

EC-4101 Datos Económicos, I-2023

Trabajo Final (Laboratorio)

Instrucciones

En este informe se busca que los estudiantes puedan realizar una presentación de estadística descriptiva, de manera que se cumplan los estándares mínimos exigidos en un documento de investigación académico, utilizando las herramientas vistas en las clases teóricas, así como en los laboratorios del curso.

Los siguientes son los lineamientos administrativos de este trabajo:

- Se trabajará en grupos de máximo 4 integrantes.
- Debe entregarse un único trabajo en pdf por grupo, así como único do-file.
- Los cálculos y gráficos se deben realizar en Stata, pero puede aplicarse el formato a los cuadros en otro programa (Word, Excel, etc.).
- Debe incluir todas las instrucciones de Stata utilizadas para contestar cada pregunta con una breve explicación.
- Se calificarán 4 dimensiones en cada pregunta: formato, contenido, comentario y procedimiento. El primero se refiere a aplicar el formato apropiado a los cuadros y gráficos, el segundo a la representación correcta de los datos solicitados en la pregunta, el tercero a la interpretación de los datos y el cuarto a la inclusión de los procedimientos utilizados para el cálculo. Puede consultar la tabla de calificación al final de esta asignación.
- Al inicio del do-file identifiquen su trabajo poniendo un encabezado con sus nombres, números de carné e información básica (fecha, título, etc).
- Se deducirá la totalidad de los puntos de las respuestas que sean muy similares entre grupos.
- El formato de los cuadros y gráficos debe seguir los lineamientos vistos en Laboratorio y expuestos en la *Guía para la presentación de información estadística* del INEC. Segunda edición. 2017.
- Fecha de entrega máxima: 30 de junio, 11:59 pm.

Para la elaboración de este proyecto se le han compartido las Encuestas de Hogares de los años 2017 y 2018. Estos archivos contienen solamente una parte de las variables de ambas encuestas, porque de esta forma son menos pesados, para compartirlos con propósitos del proyecto. Los nombres de las variables no han sido modificados esencialmente, por lo que la documentación sobre las encuestas que encuentren en el sitio web del INEC aplica.

Sección 1: Corte Transversal

Selección de los individuos que estarán en el análisis

Se va a restringir el análisis a individuos que cumplan con las siguientes características:

- Son el jefe (o la jefa) y su compañera(o) o esposa(o)
- Tienen ***Ingreso por trabajo bruto total (itrbt)*** positivo (>0) y conocido (no missing)

Con respecto a las ***Horas totales normales (HorTotNorm)*** que es la variable de horas semanales de trabajo que se usarán en este proyecto, se eliminarán todos los individuos que no cumplan lo siguiente:

- Tienen ***Horas totales normales (HorTotNorm)*** positivas (>0) y conocidas (no missing, no 999)
- Tienen ***Horas totales normales (HorTotNorm)*** mayores o iguales que 8 y menores o iguales a 60.

Al usar estos criterios quedan 8031 y 8244 observaciones en 2017 y 2018, respectivamente. En este documento se llama a estos individuos ***“perceptores principales”***.

Variables de los miembros del hogar

Además de las características de los perceptores principales, hay algunas variables de interés que deben ser creadas con la información de todos los miembros de cada **hogar**, pero tomando en cuenta que entre las personas que están enlistadas en un hogar se incluyen personas del servicio doméstico y pensionistas (personas que alquilan un cuarto), los cuales no son de interés para la creación de estas variables, por lo tanto, exclúyalas.

Las variables para crear son las siguientes:

- La cantidad total de niños de con edad menor igual a 12 años (inf12) en cada hogar.
- La cantidad de adultos mayores con edad mayor a 70 años (am70) en cada hogar.
- La cantidad total de perceptores calculado con el Ingreso total personal bruto (percep) en cada hogar.

Unir las bases de las dos Encuestas de Hogares

Una vez que se tengan bases con la información de interés personal y del hogar para los ***perceptores principales***, se deben unir las dos bases de datos para crear una sola base que tenga a todos los perceptores de los dos años. Esta base de datos debe tener:

8031+8244=16275 observaciones.

Creación y recodificación de variables

Después de unir las bases de datos 2017 y 2018, hay algunas variables adicionales que deben ser creadas:

- Crear una variable de zona (zon) que tenga valores 0 y 1, con el 1 igual a zona rural.
- Crear una variable de género (mujer) que tenga valores de 0 y 1, con el 1 igual a mujer.
- Crear una variable que separe a las personas según quienes tienen pareja (pareja=1, si están casadas o en unión libre y 0 en los otros casos)
- Crear una variable de años de escolaridad (Esc) en la cual los valores missing estén recodificados como un punto.
- Crear una variable de adulto mayor (am) que tenga valores 0 y 1, con 1 indicando que el hogar sí tiene adulto mayor.
- Crear una variable de rama de actividad, usando RamaEmpPri, que se llame rama y en la cual los valores missing estén recodificados como un punto.
- Crear una variable de experiencia laboral (con el nombre de: exp), con la siguiente formula: Edad*Escolaridad-6. La variable exp debe tener valor missing (.) si años de escolaridad es missing (Esc==.). La variable exp debe hacerse igual a cero, si diera negativa.
- Crear el Ingreso del trabajo por hora (itrh) usando el ***Ingreso por trabajo bruto total***, para calcular la cantidad de colones por hora, usando ***Horas totales normales*** y calculando 4 semanas por mes.

La base que se debe obtiene después de lo anterior debe tener las variables que están en la tabla siguiente, en la cual se ponen las estadísticas descriptivas, para que verifique si tuvo algún error.

Tabla A: Descripción de las variables de interés en el proyecto que deben estar en la base que se obtiene del manejo de datos

Variable	Descripción	Nombre	Obs	Mean	Std. Dev.	Min	Max
Identificador de hogar	Creado con UPM, CUESTIONARIO, HOGAR	idh	16275	571,734.6000	319,205.50	1,011.00	1,120,121.00
Fuente de dato por ENAHO	Puede ser 2017 o 2018	enaho	16275	2,017.5070	0.50	2,017.00	2,018.00
Zona	Urbano=1, Rural=2	ZONA	16275	1.3000	0.46	1.00	2.00
Zona	Urbano=0, Rural=2	zon	16275	0.3000	0.46	0.00	1.00
Sexo	Hombre=1 mujer=2	A4	16275	1.4103	0.49	1.00	2.00
Sexo	Hombre=0 mujer=1	mujer	16275	0.4103	0.49	0.00	1.00
Edad	Años	A5	16275	44.6979	12.06	17.00	93.00
Estado conyugal	Ver label	A6	16275	2.3805	1.44	1.00	6.00
Pareja	No tiene=0, Tiene pareja=1	pareja	16275	0.7587	0.43	0.00	1.00
Niños menores de 12 años del hogar	Cantidad	inf12	16275	0.6748	0.91	0.00	9.00
Adultos mayores más de 70 años del hogar	Cantidad	am70	16275	0.0543	0.26	0.00	3.00
Dicotómica de adulto mayor	No tiene=0, Tiene=1	am	16275	0.0474	0.21	0.00	1.00
Años de escolaridad	Años	Escolari	16275	9.4425	5.76	0.00	99.00
Años de escolaridad	Años	Esc	16249	9.2992	4.51	0.00	22.00
Efecto cruzado escolaridad	Años	Esc_muj	16249	4.2192	5.82	0.00	22.00
Experiencia	Años	exp	16249	29.4023	13.50	0.00	86.00
Horast Totales normales	Horas por semana	HorTotNorm	16275	41.8587	12.43	8.00	60.00
Rama de actividad	Ver label	RamaEmpPri	16275	9.7553	6.35	1.00	99.00
Rama de actividad	Ver label	rama	16271	9.7334	6.20	1.00	21.00
Ingreso trabajo bruto total	Colones por mes	itrbt	16275	563,302.8000	707,727.60	2,000.00	23,600,000.00

Ingreso total personal bruto	Colones por mes	itpb	16275	597,354.1000	741,483.60	2,500.00	23,800,000.00
Ingreso trabajo por hora	Colones por hora	itrh	16275	3,444.8390	4,469.77	26.04	150,000.00
Cantidad de perceptores del hogar	Cantidad	percep	16275	2.1976	1.04	1.00	10.00
Quintil de ingreso	Categoría de 1 a 5	q_ipcn	16275	3.3338	1.34	1.00	5.00

Ahora bien, se quiere describir las diferencias de ingreso de trabajo bruto total (itrbt) y por hora (itrh) entre las mujeres y los hombres que son perceptores principales, por quintil de ingreso y en cada encuesta, calculando los promedios. Con tal fin, debe preparar una tabla con el siguiente formato:

Tabla 1

Promedio Ingreso trabajo bruto total						
	2017			2018		
Quintil	Mujer	Hombre	Razón	Mujer	Hombre	Razón
1						
2						
3						
4						
5						
Promedio Ingreso de trabajo por hora						
	2017			2018		
Quintil	Mujer	Hombre	Razón	Mujer	Hombre	Razón
1						
2						
3						
4						
5						

Nota: Las razones están calculadas como el promedio de mujeres sobre el promedio de hombres.

Calcule el coeficiente de Gini a partir del ingreso de trabajo bruto total (itrbt), para cada uno de los años en la muestra y comente brevemente el resultado. Adicionalmente grafique la curva de Lorenz para el año 2018, la cual debe incluir la curva de perfecta igualdad, comente el resultado.

Genere un gráfico de caja con el ingreso de trabajo por hora (itrh) segregado por sexo, por la variable “pareja” y según año, presentándose de manera conjunta (combinada). ¿Cómo se compara la distribución de los ingresos entre hombres y mujeres dependiendo de si tienen pareja o no para cada año?

Genere un histograma (que muestre la frecuencia, no la densidad) del ingreso de trabajo por hora (itrh), para los hogares sin adultos mayores, donde solo aparezcan ingresos de trabajo por hora menores a 100.000 colones. Comente la forma de la curva, refiérase a la asimetría.

Sección 2: Series de Tiempo

Utilice la serie de Producto Interno Bruto, por Actividad Económica, a precios básicos y de mercado:

<https://gee.bccr.fi.cr/indicadoreseconomicos/Cuadros/frmVerCatCuadro.aspx?idioma=1&CodCuadro=%202992>

Importe, como se vio en clase, esta serie a Stata, y declárela como serie de tiempo.

Una vez que la declaró, genere un gráfico de línea con la serie de su elección, y su suavización con la opción window, y weights, en ambos casos para medias móviles con 2 periodos de antelación y 2 periodos posteriores, para el caso ponderado, elija los ponderadores que desee.

Determine la correlación entre tres variables de su elección como se vio en clase. Determine la correlación entre el PIB a precios de mercado y su primer rezago.

Calificación

Los puntos que se asignen a cada pregunta, se valorarán de acuerdo con la siguiente descomposición:

Dimensión	Criterio	Valor	Obtenido
Formato	Título apropiado y numeración	2.5%	
	Fuente correcta y notas al pie	2.5%	
	Títulos de eje y etiquetas de eje legibles / Columna matriz ordenada y legible	2.5%	
	Leyenda legible / Encabezado legible	2.5%	
Contenido	Utilización de la variable solicitada	20%	
Comentario	Interpretación correcta de la información representada	20%	
	Uso adecuado de las estadísticas calculadas	5%	
	Contextualización de los comentarios en términos de la submuestra	5%	
Procedimiento	Instrucciones de Stata funcionales de acuerdo a lo solicitado	40%	
Total		100%	