Subject: Selecting sample within one standard deviation in R
Posted by berhardt93 on Tue, 15 Feb 2022 20:00:29 GMT
View Forum Message <> Reply to Message

Hi,

I'm looking at the Nigeria 2018 DHS. I created a variable "tot_encounters" that calculates the number of sexual encounters reported by an individual in the past 12 months by adding the values from their most recent, second most recent, and third most recent partners. I also created the weighting variable "weight".

I found the mean of the weighted variable:

```
weighted.mean(yesNUIS$tot_encounters, yesNUIS$weight)
```

Then I found the standard deviation:

```
weighted_var <- wtd.var(yesNUIS$tot_encounters, yesNUIS$weight)
weighted_sd <- sqrt(weighted_var)
```

Weighted mean = 27.78
Standard deviation = 25.57

Now I want to select all observations that fall within one standard deviation (2.21-53.35). When I tried to do this, the sample was 80% of the original sample, not 68% (aka. the number of observations within one standard deviation of the mean):

```
sdNUIS <- yesNUIS
sdNUIS %<>%
  dplyr::filter(tot_encounters > 2.2057 & tot_encounters <  53.3527)
```

How would I make sure that this filter only includes the 68% within one standard deviation of the weighted mean?

Thanks!

Subject: Re: Selecting sample within one standard deviation in R
Posted by Bridgette-DHS on Tue, 15 Feb 2022 21:52:45 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

You have an extremely skewed distribution. The "68%" rule works for normally distributed variables, and the normal approximation doesn't work for your variable. I can think of two options.

One would be to take the log of the frequency, which will have a distribution that is more nearly normal, but there's the problem that you can't take the log of 0. Another option would be to calculate the percentiles of the distribution. If you identify the 25th and 75th percentiles, then you have the boundaries for the middle 50%. Or identify the 16th and 84th percentiles, which enclose the middle 68%.