

Ejercicios clase 15

Jorge Loría

Sept 25, 2017

Datos:

Copie y pegue en un script de R Markdown el siguiente código:

```
install.packages('randomNames')
library(randomNames)
library(dplyr)

set.seed(5)
total<- 100

datosSalario<- data.frame(Genero = sample(c('male','female'),total,replace = TRUE),
                          Antiguedad = rpois(total,15),
                          Provincia = sample(7,total,replace = TRUE),
                          Moneda = sample(c('CRC','USD'),total,replace = TRUE),
                          stringsAsFactors = FALSE) %>%

  rowwise() %>%
  mutate(Nombre = randomNames(gender = Genero,ethnicity = 'Hispanic')) %>%
  mutate(Edad = Antiguedad + 18 + rpois(n=1,5)) %>%
  mutate(Casado = rbinom(n =1,size = 1,prob = 0.3)) %>%
  mutate(Hijos = rbinom(n = 1,size = 1,prob = 0.27)) %>%
  mutate(Regimen = sample(c('IVM', 'Magisterio','PJ'),size = 1,prob = c(0.6,0.3,0.1))) %>%
  ungroup() %>%
  rowwise() %>%
  mutate(Cantidad_Hijos = ifelse(Hijos == 1,rbinom(size = 5,prob = 0.5,n = 1),0)) %>%
  ungroup()

annos<- expand.grid(Nombre = datosSalario$Nombre,
                  Anno = 2017:2021,
                  Mes = 1:12,
                  stringsAsFactors = FALSE)

datosSalario<- datosSalario %>%
  left_join(annos) %>%
  rowwise() %>%
  mutate(Salario = rnorm(n = 1,mean = 1500000,sd = 200000)) %>%
  ungroup() %>%
  group_by(Nombre) %>%
  arrange(Anno) %>%
  mutate(Antiguedad = Anno - first(Anno) + Antiguedad) %>%
  ungroup() %>%
  rowwise() %>%
  mutate(Volar = rbinom(n=1,size = 1,prob = 1/7)) %>%
  ungroup() %>%
  filter(Volar <1) %>%
  select(-Volar)
rm(annos,total)
```

Preguntas:

La idea de esta clase es que responda o verifique las siguientes preguntas utilizando los datos que se le proveen en la página anterior, aplicando las herramientas que vimos en las clases pasadas de R.

- Revise que la cantidad de hijos siempre sea cero para los casos en que las personas tienen el indicador de hijos en 0.
- Revise si las personas que se indica que tienen hijos siempre tienen hijos. En caso de que no tengan ponga que tienen 1 hijo.
- Revise cual es el salario en valor presente más alto. Usando para esto una tasa de interés anual de 5%.
- ¿Cuántas personas hay en cada uno de los regímenes?
- ¿Cuántas personas hay que estén casadas y que tengan más de 4 hijos? y Más de 3?
- ¿Cuántas personas hay de cada género por mes y por año?
- ¿Cuánto es el mayor crecimiento porcentual de salario que recibe una persona de un mes a otro? i.e. $\frac{S_t - S_{t-1}}{S_{t-1}}$
- ¿Cuál es el régimen de pensiones que tiene un mayor porcentaje de mujeres?
- ¿Cuántas personas sin hijos ganan en dólares en febrero del 2018?
- ¿Cual es el monto en valor presente que se espera que ganen todas las personas?
- ¿Cuál es la provincia con mayor cantidad de mujeres en diciembre del 2019?
- ¿Cuál es la provincia con mayor porcentaje de hombres para cada año?
- Para cada año, indique el monto total que se gana en cada régimen.
- Identifique, puede usar la función `summarise`, cuanto es la mayor antigüedad, edad y salario, para cada año.
- Verifique que la edad no incrementa con los años para al menos una persona. Asumiendo que las personas cumplen en enero de cada año. Arregle esta columna.
- ¿Cuál es la persona que recibe un mayor monto en todos los años?
- Indique qué porcentaje del total de personas cotizan cada mes.
- ¿Cuál es el nombre de las personas que tienen la mayor cantidad de hijos? Sugerencia: busque la función: `distinct`.