMBER OF UNSEEN SPECIES

There has long been a need for estimating the total number of unseen species of various types, such as the number of deer in a forest. Zhang and Stern (2009) addressed the issue of sample size determination for finding unseen species and used a Bayesian approach.

11.19 TEST RELIABILITY

Test reliability refers generally to the ability of a test instrument to measure what it purports to measure. The coefficient alpha (α) is a commonly used index of test reliability. Sample size determination for estimating alpha has been considered by Bonett (2002), with Bonett (2003) and Feldt and Ankenmann (1999) considering sample size determination for comparing two alpha coefficients. See also sample size determination for test reliability as discussed by Shoukri, Asyali, and Donner (2004) and Walter, Eliasziw, and Donner (1998).

11.20 AGREEMENT STUDIES

Agreement studies are similar to test reliability in the sense that measurement methods, say, can be compared with agreement of course being the objective. Liao (2010) proposed a method of determining sample size for such studies.

11.21 GENOME-WIDE ASSOCIATION STUDIES

The objective of a genome-wide association study is to find genetic variations associated with a particular disease. Sample size determination for genome-wide association studies has been considered by Xie, Cai, and Li (2011) and other authors. Pirinen, Donnelly, and Spencer (2012) show that the inclusion of known covariates in a model such as a logistic regression model can, in the case of rare diseases, reduce the power to detect new genetic associations, whereas the power can be increased when a disease is common. Wang and Zhao (2003) discussed sample size requirements for detecting gene–gene interactions. See also Edwards, Haynes, Levenstien, Finch, and Gordon (2005) and Kozlitina, Xing, Pertsemlidis, and Schucany (2010).

11.22 NATIONAL SECURITY

Biometrics have been used during the past several years in national security work. Wu and Wilson (2005) considered sample size determination when biometrics are used with fingerprint data.

11.23 MISCELLANEOUS

Pennington and Volstad (1991) showed that a smaller sampling unit can be more efficient than a larger unit in marine abundance surveys and determined the sample size that produces the most precise density estimate given a fixed amount of survey resources or the sample size that minimizes the cost for a given level of precision. Bochmann, Johnson, and Azuara-Blanco (2007) reported that sample size calculations were not being reported in the ophthalmology literature for diagnostic performance studies. Branscum, Johnson, and Gardner (2007) developed a Bayesian approach to determining sample size for evaluating diagnostic test accuracy and Buderer (1996) considered sample size determination for estimating the sensitivity and specificity of a diagnostic test. Carley, Dosman, Jones, and Harrison (2005) stated that sample size calculations are performed infrequently in diagnostic studies and provided some simple nomograms as aids in sample size determination. Kosinski, Chen, and Lyles (2010) provided a method for determining sample size for evaluating a diagnostic test when the gold standard for the test is missing at random. Liu, Schisterman, Mazumbar, and Hu (2005) considered power and sample size calculation of diagnostic studies with multiple correlated test results.

Cheng, Branscum, and Stamey (2010a) developed a Bayesian procedure for determining sample size and power for cross-sectional studies designed to evaluate and compare continuous medical tests. Although the worth of internal pilot studies in designed experimentation is well accepted, Gurka, Coffey, and Gurka (2010) demonstrated how they can also have value in observational studies.

Analysis of molecular variance (AMOVA) is widely used in genetic data analysis for quantifying the contribution of various levels of population structure to patterns of genetic variation. Fitzpatrick (2009) considered power and sample size determination for nested AMOVA. Gadbury, Page, Edwards, Kayo, Prolla, Weindruch, Permana, Mountz, and Allison (2004) considered sample size and power estimation in high dimensional biology. Ein-Dor, Zuk, and Domany (2006) discussed the need for thousands of samples to generate a robust gene list for predicting cancer outcomes in clinical research.

Cai, Lin, and Lee (2010) considered sample size determination for high-dimensional data analyses and Edwards, Mohai, Halverson, and DeWalle (1989) considered sample size determination for throughfall chemistry. Meadows-Shropshire, Gennings, and Carter (2005) considered the determination of sample size for detecting interactions of drugs or chemicals. Specifically, they determined

the number of observations to use at specific mixture points of interest. Hayasaka, Peiffer, Hugenschmidt, and Laurienti (2007) considered power and sample size in neuroimaging studies and Li and Fine (2004) considered sample size for sensitivity and specificity in prospective accuracy studies. Liu, Schisterman, and Teoh (2004) developed procedures for power and sample size and power calculations for planning a study comparing the accuracy of biomarkers in diagnosing diseases. Noble, Bailer, Kunkel, and Straker (2006) considered sample size determination for estimating parameters in small populations.

Voting irregularities and miscounts unfortunately often occur, necessitating some type of audit. McCarthy, Stanislevic, Lindeman, Addona, and Batcher (2008) addressed the question of whether percentage-based audits should be used, or if it is preferable to take a power-based approach to determining sample size. They defined the power of an audit as being equal to the probability of sampling at least one miscounted precinct whenever there are enough precincts with miscounts to have altered the outcome. They pointed out that percentage audits (such as a 10% audit) have limited power and urged the use of power-based audits.

The QT interval is a type of heart measurement. Zhang, Dmitrienko, and Luta (2008) and Chow, Cheng, and Cosmatos (2008a) considered sample size determination for QT studies.

Although a distribution must be specified when sample size determination software is used, Mahnken (2009) presented a method for determining sample size and power for moderate-to-large studies that use quasi-likelihood methods, with the only requirement being that the variance as a functional proportion of the mean be known.

Moon, Lee, Ahn, and Nikolova (2002) gave a Web-based simulator for determining sample size and power in animal carcinogenicity studies. Figueroa, Zeng-Treitler, Kandula, and Ngo (2012) considered sample size determination for predicting a classifier's performance and Guo, Graber, McBurney, and Balasubramanian (2010) considered sample size and power for comparing classification algorithms. Greselin and Maffenini (2007) considered minimum sample sizes for constructing confidence intervals for Gini's mean difference.

Bartolucci, Bae, and Singh (2006) considered sample size determination for stratified random sampling of pond water using a Bayesian approach.

11.24 SUMMARY

Sample size determination in conjunction with many statistical methods and tests has been presented in this chapter and previous chapters. Some familiarity with sample size determination and power is desirable for people in a wide variety of professions and tutorials have been given in the literature to provide the basics (see, e.g., Rempher and Miller, 2008).

REFERENCES

Bankier, M. D. (1988). Power allocation: Determining sample sizes for subnational areas. *The American Statistician*, **42**(3), 174–177.

Barrera-Gomez, J., D. Spiegelman, and X. Basagaña (2012). Optimal combination of number of participants and number of repeated measurements in longitudinal studies with time-varying exposure. Submitted to *Statistics in Medicine*.

Bartlett, J. E. II, J. W. Kotrik, and C. C. Higgins (2001). Organizing research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, **19**(1), 43–50.

Bartolucci, A. A., S. J. Bae, and K. P. Singh (2006). A Bayesian method of computing sample size and cost requirements for stratified random sampling of pond water. *Environmental Modelling & Software*, **21**(9), 1319–1323.

Basagaña, X. and D. Spiegelman (2010a). Power and sample size calculations for longitudinal studies estimating a main effect of a time-varying exposure. *Statistical Methods in Medical Research*, **19**(3), 1–17.

Basagaña, X. and D. Spiegelman (2010b). Power and sample size calculations for longitudinal studies comparing rates of change with a time-varying exposure. *Statistics in Medicine*, **29**(2), 181–192.

Basagaña, X, and D. Spiegelman (2011a). The design of observational longitudinal studies with a time-invariant exposure. Submitted for publication, *International Statistical Review*, July 2011.

Basagaña, X. and D. Spiegelman (2011b). The design of observational longitudinal studies. Manuscript. Part of doctoral thesis of first author, Harvard School of Public Health. (Available at `http://www.hsph.harvard.edu/donna-spiegelman/files/optixs_the_design_of_observational_longitudinal_studies.pdf`).

Basagaña, X., X. Liao, and D. Spiegelman (2011). Power and sample size calculations for longitudinal studies estimating a main effect of a time-varying response. *Statistical Methods in Medical Research*, **20**, 471–487.

Bochmann, F., Z. Johnson, and A. Azuara-Blanco (2007). Sample size in studies on diagnostic accuracy in ophthalmology: A literature survey. *British Journal of Ophthalmology*, **91**, 898–900.

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics*, **27**(4), 335–340.

Bonett, D. G. (2003). Sample size requirements for comparing two alpha coefficients. *Applied Psychological Measurement*, **27**, 72–74.

Borenstein, M., L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein (2009). *Introduction to Meta-Analysis*, Chapter 29. Hoboken, NJ: Wiley.

Branscum, A. J., W. O. Johnson, and I. A. Gardner (2007). Sample size calculations for studies designed to evaluate diagnostic test accuracy. *Journal of Agricultural, Biological, and Environmental Statistics*, **12**(1), 112–127.

Buderer, N. M. (1996). Statistical methodology: 1. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Academic Emergency Medicine*, **3**(9), 895–900.

Cai, J. and D. Zeng (2004). Sample size/power calculation for case–cohort studies. *Biometrics*, **60**, 1015–1024.

Cai, G., X. Lin, and K. Lee (2010). Sample size determination with false discovery rate adjustment for experiments with high-dimensional data. *Statistics in Biopharmaceutical Research*, **2**(2), 165–174.

Carley, S., S. Dosman, S. R. Jones, and M. Harrison (2005). Simple nomograms to calculate sample size in diagnostic studies. *Emergency Medical Journal*, **22**, 180–181.

Cheng, D., A. J. Branscum, and J. D. Stamey (2010a). A Bayesian approach to sample size determination for studies designed to evaluate continuous medical tests. *Computational Statistics and Data Analysis*, **54**, 298–307.

Cheng, D., A. J. Branscum, and J. D. Stamey (2010b). Accounting for response misclassification and covariate measurement error improves power and reduces bias in epidemiologic studies. *Annals of Epidemiology*, **20**(7), 562–567.

Chow, S.-C., B. Cheng, and D. Cosmatos (2008a). On power and sample size calculation for QT studies with recording replicates at given time point. *Journal of Biopharmaceutical Statistics*, **18**(3), 483–493.

Chow, S.-C., J. Shao, and H. Wang (2008b). *Sample Size Calculations in Clinical Research*, 2nd edition. Boca Raton, FL: Chapman & Hall/CRC.

Cohen, M. P. (2005). Sample size considerations for multilevel surveys. *International Statistical Review*, **73**(3), 279–287.

Cohn, L. D. and B. J. Becker (2003). How meta-analysis increases statistical power. *Psychological Methods*, **8**(3), 243–253.

D'Amico, E. J., T. B. Neilands, and R. Zambarano (2001). Power analysis for multivariate and repeated measures designs: A flexible approach using the SPSS MANOVA procedure. *Behavior Research Methods, Instruments, and Computers*, **33**(4), 479–484.

Dawson, J. D. (1998). Sample size calculations based on slopes and other summary statistics. *Biometrics*, **54**, 323–330.

De Santis, F., M. P. Pacifico, and V. Sambucini (2004). Optimal predictive sample size for case–control studies. *Journal of the Royal Statistical Society, Series C*, **53**, 427–441.

Dobbin, K. and R. Simon (2005). Sample size determination in microarray experiments for class comparison and prognostic classification. *Biostatistics*, **6**(1), 27–38.

Donohue, M. C., S. D. Edland, and A. C. Gamst (2010). Power for linear models of longitudinal data with applications to Alzheimer's disease Phase II study design. Manuscript.

Donner, A. and M. A. Rotondi (2010). Sample size requirements for interval estimation of the kappa statistic for interobserver agreement studies with a binary outcome and multiple raters. *The International Journal of Biostatistics*, **6**(1), Article 31.

Dupont, W. (1988). Power calculations for matched case–control studies. *Biometrics*, **44**, 1157–1168.

Edwards, B. J., C. Haynes, M. A. Levenstien, S. J. Finch, and D. Gordon (2005). Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. *BMC Genetics*, **6**(18). (Available at `http://www.biomedcentral.com/1471-2156/6/18`.)

Edwards, P. J., P. Mohai, H. G. Halverson, and D. R. DeWalle (1989). Considerations for throughfall chemistry sample-size determination. *Forest Science*, **35**(1), 173–182.

Efird, J. T. and K. Alimineti (2004). Computing exact power and sample size for Hotelling's T^2-test and related multivariate procedures. WUSS 2004 Proceedings, SAS Institute, Pasadena, CA.

Efird, J. T. and K. Alimineti (2005). Computing exact power for multivariate repeated measurements design. WUSS Proceedings, SAS Institute, San Jose, CA.

Ein-Dor, L., O. Zuk, and E. Domany (2006). Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proceedings of the National Academy of Sciences*, **103**(15), 5923–5928.

Enders, C. K. and J. L. Peugh (2004). Using an EM covariance matrix to estimate structural equation models with missing data: Choosing an adjusted sample size to improve the accuracy of inferences. *Structural Equation Modeling*, **11**, 1–19.

Falagas, M. E., V. D. Kouranos, A. Michalopoulos, S. P. Rodopoulou, A. P. Athanasoulia, and D. E. Karageorgopoulos (2010). Inadequate statistical power of published comparative cohort studies on ventilator-associated pneumonia to detect mortality differences. *Clinical Infectious Diseases*, **50**(4), 468–472.

Feldt, L. S. and R. D. Ankenmann (1999). Determining sample size for a test of the equality of alpha coefficients when the number of part tests is small. *Psychological Methods*, **4**(4), 366–377.

Ferreira, J. A. and A. H. Zwinderman (2006). Approximate sample size calculations with microarray data: An illustration. *Statistical Applications in Genetics and Molecular Biology*, **5**(1), article 25.

Figueroa, R. L., Q. Zeng-Treitler, S. Kandula, and L. H. Ngo (2012). Predicting sample size required for classification performance. *BMC Medical Informatics and Decision Making*, **12**(8), electronic journal.

Fitzpatrick, B. M. (2009). Power and sample size for nested analysis of molecular variance. *Molecular Ecology*, **18**, 3961–3966.

Flack, V. F., A. A. Afifi, P. A. Lachenbruch, and H. J. A. Schouten (1988). Sample size determinations for the two rater Kappa statistic. *Psychometrika*, **53**(3), 321–325.

Foppa, I. and D. Spiegelman (1997). Power and sample size calculations for case–control studies of gene–environment interactions with a polytomous exposure variable. *American Journal of Epidemiology*, **146**(7), 596–604.

Gadbury, G. L., G. P. Page, J. Edwards, T. Kayo, T. A. Prolla, R. Weindruch, P. A. Permana, J. D. Mountz, and D. B. Allison (2004). Power and sample size estimation in high dimensional biology. *Statistical Methods in Medical Research*, **13**(4), 325–338.

Garcia-Closas, M. and J. H. Lubin (1999). Power and sample size calculations in case–control studies of gene–environment interactions: Comments on different approaches. *American Journal of Epidemiology*, **149**, 689–692.

Gauderman, W. J. (2002). Sample size requirements for matched case–control studies of gene–environment interaction. *Statistics in Medicine*, **21**, 35–50.

Goldstein, H. (2011). *Multilevel Statistical Models*, 4th edition. Hoboken, NJ: Wiley

Greselin, F. and W. Maffenini (2007). Minimum sample sizes for confidence intervals for Gini's mean difference: A new approach for their determination. *Statistica and Applicazioni*, **5**, 103–122.

Guo, Y., A. Graber, R. N. McBurney, and R. Balasubramanian (2010). Sample size and statistical power considerations in high-dimensionality data settings: A comparative study of classification algorithms. *BMC Bioinformatics*, **11**, Article 447.

Gupta, P. L. and R. D. Gupta (1987). Sample size determination in estimating a covariance matrix. *Computational Statistics & Data Analysis*, **5**, 185–192.

Gurka, M. J., C. S. Coffey, and K. K. Gurka (2010). Internal plots for observational studies. *Biometrical Journal*, **52**(5), 590–603.

Hancock, G. R. and M. J. Freeman (2001). Power and sample size for the root mean square error of approximation test of not close fit in structural equation modeling. *Educational and Psychological Measurement*, **61**(5), 741–758.

Hanley, J. A., I. Csizmadi, and J.-P. Collet (2005). Two-stage case–control studies: Precision of parameter estimates and consideration in selecting sample size. *American Journal of Epidemiology*, **162**, 1225–1234.

Hayasaka, S., A. M. Peiffer, C. E. Hugenschmidt, and P. J. Laurienti (2007). Power and sample size calculation for neuroimaging studies by noncentral random field theory. *NeuroImage*, **37**(3), 721–730.

Hedeker, D., R. D. Gibbons, and C. Waternaux (1999). Sample size estimation for longitudinal designs with attrition: Time-related concepts between two groups. *Journal of Educational and Behavioral Statistics*, **24**(1), 70–93.

Hedges, L. V. and T. D. Pigott (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, **6**(3), 203–217.

Hedges, L. V. and T. D. Pigott (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, **9**(4), 424–445.

Hwang, D., W. A. Schmitt, Ge. Stephanopoulous, and Gr. Stephanopolous (2002). Determination of minimum sample size and discriminatory expression patterns in microarray data. *Bioinformatics*, **18**, 1184–1193.

Hwang, S.-J., T. H. Beaty, K.-Y. Liang, J. Coresh, and M. J. Khoury (1994). Minimum sample size estimation to detect gene–environment interaction in case–control designs. *American Journal of Epidemiology*, **140**, 1029–1037.

Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: A Monte Carlo investigation. *Structural Equation Modeling*, **8**(2), 205–223.

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N-q hypothesis. *Structural Equation Modeling*, **10**, 128–141.

Jackson, D. L. (2007). The effect of the number of observations per parameter in misspecified confirmatory factor analytic models. *Structural Equation Modeling*, **14**(1), 48–76.

Johnson, W. O., C.-L. Su, I. A. Gardner, and R. Christensen (2004). Sample size calculations for surveys to substantiate freedom of populations from infectious agents. *Biometrics*, **60**, 165–171.

Jung, S. H., H. Bang, and S. S. Young (2005). Sample size calculation for multiple testing in microarray data analysis. *Biostatistics*, **6**, 157–159.

Kasiulevičius, V., V. Šapoka, and R. Filipaviciute (2006). Sample size calculation in epidemiological studies. *Gerontologija*, **7**(4), 225–231.

Kerry, S. M. and J. M. Bland (1998). Statistics notes: Sample size in cluster randomization. *British Medical Journal*, **316**, 549.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. New York: Guilford.

Kosinski, A. S., Y. Chen, and R. H. Lyles (2010). Sample size calculations for evaluating a diagnostic test when the gold standard is missing at random. *Statistics in Medicine*, **29**, 1572–1579.

Kozlitina, J., C. Xing, A. Pertsemlidis, and W. S. Schucany (2010). Power of genetic association studies with fixed and random genotype frequencies. *Annals of Human Genetics*, **74**(5), 429–438.

Kubota, K. (2011). Sample-size formula for case–cohort studies. *Epidemiology*, **22**(2), 279.

Kumar, R. and A. Indrayan (2002). A nomogram for single-stage cluster-sample surveys in a community for estimation of a prevalence rate. *International Journal of Epidemiology*, **31**(2), 463–467.

Lee, M. L. and G. A. Whitmore (2002). Power and sample size for DNA microarray studies. *Statistics in Medicine*, **21**(23), 3543–3570.

Li, J. and J. Fine (2004). On sample size for sensitivity and specificity in prospective diagnostic accuracy studies. *Statistics in Medicine*, **23**, 2537–2550.

Liao, J. J. Z. (2010). Sample size calculation for an agreement study. *Pharmaceutical Statistics*, **9**, 125–132.

Liu, A., E. F. Schisterman, and E. Teoh (2004). Sample size and power calculation in comparing diagnostic accuracy of biomarkers with pooled assessments. *Journal of Applied Statistics*, **31**(1), 49–59.

Liu, A., E. F. Schisterman, M. Mazumbar, and J. Hu (2005). Power and sample size calculation of comparative diagnostic accuracy studies with multiple correlated test results. *Biometrical Journal*, **47**(2), 140–150.

Liu, G. and K. Liang (1997). Sample size calculations for studies with correlated observations. *Biometrics*, **53**, 937–947.

Liu, J.-P., H. Hsueh, and J. J. Chen (2002). Sample size requirements for evaluation of bridging evidence. *Biometrical Journal*, **44**, 969–981.

Liu, P. and J. T. G. Hwang (2007). Quick calculation for sample size while controlling false discovery rate with application to microarray analysis. *Bioinformatics*, **23**(6), 739–746.

Lohr, S. L. (1999). *Sampling: Design and Analysis*. North Scituate, MA: Duxbury.

Lu, K., D. V. Mehrotra, and G. Liu (2009). Sample size determination for constrained longitudinal data analysis. *Statistics in Medicine*, **28**, 679–699.

Luan, J. A., M. Y. Wong, N. E. Day, and N. J. Wareham (2001). Sample size determination for studies of gene–environment interaction. *International Journal of Epidemiology*, **30**(5), 1035–1040.

Lubin, J. H., M. H. Gail, and A. G. Ershow (1987). Sample size and power for case–control studies when exposures are continuous. *Statistics in Medicine*, **7**(3), 363–376.

M'Lan, C. E., L. Joseph, and D. B. Wolfson (2006). Bayesian sample size determination for case–control studies. *Journal of the American Statistical Association*, **101**(474), 760–772.

Maas, C. J. and J. J. Hox (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, **1**(3), 86–92.

Maccallum, R. C., M. W. Browne, and H. M. Sugawara (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, **1**(2), 130–149.

Mahnken, J. D. (2009). Power and sample size calculations for models from unknown distributions. *Statistics in Biopharmaceutical Research*, **1**, 328–336.

Marsh, H. W., J. R. Balla, and R. P. McDonald (1988). Goodness of fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, **103**, 391–410.

Marsh, H. W., K.-T. Hau, J. R. Balla, and D. Grayson (1998). Is more ever too much? The number of indicators per factor in exploratory factor analysis. *Multivariate Behavioral Research*, **33**, 181–220.

McCarthy, J., H. Stanislevic, H. M. Lindeman, V. Addona, and M. Batcher (2008). Percentage-based versus statistical-power-based vote tabulation audits. *The American Statistician*, **62**(1), 11–16.

Meadows-Shropshire, S. L., C. Gennings, and W. H. Carter, Jr. (2005). Sample size and power determination for detecting interactions in mixtures of chemicals. *Journal of Agricultural, Biological, and Environmental Statistics*, **10**(1), 104–117.

Moerbeek, N. and W. K. Wong (2008). Sample size formula for trials comparing group and individual treatments in a multilevel model. *Statistics in Medicine*, **27**, 2850–2864.

Moon, H., J. J. Lee, H. Ahn, and R. G. Nikolova (2002). A web-based simulator for sample size and power estimation in animal carcinogenicity studies. *Journal of Statistical Software*, **7**, 1–36.

Mukherjee, S., P. Tamayo, S. Rogers, R. Rifkin, A. Engle, C. Campbell, T. R. Golub, and J. P. Mesirov (2004). Estimating dataset size requirements for classifying DNA microarray data. *Journal of Computational Biology*, **10**(2), 119–142.

Muller, K. E. and B. L. Peterson (1984). Practical methods for computing power in testing the multivariate general linear hypothesis. *Computational Statistics and Data Analysis*, **2**, 143–158.

Muller, K. E., L. LaVange, S. L. Ramey, and C. T. Ramey (1992). Power calculations for general linear multivariate models including repeated measures applications. *Journal of the American Statistical Association*, **87**, 1209–1226.

Müller, P., G. Parmigiani, C. Robert, and J. Rousseau (2004). Optimal sample size for multiple testing: The case of gene expression microarrays. *Journal of the American Statistical Association*, **99**(468), 990–1001.

Müller, P., C. Robert, and J. Rousseau (2006). Sample size choice for microarray experiments. In *Bayesian Inference for Gene Expression and Proteomics* (K.-A. Do, P. Müller, and M. Vannucci, eds.) Cambridge, UK: Cambridge University Press.

Musonda, P., C. P. Farrington, and H. J. Whitaker (2006). Sample sizes for self-controlled case series studies. *Statistics in Medicine*, **25**, 2618–2631.

Muthen, L. K. and B. Muthen (2002). How to use a Monte Carlo study to decide on sample size and power. *Structural Equation Modeling*, **4**, 599–620.

Nam, J.-M. (1992). Sample size determination for case–control studies and the comparison of stratified and unstratified analyses. *Biometrics*, **48**, 389–395.

Noble, R. B., J. A. Bailer, S. R. Kunkel, and J. K. Straker (2006). Sample size requirements for studying small populations in gerontology research. *Health Services & Outcomes Research Methodology*, **6**, 59–67.

O'Brien, R. G. and G. Shieh (1999). Pragmatic, unifying algorithm gives power probabilities for common F tests of the multivariate general linear hypothesis. Unpublished manuscript. (Available at `www.bio.ri.ccf.org/UnifyPow.all/PowerMVGLH990908.pdf`.)

Obuchowski, N. A. (1998). Sample size calculations in studies of test accuracy. *Statistical Methods in Medical Research*, **7**, 371–392.

Obuchowski, N. A. and D. K. McClish (1997). Sample size determination for diagnostic studies involving binormal ROC curve indices. *Statistics in Medicine*, **16**(13), 1529–1542.

Olobatuyi, M. E. (2006). *A User's Guide to Path Analysis*. Lanham, MD: University Press of America.

Orr, M. and P. Liu (2009). Sample size estimation while controlling false discovery rate for microarray experiments using `ssize.fdr` package. *The R Journal*, **1**(1), 47–53.

Pennington, M. and J. H. Volstad (1991). Optimum size of sampling unit for estimating the density of marine populations. *Biometrics*, **47**, 717–723.

Pirinen, M., P. Donnelly, and C. C. A. Spencer (2012). Including known covariates can reduce power to detect genetic effects in case–control studies. *Nature Genetics*, **44**, 848–851.

Qiu, P., M. L. Moeschberger, G. E. Cooke, and P. J. Goldschmidt-Clemont (2000). Sample size to test for interaction between a specific exposure and a second risk factor in a pair-matched case–control study. *Statistics in Medicine*, **19**(7), 923–935.

Rempher, K. J. and S. Miller (2008). Making sense of statistical power. *American Nurse Today*, June.

Roy, A., D. K. Bhaumik, S. Aryal, and R. D. Gibbons (2007). Sample size determination for hierarchical longitudinal designs with differential attrition rates. *Biometrics*, **63**(3), 699–707.

Satten, G. A. and L. L. Kupper (1990). Sample size requirements for interval estimation of the odds ratio. *American Journal of Epidemiology*, **99**, 381–384.

Scheaffer, R. L., W. Mendenhall, and L. Ott (1979). *Elementary Survey Sampling*, 2nd edition. North Scituate, MA: Duxbury. (The current edition is the 6th edition, 2005.)

Schwertman, N. C. and D. K. Smith (1998). Optimal sample size for detection of an infestation. *Journal of Agricultural, Biological, and Environmental Statistics*, **3**(4), 359–369.

Self, S. and R. Mauritsen (1988). Power/sample size calculations for generalized linear models. *Biometrics*, **44**, 79–86.

Shao, Y. and C. Tseng (2007). Sample size calculation with dependence adjustment for FDR-control in microarray studies. *Statistics in Medicine*, **26**, 4219–4237.

Shieh, G. (2003). A comparative study of power and sample size calculations for multivariate general linear models. *Multivariate Behavioral Research*, **38**(3), 285–307.

Shoukri, M. M., M. H. Asyali, and A. Donner (2004). Sample size requirements for the design of reliability study: Review and new results. *Statistical Methods in Medical Research*, **13**(4), 251–271.

Smith, P. J. (1997). Power and sample size considerations for detecting deviations from secular trends in surveillance surveys. *The Statistician*, **46**(3), 423–432.

Snijders, T. A. B. (2005). Power and sample size in multilevel modeling. In *Encyclopedia of Statistics in Behavioral Science*, **3**, 1570–1573.

Snijders, T. A. B. and R. J. Bosker (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics*, **18**(3), 237–259.

Sparks, T. H., J. O. Mountford, S. J. Manchester, P. Rothery, and J. R. Treweek (1997). Sample size for estimating species lists in vegetation surveys. *The Statistician*, **46**(2), 253–260.

Taylor, J. M. G. (1986). Choosing the number of controls in a matched case–control study, some sample size, power and efficiency considerations. *Statistics in Medicine*, **5**(1), 29–36.

Tian, G.-L., M.-L. Tang, Z. Liu, M. Tan, and N.-S. Tang (2011). Sample size determination for the non-randomised triangular model for sensitive questions in a survey. *Statistical Methods in Medical Research*, **20**, 159–173.

Tibshirani, R (2006). A simple method for assessing sample sizes in microarray experiments. *BMC Bioinformatics*, **7**, 106.

Tokola, K., D. Larocque, J. Nevalainen, and H. Oja (2011). Power, sample size and sampling costs for clustered data. *Statistics and Probability Letters*, **81**, 852–860.

Tu, X. M., J. Kowalski, J. Zhang, K. G. Lynch, and P. Crits-Christoph (2004). Power analyses for longitudinal trials and other clustered designs. *Statistics in Medicine*, **23**(18), 2799–2815.

Ury, H. K. (1975). Efficiency of case–control studies with multiple controls per case: Continuous or dichotomous data. *Biometrics*, **31**, 643–649.

Wallis, W. A. (1980). The Statistical Research Group, 1942–45. *Journal of the American Statistical Association*, **75**(370), 321.

Walter, S. D., M. Eliasziw, and A. Donner (1998). Sample size and optimal designs for reliability studies. *Statistics in Medicine*, **17**(1), 101–110.

Wang, S. and H. Zhao (2003). Sample size needed to detect gene–gene interactions using association designs. *American Journal of Epidemiology*, **158**, 899–914.

Wei, C., J. Li, and R. E. Bumgartner (2004). Sample size for detecting differentially expressed genes in microarray experiments. *BMC Genomics*, **5**(1), 87.

Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications*, **9**(6), 476–487.

Wolynski, W. (2005). Minimal sample size in the group classification problem. *Journal of Classification*, **22**, 49–58.

Wu, J. C. and C. L. Wilson (2005). Using Chebyshev's inequality to determine sample size in biometric evaluation of fingerprint data. NISTIR 7273, National Institute of Technology, Gaithersburg, MD.

Xie, J., T. T. Cai, and H. Li (2011). Sample size and power analysis for sparse signal recovery in genome-wide association studies. *Biometrika*, **98**(2), 273–290.

Yamamura, K. and M. Ishimoto (2009). Optimal sample size for composite sampling with subsampling, when estimating the proportion of pecky rice grains in a field. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**(2), 135–153.

Zhang, H. and H. Stern (2009). Sample size calculation for finding unseen species. *Bayesian Analysis*, **4**, 763–792.

Zhang, L., A. Dmitrienko, and G. Luta (2008). Sample size calculations in thorough QT studies. *Journal of Biopharmaceutical Statistics*, **18**, 468–482.