

A Guide to Using the Time Use Data Files

A Guide to Using the Time Use Data Files

The time use portion of the GSS Cycle 12 collected data on the daily activities of Canadians. Information was collected by asking respondents to report their daily activities during the course of a 24 hour reference day starting at 4:00 in the morning. While multiple activities may be done simultaneously, the scope of the survey only allowed for the collection of data on one activity - the main activity as determined by the respondent.

For each activity respondents were asked the start and end time of the activity, where the activity took place and who the respondent was with at the time. The latter was not asked for sleep and most other personal care activities (codes 400, 450, 460 and 480).

The activities reported by respondents were coded into 177 individual groups and these groups were then categorized into 10 major activity groups and 24 minor groups. See Appendix L for details of the classification.

Two separate data files were created from the results of the 1998 Time use Survey: the main file and the time use episode file. The following is a guide to their use.

Three main measures of time use

The analysis done using the file will usually make use of one or more of the following measures.

1. Participation Rate

This is the proportion of the population who reported a particular activity. This is calculated as:

$$P^a = \frac{\sum_i W_i X_i^a}{\sum_i W_i}$$

where P^a = participation rate for activity a
 X_i^a = 1 if respondent reported activity a, = 0 otherwise
 W_i = weight for person i

Note that the indicator of participation is a non-zero number of episodes for that activity.

2. Average time for participants

The average time spent on an activity by all participants in that activity is calculated as:

$$TP^a = \frac{\sum_i W_i t_i^a}{\sum_i W_i X_i^a}$$

where TP^a = average time for all participants in activity a
 X_i^a = 0 or 1, indication of participation in activity a
 t_i^a = time on activity a for person i (=0 if no participation)
 W_i = weight for person i

3. Average time for total population

The average time spent on an activity by the total population (including both participants and non participants) is calculated as:

$$T^a = \frac{\sum_i W_i t_i^a}{\sum_i W_i}$$

where T^a = average time for total population in activity a
 t_i^a = time on activity a for person i (=0 if no participation)
 W_i = weight for person i

This time will always be less than the average time for participants and is equal to the time for participants if the participation rate is 100%.

The following are a number of comments that are intended to help in using the time use files:

1. The participation rates and the average times can be calculated for any subgroup of the population by including only the individuals in the subgroup.
2. The average time spent either for the participants or the entire population represent an average over a full seven day week (automatically due to the weight) unless a selection is done for a particular day of the week using variable DDAY.
3. The average time for the total population summed across all activities is equal to 1440 minutes (24 hours).
4. Average time for the total population can be added to obtain average time for a grouping of activities.
5. The participation rate can be also calculated by dividing the average time for the population by the average time for the participants. Similarly, the average for participants can be approximated by dividing the average time for the population by the participation rate.
6. Adding durations for social contacts (i.e. variables DURSOC01 to DURSOC11) will likely exceed 24 hours in most situations since time spent for a given activity with more than one type of social contact is counted each time. For example, watching television for an episode of 45 minutes with spouse and children will account for 45 minutes in DURSOC02 (spouse) as well as 45 minutes in DURSOC03 (children of the household).
7. Code 002 represents time spent on activities the respondent refused to report, while code 001 represents gaps in time when the respondent described his/her reference day. Respondents included in the results will not have more than 4 hours missing or refused in total.
8. Durations for each activity are for main activity only (as perceived by the respondent).
9. Variables on the Main File can be linked to variables on the Episode File using the variable RECID as a matching key.

Main File

In addition containing the bulk of the questionnaire responses and derived variables, the Main File provides summary time use activity information for each respondent on:

- i) the total time spent on each activity;
- ii) the total time spent at various locations;
- iii) the total time spent with various persons.
- iv) the total time spent on helping a person or an organization and its characteristics .

Note that this file summarizes the data for each respondent along each of these four dimensions of activities.

It does not, however, provide the details on individual activity episodes. For example, the Main file provides the total time spent on an activity such as T.V. watching, although the total time may have been reported on more than one episode of T.V. watching during the day. The Main file indicates the number of episodes of each activity but does not indicate when during the day they occurred.

Similarly the information for location and "who with" is the total for the day. The "who with" data do not add to 24 hours as a respondent could be with more than one person or groups of persons at a time. DURMEIN provides an unduplicated measure of time spent with the household members. There is no information on this file which links an activity with a location or who the person was with at the time. This information is provided on the detailed episode file described below.

Further summarization of the diary information produced variables for 10 major groups of activities and the 24 subcategories found in Appendix L. Other derived variables are:

- Total duration of time in transit
- Total duration of time spent with household members
- Number of activities
- Number of episodes

In order to provide control counts in using the Main File related to time use, a table is included at the end of this appendix. Users should be able to replicate these using the file.

Examples using the Main file

a) ACTIVITY TABLES

When weighted estimates for the duration of time spent at an activity, for example, employed work, by the population are required, use the variables

- WGHTFIN (weight)
- DVPAID (employed work).

When weighted estimates for the duration of time spent at an activity for participants only are required, exclude the respondents who did not report that activity, e.g., employed work,

i.e., Select respondents for whom DVPAID > 0.

The participation rate of a given activity is the percentage of the total population that reported the activity and can be derived using the formula provided.

When weighted estimates are required for a sub-group of the population, select the provided code for the desired sub-group, for example, time spent at employed work (DVPAID) for males and employed males. The variables used would be

- WGHTFIN (weight)

DVPAID (employed work)
 ACT7DAYS (main activity in the past 7 days)
 SEX (sex of respondent)

The selected subgroup would be defined as those where SEX = 1 and ACT7DAYS = 1.

DVPAID	Total Population	Total Participants ¹	Participation Rate (%)
Males	11,937,362	6,085,025	51
Employed Males	7,208,912	5,225,699	73

b) LOCATION

When weighted estimates for the duration of time spent at various locations or in various means of transit by the population are required use the following variables:

WGHTFIN (weight)
 DURLOC01 (home)
 DURLOC02 (work)
 DURLOC03 (someone else's home)
 DURLOC04 (other place)
 DURLOC05 (car as a driver)
 DURLOC06 (car as a passenger)
 DURLOC07 (walking)
 DURLOC08 (bus or subway)
 DURLOC09 (bicycle)
 DURLOC10 (other form of transit)
 DURLOC97 (missing or refused location)
 DURLOC98 (location not known)
 DURLOC99 (location not stated)

When weighted estimates for duration of time spent at various locations or in transit by participants only are required, exclude the respondents who did not report any time at that location or in transit,

i.e., Select respondents for whom DURLOC## > 0.

The participation rate of activity at a given location or given means of transit, is the percentage of the total

¹ For any activity, if sample size is less than 25 or weighted sample size is less than 35,000 (at the Canada level) then the data are not considered reliable and should be suppressed.

population that reported activity at the location or in transit and can be derived using the formula provided.

DURLOC02

Location (Work)	Total Population	Total Participants	Participation Rate (%)
Employed Males	7,208,912	4,523,220	63

DURLOC01 to DURLOC99 provides an estimate of the duration of time spent at various locations or in various means of transit. These categories are mutually exclusive, therefore the time will add to 24 hours for any given population.

c) SOCIAL CONTACTS

When weighted estimates for the duration of time spent with various social contacts for the population are required, use the following variables:

WGHTFIN	(weight)
DURSOC01	(alone)
DURSOC02	(spouse/partner)
DURSOC03.	(with household child(ren) less than 15 years of age)
DURSOC04	(with parent(s) or parent(s)-in-law who is living in the household)
DURSOC05	(with other member of the household)
DURSOC06	(with respondent's non-household child(ren) less than 15 years of age)
DURSOC07	(with respondent's non-household child(ren) 15 years of age or older)
DURSOC08	(with parent(s) or parent(s)-in-law who is not living in the household)
DURSOC09	(with other family member(s) who is not living in the household)
DURSOC10	(with friend(s) who is not living in the household)
DURSOC11	(with another person(s) who is not living in the household)
DURSOC97	(refused or missed activities)
DURSOC98	(personal activities ²)
DURSOC99	(don't know or not stated)

When weighted estimates for the duration of time spent with social contacts for participants only are required, exclude the respondents who did not report the required social contact,

i.e., Select respondents for whom DURSOC## > 0

The participation rate of activity with a given social contact is the percentage of the total population that reported some activity with the contact and can be derived using the formula provided.

Social	Total	Total	Participation
---------------	-------	-------	---------------

² Respondents were not asked for social contacts during personal care activity codes 100, 450, 460 and 480.

Contact (Spouse)	Population	Participants ³	Rate (%)
Employed			
Males	7,208,912	4,571,896	63

DURSOC01 to DURSOC11 provides estimates of the duration of time spent alone or with others. The duration of time with social contacts will not necessarily add to 24 hours because a respondent can spend time in the company of more than one person at a time.

Social contacts are further summarized in two derived variables:

DURMEIN includes total time with any household member (any of DURSOC02 to DURSOC05 is greater than zero) but with no persons outside the household.

DURMOUT includes the total time with any non-household person (any of DURSOC06 to DURSOC11 is greater than zero) but with no household member.

Although these are mutually exclusive, they do not add to 24 hours since time spent with both household and non-household members is not included.

Episode File

The episode file provides the detailed information on each activity episode reported by respondents. For each episode there is information on the start and end time of the activity, the duration of episode (derived from start and end time), the location of the episode, a set of variables that reflect who the respondent was with during the episode, and information on who an activity helped. Since there could be multiple contacts for an episode, the contact data is provided in the form of a set of variables, one for each type of contact.

The episode file consists of 221,105 records. The unit record for this file is the episode and not the respondent. Each record represents a single activity⁴ in a respondent's day, and all respondent's episodes must add up to twenty four hours (1440 minutes). For example, a respondent who has reported 26 different episodes for his/her reference day has generated 26 records on the Episode file. There is no information on the characteristics of the respondent. However each episode can be linked to the respondent using the RECID and characteristics can be obtained from the Main File. In addition, each episode includes information on the diary day and the total number of episodes for a respondent.

³ For any activity, if sample size is less than 25 or weighted sample size is less than 35,000 (at the Canada level) then the data are not considered reliable and should be suppressed.

⁴ It is not uncommon to find a string of two or more episodes with the same activity codes. These would have been reported as separate episodes when the location of the activity changed or when there was a change in the social contacts present.

Each episode has a weight, WGHTEPI. This is the weight to use when using the Episode File to make estimates based on episodes. When the episode file is used to derive a respondent characteristic, the person weight, WGHTFIN, should be used with the derived characteristic. To make this easier, WGHTFIN has been added to the Episode File on last episode for each respondent.

Examples using the Episode file

The episode file can be used for a number of different types of analysis. One use of the file is to consider a given activity (e.g. T.V. watching) and to analyze the distribution of episodes across time (time of day and/or day of week). The file can also be used to look at where various activities take place (e.g. paid work at home) or the social contacts for various activities. The file can also be used to look at the distribution of activities at any point in time (e.g. what is the population doing at 8:00 a.m., 11:00 p.m., 3:00 a.m., etc). More complicated analysis can be done by linking episodes for an individual and looking at the sequencing of different activities. The episode sequence number EPINO will facilitate this. Similarly by linking the episodes back to the characteristics of respondents, one can look at who in the population engages in various activities at different times during the day.

In view of the additional complexity of exploiting of the episode file, the GSS staff is interested in any work that is done with the file. Users are encouraged to contact the GSS staff to pass on any experiences with the data. The GSS staff will attempt to share whatever experience they have and that which other researchers have reported to them.

In cases where an analysis focuses on an activity, e.g. television viewing, that could have more than one episode in a day, the analyst must decide which weight to use. If, in the analysis, each episode should contribute separately to the estimate, then the episode weight, WGHTEPI, should be used. If, on the other hand, each respondent should contribute at most once to the estimate then the person weight, WGHTFIN, should be used with a derived person level variable.

For instance, the average length of an episode of watching television is an episode based statistic, while the average amount of time a person spends watching television in a typical day is a person based statistic. The first would be estimated as the (weighted) average over all episodes of watching television of the length of the episode. The second would be estimated by taking the weighted average over all respondents of the total length for each respondent of all episodes of watching television.

Here are some examples of the logic and algorithms that should be used when working with the Time Use Episode File. The file should for most purposes be sorted by RECID (the respondent identifier) and EPINO (the identifier of separate episodes for the respondent).

a) A PERSON BASED STATISTIC

When weighted estimates for the average amount of time spent daily at an activity, e.g., work for pay at main job, at a given location, e.g., at home, are required, the estimate is a person based one, the average time a person spends each day at an activity.

Use the variables: ACTCODE
 DURATION
 PLACE
 WGHTFIN

Select ACTCODE = 011 (Working for pay at main job)
 PLACE= 01 (Home).

Calculate the average time by summing across all records as follows:

$$\frac{\sum_k \text{WGHTFIN}_k (\sum_i \text{DURATION}_i, \text{ where ACTCODE}=011 \text{ and PLACE}=01)}{\sum_k \text{WGHTFIN}_k}$$

where DURATION_i = episode time for episode i (of respondent k).
 WGHTFIN_k = respondent weight for respondent k .

This could be done by using the episode file to create a new file with one record for each respondent and these variables: WGHTFIN, DURINT, where DURINT is the ‘duration of interest’ for the respondent, the total duration of all episodes for the respondent with ACTCODE=011 and PLACE=01. The procedure would be to set DURINT to zero, then look through the episode records for the first respondent, and whenever ACTCODE=011 and PLACE=01, add DURATION to DURINT. After examining the last episode for the first respondent, save WGHTFIN and DURINT to the new file, reset DURINT to zero and continue with the second respondent. Continue in this way until a record has been added to the new file for each respondent. Then the equation above become:

$$\frac{\sum_k \text{WGHTFIN}_k \text{DURINT}_k}{\sum_k \text{WGHTFIN}_k} = \frac{291,687,822}{24,260,137} = 12.02 \text{ minutes}$$

Interpretation: On an average day, Canadians spend 12.02 minutes working at their main job while they are at home.

Calculate the participation rate as follows:

$$\frac{\sum_k \text{WGHTFIN}_k (\text{for those with } \sum \text{DURATION}_i, \text{ where ACTCODE}=011 \text{ and PLACE}=01 \text{ not equal to zero, (i.e. DURINT not equal to zero)})}{\sum_k \text{WGHTFIN}_k}$$

$$= \frac{1,422,215}{24,260,137} = 5.9 \%$$

Interpretation: On an average day, 5.9% of Canadians do some work at their main job while they are at home.

And so the average time spent per participant is:

$$= \frac{291,687,822}{1,422,215} = 205.1 \text{ minutes}$$

Interpretation: On an average day when they do some work at home, Canadians spend 205.1 minutes working at their main job while they are at home.

b) AN EPISODE BASED STATISTIC

When weighted estimates for the average duration of a single episode of a certain activity, e.g., watching television, the estimate is an episode based one, the average length of an episode of watching t.v.

Use the variables: ACTCODE
 DURATION
 WGHTEPI

Select ACTCODE = 911, 912, 913, or 914 (Watching t.v.)

Calculate the average time by summing across all episode records as follows:

$$\frac{\sum_j \text{WGHTEPI}_j \text{ DURATION}_j}{\sum_j \text{WGHTEPI}_j}, \text{ where ACTCODE}=911, 912, 913, \text{ or } 914$$

where DURATION_j = episode time for episode j.
 WGHTEPI_j = episode weight for episode j.

This yields an estimate of:

$$\frac{3,198,292,910}{\text{-----}} = 95.30 \text{ minutes}$$

33,559,271

Statistical analysis software packages and database management software packages are currently used for exploiting these types of data files. For example, SAS and SPSS are widely used for statistical analysis of this data. While these types of packages can be used to merge information from the Main and the Episode files, intensive users of the Episode file may also want to consider bringing these files together in a relational database. Most database management system software packages provide a mechanism for easily linking and retrieving data from the two files with a one-to-many relation. This is usually based on Standard Query Language (SQL). Among these are Access, Oracle, Paradox, SQLServer.