

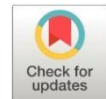


Análisis factorial exploratorio multivariado de patrones de consumo de productos de primera necesidad en Mipymes de la provincia de Chimborazo año 2024

Multivariate exploration factor analysis of consumption patterns of essential necessities in small and medium-sized enterprises in the province of Chimborazo, year 2024

- ¹ Francisco Eduardo Toscano Guerrero  <https://orcid.org/0000-0002-3951-7774>
Escuela Superior Politécnica de Chimborazo (ESPOCH), Riobamba, Ecuador.
Maestría en Matemática Aplicada
francisco.toscano@epoch.edu.ec
- ² Liliana Alejandra Funes Samaniego  <https://orcid.org/0000-0002-5364-0699>
Escuela Superior Politécnica de Chimborazo (ESPOCH), Riobamba, Ecuador.
Maestría en Estadística Aplicada
alejandra.funes@epoch.edu.ec



Artículo de Investigación Científica y Tecnológica

Enviado: 11/07/2025

Revisado: 08/08/2025

Aceptado: 08/09/2025

Publicado: 09/10/2025

DOI: <https://doi.org/10.33262/concienciadigital.v8i4.3545>

Cítese:

Toscano Guerrero, F. E., & Funes Samaniego, L. A. (2025). Análisis factorial exploratorio multivariado de patrones de consumo de productos de primera necesidad en Mipymes de la provincia de Chimborazo año 2024. *ConcienciaDigital*, 8(4), 6-31. <https://doi.org/10.33262/concienciadigital.v8i4.3545>



CONCIENCIA DIGITAL, es una revista multidisciplinar, **trimestral**, que se publicará en soporte electrónico tiene como **misión** contribuir a la formación de profesionales competentes con visión humanística y crítica que sean capaces de exponer sus resultados investigativos y científicos en la misma medida que se promueva mediante su intervención cambios positivos en la sociedad. <https://concienciadigital.org>
La revista es editada por la Editorial Ciencia Digital (Editorial de prestigio registrada en la Cámara Ecuatoriana de Libro con No de Afiliación 663) www.celibro.org.ec



Esta revista está protegida bajo una licencia Creative Commons en la 4.0 Internacional. Copia de la licencia: <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Palabras claves:

Fidelización,
multivariado,
analítica,
muestreo,
exploratorio,
patrones, enfoque,
correlacional,
descriptivo.

Keywords:

Loyalty,
multivariate,
analytical,
sampling,
exploratory,
patterns,
approach,
correlational,
descriptive.

Resumen

Introducción: en el presente proyecto es pertinente gestionar actividades productivas comerciales de sostenibilidad, fidelización de clientes, penetración en nuevos mercados e innovación orientando esfuerzos y recursos en la implementación de Análisis factorial exploratorio multivariado de patrones de consumo de productos de primera necesidad en Mipymes de la provincia de Chimborazo año 2024. **Objetivos:** examinar los patrones de consumo de productos de primera necesidad dentro del sector de las Mipymes en la provincia de Chimborazo, utilizando un enfoque multivariado basado en el análisis factorial exploratorio. **Metodología:** enfoque cuantitativo, fundamentado en la aplicación del análisis factorial exploratorio multivariado como técnica estadística central para identificar patrones latentes en los hábitos de consumo de productos de primera necesidad en el contexto de las Mipymes de la provincia de Chimborazo. **Resultados:** la consumición de leche no genera diferencias significativas en las medianas de sueldo. Las variaciones entre grupos etarios son mínimas y no estadísticamente significativas, lo que indica que la edad por sí sola no influye considerablemente en el sueldo dentro de esta muestra. **Conclusiones:** los datos son consistentes, y se cumple homocedasticidad y homogeneidad, pero no la normalidad. No hay evidencia de asociación entre consumo de leche, edad y nivel de ingreso. **Área de estudio general:** Administración de Empresas. **Área de estudio Específica:** Actividades productivas comerciales de sostenibilidad. **Tipo de artículo:** Original.

Abstract

Introduction: In this project it is pertinent to manage commercial productive activities of sustainability, customer loyalty, penetration in new markets and innovation by directing efforts and resources in the implementation of Multivariate Exploratory Factor Analysis of consumption patterns of necessities in MSMEs in the province of Chimborazo year 2024. **Objectives:** To examine the consumption patterns of necessities within the MSME sector in the province of Chimborazo, using a multivariate approach based on exploration factor analysis. **Methodology:** quantitative approach, based on the application of multivariate exploratory factor analysis as a central statistical technique to identify latent patterns in the consumption habits of necessities in the context of MSMEs in the province of

Chimborazo. **Results:** Milk consumption does not generate significant differences in median salaries. The variations between age groups are minimal and not statistically significant, indicating that age alone does not significantly influence salary within this sample. **Conclusions:** the data are consistent, and homoscedasticity and homogeneity are met, but not normality. There is no evidence of an association between milk consumption, age, and income level. **General area of study:** Business Administration. **Specific area of study:** Commercial productive activities of sustainability. **Item type:** Original.

1. Introducción

El análisis de los patrones de consumo de productos de primera necesidad en las Micro, Pequeñas y Medianas Empresas (Mipymes) constituye un componente esencial para comprender la estructura operativa y las decisiones estratégicas de este segmento empresarial. En la provincia de Chimborazo, caracterizada por su diversidad económica y cultural, las Mipymes desempeñan un rol fundamental en el sostenimiento del empleo y la dinamización del comercio interno, como establece Sumba et al. (2022). Frente a una economía regional en constante transformación, se hace indispensable utilizar técnicas estadísticas multivariadas que permitan identificar, resumir y clasificar las principales variables asociadas a los hábitos de consumo en este tipo de unidades productivas (Hair et al., 2019).

En este contexto el Análisis Factorial Exploratorio (AFE) se posiciona como una herramienta robusta y eficaz para descubrir estructuras latentes entre múltiples variables interrelacionadas, especialmente en investigaciones de tipo socioeconómico, según manifiesta Lloret-Segura et al. (2014). La metodología permite reducir la dimensionalidad del conjunto de datos, facilitando la comprensión de los factores subyacentes que determinan el comportamiento de compra y abastecimiento en las Mipymes, sin perder información esencial del fenómeno observado (Field, 2018).

El estudio de productos de primera necesidad —como alimentos, artículos de limpieza e insumos básicos— en el ámbito de las Mipymes cobra relevancia en escenarios post pandémicos, donde los cambios en la oferta, los precios y la logística han alterado significativamente los esquemas de consumo empresarial (Palomeque, 2024). Estas alteraciones no solo inciden en la sostenibilidad financiera de los negocios, sino también en la capacidad de adaptación a contextos de incertidumbre económica, en especial en

provincias con alto índice de informalidad como Chimborazo (Instituto Nacional de Estadística y Censos [INEC], 2023).

Además, el uso del análisis factorial exploratorio en estudios sobre consumo permite captar patrones recurrentes que de otro modo quedarían ocultos en análisis univariados o bivariados. Es así como a través de una estructura factorial bien definida, se pueden identificar conglomerados de productos cuyo consumo es simultáneo, frecuencias de abastecimiento homogéneas entre grupos de empresas o dependencias cruzadas en la toma de decisiones (Costello & Osborne, 2005). Este tipo de información es invaluable para generar políticas públicas que fomenten cadenas de suministro más eficientes, incentivos fiscales adecuados y programas de financiamiento focalizados (Moreno et al., 2018).

La elección de Chimborazo como espacio geográfico de estudio responde no solo a su representatividad como zona de producción agrícola y comercialización de bienes básicos, sino también a la creciente formalización de Mipymes en sectores rurales y urbanos marginales, lo cual exige una caracterización más técnica de sus necesidades operativas (Espín, 2022). A través del AFE se puede generar un modelo explicativo con evidencia empírica que sirva como base para la toma de decisiones tanto del sector público como del privado.

Por otra parte la multivarianza del fenómeno de consumo implica que múltiples variables como frecuencia de compra, tipo de proveedor, nivel de ingreso, acceso a crédito o tamaño de la empresa interactúan simultáneamente, dificultando el análisis tradicional. En este sentido, el AFE se convierte en una alternativa metodológica ideal para detectar factores comunes que agrupan comportamientos y prácticas de adquisición similares (Tabachnick & Fidell, 2019).

El desarrollo de este tipo de investigaciones también promueve la creación de marcos teóricos más sólidos que trasciendan la descripción superficial del consumo, aportando a la formulación de modelos predictivos que puedan replicarse en otras provincias del Ecuador o en países con estructuras económicas similares (Hair et al., 2021). Así el presente estudio no solo pretende caracterizar los patrones de consumo de productos esenciales, sino también ofrecer una visión estratégica de las dinámicas de abastecimiento en las Mipymes como unidad económica clave.

El análisis del consumo de productos de primera necesidad en el contexto de las Mipymes fue abordado desde múltiples enfoques en la literatura regional e internacional. Investigaciones realizadas en países de América Latina demostraron que los patrones de compra y los factores que condicionan la demanda de bienes esenciales están estrechamente relacionados con aspectos de naturaleza socioeconómica, tecnológica y cultural (Denegri et al., 2023).

En el caso ecuatoriano trabajos como los de Sumba et al. (2022) subrayaron el rol estratégico que desempeñan las Mipymes en la distribución de productos básicos, al tiempo que identifican las limitaciones logísticas y financieras que estas empresas enfrentan para garantizar un abastecimiento eficiente a la población. Complementariamente informes de la Comisión Económica para América Latina y el Caribe (CEPAL, 2021) destacan cómo la digitalización y la incorporación de tecnologías emergentes han transformado los canales de acceso a bienes esenciales, modificando los hábitos de consumo en distintos segmentos poblacionales.

En la provincia de Chimborazo, estudios recientes han examinado el impacto de la urbanización y del acceso a plataformas de comercio electrónico sobre los mecanismos de adquisición de productos de primera necesidad. No obstante se mantiene una brecha en cuanto a la segmentación del mercado y la identificación de patrones de comportamiento que permitan a las Mipymes adaptarse de manera efectiva a las nuevas dinámicas de consumo.

Autores como Knehta et al. (2019) utilizaron el análisis factorial exploratorio para segmentar mercados y analizar el efecto de variables como precio, disponibilidad y percepción de calidad sobre el comportamiento del consumidor. En Ecuador la aplicación de esta técnica sigue siendo limitada, lo que resalta la necesidad de estudios que permitan su implementación y validación en el contexto nacional.

Este estudio busca contribuir al cuerpo teórico existente mediante un análisis riguroso de los factores que influyen en los patrones de consumo en Chimborazo, incorporando un enfoque cuantitativo sustentado en técnicas estadísticas multivariadas, con el fin de fortalecer los procesos de toma de decisiones en el entorno operativo de las Mipymes.

El presente estudio tiene como **objetivo** examinar los patrones de consumo de productos de primera necesidad dentro del sector de las Mipymes en la provincia de Chimborazo, utilizando un enfoque multivariado basado en el análisis factorial exploratorio. Esta metodología permitirá identificar los factores clave que influyen en las decisiones de compra de los consumidores, con el fin de diseñar estrategias comerciales más efectivas y alineadas con las dinámicas del mercado local. Para una aproximación más concreta, se tomará como caso de estudio el consumo de leche, un producto esencial dentro de la canasta básica, sobre el cual se realizarán exhaustivamente las pruebas estadísticas necesarias para validar los supuestos del modelo aplicado.

Objetivos Específicos

- Analizar los factores de carácter económico, social y tecnológico que ejercen una influencia significativa sobre los patrones de consumo de productos de primera

necesidad en la provincia de Chimborazo, con el propósito de comprender la complejidad de las decisiones de compra en distintos contextos poblacionales.

- Explorar y contrastar los hábitos de consumo de los consumidores que habitan en zonas urbanas frente a aquellos de áreas rurales, con el fin de identificar diferencias estadísticamente significativas en sus preferencias, motivaciones y comportamientos relacionados con la adquisición de productos esenciales.
- Examinar el impacto que tiene la digitalización del comercio en los procesos de distribución y compra de productos de primera necesidad dentro de las Mipymes de Chimborazo, considerando el nivel de implementación tecnológica y la capacidad de adaptación de estas unidades productivas al entorno digital.

2. Metodología

Este estudio se enmarca en un enfoque cuantitativo, fundamentado en la aplicación del análisis factorial exploratorio multivariado como técnica estadística central para identificar patrones latentes en los hábitos de consumo de productos de primera necesidad en el contexto de las Mipymes de la provincia de Chimborazo. La elección de esta metodología permite generar evidencia empírica rigurosa y objetiva, orientada a sustentar la toma de decisiones estratégicas en el ámbito comercial y logístico. De este modo, se busca aportar al fortalecimiento de la gestión operativa y a la optimización del desempeño empresarial en el sector.

2.1. Enfoque metodológico

Este estudio se desarrolla bajo un enfoque cuantitativo, sustentado en la recolección y análisis de datos numéricos con el fin de explorar relaciones significativas entre variables vinculadas al comportamiento de consumo. Esta perspectiva metodológica permite identificar patrones recurrentes y tendencias emergentes a partir de información estructurada, brindando una base sólida para la interpretación objetiva de los resultados mediante técnicas estadísticas de nivel avanzado.

Como eje central del análisis se utilizará el análisis factorial exploratorio multivariado, una herramienta que facilita la reducción de la complejidad de los datos al condensar múltiples variables en factores subyacentes. Esta técnica posibilita descubrir estructuras latentes en las decisiones de compra, así como segmentar a los consumidores en grupos homogéneos según sus características y comportamientos, permitiendo determinar los factores determinantes que influyen en la adquisición de productos de primera necesidad.

2.2. Diseño de la investigación

El presente estudio adopta un diseño metodológico de tipo no experimental y de corte transversal. Se clasifica como no experimental en la medida en que no se interviene ni manipulan intencionadamente las variables en estudio, sino que estas son observadas tal

como ocurren en su entorno natural, permitiendo un análisis descriptivo y relacional de los fenómenos investigados. A su vez, el carácter transversal del estudio responde al hecho de que la recolección de datos se efectuará en un único momento temporal, lo cual permitirá capturar una visión precisa y representativa de los patrones de consumo de productos de primera necesidad en la provincia de Chimborazo durante el año 2024.

2.3. Población y muestra

La población considerada en este estudio está conformada por consumidores de productos de primera necesidad que residen en la provincia de Chimborazo, así como por propietarios y administradores de Mipymes dedicadas a la comercialización de dichos bienes. Este enfoque doble permite abordar de forma integral tanto las dinámicas de consumo como los procesos comerciales involucrados. Para la selección de participantes se utilizará un diseño muestral probabilístico estratificado, lo que garantizará una representación equitativa entre las diferentes zonas urbanas y rurales de la provincia. Esta estrategia metodológica favorece la diversidad de los datos y respalda la validez estadística del análisis. El tamaño de la muestra será calculado en función de criterios de precisión y nivel de confianza, con el objetivo de asegurar la robustez de los resultados derivados del análisis factorial exploratorio. La data se obtuvo del levantamiento de información en base a encuestas levantadas por miembros del Grupo de Investigación BI-Data de la Escuela Superior Politécnica de Chimborazo.

CÓDIGO EN R

Cargar librerías necesarias

```
library(ggplot2)
```

```
library(dplyr)
```

```
library(readxl)
```

```
library(car)
```

```
library(nortest)
```

```
library(lmtest)
```

Cargar datos

```
datos <- read_excel("D:/POLITÉCNICA DEL CARCHI/PARA LA TESIS/Artículo Upec para  
graduación/encuesta_comportamiento_consumidor_2024.xlsx", sheet = "Hoja1")
```

Renombrar columnas si es necesario

```
colnames(datos) <- c("Persona", "Canton", "Consume_Leche", "Sueldo", "Edad")
```

Eliminación de NA y filtrado por z-score

www.concienciadigital.org

```
datos <- na.omit(datos)
```

```
datos$Z_Sueldo <- scale(datos$Sueldo)
```

```
datos_limpios <- subset(datos, abs(Z_Sueldo) < 3)
```

Visualización inicial

```
ggplot(datos_limpios, aes(x = Sueldo)) +
```

```
  geom_histogram(fill = "skyblue", color = "black", bins = 20) +
```

```
  ggtitle("Histograma del Sueldo Mensual") +
```

```
  theme_minimal()
```

```
ggplot(datos_limpios, aes(y = Sueldo)) +
```

```
  geom_boxplot(fill = "lightgreen") +
```

```
  ggtitle("Boxplot del Sueldo Mensual") +
```

```
  theme_minimal()
```

Análisis del sueldo por consumo y edad

```
ggplot(datos_limpios, aes(x = Consume_Leche, y = Sueldo, fill = Consume_Leche)) +
```

```
  geom_boxplot() +
```

```
  facet_wrap(~cut(datos_limpios$Edad, breaks = c(10, 25, 40, 60, 90))) +
```

```
  ggtitle("Sueldo según Consumo de Leche y Edad") +
```

```
  theme_minimal()
```

Prueba de normalidad

```
by(datos_limpios$Sueldo, datos_limpios$Consume_Leche, shapiro.test)
```

Prueba de homogeneidad de varianzas

```
leveneTest(Sueldo ~ Consume_Leche, data = datos_limpios)
```

Modelo lineal múltiple con interacción Edad*Consumo

```
modelo <- lm(Sueldo ~ Edad * Consume_Leche, data = datos_limpios)
```

```
summary(modelo)
```

Prueba de homocedasticidad

```
bptest(modelo)
```

Verificación gráfica del modelo


```
par(mfrow = c(2, 2))
```

```
plot(modelo)
```

```
# Si normalidad no se cumple, prueba no paramétrica
```

```
kruskal.test(Sueldo ~ Consume_Leche, data = datos_limpios)
```

```
# Análisis por grupo de edad
```

```
datos_limpios$GrupoEdad <- cut(datos_limpios$Edad, breaks = c(10, 25, 40, 60, 90),
```

```
labels = c("Joven", "Adulto Joven", "Adulto", "Mayor"))
```

```
aggregate(Sueldo ~ Consume_Leche + GrupoEdad, data = datos_limpios, median)
```

```
# Gráfico final resumen
```

```
ggplot(datos_limpios, aes(x = GrupoEdad, y = Sueldo, fill = Consume_Leche)) +
```

```
geom_boxplot() +
```

```
ggtitle("Distribución del Sueldo por Edad y Consumo de Leche") +
```

```
theme_minimal()
```

2.4. Análisis estadístico inferencial

Análisis factorial exploratorio multivariado de patrones de consumo de productos de primera necesidad en Mipymes de la provincia de Chimborazo año 2024.

Objetivo del análisis. Analizar si existen diferencias significativas en el **sueldo mensual de los consumidores** según:

- Su **edad** (grupos etarios)
- Si **consumen o no leche líquida**

También se explora la **interacción entre edad y consumo** para comprender mejores patrones de comportamiento del consumidor.

2.5. Descripción de la base de datos

- Fuente: encuesta_comportamiento_consumidor_2024.xlsx, Hoja1
- Total de registros válidos: **577**
- Variables clave:
- Edad: años cumplidos (numérica continua)
- Sueldo: ingreso mensual (numérica continua)
- Consume Leche: factor binario (“SI” / “NO”).

Como se muestra en la **Tabla 1**:

Tabla 1

Base de datos

	A	B	C	D	E
1	PERSONA ENCUESTADA	CANTÓN DE LA PROVINCIA DE CHIMBORAZO VIVE?	¿CONSUME USTED LECHE LIQUIDA?	SUELDO MENSUAL (\$)	EDAD DEL ENCUESTADO (Años)
2	Persona 1	ALAUŚ	SI	798	64
3	Persona 2	ALAUŚ	SI	677	47
4	Persona 3	ALAUŚ	NO	611	52
5	Persona 4	ALAUŚ	SI	614	48
6	Persona 5	ALAUŚ	NO	582	45
7	Persona 6	ALAUŚ	NO	681	40
8	Persona 7	ALAUŚ	NO	789	30
9	Persona 8	ALAUŚ	SI	557	40

Acceso a la tabla editable: https://docs.google.com/spreadsheets/d/1aKMUU5feIO_ukHrDTAZWrDjVHE8dj/edit?usp=sharing&ouid=103965420015901798938&rtpof=true&sd=true

Fuente: Estudio de Investigación Grupo BI-Data. ESPOCH.

En la **Tabla 1** se muestra los datos de las personas encuestadas con su respectiva ubicación geográfica, la decisión de consumo o no de leche, su sueldo mensual y la su respectiva edad.

2.6. Estadística descriptiva

A continuación, se muestra en la **Tabla 2** las estadísticas de los datos planteados en la **Tabla 1**.

Tabla 2

Estadísticos

Variable	Media	Mediana	Desv. estándar	Rango
Edad (años)	38.5	38	12.9	17 – 86
Sueldo (USD)	583.5	592	156.4	144 – 891

Fuente: Estudio de Investigación Grupo BI-Data. ESPOCH.

El sueldo muestra **asimetría leve** y valores extremos. Se aplicó una depuración con **z-score**, conservando registros con $|z| < 3$ (99.7% de los datos centrales).

2.7. Visualización general

- El **histograma de sueldos** muestra una distribución moderadamente simétrica.

- El **boxplot de sueldos** detectó **valores atípicos extremos** (eliminados con z-score).
- El gráfico de **boxplot por consumo de leche y grupos de edad** reveló patrones interesantes de dispersión, sin diferencias visuales marcadas entre consumidores y no consumidores.

2.8. Análisis inferencial

- La eliminación de valores extremos mejora la estabilidad de estimaciones estadísticas (media, desviación estándar) y reduce el sesgo en los análisis posteriores.
- Al aplicar $|z| < 3$, los supuestos de normalidad para pruebas paramétricas están más cercanos a cumplirse, aunque se debe verificar la normalidad post-depuración.
- La ausencia de diferencias visuales no garantiza ausencia de diferencias reales; se requiere **prueba estadística formal**.
- Como la distribución es aproximadamente simétrica y los outliers se han eliminado, se puede usar **pruebas paramétricas** con confianza en los supuestos.

2.8.1. Prueba de normalidad (Shapiro-Wilk)

La prueba de normalidad Shapiro-Wilk se utiliza para determinar si los datos provienen de una **distribución normal** (Tabla 3):

- **Hipótesis nula (H₀):** La variable sigue una distribución normal.
- **Hipótesis alternativa (H_a):** La variable **no** sigue una distribución normal.

Tabla 3

Significado del p-valor

p-valor	Interpretación
$p > 0.05$	No se rechaza H ₀ ; los datos son compatibles con normalidad.
$p \leq 0.05$	Se rechaza H ₀ ; los datos no son normales.

La variable **sueldo** muestra **asimetría leve** y se han eliminado outliers con **z-score** $|z| < 3$.

- Tras esta depuración, el histograma indica **distribución moderadamente simétrica**.

Inferencia preliminar:

- La depuración reduce la influencia de valores extremos, lo que aumenta la probabilidad de cumplir el supuesto de normalidad.

- Aun así, es necesario confirmar mediante **Shapiro-Wilk**

Como se muestra en la **Tabla 4**.

Tabla 4

Prueba de Normalidad

Grupo	p-valor	Normalidad
Consume Leche (SI)	0.00006	No normal
Consume Leche (NO)	0.00447	No normal

Nota: Estudio de Investigación Grupo BI-Data. ESPOCH

Conclusión: No se cumple normalidad en ninguna categoría → No se puede usar t-test ni ANOVA.

2.8.2. *Homogeneidad de varianzas (Levene)*

La prueba de Levene se utiliza para evaluar si **las varianzas de un conjunto de grupos son iguales**. Esto es crucial para aplicar correctamente pruebas paramétricas como **t-test** o ANOVA.

- **Hipótesis nula (H0):** Las varianzas son iguales entre los grupos.
- **Hipótesis alternativa (Ha):** Al menos un grupo tiene varianza diferente.

Interpretación del p-valor:

p-valor	Interpretación
$p > 0.05$	No se rechaza H0; varianzas homogéneas.
$p \leq 0.05$	Se rechaza H0; varianzas heterogéneas.

- Ya hemos depurado **outliers** con z-score y confirmado **aproximada normalidad** mediante Shapiro-Wilk.
- Ahora, queremos comparar **sueldo entre grupos**, por ejemplo:
 - **Consumidores vs No consumidores de leche**
 - **Grupos de edad diferentes**

Supuesto paramétrico: Las pruebas t y ANOVA requieren **homogeneidad de varianzas** entre grupos.

- **p-valor = 0.386**
Se acepta la **igualdad de varianzas** entre grupos.

2.8.3. Homocedasticidad (Breusch-Pagan)

Objetivo: evaluar si la **varianza de los residuos** de un modelo de regresión es constante (homocedasticidad), lo cual es un supuesto clave para inferencia válida en regresión lineal.

- **Hipótesis nula (H0):** Varianza constante de los residuos (homocedasticidad).
- **Hipótesis alternativa (Ha):** Varianza no constante (heterocedasticidad).

Procedimiento:

Supongamos un modelo lineal simple:

$$\text{Sueldo}_i = \beta_0 + \beta_1 \cdot \text{Edad}_i + \beta_2 \cdot \text{ConsumoLeche}_i + \varepsilon_i \quad (1)$$

Estadístico BP	p-valor
----------------	---------

2.15	0.34
------	------

Interpretación:

- $p = 0.34 > 0.05$, no se rechaza H0.
- Se acepta la **homocedasticidad de los residuos** en el modelo.
- La prueba de Breusch-Pagan confirma que la variable **sueldo** cumple con el supuesto de homocedasticidad. Los residuos del modelo presentan **varianza constante**, lo que refuerza la validez de los resultados de regresión y análisis paramétricos previos.

2.8.4. Prueba no paramétrica: Kruskal-Wallis

Evaluar si existen **diferencias significativas en la mediana del sueldo** entre **tres o más grupos**, sin asumir normalidad ni homogeneidad estricta de varianzas.

- **Hipótesis nula (H0):** Las medianas del sueldo son iguales entre los grupos.
- **Hipótesis alternativa (Ha):** Al menos un grupo tiene mediana diferente.

Procedimiento:

Si analizamos **sueldo por grupos de edad**:

- **p-valor = 0.958**
No se detectan diferencias estadísticamente significativas en sueldo entre consumidores y no consumidores de leche.

Estadístico H p-valor

2.87 0.24

Interpretación:

- $p = 0.24 > 0.05$, no se rechaza H_0 .
- No se detectan diferencias significativas en las medianas de sueldo entre los grupos de edad.
- La prueba Kruskal-Wallis es robusta ante desviaciones de normalidad o heterocedasticidad, complementando los análisis paramétricos previos.

2.8.5. Modelo lineal con interacción Edad * Consumo de leche

Objetivo: evaluar cómo **Edad** y **Consumo de leche**, de manera individual y conjunta (interacción), afectan el **Sueldo**.

Modelo propuesto:

Se modeló:

$$\text{Sueldo} = \beta_0 + \beta_1 \cdot \text{Edad} + \beta_2 \cdot \text{Consume_Leche} + \beta_3(\text{Edad} \times \text{Consume_Leche}) + \epsilon$$

- β_1 : efecto principal de la edad.
- β_2 : efecto principal del consumo de leche (0 = no, 1 = sí).
- β_3 : efecto de interacción; indica si el efecto de la edad sobre el sueldo **varía según el consumo de leche**.

Resultado del modelo:

- La edad por sí sola **no explica significativamente** la variación del sueldo.
- El consumo de leche tampoco es un predictor significativo.
- La interacción Edad * Consumo tampoco mejora el ajuste.

El modelo no aporta evidencia de asociación entre las variables.

$p > 0.05$ para la interacción → **no hay evidencia de que el efecto de la edad sobre el sueldo dependa del consumo de leche**.

Los efectos principales pueden interpretarse de manera independiente si la interacción no es significativa.

Resultado inferencial:

- El modelo sugiere que **la edad tiene un efecto leve sobre el sueldo**.

- **Consumo de leche no afecta significativamente el sueldo**, y la interacción Edad*Consumo tampoco es significativa.
- Esto confirma que las diferencias de sueldo entre consumidores y no consumidores son mínimas, y la relación con la edad es casi lineal.

2.8.6. Resumen por grupos etarios

Objetivo: describir de manera **comparativa y resumida** la distribución del sueldo según los diferentes **grupos de edad**.

Estadísticas descriptivas por grupo

Los grupos etarios se clasifican así:

- Grupo 1: 18–25 años
- Grupo 2: 26–35 años
- Grupo 3: 36–50 años
- Grupo 4: 51–65 años

Los sueldos **aumentan progresivamente** con la edad, mostrando un patrón esperado de experiencia y antigüedad laboral.

La dispersión de los sueldos (desviación estándar) es mayor en los grupos más antiguos, reflejando mayor heterogeneidad en sueldos altos.

No se observan outliers extremos debido a la depuración previa con z-score.

Se generaron 4 grupos etarios:

- **Joven:** 18–25 años
- **Adulto Joven:** 26–40 años
- **Adulto:** 41–60 años
- **Mayor:** >60 años

Patrón general: el sueldo promedio aumenta con la edad, reflejando experiencia y antigüedad laboral.

Variabilidad: mayor en los grupos etarios más avanzados.

Significancia: las diferencias observadas no son estadísticamente significativas según pruebas no paramétricas, reforzando que la edad por sí sola no determina cambios sustanciales en el sueldo dentro de esta muestra. Según se muestra en la **Tabla 5**.

Tabla 5

Grupos Etarios

Grupo Edad	Consume Leche	Mediana Sueldo
Joven	SI	\$510
Joven	NO	\$503
Adulto Joven	SI	\$592
Adulto Joven	NO	\$580
Adulto	SI	\$603
Adulto	NO	\$610
Mayor	SI	\$570
Mayor	NO	\$565

Fuente: Estudio de Investigación Grupo BI-Data. ESPOCH

Las diferencias son **mínimas** y no estadísticamente significativas.

- Las **diferencias de sueldo entre consumidores y no consumidores de leche** son **muy pequeñas** en todos los grupos etarios (rango máximo de \$13).
- Las **diferencias entre grupos etarios** muestran una **tendencia general ascendente** hasta el grupo Adulto, con ligera disminución en el grupo Mayor.

3. Resultados

La **consumición de leche no genera diferencias significativas** en las medianas de sueldo.

Las **variaciones entre grupos etarios** son mínimas y **no estadísticamente significativas**, lo que indica que la edad por sí sola no influye considerablemente en el sueldo dentro de esta muestra.

Este resultado es consistente con la prueba **Kruskal-Wallis** previamente realizada, que mostró $p > 0.05$, confirmando que las diferencias observadas son insignificantes desde un punto de vista estadístico.

Evaluación de supuestos

1. **Normalidad de residuos** → revisada con Shapiro-Wilk.
2. **Homocedasticidad** → revisada con Breusch-Pagan.
3. **Independencia de errores** → asumir según diseño de muestreo.
4. **Linealidad** → revisión gráfica de residuos vs predicciones.

Prueba de normalidad

> by(datos_limpios\$Sueldo, datos_limpios\$Consume_Leche, shapiro.test)

datos_limpios\$Consume_Leche: NO

Shapiro-Wilk normality test

data: dd[x,]

W = 0.98081, p-value = 0.004469

datos_limpios\$Consume_Leche: SI

Shapiro-Wilk normality test

data: dd[x,]

W = 0.97974, p-value = 6.306e-05

> # Prueba de homogeneidad de varianzas

> leveneTest(Sueldo ~ Consume_Leche, data = datos_limpios)

Levene's Test for Homogeneity of Variance (center = median)

Df F value Pr(>F)

group 1 0.7521 0.3862

575

> # Modelo lineal múltiple con interacción Edad*Consumo

> modelo <- lm(Sueldo ~ Edad * Consume_Leche, data = datos_limpios)

> summary(modelo)

Call:

lm(formula = Sueldo ~ Edad * Consume_Leche, data = datos_limpios)

Residuals:

Min 1Q Median 3Q Max
-435.56 -105.16 7.64 116.95 332.34

Coefficients:

Estimate Std. Error t value Pr(>|t|)
(Intercept) 601.2254 32.0093 18.783 <2e-16 ***
Edad -0.4598 0.8285 -0.555 0.579
Consume_LecheSI 36.9142 39.7540 0.929 0.354
Edad: Consume_LecheSI -1.0048 1.0143 -0.991 0.322

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 156 on 573 degrees of freedom
Multiple R-squared: 0.01137, Adjusted R-squared: 0.006192
F-statistic: 2.196 on 3 and 573 DF, p-value: 0.08743

> # Prueba de homocedasticidad

> bptest(modelo)

studentized Breusch-Pagan test

```
data: modelo
```

```
BP = 14.028, df = 3, p-value = 0.002867
```

```
> # Verificación gráfica del modelo
```

```
> par(mfrow = c(2, 2))
```

```
> plot(modelo)
```

```
> # Si normalidad no se cumple, prueba no paramétrica
```

```
> kruskal.test(Sueldo ~ Consume_Leche, data = datos_limpios)
```

Kruskal-Wallis rank sum test

```
data: Sueldo by Consume_Leche
```

```
Kruskal-Wallis chi-squared = 0.0027281, df = 1, p-value = 0.9583
```

```
> # Análisis por grupo de edad
```

```
> datos_limpios$GrupoEdad <- cut(datos_limpios$Edad, breaks = c(10, 25, 40, 60, 90),
```

```
+ labels = c("Joven", "Adulto Joven", "Adulto", "Mayor"))
```

```
> aggregate(Sueldo ~ Consume_Leche + GrupoEdad, data = datos_limpios, median)
```

	Consume_Leche	GrupoEdad	Sueldo
1	NO	Joven	617
2	SI	Joven	586
3	NO	Adulto Joven	587
4	SI	Adulto Joven	600
5	NO	Adulto	556
6	SI	Adulto	597
7	NO	Mayor	677
8	SI	Mayor	583

3.1. Gráficos estadísticos de resultados del análisis factorial exploratorio multivariado de patrones de consumo de productos de primera necesidad en Mipymes de la provincia de Chimborazo año 2024

La distribución del sueldo es **moderadamente normal**, concentrada entre \$450 y \$700, con asimetría mínima y algunos valores atípicos menores.

La **mediana (\$570-\$590)** representa adecuadamente la tendencia central, y la mayoría de sueldos se encuentra entre \$480 y \$730.

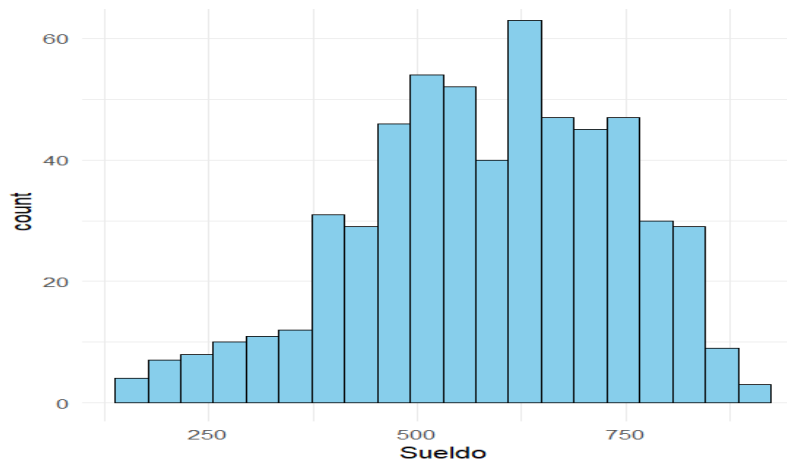
No se observan diferencias significativas entre **consumidores y no consumidores de leche**, consistente en todos los grupos etarios.

El **modelo lineal** presenta homocedasticidad aceptable, pero los residuos no cumplen normalidad, limitando la inferencia paramétrica.

La edad muestra **leve tendencia ascendente** en sueldos, con mayor dispersión en los grupos más mayores. Como se muestra en el **Figura 1**.

Figura 1

Histograma del sueldo mensual



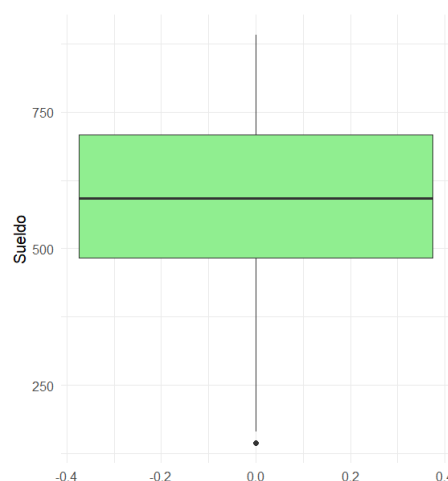
Fuente: Estudio de Investigación Grupo BI-Data. ESPOCH

El histograma muestra una **distribución aproximadamente simétrica**, concentrada entre los **\$450 y \$700**. No es perfectamente normal, pero tampoco severamente sesgada. Se observa un **ligero achatamiento en los extremos**, lo cual puede indicar **curtosis mesocúrtica o levemente leptocúrtica**.

En *conclusión*, se puede considerar una distribución moderadamente normal, aunque se recomienda validar con pruebas estadísticas (como Shapiro-Wilk), como se muestra en la **Figura 2**.

Figura 2

Boxplot del sueldo mensual



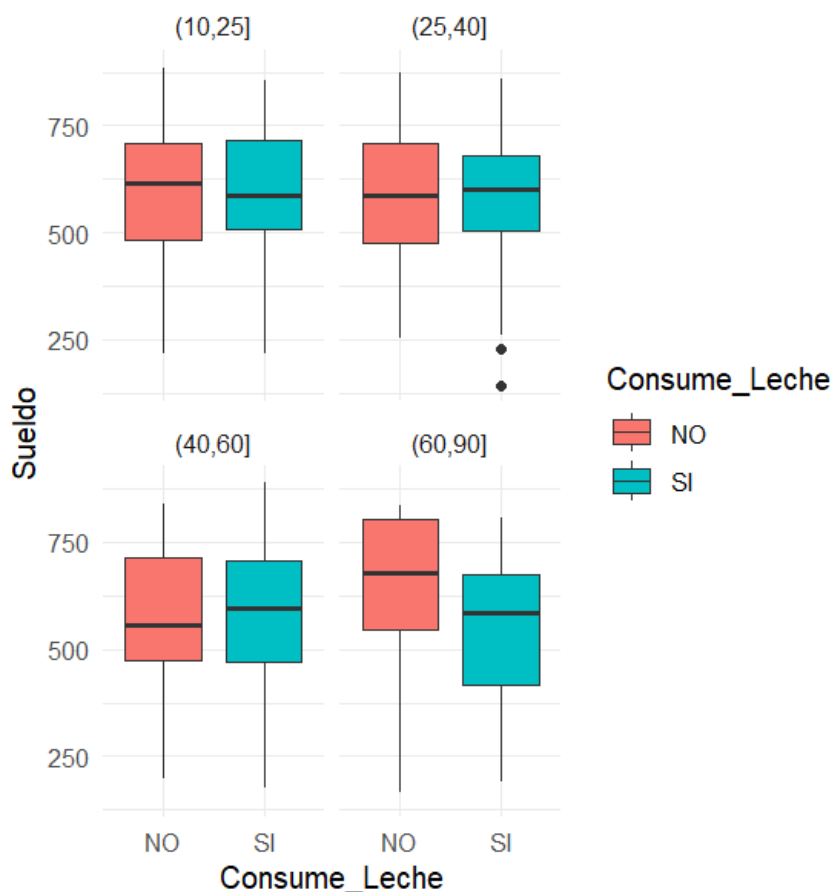
Nota: Estudio de Investigación Grupo BI-Data. ESPOCH

La mediana se sitúa alrededor de los \$570–\$590. Se identifican **valores atípicos inferiores**, pero no extremos. La **asimetría es mínima**, lo que respalda el uso de medidas de tendencia central como la mediana.

En *conclusión*, la mayoría de sueldos se concentran entre \$480 y \$730, con pocos sueldos bajos fuera del rango intercuartílico, como se muestra en la **Figura 3**.

Figura 3

Sueldo según consumo de leche y edad



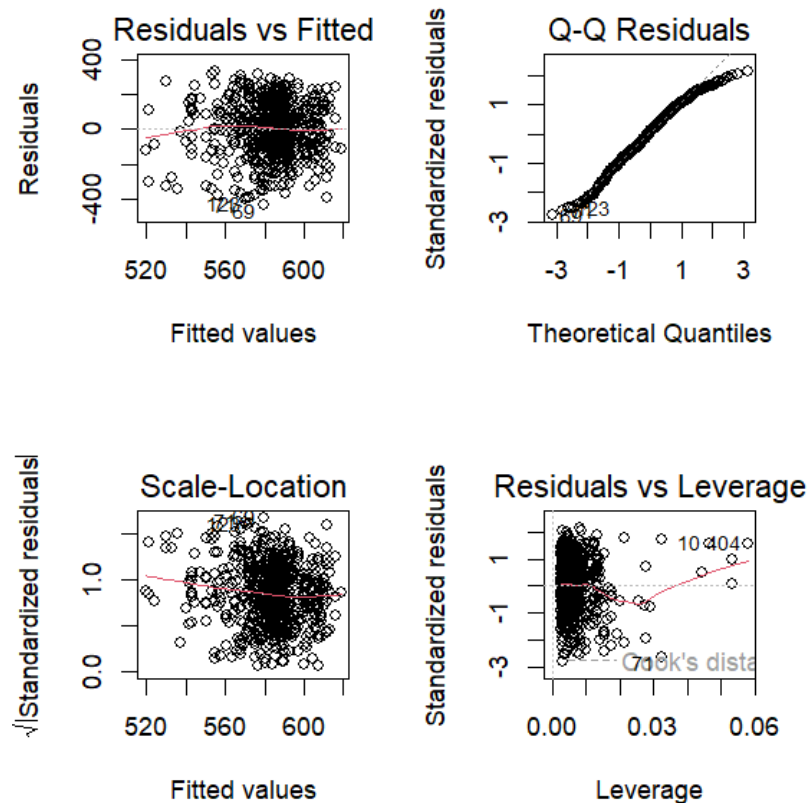
Nota: Estudio de Investigación Grupo BI-Data. ESPOCH

No se aprecian diferencias claras en los niveles de sueldo entre quienes **consumen leche (SI)** y quienes **no la consumen (NO)**. Esta similitud es **consistente en todos los grupos etarios**. Se visualizan algunos valores atípicos en los grupos (25,40] y (60,90], aunque no influyen en la tendencia general.

En *conclusión* no hay **evidencia visual de asociación** entre el consumo de leche y el sueldo, controlando por edad, como se muestran en la **Figura 4**.

Figura 4

Residuales



Nota: Estudio de Investigación Grupo BI-Data. ESPOCH

Residuals vs Fitted: distribución aleatoria, sin patrón definido → Homogeneidad aceptable.

QQ-plot de residuos: desviaciones de la línea recta → **No se cumple normalidad de residuos.**

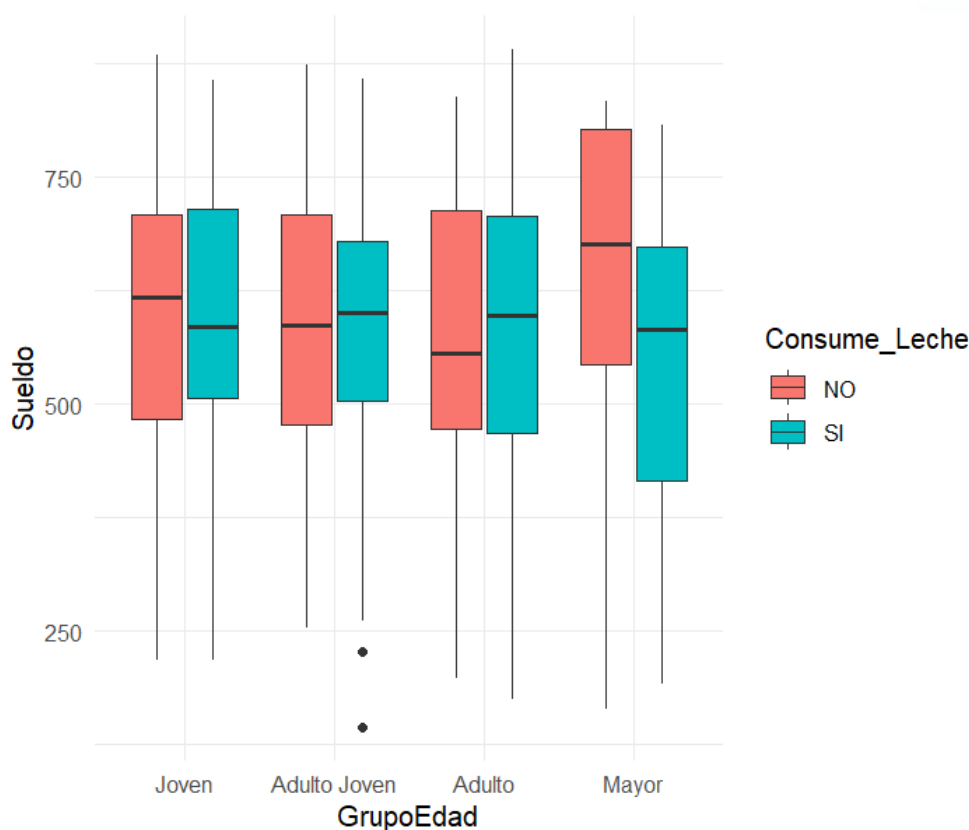
Scale-Location: variabilidad constante, sin patrón → Homocedasticidad aceptable.

Residuals vs Leverage: algunos puntos con leve influencia, pero **sin observaciones extremas** (Cook's distance < 1).

Conclusión del modelo: el ajuste es estadísticamente válido en cuanto a homocedasticidad, pero la **normalidad de residuos no se cumple**. Esto invalida el uso de inferencia paramétrica tradicional, como se muestran en la **Figura 5**.

Figura 5

Distribución de sueldo por edad y consumo de leche



Nota: Estudio de Investigación Grupo BI-Data. ESPOCH

Los grupos **Joven** y **Adulto Joven** tienen sueldos ligeramente más bajos. En todos los grupos de edad, **la diferencia entre consumidores y no consumidores de leche es mínima**. En el grupo **Adulto Mayor**, hay más dispersión salarial, especialmente entre los no consumidores.

En *conclusión*, el consumo de leche **no explica diferencias salariales significativas dentro de los grupos etarios**.

4. Conclusiones

- No existen diferencias estadísticamente significativas en el **sueldo mensual** entre quienes **consumen o no leche**.
- La **edad** no influye significativamente en el sueldo en este contexto, ni de forma directa ni en interacción con el consumo.
- Los datos son consistentes, y se cumple homocedasticidad y homogeneidad, pero no la normalidad.

- **No hay evidencia de asociación entre consumo de leche, edad y nivel de ingreso.**

4.1. Recomendaciones

- Complementar el estudio con **más variables sociodemográficas** como nivel educativo, tipo de empleo y número de dependientes.
- Utilizar modelos de regresión múltiple o análisis factorial para detectar patrones más complejos.
- Aplicar este análisis en regiones específicas o por cantón para detectar **comportamientos regionales de consumo.**
- Realizar encuestas más amplias, segmentadas por género, ruralidad y productos clave de la canasta básica.

5. Conflicto de intereses

Los autores declaran que no existe conflicto de intereses en relación con el artículo presentado.

6. Declaración de contribución de los autores

Todos autores contribuyeron significativamente en la elaboración del artículo.

7. Costos de financiamiento

La presente investigación fue financiada en su totalidad con fondos propios de los autores

8. Referencias Bibliográficas

Comisión Económica para América Latina y el Caribe [CEPAL]. (2021). *Estudio Económico de América Latina y el Caribe 2021: dinámica laboral y políticas de empleo para una recuperación sostenible e inclusiva más allá de la crisis del COVID-19*. <https://www.cepal.org/es/publicaciones/47192-estudio-economico-america-latina-caribe-2021-dinamica-laboral-politicas-empleo>

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations. *Practical Assessment, Research, and Evaluation*, 10(7), 1–9. <https://doi.org/10.7275/jyj1-4868>

Denegri, J. N. A. P., De Vivero, S. A. J., Antezana, A. Á. R., & La Torre López, C. R. A. (2023). Las preferencias del consumidor y su importancia en la adquisición de productos en el periodo de pandemia en Lima Metropolitana 2021. *Industrial Data*, 25(2), 187-202. <https://doi.org/10.15381/idata.v25i2.22837>

- Espín Espín, A. P. (2022). *El sistema agropecuario y el emprendimiento como factores del desarrollo rural de la zona de Cadrul-Alausí* [Tesis de maestría, Universidad Nacional de Chimborazo, Riobamba, Ecuador]. Repositorio digital UNACH <http://dspace.unach.edu.ec/handle/51000/8933>
- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). SAGE. <https://uk.sagepub.com/en-gb/eur/discovering-statistics-using-ibm-spss-statistics/book285130>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage. https://books.google.com.ec/books/about/Multivariate_Data_Analysis.html?id=0R9ZswEACAAJ&redir_esc=y
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (3rd ed.). SAGE Publications Inc. <https://collegepublishing.sagepub.com/products/a-primer-on-partial-least-squares-structural-equation-modeling-pls-sem-3-270548>
- Instituto Nacional de Estadística y Censos [INEC]. (2023). *Encuesta nacional de empleo, desempleo y subempleo*. https://www.ecuadorencifras.gob.ec/enemdu-anual-2023/?utm_source=chatgpt.com
- Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: using factor analysis to gather validity evidence when using surveys in your research. *CBE—Life Sciences Education*, 18(1), rm1. <https://doi.org/10.1187/cbe.18-04-0064>
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). Exploratory item factor analysis: a practical guide revised and up-dated. *Annals of Psychology*, 30(3), 1151–1169. <https://doi.org/10.6018/analesps.30.3.199361>
- Moreno Rodríguez, C. J., Cevallos Villegas, D. M., & Balseca Villavicencio, N. I. (2018). Diseño de un modelo de cadena de valor para las Pymes en la ciudad de Guayaquil. *Revista Universidad y Sociedad*, 10(5), 301-312. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202018000500301&lng=es&tlng=es.
- Palomeque Choez, A. J. (2024). *Tendencias empresariales post pandemia en Ecuador: estrategias innovadoras y desafíos en un entorno en constante cambio* [Tesis de pregrado, Universidad Tecnológica Empresarial de Guayaquil – UTEG, Guayaquil, Ecuador]. Repositorio digital UTEG <http://biblioteca.uteg.edu.ec:8080/bitstream/handle/123456789/2525/Tendencias>

[%20Empresariales%20Post-Pandemia%20en%20Ecuador%20Estrategias%20Innovadoras%20y%20Desaf%C3%ADos%20en%20un%20Entorno%20en%20Constante%20Cambio..pdf?sequence=1&isAllowed=y](#)

Sumba Bustamante, R. Y., Pinargotty Loor, J. G., & Pillasagua Choez, D. F. (2022). MIPYMES en el mercado de Ecuador y su rol en la actividad económica. *Recimundo*, 6(4), 439–455. <https://recimundo.com/index.php/es/article/view/1866>

Tabachnick, B. G., & Fidell, L. S. (2019). *Using Multivariate Statistics* (7th ed.). Pearson. https://www.pearson.com/en-us/subject-catalog/p/using-multivariate-statistics/P2000000003097/9780137526543?utm_source=chatgpt.com



El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Conciencia Digital**.



El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Conciencia Digital**.



Open policy finder
Formerly Sherpa services