

Pruebas paramétricas en muestreo de anhídrido carbónico en espacio cerrado con circulación vehicular

Parametric tests in carbon anhydride sampling In closed space with vehicular circulation

Ramiro Hernán Polanco Contreras¹

Juan Alejandro Chica García²

Ángel Alejandro Rodríguez Aya³

John Alejandro Figueredo Luna⁴

Sabina Marcela Velásquez García⁵

Universidad Nacional Abierta y a Distancia, Colombia

Resumen

Como parte de los procedimientos para el establecimiento de las bondades del diseño del medidor de calidad del aire, en particular en la determinación del nivel de CO_2 , se parte de un análisis de factores poblacionales referenciales a partir de un tamaño de muestra suficientemente grande, obtenidos con un medidor estándar certificado, es indudable tener presente que el tratamiento estadístico genera una mayor bondad y exactitud con tamaños de muestra más grandes. Se debe partir, entonces, de la especificidad en la forma de la distribución la se constituye en la base para determinar las pruebas posteriores de aceptación o rechazo de las hipótesis formuladas. En este documento se hace un recorrido por las pruebas paramétricas de aplicación a la muestra de datos obtenidos con el medidor de referencia y que serán la base de comparación con los resultados obtenidos partir del diseño de medidor de calidad del aire diseñado a bajo costo.

¹ <https://orcid.org/0000-0002-1410-6733/> ramiro.polanco@unad.edu.co

² <https://orcid.org/0000-0002-4969-2402/> juan.chica@unad.edu.co

³ <https://orcid.org/0000-0002-0498-8648/> angel.rodriguez@unad.edu.co

⁴ <https://orcid.org/0000-0003-0006-4234/> john.figueredo@unad.edu.co

⁵ <https://orcid.org/0000-0001-5697-2712/> smvelasquezga@unadvirtual.edu.co

Palabras clave: pruebas paramétricas, anhídrido carbónico, concentración.

Abstract

In the procedures for establishing a good air quality meter design, in particular in the level CO_2 determination, it is based on an analysis of population factors referential from a sufficiently large sample size. It is unquestionable to keep in mind that the statistical treatment generates greater goodness and accuracy with larger sample sizes. The data analysis must be based on the distribution form that is the basis for determining the subsequent evidence of acceptance or rejection of the hypotheses formulated. This document makes the data parametric tests tour. The results will be the basis of comparison with the results obtained from the low-cost designed air quality meters.

Keywords: Parametric tests, Carbon dioxide, Concentration.

1. Introducción

Una vez obtenidos los datos de procesos muestrales, es relevante comprender el comportamiento y las características de los mismos, para el presente trabajo se pretende la caracterización de 1.024 muestras de anhídrido carbónico, lecturas obtenidas en un espacio cerrado con circulación vehicular en la ciudad de Villavicencio, obtenidas con un medidor estándar certificado. El documento inicia con una descripción de estadígrafos de tendencia central, dispersión y gráficos, como parte del proceso muestral en forma adicional se recopiló información de las variables temperatura ambiente y humedad relativa permitiendo inferir su nivel de incidencia en los niveles del gas de interés, para ello se construyeron gráficas de dispersión y el estimado del coeficiente de correlación de Pearson, para luego aplicar tres pruebas en su orden t de Student, Prueba de Shapiro-Wilks, finalizando con la prueba de bondad de ajuste de Kolmogorov.

2. Metodología

El equipo fue instalado en las áreas de circulación peatonal dentro de un parqueadero subterráneo, mantenido en funcionamiento durante 4.320 minutos separados en nueve bloques de 480 minutos cada uno, teniendo en cuenta los periodos de mayor circulación tanto de vehículos como de peatones; el muestreo se realizó cada cuatro minutos, obteniendo así un total de 1.024 lecturas. Las mismas se procesaron usando el software Infostat para el cálculo de los diferentes estadígrafos asociados con la distribución

y la escala de medida⁶. De igual forma en los análisis se incorporaron cruces de la variable de interés con dos parámetros a saber: temperatura y humedad relativa, los cuales se considera son modificadores en la concentración de gas.

En este punto se considera importante resaltar la bondad a partir de una comparación de promedios (Distribución t de Student | | UPV, 2011)

- Distribución normal a través de la t-Student.
- Distribución no normal Wilcoxon o U Mann Whitney.

Correlación de variables cuantitativas,

- Distribución normal muestras independientes coeficiente de correlación de Pearson.
- Cuantitativas discontinuas y continuas sin distribución normal muestras independientes coeficiente de correlación de Spearman.

3. Discusión y resultados

A partir de los datos obtenidos se realiza una caracterización general de la muestra con base en las medidas resumen arrojadas por el software, así pues se cuenta con un total de 1.024 elementos muestrales con media en el nivel de CO_2 de 683,54 ppm con un grado de dispersión medio de 141,31 ppm, las varianzas tanto muestral como poblacional tienen una diferencia de 20,5 puntos indicando una buena aproximación de la muestra al comportamiento poblacional; el valor mínimo del gas fue de 457 ppm, mientras el máximo alcanzó los 1.424 ppm, el rango de variación es amplio con aproximadamente 1000 ppm entre los extremos de la muestra.

En cuanto a su distribución los datos se encuentran ligeramente por encima el valor de la media con un valor de asimetría cercano al 0,87 por ende los datos se distribuyen como los define Corporate Finance Institute (2020) uniformemente alrededor de este estadígrafo central, como se muestra en la Figura 1.

⁶ "Normally, when one hears the term measurement, they may think in terms of measuring the length of something (ie. the length of a piece of wood) or measuring a quantity of something (ie. a cup of flour). This represents a limited use of the term measurement. In statistics, the term measurement is used more broadly and is more appropriately termed scales of measurement. Scales of measurement refer to ways in which variables/numbers are defined and categorized. Each scale of measurement has certain properties which in turn determines the appropriateness for use of certain statistical analyses" (Topic #1: Introduction to measurement and statistics, s. f.).

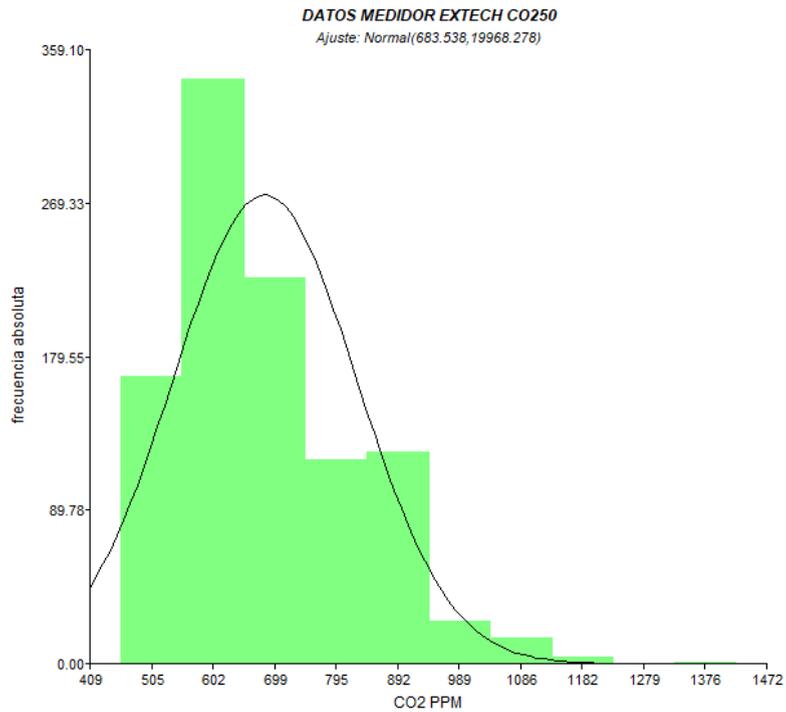


Figura 1. Distribución de datos muestrales. Fuente: los autores.

La situación anterior se corrobora con el valor de la curtosis identifica el comportamiento en comparación a la distribución normal mostrando un comportamiento leptocúrtico, en la Tabla 1, a continuación se consolidan las medidas resumen.

Tabla 1. Medidas resumen

Nueva tabla : 28/04/2021 - 12:22:16 PM - [Versión : 20/09/2019]

Medidas resumen

Variable	n	Media	D.E.	Var(n-1)	Var(n)	CV	Mín	Máx	Asimetría	Kurtosis	MAD
CO2 PPM	1024	683.54	141.31	19968.28	19948.78	20.67	457.00	1424.00	0.87	0.77	89.50

Fuente: los autores

Por su parte la distribución t de Student se construye como un cociente entre una normal y la raíz de chi cuadrado independientes ambas entre sí, el aumento de los grados de libertad en la distribución de su comportamiento se asocia al de una normal en los términos a continuación:

$$t_{n \rightarrow \infty} \rightarrow N(0, 1)$$

Si el valor de la media y la desviación típica en una muestra obtenida de una población normal el comportamiento sería una distribución t de Student con $n - 1$ grados de libertad por ello para un valor poblacional se tiene:

$$\frac{\bar{x} - m}{\frac{\sigma}{\sqrt{N}}} \sim N(0, 1)$$

Mientras que el valor muestral es:

$$\frac{\bar{x} - m}{\frac{s}{\sqrt{N}}} \sim t_{N-1}$$

Así pues la distribución t de Student cuenta con características del ser continua, con media 0 simétrica respecto a la misma, con forma acampanada, se extiende desde $-\infty$ hasta $+\infty$, y con varianza 1, con grados de libertad > 2 , lo que la hace aproximar al comportamiento de una distribución normal para grupos de datos > 30 , al aplicar la prueba t para H_0 tal que la distribución no es normal:

Si $p \leq 0,05$ se rechaza H_0

Si $p > 0,05$ no se rechaza H_0

Con los resultados de la Tabla 2 el p (*bilateral*) rechaza H_0 tal que la distribución de datos se comporta como una distribución normal.

Tabla 2. Prueba t para una media

Prueba t para una media

Valor de la media bajo la hipótesis nula: 0

Variable	n	Media	DE	LI(95)	LS(95)	T	p(Bilateral)
CO2 PPM	1024	683.54	141.31	674.87	692.20	154.79	<0.0001

Fuente: los autores.

Como factores que generan modificaciones en los niveles de CO_2 se consideran la temperatura y humedad, que durante el período testeado mostraron un comportamiento relacionado de forma directa y decreciente como se muestra en la Figura 2, con aproximadamente 17 puntos extremos, tanto inferiores como superiores, lo que indica que un 98,37 % de los pares ordenados mantienen la tendencia dejando el 1,63 % de los datos fuera del comportamiento típico, es de resaltar que la temperatura fluctuó 5 °C, siendo los valores registrados mínima 25 °C y la máxima de 30 °C; en similares circunstancias la humedad relativa registró un mínimo de 50 % y una máxima de 85 %, cuyo rango fue de 35 %; con estos datos se puede afirmar que un aumento de 1 °C refleja un descenso en humedad de 7 %; se evidencia una concentración de puntos para temperaturas superiores e 28,89 °C y menores a 26,67 °C.

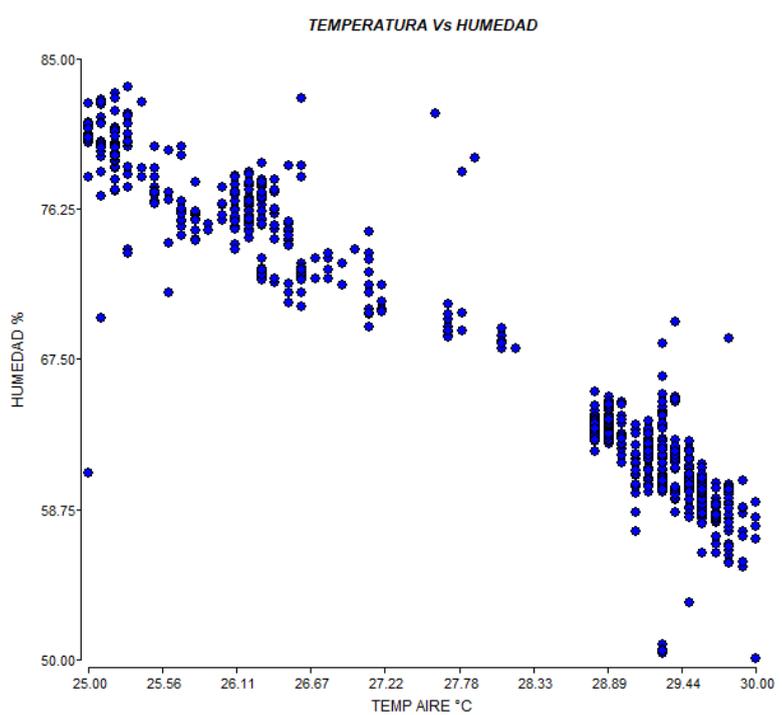


Figura 2. Dispersión temperatura y humedad.

Fuente: los autores.

El comportamiento de los datos que relacionan estas variables muestran un comportamiento no relacionado entre la temperatura y el nivel del gas, así entre 25 °C y 28 °C el comportamiento del CO_2 es amplio en un rango de 656 ppm que varía desde los 544 ppm hasta los 1.200 ppm, a pesar de la dispersión se encuentra una concentración de puntos entre los 26,1 °C y 26,67 °C con concentraciones entre los 544 ppm y 850 ppm; finalmente para temperaturas superiores a los 28,7 °C la fluctuación en la concentración de anhídrido carbónico es más amplia en la dispersión pero se evidencia mayor número de puntos en el rango de 450 ppm y 638 ppm, como se observa en la Figura 3.

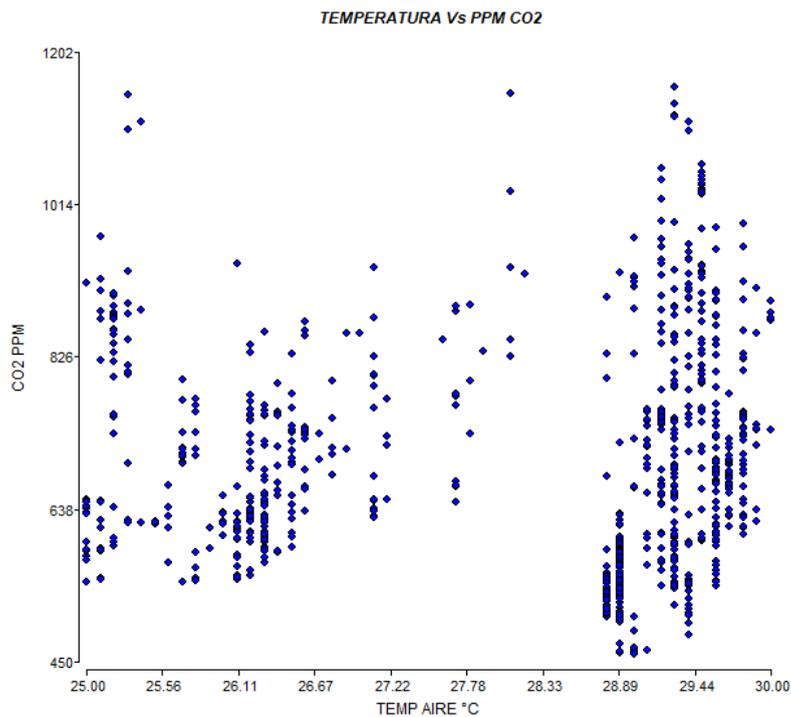


Figura 3. Dispersión temperatura y anhídrido carbónico.

Fuente: los autores.

Como punto de partida para describir el efecto de la humedad relativa, se considera la caracterización de esta por el Instituto de Hidrología, Meteorología y Estudios Ambientales (2000) para la ciudad de Villavicencio que demuestra una oscilación a lo

largo del año entre 67 % y 83 %, siendo mayor en los meses de junio y julio y menor en el primer trimestre del año, límites que permiten descartar once de las mediciones tomadas por encontrarse fuera de ellos, quedando así el 98,92 % de las mismas para su análisis. Durante el periodo de muestreo la humedad, como se observa en la Figura 4, vario en el rango entre 56 % y 85 %, sin embargo es preciso anotar la prevalencia de humedades entre los 56 % y 64 % con una concentración de anhídrido carbónico entre los 450 ppm y las 850 ppm en este rango de humedad. En humedades entre el 75 % y 85 % la concentración del gas varía entre los 550 ppm y los 950 ppm con mayores densidades de puntos entre los 550 ppm y 650 ppm.

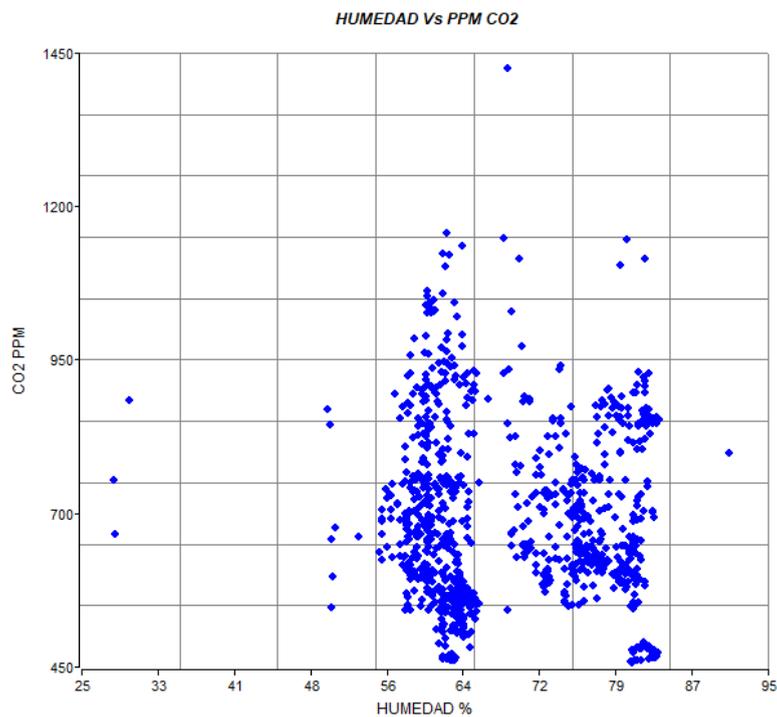


Figura 4. Dispersión humedad relativa y anhídrido carbónico

Fuente: los autores.

Estimando los valores de los coeficientes de correlación de Pearson como indicador de la relación entre variables, en la Tabla 3 se observa para la relación temperatura del aire vs concentración de anhídrido carbónico un estimado de 0,11, lo que indica una relación positiva por su dirección, así un incremento en la temperatura se asocia con un aumento en la concentración del gas, sin embargo por su magnitud en valor absoluto la relación de linealidad es baja. De igual forma al asumir un nivel de significancia α de 0,05 y comparado con el valor p cuyo resultado entre las variables arrojó 0,0005 existe evidencia concluyente sobre la significancia de la asociación entre las variables.

De igual forma para la relación entre humedad y concentración del anhídrido presenta una relación negativa con valor de -0,06 indicando que un aumento en la humedad relativa indica un descenso en la concentración del gas, nuevamente su magnitud en valor absoluto induce a una relación de linealidad baja; el valor del nivel de significación en este caso es del 0,0721 por ende existe evidencia no concluyente sobre la significancia de la asociación entre las variables. Finalmente, y como medida de control se evidencia una clara relación negativa entre la temperatura ambiente y la humedad relativa, lo que permite afirmar que un aumento en temperatura genera una reducción en la humedad, con una relación de linealidad alta al ser el valor absoluto de la correlación de Pearson 0,89 y una significancia de asociación entre variables por un valor p menor al 0,0001.

Tabla 3. Coeficientes de correlación

Nueva tabla : 28/04/2021 - 10:24:25 AM - [Versión : 20/09/2019]

Coeficientes de correlación

Correlación de Pearson

Variable(1)	Variable(2)	n	Pearson	p-valor
CO2 PPM	CO2 PPM	1024	1.00	<0.0001
CO2 PPM	HUMEDAD %	1024	-0.06	0.0721
CO2 PPM	TEMP AIRE °C	993	0.11	0.0005
HUMEDAD %	CO2 PPM	1024	-0.06	0.0721
HUMEDAD %	HUMEDAD %	1024	1.00	<0.0001
HUMEDAD %	TEMP AIRE °C	993	-0.89	<0.0001
TEMP AIRE °C	CO2 PPM	993	0.11	0.0005
TEMP AIRE °C	HUMEDAD %	993	-0.89	<0.0001
TEMP AIRE °C	TEMP AIRE °C	993	1.00	<0.0001

Fuente: los autores.

Manera de estimar si los datos se comportan bajo el modelo de distribución normal la prueba de Shapiro Wilks, plantea la hipótesis nula de que una muestra proviene de una distribución normal eligiendo nivel de significancia del 0.05, manteniendo como hipótesis alterna que la distribución no es normal, tomando en cuenta que el p – *valor* es inferior al nivel de significancia se rechaza H_1 , como se observa en los resultados de la Tabla 4.

$$H_0: X \sim N(\mu, \sigma^2)$$

$$H_1: X \neq N(\mu, \sigma^2)$$

Tabla 4. Prueba de Shapiro-Wilks

Nueva tabla : 28/04/2021 - 9:21:24 AM - [Versión : 20/09/2019]

Shapiro-Wilks (modificado)

Variable	n	Media	D.E.	W*	p(Unilateral D)
CO2 PPM	1024	683.54	141.31	0.95	<0.0001

Fuente: los autores.

El comportamiento paramétrico de los datos del gas anhídrido de carbono frente a un comportamiento normal típico, con un nivel de significación del 0,05 para la prueba de bondad de Kolmogorov tal que $\max_i |F(X_i) - S(X_i)|$, sujeto a rechazo de H_0 Si

$$\left(\sqrt{n} - 0,12 + \frac{0,11}{\sqrt{n}}\right)D > c_\alpha$$

Con un valor D de 0,51 y un el $p - valor$ de 0,0001 es inferior, se da por aceptada H_1 como se observa en los cálculos de la Tabla 5.

Tabla 5. Prueba de bondad de ajuste

Nueva tabla : 30/04/2021 - 4:36:41 PM - [Versión : 20/09/2019]

Prueba de bondad de ajuste (Kolmogorov)

Variable	Ajuste	media	varianza	n	Estadístico D	p-valor
CO2 PPM	Normal(683.84,141.31)	683.54	19968.28	1024	0.51	<0.0001

Fuente: los autores.

4. Conclusiones

El comportamiento de la muestra de anhídrido carbónico obtenida en espacios cerrados en la ciudad de Villavicencio, se inicia con una descripción general desde estadígrafos de tendencia central y dispersión, complementado con una interpretación de la gráfica obtenida, asociada a un comportamiento normal; al contar con tres variables, Temperatura ambiente, Humedad relativa y Anhídrido carbónico, se complementa el análisis con la revisión bidimensional, no solo de las gráficas de dispersión, sino a partir del coeficiente de correlación de Pearson usado debido a la independencia de las unidades de las variables, se cuenta además con una muestra significativa de 1.024 datos, permitiendo mayor exactitud de la estimación, ambas variables son continuas y su distribución se asemeja a la curva normal; con ello según el resultado de la prueba de Shapiro-Wilks se considera el comportamiento de la variable de interés como normal. Por su parte, la relación de condicionalidad de concentración del gas anhídrido carbónico por acción de la temperatura ambiente y humedad relativa no muestra una completa condicionalidad, pero sí una relación entre cada una de las variables independientes y la dependiente.

Referencias

Corporate Finance Institute. (15 de mayo de 2020,). *Kurtosis*.

<https://corporatefinanceinstitute.com/resources/knowledge/other/kurtosis/>

Dietrichson, A. (s. f.). 7.2 Prueba de Shapiro-Wilks | *Métodos Cuantitativos*. bookdown.org.

<https://bookdown.org/dietrichson/metodos-cuantitativos/test-de-normalidad.html>

Distribución t de Student | | UPV. (21 de septiembre de 2011). [Vídeo]. YouTube.

<https://www.youtube.com/watch?v=cnrpYAsM9kl>

Guasch, A., Piera, M. A., Casanovas, J. & Figueras, J. (2005). *Modelado y simulación*.

Aplicación a procesos logísticos de fabricación y servicios. Barcelona: Alfaomega.

Instituto de Hidrología, Meteorología y Estudios Ambientales. (2000). *Características climatológicas de ciudades principales y municipios turísticos*.

<http://www.ideam.gov.co/documents/21021/418894/Caracter%C3%ADsticas+de+Ciudades+Principales+y+Municipios+Tur%C3%ADsticos.pdf/c3ca90c8-1072-434a-a235-91baee8c73fc>

Minitab. (2019). Interpretar los resultados clave para Correlación - Minitab. (C) Minitab,

[LLChttps://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/how-to/correlation/interpret-the-results/key-results/](https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/how-to/correlation/interpret-the-results/key-results/)

Ríos Insua, D., Ríos Insua, S., & Martín Jiménez, J. (2009). *Simulación: métodos y aplicaciones*. Barcelona: Alfaomega.

Topic #1: Introduction to measurement and statistics. (s. f.). Cornell University.

<https://lsc.cornell.edu/wp-content/uploads/2016/01/Intro-to-measurement-and-statistics.pdf>