

Análisis de datos estadísticos con STATA

SILABO

Análisis de datos estadísticos con STATA

I. Información general

Nombre del curso	Análisis de datos estadísticos con STATA
Docente especialista	Juan Palomino
Duración del Curso	8 semanas
Dedicación del Participante	6 horas por semana
Requisitos	Estadística
Modalidad	Virtual

II. Presentación

Durante los últimos años se ha producido un desarrollo explosivo en paquetes informáticos, lo cual permite recolectar un número creciente de set de datos relacionados con las ciencias sociales, la economía y otras ciencias. Además, saber explorar, describir y analizar datos es una competencia cada día más valorada por los empleadores y más necesaria para comprender lo que nos rodea e imprescindible en un investigador o científico de datos. Dentro de los distintos programas econométricos disponibles, Stata proporciona herramientas para aplicaciones estadísticas y el uso de técnicas econométricas más importantes del análisis de datos. Este curso tiene como finalidad brindar a los estudiantes un conjunto de herramientas de programación en Stata, así como realizar análisis de datos para enfrentar la resolución de problemas de investigación en distintos tópicos económicos como educación, salud, regional, ambiental, y financiero.

III. Objetivos

El objetivo principal del curso es que el participante domine los fundamentos básicos-intermedios de programación y manejo de bases de datos en el software Stata. De tal manera, que tenga la capacidad analítica y pragmática de modificar, limpiar y analizar diferentes tipos de muestras. En cuanto a los objetivos específicos, el participante al finalizar el curso tendrá la capacidad de:

1. Aplicar herramientas técnicas de estadística y econometría, enfocado principalmente a estudiantes y profesionales interesados en el manejo de información cuantitativa aplicada.
2. Identificar metodologías adecuada para dar respuesta a los problemas que aparecen en el estudio empírico de algunos datos económicos.
3. Trabajar con distintas fuentes de encuestas nacionales entre ellas: Encuesta Nacional de Hogares (ENAHOG), Niños del Milenio, Encuesta Nacional Demográfica y de Salud Familiar (ENDES), Evaluación Censal de Estudiantes (ECE), Registro Nacional de Municipalidades (RENAMU), entre otras (Censos, Consulta Amigable MEF, datos del BCRP)

IV. CONTENIDOS

Unidad 1: INTRODUCCIÓN A STATA	
Contenidos conceptuales	Contenidos Procedimentales
<p>Tema 1: Recursos de STATA 16</p> <p>1.1 Presentación de Stata 16: ventajas y desventajas</p> <p>1.2 Interfaz de Stata 16</p> <p>1.3 Ayuda y búsqueda de comandos rápida en Stata 16</p> <p>1.4 Documentación PDF en Stata 16</p> <p>1.5 Ejemplo de base de datos incluidos en Stata 16</p> <p>1.6 Cómo descargar e instalar comandos escritos por usuarios</p> <p>1.7 La sintaxis de programación: Do file</p> <p>1.8 Interfaz del gestor de proyectos de Stata</p> <p>1.9 Identificando y eliminando observaciones duplicadas</p>	<ul style="list-style-type: none"> • Reconocer el entorno e interfaz para el uso de Stata • Comprender el uso de las diferentes ventanas de Stata. • Operar la lógica y la sintaxis de Stata 16 • Utilizar el Do-file editor para aprender a programar en Stata • Usar el gestor de proyectos para ordenar la información en Stata
UNIDAD 2: MANEJO DE BASE DE DATOS	
Contenidos conceptuales	Contenidos procedimentales
<p>Tema 1: Manejo de Base de Datos</p> <p>1.1. Organizando nuestro proyecto y establecimiento de rutas</p> <p>1.2. Tipos y formatos de variables</p> <p>1.3. Sintaxis de los comandos</p> <p>1.4. Cargando bases de datos</p> <p>1.5. Trabajando con múltiples bases de datos</p> <p>1.6. El sufijo if e in, el prefijo by y bysort</p> <p>1.7. Explorando las variables</p> <p>1.8. Etiquetando variables y valores</p> <p>1.9. Manipulando variables: creación de variables, borrando y manteniendo variables, renombrando y ordenando variables</p> <p>1.10 Manipulando observaciones: reemplazando y recodificando valores</p> <p>1.11 Estadística descriptiva y tabulación de Variables</p> <p>1.12 Guardado y eliminando archivos</p> <p>Tema 2: Uniendo Base de Datos</p> <p>1.1 Combinando dos o más archivos de datos: Merge y Append</p> <p>1.2 Agrupamiento y organización de datos: Collapse y Reshape</p>	<ul style="list-style-type: none"> • Aplicar comandos básicos y útiles para el manejo de base de datos • Elaborar resúmenes estadísticos que permitan la identificación de los valores atípicos de una variable. • Unir bases de datos a través de un identificador en común. • Manejar los tipos de estructuras de bases de datos: Wide y Long
UNIDAD 3: LOOPS Y MACROS	
Contenidos conceptuales	Contenidos procedimentales
<p>Tema 1: Macros y Loops</p> <p>1.1 ¿Qué es una macro?</p>	



<p>1.2 Tipos de Macros: Global, Local, Escalares, Return</p> <p>1.3 ¿Qué es un Loop?</p> <p>1.4 Comandos: Foreach, Forvalues</p>	<ul style="list-style-type: none"> Utilizar iteraciones en procesos repetitivos para ser eficientes en la programación.
UNIDAD 4: VISUALIZACIÓN DE DATOS Y MAPAS	
Contenidos conceptuales	Contenidos procedimentales
<p>Tema 1: Elaboración y Creación de Gráficos</p> <p>1.1 Gráficos twoway</p> <p>1.2 Gráficos de dispersión</p> <p>1.3 Gráficas de línea, área y barras</p> <p>1.4 Histogramas y Densidades de Kernels</p> <p>1.5 Gráficos de cajas (box plot)</p> <p>1.6 Gráficos de puntos</p> <p>1.7 Gráficos circulares</p> <p>1.8 Creación de mapas</p> <p>1.9 Combinación de gráficos</p>	<ul style="list-style-type: none"> Crear gráficos informativos y estadísticos que ayuden a la visualización de la información. Crear mapas didácticos con la finalidad de encontrar fenómenos focalizados en determinados territorios. Combinar gráficos y trabajar con gráficos personalizados.
UNIDAD 5: ANOVA Y ANÁLISIS DE REGRESIÓN LINEAL	
Contenidos conceptuales	Contenidos procedimentales
<p>Tema 1: ANOVA y otros métodos de comparación de muestras</p> <ul style="list-style-type: none"> Test de una y dos muestras Análisis Bivariado: <ul style="list-style-type: none"> Test de proporciones Test de varianza Test de medias Relación entre variables Coefficiente de correlación de Pearson <p>Tema 2: Análisis de Regresión Lineal</p> <p>2.1 Fundamentos y estimación del Modelo Multivariado</p> <p>2.2 Modelo con covariantes continuos</p> <p>2.3 Modelo con covariantes polinómicos</p> <p>2.4 Modelo con covariantes categóricos</p> <p>2.5 Modelo con covariantes dicotómicos e iteración</p> <p>Tema 3: Creación automatizada de documentos y reportes</p> <p>3.1 Crear informes reproducibles en Stata</p> <p>3.2 Crear informes PDF desde Stata</p> <p>3.3 Crear documentos de Word desde Stata</p>	<ul style="list-style-type: none"> Analizar una muestra o grupos de muestras con la finalidad de validar estadísticamente los grupos. Analizar la correlación entre diferentes variables Estimar diferentes tipos de variables de manera adecuada a través de un modelo MCO Realizar análisis post-estimaciones Elaborar tablas, gráficos e informes en PDF y Word desde Stata 16
UNIDAD 6: MODELOS DE ELECCIÓN DISCRETA	
Contenidos conceptuales	Contenidos procedimentales
<p>Tema 1: Modelos de Elección Discreta</p> <p>1.1 Identificación de modelos no lineales</p> <p>1.2 Modelo de probabilidad lineal</p> <p>1.3 Modelos Probit y Logit binario</p> <p>1.4 Modelos Probit y Logit ordenados</p> <p>1.5 Modelos Logit Multinomial</p>	<ul style="list-style-type: none"> Identificar cuándo se debe usar un modelo binario, ordenado y multinomial. Saber interpretar los efectos marginales de las estimaciones no lineales.
UNIDAD 7: ESTIMACIONES DE SERIES DE TIEMPO	
Contenidos conceptuales	Contenidos procedimentales

Tema 1: Series de Tiempo 1.1 Manejo de variables temporales 1.2 Estacionariedad, diferencia, Detrending, estacionalidad 1.3 Correlogramas y autocorrelación serial 1.4 Test de Raíz unitaria 1.5 Función autocorrelación (ACF) y función de autocorrelación parcial (PACF) 1.6 Metodología Box-Jenkins para seleccionar un modelo ARIMA	<ul style="list-style-type: none"> • Crear y configurar variables temporales en Stata 16. • Detectar y corregir problemas de autocorrelación. • Identificar problemas de estacionariedad a través de prueba de raíz unitaria • Identificar la selección de modelos ARIMA.
UNIDAD 8: REGRESIÓN CON DATOS DE PANEL	
Contenidos conceptuales	Contenidos procedimentales
Tema 1: Manejo de Datos de Panel 1.1 Modelo de datos de panel estático: 1.2 Regresión agrupada (Pooled) 1.3 Efectos fijos: individuales, temporales, ambos 1.4 Efectos aleatorios <ul style="list-style-type: none"> • Test de elección de modelos alternativos • Modelos de efectos individuales vs modelo pool • Modelo de efectos fijos versus efectos aleatorios 	<ul style="list-style-type: none"> • Trabajar, reorganizar y estructurar una base de datos de panel. • Realizar un análisis empírico en la estimación de los modelos de datos de panel: Pooled, fijos y aleatorios.
CONTENIDOS ACTITUDINALES	
<ul style="list-style-type: none"> • Valorar la organización y el orden en la codificación y programación. • Valorar el uso potencial del software Stata 16 para realizar estadísticos, gráficos y estimaciones • Valorar el conocimiento en la identificación de variables para el uso de modelos adecuados. • Plasmar ideas y apreciaciones en resolver preguntas de manejo y limpieza de base de datos. 	

V. Metodología

Este curso ha sido diseñado bajo la metodología “aprender-haciendo” en entornos virtuales, en el cual se incorpora las características del e-learning y el empleo de una participación activa y reflexiva sobre los temas. Los participantes cuentan con un material elaborado por el profesor para el estudio de las ocho secciones que conforman el curso. Asimismo, se han elaborado videos tutoriales para reforzar el material. Las actividades planteadas para este curso son las siguientes:

- **Foros de consultas:** permitirá al profesor y el participante intercambiar información sobre los temas del curso. Además, este espacio virtual permitirá realizar consultas y debates sobre los temas.
- **Trabajos aplicativos:** donde los participantes elaborarán, de manera individual, tres trabajos aplicativos en los cuales pueden ejercer casos económicos y financieros en una situación real de trabajo.
- **Trabajo Final:** los participantes elaborarán, de manera individual, un trabajo integrador que tendrá un seguimiento a la mitad (avance) y al término del curso. Este trabajo se basará en un caso propuesto y se tomará en cuenta la aplicación de los diferentes comandos usados en los temas tratados a lo largo del curso.



VI. Sistema de evaluación

El sistema de evaluación es formativo y permanente con el objetivo de que el participante reflexione, cuestione y relacione con sus saberes previos los contenidos planificados en el curso y las actividades correspondientes. El sistema de evaluación consiste en:

Tipo de evaluación	Porcentaje
Trabajo aplicativo 1	15%
Trabajo aplicativo 2	15%
Trabajo aplicativo 3	15%
Trabajo aplicativo 4	15%
Trabajo aplicativo 5	15%
Trabajo Final	40%
Nota final del curso	100%

La nota más baja de los cinco trabajos aplicativos se elimina.

VII. Bibliografía

- Acock, Alan
2014 *A gentle introduction to Stata*. College Station, TX: Stata Press
- Adkins, Lee
2011 *Using Stata for principles of econometrics*. Fourth Edition. Hoboken.
- Beckett, Sean
2013 *Introduction to time series using Stata*. College Station, TX: Stata Press.
- Kohler, Ulrich y Frauke Kreuter
2005 *Data analysis using Stata*. Stata press.
- Long, Scott
2009 *The workflow of data analysis using Stata*. College Station, TX: Stata Press.
- Mitchell, Michael
2008 *A visual guide to Stata graphics*. College Station, TX: Stata Press



BUSINESS REPORT

DISTANCE (MILES) PACE (TIME / MILE)



PUCP

NOTES

DATE	TIME	DISTANCE (MILES)	PACE (TIME / MILE)
APR 04	10:00	3.00	10:00
APR 04	20:00	3.00	17:00
APR 04	30:00	3.00	17:00
APR 04	40:00	3.00	17:00
APR 04	50:00	3.00	17:00
APR 04	60:00	3.00	17:00
APR 04	70:00	3.00	17:00
APR 04	80:00	3.00	17:00
APR 04	90:00	3.00	17:00
APR 04	100:00	3.00	17:00
APR 04	110:00	3.00	17:00
APR 04	120:00	3.00	17:00
APR 04	130:00	3.00	17:00
APR 04	140:00	3.00	17:00
APR 04	150:00	3.00	17:00
APR 04	160:00	3.00	17:00
APR 04	170:00	3.00	17:00
APR 04	180:00	3.00	17:00
APR 04	190:00	3.00	17:00
APR 04	200:00	3.00	17:00