

Statistics in R – Modus Paraguayensis

Explorando la Estadística desde la Perspectiva de R

“Statistics in R – Modus Paraguayensis” consiste en una serie de cuadernos técnicos que condensan un esfuerzo sistemático por digitalizar y potenciar el análisis estadístico tradicional. Propone abordar una metodología de migración desde el tratamiento estadístico convencional hacia el lenguaje de programación R, priorizando la reproducibilidad científica y la interpretación de los datos en el contexto de la economía y las ciencias sociales.

Autor: Mag. Econ. Marcos Gómez Hermosa
Facultad de Ciencias Económicas (FACE)
Universidad Católica Campus Itapúa (UCI)
Encarnación, Paraguay – 2026

Análisis de un índice para la Articulación Académico-Profesional de estudiantes en situación de empleo

Un ejemplo de aplicación de análisis descriptivo estadístico de datos agrupados

Contents

1. Introducción	1
2. Disponibilidad de Código y Datos	3
3. Creación de Intervalos de frecuencia	3
4. Medidas de Tendencia Central	5
Media aritmética	5
Media geométrica	6
Mediana	6
Moda	7
5. Medidas de Posición	9
Cuartiles	9
Deciles	10
Percentiles	11
6. Medidas de Variabilidad	12
Varianza	12
Desviación Estándar	12
Coeficiente de Variación (CV)	13
8. Medidas de Asimetría y Curtosis	14
Asimetría de Fisher	14
Asimetría de Bowley	15
Asimetría de Pearson	15
Curtosis	16
9. Discusión de Resultados	20
10. Referencias	20

1. Introducción

Suponga que a partir cierto proceso de reducción de datos, por ejemplo “Análisis de Componentes Principales” (PCA), se logra sintetizar un conjunto diverso de variables (el Promedio Acumulado, Horas de Estudio, Créditos Aprobados e IQ Académico, junto con Horas de Capacitación, Complejidad de la Tarea, Salario Mensual y Evaluación de Desempeño) en dos constructos ortogonales: el Índice de Conocimiento General (ICG) y el Índice de Especialización (IE).

Y, como síntesis de este proceso, se define el indicador compuesto “Índice de Articulación Académico-Profesional” (IAAP). Este índice representa la convergencia entre la formación técnica y la eficiencia en el entorno productivo, obteniéndose mediante la ponderación de los constructos generados a través de la norma euclidiana:

$$IAAP = \sqrt{ICG^2 + IE^2}$$

Este indicador sintético actúa como facilitador para que el desarrollo estadístico se despliegue sobre una métrica unificada que integra de manera robusta las dimensiones académica y profesional, para estudiantes que se encuentran en situación de empleo u ocupados. Una magnitud superior en el IAAP es representativa de una articulación exitosa y un desempeño sobresaliente en ambos campos.

Entonces, es éste indicador, el “IAAP”, el que se considerará para realizar una caracterización morfológica de los datos, utilizando técnicas de estadística descriptiva para datos agrupados. Esta caracterización implica la síntesis de los datos mediante el cálculo de medidas de posición y dispersión, permitiendo una interpretación de la estructura del índice y su comportamiento dentro de los intervalos de clase establecidos. Complementariamente, se abordará el análisis de asimetría y curtosis para determinar la deformación horizontal y el grado de concentración central de la distribución, factores críticos para validar la representatividad del promedio obtenido.

La implementación técnica se realizó mediante el lenguaje de programación estadística R (R Core Team, 2026). Para la manipulación de los datos multivariados y la sistematización de frecuencias, se utilizaron las funciones de la librería dplyr (Wickham et al., 2023). Asimismo, la caracterización morfológica y la arquitectura visual de los datos, incluyendo polígonos de frecuencia e histogramas, se desarrollaron bajo la gramática de gráficos de ggplot2 (Wickham, 2016), siguiendo los principios de visualización y modelado de datos propuestos por Hadley y Grolemund (2017).

```
library(readr)
# link "Raw" al archivo de datos
url_raw <- paste0("https://raw.githubusercontent.com/",
                  "elprofemarcosgh/Statistics-in-R-Modus-Paraguayensis/",
                  "main/trabajo%231/datos_para_tabla_est.csv")

df_indices <- read_csv(url_raw)

df_indices <- df_indices %>%
  mutate(IAAP = sqrt(Indice_Capacidad_General^2 + Indice_Especializacion^2))
# head() para mostrar solo los primeros 10 registros
# select() para organizar las columnas de forma lógica para el lector
library(dplyr)
vista_previa <- df_indices %>%
  select(Indice_Capacidad_General, Indice_Especializacion, IAAP) %>%
  head(10)

knitr::kable(vista_previa,
              caption = "Muestra de los primeros 10 registros de la base de datos
↵ (Variables clave e Índices)",
              digits = 2)
```

Table 1: Muestra de los primeros 10 registros de la base de datos (Variables clave e Índices)

Indice_Capacidad_General	Indice_Especializacion	IAAP
-1.03	0.25	1.06
-0.14	1.24	1.25
-0.25	-0.48	0.54
-1.34	-0.01	1.34
-1.36	1.27	1.86
1.04	-0.02	1.04
-1.56	0.20	1.57
-0.58	0.31	0.66
-0.15	-1.35	1.36
-2.07	0.02	2.07

La Tabla 2 resume las características sociodemográficas y de aptitud de los 250 individuos de la muestra.

```
library(dplyr)
library(knitr)

df_esencial <- df_indices %>%
  select(promedio_acum, horas_estudio, creditos_aprob, iq_academico,
         horas_capacit, complej_tarea, salario_mensual, eval_desempeno,
         Indice_Capacidad_General, Indice_Especializacion, IAAP)

resumen_esencial <- data.frame(
  Variable = c(
    "Promedio Acumulado", "Horas de Estudio", "Créditos Aprobados",
    "IQ Académico", "Horas Capacitación", "Complejidad Tarea",
    "Salario Mensual", "Eval. Desempeño",
    "Índice Cap. General (ICG)", "Índice Especialización (IE)", "IAAP"
  ),
  Minimo = sapply(df_esencial, min, na.rm = TRUE),
  Maximo = sapply(df_esencial, max, na.rm = TRUE),
  Media = sapply(df_esencial, mean, na.rm = TRUE),
  n = sapply(df_esencial, function(x) sum(!is.na(x)))
)

# tabla
kable(resumen_esencial,
      digits = 2,
      row.names = FALSE,
      align = c("l", "c", "c", "c", "c"),
      caption = "Estadísticos Descriptivos Básicos",
      format.args = list(big.mark = ","))
```

Table 2: Estadísticos Descriptivos Básicos

Variable	Minimo	Maximo	Media	n
Promedio Acumulado	4.47	11.49	8.00	250
Horas de Estudio	7.59	32.26	20.00	250

Variable	Minimo	Maximo	Media	n
Créditos Aprobados	18.79	101.66	60.00	250
IQ Académico	74.43	124.94	100.00	250
Horas Capacitación	4.77	18.94	12.00	250
Complejidad Tarea	0.93	11.53	6.00	250
Salario Mensual	358.86	2,359.85	1,500.00	250
Eval. Desempeño	56.85	103.20	80.00	250
Índice Cap. General (ICG)	-2.60	2.87	0.00	250
Índice Especialización (IE)	-2.68	2.99	0.00	250
IAAP	0.05	3.19	1.28	250

2. Disponibilidad de Código y Datos

El código fuente para este análisis, incluyendo los scripts de procesamiento en R, las funciones personalizadas y la base de datos utilizada, está disponible en el repositorio oficial de “Statistics in R - Modus Paraguayensis”:

<https://github.com/elprofemarcosgh/Statistics-in-R-Modus-Paraguayensis>

Nota: Los materiales específicos de este capítulo se encuentran en la carpeta /trabajo#1

3. Creación de Intervalos de frecuencia

Se aplica la Regla de Sturges para organizar la información en intervalos de frecuencia. La Regla de Sturges es un criterio matemático utilizado en estadística para determinar el número óptimo de intervalos de clase (k) necesarios para agrupar un conjunto de datos en una tabla de frecuencias (Sturges, 1926). Este procedimiento de organización permite derivar una estructura objetiva para el análisis de la dispersión de la muestra (Anderson, Sweeney & Williams, 2019).

$$k = 1 + 3.322 \log_{10}(n)$$

- k : Es el número óptimo de intervalos o clases. Representa en cuántas partes vamos a dividir el rango total de la muestra.
- 1: Es una constante de la fórmula que asegura que, incluso con una muestra pequeña, exista al menos un intervalo.
- 3.322: Es un coeficiente fijo derivado de la expresión $1/\log_{10}(2)$, que asume una distribución aproximadamente normal de los datos.
- \log_{10} : Es el logaritmo en base 10, que ayuda a suavizar el crecimiento del número de intervalos a medida que la muestra aumenta.
- n : Es el tamaño total de la muestra (en este caso, el número total de estudiantes analizados).

```
##1. CONSTRUCCION DE INTERVALOS##
# Parámetros básicos
n_obs <- nrow(df_indices)
val_min <- min(df_indices$IAAP)
val_max <- max(df_indices$IAAP)
rango <- val_max - val_min
cat("Rango:\n",
    "Rango:", round(rango, 2), "\n")
```

```
## Rango:
## Rango: 3.14
```

```
# Aplicar Sturges para obtener el número de intervalos (k)
k <- ceiling(1 + 3.322 * log10(n_obs)) # ceiling redondea hacia arriba
cat("Cantidad de intervalos necesarios:\n",
    "Sturges:", round(k, 2), "\n")
```

```
## Cantidad de intervalos necesarios:
## Sturges: 9
```

```
# Calcular la amplitud (h)
amplitud <- rango / k
cat("Amplitud de intervalos:\n",
    "Amplitud:", round(amplitud, 2), "\n")
```

```
## Amplitud de intervalos:
## Amplitud: 0.35
```

```
# Generar los cortes (breaks) automáticos
cortes_sturges <- seq(val_min, val_max, by = amplitud)
cat("Cortes de intervalos:\n",
    "Cortes:", round(cortes_sturges, 2), "\n")
```

```
## Cortes de intervalos:
## Cortes: 0.05 0.4 0.75 1.1 1.45 1.8 2.14 2.49 2.84 3.19
```

```
library(dplyr)

# 1. cut() para agrupar el IAAP según los cortes de Sturges calculado
# include.lowest = TRUE asegura que el valor mínimo entre en el primer intervalo
df_agrupado <- df_indices %>%
  mutate(Intervalo_Sturges = cut(IAAP, breaks = cortes_sturges, include.lowest = TRUE))

# 2. Construcción de la tabla dinámica
tabla_completa <- df_agrupado %>%
  group_by(Intervalo_Sturges) %>%
  summarise(Frecuencia_Absoluta = n(), .groups = 'drop') %>%
  mutate(
    # El nombre del intervalo viene de R (ej. [0.05, 0.4])
    Intervalo = as.character(Intervalo_Sturges),

    # Marca de Clase: punto medio entre cada par de cortes
    Marca_Clase = (cortes_sturges[-length(cortes_sturges)] + cortes_sturges[-1]) / 2,

    # Frecuencias y porcentajes
    Frec_Acumulada = cumsum(Frecuencia_Absoluta),
    Porcentaje = (Frecuencia_Absoluta / sum(Frecuencia_Absoluta)) * 100,
    Porc_Acumulado = cumsum(Porcentaje)
  ) %>%
  # Seleccionamos y ordenamos las columnas para el reporte
```

```
select(Intervalo, Marca_Clase, Frecuencia_Absoluta, Frec_Acumulada, Porcentaje,
  ↪ Porc_Acumulado)

# 3. Mostrar tabla
knitr::kable(tabla_completa,
  digits = 2,
  caption = "Distribución de Frecuencias del IAAP basada en la Regla de
  ↪ Sturges (k=9)")
```

Table 3: Distribución de Frecuencias del IAAP basada en la Regla de Sturges (k=9)

Intervalo	Marca_Clase	Frecuencia_Absoluta	Frec_Acumulada	Porcentaje	Porc_Acumulado
[0.0536,0.402]	0.23	14	14	5.6	5.6
(0.402,0.75]	0.58	40	54	16.0	21.6
(0.75,1.1]	0.92	53	107	21.2	42.8
(1.1,1.45]	1.27	47	154	18.8	61.6
(1.45,1.8]	1.62	43	197	17.2	78.8
(1.8,2.14]	1.97	34	231	13.6	92.4
(2.14,2.49]	2.32	9	240	3.6	96.0
(2.49,2.84]	2.67	6	246	2.4	98.4
(2.84,3.19]	3.02	4	250	1.6	100.0

El proceso de organización determinó el Rango total de la métrica IAAP, el cual resultó ser de 3.14 unidades, estableciendo así la extensión completa de la dispersión de la muestra. Sobre esta base, se aplicó la Regla de Sturges para definir de manera objetiva la cantidad de categorías necesarias para el análisis, resultando en un óptimo de 9 intervalos de clase.

Esta estructura permitió derivar una Amplitud constante de 0.35, garantizando que cada segmento posea la misma sensibilidad analítica. Finalmente, este procedimiento culminó en la fijación de los Cortes de Intervalo (desde el límite inferior de 0.05 hasta el superior de 3.19).

4. Medidas de Tendencia Central

Media aritmética

Se define como el valor promedio obtenido al sumar todos los productos de las marcas de clase por sus frecuencias y dividirlo entre el número total de datos. Representa el valor que tendrían todos los individuos si la característica medida se repartiera de forma equitativa. Es sensible a valores extremos (datos atípicos) (Levin & Rubin, 2010).

$$\bar{x} = \frac{\sum_{i=1}^k f_i \cdot x_i}{n}$$

\bar{x} : Representa el estimador puntual de la media o promedio aritmético de la muestra.

$\sum_{i=1}^k$: Es el operador sumatorio que indica la adición de todos los productos resultantes desde el primer intervalo ($i = 1$) hasta el último (k).

f_i : Es la frecuencia absoluta del i -ésimo intervalo (el número de observaciones que caen dentro de esa categoría).

x_i : Representa la marca de clase del intervalo, que funciona como el valor representativo o punto medio de cada clase.

$f_i \cdot x_i$: Es el producto ponderado; equivale a sumar la marca de clase tantas veces como observaciones existan en dicho intervalo.

n : Es el tamaño total de la muestra, definido por la sumatoria de todas las frecuencias absolutas ($\sum f_i$).

Media geométrica

Se define como la raíz enésima del producto de todos los números. En el caso de datos agrupados, la fórmula utiliza la marca de clase (x_i) elevada a su frecuencia absoluta (f_i). Es la medida de tendencia central por excelencia para tasas de variación, porcentajes y números índice (Spiegel & Stephens, 2009):

$$MG = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_k^{f_k}}$$

n (El Índice): Es el tamaño total de la muestra

x_1, x_2, \dots, x_k : Representan las marcas de clase de cada intervalo

k : Es el número total de intervalos o renglones de la tabla de frecuencias

También se puede calcular de forma más sencilla usando logaritmos:

$$\log(MG) = \frac{\sum (f_i \cdot \log(x_i))}{n}$$

$\log(MG)$: Representa el logaritmo de la Media Geométrica. Una vez que obtienes el resultado de toda la operación de la derecha, se debe aplicar la función inversa (el antilogaritmo o función exponencial e^x) para despejar la MG y volver a la unidad de medida original.

f_i : Es la frecuencia absoluta del intervalo i . Aquí actúa como un ponderador. En lugar de sumar el logaritmo de cada dato individual, multiplica el logaritmo de la marca de clase por la cantidad de personas que hay en ese grupo.

$\log(x_i)$: Es el logaritmo de la marca de clase del intervalo.

n (El Índice): Es el tamaño total de tu muestra

Mediana

Es la medida de posición central que divide a la serie de datos en dos partes porcentualmente iguales. Es el valor que deja el 50% de las observaciones por debajo y el 50% por encima. En datos agrupados, se calcula mediante interpolación lineal en el intervalo central. A diferencia de la media, no se ve afectada por valores extremadamente altos o bajos, por lo que es más robusta en distribuciones asimétricas (Lind, Marchal & Wathen, 2018).

$$\text{Mediana} = L_i + \left(\frac{\frac{n}{2} - F_{i-1}}{f_i} \right) \cdot A_i$$

L_i : Límite inferior de la clase mediana (el intervalo donde la frecuencia acumulada alcanza o supera por primera vez $n/2$).

n : Tamaño total de la muestra (número de observaciones).

F_{i-1} : Frecuencia absoluta acumulada del intervalo inmediatamente anterior a la clase mediana.

f_i : Frecuencia absoluta simple de la clase mediana. A_i : Amplitud del intervalo de clase (diferencia entre el límite superior e inferior).

Moda

Es el valor que representa la mayor densidad de frecuencia en la muestra. En tablas de frecuencias, es el valor (o punto dentro de un intervalo) donde se concentra la mayor cantidad de datos (Lind, Marchal & Wathen, 2018). Indica el “estilo” predominante o el valor más común dentro del grupo. Una distribución puede ser unimodal, bimodal o multimodal si existen varios picos de frecuencia similares.

$$\text{Moda} = L_i + \left(\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \right) \cdot A_i$$

L_i : Límite inferior de la clase modal (el intervalo con la mayor frecuencia absoluta f_i).

f_i : Frecuencia absoluta simple de la clase modal.

f_{i-1} : Frecuencia absoluta simple de la clase inmediatamente anterior a la clase modal.

f_{i+1} : Frecuencia absoluta simple de la clase inmediatamente posterior a la clase modal.

A_i : Amplitud del intervalo de clase.

```
media_agrupada <- sum(tabla_completa$Marca_Clase * tabla_completa$Frecuencia_Absoluta) /
  ↳ sum(tabla_completa$Frecuencia_Absoluta)

n <- sum(tabla_completa$Frecuencia_Absoluta)
suma_log <- sum(tabla_completa$Frecuencia_Absoluta * log(tabla_completa$Marca_Clase))
media_geo_agrupada <- exp(suma_log / n)

intervalo_mediano <- tabla_completa %>% filter(Porc_Acumulado >= 50) %>% slice(1)
mediana_agrupada <- intervalo_mediano$Marca_Clase

intervalo_modal <- tabla_completa %>% filter(Frecuencia_Absoluta ==
  ↳ max(Frecuencia_Absoluta))
moda_agrupada <- intervalo_modal$Marca_Clase
cat("Resultados de Tendencia Central (Datos Agrupados):\n",
    "Media aritmética:", round(media_agrupada, 2), "\n",
    "Mediana:", mediana_agrupada, "\n",
    "Moda:", moda_agrupada, "\n",
    "Media geométrica:", media_geo_agrupada, "\n")
```

```
## Resultados de Tendencia Central (Datos Agrupados):
## Media aritmética: 1.28
## Mediana: 1.273081
## Moda: 0.9246466
## Media geométrica: 1.110211
```

El análisis de las medidas de tendencia central revela una estructura de resultados notablemente equilibrada en la articulación estudio-trabajo. La Mediana (1.27) y la Media Aritmética (1.28) presentan una convergencia casi exacta, situándose en el tramo central de la escala; esto indica que el punto que divide a la muestra a la mitad coincide con el promedio general, sugiriendo una distribución simétrica en el núcleo de la población analizada. Por su parte, la Moda (0.92), al situarse por debajo de la media, identifica el nivel de mayor frecuencia en una etapa de consolidación inicial-intermedia, lo que matiza la interpretación: aunque el promedio es sólido, el grupo más numeroso de individuos se concentra en un rango previo a la media aritmética.

Finalmente, la Media Geométrica (1.11) actúa como el anclaje de estabilidad del modelo, al ser menos sensible a las fluctuaciones de los valores periféricos. En conjunto, estos valores sugerirían que la articulación

académico-profesional no es precaria, sino que presenta una consistencia robusta con una tendencia clara hacia la maduración del desempeño en la muestra estudiada.

```
h_max <- max(tabla_completa$Frecuencia_Absoluta)

ggplot(tabla_completa, aes(x = Marca_Clase, y = Frecuencia_Absoluta)) +
  # Barras y Polígono
  geom_bar(stat = "identity", fill = "steelblue", alpha = 0.3, color = "white") +
  geom_line(color = "darkblue", size = 1.2) +
  geom_point(color = "darkblue", size = 3) +

  # Líneas de Tendencia Central
  geom_vline(aes(xintercept = media_agrupada), color = "red") +
  geom_vline(aes(xintercept = mediana_agrupada), color = "darkgreen") +
  geom_vline(aes(xintercept = media_geo_agrupada), color = "black") +
  geom_vline(aes(xintercept = moda_agrupada), color = "darkorange") +

  # Media
  annotate("text", x = media_agrupada, y = h_max * 0.65,
    label = paste("Media:", round(media_agrupada, 2)),
    color = "red", angle = 90, vjust = -0.5, fontface = "bold") +

  # Mediana
  annotate("text", x = mediana_agrupada, y = h_max * 0.75,
    label = paste("Mediana:", round(mediana_agrupada, 2)),
    color = "darkgreen", angle = 90, vjust = 1.5, fontface = "bold") +

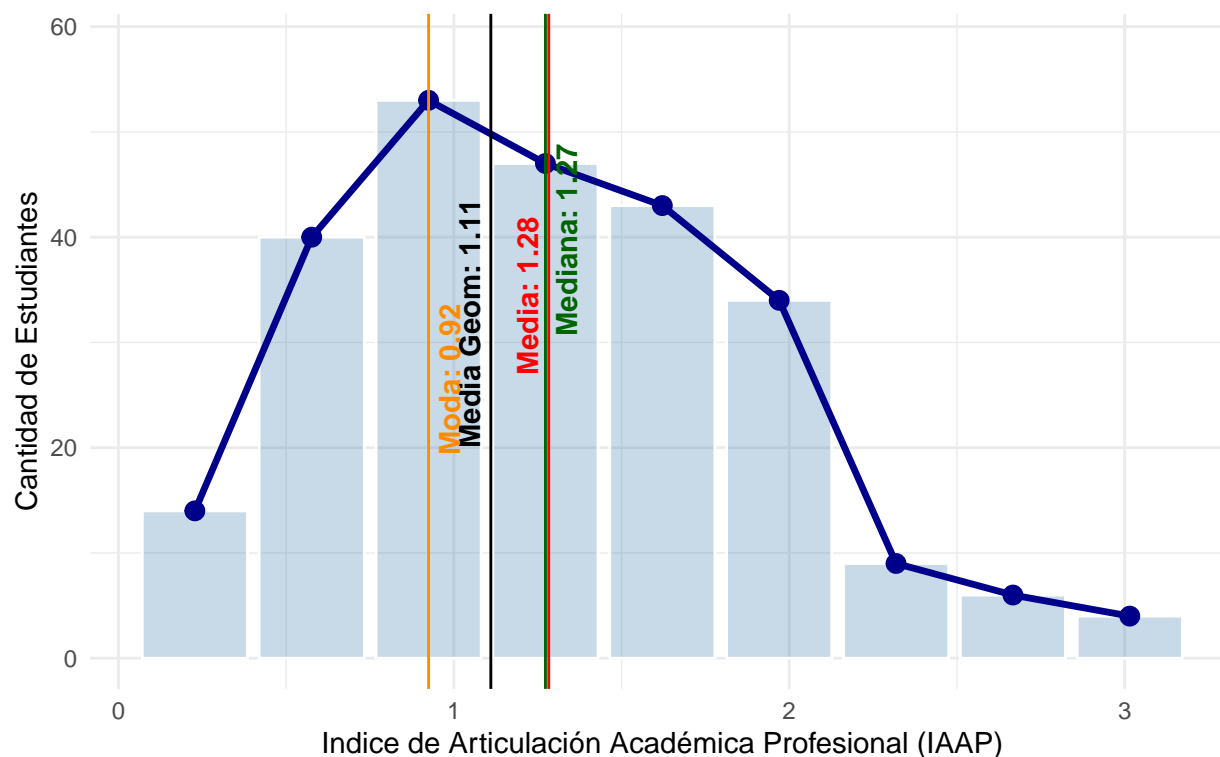
  # Media Geométrica
  annotate("text", x = media_geo_agrupada, y = h_max * 0.60,
    label = paste("Media Geom:", round(media_geo_agrupada, 2)),
    color = "black", angle = 90, vjust = -0.5, fontface = "bold") +

  # Moda
  annotate("text", x = moda_agrupada, y = h_max * 0.50,
    label = paste("Moda:", round(moda_agrupada, 2)),
    color = "darkorange", angle = 90, vjust = 1.5, fontface = "bold") +

  # Estética y límites
  scale_y_continuous(limits = c(0, h_max * 1.1)) + # un 10% de margen arriba
  labs(title = "Gráfico 1. Polígono de Frecuencias de la variable IAAP",
    subtitle = "La silueta une las marcas de clase de cada intervalo",
    x = "Índice de Articulación Académica Profesional (IAAP)",
    y = "Cantidad de Estudiantes") +
  theme_minimal()
```

Gráfico 1. Polígono de Frecuencias de la variable IAAP

La silueta une las marcas de clase de cada intervalo



El Polígono de Frecuencias revela una distribución con una mayor densidad en el segmento central-izquierdo del eje X. El pico máximo, que representa la Moda (0.92), se sitúa en el segundo tercio de la escala, indicando que el perfil de desempeño más frecuente se encuentra en una etapa de consolidación. Se observa una jerarquía de valores donde la Media Aritmética (1.28) y la Mediana (1.27) presentan una convergencia casi exacta, situándose ambas por encima de la moda y de la Media Geométrica (1.11). Visualmente, aunque el gráfico mantiene una leve prolongación o “cola” hacia los valores máximos cercanos a 3.19, el área bajo la curva muestra una consistencia robusta en los niveles intermedios.

Esta configuración confirma una distribución con una asimetría positiva leve. La mínima diferencia entre la media y la mediana (apenas 0.01 unidades) sugiere que el promedio no está siendo drásticamente distorsionado por valores extremos, sino que representa fielmente el punto de equilibrio de la muestra. A diferencia de modelos anteriores, la ubicación de la mediana en 1.27 revela que el 50% de los sujetos ya ha superado el primer tercio del índice, desplazando el foco desde niveles iniciales hacia rangos de ejecución intermedia-avanzada. En este escenario, la Media Geométrica (1.11) se ratifica como el indicador de mayor estabilidad, validando una estructura donde la progresión hacia la excelencia es constante y menos dependiente de casos aislados.

5. Medidas de Posición

Cuartiles

Los cuartiles son medidas de posición no central que dividen un conjunto de datos ordenados en cuatro partes porcentualmente iguales (25% cada una) (Black, 2019). Mientras que la mediana divide los datos en dos, los cuartiles permiten segmentar la muestra en cuatro grupos de rendimiento o magnitud.

- Q_1 (25%): El umbral del primer cuarto de los datos.

- Q_2 (50%): Coincide con la Mediana.
- Q_3 (75%): El umbral que separa al 75% inferior del 25% superior (los valores más altos).

La estructura de la fórmula para datos agrupados es casi idéntica a la de la mediana, pero el “motor” del cálculo cambia según el cuartil que busques.

Los elementos dinámicos (Lo que cambia):

- La posición ($\frac{kn}{4}$): Es lo que determina en qué fila de tu tabla de frecuencias te vas a ubicar. Para Q_1 usarías $\frac{1n}{4}$ (o simplemente $0.25n$). Para Q_3 usas $\frac{3n}{4}$ (el 75% de los datos).
- Los datos de la clase (L_i, F_{i-1}, f_i): Al cambiar la posición, te mueves de intervalo. Por lo tanto, el límite inferior (L_i) y las frecuencias (F_{i-1}, f_i) deben extraerse de la nueva fila correspondiente al 75% de la frecuencia acumulada.

Los elementos estables (Lo que no cambia):

- n : El tamaño de la muestra sigue siendo el mismo.
- A_i : La amplitud de los intervalos no varía.

Como ejemplo, si se desea calcular el cuartil 3 sería:

$$Q_3 = L_i + \left(\frac{\frac{3n}{4} - F_{i-1}}{f_i} \right) \cdot A_i$$

Q_3 : Tercer cuartil o percentil 75.

L_i : Límite inferior de la clase que contiene al tercer cuartil (donde $F_i \geq \frac{3n}{4}$).

n : Tamaño total de la muestra. $\frac{3n}{4}$: Posición teórica del tercer cuartil en la distribución acumulada.

F_{i-1} : Frecuencia absoluta acumulada del intervalo anterior a la clase de Q_3 .

f_i : Frecuencia absoluta simple de la clase que contiene al tercer cuartil.

A_i : Amplitud del intervalo de clase.

Deciles

Los deciles son las nueve medidas de posición que dividen a un conjunto de datos ordenados en diez partes porcentualmente iguales (10% cada una). Mientras que los cuartiles dividen los datos en 4 trozos, los deciles la dividen en 10. La estructura de la fórmula es similar a la del cuartil, solo se modifica el divisor de la posición (Anderson, Sweeney & Williams, 2019). Por ejemplo, el cálculo del decil 7, sería:

$$D_7 = L_i + \left(\frac{\frac{7n}{10} - F_{i-1}}{f_i} \right) \cdot A_i$$

D_7 : Séptimo decil o percentil 70.

L_i : Límite inferior de la clase que contiene al séptimo decil (donde $F_i \geq \frac{7n}{10}$).

n : Tamaño total de la muestra. $\frac{7n}{10}$: Posición del séptimo decil (70% de los datos).

F_{i-1} : Frecuencia absoluta acumulada del intervalo inmediatamente anterior a la clase de D_7 .

f_i : Frecuencia absoluta simple del intervalo donde se ubica el D_7 .

A_i : Amplitud del intervalo de clase

Percentiles

En la práctica analítica contemporánea, los percentiles son mayormente utilizados respecto a los cuartiles y deciles, debido a su resolución y granularidad. Mientras que los cuartiles dividen la muestra en cuatro grandes bloques y los deciles en diez, los percentiles segmentan la distribución en cien partes iguales, permitiendo identificar posiciones relativas con una precisión milimétrica (Black, 2019). El uso de percentiles proporciona un lenguaje común y detallado que facilita la comparación de resultados individuales frente a normativas estandarizadas, ofreciendo una visión mucho más matizada de la jerarquía de los datos que la que permiten las divisiones gruesas de los otros parámetros de posición.

$$P_{85} = L_i + \left(\frac{\frac{85n}{100} - F_{i-1}}{f_i} \right) \cdot A_i$$

P_{85} : Percentil 85 o cuantil 0.85. L_i : Límite inferior de la clase que contiene al percentil 85 (donde $F_i \geq \frac{85n}{100}$).

n : Tamaño total de la muestra. $\frac{85n}{100}$: Posición del percentil 85 (85% de la masa de datos).

F_{i-1} : Frecuencia absoluta acumulada del intervalo anterior a la clase de P_{85} .

f_i : Frecuencia absoluta simple de la clase donde se encuentra el percentil 85.

A_i : Amplitud del intervalo de clase.

```
# n = total de la muestra
n <- sum(tabla_completa$Frecuencia_Absoluta)

# --- CÁLCULO DEL CUARTIL 3 (Q3 - Percentil 75) ---
f_C3 <- tabla_completa %>% filter(Porc_Acumulado >= 75) %>% slice(1)
idx3 <- which(tabla_completa$Intervalo == f_C3$Intervalo)
# F_ant es 0 si el índice es 1
F_ant3 <- if(idx3 == 1) 0 else tabla_completa$Frec_Acumulada[idx3-1]

C3 <- cortes_sturges[idx3] + (((n * 0.75) - F_ant3) / f_C3$Frecuencia_Absoluta) *
  ↪ amplitud

# --- CÁLCULO DEL DECIL 7 (D7 - Percentil 70) ---
f_D7 <- tabla_completa %>% filter(Porc_Acumulado >= 70) %>% slice(1)
idx7 <- which(tabla_completa$Intervalo == f_D7$Intervalo)
F_ant7 <- if(idx7 == 1) 0 else tabla_completa$Frec_Acumulada[idx7-1]

D7 <- cortes_sturges[idx7] + (((n * 0.70) - F_ant7) / f_D7$Frecuencia_Absoluta) *
  ↪ amplitud

# --- CÁLCULO DEL PERCENTIL 85 (P85) ---
f_P85 <- tabla_completa %>% filter(Porc_Acumulado >= 85) %>% slice(1)
idx85 <- which(tabla_completa$Intervalo == f_P85$Intervalo)
F_ant85 <- if(idx85 == 1) 0 else tabla_completa$Frec_Acumulada[idx85-1]

P85 <- cortes_sturges[idx85] + (((n * 0.85) - F_ant85) / f_P85$Frecuencia_Absoluta) *
  ↪ amplitud

# --- MOSTRAR RESULTADOS ---
cat("RESULTADOS DE POSICIÓN (IAAP):\n",
```

```
"Cuartil 3 (Q3): ", round(C3, 3), "\n",
"Decil 7 (D7): ", round(D7, 3), "\n",
"Percentil 85 (P85):", round(P85, 3))
```

```
## RESULTADOS DE POSICIÓN (IAAP):
## Cuartil 3 (Q3): 1.719
## Decil 7 (D7): 1.617
## Percentil 85 (P85): 1.955
```

Los cálculos realizados para datos agrupados arrojan los siguientes valores de corte en la distribución:

El Cuartil 3 (Q3) se sitúa en 1.719, el Decil 7 (D7) en 1.617 y el Percentil 85 (P85) alcanza un valor de 1.955. Al contrastar estos datos con el Polígono de Frecuencias, se observa que todos estos puntos de posición se desplazan con claridad hacia la mitad derecha de la distribución, superando consistentemente a la Mediana (1.27). Esta transición indica que el núcleo del desempeño se ha movido hacia rangos de mayor consolidación.

Estos resultados permiten segmentar la población según su ejecución de manera precisa. El valor del Q3 (1.719) indica que el 75% de los evaluados obtuvo un puntaje igual o inferior a dicha cifra, situando a la cuarta parte restante en un nivel avanzado. Por su parte, el P85 (1.955) define el umbral de excelencia, marcando el inicio del “Top 15%” de la muestra; es decir, solo un grupo selecto logra aproximarse o superar la barrera de los 2.0 puntos. La distancia moderada entre el D7 (1.617) y el P85 refuerza la observación de una progresión gradual en la articulación, donde alcanzar los niveles más altos del índice sigue siendo un logro estadísticamente exigente pero más frecuente que en modelos de baja resolución.

6. Medidas de Variabilidad

Varianza

Estadísticamente, la varianza representa el promedio de los cuadrados de las desviaciones. Al elevar las diferencias al cuadrado, se garantiza que todas las distancias sean positivas y se penaliza proporcionalmente a aquellos valores que se alejan más del centro de la distribución. Para datos agrupados, el uso de f_i pondera la importancia de cada desviación según la cantidad de observaciones en ese intervalo, proporcionando una medida de la heterogeneidad del grupo (Walpole et al., 2012).

$$s^2 = \frac{\sum f_i(x_i - \bar{x})^2}{n - 1}$$

s^2 : Varianza de la muestra (estimador de la dispersión).

\sum : Operador sumatorio de todas las clases del modelo.

f_i : Frecuencia absoluta simple de la clase i .

x_i : Marca de clase (punto medio del intervalo).

\bar{x} : Media aritmética de la distribución (calculada previamente).

$(x_i - \bar{x})^2$: Desviación cuadrática de la marca de clase respecto a la media.

$n - 1$: Grados de libertad de la muestra (corrección de Bessel para un estimador insesgado)

Desviación Estándar

Es la raíz cuadrada de la varianza. Su función principal es devolver la medida de dispersión a las unidades originales de la variable (por ejemplo, si la varianza está en “puntos de IQ al cuadrado”, la desviación vuelve

a ser simplemente “puntos de IQ”). Con un valor de desviación estándar de 0.54 y una media de 1.28, por ejemplo, indicaría que, en promedio, los puntajes individuales se alejan 0.54 unidades respecto a la media aritmética (Lind, Marchal & Wathen, 2018). Esto significaría que la “zona de normalidad” o el rango típico de desempeño en la muestra se sitúa aproximadamente entre 0.74 y 1.82.

En una distribución que tiende a la normalidad, este rango ($\text{Media} \pm 1 \text{ Desviación Estándar}$) contiene aproximadamente al 68% de la población. Es decir, la gran mayoría de los sujetos evaluados presentarían un nivel de articulación que se mueve en este intervalo, lo que define el estándar de desempeño esperado para el colectivo (Walpole, Myers, Myers & Ye, 2012).

La interpretación estadística, sin embargo, trasciende el primer umbral de dispersión para explorar los límites de la normalidad operativa. Bajo la lógica de la regla empírica, el modelo extiende su validez hasta las tres desviaciones estándar, situando el límite superior de la distribución esperada en aproximadamente 2.90 unidades ($\bar{x} + 3s = 1.28 + 1.62$). Alcanzar este nivel es estadísticamente “normal” pero cualitativamente extraordinario: mientras que los perfiles situados entre una y dos desviaciones (hasta 2.36; $\bar{x} + 2s = 1.28 + 1.08$) representan un desempeño destacado y predecible, aquellos que se ubican entre la segunda y la tercera desviación (hasta 2.90) constituyen casos excepcionales de alta especialización que apenas representan al 2.1% de la población. Cuando el IAAP supera este techo, como puede ocurrir con registros máximos como ser 3.19 (> 2.90), se encuentran valores atípicos o outliers que, lejos de ser errores de medición, simbolizan la excelencia absoluta; individuos que han logrado romper la tendencia central del grupo para situarse en una categoría de articulación académico-profesional superior y fuera de lo convencional.

$$s = \sqrt{s^2}$$

s : Desviación estándar muestral.

$\sqrt{s^2}$: Operador de raíz cuadrada aplicado a la varianza.

\bar{x} : Media aritmética

Coefficiente de Variación (CV)

Es una medida de dispersión relativa. Permite comparar la variabilidad entre grupos diferentes o variables con distintas unidades (por ejemplo, comparar qué varía más: el Salario o el IQ Académico).

Mientras que la desviación estándar nos da una magnitud absoluta, el CV ofrece una proporción de error o variabilidad en relación con la media. Interpretarlo implica entender qué tan “ruidosos” son los datos: un CV bajo (típicamente $< 15\%$) sugiere un grupo homogéneo, donde la media es un reflejo fiel de casi todos los individuos. Por el contrario, un CV alto ($> 30\%$) señala un grupo heterogéneo, advirtiendo que la media es una cifra “híbrida” que promedia realidades muy distantes (Pearson, 1895; Kazmier, 2006), perdiendo así su capacidad para describir al sujeto típico.

Desde el punto de vista analítico, la implicación más profunda del CV radica en la estabilidad del modelo. Si el coeficiente es elevado, la media aritmética deja de ser el indicador más confiable para la toma de decisiones, ya que está siendo afectada por la dispersión de los datos. En términos de gestión, un CV alto implica que no existe una “receta única” para el grupo, sino que conviven subpoblaciones (líderes y rezagados) que requieren análisis segmentados. Por tanto, el CV no solo mide dispersión; mide la confianza en los indicadores centrales.

$$CV = \left(\frac{s}{\bar{x}} \right) \cdot 100\%$$

CV : Coeficiente de variación (suele expresarse en porcentaje: $CV \times 100$).

s : Desviación estándar.

$|\bar{x}|$: Valor absoluto de la media aritmética.

```

n <- sum(tabla_completa$Frecuencia_Absoluta)
# Variabilidad
varianza_agrupada <- sum(tabla_completa$Frecuencia_Absoluta * (tabla_completa$Marca_Clase
  ↪ - media_agrupada)^2) / (n - 1)
desviacion_agrupada <- sqrt(varianza_agrupada)
cv <- (desviacion_agrupada / media_agrupada) * 100

cat("MEDIDAS DE VARIABILIDAD:\n",
    "Varianza:", round(varianza_agrupada, 4), "\n",
    "Desviación Estándar:", round(desviacion_agrupada, 4), "puntos\n",
    "Coeficiente de Variación:", round(cv, 2), "%\n")

```

```

## MEDIDAS DE VARIABILIDAD:
## Varianza: 0.3875
## Desviación Estándar: 0.6225 puntos
## Coeficiente de Variación: 48.53 %

```

Al analizar la desviación estándar de 0.62, es vital contextualizarla dentro del Rango de la muestra, que en este estudio es de 3.14 unidades (calculado como la diferencia entre el valor máximo de 3.19 y el mínimo de 0.05). Que la desviación represente casi un 20% del recorrido total de los datos ($0.62/3.14 = 19.7\%$) confirma que la variabilidad no es un error de medida, sino una característica intrínseca del grupo.

Esta dispersión, reflejada en el CV del 48.53%, indica que los sujetos no están confinados en un pequeño nicho de puntajes, sino que aprovechan casi toda la extensión del rango observado. Esto implica que el IAAP tiene una alta capacidad de discriminación: es capaz de distinguir claramente entre perfiles muy diversos a lo largo de esos 3.14 puntos de diferencia, permitiendo identificar desde niveles de articulación incipientes hasta desempeños de alta competencia.

8. Medidas de Asimetría y Curtosis

Asimetría de Fisher

Es la medida basada en el tercer momento central. Evalúa la simetría de la distribución respecto a la media.

El coeficiente de Fisher se basa en el tercer momento central de la distribución, normalizado por la desviación estándar elevada al cubo (Fisher, 1922). Teóricamente, el valor de referencia fundamental es cero. Cuando el coeficiente arroja un resultado de $g_1 = 0$, se establece que la distribución es perfectamente simétrica, lo que significa que el lado izquierdo y el derecho respecto a la media son imágenes especulares entre sí. Si el valor es mayor a cero ($g_1 > 0$), se determina una asimetría positiva, mientras que un valor menor a cero ($g_1 < 0$) define una asimetría negativa. La magnitud del número indica la severidad del sesgo: valores entre -0.5 y 0.5 sugieren una asimetría leve, mientras que valores que exceden 1 o -1 indican una asimetría pronunciada.

Desde la perspectiva de la teoría estadística, el signo del coeficiente de Fisher describe hacia dónde se “estira” la cola de la distribución (Aczel & Sounderpandian, 2017). En una asimetría positiva ($g_1 > 0$), la cola derecha es más larga o gruesa, lo que implica que la mayoría de los datos se agrupan en valores menores a la media, pero existen casos extremos en la parte alta que desplazan el promedio. En una asimetría negativa ($g_1 < 0$), sucede lo contrario: la cola se extiende hacia la izquierda, indicando que la masa de los datos está en valores altos, con valores extremos bajos que “jalan” la media hacia abajo. En términos de tendencia central, la asimetría rompe la igualdad entre las medidas: en un sesgo positivo, la media suele ser mayor que la mediana, mientras que en uno negativo, la media es menor que la mediana.

$$g_1 = \frac{\frac{1}{n} \sum f_i (x_i - \bar{x})^3}{s^3}$$

g_1 : Coeficiente de asimetría de Fisher.

$\sum f_i(x_i - \bar{x})^3$: Tercer momento central (las desviaciones elevadas al cubo mantienen el signo, indicando la dirección del sesgo).

s^3 : Cubo de la desviación estándar (estandariza la medida).

n : Tamaño de la muestra.

Asimetría de Bowley

Es una medida de asimetría robusta basada exclusivamente en los cuartiles de la distribución. Es una medida de asimetría robusta basada exclusivamente en los cuartiles de la distribución, ideal para analizar la simetría del “corazón” de la muestra sin distorsiones por valores atípicos (Newbold, Carlson & Thorne, 2013).

La interpretación teórica del Coeficiente de Asimetría de Bowley (A_B) se fundamenta en la relación de distancia entre los cuartiles de una distribución, ofreciendo una perspectiva basada exclusivamente en la posición de los datos y no en sus momentos centrales

El coeficiente de Bowley utiliza como puntos de referencia el Primer Cuartil (Q_1), el Segundo Cuartil o Mediana (Q_2) y el Tercer Cuartil (Q_3). El valor de referencia teórico es cero. El cálculo se basa en comparar la distancia entre el cuartil superior y la mediana ($Q_3 - Q_2$) frente a la distancia entre la mediana y el cuartil inferior ($Q_2 - Q_1$). Si ambas distancias son idénticas, el numerador de la fórmula se anula, resultando en un coeficiente de $A_B = 0$, lo que dictamina una simetría perfecta en la zona central de la distribución. Un resultado positivo indica que el tramo superior es más extenso, mientras que un resultado negativo indica que el tramo inferior es el más largo.

Desde el punto de vista de la teoría estadística, el coeficiente de Bowley es una medida de asimetría robusta, ya que, al centrarse en los cuartiles, no se ve distorsionado por valores extremadamente alejados (valores atípicos) que sí afectarían a Fisher. Un valor de Bowley mayor a cero interpreta que el 25% de los datos con valores más altos está más disperso o alejado de la mediana que el 25% de los datos con valores más bajos; esto define una concentración de la masa de datos en la parte inferior de la escala. Por el contrario, un valor menor a cero interpreta que la dispersión es mayor en la zona de valores bajos, indicando que la masa de la población se concentra en los niveles altos. A diferencia de otras medidas, Bowley solo describe lo que ocurre dentro del 50% central de los datos (entre el percentil 25 y 75), siendo ideal para analizar la simetría del “corazón” de la muestra.

$$A_B = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

A_B : Coeficiente de asimetría de Bowley.

Q_1, Q_2, Q_3 : Primer, segundo (mediana) y tercer cuartil, respectivamente.

$(Q_3 - Q_1)$: Rango intercuartílico (denominador de normalización).

Asimetría de Pearson

Mide la distancia relativa entre la media y la moda, indicando qué tan desplazada está la “punta” de la curva o cúspide de la curva respecto al promedio (Keller, 2018).

La interpretación teórica del Coeficiente de Asimetría de Pearson (A_P) se basa en la relación de distancia que existe entre la Media Aritmética y la Moda (o la Mediana), utilizando la desviación estándar como unidad de medida para normalizar el resultado.

El parámetro de referencia fundamental para este coeficiente es el cero. Cuando el valor es $A_P = 0$, se establece que la media y la moda coinciden en el mismo punto, lo que caracteriza a una distribución perfectamente simétrica (unimodal). Si el resultado es positivo ($A_P > 0$), significa que la media es mayor que la

moda. Si el resultado es negativo ($A_P < 0$), la media es menor que la moda. La magnitud del coeficiente indica la intensidad de la deformación: se considera una asimetría leve si el valor está cerca de cero y una asimetría fuerte si se aproxima o supera el valor de 1.

Desde la perspectiva de la teoría estadística, el coeficiente de Pearson mide el grado de “desplazamiento” de la tendencia central. En una asimetría positiva, la interpretación es que los valores extremos altos están “jalando” a la media lejos del valor más frecuente (la moda), creando una cola larga hacia la derecha. En una asimetría negativa, son los valores extremadamente bajos los que arrastran la media hacia la izquierda de la moda. A diferencia de Fisher, que observa toda la forma de la curva, Pearson se enfoca en el desequilibrio entre el promedio y el punto de mayor concentración, siendo una medida muy intuitiva para entender qué tan engañosa puede ser la media aritmética respecto al valor que más se repite en la muestra.

$$A_P = \frac{\bar{x} - \text{Moda}}{s}$$

A_P : Primer coeficiente de asimetría de Pearson.

\bar{x} : Media aritmética.

Moda: Valor con mayor frecuencia absoluta.

s : Desviación estándar.

Curtosis

Mide el grado de apuntamiento o concentración de los datos alrededor de la media (la “agudeza” de la curva).

El parámetro de referencia teórico para la curtosis es cero (cuando se calcula como “exceso de curtosis”). Bajo esta métrica, una distribución con un valor igual a cero se denomina Mesocúrtica, lo que indica que su grado de concentración de datos es idéntico al de una distribución normal. Si el coeficiente arroja un valor positivo (>0), la distribución se clasifica como Leptocúrtica, caracterizada por un pico muy elevado y colas delgadas. Por el contrario, un valor negativo (<0) define una distribución Platicúrtica, la cual presenta una forma aplanada con una dispersión de datos más uniforme y menos concentrada en el centro (Triola, 2018).

Desde la perspectiva de la teoría estadística, la curtosis mide la “pesadez” de las colas y la esbeltez del pico de la distribución. Una curva Leptocúrtica interpreta una alta concentración de frecuencias alrededor de la zona central, sugiriendo que gran parte de la población comparte valores muy similares, con pocos casos que se alejan hacia los extremos. Una curva Platicúrtica interpreta una baja concentración, donde los datos están más repartidos a lo largo de la escala, indicando mayor diversidad o heterogeneidad cerca del centro. En esencia, mientras la asimetría nos dice hacia dónde se inclina la balanza, la curtosis nos dice qué tan “apretados” están los individuos en el punto de mayor frecuencia.

$$g_2 = \left(\frac{\frac{1}{n} \sum f_i (x_i - \bar{x})^4}{s^4} \right) - 3$$

g_2 : Coeficiente de exceso de curtosis.

$\sum f_i (x_i - \bar{x})^4$: Cuarto momento central (sensible a valores extremos o “colas”).

s^4 : Desviación estándar a la cuarta potencia.

−3: Constante de ajuste que centra el valor en 0 para una distribución normal (Mesocúrtica).

```
# --- CÁLCULO DE MOMENTOS (Para Fisher) ---
# Usa n calculado anteriormente y las marcas de clase de Sturges
m3 <- sum(tabla_completa$Frecuencia_Absoluta * (tabla_completa$Marca_Clase -
  ↪ media_agrupada)^3) / n
m4 <- sum(tabla_completa$Frecuencia_Absoluta * (tabla_completa$Marca_Clase -
  ↪ media_agrupada)^4) / n
```

```

# 1. ASIMETRÍA DE FISHER
asimetria_fisher <- m3 / (desviacion_agrupada^3)

# 2. ASIMETRÍA DE PEARSON
asimetria_pearson <- (media_agrupada - moda_agrupada) / desviacion_agrupada

# 3. ASIMETRÍA DE BOWLEY (Basada en Cuartiles)

# función para que use los cortes de Sturges y la amplitud real
calcular_Q_dinamico <- function(k) {
  objetivo <- n * k
  fila_idx <- which(tabla_completa$Frec_Acumulada >= objetivo)[1]

  # Usa los vectores que definiste en el bloque de Sturges
  L_inf <- cortes_sturges[fila_idx]
  F_ant <- if(fila_idx == 1) 0 else tabla_completa$Frec_Acumulada[fila_idx - 1]
  f_i <- tabla_completa$Frecuencia_Absoluta[fila_idx]
  # 'amplitud' es la variable calculada (0.35)

  return(L_inf + ((objetivo - F_ant) / f_i) * amplitud)
}

# Cálculos de Cuartiles para Bowley
Q1_b <- calcular_Q_dinamico(0.25)
Q2_b <- calcular_Q_dinamico(0.50) # Mediana interpolada
Q3_b <- calcular_Q_dinamico(0.75)

asimetria_bowley <- (Q3_b + Q1_b - 2 * Q2_b) / (Q3_b - Q1_b)

# 4. CURTOSIS (G2)
kurtosis_fisher <- (m4 / (desviacion_agrupada^4)) - 3

# --- MOSTRAR RESULTADOS ---
cat("CARACTERIZACIÓN MORFOLÓGICA (IAAP):\n",
    "-----\n",
    "Asimetría de Fisher: ", round(asimetria_fisher, 3), "\n",
    "Asimetría de Pearson: ", round(asimetria_pearson, 3), "\n",
    "Asimetría de Bowley: ", round(asimetria_bowley, 3), "\n",
    "Curtosis: ", round(kurtosis_fisher, 3))

## CARACTERIZACIÓN MORFOLÓGICA (IAAP):
## -----
## Asimetría de Fisher: 0.438
## Asimetría de Pearson: 0.575
## Asimetría de Bowley: 0.066
## Curtosis: -0.211

```

La morfología de la distribución presenta una transición hacia la estabilidad, alejándose de los sesgos extremos de modelos preliminares. Los coeficientes de asimetría arrojan valores positivos pero moderados: Fisher (0.438), Pearson (0.575) y Bowley (0.066). Asimismo, se registra una Curtosis (Exceso) de -0.211. Visualmente, esto se traduce en una silueta con un sesgo a la derecha suavizado y una curva más abierta y menos pronunciada que la normal.

La convergencia de los tres coeficientes de asimetría por encima de cero ratifica un sesgo positivo leve. Fisher (0.438) indica que persiste una “cola” hacia los valores altos, pero mucho menos extendida que en mediciones previas; Pearson (0.575) confirma una separación moderada entre la moda y el promedio, mientras que Bowley (0.066), al estar muy cerca de cero, demuestra que el 50% central de la muestra ya presenta una estructura interna casi simétrica.

Por otro lado, el exceso de curtosis negativo (-0.211) define una naturaleza platicúrtica. A diferencia de una distribución “puntiaguda” (leptocúrtica), este valor indica que los datos están más dispersos y aplanados, lo que interpreta una menor concentración masiva en un solo punto y una mayor diversidad de perfiles a lo largo del índice. En conclusión, la población no está “amontonada” en los niveles bajos, sino que se distribuye de manera más fluida y heterogénea, donde alcanzar puntajes superiores es una progresión natural y no una anomalía estadística extrema.

```
ggplot(df_indices, aes(x = IAAP)) +
  # 1. Histograma
  geom_histogram(aes(y = after_stat(density)), bins = 10,
    fill = "steelblue", alpha = 0.3, color = "white") +

  # 2. Curva Kernel (distribución real)
  geom_density(color = "darkblue", size = 1.2) +

  # 3. Curva Normal Teórica (Referencia en línea discontinua)
  stat_function(fun = dnorm,
    args = list(mean = 1.28, sd = 0.6225),
    color = "black", linetype = "dashed", size = 1.5) +

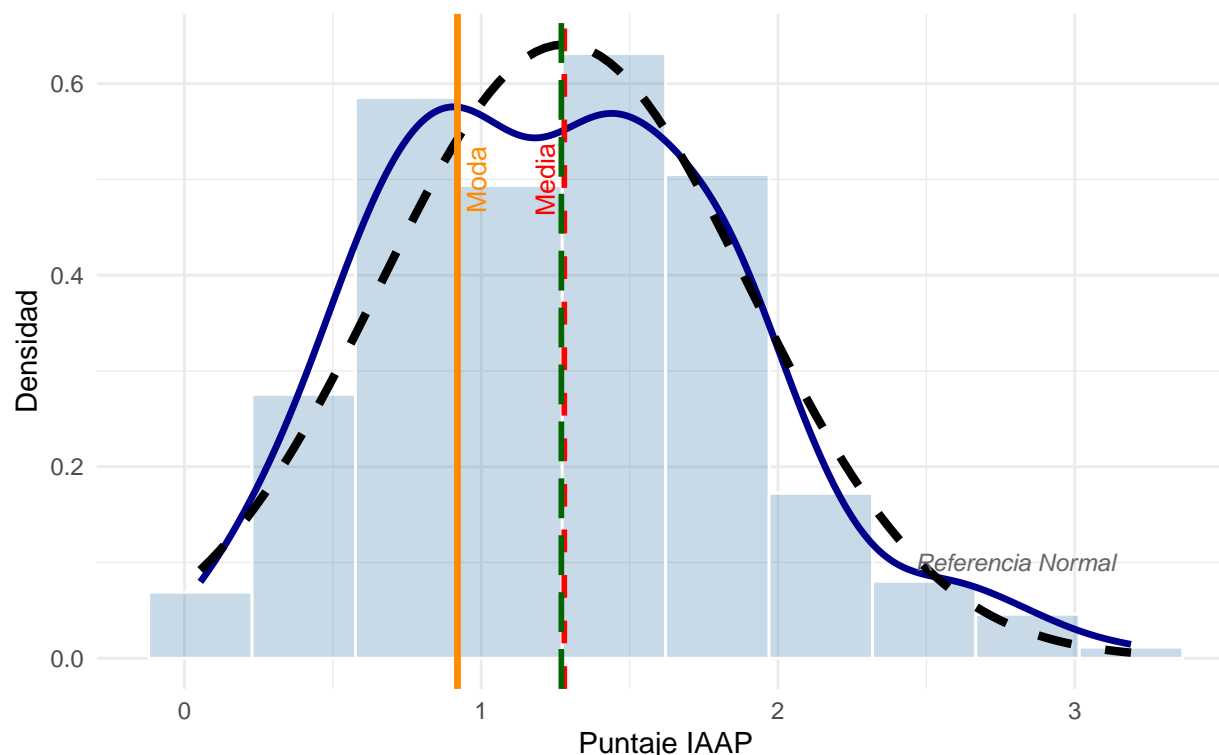
  # 4. Líneas de referencia (Tendencia Central)
  geom_vline(aes(xintercept = 1.28), color = "red", linetype = "dashed", size = 1) + #
  ↪ Media
  geom_vline(aes(xintercept = 1.27), color = "darkgreen", linetype = "longdash", size =
  ↪ 1) + # Mediana
  geom_vline(aes(xintercept = 0.92), color = "darkorange", size = 1.2) + # Moda

  # 5. Anotaciones
  annotate("text", x = 1.28, y = 0.5, label = "Media", color = "red", angle = 90, vjust =
  ↪ -0.5, size = 3.5) +
  annotate("text", x = 0.92, y = 0.5, label = "Moda", color = "darkorange", angle = 90,
  ↪ vjust = 1.5, size = 3.5) +
  # Etiqueta para la curva de referencia
  annotate("text", x = 2.8, y = 0.1, label = "Referencia Normal", color = "grey40", size
  ↪ = 3, fontface = "italic") +

  # Estética y subtítulo actualizado con los datos reales
  labs(title = "*Gráfico 2. Silueta de la Distribución de Puntajes IAAP*",
    subtitle = paste("Asimetría Fisher: 0.438 | Curtosis: -0.211 | CV: 48.53%"),
    x = "Puntaje IAAP",
    y = "Densidad") +
  theme_minimal()
```

Gráfico 2. Silueta de la Distribución de Puntajes IAAP

Asimetría Fisher: 0.438 | Curtosis: -0.211 | CV: 48.53%



El “centro de gravedad” de la muestra se localiza en una zona de desempeño intermedia-avanzada. La Media Aritmética (1.28) y la Mediana (1.27) presentan una convergencia casi total, lo que indica que el promedio general coincide con el punto que divide a la población en dos partes iguales. Sin embargo, la Moda (0.92) se sitúa por debajo de estos valores, advirtiendo que, aunque el promedio es alto, el perfil de mayor frecuencia todavía se encuentra en una etapa de consolidación inicial. La Media Geométrica (1.11) actúa como un estabilizador, validando que el núcleo de la muestra es robusto y no depende únicamente de éxitos aislados.

A pesar de la estabilidad central, la muestra se define por su heterogeneidad. Con una Desviación Estándar de 0.6225 puntos y un Coeficiente de Variación del 48.53%, queda claro que conviven realidades muy dispares. Esta variabilidad supera el umbral crítico del 30%, lo que interpreta que el IAAP es un indicador altamente sensible: es capaz de discriminar con precisión entre sujetos con una articulación incipiente y aquellos que alcanzan niveles de excelencia. La brecha entre el Decil 7 (1.61) y el Percentil 85 (1.95) confirma que el ascenso hacia los rangos superiores es exigente y marca una diferencia competitiva real.

El análisis visual de la silueta (Gráfico 2) ofrece la conclusión más reveladora. Al contrastar la curva real (azul) con la Referencia Normal (línea punteada), observamos una distribución platicúrtica (Curtosis: -0.211). Esto significa que los datos están más repartidos y son menos “puntiagudos” que una distribución normal perfecta, reflejando una diversidad de perfiles saludable y no un estancamiento masivo en un solo valor.

La Asimetría de Fisher (0.438) confirma un sesgo positivo leve; existe una “cola” hacia la derecha que se extiende hasta el máximo observado de 3.19. Esta extensión contempla individuos que han logrado romper la inercia del grupo para situarse en niveles de articulación profesional extraordinarios.

9. Discusión de Resultados

El presente análisis se basa en una muestra hipotética y recreada, diseñada para modelar el comportamiento del Índice de Articulación Académico-Profesional. Al ser un escenario simulado, los resultados permiten observar con claridad las propiedades de la escala y la sensibilidad de las medidas agrupadas bajo la regla de Sturges.

En este entorno recreado, la Media Aritmética Agrupada (1.28) y la Mediana (1.27) actúan como los pilares de estabilidad del índice, indicando un punto de equilibrio muy sólido. Al trabajar con datos agrupados, los cálculos se basan en las marcas de clase de los 9 intervalos definidos. Es importante notar que, en una muestra real, el cálculo de datos no agrupados sería ligeramente más sensible a los valores extremos, mientras que en esta recreación agrupada se obtiene una visión más suavizada y estructural de la tendencia, lo que facilita la lectura al reducir la dispersión de las puntuaciones individuales.

Para entender la morfología de esta muestra hipotética, es fundamental diferenciar entre la concentración de la masa de datos, la extensión de sus valores extremos y la simetría de su zona central. Aquí es donde los tres coeficientes nos cuentan historias distintas y complementarias:

- La “Montaña” (Sesgo de Pearson): El coeficiente de Pearson (0.575) se centra en la relación entre la media y la moda. Al ser positivo, nos indica que la “cima” o montaña de nuestra distribución está desplazada hacia la izquierda (valores bajos de 0.92). Esto significa que la mayor acumulación de sujetos en este modelo hipotético ocurre antes de llegar al promedio, creando una pendiente pronunciada en el lado de la entrada.
- La “Cola” (Sesgo de Fisher): Por su parte, el coeficiente de Fisher (0.438) analiza la simetría de las “colas” de la distribución. Su valor positivo confirma que la cola derecha es más larga y pesada que la izquierda. Esto traduce matemáticamente la existencia de un grupo selecto de casos de excelencia que “estiran” la curva hacia los 3.19 puntos, impidiendo que la distribución se cierre de forma abrupta tras el promedio.
- El “Corazón” (Sesgo de Bowley - 0.066): A diferencia de los anteriores, Bowley es un coeficiente robusto que solo utiliza los cuartiles (Q_1 , Mediana y Q_3), ignorando los valores extremos. Su valor de 0.066 es sumamente revelador: al estar tan cercano a cero, nos indica que el 50% central de la población (el “corazón” de la muestra) es prácticamente simétrico.

La variabilidad, reflejada en una Desviación Estándar de 0.6225 y un CV de 48.53%, reafirma que el modelo recreado posee una alta capacidad de discriminación. La Curtosis (-0.211), al ser negativa, nos indica que la “montaña” de datos no es un pico agudo, sino una meseta más redondeada y aplanada (platicúrtica). Esto implica que, aunque hay un sesgo hacia la izquierda, los sujetos están lo suficientemente repartidos entre las clases como para evitar una polarización extrema.

Es pertinente señalar que los cálculos realizados de forma agrupada son ideales para este tipo de modelos recreados. En una muestra recreada hipotéticamente, los datos individuales pueden presentar variaciones artificiales que no siempre responden a la lógica del fenómeno que queremos estudiar. Al utilizar la agrupación por intervalos (como los 10 rangos que definiste con Sturges), lo que se hace es aplicar un filtro de suavizado. En lugar de analizar cada “micro-variación” (el ruido), el agrupamiento permite observar la tendencia macro. Esto es ideal para validar si el índice IAAP, como teoría, funciona bien a nivel de grupos o estratos de competencia.

10. Referencias

- Aczel, A. D., & Sounderpandian, J. (2017). *Complete Business Statistics* (8th ed.). McGraw-Hill Education.
- Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2019). *Estadística para administración y economía* (13ª ed.). Cengage Learning.

- Black, K. (2019). *Business Statistics: For Contemporary Decision Making* (10th ed.). Wiley.
- Fisher, R. A. (1922). *On the mathematical foundations of theoretical statistics*. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 222(594-604), 309-368.
- Hadley, W., & Grolemund, G. (2017). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media. <https://r4ds.had.co.nz/>
- Kazmier, L. J. (2006). *Estadística aplicada a la administración y a la economía* (4^a ed.). McGraw-Hill / Interamericana.
- Keller, G. (2018). *Statistics for Management and Economics* (11th ed.). Cengage Learning.
- Levin, R. I., & Rubin, D. S. (2010). *Estadística para administración y economía* (7^a ed.). Pearson Educación.
- Lind, D. A., Marchal, W. G., & Wathen, S. A. (2018). *Estadística aplicada a los negocios y la economía* (16^a ed.). McGraw-Hill Education.
- Newbold, P., Carlson, W. L., & Thorne, B. (2013). *Estadística para administración y economía* (8^a ed.). Pearson Educación.
- Pearson, K. (1895). *Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material*. Philosophical Transactions of the Royal Society of London. A, 186, 343-414.
- R Core Team (2026). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Spiegel, M. R., & Stephens, L. J. (2009). *Estadística* (4^a ed.). McGraw-Hill.
- Sturges, H. A. (1926). *The choice of a class interval*. Journal of the American Statistical Association, 21(153), 65-66
- Triola, M. F. (2018). *Estadística* (12^a ed.). Pearson Educación.
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias* (9^a ed.). Pearson Educación.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. <https://ggplot2.tidyverse.org/>
- Wickham, H., François, R., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.2. <https://dplyr.tidyverse.org/>