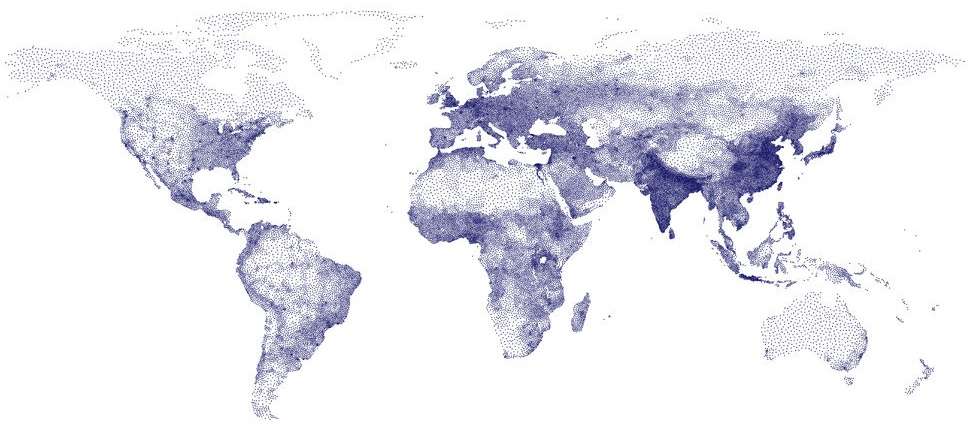


**ΤΑ ΕΘΝΗ ΤΟΥ ΚΟΣΜΟΥ ΤΟ 1981**



**ΟΝΟΜΑΤΕΠΩΝΥΜΟ : ΒΙΚΑΤΟΣ ΑΛΕΞΑΝΔΡΟΣ**

**ΑΜ : 6160011**

**ΤΜΗΜΑ : ΣΤΑΤΙΣΤΙΚΗΣ**

**ΜΑΘΗΜΑ : ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ**

**ΕΞΑΜΗΝΟ : ΕΑΡΙΝΟ 2023**

# ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΕΧΟΜΕΝΑ: ΣΕΛΙΔΑ 2

ΚΕΦΑΛΑΙΟ 1Ο ΕΙΣΑΓΩΓΗ : ΣΕΛΙΔΑ 3

ΚΕΦΑΛΑΙΟ 2Ο ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ : ΣΕΛΙΔΑ 3

ΚΕΦΑΛΑΙΟ 3Ο ΣΧΕΣΕΙΣ ΜΕΤΑΒΛΗΤΩΝ ΑΝΑ ΖΕΥΓH : ΣΕΛΙΔΑ 6

ΚΕΦΑΛΑΙΟ 4Ο ΠΡΟΒΛΕΠΤΙΚΑ & ΕΡΜΗΝΕΥΤΙΚΑ ΜΟΝΤΕΛΑ : ΣΕΛΙΔΑ 9

ΚΕΦΑΛΑΙΟ 5Ο ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ-ΒΙΒΛΙΟΓΡΑΦΙΑ : ΣΕΛΙΔΑ 12

ΠΑΡΑΡΤΗΜΑ-ΠΙΝΑΚΕΣ : ΣΕΛΙΔΑ 13

# 1.ΕΙΣΑΓΩΓΗ

Το 1981 στο Πανεπιστήμιο της Βόρειας Καρολίνα στις ΗΠΑ ο G. David Garson επιμελήθηκε μιας βάσης δεδομένων που αποτελείται από 116 μεταβλητές που η κάθε μεταβλητή παρουσιάζει και ένα χαρακτηριστικό του κόσμου μας. Στην παρακάτω εργασία θα χρησιμοποιήσουμε μερικές από τις μεταβλητές αυτές (συν μία κατηγορική μεταβλητή που φτιάχτηκε για να κατηγοριοποιήσει τα ποσοστά αστικοποίησης σε τέσσερα επίπεδα) για να προσπαθήσουμε να ερευνήσουμε διάφορα φαινόμενα στον κόσμο μας. Όπως την κατανάλωση του ηλεκτρικού ρεύματος συμφωνά με το ποσοστό αστικοποίησης κάθε χωράς , την σχέση της γυναίκας με την εκπαίδευση σε σχέση με τα πολιτικά και νομικά δικαιώματα της γυναίκας. Μετά από την μελέτη των μεταβλητών που θα δείτε αργότερα παρατηρήσαμε κάποιες ενδιαφέρουσες τιμές. Όπως ο δείκτης τιμών καταναλωτή(CPI) είχε κάποιες ακραίες τιμές κυρίως σε χώρες της Λατινικής Αμερικής. Αυτό οφείλεται στην οικονομική ύφεση του 1980 λόγω της ανόδου της τιμής του πετρελαίου και του πολέμου της Αργεντινής (1976-1983). Τέλος θα σας παρουσιάσω ένα μοντέλο το οποίο θα εκτιμά τον αναμενόμενο ρυθμό ανάπτυξης μιας χώρας βάση των παρακάτω μεταβλητών.

# 2.ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Το 01\_world81.sav είναι SPSS αρχείο ,οπότε θα χρειαστεί το foreign Package ,ώστε διαβαστούν τα δεδομένα του στην R. Βλέπω ότι το dataset μου έχει 10 μεταβλητές συν την νέα μεταβλητή NewPopurban η οποία έχει τέσσερα επίπεδα “Very Low”, “Low” , “Normal” και “High”.



**Πίνακας 1: Περιγραφικά Μέτρα ποσοτικών μεταβλητών**



Παρατηρώντας τα περιγραφικά μέτρα των ποσοτικών μεταβλητών βλέπω ότι ο μέσος πληθυσμός είναι 40448140. Ο μέσος όρος ανάπτυξης είναι 2.314, ενώ ο μέσος όρος θανάτων ανά χίλιους κατοίκους (Deathrate) είναι 13.2. Σε όλες τις μεταβλητές έχουμε ανομοιογένεια καθώς το CV>0.3.Επίσης σημαντικό να αναφέρουμε ότι η μεταβλητή CPI έχει επηρεαστεί από τις ακραίες τιμές που βρέθηκαν σε κάποιες χώρες και για αυτό βλέπουμε τόσο μεγάλο μέσο όρο. Επίσης για αυτόν το λόγο υπολογίστηκε το Trimmean του CPI για 10% του δείγματος 17581.84. Όπως βλέπουμε και στο παρακάτω διάγραμμα καμία από τις μεταβλητές δεν ακολουθεί κανονική κατανομή.

**Πίνακας 2 : Ιστογράμματα ποσοτικών μεταβλητών**

A picture containing text, diagram, screenshot, plot

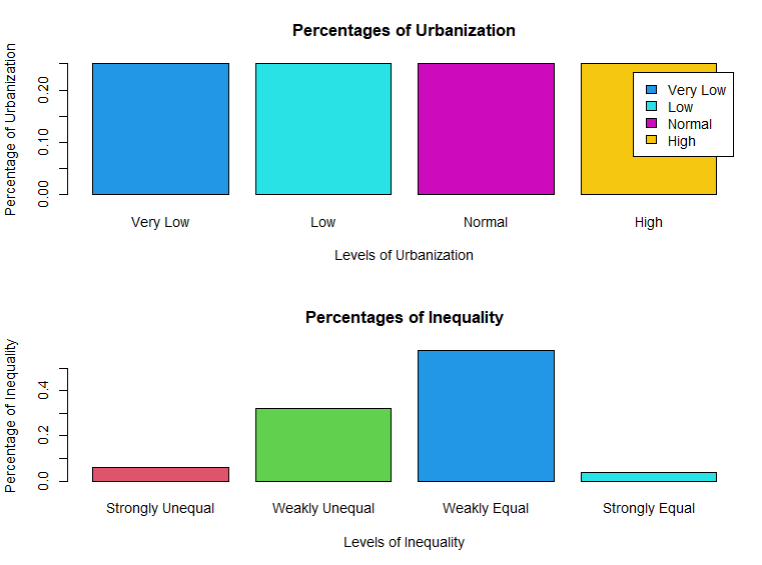
Description automatically generated

**Πίνακας 3 : Σχετικών Συχνοτήτων Κατηγορικών Μεταβλητών**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ΜΕΤΑΒΛΗΤΗ** | Strongly Equal | Weakly Unequal | Weakly Equal | Strongly Equal |
| Polegw | **6%** | **32%** | **58%** | **3%** |
| **ΜΕΤΑΒΛΗΤΗ** | Very Low | Low | Normal | High |
| NewPopurban | **25%** | **25%** | **25%** | **25%** |

Δημιουργήθηκε μία νέα μεταβλητή η NewPopurban η οποία έχει τέσσερα επίπεδα Very Low, Low, Normal και High τα οποία αφορούν το ποσοστό αστικοποίησης κάθε χώρας. Όπως διακρίνεται και στον πίνακα 3 δεν υπάρχουν διάφορες, καθώς οι τιμές αστικοποίησης είναι όλες 25%. Δεν ισχύει, όμως το ίδιο στην μεταβλητή Polegw που έχει να κάνει με πολιτική και νομική ισότητα των γυναικών, καθώς παρατηρείται ένα αυξημένο ποσοστό της τάξης του 32% που αντιπροσωπεύει την ανισότητα μεταξύ των δύο φύλων.

**Πίνακας 4**



# 3.ΣΧΕΣΕΙΣ ΜΕΤΑΒΛΗΤΩΝ ΑΝΑ ΖΕΥΓΗ

Στο προηγούμενο κεφάλαιο μελετήσαμε τις μεταβλητές μας ξεχωριστά υπολογίζοντας τα απαραίτητα μέτρα για την περιγραφική στατιστική ανάλυσή τους. Στο κεφάλαιο αυτό θα μελετήσουμε τις σχέσεις των μεταβλητών μας ανά ζεύγη . Αρχικά εξετάζουμε τις σχέσεις μεταξύ των ποσοτικών μεταβλητών ανά ζεύγη, καθώς μας ενδιαφέρει να διαπιστώσουμε αν υπάρχουν συσχετίσεις μεταξύ τους. Πρώτα πρέπει να εξεταστεί αν οι μεταβλητές ακολουθούν την κανονική κατανομή. Πραγματοποιώντας Shapiro Test για κάθε μεταβλητή καταλήγουμε στο συμπέρασμα ότι καμία μεταβλητή δεν ακολουθεί κανονική κατανομή οπότε για όλα τα ζεύγη θα γίνει μη-παραμετρικός έλεγχος Spearman.

**Πίνακας 5 : Συσχετίσεων Ποσοτικών Μεταβλητών**



Για τα ζεύγη των μεταβλητών που παρατηρείται p-value>5% δεν υπάρχει ένδειξη για συσχέτιση. Αντιθέτως για τα ζεύγη με p-value<5% απορρίπτεται Ηο ,οπότε υπάρχει ένδειξη συσχέτισης. Πιο συγκεκριμένα ενδιαφέρον παρουσιάζει η σχέση μεταξύ της μεταβλητής Electcon και Educexpc ,καθώς έχουμε **σημαντική συσχέτιση (Rho=0.90).** Τέλος η σχέση Educexpc και Popurban παρουσιάζει **σημαντική θετική συσχέτιση (Rho=0.80).**

Στη συνέχεια εξετάζεται αν η μέση κατανάλωση ηλεκτρικού ρεύματος διαφοροποιείται σε διαφορετικά επίπεδα αστικοποίησης (NewPopurban). Για τον παραπάνω έλεγχο πραγματοποιείται ανάλυση διακύμανσης. Τα κατάλοιπα της ανάλυσης διακύμανσης δεν κατανέμονται κανονικά και πραγματοποιείται το μη παραμετρικό τεστ Kruskal-Wallis και υπάρχει διαφορά στις διαμέσους (p-value<5%). Για να εντοπιστούν που προσδιορίζονται οι διαφορές πραγματοποιείται boxplot και παρατηρείται διαφορά διαμέσων στα επίπεδα High-Normal, High-Low, High-Very Low, Normal-Low και Normal-Very Low.

**Πίνακας 6: Boxplot Electcon με NewPopurban**

Εικόνα που περιέχει κείμενο, διάγραμμα, στιγμιότυπο οθόνης, ορθογώνιο παραλληλόγραμμο

Περιγραφή που δημιουργήθηκε αυτόματα

**Πίνακας 7: Boxplot(Educexpc~Polegw)**

Εικόνα που περιέχει κείμενο, διάγραμμα, στιγμιότυπο οθόνης, ορθογώνιο παραλληλόγραμμο

Περιγραφή που δημιουργήθηκε αυτόματα

Τέλος εξετάστηκε η σχέση μεταξύ των μεταβλητών Educexpc και Polegw. Πιο συγκεκριμένα έγινε έλεγχος διακύμανσης. Όπως και παραπάνω τα κατάλοιπα δεν κατανέμονται κανονικά, οπότε έγινε το μη-παραμετρικό τεστ Kruskal-Wallis και παρατηρήθηκε διαφορά στις διαμέσους (P-value<0.05). Από το αντίστοιχο Boxplot για τις 2 μεταβλητές παρατηρείται διαφορά διαμέσων και σε όλα τα επίπεδα εκτός του Strongly Equal-Weakly Equal.

# ΠΡΟΒΛΕΠΤΙΚΑ & ΕΡΜΗΝΕΥΤΙΚΑ ΜΟΝΤΕΛΑ

Στο κεφάλαιο αυτό εξετάζεται το συνολικό μοντέλο που ερμηνεύει την μεταβλητή Growth.

**Growth=Bo+B1\*population+B2\*Deathrate+B3\*Popurban+B4\*Saving+B5\*Electcon+B6\*Educexpc+B7\*CPI+B8\*Polegw+ε**

Το μοντέλο ικανοποίει όλες τις συνθήκες **Κανονικότητα , γραμμικότητα** και **ομοσκεδαστικότητα.** Ο έλεγχος για ανεξαρτησία των κατάλοιπων δεν χρειάστηκε ,καθώς τα data είναι διαστρωματικά .Επίσης στον έλεγχο πολυσυγγραμικότητας σχεδόν όλες οι μεταβλητές είχαν VIF κοντά στο 1 , ενώ οι μεταβλητές Electcon=3.69 και Educexpc=4.65, άρα δεν υπάρχει πρόβλημα πολυσυγγραμμικότητας. Το μοντέλο διαμορφώνεται ως εξής:



Multiple R-squared: 0.3041, Adjusted R-squared: 0.1776 F-statistic: 2.404 on 10 and 55 DF, p-value: 0.01894 Το μοντέλο αυτό είναι στατιστικά σημαντικό με F-statistic=2.404 και το ποσοστό της μεταβλητότητας της Growth(ανάπτυξης) που ερμηνεύεται από τις υπόλοιπες μεταβλητές 30.4%. Οι στατιστικά σημαντικές μεταβλητές είναι το saving ,deathrate και η κατηγορική μεταβλητή Polegw.

**ΔΙΑΓΡΑΜΜΑ ΓΡΑΜΜΙΚΟΤΗΤΑΣ ΜΟΝΤΕΛΟΥ**

Εικόνα που περιέχει διάγραμμα, κείμενο, τεχνικό σχέδιο, Σχέδιο

Περιγραφή που δημιουργήθηκε αυτόματα

**Πίνακας 9 : Συνθήκες Μοντέλου**

**Α)Histogram B)Έλεγχος Κανονικότητας Γ)Ομοσκεδαστικότητα**

Εικόνα που περιέχει κείμενο, διάγραμμα, γραμμή, στιγμιότυπο οθόνης

Περιγραφή που δημιουργήθηκε αυτόματα

Όπως φαίνεται και στα παρακάτω διαγράμματα Cooks Distance και Διάγραμμα μόχλευσης παρατηρείται ότι η παρατήρηση 41 είναι σημείο επιρροής για το μοντέλο. Για αυτόν το λόγο αφαιρέθηκε η παρατήρηση αυτή, ώστε να εξεταστεί η επίδραση της. Το μοντέλο που προκύπτει είναι έγκυρο ,καθώς πληρούνται όλες οι προϋποθέσεις(κανονικότητα, ομοσκεδαστικότητα και γραμμικότητα)

**Πίνακας 9 : Συνθήκες νέου Μοντέλου και Τιμή Επιρροής**

**Α)Cooks Distance B)Διάγραμμα Μόχλευσης Γ)Κανονικότητα Δ)Ομοσκεδαστικότητα**

Εικόνα που περιέχει κείμενο, διάγραμμα, γραμμή, παράλληλα

Περιγραφή που δημιουργήθηκε αυτόματα

**ΔΙΑΓΡΑΜΜΑ ΓΡΑΜΜΙΚΟΤΗΤΑΣ NEOY ΜΟΝΤΕΛΟΥ**

Εικόνα που περιέχει κείμενο, διάγραμμα, παράλληλα, Σχέδιο

Περιγραφή που δημιουργήθηκε αυτόματα

Άρα μετά την αφαίρεση της παρατήρησης 41 το μοντέλο που προκύπτει είναι το παρακάτω:



F-statistic: 2.772 on 10 and 54 DF, p-value: 0.00774

Το μοντέλο αυτό είναι στατιστικά σημαντικό με F-statistic=2.772 και το ποσοστό της μεταβλητότητας της Growth(ανάπτυξης) που ερμηνεύεται από τις υπόλοιπες μεταβλητές 33.9%. Οι στατιστικά σημαντικές μεταβλητές είναι το saving ,deathrate και η κατηγορική μεταβλητή Polegw. Μετά από την σύγκριση των δύο μοντέλων παρατηρείτε ότι η μεταβλητή Electcon στο νέο μοντέλο έχει αρνητική επίδραση στην τελική διαμόρφωση της Growth , ενώ στο αρχικό μοντέλο θετική. Επίσης μετά από σύγκριση των διαγραμμάτων γραμμικότητας φαίνεται ότι το νέο μοντέλο ικανοποιεί σε μεγαλύτερο βαθμό την συνθήκη αυτή.

# ΣΥΜΠΕΡΑΣΜΑΤΑ

Συμπεραίνεται λοιπόν , ότι μετά τους κατάλληλους ελέγχους φαίνεται να υπάρχει διαφορά στη μέση κατανάλωση του ηλεκτρικού ρεύματος ανάλογα του ποσοστού αστικοποίησης ,καθώς παρατηρείται μεγαλύτερη κατανάλωση σε χώρες που έχουν μεγάλο ποσοστό αστικοποίησης . Επίσης υπάρχει διαφορά και στα έξοδα εκπαίδευσης βάση των δικαιωμάτων των γυναικών. Όπως φαίνεται και στο σχετικό κεφάλαιο υπάρχει διαφορά στην μεταβλητότητα των εξόδων. Πιο συγκεκριμένα στις χώρες που υπάρχει νομική και πολιτική ισότητα μεταξύ αντρών και γυναικών τα έξοδα εκπαίδευσης λαμβάνουν μεγαλύτερες τιμές από οποιαδήποτε άλλη κατηγορία. Τέλος το και τα δύο μοντέλα που παρουσιάστηκαν έχουν χαμηλούς συντελεστές προσδιορισμού, οπότε δεν μπορούν να χρησιμοποιηθούν για προγνώσεις, αλλά ταυτόχρονα προσφέρουν κάποιες χρήσιμες πληροφορίες για το πως οι δοθέντες μεταβλητές επηρεάζουν την ανάπτυξη.

# ΒΙΒΛΙΟΓΡΑΦΙΑ

* <https://www.ethnos.gr/history/article/212046/topiostygnokathestostoy20oyaionaexafanisebrefhpetaxeanthropoyssththalassaapoaeroplanaektelesebiasebasanisemakelepsetohsyxotelostoyarxhgoy>
* <https://www.euro2day.gr/specials/topics/article/1214834/h-istoria-ton-kriseon-hreoys.html>

# 5.ΠΑΡΑΡΤΗΜΑ

ΕΝΤΟΛΗ 1

install.packages("foreign")

library(foreign)

world <- read.spss(file.choose(),header=T,sep=",",add.undeclared.levels="no")

suppressWarnings(world)

world=as.data.frame(world)

world

head(world)

ΕΝΤΟΛΗ 2

world$polegw==-1

##antikathistw sthn metavliti polegw to -1 me NA

world[world$polegw==-1,]$polegw=NA

world$polegw=as.factor(world$polegw)

levels(world$polegw)=c("Strongly Unequal","Weakly Unequal","Weakly Equal","Strongly Equal")

head(world)

str(world)

ΕΝΤΟΛΗ 3

###ftiaxnw thn nea metavliti

NewPopurban=c()

NewPopurban[world$popurban<27]=1

NewPopurban[world$popurban>27 &world$popurban<=46]=2

NewPopurban[world$popurban>46 &world$popurban<=67]=3

NewPopurban[world$popurban>67]=4

world$NewPopurban=NewPopurban

head(world)

world$NewPopurban=as.factor(world$NewPopurban)

levels(world$NewPopurban)=c("Very Low","Low","Normal","High")

ΕΝΤΟΛΗ 4

###PERIGRAFIKI STATISTIKI

descriptives=function(x){

n=length(x)

mu=mean(x)

med=median(x)

q1=quantile(x,0.25)

q2=quantile(x,0.75)

sd=sd(x)

cv=sd(x)/mean(x)

kurtosis=sum((x-mu)^4)/(n\*sd^4)-3

symmetry=sum((x-mu)^3)/(n\*sd^3)

d=c(mu,med,q1,q2,sd,cv,kurtosis,symmetry)

names(d)=c("mesos oros","median","Q1","Q2","SD","CV","kurtosis","symmetry")

return(d)

}

na.omit(world)

descriptives(world$population)

descriptives(world$deathrate)

descriptives(world$popurban)

descriptives(world$growth)

descriptives(world$saving)

descriptives(world$electcon)

descriptives(world$cpi)

descriptives(world$educexpc)

ΕΝΤΟΛΗ 5

##trimmean gia to cpi giati paratirw polles akraies times

#A ten percent trimmed mean, for example represents the mean of a dataset

#after the ten percent smallest and greatest values have been eliminated.

mean(world$cpi, trim=0.1)

#A ten percent trimmed mean, for example, represents the mean of a dataset

#after the 5 percent smallest and greatest values have been eliminated.

mean(world$cpi,trim=0.05)

ΕΝΤΟΛΗ 5

##perigrafika metra katigorikwn

prop.table(table(world$polegw))

prop.table(table(world$NewPopurban))

t1=prop.table(table(world$polegw))

t2=prop.table(table(world$NewPopurban))

par(mfrow=c(2,1))

barplot(t2,col=4:7,main="Percentages of Urbanization",ylab="Percentage of Urbanization",

xlab="Levels of Urbanization",legend=names(t2),beside=TRUE)

barplot(t1,col = 10:13,main="Percentages of Inequality",ylab="Percentage of Inequality",

xlab="Levels of Inequality",legend=names(t1),beside=TRUE)

ΕΝΤΟΛΗ 6

par(mfrow=c(3,3))

**##hist for population**

hist(world$population,probability = T,main="Histogram for Population",border="white",col="green")

x<-world$population

mean<-mean(x)

sd<-sd(x)

curve(dnorm(x,mean,sd),col="violetred",add=T,lwd=2)

**##hist for deathrate**

hist(world$deathrate,probability = T,main="Histogram for Deathrate",border="white",col="red")

x=world$deathrate

mean<-mean(x)

sd<-sd(x)

curve(dnorm(x,mean,sd),col="black",add=T,lwd=2)

**##hist for popurban**

hist(world$popurban,ylim=c(0,0.02),probability=T,main="Histogramfor Popurban",border="white",col="blue")

x=world$popurban

mean<-mean(x)

sd<-sd(x)

curve(dnorm(x,mean,sd),col="red",add=T,lwd=2)

shapiro.test(world$popurban) ##PROSEGKIZEI KANONIKI

**##hist for Growth**

hist(world$growth,probability = T,main="Histogram for Growth",border="white",col="orange")

x=world$growth

mean<-mean(x)

sd<-sd(x)

curve(dnorm(x,mean,sd),col="grey",add=T,lwd=2)

**##hist for saving**

hist(world$saving,probability = T,main="Histogram for Saving",border="white",col="pink")

x=world$saving

mean<-mean(x)

sd<-sd(x)

curve(dnorm(x,mean,sd),col="green",add=T,lwd=2)

**##hist for Electcon**

hist(world$electcon,probability = T,main="Histogram for Electcon",border="white",col="brown")

x=world$electcon

mean<-mean(x)

sd<-sd(x)

curve(dnorm(x,mean,sd),col="green",add=T,lwd=2)

**##hist for CPI**

hist(world$cpi,probability = T,main="Histogram for CPI",border="white",col="cadetblue")

x=world$cpi

mean<-mean(x)

sd<-sd(x)

curve(dnorm(x,mean,sd),col="green",add=T,lwd=2)

ΕΝΤΟΛΗ 7

###ELEGXOI DIAKIMANSEWN ANOVA####

#NewPopurban~Population

anova1=aov(world$population~world$NewPopurban)

summary(anova1)

#p-value>5%

shapiro.test(anova1$res)

#p-value<2.2e-16 , w=0.451

mean(anova1$res)

median(anova1$res)

kruskal.test(world$population~world$NewPopurban)

pairwise.wilcox.test(world$population,world$NewPopurban,p.adjust.method = "bonf",paired=F)

tapply(world$population,world$NewPopurban,summary)

boxplot(world$population~world$NewPopurban)

##NewPopurban~Deathrate

anova2=aov(world$deathrate~world$NewPopurban)

summary(anova2)

shapiro.test(anova2$res)

#p-value>5%

mean(anova2$res)

median(anova2$res)

kruskal.test(world$deathrate~world$NewPopurban)

pairwise.wilcox.test(world$deathrate,world$NewPopurban,p.adjust.method = "bonf",paired=F)

tapply(world$deathrate,world$NewPopurban,summary)

boxplot(world$deathrate~world$NewPopurban)

##NewPopurban~Growth

anova3=aov(world$growth~world$NewPopurban)

summary(anova3)

shapiro.test(anova3$res)

mean(anova3$res)

median(anova3$res)

kruskal.test(world$growth~world$NewPopurban)

pairwise.wilcox.test(world$growth,world$NewPopurban,p.adjust.method = "bonf",paired=F)

tapply(world$growth,world$NewPopurban,summary)

boxplot(world$growth~world$NewPopurban)

##NewPopurban~saving

anova4=aov(world$saving~world$NewPopurban)

summary(anova4)

shapiro.test(anova4$res)

mean(anova4$res)

median(anova4$res)

kruskal.test(world$saving~world$NewPopurban)

pairwise.wilcox.test(world$saving,world$NewPopurban,p.adjust.method = "bonf",paired=F)

tapply(world$saving,world$NewPopurban,summary)

boxplot(world$saving~world$NewPopurban)

####WORD\_MATERIAL####

##NewPopurban~Electcon

anova5=aov(world$electcon~world$NewPopurban)

summary(anova5)

shapiro.test(anova5$res)

mean(anova5$res)

#O MESOS DEN EINAI KATALILO METRO

median(anova5$res)

**#O MESOS DEN EINAI KATALILO METRO AFOU EINAI KONTA STO 0 KAI H DIAMESOS -110 ARA DEN PROSEGGIZEI NORMALITY**

**#KANW KRUSKAL GIA NA DW TIS DIAMESOUS**

kruskal.test(world$electcon~world$NewPopurban)

#p-value=1.895e-15 chi-squared=71.64

TukeyHSD(anova5)

#gia ta 4 epipeda astikopoihshs vlepw oti uparxei diafora metaksu twn epipedwn

#Normal-Very Low , High-Very Low kai High-Low

pairwise.wilcox.test(world$electcon,world$NewPopurban,p.adjust.method = "bonf",paired=F)

tapply(world$electcon,world$NewPopurban,summary)

boxplot(world$electcon~world$NewPopurban,main="Electricity Consumption for Each Level of Urbanization", col=8:11,ylab="Electricity Consumption",xlab="Levels of Urbanization")

##NewPopurban~cpi

anova6=aov(world$cpi~world$NewPopurban)

summary(anova6)

shapiro.test(anova6$res)

mean(anova6$res)

median(anova6$res)

kruskal.test(world$cpi~world$NewPopurban)

pairwise.wilcox.test(world$cpi,world$NewPopurban,p.adjust.method = "bonf",paired=F)

tapply(world$cpi,world$NewPopurban,summary)

boxplot(world$cpi~world$NewPopurban)

##NewPopurban~educexpc

anova7=aov(world$educexpc~world$NewPopurban)

summary(anova7)

shapiro.test(anova7$res)

mean(anova7$res)

median(anova7$res)

kruskal.test(world$educexpc~world$NewPopurban)

pairwise.wilcox.test(world$educexpc,world$NewPopurban,p.adjust.method = "bonf",paired=F)

tapply(world$educexpc,world$NewPopurban,summary)

boxplot(world$educexpc~world$NewPopurban)

##Polegw~Population

anova8=aov(world$population~world$polegw)

summary(anova8)

shapiro.test(anova8$res)

mean(anova8$res)

median(anova8$res)

kruskal.test(world$population~world$polegw)

pairwise.wilcox.test(world$population,world$polegw,p.adjust.method = "bonf",paired=F)

tapply(world$population,world$polegw,summary)

boxplot(world$population~world$polegw)

##Polegw~deathrate

anova9=aov(world$deathrate~world$polegw)

summary(anova9)

mean(anova9$res)

median(anova9$res)

kruskal.test(world$deathrate~world$polegw)

pairwise.wilcox.test(world$deathrate,world$polegw,p.adjust.method = "bonf",paired=F)

tapply(world$deathrate,world$polegw,summary)

boxplot(world$deathrate~world$polegw)

##Polegw~popurban

anova10=aov(world$popurban~world$polegw)

summary(anova10)

mean(anova10$res)

median(anova10$res)

kruskal.test(world$popurban~world$polegw)

pairwise.wilcox.test(world$popurban,world$polegw,p.adjust.method = "bonf",paired=F)

tapply(world$popurban,world$polegw,summary)

boxplot(world$popurban~world$polegw)

##Polegw~Growth

anova11=aov(world$growth~world$polegw)

summary(anova11)

mean(anova11$res)

median(anova11$res)

kruskal.test(world$growth~world$polegw)

pairwise.wilcox.test(world$growth,world$polegw,p.adjust.method = "bonf",paired=F)

tapply(world$growth,world$polegw,summary)

boxplot(world$growth~world$polegw)

##Polegw~Electcon

anova12=aov(world$electcon~world$polegw)

summary(anova12)

mean(anova12$res)

median(anova12$res)

kruskal.test(world$electcon~world$polegw)

TukeyHSD(anova12)

pairwise.wilcox.test(world$electcon,world$polegw,p.adjust.method = "bonf",paired=F)

tapply(world$electcon,world$polegw,summary)

boxplot(world$electcon~world$polegw)

##WORD\_MATERIAL##

##Polegw~Educexpc

anova13=aov(world$educexpc~world$polegw)

summary(anova13)

mean(anova13$res)

median(anova13$res)

#o mesos oros einai sxedon 0 kai h diamesos einai -31.11 ara den proseggizei normality

kruskal.test(world$educexpc~world$polegw)

TukeyHSD(anova13)

#gia ta eksoda ekpaideushs paratirite simantiki diafora stis xwres metaksu pou ta epipeda

#isotitas sta nomika kai politika dikeomata metaksu antrwn kai gunaikwn einai weakly equal-weakly unequal

#kai strongy equal- weakly unequal

pairwise.wilcox.test(world$educexpc,world$polegw,p.adjust.method = "bonf",paired=F)

tapply(world$educexpc,world$polegw,summary)

boxplot(world$educexpc~world$polegw,main="Educational Expenses ~ Gender Equality ", col=2:6,ylab="Educational Expenses",xlab="Women Equality ")

##Polegw~saving

anova14=aov(world$saving~world$polegw)

summary(anova14)

mean(anova14$res)

median(anova14$res)

kruskal.test(world$saving~world$polegw)

pairwise.wilcox.test(world$saving,world$polegw,p.adjust.method = "bonf",paired=F)

tapply(world$saving,world$polegw,summary)

boxplot(world$saving~world$polegw)

##Polegw~cpi

anova15=aov(world$cpi~world$polegw)

summary(anova15)

mean(anova15$res)

median(anova15$res)

kruskal.test(world$cpi~world$polegw)

pairwise.wilcox.test(world$cpi,world$polegw,p.adjust.method = "bonf",paired=F)

tapply(world$cpi,world$polegw,summary)

boxplot(world$cpi~world$polegw)

**#Για να είναι έγκυρο ένα μοντέλο θα πρέπει να μην έχει πολυσυγγραμμικότητα,**

**#να ικανοποιείται η κανονικότητα, η γραμμικότητα, η ομοσκεδαστικότητα,**

**#η ανεξαρτησία και ελέγχεται η ύπαρξη ακραίων τιμωνν και σημείων επιρροής.**

**###MONTELO###**

simple\_model=lm(world$growth ~ world$population + world$deathrate +

world$popurban + world$saving + world$electcon + world$educexpc + world$cpi + world$polegw)

summary(simple\_model)

##elegxos kanonikotitas katalipwn

library(nortest)

lillie.test(simple\_model$res)

shapiro.test(simple\_model$res)

##p-value=0.053 εχω κανονικοτητα

par(mfrow=c(3,1))

hist(simple\_model$res,col="cadetblue",main="Histogram of Residuals",xlab="Residuals")

qqnorm(rstandard(simple\_model),col="red")

qqline(rstandard(simple\_model),col="black")

**##elegxos omoioskedastikotitas**

quancut=function(x,digits=6){cut(x,breaks=quantile(x),include.lowest = T)}

qfits=quancut(simple\_model$fit)

library(car)

leveneTest(rstandard(simple\_model),qfits)

**#p-value>0.05 ara den aporiptoume Ho opote uparxei omoioskedastikotita**

plot(simple\_model$fit, rstandard(simple\_model), col="cadetblue",pch=15)

abline( h= 1.96, col='black', lwd=2, lty=2)

abline( h=-1.96, col='black', lwd=2, lty=2)

**#Elegxos Polysyggramikothtas**

install.packages("car")

library(car)

vif(simple\_model)

**##metaksu twn metavlitwn den uparxei idietera isxurh susxetisi metaksu twn metavlitwn mou**

**#afou oles einai konta sto 1 ektow apo Electcon Educexpc**

**##grammikotita**

residualPlots(simple\_model)#me thn paratirisi 41

install.packages("sjPlot")

library(sjPlot)

library(sjPlot)

tab\_model(simple\_model)

**##aneksartisia katalipwn den kanw giati exw diastromatika data**

**##akraies times-simia epirohs**

par(mfrow=c(2,1))

outlierTest(simple\_model)

plot(simple\_model,which = 4) ##cooks distance

abline(h=4/27,col="cadetblue")

plot(simple\_model,which=5)##shmeia epirohs

**###afairw to simeio epirohs**

worldnew=world[-41,]

simple\_model1=lm(formula = worldnew$growth ~ worldnew$population + worldnew$deathrate +

worldnew$popurban + worldnew$saving + worldnew$electcon + worldnew$educexpc +

worldnew$cpi + worldnew$polegw)

summary(simple\_model1)

**##ELEGXOS KANONIKOTITAS**

par(mfrow=c(2,2))

hist(simple\_model1$res)

qqnorm(rstandard(simple\_model1),col="red")

qqline(rstandard(simple\_model1),col="black")

shapiro.test(simple\_model1$res)

lillie.test(simple\_model1$res)

**#ARA IKANOPOIEITAI H KANONIKOTITA p-value>0.05**

**##elegxos omoioskedastikotitas**

quancut=function(x,digits=6){cut(x,breaks=quantile(x),include.lowest = T)}

qfits=quancut(simple\_model1$fit)

library(car)

leveneTest(rstandard(simple\_model1),qfits)

#p-value>0.05 ara den aporiptoume Ho opote uparxei omoioskedastikotita

plot(simple\_model1$fit, rstandard(simple\_model1), col="STEELBLUE",pch=15)

abline( h= 1.96, col='black', lwd=2, lty=2)

abline( h=-1.96, col='black', lwd=2, lty=2)

**#Elegxos Polysyggramikothtas**

install.packages("car")

library(car)

vif(simple\_model1)

**## metaksu twn metavlitwn den uparxei idietera isxurh sugkrammikotita metaksu twn metavlitwn mou**

**#afou oles oi times einai konta sto 1 ektos apo to electcon kai to educexpc**

**##grammikotita**

residualPlots(simple\_model1)#xwris thn paratirisi 41

install.packages("sjPlot")

library(sjPlot)

library(sjPlot)

tab\_model(simple\_model1)

summary(simple\_model1)

# 6.ΠΙΝΑΚΕΣ

