



REGIONAL FUND QUALITY INFRASTRUCTURE FOR BIODIVERSITY & CLIMATE PROTECTION
IN LATIN AMERICA AND THE CARIBBEAN

Workshop on Statistic, data analysis and measurement uncertainty for Meteorology

Fundamentos estadísticos para la evaluación de la incertidumbre en las calibraciones

CENAM – Querétaro, México – 02 ~ 06.12.2019

Márcio A. A. Santana – INPE/BR
marcio.santana@inpe.br

Ricardo de A. Kalid – UFSB/BR
kalid@ufsb.edu.br



i. Fundamentos estadísticos

- a) Terminología & Análisis de datos
- b) Evaluación de la incertidumbre
- c) Calibración

Fundamentos estadísticos para la evaluación de la incertidumbre en las calibraciones

Terminología & Análisis de datos

1.1

população

totalidade dos itens considerados

1.2

unidade amostral

uma das partes individuais em que uma **população** (1.1) é dividida

1.3

amostra

Subconjunto de uma **população** (1.1) composto de uma ou mais **unidades amostrais** (1.2)

1.4

valor observado

valor obtido de uma propriedade associada com um elemento de uma amostra (1.3)

1.28

intervalo de confiança

estimador de intervalo (1.25) (T_0, T_1) para o **parâmetro** (2.9) θ com a **estatística** (1.8) T_0 e T_1 como limites de intervalo e para qual tem-se $P[T_0 < \theta < T_1] \geq 1 - \alpha$

NOTA 1 A confiança reflete a proporção de casos em que o intervalo de confiança conteria o valor verdadeiro do parâmetro em uma série longa de **amostras aleatórias** (1.6) repetidas sob circunstâncias idênticas. Um intervalo de confiança não reflete a **probabilidade** (2.5) do intervalo observado conter o valor verdadeiro do parâmetro (o intervalo contém ou não o valor verdadeiro).

Jerarquia terminológica

$$\text{SI} \geq \text{VIM} \geq \text{GUM} \geq \text{ASTM}$$

- **SI:** El Sistema Internacional de Unidades
 - **VIM:** Vocabulario Internacional de Metrología
 - **GUM:** Guía para la Expresión de la Incertidumbre de Medida
 - **ASTM** D4430-00: Standard Practice for Determining the Operational Comparability of Meteorological Measurements
- Si existe un conflicto entre las definiciones, se adoptará **SI**, seguido por el establecido en **VIM**, despues **GUM** y, por último, por **ASTM**.

Por que usar o SI, VIM y GUM



- Evitar errores groseros
- Fomentar el comercio internacional
- Estandarizar unidades y terminologías
- Autores:
 - ❖ **BIPM:** Organización Internacional de Pesas y Medidas
 - ❖ **IEC:** Comisión Electrotécnica Internacional
 - ❖ **IFCC:** Federación Internacional de Quimica Clinica Y Laboratorios Medicos
 - ❖ **ILAC:** Cooperacion Internacional de Acreditación de Laboratorios
 - ❖ **ISO:** Organización Internacional de Normalización
 - ❖ **IUPAC:** Unión Internacional de Química Pura y Aplicada
 - ❖ **IUPAP:** Unión Internacional de Fisica Pura y Aplicada
 - ❖ **OIML:** Organización Internacional de Metrología Legal

El Sistema Internacional de Unidades

- “Los símbolos de las unidades se tratan como entidades matemáticas. Cuando se expresa el valor de una magnitud como producto de un valor numérico por una unidad, el valor numérico y la unidad pueden tratarse de acuerdo con las reglas ordinarias del álgebra. Este procedimiento constituye el cálculo de magnitudes, o álgebra de magnitudes. Por ejemplo, la ecuación $T = 293 \text{ K}$ puede escribirse también como $T/\text{K} = 293$. Puede resultar cómodo escribir de esta forma el cociente entre una magnitud y una unidad en la cabecera de una tabla, tal que las entradas de la tabla sean simplemente números. Por ejemplo, una tabla que presente la presión de vapor en función de la temperatura y el logaritmo neperiano de la presión de vapor en función de la temperatura, a la potencia menos uno, podría adoptar la siguiente forma:.” (SI, 2013, p. 44).

Tabla 1: Presentación de unidades en una tabla.

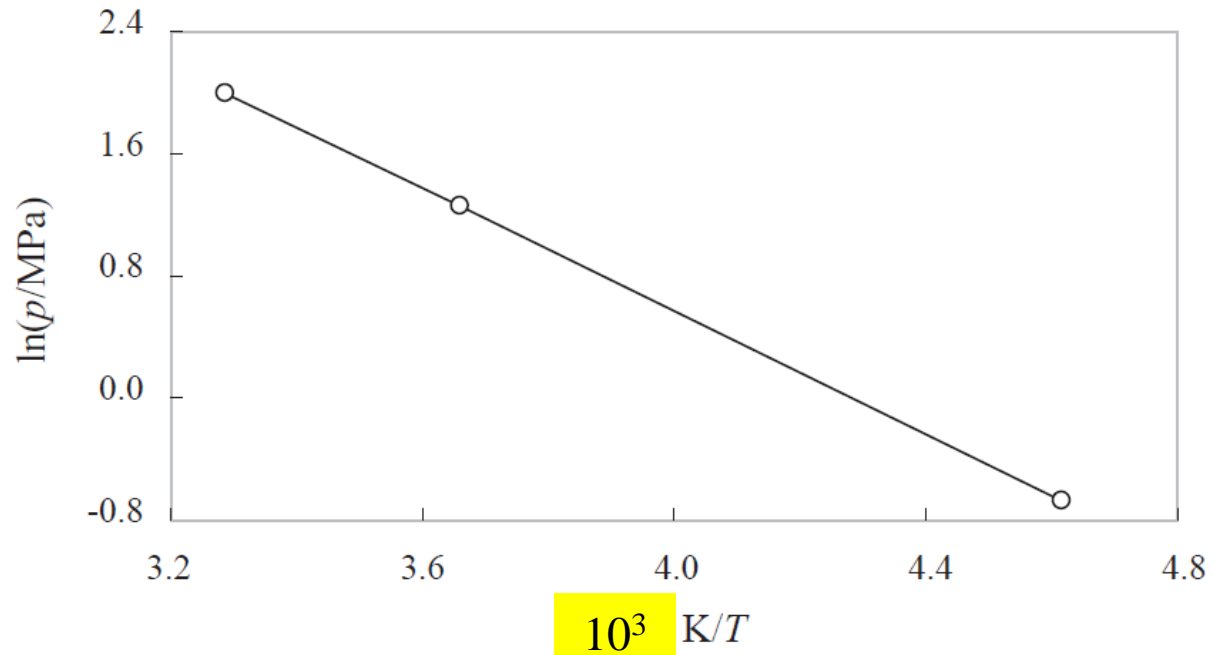
T/K	$10^3 \text{ K}/T$	p/MPa	$\ln(p/\text{MPa})$
216,55	4,6179	0,5180	-0,6578
273,15	3,6610	3,4853	1,2486
304,19	3,2874	7,3815	1,9990

Tabla 1:
Presentación
de unidades
en una tabla.

T/K	$10^3 K/T$	p/MPa	$\ln(p/\text{MPa})$
216,55	4,6179	0,5180	-0,6578
273,15	3,6610	3,4853	1,2486
304,19	3,2874	7,3815	1,9990

Figura 1:
Presentación
de unidades
en un gráfico

Los ejes de un gráfico pueden también etiquetarse de este modo, de forma que las graduaciones sean puramente numéricas, como se indica en la figura de abajo.



Desde el punto de vista algebraico podrían emplearse otras formas equivalentes a $10^3 K/T$, como por ejemplo kK/T , ó $10^3 (T/K)^{-1}$.

El Sistema Internacional de Unidades

- El símbolo utilizado para separar la parte entera de su parte decimal se denomina “separador decimal”. Desde la 22ª Conferencia General (2003, Resolución 10), “el símbolo del separador decimal puede ser el punto o la coma, en la propia línea de escritura. El separador decimal elegido será el de uso corriente en el contexto en cuestión.” (SI, 2013, p. 45).
- El símbolo utilizado para separar la parte entera de su parte decimal se denomina “separador decimal”. Desde la 22ª Conferencia General (2003, Resolución 10), “el símbolo del separador decimal puede ser el punto o la coma, en la propia línea de escritura. El separador decimal elegido será el de uso corriente en el contexto en cuestión.” (SI, 2013, p. 45).
- Ejemplo 1: 43 279.168 29 está correcto, pero 43,279.168,29 no
- Ejemplo 2: 3279.1683 está correcto, pero 3,279.168 3 no
- Ejemplo 3: -0.234 está correcto, pero -.234 no
- Ejemplo 4: $t = 30.2\text{ }^{\circ}\text{C}$ está correcto, pero $t = 30.2^{\circ}\text{C}$ o $t = 30.2^{\circ}\text{ C}$ no

El Sistema Internacional de Unidades

- **Unidades fuera del SI:** bar, atm y mmHg
- Preferiblemente escriba las ecuaciones apenas con los símbolos de las magnitudes: $x = v \cdot t$
- Si usa valores numéricos en una ecuación, también debe ingresar las unidades correspondientes en la ecuación.

Correcto: $\frac{x}{m} = 3,6^{-1} \frac{v}{km \cdot h^{-1}} \cdot \frac{t}{s}$

Incorrecto: $x = 3,6^{-1} v \cdot t$

acompañado por el texto " donde x está en metros,
 v está en kilómetros por hora y t está en segundos"

Adaptado de <https://physics.nist.gov/cuu/Units/checklist.html>

- Ejemplo 5: $t/^{\circ}C = T/K - 273.15$ está correcto, pero $t = T - 273.15$ no
- Ejemplo 6: $U_{max} = 1000 V$ está correcto, pero $U = 1000 V_{max}$ no
- Ejemplo 7: $U_{max} = 1000 V$ está correcto, pero $U_{max} = 1000 V$ no

El Sistema Internacional de Unidades

“5.3.5 Expresión de la incertidumbre de medida asociada al valor de una magnitud

La incertidumbre asociada al valor estimado de una magnitud debe evaluarse y expresarse de acuerdo con la ***Guía para la expresión de la incertidumbre de medida*** [ISO, 1995]. La incertidumbre típica, es decir, la desviación típica estimada (correspondiente a un factor de cobertura $k = 1$), asociada a una magnitud x se designa como $u(x)$. Una forma cómoda de representar la incertidumbre es, por ejemplo:

$$m_n = 1,674\,927\,28\,(29) \times 10^{-27} \text{ kg},$$

donde m_n es el símbolo de la magnitud (en este caso la masa del neutrón) y el número entre paréntesis el valor numérico de la incertidumbre típica referida a las dos últimas cifras del valor estimado de m que para este caso es:

$$u(m_n) = 0,000\,000\,29 \times 10^{-27} \text{ kg}.$$

Si se usa un factor de cobertura k distinto de 1, es necesario indicarlo.”
(SI, 2013, 46)

La incertidumbre típica es una desviación típica → no negativa

Constantes utilizadas para definir las unidades

Unidad Básica

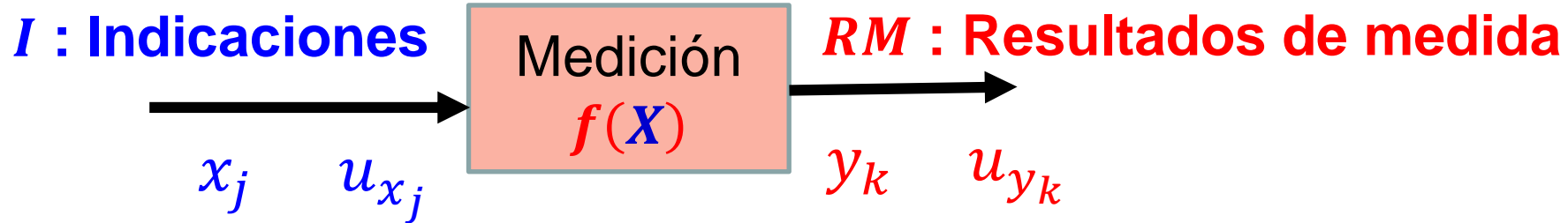
SI actual

SI revisado

segundo	s	$\Delta\nu(^{133}\text{Cs})_{\text{hfs}}$	$\Delta\nu_{\text{Cs}}$	frecuencia de la transición hiperfina del estado fundamental no perturbado del átomo de cesio 133
metro	m	c	c	velocidad de la luz en el vacío
kilogramo	kg	$m(\text{IPK})$	h	constante de Planck
amperio	A	μ_0 (permeabilidad magnética del vacío)	e	carga elemental
kelvin	K	T_{TPW}	k	constante de Boltzmann
mol	mol	$M(^{12}\text{C})$	N_{A}	constante de Avogadro
candela	cd	K_{cd}	K_{cd}	eficacia luminosa de una fuente emitiendo a 540 THz

Curso Virtual Metrología: <http://aulavirtual.cem.es/>

Función de medición



VIM 2.49

función de medición, f

función de **magnitudes** cuyo valor es un **valor medido** de la **magnitud de salida** en el **modelo de medición**, cuando se calcula mediante los **valores** conocidos de las **magnitudes de entrada en el modelo de medición**

NOTA 1 Si el **modelo de medición** $h(Y, X_1, X_2, \dots, X_n) = 0$ puede escribirse explícitamente como $Y = f(X_1, X_2, \dots, X_n)$, siendo Y la magnitud de salida en el modelo de medición, f es la función de medición. En general, f puede representar un algoritmo que, para los valores de entrada x_1, x_2, \dots, x_n , da como resultado un valor único de la magnitud de salida $y = f(x_1, x_2, \dots, x_n)$

NOTA 2 La función de medición se utiliza también para calcular la **incertidumbre de medida** asociada al valor medido de Y .

VIM – Vocabulário Internacional de Metrologia

2.9 (3.1)

resultado de medida, m

resultado de una medición, m

RM

conjunto de **valores de una magnitud** atribuidos a un **mensurando**, acompañados de cualquier otra información relevante disponible

NOTA 1 Un resultado de medida contiene generalmente información relevante sobre el conjunto de valores de una magnitud. Algunos de ellos representan el mensurando mejor que otros. Esto puede representarse como una función de densidad de probabilidad (FDP).

NOTA 2 El resultado de una medición se expresa generalmente como un **valor medido** único y una **incertidumbre de medida**. Si la incertidumbre de medida se considera despreciable para un determinado fin, el resultado de medida puede expresarse como un único valor medido de la magnitud. En muchos campos ésta es la forma habitual de expresar el resultado de medida.

NOTA 3 En la bibliografía tradicional y en la edición precedente del VIM, el término resultado de medida estaba definido como un valor atribuido al mensuran do y podía entenderse como **indicación**, resultado no corregido o resultado corregido, según el contexto.”

4.1 (3.2)

indicación, f

valor proporcionado por un **instrumento** o **sistema de medida**

I

4.20 (5.25)

sesgo instrumental, m

sesgo, m

diferencia entre la media de las **indicaciones** repetidas y un **valor de referencia**

$S = I - VR$

2.53 (3.15) (3.16)

corrección, f

compensación de un efecto sistemático estimado

C

$\left\{ \begin{array}{l} C = -S \end{array} \right.$

$$RM = I - S = I + C$$

Ejemplo: medición de temperatura

Informaciones:

- 1) Indicaciones: (24,1 ; 24,2 ; 24,0 ; 23,9 ; 24,1) °C
- 2) Certificado de calibración (abajo e con no conformidades)
- 3) Fuentes de incertidumbre significativas son indicaciones, corrección de las indicaciones y resolución del instrumento de medición, las otras fuentes de incertidumbre son insignificantes
- 4) Las fuentes de incertidumbre son independientes
- 5) La medición está en estado estacionario



RESULTADOS DA CALIBRAÇÃO:

Os resultados a seguir apresentados referem-se à situação do instrumento conforme recebido pelo Laboratório, sendo V_r o valor de referência, V_i o valor do instrumento em calibração e Erro a diferença entre a indicação do instrumento em calibração e o valor de referência.

Profundidade de imersão (mm)	Padrão Utilizado	V_r (°C)	V_i (°C)	Erro (°C)	U (°C)	Fator k	V_{eff}
155	141339/1	-25,1	-25,0	0,1	0,4	2,00	∞
155	141339/1	0,0	0,0	0,0	0,3	2,00	∞
155	141339/1	50,0	50,0	0,0	0,4	2,00	∞
155	141339/1	99,9	100,0	0,1	0,4	2,00	∞
155	141339/1	139,8	140,0	0,2	0,4	2,00	∞

Ejemplo: medición de temperatura

Función de medición: $RM_i = I_i + C_i + R_i$

RM : Resultado de medición

I : Indicación

C : Corrección sistemática

R : Resolución sistemática



Vamos a Excel: Archivo **Ejemplo_medicion_de_temperatura.xlsx**

Tabla DM-T: Datos de medición de temperatura

Unidad: °C				
Variables:	Indicación / (°C)	Corrección / (°C)	Media de la Resolución / (°C)	Resultado de medición / (°C)
Media experimental:	24,060	NA	NA	23,960
Desviación típica experimental:	0,114	NA	NA	0,114
Desviación típica experimental de la media:	0,051	NA	NA	0,051
Punto	I_i	C_i	R_i	$RM_i = I_i + C_i + R_i$
1	24,1	-0,1	0,0	24,0
2	24,2	-0,1	0,0	24,1
3	24,0	-0,1	0,0	23,9
4	23,9	-0,1	0,0	23,8
5	24,1	-0,1	0,0	24,0

Legenda: NA - No se aplica

Esperanza
Varianza
Covarianza
Auto covarianza

Esperanza $E\{X\}$ o μ_X

- Valor esperado o esperanza $E\{X\}$: Es una medidas de tendencia central de X
- Para un proceso estacionario de 1ª orden:
 - Si X es una magnitud discreta: $E\{X\} \stackrel{\text{def}}{=} \sum_{k=1}^{q_p} \{x_k \cdot p(x_k)\}$,
donde $p(x_k)$ es la función de probabilidad de x_k e q_p es el tamaño de la población
 - Si X es una magnitud continua: $E\{X\} \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} \{x \cdot f(x)\} dx$,
donde $f(x)$ es la función de densidad de probabilidad de x_k
 - Otras medidas de tendencia central son la mediana e la moda
 - Un estimador insesgado de $E\{X\}$ es la media aritmética:

$$\bar{x} = \frac{\sum_{k=1}^q \{x_k\}}{q} \rightarrow E\{X\}$$

Varianza $var\{X\}$ o $V\{X\}$

Para un proceso estacionario de 1ª orden

- Varianza $V\{X\}$: Indica la variabilidad o dispersión de la magnitud X

$$V\{X\} \stackrel{\text{def}}{=} var\{X\} = E\{(X - E\{X\})(X - E\{X\})\} = E\{(X - \mu_X)^2\}$$

- Un estimador no sesgado de $V\{X\}$ es $\hat{V}\{X\} = \frac{1}{q-1} \cdot \sum_{k=1}^q \{(x_k - \bar{x})^2\} \rightarrow V\{X\}$

Donde q es la cantidad de experimentos, \bar{x} es el promedio aritmético da magnitud X

Desviación estándar o Desviación típica σ_X

- Desviación estándar σ_X : También indica la variabilidad o dispersión de la magnitud X

Por definición σ_X es la raíz cuadrada **POSITIVA** de la varianza: $\sigma_X \stackrel{\text{def}}{=} +\sqrt{V\{X\}}$

- Un estimador no sesgado de σ_X es $\hat{\sigma}_X = \sqrt{\frac{1}{q-1} \cdot \sum_{k=1}^q \{(x_k - \bar{x})^2\}} \rightarrow \sigma_X$

desde que los valores de x_k sean independientes entre si

Covarianza $cov\{X, Y\}$

Para un proceso estacionario de 1ª orden

- Covarianza $cov\{X, Y\}$: Indica la variación entre una magnitud X cuando otra magnitud Y cambia $cov\{X; Y\} \stackrel{\text{def}}{=} E\{(X - \mu_X)(Y - \mu_Y)\}$
- Si $cov\{x; y\} = 0$ entonces o conjunto de datos de $[x]$ em análisis no cambia como conjunto de datos de $[y]$, y viceversa
- Estimador sin sesgo de $cov\{X; Y\}$: $cov\{\bar{X}; \bar{Y}\} = \frac{1}{q-1} \cdot \sum_{k=1}^q \{(x_k - \bar{x}) \cdot (y_k - \bar{y})\}$

Donde q es la cantidad de experimentos, \bar{x} es el promedio aritmético de la magnitud X

Coeficiente de correlación ρ_{XY}

$$\rho_{XY} = \frac{cov\{X, Y\}}{\sqrt{var(X)} \cdot \sqrt{var(Y)}} = \frac{cov\{X, Y\}}{\sigma(X) \cdot \sigma(Y)}$$

Donde $\sigma(X)$ es la desviación estándar de la magnitud X

- Tenga en cuenta que $\rho_{XX} = 1$ y si $\rho_{XY} = 0 \Rightarrow X$ no están correlacionados Y .

- Estimador no sesgado de ρ_{XY} :

$$\hat{\rho}_{XY} = \frac{\sum_{k=1}^q \{(x_k - \bar{x}) \cdot (y_k - \bar{y})\}}{\sqrt{\sum_{k=1}^q \{(x_k - \bar{x})^2\}} \cdot \sqrt{\sum_{k=1}^q \{(y_k - \bar{y})^2\}}}$$

- Si X no tiene relación fenomenológica con Y (X y Y son independientes) $\Rightarrow \rho_{XY} = 0$
- Pero si $\rho_{XY} = 0$ no necesariamente X y Y son magnitudes independientes

Matriz de Covarianza $cov\{X\}$

Para un proceso estacionario de 1ª orden

- También llamada
 - matriz de varianza
 - matriz de varianza-covarianza
- Considerar o vector das magnitudes $X = [X_1; X_2; \dots; X_{q_X}]$, donde q_X es la cantidad de magnitudes

$$cov\{X\} \stackrel{\text{def}}{=} E\{(X - E\{X\}) \cdot (X - E\{X\})^T\} = E\{(X - \mu_X) \cdot (X - \mu_X)^T\}$$

Matriz de Coeficientes de Correlación de X : ρ_X

$$\rho_X = \begin{bmatrix} 1 & \cdots & \rho_{X_1 X_{q_X}} \\ \vdots & \ddots & \vdots \\ \rho_{X_{q_X} X_1} & \cdots & 1 \end{bmatrix}$$

$$\rho_X = \frac{cov\{X\}}{\left[\sqrt{diag(cov\{X\})^T \times diag(cov\{X\})} \right]}$$

- La matriz de coeficiente de correlación ρ_X es simétrica
- Cuanto más cerca $\rho_{X_i X_j}$ de **+1** (**-1**), más **positivamente** (**negativamente**) correlacionadas están las magnitudes X_i con X_j

Auto covarianza γ_k

- Auto covarianza γ_k : Es la covarianza entre X_t y su valor X_{t-k} separados por k intervalos de tiempo

$$\gamma_k \stackrel{\text{def}}{=} \text{cov}\{X_t; X_{t-k}\} = E\{(X_t - \mu_X) \cdot (X_{t-k} - \mu_X)^T\}; \quad k = 0; 1; 2; \dots$$

- Estimador sin sesgo de γ_k (para grandes muestras):

$$\hat{\gamma}_k = \frac{1}{T} \cdot \sum_{t=k+1}^T \{(x_t - \bar{x}) \cdot (x_{t-k} - \bar{x})\} \rightarrow \text{cov}\{X_t; X_{t-k}\}$$

Coeficiente de auto correlación ρ_k

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{cov}\{X_t; X_{t-k}\}}{\sqrt{\text{var}(X_t)} \cdot \sqrt{\text{var}(X_{t-k})}}; \quad k = 0; 1; 2; \dots$$

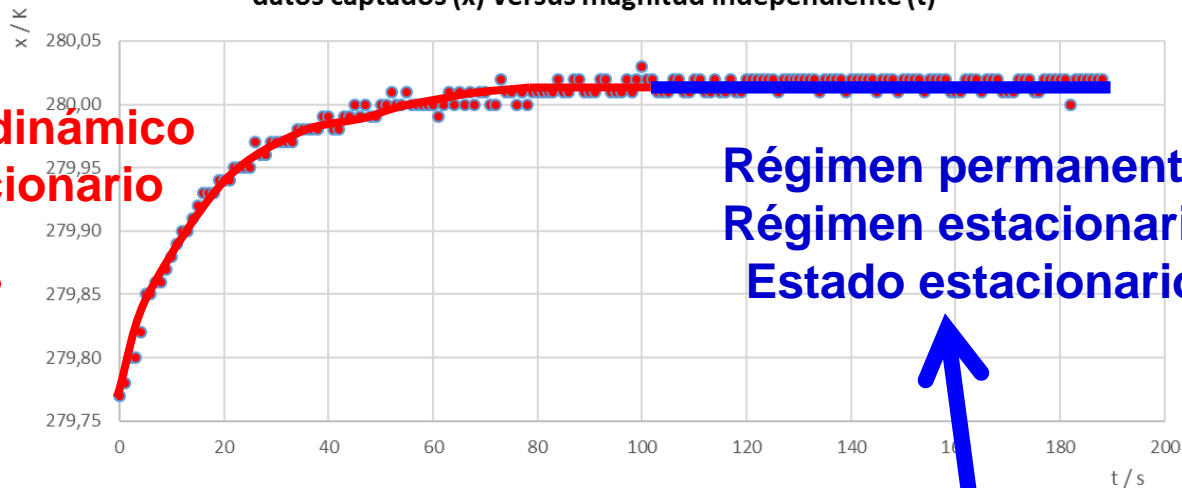
- Donde $\rho_0 = 1$ e $\rho_k = \rho_{-k}$.
- Estimador sin sesgo:

$$\hat{\rho}_k = \frac{T}{T-k} \cdot \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

Seleccionar datos en régimen estacionario

Gráfico de tendencia

Figura Dx: Gráfico de tendencia -
datos captados (x) versus magnitud independiente (t)



Régimen dinámico
No estacionario

Régimen permanente
Régimen estacionario
Estado estacionario

Auto correlación

Figura Correlograma 1
Datos: 1 a 189

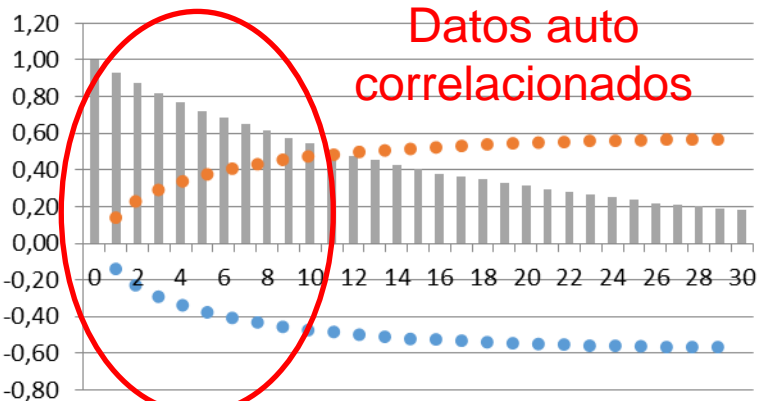
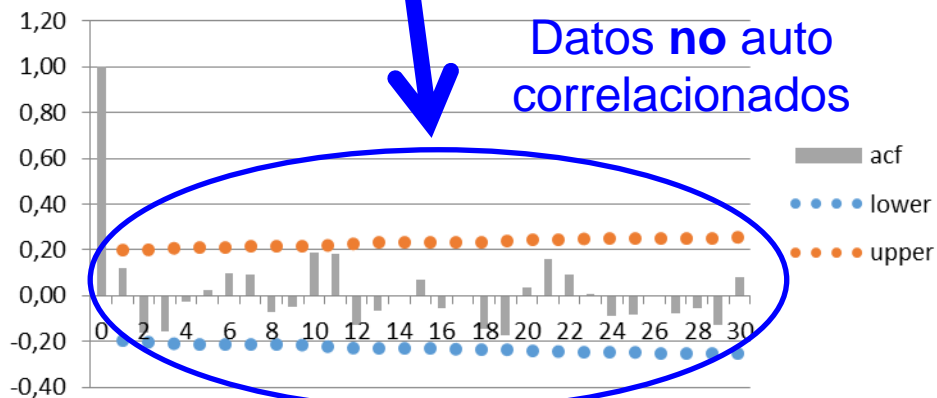


Figura Correlograma 3
Datos: 101 a 189



Vamos a Excel: archivo

Ejemplo_auto_correlacion.xlsx

Instalación

Real Statistics Resource Pack

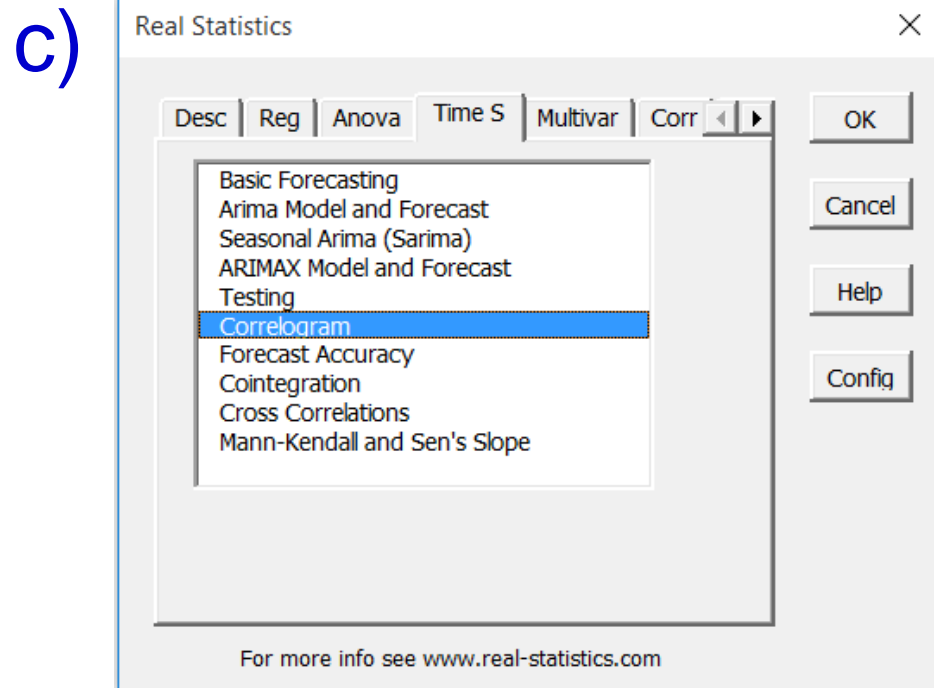
- Caja de herramientas estadísticas complementaria a las funciones estadísticas de Excel
- <http://www.real-statistics.com/>
- Instrucciones de instalación: <http://www.real-statistics.com/free-download/real-statistics-resource-pack/>

Seleccionar datos en régimen estacionario

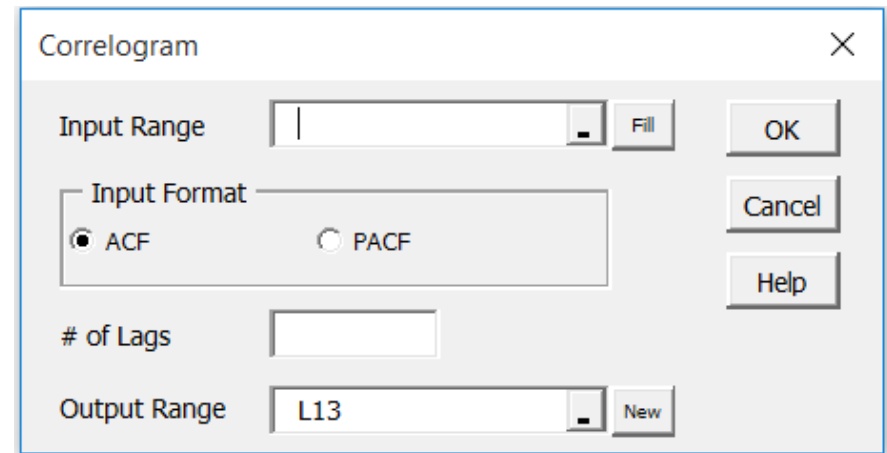
Real Statistics Resource Pack

a) Activación RSRP: Archivo > Opciones > Complementos > Complementos de Excel > Ir... > Xrealstats > Examinar > XRealStats.xlam > Aceptar

b) **Ctrl + m**



d)



Análisis de datos

Supuestos:

- Experimento bien hecho
- Régimen estacionario
- Instrumento calibrado y con trazabilidad

Procedimiento para el análisis de datos:

- 1) Poner datos en unidades SI
- 2) Seleccionar datos en régimen estacionario
- 3) Eliminar datos divergentes
(discrepantes, espurios)
- 4) Evaluar estadísticamente los datos
- 5) Elaborar un informe

2) Seleccionar datos en régimen estacionario

2.1 Elaborar gráficos

2.1.1 Gráfico e tendencia:

- Eje horizontal: tiempo o secuencia de datos
- Eje vertical: magnitudes independiente e dependiente

2.1.2 Gráfico de dispersión

- Eje horizontal magnitudes independiente (patrón)
- Eje vertical: magnitudes dependiente (objeto)

2.2 Seleccionar gráficamente la región candidata de estado estacionario

2.3 Calcular coeficientes de auto correlación

2.4 Probar estadísticamente que región candidata de régimen estacionario es sin auto covarianza

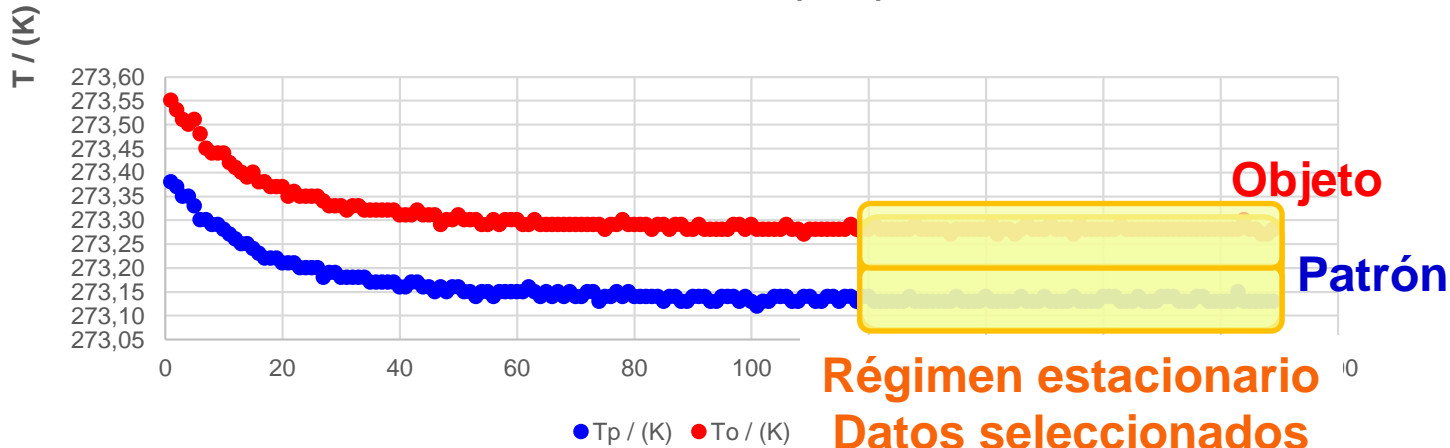
Los tiempos de captura debe ser lo mismo para el patrón e objeto



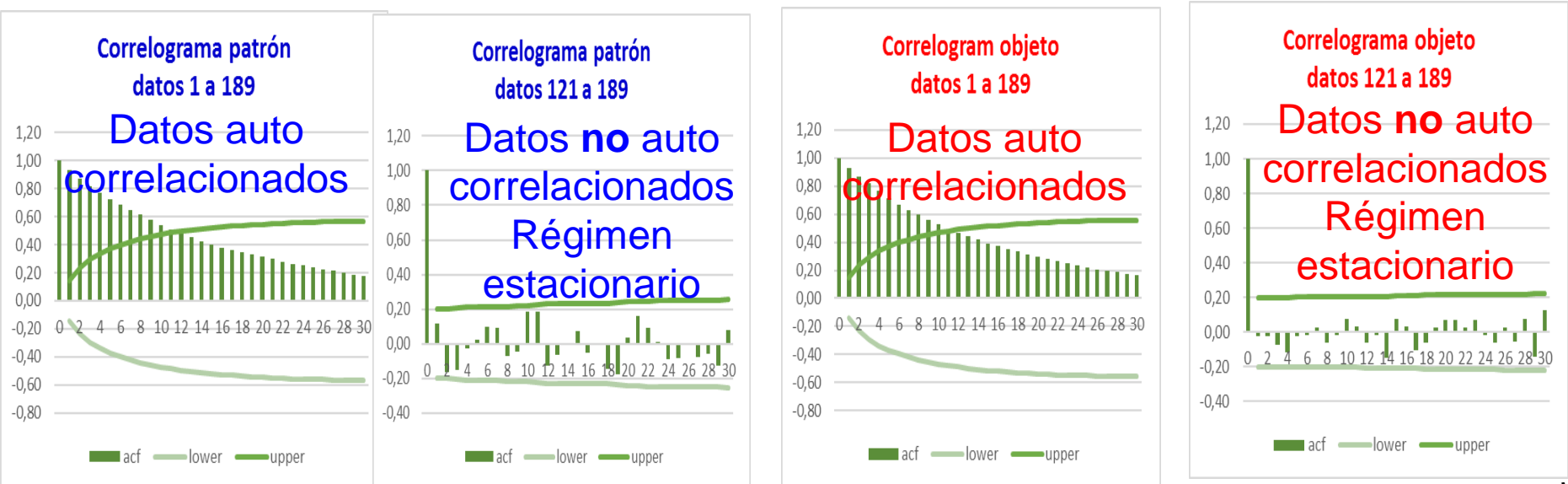
Seleccionar datos en régimen estacionario

Gráfico de tendencia

Figura T-01: Tendencia de datos originales de temperatura a 273,15 K (0 °C).



Auto correlación



i) Régimen dinámico o régimen **NO** estacionario

i.a) Rehacer el experimento

i.b) Trabajar con modelos dinámicos (modelo de serie temporales):
AR, ARX, ARMA, ARIMA, ARMAX, ARIMAX, NARMA, NARMAX:

- 1) AR: Auto Regressive
- 2) ARMA: Auto Regressive Moving Average
- 3) ARIMA: Auto Regressive Integrated Moving Average
- 4) ARMAX: Auto Regressive Moving Average with exogenous inputs
- 5) ARIMAX: Auto Regressive Integrated Moving Average with exogenous inputs
- 6) NARMA: Non linear ARMA
- 7) NARMAX: Non linear ARMA with exogenous inputs
- 8) NARIMAX: Non linear Auto Regressive Integrated Moving Average with exogenous inputs;

i.b) Usar métodos estadísticos para tratar mediciones correlacionadas (GUM 4.2.7): varianza de Allan, Estadística Bayesiana.

¡Vamos a practicar!

Detección de régimen permanente



Practica: detección de régimen permanente

Archivo: **1_Deteccion_regimen_estacionario.xlsx**



Procedimiento sugerido para la detección de datos en régimen permanente:

1. Elabore gráfico de tendencia de todas las magnitudes
2. Seleccione regiones candidatas a estado estacionario
3. Defina nivel de confianza (NC), generalmente 95 %
4. Elabore gráficos de auto correlación e calcule los coeficientes de autocorrelaciones (ρ_k)
5. Seleccione período de muestreo (T_a)
6. Todos coeficientes de auto correlaciones (ρ_k) deberían pertenecer al intervalo de confianza (IC_{ρ_k})
 - (i) altere nivel de confianza (NC); o
 - (ii) altere período de muestreo (T_a); o
 - (iii) altere NC y T_a hasta que todos $\rho_k \in IC_{\rho_k}$

La región estacionaria debe ser la misma para todas magnitudes

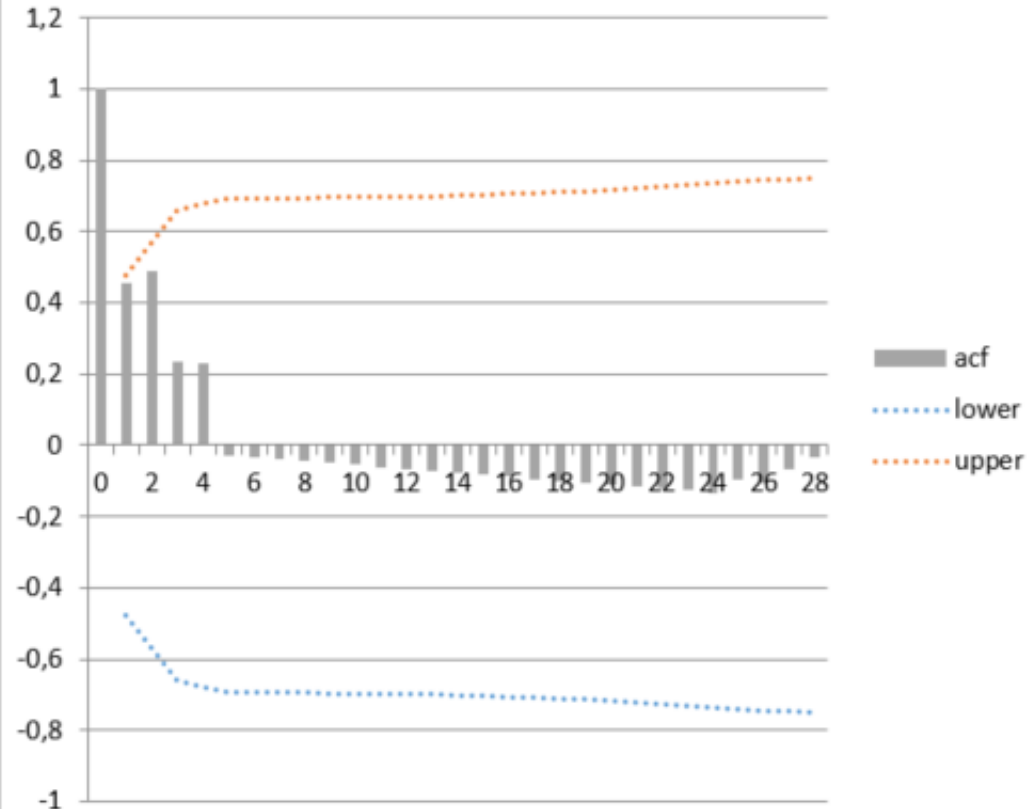
Comparación datos Régimen estacionario

Gráfico de tendencia

T / K

Correlogram	alpha	0,01
lags	acf	lower upper
0	1	
1	0,45448276	-0,47831949 0,47831949
2	0,48896552	-0,5685988 0,5685988
3	0,23344828	-0,65780387 0,65780387
4	0,22793103	-0,67649325 0,67649325
5	-0,02758621	-0,69384113 0,69384113
6	-0,03310345	-0,69409202 0,69409202
7	-0,03862069	-0,69445314 0,69445314
8	-0,04413793	-0,69494437 0,69494437
9	-0,04965517	-0,69558544 0,69558544
10	-0,05517241	-0,69639596 0,69639596
11	-0,06068966	-0,6973953 0,6973953
12	-0,0662069	-0,69860259 0,69860259
13	-0,07172414	-0,70003664 0,70003664
14	-0,07724138	-0,70171593 0,70171593
15	-0,08275862	-0,70365849 0,70365849
16	-0,08827586	-0,70588188 0,70588188
17	-0,0937931	-0,70840311 0,70840311
18	-0,09931034	-0,71123861 0,71123861
19	-0,10482759	-0,71440413 0,71440413
20	-0,11034483	-0,7179147 0,7179147
21	-0,11586207	-0,72178458 0,72178458

Auto correlación del **objeto**, NC = 99 %
21/11/17 22:20:00 datos 673 a 701



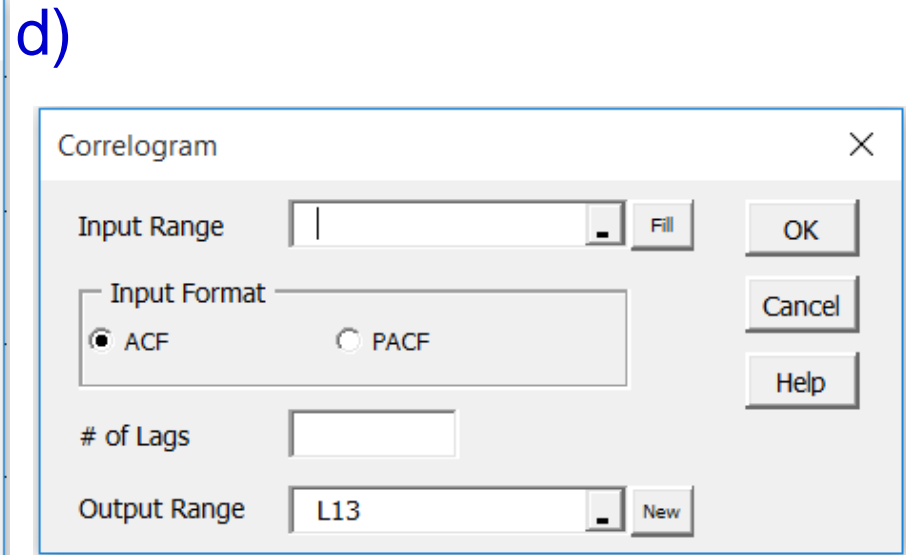
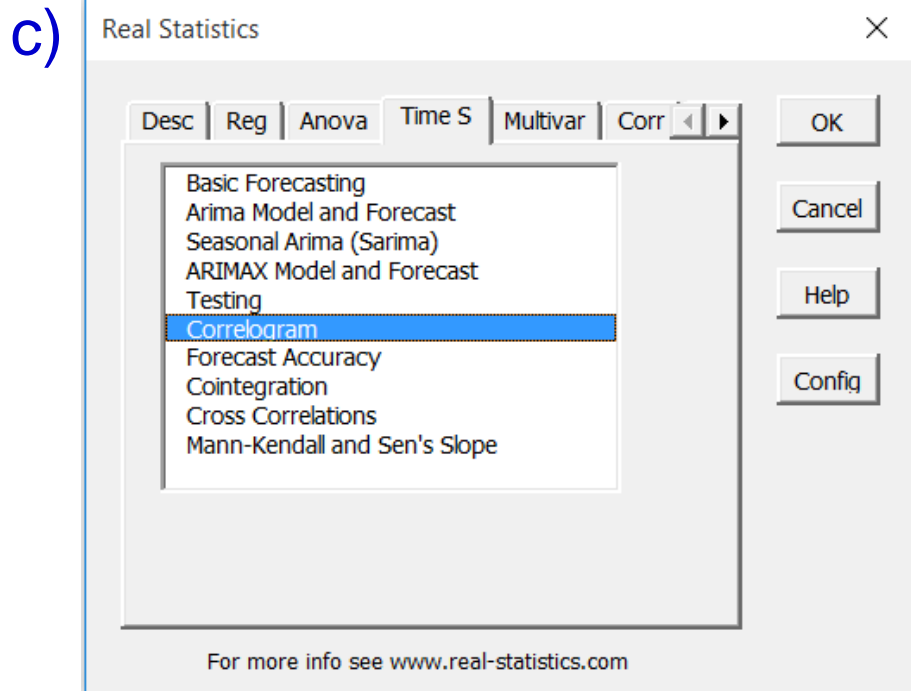
19	-0,10482759	-0,54359439	0,54359439
20	-0,11034483	-0,5462656	0,5462656
21	-0,11586207	-0,54921022	0,54921022

-0,8

2_Ejemplo_auto_correlacion_temperatura.xlsx

Real Statistics Resource Pack

- Activación RSRP: Archivo > Opciones > Complementos > Complementos de Excel > Ir... > Xrealstats > Examinar > XRealStats.xlam > Aceptar
- Ctrl + m



Análisis de datos:

3) Eliminar datos divergentes

Análisis de datos:

3) Eliminar datos divergentes



3.1 Elaborar gráficos

- a) Dispersión $X \times Y$
- b) Histogramas de X y de Y
- c) Box plot de X y de Y
- d) Identificar candidatos *para puntos divergentes*

3.2 *Probar estadísticamente* los candidatos a puntos divergentes

- a) IQR (rango intercuartílico – *Interquartile range*):
requiere simetría en los datos
- b) Prueba de Grubbs:
requiere datos que se adhieran al FDP normal

Vamos a Excel: archivo

3_Ejemplo_datos_divergentes_temperatura.xlsx



UFSB
UNIVERSIDADE FEDERAL
DO SUL DA BAHIA

Análisis de datos:



3) Eliminar datos divergentes

3.1 Elaborar gráficos

a) Tendencia t

b) Dispersión σ

c) Histogramas

d) Box plot de

e) Identificar ca

Tabla H-T: Histograma de datos

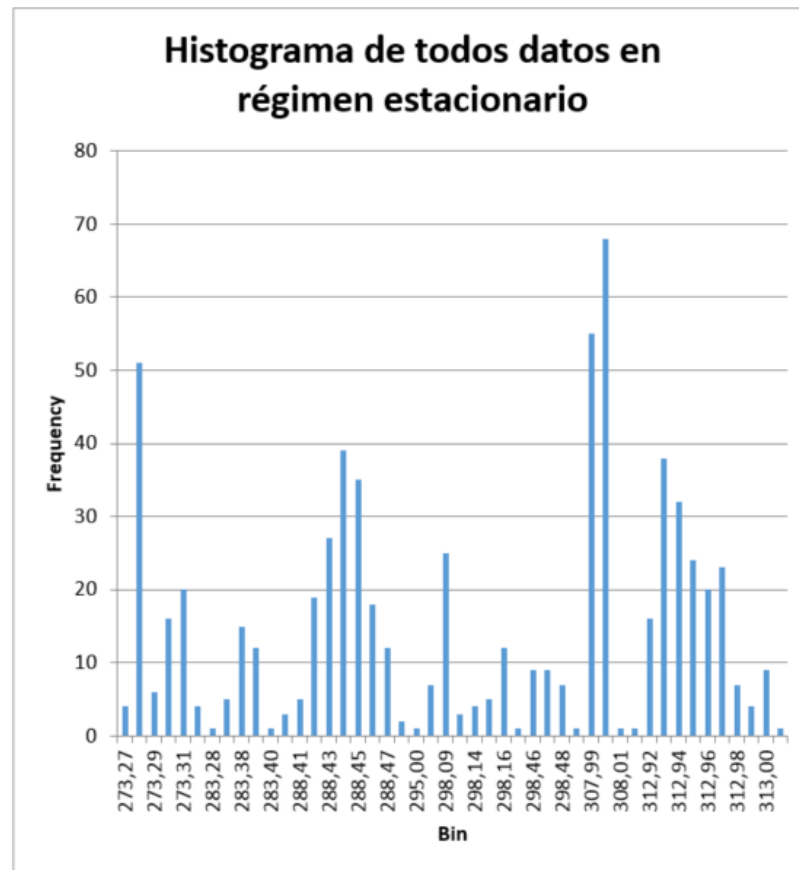
Frequency Table

item	freq
273,27	4
273,28	51
273,29	6
283,28	1

Tabla H-Tt: Histograma de todos datos en régimen estacionario.

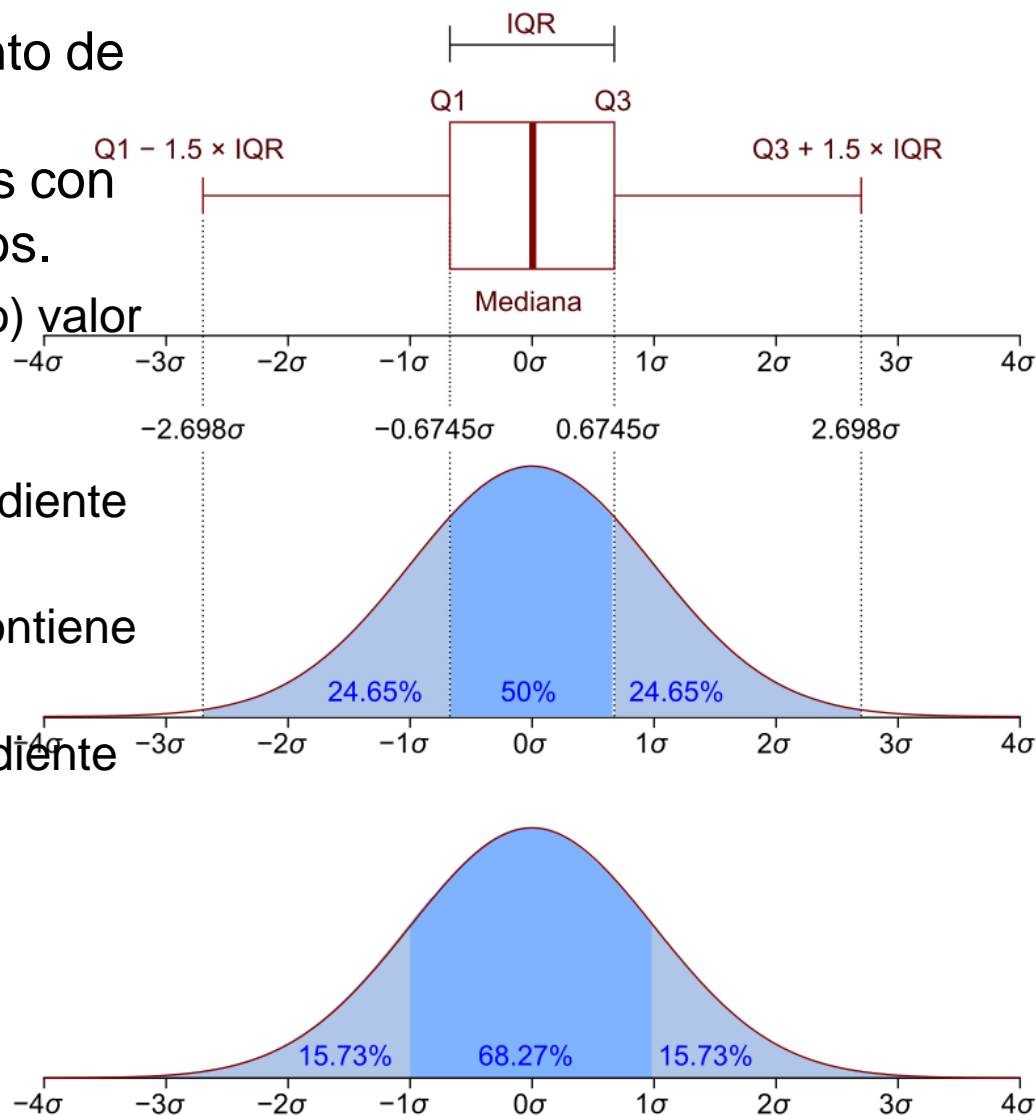
Frequency Table

item	freq	prob
273,27	4	0,0059
273,28	51	0,07522
273,29	6	0,00885
273,3	16	0,0236
273,31	20	0,0295
273,32	4	0,0059
283,28	1	0,00147
283,37	5	0,00737
283,38	15	0,02212
283,39	12	0,0177
283,4	1	0,00147
288,4	3	0,00442
288,41	5	0,00737
288,42	19	0,02802
288,43	27	0,03982
288,44	39	0,05752
288,45	35	0,05162
288,46	18	0,02655
288,47	12	0,0177
288,48	2	0,00295
295	1	0,00147
298,08	7	0,01032
298,09	25	0,03687
298,1	3	0,00442
298,14	4	0,0059
298,15	5	0,00737



Box Plot

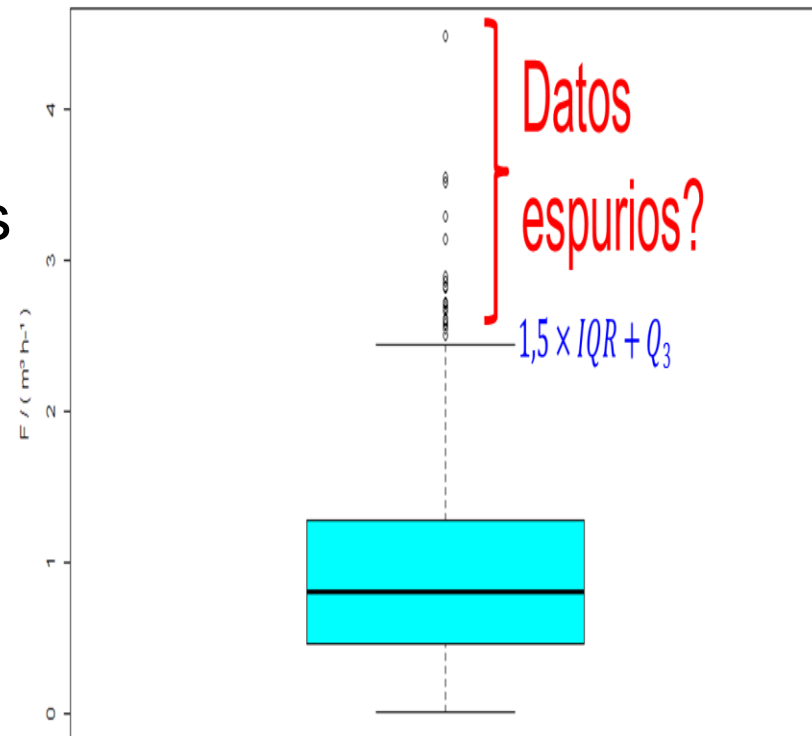
- Los cuartiles dividen un conjunto de valores ordenados, en orden ascendente, en 4 subconjuntos con un número similar de elementos.
- Q_0 (cuartil de orden cero o mínimo) valor correspondiente al 25 % de los elementos
- Q_1 (primer cuartil) valor correspondiente al 25 % de los elementos
- Q_2 (segundo cuartil o mediana) contiene el 50 % de los elementos
- Q_3 (tercer cuartil) valor correspondiente al 75 % de los elementos
- Q_4 (cuarto cuartil o máximo) valor correspondiente al 100 % de los elementos
- $IQR = IIQ = DIQ = Q_3 - Q_1$: Distancia entre cuartiles o rango intercuartílico



Detección de datos divergentes: Método *IQR*

- Premisa: datos tienen distribución simétrica (FDP simétrica)
- IQR* : Distancia entre cuartiles o rango intercuartílico
- $IQR = Q_3 - Q_1$
- Los datos espúrios o divergentes son los que están por encima de $Q_3 + 1,5 \times IQR$ o por debajo de $Q_1 - 1,5 \times IQR$

Box-plot da Vazão



Vamos a Excel: arquivo

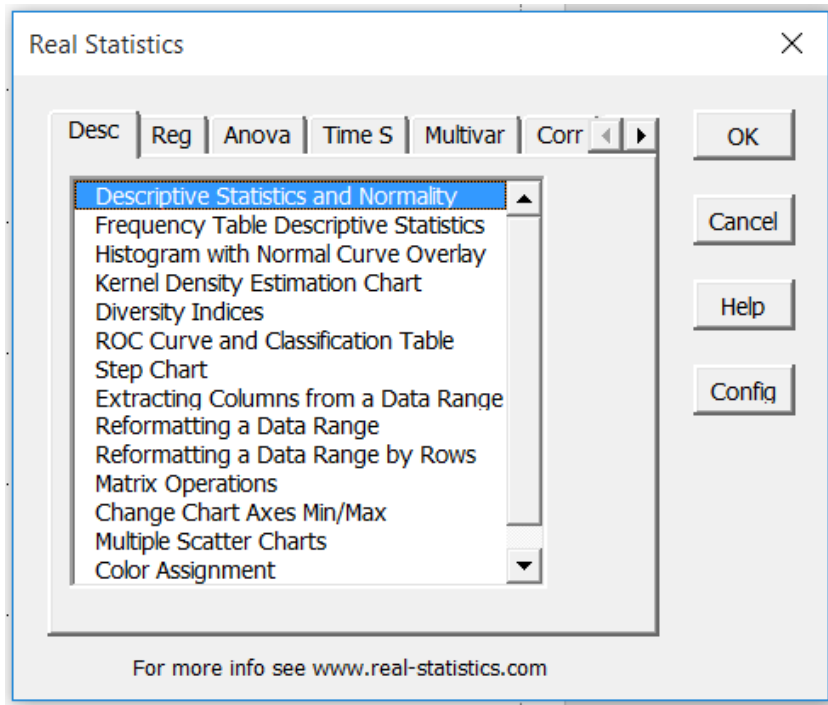
3_Ejemplo_datos_divergentes_temperatura.xlsx

Real Statistics Resource Pack

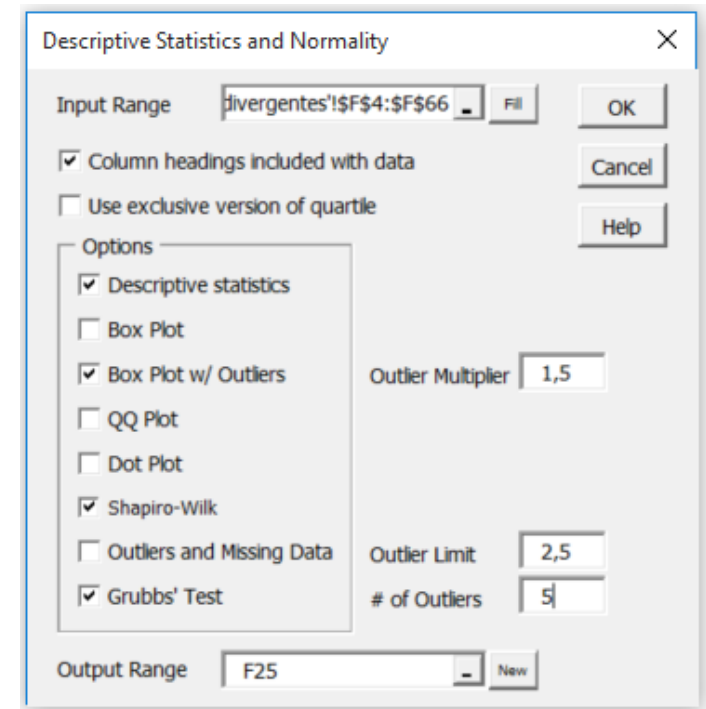
Identificación de puntos divergentes

a) **Ctrl + m**

b)



c)



3) Eliminar datos divergentes

Temperatura experimental de los datos no tratados

Box Plot datos no tratado, para temperatura 273,15 K (0 °C)



Datos a 273,15 K (0 °C):

- IQR: datos divergentes: 273,29 ; 273,29 ; 273,27 ; **283,28** ; 273,27 ; 273,27 ; 273,29 ; 273,29 ; 273,27 ; 273,29 ; 273,29
- Grubbs para nivel de significancia de 5 % ($\alpha = 5 \%$):

outlier	283,280	273,270	273,270	273,270	273,270
G	7,747	2,538	2,711	2,925	3,200
G-crit	3,212	3,206	3,200	3,193	3,187
sig	yes	no	no	no	yes

¡Vamos a practicar!

Eliminación de datos divergentes

Practica: eliminación de datos divergentes

Archivo: **4_Eliminacion_datos_divergentes.xlsx**



Procedimiento sugerido para la eliminación de datos divergentes:

1. Seleccione datos la región en régimen permanente
2. Elabore gráfico de tendencia de todas las magnitudes
3. Elabore histogramas de todas las magnitudes
4. Elabore box plot de todas las magnitudes
5. Seleccione en los gráficos puntos candidatos
6. Defina fator multiplicador da prueba *IQR*, generalmente 1,5
7. A partir do *IQ*, seleccione puntos candidatos
8. Defina nivel de confianza (*NC*), generalmente 95 %
9. A partir da prueba de Grubbs, seleccione puntos candidatos
10. Elimine los puntos que considere espurios, resalte los motivos de su decisión;

Eliminar líneas (datos de patrones y objetos) que contienen puntos espurios

Eliminar uno (1) dato a cada ciclo

Evaluar estadísticamente los datos

Análisis de datos:

4) Evaluar estadísticamente los datos

4.1 Calcular estadísticas

- a) Media (aritmética), Moda (valor más frecuente), Mediana (Q_2)
- b) Mínimo (Q_0), Máximo (Q_4),
Rango o Amplitud = Máximo - Mínimo ($Q_4 - Q_0$)
- c) $Q_0 ; Q_1 ; Q_2 ; Q_3 ; IQR = Q_3 - Q_1$
- d) Desviación típica experimental o
Desviación estándar experimental
- e) Desviación típica experimental de la media o
Desviación estándar experimental de la media o
Error estándar
- f) Desviación típica experimental relativa
- g) Desviación típica experimental relativa de la media
- h) Asimetría (*skewness*)
- i) Curtosis (*kurtosis*)
- j) AAD: promedio de las desviaciones absolutas de la media

$$AAD = \frac{\sum_{k=1}^n |x_k - \bar{x}|}{n} = DESVPROM(*)$$

Estadísticas

Media (aritmética): $\bar{x} = \frac{\sum_{k=1}^n x_k}{n}$

Moda: M_o valor más probable o valor más frecuente

Mediana: M_o valor que corresponde a 50 % puntos

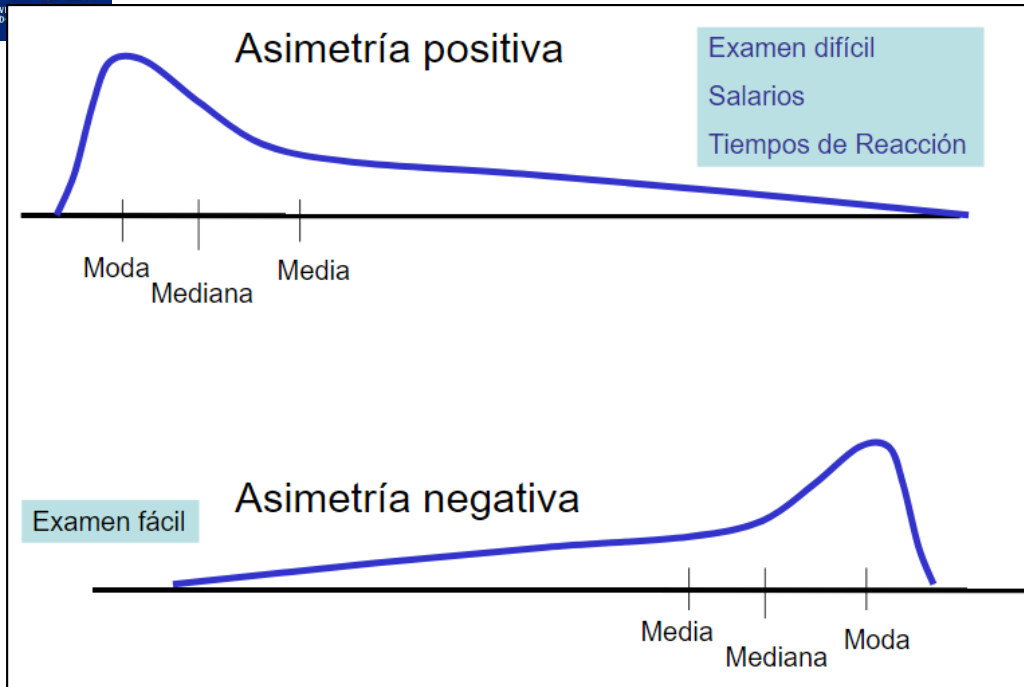
Desviación típica experimental: $s = \sqrt{\frac{\sum_{k=1}^n (x_k - \bar{x})^2}{(n-1)}}$

Desviación típica experimental de la media: $s_{\bar{x}} = \frac{s}{\sqrt{n}}$

Desviación típica experimental relativa
o coeficiente de variación: $CV = \frac{s}{\bar{x}}$

Desviación típica experimental relativa de la media
o coeficiente de variación de la media: $CV_{\bar{x}} = \frac{s_{\bar{x}}}{\bar{x}}$

Índices de asimetría (M)



Fuente: <https://slideplayer.es/slide/106388/>

Índice de asimetría de Pearson

$$M = \frac{\bar{x} - Mo}{s}$$

Excel:
$$M = \frac{n}{(n-1) \cdot (n-2)} \cdot \frac{\sum_{k=1}^n (x_k - \bar{x})^3}{s^3}$$

Ref.1: Ralph B. D'Agostino, Albert Belanger; Ralph B. D'Agostino, Jr. **A Suggestion for Using Powerful and Informative Tests of Normality**. The American Statistician, Vol. 44, No. 4 (Nov., 1990), p. 316-321.

Ref.2: D. N. Joanes and C. A. Gill. Comparing Measures of Sample Skewness and Kurtosis. Journal of the Royal Statistical Society. Series D (The Statistician). Vol. 47, No. 1 (1998), pp. 183-189. Identificador: ISSN: 0039-0526 ; E-ISSN: 1467-9884 ; DOI: 10.1111/1467-9884.00122

- Mo : Moda
- Me : Mediana

FPD simétrica $\rightarrow M = 0$

Asimetría positiva $\rightarrow M > 0$
 $\bar{x} > Me > Mo$

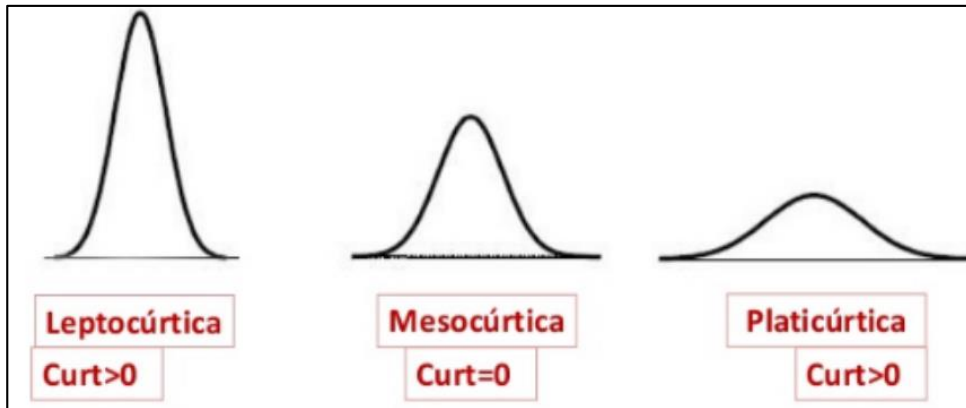
Asimetría negativa $\rightarrow M < 0$
 $\bar{x} < Me < Mo$

Índice de asimetría de Fisher

$$M = \frac{\sum_{k=1}^n (x_k - \bar{x})^3}{n \cdot s^3}$$

Sensible a valores extremos

Índices de curtosis (K) o apuntamiento



Fonte:
<https://forodeltransporteyelferrocarril.blogspot.com/2017/04/modelos-aleatorios-de-difusion-de.html?m=1>

FDP mesocúrtica (normal) $\rightarrow K = 0$

FDP leptocúrtica $\rightarrow K > 0$

FDP leptocúrtica es más probable de presentar valores más extremos que una FDP normal

FDP platicúrtica $\rightarrow K < 0$

Excel:
$$K = \frac{n \cdot (n+1)}{(n-1) \cdot (n-2) \cdot (n-3)} \cdot \frac{\sum_{k=1}^n (x_k - \bar{x})^4}{s^4} - 3 \frac{(n-1)^2}{(n-2) \cdot (n-3)}$$

Ref.1: Ralph B. D'Agostino, Albert Belanger; Ralph B. D'Agostino, Jr. **A Suggestion for Using Powerful and Informative Tests of Normality**. The American Statistician, Vol. 44, No. 4 (Nov., 1990), p. 316-321.

Ref.2: D. N. Joanes and C. A. Gill. Comparing Measures of Sample Skewness and Kurtosis. Journal of the Royal Statistical Society. Series D (The Statistician). Vol. 47, No. 1 (1998), pp. 183-189. Identificador: ISSN: 0039-0526; E-ISSN: 1467-9884 ; DOI: 10.1111/1467-9884.00122

Vamos a Excel: archivo

5_Ejemplo_analisis_datos_temperatura.xlsx

Tabla E-To: Estadísticas dos datos de inmigración, para **objeto**

Punto de inicio	1	1	63	84	121	245	278	438	464	639	Unidad
Punto final	678	62	83	120	244	277	437	463	638	678	
Media:	297,030	273,280	298,154	298,089	307,996	283,383	288,441	298,468	312,950	273,307	K
Moda:	308,000	273,280	298,160	298,090	308,000	283,380	288,440	298,460	312,930	273,310	K
Mediana:	298,150	273,280	298,160	298,090	308,000	283,380	288,440	298,470	312,950	273,310	K
Mínimo:	273,270	273,270	298,140	298,080	307,990	283,370	288,400	298,450	312,910	273,300	K
Máximo:	313,010	273,290	298,160	298,100	308,010	283,400	288,480	298,480	313,010	273,320	K
Amplitud o Rango:	39,740	0,020	0,020	0,020	0,020	0,030	0,080	0,030	0,100	0,020	K
Q0	273,270	273,270	298,140	298,080	307,990	283,370	288,400	298,450	312,910	273,300	K
Q1	288,430	273,280	298,150	298,090	307,990	283,380	288,430	298,460	312,930	273,300	K
Q2	298,150	273,280	298,160	298,090	308,000	283,380	288,440	298,470	312,950	273,310	K
Q3	312,920	273,280	298,160	298,090	308,000	283,390	288,450	298,478	312,965	273,310	K
Q4	313,010	273,290	298,160	298,100	308,010	283,400	288,480	298,480	313,010	273,320	K
Desviación estándar experimental	14,111	0,004	0,008	0,005	0,005	0,008	0,017	0,009	0,022	0,006	K
Cantidad de puntos	678	62	21	37	124	33	160	26	175	40	
Desviación estándar experimental de la media	0,542	0,001	0,002	0,001	0,000	0,001	0,001	0,002	0,002	0,001	K
Desviación estándar experimental relativo	4,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	%
Desviación estándar experimental relativo de la media	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	%
Asimetría	-0,359	0,320	-0,844	-0,065	-0,080	-0,064	-0,090	-0,061	0,678	0,380	(K) ³
Curtosis	-1,230	2,886	-0,865	0,576	-1,572	-0,445	-0,244	-0,973	-0,205	-0,636	((K) ²) ²
AAD	12,525	0,002	0,007	0,003	0,005	0,006	0,013	0,007	0,018	0,006	K
Desviación estándar experimental relativo de la media	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	%
Asimetría	-0,358	1,235	-0,453	-0,284	0,294	-0,349	-0,191	-0,220	0,675	0,310	(K) ³
Curtosis	-1,231	-0,491	-1,095	-2,032	3,015	-0,581	-0,287	-0,528	-0,182	-0,601	((K) ²) ²
AAD	12,524	0,004	0,007	0,005	0,002	0,005	0,014	0,007	0,018	0,005	K

¿La asimetría es grande o pequeña?

¿La curtosis está cerca de la DPF normal?

Para responder a estas preguntas tenemos que establecer el intervalo de confianza de las estadísticas estimadas.

Vamos a Excel: archivo **5_Ejemplo_analisis_datos_temperatura.xlsx**
hoja **Estadísticas**

Posibles escenarios

ii) Desviación estándar del patrón

no es significativamente más pequeño
que desviación estándar del objeto

ii.a) Rehacer el experimento con un patrón con menor incertidumbre

$$\left(u_p < \frac{u_o}{10}\right)$$

ii.b) **Propagar la incertidumbre de la magnitud independiente para la magnitud dependiente**

ii.c) Considerar la incertidumbre de la magnitud independiente en la regresión → regresión con simultánea reconciliación de datos

ii.d) Considerar la moda para representar magnitud independiente e magnitudes dependientes, y desviación experimental da magnitudes dependientes (no utilizar desviación experimental de la media)

ii.e) Eliminar una (o más de una) cifra decimal de las indicaciones del instrumento objeto → aumentar la incertidumbre de las indicaciones.

Vamos a Excel: archivo **5_Ejemplo_analisis_datos_temperatura.xlsx**
hoja **Incetidumbre**

Intervalo de confiança
X
Intervalo de cobertura

Nivel de confianza NC

Nivel de Significancia NS

- Está relacionado con la probabilidad del intervalo de confianza contener los valores verdaderos
- Cuanto mayor o NC mayor la probabilidad del intervalo de confianza no errar
- $NC = 100 \%$ siempre el intervalo ira contener el valor verdadero, pero en este caso el intervalo de confianza no contiene información
- NS : nivel de significancia ($p - value = NS/100$):
$$NC + NS = 100 \% \Rightarrow NC = 100 \% - NS,$$
$$p = 0,05 \Rightarrow NS = 5 \% \Rightarrow NC = 95 \%;$$
$$p = 0,01 \Rightarrow NS = 1 \% \Rightarrow NC = 99 \%.$$

Intervalo de confianza **x** Intervalo de cobertura

El **intervalo de confianza** es un rango que podría contener “(1.28) ... el valor verdadero del parámetro en una larga serie de muestras aleatorias (1.6) repetidas en circunstancias idénticas. Un intervalo de confianza no refleja la probabilidad (2.5) de que el rango observado contenga el valor verdadero del parámetro (el rango contiene el valor verdadero o no). ”ISO 3534-1

VIM 2.36

intervalo de cobertura, *m*

intervalo que contiene el conjunto de **valores verdaderos** de un **mensurando** con una probabilidad determinada, basada en la información disponible

NOTA 1 El intervalo de cobertura no necesita estar centrado en el **valor medido** elegido (véase la Guía ISO/IEC 98-3:2008/Supl. 1).

NOTA 2 El intervalo de cobertura no debería denominarse “intervalo de confianza”, evitando así confusión con el concepto estadístico (véase la Guía ISO/IEC 98-3:2008, 6.2.2).

NOTA 3 El intervalo de cobertura puede obtenerse de una **incertidumbre expandida** (véase la Guía ISO/IEC 98-3:2008, 2.3.5)."

Intervalo de confianza x Intervalo de cobertura

Estadística

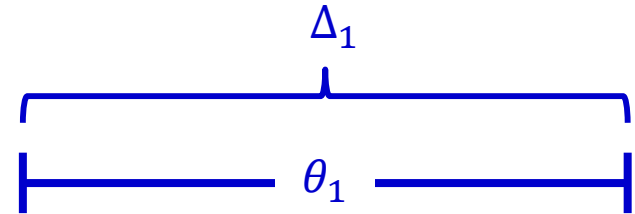
- 1) Considere solo la variabilidad de la muestra
- 2) Nivel de confianza**
(E: *confidence level* | P: *nível de confiança*)
- 3) Quantile (t)
- 4) Intervalo de confianza (Δ_e)
(*confidence interval*)
- 5) Grados de libertad:
$$GL = n - 1$$
- 6) FDP ~ Normal o Student, Δ_e para media (sistema simétrico):
$$\Delta_e: \bar{x} - t \cdot s_{\bar{x}} \leq \bar{x} \leq \bar{x} + t \cdot s_{\bar{x}}$$
$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$
- 7) Erro padrão (*standard error - se*):
$$ep = s_{\bar{x}}$$

Metrología

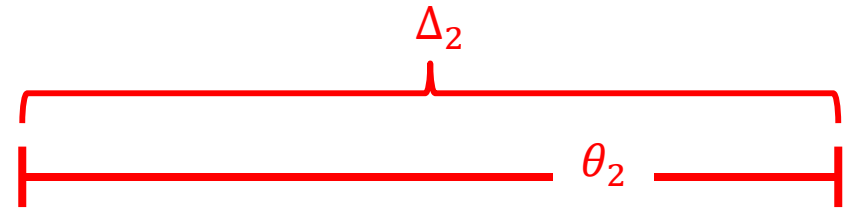
- 1) Considere la variabilidad de la muestra e del sistema de medición
- 2) Probabilidad de cobertura** (VIM, 2.37)
(E: *level of confidence* | P: *nível da confiança*)
- 3) Fator de cobertura (k) (VIM, 2.38)
- 4) Intervalo de cobertura (Δ_m) (VIM, 2.36)
(*coverage interval*)
- 5) Grados de libertad: $GL_A = n - 1$
$$GL = f(GL_A, GL_B)$$
- 6) FDP ~ Normal o Student, Δ_m para media (sistema simétrico):
$$\Delta_m: \bar{x} - k \cdot u_c \leq \bar{x} \leq \bar{x} + k \cdot u_c$$
$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = u_A ; u_c = \sqrt{u_A^2 + \sum_{k=1}^{n_B} u_{B_k}^2}$$
- 7) Incertidumbre estándar combinada: u_c

Comparación entre estimaciones o Prueba de hipótesis monovariable

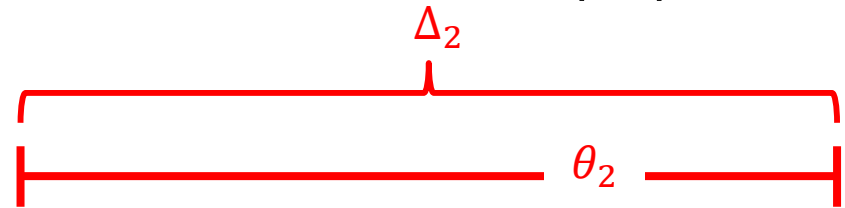
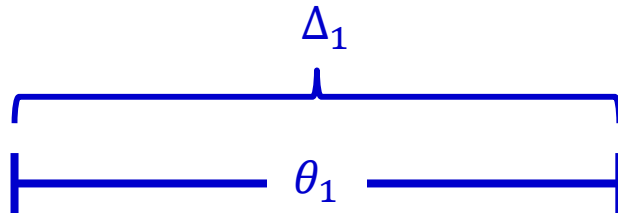
Estadística θ_1 y respectivo
intervalo de cobertura Δ_1



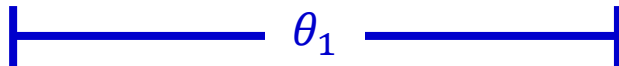
Estadística θ_2 y respectivo
intervalo de cobertura Δ_2



Estadísticas distintas: $\theta_1 \neq \theta_2$ intervalos de cobertura sin superposición



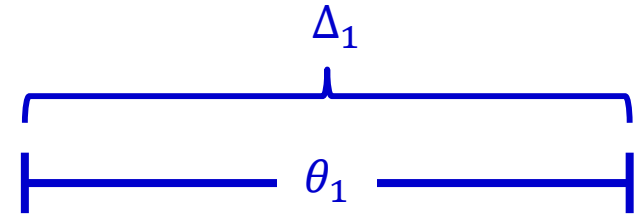
Estadísticas iguales: $\theta_1 = \theta_2$ intervalos de cobertura con superposición



Comparación entre estimaciones o Prueba de hipótesis monovariable

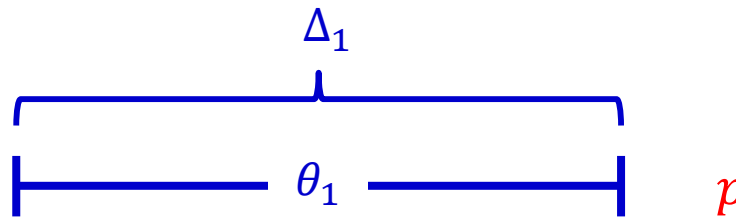


Estadística θ_1 y respectivo
intervalo de cobertura Δ_1



Parámetro fijo (constante) p

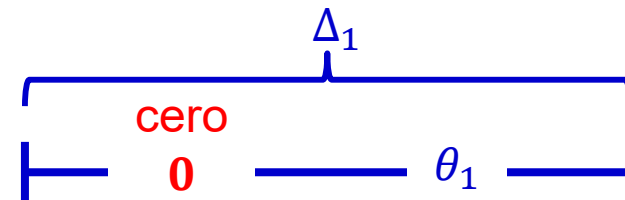
Estadísticas distintas: $\theta_1 \neq p$ intervalo de cobertura sin superposición



Estadística θ_1 igual a p : $\theta_1 = p$ intervalo de cobertura con superposición

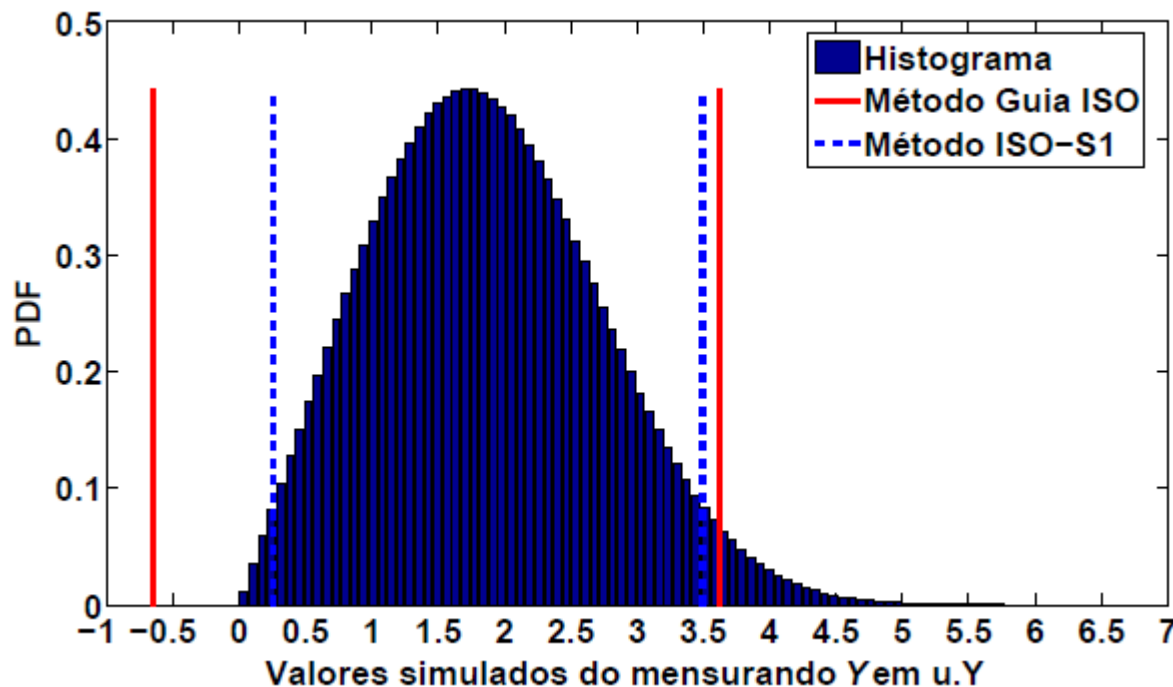


Estadísticas θ_1 sin
significado estadístico



Intervalo de cobertura para sistemas no lineal

Función de medición: $Y = \sqrt{X_1^2 + X_2^2}$



Kalid, et al. *Comparação entre os métodos linear e não linear para a avaliação da incerteza de medição.* Controle & Automação Sociedade Brasileira de Automatica. Nov. 2010, 21(6)
DOI: 10.1590/S0103-17592010000600002

$$IC_{LPI} = 2U$$

Método LPI - Ley de Propagación de la Incertidumbre
Propagación de varianza del modelo linealizado.

$$IC_{MC} = [LI; LS]$$

Método Monte Carlo o Inferencia Bayesiana
Propagación de FDP del modelo no lineal



REGIONAL FUND QUALITY INFRASTRUCTURE FOR BIODIVERSITY & CLIMATE PROTECTION
IN LATIN AMERICA AND THE CARIBBEAN

Workshop on Statistics, data analysis and measurement uncertainty for Meteorology

Fundamentos estadísticos para la evaluación de la incertidumbre en las calibraciones



Márcio A. A. Santana – INPE/BR
marcio.santana@inpe.br

Ricardo de A. Kalid – UFSB/BR
kalid@ufsb.edu.br



TECLIM-UFSB