

Sistema IoT y validación estadística para monitoreo de salud de estudiantes

IoT system and statistical validation for monitoring students' health

J. C. Gómez-Gayosso ^a, J. Suárez-Cansino ^{b,*}, V. López-Morales ^b, A. Franco-Árcega ^b

^aUniversidad Autónoma del Estado de Hidalgo, Instituto de Ciencias Básicas e Ingeniería, Área Académica de Computación y Electrónica, Maestría en Internet de las Cosas, Colonia Carboneras, CP 42184, Mineral de la Reforma, Hidalgo, México.

^bUniversidad Autónoma del Estado de Hidalgo, Instituto de Ciencias Básicas e Ingeniería, Área Académica de Computación y Electrónica, Cuerpo Académico de Computación Inteligente, Colonia Carboneras, CP 42184, Mineral de la Reforma, Hidalgo, México.

Resumen

La combinación del Internet de las Cosas y de la Estadística para Ciencias de los Datos resulta bastante interesante, sobre todo cuando se aplican dentro del área educativa. Este trabajo propone el desarrollo de una estación de salud (kiosko) y del análisis de los datos biométricos obtenidos a través de un procedimiento experimental con el cual obtener valores de variables antropométricas que sean útiles para indicar el estado de salud de estudiantes de la Preparatoria 2, perteneciente a la Universidad Autónoma del Estado de Hidalgo. En una primera fase, el trabajo propone la medición de peso y estatura de las y los estudiantes, así como el Índice de Masa Corporal, el cual se emplea tradicionalmente para determinar sobrepeso y su relación con otros tipos de enfermedades. Los resultados experimentales muestran que adquirir esta información requiere de sensores de peso y estatura, para cuyo análisis estadístico existen modelos que permiten realizar pruebas de normalidad y homogeneidad de varianzas, las cuales resultan útiles para determinar el tipo de estadística a usar.

Palabras Clave: IoT, Normalidad, Varianza, Sensor, IMC, Sobrepeso

Abstract

By blending the Internet of Things and the Statistics in Data Sciences, one can find interesting research and application areas, especially in the Educational sector. This paper proposes the implementation of a Health Station (kiosk) and the analysis of biometrical data that an experimental procedure produces. Through the biometrical data, one can get the values of anthropometric variables that are useful to indicate the health status of some students of high school number 2 at the Universidad Autónoma del Estado de Hidalgo, México. In this first stage of the project, the paper proposes the measurement of the weight and height of students, to calculate the body mass index (BMI), and then relate it to other diseases. In order to acquire this information, weight and height sensors are required and the experimental results show that there are models that allow tests of normality and homogeneity of variances, which are useful to determine the type of statistics to use.

Keywords: IoT, Normality, Variance, Sensor, BMI, Overweight

1. Introducción

En la actualidad, el uso de la Internet de las Cosas se ha expandido al ámbito del cuidado de la salud. En ello, se utilizan diversos dispositivos sensores con varias funcionalidades.

Por ejemplo, los biosensores pueden monitorear en forma remota, diversas características de personas, las cua-

les se encuentran relacionadas con variables asociadas al estado de salud de cada una de ellas (Amaraweera and Halgamuge, 2019).

En particular, en el ámbito escolar, resulta interesante de analizar ya que los estudiantes pueden presentar diversos síntomas psicosomáticos que afecten su desempeño escolar.

* Autor para correspondencia: jsuarez@uaeh.edu.mx

Correo electrónico: julio_gomez@uaeh.edu.mx (Julio César Gómez-Gayosso), jsuarez@uaeh.edu.mx (Joel Suárez-Cansino), virgilio@uaeh.edu.mx (Virgilio López-Morales), afranco@uaeh.edu.mx (Anilú Franco Árcega).

Historial del manuscrito: recibido el 14/04/2022, última versión-revisada recibida el 29/06/2022, aceptado el 02/08/2022, publicado el 31/08/2022. **DOI:** <https://doi.org/10.29057/icbi.v10iEspecial3.9004>



Las afectaciones en la salud de los estudiantes impactan de manera considerable en su rendimiento y asistencia escolar, y en ocasiones, la ausencia de monitoreo frecuente y registro oportuno, retrasa el análisis de datos clínicos para la toma de decisiones.

Con el apoyo de un sistema de sensores, que permita la revisión constante y el registro de peso y estatura dentro del servicio médico en las instituciones de enseñanza media superior, la examinación de los datos obtenidos permitirá agilizar la detección de padecimientos, así como el tratamiento oportuno del estudiante, lo cual evitará inasistencias, aumentará su rendimiento escolar y favorecerá su salud en general.

El Internet de las Cosas en la actualidad es útil en una amplia gama de aplicaciones relacionadas con la salud, con la finalidad de monitorear y evitar complicaciones en las condiciones física y mental de las personas (Rayan et al., 2021).

Las aplicaciones médicas con IoT van adquiriendo mayor relevancia en nuestros días, y con ellas se tiene la facilidad de adquirir datos que son importantes para el monitoreo del estado de salud.

Las aplicaciones de IoT, los diversos dispositivos y el software para los móviles que permiten verificar el estado de salud de manera cómoda en casa e incluso en lugares públicos, son a menudo muy frecuentes en los tiempos actuales; sin embargo, las aplicaciones de IoT dentro de las instituciones educativas aún requieren del empuje necesario, sobre todo para la atención de las necesidades propias de estas instituciones, que incluyen el monitoreo del estado de salud de los estudiantes, y la mitigación de situaciones académicas que de ello se derivan, como son las inasistencias a las aulas (Ali and Nihad, 2021).

Encuestas aplicadas a estudiantes de diferentes carreras en un estudio llevado a cabo en una institución educativa nacional, indican que la existencia de enfermedades en una alta población de alumnas y alumnos llega a ser un factor importante en la asistencia a clases (Juárez et al., 2014).

Aunque no existe aparentemente un estudio general sobre el tema a nivel nacional, seguramente esta situación se presenta en diversos centros escolares y en todos los niveles. La formalización de un estudio de este tipo debe considerar la detección de diferentes bio-indicadores relacionados con el tópico, como por ejemplo el peso y la estatura.

En particular, se necesita un método que facilite el trabajo del personal médico (normalmente reducido) en las instalaciones escolares de educación media superior, que usualmente no permite atender de manera expedita y eficaz los diferentes problemas de salud que presentan los estudiantes en esta etapa de su formación, cuyas edades oscilan entre 15 y 20 años.

La existencia de un método para aliviar la carga de trabajo de los médicos en las instituciones escolares, facilitaría la obtención y el análisis de los datos obtenidos relativos a la población de estudiantes, y ayudaría a prevenir y/o corregir la inasistencia estudiantil a los cursos, provocada por los padecimientos.

Naturalmente que las afectaciones en la salud de los estudiantes impactan de manera considerable en sus rendimientos e inasistencias escolares y esto empeora debido a la falta del monitoreo constante y registro de la población, lo cual retrasa el análisis de los datos clínicos para la toma de decisiones. De esta forma, si se implementa un sistema de sensores, para la revisión constante y el registro de peso y estatura estudiantil dentro del servicio médico de la institución, entonces la automatización en la recuperación de los datos médicos, permitirá agilizar la detección de padecimientos derivados de la desnutrición y/o sobrepeso. Así el tratamiento oportuno, ayudaría a disminuir las inasistencias y favorecerá la salud de los estudiantes.

En este trabajo se propone apoyar en la elaboración de expedientes y reportes clínicos, así como en la toma de decisiones en el servicio médico de las instituciones de educación media superior, por medio de técnicas de Internet de las Cosas (IoT) que proporcionen datos, como el peso y la estatura estudiantil, así como indicadores de la presencia de enfermedades, por ejemplo el sobrepeso, y de técnicas estadísticas para el análisis de los datos, con la idea de tener un mejor control de la salud en la comunidad en general.

Por ello, se plantea la implementación de una estación de salud (kiosco) que permita la lectura y monitoreo de la estatura y peso de manera automatizada, por medio de un dispositivo en el cual se instalará un módulo que contendrá sensores apropiados, para tener la infraestructura mínima que con IoT logre cumplir con los requerimientos necesarios para la institución educativa.

Con los datos adquiridos, se indica un análisis estadístico con una muestra significativa de la población de estudiantes en el nivel medio superior, que permita realizar el cálculo de su Índice de Masa Corporal (IMC) y mediante éste, verificar si la población padece desnutrición o sobrepeso.

En paralelo con lo anterior, se propone resguardar los datos obtenidos con los sensores IoT por medio de servicios de almacenamiento en la nube, para tener así un seguimiento e historial en apoyo a la toma de decisiones dentro del servicio médico de la institución.

2. Arreglo experimental

En la Figura 1 se muestra el arreglo experimental que se propone para la estación de salud. Este arreglo incluye algoritmos que acompañan a un microcontrolador, particularmente el Node MCU ESP 8266, con el cual se facilita la adquisición de los datos del usuario, mismos datos que son enviados a la nube mediante protocolo MQTT y almacenados en una base de datos.

Para la obtención de los datos directamente del estudiante se utilizan sensores, como parte fundamental de un sistema IoT. La lectura de aquéllos se realiza habilitando los puertos de propósito general en el microcontrolador Node MCU ESP8266, que apoyado en los algoritmos correspondientes permite la recepción, captura y almacenamiento de los datos.

Los sensores recolectan los datos, y establecen comunicación con el microcontrolador, el cual los deposita en un servidor local y posteriormente son enviados a un sistema de almacenamiento en la nube (característica esencial de un sistema IoT). Posteriormente estos datos son procesados y tratados para la elaboración del análisis estadístico correspondiente.

Los datos se obtienen por medio de dos sensores de estatura y peso. Los datos aquí presentados provienen de varios grupos estudiantiles de la Preparatoria 2, cuarto semestre de bachillerato turno vespertino. Esta preparatoria pertenece al sistema educativo de la Universidad Autónoma del Estado de Hidalgo, y se encuentra ubicada en Tullancingo, Hidalgo, México.

La estatura se obtiene haciendo una diferencia entre la distancia que existe desde el fondo del módulo que soporta al estudiante y la distancia que existe entre el sensor ultrasónico de distancia, HC-SR04, y la coronilla de la cabeza del estudiante. El peso se adquiere por medio del sensor de carga HX711.

Se ocuparon dos grupos heterogeneos (masculinos y femeninos) del mismo semestre. El servicio médico apoyó a cierta hora de clase para cada grupo y, en una primera sesión, se envió a una primera mitad de un grupo y otro día la segunda mitad del mismo grupo. Se repitió el proceso con el otro grupo. Un grupo diferente adicional compuesto de 112 estudiantes participó en la aportación de datos referentes al Índice de Masa Corporal o IMC.

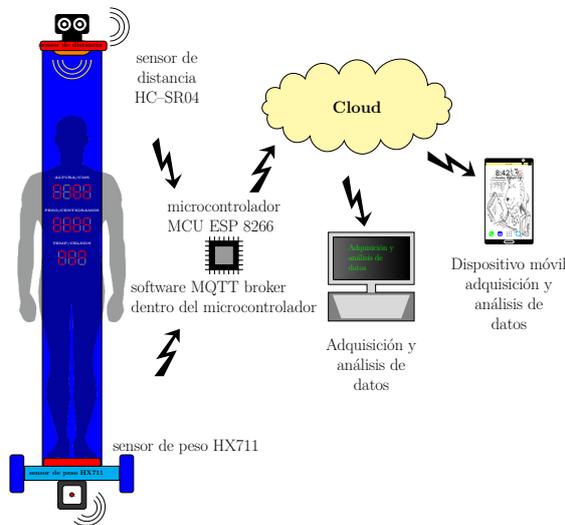


Figura 1: Arreglo experimental del kiosco, se muestran los sensores de distancia y peso, HC-SR04 y HX711 respectivamente, en comunicación con microcontrolador MCU ESP 8266 que envía datos a la nube usando protocolo MQTT.

El servicio médico universitario cuenta con el arreglo experimental mostrado en la Figura 1. Las mediciones se almacenan en forma local, en parte, y otra parte en la versión de prueba en Amazon Web Services, con lo cual se probó también la conectividad a la nube. Dicha conectividad resultó exitosa para las muestras que se guardaron en

la nube. Obviamente, este servicio funciona con restricciones en su versión gratuita.

3. Resultados

La decisión acerca de utilizar estadística paramétrica o no-paramétrica se lleva a cabo realizando pruebas de normalidad tanto para la población de pesos como para la población de estaturas de las alumnas y los alumnos, así como a su derivado denominado Índice de Masa Corporal o IMC.

En ambos casos se consideran grupos heterogeneos de estudiantes, los cuales están compuestos de varones y mujeres. La Tabla 1 muestra los resultados obtenidos al medir los pesos de 115 estudiantes, mientras que la Figura 2 da una descripción estadística de la forma en que se distribuye esta población. En esta descripción, se observa un sesgo hacia la izquierda lo que indica anormalidad de la distribución de la población de pesos. De hecho, la moda tiene el valor 50 y la mediana el valor 57.

Un primer grupo tiene 59 integrantes, mientras que el segundo grupo tiene 56 elementos. El grupo heterogeneo de pesos estudiantiles proporciona valores de media y desviación estándar $\mu = 59.631$ y $\sigma = 12.496$, respectivamente.

Tabla 1: Valores de los pesos, en kilogramos, de un primer grupo heterogeneo de 115 estudiantes. Se fusionaron los resultados de cada grupo experimental introducidos en la Sección 2.

| | | | | |
|---------|--------|---------|--------|--------|
| 85.500 | 45.000 | 100.000 | 43.200 | 49.000 |
| 51.000 | 71.000 | 100.000 | 50.500 | 56.000 |
| 63.400 | 74.700 | 56.500 | 46.800 | 58.000 |
| 44.800 | 56.589 | 60.000 | 65.200 | 46.000 |
| 64.500 | 46.600 | 58.600 | 46.000 | 79.000 |
| 64.000 | 47.000 | 77.300 | 44.500 | 52.000 |
| 100.000 | 77.000 | 90.000 | 79.000 | 52.400 |
| 60.500 | 49.100 | 85.600 | 55.000 | 60.000 |
| 79.000 | 66.800 | 50.000 | 53.600 | 55.000 |
| 67.800 | 63.500 | 50.000 | 51.600 | 70.000 |
| 69.000 | 74.000 | 64.500 | 65.000 | 56.500 |
| 56.700 | 45.300 | 46.000 | 50.000 | 57.000 |
| 45.000 | 69.200 | 47.200 | 51.000 | 64.500 |
| 58.440 | 60.500 | 50.600 | 50.000 | 50.000 |
| 40.000 | 43.500 | 75.600 | 72.400 | 58.900 |
| 63.100 | 47.000 | 61.200 | 70.900 | 54.000 |
| 56.000 | 53.500 | 68.600 | 61.800 | 53.000 |
| 43.700 | 59.500 | 54.600 | 59.000 | 73.500 |
| 62.000 | 55.000 | 48.500 | 42.100 | 64.700 |
| 60.700 | 59.600 | 77.000 | 56.500 | 50.100 |
| 58.000 | 57.000 | 55.000 | 63.100 | 53.400 |
| 56.500 | 64.510 | 45.000 | 75.700 | 56.000 |
| 56.000 | 50.100 | 57.400 | 46.800 | 68.000 |

Un comentario adicional propicio para la explicación de la Figura 2 es que debe tenerse en cuenta que el ancho

de cada barra en el programa Octave, que es uniforme para todas ellas, se debe multiplicar por cada una de las alturas de las barras y que la suma de todos los productos debe ser aproximadamente igual a la unidad, siendo la unidad el área bajo la curva normal mostrada. Esta es la razón de identificar al histograma como normalizado.

Por otro lado, la descripción estadística de la forma en que se distribuye la población de estaturas para el mismo grupo de 115 estudiantes es dada por la Tabla 2 y por la Figura 3. También aquí se aplican comentarios similares a la descripción estadística de los pesos.

Sin embargo, se puede realizar una prueba de normalidad utilizando métodos estadísticos paramétricos formales, como por ejemplo aquellas pruebas que se basan en los conceptos de “kurtosis” y/o “sesgo” o “skew”. Lo cual es bastante útil para automatizar el proceso de análisis de los datos. Sin embargo, antes de profundizar acerca de estas ideas, es conveniente dejar en claro lo que ellas significan.

Para empezar, una distribución normal $\mathcal{N}(x; \mu, \sigma)$ está definida por la función

$$\mathcal{N}(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

donde los parámetros μ y σ se conocen, respectivamente, como media y desviación estándar de la población.

La Figura 2 y la Figura 3 muestran el comportamiento de las respectivas distribuciones normales para las correspondientes poblaciones de pesos y estaturas estudiantiles.

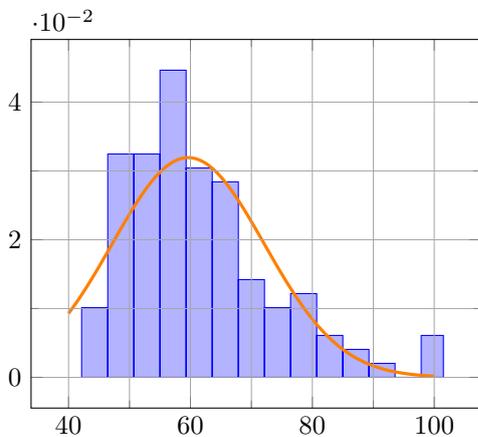


Figura 2: Histograma normalizado de pesos en kilogramos, donde $\mu = 59.631$, $\sigma = 12.496$.

La ecuación para el sesgo, contando con una muestra de tamaño n de la población, es dada como sigue (Doane

and Seward, 2011; Brown, 2022; Joanes and Gill, 1998)¹

$$\frac{n}{(n-1)(n-2)} \sum_{j=1}^n \left(\frac{x_j - \mu}{\sigma} \right)^3$$

y debe observarse que esta expresión es adimensional, y que el factor $\frac{n}{(n-1)(n-2)}$ garantiza un estimador no sesgado (la expresión para el sesgo es una variable aleatoria) (Joanes and Gill, 1998).

El tamaño de la muestra que aquí se trabaja está dentro de los estándares expresados en (Joanes and Gill, 1998).

Cabe mencionar que existe toda una familia de expresiones para calcular el sesgo, y que el integrante de esta familia que se utiliza en este trabajo ha mostrado ser aceptable para tamaños de muestra como las que aquí se presentan, además de que Excel de Office en Windows 10, y Libre Office de Ubuntu 18.04.6 LTS, calculan el sesgo de esta manera (Joanes and Gill, 1998; Brown, 2022).

El sesgo mide qué tan larga es la cola o lado derecho (izquierdo) de la función densidad de probabilidad con respecto a la cola o lado izquierdo (derecho), y si el pico de la misma, supuesta unimodal, se desvía hacia la cola o lado derecho o izquierdo.

Es claro que una distribución normal tiene un comportamiento completamente simétrico, por lo que su sesgo será nulo, aunque en el caso de trabajar con muestras de la población esto no es necesariamente cierto (Cox, 2010), ya que la simetría se puede violar dependiendo si el tamaño de la muestra está definido por un valor par o impar, y además es pequeño.

Los valores negativos del sesgo indican que existe un sesgo hacia la izquierda y valores positivos indican que hay un sesgo hacia la derecha (Aishah et al., 2011).

La técnica que se comenta líneas arriba, para calcular el sesgo, se conoce como “coeficientes β y γ de sesgo” (Trivedi, 2017).

Estos coeficientes fueron definidos por Karl Pearson, basándose en el segundo y tercer momentos centrales. Otro nombre es el de “Coeficiente ajustado de Fisher-Pearson” (NIST, 2022).

Por otro lado, el cuarto momento resulta útil para definir lo que se denomina como “curtosis”². En este sentido, existen al menos dos interpretaciones de lo que la curtosis mide o indica.

1. Mide o indica lo “picudo” de la función densidad de probabilidad, mientras que
2. Mide o indica lo “grueso” de las colas o lados de la función densidad de probabilidad.

Siguiendo de igual manera los criterios de la definición de curtosis, se asume en este trabajo, una curtosis no

¹sample skewness = $G_1 = \frac{\sqrt{n(n-1)}}{n-2} g_1$, donde $g_1 = \frac{m_3}{\sqrt{m_2^3}}$, $m_2 = \frac{n-1}{n} \sigma^2$, $m_3 = \frac{1}{n} \sum_k (x_k - \mu)^3$, por lo que $G_1 = \frac{n}{(n-1)(n-2)} \sum_{j=1}^n \left(\frac{x_j - \mu}{\sigma} \right)^3$.

²curtosis de muestra en exceso = $G_2 = \frac{K_4}{K_2^2} = \frac{\left(\frac{n^2}{(n-1)(n-2)(n-3)} \right) m_2^2 \left((n+1) \frac{m_4}{m_2^2} - 3(n-1) \right)}{\left(\frac{n}{n-1} \right)^2 m_2^2} = \frac{n-1}{(n-2)(n-3)} ((n+1)g_2 + 6)$.

sesgada, que elimina el valor 3 de la curtosis que en una distribución normal se tiene; en otras palabras,

$$\begin{aligned} \text{curtosis de la muestra por exceso} &= G_2 \\ &= \frac{K_4}{K_2^2} \\ &= \frac{n-1}{(n-2)(n-3)} ((n+1)g_2 + 6) \end{aligned}$$

donde se tiene que

$$\begin{aligned} g_2 &= \frac{m_4}{m_2^2} - 3, \\ m_2 &= \frac{1}{n} \sum_{j=1}^n (x_j - \mu)^2, \\ m_4 &= \frac{1}{n} \sum_{j=1}^n (x_j - \mu)^4, \\ K_2 &= \frac{n}{n-1} m_2, \quad \text{y} \\ K_4 &= \frac{n^2}{(n-1)(n-2)(n-3)} ((n+1)m_4 - 3(n-1)m_2^2). \end{aligned}$$

Observe que el segundo momento no se divide por $n-1$, como es usual.

En (Jambu, 1991), se define también el Coeficiente de Pearson $\frac{\mu_4}{\mu_2^2}$ y el Coeficiente de Fisher $\frac{\mu_4}{\mu_2} - 3$, para medir la curtosis.

Bajo las condiciones especificadas, es posible realizar ahora una prueba de normalidad, teniendo en mente que existe toda una familia de ellas (Kim, 2013).

Por su sencillez, en este trabajo se seleccionó la técnica de la Prueba-Z, en donde aparecen los conceptos de sesgo y curtosis en exceso.

Se obtiene un valor-Z al dividir los valores del sesgo y la curtosis en exceso por su correspondiente error estándar, respectivamente. En este caso, las hipótesis nula y alternativa son dadas como sigue,

- H_0 : la población se distribuye en forma normal,
- H_1 : la población no se distribuye en forma normal.

De acuerdo con (Kim, 2013), para tamaños de muestra con valores que satisfacen las desigualdades $50 < n < 300$, se rechaza la hipótesis nula para valores absolutos del valor-Z que se encuentren arriba de 3.29, el cual se asocia con un nivel de significancia α igual al 5%, y se concluye entonces que la muestra no se distribuye de manera normal.

Algunos autores son más exigentes y sugieren un igual nivel de significancia, pero solicitando que el valor absoluto sea mayor que 1.96 o bien 2 (Brownlee, 2020). La prueba-Z se aplica a los datos que proporcionan Tabla 1 y Tabla 2. Sin embargo, la contradicción es que, para aplicar la prueba-Z, se asume que el sesgo y la curtosis de muestras de poblaciones son variables aleatorias que siguen una distribución normal.

Tabla 2: Valores de las estaturas, en metros, de un grupo heterogeneo de estudiantes varones y mujeres. La media de la población es $\mu = 1.6469$ y su desviación estándar es $\sigma = 0.085802$. La moda es igual a 1.56 y la mediana toma el valor 1.65.

| | | | | |
|--------|--------|--------|--------|--------|
| 1.5000 | 1.5800 | 1.6500 | 1.7400 | 1.5200 |
| 1.5800 | 1.6500 | 1.7400 | 1.5300 | 1.5900 |
| 1.6500 | 1.7500 | 1.5300 | 1.6000 | 1.6600 |
| 1.7600 | 1.5400 | 1.6000 | 1.6600 | 1.7700 |
| 1.5600 | 1.6000 | 1.6600 | 1.7700 | 1.5600 |
| 1.6000 | 1.6700 | 1.7800 | 1.5600 | 1.6000 |
| 1.6800 | 1.8000 | 1.5600 | 1.6100 | 1.7000 |
| 1.8000 | 1.5600 | 1.6200 | 1.7000 | 1.8000 |
| 1.5600 | 1.6300 | 1.7000 | 1.8000 | 1.5700 |
| 1.6300 | 1.7000 | 1.8000 | 1.5700 | 1.6400 |
| 1.7000 | 1.8200 | 1.5700 | 1.6400 | 1.7000 |
| 1.8600 | 1.5800 | 1.6500 | 1.7300 | 1.4500 |
| 1.5900 | 1.6600 | 1.7100 | 1.4800 | 1.6000 |
| 1.6600 | 1.7100 | 1.4800 | 1.6000 | 1.6700 |
| 1.7100 | 1.5000 | 1.6000 | 1.6700 | 1.7200 |
| 1.5000 | 1.6100 | 1.6700 | 1.7400 | 1.5000 |
| 1.6100 | 1.6700 | 1.7400 | 1.5500 | 1.6100 |
| 1.6800 | 1.7400 | 1.5600 | 1.6200 | 1.6800 |
| 1.7400 | 1.5600 | 1.6200 | 1.6800 | 1.7600 |
| 1.5600 | 1.6200 | 1.6800 | 1.8000 | 1.5700 |
| 1.6400 | 1.6900 | 1.8000 | 1.5800 | 1.6500 |
| 1.6900 | 1.5800 | 1.6500 | 1.7000 | 1.5800 |
| 1.6500 | 1.7000 | 1.5900 | 1.6600 | 1.7100 |

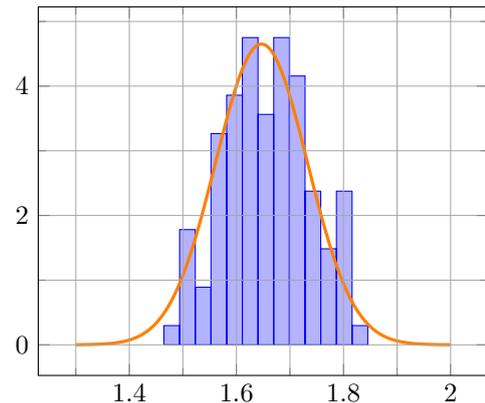


Figura 3: Histograma normalizado de alturas en metros, donde $\mu = 1.6469$, $\sigma = 0.085802$.

Para ambos estadísticos, un análisis riguroso muestra que, efectivamente, ellos siguen una distribución normal (tomando en cuenta el Teorema del Límite Central) (Hain, 2010). Por lo tanto, aplicar la prueba-Z usando sesgos y curtosis tiene bastante sentido.

Otra prueba importante es la que se refiere a la homogeneidad de la varianza para dos o más grupos experimentales. Conocer este hecho tiene mucha importancia en el estudio de la asociación entre salud y educación con cuestiones antropométricas que consideran al Índice de Masa Corporal y la estatura (Kim et al., 2020).

Así, la Prueba de Levene se usa para determinar si dos o más grupos tienen iguales varianzas. Esta es una prueba ampliamente usada en Estadística debido a que muchas pruebas estadísticas suponen que los grupos tienen varianzas iguales.

Por ejemplo, se desea saber si dos grupos diferentes de estudiantes conducen a diferentes niveles de IMCs, por lo que se seleccionan 112 estudiantes al azar y se dividen en dos grupos de 56, tomando en cuenta que existen varones y mujeres en ambos grupos, con diferentes pesos y estaturas y diferentes hábitos alimenticios.

Antes de conducir una prueba estadística para determinar si existe una diferencia en el IMC entre los dos grupos, se lleva a cabo una prueba de Levene para determinar si los dos grupos tienen o no iguales varianzas. Si el p -valor del ANOVA es menor que algún nivel de significancia igual a 0.05, entonces esto indica que los dos grupos no tienen iguales varianzas (Mara and Cribbie, 2018).

En (Mara and Cribbie, 2018) se da una explicación de la equivalencia entre la Prueba de Levene y la ANOVA F test, por lo que después de aplicar la ANOVA F Test a los datos de IMC que muestra parcialmente la Tabla 3, se encuentra que el p -valor del ANOVA unilateral es igual a 0.0000514211.

Debido a que este valor es menor que 0.05, entonces se rechaza la hipótesis nula; es decir, los dos grupos no tienen iguales varianzas. En otras palabras, no se tiene la suficiente evidencia de que las varianzas entre los dos grupos sean iguales.

Obviamente, antes de aplicar la prueba de Levene se debió realizar una prueba estadística de normalidad de la distribución de la población del IMC. La Tabla 4 muestra los resultados.

Por consiguiente, se concluye que la distribución del IMC no es normal y para cualquier prueba de hipótesis relacionada con el IMC se debe seleccionar un técnica de inferencia estadística no paramétrica, razón por la cual se ha seleccionado arriba la prueba de Levene para verificar si se tiene homogeneidad de varianzas entre los dos grupos de IMCs.

El valor del sesgo es positivo, lo que significa que existe una asimetría positiva en la que el valor de la media es más grande que la mediana y ésta a su vez es más grande que el de la moda. Los valores coinciden con la descripción dada por Figura 4.

Tabla 3: Algunos valores del índice de masa corporal (IMC), en Kgs/mts^2 , de un grupo heterogeneo de estudiantes varones y mujeres. La media de la población es $\mu = 21.718$ y su desviación estándar es $\sigma = 2.2675$, mientras que su moda y mediana adquieren los valores 18.902 y 21.218, respectivamente. Por otro lado, el sesgo adquiere el valor 1.60416 (calculado con Excel SKEW()) y la kurtosis el valor 3.50 (calculada con Excel KURT()).

| | | | | |
|-------|-------|-------|-------|-------|
| 18.90 | 18.90 | 18.90 | 19.02 | 19.10 |
| 19.13 | 19.14 | 19.20 | 19.22 | 19.22 |
| 19.23 | 19.26 | 19.31 | 19.31 | 19.39 |
| 19.42 | 19.48 | 19.86 | 19.88 | 19.92 |
| 19.93 | 20.00 | 20.00 | 20.01 | 20.03 |
| 20.03 | 20.03 | 20.07 | 20.07 | 20.16 |
| 20.17 | 20.17 | 20.23 | 20.28 | 20.28 |
| 20.31 | 20.47 | 20.68 | 20.70 | 20.75 |
| 20.75 | 20.82 | 20.83 | 20.83 | 20.86 |
| 20.90 | 20.94 | 20.94 | 20.95 | 20.96 |
| 21.04 | 21.05 | 21.06 | 21.08 | 21.12 |
| 21.12 | 21.15 | 21.22 | 21.22 | 21.26 |
| 21.26 | 21.30 | 21.33 | 21.34 | 21.34 |
| 21.34 | 21.43 | 21.48 | 21.51 | 21.52 |
| 21.64 | 21.68 | 21.71 | 21.83 | 21.83 |
| 21.95 | 22.04 | 22.04 | 22.06 | 22.06 |
| 22.13 | 22.32 | 22.49 | 22.56 | 22.86 |

Tabla 4: Prueba de normalidad para el IMC. Note que la población de IMCs no se distribuye de manera normal, por lo que la homogeneidad de las varianzas debe usar estadística no paramétrica.

| | |
|------------------------------|---|
| n | |
| Sesgo | 1.6041609875 |
| Curtosis | 3.50021547933 |
| SE Sesgo | 0.22841609629 |
| SE Curtosis | 0.45683219258 |
| Resultado basado en sesgo | Distribución no es simétrica debido a que el sesgo $1.60416098750 > 2 \times 0.228416096288064$ |
| Resultado basado en curtosis | Distribución no es simétrica debido a que curtosis $3.50021547933 > 2 \times 0.45683219258$ |

Lo mismo ocurre con el valor de la curtosis, el cual indica una distribución platocúrtica (curtosis mayor que 3), lo que reafirma el hecho de que la distribución de la población IMC no es normal.

La función densidad de probabilidad, o PDF, es definida por medio de la Ecuación (1), siguiendo las sugerencias de (Ashour and Abdel-hameed, 2010; Harvey et al., 2022;

Mudholkar and Hutson, 2000),

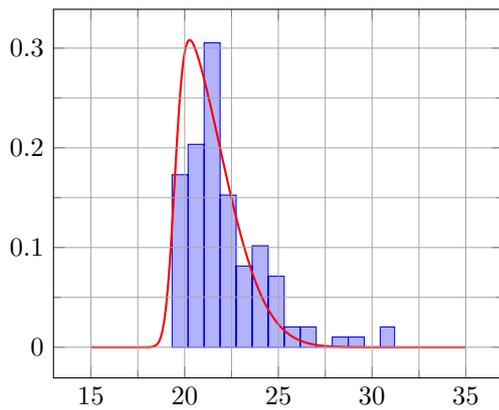


Figura 4: Histograma normalizado del IMC en Kgs/mts^2 , donde $\mu_s = 19.468$, $\sigma_s = 2.3925$. La población IMC original tiene $\mu = 21.718$, $\sigma = 2.2675$, mientras que su moda y mediana son 18.902 y 21.218, respectivamente.

$$f_X(x; \mu_s, \sigma_s, \alpha) = 2\phi_{Y(X)}(y; \mu_s, \sigma_s)\Phi_{Y(X)}(\alpha y; \mu_s, \sigma_s), \tag{1}$$

donde

$$\phi_{Y(X)}(y; \mu_s, \sigma_s) = \frac{1}{\sigma_s} f_{Y(X)}(y, 0, 1),$$

siendo $Y(X) = \frac{X - \mu_s}{\sigma_s}$ y

$$\Phi_{Y(X)}(\alpha y; \mu_s, \sigma_s) = \frac{1}{2} \left(1 + erf_{Y(X)} \left(\frac{\alpha y}{\sqrt{2}} \right) \right),$$

en la que $f_{Y(X)}$ es la función densidad de probabilidad Gaussiana, con variable aleatoria X escalada y desplazada en la forma $Y = \frac{X - \mu_s}{\sigma_s}$, para que esta nueva variable aleatoria tenga media cero y varianza igual a la unidad. La función erf es la función de error ya conocida en la literatura físico-matemática (Arfken et al., 2013).

Los valores para μ_s y σ_s , que se indican en la Figura 4, son los que se toman para la función de error incluida en la Ecuación (1), en la que también $\alpha = 6$. Nótese que la media para la función de error se parece más a la moda de la población IMC original.

La función densidad de probabilidad se encuentra normalizada, lo cual se puede corroborar numéricamente usando sencillamente el comando “trapz” de Octave; es decir, la integral sobre toda la población es igual a la unidad.

La importancia de conocer la estructura de la función densidad de probabilidad de la población de valores de índice de masa corporal, y los parámetros participantes en ella, radica en el hecho de que, a partir de ahí, se pueden aplicar con alta confianza técnicas estadísticas que vayan más allá de solamente observar los valores de las medias

de muestras, y de solamente saber hasta dónde se extiende la epidemia de obesidad (Ng et al., 2016).

A partir de la función densidad de probabilidad de la población, se pueden conocer diferentes características del comportamiento de esta población, y viceversa los valores de diferentes características influyen en la estructura y parámetros de la función densidad de probabilidad.

La interpretación del comportamiento en el tiempo de la distribución de la población recibe una fuerte influencia de la forma en que se definen los valores de corte de IMC para las categorías denominadas “Desnutrición severa”, “Desnutrición”, “Peso normal”, “Sobrepeso” y “Obesidad”. La Tabla 5 especifica esta clasificación de acuerdo a los estándares establecidos por la Organización Mundial de la Salud (Weir and Jan, 2021).

Tabla 5: La Organización Mundial de la Salud emplea el IMC para definir la existencia o no existencia de obesidad u otras enfermedades relacionadas con el peso de un individuo.

| Enfermedad | Subclases | Intervalo de valores del IMC Kg/m ² | Población |
|---------------------|-----------|--|-----------------------------|
| Desnutrición severa | | IMC < 16.5 | |
| Desnutrición | | IMC < 18.5 | |
| Peso normal | | 18.5 ≤ IMC ≤ 24.9 | |
| Sobrepeso | | 25 ≤ IMC ≤ 29.9 | |
| | | 23 ≤ IMC ≤ 24.9 | Población de Asia y Sudasia |
| Obesidad | clase I | 30 ≤ IMC ≤ 34.9 | |
| | clase II | 35 ≤ IMC ≤ 39.9 | |
| | clase III | 40 ≤ IMC (también conocida como severa, extrema, o masiva) | |
| | | 25 < IMC | Población de Asia y Sudasia |

Un estimado de la población estudiantil actual en la Preparatoria 2 perteneciente a la Universidad Autónoma del Estado de Hidalgo, y ubicada en la Ciudad de Tulancingo, Hidalgo, México, es igual al número 3,374, conociéndose datos acerca de la condición de solamente 849 estudiantes.

Al emplear la función densidad de probabilidad dada por la Ecuación (1) para determinar los porcentajes de población con peso normal, con sobrepeso y obesidad, se obtienen respectivamente los resultados 97.9%, 2.1% y 0%. Para desnutrición, el porcentaje es igual a 0.0003%.

Estas cifras indican que aproximadamente 3,303 integrantes de la población de 3,374 estudiantes tienen un peso

normal, mientras que aproximadamente 71 de los 3,374 estudiantes tienen sobrepeso, siendo esta una cantidad digna de tomar en cuenta. Por otro lado, aproximadamente cero estudiantes tienen obesidad.

El caso curioso es que aproximadamente un estudiante padezca desnutrición, pero esta es una predicción del modelo que se está utilizando. Por supuesto, se puede emplear la misma función densidad de probabilidad para determinar los intervalos de confianza para ciertos valores de significancia α' .

Datos proporcionados por el Servicio Médico Universitario, sobre la muestra de 849 estudiantes que el Servicio Médico mismo obtuvo para estudiantes de la Preparatoria 2, se encuentran dados por la Tabla 6.

La Tabla 6, se basa en los valores que proporciona la Tabla 5, aunque el servicio médico considera que el valor del IMC para la desnutrición es menor a 18.9, lo que aumenta a 13 el número de estudiantes desnutridos según el modelo de distribución empleado. Sin embargo, las comparaciones indican que hay mucho trabajo por hacer en lo que corresponde al modelado.

Tabla 6: Datos proporcionados por el Servicio Médico Universitario de la Universidad Autónoma del Estado de Hidalgo, obtenidos de una muestra de 849 estudiantes pertenecientes a la Preparatoria 2, enero–junio 2022.

| Condición | Número de alumnos | Porcentajes (%) |
|--------------------------|-------------------|-----------------|
| Peso bajo (desnutrición) | 132 | 15.55 |
| Peso normal | 551 | 64.90 |
| Sobrepeso | 139 | 16.37 |
| Obesidad leve | 27 | 3.18 |
| Obesidad media | 0 | 0.00 |
| Obesidad mórbida | 0 | 0.00 |
| Total | 849 | 100 |

No solamente termina aquí la importancia de conocer la función densidad de probabilidad ya que, con la aplicación del Teorema de Bayes, la distribución resulta de suma utilidad para realizar inferencias, actuando como una distribución *prior* que ayude a conocer distribuciones condicionales y también marginales.

En este sentido, existen estudios que relacionan cambios en IMC con cambios en patrones de alimentación, cambios en actividades físicas, cambios en algunas componentes sociales, cambios en algunas componentes psicológicas, etc., obteniéndose de esta forma una expresión para la distribución $P(IMC \leq b|C)$, donde C representa uno o más de los cambios de comportamiento personal, social, etc.

A partir de aquí, no es difícil darse cuenta que lo que interesa es $P(C|IMC \leq b)$ y que esta distribución se puede obtener conociendo la distribución $P(IMC \leq b|C)$, además de $P(IMC \leq b)$ y $P(C)$. Claramente, C puede incluir el ausentismo, entre otras cuestiones puramente escolares, pero estos tópicos están más allá del alcance de este reporte de avance de proyecto.

4. Conclusiones y trabajo futuro

La combinación de técnicas de Internet de las Cosas, en las que se hace uso de diversos dispositivos para adquisición de datos, junto con el análisis estadístico de los mismos, resulta en una tecnología sumamente interesante. Esta técnica puede ser aplicada por ejemplo en la solución de problemas relacionados con la ausencia escolar y las enfermedades de los estudiantes.

En particular, un sistema electrónico con el cual se puedan obtener datos relacionados con estatura y peso, a partir de los cuales se pueda determinar el comportamiento de la variable aleatoria denominada Índice de Masa Corporal, resulta de utilidad para el estudio del sobrepeso de las y los estudiantes.

La construcción de dicho sistema electrónico, que consiste de diversos sensores y dispositivos especiales (cómputo y comunicación), se encuentra en una etapa avanzada y la Figura 1 ilustra la forma en que interactúan estas componentes.

En una primera fase de prueba, se han obtenido diversos datos relacionados con el IMC, y las técnicas estadísticas comentadas en este trabajo son sólo algunas de las que se pueden implementar para llevar a cabo un análisis de la confiabilidad de los datos, y demostrar algunas hipótesis estadísticas con técnicas paramétricas y/o no paramétricas.

Las demostraciones de estas hipótesis permiten trabajar de manera confiable con las funciones de distribución obtenidas, y con las cuales, en una siguiente etapa de este proyecto, se podrán establecer algunas conclusiones asociadas al interés principal de conocer los efectos que tiene el IMC sobre patrones de conductas escolares (Sharifian et al., 2017).

En términos generales, el trabajo futuro debe considerar el análisis de las diferentes relaciones que pueden tener variables tales como edad, género, estado socio-económico, etc., con diferentes indicadores del estado de salud de las y los estudiantes, como por ejemplo ansiedad, obesidad, etc. (Yu et al., 2018).

Referencias

Aishah, N., Yin, T. S., Othman, A. R., and Yaacob, C. R. (2011). Sensitivity of Normality Tests to Non-normal Data (Kepekaan Ujian Kenormalan Terhadap Data Tidak Normal). *Sains Malaysiana*, 40(6):637–641.

Ali, S. M. and Nihad, M. (May 2021). Internet of Things for Education Field. *Journal of Physics: Conference Series*, 1897(1):1–9.

Amaraweera, S. P. and Halgamuge, M. N. (2019). Internet of Things in the Healthcare Sector: Overview of Security and Privacy Issues. In Mahmood, Z., editor, *Security, Privacy and Trust in the IoT Environment*, chapter 8, pages 153 – 179. Springer.

Arfken, G. B., Weber, H. J., and Harris, F. E. (2013). Chapter 9 Partial Differential Equations. In *Mathematical Methods for Physicists: A Comprehensive Guide, Seventh Edition*, pages 401–445. Elsevier.

Ashour, S. K. and Abdel-hameed, M. A. (2010). Approximate skew normal distribution. *Journal of Advanced Research*, 1:341–350.

Brown, S. (2008 – 2022). Measures of shape: Skewness and kurtosis. <https://brownmath.com/stat/shape.htm>.

- Brownlee, J. (2020). Continuous Probability Distributions. In *Probability for Machine Learning: Discover How To Harness Uncertainty with Python*, chapter 9, pages 60–72. Machine Learning Mastery Books.
- Cox, N. J. (2010). Speaking Stata: The limits of sample skewness and kurtosis. *The Stata Journal*, 10(3):482–495.
- Doane, D. P. and Seward, L. E. (2011). Measuring skewness: A forgotten statistic? *Journal of Statistics Education*, 19(2):18 pages.
- Hain, J. (2010). *Chapter 3 Moment Tests for Normality*. PhD thesis, In Comparison of Common Tests for Normality, Thesis at Julius-Maximilians-Universität Würzburg, Institut für Mathematik und Informatik, Lehrstuhl für Mathematik VIII (Statistik).
- Harvey, P. S., Gavin, H. P., and Scruggs, J. T. (2022). Probability Distributions. Lecture Notes, CEE 201L, Uncertainty, Design, and Optimization, Department of Civil and Environmental Engineering, Duke University <https://people.duke.edu/~hpgavin/cee201/ProbabilityDistributions.pdf>.
- Jambu, M. (1991). 1-D Statistical Data Analysis. In *Exploratory and Multivariate Data Analysis*, chapter 3, pages 27–62. Academic Press.
- Joanes, D. N. and Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 47(1):183–189, 7 pages.
- Juárez, E. A. O., Ávila Wall, L., and Acosta, L. R. S. (2014). Causas de Inasistencia a Clases de los Estudiantes—Caso: Estudiantes de la FCA de la UACH. In *Memorias en extenso XV Congreso Asociación de Profesores de Contaduría y Administración de México (APCAM)*, page 19. Asociación de Profesores de Contaduría y Administración de México. <http://www.fca.uach.mx/apcam/2014/04/05/Ponencia%20111-UACH.pdf>.
- Kim, H.-Y. (February 26, 2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, 38(1):52–54. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3591587/>.
- Kim, R., Pathak, P. K., Tripathi, N., and Subramanian, S. V. (2020). Heterogeneity in adult anthropometry by socioeconomic factors: Indian National Family Health Survey 2006 and 2016. *European Journal of Clinical Nutrition*, 74:953 – 960.
- Mara, C. and Cribbie, R. A. (2018). Equivalence of Population Variances: Synchronizing the Objective and Analysis. *The Journal of Experimental Education*, 86(3):442 – 457.
- Mudholkar, G. S. and Hutson, A. D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference*, 83:291–309.
- Ng, M., Liu, P., Thomson, B., and Murray, C. J. L. (2016). A novel method for estimating distributions of body mass index. *Population Health Metrics*, 14(6):1–7.
- NIST (2022). 1. Exploratory Data Analysis/1.3. EDA Techniques/1.3.5. Quantitative Techniques/1.3.5.11. Measures of Skewness and Kurtosis. In *Engineering Statistics Handbook, NIST/SEMATECH e-Handbook of Statistical Methods*, chapter 1. <http://www.itl.nist.gov/div898/handbook/>.
- Rayan, R. A., Tsagkaris, C., and Iryna, R. B. (2021). The Internet of Things for Healthcare: Applications, Selected Cases and Challenges. In *IoT in Healthcare and Ambient Assisted Living*, pages 1–15. Springer.
- Sharifian, S. A., Aminian, O., Kolahi, L., Zadeh, S. A. H. M., and Mohseni, H. (2017). The Association between Body Mass Index and Absenteeism. *Journal of Advances in Medicine and Medical Research (Past name: British Journal of Medicine and Medical Research)*, 24(11):1–8.
- Trivedi, M. (2017). Skewness and Kurtosis. Post-Graduate Diploma in Applied Statistics (PGDAST) community home page, MST – 002 Descriptive Statistics. In *Block-1 Analysis of Quantitative Data*, chapter 4. eGyankosh— a National Digital Repository, India, Indira Gandhi National Open University. <https://egyankosh.ac.in/bitstream/123456789/20444/1/Unit-4.pdf>.
- Weir, C. B. and Jan, A. (2021). BMI Classification Percentile And Cut Off Points. National Library of Medicine, National Center for Biotechnology Information, StatPearls Publishing, <https://people.duke.edu/~hpgavin/cee201/ProbabilityDistributions.pdf>.
- Yu, K., Liu, X., Alhamzawi, R., Becker, F., and Lord, J. (2018). Statistical methods for body mass index: A selective review. *Stat Methods Med Res.*, 27(3):798–811.