TECNICAS DE PRONOSTICOS

M.Sc. Christian W. Barreto Schuler

CONTENIDO

V. COVARIANZA, CORRELACION, AUTOCORRELACION, REGRESION LINEAL, TENDENCIAS, SIGNIFICANCIA

COVARIANZA

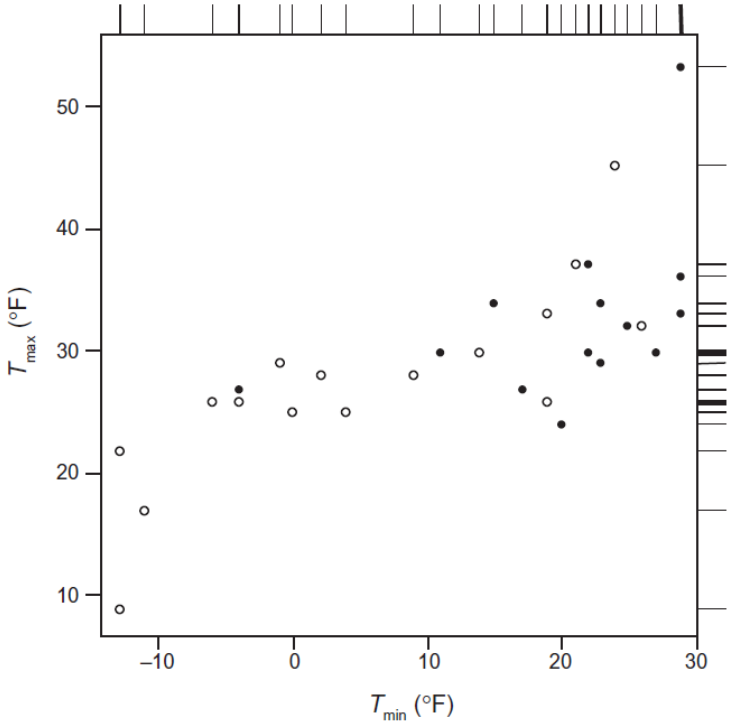


FIGURE 3.17 Scatterplot for daily maximum and minimum temperatures during January 1987 at Ithaca, New York. "Fringes" along the margins separately indicate the individual data distributions, with repeated data represented by heavier lines. Closed circles represent days with at least 0.01 in. of precipitation (liquid equivalent).

$$S_{X,Y} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{(n-1)}$$

Wilks, 2019

CORRELACION

Correlación de Pearson

Una forma de ver la correlación de Pearson es como la relación entre la covarianza muestral entre las dos variables, y el producto de las dos desviaciones estándar; donde los numeradores denotan anomalías, o resta de valores medios.

$$r_{x,y} = \frac{\mathrm{Cov}(x,y)}{s_x s_y} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}[(x_i-\bar{x})(y_i-\bar{y})]}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{1/2}\left[\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2\right]^{1/2}},$$

$$= \frac{\sum_{i=1}^{n}(x_i' y_i')}{\left[\sum_{i=1}^{n}(x_i')^2\right]^{1/2}\left[\sum_{i=1}^{n}(y_i')^2\right]^{1/2}}$$

Varianza (S^2)	$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$
Desviación estándar (S)	$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$

$$r_{x,y} = \frac{1}{n-1}\sum_{i=1}^{n}\left[\frac{(x_i-\bar{x})}{s_x}\frac{(y_i-\bar{y})}{s_y}\right] = \frac{1}{n-1}\sum_{i=1}^{n} z_{x_i} z_{y_i},$$

La correlación de Pearson es (casi) el producto promedio de las variables después de la conversión a anomalías estandarizadas.

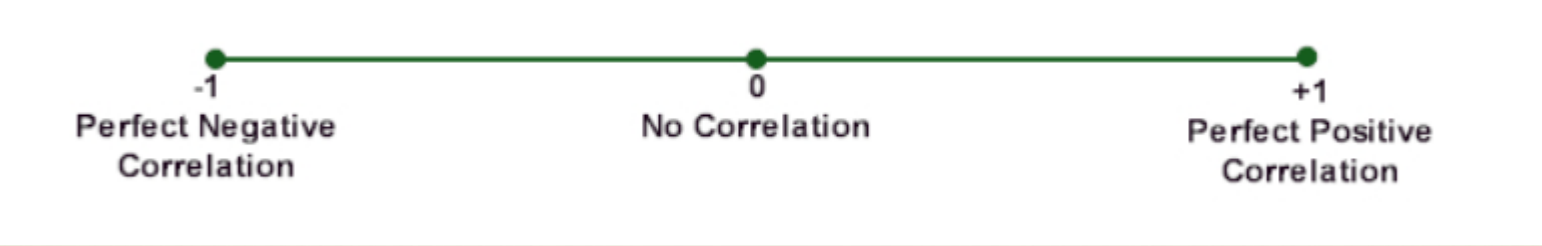
Wilks, 2019

CORRELACION

Correlación de Pearson

El Coeficiente de Correlación de Pearson (r), es un índice que mide el grado de relación entre 2 variables cuantitativas, pudiendo variar en el intervalo de -1 a +1.

$$-1 \leq r \leq 1$$



Otra de sus propiedades, es que el cuadrado de la correlación de Pearson (r^2), el cual especifica la proporción de la variabilidad de ya sea "*x*" o "*y*" que es linealmente descrita por el otro. (Su interpretación es engañosa).

CORRELACION

Correlación de Pearson

Limitaciones

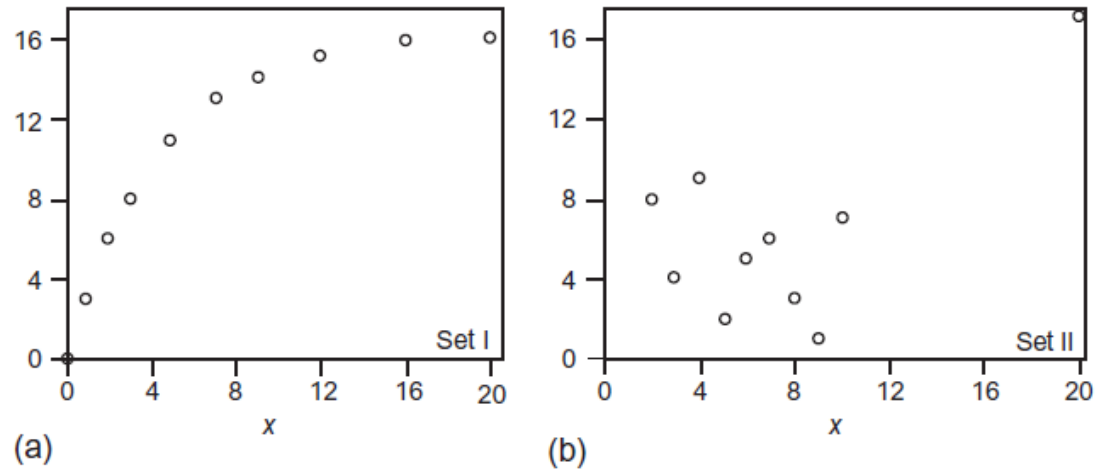


FIGURE 3.20 Scatterplots of the two artificial sets of paired data in Table 3.4. The Pearson correlation for the data in panel (a) (Set I in Table 3.4) of only 0.88 underrepresents the strength of the relationship, illustrating that this measure of correlation is not robust to nonlinearities. The Pearson correlation for the data in panel (b) (Set II) is 0.61, reflecting the overwhelming influence of the single outlying point, and illustrating lack of resistance.

TABLE 3.4 Artificial Paired Data Sets for Correlation Examples

Set I		Set II	
x	*y*	*x*	*y*
0	0	2	8
1	3	3	4
2	6	4	9
3	8	5	2
5	11	6	5
7	13	7	6
9	14	8	3
12	15	9	1
16	16	10	7
20	16	20	17

Wilks, 2019

CORRELACION

Correlación de rangos de Spearman

Existen alternativas disponibles, robustas y resistentes, al coeficiente de correlación de Pearson. El primero de ellos se conoce como **coeficiente de correlación de rango de Spearman**. La correlación de Spearman es simplemente el coeficiente de correlación de Pearson calculado utilizando los rangos de los datos.

$$r_{rank} = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)}, \tag{3.34}$$

where D_i is the difference in ranks between the ith pair of data values. In cases of ties, where a particular data value appears more than once, all of these equal values are assigned their average rank before computing the D_i's.

Wilks, 2019

Correlación Tau-Kendall

$$\tau = \frac{N_C - N_D}{n(n-1)/2}.$$

Wilks, 2019

CORRELACION

Significancia con p-valor

p-value: es la representación de la significancia estadística.

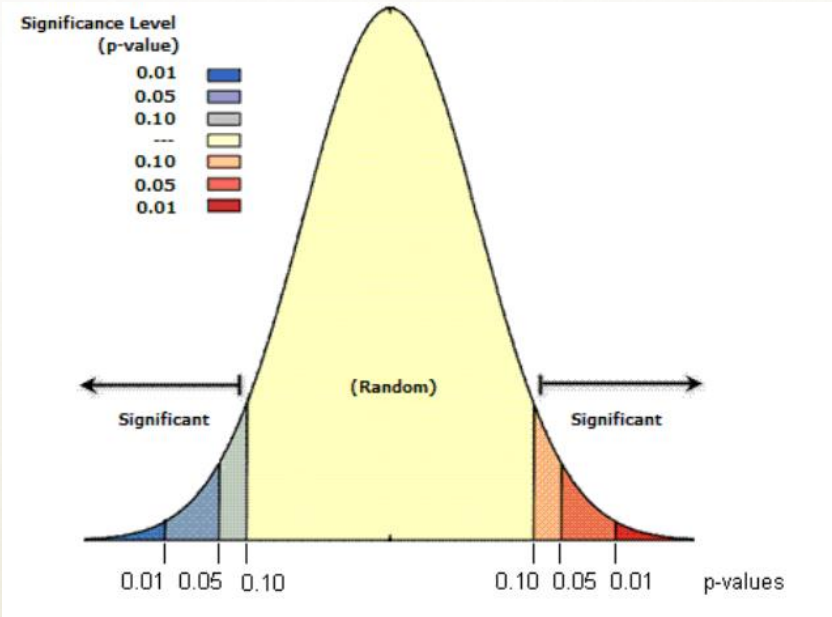
Normalmente, trabajamos con p-value < 0.05 para determinar significancias.

HIPOTESIS NULA (H_0)

CORRELACION NO ES SIGNIFICATIVA
p-value >= 0.05

HIPOTESIS ALTERNA(H_1)

CORRELACION ES SIGNIFICATIVA
p-value < 0.05



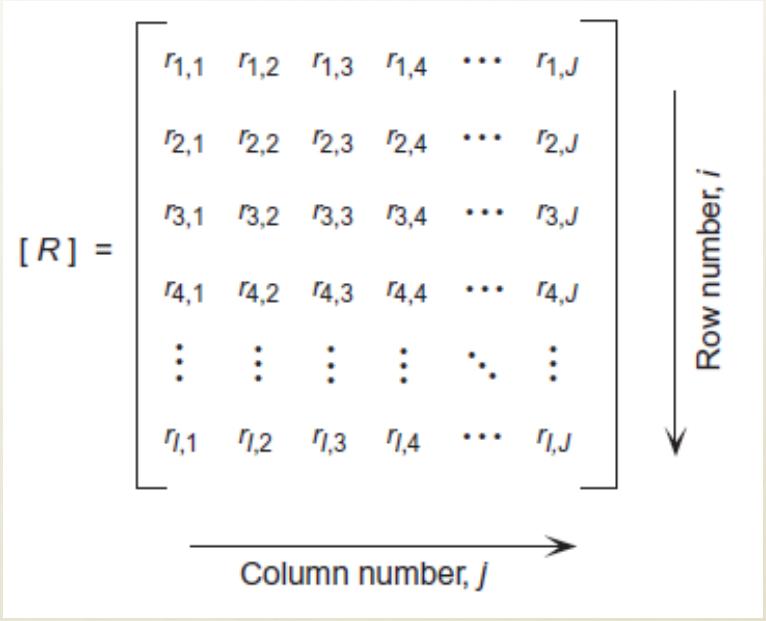
CORRELACION

Matriz de correlación

La matriz de correlación es un dispositivo muy útil para mostrar simultáneamente correlaciones entre más de dos lotes de datos emparejados. Por ejemplo, el conjunto de datos de la **Tabla A.1** contiene datos emparejados para seis variables. Los coeficientes de correlación se pueden calcular para cada uno de los 15 pares distintos de estas seis variables. En general, para K variables, hay (K) (K-1)/2 pares distintos, y las correlaciones entre ellos se pueden organizar sistemáticamente en una matriz cuadrada, con tantas filas y columnas como variables de datos coincidentes cuyas relaciones haya son resumidos. Cada entrada en la matriz, ri,j, está indexada por los dos subíndices, i y j, que apuntan a la identidad de las dos variables cuya correlación está representada. Por ejemplo, **r2,3** denotaría la correlación entre la segunda y tercera variables en una lista. Las filas y columnas de la matriz de correlación se numeran de manera correspondiente, de modo que las correlaciones individuales se organizan como se muestra en la Figura de abajo.

TABLE A.1 Daily Precipitation (in.) and Temperature (°F) Observations at Ithaca and Canandaigua, New York, for January 1987

	Ithaca			Canandaigua		
Date	Precipitation	Max Temp.	Min Temp.	Precipitation	Max Temp.	Min Temp.
1	0.00	33	19	0.00	34	28
2	0.07	32	25	0.04	36	28
3	1.11	30	22	0.84	30	26
4	0.00	29	−1	0.00	29	19
5	0.00	25	4	0.00	30	16
6	0.00	30	14	0.00	35	24
7	0.00	37	21	0.02	44	26
8	0.04	37	22	0.05	38	24
9	0.02	29	23	0.01	31	24
10	0.05	30	27	0.09	33	29
11	0.34	36	29	0.18	39	29
12	0.06	32	25	0.04	33	27
13	0.18	33	29	0.04	34	31
14	0.02	34	15	0.00	39	26
15	0.02	53	29	0.06	51	38
16	0.00	45	24	0.03	44	23
17	0.00	25	0	0.04	25	13
18	0.00	28	2	0.00	34	14
19	0.00	32	26	0.00	36	28
20	0.45	27	17	0.35	29	19
21	0.00	26	19	0.02	27	19
22	0.00	28	9	0.01	29	17
23	0.70	24	20	0.35	27	22
24	0.00	26	−6	0.08	24	2
25	0.00	9	−13	0.00	11	4
26	0.00	22	−13	0.00	21	5
27	0.00	17	−11	0.00	19	7



Wilks, 2019

CORRELACION

Matriz de correlación

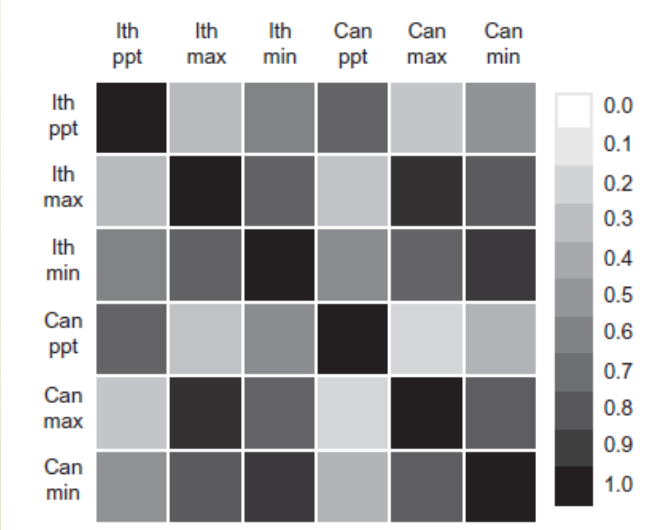
TABLE 3.5 Correlation Matrices for the Data in Table A.1

	Ith. Ppt	Ith. Max	Ith. Min	Can. Ppt	Can. Max	Ith. Ppt	Ith. Max	Ith. Min	Can. Ppt	Can. Max
Ith. Max	−0.024					0.319				
Ith. Min	0.287	0.718				0.597	0.761			
Can. Ppt	0.965	0.018	0.267			0.750	0.281	0.546		
Can. Max	−0.039	0.957	0.762	−0.015		0.267	0.944	0.749	0.187	
Can. Min	0.218	0.761	0.924	0.188	0.810	0.514	0.790	0.916	0.352	0.776

Only the lower triangle of the matrices is shown, to omit redundancies and the uninformative diagonal values. The left matrix contains Pearson product-moment correlations, and the right matrix contains Spearman rank correlations.

Wilks, 2019

Heatmaps



Wilks, 2019

CORRELACION

Scatterplot Matrix (Matriz de diagramas de dispersión)

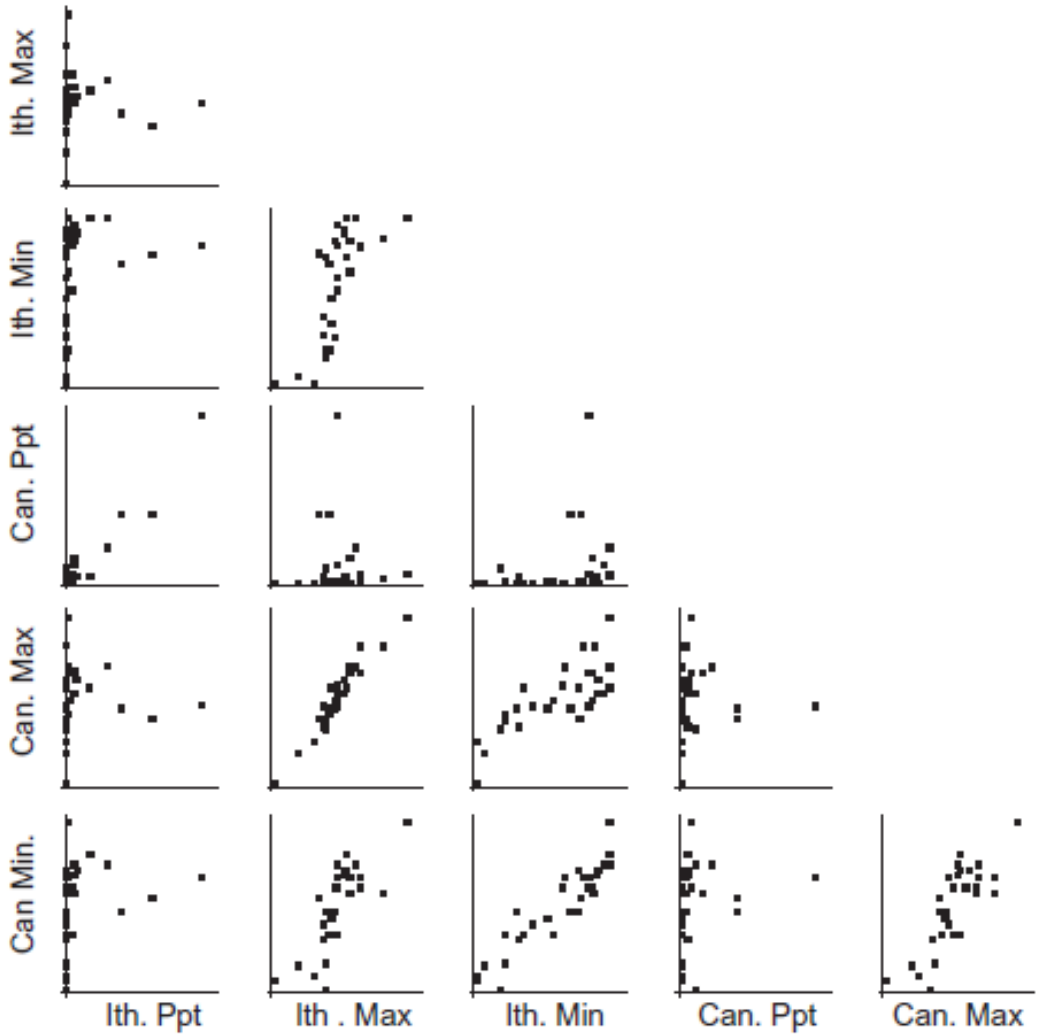


FIGURE 3.31 Scatterplot matrix for the January 1987 data in Table A.1 of Appendix A.

Wilks, 2019

CORRELACION

Mapas de correlación

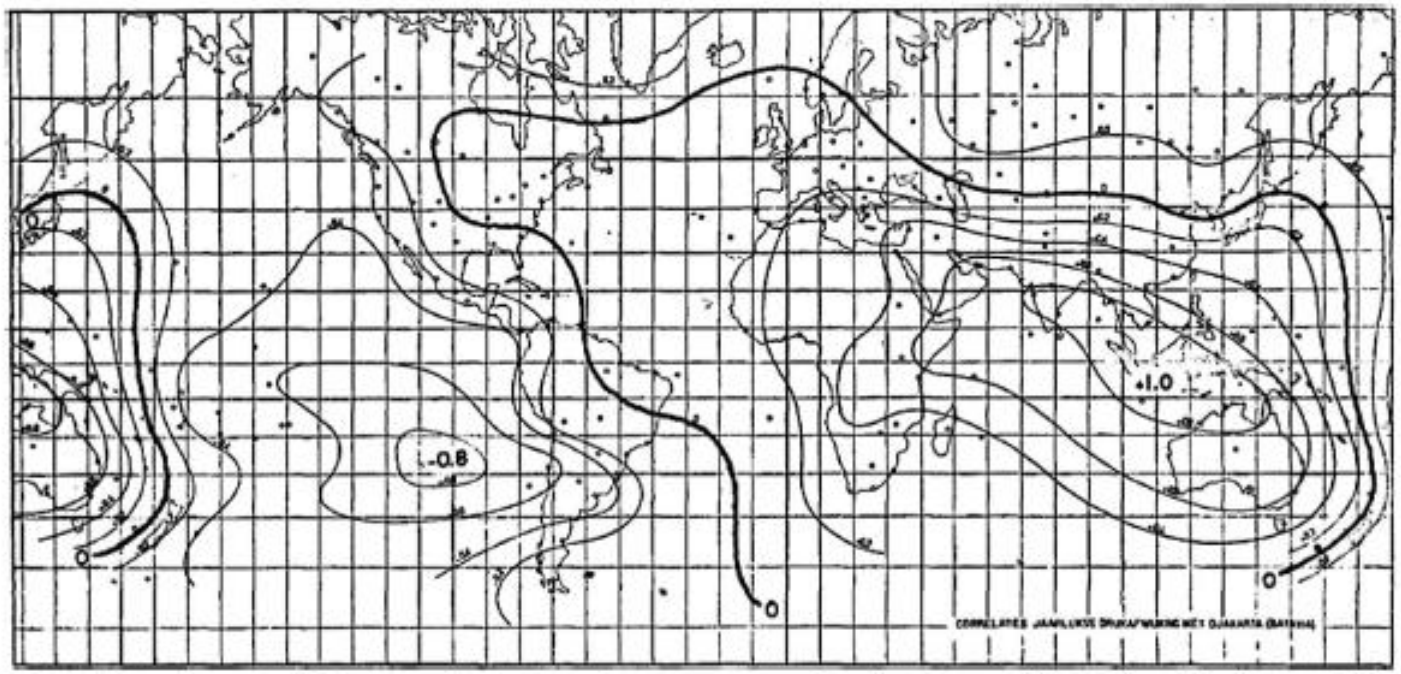


FIGURE 3.32 One-point correlation map of annual surface pressures at locations around the globe with those at Djakarta, Indonesia. The strong negative correlation of −0.8 at Easter Island reflects the atmospheric component of the El Niño-Southern Oscillation phenomenon. *From Bjerknes (1969). © American Meteorological Society. Used with permission.*

Wilks, 2019

PRACTICA

AUTOCORRELACION

La persistencia meteorológica, o la tendencia del clima en períodos sucesivos de tiempo a ser similar, se ilustró anteriormente en términos de probabilidades condicionales para los dos eventos discretos "precipitación" y "sin precipitación".

Para las variables continuas (por ejemplo, temperatura), la persistencia se caracteriza típicamente en términos de correlación en serie o autocorrelación temporal. El prefijo "auto" en autocorrelación denota la correlación de una variable consigo misma, de modo que la autocorrelación temporal indica la correlación de una variable con sus propios valores futuros y pasados.

A veces, estas correlaciones se denominan correlaciones con lag.

Casi siempre, **las autocorrelaciones se calculan como coeficientes de correlación de Pearson**, aunque no hay ninguna razón por la que no se puedan calcular también otras formas de correlación con *lag*.

Wilks, 2019

AUTOCORRELACION

El proceso de calcular las autocorrelaciones se puede visualizar imaginando que se escriben dos copias de una secuencia de valores de datos, con una de las series desplazada en una unidad de tiempo. Este cambio se ilustra en la Figura 3.21, utilizando los datos de temperatura máxima de Ithaca de enero de 1987.

Esta serie de datos se ha reescrito, con la parte media del mes representada por puntos suspensivos, en la primera línea. El mismo registro se ha vuelto a copiar en la segunda línea, pero se ha desplazado un día hacia la derecha.

Este proceso da como resultado 30 pares de temperaturas dentro de la caja, que están disponibles para el cálculo de un coeficiente de correlación.

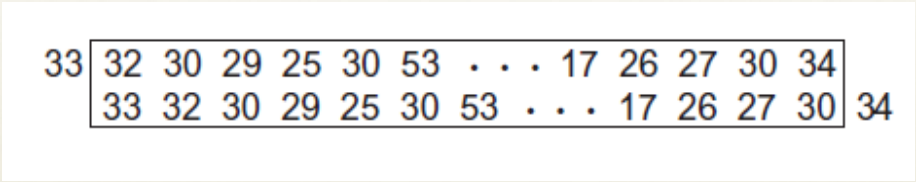


FIGURE 3.21 Construction of a shifted time series of January 1987 Ithaca maximum temperature data. Shifting the data by one day leaves 30 data pairs (enclosed in the box) with which to calculate the lag-1 autocorrelation coefficient.

Wilks, 2019

AUTOCORRELACION

Las autocorrelaciones se calculan sustituyendo los pares de datos rezagados en la fórmula de la correlación de Pearson.

$$r_1 = \frac{\sum_{i=1}^{n-1} [(x_i - \bar{x}_-)(x_{i+1} - \bar{x}_+)]}{\left[\sum_{i=1}^{n-1} (x_i - \bar{x}_-)^2\right]^{1/2} \left[\sum_{i=2}^{n} (x_i - \bar{x}_+)^2\right]^{1/2}}$$

Para la autocorrelación de retardo -1 (lag-1), hay n-1 pares. La única confusión real surge porque los valores promedios de las dos series serán, en general, ligeramente diferentes.

En la Figura anterior 3.21, por ejemplo, la media de los 30 valores encuadrados en la serie superior es 29,77°F, y la media de los valores encuadrados en la serie inferior es 29,73°F. Esta diferencia surge porque la serie superior no incluye la temperatura del 1 de enero y la serie inferior no incluye la temperatura del 31 Enero.

33	32	30	29	25	30	53	· · ·	17	26	27	30	34	
	33	32	30	29	25	30	53	· · ·	17	26	27	30	34

Wilks, 2019

AUTOCORRELACION

La autocorrelación de retardo-1 (lag-1) es la medida de persistencia más comúnmente calculada, pero a veces también es interesante calcular autocorrelaciones en retardos más largos. Conceptualmente, esto no es más difícil que el procedimiento para la autocorrelación de retardo-1, y computacionalmente la única diferencia es que las dos series están desplazadas en más de una unidad de tiempo.

Por supuesto, a medida que una serie de tiempo se desplaza cada vez más con respecto a sí misma, hay cada vez menos datos superpuestos con los que trabajar.

La ecuación siguiente es la generalización al coeficiente de autocorrelación lag-k.

$$r_k = \frac{\sum_{i=1}^{n-k}\left[(x_i-\bar{x}_-)(x_{i+k}-\bar{x}_+)\right]}{\left[\sum_{i=1}^{n-k}(x_i-\bar{x}_-)^2\right]^{1/2}\left[\sum_{i=k+1}^{n}(x_i-\bar{x}_+)^2\right]^{1/2}}.$$

$$r_k \approx \frac{\sum_{i=1}^{n-k}\left[(x_i-\bar{x})(x_{i+k}-\bar{x})\right]}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\sum_{i=1}^{n-k}(x_i x_{i+k}) - \frac{n-k}{n^2}\left(\sum_{i=1}^{n}x_i\right)^2}{\sum_{i=1}^{n}x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2}.$$

Wilks, 2019

AUTOCORRELACION

Función de autocorrelación

En conjunto, la colección de autocorrelaciones calculadas para varios ***lags*** se denomina función de autocorrelación.

A menudo, las funciones de autocorrelación se muestran gráficamente, con las autocorrelaciones trazadas como una función del retraso.

La Figura 3.22 muestra los primeros siete valores de la función de autocorrelación de muestra para los datos de temperatura máxima de Ithaca de enero de 1987. Una función de autocorrelación siempre comienza con r=1, ya que cualquier serie de datos no desplazada exhibirá una correlación perfecta consigo misma.

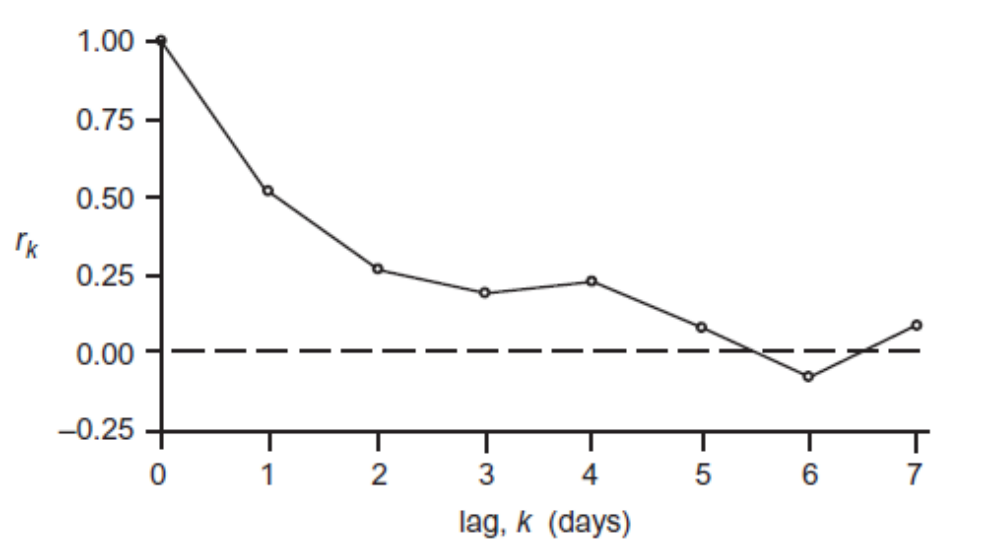


FIGURE 3.22 Sample autocorrelation function for the January 1987 Ithaca maximum temperature data. The correlation is 1 for $k=0$, since the unlagged data are perfectly correlated with themselves. The autocorrelation function decays to essentially zero for $k \geq 5$.

Wilks, 2019

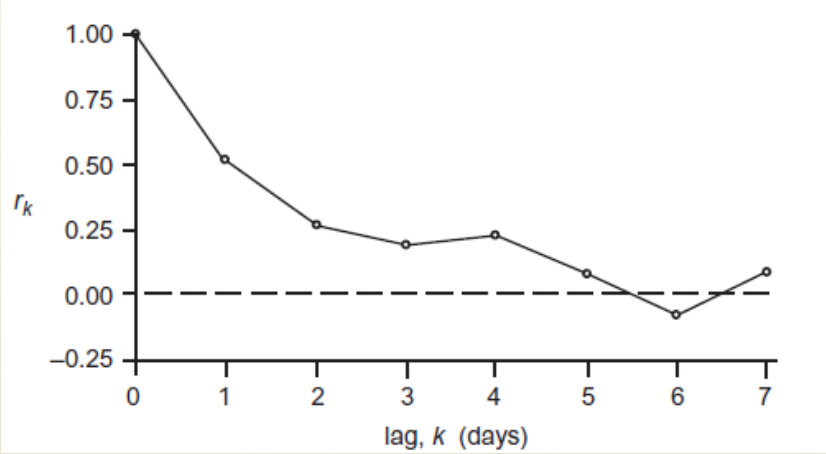
AUTOCORRELACION

Función de autocorrelación

Es típico que una función de autocorrelación muestre una disminución más o menos gradual hacia cero a medida que aumenta el retardo k, lo que refleja las relaciones estadísticas generalmente más débiles entre pares de puntos de datos más alejados entre sí en el tiempo.

Es instructivo relacionar esta observación con el contexto de la predicción meteorológica. Si la función de autocorrelación no decayera hacia cero después de unos días, sería muy fácil hacer pronósticos razonablemente precisos en ese rango:

"simplemente pronosticar con la observación de hoy (pronóstico de persistencia), o alguna modificación de la observación de hoy daría buenos resultados."



Wilks, 2019

PRACTICA

REGRESION LINEAL

Gran parte de la predicción meteorológica estadística se basa en el procedimiento conocido como regresión lineal por mínimos cuadrados.

La regresión se entiende más fácilmente en el caso de la regresión lineal simple, que describe la relación lineal entre dos variables, digamos x e y. Convencionalmente, el símbolo x se usa para la variable independiente o predictora, y el símbolo "y" se usa para la variable dependiente.

Los términos variable dependiente e independiente son comunes en la literatura de estadística y otras disciplinas, mientras que los términos predictor y predictante se utilizan principalmente en las ciencias atmosféricas y afines, habiendo aparentemente sido introducidos por Gringorten (1949).

REGRESION LINEAL

Esencialmente, la regresión lineal simple busca resumir la relación entre x e y, mostrada gráficamente en su diagrama de dispersión, usando una sola línea recta.

El procedimiento de regresión elige la línea que produce el menor error para las predicciones de y dadas las observaciones de x, dentro del conjunto de datos (x, y) utilizado para definir esa relación.

Exactamente lo que se define como el error mínimo puede depender del contexto, pero el criterio de error más común es la minimización de la suma (o, de manera equivalente, el promedio) de los errores al cuadrado.

La elección del criterio de error al cuadrado es la base del nombre regresión de mínimos cuadrados o regresión de Mínimos Cuadrados Ordinarios (MCO).

Es la minimización de la suma de los residuos al cuadrado lo que define la línea de mejor ajuste.

$$\hat{y} = a + bx, \tag{7.1}$$

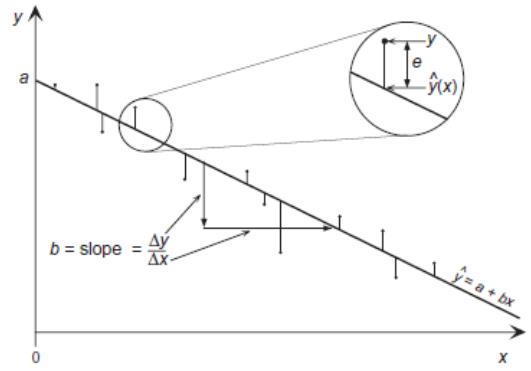


FIGURE 7.1 Schematic illustration of simple linear regression. The regression line, $\hat{y} = a + bx$, is chosen as the one minimizing some measure of the vertical differences (the residuals) between the points and the line. In least-squares regression that measure is the sum of the squared vertical distances. The inset shows a residual, e, as the difference between a data point and the regression line.

$$y_i = \hat{y}_i + e_i = a + bx_i + e_i,$$

Wilks, 2019

REGRESION LINEAL

$$y_i = \hat{y}_i + e_i = a + bx_i + e_i,$$

$$\sum_{i=1}^{n} (e_i)^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - [a + bx_i])^2,$$

$$\frac{\partial \sum_{i=1}^{n} (e_i)^2}{\partial a} = \frac{\partial \sum_{i=1}^{n} (y_i - a - bx_i)^2}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

$$\frac{\partial \sum_{i=1}^{n} (e_i)^2}{\partial b} = \frac{\partial \sum_{i=1}^{n} (y_i - a - bx_i)^2}{\partial b} = -2 \sum_{i=1}^{n} x_i [(y_i - a - bx_i)] = 0.$$

$$\sum_{i=1}^{n} y_i = na + b \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i y_i = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} (x_i)^2.$$

$$b = \frac{\sum_{i=1}^{n} [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$a = \bar{y} - b\bar{x}.$$

Wilks, 2019

REGRESION LINEAL

Medida de bondad de ajuste

R^2: Se puede interpretar como la proporción de la variación de la predicción y.

Para una regresión perfecta, $R^2=1$, y de manera opuesta, una regresión complatemente inútil, $R^2 =0$

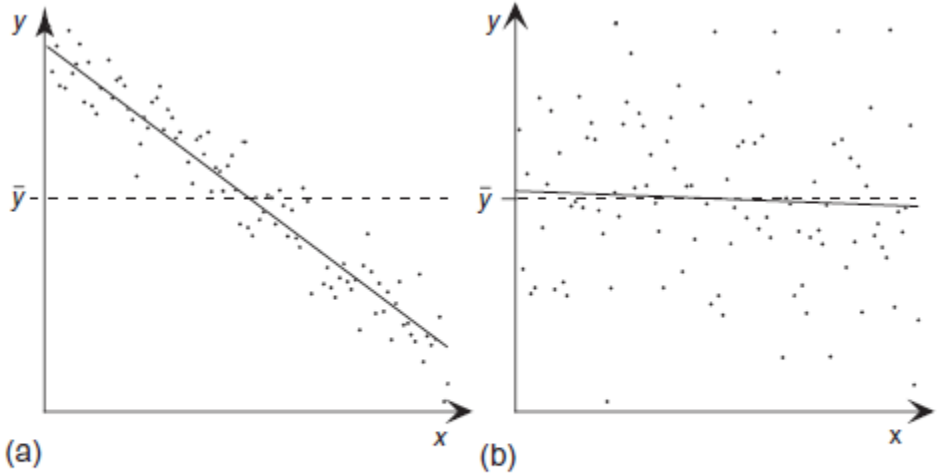


FIGURE 7.3 Illustration of the distinction between a fairly good regression relationship (a) and an essentially useless relationship (b). The points in panel (a) cluster closely around the regression line (solid), indicating small MSE, and the line deviates strongly from the average value of the predictand (dashed), producing a large SSR. In panel (b) the scatter around the regression line is large, and the regression line is almost indistinguishable from the mean of the predictand.

Wilks, 2019

REGRESION LINEAL

Recent change of vegetation growth trend in China

Shushi Peng[1], Anping Chen[2], Liang Xu[3], Chunxiang Cao[4], Jingyun Fang[1], Ranga B Myneni[3], Jorge E Pinzon[5], Compton J Tucker[5] and Shilong Piao[1]

[1] College of Urban and Environmental Sciences, Peking University, Beijing 100871, People's Republic of China
[2] Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA
[3] Department of Geography and Environment, Boston University, 675 Commonwealth Avenue, Boston, MA 02215, USA
[4] Institute of Remote Sensing Application of Chinese Academy of Sciences, Datun Road, 10 Beijing, People's Republic of China
[5] NASA/Goddard Space Flight Center, Greenbelt, MD 20771, USA

E-mail: slpiao@pku.edu.cn

Received 19 October 2011
Accepted for publication 30 November 2011
Published 22 December 2011
Online at stacks.iop.org/ERL/6/044027

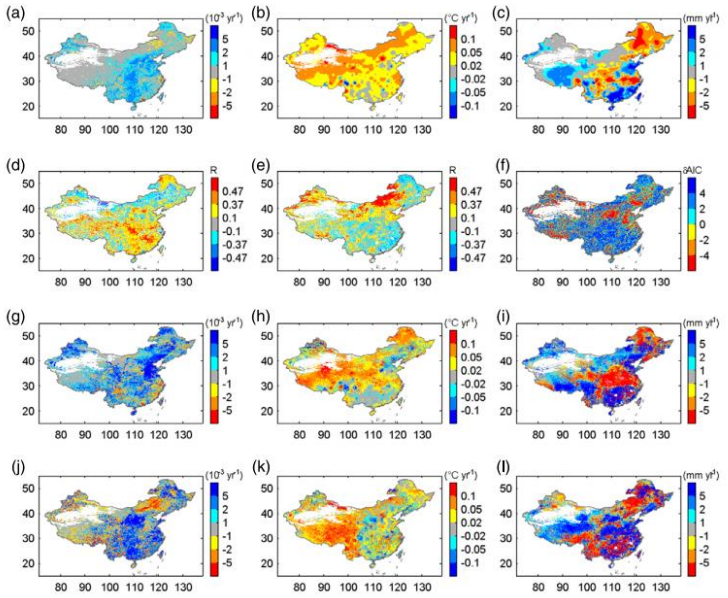


Figure 3. Spatial distribution of the results by the linear regression model and the piecewise regression model on growing season (April–October) NDVI and climate. Linear regression trends in growing season (a) NDVI, (b) temperature and (c) precipitation from 1982 to 2010. The correlation coefficients between growing season NDVI and (d) temperature and (e) precipitation. (f) Difference in AIC between piecewise regression and linear regression (δAIC). Trends in growing season NDVI (g) before and (j) after its TP. Trends in growing season temperature (h) before and (k) after the TP of the growing season NDVI trend. Trends in growing season precipitation (i) before and (l) after the TP of the growing season NDVI trend. $R = 0.37$ and $R = 0.47$ correspond statistically to 5% and 1% significance levels, respectively.

PRACTICA

TENDENCIAS

TENDENCIA LINEAL

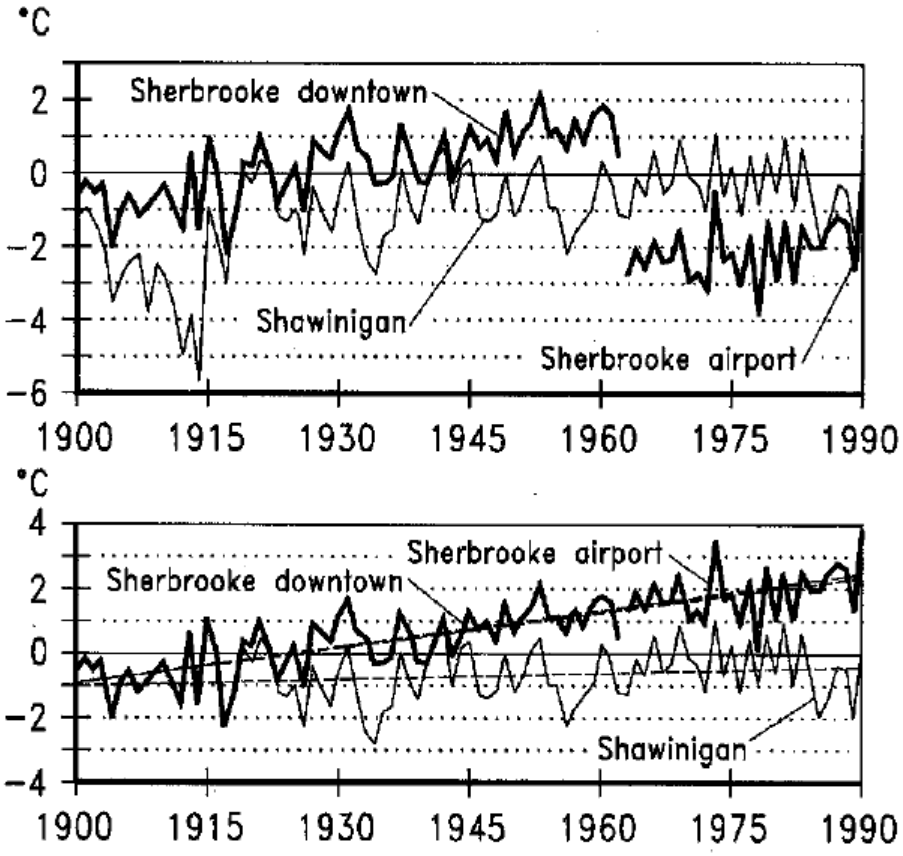


Figure 1.9: *Annual mean daily minimum temperature time series at two neighbouring sites in Quebec. Sherbrooke has experienced considerable urbanization since the beginning of the century whereas Shawinigan has maintained more of its rural character.*
Top: The raw records. The abrupt drop of several degrees in the Sherbrooke series in 1963 reflects the move of the instrument from downtown Sherbrooke to its suburban airport. The reason for the downward dip before 1915 in the Shawinigan record is unknown.
Bottom: Corrected time series for Sherbrooke and Shawinigan. The Sherbrooke data from 1963 onward are increased by 3.2 °C. The straight lines are trend lines fitted to the corrected Sherbrooke data and the 1915–90 Shawinigan record.
Courtesy L. Vincent, AES Canada.

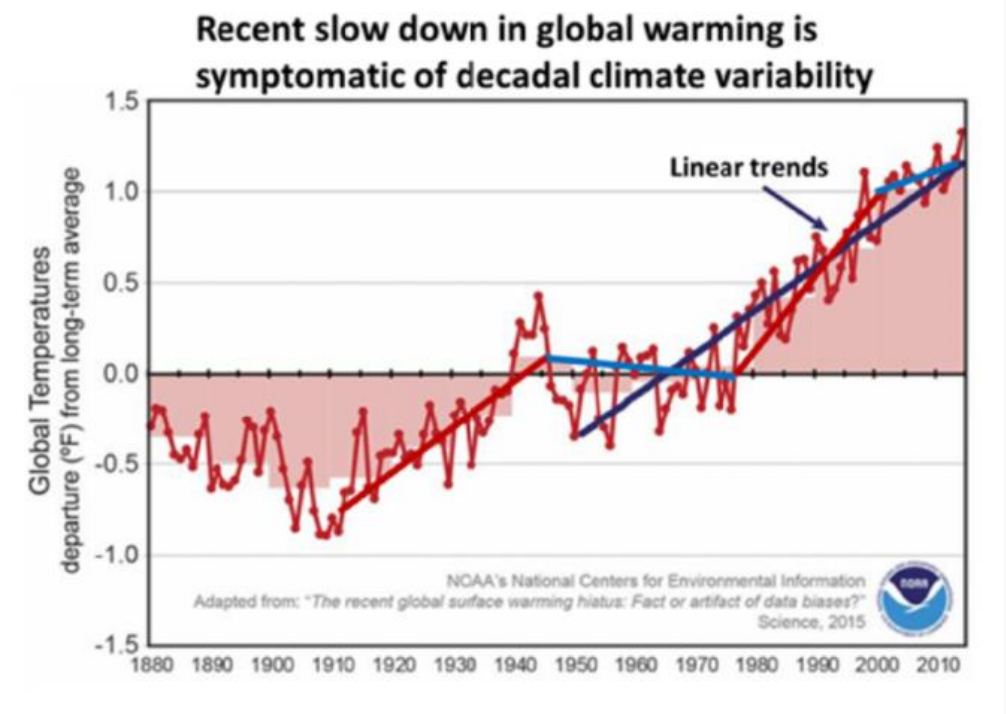
Storch &Zwiers, 1999

TENDENCIAS

TENDENCIA LINEAL

Si la serie de tiempo cumple los supuestos del modelo de regresión simple, se puede ajustar un modelo de regresión lineal considerando el valor de la serie de tiempo como variable dependiente y el paso de tiempo como variable independiente.

$$X(t) = a + bt$$



https://www.nap.edu/read/23552/chapter/4#10

PRACTICA

TENDENCIAS

SIGNIFICANCIA: TEST DE MANN KENDALL

La prueba de Mann-Kendall es una prueba no paramétrica que identifica la tendencia en la serie temporal. Al ser una prueba no paramétrica, la prueba se aplica ampliamente para detectar tendencias en series de tiempo siguiendo cualquier distribución de probabilidad. El estadístico de Mann-Kendall se define como:

$$S=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\operatorname{sgn}(x_j-x_i)=\sum_{i<j}\operatorname{sgn}(x_j-x_i),$$

$$\operatorname{sgn}(\Delta x)=\begin{cases}+1, & \Delta x>0\\ 0, & \Delta x=0.\\ -1, & \Delta x<0\end{cases}$$

$$\operatorname{Var}(S)=\frac{n(n-1)(2n+5)-\sum_{j=1}^{J}t_j(t_j-1)(2t_j+5)}{18}.$$

$$z=\begin{cases}\dfrac{S-1}{[\operatorname{Var}(S)]^{1/2}}, & S>0\\ \dfrac{S+1}{[\operatorname{Var}(S)]^{1/2}}, & S<0\end{cases}.$$

Wilks, 2019

where g is the number of tied groups and t_i represents the number of observations in the tied group. Tied groups are groups having members tied, or, in other words if the frequency of a value is greater than 1 in the frequency table, it constitutes the tied group. For example, in the data set {15, 11, 10, 12, 10, 15, 13, 15} there are two tied groups (10 and 15). Tied group for 10 has 2 members and tied group for 15 has 3 members. However, continuous hydroclimatic variables like precipitation, stream flow, temperature may have very less or no tied group. Under the assumption that there is no tied group, the variance of S statistic becomes:

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

Maity, 2018

TENDENCIAS

TEST DE MANN KENDALL

$$z = \begin{cases} \dfrac{S-1}{[\mathrm{Var}(S)]^{1/2}}, & S>0 \\ \dfrac{S+1}{[\mathrm{Var}(S)]^{1/2}}, & S<0 \end{cases}.$$

Wilks, 2019

The test statistics u_c is given by:

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

u_c statistic follows standardized normal distribution. The null hypothesis of no trend can be rejected if $|u_c| > Z_{(\alpha/2)}$, where $Z_{(\alpha/2)}$ is standardized normal variate for the non-exceedance probability of $(1 - \alpha/2) \times 100\%$ and α is the level of significance. For no tied group, the test is valid for $N > 10$.

Maity, 2018

TENDENCIAS

$$Y_i = \beta_0 + \beta_1 X_i$$

THEIL-SEN SLOPE

$$\hat{\beta}_1 = median\left\{\tilde{B}\right\},\ \tilde{B} = \left\{b_{ij}\ \middle|\ b_{ij} = \frac{y_j - y_i}{x_j - x_i}, x_i \neq x_j, 1 \leq i < j < n\right\}.$$

$$\hat{\beta}_0 = Y_{median} - \hat{\beta}_1 \times X_{median}$$

Chervenkov & Slavov, 2018

Chervenkov, Hristo & Slavov, Kiril. (2019). Theil–Sen estimator vs. Ordinary least squares — Trend analysis for selected ETCCDI climate indices. Comptes rendus de l'Académie bulgare des sciences: sciences mathématiques et naturelles. 72. 47-54. 10.7546/CRABS.2019.01.06.

PRACTICA