COMPARING 2001 NHTS DATA TO DATA FROM OTHER SOURCES

It is inevitable that the 2001 NHTS data will be compared to data from other sources. For example, the 2001 NHTS data will be compared to data from previous surveys to address questions of whether travel behavior has changed over time, and whether the changes are real or an artifact. Also, the 2001 NHTS data collected from New York residents will be compared to those collected from the rest of the country to determine whether New York residents have similar travel behavior than the rest of the country.

There are two points to remember when comparing 2001 NHTS data to data from other sources. First, one should always use the **weighted** statistics. Second, one should always draw conclusions by taking into account the variability in the data. In general, caution needs to be exercised when comparing estimates based on travel surveys or other sampling approaches. Since the 2001 New York NHTS did not interview every New York resident regarding his/her travel patterns, travel statistics/rates reported in this report are estimated based on data collected from a representative sample of New York residents. Whenever estimates are derived based on data from a sample, they are subject to statistical errors.

A statistical test(s) should be conducted to ascertain whether differences between the 2001 NHTS data and those of previous years (or those of other sources) are real, or within the statistical noise. The following example is used to illustrate such a test. It is estimated that New York residents drove a total of 95,207 million miles in their vehicles (vehicle miles traveled (VMT)) in 2001. This figure was estimated to be 95,571 million miles in 1995. The question then becomes whether this decrease in driving is real or within the sampling errors. To answer this question with a statistical certainty, one needs first to estimate the variances for these VMT estimates. They are $\sigma^2_{2001}$ and $\sigma^2_{1995}$ million in 2001 (denoted by $\hat{\sigma}^2_{2001}$) and 2,200$^2$ million in 1995 (denoted by $\hat{\sigma}^2_{1995}$). If the $z$ test statistic is less than -1.96 (or greater than 1.96), then we are 95% confident to conclude that the 2001 VMT was actually less than (or greater than) that in 1995. The test statistics is computed as:
\[
\frac{VMT_{2001} - VMT_{1995}}{\hat{\sigma}_{VMT_{2001} - VMT_{1995}}} = \frac{VMT_{2001} - VMT_{1995}}{\sqrt{\hat{\sigma}^2_{2001} + \hat{\sigma}^2_{1995}}} = \frac{95,207 - 95,571}{\sqrt{1,844^2 + 2,200^2}} = -0.127
\]

Since -0.127 is greater than -1.96 (i.e., less negative), we are about 95% confident that the decrease in driving from 1995 to 2001 is within the statistical noise.

Alternatively, one could compare the confidence intervals\(^1\) of these \(VMT\) estimates. If the intervals do not overlap, then one can conclude that the VMT difference between the two years is real. However, if the intervals do overlap, then it is not necessarily true that the difference is real. This is because the confidence-interval approach overestimates the standard error of the difference (e.g., \(VMT_{2001} - VMT_{1995}\)). Therefore, we recommend that the \(z\) test be used when comparing 2001 data with data from earlier surveys or from other sources.

Standard errors were calculated for 2001 NHTS estimates on: 1) vehicle occupancy, 2) the total vehicle miles traveled, 3) the total number of vehicle trips, 4) the total person miles traveled, 5) the total number of person trips, and 6) the average length of person trips. All of these estimates are categorized by mode (e.g., car, air, transit) and trip purpose (e.g., to work, religion related activities, shopping). This information is available on the web site:

http://www.dot.state.ny.us/ttss/index.html

Similar tables for 1995 NPTS data are also available on this site. If these published tables do not include standard errors for particular 2001 NHTS data, then one can use the web tools on www.nhts.ornl.gov to calculate the errors.

---

\(^1\) Confidence intervals represent a range of values within which the true population parameter is likely to lie with a pre-determined confidence.