

Tabla de contenido

1. Objetivo de la unidad 1.....	pág 2
2. Tema 3: Análisis bidimensionales.....	pág 2
2.1 Diagramas de dispersión.....	pág 2
2.2 ¿Qué es la correlación?.....	pág 3
2.3 Recta de regresión.....	pág 6
2.4 Tablas de contingencia.....	pág 10
2.5 Autoevaluación.....	pág 11
3. Bibliografía.....	pág 11

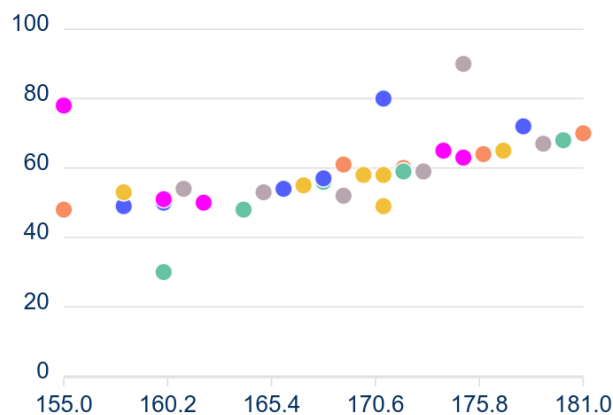
1. Objetivo de la unidad 1

Apropiar conceptos básicos de la Estadística Descriptiva, para realizar análisis estadístico más sofisticados que facilite la toma de decisiones basadas en datos.

2. Tema 3: Análisis bidimensionales

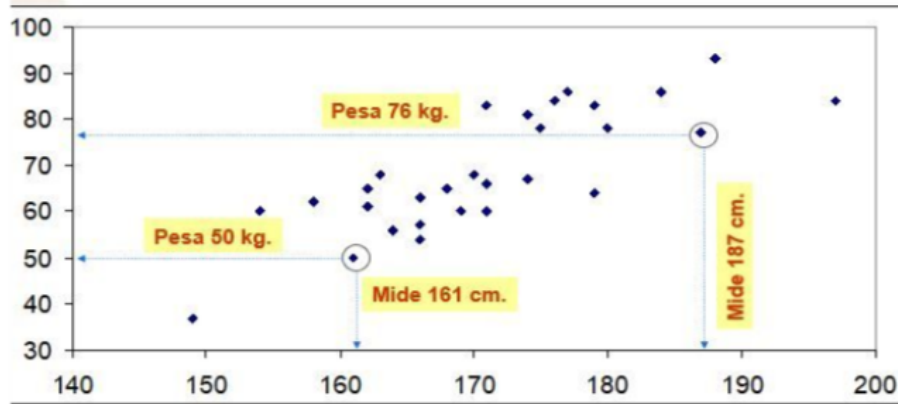
2.1 Diagramas de dispersión

Los gráficos de dispersión representan puntos en un sistema de coordenadas cartesianas para visualizar los valores de dos variables. Al analizar los patrones en estos gráficos, es posible interpretar diferentes tipos de correlación. Esta correlación puede ser positiva, negativa o nula. Además, los gráficos de dispersión permiten evaluar la intensidad de la correlación entre las variables, determinando si es fuerte, débil o inexistente.



Ejemplo:

En la imagen siguiente tenemos una posible manera de identificar la relación entre dos variables Estatura (x) y Peso (y) de los individuos.

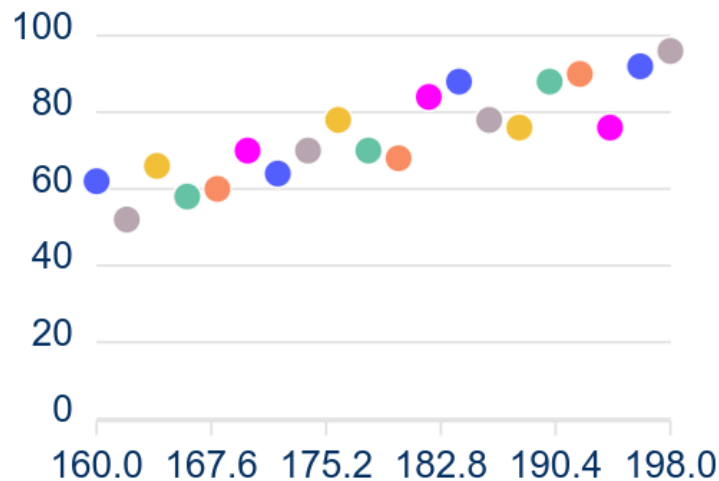


2.2 ¿Qué es la correlación?

La correlación es una medida que indica el grado y la dirección de la relación lineal entre dos variables cuantitativas. Se utiliza para determinar si, y en qué medida, las variaciones en una variable están asociadas con las variaciones en otra variable.

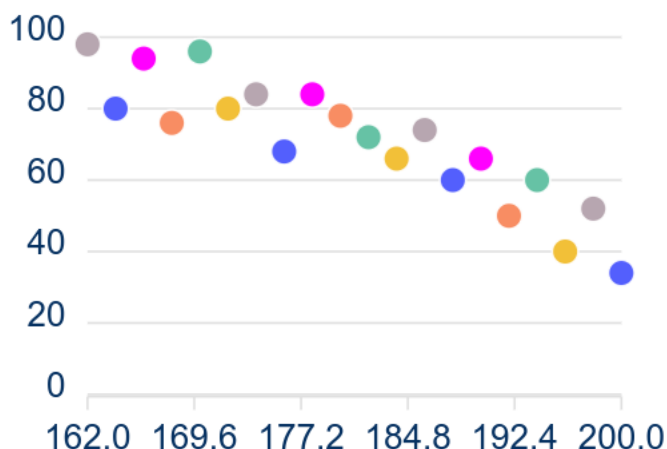
Correlación positiva

La correlación positiva se refiere a una relación directa entre dos variables, lo que significa que cuando una variable aumenta, la otra variable también tiende a aumentar. De manera similar, cuando una variable disminuye, la otra variable tiende a disminuir.



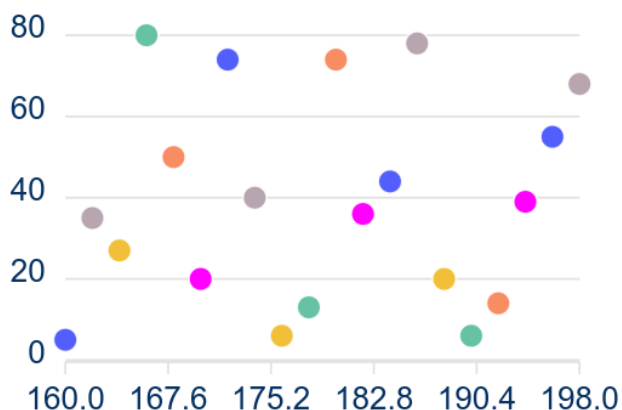
Correlación negativa

La correlación negativa se refiere a una relación inversa entre dos variables, lo que significa que cuando una variable aumenta, la otra variable tiende a disminuir. De manera similar, cuando una variable disminuye, la otra variable tiende a aumentar.



Correlación nula

La correlación nula, también conocida como correlación cero, se refiere a la ausencia de una relación lineal entre dos variables. Esto significa que no hay una relación directa ni inversa entre las dos variables; los cambios en una variable no están asociados con cambios consistentes en la otra variable.



Coefficiente de correlación lineal de Pearson

El coeficiente de correlación lineal de Pearson, R , es una herramienta estadística ampliamente utilizada para evaluar la fuerza y dirección de la relación lineal entre dos variables cuantitativas.

Este coeficiente varía entre -1 y 1.

- Un valor de 1 indica una correlación positiva perfecta.
- Un valor de -1 indica una correlación negativa perfecta.
- Un valor de 0 indica una correlación nula o la ausencia de una relación lineal.

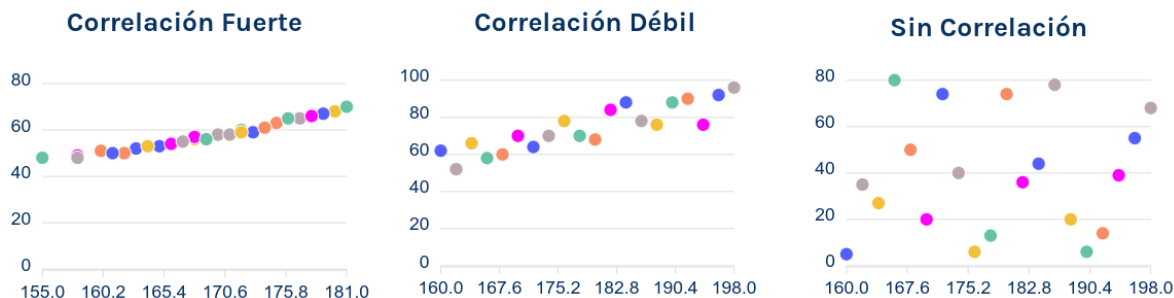
El coeficiente de correlación lineal de Pearson sirve para cuantificar tendencias lineales. La correlación nos permite medir el signo y magnitud de la tendencia entre dos variables:

Interpretación	Valores R (+)	Valores R (-)
Correlación Fuerte	$0.75 < R < 1$	$-1 < R < -0.75$
Correlación moderada	$0.5 < R < 0.75$	$-0.75 < R < -0.5$
Correlación débil	$0.15 < R < 0.5$	$-0.5 < R < -0.15$
Correlación nula	$0 < R < 0.15$	$-0.15 < R < 0$

- **Importante:** Es importante tener en cuenta que la correlación no implica causalidad. Es decir, solo porque dos variables estén correlacionadas, no significa que un cambio en una variable cause un cambio en la otra variable. El coeficiente de correlación de Pearson es útil en estadísticas, investigación, finanzas, medicina y muchas otras áreas para entender y cuantificar las relaciones entre variables.

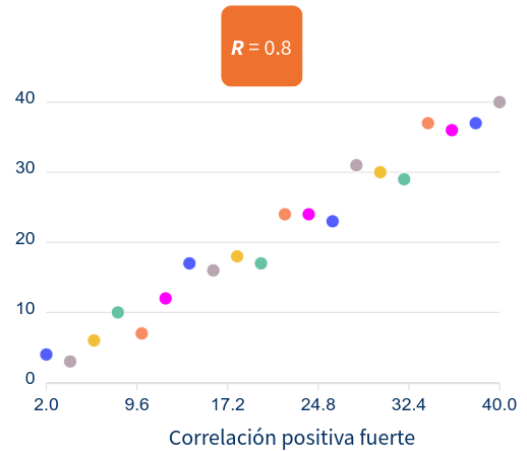
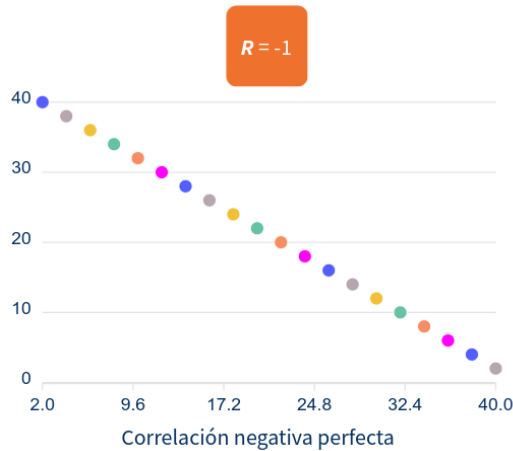
Ejemplo:

Los siguientes gráficos ilustran como lucen las correlaciones fuertes, débiles y sin correlación.



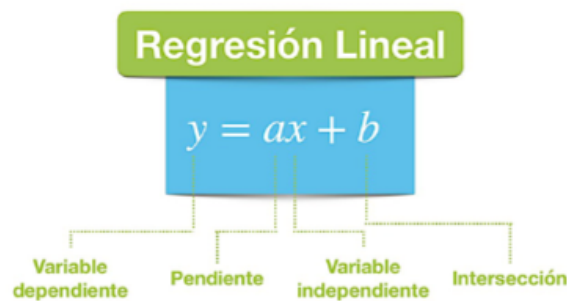
Ejemplo:

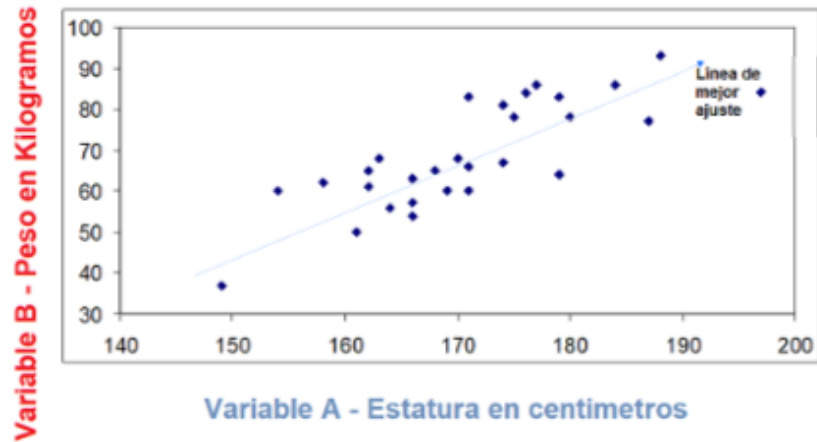
Los siguientes gráficos ilustran como lucen las correlaciones negativa perfecta y positiva fuerte.



2.3 Recta de regresión

La regresión es un método de análisis de datos que predice el valor de datos no conocidos basándose en otro valor de datos que sí se conoce y está relacionado con ellos. Este proceso implica la creación de un modelo matemático que describe la relación entre la variable desconocida (dependiente) y la variable conocida (independiente). Por lo general, se emplea un modelo lineal para esta tarea, aunque en función de la distribución de los datos, también pueden utilizarse modelos cuadráticos o cúbicos.



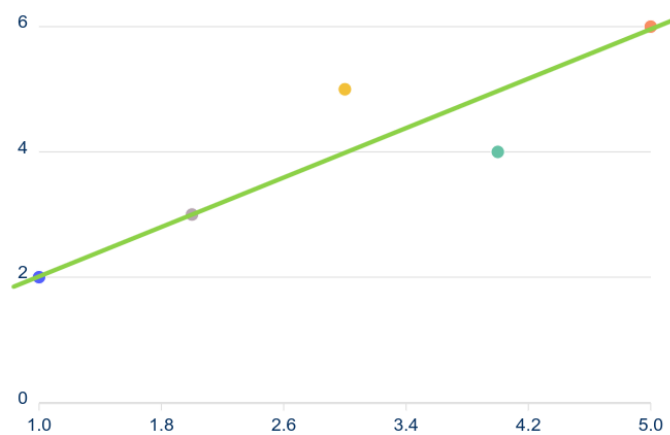


Ejemplo:

Supongamos que tenemos los siguientes datos de estudio, que representan las horas estudiadas por un estudiante (X) y sus respectivas calificaciones (Y):

Horas estudiadas (X)	Calificación (Y)
1	2
2	3
3	5
4	4
5	6

La imagen muestra el gráfico de dispersión junto con la recta de regresión que ha sido calculada como $Y=X+1$.



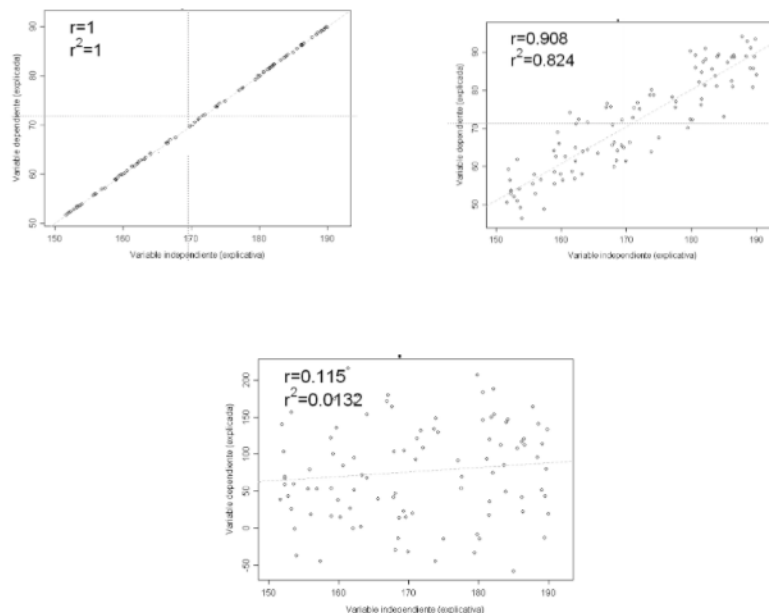
Coeficiente de determinación

El coeficiente de determinación, llamado R^2 es una medida estadística que proporciona información sobre la proporción de la varianza en la variable dependiente que es predecible a partir de la variable independiente en un modelo de regresión. En otras palabras, muestra qué parte de la variación total de la variable dependiente puede explicarse mediante el modelo de regresión.

Un R^2 cercano a 1 indica que una gran proporción de la variabilidad de la variable dependiente está explicada por el modelo de regresión. En general, un R^2 alto sugiere un buen ajuste del modelo a los datos.

Un R^2 cercano a 0 indica que el modelo no es adecuado para explicar la variabilidad de la variable dependiente. Esto puede ser debido a que el modelo es incorrecto o porque la relación entre las variables es no lineal.

El coeficiente de determinación no indica la dirección o la causa de la relación. Un R^2 alto no garantiza un buen modelo, ya que un modelo demasiado complejo puede sobre ajustarse a los datos y no ser generalizable a nuevos datos.



Ejemplo:

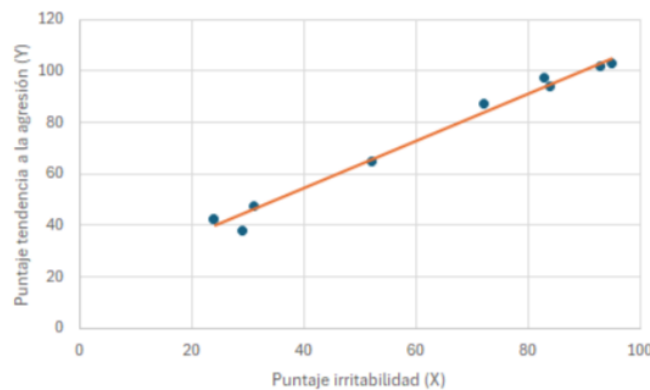
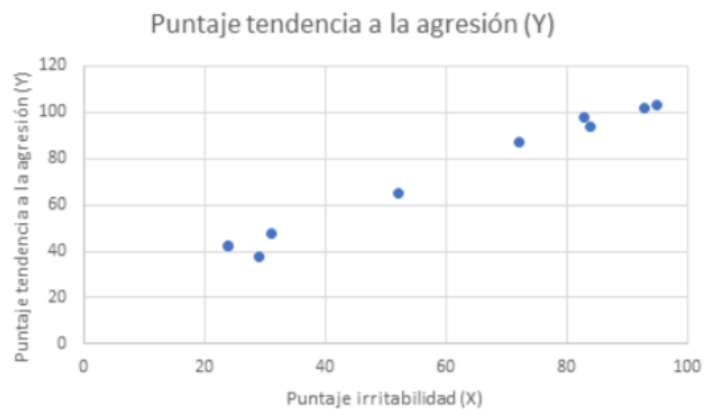
Un grupo de investigadores está interesado en explicar la relación existente entre la irritabilidad (X), y la tendencia a agredir a una persona (Y), para ello se tomó una muestra aleatoria de 10 individuos,

el siguiente grafico muestra la relación entre la irritabilidad y la tendencia a la agresión.

Cada punto es un dato y se observa que la relación es positiva y se ajusta a una línea recta.

Calculando el coeficiente de correlación de Pearson, se obtiene:

0.993680029. Lo que indica que tiene una alta correlación positiva, que indicará que podemos proseguir con el modelado matemático de tipo lineal. Usando el método de mínimos cuadrados se obtiene la recta del gráfico.



La ecuación de la recta regresión es la siguiente: **$y=0.9152x+18.235$** , y su coeficiente de determinación R^2 es: **0.9874**. Por lo tanto, el modelo lineal tiene una eficacia de aproximadamente el **98.7%**.

2.4 Tablas de continencia

Una tabla de contingencia o tabla cruzada permite mostrar la frecuencia con la que ocurren combinaciones específicas de categorías para las variables en estudio, lo que permite analizar las relaciones entre ellas. En una tabla de contingencia, las variables se colocan en los encabezados de las filas y columnas, y en las celdas se presenta el recuento de observaciones que pertenecen a la combinación de categorías correspondiente a esa fila y columna.

Genero \ Aprobación	Aprobado	Reprobado
Masculino (M)	50	15
Femenino (F)	30	20

Ejemplo:

Se realiza un estudio sobre el rendimiento académico de los estudiantes en una escuela, y se desea investigar si existe una relación entre el género del estudiante y su calificación final en un curso de programación:

Género \ Aprobación	Aprobado	Reprobado	Total
Masculino (M)	50	15	65
Femenino (F)	30	20	50
Total	80	35	115

Una de las interpretaciones de la tabla de contingencia podría ser la siguiente:

De acuerdo con los resultados de la tabla podemos visualizar que 50 estudiantes de género masculino han aprobado el curso de programación.

2.5 Autoevaluación

Pregunta 1:

El propósito principal de un diagrama de dispersión en estadística es representar la relación entre dos variables cuantitativas.

Opción 1: Verdadero

Opción 2: Falso

Pregunta 2:

Un diagrama de Dispersión describe una correlación negativa cuando los puntos forman una línea recta inclinada hacia arriba.

Opción 1: Verdadero

Opción 2: Falso

Pregunta 3:

¿Cuál es el objetivo principal de utilizar una recta de regresión en estadística?

Opción 1: Predecir valores futuros de una variable a partir de otra.

Opción 2: Calcular la media de un conjunto de datos.

Opción 3: Determinar la modificación de una variable por otra.

Opción 4: Identificar la moda de un conjunto de datos.

3. Bibliografía

- ✓ Romero Ramos, E. (2016). Estadística para todos: análisis de datos: estadística descriptiva, teoría de la probabilidad e inferencia. Difusora Larousse - Ediciones Pirámide, (pp. 22-30).
<https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/49136?page=1>
- ✓ Romero Ramos, E. (2016). Estadística para todos: análisis de datos: estadística descriptiva, teoría de la probabilidad e inferencia. Difusora Larousse - Ediciones Pirámide, (pp. 31-39).
<https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/49136?page=1>
- ✓ Romero Ramos, E. (2016). Estadística para todos: análisis de datos: estadística descriptiva, teoría de la probabilidad e inferencia. Difusora Larousse - Ediciones Pirámide, (pp. 41-55).

<https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/49136?page=1>

- ✓ Romero Ramos, E. (2016). Estadística para todos: análisis de datos: estadística descriptiva, teoría de la probabilidad e inferencia. Difusora Larousse - Ediciones Pirámide, (pp. 57-69).
<https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/49136?page=1>
- ✓ Romero Ramos, E. (2016). Estadística para todos: análisis de datos: estadística descriptiva, teoría de la probabilidad e inferencia. Difusora Larousse - Ediciones Pirámide, (pp. 93-100).
<https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/49136?page=1>
- ✓ Cely, L. A. (2020). Tipos de muestreo y error muestral. [Objeto_virtual_de_Informacion_OVI]. Repositorio Institucional UNAD. <https://repository.unad.edu.co/handle/10596/38474>
- ✓ Ortegon, M. F. & Cabrera, F. (2018). Tablas de Frecuencia. [Objeto_virtual_de_Informacion_OVI]. Repositorio Institucional UNAD. <https://repository.unad.edu.co/handle/10596/20458>
- ✓ Ortegon, M. F. (2019). Medidas de Tendencia Central. [Objeto_virtual_de_aprendizaje_OVA]. Repositorio Institucional UNAD. <https://repository.unad.edu.co/handle/10596/33842>
- ✓ Camargo, I. (2022). *Medidas estadísticas univariantes: Muestreo*. [Objeto_virtual_de_Informacion_OVI]. Repositorio Institucional UNAD.
<https://repository.unad.edu.co/handle/10596/59205>
- ✓ Camargo, I. (2022). Medidas univariantes para variables cuantitativas. [Objeto_virtual_de_aprendizaje_OVA]. Repositorio Institucional UNAD.
<https://repository.unad.edu.co/handle/10596/50133>
- ✓ Camargo, I. (2023). Muestreo y caracterización de variables cualitativas. [Objeto_virtual_de_aprendizaje_OVA]. Repositorio Institucional UNAD.
<https://repository.unad.edu.co/handle/10596/59206>