(*Brevis*) Introducción al análisis de series temporales

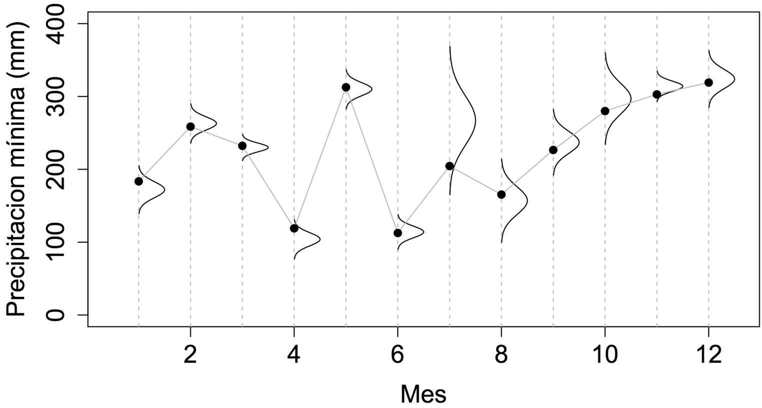
Antonio J. Pérez-Luque

ajperez@ugr.es // ajperez@go.ugr.es

Análisis de datos de seguimiento en la Red de Parques Nacionales

¿Qué es una **serie temporal**?

- Conjunto de observaciones registradas a intervalos regulares de tiempo.
- En cada instante t_i la observación proviene de una variable que puede tener igual o diferente distribución
- El orden de llegada de los datos es importante

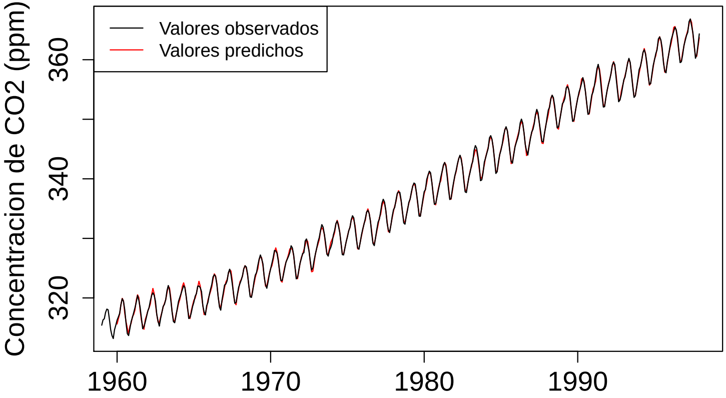


¿Por qué nos interesa analizar una **serie temporal**

- Para conocer como se repite a lo largo del tiempo (ciclo)
- Analizar si los datos se ajustan bien a un modelo teórico
- Analizar fenómenos espacio-temporales (avanzado)

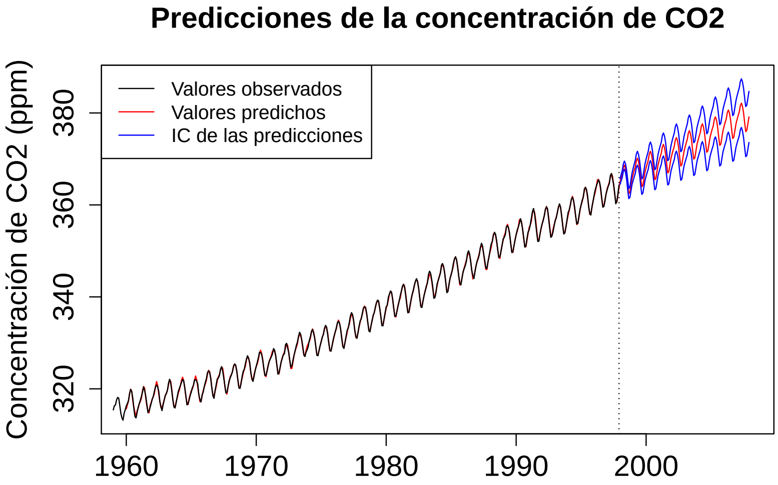
¿Por qué nos interesa analizar una **serie temporal**

- Explicar la evolución de un fenómeno a lo largo del tiempo

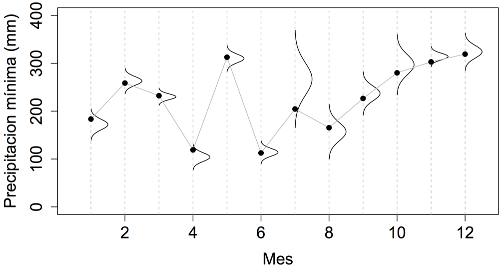


¿Por qué nos interesa analizar una **serie temporal**

- Predecir su valores en el futuro

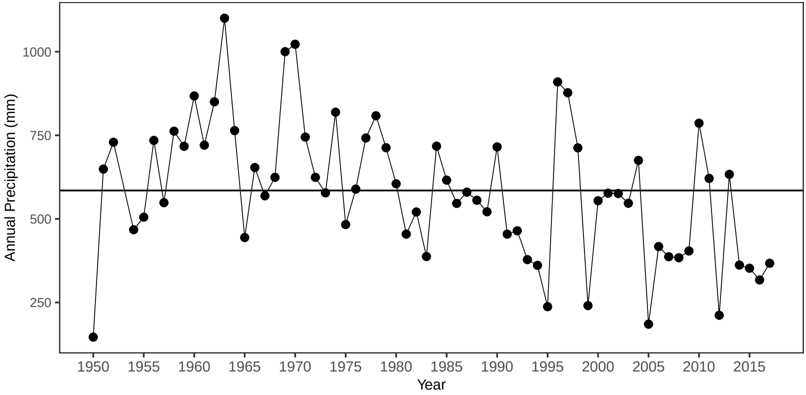


- Falta de información sobre las distribuciones subyacentes (cada una con sus paráametros) de las variables que analizamos. Para cada variable solamente disponemos de un dato observado.



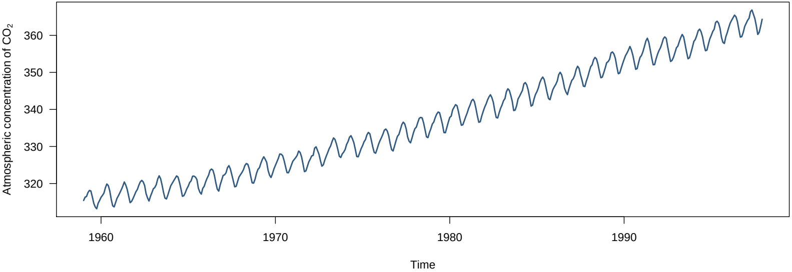
- **Solución**: Imponer condiciones a la serie de datos.

Tipos: Estacionarias

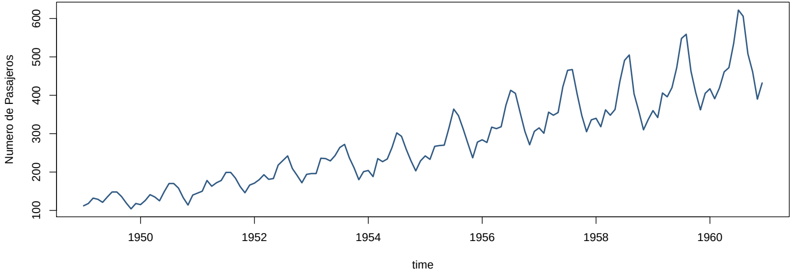


- Los datos varían todo el tiempo alrededor del mismo **valor medio** y con la **misma variabilidad**
- La relación entre las observaciones en dos momentos del tiempo diferentes sólo depende del número de observaciones que distan entre ambas

No se cumplen las condiciones de estacionariedad



- Serie no estacionaria estacional con **tendencia**



- Aumenta la variabilidad con el tiempo (serie no estacionaria con tendencia y aumento de variabilidad)

Trabajo con series temporales en R

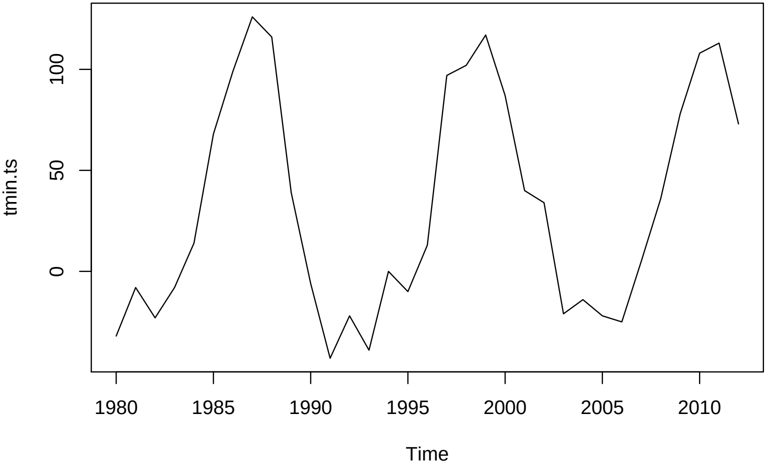
- Existen diferentes funciones para analizar series temporales en R https://cran.r-project.org/web/views/TimeSeries.html
- Objeto **ts**: Tipo especial de objeto dentro de R para las series temporales

```r
clima <- read.csv(here::here("/datos/temp_guadarrama.csv"),
                  header=TRUE, sep=";")
names(clima)[2] <- "YEAR"

# Seleccionar datos de la estacion Navacerrada
navacerrada <- subset(clima, NOMBRE == "NAVACERRADA,PUERTO")

# Seleccionar variables
tmin <- navacerrada[, c("YEAR", "TM_MIN")]

# Crear un objeto de tipo ts
tmin.ts <- ts(data = tmin$TM_MIN,
              start = min(tmin$YEAR),
              end = max(tmin$YEAR))
```

```
tmin.ts
```

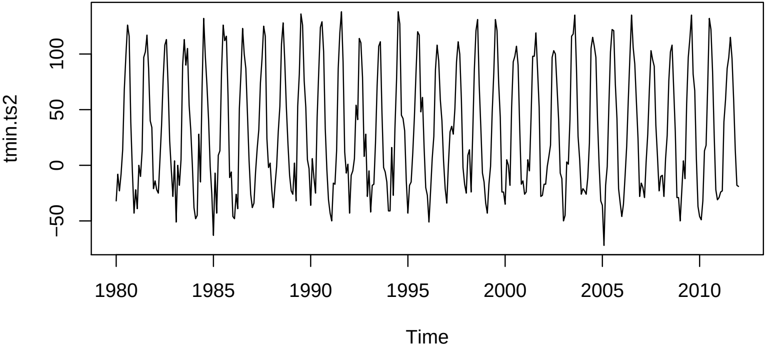
```
Time Series:
Start = 1980
End = 2012
Frequency = 1
 [1] -32  -8 -23  -8  14  68  99 126 116  39  -6 -43 -22 -39   0
[18]  97 102 117  87  40  34 -21 -14 -22 -25   5  36  78 108 113
```

Trabajo con series temporales en R

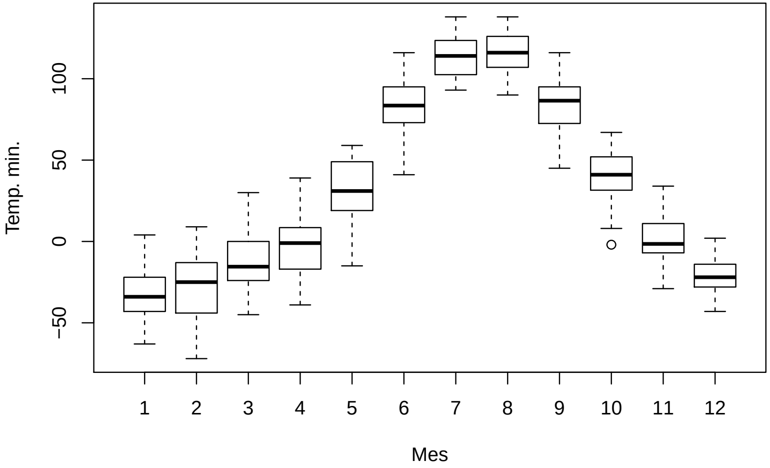


```
tmin.ts2 <- ts(data = tmin$TM_MIN,
               start = min(tmin$YEAR),
               end = max(tmin$YEAR),
               frequency = 12)
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1980	-32	-8	-23	-8	14	68	99	126	116	39	-6	-43
1981	-22	-39	0	-10	13	97	102	117	87	40	34	-21
1982	-14	-22	-25	5	36	78	108	113	73	22	-4	-28
1983	4	-51	0	-18	5	90	113	90	105	54	32	0
1984	-38	-48	-45	28	-15	70	132	98	72	41	-1	-26
1985	-63	-7	-43	9	13	84	126	112	116	63	-11	-6
1986	-46	-48	-26	-39	51	83	123	100	87	46	4	-26
1987	-38	-34	-6	15	32	74	96	125	116	24	-2	2
1988	-22	-38	-18	-1	30	53	109	128	94	52	19	-9
1989	-23	-26	2	-32	53	82	136	126	76	52	5	-3
1990	-36	6	-11	-25	43	85	124	129	102	33	-4	-31
1991	-43	-50	-16	-17	13	87	119	138	93	12	-7	1
1992	-43	-9	-5	6	54	41	114	110	79	8	28	-28
1993	-5	-42	-18	-17	20	71	107	111	47	-2	-6	-15
1994	-41	-41	16	-27	33	77	138	127	45	42	31	-14
1995	-43	-18	-15	11	43	83	120	117	48	61	16	-20
1996	-27	-51	-23	6	26	86	108	94	59	40	4	-21
1997	-34	1	30	35	28	51	93	111	100	59	-2	-17
1998	-25	9	14	-24	28	86	121	131	72	34	-7	-15



tmin.ts2
−50
0
50
100
1980
1985
1990
1995
2000
2005
2010
Time


Temp. min.
100
50
0
−50
1
2
3
4
5
6
7
8
9
10
11
12
Mes

Aproximaciones al estudio de las series temporales

- Modelos lineales (deterministas)
- Alisados exponenciales
- Extracción de señales
  - Descomposicion de una serie temporal
  - LOESS (*local polynomial regression fitting*)
- Modelos paramétricos ARIMA (Autoregressive Integrated Moving Average):
  - modelos AR (Autoregressive)
  - MA (Moving Averages)
  - ARMA (Autoregressive Moving Averages) para series estacionarias
  - Modelos integrados para las series no estacionarias

Descomposición aditiva de una serie temporal (extracción de señales)

$$y_t = T_t + S_t + I_t$$

- Tendencia $(T - t)$ representa un movimiento suave a lo largo del tiempo que puede ser constante o variable
- Estacionalidad (S_t) (*Seasonality*) supone una oscilación dependiente de la estación
- Componente irregular (I_t) lo componen variaciones aleatorias no explicadas por las otras componentes

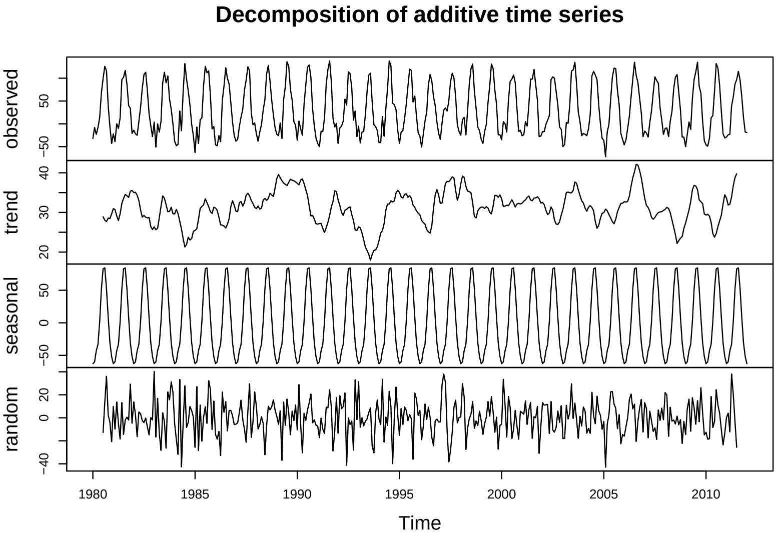
Procedimiento:

1. Se extrae la tendencia y se calculan los residuos (observaciones - tendencia). Los residuos son la serie sin tendencia, que contiene la estacionalidad y el componente irregular.
2. Se estima la estacionalidad de la serie y se resta a la serie sin tendencia, se obtiene la **serie desestacionalizada**. La serie desestacionalizada no debe contener ninguna estructura aparente y debe variar en torno a un valor constante, que es el **componente irregular**.
3. La predicción de la serie se realiza agregando al valor medio del componente irregular la predicción de tendencia y el componente estacional.

```r
d.tmin <- decompose(tmin.ts2)
str(d.tmin)
```

```
List of 6
 $ x       : Time-Series [1:385] from 1980 to 2012: -32 -8 -23 -
 $ seasonal: Time-Series [1:385] from 1980 to 2012: -62.8 -59.9
 $ trend   : Time-Series [1:385] from 1980 to 2012: NA NA NA NA
 $ random  : Time-Series [1:385] from 1980 to 2012: NA NA NA NA
 $ figure  : num [1:12] -62.8 -59.9 -42.9 -33.1 1.8 ...
 $ type    : chr "additive"
 - attr(*, "class")= chr "decomposed.ts"
```

- Valores predichos para la estacionalidad (`$seasonal`), la tendencia (`$trend`) y la componente irregular (`$random`)
- Valores promedios estimados para la componente estacional (`$figure`).


Decomposition of additive time series
observed
trend
seasonal
random
Time
1980
1985
1990
1995
2000
2005
2010
−50
50
20
30
40
0
−40

Alisados exponenciales

- Es una aproximación determinista al tratamiento de series temporales
- Permiten predecir nuevos valores de la serie
- Están basados en modelos paramétricos deterministas que se ajustan a la evolución de la serie
- Estos modelos permiten ajustar niveles y comportamientos tendenciales y estacionales que evolucionan en el tiempo, de manera que las observaciones más recientes tienen más peso en el ajuste que las más alejadas
- Tipos:
  - Alisado exponencial simple (no tendencia; no estacionalidad)
  - Alisado de Holt (tendencia; no estacionalidad)
  - Alisado de Holt-Winters (tendencia; estacionalidad)

```r
ordesa <- read.csv(file=here::here("datos/ordesa_ord.csv"))

# Seleccionar un municipio
panticosa <- subset(ordesa, Municipio == "Panticosa")
head(panticosa)
```

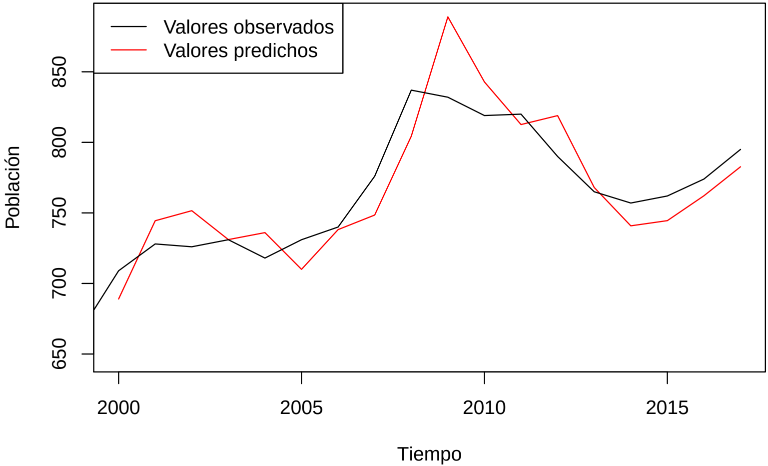
	Codigo	Municipio	valle	year	Habitantes
281	22170	Panticosa	TEN	1998	647
282	22170	Panticosa	TEN	1999	668
283	22170	Panticosa	TEN	2000	709
284	22170	Panticosa	TEN	2001	728
285	22170	Panticosa	TEN	2002	726
286	22170	Panticosa	TEN	2003	731

```
panticosa_ts <- ts(data = panticosa$Habitantes,
                   start= min(panticosa$year),
                   end = max(panticosa$year))
panticosa_holt <- HoltWinters(panticosa_ts, gamma=FALSE)

panticosa_holt
```

```
Holt-Winters exponential smoothing with trend and without season

Call:
HoltWinters(x = panticosa_ts, gamma = FALSE)

Smoothing parameters:
 alpha: 1
 beta : 0.7224186
 gamma: FALSE

Coefficients:
       [,1]
a 795.00000
```

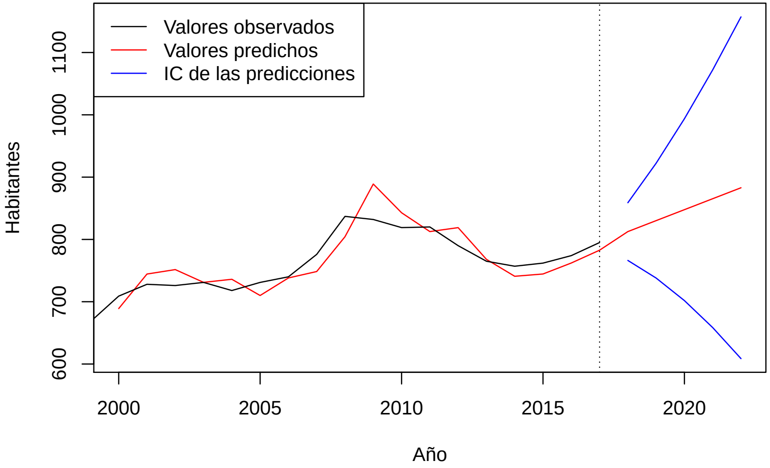
```
panticosa_holt
```

```
Holt-Winters exponential smoothing with trend and without season

Call:
HoltWinters(x = panticosa_ts, gamma = FALSE)

Smoothing parameters:
 alpha: 1
 beta : 0.7224186
 gamma: FALSE

Coefficients:
       [,1]
a 795.00000
b  17.58844
```


Valores observados
Valores predichos
Población
650
700
750
800
850
2000
2005
2010
2015
Tiempo

```r
panticosa_pred <- predict(panticosa_holt, n.ahead=5,
                              prediction.interval = TRUE)
panticosa_pred
```

```
Time Series:
Start = 2018
End = 2022
Frequency = 1
          fit       upr      lwr
2018 812.5884  858.8794 766.2975
2019 830.1769  922.3730 737.9808
2020 847.7653  993.7395 701.7912
2021 865.3537 1072.2463 658.4612
2022 882.9422 1157.2139 608.6705
```

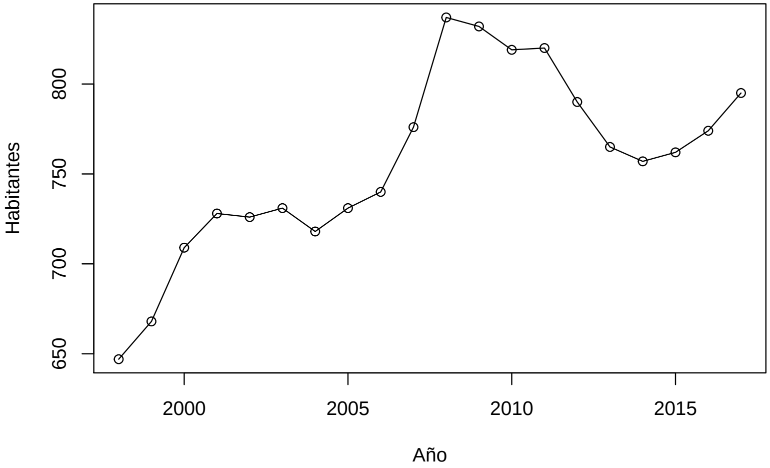

Valores observados
Valores predichos
IC de las predicciones
Habitantes
600
700
800
900
1000
1100
2000
2005
2010
2015
2020
Año

Análisis de tendencias

¿Ha aumentado significativamente la población de Panticosa en los últimos 20 años?

	Codigo	Municipio	valle	year	Habitantes
281	22170	Panticosa	TEN	1998	647
282	22170	Panticosa	TEN	1999	668
283	22170	Panticosa	TEN	2000	709
284	22170	Panticosa	TEN	2001	728
285	22170	Panticosa	TEN	2002	726
286	22170	Panticosa	TEN	2003	731

¿Ha aumentado significativamente la población de Panticosa en los últimos 20 años?



¿Ha aumentado significativamente la población de Panticosa en los últimos 20 años?

1. Análisis de tendencia (**test de Mann-Kendall**)
2. Magnitud del cambio (**Estimación de Theil-Sen Slope**)

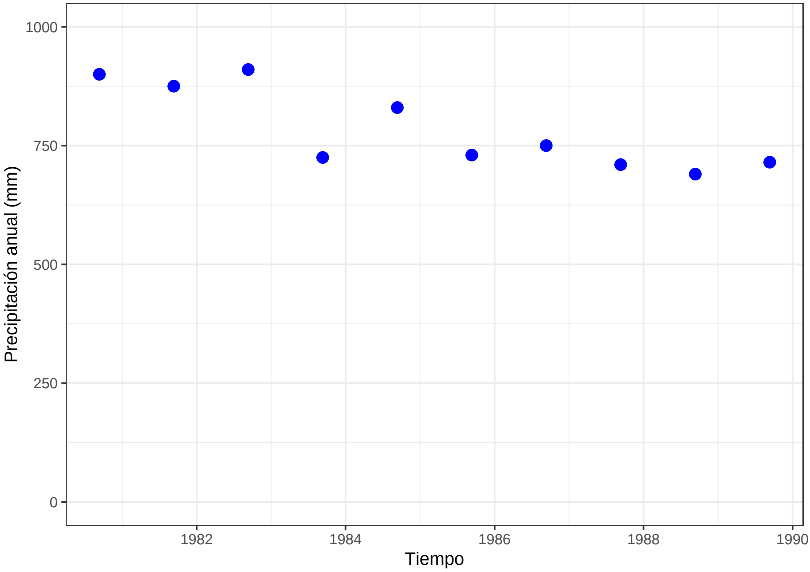
1. Análisis de tendencias: **Mann-Kendall test**

- Es un caso especial del test de correlación de Kendall (τ)
- Mide la asociación entre dos variables
- Basado en el ranking relativo de los datos (no sobre los datos en sí)
- Hipótesis:
  - H_0 : No existe una tendencia en los datos (los dataos son independientes y etán ordenados aleatoriamente)
  - H_1 : Existe una tendencia monótona (no necesariamente lineal)

Supongamos que tenemos un conjunto de observaciones de dos variable aleatorias, x (tiempo) e y (ej. Precipitación):

$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

year	prec
1980	900
1981	875
1982	910
1983	725
1984	830
1985	730
1986	750
1987	710
1988	690
1989	715



Precipitación anual (mm)
1000
750
500
250
0
1982
1984
1986
1988
1990
Tiempo

- Se calcula el valor de un estadístico **tau**

$$\tau = \frac{(\text{n pares concordantes}) - (\text{n pares discordantes})}{\frac{1}{2}n(n-1)}$$

Clasificamos los pares de observaciones (x_i, y_i) y (x_j, y_j) en:

- **concordantes** si los rangos de ambos elementos concuerdan
  - $x_i > x_j$ y $y_i > y_j$ o
  - $x_i < x_j$ y $y_i < y_j$
- **discordantes** cuando no concuerdan
  - $x_i > x_j$ y $y_i < y_j$ o
  - $x_i < x_j$ y $y_i > y_j$

Construcción del estadístico y del test

$$S = \sum_{i=1}^{n-1} \sum_{i+1}^{n} sign(x_{i+1} - x_i)$$

donde:

- $sign(x_{i+1} - x_i) = 1$ cuando $x_{i+1} - x_i > 0$
- $sign(x_{i+1} - x_i) = 0$ cuando $x_{i+1} - x_i = 0$
- $sign(x_{i+1} - x_i) = -1$ cuando $x_{i+1} - x_i < 0$

Un valor alto de S indica una tendencia creciente, mientras que un valor negativo indica una tendencia decreciente

Construcción del estadístico (demostración)

	year	prec
6	1985	730
7	1986	750
8	1987	710
9	1988	690
10	1989	715

- $715 > 690$ (**1**) // $715 > 710$ (**1**) // $715 < 750$ (-**1**) // $715 < 730$ (-**1**)
- $690 < 710$ (-**1**) // $690 < 750$ (-**1**) // $690 < 730$ (-**1**)
- $710 < 750$ (-**1**) // $710 < 730$ (-**1**)
- $750 > 730$ (**1**)

Construcción del estadístico (demostración)

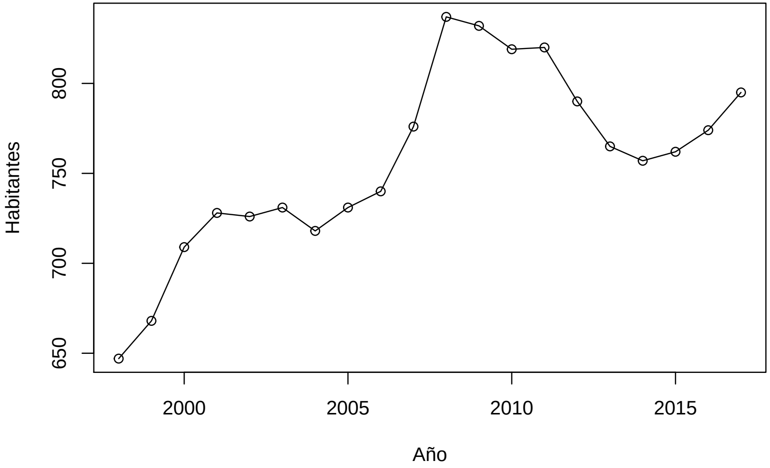
$$\tau = \frac{S}{\frac{1}{2}n(n-1)}$$

$$\tau = \frac{-4}{\frac{1}{2}5(5-1)} = -0.4$$

```r
library(trend)
mk.test(gg$prec)
```

```
    Mann-Kendall trend test

data:  gg$prec
z = -0.73485, n = 5, p-value = 0.4624
alternative hypothesis: true S is not equal to 0
sample estimates:
       S     varS      tau
-4.00000 16.66667 -0.40000
```

Panticosa



```
    Mann-Kendall trend test

data:  panticosa$Habitantes
z = 3.3111, n = 20, p-value = 0.0009294
alternative hypothesis: true S is not equal to 0
sample estimates:
          S        varS         tau
103.0000000 949.0000000   0.5435375
```

Contraste de Hipótesis para el parámetro $\hat{\tau}$

- Hipótesis $H_0 : \tau = 0$
- Calculamos el estadístico

$$z = \frac{S - sign(S)}{\sqrt{Var(S)}}$$

donde

$$Var(S) = \frac{n(n-1)(2n+5)}{18}$$

- Bajo la Hipótesis nula, el valor de z sigue aproximadamente una distribución normal

Algunas ventajas de utilizar este test

- El valor de $\hat{\tau}$ varía entre -1 y 1 (similar al coeficiente de correlación de Kendall)
- Informa de la dirección de la tendencia y de la magnitud
- Test estadístico muy utilizado en análisis de tendencias temporales para datos climatológicos, hidrológicos, evolución de población, etc.
- Es un test no-paramétrico que no requiere que los datos cumplan normalidad
- Se puede aplicar a datos estacionales, mensuales, etc (diferentes períodos de tiempo definidos por el usuario)
- Presenta una baja sensibilidad a los cambios abruptos en las series de datos no homogéneas

2. Theil-Sen Slope: Estimación de la tasa de cambio

- Estimación no paramétrica de la pendiente (tasa de cambio) para el intervalo analizado
- Método para ajustar una línea a un conjunto de puntos
- Mediana de las pendientes de todos los pares de puntos

$$\hat{\beta} = Median(\frac{Y_j - Y_i}{X_j - X_i})$$

```
    Sen's slope

data:  panticosa$Habitantes
z = 3.3111, n = 20, p-value = 0.0009294
alternative hypothesis: true z is not equal to 0
95 percent confidence interval:
  3.066667 10.214286
sample estimates:
Sen's slope
   5.755682
```

Consideraciones sobre el análisis de la tendencia

- Clasificación de tendencias. Ejemplo seguimiento de Aves (TRIM)

Table 9. Classification of the trend estimates. See text for details.

	Greater than 20% change in a 20-year period		Less than 20% change in a 20-year period	
	Significantly so	Not significantly so	Not significantly so	Significantly so
Significantly different from zero	(1) substantial decline or increase	(3) decline or increase	(3) decline or increase	(2) non-substantial decline or increase
Not significantly different from zero	(impossible)	(5) poorly known	(5) poorly known	(4) stable

Análisis de tendencias de parámetros climáticos en Sierra Nevada

- Variación temporal de la tendencia

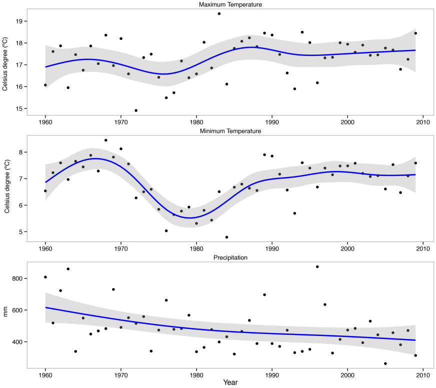
Tabla 1

		Píxeles		Píxeles significativos	
Variable	**Tendencia**	***n***	**%**	***n***	**%**
Precipitación	Positiva Negativa	298 171.460	0,17 99,79	0 74.516	0 43,37
Temperatura máxima	Positiva Negativa	141.757 29.551	82,50 17,19	23.417 0	13,63 0
Temperatura mínima	Positiva Negativa	129.759 41.762	75,52 24,30	7 0	<0,01 0

Resultados del análisis de las tendencias (test de *Mann-Kendall*) anuales en los últimos 50 años para la precipitación, la media de las temperaturas máximas y la media de las temperaturas mínimas. Para cada variable se muestran el número de pixeles (n) con tendencias negativas ($\tau < 0$) y positivas ($\tau > 0$) así como el número de pixeles significativos (p-valor $< 0{,}05$).

Consideraciones sobre el análisis de la tendencia

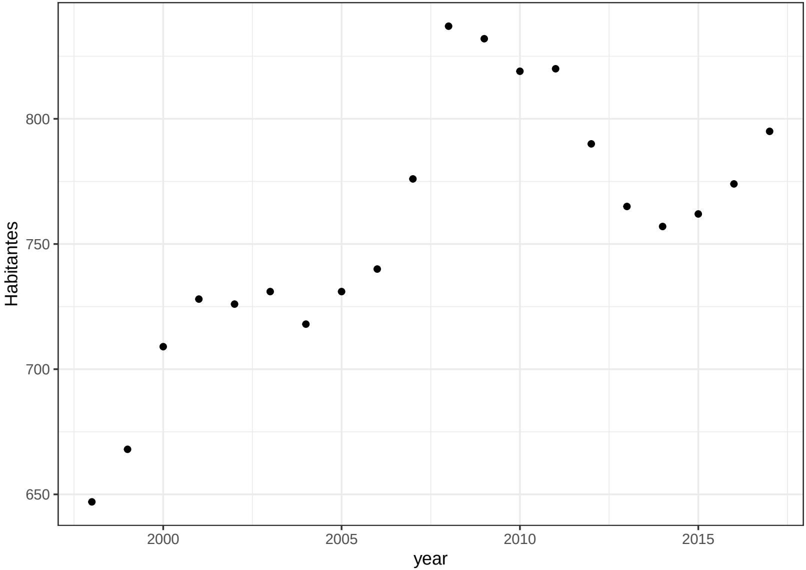
… o sobre la importancia de realizar un Análisis exploratorio de los datos



- Precaución al analizar tendencias globales
- Comparación de periodos, Modelos GAM

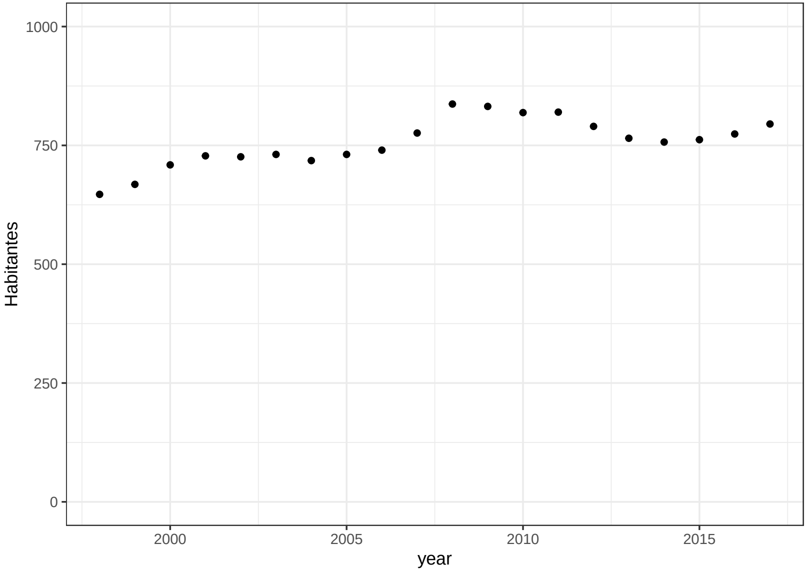
Consideraciones sobre el análisis de la tendencia

¡Ojo con el análisis exploratorio de datos!

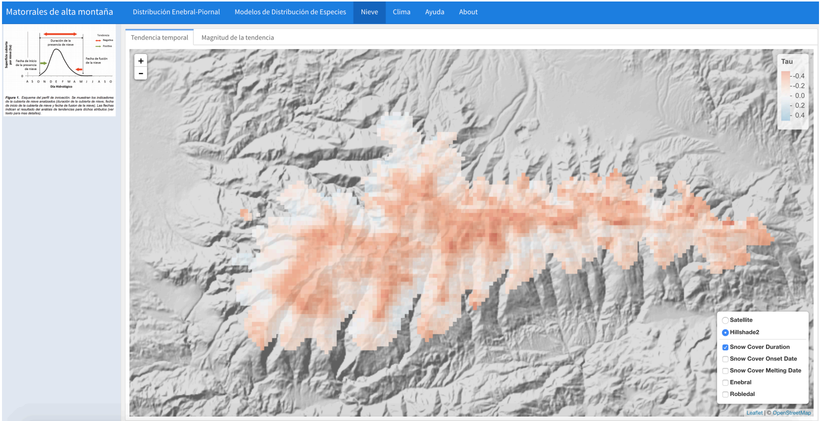


Consideraciones sobre el análisis de la tendencia

¡Ojo con el análisis exploratorio de datos!



- Análisis de tendencias cubierta de nieve y clima en SN https://sl.ugr.es/enebral_dist



Recursos (algunos)

- Falk M (2012). A First Course on Time Series Analysis - Examples with SAS. Chair of Statistics, University of Wurzburg. http://www.statistik-mathematik.uni-wuerzburg.de/wissenschaftforschung/time_series/the_book/
- CRAN Task View: Time Series Analysis https://cran.r-project.org/web/views/TimeSeries.html
- Helsel, D.R., and R. M. Hirsch. (2002). Statistical Methods in Water Resources. Techniques of Water Resources Investigations, Book 4, chapter A3. U.S. Geological Survey. http://pubs.usgs.gov/twri/twri4a3/pdf/twri4a3-new.pdf.