

Curso Introducción a R en investigación biomédica

Solución ejercicios prácticos Modulo4 - Manipulación y reestructuración de datos:
Introducción al tidyverse

Unidad de Apoyo Metodológico IIS Biogipuzkoa - Lore Zumeta & Jone Renteria

Noviembre 2025



Ejercicios modulo 4

```
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.2.3
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##     filter, lag
## The following objects are masked from 'package:base':
##     intersect, setdiff, setequal, union
# Cargar datos desde CSV (ajusta la ruta al archivo descargado)
diabetes <- read.csv("diabetes.csv")
```

1. dplyr::mutate()

Genera nuevas variables de peso en kg y altura en cm utilizando la función `mutate()`. La relación entre variables es la siguiente:

- Peso: 1 lb = 0.453592 kg
- Altura: 1 in = 2.54 cm

```
diabetes2 <- diabetes %>%
  mutate(peso_kg = weight * 0.453592,
        altura_cm = height * 2.54)

head(diabetes2)
```

2. dplyr::filter()

```
##      id chol stab.glu hdl ratio glyhb    location age gender height weight frame
## 1 1000  203          82  56   3.6  4.31 Buckingham 46 female   62   121 medium
## 2 1001  165          97  24   6.9  4.44 Buckingham 29 female   64   218 large
## 3 1002  228          92  37   6.2  4.64 Buckingham 58 female   61   256 large
## 4 1003   78          93  12   6.5  4.63 Buckingham 67 male    67   119 large
## 5 1005  249          90  28   8.9  7.72 Buckingham 64 male    68   183 medium
## 6 1008  248          94  69   3.6  4.81 Buckingham 34 male    71   190 large
##      bp.1s bp.1d bp.2s bp.2d waist hip time.ppn peso_kg altura_cm
## 1    118    59     NA     NA    29   38      720  54.88463  157.48
## 2    112    68     NA     NA    46   48      360  98.88306  162.56
## 3    190    92   185    92    49   57     180 116.11955  154.94
## 4    110    50     NA     NA    33   38      480  53.97745  170.18
## 5    138    80     NA     NA    44   41     300  83.00734  172.72
## 6    132    86     NA     NA    36   42     195  86.18248  180.34
```

2. dplyr::filter()

Filtrar a los individuos con peso mayor a 100kg

```
diabetes3 <- diabetes2 %>%
  filter(peso_kg > 100)

summary(diabetes3$peso_kg)
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	100.7	104.3	113.4	114.3	119.7	147.4

3. dplyr::filter() y operador AND (&)

Filtrar a los individuos con peso mayor a 100kg y altura menor a 180cm

```
diabetes4 <- diabetes2 %>%
  filter(peso_kg > 100 & altura_cm < 180)

head(diabetes4)

##      id chol stab.glu hdl ratio glyhb    location age gender height weight frame
## 1 1002  228          92  37   6.2  4.64 Buckingham 58 female   61   256 large
## 2 1253  234          65  76   3.1  4.67 Buckingham 47 male    67   230 large
## 3 1254  203          299 43   4.7 12.74 Buckingham 38 female   69   288 large
## 4 1280  232          87  30   7.7  5.10 Buckingham 37 male    68   252 large
## 5 1501  243          52  59   4.1  4.41 Buckingham 37 female   64   233 medium
## 6 2763  193          106 63   3.1  6.35 Buckingham 20 female   68   274 small
##      bp.1s bp.1d bp.2s bp.2d waist hip time.ppn peso_kg altura_cm
## 1    190    92   185    92    49   57     180 116.1196  154.94
## 2    137   100   149   110    45   46     480 104.3262  170.18
## 3    136    83     NA     NA    48   55     240 130.6345  175.26
## 4    140    95     NA     NA    43   47     420 114.3052  172.72
## 5    110    82     NA     NA    49   57     90 105.6869  162.56
## 6    165   110   153   100    49   58     60 124.2842  172.72

summary(diabetes4$peso_kg)
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	100.7	105.2	111.1	114.0	119.3	147.4

4. Cálculos matemáticos

```
summary(diabetes4$altura_cm)

##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##    154.9   162.6  167.6  167.8  172.7  177.8
```

4. Cálculos matemáticos

Calcula el promedio de peso en kg por sexo

```
diabetes5 <- diabetes2 %>%
  group_by(gender) %>%
  summarise(media_peso = mean(peso_kg, na.rm = TRUE)) %>%
  ungroup()

diabetes5

## # A tibble: 2 x 2
##   gender media_peso
##   <chr>     <dbl>
## 1 female     79.1
## 2 male       82.5
```

5. Generar una nueva variable

Crea una nueva variable de IMC usando peso en kg y altura en metros (para ello, primero recuerda transformar la altura de cm a m). La fórmula del IMC es la siguiente: - IMC: peso_kg/(altura_m^2)

```
diabetes6 <- diabetes2 %>%
  mutate(altura_m = altura_cm / 100,
        imc = peso_kg / (altura_m^2))

head(diabetes6)

##      id chol stab.glu hdl ratio glyhb location age gender height weight frame
## 1 1000  203          82  56   3.6  4.31 Buckingham 46 female    62    121 medium
## 2 1001  165          97  24   6.9  4.44 Buckingham 29 female    64    218 large
## 3 1002  228          92  37   6.2  4.64 Buckingham 58 female    61    256 large
## 4 1003    78          93  12   6.5  4.63 Buckingham 67 male     67    119 large
## 5 1005  249          90  28   8.9  7.72 Buckingham 64 male     68    183 medium
## 6 1008  248          94  69   3.6  4.81 Buckingham 34 male     71    190 large
##      bp.1s bp.1d bp.2s bp.2d waist hip time.ppn peso_kg altura_cm altura_m
## 1     118    59    NA    NA     29    38      720  54.88463    157.48   1.5748
## 2     112    68    NA    NA     46    48      360  98.88306    162.56   1.6256
## 3     190    92   185    92     49    57      180 116.11955    154.94   1.5494
## 4     110    50    NA    NA     33    38      480  53.97745    170.18   1.7018
## 5     138    80    NA    NA     44    41      300  83.00734    172.72   1.7272
## 6     132    86    NA    NA     36    42      195  86.18248    180.34   1.8034
##      imc
## 1 22.13094
## 2 37.41920
## 3 48.37024
## 4 18.63783
## 5 27.82475
```

6. Todo junto

```
## 6 26.49933
```

6. Todo junto

Analiza si la persona es diabética o no. Según los autores de la base de datos, utilizando el corte de condición diabética cuando los valores del ratio, variable ratio, son iguales o superiores a 7. Para ello, primero crea la variable diabetico. Además, selecciona las variables diabetico, ratio, edad, peso_kg y peso original y renombra la variable age e indica edad.

```
diabetes7 <- diabetes6 %>%
  mutate(diabetico = if_else(ratio >= 7, "Sí", "No")) %>%
  select(diabetico, ratio, age, peso_kg, weight) %>%
  rename(edad = age)

head(diabetes7)
```

```
##   diabetico ratio edad  peso_kg weight
## 1      No    3.6   46  54.88463   121
## 2      No    6.9   29  98.88306   218
## 3      No    6.2   58 116.11955   256
## 4      No    6.5   67  53.97745   119
## 5      Sí    8.9   64  83.00734   183
## 6      No    3.6   34  86.18248   190
```