



UNIVERSIDAD TÉCNICA DE AMBATO



FACULTAD DE INGENIERÍA CIVIL Y MECÁNICA

CARRERA DE INGENIERÍA MECÁNICA

DOCENTE: Ingeniero Federico Zertuche

ALUMNO: Patricio Pinos

CATEDRA: Estadística

CURSO: Tercer Semestre

PARALELO: "A"

CONTENIDOS: Ejercicios

EJERCICIOS

EJERCICIO 1

Pregunta 1 Cuáles de las siguientes expresiones valen 99 para $x = 10$ en R? Analicen la sintaxis como si estuvieran programando. $10x - 1$ $(x)(x) - 1$ $\text{abs}(x*x) - \text{abs}(9-x)$ $11 * x - x + 1$

Pregunta 1

- `10*X-1`
[1] 99
- `X*X-1`
[1] 99
- `abs(X*X)-abs(9-X)`
[1] 99
- `11*X-X+1`
[1] 101

Pregunta 2 Un vector contiene una serie de ganancias ordenadas de manera creciente. Escriban el código que genera: La suma de todas las ganancias. La segunda ganancia más grande. La diferencia más grande entre las ganancias. 1 Un booleano que responda a la pregunta: La más grande diferencia entre dos ganancias es mayor a 10? La menor diferencia positiva entre dos ganancias. El máximo número de ganancias que pueden sumar sin pasar de 10000.

Pregunta 2

`x=c(1100, 1200, 1300, 1800,2800)`

- `sum(x)`
[1] 8200
`x[c(4)]`
[1] 1800
`x[c(5)]- x[c(1)]`
[1] 1700
`Cummin(x)>10`

```
[1] TRUE TRUE TRUE TRUE TRUE
x[c(2)]- x[c(1)]
[1] 100
x[c(3)]- x[c(2)]
[1] 100
Cumsum(x)<1000
[1] FALSE FALSE FALSE FALSE FALSE
```

EJERCICIO 2

Vamos a estudiar los datos de los vuelos locales en Estados Unidos durante el 2011. Usen los verbos: `select()` `filter()` `mutate()` `arrange()` `group_by()` `summarise()` para manipular 3 data frames: Uno con los vuelos (`flights`), uno con los aviones (`planes`) y otro con el clima (`weather`).

PREGUNTA 3

Pregunta 3 Instalen la librería `nycflights13`. Escriban el código para encontrar todos los vuelos que: Fueron de SFO (San Francisco) hasta OAK (Oakland). Salieron en Enero. 2 Tienen demoras de más de una hora (las demoras están en minutos). Salieron entre medianoche y las 5 a.m. Tuvieron una demora de llegada 2 veces más grande que la de salida.

- `flights %>% filter(origin=='SFO', dest=='OAK') %>%
View()`
- `flights %>% filter(month=='1') %>%
View()`
- `flights %>% filter(dep_delay>=1) %>%
View()`
- `flights %>% filter(hour==24, hour==5) %>%
View()`
- `flights %>% filter(arr_delay==2*dep_delay) %>%
View()`

PREGUNTA 4

Pregunta 4 Lean la ayuda de `select()`. Escriban 2 formas de seleccionar las dos variables de retraso.

- `flights %>% select(arr_delay, -starts_with("-"))`

- `flights %>% select(dep_delay)`

PREGUNTA 5

Pregunta 5 Ordenen la tabla por fecha de salida y tiempo. ¿Cuáles fueron los vuelos que sufrieron las mayores demoras? ¿Cuáles recuperaron la mayor cantidad de tiempo durante el vuelo?

- `flights %>% arrange(dep_day, month, time, year) %>%
View ()`

PREGUNTA 6

Pregunta 6 Calculen la velocidad en mph usando el tiempo (que está en minutos) y la distancia (que está en millas). Cuál fue el avión que voló más rápido?

- `flights %>% filter(air_time, distance) %>%`
- `mutate(Velocidad=air_time*distance, air_time=air_time/60) %>%
View ()`

PREGUNTA 7

Pregunta 7 En dplyr el comando pipeline `%>%` se lee entonces. Significa: $x \% \> \% f(y) \longrightarrow f(x, y)$. Es decir pasa x como primer argumento de f. Qué significan las siguientes líneas de código: `flights %>% filter(! is.na(dep_delay)) %>% group_by(date, hour)
%>% summarise(delay = mean(dep_delay), n = n()) %>% filter(n > 10)`

- `flights %>% filter(! is.na(dep_delay))` Pasa los valores de síntesis en los atrasos de salida que se están desperdiciando.
- `group_by(date, hour)` Indica los valores colectivos entre la fecha y la hora.
- `summarise(delay=mean(dep_delay, n=n()))` Hace un sinopsis del promedio de los demoras.
- `filter(n > 10)` En esta función apareció error.
View ()

PREGUNTA 8

Pregunta 8 ¿Cuál es la destinación que tiene las demoras promedio más grandes? ¿Cuántos vuelos diarios hay? ¿Cuál es la mejor hora para viajar sin retraso?

- `flights %>% group_by(dest, arr_delay) %>%`
- `summarize(mean(arr_delay, na.rm=TRUE)) %>%`

View ()

EJERCICIO 3

Ejercicio 3 6,000 Años de Urbanización Global. En el artículo Spatializing 6,000 years of global urbanization from 3700 BC to AD 2000 los autores reúnen una serie de datos sobre la población de las ciudades del mundo desde 3700 A.C. hasta el 2000 D.C. Usemos estos datos para tratar de entender algunas cosas sobre la distribución espacial de las personas a través del tiempo. Para hacer esto vamos a tener que manipular dos data frames - uno para la población y otro con los nombres de los países y continentes - usando dplyr para eventualmente unirlos. En este ejercicio vamos a usar algunos verbos para unir data frames. `inner_join(d1, d2)` contiene solo las filas de d1 comunes a d2. `left_join(d1, d2)` contiene todas las filas de d1 y NA en las filas de d2 que no están en d1. `semi_join(d1, d2)` no añade columnas a d1. Contiene las filas de d1 comunes a d2.

PREGUNTA 9

Pregunta 9 ¿Cuántas personas han habitado Sur América? ¿Hay alguna forma de usar los datos para saber cuál es la región que tiene las poblaciones más antiguas? ¿Pueden usar los datos para tratar de entender si hay un patrón migratorio a lo largo de la historia? **(OPCIONAL)**