

STATA®

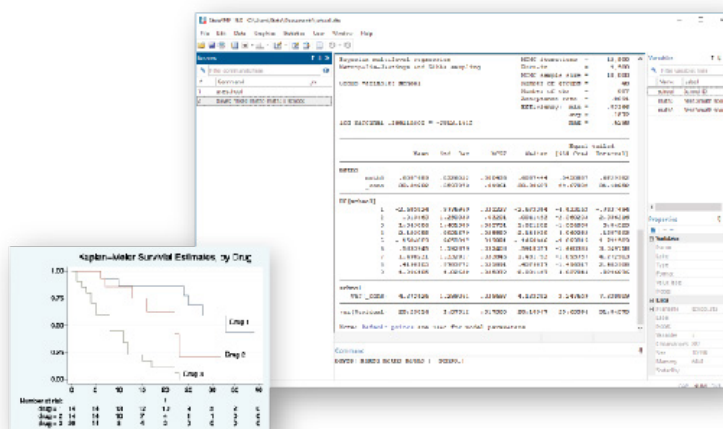
17

Características y Funcionalidades de Stata

A continuación se referencian de manera detallada más no exhaustiva las características y funcionalidades de Stata.

Interfaz de usuario gráfica

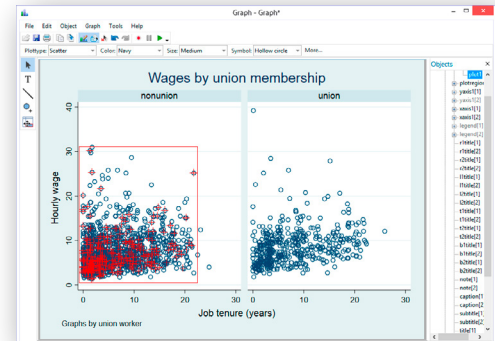
- Menús y diálogos para todas las características
- Editor de datos
- Administrador de variables
- Editor de gráficos
- Project Manager
- Editor Do-file
- Vista previa (clipboard)
- Múltiples ajustes de preferencia



Gráficos

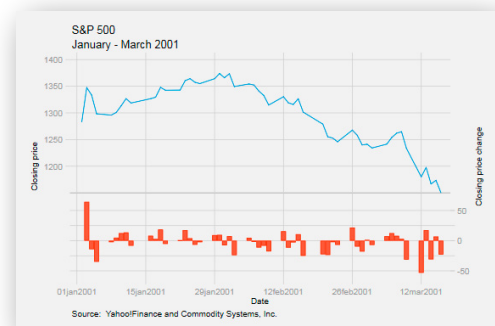
- Líneas
- Dispersión
- Barras
- Pie
- Máximo-mínimo
- Contorno
- GUI editor
- Gráficos de diagnóstico de regresión
- Gráficos de supervivencia
- Suavizadores no paramétricos
- Distribuciones Q-Q
- Geo-referenciación*

Requiere de mapas con formatos ESRI shapefiles o MapInfo Interchange Format. Disponible a través de del comando `smap` disponible en un download gratuito



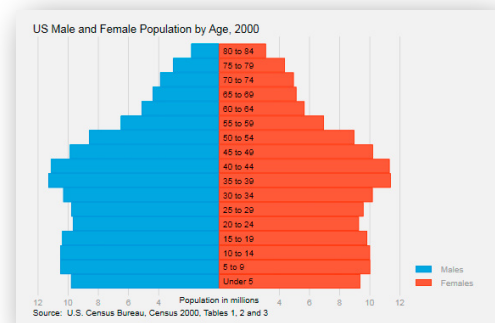
Funciones

- Estadísticas
- Números aleatorios
- Matemáticas
- Texto
- Fecha y tiempo



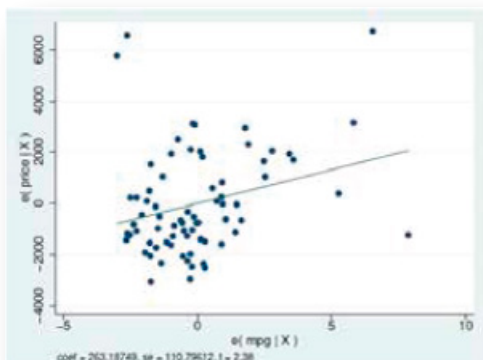
Estadísticas básicas

- Resúmenes
- Tabulaciones cruzadas
- Correlaciones
- Pruebas t
- Pruebas de igualdad de varianza
- Pruebas de proporciones
- Intervalos de confianza
- Variables de factor



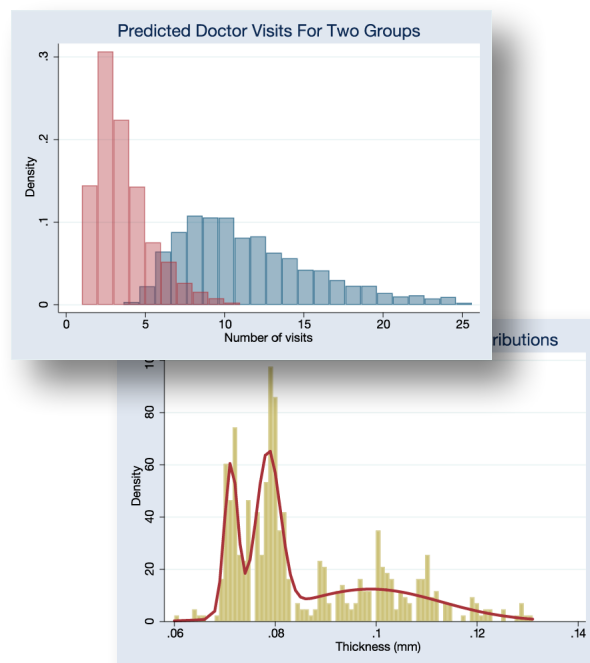
Modelos lineales

- Regresión
- Respuestas censuradas
- Regresoras endógenas
- Bootstrap, jackknife
- Regresiones robustas y robustas en cluster
- Variables instrumentales
- Mínimos cuadrados tri-etápicas
- Restricciones
- Regresión cuantílica
- Mínimos cuadrados generalizados



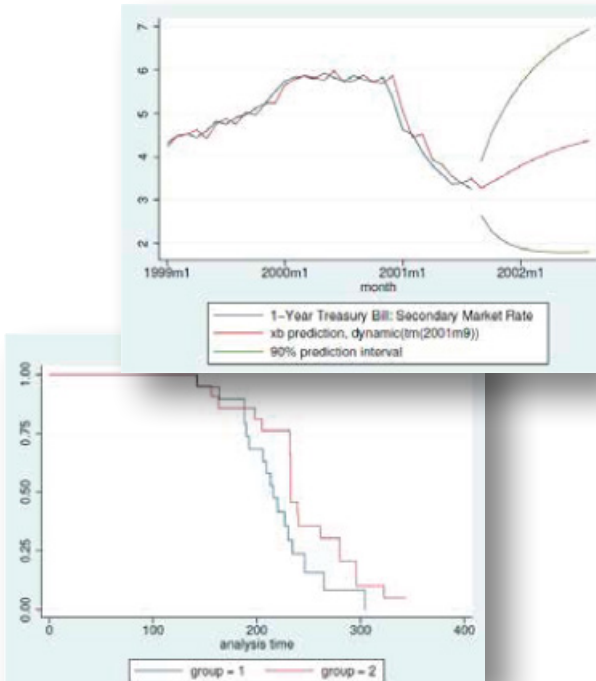
Serie de tiempo

- ARIMA
- ARFIMA
- ARCH/GARCH
- VAR
- VECM
- GARCH multivariado
- Modelos de componentes no observables
- Factores dinámicos
- Modelos de espacio-estado (space-state models)
- Calendario de negocios
- Correlogramas
- Periodogramas
- Pronósticos
- Funciones de impulso-respuesta
- Pruebas de raíz unitaria
- Filtros y suavizadores
- Estimaciones "rolling" y recursivas



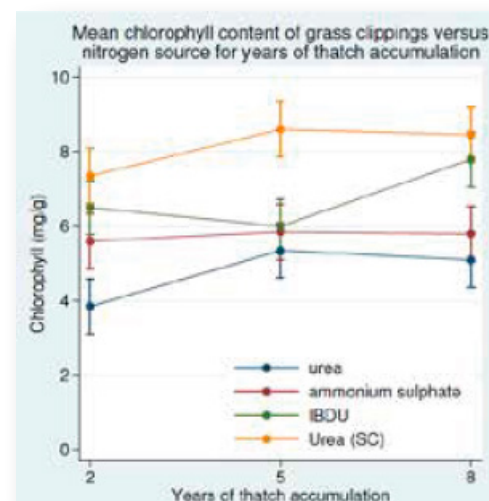
Gestión de datos

- Transformaciones de datos
- Match-merge-append
- Importación y exportación data
- ODBC
 - Importar de Oracle, SQL, Access, Excel, MySQL y DB2
 - Exportar datos a tablas ODBC nuevas o existentes
 - Ejecución personalizada de comandos SQL de manera individual o en batches
 - Personalización de strings de conexión ODBC
 - Soporte para ODBC
 - Soporte para VARCHARs/CLOBs and BLOBs
- SQL
- XML
- Ordenamiento
- Transposición
- Rotulación
- Haver Analytics database



Datos longitudinales y de panel

- Efectos fijos y aleatorios con errores estándar robustos
- Modelos mixtos lineales
- Probit de efectos aleatorios
- GEE (estimación de ecuaciones generalizada)
- Poisson de efectos fijos y aleatorios
- Modelos dinámicos de datos panel
- Variables instrumentales
- Pruebas de raíz unitaria para paneles

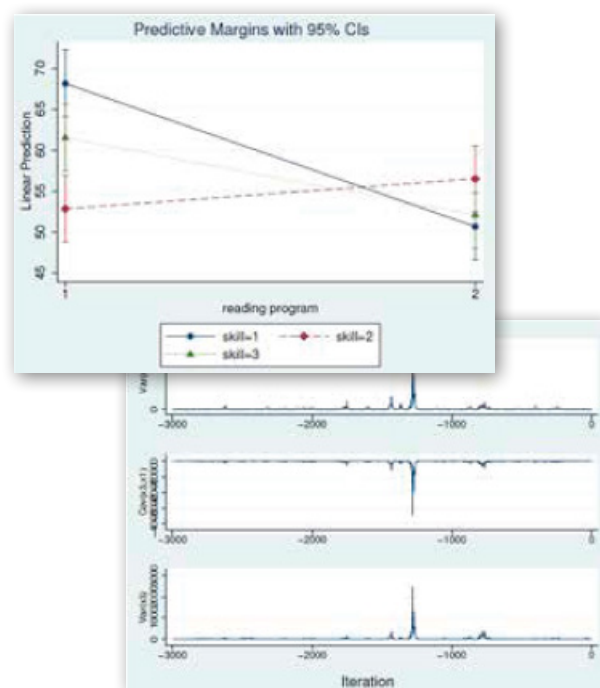


Análisis de supervivencia

- Estimadores de Kaplan-Meier y Nelson-Aalen
- Regresión Cox
- Modelos paramétricos
- Riesgo competitivo (competing risk)
- Hazards
- Covariables variantes en el tiempo
- Censura izquierda y derecha
- Modelos Weibull, Exponencial y Gompertz

Modelos multinivel y de efectos mixtos

- Continuos, binarios y de conteo
- Modelos de 2, 3 y múltiples niveles
- Modelos lineales generalizados
- Interceptos aleatorios
- Pendientes aleatorias
- Efectos aleatorios cruzados
- BLUPs de los efectos y los valores estimados
- Modelos jerárquicos
- Estructuras residual-error
- Soporte de información de encuestas en modelos lineales



```
. svydescribe
Survey: Describing stage 1 sampling units
  pweight: finalwgt
    VCE: linearized
Single unit: missing
Strata 1: stratid
SU 1: psuid
FPC 1: <zero>
```

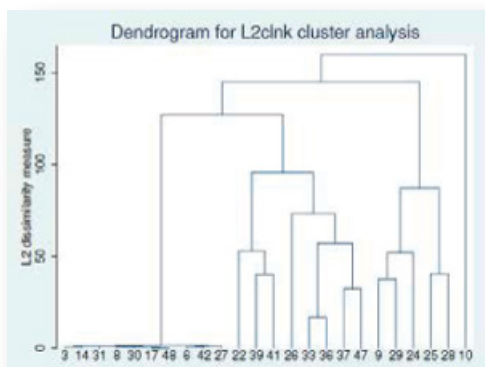
Stratum	#Units	#Obs	#Obs per Unit		
			min	mean	max
1	2	380	165	190.0	215
2	2	185	67	92.5	118
3	2	348	149	174.0	199
(output omitted)					
17	2	393	180	196.5	213
18	2	359	144	179.5	215
20	2	285	125	142.5	160
21	2	214	102	107.0	112
(output omitted)					
31	2	308	143	154.0	165
32	2	450	211	225.0	239
31	62	10351	67	167.0	288

Potencia y tamaño de la muestra

- Potencia y tamaño de la muestra
- Efecto del tamaño
- Efecto mínimo detectable
- Medias, Proporciones y Varianzas
- Correlaciones
- ANOVA
- Estudio caso-control
- Estudios de cortes
- Análisis de supervivencia
- Diseños balanceados y no balanceados
- Resultados en tablas o gráficos

Modelos de respuesta binaria, limitada y de conteo

- Logit
- Probit
- Tobit
- Poisson y binomial negativa
- Logístico condicional, multinomial, anidado, ordenada, ordenado y estereotípico
- Probit multinomial
- Modelos de conteo truncados
- Modelos de selección
- Efectos marginales

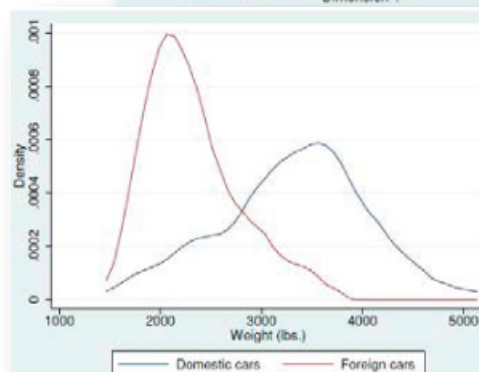
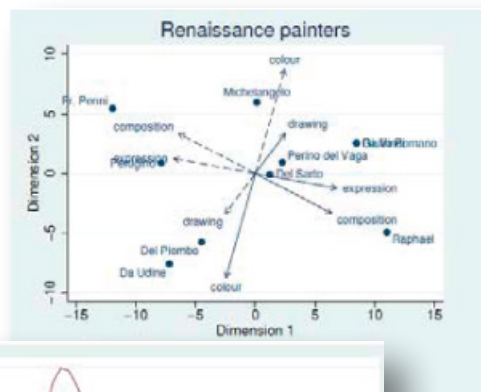


Efectos de tratamiento

- IPW (ponderación de probabilidad inversa)
- Probit
- Métodos robustos dobles
- Propensity score matching
- Ajustes de regresión
- Matching de covariables
- Tratamientos multinivel
- ATE (efectos promedio del tratamiento)
- ATETs (efectos promedio del tratamiento en el tratado)
- POMs (potential-outcome means)

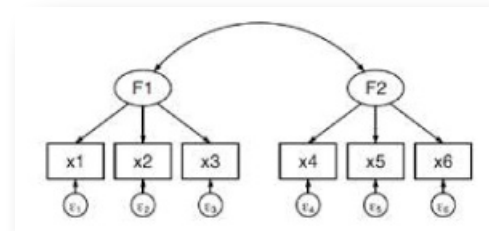
Modelos lineales generalizados (GLMs)

- 10 funciones de vinculación
- Vínculos definidos por el usuario
- Estimación ML e IRLS
- 9 estimadores de varianza
- 7 residuos



Modelación de ecuaciones estructurales (SEM)

- Constructor gráfico de diagramas
- Estimaciones estandarizadas y no estandarizadas
- Índices de modificación
- Efectos directos e indirectos
- Respuesta continuas, binarias, de conteo y ordinales
- Modelos multinivel
- Interceptos y pendientes aleatorias
- Factors scores
- Predictores empíricos de Bayes y otras
- Grupos y pruebas de invariancia
- Bondad de ajuste
- Datos MAR
- Datos correlacionados



```
2.      if mod('x',2) {  
3.          display "'x' is odd"  
4.          continue  
5.      }  
6.      display "'x' is even"  
7. }  
1 is odd  
2 is even  
3 is odd  
4 is even
```

Otros métodos estadísticos

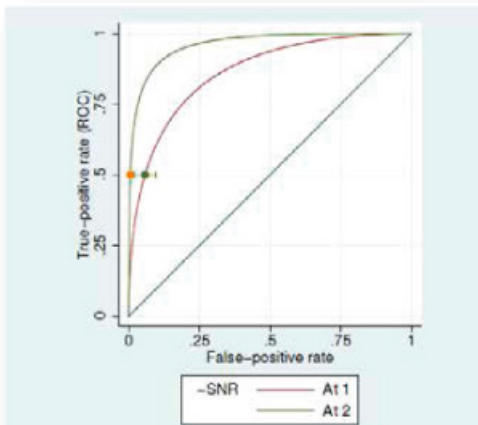
- Medida de Kappa para acuerdos
- Alfa de Cronbach
- Regresión stepwise
- Pruebas de normalidad

ANOVA/MANOVA

- Diseños balanceados y no balanceados
- Diseños factoriales, anidados y mixtos
- Medidas repetidas
- Medias marginales
- Contrastes

Imputación múltiple

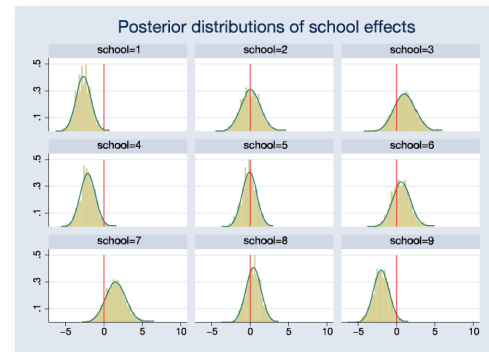
- 9 métodos univariados de imputación
- Imputación normal multivariada
- Ecuaciones encadenadas
- Exploración de patrones de pérdida
- Gestión de bases de datos imputadas



- Ajuste de modelos y resultados agrupados
- Parámetros de transformación
- Pruebas conjuntas de las estimaciones de los parámetros
- Predicciones

Estadísticas exactas

- Regresiones Logística y de Poisson exactas
- Estadísticas exactas para caso-control
- Pruebas binomiales
- Prueba exacta de Fisher para tablas $r \times c$



Pruebas, predicciones y efectos

- Tests de Wald
- LR tests
- Combinaciones lineales y no lineales
- Predicciones y predicciones generalizadas
- Medias marginales
- Medias ajustadas
- Efectos marginales y parciales
- Modelos de pronóstico
- Test de Hausman

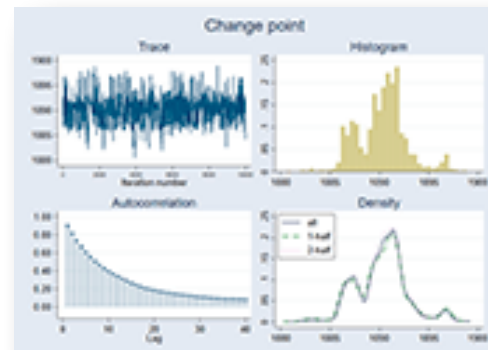
Métodos de encuesta

- Diseños multietápicos
- Estimación de varianza Bootstrap, BRR, jackknife, linealizada y SDR
- Post-estratificación
- DEFF
- Márgenes predictivos
- Medias, proporciones, razones y totales
- Tablas de resumen
- Regresión de variables instrumentales, Probit y Cox



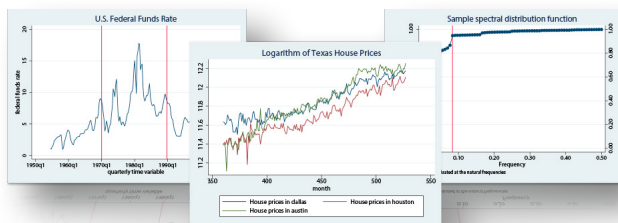
Contrastes, comparaciones pareadas y márgenes

- Comparación de medias, interceptos o pendientes
- Comparación respecto una categoría de referencia
- Categoría adyacente, gran media, etc.
- Polinomios ortogonales
- Ajustes de comparación múltiple
- Gráficos de medias estimadas y contrastes
- Gráficos de interacción



Análisis de conglomerados (cluster analysis)

- Conglomerados jerárquicos
- Agrupamiento no jerárquico para Kmeans y kmedian
- Dendogramas
- Reglas de parada
- Análisis extensibles al usuario

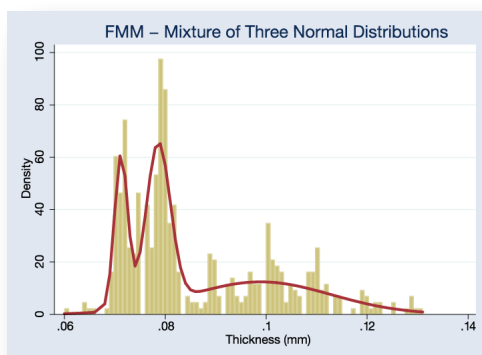


Métodos multivariados

- Análisis factorial
- Componentes principales
- Análisis discriminante
- Rotación
- Escalamiento multidimensional
- Biplots
- Dendogramas
- Análisis extensibles al usuario

GMM y regresión no lineal

- Análisis factorial
- GMM (método generalizado de momentos)
- Regresión no lineal

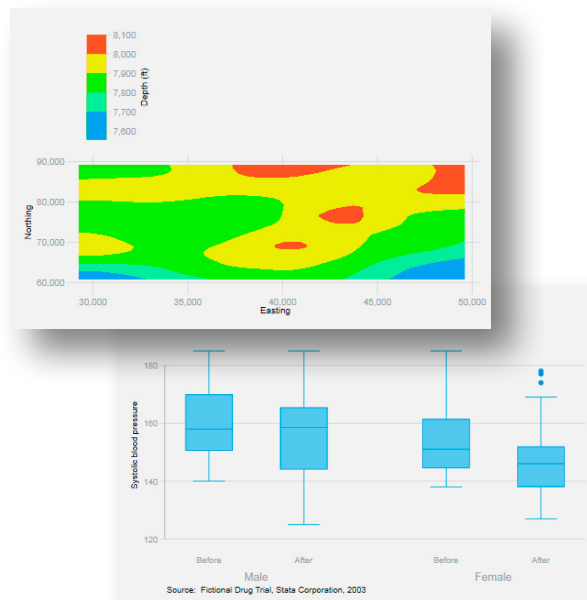


Máxima verosimilitud simple

- Uso de expresiones simples para especificar verosimilitud
- No se requiere programación
- Datos de encuestas
- Estimadores matriciales

Métodos no paramétricos

- Wilcoxon-Mann-Whitney
- Rango de signos de Wilcoxon
- Kruskal-Wallis test
- Correlaciones de Spearman y Kendall
- Pruebas de Kolmogorov-Smirnov
- Intervalos de confianza binomiales exactos
- Datos de supervivencia
- Análisis ROC
- Suavizamiento
- Bootstrapping

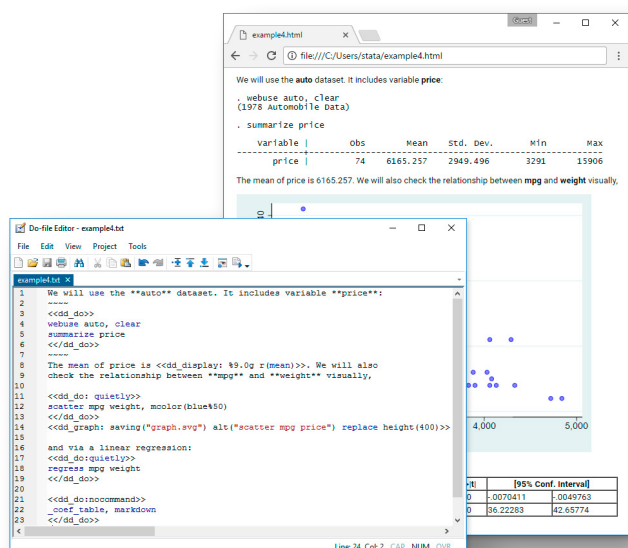


Remuestreo y métodos de simulación

- Bootstrap
- Jackknife
- Simulación de Monte Carlo
- Pruebas de permutación

Epidemiología

- Estandarización de tasas
- Caso-control
- Cohortes
- Matched caso-control
- Mantel-Haenszel
- Farmacocinética
- Análisis ROC
- ICD-9C-M



Característica de programación

- Adición de nuevos comandos
- Command scripting
- Programación orientada a objetos
- Menú y cuadros de diálogo para programación
- Project Manager
- Plugins

Programación de matrices – Mata

- Sesiones interactivas
- Desarrollo de proyectos de gran escala
- Optimización
- Inversión de matrices
- Descomposiciones
- Eigenvalores y eigenvectores
- Motor LAPACK
- Números reales y complejos
- Matrices de texto
- Interface a bases de datos de Stata y matrices
- Derivaciones numéricas
- Programación orientada a objetos

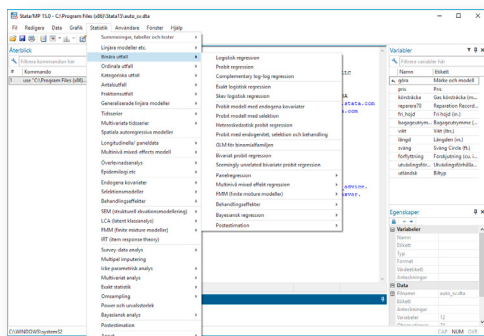
```

1  We will use the **auto** dataset. It includes variable **price**
2  ----
3  <<dd do>>
4  webuse auto, clear
5  summarize price
6  <</dd do>>
7  ----
8  The mean of price is <<dd display: %9.0g r(mean)>>. We will also
9  check the relationship between **mpg** and **weight** visually,
10
11 <<dd do: quietly>>
12 scatter mpg weight, mcolor(blue*50)
13 <</dd do>>
14 <<dd_graph: saving("graph.svg") alt("scatter mpg price") replace height(400)>>
15
16 and via a linear regression:
17 <<dd do: quietly>>
18 regress mpg weight
19 <</dd do>>
20
21 <<dd do: nocommand>>
22 _coef_table, markdown
23 <</dd do>>

```

Percentiles		Smallest			
1%	1.47	4.13			
5%	2.75	1.43			
10%	2.90	1.3	Obs	526	
25%	3.30	1.3	Sum of Wgt.	526	
50%	4.85		Mean	5.896103	
75%	6.30	Largest	Std. Dev.	3.493086	
90%	10	21.20	Variance	12.43888	
95%	17	22.84	Skewness	3.067326	
99%	20	24.30	Kurtosis	7.970023	

Percentiles		Smallest			
1%	4	0			
5%	8	0			
10%	9	0	Obs	526	
25%	11	0	Sum of Wgt.	526	
50%	12		Mean	12.56274	
75%	14	10	Std. Dev.	3.769022	
90%	16	10	Variance	7.667665	
95%	17	10	Skewness	-4.195761	
99%	19	10	Kurtosis	4.814246	



Capacidades de internet

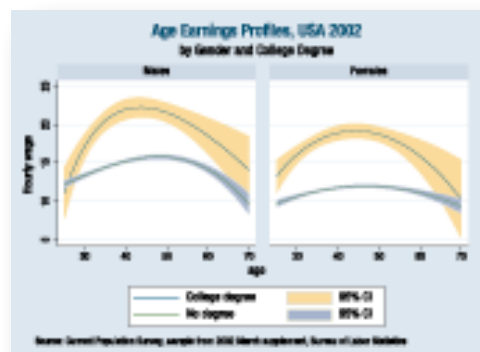
- Habilidad para la instalación de nuevos comandos
- Actualización web
- Noticias de Stata recientes
- Compartir bases de datos y programas
- Ayuda

Accesibilidad

- Acceso a personas con discapacidades

Documentación

- 27 manuales (Más de 14,000 páginas, todos en PDF e interactivos)
- Navegación por hipervínculos
- Miles de ejemplos trabajados
- Métodos y fórmulas
- Referencias



Tablas

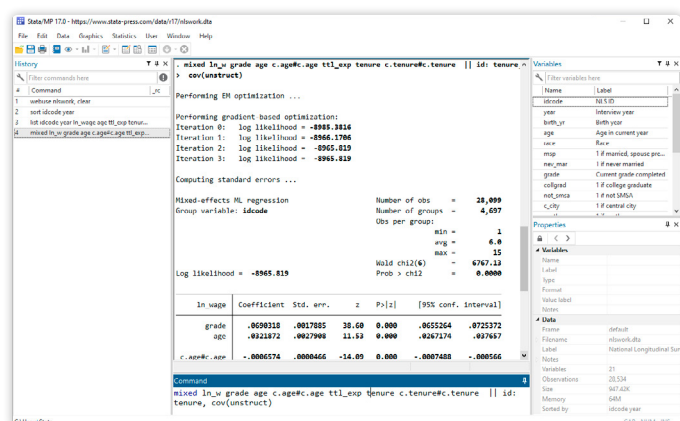
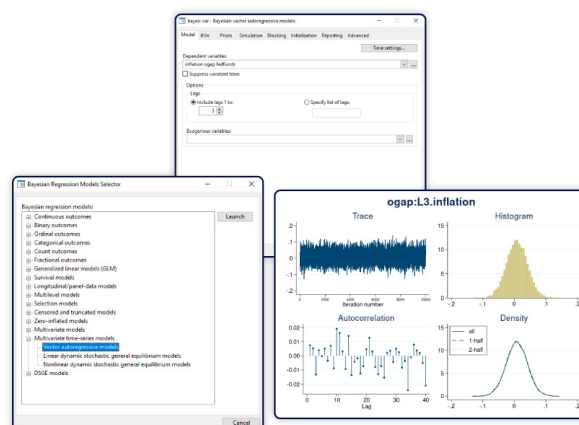
Se incluye un módulo para crear fácilmente tablas que comparan resultados de regresión o estadísticas de resumen, en donde es posible crear estilos de presentación y aplicarlos, además, se mejoran las opciones de exportación de tablas a MS Word®, PDF, HTML, LaTeX, MS Excel® o Markdown para incluirlos en informes.

Econometría bayesiana

Uno de los atractivos del uso de métodos bayesianos en la modelización econométrica es incorporar la información sobre los parámetros del modelo, esta información puede provenir de datos históricos, o del conocimiento de un proceso económico.

Las nuevas funciones de Econometría Bayesiana de Stata se ajustan a muchos modelos bayesianos, como modelos de corte transversal, datos de panel, multinivel y series de tiempo. Compare modelos utilizando factores de Bayes y genere pronósticos. Estas son algunas de las funciones que encontrará en esta versión:

- Modelos Bayesianos VAR.
- Análisis bayesiano de IRF y FEVD.
- Pronóstico dinámico bayesiano.
- Modelos de datos de panel / longitudinales bayesianos.
- Modelos DSGE lineales y no lineales bayesianos.



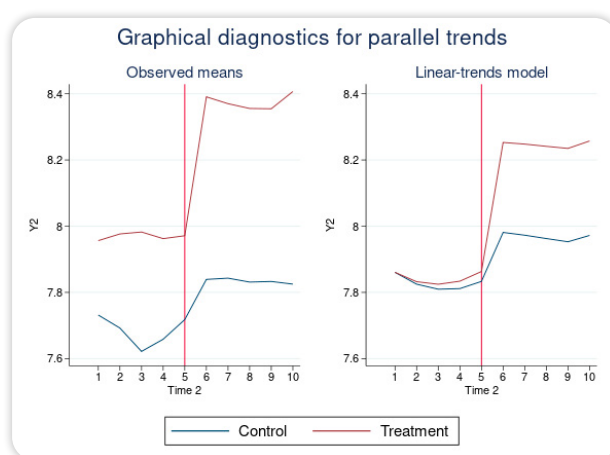
Mayor rapidez en Stata

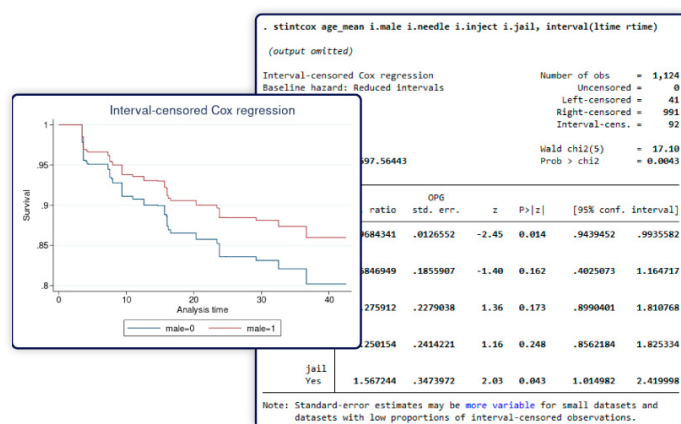
Stata se caracteriza por la velocidad y precisión, Stata 17, ofrece lo mejor de ambas opciones, en esta versión se han incluido mejoras y actualización en los algoritmos **sort** y **collapse**. De igual forma, se ofrecen novedades en algunos comandos de estimación, por ejemplo para modelos multinivel con efectos mixtos.

Modelos de diferencia en diferencias

En esta versión aparecen los comandos de estimación `didregress` y `xtdidregress` que se ajustan a modelos de diferencias en diferencias (DID) y diferencias triples (DDD) con datos de medidas repetidas.

Así, el comando `didregress` permite el trabajo con datos transversales repetidos y el comando `xtdidregress` facilita opciones para datos longitudinales o panel.



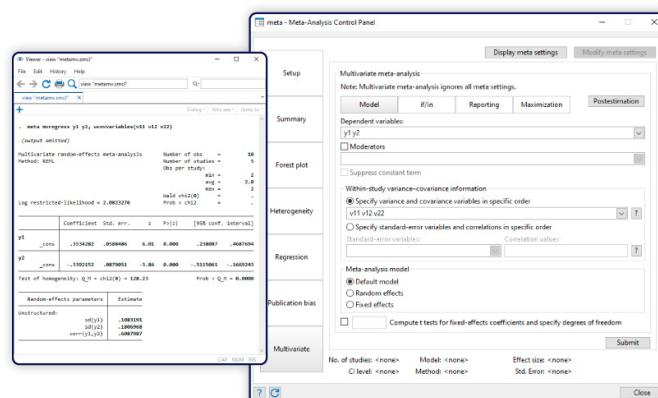


Modelo de regresión COX para intervalos

Stata 17, incluye el comando de estimación `stntcox`, una opción que ajusta el modelo Cox a datos de tiempo censurados por intervalos, este procedimiento es útil cuando el tiempo de un evento de interés no es observado directamente, pero se sabe que se encuentra dentro de un intervalo. Esta nueva versión permite hacer los ajustes necesarios para revisar del mejor modo sus datos de interés.

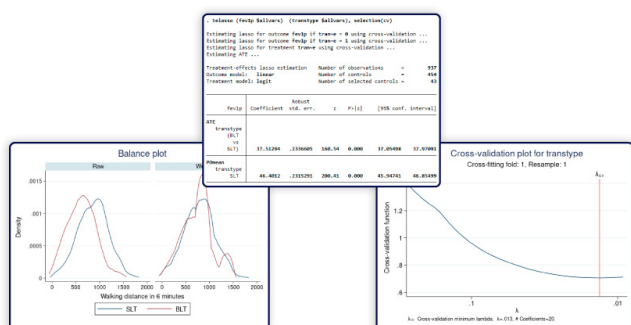
Meta - análisis multivariado

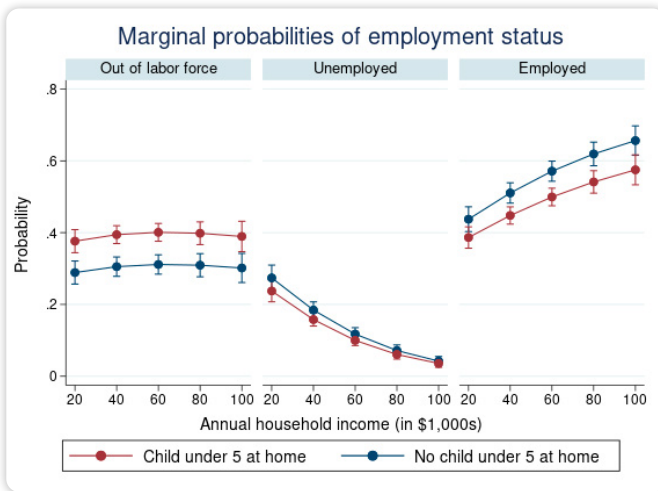
Los estudios presentan múltiples tamaños de efecto, en general es probable que estén correlacionados. Hasta ahora, era posible en Stata trabajar con opciones de Meta-análisis de forma general y con efectos separados, Stata 17 incluye el comando `mvregress` para realizar un meta-análisis multivariado que facilite la revisión de la correlación de sus datos.



Estimación Lasso: tratamiento-efecto

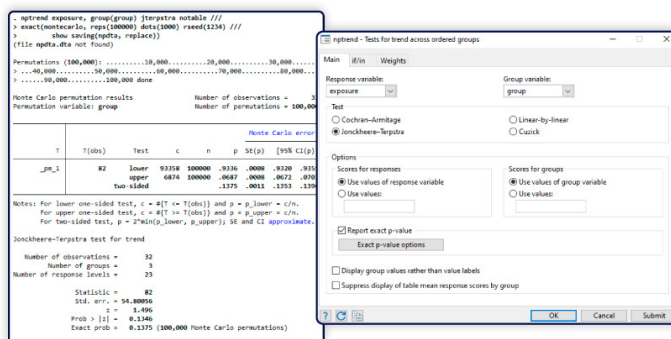
Usar la estimación Lasso le permite la revisión de cientos de covariables, así mismo, **teffects** permite estimar los efectos del tratamiento aplicado, en Stata 17, con la nueva función **telasso** podrá estimar el tratamiento-efecto y control de muchas covariables.





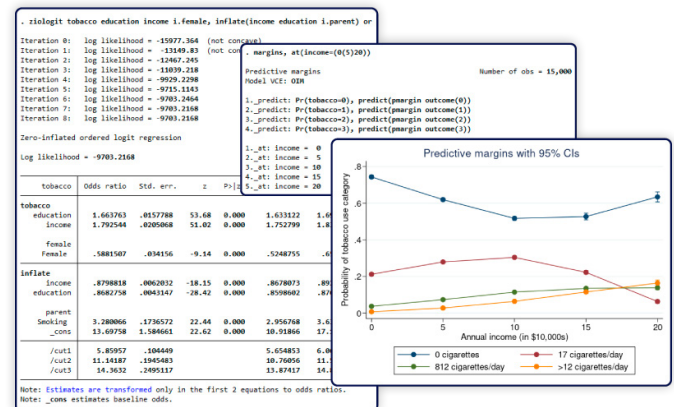
Modelo logit ordenado con ceros inflados

El nuevo comando de estimación *ziologit* se ajusta a los modelos de regresión logística ordenados con ceros inflados. La inflación cero se contabiliza asumiendo que estos provienen tanto de un modelo logístico como de un modelo logístico ordenado, cada uno puede tener diferentes covariables y los resultados se pueden mostrar como razones de probabilidades en lugar de los coeficientes predeterminados.



Modelo logit multinomial para datos de panel

El nuevo comando de estimación de Stata, *xtmlogit*, ajusta los modelos logit multinomial (MNL) de datos de panel a resultados observados a lo largo del tiempo, permite modelar las características individuales directamente y, por lo tanto, puede producir resultados más eficientes, así mismo, puede dar cuenta adecuadamente de las características que podrían estar correlacionadas con las covariables.

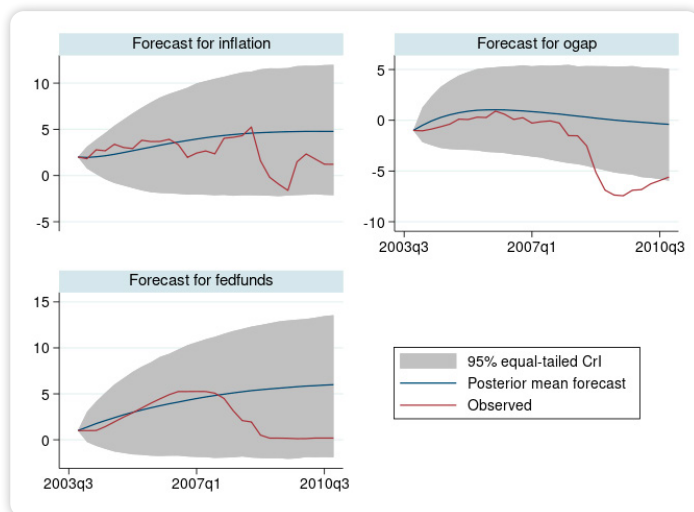


Pruebas no paramétricas a partir de la tendencia

El comando *nptrend* ahora admite cuatro pruebas de tendencia en grupos ordenados, por lo que podrá elegir entre la prueba de Cochran-Armitage, la prueba de Jonckheere-Terpstra, la prueba de tendencia lineal por lineal y la prueba de Cuzick usando rangos. Las primeras tres pruebas son nuevas y la cuarta prueba venía incluida en *nptrend* en versiones anteriores.

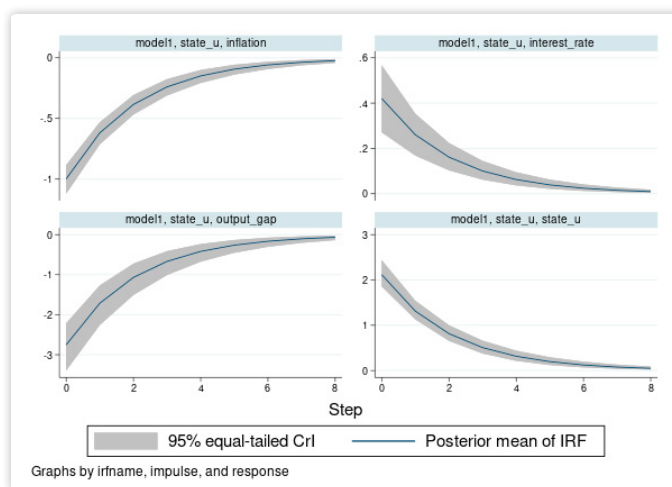
Pronóstico dinámico bayesiano

El pronóstico dinámico es una herramienta usualmente empleada luego de ajustar modelos de series de tiempo multivariantes, como modelos vectoriales autorregresivos (VAR). Stata 17 incluye el comando **bayesfcst** para calcular pronósticos dinámicos bayesianos después de ajustar un modelo VAR bayesiano. Los pronósticos dinámicos bayesianos producen una muestra completa de valores predichos en lugar de una sola predicción como en el análisis clásico. Esta muestra se puede utilizar para responder a varias preguntas de modelado, como qué tan bien el modelo predice observaciones futuras sin hacer el supuesto de normalidad asintótica.

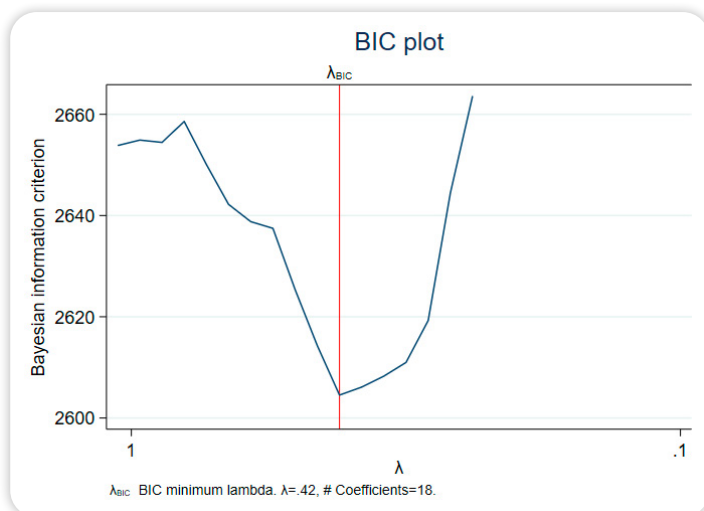


Análisis Bayesiano IRF y FEVD

Las funciones de impulso-respuesta (IRF), funciones multiplicadoras dinámicas y varianza del error de pronóstico y descomposiciones (FEVD) se utilizan comúnmente para describir los resultados de series de tiempo multivariadas, por su parte, los modelos VAR tienen muchos parámetros que pueden ser difíciles de interpretar. Los IRF y otras funciones combinan el efecto de múltiples parámetros en un resumen (por período de tiempo), de igual manera, pueden proporcionar mayor confianza en las estimaciones para pequeños conjuntos de datos porque incorporan información previa sobre los parámetros del modelo.



Selección de penalización por Lasso: BIC



La selección de un parámetro de penalización es fundamental para el análisis Lasso. La estimación Lasso proporciona varios métodos de selección de penalizaciones, incluida la validación cruzada, adaptable y plugin. En esta versión podrá utilizar el criterio de información bayesiano (BIC) para seleccionar la penalización - parámetro después de obtener una predicción Lasso y Lasso para inferencia especificando la opción de selección (BIC). De igual forma, el nuevo comando de post estimación bicplot traza los valores BIC como una función de un parámetro de penalización, con esto se obtiene una representación gráfica conveniente del valor de la penalización.

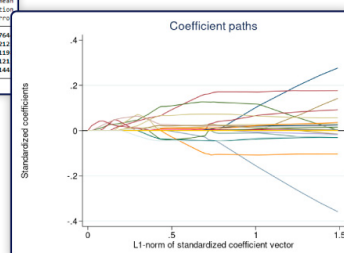
Lasso para datos agrupados

Ahora puede tener en cuenta los datos agrupados en su análisis Lasso y revisar la presencia de correlación entre las observaciones dentro del mismo grupo.

```
. lasso linear in_age $controls, cluster(idcode) rseed(1234)
20-fold cross-validation with 100 lambdaes ...
Grid value 1: lambda = .2337718 no. of nonzero coef. = 0
(output omitted)
... change in deviance stopping tolerance reached ... last lambda selected
lasso linear model      No. of obs = 28,093
                        No. of clusters = 43
Cluster: 1 idcode      No. of clusters = 4,699
Selection: Cross-validation No. of CV folds = 10
```

ID	Description	lambda	No. of nonzero coef.	Out-of- sample R-squared	CV mean predicting error
1	first lambda	.2337718	0	0.0007	.752764
10	lambda before	.0001503	24	0.3591	.582119
* 21	selected lambda	.0001359	24	0.3591	.582119
22	lambda after	.0001248	24	0.3591	.582121
30	last lambda	.0000013	24	0.3598	.582144

* lambda selected by cross-validation.



Mejoras en el editor de archivos Do: Control de navegación y marcadores mejorados

El editor de archivos Do en Stata 17 incluye las siguientes mejoras:

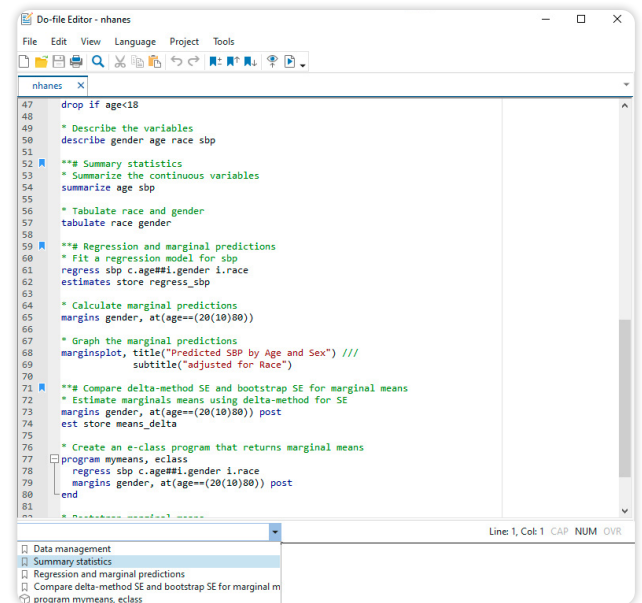
- Marcadores: ahora se guardan con los **archivos Do**.
- Nuevo control de navegación: navegue por los **archivos Do** más fácilmente.
- El soporte de resaltado de sintaxis ahora incluye Java y XML.
- Autocompletado de comillas, paréntesis y corchetes alrededor de una selección.

Marcadores:

Los marcadores se utilizan para marcar líneas de interés y volver a ellas más tarde con mayor facilidad, son particularmente útiles para navegar por **archivos Do** extensos.

Control de navegación:

Stata 17 facilita la navegación en los **archivos Do** con el nuevo Control de Navegación, que muestra una lista de marcadores y sus etiquetas.



Nuevas funciones para fechas y horas

Stata en su versión 17 ha agregado nuevas funciones para el manejo de fechas y horas tanto en Stata como en Mata.

Las nuevas funciones tienen en cuenta los años y meses bisiestos, según corresponda. Estas se pueden agrupar en tres categorías:

- **Duraciones de fecha y hora:** funciones diseñadas para obtener duraciones, como edades.
- **Fechas relativas:** funciones que devuelven fechas basadas en otras fechas, como el próximo cumpleaños.
- **Componentes de fecha y hora:** funciones que extraen diferentes componentes de los valores de fecha y hora.

What is the date?

```
. display %td today()  
20apr2021
```

How many days until summer?

```
. display datediff(today(), td(20jun2021), "day")  
61
```

How many days are in this month?

```
. display daysinmonth(tm(2021m4))  
31
```

$$A = GG'$$

$$A(A^*)A = A$$

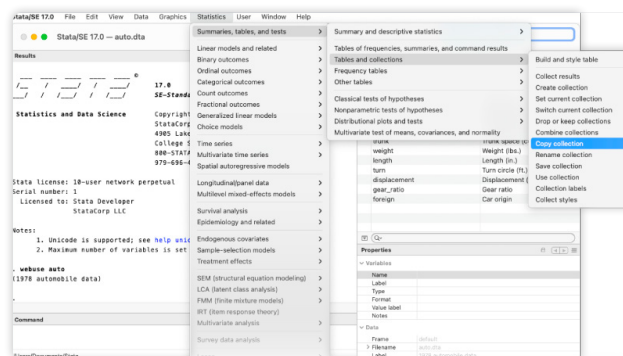
$$(A - \lambda I)x = 0$$

Intel Math Kernel Library (MKL)

Stata 17 introduce el uso de Intel Math Kernel Library (MKL) en hardware compatible (para cualquier procesador Intel y Computadoras de 64 bits basadas en AMD), que proporcionan rutinas LAPACK profundamente optimizadas.

Stata sobre Apple Silicon

Stata 17 para Mac está disponible como una aplicación universal que se ejecutará de forma nativa tanto en Mac con Apple Silicon como en Mac con procesadores Intel.



JDBC

Conectar Stata con bases de datos se ha vuelto aún más fácil puesto que Stata 17 agrega soporte para JDBC (Conectividad a bases de datos de Java). JDBC es un estándar multiplataforma para intercambiar datos entre programas y bases de datos.

Puede importar datos desde algunos de los proveedores de bases de datos más populares, como Oracle, MySQL, Amazon Redshift, Snowflake, Microsoft SQL Server y muchos más, esta opción corresponde a una solución multiplataforma, por lo que, funciona de la misma manera para los sistemas Windows, Mac y Unix.

Integración de Java

Ahora puede insertar y ejecutar código Java directamente en Stata. En versiones anteriores era posible usar complementos de Java, sin embargo, eso requería compilar y empaquetar en un archivo Jar.

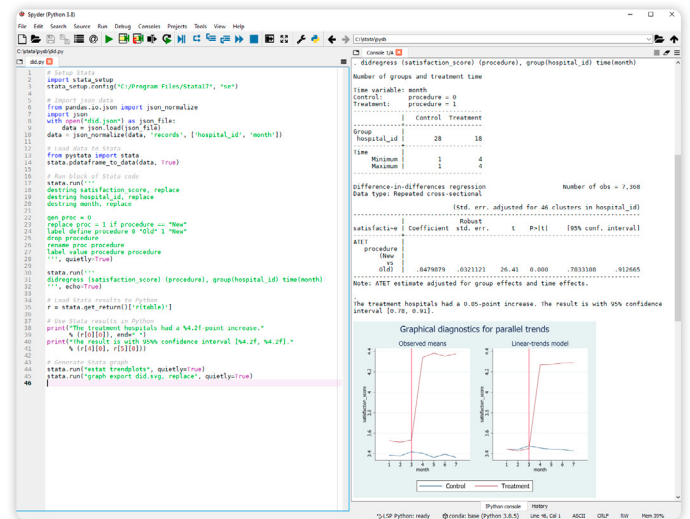
La ejecución de Java en un **archivo Do** le da la libertad de ejecutar código Java vinculado directamente a su código de Stata.

Integración de H2O

Con la integración de H2O, puede iniciar, conectarse y consultar un clúster de H2O de Stata. Además, el software proporciona un conjunto de comandos para manipular datos (marcos H2O) en el clúster.

PyStata

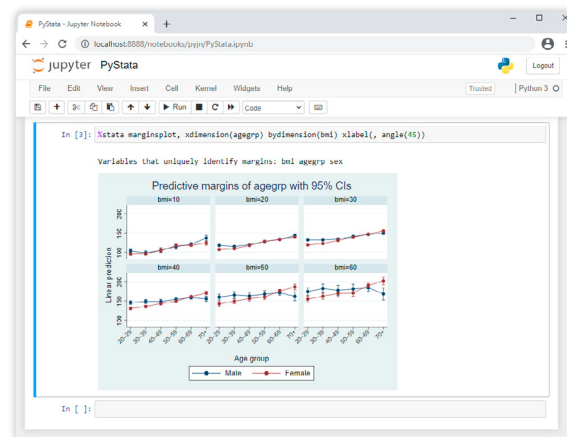
pystata es un término que abarca todas las formas en que Stata y Python puede interactuar, este término es introducido en Stata 17. El paquete **pystata** Python permite llamar a Stata desde su entorno Python independiente. Además, acceda a Stata y Mata cómodamente en un entorno basado en kernel de Python.



Jupyter Notebook con Stata

Jupyter Notebook es una aplicación web potente y fácil de usar que le permite combinar código ejecutable, visualizaciones, ecuaciones y fórmulas matemáticas, texto narrativo, entre otros, en un documento único o cuaderno para el desarrollo.

Esta opción es ampliamente utilizada por investigadores y científicos, para compartir sus ideas y resultados. También, IPython permite llamar a Stata y Mata desde Jupyter Notebook.





*Ciencia y tecnología al
servicio de la investigación y
el desarrollo*

Argentina

+54 (11) 5077 9516

Brasil

+55 (21) 9357 1215

Chile

+56 (2) 2656 2790

Colombia

+57 (1) 619 4000

México

+52 (555) 351 1755

Perú

+51 (1) 706 8197

USA

+1 (425) 996 0636

Venezuela

+58 (212) 335 0588

Manténgase conectado con nosotros

