

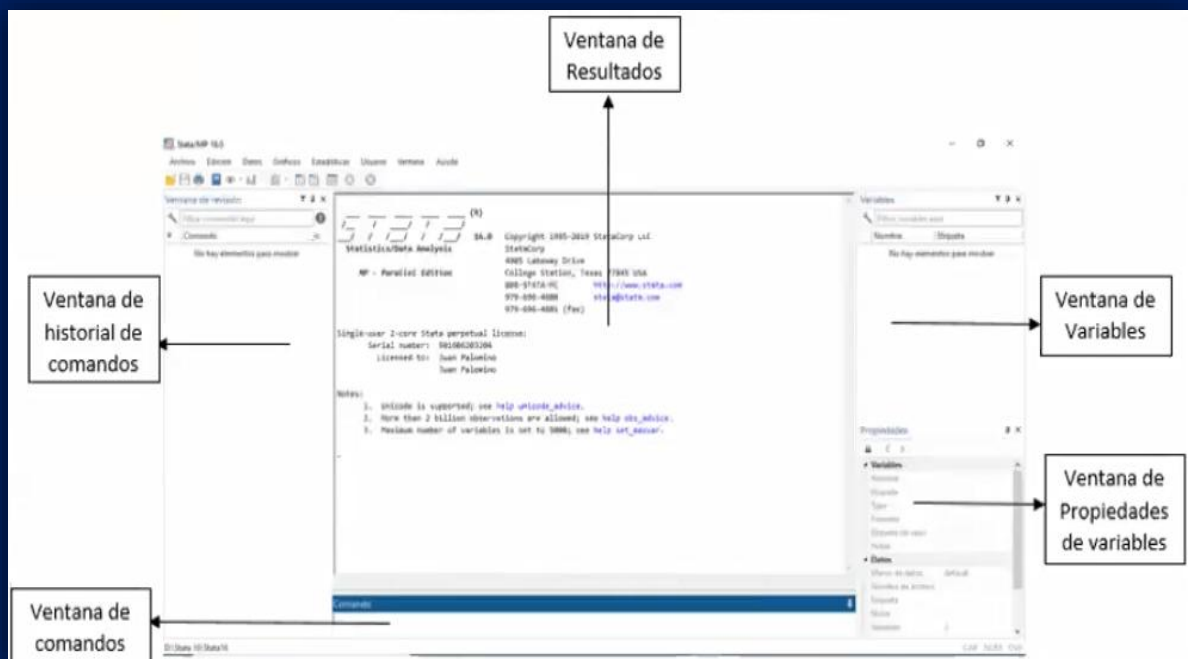
# EL ABC DE

# STATA

**MSc. Ciro Ivan Machacuay Meza**

*STATA es un paquete de software estadístico creado en 1985 por Stata Corp. Es utilizado principalmente por instituciones académicas y empresariales dedicadas a la investigación, especialmente en economía, sociología, ciencias políticas, biomedicina, geoquímica y epidemiología, etc.*

*STATA permite, entre otras funcionalidades la gestión de datos, el análisis estadístico, el trazado de graficas y los simuladores. Tiene la ventaja de ejecutar los análisis y tareas mediante pestañas desplegadas o mediante comandos que lo hace muy amigable*



*Tipos de Datos en Stata*

Stata tiene dos categorías principales de tipos de datos: numéricos y de texto (string). La elección del tipo correcto es crucial para la eficiencia de la memoria y la precisión de los cálculos.

*1. Datos Numéricos*

Los tipos numéricos se dividen en **enteros** y **de punto flotante**. Su principal diferencia es el rango de valores que pueden almacenar y la cantidad de memoria que utilizan.

*A. Enteros (Integer)*

Almacenan números enteros (sin decimales). Son más eficientes en memoria que los de punto flotante.

Tipo	Bytes de Memoria	Rango (límites inclusive)	Uso Típico
byte	1	-127 to 100	Variables categóricas o indicadoras (0, 1), edades de niños, años pequeños.
int	2	-32,767 to 32,740 (aproximadamente $\pm 32,700$ )	Edades, años (ej. 1990, 2020), conteos medianos (ej. número de habitaciones).
long	4	-2,147,483,647 to 2,147,483,620 (aproximadamente $\pm 2.14e9$ )	Conteos muy grandes (ej. población de un país grande en unidades), ID de números grandes.

*B. Punto Flotante (Floating Point)*

Almacenan números reales (con decimales). Son esenciales para cálculos científicos y financieros, pero usan más memoria.

Tipo	Bytes de Memoria	Precisión (dígitos significativos)	Uso Típico
float	4	~6-7 dígitos	Precios, medidas corporales (peso, altura), porcentajes. Suficiente para la mayoría de aplicaciones.
double	8	~15-16 dígitos	Cálculos científicos y financieros de alta precisión, coordenadas GPS muy detalladas. Evita errores de redondeo en cálculos complejos.

*Nota: Usa float a menos que necesites explícitamente la precisión de double. Un double ocupa el doble de memoria que un float.*

*2. Datos de Texto (String)*

Característica	Descripción
Definición	Almacenan texto: letras, números, símbolos, espacios. Todo lo que no sea un número para hacer cálculos.
Longitud	Longitud fija. Cuando se crea una variable string, se debe definir su longitud máxima (ej. str5, str20, str244).
Memoria	Cada carácter ocupa 1 byte de memoria. Una variable str20 ocupa 20 bytes por observación, siempre.
Valores Missing	Se representan con "" (una cadena vacía).

### *Ejemplo de longitud fija:*

Si defines una variable como str5 ("hola") y almacenas la palabra "hola" (4 caracteres), Stata guardará "hola " (agregando un espacio para completar los 5 caracteres). Si intentas guardar "hola mundo" (10 caracteres), lo truncará a "hola m".

### *Resumen:*

- **byte/int/long:** Para números enteros. Elige el más pequeño que pueda contener tus valores.
- **float:** Para la mayoría de los números con decimales. Buen equilibrio entre precisión y memoria.
- **double:** Para máxima precisión en cálculos críticos.
- **str#:** Para texto, códigos alfanuméricos (ej. ID de cuestionario "A15B") o números que no se usan en cálculos (ej. teléfonos, códigos postales). Elige una longitud suficiente para evitar truncamientos.