How to use data on Time Use in the GSS

Geneviève Vézina

March 21st 2019

Delivering insight through data for a better Canada

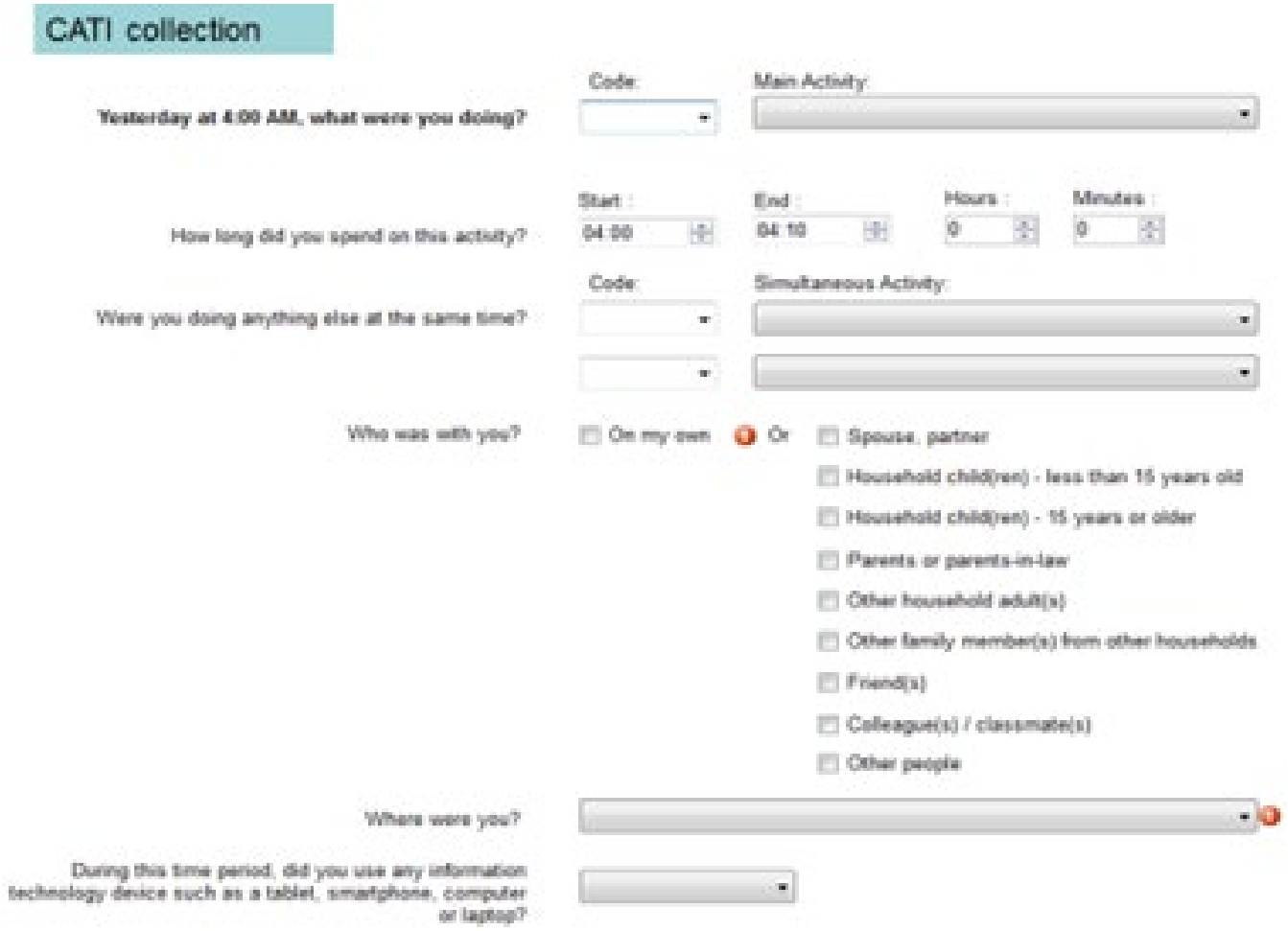
Outline Presentation

- 1. Description of the survey
- 2. Variables on the episode file
- 3. Link between the main file and the episode file
- 4. Person weights vs episode weights
- 5. Measures of time use
  - 5.1 Main file
  - 5.2 Episode file
- 6. Examples
- 7. Conclusion

1. Description of the survey

- Particularity of this cycle: the diary
- Way to get accurate information about people's time use
- Started at 4:00 am for 24 hours
- One reference day
- Questions on the nature of the main activity, the duration, the location, who the respondent was with, whether information technology was used
- 2 analytical files: main file (17,390 records in 2015) and episode file (274,108 records in 2015)

1. Description of the survey



1. Description of the survey

- Tracks patterns of time use
- Better understand what contributes to their well-being and stress levels
- Data collected is used by governments to make decisions
- Data to inform about the changing nature of time use

2. Variables on the episode file

Variables	Definition
TUI_01	Activity code of the episode
TUI_06A to TUI_06J	Indicates who the respondent was with during the activity mentioned at TUI_01 (up to 10 people)
TUI_03A and TUI_03B	First and second simultaneous activities (max of two)
TUI_07	Indicates whether or not technology was used during the activity mentioned at TUI_01
TUI_10	Level of subjective well-being during the activity mentioned in TUI_01 (asked for only two activities in the day)
DURATION	Duration, in minutes, of the episode (derived from the variables STARTIME, ENDTIME, STARTMIN and ENDMIN)
LOCATION	Location where the activity took place
DDAY	Diary- reference day
WGHT_EPI	Weight of episode mentioned at TUI_01
PUMFID	Record identification number (not related to the episode)

3. Link between the main file and the episode file

- The main file provides summary time use activity information for each respondent
  - The total time spent on each activity
  - The total time spent at various locations
  - The total time spent with various persons
- Does not provide the details of individual activity episodes, but indicates the number of episodes of each activity
- Similarly for the location of the episodes
- Similarly for who the respondent was with during the episode

3. Link between the main file and the episode file

- The episode file provides the detailed information on each activity episode reported by respondents
  - Start and end time of the activity
  - Duration of episode
  - Location of the episode
  - Simultaneous activities
  - Who the respondent was with during the episode
- The unit record for this file is the episode and not the respondent

3. Link between the main file and the episode file

- Several variables from the main file were derived from the episode file
- Variables DURXX linked to TUI_01 where XX is the activity code
- Variables DURLXXX linked to LOCATION where XXX is the location of the activity
- Variables DURSXXX linked to TUI_06A to TUI_06J where XXX indicates with who the respondent was with

3. Link between the main file and the episode file

- Extracted from the episode file for one respondent

PUMFID	TUI_01	DURATION	LOCATION	TUI_06A	TUI_06B	TUI_06C	TUI_06D	TUI_06E	TUI_06F	TUI_06G	TUI_06H	TUI_06I	TUI_06J
10000	1	120	At home or on property	No	No	No	No	No	No	No	No	No	Yes
10000	6	60	At home or on property	No	No	No	No	No	No	No	No	No	Yes
10000	60	30	At home or on property	No	No	No	No	No	No	No	No	No	Yes
10000	1	180	At home or on property	No	No	No	No	No	No	No	No	No	Yes
10000	7	10	Travel - Car (Driver)	Yes	No	No	No	No	No	No	No	No	No
10000	7	80	Travel - Car (Driver)	Yes	No	No	No	No	No	No	No	No	No
10000	7	15	Travel - Car (Passenger)	No	No	No	No	No	No	No	Yes	No	No
10000	6	60	Restaurant, bar or club	No	No	No	No	No	No	No	Yes	No	No
10000	7	45	Travel - Car (Passenger)	No	No	No	No	No	No	No	Yes	No	No
10000	60	120	At home or on property	Yes	No	No	No	No	No	No	No	No	No
10000	21	120	At home or on property	Yes	No	No	No	No	No	No	No	No	No
10000	2	30	At home or on property	Yes	No	No	No	No	No	No	No	No	No
10000	5	60	At home or on property	No	No	No	No	No	No	No	No	No	Yes
10000	50	30	Outdoors	Yes	No	No	No	No	No	No	No	No	No
10000	60	150	At home or on property	Yes	No	No	No	No	No	No	No	No	No
10000	50	20	Outdoors	Yes	No	No	No	No	No	No	No	No	No
10000	60	100	At home or on property	No	No	No	No	No	No	No	No	No	Yes
10000	1	210	At home or on property	Yes	No	No	No	No	No	No	No	No	No

3. Link between the main file and the episode file

- If we want to derive the variables of the activities duration, we need to:
  - Sort the file by the TUI_01 variable for each respondent
  - Sum up the durations using the DURATION variable for each activity by respondent
  - The sums are the value of each DURXX variables, where XX is the activity code

3. Link between the main file and the episode file

PUMFID	TUI_01	DURATION
10000	1	210
10000	1	120
10000	1	180
10000	2	30
10000	5	60
10000	6	60
10000	6	60
10000	7	10
10000	7	80
10000	7	15
10000	7	45
10000	21	120
10000	50	30
10000	50	20
10000	60	120
10000	60	150
10000	60	30
10000	60	100

210+120+180=510 minutes for activity 1

60+60=120 minutes for activity 6

10+80+15+45=150 minutes for activity 7

30+20=50 minutes for activity 50

120+150+30+100=400 minutes for activity 60

3. Link between the main file and the episode file

On the main file we have the following variables for our respondent

PUMFID	DUR01	DUR02	DUR05	DUR06	DUR07	DUR21	DUR50	DUR60
10000	510	30	60	120	150	120	50	400

All the other DURXX variables are equal to 0. The sum of the DURXXs has to equal 1440. There are also variables to indicate the frequency of episodes for each activity, EPIXX, where XX is the activity code

PUMFID	EPI01	EPI02	EPI05	EPI06	EPI07	EPI21	EPI50	EPI60
10000	3	1	1	2	4	1	2	4

3. Link between the main file and the episode file

- We can do the same thing for the location variable :
  - Sort the file by the LOCATION variable by respondent
  - Sum up the duration using the DURATION variable for each location by respondent
  - The sums are the value for the DURLXXX variables where XXX is the code of the location

3. Link between the main file and the episode file

PUMFID	DURATION	LOCATION
10000	210	At home or on property
10000	120	At home or on property
10000	180	At home or on property
10000	30	At home or on property
10000	60	At home or on property
10000	60	At home or on property
10000	120	At home or on property
10000	120	At home or on property
10000	150	At home or on property
10000	30	At home or on property
10000	100	At home or on property
10000	30	Outdoors
10000	20	Outdoors
10000	60	Restaurant, bar or club
10000	10	Travel - Car (Driver)
10000	80	Travel - Car (Driver)
10000	15	Travel - Car (Passenger)
10000	45	Travel - Car (Passenger)

Sum of DURATION = 1180 minutes

Sum of DURATION = 50 minutes

Sum of DURATION = 90 minutes

Sum of DURATION = 60 minutes

3. Link between the main file and the episode file

On the main file, we have the following variables for our respondent

PUMFID	DURL300	DURL305	DURL309	DURL313	DURL314
10000	1180	50	60	90	60

All the other DURLXXX variables are equal to 0. The sum of the DURLXXXs has to equal 1,440.

3. Link between the main file and the episode file

- We can do the same for the social contact variable of the activity. However, since each social contact is represented by a different variable on the episode file, we have to calculate the sum of the duration for each variable TUI_06A to TUI_06J.
- The sums are the values of the DURSXXX variables where XXX is the code of the social contact of the activity.

3. Link between the main file and the episode file

PUMFID	DURATION	TUI_06A
10000	120	No
10000	180	No
10000	60	No
10000	60	No
10000	60	No
10000	15	No
10000	45	No
10000	30	No
10000	100	No
10000	210	Yes
10000	30	Yes
10000	10	Yes
10000	80	Yes
10000	120	Yes
10000	30	Yes
10000	20	Yes
10000	120	Yes
10000	150	Yes

$$\sum DURATION$$

PUMFID	DURATION	TUI_06H
10000	210	No
10000	120	No
10000	180	No
10000	30	No
10000	60	No
10000	60	No
10000	10	No
10000	80	No
10000	120	No
10000	30	No
10000	20	No
10000	120	No
10000	150	No
10000	30	No
10000	100	No
10000	60	Yes
10000	15	Yes
10000	45	Yes

PUMFID	DURATION	TUI_06J
10000	210	No
10000	30	No
10000	60	No
10000	10	No
10000	80	No
10000	15	No
10000	45	No
10000	120	No
10000	30	No
10000	20	No
10000	120	No
10000	150	No
10000	120	Yes
10000	180	Yes
10000	60	Yes
10000	60	Yes
10000	30	Yes
10000	100	Yes

3. Link between the main file and the episode file

On the main file, we have the following variables for our respondent

PUMFID	DURS200	DURS207	DURS209
10000	770	120	550

All the other variables DURSXXX are equal to 0. The sum of the DURSXXXs won't necessarily be 1,440 since we can be with more than one person during an activity.

3. Link between the main file and the episode file

- We can create a combined file using the PUMFID that allows us to have all the variables from both files on the same file.

```
proc sql;
        create table GSS29 as
        select a.*, b.*
        from GSS29pumfm as a, GSS29pumfe as b
        where a.PUMFID=b.PUMFID
        order by a.PUMFID;
quit;
```

SAS

```
sysuse main_e, clear
merge 1:m PUMFID using epi_e
save "c:\GSS29_e.dta"
```

STATA

3. Link between the main file and the episode file

WARNING!!

Episodes file

PUMFID	TUI_01	DURATION	LOCATION	WGHT_EPI
10000	1	10	300	10
10000	4	30	306	10
10000	8	90	302	10
20000	1	60	300	15
20000	5	15	300	15
20000	9	45	301	15
20000	1	20	303	15
30000	6	50	302	2
30000	2	5	302	20

Main file

PUMFID	AGE	SEX	PRV	WGHT_PER
10000	55	1	24	10
20000	32	2	35	15
30000	78	2	59	20

Combined file

PUMFID	TUI_01	DURATION	LOCATION	WGHT_EPI	AGE	SEX	PRV	WGHT_PER
10000	1	10	300	10	55	1	24	10
10000	4	30	306	10	55	1	24	10
10000	8	90	302	10	55	1	24	10
20000	1	60	300	15	32	2	35	15
20000	5	15	300	15	32	2	35	15
20000	9	45	301	15	32	2	35	15
20000	1	20	303	15	32	2	35	15
30000	6	50	302	2	78	2	59	20
30000	2	5	302	20	78	2	59	20

4. Person weights vs episode weights

Person weight → WGHT_PER

- This is the basic weighting factor for analysis at the person level, i.e. to calculate estimates of the number of persons (non-institutionalized and aged 15 or over) having one or several given characteristics.
- We need to use WGHT_PER for estimation at the person level.

$$\sum_{i=1}^{17,390} \text{WGHT\_PER} = 29,766,399^{1}$$

1 Estimate of the number of persons aged 15 and over in the population

4. Person weights vs episode weights

Episode weights → WGHT_EPI

- This is the basic weighting factor for the analysis at the episode level i.e. to calculate estimates on the number of times an activity is done by the Canadian population.
- It indicates the number of time use episodes that a record on the Episode File represents.

$$\sum_{i=1}^{274,108} \text{WGHT\_EPI} = 461,837,622^{1}$$

1 Estimated total number of activities performed in a day by the population 15 years of age or older.

4. Person weights vs episode weights

- Weights to use during the analysis
  - Always ask yourself what is the unit of interest for the estimate, is it the person or the episode
  - Always use weights during the analysis

Person	Duration	Weight_EPI
A	15	10
A	10	10
A	35	10
B	20	20
C	10	15
D	15	30
D	15	30

Summary at the person level

Person	Duration	Weight_PER
A	60	10
B	20	20
C	10	15
D	30	30

4. Person weights vs episode weights

Average at the episode level

$$
\begin{aligned}
Mean &= \frac{\sum Duration\_Epi \times Weight\_Epi}{\sum Weight\_Epi} \\
&= \frac{(15 \times 10) + (10 \times 10) + (35 \times 10) + (20 \times 20) + (10 \times 15) + (15 \times 30) + (15 \times 30)}{10 + 10 + 10 + 20 + 15 + 30 + 30} \\
&= \frac{2050}{125} \\
&= 16.4
\end{aligned}
$$

4. Person weights vs episode weights

Average at the person level

$$
\begin{aligned}
Mean &= \frac{\sum Duration\_Per \times Weigth\_Per}{\sum Weight\_Per} \\
&= \frac{(60 \times 10) + (20 \times 20) + (10 \times 15) + (30 \times 30)}{10 + 20 + 15 + 30} \\
&= \frac{2050}{75} \\
&= 27.3
\end{aligned}
$$

Therefore, an episode has an average duration of 16.4 minutes, but a person does the activity for 27.3 minutes on average.

4. Person weights vs episode weights

Main file or episode file?

- What do we want to analyze?
- Analysis of one characteristic/variable from the diary: use either file
- Analysis of two characteristics/variables or more from the diary : use the episode file
- Example:
  - Using the main file, we could estimate the total time spent sleeping.
  - Using the episode file, we could estimate the total time spent sleeping not at home

5. Measures of time use

- Three measures are frequently used to analyse Time Use data
  - Participation rate
  - Average time spent on activities by participant
  - Average time spent on activities by the total population

5. Measures of time use

Participation rate

- A participant in an activity is a person who has reported as least one occurrence of the activity on their reference day. The participation rate is the percentage of the population having reported the activity.

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

where p^a = participation rate for activity a

X_i^a = if respondent reported activity a, = 0 otherwise

W_i = weight for person i

5. Measures of time use

Average time spent on activities by participant

- Average time is obtained by dividing the estimated total time spent per day on the activity by the estimated total number of persons who reported this activity.

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

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

5. Measures of time use

Average time spent on activities by the total population

Average time spent on activities is obtained by dividing the estimated total time spent per day on the activity by the estimated total number of persons in a given population.

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

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

5. Measures of time use

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 on activities is usually calculated based on a 24-hour period, over a 7-day week unless a selection is done for a particular day of the week using variable DDAY.

3- For activities like paid work which are normally considered over a 5-day period, a simple conversion will reconstruct activities to a 5-day average. Multiply the daily average by 7 for a weekly average and divide by 5. For example, a paid workday of 5.7 hours (averaged over 7 days) will convert to an 8.0 hour day (averaged over 5 days).

5. Measures of time use

4- The average time for the total population summed across all activities is equal to 1440 minutes (24 hours).

5- Average time for all activities for the total population can be added to obtain average time for a grouping of activities.

6- The participation rate can be calculated by dividing the average time for the population by the average time for the participants. Similarly, the average for participants can be calculated by dividing the average time for the population by the participation rate

5. Measures of time use

7- Adding durations for social contacts (i.e. variables DURS200 to DURS209) 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 DURS201 (spouse) as well as 45 minutes in DURS202 (household children less than 15 years of age).

8- Code 95 represents time spent on activities the respondent refused to report, were unknown or uncodable.

9- Variables on the Main File can be linked to variables on the Episode File using the variable PUMFID as a matching key.

5.1 Main file

Participation rate to an activity

- Participation rate is the percentage of the total population that reported the activity
- Example: We want to know the participation rate of employed men who did paid work compared to paid work for men in general
- Variables:
  - WGHT_PER (weight)
  - PDWKDUR (employed work)
  - ACT7DAYS (main activity in the past 7 days)
  - SEX (sex of respondent)

5.1 Main file

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

First participation rate

- Calculate the sum of the weights for respondents that are men and have PDWKDUR > 0
- Calculate the sum of the weights for all respondents that are men

Second participation rate

- Calculate the sum of the weights for respondents that are employed men and have PDWKDUR > 0
- Calculate the sum of the weights for all respondents that are employed men (ACT7DAYS=1)

5.1 Main file

PDWKDUR	Total Population	Total Participants	Participation Rate(%)
Males	14,689,652	6,741,792	46
Employed Males	8,780,694	5,991,536	68

$$\text{Participation Rate} = \frac{Total\ \ Participants}{Total\ Population} \times 100 = \frac{6{,}741{,}792}{14{,}689{,}652} \times 100 \approx 46$$

Interpretation: On an average day, 46% of **males** have spent time at employed work compared to 68% of **employed males** who have spent time at employed work. Because the GSS on Time Use only ask the details for one day, it is possible that an employed male has not reported time spent at employed work during his reference day.

5.1 Main file

```
* Total population of men;
proc means data=main sum;
        var WGHT_PER;
        where sex=1;
run;
* Total number of employed men;
proc means data=main sum;
        var WGHT_PER;
        where sex=1 and PDWKDUR>0;
run;
```

5.1 Main file

```
cd "T:\CRAD-DARC\00INTERNAL\DARC-Consults\CONSULTATIONS_ISABELLE\STATA\Conversion_codeSAS_vers_codeSTATA\STATA"

sysuse main_e, clear

*Create an indicator variable for the calculation of the frequencies
generate indicator = 1

*Total population of employed men
tabulate indicator [iweight=WGHT_PER] if SEX==1 & ACT7DAYS==1

*Number of employed men who did paid work
tabulate indicator [iweight=WGHT_PER] if SEX==1 & ACT7DAYS==1 & PDWKDUR>0
```

5.2 Episode file

- Each episode has a weight, WGHT_EPI. 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, WGHT_PER, should be used with the derived characteristic.
- In cases where an analysis focuses on an activity 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, WGHT_EPI, should be used.
  - If, on the other hand, each respondent should contribute at most once to the estimate then the person weight, WGHT_PER, should be used.

5.2 Episode file

Person-based statistics and estimates

- When weighted estimates for the average amount of time spent daily at an activity at a given location are required, the estimate is person-based, the average time a person spends each day at an activity.
- Weighted estimates for the average amount of time spent daily on paid work at main job at home
- Variables:
  - TUI_01=8 (Working for pay job)
  - DURATION
  - LOCATION=300 (Home)
  - WGHT_PER

5.2 Episode file

$$\frac{\sum_k WGHT\_PER_k(\sum_i DURATION_i, \text{ where TUI\_01=8 and LOCATION=300})}{\sum_k WGHT\_PER_k}$$

Where DURATION$_i$ = time for episode i for respondent k

WGHT_PER $_k$= weight for respondent k

$$= \frac{\sum_k WGHT\_PER_k DURINT_k}{\sum_k WGHT\_PER_k}$$

- Need to derive a total (DURINT) for the episodes of interest, TUI_01=8 and LOCATION=300, by respondent
- We can do this by creating a counter
- Keep the last row for each respondent with the total of the counter and keep only variables related to the person and not the episode

5.2 Episode file

Example of an episodes file

PUMFID	TUI_01	DURATION	WGHT_EPI	LOCATION
1000	8	100	10	300
1000	2	25	10	300
1000	8	45	10	300
1000	7	120	10	305
1000	5	60	10	302
2000	8	30	20	300
2000	1	15	20	306
2000	5	20	20	304
2000	8	150	20	301
3000	8	90	15	300
3000	3	180	15	305
3000	8	75	15	300

Creation of a counter DURINT for TUI_01=8 and LOCATION=300

PUMFID	TUI_01	DURATION	WGHT_EPI	LOCATION	DURINT
1000	8	100	10	300	100
1000	2	25	10	300	100
1000	8	45	10	300	145
1000	7	120	10	305	145
1000	5	60	10	302	145
2000	1	15	20	306	0
2000	8	30	20	300	30
2000	5	20	20	304	30
2000	8	150	20	301	30
3000	8	90	15	300	90
3000	3	180	15	305	90
3000	8	75	15	300	165

Keep the person level variables, WGHT_EPI becomes WGHT_PER

PUMFID	WGHT_PER	DURINT
1000	10	145
2000	20	30
3000	15	165

Keep the last row for each respondent with the total of DURINT

PUMFID	TUI_01	DURATION	WGHT_EPI	LOCATION	DURINT
1000	5	60	10	302	145
2000	8	150	20	301	30
3000	8	75	15	300	165

5.2 Episode file

$$\frac{\sum_k WGHT\_PER_k DURINT_k}{\sum_k WGHT\_PER_k}$$

$$\frac{565{,}489{,}635}{29{,}766{,}399} = 19 \text{ minutes}$$

Interpretation: On an average day, Canadians spend 19 minutes working at their main job while they are at home.

5.2 Episode file

```
* Create variable DURINT (counter for the sum of duration);
proc sort data=epi; by PUMFID; run;
data epi_durint;
    set epi;
    by PUMFID;
    retain DURINT 0; /*Counter for the total of time spent working at home */

    if first.PUMFID then do;
        if TUI_01=8 and LOCATION=300 then DURINT=DURATION; else DURINT=0;
    end;
    else if first.PUMFID ne 1 then do;
        if TUI_01=8 and LOCATION=300 then DURINT=DURINT+DURATION; else DURINT=DURINT;
    end;

    if last.PUMFID; /* keep the last row the total of the counter for each respondent */
    keep PUMFID DURINT WGHT_EPI; /* keep variables of interest */
    rename WGHT_EPI=WGHT_PER;
run;
```

5.2 Episode file

```
* Weighted sum of the minutes spent at working at home;
proc means data=epi_durint sum;
    var DURINT;
    weight WGHT_PER;
run;
* Total population;
proc means data=epi_durint sum;
    var WGHT_PER;
run;

 * If we want to calculate the mean in one step;
 proc means data=epi_durint mean;
    var DURINT;
    weight WGHT_PER;
 run;
```

5.2 Episode file

```
cd "T:\CRAD-DARC\00INTERNAL\DARC-Consults\CONSULTATIONS_ISABELLE\STATA\Conversion_codeSAS_vers_codeSTATA\STATA"

sysuse epi_e, clear
svyset PUMFID [pweight=WGHT_EPI], bsrweight(WEPI_001-WEPI_500) vce(bootstrap) mse

*Create variable DURINT
generate DURINT_temp=DURATION if TUI_01 == 8 & LOCATION == 300
replace DURINT_temp=0 if DURINT_temp!=DURATION
sort PUMFID
by PUMFID: egen DURINT = total(DURINT_temp)

*Keep the last row with the total for each respondent
sort PUMFID
by PUMFID: keep if _n == _N
```

5.2 Episode file

generate indicator=1 if DURINT>0

replace indicator=0 if DURINT<=0

generate WGHT_PER=WGHT_EPI

generate one=1

*Weighted sum of the minutes spent at working at home

svy: total DURINT

*Total population

svy: total one

*If we want to calculate the mean in one step

svy: mean DURINT

5.2 Episode file

- Participation rate

$$\frac{\sum_k WGHT\_PER_k \left(\begin{array}{c} for\ those\ with\ \sum DURATION_i,\ where\ TUI\_01 = 8\ and\ LOCATION = 300 \\ not\ equal\ to\ zero.\ (i.e.\ DURINT\ not\ equal\ to\ zero) \end{array} \right)}{\sum_k WGHT\_PER_k}$$

$$= \frac{1{,}735{,}589}{29{,}766{,}399} = 5.8\%$$

Interpretation: On an average day, 5.8 % of Canadians do some work at their main job while they are at home.

5.2 Episode file

- Average time spent per participant

$$\frac{\sum_k WGHT\_PER_k(\sum_i DURATION_i, \text{ where TUI\_01=8 and LOCATION=300})}{\sum_k WGHT\_PER_k(\text{for tose with } \sum DURATION_i, \text{ where } TUI\_01 = 8 \text{ and } LOCATION = 300 \text{ not equal to zero, } (i.e. DURINT \text{ not equal to zero}))}$$

$$= \frac{565{,}489{,}635}{1{,}735{,}589} = 326 \text{ minutes}$$

Interpretation: On an average day when they do some work at home, Canadians spend 326 minutes working at their main job while they are at home.

5.2 Episode file

Episode-based statistics and estimates

- Weighted estimates for the average duration of a single episode of a certain activity. We are looking for the average time of one episode of the activity and not the average time Canadians spent on the activity
- Weighted estimates for the average duration of a single episode of watching television.
- Variables:
  TUI_01=60 (Watching t.v.)
  DURATION
  WGHT_EPI

5.2 Episode file

$$\frac{\sum_j WGHT\_EPI_j DURATION_j \ , \quad where\ TUI\_01{=}60}{\sum_j WGHT\_EPI_j \ , \quad where\ TUI\_01{=}60}$$

Où $DURATION_j$ = time for episode j

$WGHT\_EPI_{j\,j}$ = weight for episode j

$$= \frac{3{,}739{,}440{,}834}{32{,}386{,}351} = 115 \ \ minutes$$

Interpretation: On an average day, when an episode of watching television is reported, 115 minutes on average is spent on it.

5.2 Episode file

```
/* Numerator */
proc means data=epi sum;
        var DURATION;
        weight WGHT_EPI;
        where TUI_01=60;
run;
/* Denominator */
proc means data=epi sum;
        var WGHT_EPI;
        where TUI_01=60;
run;

/* If we want to do it in one step*/
proc means data=epi mean;
        var DURATION;
        weight WGHT_EPI;
        where TUI_01=60;
run;
```

5.2 Episode file

```
cd "T:\CRAD-DARC\00INTERNAL\DARC-Consults\CONSULTATIONS_ISABELLE\STATA\Conversion_codeSAS_vers_codeSTATA\STATA"

sysuse epi_e, clear
svyset PUMFID [pweight=WGHT_EPI], bsrweight(WEPI_001-WEPI_500) vce(bootstrap) mse

generate one=1

* Numerator
svy: total DURATION if TUI_01==60
* Denominator
svy: total one if TUI_01==60
* If we want to do it in one step
svy: mean DURATION if TUI_01==60
```

6. Examples

- 1- Average time that Canadians spent per day eating
  - We can use the main file or the episode file
- 2- Number of times, on average, that Canadians ate and how much time it took them on average each time
  - We can use the main file or the episode file
- 3- Proportion of Canadians who ate all their meals alone or without their spouse/partner
  - We need to use the episode file or the combined file
- 4- Proportion of Canadians who watched t.v. while eating
  - We need to use the episode files for simultaneous activities

7. Conclusion

- Always determine what estimate you are looking
- Determine your unit of interest
- Find your population of interest
- Determine which file you need to work with

Thank you!

Questions?

Genevieve.vezina@Canada.ca