

NOTA METODOLÓGICA

CLASIFICACIÓN DE LA INFORMACIÓN DE RENTA DE PERSONAS NATURALES EN CUANTILES DE INGRESO BRUTO Y DE PATRIMONIO BRUTO

Subdirección de Estudios Económicos
Dirección de Gestión Estratégica y de Analítica



CLASIFICACIÓN DE LA INFORMACIÓN DE RENTA DE PERSONAS NATURALES EN CUANTILES DE INGRESO BRUTO Y DE PATRIMONIO BRUTO

NOTA METODOLÓGICA

Para la generación de información estadística de las declaraciones de renta de personas naturales, agregada en grupos de acuerdo con las medidas de posición (cuantiles), ésta se ha construido a partir de los valores registrados por las personas naturales en los registros administrativos: formularios 210 (años 2014 a 2020), 230 y 240 (años 2014 a 2016).

Como primera medida, toda la información original ha pasado por un proceso de revisión que busca validar la veracidad de las cifras, de tal manera que los datos publicados tengan el mayor grado de confiabilidad. Se precisa que en algunos casos se realizaron modificaciones (permitidas por metodología estadística) a la información de declaraciones de renta de las personas naturales, con el propósito de reflejar la consistencia estadística en el agregado año a año. Estas modificaciones no pretenden alterar las cifras registradas por los contribuyentes en los formularios de declaración, el objetivo es poner a disposición del usuario interno un insumo estadístico confiable para los análisis y estudios que se realizan al interior de la DIAN y, así mismo, permitir su consulta mediante publicación en la página web de la DIAN para el uso de los centros de investigación, las universidades, demás grupos de interés en materia tributaria y la ciudadanía en general.

Adicionalmente, este documento presenta una descripción de los recursos utilizados para obtener la información agregada por medidas de posición; se presenta el procedimiento que permite obtener las estadísticas por cuantiles a partir de dos variables utilizadas como criterio de clasificación: “*Total ingresos brutos*” y “*Total patrimonio bruto*”. También se incluye una muestra de la información, tal como aparece organizada en los cuadros que se han construido siguiendo el formato de publicación de agregados de renta de personas naturales.

Período de publicación:

La información ha sido generada a partir de los formularios de declaración de renta de personas naturales no obligadas a llevar contabilidad (formularios 210, 230 y 240). Se ha tenido en cuenta el período comprendido por los años gravables 2014 a 2020, siendo este último el año más reciente para el cual se cuenta con información revisada y actualizada de los formularios.

Las actualizaciones se realizan de acuerdo con la disponibilidad de la información recibida en los formularios de declaración de renta de personas naturales (última declaración presentada por el declarante a la fecha de corte), la cual se presenta de acuerdo con el calendario tributario y se publica después de realizar el proceso de revisión y validación de cifras.

1. Insumos y Recursos

Información utilizada:

A continuación, se presenta la relación de los formularios de declaración de renta de las personas naturales no obligadas a llevar contabilidad, de los cuales se tomó la información para generar las cifras agregadas por cuantiles. Frente a cada formulario relacionado se encuentra el número de casos que conforman la población de declarantes en ese formulario para el respectivo año gravable.

Cuadro 1. Información de renta de personas naturales utilizada en el ejercicio

Año	Número Formulario	Nombre formulario	Número de casillas	Número de casos
2014	210	Declaración de Renta y Complementarios Personas Naturales y Asimiladas No Obligadas a Llevar Contabilidad	73	1.627.084
	230	Declaración Anual de Impuesto Mínimo Alternativo Simple (IMAS) para Empleados	38	483.934
	240	Declaración Anual de Impuesto Mínimo Alternativo Simple (IMAS) para Trabajadores por Cuenta Propia	35	38.270
2015	210	Declaración de Renta y Complementarios Personas Naturales y Asimiladas No Obligadas a Llevar Contabilidad	73	1.860.868
	230	Declaración Anual de Impuesto Mínimo Alternativo Simple (IMAS) para Empleados	38	430.417
	240	Declaración Anual de Impuesto Mínimo Alternativo Simple (IMAS) para Trabajadores por Cuenta Propia	35	43.518
2016	210	Declaración de Renta y Complementarios Personas Naturales y Asimiladas No Obligadas a Llevar Contabilidad	73	2.025.473
	230	Declaración Anual de Impuesto Mínimo Alternativo Simple (IMAS) para Empleados	38	444.878
	240	Declaración Anual de Impuesto Mínimo Alternativo Simple (IMAS) para Trabajadores por Cuenta Propia	35	42.827
2017	210	Declaración de Renta y Complementario Personas Naturales y Asimiladas de Residentes y Sucesiones Ilíquidas de Causantes Residentes	76	3.059.135
2018	210	Declaración de Renta y Complementario Personas Naturales y Asimiladas de Residentes y Sucesiones Ilíquidas de Causantes Residentes	76	3.289.969
2019	210	Declaración de Renta y Complementario Personas Naturales y Asimiladas de Residentes y Sucesiones Ilíquidas de Causantes Residentes	80	3.621.027
2020	210	Declaración de Renta y Complementario Personas Naturales y Asimiladas Residentes y Sucesiones Ilíquidas de Causantes Residentes	111	3.975.418

Fuente: Elaboración Subdirección de Estudios Económicos

De cada uno de los formularios de declaración, relacionados en el cuadro anterior, se ha tomado la totalidad de las casillas (variables) para mostrar la información de cada año gravable. Se debe tener en cuenta que, como ya se mencionó, la clasificación de la información se realiza a partir de las variables *Total ingresos brutos* y *Total patrimonio bruto*, obtenida mediante la aplicación de dos ejercicios diferentes.

En el caso de los ingresos brutos, esta variable aparece con este nombre en los formularios de los años gravables 2014, 2015 y 2016 de manera que se tomó el valor la casilla para realizar la clasificación. Para los años gravables 2017, 2018, 2019 y 2020, la variable Ingresos brutos se calcula a partir de la sumatoria de los ingresos brutos de todas las rentas cedulares, tal como se muestra en el cuadro 2.

Cuadro 2. Cálculo de la variable Total ingresos brutos para los años 2017 a 2020

Año	Cálculo de la variable <i>Total ingresos brutos</i>
2017	Total ingresos brutos = (Ingresos brutos por rentas de trabajo + Ingresos brutos por rentas de pensiones del país y del exterior + Ingresos brutos rentas de capital + Ingresos brutos rentas no laborales + Dividendos y participaciones 2016 y anteriores, y otros + 1a. Subcédula año 2017 y siguientes numeral 3 art. 49 del E.T. + 2a. Subcédula año 2017 y siguientes parágrafo 2 art. 49 del E.T. + Ingresos por ganancias ocasionales del país y del exterior)
2018	Total ingresos brutos = (Ingresos brutos por rentas de trabajo + Ingresos brutos por rentas de pensiones del país y del exterior + Ingresos brutos rentas de capital + Ingresos brutos rentas no laborales + Dividendos y participaciones 2016 y anteriores, y otros + 1a. Subcédula año 2017 y siguientes numeral 3 art. 49 del E.T. + 2a. Subcédula año 2017 y siguientes parágrafo 2 art. 49 del E.T. + Ingresos por ganancias ocasionales del país y del exterior)
2019	Total ingresos brutos = (Ingresos brutos por rentas de trabajo + Ingresos brutos rentas de capital + Ingresos brutos rentas no laborales + Ingresos brutos por rentas de pensiones del país y del exterior + Dividendos y participaciones 2016 y anteriores, y otros + 1a. Subcédula año 2017 y siguientes numeral 3 art. 49 del E.T. + 2a. Subcédula año 2017 y siguientes parágrafo 2 art. 49 del E.T. + Ingresos por ganancias ocasionales del país y del exterior)
2020	Total ingresos brutos = (Ingresos brutos Trabajo + Ingresos brutos Honorarios + Ingresos brutos Capital + Ingresos brutos No laboral + Ingresos brutos por rentas de pensiones del país y del exterior + Dividendos y participaciones 2016 y anteriores, y otros + 1a. Subcédula año 2017 y siguientes numeral 3 art. 49 del E.T. + 2a. Subcédula año 2017 y siguientes parágrafo 2 art. 49 del E.T. + Ingresos por ganancias ocasionales del país y del exterior)

Fuente: Elaboración Subdirección de Estudios Económicos

Recursos informáticos:

El ejercicio de clasificación por cuantiles se realizó haciendo uso del programa IBM SPSS Modeler, Versión 18.2., y la información de los cuadros de salida obtenidos en Modeler se exportó a archivos Excel.

En los archivos Excel se organizaron las estadísticas para cada uno de los mil cuantiles construyendo tablas dinámicas. Con la organización de la información en los archivos Excel se visualizan las cifras agregadas, con el fin de realizar una revisión a partir de controles para verificar la consistencia de los datos obtenidos en cada ejercicio de clasificación. También se utilizó el formato Excel para construir los cuadros finales en los cuales se presenta la información publicada en la página Web de la DIAN.

2. Procedimiento para obtener las estadísticas por cuantiles

La información relacionada en el cuadro 1 se tomó de los archivos de datos de la Subdirección de Estudios Económicos. Estos archivos contienen la información registrada en los formularios de declaración de renta de las personas naturales no obligadas a llevar contabilidad. Los datos han sido previamente revisados, para lo cual se ha verificado, entre otras cosas, la consistencia de las cifras en el tiempo, la validez de los resultados de las operaciones registradas por los contribuyentes en los formularios y los requisitos establecidos en estatuto tributario.

A continuación, se describe detalladamente el procedimiento para obtener la información, utilizando como ejemplo el ejercicio desarrollado para la seccional 32 (Bogotá) del formulario 240, año 2016.

2.1. Cargue de información

La información, previamente revisada para cada formulario de declaración, se importó utilizando el programa SPSS Modeler, indicando la lectura desde archivos almacenados en formato SPSS Statistics Data Document (.sav).

2.2. Organización de variables y lectura de datos

2.2.1 De la información registrada en los formularios de declaración de renta de las personas naturales se tomaron todos los registros (casos) teniendo en cuenta las variables: *Total ingresos brutos* (variable calculada como se indica en el cuadro 2) y *Total patrimonio bruto*.

2.2.2 A continuación, se realizó la lectura completa de la totalidad de los datos que componen cada base de declaraciones a analizar, y se definió la variable tenida en cuenta para el ejercicio de clasificación. En este caso, se tiene presente que se debe realizar una clasificación por la variable *Total ingresos brutos* y una clasificación por la variable *Total patrimonio bruto*.

2.3. Registros ordenados

Una vez se ha realizado la importación de la información y la lectura de datos para la totalidad de las variables, en el siguiente paso se ordenan de forma ascendente los datos de la variable de clasificación seleccionada (*Total ingresos brutos* o *Total patrimonio bruto*). En este paso se debe garantizar que, luego del arreglo ascendente, se conserven los datos originales para cada uno de los registros (NIT) en las operaciones posteriores.

2.4. Construcción de conjuntos (intervalos)

La variable ordenada de forma ascendente permite un ejercicio de clasificación de cada uno de los registros en un determinado intervalo (en este caso, en cada uno de los mil cuantiles). En el programa se realizan 999 cortes a la variable ordenada para obtener así los mil grupos de clasificación. De esta forma, se genera un conjunto de información que indica a qué grupo dentro los mil cuantiles de la distribución pertenece cada caso asociado a un NIT.

2.5. Elaboración de tabla con registros clasificados en un intervalo

La información clasificada en cuantiles de ingreso bruto o de patrimonio bruto se guarda internamente en el panel de resultados del programa. Con el fin de obtener un formato de visualización de las cifras generadas, se elabora en el programa una tabla con los campos que se importaron (respetando los valores originales de cada registro) y, adicionalmente, con un nuevo campo que indica el número del cuantil al cual pertenece cada registro en cada variable del formulario de declaración de renta de personas naturales.

A continuación, se presenta una muestra para visualizar la forma en que se va estructurando la base. El cuadro 3 contiene información de un segmento de la tabla generada por SPSS Modeler luego de ordenar de forma ascendente la variable *Total Patrimonio Bruto*, tomada del Formulario 230, año gravable 2014. Los valores del cuadro 3 son ilustrativos y no necesariamente corresponden con lo registrado en el formulario; esto con el fin de guardar la reserva de la información.

Cuadro 3. Muestra de información ordenada por Patrimonio bruto

Número de registro	Total Patrimonio Bruto	Deudas	Total Patrimonio Líquido	.	.	.	Total saldo a favor	Cuantil
1	0	29.764.000	0	.	.	.	1.291.000	1
2	0	15.000.000	0	.	.	.	5.298.000	2
3	1.000	0	0	.	.	.	0	3
4	2.000	0	0	.	.	.	0	4
.
.
.
483.931	1.234.567.890	123.456.789	1.111.111.101	.	.	.	0	997
483.932	1.234.567.891	123.456.793	1.111.111.098	.	.	.	123.456	998
483.933	1.234.567.892	0	1.234.567.892	.	.	.	897.533	999
483.934	1.234.567.893	123.456	1.234.444.437	.	.	.	2.189.755	1.000

Fuente: Elaboración SEE-DIAN con base en información de las declaraciones de renta personas naturales (Análisis de Operaciones. SEE. DGEA. DIAN).

2.6. Exportación de información a formato Excel y cálculos de estadísticas por cuantil

Las tablas obtenidas con la información clasificada en cuantiles, tal como se muestra en el cuadro 3, son exportadas a archivos Excel en los que se organizan tablas dinámicas para calcular las estadísticas por cada grupo de ingreso bruto o de patrimonio bruto: Conteo de casos, Suma y Participación del valor de cada cuantil en el total de cada variable.

Los pasos del algoritmo presentado en los numerales 2.1 a 2.6 se deben ejecutar para cada uno de los formularios, y por cada año gravable, teniendo en cuenta la información relacionada en el cuadro 1 de este documento.

2.7. Organización de información

En archivos Excel se construyen los cuadros de salida con la información revisada, verificada y organizada que se entrega como producto, y que aparece publicada y disponible para el usuario en la página WEB de la DIAN. Se presenta un cuadro con información para cada formulario y por cada año gravable en el período 2014 – 2020.

3. Información publicada

3.1 Sitio de publicación

En el micrositio de las estadísticas de los tributos administrados por la DIAN: <https://www.dian.gov.co/dian/cifras/Paginas/TributosDIAN.aspx> Sección: *Impuesto de renta personas naturales no obligadas a llevar contabilidad* se encuentran dos carpetas

identificadas con los nombres: *Clasificación_por_Cuantiles_Ingreso_bruto* y *Clasificación_por_Cuantiles_Patrimonio_bruto*.

3.2 Contenido de las carpetas

a) La carpeta *Clasificación_por_Cuantiles_Ingreso_bruto* contiene 13 archivos en formato Excel. Cada uno de estos archivos presenta la información clasificada en mil cuantiles de ingreso bruto para los formularios relacionados en el cuadro 1.

b) La carpeta *Clasificación_por_Cuantiles_Patrimonio_bruto* contiene 13 archivos en formato Excel. Cada uno de estos archivos presenta la información clasificada en mil cuantiles de patrimonio bruto para los formularios relacionados en el cuadro 1.

3.3 Contenido de los archivos

Cada uno de los archivos Excel contiene 2 hojas en las que el usuario puede encontrar la siguiente información:

a) Hoja 1: “*Valor cuantiles PN 2020*”. Con este nombre se identifica la hoja en la que aparece el cuadro con la información de las declaraciones de renta de las personas naturales no obligadas a llevar contabilidad, presentando el valor totalizado para cada uno de los cuantiles y para cada variable del formulario al que hace referencia el archivo.

Cuadro 4. Muestra de un segmento de la Hoja 1 para el año gravable 2020.
Clasificación por Ingreso Bruto

Cuantil	Número de casos	Total ingresos brutos	PG	. . .	HB	PAGO_TOTAL
12	47.251	0	16.343.673	. . .	28.479	87
13	3.724	5	849.473	. . .	2.532	3
14	4.614	37	1.289.507	. . .	3.015	4
15	3.892	196	1.161.630	. . .	3.162	3
.
.
.
997	3.975	7.439.263	11.870.272	. . .	12.050	56
998	3.976	9.423.089	14.026.016	. . .	15.368	19
999	3.975	13.433.235	16.605.206	. . .	21.576	12
1000	3.975	42.526.444	45.244.687	. . .	255.417	1
Total	3.975.418	473.484.853	1.441.524.902	. . .	3.868.374	60.569

Fuente: Elaboración SEE-DIAN con base en información de las declaraciones de renta personas naturales (Análisis de Operaciones. SEE. DGEA. DIAN).

b) Hoja 2: “Renglones”. En esta hoja se encuentran los nombres de las variables correspondientes a la nomenclatura asignada para etiquetar cada columna que describe las variables de la hoja 1, y que representan las casillas del formulario al que hace referencia el archivo.

Cuadro 5. Muestra de un segmento de la Hoja 2 para el año gravable 2020

Sección	Nombre del campo	Descripción del Campo	Número de Renglón
Patrimonio	PG	Total Patrimonio Bruto	29
	PH	Deudas	30
	PI	Total Patrimonio Líquido	31
Rentas de trabajo	IB_TRABAJO	Ingresos brutos	32
	INCR_TRABAJO	Ingresos no constitutivos de renta	33
	RLIQ_TRABAJO	Renta líquida	34
	AP_VOL_TRABAJO	Aportes voluntarios AFC, FVP y/o AVC	35
	OTRAS_EXEN_TRABAJO	Otras rentas exentas	36
	TOTAL_EXEN_TRABAJO	Total rentas exentas	37
	INTERÉS_VIVIENDA_TRABAJO	Intereses de vivienda	38
	OTRAS_DEDUCC_TRABAJO	Otras deducciones imputables	39
	TOTAL_DEDUCC_TRABAJO	Total deducciones imputables	40
	REXEN_LIM_TRABAJO	Rentas exentas y/o deducciones imputables (limitadas)	41
	RLIQ_CED_TRABAJO	Renta líquida ordinaria	42

Fuente: Elaboración SEE-DIAN con base en información de las declaraciones de renta personas naturales (Análisis de Operaciones. SEE. DGEA. DIAN).

3.2 Notas aclaratorias

3.2.1 Información del formulario 240

La información de agregados que aparece en los archivos del formulario 240, para los años 2014, 2015 y 2016, se ha clasificado en percentiles (100 grupos). Se define esta clasificación teniendo en cuenta que, por ejemplo, el número máximo de casos que conforman el universo en este formulario es de 43.518 (corresponde al año 2015). Esto implica que, al dividir la información en mil cuantiles de igual frecuencia, cada uno de los grupos estaría conformado en promedio por 43 casos; en ese caso es necesario definir un nuevo criterio para determinar el número máximo de conjuntos (cuantiles), de tal manera que sea posible garantizar la reserva de la información.

3.2.2 Construcción de los cuantiles

Como se observa en el cuadro 4, presentado en la Sección 3.3 de este documento, el primer grupo de clasificación se identifica con el número 12 en la columna *Cuantil* (cuantil 12). Esto quiere decir que el programa genera una tabla que no muestra información para los primeros 11 cuantiles la razón se presenta a continuación.

Como se explicó en el procedimiento detallado, la clasificación por cuantiles se obtiene al ordenar de forma ascendente la variable utilizada como criterio de clasificación (en este caso *Total ingresos brutos* o *Total patrimonio bruto*), y luego se realizan los cortes para obtener los percentiles mediante la opción “cuantiles de igual frecuencia”, incluida en el programa SPSS - Modeler. De esta forma, de acuerdo con la distribución de la variable, hay algunos cuantiles (en el ejemplo, los primeros 11) que acumulan un valor cero (\$0) y que corresponden a los contribuyentes que no registraron un valor de ingresos brutos o de patrimonio (según el caso) mayor a cero durante el año gravable.