Test of a Cross-sectional Sampling Timer

R. Mark Sharp

2021-09-18

Contents

1 Purpose

A cross-sectional time on task estimation method is presented, compaired to use of time logging, illustrated with an example, and its accuracy and precision demonstrated with a simulation. Embedded within this package are the R functions used to create this document including the simulation software. The specific example, which is the the husbandry and veterinary care of colony animals is provided by setup section of the ".Rmd" source file and a comma separated value (CSV) file with the actor, activity, and activity descriptions.

Management of a value stream requires adequate estimate of the resources used to create that value with time on task being a direct measure of the primary resource for many value streams. The need to obtain quality estimates of these times on the various tasks in a value stream motivated me to provide and advocating the use of this time on task method

In this example, we want to gain insight into how much time is spent on specific activities by various actors[1] responsible for the success of the husbandry and veterinary care of colony animals. It is anticipated that having time on task data will provide insight in those areas where where improvement opportunities exist.

Detailed planning regarding questions to be addressed is required to design which activities[2] should be measured.

A benefit of time on task studies is that they provide hard data regarding how time is spent and how much of it directly contributes to our primary goals. A second benefit is that these data allow each class of actor to advocate for modifying the way their time is being used to improve their work environment and productivity.

[1] animal_caretaker and veterinary_tech

[2] We expect that some actors will work on more than one value stream and all will have additional responsibilities. Thus, some activities recorded are intentionally defined as various unrelated activities. Planning how much detail to collect and how the data will be analyzed should occur prior to any data being collected.

Time on task estimation methods that call for representative actors within each group or type of actors to log time on tasks during their work day have the problems listed below.

- Failure rate in logging activities can be high.
- Recall of the amount of time spent is often wrong this will artificially inflate the variance.
- Preconceptions of how time should be spent biases recalled durations.
- Keeping, collecting, and analyzing logs requires significant administrative overhead.
- The biases introduced from earlier points makes interpretation of results difficult.
- The longitudinal collection of data requires a long data collection phase before any analysis can begin.
- Variance among different projects is anticipated to be high, which further complicates analysis.

The cross-sectional design assumes that we have the ability to send a question to selected individuals at selected times and collect that response.

The cross-sectional design presented and tested below has the following advantages.

- The failure to respond rate has far less impact since failure to respond simply initiates another response request.
- Participants do not need to recall time spent.
- Administrative overhead is greatly reduced.
- Bias is minimized.
- Data can be analyzed at any time.
- Variance among projects could be assessed and more accurately measured if desired. However, I recommend that project differences be ignored during initial timeing studies.

2 Simulation

2.1 Design

The remainder of this document presents an example data collection, a simple simulation study using a realistic design that provides sufficient detail to assess expected data quality.

The example has a set of representative activities for animal_caretakers.[3]. It simulates asking a set of animal_caretakers to indicate which of the activities listed on the questionnaire they were doing at the time the question popped up on their screen. Once the animal_caretaker makes a selection, that selection is returned as a result to the collection software, which cumulates the responses for later analysis.

In the simulation model, these queries and responses can all be processed in less than a second but, as described below, this simulation is constructed so that each simulated animal_caretaker responds five times during the entire duration of the experiments illustrated below.

The manner in which the questions are presented and the timing of when the questions are presented are critical aspects of the study design but are not discusses further.

2.2 Simulation Activity Frequencies

The animal_caretakers' activities are programmed to occur with the following frequencies.

[3] The software is has the ability to examine multiple job types, job specific activity lists, and corresponding expected frequencies.

Table 1: List of possible activities for the animal_caretaker with assigned frequencies to be used in simulations.

Activity	Frequency
administrative work	0.10
animal care correspondence	0.01
animal care meeting	0.05
animal care record keeping	0.05
animal care record use	0.05
food preparation	0.15
gang cage animal feeding	0.20
gang caged animal inventory and health check	0.05
group animal transfer	0.05
individual animal feeding	0.05
individual animal transfer	0.01
individual cage cleaning	0.05
individual cage sanitation	0.05
individual caged animal inventory and health check	0.05
ordering of major equipment	0.01
ordering of supplies and minor equipment	0.01
personal time	0.05
other	0.01

For clarity, we demonstrate one and three samples.

2.3 Data Collection Example

A single sample: gang cage animal feeding
Three samples: food preparation, individual cage cleaning, and individual animal feeding.

Lets see how precise our estimates would be if we included 20 animal_caretakers sampled just 5 times a month for each of 12 months.

Since the algorithm is not actually modeling individual animal_caretakers, the order of sampling has no impact on results, which means we simply take 1200 (20 * 5 * 12) samples per result set and compare each to the expected values.

Table 2: List of activities, the frequency in which they were observed, the simulated frequency of the activity (expected), and the absolute difference between the observed and expected frequencies.

	Duration		
Description	Observed	Expected	Delta
administrative work	0.096	0.10	0.004
animal care correspondence	0.008	0.01	0.002
animal care meeting	0.038	0.05	0.012
animal care record keeping	0.059	0.05	0.009
animal care record use	0.042	0.05	0.008
food preparation	0.135	0.15	0.015
gang cage animal feeding	0.208	0.20	0.008
gang caged animal inventory and health check	0.054	0.05	0.004
group animal transfer	0.048	0.05	0.002
individual animal feeding	0.057	0.05	0.007
individual animal transfer	0.012	0.01	0.002
individual cage cleaning	0.059	0.05	0.009
individual cage sanitation	0.052	0.05	0.002
individual caged animal inventory and health check	0.050	0.05	0.000
ordering of major equipment	0.009	0.01	0.001
ordering of supplies and minor equipment	0.009	0.01	0.001
other	0.010	0.01	0.000
personal time	0.052	0.05	0.002

We can estimate the duration in minutes on each activity if we allocate the 1200 observations into one 8 hour day which has 560 minutes.

Table 3: List of activities and the estimated number of minutes spent on the activity if represented in a single day.

Description	Minutes
administrative work	53.7
animal care correspondence	4.7
animal care meeting	21.5
animal care record keeping	33.1
animal care record use	23.3
food preparation	75.6
gang cage animal feeding	116.7
gang caged animal inventory and health check	30.3
group animal transfer	27.1
individual animal feeding	31.7
individual animal transfer	6.5
individual cage cleaning	33.1
individual cage sanitation	29.4
individual caged animal inventory and health check	28.0
ordering of major equipment	5.1
ordering of supplies and minor equipment	5.1
other	5.6
personal time	29.4

I have run this experiment several times and usually the estimates are shown to be very close as in seen in the *Delta* column (within about 1 percent). However, it is more instructive to simulation this experiment

many times to learn what precision we can expect.

2.4 Results

Figure 1 shows a histogram plot of the results of repeating the simulation of 20 animal caretakers providing 5 responses in each of the 12 months of a year 5000 times. Similarly, Figure **??** shows a histogram plot of the similar simulation with 40 animal caretakers instead of 20.

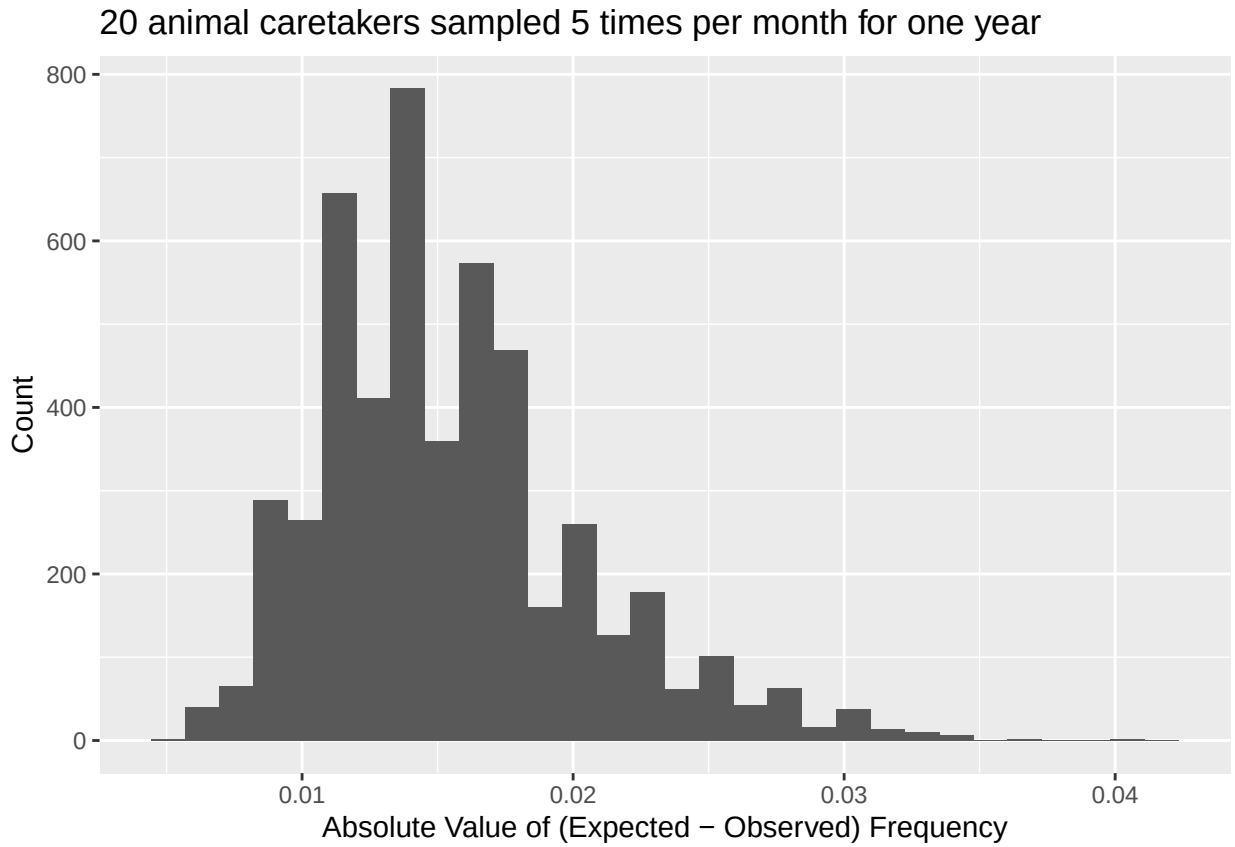
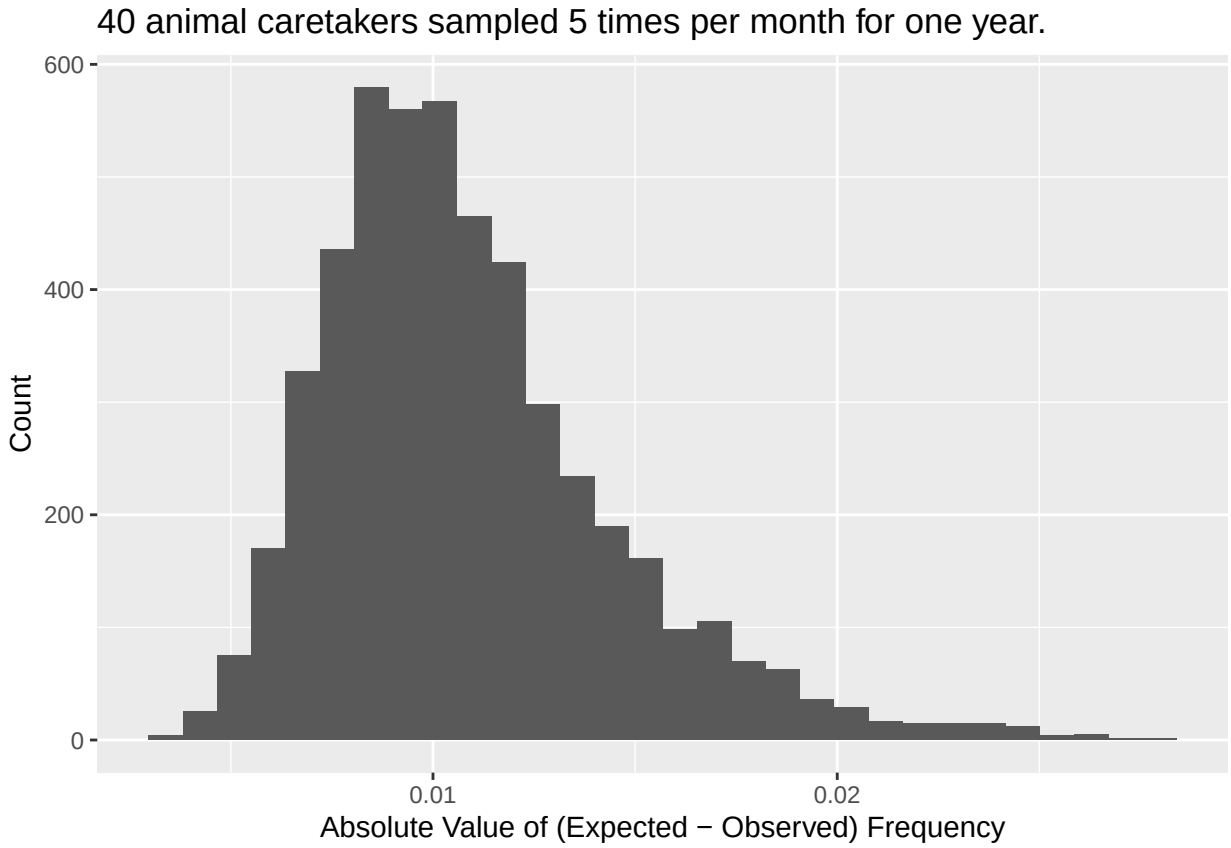


Figure 1: Simulation of 5,000 iterations with 20 animal caretakers found 17.4% of durations were greater than 2 percent away from the expected value and 1.04% of durations were greater than 3 percent away from the expected value with the largest absolute delta value being 0.042.


40 animal caretakers sampled 5 times per month for one year.
600
400
200
0
Count
0.01
0.02
Absolute Value of (Expected − Observed) Frequency

2.5 Conclusions

Prior experiments with this cross-sectional activity time estimation have shown that about 250 samples will provide time estimates for up to 20 activities that are all accurate and precise within ±2 percent, which should be precise enough for most management activities. I do not think any amount of sampling using actor logs can be that accurate because of inherent cognative bias introduced by the actors' own expectations.

2.6 Appendix

See the setup code chunk of this vignette to see how to customize this document to examine results you can expect from your own time on task activity analysis. You will need to provide the CSV file[4] with your own actors[5], activities, and activity descriptions in addition to providing your definitions for the setup chunk definitions indicated below.

```
activity_file <-
  system.file("extdata",
              "colony_management_defined_activities.csv",
              package = "crosssectiontimer", lib.loc = NULL,
              mustWork = FALSE)
value_stream_descriptor <- stri_c("the husbandry and veterinary care of ",
                                  "colony animals")
actor <- "animal_caretaker" # can be any value in the actor column of the
                            # activity file

# The product of n_actors, times_per_month, and n_months is the size.
# The size should be about 250 for about 20 activities to get the precision
# illustrated herein
n_actors <- 20
times_per_month <- 5
n_months <- 12

iterations <- 5000          # 5000 seems more than adequate for stable results
```

[4] Comparable to colony_management_defined_activities.csv

[5] Actors are used as variable names so must follow rules for naming variables in R. They must start with a letter, contain only letters, numerals, ".", and "_".