**Maestría en Ciencia de Datos v.2**

**Proyecto Final de Análisis Estadistico I**

**“Estudio del padrón empresarial Boliviano”**

**Profesor:**

**Lic. Carlos Lopez**

**Módulo:**

**Análisis Estadistico I**

**Ivan Fernando Mujica Mamani**

**1. RESUMEN**

En el entormo mundial actual de los negocios y de la economía, tenemos acceso a grandes cantidad de información estadística y es transversal a diferentes ramas profesionales como ser la Contaduria, Finanzas, Marketing, Producción y Economía entre otros. En el presente documento se pretende aplicar técnicas de Estadística Descriptiva e Inferencial al padrón empresarial de Bolivia, misma serie temporal esta disponible 2005 – 2017 para el presente proyecto, como ser estadisticos descriptivos, analisis de varianza y regresión al número de empresas, primeramente se hará un proceso manual para hallar los estadisticos y se implementaran las mismas con el Lenguaje R.

**2. INTRODUCCIÓN**

En el marco de la asignatura Análisis Estadístico I, se toma como caso de estudio el padron empresarial boliviano, cabe recordar que actualmente la custodia y actualización del padrón se encuentra tercializado a FUNDEMPRESA, esta misma se encuentra bajo tuición del Ministerio de Desarrollo Productivo y Economía Plural .

FUNDEMPRESA es una fundación sin fines de lucro que, en el marco de un Contrato de Concesión suscrito con el estado Plurinacional, opera el Registro de Comercio en todo el país (Ley 2064, Ley 2196, Decreto Supremo 26215 y Decreto Supremo 26335) ([www.fundempresa.gob.bo](http://www.fundempresa.gob.bo/)). El nombre formal de esta función del estado y según normativa vigente es “Registro de Comercio de Bolivia” y esta sujeta tambien a la Autoridad de Fiscalización de empresas.

La función principal de FUNDEMPRESA es la actualización de la base de datos de empresas y la otorgación de la matrícula de comercio a las empresas, para que estas puedan operar ya sea en el sector publico o privado.

Para mayor claridad vamos a dividir el presente estudio en dos partes, en la primera describiremos los datos por gestión y presentar las medidas de tendencia y dispersion tradicionales, asi tambien las tablas de distribución de frecuencias. En la segunda parte aplicaremos un analisis de varianza en función al modelo de regresion obtenido.

**3. OBJETIVOS**

**3.1 OBJETIVO GENERAL**

El objetivo general del presente trabajo es aplicar técnicas Estadistica Descriptiva e Inferencial a la base de datos del padrón empresarial.

**3.2 OBJETIVO ESPECIFICO**

Aplicar Estadisticos de Dispersion y de Tendencia Central a la base de datos.

Hallar Intervalos de confianza para las medias, varianzas y proporciones para ambas poblaciones.

Contrastar los temas metodológicos estadisticos regresión y análisis de varianza aplicados a la base de datos padrón empresarial, cuyo número de empresas es analizado por tiempo tipo de empresa.

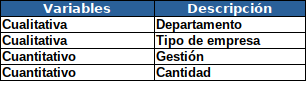
**DESAROLLO DE PROCEDIMIENTO E INVESTIGACIÓN**

Para el presente trabajo se usa una serie temporal de 13 gestiones, 2005-2017, esta base de datos tiene como variable principal el número de empresas por gestión, desagregadas por tipo de empresa y departamento. La base de datos tiene un carácter acumulado, es decir, el número de empresas de una gestion *n* se acumulara respecto de una gestión anterior *n-1*. Para ver la especificación de columnas ver el **ANEXO 1**.

**Definición de las variables de estudio.**

Para el presente estudio, se presentan las siguientes variables de interes, podemos ver que contamos con dos variables cualitativas y dos variables cuantitativas.

La variable independiente sera el número de empresas

La variable dependiente será la gestión.

**Departamento** se refiere a los 9 departamenos del pais por lo tanto los valores que puede tener son



**Tipo de empresa** se refiere a los tipos de empresa constituidos y registrados.



**Procedimiento de análisis descriptivo**

Este método se basa en una o varias preguntas de investigación y no tiene una hipótesis. Además, incluye la recopilación de datos relacionados, posteriormente, los organiza, tabula y describe el resultado. Un análisis básico descriptivo implica el calcular las medidas simples de composición y distribución de variables.

El procedimiento de análisis descriptivo corresponde al número de empresas en el tiempo y se usa para resumir los datos.

Tabla de distribución de frecuencias

**Medidas de localización**

La media es la medida de localizacion mas importante, llamado tambien el promedio de una variable. La media proporciona una medida de localización central de los datos.

La mediana es otra medida de localización central. Es el valor de enmedio en los datos orde-

nados de menor a mayor (en forma ascendente). Cuando tiene un número impar de observaciones, la mediana es el valor de enmedio. Cuando la cantidad de observaciones es par, no hay un número enmedio.

La moda es el valor que se presenta con mayor frecuencia.

**Medidas de variabilidad**

La **varianza** es una medida de variabilidad que utiliza todos los datos. La varianza está basada en la diferencia entre el valor de cada observación (*Xi*) y la media. A la diferencia entre cada valor *Xi* y la media ( cuando se trata de una muestra, μ cuando se trata de una población) se le llama desviación respecto de la media.

La **desviación** **estándar** se define como la raíz cuadrada positiva de la varianza. Continuando con la notación adoptada para la varianza muestral y para la varianza poblacional, se emplea para denotar la desviación estándar muestral y σ para denotar la desviación estándar poblacional.

Se realiza un análisis de regresion con respecto al numero de empresas en funciones al tiempo.

Realizar anova en funcion del modelo de regresion obtenido

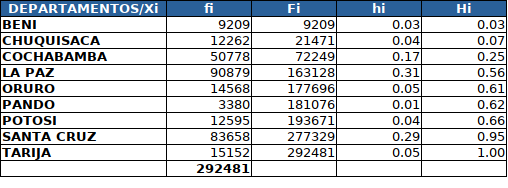
descriptivo

regresion pronosticar para dos anios numero de empresa por tipo

anova para realizar un analisis del grado de explicacion del modelo

Para el estudio de este trabajo se dispondrá el número de empresas por departamento y tipo de empresa desde la gestión 2007.

**Elaboración de tablas de frecuencia de las variables de estudio para la gestión 2017**



**sacar estadisticos.**

**PRESENTACIÓN DE RESULTADOS DEL PADRON EMPRESARIAL BOLIVIANO**

**Tablas**

**graficas**

**citar al anexo**

1. **CONCLUSIONES**

**de acuerdo al modelo de regresion y anova se puede apreciar que …..**

1. **RECOMENDACIONES**
2. **BIBLIOGRAFIA**

[1] David Anderson, Dennis Sweeney, Thomas A. Williams, Statistics for business and economy, 10th ed. Col. Santa Cruz Manca, Santa Fe: Cengage Learning Editores, pp. 2-13, 2008.

[2] https://www.r-project.org/ “The R Project for Statistical Computing”[online] Available: https://www.r-project.org/ 2020.

1. **ANEXOS**

**ANEXO 1**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | **Tipo** |
| Gestion | Gestión de registro | Númerico |
| Departamento | Departamento de la empresa | Cadena |
| tipo | Tipo de empresa | Cadena |
| Cantidad | Cantidad de empresas | Númerico |