UNIVERSIDAD NACIONAL DEL ALTIPLANO

ESCUELA DE POSGRADO

PROGRAMA DE DOCTORADO

DOCTORADO EN CIENCIAS DE LA COMPUTACIÓN

TESIS

ANÁLISIS PREDICTIVO PARA LOS PROCESOS DE ADMISIÓN DE LA UNIVERSIDAD NACIONAL DEL ALTIPLANO – PUNO

PRESENTADA POR:

ADOLFO CARLOS JIMÉNEZ CHURA

PARA OPTAR EL GRADO ACADÉMICO DE:

DOCTORIS SCIENTIAE EN CIENCIAS DE LA COMPUTACIÓN

PUNO, PERÚ

2017

UNIVERSIDAD NACIONAL DEL ALTIPLANO

ESCUELA DE POSGRADO

PROGRAMA DE DOCTORADO

DOCTORADO EN CIENCIAS DE LA COMPUTACIÓN

TESIS

ANÁLISIS PREDICTIVO PARA LOS PROCESOS DE ADMISIÓN DE LA UNIVERSIDAD NACIONAL DEL ALTIPLANO – PUNO

PRESENTADA POR:

ADOLFO CARLOS JIMÉNEZ CHURA

PARA OPTAR EL GRADO ACADÉMICO DE:

DOCTORIS SCIENTIAE EN CIENCIAS DE LA COMPUTACIÓN

APROBADA POR EL SIGUIENTE JURADO:

PRESIDENTE

.....

Dr. BERNABÉ CANQUI FLORES

PRIMER MIEMBRO

.....

Dr. HENRY IVAN CONDORI ALEJO

SEGUNDO MIEMBRO

.....

Dr. ALFREDO PELAYO CALATAYUD MENDOZA

ASESOR DE TESIS

.....

Dr. PERCY HUATA PANCA

Puno, 8 de junio del 2017

ÁREA: Ciencias de la computación
TEMA: Aplicación de la informática en la educación
LÍNEA: Informática y sociedad

DEDICATORIA

A mis seres más queridos,
mis hijos André, Anderson y Andrea.

AGRADECIMIENTOS

- A la Universidad Nacional del Altiplano por la formación tanto a nivel de pregrado y posgrado que he recibido.
- A los miembros del Jurado Dr. Bernabe Canqui Flores, Dr. Henry Ivan Condori Alejo, Dr. Alfredo Pelayo Calatayud Mendoza y al Dr. Percy Huata Panca; por todas sus sugerencias y correcciones en el trabajo de investigación.
- A los docentes que por haber impartido su conocimientos con nosotros los estudiantes.

ÍNDICE GENERAL

ÍNDICE DE CUADROS

ÍNDICE DE FIGURAS

RESUMEN

El presente proyecto de tesis se desarrolló en la ciudad de Puno, entre los años 2016-2017, cuyo objetivo principal es predecir la tendencia de postulantes e ingresantes a las Escuelas Profesionales y su formación en las escuelas de educación secundaria, públicas y privadas, y en función a estos resultados establecer políticas adecuadas en las Escuelas Profesionales de la Universidad Nacional del Altiplano y las escuelas de educación secundaria de la región de Puno. Se usó como referencia la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), para los modelos de predicción se usó el software R y paquetes adicionales como RMySQL, dplyr, ggplot, polynom, entre otros. Estos permitieron procesar, analizar, graficar e interpretar la información acerca de los postulantes e ingresantes en los procesos de admisión general y cepreuna. El resultado obtenido, con los modelos lineales y polinómicos permitieron predecir y confirmar el nivel de crecimiento de las Escuelas Profesionales como Ing. Civil, Ciencias Contables, etc. La escuela de educación secundaria de la Gran Unidad Escolar San Carlos cuenta con una mayor cantidad de postulantes, sin embargo, la mayor cantidad de ingresantes es de la escuela Santa Rosa, lo que indica que sus estudiantes poseen una mejor formación.

Palabras claves: estadística, librerías, predicción, procesos de admisión, tecnología.

ABSTRACT

This thesis was developed in Puno-Perú, between 2016 and 2017, where the main objective is to predict the preference trends and the training level of the applicants and entrants to the university undergrad courses and what public or private high schools comes from. Later, in function of these results to take the appropriate policies for the National University of the Altiplano undergrad courses, and high schools of the region. The CRISP-DM (Cross Industry Standard Process for Data Mining) methodology was used as reference, software R and additional packages such as RMySQL, dplyr, ggplot, polynom, among others were used for the prediction models. These allowed to process, analyze, graph and interpret the information about the applicants and entrants in the processes of general admission and cepreuna. The results obtained with the linear and polynomial models allowed us to predict and confirm the level of growth of the undergrad courses such as Civil Engineering, Accounting Sciences, etc. The high school of the San Carlos High School Unit counts on a greater number of applicants, however, the greater number of entrants is from the Santa Rosa school, which indicates that its students have a better formation.

Keywords: statistics, libreries, prediction, admission processes, technology.

INTRODUCCIÓN

La Universidad Nacional del Altiplano capta gran cantidad de información de postulantes e ingresantes en cada proceso de admisión que lleva a cabo la Comisión Central de Admisión, la base de datos se alimenta de información en cada proceso y esto conlleva a realizar un análisis detallado con las técnicas adecuadas de minería de datos. La minería de datos se aplica en diferentes ámbitos del conocimiento como apoyo a la solución de problemas específicos buscando patrones de interés en una determinada forma de representación.

El objetivo de éste estudio está orientado a predecir la tendencia del futuro de las Escuelas Profesionales y la formación educativa brindada en las escuelas de educación secundaria tanto públicas como privadas de la región de Puno.

El trabajo desarrollado comprende cinco capítulos, desarrollados de la siguiente manera:

El capítulo I, se ha considerado la problemática de la investigación, la cual permitió conocer la situación real en la que se encuentra las Escuelas Profesionales y escuelas de la región de Puno; asimismo, se planteó el problema principal; se indican los objetivos de la investigación, para luego exponer los motivos que justificaron la realización de éste estudio y su importancia; además de las limitaciones.

El capítulo II, se aborda el marco teórico, en el cual se ha desarrollado los antecedentes de la investigación, la metodología CRISP-DM empleada, los paquetes RMySQL, dplyr, ggplot empleados en el desarrollo de la investigación, también se cuenta con la definición de términos básicos empleado en la investigación.

El capítulo III, se planteó la metodología usada, las herramientas que se usan en la investigación y la población de los postulantes de los diferentes procesos de admisión.

El capítulo IV, se desarrollaron todas las etapas de la metodología CRISP-DM, la implementación de las funciones que hacen uso del paquete dplyr, ggplot, polynomic, hasta la presentación de los modelos.

Finalmente, las conclusiones, recomendaciones y bibliografía.

CAPÍTULO I

PROBLEMA DE INVESTIGACIÓN

1.1. PLANTEAMIENTO DEL PROBLEMA

La Universidad Nacional del Altiplano cuenta con 38 Escuelas Profesionales en sus tres áreas de estudio (biomédicas, sociales e ingenierías), y actualmente existe gran demanda por ciertas Escuelas Profesionales, sin embargo, en ciertas escuelas se observa una tendendia de postulantes/ingresantes va decayendo, tal es el caso de la Escuela Profesional de Antropología que tiene una ecuación lineal para ingresantes de Y = -1.4727X + 31.018, lo que indica que el siguiente examen general del 2017 se espera tener trece (13) ingresantes, tambien se observa que la tendencia de elección de ésta Escuela Profesional va decayendo. Sucede lo mismo Ingeniería Estadístia e Informática, Ingeniería Química, etc.

Para el caso de los colegios de la Región de Puno, el Colegio Nacional Nuestra Señora de Alta Gracia del distrito de Ayaviri, en el proceso de admisión general 12 de enero del 2014 cuenta con una cantidad de 180 postulantes y 31 ingresantes; en el proceso de admisión general 18 de enero del 2015 cuenta con una cantidad de 172 postulantes y 11 ingresantes; en el proceso de admisión general 31 de enero del 2016

cuenta con una cantidad de 142 postulantes y 7 ingresantes, se puede observar que la cantidad de postulantes/ingresantes va decayendo; el colegio particular Nuestra Señora de la Merced del distrito de Puno, en el proceso general del 12 de enero del 2014 cuenta con 69 postulantes y 6 ingresantes ; en el proceso general del 18 de enero del 2015 cuenta con 83 postulantes y 6 ingresantes; en el proceso general 31 de enero del 2016 cuenta con 61 postulantes y 6 ingresantes. Esta información indica que la cantidad de ingresantes/postulantes disminuye, entonces ¿la formación brindada a los escolares de los colegios público y privados está decayendo?

1.2. FORMULACIÓN DEL PROBLEMA

¿Cuál es el efecto del análisis predictivo sobre la información de postulantes e ingresantes de las Escuelas Profesionales y la formación brindada a los escolares en los colegios de educación secundaria, públicos y privados de la región de Puno?

1.3. OBJETIVOS

1.3.1. Objetivo general

Realizar un análisis de la información de postulantes e ingresantes para predecir la tendencia del futuro de las diferentes Escuelas Profesionales y la formación brindada en las escuelas de educación secundaria, públicas y privadas de la región de Puno.

1.3.2. Objetivos específicos

- ✓ Recabar información de postulantes e ingresantes de las

diferentes Escuelas Profesionales de los diferentes procesos de admisión de la Universidad Nacional del Altiplano.

- ✓ Recabar información de ingresantes de los diferentes colegios públicos y privados de educación secundaria de la región de Puno en las diferentes modalidades de ingreso a la Universidad Nacional del Altiplano.
- ✓ Aplicar la metodología CRISP-DM para la gestión de los datos en sus diferentes fases y usar el software R, RMySql y paquetes de manipulación de datos.
- ✓ Interpretar y graficar la información de postulantes e ingresantes a las diferentes Escuelas Profesionales y de colegios públicos y privados de educación secundaria de la región de Puno.

1.4. JUSTIFICACIÓN

En diversos sectores institucionales es de gran interés la implementación de estrategias para la toma de decisiones importantes a través del análisis de datos.

El avance de la tecnología en el campo de la informática trae como consecuencia adelantos muy significativos para las instituciones, tal es el caso de la minería de datos y las técnicas empleadas para extraer información y descubrir patrones ocultos de información, sin embargo, esta información contenida en la base de datos es usada tangencialmente y no dándole la importancia debida y los beneficios que ésta pueda ofrecer a las Escuelas Profesionales y escuelas de educacion secundaria, públicas y privadas de la región de Puno.

Para este caso, es conveniente llevar a cabo esta investigación, ya que permitirá tomar decisiones en las diferentes Escuelas Profesionales de la universidad y escuelas de educación secundaria gracias al uso de éstas técnicas de minería de datos, obteniendo beneficio en dichas instituciones dándole la posibilidad de tomar las medidas correctivas que requieren dichas entidades del estado.

Por la razones expuestas es indispensable realizar un análisis de los datos de postulantes e ingresantes de las diferentes Escuelas Profesionales de la Universidad Nacional del Altiplano y escuelas de educación secundaria, públicas y privadas de nuestra región y asi implementar las medidas correctivas lo que conducirá a cumplir los objetivos de dichas instituciones.

1.5. LIMITACIONES DE LA INVESTIGACIÓN

Restricción de los datos: Los datos a procesar serán todos los postulantes procedentes del departamento de Puno.

Espacio físico – geográfico: Comisión Central de Admisión de la Universidad Nacional del Altiplano – Puno

Espacio semántico:

Análisis predictivo: conjunto de datos de los procesos de admisión que serán analizados, procesados e interpretados con paquetes adicionales basados en R.

Proceso de Admisión: modalidad en la que un postulante se presenta a la Universidad Nacional del Altiplano para alcanzar una vacante de ingreso.

CAPÍTULO II

MARCO TEÓRICO

2.1. ANTECEDENTES DE ESTUDIO

A continuación se mencionan los antecedentes de trabajos de investigación realizados a nivel nacional e internacional.

(Velasco, 2017), realiza el análisis exploratorio de datos del mercado eléctrico en España, cuyo objetivo es clasificar, representar gráficamente y resumir la información contenida en un fichero de datos, realizar un análisis exploratorio de datos desarrollado con R y aplicado al mercado eléctrico español. Han utilizado los datos públicos del MIBEL de los años 2011, 2012 y primer trimestre del 2013 disponibles en www.omelholding.es. En primer lugar se introducen los conceptos y características esenciales necesarias para comprender el mercado de la energía en España. Para poder realizar un modelo matemático correcto, se requiere de un análisis de los datos previo donde se determinen los valores atípicos, la normalidad y las relaciones entre las variables. Los códigos de los scripts de R programados para este estudio y las funciones específicas de librerías de R utilizadas se incluyen se enumeran, se comentan en el trabajo y se incluyen en un anexo. La gran cantidad de gráficos ofrece al lector una mejor visualización de los

datos y por tanto una mejor interpretación de los resultados.

(Diaz, 2015), realiza un análisis estadístico sobre una base de datos de las tarjetas de Millennium, representando en un mapa de densidad demográfica de Coruña y sobre ese mapa representar la red de autobuses de A Coruña coloreando cada tramo entre dos stops por la cantidad de pasajeros, con el objetivo de identificar si hay zonas que necesitan más autobuses o más líneas. Para lograr ésta representación se usó paquetes para integrar mapas de ciudades en R. Se usó el paquete Osmar que proporciona infraestructura para acceder a los datos de OpenStreetMap de diferentes fuentes, permitiendo trabajar con los datos OSM en el lenguaje R y convertir los datos en objetos basados en los paquetes existentes de R. También usaron, el paquete RgoogleMaps.

(Velasquez y Cataño, 2010), en us artículo presenta ejemplos de por qué el lenguaje R es de interés para profesionales e investigadores pertenecientes al área de las ciencias de la computación. Finalmente, se argumenta por qué el lenguaje R es una herramienta interesante para desarrollar software en el campo de la inteligencia computacional.

(Pereira, 2010), en su investigación concluye que el análisis de datos es muy útil para estudiar y ajustar de manera eficiente el comportamiento de un sistema dinámico lineal o no lineal a partir de las medidas discretas de sus variables. Por tanto, el objetivo principal de un modelo de regresión generado a partir de un análisis predictivo es obtener una ecuación matemática que nos permita “predecir” con el mínimo error posible el valor de una variable dependiente Y una vez conocido los valores de X1, X2, …,

Xn o variables independientes predictoras. Dicha ecuación servirá como modelo o función de aproximación para la predicción de futuras aproximaciones.

Cuando las variables predictoras están muy correlacionadas, los coeficientes de regresión resultantes de un ajuste por mínimos cuadrados ordinarios (MCO) pueden llegar a ser muy erráticos e imprecisos, debido a los efectos desastrosos que la multicolinealidad tiene sobre su varianza. Estos coeficientes originan predicciones erróneas a la hora de vaticinar nuevas respuestas correspondientes a entradas similares que deberían pronosticar salidas similares. La técnica Ridge Regression (RR) trata estas colinealidades minimizando el problema al contraer los coeficientes de regresión de MCO mediante la introducción de un sesgo, logrando coeficientes ajustados con menor varianza, dando estabilidad así a la predicción del modelo y solucionando dicho problema.

Finalmente, se han aplicado estas técnicas predictivas a diferentes series temporales no linales. Para ello se compararon los resultados en presencia de dos tipos de intensidades de ruido gaussiano añadido, con los resultados obtenidos en ausencia de ruido (datos brutos originales), concluyendo que el uso del kernel lineal mediante la solución dual de RR es el que mejor rendimiento proporciona en términos de mínimo error en el ajuste. Además la forma estructural de la serie temporal esperada seguía conservándose, incluso en presencia de ruido gaussiano moderado.

(Prieto, 2010), en el resumen de su artículo manifiesta que la reprobación escolar, específicamente en el nivel superior, es un fenómeno altamente

indicativo de la crisis por la que atraviesa la sociedad en general y la educación. Se entiende que la reprobación como parte del fracaso escolar es preocupante en todos los niveles educativos. Se estima que la eficiencia terminal en educación superior en México oscila entre 53% y 63%. En este trabajo se llevó a cabo el análisis de los datos que nos permitirán generar un modelo que ayude a predecir, desde que los alumnos ingresan a la Universidad, las causas que los llevarán a reprobar, así como las materias con mayor riesgo. Se recolectaron los datos relevantes que inciden en la reprobación por alumno, resultando un repositorio denominado datawarehouse, sobre él se está trabajando para diseñar el modelo predictivo. Finalmente, se implementará en una interfaz para que el usuario pueda capturar y observar los resultados.

(Cortes y Level, 2008), en us artículo cuyo objetivo de esta investigación fue conocer la validez predictiva del proceso de admisión en el rendimiento académico, en el primer año de la licenciatura en una universidad privada de la Ciudad de México. Se consideraron como variables predictoras del rendimiento las calificaciones en el Examen Nacional de Ingreso a la Educación Superior (EXANI II), el promedio general de preparatoria y el puntaje obtenido en el cuestionario sobre problemas sociales (DIT). Participaron 240 alumnos de ambos sexos, inscritos en la carrera de Psicología, que tenían en promedio 20 años. Los resultados permitieron observar que las calificaciones más altas que se obtuvieron en el EXANI-II fueron en las áreas de razonamiento verbal y numeric, y después en el área de español. Asimismo, se encontró que el puntaje en el EXANI-II, el promedio de bachillerato y el desarrollo moral permitieron predecir el

rendimiento académico en el primer año de la carrera.

(Rodriguez *et al*., 2008), en su artículo verifican la capacidad predictiva de las pruebas de aptitud, el índice académico y los exámenes de ingreso, en relación con el rendimiento académico en la carrera de Medicina, se diseñó una investigación de seguimiento longitudinal de los estudiantes que ingresaron en el curso 91-92, utilizando los datos obtenidos durante el proceso selectivo y sus calificaciones durante los primeros 4 años de la carrera para la totalidad de las asignaturas del plan de estudio. Dada la naturaleza continua de la variable dependiente, el análisis estadístico descansó en lo esencial en el modelo de regresión múltiple. Se confirmó la relevancia del índice académico como predictor del rendimiento y se han aportado, tal vez, las primeras evidencias de la capacidad predictiva de los exámenes de ingreso. La capacidad pronóstica de las variables que se registran en el momento del ingreso se disipa a medida que el alumno transita de un curso a otro y su lugar lo ocupan los propios indicadores de rendimiento parcial.

2.2. SUSTENTO TEÓRICO

Para el conocimiento, análisis se ha consultado las diferentes teorías, definiciones y evaluaciones de los autores que se cita a continuación:

2.2.1. El análisis predictivo

(Espino, 2017), indica que el análisis predictivo es un área de la minería de datos que consiste en la extracción de información existente en los datos y su utilización para predecir tendencias y patrones de comportamiento, pudiendo aplicarse sobre cualquier evento desconocido, ya sea en el

pasado, presente o futuro. El análisis predictivo se fundamenta en la identificación de relaciones entre variables en eventos pasados, para luego explotar dichas relaciones y predecir posibles resultados en futuras situaciones. Ahora bien, hay que tener en cuenta que la precisión de los resultados obtenidos depende mucho de cómo se ha realizado el análisis de los datos, así como de la calidad de las suposiciones.

En un principio puede parecer que el análisis predictivo es lo mismo que hacer un pronóstico (que hace predicciones a un nivel macroscópico), pero se trata de algo completamente distinto. Mientras que un pronóstico puede predecir cuántos helados se van a vender el mes que viene, el análisis predictivo puede indicar qué individuos es más probable que se coman un helado. Esta información, si se utiliza de la forma correcta, supone un cambio radical en el juego, ya que permite orientar los esfuerzos para ser más productivos en la consecución de los objetivos.

Para llevar a cabo el análisis predictivo es indispensable disponer de una considerable cantidad de datos, tanto actuales como pasados, para poder establecer patrones de comportamiento y así inducir conocimiento. Por ejemplo, en el caso comentado en el párrafo anterior, acerca de quién es más probable que se coma un helado, si se cruzan datos acerca de la temperatura registrada, la época del año y si es fin de semana o festivo se puede inferir qué perfil de persona comerá helado. Este proceso se realiza gracias al aprendizaje computacional. Los ordenadores pueden “aprender” de manera autónoma y de esta forma desarrollar nuevo conocimiento y capacidades, para ello basta con proporcionarles el más potente y gran recurso natural de la sociedad moderna: los datos.

2.2.2. El lenguaje de programación R

Según (Ocoña, 2017), R es un lenguaje de programación muy flexible orientado a la estadística computacional, el análisis de datos y el desarrollo de gráficos. Es un software libre y gratuito desarrollado bajo las condiciones GNU General Public License (www.gnu.org) por el equipo central de la R Foundation for Statistical Computing (www.r-project.org).

Como en cualquier lenguaje de programación, el usuario debe conocer bien el entorno de trabajo y las funciones básicas implementadas en R para, a partir de ellas, realizar el análisis estadístico deseado o desarrollar nuevas funciones. En el ámbito informático, una función es un grupo de instrucciones que procesa los datos introducidos y devuelve un valor final. Así, para calcular la media de una serie de valores en R será necesario programar una función que contenga las siguientes instrucciones:

```
function (x, trim = 0, na.rm = FALSE, ...) {
    if (!is.numeric(x) && !is.complex(x) && !is.logical(x)) {
        warning("argument is not numeric or logical: returning NA")
        return(as.numeric(NA))
    }
    if (na.rm)
        x <- x[!is.na(x)]
    trim <- trim[1]
    n <- length(x)
    if (trim > 0 && n > 0) {
        if (is.complex(x))
            stop("trimmed means are not defined for complex data")
        if (trim >= 0.5)
```

```
            return(median(x, na.rm = FALSE))
        lo <- floor(n * trim) + 1
        hi <- n + 1 - lo
        x <- sort(x, partial = unique(c(lo, hi)))[lo:hi]
        n <- hi - lo + 1
    }
    if (is.integer(x))
        sum(as.numeric(x))/n
    else sum(x)/n
}
```

Todas estas instrucciones forman una función denominada mean, que por defecto ya viene implementada en R. Cuando el usuario hace uso de ella, la función solicita unos datos y devuelve el valor medio de los mismos. Así, al escribir mean(c(0,2,4)) se procesarán automáticamente las instrucciones anteriores y se obtendrá como resultado 2, media aritmética de los valores 0, 2 y 4.

Al igual que mean, existen otras funciones que ya han sido programadas por el equipo de desarrollo de R y están disponibles para su uso inmediato, como las funciones min y max que devuelven respectivamente el valor mínimo y máximo de los datos introducidos. Así, min(c(0,2,4)) dará como resultado 0 y max(c(0,2,4)) devolverá el valor 4. Estas funciones forman parte de procedimientos estadísticos básicos, por lo que están agrupadas en un paquete de funciones denominado Base. Este paquete, junto a otros que contienen funciones más avanzadas, ha sido desarrollado por el equipo central de R y viene incorporado en su instalación.

En general, las instrucciones que permiten realizar un cálculo determinado se programan en una función y a su vez las funciones se agrupan en paquetes temáticos para facilitar su localización. La estructura es similar a la descrita en el siguiente gráfico:

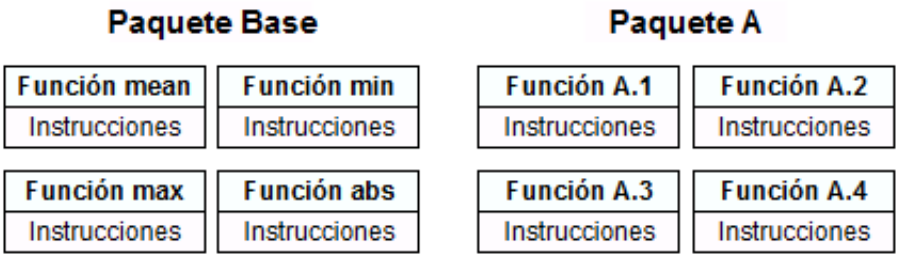


Figura 1. Estructura de los paquetes

Actualmente, quizá R sea uno de los lenguajes de programación con más funciones implementadas para el análisis de datos. Además, su flexibilidad permite programar e incorporar nuevos modelos que han sido desarrollados en el campo de la teoría matemática, cualidad que lo ha convertido en un software muy popular entre estadísticos y matemáticos especializados en estadística computacional.

A pesar de sus cualidades técnicas, el uso de R puede resultar complejo para personas que no están familiarizadas con los lenguajes de programación. La necesidad de escribir instrucciones y comandos para realizar análisis estadísticos simples hace que R no sea el software elegido por profesionales no especializados en estadística para llevar a cabo proyectos de investigación aplicada.

2.2.3. Package RMySQL

(James, 2016), indica que las tablas de MySQL se leen en R como data.frames, pero sin coerción de caracteres o datos lógicos en factores.

Del mismo modo al exportar data.frames, los factores(factor) se exportan como vectores de caracteres. Las columnas enteras se importan normalmente como R integer vectors, excepto para casos como BIGINT o UNSIGNED INTEGER que son los vectores de doble precisión de R para evitar el truncamiento (actualmente los números enteros de R son cantidades de 32 bits). Las variables de tiempo se importan/exportan como datos de caracteres, por lo que es necesario convertirlos a su representación de fecha / hora favorita.

Actualmente no hay instalaciones para importar/exportar BLOBs. Las tablas en una base de datos relacional sólo son superficialmente similares a los datos de R.frames (por ejemplo, tablas como conjuntos no ordenados de filas en comparación con los data.frames como conjuntos ordenados, tablas con restricciones referenciales, índices, etc.).

Para establecer la conexión se debe instalar el paquete RMySQL con el siguiente comando: install.packages(“RMySQL”) y para la carga del paquete se usa el siguiente comando: library(“RMySQL”).

Teniendo la información en la base de datos se establece la conexión con los siguientes comandos:

```
library(RMySQL)
m ← dbDriver("MySQL");
con ←  dbConnect(m,user=root',
        password=123',
        host=localhost',
        dbname=prueba');
res ←  dbSendQuery(con, "select * from mitabla")
```

genes ← fetch(res, n = -1)

2.2.4. WaterML R package for managing ecological experiment data on a CUAHSI HydroServer

(Kadlec, 2015), en éste artículo trata sobre la recolección y almacenamiento de datos ecológicos empleando el paquete WaterML, existen grupos independientes de investigadores que dirigen sus laboratorios y no sólo se centran en sus preguntas científicas ahora deben aprender técnicas de procesamiento de datos y desarrollo de base de datos para que el trabajo no se frustre por mucho tiempo frente al mal almacenamiento o acceso de los datos. Aquellos que están realizando un experimento y aquellos que esperan comprender los datos provenientes del experimento pueden no tener los medios financieros necesarios para desarrollar su propio sistema de gestión de datos. Los sitios web de alojamiento de datos compartidos que utilizan estándares internacionales (WaterML) y soluciones de software de código abierto para la administración de datos (HydroServer para Windows o HydroServer Lite para Linux), el archivo (base de datos ODM) y la publicación (servicio web WaterOneFlow) pueden ser un medio eficaz para investigadores independientes siguen siendo competitivos en un mundo de datos en la investigación científica. El Consorcio de Universidades para el Avance de la Ciencia Hidrológica (CUAHSI) brinda apoyo a científicos y grupos independientes de laboratorio de ecología y ayuda a administrar y organizar sus datos experimentales utilizando tecnología de código abierto.

Un problema particular que enfrentan los investigadores en ecología es

integrar un sistema de gestión de datos con un entorno de análisis computacional como Matlab, stata o R. Una característica común de estos entornos de análisis computacional es que proporcionan capacidades para el análisis exploratorio (gráficos, gráficos) y Inferencia estadística (prueba de hipótesis). Por lo general, los pasos de análisis de datos se registran en un script, haciendo que los pasos sean reproducibles. Un sistema que vincula el software de análisis computacional con la gestión de datos basados en estándares que la nube permitiría a los investigadores automatizar la recuperación de datos sin procesar o datos previamente procesados directamente desde el sistema de gestión de datos en su entorno analítico. El sistema también permitiría a los investigadores publicar los datos y los resultados del análisis en el sistema para fines de archivo y de compartición.

Se han construido una serie de herramientas existentes que cumplen algunas de las metas generales mencionadas anteriormente. Por ejemplo, recientemente se han introducido dos paquetes R para recuperar datos de cantidad y calidad del agua del Sistema Nacional de Información sobre el Agua del USGS (NWIS) en el repositorio de paquetes R Comprehensive R Archive Network (CRAN), incluyendo los paquetes "dataRetrieval" y "waterData". Estos paquetes R proporcionan funciones útiles de descarga de datos que podrían apoyar la investigación ecológica en términos de la información nacional de agua de los Estados Unidos, sin embargo, no están destinados a la carga de datos ni a la gestión de datos asociados con la investigación de laboratorio. Para la investigación de laboratorio, es posible utilizar, controladores de base de datos que vinculan el paquete de software

estadístico R a las principales plataformas de bases de datos relacionales. El paquete "RObsDat" (Reusser, 2014) es uno de esos drivers diseñado específicamente para conectarse a cualquier base de datos de observaciones ambientales compatible con el esquema de modelo de datos de observaciones (ODM) utilizando el estándar SQL (Structured Query Language). Otros ejemplos más generales de los paquetes que vinculan R a una base de datos relacional utilizando SQL son "RMySQL" (James y DebRoy, 2012) y "RSQLite" (James y Falcon, 2011).

Los drivers indicados anteriormente requieren una conexión SQL directa a una base de datos ODM asociada utilizando una dirección IP y un número de puerto. El problema con una conexión directa de SQL usando la dirección IP y el número de puerto es que los firewalls institucionales bloquean los puertos necesarios en la mayoría de los casos, haciendo que la conexión sólo sea posible dentro de la red local de la institución. En el caso común de las colaboraciones multiinstitucionales, tales firewalls pueden restringir el acceso directo a la base de datos, por lo que se requiere otro enfoque. Además, no todas las instancias de HydroServer utilizan el esquema de la base de datos ODM. Una solución a estos problemas es abstraer la base de datos física exponiendo sólo una capa de servicios web (también llamada Interfaz de Programación de Aplicaciones Web o API Web). Normalmente, la API web utiliza el Protocolo de transferencia de hipertexto (HTTP) para pasar información entre una herramienta de cliente y una base de datos utilizando JavaScript (JSON) o texto codificado XML (Extensible Markup Language). Dicho servicio web resuelve el problema del cortafuegos y permite el acceso a los datos a través de las instituciones,

aunque comparado con el poder expresivo de SQL, el servicio Web típicamente sólo permite un conjunto limitado de consultas predefinidas. Si está bien definido (es decir, como en el caso de los servicios web HIS de CUAHSI), este subconjunto limitado de consultas puede satisfacer fácilmente los requisitos de la mayoría de los casos de uso de la administración de bases de datos.

El método empleado en el diseño se eligió una solución de código abierto utilizando la base de datos relacional MySQL y el software HydroServer Lite. Se puede instalar y alojar en cualquier servidor web o cuenta webhosting compartida que soporte PHP (versión 5 o superior) y MySQL. Hemos alojado la base de datos y el servidor web en el sitio web de alojamiento compartido http://worldwater.byu.edu, que proporciona espacio web y espacio de base de datos para cualquier grupo de investigación independiente para publicar sus datos de acceso abierto. Una base de datos centralizada basada en web como esta tiene el beneficio de permitir a los ecologistas tanto almacenar datos como compartirlos fácilmente con compañeros de trabajo en el proyecto. Como método principal para el análisis estadístico de los resultados experimentales se eligió el entorno computacional R porque también es de código abierto, multiplataforma y ampliamente utilizado en la investigación ecológica. En R, cada paso del procesamiento de datos se almacena como una secuencia de comandos, lo que permite a los compañeros de trabajo para reproducir el análisis. Su naturaleza basada en scripts también hace que R sea un buen candidato para construir una base de datos integrada y un sistema de análisis, ya que los servicios web de base de datos también deberán utilizarse a través de

código de script.

2.2.5. Inteligencia de negocios

(Martín, 2015), manifiesta que el análisis predictivo es la disciplina que tiene el mayor potencial de valor/recompensa, muy por encima de la Inteligencia de Negocios tradicional que audita actividades o transacciones realizadas en el pasado brindando un análisis retrospectivo del negocio, en vez de proveer una visión de la información que permita tomar decisiones relacionadas con tendencias y situaciones que pueden presentarse en el futuro.

El análisis predictivo crea modelos analíticos en los niveles más detallados del negocio como, por ejemplo, al nivel de cada cliente individual, los productos, las campañas, el movimiento en las tiendas y en los puntos de venta y busca por comportamientos predictivos, tendencias y reglas de negocios que puedan expresarse en fórmulas matemáticas que puedan utilizarse para predecir la probabilidad de que sucedan ciertos comportamientos y acciones.

2.2.6. Big Data ¿evolución o revolución?

(Pautasio, 2014), Big data es como todas las tecnologías capaces de analizar y procesar grandes volúmenes de información, de diversas fuentes, con diferentes estructuras y a gran velocidad.

Big data no es tal si no soluciona un problema específico de un cliente. De nada sirve tener acumulados grandes cantidades de información si no se puede hacer uso de ellas. Hay que saber qué dato es valioso y para ello

exista una tecnología útil para resolver los problemas.

Big data es importante porque, además, es un concepto que engloba, se conecta y dialoga con todas las tendencias IT: cloud computing, movilidad, Internet de Todo (IoE, por sus siglas en inglés), ciberseguridad, analytics son algunas de las palabras claves por donde se mueve el mundo de la tecnología y en todas ellas hay espacio para big data.

Big data es sin duda una revolución. Pero es también la evolución lógica de las herramientas de business intelligence ante un mundo hiperconectado que genera cada vez más cantidad de información. Las empresas deberán aprovechar esta marea de información para lograr un diferencial competitivo y no quedarse detrás en la carrera.

Las redes sociales, los dispositivos móviles, las aplicaciones y todo el nuevo universo de dispositivos que se abre a través de Internet de Todo –que comprende las conexiones entre personas, procesos y máquinas– son nuevas fuentes de información que se deberán incorporar para tomar mejores decisiones.

Según información divulgada por IBM, cada día generamos 2,5 trillones de bytes de datos de una variedad casi infinita de fuentes.

El último informe de EMC sobre el universo digital, presentado a mediados de abril, reveló que en 2013 se generaron 4,4 trillones de gigabytes en todo el mundo. El estudio de EMC intenta clasificar y pronosticar la cantidad de datos producidos anualmente a escala mundial. Para poner en perspectiva, esta información comprendida en el universo digital representaría una pila de tabletas iPad Air de 128 Gb tan alta que cubriría dos tercios de la

distancia hacia la Luna, que es de 253.704 kilómetros. El informe augura que este universo crecerá cerca de 40% año a año y alcanzará 44,4 trillones de gigabytes para 2020.

Es interesante destacar que actualmente 70% de la información del universo digital es generada por seres humanos. De este total, 85% es responsabilidad de organizaciones, por ejemplo, un correo electrónico laboral.

2.2.7. Exploración de patrones de datos

Según indica (Jhon y Reitsch, 1996), se debe reunir datos que sean aplicables para la tarea de pronóstico y que contenga información que pueda producir pronósticos precisos. Las técnicas de pronóstico cuantitativas se utilizan cuando existen suficientes datos históricos disponibles y cuando se juzga que estos datos son representativos de un futuro desconocido. Esta técnica se apoya en la suposición de que el pasado puede extenderse hacia el futuro de manera significativa para proporcionar pronósticos precisos.

Las técnicas estadísticas se enfocan completamente en patrones, cambios en los patrones y perturbaciones causadas por influencias aleatorias, éstas técnicas son los promedios moviles y la atenuación exponencial, descomposición de series de tiempo y proyecciones de tendencia y la metodología Box-Jenkins.

2.2.8. El costo de no hacer nada en el análisis predictivo

(Velez, 2014), ¿Cuánto cuesta no hacer nada? En principio, si no haces

algo no cuesta, ¿no? Veámoslo desde otro punto de vista. Si vamos a cualquier comercio, y no hallamos los productos que esperábamos encontrar, o la oferta existente no coincide con nuestras expectativas, ¿qué coste tiene esta situación para el establecimiento?, ¿volveríamos al comercio en un futuro?

De forma análoga, si estamos al frente de la gestión de una compañía y no invertimos recursos en predecir escenarios futuros, que nos permitan establecer una estrategia adecuada, ¿supone esto un ahorro? ¿es gratuito el dirigir la compañía a ciegas?

Resulta cuanto menos llamativo que únicamente una de cada ocho empresas disponga de sistemas con capacidad suficiente como para realizar análisis predictivos, quedándose las otras siete ancladas en el estudio de datos históricos que ya han acontecido, y sobre los que no pueden reaccionar. Evidentemente su implantación implica una inversión, tanto en recursos económicos como humanos, pero sin lugar a dudas tiene un retorno claro. Pensemos en una entidad financiera que no realiza un estudio antes de proporcionar un crédito significativo a uno de sus clientes, a partir de datos históricos y tendencias del momento. El riesgo en el que incurriría superaría con creces el coste de dicho estudio.

Dentro del análisis de información, podemos decir que existen tres niveles:

- ✓ Descriptivo
- ✓ Predictivo
- ✓ Prescriptivo

El análisis descriptivo pretende, mediante técnicas de reporting tradicional,

AdHoc, y multidimensional, proporcionar una visión sobre lo que ha ocurrido, en qué medida, y al nivel de detalle deseado. Es el más sencillo, pero el que otorga una menor ventaja competitiva dado que prácticamente todas las empresas lo han alcanzado.

El predictivo va un paso más allá, incorporando alertas, presupuestación y herramientas de simulación, de forma que permite identificar qué acciones son necesarias en un determinado momento, que ocurrirá si las tendencias se mantienen en el tiempo, y qué podría ocurrir ante determinados cambios en el modelo. Es, de los tres niveles, el más equilibrado en términos de complejidad y competitividad.

El prescriptivo es el más sofisticado y complejo, y se basa en la optimización tradicional y estocástica. Estudia las condiciones que se han de dar para conseguir el mejor resultado, llegando a considerar el efecto de la variabilidad dentro del modelo. Es el que va a proporcionar las mayores ventajas competitivas de los tres, pero no todas las compañías están preparadas para implementarlo debido al estado actual de sus sistemas de información.

Las aplicaciones informáticas de business analytics tienen, en general, más que superado el primer nivel. La tendencia del momento consiste en permitir que las compañías evolucionen al menos al segundo nivel, sin necesidad de realizar fuertes inversiones para conseguirlo, y con autonomía suficiente como para poder abordar su implantación internamente, con las debidas herramientas informáticas y el soporte de su proveedor tecnológico.

El reto del futuro no muy lejano de aquellas empresas que quieran seguir siendo competitivas en el mercado será abordar el tercer nivel, pero a éste sólo llegarán aquellas que hayan dado los pasos previos, con lo que es momento de considerar si realmente es gratuito no hacer nada.

2.2.9. R and Data Mining: Examples and Case of Studies

(Zhao, 2013), indica que las principales técnicas para la mineria de datos incluyen la clasificación, la predicción, la agrupación, reglas de asociación, análisis de secuencia, análisis de series de tiempo y mineria de texto y nuevas técnicas como análisis de redes sociales y análisis de sentimientos. En aplicaciones reales un proceso de mineria de datos se divide en seis fases principales: la comprensión del negocio, la comprensión de los datos, el modelado, evaluación y despliegue, según lo definido por CRISP-DM (Cross Industry Standard Process for Data Mining).

2.3. MODELO DE REFERENCIA CRISP-DM

El modelo de referencia CRISP-DM (Chapman *et al*., 2000) propone una metodología de minería de datos estandarizada, que es la más utilizada por su flexibilidad y capacidad de personalizarse para su aplicación en diferentes dominios fácilmente.

Esta metodología está descrita en términos de un modelo de procesos jerárquico, consiste de un conjunto de tareas descritas en cuatro niveles de abstracción: fase, tarea genérica, tarea especializada e instancia de proceso (Iturbide, 2013) y organizadas en forma gerarquica que van desde el nivel más general, hasta los casos más específicos y organiza el desarrollo de un proyecto de minería de datos en seis fases:

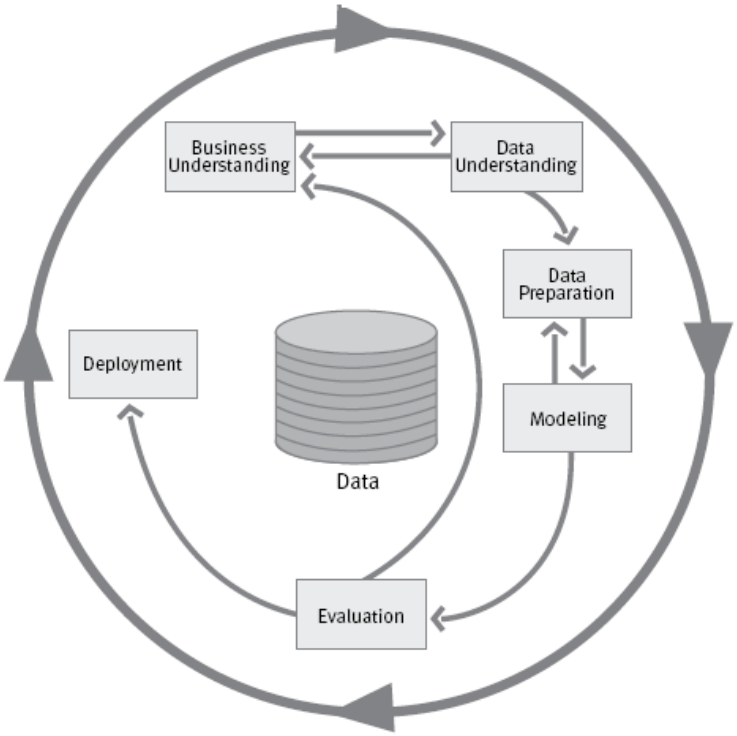


Figura 2. Fases del modelo CRISP-DM
Fuente: Extraído de [Chapman *et al*., 2000]

2.3.1. Fase de comprensión del negocio

Las principales tareas de esta fase son los siguientes:

Determinar los objetivos del negocio: se tiene que determinar cuál sería el problema que se desea resolver, porqué se usa minería de datos para dicho propósito y definir los criterios de éxito. En cuanto a este último, pueden ser de tipo cualitativo o de tipo cuantitativo, por ejemplo, si el problema es detectar fraude en el uso de tarjetas de crédito, el criterio de éxito cuantitativo seria el número de detecciones de fraude.

Evaluar la situación actual: en esta tarea se debe evaluar antecedentes y requisitos del problema, tanto en términos del negocio como en términos de la minería de datos. Algunos de los aspectos a tomar en cuenta pueden ser: el conocimiento previo acerca del tema, la cantidad de datos requeridas para resolver el problema, ventajas de aplicar minería de datos al problema,

entre otros.

Determinar los objetivos de la minería de datos: el objetivo de esta tarea es representar los objetivos del negocio en términos de las metas del proyecto de minería de datos. Por ejemplo, si el objetivo del negocio es el desarrollo de una campaña publicitaria para incrementar asignación de créditos hipotecarios, la meta de la minería de datos seria determinar el perfil de los clientes respecto de su capacidad de endeudamiento.

Producir de un plan de proyecto: la última tarea de esta fase tiene como objetivo desarrollar el plan de proyecto considerando los pasos a seguir y los métodos a emplear en cada paso.

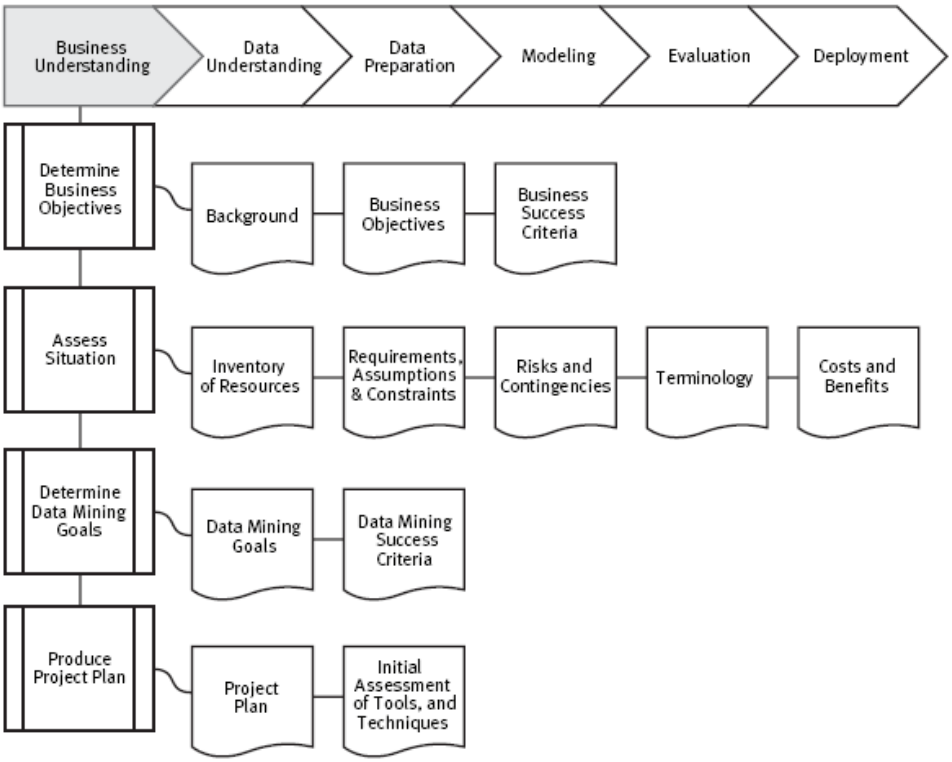


Figura 3. Comprensión del negocio
Fuente: Extraído de [Chapman *et al.*, 2000]

2.3.2. Fase de comprensión de los datos

Recolectar datos iniciales: tiene como objetivo principal la recolección de

datos iniciales y adecuación de datos para su posterior procesamiento. Se debe elaborar informes con una lista de los datos adquiridos, su localización, las técnicas utilizadas en su recolección y los problemas y soluciones inherentes a este proceso.

Describir los datos: se debe describir los datos iniciales obtenidos, tales como número de registros y campos por registro, su identificación, el significado de cada campo y la descripción del formato inicial.

Explorar los datos: Su finalidad es descubrir una estructura general para los datos. Involucra la aplicación de pruebas estadísticas básicas, que revelen propiedades en los datos, se crean tablas de frecuencia y se construyen gráficos de distribución. Se crea un informe de exploración de datos.

Verificar la calidad de los datos: se realiza la verificación de los datos para determinar la consistencia de los valores de los campos, la cantidad y distribución de los valores nulos, encontrar valores fuera de rango que pueden ser ruido para el proceso. Se tiene como objetivo asegurar la completitud y corrección de los datos.

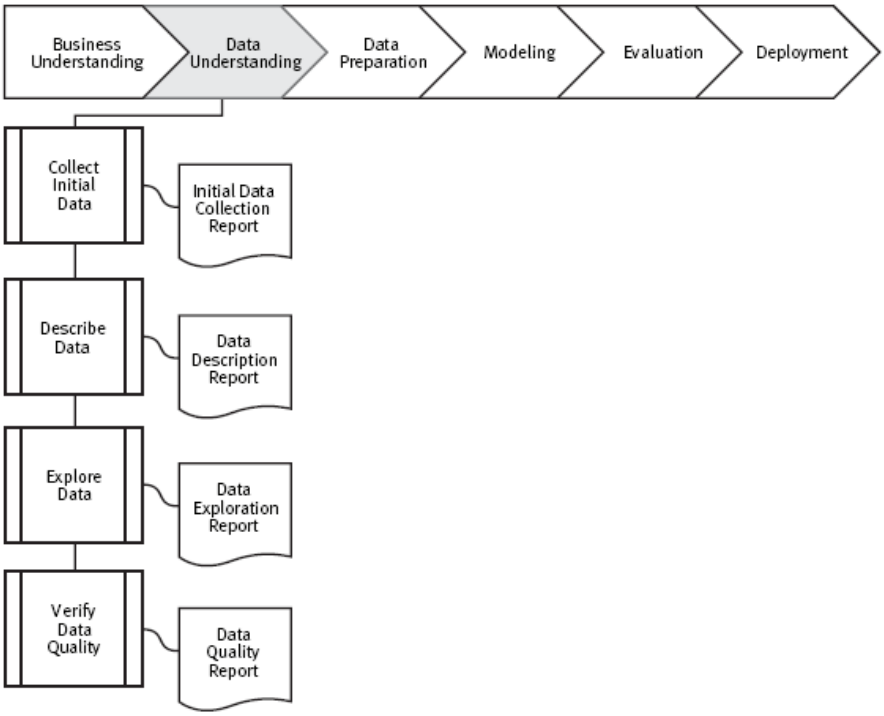


Figura 4. Comprensión de los datos
Fuente: Extraído de [Chapman *et al.*, 2000]

2.3.3. Fase de preparación de los datos

Seleccionar los datos: se selecciona un subconjunto de datos considerando la calidad de los datos, la limitación en el volumen o en los tipos de datos que están relacionadas con las técnicas de minería de datos seleccionadas.

Limpiar los datos: existe una diversidad de técnicas aplicables a esta tarea con el fin de optimizar la calidad de los datos para prepararlos para la fase de modelación. Algunas de las técnicas pueden ser: la normalización de los datos, discretización de campos numéricos, tratamiento con valores vacios, reducción del volumen de datos, etc.

Estructurar los datos: algunas de las operaciones a realizar en esta tarea puede ser la generación de nuevos atributos a partir de atributos ya existentes, integración de nuevos registros o transformación de valores

para atributos existentes.

Integrar los datos: involucra la creación de nuevas estructuras, por ejemplo, crear nuevos campos, nuevos registros, fusión de tablas o nuevas tablas.

Formatear los datos: consiste principalmente en transformar sintácticamente los datos sin modificar su significado con el fin de permitir o facilitar el empleo de alguna técnica de minería de datos en particular. Por ejemplo, eliminar comas, tabuladores, caracteres especiales, espacios, máximos y mínimos para las cadenas de caracteres, etc.

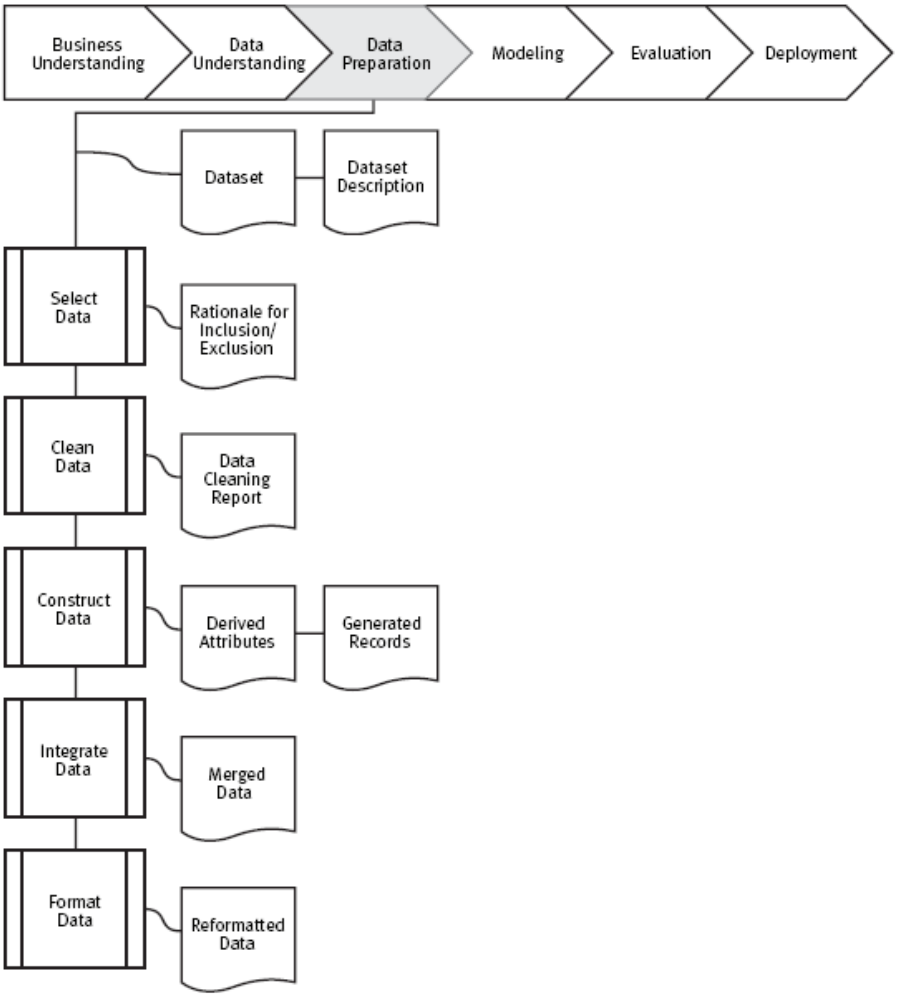


Figura 5. Preparación de los datos
Fuente: Extraído de [Chapman *et al.*, 2000]

2.3.4. Fase del modelado

Seleccionar la técnica del modelado: se debe elegir una técnica de modelado más apropiado para el proyecto específico. Se pueden elegir de acuerdo a los siguientes criterios:

- ✓ Ser apropiada al problema
- ✓ Disponer de los datos adecuados
- ✓ Cumplir requisitos del problema
- ✓ Tiempo adecuado para obtener un modelo
- ✓ Conocimiento de la técnica

Por ejemplo, si el problema es de clasificación se puede elegir de entre arboles de decisión, k-nearest neigbour o razonamiento basado en caos (CBR).

Generar plan de prueba: se debe generar un plan para probar la calidad y validez del modelo construido. Por ejemplo, en una tarea como la clasificación es posible usar la razón de error como medida de la calidad. Entonces, típicamente se separan los datos en dos conjuntos, uno de entrenamiento y otro de prueba.

Construir el modelo: se ejecuta la técnica seleccionada sobre los datos preparados para generar uno o más modelos. Todas las técnicas del modelado tienen un conjunto de parámetros que determinan características del modelo a generar. La tarea de selección de los mejores parámetros es iterativa basado en los resultados generados. Estos deben ser interpretados y su rendimiento justificado.

Evaluar el modelo: se debe interpretar los modelos de acuerdo al conocimiento del dominio y los criterios de éxitos preestablecidos.

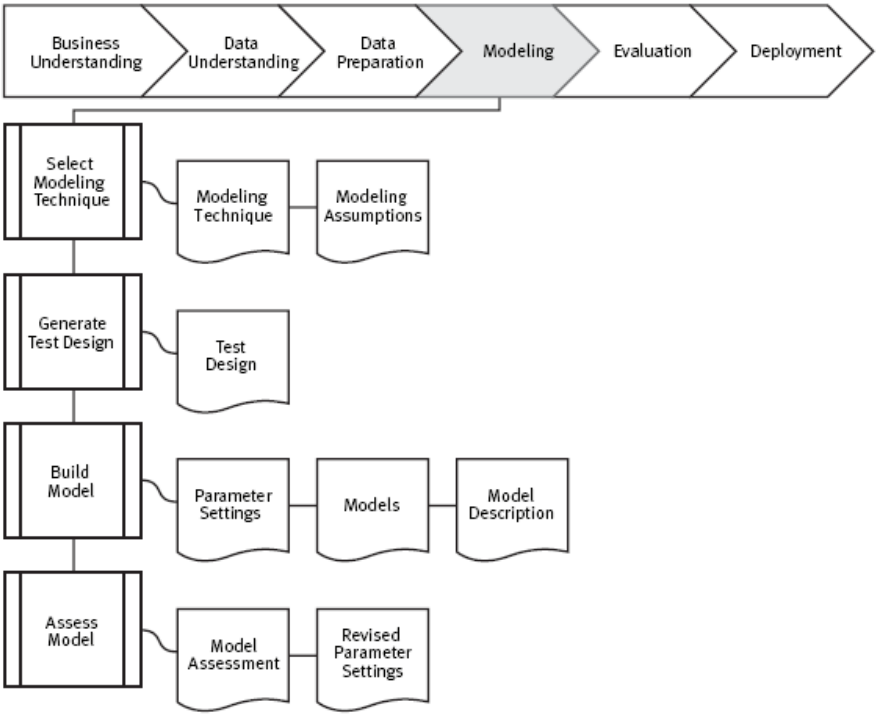


Figura 6. Modelado
Fuente: Extraído de [Chapman *et al*., 2000]

2.3.5. Fase de evaluación

Evaluar los resultados: esta tarea involucra la evaluación del modelo en relación a los objetivos del negocio y busca determinar si es aconsejable probar el modelo o determinar si hay alguna razón de negocio para el cual, el modelo es deficiente.

Revisar el proceso: consiste en la calificación del proceso entero de la minería de datos, con el objeto de identificar elementos que pudieran ser mejorados.

Determinar próximos pasos: en caso de que no se han generado resultados satisfactorios, se podría decidirse por otra iteración desde la fase de preparación de datos o modelación con otros parámetros.

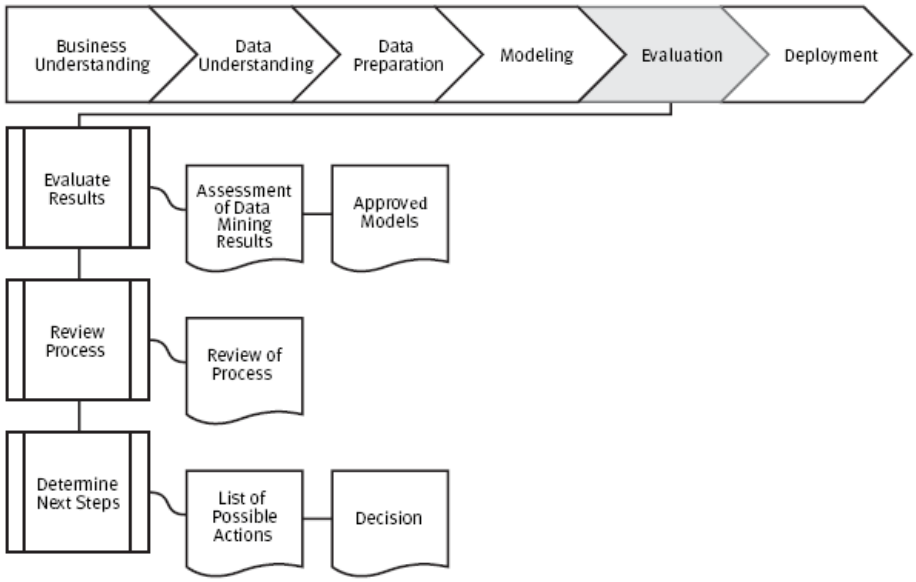


Figura 7. Evaluación
Fuente: Extraído de [Chapman *et al*., 2000]

2.3.6. Fase de implementación

Planear la implementación: esta tarea toma los resultados de la evaluación y concluye una estrategia para su implementación. Si un procedimiento general se ha identificado para crear el modelo, este procedimiento debe estar documentado para su posterior implementación.

Monitorizar y mantener: se debe preparar estrategias de monitorización y mantenimiento para ser aplicada sobre los modelos.

Informe final: dependiendo del plan de implementación, esta puede ser un resumen de los puntos importantes del proyecto y la experiencia lograda o puede ser una presentación final que incluya y explique los resultados logrados con el proyecto.

Revisar el proyecto: se evalúa lo correcto y lo incorrecto.

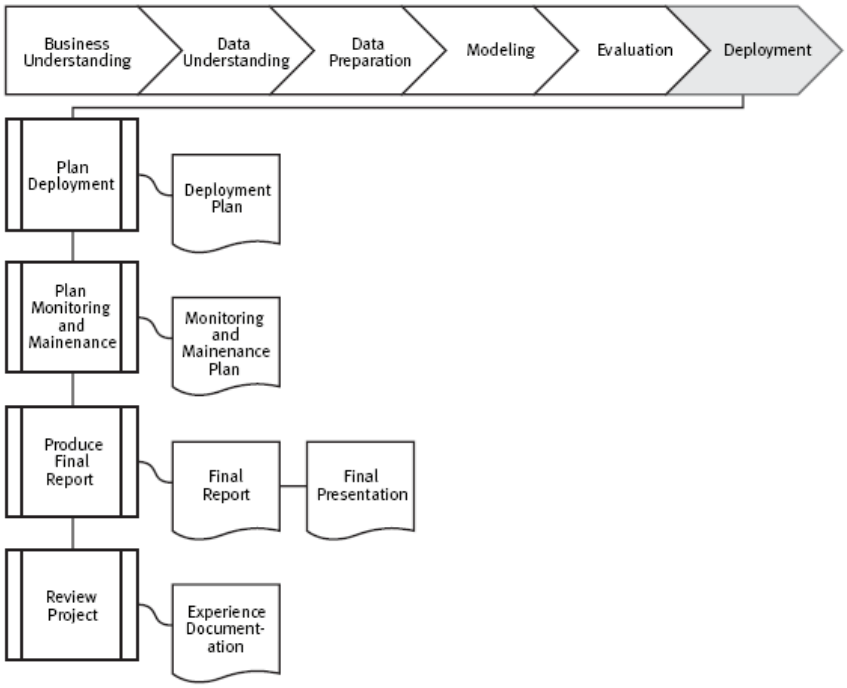


Figura 8. Implantación
Fuente: Extraído de [Chapman et al., 2000]

2.4. EQUIDAD Y CALIDAD EN LOS PROCESOS DE ADMISIÓN DE LA EDUCACIÓN SUPERIOR

(Blanco, *et al*., 2010), el presente trabajo tiene como propósito analizar los procesos de admisión de la Universidad Simón Bolívar (USB) y de la Universidad Central de Venezuela (UCV). Dicho análisis se realiza en función de una propuesta que se nutre de los postulados del igualitarismo contemporáneo e identifica las características de un proceso de admisión equitativo y, a su vez, de calidad. Para lograr tal propósito se realizó una revisión de los principios rectores de la política de admisión de cada universidad y un estudio de campo que permitió identificar el perfil del estudiante de la cohorte 2007 en los tipos de carreras comunes a ambas instituciones, esto es: Ingeniería, Arquitectura y Urbanismo y Ciencias Básicas. Los resultados indican que ambas universidades presentan acciones diferenciadas de acceso que promueven la inclusión con calidad.

Sin embargo, aún deben articular mejor los principios y valores de sus políticas de admisión con los mecanismos de ingresos, que ejecutan. Igualmente, el hecho de que estos programas tengan como base la nivelación de conocimientos da cuenta de la necesidad de actuar en la Educación Media Oficial para contribuir con la distribución equitativa de las condiciones de ingreso. Al estudiar la composición social de los estudiantes de nuevo ingreso, las hipótesis planteadas arrojaron que la USB y la UCV difieren en la estratificación social de sus estudiantes sólo para el tipo de carrera de Ingeniería. Finalmente, la asignación poco efectiva en el tipo de carrera en Ciencias Básicas evidencia la necesidad de reorientar la demanda social en los estudios superiores.

2.4.1. Descubrimiento de conocimientos en base de datos

(León, 2006), indica que el descubrimiento de conocimiento en bases de datos (*Knowledge Discovery in Database*) implica un proceso interactivo, que comprende la aplicación de métodos de minería de datos para extraer o identificar aquello que se considera conocimiento, a partir de la especificación de ciertos parámetros en una base de datos. La meta de este proceso es justamente procesar automáticamente grandes cantidades de datos en bruto, identificar los patrones más significativos y presentarlos como conocimiento apropiado para satisfacer las metas del usuario. El proceso de descubrimiento del conocimiento en bases de datos requiere de varios pasos:

- ✓ Entender el dominio de aplicación, el conocimiento relevante a utilizar y las metas del usuario.

- ✓ Seleccionar el conjunto de datos y enfocar la búsqueda hacia los subconjuntos de variables o muestras de datos donde se realizará el proceso de descubrimiento.
- ✓ Filtrar y preprocesar datos, diseñar una estrategia adecuada para manejar ruido, valores incompletos, secuencias de tiempo, etcétera.
- ✓ Reducir datos y proyecciones para disminuir el número de variables a considerar.
- ✓ Seleccionar la tarea de descubrimiento a realizar (clasificación, agrupamiento, regresión).
- ✓ Seleccionar el o los algoritmos a utilizar.
- ✓ Realizar el proceso de minería de datos.
- ✓ Interpretar los resultados.
- ✓ Incorporar el conocimiento descubierto al sistema.

Los algoritmos de la minería de datos realizan por lo general tareas de predicción (de datos desconocidos) y descripción de patrones mediante algoritmos de aprendizaje y estadísticos como:

- ✓ Análisis de dependencias.
- ✓ Identificación de clases (agrupamiento de registros en clases o clustering).
- ✓ Descripción de conceptos.
- ✓ Detección de desviaciones, casos extremos y anomalías.

Entre los componentes básicos de los métodos de minería de datos están:

- ✓ El lenguaje de representación del modelo.
- ✓ Evaluación del modelo.
- ✓ Método de búsqueda.

La minería de datos ha surgido del análisis potencial de grandes volúmenes de información, con el fin de obtener resúmenes y conocimiento que apoyen la toma de decisiones. Por ello, la minería de datos puede clasificarse según las siguientes variantes:

a) Las técnicas aplicadas:

- ✓ Sin algoritmos de aprendizaje
- ✓ Consultas SQL (Structured Query Language)
- ✓ Procesamiento analítico en línea OLAP (*On- line Transactional Processing*)
- ✓ Análisis estadístico (correlación, regresiones)

b) Las funciones que realizan:

- ✓ Redes neuronales y algoritmos genéticos
- ✓ Inducción de árboles y reglas

c) Nuevos algoritmos:

- ✓ Inducción de reglas de asociación
- ✓ Inducción de clasificadores bayesianos

Las diferentes técnicas permiten realizar asociaciones, clasificaciones, agrupamientos y el establecimiento de patrones secuenciales.

Aunque los diferentes campos de aplicación de la minería de datos demandan el desarrollo de poderosas y costosas herramientas para crear métodos de búsqueda de patrones, no es el único camino existente. El Web, como se conoce hoy, requiere una visión más integral de los problemas de organización y recuperación de información, sobre todo, si se considera que se encuentra estructurado mediante lenguajes de etiquetado que prácticamente describen sólo la forma en que la información debe presentarse al usuario (colores, maquetación, tipografía, etcétera) y dicen muy poco sobre su significado: semántica.

El proyecto denominado Web semántica (Semantic Web) busca que la información pueda reunirse de forma que un buscador pueda comprenderla en lugar de ponerla simplemente en una lista, donde el trabajo que hasta hoy se realizaba en función del usuario (el humano), se centrará en otro tipo de usuario que se valdrá de grandes cúmulos de información, clasificada, descrita y estructurada para una eficiente recuperación: el agente inteligente.

2.4.2. Ingeniería del proceso de software

Para lograr el entendimiento del proceso de software, es necesario, en primer lugar, definir que es un proceso. Un proceso “es una colección de actividades que toman uno o más tipos de entradas y crea una salida que es de valor para el cliente” (Borges de Barros Pereira, 2002). Bajo este concepto, un proceso de software se define como un marco de trabajo para las tareas que se requieren en la construcción del software de alta calidad definiendo métodos y técnicas para su construcción (Pressman R. , 1998).

2.4.2.1. Modelo de análisis

El modelo de análisis es una representación de los requisitos en un momento determinado; conforme el modelo de análisis evoluciona ciertos elementos se volverán relativamente estables.

Elementos del modelo de análisis

Existen muchas maneras de buscar los requisitos para un sistema basado en computadora, entre los más genéricos se tienen: elementos basados en escenarios (diagramas de casos de uso), elementos basados en clases (diagramas de clases), elementos de comportamiento (diagrama de estados) y elementos orientados al flujo.

a. Modelo de datos

En el modelo de datos se define todos los objetos de datos que se procesan dentro del sistema y las relaciones entre los objetos de datos mediante una notación gráfica, definiendo los objetos de datos, atributos y relaciones.

a.1. Modelo conceptual, Entidad-Relación-Atributos

El modelo E/R. La técnica más sencilla que se utiliza a la hora de analizar las necesidades de la base de datos.

Entidades. Las entidades creadas en cualquier base de datos son esencialmente los nombres o elementos que describe cuando habla de un proceso. Una persona, un lugar, o una cosa puede ser una entidad en un modelo de datos. Las entidades son en esencia tablas

en el diseño de la base de datos.

Relaciones. Las relaciones son la forma en que las entidades se relacionan entre sí, de forma que se puedan asociar registros en consultas o en definiciones de vista. La entidad primaria tiene un valor clave (clave principal) que identifica un registro en una tabla de forma univoca con otras tablas o entidades (hijos).

Atributos. Los atributos (o columnas) describen la entidad. Contienen detalles de la entidad y hacen que cada registro sea único con respecto a otros registros de la misma tabla (Dalton P. y Whitehead, 2000).

a.2. Modelo relacional

El elemento central del modelo relacional es la Relación. Una relación tiene un nombre, un conjunto de atributos que representan sus propiedades y un conjunto de tuplas que incluyen los valores que cada uno de los atributos toma para cada elemento de la relación. Una relación se representa como una tabla de dos dimensiones (las columnas son los atributos de la relación y las filas son las tuplas) con un único valor en cada celda intersección (De Miguel Castaño, 2012).

2.4.3. Estadística con R project

(Velasquez, 2008), dice: “el entorno de R-Project se formó con un conjunto de programas de datos, cálculos y gráficos. Algunas de sus características son el almacenamiento y manipulación de datos; operadores para cálculos en vectores y matrices; una variedad de herramientas para el análisis de datos; gráficas de estos; y un lenguaje de programación que incluye:

condicionales, ciclos, funciones recursivas y la posibilidad de importar y exportar datos".

Todo esto logra que R-Project sea un vehículo para el desarrollo de nuevos métodos de análisis interactivo de datos, ya que es muy dinámico y posee muchas técnicas estadísticas, algunas incluidas en el entorno base y otras como bibliotecas (packages).

2.4.4. Ggplot

Ggplot es un paquete de datos, creado por Hadley Wickham y Winston Chang, que se ejecuta en el software libre R. La principal característica de este paquete es ofrecer una forma fácil y estilizada de crear gráficos.

Estructura básica

Cuenta con una estructura básica con cuatro argumentos:

- ✓ El nombre del dataframe de donde se van a obtener los datos
- ✓ Los datos (variable o variables) que se van a graficar en el eje X
- ✓ Los datos (variable o variables) que se van a graficar en el eje Y
- ✓ El tipo de gráfico que se va a utilizar.

Al emplear este paquete es posible utilizar dos funciones: qplot y ggplot. La principal diferencia entre estas dos funciones es que con la segunda se pueden presentar diferentes datos y gráficos en un mismo plano cartesiano; mientra que con la primera sólo se puede generar un gráfico por plano cartesiano. En la Figura siguiente se observa un ejemplo donde se presenta esta diferencia. En el gráfico del panel izquierdo se incluye una gráfica con qplot y en el derecho una con ggplot.

En el ejemplo, la segunda gráfica muestra la misma información de la primera, y además, la de otras líneas que presentan la tendencia, en el mismo periodo, pero para diferentes ítems (las ciudades).

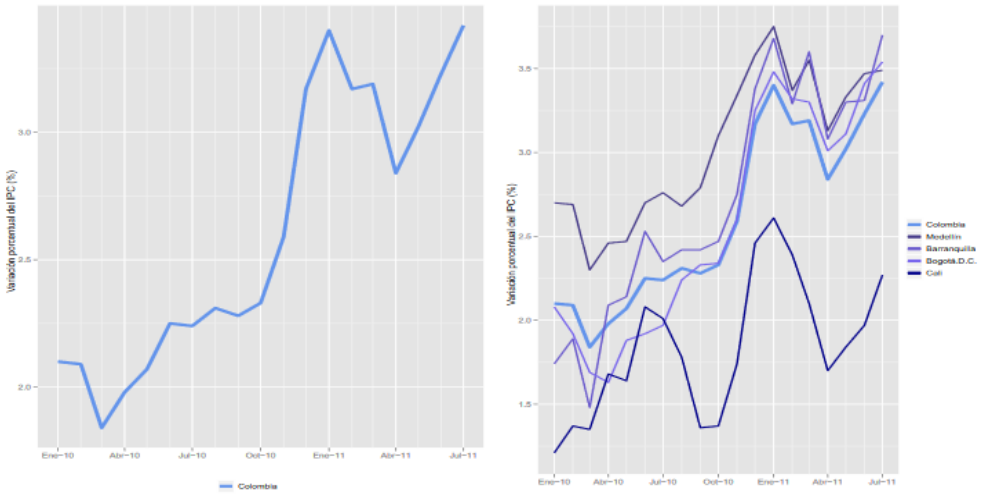


Figura 9. Diferencia entre qplot y ggplot
Fuente: Gonzales, A. (2012). Ggplot: gráficos de alta calidad [figura]. Recuperado de: https://repository.icesi.edu.co/biblioteca_digital/bitstream/10906/68007/5/ggplot_graficos_calidad.pdf

2.4.5. Package dplyr

El paquete dplyr fue desarrollado por Hadley Wickham de RStudio y es una versión optimizada de su paquete plyr. El paquete dplyr no proporciona ninguna nueva funcionalidad a R, en el sentido que todo aquello que se puede hacer con dplyr se puede hacer con la sintaxis básica de R.

Una importante contribución del paquete dplyr es que proporciona una "gramática" (particularmente verbos) para la manipulación y operaciones con data frames. Con esta gramática se puede comunicar mediante el código que es lo que se está haciendo en los data frames a otras personas (asumiendo que conozcan la gramática). Esto es muy útil, ya que proporciona una abstracción que anteriormente no existía. Por último, cabe destacar que las funciones del paquete dplyr son muy rápidas, puesto que

están implementadas con el lenguaje C++.

2.4.5.1. La gramática de dplyr

Algunas de los principales "verbos" del paquete dplyr son:

- ✓ select: devuelve un conjunto de columnas
- ✓ filter: devuelve un conjunto de filas según una o varias condiciones lógicas
- ✓ arrange: reordena filas de un data.frame
- ✓ rename: renombra variables en un data.frame
- ✓ mutate: añade nuevas variables/columnas o transforma variables existentes
- ✓ transmútate: crea variables y además borra todas las otras que no estén especificadas
- ✓ summarise/summarize: genera resúmenes estadísticos de diferentes variables en el data.frame
- ✓ slice: para hacer una selección de casos con base en la posición de las observaciones, por ejemplo, las 10 primeras, o entre la observación 50 y 100.
- ✓ distinct: retiene solo valores únicos de un grupo de datos
- ✓ _%>%: el operador "pipe" es usado para conectar múltiples acciones en una única "pipeline" (tubería)

2.4.5.2. Argumentos comúnes en las funciones plyr

- ✓ El primer argumento es el data.frame

- ✓ Los otros argumentos describen que hacer con el data.frame especificado en el primer argumento, se puede referir a las columnas en el data.frame directamente sin utilizar el operador $, es decir sólo con el nombre de la columna/variable.
- ✓ El valor de retorno es un nuevo data.frame.
- ✓ Los data.frames deben estar bien organizados/estructurados, es decir, debe existir una observación por columna y, cada columna representar una variable, medida o característica de esa observación. Para ello, es muy útil es uso del paquete tidy.

2.4.5.3. Instalación del paquete dplyr

Se instala el paquete desde CRAN con el siguiente comando:

```
## Instalación desde CRAN

install.packages("dplyr")
```

después de la instalación es necesario cargar el paquete con el siguiente comando:

```
library(dplyr)
```

para consultar la documentación de las funciones se usa:

```
?select

?filter

?arrange

?mutate

?summarise

?group_by
```

2.4.5.4. Encadenar operaciones

El paquete dplyr tiene una forma de encadenar operaciones para hacerla en una sola sentencia, de esta forma sólo se hace referencia una sola vez al data.frame el cual no se debe mencionar en las sucesivas funciones de dplyr. El operador para hacer el enlace entre funciones es %>%.

```
newdata ← datos %>%
    select (mpg, cyl, am) %>%
    mutate(log_mpg == log(mpg)) %>%
    arrange(mpg) %>%
    filter(am == 1)
```

Su representación equivalente:

```
newdata ← select (datos, mpg, cyl, am)
newdata ← mutate(newdata, log_mpg == log(mpg))
newdata ← arrange(newdata, mpg)
newdata ← filter(newdata, am == 1)
```

2.5. DEFINICIÓN DE TÉRMINOS BÁSICOS

2.5.1. ¿Qué es R?

R es un dialecto de S. S es un lenguaje que se desarrolló para el análisis de datos, cálculos estadísticos, simulación y gráficos. Además, es un lenguaje de programación de tipo general. S-Plus es la versión comercial de S. R es la versión en código abierto y gratuita de S.

2.5.2. Variables

Las más elementales que nos encontraremos contendrán números (enteros o reales) o tiras de caracteres. Los nombres de las variables empiezan por una letra que puede ir seguida de más letras, dígitos o los símbolos punto (.) y subrayado (_). Las letras mayúsculas y minúsculas son tratadas como caracteres distintos. Para dar un valor a una variable se usa el símbolo de asignación <-. En R, a diferencia de otros lenguajes, no es necesario declarar a priori el nombre y el tipo de las variables. Para crear una variable denominada r que valga 2.5 basta con escribir: r <- 2.5

2.5.3. Proyecto R

R es un software para el análisis estadístico de datos considerado como uno de los más interesantes. Apoyan esta opinión la vasta variedad de métodos estadísticos que cubre, las capacidades gráficas que ofrece y, también muy importante, el hecho de ser un software libre, es decir, gratuito. El mayor inconveniente que podría presentar frente al software más utilizado en nuestro medio es el hecho de funcionar mediante comandos, lo que para algunos usuarios puede resultar engorroso. Para solventar esta dificultad existe un paquete llamado **R Commander** que permite utilizar R sin tener que escribir los comandos.

2.5.4. Paquetes CRAN

R funciona con paquetes de programación, los cuales están disponibles en una Red Comprehensiva de Archivos R (Comprehensive Archive Network, CRAN) en sitios Web llamados MIRROR – sitios que contienen réplicas exactas de R – desde los cuales los usuarios pueden descargarlos.

Actualmente hay 10563 paquetes disponibles para ser descargados y 152 MIRROS en 49 paises de los cinco continentes, con siete paises de américa latina: Argentina, Brazil, Chile, Colombia, Ecuador, México y Venezuela.

2.5.5. Base de datos

Una Base de Datos es una colección de archivos relacionados que almacenan tanto una representación abstracta del dominio de un problema del mundo real cuyo manejo resulta de interés para una organización, como los datos correspondientes a la información acerca del mismo. Tanto la representación como los datos están sujetos a una serie de restricciones, las cuales forman parte del dominio del problema y cuya descripción está también almacenada en esos archivos.

2.5.6. MySQL

MySQL es un sistema de gestión de bases de datos relacionales. Una base de datos relacional almacena datos en tablas separadas en lugar de poner todos los datos en un gran almacén. Esto añade velocidad y flexibilidad. La parte SQL de “MySQL” se refiere a “Structured Query Language”. SQL es el lenguaje estandarizado más común para acceder a bases de datos y está definido por el estándar ANSI/ISO SQL. El estándar SQL ha evolucionado desde 1986 y existen varias versiones. “SQL-92” se refiere al estándar del 1992, “SQL: 1999” se refiere a la versión de 1999, y “SQL: 2003” se refiere a la versión actual del estándar, MySQL AB (2006).

2.5.7. Script

es un documento que contiene instrucciones, escritas en códigos de

programación. El script es un lenguaje de programación que ejecuta diversas funciones en el interior de un programa de computador.

Los scripts se encargan de cumplir las siguientes funciones:

- ✓ Combinar componentes
- ✓ Interactuar con el sistema operativo o con el usuario
- ✓ Controlar un determinado progama o aplicación
- ✓ Configurar o instalar sistemas operacionales, especialmente en los juegos, se usa para controlar las acciones de los personajes

2.5.8. SQL

SQL (Structured Query Language) es un lenguaje de programación estándar e interactivo para la obtención de información desde una base de datos y para actualizarla. Aunque SQL es a la vez un ANSI y una norma ISO, muchos productos de bases de datos soportan SQL con extensiones propietarias al lenguaje estándar. Las consultas toman la forma de un lenguaje de comandos que permite seleccionar, insertar, actualizar, averiguar la ubicación de los datos, y más. También hay una interfaz de programación.

2.5.9. RStudio

Es un entorno de desarrollo integrado (IDE) para R, que incluye una consola, un editor resaltado de sintaxis que soporta la ejecución directa de código, así como herramientas para representar los datos gráficamente, depurar y gestionar el espacio de trabajo.

2.5.10. Análisis de series de tiempo

Análisis de una secuencia de medidas hechas a intervalos específicos. El tiempo es usualmente la dimensión dominante de los datos.

2.5.11. Análisis prospectivo de datos

Análisis de datos que predice futuras tendencias, comportamientos o eventos basado en datos históricos.

2.5.12. Análisis exploratorio de datos

Uso de técnicas estadísticas tanto gráficas como descriptivas para aprender acerca de la estructura de un conjunto de datos.

2.5.13. Análisis retrospectivo de datos

Análisis de datos que provee una visión de las tendencias, comportamientos o eventos basado en datos históricos.

2.5.14. Clasificación

Proceso de dividir un conjunto de datos en grupos mutuamente excluyentes de tal manera que cada miembro de un grupo esté lo "más cercano" posible a otro, y grupos diferentes estén lo "más lejos" posible uno del otro, donde la distancia está medida con respecto a variable(s) específica(s) las cuales se están tratando de predecir. Por ejemplo, un problema típico de clasificación es el de dividir una base de datos de compañías en grupos que son lo más homogéneos posibles con respecto a variables como "posibilidades de crédito" con valores tales como "Bueno" y "Malo".

2.5.15. Modelo predictivo

Estructura y proceso para predecir valores de variables especificadas en un conjunto de datos.

2.5.16. Regresión lineal

Técnica estadística utilizada para encontrar la mejor relación lineal que encaja entre una variable seleccionada (dependiente) y sus predicados (variables independientes).

2.5.17. Datos

Los datos son la fuente de la que se obtienen las variables, las relaciones entre ellas, el conocimiento inducido o los patrones de comportamiento identificados, convirtiéndose en un elemento vital de todo análisis predictivo.

CAPÍTULO III

METODOLOGÍA

3.1. TÉCNICAS Y MATERIALES

Dado que la investigación implica la producción de un nuevo conocimiento en la solución de problemas prácticos (Arias, 2004), ésta se enmarca dentro de la investigación aplicada, y la metodología a usar es CRISP-DM por su uso generalizado en proyectos de minería de datos y que abarca precisamente los objetivos de la investigación.

3.2. HERRAMIENTAS

Para el desarrollo del análisis predictivo se utilizó las siguientes herramientas:

R Statictics

RStudio

Package ggplot

Package dplyr

SQL Manager 2005 for MySQL

MySQL Database versión 5.0.16

3.2.1. El entorno de desarrollo

Básicamente toda la implementación de los scripts se desarrollaron con el software RStudio que incluye un editor de código, herramientas de visualización y depuración. Fue descargado de https://www.rstudio.com/.

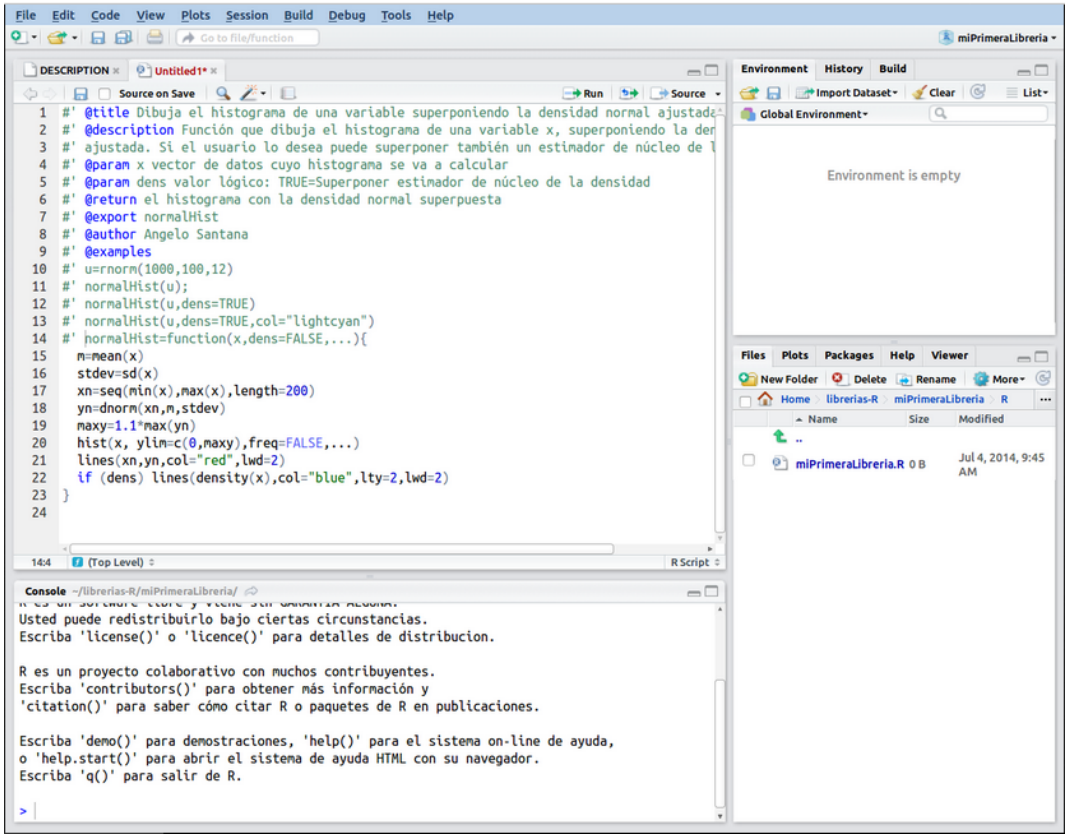


Figura 10. Entorno de RStudio
Fuente: (Santana & Hernadez, 2016). Creación de paquetes R y RStudio. Recuperado de: http://www.dma.ulpgc.es/profesores/personal/stat/cursoR4ULPGC/16-crearPaquetes.html

3.3. POBLACIÓN

La población estuvo conformada por 135,836 observaciones comprendidos desde el examen cepreuna con fecha 31 de marzo del 2013 al examen general del 22 de enero del 2017 de los diferentes procesos de admisión del departamento de Puno. Los procesos de admisión considerados son ordinario (general y cepreuna).

Cuadro 1. Población de los procesos ordinarios

	Modalidad	Fecha	Total
2	CEPREUNA	31_03_2013	3281
3	GENERAL	07_04_2013	8957
4	CEPREUNA	14_07_2013	2617
5	GENERAL	18_08_2013	8453
6	CEPREUNA	08_09_2013	1146
7	CEPREUNA	22_12_2013	2231
8	GENERAL	12_01_2014	8069
10	CEPREUNA	16_03_2014	2565
11	GENERAL	01_06_2014	8097
12	CEPREUNA	15_06_2014	2203
13	CEPREUNA	31_08_2014	1988
14	GENERAL	14_09_2014	7680
15	CEPREUNA	14_12_2014	2089
16	GENERAL	18_01_2015	8247
18	CEPREUNA	29_03_2015	3507
19	GENERAL	07_06_2015	7930
20	CEPREUNA	09_08_2015	2822
21	GENERAL	06_09_2015	6889
22	CEPREUNA	08_11_2015	1873
23	CEPREUNA	10_01_2016	1492
24	GENERAL	31_01_2016	7833
26	CEPREUNA	20_03_2016	2935
27	GENERAL	15_05_2016	6267
28	CEPREUNA	26_06_2016	2851
29	GENERAL	21_08_2016	8677
30	CEPREUNA	18_09_2016	2342
31	CEPREUNA	18_12_2016	2857
32	GENERAL	22_01_2017	8665

3.4. MÉTODO DE LOS MÍNIMOS CUADRADOS

La regresión lineal es una técnica para determinar la mejor línearecta que pasa por un conjunto de observaciones definidas por puntos (x_1, x_1), (x_2, x_2), …, (x_n, x_n). Para algunas casos se usará la siguiente ecuación lineal:

$y = a_0 + a_1x + e$

Donde:

y → es el valor verdaderdo

a_0 y a_1 → son la ordenada al origen y la pendiente de la línea recta respectivamente.

e → es el error y diferencia entre el modelo y las observaciones, el cual se representa al reordenar la ecuación como: $e = y - a_0 – a_1x$

$a_0 + a_1x$ → el valor pronosticado de la variable dependiente.

Cuantificación del error

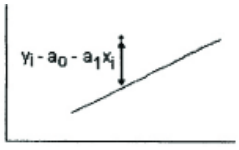
$$\bar{y} = \frac{\sum y_i}{n} \quad \text{de } i = 1 \ldots n$$

Desviación estándar

$$S_y = \sqrt{\frac{S_t}{n-1}} \quad \text{donde} \quad S_t = \sum (y_i - \bar{y})^2$$

Cuantificación del error de la línea de regresión

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 X)^2$$



La desviación estándar para la línea de regresión se puede dererminar por la siguiente fórmula:

$$S_{y/x} = \sqrt{\frac{S_r}{n-2}}$$

Donde $S_{x/y}$ es llamado el error estándar del estimado, la notación del suíndice y/x designa que el error es para un valor predicho de Y correspondiente a un valor de X.

Coeficiente de correlación

El coeficiente de correlación cuantifica la mejora o reducción del error originado por la representación de los datos por medio de una línea recta en vez de como un valor promedio.

$$r = \sqrt{\frac{S_t - S_r}{S_t}}$$

R^2 es el coeficiente de determinación y si es cercano a una unidad indica un buen ajuste, mientras que si es próximo a cero, el ajuste es pobre.

3.5. MÉTODO DE LOS MÍNIMOS CUADRADOS PARA EL CASO POLINOMIAL

Aunque algunos datos exhiben un patrón marcado, son pobremente representados por una línea recta, entonces una curva será la más adecuada para ajustarse a los datos; una alternativa es ajustar polinomios a los datos mediante regresión polinomial. La ecuación polinomial de grado n es:

$$y = a_0 + a_1 x + a_2 x^2 + \ldots\ldots + a_n x^n = \sum_{i=0}^{n} a_i x^i$$

Aplicando el método de los mínimos cuadrados la curva propuesta es:

$$y_p = a_0 + a_1 x + a_2 x^2 + \ldots\ldots + a_n x^n + e$$

Donde a_i son coeficientes y e es el error. La estrategia para minimizar la suma de los cuadrados de los residuos (S_r), entre la y medida y la y calculada con el modelo lineal, está dado por:

$$S_r = \sum e_i^2 = \sum (y_{i,\,medida} - y_{i,\,modelo})^2 = \sum (y_i - a_0 - a_1 x_i - a_2 x_i^2 - \ldots\ldots - a_n x_i^n)^2$$

El error estándar del estimado se formula como:

$$s_{y/x} = \sqrt{\frac{S_r}{m - (n + 1)}}$$

Donde m es el número de puntos.

El sistema de ecuaciones normales es de la forma:

$$S_x \, a = S_{xy}$$

Donde S_x es la matriz de sumatorias de potencias de x.

$$S_x = \begin{bmatrix} m & \sum x & \sum x^2 & \ldots & \sum x^n \\ \sum x & \sum x^2 & \sum x^3 & \ldots & \sum x^{n+1} \\ \sum x^2 & \sum x^3 & \sum x^4 & \ldots & \sum x^{n+2} \\ . & & & & \\ . & & & & \\ \sum x^n & \sum x^{n+1} & \sum x^{n+2} & \ldots & \sum x^{2n} \end{bmatrix}$$

$$a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ . \\ . \\ a_n \end{bmatrix} \qquad S_{xy} = \begin{bmatrix} \sum y \\ \sum yx \\ \sum yx^2 \\ . \\ . \\ \sum yx^n \end{bmatrix}$$

A → vector de coefientes, las constantes del polinomio

Sxy → vector de sumatorias de potencias de x con y's.

3.6. METODOLOGÍA CRISP-DM

En ésta etapa se desarrollarán cada una de las fases de la metodología CRISP-DM al problema planteado. Estas fases se enumeran a continuación tal como se indican en el marco teórico.

Comprensión del negocio

Entendimiento de los objetivos y requerimientos del proyecto.

Definición del problema de minería de datos.

Comprensión de los datos

Obtención del conjunto de datos iniciales.

Exploración del conjunto de datos.

Identificar las características de calidad de los datos.

Identificar los resultados iniciales obvios.

Preparación de los datos

Selección de los datos.

Limpieza de los datos.

Modelamiento

Implementación con dplyr, gglot de análisis de datos.

Evaluación

Determinar si los resultados coinciden con los objetivos del negocio

Identificar los temas del negocio que deberían haberse abordado

Implementación

Instalar los modelos resultantes en la práctica.

Configuración para minería de datos de forma repetida ó continua.

CAPÍTULO IV

RESULTADO Y DISCUSIÓN

4.1. APLICACIÓN DE LA METODOLOGÍA CRISP-DM

En esta parte se aplicará cada una de las fase de la metodología CRISP-DM al problema de estudio, se desarrollará la metodología tal como se indica en el documento original.

4.1.1. Comprensión del negocio

4.1.1.1. Determinar los objetivos del negocio

El objetivo es realizar las predicciones de la tendencia de las Escuelas Profesionales y que éstas sean lo más precisas posibles a partir de los datos de los postulantes e ingresantes de la Universidad.

Contexto

Considerando la situación del negocio (universidad) el presente trabajo tiene lugar en la oficina de la Comisión Central Admisión de la Universidad Nacional del Altiplano, donde la información es exclusivamente de los procesos de admisión, no se considera información de estudiantes que estén llevando cursos de pregrado.

Objetivos del negocio

El objetivo es realizar un análisis de la información de postulantes e ingresantes para predecir la tendencia del futuro de las diferentes Escuelas Profesionales y del rendimiento brindado en las escuelas de educación secundaria, públicas y privadas de la región de Puno.

Esta información puede ser útil para las Escuelas Profesionales ya que permitirá tomar algunas medidas para captar más postulantes, del mismo modo, los colegios de educación secundaria preocuparse por una mejor formación de los escolares.

Criterios de éxito

Desde el punto de vista del negocio se considera la posibilidad de éxito el predecir la cantidad de postulantes e ingresantes de cada Escuela Profesional, igualmente, elevar el nivel de ingresantes de las escuelas de educación secundaria.

4.1.1.2. Evaluación de la situación

Se cuenta con una base de datos desarrollada con el gestor MySQL detallada de los postulantes e ingresantes administrados por la sub comisión de Cómputo de la oficina de la Comisión Central de Admisión desde el proceso de admisión extraordinario del 17 de marzo del 2013 hasta el proceso de admisión general 22 de enero del 2017. Ésta información incluye datos como: generales de ley, lugar de procedencia, colegio de procedencia, tipo de colegio, género, escuela profesional a la que se presentó y otros datos personales que puedan ser útiles a la hora de realizar minería de datos.

Inventario de recursos

En cuanto a los recursos de software, se dispone del software R y el entorno RStudio de uso libre, para realizar las tareas de minería de datos sobre la base de datos MySQL donde se encuentra la información.

La fuente de datos de postulantes e ingresantes es desde marzo del 2013 hasta enero del 2017

Requisitos, supuestos y restricciones

No existe ningún tipo de restricción frente a la información de postulantes, ya que se va a trabajar con agrupaciones donde no se va a involucrar información personal del postulante.

Costos y beneficios

La información de ésta investigación no supone ningún costo adicional a la Universidad ya que éstos datos corresponden a la misma desde el momento en que el postulante registra sus datos vía la página Web de la Comisión Central de Admisión para cada proceso de admisión.

Frente a los beneficios, ésta no genera ningún tipo de beneficio económico, el único beneficio es lograr algún tipo de medida correctiva para ver la situación de postulantes e ingresantes en cada Escuela Profesional, igualmente, las escuelas de educación secundaria ver la situación de la institución educativa frente a la cantidad de ingresantes y tomar alguna medida que contribuya a la formación de los escolares.

4.1.1.3. Determinar los objetivos de la minería de datos

Entre los objetivos de minería de datos se tiene:

- ✓ Predecir la cantidad de postulantes que tendrán aquellas Escuelas Profesionales que van decayendo en número de postulantes e ingresantes.
- ✓ Predecir la cantidad de postulantes e ingresantes por tipo de colegio, zonas urbanas y rurales, de las escuelas públicas y privadas.
- ✓ Identificar aquellas escuelas de educación secundaria que tienen una mejor formación y por ende, mayor cantidad de ingresantes y predecir el número de ingresantes en los siguientes procesos de admisión.

4.1.1.4. Realizar el plan del proyecto

El proyecto se dividió en las siguientes etapas para tener una mejor organización y cumplir con las metas establecidas:

Etapa 1: análisis de la estructura de los datos y la información de las tablas de la base de datos.

Etapa 2: ejecución de consultas para tener muestras representativas de los datos.

Etapa 3: preparación de los datos que comprende selección, limpieza, conversión y formateo, esto para facilitar la minería de datos sobre la información.

Etapa 4: elección de las técnicas de modelado y ejecución de las

mismas con la información obtenida.

Etapa 5: análisis de los datos obtenidos.

Etapa 6: generar informes con los resultados obtenidos en función a los objetivos del negocio y los criterios de éxito establecidos.

Etapa 7: preparación de los resultados finales.

4.1.1.5. Evaluación inicial de herramientas y técnicas

La herramienta que se usó para llevar a cabo ésta minería de datos es el paquete RMySQL y en cuanto a las técnicas que se van a emplear para la extracción de conocimiento es:

- ✓ Predictivas
  - ✓ Clasificación
  - ✓ Regresión
- ✓ Descriptivas
  - ✓ Agrupamiento (clustering)
  - ✓ Reglas de asociación

4.1.2. Comprensión de los datos

En esta segunda fase de la metodología CRISP-DM se realiza la recolección inicial de los datos para poder establecer un primer contacto con el problema, familiarizarse con los datos y averiguar su calidad, así como identificar las relaciones más evidentes para formular las primeras hipótesis.

4.1.2.1. Recolección de datos iniciales

Los datos iniciales de ésta investigación son referentes a postulantes e ingresantes que incluyen dni, apellidos, nombres, fecha de nacimiento, ubigeo de nacimiento, género, escuela profesional a la que postula, nombre del colegio donde estudió, ubigeo del colegio, tipo de colegio, etc., ésta información es fiable ya que los datos son precisos y sin errores lo que permitirá realizar las predicciones lo más reales posibles.

A continuación, se detallan los datos adquiridos:

Área de la Escuela Profesional

Tiene un código numérico único y un nombre.

Escuela Profesional

Cada Escuela Profesional está identificado por un código y el área al que pertenece, así como el nombre de la carrera y se encuentra relacionada con el área al que pertenece la Escuela Profesional.

Postulante

Cada postulante tiene un código el cual es único, éste es el documento nacional de identidad; además se encuentra relacionada con la Escuela Profesional, el ubigeo de nacimiento y el ubigeo de la escuela donde estudio.

Departamento

- ✓ Cada departamento está identificado por un código alfanúmerico y el nombre del departamento.

Provincia

- ✓ Cada provincia está identificada por un código alfanúmerico y está relacionada con el departamento.

Distrito

- ✓ Cada distrito está identificado por un código alfanúmerico y está relacionada con la provincia.

Escuela de educación secundaria

- ✓ Cada escuela está identificada por un código único que es el que se encuentra en la página Web del ministerio de educación (ESCALE), además cuenta con el ubigeo del colegio, el tipo de colegio y la zona a la que pertenece.

Modalidad del proceso de admisión

- ✓ Existen dos procesos de admisión que es ordinario y extraordinario.

Los atributos con las que cuenta ésta información y que serán útiles para realizar la minería de datos son:

- ✓ Código de la Escuela Profesional
- ✓ Código del área de la Escuela Profesional
- ✓ Ubigeo del colegio
- ✓ Tipo de colegio
- ✓ Zona del colegio
- ✓ Género

- ✓ Ubigeo de nacimiento
- ✓ Modalidad del proceso de admisión

Las tablas de las que se recogen los datos necesarios para la minería de datos son:

- ✓ c_carrera
- ✓ c_postulante_fecha
- ✓ c_area
- ✓ c_colegioperu
- ✓ c_departamento
- ✓ c_provincia
- ✓ c_distrito
- ✓ c_modalidad
- ✓ c_modalidad_carrera
- ✓ c_usuarios

4.1.2.2. Descripción de los datos

El modelo de datos fue el punto de partida para representación de la base de datos y se realizó con el modelo E/R con sus atributos, en las siguientes figuras se puede apreciar el modelo entidad-relación (figura 11) y el esquema relacional de ésta base de datos (figura 12). Para generar ésta relación se ha empleado la herramienta EMS MySQL, que además de administrar la base de datos también puede generar modelos físicos de la base de datos.

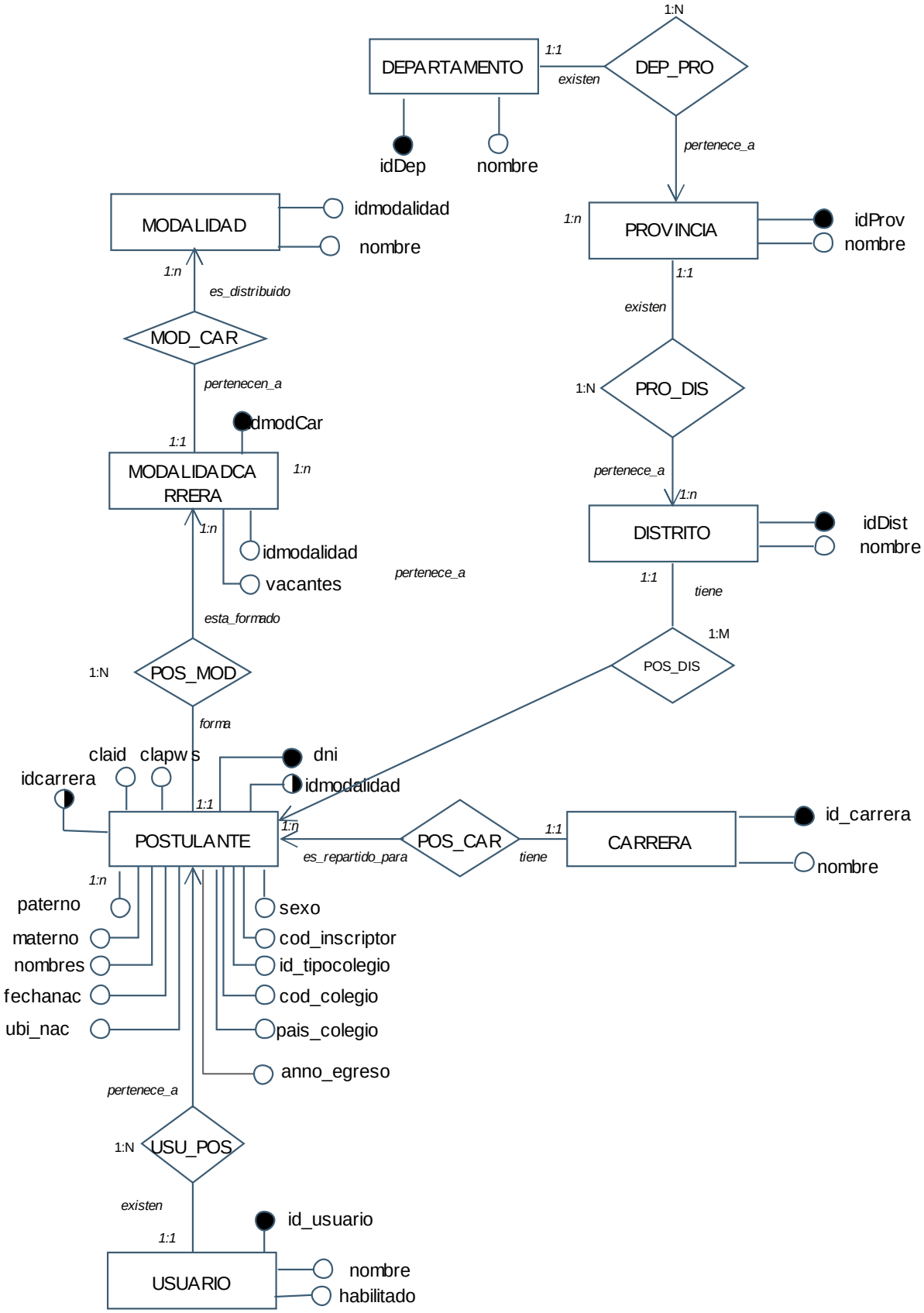


Figura 11. Modelo físico de la base de datos

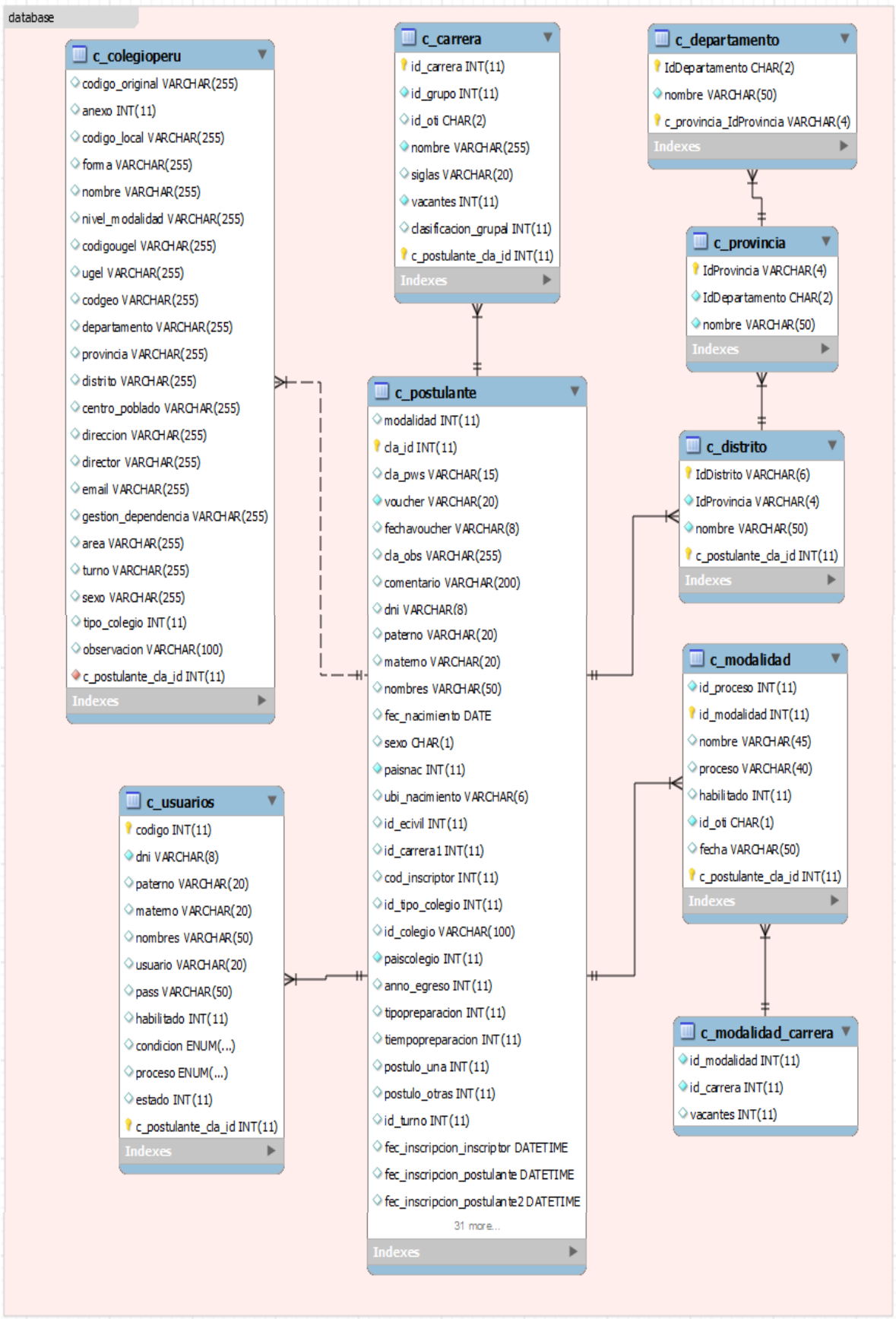


Figura 12. Modelo físico de la base de datos

El estándar que se siguió para la elaboración del modelo físico de la base de datos fue la siguiente:

TABLAS: Alias sistema +_+ nombre de la tabla

Ejemplo:

- ✓ c_carrera: será la tabla donde se almacenará información de las Escuelas Profesionales de la Universidad Nacional del Altiplano.

El diccionario de datos contiene “datos acerca de los datos” es decir, definiciones de otros objetos del sistema, en lugar de simples datos en bruto (Date, 2001). El detalle del diccionario de datos de cada uno de los campos de las nueve tablas de la base de datos se describe en el anexo 1.

4.1.2.3. Exploración de datos

En esta etapa se realiza la exploración de datos sobre la información extraída desde marzo del 2013 a enero del 2017. Los primeros resultados van a mostrar información estadística y principalmente sirve para determinar la consistencia y completitud de datos.

Para realizar ésta exploración se ha empleado el paquete RMySQL, que establece un puente un puente de conexión entre la base de datos MySQL y el software estadístico R. Para la conexión se usó la función dbConnect de la librería RMySQL.

```
Library(RMySQL)
con <- dbConnect(RMySQL::MySQL(),
                 dbname = "database",
                 user="root",
                 password = "123",
                 port = 3306)
```

Nombre de las variables globales que contienen los nombres de todas las tablas de la base de datos:

```
proc_general<-
c("c_postulante_01g_22_01_2017",
  "c_postulante_01g_21_08_2016",
  …)
proc_cepreuna<-
c("c_postulante_02c_18_12_2016",
  "c_postulante_02c_18_09_2016",
  …)
proc_extraordinario<-
c("c_postulante_03e_12_03_2016",
  "c_postulante_03e_14_03_2015",
  …)
```

Se implementó la función que retorna las escuelas profesionales y el área al que pertenecen a través de consultas SQL inscrustadas.

```
function()
{
  SQL <- paste("SELECT id_carrera, id_grupo, nombre",
               "FROM c_carrera ORDER BY orden", sep = " ")
  return (data.frame(dbGetQuery(con, SQL)))

}
```

Instrucciones que usan el paquete dplyr, ggplot para generar la cantidad de postulantes e ingresantes del departamento de Puno por proceso de admisión.

```
 7  vingreso<-"NO"
 8  vcarrera <- 1
 9  vproceso <- c("CEPREUNA")
10  grado <- 6
11  # estado indica si son postulantes o ingresantes
12  if(vingreso=="NO"){
13    estado<-c("SI","NO")
14  }else{
15    estado<-c("SI")
16  }
17  if(length(estado)==1){
18    tipo<-"ingresantes"
19  }else{
20    tipo<-"postulantes"
21  }
```

```
reg<-todo %>%
  filter(depart_nac=="PUNO" & ingreso %in% estado) %>%
  select(num, modalidad_tipo, id_carrera, fecha,
         escuela_profesional, ingreso) %>%
  group_by(num, modalidad_tipo, fecha, escuela_profesional,
           id_carrera) %>%
  summarise(total=n()) %>%
  arrange(num, fecha, id_carrera)  %>%
  select(num, modalidad_tipo, fecha, id_carrera,
         escuela_profesional, total)

datacarrera<-reg %>% filter(id_carrera==vcarrera &
                              modalidad_tipo %in% vproceso)
if(grado==0){
  miformula<-y~x
}else{
  miformula<-y~poly(x,grado)
}
# convirtiendo texto a fecha
datacarrera$fecha<-as.Date(datacarrera$fecha,"%d_%m_%Y")

p<-ggplot(data = datacarrera, aes(x=reorder(fecha, num), y=total,
                                  color=modalidad_tipo)) + theme_bw()
p<-p + geom_point(shape=1)
p<-p + theme(plot.title=element_text(hjust=0.5, size=8, face="bold"),
             text = element_text(size = 9),
             axis.text.x = element_text(angle=90, hjust=0.5,size=8))

p<-p + geom_point(aes(colour=factor(modalidad_tipo))) +
  geom_line(aes(group = modalidad_tipo))
p<-p + geom_smooth(aes(group=modalidad_tipo), method="lm",
                   formula = miformula,
                   colour="blue", fill=NA)
p<-p + ggtitle(paste("Cantidad de ", tipo,
                     datacarrera$escuela_profesional))
p<-p + labs(x = "", y=tipo, colour = "Proceso")
p<-p + geom_text(aes(label=total),hjust=0.5, vjust=-1.0, size=2.5)

m <- lm(miformula, data = data.frame("x"= as.numeric(datacarrera$fecha),
                                     "y"= datacarrera$total))
my.eq <- as.character(signif(as.polynomial(coef(m)), 3))
label.text <- paste(gsub("x", "~italic(x)", my.eq, fixed = TRUE),
                    paste("italic(R)^2",
                          format(summary(m)$r.squared, digits = 2),
                          sep = "~`=`~"),
                    sep = "~~~~")

#p + facet_wrap(~datacarrera$modalidad_tipo)

p + annotate(geom = "text", x = 1.0, y = max(datacarrera$total)+1,
             label = label.text,
             family = "serif", hjust = 0, parse = TRUE, size = 3)
```

Cantidad de postulantes e ingresantes de los procesos general y cepreuna de las Escuelas Profesionales del área biomédicas

Cuadro 2. Med. Vet. y Zoot., porcentaje de ingresantes, general

	modalidad	fecha	escuela	post	ing	porc.ing
8	GENERAL	06_09_2015	Medicina Veterinaria y Zootecnia	81	33	40.74 %
5	GENERAL	14_09_2014	Medicina Veterinaria y Zootecnia	105	30	28.57 %
4	GENERAL	01_06_2014	Medicina Veterinaria y Zootecnia	113	27	23.89 %
3	GENERAL	12_01_2014	Medicina Veterinaria y Zootecnia	83	16	19.28 %
1	GENERAL	07_04_2013	Medicina Veterinaria y Zootecnia	110	21	19.09 %
11	GENERAL	21_08_2016	Medicina Veterinaria y Zootecnia	85	13	15.29 %
12	GENERAL	22_01_2017	Medicina Veterinaria y Zootecnia	98	13	13.27 %
9	GENERAL	31_01_2016	Medicina Veterinaria y Zootecnia	86	9	10.47 %
10	GENERAL	15_05_2016	Medicina Veterinaria y Zootecnia	73	6	8.22 %
2	GENERAL	18_08_2013	Medicina Veterinaria y Zootecnia	132	8	6.06 %
6	GENERAL	18_01_2015	Medicina Veterinaria y Zootecnia	108	3	2.78 %
7	GENERAL	07_06_2015	Medicina Veterinaria y Zootecnia	100	1	1 %

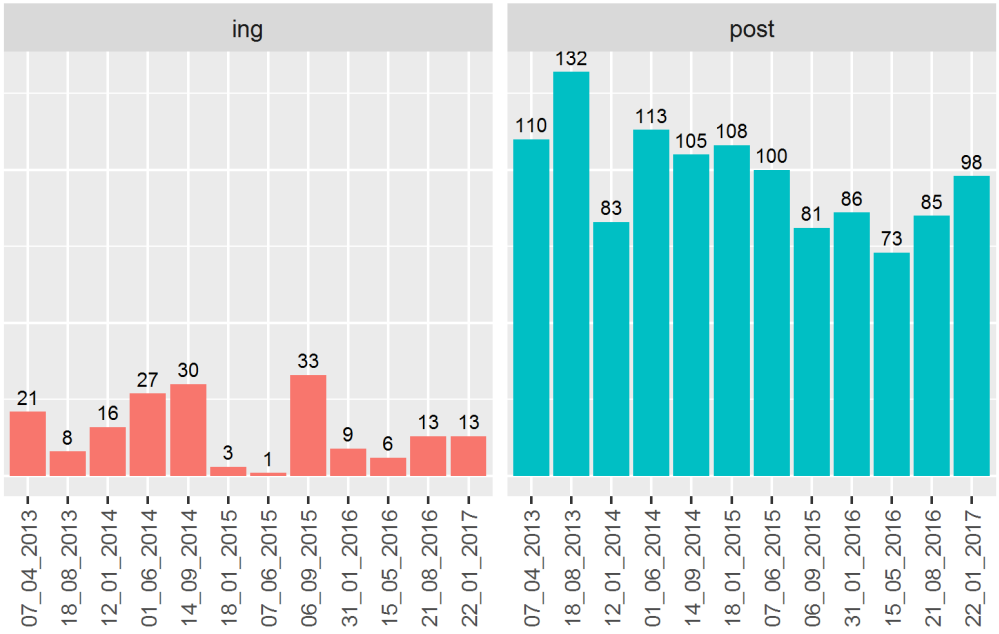


Figura 13. Med. Vet. Zoot, cant. postulantes e ingresantes, examen general

De la figura 13 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 40% y el mínimo fue del 1%.

Cuadro 3. Med. Vet. y Zoot., porcentaje de ingresantes, cepreuna

	modalidad	fecha	escuela	post	ing	porc.ing
3	CEPREUNA	08_09_2013	Medicina Veterinaria y Zootecnia	7	7	100 %
6	CEPREUNA	15_06_2014	Medicina Veterinaria y Zootecnia	25	19	76 %
14	CEPREUNA	26_06_2016	Medicina Veterinaria y Zootecnia	20	14	70 %
12	CEPREUNA	10_01_2016	Medicina Veterinaria y Zootecnia	14	8	57.14 %
2	CEPREUNA	14_07_2013	Medicina Veterinaria y Zootecnia	23	13	56.52 %
4	CEPREUNA	22_12_2013	Medicina Veterinaria y Zootecnia	23	12	52.17 %
15	CEPREUNA	18_09_2016	Medicina Veterinaria y Zootecnia	20	9	45 %
11	CEPREUNA	08_11_2015	Medicina Veterinaria y Zootecnia	20	8	40 %
13	CEPREUNA	20_03_2016	Medicina Veterinaria y Zootecnia	25	10	40 %
16	CEPREUNA	18_12_2016	Medicina Veterinaria y Zootecnia	22	8	36.36 %
1	CEPREUNA	31_03_2013	Medicina Veterinaria y Zootecnia	32	11	34.38 %
8	CEPREUNA	14_12_2014	Medicina Veterinaria y Zootecnia	18	6	33.33 %
5	CEPREUNA	16_03_2014	Medicina Veterinaria y Zootecnia	41	12	29.27 %
9	CEPREUNA	29_03_2015	Medicina Veterinaria y Zootecnia	31	9	29.03 %
7	CEPREUNA	31_08_2014	Medicina Veterinaria y Zootecnia	14	3	21.43 %
10	CEPREUNA	09_08_2015	Medicina Veterinaria y Zootecnia	28	6	21.43 %

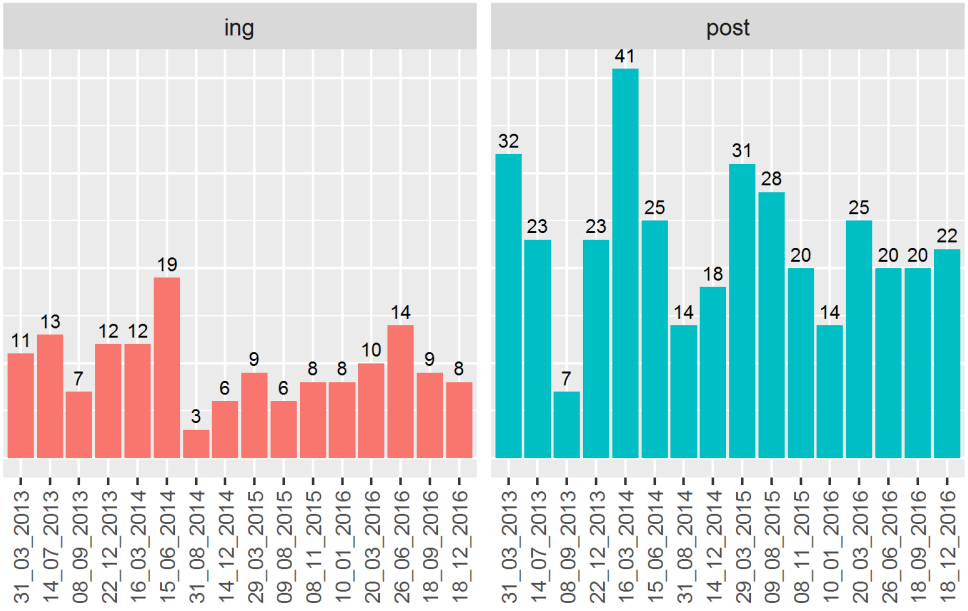


Figura 14. Med. Vet. Zoot, número postulantes e ingresantes, examen cepreuna

De la figura 14 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 100% en el proceso de admisión de fecha setiembre del 2013 y el mínimo fue del 21.43% de fecha agosto del 2015.

Cuadro 4. Biología, porcentaje de ingresantes, general

	modalidad	fecha	escuela	post	ing	porc.ing
5	GENERAL	14_09_2014	Biología	80	28	35 %
4	GENERAL	01_06_2014	Biología	72	21	29.17 %
8	GENERAL	06_09_2015	Biología	62	18	29.03 %
2	GENERAL	18_08_2013	Biología	64	17	26.56 %
3	GENERAL	12_01_2014	Biología	95	25	26.32 %
9	GENERAL	31_01_2016	Biología	69	17	24.64 %
10	GENERAL	15_05_2016	Biología	59	11	18.64 %
12	GENERAL	22_01_2017	Biología	94	16	17.02 %
11	GENERAL	21_08_2016	Biología	74	12	16.22 %
6	GENERAL	18_01_2015	Biología	96	13	13.54 %
1	GENERAL	07_04_2013	Biología	111	14	12.61 %
7	GENERAL	07_06_2015	Biología	66	3	4.55 %

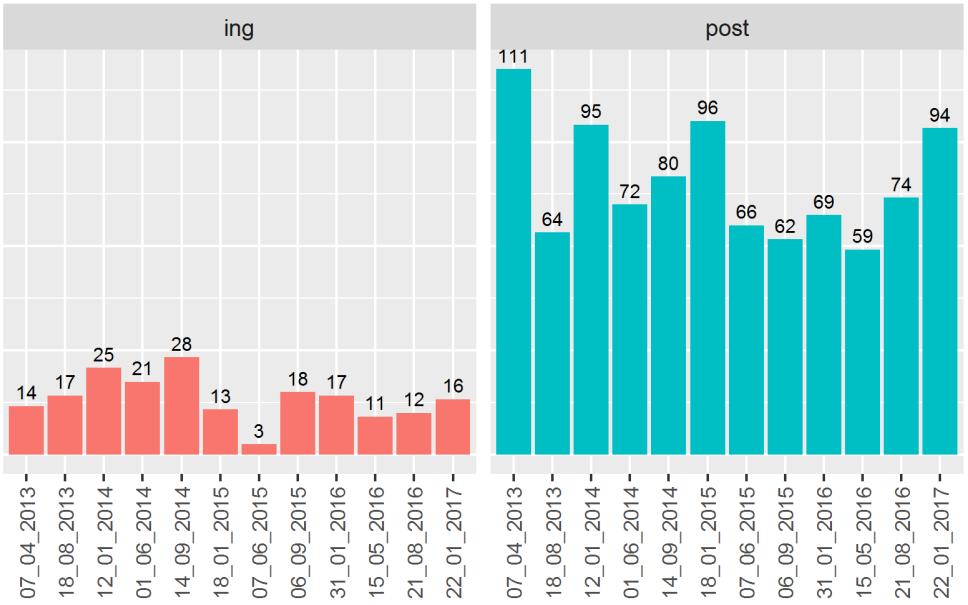


Figura 15. Biología, número postulantes e ingresantes, examen general

De la figura 15 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 35% en el proceso de admisión de fecha setiembre del 2014 y el mínimo fue del 4.55% de fecha junio del 2015.

Cuadro 5. Medicina Humana, porcentaje de ingresantes, general

	modalidad	fecha	escuela	post	ing	porc.ing
4	GENERAL	01_06_2014	Medicina Humana	439	19	4.33 %
2	GENERAL	18_08_2013	Medicina Humana	413	17	4.12 %
3	GENERAL	12_01_2014	Medicina Humana	467	19	4.07 %
10	GENERAL	15_05_2016	Medicina Humana	366	7	1.91 %
7	GENERAL	07_06_2015	Medicina Humana	438	8	1.83 %
8	GENERAL	06_09_2015	Medicina Humana	386	7	1.81 %
6	GENERAL	18_01_2015	Medicina Humana	373	6	1.61 %
5	GENERAL	14_09_2014	Medicina Humana	382	6	1.57 %
11	GENERAL	21_08_2016	Medicina Humana	465	7	1.51 %
9	GENERAL	31_01_2016	Medicina Humana	400	6	1.5 %
1	GENERAL	07_04_2013	Medicina Humana	326	4	1.23 %
12	GENERAL	22_01_2017	Medicina Humana	435	5	1.15 %

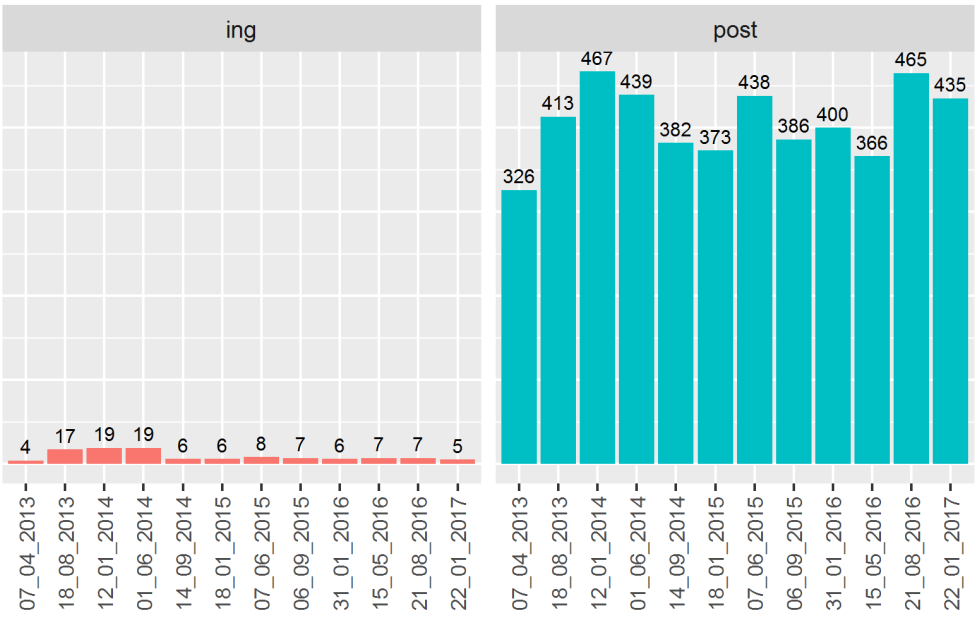


Figura 16. Medicina Humana, número postulantes e ingresantes, examen general

De la figura 16 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 4.33% en el proceso de admisión de fecha junio del 2014 y el mínimo fue del 1.15% de fecha enero del 2017.

Cuadro 6. Medicina Humana, porcentaje de ingresantes, cepreuna

	modalidad	fecha	escuela	post	ing	porc.ing
4	CEPREUNA	22_12_2013	Medicina Humana	195	7	3.59 %
3	CEPREUNA	08_09_2013	Medicina Humana	142	5	3.52 %
11	CEPREUNA	08_11_2015	Medicina Humana	176	5	2.84 %
2	CEPREUNA	14_07_2013	Medicina Humana	183	5	2.73 %
12	CEPREUNA	10_01_2016	Medicina Humana	153	4	2.61 %
6	CEPREUNA	15_06_2014	Medicina Humana	197	5	2.54 %
7	CEPREUNA	31_08_2014	Medicina Humana	199	5	2.51 %
15	CEPREUNA	18_09_2016	Medicina Humana	166	4	2.41 %
9	CEPREUNA	29_03_2015	Medicina Humana	302	7	2.32 %
5	CEPREUNA	16_03_2014	Medicina Humana	309	7	2.27 %
14	CEPREUNA	26_06_2016	Medicina Humana	197	4	2.03 %
16	CEPREUNA	18_12_2016	Medicina Humana	212	4	1.89 %
1	CEPREUNA	31_03_2013	Medicina Humana	268	5	1.87 %
8	CEPREUNA	14_12_2014	Medicina Humana	222	4	1.8 %
10	CEPREUNA	09_08_2015	Medicina Humana	226	4	1.77 %
13	CEPREUNA	20_03_2016	Medicina Humana	286	4	1.4 %

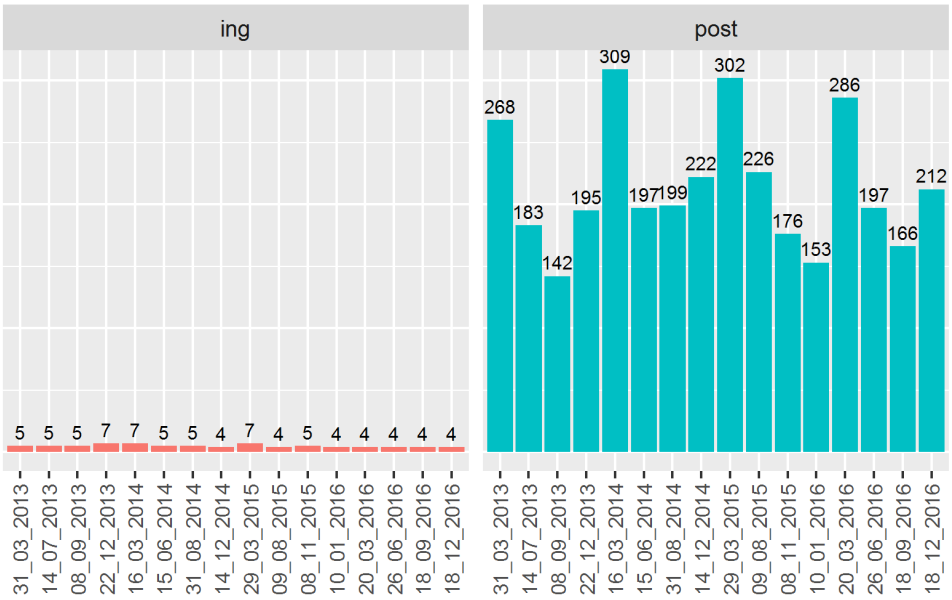


Figura 17. Medicina Humana, número postulantes e ingresantes, examen cepreuna

De la figura 17 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 3.59% en el proceso de admisión de fecha diciembre del 2013 y el mínimo fue del 1.4% de fecha marzo del 2016.

Cantidad de postulantes e ingresantes de los procesos general y cepreuna de las Escuelas Profesionales del área sociales

Cuadro 7. Educación Primaria, porcentaje de ingresantes, general

	modalidad	fecha	escuela	post	ing	porc.ing
8	GENERAL	06_09_2015	Educación Primaria	30	18	60 %
5	GENERAL	14_09_2014	Educación Primaria	34	18	52.94 %
2	GENERAL	18_08_2013	Educación Primaria	33	17	51.52 %
11	GENERAL	21_08_2016	Educación Primaria	42	14	33.33 %
12	GENERAL	22_01_2017	Educación Primaria	60	19	31.67 %
4	GENERAL	01_06_2014	Educación Primaria	40	12	30 %
1	GENERAL	07_04_2013	Educación Primaria	46	11	23.91 %
3	GENERAL	12_01_2014	Educación Primaria	34	5	14.71 %
10	GENERAL	15_05_2016	Educación Primaria	37	5	13.51 %
9	GENERAL	31_01_2016	Educación Primaria	45	6	13.33 %
6	GENERAL	18_01_2015	Educación Primaria	28	3	10.71 %
7	GENERAL	07_06_2015	Educación Primaria	34	2	5.88 %

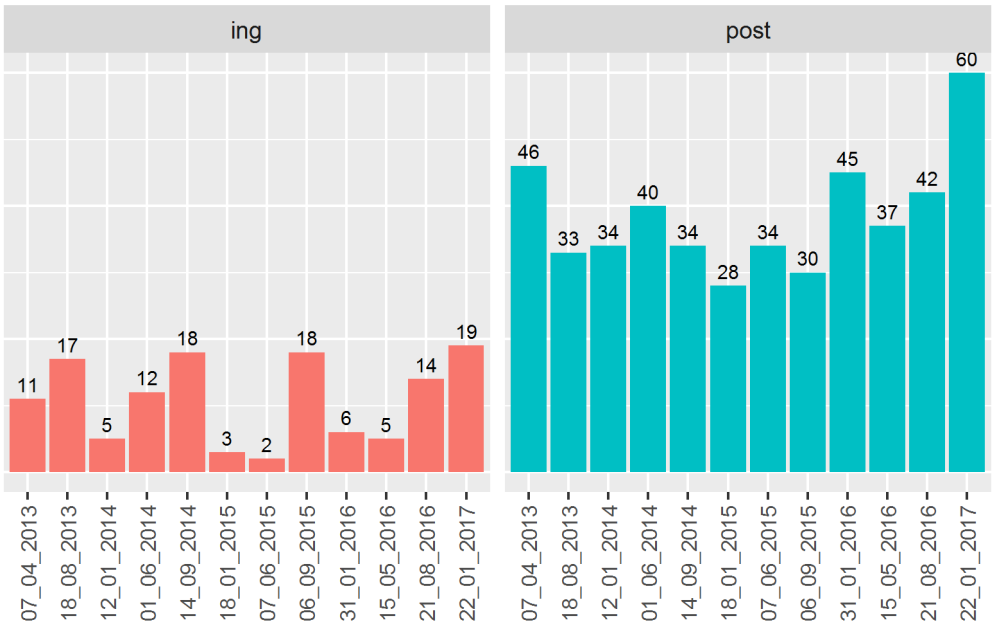


Figura 18. Educación Primaria, número postulantes e ingresantes, examen general

De la figura 18 se puede observar que la mayor cantidad de ingresantes se dió en un 60% en el proceso de admisión de fecha setiembre del 2015 y el mínimo fue del 5.88% de fecha junio del 2015.

Cuadro 8. Educación Primaria, porcentaje de ingresantes, cepreuna

	modalidad	fecha	escuela	post	ing	porc.ing
3	CEPREUNA	08_09_2013	Educación Primaria	4	4	100 %
5	CEPREUNA	16_03_2014	Educación Primaria	8	8	100 %
12	CEPREUNA	10_01_2016	Educación Primaria	5	5	100 %
13	CEPREUNA	20_03_2016	Educación Primaria	11	10	90.91 %
9	CEPREUNA	29_03_2015	Educación Primaria	7	5	71.43 %
2	CEPREUNA	14_07_2013	Educación Primaria	15	10	66.67 %
4	CEPREUNA	22_12_2013	Educación Primaria	9	6	66.67 %
6	CEPREUNA	15_06_2014	Educación Primaria	15	10	66.67 %
1	CEPREUNA	31_03_2013	Educación Primaria	13	8	61.54 %
10	CEPREUNA	09_08_2015	Educación Primaria	16	9	56.25 %
16	CEPREUNA	18_12_2016	Educación Primaria	20	10	50 %
14	CEPREUNA	26_06_2016	Educación Primaria	20	9	45 %
7	CEPREUNA	31_08_2014	Educación Primaria	9	4	44.44 %
15	CEPREUNA	18_09_2016	Educación Primaria	22	9	40.91 %
11	CEPREUNA	08_11_2015	Educación Primaria	4	1	25 %
8	CEPREUNA	14_12_2014	Educación Primaria	5	1	20 %

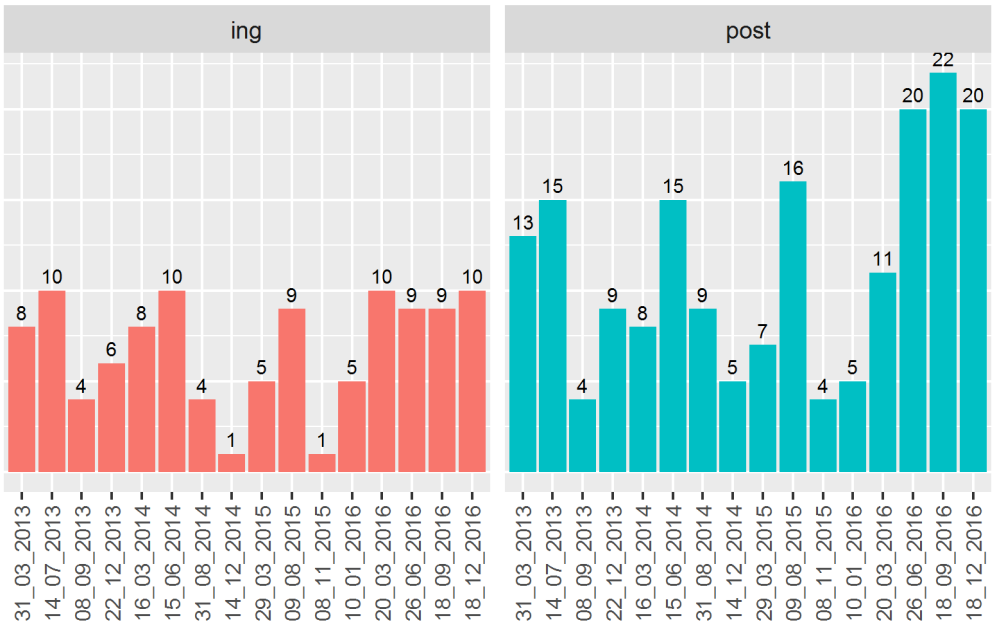


Figura 19. Educación Primaria, número postulantes e ingresantes, examen cepreuna

De la figura 19 se observa que la mayor cantidad de ingresantes se dio en un 100% en el proceso de admisión de fecha setiembre del 2013 y el mínimo fue del 20% de fecha diciembre del 2014.

Cuadro 9. Educación Inicial, porcentaje de ingresantes, general

	modalidad	fecha	escuela	post	ing	porc.ing
2	GENERAL	18_08_2013	Educación Inicial	73	28	38.36 %
5	GENERAL	14_09_2014	Educación Inicial	113	37	32.74 %
4	GENERAL	01_06_2014	Educación Inicial	92	29	31.52 %
3	GENERAL	12_01_2014	Educación Inicial	96	29	30.21 %
8	GENERAL	06_09_2015	Educación Inicial	95	27	28.42 %
11	GENERAL	21_08_2016	Educación Inicial	177	21	11.86 %
10	GENERAL	15_05_2016	Educación Inicial	161	19	11.8 %
1	GENERAL	07_04_2013	Educación Inicial	82	9	10.98 %
12	GENERAL	22_01_2017	Educación Inicial	207	20	9.66 %
9	GENERAL	31_01_2016	Educación Inicial	195	14	7.18 %
7	GENERAL	07_06_2015	Educación Inicial	116	8	6.9 %
6	GENERAL	18_01_2015	Educación Inicial	146	4	2.74 %

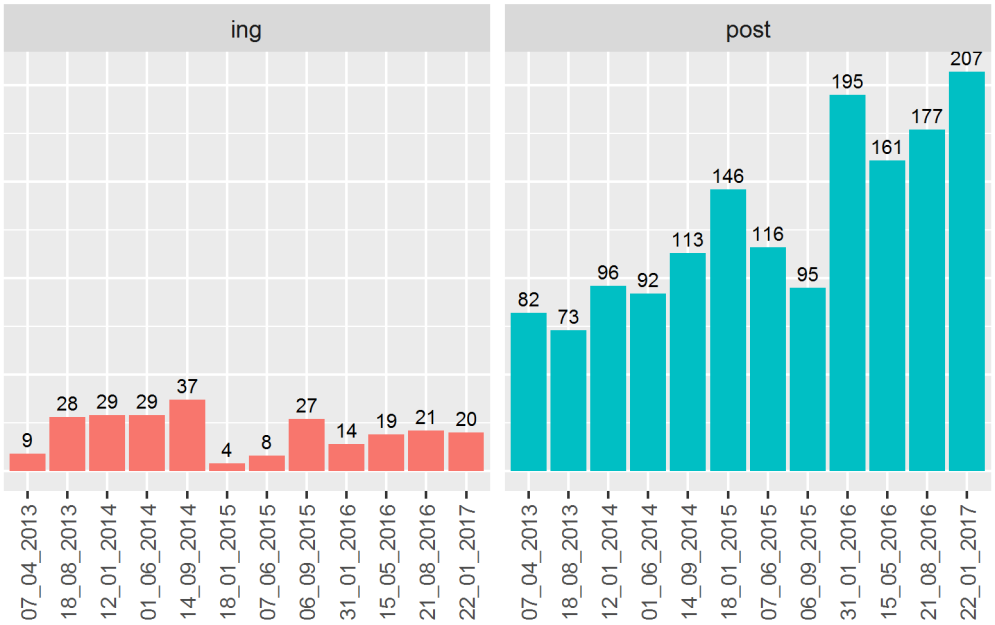


Figura 20. Educación Inicial, número postulantes e ingresantes, examen general

De la figura 20 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 38.36% en el proceso de admisión de fecha agosto del 2013 y el mínimo fue del 2.74% de fecha enero del 2015.

Cuadro 10. Educación Inicial, porcentaje de ingresantes, cepreuna

	modalidad	fecha	escuela	post	ing	porc.ing
3	CEPREUNA	08_09_2013	Educación Inicial	5	5	100 %
5	CEPREUNA	16_03_2014	Educación Inicial	7	6	85.71 %
6	CEPREUNA	15_06_2014	Educación Inicial	6	5	83.33 %
8	CEPREUNA	14_12_2014	Educación Inicial	23	12	52.17 %
4	CEPREUNA	22_12_2013	Educación Inicial	14	6	42.86 %
1	CEPREUNA	31_03_2013	Educación Inicial	20	8	40 %
12	CEPREUNA	10_01_2016	Educación Inicial	31	10	32.26 %
11	CEPREUNA	08_11_2015	Educación Inicial	30	9	30 %
15	CEPREUNA	18_09_2016	Educación Inicial	31	9	29.03 %
2	CEPREUNA	14_07_2013	Educación Inicial	29	8	27.59 %
7	CEPREUNA	31_08_2014	Educación Inicial	36	9	25 %
10	CEPREUNA	09_08_2015	Educación Inicial	46	10	21.74 %
13	CEPREUNA	20_03_2016	Educación Inicial	65	10	15.38 %
16	CEPREUNA	18_12_2016	Educación Inicial	69	10	14.49 %
9	CEPREUNA	29_03_2015	Educación Inicial	43	6	13.95 %
14	CEPREUNA	26_06_2016	Educación Inicial	72	9	12.5 %

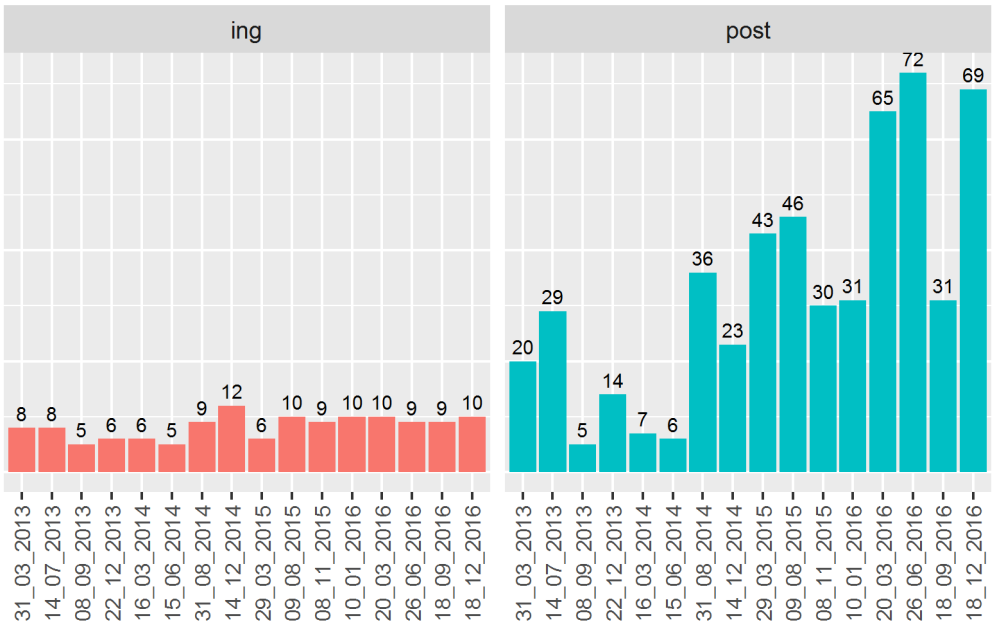


Figura 21. Educación Inicial, número postulantes e ingresantes, examen cepreuna

De la figura 21 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 100% en el proceso de admisión de fecha setiembre del 2013 y el mínimo fue del 12.5% de fecha junio del 2016.

Cuadro 11. Educ.Sec. Ciencias Sociales, porcentaje de ingresantes, general

	modalidad	fecha	escuela	post	ing	porc.ing
2	GENERAL	18_08_2013	Educ. Sec. Ciencias Sociales	26	22	84.62 %
11	GENERAL	21_08_2016	Educ. Sec. Ciencias Sociales	34	22	64.71 %
4	GENERAL	01_06_2014	Educ. Sec. Ciencias Sociales	18	10	55.56 %
8	GENERAL	06_09_2015	Educ. Sec. Ciencias Sociales	32	16	50 %
5	GENERAL	14_09_2014	Educ. Sec. Ciencias Sociales	28	13	46.43 %
1	GENERAL	07_04_2013	Educ. Sec. Ciencias Sociales	27	12	44.44 %
12	GENERAL	22_01_2017	Educ. Sec. Ciencias Sociales	30	12	40 %
3	GENERAL	12_01_2014	Educ. Sec. Ciencias Sociales	15	5	33.33 %
9	GENERAL	31_01_2016	Educ. Sec. Ciencias Sociales	29	6	20.69 %
7	GENERAL	07_06_2015	Educ. Sec. Ciencias Sociales	35	3	8.57 %
6	GENERAL	18_01_2015	Educ. Sec. Ciencias Sociales	28	0	0 %
10	GENERAL	15_05_2016	Educ. Sec. Ciencias Sociales	24	0	0 %

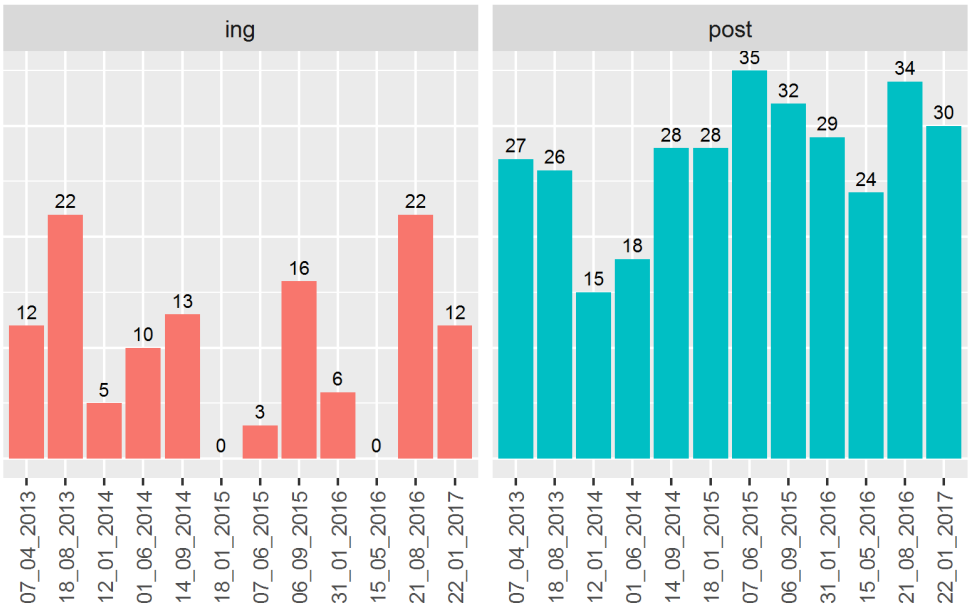


Figura 22. Educ.Sec. Ciencias Sociales, número postulantes e ingresantes, examen general

Del gráfico anterior se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 84.62% en el proceso de admisión de fecha agosto del 2013 y el mínimo fue del 0% de fecha junio del 2015.

Cuadro 12. Educ.Sec. Ciencias Sociales, porcentaje de ingresantes, cepreuna

	modalidad	fecha	escuela	post	ing	porc.ing
3	CEPREUNA	08_09_2013	Educ. Sec. Ciencias Sociales	3	3	100 %
10	CEPREUNA	09_08_2015	Educ. Sec. Ciencias Sociales	7	7	100 %
12	CEPREUNA	10_01_2016	Educ. Sec. Ciencias Sociales	7	7	100 %
14	CEPREUNA	26_06_2016	Educ. Sec. Ciencias Sociales	10	10	100 %
6	CEPREUNA	15_06_2014	Educ. Sec. Ciencias Sociales	8	7	87.5 %
5	CEPREUNA	16_03_2014	Educ. Sec. Ciencias Sociales	6	5	83.33 %
8	CEPREUNA	14_12_2014	Educ. Sec. Ciencias Sociales	5	4	80 %
16	CEPREUNA	18_12_2016	Educ. Sec. Ciencias Sociales	10	8	80 %
2	CEPREUNA	14_07_2013	Educ. Sec. Ciencias Sociales	12	9	75 %
11	CEPREUNA	08_11_2015	Educ. Sec. Ciencias Sociales	7	5	71.43 %
13	CEPREUNA	20_03_2016	Educ. Sec. Ciencias Sociales	9	6	66.67 %
4	CEPREUNA	22_12_2013	Educ. Sec. Ciencias Sociales	8	5	62.5 %
15	CEPREUNA	18_09_2016	Educ. Sec. Ciencias Sociales	16	10	62.5 %
9	CEPREUNA	29_03_2015	Educ. Sec. Ciencias Sociales	7	4	57.14 %
1	CEPREUNA	31_03_2013	Educ. Sec. Ciencias Sociales	8	4	50 %
7	CEPREUNA	31_08_2014	Educ. Sec. Ciencias Sociales	4	1	25 %

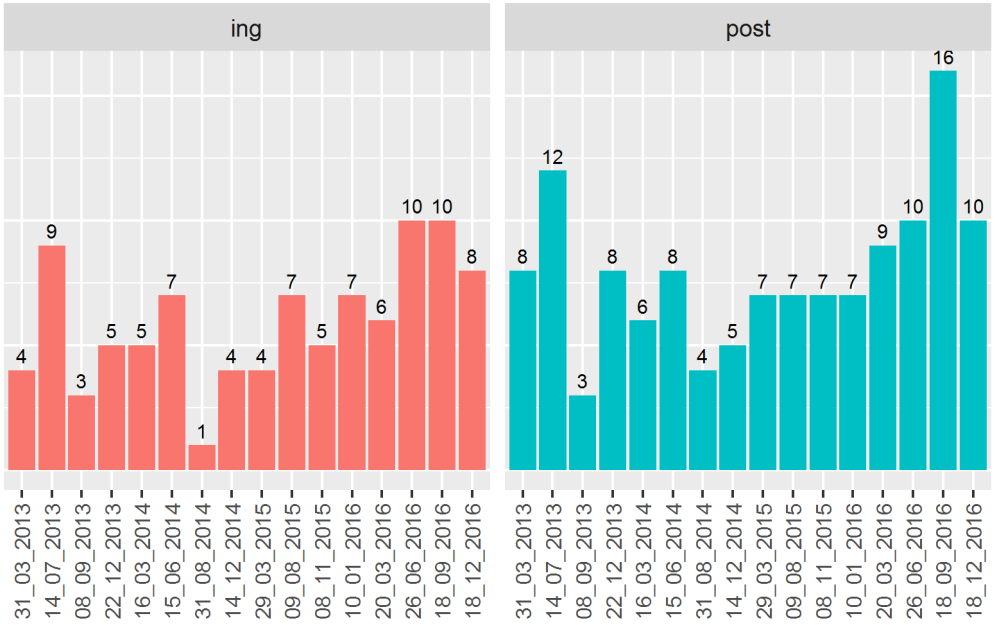


Figura 23. Educ.Sec. Ciencias Sociales, número postulantes e ingresantes, examen cepreuna

De la figura 23 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 100% en el proceso de admisión de fecha setiembre del 2013 y el mínimo fue del 25% de fecha agosto del 2014.

Cuadro 13. Administración, porcentaje de ingresantes, general

	modalidad	fecha	escuela	post	ing	porc.ing
10	GENERAL	15_05_2016	Administración	560	23	4.11 %
2	GENERAL	18_08_2013	Administración	469	19	4.05 %
3	GENERAL	12_01_2014	Administración	507	20	3.94 %
4	GENERAL	01_06_2014	Administración	498	19	3.82 %
6	GENERAL	18_01_2015	Administración	710	25	3.52 %
7	GENERAL	07_06_2015	Administración	738	26	3.52 %
8	GENERAL	06_09_2015	Administración	593	20	3.37 %
11	GENERAL	21_08_2016	Administración	750	25	3.33 %
5	GENERAL	14_09_2014	Administración	587	19	3.24 %
9	GENERAL	31_01_2016	Administración	703	18	2.56 %
12	GENERAL	22_01_2017	Administración	783	19	2.43 %
1	GENERAL	07_04_2013	Administración	829	17	2.05 %

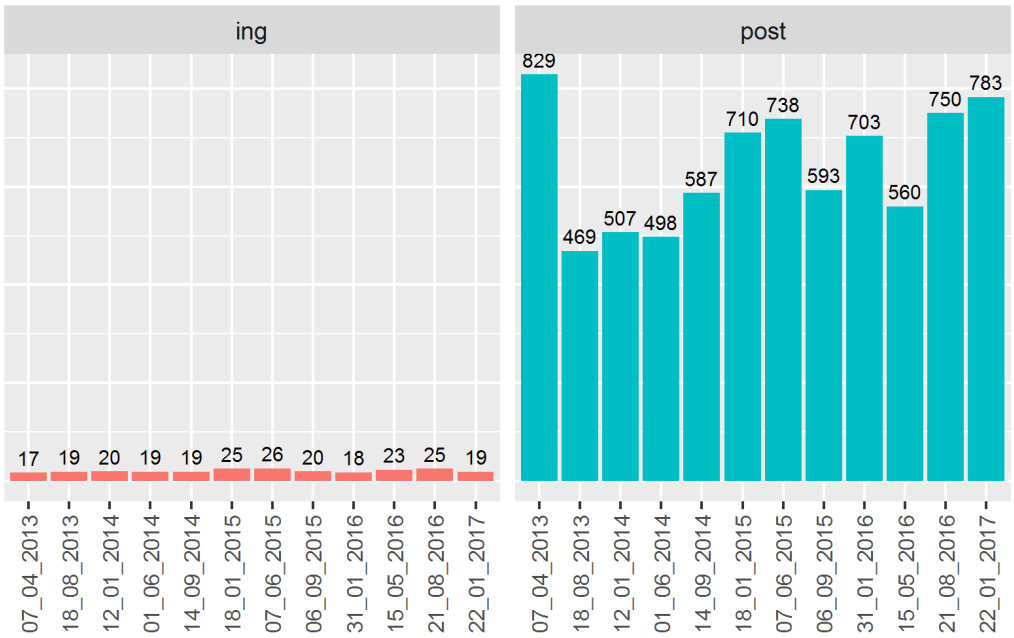


Figura 24. Administración, número postulantes e ingresantes, examen general

De la figura 24 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 4.11% en el proceso de admisión de fecha mayo del 2016 y el mínimo fue del 2.05% de fecha abril del 2013.

Cuadro 14. Administración, porcentaje de ingresantes, cepreuna

	modalidad	fecha	escuela	post	ing	porc.ing
5	CEPREUNA	16_03_2014	Administración	37	10	27.03 %
3	CEPREUNA	08_09_2013	Administración	77	9	11.69 %
12	CEPREUNA	10_01_2016	Administración	127	11	8.66 %
7	CEPREUNA	31_08_2014	Administración	167	11	6.59 %
8	CEPREUNA	14_12_2014	Administración	166	10	6.02 %
11	CEPREUNA	08_11_2015	Administración	174	9	5.17 %
4	CEPREUNA	22_12_2013	Administración	162	8	4.94 %
6	CEPREUNA	15_06_2014	Administración	186	9	4.84 %
2	CEPREUNA	14_07_2013	Administración	216	10	4.63 %
10	CEPREUNA	09_08_2015	Administración	229	10	4.37 %
15	CEPREUNA	18_09_2016	Administración	221	9	4.07 %
16	CEPREUNA	18_12_2016	Administración	247	10	4.05 %
13	CEPREUNA	20_03_2016	Administración	249	10	4.02 %
14	CEPREUNA	26_06_2016	Administración	269	10	3.72 %
9	CEPREUNA	29_03_2015	Administración	313	10	3.19 %
1	CEPREUNA	31_03_2013	Administración	305	8	2.62 %

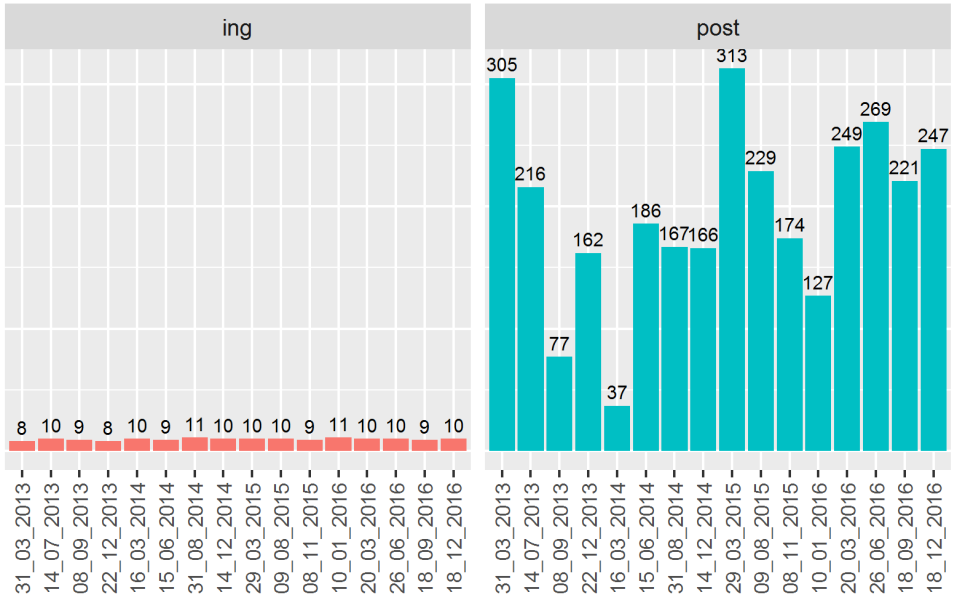


Figura 25. Administración, número postulantes e ingresantes, examen cepreuna

De la figura 25 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 27.03% en el proceso de admisión de fecha marzo del 2014 y el mínimo fue del 2.62% de fecha marzo del 2013.

Cantidad de postulantes e ingresantes de los procesos general y cepreuna de las Escuelas Profesionales del área ingenierías

Cuadro 15. Ingeniería Económica, porcentaje de ingresantes, general

	modalidad	fecha	escuela	post	ing	porc.ing
4	GENERAL	01_06_2014	Ingeniería Económica	757	54	7.13 %
2	GENERAL	18_08_2013	Ingeniería Económica	794	54	6.8 %
3	GENERAL	12_01_2014	Ingeniería Económica	699	44	6.29 %
8	GENERAL	06_09_2015	Ingeniería Económica	619	35	5.65 %
5	GENERAL	14_09_2014	Ingeniería Económica	731	37	5.06 %
10	GENERAL	15_05_2016	Ingeniería Económica	637	32	5.02 %
6	GENERAL	18_01_2015	Ingeniería Económica	682	33	4.84 %
7	GENERAL	07_06_2015	Ingeniería Económica	710	30	4.23 %
11	GENERAL	21_08_2016	Ingeniería Económica	875	35	4 %
9	GENERAL	31_01_2016	Ingeniería Económica	669	25	3.74 %
12	GENERAL	22_01_2017	Ingeniería Económica	764	28	3.66 %
1	GENERAL	07_04_2013	Ingeniería Económica	613	21	3.43 %

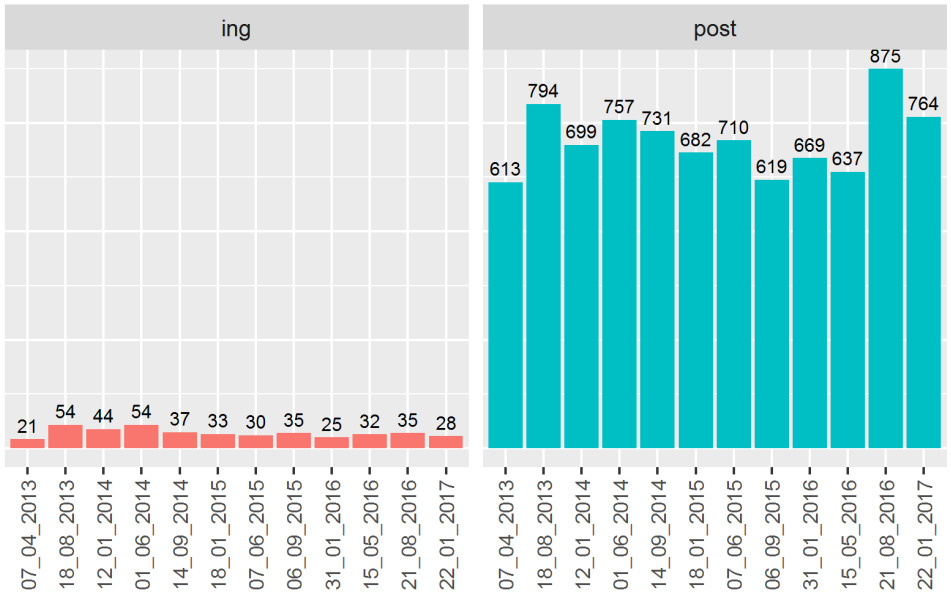


Figura 26. Ingeniería Económica, número postulantes e ingresantes, examen general

De la figura 26 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 7.13% en el proceso de admisión de fecha junio del 2014 y el mínimo fue del 3.43% de fecha abril del 2013.

Cuadro 16. Ingeniería Económica, porcentaje de ingresantes, cepreuna

	modalidad	fecha	escuela	post	ing	porc.ing
3	CEPREUNA	08_09_2013	Ingeniería Económica	71	13	18.31 %
6	CEPREUNA	15_06_2014	Ingeniería Económica	141	15	10.64 %
12	CEPREUNA	10_01_2016	Ingeniería Económica	97	10	10.31 %
7	CEPREUNA	31_08_2014	Ingeniería Económica	147	15	10.2 %
4	CEPREUNA	22_12_2013	Ingeniería Económica	158	14	8.86 %
8	CEPREUNA	14_12_2014	Ingeniería Económica	177	15	8.47 %
2	CEPREUNA	14_07_2013	Ingeniería Económica	213	15	7.04 %
11	CEPREUNA	08_11_2015	Ingeniería Económica	128	9	7.03 %
15	CEPREUNA	18_09_2016	Ingeniería Económica	172	10	5.81 %
5	CEPREUNA	16_03_2014	Ingeniería Económica	248	14	5.65 %
9	CEPREUNA	29_03_2015	Ingeniería Económica	293	15	5.12 %
1	CEPREUNA	31_03_2013	Ingeniería Económica	235	12	5.11 %
16	CEPREUNA	18_12_2016	Ingeniería Económica	206	10	4.85 %
10	CEPREUNA	09_08_2015	Ingeniería Económica	233	10	4.29 %
13	CEPREUNA	20_03_2016	Ingeniería Económica	234	10	4.27 %
14	CEPREUNA	26_06_2016	Ingeniería Económica	247	10	4.05 %

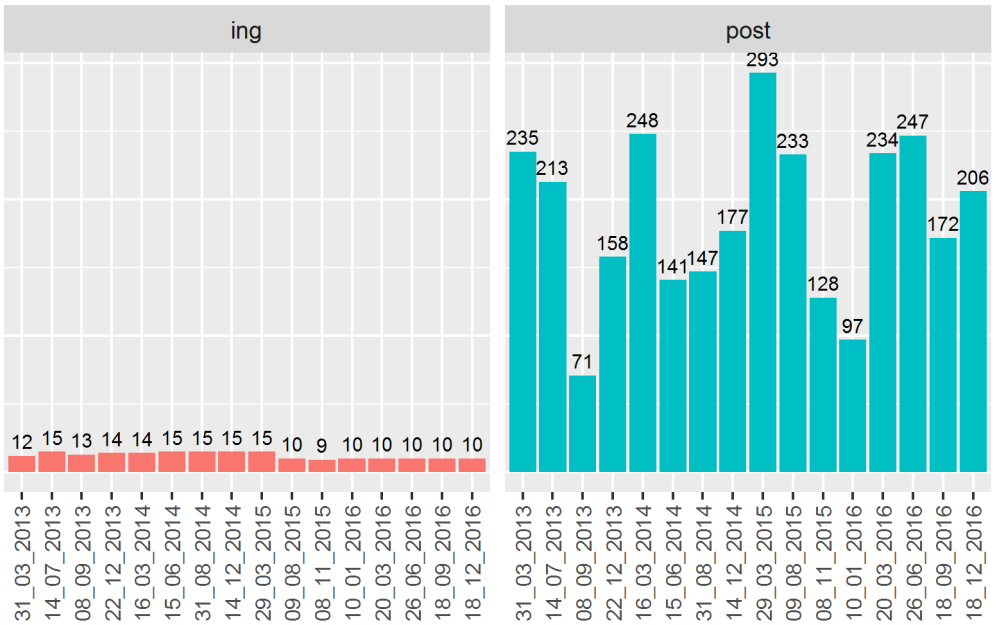


Figura 27. Ingeniería Económica, número postulantes e ingresantes, examen cepreuna

De la figura 27 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 18.31% en el proceso de admisión de fecha setiembre del 2013 y el mínimo fue del 4.05% de fecha junio del 2016.

Cuadro 17. Ingeniería de Minas, porcentaje de ingresantes, general

	modalidad	fecha	escuela	post	ing	porc.ing
10	GENERAL	15_05_2016	Ingeniería de Minas	228	24	10.53 %
8	GENERAL	06_09_2015	Ingeniería de Minas	260	21	8.08 %
7	GENERAL	07_06_2015	Ingeniería de Minas	341	23	6.74 %
4	GENERAL	01_06_2014	Ingeniería de Minas	407	26	6.39 %
11	GENERAL	21_08_2016	Ingeniería de Minas	314	20	6.37 %
6	GENERAL	18_01_2015	Ingeniería de Minas	362	22	6.08 %
12	GENERAL	22_01_2017	Ingeniería de Minas	292	17	5.82 %
3	GENERAL	12_01_2014	Ingeniería de Minas	424	23	5.42 %
5	GENERAL	14_09_2014	Ingeniería de Minas	379	19	5.01 %
2	GENERAL	18_08_2013	Ingeniería de Minas	479	23	4.8 %
9	GENERAL	31_01_2016	Ingeniería de Minas	265	11	4.15 %
1	GENERAL	07_04_2013	Ingeniería de Minas	468	14	2.99 %

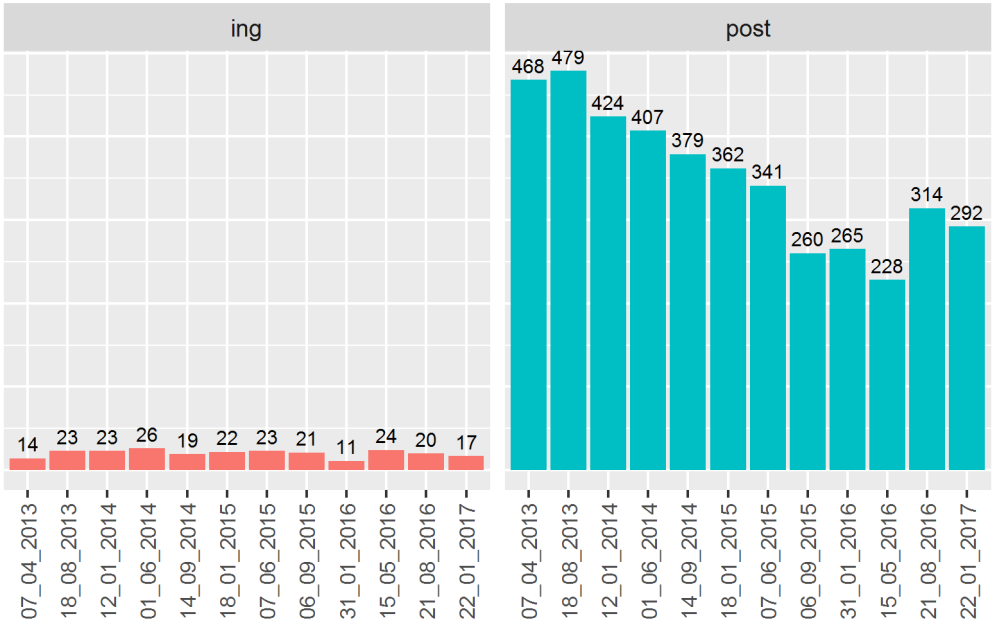


Figura 28. Ingeniería Económica, número postulantes e ingresantes, examen general

De la figura 28 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 10.53% en el proceso de admisión de fecha mayo del 2016 y el mínimo fue del 2.99% de fecha abril del 2013.

Cuadro 18. Ingeniería de Minas, porcentaje de ingresantes, cepreuna

	modalidad	fecha	escuela	post	ing	porc.ing
12	CEPREUNA	10_01_2016	Ingeniería de Minas	58	15	25.86 %
3	CEPREUNA	08_09_2013	Ingeniería de Minas	70	15	21.43 %
11	CEPREUNA	08_11_2015	Ingeniería de Minas	82	14	17.07 %
6	CEPREUNA	15_06_2014	Ingeniería de Minas	88	15	17.05 %
7	CEPREUNA	31_08_2014	Ingeniería de Minas	88	14	15.91 %
13	CEPREUNA	20_03_2016	Ingeniería de Minas	103	15	14.56 %
10	CEPREUNA	09_08_2015	Ingeniería de Minas	109	15	13.76 %
14	CEPREUNA	26_06_2016	Ingeniería de Minas	125	14	11.2 %
9	CEPREUNA	29_03_2015	Ingeniería de Minas	115	12	10.43 %
8	CEPREUNA	14_12_2014	Ingeniería de Minas	87	9	10.34 %
4	CEPREUNA	22_12_2013	Ingeniería de Minas	111	11	9.91 %
15	CEPREUNA	18_09_2016	Ingeniería de Minas	95	8	8.42 %
2	CEPREUNA	14_07_2013	Ingeniería de Minas	149	12	8.05 %
16	CEPREUNA	18_12_2016	Ingeniería de Minas	115	9	7.83 %
5	CEPREUNA	16_03_2014	Ingeniería de Minas	158	11	6.96 %
1	CEPREUNA	31_03_2013	Ingeniería de Minas	175	10	5.71 %

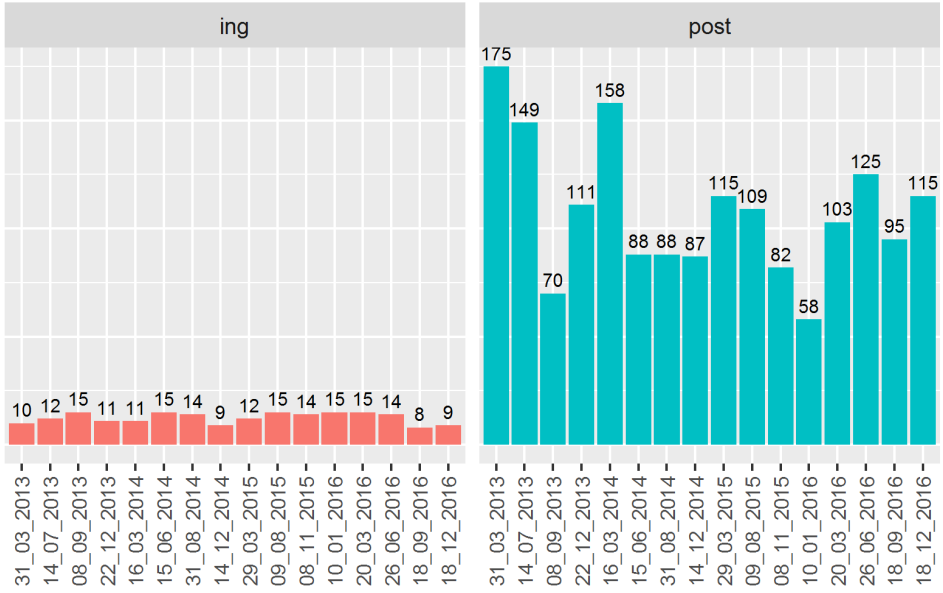


Figura 29. Ingeniería de Minas, número postulantes e ingresantes, examen cepreuna

De la figura 29 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 25.86% en el proceso de admisión de fecha enero del 2016 y el mínimo fue del 5.71% de fecha marzo del 2013.

Cuadro 19. Ingeniería Química, porcentaje de ingresantes, general

	modalidad	fecha	escuela	post	ing	porc.ing
3	GENERAL	12_01_2014	Ingeniería Química	21	18	85.71 %
2	GENERAL	18_08_2013	Ingeniería Química	31	24	77.42 %
5	GENERAL	14_09_2014	Ingeniería Química	33	25	75.76 %
8	GENERAL	06_09_2015	Ingeniería Química	20	14	70 %
4	GENERAL	01_06_2014	Ingeniería Química	19	12	63.16 %
11	GENERAL	21_08_2016	Ingeniería Química	44	23	52.27 %
12	GENERAL	22_01_2017	Ingeniería Química	34	15	44.12 %
1	GENERAL	07_04_2013	Ingeniería Química	9	3	33.33 %
10	GENERAL	15_05_2016	Ingeniería Química	29	6	20.69 %
6	GENERAL	18_01_2015	Ingeniería Química	30	5	16.67 %
9	GENERAL	31_01_2016	Ingeniería Química	32	3	9.38 %
7	GENERAL	07_06_2015	Ingeniería Química	32	2	6.25 %

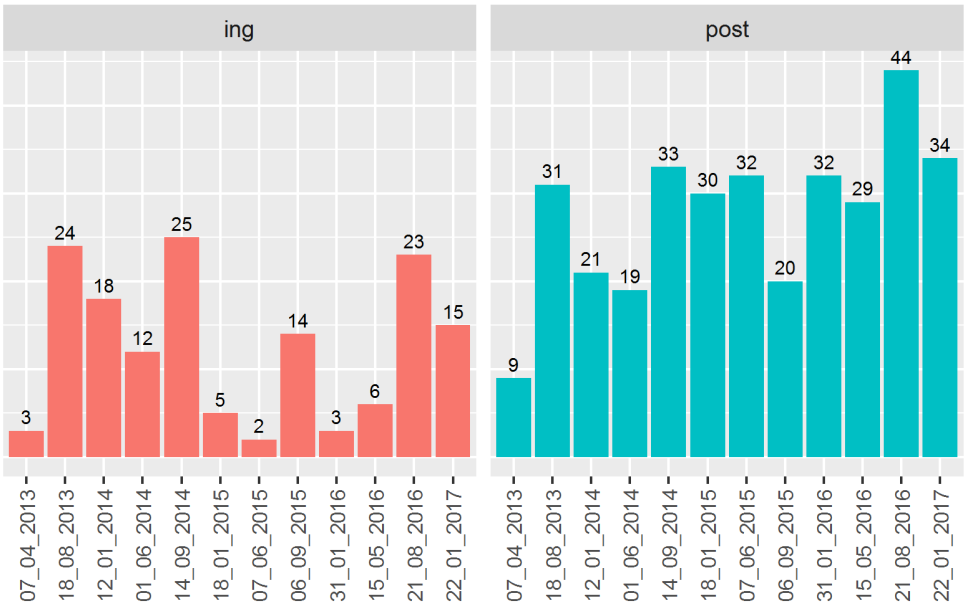


Figura 30. Ingeniería Química, número postulantes e ingresantes, examen general

De la figura 30 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 85.71% en el proceso de admisión de fecha enero del 2014 y el mínimo fue del 6.25% de fecha junio del 2015.

Cuadro 20. Ingeniería Química, porcentaje de ingresantes, cepreuna

	modalidad	fecha	escuela	post	ing	porc.ing
1	CEPREUNA	31_03_2013	Ingeniería Química	1	1	100 %
2	CEPREUNA	14_07_2013	Ingeniería Química	4	4	100 %
3	CEPREUNA	08_09_2013	Ingeniería Química	4	4	100 %
4	CEPREUNA	22_12_2013	Ingeniería Química	4	4	100 %
5	CEPREUNA	16_03_2014	Ingeniería Química	8	8	100 %
16	CEPREUNA	18_12_2016	Ingeniería Química	8	8	100 %
14	CEPREUNA	26_06_2016	Ingeniería Química	11	10	90.91 %
12	CEPREUNA	10_01_2016	Ingeniería Química	4	3	75 %
11	CEPREUNA	08_11_2015	Ingeniería Química	3	2	66.67 %
15	CEPREUNA	18_09_2016	Ingeniería Química	6	4	66.67 %
10	CEPREUNA	09_08_2015	Ingeniería Química	5	3	60 %
6	CEPREUNA	15_06_2014	Ingeniería Química	2	1	50 %
9	CEPREUNA	29_03_2015	Ingeniería Química	9	4	44.44 %
13	CEPREUNA	20_03_2016	Ingeniería Química	7	3	42.86 %
7	CEPREUNA	31_08_2014	Ingeniería Química	6	1	16.67 %
8	CEPREUNA	14_12_2014	Ingeniería Química	4	0	0 %

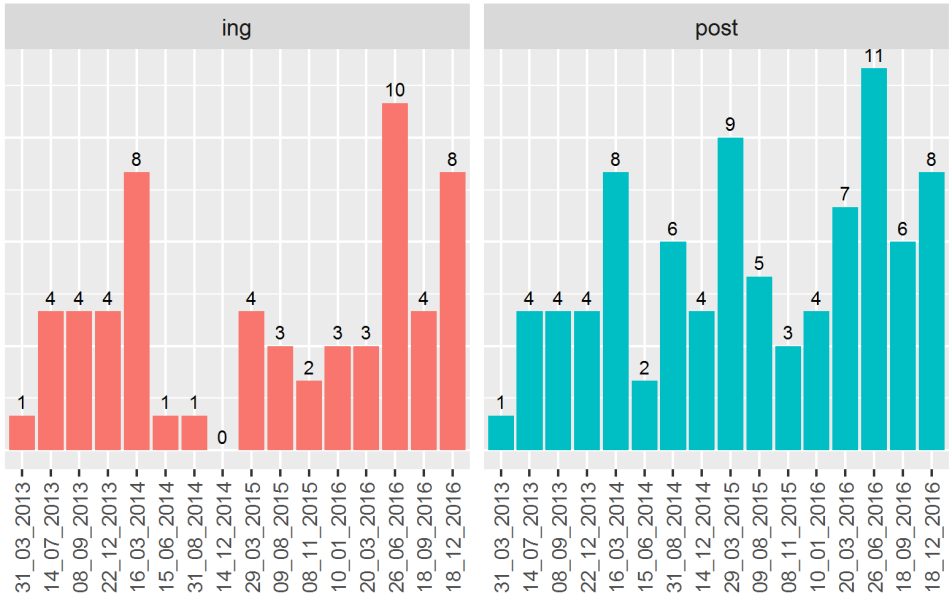


Figura 31. Ingeniería Química, número postulantes e ingresantes, examen cepreuna

De la figura 31 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 100% en el proceso de admisión de fecha marzo, julio, setiembre y diciembre del 2013 y el mínimo fue del 0% de fecha diciembre del 2014.

Cuadro 21. Ingeniería Estadística e Informática, porcentaje de ingresantes, general

	modalidad	fecha	escuela	post	ing	porc.ing
5	GENERAL	14_09_2014	Ingeniería Estadística e Informática	19	16	84.21 %
4	GENERAL	01_06_2014	Ingeniería Estadística e Informática	25	19	76 %
2	GENERAL	18_08_2013	Ingeniería Estadística e Informática	32	22	68.75 %
3	GENERAL	12_01_2014	Ingeniería Estadística e Informática	14	7	50 %
11	GENERAL	21_08_2016	Ingeniería Estadística e Informática	23	11	47.83 %
12	GENERAL	22_01_2017	Ingeniería Estadística e Informática	34	15	44.12 %
8	GENERAL	06_09_2015	Ingeniería Estadística e Informática	16	6	37.5 %
1	GENERAL	07_04_2013	Ingeniería Estadística e Informática	28	7	25 %
6	GENERAL	18_01_2015	Ingeniería Estadística e Informática	30	4	13.33 %
7	GENERAL	07_06_2015	Ingeniería Estadística e Informática	33	3	9.09 %
9	GENERAL	31_01_2016	Ingeniería Estadística e Informática	23	2	8.7 %
10	GENERAL	15_05_2016	Ingeniería Estadística e Informática	27	1	3.7 %

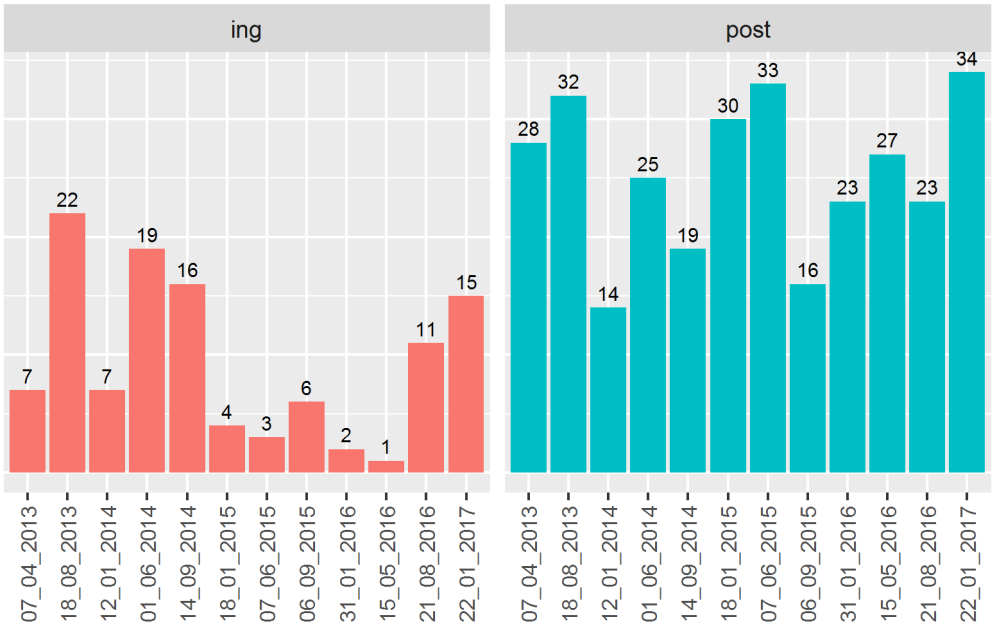


Figura 32. Ingeniería Química, número postulantes e ingresantes, examen cepreuna

De la figura 32 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 84.21% en el proceso de admisión de fecha setiembre del 2014 y el mínimo fue del 3.7% de fecha mayo del 2016.

Cuadro 22. Ingeniería Estadística e Informática, porcentaje de ingresantes, cepreuna

	modalidad	fecha	escuela	post	ing	porc.ing
3	CEPREUNA	22_12_2013	Ingeniería Estadística e Informática	3	3	100 %
5	CEPREUNA	15_06_2014	Ingeniería Estadística e Informática	5	5	100 %
14	CEPREUNA	18_09_2016	Ingeniería Estadística e Informática	6	6	100 %
15	CEPREUNA	18_12_2016	Ingeniería Estadística e Informática	8	8	100 %
2	CEPREUNA	14_07_2013	Ingeniería Estadística e Informática	9	8	88.89 %
10	CEPREUNA	08_11_2015	Ingeniería Estadística e Informática	9	8	88.89 %
1	CEPREUNA	31_03_2013	Ingeniería Estadística e Informática	8	7	87.5 %
13	CEPREUNA	26_06_2016	Ingeniería Estadística e Informática	10	8	80 %
9	CEPREUNA	09_08_2015	Ingeniería Estadística e Informática	4	3	75 %
4	CEPREUNA	16_03_2014	Ingeniería Estadística e Informática	9	6	66.67 %
11	CEPREUNA	10_01_2016	Ingeniería Estadística e Informática	3	2	66.67 %
8	CEPREUNA	29_03_2015	Ingeniería Estadística e Informática	5	2	40 %
12	CEPREUNA	20_03_2016	Ingeniería Estadística e Informática	10	3	30 %
6	CEPREUNA	31_08_2014	Ingeniería Estadística e Informática	5	1	20 %
7	CEPREUNA	14_12_2014	Ingeniería Estadística e Informática	2	0	0 %

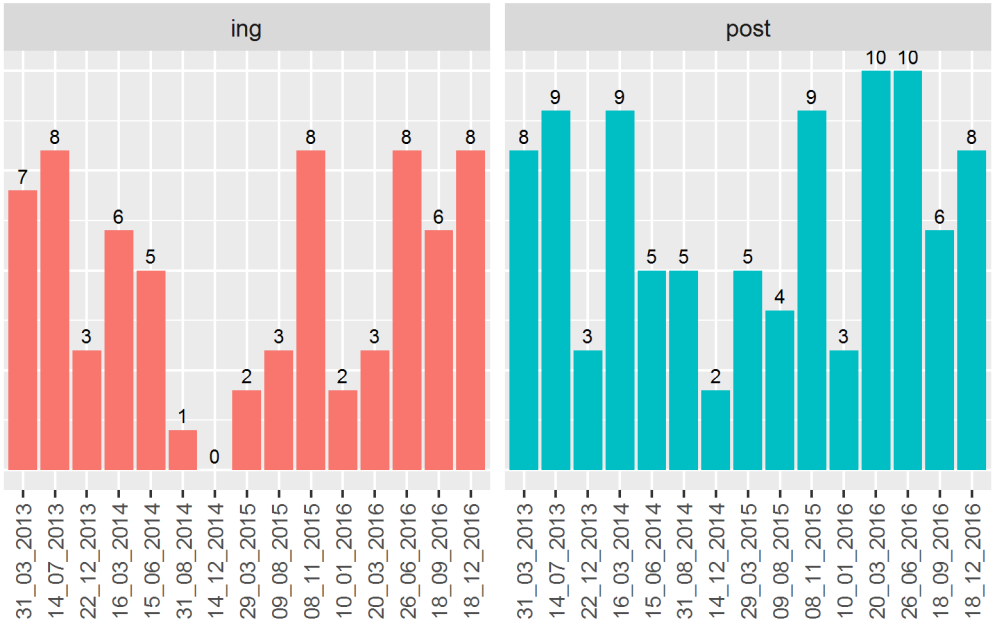


Figura 33. Ingeniería Estadística e Informática, número postulantes e ingresantes, examen cepreuna

De la figura 33 se observa que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 100% en el proceso de admisión de fecha diciembre 2013, junio 2014, setiembre y diciembre 2016 y el mínimo fue del 0% de fecha diciembre 2014.

Cuadro 23. Ingeniería de Sistemas, porcentaje de ingresantes, general

	modalidad	fecha	escuela	post	ing	porc.ing
4	GENERAL	01_06_2014	Ingeniería de Sistemas	188	29	15.43 %
2	GENERAL	18_08_2013	Ingeniería de Sistemas	192	28	14.58 %
3	GENERAL	12_01_2014	Ingeniería de Sistemas	216	29	13.43 %
10	GENERAL	15_05_2016	Ingeniería de Sistemas	188	22	11.7 %
6	GENERAL	18_01_2015	Ingeniería de Sistemas	260	29	11.15 %
5	GENERAL	14_09_2014	Ingeniería de Sistemas	211	22	10.43 %
8	GENERAL	06_09_2015	Ingeniería de Sistemas	232	22	9.48 %
11	GENERAL	21_08_2016	Ingeniería de Sistemas	308	22	7.14 %
7	GENERAL	07_06_2015	Ingeniería de Sistemas	249	17	6.83 %
1	GENERAL	07_04_2013	Ingeniería de Sistemas	237	16	6.75 %
9	GENERAL	31_01_2016	Ingeniería de Sistemas	276	13	4.71 %
12	GENERAL	22_01_2017	Ingeniería de Sistemas	274	12	4.38 %

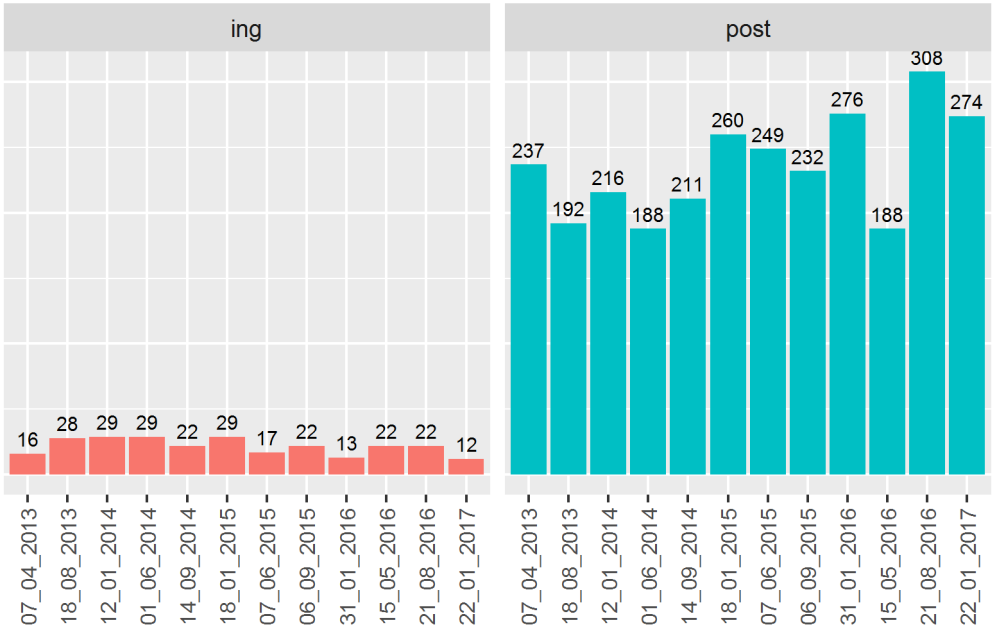


Figura 34. Ingeniería de Sistemas, número postulantes e ingresantes, examen general

De la figura 34 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 15.43% en el proceso de admisión de fecha junio del 2014 y el mínimo fue del 4.38% de fecha enero del 2017.

Cuadro 24. Ingeniería de Sistemas, porcentaje de ingresantes, cepreuna

	modalidad	fecha	escuela	post	ing	porc.ing
3	CEPREUNA	08_09_2013	Ingeniería de Sistemas	21	11	52.38 %
4	CEPREUNA	22_12_2013	Ingeniería de Sistemas	59	20	33.9 %
6	CEPREUNA	15_06_2014	Ingeniería de Sistemas	70	23	32.86 %
12	CEPREUNA	10_01_2016	Ingeniería de Sistemas	47	15	31.91 %
2	CEPREUNA	14_07_2013	Ingeniería de Sistemas	69	22	31.88 %
11	CEPREUNA	08_11_2015	Ingeniería de Sistemas	50	14	28 %
10	CEPREUNA	09_08_2015	Ingeniería de Sistemas	70	15	21.43 %
13	CEPREUNA	20_03_2016	Ingeniería de Sistemas	82	16	19.51 %
16	CEPREUNA	18_12_2016	Ingeniería de Sistemas	79	15	18.99 %
15	CEPREUNA	18_09_2016	Ingeniería de Sistemas	84	15	17.86 %
9	CEPREUNA	29_03_2015	Ingeniería de Sistemas	103	18	17.48 %
5	CEPREUNA	16_03_2014	Ingeniería de Sistemas	101	17	16.83 %
14	CEPREUNA	26_06_2016	Ingeniería de Sistemas	84	14	16.67 %
8	CEPREUNA	14_12_2014	Ingeniería de Sistemas	38	5	13.16 %
1	CEPREUNA	31_03_2013	Ingeniería de Sistemas	95	11	11.58 %
7	CEPREUNA	31_08_2014	Ingeniería de Sistemas	61	2	3.28 %

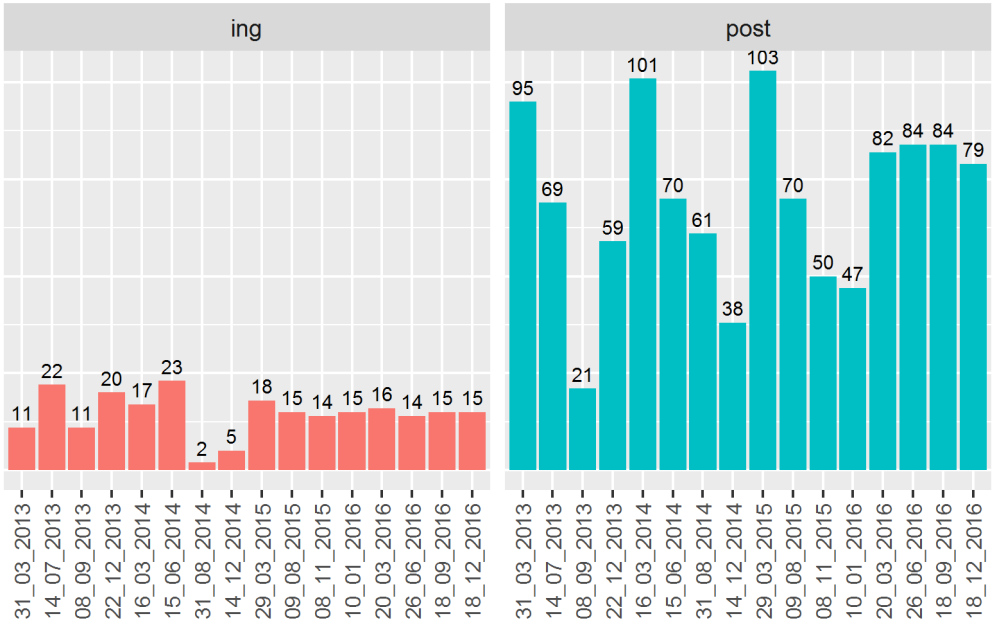


Figura 35. Ingeniería de Sistemas, número postulantes e ingresantes, examen cepreuna

De la figura 35 se puede observar que la mayor cantidad de ingresantes en función a la cantidad de postulantes se dio en un 52.38% en el proceso de admisión de fecha setiembre del 2013 y el mínimo fue del 3.28% de fecha agosto del 2014.

Cantidad de postulantes e ingresantes del departamento de Puno de los colegios públicos de educación secundaria ordenados en forma descendente en función a la cantidad de postulantes

Líneas de código que genera la relación y el gráfico con la cantidad de postulantes e ingresantes de las escuelas públicas y privadas de educación secundaria.

```
v_dep <- "PUNO"
v_cole <- "PARTICULAR"
# relacion postulantes del departamento de puno - estatal (116231)
post<-todo %>% filter(depart_colegio==v_dep &
                        tipo_colegio==v_cole)
# cant. post. por colegio (estatal) - ordenados de mayor a menor
resppost<-post %>% group_by(depart_colegio, prov_colegio,
                            dist_colegio, nombre_colegio) %>%
  summarise(total=n()) %>% arrange(desc(total))
#View(resppost)
# relacion ingresantes del departamento de puno - estatal (116231)
ing<-todo %>% filter(depart_colegio==v_dep &
                       tipo_colegio==v_cole &
                       ingreso=="SI")
# cant. ing. por colegio (estatal) - ordenados de mayor a menor
resping<-ing %>% group_by(depart_colegio, prov_colegio,
                          dist_colegio, nombre_colegio) %>%
  summarise(total=n()) %>% arrange(desc(total))
#View(resping)
# juntando cant. de postulantes en ingresantes, x nomb Colegio del d
post.ing<-full_join(resppost, resping,
                    by=c("depart_colegio", "prov_colegio",
                         "dist_colegio", "nombre_colegio"))
# reemplazando los valores NULL con CERO
post.ing[is.na(post.ing)]<-0
# renombrando columnas
post.ing<-rename(post.ing, depart = depart_colegio, prov = prov_colegio,
                 dist = dist_colegio, post = total.x, ing = total.y)

# agrego columna de porcentaje
post.ing$porc.ing<-round(post.ing$ing*100/post.ing$post,2)
post.ing<-data.frame(post.ing %>% arrange(desc(post)))
post.ing<-post.ing %>% mutate(porc.ing = paste(porc.ing,"%"))
View(post.ing)
df.long<-melt((post.ing01))
ggplot(df.long, aes(x=reorder(paste(prov," - ",nombre_colegio, sep = ""),
                              value), y=value, fill=variable)) +
  geom_bar(stat="identity",position="dodge") +
  scale_y_continuous(limits=c(0,1700)) + # TENER CUIDADO, ME SALE ADVERTENCI
  theme(axis.title.x=element_blank(),
        axis.title.y = element_blank(),
        axis.ticks.x=element_blank(),
        plot.title=element_text(hjust=0.5, size=8,
                                face="bold"),
        text = element_text(size = 8),
        axis.text.x=element_blank(),
        plot.margin=unit(c(-0.5, 1, 0.5, 0.5), units="line"),
        legend.position="top",
        plot.background=element_rect(fill="white"),
        legend.margin=unit(0, "lines")) +
  guides(fill=guide_legend(title.position="top"))+
  coord_flip() + geom_text(aes(label=value), hjust=-0.2,
                           size=1.8, position = position_dodge(width = 1))
  scale_fill_discrete(name="",
                      labels=c("Postulante","Ingresante"))
```

Las 30 escuelas de educación secundaria con una cantidad de postulantes mayor a 1000.

Cuadro 25. Porcentaje de ingresantes colegios públicos

	depart	prov	dist	nombre_colegio	post	ing	porc.ing
1	PUNO	PUNO	PUNO	GRAN UNIDAD ESCOLAR SAN CARLOS	6923	838	12.1 %
2	PUNO	SAN ROMAN	JULIACA	JOSÉ ANTONIO ENCINAS	5834	568	9.74 %
3	PUNO	PUNO	PUNO	GLORIOSO SAN CARLOS	4666	545	11.68 %
4	PUNO	PUNO	PUNO	SANTA ROSA	4463	677	15.17 %
5	PUNO	EL COLLAO	ILAVE	NUESTRA SEÑORA DEL CARMEN	4083	409	10.02 %
6	PUNO	PUNO	PUNO	SECUNDARIA MARIA AUXILIADORA	3687	400	10.85 %
7	PUNO	PUNO	PUNO	45 EMILIO ROMERO PADILLA	3423	350	10.22 %
8	PUNO	MELGAR	AYAVIRI	NUESTRA SEÑORA DE ALTA GRACIA	2774	301	10.85 %
9	PUNO	SAN ROMAN	JULIACA	S/N G.U.E. LAS MERCEDES	2303	179	7.77 %
10	PUNO	MELGAR	AYAVIRI	SECUNDARIA MARIANO MELGAR	2116	282	13.33 %
11	PUNO	SAN ROMAN	JULIACA	POLITECNICO REGIONAL LOS ANDES	2116	173	8.18 %
12	PUNO	SAN ROMAN	JULIACA	COMERCIO 32 MARIANO H. CORNEJO	2079	169	8.13 %
13	PUNO	YUNGUYO	YUNGUYO	I.E.S. JOSE GALVEZ	2007	199	9.92 %
14	PUNO	AZANGARO	AZANGARO	I.E.S. PEDRO VILCAPAZA ALARCON	1938	165	8.51 %
15	PUNO	AZANGARO	AZANGARO	28 PERU BIRF	1837	181	9.85 %
16	PUNO	SAN ROMAN	JULIACA	PERU BIRF	1670	164	9.82 %
17	PUNO	EL COLLAO	ILAVE	JOSE CARLOS MARIATEGUI	1641	154	9.38 %
18	PUNO	PUNO	PUNO	INDEPENDENCIA NACIONAL	1575	182	11.56 %
19	PUNO	PUNO	PUNO	32 INDUSTRIAL	1520	126	8.29 %
20	PUNO	CHUCUITO	JULI	SECUNDARIA TELESFORO CATACORA JULI	1478	165	11.16 %
21	PUNO	SAN ROMAN	JULIACA	PEDRO VILCAPAZA	1314	84	6.39 %
22	PUNO	CHUCUITO	JULI	MARIA ASUNCION GALINDO	1306	148	11.33 %
23	PUNO	CARABAYA	MACUSANI	JOSE MACEDO MENDOZA	1285	127	9.88 %
24	PUNO	PUNO	PUNO	IES. CARLOS RUBINA BURGOS PUNO	1284	110	8.57 %
25	PUNO	HUANCANE	HUANCANE	VARONES	1282	111	8.66 %
26	PUNO	SAN ROMAN	CABANILLAS	CABANILLAS	1173	112	9.55 %
27	PUNO	HUANCANE	TARACO	TARACO	1160	85	7.33 %
28	PUNO	CHUCUITO	DESAGUADERO	IES. TAWANTINSUYO	1092	90	8.24 %
29	PUNO	EL COLLAO	ILAVE	POLITECNICO REGIONAL DON BOSCO	1034	79	7.64 %
30	PUNO	PUNO	PUNO	JOSE CARLOS MARIATEGUI APLICACIÓN UNA-P	1030	106	10.29 %

Del cuadro 25 se aprecia que la mayor cantidad de ingresantes lo tiene la escuela SANTA ROSA de la ciudad de Puno en un 15.17%, seguido de la escuela MARIANO MELGAR DE AYAVIRI en un 13.33%, seguido de la escuela GUE SAN CARLOS con un 12.1%, seguido de GLORIOSO SAN CARLOS con un 11.68%, seguido de la escuela INDEPENDENCIA de la ciudad de Puno 11.56%, etc.

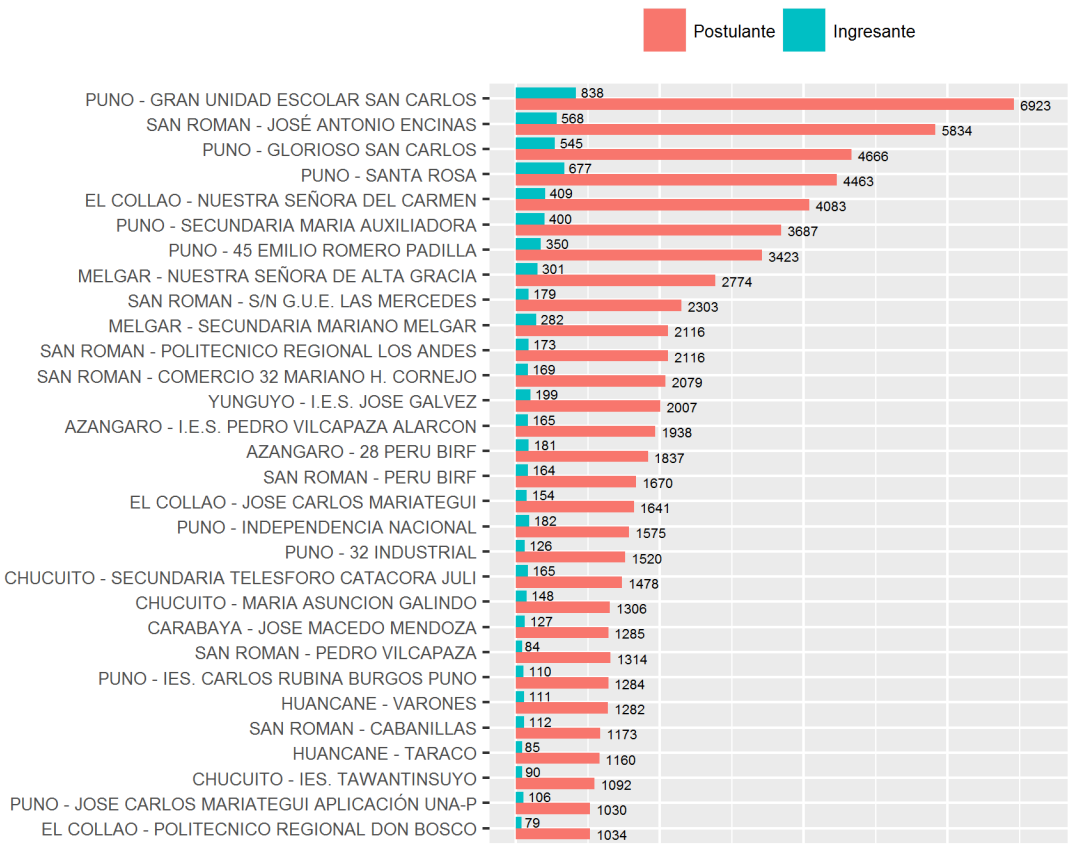


Figura 36. Cantidad de postulantes ingresantes de colegios públicos

La figura 36 muestra el ordenamiento descendente por la cantidad de postulantes de las escuelas estatales.

Cantidad de postulantes e ingresantes del departamento de Puno de los colegios privados de educación secundaria ordenados en forma descendente en función a la cantidad de postulantes.

Las 20 escuelas de educación secundaria privados con una cantidad de postulantes mayor a 260.

Cuadro 26. Porcentaje de ingresantes colegios privados

	depart	prov	dist	nombre_colegio	post	ing	porc.ing
1	PUNO	SAN ROMAN	JULIACA	CLAUDIO GALENO	1553	192	12.36 %
2	PUNO	PUNO	PUNO	CLAUDIO GALENO	1406	156	11.1 %
3	PUNO	PUNO	PUNO	NUESTRA SEÑORA DE LA MERCED	1310	170	12.98 %
4	PUNO	PUNO	PUNO	SAN IGNACIO DE LOYOLA	1037	139	13.4 %
5	PUNO	SAN ROMAN	JULIACA	JAMES BALDWIN	870	85	9.77 %

6	PUNO	PUNO	PUNO	I.E.N.E. DIVINO MAESTRO	849	107	12.6 %
7	PUNO	PUNO	PUNO	ADVENTISTA PUNO	779	85	10.91 %
8	PUNO	SAN ROMAN	JULIACA	TUPAC AMARU	705	88	12.48 %
9	PUNO	SAN ROMAN	JULIACA	SANTA CATALINA	601	66	10.98 %
10	PUNO	PUNO	PUNO	PARROQUIAL VILLA FATIMA	581	76	13.08 %
11	PUNO	SAN ROMAN	JULIACA	FRANCISCANO SAN ROMAN	557	55	9.87 %
12	PUNO	PUNO	PUNO	IEP CRAMER	499	92	18.44 %
13	PUNO	PUNO	PUNO	PARROQUIAL LA INMACULADA - PUNO	494	65	13.16 %
14	PUNO	SAN ROMAN	JULIACA	GREGOR MENDEL	468	63	13.46 %
15	PUNO	PUNO	PUNO	CHAMPAGNAT DEL NIÑO DIVINO JESUS	456	39	8.55 %
16	PUNO	SAN ROMAN	JULIACA	SIGMA	440	41	9.32 %
17	PUNO	SAN ROMAN	JULIACA	LUZ ANDINA REINA DE LAS AMERICAS	414	33	7.97 %
18	PUNO	EL COLLAO	ILAVE	ANTONIO RAIMONDI	370	49	13.24 %
19	PUNO	PUNO	PUNO	ALEXANDER FLEMING	311	44	14.15 %
20	PUNO	SAN ROMAN	JULIACA	ADVENTISTA TITICACA	263	29	11.03 %

Del cuadro 26 la mayor cantidad de ingresantes lo tiene la escuela IEP CRAMER de la ciudad de Puno en un 18.44%, seguido de la escuela ALEXANDER FLEMING de la ciudad de Puno en un 14.15%, seguido de la escuela GREGOR MENDEL de la ciudad de Juliaca con un 13.46%, etc.

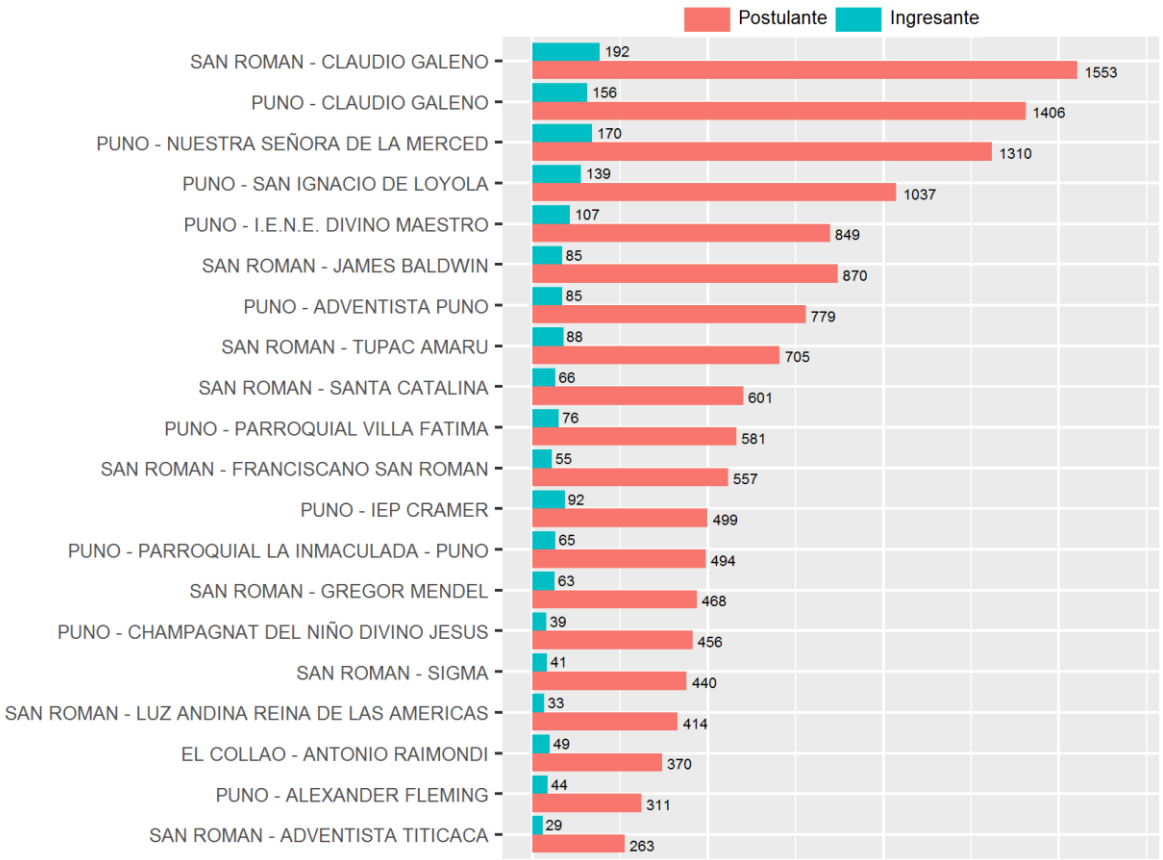


Figura 37. Cantidad de postulantes ingresantes de colegios privados

Cantidad de postulantes e ingresantes de la escuela estatal GRAN UNIDAD SAN CARLOS del proceso de admisión general

Cuadro 27. Porcentaje de ingresantes de la Gran Unidad Escolar San Carlos

	modalidad	fecha	escuela	post	ing	porc.ing
1	CEPREUNA	31_03_2013	GRAN UNIDAD ESCOLAR SAN CARLOS	167	22	13.17 %
2	GENERAL	07_04_2013	GRAN UNIDAD ESCOLAR SAN CARLOS	443	32	7.22 %
3	CEPREUNA	14_07_2013	GRAN UNIDAD ESCOLAR SAN CARLOS	150	26	17.33 %
4	GENERAL	18_08_2013	GRAN UNIDAD ESCOLAR SAN CARLOS	416	57	13.7 %
5	CEPREUNA	08_09_2013	GRAN UNIDAD ESCOLAR SAN CARLOS	81	19	23.46 %
6	CEPREUNA	22_12_2013	GRAN UNIDAD ESCOLAR SAN CARLOS	144	22	15.28 %
7	GENERAL	12_01_2014	GRAN UNIDAD ESCOLAR SAN CARLOS	405	54	13.33 %
8	CEPREUNA	16_03_2014	GRAN UNIDAD ESCOLAR SAN CARLOS	146	19	13.01 %
9	GENERAL	01_06_2014	GRAN UNIDAD ESCOLAR SAN CARLOS	420	57	13.57 %
10	CEPREUNA	15_06_2014	GRAN UNIDAD ESCOLAR SAN CARLOS	115	25	21.74 %
11	CEPREUNA	31_08_2014	GRAN UNIDAD ESCOLAR SAN CARLOS	135	14	10.37 %
12	GENERAL	14_09_2014	GRAN UNIDAD ESCOLAR SAN CARLOS	389	50	12.85 %
13	CEPREUNA	14_12_2014	GRAN UNIDAD ESCOLAR SAN CARLOS	131	12	9.16 %
14	GENERAL	18_01_2015	GRAN UNIDAD ESCOLAR SAN CARLOS	414	30	7.25 %
15	CEPREUNA	29_03_2015	GRAN UNIDAD ESCOLAR SAN CARLOS	190	32	16.84 %
16	GENERAL	07_06_2015	GRAN UNIDAD ESCOLAR SAN CARLOS	433	29	6.7 %
17	CEPREUNA	09_08_2015	GRAN UNIDAD ESCOLAR SAN CARLOS	150	30	20 %
18	GENERAL	06_09_2015	GRAN UNIDAD ESCOLAR SAN CARLOS	366	42	11.48 %
19	CEPREUNA	08_11_2015	GRAN UNIDAD ESCOLAR SAN CARLOS	92	15	16.3 %
20	CEPREUNA	10_01_2016	GRAN UNIDAD ESCOLAR SAN CARLOS	77	16	20.78 %
21	GENERAL	31_01_2016	GRAN UNIDAD ESCOLAR SAN CARLOS	330	28	8.48 %
22	CEPREUNA	20_03_2016	GRAN UNIDAD ESCOLAR SAN CARLOS	153	18	11.76 %
23	GENERAL	15_05_2016	GRAN UNIDAD ESCOLAR SAN CARLOS	315	35	11.11 %
24	CEPREUNA	26_06_2016	GRAN UNIDAD ESCOLAR SAN CARLOS	149	22	14.77 %
25	GENERAL	21_08_2016	GRAN UNIDAD ESCOLAR SAN CARLOS	425	45	10.59 %
26	CEPREUNA	18_09_2016	GRAN UNIDAD ESCOLAR SAN CARLOS	119	25	21.01 %
27	CEPREUNA	18_12_2016	GRAN UNIDAD ESCOLAR SAN CARLOS	150	14	9.33 %
28	GENERAL	22_01_2017	GRAN UNIDAD ESCOLAR SAN CARLOS	391	39	9.97 %

Del cuadro 27 se aprecia que la mayor cantidad de ingresantes se dio en un 13.7% en el proceso de admisión general del 18 de agosto del 2013 y un 23.46% en el cepreuna del 08 setiembre 2013.

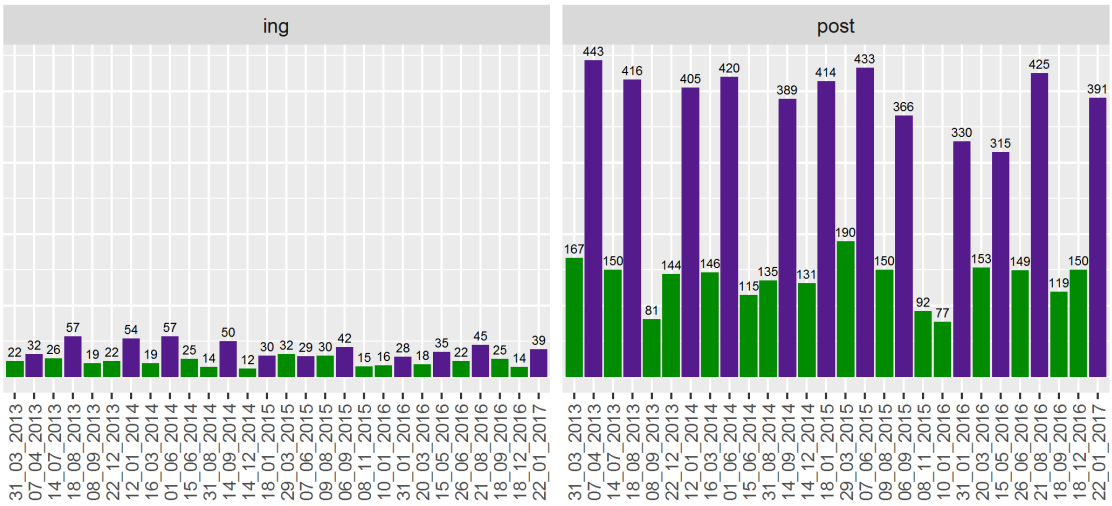


Figura 38. Cantidad de postulantes ingresantes de la Gran Unidad Escolar San Carlos

Cantidad de postulantes e ingresantes de la escuela estatal NUESTRA SEÑORA DE ALTA GRACIA del proceso de admisión General y Cepreuna

Cuadro 28. Porcentaje de ingresantes de Nuestra Señora de Alta Gracia

	modalidad	fecha	escuela	post	ing	porc.ing
1	CEPREUNA	31_03_2013	NUESTRA SEÑORA DE ALTA GRACIA	104	9	8.65 %
2	GENERAL	07_04_2013	NUESTRA SEÑORA DE ALTA GRACIA	184	16	8.7 %
3	CEPREUNA	14_07_2013	NUESTRA SEÑORA DE ALTA GRACIA	61	15	24.59 %
4	GENERAL	18_08_2013	NUESTRA SEÑORA DE ALTA GRACIA	207	23	11.11 %
5	CEPREUNA	08_09_2013	NUESTRA SEÑORA DE ALTA GRACIA	7	0	0 %
6	CEPREUNA	22_12_2013	NUESTRA SEÑORA DE ALTA GRACIA	43	9	20.93 %
7	GENERAL	12_01_2014	NUESTRA SEÑORA DE ALTA GRACIA	180	31	17.22 %
8	CEPREUNA	16_03_2014	NUESTRA SEÑORA DE ALTA GRACIA	42	13	30.95 %
9	GENERAL	01_06_2014	NUESTRA SEÑORA DE ALTA GRACIA	159	16	10.06 %
10	CEPREUNA	15_06_2014	NUESTRA SEÑORA DE ALTA GRACIA	52	12	23.08 %
11	CEPREUNA	31_08_2014	NUESTRA SEÑORA DE ALTA GRACIA	48	2	4.17 %
12	GENERAL	14_09_2014	NUESTRA SEÑORA DE ALTA GRACIA	187	24	12.83 %
13	CEPREUNA	14_12_2014	NUESTRA SEÑORA DE ALTA GRACIA	44	5	11.36 %
14	GENERAL	18_01_2015	NUESTRA SEÑORA DE ALTA GRACIA	172	11	6.4 %
15	CEPREUNA	29_03_2015	NUESTRA SEÑORA DE ALTA GRACIA	55	8	14.55 %
16	GENERAL	07_06_2015	NUESTRA SEÑORA DE ALTA GRACIA	167	10	5.99 %
17	CEPREUNA	09_08_2015	NUESTRA SEÑORA DE ALTA GRACIA	61	9	14.75 %
18	GENERAL	06_09_2015	NUESTRA SEÑORA DE ALTA GRACIA	132	21	15.91 %

19	CEPREUNA	08_11_2015	NUESTRA SEÑORA DE ALTA GRACIA	32	7	21.88 %
20	CEPREUNA	10_01_2016	NUESTRA SEÑORA DE ALTA GRACIA	31	3	9.68 %
21	GENERAL	31_01_2016	NUESTRA SEÑORA DE ALTA GRACIA	142	7	4.93 %
22	CEPREUNA	20_03_2016	NUESTRA SEÑORA DE ALTA GRACIA	24	3	12.5 %
23	GENERAL	15_05_2016	NUESTRA SEÑORA DE ALTA GRACIA	145	14	9.66 %
24	CEPREUNA	26_06_2016	NUESTRA SEÑORA DE ALTA GRACIA	50	8	16 %
25	GENERAL	21_08_2016	NUESTRA SEÑORA DE ALTA GRACIA	166	13	7.83 %
26	CEPREUNA	18_09_2016	NUESTRA SEÑORA DE ALTA GRACIA	38	2	5.26 %
27	CEPREUNA	18_12_2016	NUESTRA SEÑORA DE ALTA GRACIA	60	2	3.33 %
28	GENERAL	22_01_2017	NUESTRA SEÑORA DE ALTA GRACIA	174	12	6.9 %

Del cuadro 28 se aprecia que la mayor cantidad de ingresantes se dio en un 17.22% en el proceso de admisión general del 12 de enero del 2014.

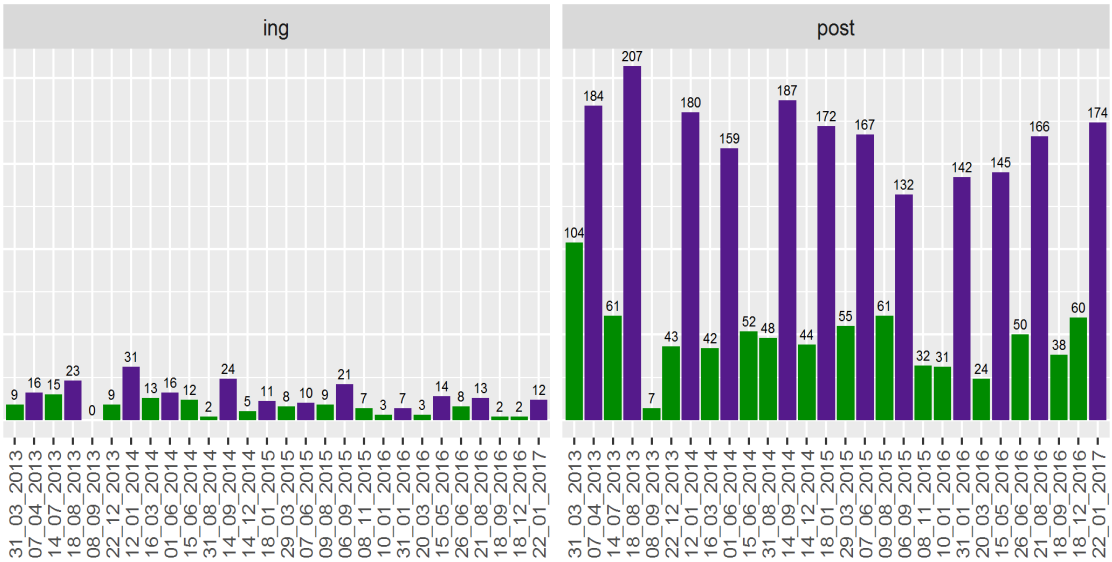


Figura 39. Cantidad de postulantes ingresantes de la escuela Nuestra Señora de Alta Gracia

Cantidad de postulantes e ingresantes de la escuela privada NUESTRA SEÑORA DE LA MERCED del proceso de admisión General y Cepreuna

Cuadro 29. Porcentaje de ingresantes de Nuestra Señora de la Merced

	modalidad	fecha	escuela	post	ing	porc.ing
1	CEPREUNA	31_03_2013	NUESTRA SEÑORA DE LA MERCED	42	8	19.05 %
2	GENERAL	07_04_2013	NUESTRA SEÑORA DE LA MERCED	79	14	17.72 %
3	CEPREUNA	14_07_2013	NUESTRA SEÑORA DE LA MERCED	23	3	13.04 %
4	GENERAL	18_08_2013	NUESTRA SEÑORA DE LA MERCED	59	11	18.64 %
5	CEPREUNA	08_09_2013	NUESTRA SEÑORA DE LA MERCED	16	7	43.75 %
6	CEPREUNA	22_12_2013	NUESTRA SEÑORA DE LA MERCED	26	5	19.23 %
7	GENERAL	12_01_2014	NUESTRA SEÑORA DE LA MERCED	69	6	8.7 %
8	CEPREUNA	16_03_2014	NUESTRA SEÑORA DE LA MERCED	36	9	25 %
9	GENERAL	01_06_2014	NUESTRA SEÑORA DE LA MERCED	83	15	18.07 %
10	CEPREUNA	15_06_2014	NUESTRA SEÑORA DE LA MERCED	27	3	11.11 %
11	CEPREUNA	31_08_2014	NUESTRA SEÑORA DE LA MERCED	22	1	4.55 %
12	GENERAL	14_09_2014	NUESTRA SEÑORA DE LA MERCED	72	8	11.11 %
13	CEPREUNA	14_12_2014	NUESTRA SEÑORA DE LA MERCED	26	1	3.85 %
14	GENERAL	18_01_2015	NUESTRA SEÑORA DE LA MERCED	83	6	7.23 %
15	CEPREUNA	29_03_2015	NUESTRA SEÑORA DE LA MERCED	38	3	7.89 %
16	GENERAL	07_06_2015	NUESTRA SEÑORA DE LA MERCED	93	10	10.75 %
17	CEPREUNA	09_08_2015	NUESTRA SEÑORA DE LA MERCED	28	5	17.86 %
18	GENERAL	06_09_2015	NUESTRA SEÑORA DE LA MERCED	60	10	16.67 %
19	CEPREUNA	08_11_2015	NUESTRA SEÑORA DE LA MERCED	12	0	0 %
20	CEPREUNA	10_01_2016	NUESTRA SEÑORA DE LA MERCED	21	0	0 %
21	GENERAL	31_01_2016	NUESTRA SEÑORA DE LA MERCED	61	6	9.84 %
22	CEPREUNA	20_03_2016	NUESTRA SEÑORA DE LA MERCED	30	1	3.33 %
23	GENERAL	15_05_2016	NUESTRA SEÑORA DE LA MERCED	54	6	11.11 %
24	CEPREUNA	26_06_2016	NUESTRA SEÑORA DE LA MERCED	29	3	10.34 %
25	GENERAL	21_08_2016	NUESTRA SEÑORA DE LA MERCED	78	13	16.67 %
26	CEPREUNA	18_09_2016	NUESTRA SEÑORA DE LA MERCED	31	2	6.45 %
27	CEPREUNA	18_12_2016	NUESTRA SEÑORA DE LA MERCED	30	2	6.67 %
28	GENERAL	22_01_2017	NUESTRA SEÑORA DE LA MERCED	70	5	7.14 %

Del cuadro 29 se aprecia que la mayor cantidad de ingresantes se dio en un 19.05% en el proceso de admisión cepreuna del 31 de marzo del 2013.

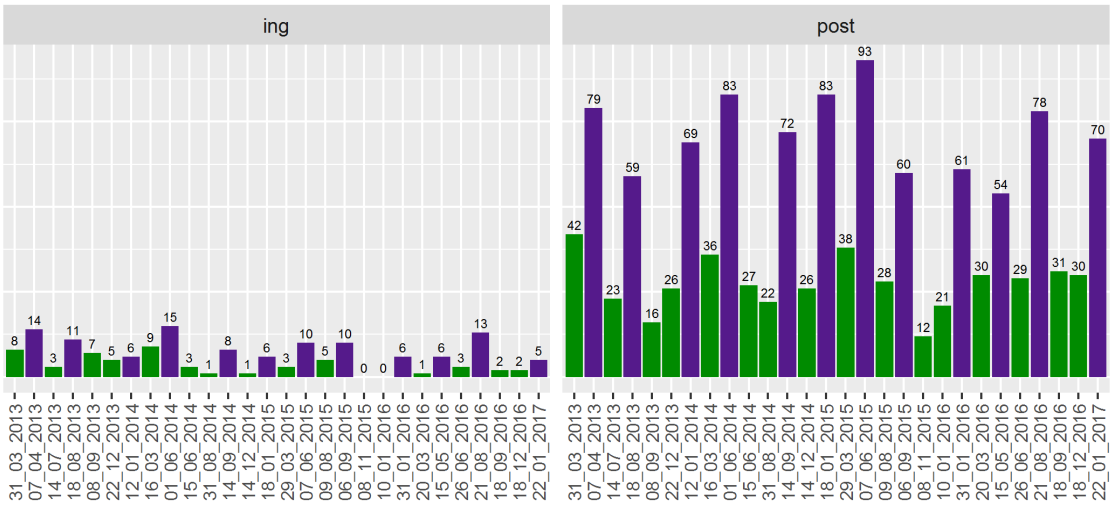


Figura 40. Cantidad de postulantes ingresantes de la escuela Nuestra Señora de la Merced

4.1.3. Verificación de los datos

Finalizado la exploración de los datos iniciales se pueden concluir que estos son completos, es decir, no tienen valores NULL ya que fueron reemplazados con cero (0) para aquellas Escuelas Profesionales que no obtuvieron ingresantes además se agregó nuevas columnas con la sintaxis que ofrece el lenguaje de programación R. Los datos son precisos en este punto lo que garantiza la completitud de la información.

4.2. PREPARACIÓN DE LOS DATOS

En esta fase se preparará los datos para adecuarlos a las técnicas de agrupamiento usando el paquete dplyr. Esto implica seleccionar el conjunto de datos que se va a usar, limpiarlos para mejorar su calidad, añadir nuevos datos a partir de los existentes y darles el formato requerido.

4.2.1. Seleccionar los datos

Se ha utilizado todos los registros de la base de datos, sin embargo, existen campos que no se han usado. Los campos usados para la selección de

datos son los siguientes:

Tabla c_carrera: id_carrera, id_grupo, nombre

Tabla c_postulante: c_idcarrera, ubi_nacimiento, codigo_colegio

Tabla c_colegio: codigo, tipo_colegio, area

Tabla c_departamento: idDep, nombre

Tabla c_provincia: idPRov, idDep, nombre

Tabla c_distrito: idDist, idProv, nombre

4.2.2. Limpiar los datos

Como se mencionó, se tiene acceso a toda la base de datos, sin embargo, se a detectado que existen nombres de postulantes que difieren entre procesos, esto debido a que el postulante al momento de su inscripción cometieron el error de digitar mal sus datos. Esta información fue limpiada consultando el dni del postulante en la página Web de la RENIEC.

Para general el modelo relacionado con la predicción de la tendencia de postulantes e ingresantes se uso la función filter del paquete dplyr, que es muy similar al WHERE de la instrucción SELECT.

4.2.3. Construir los datos

En esta fase no se ha agregado ningún campo adicional a la base de datos ya que los verbos del paquete dplyr permitieron agregar campos nuevos sobre las variables donde se capturaron los datos.

4.2.4. Integrar los datos

No ha sido necesaria la creación de nuevas estructuras (campos, registros,

etc.), ni la fusión entre distintas tablas de la base de datos, ya que el paquete dplyr se encarga de realizar estas tareas usando la función mutate; finalizada esta etapa se tiene lista los datos para ser usados en el modelo de acuerdo a las funciones creadas anteriormente.

4.3. MODELADO

En esta fase se escogió las técnicas adecuadas que cumplieron los objetivos del presente trabajo, ésta técnica se aplicó sobre los datos para generar el modelo y por último se avalúo si dicho modelo cumplió los criterios de éxito.

4.3.1. Escoger la técnica de modelado

R provee una serie de funciones implementados para generar modelos. La función que se usó fue **lm** (regresión lineal) que almacena toda la información del modelo, para ver su contenido se empleó la función **names** y para visualizar los principales parámetros del modelo generado se usó la función **summary,** igualmente se usó la función **poly** (polinómico) del paquete polynom, ambas funciones se adaptan a los objetivos del proyecto.

4.3.2. Construir el modelo

Se ejecutó el modelo elegido sobre los datos de las agrupaciones de los postulantes e ingresantes calculando los coeficientes del modelo lineal y polinómico. Dado que las fechas de los procesos de admisión no tienen una distancia constante, esta fue convertida a su valor numérico tal como se muestra en el código implementado. Se generó

el modelo para ciertas Escuelas Profesionales.

```
data <- datacarrera %>% filter(modalidad_tipo == vproceso)

print(paste(vproceso, ". ",data$escuela_profesional[1], sep = ""))
if(grado==0){
  modelo <- lm(data$total~data$fecha)
}else{
  modelo <- lm(data$total~poly(data$fecha, degree = grado))
}
coeff <- coef(modelo)
# coeficiente de correlacion
coef.correl <- cor(as.numeric(data$fecha),data$total,
                   use = "everything", method = "pearson")
print(paste("coef.correlación: ", coef.correl))
# coeficiente de determinacion
coed.determ<-summary(modelo)$r.squared
print(paste("coef.determinación: ", coed.determ))
print(summary(modelo))
```

Escuela Profesional de Medicina, Veterinaria y Zootecnia

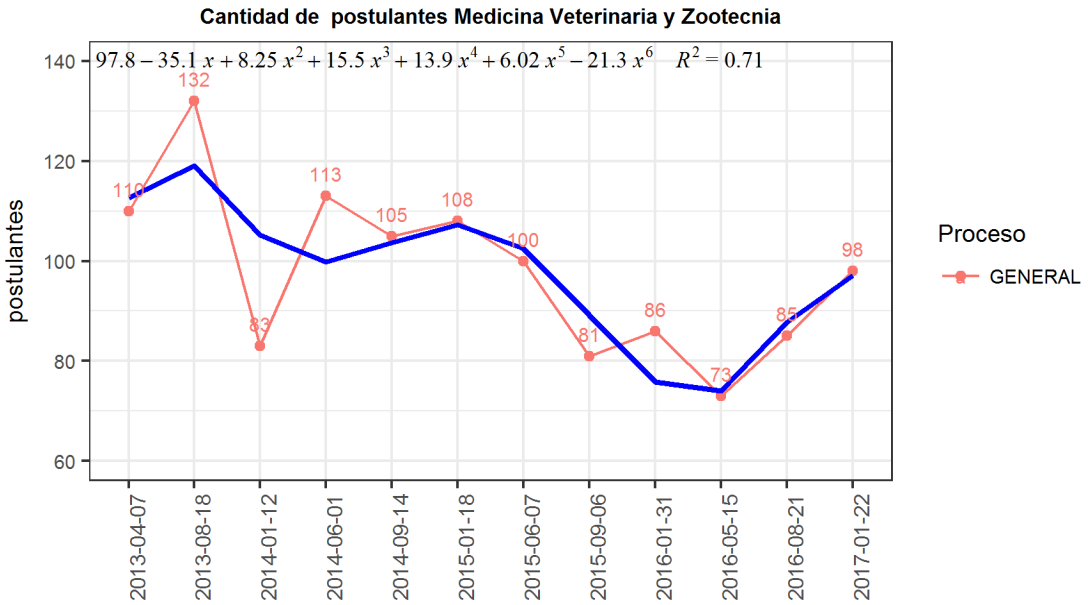


Figura 41. Modelo y resultados de la cantidad de postulantes de Med. Vet. y Zoot. - general

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
      1       2       3       4       5       6       7       8
-2.1037  9.9278 -19.0987  14.2233  0.5171 -0.9391 -1.1204 -8.8107
      9      10      11      12
11.9348 -2.2968 -3.0525  0.8187

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       97.833      3.890  25.152 1.85e-06 ***
poly(data$fecha, degree = grado)1 -35.072     13.474  -2.603   0.0481 *
poly(data$fecha, degree = grado)2   8.249     13.474   0.612   0.5672
poly(data$fecha, degree = grado)3  15.490     13.474   1.150   0.3023
poly(data$fecha, degree = grado)4  13.866     13.474   1.029   0.3506
poly(data$fecha, degree = grado)5   6.022     13.474   0.447   0.6736
poly(data$fecha, degree = grado)6 -21.339     13.474  -1.584   0.1741
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.47 on 5 degrees of freedom
Multiple R-squared:  0.7099,    Adjusted R-squared:  0.3619
F-statistic:  2.04 on 6 and 5 DF,  p-value: 0.2256
```

De la figura 41 se puede observar que Residual Standard Error (RSE) se aleja 13.47 unidades del verdadero valor de la cantidad de postulantes; además, el R-squared (R^2) indica que el predictor poly(data$fecha, degree=grado) es capaza de explicar el 70.9% de la variabilidad observada en la cantidad de postulantes.

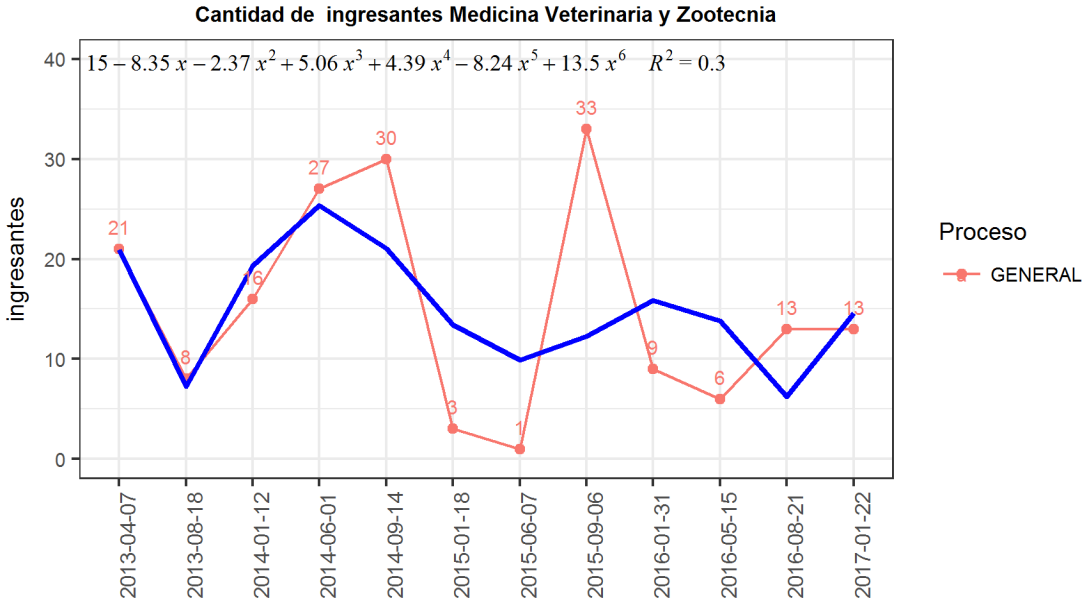


Figura 42. Modelo y resultados de la cantidad de ingresantes de Med. Vet. y Zoot. - general

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
      1       2       3       4       5       6       7       8
 -0.2246  1.6287 -4.4960  2.0091  9.1545 -10.9274  -9.9940  20.5579
      9      10      11      12
 -6.1528 -6.7076  6.0004 -0.8482

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                         15.000      3.813   3.934    0.011 *
poly(data$fecha, degree = grado)1   -8.352     13.208  -0.632    0.555
poly(data$fecha, degree = grado)2   -2.372     13.208  -0.180    0.865
poly(data$fecha, degree = grado)3    5.058     13.208   0.383    0.718
poly(data$fecha, degree = grado)4    4.394     13.208   0.333    0.753
poly(data$fecha, degree = grado)5   -8.238     13.208  -0.624    0.560
poly(data$fecha, degree = grado)6   13.550     13.208   1.026    0.352
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.21 on 5 degrees of freedom
Multiple R-squared:  0.2988,    Adjusted R-squared:  -0.5426
F-statistic: 0.3552 on 6 and 5 DF,  p-value: 0.8801
```

De la figura 42 se puede observar que Residual Standard Error (RSE) se aleja 13.47 unidades del verdadero valor de la cantidad de ingresantes; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 30% de la variabilidad observada en la cantidad de ingresantes.

Educación Primaria

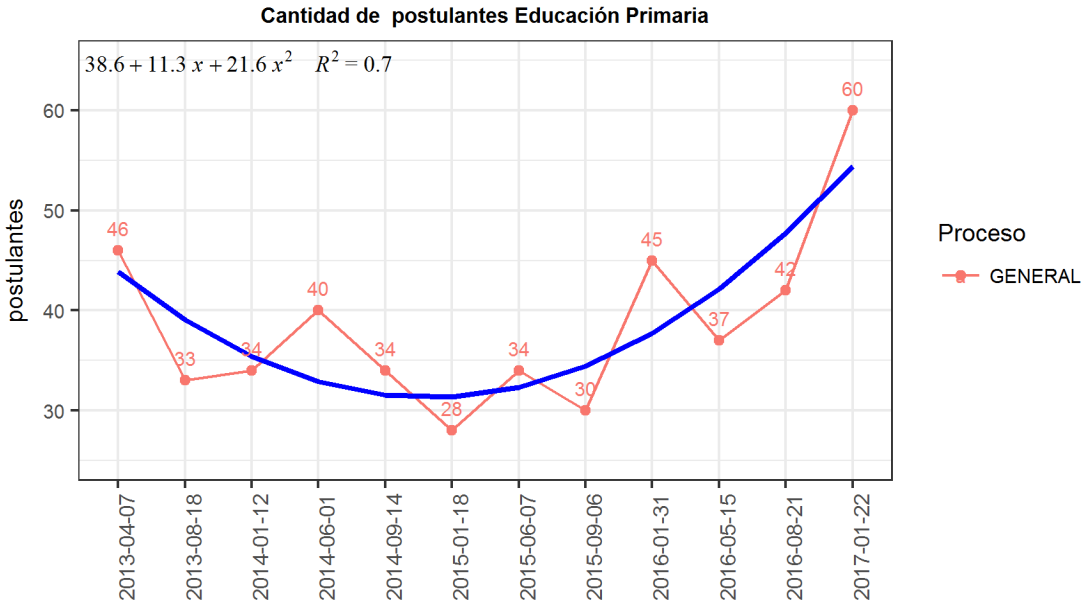


Figura 43. Modelo y resultados de la cantidad de postulantes de Educación Primaria - general

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
    Min      1Q  Median      3Q     Max
-6.0114 -4.5292  0.2198  3.1980  7.7079

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       38.583      1.543  25.010 1.25e-09 ***
poly(data$fecha, degree = grado)1   11.335      5.344   2.121  0.06292 .
poly(data$fecha, degree = grado)2   21.573      5.344   4.037  0.00294 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.344 on 9 degrees of freedom
Multiple R-squared:  0.6979,    Adjusted R-squared:  0.6308
F-statistic:  10.4 on 2 and 9 DF,  p-value: 0.004576
```

De la figura 43 se puede observar que Residual Standard Error (RSE) se aleja 5.34 unidades del verdadero valor de la cantidad de postulantes; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 70% de la variabilidad observada en la cantidad de postulantes.

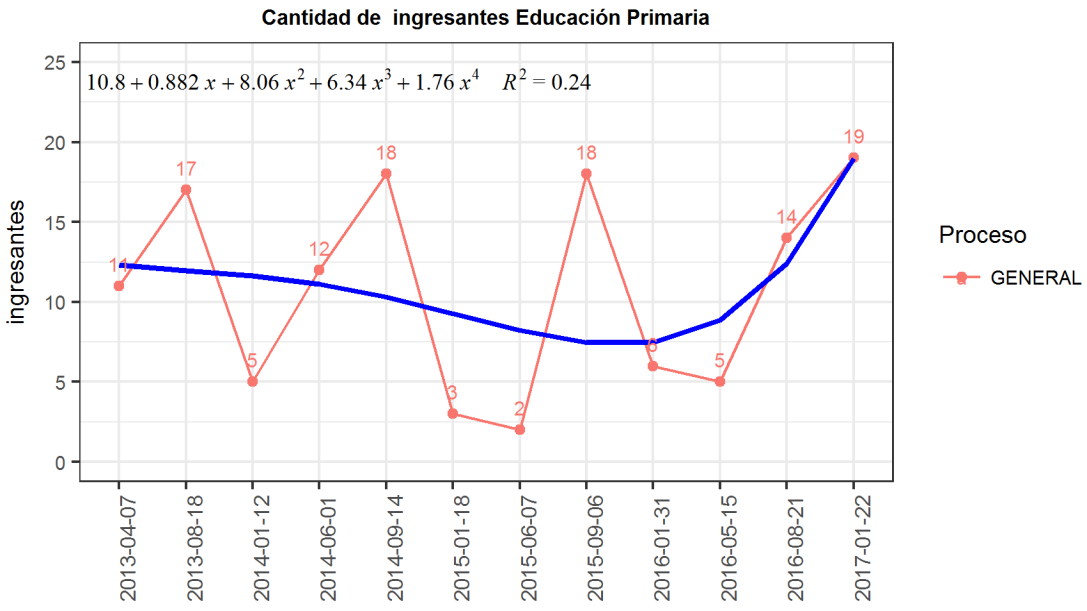


Figura 44. Modelo y resultados de la cantidad de ingresantes de Educación Primaria -general

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
   Min     1Q Median     3Q    Max
-6.926 -4.347 -0.717  3.366  9.936

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      10.8333     2.1692   4.994  0.00247 **
poly(data$fecha, degree = grado)1   0.8821     7.5144   0.117  0.91039
poly(data$fecha, degree = grado)2   8.0558     7.5144   1.072  0.32491
poly(data$fecha, degree = grado)3   6.3360     7.5144   0.843  0.43145
poly(data$fecha, degree = grado)4   1.7567     7.5144   0.234  0.82293
poly(data$fecha, degree = grado)5   1.4031     7.5144   0.187  0.85803
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.514 on 6 degrees of freedom
Multiple R-squared:  0.2466,    Adjusted R-squared:  -0.3813
F-statistic: 0.3927 on 5 and 6 DF,  p-value: 0.838
```

De la figura 44 se puede observar que Residual Standard Error (RSE) se aleja 7.51 unidades del verdadero valor de la cantidad de ingresantes; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 24% de la variabilidad observada en la cantidad de ingresantes.

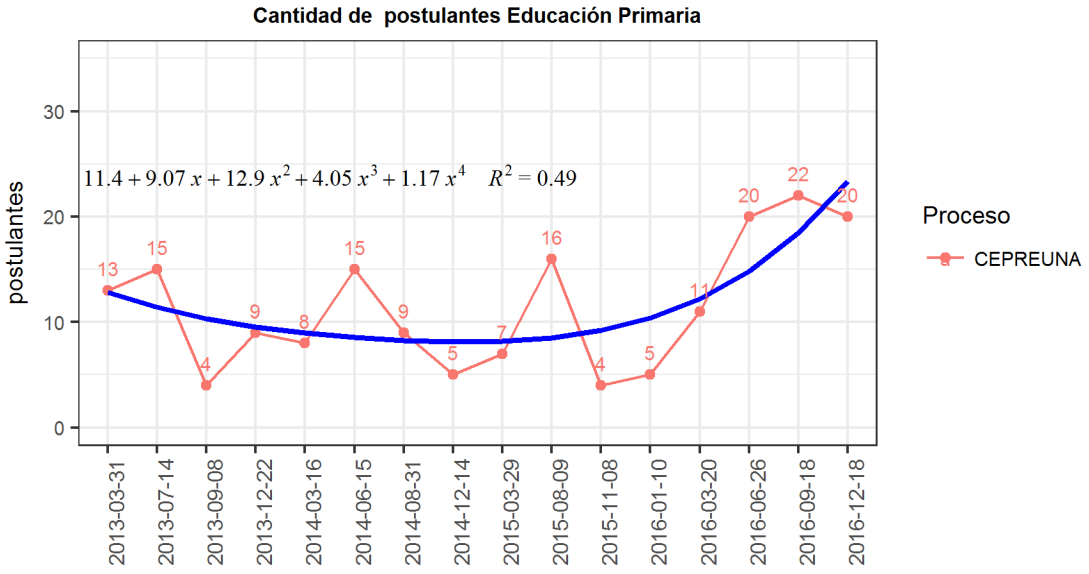


Figura 45. Modelo y resultados de la cantidad de postulantes de Educación Primaria - cepreuna

```
Residuals:
    Min      1Q  Median      3Q     Max
-6.5585 -3.1175 -0.8208  3.7062  7.4201

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                        11.438      1.257   9.096 1.89e-06 ***
poly(data$fecha, degree = grado)1   9.071      5.029   1.804   0.0987 .
poly(data$fecha, degree = grado)2  12.870      5.029   2.559   0.0266 *
poly(data$fecha, degree = grado)3   4.050      5.029   0.805   0.4377
poly(data$fecha, degree = grado)4   1.167      5.029   0.232   0.8208
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.029 on 11 degrees of freedom
Multiple R-squared:  0.4884,    Adjusted R-squared:  0.3024
F-statistic: 2.626 on 4 and 11 DF,  p-value: 0.09237
```

De la figura 45 se puede observar que Residual Standard Error (RSE) se aleja 5.02 unidades del verdadero valor de la cantidad de ingresantes; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 48.8% de la variabilidad observada en la cantidad de ingresantes.

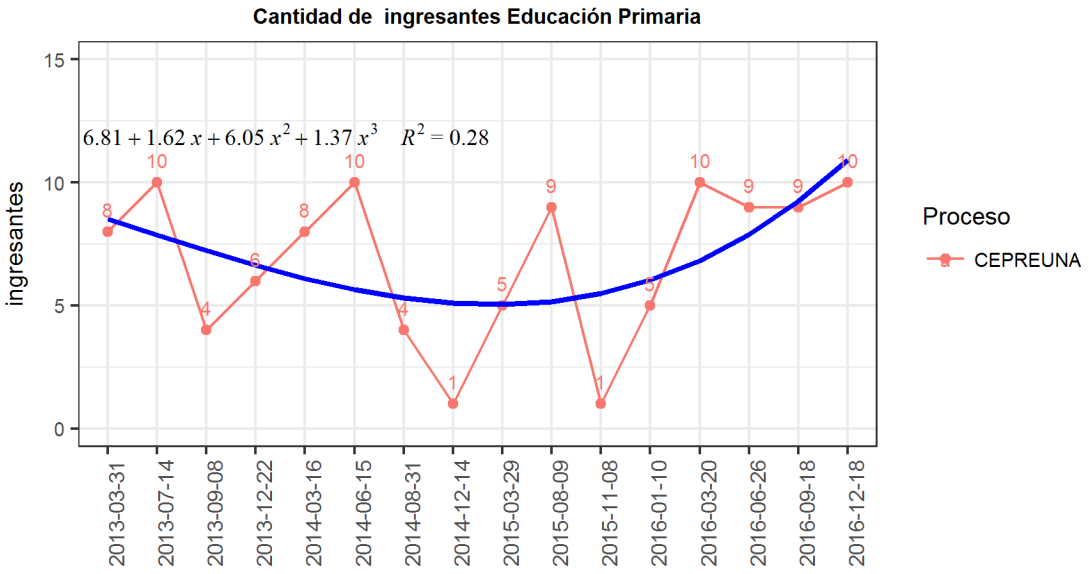


Figura 46. Modelo y resultados de la cantidad de ingresantes de Educación Primaria - cepreuna

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
    Min      1Q  Median      3Q     Max
-4.6383 -1.1525 -0.2575  1.9421  4.2821

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       6.8125     0.7477   9.112 9.69e-07 ***
poly(data$fecha, degree = grado)1   1.6205     2.9906   0.542   0.5978
poly(data$fecha, degree = grado)2   6.0503     2.9906   2.023   0.0659 .
poly(data$fecha, degree = grado)3   1.3705     2.9906   0.458   0.6549
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.991 on 12 degrees of freedom
Multiple R-squared:  0.277,     Adjusted R-squared:  0.0962
F-statistic: 1.532 on 3 and 12 DF,  p-value: 0.2568
```

De la figura 46 se puede observar que Residual Standard Error (RSE) se aleja 2.99 unidades del verdadero valor de la cantidad de ingresantes; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 27.7% de la variabilidad observada en la cantidad de ingresantes.

Educación Inicial

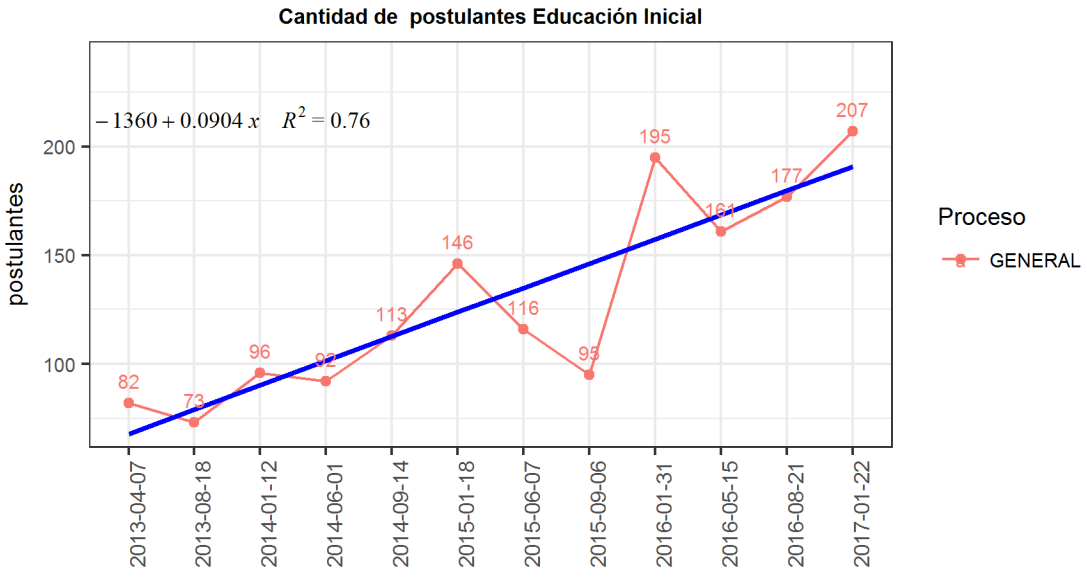


Figura 47. Modelo y resultados de la cantidad de postulantes de Educación Inicial - general

```
Call:
lm(formula = data$total ~ data$fecha)

Residuals:
    Min      1Q  Median      3Q     Max
-50.399  -8.296  -0.083  16.108  36.308

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.363e+03  2.623e+02  -5.198 0.000403 ***
data$fecha   9.043e-02  1.588e-02   5.693 0.000200 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23.51 on 10 degrees of freedom
Multiple R-squared:  0.7642,    Adjusted R-squared:  0.7406
F-statistic: 32.41 on 1 and 10 DF,  p-value: 0.0002002
```

De la figura 47, la información devuelta por el summary se observa que el p-value del estadístico F es pequeño, indicando que al menos el predictor del modelo está significativamente relacionado con la variable respuesta (postulantes). El R-squared (R^2) empleado en el modelo es capaz de explicar el 76.4% de la variabilidad observada en la cantidad de postulantes.

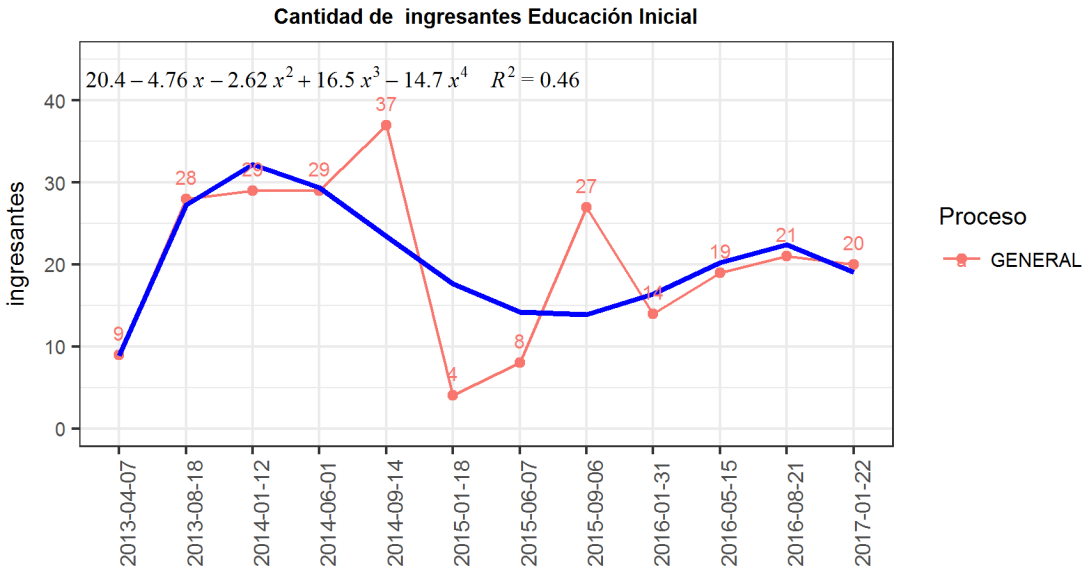


Figura 48. Modelo y resultados de la cantidad de ingresantes de Educación Inicial - general

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
     Min       1Q   Median       3Q      Max
-13.9525  -2.8429  -0.1628   0.3798  13.4647

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       20.417      2.681   7.616 0.000125 ***
poly(data$fecha, degree = grado)1  -4.761      9.286  -0.513 0.623901
poly(data$fecha, degree = grado)2  -2.624      9.286  -0.283 0.785705
poly(data$fecha, degree = grado)3  16.490      9.286   1.776 0.119015
poly(data$fecha, degree = grado)4 -14.691      9.286  -1.582 0.157650
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.286 on 7 degrees of freedom
Multiple R-squared:  0.4615,    Adjusted R-squared:  0.1538
F-statistic:   1.5 on 4 and 7 DF,  p-value: 0.2997
```

De la figura 48 se puede observar que Residual Standard Error (RSE) se aleja 9.28 unidades del verdadero valor de la cantidad de ingresantes; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 46% de la variabilidad observada en la cantidad de ingresantes.

Ingeniería Estadística e Informática

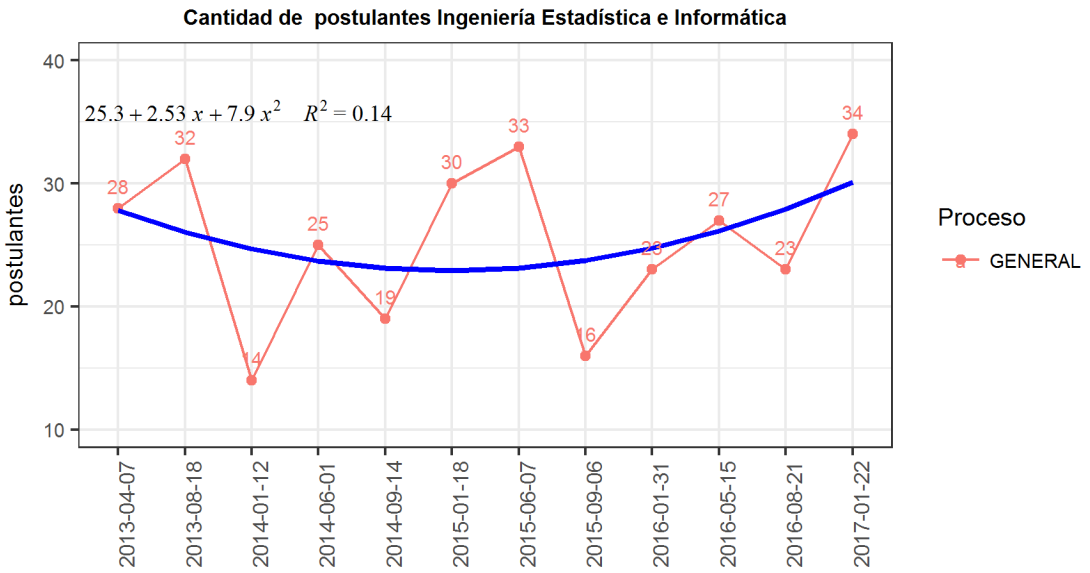


Figura 49. Modelo y resultados de la cantidad de postulantes de Ing. Estadística e Informática - general

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
     Min       1Q   Median       3Q      Max
-10.4127  -4.1056   0.2867   3.9992   9.8930

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       25.333      1.943  13.037 3.79e-07 ***
poly(data$fecha, degree = grado)1  2.530      6.731   0.376    0.716
poly(data$fecha, degree = grado)2  7.903      6.731   1.174    0.271
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.731 on 9 degrees of freedom
Multiple R-squared:  0.1445,    Adjusted R-squared:  -0.04567
F-statistic: 0.7598 on 2 and 9 DF,  p-value: 0.4956
```

De la figura 49 se observa que Residual Standard Error (RSE) se aleja 6.73 unidades del verdadero valor de la cantidad de postulantes; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 14.46% de la variabilidad observada en la cantidad de postulantes.

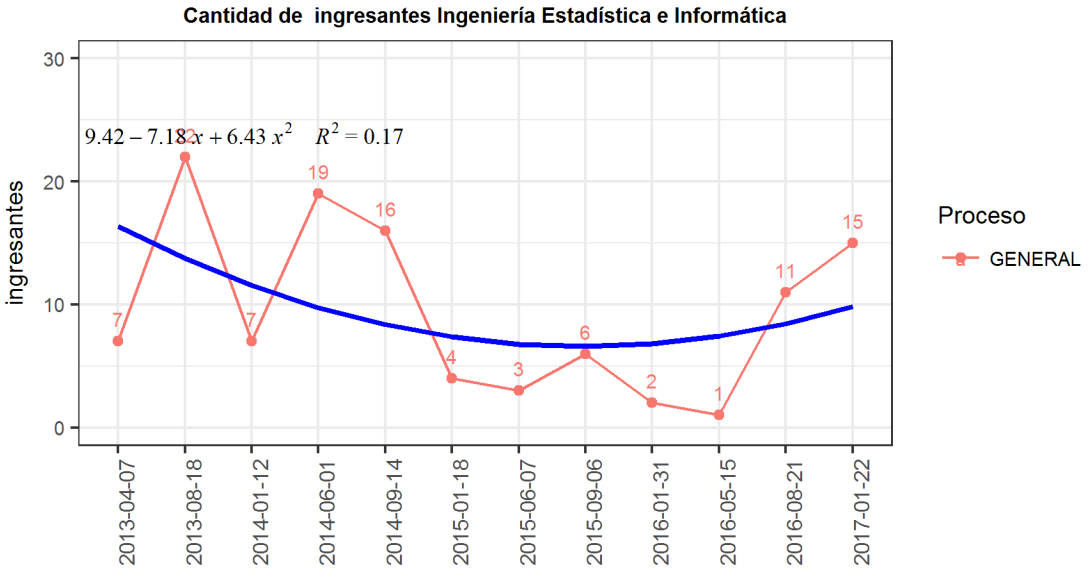


Figura 51. Modelo y resultados de la cantidad de ingresantes de Ing. Estadística e Informática - general

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
   Min     1Q Median     3Q    Max
-9.127 -4.512 -2.290  6.004  9.424

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                        9.417      2.050   4.593   0.0013 **
poly(data$fecha, degree = grado)1  -7.178      7.103  -1.011   0.3386
poly(data$fecha, degree = grado)2   6.430      7.103   0.905   0.3889
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.103 on 9 degrees of freedom
Multiple R-squared:  0.1698,    Adjusted R-squared:  -0.01467
F-statistic: 0.9205 on 2 and 9 DF,  p-value: 0.4328
```

De la figura 51 se observa que Residual Standard Error (RSE) se aleja 7.1 unidades del verdadero valor de la cantidad de ingresantes; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 16.9% de la variabilidad observada en la cantidad de postulantes.

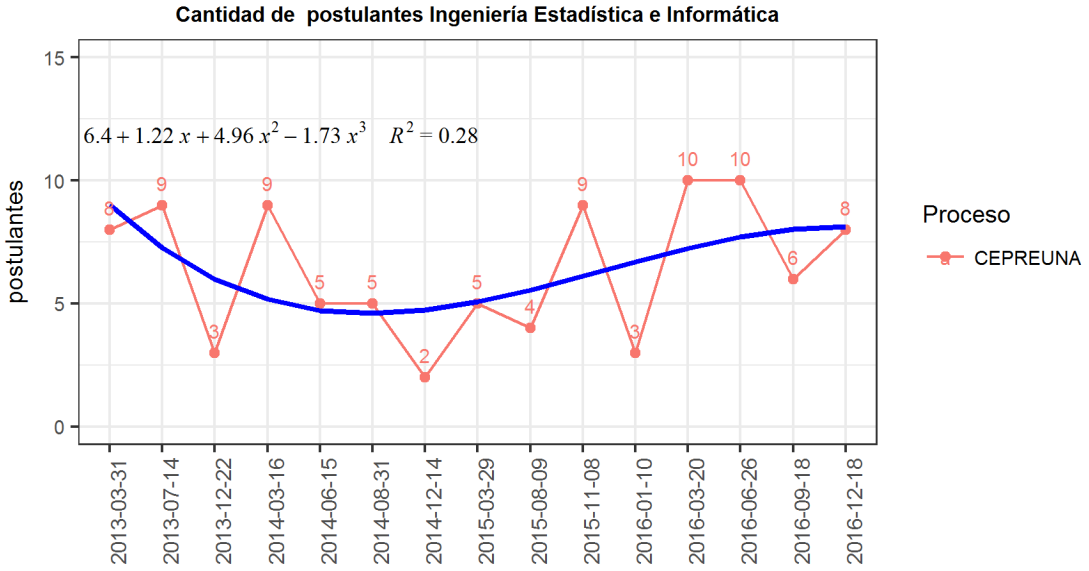


Figura 52. Modelo y resultados de la cantidad de postulantes de Ing. Estadística e Informática - cepreuna

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
    Min      1Q  Median      3Q     Max
-3.5731 -1.8411 -0.0897  2.0259  3.8876

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       6.4000     0.6807   9.402 1.36e-06 ***
poly(data$fecha, degree = grado)1   1.2223     2.6365   0.464   0.6520
poly(data$fecha, degree = grado)2   4.9634     2.6365   1.883   0.0865 .
poly(data$fecha, degree = grado)3  -1.7346     2.6365  -0.658   0.5241
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.636 on 11 degrees of freedom
Multiple R-squared:  0.2759,    Adjusted R-squared:  0.07845
F-statistic: 1.397 on 3 and 11 DF,  p-value: 0.2954
```

De la figura 52 se observa que Residual Standard Error (RSE) se aleja 2.63 unidades del verdadero valor de la cantidad de postulantes; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 27.59% de la variabilidad observada en la cantidad de postulantes.

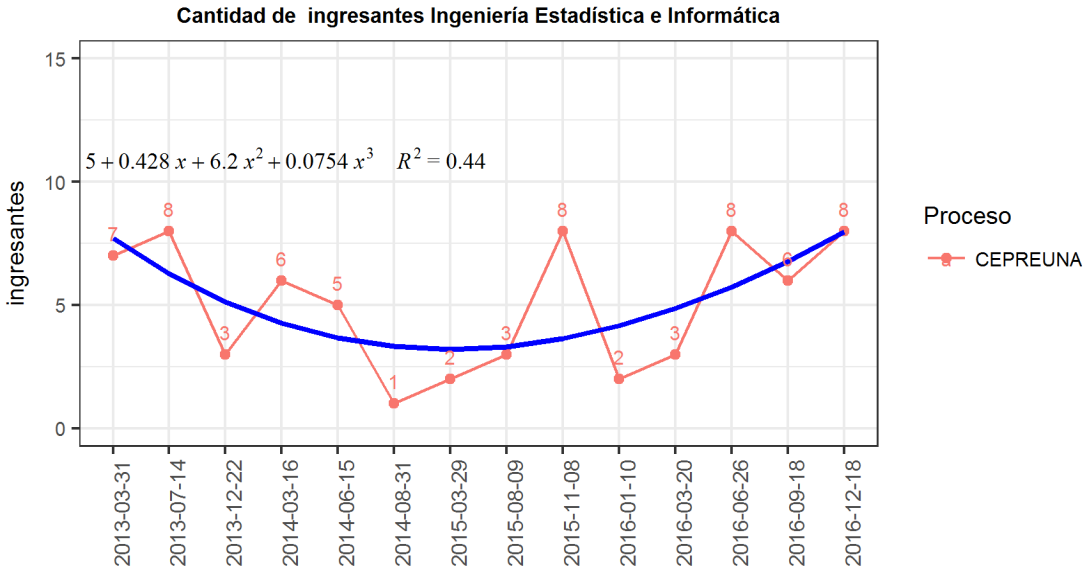


Figura 53. Modelo y resultados de la cantidad de ingresantes de Ing. Estadística e Informática - cepreuna

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
    Min      1Q  Median      3Q     Max
-2.2548 -1.5997 -0.5777  1.4914  4.2011

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      5.00000    0.59382   8.420  7.5e-06 ***
poly(data$fecha, degree = grado)1  0.42769    2.22187   0.192    0.8512
poly(data$fecha, degree = grado)2  6.20034    2.22187   2.791    0.0191 *
poly(data$fecha, degree = grado)3  0.07539    2.22187   0.034    0.9736
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.222 on 10 degrees of freedom
Multiple R-squared:  0.439,     Adjusted R-squared:  0.2707
F-statistic: 2.609 on 3 and 10 DF,  p-value: 0.1095
```

De la figura 53 se observa que Residual Standard Error (RSE) se aleja 2.22 unidades del verdadero valor de la cantidad de ingresantes; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 43.9% de la variabilidad observada en la cantidad de ingresantes.

Ingeniería Química

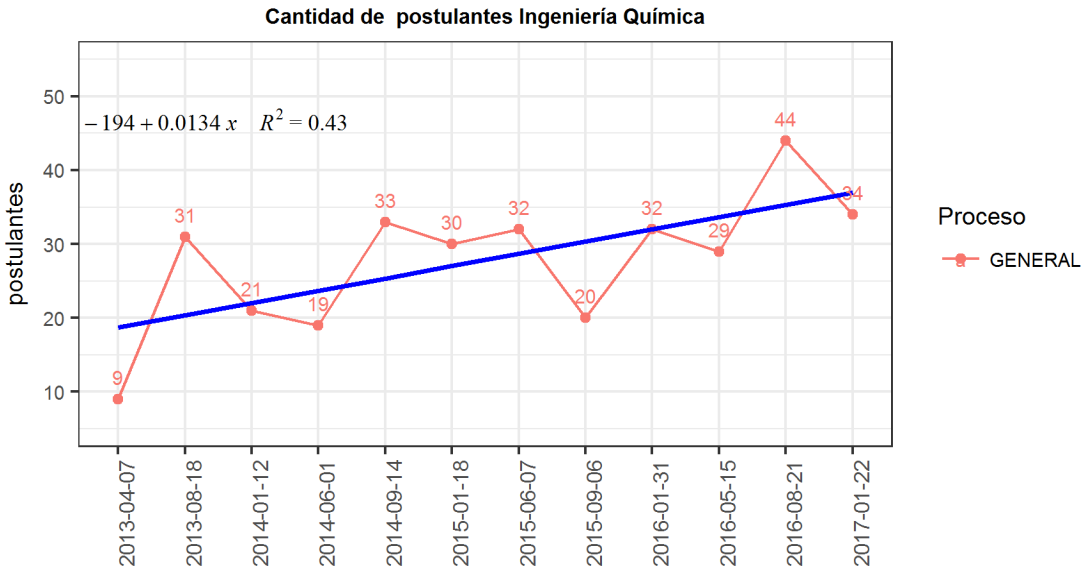


Figura 54. Modelo y resultados de la cantidad de postulantes de Ing. Química - general

```
Call:
lm(formula = data$total ~ data$fecha)

Residuals:
     Min       1Q   Median       3Q      Max
-10.2041  -4.6899  -0.6526   4.1585  10.8422

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.936e+02  8.073e+01  -2.398   0.0374 *
data$fecha   1.341e-02  4.889e-03   2.744   0.0207 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.236 on 10 degrees of freedom
Multiple R-squared:  0.4295,    Adjusted R-squared:  0.3724
F-statistic: 7.527 on 1 and 10 DF,  p-value: 0.0207
```

De la figura 54 la información devuelta por el summary se observa que el p-value del estadístico F es pequeño, indicando que al menos el predictor del modelo está relacionado con la variable respuesta (postulantes). El R-squared (R^2) empleado en el modelo es capaz de explicar el 42.9% de la variabilidad observada en la cantidad de postulantes.

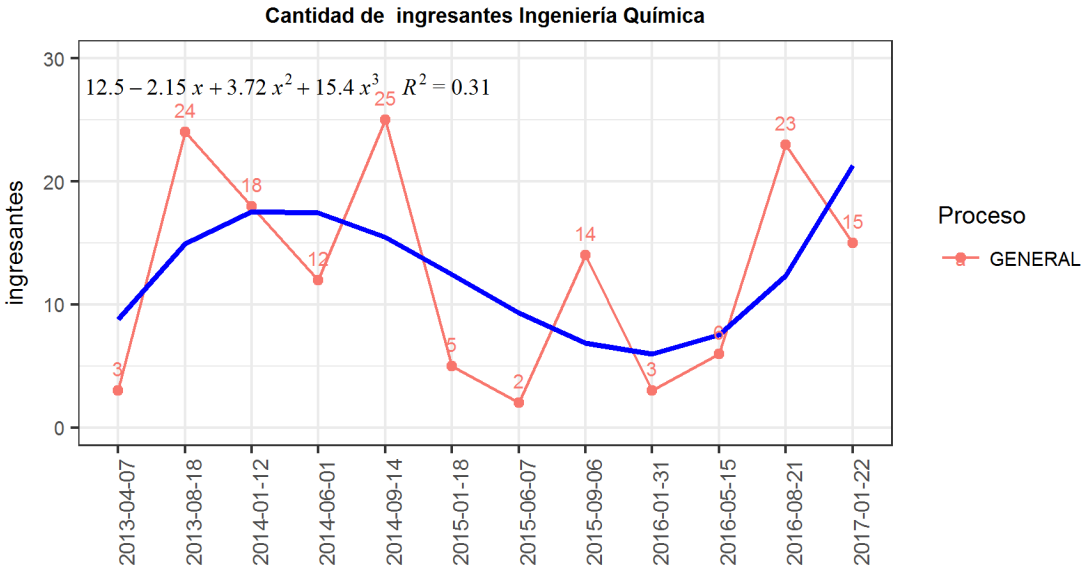


Figura 55. Modelo y resultados de la cantidad de ingresantes de Ing. Química - general

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
   Min     1Q Median     3Q    Max
-7.523 -5.630 -2.640  7.123 11.979

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       12.500      2.440   5.124 0.000903 ***
poly(data$fecha, degree = grado)1  -2.154      8.451  -0.255 0.805212
poly(data$fecha, degree = grado)2   3.715      8.451   0.440 0.671818
poly(data$fecha, degree = grado)3  15.403      8.451   1.823 0.105821
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.451 on 8 degrees of freedom
Multiple R-squared:  0.3092,    Adjusted R-squared:  0.05012
F-statistic: 1.193 on 3 and 8 DF,  p-value: 0.3721
```

De la figura 55 se observa que Residual Standard Error (RSE) se aleja 8.45 unidades del verdadero valor de la cantidad de ingresantes; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 30.9% de la variabilidad observada en la cantidad de ingresantes.

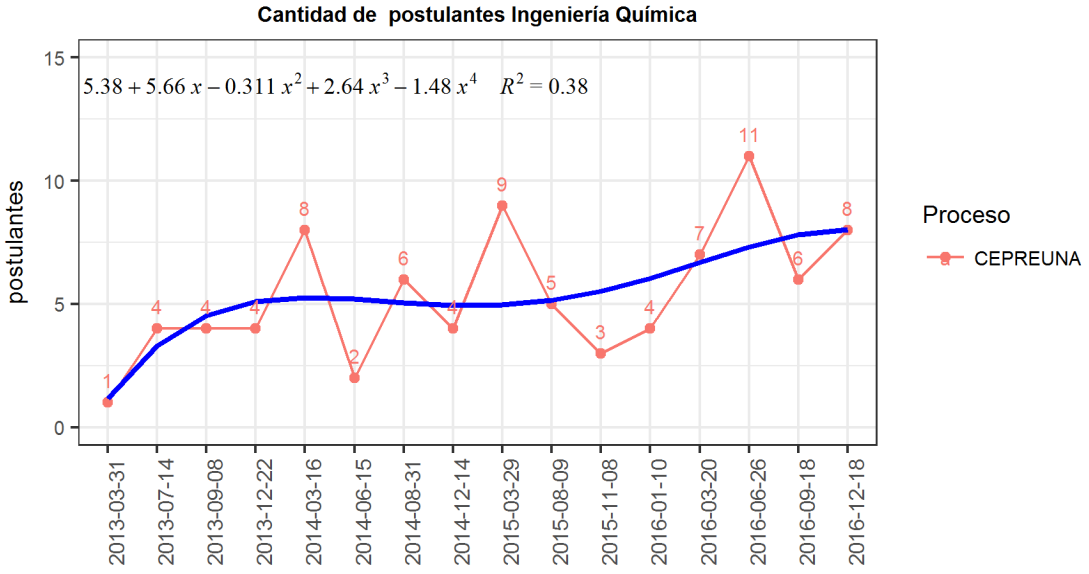


Figura 56. Modelo y resultados de la cantidad de postulantes de Ing. Química - cepreuna

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
    Min      1Q  Median      3Q     Max
-3.2892 -1.2344 -0.1884  0.6156  4.0068

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       5.3750     0.6142   8.751 2.76e-06 ***
poly(data$fecha, degree = grado)1   5.6647     2.4569   2.306   0.0416 *
poly(data$fecha, degree = grado)2  -0.3108     2.4569  -0.127   0.9016
poly(data$fecha, degree = grado)3   2.6406     2.4569   1.075   0.3055
poly(data$fecha, degree = grado)4  -1.4798     2.4569  -0.602   0.5592
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.457 on 11 degrees of freedom
Multiple R-squared:  0.3837,    Adjusted R-squared:  0.1596
F-statistic: 1.712 on 4 and 11 DF,  p-value: 0.217
```

De la figura 56 se observa que Residual Standard Error (RSE) se aleja 2.45 unidades del verdadero valor de la cantidad de ingresantes; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 38.3% de la variabilidad observada en la cantidad de ingresantes.

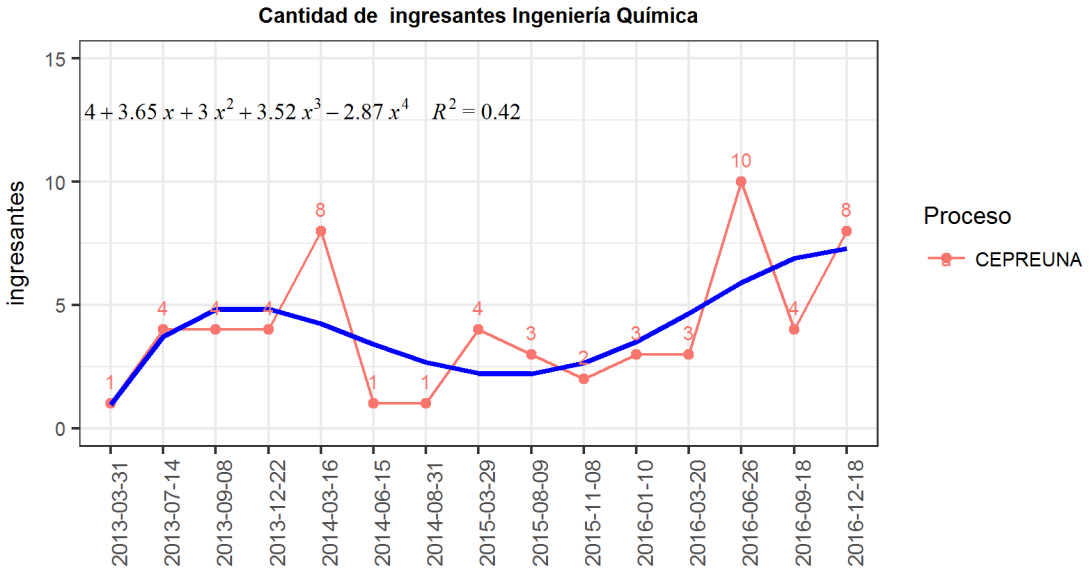


Figura 57. Modelo y resultados de la cantidad de ingresantes de Ing. Química - cepreuna

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
    Min      1Q  Median      3Q     Max
-2.8316 -1.3054 -0.4374  0.7090  4.0943

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       4.0000     0.6272   6.378 8.06e-05 ***
poly(data$fecha, degree = grado)1   3.6539     2.4291   1.504    0.163
poly(data$fecha, degree = grado)2   3.0018     2.4291   1.236    0.245
poly(data$fecha, degree = grado)3   3.5244     2.4291   1.451    0.177
poly(data$fecha, degree = grado)4  -2.8658     2.4291  -1.180    0.265
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.429 on 10 degrees of freedom
Multiple R-squared:  0.4215,    Adjusted R-squared:  0.1901
F-statistic: 1.822 on 4 and 10 DF,  p-value: 0.2013
```

Del a figura 57 se observa que Residual Standard Error (RSE) se aleja 2.429 unidades del verdadero valor de la cantidad de ingresantes; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 42.15% de la variabilidad observada en la cantidad de ingresantes.

Modelo para las Escuelas públicas y privadas

Gran Unidad Escolar San Carlos

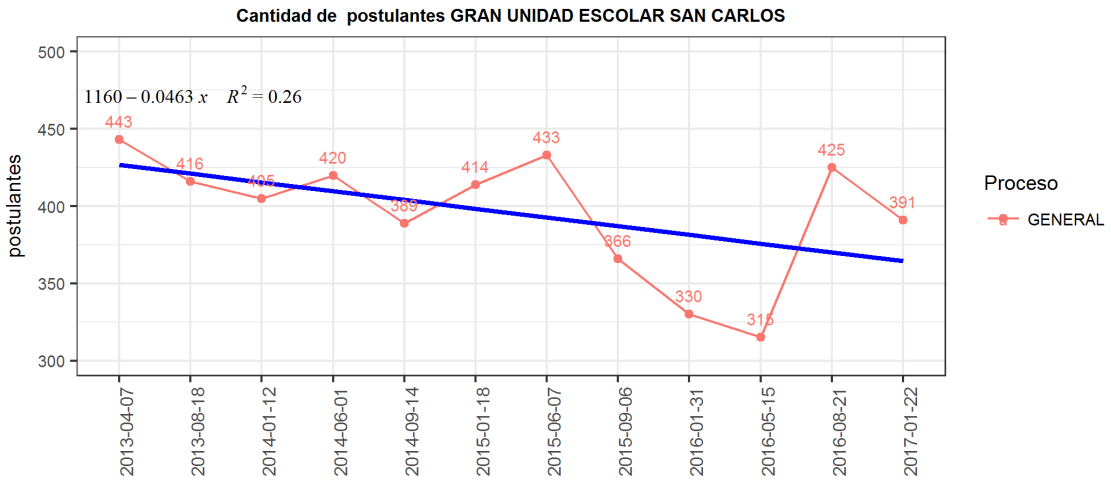


Figura 58. Modelo y resultados de los postulantes, escuela Gran Unidad Escolar San Carlos - general

```
Call:
lm(formula = data$total ~ data$fecha)

Residuals:
    Min      1Q  Median      3Q     Max
-60.750 -16.543   2.583  18.657  53.783

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1159.18087  402.28409   2.881   0.0163 *
data$fecha    -0.04626    0.02436  -1.899   0.0868 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36.06 on 10 degrees of freedom
Multiple R-squared:  0.265,     Adjusted R-squared:  0.1915
F-statistic: 3.605 on 1 and 10 DF,  p-value: 0.0868
```

De la figura 58 la información devuelta por el summary se observa el R-squared (R^2) empleado en el modelo es capaz de explicar el 26.5% de la variabilidad observada en la cantidad de postulantes.

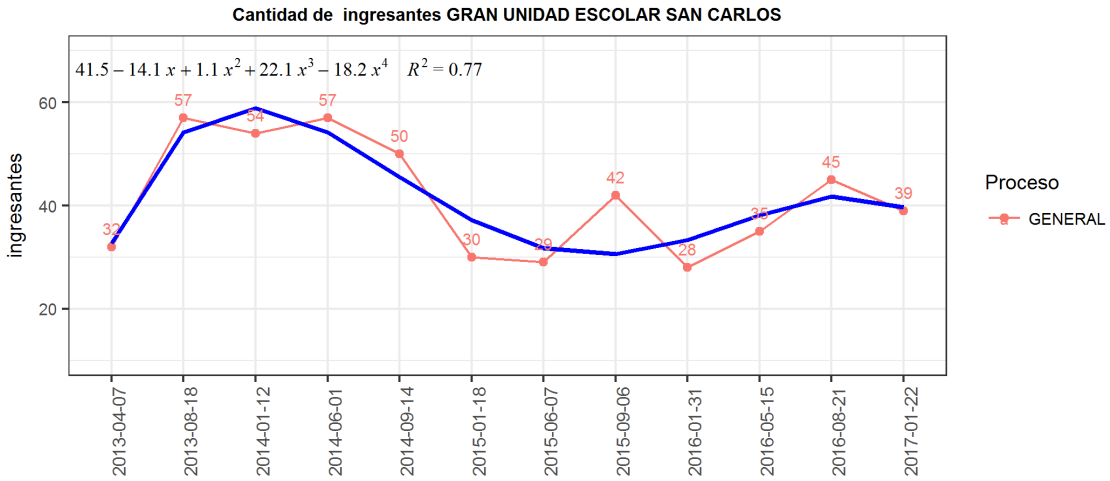


Figura 59. Modelo y resultados de los ingresantes, escuela Gran Unidad Escolar San Carlos - general

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
    Min      1Q  Median      3Q     Max
-7.4959 -3.6398 -0.5877  4.1499 10.7916

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       41.500      1.928  21.526 1.18e-07 ***
poly(data$fecha, degree = grado)1 -14.063      6.678  -2.106   0.0732 .
poly(data$fecha, degree = grado)2   1.104      6.678   0.165   0.8733
poly(data$fecha, degree = grado)3  22.104      6.678   3.310   0.0129 *
poly(data$fecha, degree = grado)4 -18.200      6.678  -2.725   0.0295 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.678 on 7 degrees of freedom
Multiple R-squared:  0.7654,    Adjusted R-squared:  0.6314
F-statistic: 5.711 on 4 and 7 DF,  p-value: 0.023
```

De la figura 59 la información devuelta por el summary se observa el R-squared (R^2) empleado en el modelo es capaz de explicar el 76.5% de la variabilidad observada en la cantidad de ingresantes.

ESCUELA NUESTRA SEÑORA DE ALTA GRACIA

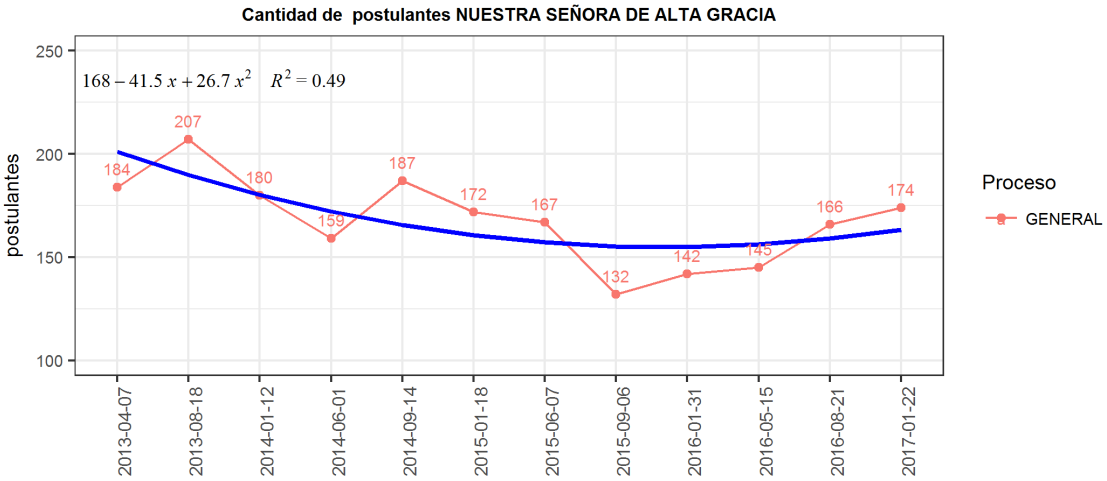


Figura 60. Modelo y resultados de los postulantes, escuela Nuestra Señora de Alta Gracia - general

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
   Min     1Q Median     3Q    Max
-24.18 -12.30   4.28  11.03  21.22

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      167.917      4.797  35.004 6.26e-11 ***
poly(data$fecha, degree = grado)1  -41.536     16.618  -2.499   0.0339 *
poly(data$fecha, degree = grado)2   26.653     16.618   1.604   0.1432
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.62 on 9 degrees of freedom
Multiple R-squared:  0.4949,    Adjusted R-squared:  0.3827
F-statistic:  4.41 on 2 and 9 DF,  p-value: 0.04624
```

De la figura 60 se observa que Residual Standard Error (RSE) se aleja 16.62 unidades del verdadero valor de la cantidad de postulantes de la escuela Nuestra Señora de Alta Gracia; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 49.49% de la variabilidad observada en la cantidad de postulantes.

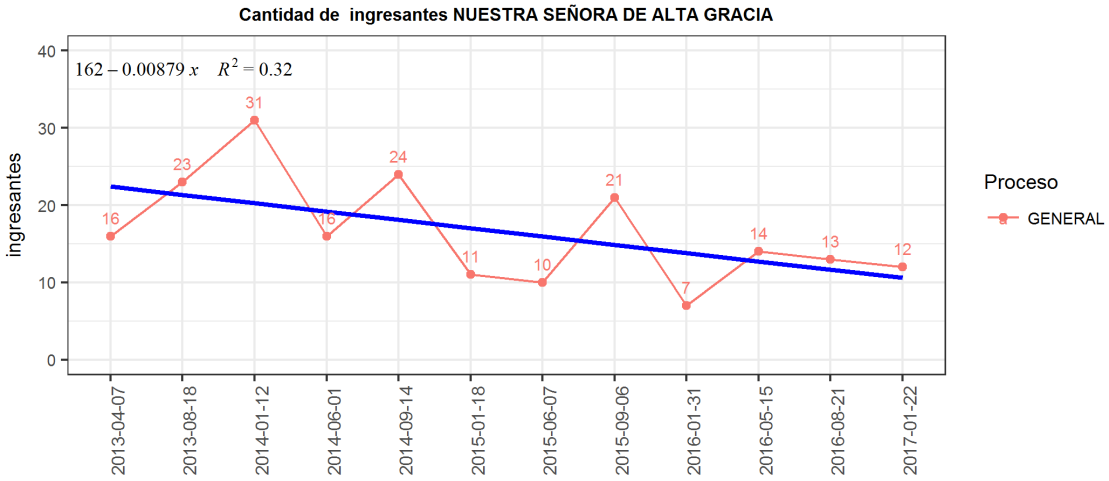


Figura 61. Modelo y resultados de los ingresantes, escuela Nuestra Señora de Alta Gracia - general

```
Call:
lm(formula = data$total ~ data$fecha)

Residuals:
   Min     1Q Median     3Q    Max
-6.701 -5.804  1.200  2.593 10.761

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 161.636940  67.094296   2.409   0.0367 *
data$fecha   -0.008792   0.004063  -2.164   0.0557 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.014 on 10 degrees of freedom
Multiple R-squared:  0.3189,    Adjusted R-squared:  0.2508
F-statistic: 4.682 on 1 and 10 DF,  p-value: 0.05574
```

De la figura 61 se observa que Residual Standard Error (RSE) se aleja 6.01 unidades del verdadero valor de la cantidad de ingresantes de la escuela Nuestra Señora de Alta Gracia; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 31.89% de la variabilidad observada en la cantidad de ingresantes.

ESCUELA NUESTRA SEÑORA DE LA MERCED

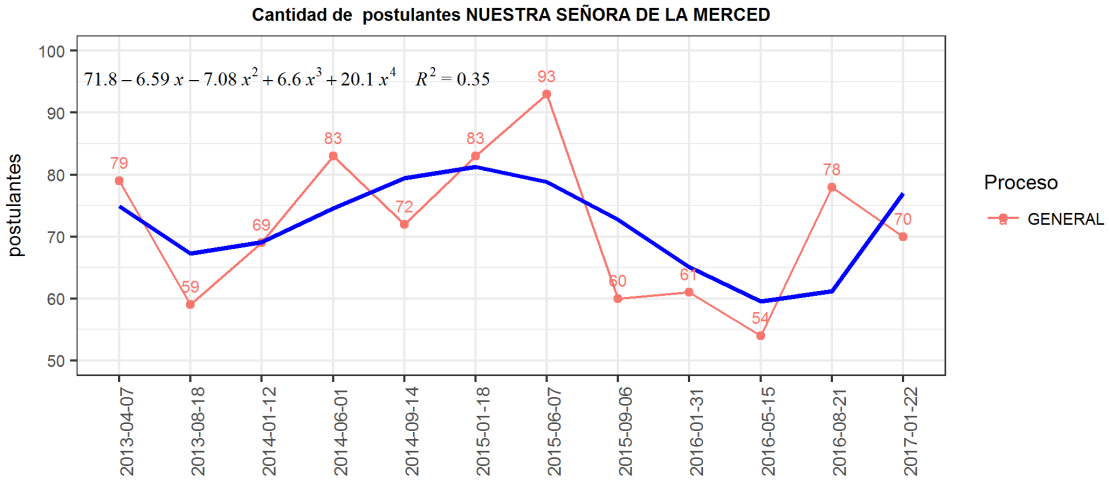


Figura 62. Modelo y resultados de los postulantes, escuela Nuestra Señora de la Merced - general

```
Call:
lm(formula = data$total ~ poly(data$fecha, degree = grado))

Residuals:
    Min      1Q  Median      3Q     Max
-14.178  -7.109  -2.131   4.590  17.253

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       71.750      3.448  20.811 1.49e-07 ***
poly(data$fecha, degree = grado)1  -6.592     11.943  -0.552    0.598
poly(data$fecha, degree = grado)2  -7.078     11.943  -0.593    0.572
poly(data$fecha, degree = grado)3   6.599     11.943   0.553    0.598
poly(data$fecha, degree = grado)4  20.066     11.943   1.680    0.137
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.94 on 7 degrees of freedom
Multiple R-squared:  0.3509,    Adjusted R-squared:  -0.02004
F-statistic: 0.946 on 4 and 7 DF,  p-value: 0.491
```

De la figura 62 se observa que Residual Standard Error (RSE) se aleja 11.94 unidades del verdadero valor de la cantidad de postulantes de la escuela Nuestra Señora de la Merced; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 35.09% de la variabilidad observada en la cantidad de postulantes.

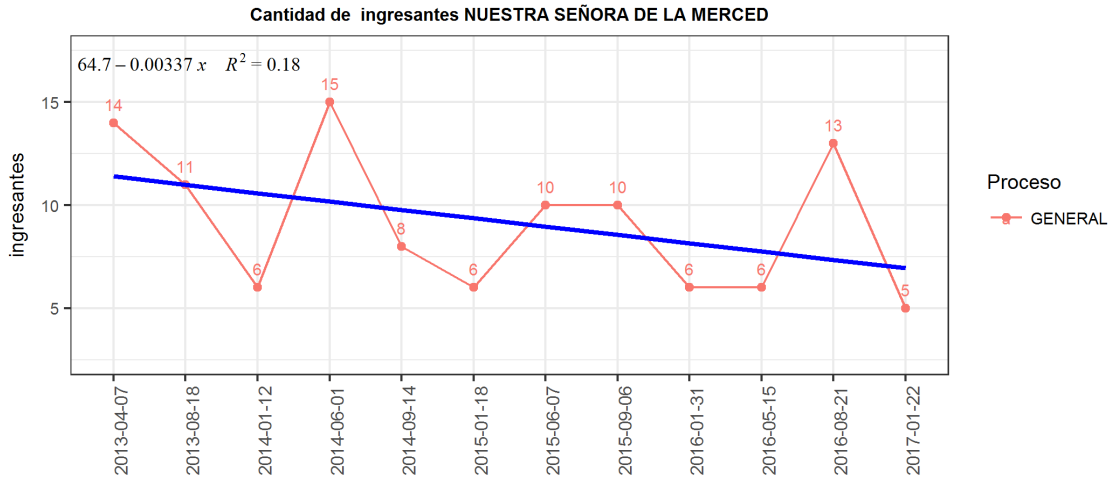


Figura 63. Modelo y resultados de los ingresantes, escuela Nuestra Señora de la Merced - general

```
Call:
lm(formula = data$total ~ data$fecha)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5982 -1.9254 -0.9082  1.6861  5.6066

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 64.736998  37.143961   1.743    0.112
data$fecha  -0.003366   0.002249  -1.497    0.165

Residual standard error: 3.329 on 10 degrees of freedom
Multiple R-squared:  0.183,     Adjusted R-squared:  0.1013
F-statistic:  2.24 on 1 and 10 DF,  p-value: 0.1654
```

De la figura 63 se observa que Residual Standard Error (RSE) se aleja 3.32 unidades del verdadero valor de la cantidad de postulantes de la escuela Nuestra Señora de la Merced; además, el R-squared (R^2) indica que el predictor es capaza de explicar el 18.3% de la variabilidad observada en la cantidad de postulantes.

4.4. EVALUACIÓN

En esta fase se evaluó los modelos creados con predicciones para predecir la tendencia de la información, para ello se usó la siguiente función predict del software R con un 95% de confianza.

```
predict(object = modelo,
  newdata = data.frame(fecha = c(as.numeric(as.Date("2013/04/07")),
                                 as.numeric(as.Date("2013/08/18")),
                                 as.numeric(as.Date("2014/01/12")),
                                 as.numeric(as.Date("2014/06/01")),
                                 as.numeric(as.Date("2014/09/14")),
                                 as.numeric(as.Date("2015/01/18")),
                                 as.numeric(as.Date("2015/06/07")),
                                 as.numeric(as.Date("2015/09/06")),
                                 as.numeric(as.Date("2016/01/31")),
                                 as.numeric(as.Date("2016/05/15")),
                                 as.numeric(as.Date("2016/08/21")),
                                 as.numeric(as.Date("2017/01/11"))
       )),
       interval = "prediction", level = 0.95)
```

Éste modelo se probó con la información de la escuela de educación secundaria NUESTRA SEÑORA DE LA MERCED obteniendo los siguientes resultados predictivos.

```
        fit      lwr      upr
1  75.55448 36.58764 114.52133
2  66.25129 32.76515  99.73744
3  68.53877 35.03793 102.03960
4  74.97721 43.07362 106.88080
5  79.11096 47.48467 110.73725
6  81.06178 48.98926 113.13431
7  78.35362 46.41795 110.28928
8  74.17788 42.62114 105.73462
9  65.72371 33.75707  97.69036
10 61.10868 28.23316  93.98419
11 60.74676 27.73368  93.75984
12 75.39486 36.12693 114.66278
```

Se puede observar que la función predict muestra la cantidad de postulantes muy cercanos a la cantidad de postulantes de los doce procesos de admisión. La columna fit se puede comparar con los datos de la figura 62, para el primer proceso realizado el 7 de abril del 2013 se tiene 79 postulantes, y el modelo muestra 75.55 postulantes; para el procesos 18 de agosto del 2013 se tiene 59 postulantes, y el modelo muestra 66.25 postulantes; etc. aquí se puede apreciar que se cumple el error estándar residual que esta 11.9 unidades por debajo y encima de la información real.

Cálculo de las predicciones de las Escuelas Profesionales para tres procesos de admisión del examen cepreuna. Esto se aprecia en la fila 17, 18 y 19.

Cuadro 30. Predicciones de las Escuelas Profesionales - cepreuna

		Med. Vet. Zoo.		Enfermería		Biología		Med. Humana		Nut. Humana		Odontología		Contables Ciencias		Trabajo Social		Sociología		Educ.Primaria		Educ.Inicial	
	fecha	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing
1	3/31/2013	32	11	143	13	34	17	268	5	29	7	111	6	436	14	63	12	52	12	13	8	20	8
2	7/14/2013	23	13	91	19	46	19	183	5	31	11	76	4	294	10	76	27	49	10	15	10	29	8
3	9/8/2013	7	7	36	20	11	6	142	5	13	10	37	4	135	12	43	29	11	9	4	4	5	5
4	12/22/2013	23	12	74	14	23	11	195	7	25	12	50	6	240	13	79	23	29	10	9	6	14	6
5	3/16/2014	41	12	134	16	37	13	309	7	63	12	125	7	68	12	33	23	11	10	8	8	7	6
6	6/15/2014	25	19	78	19	26	17	197	5	36	12	63	5	251	13	76	28	28	8	15	10	6	5
7	8/31/2014	14	3	44	5	23	8	199	5	27	7	47	5	220	14	53	13	28	10	9	4	36	9
8	12/14/2014	18	6	54	16	32	12	222	4	24	10	51	3	238	10	66	24	18	9	5	1	23	12
9	3/29/2015	31	9	125	17	35	12	302	7	62	10	88	6	428	12	101	24	29	10	7	5	43	6
10	8/9/2015	28	6	87	18	24	8	226	4	45	10	68	4	366	21	94	21	32	10	16	9	46	10
11	11/8/2015	20	8	56	8	18	5	176	5	33	10	34	4	233	10	55	9	40	9	4	1	30	9
12	1/10/2016	14	8	39	8	8	5	153	4	24	8	34	4	187	9	34	10	27	10	5	5	31	10
13	3/20/2016	25	10	97	10	26	9	286	4	45	9	79	5	339	10	48	9	32	8	11	10	65	10
14	6/26/2016	20	14	114	18	32	8	197	4	46	8	50	5	365	20	92	18	26	7	20	9	72	9
15	9/18/2016	20	9	90	10	31	9	166	4	33	10	41	6	247	10	79	9	52	9	22	9	31	9
16	12/18/2016	22	8	100	10	31	8	212	4	52	8	59	7	295	10	103	9	70	10	20	10	69	10
17	***12/3/2017***	***20***	***8***	***82***	***9***	***24***	***4***	***203***	***4***	***49***	***8***	***38***	***5***	***308***	***12***	***84***	***7***	***45***	***8***	***17***	***8***	***71***	***11***
18	***6/25/2017***	***21***	***8***	***90***	***9***	***25***	***5***	***210***	***4***	***47***	***8***	***46***	***5***	***325***	***13***	***83***	***7***	***46***	***9***	***17***	***8***	***66***	***11***
19	***9/24/2017***	***21***	***8***	***93***	***9***	***27***	***6***	***207***	***3***	***48***	***8***	***45***	***5***	***343***	***12***	***86***	***5***	***51***	***9***	***18***	***8***	***72***	***11***

		Ing. de Minas		Ing. Económica		Ing. Agronómica		Arte		Administración		Ciencias de la Comunicación Social		Turismo		Derecho		Antropología		Educ.Sec. Ciencias Sociales		Filosof. Educ. Sec. Lengua, Lit. Psicol. y		Educ.Física	
	fecha	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing
	3/31/2013	11	9	12	6	8	4	21	9	78	9	25	10	305	8	191	7	21	9	8	4	12	6	11	9
	7/14/2013	7	7	12	9	12	9	42	24	61	15	34	9	216	10	140	10	23	17	3	9	5	3	12	7
	9/8/2013	5	5	5	3	3	3	23	17	31	15	19	10	77	9	60	9	52	31	8	5	14	9	15	5
	12/22/2013	5	5	14	9	8	5	52	31	71	13	38	15	162	8	109	14	26	25	6	5	9	7	16	5
	3/16/2014	4	4	9	7	6	5	26	25	15	12	17	13	37	10	44	13	58	35	8	5	16	10	19	10
	6/15/2014	3	3	9	7	8	7	58	35	40	15	16	10	186	9	144	10	39	13	4	1	10	6	14	9
	8/31/2014	5	2	10	6	4	1	39	13	46	13	15	6	167	11	126	10	46	26	5	4	11	5	17	10
	12/14/2014	9	5	11	5	5	4	46	26	36	9	26	13	166	10	144	8	46	26	5	4	11	5	16	9
	3/29/2015	9	5	18	8	7	4	42	21	76	13	42	13	313	10	251	14	42	21	7	4	12	6	17	10
	8/9/2015	11	6	14	10	7	7	35	20	49	10	25	9	229	10	199	7	35	20	7	7	12	8	16	9
	11/8/2015	10	7	13	10	7	5	27	9	41	10	30	9	174	9	116	8	27	9	7	5	12	8	16	9
	1/10/2016	5	5	8	8	7	7	18	9	29	10	8	7	127	11	87	9	18	9	7	7	13	9	17	10
	3/20/2016	15	9	12	8	9	6	23	8	60	10	20	8	249	10	240	5	23	8	3	6	13	8	19	9
	6/26/2016	11	9	30	9	10	10	21	19	49	10	17	5	269	10	148	8	21	19	10	10	15	10	20	10
	9/18/2016	14	10	11	9	16	10	36	9	56	10	22	9	221	9	136	9	36	9	10	10	15	10	19	10
	12/18/2016	19	10	22	9	10	8	58	10	67	9	44	8	247	10	145	8	58	10	8	8	11	7	20	9
	12/3/2017	**16**	**9**	**20**	**10**	**11**	**9**	**35**	**8**	**51**	**8**	**25**	**7**	**247**	**10**	**175**	**7**	**35**	**8**	**9**	**9**	**12**	**5**	**21**	**10**
	6/25/2017	**16**	**9**	**19**	**10**	**11**	**9**	**33**	**8**	**54**	**8**	**25**	**7**	**256**	**10**	**178**	**7**	**33**	**8**	**9**	**9**	**13**	**5**	**21**	**10**
	9/24/2017	**17**	**10**	**21**	**10**	**12**	**9**	**33**	**6**	**57**	**8**	**26**	**6**	**273**	**10**	**186**	**7**	**33**	**6**	**9**	**9**	**14**	**5**	**22**	**10**

Ciencias Físico Matemáticas		Arquitectura		Ing. Electrónica		Ing. Mec. Eléctrica		Ing. de Sistemas		Ing. Covil		Ing. Agrícola		Ing. Agroindustrial		Ing. Topográfica		Ing. Est. E Infor.		Ing. Química		Ing. Metalúrgica		Ing. Geológica		
ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	fecha
5	10	12	173	9	29	10	77	11	95	10	384	9	42	8	23	9	50	7	8	1	1	8	13	9	96	3/31/2013
10	17	18	174	15	31	15	62	22	69	20	251	14	33	8	14	18	38	8	9	4	4	4	4	10	84	7/14/2013
9	14	19	53	3	4	11	22	11	21	19	110	13	20	3	7	14	28	3	3	4	4	4	6	10	34	9/8/2013
7	12	19	164	10	16	16	42	20	59	20	223	9	39	11	17	13	74	6	9	5	4	4	8	10	73	12/22/2013
8	17	20	266	10	30	15	77	17	101	18	373	9	53	11	21	15	80	5	5	1	2	8	10	17	93	3/16/2014
9	12	18	141	10	14	13	51	23	70	18	208	15	47	9	23	20	51	1	5	1	0	4	5	4	66	6/15/2014
2	8	16	137	0	9	1	51	2	61	19	206	0	30	0	11	1	44	0	2	1	0	1	2	4	53	8/31/2014
0	6	18	141	2	14	9	31	5	38	12	246	0	15	1	9	9	61	2	5	1	0	2	4	9	45	12/14/2014
2	6	20	244	6	26	10	63	18	103	20	364	16	51	0	13	15	89	3	4	3	2	4	9	2	70	3/29/2015
7	10	8	186	8	20	13	70	15	70	9	268	10	50	1	10	9	76	8	9	3	5	4	7	4	79	8/9/2015
7	8	9	102	10	12	4	28	14	50	9	200	9	18	2	4	10	39	2	3	2	3	3	8	9	50	11/8/2015
4	6	10	89	11	18	6	27	15	47	11	155	10	22	1	7	10	32	3	10	3	4	4	7	10	58	1/10/2016
7	10	10	185	13	23	6	42	16	82	7	333	10	22	6	14	9	47	8	10	3	7	4	11	7	96	3/20/2016
10	13	9	187	10	21	13	70	14	84	10	216	10	38	10	15	10	54	6	6	10	11	6	8	10	126	6/26/2016
5	5	8	154	14	29	5	40	15	84	10	206	7	37	10	10	10	39	8	8	4	6	6	10	9	92	9/18/2016
7	6	10	179	14	30	6	42	15	79	8	241	10	40	8	11	9	66		8	8	8	8	13	10	143	12/18/2016
5	**5**	**6**	**166**	**12**	**24**	**4**	**39**	**14**	**77**	**6**	**219**	**8**	**32**	**5**	**7**	**7**	**55**	**5**	**8**	**6**	**9**	**6**	**7**	**10**	**108**	**12/3/2017**
5	**5**	**6**	**167**	**12**	**24**	**5**	**43**	**14**	**80**	**6**	**238**	**8**	**33**	**5**	**9**	**7**	**54**	**6**	**8**	**6**	**8**	**6**	**12**	**10**	**108**	**6/25/2017**
5	**5**	**5**	**170**	**14**	**27**	**4**	**44**	**14**	**82**	**5**	**239**	**9**	**32**	**5**	**8**	**7**	**52**	**6**	**8**	**6**	**8**	**8**	**12**	**10**	**115**	**9/24/2017**

Cálculo de las predicciones de las Escuelas Profesionales para tres procesos de admisión del examen general. Esto se aprecia en la fila 17, 18 y 19.

Cuadro 31. Predicciones de las Escuelas Profesionales - general

		Med. Vet. Zoo.		Enfermería		Biología		Med. Humana		Nut. Humana		Odontología		Ciencias Contables		Trabajo Social		Sociología		Educ.Primaria		Educ.Inicial		Educ.Física		Filosof. Educ. Sec. Lengua, Lit. Psicol. y	
	fecha	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing
1	4/7/2013	110	21	377	23	111	14	326	4	119	13	169	5	1323	29	212	16	163	21	46	11	82	9	22	11	43	13
2	8/18/2013	132	8	306	25	64	17	413	17	99	9	124	9	1834	79	177	28	75	27	33	17	73	28	31	26	37	25
3	1/12/2014	83	16	304	36	95	25	467	19	101	24	185	17	1289	65	165	29	69	20	34	5	96	29	40	22	46	10
4	6/1/2014	113	27	272	37	72	21	439	19	123	25	142	8	1504	73	158	25	85	27	40	12	92	29	31	19	41	15
5	9/14/2014	105	30	271	27	80	28	382	6	108	22	121	5	1031	32	172	21	73	24	34	18	113	37	44	24	41	22
6	1/18/2015	108	3	300	21	96	13	373	6	119	7	120	7	1158	36	134	2	105	4	28	3	146	4	38	0	56	3
7	6/7/2015	100	1	274	9	66	3	438	8	111	5	148	7	1211	37	137	11	85	9	34	2	116	8	49	3	32	4
8	9/6/2015	81	33	221	31	62	18	386	7	98	25	84	6	1056	36	146	25	78	35	30	18	95	27	34	13	37	22
9	1/31/2016	86	9	270	19	69	17	400	6	130	12	132	8	1035	30	199	18	103	19	45	6	195	14	57	2	48	5
10	5/15/2016	73	6	237	22	59	11	366	7	102	11	101	6	868	39	132	19	85	15	37	5	161	19	44	5	44	10
11	8/21/2016	85	13	337	23	74	12	465	7	126	19	120	9	1266	39	197	17	133	24	42	14	177	21	60	19	38	25
12	1/22/2017	98	13	316	16	94	16	435	5	144	17	116	9	1007	28	243	19	137	20	60	19	207	20	87	17	67	18
17	***5/30/2017***	***79***	***10***	***265***	***18***	***68***	***12***	***427***	***4***	***127***	***15***	***100***	***7***	***898***	***27***	***181***	***16***	***108***	***18***	***45***	***11***	***203***	***18***	***70***	***8***	***50***	***14***
18	***8/27/2017***	***76***	***10***	***276***	***15***	***72***	***10***	***416***	***2***	***131***	***15***	***99***	***7***	***827***	***18***	***192***	***14***	***124***	***18***	***48***	***11***	***214***	***15***	***73***	***6***	***52***	***14***
19	***1/12/2018***	***76***	***8***	***280***	***13***	***68***	***8***	***415***	***0***	***134***	***13***	***90***	***6***	***840***	***17***	***202***	***14***	***131***	***19***	***51***	***13***	***228***	***14***	***78***	***6***	***53***	***16***

		Educ.Sec. Ciencias Sociales		Antropología		Derecho		Turismo		Ciencias de la Comunicación Social		Administración		Arte		Ing. Agronómica		Ing. Económica		Ing. de Minas		Ing. Geológica		Ing. Metalúrgica	
	fecha	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing
1	4/7/2013	27	12	60	13	542	15	213	15	110	15	829	17	44	17	60	20	613	21	468	14	310	13	54	17
2	8/18/2013	26	22	84	35	439	20	158	22	102	28	469	19	72	52	28	19	794	54	479	23	209	19	65	37
3	1/12/2014	15	5	125	31	571	34	131	19	81	19	507	20	62	45	38	23	699	44	424	23	178	17	52	28
4	6/1/2014	18	10	78	36	532	24	130	21	81	29	498	19	42	40	31	23	757	54	407	26	193	19	42	21
5	9/14/2014	28	13	63	22	606	19	150	25	73	18	587	19	60	37	42	26	731	37	379	19	206	18	35	22
6	1/18/2015	28	0	49	3	626	18	190	12	78	14	710	25	57	3	64	5	682	33	362	22	194	16	45	7
7	6/7/2015	35	3	61	2	614	16	141	14	62	8	738	26	39	2	40	1	710	30	341	23	179	13	40	1
8	9/6/2015	32	16	51	29	530	16	133	22	58	22	593	20	40	15	31	15	619	35	260	21	208	17	39	18
9	1/31/2016	29	6	57	15	563	19	176	20	75	14	703	18	70	8	57	3	669	25	265	11	278	14	65	2
10	5/15/2016	24	0	50	11	418	15	116	18	55	18	560	23	39	6	44	1	637	32	228	24	222	16	20	2
11	8/21/2016	34	22	90	19	587	18	160	18	81	18	750	25	51	15	37	15	875	35	314	20	267	18	34	11
12	1/22/2017	30	12	65	17	694	19	178	18	89	18	783	19	60	18	37	6	764	28	292	17	272	15	37	9
17	***5/30/2017***	***33***	***9***	***57***	***12***	***604***	***16***	***148***	***18***	***62***	***15***	***721***	***23***	***50***	***3***	***41***	***2***	***739***	***28***	***217***	***19***	***243***	***16***	***31***	***0***
18	***8/27/2017***	***34***	***9***	***51***	***8***	***613***	***14***	***157***	***17***	***64***	***14***	***784***	***23***	***48***	***0***	***45***	***0***	***726***	***23***	***201***	***17***	***267***	***15***	***29***	***0***
19	***1/12/2018***	***36***	***11***	***48***	***8***	***603***	***13***	***162***	***17***	***66***	***14***	***803***	***23***	***50***	***0***	***43***	***0***	***743***	***22***	***186***	***17***	***281***	***15***	***30***	***0***

	Ciencias Físico Matemáticas		Arquitectura		Ing. Electrónica		Ing. Mec. Eléctrica		Ing. de Sistemas		Ing. Civil		Ing. Agrícola		Ing. Agroindustrial		Ing. Topográfica		Ing. Est. E Infor.		Ing. Química		
	fecha	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post
1	4/7/2013	4	24	20	475	3	84	16	253	16	237	24	963	10	140	3	77	19	222	7	28	3	9
2	8/18/2013	15	50	27	392	17	59	14	146	28	192	26	819	45	195	13	57	28	153	22	32	24	31
3	1/12/2014	15	40	23	455	12	79	18	173	29	216	29	784	36	209	13	60	29	152	7	14	18	21
4	6/1/2014	14	36	25	453	24	63	16	180	29	188	29	788	51	195	16	59	27	152	19	25	12	19
5	9/14/2014	11	22	25	510	27	78	12	132	22	211	16	722	40	150	16	58	37	194	16	19	25	33
6	1/18/2015	8	33	26	558	11	86	18	158	29	260	16	722	21	185	5	67	27	242	4	30	5	30
7	6/7/2015	3	31	17	470	6	62	7	158	17	249	19	752	6	128	4	50	9	153	3	33	2	32
8	9/6/2015	14	25	20	429	24	68	21	133	22	232	18	661	36	101	20	41	32	145	6	16	14	20
9	1/31/2016	2	25	18	448	5	79	11	185	13	276	20	688	7	89	2	41	7	159	2	23	3	32
10	5/15/2016	5	22	18	351	4	51	16	118	22	188	16	534	13	88	1	32	7	98	1	27	6	29
11	8/21/2016	15	26	19	556	12	82	16	164	22	308	14	744	19	107	13	36	19	154	11	23	23	44
12	1/22/2017	14	24	17	510	13	93	14	170	12	274	16	813	18	121	13	36	19	174	15	34	15	34
17	***5/30/2017***	***10***	***21***	***17***	***485***	***11***	***77***	***14***	***138***	***16***	***275***	***13***	***640***	***13***	***84***	***9***	***29***	***13***	***141***	***5***	***27***	***11***	***39***
18	***8/27/2017***	***8***	***17***	***15***	***492***	***8***	***80***	***14***	***150***	***13***	***287***	***12***	***653***	***6***	***66***	***7***	***27***	***10***	***146***	***4***	***28***	***9***	***38***
19	***1/12/2018***	***8***	***17***	***14***	***483***	***7***	***79***	***14***	***146***	***12***	***295***	***10***	***648***	***2***	***52***	***7***	***22***	***7***	***139***	***5***	***30***	***10***	***41***

Cálculo de las predicciones de escuelas públicas del examen general. Esto se aprecia en la fila 13, 14 y 15

Cuadro 32. Predicciones de colegios públicos – general

		GUE San Carlos - Puno		José Antonio Encinas - San Román		Glorioso San Carlos - Puno		Santa Rosa - Puno		Nuestra Señora del Carmen - EL Collao		María Auxiliadora - Puno		45 Emilio Romero Padilla - Puno		Nuestra Señora de Alta Gracia - Melgar		Las Mercedes - San Román		Mariano Melgar - Melgar		Politecnico Regional Los andes - San Román	
	fecha	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing
1	07/04/2013	443	32	51	2	332	29	292	29	266	17	318	21	254	10	184	16	122	8	167	14	128	6
2	18/08/2013	416	57	43	7	358	43	255	50	267	32	263	36	252	32	207	23	133	9	172	26	123	7
3	12/01/2014	405	54	57	3	317	46	268	40	236	26	236	28	227	24	180	31	141	6	138	19	135	12
4	01/06/2014	420	57	43	6	306	31	291	72	195	21	227	29	213	33	159	16	138	7	184	29	106	8
5	14/09/2014	389	50	41	2	311	49	245	32	277	16	228	27	202	24	187	24	133	15	122	17	106	4
6	18/01/2015	414	30	46	0	316	34	281	26	206	8	207	12	205	7	172	11	150	3	135	16	123	8
7	07/06/2015	433	29	45	0	272	13	292	31	232	8	221	11	210	8	167	10	133	4	117	8	103	4
8	06/09/2015	366	42	36	1	230	31	213	33	201	16	184	29	183	25	132	21	109	13	105	14	95	14
9	31/01/2016	330	28	51	1	233	15	241	31	215	7	153	9	182	8	142	7	153	3	110	12	138	4
10	15/05/2016	315	35	48	6	184	21	203	26	167	7	168	12	143	18	145	14	91	5	95	10	101	6
11	21/08/2016	425	45	46	1	268	27	271	37	270	12	195	8	209	15	166	13	121	6	110	16	119	6
12	22/01/2017	391	39	55	4	234	16	248	38	218	20	195	10	185	12	174	12	158	10	110	9	140	7
13	***30/05/2017***	***358***	***34***	***48***	***2***	***201***	***15***	***233***	***30***	***203***	***8***	***148***	***6***	***161***	***11***	***145***	***9***	***131***	***6***	***87***	***9***	***117***	***6***
14	***27/08/2017***	***357***	***30***	***49***	***1***	***188***	***11***	***233***	***26***	***203***	***6***	***150***	***2***	***158***	***8***	***142***	***7***	***128***	***6***	***81***	***6***	***119***	***6***
15	***12/01/2018***	***351***	***28***	***49***	***2***	***177***	***8***	***225***	***25***	***205***	***6***	***144***	***1***	***155***	***7***	***143***	***5***	***127***	***6***	***76***	***5***	***121***	***5***

		Maria Asunción Galindo - Chucuito		Pedro Vilccapaza - San Román		Sec. Telesforo Catacora Juli - Chucuito		32 Industrial - Puno		Independencia Nacional - Puno		José Carlos Mariategui - El collao		Perú Birf - San Román		28 Perú Birf - Azángaro		Pedro Vilcapaza Alarcon - Azángaro		I.E.S. Jose galvez - Yunguyo		Comercio 32 Mariano H. Cornejo - San Román			
	fecha	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing		
1	07/04/2013	82	4	78	3	76	1	107	7	147	13	189	8	119	7	128	12	134	4	136	7	150	7		
2	18/08/2013	96	11	41	3	86	15	109	8	122	13	156	17	95	11	111	16	116	8	102	12	121	16		
3	12/01/2014	77	13	89	6	65	3	90	8	101	15	156	16	97	11	74	10	104	13	107	19	134	11		
4	01/06/2014	89	6	70	4	85	11	80	9	97	15	171	16	110	15	117	13	128	12	115	16	109	10		
5	14/09/2014	67	8	62	1	103	10	110	10	81	12	159	17	79	8	83	8	106	11	100	6	99	9		
6	18/01/2015	69	2	76	2	68	4	108	7	107	1	172	7	136	7	98	3	106	7	110	6	125	4		
7	07/06/2015	79	3	70	3	98	1	102	2	98	1	158	3	100	3	111	2	115	7	130	3	107	2		
8	06/09/2015	54	9	69	7	64	7	73	13	74	10	132	13	71	7	97	16	113	3	105	13	101	13		
9	31/01/2016	70	5	96	0	88	5	79	2	73	6	135	6	103	4	132	3	128	4	113	4	120	4		
10	15/05/2016	66	5	69	1	80	9	56	0	80	6	122	3	53	4	74	5	83	5	89	4	93	2		
11	21/08/2016	71	6	84	5	98	9	90	4	97	8	184	9	93	6	97	4	111	6	140	14	139	7		
12	22/01/2017	67	7	99	6	90	4	96	4	90	6	141	9	108	9	137	10	111	11	110	10	141	8		
13	***30/05/2017***	***60***	***4***	***90***	***4***	***90***	***6***	***77***	***3***	***70***	***3***	***138***	***5***	***85***	***5***	***107***	***4***	***103***	***6***	***112***	***7***	***115***	***4***		
14	***27/08/2017***	***58***	***4***	***94***	***4***	***91***	***5***	***76***	***2***	***73***	***3***	***139***	***3***	***86***	***4***	***119***	***4***	***105***	***5***	***116***	***6***	***120***	***3***		
15	***12/01/2018***	***58***	***4***	***93***	***4***	***93***	***6***	***75***	***1***	***73***	***2***	***135***	***2***	***83***	***3***	***119***	***4***	***104***	***4***	***116***	***5***	***122***	***3***		

Cálculo de las predicciones de escuelas privadas del examen cepreuna. Esto se aprecia en la fila 17, 18 y 18.

Cuadro 33. Predicciones de colegios privados – cepreuna

		San Román, Puno Claudio Galeno –		Puno Merced – Nuestra Sra. De la		Puno San Ignacio de Loyola –		Puno Divino Maestro –		San Román James Baldwin –		Puno Adventista Puno –		San Román Tupac Amaru –		San Román Santa Catalina –		- Puno Fátima Parroquial Villa		- San Román Franciscano San Román		IEP Cramer - Puno		Puno Inmaculada – Parroquial La		San Román Gregor Mendel –	
	Fecha	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing	post	ing
1	31/03/2013	59	3	42	8	9	0	4	1	14	0	21	1	41	5	18	1	18	1	22	2	15	4	11	1	14	1
2	14/07/2013	54	8	23	3	13	2	10	0	11	3	14	0	27	7	13	2	13	3	14	0	12	3	16	1	8	0
3	08/09/2013	30	10	16	7	5	2	8	0	5	2	11	5	12	5	3	0	6	2	5	0	3	1	9	1	1	0
4	22/12/2013	36	3	26	5	30	4	9	1	6	1	11	1	19	5	5	0	11	4	10	1	3	1	13	5	5	2
5	16/03/2014	53	12	36	9	22	4	15	1	12	4	26	6	33	6	17	2	17	4	16	0	16	4	10	1	5	1
6	15/06/2014	33	7	27	3	14	2	10	0	12	2	17	3	26	6	7	1	5	0	8	0	2	0	19	5	6	0
7	31/08/2014	46	2	22	1	7	2	11	3	14	1	15	3	20	3	13	2	9	0	6	0	4	1	9	0	8	1
8	14/12/2014	33	1	26	1	20	2	13	2	12	2	13	1	22	4	10	1	11	3	12	1	9	0	9	0	10	3
9	29/03/2015	59	7	38	3	21	1	23	5	21	0	37	3	37	0	27	2	21	1	18	2	14	3	10	0	10	0
10	09/08/2015	56	2	28	5	23	1	20	2	23	3	13	1	24	3	8	0	9	1	10	1	9	0	9	1	12	1
11	08/11/2015	38	7	12	0	19	2	9	1	15	3	7	0	13	5	6	0	7	0	8	0	4	0	6	0	6	2
12	10/01/2016	35	6	21	0	12	1	6	2	7	1	9	1	12	1	8	1	6	1	7	1	6	1	4	1	8	2
13	20/03/2016	85	9	30	1	29	3	25	2	18	1	22	4	27	1	18	1	17	1	15	1	13	4	8	1	11	6
14	26/06/2016	57	5	29	3	10	2	16	1	27	3	13	1	32	8	12	1	15	1	8	1	5	0	6	2	7	0
15	18/09/2016	53	9	31	2	18	2	9	0	16	4	13	2	30	6	6	2	6	0	5	0	5	1	7	1	5	0
16	18/12/2016	76	8	30	2	21	4	14	2	14	2	11	1	19	4	6	0	12	1	9	1	9	1	8	2	15	3
17	***18/03/2017***	***62***	***7***	***27***	***1***	***21***	***2***	***17***	***2***	***20***	***3***	***13***	***1***	***22***	***3***	***10***	***1***	***10***	***0***	***7***	***1***	***7***	***1***	***5***	***1***	***10***	***2***
18	***25/06/2017***	***66***	***6***	***29***	***1***	***21***	***2***	***17***	***2***	***21***	***2***	***13***	***1***	***24***	***3***	***10***	***1***	***11***	***2***	***8***	***1***	***7***	***1***	***4***	***1***	***11***	***3***
19	***21/09/2018***	***79***	***7***	***30***	***0***	***22***	***2***	***18***	***2***	***25***	***3***	***12***	***0***	***25***	***3***	***10***	***1***	***12***	***1***	***8***	***1***	***8***	***1***	***2***	***0***	***14***	***3***

CONCLUSIONES

- Para recabar información de postulantes se debe contar con la base de datos MySQL libre de errores ya que esto permitió obtener todas las variables y observaciones en data.frames, tal como indica (Kadlec, 2015), aptos para el análisis de datos, además sólo deben ser consideradas todas aquellas variables a ser analizadas.
- La metodología CRISP–DM, según (Zhao, 2013), es una guía práctica que permitió analizar la información de postulantes de las diferentes Escuelas Profesionales y escuelas de educación secundaria, además, se obtuvo los primeros datos estadísticos con su representación visual para finalmente llegar a establecer los modelos para las Escuelas Profesionales y escuelas de educación secundaria.
- Los paquetes dplyr, ggplot creado por Hadley Wickham, permitieron obtener las agrupaciones de postulantes e ingresantes y adicionar nuevas variables a ser analizadas como parte de la información.
- Se puede observar que el patrón de crecimiento de las escuelas educación inicial, entre otras tiene una pendiente positiva, ésta información del comportamiento de las cantidades de postulantes e ingresantes en la exploración de datos permite identificar claramente que se debe revertir esta

situación definiendo políticas de captación de postulantes e ingresantes.

- Para el tratamiento de la información de los colegios de educación secundaria se puede apreciar que no siempre las escuelas de educación secundaria pública con mayor cantidad de postulantes tienen la mayor cantidad de ingresantes, esto se puede observar que la escuela de educación secundaria Santa Rosa tiene la mayor cantidad de ingresantes en un 15.17% frente al 12.1% de la Gran Unidad Escolar San Carlos, de la misma forma, la escuela privada IEP Cramer cuenta con 18.44% de ingresantes frente a un 12.36% de Claudio Galeno.

RECOMENDACIONES

- Se recomienda el uso de la metodología CRISP-DM ya que provee una serie de fases que permiten identificar claramente los objetivos del proyecto esto para garantizar que los resultados finales del análisis sean técnicamente válidos para tomar decisiones.
- Se recomiendo usar el software estadístico R ya que ofrece un gran aporte al análisis de datos llevando una clara ventaja frente a los programas de uso comercial, como SAS, SPSS y Excel; sin embargo, existe una limitante para aquellos profesionales que no tienen conocimientos sobre éste lenguaje de programación por lo que su aprendizaje se torna lento al inicio y fluido con el tiempo.
- Se recomienda usar el paquete RMySQL como interfaz entre la base de datos y el software estadístico R para evitar realizar consultas SQL innecesarias e integrado con el paquete dplyr con sus verbos summarise, arrange, filter, select, unique, group_by, etc. se hace fácil la manipulación de las observaciones y variables devueltas en un data.frame para posteriormente realizar la graficación e interpretación con el paquete ggplot.
- Se recomienda a las Escuelas Profesionales que tienen menos postulantes

tomar las políticas adecuadas para revertir esta situación, ya que se debe cumplir con los estándares para que la escuela esté debidamente acreditada y si no cuenta con postulantes y menos ingresantes no estaría cumpliendo con la normatividad de formación profesional.

- Se recomienda a las escuelas de educación secundaria poner mayor énfasis en la enseñanza, puesto que los resultados indican que la cantidad de ingresantes está disminuyendo, tanto en públicas como privadas. Se puede observar que la escuela privada Alexander Fleming ha obtenido 29 ingresantes en 32 procesos de admisión, esto también puede deberse a que no estén postulando a la Universidad o estén mal formados.

BIBLIOGRAFÍA

Arias, F. G. (2004). *El proyecto de investigación, Introducción a la metodología científica.* C.A.: Episteme.

Blanco Perez, F., Flores Paredes, E., & Giménez Mercado, C. (2010). La equidad y la calidad en los procesos. *Revista de pedagogía*, 276.

Borges de Barros Pereira, H. (15 de abril de 2002). *TDR Tesis Doctorales en Red.* Recuperado el 22 de diciembre de 2014, de http://www.tdx.cbuc.es/handle/10803/6542;jsessionid=9DF2DA9FA32C0A6AF2C25DD6E04FC253.tdx1

Chapman, P., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & otros, y. (2000). *CRISP-DM 1.0 Step-by-step dataminig guide.*

Cortes Flores, A., y Level, J. P. (2008). El proceso de admisión como predictor del rendimiento académico en la educación superior. *PEPSIC*.

Dalton P. y Whitehead, P. (2000). *Sql Server 2000.* España.

De Miguel Castaño, A. (2012). *Diseño de base de datos relacionales.* México.

Diaz Lopez, J. (2015). *Análisis de las líneas de autobuses urbanos de A Coruña.* España.

E. Prieto, M. (2010). *Recursos digitales para la educación y la cultura.* Mexico: Merida.

Espino Timon, C. (16 de enero de 2017). *Universitat Oberta de Catalunya*. Obtenido de http://openaccess.uoc.edu

Iturbide Dominguez, G. (2013). *Metodología de Preparación de Datos Orientada a Aplicaciones de Epidemiología Basada en el Modelo CRISP-DM.* México: Morelos.

James, D. (12 de agosto de 2016). *R interface to the MySQL database.* Obtenido de CRAN.

Jhon, H., y Reitsch, A. (1996). *Pronósticos en los negocios.* México: McGrawHill.

Kadlec, J. (2015). WaterML R package for managing ecological experiment data on a CUAHSI HydroServer. *ELSEVIER*, 10.

Martín, F. (5 de octubre de 2015). *E&N.* Obtenido de http://www.estrategiaynegocios.net/opinion/851122-345/inteligencia-de-negocios-predictiva

Ocoña Riola, R. (2 de marzo de 2017). *Escuela Andaluza de Salud Pública.* Obtenido de Escuela Andaluza de Salud Pública: http://www.easp.es/project/descubriendo-r-commander/

Pautasio, L. (16 de julio de 2014). *Mercado.* Obtenido de http://www.mercado.com.ar/notas/informes/8015960/big-data-evolucin-o-revolucin

Pereira Gonzalez, A. (2010). *Análisis predictivo de datos mediante técnicas de regresión estadística.* España.

Pressman, R. (1998). *Ingeniería de Softaware un enfoque práctico.* España: McGrawhill.

Rodriguez Fuentes, R., Gallestey, J. B., Rodriguez, P. D., & Lazo, M. M. (2008).

Valor predictivo de algunos criterios de selección para el ingreso a la carrera de medicina. *Scielo*.

SAP. (2013). Predecir el futuro de los análisis predictivos. *LOUDHOUSE*, 12.

Velasco Cabo, J. (2017). Análisis exploratorio de datos del mercado eléctrico español con R. *Invurnus, en busca del conocimiento*, 12-17.

Velasquez Uribe, M. T. (2008). *Aprender demografía con R-project.* México.

Velasquez, J., Montoya, O., & Cataño, N. (2010). ¿Es el proyecto R para la computación estadística apropiado para la inteligencia computacional? *Ingeniería y Competitividad*, 81-94.

Velez, I. (19 de juni de 2014). *El coste de no hacer nada en el análisis predictivo de la información*. Obtenido de http://www.muycomputerpro.com/2014/06/19/coste-analisis-predictivo-informacion

Zhao, Y. (26 de abril de 2013). *R contributed Documentation*. Obtenido de R contributed Documentation: https://cloud.r-project.org/doc/contrib/Zhao_R_and_data_mining.pdf

ANEXOS

Anexo 1. Diccionario de datos de las tablas de la base de datos

Tabla c_carrera

Nombre:	C_CARRERA
Descripción:	Información de las Escuelas Profesionales
Clave primaria:	Id_carrera
Clave foránea:	
Índice:	Id_carrera

Campo	Tipo	Descripción
Id_carrera	Integer 2 not null	Código de Escuela Profesional
Nombre	varchar(40) not null	Nombre de la Escuela Profesional

Tabla c_modalidad

Nombre:	C_MODALIDAD
Descripción:	Información de las modalidades de ingreso
Clave primaria:	Id_modalidad
Clave foránea:	

Campo	Tipo	Descripción
Id_modalidad	Integer	Código de modalidad
nombre	varchar(20)	Nombre de la modalidad

Tabla c_modalidad_carrera

Nombre:	C_MODALIDAD_CARRERA
Descripción:	Información de las vacantes por modalidad
Clave primaria:	Id_modalidadcarrera
Clave foránea:	Id_modalidad, referencia C_MODALIDAD

Campo	Tipo	Descripción
Id_modalidadcarrera	Integer	Código de modalidad
Id_carrera	Integer	Código de la Escuela Profesional

Tabla c_usuario

Nombre:	C_ USUARIO
Descripción:	Información de los usuarios
Clave primaria:	codigo
Clave foránea:	

Campo	Tipo	Descripción
Codigo	varchar(7) not null	Código de la U.O.
Nombres	varchar(40) not null	Nombres del usuario
Paterno	varchar(20) not null	Apellido materno del usuario
Materno	varchar(20) not null	Apellido paterno del usuario
User	varchar(20) not null	Nombre de acceso al sistema
Passwd	varchar(32) not null	Clave de acceso

Tabla c_modalidad_carrera

Nombre:	C_USUARIO	
Descripción:	Información de las vacantes por modalidad	
Clave primaria:	Id_modalidadcarrera	
Clave foránea:	Id_modalidad, referencia C_MODALIDAD	
Campo	**Tipo**	**Descripción**
Id_modalidadcarrera	Integer	Código de modalidad
Id_carrera	Integer	Código de la Escuela Profesional

Tabla c_departamento

Nombre:	C_DEPARTAMENTO	
Descripción:	Información de los departamentos del Perú	
Clave primaria:	IdDepartamento	
Clave foránea:		
Campo	**Tipo**	**Descripción**
IdDepartamento	Char(2)	Código del departamento
Nombre	Varchar(20)	Nombre del departamento

Tabla c_provincia

Nombre:	C_PROVINCIA	
Descripción:	Información de las provincias del Perú	
Clave primaria:	IdProvincia	
Clave foránea:	IdDepartamento, tabla C_DEPARTAMENTO	
Campo	**Tipo**	**Descripción**
IdDepartamento	Char(2)	Código del departamento
Nombre	Varchar(20)	Nombre del departamento
idProvincia	Varchar(4)	Codigo de la provincia

Tabla c_distrito

Nombre:	C_PROVINCIA	
Descripción:	Información de las provincias del Perú	
Clave primaria:	IdDistrito	
Clave foránea:	IdProvincia, tabla C_PROVINCIA	
Campo	**Tipo**	**Descripción**
IdDistrito	Char(6)	Código del distrito
Nombre	Varchar(20)	Nombre del distrito
idProvincia	Varchar(4)	Codigo de la provincia

Tabla c_colegioperu

Nombre:	C_COLEGIOPERU
Descripción:	Información de los colegios del Perú
Clave primaria:	Codigo_original
Clave foránea:	

Campo	Tipo	Descripción
Codigo_original	Char(6)	Código del distrito
Nombre	Varchar(20)	Nombre del distrito
Departamento	Varchar(20)	Nombre del departamento del colegio
Provincia	Varchar(20)	Nombre de la provincia del colegio
Distrito	Varchar(20)	Nombre de la provincia del colegio
area	Varchar(6)	Área al que pertenece el colegio
Tipo_colegio	Integer	Es el tipo de colegio (estatal, particular)

Tabla c_postulante

Nombre:	C_POSTULANTE
Descripción:	Información de los postulantes
Clave primaria:	Dni
Clave foránea:	Id_carrera, tabla C_CARRERA, codigo, tabla C_USUARIOS, ubi_nac, TABLA C_DISTRITO, codigo_colegio, tabla C_COLEGIOPERU

Campo	Tipo	Descripción
Dni	Char(8)	Dni del postulante
Nombre	Varchar(20)	Nombre del distrito
Paterno	Varchar(20)	Apellido paterno del postulante
Materno	Varchar(20)	Apellido materno del postulante
Nombres	Varchar(20)	Nombres del postulante
Id_carrera1	Integer	Codigo de la escuela profesional
Cla_id	Varchar(6)	Clave del folder del postulante
Cla_pass	Varchar(6)	Clave del folder
Sexo	Char(1)	Sexo del postulante
Id_colegio	Varchar(6)	Codigo del colegio
Fecha_nac	Date	Fecha de nacimiento
Ubi_nac	Varchar(6)	Lugar de nacimiento del postulante
Fechavoucher	Varchar(10)	Fecha de pago del voucher
Voucher	Varchar(6)	Codigo del voucher
Cod_inscriptor	Integer	Codigo del inscriptor
Cla_obs	Varchar(6)	Alguna observacion acerca de la clave
Comentario	Varchar(100)	Comentario acerca de la modalidad
Id_tipo_colegio	Integer	Es es tipo de colegio
Fecha_insc_inscrip	Date	Es la fecha de inscripcion del inscriptor
Fecha_incs_post1	Date	Fecha de inicio de insc. postulante