



Proyecto Final de Análisis Estadístico I
“Estimación del padrón empresarial Boliviano”

Profesor:

Lic. Carlos Lopez

Módulo:

Análisis Estadístico I

Ivan Fernando Mujica Mamani

Índice

1. RESUMEN.....	3
2. INTRODUCCIÓN.....	3
3. OBJETIVOS.....	4
3.1 Objetivo General.....	4
3.2 Objetivos Específicos.....	4
4. DESAROLLO DE PROCEDIMIENTO E INVESTIGACIÓN.....	4
4.1 Definición de las variables de estudio.....	4
4.2 Procedimiento de análisis descriptivo.....	5
4.3 Tabla de distribución de frecuencias.....	5
4.4 Medidas de localización.....	6
4.5 Medidas de variabilidad.....	6
4.6 Procedimiento de Análisis de Regresión Lineal.....	6
5. PRESENTACIÓN DE RESULTADOS DEL PADRON EMPRESARIAL BOLIVIANO.....	7
5.1 Medidas de tendencia central.....	10
5.2 Medidas de dispersión.....	10
5.3 Análisis de regresión lineal.....	11
5.4 Análisis de Varianza.....	13
6. CONCLUSIONES.....	14
7. RECOMENDACIONES.....	14
8. BIBLIOGRAFÍA.....	14

1. RESUMEN

En el entorno 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 Contaduría, 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 estadísticos descriptivos, análisis de varianza y regresión al número de empresas, primeramente se hará un proceso manual para hallar los estadísticos 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 padrón 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). El nombre formal de esta función del estado y según normativa vigente es “Registro de Comercio de Bolivia” y esta sujeta también a la Autoridad de Fiscalización de empresas.[3]

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 público 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 dispersión tradicionales, así también las tablas de distribución de frecuencias. En la segunda parte aplicaremos un análisis de varianza en función al modelo de regresión obtenido.

3. OBJETIVOS

3.1 Objetivo General

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

3.2 Objetivos Específicos

- Aplicar Estadísticos de Dispersión y de Tendencia Central a la base de datos.

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

4. 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 gestión n se acumulara respecto de una gestión anterior $n-1$.

Tabla 1. Diccionario de datos

Variable	Descripción	Tipo
Gestion	Gestión de registro	Númerico
Departamento	Departamento de la emp	Cadena
tipo	Tipo de empresa	Cadena
Cantidad	Cantidad de empresas	Númerico

4.1 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 la gestión.

La variable dependiente será el número de empresas.

Variables	Descripción
Cualitativa	Departamento
Cualitativa	Tipo de empresa
Cuantitativo	Gestión
Cuantitativo	Cantidad

Tabla 2. Tipos de variables

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

DEPARTAMENTOS
BENI
CHUQUISACA
COCHABAMBA
LA PAZ
ORURO
PANDO
POTOSI
SANTA CRUZ
TARIJA

Tabla 3. Valores de Departamento

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

TIPO
EMPRESA UNIPERSONAL
ENTIDAD FINANCIERA DE VIVIENDA
SOCIEDAD ANONIMA
SOCIEDAD ANONIMA MIXTA
SOCIEDAD COLECTIVA
SOCIEDAD CONSTITUIDA EN EL EXTRANJERO
SOCIEDAD DE RESPONSABILIDAD LIMITADA
SOCIEDAD EN COMANDITA POR ACCIONES
SOCIEDAD EN COMANDITA SIMPLE
SUCURSAL DE SOCIEDAD CONSTITUIDA EN EL EXTRANJERO

Tabla 4. Valores de Tipo de Empresa

4.2 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.

4.3 Tabla de distribución de frecuencias

Las distribuciones de frecuencias son tablas en que se dispone las modalidades de la variable por filas. En las columnas se dispone el número de ocurrencias por cada valor, porcentajes, etc. La finalidad de las agrupaciones en frecuencias es facilitar la obtención de la información que contienen los datos. Se realizara una distribución del número de inscritos por a lo largo de los años.

4.4 Medidas de localización

La media es la medida de localización mas importante, llamado tambien el promedio de una variable. La media proporciona una medida de localización central de los datos en el tiempo según las gestiones 2005 2017, tambien se realizara el calculo de la tasa de crecimiento anual asi como el calculo de tiempo de duplicación de empresas.

4.5 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 (X_i) y la media. A la diferencia entre cada valor X_i y la media (cuando se trata de una muestra, μ cuando se trata de una población) se le llama desviación respecto de la media[1].

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. La medición de la dispersión de las empresas creadas anualmente en promedio desde la gestion 2005 al 2017[1].

4.6 Procedimiento de Análisis de Regresión Lineal

El análisis de regresión lineal es una técnica estadística utilizada para estudiar la relación entre variables. Se adapta a una amplia variedad de situaciones. En la investigación social, el análisis de regresión se utiliza para predecir un amplio rango de fenómenos, desde medidas económicas hasta diferentes aspectos del comportamiento humano. En el contexto de la investigación de mercados puede utilizarse para determinar en cuál de diferentes medios de comunicación puede resultar más eficaz invertir; o para predecir el número de ventas de un determinado producto.

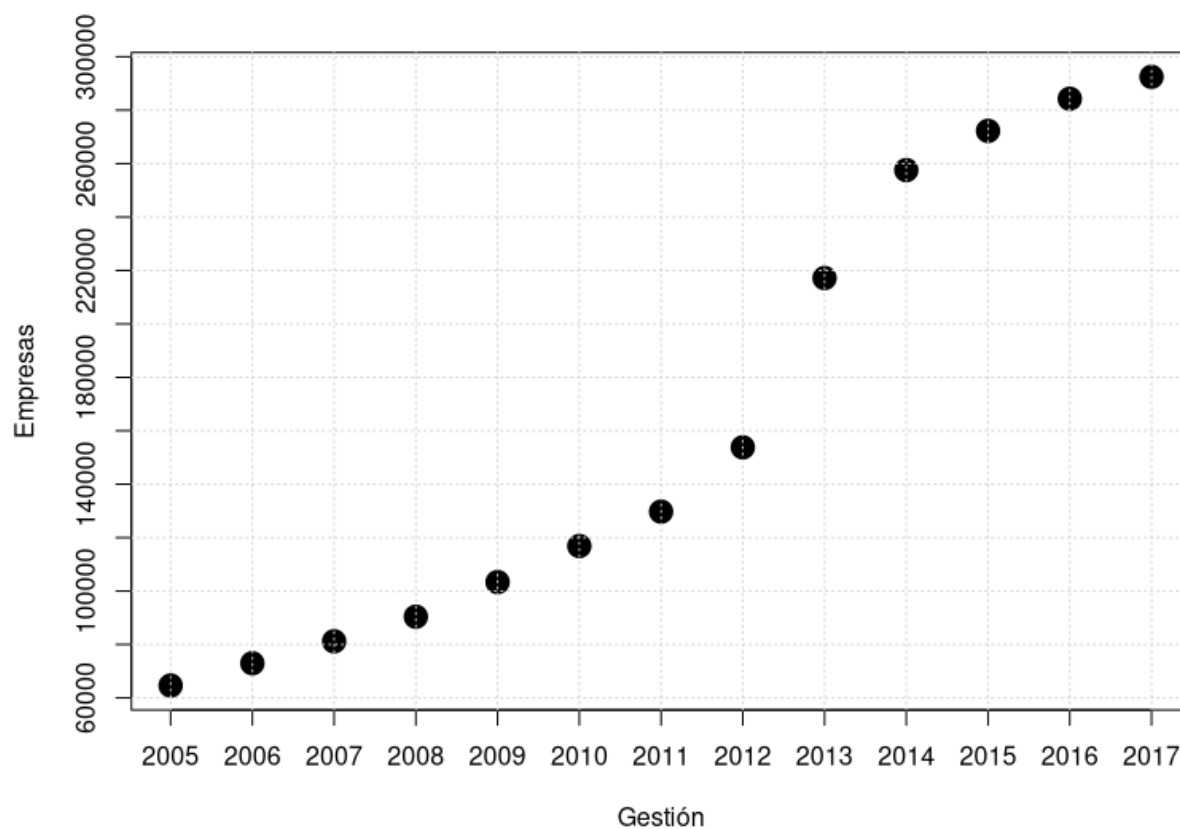
Se realiza un análisis de regresion con respecto al numero de empresas en funcion al tiempo con los datos de la gestion 2005 a la gestion 2017 explicados en los acapites anteriores. Este analisis tiene por objetivo estimar el número de empresas para los siguientes 5 años, es decir, desde la gestion 2018 hasta la gestion 2022.

5. PRESENTACIÓN DE RESULTADOS DEL PADRON EMPRESARIAL BOLIVIANO

Padron empresarial boliviano desde el 2005 al 2017, adicionalmente en la tercera columna mostramos el crecimiento del padrón.

Gestion	Cantidad	Crecimiento %
2005	64633	
2006	72969	11.42
2007	81191	10.13
2008	90398	10.18
2009	103353	12.53
2010	116857	11.56
2011	129724	9.92
2012	153792	15.65
2013	217164	29.18
2014	257564	15.69
2015	272249	5.39
2016	284271	4.23
2017	292481	2.81

Tabla 5. Número de empresas por año y crecimiento porcentual



Crecimiento del número de empresas por año.

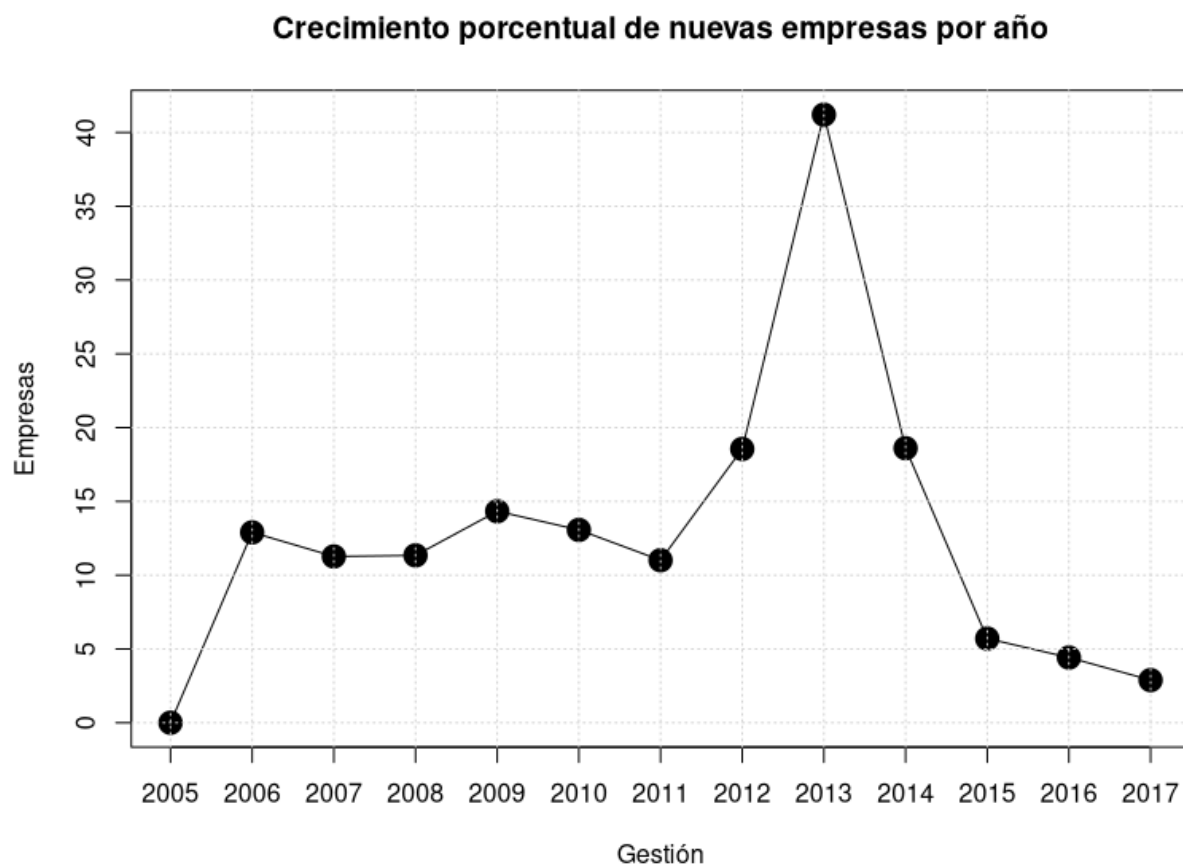


Fig. 2 Crecimiento porcentual

En el año 2013 hubo el mayor incremento de registros de empresas con un 41% de incremento.

Elaboración de tablas de frecuencia de las variables de estudio para la gestión 2017 desagregado por departamento

DEPARTAMENTOS/Xi	fi	Fi	hi	Hi
BENI	9209	9209	0.03	0.03
CHUQUISACA	12262	21471	0.04	0.07
COCHABAMBA	50778	72249	0.17	0.25
LA PAZ	90879	163128	0.31	0.56
ORURO	14568	177696	0.05	0.61
PANDO	3380	181076	0.01	0.62
POTOSI	12595	193671	0.04	0.66
SANTA CRUZ	83658	277329	0.29	0.95
TARIJA	15152	292481	0.05	1.00
	292481			

Tabla 6. Distribución de Frecuencias de la gestión 2017

Participación porcentual de los departamentos en la gestión 2017

Participación porcentual según departamento 2017

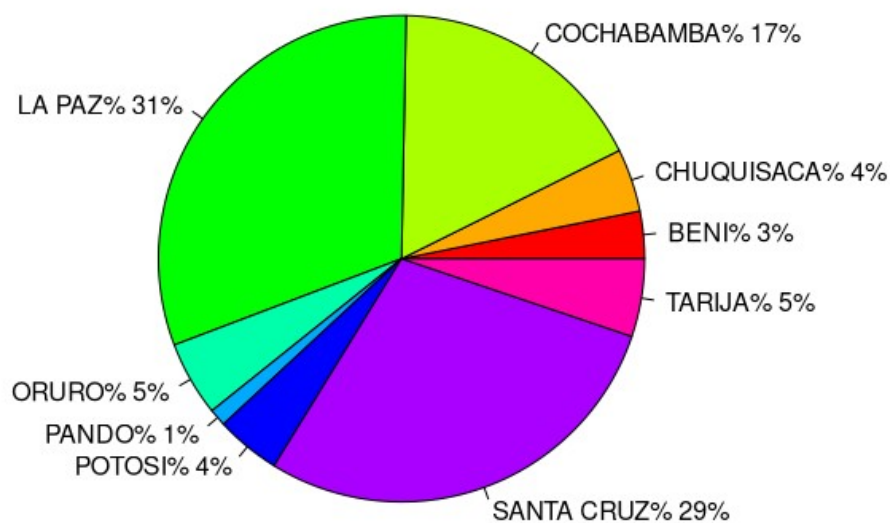


Fig. 3 Participación porcentual de cada Departamento

Para la determinación de los estadísticos usaremos R y las librerías implementadas que este lenguaje ofrece.

5.1 Medidas de tendencia central

La media esta dada por : $\bar{X} = \frac{\sum_{i=0}^n X_i}{n}$

aplicando la función *summary* en R tenemos:

```
> datos = read.csv("./padron_anios.csv", header=TRUE)
> summary(datos$Cantidad)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 64633  90398  129724 164357  257564  292481
```

De donde podemos interpretar que la *media* es 164357 empresas y tenemos una mediana de 129724 empresas registradas en el padrón.

5.2 Medidas de dispersión

La varianza esta dada por:

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{n} \text{ y la desviación típica o estandar por } \sigma = \sqrt{(\sigma^2)}$$

Aplicación en R:

```
> datos = read.csv("./padron_anios.csv", header=TRUE)
> var(datos$Cantidad)
[1] 7656991368
> sd(datos$Cantidad)
[1] 87504.24
```

De donde se puede apreciar que la *desviación estandar* es 87504.24 al rededor de la media.

5.3 Análisis de regresión lineal

Dado que $y = a + b \cdot x$ y que a y b están definidas de la siguiente forma:

$$a = \frac{\sum Y \cdot \sum X^2 - \sum X \cdot \sum (X \cdot Y)}{n \cdot \sum X^2 - (\sum X)^2}$$

$$b = \frac{n \cdot \sum (X \cdot Y) - \sum X \cdot \sum Y}{n \cdot \sum X^2 - (\sum X)^2}$$

Primeramente encontramos $X \cdot Y$ y X^2 para cada valor en la siguiente tabla.

X	Y	X*Y	X*X
2005	64633	129589165	4020025
2006	72969	146375814	4024036
2007	81191	162950337	4028049
2008	90398	181519184	4032064
2009	103353	207636177	4036081
2010	116857	234882570	4040100
2011	129724	260874964	4044121
2012	153792	309429504	4048144
2013	217164	437151132	4052169
2014	257564	518733896	4056196
2015	272249	548581735	4060225
2016	284271	573090336	4064256
2017	292481	589934177	4068289
26143	2136646	4300748991	52573755

Tabla 7. Cálculo de $X \cdot Y$ y X^2

Utilizamos la definición de a y b con los datos obtenidos en la tabla anterior

$$a = \frac{\sum Y \cdot \sum X^2 - \sum X \cdot \sum (X \cdot Y)}{n \cdot \sum X^2 - (\sum X)^2} = \frac{2136646 \cdot 52573755 - 26143 \cdot 4300748991}{13 \cdot 52573755 - 26143^2} = -43523899.4010$$

$$b = \frac{n \cdot \sum (X \cdot Y) - \sum X \cdot \sum Y}{n \cdot \sum X^2 - (\sum X)^2} = \frac{13 \cdot 4300748991 - 26143 \cdot 2136646}{13 \cdot 52573755 - (26143)^2} = 21724.6428$$

Primeramente del gráfico de crecimiento podemos observar que existe una anomalía en el año 2013, un crecimiento del 43% que está muy alejada de la media del crecimiento, para evitar valores beta falso (21724.64) no es real, aplicaremos logaritmo a las cantidades. La tabla quedaría de la siguiente manera.

X	Y	X*Y	X*X
2005	11.0765	22208.3432	4020025
2006	11.1978	22462.7667	4024036
2007	11.3046	22688.2513	4028049
2008	11.4120	22915.2507	4032064
2009	11.5459	23195.7243	4036081
2010	11.6687	23454.0996	4040100
2011	11.7732	23675.8336	4044121
2012	11.9434	24030.0329	4048144
2013	12.2884	24736.5655	4052169
2014	12.4590	25092.4734	4056196
2015	12.5145	25216.6618	4060225
2016	12.5577	25316.2895	4064256
2017	12.5862	25386.2746	4068289
26143	154.3277	310378.5670	52573755

Tabla 8. Cálculo de $X*Y$ y X^2 con la base de logaritmos

$$a = \frac{\sum Y * \sum X^2 - \sum X * \sum (X * Y)}{n * \sum X^2 - (\sum X)^2} = \frac{154.3277 * 52573755 - 26143 * 310378.5670}{13 * 52573755 - 26143^2} = -270.97201352712$$

$$b = \frac{n * \sum (X * Y) - \sum X * \sum Y}{n * \sum X^2 - (\sum X)^2} = \frac{13 * 310378.5670 - 154.3277 * 26143}{13 * 52573755 - (26143)^2} = 0.140648122179375$$

Aplicación en R:

```
> mod <- lm(log(cantidad) ~ gestion, data = data)
> summary(mod)
```

```
Call:
lm(formula = log(cantidad) ~ gestion, data = data)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.12909 -0.06201 -0.01692  0.04901  0.16572
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.710e+02  1.379e+01  -19.65 6.47e-10 ***
gestion      1.406e-01  6.859e-03   20.51 4.09e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.09253 on 11 degrees of freedom
Multiple R-squared:  0.9745, Adjusted R-squared:  0.9722
F-statistic: 420.5 on 1 and 11 DF, p-value: 4.086e-10
```

El coeficiente de determinación (R cuadrado) es de 97.5% significa que el modelo explica el 97.5% del número de empresas desde el 2005 al 2017.

El nuevo modelo de regresión y final $Y = -270.9720 + 0.1406X$ donde:
 X : es el tiempo, Y: cantidad, el valor 0.1406 se interpreta como porcentaje de incremento

Una vez obtenido el modelo de regresión procedemos a estimar el número de empresas desde la gestión 2016 al 2022:

```
> gest <- seq(2016,2022,1)
> dat<- data.frame(gestion=gest)
> pronostico <- exp(predict(mod,dat))
> names(pronostico) <- gest
> pronostico
      2016      2017      2018      2019      2020      2021      2022
289121.1 332784.0 383040.9 440887.6 507470.2 584108.1 672319.9
>
```

Gestión	Cantidad
2016	289121
2017	332784
2018	383040
2019	440887
2020	507470
2021	584108
2022	672319

Tabla 9. Estimación hasta el 2022

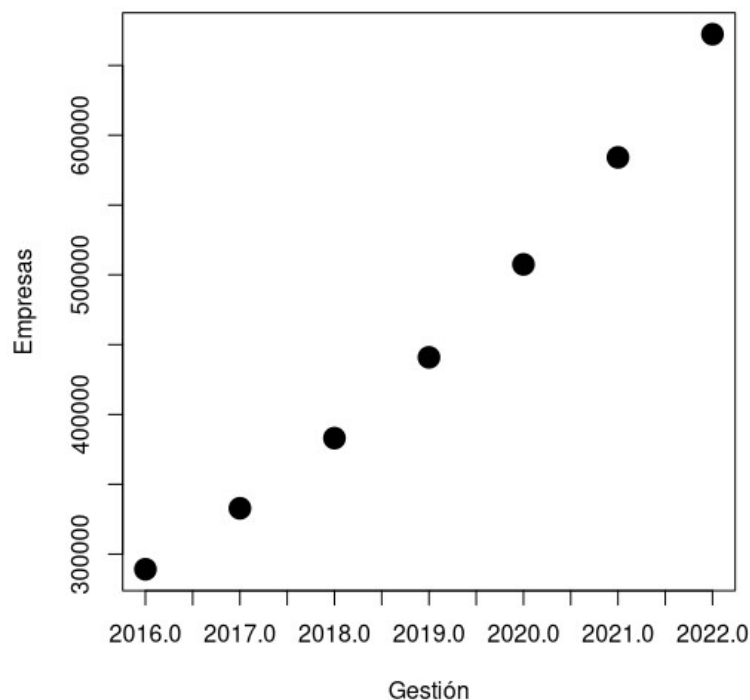


Fig.4 Estimación de Número de empresas hasta el 2022

5.4 Análisis de Varianza

```
> # analisis de regresion
> # tabla de regresión anova
> summary(aov(mod))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gestion	1	3.600	3.600	420.5	4.09e-10 ***
Residuals	11	0.094	0.009		

```
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Se propone que la hipótesis es que los betas (coeficientes de la regresión) son iguales a cero versus que no lo sean. Como el pvalor es 4.09e-10 significa que se rechaza la hipótesis nula.

Existe evidencia estadística que los coeficientes son distintos a cero y por lo tanto son significativos.

6. CONCLUSIONES

Por lo expuesto en el transcurso del presente trabajo se han calculado las medidas de tendencia central y dispersión tradicionales y luego se ha realizado un análisis de regresión lineal para estimar el crecimiento del padrón en los siguientes años.

7. RECOMENDACIONES

De acuerdo a lo mostrado en el presente trabajo es recomendable aplicar las mismas técnicas a los datos desagregados de la base de datos.

8. BIBLIOGRAFÍA

[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.

[3] <https://www.fundempresa.com.bo/> "Fundempresa - Registro de comercio de Bolivia"[online] Available: <https://www.fundempresa.com.bo>.