

# Estudio de la homogeneidad de la infección para los datos de España

## Instrucciones

En esta tarea vamos a realizar un estudio de la homogeneización de la infección por comunidades autónomas dentro de España. Es decir, a partir del test  $\chi^2$  vamos a ver si el virus se expandió por igual en todas las comunidades autónomas de España.

## Descripción de la tarea

Instrucciones de tareas

En el recurso descargable ficheros.zip tenéis dos ficheros:

- El archivo `datosCCAA.csv` donde estan los datos del número de infectados, hospitalizados, fallecidos y pacientes en las UCI (ver la variable o columna tipo) de las 17 comunidades autónomas (faltan Ceuta y Melilla) de España.
- El archivo `Fichero_poblaciónCCAA.csv` donde están los datos de la población o del número de habitantes de las 17 comunidades autónomas.

## Pregunta 1

En primer lugar, usando la función `select` del paquete `tidyverse`, transformar la tabla de datos anterior en una tabla de datos donde la primera columna sea la fecha, y a partir de la segunda columna hasta la última aparezcan el número de infectados de cada comunidad autónoma. Es decir debería aparecer una estructura como la que sigue:

fecha tipo Andalucía Aragón Asturias Baleares Canarias Cantabria **Castilla-La Ma...**Castilla y Leó...  
Cataluña

1 2020-02-27 casos 1 0 0 1 6 0 0 0 2

2 2020-02-28 casos 6 1 0 1 6 0 0 2 3

3 2020-02-29 casos 8 1 0 2 6 0 0 2 5

4 2020-03-01 casos 12 0 1 2 7 1 1 3 6

5 2020-03-02 casos 12 0 1 2 7 10 3 3 15

6 2020-03-03 casos 13 0 1 2 7 10 7 8 15

donde faltan 7 columnas más correspondientes a las comunidades autónomas de Valencia, Extremadura, Galicia, Madrid, Murcia, Navarra y País Vasco.

Indicación: primero filtrar los datos seleccionando las variables de interés y en un segundo filtrado, usar la opción `pivot_wider` para obtener lo que se pide.

## Solución

Primero tenéis que cargar los datos del fichero `DatosCCAA.csv`:

```
datos = read.csv("DatosCCAA.csv")
```

Luego cargamos el paquete `tidyverse` y filtramos los datos con la opción `pivot_wider`:

```
library(tidyverse)

datos.filtrados= datos %>% select(fecha,CCAA,total,tipo)

datos.filtrados2= datos.filtrados %>%
  pivot_wider(names_from=CCAA,values_from=total)

# Como sólo nos interesan el número de infectados, hacemos lo siguiente:

tabla.casos = datos.filtrados2[datos.filtrados2$tipo=="casos",]
```

## Pregunta 2

Usando el paquete `ggplot2`, realizar un gráfico del número de infectados donde cada comunidad autónoma aparezca de un color distinto.

## Solución

Primero cargamos el paquete:

```
library(ggplot2)
```

El gráfico que nos piden es el siguiente:

```
x=as.Date(tabla.casos$fecha)

ggplot(tabla.casos, aes(x)) +
  geom_line(aes(y=tabla.casos$Andalucía, colour="Andalucía")) +
  geom_line(aes(y=tabla.casos$Aragón, colour="Aragón")) +
  geom_line(aes(y=tabla.casos$Asturias, colour="Asturias")) +
  geom_line(aes(y=tabla.casos$Balears, colour="Balears")) +
  geom_line(aes(y=tabla.casos$Canarias, colour="Canarias")) +
  geom_line(aes(y=tabla.casos$"Castilla-La Mancha", colour="Castilla-La Mancha")) +
  geom_line(aes(y=tabla.casos$"Castilla y León", colour="Castilla y León")) +
  geom_line(aes(y=tabla.casos$Cataluña, colour= "Cataluña")) +
  geom_line(aes(y=tabla.casos$"C. Valenciana", colour= "C. Valenciana")) +
  geom_line(aes(y=tabla.casos$Extremadura, colour= "Extremadura")) +
  geom_line(aes(y=tabla.casos$Galicia, colour= "Galicia")) +
  geom_line(aes(y=tabla.casos$Madrid, colour= "Madrid")) +
  geom_line(aes(y=tabla.casos$Murcia, colour= "Murcia")) +
  geom_line(aes(y=tabla.casos$Navarra, colour= "Navarra")) +
  geom_line(aes(y=tabla.casos$"País Vasco", colour= "País Vasco")) +
  xlab("Fecha") + ylab("Frecuencias")
```

## Pregunta 3

Para el día **15 de marzo de 2020** y suponiendo que la infección se ha expandido por igual en todas las comunidades autónomas hallar el **número de infectados esperado y empírico** para todas las comunidades autónomas.

## Solución

Primero fijamos la fecha:

```
fecha="2020-03-15"
```

Las frecuencias empíricas son:

```
frecuencias.empiricas=tabla.casos[tabla.casos$fecha==fecha & tabla.casos$tipo=="casos",]
```

Quitamos los dos primeros resultados ya que no son numéricos

```
frecuencias.empiricas=frecuencias.empiricas[c(-1,-2)]
```

Para calcular las frecuencias teóricas, necesitamos la población de cada CCAA. Primero leemos el fichero Fichero\_poblaciónCCAA.csv:

```
fichero.poblaciones = read.csv("Fichero_poblaciónCCAA.csv")
```

Las poblaciones de cada CCAA serán:

```
poblaciones = fichero.poblaciones$Pob_CCAA_2019
```

A continuación calculamos la población total española y los infectados totales:

```
población.total=sum(poblaciones)
```

```
infectados.totales = sum(frecuencias.empiricas)
```

Las frecuencias teóricas o esperadas serán:

```
frecuencias.teóricas = infectados.totales*poblaciones/población.total
```

#### Pregunta 4

Realizar el test  $\chi^2$  correspondiente para testear si nuestra hipótesis de infección por igual es cierta o no y comentar los resultados obtenidos.

#### Solución

El test chi2 será el siguiente:

```
chisq.test(frecuencias.empiricas,p=frecuencias.teóricas/sum(frecuencias.teóricas))
```

Observamos que el p-valor es muy pequeño.

Por tanto, concluimos que el virus no se expandió por igual en todas las comunidades autónomas el día 15 de marzo de 2020.

#### Pregunta 5

Realizar las preguntas 3) y 4) pero para el **día 4 de abril**.

#### Solución

Basta seguir todos los pasos de las respuestas a las preguntas 3 y 4 cambiando la variable fecha por:

```
fecha="2020-04-04"
```

#### Pregunta 6

Realizar las preguntas 1), 2), 3) y 4) pero en lugar de considerar el número de infectados, considerar el **número de fallecidos**.

### Solución

Basta seguir todos los pasos anteriores cambiando la tabla `tabla.casos` en la respuesta de la pregunta 1 por

```
tabla.fallecidos = datos.filtrados2[data_wide_filtrado$tipo=="fallecidos",]
```

y en lugar de `tabla.casos` escribir `tabla.fallecidos`.

### Pregunta 7

Realizar las preguntas 1), 2), 3) y 4) pero en lugar de considerar el número de infectados, considerar el **número de hospitalizados**.

### Solución

Basta seguir todos los pasos anteriores cambiando la tabla `tabla.casos` en la respuesta de la pregunta 1 por

```
tabla.hospitalizados = datos.filtrados2[data_wide_filtrado$tipo=="hospitalizados",]
```

y en lugar de `tabla.casos` escribir `tabla.hospitalizados`.