lab3.1_estes

Andrew Estes

11/8/2021

he code below creates a table of 12 observations with 3 variables. The first column is "ID" which is a unique identifier between 1 and 12, only using whole numbers - much like a ranking.

An additional observation column is created with the "oops" row of coding. The "oops" column calculates a random number following a normal distribution with the mean of 0 and standard deviation of 1. That means 99.7% of all datapoints that are randomly created will fall between -3 and 3.

The final column is called IDSum. This simply adds the "ID" column to the "oops" column for a final number.

```
library(dplyr)
M <- 12
biglist <- tibble(ID = seq(1,M),
     oops=c(rnorm(M, mean=0, sd=1)),
     IDSum = (ID + oops))
biglist
```

Now we will change it from a tibble of 12 observations to 120k observations.

```
M <- 120000
biglist <- tibble(ID = seq(1,M),
     oops=c(rnorm(M, mean=0, sd=1)),
     IDSum = (ID + oops))
write.csv(biglist, "biglist120.csv")
```

Here are the last 9 rows of this dataset: 119991 119991 -0.201401801 119990.7986 119992 119992 0.220555834 119992.2206 119993 119993 1.073082965 119994.0731 119994 119994 0.059058134 119994.0591 119995 119995 1.22591248 119996.2259 119996 119996 0.232277739 119996.2323 119997 119997 -0.672915803 119996.3271 119998 119998 1.559956798 119999.56 119999 119999 -0.919454481 119998.0805 120000 120000 -0.117597277 119999.8824

They were obviously the same for both the Excel and Google Sheets versions. However the Google Sheets was a much more complicated process.