

ICO 187 ANÁLISIS DE DATOS

2021

CLASE 04: Manipulación de datos numéricos y no numéricos

Profesor: **Sebastián Egaña**

1. Data No Numérica

Recordar, que para el entorno de programación y análisis de datos, variable refiere a un elemento dentro del entorno de programación utilizado. Por otra parte, el término variable utilizado en estadística se corresponde con los tipos de datos en programación.

La data no numérica, se puede categorizar en el enfoque estadístico dentro del grupo de variables cualitativas:

1. Nominal
2. Ordinal
3. Binaria

Donde la variable nominal refiere a las variables que no posee un orden interno, a diferencia de la variable ordinal. En el caso de la variable binaria, corresponde a categorías dicotómicas o contrapuestas y vimos que puede ser representada de forma numérica.

Desde el enfoque de datos, en base al lenguaje R, existen los siguientes tipos de datos no numéricos:

1. Character: Corresponde a cadenas de caracteres en general.
2. Logical: Refiere a un tipo de dato lógico valores TRUE o FALSE. Se conoce también como booleanos.

Como ya se analizó, las fechas corresponden a un caso particular: de manera interna, se almacenan como números pero se muestran en la forma correspondiente una fecha. En este sentido, son muy parecidas a un texto. Para el caso de R, el formato de fecha más utilizado corresponde a POSIXct y Date, en donde el primero posee la cualidad de poder almacenar fecha y hora.

Para este tipo de datos, lo más importante para este tipo de datos corresponde a la extracción y transformación de los mismos. Ya revisamos los tipos de datos, pero en este caso la extracción y transformación de datos no se limita solo a los datos que son cadenas de caracteres; las fechas y algunos formatos de número también pueden ser tratados como

cadenas de caracteres debido a la forma en como trabaja Excel. Veamos un ejemplo de esto:



Ejercicio: Descarguemos algunos datos de la siguiente página [Enlace acá](#) e intentemos reconstruir la fecha de cada transacción en la bolsa.

Respuesta:

Podemos utilizar las siguientes funciones: FECHA, DIA, MES y AÑO, IZQUIERDA, DERECHA y EXTRAE.

```
=FECHA(DERECHA(A2;4);EXTRAE(A2;4;2);IZQUIERDA(A2;2))
```

Otra opción, podría ser usar REEMPLAZAR, y FECHA NUMERO.

```
=REEMPLAZAR(A2;3;1;"/")
```

```
luego =REEMPLAZAR(C2;6;1;"/")
```

```
y por último =FECHANUMERO(D2)
```

1.1. Aplicación en relación con los contenidos de la clase pasada

Utilizando la siguiente clase_04 fifa.csv, realicemos un ejercicio relacionado con la función SUMAR.SI, CONTAR.SI y PROMEDIO.SI. Utilicemos como agrupamiento el año y el mes. ¿Podríamos hacer esto de una manera distinta?

1.2. Funciones para caracteres

- IZQUIERDA
- DERECHA
- EXTRAE
- ESTEXTO
- ENCONTRAR
- REEMPLAZAR

1.3. Funciones para fechas

- FECHA
- DIAS

- MES
- MINUTO
- AÑO

2. Desafío clase pasada

- Genere una función que retorne el primer valor no blanco (vacío de un vector).
Resolvamos dicho ejercicio. Para esto, considerar las siguientes funciones:
 - INDICE
 - COINCIDIR
 - ESBLANCO

Respuesta:

Se debe aplicar funciones de la siguiente manera:

=INDICE(C4:N4;1;COINCIDIR(1;INDICE(1-ESBLANCO(C4:N4);1;0);0))

Veamos esto parte por parte.