**Visualizar la variabilidad: una propuesta didáctica para el desarrollo del pensamiento estadístico con ANOVA**

***Resumen***

La variabilidad es un concepto central en Estadística y una piedra angular para el desarrollo del pensamiento estadístico. Sin embargo, su comprensión resulta especialmente difícil para muchos estudiantes, que tienden a centrarse únicamente en medidas de centralización (media, mediana, etc). Esta dificultad se debe a múltiples factores tales como la tradición determinista de la enseñanza matemática, la naturaleza abstracta del concepto de variabilidad, y la falta de experiencias directas y adecuadas con datos reales y procesos aleatorios.

En este trabajo se analiza la importancia de la variabilidad en la Estadística, se revisan algunas de las razones didácticas que explican su difícil asimilación en el aula y se propone una aproximación basada en herramientas visuales e interactivas construidas con R, orientadas a hacer visible la existencia y comportamiento de la variabilidad en los datos y facilitar su análisis.

Como caso de aplicación, se plantea una propuesta didáctica centrada en el Análisis de la Varianza (ANOVA), utilizando visualizaciones dinámicas para ilustrar la descomposición de la variabilidad total en variabilidad explicada por el factor y variabilidad residual. Esta propuesta se contextualiza en la asignatura *Modelos Estadísticos para la Toma de Decisiones II* del *Grado en Ciencia de Datos* de la *Universitat Politècnica de València*, y tiene como objetivo reforzar la comprensión conceptual del ANOVA, promover el razonamiento estadístico y facilitar la toma de decisiones basada en datos.

**Palabras clave**:  
Variabilidad

Estadística

Pensamiento estadístico

Aprendizaje

Educación superior

ANOVA

Visualización

Herramientas interactivas

Lenguaje R

***Abstract***

Variability is a central concept in Statistics and a cornerstone for the development of statistical thinking. However, it remains particularly challenging for many students, who tend to focus solely on measures of central tendency (mean, median, etc.). This difficulty arises from multiple factors, such as the deterministic tradition of mathematics teaching, the abstract nature of the concept of variability, and the lack of direct and appropriate experiences with real data and random processes.

This paper explores the importance of variability in Statistics, reviews several didactic reasons that help explain its difficult assimilation in the classroom, and proposes an approach based on visual and interactive tools developed in R. These tools aim to make the existence and behaviour of variability in data more visible and facilitate its analysis.

As a case study, we present a didactic proposal focused on **Analysis of Variance (ANOVA)**, using dynamic visualisations to illustrate the decomposition of total variability into variability explained by the factor and residual variability. This proposal is contextualised within the course Statistical Models for Decision Making II of the Bachelor’s Degree in Data Science at the Universitat Politècnica de València, and aims to strengthen conceptual understanding of ANOVA, foster statistical reasoning, and support data-driven decision-making.

**Keywords**:  
Variability

Statistics

Statistical Thinking

Learning

Higher Education

ANOVA

Visualization

Interactive Tools

R Language

# **Introducción**

La variabilidad es un concepto central en la Estadística y, al mismo tiempo, uno de los más difíciles de comprender por parte del alumnado. A diferencia de otras disciplinas con un enfoque más determinista~~s~~, la estadística se ocupa de fenómenos aleatorios, donde no hay una única respuesta ~~correcta~~, sino una distribución de posibles resultados. En este contexto, comprender cómo varían los datos es clave para interpretar la información y tomar decisiones fundamentadas.

Este trabajo parte de la premisa —apoyada por autores como Carmen Batanero— de que muchos estudiantes y docentes centran su atención exclusivamente en los promedios, descuidando el análisis de la dispersión de los datos. Esta visión limitada reduce el alcance del pensamiento estadístico y obstaculiza la comprensión profunda de la incertidumbre inherente a los datos reales.

Con el objetivo de contribuir a una mejor enseñanza y aprendizaje de la variabilidad, se presentan aquí tres elementos:

* Una revisión conceptual del término y de su papel fundacional en la Estadística.
* Un análisis de las dificultades que presenta su comprensión en el aula.
* Una propuesta de herramientas visuales e interactivas que permiten abordarla de forma didáctica y experiencial.

A continuación, se presenta el concepto de variabilidad y su importancia como puerta de entrada a la Estadística.

# **El concepto de variabilidad y su importancia en Estadística**

## ¿Qué es la variabilidad?

La **variabilidad** es la tendencia natural de los datos a diferir entre sí. Es decir, describe cuánto cambian los valores de una variable respecto a un valor central, como la media. Esta característica es inherente a la mayoría de los fenómenos reales: pocas veces encontramos que todos los datos sean iguales.

## ¿Cómo se cuantifica?

En Estadística, la variabilidad se mide mediante distintas aproximaciones al concepto, entre las que destacan:

* **Varianza**: medida del promedio de las desviaciones al cuadrado respecto a la media.
* **Desviación típica (o estándar)**: raíz cuadrada de la varianza.
* **Rango** y **rango intercuartílico**: medidas basadas en la posición de los datos.
* **Coeficiente de variación**: útil para comparar variabilidades relativas entre variables con escalas diferentes.

## ¿Por qué es tan importante en Estadística?

La variabilidad es un concepto **central y distintivo** de la Estadística frente a otras disciplinas. Mientras que en las ciencias deterministas (como la física clásica) se buscan leyes que describen fenómenos sin ambigüedad, la Estadística se ocupa de fenómenos **inciertos y aleatorios**, donde no hay una respuesta única, sino distribuciones de respuestas.

Algunas razones clave por las que la variabilidad es fundamental en Estadística:

* **Sin variabilidad no habría estadística**: si todos los datos fueran iguales, no necesitaríamos técnicas estadísticas para analizarlos.
* **Permite entender la incertidumbre**: conocer cómo varían los datos es esencial para hacer inferencias válidas.
* **Está presente en todo el proceso estadístico**: desde la recogida de datos hasta la interpretación de resultados.
* **Ayuda a distinguir entre patrones reales y fluctuaciones aleatorias**: comprender si una diferencia observada es significativa o simplemente fruto del azar depende de analizar la variabilidad.

# **Dificultades en la comprensión del concepto de variabilidad**

Según autores en educación estadística, los estudiantes presentan dificultades para comprender el concepto de **variabilidad** debido a diversas razones interrelacionadas, tanto **cognitivas** como **didácticas**:

**1. Primacía de lo determinista en la educación matemática previa**

La enseñanza tradicional de las matemáticas está centrada en:

* **Relaciones exactas y resultados únicos**, sin considerar la incertidumbre.
* **Problemas cerrados** con una única respuesta correcta.

Esto **contrasta radicalmente con la estadística**, donde:

* Se trabaja con **datos reales** y **resultados variables**.
* El foco no está solo en “cuánto vale” sino en “cuánto varía”.

Así, los estudiantes llegan a la estadística **con un marco de pensamiento determinista** que dificulta asumir la variabilidad como algo normal y valioso.

**2. Énfasis excesivo en los promedios**

En muchos currículos escolares, se enseña la estadística reducida a:

* El cálculo de **medidas centrales** (media, mediana, moda).
* La interpretación de gráficos simples.

De modo que la variabilidad suele tratarse superficialmente, o bien se omite del todo.

Todo esto lleva al alumnado a una **falsa idea de que una media resume suficientemente un conjunto de datos**, y no entran a reflexionar sobre la dispersión de los mismos.

**3. Naturaleza abstracta del concepto de variabilidad**

* La variabilidad **no es una cantidad fija**: es un patrón que emerge de comparar datos entre sí.
* Requiere **pensamiento comparativo y relacional**: ¿es esta diferencia “grande”? ¿está dentro de lo esperable?
* Cuantificarla (varianza, desviación típica) implica cálculos que no tienen una interpretación tan intuitiva y directa como otros parámetros.

A diferencia de “sumar” o “contar”, **variar** es una acción menos concreta cognitivamente.

**4. Falta de experiencias directas con datos reales y aleatoriedad**

Muchos estudiantes, por ejemplo:

* No han generado o recolectado jamás datos propios.
* No han vivido procesos donde **el mismo experimento genera resultados distintos**.
* No han visualizado con simulaciones cómo la variabilidad **aparece incluso cuando todo parece igual**.

Sin estas vivencias, es difícil construir una imagen mental rica del concepto.

**5. Confusión entre error y variabilidad**

Algunos estudiantes asocian la variabilidad con "error" o "fallo", y por tanto tienden a **minimizarla o ignorarla** en lugar de interpretarla como **parte esencial del fenómeno observado**.

Por ello viversos autores (citar), abogan por

* Un **enfoque didáctico** que incluya actividades centradas específicamente en la variabilidad.
* La **utilización de gráficos, simulaciones y contextos reales**, que hagan visible esta variabilidad y permitan trabajarla en profundidad.
* La **insistencia en la idea de que sin variabilidad no hay necesidad de Estadística**, lo que resalta su papel fundacional en la disciplina.

Fomentar una comprensión profunda de la variabilidad es clave para desarrollar un verdadero **pensamiento estadístico**, tanto en el alumnado como en el profesorado.

# **Propuestas didácticas basadas en herramientas visuales e interactivas**

Las **herramientas visuales e interactivas** pueden ser decisivas para mejorar la comprensión de la *variabilidad* en el aula, ya que permiten **hacer visible lo que normalmente permanece oculto tras los resúmenes numéricos**. En sintonía con la propuesta de Batanero, este tipo de recursos permiten llevar a cabo un enfoque más profundo, experiencial y significativo. A continuación, se detalla cómo contribuyen a cada uno de los tres ejes que plantea:

**1. Actividades centradas específicamente en la variabilidad**

Las herramientas visuales permiten **centrar la atención del alumnado en la dispersión y no solo en la media**. Algunos ejemplos:

* **Diagramas interactivos, de puntos o de cajas**: al modificar los datos, se observa cómo cambia (o no) la media y cómo se amplía o reduce la dispersión.
* **Medidas de dispersión dinámicas**: representar visualmente la **desviación típica** como "barras" de distancia respecto a la media.
* **Comparación de grupos**: dos conjuntos con la misma media, pero distinta variabilidad (por ejemplo, con simulaciones de lanzamientos o notas de exámenes).

Esto permite desarrollar intuiciones del tipo:  
"Aquí la media es la misma, pero hay más variabilidad, por tanto, hay más incertidumbre en los resultados individuales."

**2. Gráficos, simulaciones y contextos reales para *hacer visible* la variabilidad**

Este tipo de recursos **rompen la abstracción** del concepto y lo conectan con la experiencia:

* **Simulaciones de fenómenos aleatorios** (e.g., lanzar monedas, tirar dados, medir tiempos): muestran cómo se generan diferentes resultados *bajo las mismas condiciones*, ilustrando la **variabilidad intrínseca** de los procesos reales.
* **Animaciones de muestreo**: permiten ver cómo varían las medias y desviaciones de muestras diferentes tomadas de la misma población.
* **Juegos o applets** donde el alumnado debe **predecir o controlar la variabilidad** (por ejemplo, ajustando la dispersión de una distribución para ganar un juego de puntería).
* **Datos reales con variabilidad evidente**: series meteorológicas, resultados deportivos, precios de productos, ...

Estos recursos ayudan a pasar **de lo numérico a lo visual, de lo estático a lo dinámico, y de lo abstracto a lo vivido**.

**3. Insistir en que sin variabilidad no hay necesidad de estadística**

Las herramientas interactivas permiten experimentar con esta idea de forma directa:

* **Situaciones con variabilidad nula**: mostrar conjuntos donde todos los datos son iguales y el análisis estadístico se vuelve trivial.
* **Modelos donde introducir variabilidad cambia el problema**: por ejemplo, en decisiones bajo incertidumbre o en comparación de tratamientos.

Al mostrar cómo la Estadística **cobra sentido solo cuando hay incertidumbre o dispersión**, estas herramientas fortalecen una comprensión epistemológica de la disciplina.

REVISAR HERRAMIENTAS EXISTENTES

# **Propuesta didáctica: visualizar la descomposición de la variabilidad mediante ANOVA con R**

Esta propuesta se enmarca en la asignatura **Modelos Estadísticos para la Toma de Decisiones II (MET II)**, impartida en el **Grado en Ciencia de Datos** de la **Universitat Politècnica de València**. En ella, se introducen modelos estadísticos paramétricos con aplicaciones en contextos reales, siendo el **Análisis de la Varianza (ANOVA)** uno de los bloques clave del temario.

El ANOVA es una herramienta estadística especialmente útil para comprender cómo se descompone la **variabilidad total de un fenómeno** en distintas fuentes: variabilidad explicada por factores controlados y variabilidad residual. No obstante, su tratamiento habitual —centrado en el cálculo de sumas de cuadrados y el contraste de hipótesis— puede resultar abstracto y mecánico para los estudiantes, dificultando la comprensión conceptual de la variabilidad.

Una **comprensión profunda del concepto de variabilidad** es esencial para entender el fundamento del **Análisis de la Varianza (ANOVA)**. Lejos de ser sólo una técnica de contraste de medias, el ANOVA se basa en una idea central: **la variabilidad total de los datos puede dividirse en partes atribuibles a distintas fuentes**. Esta descomposición no sólo permite evaluar si un factor tiene efecto sobre una variable, sino que ilustra de forma clara cómo y por qué **la estadística estudia la variabilidad**.

La ecuación fundamental del ANOVA lo expresa de manera sencilla:

SCT=SCF+SCR

donde:

* **SCT (Suma de Cuadrados Total)** mide la variabilidad global de los datos respecto a la media general.
* **SCF (Suma de Cuadrados del Factor)** representa la parte de esa variabilidad explicada por las diferencias entre los grupos (por ejemplo, distintos tratamientos, proveedores, condiciones...).
* **SCR (Suma de Cuadrados Residual)** es la variabilidad restante: la que no se explica por el factor, atribuida al azar o a otros elementos no considerados.

Desde un punto de vista **visual** e **intuitivo**, esta ecuación puede interpretarse como:

* **SCT**: qué tan dispersos están los datos con respecto a la media general (una nube global de puntos).
* **SCF**: cuánto se alejan las medias de cada grupo de esa media general (diferencias entre “líneas de grupo” y la línea horizontal global).
* **SCR**: cuánto se dispersan los datos dentro de cada grupo respecto a su propia media (la anchura vertical de cada grupo).

Por ello, se propone el uso de **herramientas visuales e interactivas desarrolladas con R y Shiny**, que permitan al alumnado:

* Explorar dinámicamente cómo cambia la **suma de cuadrados total (SCT)** en función de los datos existentes.
* Visualizar la descomposición de la SCT (total) en **SCF (factor) + SCR (residual)**.
* Observar el impacto de diferentes configuraciones de datos (variabilidad entre grupos, dentro de grupos, presencia de valores atípicos).
* Relacionar los resultados del **test F** con la representación gráfica de la variabilidad.

**Objetivos didácticos**

La propuesta tiene como finalidad:

* Facilitar la comprensión conceptual del **papel de la variabilidad** en la inferencia estadística.
* Asociar gráficamente los componentes del modelo ANOVA con sus fórmulas algebraicas.
* Fortalecer el pensamiento estadístico mediante la experimentación directa con datos.

**Características de la herramienta**

La app, desarrollada en R y Shiny, incluirá:

* **Carga de datos propios** o selección de conjuntos predefinidos.
* **Visualización simultánea** de:
  + Gráficos de puntos por grupo.
  + Líneas de medias de grupo y media global.
  + Áreas que representan SCF, SCR y SCT.
* **Interacción con los datos**: modificar valores, añadir ruido, introducir o eliminar valores atípicos.
* **Resultados del ANOVA actualizados en tiempo real**, acompañados de explicaciones interpretativas adaptadas al nivel del alumnado.

**Ejemplo aplicado: el caso de los proveedores**

Uno de los ejemplos incluidos es el del **análisis del equilibrado dinámico de cigüeñales suministrados por distintos proveedores**, extraído del manual docente de MET I y MET II (Romero & Zúnica, 2013). Este caso permite observar:

* Variabilidad total entre las observaciones.
* Diferencias entre las medias de los proveedores (SCF).
* Dispersión interna dentro de cada proveedor (SCR).
* Resultado del test F e interpretación de su significado.
* Importancia de los valores atípicos en el resultado del test F.

# **Conclusiones**

# **Trabajos futuros**

# **Referencias**