SAMPLING METHODS

GLOSSARY OF TERMS

The following terms include statistical terms that are encountered in sampling as well as specific terms included in this chapter.

1. Absolute Value - disregards all negative values of numbers; considers all numbers positive.

2. Adequate Sample - pertains most commonly to the size of a sample. A sample is adequate if its size is large enough to give the degree of precision or reliability required in a given sample estimate.

3. Alpha - the allowable probability associated with observed differences attributed to chance. If the probability associated with sample differences is less than alpha, we can reasonably conclude that a real difference between samples exists.

4. Bias - systematic error, leading to distortion in one direction of a statistical result; distinct from random error, where distortion in both directions may be largely self-canceling.

5. Caseload - the "target" population, comprised of only those cases included in the QC system for Medicaid.

6. Confidence Interval - the interval between two sample values, known as confidence limits, within which it may be asserted with a specified degree of confidence that the true population value lies.

7. Confidence Limits - the values which form the upper and lower limits of the confidence interval.

8. Equal Probability of Selection - selection of a sample where every case has an independent and equal chance of inclusion in the sample (also called self-weighted sample).

9. Frame - the list of cases from which the sample is actually selected; also known as the sample selection list.

10. Listed in Error - cases included in the sample selection list that are not included in the population of interest.

11. Mean - a measure of the central tendency of data; the sum of the values divided by the number of values.

12. Nonsampling Error - the error or deviation from the true population value in sample estimates which cannot be attributed to chance sampling variations. Examples are errors resulting from imperfections in the selection of sample units, bias in the estimating procedures used, mistakes in arithmetical calculations, inconsistent review procedures, etc.

Rev. 41 7-2-35

SAMPLING METHODS

13. Normal Distribution - a symmetrical, bell-shaped curve which describes the sampling distribution of many common sample statistics. While the sampling distributions of proportions and "percents in error" as used in QC are more correctly described by the binomial distribution, they are often closely approximated by the normal distribution, and it is common practice to use the normal distribution for this purpose. The normal distribution provides the theoretical basis for the determination of confidence limits and QC limits for the specification of particular levels or degrees of confidence involved in making sample estimates and in evaluating sampling error.

14. Oversampling - selecting more sample cases than required to compensate for cases that will have to be dropped.

15. Parameter - a value, property, or characteristic of a population which is estimated from a sample. Examples are a mean, proportion or percentage, total, range, or standard deviation.

16. Population of Interest - those units about which we wish to form conclusions from which a sample is selected and estimates made.

17. Precision - (See definition for reliability.) The degree to which a sample estimate approximates the true values; the sampling error of a sample estimate.

18. Probability - relative frequency of occurrence; the probability of an event is the relative frequency of occurrence of the event in an indefinitely large number of series of observations.

19. Probability Sampling - any method of sample selection which is based on the theory of probability. Probability sampling, which requires that at any stage of selection the probability of any unit or set of units being selected must be known, is the only general method of sampling which makes it possible to obtain a mathematical measure of the precision of the sample estimate.

20. Random Numbers - series of digits, each occurring independently of each other. Each digit tends to appear as many times as any other, in any progression, if the series selected is large.

21. Random Sampling - the process of selecting a sample from a population so that every unit in the population has a known chance of being included in the sample.

22. Random Start - In selecting a systematic random sample at intervals of some specified number of items in an ordered frame, it is mandatory to select the first item completely without bias. Such selection is then said to have given the sample "a random start."

23. Range - the largest minus the smallest of a group of values.

7-2-36 Rev. 41

SAMPLING METHODS

24. Reliability - (synonymous with precision) - the uniformity of sample results when obtained from repeated samples of the same size and type from the same population; the degree to which a sample estimate approximates the true value.

25. Risk - as used here, refers to the degree of risk associated with given degrees of confidence. For example, if a statement is made "with 95-percent confidence" that the true population parameter lies within a specified interval, there is a "5-percent risk" that the parameter actually lies outside that interval (also called alpha).

26. Sample - part of a universe; a limited or finite number of items selected from a universe, by a prescribed procedure, with the objective of estimating certain values (mean, total proportion, etc.) of the parent universe or of testing in respect to particular properties of the universe.

27. Sample Selection List - the list of cases from which the sample is actually selected; also known as the sample frame.

28. Sample Size - the number of items in the sample.

29. Sampling Distribution - the distribution of a (sample) statistic, such as a sample mean or a sample proportion or percentage, that would be formed by obtaining such statistics from all possible samples of a given fixed size selected by some specified sampling procedures; a population of all possible sample values of the statistic under consideration.

30. Sampling Error - that part of the difference between a universe value and an estimate of that value obtained from a random sample which is due solely to the fact that only a sample of values is observed; to be distinguished from non-sampling error which is due to biased or imperfect sample selection or real difference due to changes over time, error of observation, recording calculation, etc.

31. Sample Interval - in systematic sampling the number of cases between selections on the sampling frame.

32. Significant Difference - A difference is statistically significant if it can be concluded from a sample, with a given degree of risk, that the difference actually exists in the universe. A difference observed in a sample is judged not statistically significant if it could easily have occurred purely as a result of random sampling variations.

33. Simple Random Sample - a probability sample selected in such a way that each unit of the frame has an equal and independent chance of being included in the sample; for samples of any given size all possible combinations of units that could form samples of that size must have the same probability of selection (usually uses random digits for item selection).

Rev. 41 7-2-37

SAMPLING METHODS

34. Standard Deviation - the most widely used measure of the dispersion (scatter or variability) of frequency distributions from their arithmetic means. The standard deviation of the sampling distribution of any given statistic is also known as the "standard error" of that statistic.

35. Standard Error - the standard deviation of the sampling distribution of a given statistic; used in measuring precision of an estimate.

36. Stratified Random Sampling - random sampling of a universe which has been divided into a number of subuniverses according to some predetermined criterion (geographic location, characteristic, etc.). The percentage size of each sample must be equal or have individual weighing factors taken into account before the subuniverse sample results can be combined.

37. Stratum - a segment of the universe for which separate estimates are computed for some special reason. All strata must be combined if an estimate of the total universe is to be made.

38. Systematic Random Sample - a sample attained by selecting from a file, list, or computer tape individual items at equally spaced intervals (as every 10th, 140th, 850th, etc., item, as required to obtain a total sample of a given size), with the starting point within the first such interval being determined by random selection.

39. Tolerance - the proportion of sampling error which has been determined to be acceptable.

40. Universe (also called a population) - all units about which information is desired; a probability sample of these units yields an estimate of universe values within certain limits of reliability.

41. Weighted Sample - a sample in which the probability of selection is not equal, thereby requiring weighing by various factors so that no segment of the caseload is overrepresented.

7-2-38 Rev. 41

03-93 SAMPLING METHODS APPENDIX A

**MEQC ERROR RATE CALCULATION**

The MEQC fiscal year error rate and lower limit are based on the following:

RFYLL = RFY - 1.96 VAR(RFY)

D1 R1 + D2 R2

RFY = ─────────────

D1 + D2

D12 VAR(R1) + D22 VAR(R2)

VAR(RFY) = ────────────────────────

( D1 + D2 )2

Where Dp is the AFDC + MAO universe dollars for review period, P and Rp is the error rate for review period P. For each review period, the error rates and variances are computed as follows:

L \_ L \_

RP = (  wh xh ) **/** (  wh uh )

L

VAR(RP) =  wh2 VAR(Rh)

nh nh nh

 (xhi2) - 2 RP  (xhi uhi) + RP2  (uhi2)

VAR(Rh )

\_

uh2  nh (nh - 1)

Where ( For strata from h=1 to L, and assuming all sustained Federal findings have been substituted for State findings) :

xhi = The error amount for the ith case in the hth stratum

\_

xh = The full sample average error dollars per case in stratum h

uhi = The payment amount for the ith case in the hth stratum

\_

uh = The full sample average State dollars paid per case in stratum h

nh = The number of completed sample reviews in stratum h

wh = Nh / NP = The universe case weight for stratum h

Rev. 48 7-2-39