The striking criterion whether variance calculation requires dividing the sum of squares by the number of summands or by that number less one

by

Martin Bachmaier

Technische Universität München
Agrricultural Systems Engineering
Am Staudengarten 2, D-85 354 Freising, Germany

Phone: +49 8161 71 3088

Email: bachmai@wzw.tum.de

Abstract The objective of this article is to address the frequently found academic opinion that in order to calculate the population variance, the sum of the squared deviations from the mean should be divided by the non-reduced number of summands, and not, as is the case for the sample variance, by one less. In reality, the question regarding the subtraction of one has nothing to do with whether it concerns a population or a sample; the decisive factor is whether the sum of squares refers to all measurement values or to characteristic values that are equally probable (or frequent). In the latter case, the full number is the divisor. However, if the sum of squares contains all measurement values of a sample or a population, one must be deducted from the overall number before it can be considered suitable as a divisor. Nevertheless, and precisely because of this subtraction, the variance becomes an average spread in the sense that it is not subject to a trend with regard to the number of data points, so that variances of populations of different sizes can be compared to each other.

Keywords: variance definition, population variance, sample variance, unbiased estimate, sampling techniques, variogram.

1 Introduction

Conventional wisdom dictates that in order to calculate the population variance, the sum of squared deviations from the mean should not be divided by $N-1$, but rather by the number of all individual items, N. At a first glance, a divisor N may appear uncontroversial for population variance, since it is believed that this method achieves an average measure. But this is not true. As shall be shown in Section 3, an average measure of statistical dispersion only results from the subtraction of 1 for the divisor of the sum of squares; an average measure in the sense that it is not subject to a trend with regard to the data quantity.

However, if we also use the divisor $N-1$ for the population variance, are we not creating a conflict with the probability theory? The variance of a random variable or distribution is defined by $\mathrm{Var}(X) = \mathrm{E}([X - E(X)]^2)$, which in a discrete uniform distribution leads to a division of the sum of squares by the number of the different characteristic occurrences, and there is no hint of a subtraction of 1!

Likewise to Bachmaier (2010), this article provides the striking criterion when the subtraction of 1 is necessary and when it is not. This criterion is independent of whether the variance relates to a population or a sample. It only depends on the meaning of the x_i in the variance formula, a simple thing, so simple that people are not aware of it. So, let us first take a look at the text book literature.

2 Disagreement in the literature

The variance of a random variable X is defined as $\mathrm{E}([X - \mathrm{E}(X)]^2)$, which results in dividing the square sum by the number of summands if X follows a discrete distribution with equal probability mass on each support point. There is always complete consensus in the literature when the variance refers to a random variable or its distribution. However, this consensus ends when it refers to a finite set of data, as in the example above.

With regard to the population variance, many authors avoid the question of the divisor of the deviation square sum from the mean by indicating only the variance of a random variable (or its distribution) as well as a sample variance. However, these authors already cannot agree with regard to the sample variance. In this vein, Bosch (1993), Cox and Cohen (1985), Kreyszig (1988), Snedecor and Cochran (1989), as well as Warren and Grant (2001) use divisor $n-1$, while Bortz (1993), Burkschat et al. (2004), Fahrmeier et al. (2007) and Fowler (2009) favor divisor n, the sample size. The latter apparently interpret the variance as an average square deviation from the mean, while the former invoke arguments such as unbiasedness and the associated degrees of freedom, whereby their unbiasedness refers to the variance σ^2 of a random variable or its distribution, but not of a finite population.

There is also disagreement among authors who do not differentiate between population and sample. For example, Crawshaw and Chambers (2001) calculate the variance of a "set of numbers" with its number as the divisor, while in Moore et al. (2009), the divisor of the variance of a "set of observations" is smaller by 1. Wonnacott and Wonnacott (1985) allow for both, but speak about the former as "mean squared deviation" and only call the latter "variance".

The two types of definitions for variance are more frequently found with finite populations, which was not explicitly considered by the authors that have been mentioned so far. Cochran (1977), Hansen et al. (1993) and Kish (1995) define its variance both via the divisor N, the number of elements of the population, as well as via the divisor $N-1$. All of these authors describe it as σ^2 in the first case, and S^2 in the latter. A preference for one of the two definitions is not provided. Divisor $N-1$ seems to derive its justification from consistency with the sample variance, where these authors favored using divisor $n-1$, or, as expressed by Cochran (1977), to "approach sampling theory by means of the analysis of variance." On the other hand however, the argument of an unbiased estimate, on the basis of which the divisor of the sample variance has been reduced by 1, has fallen by the wayside, since no estimate is required in the case of a population, so why reduce the divisor N?

In contrast, Devore and Peck (1994), Lohr (1999) and Thompson (2002) follow a consistent line. Similar to Moore et al. (2009), the divisor of the square sum is 1 less than the data number, regardless of whether that number originates from a sample or a population. The latter criterion only decides whether the divisor is described as $N-1$ or $n-1$. No reason is indicated as to why 1 is subtracted for the population variance. Divisor $N-1$ is also used by Sampath (2005) who however does not mention the sample variance.

Those authors who use different definitions for population and sample variance without compromise garner the most attention. These include Anderson et al. (2002), Clarke et al. (2005), Levy and Lemeshow (1999), Sachs (1992), Scheaffer et al. (2006) as well as Som (1996). They use N as the divisor as soon as it concerns a population, and $n-1$ if it refers to a sample. For Scheaffer et al. (2006), it is only possible to detect the use of N on the basis of an example on page 53. Sachs (1992) effectively abandons the subtraction of 1 in the divisor if the average value is known. The use of different divisors is surprising with regard to Levy and Lemeshow (1999) and Som (1996) to the extent that they mention that $[(N-1)/N]s^2$ provides an unbiased estimate of their σ^2 defined with the divisor N, or s^2 estimates $[N/(N-1)]$-times of their σ^2 without bias. At the same time, Som (1996) also mentions that some textbooks also use the divisor $N-1$ for population variances. Clarke et al. (2005) on the other hand justify divisor $n-1$ for a sample variance with unbiasedness for the variance of a random variable, but in doing so they do not provide a reason for why they use divisor N to calculate the variance of a finite population. On the other hand, Anderson et al. (2002) contend that divisor $n-1$ of a sample variance leads to an unbiased estimate for the population variance as defined with divisor N. They do not mention that this only applies to the rather unusual case of sampling with replacement. Hedayat and Sinha (1991) also use a divisor N for the population variance, without providing any reasons; the estimation of this variance using a sample is not part of the topic of their book.

The less than unanimous variance definitions found in this overview of literature are unlikely to assist our student, especially since his desire for a fair variance comparison with different data numbers has not even been touched.

Moreover, this overview of literature leaves one with the impression as if there were no definition per se; each definition seems to have advantages and disadvantages. However, all the discrepancies mentioned above are easily solved. There is always a way of appropriately defining variance. To demonstrate this, we begin with an alternative definition of variance that is based on pairwise differences and just on this basis already offers better insights into the nature of the variance.

3 Variance definitions based on pairwise differences

As already mentioned in Bachmaier (2010), the idea of a measure of statistical dispersion is better defined by way of pairwise differences, since this does not require an estimate of a nuisance parameter, such as the mean. However, this clarifying definition, which also forms the foundation of an empirical variogram (Bachmaier and Backes, 2008, 2011), seems to be little known; with regard to all the textbooks referenced in the present article, it is only mentioned by Hedayat and Sinha (1991).

The following applies to the usual sample variance s^2 and the maximum likelihood estimate s_*^2 (Hedayat and Sinha, 1991, *with proof*):

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{1}$$

$$= \frac{1}{2}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(x_i - x_j)^2 = \frac{1}{2}\frac{1}{\binom{n}{2}}\sum_{i=1}^{n}\sum_{j=i+1}^{n}(x_i - x_j)^2 \tag{2}$$

$$s_*^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{3}$$

$$= \frac{1}{2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2 \tag{4}$$

If one refers to characteristic values $x_1, x_2, \ldots, x_N$ of a population with N individual items, the definition with a divisor N is often considered as the actual one; however, for the sake of consistency with the above definitions we will refer to it as σ_*^2, while the actual variance name σ^2 receives the definition with the divisor $N-1$ (Hedayat and Sinha, 1991, *with proof*):

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{5}$$

$$= \frac{1}{2}\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}(x_i - x_j)^2 = \frac{1}{2}\frac{1}{\binom{N}{2}}\sum_{i=1}^{N}\sum_{j=i+1}^{N}(x_i - x_j)^2 \tag{6}$$

$$\sigma_*^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{7}$$

$$= \frac{1}{2}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(x_i - x_j)^2 \tag{8}$$

	Goalie	Defencemen		Forwards		
Squared age differences	Hähnel * 1992	Moburg * 1977	Pufal * 1981	Michl * 1989	Welz * 1976	Möhle * 1989
Hähnel * 1992	—	225	121	9	256	9
Moburg * 1977	225	—	16	144	1	144
Pufal * 1981	121	16	—	64	25	64
Michl * 1989	9	144	64	—	169	0
Welz * 1976	256	1	25	169	—	169
Möhle * 1989	9	144	64	0	169	—

Table 1: Squared pairwise age differences of six Landshut Cannibals players

Example 1: The difference between the divisors N and $N-1$ will now be demonstrated by means of the average age of six ice hockey players at Landshut Cannibals. For simplicity's sake, we will use the year of birth of the players as a basis, so that their ages will only be accurate to the year.

As the team currently playing on the ice, the selection of $N = 6$ players forms a population. On the other hand, it constitutes a sample of size $n = 6$, for instance to make inferences about the variance in ages of the entire team. These two perspectives suffice to make it inappropriate to calculate the variance for the population and the sample in different ways.

Table 1 illustrates the difference between the two variance calculations. While for s_*^2 and σ_*^2 in (4) and (8) the age differences between self-identical players, which are zero by default, are considered just as all other differences — whereby the dash "—" in the table is to be interpreted as zero for this purpose — these zero differences are not considered for s^2 and σ^2 in (2) and (6). The dash "—" in this case means "does not apply". At the same time, zero differences that result from players of the same age flow into the variance calculation. While the fact that Möhle is the same age as himself does not offer any information with regard to the spread of the age, the same age of forwards Möhle and Michl points to the age homogeneity of the team, which should bring the variance closer to zero.

The table shows that the share of self-identical pairwise comparisons on the main diagonal increases as fewer players are compared. The larger the share of these zero differences that are devoid of information, the more downward pressure it exerts on the variance σ_*^2. If only one player is examined, e.g., the goal keeper, this share even increases to 100 % and the variance σ_*^2 of the age is 0, even though no information regarding a dispersion is available with one player. In contrast to this, variance σ^2 for $N = 1$ remains undefined. If we now examine the two defenders playing on the ice as a population, the mentioned share still makes up half, while for the three forwards only a third of comparisons result in self-identical differences. When we expand our view to all of the world's ice hockey teams, we would conclude that with regard to age, a forward line varies more than a defensive line, as long as the 'varying' is measured with the σ_*^2 variance. Debates could start as to why trainers use more homogeneous players as regards age in defensive rather than offensive lines. However, if one measures the heterogeneity of age with σ^2, these discrepancies do not occur.

We can see that the question regarding a fair variance comparison for different data numbers clearly speaks in favor of a divisor $N - 1$. A variance calculation using divisor N instead of $N - 1$ is comparable with the range for the measurement of the dispersion. Qualitatively, the resulting misrepresentation would be the same in both cases.

4 Which population variance is estimated without bias by the sample variance?

Is the population variance σ_*^2 in (7) at least justified by the fact that it is estimated without bias by the usual sample variance s^2 in (1)? Can a sample of $n = 5$ result in an unbiased estimate for the variance of a population with $N = 6$ by dividing the sum of squares of the sample variance by $n - 1 = 4$, and that of the population variance by $N = 6$? Of course not, unless the sample is taken with replacement. The variance s^2 of a sample without replacement does not estimate σ_*^2, with the divisor N, in an unbiased manner; rather $[N/(N-1)]\sigma_*^2$, which is σ^2, where the divisor is also 1 less than the data quantity (Cochran, 1977; Kish, 1995; Levy and Lemeshow, 1999; Som, 1996). The variance definitions based on pairwise difference creation show how trivial this is:

Proof that $E(s^2) = \sigma^2$: Vector $(x_1, x_2, \ldots, x_N)$ with length N describes the data of the population. To calculate the expected value of s^2 (which is here a random variable), it is helpful to illustrate the random vector $(X_1, X_2, \ldots, X_n)$ with length $n \leq N$, which represents the sample without replacement, via a random vector $(\pi_1, \pi_2, \ldots, \pi_n)$ which indicates the indices of selected individuals. Each index vector would therefore appear with the same probability. It is $1/[N(N-1)(N-2)\cdots(N-n+1)]$. The following then applies to the random variable, which indicates the i^{th} sample value:

$$X_i = x_{\pi_i} \tag{9}$$

whereby all X_i are identically distributed, but are not independent. The latter is not required for the linearity of the expected value. Hence it can be quickly shown:

$$\begin{aligned}
\mathrm{E}(s^2) &= \mathrm{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] && (10)\\
&= \mathrm{E}\left[\frac{1}{2}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(X_i - X_j)^2\right] && (11)\\
&= \frac{1}{2}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathrm{E}\left[(X_i - X_j)^2\right] && (12)\\
&= \frac{1}{2}\frac{1}{n(n-1)}n(n-1)\mathrm{E}\left[(X_1 - X_2)^2\right] && (13)\\
&= \frac{1}{2}\mathrm{E}\left[(x_{\pi_1} - x_{\pi_2})^2\right] && (14)\\
&= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\left[(x_i - x_j)^2 P((\pi_1, \pi_2) = (i,j))\right] && (15)\\
&= \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\left[(x_i - x_j)^2\frac{1}{N(N-1)}\right] && (16)\\
&= \frac{1}{2}\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\left[(x_i - x_j)^2\right] && (17)\\
&= \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 && (18)\\
&= \sigma^2 && (19)
\end{aligned}$$

Hence, for a normal case of sampling without replacement, the argument of unbiasedness also clearly speaks for divisor $N-1$.

5 The variance of a discrete distribution with equal probability mass on each support point

Are we not creating a conflict with the probability theory when using the divisor $N-1$ even for the population variance? The variance of a random variable X or its distribution is defined by $\mathrm{Var}(X) = \mathrm{E}([X - E(X)]^2)$, which in a discrete distribution with equal probability mass on each support point[1] leads to a division of the sum of squares by the number of support points, and there is no hint of a subtraction of 1!

However, a probability distribution does not relate to frequencies of a characteristic of a finite population, but to a set of characteristic values whose occurrences are described by a probability measure, and to approach a probability by a frequency, the population size must tend to infinity. In particular, a discrete distribution with equal probability mass on each support point x_i, $i = 1, \dots, M$, is not comparable to the characteristic values of a finite population, $x_1, \dots, x_M$, but to those of an infinitely large imaginary population, $x_1, \dots, x_M,\ x_1, \dots, x_M,\ x_1, \dots, x_M,\ \dots$, where each x_i occurs an infinite number of times and with the same frequency. The variance of this imaginary population of size $N = \infty$ can be computed as the following limit of the population variance formula in (5), which is based on divisor $N-1$:

$$\mathrm{Var}(\text{imag.pop.}) = \lim_{k\to\infty} \frac{1}{\underbrace{kM}_{N}-1} \sum_{i=1}^{M} k(x_i - \bar{x})^2 = \frac{1}{M} \sum_{i=1}^{M} (x_i - \bar{x})^2 = \mathrm{Var}(X). \tag{20}$$

This shows that the population variance formula based on divisor $N-1$ is compatible with the variance of a random variable. As far as the theory of probability is concerned, it does not play a role whether divisor N or $N-1$ is used.

[1] To avoid misunderstandings, such a distribution should not be called a discrete uniform distribution, unless the support points x_i are equidistant.

6 The significance of x_i — one-time measurements or recurring occurrences of characteristics

The preceding section also shows that the discrepancies with regard to different variance definitions are based on the disregard of the exact significance of x_i, and yet no one seems to be aware of this. Both one-time measurements and recurring occurrences of characteristics are denoted by x_i, and different denotations for their number, such as N and M in the preceding section, are usually not used. It seems as if many authors imagined a finite population as a discrete distribution with equal probability mass on each support point, and hence, they adopt the variance formula of this distribution when calculating the variance of a finite population by using divisor N instead of $N-1$.

The next example also concerns the same difference on the basis of a sample instead of a population. In this case, nobody would make this mistake of disregarding the number of occurrences.

Example 2: We are examining the coin toss. Heads win one Euro, while tails loses one Euro. Let us assume someone tossed a coin 2000 times and has obtained approximately 1000 tails and 1000 heads. The sample only recognizes $m = 2$ different wins, $x_1 = +1$ and $x_2 = -1$. Both appear frequently and at approximately the same frequency, so that the sample variance of the win can be approximated with $s^2 \approx \frac{1}{m}\sum_{i=1}^{m}(x_i - \bar{x})^2 = \frac{1}{2}[(+1-0)^2+(-1-0)^2] = 1$. Calculating with $1/(m-1)$ instead of $1/m$ would have yielded 2 as a result, which is completely wrong. To accurately obtain the sample variance, one cannot bypass the formula in (1), which includes all $n = 2000$ measurement values x_i in the sum and hence has the divisor $n - 1 = 1999$.

7 Sampling with replacement

When sampling with replacement, the ordinary sample variance, s^2 in (1), unbiasedly estimates σ_*^2 in (7) with divisor N, and at a first glance, this seems to favor defining the population variance with divisor N. However, drawing and replacing involves pretending that each of the N individuals in the population is present an infinite number of times, and hence, we have arrived at the imaginary population or discrete distribution in Section 5 with the only exception that now identical x_i in the original population and thus discrete distributions with unequal probability masses, $P(X = x_i)$, are also possible. For example, drawing with replacement from a finite population with the measurements 2, 3, 7, 7, 9 ($N = 5$) corresponds to drawing from a discrete distribution with support $\{2, 3, 7, 9\}$ and probability masses 0.2, 0.2, 0.4, and 0.2. Thus, what one unbiasedly estimates when applying the ordinary sample variance to a sample based on drawing and replacing is not the population variance, but the variance of a discrete distribution whose probabilities, $P(X = x_i)$, equal the frequencies of the corresponding x_i in the original, finite population. The latter variance can be computed according to the last equation in (20), where the x_i are the measurements of the original population, which can be equal each other, and the divisor M equals the size N of this finite population, whose variance should actually be estimated.

This population variance continues to be represented by σ^2 in (5). Its divisor is $N-1$, and hence, the factor by which it is greater than the variance of the aforementioned discrete distribution is $N/(N-1)$. Therefore, to obtain an unbiased estimate of the population variance with data from sampling with replacement, the sample variance must be corrected by this factor:

$$s^2_{\text{with repl.}} = \frac{N}{N-1}\,\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{21}$$

8 Summary of results and conclusions

In the case of a finite number of data, only s^2 and σ^2 in (1) and (5), where the sum of deviances from the mean is divided by $n-1$ or $N-1$, are appropriate for a definition of a variance. There are three reasons for this:

- They are the true average squared dispersion measures, as their averaging of pairwise squared differences in the alternative formulae in (2) and (6) is not distorted by a nuisance parameter.
- They enable a fair comparison of samples or populations of different size.
- They match with respect to unbiasedness in the usual case of sampling without replacement: Neither is $\mathrm{E}(s_*^2) = \sigma_*^2$ nor $\mathrm{E}(s^2) = \sigma_*^2$ but $\mathrm{E}(s^2) = \sigma^2$.

Moreover, it is a lot easier to teach students to use the same variance formula for sample and population than to use different formulae, especially as it is not always easy to differ between sample and population. The variance estimate in (21) is an exception, since drawing and replacing involves different conditions for sampling and finite population.

The question of whether the sum of squares is to be divided by the number of summands or by one less,

- does not depend on the existence of either a population or sample,
- but rather depends on whether the sum of squares refers only to the different occurrences of a characteristic or to all measurement values.

If it refers to the latter, the divisor is one less than the number of summands. However, where a characteristic is discrete and its M different values occur at the same frequency and in great number, and more accurately even in infinite number, which can be described by a discrete distribution with equal probability for each occurrence, the variance calculation can be shortened by exclusive reference to the different occurrences pursuant to the limit result in (20). In that case however the divisor is not $M-1$ but rather M.

References

Anderson, D.R., Sweeney, D.J., & Williams, T.A. (2002) Statistics for Business and Economics. 8^{th} *ed., South Western, a division of Thomson Learning, Cincinnati.*

Bachmaier, M. (2010) Why the square sum must be divided by $N-1$ even for a population variance. *Journal of Statistics: Advances in Theory and Application,* **4**, 25–39.

Bachmaier, M., Backes, M. (2008) Variogram or Semivariogram? — Understanding the Variances in a Variogram. *Precision Agriculture,* **9**, 173–175.

Bachmaier, M., Backes, M. (2011) Variogram or Semivariogram? Variance or Semivariance? Allan Variance or introducing a new term? *Mathematical Geosciences, doi: 10.1007/s11004-011-9348-3.* Early view: http://www.springerlink.com/content/y37987lk358rn716/

Bortz, J. (2005) Statistik für Human- und Sozialwissenschaftler. 6^{th} *ed., Springer Medizin Verlag, Heidelberg.*

Bosch, K. (1993) Statistik-Taschenbuch. *R. Oldenbourg Verlag, München, Wien.*

Burkschat, M., Cramper, E., Kamps, U. (2004) Beschreibende Statistik. Grundlegende Methoden. *Springer-Verlag, Berlin, Heidelberg.*

Cochran, W.G. (1977) Sampling Techniques. 3rd *ed.,* 14th *print, John Wiley & Sons, Inc., New York.*

Clarke, G.M., & Cooke, D. (2005) A Basic Course in Statistics. 5th *ed., Hodder Arnold, Oxford University Press, London, New York.*

Cox, B.G., & Cohen, S.B. (1985) Methological Issues for Health Care Surveys. *Marcel Dekker, Inc., New York, Basel.*

Crawshaw, J., & Chambers, J. (2001) A Concise Course in Advanced Level Statistics. 4th *ed., Nelson Thornes Limited, Cheltenham.*

Devore, J., & Peck, R. (1994) Introductory Statistics. 2nd *ed., West Publishing Company, Minneapolis, New York.*

Fahrmeir, Künstler, Pigeot, Tutz (2007) Statistik. Der Weg zur Datenanalyse. *Springer Verlag, Berlin, Heidelberg.*

Fowler, F.J. (2009) Survey Research Methods. 4th *ed., Sage, Los Angeles.*

Hansen, M.H., Hurwitz, W.N., & Madow, W.G. (1993) Sample Survey Methods and Theory. Volume I, Methods and Applications. *John Wiley & Sons, Inc., New York.*

Hedayat, A.S., & Sinha, B.K. (1991) Design and Inference in Finite Population Sampling. *John Wiley & Sons, Inc., New York.*

Kish, L. (1995) Survey Sampling. *John Wiley & Sons, Inc., New York.*

Kreyszig, E. (1988) Statistische Methoden und ihre Anwendungen. *Vandenhoeck & Ruprecht in Göttingen.*

Levy, P.S., & Lemeshow, S. (1999) Sampling of Populations. Methods and Applications. *3*rd *ed., John Wiley & Sons, Inc., New York.*

Lohr, S.L. (1999) Sampling: Design and Analysis. *Duxbury Press, Pacific Grove, California.*

Moore, D.S., McCabe, G.P., & Craig, B.A. (2009) Introduction to the Practice of Statistics. 6th *ed., W.H. Freeman and Company, New York.*

Sachs, L. (1989) Angewandte Statistik. *1992* 7th ed., Springer-Verlag Berlin, Heidelberg, New York.

Sampath, S. (2005) Sampling Theory and Methods. 2nd *ed., Alpha Science International Ltd., Harrow, UK.*

Scheaffer, R.L., Mendenhall, W., & Ott, R.L. (2006) Elementary Survey Sampling. 6th *ed., Duxbury, Thomson Books/Cole, Belmont, USA.*

Snedecor, G.W., & Cochran, W.G. (1989) Statistical Methods. 8th *ed., Iowa State University Press / Ames.*

Som, R.K. (1996) Practical Sampling Techniques. *2nd ed., Marcel Dekker, Inc., New York.*

Thompson, S.K. (2002) Sampling. *John Wiley & Sons, Inc., New York.*

Warren, J.E., & Grant, G.R. (2001) Statistical Methods in Bioinformatics. *Springer Verlag New York, Berlin, Heidelberg.*

Wonnacott, R.J., & Wonnacott, T.H. (1985) Introductionary Statistics. 4th *ed., John Wiley & Sons, Inc., New York.*