***Universidad‌ ‌Nacional‌ ‌del‌ ‌Litoral***

***Facultad‌ ‌de‌ ‌Ingeniería‌ ‌y‌ ‌Ciencias‌ ‌Hídricas*** *‌*

**

**Estadística‌**

**‌**

**TRABAJO FINAL**

**‌**

**Carrera‌**:‌‌ ‌Ingeniería‌ ‌Ambiental‌,

**‌**

**Asignatura‌**:‌ ‌Es‌tadística

**Alumna:‌ ‌**

* Flores Tortul, Aitana

**Profesores:‌ ‌ ‌**

* Vanlesberg, Susana
* Taulamet, Juan Pablo
* Llop, Maria José **‌**

**‌**

**Fecha‌ ‌de‌ ‌entrega:‌‌** 14/12/2021

**INTRODUCCIÓN**

En el siguiente trabajo se pretende aplicar los conocimientos sobre estadística adquiridos durante el cursado de la asignatura a un tema en específico: incendios en los Parques Nacionales.

La pérdida de bosque causada por incendios, en especial de bosques nativos, ha causado gran preocupación en los últimos años ya tiene grandes consecuencias como la extinción de especies autóctonas, contaminación del agua, contaminación del aire, afectando la salud de poblaciones cercanas, entre otras. Debido a esto se ha decidido estudiar esta temática.

Los datos a utilizar es superficie incendiada y cantidad de incendios; los cuales fueron obtenidos de la página web de la Secretaría de Ambiente y Desarrollo Sostenible de la Nación.

**OBJETIVOS**

En este trabajo se desea obtener y mostrar información sobre la superficie incendiada en los Parques Nacionales, en su totalidad, por sectores y en relación con la cantidad de incendios en un periodo de 31 años (1990-2020), utilizando las herramientas aprendidas en el curso de estadística. Se trabajará con datos de hectáreas incendiadas, variable continua.

* Identificación de la variable.
* Análisis exploratorio a través de estadística descriptiva.
* Una vez obtenida una estimación puntual de la media poblacional se realizará un test de hipótesis para tomar conclusiones acerca de la misma.
* La estimación por intervalos de confianza de la varianza poblacional a través de su estimador puntual que es la varianza muestral.
* Conocer la probabilidad de que, en 60 años, 10 tengan menos de 100ha de superficie de bosque nativo incendiada.
* Estudiar la relación entre las hectáreas afectadas y la cantidad de incendios.

**DESARROLLO**

Se comienza analizando la muestra con un análisis exploratorio a partir de estadística descriptiva.

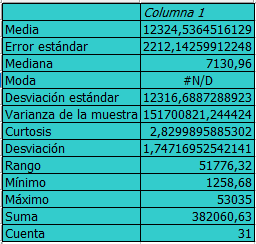
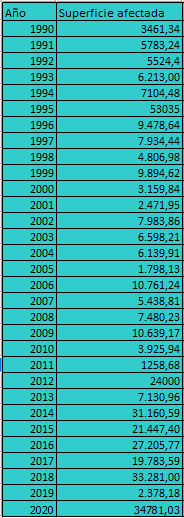
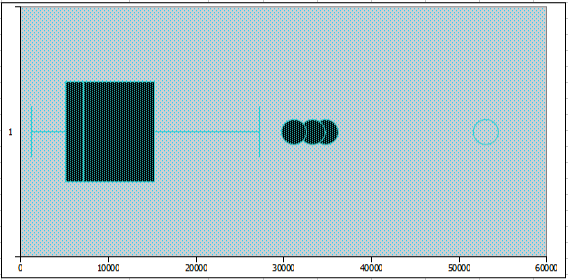


Gráfico de box plot:



*Datos*:

* Media muestral (ẋ): 12.324,54
* Cantidad de muestra: n= 31
* Desvío estándar (S’): 12.316,69
* Desvío estándar de la población (𝜎): desconocido
* 1-⍺=95%
* ⍺= 5%
* Grados de libertad: 30

En primera instancia se procede a estimar el intervalo de confianza en el que se encontrará el parámetro de la población (µ). Para ello se utilizan los datos obtenidos a través de la estadística descriptiva en el programa de cálculos *gnumeric*.

Teniendo en cuenta que la muestra es menor a 30 se obtendrá el intervalo de confianza por medio de la distribución normal, desvío parámetro desconocido (muestra >30)

A través de dicho programa calculamos el valor de “z” para luego reemplazarlo en la ecuacion antes expuesta.

Valor de Z: =normsinv(0,95)

***Z=*** ***1,645***

Para tomar una decisión respecto al valor del parámetro (µ) haremos un test de hipótesis; en el cual partiremos del siguiente supuesto, con un nivel de significación del 5%.

H0: µ 10.000

H1: µ > 10.000

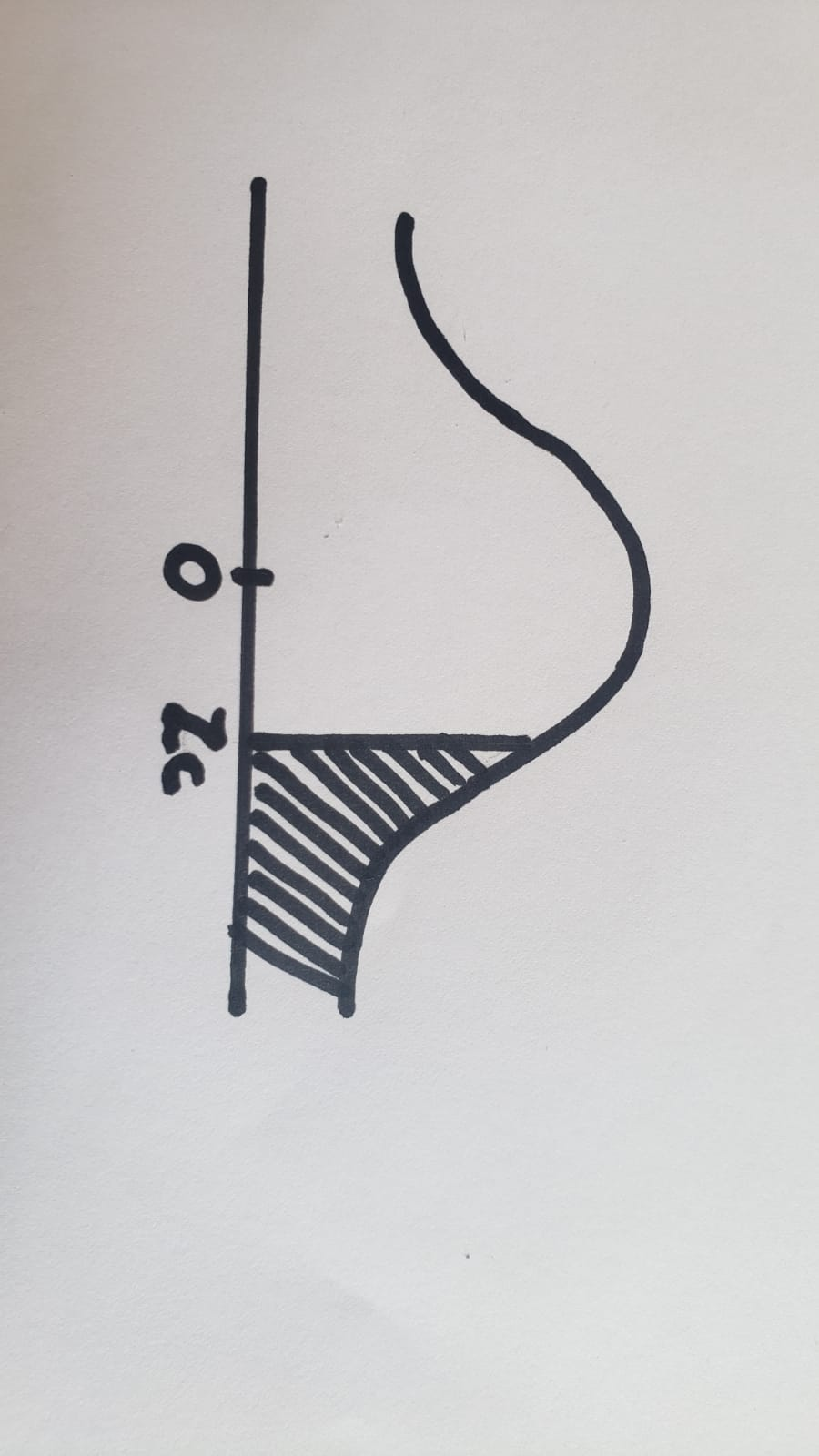
α = 5%

Se calcula por gnumeric el valor critico que divide la región de aceptación de la región de rechazo. En este caso se denomina z crítico.

*Zcritico*: = normsinv(0,95)

Luego se obtiene el *t* observado por medio de la siguiente ecuación:

Se trata de un test unilateral por derecha y se obtiene el siguiente intervalo

****

Se puede concluir que como el tobs ˂ crítico no hay evidencia suficiente para rechazar H0 porque nos encontramos en un test de cola derecha.

Para lograr un estadístico de prueba igual o más exacto al antes calculado se obtiene la probabilidad del valor *p*, que se conoce como nivel de significación observado, el cual es el mínimo nivel por el cual H0 puede no ser aceptada.

Valor de *p*:

***p = 0,8492***

Al ser el valor de ***p ˃ α***, concluimos que no hay evidencia para rechazar H0.

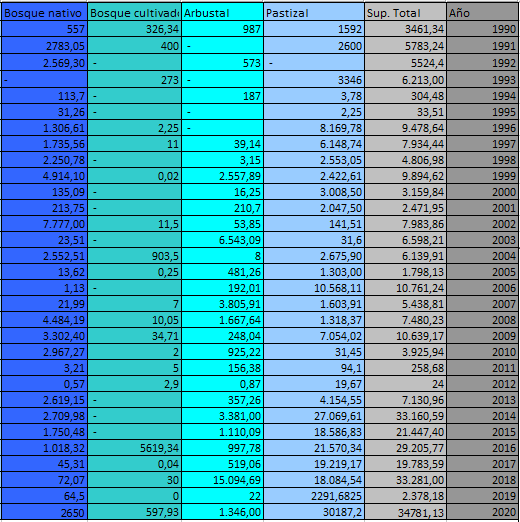
En segunda instancia se estima la varianza de la población mediante un intervalo de confianza del 95% para mejorar la estimación puntual, es decir, como varían las variables respecto de la media. Para obtener esta el intervalo utilizaremos la distribución chi-cuadrado.

*Datos:*

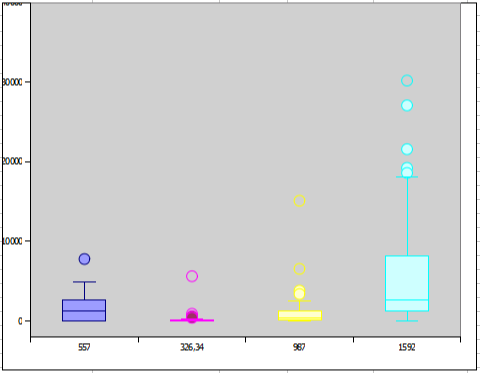
* n=31
* S’= 12.316,69
* 1-α= 95%

**280090913**

Luego es de interés analizar los incendios por sectores, para lo cual se comienza nuevamente realizando estadística descriptiva.







Debido a la existente preocupación por la pérdida de bosque nativo que lleva a la extinción de especies autóctonas, y con los datos de las áreas de vegetación, se desea determinar la probabilidad de que de los siguientes 60 años en 10 de ellos la superficie de bosque nativo incendiado no pase las 100ha.

Para comenzar, se utilizó probabilidad a posteriori:

,

siendo la cantidad de casos favorables (años en los que se incendia menos de 100 hectáreas de bosque nativo) y el n° de casos igualmente probables (siendo este el total hectáreas de bosque nativo incendiadas por año)

En los 31 años que se tomaron como muestra, en 6 se incendiaron menos de 100 hectáreas de bosque nativo, obteniendo una probabilidad del 19%.

Luego se empleó el Modelo Binomial, este modelo es útil cuando se realizan una serie de pruebas de tipo Bernoulli, en donde los resultados son mutuamente independientes y se supone que la probabilidad de éxito no cambia (en este caso, tomaremos como éxito cuando la cantidad de hectáreas incendiadas esté por debajo de 100). Teniendo en cuenta la función de cuantía de este modelo:

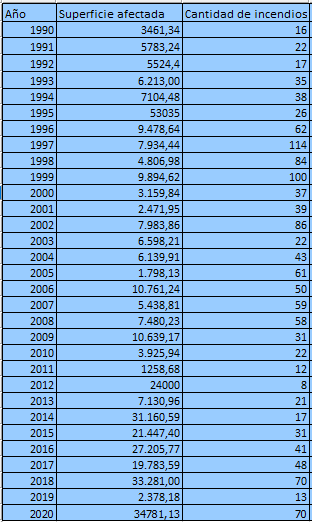
Siendo el número de éxitos, la cantidad de pruebas y la probabilidad de éxitos.

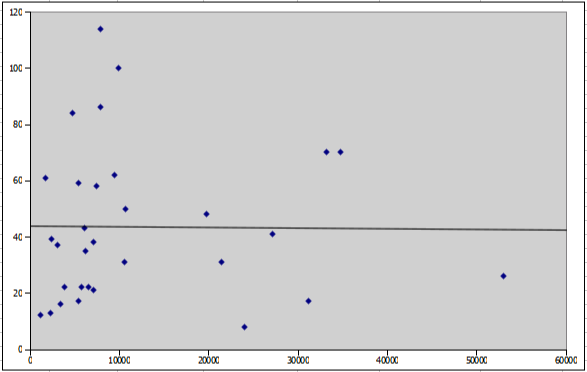
Con estos datos y suponiendo que hay 19% probabilidad se planteó el problema de la siguiente manera:

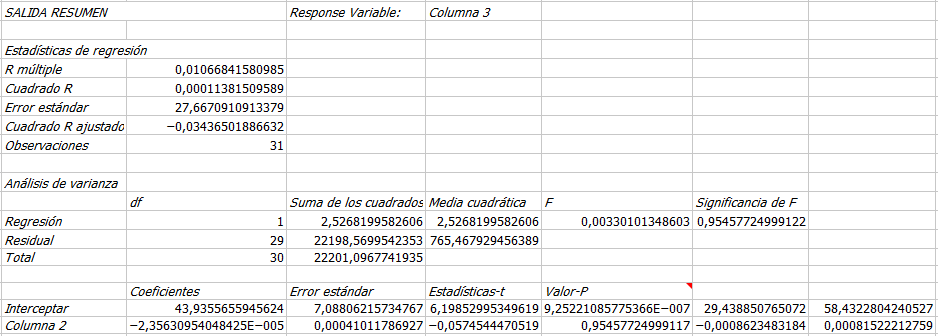
"cantidad de años en los que se incendian menos de 100 ha de bosque nativo ".

Reemplazando con los datos disponibles obtenemos:

Se concluye que existe una probabilidad de 0,000023% de que en 10 de 60 años las hectáreas de bosque nativo incendiadas sea menor a 100, la cual es muy baja.

RELACIÓN ENTRE HECTÁREAS INCENDIADAS Y CANTIDAD DE INCENDIOS 





Resultó de interés estudiar conjuntamente la superficie afectada por los incendios y la cantidad de incendios ya que generalmente se cree que estas dos variables mantienen una relación lineal. Mediante un análisis de regresión donde se realizó un gráfico de dispersión, su línea de tendencia y su información obtenida en gnumeric se quiso plantear la posibilidad de que exista una relación lineal entre ambas variables; se comprobó que no ya que al examinar el valor de R cuadrado, es decir, del coeficiente de determinación que indica la porción de la variabilidad total que es explicada por la regresión, se observa que se obtuvo un valor de aproximadamente 0,000114 sabiendo que el R cuadrado oscila entre el 0 (el modelo no explica variabilidad de los datos en torno a su media) y el 1 (se explica perfectamente), cuanto más cercano al 1, mejor se ajusta el modelo. A su vez, el valor del coeficiente de correlación, que se relaciona con la covarianza, mide la dependencia lineal entre las variables, tuvo un valor de 0,0107, al ser cercano a 0 puede decirse que la dependencia no es lineal, pero no puede asegurar que las variables son independientes.

La recta ajustada en el dispersiograma es horizontal y esto se debe a que las observaciones se dispersan alrededor del valor medio en forma aleatoria, esto no quiere decir que el modelo esté bien ajustado ya que como se observa en el cuadro de salida de regresión, R cuadrado es un valor cercano a 0 en otras palabras esto significa que la explicación de la regresión respecto de la variabilidad total es pobre.

Para confirmar si el modelo está bien ajustado o no debería realizarse un análisis de residuos para verificar que se cumplan los supuestos, pero en este caso como el coeficiente de determinación y correlación son cercanos a 0 lo que convendría sería buscar otro tipo de modelo que se ajuste correctamente

**CONCLUSIÓN**

A lo largo de este trabajo, se logró aplicar los conceptos teóricos aprendidos durante el cursado, además de que su aplicación en un fenómeno real nos ayudó a integrar su práctica de una manera óptima. A través de los valores obtenidos del software, se puede concluir que si bien los coeficientes que describen la relación entre las variables estudiadas no fueron los esperados puesto que se creía que existía una relación lineal entre ellas, debe tenerse en cuenta que se están tratando fenómenos complejos de la vida real y aquí puede que existan distintos factores que modifican estos valores.

**BIBLIOGRAFÍA**

* <http://snmb.ambiente.gob.ar/develop/>

* Apuntes de la cátedra “Estadística” – Fich UNL