

Anàlisi Estadístic en el Projecte

Carla Viñas Templado

2022-05-22

ANÀLISI ESTADÍSTIC DEL PROJECTE

Carrega,filtratge i creació dels datasets

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.6      v purrr  0.3.4
## v tibble  3.1.7      v dplyr  1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(dplyr)
library(tidyr)
library(plotly)

##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##   last_plot

## The following object is masked from 'package:stats':
##
##   filter

## The following object is masked from 'package:graphics':
##
##   layout

library(shiny)

Edat_dataset <- read.csv("~/Desktop/Projecte VD/Datasets Filtrados/Edat_dataset/Edat_dataset.csv")
dataset_final <- read.csv("~/Desktop/Projecte VD/dataset_final.csv")
dataset_final <- dataset_final%>%select(state, year, hdi, gp_ratio, income_index)
Estadistica_dataset <- merge(x = Edat_dataset, y = dataset_final, by = c("state", "year"))
Estadistica_dataset <- select(Estadistica_dataset, -X)
colnames(Estadistica_dataset)[9] = "ratio_deads"

Estadistica_2015 <- Estadistica_dataset%>%filter(year==2015)
```

```

Estadistica_2016 <- Estadistica_dataset%>%filter(year==2016)
Estadistica_2017 <- Estadistica_dataset%>%filter(year==2017)
Estadistica_2018 <- Estadistica_dataset%>%filter(year==2018)
Estadistica_2019 <- Estadistica_dataset%>%filter(year==2019)

```

RELACIÓ entre el ratio de les morts amb el índex en educació dels estats

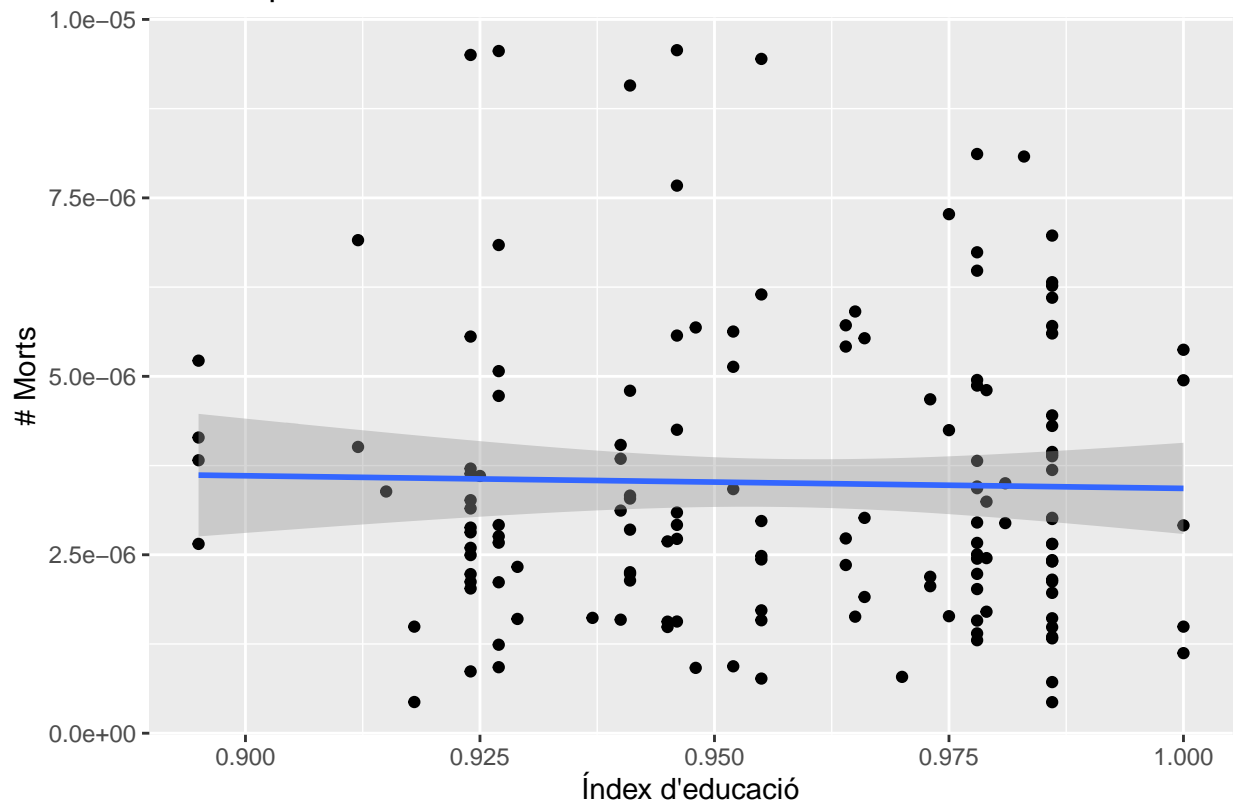
Aquest és el gràfic sense tenir en compte els anys:

```

Estadistica_dataset%>%ggplot(aes(x=income_index, y=ratio_deads))+geom_point()+geom_smooth(method='lm')+
## `geom_smooth()` using formula 'y ~ x'

```

Scatter plots de la relació entre morts i índex educacional en els diferents

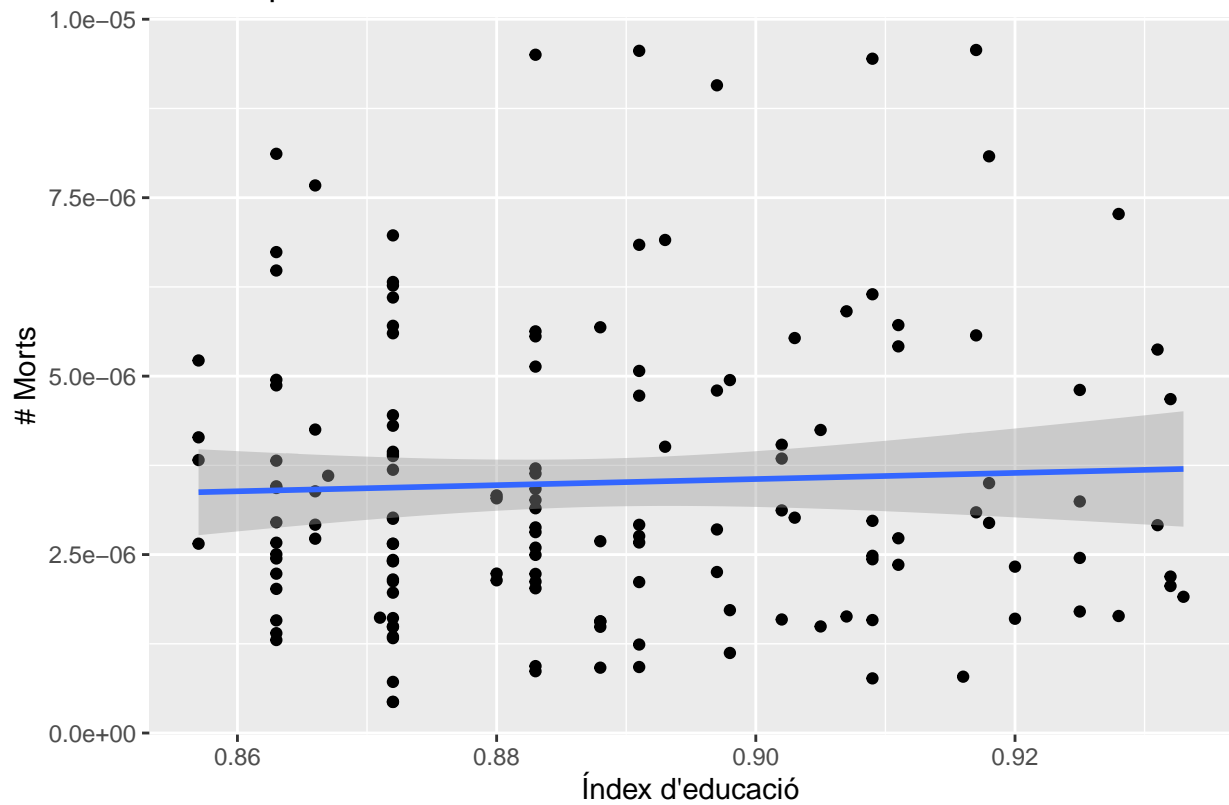


```

Estadistica_dataset%>%ggplot(aes(x=educational_index, y=ratio_deads))+geom_point()+geom_smooth(method='lm')+
## `geom_smooth()` using formula 'y ~ x'

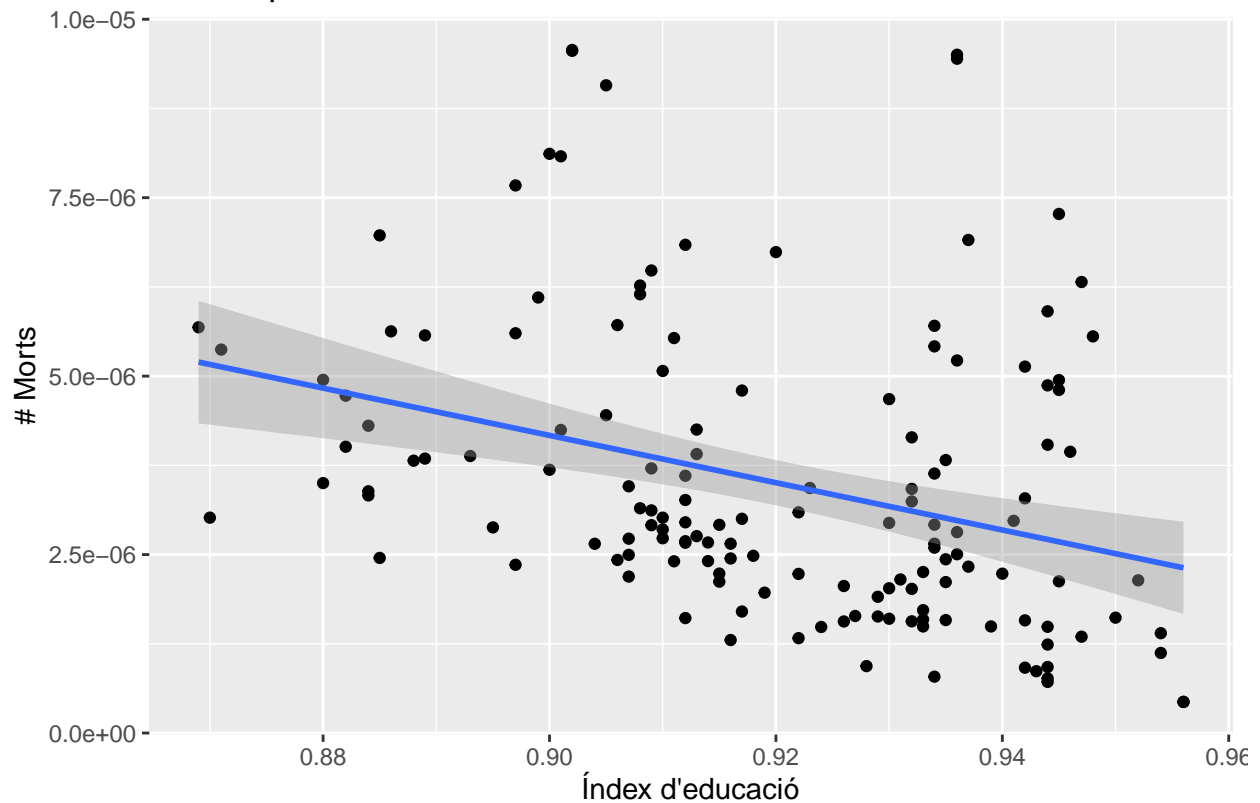
```

Scatter plots de la relació entre morts i index educacional en els diferents països



```
Estadistica_dataset%>%ggplot(aes(x=hdi, y=ratio_deads))+geom_point()+geom_smooth(method='lm')+ggtitle("Scatter plot de la relació entre morts i index educacional en els diferents països")
## `geom_smooth()` using formula 'y ~ x'
```

Scatter plots de la relació entre morts i index educacional en els diferents



Ara ho farem de forma estadística: Obtenim 3 variables explicatives i una mostra de 147, estudiarem exhaustivament el model (supòsits del model, validació global, validació individual, i residuals en el cas que el model sigui significatiu) i crearem les diferents gràfiques.

Plantejament de l'exercici: 1. Per a la validació del model s'utilitzarà: $H_0: \beta(0) = \beta(1)$ $H_1: \beta(0) \neq \beta(1), i \neq j$

Aquesta hipòtesi s'utilitzarà durant tot el treball estadístic. I sempre la nostra $\alpha = 0.05$

A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts:

```
MR_Global = lm(ratio_deads ~ income_index, data = Estadistica_dataset)
summary(MR_Global)
```

```
##
## Call:
## lm(formula = ratio_deads ~ income_index, data = Estadistica_dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.138e-06 -1.431e-06 -5.199e-07  1.232e-06  6.044e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.199e-06  6.154e-06   0.845   0.400
## income_index -1.769e-06  6.427e-06  -0.275   0.784
##
## Residual standard error: 2.062e-06 on 145 degrees of freedom
## Multiple R-squared:  0.000522, Adjusted R-squared: -0.006371
## F-statistic: 0.07573 on 1 and 145 DF, p-value: 0.7836
```

Obtenim que $Y_t = 5.199e-06 - (1.769e-06)income_index$

En augmentar la ràtio de morts de l'estat en una unitat, l'`income_index` disminueix un $1.769e-06$. Suposarem que $\alpha = 0.05$. Com que p-value de l'`income_index` és $0.7836 > \alpha$, la variable `income_index` no és significativa per al model. Per tant, el model no és estadísticament significatiu. Rebutgem H_0 .

```
MR_Global= lm(ratio_deads~educational_index, data= Estadistica_dataset)
summary(MR_Global)
```

```
##
## Call:
## lm(formula = ratio_deads ~ educational_index, data = Estadistica_dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.003e-06 -1.392e-06 -6.026e-07  1.174e-06  6.038e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.954e-07  7.302e-06  -0.040    0.968
## educational_index  4.282e-06  8.224e-06   0.521    0.603
##
## Residual standard error: 2.06e-06 on 145 degrees of freedom
## Multiple R-squared:  0.001866, Adjusted R-squared:  -0.005018
## F-statistic: 0.2711 on 1 and 145 DF, p-value: 0.6034
```

Obtenim que $Y_t = -2.954e-07 + (4.282e-06)educational_index$

En augmentar la ràtio de morts de l'estat en una unitat, l'`educational_index` augmenta un $4.282e-06$. Suposarem que $\alpha = 0.05$. Com que p-value de l'`income_index` és $0.6034 > \alpha$, la variable `educational_index` no és significativa per al model. Per tant, el model no és estadísticament significatiu. Rebutgem H_0 .

```
MR_Global= lm(ratio_deads~hdi, data= Estadistica_dataset)
summary(MR_Global)
```

```
##
## Call:
## lm(formula = ratio_deads ~ hdi, data = Estadistica_dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.336e-06 -1.363e-06 -6.890e-07  1.014e-06  6.526e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.399e-05  7.280e-06   4.668 6.86e-06 ***
## hdi         -3.313e-05  7.911e-06  -4.188 4.87e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.948e-06 on 145 degrees of freedom
## Multiple R-squared:  0.1079, Adjusted R-squared:  0.1017
## F-statistic: 17.54 on 1 and 145 DF, p-value: 4.87e-05
```

Obtenim que $Y_t = 3.399e-05 - (3.313e-05)hdi$

En augmentar la ràtio de morts de l'estat en una unitat, l'`educational_index` augmenta un $4.282e-06$. Suposarem que $\alpha = 0.05$. Com que p-value de l'`income_index` és $4.87e-05 < \alpha$, la variable `educational`

index Sí que és significativa per al model. Per tant, el model és estadísticament significatiu. Acceptem H0.

A partir d'aquí, podrem observar la bondat d'ajust. La bondat d'ajust (R), la obtenim a la sortida anterior del summary del model com a Multiple R-squared= 0.1079 El model per tant veiem que és capaç d'explicar el 10.79% de la ràtio de morts dels estats respecte a les variables hdi que s'ha acabat escollint en el darrer model. Fent l'arrel de R obtenim el coeficient de correlació de Pearson (r), que dóna informació del grau d'associació entre les variables explicatives amb la variable explicada, en aquest cas de 0.3284. Per tant, totes dues variables tenen una associació de 32.84%. Observem que no és gaire bona associació, ja que és un valor baix.

RELACIÓ tenint en compte els diferents anys

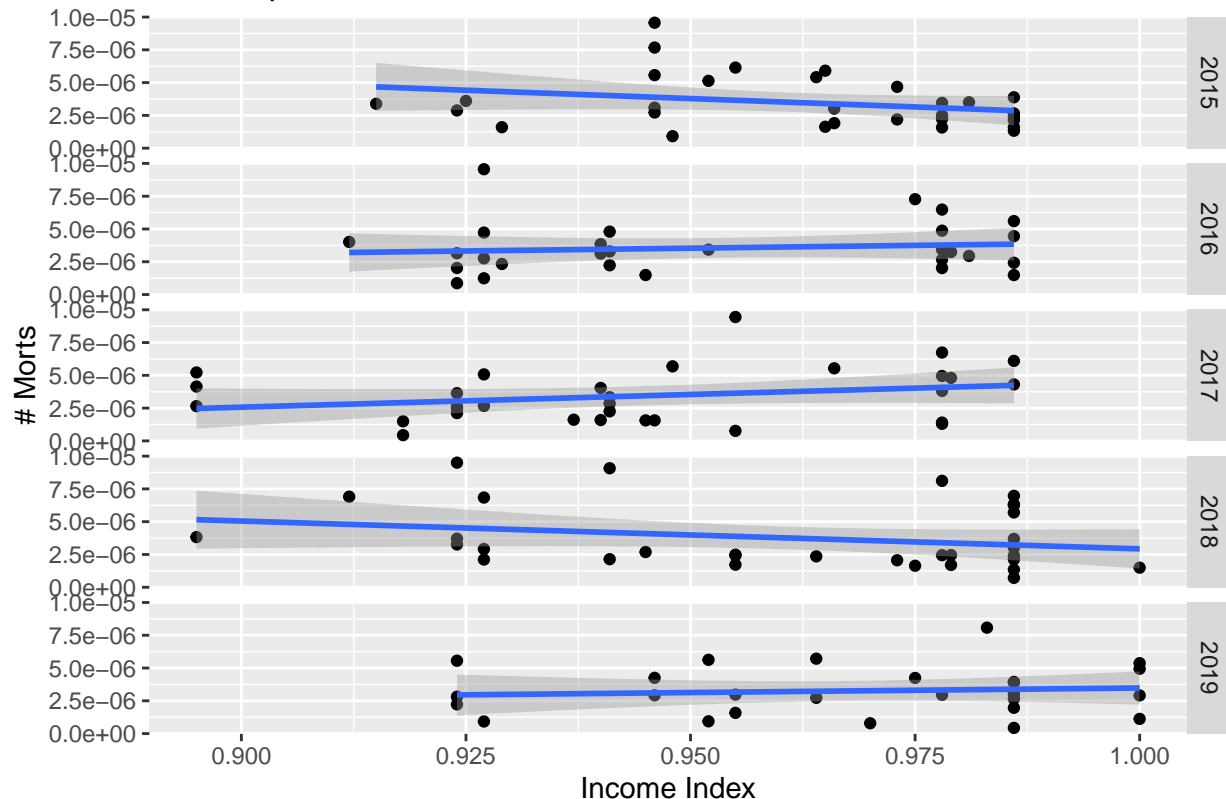
Aquests són els gràfics amb recta les rectes de regressió tenint en compte els anys:

```
Estadistica_dataset%>%ggplot(aes(x=income_index, y=ratio_deads))+geom_point()+geom_smooth(method='lm')+

```

```
## `geom_smooth()` using formula 'y ~ x'
```

Scatter plots de la relació entre morts i income index en els diferents es:

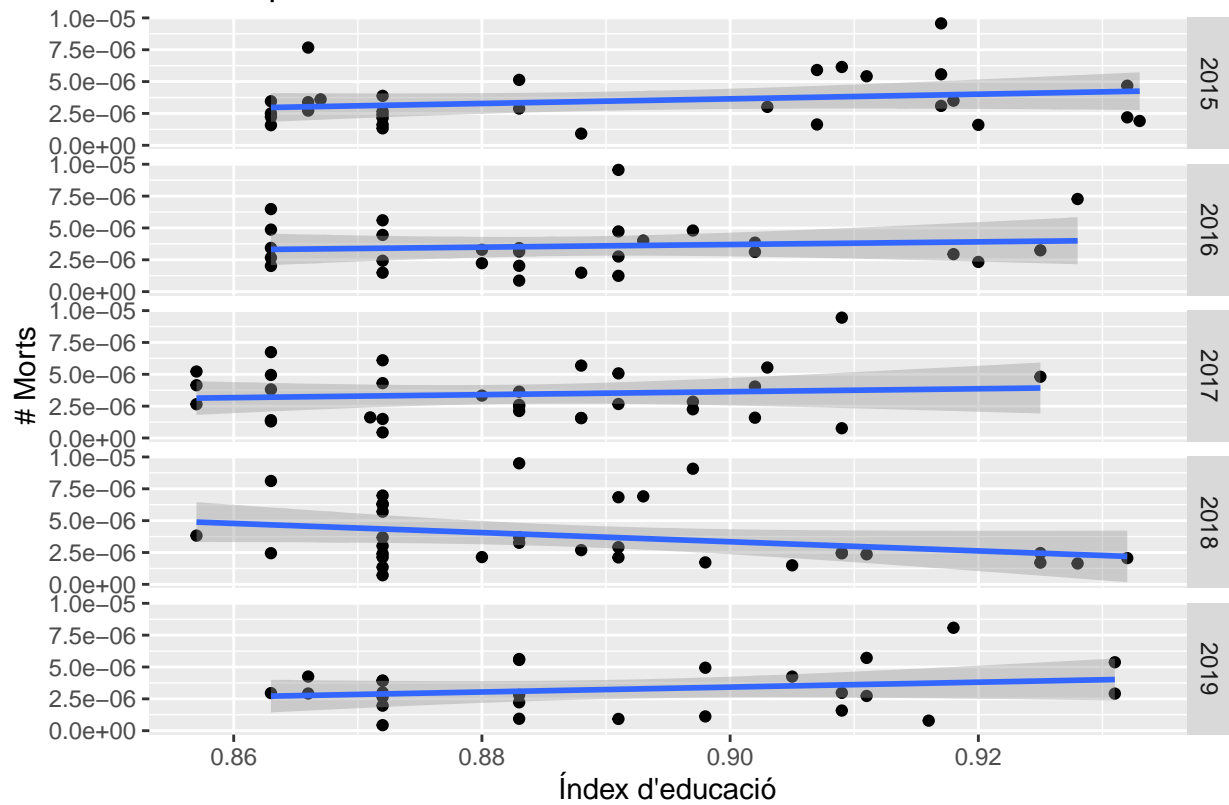


```
Estadistica_dataset%>%ggplot(aes(x=educational_index, y=ratio_deads))+geom_point()+geom_smooth(method='lm')+

```

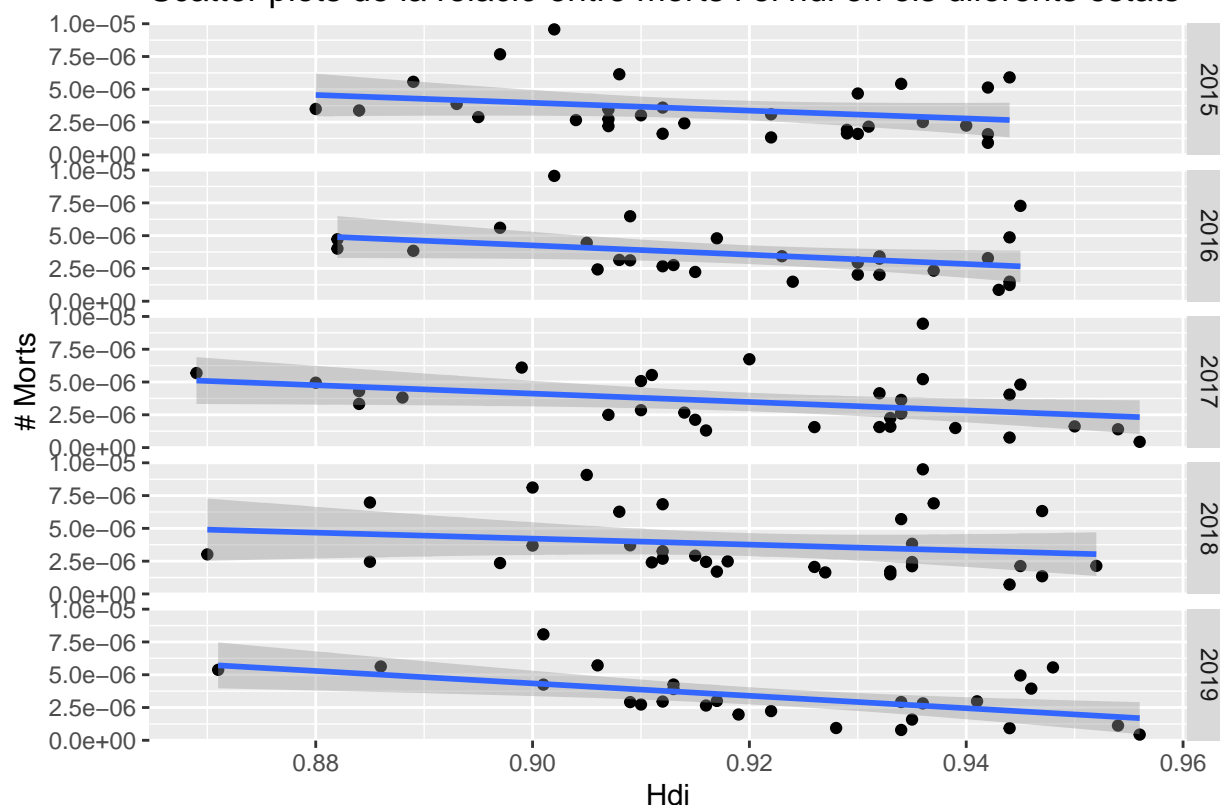
```
## `geom_smooth()` using formula 'y ~ x'
```

Scatter plots de la relació entre morts i index educacional en els diferents



```
Estadistica_dataset%>%ggplot(aes(x=hdi, y=ratio_deads))+geom_point()+geom_smooth(method='lm')+ggtitle("
## `geom_smooth()` using formula 'y ~ x'
```

Scatter plots de la relació entre morts i el hdi en els diferents estats



Any 2015

Income Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts com hem fet anteriorment amb el model global:

```
MR_2015_incomeIndex = lm(ratio_deads~income_index, data= Estadistica_2015)
summary(MR_2015_incomeIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ income_index, data = Estadistica_2015)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.919e-06 -1.281e-06 -6.398e-07  1.304e-06  5.683e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.800e-05  1.670e-05   1.677   0.105
## income_index -2.549e-05  1.735e-05  -1.469   0.153
##
## Residual standard error: 1.969e-06 on 28 degrees of freedom
## Multiple R-squared:  0.07156,    Adjusted R-squared:  0.0384
## F-statistic: 2.158 on 1 and 28 DF,  p-value: 0.153
```

Obtenim que $Y_t = 2.800e-05 - 2.549e-05 \text{ income_index}$

Observem que tant per inferència global del model el p-value > 0,05, per això es rebutja H_0 . Així, podem

concloure que el model no és estadísticament significatiu.

Educational Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts per a la variable Index en educació dels estats:

```
MR_2015_educationalIndex = lm(ratio_deads~educational_index, data= Estadistica_2015)
summary(MR_2015_educationalIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ educational_index, data = Estadistica_2015)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.510e-06 -1.289e-06 -4.686e-07  7.003e-07  5.613e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.279e-05  1.322e-05  -0.967    0.342
## educational_index  1.826e-05  1.484e-05   1.230    0.229
##
## Residual standard error: 1.99e-06 on 28 degrees of freedom
## Multiple R-squared:  0.05129,    Adjusted R-squared:  0.01741
## F-statistic: 1.514 on 1 and 28 DF,  p-value: 0.2288
```

Obtenim que $Y_t = -1.279e-05 + (1.826e-05)\text{educational_index}$

Observem que tant per inferència global del model el p-value > 0,05, per això es rebutja H_0 . Així podem concloure que el model tampoc no és estadísticament significatiu.

Hdi Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts:

```
MR_2015_hdiIndex = lm(ratio_deads~hdi, data= Estadistica_2015)
summary(MR_2015_hdiIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ hdi, data = Estadistica_2015)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.001e-06 -1.198e-06 -7.740e-07  9.524e-07  5.659e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.084e-05  1.778e-05   1.734  0.0938 .
## hdi         -2.986e-05  1.940e-05  -1.539  0.1350
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.962e-06 on 28 degrees of freedom
## Multiple R-squared:  0.078,    Adjusted R-squared:  0.04507
## F-statistic: 2.369 on 1 and 28 DF,  p-value: 0.135
```

Obtenim que $Y_t = 3.084e-05 - (2.986e-05)\text{hdi}$

Observem que per inferència global del model el p-value $> 0,05$, per això es rebutja H_0 . Així, podem concloure que el model no és estadísticament significatiu.

Hem observat que l'any 2015 cap d'aquests 3 factors referents a la població no influeixen en la ràtio de les morts als estats.

Any 2016

Income Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts:

```
MR_2016_incomeIndex = lm(ratio_deads~income_index, data= Estadistica_2016)
summary(MR_2016_incomeIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ income_index, data = Estadistica_2016)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.446e-06 -1.239e-06 -3.361e-07  8.754e-07  6.220e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.619e-06  1.402e-05  -0.329    0.744
## income_index  8.583e-06  1.470e-05   0.584    0.564
##
## Residual standard error: 1.959e-06 on 26 degrees of freedom
## Multiple R-squared:  0.01294,    Adjusted R-squared:  -0.02503
## F-statistic: 0.3407 on 1 and 26 DF,  p-value: 0.5644
```

Obtenim que $Y_t = -4.619e-06 + (8.583e-06)income_index$

S'observa que el p-value (0,5644) del model és més gran que 0,05, la nostra alpha. Per això, podem concloure que el model no és estadísticament significatiu i que la variable no influeix en la ràtio de morts a l'estat.

Educational Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts:

```
MR_2016_educationalIndex = lm(ratio_deads~educational_index, data= Estadistica_2016)
summary(MR_2016_educationalIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ educational_index, data = Estadistica_2016)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.657e-06 -1.267e-06 -4.872e-07  1.064e-06  5.952e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5.675e-06  1.760e-05  -0.322    0.750
## educational_index  1.042e-05  1.984e-05   0.525    0.604
##
## Residual standard error: 1.961e-06 on 26 degrees of freedom
```

```
## Multiple R-squared:  0.01049,    Adjusted R-squared:  -0.02757
## F-statistic: 0.2756 on 1 and 26 DF,  p-value: 0.6041
```

Obtenim que $Y_t = -5.675e-06 + (-5.675e-06)\text{educational_index}$

S'observa que el p-value del model és més gran que 0,05, la nostra alpha. Per això, podem concloure que el model no és estadísticament significatiu i que la variable no influeix en la ràtio de morts a l'estat.

Hdi Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts per a l'índex hdi:

```
MR_2016_hdiIndex = lm(ratio_deads~hdi, data= Estadistica_2016)
summary(MR_2016_hdiIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ hdi, data = Estadistica_2016)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.920e-06 -1.165e-06 -7.078e-07  4.102e-07  5.372e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.618e-05  1.700e-05   2.128  0.0429 *
## hdi         -3.548e-05  1.848e-05  -1.919  0.0660 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.845e-06 on 26 degrees of freedom
## Multiple R-squared:  0.1241, Adjusted R-squared:  0.09041
## F-statistic: 3.684 on 1 and 26 DF,  p-value: 0.06599
```

Obtenemos que $Y_t = 3.618e-05 - (3.548e-05)\text{hdi}$

Se observa que el p-value del modelo es más grande que 0.05, nuestra alpha. Por ello, podemos concluir que el modelo no es estadísticamente significativo y que la variable no influye en el ratio de muertes en el estado.

Año 2017

Income Index A continuación, crearemos el modelo de regresión múltiple y analizaremos los resultados obtenidos:

```
MR_2017_incomeIndex = lm(ratio_deads~income_index, data= Estadistica_2017)
summary(MR_2017_incomeIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ income_index, data = Estadistica_2017)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.871e-06 -1.549e-06 -2.673e-07  1.267e-06  5.810e-06
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.488e-05  1.268e-05  -1.173    0.25
## income_index  1.938e-05  1.343e-05   1.443    0.16
##
## Residual standard error: 1.997e-06 on 29 degrees of freedom
## Multiple R-squared:  0.06703,    Adjusted R-squared:  0.03486
## F-statistic: 2.083 on 1 and 29 DF,  p-value: 0.1596
```

Obtenim que $Y_t = -1.488e-05 - (1.938e-05)income_index$

S'observa que el p-value del model és més gran que 0,05, la nostra alpha. Per això, podem concloure que el model no és estadísticament significatiu i que la variable no influeix en la ràtio de morts a l'estat.

Educational Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts:

```
MR_2017_educationalIndex = lm(ratio_deads~educational_index, data= Estadistica_2017)
summary(MR_2017_educationalIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ educational_index, data = Estadistica_2017)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.972e-06 -1.742e-06 -4.810e-07  1.277e-06  5.710e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.820e-06  1.865e-05  -0.366    0.717
## educational_index  1.161e-05  2.113e-05   0.550    0.587
##
## Residual standard error: 2.057e-06 on 29 degrees of freedom
## Multiple R-squared:  0.01031,    Adjusted R-squared:  -0.02382
## F-statistic: 0.302 on 1 and 29 DF,  p-value: 0.5868
```

Obtenim que $Y_t = -6.820e-06 + (1.161e-05)educational_index$

S'observa que el p-value del model és més gran que 0,05, la nostra alpha. Per això, podem concloure que el model no és estadísticament significatiu i que la variable no influeix en la ràtio de morts a l'estat.

Hdi Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts per a l'índex hdi:

```
MR_2017_hdiIndex = lm(ratio_deads~hdi, data= Estadistica_2017)
summary(MR_2017_hdiIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ hdi, data = Estadistica_2017)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.301e-06 -1.386e-06 -6.865e-07  1.163e-06  6.483e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
##
```

```
## (Intercept)  3.297e-05  1.406e-05   2.345   0.0261 *
## hdi         -3.205e-05  1.525e-05  -2.102   0.0443 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.926e-06 on 29 degrees of freedom
## Multiple R-squared:  0.1322, Adjusted R-squared:  0.1023
## F-statistic: 4.419 on 1 and 29 DF,  p-value: 0.04434
```

Obtenim que $Y_t = 3.297e-05 - (3.205e-05)hdi$

S'observa que el p-value del model és més petit que 0,05, la nostra alpha. Per això, podem concloure que el model Sí és estadísticament significatiu i que l'hdi de l'estat el 2017 influeix en la ràtio de morts.

A partir d'aquí, podrem observar la bondat d'ajust. La bondat d'ajust (R), la obtenim a la sortida anterior del summary del model com a Multiple R-squared= 0.1322 El model, per tant, veiem que és capaç d'explicar el 13.22% de la ràtio de morts dels estats respecte a les variables hdi en aquest model. Fent l'arrel de R obtenim el coeficient de correlació de Pearson (r), que dona informació del grau d'associació entre les variables explicatives amb la variable explicada, en aquest cas de 0.3635. Per tant, totes dues variables tenen una associació de 36.35%. Observem que tampoc no és una associació molt bona entre variables.

Año 2018

Income Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts per a l'any 2018 a les 3 variables que tenim:

```
MR_2018_incomeIndex = lm(ratio_deads~income_index, data= Estadistica_2018)
summary(MR_2018_incomeIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ income_index, data = Estadistica_2018)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.508e-06 -1.476e-06 -1.187e-06  2.176e-06  4.963e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.415e-05  1.463e-05   1.651   0.109
## income_index -2.122e-05  1.522e-05  -1.394   0.174
##
## Residual standard error: 2.399e-06 on 30 degrees of freedom
## Multiple R-squared:  0.06084,    Adjusted R-squared:  0.02954
## F-statistic: 1.944 on 1 and 30 DF,  p-value: 0.1735
```

Obtenim que $Y_t = 2.415e-05 - (2.122e-05)income_index$

S'observa que el p-value del model és més gran que 0,05, la nostra alpha. Per això, podem concloure que el model no és estadísticament significatiu i que la variable no influeix en la ràtio de morts a l'estat.

Educational Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts:

```
MR_2018_educationalIndex = lm(ratio_deads~educational_index, data= Estadistica_2018)
summary(MR_2018_educationalIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ educational_index, data = Estadistica_2018)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.633e-06 -1.581e-06 -6.749e-07  1.497e-06  5.624e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.568e-05  1.830e-05   1.950  0.0606 .
## educational_index -3.593e-05  2.060e-05  -1.744  0.0913 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.359e-06 on 30 degrees of freedom
## Multiple R-squared:  0.09209,    Adjusted R-squared:  0.06183
## F-statistic: 3.043 on 1 and 30 DF,  p-value: 0.09132
```

Obtenim que $Y_t = 3.568e-05 - (3.593e-05)\text{educational_index}$

S'observa que el p-value del model és més petit que 0,05, la nostra alpha. Per això, podem concloure que el model no és estadísticament significatiu i que la variable no influeix en la ràtio de morts a l'estat.

Hdi Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts per a l'índex hdi:

```
MR_2018_hdiIndex = lm(ratio_deads~hdi, data= Estadistica_2018)
summary(MR_2018_hdiIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ hdi, data = Estadistica_2018)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.499e-06 -1.758e-06 -1.027e-06  2.240e-06  6.105e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.473e-05  1.991e-05   1.242  0.224
## hdi         -2.279e-05  2.164e-05  -1.053  0.301
##
## Residual standard error: 2.431e-06 on 30 degrees of freedom
## Multiple R-squared:  0.03566,    Adjusted R-squared:  0.003517
## F-statistic: 1.109 on 1 and 30 DF,  p-value: 0.3006
```

Obtenim que $Y_t = 2.473e-05 - (2.279e-05)\text{hdi}$

S'observa que el p-value del model és més gran que 0,05, la nostra alpha. Per això, podem concloure que el model torna a no ser estadísticament significatiu i que la variable hdi l'any 2018 no influeix en la ràtio de morts als estats.

Any 2019

Income Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts:

```
MR_2019_incomeIndex = lm(ratio_deads~income_index, data= Estadistica_2019)
summary(MR_2019_incomeIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ income_index, data = Estadistica_2019)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.947e-06 -1.244e-06 -2.838e-07  1.096e-06  4.718e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -3.483e-06  1.459e-05  -0.239    0.813
## income_index   6.963e-06  1.507e-05   0.462    0.648
##
## Residual standard error: 1.892e-06 on 24 degrees of freedom
## Multiple R-squared:  0.008811,    Adjusted R-squared:  -0.03249
## F-statistic: 0.2134 on 1 and 24 DF,  p-value: 0.6483
```

Obtenim que $Y_t = -3.483e-06 + (6.963e-06)income_index$

S'observa que el p-value del model és més gran que 0,05, la nostra alpha. Per això, podem concloure que el model no és estadísticament significatiu i que la variable no influeix en la ràtio de morts a l'estat.

Educational Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts:

```
MR_2019_educationalIndex = lm(ratio_deads~educational_index, data= Estadistica_2019)
summary(MR_2019_educationalIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ educational_index, data = Estadistica_2019)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.941e-06 -1.059e-06 -5.870e-08  1.279e-06  4.311e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.385e-05  1.591e-05  -0.871    0.392
## educational_index  1.920e-05  1.785e-05   1.076    0.293
##
## Residual standard error: 1.856e-06 on 24 degrees of freedom
## Multiple R-squared:  0.04599,    Adjusted R-squared:  0.006241
## F-statistic: 1.157 on 1 and 24 DF,  p-value: 0.2928
```

Obtenim que $Y_t = -1.385e-05 + (1.920e-05)educational_index$

S'observa que el p-value del model és més petit que 0,05, la nostra alpha. Per això, podem concloure que el model no és estadísticament significatiu i que la variable no influeix en la ràtio de morts a l'estat.

Hdi Index A continuació, crearem el model de regressió múltiple i analitzarem els resultats obtinguts per a l'índex hdi:

```
MR_2019_hdiIndex = lm(ratio_deads~hdi, data= Estadistica_2019)
summary(MR_2019_hdiIndex)
```

```
##
## Call:
## lm(formula = ratio_deads ~ hdi, data = Estadistica_2019)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.077e-06 -1.093e-06 -4.380e-07  5.621e-07  3.786e-06
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.697e-05  1.401e-05   3.352  0.00265 **
## hdi         -4.736e-05  1.518e-05  -3.121  0.00465 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.603e-06 on 24 degrees of freedom
## Multiple R-squared:  0.2887, Adjusted R-squared:  0.259
## F-statistic:  9.74 on 1 and 24 DF,  p-value: 0.004649
```

Obtenim que $Y_t = 4.697e-05 - (4.736e-05)hdi$

S'observa que el p-value del model és més petit que 0,05, la nostra alpha. Per això, podem concloure que el model Sí que és estadísticament significatiu i que la variable hdi sí que influeix en la ràtio de morts als estats el 2017.

A partir d'aquí, com que el model és estadísticament significatiu, podrem observar la bondat d'ajust. La bondat d'ajust (R), la obtenim a la sortida anterior del summary del model com a Multiple R-squared= 0.2887 El model, per tant, veiem que és capaç d'explicar el 28.87% de la ràtio de morts dels estats respecte a les variables hdi en aquest model. Fent l'arrel de R obtenim el coeficient de correlació de Pearson (r), que dóna informació del grau d'associació entre les variables explicatives amb la variable explicada, en aquest cas de 0.5373. Per tant, totes dues variables tenen una associació de 53.73%. Observem que tampoc és una associació molt bona entre variables, però sí que és més gran que la obtinguda l'any 2017.