

Serie temporal de turismo en Países Bajos 1990-2022

Turismo nacional y extranjero en campings, aparcamientos de autocaravanas y similares

Ágatha del Olmo Tirado

20/02/2024

Introducción

Este trabajo se centra en analizar los datos de alojamiento turístico en los Países Bajos desde 1990 hasta 2022 obtenidos del Eurostat, específicamente enfocándose en el total de turistas, tanto nacionales como extranjeros, que se han alojado en aparcamientos de caravanas, campings y similares. Este estudio es especialmente interesante teniendo en cuenta que esta es una nación caracterizada por sus canales famosos y pintorescos paisajes, convirtiéndose en un destino ideal para explorar en caravana.

Para estudiar el comportamiento de la serie vamos a analizar por separado cada uno de sus componentes: tendencia, estacionalidad, ciclo, intervención y residuo. Tras cargar las librerías necesarias y fechar la base de datos, hemos cortado la serie hasta diciembre 2019 para evitar el efecto del Covid sobre el estudio general.

```
dataCompleta <- read.csv2("./Netherlands_TO_OT.csv",  
                           header = TRUE)
```

```
dataCompleta <- ts(dataCompleta[, 1],  
                   start = 1990,  
                   frequency = 12)
```

```
data <- window(dataCompleta, end = c(2019, 12))
```

Después de preparar los datos, podemos ver el esquema de la serie temporal, que puede ser aditiva o multiplicativa. Las series aditivas implican que cada componente suma su efecto sobre las demás, $y_t = T_t + S_t + C_t + I_t + R_t$, en cambio, las series multiplicativas suponen un incremento porcentual sobre el resto de componentes, $y_t = T_t \cdot S_t \cdot C_t \cdot I_t \cdot R_t$.

Figura 1. Turismo en Países bajos 1990-2019

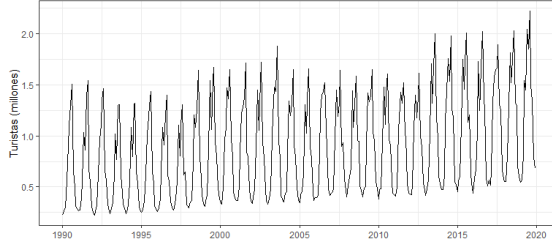
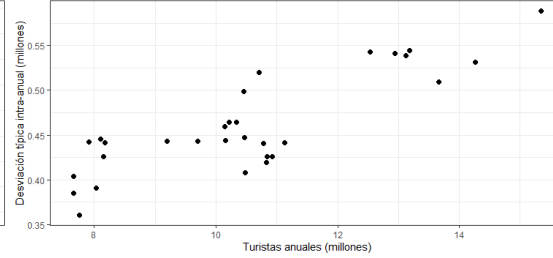


Figura 2. Dispersión de turismo frente a su desviación típica



Al no quedar del todo claro el esquema de la serie en la primera figura, analizamos una segunda, donde es evidente que la serie es de esquema multiplicativo, es decir, a mayor número de turistas hay una mayor desviación típica o dispersión. Conocer cuál es su esquema es esencial para saber qué métodos podemos usar y cómo interpretarlos.

Tendencia

La tendencia es el comportamiento de la serie a lo largo del tiempo (siempre a más de un año). Agregamos la serie de mensual a anual, de forma que el gráfico tiene menos ruido y queda más clara su tendencia.

Figura 3. Tendencia de la serie de turismo Países Bajos 1990-2019



Vemos una tendencia creciente general. En concreto, la primera década estuvo estancada en los 8 millones de turistas, pero en 1997 se disparan las cifras y se mantiene en los 11 millones durante la siguiente década, hasta 2012, que crece estrepitosamente hasta el 2019.

Para explicar esta tendencia hay dos posibles causas simples en cuanto al turismo nacional. Para empezar, aproximadamente ha crecido en población el país en 2.5 millones de personas desde 1990 hasta 2019, de forma que por razones de medidas absolutas es fácil que el turismo nacional haya crecido en número. Además, el PIB per cápita ha subido unos 30000€ durante el mismo periodo, de forma que el turismo interno ha podido crecer gracias a la capacidad económica y el tiempo libre del neerlandés moderno. De todas formas, necesitaríamos más información para saber si

el turismo interno representa la suficiente proporción sobre el total como para que estos eventos tengan efecto.

Estacionalidad

La estacionalidad es el cambio repetitivo o cíclico de un comportamiento de la serie (a menos de un año). Para estudiar la estacionalidad de forma rigurosa debemos descomponer la serie en las diferentes partes que la componen, así podremos extraerla.

Hay diversos métodos de obtención de la descomposición en RStudio: `tapply`, `decompose` (medias móviles) y `stl` (regresiones locales ponderadas). La primera opción siempre quedará en la sombra frente a las otras dos ya que es la más pobre o descuidada, la segunda se utiliza con series sin intervención o datos atípicos y puede usarse con tanto esquemas aditivos como multiplicativos, y la última sí puede tener en cuenta la intervención pero solo es válida para esquemas aditivos (sin realizar “trucos”).

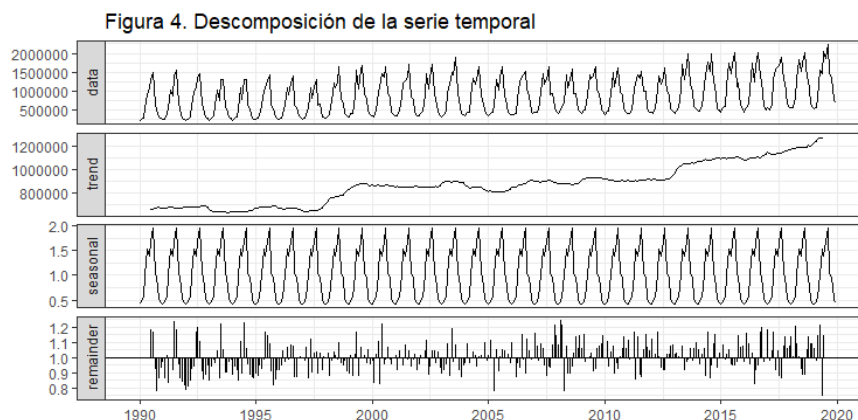
Previamente habíamos previsto el problema de la intervención eliminando los últimos años que contenían el efecto del Covid, de forma que podemos usar sin problemas la opción de `decompose`, especificando que es de tipo multiplicativo. Comprobamos que la hemos realizado correctamente sumando la estacionalidad de cada mes, que debe ser igual al orden, en este caso 12 (meses del año).

```
dataDes <- decompose(data,
                      type = "mult")
round(dataDes$figure, 2)

[1] 0.42 0.47 0.56 1.00 1.53 1.39 1.66 1.94 1.06 0.95 0.54 0.46

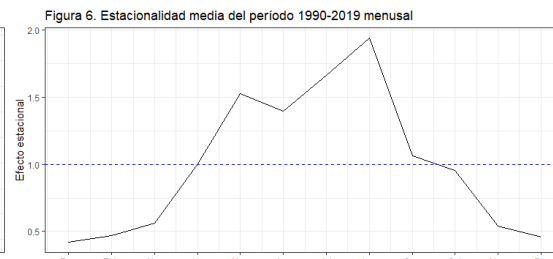
sum(dataDes$figure)

[1] 12
```



Este gráfico nos muestra cada uno de los componentes descompuestos de la serie de forma graficada. Ahora mismo nos centramos en el tercero, la estacionalidad.

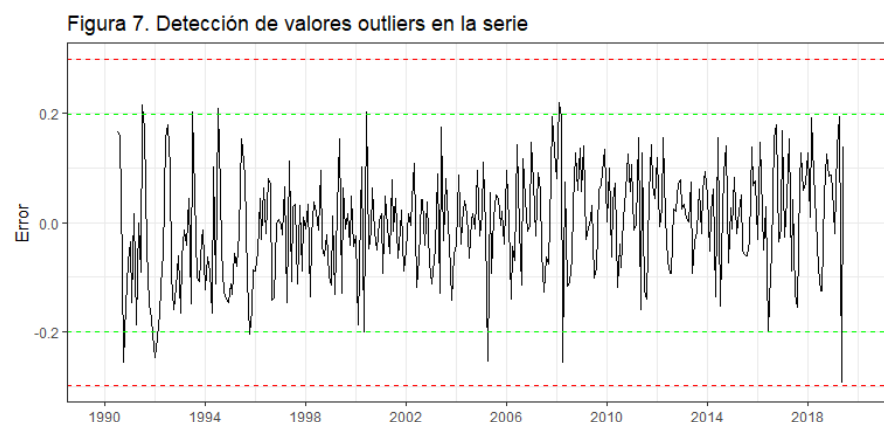
Las componentes estacionales representan cuánto se desvía cada mes de la media anual y están expresados alrededor de 1. Destacan los meses de invierno - especialmente enero donde sobrepasa la mitad de lo habitual- con el menor número de turistas y verano -particularmente agosto donde se duplica- con el mayor turismo. Podemos verlo de forma más evidente en los siguientes gráficos:



Intervención y residuo

La intervención y el residuo (componente estocástica de la serie) se estudian de forma conjunta, y se puede medir si la causa de una subida o bajada aparentemente notoria en la serie es causa de la aleatoriedad del error o de una intervención. El método consiste en que cuando un error sobrepasa mínimo 3 veces su desviación típica, se considera un valor *outlier*.

```
error <- log(remainder(dataDes))
sderror <- sd(error, na.rm = TRUE)
```



```
time(error)[!is.na(error) & abs(error) > 2.9 * sderror]
[1] 2019.333
```

Como vemos, no hay ningún error que sobrepase 3 veces su desviación típica (cabe mencionar que de haber usado la serie completa habría una clara intervención

provocada por el Covid), pero en el gráfico vemos que al final de la serie hay un mes en el que bajan notoriamente las visitas al país a pesar de no llegar a pasar de la línea roja. Para saber exactamente cuándo ocurre, en vez de sobrepasar 3 veces la desviación, pedimos que sobrepase 2.9 veces, y R detecta abril de 2019.

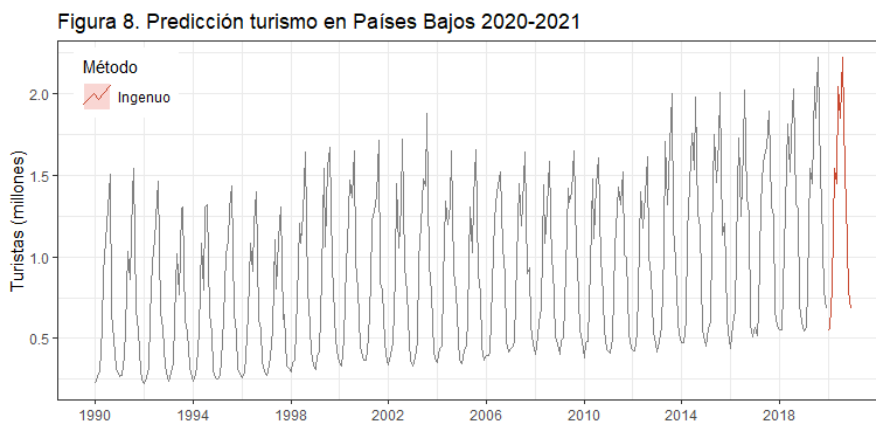
Esta bajada de turismo en los Países Bajos puede darse a un evento trágico que ocurrió la mañana del 18 de marzo de 2019: el atentado de Utrecht, donde tres personas murieron y siete quedaron gravemente heridas. Este atentado se consideró el peor ataque terrorista islámico que el país había vivido hasta el momento, y el nivel de amenaza de la ciudad subió a grado 5. Siendo Utrecht la cuarta ciudad más importante del país, es comprensible que el efecto que tuvo el atentado sobre las visitas al país las siguientes semanas fuera notorio.

Métodos sencillos de predicción

Tras estudiar las características de la serie podemos predecir con el método sencillo más adecuado, y si algo tenemos muy claro del turismo en Países Bajos es que hay una fuerte componente estacional. Esto implica que debemos usar el método ingenuo con estacionalidad.

Este método consiste en usar la última observación del mismo mes que haya disponible para predecir el siguiente, $\hat{y}_{T+h} = y_{T-m(k+1)}$. Por ejemplo, para predecir junio 2022 se usaría junio 2019 ya que es el último dato disponible en nuestra serie.

```
snaivedata <- snaive(data, h = 12)
```



Tras predecir, es interesante medir si el modelo que hemos utilizado nos prevé de unas buenas predicciones o no.

Evaluación de las predicciones

Un buen modelo no debe tener sesgo, debe tener una buena calidad de ajuste, unos intervalos de confianza fidedignos y debe mejorar respecto del método más sencillo.

- Sesgo: ME y MPE -> menor que 1% no tiene sesgo
- Calidad de ajuste: RMSE, MAPE y MAE -> han de acercarse a 0
- Intervalos de confianza: ACF1 (correlación entre \hat{e}_t y \hat{e}_{t-1}) -> menor que 0.1
- Mejora: MASE -> si es mayor que 1 sería preferible un método más sencillo

```
accuracy(snaivedata)
```

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	21377.64	127584.6	84734.29	1.897998	9.120059	1	-0.1737397

En cuanto al sesgo, vemos que hay un 1.9%, de forma que sí que hay algo de sesgo, y al ser positivo sabemos que es sesgo por abajo, es decir, las predicciones realizadas tienden a ser sistemáticamente menores que los valores reales.

Respecto a la calidad de ajuste, se equivoca por 127584 visitantes según el RMSE, y según el MAE 84734 (MAE siempre es más pequeña y no comparable al RMSE), es decir, un 9.12%, lo cual representa bastante parte del total.

Acerca de los intervalos de confianza, no nos podemos fiar, ya que el ACF1 está por encima del 0.1 en valor absoluto, de forma que deberíamos usar bootstrapping.

Por último, el MASE es justo 1 porque se compara consigo mismo, ya que el método más sencillo en este caso es precisamente el ingenuo con estacionalidad.

Origen de predicción móvil

Para mejorar la estimación de la calidad, podemos obtener la medida MAPE (error porcentual absoluto medio), usando el método de origen de predicción móvil, también conocido como validación cruzada (*cross validation*). Este método consiste en ajustar y predecir para una parte de la muestra (ajusta con el training set y evalúa con el test set) que se desplaza hacia delante y repite el proceso k veces.

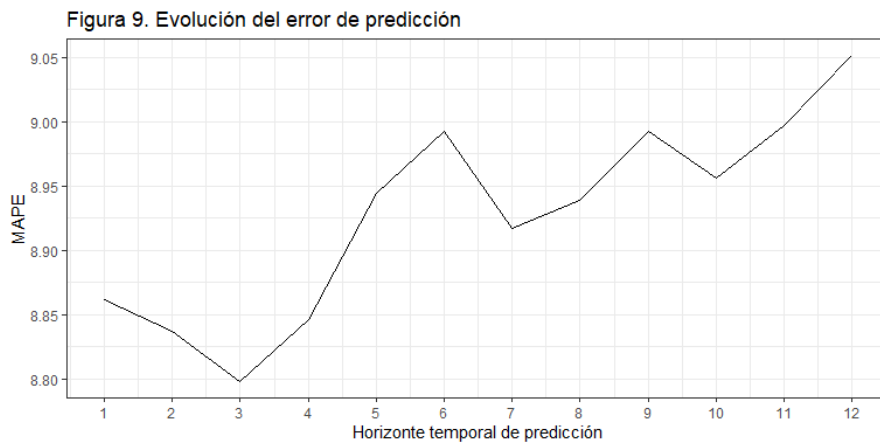
```
k <- 144
h <- 12
TT <- length(data)
s <- TT - k - h

mapeSnaive <- matrix(NA, s + 1, h)

for (i in 0:s) {
  train.set <- subset(data, start = i + 1, end = i + k)
  test.set <- subset(data, start = i + k + 1, end = i + k + h)
  fit <- snaive(train.set, h = h)
  mapeSnaive[i + 1,] <- 100*abs(test.set - fit$mean)/test.set
}

mapeSnaive <- colMeans(mapeSnaive)
round(mapeSnaive, 2)
```

```
[1] 8.86 8.84 8.80 8.85 8.94 8.99 8.92 8.94 8.99 8.96 9.00 9.05
```



A pesar de que posteriormente vayamos a calcular el error real cometido, de esta forma tenemos una idea de cómo sería nuestro error en predicciones extramuestrales sin que hubiera afectado el Covid, lo cual es muy útil ya que no esperamos un nuevo evento similar próximamente.

Vemos que los errores a corto plazo son muy similares a los de largo plazo, a 1 mes vista, se equivoca un 8.86%, mientras que a 12 meses vista, se equivoca en un 9.05%. Pero parece que tiende hacia arriba, de forma que cuanto más lejano esté el horizonte temporal, parece que la predicción será peor.

Predicción vs Covid

Para obtener el efecto del Covid, restamos la predicción que hemos realizado ignorando el Covid a lo que debiera haber pasado (la serie real).

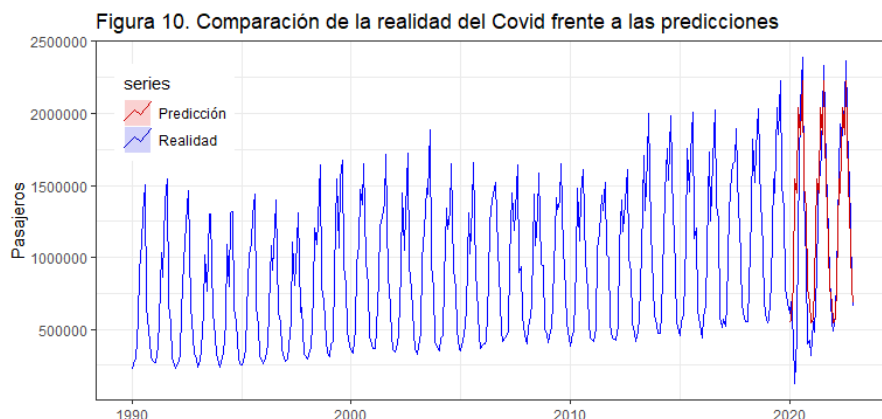
```
snaivedata_nocovid <- snaive(data, h = 36)

aggregate(dataCompleta - snaivedata_nocovid$mean, FUN = sum)/1000000

Time Series:
Start = 2020
End = 2022
Frequency = 1
[1] -4.209511 -1.086880 1.102863
```

Los resultados están expresados en millones: en 2020 hubo 4 millones de turistas menos de los que se esperaba, en 2021 un millón menos, y en 2022, se recupera con un millón de turistas por encima de lo esperado. Esta mejora en 2022 probablemente se da al hecho de que el método ingenuo con estacionalidad no tiene en cuenta que la serie tenía una tendencia creciente, de forma que era de esperar que tras recuperarse del Covid, las cifras de turismo fueran a ser mayores de lo previsto.

Para ver si respecto al resto de los años el efecto del Covid es considerable podemos verlo graficado de forma más clara:



Como vemos, el descenso del Covid no es muy exagerado, no parece que su efecto fuera muy notorio en los Países Bajos. De hecho, el año 2020 se asemeja bastante a las cifras de los años 90, mientras que en algunos países europeos la pandemia causó estragos record muy por debajo de los históricos. Debe tenerse en cuenta que si solo consideráramos los turistas extranjeros, la diferencia sería probablemente mayor.

Conclusión

Se destaca un crecimiento constante del turismo en Países Bajos desde 2019 hasta la actualidad, una estacionalidad marcada con más turismo en verano, y una intervención destacada en abril de 2019 por el atentado de Utrecht. Este trabajo subraya la necesidad de comprender y abordar las fluctuaciones del turismo para la toma de decisiones informadas como la mejora de infraestructuras destinadas al turismo en caravanas y campings, y la implementación de políticas de marketing con el fin de atraer un mayor número de visitantes al país.

Webgrafía

Wikipedia contributors. (2023, September 27). 2019 in the Netherlands. Wikipedia.https://en.wikipedia.org/wiki/2019_in_the_Netherlands

Simcox, R. (2019, March 25). The Netherlands' luck is running out. Foreign Policy. <https://foreignpolicy.com/2019/03/25/the-netherlands-luck-is-running-out/>

Guía completa markdown y su integración con R. (n.d.). https://rstudio-pubs-static.s3.amazonaws.com/330387_5a40ca72c3b14824acedceb7d34618d1.html

¿Qué tan rápido se recupera el turismo del COVID-19? (2022, April 6). Foro Económico Mundial. <https://es.weforum.org/agenda/2022/04/que-tan-rapido-se-recupera-el-turismo-de-covid-19/>