

Métodos de Contraste, Análisis Estadístico de Sismos en Chile

entre los años 2003 y 2004

Oscar Henríquez

10/09/2020

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 realizar nuestro estudio y contrastar los resultados obtenidos de ambos años de acuerdo a nuestra hipótesis.

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.

También creamos un 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 WebScraping 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. |
| 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). |

| VARIABLE | TIPO | DESCRIPCIÓN |
|---------------------|--------------|---|
| Latitud | Cuantitativa | Latitud geografica del sismo. |
| | Nominal | |
| Longitud | Cuantitativa | Longitud geografica del sismo. |
| | Nominal | |
| Profundidad(Km) | Cuantitativa | Medida en Kilometros desde el epicentro hacia el centro de la tierra. |
| | Discreta | |
| Magnitud | Cuantitativa | Cantidad otorgada por el sistema de medicion utilizado |
| | Discreta | |
| Escala | Cualitativa | Tipo de medida utilizada (Mb, Mc, Ml, Ms, Mw) |
| | Nominal | |
| Lugar de Referencia | Cuantitativa | Comuna o ciudad cercana al sismo |
| | Nominal | |

Hipótesis

Para realizar nuestros analisis, plantearemos una hipotesis nula (H0) y una hipótesis alternativa (H1):

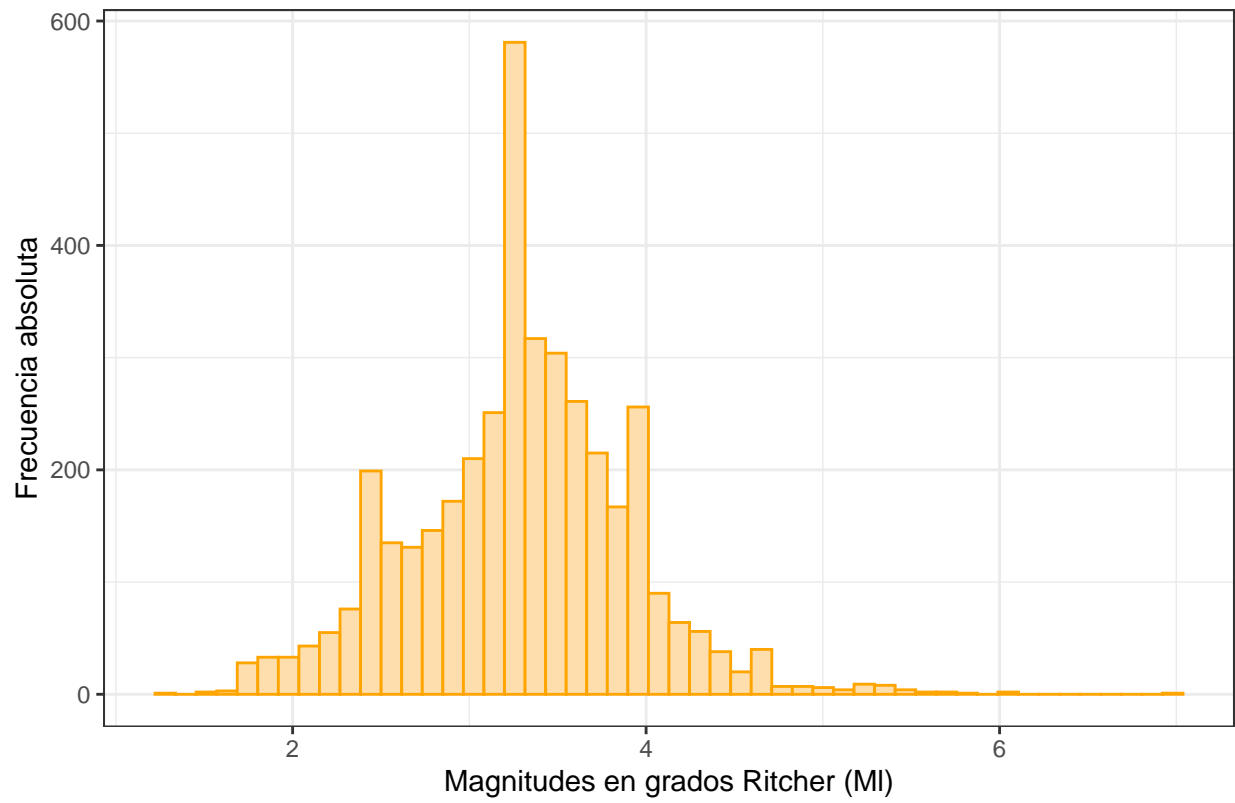
H0) Hipótesis Nula: Los sismos de menor profundidad tienen una mayor magnitud.

H1) Hipótesis Alternativa: Los sismos de menor profundidad no tienen una mayor magnitud.

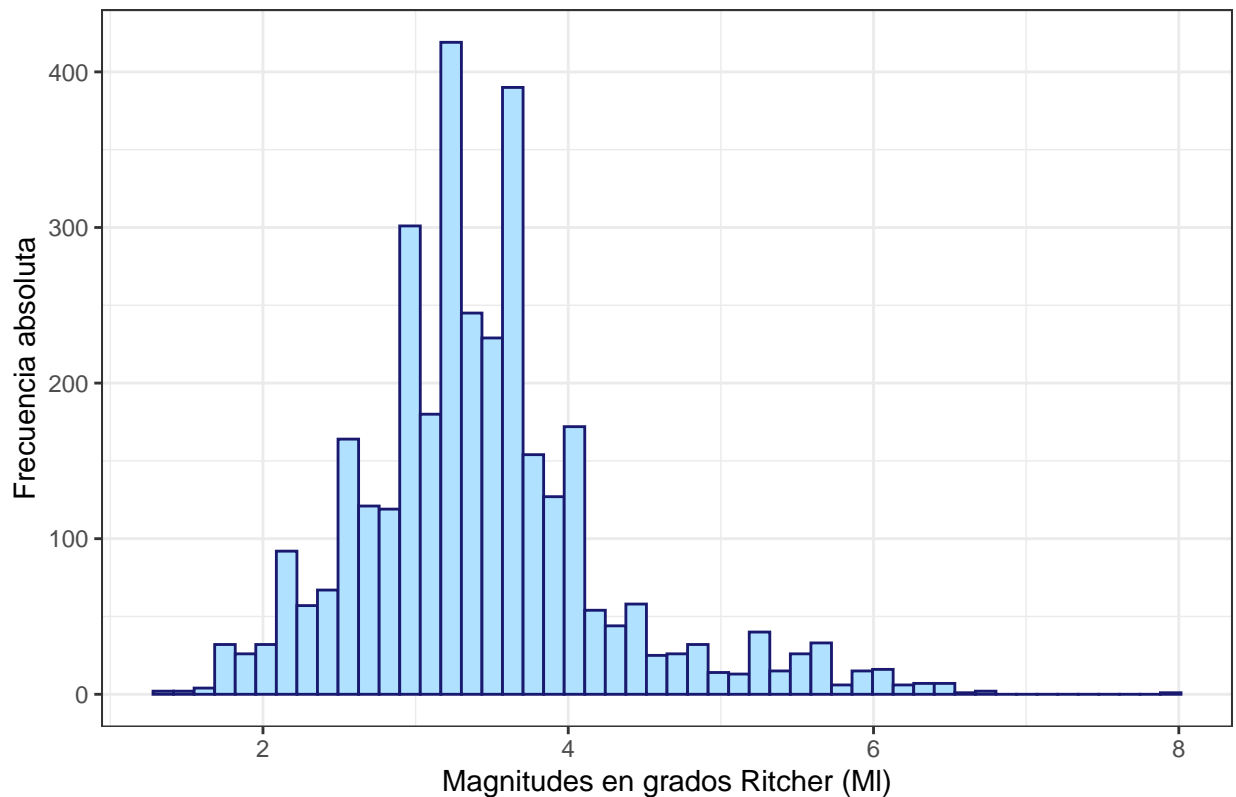
Histograma

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

Histograma de Sismos en Chile en 2003



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

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, paramétrica o no paramétrica.

Analisis de Normalidad

Para determinar si debo realizar un analisis estadistico parametrico o no parametrico, debo determinar si existe normalidad en mis datos.

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, entonces 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 t SF no se pudieron aplicar ya que la muestra tiene mas de 5000 registros.

Análisis de la homogeneidad de varianza

Ddddddffdfdfdf: Si existe homogeneidad entre mis 2 grupos utilizo test parametrico, sino no.

El siguiente test me ayuda a determinar la Homogeneidad de mis datos

Análisis de la homogeneidad de varianza Test de Bartlett: `#{r, echo = FALSE}
#prob=ols_test_bartlett(hsb, "read", group_var = "female") #prob=ols_test_bartlett(sismo
#print(prob$pval) #`

El resultado del analisis me indica que mi varianza no es homogenea dentro de los grupos estudiados ya que su valor es mayor/menor 0,05

Conclusión de Analisis realizados

Ya que mis datos NO cumplen una distribucion normal y su varianza tampoco es homogenea, debo utilizar estadistica NO paramétrica.

Sources

Los fuentes y datos de este reporte los puede encontrar en el siguiente repositorio: <https://github.com/oscarhenriquezg/Estadistica/tree/master/MetodosContraste>