

1 Análisis del Codeo y Estimaciones de Modelo

1.1 Matriz de Datos Espaciales

```
#panel
panel <- read_excel("Base.xlsx")

#Descargo el archivo con los poligonos de https://gadm.org/
arg <- raster::getData("GADM", country = "Argentina", level = 1)
summary(arg)

## Object of class SpatialPolygonsDataFrame
## Coordinates:
##           min           max
## x -73.57778 -53.59184
## y -55.06153 -21.77855
## Is projected: FALSE
## proj4string : [+proj=longlat +datum=WGS84 +no_defs]
## Data attributes:
##      GID_0          NAME_0          GID_1          NAME_1
## Length:24      Length:24      Length:24      Length:24
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##  VARNAME_1      NL_NAME_1          TYPE_1          ENGTYPE_1
## Length:24      Length:24      Length:24      Length:24
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##      CC_1          HASC_1
## Length:24      Length:24
## Class :character Class :character
## Mode :character Mode :character

#vecinos
arg_nb <- poly2nb(arg, queen=TRUE)

#coordenadas que me sirven para el plot
coords <- coordinates(arg)

#plot
plot(arg,border="grey",lwd=1.5)
plot(arg_nb, coordinates(arg), add=TRUE, col="darkred")
```

```

#tierra del fuego queda aislada por este
#criterio, sino uno queda aislado R
#directamente no te deja usar la función para la matriz

##los k=4 vecinos más cercanos (con este funciona)
arg_k <- knn2nb((knearneigh(coords,k=4)))
arg_k

```

```

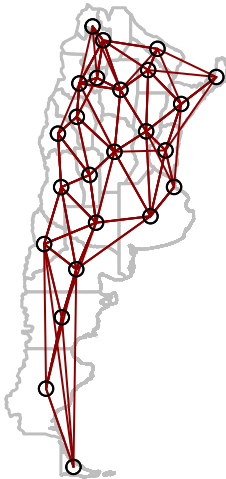
## Neighbour list object:
## Number of regions: 24
## Number of nonzero links: 96
## Percentage nonzero weights: 16.66667
## Average number of links: 4
## Non-symmetric neighbours list

```

```

warg <- nb2listw(arg_k)#funciona
plot(arg_k, coordinates(arg), add=TRUE, col="darkred")

```



```

summary(arg_k)

```

```

## Neighbour list object:
## Number of regions: 24
## Number of nonzero links: 96
## Percentage nonzero weights: 16.66667
## Average number of links: 4
## Non-symmetric neighbours list
## Link number distribution:

```

```
##
## 4
## 24
## 24 least connected regions:
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 with 4 links
## 24 most connected regions:
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 with 4 links
```

```
summary(warg)
```

```
## Characteristics of weights list object:
## Neighbour list object:
## Number of regions: 24
## Number of nonzero links: 96
## Percentage nonzero weights: 16.66667
## Average number of links: 4
## Non-symmetric neighbours list
## Link number distribution:
##
## 4
## 24
## 24 least connected regions:
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 with 4 links
## 24 most connected regions:
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 with 4 links
##
## Weights style: W
## Weights constants summary:
##      n  nn S0      S1      S2
## W 24 576 24 10.375 101.625
```

Sean Z_1, Z_2, \dots, Z_K variables aleatorias y sea S la ubicación del punto o área. La matriz de datos espaciales viene generada por:

$$\left(\begin{array}{cccc|c} z_1(1) & z_2(1) & \dots & z_K(1) & s(1) \\ z_1(2) & z_2(2) & \dots & z_K(2) & s(2) \\ \vdots & \vdots & \ddots & \vdots & \\ z_1(n) & z_2(n) & \dots & z_K(n) & s(n) \end{array} \right)$$

la cual puede ser expresada de la siguiente forma:

$$\{z_1(i), z_2(i), \dots, z_K(i) | s(i)\}_{i=1 \dots n}$$

donde z_k denota una realización muestral (valor de datos real) de la variable $Z_k(k = 1, \dots, K)$ mientras que el símbolo i dentro de los paréntesis hace referencia al caso u observación particular (punto, área, polígono o zona espacial). Para cada caso, $i = 1, \dots, n$, $s(i)$ representa la ubicación del objeto espacial. La referencia implicará dos coordenadas geográficas. Genealmente nos referiremos a un espacio bidimensional, por tanto $s(i) = (s_1(i), s_2(i))^t$.

En el caso de datos referentes a objetos de puntos en un espacio bidimensional, la ubicación del punto i -ésimo puede estar dada por un par de coordenadas cartesianas (ortogonales). Los ejes del sistema de coordenadas generalmente se han construido para el conjunto de datos particular, pero se puede usar cualquier sistema de referencia. En el caso de datos referentes a objetos de área de forma irregular, una opción para representar la ubicación es seleccionar un punto representativo de cada área denominado centroide y luego usar el mismo procedimiento que para un objeto de punto para identificar $s(i)$. Hay situaciones en las que la información de georreferencia proporcionada por $\{s(i)\}$ tiene que completarse con información de vecindad que define no solo qué pares de áreas son adyacentes entre sí, sino que también puede cuantificar la cercanía de esa adyacencia. Esta información es necesaria para la especificación de muchos modelos estadísticos espaciales, como los modelos de regresión espacial. Dicha información está recogida en el concepto de autocorrelación espacial.

El criterio de contigüidad que se utilizó es el de los 4-vecinos más cercanos, donde esto a todos los centroides un mismo número de conexiones y además obliga a los polígonos que se encuentran aislados geográficamente a relacionarse o conectarse con los demás (tierra del fuego quedaba aislada)

1.2 Modelo

SARAR: Spatial AutoRegressive with AutoRegressive error

SDM: Spatial Durbin Model

SDEM: Spatial Durbin Error Model

SLM: Spatial Lag Model

SLX: Spatial Lag in X's

SEM: Spatial Error Model: El otro modelo con buena presencia en los inicios fue el modelo SEM que permite una ganancia de eficiencia respecto al modelo no-espacial. Una crítica empírica al modelo SEM sucede cuando la autocorrelación espacial es débil: los resultados tienden a ser similares a los obtenidos por una estimación no-espacial robusta.

pooled (OLS): Para pasar de información de corte transversal a información de datos de panel solo es necesario incorporar un sub-índice t , que va de $t = 1, \dots, T$, apilando las observaciones del corte transversal periodo a periodo. Sin modificar ningún supuesto

en el término de error, este tipo de extensión espacio-temporal da lugar a los modelos pooled.

1.3 Autocorrelación Espacial.

Para Cliff y Ord (1973) es útil ver si la distribución de cierta cualidad o cantidad en los condados o estados de un país hace que su presencia en los condados vecinos sea más o menos probable. En tal caso, exponen que el fenómeno exhibe autocorrelación espacial. Sokal y Oden (1978) argumentaron que el análisis de autocorrelación espacial prueba si el valor observado de una variable nominal, ordinal o de intervalo en una localidad es independiente de los valores de esa misma variables en las localidades vecinas.

Upton y Fingleton (1985) la definen como una propiedad que los datos poseen cuando muestran un patrón de comportamiento. Estos autores exponen que la autocorrelación espacial existe siempre que haya una variación espacial sistemática en los valores a lo largo de un mapa, o patrones en los valores registrados en las localizaciones. Goodchild (1987) dice que en su sentido más general la autocorrelación espacial se refiere al grado en que los objetos o actividades en algún lugar de la superficie son similares a otros objetos o actividades ubicados cerca y refleja la primera ley de geografía de Tobler “todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las cosas distantes”.

Entonces se puede llegar a la conclusión de que si el valor de una o varias variables en una ubicación son similares a los valores de dichas variables en ubicaciones cercanas, entonces se dice que el patrón en conjunto exhibe una autocorrelación espacial positiva. Por el contrario, se dice que existe autocorrelación espacial negativa cuando las observaciones que están cerca en el espacio tienden a ser más diferentes en los valores de las variables que las observaciones que están más separadas. La autocorrelación cero se produce cuando los valores variables son independientes de la ubicación.

En los análisis de autocorrelación espacial se necesita una medida de contigüidad que podemos definirla de manera general como una relación de vecindad, estas pueden ser de tres tipos, caso de torre, caso de alfil y caso de la reina.

Un aspecto crucial de la definición de la autocorrelación espacial es la determinación de ubicaciones cercanas, es decir, aquellas ubicaciones que rodean un punto de datos dado que podría considerarse que influyen en la observación en ese punto de datos. Sin embargo, la determinación de este vecindario tiene un cierto grado de arbitrariedad. El número de observaciones en el vecindario establecido para cada ubicación puede expresarse mediante una matriz de ponderaciones W :

$$\begin{bmatrix} W_{11} & W_{12} & \dots & W_{1n} \\ W_{21} & W_{22} & \dots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \dots & W_{nn} \end{bmatrix}$$

donde n representa el número de ubicaciones. La entrada en la fila i y columna j , denotado como W_{ij} corresponde al par (i, j) de ubicaciones. Los elementos diagonales de la matriz son cero, por convenio, mientras que los demás elementos donde $i \neq j$ toman valores distintos de cero cuando las ubicaciones se consideran vecinas.

Para los datos de áreas, n representaría el número de áreas donde cada área se identifica con un punto (centroide) del que se conocen sus coordenadas cartesianas y cada elemento de la matriz W corresponde a la relación de dos áreas.

Una forma de representar las relaciones espaciales con datos de áreas es a través del concepro de contigüidad. Los vecinos contiguos de primer orden se definen como áreas que tienen un límite común. Formalmente

$$W_{ij} = \begin{cases} 1 & \text{si el área } j \text{ comparte un límite común con el área } i \\ 0 & \text{c.c} \end{cases}$$

Alternativamente, dos áreas i y j pueden definirse como vecinas cuando la distancia d_{ij} entre sus centroides es menor que un valor crítico dado, pongamos d , donde las distancias se calculan a partir de la información sore latitud y longitud, $s(i)$, de las ubicaciones del centroide:

$$W_{ij} = \begin{cases} 1 & \text{si } d_{ij} < d. (d < 0) \\ 0 & \text{c.c} \end{cases}$$

Esta especificación de la matriz de ponderaciones depende de un valor de distancia crítica dado, d . Sin embargo, cuando hay un alto grado de heterogeneidad en el tamaño de las unidades de área, puede ser difícil encontrar una distancia crítica satisfactoria. En tales circunstancias, una pequeña distancia tenderá a conducir a muchas “islas”, mientras que una distancia elegida para garantizar que cada unidad de área tenga al menos un vecino puede producir un tamaño inaceptablemente grande de número de vecinos para las unidades de áreas más pequeñas. Una solución común a este problema es restringir la estructura ccontigua a lso k -vecinos más cercanos, y por lo tanto excluir las “islas” (áreas que por no estar a una distancia d de otra áreas se podría decir que no tiene vecinos) y forzar a cada unidad de área a tener el mismo número k de vecinos. Formalmente:

$$W_{ij} = \begin{cases} 1 & \text{si el centroide de } j \text{ es uno de los } k \text{ centroides más cercanos al de } i \\ 0 & \text{c.c} \end{cases}$$

En este caso el número de vecinos, k , es el parámetro de este esquema de ponderación. También podemos cambiar la ponderación para que los vecinos más distantes obtengan menos peso introduciendo un parámetro θ que permita indicar la tasa de disminución de los pesos.

Un esquema de ponderación continua comúnmente utilizado se basa en la función de distancia inversa, de modo que los pesos están inversamente relacionados con el área de separación de distancia i y el área j donde el parámetro θ se estima o se establecen a priori:

$$W_{ij} = \begin{cases} 1 & d_{ij}^{-\theta} \text{ si la distancia entre centroides } d_{ij} < d (d > 0, \theta > 0) \\ 0 & \text{c.c} \end{cases}$$

Una vez definida la matriz de contactos, pueden utilizarse diferentes estadísticos univariantes que permiten detectar autocorrelación espacial. Los mismos pueden clasificarse como medidas de dependencia globales o locales. Las medidas globales utilizan la información completa del conjunto de datos con el fin de obtener un valor promedio para todo el espacio geográfico. Al resumir en un único valor toda la información, no es posible detectar la variabilidad de la dependencia ni la localización de estos patrones. Por su parte, las medidas locales examinan la autocorrelación espacial en un subconjunto de datos.

A la hora de estudiar los datos espaciales, está puede estar presente en variables explicativas, variable dependiente o en los residuos (términos de error). Cuando la dependencia espacial se encuentra en la variable dependiente los modelos se denominan modelos de retardo espacial mientras que si está en los residuos se denominan modelos de error espacial. Cuando está presente en las variables explicativas se llaman modelos de regresión cruzada o modelos X espacialmente retardados. Entonces, lo que primero se hace es armar el modelo como si fuera una regresión lineal y ver donde está la autocorrelación espacial y ver cual es el modelo que mejor se ajusta

1.3.1 Interpretaciones de los test

Estos contrastes permiten incorporar elementos espaciales de acuerdo al rechazo o no de cada una de las hipótesis nulas. Tal es así que puede establecerse una estrategia de especificación:

- Si I de Moran rechaza $H_0 \rightarrow$ Evidencia a favor de inclusión de elementos espaciales.
- Si LM_{ERROR} y LM_{ERROR}^* rechazan $H_0 \rightarrow$ Evidencia a favor de un modelo de error espacial (SEM).

- Si LM_{LAG} y LM_{LAG}^* rechazan $H_0 \rightarrow$ Evidencia a favor de un modelo de rezago espacial (SLM).
- Si no se rechaza H_0 bajo ninguno de los contrastes \rightarrow Evidencia a favor del modelo lineal general no espacial.
- Si ambos contrastes robustos, LM_{ERROR} y LM_{LAG}^* , rechazan H_0 entonces se deberán incorporar elementos espaciales en la parte sistemática (Wy) y aleatoria (Wu).
- es la prueba más robusta

Para seleccionar la más adecuada, partimos del modelo sin tener en cuenta la autocorrección espacial. Codigos y resultados

```
eq <- log(Pbg) ~ log(Pob) + num_uni

verdoon_pooled <- plm(eq, data=panel, model="pooling")

verdoorn_fe1 <- plm(eq, data = panel,
                    model = "within", effect="individual")

verdoorn_re1 <- plm(eq, data = panel,
                    model = "random", effect="individual")

summary(verdoon_pooled)
```

```
## Pooling Model
##
## Call:
## plm(formula = eq, data = panel, model = "pooling")
##
## Balanced Panel: n = 24, T = 26, N = 624
##
## Residuals:
##      Min.   1st Qu.   Median   3rd Qu.    Max.
## -1.88402 -0.98996 -0.22338  0.82310  3.09363
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept)  4.418634   0.344416  12.829 < 2.2e-16 ***
## log(Pob)      0.682238   0.054504  12.517 < 2.2e-16 ***
## num_uni       0.264833   0.024116  10.982 < 2.2e-16 ***
```



```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1684.4
## Residual Sum of Squares: 888.16
## R-Squared:              0.47272
## Adj. R-Squared: 0.47102
## F-statistic: 278.367 on 2 and 621 DF, p-value: < 2.22e-16
```

```
summary(verdoorn_fe1)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = eq, data = panel, effect = "individual", model = "within")
##
## Balanced Panel: n = 24, T = 26, N = 624
##
## Residuals:
##      Min.   1st Qu.   Median   3rd Qu.    Max.
## -1.17952 -0.45140 -0.12693  0.33672  2.44499
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## log(Pob)  8.712956    0.245462  35.4961 < 2.2e-16 ***
## num_uni    0.106046    0.021401   4.9553 9.416e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    975.29
## Residual Sum of Squares: 260.45
## R-Squared:              0.73296
## Adj. R-Squared: 0.72179
## F-statistic: 820.664 on 2 and 598 DF, p-value: < 2.22e-16
```

```
summary(verdoorn_re1)
```

```
## Oneway (individual) effect Random Effect Model
##      (Swamy-Arora's transformation)
##
## Call:
```

```
## plm(formula = eq, data = panel, effect = "individual", model = "random")
##
## Balanced Panel: n = 24, T = 26, N = 624
##
## Effects:
##               var std.dev share
## idiosyncratic 0.4355  0.6599  0.64
## individual    0.2450  0.4950  0.36
## theta: 0.747
##
## Residuals:
##      Min.   1st Qu.   Median   3rd Qu.    Max.
## -1.74845 -0.89042 -0.14512  0.69192  2.64971
##
## Coefficients:
##               Estimate Std. Error z-value Pr(>|z|)
## (Intercept) -3.422275   1.053974 -3.2470  0.001166 **
## log(Pob)     1.842667   0.157855 11.6731 < 2.2e-16 ***
## num_uni       0.273810   0.031376  8.7267 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1020.7
## Residual Sum of Squares: 679.62
## R-Squared:      0.33415
## Adj. R-Squared: 0.332
## Chisq: 311.64 on 2 DF, p-value: < 2.22e-16
```

Luego implementamos las pruebas de estándar de Hausman y su versión robusta. Los resultados para determinar la autocorrelación espacial, conducen al rechazo de la hipótesis nula sobre la ausencia de correlación entre los efectos individuales y las variables explicativas. Por lo que se elige un modelo de efectos fijos. (SAR > SEM en los estadísticos y en el nivel de significancia)

```
## Prueba de hausman
```

```
hausman_panel <- phtest(eq, data = panel)

spat_hausman_ML_SEM <- sphtest(eq,data=panel, listw=warg,
                              spatial.model = "error", method="ML")

spat_hausman_ML_SAR<-sphtest(eq,data=panel, listw =warg,
                              spatial.model = "lag", method="ML")
```

```
hausman_panel
```

```
##  
## Hausman Test  
##  
## data: eq  
## chisq = 1316.3, df = 2, p-value < 2.2e-16  
## alternative hypothesis: one model is inconsistent
```

```
spat_hausman_ML_SEM
```

```
##  
## Hausman test for spatial models  
##  
## data: x  
## chisq = 35.043, df = 2, p-value = 2.458e-08  
## alternative hypothesis: one model is inconsistent
```

```
spat_hausman_ML_SAR
```

```
##  
## Hausman test for spatial models  
##  
## data: x  
## chisq = 17.048, df = 2, p-value = 0.0001986  
## alternative hypothesis: one model is inconsistent
```

Los resultados de las pruebas del multiplicador de Lagrange en un modelo de efectos fijos fomentan favorecer una Especificación SAR. Los estadísticos de la prueba para tomar la autocorrelación espacial por SAR (Prueba 1) o SEM (Prueba 2) confirman el rechazo de la hipótesis de que estos dos términos (tomados independientemente) son nulos, la lectura simultánea nos permite concluir sobre la especificación más adecuada para tener en cuenta la autocorrelación espacial (estas dos pruebas no están incluídas). Para concluir de una manera más creíble, se utilizan pruebas robustas en la presencia de la especificación alternativa de autocorrelación espacial (Pruebas 3 y 4).

La versión robusta de RLMerr no es significativa (Prueba 4) mientras que RLMlag si lo es (Prueba 3). Por lo tanto, es conveniente estimar un modelo Fixed-effect SAR (Ver Interpretaciones de los test). Comentario: En algunos casos, estas dos últimas pruebas robustas no permiten discriminar entre un SAR y un SEM. Son posibles varias

posibilidades. El primero consiste en estimar un modelo que contiene ambos términos espaciales (SARAR). Comentario: El segundo consiste en discriminar entre dos especificaciones sobre la base de las estadísticas de prueba RLMerr y RLMlag (utilizando la especificación con las estadísticas asociadas más altas) o comparando los criterios de Akaike de las dos especificaciones.

```
# Fixed effects model
```

```
# Test 1
```

```
slmtest(eq, data=panel, listw = warg, test="lml", model="within")
```

```
##
```

```
## LM test for spatial lag dependence
```

```
##
```

```
## data: formula (within transformation)
```

```
## LM = 807.74, df = 1, p-value < 2.2e-16
```

```
## alternative hypothesis: spatial lag dependence
```

```
# Test 2
```

```
slmtest(eq, data=panel, listw = warg, test="lme", model="within")
```

```
##
```

```
## LM test for spatial error dependence
```

```
##
```

```
## data: formula (within transformation)
```

```
## LM = 541.03, df = 1, p-value < 2.2e-16
```

```
## alternative hypothesis: spatial error dependence
```

```
# Test 3
```

```
slmtest(eq, data=panel, listw = warg, test="rlml", model="within")
```

```
##
```

```
## Locally robust LM test for spatial lag dependence sub spatial error
```

```
##
```

```
## data: formula (within transformation)
```

```
## LM = 267.46, df = 1, p-value < 2.2e-16
```

```
## alternative hypothesis: spatial lag dependence
```

```
# Test 4
```

```
slmtest(eq, data=panel, listw = warg, test="rlme", model="within")
```

```
##
## Locally robust LM test for spatial error dependence sub spatial lag
##
## data: formula (within transformation)
## LM = 0.74516, df = 1, p-value = 0.388
## alternative hypothesis: spatial error dependence
```

Modelos: Codigos y resultados variando parametros que permite la función

```
# Table

# Likelihood Maximum estimation

verdoorn_SAR_pool <- spml(eq, data = panel, listw = warg, lag=TRUE,
                          ,model="pooling")
summary(verdoorn_SAR_pool)
# Fixed-effect SAR
verdoorn_SAR_FE1 <- spml(eq, data = panel, listw = warg, lag=TRUE,
                        model="within", effect="individual", spatial.error="b")
summary(verdoorn_SAR_FE1)

verdoorn_SAR_FE2 <- spml(eq, data = panel, listw = warg, lag=TRUE,
                        model="within", effect="individual", spatial.error="kkp")
summary(verdoorn_SAR_FE1)
```

El modelo de rezago espacial, incorpora un rezago espacial de la variable dependiente Wy , como explicativa:

$$y = \lambda Wy + X\beta + \mu$$

donde λ es un parámetro espacial autoregresivo, μ es un vector de errores de dimensión $(n \times 1)$ y $\mu \sim (0, \sigma^2 I_n)$.

Resultados del modelo, acá hay tres posibles opciones dependiendo de diferentes criterios. La ecuación con la que fueron estimados los modelos es $\log(Pbg) = \log(Pob) + \text{num}_u ni$

pooling: resultados

```
coefficients:
              Estimate Std. Error t-value Pr(>|t|)
(Intercept)  6.408201    0.271333  23.6175 < 2.2e-16 ***
log(Pob)      0.863431    0.020672  41.7686 < 2.2e-16 ***
num_uni       0.068693    0.008178   8.3997 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Luego en model Within es para fixed effects y se le puede variar el argumento de la función `spatial.error` con `b` (Baltagi) o `kkp` (Kapoor). Resultados (dan iguales)

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t)	
log(Pob)	0.0718612	0.0393266	1.8273	0.0676558	.
num_uni	0.0160614	0.0042649	3.7659	0.0001659	***

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

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t)	
log(Pob)	0.0718612	0.0393266	1.8273	0.0676558	.
num_uni	0.0160614	0.0042649	3.7659	0.0001659	***

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