

---

**Pronóstico de la tasa de cobertura de educación superior del departamento de Casanare desde 2019 hasta 2022 mediante un modelo SARIMA en series de tiempo.**

---



**LOS LIBERTADORES**  
FUNDACIÓN UNIVERSITARIA

---

**Pronóstico de la tasa de cobertura de  
educación superior del departamento de  
Casanare desde 2019 hasta 2022 mediante un  
modelo SARIMA en series de tiempo.**

---



Presentado por

**Cesar Miguel Palacios**

en cumplimiento parcial de los requerimientos para optar al título

de

**Especialista en Estadística Aplicada**

*Dirigida por*

**MSc. Sebastián Lozano Forero**

Profesor

**Notas de aceptación**

---

---

---

---

---



---

Firma del presidente del jurado

**LOS LIBERTADORES**  
FUNDACIÓN UNIVERSITARIA

---

Firma del jurado

---

Firma del jurado

Bogotá DC, Diciembre de 2018.



# LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Las directivas de la Fundació Universitaria Los Libertadores, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento. Estos corresponden únicamente a los autores y a los resultados de su trabajo.

## DEDICATORIA



# LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Dedicado a mi madre por  
su incondicional apoyo  
durante toda mi formación  
académica.

## AGRADECIMIENTOS



# LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Agradezco a los docentes de la Fundación Universitaria Los Libertadores por su adecuada orientación profesional durante el tiempo en que cursé esta especialización.

# Índice general

## Contenido

<b>Capítulo 1. Introducción .....</b>	<b>11</b>
<b>Capítulo 2. Planteamiento del Problema .....</b>	<b>13</b>
Para avanzar en la resolución de la pregunta de investigación se plantearon los siguientes objetivos:.....	15
Objetivo General.....	15
Objetivos Específicos.....	15
Justificación .....	16
<b>Capítulo 3. Marco Teórico .....</b>	<b>17</b>
Tasa de cobertura de educación superior .....	17
Series de tiempo .....	19
Componentes de una serie de tiempo .....	19
Clasificación de las series de tiempo.....	20
Proceso estacional, autorregresivo, integrado y de media móvil SARIMA (p,d,q) × (P,D,Q) <sub>s</sub> .....	20
Pasos a seguir para la identificación de un modelo SARIMA adecuado .....	21
<b>Capítulo 4. Marco Metodológico.....</b>	<b>23</b>
Tipo de investigación .....	23
Investigación exploratoria .....	23
Investigación descriptiva.....	23
Diseño metodológico .....	23
Tipo de estudio .....	23
Área de estudio .....	24
Universo y muestra .....	24
Métodos e instrumentos de recolección de datos .....	24
Plan de tabulación y análisis .....	24
Enfoque.....	24
<b>Capítulo 5. Análisis y Resultados .....</b>	<b>26</b>

Construcción de la base de datos .....	26
Análisis de los datos bajo en el enfoque Box-Jenkins.....	28
Obtención de modelos SARIMA .....	30
Verificación de supuestos .....	31
Predicción de la tasa de cobertura de educación superior del departamento de Casanare .....	32
<b>Capítulo 6. Conclusiones .....</b>	<b>34</b>
<b>Recomendaciones .....</b>	<b>35</b>
<b>Bibliografía .....</b>	<b>36</b>
<b>Apéndice A. Código en R .....</b>	<b>37</b>



## Lista de gráficas

Gráfica 1 Tasa de cobertura de educación superior de Casanare.....	27
Gráfica 2. Primera diferencia de la serie de datos.....	28
Gráfica 3. Segunda diferencia de la serie de datos .....	29
Gráfica 4. Tercera diferencia de la serie de datos .....	30
Gráfica 5. Gráficas de los errores de la serie ajustada .....	31
Gráfica 6. Predicciones de la tasa de cobertura de educación superior de Casanare .....	33

## Lista de tablas

Tabla 1. Cálculo de la tasa de cobertura de educación superior de Casanare basado en estadísticas del SNIES y del DANE. ....	27
Tabla 2. Modelos SARIMA ajustados .....	31
Tabla 3. Pronósticos de la tasa de cobertura de educación superior de Casanare con un 80% de confianza .....	33

# **Pronóstico de la tasa de cobertura de educación superior del departamento de Casanare desde 2019 hasta 2022 mediante un modelo SARIMA en series de tiempo.**

## **Resumen**

**Palabras claves:** Tasa de cobertura, educación, serie de tiempo, estacionariedad, estacional, SARIMA.

Este documento tiene como objetivo analizar la serie de tiempo de la tasa de cobertura de educación superior del departamento de Casanare bajo la metodología propuesta por Box y Jenkins, que permita obtener un modelo SARIMA para la predicción del comportamiento de este indicador en el corto plazo. Se pretende que con este pronóstico la opinión pública en general y las entidades gubernamentales cuenten con datos fiables respecto de la efectividad de las políticas en materia de cobertura de educación superior en la región y de esta forma contar con evidencias confiables para asumir posturas críticas. Luego de ajustar diversos modelos se estableció que aquel que mejor describe el comportamiento de este indicador es un SARIMA(1,1,0)×(0,1,1)<sub>2</sub>. Los resultados del pronóstico para el período comprendido entre 2019 y 2022 muestran que existe una tendencia creciente pero con fluctuaciones. Se espera que durante este período de tiempo se supere la barrera del 30% de cobertura llegando a un 31,30% en 2022 con un 80% de confianza.

# **Forecast of the coverage rate of college education in the department of Casanare, from 2019 to 2022 through an ARIMA model in time series.**

## **SUMMARY**

**Keywords:** Coverage rate, education, time series, stationarity, seasonal, SARIMA.

The objective of this document is to analyze the time series of the college education coverage rate of the Department of Casanare, under the methodology proposed by Box and Jenkins, which allows obtain a SARIMA model for the prediction of the behavior of this indicator in the short term. It is intended that with this forecast public opinion in general and government entities have reliable data regarding about the effectiveness of policies in terms of coverage of college education in the region and in this way have reliable evidence to assume critical positions. After adjusting various models, it was established that the one that best describes the behavior of this indicator is a SARIMA (1,1,0) × (0,1,1)<sub>2</sub>. The results of the forecast for the period between 2019 and 2022 show that there is a growing trend but with fluctuations. It is expected that during this period of time the barrier of 30% coverage will be overcome, reaching 31.30% in 2022 with 80% confidence.

## Capítulo 1. Introducción

La cobertura de educación superior constituye un aspecto de especial relevancia dado su carácter de indicador educativo, por lo que a las entidades gubernamentales y a la opinión pública en general le resulta interesante conocer dicha cifra año a año, lo que les permite asumir una posición crítica respecto del comportamiento de dicha variable a lo largo del tiempo.

Dicho indicador se ve afectado por dos aspectos principalmente: el número de jóvenes que se matriculan en las instituciones de educación superior de determinada región y la población entre 17 y 21 años que habita tal lugar. Para el caso del departamento de Casanare, se ha presentado un incremento a lo largo de los últimos años en el número de habitantes y en la demanda educativa por parte de los jóvenes, situación que ha conllevado a la aparición de nuevas instituciones y al aumento de la oferta. Tal situación haría suponer un incremento en la tasa de cobertura de educación superior, sin embargo, se observa que el número de jóvenes que se matriculan a un programa de educación superior presenta fluctuaciones dependiendo de la época del año, para este caso, primer y segundo semestre académico.

De igual forma, existen otros aspectos que influyen en la tasa de cobertura, como son la emigración de jóvenes hacia otras regiones, aspectos económicos que impiden que algunas personas puedan realizar sus estudios superiores y la falta de una oferta amplia que satisfaga las expectativas educativas de la juventud casanareña. Se podría decir que son aspectos que juegan en contra y que podrían contribuir a que la tasa de cobertura en educación superior no sea la esperada.

A lo largo de los últimos años el ministerio de educación ha publicado las cifras de cobertura de educación superior tanto a nivel nacional como departamental. Sin embargo, luego de realizarse una revisión de la literatura no se encontró que se hayan realizado estudios de predicción de indicadores de educación superior, en específico, en el departamento de Casanare. Por tal motivo, se considera pertinente realizar un estudio mediante series de tiempo, bajo el enfoque propuesto por Box y Jenkins, que permita proyectar, con un nivel de confianza alto, el comportamiento de la cobertura de educación superior en el departamento, y de esta forma, contar con cifras que permitan tomar decisiones tendientes a mejorar cada vez más este índice, así como para emitir juicios de valor respecto de la gestión de los entes gubernamentales sobre este tema.

Es así como el lector podrá encontrar en el capítulo dos, el planteamiento del problema y la formulación de la pregunta de investigación que se pretende sea resuelta a través de este estudio. De igual forma, se podrán encontrar los objetivos propuestos y la justificación del trabajo, donde se desglosan aquellos aspectos que motivaron la realización de este proyecto.

En el capítulo tres, el lector podrá evidenciar un desarrollo teórico referente a las series de tiempo, su clasificación y sus respectivas componentes, de igual forma se explicará de una forma sencilla, clara y comprensible el procedimiento a seguir para encontrar el mejor modelo SARIMA que permita la descripción de los datos, procedimiento compuesto por las fases de identificación, modelación, validación y pronóstico. También, se profundizará más

en lo relacionado con la tasa de cobertura de educación superior, dando a conocer algunas cifras de orden internacional, nacional y regional. Por otra parte, en el capítulo cuatro, se especificará los tipos de investigación aplicados (exploratoria y descriptiva), el diseño metodológico y el enfoque utilizado.

Posteriormente, en el capítulo cinco, se realizará un análisis estadístico de la serie de tiempo de la variable en cuestión, bajo el enfoque Box Jenkins, lo que permitirá obtener varios modelos estadísticos que expliquen el comportamiento de la cobertura a lo largo del período comprendido entre los años 2001 y 2017. Una vez obtenidos dichos modelos se seleccionará el que mejor se ajuste y que cumpla con supuestos de normalidad en los errores, así como ausencia de autocorrelación serial.

Con el modelo ideal seleccionado, se podrá realizar una predicción en el corto plazo sobre el comportamiento que tendrá la variable cobertura de educación superior en Casanare durante los próximos cinco años. De esta forma, se podrá contar con información confiable que permita asumir una postura crítica frente a este indicador educativo, y las entidades gubernamentales e instituciones de educación superior que decidan examinar los resultados tendrán información que les oriente hacia la toma de decisiones en pro del mejoramiento de sus cifras de cobertura en la región.

Finalmente, en el capítulo seis, el lector podrá consultar las conclusiones derivadas del pronóstico realizado mediante el modelo SARIMA ajustado a los datos. De igual forma, en este mismo apartado se presentan algunas sugerencias para posibles estudios de la misma naturaleza que se vayan a llevar a cabo.

## Capítulo 2. Planteamiento del Problema

La cobertura de educación superior en el departamento de Casanare se ha venido incrementando durante los últimos años gracias a la llegada de algunas instituciones cuya oferta académica se ha ido adaptando parcialmente a las necesidades de formación de la población y como respuesta a la demanda de profesionales, tecnólogos y técnicos por parte de sectores como el industrial, empresarial, comercial, bancario, salud y educación. Tal y como afirma Rama (2010) “hay una multiplicidad de factores impulsores del proceso de ampliación de la matrícula. No se trata del mero resultado de la existencia de una masa de bachilleres que han concluido sus estudios medios. Tampoco sólo una derivación de la medición entre los beneficios presentes del ingreso a los mercados laborales. Se asocia a un cambio en la estructura de los procesos productivos y de los mercados de trabajo que demandan más profesionales y técnicos superiores, asociado a un específico modelo de acumulación de capitales”.

Sin embargo, ha sido evidente que las familias que tienen mayor capacidad económica prefieren que sus hijos realicen los estudios superiores en instituciones de alta calidad en otros departamentos o en la capital del país, en el otro extremo se encuentran las familias de escasos recursos económicos quienes encuentran dificultades para que sus hijos puedan continuar su formación una vez culminada la educación básica secundaria y media, por lo que se ven abocados bien sea a suspender su proceso educativo o a ingresar a instituciones más económicas o de menor calidad. Lo anterior debido a que “la incidencia más compleja de la creciente masificación de la educación superior en la región ha sido en la diferenciación y en la calidad de la educación en general, y de algunas instituciones en particular. Tanto en las instituciones de mercado que se focalizan en la captación de la demanda de sectores sociales de menos recursos y menor capital cultural, así como en las instituciones sin niveles selectivos de ingresos o fuertes mecanismos internos o externos de aseguramiento de la calidad, la masificación tendió a reducir y relajar sus niveles de calidad” (Rama, 2010).

La tasa de cobertura de educación superior es un indicador utilizado por el Ministerio de Educación Nacional para medir los alcances y la efectividad de sus políticas educativas. Dicha tasa corresponde a la relación entre el número de estudiantes matriculados en las instituciones de educación superior de determinada región y el número de jóvenes entre 17 y 21 años que la habitan. En tal sentido, cabe mencionar que la población del departamento de Casanare, y en especial la de su capital, Yopal, tuvo una tasa de crecimiento considerablemente alta durante la segunda mitad del siglo XX, siendo una de las ciudades con mayor crecimiento demográfico en el país. Este fenómeno suscitó la llegada de algunas

universidades, en su mayoría del sector privado, así como el aumento de la oferta educativa de instituciones de carácter público como el SENA. De igual forma, el auge de la educación virtual y a distancia ha permitido ampliar un poco el abanico de posibilidades para los jóvenes casanareños.

Sin embargo, el hecho de que algunas carreras profesionales no sean ofertadas en la región, así como la calidad de las instituciones y de sus programas, en muchos de los casos incide en la decisión de desarrollar los estudios superiores por fuera del departamento, tal como lo afirma (Avila, 2008) “la educación en Casanare, se enfrenta hoy a una creciente tasa de población con necesidad de educación superior, una progresiva masa de jóvenes graduados requieren estudiar en el departamento, sin embargo no poseen los recursos necesarios para realizarlo. Aquellos que sí poseen los recursos económicos, envían sus hijos a estudiar a universidades de otras regiones, como es el caso prioritario con la U.P.T.C y la UNIBOYACÁ en Tunja y las Universidades Nacional de Colombia y los Andes en Bogotá, sector oficial y no oficial respectivamente.”

Dicha situación ha provocado que a pesar del incremento en la cobertura de educación superior, el índice del departamento de Casanare haya permanecido por debajo de las cifras nacionales. El Ministerio de Educación Nacional actualiza año a año sus estadísticas e indicadores de educación superior a partir de las cifras registradas en el Sistema Nacional de Información de la Educación Superior SNIES, lo que le permite a las entidades gubernamentales conocer cuál ha sido el comportamiento y la efectividad de sus políticas en dicho aspecto. Sin embargo, no se cuenta con estimaciones confiables que permitan predecir cuál será el comportamiento de tales indicadores, y en específico de la tasa de cobertura de educación superior en el departamento. Dado que es posible predecir, con un alto grado de confianza, el valor de una variable a partir de los datos históricos que le anteceden, las autoridades que dirigen el sector educativo del departamento y del país podrían contar con información estadística valiosa que les permita tomar decisiones acertadas y orientadas hacia el mejoramiento de la cobertura.

Por tales motivos se formula la siguiente pregunta problema:

¿Cuál es la proyección en el corto plazo de la tasa de cobertura de educación superior en el departamento de Casanare a partir de los datos históricos registrados entre los años 2001 y 2017?

## **Objetivos**

Para avanzar en la resolución de la pregunta de investigación se plantearon los siguientes objetivos:

### **Objetivo General**

Realizar un pronóstico en el corto plazo de la tasa de cobertura de educación superior en el departamento de Casanare a partir de los datos históricos registrados entre los años 2.001 y 2.017

### **Objetivos Específicos**

- Construir una base de datos con los registros históricos de la tasa de cobertura de educación superior en el departamento de Casanare.
- Obtener un modelo SARIMA, a partir de varios modelos ajustados, que explique con un alto grado de confianza el comportamiento de la tasa de cobertura de educación superior en el departamento de Casanare a través del tiempo.
- Validar el modelo SARIMA generado, a partir de supuestos de normalidad y ausencia de autocorrelación serial en los errores.

## **Justificación**

El uso de proyecciones mediante modelos estadísticos de series de tiempo no es muy común en el sector educativo colombiano, y en especial en el pronóstico de indicadores de educación superior. Sin embargo, dado que a lo largo de este siglo se han llevado a cabo registros históricos completos y organizados sobre diversas estadísticas del sector, gracias a los datos que proporcionan las instituciones educativas al Sistema Nacional de Instituciones de Educación Superior (SNIES), es posible realizar este tipo de estudios que pueden resultar muy pertinentes por diversos motivos.

Contar con proyecciones confiables sobre el comportamiento de indicadores de educación superior, y en específico de la tasa de cobertura en educación superior, le puede otorgar herramientas a las entidades gubernamentales para evaluar la efectividad que tendrán a futuro sus políticas educativas y tomar decisiones acertadas en la búsqueda de estrategias de mejoramiento en el caso en que las estimaciones no sean las esperadas.

De igual forma, le puede resultar interesante a la opinión pública conocer los resultados de dichas proyecciones para tener un criterio objetivo acerca del manejo que los dirigentes públicos de orden nacional y departamental le dan al sector educativo respecto de su cobertura.

Específicamente, en el departamento de Casanare, es oportuno conocer cuál será el comportamiento de la tasa de cobertura de educación superior, puesto que el número de jóvenes cada vez es mayor, y aspectos de índole económico, que han afectado a la región durante los últimos años, obligan a un número considerable de familias a poner su mirada en las instituciones de la región por encima de las de otros lugares del país. Es evidente que la caída del precio del petróleo ha tenido una incidencia negativa en la economía de la región, afectando de paso al sector comercial, y por ende los ingresos de muchas familias, por lo que un buen número de personas encuentran dificultad para enviar sus hijos a estudiar a otras ciudades de Colombia. Tal conjetura genera como interrogante saber si realmente la cobertura de educación superior del departamento se incrementará o si por el contrario tenderá a disminuir ante la posible reducción de jóvenes migrantes casanareños.



## Capítulo 3. Marco Teórico

### Tasa de cobertura de educación superior

Es considerado uno de los indicadores más importantes del sector educativo en Colombia, pues mide el porcentaje de jóvenes que ingresan a las instituciones de educación superior. A lo largo de la últimas décadas ha existido preocupación por parte de las entidades gubernamentales por incrementar dicho índice, y es así como en los últimos veinte años se ha presentado un incremento razonable, sin embargo, no es el esperado, pues aún existe un considerable porcentaje de jóvenes que no logran ingresar a cursar carreras, ya sea técnicas, tecnológicas o profesionales, una vez finalizan su educación básica secundaria y media.

Tal y como afirman (Melo, Ramos, & Hernández, 2014), en Colombia “durante los últimos 20 años, el acceso a la educación superior aumentó a un mayor ritmo, lo que se tradujo en una tasa de cobertura bruta de 24,0% en el año 2000 y de 42,4% al final de 2012. Durante este período, la población matriculada a nivel de pregrado ascendió de 487.448 estudiantes en 1990 a 1.841.282 en 2012, lo que significó una ampliación de los cupos cercana a 278%”. Gran parte de ese incremento en la tasa de cobertura se debió a la prioridad que se le dio a la formación técnica y tecnológica, especialmente a finales de la primera década y comienzos de la segunda década del presente siglo.

Si bien se ha presentado una mejoría en las cifras nacionales con respecto a este indicador a lo largo de los últimos años, la realidad de las regiones es muy heterogénea, presentándose departamentos, como el caso de Casanare, en donde las cifras son muy inferiores a los índices nacionales. Por otra parte, a pesar del incremento en la cobertura, al comparar las cifras del país con las de otros países se evidencia que aún falta mucha inversión y gestión por parte del gobierno para igualar las cifras internacionales. Según afirman (Melo, Ramos, & Hernández, 2014) “la tasa de cobertura de Colombia es relativamente baja cuando se compara con países desarrollados, como Estados Unidos, Finlandia, España, Nueva Zelanda, Australia, Noruega y con un grupo de países latinoamericanos como Argentina, Chile, Cuba, Uruguay y Puerto Rico, cuyas tasas de cobertura superan el 60%”.

En el caso específico del departamento de Casanare, la tasa de cobertura ha presentado un incremento considerable en las dos últimas décadas, mientras que en el año 2001 apenas se presentaba una cobertura del 3,27%, en el año 2017 se llegó a tener una cobertura del 27%, que si bien no es una cifra ideal, sí constituye un incremento significativo en esta región. El aumento de la cobertura de educación superior en dicho departamento se vio potenciado principalmente a partir del año 2007, en donde la cifra prácticamente se duplicó respecto del año anterior, lo que probablemente tenga que ver con el impulso que le dieron el gobierno nacional y departamental a las carreras técnicas y tecnológicas ofertadas por el SENA, así como también con la apertura de nuevos programas y mejoras de infraestructura en algunas universidades de la región.

Algunos factores que pueden influir directamente en la decisión de los jóvenes por desarrollar sus estudios superiores en el departamento de Casanare tienen que ver con la capacidad económica y el hecho de que no sean ofertadas algunas carreras en la región, lo que los obliga a emigrar hacia otros departamentos del país. Entre quienes presentan obstáculos económicos para poder estudiar, se encuentran aquellos que deciden no acceder al sistema educativo y aquellos que deciden no emigrar y matricularse en las instituciones de la región para disminuir costos.

Por su parte, existe otro grupo de personas que deciden emigrar del departamento buscando acceder a programas que no son ofertados en su región. Dicha emigración de jóvenes ha contribuido a que la cobertura educativa en las regiones del país sea muy desigual, según afirman (Ospina Londoño, Canavire Bacarreza, Bohórquez, & Cuartas, 2015) “la dinámica económica y social de las principales ciudades ha resultado en un crecimiento heterogéneo de la cobertura educativa del país. Así, la oferta de educación superior se ha concentrado históricamente en las ciudades de Bogotá, Medellín, Cali y Barranquilla, lo que las ha convertido en polos de atracción de estudiantes, que se han visto forzados a migrar asumiendo los costos económicos de esta decisión y creando asignaciones ineficientes en el acceso a la educación superior”.

El número de jóvenes que deciden no estudiar, como aquellos que deciden emigrar, influye de manera directa en la baja tasa de cobertura de educación superior del departamento de Casanare, en comparación con las cifras de otras regiones. Para corregir tal situación es necesario la reconducción de las políticas educativas, tal como afirman (Ospina Londoño, Canavire Bacarreza, Bohórquez, & Cuartas, 2015) “la migración puede interpretarse como resultado de una asignación ineficiente de la oferta en programas de educación superior. Es decir, si un individuo migra para acceder a programas de educación superior, debe ser porque en su lugar de origen no tiene la oferta que está buscando. Esto puede interpretarse como una falla de las políticas de expansión y regionalización de la educación superior de un país, ya que no se cumpliría el objetivo del desarrollo de los mercados laborales locales”.

Adicionalmente, existen otros factores que también influyen en la decisión de estudiar o no en la región de origen, tal y como afirman (Ospina Londoño, Canavire Bacarreza, Bohórquez, & Cuartas, 2015) “los determinantes de acceder a la educación superior y de migrar o no dependen no sólo de la existencia de oferta de programas en el sitio de origen, sino también de la calidad y tipo de programas ofertados y las preferencias por una IES en particular”. Por otra parte, según afirma Flint (1992) citado por (Ospina Londoño, Canavire Bacarreza, Bohórquez, & Cuartas, 2015), el entorno familiar también ha sido considerado en algunos estudios, los padres de familia desempeñan un papel importante en la elección del programa académico, pues ellos consideran variables como el prestigio, la proximidad al hogar y los costos. Para el caso del departamento de Casanare, la mayoría de jóvenes que emigran prefieren estudiar en universidades de las principales ciudades del país, por ejemplo Bogotá o Medellín, dada la calidad de las universidades y su amplia oferta. De igual forma, un buen número de jóvenes han emigrado hacia el departamento de Boyacá dada la cercanía y calidad de algunas de sus universidades, por ejemplo la UPTC y la UNIBOYACÁ.

## **Series de tiempo**

De acuerdo con Villavicencio (s.f) “una serie de tiempo es una secuencia de observaciones, medidas en determinados momentos del tiempo, ordenados cronológicamente y, espaciados entre sí de manera uniforme, así los datos usualmente son dependientes entre sí”. Su finalidad principal es la generación de pronósticos respecto de la variable en cuestión, según Giraldo (2006) “pueden mencionarse varios motivos por los cuales es útil analizar las series de tiempo. Por ejemplo, para obtener pronósticos de valores futuros de la serie, con el fin de ayudar a tomar decisiones que tienen consecuencias importantes. O para entender mejor el mecanismo de generación de los datos, que puede no ser claro inicialmente en una investigación”.

## **Componentes de una serie de tiempo**

En una serie de tiempo se pueden identificar tres componentes principales:

### **Tendencia**

Intuitivamente, la tendencia “se puede definir como un cambio a largo plazo que se produce en relación al nivel medio, o el cambio a largo plazo de la media” (Villavicencio, Introducción a Series de Tiempo). Desde un punto de vista más formal, (Giraldo, 2006) define la tendencia como “una función  $T_t$  de  $t$  que describe la evolución lenta y a largo plazo del nivel medio de la serie”. Es decir, la tendencia se puede identificar como el cambio que se presenta en el tiempo, ya sea lineal, cuadrático, cúbico, exponencial o logístico, de la línea media de los datos, que puede ser creciente o decreciente.

### **Componente estacional**

Se refiere a la variación periódica (trimestral, semestral, mensual, anual) que presentan ciertos fenómenos, lo que hace que se presente un comportamiento similar entre período y periodo, y que la mayoría de las veces se puede identificar gráficamente pues se observa el mismo patrón en cada uno de los intervalos de tiempo medidos. Para el caso de este proyecto, el componente estacional se puede identificar en un incremento de la tasa de cobertura de educación superior durante el primer semestre de cada año y un descenso en el segundo semestre. Lo anterior debido a que es mayor el número de jóvenes que se matriculan en el primer semestre que en el segundo.

### **Componente aleatorio**

Como su nombre lo indica, esta componente se debe al azar que interviene en todos los procesos estocásticos, es decir que “...no responde a ningún patrón de comportamiento, sino que es el resultado de factores fortuitos o aleatorios que inciden de forma aislada en una serie de tiempo” (Villavicencio, Introducción a Series de Tiempo).

## **Clasificación de las series de tiempo**

Por otra parte, Las series de tiempo se clasifican en dos grandes grupos:

### **Series estacionarias**

Se dice que una serie es estacionaria cuando presenta media y varianza constantes en intervalos de tiempo de la misma amplitud. En algunas ocasiones la estacionariedad se puede identificar mediante la observación de la gráfica de la serie, según (Forero, Series de tiempo I, 2017) una serie se podría considerar estacionaria si “la gráfica no presenta picos que sobresalgan demasiado, un aumento perceptible en la varianza, o donde pueda percibirse un cambio en la tendencia de la serie. Así mismo, es necesario verificar que exista una cantidad razonablemente parecida de picos en intervalos de igual tamaño”. Sin embargo, para tener evidencia estadística confiable que permita determinar si una serie es estacionaria se requiere aplicar algunas pruebas de raíz unitaria, como la prueba de Dickey-Fuller, sobre la cual se hará referencia más adelante.

### **Series no estacionarias**

Son aquellas series que muestran tener una tendencia, ya sea creciente o decreciente, por lo que la media y la varianza no son constantes en el tiempo. En estas series “los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante” (Villavicencio, Introducción a Series de Tiempo). La no estacionariedad se puede establecer, por ejemplo, cuando no se rechaza la hipótesis nula en una prueba de raíz unitaria como la de Dickey-Fuller.

## **Proceso estacional, autorregresivo, integrado y de media móvil SARIMA (p,d,q) × (P,D,Q)<sub>s</sub>**

Se habla de un modelo SARIMA “...cuando una serie de tiempo en estudio tiene intervalos de observación menores a un año, entonces es frecuente que estas tengan variaciones o patrones sistemáticos cada cierto periodo. Estas variaciones sistemáticas inferiores a un año, por ejemplo semestral, mensual, diario, etc., deben ser captadas en los llamados factores estacionales, dentro de la estructura del modelo a construirse” (Villavicencio, Introducción a Series de Tiempo). Para el caso de este proyecto, como se mencionaba anteriormente, se observa que la serie tiene picos altos el primer semestre de cada año y picos bajos durante el segundo semestre, presentando características de un modelo SARIMA.

Según este Forero (2017), una manera sencilla de entender el modelo SARIMA es “considerar un modelo ARIMA en el que cada dato es un periodo de estacionalidad, es decir, otro modelo ARIMA”. Según este mismo autor, el modelo SARIMA tiene la siguiente notación:

## **SARIMA (p, d, q) × (P, D, Q)<sub>s</sub>**

Donde los parámetros p, d, q minúsculos corresponden a la componente no estacional, los parámetros P, D, Q mayúsculos corresponden a la componente estacional y s es el número de períodos por estación.

De igual forma, cabe destacar que los parámetros p y P corresponden a la parte autorregresiva del modelo, los parámetros d y D a la parte integrada y los parámetros q y Q a la parte de medias móviles, tanto de la componente no estacional como de la estacional respectivamente.

### **Pasos a seguir para la identificación de un modelo SARIMA adecuado**

Villavicencio (s.f) sugiere el siguiente protocolo para identificar el modelo SARIMA más adecuado que permita describir el comportamiento de un conjunto de datos:

#### **Identificación**

El objetivo principal en este paso es identificar si la serie es estacionaria o no. Un primer indicio lo da la gráfica de los datos, si se observan media y varianza constantes en el tiempo se puede sospechar que la serie es estacionaria. Sin embargo, la mejor manera de tener evidencia estadística respecto de la estacionariedad de la serie es aplicando una prueba de raíz unitaria al conjunto de datos. En este proyecto se utilizará la prueba de Dickey Fuller para tal fin. La misma consiste en una prueba de hipótesis estadística en donde la hipótesis nula  $H_0$  indica que la serie es integrada de orden 1 (no estacionaria) y la hipótesis alternativa indica que la serie es integrada de orden 0 (estacionaria).

Prueba de Dickey Fuller:

$$\begin{cases} H_0: y_t \sim I(1) \\ H_1: y_t \sim I(0) \end{cases}$$

Para aplicar esta prueba en RStudio se puede utilizar el comando `adf.test` del paquete `Tseries`, el cual arroja como resultado el p-valor de la prueba. En ese caso el criterio de decisión es el siguiente: se rechaza la hipótesis nula si el p-valor es menor que el nivel de significancia (generalmente de 0,01 o 0,05), de lo contrario no se rechaza la hipótesis nula y se concluye que la serie no es estacionaria.

En caso que la serie resulte no estacionaria, se debe diferenciar y verificar nuevamente a través del test de Dickey Fuller que la hipótesis nula sea rechazada. El mismo procedimiento se debe repetir hasta verificar que se cumple la estacionariedad.

#### **Estimación**

Según Villavicencio (s.f) se debe tener en cuenta que “observando las dos gráficas del ACF y

PACF de la serie transformada podemos hacernos una idea del modelo que describe nuestra serie, o al menos de cuáles son los candidatos que debemos probar”. A partir de los indicios de dichas gráficas se pueden ajustar varios modelos SARIMA realizando variaciones en cada uno de los parámetros. Una vez identificados los modelos que se ajusten mejor a la gráfica de los datos, se comparan los valores en los criterios AIC y BIC, siendo un buen modelo aquel que tenga los valores más cercanos a cero en dichos criterios.

## **Validación**

En esta etapa se deben analizar algunas condiciones relacionadas con los errores del modelo frente a los datos originales, según (Forero, Series de tiempo II, 2017) “la etapa de verificación de supuestos pasa por verificar ciertas condiciones sobre tales residuos: errores no autocorrelacionados, errores con distribución normal y errores independientes”. De igual forma, se puede verificar gráficamente si el modelo se ajusta en un alto porcentaje a la serie de tiempo original. Si el modelo cumple las anteriores condiciones se puede utilizar en la siguiente fase del proceso. De lo contrario hay que regresar de nuevo a la etapa de estimación y repetir el proceso.

## **Predicción**

Una vez se ha seleccionado el modelo que mejor se ajusta al comportamiento de los datos, se procede a realizar el pronóstico de la variable en cuestión para determinado lapso de tiempo. En RStudio se puede utilizar el comando `sarima.for` del paquete `forecast`, en donde el número de períodos a estimar se define con el argumento `n.ahead`. El software generará una gráfica con la estimación puntual de la predicción y dos bandas de confianza, una del 80% y otra del 95%, que aparecen por defecto.

## Capítulo 4. Marco Metodológico

### Tipo de investigación

Para resolver el problema y cumplir con los objetivos trazados se determinó que el tipo de investigación a utilizar es de carácter mixto, ya que realizar un análisis de la cobertura de educación superior por medio de las series de tiempo se combinan los métodos: exploratorio y descriptivo, lo que permitió recabar los datos, describirlos en el tiempo y analizarlos.

### Investigación exploratoria

Este proyecto encaja dentro de la investigación exploratoria puesto que se va a abordar una temática poco estudiada. En este caso, se trata de la proyección de la cobertura de la educación superior en el departamento de Casanare. Se utiliza la investigación exploratoria “cuando la revisión de la literatura reveló que únicamente hay guías no investigadas e ideas vagamente relacionadas con el problema de estudio” (Hernández Sampieri, Fernández-Collado, & Baptista Lucio, Metodología de la investigación). Al respecto, una vez realizada la revisión de la literatura no se encontraron estudios que realicen pronósticos, simplemente se encontraron documentos que describen el comportamiento de dicha variable durante pequeños lapsos de tiempo en el pasado. Inclusive, no se encontraron estudios que utilicen la metodología Box Jenkins para realizar proyecciones sobre este tipo de indicadores de educación superior en Colombia.

### Investigación descriptiva

De igual forma, este proyecto también se enmarca dentro de la investigación descriptiva, puesto que se pretende medir con la mayor precisión posible, y mediante técnicas estadísticas, el comportamiento que tendrá a futuro la variable cobertura de educación superior en el departamento de Casanare, sin establecer relaciones de dicha variable con cualquiera otra, es decir, sin abordar las correlaciones existentes con variables que podrían, quizás, influir en el comportamiento de la misma, dicho de otra forma “en un estudio descriptivo se selecciona una serie de cuestiones y se mide cada una de ellas independientemente, para así – y valga la redundancia – describir lo que se investiga.” (Hernández Sampieri, Fernández - Collado, & Baptista Lucio, Metodología de la investigación)

### Diseño metodológico

#### Tipo de estudio

Este estudio corresponde a una predicción estadística utilizando la serie de tiempo de la tasa de cobertura de educación superior en el departamento de Casanare, para lo cual se aplicará

la metodología propuesta por Box y Jenkins.

## **Área de estudio**

El área de estudio de este proyecto corresponde al sector de educación superior en el departamento de Casanare, utilizando una variable de especial importancia como lo es la tasa de cobertura, la cual permite medir la efectividad de las políticas públicas que buscan la inclusión de los jóvenes en la formación técnica, tecnológica y profesional, y de esa forma ampliar su abanico de oportunidades de desempeño a nivel laboral.

## **Universo y muestra**

El universo de este estudio corresponde al total de jóvenes entre 17 y 21 años que habitan el departamento de Casanare.

Para efectos de este estudio no se selecciona una muestra, sino que se tiene en cuenta el total de la población, sin embargo, se debe aclarar que el análisis de la serie de tiempo está limitado para el período comprendido entre el año 2001 y el año 2017, pues es la época de la cual se dispone de información concreta respecto de la variable en cuestión por parte del Sistema Nacional de Información de la Educación Superior SNIES.

## **Métodos e instrumentos de recolección de datos**

Para el cálculo de la tasa de cobertura de educación superior en el departamento de Casanare a lo largo de los últimos años, se tuvo en cuenta la información disponible en el SNIES y en la página del DANE. Respecto a la primera, se consideró la información referente al número de estudiantes matriculados en cada uno de los semestres a partir del año 2001 y hasta el año 2017. Respecto a la segunda fuente de información se tuvieron en cuenta las proyecciones de población del departamento de Casanare, entre 17 y 21 años, para el mismo período de tiempo. La tasa de cobertura corresponde al cociente entre las variables mencionadas y multiplicando el resultado por cien.

## **Plan de tabulación y análisis**

Una vez conformada la base de datos, se procede a realizar su tratamiento estadístico, utilizando el software RStudio, bajo el enfoque o metodología Box Jenkins. De esta forma se pueden obtener varios modelos SARIMA que describan el comportamiento de la variable en cuestión y seleccionar aquel que se ajuste mejor, para posteriormente generar pronósticos confiables sobre el comportamiento de la tasa de cobertura en el corto plazo.

## **Enfoque**

El enfoque de este proyecto presenta en su mayoría características cuantitativas, puesto que la variable principal, tasa de cobertura en educación superior, se mide en escala de razón. De igual forma, el tratamiento estadístico que se le realiza a los datos garantiza tal hecho. Sin embargo, no se puede desconocer que alrededor de dicha variable, y para efectos de este



trabajo, se tienen en cuenta aspectos subjetivos que buscan explicar, desde el punto de vista del autor, la relación e influencia de factores exógenos en el comportamiento de la variable. Por tal motivo, existen aspectos cualitativos que también son considerados. Luego se puede afirmar que el enfoque del proyecto es mixto, puesto que combina tanto el enfoque cualitativo como el cuantitativo.

## Capítulo 5. Análisis y Resultados

### Construcción de la base de datos

Para construir la base de datos de la cobertura de educación superior en Casanare se tomaron los datos sobre el número de estudiantes matriculados por semestre en las instituciones de educación superior del departamento, información disponible en el Sistema Nacional de Información de la Educación Superior (SNIES), para el período comprendido entre los años 2001 y 2017, pues es el intervalo de tiempo del cual se dispone de información completa al respecto. De igual forma se consideraron las proyecciones de población del departamento, entre 17 y 21 años, que se encuentra disponible en la página del DANE. La tasa de cobertura (TC) se calcula mediante la siguiente fórmula:

$$TC = \frac{\text{Población efectivamente matriculada en educación superior en el semestre } n}{\text{Población entre 17 y 21 años en el semestre } n} \times 100$$

Para este proyecto se tuvo en cuenta que existe una variación en el número de estudiantes matriculados entre el primer y segundo semestre de cada año, siendo mayor el número de matriculados en primer semestre, por lo cual se realizó el cálculo de la tasa de cobertura con una periodicidad semestral y no anual.

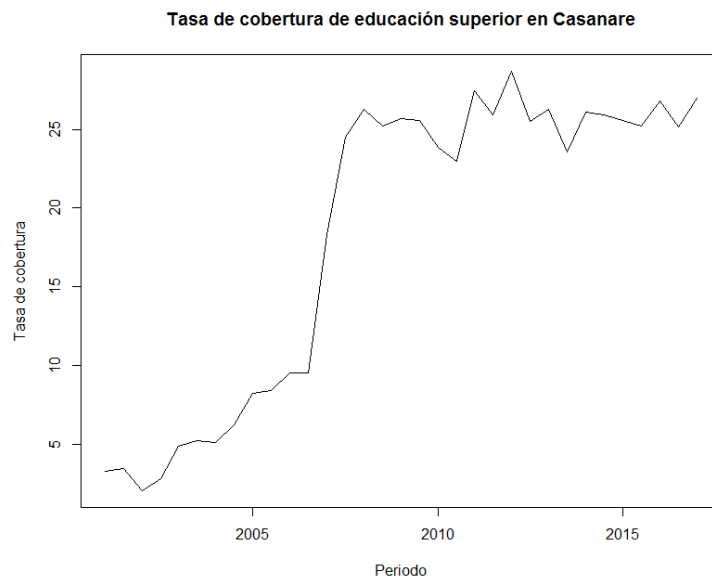
Una vez calculado el indicador en cuestión, se conformó la base de datos que sirve de insumo para el análisis estadístico. A continuación se muestran los datos obtenidos:

SEMESTRE	MATRICULADOS	POBLACIÓN 17-21 AÑOS	TASA DE COBERTURA
2001-1	821	25069	3,27%
2001-2	860	25069	3,43%
2002-1	527	25679	2,05%
2002-2	713	25679	2,78%
2003-1	1277	26401	4,84%
2003-2	1369	26401	5,19%
2004-1	1381	27206	5,08%
2004-2	1697	27206	6,24%
2005-1	2337	28542	8,19%
2005-2	2395	28542	8,39%
2006-1	2839	29751	9,54%
2006-2	2825	29751	9,50%
2007-1	5674	30959	18,33%
2007-2	7585	30959	24,50%
2008-1	8441	32117	26,28%

2008-2	8107	32117	25,24%
2009-1	8516	33180	25,67%
2009-2	8481	33180	25,56%
2010-1	8120	34051	23,85%
2010-2	7830	34051	22,99%
2011-1	9527	34696	27,46%
2011-2	8987	34696	25,90%
2012-1	10075	35134	28,68%
2012-2	8968	35134	25,53%
2013-1	9304	35382	26,30%
2013-2	8329	35382	23,54%
2014-1	9263	35531	26,07%
2014-2	9204	35531	25,90%
2015-1	9116	35643	25,58%
2015-2	8988	35643	25,22%
2016-1	9598	35772	26,83%
2016-2	9009	35772	25,18%
2017-1	9708	35959	27,00%

*Tabla 1. Cálculo de la tasa de cobertura de educación superior de Casanare basado en estadísticas del SNIES y del DANE.*

La gráfica de la serie de tiempo generada en RStudio es la siguiente:



*Gráfica 1 Tasa de cobertura de educación superior de Casanare*

Como se observa, a simple vista se podría inferir que la serie no es estacionaria, sin embargo,

se aplica la prueba de Dickey-Fuller para corroborar este hecho. El resultado arrojado por RStudio es el siguiente:

```
> adf.test(datos)

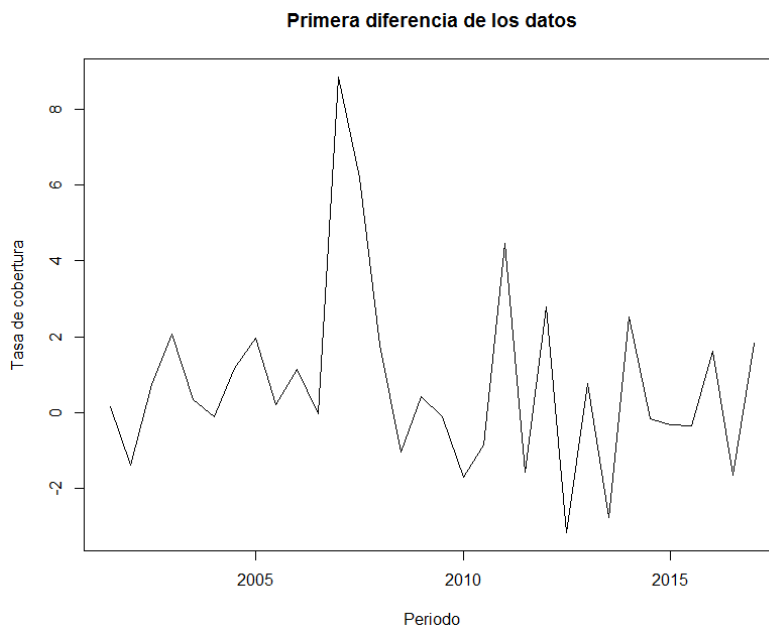
Augmented Dickey-Fuller Test

data: datos
Dickey-Fuller = -1.0436, Lag order = 3, p-value = 0.9165
alternative hypothesis: stationary
```

Como se observa, el p-valor es mayor que cualquier nivel de significancia, por lo que en este caso no se rechaza la hipótesis nula y se concluye que no existe evidencia estadísticamente significativa para afirmar que la serie es estacionaria. Por tal motivo, la serie debe ser diferenciada y aplicarse de nuevo el test de raíz unitaria.

### **Análisis de los datos bajo en el enfoque Box-Jenkins**

La gráfica de la serie diferenciada se muestra a continuación:



*Gráfica 2. Primera diferencia de la serie de datos*

Al aplicar nuevamente el test de Dickey-Fuller se obtiene el siguiente resultado:

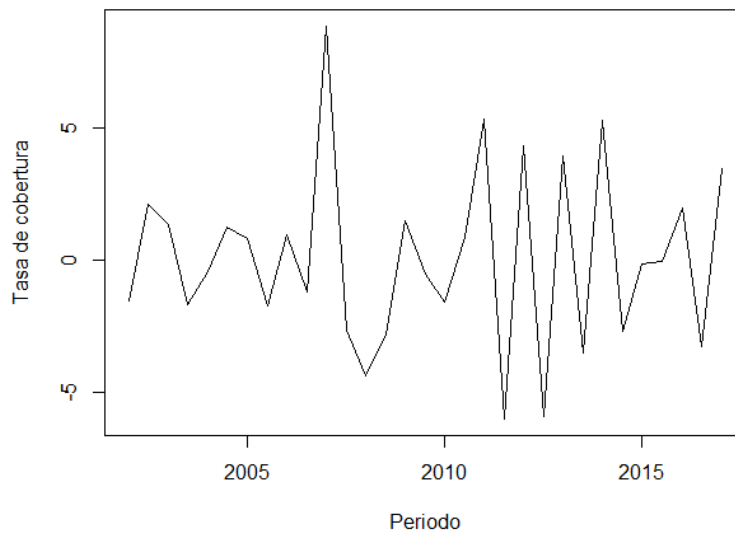
```
> adf.test(datos.diff)
```

Augmented Dickey-Fuller Test

```
data: datos.diff  
Dickey-Fuller = -2.3808, Lag order = 3, p-value = 0.4259  
alternative hypothesis: stationary
```

Como se observa, el p-valor sigue siendo mayor que cualquier nivel de significancia, luego se concluye que la serie diferenciada no es estacionaria. En este caso, se requiere diferenciar nuevamente la serie y verificar mediante la prueba de raíz unitaria otra vez. La serie diferenciada por segunda vez se muestra a continuación:

**Segunda diferencia de los datos**



*Gráfica 3. Segunda diferencia de la serie de datos*

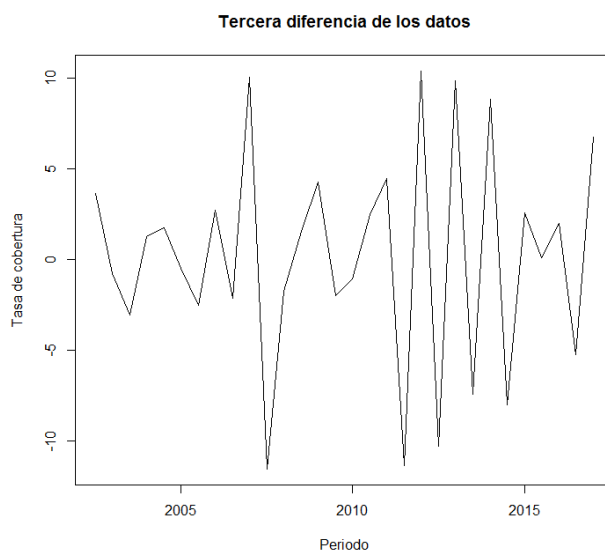
El resultado de la prueba de Dickey-Fuller es el siguiente:

```
> adf.test(datos.diff2)
```

Augmented Dickey-Fuller Test

```
data: datos.diff2  
Dickey-Fuller = -3.4471, Lag order = 3, p-value = 0.06887  
alternative hypothesis: stationary
```

En este caso el p-valor sigue siendo superior al cualquier nivel de significancia aceptable, luego se debe repetir el procedimiento nuevamente. La serie diferenciada por tercera vez se muestra a continuación:



Gráfica 4. Tercera diferencia de la serie de datos

El resultado del test de Dickey-Fuller para este caso es el siguiente:

```
> adf.test(datos.diff3)

Augmented Dickey-Fuller Test

data: datos.diff3
Dickey-Fuller = -4.7154, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

Como se observa, el p-valor es menor que cualquier nivel de significancia, luego existe evidencia estadísticamente significativa para afirmar que la tercera diferencia de la serie de datos es estacionaria.

El siguiente paso en el proceso es generar varios modelos SARIMA que expliquen el comportamiento de los datos y seleccionar aquel que se ajuste mejor según los criterios AIC y BIC.

## Obtención de modelos SARIMA

A través de la función `sarima` en RStudio se realizó el ajuste de varios modelos que expliquen el comportamiento de los datos de la variable tasa de cobertura de educación superior en Casanare. A continuación se presentan los modelos ajustados y sus respectivos valores para los criterios AIC, AICc y BIC.

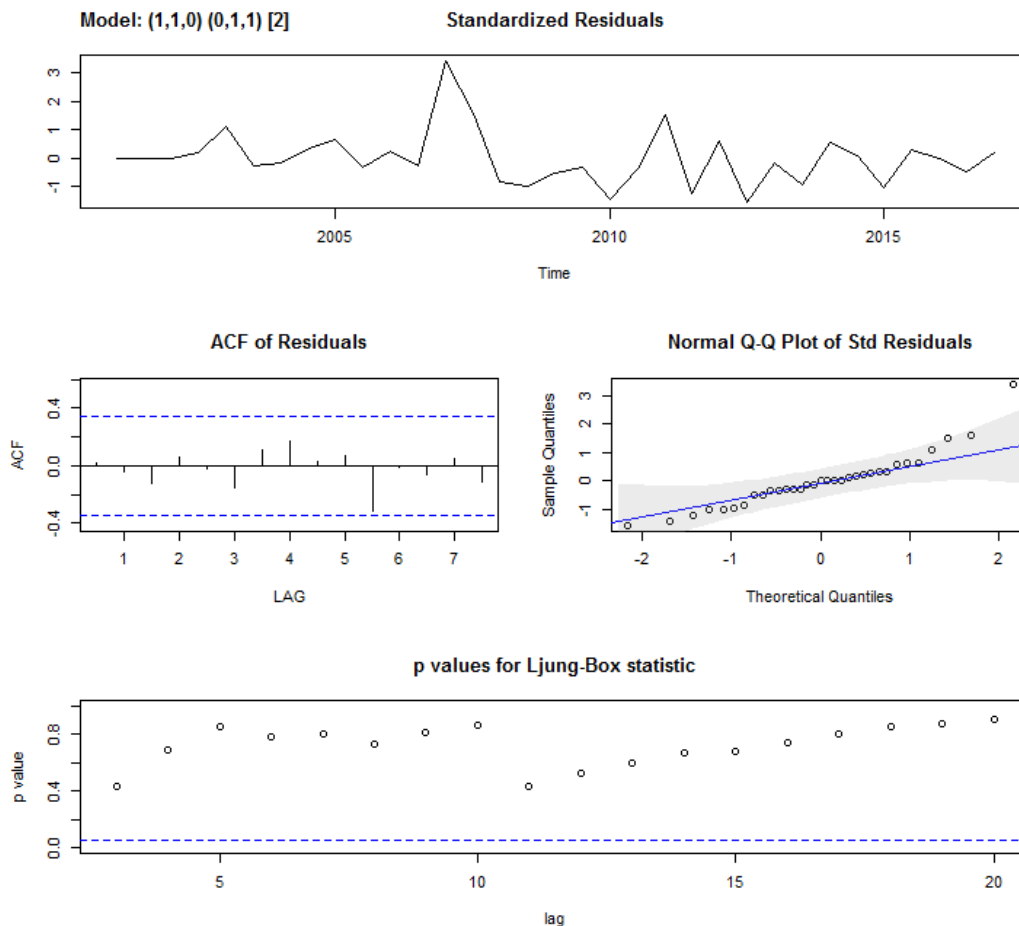
MODELO	AIC	AICc	BIC
SARIMA (1,1,1)× (1,1,1) <sub>2</sub>	2,9247	3,0527	2,1061
SARIMA (1,1,0)× (1,0,1) <sub>2</sub>	2,8949	3,0229	2,0763
SARIMA (1,2,0)× (1,0,1) <sub>2</sub>	2,9073	3,0112	2,0433
SARIMA (2,2,0)× (1,0,1) <sub>2</sub>	2,9176	3,0455	2,0989
SARIMA (1,1,0)× (0,1,1) <sub>2</sub>	2,8021	2,8878	1,8928
SARIMA (1,1,0)× (1,1,1) <sub>2</sub>	2,8647	2,9686	2,0007

Tabla 2. Modelos SARIMA ajustados

Teniendo en cuenta lo anterior, el modelo más adecuado entre los considerados es un **SARIMA (1,1,0)×(0,1,1)<sub>2</sub>** pues el que presenta los valores más bajos en los criterios AIC, AICc y BIC.

## Verificación de supuestos

A continuación se muestran algunas gráficas de interés sobre dicho modelo:



Gráfica 5. Gráficas de los errores de la serie ajustada

Como se observa, la gráfica de los residuos muestra tener aproximadamente media y varianza constantes, el ACF sugiere que no existe autocorrelación serial pues los rezagos se encuentran dentro del intervalo deseado, los puntos en el gráfico Q-Q Plot se encuentran en la banda de confianza considerando el intervalo  $[-2, 2]$ , y los p-valores para el estadístico Ljung Box son mayores al nivel de significancia, lo que sugiere que no existe autocorrelación serial en los errores.

## Predicción de la tasa de cobertura de educación superior del departamento de Casanare

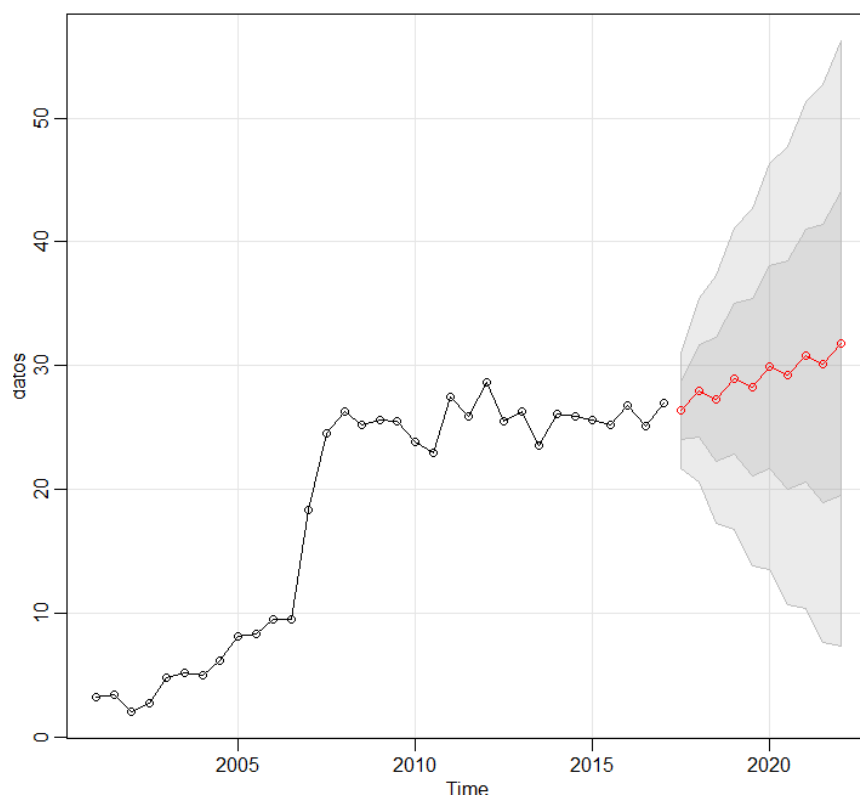
Una vez ajustado el modelo que describe con un alto grado de confianza el comportamiento de la variable en cuestión, y de verificar que se cumplen ciertos supuestos de normalidad en los residuos, así como ausencia de autocorrelación serial, se procede a realizar el pronóstico de la tasa de cobertura para los próximos diez semestres. En este caso, se estimaron los valores correspondientes al segundo semestre de 2017 y los del año 2018 puesto que el SNIES aún no había liberado los datos correspondientes a dichos períodos para el momento en que se desarrolló el proyecto. Para realizar el pronóstico se utilizó la función `sarima.for` del paquete `astsa` en RStudio. Los resultados son los siguientes:

```
> sarima.for(datos,1,1,0,0,1,1,2,n.ahead = 10)
$pred
Time series:
Start = c(2017, 2)
End = c(2022, 1)
Frequency = 2
 [1] 26.36787 28.00722 27.33076 28.95923 28.28010 29.90792 29.22863 30.85640
 [9] 30.17710 31.80487

$se
Time series:
Start = c(2017, 2)
End = c(2022, 1)
Frequency = 2
 [1] 2.319815 3.704916 5.000725 6.088766 7.193869 8.191932 9.237545 10.212508
 [9] 11.243854 12.221483
```

La gráfica que muestra las predicciones con intervalos de confianza del 80% y del 95% es la siguiente:





Gráfica 6. Predicciones de la tasa de cobertura de educación superior de Casanare

La siguiente tabla presenta de una forma más clara los pronósticos obtenidos, redondeados a dos cifras decimales, con un nivel de confianza del 80%:

SEMESTRE	Estimación puntual	Intervalo de confianza del 80%	
		Límite inferior	Límite superior
2017-2	26,37%	24,05%	28,69%
2018-1	28,00%	24,30%	31,71%
2018-2	27,33%	22,33%	32,33%
2019-1	28,96%	22,87%	35,05%
2019-2	28,28%	21,09%	35,47%
2020-1	29,91%	21,72%	38,10%
2020-2	29,23%	19,99%	38,47%
2021-1	30,86%	20,64%	41,07%
2021-2	30,18%	18,93%	41,42%
2022-1	31,30%	19,58%	44,03%

Tabla 3. Pronósticos de la tasa de cobertura de educación superior de Casanare con un 80% de confianza

## Capítulo 6. Conclusiones

- Teniendo en cuenta los modelos SARIMA que se ajustaron a los datos de la tasa de cobertura de educación superior del departamento de Casanare, se puede afirmar que el modelo que mejor explica el comportamiento de dichos datos es un SARIMA  $(1,1,0) \times (0,1,1)_2$ , pues fue el que presentó valores más cercanos al cero en los criterios AIC, AICc y BIC. Además, su validación resultó coherente pues se evidenció la ausencia de autocorrelación serial en los errores.
- La estimación puntual de la tasa de cobertura de educación superior del departamento de Casanare, para el período comprendido entre el primer semestre de 2019 y el primer semestre de 2022, permite concluir que dicho indicador educativo tendrá una leve tendencia creciente en el tiempo, logrando sobrepasar el 30% de cobertura desde el año 2021 de acuerdo con la estimación puntual, sin embargo, se mantienen las fluctuaciones de los datos, presentándose un incremento durante el primer semestre de cada año y un descenso en el segundo semestre.
- De acuerdo con los pronósticos realizados a través del modelo SARIMA ajustado se puede concluir que la tasa de cobertura de educación superior del departamento de Casanare continuará presentando cifras inferiores a las de otras regiones del país y a la tasa de cobertura nacional, al menos en el período comprendido entre los años 2019 y 2022.
- El incremento de la tasa de cobertura de educación superior del departamento de Casanare en los próximos años, puede hacer suponer que factores como la disminución de la emigración de los jóvenes hacia otros departamentos y la ampliación de la oferta educativa, sean factores que favorezcan dicha tendencia, que aunque pequeña, es positiva.
- La aplicación de la metodología Box Jenkins en la predicción de indicadores de educación superior, que se miden de manera semestral u anual, resulta confiable únicamente en el corto plazo, pues en la medida que se aumenta el margen de pronóstico, se hace más amplio el intervalo de confianza donde se supone que se encontrará el estadístico estimado con cierto nivel de confianza.

## **Recomendaciones**

- Para tener una mayor precisión en las estimaciones e intervalos de confianza más estrechos se sugiere contar con el mayor número de datos posibles. Para el caso de este proyecto se tuvo como limitante el hecho de que las estadísticas sobre educación superior, que han sido liberadas por el SNIES, se encuentran a partir del año 2.000. Se evidencia la importancia de llevar de una manera organizada y metódica los datos de cualquier entidad, institución o empresa a lo largo del tiempo.
- Para este tipo de estudios también se podría tener en cuenta la influencia de variables exógenas en el comportamiento de los datos, para lo cual sería ideal ajustar modelos SARIMAX. En el caso de este estudio se pudieron haber tenido en cuenta variables como el estrato socioeconómico, el comportamiento del precio del petróleo (sector industrial de especial incidencia en la economía de Casanare), la calidad de las carreras y de las instituciones educativas que las ofrecen, entre otras.

## Bibliografía

- Avila, C. (2008). *La dinámica de acumulación de capital humano en el departamento de Casanare (Yopal)*. Bogotá.
- Forero, S. L. (2017). *Series de tiempo I*. Bogotá: Fundación Universitaria Los Libertadores.
- Forero, S. L. (2017). *Series de tiempo II*. Bogotá: Fundación Universitaria Los Libertadores.
- Giraldo, N. (2006). *Notas de Clase. Series de tiempo con R*. Medellín: Universidad Nacional de Colombia.
- Hernández Sampieri , R., Fernández-Collado, C., & Baptista Lucio, P. (s.f.). *Metodología de la investigación*. Mc Graw Hill.
- Melo, L. A., Ramos, J. E., & Hernández, P. O. (2014). La educación superior en Colombia: situación actual y análisis de eficiencia. *Borradores de economía*, 10, 12.
- Ospina Londoño, M., Canavire Bacarreza, G., Bohórquez, S., & Cuartas , D. (2015). Expansión de la educación superior y sus efectos en matriculación y migración: evidencia de Colombia. *Desarrollo y Sociedad*, 317 - 318.
- Rama, C. (2010). La tendencia a la masificación de la cobertura de la educación superior en América Latina. *Revista Iberoamericana de Educación a Distancia*, 177, 192 y 193.
- Villavicencio, J. (s.f.). *Introducción a Series de Tiempo*. Obtenido de [http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4\\_BxecUaZmg%3D](http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D)

## Apéndice A. Código en R

```
#Carga de librerías

library(forecast)
library(tseries)
library(urca)
library(ggplot2)
library(stats)
library(astsa)

#Lectura de la base de datos y conversión a serie de tiempo

datos<-read.table("tasa de cobertura casanare.txt",header = FALSE)
datos<-ts(datos,start=c(2001,1),frequency = 2)

#Gráfica de los datos

plot(datos,main="Tasa de cobertura de educación superior en
Casanare",xlab="Periodo",ylab="Tasa de cobertura")

#Prueba de Dickey Fuller para los datos

adf.test(datos)

#Primera diferencia de los datos

datos.diff<-diff(datos)

#Prueba de Dickey Fuller para la primera diferencia de la serie

adf.test(datos.diff)

#Gráfica de la primera diferencia

plot(datos.diff,main="Primera diferencia de los datos",xlab="Periodo",ylab="Tasa de
cobertura")

#Segunda diferencia de los datos

datos.diff2<-diff(datos.diff)
```

```

#Gráfica de la segunda diferencia
plot(datos.diff2,main="Segunda diferencia de los datos",xlab="Periodo",ylab="Tasa de
cobertura")

#Prueba de Dickey Fuller para la segunda diferencia
adf.test(datos.diff2)

#Tercera diferencia de los datos
datos.diff3<-diff(datos.diff2)

#Gráfica de la tercera diferencia
plot(datos.diff3,main="Tercera diferencia de los datos",xlab="Periodo",ylab="Tasa de
cobertura")

#Prueba de Dickey Fuller para la tercera diferencia
adf.test(datos.diff3)

#Ajuste de los datos a un modelo SARIMA
ajuste1<-sarima(datos,1,1,0,1,1,1,2)

#Pronóstico para los próximos seis semestres
sarima.for(datos,1,1,0,0,1,1,2,n.ahead = 10)

```