A Comparison of the New Ratings Reliability Estimator and the Previous Reliability Model

Overview

Since it is impractical to survey the entire population, we use samples to estimate the ratings. The Ratings Reliability Estimator (RRE) is an updated tool designed to help clients determine the reliability of the reported estimates. It provides reliability estimates (standard errors and confidence intervals) to help clients understand the potential range of ratings we might have derived had we conducted many surveys with the same sample size. This document describes the enhancements to the updated tool and outlines the differences between it and the previous version.

Why the New Ratings Reliability Estimator?

The previous tool (sometimes called the A/B model) provided reliability estimates only for those estimates that appeared in Arbitron ebook℠. Many clients use software services from Arbitron that report estimates beyond those in Arbitron ebook. We created the new reliability tool (called the Ratings Reliability Estimator or RRE) to estimate the reliability for any reported audience estimate including custom dayparts, custom demographics and custom geographies. The RRE is based on more updated statistical models and it is designed to yield more precise estimates of reliability than the A/B model.

Key Components of Radio Reliability Estimator

This section describes the key components of the Radio Reliability Estimator and how they may differ from the previous reliability model.

In-Tab

Sample size is a key input in estimating audience estimate reliability in both the RRE and the previous A/B method. However, the RRE makes a distinction between Diary and PPM™ in-tab effects while the A/B method treated both the same. The RRE PPM model is more refined because it includes factors for average in-tab (both daily or weekly) and unique in-tab for the entire survey, while the A/B method used only a single factor. This refinement provides more precise estimates of reliability for PPM-based ratings. The Diary in-tab factor used in the RRE model is similar to the A/B method.

Variation of the Weights

Reliability is also impacted by the extent and variation of weighting. Less weighting yields more reliable estimates and more weighting diminishes the reliability of the estimates. The RRE (like the A/B method) uses the relative variance of the weights as a factor in the reliability model.

Demographic Clustering Within Households

Audience estimates from households comprised of two or more generations of people who tune to a wider variety of stations will be more reliable than estimates from homes with people of similar age and listening characteristics. The new RRE model uses a new input to determine

continued...

the extent of demographic clustering within households that enables it to be used for any reported daypart/demo/geography, while the factor used in the A/B table limited it to only standard demos, dayparts and geographies.

Repeated Measurements Effect

More observations from the same people yield more reliable audience estimates. Audience estimates for broader dayparts are more reliable than narrower dayparts since the broader time periods have more quarter-hours to observe listening behavior. The RRE is more sensitive to differences in daypart quarter-hours than the previous A/B method. The number of quarter-hours per week in the daypart is multiplied by the number of weeks of measurement for each respondent to reflect the daypart effect in the RRE. Using the number of weeks as an input for the model reflects a difference between the PPM and Diary methodologies. For Diary reliability measures, each Diary respondent contributes only one week of listening data to the estimates. For PPM, factoring the number of weeks in the reporting period into the model reflects the fact that PPM panelists can contribute more than one week of exposure data to the estimates. The A/B method did not factor in the number of weeks and it was limited to Arbitron ebook demos/dayparts only.

Variation and Clustering in Listening Across Respondents

The RRE is sensitive to differences in reliability driven by the concentration or dispersion of listening. All else being equal, a given AQH rating will be more reliable if it is based on a wider number of different listeners and less reliable if it comes from a smaller and more concentrated number of listeners. Therefore, the RRE may produce two separate estimates of reliability for two stations with the same AQH rating. Also, estimates that are based on a given number of listeners within a small set of households are generally less reliable than those based on the same number of listeners spread over many households. The RRE method captures these factors and the A/B method did not. Note: This factor applies only to AQH reliability measures.

Asymmetric Confidence Intervals for Smaller Ratings

A confidence interval is an estimate of the range reliability around a published estimate for a given level of statistical precision. For example, a reliability estimate for a published rating of 1.0 might have a confidence interval of 0.8 to 1.2 with 90% confidence. In other words, if we survey the entire population under the same circumstances we could be 90% confident that the rating would fall between a range of a 0.8 rating and a 1.2 rating. Most clients are used to seeing this expressed as the published rating +/- a given number (sometimes called margin of error) with the high and low end of the range being the same distance from the published estimate. In this example, the width of the confidence interval is 0.4 and the distance from the published estimate for both the upper end and the lower end of the interval is 0.2. This pattern of equal distribution of margin of error around the published estimate occurs in most cases with AQH ratings. However, as ratings get smaller (e.g., a 0.2 AQH rating), the RRE may estimate the upper end of the confidence interval to be higher than the published rating plus the margin of error and the lower end may be less than the rating minus the margin of error. This refinement will produce confidence intervals that are more likely to reflect the true range of potential sampling error for smaller audience estimates.

PPM ratings are based on audience estimates and are the opinion of Arbitron and should not be relied on for precise accuracy or precise representativeness of a demographic or radio market.

ARBITRON

24/7 Customer Service & Support (866) 776-8300 • www.ArbitronTraining.com • www.arbitron.com

Headquarters	Atlanta	Chicago	Dallas	Los Angeles	New York
(410) 312-8000	(888) 880-7810	(312) 542-1900	(972) 385-5388	(310) 824-6600	(212) 887-1300

RMR-10-05325 9/10