

Métodos de Contraste, análisis Estadístico de Sismos en Chile

años 2003 y 2004

Oscar Henríquez

10/09/2020

Introducción

En este trabajo de metodos de contraste estadisticos hemos tomado informacion de sismos ocurridos entre los años 2003 y 2004 en Chile, con el objeto de determinar si sus magnitudes van en aumento. Esto con fines didacticos para estudiar métodos de contraste en estadística.

Sobre la Base de datos

En este caso hemos tomado la base de datos y la hemos dividido en 2 años (2003 y 2004) para mayor comodidad en la creacion de Histogramas.

El archivo de base de datos Sismos_2003.csv contiene sismos desde el 01/01/2003 hasta el 31/12/2003 con un total de 3980 registros. El archivo de base de datos Sismos_2004.csv contiene sismos desde el 01/01/2004 hasta el 31/12/2004 con un total de 3377 registros.

Tambien creamos una archivo que contiene toda la data llamado Sismos_Total.csv y contiene sismos desde el 01/01/2003 hasta el 31/12/2004 con un total de 7357 registros.

Todos los datos han sido obtenidos desde la BD de sismologia.cl del Centro Sismológico Nacional de la Universidad de Chile. Obtenidos a través de WebScrapping como se muestra en esta web: <https://benjad.github.io/2015/08/21/base-de-datos-sismos-chile>.

Variables

VARIABLE	TIPO	DESCRIPCIÓN
Fecha Local	Cuantitativa Nominal	fecha horario de chile utc -3 o utc -4 según corresponde.

VARIABLE	TIPO	DESCRIPCIÓN
Hora Local	Cuantitativa Nominal	hora en horario de chile utc -3 o utc -4 según corresponde.
Fecha UTC	Cuantitativa Nominal	fecha y considerando el tiempo universal coordinado (UTC).
Hora UTC	Cuantitativa Nominal	hora en El tiempo universal coordinado (UTC).
Latitud	Cuantitativa Nominal	Latitud geografica del sismo.
Longitud	Cuantitativa Nominal	Longitud geografica del sismo.
Profundidad(Km)	Cuantitativa Discreta	Medida en Kilometros desde el epicentro hacia el centro de la tierra.
Magnitud	Cuantitativa Discreta	Cantidad otorgada por el sistema de medicion utilizado
Escala	Cualitativa Nominal	Tipo de medida utilizada (Mb, Mc, Ml, Ms, Mw)
Lugar de Referencia	Cuantitativa Nominal	Comuna o ciudad cercana al sismo

Hipótesis

Para realizar nuestros analisis, plantearemos una hipotesis nula (H_0) y una hipótesis alternativa (H_1):

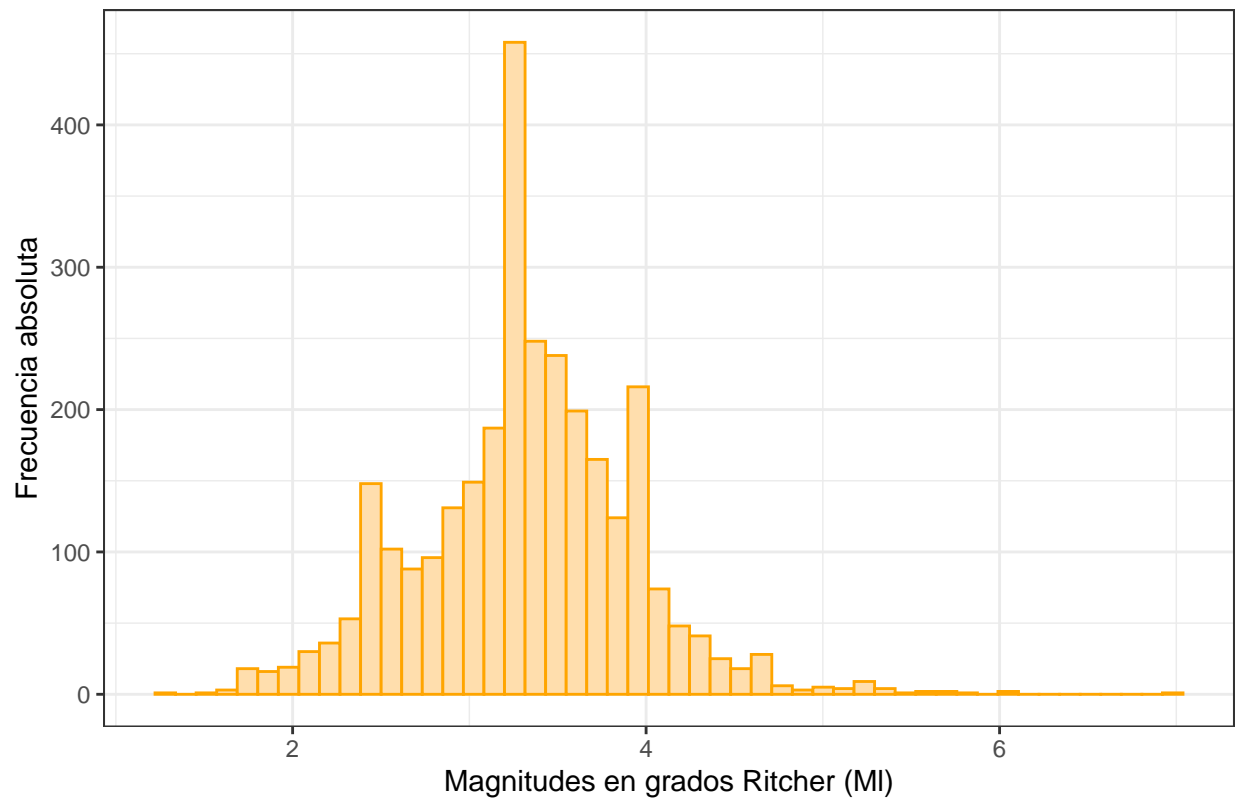
H_0) Hipótesis Nula: Los sismos en Chile aumentan cada año su magnitud.

H_1) Hipótesis Alternativa: Los sismos en Chile NO aumentan cada año su magnitud.

Histogramas

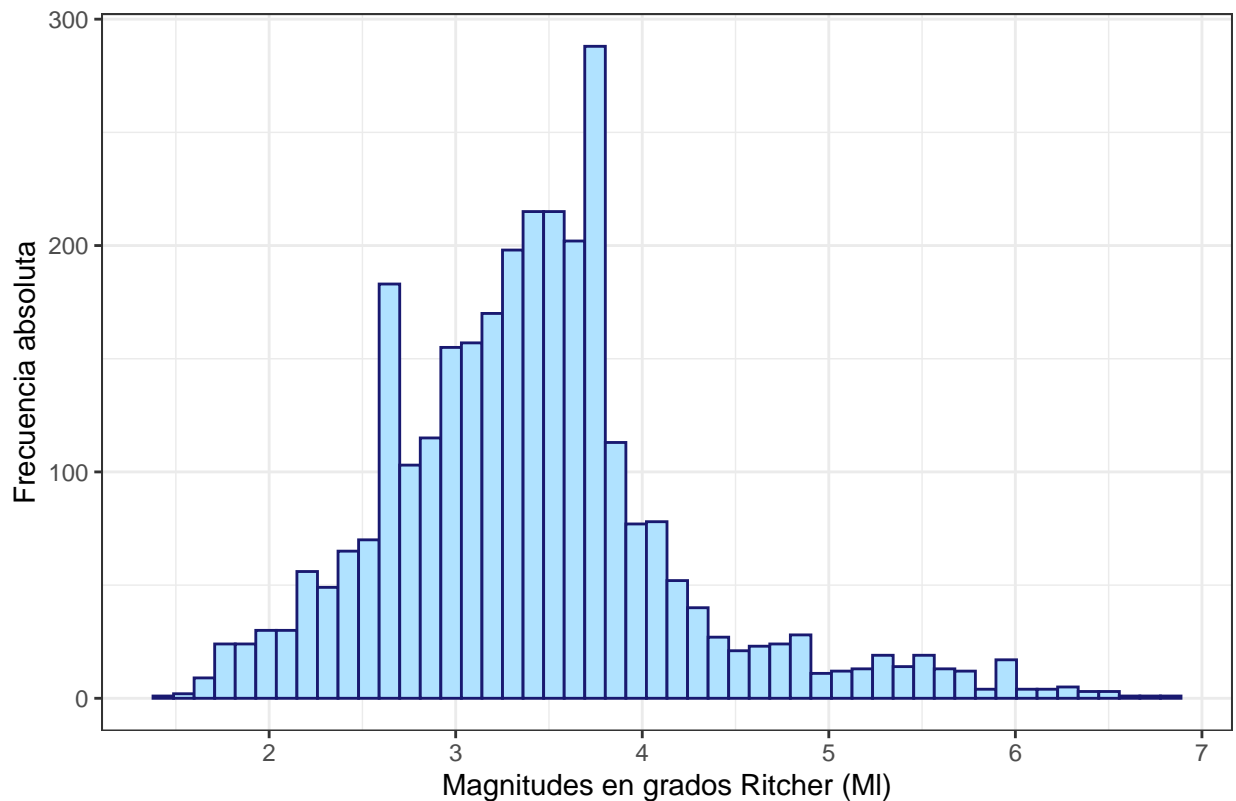
Representacion grafica de la tabla de frecuencia. En el histograma podemos observar que la mayoria de los sismos se encuentran entre los 2.0 y 5.0 grados.

Histograma de Sismos en Chile en 2003



Representacion grafica de la tabla de frecuencia. En el histograma podemos observar que la mayoria de los sismos se encuentran entre los 2.0 y 4.0 grados.

Histograma de Sismos en Chile en 2004



##Definiciones

Escala MI o de Richter

- Menos de 3.5: Generalmente no se siente, pero es registrado.
- 3.5 - 5.4: A menudo se siente, pero sólo causa daños menores.
- 5.5 - 6.0: Ocasiona daños ligeros a edificios.
- 6.1 - 6.9: Puede ocasionar daños severos en áreas muy pobladas.
- 7.0 - 7.9: Terremoto mayor. Causa graves daños.
- 8 o mayor: Gran terremoto. Destrucción total a comunidades cercanas.

4. Medidas de localización o centralidad

Las siguientes medidas de centralidad se calculan tomando el conjunto de datos completos de los años 2003 y 2004 de sismos registrados en Chile.

Media: La media aritmetica de nuestra basede datos de registro de sismos, nos muestra el promedio de la intesidad de los sismos, su valor es el siguiente:

```
## La media de intensidad de los sismos es : 3.3 grados
```

Mediana: La mediana que corresponde al valor central de toda nuestra muestra de sismos es:

```
## La mediana de intensidad de los sismos es: 3 grados
```

Moda: La moda es el caso que mas se repite, en nuestro estudio la intensidad de sismos que mas se ha repetido.

```
## La intensidad de sismos que más se repite (moda) es: 3.4 grados,  
## en 562 registros de sismos.
```

Desviación Estandar: La desviación estandar nos indica qué tan dispersos están los datos con respecto a la media, en nuestro caso su valor es:

```
## Desviación Estandar: 0.719610288661123 .
```

Estadística Inferencial

Desde este punto nuestro trabajo se enfocara en análisis estadísticos inferenciales, utilizando los anteriores análisis de estadística descriptiva como referencia para determinar que tipo de estadística utilizaremos en nuestro hipótesis a estudiar, paramétrica o no paramétrica.

Análisis de Normalidad

Para determinar si debo realizar un análisis estadístico paramétrico o no paramétrico, debo determinar si existe normalidad en mis datos.

#H0: Los datos se distribuyen de forma normal #H1: Los datos no se distribuyen de forma normal

Los siguientes test me ayudan a determinar la normalidad de mis datos:

```
lillie.test(sismosTotal$Magnitud)$p.value
```

```
## [1] 4.101597e-155
```

```
ad.test(sismosTotal$Magnitud)$p.value
```

```
## [1] 3.7e-24
```

```
pearson.test(sismosTotal$Magnitud)$p.value
```

```
## [1] 0
```

```
cvm.test(sismosTotal$Magnitud)$p.value
```

```
## Warning in cvm.test(sismosTotal$Magnitud): p-value is smaller than 7.37e-10,  
## cannot be computed more accurately
```

```
## [1] 7.37e-10
```

```
#NO soportan mas de 5000 datos para hacer el test  
#shapiro.test(sismosTotal$Magnitud)$p.value  
#sf.test(sismosTotal$Magnitud)$p.value
```

De los 4 test aplicados 3 me arrojan valores menores a 0,05 y 1 (Test CVM) me indica que no logra calcular un p value exactamente, a pesar de esto, con los resultados obtenidos podriamos decir que se tiene seguridad de que no existe distribucion normal en los datos ya que los P VALUE obtenidos son menores a 0.05.

*Los test de Saphiro y SF no se pudieron aplicar ya que la muestra tiene mas de 5000 registros.

Análisis de la homogeneidad de varianza

El siguiente test me ayuda a determinar la Homogeneidad de mis datos Prueba de homogeneidad de varianza con Bartlett H0: Varianza es homogenea H1: Variancia no es homogenea Si existe homogeneidad entre mis 2 grupos utilizo test paramétrico, sino no.

Análisis de la homogeneidad de varianza

```
bartlett.test(sismosTotal$Magnitud,  
              substring(sismosTotal$`Fecha Local`,7,8),  
              data=sismosTotal)
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data: sismosTotal$Magnitud and substring(sismosTotal$`Fecha Local`, 7, 8)  
## Bartlett's K-squared = 302.5, df = 1, p-value < 2.2e-16
```

```
#Aplico Substring para distinguir entre 03 y 04 el año  
#substring(sismosTotal$`Fecha Local`,7,8)
```

El resultado del análisis me indica que mi varianza no es homogenea dentro de los grupos estudiados ya que su P VALUE es menor a 0,05.

Conclusiones Preeliminares de Análisis realizados

Ya que mis datos NO cumplen una distribucion normal y su varianza tampoco es homogenea, debo utilizar estadística NO paramétrica. Y dado que los datos analizados corresponden a un grupo relacionado (tomando a Chile como sujeto) ya que todas las mediciones son realizadas en territorio nacional, determinamos que el test a realizar más apropiado es Wilcoxon.

```
wilcox.test(sismosTotal$Magnitud ~ substring(sismosTotal$`Fecha Local`,7,8),  
            data = sismosTotal, paired= F)
```

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data:  sismosTotal$Magnitud by substring(sismosTotal$`Fecha Local`, 7, 8)  
## W = 6211610, p-value = 2.28e-08  
## alternative hypothesis: true location shift is not equal to 0
```

Conclusiones

De acuerdo a los analisis realizados y considerando que el P VALUE de nuestro test de Wilcoxon es menor que 0.05 podemos indicar que nuestra hipotesis alternativa esta en lo correcto y esta nos indica que:

H1) Hipótesis Alternativa: Los sismos en Chile NO aumentan cada año su magnitud.

Sources

Los fuentes y datos de este reporte los puede encontrar en el siguiente repositorio:

<https://github.com/oscarhenriquezg/Estadistica/tree/master/MetodosContraste>