

Documentation file for Do-file and Datasets: "Jadenkä Panamá 2018"

Por: Gabriella Wong, Innovations for Poverty Action

Última versión: 03 de abril, 2019

Este archivo contiene información sobre las carpetas, estructura de documentos, y especificaciones técnicas necesarias para replicar los resultados de "*Jadenkä: Matemática temprana para educación inicial, Panamá*" por Emma Naslünd-Hadley y Juan Manuel Hernández-Agramonte.

Versión de stata y user-written commands

- Versión de stata 14.0
- User written commands instalados: orth_out, cplot, codebook, grstyle, palettes, spineplot

Estructura de archivos

Para realizar la evaluación de Jadenkä, se trabajó todos los archivos relacionados a la recolección de datos en la carpeta de '01 data analysis', de '10 Analysis & results'. Si bien los raw datasets de línea base y línea final se recogieron durante campo en la carpeta '07 Questionnaires & Data', se decidió exportar los raw datasets a la carpeta 10 para simplificar el proceso de réplica de información. No obstante, dentro de la carpeta 07 se puede visualizar la información ligado con las primeras versiones del data cleaning, los cuales se utilizaron durante la fase de fieldwork.

Dentro de la carpeta en la que se trabajó, '01 Data analysis', se cuenta con los siguientes sub-carpetas:

- **adofile**: programación en stata.
- **archive**: documentos varios.
- **data**: data cruda, intermedia, y final.
- **dofiles**: do-files de limpieza, preparación de datos, y análisis.
- **results**: tablas de balance, descriptivas, y resultados.

En este reporte, se explica lo que contiene la carpeta de **data** y **dofiles**.

Datasets

Los datasets han sido guardados en dos archivos distintos: raw y dta.

Parte 1: Raw Data

Raw data contiene bases de datos con información "recién salida" de la recolección de datos. En tal sentido, contiene toda la información subida durante campo a través de este programa. Cabe mencionar que toda la carpeta se encuentra **encriptada**, y solo los que cuentan con acceso a ésta podrán acceder.

A. Dentro de la carpeta "**BL**" se encuentra:

1. **Estudiantes**: Incluye información levantada a través del aplicativo surveyCTO. (file: Formulario de ESTUDIANTES)
2. **Maestros parte 1**: Incluye información levantada en papel. (file: Maestro_I_clean")
3. **Maestros parte 2**: Incluye dos bases de datos. La información realizada en tablet, y la información levantada en papel. (file SurveyCTO: Formulario de DOCENTES) (file papel: Maestro_II_clean)
4. **Directores**: Incluye dos bases de datos. La información realizada en tablet, y la información levantada en papel. (file SurveyCTO: Formulario de DIRECTOR) (file papel: Director_clean)

B. Dentro de la carpeta "**LF**" se encuentra:

- 1) **Estudiantes**: Incluye información levantada a través del aplicativo surveyCTO.
- 2) **Maestros endline**: Incluye información levantada en SurveyCTO y papel (file surveyCTO: Formulario de docentes) (file papel: maestros_endline_clean)

C. **complementary**

1. **'4_overal_teacher_survey_treat_30oct'** (CSV): lista de códigos únicos de maestros con el que se trabajó para la recolección de línea final. Primero se crearon códigos usando solo maestros parte II, luego, la editora de datos, al empezar a editar encuestas parte 1, fue agregando códigos consecutivos a los maestros que solo aparecían en maestros parte 1, pero no en parte 2. De esta forma se crea este documento.
2. **'encuestas_papel'** (XLS): documento que especifica las encuestas que fueron recolectadas en papel durante BASELINE uso: ser indicador de encuestas papel para fusionarlo con las bases de datos de tablet
3. **'listas_estudiantes_LONG_v2'** (CSV): lista de códigos únicos de estudiantes que se trabajó al inicio de endline uso: corregir creación de código de estudiantes al recrear código
4. **'Reporte de visitas BL 9 Dec 2018'** (XLS): base de datos de llamadas telefónicas donde se identificó a los estudiantes que no pertenecían al estudio.
5. **'muestra_aleatorizada_032018'** (DTA): uso: esta base la utilizamos para fusionar la asignación al tratamiento, así como las áreas geográficas. Este es el único documento no encriptado, pues más adelante se utiliza dentro de dofiles de preparación de datos.

Parte 2: DTA data

1. PII information

La carpeta PII_info_encrypted cuenta con información PII de las observaciones que ha sido dividida utilizando los dofiles de limpieza por cada tipo de actor. Cada una cuenta con un código único que identifica a la observación.

2. BL y EL

Las carpetas BL y EL, respectivamente, cuentan con los datasets ex-post división de información PII. De este modo, son datasets seguros de utilizar y compartir, pues solo cuentan con códigos únicos por observación. Los archivos que terminan en "_clean" son productos de los dofiles de limpieza; mientras que los que terminan en "_prep" son producto de haber generado variables para el análisis de datos.

2.1 Archivos temporales BL

- **BL_dropped_students** (DTA): Dataset que compila el total de estudiantes que inicialmente se contaban como una observación en el dataset de estudiantes LB, pero que por motivos varios (duplicado, encuesta de entrenamiento, falsificado), no deben pertenecer, y por lo tanto, fueron eliminados durante la limpieza de datos.
- **Correcciones_en_estudiantes** (CSV): Documento que forma parte del dofile de limpieza de estudiantes BL, en el cual se corrige grados, turnos, códigos mal digitados durante fieldwork, nombres de estudiantes, entre otros.

2.2 Archivos temporales LF

- **EL_dropped_students** (DTA): Dataset que compila el total de estudiantes que inicialmente se contaban como una observación en el dataset de estudiantes LF, pero que por motivos varios (duplicados, falsificados, otros) no deben pertenecer, y por lo tanto, fueron eliminados durante la limpieza de datos.
- **Correcciones_en_estudiantes** (CSV): Documento que forma parte del dofile de limpieza de estudiantes LF, en el cual se corrige grados, turno, códigos mal digitados durante fieldwork, entre otros.
- **Ausentismo** (DTA): Dataset que incluye todas las observaciones que no pudieron ser entrevistados durante la línea final. Incluye el motivo de ausencia durante la visita.
- **Ausentismo_razones** (CSV): Dentro de los motivos por los cuales se podía indicar el motivo de ausencia, incluimos la opción "Otros". Luego de revisarlo, se logró re-codificarla e incluirla de manera categórica, junto con las previas categorías de ausentismo. Se utiliza para hacer un readreplace en la variable "why_salido" dentro del dofile de descripción general.

3. Master data

Los datasets que contienen el archivo "Master_data" son producto de combinar las bases de datos con el fin de ser utilizados para el análisis final. En total, se produjeron 4 datasets:

- **Estudiantes_merged**: Estudiantes LB + LF
- **Estudiantes_docentes**: Estudiantes_merged + Docentes LF

- **Estudiantes_padres:** Estudiantes_merged + Padres LB
- **Docentes_merged:** Docentes LB1 + LB2 + LF

Do-files

0. Master file

El do-file principal se llama "**MASTER_jadenka**". Llama a los otros do-files uno por uno. En este do-file, se tendrá que cambiar el rootpath para el análisis.

El master data se encarga de generar las variables para el análisis y luego fusionar bases de datos para poder realizar los respectivos análisis de heterogeneidad. Finalmente, corre los dofiles de desgaste, balance, descripción de data, y análisis exploratorio.

1. Clean do-files (1_clean)

Los dofiles de limpieza realizan lo siguiente: corrige formato de fechas, elimina data que no debe ser considerada parte del estudio, corrige valores utilizando readreplaces. Por cada actor y medición, se cuenta con un dofile.

2. Prep do-files (2_aggregate)

Los dofiles de generación de variables procesan la información limpia de los datasets, y crea nuevas variables que serán utilizadas para el análisis de datos. De este modo, se modifican los formatos, etiqueta de valores y variables, crean nuevas variables, generan índices, estandarizan información, entre otros. Por cada actor y medición, se cuenta con un dofile.

4. Análisis do-files (3_analysis)

Una vez tratada la información recolectada en campo, eliminando información que no forma parte del estudio, así como creado las variables para el análisis, se procede a describir esta información.

- **Reporte_DataManagement** (stmd y do-file): STMD es el formato por el cual se trabajó para generar el reporte de manejo de datos. Combina código de Stata y edición de texto, utilizando el comando **markstat**, el cual permite generar documentos dinámicos.
- **Attrition_test:** Durante la línea final no se pudo recolectar el total de estudiantes esperados, lo cual generó un desgaste de la muestra. Para analizar la correlación entre dicho desgaste y el tratamiento, se utiliza este do. El resultado es las regresiones respectivas de análisis de desgaste, y
- **Balance_test:** Tablas de ortogonalidad generales de línea de base y línea final.
- **Descripcion_General:** Tabla general que resume el total de data recolectada durante la línea de base, y línea final.
- **Descripcion_Desagregada:** Tablas descriptivas desagregadas por actor.
- **Analisis_exploratorio:** Modelos sobre el efecto global y desagregado del programa, incluido heterogeneidades.
- **BL_directores_analysis:** Tablas de ortogonalidad por sección del cuestionario.
- **BL_maestros_I_analysis:** Tablas de ortogonalidad por sección del cuestionario.
- **BL_maestros_II_analysis:** Tablas de ortogonalidad por sección del cuestionario.
- **BL_padres_analysis:** Tablas de ortogonalidad por sección del cuestionario.
- **EL_docentes_analysis:** Tablas de ortogonalidad por sección del cuestionario.
- **EL_estudiantes_analysis:** Tablas de ortogonalidad por sección del cuestionario.