Cable Penetration Projections

Methodology Used to Create Data and Report

J.D. Welch
Chief Technology Officer, BIGPIPE, Golden, CO
October 25, 2001bb
Revised November 6, 2001

Objective

Our goal was to produce a report that projects the likely number of households passed by cable, likely number of households NOT passed by cable, and likely number of households subscribing to cable, in each US county based on county size in each state organized into geographic regions.

Report Content

Geographic Breakdown:

- by region of US:
    - Northeast
    - Southeast
    - Midwest
    - Southwest
    - West Coast
- by state, and within each state,
    - by county, and for each county
        - designated by "size" – A, B, C or D

Data Elements:

- the number of households
- the projected percentage and number of households passed by cable,
- the projected percentage and number of households subscribing to cable, and
- the projected percentage and number of households NOT passed by cable.

Methodology

The methodology was based on Homes Passed (HP) and Cable Subscribers (Subs) of approximately 7,500 ZIP codes, representing a random selection of cable systems within the BIGPIPE cable system directories (proprietary). The data, provided directly by cable system operators (MSOs), included:

- ZIP code
- households passed
- households subscribing to cable

The MSO data by ZIP code (MSO Zips) represented a "true" random selection of cable systems from a variety of counties (of all four size designations) within states within each region.

 ©2001, BIGPIPE

The baseline ZIP code data, based on September, 2001 US Postal Service data, included:

- ZIP code
- ZIP code name (city)
- Nielsen County
- Nielsen State
- Nielsen County Size (A, B, C or D)
- Region
- Number of households (HHs)

(Note: We found that using USPS data yielded more accurate results than 2000 Census data, which had been the previous basis for determining HHs. Census data may not accurately represent all of the households eligible or able to receive cable service. Whereas if a non-PO Box USPS location does not accurately receive its mail, the USPS is usually promptly notified by the resident(s) of that location. We reasoned that if a USPS "household" could receive mail, it could receive a coaxial cable. Subsequent research bore this assumption out.)

Verification, Validation and Sanity Checks

The first thing we did with the "master" ZIP data was to assign a numeric "BPFIPS" to facilitate joins and searches.

Joining the MSO ZIPs to the USPS ZIPs by ZIP Code, the MSO ZIPs were assigned the corresponding values of:

- BPFIPS
- Nielsen State
- Nielsen County
- Region
- ZipCodeName
- HHs

However, out of 14,635 MSO ZIPs available to us, 1,658 did not have a corresponding record in the USPS ZIP data, and therefore could not be assigned any of the USPS ZIP values. These records were deleted from the MSO ZIPs data. These 1,658 MSO ZIPs, however, only represented 1.26% of the total HP and 1.06% of the total Subs available to us, so this should not have a significant impact on the calculated projections.

Next, all MSO records where HP was null or 0 were deleted. This was 384 MSO ZIPs. The result was 11,782 records of MSO data by ZIP where HP was >0 and a corresponding "master" ZIP code record existed, and 29,760 "master" ZIP records with HHs.

MSO ZIP data was then aggregated by ZIP Code, eliminating any MSO distinctions and resulting in a single record for each ZIP of MSO data. The data for MSO and master

©2001, BIGPIPE

ZIPs were then aggregated into counties, identified by unique BPFIPS. The result was a table of 3,134 master counties and a table of 1,887 counties of MSO data (at least 1 HP) which could be joined by BPFIPS, and each retaining the County, State, CountySize and Region attributes, but losing the distinctiveness of ZIP Code. The master ZIPs counties had HHs aggregated into a total number of household for the county, called [Total HHs]. The MSO data had HP and Subs aggregated, as well as the aggregate number of HHs for those ZIPs of MSO data in each county. This attribute is called [MSO HHs], representing the number of HHs that correspond to the HP and Subs data available for each county, and facilitates what should be the comparison of HP to HHs for the same ZIPs for which we have HP data. Total HHs for each county, [Total HHs], was also populated in the MSO table to facilitate comparisons of HP to total HHs in each county. [Total HHs] represents the total aggregate number of households in the county, as opposed to [MSO HHs] which is just the aggregated of households for those ZIPs in the county for which we also have HP and Subs. Both [MSO HHs] and [Total HHs] were included for comparison with HP to account for the differences in the sources of data, as illustrated by the occurrences where HP was higher than [MSO HHs].

Adjusting HP and Subs

The next stage was to determine where the aggregate number of HP was greater than the total number of HHs for each County. The MSO data covered several overlapping months as a result of the reporting parameters of the different MSOs providing the HP and Subs data. In addition, cable systems frequently change ownership, and there is always some amount of ownership change taking place within any given time period. As such, it was possible that a ZIP included in one MSO's data could also be included in a different MSO's data from an earlier or later reporting period, where the system providing service to that ZIP (or group of ZIPs) had changed ownership from one MSO to the other between the times the MSOs reported their data to us. In these cases the HP were significantly higher than the HHs. In other situations, HP was only slightly higher than HHs. This could be attributed to HP and Subs being counted in one ZIP by an MSO when the homes were actually in an adjacent ZIP, but for whatever reason the MSO decided to report the data in the first ZIP. To reduce the chance of incorrectly reducing one ZIP's HP and Subs at the expense of the (accurate) total HP and Subs for the entire county, the aggregated MSO data was used.

If HP were greater than the HHs of the ZIPs that had MSO data in a county, HP were adjusted down by the difference between HP and total HHs for that county, plus 1%. In other words, for each county where the aggregate HP was greater than [MSO HHs], HP was adjusted to equal:

Adjusted HP = [HP] – (([HP] – [MSO HHs]) * 1.01)

Subs were also adjusted down, by 99% of the ratio of the difference between HP and [MSO HHs] and [MSO HHs], or:

Adjusted Subs = ([Subs] * ((((([HP] – [MSO HHs]) / [HP]) – 0.99) * –1))

©2001, BIGPIPE