ARBITRON REPLICATION II

A Study of the Reliability of Radio Ratings

ARBITRON RATINGS

About the Authors

Michael G. Occhiogrosso is Vice President of Statistical Services for the Arbitron Ratings Company. He provides general research and technical statistical support for the company and has been the chief architect of a number of major research advances.

Mr. Occhiogrosso has twenty-three years of professional research and statistical experience. Before joining Arbitron, he was Vice President, Associate Director of Communications Research for the SSC&B Advertising Agency. Mr. Occhiogrosso has published many research documents under the auspices of the various companies with which he has been associated. He has appeared as a speaker before various industry groups, including the Association for Consumer Research, the New York Advertising Agency Media Research Directors and the New York Modellers Discussion Group.

Mr. Occhiogrosso earned his B.B.A. and M.B.A. degrees, with a major in statistics, from the City University of New York. He is a member of the American Statistical Association, the American Marketing Association, the American Association for Public Opinion Research and the American Psychological Association (Division of Consumer Psychology).

(Continued on back cover)

ARBITRON REPLICATION II

A Study of the Reliability of Radio Ratings

Michael G. Occhiogrosso

Martin R. Frankel

A Special Study by the
Statistical Services Department
Arbitron Ratings Company
New York, 1982.
A Control Data Company.

Published by
Fishergate Publishing Company, Inc.
2521 Riva Road, Annapolis, Maryland 21401

ISBN 0-942720-03-2
Library of Congress Catalog Card Number: 82-83801

Manufactured in the United States of America.

Contents

Section IV: Applications

Section V: Appendices

List of Tables and Figures

Section I: Study Overview

Arbitron Replication II: A Study Of The Reliability Of Radio Ratings

Introduction

Arbitron Radio audience data are used by advertisers, advertising agencies, station general managers and programming directors, and others to make critical decisions concerning the radio medium. This report is extremely relevant to all persons who use Arbitron Radio audience data because it provides the answers to such questions as:

- How reliable are Arbitron Radio audience ratings and projections?
- If Arbitron had used a different sample of persons, how much different might the reported audience have been?
- Can a simple method be devised to determine the margin of statistical sampling error in any given audience estimate?
- How confident can one be that a reported audience increase or decrease between two Arbitron surveys represents a real change and not just "statistical bounce"?

To answer these types of questions, Arbitron commissioned a special study that cost hundreds of thousands of dollars and took two years to complete. In addition to making an overall reliability evaluation, the study resulted in the development of a new, quick and easy method for determining the reliability of any audience rating or projection published in Arbitron Standard or Condensed Local Market Radio Reports. With the concurrence of the Electronic Media Rating Council (formerly The Broadcast Rating Council), the new "Reliability Procedure" was introduced in the Fall 1981 Reports.

Chapter 1 of this document presents a non-technical overview of the objectives, methods and findings of the *Arbitron Replication II* study. Examples illustrating how the results of the study may be applied as an aid to decision making are given in Chapter 8. The remaining sections are designed to give researchers an in-depth knowledge of the statistical concepts, study design and research methods used in this special study. These sections are necessarily more technical in nature.

We believe that the more accurate and easier to use reliability information provided by *Arbitron Replication II* represents a major contribution to users of Arbitron radio audience data. It should prove invaluable to all who use Arbitron radio audience ratings and projections in making programming and buying decisions.

Chapter 1

Study Highlights

A. Background

Arbitron Replication II: A Study of the Reliability of Radio Ratings is a special study that was initiated by Arbitron in the winter of 1979 and completed during the fall of 1981. Its completion represents, for the broadcast industry, a major breakthrough in measuring and calculating the reliability of Arbitron Radio Ratings.

This new study is another major research effort by Arbitron in evaluating the reliability of broadcast audience ratings through a sophisticated technique called "replication." A previous study[1] completed in 1974, focused primarily on television ratings, but it also included a pilot investigation of the reliability of radio ratings. *Replication II* represents the next logical step in this fundamental research — a full-scale study devoted entirely to examining the reliability of radio ratings.

The necessity of the full scale study was apparent from the pilot. The pilot indicated that more sophisticated methods than simple textbook formulae or even the Arbitron Nomograph Procedure must be applied to yield accurate estimates of the reliability of radio audience estimates. Against that background, the purpose of this study was: (1) to evaluate the reliability of Arbitron ratings, and (2) to develop a new, quick and easy method for reliability determinations.

B. Quality of the Study

Costing Arbitron hundreds of thousands of dollars and taking two years of research effort to complete, *Replication II* is one of the most complex and important undertakings of its kind. Its importance is reflected in the fact that Arbitron contracted a well-known expert in

[1] *Arbitron Replication: A Study of the Reliability of Broadcast Ratings.*

the field of survey sampling, Dr. Martin R. Frankel, as a statistical consultant to design the study together with Arbitron's Vice President of Statistical Services, Michael G. Occhiogrosso. Overall guidance for the project was provided by Marshall L. Snyder, Arbitron's Vice President and General Manager of Product and Research Services.

In *Replication II*, Arbitron applied the latest statistical advances in survey sampling methodology to empirically determine reliability levels for radio ratings. Using a technique called "Jackknife Replication," Arbitron analyzed the reliability levels for more than one million of its published audience ratings for radio. From these extensive analyses, a new mathematical model and a new method for determining the reliability of radio ratings were developed. Significantly, to Arbitron's knowledge, this new method is substantially more sophisticated than any other currently in general use by commercial research firms or governmental bureaus such as the U.S. Census.

C. Key Findings

The two stated objectives for the *Replication II* study were successfully achieved. Key findings are summarized as follows:

1. *Ratings Are Significantly More Reliable*—Across the more than one million published audience estimates examined in the *Replication II* study, it was found that the formerly used *Nomograph Procedure* overstated the size of the actual sampling error of Arbitron Radio Ratings by 24% on average. This finding means that Arbitron's radio audience estimates are much more reliable (that is, have smaller sampling error) and can be used with greater confidence than previously thought.

2. *Effective Sample Bases Are Larger*—Another important measure frequently used by broadcast researchers to gauge reliability is the quantity called *Effective Sample Base (ESB)*.[2] In terms of this index, the *Replication II* study found that

[2]For textbook *simple random samples*, reliability is, in great part, dependent upon the size of the sample used for the survey: the larger the sample, the greater the reliability of the survey estimates. The ESB value is an analogous measure frequently used by Broadcast Researchers to gauge the reliability of complex surveys. The ESB is an adjusted sample size value. It takes into account the influence on reliability of such factors as sample stratification and weighting, which impact complex survey designs such as Arbitron's. The larger the ESB, the greater the reliability of the survey estimates.

Arbitron's average Effective Sample Bases are 53% higher than previously reported. Again, this means that Arbitron radio report users can have more confidence in the reliability of published ratings than was previously believed.

3. *Reliability Varies Significantly By Several Factors And Must Be Determined On A Custom Basis For Each Local Market Radio Report*—The *Replication II* study found differences in reliability by *estimate types* (e.g., Cume versus Average Quarter Hour estimates), by *regions* (Metro, TSA and ADI), and by *demographic groups* (e.g., Males 18-34 versus Total Persons 12+). Furthermore, there exist influences on reliability, such as sample sizes and sample weighting adjustments, that are unique to a particular market and report. As a result, Arbitron concluded that reliability (sampling error) parameters must be custom-determined for each Local Market Radio Report and its regional, demographic and estimate type groupings.

4. *New Highly Sophisticated Radio Reliability Model Needed To Determine Reliability More Accurately*—To reflect the varying influences on reliability as measured in the *Replication II* study, Arbitron developed a new highly sophisticated mathematical model (formula) for determining the reliability of radio audience ratings and projection estimates. The new Radio Reliability Model represents a substantial benefit to broadcasters and advertisers because sampling error estimates can now be determined with *significantly greater accuracy* than was previously possible.

   The Electronic Media Rating Council has reviewed the new Radio Reliability Model and, with the Council's concurrence, the improved procedure was introduced beginning with the Fall, 1981 Standard and Condensed Local Market Radio Reports. The Electronic Media Rating Council is planning to complete future independent testing of the new Model and may suggest further refinements which may be incorporated in future reports.

5. *Reliability Can Now Be Determined With Greater Ease and Speed*—To determine reliability accurately requires a complicated mathematical model. However, despite the complexities of the new Radio Reliability Model, Arbitron has devised a unique method that relieves the report user of the need to make complex calculations. This method permits Arbitron to perform the burdensome calculations that must be custom-generated for

every individual survey and local market report so that the report user needs only to perform a simple two-step arithmetic procedure.

For the report user, the new procedure consists simply of selecting an appropriate value from each of two reference tables and dividing the value from Table A by the value from Table B to obtain the size of the "sampling error" for any published audience rating. The Table A and Table B values are now supplied by Arbitron (along with instructions on how to use them) in every Standard and Condensed Local Market Radio Report beginning with the Fall, 1981 survey period.

Appendix A contains an example of an actual Table A and Table B that appeared in the Fall 1981, New York Local Market Radio Report. Note that the values for Table B are individually, custom-derived for each market and report period; hence, the values in Table B will vary from report to report. The Table B shown in Appendix A is for illustrative purposes and the values shown in it *do not universally apply* to every Local Market Radio Report. Table A, however, is *universal;* the values in Table A will be the same across all Arbitron radio reports.

D. Benefits to Report Users

Accurate and easily available information about the reliability of an audience rating or projection is important when the rating or projection is used to help make a radio programming or buying decision. Such information can help minimize the chance of drawing inappropriate conclusions from the survey results because it serves as a basis for deciding how much allowance should be made for the possibility of random statistical error (or "statistical bounce" as some prefer to call it).

For example, when an increase in a station's rating is observed from one report to the next, it is important to know the likelihood of whether this increase is "real" or just due to "random sampling error."

There are a number of specific ways in which report users can apply the information provided by the new Radio Reliability Model. In Chapter 8, specific examples are given illustrating how reliability calculations may be used as an aid to decision-making in several situations.

It must be noted that Arbitron has not changed its basic methodology for producing radio audience estimates. The new reliability procedures affect only the user's ability to assess easily and accurately how reliable the audience estimates are.

E. Summary

The *Replication II* study successfully accomplished its two major objectives. The study evaluated the actual reliability levels of Arbitron's published radio audience estimates and determined that they are significantly more reliable than had been previously indicated by the approximation procedures formerly used for gauging reliability. This evaluation made it possible to quantify the combined effects of various key factors that influence reliability. As a result, Arbitron was able to develop a highly sophisticated new Radio Reliability Model which offers two benefits to report users:

1. Greater accuracy in determining reliability
2. Improved ease and speed in making reliability calculations.

The *Replication II* study is indeed, a success, and Arbitron would like to thank the Electronic Media Rating Council for its encouragement and guidance in this project. Believing strongly that users of radio ratings will be greatly aided by the new procedure, Arbitron urges them to become familiar with its value as a significant tool in rating analysis.

Section II: Study Details

Chapter 2

Background

A. Arbitron Annually Publishes Millions of Estimates of Radio Audience Sizes

Through the course of each calendar year, Arbitron publishes many hundreds of syndicated Local Market Radio Reports.[3] These reports contain various types of estimates of radio audience size. For example, some estimates are given in terms of "projections" (defined as: the number of listeners), some are given in the form of "ratings" (defined as: the number of listeners expressed as a percentage of the population), etc. The audience estimates are reported by individual stations for a large number of different Daypart, Age/Sex, and Geographic groupings. A single Arbitron Local Market Radio Report may contain as many as 200,000 or more different audience size estimates. Across all Local Market Radio Reports, Arbitron produces about forty million audience size estimates annually.

B. Radio Audience Estimates are Based on Samples

Important programming and buying decisions are made on the basis of Arbitron audience estimates. Hence, Arbitron strives to produce estimates that are as accurate as practical, and seeks continually to improve its methods for generating audience estimates. Arbitron's audience estimates are based upon surveys conducted among "samples" of persons. As a result, radio audience estimates (like all survey estimates) are subject to inherent error due to the fact that different samples will tend to generate somewhat different results. The margin of possible error due to this factor is commonly referred to as

[3]In addition to its syndicated Local Market Radio Reports, which are the focus of this document, Arbitron also provides the broadcast industry with many special types of radio reports, as well as reports that measure television audiences.

"sampling error" (or sometimes as "statistical bounce" in non-statistical language).

Another way of describing this margin of sampling error is to speak in terms of the "reliability" of a sample. Samples having *smaller* sampling errors may be thought of as having *higher* reliability and vice versa. In the ensuing text, depending upon the particular context, sometimes the term "reliability" is used and other times the term "sampling error" is used. These two terms are different ways of describing the same phenomenon.

C. Reliability Is Measured by a Statistical Quantity Called the "Standard Error"

Sampling error and reliability are conceptual terms. The actual size of the sampling error in numerical terms (e.g., 300 persons, or ¼ of a rating point) is measured by a statistical quantity called the "standard error." The standard error quantity can be used to ascertain upper and lower bounds about the published audience estimate to produce a "confidence interval" within which the audience size would be expected to fall *if the entire population frame used by Arbitron (instead of a sample)* were surveyed under the same conditions.

For example, suppose a station had a reported Metro Survey Area rating of 1.0 among Men 18+ for the average quarter hour during the weekday morning drive time 6AM–10AM. Based upon knowledge of the size of the standard error, a 90% confidence interval of 0.8 to 1.2 could be calculated as the lower and upper boundary values for the rating. This would inform the user that there is a 90% likelihood (9:1 odds) that the particular station's published rating could be as low as .8, or as high as 1.2, if the total population frame from which the sample were drawn had been surveyed by Arbitron under identical conditions.

Upper and lower bounds for other degrees of confidence (e.g., 68%, 80%, 95%, 99%) can also be calculated from knowledge of the standard error. The mathematics for completing these calculations is relatively simple and is described later in this report. The point here is that knowledge concerning the size of the standard error is important because it can then be taken into consideration when making radio programming and buying decisions, thereby minimizing the chances of drawing inappropriate conclusions from the survey results. When an increase in a station's rating is observed from one report to the next, it is important to know the likelihood of whether this increase is "real" or just due to "sampling". In Chapter 8 specific examples are given of how knowledge concerning the size of the standard error can be used as an important aid to decision-making.

D. Many Factors That Influence Reliability Can Not Be Measured By Simple Textbook Formulae

The size of the standard error, hence reliability, is influenced by many factors. There is no simple rule-of-thumb, or single number, that can accurately describe the size of the standard error. Theoretically, each one of the millions of audience estimates published each year by Arbitron is subject to its own individual (unique) standard error. If Arbitron's audience estimates were derived solely on the basis of a simple textbook random sample, then the size of the standard error for any particular audience rating would be influenced by just two factors—(i) *the size of the sample upon which the rating was based ("n")*, and (ii) *the numerical value of the rating itself ("R")*. The impact of these two factors on the size of the standard error for a textbook simple random sample is given by the following simple formula expressed in terms of a *one unit* (also called *one sigma*) standard error value:[4]

$$\begin{array}{r}\text{One sigma standard error value}\\ \text{for a simple random sample}\end{array} = \sqrt{\frac{\mathrm{R}(100\text{-}\mathrm{R})}{\mathrm{n}}}$$

Arbitron's surveys, however, like most "real world" surveys are considerably more complex, and other important additional factors influence the size of the standard error besides the sample size and rating level. As a result, simple textbook formulae do not accurately measure the impact of the additional factors that affect complex surveys such as Arbitron's. These additional factors can be grouped into four categories:

1. *Pre-Stratification Factors*—This category refers to the types of controls exercised in the initial sample selection and allocation

---

[4]From a statistical and numerical standpoint, audience ratings and projections are related (as are their respective standard errors) in an exact one-to-one fashion by a uniform constant (the population or universe size) within any given geographic and demographic grouping. Hence, while this document sometimes focuses on ratings (instead of projections) for simplification purposes, no loss of information or generality is intended.

process to help insure that certain demographic and geographic compositions are attained.

2. *Post-Stratification Factors*—This category refers to the weighting (sample balancing) applied at the tabulation stage of a survey to insure conformity of the sample configuration with known Universe or Census Population profiles.
3. *Sample Clustering Factors*—This category refers to the types of statistical dependence among sample respondents built into the design as when, for example, members of the same household are included in the survey tabulation, or multiple measurements are made on the same respondents.
4. *Random Data Errors*—This category refers to the types of random errors that might inadvertently be introduced during the various stages of the data collection and later tabulation processes.

Depending upon the circumstances and certain statistical relationships evidenced in the data, these factors will have a varying influence on reliability—that is, they may increase (or decrease) reliability.

E. Reliability Estimation Requires "Statistical Modelling"

The many factors that have an influence on the size of the standard error and its uniqueness for any particular audience rating makes the accurate estimation of these millions of standard errors a complicated problem. The resolution of this problem requires the development of "statistical models" that can be used as a vehicle for standard error estimation. Statistical models are mathematical formulae developed on the basis of empirical data analyses and theoretical statistical considerations.

In its past reports, Arbitron provided a method for standard error determination known as the *Nomograph Procedure*. This past method was not developed through extensive statistical modelling, but, for its time, the Nomograph Procedure provided a better estimate of the standard error, and represented a statistical advantage, over simple approximations. With the completion of the Replication II study, Arbitron now has available a new and improved statistical model for standard error estimation.

Chapter 3

Study Description

A. Study Objectives

It is important for the broadcast industry to have accurate and easily available information concerning the reliability of audience ratings. However, the results of Arbitron's earlier pilot study indicated that *simple textbook formulae* (which are often used by broadcast and advertising researchers) yield inadequate approximations of the true reliability levels of radio audience estimates. Furthermore, the *Nomograph Procedure*—which formerly appeared in Arbitron Reports—while generally superior to simple textbook formulae approximations, warranted improvement in two areas: statistical refinements and computational ease. Therefore, with the encouragement of the Electronic Media Rating Council, Arbitron set two basic objectives for its *Replication II* study:

> *Objective #1*—To develop new statistical procedures that would allow Arbitron report users to determine the reliability (sampling error[5]) of any rating (or projection) contained in Arbitron Local Market Radio Reports with:
>
> > Greater accuracy, and
> >
> > Greater ease and speed
>
> than any procedure currently available.
>
> *Objective #2*—To make an *overall evaluation* of the reliability of Arbitron audience ratings and projections, and to quantify certain key factors that have a significant influence on their reliability in order to enable the accomplishment of Objective #1.

---

[5]Arbitron's radio audience estimates are based upon surveys conducted among "samples" of persons. Different samples will tend to produce different results. These differences are commonly referred to as "sampling error." Reliability measures the size of the sampling error. The smaller the *sampling error*, the greater the *reliability*.

To meet these objectives, many advanced statistical analyses, in addition to the "Jackknife Replication" technique mentioned earlier, were applied.

It should also be noted that the two objectives are inter-related. An understanding of the factors that have a significant influence on reliability (Objective #2) is an essential component that helps make possible the achievement of Objective #1. In fact, one of the advantages of the "statistical modelling" approach cited previously is that it helps to identify the impact of the key factors that influence the size of the standard error.

B. Methodology

I. Conceptual Depiction

In simplified terms, the *Replication II* Study used a concept *akin* to the empirical intuitive approach described below:

- If the original study were replicated (repeated) several times under identical conditions using different samples of respondents, then the results of each replicate could be compared for consistency.
- If the different replicates showed very consistent results (as illustrated with eight replicates in Figure 1), then one would have a great deal of confidence in the audience estimates derived from the original study.
- By contrast, if the audience estimates from the replicates were very disparate (as illustrated in Figure 2), then this would indicate that the audience estimates from the original study may be subject to considerable statistical fluctuation.
- By gauging *empirically* how much the corresponding audience estimates vary across the several replicates, the reliability (standard errors) of the audience estimates from the original Local Market Radio Reports could be ascertained with the help of the appropriate mathematical formulae.

II. Summary of Methodological Details

The *Replication II* study used a more highly sophisticated statistical variation of the empirical intuitive approach outlined in the preceding

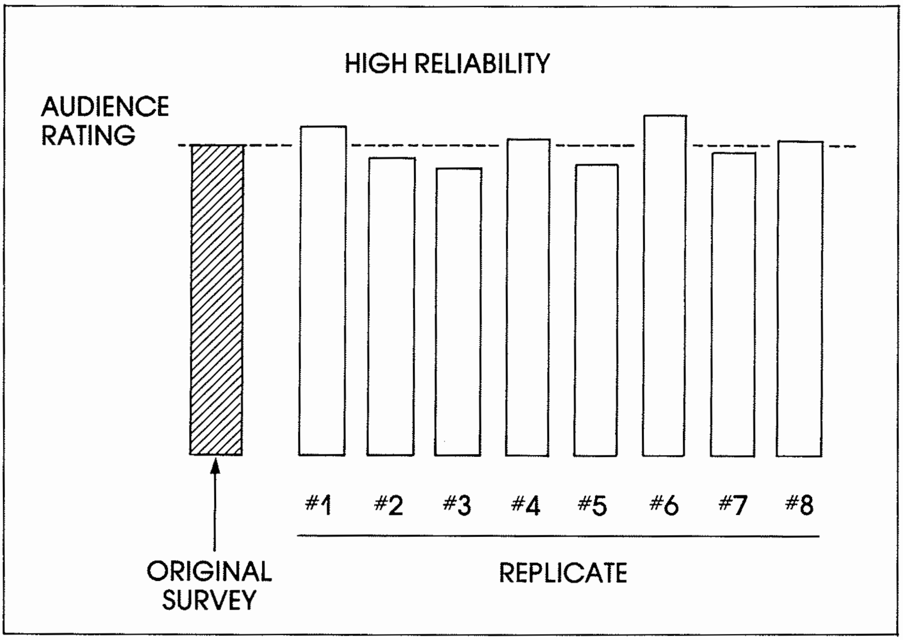


Figure 1. High reliability (graphic illustration).

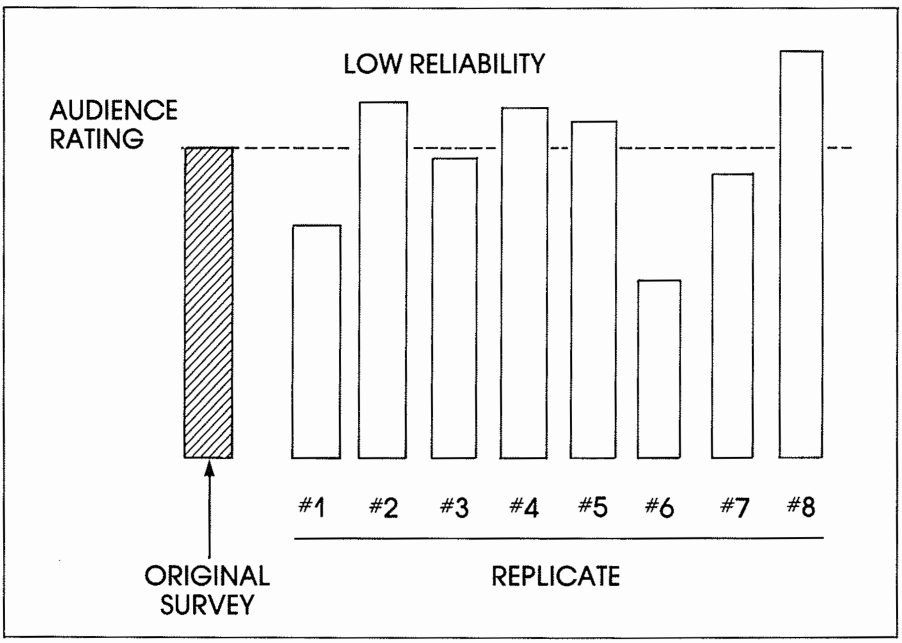


Figure 2. Low reliability (graphic illustration).

conceptual depiction. The "replicates" used for analyses were created by a special technique called *Jackknife Replication*. Statistically, the Jackknife Replication technique represents the advanced "state-of-the-art" in survey sampling methodology, and it is an improvement over the Simple Replication technique used in earlier reliability studies. The methodological details for the *Replication II* study are given in later chapters, but they can be summarized here into six steps as described below:

Step 1: Selection of Local Market Reports
A set of nineteen Arbitron Radio Local Market Reports were selected to reflect the presence of a full range of market and report conditions such as:

- Population size
- Market size from the standpoint of: (a) numbers of stations "home" to the Metro Survey Area, and (b) the total number of stations reported for the market
- Geographic region of the country
- Season of the year (Report Survey Period)
- Arbitron survey sample sizes (In-Tab) for the Metro Survey Area (Metro) and the Total Survey Area (TSA)
- Arbitron's use (or non-use) of special ethnic sampling and/or weighting adjustment procedures

Step 2: Selection of Eight Basic Subsamples For Each Report
Eight non-overlapping subsamples were separately drawn from the original total report sample for each of the nineteen reports. In effect, the initial total report sample was subdivided into eight parts (subsamples) of approximately equal size, and thus, each subsample is about one-eighth the size of the original report sample. In total, 152 subsamples were selected (8 per each of 19 separate reports). Their selection was accomplished via the use of appropriate mathematical randomization procedures. Great care was exercised during the selection process to duplicate all of the pre-stratification and sample-clustering design features that were used in the selection of the initial total report sample.

Step 3: Formation of Eight "Jackknife Replicates"
For each of the nineteen reports, the eight basic subsamples were combined in a special way to form the eight "replicates" that were used to perform an empirical analysis analagous to the one described on the preceding two pages. The special manner by which the sub-

samples were combined to form the eight "replicates" is called "Jackknife Replication." This technical aspect is explained in Chapters 6 and 7 of this report.

Step 4: Generation Of Audience Reports For Each Jackknife Replicate
The eight Jackknife Replicate samples for each of the nineteen selected markets surveys were then separately weighted and tabulated by the standard Arbitron procedures to produce eight "new" separate audience reports for each market. Thus, in addition to the nineteen *published* reports, 152 "new" reports (8 Jackknife Replicates × 19 markets) had to be generated by Arbitron solely for the purposes of this special study. In total then, 171 reports (19 published reports and 152 new Jackknife Replicate reports) were used for analysis purposes as explained in Steps 5 and 6.

Step 5: Determination Of Reliability Based On Jackknife Replication
For each market, the audience ratings derived for each *Jackknife Replicate report* were compared to the corresponding sets of audience ratings shown in Arbitron's *published reports.* Using the appropriate mathematical formulae, the differences observed between the ratings generated for each Jackknife Replicate report and the corresponding published report rating were expressed as a standard error quantity.

The observed differences between the corresponding ratings of each Jackknife Replicate versus the published report rating reflect, in part, the fact that the Jackknife Replicate sample sizes are *smaller* than the original published report samples. The larger sample sizes for the published report full sample (versus the Jackknife Replicate sample sizes) are factored into the proper mathematical formulae when calculating the standard errors of published report audience ratings. (This factoring is necessary to arrive at valid inferences concerning the reliability of the audience ratings in the *full sample* published report.)

In total, over one million standard errors were calculated via the Jackknife Replication empirical analysis. These calculations were the basis for determining the actual reliability levels of the ratings published in Arbitron's Local Market Radio Reports.

Step 6: Quantification Of Key Factors Impacting Reliability And Development Of Arbitron's New Radio Reliability Model
The more than *one million standard errors* calculated in Step 5 were

based upon an analysis of over *twelve million ratings estimates.*[6] These estimates covered a wide range of market types as described in Step 1.

A full range of audience ratings were also represented in the standard error calculations including: rating sizes, demographic groups, geographic areas, and estimate types (e.g., Cume and Average Quarter Hour) as published in Arbitron reports. This expansive data base permitted Arbitron to make the needed assessment of the varying influences on the reliability of radio audience ratings.

A complex statistical analysis (as described in the Methodology Section) was completed in order to isolate and measure the numerical impact on reliability of certain factors that can be viewed as either:

(a) *Common Factors:* Factors having essentially a "common" influence across all markets and reports (e.g., the impact on the standard error due to the collection of multiple observations on the same respondents and other sample clustering elements)

or

(b) *Unique Factors:* Factors having a "unique" influence specific to a particular market and report (e.g., sample balancing).

Importantly, the quantification of these varying influences on reliability enabled Arbitron to complete the statistical modelling process for the development of the *new and improved Radio Reliability model.* The technical formulation of Arbitron's new Radio Reliability Model is described in Chapter 5. The advantages and benefits to the user of Arbitron reports are described in the next chapter.

[6]*Technical Note:* The number of standard errors derived (over *one* million) differs from the number of audience estimates (over *twelve* million) used in the standard error calculations. When the numerical value for a rating contained in the published report is compared to the value generated for each of the eight Jackknife Replicates for the corresponding audience estimate, those *eight* numerical differences are summarized into *one* standard error quantity (via use of the appropriate mathematical formula for the standard error calculation).

Chapter 4

Study Results

An extensive and relatively complicated statistical analysis made it possible for Arbitron to successfully achieve the two study objectives that were cited in Chapter 3. The study results, as they relate to the successful achievement of each basic objective, are summarized in separate sections below:

A. Achievement of Objective #1 — *To develop procedures to enable report users to determine standard errors with greater accuracy, ease and speed*

1. Greater Accuracy

With Arbitron's new Radio Reliability Model, standard error estimates can now be calculated with *significantly greater accuracy* than was previously possible. The new Radio Reliability Model eliminates the persistent inaccuracies (biases) in standard error estimation that are associated with simple random sample formula approximations, as well as those associated with the Nomograph Procedure. The Nomograph Procedure, for example, overstates the size of the actual standard errors by 24 percent on average across the more than one million published estimates examined in the *Replication II* study.

Perhaps even more important is the fact that the biases related to the former cited approximation methods are not constant — that is, they vary dramatically by estimate type (Cume versus Average Quarter Hour), region (Metro, TSA and ADI) and demographic groups (e.g., Males 18-34 versus Total Persons 12+). This finding is described in greater detail in Section B of this chapter. The relevant point here is that Arbitron's new Radio Reliability Model eliminates these varying biases. In other words, as compared to previous approximation methods, *the new Radio Reliability Model yields significantly better estimates of the true reliabilities of Arbitron's published audience ratings and projections for radio.*

When applying this new model to determine the standard error of an audience rating published in any *future* Arbitron Local Market Radio Report, numerical inputs for the mathematical variables and parameters of the new model come from two sources: (1) Information derived from the *Replication II* study, per se, and (2) Information obtained by making specific calculations based upon the individual data base associated with each particular *future* Arbitron Local Market Radio Report for which standard error evaluations are to be made.

The model information derived from the *Replication II* study consists of a set of numerical values—over 600 of them—that reflect the effects on reliability of certain factors that can be viewed as *common* across all Arbitron Local Market Radio Reports as described in Chapter 3. These are called *Common Design Factors* in the technical description given in Chapter 5 of this report. In Chapter 5, the calculations that must be performed individually for all specific future report data bases are referred to as *Unique Design Factors.* These specific report calculations are designed to measure the impact on reliability of factors that are *unique* to a particular Market Report (for example, the effect of the specific "sample balancing" process applied to a particular report).

2. Greater Ease and Speed for Report User

The above outlined complexities are required in order to gain a significant improvement in the accuracy of standard error estimation; however, these complexities posed a challenging dilemma in terms of meeting the second aspect of Objective #1—the development of a procedure whereby Arbitron report users could calculate standard errors with *greater ease* and *greater speed* than any procedure currently available. Arbitron successfully solved this dilemma by reformulating its conceptual model in a way that maintained the model's validity, but, at the same time, allowed Arbitron to take responsibility for the burdensome calculations.

The mathematical structure of the "conceptual" and the "reformulated" models are given in Chapter 5 of this report, but from the report users perspective, the new method consists of a simple two-step arithmetic procedure. Use of the new method involves selecting an appropriate value from each of two reference tables and dividing the value from Table A by the value from Table B. The two reference tables will be provided by Arbitron, and they will be included in every Standard and Condensed Arbitron Local Market Radio Report beginning with the Fall, 1981 Reports.

The values contained in *Table A will be the same across all Arbitron Radio Reports.* Table A contains "look up" values for each possible rating betweeen 0.1 and 99.9.

In Table B, the values are individually derived for each report, and therefore, the values in *Table B will vary from report to report.* The Table B value is obtained by reference to the appropriate section of the table that defines the particular published audience rating (for which the standard error determination is being made) in terms of its *geographic region* (e.g., Metro Survey Area), *demographic group* (e.g., Men 18+) and *estimate type* (e.g., Monday–Friday 6–10 AM Average Quarter Hour rating).

An Example

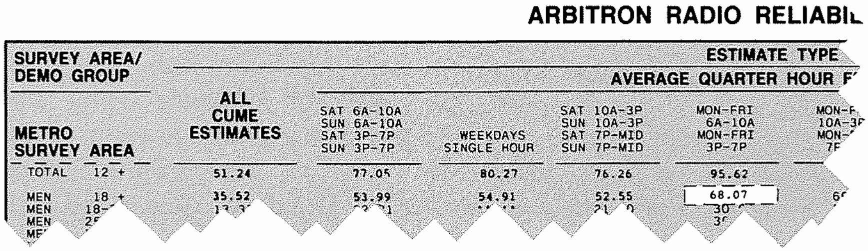
As an example, suppose that a station had a Monday–Friday 6–10AM published AQH rating of 1.0 among Men 18+ in the Metro Survey Area. The "value" listed in Table A for a Rating of 1.0 is 9.95 as shown in the abbreviated illustration of Table A in Figure 3 below.

ARBITRON RADIO RELIABILITY — TABLE A

RATING	VALUE	RATING	VALUE	RATING	VALUE	RATING	VALUE	RATING	VALUE
0.1	3.16	5.1	22.00	10.1	30.13	15.1	35.80	20.1	40.07
0.2	4.47	5.2	22.20	10.2	30.26	15.2	35.90	20.2	40.15
0.3	5.47	5.3	22.40	10.3	30.40	15.3	36.00	20.3	40.[illegible]
0.4	6.31	5.4	22.60	10.4	30.53	15.4	36.09	20.4	40.30
0.5	7.05	5.5	22.80	10.5	30.66	15.5	36.19	20.5	40.37
0.6	7.72	5.6	22.99	10.6	30.78	15.6	36.29	20.6	40.[illegible]
0.7	8.34	5.7	23.18	10.7	30.91	15.7	36.38	20.7	40.5[illegible]
0.8	8.91	5.8	23.37	10.8	31.04	15.8	36.47	20.8	40.59
0.9	9.44	5.9	23.56	10.9	31.16	15.9	36.57	20.9	40.6[illegible]
$\boxed{1.0}$	$\boxed{9.95}$	6.0	23.75	11.0	31.29	16.0	36.66	21.0	40.[illegible]
1.1	10.43	6.1	23.93	11.1	31.41	16.1	36.75	21.1	40.80
1.2	10.89	6.2	24.12	11.2	31.54	16.2	36.85	21.2	40.87
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Figure 3. Portion of Table A.

To find the appropriate Table B value, locate the *demographic group* ("Men 18+") and *estimate type* ("Average Quarter Hour Mon–Fri 6–10AM") in the "Metro Survey Area" section of Table B. In this example, the Table B value is *68.07* as shown in the abbreviated illustration of Table B in Figure 4 below.



ARBITRON RADIO RELIABI

SURVEY AREA/ DEMO GROUP		ESTIMATE TYPE					
		AVERAGE QUARTER HOUR F					
METRO SURVEY AREA	ALL CUME ESTIMATES	SAT 6A-10A SUN 6A-10A SAT 3P-7P SUN 3P-7P	WEEKDAYS SINGLE HOUR	SAT 10A-3P SUN 10A-3P SAT 7P-MID SUN 7P-MID	MON-FRI 6A-10A MON-FRI 3P-7P	MON-F 10A-3P MON- 7P
TOTAL 12 +	51.24	77.05	80.27	76.26	95.62	
MEN 18 +	35.52	53.99	54.91	52.55	68.07	6

Figure 4. Portion of Table B.

The standard error (one sigma value) is obtained simply by dividing the Table A value (9.95) by the value obtained from Table B (68.07). In this example, the one sigma standard error value is approximately equal to one-tenth of a rating point, as shown below:

$$\frac{\text{Standard Error}}{\text{(One Sigma Value)}} = \frac{\text{Table A value}}{\text{Table B value}} = \frac{9.95}{68.07} \cong 0.146 \cong 0.1$$

The upper and lower confidence interval bounds discussed earlier are obtained by adding and subtracting an appropriate multiple of the standard error to the published rating. For the 90% Confidence Interval, the appropriate standard error multiple is 1.64. The calculations are demonstrated below where it is assumed that the published rating is 1.0:

90% Confidence Interval Calculations

Lower Boundary: $1.0 - (0.146 \times 1.64) \cong 0.761 \cong 0.8$

Upper Boundary: $1.0 + (0.146 \times 1.64) \cong 1.239 \cong 1.2$

This confidence interval calculation would indicate that the particular station's published rating could have been as low as .8, or as high as 1.2, if the total population frame from which the sample were drawn had been surveyed by Arbitron under identical conditions.

The appropriate standard error multiples for computing lower and upper bounds for other degrees of confidence (i.e., 68%, 80%, 95% and 99%), are given in Chapter 8. They also are shown in Appendix A, along with an example of an actual Table A and Table B which appeared in the Fall 1981, New York Local Market Radio Report. Appendix A also contains an illustration which explains how to obtain the standard error of a projected audience *number* (as contrasted to a *rating*).

3. Additional Benefit: More Refined Estimates of Effective Sample Bases (ESB's)

The Effective Sample Base is a commonly used summary index of reliability adopted by broadcast researchers. It provides a summarization of reliability levels in terms of a benchmark called a "simple random sample." The Effective Sample Base is the size of a "simple random sample" necessary to provide the same degree of reliability that is associated with an audience rating derived from a complex sample design such as Arbitron's.

Prior to the Fall 1981 radio reports, Arbitron provided Effective Sample Base (ESB) information, but it was not as statistically refined or as extensive as the ESB information provided with the new Reliability Model.

Now, in addition to the greater accuracy, ease and speed in making standard error determinations, Arbitron's new Reliability Model formulation has an extra benefit for the more research-oriented report user. By simply squaring the values contained in Table B, the *Effective Sample Bases* (ESB's) are obtained for *every* audience rating and projection published in the corresponding report. For instance, in the example above, the Table B value for *"Men 18+" in the "Metro Survey Area" for the estimate type "Average Quarter Hour Mon–Fri 6–10AM"* is 68.07. To obtain the corresponding ESB for audience estimates meeting these criteria, simply square the Table B value to obtain, in this instance, an ESB of 4634 (68.07 squared). Thus, the statistically more refined estimates of ESB provided by Arbitron's new Reliability Model are an important additional benefit.

B. Achievement of Objective #2—*To make an overall evaluation of the reliability levels of Arbitron radio audience estimates*

1. Average "Actual" Reliability Better Than Previously Estimated

To determine how accurately the Nomograph Procedure approximated actual standard errors, "actual" standard errors were computed empirically by the Jackknife Replication procedure for each of the 1,377,296 published audience estimates examined in this study. These "actual" standard errors were compared to the corresponding standard error estimates that were derived by using the Nomograph Procedure. A Comparison Ratio was formed as shown in Equation

4-1 which follows:

$$\text{Comparison Ratio} = \frac{\text{(Standard error estimated via Nomograph Procedure}^{7}\text{)}}{\text{("Actual" standard error determined empirically by Jackknife Replication)}} \tag{4-1}$$

Across all 1,377,296 audience estimates, the average Comparison Ratio was estimated to be 1.238. Therefore, it was determined that the Nomograph Procedure overstated the size of the "actual" standard error by about 24% on average across the more than one million estimates examined.[8] This finding means that *Arbitron Radio audience estimates are, on average, significantly more reliable (that is, have smaller sampling error) and can be used with greater confidence than previously believed.*

Since standard errors vary inversely with the square root of ESB (Effective Sample Base) values, the 24% average overstatement by the Nomograph of "actual" standard errors also implies that *the average ESB of Arbitron Radio audience estimates is 53% higher than previously reported.*[9]

Arbitron also made an analysis of the Comparison Ratio defined by Equation 4-1 for the separate groupings defined by 32 *demographic* categories, 11 *estimate type* categories, and 3 *geographic region* categories. The categories for each variable are defined in Tables 1 2, and 3.

---

[7]To use the Nomograph Procedure, an Effective Sample Base (ESB) value is required. Prior to the Fall 1981 reports, ESB values were provided for the total sample (Total Persons 12+) in the Metro area, the TSA area and, if applicable, the ADI area. Calculation of the standard error for age-sex groups required the determination of ESB estimates for the age-sex groups. The required ESB for age-sex groups was obtained by multiplying the total sample ESB for the appropriate region by the In-Tab sample proportion represented by a particular age-sex group within the region.

The Nomograph Procedure expressed reliability as a *two* standard error quantity. Hence, the Nomograph reliability estimates must be divided by two in order to adjust them to the *one* standard error value.

[8]For the sake of calculation simplicity and to minimize the possible undue influence of a few large individual ratios, the average amount of overstatement was derived by taking the ratio of the average Design Factor for Nomograph standard errors to the average Design Factor calculated for "actual" standard errors. The derivation of the average Design Factor quantities is described in Section D of Chapter 7.

[9]The 53% higher ESB is determined from the average Comparison Ratio (1.238, as cited earlier) as follows:

$$\text{ESB Increase} = [(1.238)^2 - 1.000] \times 100 = 53.2664 \text{ (or approximately 53\%)}$$

TABLE 1. DEFINITION OF DEMOGRAPHIC CATEGORIES

1. Men 18+	13. Women 18+	25. Adults 18+
2. Men 18-24	14. Women 18-24	26. Adults 18-34
3. Men 25-34	15. Women 25-34	27. Adults 18-49
4. Men 35-44	16. Women 35-44	28. Adults 25-49
5. Men 45-54	17. Women 45-54	29. Adults 25-54
6. Men 55-64	18. Women 55-64	30. Adults 35-64
7. Men 18-34	19. Women 18-34	31. Teens 12-17
8. Men 18-49	20. Women 18-49	32. Total Persons 12+
9. Men 25-44	21. Women 25-44	
10. Men 25-49	22. Women 25-49	
11. Men 25-54	23. Women 25-54	
12. Men 35-64	24. Women 35-64	

TABLE 2. DEFINITION OF ESTIMATE TYPE CATEGORIES

Estimate Type	Description	Number of Quarter Hours
1. Cume	• All Cume Estimates	Not Applicable
Average Quarter Hour (AQH):		
2. Type A	• Sat 6A–10A; Sun 6A–10A; Sat 3P–7P; Sun 3P–7P	16
3. Type B	• Weekdays Single Hour	20
4. Type C	• Sat 10A–3P; Sun 10A–3P; Sat 7P–MID; Sun 7P–MID	20
5. Type D	• Mon-Fri 6A–10A; Mon-Fri 3P–7P	80
6. Type E	• Mon-Fri 10A–3P; Mon-Fri 7P–MID	100
7. Type F	• Sat-Sun 6A–MID	144
8. Type G	• Mon-Fri 6A–10A + Mon-Fri 3P–7P	160
9. Type H	• Mon-Fri 6A–7P	260
10. Type I	• Mon-Fri 6A–MID	360
11. Type J	• Mon-Sun 6A–MID	504

TABLE 3. DEFINITION OF GEOGRAPHIC REGION CATEGORIES

1. Metro Survey Area (Metro)
2. Total Survey Area (TSA)
3. Area of Dominant Influence (ADI)

The Comparison Ratios can vary dramatically by the individual "cells" that are defined by any particular combination of demographic, estimate type and geographic region categories. The Comparison Ratios that were tabulated for these cells are provided in Appendix E. In the Metro area, for example, the average Comparison Ratio ranges from a 0.83 (for Total Persons 12+, Cume estimates) to a 2.79 (for Teen, AQH Mon–Sun 6A–Mid estimates). This is a rather extreme example of the range of values among these ratios. In some instances, the average Comparison Ratio was very close to 1.00 (e.g., the average Comparison Ratios for Cume estimates among the Men and Women subgroups).

In terms of generalized conclusions, the following important findings can be cited:

1. Across the 744,478 *AQH audience estimates* examined in this study, the average Comparison Ratio is 1.61 (i.e., the Nomograph Procedure overstates "actual" standard errors of AQH audience estimates by 61% on average across the 744,478 cases examined).

2. Across the 632,828 *Cume audience estimates* examined in this study, the average Comparison Ratio is 0.97 (i.e., a 3% understatement of "actual" standard errors by the Nomograph Procedure).

3. Across the 1,302,971 *Demographic Subgroup audience estimates* examined in this study, the average Comparison Ratio is 1.25 (i.e., a 25% overstatement of "actual" standard errors by the Nomograph Procedure).

4. Across the 74,505 *Total Persons 12+ audience estimates* examined in this study, the average Comparison Ratio is 1.08 (i.e., an 8% overstatement of "actual" standard errors by the Nomograph Procedure).

It can be seen from the preceding findings, and from a study of the Comparison Ratios given by cell in Appendix E, that the Nomograph approximation is reasonably accurate on average only for Cume estimates. By contrast, an analysis of the 534 AQH estimate cells in Appendix E shows a persistent bias in the Nomograph Procedure approximation toward overstating the size of the "actual" standard error for AQH estimates. This overstatement of "actual" standard errors for AQH estimates is larger for demographic subgroups than it is for the total sample (i.e., the Total Persons 12+ group).

The significance of these findings is that it demonstrates that the degree of disparity between the "actual" standard errors and those estimated by the Nomograph Procedure is *not constant*. The disparity varies substantially among cells in the 3 dimensional array defined in Appendix E by the 32 *demographic*, 3 *geographic* and 11 *estimate type* categories for which Arbitron publishes audience estimates in its Local Market Radio Reports.

This fact has important implications in terms of building an improved reliability model. It clearly indicates *that no single adjustment factor* can adequately correct the inaccuracies in the former approximation method. As a result, Arbitron's new Reliability Model is considerably more complex than the model on which the Nomograph Procedure is based. The new model is designed to eliminate the persistent overstatement and sometimes, understatement of reliability of the former approximation method.

The statistical structure of Arbitron's new Reliability Model and its contrast to the Nomograph and Simple Random Sample approximation models, is described in Chapter 5. A full understanding of the structure of the new model—and why it offers a significant improvement in the accuracy of standard error determinations—requires a discussion of several other key findings in this study. These additional findings are centered about the concept of "statistical efficiency" and they are described in the following sections of this chapter.

2. Statistical Efficiency of Arbitron Samples

Because the standard error quantity is specific to individual ratings and projections, other measures are required in order to summarize the behavior of Arbitron audience ratings and projections in terms of their reliability. One index that is particularly useful for making comparative global evaluations of the reliability behavior of a set of ratings or projections is a measure called *Statistical Efficiency*. The Statistical Efficiency measure expresses the ESB value *relative* to the actual sample size (In-tab). For example, if the ESB value is 1400 and the actual sample size (In-tab) is 1600, the Statistical Efficiency is .875 (1400 ÷ 1600 = .875)—or 87.5 percent, when expressed as a percent. Similarly, if the ESB is 3600 and the actual sample size is 1600, the Statistical Efficiency is 2.25 (3600 ÷ 1600 = 2.25)—or 225.0 percent when expressed as a percent. A Statistical Efficiency value greater than 100 percent means that the sample used in the actual survey design is more reliable than a Simple Random Sample of corresponding size.

a. *Demographic Subgroups Have Higher Statistical Efficiencies Compared To Total Sample.* Based upon the Jackknife Replication analysis, actual average Statistical Efficiencies were calculated for the total sample (Total Persons 12+) and for each demographic group reported in Arbitron Local Market Radio Reports. For ease of comparison, the demographic subgroup Statistical Efficiencies were indexed to the Total Persons 12+ values. This indexing was done separately by each of the eleven estimate types (defined previously in Table 2) within each of the three geographic regions (Metro, TSA and ADI). The results are presented in Tables 4a (Metro), 4b (TSA), and 4c (ADI) on the following pages.

It is readily apparent from an examination of Tables 4a, 4b, and 4c that the Statistical Efficiencies for demographic subgroups are generally much higher than those for the total sample (i.e., the Total Persons 12+). This holds true for each *geographic region* and each *estimate type*. There is also a strong general tendency for Statistical Efficiencies to increase as the definition of the demographic subgroup becomes more narrowly delineated (e.g., "Men 18–24" have higher Statistical Efficiencies than do "Men 18+").

The findings of increased Statistical Efficiency for demographic subgroups is explained by the statistical effect of "clustering". More specifically, Arbitron samples include more than one respondent from the same household. As a result, a household can be viewed as a "cluster" of respondents. Any positive correlation in listening behavior among respondents within the same households tends to produce a lower Statistical Efficiency for the sample. Such positive correlations within households serve to reduce the Effective Sample Base (ESB) for the survey audience estimates.

As compared to the Total Persons 12+ audience estimates, demographic subgroups are subject to the least amount of clustering. As a result, when estimates are based upon demographic subgroups, rather than upon the total sample, they tend to be statistically more efficient on a per-case basis since the amount of clustering is lower. This is particularly true for the more narrowly defined age-sex subgroups. For example, many households contain both an adult male and an adult female. This results in a *within-household* clustering effect for any audience estimates reported for Total Adults. By contrast, the average within-household cluster size is significantly reduced for audience estimates reported in terms of "Men Only" or "Women Only".

The fact that demographic subgroups have higher Statistical Efficiencies compared to the total sample has two important implications.

TABLE 4a DEMOGRAPHIC SUBGROUP STATISTICAL EFFICIENCIES INDEXED* TO THE TOTAL PERSONS 12 + STATISTICAL EFFICIENCIES FOR EACH OF THE ELEVEN ESTIMATE TYPES DEFINED IN TABLE 2 (Metro)

	Estimate Type (See Table 2 for Detailed Definitions)										
Demographic Group	#1 All Cume Est.	#2 AQH-Type A (16)**	#3 AQH-Type B (20)**	#4 AQH-Type C (20)**	#5 AQH-Type D (80)**	#6 AQH-Type E (100)**	#7 AQH-Type F (144)**	#8 AQH-Type G (160)**	#9 AQH-Type H (260)**	#10 AQH-Type I (360)**	#11 AQH-Type J (504)**
Total Persons 12+	100	100	100	100	100	100	100	100	100	100	100
Men 18+	130	133	127	129	137	120	135	127	128	133	136
Men 18-24	131	162	***	149	199	171	179	179	197	198	199
Men 25-34	147	162	***	160	176	157	184	159	164	177	184
Men 35-44	148	154	153	159	170	145	177	148	159	158	166
Men 45-54	145	142	142	153	159	139	157	141	149	154	158
Men 55-64	152	152	***	160	177	157	168	160	184	189	195
Men 18-34	138	151	151	144	166	144	***	***	***	***	171
Men 18-49	134	143	137	138	149	124	***	***	***	***	147
Men 25-44	***	***	147	***	***	***	***	***	***	***	***
Men 25-49	148	150	143	149	155	130	***	***	***	***	154
Men 25-54	146	144	***	145	150	125	***	***	***	***	147
Men 35-64	148	139	***	148	153	129	***	***	***	***	152
Women 18+	133	127	124	127	124	125	134	124	124	123	125
Women 18-24	142	163	***	149	176	165	182	158	170	172	185
Women 25-34	149	158	***	159	171	157	171	157	169	175	173
Women 35-44	147	152	157	157	166	152	175	143	161	163	173
Women 45-54	147	147	141	146	149	139	154	135	147	150	147
Women 55-64	156	143	***	146	152	149	151	137	156	160	157
Women 18-34	145	152	153	145	158	146	***	***	***	***	163
Women 18-49	135	140	138	136	142	128	***	***	***	***	142
Women 25-44	***	***	150	***	***	***	***	***	***	***	***
Women 25-49	148	144	144	146	151	134	***	***	***	***	152
Women 25-54	147	141	***	143	147	128	***	***	***	***	147
Women 35-64	149	138	***	141	142	127	***	***	***	***	140
Adult 18+	106	103	***	102	102	95	***	***	***	***	102
Adult 18-34	122	129	***	123	136	125	***	***	***	***	134
Adult 18-49	114	118	***	114	118	105	***	***	***	***	117
Adult 25-49	122	121	***	121	124	109	***	***	***	***	123
Adult 25-54	120	116	***	118	120	105	***	***	***	***	120
Adult 35-64	123	113	***	118	121	107	***	***	***	***	119
Teens	130	154	181	163	186	223	172	187	228	214	201

*Index = (Subgroup Statistical Efficiency ÷ Total Persons 12+ Statistical Efficiency) × (100)
**Number in "( )" equals number of quarter hours averaged in each AQH (Average Quarter Hour) estimate type.
***Arbitron's Local Market Radio Reports do not include audience estimates for this cell; therefore, index not available.

TABLE 4b DEMOGRAPHIC SUBGROUP STATISTICAL EFFICIENCIES INDEXED* TO THE TOTAL PERSONS 12+ STATISTICAL EFFICIENCIES FOR EACH OF THE ELEVEN ESTIMATE TYPES DEFINED IN TABLE 2 (TSA)

	Estimate Type (See Table 2 for Detailed Definitions)										
Demographic Group	#1 All Cume Est.	#2 AQH-Type A (16)**	#3 AQH-Type B (20)**	#4 AQH-Type C (20)**	#5 AQH-Type D (80)**	#6 AQH-Type E (100)**	#7 AQH-Type F (144)**	#8 AQH-Type G (160)**	#9 AQH-Type H (260)**	#10 AQH-Type I (360)**	#11 AQH-Type J (504)**
Total Persons 12+	100	100	100	100	100	100	100	100	100	100	100
Men 18+	132	134	129	131	140	120	138	129	132	138	137
Men 18-24	143	174	***	165	212	176	193	184	206	211	213
Men 25-34	151	177	***	174	185	162	198	160	174	186	194
Men 35-44	154	166	161	171	173	150	191	147	161	165	173
Men 45-54	140	143	141	153	155	134	157	133	141	147	149
Men 55-64	151	156	***	165	175	159	167	158	190	189	192
Men 18-34	143	161	160	157	177	146	***	***	***	***	183
Men 18-49	139	147	141	145	152	124	***	***	***	***	150
Men 25-44	***	***	152	***	***	***	***	***	***	***	***
Men 25-49	147	155	145	155	155	130	***	***	***	***	154
Men 25-54	145	148	***	150	151	125	***	***	***	***	145
Men 35-64	146	142	***	150	152	129	***	***	***	***	151
Women 18+	130	128	123	126	122	123	136	123	122	121	125
Women 18-24	134	169	***	149	171	153	185	147	161	162	176
Women 25-34	151	164	***	170	180	161	184	158	179	187	183
Women 35-44	151	164	164	165	169	154	187	147	167	168	176
Women 45-54	145	153	144	156	150	140	164	131	148	151	149
Women 55-64	153	149	***	154	156	151	158	136	166	162	157
Women 18-34	142	158	152	147	160	140	***	***	***	***	164
Women 18-49	137	144	137	137	141	122	***	***	***	***	143
Women 25-44	***	***	155	***	***	***	***	***	***	***	***
Women 25-49	147	149	149	152	154	135	***	***	***	***	157
Women 25-54	147	146	***	149	150	129	***	***	***	***	152
Women 35-64	146	145	***	146	143	127	***	***	***	***	144
Adult 18+	105	103	***	102	104	93	***	***	***	***	103
Adult 18-34	123	134	***	129	140	120	***	***	***	***	142
Adult 18-49	116	119	***	116	122	103	***	***	***	***	122
Adult 25-49	122	123	***	124	126	110	***	***	***	***	125
Adult 25-54	120	119	***	121	121	106	***	***	***	***	120
Adult 35-64	121	115	***	121	121	109	***	***	***	***	120
Teens	128	153	180	164	173	217	175	172	215	200	187

*Index = (Subgroup Statistical Efficiency ÷ Total Persons 12+ Statistical Efficiency) × (100)

**Number in "( )" equals number of quarter hours averaged in each AQH (Average Quarter Hour) estimate type.

***Arbitron's Local Market Radio Reports do not include audience estimates for this cell; therefore, value not available.

TABLE 4c DEMOGRAPHIC SUBGROUP STATISTICAL EFFICIENCIES INDEXED* TO THE TOTAL PERSONS 12+ STATISTICAL EFFICIENCIES FOR EACH OF THE ELEVEN ESTIMATE TYPES DEFINED IN TABLE 2 (ADI)

	Estimate Type (See Table 2 for Detailed Definitions)										
Demographic Group	#1 All Cume Est.	#2 AQH-Type A (16)**	#3 AQH-Type B (20)**	#4 AQH-Type C (20)**	#5 AQH-Type D (80)**	#6 AQH-Type E (100)**	#7 AQH-Type F (144)**	#8 AQH-Type G (160)**	#9 AQH-Type H (260)**	#10 AQH-Type I (360)**	#11 AQH-Type J (504)**
Total Persons 12+	100	***	***	***	***	***	***	***	***	***	100
Men 18+	161	****	***	****	****	****	***	***	***	***	133
Men 18-24	146	***	***	***	***	***	***	***	***	***	186
Men 25-34	173	***	***	***	***	***	***	***	***	***	180
Men 35-44	170	***	***	***	***	***	***	***	***	***	171
Men 45-54	170	***	***	***	***	***	***	***	***	***	148
Men 55-64	164	***	***	***	***	***	***	***	***	***	188
Men 18-34	173	****	***	****	****	****	***	***	***	***	***
Men 18-49	169	****	***	****	****	****	***	***	***	***	***
Men 25-44	***	***	***	***	***	***	***	***	***	***	***
Men 25-49	181	****	***	****	****	****	***	***	***	***	***
Men 25-54	178	****	***	****	****	****	***	***	***	***	***
Men 35-64	179	****	***	****	****	****	***	***	***	***	***
Women 18+	162	****	***	****	****	****	***	***	***	***	129
Women 18-24	164	***	***	***	***	***	***	***	***	***	184
Women 25-34	162	***	***	***	***	***	***	***	***	***	175
Women 35-44	159	***	***	***	***	***	***	***	***	***	172
Women 45-54	174	***	***	***	***	***	***	***	***	***	152
Women 55-64	186	***	***	***	***	***	***	***	***	***	155
Women 18-34	177	****	***	****	****	****	***	***	***	***	***
Women 18-49	171	****	***	****	****	****	***	***	***	***	***
Women 25-44	***	***	***	***	***	***	***	***	***	***	***
Women 25-49	179	****	***	****	****	****	***	***	***	***	***
Women 25-54	180	****	***	****	****	****	***	***	***	***	***
Women 35-64	182	****	***	****	****	****	***	***	***	***	***
Adult 18+	131	****	***	****	****	****	***	***	***	***	***
Adult 18-34	151	****	***	****	****	****	***	***	***	***	***
Adult 18-49	142	****	***	****	****	****	***	***	***	***	***
Adult 25-49	149	****	***	****	****	****	***	***	***	***	***
Adult 25-54	146	****	***	****	****	****	***	***	***	***	***
Adult 35-64	151	****	***	****	****	****	***	***	***	***	***
Teens	162	****	***	****	****	****	***	***	***	***	194

*Index = (Subgroup Statistical Efficiency ÷ Total Persons 12+ Statistical Efficiency) × (100)

**Number in "( )" equals number of quarter hours averaged in each AQH (Average Quarter Hour) estimate type.

***Arbitron's Local Market Radio Reports do not include audience estimates for this cell; therefore, value not available.

****Index can not be calculated since Arbitron's Local Market Radio Reports do not include Total Person 12+ audience estimates for this estimate type.

First, it means that the diminution in reliability due to smaller sample sizes for subgroups is partially offset by the per-case increase in Statistical Efficiency. *Second*, this finding clearly indicates that any model developed for the determination of standard errors must employ separate parameters for each individual demographic grouping. Separate parameters are necessary in order to accurately reflect the noted differences in Statistical Efficiencies among the various sub-bases for which audience estimates are reported.

Unlike the former Nomograph Procedure, *Arbitron's new Reliability Model accurately reflects the differences in Statistical Efficiency values by demographic group*. How the new model accomplishes this important objective is explained in Chapter 5.

b. *Statistical Efficiencies for Average Quarter Hour (AQH) Audience Estimates Improve As The Number of Quarter Hours In the Reporting Period Increases.* Another major reason why the Nomograph Procedure typically understates the true reliability of Arbitron audience estimates is its implied statistical assumption that each individual in the sample contributes only a single observation to any reported audience estimate. In a statistical sense, this is true for Cume audience estimates, but it is *not true* for Average Quarter Hour ratings and projections. Average Quarter Hour audience estimates are derived by averaging multiple observations on the same individual.

At a minimum, each individual in the In-tab sample contributes 16 observations to each AQH Audience rating and projection included in Arbitron Local Market Radio Reports. In particular, the 16 observations relates to the "Type A" audience estimates defined previously in Table 2. As shown in Table 2, most AQH audience estimates provided in Arbitron reports include more than 16 quarter-hour observations. The maximum number of observations contributed by each individual in the sample is reflected in the AQH estimate identifed in Table 2 as "Type J" (Mon–Sun 6AM–Midnight). Type J includes 504 quarter-hour measurements.

These multiple observations from each individual have the effect of increasing the Effective Sample Base of Arbitron samples, and in turn, their Statistical Efficiencies. In other words, in terms of reliability, the multiple observations are somewhat similar to actual increases in in-tab sample sizes. However, the multiple observations are not statistically independent; instead they are correlated—e.g., if the individual is listening to Station "X" between 9:30 AM and 9:45 AM, there is a higher than average likelihood that this person will still be listening to Station "X" during the next quarter-hour time segment (9:45 AM–10:00 AM).

As a result, the multiple observations are not completely equivalent to corresponding increases in actual in-tab sample sizes. In other words, 504 multiple observations made on the *same* individual are not quite equal statistically to 504 observations made on *different* individuals. Nevertheless, even with the noted correlation, the result of the multiple observations is an appreciable gain in overall reliability.

In general, for AQH ratings and projections, as the number of quarter hours in the reporting period increases so does the Statistical Efficiency associated with the corresponding audience estimate. This general improvement in Statistical Efficiency which accompanies AQH audience estimates as the number of quarter hours included therein increases holds true for demographic subgroups, as well as for the Total Persons 12+ group. It also holds true for Metro, TSA, and ADI audience estimates. These findings are demonstrated by the indices given in Tables 5a (Metro), 5b (TSA), and 5c (ADI).

To facilitate comparisons, Tables 5a, 5b, and 5c show the AQH Statistical Efficiency values *indexed* to the Statistical Efficiency values of a Cume audience estimate (for the corresponding demographic group). These exhibits are useful for two purposes. *First*, they clearly support the fact that the Statistical Efficiency for AQH audience estimates show a pattern of increasing values as the number of quarter hours contained in the AQH audience estimate increases. For example, in the Metro area (Table 5a), the index value among "Men 35–44" for the "Type J" AQH audience estimate is *629*. [The "Type J" audience estimates includes 504 quarter hours as explained previously in Table 2.] By contrast, the "Type A" audience estimate contains just 16 quarter-hour observations, and the index among "Men 35–44" is *236* (only about one-third the size of the 629 index for "Type J" audience estimate).[10]

Again, unlike the former Nomograph Procedure, *Arbitron's new Reliability Model accurately reflects the differences in Statistical Efficiency values among the various types of Average Quarter Hour estimates.* The manner in which the new model accomplishes this important objective is explained in Chapter 5.

The *second* purpose of Tables 5a, 5b, and 5c is to contrast the Statistical Efficiencies of AQH audience estimates and Cume audience estimates. This contrast is discussed in the following section.

---

[10]See *Appendix B* for a further technical discussion concerning the Per-Case Statistical Efficiency of AQH audience estimates.

TABLE 5a AQH STATISTICAL EFFICIENCIES INDEXED* TO CUME STATISTICAL EFFICIENCIES BY DEMOGRAPHIC GROUP

(Metro)

Demographic Group	#1 All Cume Est.	#2 AQH-Type A (16)**	#3 AQH-Type B (20)**	#4 AQH-Type C (20)**	#5 AQH-Type D (80)**	#6 AQH-Type E (100)**	#7 AQH-Type F (144)**	#8 AQH-Type G (160)**	#9 AQH-Type H (260)**	#10 AQH-Type I (360)**	#11 AQH-Type J (504)**
	Estimate Type (See Table 2 for Detailed Definitions)										
Total Persons 12+	100	226	245	221	348	354	486	488	415	490	562
Men 18+	100	231	239	219	367	326	505	475	409	500	586
Men 18-24	100	279	***	251	530	463	664	665	624	741	853
Men 25-34	100	249	***	240	415	378	607	526	463	589	703
Men 35-44	100	236	254	238	401	348	580	489	446	525	629
Men 45-54	100	221	240	232	381	340	526	474	425	520	612
Men 55-64	100	226	***	234	407	368	537	514	505	611	723
Men 18-34	100	247	268	230	419	369	***	***	***	***	693
Men 18-49	100	241	250	227	387	326	***	***	***	***	616
Men 25-44	***	***	****	***	***	***	***	***	***	***	***
Men 25-49	100	229	237	222	365	311	***	***	***	***	583
Men 25-54	100	223	***	220	358	302	***	***	***	***	567
Men 35-64	100	213	***	222	360	309	***	***	***	***	577
Women 18+	100	216	228	211	325	334	488	456	386	454	529
Women 18-24	100	260	***	232	433	413	625	544	499	597	735
Women 25-34	100	240	***	237	400	373	560	516	472	576	655
Women 35-44	100	234	262	237	394	368	578	476	457	545	663
Women 45-54	100	226	235	221	355	335	511	451	415	502	562
Women 55-64	100	208	***	208	340	339	471	432	417	505	566
Women 18-34	100	237	259	221	379	356	***	***	***	***	629
Women 18-49	100	235	251	224	366	336	***	***	***	***	593
Women 25-44	***	***	****	***	***	***	***	***	***	***	***
Women 25-49	100	220	240	219	355	320	***	***	***	***	578
Women 25-54	100	215	***	215	349	308	***	***	***	***	562
Women 35-64	100	209	***	210	333	302	***	***	***	***	528
Adult 18+	100	218	***	212	335	315	***	***	***	***	540
Adult 18-34	100	238	***	223	388	362	***	***	***	***	618
Adult 18-49	100	233	***	220	362	325	***	***	***	***	578
Adult 25-49	100	224	***	220	355	316	***	***	***	***	567
Adult 25-54	100	219	***	218	350	310	***	***	***	***	564
Adult 35-64	100	208	***	213	343	310	***	***	***	***	547
Teens	100	268	343	278	501	610	646	706	731	811	869

*Index = (AQH Statistical Efficiency ÷ Total Persons 12+ Statistical Efficiency) × (100)

**Number in "( )" equals number of quarter hours averaged in each AQH (Average Quarter Hour) estimate type.

***Arbitron's Local Market Radio Reports do not include audience estimates for this cell; therefore, value not available.

****Index can not be calculated since Arbitron's Local Market Radio Reports do not include Total Person 12+ audience estimates for this estimate type.

TABLE 5b AQH STATISTICAL EFFICIENCIES INDEXED* TO CUME STATISTICAL EFFICIENCIES BY DEMOGRAPHIC GROUP

(TSA)

	Estimate Type (See Table 2 for Detailed Definitions)										
Demographic Group	#1 All Cume Est.	#2 AQH-Type A (16)**	#3 AQH-Type B (20)**	#4 AQH-Type C (20)**	#5 AQH-Type D (80)**	#6 AQH-Type E (100)**	#7 AQH-Type F (144)**	#8 AQH-Type G (160)**	#9 AQH-Type H (260)**	#10 AQH-Type I (360)**	#11 AQH-Type J (504)**
Total Persons 12+	100	215	235	207	335	347	458	481	398	472	541
Men 18+	100	218	230	206	357	315	480	469	398	493	564
Men 18-24	100	263	***	239	499	428	620	620	574	698	806
Men 25-34	100	251	***	239	409	371	599	509	456	580	692
Men 35-44	100	232	246	230	378	339	569	460	415	506	610
Men 45-54	100	219	236	227	370	331	513	454	400	496	573
Men 55-64	100	221	***	227	388	365	504	503	499	591	686
Men 18-34	100	242	262	228	414	354	***	***	***	***	691
Men 18-49	100	228	239	216	368	309	***	***	***	***	585
Men 25-44	***	***	****	***	***	***	***	***	***	***	***
Men 25-49	100	226	232	219	353	306	***	***	***	***	566
Men 25-54	100	219	***	215	348	298	***	***	***	***	541
Men 35-64	100	209	***	213	351	308	***	***	***	***	559
Women 18+	100	211	222	200	315	327	478	454	372	440	518
Women 18-24	100	271	***	230	427	397	631	526	477	570	710
Women 25-34	100	233	***	233	399	369	556	503	472	584	653
Women 35-44	100	234	255	227	376	353	566	469	441	526	631
Women 45-54	100	228	234	223	347	335	518	436	406	493	558
Women 55-64	100	209	***	209	342	341	474	428	431	499	556
Women 18-34	100	239	253	215	379	343	***	***	***	***	625
Women 18-49	100	224	234	206	345	307	***	***	***	***	562
Women 25-44	***	***	****	***	***	***	***	***	***	***	***
Women 25-49	100	218	238	215	351	318	***	***	***	***	578
Women 25-54	100	213	***	210	341	303	***	***	***	***	559
Women 35-64	100	213	***	208	329	301	***	***	***	***	534
Adult 18+	100	210	***	202	332	306	***	***	***	***	532
Adult 18-34	100	233	***	218	382	338	***	***	***	***	624
Adult 18-49	100	222	***	208	354	308	***	***	***	***	571
Adult 25-49	100	217	***	212	347	312	***	***	***	***	554
Adult 25-54	100	213	***	209	339	306	***	***	***	***	542
Adult 35-64	100	205	***	207	336	313	***	***	***	***	538
Teens	100	256	329	265	453	587	623	642	666	735	788

*Index = (AQH Statistical Efficiency ÷ Cume Statistical Efficiency) × (100)

**Number in "( )" equals number of quarter hours averaged in each AQH (Average Quarter Hour) estimate type.

***Arbitron's Local Market Radio Reports do not include audience estimates for this cell; therefore, value not available.

****Index can not be calculated since Arbitron's Local Market Radio Reports do not include Cume audience estimates for this demographic subgroup.

TABLE 5c AQH STATISTICAL EFFICIENCIES INDEXED* TO CUME STATISTICAL EFFICIENCIES BY DEMOGRAPHIC GROUP

(ADI)

Demographic Group	#1 All Cume Est.	#2 AQH-Type A (16)**	#3 AQH-Type B (20)**	#4 AQH-Type C (20)**	#5 AQH-Type D (80)**	#6 AQH-Type E (100)**	#7 AQH-Type F (144)**	#8 AQH-Type G (160)**	#9 AQH-Type H (260)**	#10 AQH-Type I (360)**	#11 AQH-Type J (504)**
	Estimate Type (See Table 2 for Detailed Definitions)										
Total Persons 12+	100	***	***	***	***	***	***	***	***	***	714
Men 18+	100	219	***	210	355	306	***	***	***	***	590
Men 18-24	100	***	***	***	***	***	***	***	***	***	909
Men 25-34	100	***	***	***	***	***	***	***	***	***	743
Men 35-44	100	***	***	***	***	***	***	***	***	***	721
Men 45-54	100	***	***	***	***	***	***	***	***	***	623
Men 55-64	100	***	***	***	***	***	***	***	***	***	816
Men 18-34	100	235	***	222	391	343	***	***	***	***	***
Men 18-49	100	223	***	215	353	306	***	***	***	***	***
Men 25-44	***	***	***	***	***	***	***	***	***	***	***
Men 25-49	100	224	***	221	352	318	***	***	***	***	***
Men 25-54	100	218	***	217	348	310	***	***	***	***	***
Men 35-64	100	210	***	218	356	312	***	***	***	***	***
Women 18+	100	216	***	208	323	323	***	***	***	***	571
Women 18-24	100	***	***	***	***	***	***	***	***	***	802
Women 25-34	100	***	***	***	***	***	***	***	***	***	768
Women 35-44	100	***	***	***	***	***	***	***	***	***	775
Women 45-54	100	***	***	***	***	***	***	***	***	***	624
Women 55-64	100	***	***	***	***	***	***	***	***	***	596
Women 18-34	100	239	***	220	378	360	***	***	***	***	***
Women 18-49	100	225	***	212	350	329	***	***	***	***	***
Women 25-44	***	***	***	***	***	***	***	***	***	***	***
Women 25-49	100	217	***	217	357	333	***	***	***	***	***
Women 25-54	100	212	***	211	347	316	***	***	***	***	***
Women 35-64	100	213	***	210	331	321	***	***	***	***	***
Adult 18+	100	210	***	204	332	315	***	***	***	***	***
Adult 18-34	100	230	***	217	370	342	***	***	***	***	***
Adult 18-49	100	220	***	209	345	316	***	***	***	***	***
Adult 25-49	100	216	***	213	346	327	***	***	***	***	***
Adult 25-54	100	211	***	211	342	322	***	***	***	***	***
Adult 35-64	100	205	***	208	340	324	***	***	***	***	***
Teens	100	257	***	264	475	568	***	***	***	***	855

*Index = (AQH Statistical Efficiency ÷ Cume Statistical Efficiency) × (100)

**Number in "( )" equals number of quarter hours averaged in each AQH (Average Quarter Hour) estimate type.

***Arbitron's Local Market Radio Reports do not include audience estimates for this cell; therefore, value not available.

c. *Average Quarter Hour (AQH) Audience Estimates Have Higher Statistical Efficiencies Compared to Cume Audience Estimates.* Cume audience ratings and projections are not only different from AQH audience estimates in terms of what they measure for the broadcast industry, but they differ from a statistical standpoint as well. *For an AQH audience estimate*, each individual in Arbitron's in-tab sample contributes their listening behavior *averaged* over all of the quarter-hour time segments that define the particular AQH audience estimate.

For Cume audience estimates, an individual respondent (diarykeeper) contributes a single "yes" or "no" (from a statistical viewpoint) for the entire time period defined by the Cume estimate—even though the diarykeeper actually records his or her viewing on a quarter-hour by quarter-hour basis.[11] For Cume Audience Estimates, the quarter-hour listening information is not averaged; instead, it is statistically summarized to answer a single question, namely, "Did the individual listen *at least once* during the defined reporting period?"

The multiple observations for a single person, which elevates Statistical Efficiency levels for AQH audience estimates, count statistically only as a single observation in the case of Cume audience estimates. Hence, statistical theory dictates that Cume audience estimates should show lower Statistical Efficiency levels than audience estimates calculated on an average quarter-hour basis.

The indices provided in Tables 5a, 5b, and 5c show that the empirical findings of this study are consistent with statistical theory. In every instance across the three Tables, the index for each AQH audience estimate is greater than the Cume audience benchmark which is indexed at 100. For instance, the index value is 629 *in the "Metro area" for the "Type J" AQH audience estimates among "Men 35–44"*. This means that the actual Statistical Efficiency for "*Type J*" *AQH audience estimates in the "Metro area" among "Men 35–44"* is approximately 6.29 times more efficient statistically, as compared to the corresponding Cume audience estimate for the same geographic and demographic group.

[11]In more precise terms, a *Cume Rating* is defined as the estimated number of Cume Persons expressed as a percentage of the total number of persons in the particular demographic group and geographic area being reported. *Cume Persons* is defined as the estimated number of different persons who listened at home and away from home to a particular station for a minimum of five minutes in any quarter-hour within the daypart defined by the reporting period.

An illustration of the computation of the comparative index of relative Statistical Efficiencies for the case just described is given in Table 6.

TABLE 6. ILLUSTRATION OF AN INDEX COMPUTATION FOR A TYPE J AQH ESTIMATE BASED ON MEN 35-44 IN THE METRO SURVEY AREA

	Statistical Efficiency
Type J—AQH Estimates	5.41
Cume estimates	.86

Index = $(5.41 \div 0.86) \times (100) = 629$ (appears in table 5a).

A complete set of the "actual" Statistical Efficiency values that were used to compute the indices in Tables 4 and 5 is given in Appendix C. The Statistical Efficiencies in Appendix C are average values derived empirically from the same nineteen Local Market Radio Report data base that was used in the Jackknife Replication analysis. These average values were *conservatively* derived via use of a Harmonic Mean.[12] A standard arithmetic mean of the individual Statistical Efficiency values (1,377,296 individual values in total) would yield considerably higher Statistical Efficiency averages for each cell entry in Appendix C. The Statistical Efficiency averages shown in Appendix C were calculated on a *conservative basis* for the sake of statistical rigor.

Before closing this section, it should be noted that Arbitron's new Radio Reliability Model is designed to properly reflect the differences in Statistical Efficiency between Cume vs. AQH audience estimates. The next Chapter describes how this is accomplished.

[12]Further technical details (concerning the derivation of the average Statistical Efficiency values in Appendix C) are given in Section D of Chapter 7.

Section III: Methodology

Chapter 5

Statistical Description of Arbitron's New Radio Reliability Model

As explained earlier, the primary objective of the *Replication II* study was to develop a procedure that would allow Arbitron report users to quickly and easily determine the reliability (standard error) of any audience rating or projection contained in Arbitron Local Market Radio Reports, and to accomplish this with *greater accuracy*, and *greater ease* and *greater speed* than was previously possible. This chapter gives a detailed statistical description of Arbitron's new Radio Reliability Model. In order to place the new model in proper perspective, a contrast is provided between the structure of the new model versus the present state-of-the-art including the Nomograph Procedure model.

A. Present State-Of-The-Art

Most techniques for providing standard error information to survey users rely on a two-component statistical model. One of the components is the formula for calculating the standard error under the assumption of simple random element sampling (i.e., the textbook *Simple Random Sample, [SRS]* mentioned previously). The second component, called the Design Factor (DF), is a correction term which attempts to adjust for the differences in the standard errors between the actual sample design and a Simple Random Sample design due to such factors as pre-stratification, post-stratification ("weighting" or "sample balancing"), sample clustering and random data errors.[13] More specifically, the two-component model can be expressed in the

[13]In the statistical sampling literature, the Design Factor (DF) often times appears in its squared form. In this case, it is usually called the Design Effect (DEFF). In this document, the Design Effect is designated by the abbreviated notation "DE." Thus, the Design Effect equals the square of the Design Factor—i.e., $DE = DF^2$.

form of equation 5-1:

$$\begin{array}{lclcll} \text{Standard error} & & \text{Standard error of} & & \text{Design} & \\ \text{associated with} & = & \text{simple random} & \times & \text{factor} & (5\text{-}1) \\ \text{actual design} & & \text{sample design} & & \text{adjustment} & \end{array}$$

For convenience in later reference, equation 5-1 can be re-written in more compact notation as shown below in equation 5-2.

$$\text{ste: actual} = \text{ste: srs} \times \text{DF} \qquad (5\text{-}2)$$

Most models in current use employ a single Design Factor value to estimate the size of the standard error for all estimates generated by the survey. For large scale personal-interview surveys, typically a number in the 1.2 to 1.5 range is used for the single Design Factor value. For most studies, the single Design Factor value is derived based upon a limited number of empirical standard error computations performed on the particular survey's data bank. In other instances, the Design Factor value is simply based upon a "knowledgeable guess" derived from general past experience with similar types of surveys.

B. The Nomograph Procedure Model

In an attempt to be more precise, sometimes several different Design Factor values are provided for use with different sample geographic components or demographic subclasses. The Arbitron Nomograph procedure falls into this category. In an attempt to advance the state-of-the art, it recognized that: (1) different Design Factors were required across the different Local Market Reports, and (2) within any single Local Market Report, several different Design Factors were required to account for the overall differences in the pre and post-stratification elements for the three geographic areas (Metro, TSA, ADI) shown in Arbitron reports.

The Design Factor structure of the Nomograph Procedure Model can be symbolically described as:

DF (mr, g)

The "mr" notation stands for *market report*, and the "g" stands for *geographic area within each market report* (Metro, TSA, and ADI). Thus, it can be seen that the Nomograph Procedure model was considerably more refined than the more traditional single Design Factor model. Furthermore, under the Nomograph Procedure, the Design

Factors were not determined simply from a generalized past experience, but instead, they were specifically calculated for each separate market and each individual survey report. For each report, three Design Factors were calculated—one for the Metro area, one for the TSA area, and one for the ADI area (for those reports providing ADI audience estimates). These Design Factors were calculated by using a well-known formula for evaluating the Design Factor associated with a weighted sample.[14]

C. Conceptual Structure of Arbitron's New Radio Reliability Model

On the basis of over one million empirical standard error calculations made via the Jackknife Replication analysis, it was quite clear that any simple model which used only a small number of different Design Factors would not adequately estimate the Standard Errors appropriate for all different estimate types, demographic groups and geographic areas that appear in Arbitron Local Market Radio Reports. In fact, based on the findings described in Chapter 4, it was quite evident that any improved model would have to include provision for the calculation of Design Factors separately for the more than 600 individual "cells" in the three-dimensional array defined by the various reporting categories for: (1) geographic area, (2) demographic group, and (3) estimate type, as shown earlier (see Appendix C).

The need for calculating separate Design Factors by cell follows directly from the fact that the sample Statistical Efficiencies cannot be assumed to be equal for each reporting cell, as was noted in Chapter 4. Hence, the Design Factors cannot be assumed to be equal for the different reporting cells because the Design Factor quantity is mathematically related to the Statistical Efficiency value. For any individual audience estimate, the relationship between the Statistical

[14]The actual formula is:

$$\sqrt{1 + RV\ (mr,\ g,\ p12+)}$$

where RV stands for Rel-Variance and equals the squared Coefficient of Variation of the individual diary weights among the Total Persons 12+ sample ("p12+") within a specific geographic ("g") region (Metro, TSA or ADI) for a specific market and survey report ("mr"). The Coefficient of Variation (CV) is defined by the following ratio:

$$CV = \frac{\text{Standard Deviation of Diary Weights}}{\text{Average (Mean) Diary Weight}} = \sqrt{RV}$$

Efficiency and the Design Factor quantities is very straightforward as shown in equation 5-3:

$$\text{Statistical Efficiency} = 1/(\text{Design Factor})^2 \qquad (5\text{-}3)$$

In other words, the *Statistical Efficiency* for any individual audience estimate is equal to the reciprocal of the square of the *Design Factor.*[15]

The number of different Design Factors that are necessary to produce appropriate standard errors for all audience estimates which appear in the Local Market Reports was determined by examining the sources of variation in the sizes of the actual standard errors. After taking into account both the in-tab sample size and the size of the audience (rating level), it was found that the size of the standard error varies with the following dimensions:

1. The specific *geographic region* (Metro, TSA, ADI) for which the estimate is reported.
2. The specific *demographic group* for which the estimate is reported (there are 32 different age/sex groupings in a standard Local Market Radio Report—See Table 1 in Chapter 4 for definitions).
3. The specific *estimate type:* Cume versus AQH; and for AQH estimates, the number of quarter hours included in the reported time period (see Table 2 in Chapter 4 for definition of the 11 different basic estimate types included in Arbitron Local Market Reports).

Taken together, this implies that for any standard market report, Design Factors might be defined for more than a thousand different cells (since 3 geographic regions by 32 demographic groups by 11 estimate types = 1056 cells). However, as indicated in Appendix C, not all estimate types are reported by all demographic groups in all geographic regions in Arbitron's Local Market Radio Reports. Hence, the number of cells for which audience ratings or projections are reported is 624 for the Standard Local Market Radio Report. (Arbitron's Condensed Local Market Radio Reports contain fewer cells— 144). Thus, for each Standard Report, Design Factors must be

[15]See Chapter 7, Section D for further commentary concerning the relationship between these two statistical quantities, and its pertinence to the derivation of the Statistical Efficiency values given in Appendix C.

calculated for 624 cells; only 144 are required for Arbitron's Condensed Reports.

Once the necessary number of different Design Factors was determined, the next step was to develop a method for calculating these factors for each Local Market and each survey period. Statistical sampling theory supported by the empirical analysis of "actual" standard errors calculated by the Jackknife Replication Procedure[16] led to the conclusion that each of the more than 600 Design Factors is influenced by some sources which are unique to each market report and other sources which can be viewed as common across all market reports.

The sources which can be viewed as unique to each report include:

1. The pre-stratification used for sample selection and the sample size allocations used among the different geographic strata.
2. The post-stratification ("weighting"/"sample balancing") used to compensate for non-proportionate sample allocation and differential sample recovery/cooperation rates among demographic and geographic subclasses.

The sources which can be viewed as common across all market reports include:

3. The clustering and correlation of listening behavior *among* respondents within the same households.
4. The impact on the standard error due to the collection of multiple observations on the same respondents and, the clustering and correlation of listening behavior *within* individual respondents across quarter-hour time units (NOTE: This type of intra-individual clustering of listening behavior impacts the standard error for AQH estimates, but not Cume estimates).

On this basis, the Design Factor for each geo-demo-estimate type cell[17] in a market report may be viewed as consisting of two elements—a "unique" component related to sources 1 and 2, and a "common" component related to sources 3 and 4. Thus, the Total

---

[16]A complete description of the Jackknife Replication procedure is given in Chapters 6 and 7.

[17]As explained previously, the *Standard* Local Market Radio Reports contain 624 reporting cells, and the *Condensed* Local Market Radio Reports contain 144 reporting cells.

Design Factor for any given cell can be evaluated by the following multiplicative model:

$$\text{Total DF (mr, g, d, e)} = \text{Unique DF (mr, g, d, e)} \times \text{Common DF (g, d, e)} \qquad (5\text{-}4)$$

where
mr = market survey report
g = geographic region
d = demographic group
e = estimate type
DF = Design Factor

Equation 5-4 highlights the fact that for each market and survey report, a separate Total Design Factor is calculated for each of the more than 600 individual cells defined by the various combinations of geographic, demographic and estimate type variables described previously.

In the new model, the Unique Design Factor as specified in equation 5-4 is calculated for each individual geographic by demographic by estimate type cell for each individual market and survey report period.[18] This calculation is made by using the well-known formula described earlier in Footnote 14 of this Chapter—namely, the square root of the quantity (1 + Rel-Variance of the diary weights). However, in Arbitron's new model, the Unique Design Factor is computed separately for each cell within a report;[19] therefore, the calculation can be expressed as shown in equation 5-5 below in order to highlight this fact:

$$\text{Unique DF (mr, g, d, e)} = \sqrt{1 + \text{RV (mr, g, d, e)}} \qquad (5\text{-}5)$$

The Common Design Factors for the more than 600 geographic by demographic by estimate type cells were determined from the *Arbitron Replication II* study. They are used as parameter values (constants) for all market reports. The Common Design Factors were

[18]The Unique Design Factor within a particular market report and a particular geographic by demographic cell has the same value for all eleven basic estimate types since all audience estimates for that cell are derived from the same set of diaries.

[19]Under the Nomograph procedure, this quantity was also computed separately for each report, but only for the Total Persons 12 + demographic group within each region (Metro, TSA and ADI). The Nomograph procedure assumes that this calculation is an adequate description of the Total Design Factor, and furthermore, that the Total Persons 12 + Design Factor can be used as a surrogate (in the corresponding region) for separate Design Factors calculated by demographic subgroup and varying estimate types. The new model is more sophisticated, and therefore, the new model provides more accurate estimates of the actual standard errors.

determined by using the actual Total Design Factors found by the Jackknife Replication analysis in conjunction with the Unique Design Factors calculated for the 19 market reports included in the study.[20] Specifically, for each cell, the Common Design Factor was calculated by taking the ratio of: (1) the actual average Total Design Factor determined empirically from the Jackknife Replication analysis of over one million standard errors, to (2) the average Unique Design Factor calculated across the 19 market reports included in the study. This calculation can be expressed algebraically as shown in Equation 5-6:

$$\text{Common DF (g, d, e)} = \frac{\text{Average jk. Total DF (g, d, e)}}{\text{Average Unique DF (g, d, e)}} \qquad (5\text{-}6)$$

The "g", "d" and "e" notation have the same meaning as defined earlier. They serve to highlight the fact that the Common Design Factor is separately computed for each reporting cell defined by the 3-dimensional array generated by the varying combinations of geographic, demographic, and estimate type groupings as illustrated by the structure of the Tables given in Appendix C. The averaging procedure used to calculate the numerator and denominator on the right-hand side of Equation 5-6 was a conservative one and is explained in Section D of Chapter 7. Putting that aside for the moment, the new Arbitron Radio Reliability Model can now be fully specified in the form of the following equation—*The New Arbitron Radio Reliability Model Equation:*

$$\begin{aligned} &\text{ste: actual (mr, g, d, e, i)} = \\ &\quad \text{ste: srs (mr, g, d, e, i)} \times \text{Total DF (mr, g, d, e)} \end{aligned} \qquad (5\text{-}7)$$

where The letters contained in the parenthesis highlight the fact that Equation 5-7 really represents a whole series of equations— in effect, a separate equation for each audience rating presented in Arbitron reports now and into the future. The "i" stands for any individual (but particular) audience rating contained in any particular Market Report ("mr")/for any particular Geographic Region ("g")/Demographic Group ("d")/ and Estimate Type ("e").

and "Total DF" equals the product of the Unique DF and the Common DF as specified earlier in Equation 5-4.

[20]See Chapter 7 for further discussion concerning the calculation details for these and other related statistics.

and "ste: srs" stands for the corresponding *Standard Error of a Simple Random Sample* with the same in-tab sample size as was used to derive the reported audience rating from the actual sample design. The standard error for a Simple Random Sample is defined by equation 5-8 below:

$$\begin{matrix}\text{The Standard Error}\\ \text{For A}\\ \text{Simple Random Sample}\end{matrix} = \sqrt{\frac{R\,(100\text{-}R)}{n}} \tag{5-8}$$

where
"R" equals the numerical value of the rating
and
"n" equals the size of the sample (in-tab) upon which the rating was based.

For notational consistency with Equation 5-7, the "ste:srs" formula can be expressed with the appropriate subscripts (mr, g, d, e, i) as Equation 5-9 which follows:

$$\sqrt{\frac{R(mr,\ g,\ d,\ e,\ i) \times [100 - R(mr,\ g,\ d,\ e,\ i)]}{n\ (mr,\ g,\ d)}} \tag{5-9}$$

NOTE: the "e" and "i" subscripts need not appear with the n value since the in-tab sample size is constant for all individual estimates and estimate types contained within a particular report for a particular demographic and geographic group.

Summary Analysis

A careful analysis of the structural form of Equation 5-7 clearly highlights the implications of the new Arbitron model as an improved tool for determining the size of the standard error for any audience rating. In contrast to the Nomograph Procedure and other less complex models, the new model properly accounts for differences in Design Factors (and the related differences in Statistical Efficiencies) by the important geographic, demographic and estimate type variables as defined by the more than 600 reporting cells included in Arbitron Local Market Radio Reports. As a result, *the new model provides a substantial improvement in reliability (standard error) estimation* because it eliminates the persistent biases (consistent under or over-statements of reliability) that are associated with the simpler-type models.

D. Algebraic Reformation of Arbitron's New Model To Enhance Its Usefulness to Report Users

The final step in the development of Arbitron's new Radio Reliability Model consisted of transforming a complex mathematical structure into a form that would be easily accessible to report users. This was accomplished by re-expressing the initial formulation of the new model as given by Equation 5-7 into the quotient of two terms that could be calculated by Arbitron and placed into two tables whereby report users could simply "look up" the appropriate values.

By algebraic reformulation, the conceptual model given in Equation 5-7 can be re-expressed in the form of Equation 5-10 below:

$$\text{ste: actual (mr, g, d, e, i)} = \frac{\text{Table A Value}}{\text{Table B Value}} \tag{5-10}$$

where

$$\text{Table A Value} = \sqrt{\text{R(mr, g, d, e, i)} \times [100 - \text{R (mr, g, d, e, i)}]}$$

and

$$\text{Table B Value} = \frac{\sqrt{\text{n(mr, g, d)}}}{\text{Total DF(mr, g, d, e)}}$$

The Table A values depend only on the rating size. Thus, a single table can be pre-printed to show the values of term A for all possible rating sizes. An actual Table A from a recent report is shown in Appendix A.

The Table B values depend upon the estimate type, the demographic group, and the geographic area. These values will differ for each report. Hence, a separate set of Table B values must be computed when processing each Market Report. Thus, a separate Table B will be provided for each Market Report. For illustrative purposes, Appendix A shows the Table B that appeared in the New York, Fall 1981 Local Market Radio Report.

While Arbitron's new Radio Reliability Model is quite complex, it has been algebraically reformulated to enhance its usefulness to report users. The standard errors for any published rating can be determined by looking up two values (one from Table A and one from Table B), and then performing a simple division (A ÷ B). The standard error of a projected audience number can easily be determined by first converting the audience projection number to a rating, and then converting the rating's standard error to the standard error for the audience projection number. Examples of those calculations are given in Chapter 8, as well as in Appendix A.

Chapter 6

Empirical Determination of Standard Errors By Replication Technique—Background Discussion

A. Introduction

Methods of determining standard errors for survey estimates by replicated subsamples were first used in the 1940's. It was not until 1960, however, that these methods gained widespread acceptance and use in the United States. This acceptance was due, for the most part, to the publication of a text *Sample Design for Business Research*, by Professor W. Edwards Deming of New York University.[21]

During the same time period, Professor John Tukey of Princeton University described a method (which he named the "Jackknife") for reducing statistical bias in non-linear estimates derived from relatively small samples. In a short abstract,[22] published in 1958, Professor Tukey conjectured that the Jackknife procedure might also find use in the estimation of standard errors.

The Jackknife method remained relatively obscure for some time, but in 1966, Dr. David Brillinger, one of Tukey's students, published a detailed explanation of a method of standard error determination in complex survey samples which synthesized the concepts of Replication and the Jackknife.[23] This method is called "Jackknife Replication."

---

[21]Deming, W. E., *Sample Design in Business Research.* New York: John Wiley and Sons, 1960.

[22]Tukey, J. W., "Bias and Confidence in Not-Quite Large Samples: Abstract." *Annals of Mathematical Statistics*, Vol. 29, (1958), p. 614.

[23]Brillinger, D. R., "The Application of the Jackknife to the Analysis of Sample Surveys." *Commentary*, (a Journal of the British Market Resarch Society) Vol. 8, (1966), p. 74–80.

B. Simple Replication

In order to understand why Jackknife Replication provides appropriate, empirically determined, standard errors for ratings and audience projections in Arbitron Radio Market Reports it is necessary to first understand the method of Simple Replication (i.e., replicated subsamples).

In Simple Replication the sample is either selected, or viewed as if it had been selected, in g independent replications or repetitions of the basic sample design (where g is the number of replicates). Each of these g "Basic Replicates" is projected (weighted) separately to the applicable universe. Letting R_i denote the estimate of a specific rating obtained from the i^{th} Basic Replicate, the Simple Replication standard error of R (the estimate based on the total sample) is defined as:

$$\text{sterep}(R) = \sqrt{(1-f)\,\frac{1}{g\,(g-1)}\,\Sigma\,(R_i - \bar{R})^2} \qquad (6\text{-}1)$$

where $\bar{R} = \frac{1}{g}\,\Sigma\,R_i$ (6-2)

g = the number of Basic Replicates

f = the overall sampling fraction (the size of the sample relative to the population) across all g Basic Replicates

Σ = the sum across all g Basic Replicates

In those instances where the value of f is small (e.g., below 0.01) the factor (1 – f) is often eliminated, and the formula may be written as:

$$\text{sterep}(R) = \sqrt{\frac{1}{g\,(g-1)}\,\Sigma\,(R_i - \bar{R})^2} \qquad (6\text{-}3)$$

This convention will be followed through the remainder of the chapter.

If the sample selection and estimation process do not involve complex clustering or complex weighting procedures, the estimate R (which is derived from all the Basic Replicates combined—i.e., the total report sample) will be very close in value to the estimate $\bar{R}$ (the value obtained by first separately weighting each Basic Replicate and then producing the estimates R_i, and finally taking their arithmetic mean). In this case, the Simple Replication standard error, defined by sterep(R) in Formula 6-1 or 6-3, will be an appropriate estimate of the standard error of R.

C. Jackknife Replication

When differences between the values R and $\bar{R}$ are not insignificant, a more appropriate standard error for R may be found by the method of Jackknife Replication. This can occur if the sample design involves complex clustering and/or weighting. Under these circumstances, the sample estimate may not be *linear* at the elementary selection-unit level, or at the elementary analytical-unit level. In this case, the resulting value of $\bar{R}$ may be quite different from the value of R.

This type of non-linearity is present in the samples used for Arbitron Radio Reports. Arbitron makes use of a weighting procedure known as "sample balancing." The basic statistical property of sample balancing is its ability to assure that the weighted sample will conform to a pre-specified set of distributions on various characteristics, while minimizing the variation in weights that must be applied among survey respondents. These weights are determined by an iterative procedure which is not a simple linear function. As a result, standard errors produced by the Jackknife Replication formulas described below will produce more appropriate estimates of the actual standard errors of Arbitron audience ratings and projections.

The *first step* in the method of Jackknife Replication is exactly the same as the first step used in the method of Simple Replication. The basic sample is either selected as g independent Basic Replicates, or is subdivided into g independent Basic Replicates on a post-hoc (after selection) basis. These g Basic Replicates are then formed into g "Jackknife Replicates" by defining the i^{th} Jackknife Replicate as: *all Basic Replicates combined with the exception of the i^{th} Basic Replicate.* In other words, the i^{th} Jackknife Replicate is simply the complement (with respect to the total sample) of the i^{th} Basic Replicate. This process is illustrated in Figure 5 for the case where *eight* Jackknife Replicates are formed from *eight* Basic Replicates to depict the procedure used in the *Replication II* study. For example, from Figure 5 it can be seen that Jackknife Replicate #3 is formed by adding together Basic Replicates 1, 2, 4, 5, 6, 7, 8 and excluding Basic Replicate #3.

In the *second step*, each Jackknife Replicate is projected (weighted) to conform to the total population distribution. If the estimate of an audience rating (or projection) based on the i^{th} Jackknife Replicate is denoted as R'_i and the mean of the g Jackknife Replicate R'_i values is denoted as $\bar{R}'$ then the traditional Jackknife Replication standard

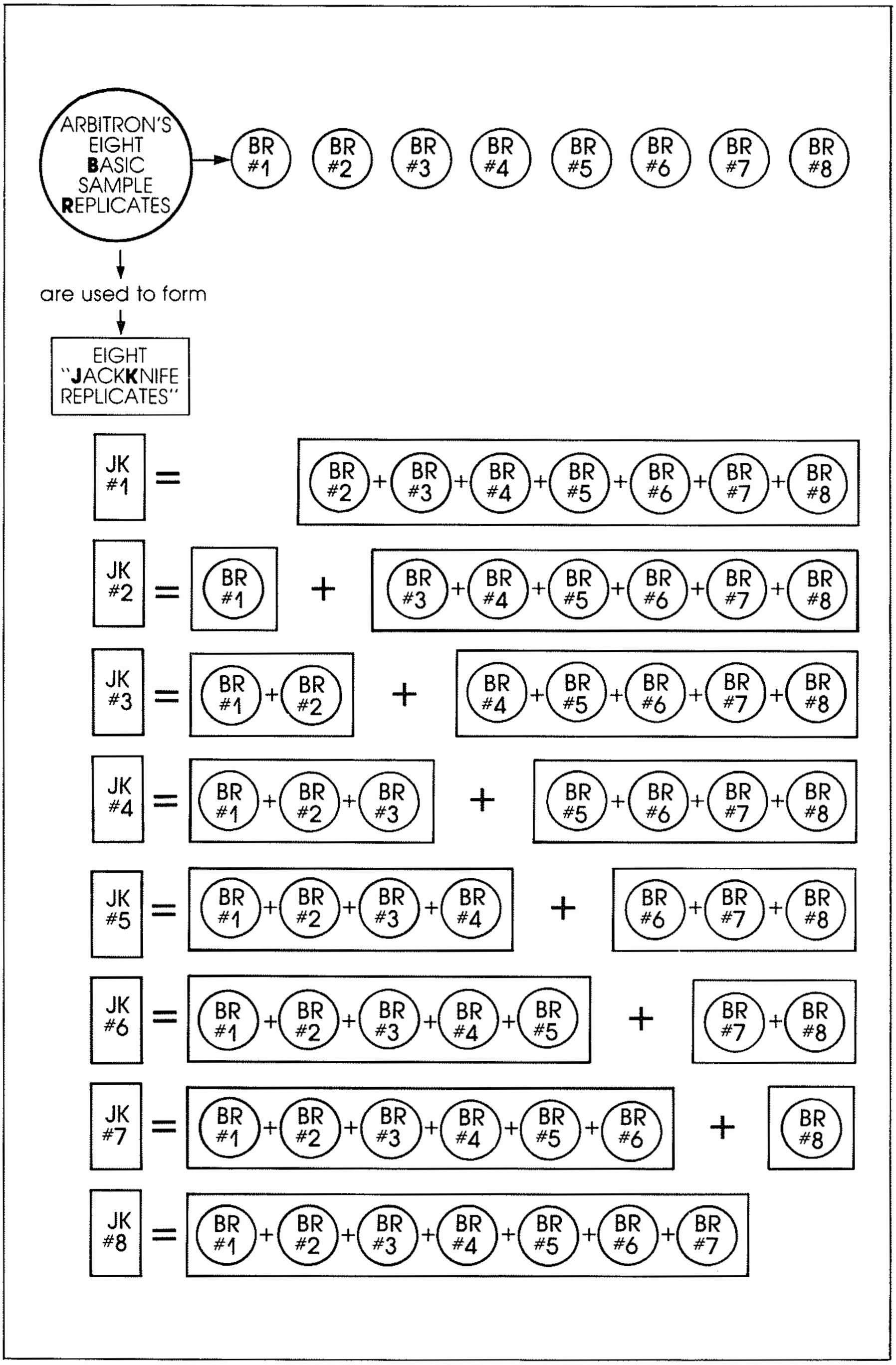


Figure 5. How eight "Jackknife Replicate" samples are constructed from Arbitron's eight "Basic Replicate" samples.

error of R is defined by Formula 6-4 below:

$$*\text{stej}(R) = \sqrt{\frac{g-1}{g} \sum (R'_i - \bar{R}')^2} \tag{6-4}$$

where $$\bar{R}' = \frac{1}{g} \sum R'_i \tag{6-5}$$

Σ = Sum across all g Jackknife Replicates

*Note: The stej(R) quantity is expressed in Appendix D using the notation: stej(R_q).

Formula 6-4 is quite similar to Formula 6-3, except that R'_i and R' replace R_i and $\bar{R}$, and the factor $1/g(g-1)$ has been replaced by the factor $(g-1)/g$. The mathematical justification of this difference in factors for Simple and Jackknife Replication is discussed in Appendix D. It should be noted that when there are only two replicates (i.e., $g=2$), Formulas 6-3 and 6-4 become identical. For values of g greater than 2, this identity does not hold.

The traditional Jackknife Replication estimate of the standard error of R (as given by Formula 6-4) provides a more appropriate estimate of the standard error R, than does the Simple Replication procedure (defined by Formula 6-3), whenever $\bar{R}'$ is closer to R than is $\bar{R}$. For Arbitron's audience data $\bar{R}'$ would be expected to be closer to R, than $\bar{R}$, for the reason explained in what follows. The *Arbitron Replication II* study used Jackknife Replication with $g=8$. Thus, each R'_i was formed by application of the sample balancing process to seven-eighths (7/8) of the total sample. The functional form of the sample balancing weights using these Jackknife Replicates is very close to the functional form of the sample balancing weights that were used for the original report total sample. This would not have been the case, if each replicate had been only one-eighth (1/8) of the total sample as when a Simple Replication approach is used (with $g=8$). As a result, values of $\bar{R}'$ (from Jackknife Replication samples) will more closely conform with the values of R (from total sample), than would values of $\bar{R}$ (from Simple Replication samples).

In recognition of the possibility that $\bar{R}'$ may differ from R by a very small amount, the standard error computations in the *Replication II* study were actually accomplished by using Formula 6-6 given below:

$$\text{stejk}(R) = \sqrt{\frac{g-1}{g} \sum (R'_i - R)^2} \tag{6-6}$$

Formula 6-6 differs from Formula 6-4 in that the value of R is used in place of $\bar{R}'$. *Conservative*[24] estimates of standard errors are produced by this change because the numerical values produced by Formula 6-6 will always be equal to, or greater than, the values produced by Formula 6-4. Differences between Formula 6-6 and Formula 6-4 are related to the size (if non-zero) of the difference between $\bar{R}'$ and R. Use of Formula 6-6, rather than Formula 6-4, has become the preferred practice in survey sampling applications whenever conservative estimates are desired.[25]

D. Simple Replication versus Jackknife Replication: Summary of Issues

1. Simple (Non-Complex) Surveys

For surveys which involve simple sample designs and simple estimation procedures, *Simple Replication* is usually employed to estimate standard errors for the following reasons:

- Simple Replication is easier to implement than Jackknife Replication.
- Compared to Jackknife Replication, Simple Replication is easier to explain to non-statisticians.
- Simple Replication can provide appropriate estimates of the standard error of R, when the overall survey estimate R can be expected to generally exhibit the same statistical behavior as $\bar{R}$ (the arithmetic mean of the Basic Replicate estimates R_i).

2. Complex Surveys

When survey designs involve complex sampling procedures (clustering and pre-stratification) and complex estimation procedures (sample balancing), Simple Replication may somewhat overstate (or understate) the standard error of R (the overall survey estimate). This is because the sample design and sample balancing complexities result in a lack of equality between R, (the overall survey estimate), and $\bar{R}$

---

[24]Arbitron used the Formula 6-6 *conservative* approach so that if any differences occurred, Arbitron would slightly overstate the size of the actual standard errors rather than understate them.

[25]Wolter, K. M., *Introduction to Variance Estimation.* Unpublished Manuscript. 1980.

(the mean of the Basic Replicate estimates). This lack of equality may be somewhat reduced by using a smaller g (the number of Basic Replicates), but this has two disadvantages:

- A lowering of g will lower the stability of the standard error estimate.
- A lowering of g to its theoretical minimum of "two" may still *not* adequately reduce the discrepancy between $\bar{R}$ and R.

Jackknife Replication solves the problem because Jackknife Replication does not require a lowering of g to achieve near equality between the overall survey estimate, R, and the arithmetic mean of the Jackknife Replicate estimates, $\bar{R}'$. In fact, as g is made large, the Jackknife Replication estimate of the standard error of R improves and has two advantages:

- Increasing g results in an increase in the stability of the Jackknife Replication estimate of the standard error of R, and
- Increasing g tends to decrease any difference between $\bar{R}'$, (the mean of the Jackknife Replication estimates), and R (the overall survey estimate), thereby reducing the degree of any possible mathematical bias in the standard error estimator. [Note: This situation is reversed for Simple Replication].

* * * * * * * * * *

As can be seen from the above summary, the Jackknife Replication approach to standard error determination offers substantial benefits in the case of *complex* surveys. In the context defined above, Arbitron Radio Audience Surveys can be classified as *complex*. Hence, for the beneficial reasons already outlined, the Jackknife Replication approach was chosen as the preferred methodology for the *Arbitron Replication II* study.

Chapter 7

Sample Selection and the Formation of Jackknife Replicates in the Arbitron Replication II Study

A. Selection of Markets and Survey Reports

During the initial planning stages of the *Arbitron Replication II* study design, it was decided that the actual market reports which would provide the basic standard error information for use in the development of the new Arbitron Radio Reliability Model should be selected so as to produce conservative model parameters. When used in a statistical context, the term conservative indicates that any necessary approximations and decisions should be structured to produce an *under*statement (rather than an *over*statement) of actual reliability levels. In terms of standard errors, this means that to the extent that the model construction involved approximations, *over*statement of standard error levels was viewed as being preferable to *under*statement of these levels.

A related consideration in the selection of market reports was the knowledge that the greatest stresses on reliability would occur at the extremes of the market size distributions—largest markets vs. smallest markets. (This phenomenon was detected during an earlier reliability study which focused on the development of standard error estimates for Arbitron's Television Market Reports). This stress phenomenon is quite common in most modeling tasks. For this reason, it was decided that market selection should not involve a simple random (or proportionate) selection of markets, but rather it should ensure that a sufficient number of markets are included at the extremes (as well as the center) of the market size range.

A similar philosophy was felt appropriate with respect to other important variables such as: the in-tab sample size at both the Metro and TSA level, the use (or non-use) of special ethnic sampling and weighting procedures, the regional distribution of markets, the number of reported stations in a report, and finally the survey period itself. In order to provide a data base which would be suitable for special studies, it was also decided that actual standard error information should be available for several markets over different survey periods. This was accomplished by including several survey report periods for the same market in the final selection of a sample of market reports.

All of the above conditions were met by selecting a purposive sample of *19* Local Market Radio Reports, involving *12* different markets. Since the Jackknife Replication formula (see Chapter 6) that was used to calculate standard errors required 8 Jackknife Replicate estimates *plus* the audience estimates published in the original report, the 19 report selections produced a working data base for this study that included a total of 171 (19×9) audience reports. These reports contained a total of 12,395,664 audience ratings and projection numbers that were used for analysis purposes to produce 1,377,296 Jackknife Replication standard error estimates. The distribution of the original 19 selected reports with respect to the dimensions discussed above appears in Table 7.

B. Selection of Basic Replicates

In order to form the eight Jackknife Replicates for each selected market report, the original total report in-tab sample for each of the nineteen reports was first divided into eight non-overlapping subsamples called "Basic Replicates". In effect, the initial total report sample was systematically subdivided into eight parts (Basic Replicates) of approximately equal size. Hence, each Basic Replicate is about one-eighth the size of the original report sample. In total, 152 Basic Replicates were selected (8 per each of 19 separate market reports). Their selection was accomplished via the use of appropriate mathematical randomization procedures.

In forming the Basic Replicates, appropriate care was exercised during the selection process to duplicate all of the pre-stratification and sample-clustering design features that were used in the selection of the initial total report sample. The Basic Replicates were formed by first re-constructing the original total sample selection structure. Since Arbitron's standard sampling methodology is described in several other Arbitron publications,[26] it need be only briefly outlined here.

[26]Arbitron's Sampling Methodology is described in Arbitron Radio's Description of Methodology document and also in the "Description of Methodology" section (page i) of every Local Market Radio Report.

TABLE 7. CHARACTERISTICS OF SELECTED MARKETS AND SURVEY REPORTS

1. Number of Markets	12
2. Number of Survey Reports	19
3. Number of Survey Periods	4
4. Number of Markets with Special Ethnic Samples	7
5. Number of Markets with Ethnic sample balancing	8
6. Regional distribution of Markets:	
Northeast	2
Southeast	3
Southwest	1
North Central	3
West Pacific	1
Northwest	2

7. Distribution of Survey Reports

by Number of Reporting Stations

Home to Metro		Total Stations	
40 or more	4	40 or more	4
25–39	4	35–39	5
20–24	3	30–35	3
11–19	5	25–29	5
10 or less	3	24 or less	2

by In-Tab Sample Sizes

Total Survey Area		Metro Survey Area	
6000 or more	3	3000 or more	4
3500–5999	2	1500–2999	4
2500–3499	5	1000–1499	3
1000–2499	4	500– 999	3
Under 1000	5	Under 500	5

The standard sampling methodology used by Arbitron for its Local Market Radio Reports involves allocation units known as sampling areas. These sampling areas are defined within the Metro, TSA and ADI portions of the market on a geographic basis, usually at the county level. In those instances where a county contains more than one sampling area (e.g., ethnic and non-ethnic locations), the sampling areas are defined on the basis of zip codes. Sampling within a market will always take place from a frame of listed telephone numbers which have been numerically sorted within each sampling area by zip code.[27] Sampling from this listed frame is accomplished separately within each sampling area through the use of Systematic Sampling (every nth selection) beginning from a random starting point.

Sample allocation and diary placement rates to each sampling area are determined by several factors including the sampling area's population size and Arbitron's experience in getting back complete, usable diaries. These latter considerations take into account Arbitron's past experiences in terms of respondent response and diary usability rates, as well as the average household size for each sampling area. In markets where Arbitron uses an Expanded Sample Frame (ESF) to incorporate unlisted telephone homes into the sample, telephone numbers and sample allocations are generated for the entire ESF area as a whole. After the basic selections are completed for both the listed and unlisted (where applicable) portions of the sampling frame, those selections are then systematically (every nth) assigned to the weekly time periods that span the particular survey report.

The stratification elements in Arbitron's standard sampling methodology as just described were preserved when forming the eight Basic Replicates for this study. Specifically, the in-tab sample cases were sort ordered by broad geographic area (Metro, TSA balance, ADI balance), then by sampling area, then by time period, then by original selection order within time period. All in-tab cases that were associated with an initial selection unit were grouped together. That is, they were linked to their appropriate *ultimate sampling unit*. For the most part, the formation of ultimate sampling units involved grouping together diaries for all persons residing within the same households. In those instances where data collection was accom-

---

[27]The actual sampling frame is created by the Metro Mail Corporation using the specifications designed by Arbitron. The Metro Mail Corporation is one of the nation's leading suppliers of listed telephone households.

plished by personal interviewing[28], all in-tab cases associated with the interviewing cluster were grouped together in a simple ultimate sampling unit to reflect the method used in initial sample selection.

Eight Basic Replicates were then formed by the systematic assignment of ultimate sampling units. For example, the first Basic Replicate consisted of all in-tab cases associated with ultimate sampling units 1, 9, 17, 25, etc. The second Basic Replicate consisted of all in-tab cases associated with ultimate sampling units 2, 10, 18, 26, etc.

C. Formation of Jackknife Replicates

The eight Jackknife Replicates were then formed from the eight Basic Replicates by the process described in Chapter 6 (and illustrated in Figure 5). For example, the first Jackknife Replicate consisted of all in-tab cases associated with Basic Replicates 2, 3, 4, 5, 6, 7, and 8; while the second Jackknife Replicate consisted of all in-tab cases associated with Basic Replicates 1, 3, 4, 5, 6, 7, and 8.

The final step in preparation for the standard error determinations was the sample balancing (weighting) of each of the Jackknife Replicates. Each of the 19 market reports had already been processed and released prior to the start of processing for the *Replication II* study. As a result, the specific procedures and parameters used in the sample balancing process had already been determined. These procedures and parameters were applied in exactly the same way to each of the eight Jackknife Replicates associated with the market report. Final output for each of the 8 Jackknife Replicate reports was prepared in the form of computer tapes. These files were combined with already existing files which had been produced in conjunction with the release of the original market report.

D. Standard Error and Other Related Computations

After all the required Jackknife Replicate reports were produced, "actual" standard errors were empirically computed by using the Jackknife Replication method specified by Formula 6-6 in Chapter 6. Standard errors were individually calculated for all published audi-

---

[28]See references given in Footnote 26 for description of Arbitron's special interviewing procedures in High Density Hispanic Areas.

ence estimates contained in the 19 market report books used for this study.

In order to retain a uniformity in the structure of the standard error computations, all standard errors were calculated in terms of ratings (instead of projections). In other words, in those instances where a published report provides audience projection numbers (but not ratings), the projection numbers were first converted to a rating; then, the standard error was calculated for the corresponding rating. This procedure does not result in any loss of information from either a statistical or numerical standpoint, because audience ratings and projections are related in an exact one-to-one fashion by a uniform constant (the population, or universe value for the given geographic and demographic group).

Across the 19 market reports used in this study, a total of 1,377,296 individual Jackknife standard errors were calculated by use of the Formula 6-6 defined in Chapter 6. The 1,377,296 standard errors were generated from a total of 12,395,664 ratings estimates.

> [**Note:** When the numerical value of the rating published in a report is compared to the corresponding rating derived for each of the eight Jackknife Replicates, the *eight* numerical comparisons are summarized into *one* standard error quantity. This explains why the "number" of calculated standard errors (slightly over *one* million) differs from the "number" of audience ratings (slightly over *twelve* million) used in the analysis.]

For the purpose of model development and parameterization and for the purpose of performing various comparative analyses such as those described in Chapter 4, standard errors were also calculated by the Nomograph Procedure and by the Simple Random Sample formula. In addition, the following Design Effect (DE) ratios were calculated for each audience rating:

$$DE_{jk}(R) = \frac{[\text{stejk}(R)]^2}{[\text{ste: srs}(R)]^2} \tag{7-1}$$

$$DE_{nomo}(R) = \frac{[\text{ste: nomo}(R)]^2}{[\text{ste: srs}(R)]^2} \tag{7-2}$$

where stejk(R) = the "actual" standard error as calculated by *Jackknife Replication* ste: srs (R) = the corresponding *Simple Randon Sample* standard error

ste: nomo (R) = the corresponding standard error as estimated by the *Nomograph* Procedure

and R stands for the particular *rating* for which the standard error is being determined.

It can be seen from the preceding formulae that the Design Effect statistic is defined as the ratio of two Variances[29] with the Variance in the denominator of the ratio being that for a Simple Random Sample.

For the purpose of developing parameters for the new Radio Reliability Model, Unique Design Effects (Unique DE) were computed within each of the 19 market reports for the total sample and each geo-demographic group by the following formula:

$$\text{Unique DE}_j = 1 + RV_j \tag{7-3}$$

where

$$RV_j = \frac{VAR_j}{\overline{W}_j^2} \tag{7-4}$$

and

$$VAR_j = \frac{\Sigma(W_{ji} - \overline{W}_j)^2}{n_j - 1}$$

$$\overline{W} = \frac{1}{n_j} \Sigma\, W_{ji}$$

W_{ji} = the diary projection weight associated with the i^{th} in-tab respondent within the j^{th} geo-demographic subgroup.

n_j = the number of cases (diaries) in the j^{th} geo-demographic subgroup.

The Design Effect ratios defined by Equations 7-1, 7-2 and 7-3 were aggregated (summed and averaged) separately by each of the 19 market reports and by each of the more than 600 "cells" defined in Appendix C in terms of 11 estimate types, 32 demographic groups and 3 geographic areas (2 geographic areas for those market reports which do not contain ADI audience information). The statistical properties which dictate the use of the Design Effect statistic in initial aggrega-

[29]The "Variance" expression is, of course, the standard nomenclature for the *square of the Standard Error.*

tion, rather the Design Factor form,[30] are linked to the well-known behavior of the F distribution.[31] The average (mean) Design Effect produced by this initial aggregation by "cell" within each market report can be denoted by the symbol $\overline{DE}$; then, the manner in which the average Design Factor was computed for each cell within each market report can be defined as follows:

$$\text{Average Design Factor} = \sqrt{\overline{DE}} \tag{7-5}$$

The procedure in Equation 7-5 produces a *conservative* estimate for the average Design Factor since in general, a straight arithmetic mean of the individual Design Factors would yield a lower value.[32] The results obtained for each cell within each market report were then averaged across all 19 market reports separately by cell to produce the final average Design Factor values for each cell as defined in Appendix C. The averaging across markets was accomplished by taking a weighted mean of the 19 market report Design Factors for each corresponding cell. The weights were determined by the number of cases contained in each market report for any given cell.

The overall result of this process was used in three ways:

1. To determine the Common Design Factor parameter values for the new Reliability Model.
2. To calculate the Statistical Efficiency values presented in Appendix C.
3. To form the Comparison Ratios of "actual" versus "Nomograph" standard errors given in Appendix E.

These three uses are discussed in the following three subsections.

Determination of the Common Design Factors

Equation 7-6 below was used to determine the Common Design Factors for each of the more than 600 geo-demo-estimate type cells for

---

[30]The general relationship between these two forms for any individual rating is:
$(\text{Design Factor})^2 = \text{Design Effect}$

[31]Kendall, M. G., and Stuart, A. *The Advanced Theory of Statistics* (1958), London: Griffin and Co.

[32]A lower value for the Design Factor implies a smaller standard error. Hence, the larger value obtained by use of Equation 7-5 is more conservative in that, if any differences occur, Arbitron would overstate (rather than understate) the size of the actual standard errors.

which Arbitron publishes audience estimates in its Local Market Radio Reports:

$$(\text{Common Design Factor})_j = \frac{(\text{Average Jackknife Replication Design Factor})_j}{(\text{Average Unique Design Factor})_j} \quad (7\text{-}6)$$

where "j" stands for any particular cell (one of the more than 600 geo-demo-estimate type cells shown in Appendix C)

Calculation of Statistical Efficiency Values

For any individual rating, the general relationship between the Statistical Efficiency value and the Design Factor value is given by Equation 7-7 below:

$$\text{Statistical Efficiency}_i = 1/(\text{Design Factor}_i)^2 \quad (7\text{-}7)$$

where i = an individual rating

In order to maintain this straight-forward relationship in the averaging process, the average Statistical Efficiency values given in Appendix C were derived by taking the recriprocal of the average Jackknife Replication Design Factor. This was done separately for each geo-demo-estimate type cell. The net result of this procedure is that it yields *conservative* (lower) estimates for the average Statistical Efficiency values for each cell, as compared to a standard Arithmetic Mean of the Statistical Efficiency values of all of the individual cases contained within any particular cell.

The *conservativeness* of the average Statistical Efficiency values given in Appendix C derives from the mathematical relationship beween the Arithmetic and Harmonic Means as explained in the footnote below.[33] With respect to Statistical Efficiency, the process of

[33]For any set of numbers $X_1, X_2, \ldots X_m$,

$$\text{Harmonic Mean} = \frac{m}{\sum_{i=1}^{m} \left(\frac{1}{X_i}\right)}$$

$$\text{Arithmetic Mean} = \sum_{i=1}^{m} (X_i)/m$$

In general, when the X_i are not all equal, the Harmonic Mean will always yield a smaller value than the Arithmetic Mean. See, for example, Croxton and Cowden's book entitled *Applied General Statistics*, McGraw-Hill (1955).

averaging individual Design Effects within a cell in a Local Market Report corresponds to obtaining a Harmonic Mean for the Statistical Efficiency for that market cell. By the well-known relationship between Arithmetic and Harmonic Means, this produces a more conservative (lower) average Statistical Efficiency value than would a straight arithmetic average of the individual Statistical Efficiencies. Furthermore, the process of taking a weighted average of Design Factors across the nineteen Local Market Reports for each cell corresponds to obtaining a weighted Harmonic Mean for the square root of the Statistical Efficiency values. This second averaging also produces a more conservative (lower)value for the average Statistical Efficiency than would a weighted arithmetic mean.

Formation of Comparison Ratios of "Actual" versus "Nomograph" Standard Errors

For any individual rating R, the Comparison Ratio is defined to be:

$$\text{Comparison Ratio (R)} = \frac{[\text{ste: nomo (R)}]}{[\text{stejk (R)}]} \tag{7-8}$$

From equations 7-1 and 7-2, it can be seen that:

$$\text{ste: nomo (R)} = \sqrt{\text{DE}_{\text{nomo}}(\text{R})} \times [\text{ste:srs (R)}]$$

and

$$\text{stejk (R)} = \sqrt{\text{DE}_{\text{jk}}(\text{R})} \times [\text{ste:srs (R)}]$$

Thus, the Comparison Ratio can be conveniently re-expressed as:

$$\text{Comparison Ratio (R)} = \frac{\sqrt{\text{DE}_{\text{nomo}}(\text{R})}}{\sqrt{\text{DE}_{\text{jk}}(\text{R})}} \tag{7-9}$$

$$= \frac{\text{DF}_{\text{nomo}}(\text{R})}{\text{DF}_{\text{jk}}(\text{R})}$$

In order to preserve this simple relationship between the Comparison Ratio and the two Design Factors, each average Comparison Ratio displayed in Appendix E is calculated by taking the ratio of the average Nomograph Design Factor to the average Jackknife Replication Design Factor for that particular cell as shown below:

$$\text{Comparison Ratio}_{\text{j}} = \frac{(\text{Average Nomograph Design Factor}_{\text{j}})}{(\text{Average Jackknife Replication Design Factor}_{\text{j}})} \tag{7-10}$$

where "j" stands for any one cell

The average Comparison Ratio obtained in this manner is statistically more *robust* than a straight arithmetic mean of the Comparison Ratios for the ratings contained in that particular cell. Robustness in this context means that the averge Comparison Ratio should not be unduly affected by the presence of a few inordinately large individual Ratios (known as "outliers").

Section IV: Applications

Chapter 8

How to Use Reliability Calculations As An Aid To Decision-Making

Audience rating and projection estimates are used by the broadcasting and advertising industry to make important buying and selling decisions related to the radio medium. They are also used as criteria for measuring the success of specific programming efforts. Therefore, accurate information about the reliability of radio audience estimates should be viewed as an important decision-making aid to users of this information. For example, when an increase or decrease in a station's audience is observed from one report to the next, it is important to know the likelihood of whether the noted increase or decrease is "real", or just due to the "statistical bounce" commonly known as sampling error.

In the earlier chapters, it was explained that the size of this statistical bounce can be measured in a formal manner by a mathematical quantity called the *standard error*. It was also explained how Arbitron developed a new Radio Reliability Model to make it easy for Arbitron report users to quickly determine the size of the standard error, as well as statistical confidence boundaries for any single audience rating or audience projection number published in Arbitron Local Market Radio Reports.[34] The purpose of this chapter is to explain in a standardized "How To Do It" format the earlier cited applications, *plus* new applications that were not previously discussed.

The first two sections of this chapter explain the following specific areas of application:

A. *Evaluating Book-to-Book Changes in Reported Audiences.* That is, determining the likelihood (probability) that a reported in-

[34] An example of this calculation is given in Chapter 4, Section A.2, and also in Appendix A.

crease or decrease in a station's audience from one survey period to the next is "real", and not just due to "statistical bounce" in the survey audience estimates:

1. For estimates expressed as a *Rating* (p. 76).

2. For estimates expressed as a *Projection Number* (p. 79).

B. *Determining Statistical Confidence Limits for Reported Audience Estimates.* That is, determining upper and lower boundaries of the possible error range that may be associated with reported audience estimates, due to the fact that Arbitron surveys a sample of respondents. The application examples show how to determine statistical confidence limits:

- For any *single* audience estimate reported as a:
  1. *Rating* (p. 83).
  2. *Projection Number* (p. 85).

- For an *aggregated audience estimate* averaged across several Reports (Survey Periods) and expressed as a:
  3. *Rating* (p. 86).
  4. *Projection Number* (p. 89).

Section C contains some commentary concerning more complex applications. The final section of this chapter, Section D, contains a general technical discussion regarding the differences between "Classical" and "Bayesian" statistical inference interpretations. This discussion is relevant since the statistical inference interpretations adopted in the application examples in this chapter are Bayesian.

A. Evaluating Book-to-Book Changes In Reported Audiences

1. Changes In A *Rating* From Book-To-Book

a. Illustrative Application Situation—Station KBBB in Market X made a number of changes in its afternoon drive time programming with the expectation of increasing its Metro area Average Quarter Hour (AQH) audience among Males 18+. In the survey prior to the programming change, KBBB's reported audience rating was 0.4. In the new survey, after the programming changes had been in effect for a time, a rating of 0.7 was reported. How certain can the station be that this reported *in-*

crease in ratings was not just the result of statistical sampling error?[35]

b. Computational Details[36]

Step 1. Assemble the Required Basic Data

Data Item	Data Value	Data Source
Rating #1 (Last Time)	0.4	Arbitron Report #1: Mkt X
Rating #2 (This Time)	0.7	Arbitron Report #2: Mkt X
Standard Error of Rating #1	0.079	Arbitron Report # 1: Mkt X (Table A ÷ Table B)[37]
Standard Error of Rating #2	0.104	Arbitron Report #2: Mkt X (Table A ÷ Table B)[37]

Step 2. Determine The Probability (and Odds) That A Real Change Occurred In The Direction Indicated By The Rating Difference Between the Two Reports By Performing Three Simple Calculations Plus A Table "Look Up."

1. Compute the Rating Difference:

Rating Difference = (Rating #2) − (Rating #1)
Rating Difference = (0.7) − (0.4)
Rating Difference = +0.3

[35]The analysis in the *Computational Details* can be applied to any two reports; the reports need *not* be immediately successive in terms of Arbitron survey period chronology.

[36]For persons who have access to a micro computer, an interactive program (written in Microsoft BASIC-80) is provided in Appendix G to perform the calculations described. This program can be easily adapted to other versions of BASIC.

[37]See Chapter 4, Section A.2, for a general explanation of Table A and Table B. Market X Tables A and B are *not* shown here since the example is just hypothetical. Also, for pre-Fall 1981 Survey Reports which do not include a Table A or Table B, see Appendix F to estimate the standard error of an audience rating.

2. Compute the Standard Error of the Difference (Std. Error of Diff.):

$$\text{Std. Error of Diff.} = \sqrt{\left(\begin{matrix}\text{Std. Error}\\ \text{of Rating \#2}\end{matrix}\right)^2 + \left(\begin{matrix}\text{Std. Error}\\ \text{of Rating \#1}\end{matrix}\right)^2}$$

$$\text{Std. Error of Diff.} = \sqrt{(0.104)^2 + (0.079)^2}$$

$$\text{Std. Error of Diff.} = \sqrt{(0.010816) + (0.006241)}$$

$$\text{Std. Error of Diff.} = \sqrt{0.017056}$$

$$\text{Std. Error of Diff.} = 0.1306$$

3. Compute the Standardized Difference (Stdz. Diff.):

$$\text{Stdz. Diff.} = \frac{(\text{Rating Difference})}{(\text{Std. Error of Diff.})}$$

$$\text{Stdz. Diff.} = \frac{(+0.3)}{(0.1306)}$$

$$\text{Stdz. Diff.} = +2.29709$$

$$\text{Stdz. Diff.} = +2.3 \text{ (rounded to 1 decimal place)}$$

4. Use the Standardized Difference value (+2.3) as the reference point in Table 8 to obtain Probability and Odds that a real increase in ratings has occurred for station KBBB.

c. **Interpretation**[38]—By using the +2.3 Standardized Difference value to reference Table 8, the probability that a real *increase* occurred in the ratings for station KBBB is determined to be .989. This probability level is equivalent to 92.2 to 1.0 odds that a real change occurred in the *upward direction*. Certainly, in this case, KBBB can be quite sure that the programming changes made for the afternoon drive time did, in fact, *increase* their actual average audience among Males 18+ in the Metro area.

IMPORTANT: By definition, probability values must sum to 1.000. Therefore, there is also a probability that the noted upward increase in the report rating is *not* indicative of an actual increase, but rather it is the result of random statistical error. This latter probability is determined simply by subtracting the

[38]Readers who are familiar with the methods of statistical inference will recognize that the interpretation of the probability statements associated with the methods illustrated here are "Bayesian" rather than "Classical." A further discussion concerning the technical aspects of these two approaches is given in Section D of this chapter.

TABLE 8. PROBABILITIES AND ODDS THAT A REAL CHANGE HAS OCCURRED IN THE DIRECTION INDICATED BY THE DIFFERENCE IN RATINGS OR PROJECTION NUMBERS OBSERVED BETWEEN TWO REPORTS.

Stdz. Diff.	Probability	Odds	Stdz. Diff.	Probability	Odds
+3.0	.999	739.8 to 1.0	−0.1	.540	1.2 to 1.0
+2.9	.998	534.9 to 1.0	−0.2	.579	1.4 to 1.0
+2.8	.997	390.4 to 1.0	−0.3	.618	1.6 to 1.0
+2.7	.997	287.4 to 1.0	−0.4	.655	1.9 to 1.0
+2.6	.995	213.5 to 1.0	−0.5	.691	2.2 to 1.0
+2.5	.994	160.0 to 1.0	−0.6	.726	2.6 to 1.0
+2.4	.992	121.0 to 1.0	−0.7	.758	3.1 to 1.0
+2.3	.989	92.2 to 1.0	−0.8	.788	3.7 to 1.0
+2.2	.986	70.9 to 1.0	−0.9	.816	4.4 to 1.0
+2.1	.982	55.0 to 1.0	−1.0	.841	5.3 to 1.0
+2.0	.977	43.0 to 1.0	−1.1	.864	6.4 to 1.0
+1.9	.971	33.8 to 1.0	−1.2	.885	7.7 to 1.0
+1.8	.964	26.8 to 1.0	−1.3	.903	9.3 to 1.0
+1.7	.955	21.4 to 1.0	−1.4	.919	11.4 to 1.0
+1.6	.945	17.2 to 1.0	−1.5	.933	14.0 to 1.0
+1.5	.933	14.0 to 1.0	−1.6	.945	17.2 to 1.0
+1.4	.919	11.4 to 1.0	−1.7	.955	21.4 to 1.0
+1.3	.903	9.3 to 1.0	−1.8	.964	26.8 to 1.0
+1.2	.885	7.7 to 1.0	−1.9	.971	33.8 to 1.0
+1.1	.864	6.4 to 1.0	−2.0	.977	43.0 to 1.0
+1.0	.841	5.3 to 1.0	−2.1	.982	55.0 to 1.0
+0.9	.816	4.4 to 1.0	−2.2	.986	70.9 to 1.0
+0.8	.788	3.7 to 1.0	−2.3	.989	92.2 to 1.0
+0.7	.758	3.1 to 1.0	−2.4	.992	121.0 to 1.0
+0.6	.726	2.6 to 1.0	−2.5	.994	160.0 to 1.0
+0.5	.691	2.2 to 1.0	−2.6	.995	213.5 to 1.0
+0.4	.655	1.9 to 1.0	−2.7	.997	287.4 to 1.0
+0.3	.618	1.6 to 1.0	−2.8	.997	390.4 to 1.0
+0.2	.579	1.4 to 1.0	−2.9	.998	534.9 to 1.0
+0.1	.540	1.2 to 1.0	−3.0	.999	739.8 to 1.0
0.0	No change in report ratings or projections noted; therefore, odds of "actual" increase versus decrease are 1:1.				

.989 probability value from 1.000 to obtain .011. This means that there are only eleven chances out of one thousand that a real change did *not* occur in an upward direction.

2. Changes In A *Projection Number* From Book-To-Book

a. **Illustrative Application Situation**—In the previous survey (Report #1) station WBBB in Market Y had a reported Total Survey Area Cume audience projection of 150,000 adults 18-34 in the Monday–Friday, 7PM–Midnight time period. In the current survey period (Report #2), their audience projection was reported as 140,000 – a drop of 10,000. To what extent does this reported change indicate a real decline in WBBB's audience? What are the odds that this reported decline could be nothing more than the result of random statistical bounce (i.e., sampling error)?[39]

b. Computational Details[40] –

Step 1. Assemble the Required Basic Data

Data Item	Data Value	Data Source
Audience Projection #1	150,000	Arbitron Report #1: Mkt Y
Audience Projection #2	140,000	Arbitron Report #2: Mkt Y
Standard Error of Audience Projection #1	20,000	Arbitron Report # 1: Mkt Y (via use of Table A and B)[41]
Standard Error of Audience Projection #2	18,500	Arbitron Report #2: Mkt Y (via use of Table A and B)[41]

Step 2. Determine The Probability (and Odds) That A Real Change Occurred In The Direction Indicated By The Projection Number Difference Between the Two Reports By Performing Three Simple Calculations Plus A Table "Look Up."

1. Compute the Projection Number Difference:

Proj. # Difference = (Proj. #2) – (Proj. #1)
Proj. # Difference = (140,000) – (150,000)
Proj. # Difference = – 10,000

[39]See Note 35

[40]See Note 36

[41]See Chapter 4, Section A.2, for a general explanation of Table A and Table B. Also, see Appendix A for general instructions for how to use Table A and Table B to obtain the standard error of an audience projection number. Market Y Tables A and B are *not* shown here since example is just hypothetical. Also for pre-Fall 1981 Survey Reports which do not include a Table A or Table B, see Appendix F to estimate the standard error of an audience projection number.

2. Compute the Standard Error of the Difference in Projection Numbers:

$$\text{Std. Error of Diff.} = \sqrt{\left(\begin{matrix}\text{Std. Error}\\ \text{of Rating \#2}\end{matrix}\right)^2 + \left(\begin{matrix}\text{Std. Error}\\ \text{of Rating \#1}\end{matrix}\right)^2}$$

$$\text{Std. Error of Diff.} = \sqrt{(18{,}500)^2 + (20{,}000)^2}$$

$$\text{Std. Error of Diff.} = \sqrt{(342{,}250{,}000) + (400{,}000{,}000)}$$

$$\text{Std. Error of Diff.} = \sqrt{742{,}250{,}000}$$

$$\text{Std. Error of Diff.}^{42} = 27{,}\ 244.265$$

3. Compute the Standardized Difference (Stdz. Diff.):

$$\text{Stdz. Diff.} = \frac{\text{(Projection Number Difference)}}{\text{(Std. Error of Diff.)}}$$

$$\text{Stdz. Diff.} = \frac{(-10.000)}{(27{,}244.265)}$$

$$\text{Stdz. Diff.} = -0.36705$$

$$\text{Stdz. Diff.} = -0.4 \text{ (rounded to 1 decimal place)}$$

4. Use the Standardized Difference value (−0.4) as the reference point in Table 8 to obtain Probability and Odds that a "real decrease" in audience has occurred for station WBBB.

c. **Interpretation**[43]—By using the −0.4 Standardized Difference value to reference Table 8 (p. 79), the probability that a real *decrease* occurred in the ratings for station WBBB is determined to be .655. This probability level is equivalent to 1.9 to 1.0 odds that a real change occurred in the *downward direction*. Thus, there is only weak evidence that a real decline actually occurred WBBB's Mon–Fri 7PM–Midnight Cume audience among Adults 18-34. There is a reasonably high likelihood that the reported decline can be explained by normal statistical bounce (sampling error) in the reported audience numbers. In fact, there is a 34.5% chance that the reported change is not indicative of a true decline in audience for WBBB.

---

[42]For calculators with a limited display, the quantity $\sqrt{742{,}250{,}000}$ can be evaluated by breaking this down into two parts which can then be multiplied to obtain the required answer. For example, on an 8 digit display calculator, the $\sqrt{742{,}250{,}000}$ can be evaluated as follows:

$$\sqrt{742{,}250{,}000} = \sqrt{742{,}250} \times \sqrt{1{,}000} = (861.53932 \times 31.622777) = 27{,}244.265$$

[43]See Note 38.

IMPORTANT: By definition, probability values must sum to 1.000. Therefore, by subtracting the .655 probability from 1.000, a probability of .345 (34.5% chance) is obtained to support the contention that the reported change is not truly indicative of a real decline in audience for WBBB.

B. Determining Statistical Confidence Limits For Reported Audience Estimates

In many situations, knowledge about the sampling error (statistical bounce) associated with any particular audience rating or projection number is usefully summarized in the form of the quantity called a Standard Error. Based upon Arbitron's new Radio Reliability Model, the determination of the standard error is now an easy task for report users. The procedure is described in Appendix A, as well as in the text in Chapter 4, Section A.2. In some practical situations, no further information is required to make a satisfactory assessment of reliability. For example, if a specific audience projection is 90,000 and its standard error is determined to be 3,000, most users would agree that the projection is quite reliable.

On some occasions, however, users may wish to approach the question of reliability assessment in a more formal manner. The purpose of this section is to illustrate how upper and lower boundaries can be formally determined at pre-chosen confidence levels to describe the size of the possible error range that may be associated with reported audience estimates due simply to the fact that Arbitron surveys a sample of respondents, instead of surveying its entire population frame with identical procedures to those used for the original report.

In the procedures that follow, the report user must select a "confidence level" (i.e., probability, or degree of certainty) to use in determining the potential error range associated with any particular audience estimate as a result of the sampling error phenomenon. In the examples given in the following pages, the confidence level value has been set at 90% for illustrative purposes. An error range determined at the 90% confidence level means that there is a .90 probability (9:1 odds) that, if the survey were conducted again (in the same way and under identical conditions) but among the entire population frame used originally, the expected audience estimate value of this comparable complete coverage would lie within the calculated error range.

The choice of what confidence level to use in the context of a particular problem depends upon an evaluation of the relative risks involved in terms of what statisticians call the "alpha " and "beta" errors. A discussion of these issues can be found in most standard statistical texts, and it is therefore beyond the scope of this report.[44] It is sufficient to state here that the choice of what confidence level to use has to be balanced against other considerations. Thus, choosing a confidence level involves a trade-off. A choice of too high a confidence level will produce an unduly wide error range statement that might prove to have little utility as an aid to decision-making.

For the convenience of the reader interested in using confidence levels other than 90%, Table 9 shows the various commonly used confidence levels and their corresponding "Z" values. The Z value is the standard error multiplier factor that is required in order to determine the size of the error range at the chosen confidence level. The examples that follow will illustrate how the Z value is used for making calculations of confidence limits.

TABLE 9. Z VALUES CORRESPONDING TO VARIOUS CONFIDENCE LEVELS

Confidence Level (Percent)	Z Value
50	0.67
60	0.84
68	1.00
70	1.04
80	1.28
90	1.64
95	1.96
98	2.33
99	2.58

1. Confidence Limits For A Single Audience *Rating*

a. **Illustrative Application Situation** — Station WCCC in Market X has a reported Average Quarter Hour estimated rating of 1.0 among Men 18+ in the Metro Survey Area for the Mon–Fri 6AM–10 AM daypart. What are the 90% confidence limits for WCCC's estimated rating? In other words, what is the range within which it can be expected with 90% certainty that Station WCCC's rating would fall if Arbitron had used a different sample of respondents for its survey, or had attempted to survey the entire population frame instead of a sample?

[44] See, for example, Walker, Helen W. and Lev, Joseph, *Statistical Inference*, Henry Holt and Company, Inc. (1953). In particular, see Chapter 3, pages 60-66.

b. Computational Details—

Step 1. Assemble the Required Basic Data

Data Item	Data Value	Data Source
Rating	1.0	Arbitron Report: Mkt X
Standard Error of Rating	0.146	Arbitron Report: Mkt X (Table A ÷ Table B)[45]
Z value for 90% Confidence Level	1.64	Table 9 (p. 83)

Step 2. Determine 90% Confidence Limits As Follows:

- Lower 90% Confidence Limit
  = Rating − (Z value × Standard Error)
  = 1.0 − (1.64 × 0.146)
  = 1.0 − 0.23944
  = 0.76056
  = 0.8 (rounded to 1 decimal place)
- Upper 90% Confidence Limit
  = Rating + (Z value × Standard Error)
  = 1.0 + (1.64 × 0.146)
  = 1.0 + 0.23944
  = 1.23944
  = 1.2 (rounded to 1 decimal place)

c. Interpretation—The calculation yields 90% confidence limits of 0.8 and 1.2 as the lower and upper boundary values for the rating. This would inform the station that there is a 90% probability (9:1 odds) that if Arbitron had used a different sample of respondents for its survey, WCCC's reported rating would be within the 0.8 to 1.2 range (with 1.0 being the most likely value). Alternatively, it can be stated that, if the total population frame from which the sample were drawn had been surveyed by Arbitron under identical conditions, the expected rating value of this comparable complete coverage would fall in the range of 0.8 to 1.2 with 90% certainty.

[45]See Chapter 4, Section A.2, for general explanation of Tables A and B. Market X Tables A and B are *not* shown here since example is just hypothetical.

2. Confidence Limits For A Single Audience *Projection Number*

a. Illustrative Application Situation — In the latest report for Market Y, Station WAAA is found to have a 6–10AM Monday–Friday Cume audience estimate of 150,000 Adults 18 + in the Total Survey Area. What are the 90% confidence limits for WAAA's estimated audience? In other words, if Arbitron had attempted to survey the entire population frame rather than a sample, within what range would the result be expected to fall with 90% certainty?

b. Computational Details —

Step 1. Assemble the Required Basic Data

Data Item	Data Value	Data Source
Cume Audience Projection #	150,000	Arbitron Report: Market Y
TSA Population (Adult 18 +)	3,000,000	Arbitron Report: Market Y
Cume Rating	5.0	(Cume Audience ÷ TSA Pop.) × 100 = (150,000 ÷ 3,000,000) × 100 = 5.0
Standard Error of Cume Rating	.508	Arbitron Report: Market Y (Table A ÷ Table B)[46]
Z value for 90% Confidence Level	1.64	Table 9 (p. 83)

Step 2. Calculate Standard Error of Cume Projection Number

Std. Error of Cume Proj. # = (Std. Error of Cume Rating × TSA Pop) ÷ 100

Std. Error of Cume Proj. # = (.508 × 3,000,000) ÷ 100

Std. Error of Cume Proj. # = 1,524,000 ÷ 100

Std. Error of Cume Proj. # = 15,240 (rounded to a whole number)

---

[46]See Note 45.

Step 3. Determine 90% Confidence Limits

- Lower Confidence Limit
  = Proj. # − (Z value × Standard Error of Proj. #)
  = 150,000 − (1.64 × 15,240)
  = 150,000 − 24,993.6
  = 125,006.4
  = 125,006 (rounded to whole #)
- Upper Confidence Limit
  = Proj. # + (Z value × Standard Error of Proj. #)
  = 150,000 + (1.64 × 15,240)
  = 150,000 + 24,993.6
  = 174,993.6
  = 174,994 (rounded to whole #)

c. **Interpretation** — The calculation yields 90% confidence limits of 125,006 to 174,994 as the lower and upper boundary values for WAAA's TSA Cume audience projection among Adults 18+ during Mon–Fri morning drive time. This confidence interval means that, if Arbitron had attempted to survey the entire population frame rather than a sample, WAAA's TSA Cume audience for Mon–Fri 6–10AM among Adults 18+ would be expected to be between 125 thousand and 175 thousand (with 90% certainty). It should also be noted that a tighter interval could be calculated if, for example, the 68% confidence limits were calculated by using a Z value of 1.00 (instead of the 1.64 used for the 90% confidence limits). In this instance, the 68% confidence interval is calculated to be 134,760 to 165,240. This means that there are slightly over 2:1 odds (68% ÷ 32% = 2.125:1) that WAAA's TSA Cume morning drive time audience among Adults 18+ would be between about 135 to 165 thousand if Arbitron attempted to survey the entire population frame, rather than a sample.

3. Confidence Limits For An *Aggregated Audience Rating* Averaged Across Several Reports

a. **Illustrative Application Situation** — Station KSSS in Market X wants to evaluate its average performance over the past four survey report periods in terms of its Saturday 10:00AM–3:00 PM Average Quarter Hour (AQH) listening among Women 18–34 in the Metro Survey Area. The arithmetic mean of the station's AQH Ratings in the four survey reports is used to evaluate the station's average performance. The mean rating is calculated to be 3.7. What are the 90% confidence limits of this mean rating?

b. Computational Details—

Step 1. Assemble the Required Basic Data

Survey Period	Rating[47]	Standard Error of Rating[48]
Fall (Report #1)	2.2	0.199
Winter (Report #2)	4.9	0.287
Spring (Report #3	3.4	0.245
Summer (Report #4)	4.3	0.262
Mean Rating	3.7	To be determined in Step 2

Step 2. Determine the Standard Error of Mean Rating

The general formula for the Standard Error of a Mean Rating is:

$$S_{mean^*} = \frac{\sqrt{\Sigma S_i^2}}{n}$$

where n = the number of report ratings being averaged

S_i = the Standard Error for the rating in the i^{th} report where i stands for any single report

Σ = the sum across all "n" reports

S_{mean} = the Standard Error of the Mean Rating.

*This formula applies only when the average rating is

[47]From Arbitron Local Market Radio Reports.

[48]Calculated as Table A ÷ Table B from each of the Arbitron Reports. These values are not shown here for the four Market X Survey Reports since this example is just hypothetical. See Chapter 4, Section A.2, for general explanation of Table A and Table B.

Note: Table A and B do not appear in survey reports prior to Fall 1981. A quick estimate of the standard error of an audience rating from a pre-Fall 1981 survey report is obtained by using the Tables A and B values contained in the more recent reports. A more accurate method is described in Appendix F.

computed by using a standard unweighted mean of the individual report ratings defined as follows:

$$R_{mean} = \frac{\Sigma R_i}{n}$$

where R_i = the rating value in the i^{th} Report
R_{mean} = the unweighted mean across all "n" reports

Applying the above general formula to determine the Standard Error of the 3.7 Mean Rating in the example yields the following:

$$S_{mean} = \frac{\sqrt{(0.199)^2 + (0.287)^2 + (0.245)^2 + (0.262)^2}}{4}$$

$$S_{mean} = \frac{\sqrt{0.039601 + 0.082369 + 0.060025 + 0.068644}}{4}$$

$$S_{mean} = \frac{\sqrt{0.250639}}{4}$$

$$S_{mean} = \frac{0.5006386}{4}$$

$S_{mean} = 0.1251597 = 0.125$ (rounded to 3 decimal places)

Step 3. Determine 90% Confidence Limits

- Lower Confidence Limit
  = (Mean Rating) − (Z Value* × S_{mean})
  = $3.7 - (1.64 \times 0.125)$
  = $3.7 - 0.205$
  = $3.495 = 3.5$ (rounded to 1 decimal place)

- Upper Confidence Limit
  = (Mean Rating) + (Z Value* × S_{mean})
  = $3.7 + (1.64 \times 0.125)$
  = $3.7 + 0.205$
  = $3.905 = 3.9$ (rounded to 1 decimal place)

*The Z value is obtained from Table 9 (p. 83). It is the Standard Error multiplier factor that corresponds to the 90% Confidence Limits. See explanation on p. 82.

c. **Interpretation** – The average (mean) rating across the past four survey reports for station KSSS is 3.7 for its Saturday 10:00 AM–3:00 PM AQH listening among Women 18–34. The rating in each of the individual survey reports is subject to a certain degree of statistical bounce due to the fact that a sample was surveyed instead of the entire population frame. Therefore, the *mean* of these four report ratings is, in turn, also subject to a certain degree of statistical bounce. From the above calculations, the researcher has quantified the amount of potential statistical bounce around the 3.7 mean rating. The researcher has determined that, at the 90% confidence level (9:1 odds), the lower limit is 3.5 and the upper limit is 3.9.

4. Confidence Limits For An *Aggregated Audience Projection Number* Averaged Across Several Reports

a. **Illustrative Application Situation** – Station WJJJ in Market Y wishes to estimate the average reach of its station over the past two survey report periods for the Mon–Sun 6AM–Midnight daypart among Teens in the Total Survey Area. The average (arithmetic mean) Cume audience projection number across the two reports is 19,500. At the 90% confidence level, how much allowance should be made for the statistical sampling error about this 19,500 average projection?

b. **Computational Details** –

Step 1. Assemble the Required Basic Data

Survey Period	Cume Projection[49]	Standard Error of Cume Proj.[50]
Fall (Report #1)	15,700	1,238
Spring (Report #2)	23,300	1,871
Mean Projection	19,500	To be determined in Step 2

[49]From Arbitron Local Market Radio Reports.

[50]Calculated with Table A and Table B using the procedure explained in Appendix A and in an earlier example: B-2 (p. 85). The Tables A and B are not shown for the two Market "Y" survey reports since this example is just hypothetical. See Chapter 4, Section A.2, for general explanation of Table A and Table B.

Note: Table A and B do not appear in survey reports prior to Fall 1981. A quick estimate of the standard error of an audience rating from a pre-Fall 1981 survey report is obtained by using the Tables A and B values contained in the more recent reports. A more accurate method is described in Appendix F.

Step 2. Determine the Standard Error of Mean Projection Number

The general formula for the Standard Error of a Mean Projection Number is:

$$S'_{mean}{}^{*} = \frac{\sqrt{\Sigma S'^{2}_{i}}}{n}$$

where

n = the number of report projections being averaged

S'_i = the Standard Error for the projection number in the i^{th} report where i stands for any single report

Σ = the sum across all "n" reports

S'_{mean} = the Standard Error of the Mean Projection Number.

*This formula applies only when the average projection number is computed by using a standard unweighted mean of the individual report projection numbers defined as:

$$PN_{mean} = \frac{\Sigma PN_i}{n}$$

where

PN_i = the projection number in the i^{th} Report

PN_{mean} = the unweighted mean across all "n" reports

Applying the above general formula to determine the Standard Error of the 19,500 projection number in the example yields the following:

$$S'_{mean} = \frac{\sqrt{(1238)^2 + (1871)^2}}{2}$$

$$S'_{mean} = \frac{\sqrt{1,532,644 + 3,500,641}}{2}$$

$$S'_{mean} = \frac{\sqrt{5,033,285}}{2}$$

$$S'_{mean} = \frac{2243.4984}{2}$$

$$S'_{mean} = 1121.7492$$

Step 3. Determine 90% Confidence Limits for Mean Projection Number

- Lower Confidence Limit
  = (Mean Proj. No.) − (Z Value* × S'_{mean})
  = 19,500 − (1.64 × 1121.7492)
  = 19,500 − 1839.6687
  = 17,660.3310
  = 17,660 (rounded to a whole number)
- Upper Confidence Limit
  = (Mean Proj. No.) + (Z Value* × S'_{mean})
  = 19,500 + (1.64 × 1121.7492)
  = 19,500 + 1839.6687
  = 21,339.669
  = 21,340 (rounded to a whole number)

*The Z value is obtained from Table 9 (p. 83). It is the Standard Error multiplier factor that corresponds to the 90% Confidence Limits.

c. **Interpretation** — The average (mean) projection number across the two past survey reports for station WJJJ is 19,500 for its 6AM–Midnight Cume audience among Teens in the Total Survey Area. The Cume projection number in each of the individual survey reports is subject to statistical sampling error. Therefore, the mean projection number is, in turn, also subject to a certain degree of statistical sampling error. Based upon the above calculations, the 90% confidence interval for the mean projection number was determined to be 17,660 to 21,340.

C. Some Commentary Concerning More Complex Applications

The applications illustrated in Sections A and B of this chapter involved the use of just one audience estimate per survey report; therefore, each estimate in the comparison (or average) is based upon a *different* sample of respondents (diary keepers). When an application situation involves the use of more than one rating[51] taken from the *same* survey report, the determination of the Standard Errors of

[51]The concepts described here also apply to projection numbers.

rating differences, rating averages, or sums of ratings [52] is considerably more complicated.

The complication stems from the fact that, in these instances, the estimates being compared (or averaged, or summed) are all being generated from the *same* single sample of respondents. As a result, the procedures and formulae described in Section A (Book-to-Book Changes in Reported Audiences) *are not applicable* for comparing differences between two ratings that are contained in the *same* report. Similarly, the procedures and formulae described in Sections B-3 and B-4 (Confidence Limits for aggregated audience estimates averaged across several Reports) *are not applicable* to averages derived from several ratings that are obtained from the *same* report.

The formulae appropriate to the situations where more than one rating in the analysis come from the *same* survey report all require the use of a quantity called the Correlation Coefficient, denoted by the symbol "rho". For example, the formula for the Standard Error of the difference between two ratings taken from the *same* survey report is as follows:

$$\sqrt{\begin{pmatrix}\text{Standard}\\ \text{Error of}\\ \text{Rating \#1}\end{pmatrix}^2 + \begin{pmatrix}\text{Standard}\\ \text{Error of}\\ \text{Rating \#2}\end{pmatrix}^2 - 2 \times \text{rho} \times \begin{pmatrix}\text{Standard}\\ \text{Error of}\\ \text{Rating \#1}\end{pmatrix} \times \begin{pmatrix}\text{Standard}\\ \text{Error of}\\ \text{Rating \#2}\end{pmatrix}}$$

The Correlation Coefficient can assume numerical values from -1 to $+1$. The value will vary depending upon the type of audience duplication or lack of duplication found among the sample of diary keepers for the days and time periods involved in the audience estimates being evaluated. Therefore, the value of the Correlation Coefficient must be estimated by special methods. Sometimes it is "guesstimated" from historical experience.

In general, the relationship between the Correlation Coefficient and audience duplication is roughly as follows:

- When audience duplication is heavy, the Correlation Coefficient will be positive. This might be expected to occur, for exam-

[52]The Gross Rating Point statistic for a radio spot schedule is an example of an application that involves *sums of ratings*.

ple, when comparing the same station for consecutive time periods.

- When audience duplication is virtually non-existent, the Correlation Coefficient will be negative. This can be expected to occur for example, when comparing different stations of opposite formats in the same time period.

The value of the Correlation Coefficient between the two ratings will vary with the factors involved in the rating comparison. For example, does the comparison involve the same demographic group, or different demographic groups?, the same station or different stations?, etc.

Another type of application example that involves the use of more than one audience estimate from a single report is the evaluation of advertising schedules on the basis of the sum of the ratings of the individual announcements—i.e., Gross Rating Points. Here again, the ratings typically originate from the *same* survey report, and, therefore, audience duplication must be considered. A proper calculation of the Standard Error of a Gross Rating Point statistic requires not only the Standard Errors associated with the individual ratings that make up the schedule, but also the Correlation Coefficient between each *pair* of ratings that comprise the schedule. To ignore this consideration can lead to improper inferences.

The complexities in determining the Standard Errors in the types of situations just outlined are by no means unique to Arbitron Surveys. These complexities, which center about determining the value of the Correlation Coefficient, are common to all surveys that use multiple measurements on the same sample of respondents. Thus, to omit this consideration can lead to incorrect decisions.

Special study of the analysis of the Standard Errors for audience estimates drawn from the *same* survey report is beyond the scope of the present study. However, any such study must draw upon accurate estimates of the Standard Errors of individual ratings as the starting point. Hence, the *Arbitron Replication II* study provides a firm foundation for these more complex analyses.

D. Technical Discussion: "Classical" vs. "Bayesian" Theories of Statistical Inference

At the present time there are two basic theories of statistical inference in widespread use: Classical and Bayesian. From a mathematical standpoint, these theories differ in the way that they view the joint distribution of the sample estimate and the corresponding population parameter. More specifically, Classical inference is based upon the functional distribution of the sample estimate given (or holding fixed) the population parameter, while Bayesian inference is based upon the functional distribution of the population parameter holding the sample estimate fixed.

From a practical users standpoint, the techniques and methods of Classical and Bayesian (diffuse prior) inference for samples of the sizes used in Arbitron Radio Reports, are essentially the same in their computational form. They differ in the way they interpret the results they produce. This difference is linked to the way these two theories look at probability.

This difference is perhaps best illustrated by an example of the interpretation of a simple confidence interval of the type described in Example B-1. In that example, a 90% confidence interval was computed for a reported rating of 1.0. This audience estimate was found to have a lower 90% confidence limit of 0.8 and an upper 90% confidence limit of 1.2.

Using the Bayesian approach to inference these numbers may be interpreted as follows: If Arbitron had attempted to survey the entire population frame rather than a sample, then the probability is 90% that the result would fall somewhere in the range 0.8 to 1.2.

The Classical interpretation of this confidence interval (i.e., the same numbers shown above) requires a somewhat longer explanation. The audience rating that would result if the survey had been undertaken for the frame population, rather than a sample, is a fixed, but unknown quantity. Since this fixed, but unknown quantity either falls within the range 0.8 to 1.2 or it does not fall within this range, the corresponding probability (also unknown) is either zero or one. In other words, in the Classical interpretation, one cannot talk about a probability that the range 0.8 to 1.2 includes the total population frame rating. It either does or it doesn't. Which alternative is true remains unknown.

In the Classical interpretation, the 90% confidence interval that is produced from the sample must be interpreted using the following scenario: Suppose that rather than a single study, a very large number of identical studies were undertaken — using the same methods, at exactly the same time. Further, suppose that for each of these studies a set of upper and lower limits at the 90% confidence level were computed. These upper and lower limits (as well as their midpoint — i.e., the survey rating estimate) will differ among the different surveys. For 90% of the surveys, the upper and lower limits will include the rating that would be obtained from a study undertaken for the entire population frame.

Given the complexity of the Classical interpretation of confidence intervals, and the simple and straightforward nature of the Bayesian interpretation, it is not surprising that many applied statisticians have preferred the Bayesian theory of inference. Although it first may appear that the Bayesian and Classical interpretation of confidence intervals are essentially the same, they are, in fact, quite different. Most people who learned the Classical interpretation of probability find it more useful to adopt the Bayesian approach when they are in a decision-making situation.

In addition to its different interpretation of probability, the Bayesian method of inference allows for the type of direct odds determination used in Examples A-1 and A-2. The odds determination in those examples are called Posterior Odds Assessment because they make use of the posterior distribution of the difference between two population parameters (i.e., the population frame rating for two different time periods). Given the sample sizes used by Arbitron and under the assumption of a diffuse prior distribution, the posterior distribution of the difference between the two population frame parameters will be approximately normal with a standard deviation equal to the Standard Error of the Difference. The mean of this distribution is the estimated sample difference (Estimate 2 Minus Estimate 1). When the Standardized Difference value is positive, the area of this distribution above zero gives the probability that the change is in the positive direction shown by the samples, while the area below zero provides the probability that the change is in the opposite direction. When the Standardized Difference value is negative, the area of this distribution below zero gives the probability that the change is in the negative direction shown by the samples, while the area above zero provides the probability that the change is in the opposite direction. These areas are easily computed from standard normal tables, or by the approximations used in the BASIC program provided in Appendix G of this document.

Section V: Appendices

Appendix A

An Example of An Actual Table A and Table B As They Appear In Arbitron's Local Market Radio Reports

NOTE

The values for Table B are individually, custom-derived for each market and report period. Hence, the Table B that follows applies only to the New York, Fall 1981 Report. Table A values, however, are universally applied across all Arbitron Radio Reports. For further explanation, see text (Chapter 4, Section A.2).

INSTRUCTIONS FOR THE DETERMINATION OF RELIABILITY AND EFFECTIVE SAMPLE BASES (ESB'S)

To Obtain Standard Error of a Rating

The approximate Standard Error (one sigma level) for specific ratings published in this Report may be determined by dividing a value obtained from Table A by a value obtained from Table B.

The Table A value depends only on the value of the specific rating. For example, if the rating is 1.2, the Table A value is 10.89.

Table B values are shown for each of the geographic areas included in this Report; Metro, TSA and ADI (if applicable). Using the appropriate geographic area, find the Table B value on the basis of the demographic group and estimate type for the specific rating. For example, for a weekday, 6AM-10AM average quarter-hour rating among Males 18-34 in the Metro area, the Table B value would be determined from the appropriate row (Males 18-34) and column (estimate type Monday-Friday 6AM-10AM or Monday-Friday 3PM-7PM) of the Metro section of Table B.

The Standard Error (one sigma level) is obtained by dividing the Table A value by the Table B value. For example, if the Table A value was found to be 7.05 and the Table B value was found to be 28.20, the Standard Error would be 7.05 ÷ 28.20 = 0.25. ***Note: Table B values differ from report to report; hence, the 28.20 value is used here for illustrative purposes only.***

It should be noted that columns 2-11 of Table B refer to average quarter-hour ratings only. Column 1 of Table B is to be used for any cume rating.

To Obtain Standard Error of a Projected Audience Number

In order to determine the Standard Error for audiences expressed in terms of projected numbers of persons rather than ratings, first calculate the corresponding ratings. (The rating is calculated by dividing the projected audience by the corresponding population base and multiplying the result by 100.0).

After the rating is calculated, proceed as described above for determining the Standard Error for the audience expressed as a rating. ***Then, multiply the resulting Standard Error of the rating by the corresponding population base and divide the result by 100, to produce the Standard Error for the projected number of persons.***

For example, if the projected audience is 1,300 and the corresponding population is 100,000, the corresponding rating would be (1,300 ÷ 100,000) x 100 = 1.3. If the Standard Error for this rating were found to be 0.30, the Standard Error for the projected number of persons would be (0.30 x 100,000) ÷ 100 = ± 300.

To Convert Standard Errors to Confidence Intervals

Standard Error levels determined as described above may be used to determine 68%, 80%, 90%, 95% and 99% confidence intervals by adding and subtracting the appropriate multiple of the Standard Error to and from the corresponding rating or audience estimate. The multiples are 1.00, 1.28, 1.64, 1.96 and 2.58 respectively.

For example, if the estimated rating is 2.0 and its estimated Standard Error is 0.30, the lower 95% confidence limit is 2.0 ***minus*** (1.96 x 0.30) which equals 1.412, and the upper 95% confidence limit is 2.0 ***plus*** (1.96 x 0.30) which equals 2.588.

Note: Prior to the Fall 1981 Reports, Standard Error levels were derived via the Nomograph procedures. The Nomograph procedures yielded values that were Standard Errors multiplied by a factor of 2.00.

To Obtain Effective Sample Base (ESB) Sizes

Users may determine the approximate ESB for various age-sex groups and estimate types (including different geographic areas) by squaring the values shown in Table B. (Also see Section 38, page ii, for further discussion concerning ESB values.)

For example, if the Table B value shown for Males 18+ for Monday-Friday 6AM-10AM average quarter-hour estimates in the Metro area is 50, the corresponding ESB value is 50 squared which equals 2500.

To Obtain Additional Methodology Details

The derivation of these tables and the above procedures for

RATING	VALUE	RATING	VALUE	RATING	VALUE
0.1	3.16	5.1	22.00	10.1	30.13
0.2	4.47	5.2	22.20	10.2	30.26
0.3	5.47	5.3	22.40	10.3	30.40
0.4	6.31	5.4	22.60	10.4	30.53
0.5	7.05	5.5	22.80	10.5	30.66
0.6	7.72	5.6	22.99	10.6	30.78
0.7	8.34	5.7	23.18	10.7	30.91
0.8	8.91	5.8	23.37	10.8	31.04
0.9	9.44	5.9	23.56	10.9	31.16
1.0	9.95	6.0	23.75	11.0	31.29
1.1	10.43	6.1	23.93	11.1	31.41
1.2	10.89	6.2	24.12	11.2	31.54
1.3	11.33	6.3	24.30	11.3	31.66
1.4	11.75	6.4	24.48	11.4	31.78
1.5	12.16	6.5	24.65	11.5	31.90
1.6	12.55	6.6	24.83	11.6	32.02
1.7	12.93	6.7	25.00	11.7	32.14
1.8	13.30	6.8	25.17	11.8	32.26
1.9	13.65	6.9	25.35	11.9	32.38
2.0	14.00	7.0	25.51	12.0	32.50
2.1	14.34	7.1	25.68	12.1	32.61
2.2	14.67	7.2	25.85	12.2	32.73
2.3	14.99	7.3	26.01	12.3	32.84
2.4	15.30	7.4	26.18	12.4	32.96
2.5	15.61	7.5	26.34	12.5	33.07
2.6	15.91	7.6	26.50	12.6	33.18
2.7	16.21	7.7	26.66	12.7	33.30
2.8	16.50	7.8	26.82	12.8	33.41
2.9	16.78	7.9	26.97	12.9	33.52
3.0	17.06	8.0	27.13	13.0	33.63
3.1	17.33	8.1	27.28	13.1	33.74
3.2	17.60	8.2	27.44	13.2	33.85
3.3	17.86	8.3	27.59	13.3	33.96
3.4	18.12	8.4	27.74	13.4	34.07
3.5	18.38	8.5	27.89	13.5	34.17
3.6	18.63	8.6	28.04	13.6	34.28
3.7	18.88	8.7	28.18	13.7	34.38
3.8	19.12	8.8	28.33	13.8	34.49
3.9	19.36	8.9	28.47	13.9	34.59
4.0	19.60	9.0	28.62	14.0	34.70
4.1	19.83	9.1	28.76	14.1	34.80
4.2	20.06	9.2	28.90	14.2	34.91
4.3	20.29	9.3	29.04	14.3	35.01
4.4	20.51	9.4	29.18	14.4	35.11
4.5	20.73	9.5	29.32	14.5	35.21
4.6	20.95	9.6	29.46	14.6	35.31
4.7	21.16	9.7	29.60	14.7	35.41
4.8	21.38	9.8	29.73	14.8	35.51
4.9	21.59	9.9	29.87	14.9	35.61
5.0	21.79	10.0	30.00	15.0	35.71

Note: When rating is greater than 50.0, use the "value" given

RADIO RELIABILITY — TABLE A

RATING	VALUE	RATING	VALUE	RATING	VALUE	RATING	VALUE	RATING	VALUE	RATING	VALUE	RATING	VALUE
15.1	35.80	20.1	40.07	25.1	43.36	30.1	45.87	35.1	47.73	40.1	49.01	45.1	49.76
15.2	35.90	20.2	40.15	25.2	43.42	30.2	45.91	35.2	47.76	40.2	49.03	45.2	49.77
15.3	36.00	20.3	40.22	25.3	43.47	30.3	45.96	35.3	47.79	40.3	49.05	45.3	49.78
15.4	36.09	20.4	40.30	25.4	43.53	30.4	46.00	35.4	47.82	40.4	49.07	45.4	49.79
15.5	36.19	20.5	40.37	25.5	43.59	30.5	46.04	35.5	47.85	40.5	49.09	45.5	49.80
15.6	36.29	20.6	40.44	25.6	43.64	30.6	46.08	35.6	47.88	40.6	49.11	45.6	49.81
15.7	36.38	20.7	40.52	25.7	43.70	30.7	46.12	35.7	47.91	40.7	49.13	45.7	49.81
15.8	36.47	20.8	40.59	25.8	43.75	30.8	46.17	35.8	47.94	40.8	49.15	45.8	49.82
15.9	36.57	20.9	40.66	25.9	43.81	30.9	46.21	35.9	47.97	40.9	49.16	45.9	49.83
16.0	36.66	21.0	40.73	26.0	43.86	31.0	46.25	36.0	48.00	41.0	49.18	46.0	49.84
16.1	36.75	21.1	40.80	26.1	43.92	31.1	46.29	36.1	48.03	41.1	49.20	46.1	49.85
16.2	36.85	21.2	40.87	26.2	43.97	31.2	46.33	36.2	48.06	41.2	49.22	46.2	49.86
16.3	36.94	21.3	40.94	26.3	44.03	31.3	46.37	36.3	48.09	41.3	49.24	46.3	49.86
16.4	37.03	21.4	41.01	26.4	44.08	31.4	46.41	36.4	48.11	41.4	49.25	46.4	49.87
16.5	37.12	21.5	41.08	26.5	44.13	31.5	46.45	36.5	48.14	41.5	49.27	46.5	49.88
16.6	37.21	21.6	41.15	26.6	44.19	31.6	46.49	36.6	48.17	41.6	49.29	46.6	49.88
16.7	37.30	21.7	41.22	26.7	44.24	31.7	46.53	36.7	48.20	41.7	49.31	46.7	49.89
16.8	37.39	21.8	41.29	26.8	44.29	31.8	46.57	36.8	48.23	41.8	49.32	46.8	49.90
16.9	37.48	21.9	41.36	26.9	44.34	31.9	46.61	36.9	48.25	41.9	49.34	46.9	49.90
17.0	37.56	22.0	41.42	27.0	44.40	32.0	46.65	37.0	48.28	42.0	49.36	47.0	49.91
17.1	37.65	22.1	41.49	27.1	44.45	32.1	46.69	37.1	48.31	42.1	49.37	47.1	49.92
17.2	37.74	22.2	41.56	27.2	44.50	32.2	46.72	37.2	48.33	42.2	49.39	47.2	49.92
17.3	37.82	22.3	41.63	27.3	44.55	32.3	46.76	37.3	48.36	42.3	49.40	47.3	49.93
17.4	37.91	22.4	41.69	27.4	44.60	32.4	46.80	37.4	48.39	42.4	49.42	47.4	49.93
17.5	38.00	22.5	41.76	27.5	44.65	32.5	46.84	37.5	48.41	42.5	49.43	47.5	49.94
17.6	38.08	22.6	41.82	27.6	44.70	32.6	46.87	37.6	48.44	42.6	49.45	47.6	49.94
17.7	38.17	22.7	41.89	27.7	44.75	32.7	46.91	37.7	48.46	42.7	49.46	47.7	49.95
17.8	38.25	22.8	41.95	27.8	44.80	32.8	46.95	37.8	48.49	42.8	49.48	47.8	49.95
17.9	38.34	22.9	42.02	27.9	44.85	32.9	46.98	37.9	48.51	42.9	49.49	47.9	49.96
18.0	38.42	23.0	42.08	28.0	44.90	33.0	47.02	38.0	48.54	43.0	49.51	48.0	49.96
18.1	38.50	23.1	42.15	28.1	44.95	33.1	47.06	38.1	48.56	43.1	49.52	48.1	49.96
18.2	38.58	23.2	42.21	28.2	45.00	33.2	47.09	38.2	48.59	43.2	49.54	48.2	49.97
18.3	38.67	23.3	42.27	28.3	45.05	33.3	47.13	38.3	48.61	43.3	49.55	48.3	49.97
18.4	38.75	23.4	42.34	28.4	45.09	33.4	47.16	38.4	48.64	43.4	49.56	48.4	49.97
18.5	38.83	23.5	42.40	28.5	45.14	33.5	47.20	38.5	48.66	43.5	49.58	48.5	49.98
18.6	38.91	23.6	42.46	28.6	45.19	33.6	47.23	38.6	48.68	43.6	49.59	48.6	49.98
18.7	38.99	23.7	42.52	28.7	45.24	33.7	47.27	38.7	48.71	43.7	49.60	48.7	49.98
18.8	39.07	23.8	42.59	28.8	45.28	33.8	47.30	38.8	48.73	43.8	49.61	48.8	49.99
18.9	39.15	23.9	42.65	28.9	45.33	33.9	47.34	38.9	48.75	43.9	49.63	48.9	49.99
19.0	39.23	24.0	42.71	29.0	45.38	34.0	47.37	39.0	48.77	44.0	49.64	49.0	49.99
19.1	39.31	24.1	42.77	29.1	45.42	34.1	47.40	39.1	48.80	44.1	49.65	49.1	49.99
19.2	39.39	24.2	42.83	29.2	45.47	34.2	47.44	39.2	48.82	44.2	49.66	49.2	49.99
19.3	39.47	24.3	42.89	29.3	45.51	34.3	47.47	39.3	48.84	44.3	49.67	49.3	50.00
19.4	39.54	24.4	42.95	29.4	45.56	34.4	47.50	39.4	48.86	44.4	49.69	49.4	50.00
19.5	39.62	24.5	43.01	29.5	45.60	34.5	47.54	39.5	48.89	44.5	49.70	49.5	50.00
19.6	39.70	24.6	43.07	29.6	45.65	34.6	47.57	39.6	48.91	44.6	49.71	49.6	50.00
19.7	39.77	24.7	43.13	29.7	45.69	34.7	47.60	39.7	48.93	44.7	49.72	49.7	50.00
19.8	39.85	24.8	43.19	29.8	45.74	34.8	47.63	39.8	48.95	44.8	49.73	49.8	50.00
19.9	39.92	24.9	43.24	29.9	45.78	34.9	47.67	39.9	48.97	44.9	49.74	49.9	50.00
20.0	40.00	25.0	43.30	30.0	45.83	35.0	47.70	40.0	48.99	45.0	49.75	50.0	50.00

for a rating equal to 100.0 minus the original rating. For example, if the rating were 87.3, use the "value" shown for a rating of 12.7 (100.0-87.3)

the determination of Reliability and Effective Sample Bases is described in the Arbitron publication entitled ***Arbitron Replication II: A Study of the Reliability of Radio Ratings.***

Limitations

Although Arbitron believes that the above described procedures provides report users with useful estimates of Standard Errors, the reader should note the Limitations described in paragraphs 37, 38, and 40 herein on page ii.

ARBITRON RADIO

SURVEY AREA/ DEMO GROUP	ESTIMATE				
		AVERAGE QUARTER			
	ALL CUME ESTIMATES	SAT 6A-10A SUN 6A-10A SAT 3P-7P SUN 3P-7P	WEEKDAYS SINGLE HOUR	SAT 10A-3P SUN 10A-3P SAT 7P-MID SUN 7P-MID	MON-FRI 6A-10A MON-FRI 3P-7P
METRO SURVEY AREA					
TOTAL 12 +	51.41	77.30	80.54	76.51	95.93
MEN 18 +	36.16	54.96	55.89	53.49	69.29
MEN 18-24	13.33	22.31	*****	21.10	30.64
MEN 25-34	18.12	28.60	*****	28.14	36.90
MEN 35-44	14.89	22.99	23.78	23.11	29.85
MEN 45-54	14.63	21.77	22.74	22.40	28.59
MEN 55-64	15.67	23.67	*****	24.07	31.69
MEN 18-34	22.67	35.56	37.06	34.33	46.40
MEN 18-49	28.28	43.88	44.72	42.52	55.64
MEN 25-44	*****	*****	36.76	*****	*****
MEN 25-49	25.33	38.35	39.03	37.80	48.32
MEN 25-54	27.69	41.37	*****	41.11	52.36
MEN 35-64	26.08	38.11	*****	38.92	49.44
WOMEN 18 +	41.35	60.87	62.51	60.16	74.45
WOMEN 18-24	16.71	26.94	*****	25.51	34.82
WOMEN 25-34	21.21	32.92	*****	32.70	42.44
WOMEN 35-44	18.03	27.68	29.27	27.87	35.83
WOMEN 45-54	18.22	27.50	27.99	27.18	34.38
WOMEN 55-64	17.06	24.71	*****	24.73	31.50
WOMEN 18-34	27.23	41.96	43.78	40.53	53.00
WOMEN 18-49	33.82	51.86	53.59	50.66	64.65
WOMEN 25-44	*****	*****	44.12	*****	*****
WOMEN 25-49	30.66	45.51	47.48	45.41	57.70
WOMEN 25-54	33.46	49.12	*****	49.06	62.42
WOMEN 35-64	30.92	44.79	*****	44.85	56.39
ADULTS 18 +	49.49	73.02	*****	72.02	90.53
ADULTS 18-34	32.94	50.83	*****	49.19	64.81
ADULTS 18-49	40.51	61.79	*****	60.14	77.04
ADULTS 25-49	35.82	53.69	*****	53.22	67.40
ADULTS 25-54	38.97	57.72	*****	57.57	72.80
ADULTS 35-64	36.66	52.93	*****	53.50	67.83
TEENS	21.09	34.55	39.13	35.21	47.23
TOTAL SURVEY AREA					
TOTAL 12 +	60.48	88.65	92.78	87.16	110.70
MEN 18 +	42.88	63.34	65.08	61.64	80.98
MEN 18-24	16.48	26.79	*****	25.55	36.76
MEN 25-34	21.73	34.44	*****	33.69	43.88
MEN 35-44	18.52	28.33	29.13	28.33	35.93
MEN 45-54	17.13	25.40	26.35	25.90	32.95
MEN 55-64	18.25	27.23	*****	27.55	35.95
MEN 18-34	27.16	42.28	43.97	41.02	55.17
MEN 18-49	34.00	51.35	52.54	50.06	65.15
MEN 25-44	*****	*****	43.68	*****	*****
MEN 25-49	30.11	45.35	45.87	44.64	56.45
MEN 25-54	32.86	48.66	*****	48.24	61.18
MEN 35-64	30.74	44.47	*****	45.05	57.47
WOMEN 18 +	47.89	69.66	71.32	67.89	84.89
WOMEN 18-24	19.00	31.31	*****	28.86	39.27
WOMEN 25-34	25.33	38.70	*****	38.79	50.61
WOMEN 35-44	21.52	32.99	34.44	32.57	41.72
WOMEN 45-54	20.93	31.70	32.14	31.45	39.07
WOMEN 55-64	20.21	29.26	*****	29.21	37.32
WOMEN 18-34	31.58	48.81	50.18	46.34	61.39
WOMEN 18-49	39.96	59.89	61.12	57.47	74.12
WOMEN 25-44	*****	*****	51.82	*****	*****
WOMEN 25-49	35.87	52.93	55.35	52.68	67.09
WOMEN 25-54	39.05	57.04	*****	56.67	72.03
WOMEN 35-64	35.68	52.19	*****	51.57	64.60
ADULTS 18 +	57.66	83.64	*****	81.97	104.95
ADULTS 18-34	38.80	59.17	*****	57.27	75.77
ADULTS 18-49	47.93	71.36	*****	69.25	90.15
ADULTS 25-49	42.34	62.47	*****	61.73	78.71
ADULTS 25-54	46.00	67.16	*****	66.63	84.60
ADULTS 35-64	42.81	61.32	*****	61.65	78.43
TEENS	25.24	40.47	45.84	41.17	53.71

***** REPORT DOES NOT INCLUDE DATA FOR THIS DEMOGRAPHIC GROUP

NEW YORK
FALL 1981

RELIABILITY-TABLE B

TYPE

HOUR ESTIMATES

MON-FRI 10A-3P MON-FRI 7P-MID	SAT-SUN 6A-MID	MON-FRI 6A-10A+ MON-FRI 3P-7P	MON-FRI 6A-7P	MON-FRI 6A-MID	MON-SUN 6A-MID
96.84	113.30	113.57	104.74	113.81	121.82
65.26	81.20	78.79	73.15	80.83	87.49
28.69	34.32	34.28	33.27	36.23	38.88
35.28	44.71	41.55	38.96	43.97	48.05
27.85	35.99	32.96	31.47	34.13	37.40
27.10	33.62	31.89	30.20	33.37	36.25
30.19	36.44	35.62	35.30	38.84	42.30
43.52	.....	.....	.....	.....	59.68
51.02	.....	.....	.....	.....	70.16
.....	.....	.....	.....	.....	.....
44.66	.....	.....	.....	.....	61.11
48.17	.....	.....	.....	.....	65.90
45.86	.....	.....	.....	.....	62.56
75.69	91.28	88.35	81.15	88.00	95.05
33.96	41.74	39.02	37.33	40.79	45.29
41.05	50.30	48.22	46.10	50.97	54.34
34.71	43.41	39.36	38.54	42.16	46.50
33.45	41.24	38.74	37.14	40.89	43.26
31.51	37.18	35.52	34.89	38.38	40.72
51.34	.....	.....	.....	.....	68.27
62.02	.....	.....	.....	.....	82.35
.....	.....	.....	.....	.....	.....
54.90	.....	.....	.....	.....	73.68
58.75	.....	.....	.....	.....	79.27
53.81	.....	.....	.....	.....	71.06
87.87	.....	.....	.....	.....	114.95
62.64	.....	.....	.....	.....	81.88
72.98	.....	.....	.....	.....	97.38
63.68	.....	.....	.....	.....	85.15
68.67	.....	.....	.....	.....	92.51
64.59	.....	.....	.....	.....	85.64
52.21	53.68	56.12	57.10	60.16	62.27
112.69	129.34	132.52	120.53	131.37	140.63
76.14	93.89	92.81	85.48	95.19	101.81
34.09	41.04	40.99	39.44	43.43	46.69
41.85	53.28	48.98	46.32	52.25	57.09
34.19	44.27	39.64	37.64	41.57	45.69
31.25	38.86	36.48	34.27	38.11	41.01
34.95	41.05	40.93	40.72	44.38	47.86
51.01	.....	.....	.....	.....	71.30
59.75	.....	.....	.....	.....	82.10
.....	.....	.....	.....	.....	.....
52.72	.....	.....	.....	.....	71.48
56.79	.....	.....	.....	.....	76.34
53.98	.....	.....	.....	.....	72.62
86.73	104.60	102.04	92.29	100.42	108.88
37.87	47.74	43.56	41.46	45.33	50.57
48.72	59.81	56.78	54.99	61.22	64.70
40.58	51.27	46.60	45.19	49.39	54.12
38.47	47.65	43.76	42.19	46.53	49.48
37.39	43.99	41.79	41.87	45.12	47.66
58.47	.....	.....	.....	.....	78.85
70.07	.....	.....	.....	.....	94.70
.....	.....	.....	.....	.....	.....
64.04	.....	.....	.....	.....	86.17
68.05	.....	.....	.....	.....	92.24
62.03	.....	.....	.....	.....	82.32
100.94	.....	.....	.....	.....	132.98
71.31	.....	.....	.....	.....	96.81
84.16	.....	.....	.....	.....	114.48
74.79	.....	.....	.....	.....	99.55
80.47	.....	.....	.....	.....	106.93
75.82	.....	.....	.....	.....	99.25
61.20	63.03	64.04	65.21	68.55	70.98

ARBITRON RADIO

SURVEY AREA/ DEMO GROUP	ESTIMATE				
		AVERAGE QUARTER			
ADI SURVEY AREA	ALL CUME ESTIMATES	SAT 6A-10A SUN 6A-10A SAT 3P-7P SUN 3P-7P	WEEKDAYS SINGLE HOUR	SAT 10A-3P SUN 10A-3P SAT 7P-MID SUN 7P-MID	MON-FRI 6A-10A MON-FRI 3P-7P
TOTAL 12 +	51.24	*****	*****	*****	*****
MEN 18 +	40.09	59.27	*****	58.16	75.57
MEN 18-24	14.27	*****	*****	*****	*****
MEN 25-34	19.59	*****	*****	*****	*****
MEN 35-44	16.26	*****	*****	*****	*****
MEN 45-54	16.01	*****	*****	*****	*****
MEN 55-64	16.09	*****	*****	*****	*****
MEN 18-34	25.35	38.80	*****	37.75	50.12
MEN 18-49	31.75	47.43	*****	46.52	59.61
MEN 25-44	*****	*****	*****	*****	*****
MEN 25-49	28.09	42.02	*****	41.74	52.63
MEN 25-54	30.75	45.37	*****	45.29	57.31
MEN 35-64	28.81	41.71	*****	42.59	54.31
WOMEN 18 +	45.29	66.48	*****	65.34	81.37
WOMEN 18-24	18.07	*****	*****	*****	*****
WOMEN 25-34	22.09	*****	*****	*****	*****
WOMEN 35-44	18.76	*****	*****	*****	*****
WOMEN 45-54	19.46	*****	*****	*****	*****
WOMEN 55-64	18.54	*****	*****	*****	*****
WOMEN 18-34	29.89	46.19	*****	44.31	58.12
WOMEN 18-49	37.79	56.62	*****	55.04	70.75
WOMEN 25-44	*****	*****	*****	*****	*****
WOMEN 25-49	33.53	49.39	*****	49.43	63.38
WOMEN 25-54	36.57	53.18	*****	53.19	68.08
WOMEN 35-64	33.77	49.26	*****	48.92	61.42
ADULTS 18 +	54.63	79.25	*****	77.94	99.48
ADULTS 18-34	36.39	55.13	*****	53.62	70.02
ADULTS 18-49	44.91	66.65	*****	64.93	83.49
ADULTS 25-49	39.53	58.10	*****	57.71	73.45
ADULTS 25-54	42.99	62.50	*****	62.39	79.46
ADULTS 35-64	40.46	57.87	*****	58.34	74.55
TEENS	23.98	38.47	*****	38.95	52.22

***** REPORT DOES NOT INCLUDE DATA FOR THIS DEMOGRAPHIC GROUP

RELIABILITY-TABLE B

TYPE

HOUR ESTIMATES

MON-FRI 10A-3P MON-FRI 7P-MID	SAT-SUN 6A-MID	MON-FRI 6A-10A+ MON-FRI 3P-7P	MON-FRI 6A-7P	MON-FRI 6A-MID	MON-SUN 6A-MID
.....	.....	.....	.....	.....	136.93
70.14	.....	.....	.....	.....	97.40
.....	.....	.....	.....	.....	43.00
.....	.....	.....	.....	.....	53.38
.....	.....	.....	.....	.....	43.69
.....	.....	.....	.....	.....	40.04
.....	.....	.....	.....	.....	45.99
46.85	.....	.....	.....	.....	.....
55.54	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
50.11	.....	.....	.....	.....	.....
54.14	.....	.....	.....	.....	.....
50.92	.....	.....	.....	.....	.....
81.47	.....	.....	.....	.....	108.23
.....	.....	.....	.....	.....	51.10
.....	.....	.....	.....	.....	61.23
.....	.....	.....	.....	.....	52.23
.....	.....	.....	.....	.....	48.61
.....	.....	.....	.....	.....	45.29
56.76	.....	.....	.....	.....	.....
68.55	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
61.16	.....	.....	.....	.....	.....
65.07	.....	.....	.....	.....	.....
60.48	.....	.....	.....	.....	.....
96.92	.....	.....	.....	.....	.....
67.25	.....	.....	.....	.....	.....
79.88	.....	.....	.....	.....	.....
71.53	.....	.....	.....	.....	.....
77.13	.....	.....	.....	.....	.....
72.81	.....	.....	.....	.....	.....
57.18	.....	.....	.....	.....	70.13

Appendix B

Per-Case Statistical Efficiency of Average Quarter Hour (AQH) Audience Estimates*

While the overall trend of the indices given in Tables 5a, 5b and 5c (See Chapter 4) clearly is increasing as the number of quarter hours included in the AQH audience estimates gets larger, the relationship is not a simple linear one — nor is it necessarily monotonic. This is due to the varying degrees of intra-person correlation in listening behavior that occurs among the different estimate types. For AQH estimates, each individual provides a cluster of "t" observations, where "t" is the number of quarter hours included in the reported time period. The Per-Case Statistical Efficiency that results from the "t" observations from a randomly selected individual may be expressed as follows:

$$\text{Per-Case Statistical Efficiency} = \frac{t}{[1 + (t - 1)\,\text{roh}]}$$

where roh is the *intraclass (intra-person) correlation* of listening behavior within the reported time period.

If the intra-person listening among the various quarter hours included in the particular AQH estimate behaves as if it were totally random, roh would be equal to zero, and the Per-Case Statistical Efficiency due to multiple observations would be equal to "t". In other words, in terms of reliability, the sample would perform as if it contained "t" times the actual in-tab sample size. However, if the intra-person listening within the reported time period were entirely homogeneous, roh would equal a value of 1.0, and therefore, the Per-Case Statistical Efficiency would be equal to 1.0. In this instance,

* See text, Chapter 4 (Section B.2.b) for further discussion of this point.

there would be no gain in overall reliability due to multiple observations on the same individual.

The data in this study show that intra-person listening behavior is not random. As might be suspected, the intra-person listening behavior tends to be moderately homogeneous within the reporting time periods, but the degree of homogeneity varies somewhat for each reporting time period. This means that the values of roh would realistically fall somewhere between 0.0 and 1.0. As a result, the Per-Case Statistical Efficiency levels fall somewhere between 1 and "t" depending upon the particular AQH estimate type. AQH audience estimates which contained 20 quarter-hour observations were classified into two separate categories (Type B and Type C) as shown in Table 2 of Chapter 4. It was felt that their Per-Case Statistical Efficiency would differ widely enough to justify their being treated as separate estimate types, because of the substantial differences in the dayparts and days included in their respective definitions. On the other hand, the subcategories included within each of the other estimate types defined in Table 2 were judged to be similar enough to be combined into one grouping. These groupings help to simplify the structure of Table B which is included in the back of each Local Market Radio Report. In turn, this simplifies the report users' task in determining standard errors.

Appendix C

Average Statistical Efficiency Values As Empirically Determined From Jackknife Replication Analysis

NOTE

The Statistical Efficiencies contained in this Appendix represent *average values* derived in this study. For each market and report period, the Statistical Efficiencies must be uniquely determined. This can be accomplished through use of the Table B values contained in the back of each individual Local Market Radio Report (pages iv and v). The relationship is as follows:

$$(\text{Statistical Efficiency})_{g,d,e} = \frac{(\text{Table B Value})^2_{g,d,e}}{(\text{In-tab Sample Size})_{g,d,e}}$$

where g = particular geographic region
d = particular demographic region
e = particular estimate type

TABLE C-1 AVERAGE* STATISTICAL EFFICIENCY VALUES AS EMPIRICALLY DETERMINED FROM JACKKNIFE REPLICATION ANALYSIS

(Metro)

Demographic Group	#1 All Cume Est.	#2 AQH-Type A (16)**	#3 AQH-Type B (20)**	#4 AQH-Type C (20)**	#5 AQH-Type D (80)**	#6 AQH-Type E (100)**	#7 AQH-Type F (144)**	#8 AQH-Type G (160)**	#9 AQH-Type H (260)**	#10 AQH-Type I (360)**	#11 AQH-Type J (504)**
	Estimate Type (See Table 2 in Chapter 4 of Text For Detailed Definitions)										
Total Persons 12+	.58	1.31	1.43	1.29	2.03	2.06	2.83	2.84	2.42	2.85	3.27
Men 18+	.76	1.75	1.81	1.66	2.78	2.47	3.82	3.60	3.10	3.79	4.44
Men 18-24	.76	2.13	***	1.91	4.04	3.53	5.06	5.07	4.76	5.65	6.51
Men 25-34	.86	2.13	***	2.06	3.56	3.24	5.21	4.51	3.97	5.05	6.03
Men 35-44	.86	2.03	2.18	2.05	3.45	2.99	4.99	4.21	3.84	4.51	5.41
Men 45-54	.85	1.86	2.03	1.96	3.22	2.88	4.45	4.01	3.59	4.39	5.17
Men 55-64	.88	2.00	***	2.06	3.59	3.24	4.73	4.53	4.46	5.39	6.38
Men 18-34	.81	1.99	2.16	1.85	3.37	2.97	***	***	***	***	5.58
Men 18-49	.78	1.88	1.96	1.77	3.03	2.55	***	***	***	***	4.81
Men 25-44	***	***	2.10	***	***	***	***	***	***	***	***
Men 25-49	.86	1.97	2.04	1.91	3.14	2.68	***	***	***	***	5.02
Men 25-54	.85	1.89	***	1.87	3.04	2.57	***	***	***	***	4.82
Men 35-64	.86	1.83	***	1.91	3.09	2.65	***	***	***	***	4.95
Women 18+	.77	1.67	1.77	1.63	2.51	2.59	3.78	3.53	2.99	3.51	4.10
Women 18-24	.82	2.14	***	1.91	3.57	3.40	5.15	4.48	4.11	4.92	6.05
Women 25-34	.86	2.08	***	2.05	3.46	3.23	4.85	4.47	4.08	4.99	5.67
Women 35-44	.85	1.99	2.24	2.03	3.37	3.14	4.93	4.07	3.90	4.66	5.66
Women 45-54	.85	1.93	2.00	1.88	3.03	2.86	4.36	3.85	3.54	4.28	4.80
Women 55-64	.90	1.88	***	1.88	3.08	3.07	4.26	3.90	3.77	4.56	5.12
Women 18-34	.84	2.00	2.18	1.87	3.20	3.01	***	***	***	***	5.32
Women 18-49	.78	1.84	1.97	1.76	2.87	2.64	***	***	***	***	4.65
Women 25-44	***	***	2.14	***	***	***	***	***	***	***	***
Women 25-49	.86	1.89	2.06	1.88	3.05	2.76	***	***	***	***	4.97
Women 25-54	.86	1.85	***	1.84	2.99	2.64	***	***	***	***	4.82
Women 35-64	.87	1.82	***	1.82	2.89	2.62	***	***	***	***	4.58
Adult 18+	.62	1.35	***	1.31	2.07	1.95	***	***	***	***	3.34
Adult 18-34	.71	1.69	***	1.58	2.75	2.57	***	***	***	***	4.39
Adult 18-49	.66	1.54	***	1.46	2.40	2.16	***	***	***	***	3.84
Adult 25-49	.71	1.59	***	1.56	2.51	2.24	***	***	***	***	4.01
Adult 25-54	.70	1.53	***	1.52	2.43	2.16	***	***	***	***	3.93
Adult 35-64	.71	1.49	***	1.52	2.45	2.22	***	***	***	***	3.90
Teens	.75	2.02	2.59	2.09	3.78	4.60	4.87	5.32	5.51	6.12	6.55

*These average values were derived from the nineteen Market Report data base included in the *Arbitron Replication II* study. These represent *conservative estimates* of the true Statistical Efficiency because of the manner in which they were derived (See Section D of Chapter7). A standard arithmetic mean of the individual Statistical Efficiencies for the 1,377,296 audience estimates examined would yield considerably higher Statistical Efficiency values.

**Number in "( )" equals number of quarter hours averaged in each AQH (Average Quarter Hour) estimate type.

***Arbitron's Local Market Radio Reports do not include audience estimates for this cell; therefore, value not available.

TABLE C-2 AVERAGE* STATISTICAL EFFICIENCY VALUES AS EMPIRICALLY DETERMINED FROM JACKKNIFE REPLICATION ANALYSIS

(TSA)

	Estimate Type (See Table 2 in Chapter 4 of Text For Detailed Definitions)										
Demographic Group	#1 All Cume Est.	#2 AQH-Type A (16)**	#3 AQH-Type B (20)**	#4 AQH-Type C (20)**	#5 AQH-Type D (80)**	#6 AQH-Type E (100)**	#7 AQH-Type F (144)**	#8 AQH-Type G (160)**	#9 AQH-Type H (260)**	#10 AQH-Type I (360)**	#11 AQH-Type J (504)**
Total Persons 12+	.54	1.16	1.27	1.12	1.81	1.88	2.48	2.60	2.15	2.55.	2.93
Men 18+	.71	1.55	1.64	1.47	2.54	2.25	3.42	3.34	2.84	3.52	4.02
Men 18-24	.77	2.03	***	1.85	3.85	3.31	4.79	4.79	4.43	5.39	6.22
Men 25-34	.82	2.05	***	1.96	3.35	3.03	4.91	4.17	3.73	4.75	5.66
Men 35-44	.83	1.93	2.04	1.92	3.14	2.82	4.73	3.83	3.45	4.21	5.08
Men 45-54	.76	1.66	1.79	1.72	2.81	2.51	3.89	3.45	3.04	3.77	4.35
Men 55-64	.82	1.81	***	1.85	3.17	2.99	4.12	4.11	4.08	4.84	5.61
Men 18-34	.77	1.88	2.03	1.76	3.20	2.74	***	***	***	***	5.35
Men 18-49	.75	1.71	1.79	1.62	2.76	2.32	***	***	***	***	4.39
Men 25-44	***	***	1.93	***	***	***	***	***	***	***	***
Men 25-49	.80	1.81	1.85	1.74	2.81	2.44	***	***	***	***	4.51
Men 25-54	.79	1.72	***	1.69	2.73	2.34	***	***	***	***	4.25
Men 35-64	.79	1.65	***	1.68	2.76	2.42	***	***	***	***	4.41
Women 18+	.70	1.49	1.56	1.41	2.22	2.31	3.36	3.20	2.62	3.10	3.65
Women 18-24	.72	1.96	***	1.67	3.09	2.88	4.57	3.81	3.46	4.13	5.15
Women 25-34	.82	1.90	***	1.91	3.27	3.02	4.55	4.11	3.86	4.78	5.34
Women 35-44	.82	1.91	2.08	1.86	3.07	2.89	4.62	3.83	3.60	4.30	5.15
Women 45-54	.78	1.78	1.83	1.74	2.72	2.62	4.06	3.41	3.17	3.86	4.37
Women 55-64	.83	1.73	***	1.73	2.83	2.83	3.92	3.55	3.57	4.14	4.60
Women 18-34	.77	1.83	1.94	1.65	2.91	2.63	***	***	***	***	4.80
Women 18-49	.74	1.67	1.74	1.53	2.56	2.29	***	***	***	***	4.18
Women 25-44	***	***	1.98	***	***	***	***	***	***	***	***
Women 25-49	.80	1.73	1.89	1.71	2.79	2.53	***	***	***	***	4.60
Women 25-54	.80	1.70	***	1.67	2.72	2.42	***	***	***	***	4.46
Women 35-64	.79	1.68	***	1.64	2.59	2.38	***	***	***	***	4.21
Adult 18+	.57	1.19	***	1.14	1.88	1.74	***	***	***	***	3.02
Adult 18-34	.67	1.55	***	1.45	2.55	2.26	***	***	***	***	4.16
Adult 18-49	.63	1.39	***	1.30	2.21	1.93	***	***	***	***	3.57
Adult 25-49	.66	1.43	***	1.39	2.28	2.05	***	***	***	***	3.65
Adult 25-54	.65	1.38	***	1.36	2.20	1.99	***	***	***	***	3.52
Adult 35-64	.65	1.34	***	1.35	2.20	2.05	***	***	***	***	3.52
Teens	.69	1.77	2.29	1.84	3.14	4.07	4.32	4.46	4.62	5.10	5.47

*These average values were derived from the nineteen Market Report data base included in the *Arbitron Replication II* study. These represent *conservative estimates* of the true Statistical Efficiency because of the manner in which they were derived (See Section D of Chapter 7). A standard arithmetic mean of the individual Statistical Efficiencies for the 1,377,296 audience estimates examined would yield considerably higher Statistical Efficiency values.

**Number in "( )" equals number of quarter hours averaged in each AQH (Average Quarter Hour) estimate type.

***Arbitron's Local Market Radio Reports do not include audience estimates for this cell; therefore, value not available.

TABLE C-3 AVERAGE* STATISTICAL EFFICIENCY VALUES AS EMPIRICALLY DETERMINED FROM JACKKNIFE REPLICATION ANALYSIS

(ADI)

	Estimate Type (See Table 2 in Chapter 4 of Text For Detailed Definitions)										
Demographic Group	#1 All Cume Est.	#2 AQH-Type A (16)**	#3 AQH-Type B (20)**	#4 AQH-Type C (20)**	#5 AQH-Type D (80)**	#6 AQH-Type E (100)**	#7 AQH-Type F (144)**	#8 AQH-Type G (160)**	#9 AQH-Type H (260)**	#10 AQH-Type I (360)**	#11 AQH-Type J (504)**
Total Persons 12+	.47	***	***	***	***	***	***	***	***	***	3.35
Men 18+	.76	1.65	***	1.59	2.69	2.32	***	***	***	***	4.47
Men 18-24	.69	***	***	***	***	***	***	***	***	***	6.23
Men 25-34	.81	***	***	***	***	***	***	***	***	***	6.03
Men 35-44	.80	***	***	***	***	***	***	***	***	***	5.74
Men 45-54	.80	***	***	***	***	***	***	***	***	***	4.97
Men 55-64	.77	***	***	***	***	***	***	***	***	***	6.29
Men 18-34	.81	1.90	***	1.80	3.18	2.78	***	***	***	***	***
Men 18-49	.79	1.77	***	1.71	2.80	2.43	***	***	***	***	***
Men 25-44	***	***	***	***	***	***	***	***	***	***	***
Men 25-49	.85	1.90	***	1.87	2.98	2.70	***	***	***	***	***
Men 25-54	.84	1.82	***	1.82	2.91	2.60	***	***	***	***	***
Men 35-64	.84	1.76	***	1.83	2.99	2.62	***	***	***	***	***
Women 18+	.76	1.64	***	1.58	2.45	2.46	***	***	***	***	4.34
Women 18-24	.77	***	***	***	***	***	***	***	***	***	6.17
Women 25-34	.76	***	***	***	***	***	***	***	***	***	5.85
Women 35-44	.75	***	***	***	***	***	***	***	***	***	5.78
Women 45-54	.82	***	***	***	***	***	***	***	***	***	5.10
Women 55-64	.87	***	***	***	***	***	***	***	***	***	5.20
Women 18-34	.83	1.98	***	1.83	3.14	2.99	***	***	***	***	***
Women 18-49	.80	1.80	***	1.70	2.81	2.64	***	***	***	***	***
Women 25-44	***	***	***	***	***	***	***	***	***	***	***
Women 25-49	.84	1.83	***	1.83	3.01	2.80	***	***	***	***	***
Women 25-54	.85	1.79	***	1.79	2.93	2.68	***	***	***	***	***
Women 35-64	.86	1.82	***	1.80	2.83	2.75	***	***	***	***	***
Adult 18+	.62	1.30	***	1.25	2.04	1.94	***	***	***	***	***
Adult 18-34	.71	1.62	***	1.54	2.62	2.42	***	***	***	***	***
Adult 18-49	.66	1.46	***	1.39	2.29	2.10	***	***	***	***	***
Adult 25-49	.70	1.51	***	1.48	2.41	2.28	***	***	***	***	***
Adult 25-54	.69	1.45	***	1.45	2.35	2.21	***	***	***	***	***
Adult 35-64	.71	1.45	***	1.47	2.40	2.29	***	***	***	***	***
Teens	.76	1.96	***	2.01	3.62	4.33	***	***	***	***	6.51

*These average values were derived from the nineteen Market Report data base included in the *Arbitron Replication II* study. These represent *conservative estimates* of the true Statistical Efficiency because of the manner in which they were derived (See Section D of Chapter 7). A standard arithmetic mean of the individual Statistical Efficiencies for the 1,377,296 audience estimates examined would yield considerably higher Statistical Efficiency values.

**Number in "( )" equals number of quarter hours averaged in each AQH (Average Quarter Hour) estimate type.

***Arbitron's Local Market Radio Reports do not include audience estimates for this cell; therefore, value not available.

Appendix D

Mathematical Properties of Jackknife Replication Estimates and Their Relationship To Simple Replication Forms

General Discussion

The method of Jackknife Replication for the estimation of standard errors was first proposed by Professor John Tukey of Princeton University. In a 1958 abstract,[1] Tukey suggested that the subsample functions originally proposed by Quenouille in 1949,[2] as a device for reducing bias in the estimation of serial correlations, could be regarded as independently and identically distributed random variables. Under these conditions, the Quenouille estimates (called "pseudo values") serve as a vehicle for determining both variance and standard error of the overall sample estimate.

We assume that a sample is selected, or may be viewed as having been selected, in the form of g independent replications of a probability sampling process.

Let R, be defined as the estimate of a population parameter P, which is based on the entire sample (i.e., all data from all g replicates or subsamples). Let R'_i, be defined as the estimate of the population parameter P, which is based on the set of observations which result by considering the entire sample and omitting all observations from the i^{th} replicate.

The i^{th} "pseudo value" $R'_{(i)}$ is defined as:

$$R'_{(i)} = gR - (g-1)\,R'_i \qquad \text{(D1)}$$

[1] Tukey, J. W., "Bias and Confidence in Not-Quite Large Samples: Abstract." *Annals of Mathematical Statistics*, 29 (1958), p. 614.

[2] Quenouille, M. H., "Problems in Plane Sampling." *Annals of Mathematical Statistics*, 20 (1949), p. 355–375. Also see Quenouille, M. H., "Notes on Bias in Estimation." *Biometrika*, 43 (1956), p. 355–360.

The mean of the g pseudo values $R'_{(i)}$ is defined as the Quenouille's estimator R_q where:

$$R_q = \frac{1}{g} \Sigma R'_{(i)} \tag{D2}$$

Tukey proposed that the variance of R_q could be estimated by $\mathrm{var}_j(R_q)$, where:

$$\mathrm{var}_j(R_q) = \frac{1}{g\,(g-1)} \Sigma (R'_{(i)} - R_q)^2 \tag{D3}$$

The standard error of R_q is simply the square root of the expression above, and can be expressed as:

$$\mathrm{ste}_j(R_q) = \sqrt{\mathrm{var}_j(R_q)} \tag{D4}$$

In actual computations it is not necessary to determine pseudo values $R'_{(i)}$ because $\mathrm{var}_j(R_q)$ may be re-expressed as:

$$\mathrm{var}_j(R_q) = \frac{(g-1)}{g} \Sigma (R'_i - \bar{R}')^2 \tag{D5}$$

where $\bar{R}'$ is defined as:

$$\bar{R}' = \frac{1}{g} \Sigma R'_i \tag{D6}$$

This computation form is obtained from the definitional form of (D3) as follows:

The estimate R_q, defined in (D2) may be re-expressed as:

$$R_q = \frac{1}{g} \Sigma R'_{(i)} \tag{D7}$$

$$= \frac{1}{g} \Sigma [gR - (g-1) R'_i] \tag{D8}$$

$$= \frac{1}{g} \Sigma gR - \frac{1}{g} \Sigma (g-1) R'_i \tag{D9}$$

$$= gR - (g-1) \bar{R}' \tag{D10}$$

From (D10) and (D1) the terms $(R'_{(i)} - R_q)$ appearing in Formula (D3) may then be written as:

$$(R'_{(i)} - R_q) = [gR - (g-1) R'_i] - [gR - (g-1) \bar{R}'] \tag{D11}$$

$$= (g-1) (\bar{R}' - R'_i) \tag{D12}$$

Thus, expression (D3) may be rewritten as:

$$\mathrm{var_j}(R_q) = \frac{1}{g\,(g-1)} \Sigma\,(R'_{(i)} - R_q)^2 \tag{D13}$$

$$= \frac{1}{g\,(g-1)} \Sigma\,[(g-1)^2\,(\bar{R}' - R'_i)^2] \tag{D14}$$

$$= \frac{(g-1)^2}{g\,(g-1)} \Sigma\,(R'_i - \bar{R}')^2 \tag{D15}$$

$$= \frac{(g-1)}{g} \Sigma\,(R'_i - \bar{R}')^2 \tag{D16}$$

The form $\mathrm{var_j}(R_q)$ may also be used to estimate the variance of R, the estimate based on the total sample. In practice, however, the variance of R is usually estimated by varjk(R) which is defined by replacing $\bar{R}'$ by R in $\mathrm{var_j}(R_q)$. That is, it yields values that are always greater than, or equal to, the values produced by $\mathrm{var_j}(R_q)$. This may be shown as follows:

$$\mathrm{varjk}(R) = \frac{(g-1)}{g} \Sigma\,(R'_{(i)} - R)^2 \tag{D17}$$

$$= \frac{(g-1)}{g} \Sigma\,[(R'_i - \bar{R}') + (\bar{R}' - R)]^2 \tag{D18}$$

$$= \frac{(g-1)}{g} \Sigma\,(R'_i - \bar{R}')^2 + \frac{(g-1)}{g} \Sigma\,(\bar{R}' - R)^2 + \frac{2(g-1)}{g} \Sigma\,(R'_i - \bar{R}')\,(\bar{R}' - R) \tag{D19}$$

The third term in (D19) is zero since $(\bar{R}_i - R)$ is a constant and by (D6) the term $(R'_i - \bar{R}')$ sums to zero. Thus, we have:

$$\mathrm{varjk}(R) = \frac{(g-1)}{g} \Sigma\,(R'_i - \bar{R}')^2 + (g-1)\,(\bar{R}' - R)^2 \tag{D20}$$

$$\mathrm{varjk}(R) = \mathrm{var_j}(R_q) + (g-1)(\bar{R}' - R)^2 \tag{D21}$$

Since the term $(\bar{R}' - R)^2$ will never be negative, we have:

$$\mathrm{varjk}(R) \geq \mathrm{var_j}(R_q) \tag{D22}$$

And, for estimates of standard error, we have:

$$\text{stejk}(R) \geq \text{ste}_j(R_q) \tag{D23}$$

Special Case of Linear Estimates

Finally, in the situation where the estimate R is a simple linear function of the Basic Replicates, the forms varjk(R) and $\text{var}_j(R_q)$ are equal to each other (as well as being equal to the Simple Replication variance).

Let R_i denote the sample estimate computed from the i^{th} replicate. The mean of the g simple replicate estimates R_i is:

$$\bar{R} = \frac{1}{g} \Sigma (R_i) \tag{D24}$$

The estimate R is defined as a simple linear function of the replicate values R_i, if R is equal to $\bar{R}$ over the entire sample space—that is if:

$$R = \bar{R}, \text{ over the entire sample space} \tag{D25}$$

Under this assumption, for all i we have:

$$R = \bar{R_i} = \frac{1}{g} R + \frac{(g-1)}{g} R'_i \tag{D26}$$

Multiplying both sides of this expression by g, we obtain for i = 1, . . ., g the following expression:

$$g\,R = R_i + (g-1)\,R'_i \tag{D27}$$

Thus,

$$R_i = -(g-1)\,R'_i + g\,R \tag{D28}$$

The simple replication variance estimate of R is defined as:

$$\text{varrep}(R) = \frac{1}{g\,(g-1)} \Sigma (R_i - R)^2 \,{}^* \tag{D29}$$

With corresponding standard error estimates as follows:

$$\text{sterep}(R) = \sqrt{\text{varrep}(R)} \tag{D30}$$

---

*Note: In D 29, we have used R rather than $\bar{R}$, since it is assumed here that $R = \bar{R}$

From (D28), varrep(R) may be written as:

$$\text{varrep}(R) = \frac{1}{g\,(g-1)} \Sigma\,(-(g-1)\,R'_i + g\,R - R)^2 \tag{D31}$$

$$= \frac{1}{g\,(g-1)} \Sigma\,(-(g-1)\,R'_i + (g-1)\,R)^2 \tag{D32}$$

$$= \frac{(g-1)^2}{g\,(g-1)} \Sigma\,(-R'_i + R)^2 \tag{D33}$$

$$= \frac{(g-1)}{g} \Sigma\,(-R'_i + R)^2 \tag{D34}$$

Noting that,

$$(-R'_i + R)^2 = (R'_i - R)^2 \tag{D35}$$

the expression given by (D34) may be written as:

$$\text{varrep}(R) = \frac{(g-1)}{g} \Sigma\,(R'_i - R)^2 \tag{D36}$$

Thus, when R is a simple linear function of the simple replicate values R_i, we have:

$$\text{varrep}(R) = \text{varjk}(R) \tag{D37}$$

and,

$$\text{sterep}(R) = \text{stejk}(R) \tag{D38}$$

Under this condition, the equality of varjk(R) and $\text{var}_j(R_q)$ follows from the equality of R and $\bar{R}'$. This equality is shown as follows:

From (D27) we may write:

$$(g-1)\,R'_i = g\,R - R_i \tag{D39}$$

Taking the sum over $i = 1, \ldots, g$ and dividing by g on both sides of (D39), we have:

$$\frac{1}{g} \Sigma\,(g-1)R'_i = \frac{1}{g} \Sigma\,gR - \frac{1}{g} \Sigma\,R_i \tag{D40}$$

$$(g-1)\bar{R}' = g\,R - \bar{R} \tag{D41}$$

By condition (D25) we may replace $\bar{R}$ by R, and thus:

$$(g-1)\bar{R}' = g\,R - R \tag{D42}$$

$$(g-1)\bar{R}' = (g-1)\,R \tag{D43}$$

$$\bar{R}' = R \tag{D44}$$

Appendix E

Average Comparison Ratio ("Nomograph" Standard Errors Divided by "Actual" Standard Errors)

TABLE E-1 AVERAGE COMPARISON RATIO*
"NOMOGRAPH" STANDARD ERRORS ÷ "ACTUAL" STANDARD ERRORS
(Metro)

	Estimate Type (See Table 2 in Chapter 4 of Text For Detailed Definitions)										
Demographic Group	#1 All Cume Est.	#2 AQH-Type A (16)**	#3 AQH-Type B (20)**	#4 AQH-Type C (20)**	#5 AQH-Type D (80)**	#6 AQH-Type E (100)**	#7 AQH-Type F (144)**	#8 AQH-Type G (160)**	#9 AQH-Type H (260)**	#10 AQH-Type I (360)**	#11 AQH-Type J (504)**
Total Persons 12+	.83	1.25	1.30	1.24	1.55	1.56	1.83	1.83	1.69	1.84	1.97
Men 18+	.95	1.44	1.47	1.40	1.82	1.71	2.13	2.07	1.92	2.12	2.29
Men 18-24	.95	1.59	***	1.51	2.19	2.05	2.46	2.46	2.38	2.60	2.78
Men 25-34	1.01	1.59	***	1.57	2.06	1.96	2.49	2.32	2.17	2.45	2.68
Men 35-44	1.01	1.56	1.61	1.57	2.03	1.89	2.44	2.24	2.14	2.32	2.54
Men 45-54	1.00	1.49	1.56	1.53	1.96	1.85	2.30	2.19	2.07	2.29	2.48
Men 55-64	1.03	1.54	***	1.57	2.07	1.97	2.38	2.33	2.30	2.54	2.76
Men 18-34	.98	1.54	1.60	1.48	2.00	1.88	***	***	***	***	2.57
Men 18-49	.96	1.50	1.53	1.45	1.89	1.74	***	***	***	***	2.39
Men 25-44	***	***	1.58	***	***	***	***	***	***	***	***
Men 25-49	1.01	1.53	1.56	1.51	1.93	1.78	***	***	***	***	2.44
Men 25-54	1.01	1.50	***	1.49	1.90	1.75	***	***	***	***	2.39
Men 35-64	1.01	1.48	***	1.51	1.92	1.78	***	***	***	***	2.42
Women 18+	.96	1.41	1.45	1.39	1.73	1.75	2.12	2.05	1.88	2.04	2.20
Women 18-24	.99	1.60	***	1.51	2.06	2.01	2.48	2.31	2.21	2.42	2.68
Women 25-34	1.01	1.57	***	1.56	2.03	1.96	2.40	2.30	2.20	2.43	2.60
Women 35-44	1.01	1.54	1.63	1.55	2.00	1.94	2.42	2.20	2.15	2.36	2.60
Women 45-54	1.01	1.52	1.55	1.50	1.90	1.85	2.28	2.14	2.05	2.26	2.39
Women 55-64	1.04	1.50	***	1.50	1.92	1.91	2.26	2.16	2.12	2.33	2.47
Women 18-34	1.00	1.54	1.61	1.49	1.95	1.89	***	***	***	***	2.51
Women 18-49	.97	1.48	1.53	1.45	1.85	1.77	***	***	***	***	2.35
Women 25-44	***	***	1.60	***	***	***	***	***	***	***	***
Women 25-49	1.01	1.50	1.57	1.50	1.90	1.81	***	***	***	***	2.43
Women 25-54	1.01	1.48	***	1.48	1.88	1.77	***	***	***	***	2.39
Women 35-64	1.02	1.47	***	1.47	1.85	1.77	***	***	***	***	2.33
Adult 18+	.86	1.26	***	1.25	1.57	1.52	***	***	***	***	1.99
Adult 18-34	.92	1.42	***	1.37	1.81	1.75	***	***	***	***	2.28
Adult 18-49	.89	1.36	***	1.32	1.69	1.60	***	***	***	***	2.13
Adult 25-49	.92	1.37	***	1.36	1.73	1.63	***	***	***	***	2.18
Adult 25-54	.91	1.35	***	1.34	1.70	1.60	***	***	***	***	2.16
Adult 35-64	.92	1.33	***	1.34	1.70	1.62	***	***	***	***	2.15
Teens	.95	1.55	1.75	1.58	2.12	2.34	2.40	2.51	2.56	2.69	2.79

*These average values were derived from the nineteen Market Report data base included in the *Arbitron Replication II* study. (See text in Chapter 4, Section B.1 for discussion of results.)

**Number in "( )" equals number of quarter hours averaged in the particular AQH (Average Quarter Hour) estimate type.

***Arbitron's Local Market Radio Reports do not include audience estimates for this cell; therefore, Comparison Ratio not available.

TABLE E-2 AVERAGE COMPARISON RATIO*
"NOMOGRAPH" STANDARD ERRORS ÷ "ACTUAL" STANDARD ERRORS
(TSA)

Demographic Group	Estimate Type (See Table 2 in Chapter 4 of Text For Detailed Definitions)										
	#1 All Cume Est.	#2 AQH-Type A (16)**	#3 AQH-Type B (20)**	#4 AQH-Type C (20)**	#5 AQH-Type D (80)**	#6 AQH-Type E (100)**	#7 AQH-Type F (144)**	#8 AQH-Type G (160)**	#9 AQH-Type H (260)**	#10 AQH-Type I (360)**	#11 AQH-Type J (504)**
Total Persons 12+	.84	1.23	1.29	1.21	1.53	1.56	1.79	1.84	1.67	1.82	1.95
Men 18+	.96	1.42	1.46	1.38	1.82	1.71	2.11	2.08	1.92	2.14	2.29
Men 18-24	1.01	1.64	***	1.56	2.25	2.09	2.51	2.51	2.41	2.66	2.86
Men 25-34	1.03	1.64	***	1.60	2.09	1.99	2.53	2.33	2.20	2.49	2.72
Men 35-44	1.04	1.59	1.64	1.59	2.02	1.92	2.49	2.23	2.12	2.34	2.57
Men 45-54	1.00	1.48	1.54	1.51	1.92	1.82	2.26	2.12	2.00	2.22	2.39
Men 55-64	1.03	1.54	***	1.56	2.04	1.98	2.33	2.32	2.31	2.52	2.71
Men 18-34	1.01	1.57	1.63	1.52	2.04	1.89	***	***	***	***	2.64
Men 18-49	.99	1.49	1.53	1.46	1.89	1.74	***	***	***	***	2.39
Men 25-44	***	***	1.59	***	***	***	***	***	***	***	***
Men 25-49	1.02	1.54	1.55	1.51	1.91	1.79	***	***	***	***	2.42
Men 25-54	1.01	1.50	***	1.48	1.88	1.75	***	***	***	***	2.35
Men 35-64	1.01	1.47	***	1.48	1.90	1.78	***	***	***	***	2.40
Women 18+	.96	1.39	1.43	1.36	1.70	1.73	2.09	2.04	1.84	2.01	2.18
Women 18-24	.97	1.60	***	1.48	2.01	1.94	2.45	2.23	2.12	2.32	2.59
Women 25-34	1.03	1.58	***	1.58	2.06	1.99	2.44	2.32	2.24	2.50	2.64
Women 35-44	1.03	1.58	1.65	1.56	2.00	1.95	2.46	2.24	2.17	2.37	2.60
Women 45-54	1.01	1.53	1.55	1.52	1.89	1.86	2.30	2.11	2.04	2.25	2.39
Women 55-64	1.04	1.51	***	1.51	1.92	1.93	2.27	2.15	2.16	2.32	2.45
Women 18-34	1.00	1.55	1.59	1.47	1.94	1.85	***	***	***	***	2.50
Women 18-49	.98	1.48	1.51	1.42	1.83	1.73	***	***	***	***	2.33
Women 25-44	***	***	1.61	***	***	***	***	***	***	***	***
Women 25-49	1.02	1.50	1.57	1.50	1.91	1.82	***	***	***	***	2.45
Women 25-54	1.02	1.49	***	1.48	1.88	1.78	***	***	***	***	2.41
Women 35-64	1.01	1.48	***	1.46	1.84	1.76	***	***	***	***	2.34
Adult 18+	.86	1.25	***	1.22	1.56	1.50	***	***	***	***	1.98
Adult 18-34	.93	1.42	***	1.38	1.82	1.71	***	***	***	***	2.33
Adult 18-49	.90	1.34	***	1.30	1.70	1.58	***	***	***	***	2.15
Adult 25-49	.93	1.37	***	1.35	1.72	1.64	***	***	***	***	2.18
Adult 25-54	.92	1.34	***	1.33	1.69	1.61	***	***	***	***	2.14
Adult 35-64	.92	1.32	***	1.33	1.69	1.63	***	***	***	***	2.14
Teens	.95	1.52	1.73	1.55	2.03	2.31	2.38	2.42	2.46	2.58	2.68

*These average values were derived from the nineteen Market Report data base included in the *Arbitron Replication II* study. (See text in Chapter 4, Section B.1 for discussion of results.)

**Number in "( )" equals number of quarter hours averaged in each AQH (Average Quarter Hour) estimate type.

***Arbitron's Local Market Radio Reports do not include audience estimates for this cell; therefore, Comparison Ratio not available.

TABLE E-3 AVERAGE COMPARISON RATIO*
"NOMOGRAPH" STANDARD ERRORS ÷ "ACTUAL" STANDARD ERRORS
(ADI)

	Estimate Type (See Table 2 in Chapter 4 of Text For Detailed Definitions)										
Demographic Group	**#1 All Cume Est.**	**#2 AQH-Type A (16)****	**#3 AQH-Type B (20)****	**#4 AQH-Type C (20)****	**#5 AQH-Type D (80)****	**#6 AQH-Type E (100)****	**#7 AQH-Type F (144)****	**#8 AQH-Type G (160)****	**#9 AQH-Type H (260)****	**#10 AQH-Type I (360)****	**#11 AQH-Type J (504)****
Total Persons 12+	.77	•••	•••	•••	•••	•••	•••	•••	•••	•••	2.05
Men 18+	.97	1.44	•••	1.41	1.84	1.70	•••	•••	•••	•••	2.37
Men 18-24	.93	•••	•••	•••	•••	•••	•••	•••	•••	•••	2.79
Men 25-34	1.01	•••	•••	•••	•••	•••	•••	•••	•••	•••	2.75
Men 35-44	1.00	•••	•••	•••	•••	•••	•••	•••	•••	•••	2.68
Men 45-54	1.00	•••	•••	•••	•••	•••	•••	•••	•••	•••	2.50
Men 55-64	.98	•••	•••	•••	•••	•••	•••	•••	•••	•••	2.81
Men 18-34	1.01	1.54	•••	1.50	1.99	1.87	•••	•••	•••	•••	•••
Men 18-49	1.00	1.49	•••	1.46	1.87	1.75	•••	•••	•••	•••	•••
Men 25-44	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
Men 25-49	1.03	1.54	•••	1.53	1.93	1.84	•••	•••	•••	•••	•••
Men 25-54	1.02	1.51	•••	1.51	1.91	1.80	•••	•••	•••	•••	•••
Men 35-64	1.03	1.49	•••	1.52	1.94	1.81	•••	•••	•••	•••	•••
Women 18+	.98	1.43	•••	1.41	1.75	1.76	•••	•••	•••	•••	2.33
Women 18-24	.98	•••	•••	•••	•••	•••	•••	•••	•••	•••	2.78
Women 25-34	.98	•••	•••	•••	•••	•••	•••	•••	•••	•••	2.71
Women 35-44	.97	•••	•••	•••	•••	•••	•••	•••	•••	•••	2.69
Women 45-54	1.01	•••	•••	•••	•••	•••	•••	•••	•••	•••	2.53
Women 55-64	1.05	•••	•••	•••	•••	•••	•••	•••	•••	•••	2.55
Women 18-34	1.02	1.58	•••	1.51	1.98	1.94	•••	•••	•••	•••	•••
Women 18-49	1.00	1.50	•••	1.46	1.88	1.82	•••	•••	•••	•••	•••
Women 25-44	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
Women 25-49	1.03	1.51	•••	1.51	1.94	1.87	•••	•••	•••	•••	•••
Women 25-54	1.03	1.50	•••	1.50	1.92	1.83	•••	•••	•••	•••	•••
Women 35-64	1.04	1.51	•••	1.50	1.88	1.85	•••	•••	•••	•••	•••
Adult 18+	.88	1.27	•••	1.25	1.60	1.56	•••	•••	•••	•••	•••
Adult 18-34	.94	1.43	•••	1.39	1.81	1.74	•••	•••	•••	•••	•••
Adult 18-49	.91	1.35	•••	1.32	1.70	1.62	•••	•••	•••	•••	
Adult 25-49	.93	1.37	•••	1.36	1.74	1.69	•••	•••	•••	•••	•••
Adult 25-54	.93	1.35	•••	1.35	1.72	1.67	•••	•••	•••	•••	•••
Adult 35-64	.94	1.35	•••	1.36	1.74	1.69	•••	•••	•••	•••	•••
Teens	.98	1.57	•••	1.59	2.13	2.33	•••	•••	•••	•••	2.85

*These average values were derived from the nineteen Market Report data base included in the *Arbitron Replication II* study. (See text in Chapter 4, Section B.1 for discussion of results.)

**Number in "( )" equals number of quarter hours averaged in each AQH (Average Quarter Hour) estimate type.

***Arbitron's Local Market Radio Reports do not include audience estimates for this cell; therefore, Comparison Ratio not available.

Appendix F

Approximating the Standard Error of A Rating or Projection Number From an Earlier Survey Report Issued Prior to the New Reliability Procedure

Survey Reports issued prior to Fall 1981 do not contain the Tables A and B that are used in conjunction with the new Arbitron Reliability Method. The following procedure can be used to obtain estimates for the Standard Error of a rating or a projection number in such a situation. The procedure assumes that the Statistical Efficiency Values of the earlier report are the same as those of the Fall 1981 Survey Report for the comparable estimate (i.e., same estimate type, and same demographic and geographic group). This is demonstrated via an example.

Situation

A researcher is making certain historical comparisons and finds out that he needs the Standard Error for a Metro Men 25–34, Mon-Fri 6–10AM Cume audience rating published in the Spring 1981 Report. That particular report does not contain the Tables A and B of the new reliability procedure.

Step 1. Assemble The Required Basic Data

Data Item	Data Value	Data Source
Spring 1981 Rating	1.3	From Spring 1981 Arbitron Survey Report
Spring 1981 In-Tab	920	From Spring 1981 Arbitron Survey Report*
Fall 1981 In-Tab	940	From Fall 1981 Arbitron Survey Report*
Fall 1981 Table B Value	30.11	From Fall 1981 Arbitron Survey Report

*Obtained as: (Metro Total In-Tab Diaries) ×

$$\frac{\text{(Men 25–34 Metro Percent of Unweighted In-Tab Sample)}}{100}$$

Step 2. Actual Calculations

The procedure consists of obtaining the *Statistical Efficiency* (for a Cume Mon-Fri 6–10AM audience estimate among Metro Men 25–34) from the Fall 1981 Survey Report and using it to obtain an estimate of the Table B value for the Spring 1981 Survey Report for the corresponding estimate type, demographic and geographic group.

a. Stat. Eff. for Fall 1981 Report

$$= \frac{\text{(Fall 1981 Effective Sample Base)}}{\text{(Fall 1981 In-Tab)}}$$

$$= \frac{\text{(Fall 1981 Table B Value)}^2}{\text{(Fall 1981 In-Tab)}}$$

$$= \frac{(30.11)^2}{940}$$

$$= \frac{906.6121}{940}$$

$$= 0.964481$$

b. Estimated Spring 1981 Table B Value (for a Cume Mon-Fri 6–10AM audience estimate among Metro Men 25–34)

$$= \sqrt{\text{(Estimated Spring 1981 Effective Sample Base)}}$$

$$= \sqrt{\text{(Spring 1981 In-Tab)} \times \text{(Fall 1981 Stat. Eff.)}}$$

$$= \sqrt{(920) \times (0.964481)}$$

$$= \sqrt{869.2252}$$

$$= 29.48 \text{ (rounded to 2 decimal places)}$$

c. Table A Value corresponding to Spring 1981 Rating of 1.3

$$= 11.33$$

d. Standard Error for Spring 1981 Rating

$$= \frac{\text{(Table A Value)}}{\text{(Estimated Spring 1981 Table B Value)}}$$

$$= \frac{11.33}{29.48}$$

$$= 0.3843 = 0.4 \text{ (rounded to 1 decimal place)}$$

NOTE—After the Standard Error for the Spring 1981 Rating is determined as shown above, if desired, the Standard Error for the corresponding projection number can be obtained by using the procedure already described. (See example given in the instructions in Appendix A).

Appendix G

Computer Programs for "Posterior Odds Assessment"

Program #1: For Rating Data
Program #2: For Audience Projection Data

NOTE

The program that follows is written to *Microsoft BASIC–80.** It can easily be adapted to other versions of BASIC.

SOURCE LISTING FOR PROGRAM #1: FOR RATINGS DATA

```
10 REM ***********************************************************************
20 REM  PROGRAM TO COMPUTE ODDS THAT A REAL AUDIENCE CHANGE OCCURRED
30 REM  IN THE DIRECTION INDICATED BY THE REPORT TO REPORT RATING
40 REM  CHANGE.
50 REM
60 REM      THIS PROGRAM IS WRITTEN IN MICROSOFT BASIC-80
70 REM
80 REM      FOR SOME MACHINES IT WILL BE NECESSARY TO REMOVE
90 REM      THE PRINT USING STATEMENTS AND REPLACE THEM WITH
100 REM     SIMPLE PRINT STATEMENTS.  IT MAY ALSO BE NECESSARY TO CHANGE
110 REM     VARIABLE NAMES SO THAT NAME CONTAINS FEWER CHARACTERS.
120 REM
130 REM     FOR MORE INFORMATION CONTACT ARBITRON STATISTICAL SERVICES
140 REM
150 REM     NOTE: THIS PROGRAM ONLY REQUIRES TABLE B VALUES
160 REM     IT AUTOMATICALLY COMPUTES ALL REQUIRED STANDARD ERRORS
170 REM
180 REM **********************************************************************
190 DEFDBL A-H,O-Z
200 PRINT
210 INPUT "ENTER LAST TIME REPORT RATING AND TABLE B VALUE";RAT1,TABB1
211 IF RAT1<.0001 OR RAT1>99.9999 THEN 580
220 PRINT
230 INPUT "ENTER THIS TIME REPORT RATING AND TABLE B VALUE";RAT2,TABB2
231 IF RAT2<.0001 OR RAT2>99.9999 THEN 580
240 PRINT:PRINT
260 MU=RAT2-RAT1
270 SIGSQ1=((100#-RAT1)*RAT1)/(TABB1*TABB1)
280 SIGSQ2=((100#-RAT2)*RAT2)/(TABB2*TABB2)
290 SIGMA=SQR(SIGSQ1+SIGSQ2)
300 Z1=MU/SIGMA
310 XX=Z1
320 GOSUB 600
330 P1=P
340 P3=1#-P1
350 IF MU<0# THEN  A$="DOWN":B$=" UP "
360 IF MU>0#  THEN A$=" UP ":B$="DOWN"
370 IF ABS(MU)<.00001 THEN GOTO 550
380 PRINT "YOUR REPORTED RATING WENT ";A$
390 IF P1<P3 THEN ODDS = P3/P1
400 IF P1>P3 THEN ODDS = P1/P3
410 PRINT "ODDS THAT REAL CHANGE OCCURRED IN THIS DIRECTION ARE ";
420 PRINT USING "#####.#";ODDS;
430 PRINT " TO 1.0"
440 PRINT
450 REM FLIP P1,P3 IF NECESSARY
460 IF P1>=P3 THEN 480
470 P2=P1:P1=P3:P3=P2
480 PRINT "PROBABILITY THAT AUDIENCE REALLY WENT ";A$;" IS = ";
490 PRINT USING "#.####";P1
500 PRINT "PROBABILITY THAT AUDIENCE REALLY WENT ";B$;" IS = ";
510 PRINT USING "#.####";P3
520 PRINT:PRINT
530 GOTO 200
540 PRINT
550 PRINT "NO CHANGE IN REPORT RATINGS;  "
551 PRINT "THEREFORE, ODDS OF ACTUAL INCREASE VERSUS DECREASE ARE 1:1."
560 PRINT:PRINT
570 GOTO 200
580 PRINT "IMPOSSIBLE VALUE ENTERED: RATINGS MUST BE BETWEEN 0 AND 100."
590 GOTO 200
600 REM SUBROUTINE FOR NORMAL AREA UPTO X
610 AXX=XX: IF AXX<0# THEN AXX=AXX*(-1#)
620 TT=1#/(1#+(.231642*AXX))
630 DD=.398942*EXP(-XX*XX/2#)
640 P=1#-DD*TT*((((1.330274429#*TT-1.821255978#)*TT+1.781477937#)*TT-.356563782#)*TT+.31938153#)
650 IF XX<0# THEN P=1#-P
660 RETURN
```

SAMPLE RUN FOR PROGRAM #1: FOR RATINGS DATA

```
RUN

ENTER LAST TIME REPORT RATING AND TABLE B VALUE? .4,80.10

ENTER THIS TIME REPORT RATING AND TABLE B VALUE? .7,80.20

YOUR REPORTED RATING WENT  UP
ODDS THAT REAL CHANGE OCCURRED IN THIS DIRECTION ARE    92.2 TO 1.0

PROBABILITY THAT AUDIENCE REALLY WENT  UP  IS = 0.9893
PROBABILITY THAT AUDIENCE REALLY WENT DOWN IS = 0.0107


ENTER LAST TIME REPORT RATING AND TABLE B VALUE?
```

SOURCE LISTING FOR PROGRAM #2: FOR AUDIENCE PROJECTION DATA

```
10 REM ****************************************************************
20 REM  PROGRAM TO COMPUTE ODDS THAT A REAL AUDIENCE CHANGE OCCURRED
30 REM  IN THE DIRECTION INDICATED BY THE REPORT TO REPORT AUDIENCE
40 REM  PROJECTION CHANGE.
50 REM
60 REM      THIS PROGRAM IS WRITTEN IN MICROSOFT BASIC-80
70 REM
80 REM      FOR SOME MACHINES IT WILL BE NECESSARY TO REMOVE
90 REM      THE PRINT USING STATEMENTS AND REPLACE THEM WITH
100 REM     SIMPLE PRINT STATEMENTS.  IT MAY ALSO BE NECESSARY TO CHANGE
110 REM     VARIABLE NAMES SO THAT NAME CONTAINS FEWER CHARACTERS.
120 REM
130 REM     FOR MORE INFORMATION CONTACT ARBITRON STATISTICAL SERVICES
140 REM
150 REM     NOTE: THIS PROGRAM ONLY REQUIRES TABLE B VALUES
160 REM     IT AUTOMATICALLY COMPUTES ALL REQUIRED STANDARD ERRORS
170 REM
180 REM ****************************************************************
190 DEFDBL A-H,O-Z
200 PRINT
210 INPUT "ENTER LAST TIME REPORT AUDIENCE PROJECTION AND TABLE B VALUE";AUD1,TABB1
211 INPUT "ENTER LAST TIME TOTAL POPULATION";POP1
212 RAT1=100#*AUD1/POP1
213 IF AUD1<1 OR AUD1>=POP1 THEN 580
220 PRINT
230 INPUT "ENTER THIS TIME REPORT AUDIENCE PROJECTION AND TABLE B VALUE";AUD2,TABB2
231 INPUT "ENTER LAST TIME TOTAL POPULATION";POP2
232 RAT2=100#*AUD2/POP2
233 IF AUD2<1 OR AUD2>=POP2 THEN 580
240 PRINT:PRINT
260 MU=RAT2-RAT1
270 SIGSQ1=((100#-RAT1)*RAT1)/(TABB1*TABB1)
280 SIGSQ2=((100#-RAT2)*RAT2)/(TABB2*TABB2)
290 SIGMA=SQR(SIGSQ1+SIGSQ2)
300 Z1=MU/SIGMA
310 XX=Z1
320 GOSUB 600
330 P1=P
340 P3=1#-P1
350 IF MU<0# THEN  A$="DOWN":B$=" UP "
360 IF MU>0#  THEN A$=" UP ":B$="DOWN"
370 IF ABS(MU)<.00001 THEN GOTO 550
380 PRINT "YOUR REPORTED AUDIENCE WENT ";A$
390 IF P1<P3 THEN ODDS = P3/P1
400 IF P1>P3 THEN ODDS = P1/P3
410 PRINT "ODDS THAT REAL CHANGE OCCURRED IN THIS DIRECTION ARE ";
420 PRINT USING "#####.#";ODDS;
430 PRINT " TO 1.0"
440 PRINT
450 REM FLIP P1,P3 IF NECESSARY
460 IF P1>=P3 THEN 480
470 P2=P1:P1=P3:P3=P2
480 PRINT "PROBABILITY THAT AUDIENCE REALLY WENT ";A$;" IS = ";
490 PRINT USING "#.####";P1
500 PRINT "PROBABILITY THAT AUDIENCE REALLY WENT ";B$;" IS = ";
510 PRINT USING "#.####";P3
520 PRINT:PRINT
530 GOTO 200
540 PRINT
550 PRINT "NO CHANGE IN REPORT AUDIENCE PROJECTIONS: THEREFORE,"
551 PRINT "ODDS OF ACTUAL INCREASE OR DECREASE ARE 1:1."
560 PRINT:PRINT
570 GOTO 200
580 PRINT "IMPOSSIBLE VALUE ENTERED:"
581 PRINT "AUDIENCE PROJECTION MUST BE BETWEEN 0 AND TOTAL POPULATION"
590 GOTO 200
600 REM SUBROUTINE FOR NORMAL AREA UPTO X
610 AXX=XX: IF AXX<0# THEN AXX=AXX*(-1#)
620 TT=1#/(1#+(.231642*AXX))
630 DD=.398942*EXP(-XX*XX/2#)
640 P=1#-DD*TT*((((1.330274429#*TT-1.821255978#)*TT+1.781477937#)*TT-.356563782#)*TT+.31938153#)
650 IF XX<0# THEN P=1#-P
660 RETURN
```

SAMPLE RUN FOR PROGRAM #2: FOR AUDIENCE PROJECTION DATA

```
RUN

ENTER LAST TIME REPORT AUDIENCE PROJECTION AND TABLE B VALUE? 150000,17.85
ENTER LAST TIME TOTAL POPULATION? 1000000

ENTER THIS TIME REPORT AUDIENCE PROJECTION AND TABLE B VALUE? 140000,18.76
ENTER LAST TIME TOTAL POPULATION? 1000000


YOUR REPORTED AUDIENCE WENT DOWN
ODDS THAT REAL CHANGE OCCURRED IN THIS DIRECTION ARE     1.8 TO 1.0

PROBABILITY THAT AUDIENCE REALLY WENT DOWN IS = 0.6432
PROBABILITY THAT AUDIENCE REALLY WENT  UP  IS = 0.3568


ENTER LAST TIME REPORT AUDIENCE PROJECTION AND TABLE B VALUE?
```

(<u>NOTE</u>: The above odds and probability results differ slightly from those reported on page 79 in the main text. The differences are due to rounding -- the computer program carries the intermediate computations to more decimal points).

Appendix H

Glossary of Terms*

Area of Dominant Influence (ADI)—Arbitron Television's exclusive geographic area which defines each television market as a collection of counties on the basis of measurable viewing patterns. Every county in the United States (exclusive of Alaska and Hawaii) is allocated exclusively to one ADI.

Average Quarter-Hour Estimates—There are two types used in this report:

> **Average Quarter-Hour Persons**—This can be calculated by first tabulating the number of persons who listen to a station for at least five minutes in each quarter hour of a daypart; then sum and divide the result by the number of quarter-hours the station was on the air in that daypart.
>
> **Average Quarter-Hour Rating**—The Average Quarter-Hour Persons estimate expressed as a percentage of the total number of persons in the particular group and geographic area being reported.

Common Design Factor—See Design Factor.

Confidence Interval—Arbitron uses a sample of respondents instead of surveying the entire population frame, and the audience estimates so derived are subject to sampling error. The Confidence Interval describes, at a pre-chosen Confidence Level, the size of the possible error range associated with any sample statistic such as an audience

---

* For additional information, the reader is directed to "Standard Definitions of Broadcast Research Terms," published by the National Association of Broadcasters, 1771 N. Street, N.W., Washington, DC 20036. See also "Glossary of Selected Arbitron Terms" included in every Arbitron Radio Market Report.

rating or projection. The Confidence Interval is given in terms of lower and upper confidence limits. The computations require the sample statistic, its associated standard error, and also a multiplier or Z value corresponding to the pre-chosen Confidence Interval. Then

$$\begin{array}{c}\text{Lower}\\ \text{Confidence}\\ \text{Limit}\end{array} = \begin{array}{c}\text{Sample}\\ \text{Statistic}\end{array} - (\text{Z value} \times \text{standard error})$$

$$\begin{array}{c}\text{Upper}\\ \text{Confidence}\\ \text{Limit}\end{array} = \begin{array}{c}\text{Sample}\\ \text{Statistic}\end{array} + (\text{Z value} \times \text{standard error})$$

The Confidence Interval is interpreted in the following way: if the survey had been based upon the population frame rather than a sample, then the result would fall within the Confidence Interval with probability equal to the Confidence Level.

Confidence Level—See Confidence Interval.

Confidence Limits—See Confidence Interval.

Cume Estimates—There are two types referred to in this report:

Cume Persons—The estimated number of different persons who listened to a particular station for a minimum of five minutes in any quarter-hour within the daypart defined by the reporting period.

Cume Ratings—The estimated number of Cume Persons expressed as a percentage of the total number of persons in the particular demographic group and geographic area being reported.

Daypart—A given part of a day (e.g., 6–10AM, 7PM–Midnight).

Design Effect—This is related to the Design Factor by the following equation:

$$\text{Design Effect} = (\text{Design Factor})^2$$

See Design Factor and also Statistical Efficiency.

Design Factor—In analyzing standard errors derived from complex surveys, the following decomposition is often convenient:

$$\begin{array}{l}\text{Standard Error}\\ \text{Associated with}\\ \text{Actual Design}\end{array} = \begin{array}{c}\text{Standard Error}\\ \text{of Simple Random}\\ \text{Sample Design}\end{array} \times \begin{array}{c}\text{Design}\\ \text{Factor}\end{array}$$

The Design Factor is, therefore, a correction factor used to adjust for differences between the actual sample design and a Simple Random Sample design due to factors such as pre-stratification, post-stratification ("weighting" or "sample balancing"), sample clustering and random data errors. Arbitron's new Radio Reliability Model defines the Design Factor as the product of two elements:

- The Unique Design Factor, which reflects the pre-stratification and post-stratification that are unique to each market
- The Common Design Factor, which reflects general phenomena such as the clustering of listening behavior among and within individuals that are common across all markets.

Diary — A survey measuring instrument in which a respondent records his or her radio listening activity. Arbitron uses one-week individual diaries to gather listening information. Bilingual (Spanish-English) diaries are provided to all survey participants in a High Density Hispanic Area. In addition, bilingual diaries (Spanish-English) are sometimes given where a need is apparent.

Effective Sample Bases (ESB) — Estimates of Effective Sample Bases indicate the size of a simple random sample that would be required to provide the same degree of reliability (amount of sampling error) as the sample for a complex survey such as Arbitron's.

For example, in a given situation a 1% sampling error from an Arbitron survey with an in-tab sample of 1000, might require a simple random sample of 2000 to achieve the same degree of realibility. In this case, Arbitron's 1000 in-tab sample would be equal to an ESB value of 2000.

Ethnic — As employed by Arbitron, a reference to the Black and Hispanic segments of the population.

Expanded Sample Frame (ESF) — This technique is employed in certain markets in order to include in Arbitron's surveys households with unlisted telephones. The sampling universe for the ESF sample is a list of potential telephone numbers from which known listed numbers, known business numbers, non-residential exchanges, unassigned number blocks have been eliminated. From this list, supplied by Metromail, a sample of households is drawn through the use of a systematic interval technique. By Spring 1982, the ESF technique was employed in all markets with the exception of one market (McAllen-Brownsville, TX) where all information is gathered by personal diary placement.

Gross Rating Points (GRP's) — The sum of all ratings points achieved for a particular time span or spot schedule.

High-Density Black Area (HDBA) — In general, an area composed of a heavy concentration of Black households. (See Arbitron's "1981–1982 Radio Ethnic Control Directory" for further details.)

High-Density Hispanic Area (HDHA) — In general, an area composed of a heavy concentration of Hispanic households. (See Arbitron's "1981–1982 Radio Ethnic Control Directory" for further details.)

In-Tab Sample — The number of usable diaries returned and actually tabulated in producing a report.

Jackknife Replication — A specific form of Replication used to determine standard errors. See Chapter 3, Section B, *and* Chapters 6 and 7.

Metro Survey Area (MSA) — Metro Survey Areas generally correspond to Standard Metropolitan Statistical Areas (SMSA's) as defined by the U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, subject to exceptions dictated by historical industry usage and other marketing considerations.

Nomograph Procedure — Prior to the Fall 1981 Local Market Reports, standard errors of audience estimates were supplied by means of a Nomograph in each report. See Chapter 4, Footnote 7 for more details.

Population — Estimates are based upon both U.S. Census and independent sources, and are obtained for key demographic subgroups (e.g. age-sex) in a geographic survey area. Population is also referred to as Universe.

Projection — The general definition would be the total number of persons within a survey area who listen to a specific radio station. It is related to the rating as follows:

$$\text{Projection} = (\text{Population of Survey Area}) \times \frac{\text{Rating}}{100}$$

Also known as Projected Audience or Audience Projection, or Audience Projection Number. More specific usage is in terms of Average

Quarter-Hour Projection and Cume Projection. (See definitions under Average Quarter-Hour and Cume estimates).

Rating—The general definition would be the percent of all people within a survey area who listen to a specific radio station:

$$\text{Rating } (\%) = \frac{\text{Listeners to a Specific Station}}{\text{Population of Survey Area}} \times 100$$

More specific usage is in terms of Average Quarter-Hour Rating and Cume Ratings (See definitions under Average Quarter-Hour and Cume estimates.)

Reliability—A statement regarding the margin of "sampling error" of audience estimates. Estimates having smaller "sampling errors" may be thought of as having higher reliability (and vice versa). Thus, if the statement were made that "Arbitron's Radio estimates have higher reliability", it would be the same as saying that "Arbitron Radio estimates have smaller sampling errors."

Replication—A method for determining standard errors. See Chapter 3, Section B, *and* Chapters 6 and 7.

Sampling Error—Sample survey estimates are subject to inherent error due to the fact that different samples will tend to generate somewhat different results. The margin of possible error due to this factor is commonly referred to as Sampling Error. In non-statistical language, this is sometimes known as "statistical bounce". (Also see Reliability.)

Sampling Unit—A sampling unit normally is one county, although some counties have been divided into two or more sampling units because of population distribution, terrain or special interviewing technique areas.

Simple Random Sample—A Simple Random Sample of size n from a population frame of size N is one in which each possible combination of n individuals out of N has the same probability of being selected for the sample.

Simple Replication—A specific form of replication used to determine standard errors. See Chapter 3, Section B, *and* Chapters 6 and 7.

Standard Error—The reliability (sampling error) of an estimate is measured numerically by the statistical quantity called Standard Error. For a rating, R, based on a Simple Random Sample, the standard error is given by

$$\sqrt{\frac{R(100-R)}{\text{In-Tab (Size)}}}$$

For complex surveys this formula does not apply (see Design Factor). The Standard Error can be used to calculate Confidence Intervals. For example, a 68% Confidence Interval is plus and minus one unit (sigma) of the Standard Error from the sample estimate.

Statistical Efficiency—This reliability measure is defined by

$$\text{Statistical Efficiency} = \frac{\text{Effective Sample Base}}{\text{Actual Sample Size (In-tab)}}$$

For example if the ESB value is 1,400 and the actual sample size (in-tab) is 1,600, the Statistical Efficiency is 0.875 (1400 ÷ 1600)—or 87.5%, when expressed as a percentage. This measure reflects the relative 'efficiency' of the actual complex survey, relative to the benchmark of a Simple Random Sample. This concept is related to Design Factor and Design Effect by:

$$\text{Statistical Efficiency} = 1/(\text{Design Factor})^2 = 1/(\text{Design Effect})$$

Total Survey Area (TSA)—Where applicable, a geographic area that includes the Metro Survey Area plus certain counties located outside the MSA.

The first time a market area is surveyed by Arbitron, a Total Survey Area is designated from an analysis of diary data available from previous surveys in adjacent markets. A county is included or excluded on the basis of listening data from these previous surveys. The procedure for evaluation of listening records from previous surveys, for the purpose of initial market definition is the same as the procedure for updating the Total Survey Area definitions of previously surveyed markets. The criteria for inclusion of a county are based on specific numbers of mentions to the home station(s) in all diaries in-tab from the county under consideration. For purposes of these tests "mentions" is defined as the number of different diaries having entries of five or more minutes of listening within a single quarter-hour, at any time during a survey week. Also, to qualify a county for inclusion, the ratio of the cumulative mentions to home station(s) expressed as a percent of all station mentions from the county, must equal or exceed ten percent (10%). Additionally, the home station(s) cumulatively must achieve at least ten mentions. A county will be excluded if either of these conditions is not met. Updates are performed periodically.

Unique Design Factor—See Design Factor.

Universe—The estimated number of persons in a geographic survey area. Estimates are based upon both U.S. Census and independent sources, and are obtained for key demographic subgroups (e.g. age-sex) in a geographic survey area. Universe is also referred to as Population.

DR. MARTIN R. FRANKEL, a statistical consultant to Arbitron Ratings, is Professor of Statistics at Baruch College, City University of New York. Professor Frankel holds a Ph.D. from the University of Michigan and is the author of two books on the statistical reliability and validity of survey samples. His research articles on these subjects have appeared in the Journal of Marketing Research, the Journal of the American Statistical Association and the Journal of the Royal Statistical Society.

Over the past fifteen years, Dr. Frankel has been engaged by various corporations, trade associations, universities and governmental bureaus as a statistical consultant. Since 1971, he has been involved heavily in audience measurement research for both the broadcast and print media.

Dr. Frankel is a Fellow of the American Statistical Association and has served as Chairman of the Association's Section on Survey Research Methods and as Chairman of the American Statistical Association's Advisory Committee to the U.S. Census. He is a member of the editorial boards of the Public Opinion Quarterly, Sociological Research and Methods and the Encyclopedia of Statistical Sciences.