

#####

## ## EJERCICIO 3 – PREPARACION DE DATOS EN EXCEL

#####

Copiar y pegar cada fórmula a la vez, sin retorno de carro.

En esta versión, los separadores de los parámetros en las fórmulas son punto y coma.

#1 ABRE EL ARCHIVO « Datos\_paso1.xlsx »

#2 CALCULAR LA FECHA DE INCIDENCIA – COLUMNA « F »

#2.1 Reemplazar cadena de texto '19930104' en fecha '1993-01-04'

=CONCATENAR (EXTRAE (A2;1;4) ; "-" ; EXTRAE (A2;5;2) ; "-" ; EXTRAE (A2;7;2) )

#2.2 Si el mes es igual a « 99 » y luego reemplazarlo por « 07 »

# Y si el día es igual a « 99 » entonces reemplace con « 15 »

=CONCATENAR (EXTRAE (A2;1;4) ; "-" ;  
 " ; SI (EXTRAE (A2;5;2) <> "99" ; EXTRAE (A2;5;2) ; "07" ) ; "-" ;  
 " ; SI (EXTRAE (A2;7;2) <> "99" ; EXTRAE (A2;7;2) ; "15" ) )

# En la columna "INC\_DATE", copie y pegue la fórmula anterior.

## CONSEJO: Copiar y pegar una fórmula

CTRL-C luego CTRL-SHIFT-END (para llegar al final de la columna) y luego CTRL-V

O haga doble clic en el cuadro negro en la parte inferior derecha de la primera celda con la fórmula (para que Excel se rellene automáticamente).

	A	B	C	D	E	F	G	H	I	J
1	INCID	SEX	BIRTHD	I10	TOP	INC DATE	BTH DATE	AGE	AGEGRP	YEAR
2	19930101		1 19379999	C049	49	1993-01-04				
3	19930101		2 19309999	C23	239					
4	19930101		1 19219999	C250	250					

## #3 CALCULAR LA FECHA DE NACIMIENTO – COLUMNA « G »

#3.1 Reemplazar cadena de texto '19370104' en fecha '1937-01-04'

=CONCATENAR (EXTRAE (C2;1;4) ; "-" ; EXTRAE (C2;5;2) ; "-" ; EXTRAE (C2;7;2) )

#3.2 Si el mes es igual a « 99 » y luego reemplazarlo por « 07 »

# Y si el día es igual a « 99 » entonces reemplace con « 15 »

=CONCATENAR (EXTRAE (C2;1;4) ; "-"  
 " ; SI (EXTRAE (C2;5;2) <> "99" ; EXTRAE (C2;5;2) ; "07" ) ; "-"  
 " ; SI (EXTRAE (C2;7;2) <> "99" ; EXTRAE (C2;7;2) ; "15" ) )

## #4 CALCULAR LA EDAD – COLUMNA « H »

# Hacer la diferencia entre las dos fechas y expresarlo en años

=SIFECHA (G2;F2;"Y")

## #5 CONVERTIR LA EDAD EN GRUPOS DE EDAD – COLUMNA « I »

#5.1 Si la edad está entre 0 y 4 entonces el grupo de edad es igual a 1

# Si la edad es entre 5 y 9 entonces el grupo de edad es igual a 2, etc...

=BUSCAR (H2;{0;5;10;15;20;25;30;35;40;45;50;55;60;65;70;75;80;85;115}  
 ;{1;2;3;4;5;6;7;8;9;10;11;12;13;14;15;16;17;18})

#5.2 Si la fecha de nacimiento es "9999", entonces el grupo de edad se establece en "19"

=SI (EXTRAE (G2;1;4) ="9999" ; 19 ; BUSCAR (H2;{0;5;10;15;20;25;30;35;40;45;  
 50;55;60;65;70;75;80;85;115};{1;2;3;4;5;6;7;8;9;10;11;12;13;14;15;16;  
 ;17;18}))

## #6 CREAR EL AÑO DE INCIDENCIA – COLUMNA « J »

# Recuperar los primeros 4 dígitos de la variable "INCID"

=VALOR (EXTRAE (A2;1;4) )

## #7 CREAR LA VARIABLE DE CODIGO ICD – COLUMNA « K »

# Recuperar los primeros 3 dígitos de la variable "I10"

=EXTRAE (D2;1;3)

## #8 CONVERTIR LOS CODIGOS ICD EN ETIQUETAS – COLUMNA « L »

#8.1 Crear una hoja denominada "ICD" con el archivo de datos "Datos-icd.txt"

#8.2 Reemplace cada código con su etiqueta correspondiente

`=BUSCAR (K2; ICD!$A$2:$A$99; ICD!$B$2:$B$99)`

## #9 CREAR UNA "CASES" VARIABLE – COLUMNA « M »

# Crear una variable "CASES" igual a "1" para cada fila

# Será necesario para futuros pasos

## #10 REEMPLAZAR LAS FORMULAS CON LOS VALORES REALES

# Copiar las columnas B, I, J, L y M luego en Pegado especial seleccionar la opción valores

# (es decir, "SEX", "AGEGRP", "YEAR", "CANCER" y "CASES")

## #11 ABRE EL ARCHIVO « Datos\_paso2.xlsx »

# Este archivo es el resultado de la manipulación anterior; Con sólo las columnas "SEX", "AGEGRP", "YEAR", "CANCER" y "CASES" restantes.

## #12 PREESTABLECER LOS DATOS AGREGADOS

#12.1 Crear una hoja denominada "aggreg", copiar y pegar en esta hoja todas las columnas con datos de la hoja "crude"

#12.2 En la hoja "aggreg", elimine las combinaciones duplicadas en todas las columnas (utilizar el comando « Quitar duplicados » está localizado en la pestaña « Datos », dentro del grupo « *Herramientas de datos* »)

#12.3 Ordenar los datos por "SEX", "YEAR", "CANCER" y "AGEGRP" (utilizar el menú "Datos" y luego "Ordenar")

## #13 CALCULAR LOS DATOS AGREGADOS – COLUMNA « E »

# Actualizar la variable “CASES” de la hoja “aggreg” con la suma de casos de cada combinación en la hoja “crude”

`=SUMAPRODUCTO ( (crude!$A$2:$A$102571=A2) * (crude!$B$2:$B$102571=B2) * (crude!$C$2:$C$102571=C2) * (crude!$D$2:$D$102571=D2) )`

## #14 REEMPLAZAR LA FORMULA CON LOS VALORES REALES

# Copiar la columna E luego en Pegado especial seleccionar la opción valores (es decir, "CASES")

## #14 ABRE EL ARCHIVO « Datos\_paso3.xlsx »

# Este archivo es el resultado de la manipulación anterior; Con sólo la hoja « aggreg » que tiene la variable "CASES" con valores reales.

## #15 ASIGNAR LA INFORMACION DE POBLACION – COLUMNA « F »

#15.1 Crear una hoja denominada "POP" con el archivo de datos "Datos-poblacion.txt"

#15.2 Combinar el valor de la variable "POP" con cada combinación de "SEX", "AGEGRP" y "YEAR"

=SUMA (SI (A2=POP!\$A\$2:\$A\$571;SI (B2=POP!\$B\$2:\$B\$571;SI (C2=POP!\$C\$2:\$C\$571;POP!\$D\$2:\$D\$571) ) ) )

#Tenga cuidado: Es un cálculo de producto matricial, entonces usted debe escribir CTRL-SHIFT-ENTER para validar la fórmula ! De lo contrario, "#N/A" aparece en la celda...

#Esto es lo que debería ver en la celda de la fórmula :

{=SUMA (SI (A2=POP!\$A\$2:\$A\$571;SI (B2=POP!\$B\$2:\$B\$571;SI (C2=POP!\$C\$2:\$C\$571;POP!\$D\$2:\$D\$571) ) ) ) }

#Una solución alternativa para la fórmula es la siguiente:

{=INDICE (POP!\$D\$2:\$D\$571;COINCIDIR (1; (A2=POP!\$A\$2:\$A\$571) \* (B2=POP!\$B\$2:\$B\$571) \* (C2=POP!\$C\$2:\$C\$571) ; 0) ) }

#####

## Tabla de traducción de fórmulas Excel en español a inglés

CONCATENAR	CONCATENATE
EXTRAER	MID
SI	IF
FECHA	DATE
SIFECHA	DATEDIF
BUSCAR	LOOKUP
BUSCARV	VLOOKUP
VALOR	VALUE
SUMA	SUM
SUMAPRODUCTO	SUMPRODUCT
PENDIENTE	SLOPE
EXP	EXP

LN	LN
SUMAR.SI	SUMIF
INDICE	INDEX
COINCIDIR	MATCH

Problemas potenciales:

-) punto y coma “;” en lugar de coma “,” para separar los parámetros

-) EXTRAER = MED en las versiones anteriores de EXCEL

```
#####
```

```
## EJERCICIO 3 – PREPARACION DE DATOS EN R
```

```
#####
```

```
# ESTABLECER EL DIRECTORIO DE TRABAJO
```

```
setwd("D:/LMICourse/Ejercicios/3-Manipulacion")
```

```
# LEER EL ARCHIVO DE DATOS
```

```
dat = read.csv(file="Ejercicio3_datos.csv")
```

```
# CALCULAR LA FECHA DE INCIDENCIA
```

```
> dat$INCID[substr(dat$INCID,5,8)=="9999"] <-  
paste(substr(dat$INCID[substr(dat$INCID,5,8) ==  
"9999"],1,4),"0715",sep="")
```

```
dat$INC_DATE = strptime(dat$INCID, format = "%Y%m%d")
```

```
# CALCULAR LA FECHA DE NACIMIENTO
```

```
> dat$BIRTHD[substr(dat$BIRTHD,5,8)=="9999"] <-  
paste(substr(dat$BIRTHD[substr(dat$BIRTHD,5,8) ==  
"9999"],1,4),"0715",sep="")
```

```
dat$BTH_DATE = strptime(dat$BIRTHD, format = "%Y%m%d")
```

```
# CALCULAR LA EDAD
```

```
> dat$AGE =  
as.numeric(round(difftime(dat$INC_DATE,dat$BTH_DATE,units = "days")  
/ 365.25))
```

```
dat$AGE[substr(dat$BIRTHD,1,4)=="9999"] <- NA
```

```
#CONVERTIR LA EDAD EN GRUPOS DE EDAD
```

```
> dat$AGEGRP = as.integer(cut(dat$AGE, breaks=c(seq(0, 85,  
by=5),115), right=FALSE))
```

```
dat$AGEGRP[is.na(dat$AGEGRP)] <- 19
```

```
#CREAR EL AÑO DE INCIDENCIA
```

```
dat$YEAR = as.numeric(substr(dat$INCID,1,4))
```

```
#CREAR LA VARIABLE DE CODIGO ICD
```

```
dat$ICD = substr(dat$I10,1,3)
```

```
#CONVERTIR LOS CODIGOS ICD EN ETIQUETAS
```

```
> icd = read.table(file="Datos-icd.txt", header=TRUE,  
stringsAsFactors=FALSE)
```

```
dat = merge(x=dat, y=icd, by="ICD", all.x=TRUE)
```

```
dat$CANCER[is.na(dat$CANCER)] <- "OTHER"
```

```
#CREAR UNA "CASES" VARIABLE
```

```
dat$CASES = 1
```

```
#CALCULAR LOS DATOS AGREGADOS
```

```
datag = aggregate(CASES ~ SEX+AGEGRP+YEAR+CANCER, data=dat, sum)
```

```
#ORDENAR LOS DATOS POR "SEX", "YEAR", "CANCER" y "AGEGRP"
```

```
> datag = datag[order(datag$SEX, datag$YEAR, datag$CANCER,  
datag$AGEGRP),]
```

```
#ASIGNAR LA INFORMACION DE POBLACION
```

```
pop = read.table(file="Datos-poblacion.txt", header=TRUE)
```

```
> datag = merge(x=datag, y=pop, by=c("SEX", "AGEGRP", "YEAR"),  
all.x=TRUE, sort = TRUE)
```

```
#ESCRIBIR UN ARCHIVO DE DATOS DE SALIDA
```

```
> write.table(datag, "Ejercicio3_salida.csv", sep = ",", row.names =  
FALSE)
```