

Proyección del impacto de futuras medidas tributarias

2023

Proyección del impacto de futuras medidas tributarias

Rodrigo Luis González Cao

Universidad de Buenos Aires

Facultad de Ciencias Económicas

Centro de Estudios en Administración Tributaria

Buenos Aires, Argentina

Noviembre de 2023.

Nota del Autor

Las opiniones y conclusiones expresadas en los contenidos que presenta este trabajo de investigación son de exclusiva responsabilidad del autor y pueden no coincidir con las del CEAT.

© CEAT: El CEAT se siente complacido de que terceros citen los contenidos de sus investigaciones para su uso personal, sin fines comerciales, sin ningún derecho a revender o redistribuir las mismas. Agradeceremos que siempre se aclare que la fuente de información es el presente material.

Todas las publicaciones del CEAT se encuentran disponibles en
http://www.economicas.uba.ar/extension_centros/ceat



Agradecimientos

Al Centro de Estudios en Administración Tributaria (CEAT) de la Facultad de Ciencias Económicas de la Universidad de Buenos Aires (UBA), y a sus autoridades, por generar espacios para la investigación académica sobre la gestión fiscal en la universidad pública.

Resumen

En este trabajo presentamos una introducción al estado del arte en metodologías de previsión y proyección tributaria, así como también en herramientas de análisis de ingresos públicos encaminadas a medir el impacto de posibles medidas fiscales futuras originados en cambios en la normativa impositiva.

En tiempos inciertos, es un desafío clave para cualquier gobierno proyectar correctamente la recaudación futura de los ingresos públicos, analizar correctamente el desempeño fiscal y anticipar los impactos que puedan generar las medidas de modificación tributaria que se impulsen en el equilibrio del presupuesto nacional.

A continuación, se presentan algunos conceptos básicos de herramientas y métodos estadísticos que pueden resultar útiles para la proyección de los ingresos públicos, así como para el análisis de la política fiscal.

Ofrecemos una introducción práctica que ayudará a generar conocimientos fundamentales sobre los diferentes modelos y técnicas que pueden utilizarse para proyectar la recaudación impositiva y realizar un análisis certero de la política fiscal.

Palabras Clave

Administración Tributaria, Impuestos, Recaudación Tributaria, Plan estratégico.

Abstract

In this work we present an introduction to the state of the art in tax forecasting and projection methodologies, as well as public revenue analysis tools aimed at measuring the impact of possible future fiscal measures originating from changes in tax regulations.

In uncertain times, it is a key challenge for any government to correctly project the future collection of public revenues, properly analyze fiscal performance and anticipate the impacts that tax modification measures that are promoted on the balance of the national budget may generate.

Below are some basic concepts of statistical tools and methods that can be useful for the projection of public revenues, as well as for the analysis of fiscal policy.

We offer a practical introduction that will help generate fundamental knowledge about the different models and techniques that can be used to project tax collection and perform an accurate analysis of fiscal policy.

Keywords

Tax Administration, Taxes, Tax Collection, Strategic Plan.

Medición del Impacto de Futuras Medidas Tributarias

Esta publicación forma parte de una serie de investigaciones que buscan analizar en tiempos inciertos el buen impuesto, los cambios de política fiscal y el impacto de dichas modificaciones en la futura recaudación tributaria.

La utilización de la información prospectiva para proyectar el impacto de potenciales medidas de política fiscal sobre la futura recaudación de ingresos públicos requiere como primer paso ponernos de acuerdo sobre la terminología que utilizaremos.

Siguiendo una posible clasificación propuesta en la revista “*Contabilidad y Auditoría*”¹ de esta Facultad de Ciencias Económicas (VIEGAS, RIAL, & GAJST, 2000) podemos diferenciar entre:

¹ Viegas, J., Rial, L., & Gajst, I. (1). *La Información prospectiva: proyecciones o pronósticos*. Contabilidad Y Auditoría, (12), 26 Págs. Recuperado a partir de <https://ojs.econ.uba.ar/index.php/Contyaudit/article/view/188>

- **Estimación:** información proyectada respecto a un período en curso para el cual aún no se han publicado los resultados.
- **Pronóstico:** información proyectada respecto a un período en curso para el cual se posee certeza del comportamiento de las variables.
- **Proyección:** información proyectada respecto a un período futuro sobre el cual hay confianza en que continúe la tendencia de las variables, pero no hay plena certeza.
- **Ilustración:** información proyectada con respecto a un período futuro que se basa en una serie de suposiciones de eventos futuros, sobre los cuales no es posible tener suficiente confianza en que el comportamiento de las variables siga la actual tendencia.

En la presente publicación nos enfocaremos en las metodologías de proyección de los futuros ingresos tributarios. Como señala la CEPAL (MARTIN, RODRIGUEZ CABELLO, & GONZALEZ, Modelo de proyección de las cuentas fiscales y análisis de la sostenibilidad fiscal, 2012) el proceso de proyección tributaria recorre tres etapas:

- Simulación
- Predicción
- Meta

En la primera etapa de simulación las proyecciones que se formulan buscan encontrar el valor que tendrá una variable, siempre y cuando se cumplan los supuestos previstos de ciertos factores determinantes.

El pase a la etapa de predicción se concreta cuando se posee certeza de que la simulación se concretará en un determinado valor de la futura recaudación (MARTIN, RODRIGUEZ CABELLO, & GONZALEZ, Modelo de proyección de las cuentas fiscales y análisis de la sostenibilidad fiscal, 2012).

La predicción alcanza la etapa final de meta cuando se incorpora al presupuesto de ingresos y recursos del ejercicio (GONZÁLEZ CAO, Previsión de impuestos, 2023).

En todo proceso de proyección de ingresos públicos como paso previo a la proyección cuantitativa es de buena práctica realizar una evaluación cualitativa del sistema tributario. Para ello nos valemos del marco del *buen impuesto* que hemos presentado en nuestra publicación previa (GONZÁLEZ CAO, El buen impuesto, 2023).

En los actuales tiempos de cambios exponenciales donde la volatilidad, la perplejidad, la confusión y la indeterminación dificultan la proyección de la recaudación de impuestos, el análisis cualitativo del sistema tributario se vuelve imprescindible como primer acercamiento a la cuestión. En dicho marco cualitativo, la formulación de una política tributaria debe evaluarse a partir de la eficiencia, equidad, simplicidad, flexibilidad, transparencia y eficacia (GONZÁLEZ CAO, El buen impuesto, 2023).

Como hemos señalado, los ingresos fiscales pueden verse influidos por varios factores a nivel macro, entre los cuales podemos enumerar los cambios en el entorno económico, el impacto de la digitalización de la economía de vigilancia, fluctuaciones en los precios de las materias primas, auges y caídas financieras. Pero no deben descartarse los factores que también inciden a nivel micro como los cambios en el comportamiento de los contribuyentes o las modificaciones de la normativa fiscal (GONZÁLEZ CAO, Nociones básicas de desempeño fiscal, 2023).

Los conjuntos de datos que se utilizan para realizar proyecciones suelen ser imperfectos y se publican con mucho retraso por lo cual los errores de proyección son comunes e inevitables. Por lo tanto, es crucial revisar críticamente tanto las fuentes de datos que sirven de insumo como los resultados de cualquier modelo de proyección de ingresos (GONZÁLEZ CAO, Previsión de impuestos, 2023).

A partir del recorrido realizado en las publicaciones previas cuyo foco estuvo en los aspectos cualitativos, a continuación vamos a centrarnos en los aspectos técnicos de la proyección cuantitativa de ingresos públicos con el objetivo de proporcionar una descripción general de los métodos que se pueden utilizar para proyectar la trayectoria de los recursos fiscales futuros.

La elección de qué modelo de proyección usar dependerá de una variedad de factores, entre los cuales uno de los más importantes son las fuentes de datos que están disponibles. Sin embargo, es importante dejar claro de entrada que, desde una perspectiva técnica, no existe un modelo de proyección preferido por sobre los otros.

El proceso de proyección tributaria comienza identificando las fuentes de datos y cómo evaluar la calidad de dichos datos. Para su análisis, en muchos casos, recurriremos a herramientas estadísticas. En este marco entendemos a la estadística como la disciplina metodológica que nos permite responder ciertas preguntas siguiendo un procedimiento. Por lo tanto, en líneas generales, seguiremos un método compuesto por una secuencia lógica de pasos:

- Planteo de las preguntas que nos guiarán,
- Planificación y realización del estudio,
- Recolección de los datos,
- Análisis de la información y,
- Formulación de las conclusiones.

A partir de los datos recolectados y de las conclusiones estadísticas a las que arribemos, luego debemos aplicar las técnicas de proyección, las cuales presentaremos describiendo el uso de modelos incondicionales para proyectar los ingresos esperados.

Por último, presentaremos una breve introducción a los modelos condicionales y un preámbulo al análisis de regresión lineal, mencionando de manera superficial cómo se puede usar lo que llamaremos “*elasticidad fiscal*” para proyectar los ingresos.

Hemos dicho que la estadística es una disciplina metodológica que nos permite responder preguntas. Sin embargo, para formular las preguntas adecuadas se requiere comprender previamente el tema, conocer qué datos se buscarán, cuáles de ellos están disponibles y en qué fuente, qué uso se les dará a los datos recolectados y cómo se analizarán luego para inferir potenciales tendencias y posibles comportamientos. Aquí entra en juego la **analítica de datos**, también llamada en inglés como “*Business Analytics*”.

La disciplina estadística tiene su propio vocabulario y, para evitar confusiones debemos nombrar unívocamente a cada concepto con ciertas palabras específicas que faciliten su identificación. Por eso, comenzaremos explicando la diferencia entre población y muestra.

Población

El primer paso es determinar el grupo de sujetos u objetos a ser estudiados. O sea, determinar la **población** bajo estudio. A los efectos de la presente publicación al referirnos a la **población** en sentido estadístico estaremos enfocándonos en el conjunto que forman todos los elementos que son objeto de un análisis descriptivo.

Por lo tanto, llamaremos **población** a todo el grupo de **unidades muestrales** (generalmente son individuos) que interesa estudiar con el fin de responder una pregunta de investigación. Por ejemplo, todos los contribuyentes de un país o todas las inspecciones que la Administración Tributaria realiza en un año. Las poblaciones, sin embargo, pueden ser difíciles de definir. En un buen estudio, se debe definir la población con toda claridad.

Muestras

En ciertas ocasiones, cuando la población es muy grande, los estudios estadísticos recurren al empleo de **muestras**. Las muestras son subconjuntos representativos de la población y su armado va a depender del tamaño de la población y de la variabilidad que presenten sus componentes.

Para cualquier pregunta que interese responder, primero es necesario dirigir la atención a un grupo particular de unidades muestrales. Por ejemplo: contribuyentes o establecimientos. Una vez definido el tema, se plantea una pregunta y se puede identificar uno o más grupos específicos de unidades que resultan de interés particular para el estudio que se realizará.

Una vez que tenemos identificados cada uno de los elementos de la población o de la muestra, según sea el objetivo del estudio es necesario medir u observar algunas características específicas. Como resultado de dichas observaciones se deberá realizar un registro cuyo resultado más inmediato será una serie de datos. Como veremos más adelante, estos datos pueden ser de tipo categórico o numérico.

Muchas veces, se quiere estudiar una población amplia y sacar conclusiones sobre ella, pero por limitaciones de tiempo, de volumen de cómputo o de fondos solamente se realiza el estudio sobre una población restringida. Esta limitación, si no se observan algunos recaudos, puede conducir a conclusiones sesgadas o erróneas.

Supongamos que se quiere estudiar si los contribuyentes que prestan servicios son más propensos a evadir que los que realizan ventas. Para hacerlo basa su estudio en un grupo de contribuyentes que se inscribieron en el último año. Este grupo de contribuyentes constituye una muestra. Pero los resultados a los cuales se arribe, ¿pueden generalizarse a toda la población de contribuyentes? ¿Es dicha muestra representativa de la totalidad de la población? Más adelante presentaremos el concepto de representatividad de una muestra y analizaremos los conceptos de sesgo y error de estimación que pueden incidir en las conclusiones a las que arribemos y tergiversar cualquier proyección que intentemos realizar.

Análisis de los Datos

El estudio del uso que se le puede dar a los datos en el ámbito de la proyección tributaria enmarca en lo que llamamos “*Business Analytics*” (o su traducción al español como “*analítica de datos*”). Dicha tarea de análisis puede tener como finalidad la descripción, la prescripción o la predicción.

En el ámbito de la gestión fiscal, el Banco Interamericano de Desarrollo define al proceso de *analítica de datos* como aquel que comprende la recolección, transformación y organización de los datos para responder preguntas, extraer conocimientos, realizar proyecciones e impulsar la toma de decisiones por medio de la utilización de herramientas, técnicas y métodos, generalmente estadísticos, que tienen como objetivo agregar valor a dichos datos para transformarlos en información valiosa (Banco Interamericano de Desarrollo, 2021).

Como explicaremos en breve, a partir de los datos históricos podemos identificar tendencias y proyectar futuros comportamientos del desempeño fiscal. Para ello nos valdremos de ciertas herramientas de la ***estadística descriptiva*** como tablas de distribución de frecuencias y gráficos que representen las variables analizadas, así como su distribución y dispersión.

Herramientas Estadísticas

El objeto de la ***estadística descriptiva*** es estudiar la variabilidad de algunas propiedades que caracterizan a un conjunto de elementos. Esos elementos pueden ser sujetos u objetos. La ***estadística descriptiva*** es la rama de la Estadística que proporciona las herramientas necesarias para realizar el procesamiento de datos, con la finalidad de organizarlos, facilitar su presentación, sintetizar sus principales propiedades e identificar tendencias o patrones de comportamiento.

El tipo de análisis que se puede realizar depende de los datos disponibles (tipo y cantidad) y del objetivo específico del estudio. Por lo general los datos brutos que se recolectan se ordenan y agrupan, organizándolos en ***tablas de frecuencias***, para luego poder exponer el resultado de este agrupamiento mediante ***gráficos*** apropiados.

Si los datos son numéricos se calculan cantidades estadísticas representativas (como promedios o rangos) que resumen sus principales propiedades. Este tipo de procesamiento permite transformar a los “*datos en bruto*” en “*información*” con valor agregado. Es decir, transformar números sueltos en conocimiento práctico que permita orientar la toma de decisiones relacionada con la población bajo estudio.

Al recolectar los datos aparecen varios interrogantes que nos permiten presentar a continuación más términos específicos de la disciplina estadística. La primera pregunta sería: ¿cuáles individuos serán seleccionados?

A su vez, la respuesta a dicho interrogante nos plantea nuevas preguntas, como, por ejemplo: ¿cuál será la muestra?, ¿se los elegirá en forma aleatoria?, ¿todos los elementos participantes tendrán la misma oportunidad de ser seleccionados?, ¿qué variables serán importantes en relación al tema (actividad económica, ingresos, tamaño de la nómina de personal)?

Unidades Muestrales

A los objetos de interés de un estudio se los denomina **unidades muestrales** o simplemente unidades. Muchas veces, las unidades muestrales son individuos (pueden ser personas, empresas o mercaderías). Sin embargo, en ciertas ocasiones, las unidades están compuestas por muchos individuos. Al respecto, es importante no confundir una unidad muestral como objeto completo y diferenciado que se encuentra dentro de un conjunto (una docena tiene doce unidades) con las unidades que se utilizan para valorar una magnitud (el metro es una unidad de longitud y no es un conjunto de 100 centímetros).

Variables

Las **variables** son características que pueden cambiar de una unidad muestral a otra. Por ejemplo, los ingresos de las personas, la cantidad de contribuyentes de cada jurisdicción territorial, el porcentaje de fiscalizaciones finalizadas con ajuste determinado en cada año.

También se puede aplicar a otros ejemplos como la intensidad de emisión de rayos de cada escáner de la Aduana, la capacidad de almacenamiento de un disco rígido o el gasto tributario originado en una medida de alivio fiscal.

¿Valores Relativos o Valores Absolutos?

Los procedimientos y resultados estadísticos que se utilicen para describir cantidades pueden hacer una diferencia respecto a las conclusiones que se obtienen e, incluso, pueden inducir a proyecciones erróneas. Por eso, a continuación, introduciremos los conceptos de razón, tasa y porcentaje, así como el análisis sobre la expresión de resultados en valores absolutos o relativos.

Razón

Una razón es el cociente entre dos cantidades. Por ejemplo, cuando expresamos que *“la razón de monotributistas a medianos contribuyentes es de 3 a 2”* significa que hay 3 pequeños por cada 2 medianos contribuyentes. No debe entenderse que solamente hay 3 pequeños contribuyentes y 2 medianos contribuyentes en la población o la muestra bajo estudio. Las razones se expresan utilizando los términos más bajos para simplificar lo más posible. Así, esta razón puede expresar la situación de una muestra de 25 contribuyentes con 15 monotributistas y 10 medianos, o de una población objetivo de 10 millones de sujetos compuesta por 6.000.000 de pequeños y 4.000.000 de medianos contribuyentes.

Tasa

Una tasa es un cociente que refleja una cierta cantidad por unidad. Por ejemplo, la tasa de cumplidores en término de declaraciones juradas determinativas presentadas al vencimiento es de 82 por cada 100 contribuyentes (la unidad es 100 contribuyentes). Pero también es un ejemplo de tasa la velocidad cuando decimos que un vehículo en la autopista se desplaza a 120 km por hora (la unidad es una hora).

Porcentaje y Proporción

Un porcentaje es un número entre 0 y 100 que mide la proporción de un total. A su vez, una proporción es un número entre 0 y 1 que representa la participación en el total (el total es igual a 1). Por ejemplo, cuando decimos que los ingresos tributarios son el 66% del total de ingresos públicos recaudados, decimos que representa 66 de cada 100 pesos percibidos (GONZÁLEZ CAO, El buen impuesto, 2023).

Un porcentaje del 66% es lo mismo que una proporción de 0,66:

- Para convertir un porcentaje en una proporción, se divide al porcentaje por 100.
- Para convertir una proporción en un porcentaje, se multiplica la proporción por 100.

Variaciones Relativas

Cuando un porcentaje se utiliza para determinar un aumento o reducción relativa (relativa al valor inicial), se denomina **variación porcentual**. Para calcular la variación porcentual se toma el valor “después de” y se le resta el “antes de”, luego se divide ese resultado por el “antes de”. Así, se obtiene una proporción. Para transformarla en un porcentaje se multiplica el resultado por 100.

Veamos a continuación un ejemplo real:

Tabla N.º 1. Variación Interanual de Aportes Personales. En millones de pesos. Setiembre de 2023

	Antes de	Después de	(Después- Antes)	(Variación Absoluta/Antes)	(Proporción por 100)
Concepto	Setiembre 2022	Setiembre 2023	Variación Absoluta	Proporción Variación	Variación porcentual
Aportes personales	\$148.503	\$332.706	\$184.203	1,2403	124,03%

Fuente: Elaboración propia a partir de “AFIP. COMPARATIVO DE LA RECAUDACION AÑOS 2023 Y 2022. Setiembre 2023 y 2022. Millones de pesos corrientes”. Setiembre de 2023. AFIP.

Recuperado de:

<https://contenidos.afip.gob.ar/institucional/estudios/archivos/comparativo.setiembre23.xls>

Como podemos apreciar, las variaciones relativas se pueden expresar como variaciones porcentuales o proporciones. Los aumentos se reflejan en variaciones porcentuales positivas. Las reducciones se reflejan en variaciones porcentuales negativas. Volveremos sobre el concepto de valores relativos al analizar las frecuencias (frecuencia relativa) y lo retomaremos al presentar las proyecciones basadas en regresión lineal al final de la publicación.

Muestras y Muestreo

Las **muestras** son subconjuntos representativos de la Población y su armado va a depender del tamaño de la población y de la variabilidad que presenten sus componentes. La forma en que se realiza la selección puede hacer la diferencia. A continuación vamos a presentar el concepto de **muestreo** y las características de una “**buena muestra**”, es decir, una muestra representativa.

Muestreo

Al definir el alcance de un análisis que realizaremos debemos definir con claridad qué composición tendrá la muestra y si la misma es representativa de la población total o tiene un sesgo que sobre representa o sub representa a un grupo determinado.

Por ejemplo, supongamos que se quiere averiguar algo sobre el comportamiento fiscal de los contribuyentes, pero no contamos con tiempo suficiente ni el presupuesto necesario para estudiar a todos los millones de individuos que conforman la población.

Para salvar esta limitación se selecciona una cantidad más reducida de unidades muestrales de la población (esto se llama una **muestra**), se estudian esas unidades, generalmente individuos, y se utiliza dicha información para sacar conclusiones que resulten extrapolables a toda la población.

Existen diferentes tipos de **muestreos** y la correcta elección de la muestra asegurará que las comprobaciones se realicen de manera eficiente sin resignar certeza en los controles. A continuación presentaremos cuatro posibles métodos de muestreo:

- Muestreo sistemático
- Muestreo aleatorio simple
- Muestreo aleatorio estratificado
- Muestreo por bloques, cuotas, aglomerados o conglomerados

Muestreo Sistemático

El **muestreo sistemático** comienza con una unidad elegida al azar y a partir de allí continúa cada cierta cantidad de unidades. A dicha cantidad la definiremos como “*k*”.

Para definir cada cuanta unidad vamos a proceder a seleccionar a un individuo recurrimos a la siguiente lógica: si “ n ” es el tamaño muestral y “ N ” es el tamaño de la población, entonces “ k ” será el número entero igual al cociente que resulta de dividir N/n . Veámoslo con un ejemplo concreto:

- $N = 533.000$ empleadores (población total)
- $n = 1.000$ (muestra)
- $k = 533.000/1.000 = 533$

Es decir que, del listado total de empleadores inscriptos ante la Administración Tributaria, seleccionaremos uno de cada 533.

Este tipo de muestreo permite evitar el sesgo personal y es el método más sencillo. Sin embargo, debemos advertir que solamente es útil cuando la población está ordenada naturalmente. Por lo tanto, cuando la población no está ordenada naturalmente, antes de utilizar este tipo de muestreo es necesario ordenarla. Al respecto, es importante señalar que, al ordenarla, se pierden las ventajas de simplicidad e inmediatez que tiene este tipo de muestreo.

Por su parte, este muestreo no es adecuado cuando el período de la selección está relacionado con alguna característica que nos interesa evaluar. Al realizar un muestreo sistemático es importante estar alerta para identificar los factores que puedan estar invalidando los resultados.

Muestreo Aleatorio Simple

Una **muestra aleatoria simple** es aquella que se obtiene a partir de un mecanismo que le asigna a cada una de las unidades muestrales la misma probabilidad de ser elegida. Una garantía inicial para que la muestra sea representativa es la selección aleatoria de los elementos de la población que tendrán oportunidad de ser elegidos. Para que cada uno de los individuos de una población tenga la misma oportunidad de ser seleccionados:

- Debe utilizarse algún mecanismo probabilístico para elegirlos.

- Quienes participan de la muestra no se postulan a sí mismos para participar.
- No debe existir ningún favoritismo ni sesgo en el proceso de selección de individuos.

El proceso por el cual se obtiene una muestra aleatoria comienza con la confección de una lista de unidades muestrales de la cual se extraerá la muestra. Esta lista se llama **marco muestral** y al proceso lo denominamos **muestreo aleatorio**. El muestreo aleatorio debiera proveer de una muestra representativa de la población. Por lo tanto, idealmente el marco muestral debería contener la lista de la totalidad de las unidades muestrales.

El **muestreo aleatorio simple** tiene dos propiedades que lo convierten en un procedimiento ampliamente preferido para la obtención de muestras:

- Todas las unidades tienen la misma oportunidad de ser elegidas (no está sesgado).
- La elección de una unidad no influye sobre la elección de otra (independencia).

Antes de avanzar, es importante aclarar que, en ciertas ocasiones, una unidad muestral puede contener “*en su interior*” muchos individuos. Por ejemplo, un empleador puede ser una unidad muestral, que contiene en su interior muchas personas que integran la nómina de su personal asalariado. En este caso el objeto de estudio puede ser tanto:

- los empleadores (por ejemplo: interesa conocer el tamaño de sus nóminas o alguna caracterización de los contribuyentes) o
- pueden ser las personas asalariadas que trabajan en relación de dependencia (por ejemplo: interesa conocer el monto de remuneración de cada persona que trabaja por cuenta ajena para un dador de trabajo).

Las muestras se seleccionan con base en la relación entre cifras generadas al azar y los datos dentro del conjunto.

Muestreo Aleatorio Estratificado

En un muestreo estratificado la población se divide en grupos homogéneos llamados estratos y las muestras se organizan conforme a dichos grupos estratificados. Luego se realiza un muestreo aleatorio simple de unidades muestrales dentro de cada estrato. Los elementos son extraídos de cada estrato en forma proporcional a su presencia en el conjunto total.

Los estratos se eligen de acuerdo con los valores conocidos de algunas variables, de manera que haya poca variabilidad dentro del estrato (homogeneidad interna) pero que haya mucha variabilidad entre estratos (heterogeneidad entre grupos).

Por lo tanto, los valores de dichas variables para las unidades de un estrato particular difieren poco, pero los valores de dichas variables para las unidades de distintos estratos difieren mucho. Para realizar un muestreo aleatorio estratificado deberemos seguir tres pasos:

- Paso 1: las unidades se agrupan en estratos. Los estratos se eligen teniendo en cuenta que estos grupos tienen un interés especial dentro de la población, o porque los individuos en el estrato se parecen mucho.
- Paso 2: se establece la proporción de unidades, o fracción de muestreo, que se incluirá para cada estrato.
- Paso 3: dentro de cada estrato se realiza un muestreo aleatorio simple y la proporción de individuos que se incluye en la muestra es la establecida en el paso 2. La unión de las muestras de cada estrato constituye la muestra completa.

Puede ser posible estructurar la muestra para que represente un alto porcentaje del valor de los datos, y al mismo tiempo un bajo número de elementos.

Es importante tener en cuenta que, en este tipo de muestreo, lleva mucho tiempo seleccionar las muestras si no se cuenta con los datos digitalizados electrónicamente y con un programa de computación adecuado para la tarea.

Muestreo por Bloques o Conglomerados

En este tipo de muestreo se divide a la población en grupos heterogéneos llamados bloques o conglomerados. Luego se realiza un muestreo aleatorio simple en el cual las unidades muestrales son los conglomerados. Las muestras se seleccionan con base en las características alfabéticas, numéricas o cronológicas de los datos.

La idea del agrupamiento para un muestreo aleatorio por bloques es opuesta a la del muestreo estratificado. Interesa que los individuos que componen cada aglomerado sean lo más heterogéneos posibles y se espera que cada bloque sea lo más representativo posible de la población.

Los bloques son las unidades del muestreo, pero las unidades de interés son los individuos dentro de los bloques. Otra manera de armar las “cuotas” o bloques es elegir a todos los que comparten una misma característica. Luego se selecciona una muestra aleatoria de bloques, y se observan todos los individuos dentro de cada bloque o se selecciona una muestra aleatoria simple dentro del mismo.

Este tipo de muestreo puede tener mejor rendimiento costo-efectividad que un muestreo aleatorio simple porque es fácil seleccionar las muestras y es simple de utilizar. A su vez, los resultados son menos susceptibles de ser representativos del conjunto total de datos.

Variabilidad Muestral

Para introducir el concepto de **variabilidad muestral** y antes de referirnos a **sesgo** y **error de estimación**, debemos ponernos de acuerdo en la diferencia entre **parámetro** y **estimado**.

Definimos como **parámetro** a un número que describe a la población. Por lo tanto, cuando el conjunto de datos proviene de la población completa, el valor del **estadístico** es un **parámetro**. Sin embargo, en la práctica casi nunca podremos conocer con exactitud cuál es ese número porque no es posible conocer perfecta y acabadamente a toda la población debido a limitaciones de tiempo, de presupuesto o de capacidad de cómputo.

Así, cuando el conjunto de datos proviene de una muestra, el número obtenido es el **estadístico** que se utiliza como una **estimación** del parámetro.

La diferencia entre el parámetro y el estadístico de la estimación es el **error de estimación**.

La primera ventaja de las muestras aleatorias es que eliminan el **sesgo** del procedimiento de selección de una muestra. Aun así, suele no coincidir el resultado con el verdadero valor, debido a la variabilidad que resulta de la selección al azar. Este tipo de variabilidad es llamada **variabilidad muestral**. Que la **variabilidad muestral** sea muy grande, significa que el valor del estadístico cambia mucho entre muestra y muestra. En estas situaciones, no podemos confiar en el resultado que obtenemos con una muestra en particular y, seguramente, deberemos repetir varias veces el muestreo.

Frente a ello, las muestras aleatorias tienen una propiedad de la cual podemos valernos: la variación entre muestra y muestra (de un mismo tamaño) seguirá un **patrón** predecible. Este patrón predecible muestra que los resultados de muestras de mayor tamaño son menos variables que los resultados de muestras más chicas.

Mediciones Válidas

Una variable es una medida válida de un concepto siempre y cuando lo represente adecuadamente. Medir una característica de un sujeto o un objeto significa asignarle un número expresable en distintas unidades que represente dicha variable. El resultado de esa medición no es

constante y toma diferentes valores dependiendo del individuo a quien se le está realizando la medición. La medición requiere de:

- Un proceso previo de transformación de conceptos (longitud, desempleo, nivel socioeconómico, etc.) en variables definidas con precisión.
- La elección del instrumento adecuado para medirlas.

La utilización de una cinta métrica para transformar la idea “*longitud*” en un número es directa, porque sabemos exactamente qué queremos decir con longitud. Sin embargo, en otros casos como ciertas variables que pueden incidir en los ingresos tributarios nos puede resultar mucho más complicado establecer el concepto a medir y el instrumento adecuado para hacerlo.

A su vez, aunque en muchas ocasiones no disponemos de mecanismos o instrumentos de medición adecuados para abordar una medición que nos resulta de interés para formular una proyección tributaria, sin embargo, podemos utilizar datos abiertos provistos por organismos del estado o información provista por universidades, organismos internacionales o empresas privadas para aproximarnos. Al utilizarlos es importante evaluar cómo están definidos, y si se trata de medidas válidas para describir las propiedades que se pretenden medir. Aun así, a veces no es posible obtener mediciones claramente válidas. Y también debemos alertar que para un mismo concepto se puede obtener más de una medición válida como veremos más adelante.

Buenas Muestras: Muestra Representativa

Para llegar a conclusiones acertadas necesitamos obtener “*buenas muestras*”. Definimos como una “buena muestra” a aquella que es representativa de la población. Esto significa, que todas las características importantes de la población tienen que estar presentes en la muestra en la misma proporción que en la población. Es decir que una muestra tiene las características más parecidas posibles a la población que representa, pero en un tamaño más reducido.

A medida que aumenta el tamaño de la población a analizar es más difícil que una muestra tenga proporciones idénticas a las poblacionales. Se debe considerar esta complejidad al diseñar muestras que sean lo más parecidas posibles en todas las características que se puedan conocer.

Si se quiere realizar un estudio para averiguar qué empleadores subdeclaran personal en sus nóminas salariales, la muestra debería tener una distribución geográfica, por provincia o región, similar a la del último censo económico disponible del INDEC. Sin embargo, si se considera que hay otros factores que pueden influir además de la ubicación, como por ejemplo el sector económico, el tamaño de la nómina salarial o la antigüedad del contribuyente, la muestra también tendrá que ser representativa de esos otros factores adicionales.

Malas Muestras: Sesgo

Pasemos en limpio algunas ideas. Una muestra es mala cuando:

- Quienes participan de la muestra se postulan a sí mismos para participar.
- Hay favoritismo o sesgo en el proceso de selección de individuos.

El sesgo es un favoritismo presente en alguna etapa del proceso de recolección de datos que beneficia algunos resultados, perjudicando a otros y desviando las conclusiones en direcciones equivocadas o forzadas. Es importante tener presente que una muestra grande no corrige el error por sesgo, ni tampoco lo diluye. Solamente lo repite más veces.

Sesgo por Elección de la Muestra

Entre los sesgos por elección de la muestra podemos identificar:

- **Muestras sesgadas por conveniencia:** elegir una muestra entre los casos registrados en una planilla de cálculo es rápido y sencillo, pero los sujetos seleccionados pueden no ser representativos de la mayoría de la población.
- **Muestras con sesgo personal:** por gusto, interés o por conocimiento mayor de una temática sectorial que de otras, quien realiza la selección puede preferir elegir a cierto tipo de contribuyentes y no a otros.
- **Muestras de respuesta voluntaria:** cuando la muestra se conforma por participantes voluntarios que, por algún motivo, decidieron participar, debemos ser cuidadosos de que se encuentren representadas las características de la totalidad de la población.

Sesgo de Respuesta

Entre los sesgos de respuesta podemos identificar como origen:

- **La presentación de las preguntas:** la elección incorrecta de las palabras con las cuales se formula la pregunta que se responderá puede incidir en el resultado. En ciertas ocasiones el sesgo en la presentación de las preguntas puede ser intencional o condicionado por una particular visión de las cosas. Claramente no es lo mismo preguntar:
 - ¿está usted harto de pagar impuestos para no recibir nada a cambio?
 - ¿le parece importante que se paguen impuestos para impulsar el desarrollo del país?
- **El tratar de agradar o no generar rechazo:** cuando la respuesta no es anónima, puede estar condicionada por un intento de ocultar un prejuicio, ser “*políticamente correcto*” al responder o evitar ser juzgado socialmente.
- **El recuerdo impreciso:** cuando la pregunta se refiere a un hecho pasado, la respuesta puede estar condicionada por la imprecisión del recuerdo.
- **La no respuesta:** cuando un grupo importante prefiere no responder, los resultados pueden arrojar sesgos, ya que quienes responden pueden ser muy diferentes a quienes sí lo hacen.

- **El subcubrimiento:** una selección de aquellos contribuyentes que tienen determinadas características puede estar ignorando a contribuyentes que no reúnan algunos de esos requisitos.

Tipos de Errores

Un proceso de medición tiene:

- **Error aleatorio**, si mediciones realizadas sobre un mismo objeto dan resultados diferentes.
- **Sesgo**, si sistemáticamente sobreestima o subestima la variable que mide.

Cuando el **error aleatorio** es pequeño se trata de mediciones precisas. Cuando las mediciones no tienen sesgo decimos que son mediciones exactas. Es importante recordar que el error aleatorio nunca se puede eliminar totalmente. Ambos tipos de errores suelen estar presentes en los procesos de medición, y los podemos expresar como:

Modelo de medición: valor medido = parámetro + sesgo + error aleatorio

El sesgo puede provenir de la calidad del instrumento y de las opciones realizadas al realizar el muestreo. Hemos dicho que el sesgo es un favoritismo que incide en alguna etapa del proceso de recolección de datos. Ese favoritismo producirá una subestimación o una sobrestimación sistemática de aquella característica de la población que intentamos medir.

A veces, la diferencia se encuentra en la interpretación del término del error. Cuando se seleccionan individuos de una población y se observa el valor de una variable, el término llamado "**error aleatorio**" representa las diferencias entre el valor individual (si se seleccionara sin sesgo) y la media poblacional (μ). El modelo general será:

$$\text{Valor individual} = \mu + \text{sesgo} + \text{error aleatorio}$$

Si los individuos se seleccionan mediante un muestreo aleatorio simple, eliminamos el sesgo y el modelo resulta así:

$$\text{Valor individual} = \mu + \text{error aleatorio}$$

Cuando se trabaja con datos es importante preguntarse: ¿Cómo se obtuvieron esos números?

Si se trata de mediciones sobre muchos individuos, tendremos valores de variables describiendo a cada uno de ellos. Debemos saber cómo está definida exactamente cada variable y si se trata de variables válidas como mediciones numéricas de los conceptos que estamos analizando. También es necesario conocer si los datos tienen errores de medición que puedan reducir su utilidad.

En ciertas ocasiones, algunos procedimientos de medición pueden introducir sesgo, en ese caso es necesario utilizar un instrumento mejor. Si medir al mismo individuo produce resultados diferentes, de manera que los valores no son confiables, se puede mejorar la confiabilidad repitiendo la medición varias veces y utilizando su promedio.

Es importante comprender que ningún proceso de medición es perfectamente preciso. Los procesos estadísticos recurren a los promedios para mejorar la precisión. Pero la precisión no significa validez ni invalidez. La validez pasa por otro lado. La validez, en términos generales, se refiere al grado en que una variable representa realmente la característica a medir. Muchas veces el problema de la validez de una variable para medir un concepto se encuentra en la naturaleza misma del concepto que se intenta analizar. Esta cuestión quedará más clara con un ejemplo ajeno a la proyección tributaria.

Supongamos que nos interesa medir la inteligencia de un grupo de sujetos y se utiliza la memoria como medida. Claramente la memoria no es una medida válida de inteligencia. Por otra parte, si bien es válido medir el diámetro del cráneo con una cinta métrica, tampoco es válido utilizar el diámetro craneal como medida de la inteligencia.

Datos y Variables

Al realizar el análisis exploratorio de los datos, comprender su estructura es el paso inicial para poder realizar luego una minería de esos datos ("*data mining*").

La minería de datos es un proceso técnico que analiza grandes cantidades de datos para descubrir potenciales patrones y convertirlos en información. La minería de datos está orientada a la extracción no trivial de patrones implícitos, previamente desconocidos que pueden resultar de potencial utilidad para detectar y proyectar comportamientos que se repiten de manera consistente.

Para avanzar con el análisis de datos y variables desde lo más simple a lo más complejo, vamos a comenzar con el análisis univariado, analizando las variables de a una por vez. Más adelante retomaremos el concepto de análisis exploratorio de los datos.

Datos Numéricos y Categóricos

Las variables son características que pueden tomar valores diferentes de una unidad a otra, como los ingresos de los contribuyentes, la cantidad de asalariados que integran la nómina de personal de cada empleador o la duración de un contrato.

Las variables pueden ser numéricas o categóricas. Los datos son los valores observados de las variables.

Datos Numéricos

Entre las variables numéricas encontramos:

- **Variables Discretas:** son los números enteros. No existen valores intermedios entre dos valores consecutivos.
- **Variables Continuas:** son los números reales. Existen infinitos valores intermedios entre dos valores. Por ejemplo: montos de ingresos, pesos de una mercadería.

Datos Categóricos

Entre las variables categóricas encontramos:

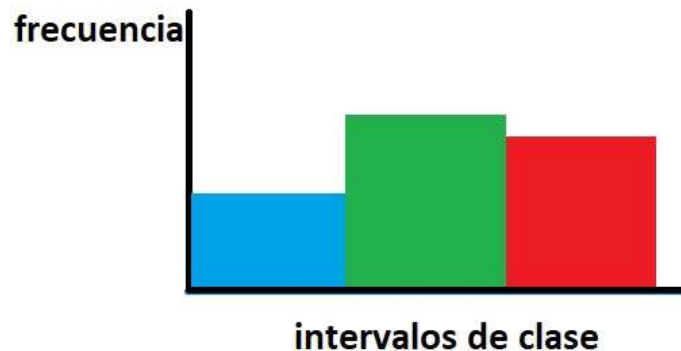
- **Variables Nominales:** no tienen un orden evidente. Por ejemplo: el estado civil de una persona, los apellidos y nombres, la denominación de las provincias.
- **Variables Ordinales:** siguen un orden, pero sin establecer una distancia. Por ejemplo: las categorías del Régimen Simplificado para Pequeños Contribuyentes (A, B, C, D, E, ...).

Distribución de Frecuencias de una Variable Numérica

Una forma de presentar un grupo de observaciones es mediante las **tablas de distribución de frecuencias**. La **distribución** de una variable nos dice cuáles son los valores que puede tomar. Por su parte, la **frecuencia** nos indica cuántas veces ocurre cada uno de sus valores.

Las tablas de frecuencias permiten conocer la **distribución** (ya sea en una población o en una muestra) de los valores de una variable.

Para representar gráficamente la distribución de los datos correspondientes a una variable numérica (discreta o continua) se utilizan tablas de frecuencias y un gráfico: el **histograma**.



Un **histograma** representa la distribución de una variable numérica en una población o en una muestra. En el eje horizontal el histograma representa los valores de una variable numérica divididos en intervalos de clase.

Los intervalos de clase de una variable discreta están centrados en sus valores posibles y tienen la misma longitud. Su representación gráfica tiene una barra sobre cada intervalo cuya altura indica la cantidad (frecuencia) o proporción (frecuencia relativa) de datos. No se deja espacio entre las barras o rectángulos.

Cuando los valores posibles de la variable numérica son pocos, la altura de cada rectángulo del histograma muestra directamente la cantidad o proporción de veces que cada uno de los valores ocurrió. Cuando son muchos, es necesario agruparlos definiendo previamente los intervalos.

La **frecuencia** es el número de veces que se repite cada valor o rango de valores de una variable. Vamos a diferenciar cuatro subtipos de frecuencia:

- **Frecuencia Absoluta** (f_i): La frecuencia absoluta es el número de repeticiones que presenta cada valor de una variable. La suma de todas las frecuencias da como resultado el número total de datos de la muestra.
- **Frecuencia Absoluta Acumulada** (F_i): La frecuencia absoluta acumulada es la suma de las frecuencias absolutas para todos los eventos iguales o anteriores que un cierto valor, en una lista ordenada de eventos.

- **Frecuencia Relativa** (h_i): La Frecuencia relativa es la proporción de cada frecuencia absoluta. Es decir, el número de veces que se produce ese resultado (frecuencia absoluta) dividido por el número total de datos observados.
- **Frecuencia Relativa Acumulada** (H_i): La frecuencia relativa acumulada es el resultado de dividir cada frecuencia absoluta acumulada por el número total de datos.

Veamos estos conceptos con un ejemplo concreto. Se consulta a 24 contribuyentes cuál es su ingreso anual en millones de pesos. Los resultados se exponen en la tabla a continuación:

2	3	4	5	5	5	6	7
3	4	4	5	5	5	6	7
3	4	4	5	5	6	6	8

A partir de dichas respuestas podemos armar la siguiente tabla de distribución de frecuencias:

Tabla N.º 2. Ingresos. En millones de pesos.

Ingreso anual	Frecuencia absoluta (f_i)	Frecuencia absoluta acumulada (F_i)	Frecuencia Relativa (h_i)	Frecuencia Relativa Acumulada (H_i)
2	1	1	$1 / 24 = \mathbf{0,042}$	0,042
3	3	$1 + 3 = \mathbf{4}$	$3 / 24 = \mathbf{0,125}$	$0,042 + 0,125 = \mathbf{0,167}$
4	5	$4 + 5 = \mathbf{9}$	$5 / 24 = \mathbf{0,208}$	$0,167 + 0,208 = \mathbf{0,375}$
5	8	$9 + 8 = \mathbf{17}$	$8 / 24 = \mathbf{0,333}$	$0,375 + 0,333 = \mathbf{0,708}$
6	4	$17 + 4 = \mathbf{21}$	$4 / 24 = \mathbf{0,166}$	$0,708 + 0,166 = \mathbf{0,875}$
7	2	$21 + 2 = \mathbf{23}$	$2 / 24 = \mathbf{0,083}$	$0,875 + 0,083 = \mathbf{0,958}$
8	1	$23 + 1 = \mathbf{24}$	$1 / 24 = \mathbf{0,042}$	$0,958 + 0,042 = \mathbf{1}$
Total	24	-	1,00	-

Variables Numéricas

Variables Numéricas Discretas

Una variable numérica es **discreta** cuando únicamente puede tomar valores dentro de una sucesión determinada de números.

Los histogramas tienen la virtud de representar gráficamente la distribución de una variable numérica bajo análisis. El eje horizontal del histograma representa los valores de la variable numérica divididos en intervalos de clase. La altura de cada barra nos indicará la "**frecuencia absoluta**" de la variable para cada intervalo de clase.

Si la muestra es representativa de los contribuyentes totales en ese momento, podremos considerar al histograma como una estimación de la distribución de la variable.

Una variable numérica es **discreta** cuando únicamente puede tomar valores dentro de una sucesión determinada de números. Por ejemplo, si medimos la cantidad de contribuyentes que componen un sector económico presentamos una variable discreta porque nunca vamos a encontrarnos con 1563,52 contribuyentes.

Variables Numéricas Continuas

En cambio, una variable numérica es **continua** cuando, dados dos valores posibles de la variable, ésta siempre puede tomar cualquier valor intermedio. El ingreso de los contribuyentes es una variable numérica continua, puede tomar valores como \$3.330.520 o \$3.330.521 pero también \$3.330.520,58.

Intervalo Abierto y Cerrado

Al medir una variable numérica continua, dados dos valores posibles de la variable, ésta siempre puede tomar cualquier valor intermedio.

El intervalo se representa mediante una expresión del tipo:

- $a < x < b$
- $(a; b)$
- $(a-b)$.

Para definir la exclusión o inclusión de los valores de cada extremo del intervalo vamos a recurrir a un otro concepto adicional: intervalo abierto e intervalo cerrado.

Para expresar esta diferencia recurriremos a la utilización de corchetes y paréntesis para indicar si el intervalo incluye o no a los valores extremos. Veamos un ejemplo concreto: el intervalo $[a-b)$. En este caso, donde “a” y “b” son números reales cualesquiera, con “a” menor que “b”, el intervalo es cerrado en “a” (incluye el valor “a”) y es abierto en “b” (no incluye el valor “b”).

Un intervalo abierto es aquel que no incluye los extremos entre los cuales está comprendido, pero sí todos los valores ubicados entre estos.

Variables Categóricas

Los datos categóricos son valores de variables categóricas. Al presentar las variables categóricas surgen múltiples inquietudes: ¿Cómo conviene mostrar los datos categóricos?, ¿expresados en cantidades o en porcentajes? En consecuencia, ¿cómo se aprecian mejor gráficamente sus magnitudes?: ¿con gráficos circulares o con gráficos de columnas?

Gráficos Circulares

El **gráfico circular** nos muestra las relaciones proporcionales entre las diferentes variables y nos permite ver a simple vista la importancia de cada elemento respecto del resto.

A continuación, vamos a presentar un ejemplo basado en datos reales de la presión tributaria. Como componentes de la “**presión tributaria**” hemos identificado seis variables categóricas relacionadas con la clasificación internacional de los ingresos públicos recaudados por la Administración Tributaria:

- Impuestos sobre bienes y servicios
- Impuestos sobre el ingreso, las utilidades y las ganancias de capital
- Aportes y contribuciones a la Seguridad Social
- Impuestos sobre el comercio exterior y las transacciones internacionales
- Impuestos sobre la propiedad
- Otros Impuestos

En la siguiente tabla se presenta el peso que tiene cada categoría dentro de la medición de la presión tributaria como porcentaje del producto interno bruto (PBI):

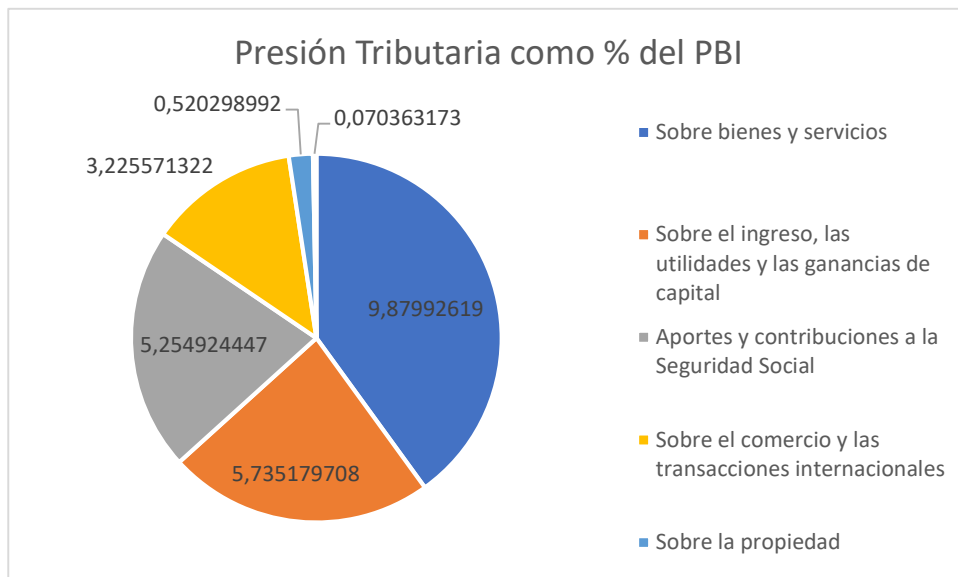
Tabla N.º 3. Presión Tributaria. En % del PIB (Año Base 2004). Año 2022

Sobre bienes y servicios	Sobre el ingreso, las utilidades y las ganancias de capital	Aportes y contribuciones a la Seguridad Social	Sobre el comercio y las transacciones internacionales	Sobre la propiedad	Otros Impuestos
9,87992619	5,735179708	5,254924447	3,225571322	0,520298992	0,070363173

Fuente: Elaboración propia a partir de “*Presión Tributaria según clasificación internacional*”. DNIAF. Recuperado de:
<https://www.argentina.gob.ar/economia/ingresospublicos/recaudaciontributaria>

Esta relación de proporcionalidad entre cada una de las categorías respecto del total de la presión tributaria lo podremos apreciar mejor si lo expresamos con un gráfico circular como el siguiente:

Gráfico N.º 1. Presión Tributaria. En % del PIB (Año Base 2004). Año 2022



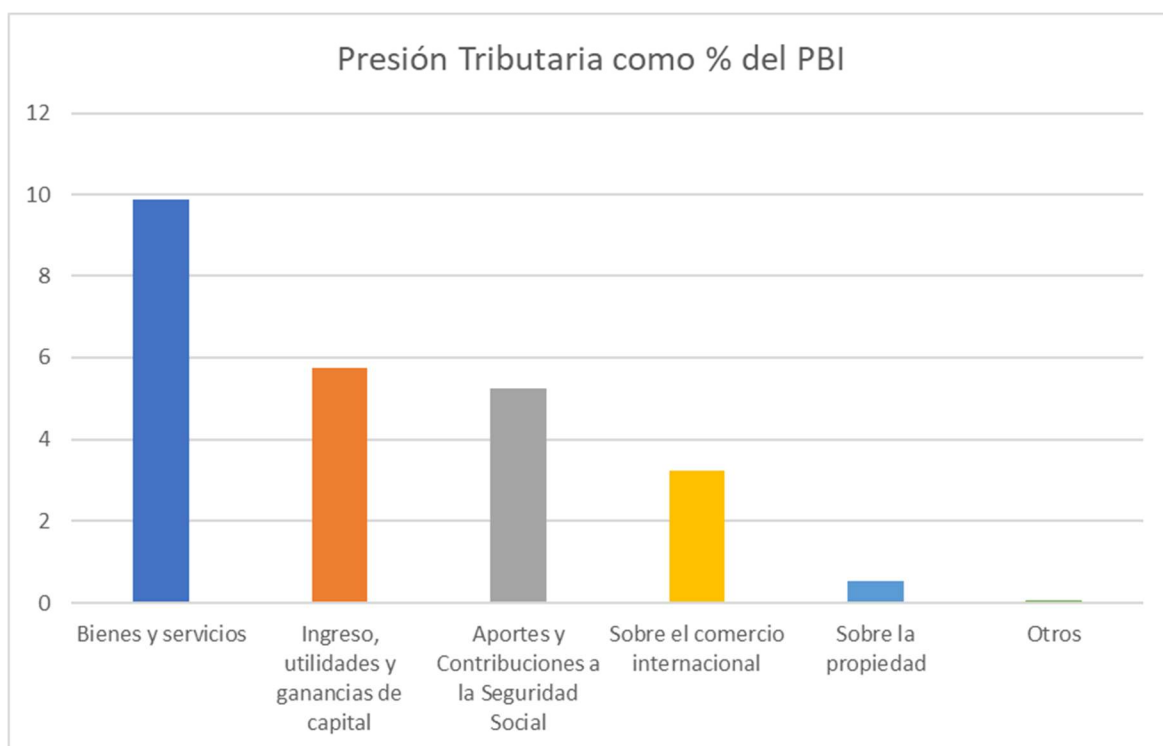
Gráficos de Columnas

Por su parte, los **gráficos de columnas** también pueden resultarnos de utilidad para representar la distribución de los valores de una variable categórica. Las categorías se representan en el eje horizontal y la cantidad de datos en el eje vertical. La altura de las columnas sobre cada categoría representa la cantidad de datos de cada una de ellas.

Tal como ocurre con los gráficos circulares, divide a los datos en grupos correspondientes a las categorías y muestra cuántos, o qué porcentaje de individuos pertenecen a cada categoría.

A continuación, vamos a utilizar los valores de las categorías de la Tabla N.º 3 presentada previamente para representar mediante un gráfico de columnas la distribución de dicha variable categórica que estamos analizando (tipo de impuesto que compone la presión tributaria según la clasificación internacional).

Gráfico N.º 2. Presión Tributaria. En % del PIB (Año Base 2004). Año 2022



Como se puede apreciar, con el gráfico de columnas también podemos comparar categorías.

Gráfico de Tiempo

Hasta aquí hemos desarrollado el análisis univariado (de una sola variable). Vamos ahora a empezar a transitar el análisis de múltiples variables. Por ejemplo, podemos comparar los valores de

varios trimestres de un mismo año o de diferentes años. Para estos casos podemos graficar la distribución por medio de un **gráfico de tiempo**. Este tipo de gráfico se utiliza para examinar la evolución a lo largo del tiempo de una o más variables.

Un **gráfico de tiempo** tiene una unidad de tiempo en el eje horizontal (como meses, trimestres o años) y en el eje vertical alguna cantidad (ingresos de los contribuyentes, tasa de ausentismo, ventas totales, etc.) o porcentajes como veremos en el ejemplo del gráfico a continuación.

En un gráfico de tiempo, para cada período la cantidad está representada por un punto, y el conjunto de los puntos están conectados por líneas. El tipo de datos utilizados en la proyección de ingresos se conoce como **datos de series temporales**. Esto significa que nuestros datos contienen observaciones durante varios períodos de tiempo para una variable específica, como los ingresos fiscales, el producto bruto interno (PBI) o un precio. Por ejemplo, a continuación, presentaremos la presión tributaria como porcentaje del PBI:

Tabla N.º 4. Presión Tributaria. En % del PBI (Año Base 2004). Años 2004 a 2022

Año	Sobre bienes y servicios	Sobre el ingreso, las utilidades y las ganancias de capital	Aportes y contribuciones a la Seguridad Social	Sobre el comercio y las transacciones internacionales	Sobre la propiedad	Otros Impuestos
2004	9,94206361	4,856533452	2,80315797	2,81221941	0,359443588	0,142887972
2005	9,81277868	5,013046464	2,986994281	2,802957804	0,324127457	0,150597047
2006	9,92832845	4,855907803	3,456775927	2,806123667	0,299312898	0,195428022
2007	10,2518446	4,930507825	4,081680819	3,0858178	0,285778798	0,122819105
2008	10,2087466	4,759910012	4,572260528	3,940866363	0,30006451	0,1385344
2009	10,3396966	4,558492698	6,188720386	3,205014391	0,330709356	0,254786442
2010	10,3173536	4,719620926	6,165891847	3,453036069	0,316004955	0,143604279
2011	10,2735270	5,053549172	6,295765343	3,182050146	0,27617498	0,169525906
2012	10,5452791	5,311314221	6,815101215	2,982559615	0,280855338	0,180420171
2013	10,7102461	5,545733292	7,050478433	2,387461249	0,312734705	0,092886513
2014	10,5340862	5,892196518	6,718729009	2,517592062	0,31830037	0,122282063

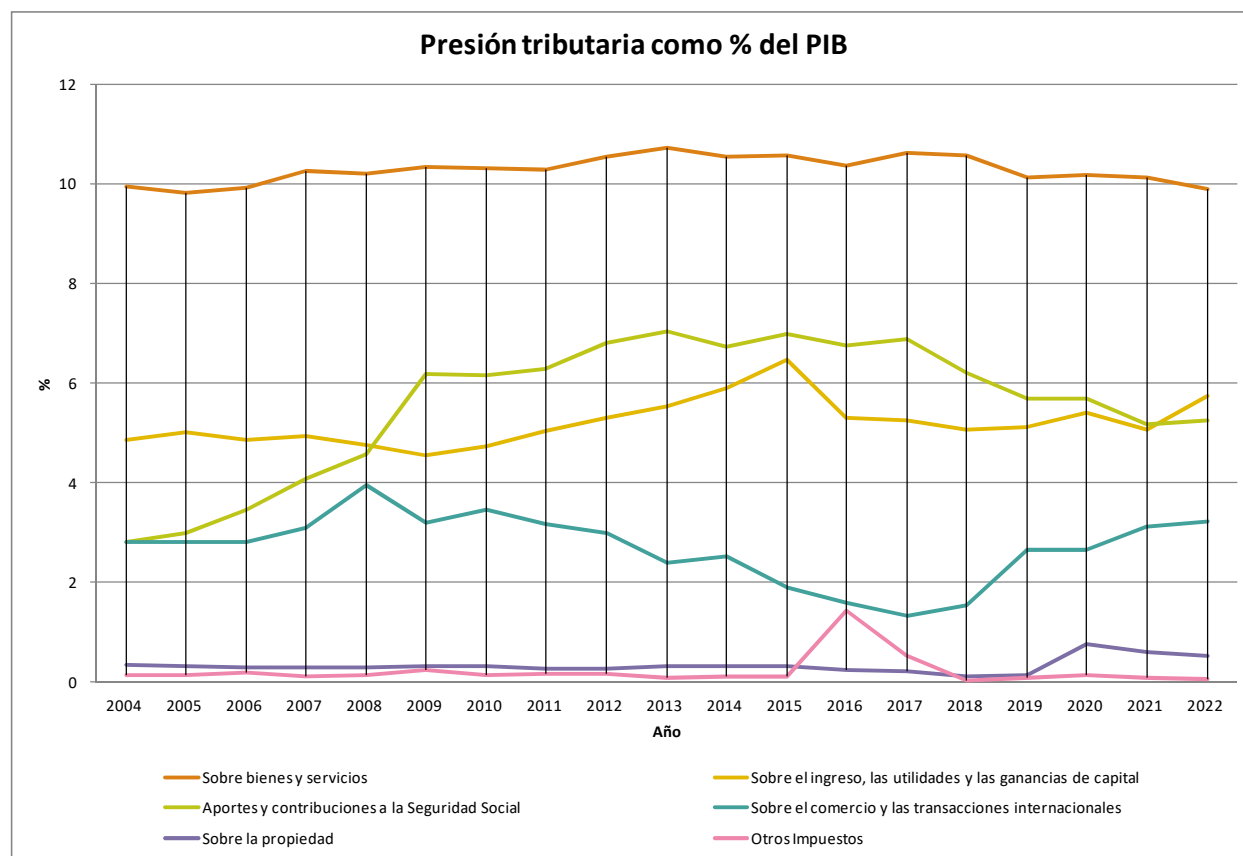
Proyección del impacto de futuras medidas tributarias

37

2015	10,5563960	6,457900859	6,976392318	1,898608701	0,311325198	0,116391376
2016	10,3608887	5,310146112	6,758097983	1,582651343	0,242776484	1,429165387
2017	10,6271288	5,250877083	6,880966811	1,32616767	0,213750066	0,531658005
2018	10,5739415	5,066147497	6,201747442	1,542422637	0,10373705	0,050232849
2019	10,1140155	5,116239624	5,685306313	2,647077139	0,15021141	0,090754291
2020	10,1757262	5,40374264	5,689316366	2,667321447	0,771518693	0,13003973
2021	10,1114369	5,07484279	5,174371025	3,128995363	0,611634311	0,079319814
2022	9,87992619	5,735179708	5,254924447	3,225571322	0,520298992	0,070363173

Fuente: Elaboración propia a partir de "Presión Tributaria según clasificación internacional". DNIAF.
Recuperado de: <https://www.argentina.gov.ar/economia/ingresospublicos/recaudaciontributaria>

Gráfico N.º 3. Gráfico de tiempo de relación de la presión tributaria con el PIB. 2004 a 2022



Fuente: Elaboración propia a partir de “Presión Tributaria según clasificación internacional”.

DNIAF. Recuperado de:

<https://www.argentina.gob.ar/economia/ingresospublicos/recaudaciontributaria>

En el gráfico de tiempo de la presión tributaria como porcentaje del PIB vemos la evolución de las seis variables categóricas (tipo de impuesto) desde el año 2004 hasta el año 2022. El valor de cada variable para cada año se define como un punto cuya altura se mide a partir de la escala de valores del eje vertical. Luego se unen los puntos que representan los valores para cada año, mostrando picos y valles que siguen la evolución de cada variable categórica.

Como veremos más adelante, el análisis de las series temporales por medio de los gráficos de tiempos es el primer paso es inspeccionar el comportamiento de los datos cuando deseamos evaluar las características clave de una muestra y determinar el enfoque de la proyección que vamos a utilizar. Gracias a los gráficos de temporales con un primer golpe de vista podemos identificar aumentos o disminuciones (picos y valles), pero también podremos detectar regularidades o patrones en los datos (tendencias o ciclos).

Distribución y Posición Relativa

Tendencia Central o Valor Central

Cualquier conjunto de datos tiene dos propiedades importantes: un **valor central** y la **dispersión** alrededor de ese valor.

Para comenzar desde lo más simple vamos a apreciar dichas propiedades con un caso real y en la siguiente tabla presentaremos la cantidad de personas empleadas según el tamaño de la nómina de sus empleadores.

A los efectos prácticos de este ejemplo en la tabla que expondremos a continuación no incluiremos a las personas que trabajan en grandes empleadores que posean nóminas de más de 5000 dependientes.

Tabla N.º 5. Trabajadores según tamaño de la nómina de los empleadores. Cápitas. Empleadores de hasta 5000 dependientes. Junio de 2023

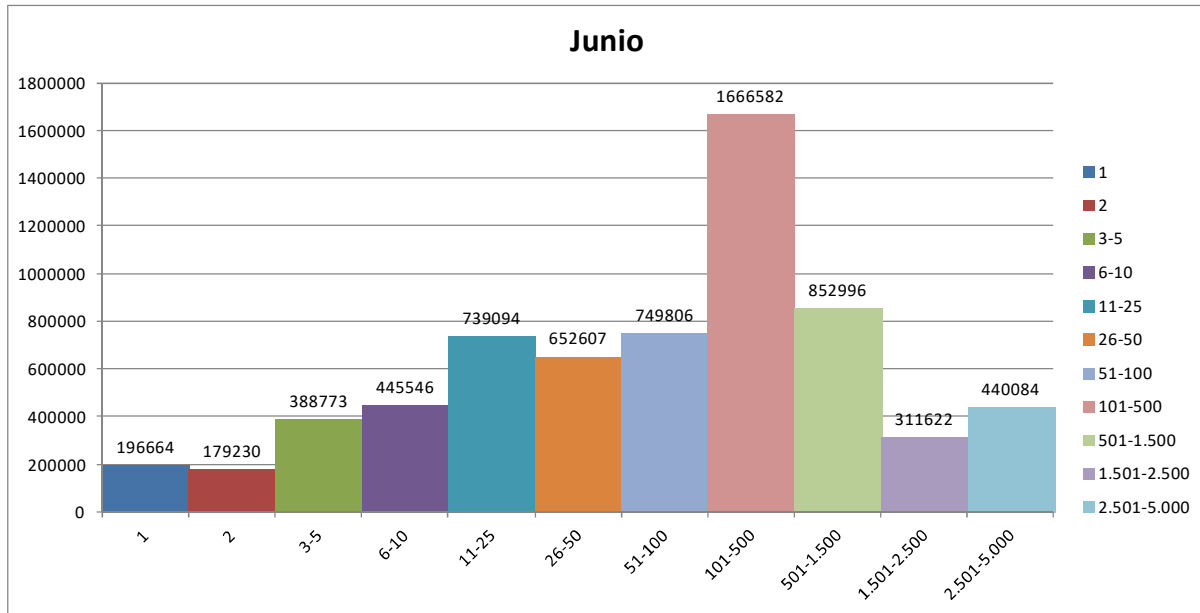
Cantidad de trabajadores	Junio/2023
1	196.664
2	179.230
3-5	388.773
6-10	445.546
11-25	739.094
26-50	652.607
51-100	749.806
101-500	1.666.582
501-1.500	852.996
1.501-2.500	311.622
2.501-5.000	440.084

Fuente: Elaboración propia a partir de “Boletín anual de Seguridad Social. Cuadro 3.2. Trabajadores según tamaño de empleador. Información SIPA”. Junio de 2023. AFIP. Recuperado de:

<https://contenidos.afip.gob.ar/institucional/estudios/archivos/Boletin-anual-de-Seguridad-Social-2023-con-apertura.xls>

Para poder apreciar mejor la distribución de los datos, vamos a exponer los valores de la tabla previa en un histograma que agrupa cada intervalo de clase en una barra de un color diferente:

Gráfico N.º 4. Histograma de trabajadores según tamaño de la nómina de los empleadores. Empleadores de hasta 5000 trabajadores. Abril de 2023



En el ejemplo vemos el valor central en el grupo de 101 a 500 y luego la dispersión de valores alrededor de ese valor hacia izquierda y derecha. El promedio define el valor característico o central de un conjunto de números. Sin embargo, como existen varios métodos para calcular el promedio, el método utilizado puede influir en las conclusiones. Cuando nos referimos al concepto de **“tendencia central”** estamos hablando de la ubicación del **“centro”** de la distribución de las variables. Las formas principales para obtener un valor central o promedio son tres:

- **Media:** Se obtiene sumando todos los valores del conjunto de datos y dividiendo la suma por la cantidad de datos en ese conjunto. Es el promedio de los datos obtenidos de la muestra. En su cálculo intervienen todos los valores observados de la variable.
- **Mediana:** Es el valor central del conjunto de datos ordenados. Deja al 50% de las observaciones por debajo de sí y al otro 50% de las observaciones por encima. Es

robusta y no es fácilmente influenciada por la presencia de valores atípicos extremos (outliers).

- **Moda:** Es el valor observado de mayor frecuencia. Un mismo grupo de datos puede tener más de una moda.

Supongamos que tenemos un conjunto con n observaciones representadas por n datos. Los valores se leen equis uno, equis dos, equis ene y se pueden representar en una tabla:

Observación	1	2	3	...	n
Valor	X_1	X_2	X_3	...	X_n

Bajemos la teoría a un ejemplo concreto. Le preguntamos a 5 contribuyentes ($n = 5$) cuántos millones facturó en el último año y obtenemos como respuesta:

- $x_1 = 13$,
- $x_2 = 17$,
- $x_3 = 8$,
- $x_4 = 29$,
- $x_5 = 18$

Contribuyente	N.º 1	N.º 2	N.º 3	N.º 4	N.º 5
Facturación	13	17	8	29	18

Para definir cuál es el “valor central” de estos datos llegaremos a la respuesta utilizando la media y la mediana.

Media

La **media** se obtiene sumando todos los datos y dividiendo por la cantidad total **n** de observaciones. La media² se representa por μ .

La suma de todos los datos ($x_1 + x_2 + \dots + x_n$) se suele representar con la letra griega sigma mayúscula (comúnmente llamada sumatoria): Σ .

$$\mu = \frac{1}{n} \sum_{t=1}^n X_t = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Usemos el ejemplo previo y calculemos la media con los datos que presentamos:

$$\mu = \Sigma / 5 = (13 + 17 + 8 + 29 + 18) / 5 = 85 / 5 = 17$$

A partir de la **media**, la respuesta del valor central de los 5 datos recolectados sería **17**.

Mediana

La **mediana** es otro tipo de centro. La mediana es el punto central de los datos. La **mediana** deja la misma cantidad de datos a cada lado. Si volvemos al ejemplo previo de los 5 contribuyentes a los cuales les consultamos el volumen de su facturación anual, para hallar la mediana del conjunto de datos (13, 17, 8, 29, 18) del ejemplo, lo primero que debemos hacer es ordenar los datos de menor a mayor:

Contribuyente	N.º 1	N.º 2	N.º 3	N.º 4	N.º 5
---------------	-------	-------	-------	-------	-------

² Además de μ , la media también se puede representar con una “x” con una raya encima.

Facturación	8	13	17	18	29
--------------------	---	----	-----------	----	----

- Luego, la mediana es esta muestra será el valor central: 17.

Antes de avanzar vamos a introducir un concepto adicional. Supongamos que solamente consultamos a cuatro contribuyentes. La regla general para calcular la mediana de n datos ordenados es:

- Si la cantidad de datos es impar: la mediana es el valor del centro. Es decir, se encuentra en la posición $(n+1) / 2$
- Si la cantidad de datos es par: la mediana es el promedio de los dos valores centrales. Es decir, se encuentra en la posición entre $(n/2)$ y $(n/2) + 1$

Así con cuatro respuestas:

Contribuyente	N.º 1	N.º 2	N.º 3	N.º 4
Facturación	8	13	17	18

- Luego, la mediana es esta muestra será el valor: $(13 + 17) / 2 = 15$.

Valores Atípicos

Tanto la media como la mediana tienen ventajas y desventajas. La media utiliza todos los datos para su cálculo. Si los datos presentan un histograma simétrico calcular la media es lo mejor para obtener el centro de los datos, en este caso la mediana será muy parecida.

Siguiendo con el ejemplo del tamaño de facturación de los 5 contribuyentes, la media y la mediana serán muy parecidas. Pero, ¿qué pasa si aparece un valor atípico que se aleja mucho de los demás? Si se presenta un valor atípico la mediana no se verá afectada, pero la media sí.

Para que podemos apreciar esta situación, vamos a hacer un cambio en el ejemplo que venimos usando. Supongamos que el quinto contribuyente en lugar de responder 29, nos dice 99. De repente nos aparece un valor que se aleja del patrón general de distribución que veníamos observando (en inglés se los llama **outliers**) y nosotros lo denominaremos como “**valor atípico**”. El valor atípico puede estar originado en un error en la captura de los datos, pero también puede explicarse porque elegimos un individuo con una conducta diferente (por ejemplo, un contribuyente más grande que los otros).

La mediana seguirá siendo 17:

Observación	1	2	3	4	5
Valor	8	13	17	18	99

Pero la media será **31**:

$$x = (8 + 13 + 17 + 18 + 99) / 5 = 155 / 5 = 31$$

¿Es razonable decir que 31 es el tamaño promedio del volumen de facturación de los contribuyentes? Claramente, cuando aparece un valor atípico, la media ya no representa a la mayoría de los datos, por eso, decimos que la media es sensible ante la presencia de valores atípicos.

Medidas Resumen y Dispersión

A medida que aumenta el volumen de datos a analizar se vuelve más difícil comprender intuitivamente la información. Como hemos visto en las tablas y gráficos presentados hasta aquí, captar la información que aportan diez números es comprensible, cuando ya hablamos de cien se nos hace difícil y cuando queremos presentar mil estamos en graves problemas.

Una manera de achicar la cantidad de datos a exponer es recurrir a las llamadas “**medidas resumen**”, que de alguna manera deben describir las características más sobresalientes del conjunto que se está analizando.

Una medida resumen es un número. Se obtiene a partir de una muestra y, en cierta forma, la caracteriza. Es el valor de un **estadístico**. Por ejemplo, un porcentaje o una proporción son medidas resumen. Las medidas resumen permiten tener una idea rápida de cómo son los datos. Pero, un estadístico mal utilizado puede dar una idea equivocada respecto de las características generales que interesa mostrar.

El cálculo de medidas resumen es el primer paso; se realiza cuando se recolectan los datos en un estudio para tener una idea de qué está pasando y cuál es el comportamiento de los datos de la muestra. Posteriormente, se pondrán a prueba las hipótesis respecto a algún parámetro poblacional, se estimarán características de la población y se estudiarán posibles relaciones entre las variables. Cuando se presenta el informe final con las conclusiones (en nuestro caso las proyecciones de la recaudación), las medidas resumen muestran los resultados en forma concisa y clara, volviendo a tener importancia.

En principio, se pueden obtener muchísimas formas de resumir los valores de un conjunto de datos numéricos. Es importante que sean fáciles de interpretar. La **dispersión** busca obtener el grado de concentración de los datos respecto del “valor central”. La **media** y el **desvío estándar** son las medidas resumen más utilizadas.

Varianza

La **varianza** es la suma de los cuadrados de los desvíos con respecto a la **media muestral**, dividida por (n-1). La varianza también se puede representar con la letra griega sigma elevada al cuadrado: σ^2

La fórmula para calcular la varianza depende de si se trata de una muestra o de una población. Para una población, la fórmula es:

$$\sigma^2 = (\sum (x-\mu)^2) / N$$

Donde σ^2 es la varianza, \sum es la suma, X es el valor de la variable, μ es la media y N es el tamaño de la población.

En cambio, si se está trabajando con una muestra, la fórmula para calcular la varianza es:

$$\sigma^2 = [\sum_{i=1}^n (x_i - \mu)^2] / (n-1)$$

Donde σ^2 es la varianza muestral, X_i es el valor de la i -ésima observación, μ es la media muestral y n es el tamaño de la muestra.

En síntesis, aunque parezca un trabalenguas, podemos decir que la varianza sintetiza la variabilidad de una variable respecto al valor medio.

Desvío Estándar

La **mediana** y los **cuartiles** son muy sencillos de calcular a mano cuando la cantidad de datos es relativamente pequeña. Pero cuando se tienen muchos datos es muy difícil ordenarlos. Por esa razón, aunque la mediana era conocida casi no se utilizaba antes de la generalización en el uso de las computadoras para el análisis de datos.

La **media** y el **desvío estándar** son muy buenos para resumir datos con histogramas razonablemente **simétricos** y **sin valores atípicos**.

La **media**, es mucho más fácil de calcular a mano cuando hay muchos datos. Sólo requiere del uso de operaciones aritméticas, para hallar un número representativo de la mayoría de los datos.

El **desvío estándar** es una medida de dispersión basada en la media y utiliza todos los datos. El desvío estándar representa una distancia típica de cualquier punto del conjunto de datos a su centro (medido por la media). Es una distancia promedio de cada observación a la media. El desvío estándar de los datos de toda una población (desvío estándar poblacional) se denota con la letra griega (sigma minúscula): σ .

$$\sigma = \sqrt{\frac{1}{n} \sum_{t=1}^n (X_t - \mu)^2} = \sqrt{\frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \dots + (X_n - \mu)^2}{n}}$$

Desvío Estándar Muestral

Muchas veces los parámetros poblacionales son desconocidos y, por lo tanto, se recurre al cálculo de un **estimador** utilizando una muestra. Llamaremos "**S**" al **desvío estándar muestral**.

La distinción entre el desvío estándar poblacional y el desvío estándar muestral vale para todos los estadísticos descriptos (media, mediana, cuartiles, distancia Inter cuartil, etc.).

Tal como vimos previamente, si el cálculo de un estadístico se realiza utilizando una muestra para estimar un parámetro, el resultado tendrá un error de muestreo.

El desvío estándar se calcula promediando la diferencia entre cada dato y la media, elevadas al cuadrado. Como este resultado tiene las unidades al cuadrado, luego se saca la raíz cuadrada. Para un conjunto de **n** datos:

- Se calcula la distancia de cada dato a la media: $X_i - \mu$
- Se eleva al cuadrado: $(x_i - \mu)^2$

- Se promedia dividiendo por $n-1$ y, así, se obtiene la varianza muestral (S^2)
- Por último, se calcula la raíz cuadrada.

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2}$$

Varianza Muestral y Dispersión

El **desvío estándar** es la raíz cuadrada de la varianza.

La **varianza** es la suma de los cuadrados de los desvíos con respecto a la **media muestral**, dividida por ($n-1$). Cuanto más grande es la varianza muestral, más dispersos están los datos.

Una medida de dispersión debe tener las mismas unidades que los datos. La **varianza muestral** está elevada al cuadrado, entonces por supuesto, debemos sacar la raíz cuadrada para obtener el **desvío estándar**.

Posición Relativa: Dispersión y Variabilidad

Los métodos estadísticos son imprescindibles para analizar los datos debido a su variabilidad.

Para tener medidas que capten dicha variabilidad de la mejor manera posible, además de conocer el punto central de un conjunto de datos, también nos interesa describir su dispersión, es decir

cuán lejos tienden a estar los datos de su centro para poder dividir una muestra ordenada en distintas partes iguales (cuantiles). La variabilidad está presente en todos los conjuntos de datos. Sea cual fuere la característica, es muy difícil que dos mediciones sean idénticas. Esto se debe a que:

- Diferentes individuos tienen diferentes características, al cuantificarlas resultan en valores diferentes de las variables correspondientes.
- Diferentes mediciones de una misma característica dan como resultado diferentes valores debido al inevitable error de medición.

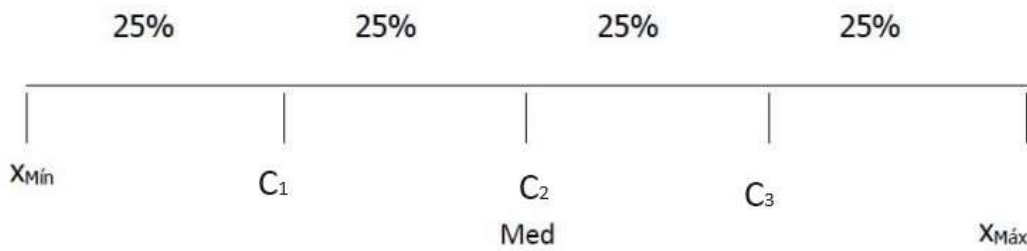
Rangos de Valores: Mínimo y Máximo

El **rango** es la diferencia entre el valor máximo observado y el valor mínimo observado. El rango de valores donde se encuentran los datos permite apreciar su variabilidad o dispersión (cuán desparrramados están). La medida natural para evaluar dicha dispersión es la distancia entre el valor máximo y el valor mínimo de los datos. El rango no distingue dos conjuntos de datos con diferentes dispersiones. El rango como medida tiene algunos inconvenientes:

- Es muy sensible a la presencia de valores atípicos.
- Como utiliza solamente dos datos, no puede distinguir dos conjuntos con máximos y mínimos coincidentes donde uno tiene la mayoría de sus valores mucho más concentrados que el otro.

Cuantiles y Distancia Inter cuantil

Las medidas de posición relativa se llaman en general **cuantiles**. La división en cuantiles se utiliza para corregir ciertos problemas de la distribución. Cada parte contiene el mismo número de elementos o igual tamaño de intervalo y divide a una muestra ordenada en distintas partes iguales. Por ejemplo, si se divide en cuatro tendremos 4 cuantiles:



Ya conocemos la mediana y, al presentar el concepto de “**rango**” también hemos presentado los valores mínimos y máximos observados. Para calcular los cuartiles se utiliza la distancia entre el valor mínimo y el valor máximo del 50% central de los datos. A dicha distancia la llamaremos **distancia Inter cuartil**.

100% de los datos			
		50% de los datos	
Cuartil 1	Cuartil 2	Cuartil 3	Cuartil 4
	C₁	C₂	C₃

¿Cómo se Calcula la Distancia Inter cuartil?

- Se ordenan los datos.
- Se calcula la mediana (**C₂**), que los divide en dos partes con igual cantidad de datos de cada lado.
- Se calcula la mediana de la mitad más baja (grupo inferior): es el cuartil inferior (**C₁**).
- Se calcula la mediana de la mitad más alta (grupo superior): es el cuartil superior (**C₃**).
- La distancia Inter cuartil (DIC) es la diferencia entre el cuartil superior y el cuartil inferior: **DIC = C₃ - C₁**.

Cuando la mediana coincide con uno de los datos se la puede considerar parte de los dos grupos, el superior y el inferior (esta regla es arbitraria y algunos autores no la cuentan en ninguno de los dos).

¿Qué Mide la Distancia Inter cuartil?

Como medida de dispersión, la distancia Inter cuartil mide la longitud del intervalo en el cual se encuentra el 50% central de los datos. Cuanto más dispersos estén los datos, mayor será la distancia Inter cuartil.

Números Resumen y Gráfico de Caja y Brazos

Cuando hay valores atípicos presentes o los histogramas son asimétricos es mejor utilizar la **mediana** y la **distancia Inter cuartil**, y mejor aún, los 5 **números resumen**. Hay cinco números que nos dan una idea de cómo está distribuido un conjunto de datos:

- X_{\min} : Valor Mínimo
- C_1 : mediana de la mitad más baja (grupo inferior), es el cuartil inferior.
- C_2 : Mediana
- C_3 : mediana de la mitad más alta (grupo superior), es el cuartil superior
- X_{\max} : Valor Máximo.

A estos cinco números se los llama: “**números resumen**”. Si suponemos que los cinco números resumen asumen los siguientes valores:

Mínimo	C ₁	M	C ₃	Máximo
37	51	58	67	85

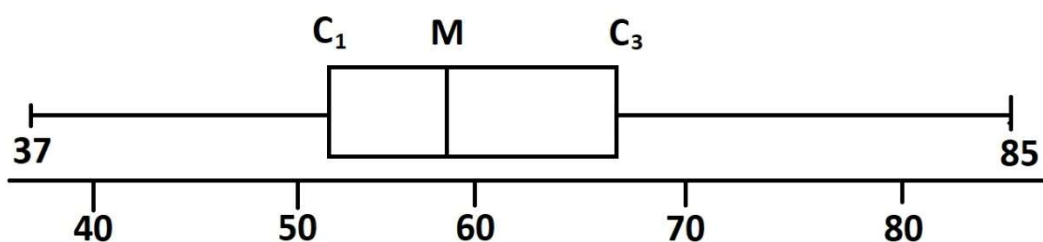
El 50% de los datos se encuentran entre el cuartil inferior y el cuartil superior.

100% de los datos						
		50% de los datos				
Cuartil inferior		Cuartil 2		Cuartil 3		Cuartil superior
51		58		67		

John Tukey creó el gráfico de caja y brazos (*boxplot*) en 1977. Este gráfico se utiliza para resumir información utilizando las 5 medidas estadísticas llamadas “**números resumen**”:

- X_{\min} : **Valor mínimo**: El valor más pequeño en el conjunto de datos.
- C_1 : **Primer cuartil** (Q1 o percentil 25): El valor que divide los datos en dos partes, donde el 25% de los datos son menores que este valor. Es la mediana de la mitad más baja (grupo inferior), es el cuartil inferior.
- C_2 : **Mediana** (Q2 o percentil 50): El valor medio del conjunto de datos.
- C_3 : **Tercer cuartil** (Q3 o percentil 75): El valor que divide los datos en dos partes, donde el 75% de los datos son menores que este valor. Es la mediana de la mitad más alta (grupo superior), es el cuartil superior.
- X_{\max} : **Valor máximo**: El valor más grande en el conjunto de datos.

Si los alineamos en una escala numérica, podemos representarlos en un Gráfico de caja y brazos de la siguiente manera:



Analicemos el gráfico:

- Los cuartiles forman los bordes de la caja (51 y 67),
- La mediana está dentro de la caja (58),
- Los brazos se extienden desde cada borde de la caja hasta los valores mínimo y máximo respectivamente (37 y 85).

Estas medidas se representan en una sola dimensión en el gráfico de caja y brazos. Los cuartiles son los valores que dividen los datos en cuatro partes iguales o cuartos. La longitud de la caja representa la distancia entre los percentiles 25 y 75, es decir, el rango *intercuartílico*. Esto incluye el 50% de las observaciones centrales.

Los gráficos de caja y brazos muestran un resumen de la distribución de los datos en muy poco espacio. Además, las líneas verticales (brazos o bigotes) representan los valores máximos y mínimos del conjunto de datos. Sirven como referencia para ubicar las observaciones que están por fuera del 50% central de la distribución. Por tanto, estos diagramas descriptivos son muy valiosos para observar los valores atípicos (*'outliers'*) y las dispersiones de los puntos con respecto a los percentiles.

El gráfico de caja y brazos (también llamado “*boxplot*” por su nombre en inglés) es la representación más visible y útil para graficar los cinco números resumen. El gráfico de caja y brazos describe más precisamente el rango donde se encuentran los datos. El rango Inter cuartil que forma la caja contiene el 50% de los datos y los brazos se extienden hasta el último dato de cada lado. Si tuviéramos datos atípicos, estos serán graficados fuera de la caja.

Los gráficos de caja y brazos sirven especialmente cuando queremos comparar varios conjuntos de datos. Por ejemplo, es útil para comparar cotizaciones de diferentes acciones o Commodities transables como los cereales o el barril de petróleo (valor de apertura, valor de cierre y cotizaciones durante la jornada).

Percentiles

Como hemos señalado, el cuartil inferior es el percentil 25 y la mediana es el percentil 50. En general hablaremos del percentil K. Si decimos percentil 25, eso significa $K = 25$.

100% de los datos						
	50% de los datos					
Cuartil inferior	Cuartil 2		Cuartil 3		Cuartil superior	
	51		58		67	

¿Cómo se Calcula un Percentil en un Conjunto de Datos No Agrupados?

Veamos ahora una forma general para hallar el percentil K para cualquier conjunto de datos:

- Paso 1. Ordene los datos de menor a mayor
- Paso 2. Calcule $K/100$ y multiplíquelo por la cantidad total de datos n.
- Paso 3. Redondee $n \cdot K/100$ al entero más cercano.

- Paso 4. Cuente desde el dato más chico hacia el más grande tantos lugares como el número hallado en el paso 3.

¿Cómo se Calcula un Percentil en un Conjunto de Datos Agrupados?

Cuando los datos tienen muchos valores repetidos es más conveniente utilizar una tabla de frecuencias para calcular los percentiles.

La frecuencia relativa acumulada (H_i) nos indicará el percentil que estamos buscando.

Para verlo en un ejemplo concreto, recordemos la Tabla N.º 2 que presentamos previamente con los ingresos de 24 contribuyentes:

Tabla N.º 2. Ingresos. En millones de pesos.

Ingreso anual	Frecuencia absoluta (f_i)	Frecuencia absoluta acumulada (F_i)	Frecuencia Relativa (h_i)	Frecuencia Relativa Acumulada (H_i)
2	1	1	0,042	0,042
3	3	4	0,125	0,167
4	5	9	0,208	0,375
5	8	17	0,333	0,708
6	4	21	0,166	0,875
7	2	23	0,083	0,958
8	1	24	0,042	1
Total	24	-	1,00	-

Si buscamos el percentil del 70% y nos guiamos por la columna de la H_i . El valor más cercano es 0,708 (o lo que equivale al 70,8%) que corresponde a un ingreso anual de 5 millones de pesos.

Formas que Describen las Distribuciones

Cuando presentamos los histogramas mediante curvas podemos descubrir algunos patrones especiales. Existe una diferencia para resaltar entre histogramas y **curvas de densidad**. Las curvas de densidad se grafican en escala de densidad y los histogramas en escala de frecuencias o frecuencias relativas.

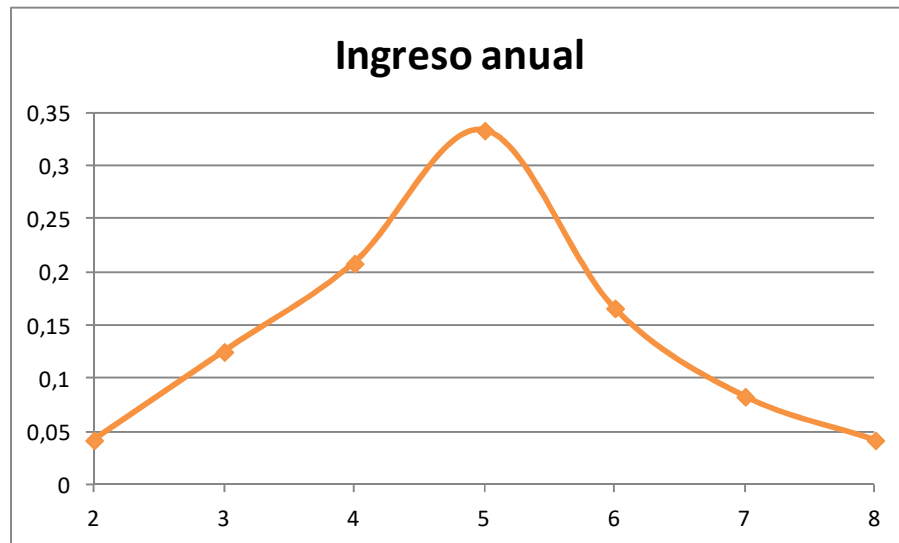
Histograma	Curva de densidad
Escala de frecuencias o frecuencias relativas	Escala de densidad
Área es proporcional a la frecuencia relativa	Área es igual a la frecuencia relativa

La mayoría de los histogramas muestran la cantidad (frecuencia) o proporción (frecuencia relativa) de observaciones de cada intervalo de clase mediante la altura del rectángulo. De esta manera, el área de cada rectángulo es proporcional a la frecuencia relativa.

En una escala de densidad, el área de cada rectángulo es IGUAL a la frecuencia relativa, y se obtiene graficando en el eje vertical la frecuencia relativa dividida la longitud del intervalo de clase. En escala de densidad, el área total de los rectángulos del histograma es 1.

Las expresiones de dichas curvas son descripciones matemáticas idealizadas y constituyen poderosas herramientas para representar la distribución de los datos. Son especialmente útiles cuando se trata de describir una cantidad muy grande de observaciones.

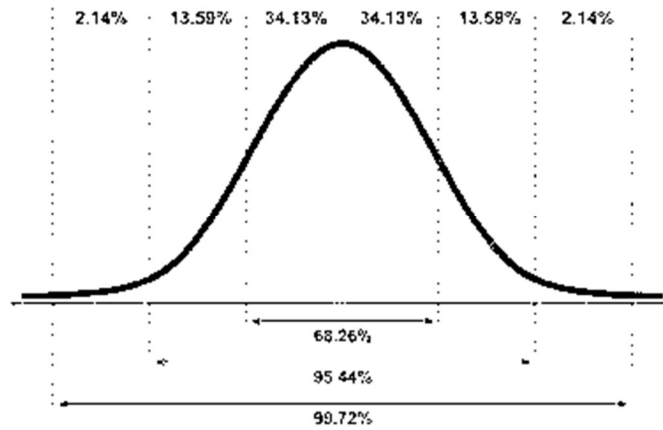
Si representáramos en un gráfico los valores de frecuencia relativa (h_i) de la Tabla N.º 2 tendríamos la siguiente curva:



Distribución Simétrica

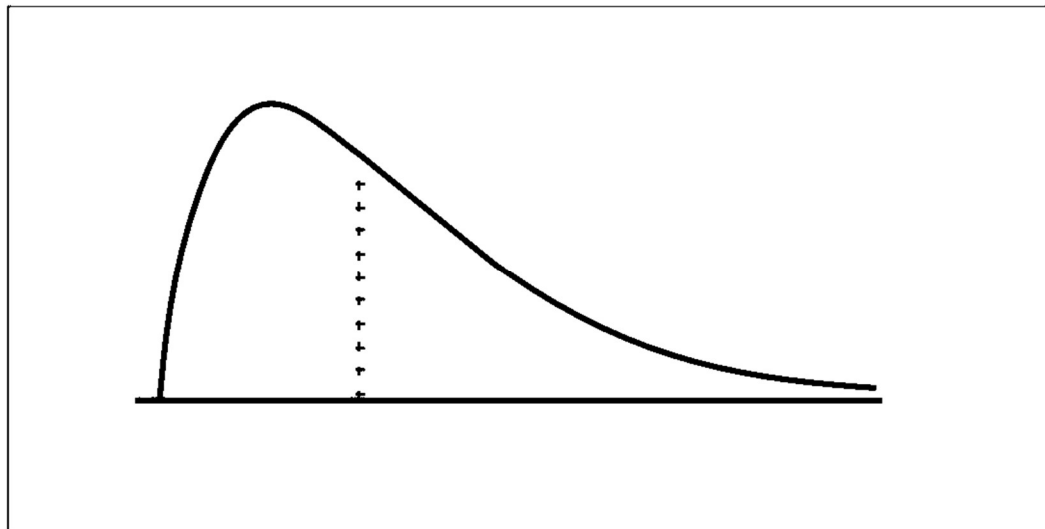
Una distribución es simétrica cuando sus dos mitades son imágenes reflejadas en un espejo.

Una distribución con forma de campana es una forma particular de una distribución simétrica con un montículo en el centro y dos caídas como toboganes hacia los costados. La distribución normal de Gauss es una de las distribuciones de datos que tal vez aparezca con más frecuencia y es la más estudiada.



Distribución Asimétrica

Una distribución es asimétrica cuando una de sus dos mitades es más fina y más larga que la otra. Por ejemplo, la distancia de los domicilios de los contribuyentes respecto de la agencia en la cual se encuentran inscriptos mostrará muchos valores pequeños, en la mitad izquierda del histograma, vemos los contribuyentes con domicilio cercano a la dependencia, mientras que habrá pocos valores grandes de los contribuyentes con domicilios más lejanos.



Por consiguiente, una distribución es asimétrica a izquierda cuando la mitad izquierda es más fina y más larga. Por ejemplo, cuando se miran los resultados de un examen que tenía preguntas fáciles, la mayoría de las notas serán altas y estarán amontonadas del lado derecho, con unas pocas notas bajas (las del lado izquierdo).

Curvas de Densidad

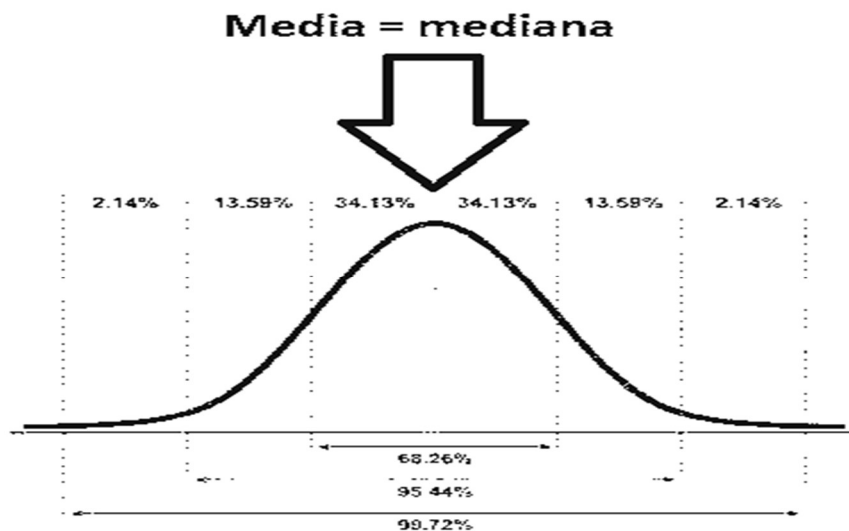
Hasta aquí hemos presentado un conjunto de herramientas para describir la distribución de los datos: tablas de frecuencias, histogramas, cálculo de medidas resumen (media, mediana, desvío estándar, distancia Inter cuartil, percentiles) y el gráfico de caja y brazos. Algunas veces estas herramientas tienen inconvenientes:

- las tablas de frecuencias, así como sus representaciones gráficas (los histogramas), eliminan los detalles y dependen de la longitud de los intervalos de clase.
- las medidas resumen (media, mediana, desvío estándar, distancia Inter cuartil, percentiles) muestran aspectos parciales de los datos.

Ante esta diversidad surge el siguiente interrogante: ¿Es posible describir la distribución de los datos en forma completa mediante una única expresión?

Una primera respuesta sería que depende. Es decir que la respuesta podría ser sí si estamos dispuestos a describir el patrón general de los datos omitiendo los atípicos. Esa respuesta para la distribución de los datos la provee la **curva de densidad**. Las curvas de densidad son histogramas idealizados. Las medidas de posición y dispersión se aplican tanto a curvas de densidad como a conjuntos de datos.

Comencemos por los percentiles previamente presentados. Sabemos que una proporción “*p*” de observaciones está por debajo del percentil *k*. En una curva de densidad simétrica es fácil visualizar donde se encuentra la **mediana**, el punto que divide al área en dos partes iguales o **percentil 50**. Para una curva simétrica, la media, el punto de equilibrio coincide con la mediana que divide el área en dos partes iguales.



Como parte de la idealización inherente a un modelo matemático, las curvas de densidad simétricas son “*perfectamente simétricas*” aunque los datos reales rara vez presenten una simetría perfecta.

Para cualquier otra curva general no siempre es fácil visualizar a simple vista la mediana, la media y los percentiles, pero es posible utilizar integrales para obtenerlos.

Las integrales son herramientas de análisis matemático que permiten obtener el área por debajo de una curva cuando se conoce la expresión de la misma.

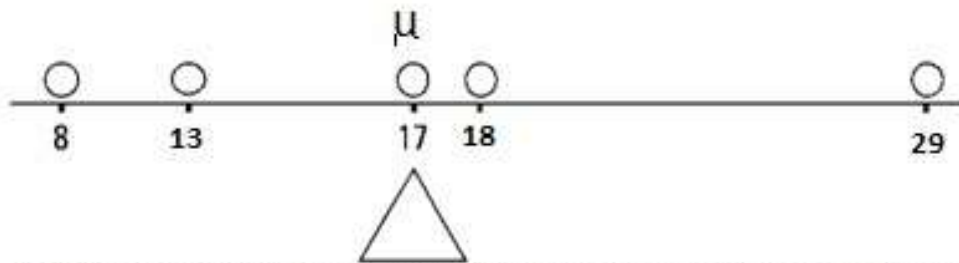
Si contamos con 5 observaciones con los siguientes valores:

Observación	1	2	3	4	5
Valor	8	13	17	18	29

La media será:

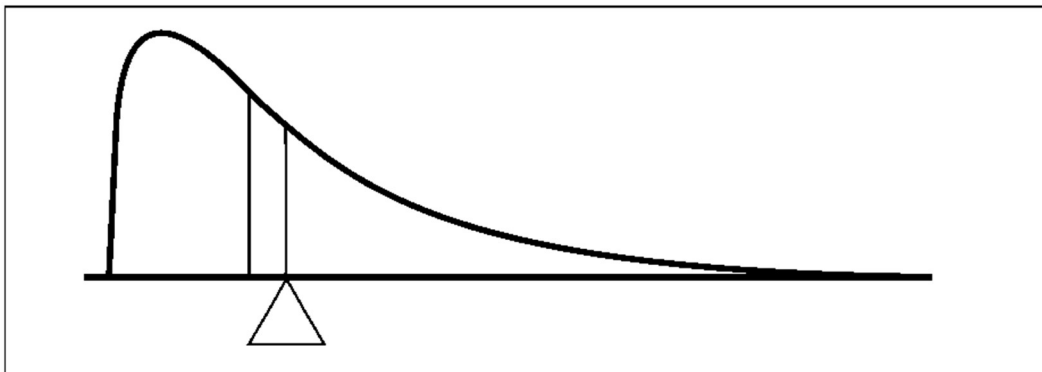
$$\mu = \Sigma / 5 = (13 + 17 + 8 + 29 + 18) / 5 = 85 / 5 = 17$$

La media es el punto de equilibrio de una vara sin peso sobre la que se colocan en cada punto correspondiente al valor de cada dato, pesos idénticos.



El triángulo indica el punto en el cual la distribución está en equilibrio

En una curva asimétrica la media (el punto de equilibrio) es arrastrado hacia la cola larga de la distribución más que la mediana.



Hallar a ojo la media en una curva asimétrica es más difícil que la mediana, pero la podemos obtener mediante integrales. Aunque no haremos aquí el cálculo de integrales, podemos decir que la media es el punto donde la curva de densidad quedaría en equilibrio. Pasando en limpio:

- La mediana es el punto que divide el área bajo la curva en dos partes iguales.
- La media es el punto de equilibrio o centro de gravedad, sobre el cual quedaría en equilibrio si se construyera con un material sólido.

Para calcular la mediana a ojo tratamos de dividir el área en dos partes iguales. En cambio, para hallar los cuartiles, tratamos de dividir el área por debajo de la curva de densidad en 4 partes iguales. La distancia Inter cuartil es la diferencia entre el cuartil superior y el inferior (también llamados tercer y primer cuartil). Los cuartiles, por lo tanto, la mediana y la distancia Inter cuartil, pueden calcularse en forma aproximada a ojo para cualquier curva de densidad.

Esto no ocurre con el desvío estándar, que no es una medida natural para la mayoría de las distribuciones. Cuando es necesario, el desvío estándar correspondiente a una curva de densidad, también (como dijimos para los percentiles), puede calcularse utilizando integrales.

¿Para qué sirve tener un conjunto de datos cuyo histograma es aproximadamente Normal? Cuando los datos tienen una distribución Normal, su distribución se puede reducir a dos números:

- la media y
- el desvío

Cuando tenemos que analizar muchos datos, a medida que se achica la longitud de los intervalos de clase mejora la aproximación a la forma de campana simétrica de Gauss. Cuando un histograma se grafica utilizando las frecuencias en el eje vertical, la escala depende de la cantidad de datos. Si se utilizan frecuencias relativas o porcentajes esto es menos arbitrario y el área del rectángulo es proporcional a la frecuencia relativa. Es más natural que el área del rectángulo sea igual a la frecuencia relativa. Lo logramos si en el eje vertical graficamos la frecuencia relativa dividida la longitud

del intervalo. Esto se llama **escala de densidad** y permite tener la misma escala vertical, aunque cambiemos la longitud de los intervalos y el área total de los rectángulos del histograma siempre será 1 (o 100 si las frecuencias relativas están expresadas como porcentajes).

La curva que describe la forma de la distribución se llama **curva de densidad** y tiene área igual a 1. El área bajo la curva sobre cualquier intervalo de valores del eje horizontal es la proporción de observaciones que caen en ese intervalo.

Otra propiedad importante de la curva de densidad Normal es poder localizar el desvío estándar a ojo: a medida que nos movemos en ambas direcciones desde el centro μ de la curva, ésta aumenta su pendiente hasta un punto de inflexión, a partir del cual la pendiente empieza a disminuir. Los dos puntos en los cuales ocurre este cambio de curvatura están localizados a una distancia σ a cada lado del centro μ .

Recuerde que σ es el desvío y μ es la media. Tanto μ como σ por sí solos no determinan la forma de una distribución en general.

Sin embargo, cuando los valores de una variable tienen una distribución Normal, los valores de μ y σ son los parámetros de la distribución y con ambos números se puede determinar la distribución de todos los valores.

Teorema del Límite Central (TLC)

Algunos modelos de probabilidad pueden ser entendidos como sumas de variables aleatorias. En esos casos, el **Teorema del Límite Central** explica que cuando la suma es suficientemente grande, dicho modelo tiende a comportarse como una Normal. El Teorema del Límite Central es una propiedad muy importante en la estadística que permite resolver diversos problemas claves.

Enunciado del Teorema

Se tienen n variables aleatorias, independientes entre sí, todas con la misma distribución de probabilidad (a la cual se identifica con la letra F), con media μ y varianza σ^2 .

Si las variables son X_1, X_2, \dots, X_n , entonces se tiene que: $X_i \sim F(\mu, \sigma^2)$.

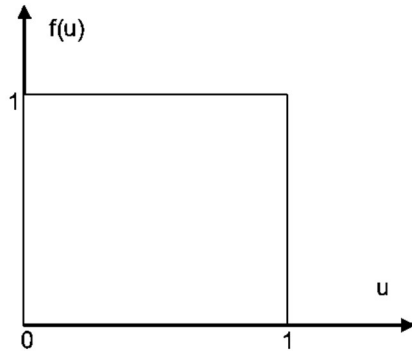
Si se suman dichas variables del siguiente modo: $\sum = X_1 + X_2 + \dots + X_n$

Cuando n tiende a infinito, la suma \sum tiende a comportarse como una Normal con media igual a la suma de las medias y varianza igual a la suma de las varianzas. Esto es: $\sum \rightarrow N(n\mu, n\sigma^2)$.

Lo interesante es que el **Teorema del Límite Central** se verifica aceptablemente aun cuando no se cumplan estrictamente las condiciones de partida. Esto es, la normalidad puede aparecer aun cuando la cantidad de términos que se suman no sea muy grande e incluso, cuando las variables sumandos no tienen estrictamente las mismas distribuciones de probabilidad. Volveremos a la suma de múltiples variables hacia el final cuando tratemos la proyección con base a la regresión lineal de diversas variables que inciden en la recaudación tributaria.

Ejemplos de Sumas de Variables

Con ayuda de una planilla de cálculo podemos simular distintas sumas de variables aleatorias. Por ejemplo, un caso típico es el de las variables rectangulares o uniformes, en el intervalo de cero a uno. Para una definición formal se debe especificar que una variable aleatoria U tiene esta distribución cuando su función de densidad es de la siguiente forma:



La media y la varianza de esta variable U, son las siguientes:

$$\mu (U) = 0,5; V (U) = 1/12$$

Ahora bien, si se realiza un experimento que consiste en sumar doce valores de esta variable U, la suma resultante debería tener distribución Normal (conforme a lo que plantea el teorema), con las siguientes media y varianza:

$$\mu (S) = 12 * 0,5 = 6 \quad V(S) = 12 * 1/12 = 1$$

Los programas de planilla de cálculo permiten generar valores de una variable como esta, con la función ALEATORIO. Con este recurso, es posible entonces construir una “muestra” de posibles resultados. Muchas variables pueden concebirse como suma de efectos. Cuando esto ocurre, conforme al Teorema del Límite Central, la variable debe tener distribución Normal.

Si analizamos, por ejemplo, el caso de la variable “comportamiento fiscal”, veremos que la misma es consecuencia de una gran cantidad de efectos como cumplimiento de la presentación de declaraciones, mora en el pago, litigios contra el Fisco, ajustes intimados en inspecciones e incluso el sector económico donde se desempeña el contribuyente. En esas condiciones, la variable sólo puede ser Normal.

Pero no es solamente una característica del comportamiento de los contribuyentes, si se habla de otro tipo de variables, la aproximación se mantiene. Por ejemplo, las ventas diarias de un contribuyente en un “Shopping center” (mall o paseo de compras), pueden tener cualquier distribución. Pero las ventas diarias de TODOS los contribuyentes del complejo comercial, deben tender a comportarse como normales.

Fuentes de Datos

Para producir buenas proyecciones es esencial tener acceso a datos de la mejor calidad posible. Sin datos confiables, no podemos usar técnicas analíticas para proyectar ingresos, y nuestra mejor intuición es nuestra única suposición.

El tipo de datos utilizados en la previsión de ingresos se conoce como datos de series temporales. Esto significa que nuestros datos contienen observaciones para una variable específica, como los ingresos fiscales, el PIB o un precio, durante varios períodos de tiempo.

Hay varias fuentes confiables disponibles para obtener datos macroeconómicos a nivel de país o entre países. Además de los datos aportados por la misma Administración Tributaria³, los datos de series temporales pueden surgir de fuentes oficiales como la Dirección Nacional de Investigaciones y Análisis Fiscal (DNI AF)⁴, el Instituto Nacional de Estadística y Censos (INDEC)⁵ o el Banco Central (BCRA)⁶, pero también de organismos internacionales como la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD)⁷, el Fondo Monetario Internacional (FMI)⁸, el Banco Mundial⁹ o la Organización para la Cooperación y el Desarrollo Económico (OCDE)¹⁰.

³ <https://www.afip.gob.ar/institucional/estudios/>

⁴ <https://www.argentina.gob.ar/economia/ingresospublicos/dniaf>

⁵ <https://www.indec.gob.ar/>

⁶ https://www.bcra.gob.ar/PublicacionesEstadisticas/Principales_variables.asp

⁷ <http://unctadstat.unctad.org/EN/Index.html>

⁸ <http://data.imf.org/revenues>

⁹ <http://datatopics.worldbank.org/world-development-indicators>

¹⁰ <https://data.oecd.org>

Inspección de los Datos

Una vez que se seleccionó la serie temporal que se analizará, el siguiente paso es inspeccionar visualmente sus datos. La visualización de datos ayuda a evaluar las características clave de los datos y advertir tempranamente problemas que podrían complicar su análisis posterior. La inspección de datos permite detectar problemas potenciales antes de comenzar a aplicar técnicas estadísticas más sofisticadas.

Visualización de Datos en Series Temporales

Al analizar series temporales el primer paso es inspeccionar el comportamiento de los datos para evaluar las características clave y determinar el enfoque de proyección a utilizar.

El primer golpe de vista nos permite visualizar aumentos o disminuciones, pero también podremos detectar regularidades o patrones en los datos (tendencias o ciclos).

Otra característica que se puede apreciar es la volatilidad de los datos. Es decir, cuánto varía una serie en el tiempo. Por ejemplo, los ingresos de un impuesto directo sobre la renta de las empresas (impuesto a las Ganancias de personas jurídicas) son generalmente más volátiles que los recursos provenientes de impuestos indirectos relacionados con el consumo como el IVA.

Otra cuestión a advertir es la presencia que cualquier anomalía que pueda estar originada en valores atípicos (extremos y posiblemente inusuales o incorrectos como los outliers previamente presentados).

En ciertas ocasiones, la falta de datos en la serie puede manifestarse como lagunas. Aunque también podríamos encontrarnos con un **quiebre estructural** originado en una eventualidad que haya

causado un cambio en la serie. Por ejemplo, un cambio normativo que modifique el hecho o la base imponible. El omitir estas señales puede llevarnos a proyecciones erróneas.

También puede resultarnos de utilidad trazar varias series de tiempo superpuestas para comprender la correlación entre ellas (más adelante presentaremos la superposición de las series de tiempo de la “*presión tributaria*” y del “*producto interno bruto*” para intentar entender su correlación). La inspección conjunta de varias series temporales al mismo tiempo nos permite comprender si otras variables posiblemente se están moviendo junto con la serie bajo análisis. La correlación entre dos series puede resultarnos de utilidad para generar proyecciones.

En síntesis, la visualización, como primer paso, tiene una utilidad orientada a advertir tendencias, volatilidad, valores atípicos, lagunas, quiebres estructurales y correlaciones.

Por lo tanto, los gráficos de series temporales nos permitirán visualizar las suposiciones preliminares sobre tendencias y patrones en los datos, todo lo cual nos ayudará a precisar el posterior proceso de proyección y la elección del modelo a utilizar más adecuado.

Diagnóstico Básico de Datos

Después de graficar los datos, el próximo paso es utilizarlos para producir algunas estadísticas descriptivas básicas. Antes de hacerlo, es útil repasar los cuatro componentes de una serie temporal. Estos no solamente son importantes para tener en cuenta al visualizar sus datos, sino que también ayudarán a determinar qué técnica de proyección podría ser la más útil para aplicar.

Cada serie de tiempo se puede descomponer en cuatro componentes:

- **Tendencia:** es la tendencia a largo plazo de una serie temporal.
- **Variaciones estacionales:** patrones repetidos que ocurren todos los años en ciertas épocas de cada año (por ejemplo, un pico en el consumo de los hogares al cobrar el

medio aguinaldo o un incremento en las liquidaciones relacionadas con la cosecha gruesa).

- **Variaciones cíclicas:** los ciclos son fluctuaciones periódicas en períodos superiores a un año. Por ejemplo, un ciclo económico, el cual suele durar varios años.
- **Variaciones irregulares:** son fluctuaciones resultantes de choques aleatorios e intrínsecos. Tales fluctuaciones a corto plazo no son ni sistemáticas ni predecibles.

Teniendo presentes estos cuatro componentes de series temporales, podemos avanzar con el cálculo de estadísticas descriptivas, como la media, la mediana y la desviación estándar, las cuales, como hemos adelantado, nos proveen información de resumen que nos resultará de utilidad para comprender la distribución de los datos y su dispersión. Como recordarán, la media y la mediana nos resultarán de utilidad para reconocer un valor de datos "**central**" de la serie (formalmente se conocen como medidas de "**tendencia central**"). Repasemos conceptos:

Media

Para una variable de serie de tiempo dada, X_t , la **media**, μ , se calcula de la siguiente manera:

$$\mu = \frac{1}{n} \sum_{t=1}^n X_t = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Desvío Estándar

El **desvío estándar**, σ , es una medida de la dispersión de datos alrededor de la media (μ) y se calcula de la siguiente manera:

$$\sigma = \sqrt{\frac{1}{n} \sum_{t=1}^n (X_t - \mu)^2} = \sqrt{\frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \dots + (X_n - \mu)^2}{n}}$$

Coefficiente de Variación

Una medida de dispersión alternativa pero relacionada es el **coeficiente de variación** (CV):

$$CV = \frac{\sigma}{\mu}$$

El **coeficiente de variación** (CV) es el resultado del cociente del desvío estándar dividido por la media. La normalización de la desviación estándar por la media permite realizar comparaciones entre series.

Por regla general, si el valor absoluto del **coeficiente de variación** es mayor que 1 (es decir, independientemente de que el signo sea positivo o negativo), se considera que la serie presenta una alta variación. Por el contrario, si el **coeficiente de variación** es menor que 1 en términos absolutos, la serie presenta poca variación.

Detección de Quiebres Estructurales

Un **quiebre estructural** ocurre cuando el comportamiento de una serie cambia abruptamente en un momento determinado. Cuando analizamos datos de series temporales debemos prestar especial atención a los posibles quiebres estructurales.

Dichos quiebres se derivan generalmente de shocks originados en aumento del precio de las materias primas o la energía, conflictos internacionales que alteran la cadena de suministros, cambios políticos o modificaciones normativas. A los fines de realizar una buena proyección se deben tener en cuenta dichos cambios estructurales (GONZÁLEZ CAO, Previsión de impuestos, 2023).

En el análisis de proyección, uno de los supuestos clave es que los parámetros, como la media y la desviación estándar, son relativamente estables a lo largo del tiempo. Sin embargo, los cambios estructurales generalmente distorsionan estos estadísticos. Como resultado, es posible que los parámetros ya no caractericen con precisión las propiedades de la serie y que los métodos utilizados produzcan estimaciones sesgadas o inconsistentes y, como resultado, malas proyecciones (GONZÁLEZ CAO, Previsión de impuestos, 2023).

Cuando se trabaja con datos macroeconómicos, la detección de posibles quiebres estructurales es importante porque se supone que las características estructurales de una economía son estables en el tiempo. Por lo tanto, si un shock genera un cambio que perdurará, deberemos corregir los modelos estadísticos para que las proyecciones sigan siendo confiables y no induzcan a interpretaciones erróneas.

Para detectar este tipo de quiebres debemos comenzar con la visualización y examen de los datos como primer paso. Al graficar los datos debemos estar atentos a los cambios repentinos en el comportamiento de una serie. Este análisis visual debe estar respaldado por un conocimiento general del entorno macroeconómico pasado:

- ¿Pasó el país por una crisis financiera?
- ¿Fue golpeado por un desastre natural, una sequía extraordinaria o una inundación de grandes dimensiones?

- ¿Osciló repentinamente el precio de sus exportaciones o importaciones más importantes como la energía o los cereales?

Además de este tipo de impactos también se debe prestar atención a los cambios en la política fiscal o a las reformas de la administración tributaria.

Consideremos un posible ejemplo. Se proyecta la evolución de los ingresos por impuestos sobre la renta de las personas (impuesto a las Ganancias de las personas humanas). Al empezar los cálculos se advierte que el Congreso está pronto a sancionar una reforma legislativa que incrementa el piso mínimo a partir del cual los empleados en relación de dependencia, los jubilados y pensionados resultarán obligados a realizar el ingreso del impuesto a las ganancias de la cuarta categoría. Esto genera un quiebre estructural. Lo mismo si se lo convierte en un impuesto cedular para los altos ingresos personales.

Esta novedad no puede ser obviada porque de no estimar y extrapolar el impacto de dichos cambios, los resultados de la proyección basados en los antecedentes de comportamiento del tributo con la legislación previa brindarán una perspectiva engañosa para los ingresos futuros cuando se apruebe la reforma normativa. Ignorar este cambio estructural originado en el cambio normativo nos conducirá a previsiones de ingresos erradas.

En algunos casos, en los **modelos incondicionales** básicos, podemos abordar la inestabilidad estructural desechando los datos asociados con el quiebre estructural y extrapolar la tendencia sobre la muestra restante. En las regresiones lineales, se suele incluir variables ficticias para eventos excepcionales o cambios de política que puedan causar un quiebre en los datos. Estas artimañas permiten preservar la integridad de los datos mientras se capturan los efectos de la intervención.

La visualización de datos, reforzada por el conocimiento general del contexto macroeconómico del país, puede ser una herramienta fundamental para detectar tales quiebres.

Descripción General de las Técnicas de Proyección de Ingresos

Como explica Fernando Martín los modelos de proyección parten del supuesto de que el valor de una variable depende de ciertos factores determinantes a los cuales se los denomina **variables explicativas** (MARTIN, El Método Directo de Proyección de los Ingresos Tributarios, 2004).

Matemáticamente podemos presentar el comportamiento de dichos modelos así:

$$X_t = f(a, b, c, \dots z)_{t, t-n}$$

Donde el valor de la variable X en el período t está determinado por los valores de los factores a, b, c, z, en el mismo o en anteriores períodos de tiempo.

A partir de dicha definición podemos emprender dos caminos:

- Recurrir a los registros estadísticos de la recaudación y de sus variables explicativas, buscando definir las relaciones entre ambas.
- Encarar la proyección de las variables explicativas.

En el segundo caso, la proyección surgirá de la siguiente ecuación:

$$X_{t+1} = X_t + \text{var}(a, b, c, \dots z)$$

En el modelo proyectivo señalado la recaudación se comportará como una variable dependiente (X). Por lo tanto su valor en el período t+1 será una función del valor de la recaudación en el período previo (t) y de las modificaciones de sus variables explicativas (MARTIN, El Método Directo de Proyección de los Ingresos Tributarios, 2004).

Habiendo presentado el concepto general de proyección de ingresos, el paso siguiente es presentar algunas técnicas estándar de proyección. Para ello diferenciaremos entre:

- **Modelos incondicionales** (análisis de tendencias, extrapolación, medias móviles): Los modelos incondicionales utilizan datos de ingresos fiscales pasados para predecir los ingresos fiscales futuros. La ventaja de estas técnicas es que son relativamente rápidas y fáciles de implementar ya que no dependen del análisis de relaciones entre la variable de interés (ingresos fiscales) y otros indicadores económicos.
- **Modelos condicionales** (modelo de tasa impositiva efectiva, regresión lineal): los modelos condicionales son generalmente más complejos. Estos modelos estiman la relación entre los ingresos fiscales y otras variables económicas relevantes relacionadas y utilizan los parámetros estimados de esta relación para pronosticar los ingresos fiscales. Como son condicionales dependen de variables distintas a los ingresos fiscales.

Modelos Incondicionales de Proyección Tributaria

Los **modelos incondicionales** se enfocan en el componente de tendencia de los datos de series temporales. Esto se debe a que establecer tendencias es un primer paso importante para proyectar los ingresos. La idea principal es utilizar los datos de los ingresos fiscales pasados para predecir los ingresos fiscales futuros. La ventaja de estas técnicas es que no dependen del análisis de relaciones entre la variable de interés (ingresos fiscales) y otros indicadores económicos. Parte de la tarea del pronosticador es evaluar adecuadamente qué forma de tendencia pueden exhibir los datos.

En general, muchas de las herramientas utilizadas en la pronóstico funcionan mejor cuando la tendencia muestra cierto grado de linealidad. Si bien esto puede ser cierto en ciertos casos, otros casos pueden requerir la transformación de los datos para garantizar dicha linealidad. Por ejemplo, una serie de tiempo podría exhibir una tendencia exponencial que necesitaría ser "*linealizada*" usando una transformación logarítmica, antes de ser analizada usando técnicas estadísticas estándar (dicha transformación logarítmica puede ser realizada con herramientas de planillas de cálculo).

Además, como se mencionó, se deben tener en cuenta los cambios estructurales que pueden hacer que las tendencias varíen con el tiempo. La idea detrás de los modelos incondicionales (tendencia)

es extender una serie de tiempo hacia el futuro utilizando la información recopilada sobre esa serie a partir de observaciones históricas. Si se pueden identificar tendencias y patrones anteriores subyacentes, éstos se pueden usar para extrapolar la serie temporal hacia adelante para generar proyecciones. Implícitamente, encontrar tales tendencias implica suavizar los datos para limitar los efectos de las variaciones aleatorias.

Este tipo de análisis de **tendencia incondicional** es adecuado para los casos en los cuales tanto los ingresos fiscales como las leyes tributarias son relativamente estables. Por ejemplo, es útil para calcular en el presupuesto anual los demás ingresos públicos no tributarios:

- Rentas de la propiedad
- Transferencias de capital
- Transferencias corrientes
- Recursos propios de capital

Proyección Aditiva en Función de los Ingresos Pasados

En el análisis de tendencias incondicionales la proyección de ingresos fiscales (T_{t+1}^f) es una función de las observaciones pasadas de ingresos fiscales (T_{t-s}) donde $n \leq s \leq t$. Matemáticamente, la proyección se expresa de la siguiente manera:

$$T_{t+1}^f = F(T_t, T_{t-1}, \dots, T_{t-n})$$

En los modelos de tendencia aditivos, los ingresos previstos equivalen a los ingresos del año pasado aumentados por la variación de ingresos del último período. En su forma más simple, la variación de la recaudación de impuestos del último período simplemente es igual a la diferencia entre T_t y T_{t-1} . Matemáticamente:

$$T_{t+1}^f = T_t + (T_t - T_{t-1})$$

Media Móvil de los Ingresos Pasados

La referida proyección aditiva también se puede expresar como la variación promedio sobre un período de tiempo dado. El uso de una **media móvil** minimizará el impacto de los elementos aleatorios en el pasado y naturalmente suavizará los datos:

$$T_{t+1}^f = T_t + \frac{(T_t - T_{t-1}) + (T_{t-1} - T_{t-2}) + \dots + (T_{t-(n-1)} - T_{t-n})}{n}$$

Donde n es el número de años que se promedian.

Este tipo de promedio se denomina **promedio móvil** porque el subconjunto sobre el cual se promedian los datos cambia junto con las proyecciones. En este escenario, en $t+1$, la proyección usa el promedio de la serie relevante desde $t-n$ hasta t . En $t+2$, la proyección usará el promedio de $t-n+1$ a $t+1$, y así sucesivamente.

Ponderaciones Diferenciadas de los Ingresos Pasados

También pueden usarse **ponderaciones diferenciadas** para aumentar la importancia de las observaciones más recientes.

Al aplicar ponderaciones mayores a períodos más recientes, se asume que las observaciones recientes tienen más poder predictivo para los ingresos futuros. En otras palabras, es probable que los

puntos de datos más recientes en los ingresos fiscales tengan más probabilidades de afectar los ingresos futuros.

Todas estas variantes de modelos de tendencia aditiva son relativamente fáciles de implementar, incluso con datos limitados. Sin embargo, asumen implícitamente que la diferencia en los ingresos fiscales de un período a otro es algo estable en el tiempo, lo cual puede no ser así.

Proyección en Función de Tendencias Multiplicativas

Como las diferencias pasadas son menos confiables y menos informativas que las tasas de crecimiento se usan con mucha más frecuencia los modelos de **tendencias multiplicativas**.

Los modelos de tendencia multiplicativa asumen que la tasa de crecimiento de un período al siguiente es relativamente estable en el tiempo. De esta manera, los valores pasados del crecimiento de los ingresos fiscales se consideran un predictor fiable del crecimiento futuro de los ingresos fiscales y, por lo tanto, pueden utilizarse para pronosticar los niveles futuros de ingresos fiscales:

$$\frac{T_{t+1}^f}{T_t} - 1 = \frac{T_t}{T_{t-1}} - 1$$
$$\Rightarrow T_{t+1}^f = T_t \times \frac{T_t}{T_{t-1}}$$

Proyección en Función de Promedio Móvil de las Tasas de Crecimiento

También se puede generar una proyección del crecimiento futuro de los ingresos fiscales utilizando el promedio de las tasas de crecimiento pasadas durante un período de tiempo determinado. Por ejemplo, un promedio móvil crea una serie de promedios durante períodos de tiempo específicos y es una forma práctica de suavizar los datos:

$$T_{t+1}^f = T_t \times \left[\frac{\frac{T_t}{T_{t-1}} + \frac{T_{t-1}}{T_{t-2}} + \dots + \frac{T_{t-(n-1)}}{T_{t-n}}}{n} \right]$$

Donde n es el número de años durante los cuales se promedian las tasas de crecimiento pasadas.

Esta fórmula utiliza los promedios no ponderados de las tasas de crecimiento pasadas. Sin embargo, se pueden aplicar ponderaciones a las tasas de crecimiento pasadas para aumentar la importancia del crecimiento de los ingresos tributarios experimentado en ciertos períodos de tiempo (por ejemplo, el pasado más reciente) sobre otros.

Los modelos de tendencias multiplicativas se pueden usar con relativa facilidad y rapidez para generar diferentes conjuntos de proyecciones bajo diferentes supuestos, incluso cuando hay datos limitados.

Evaluación de Errores de Proyección

Proyectar el comportamiento de la economía es complejo en el actual contexto de volatilidad e incertidumbre. Por consiguiente, esta complejidad en la predicción debe intentar comprender un amplio abanico de micro y macro factores que pueden afectar los ingresos fiscales.

Entre los macro factores podemos enumerar el crecimiento del producto interno bruto (PBI), la tasa de desocupación, la proporción de empleo asalariado informal, la inflación, los tipos de cambio o los vínculos macro fiscales.

Pero también hay micro factores como el grado de cumplimiento de los contribuyentes.

Sumado a todo ello, también juegan un rol importante los impactos externos, los desastres naturales y las crisis financieras internacionales.

A su vez, también son importantes ciertos factores institucionales como la solidez del sistema legal.

Finalmente, se deben considerar adicionalmente ciertos factores metodológicos como las restricciones de datos, las limitaciones de las técnicas de modelado estadístico y la necesidad de corregir el posible sesgo de estimación.

Como vamos viendo es increíblemente complejo superar todos estos condicionantes y producir proyecciones de ingresos precisos, razón por la cual se debe intentar minimizar los errores de proyección por medio de la elección del modelo de proyección más adecuado. Una buena práctica para ello es utilizar diferentes modelos, así como diferentes especificaciones del mismo modelo para generar varias proyecciones de ingresos.

Para comparar la calidad de estas diferentes proyecciones, tenemos a nuestra disposición varias estadísticas de evaluación de proyecciones que se pueden usar.

Ex ante se puede evaluar un modelo con una fracción acotada de los datos disponibles (muestra), usando el resto para evaluar el desempeño de la proyección fuera de la muestra. Por supuesto, también se puede evaluar un modelo ex post. Para ello, una vez que los resultados reales son definitivos, se pueden comparar con las previsiones previas utilizadas en el proceso de elaboración de la proyección. Para evaluar el desempeño de la proyección se pueden usar diferentes estadísticos.

Errores de Proyección (EP)

Se pueden rastrear los errores de proyección a medida que ingresan los datos de ingresos reales. Estos errores se pueden usar para producir estadísticas de evaluación de proyecciones, que

pueden ayudar a mejorar el modelo y la metodología subyacentes. Los errores de proyección, EP_t , se definen simplemente como la diferencia entre el valor de proyección de una variable, Y_t^f , y el valor real de esa variable, Y_t . Matemáticamente:

$$EP_t = Y_t^f - Y_t$$

EP_t puede definirse como la diferencia entre los ingresos fiscales previstos, Y_t^f , y la recaudación real, Y_t .

Error Medio de Proyección (EMP)

El sesgo es el error de proyección promedio durante un período de tiempo determinado. Se calcula sumando los errores de proyección y dividiéndolos por el número de observaciones. Idealmente, la medida del sesgo es cero, lo que significa que las proyecciones son imparciales. En otras palabras, las previsiones de ingresos no son sistemáticamente optimistas (positivas) ni pesimistas (negativas). Matemáticamente:

$$EMP = \frac{1}{n} \sum_{t=1}^n EP_t = \frac{EP_1 + EP_2 + \dots + EP_n}{n}$$

Por ejemplo, si el sesgo es -\$100, la proyección de ingresos puede considerarse pesimista. Es decir, la recaudación real supera las previsiones de ingresos en \$100 en promedio.

Error Absoluto Medio (EAM)

Otra medida de precisión es el error absoluto medio (EAM):

$$\text{EAM} = \frac{1}{n} \sum_{t=1}^n |EP_t| = \frac{|EP_1| + |EP_2| + \dots + |EP_n|}{n}$$

En la ecuación la expresión “|...|” denota valor absoluto. Mientras que el error medio de proyección (EMP) es una medida del sesgo del modelo de proyección, el error absoluto medio (EAM) indica el tamaño absoluto de los errores de proyección en promedio.

Error Porcentual Medio (EPM)

A menudo es más práctico expresar el error de proyección como un porcentaje de la recaudación real, en lugar de niveles. Entonces podemos calcular el error porcentual medio (EPM):

$$\text{EPM} = \frac{1}{n} \sum_{t=1}^n \frac{E_t}{T_t} = \frac{\frac{EP_1}{T_1} + \frac{EP_2}{T_2} + \dots + \frac{EP_n}{T_n}}{n}$$

Por ejemplo, si el MPE es igual a -0,05, el modelo genera proyecciones que tienden a ser pesimista: la serie de proyección es, en promedio, un 5 por ciento más baja que la serie real.

Error Porcentual Absoluto Medio (EPAM)

Tal desviación porcentual entre los errores de proyección y los valores reales también se puede expresar en términos absolutos utilizando el error porcentual absoluto medio (EPAM):

$$EPAM = \frac{1}{n} \sum_{t=1}^n \frac{|EP_t|}{T_t} = \frac{\frac{|EP_1|}{T_1} + \frac{|EP_2|}{T_2} + \dots + \frac{|EP_n|}{T_n}}{n}$$

El error porcentual absoluto medio (EPAM) indica la desviación absoluta promedio (en términos porcentuales) de la serie real. Por ejemplo, si el error porcentual absoluto medio (EPAM) es igual a 0,08, el error de proyección promedio, ya sea positivo o negativo, es del 8 por ciento. En promedio, el modelo falla su predicción en un 8 por ciento. Por construcción, el Error Porcentual Absoluto Medio (EPAM) será necesariamente mayor o igual que el Error Porcentual Medio (EPM) porque promedia el valor absoluto de los errores de proyección. En cierto sentido, el Error Porcentual Absoluto Medio (EPAM) refleja el error de proyección promedio, ya sea positivo o negativo, mientras que el Error Porcentual Medio (EPM) refleja si los errores de proyección están sesgados.

Error Estándar del Proyección (EEP)

Idealmente, la variación de los errores de proyección debería ser pequeña, lo que significa que la volatilidad de los errores de proyección debería ser pequeña. Para entender esto, podemos calcular la dispersión de los errores de proyección usando el error estándar de la proyección (EEP):

$$EEP = \sqrt{\frac{1}{n} \sum_{t=1}^n (EP_t - EMP)^2} = \frac{(EP_1 - EMP)^2 + (EP_2 - EMP)^2 + \dots + (EP_n - EMP)^2}{n}$$

Error Cuadrático Medio (ECM)

El Error Cuadrático se utiliza para calcular el siguiente estadístico que se usa para evaluar las proyecciones, el Error Cuadrático Medio (ECM):

$$\text{ECM} = \sqrt{\text{EEP}^2 + \text{EMP}^2} = \sqrt{\frac{1}{n} \sum_{t=1}^n \text{EP}_t^2} = \sqrt{\frac{\text{EP}_1^2 + \text{EP}_2^2 + \dots + \text{EP}_n^2}{n}}$$

El Error Cuadrático Medio (ECM) nunca puede ser negativo porque es el resultado de una raíz cuadrada. A su vez, un valor de 0 indica un ajuste perfecto a los datos.

En general, un error cuadrático medio (ECM) más bajo es mejor que uno más alto. Sin embargo, las comparaciones entre diferentes tipos de datos no serían válidas porque la medida depende de la escala de los números utilizados.

Además, como muestra la fórmula, el Error Cuadrático Medio (ECM) es la raíz cuadrada del promedio de los errores de proyección al cuadrado. El efecto de cada error de proyección en el Error Cuadrático Medio (ECM) es, por lo tanto, proporcional al tamaño del Error Cuadrático. Por lo tanto, los errores más grandes tienen un efecto desproporcionadamente grande en el Error Cuadrático Medio (ECM), lo que hace que esta medida sea sensible a los valores atípicos en la serie temporal.

Modelos Condicionales de Proyección Tributaria

Los **modelos condicionales** son otra categoría de modelos de proyección. Los modelos condicionales están diseñados para determinar si los ingresos fiscales están asociados con otros tipos de variables, por ejemplo, medidas de la base imponible u otros elementos de la economía, y luego usan estas relaciones para generar proyecciones de ingresos fiscales.

De esta forma, las previsiones de ingresos quedan condicionadas a las proyecciones de otros factores económicos, como el producto bruto interno (PBI), las importaciones o el consumo (más adelante presentaremos una tabla que detalla las variables que inciden sobre los ingresos tributarios).

Para poder avanzar primero vamos a presentar una diferencia entre la tasa impositiva legal (TIL) y la tasa impositiva efectiva (TIE). La tasa impositiva legal (TIL, en inglés STR) es el porcentaje de impuestos que se fija como un alícuota establecida por ley, mientras que la tasa impositiva efectiva (TIE) es la carga impositiva total o promedio, después de calcular todas las tasas impositivas.

La recaudación fiscal generada por el sistema impositivo en su conjunto es la suma de las tasas impositivas legales (TIL) para cada tipo de tributo, multiplicada por su base imponible. Las proyecciones basadas en las tasas impositivas legales (TIL) son complejas porque requieren mucha información.

Por ejemplo, para un impuesto a la renta de personas físicas con régimen progresivo (impuesto a las Ganancias de personas humanas con una tabla progresiva de alícuotas de tributación en función de los ingresos), se necesita tener a su disposición, entre otras cosas, la alícuota asociada con cada tramo de ingresos, datos confiables sobre la distribución del ingreso, un catálogo completo de todas las deducciones y créditos fiscales, amplia información sobre el desglose entre las diferentes categorías de ingresos imponibles -tales como salarios, dividendos, rentas o ganancias de capital-- todos los cuales pueden estar sujetos a diferentes tratamientos fiscales. Como esto puede ser un desafío complejo, el método de **tasa impositiva efectiva (TIE)** representa una alternativa viable.

Método de la tasa impositiva efectiva (TIE)

El método de **tasa impositiva efectiva (TIE)** es una herramienta práctica para pronosticar ingresos de impuestos individuales.

La base imponible indirecta es una variable económica que está estrechamente relacionada con la base imponible real. Si la tasa impositiva efectiva y la base impositiva indirecta se pueden pronosticar correctamente, entonces se pueden pronosticar los ingresos generados por un impuesto determinado.

Este modelo en particular no requiere muchos datos y no requiere un conocimiento completo de la compleja lista de bases, alícuotas y exenciones del sistema tributario. Por lo tanto, los pasos clave para implementar el método de Tasa Impositiva Efectiva serían:

- Seleccionar la base imponible indirecta.
- Recopilar una proyección de la base impositiva indirecta.
- Calcular y pronosticar la tasa impositiva efectiva.
- Usar el modelo de tasa impositiva efectiva para generar una proyección de ingresos.

Paso 1: Seleccionar la Base Imponible Indirecta.

El primer paso consiste en seleccionar la base impositiva indirecta adecuada.

La elección de una base impositiva indirecta adecuada debe basarse en la intuición económica y la información de las leyes fiscales pertinentes. Además, la base imponible indirecta debe ser un indicador económico fuertemente correlacionado con la base imponible real.

Los siguientes son ejemplos de bases imponibles indirectas:

impuestos agregados	PIB
Impuesto a las Ganancias – Renta de Personas Humanas (IRPF)	Renta de los hogares, PBI.
Impuesto a las Ganancias – Renta de Personas Jurídicas (CIT)	Beneficios corporativos, excedente bruto de explotación a partir de datos de cuentas nacionales.
Impuesto al Valor Agregado (IVA)	Consumo de los hogares, PBI.
Impuestos internos especiales	Consumo de bienes relevantes (tabaco, alcohol, combustible), consumo de los hogares, PBI.
Derechos de importación y exportación (derechos aduaneros al comercio exterior)	Importaciones y exportaciones.

Paso 2: Obtener un Proyección de la Base Tributaria Indirecta

Las previsiones de las bases imponibles indirectas y sus componentes suelen estar disponibles en informes que elabora el Ministerio de Economía y también se pueden utilizar los datos de las proyecciones del FMI para el PBI de cada país¹¹.

Paso 3: Calcular y Pronosticar la Tasa Impositiva Efectiva

Utilizando la base impositiva indirecta elegida junto con los datos de recaudación de ingresos se puede calcular la **tasa impositiva efectiva (TIE)**.

Una vez recopilados los datos de la serie histórica se debe analizar si es estable, si muestra una tendencia reconocible y si se detectan quiebres estructurales. La elección del método más adecuado para pronosticar la **tasa impositiva efectiva (TIE)** depende de las conclusiones que resulten del análisis exploratorio de las fuentes de datos.

El catálogo de métodos sobre los cuales elegir el más adecuado no difiere de los señalados previamente como el promedio móvil (ponderado o no ponderado) de las últimas observaciones.

Habiendo calculado el potencial escenario de referencia, se deben tener en cuenta los factores que pueden afectar la **tasa impositiva efectiva (TIE)**:

- **Tasas impositivas legales (TIL)**: Las tasas impositivas legales (TIL) se establecen en la ley. Puede diferir de la tasa impositiva efectiva (TIE) por varios motivos: exenciones, deducciones e incumplimiento.
- **Cumplimiento**: La brecha entre los ingresos reales y los ingresos potenciales generalmente refleja la efectividad de la administración tributaria en la recaudación de impuestos.

¹¹ Perspectivas de la economía mundial: <https://www.imf.org/en/Publications/WEO>

- **Ejecución**: El acto de obligar al cumplimiento de la obligación de pagar impuestos por medio de la cobranza coactiva y la imposición de sanciones.
- **Cambios en la Base Imponible Legislada**: Además de los cambios naturales en la base imponible resultantes de cambios económicos, el Congreso puede sancionar modificaciones legislativas que modifican la bases imponible como la inclusión de nuevas exenciones o la aplicación de alícuotas diferenciadas.

Paso 4: Usar el Modelo de Tasa Impositiva Efectiva (TIE) para Generar una Proyección

Finalmente, conectamos la base imponible y las previsiones de tasa impositiva efectiva (TIE) en el modelo para generar una previsión de ingresos:

$$PI^f = TIE^f \times TII^f$$

En la referida ecuación los términos son:

- PI = proyección de ingresos
- TIE = tasa impositiva efectiva
- TII = tasa impositiva indirecta

Ventajas y Desventajas del Método de la Tasa Impositiva Efectiva (TIE)

La utilización del enfoque de la tasa impositiva efectiva (TIE) tiene ventajas y desventajas. Entre las ventajas podemos señalar que:

- se puede implementar con datos limitados; en particular, no requiere información detallada sobre exenciones, tasas legales, etc.
- La tasa impositiva efectiva (TIE) tiende a ser más estables en el tiempo que las elasticidades.

- Ofrece la posibilidad de incorporar importantes cambios de política; por ejemplo, un recorte en la tasa impositiva estándar debería conducir a una caída comparable en la tasa impositiva efectiva (TIE), suponiendo que todo lo demás sea constante.

Por su parte, sus principales desventajas son que:

- las proyecciones pueden ser sensibles a los supuestos sobre la tasa impositiva efectiva (TIE).
- puede ser un desafío dar cuenta de todos los factores que afectan la tasa impositiva efectiva (TIE) como las tasas impositivas legales (TIL), el nivel de incumplimiento, la aplicación.

Método del Factor de Cobertura

El enfoque del factor de cobertura es una extensión natural del modelo de tasa impositiva efectiva (TIE). Además de la información sobre la base impositiva indirecta, este enfoque incorpora datos sobre las tasas impositivas legales (TIL).

El principal objetivo es diferenciar los efectos de los cambios discretos de política de los de otros factores potenciales, a los que nos referiremos como el **factor de cobertura**.

Naturalmente, este método es una poderosa herramienta de proyección para impuestos con tasas impositivas legales (TIL) fácilmente identificables, como el IVA o los aranceles aduaneros.

Cálculo

Se basa en el método de tasas impositivas efectivas donde:

- T es el ingreso tributario,
- B es el valor de la base imponible o base imponible indirecta.

- La tasa impositiva efectiva (TIE) se calculan por separado sus dos subcomponentes:
 - tasas impositivas legales (TIL), representada por la letra “tau”, que es un componente de política fiscal, y
 - el factor de cobertura, representado por la letra “theta”, que es un componente residual.

Como resultado, los factores que influyen en el factor de cobertura son los mismos que influyen en la tasa impositiva efectiva, excepto las tasas impositivas legales (TIL). Incluyen el grado de cumplimiento de los contribuyentes, el nivel de aplicación y cualquier diferencia entre la base imponible real y la base imponible indirecta, como las diversas exenciones y deducciones que afectan a la base.

Por tanto, el factor de cobertura, como su nombre lo indica, refleja cómo el sistema tributario cubre la base imponible potencial y, por lo general, varía entre cero y uno. Si el factor de cobertura theta es igual a cero, entonces la cobertura es inexistente, ya que no se recaudan ingresos fiscales. Por su parte, si theta es igual a uno, la cobertura es prácticamente perfecta, porque:

- la administración tributaria recauda el 100% de los ingresos potenciales, y
- la base impositiva indirecta y la base impositiva real coinciden perfectamente.

Por ejemplo, si la tasa impositiva es del 10% y las bases impositivas real e indirecta son \$100, un factor de cobertura de uno implica que las autoridades recaudan \$10, lo que implica una cobertura total.

Llegados a este punto es importante aclarar que, eventualmente, es posible tener un factor de cobertura superior a uno. Esto puede suceder con el IVA cuando una administración tributaria retrasa o deniega los reintegros a exportadores.

Como podemos apreciar los pasos a seguir para implementar el enfoque del **factor de cobertura** son muy similares a los observados para el modelo de tasa impositiva efectiva (TIE). La diferencia clave es que el factor de cobertura requiere datos sobre las tasas impositivas legales (TIL).

En consecuencia, se deberá seguir una secuencia de cinco pasos:

- Seleccionar la base imponible indirecta.
- Recopilar una proyección de la base impositiva indirecta.
- Recopilar datos sobre las tasas impositivas legales (TIL) actuales y futuras.
- Calcular y pronosticar el factor de cobertura.
- Usar el modelo de factor de cobertura para generar una proyección de ingresos.

Paso 1: Seleccionar la Base Imponible Indirecta.

impuestos agregados	PIB
Impuesto a las Ganancias – Renta de Personas Jurídicas (CIT)	Beneficios corporativos, excedente bruto de explotación a partir de datos de cuentas nacionales.
Impuesto al Valor Agregado (IVA)	Consumo de los hogares, PBI.
Derechos de importación y exportación (derechos aduaneros al comercio exterior)	Importaciones y exportaciones.

Paso 2: Obtener una Proyección de la Base Tributaria Indirecta

Las previsiones de las bases imponibles indirectas y sus componentes suelen estar disponibles en informes que elabora el Ministerio de Economía y también se pueden utilizar los datos de las proyecciones del FMI para el PBI de cada país.

Paso 3: Recopilar Datos sobre las Tasas Legales Actuales y Futuras

Para implementar este enfoque, se deben tener datos confiables sobre el historial de las tasas impositivas legales (TIL) que permita considerar los cambios de política fiscal que fueron sucediéndose.

Paso 4: Calcular y Pronosticar el Factor de Cobertura

Una vez que se recopilan los datos sobre los ingresos fiscales, la base imponible indirecta y las tasas impositivas legales (TIL), el factor de cobertura, θ_t , se calcula de la siguiente manera:

$$\theta_t = \frac{T_t}{B_t \times \tau_t}$$

Donde:

- T es el ingreso tributario,
- B es el valor de la base imponible o base imponible indirecta
- tau son las tasas impositivas legales (TIL)

El siguiente paso es graficar y analizar la serie del factor de cobertura. El factor de cobertura puede exhibir una tendencia a la baja si la administración tributaria mejora la recaudación de ingresos con el tiempo.

La serie también puede presentar un quiebre estructural si, por ejemplo, la administración tributaria modifica su sistema de recaudación y ello genera ingresos adicionales.

En economías con fuertes capacidades estatales y cierta estabilidad de su economía, los factores de cobertura son relativamente estables y, por lo tanto, es seguro asumir que θ_t permanecerá constante. Si no fuera así la serie podría ser más volátil.

Luego, se puede pronosticar la serie del factor de cobertura utilizando, por ejemplo, técnicas de extrapolación de tendencias incondicionales que explicaremos más adelante, mientras se tienen en cuenta los posibles cambios estructurales en los datos.

Paso 5: Utilizar el Enfoque del Factor de Cobertura para Generar una Proyección

Finalmente, conectamos las proyecciones de la base imponible y del factor de cobertura en el modelo para generar una proyección de ingresos de la siguiente manera:

$$T_{t+1}^f = B_{t+1}^f \times \tau_{t+1} \times \theta_{t+1}^f$$

Ventajas y Desventajas del Método del Factor de Cobertura

La utilización del enfoque de factor de cobertura tiene ventajas y desventajas. Entre las ventajas podemos señalar que:

- se puede implementar con datos limitados,
- los factores de cobertura tienden a ser relativamente predecibles a lo largo del tiempo,
- facilita la incorporación de cambios discrecionales de política en las tasas legales.

Por su parte, podemos señalar como desventajas que:

- las proyecciones pueden ser sensibles a los supuestos sobre el factor de cobertura,
- si bien es sencillo incorporar cambios en las tasas impositivas legales (TIL), sigue siendo complejo pronosticar los efectos de los cambios en las tasas impositivas sobre las otras variables del modelo, en particular la base imponible.

Proyección Mediante Elasticidades

Los modelos basados en la elasticidad se encuentran entre los métodos de proyección más utilizados. Se pueden aplicar para pronosticar ingresos fiscales agregados, así como ingresos de

impuestos específicos (por ejemplo, impuesto sobre la renta de sociedades, impuesto sobre la renta personal, derechos aduaneros e IVA).

Estos métodos se basan en una idea central: la relación entre los ingresos de un impuesto determinado y su base imponible es relativamente estable. Por lo tanto, manteniendo todo lo demás constante, los cambios en los ingresos fiscales reflejan principalmente cambios en la base imponible.

Se puede explotar esta relación para proyectar los ingresos fiscales porque se dispone de proyecciones confiables de la base imponible y las elasticidades se basan en que existe una relación medible entre los ingresos fiscales y su base imponible.

En consecuencia, podemos expresar la elasticidad de los ingresos tributarios con respecto a su base imponible como una ecuación donde épsilon (ϵ) mide el cambio porcentual en los ingresos tributarios (T) ante cambios en la base imponible (B):

$$\epsilon_{T/B} = \frac{\% \Delta T_t}{\% \Delta B_t}$$

Es decir que, si la base imponible aumenta en un 1%, entonces los ingresos fiscales aumentan en un porcentaje épsilon (ϵ). A partir del valor de la elasticidad fiscal podemos definir si un impuesto es progresivo o regresivo. Veámoslo con ejemplos concretos:

- En un **impuesto progresivo** como podría ser un impuesto sobre la renta personal con bandas de tributación y alícuotas progresivas la elasticidad fiscal será mayor que 1 porque los ingresos fiscales crecen más rápido que la base.
- En contrapartida, en un **impuesto regresivo** como podrían ser ciertos Impuestos Internos en los cuales se paga una cantidad fija por cada unidad vendida, la elasticidad fiscal será inferior a 1 y los ingresos fiscales crecerán más lentamente que la base.

- Por otra parte, si la elasticidad fiscal tendiera a 1, los ingresos fiscales crecerían a la misma tasa que la base. Este sería un ejemplo de **impuesto proporcional**. Si bien no es fácil de encontrar ejemplos reales, podría ser eventualmente el caso de un IVA con una alícuota única y sin exenciones.

Ciertos factores influyen sobre la **elasticidad fiscal**:

- **Inflación**: cuando las escalas de tributación de un impuesto progresivo no están indexadas, la inflación puede aumentar la elasticidad del impuesto sobre la renta.
- **Gestión del Cumplimiento**: para un nivel dado de ingresos, un grado alto de cumplimiento tributario incidirá en una mayor recaudación tributaria, comportamiento que se traducirá en una mayor elasticidad.
- **Mora y postergaciones de vencimientos**: En contrapartida tanto la mora como el otorgamiento de postergaciones de vencimientos reducirá las elasticidades.
- **Eficiencia de la Administración tributaria**: finalmente, si los factores previos no varían significativamente, una mejora en los procesos de administración tributaria aumentará la recaudación de una mayor proporción de la base impositiva, lo que incidirá positivamente en una mayor elasticidad.

Flotabilidad y Elasticidad fiscal

Hasta aquí hemos considerado la proyección de la recaudación futura como el resultado de la tasa de crecimiento proyectada de la base imponible multiplicada por la elasticidad fiscal estimada. Para hacerlo hemos supuesto que la base imponible se mantiene constante. El siguiente paso es introducir el concepto de **flotabilidad fiscal** para incluir en nuestras proyecciones los cambios en la base imponible.

La **flotabilidad fiscal** mide la capacidad de respuesta de los ingresos fiscales a los cambios en la base imponible. La flotabilidad se basa en cambios en los ingresos, que incluyen tanto cambios en la política fiscal y en la gestión de la administración tributaria, como modificaciones de las alícuotas impositivas, cambios en la definición de la base imponible y variaciones en la aplicación.

La **elasticidad tributaria** mide la respuesta automática de los ingresos tributarios a los cambios en la base imponible. Así, la elasticidad tributaria se basa en cambios en los ingresos corregidos por cambios legislativos en el sistema tributario; excluye el efecto de cambios discrecionales en el sistema tributario. Como tal, considera únicamente el efecto de los cambios espontáneos en la base imponible.

Para propósitos de proyección, las elasticidades son más relevantes que las flotabilidades. Sin embargo, requieren que los datos sobre el efecto de los cambios en el sistema tributario sobre los ingresos fiscales se recopilen sistemáticamente. Adicionalmente, sería deseable que el Ministerio de Economía recopile datos sobre el PBI y los precios, ajustados por cambios en la política fiscal.

Si estos datos están disponibles, el enfoque de ajuste proporcional permite construir una serie de ingresos fiscales que excluye los efectos de los cambios discrecionales de la política fiscal a lo largo del tiempo.

No obstante, en la mayoría de los casos, no se tendrá acceso a datos sobre ingresos fiscales que hayan sido corregidos por cambios en la política fiscal y tendrán que realizar investigaciones empíricas utilizando datos sin procesar. Por lo tanto, los ejemplos y evaluaciones que desarrollaremos a continuación emplearán datos de ingresos fiscales no ajustados. Esto implica que utilizaremos el término "**elasticidad fiscal**" como sinónimo de "**flotabilidad fiscal**" solamente con un ánimo didáctico de simplificación atendiendo al tenor introductorio de la presente publicación.

Como consecuencia de dicha simplificación somos conscientes que las proyecciones calculadas con datos no ajustados deben interpretarse con cautela porque incorporan tanto los efectos de los cambios endógenos en la economía como los impactos de política exógenos. Por lo tanto, dado que utilizaremos **flotabilidad fiscal** en lugar de **elasticidades**, nuestra evaluación de la relación entre los ingresos fiscales y su base imponible será imperfecta en el actual contexto de incertidumbre.

Estimación de las Elasticidades como Razones de las Tasas de Crecimiento

Se puede calcular una elasticidad tomando la relación de las tasas de crecimiento. La elasticidad fiscal es simplemente una relación de la tasa de crecimiento de los ingresos fiscales de un período al siguiente dividido por la tasa de crecimiento de la base imponible subyacente durante el mismo período:

$$\varepsilon_{T/B} = \frac{\% \Delta T_t}{\% \Delta B_t} = \frac{\Delta T_t / T_{t-1}}{\Delta B_t / B_{t-1}}$$

Ventajas y Desventajas del Método

La proyección mediante elasticidades tiene ventajas y desventajas. Como ventajas podemos señalar que:

- este método es intuitivo y simple de implementar.
- las proyecciones se pueden construir incluso con datos limitados.

Por su parte como desventajas reconocemos que:

- este método no permite controlar los factores externos (es decir, cambios en los indicadores económicos que no sean la base impositiva indirecta).
- las elasticidades tributarias tienden a variar de un período a otro, lo que significa que las previsiones de ingresos pueden mostrar una amplia gama de resultados según la muestra utilizada para estimar las elasticidades.
- no es sencillo incorporar cambios pasados o futuros en la política fiscal.

Estimación de Elasticidad Mediante Regresión Lineal

También se puede recurrir a los modelos de **regresión lineal** para proyectar los futuros ingresos tributarios. Sin embargo, se deben tomar las precauciones adecuadas al aplicar estas técnicas porque no siempre se cuenta con suficientes datos disponibles para poder modelar las regresiones.

Como explicamos al desarrollar el concepto "**buenas muestras**", a medida que crece el volumen de la muestra las propiedades de la proyección serán más representativas.

Al realizar análisis mediante regresión una cuestión que no se debe pasar por alto son los quiebres estructurales y la estacionalidad.

Las proyecciones basadas en regresión son un delicado equilibrio entre las técnicas estadísticas previamente presentadas y la aplicación del juicio del pronosticador. Avancemos un poco con el concepto de correlación de variables.

Correlación

El análisis de relaciones entre variables es la parte de la Estadística que estudia el comportamiento conjunto de dos o más variables, con el objetivo de establecer cómo es la relación que las vincula, calcular el nivel de asociación y modelar la relación funcional.

Para evaluar si dos series de tiempo están correlacionadas, si ambas comparten la escala de medida, podemos trazar ambas series en el mismo gráfico de series temporales y observar hasta qué punto covarían a lo largo del tiempo con el objetivo de detectar cualquier similitud. Veamos un ejemplo

con el comportamiento del Producto interno bruto (PBI) y la presión tributaria neta total¹² medidas en miles de millones de pesos (a precios de 2004):

Tabla N.º 6. Producto Interno Bruto y Presión Tributaria. Valores en Miles de Millones de Pesos a Precios de 2004. 2013 a 2022

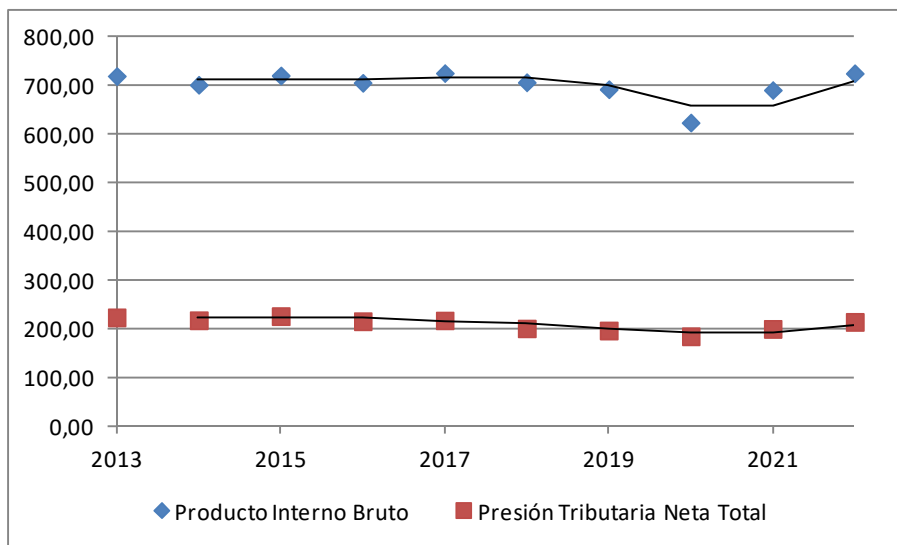
Año	Producto Interno Bruto	Variación PBI	Presión Tributaria Neta Total	Variación Presión Tributaria
2013	720,41	-	224,62	-
2014	702,31	-2,51%	218,58	-2,69%
2015	721,49	2,73%	226,93	3,82%
2016	706,48	-2,08%	216,69	-4,51%
2017	726,39	2,82%	218,24	0,71%
2018	707,38	-2,62%	201,70	-7,58%
2019	693,22	-2,00%	197,59	-2,03%
2020	624,59	-9,90%	185,79	-5,97%
2021	691,54	10,72%	200,88	8,12%
2022	725,81	4,96%	215,09	7,07%

Fuente: Elaboración propia a partir de “Presión Tributaria según clasificación internacional” y “Recaudación Nominal según clasificación internacional”. DNIAF. Recuperado de: <https://www.argentina.gob.ar/economia/ingresospublicos/recaudaciontributaria>

La regresión lineal se basa en la noción de que una variable (por ejemplo, el PBI) puede ayudar a explicar otra (la presión tributaria) y, por lo tanto, ambas variables tienen algún tipo de relación lineal, es decir, tienden a estar “**correlacionadas**”.

¹² La presión tributaria neta total surge de sumarle a la presión neta de impuestos nacionales (restados las deducciones e reintegros a la exportación) la presión de impuestos provinciales. La fuente de los datos es la Dirección Nacional de Investigaciones y Análisis Fiscal del Ministerio de Economía en base a datos de la AFIP, ANSES, Dirección Nacional de Asuntos Provinciales y Oficina Nacional de Presupuesto.

Gráfico N.º 5. Producto Interno Bruto y Presión Tributaria. 2013 a 2022

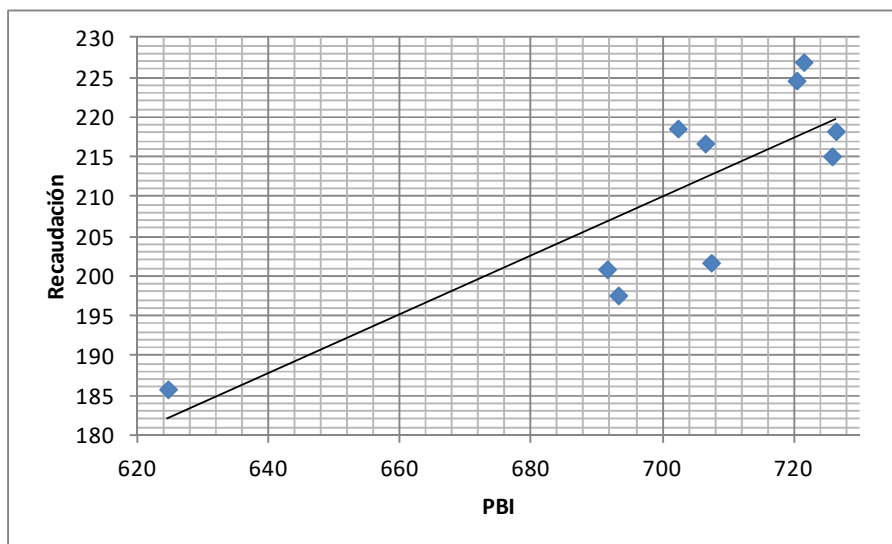


En esta primera aproximación podemos visualizar que ambas variables siguen un comportamiento similar, cuando una disminuye la otra también lo hace y viceversa. Sin embargo, como nos muestra la curva que une los puntos de cada variable en los diez años que hemos tomado como muestra, la “panza” de dicha curva no es idéntica. Esa “imperfección” en el comportamiento de la correlación lo podemos también apreciar en las diferencias del porcentaje de variación de ambas variables que exponemos en la tabla previa.

El paso siguiente para analizar la correlación entre ambas variables es elaborar un gráfico al que llamaremos “**Diagrama de Dispersión**”. En él expresaremos visualmente el grado de correlación, si existe, entre ambas series.

Para explorar el tamaño de esta posible correlación utilizando un diagrama de dispersión representaremos los datos de la tasa de crecimiento de la presión tributaria neta total y del PBI, en un par de ejes de coordenadas, colocando en las abscisas la variación del PBI (eje x) y en las ordenadas el crecimiento de la presión tributaria (eje y).

Gráfico N.º 6. Correlación entre Producto Interno Bruto y Presión Tributaria. 2013 a 2022



En el caso del ejemplo ambas variables parecen evolucionar en torno a una tendencia compartida, por lo cual, estas dos series parecería que están correlacionadas positivamente. Podemos ilustrar esta relación positiva dibujando una línea con una pendiente positiva mayor que uno.

Como próximo paso realizaremos un estudio analítico de correlación (grado de asociación) y de regresión (relación funcional). La correlación consiste en examinar la naturaleza y magnitud de la asociación lineal entre las dos variables con el objetivo de establecer el grado de asociación que existe entre ambas variables. Como la correlación estadística nos indica si dos variables están relacionadas, el valor del índice de correlación varía en el intervalo $[-1, 1]$. El signo indica el sentido de la relación:

- Si $r = 1$: existe una correlación positiva perfecta.
- Si $0 < r < 1$: existe una correlación positiva.
- Si $r = 0$: no existe una relación lineal.
- Si $-1 < r < 0$: existe una correlación negativa.
- Si $r = -1$: existe una correlación negativa perfecta.

Estimaremos una recta de regresión de la recaudación en función del PBI. A partir del trazado de dicha recta podremos proyectar cuál podría ser la recaudación si el PBI tuviera determinado valor. El comportamiento de la recta lo podemos expresar con el siguiente modelo lineal:

$$Y = \alpha + \beta X + \varepsilon$$

Al visualizar los datos, podemos apreciar que Y (presión tributaria) varía de acuerdo a cómo varíe X (PBI). Como dato adicional representamos con ε el componente aleatorio que suponemos se comportará como una distribución normal con media cero y un desvío estándar propio.

Como vemos en el gráfico previo no es una correlación positiva perfecta. Nos damos cuenta de ello, como regla general, porque si dicha curva es menos empinada que una línea de 45 grados el crecimiento de la recaudación tributaria reacciona menos de uno a uno a los cambios en el crecimiento del PIB. En otras palabras, **la elasticidad de los ingresos tributarios con respecto al PIB es menor a uno**.

Los principios básicos de las regresiones lineales usan mínimos cuadrados ordinarios (MCO). Por su complejidad, dado el carácter introductorio de la presente publicación, no nos extenderemos sobre estos modelos en esta ocasión.

Solamente señalaremos que también podemos ampliar la fórmula de regresión lineal para que sea multivariada. Es decir que podemos intentar cuantificar la relación lineal entre nuestra variable dependiente (por ejemplo la presión tributaria) y varias variables macroeconómicas independientes diferentes. Además del PIB también podríamos querer incluir una medida para la inflación.

La fórmula para los coeficientes de regresión en un caso de múltiples variables será más complicada y se deberá recurrir a paquetes estadísticos de software.

Al especificar una ecuación de regresión lineal, se deben considerar la cantidad de variables independientes que se desean incluir en la ecuación de regresión. Una variable independiente es una variable en el lado derecho de la ecuación y se la incluye porque se cree que está relacionada con la variable dependiente en el lado izquierdo (en nuestro ejemplo la presión tributaria) y, por lo tanto,

podría ayudar a explicar su variación. Pero una variable independiente solamente debe incluirse si existe alguna justificación previa sobre qué factores económicos influyen en la variable dependiente y, por un principio de simplicidad, no se debieran incluir variables independientes que estén altamente correlacionadas entre sí porque no brindan ningún poder explicativo adicional. En definitiva, la elección de las variables independientes es clave.

Al proyectar los ingresos tributarios nuestra variable dependiente se puede relacionar con diferentes variables independientes como las que sugerimos a continuación:

Variable Dependiente	Posibles variables independientes	
Ingresos Fiscales Totales Agregados	PBI	
Impuesto a las Ganancias de Personas Jurídicas	Beneficios empresariales	Precios comparados
IVA	Consumo de los hogares	
Derechos De Importación	Importaciones	Tipo de cambio real

Conclusiones

Hemos señalado que las seis características clave de un buen impuesto son eficiencia, equidad, simplicidad, flexibilidad, transparencia y eficacia. En términos generales, ninguna política tributaria puede perseguir todos estos objetivos al mismo tiempo, lo que significa que abundan las compensaciones en el diseño de un sistema tributario bien equilibrado. El actual diseño tributario grava una amplia gama de actividades y pone demasiado énfasis sobre los impuestos indirectos (GONZÁLEZ CAO, El buen impuesto, 2023). Como hemos visto al analizar las series temporales de la presión tributaria el nivel de tributación y la elección de actividades gravables difieren sustancialmente según las prioridades de cada gestión de gobierno.

En el actual contexto de restricción de acceso al crédito, la proyección confiable de los futuros ingresos fiscales es de vital importancia, porque los impuestos son la principal fuente de ingresos del Estado y el reaseguro de los equilibrios presupuestarios. Las proyecciones de ingresos confiables y el análisis del impacto de las futuras medidas tributarias ayudan a los gobiernos a alinear los gastos con los ingresos y, por lo tanto, son la base de una política sólida y sostenible de equilibrio fiscal (GONZÁLEZ CAO, Nociones básicas de desempeño fiscal, 2023).

A lo largo de las publicaciones que integran esta serie del CEAT hemos intentando presentar la importancia de la proyección de ingresos tributarios para los responsables de la formulación de la política fiscal y su rol clave como insumo en el proceso de elaboración del presupuesto. En tal sentido hemos señalado que el proceso de proyección de ingresos se basa en dos pilares, que son igualmente importantes: el marco institucional que respalda el proceso de proyección de ingresos y las técnicas de proyección. En cuanto a los aspectos institucionales, desde el punto de vista cualitativo los procesos de proyección de ingresos deben presentar tres características: transparencia, simplicidad y formalidad (GONZÁLEZ CAO, Previsión de impuestos, 2023).

En cuanto a la proyección de ingresos consideramos que el uso de técnicas cuantitativas adecuadas para pronosticar los ingresos fiscales permite usar las fuentes de datos disponibles para proyectar los ingresos fiscales futuros e, incluso, analizar el impacto de las potenciales futuras medidas que se impulsen.

Luego de realizar un repaso de las principales herramientas estadísticas hemos presentado los modelos incondicionales que se pueden usar para realizar proyecciones. Al respecto hemos aclarado que si bien estos modelos son fáciles de implementar incluso con datos limitados, no permiten que se controlen los efectos de ciertos factores externos como los cambios económicos o las modificaciones de la política fiscal ante un cambio de gobierno que tenga otras prioridades que su predecesor.

Para abordar esta complejidad, se ha sugerido la utilización de modelos condicionales que formulen sus proyecciones correlacionando otras variables que pueden incidir en los ingresos fiscales.

El primer modelo de proyección condicional que presentamos fue el método de la tasa impositiva efectiva, el cual luego lo ampliamos al enfoque del factor de cobertura.

Finalmente realizamos una primera aproximación al concepto de elasticidad fiscal y flotabilidad tributaria utilizando razones simples de tasas de crecimiento y tanteamos, apenas superficialmente, una posible estimación de estas elasticidades por medio de la aplicación métodos más avanzados como las técnicas de regresión lineal.

En vista de ello, las regresiones lineales nos permiten incluir variables independientes para mejorar la estimación o, incluso, una variable ficticia que intente reemplazar los potenciales impactos de futuras medidas de política fiscal.

Como conclusión final podemos acordar que la proyección de ingresos tributarios no es un mero ejercicio estadístico, sino que también requiere la aplicación de criterio. Lo cual demanda una amplia comprensión del comportamiento de los datos¹³ para evitar proyecciones inexactas o divorciadas del contexto macroeconómico del país. Comprender el comportamiento de los datos influirá en la elección tanto del modelo de proyección más adecuado como de la muestra de datos a utilizar.

¹³ Graficar los datos, visualizarlos, encontrar medidas resumen, analizar dispersión y varianza, identificar tendencias y quiebres estructurales.

Conceptos Clave

A los fines del presente trabajo utilizaremos los siguientes conceptos clave en el sentido que aquí se definen:

- **Administración Tributaria:** También llamada “Agencia tributaria”. Es el órgano con carácter de autoridad fiscal, que tiene por objeto la realización de una actividad estratégica del Estado consistente en la determinación, liquidación y recaudación de impuestos, aportes y contribuciones a la seguridad social y de sus accesorios para el financiamiento del gasto público. En algunos países también asume el control del comercio exterior y la percepción de la renta aduanera.
- **Autoridad Fiscal:** Es el representante del poder público que está facultado para recaudar impuestos, controlar a los sujetos obligados y contribuyentes, imponer sanciones previstas por los regímenes punitivos y de procedimiento tributario, e interpretar disposiciones de la ley, entre otros. En la República Argentina el órgano con carácter de autoridad fiscal a nivel nacional es la Administración Federal de Ingresos Públicos (AFIP).
- **Contribuyente:** Es la persona humana o jurídica que es susceptible de contraer obligaciones fiscales y que se encuentra obligada a ingresar las referidas obligaciones tributarias de conformidad con las leyes correspondientes. Es el sujeto pasivo del tributo, respecto de quien se verifica el hecho imponible.
- **Economía de Vigilancia:** Es la nueva organización de la economía en el marco de la digitalización y la plataformización. Se caracteriza por apropiarse de las experiencias humanas como materia prima gratuita para alimentar procesos de predicción y comercialización con el objetivo de modificar conductas y generar súper rentas excepcionales para un oligopolio de plataformas virtuales.
- **Evasión Fiscal:** es el comportamiento consciente y voluntario de los contribuyentes, consistente en ocultar su capacidad contributiva o disimular el hecho imponible, con el objetivo de erosionar la base imponible definida en la legislación o eludir las obligaciones que surgen de la normativa tributaria y de los Recursos de la Seguridad Social. Es un delito de defraudación a la Hacienda Pública consistente en la ocultación de ingresos, simulación o exageración de gastos deducibles, aplicación de desgravaciones y subvenciones injustificadas, entre otros, con la finalidad de evitar el pago de los tributos que por ley le correspondan a un sujeto contribuyente.

- **Fiscalización**: es el procedimiento que comprende la revisión, control y verificación que realiza la autoridad fiscal a los contribuyentes, respecto de los tributos que administra. Tiene como fin verificar el correcto cumplimiento de las obligaciones por medio del ejercicio de las facultades de inspección, verificación, auditoría fiscal y control que la ley le otorga a la Administración Tributaria. El término significa, cuidar y comprobar que se proceda con apego a la ley y a las normas establecidas al efecto.
- **Impuesto**: El Modelo de Código Tributario del CIAT (Centro Interamericano de Administraciones Tributarias, CIAT, 2015) lo define como la obligación que tiene como hecho generador y como fundamento jurídico una situación independiente de toda actividad estatal relativa al contribuyente que pone de manifiesto una determinada capacidad contributiva del mismo. Es decir que el pago del impuesto no origina por parte del Estado una contraprestación directa e inmediata en favor del sujeto contribuyente.
- **Tributo**: El Modelo de Código Tributario del Centro Interamericano de Administraciones Tributarias (Centro Interamericano de Administraciones Tributarias, CIAT, 2015) lo define como la prestación en dinero que el Estado exige, mediante el ejercicio de su poder de imperio, con el objeto de obtener recursos para financiar el gasto público y, en su caso, para el cumplimiento de otros fines de interés general. Dentro de la categoría de tributo encontramos tres subcategorías: impuestos, tasas y contribuciones especiales.

Referencias

- Banco Interamericano de Desarrollo, B. (15 de Diciembre de 2021). *Blogs del BID*. (A. Yarygina, L. Alejos, & U. González de Frutos, Edits.) Recuperado el 15 de 11 de 2023, de Gestión Fiscal: <https://blogs.iadb.org/gestion-fiscal/es/que-es-la-analitica-de-datos-y-como-puede-contribuir-a-mejorar-el-cumplimiento-tributario-en-america-latina-y-el-caribe/>
- Centro Interamericano de Administraciones Tributarias, CIAT. (2015). *Modelo de Código Tributario*. Ciudad de Panamá: CIAT.
- GONZÁLEZ CAO, R. L. (2023). *Mejores impuestos para mejores tiempos: el buen impuesto*. Buenos Aires, Argentina: Centro de Estudios en Administración Tributaria, Facultad de Ciencias Económicas, Universidad de Buenos Aires.
- GONZÁLEZ CAO, R. L. (2023). *Nociones básicas de desempeño fiscal*. Buenos Aires: Centro de Estudios en Administración Tributaria, Facultad de Ciencias Económicas, Universidad de Buenos Aires.
- GONZÁLEZ CAO, R. L. (2023). *Previsión de impuestos en tiempos imprevisibles*. Buenos Aires, Argentina: Centro de Estudios en Administración Tributaria, CEAT.
- MARTIN, F. (Enero de 2004). *El Método Directo de Proyección de los Ingresos Tributarios: un modelo para su aplicación*. Recuperado el 23 de Noviembre de 2023, de Asociación Internacional de Presupuesto Público, ASIP: <https://asip.org.ar/el-metodo-directo-de-proyeccion-de-los-ingresos-tributarios-un-modelo-para-su-aplicacion/>
- MARTIN, F., RODRIGUEZ CABELLO, J., & GONZALEZ, I. (2012). *Modelo de proyección de las cuentas fiscales y análisis de la sostenibilidad fiscal*. Obtenido de Comisión Económica para América Latina y el Caribe, CEPAL: <https://repositorio.cepal.org/server/api/core/bitstreams/f0d7fb1b-547e-4e8b-a2b9-87357f2115a0/content>
- Oficina Nacional de Presupuesto. (Octubre de 2022). *El Presupuesto en Cifras*. Recuperado el 6 de Julio de 2023, de sitio de Presupuesto Ciudadano: https://www.economia.gob.ar/onp/presupuesto_ciudadano/seccion3.php
- Oficina Nacional de Presupuesto, Subsecretaría de Presupuesto. (2016). *Manual de Clasificaciones Presupuestarias para el Sector Público Nacional* (Sexta edición ed.). (S. d. Hacienda, Ed.) Buenos Aires, Argentina: Ministerio de Hacienda y Finanzas Públicas.
- VIEGAS, J. C., RIAL, L. N., & GAJST, I. F. (2000). La información prospectiva: proyecciones o pronósticos. (F. d. Económicas, Ed.) *Contabilidad y Auditoría*, 1(12), 26.

Adicionalmente, si es de tu interés profundizar en el conocimiento de herramientas estadísticas puedes acceder a la siguiente bibliografía de consulta que puede resultarte de gran utilidad:

- APEZTEGUÍA & FERRARIO (2015). “Probabilidades y Estadística. Análisis de datos”. Universidad Nacional de La Plata. Recuperado desde:

http://sedici.unlp.edu.ar/bitstream/handle/10915/74877/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y

- KELMANSKY, Diana (2009). “Estadística para todos. Estrategias de pensamiento y herramientas para la solución de problemas”. Colección “Las ciencias naturales y la matemática”. Instituto Nacional de Educación Tecnológica. Ministerio de Educación. Recuperado desde: <https://www.inet.edu.ar/wp-content/uploads/2023/06/Estadistica-para-todos.pdf>
- MARONNA, Ricardo (1995). “Probabilidad y estadística elementales para estudiantes de ciencias”. Facultad de Ciencias Exactas. Universidad de La Plata. Recuperado de: http://sedici.unlp.edu.ar/bitstream/handle/10915/94969/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y

Publicaciones del Autor

Serie "Mejores Impuestos"

- GONZÁLEZ CAO, Rodrigo Luis (2023d). “*Mejores impuestos para mejores tiempos: el buen impuesto*”. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <https://www.economicas.uba.ar/wp-content/uploads/2023/07/el-buen-impuesto.pdf>
- GONZÁLEZ CAO, Rodrigo Luis (2023e). “*Nociones básicas de desempeño fiscal*”. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <https://www.economicas.uba.ar/wp-content/uploads/2023/08/Nociones-basicas-de-desepe%C3%B1o-fiscal.pdf>
- GONZÁLEZ CAO, Rodrigo Luis (2023f). “*Previsión de Impuestos en Tiempos Imprevisibles*”. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <https://www.economicas.uba.ar/wp-content/uploads/2023/10/Prevision-de-impuestos.pdf>

Serie "Administración Tributaria del Siglo XXI"

- GONZÁLEZ CAO, Rodrigo Luis (2022b). *"Procesos críticos y buena gobernanza"*. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <https://www.economicas.uba.ar/wp-content/uploads/2022/10/procesos-criticos.pdf>
- GONZÁLEZ CAO, Rodrigo Luis (2022a). *"Transformación organizacional: gestión del talento humano e innovación basada en datos"*. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <https://www.economicas.uba.ar/wp-content/uploads/2022/08/gestion-talento-e-innovacion-basada-en-datos.pdf>
- GONZÁLEZ CAO, Rodrigo Luis (2021a). *"Recursos de la Seguridad Social y Administración Tributaria. Su contribución conjunta a la cohesión social"*. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <http://www.economicas.uba.ar/wp-content/uploads/2021/03/CEAT-RecaudacionImpuestos-y-SeguridadSocial.pdf>
- GONZÁLEZ CAO, Rodrigo Luis (2018c). *"Fortalecimiento de las Capacidades de Gestión de Riesgo de los Recursos de la Seguridad Social"*. Instituto de Estudios de las Finanzas Públicas Americanas IEFPA. XXVII Encuentro Internacional de Administradores Fiscales, noviembre de 2018, Villa Carlos Paz, Provincia de Córdoba. Recuperado desde: (academia.edu) https://www.academia.edu/44926534/Fortalecimiento_de_las_Capacidades_de_Gesti%C3%B3n_de_Riesgo_de_los_Recursos_de_la_Seguridad_Social (versión en español en archivo IEFPA) http://old.amfeafip.org.ar/seminario/2018/campus/gestion_riesgo_seguridad_social.pdf

Serie "Gestión de personas y del capital humano"

- GONZÁLEZ CAO, Rodrigo Luis (2023a). *"Gestión estratégica de personas y del capital humano"*. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <https://www.economicas.uba.ar/wp-content/uploads/2023/04/gestion-estrategica-de-personas-y-capital-humano.pdf>
- GONZÁLEZ CAO, Rodrigo Luis (2023b). *"Organizar la función de gestión de personas y del capital humano en una Administración Tributaria"*. Centro de Estudios en Administración Tributaria –

CEAT. Universidad de Buenos Aires. Recuperado desde: <https://www.economicas.uba.ar/wp-content/uploads/2023/05/Organizar-la-funcion-de-gestion-de-personas-y-capital-humano-en-una-AATT.pdf>

- GONZÁLEZ CAO, Rodrigo Luis (2023c). “*La organización del trabajo en una Administración Tributaria*”. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <https://www.economicas.uba.ar/wp-content/uploads/2023/06/La-organizacion-del-trabajo-en-una-AATT.pdf>

Serie "Futuro del trabajo y trabajo del futuro"

- GONZÁLEZ CAO, Rodrigo Luis (2022c). “*Empleos atípicos: la tercera posición, ¿liberación o dependencia?*”. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <https://www.economicas.uba.ar/wp-content/uploads/2022/12/Empleos-atipicos.pdf>

- GONZÁLEZ CAO, Rodrigo Luis; ALVAREZ, Shirdi Sebastián y REAL, Karina (2021). “*Influencers, unboxers y eSports. Los empleos atípicos en las redes sociales y su impacto en los mercados laborales y en los Recursos de la Seguridad Social*”. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <https://www.economicas.uba.ar/wp-content/uploads/2021/09/2c1-CEAT-Influencers-Unboxers-eSports.pdf>

- GONZÁLEZ CAO, Rodrigo Luis; BOGGERO, Geraldine (2021). “*Externalidades sociales de la plataformización económica. Impacto en el trabajo y la Seguridad Social y potenciales herramientas de mitigación*”. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <https://www.economicas.uba.ar/wp-content/uploads/2021/07/Externalidades.pdf>

- GONZÁLEZ CAO, Rodrigo Luis; ROCA, Guillermo (2021). “*Plataformización de la economía y plataformas digitales. Su impacto en las relaciones laborales y los Recursos de la Seguridad Social*”. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <https://www.economicas.uba.ar/wp-content/uploads/2021/05/Plataformizacion-de-la-economia-y-plataformas-digitales.pdf>

- GONZÁLEZ CAO, Rodrigo Luis (2021b). *“La economía de vigilancia y las plataformas. Su evolución y la mitigación de sus externalidades negativas a través de la regulación y la fiscalidad”*. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <https://www.economicas.uba.ar/wp-content/uploads/2021/03/economia-de-vigilancia-y-plataformas.pdf>
- GONZÁLEZ CAO, Rodrigo Luis (2020b). *“Los desafíos de las administraciones tributarias ante el futuro del trabajo en la nueva normalidad”*. Centro Interamericano de Administraciones Tributarias. Revista N.º 46 CIAT (Octubre/2020. Recuperado desde: (versión en español) https://www.ciat.org/Biblioteca/Revista/Revista_46/Espanol/2020_RAT_46_ebook_es.pdf (english version) Tax Administration Review CIAT/AEAT/ IEF No. 46 (November, 2020) https://www.ciat.org/Biblioteca/Revista/Revista_46/Ingles/2020_TR46_ebook_ing.pdf
- GONZÁLEZ CAO, Rodrigo Luis (2020a). *“Trabajo forzoso y trabajo infantil. El combate a la esclavitud del siglo XXI desde la Administración Tributaria”*. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado desde: <http://www.economicas.uba.ar/wp-content/uploads/2020/09/Trabajo-forzoso-y-trabajo-infantil..pdf>
- VADELL, GONZÁLEZ CAO et al. (2020). *“Fiscalidad de las criptomonedas y la economía digital”*. Consejo Profesional de Ciencias Económicas C.A.B.A. Ed. Edicon. Buenos Aires. Recuperado de: <http://edicon.org.ar/wp-content/uploads/2020/11/FISCALIDAD-DE-LAS-CRIPTOMONEDAS-Y-DE-LA-ECONOM%C3%8DA-DIGITAL.-web.pdf>
- GONZÁLEZ CAO, Rodrigo Luis (2019). *“El futuro del trabajo en la economía del conocimiento”*. Instituto de Estudios de las Finanzas Públicas Americanas IEFPA. XXVIII Encuentro Internacional de Administradores Fiscales, noviembre de 2019, San Rafael, Provincia de Mendoza. Recuperado de: (academia.edu) https://www.academia.edu/44926422/El_Futuro_del_trabajo_en_la_econom%C3%ADa_del_conocimiento (versión en español en archivo IEFPA) http://old.amfeafip.org.ar/seminario/2019/documentos/rodrigo_gc.pdf
- GONZÁLEZ CAO, Rodrigo Luis (2018a). *“Los trabajadores de plataforma y la evolución de las relaciones laborales”*. Presentación en VIII Congreso Nacional de Entes Recaudadores. Centro de Estudios en Administración Tributaria – CEAT. Universidad de Buenos Aires. Recuperado

desde: http://www.economicas.uba.ar/wp-content/uploads/2018/10/CEAT_Trabajadores_de_plataforma.docx.pdf

- GONZÁLEZ CAO, Rodrigo Luis (2018b). *“La futura recaudación y fiscalización de las cotizaciones a la seguridad social ante los cambios de la economía digital”*. Centro Interamericano de Administraciones Tributarias. Revista N.º 44 CIAT (Octubre/2018, en español). Recuperado desde: https://www.ciat.org/Biblioteca/Revista/Revista_44/espanol/2018_RAT44_ebook_es.pdf (English version) Tax Administration Review CIAT/AEAT/ IEF No. 44 (October, 2018) <https://biblioteca.ciat.org/opac/book/5625>

Tabla de Contenido**Tabla de contenido**

Nota del Autor	3
Agradecimientos	3
Resumen	4
Palabras Clave	4
Abstract.....	4
Keywords.....	5
Medición del Impacto de Futuras Medidas Tributarias.....	5
Análítica de Datos, Estadística Descriptiva y Herramientas	8
Población.....	9
Muestras	10
Análisis de los Datos.....	11
Herramientas Estadísticas.....	11
Unidades Muestrales	12
Variables.....	13
¿Valores Relativos o Valores Absolutos?	13
Razón.....	13
Tasa	14
Porcentaje y Proporción	14
Variaciones Relativas	14
Muestras y Muestreo.....	15
Muestreo.....	16
Muestreo Sistemático	16
Muestreo Aleatorio Simple	17
Muestreo Aleatorio Estratificado.....	19
Muestreo por Bloques o Conglomerados	20
Variabilidad Muestral.....	20
Mediciones Válidas	21
Buenas Muestras: Muestra Representativa.....	22

Malas Muestras: Sesgo	23
Tipos de Errores	25
Datos y Variables.....	27
Datos Numéricos y Categóricos	27
Datos Numéricos.....	27
Datos Categóricos	28
Distribución de Frecuencias de una Variable Numérica	28
VARIABLES NUMÉRICAS	31
Variables Numéricas Discretas.....	31
Variables Numéricas Continuas	31
VARIABLES CATEGÓRICAS.....	32
Gráficos Circulares	33
Gráficos de Columnas	34
Gráfico de Tiempo.....	35
Distribución y Posición Relativa	38
Tendencia Central o Valor Central.....	38
Media	42
Mediana	42
Valores Atípicos.....	43
Medidas Resumen y Dispersión	44
Varianza.....	45
Desvío Estándar.....	46
Desvío Estándar Muestral	47
Varianza Muestral y Dispersión	48
Posición Relativa: Dispersión y Variabilidad	48
Rangos de Valores: Mínimo y Máximo	49
Cuantiles y Distancia Inter cuantil.....	49
Números Resumen y Gráfico de Caja y Brazos	51
Percentiles.....	54
¿Cómo se Calcula un Percentil en un Conjunto de Datos No Agrupados?	54
¿Cómo se Calcula un Percentil en un Conjunto de Datos Agrupados?.....	55
Formas que Describen las Distribuciones	56
Distribución Simétrica.....	57

Distribución Asimétrica.....	58
Curvas de Densidad.....	59
Teorema del Límite Central (TLC)	63
Enunciado del Teorema	64
Ejemplos de Sumas de Variables	64
Fuentes de Datos	66
Inspección de los Datos.....	67
Visualización de Datos en Series Temporales.....	67
Diagnóstico Básico de Datos	68
Detección de Quiebres Estructurales.....	71
Descripción General de las Técnicas de Proyección de Ingresos.....	73
Modelos Incondicionales de Proyección Tributaria.....	74
Proyección Aditiva en Función de los Ingresos Pasados.....	75
Media Móvil de los Ingresos Pasados	76
Ponderaciones Diferenciadas de los Ingresos Pasados	76
Proyección en Función de Tendencias Multiplicativas	77
Proyección en Función de Promedio Móvil de las Tasas de Crecimiento.....	77
Evaluación de Errores de Proyección.....	78
Errores de Proyección (EP).....	79
Error Medio de Proyección (EMP)	80
Error Absoluto Medio (EAM)	81
Error Porcentual Medio (EPM).....	81
Error Porcentual Absoluto Medio (EPAM).....	82
Error Estándar del Proyección (EEP)	82
Error Cuadrático Medio (ECM).....	83
Modelos Condicionales de Proyección Tributaria	83
Método de la tasa impositiva efectiva (TIE).....	84
Método del Factor de Cobertura	88
Proyección Mediante Elasticidades	92
Flotabilidad y Elasticidad fiscal	94
Estimación de las Elasticidades como Razones de las Tasas de Crecimiento.....	96
Estimación de Elasticidad Mediante Regresión Lineal.....	97
Correlación.....	97

Conclusiones	103
Conceptos Clave.....	105
Referencias.....	107
Publicaciones del Autor	108
Serie "Mejores Impuestos"	108
Serie "Administración Tributaria del Siglo XXI".....	109
Serie "Gestión de personas y del capital humano".....	109
Serie "Futuro del trabajo y trabajo del futuro".....	110
Tabla de Contenido	113
Información del Autor.....	117

Información del Autor

Rodrigo Luis GONZÁLEZ CAO se recibió de Contador Público y Licenciado en Administración en la Pontificia Universidad Católica Argentina y ha cursado la Maestría en Administración Pública en la Universidad de Buenos Aires.

En el ámbito de las administraciones tributarias ha ejercido roles de jefatura en áreas de investigación, fiscalización, control de gestión y operaciones en la ADMINISTRACIÓN FEDERAL DE INGRESOS PÚBLICOS (AFIP) de Argentina.



Universidad de Buenos Aires
Facultad de Ciencias Económicas

CEAT



Centro de Estudios en Administración Tributaria

Investigación y capacitación aplicados a los ingresos públicos