

63

Fecha de presentación: octubre, 2021

Fecha de aceptación: diciembre, 2021

Fecha de publicación: febrero, 2022

ANÁLISIS ESTADÍSTICO

CUANDO NO SE CUMPLEN LOS SUPUESTOS DE LAS PRUEBAS PARAMÉTRICAS, EN EL CONTEXTO DE LA INVESTIGACIÓN DE LA CULTURA FÍSICA

STATISTICAL ANALYSIS WHEN THE ASSUMPTIONS OF PARAMETRIC TESTS ARE NOT MET, IN THE CONTEXT OF PHYSICAL CULTURE RESEARCH

Yury Douglas Barrios Palacios¹

E-mail: yurybarriosp@ug.edu.ec

ORCID: <https://orcid.org/0000-0001-6759-8076>

Zoila Eugenia Guerrero Ávila¹

E-mail: zcoila.guerreroa@ug.edu.ec

ORCID: <https://orcid.org/0000-0002-5611-2496>

Dolores Francisca Zambrano Miranda¹

E-mail: dolores.zambranom@ug.edu.ec

ORCID: <https://orcid.org/0000-0002-5276-3611>

Henry Xavier Ponce Solórzano¹

E-mail: henry.ponces@ug.edu.ec

ORCID: <https://orcid.org/0000-0003-2029-8601>

¹ Universidad de Guayaquil. Ecuador.

Cita sugerida (APA, séptima edición)

Cabrera Albert, J. S., Fariñas León, G., Hernández Becerra, B., & Navarro Guzmán, J. (2022). Análisis estadístico cuando no se cumplen los supuestos de las pruebas paramétricas, en el contexto de la investigación de la Cultura Física. *Revista Universidad y Sociedad*, 14(S1), 591-600.

RESUMEN

La verificación de los supuestos de las pruebas paramétricas constituye uno de los pasos fundamentales para desarrollar un análisis estadístico acertado. El objetivo de la presente investigación es ilustrar, mediante un estudio real, las implicaciones y formas de tratar el incumplimiento de los supuestos de normalidad y homogeneidad de las pruebas paramétricas t de Student y el análisis de varianza, en el contexto de las investigaciones de la Cultura Física. Para esto se realizó un análisis en ejercicios de salto con contramovimiento. Participaron 116 estudiantes de la carrera de Pedagogía de la actividad física y deporte de la universidad de Guayaquil. Las variables altura del salto y RSI seleccionadas para las pruebas ANOVA y t, respectivamente, verificaron incumplimiento del supuesto de normalidad. Las transformaciones matemáticas no fueron suficientes para corregir el incumplimiento de la primera, aunque el RSI fue normalizado mediante una transformación logarítmica.

Palabras clave: Prueba paramétrica, normalidad, homogeneidad de varianzas, transformación logarítmica.

ABSTRACT

The verification of the assumptions of parametric tests constitutes one of the fundamental steps to develop a successful statistical analysis. The objective of the present research is to illustrate, by means of a real study, the implications and ways of dealing with the non-compliance of the assumptions of normality and homogeneity of the parametric Student's t-tests and the analysis of variance, in the context of Physical Culture research. For this purpose, an analysis was carried out in jumping exercises with countermovement. A total of 116 students of the Pedagogy of Physical Activity and Sport career of the University of Guayaquil participated in the study. The variables jump height and RSI selected for the ANOVA and t-tests, respectively, verified non-compliance with the normality assumption. The mathematical transformations were not sufficient to correct the non-compliance of the former, although the RSI was normalized by means of a logarithmic transformation.

Keywords: Parametric test, normality, homogeneity of variances, log transformation.

INTRODUCCIÓN

La estadística se ha convertido hoy en día en un campo de conocimiento presente en diversos aspectos de la vida cotidiana. También se ha convertido en una herramienta de investigación básica para muchos campos y disciplinas científicas (Borovkov & Moullagaliev, 2019). Como elemento del cálculo de probabilidades, se utiliza con éxito en ciencias puras como la física y la química. Es indispensable, en las ciencias técnicas (p.e, control de calidad, resistencia de los materiales), en las ciencias económicas y en las ciencias naturales.

Según lo planteado por Salas (2018), "En la actualidad, los métodos estadísticos se hacen indispensables como un procedimiento eficaz para la interpretación exacta de información de todas las ciencias y sirve como una herramienta de enlace para cada información" (p. 6).

Permiten, además, resumir y presentar fácilmente los resultados de los estudios empíricos, lo que hace que sea utilizado a menudo también por los investigadores que se ocupan de la cultura física y el deporte (Abidovoch, 2020).

Muchos trabajos que abordan cuestiones en el ámbito de la cultura física se refieren a la estructura y la composición del cuerpo humano, a la fuerza muscular de determinados grupos de músculos, así como a las capacidades motoras (Di Domenico & Raiola, 2021; McBurnie et al., 2021; Pérez-Martínez et al., 2020). Más específicamente, según plantean Sánchez-Sixto et al., (2019), en los estudios relacionados con los saltos, la determinación de los parámetros que tiene mayor influencia en la diferencia de rendimiento de salto o incluso entre los diferentes tipos o clases de salto puede ayudar a los investigadores y entrenadores a comprender el papel de cada parámetro en el rendimiento del salto vertical. En cualquiera de estos casos, los resultados de las mediciones obtenidas requieren análisis estadísticos e interpretaciones adecuadas, como una de las condiciones para dar validez científica a la investigación.

Una amplia gama de programas informáticos estadísticos disponibles permite realizar análisis a menudo muy complejos de forma muy sencilla, casi intuitiva y rápida. Sin embargo, el mero conocimiento de cómo utilizar un programa estadístico no es suficiente para comprender plenamente los análisis realizados, así como para interpretar correctamente los resultados. El utilizar pruebas estadísticas paramétricas conocidas sin comprobar sus supuestos relativos a la distribución normal y la homogeneidad de la varianza de las características probadas, constituye un error habitual dentro de la comunidad científica. Este comportamiento suele contribuir a la publicación de

resultados erróneos o sesgados y la afirmación de conclusiones equivocadas sobre el problema investigado y las hipótesis planteadas.

Cuando las suposiciones sobre la distribución normal y la homogeneidad (homogeneidad) de la varianza no se cumplen, y entonces no se tienen en cuenta en los cálculos estadísticos posteriores, el diseño original de la prueba se ve alterado, y como consecuencia el resultado puede estar muy sesgado (Sánchez, 2015). Si no se cumplen los supuestos de la prueba, hay que intentar modificar los datos (por ejemplo, rechazando los valores extremos), cambiar el modelo (por ejemplo, excluyendo uno de los varios grupos de las comparaciones) o introducir pequeñas transformaciones de los datos (raíz cuadrada, inversa, etc.) que no alteren el resultado de las comparaciones.

La prueba t de Student y el análisis de varianza ANOVA son pruebas paramétricas, lo que significa que su uso supone que los datos analizados cumplen una serie de supuestos (Moder, 2016). Esto se debe a que las pruebas paramétricas se ocupan de los valores estadísticos de la población, como la media, a diferencia de las pruebas no paramétricas, que se ocupan de la forma funcional de la distribución. Los supuestos más básicos son la existencia de una distribución normal de la variable de prueba en cada uno de los grupos considerados y la homogeneidad de la varianza de estas distribuciones. Sin embargo, hay que recordar que las pruebas paramétricas también tienen otros supuestos (Palacio et al., 2020), que a menudo se omiten o se dan por cumplidos, a menudo por desconocimiento del investigador.

Un supuesto muy importante es que las pruebas paramétricas sólo pueden utilizarse con datos medidos en una escala de intervalo o cociente, es decir, que nos permite operar con datos como la suma o la división. Es inaceptable comparar entre sí con las pruebas mencionadas grupos de naturaleza ordinal (p. e, la medición de la satisfacción del cliente con el servicio realizado, clasificado del 1 al 10) o nominal, asignando objetos a grupos particulares sin relaciones de jerarquía entre las categorías, como el color de los ojos o el género. Un supuesto tan importante como la mensurabilidad de los datos es la aleatoriedad de la muestra, es decir, el llamado principio de aleatorización (Cortés, 2018).

Hay que recordar que una muestra bien elegida debe representar un cierto reflejo de la población que se requiere examinar, por ejemplo, al indagar sobre la condición física de los estudiantes de cierta carrera, no se puede seleccionar una muestra formada sólo por alumnos de un único semestre, que practican deporte. Los resultados

de este análisis estadístico podrían ser significativamente diferentes de la aptitud física real del polaco medio. Un grupo bien seleccionado garantiza el cumplimiento de otro supuesto, que es la independencia de las mediciones (Luna, 2019), lo que significa que el conocimiento de una medición no debe dar ninguna indicación sobre los valores de las demás mediciones.

La normalidad de la distribución y la homogeneidad de la varianza son supuestos clave necesarios para realizar una prueba t-Student o un ANOVA. En otras palabras, estas pruebas comparan los resultados de las pruebas obtenidas en grupos individuales siempre que tengan distribuciones normales con una dispersión similar de los datos. Cuando se miden los parámetros de la composición corporal y la fuerza muscular de los grupos musculares seleccionados, así como las capacidades motrices de una muestra pequeña, puede ocurrir que no se cumplan estos supuestos. Si este es el caso, no significa necesariamente que se abandone el uso de una prueba paramétrica. Un método sencillo para estabilizar la varianza y normalizar los datos es transformarlos (Teijón-Alcalá & Serrano-Maillo, 2019). La más popular es la transformación logarítmica, que permite normalizar los datos de una variable mediante la logaritmación de los datos originales. Otras transformaciones alternativas son la transformación de raíz, la transformación inversa o la transformación cuadrática.

Si todos los intentos de normalizar la distribución o estabilizar la varianza fracasan, no significa que es necesario renunciar por completo a comparar las muestras. Cuando en la práctica no se conoce la distribución de la población de la que procede la muestra de datos el investigador debe conformarse con la información contenida en las observaciones. Un histograma puede dar una pista sobre la forma de la distribución. Si parece muy sesgada (si hay uno o más valores extremos), o si son tan pocas observaciones que es difícil decir algo sobre la forma de la distribución. Entonces cobran especial relevancia las pruebas no paramétricas, como la prueba de Mann-Whitney, cuando se comparan dos grupos independientes, la prueba de Kruskal-Wallis -como equivalente al análisis de varianza de una vía- y la prueba de orden de rango de Wilcoxon, para dos muestras dependientes (Jiménez-Miranda, 2021).

Por otra parte, cuando necesita comparar dos muestras, un método no paramétrico requiere que las dos distribuciones a comparar tengan la misma dispersión y forma. Este supuesto también se aplica a una prueba t cuando se compara la media de dos grupos no relacionados, pero la prueba t es algo menos vulnerable a la dispersión

diferente que una prueba no paramétrica, al menos cuando las dos muestras son aproximadamente del mismo tamaño.

La prueba t tradicional tolera relativamente bien desviaciones del orden de hasta dos o tres veces mayor desviación estándar en un grupo que en el otro. Muchos paquetes estadísticos han implementado una variante de la prueba t (prueba Welch) que tiene en cuenta que la dispersión puede ser diferente en los dos grupos. Por lo tanto, podemos tratar con relativa facilidad los incumplimientos de la hipótesis de dispersión igual cuando se utilizan pruebas paramétricas.

Además, la prueba de Kruskal-Wallis, al igual que el análisis de varianza de una vía, permite realizar pruebas post-hoc (p. e, la prueba de Dunn) cuando se rechaza la hipótesis nula de igualdad de medias en los grupos considerados. Estas pruebas generalmente no son demasiado sensibles a suposiciones iniciales y pueden utilizarse para datos ordinales y cualitativos, así como para grupos muy pequeños de datos continuos, para los que las pruebas de normalidad no resultan concluyentes. Debido a que la potencia de las pruebas no paramétricas es menor que la de las paramétricas, sólo deberían utilizarse cuando no sea posible aplicar una prueba paramétrica.

Sin embargo, autores como Glantz et al., (2016), advierten que el ANOVA es robusto frente a las violaciones del supuesto de homogeneidad de varianza. Sin embargo, especialmente cuando los tamaños de las muestras son desiguales, las grandes diferencias de varianza entre los grupos de tratamiento pueden conducir a estadísticas que se asocian con valores P que sobrestiman o subestiman el riesgo real de una conclusión falsa positiva (error de tipo I).

Una recomendación común, cuando se registran varianzas desiguales en ANOVA, es utilizar un estadístico no paramétrico, como el de Kruskal-Wallis, sin tener en consideración algo verdaderamente relevante y que muchas veces es ignorado por los investigadores. Esto es el hecho de que, aunque los estadísticos no paramétricos son libres de distribución, porque no asumen una distribución normal, erróneamente se suele pensar que no suponen varianzas iguales.

Sin embargo, las pruebas estadísticas no paramétricas suponen que las distribuciones de las distintas muestras y submuestras tienen la misma forma y, por lo tanto, asumen que hay una dispersión igual o, al menos, en sus distribuciones. Por lo tanto, las estadísticas no paramétricas, o basadas en rangos, pueden no ser una alternativa adecuada al ANOVA cuando las varianzas no resultan ser iguales (Glantz et al., 2016). Por tanto, resulta necesario

verificar la homogeneidad de varianzas antes de optar por una no paramétrica alternativa a ANOVA.

El objetivo de este artículo es debatir los problemas asociados al incumplimiento de los supuestos de las pruebas estadísticas que se suelen utilizar en la investigación biométrica en el deporte, a saber, la prueba t de Student y el ANOVA unidireccional.

Sobre la base de las propias investigaciones de los autores, se han realizado análisis estadísticos ejemplares junto con una discusión de las dificultades que surgen durante los análisis. Se ha prestado especial atención a la validez de la comprobación de los supuestos de ambas pruebas y a la indicación de los métodos que deben utilizarse en caso de que no se cumplan.

MATERIALES Y MÉTODOS

En el estudio de caso para el análisis estadístico, participaron estudiantes (del segundo al cuarto semestre), de la carrera de Pedagogía de la Actividad Física y Deporte, de la Facultad de Educación Física, Deportes y Recreación, de la Universidad de Guayaquil. Se aplicó la fórmula para el cálculo del tamaño de muestra para población finita (1), como aparece en Navia et al., (2018). Se utilizó un muestreo estratificado a partir de cada semestre para luego seleccionar las submuestras de cada estrato según muestreo aleatorio simple.

$$n = \frac{Z^2 pqN}{e^2(N-1) + Z^2 pq} \quad (1)$$

Donde:

n: Tamaño de muestra

Z: Nivel de confianza

N: Tamaño de población

p: Fracción de la población con el atributo deseado

q: Fracción de la población sin el atributo deseado

e: Máximo error de estimación

A partir de la matrícula de cada semestre se pudo determinar el tamaño de las submuestras como se puede observar en la tabla 1.

Tabla 1. Composición de la muestra seleccionada por estrato.

Semestre	Matrícula	Porcentaje	Submuestra
Segundo	136	35%	41

Tercero	124	32%	37
Cuarto	129	33%	38
Total	389	100%	116

Se midieron el peso corporal y la estatura de cada estudiante, con el fin de comprobar su nivel de relación con el índice de fuerza de reacción (RSI, por sus siglas en inglés). El RSI se asume como la relación entre la altura de salto (A) y el tiempo de contacto del pie con la alfombra del dinamómetro (TC), como propone Flanagan (2016).

Para realizar las mediciones necesarias de A y TC, se utilizó en el estudio una plataforma de contacto (Axon Jump). Axon Jump es instrumento comandado por software, que tiene incorporado un cronómetro de alta resolución para calcular la altura y la velocidad de los saltos y que, en saltos técnicamente bien ejecutados, alcanza una alta precisión de las mediciones (Malagón et al., 2019). Los alumnos realizaron una serie de 5 saltos de contramovimiento, (CMJ, por sus siglas en inglés), consistente en un salto vertical con extensiones simultáneas de brazos, con pausa de 1 minuto entre saltos.

Para los cálculos se utilizaron los datos promediados de los tres saltos de mayor calidad técnica en la ejecución. El género no se incluyó intencionadamente estratificación inicial, ya que la separación adicional de hombres y mujeres podría introducir grupos demasiado pequeños, lo que haría imposible realizar un análisis de la varianza adecuado.

Este procedimiento se consideró aceptable debido al carácter ilustrativo de este estudio y solo se registró el sexo de los participantes con el objetivo de aplicar la prueba t-Student de comparación de medias respecto al RSI. Para lo cual se supuso la hipótesis nula de que el sexo no está relacionado con el RSI calculado para cada participante. Mientras que, con el análisis de varianza, se pretende determinar la posible relación entre el semestre que cursa el estudiante y los resultados de altura del salto alcanzados. Los análisis estadísticos se realizaron con el programa IBM SPSS-Statistics 25 y en todas las pruebas se asumió un nivel de significación de 0,05. El análisis estadístico comenzó con la estimación de estadísticas descriptivas básicas como: media, mediana, desviación estándar, rango intercuartil, simetría y curtosis.

Se realizó la prueba t propiamente dicha, aunque su resultado no es objeto de esta disertación, pero se centró la atención en la comprobación de varios supuestos, a saber, la normalidad de la distribución, que se comprobó mediante el método de Shapiro-Wilk y la prueba de Kolmogorov-Smirnov, como complemento. La homogeneidad de la varianza se verificó mediante la prueba de

Levene. En caso del incumplimiento de los supuestos, se utilizaron algunas de las transformaciones mencionadas en la introducción del presente artículo, y como alternativa a la imposibilidad de corrección del incumplimiento, se aplicó la prueba no paramétrica equivalente.

RESULTADOS Y DISCUSIÓN

En la tabla 2 se muestran los valores de los estadísticos descriptivos seleccionados en esta investigación con el objetivo de revisar el comportamiento de cada variable continua (A, TC y RSI), en cuanto a tendencia central, dispersión, rango de valores e intercuartílico, asimetría y curtosis de su distribución, como medidas iniciales del análisis.

Tabla 2. Estadísticas descriptivas para la muestra total sin desagregación por estratos.

Estadísticos	Variables		
	Altura del salto (cm)	Tiempo de contacto (s)	RSI (cm/s)
Media	31,031	0,137	229,288
Mediana	31,144	0,134	224,917
Desv. Desviación	2,949	0,017	31,723
Mínimo	26,140	0,110	171,813
Máximo	35,947	0,165	309,818
Rango intercuartil	5,440	0,030	47,042
Asimetría	-0,038	0,163	0,396
Curtosis	-1,283	-1,341	-0,395

Como se puede apreciar, resulta destacable que los valores observados de cada variable muestran resultados similares para sus medias aritméticas y el valor central de la distribución (mediana), lo cual indica cierto nivel de simetría (lo cual se confirma en dicho estadístico para la mayoría de las variables) y valores relativamente bajos de dispersión. La altura del salto varía en un rango de casi 10 cm mientras que el tiempo de contacto muestra una dispersión de 0,017 segundos.

Antes de realizar un análisis de comparación de medias de la variable "Altura del salto" entre los grupos (semestres) mediante un análisis de varianza, se procede a realizar una prueba de normalidad para los valores estratificados ya que se precisa demostrar que cada una de las submuestras de los estratos debe proceder de una población normal, con varianzas homogéneas, antes de realizar el ANOVA. En la tabla 3 se muestran los resultados de las pruebas realizadas.

Tabla 3. Pruebas de normalidad para la comparación de medias de la variable A.

Variable	Grupo	Kolmogorov-Smirnova			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
Altura del salto	Segundo semestre	0,108	41	,200*	0,926	41	0,010
	Tercer semestre	0,123	37	0,168	0,927	37	0,018
	Cuarto semestre	0,084	38	,200*	0,955	38	0,126

Según los resultados de la prueba de Kolmogorov-Smirnov, las submuestras proceden de poblaciones normales, con significaciones superiores a 0,16. Sin embargo, como las submuestras tiene un tamaño menor que 50, se presta mayor atención a los resultados de la prueba de Shapiro-Wilk, la cual resulta más potente en estos casos (Mishra et al., 2019). Según dicha prueba, los valores de A para el segundo y tercer semestre no siguen una distribución normal para el valor de significación seleccionado (0,05), pues presentan significatividades asintóticas respectivas de 0,010 y 0,018. No así en el caso del cuarto semestre, el cual presenta una significatividad de la prueba de 0,126.

Por tanto, se necesita transformar esta variable. Primeramente se aplica la transformación de logaritmo neperiano de los datos de A y se procede a repetir las pruebas de normalidad. La variable transformada se denota en lo adelante como LnA ($\text{LnA} = \ln A$). Los nuevos resultados se muestran en la tabla 4.

Tabla 4. Pruebas de normalidad para la comparación de medias de la variable transformada LnA

Variable	Grupo	Kolmogorov-Smirnova			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
LnA	Segundo semestre	0,110	41	,200*	0,922	41	0,008
	Tercer semestre	0,119	37	,200*	0,927	37	0,018
	Cuarto semestre	0,088	38	,200*	0,953	38	0,112

Como se puede apreciar en la tabla 4, las significatividades de la prueba para los dos primeros semestres siguen siendo inferiores a 0,05, por lo que se rechaza la hipótesis nula de normalidad de la submuestra respectiva. A continuación, se aplica la transformación inversa para obtener la variable transformada IA ($IA=1/A$), para proceder a repetir el proceso de verificación de normalidad como se muestra en la tabla 5.

Tabla 5. Pruebas de normalidad para la comparación de medias de la variable transformada IA

Variable	Grupo	Kolmogorov-Smirnova			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
IA	Segundo semestre	0,119	41	0,150	0,916	41	0,005
	Tercer semestre	0,114	37	,200*	0,925	37	0,016
	Cuarto semestre	0,098	38	,200*	0,948	38	0,078

Esta transformación inversa generó un efecto contrario al deseado, ya que, en esta ocasión, incluso se pierde la normalidad de los datos relacionados con el cuarto semestre (significación = 0,078), por lo que se realiza una última transformación, esta vez la cuadrática, que genera la variable SA ($SA=A^2$). Y se repite el procedimiento actual, cuyos resultados aparecen en la tabla 6.

Tabla 6. Pruebas de normalidad para la comparación de medias de la variable transformada IA

Variable	Grupo	Kolmogorov-Smirnova			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
SA	Segundo semestre	0,108	41	,200*	0,924	41	0,009
	Tercer semestre	0,121	37	0,186	0,927	37	0,018
	Cuarto semestre	0,082	38	,200*	0,954	38	0,123

Mediante la transformación cuadrática se logra obtener solo una submuestra con distribución normal (cuarto semestre, con significatividad de 0,123), sin embargo, no se pudo corregir la falta de normalidad en las restantes. Por tanto, se hace necesario aplicar una prueba equivalente. Pero antes debe verificarse el supuesto de homogeneidad de varianzas para no incurrir en los errores tratados por Gantz et al., (2016). Los resultados de la aplicación de la prueba de homogeneidad de varianza de Levene se muestran en la tabla 7.

Tabla 7. Prueba de homogeneidad de varianza

		Estadístico de Levene	gl1	gl2	Sig.
Altura del salto	Se basa en la media	0,333	2	113	0,718
	Se basa en la mediana	0,244	2	113	0,784
	Se basa en la mediana y con gl ajustado	0,244	2	111,861	0,784
	Se basa en la media recortada	0,329	2	113	0,720

Los resultados obtenidos en SPSS, para la prueba de Levene, en cada una de sus categorías, permite decidir si las varianzas poblacionales son iguales. Como la probabilidad asociada al estadístico de Levene es, en todos los casos, muy superior a la significatividad de la prueba (0,05), se puede afirmar que no existen evidencias estadísticas para rechazar la hipótesis nula en la que se asume que las varianzas poblacionales son iguales.

Se puede, por tanto, aplicar la prueba no paramétrica alternativa equivalente de Kruskal-Wallis para muestras independientes. Se parte de la hipótesis nula que expresa que la altura del salto es la misma entre las categorías de los grupos (semestres). En la figura 1 se puede observar el gráfico de caja y sesgo de la prueba realizada.

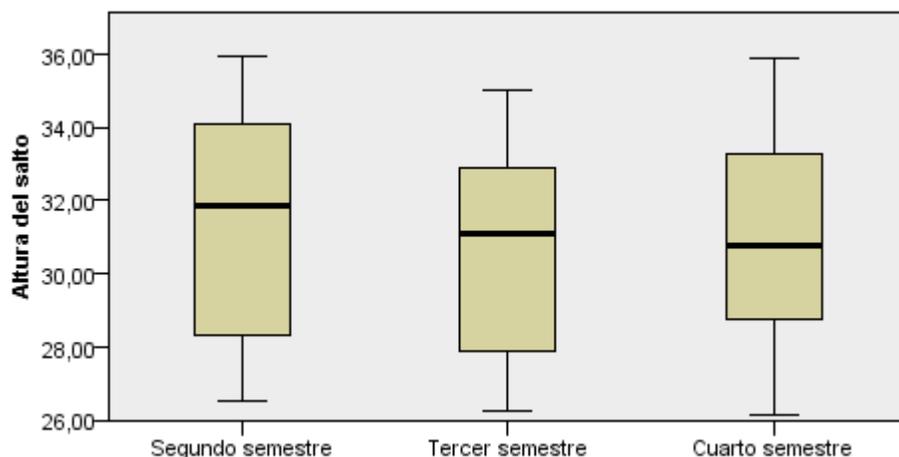


Figura 1. Gráfico de caja y sesgo de la prueba de Kruskal-Wallis para la altura del salto.

Resulta oportuno significar que, en la presentación de los resultados en los trabajos científicos, es tan importante como el análisis estadístico correctamente realizado, la decisión sobre el tipo de gráfico de caja y sesgo a manejar. Por ejemplo, cuando se utilizan pruebas no paramétricas, es importante que se trabaje en una escala ordinal, por lo que, al presentar los resultados de forma gráfica, no se deben crear diagramas de caja y bigotes con valores de media - error estándar - $1,96 \cdot \text{error estándar}$, sino utilizar valores de mediana - 25/75 % - mínimo/máximo.

Se aprecia que para cada categoría las medianas son similares y que, incluso la dispersión, analizada a partir del rango intercuartílico observado, es similar para cada categoría analizada, las cuales se mueven dentro del rango 26-36. Para confirmar lo observado, se muestran en la tabla 8, los resultados de la prueba.

Tabla 8. Prueba de Kruskal-Wallis para la altura del salto

N total	Estadístico de contraste	Grados de libertad	Sig. Asintótica (prueba bilateral)
116	0,861	2	0,65

Con una alta significación ($0,65 > 0,05$), no se rechaza la hipótesis nula de igualdad de medianas, por lo que se puede concluir que no existen diferencias significativas entre los grupos en cuanto a la altura del salto registrada para cada estudiante. De manera que el semestre que esté cursando, no tiene una influencia significativa en la altura del salto ejecutado.

Para determinar si el sexo del estudiante es determinante en el RSI logrado, se debe realizar una prueba t-Student de comparación de medias para muestras independientes, para lo cual se requiere verificar la normalidad de los datos, así como la homogeneidad de las varianzas de las submuestras. Los resultados de las pruebas de normalidad se muestran en la tabla 9.

Tabla 9. Pruebas de normalidad para la comparación de medias de la variable RSI

Variable	Sexo Estadístico	Kolmogorov-Smirnova			Shapiro-Wilk		
		gl	Sig.	Estadístico	gl	Sig.	
RSI	Femenino	0,130	62	0,011	0,955	62	0,024
	Masculino	0,094	54	,200	0,977	54	0,384

En el caso de las estudiantes, se rechaza la hipótesis nula de normalidad pues la significatividad de la prueba de Kolmogorov-Smirnov (0,011) es inferior a 0,05, pero para los datos de RSI de los del sexo masculino, se puede afirmar que sigan una distribución normal (0,200>0,05, no se rechaza la hipótesis nula de normalidad). Debido a esto, se realiza la transformación logarítmica de los datos para corregir el incumplimiento del supuesto de normalidad necesario. Se genera por tanto la variable LnRSI ($\text{LnRSI}=\ln\text{RSI}$). En la tabla 10 se pueden consultar los resultados obtenidos de las pruebas de normalidad de la variable transformada.

Tabla 10. Pruebas de normalidad para la comparación de medias de la variable LnRSI

Variable	Sexo	Kolmogorov-Smirnov			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
LnRSI	Femenino	0,101	62	0,191	0,972	62	0,159
	Masculino	0,095	54	,200*	0,982	54	0,583

En esta ocasión, la transformación logarítmica fue suficiente para corregir la violación del supuesto de normalidad pues, para ambas categorías, se obtienen significatividades superiores a 0,05 de la prueba de Kolmogorov-Smirnov. Solo resta verificar el supuesto de homogeneidad de varianzas, mediante la prueba de Levene. La prueba de Levene arrojó un valor del estadístico F de 1,512 y una significatividad de 0,221, por lo que se pueden asumir varianzas iguales para ambas submuestras. Cabe mencionar que, aunque no se encontraron problemas con el cumplimiento del supuesto de homogeneidad de la varianza, esto puede deberse a la transformación aplicada, la cual puede haberla corregido o estabilizado en caso de que hubiese existido anteriormente.

Al aplicar la prueba t-Student de comparación de medias para muestras independientes, se obtuvieron los resultados q se muestran en la tabla 11.

Tabla 11. Prueba t-Student de comparación de medias para muestras independientes de la variable LnRSI

Variable	Igualdad de varianzas	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
							Inferior	Superior
LnRSI	Se asumen varianzas iguales	-1,470	114	0,144	-0,03736	0,02541	-0,08769	0,01297
	No se asumen varianzas iguales	-1,492	113,399	0,138	-0,03736	0,02504	-0,08697	0,01225

De acuerdo con estos resultados, no se rechaza la hipótesis nula de igualdad de medias para varianzas iguales, pues la significatividad de la prueba (0,144), es mayor que la significación asumida. Además, esto se corrobora a partir de que el intervalo de confianza de la diferencia (-0,087:0,012) incluye el valor cero. Por lo que se puede asumir que no existen diferencias significativas entre el promedio de los valores de RSI entre las hembras y los varones.

A partir de los elementos hasta aquí expuestos, se puede afirmar que, conociendo los principios de la selección correcta de las herramientas estadísticas y de la comprobación y realización de las pruebas, la verificación de las hipótesis relativas a la comparación de las capacidades motrices de los alumnos en función de diversos factores se

convierte en una tarea nada difícil. Al aplicar algunas reglas básicas del análisis estadístico, la comprobación de las regularidades que se producen en el deporte se reduce a un esquema de conducta determinado, que, con la ayuda de un software especializado, pueden llegar a aplicarse con relativa facilidad.

El conocimiento de un procedimiento alternativo en forma de introducción de transformaciones, permite a menudo utilizar las pruebas paramétricas a pesar de los supuestos iniciales no cumplidos. Gracias a este procedimiento, la potencia de la prueba se conserva e incluso aumenta, siendo a menudo superior a la de las pruebas no paramétricas equivalentes, y por tanto la probabilidad de cometer un error del segundo tipo (no rechazar una hipótesis nula falsa) es menor.

Una variable que, incluso después de las transformaciones realizadas, no mostró una distribución normal, es la altura del salto. Algunos investigadores consideran que el análisis de la varianza es ampliamente robusto a las desviaciones de la normalidad (Jiménez, 2021), especialmente para grupos iguales, pero con la condición de que las medias y las desviaciones estándar no estén correlacionadas. Si se quiere hacer referencia a esta regularidad, la misma debe comprobarse, por ejemplo, mediante un gráfico de dispersión de las desviaciones estándar frente a las medias.

CONCLUSIONES

En el análisis de contrastes de hipótesis de comparación de medias de datos relacionados con investigaciones deportivas y de la cultura física, se utilizan dos grandes familias de pruebas, las paramétricas (supuestos de normalidad y homogeneidad de varianzas poblacionales) y las no paramétricas. La correcta verificación y tratamiento (transformaciones de variables según se requiera) de los supuestos asociados a estas, conducen a resultados estadísticamente fiables pues garantizan la potencia de las pruebas. Entre las paramétricas destacan, la prueba t de Student y el análisis de varianza (ANOVA de una vía), utilizadas frecuentemente en la investigación biométrica deportiva.

En esta investigación, la variable Altura del salto no pudo ser normalizada mediante las transformaciones propuestas en la bibliografía, elemento que imposibilitó la aplicación de la prueba ANOVA. Sin embargo, sí se pudo verificar la homogeneidad de varianzas, por lo que se hizo posible recurrir a la prueba de Kruskal-Wallis para comparar los tres subgrupos estudiados. Los resultados de la prueba permitieron concluir que los estudiantes de los tres semestres alcanzan alturas de salto similares.

La variable Índice de la fuerza de reacción también verificó el incumplimiento del supuesto de normalidad, lo cual pudo corregirse mediante una transformación logarítmica para aplicar luego la prueba t de comparación de medias para muestras independientes, luego de verificar además la igualdad de varianzas. La comparación de las medias permitió asumir que los estudiantes de ambos sexos registran un índice de fuerza de reacción similar.

REFERENCIAS BIBLIOGRÁFICAS

- Abidovoch, T. O. (2020). Modern pedagogical mechanisms of the growth of physical culture among the students trained in the higher education system of uzbekistan. *European Journal of Research and Reflection in Educational*, 8(7).
- Borovkov, A. A., & Moullagaliev, A. (2019). Mathematical statistics. <https://www.taylorfrancis.com/books/mono/10.1201/9780203749326/mathematical-statistics-borovkov-moullagaliev>
- Cortés, F. (2018). Observación, causalidad y explicación causal. *Perfiles latinoamericanos*, 26(52), 0-0.
- Di Domenico, F., & Raiola, G. (2021). Muscle strength and the complexity of human movement. <http://rua.ua.es/dspace/handle/10045/116308>
- Jiménez-Miranda, V. (2021). La influencia del tamaño muestral en la selección de pruebas paramétricas o no paramétricas: estudio comparativo mediante simulación de Monte Carlo (Tesis de Maestría, Universitat Oberta de Catalunya).
- Flanagan, E. (2016). The reactive strength index revisited. *Train With Push*. <http://www.fitstoronto.com/wp-content/uploads/2017/12/The-Reactive-Strength-Revisited-Part-3.pdf>
- Glantz, S; Slinker, B & Neilands, T. Neilands. (2016). *Primer of applied regression & analysis of variance*. McGraw-Hill Education / Medical.
- Luna Orozco, K. (2019). *Asociación entre la discapacidad física y la incidencia de síntomas depresivos en adultos de 50 años o más en México, 2001-2003* (Tesis de Maestría, Universidad del Norte).
- Malagón, C. E. S., Tova, P. J. A., & Vargas, D. S. B. (2019). Correlación entre potencia y resistencia en jugadores de fútbol de Boyacá. *Revista Digital: Actividad Física Y Deporte*, 5(1), 26-33.

- Mancha-Triguero, D., Reina, M., García-Rubio, J., & Ibáñez, S. J. (2021). Does physical fitness influence the technical-tactical performance indicators in a professional female basketball team?[¿ Influye la condición física en los indicadores de rendimiento técnico-táctico en un equipo profesional de baloncesto femenino?]. *RICYDE. Revista Internacional de Ciencias del Deporte*. doi: 10.5232/ricyde, 17(64), 174-188.
- McBurnie, A. J., Harper, D. J., Jones, P. A., & Dos'Santos, T. (2021). Deceleration Training in Team Sports: Another Potential 'Vaccine'for Sports-Related Injury?. *Sports medicine*, 1-12.
- Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, 22(1), 67.
- Moder, K. (2016). How to keep the Type I Error Rate in ANOVA if Variances are heteroscedastic. *Austrian Journal of Statistics*, 36(3), 179–188. <https://doi.org/10.17713/ajs.v36i3.329>
- Navia, I., Chávez, A., & Bravo, G. (2018). Niveles de estrés y rendimiento académico en los estudiantes. *Atlante Cuadernos de Educación y Desarrollo*, (Mayo). <https://www.eumed.net/rev/atlante/2018/05/niveles-estres-estudiantes.html>
- Palacio, E. S., Recuero, J. A., Moreno, A. M., & Duarte, M. M. (2020). Autocontrol, el antídoto contra la violencia en el deporte. *Revista iberoamericana de psicología del ejercicio y el deporte*, 15(3), 135-139.
- Peláez, I. M. (2012). Comparación de medias. *Revista SEDEN. Recuperado de <https://revistaseden.org/files/12-CAP>*.
- Pérez-Martínez, I., Martínez-García, M., & Quintana-Díaz, A. (2020). Introducción al estudio de variables relacionadas con la velocidad del lanzamiento en el béisbol. *Podium. Revista de Ciencia y Tecnología en la Cultura Física*, 15(1), 84-98.
- Sánchez-Sixto, A.; Harrison, A.J. y Floría, P. (2019) La importancia de la profundidad del contramovimiento en el ciclo estiramiento-acortamiento / Importance of Countermovement Depth in Stretching and Shortening Cycle Analysis. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte* vol. 19 (73) pp. 33-44
- Salas, A. (2018). *Métodos estadísticos para la Investigación Científica*. Editorial Grupo Compás.
- Sánchez Turcios, R. A. (2015). t-Student: Usos y abusos. *Revista mexicana de cardiología*, 26(1), 59-61.
- Teijón-Alcalá, M., & Serrano-Maillo, A. (2019). Bloqueo de oportunidades y emociones negativas en la causación de infracciones deportivas: un test de la teoría general de la frustración. <http://digital.casalini.it/4578070>