

ΕΡΓΑΣΙΑ ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ

Δείκτες Παγκόσμιας Ανάπτυξης

Μία Στατιστική Ανάλυση για τον Ελέγχο Διαφθοράς



Άγγελος Ντούσης

A.M. 1112202000156

Τμήμα Μαθηματικών
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Αθήνα 2024

Abstract

Η παρούσα εργασία εστιάζει στη μελέτη των δεικτών παγκόσμιας ανάπτυξης και του ελέγχου της διαφθοράς μέσω στατιστικής ανάλυσης. Αρχικά, μελετήθηκαν τα δεδομένα από διάφορες χώρες, καθώς και οι διάφοροι δείκτες που σχετίζονται με την παγκόσμια ανάπτυξη. Επιπλέον, πραγματοποιήθηκε η κατηγοριοποίηση κάποιων δεδομένων σε ποιοτικές και ποσοτικές μεταβλητές, με στόχο την εξαγωγή συμπερασμάτων σχετικά με τον μέσο όρο, την διασπορά και παράλληλα γίνεται έλεγχος αν αυτά είναι κανονικά κατανεμημένα. Έπειτα, υλοποιήθηκε ένα γραμμικό μοντέλο με τον έλεγχο της διαφθοράς ως εξαρτώμενη μεταβλητή, εξετάζοντας τη σχέση της με άλλους δείκτες παγκόσμιας ανάπτυξης. Σημαντικό βήμα ήταν η επαλήθευση των υποθέσεων του μοντέλου, προκειμένου να διασφαλιστεί η εγκυρότητα των αποτελεσμάτων. Τέλος, πραγματοποιήθηκε ανάλυση διασποράς ANOVA, με εξαρτημένη μεταβλητή τον έλεγχο διαφθοράς και με ανεξάρτητες μεταβλητές τις κατηγορίες γονιμότητας και πληθωρισμού με σκοπό να εξεταστεί αν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των μέσων των ανεξάρτητων μεταβλητών.

Λέξεις κλειδιά: Έλεγχος Διαφθοράς, Γραμμικά Μοντέλα, Δείκτες Παγκόσμιας Ανάπτυξης, ANOVA

Περιεχόμενα

Εισαγωγή	3
Δεδομένα	3
Μελέτη και σχολιασμός των περιγραφικών στατιστικών μέτρων	8
Έλεγχος Διαφθοράς	8
Πολιτική Σταθερότητα	9
Φωνή και Λογοδοσία	11
Κατηγορία Γονιμότητας	12
Ρυθμός Πληθωρισμού	13
Ανάλυση Περιγραφικών Στατιστικών ανά Κατηγορία Γονιμότητας	14
Εισαγωγή	14
Αποτελέσματα	14
Σχολιασμός	15
Μέθοδοι	15
Γραμμικό Μοντέλο	15
Ανάλυση Διακύμανσης (ANOVA)	16
Αποτελέσματα	17
Ανάλυση Γραμμικού Μοντέλου	17
Έλεγχος Στατιστικής Σημαντικότητας των ανεξάρτητων μεταβλητών ElectrAccess και PopGrowth	21
Έλεγχος Υποθέσεων και Διαγνωστικά του Μοντέλου	21
Μέσα και Ατομικά Διαστήματα εμπιστοσύνης για την CorControl	22
Ανάλυση ANOVA	23
Συζήτηση	24
Αναφορές	25

Εισαγωγή

Οι Δείκτες Παγκόσμιας Ανάπτυξης (World Development Indicators - WDIs) αποτελούν μία πηγή πληροφοριών ζωτικής σημασίας για την παρακολούθηση της προόδου μίας χώρας, τον εντοπισμό τάσεων, και τη διαμόρφωση πολιτικών. Οι WDIs περιλαμβάνουν ένα ευρύ φάσμα κοινωνικοοικονομικών δεικτών που εκτείνονται σε διάφορους τομείς όπως η εκπαίδευση, η υγεία, το περιβάλλον, η οικονομία και η διακυβέρνηση, παρέχοντας μια συνολική εικόνα της παγκόσμιας αναπτυξιακής δυναμικής. Σε αυτή τη μελέτη θα χρησιμοποιήσουμε τη βάση δεδομένων WDI της Παγκόσμιας Τράπεζας. Συγκεκριμένα, δίνεται ένα υποσύνολο των δεδομένων, το οποίο αποτελείται από 27 μεταβλητές για 120 χώρες, το έτος 2018.

Δεδομένα

Πίνακας 1: Δείκτες Παγκόσμιας Ανάπτυξης

Συντομογραφία	Σημασία	Περιγραφή
ElectrAccess	Πρόσβαση στην ηλεκτρική ενέργεια	Το ποσοστό του πληθυσμού που έχει πρόσβαση στην ηλεκτρική ενέργεια.
CookAccess	Πρόσβαση σε καθαρά καύσιμα	Το ποσοστό του συνολικού πληθυσμού που χρησιμοποιεί κατά κύριο λόγο καθαρά καύσιμα και τεχνολογίες για το μαγείρεμα. Σύμφωνα με τον ΠΟΥ, η κηροζίνη εξαιρείται από τα καθαρά καύσιμα μαγειρέματος.
AgriLand	Γεωργική γη	Η γεωργική γη αναφέρεται στο μερίδιο της έκτασης της γης που είναι καλλιεργήσιμη, υπό μόνιμες καλλιέργειες και υπό μόνιμους βοσκότοπους.
BirthRate	Ποσοστό γεννήσεων	Ο δείκτης γεννήσεων δείχνει τον αριθμό των γεννήσεων που συμβαίνουν κατά τη διάρκεια του έτους, ανά 1.000 κατοίκους που υπολογίζονται στα μέσα του έτους. Η αφαίρεση του ακατέργαστου ποσοστού θανάτων από το ακατέργαστο ποσοστό γεννήσεων παρέχει το ποσοστό φυσικής αύξησης, το οποίο ισούται με το ποσοστό μεταβολής του πληθυσμού ελλείψει μετανάστευσης.
CO2	Εκπομπές CO2	Οι εκπομπές διοξειδίου του άνθρακα είναι αυτές που προέρχονται από την καύση ορυκτών καυσίμων και την παραγωγή τσιμέντου.
CompEdu	Υποχρεωτική εκπαίδευση	Η διάρκεια της υποχρεωτικής εκπαίδευσης είναι ο αριθμός των ετών που τα παιδιά είναι νομικά υποχρεωμένα να φοιτούν στο σχολείο.

Συνεχίζεται στην επόμενη σελίδα

Πίνακας 1 – συνέχεια από την προηγούμενη σελίδα

Συντομογραφία	Σημασία	Περιγραφή
DeathRate	Ποσοστό θανάτου	Ο δείκτης θνησιμότητας υποδηλώνει τον αριθμό των θανάτων κατά τη διάρκεια του έτους, ανά 1.000 κατοίκους που εκτιμάται στα μέσα του έτους.
FoodExports	Εξαγωγές τροφίμων	Τα τρόφιμα περιλαμβάνουν τα εμπορεύματα των τμημάτων SITC 0 (τρόφιμα και ζώα), 1 (ποτά και καπνός) και 4 (ζωικά και φυτικά έλαια και λίπη) και του τμήματος SITC 22 (ελαιούχοι σπόροι, ελαιούχοι καρποί και ελαιοπυρήνες).
Telephone	Συνδρομές σταθερής τηλεφωνίας	Οι σταθερές τηλεφωνικές συνδρομές αναφέρονται στο άθροισμα του ενεργού αριθμού των αναλογικών σταθερών τηλεφωνικών γραμμών, των συνδρομών voice-over-IP (VoIP), των συνδρομών σταθερού ασύρματου τοπικού βρόχου (WLL), των ισοδύναμων φωνητικών καναλιών ISDN και των σταθερών δημόσιων καρτοτηλεφώνων.
Internet	Άτομα που χρησιμοποιούν το Διαδίκτυο	Οι χρήστες του Διαδικτύου είναι άτομα που χρησιμοποίησαν το Διαδίκτυο (από οποιαδήποτε τοποθεσία) τους τελευταίους 3 μήνες. Το Διαδίκτυο μπορεί να χρησιμοποιηθεί μέσω υπολογιστή, κινητού τηλεφώνου, ψηφιακής τηλεόρασης κ.λπ.
PopGrowth	Αύξηση του πληθυσμού	Ο ετήσιος ρυθμός αύξησης του πληθυσμού για το έτος t είναι ο εκθετικός ρυθμός αύξησης του πληθυσμού στα μέσα του έτους από το έτος $t-1$ έως το t , εκφρασμένος ως ποσοστό.
BusinessTime	Χρόνος δημιουργίας επιχείρησης	Ο χρόνος που απαιτείται για την έναρξη μιας επιχείρησης είναι ο αριθμός των ημερολογιακών ημερών που απαιτούνται για την ολοκλήρωση των διαδικασιών για τη νόμιμη λειτουργία μιας επιχείρησης.
AdolFertRate	Ποσοστό γονιμότητας εφήβων	Ο δείκτης γονιμότητας εφήβων είναι ο αριθμός των γεννήσεων ανά 1.000 γυναίκες ηλικίας 15-19 ετών.
Precipitation	Μέση βροχόπτωση σε βάθος (mm ανά έτος)	Η μέση βροχόπτωση είναι ο μακροχρόνιος μέσος όρος σε βάθος της ετήσιας βροχόπτωσης στη χώρα. Ως βροχόπτωση ορίζεται κάθε είδος νερού που πέφτει από τα σύννεφα ως υγρό ή στερεό.

Συνεχίζεται στην επόμενη σελίδα

Πίνακας 1 – συνέχεια από την προηγούμενη σελίδα

Συντομογραφία	Σημασία	Περιγραφή
CorControl	Έλεγχος της διαφθοράς	Ο έλεγχος της διαφθοράς καταγράφει τις αντιλήψεις σχετικά με τον βαθμό στον οποίο η δημόσια εξουσία ασκείται για ιδιωτικό όφελος, συμπεριλαμβανομένων τόσο των μικροδιαφθορών όσο και των μεγάλων μορφών διαφθοράς, καθώς και της "άλωσης" του κράτους από τις ελίτ και τα ιδιωτικά συμφέροντα. Η εκτίμηση δίνει τη βαθμολογία της χώρας στον συνολικό δείκτη.
EmployerPerc	Εργοδότες	Εργοδότες είναι οι εργαζόμενοι που, εργαζόμενοι για δικό τους λογαριασμό ή με έναν ή λίγους εταίρους, κατέχουν το είδος των θέσεων εργασίας που ορίζονται ως αυτοαπασχόληση", δηλαδή θέσεις εργασίας στις οποίες η αμοιβή εξαρτάται άμεσα από τα κέρδη που προέρχονται από τα αγαθά και τις υπηρεσίες που παράγονται).
FertRate	Ποσοστό γονιμότητας	Ο συνολικός δείκτης γονιμότητας αντιπροσωπεύει τον αριθμό των παιδιών που θα γεννηθούν από μια γυναίκα αν ζήσει μέχρι το τέλος της αναπαραγωγικής της ηλικίας και να γεννήσει παιδιά σύμφωνα με τους ειδικούς κατά ηλικία δείκτες γονιμότητας του συγκεκριμένου έτους.
GDPperc	Αύξηση του κατά κεφαλήν ΑΕΠ	Ετήσιος ποσοστιαίος ρυθμός αύξησης του κατά κεφαλήν ΑΕΠ με βάση το σταθερό τοπικό νόμισμα. Το κατά κεφαλήν ΑΕΠ είναι το ακαθάριστο εγχώριο προϊόν διαιρούμενο με τον πληθυσμό στο μέσο του έτους. Το ΑΕΠ σε τιμές αγοραστική είναι το άθροισμα της ακαθάριστης προστιθέμενης αξίας όλων των κατοίκων παραγωγών της οικονομίας συν τους φόρους επί των προϊόντων και μείον τις επιδοτήσεις που δεν περιλαμβάνονται στην αξία των προϊόντων. Υπολογίζεται χωρίς αφαιρέσεις για την απόσβεση των κατασκευασμένων περιουσιακών στοιχείων ή για την εξάντληση και υποβάθμιση των φυσικών πόρων.

Συνεχίζεται στην επόμενη σελίδα

Πίνακας 1 – συνέχεια από την προηγούμενη σελίδα

Συντομογραφία	Σημασία	Περιγραφή
GDPdollars	Κατά κεφαλήν ΑΕΠ	Το κατά κεφαλήν ΑΕΠ είναι το ακαθάριστο εγχώριο προϊόν διαιρούμενο με τον πληθυσμό στο μέσο του έτους. Το ΑΕΠ είναι το άθροισμα της ακαθάριστης προστιθέμενης αξίας όλων των κατοίκων παραγωγών της οικονομίας συν τους φόρους επί των προϊόντων και μείον τις επιδοτήσεις που δεν περιλαμβάνονται στην αξία των προϊόντων. Υπολογίζεται χωρίς να γίνονται αφαιρέσεις για την απόσβεση των κατασκευασμένων περιουσιακών στοιχείων ή για την εξάντληση και την υποβάθμιση των φυσικών πόρων. Τα στοιχεία είναι σε τρέχοντα δολάρια ΗΠΑ.
Inflation	Πληθωρισμός, τιμές καταναλωτή	Ο πληθωρισμός, όπως μετράται από τον δείκτη τιμών καταναλωτή, αντικατοπτρίζει την ετήσια ποσοστιαία μεταβολή του κόστους για τον μέσο καταναλωτή της απόκτησης ενός καλαθιού αγαθών και υπηρεσιών που μπορεί να είναι σταθερό ή να μεταβάλλεται σε συγκεκριμένα χρονικά διαστήματα, όπως το έτος. Γενικά χρησιμοποιείται ο τύπος Laspeyres.
WaterStress	Εξάντληση υδάτινων πόρων	Η ανάλυση γλυκού νερού ως ποσοστό των διαθέσιμων πόρων γλυκού νερού είναι ο λόγος μεταξύ του συνολικού γλυκού νερού που αποσύρεται από όλους τους κύριους τομείς και των συνολικών ανανεώσιμων πόρων γλυκού νερού, αφού ληφθούν υπόψη οι περιβαλλοντικές απαιτήσεις σε νερό.
LifeExp	Προσδόκιμο ζωής κατά τη γέννηση	Το προσδόκιμο ζωής κατά τη γέννηση υποδηλώνει τον αριθμό των ετών που θα ζούσε ένα νεογέννητο βρέφος εάν τα επικρατούντα πρότυπα θνησιμότητας κατά τη στιγμή της γέννησής του παρέμεναν τα ίδια καθ' όλη τη διάρκεια της ζωής του.
MortRate	Βρεφική Θνησιμότητα	Ο δείκτης βρεφικής θνησιμότητας είναι ο αριθμός των βρεφών που πεθαίνουν πριν από τη συμπλήρωση ενός έτους, ανά 1.000 γεννήσεις ζώντων σε ένα δεδομένο έτος.

Συνεχίζεται στην επόμενη σελίδα

Πίνακας 1 – συνέχεια από την προηγούμενη σελίδα

Συντομογραφία	Σημασία	Περιγραφή
PolStab	Πολιτική σταθερότητα και απουσία βίας/τρομοκρατίας	Η πολιτική σταθερότητα και η απουσία βίας/τρομοκρατίας μετρά τις αντιλήψεις σχετικά με την πιθανότητα πολιτικής αστάθειας και/ή βίας με πολιτικά κίνητρα, συμπεριλαμβανομένης της τρομοκρατίας. Η εκτίμηση δίνει τη βαθμολογία της χώρας στο συνολικό δείκτη, σε μονάδες μιας τυπικής κανονικής κατανομής, δηλαδή κυμαίνεται περίπου από -2,5 έως 2,5.
PopPerc14	Πληθυσμός ηλικίας 0-14 ετών	Πληθυσμός ηλικίας 0 έως 14 ετών ως ποσοστό του συνολικού πληθυσμού.
VoiceAcc	Φωνή και Λογοδοσία	Η "Φωνή και Λογοδοσία" καταγράφει τις αντιλήψεις σχετικά με τον βαθμό στον οποίο οι πολίτες μιας χώρας μπορούν να συμμετέχουν στην επιλογή της κυβέρνησής τους, καθώς και την ελευθερία της έκφρασης, την ελευθερία του συνεταιρίζεσθαι και τα ελεύθερα μέσα ενημέρωσης. Η εκτίμηση δίνει τη βαθμολογία της χώρας στον συνολικό δείκτη, σε μονάδες μιας τυπικής κανονικής κατανομής, δηλαδή κυμαίνεται περίπου από -2,5 έως 2,5.
WomBusiness	Δείκτης γυναικών στον νόμο και στις επιχειρήσεις	Ο δείκτης μετρά τον τρόπο με τον οποίο οι νόμοι και οι κανονισμοί επηρεάζουν τις οικονομικές ευκαιρίες των γυναικών. Οι συνολικές βαθμολογίες υπολογίζονται με τη μέση βαθμολογία κάθε δείκτη (Κινητικότητα, Χώρος εργασίας, Αμοιβή, Γάμος, Γονιμότητα, Επιχειρηματικότητα, Περιουσιακά στοιχεία και Σύνταξη), με το 100 να αντιπροσωπεύει την υψηλότερη δυνατή βαθμολογία.

Μελέτη και σχολιασμός των περιγραφικών στατιστικών μέτρων

Έλεγχος Διαφθοράς

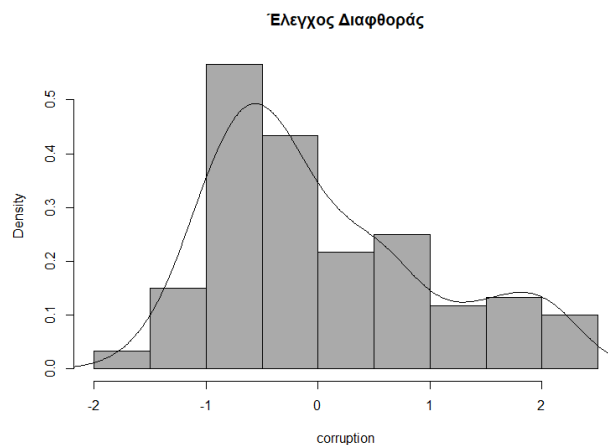
Τα περιγραφικά στατιστικά μέτρα για τη μεταβλητή "Έλεγχος Διαφθοράς-CorControl", η οποία είναι μια ποσοτική μεταβλητή, είναι τα εξής:

Πίνακας 2: Περιγραφικά Στατιστικά Μέτρα για τον Έλεγχο Διαφθοράς

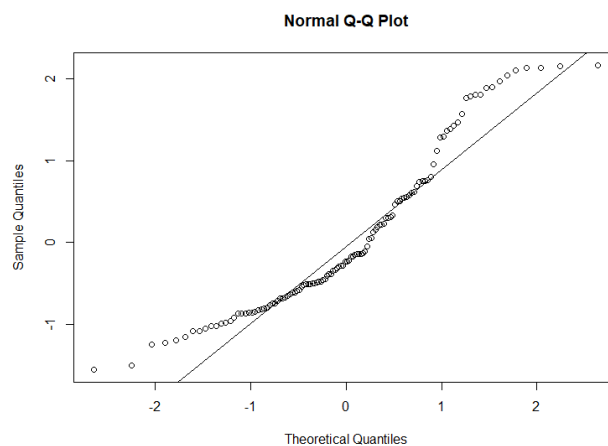
Στατιστικό Μέτρο	Τιμή
Διάμεσος	-0.2312463
1ο Ποσοστημόριο	-1.5588410
2ο Ποσοστημόριο	-0.6882838
3ο Ποσοστημόριο	-0.2312463
4ο Ποσοστημόριο	0.5830534
5ο Ποσοστημόριο	2.1713212

Παρακάτω παρουσιάζονται τα γραφήματα για τη μεταβλητή "Έλεγχος Διαφθοράς":

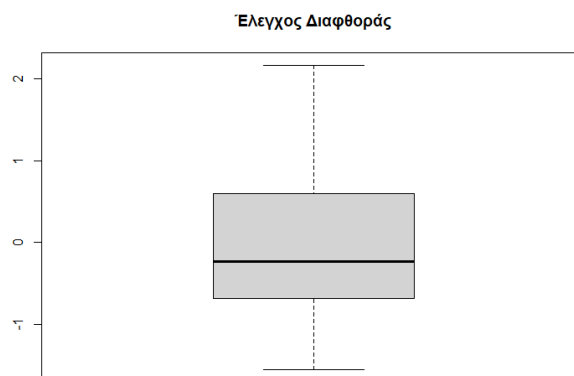
- Ιστόγραμμα:



- QQ Plot:



- Boxplot:



Σχολιασμός: Παρατηρούμε από τα γραφήματα ότι η μεταβλητή CorControl δεν είναι κανονικά κατανοημένη. Για να το επιβεβαιώσουμε κάνουμε το τεστ κανονικότητας Shapiro-Wilk το οποίο μας βγάζει αποτέλεσμα $2.181 \cdot 10^{-6}$. Το p-value είναι 0.003363745 και σε επίπεδο σημαντικότητας 5% απορρίπτω τη μηδενική υπόθεση. Άρα η μεταβλητή CorControl δεν είναι κανονικά κατανοημένη. Δεδομένου αυτού του συμπεράσματος, είναι προτιμότερο να χρησιμοποιήσουμε τη διάμεσο και τα ποσοστημόρια ως περιγραφικά στατιστικά μέτρα. Αυτό συμβαίνει γιατί η μέση τιμή και η διασπορά μπορεί να επηρεαστούν από ακραίες τιμές και να μας δώσουν παραπλανητικά αποτελέσματα. Χρησιμοποιώντας τη διάμεσο και τα ποσοστημόρια, μπορούμε να έχουμε μια πιο αξιόπιστη και ακριβή περιγραφή της κατανομής των δεδομένων.

Πολιτική Σταθερότητα

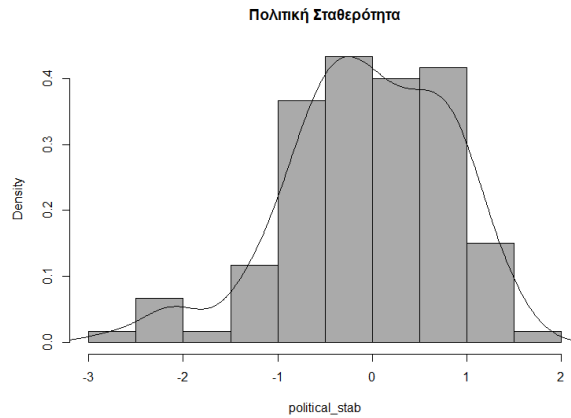
Τα περιγραφικά στατιστικά μέτρα για τη μεταβλητή "Πολιτική Σταθερότητα-PolStab", η οποία είναι μια ποσοτική μεταβλητή, είναι τα εξής:

Πίνακας 3: Περιγραφικά Στατιστικά Μέτρα για την Πολιτική Σταθερότητα

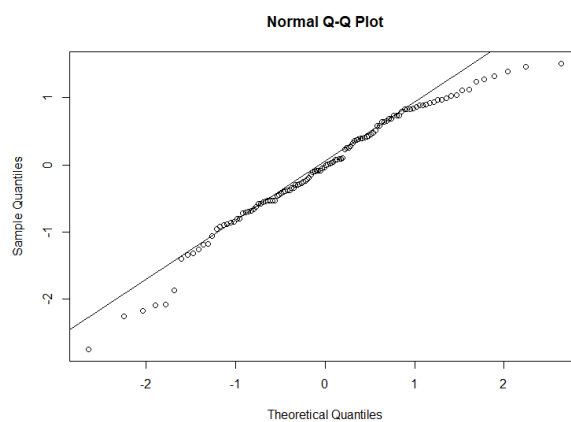
Στατιστικό Μέτρο	Τιμή
Διάμεσος	-0.01948036
1ο Ποσοστημόριο	-2.75326204
2ο Ποσοστημόριο	-0.54483178
3ο Ποσοστημόριο	-0.01948036
4ο Ποσοστημόριο	0.64384830
5ο Ποσοστημόριο	1.51216090

Παρακάτω παρουσιάζονται τα γραφήματα για τη μεταβλητή "Πολιτική Σταθερότητα":

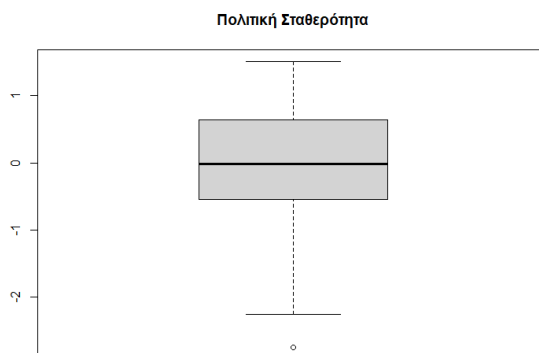
- Ιστόγραμμα:



- QQ Plot:



- Boxplot:



Σχολιασμός: Παρατηρούμε από τα γραφήματα ότι η μεταβλητή *CorControl* δεν είναι κανονικά κατανοημένη. Για να το επιβεβαιώσουμε κάνουμε το τεστ κανονικότητας Shapiro-Wilk το οποίο μας βγάζει αποτέλεσμα 0.009249. Το p-value είναι 0.009248762 και σε επίπεδο σημαντικότητας 5% απορρίπτω τη μηδενική υπόθεση. Άρα η μεταβλητή *PolStab* δεν είναι κανονικά κατανοημένη. Ομοίως χρησιμοποιούμε την διάμεσο και τα ποσοστημόρια, καθώς αυτά προσφέρουν μια πιο αξιόπιστη και ακριβή περιγραφή της κατανομής των δεδομένων όταν αυτά δεν ακολουθούν την κανονική κατανομή.

Φωνή και Λογοδοσία

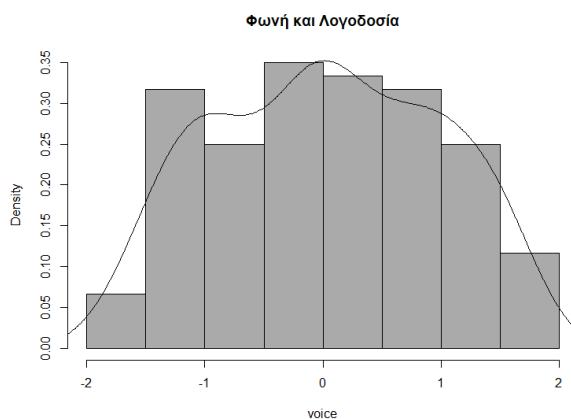
Τα περιγραφικά στατιστικά μέτρα για τη μεταβλητή "Φωνή και Λογοδοσία-VoiceAcc", η οποία είναι ποσοτική μεταβλητή, είναι τα εξής:

Πίνακας 4: Περιγραφικά Στατιστικά Μέτρα για την Πολιτική Σταθερότητα

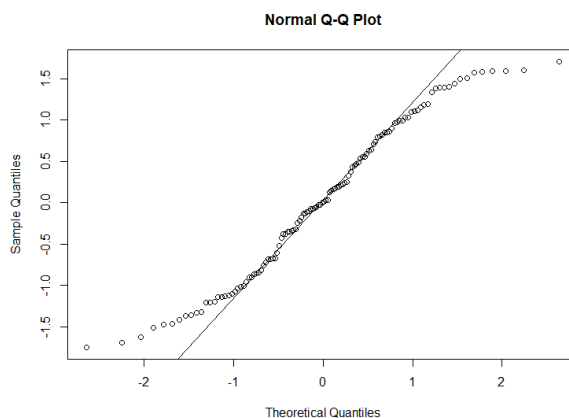
Στατιστικό Μέτρο	Τιμή
Διάμεσος	0.005002012
1ο Ποσοστημόριο	-1.750232816
2ο Ποσοστημόριο	-0.772283003
3ο Ποσοστημόριο	0.005002012
4ο Ποσοστημόριο	0.828630462
5ο Ποσοστημόριο	1.708544374

Παρακάτω παρουσιάζονται τα γραφήματα για τη μεταβλητή "Πολιτική Σταθερότητα":

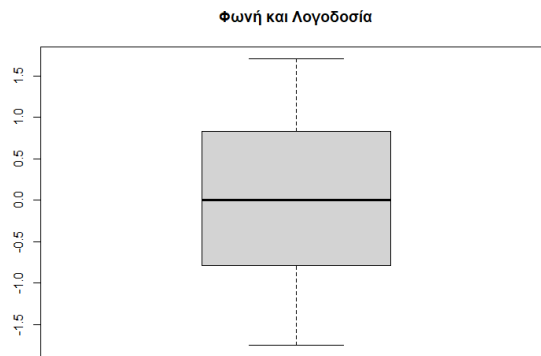
- Ιστόγραμμα:



- QQ Plot:



- Boxplot:



Σχολιασμός: Παρατηρούμε από τα γραφήματα ότι η μεταβλητή VoiceAcc δεν είναι κανονικά κατανοημένη. Για να το επιβεβαιώσουμε κάνουμε το τεστ κανονικότητας Shapiro-Wilk το οποίο μας βγάζει αποτέλεσμα 0.003364. Το p-value είναι 0.003363745 και σε επίπεδο σημαντικότητας 5% απορρίπτω τη μηδενική υπόθεση. Άρα η μεταβλητή VoiceAcc δεν είναι κανονικά κατανοημένη.

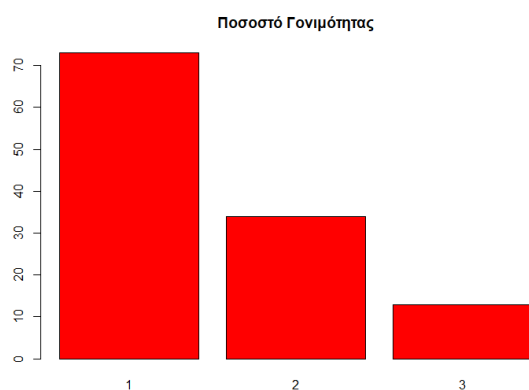
Κατηγορία Γονιμότητας

Για την κατηγοριοποίηση της "Κατηγορίας Γονιμότητας" (Fertility Category), η οποία θα λάβει το όνομα FertCat, θα χρησιμοποιηθούν τα εξής κριτήρια βασισμένα στον Ρυθμό Εφηβικής Γονιμότητας (Adolescent Fertility Rate – AdolFertRate) της εκάστοτε χώρας:

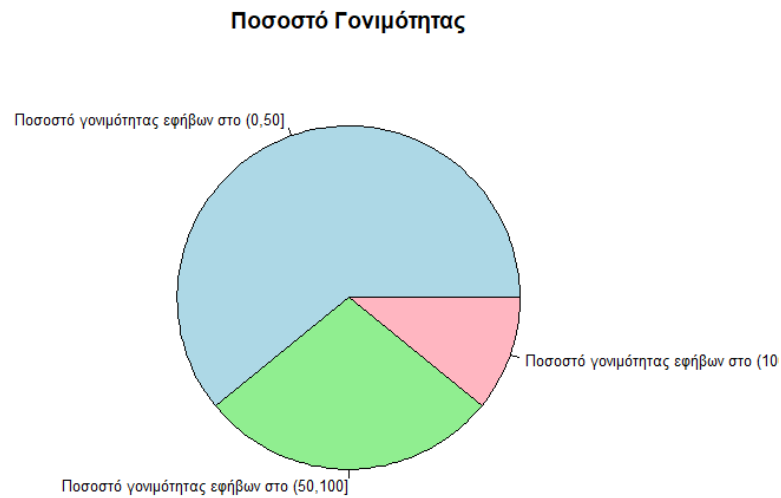
- Η τιμή 1 θα αποδίδεται όταν ο Ρυθμός Εφηβικής Γονιμότητας βρίσκεται στο διάστημα (0, 50].
- Η τιμή 2 θα αποδίδεται όταν ο Ρυθμός Εφηβικής Γονιμότητας βρίσκεται στο διάστημα (50, 100].
- Η τιμή 3 θα αποδίδεται όταν ο Ρυθμός Εφηβικής Γονιμότητας βρίσκεται στο διάστημα (100, ∞).

Με βάση αυτά τα κριτήρια, κάθε χώρα θα κατηγοριοποιείται σε μία από τις τρεις κατηγορίες γονιμότητας, ανάλογα με τον Ρυθμό Εφηβικής Γονιμότητας της. Τα περιγραφικά στατιστικά μέτρα για τη μεταβλητή "Κατηγορία Γονιμότητας-FertCat", η οποία είναι ποιοτική μεταβλητή, είναι τα εξής:

- Barplot:



- Pie Chart:



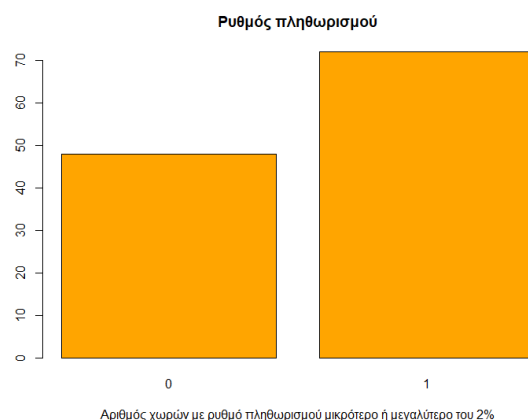
Ρυθμός Πληθωρισμού

Για την κατηγοριοποίηση του "Ρυθμού Πληθωρισμού" (Inflation Category), η οποία θα λάβει το όνομα InflCat, θα χρησιμοποιηθούν τα εξής κριτήρια βασισμένα στον πληθωρισμό (Inflation) της εκάστοτε χώρας:

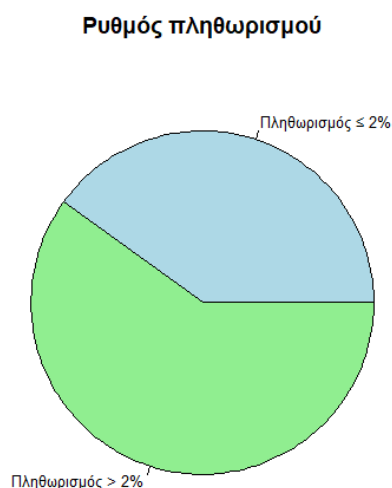
- Η τιμή 0 θα αποδίδεται όταν ο πληθωρισμός της χώρας είναι μικρότερος ή ίσος με 2%.
- Η τιμή 1 θα αποδίδεται όταν ο πληθωρισμός της χώρας είναι μεγαλύτερος από 2%.

Με βάση αυτά τα κριτήρια, κάθε χώρα θα κατηγοριοποιείται σε μία από τις δύο κατηγορίες πληθωρισμού, ανάλογα με το ποσοστό πληθωρισμού της. Τα περιγραφικά στατιστικά μέτρα για τη μεταβλητή "Κατηγορία Πληθωρισμού-InflCat", η οποία είναι ποιοτική μεταβλητή, είναι τα εξής:

- Barplot:



- Pie Chart:



Ανάλυση Περιγραφικών Στατιστικών ανά Κατηγορία Γονιμότητας

Εισαγωγή

Χρησιμοποιήσαμε το πακέτο `psych` και τη συνάρτηση `describeBy` για να υπολογίσουμε βασικά περιγραφικά στατιστικά μέτρα (δειγματικό μέσο και τυπική απόκλιση) για τις μεταβλητές Έλεγχος Διαφθοράς (CorControl), Πολιτική Σταθερότητα (PolStab), και Φωνή και Υπευθυνότητα (VoiceAcc) ανά Κατηγορία Γονιμότητας (FertCat).

Αποτελέσματα

Τα περιγραφικά στατιστικά μέτρα παρουσιάζονται στους παρακάτω πίνακες.

Πίνακας 5: Περιγραφικά Στατιστικά για τον Έλεγχο Διαφθοράς (CorControl) ανά Κατηγορία Γονιμότητας (FertCat)

Κατηγορία	Δείγμα	Μέσος Όρος	Τυπ. Απόκλ.	Διάμεσος	Ελάχιστο	Μέγιστο	Ασυμμετρία
1	73	0.4325	0.9790	0.2978	-1.0824	2.1713	0.34
2	34	-0.4912	0.5543	-0.6039	-1.5588	0.7556	0.45
3	13	-0.8115	0.3420	-0.7657	-1.2332	-0.1312	0.36

Πίνακας 6: Περιγραφικά Στατιστικά για την Πολιτική Σταθερότητα (PolStab) ανά Κατηγορία Γονιμότητας (FertCat)

Κατηγορία	Δείγμα	Μέσος Όρος	Τυπ. Απόκλ.	Διάμεσος	Ελάχιστο	Μέγιστο	Ασυμμετρία
1	73	0.2175	0.8213	0.3860	-2.2539	1.5122	-0.75
2	34	-0.2730	0.6619	-0.2475	-2.7533	0.8331	-1.22
3	13	-0.9508	0.7687	-0.7017	-2.1721	0.0825	-0.45

Πίνακας 7: Περιγραφικά Στατιστικά για τη Φωνή και Υπευθυνότητα (VoiceAcc) ανά Κατηγορία Γονιμότητας (FertCat)

Κατηγορία	Δείγμα	Μέσος Όρος	Τυπ. Απόκλ.	Διάμεσος	Ελάχιστο	Μέγιστο	Ασυμμετρία
1	73	0.1928	1.0578	0.4463	-1.6913	1.7085	-0.28
2	34	-0.1211	0.6966	0.0193	-1.7502	1.1191	-0.40
3	13	-0.4826	0.3794	-0.3362	-1.2035	-0.0610	-0.75

Σχολιασμός

Παρατηρείται ότι στις χώρες με μεγάλο ρυθμό αναπαραγωγής σε εφηβική ηλικία υπάρχει μικρότερος έλεγχος διαφθοράς. Αυτό φαίνεται από το γεγονός ότι η μέση τιμή της CorControl μειώνεται καθώς προχωράμε σε επίπεδα της FertCat που αντιστοιχούν σε διαστήματα με μεγαλύτερες τιμές της AdolFertRate. Συγκεκριμένα, ο μέσος όρος της CorControl είναι 0.4325 για την πρώτη κατηγορία γονιμότητας, -0.4912 για τη δεύτερη κατηγορία, και -0.8115 για την τρίτη κατηγορία. Αντίστοιχες τάσεις παρατηρούνται και για τις άλλες δύο μεταβλητές (PolStab και VoiceAcc), όπου οι μέσες τιμές μειώνονται καθώς αυξάνεται η κατηγορία γονιμότητας.

Μέθοδοι

Αυτό η ενότητα αφορά τις μεθόδους που χρησιμοποιήθηκαν για τη στατιστική ανάλυση.

Γραμμικό Μοντέλο

Για την εξήγηση και πρόβλεψη του Ελέγχου Διαφθοράς από τους άλλους Δείκτες Παγκόσμιας Ανάπτυξης, διενεργήθηκε ανάλυση πολλαπλής παλινδρόμησης. Συγκεκριμένα, η επιλογή του βέλτιστου μοντέλου έγινε με το κριτήριο AIC και τη διαδικασία βηματικής επιλογής (backward stepwise).

Το γραμμικό μοντέλο έχει τη μορφή:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon, \quad (1)$$

όπου Y είναι η εξαρτημένη μεταβλητή (Έλεγχος Διαφθοράς), X_i είναι οι ανεξάρτητες μεταβλητές (Δείκτες Παγκόσμιας Ανάπτυξης), β_i είναι οι συντελεστές παλινδρόμησης, και ϵ είναι ο όρος σφάλματος. Η εκτίμηση των συντελεστών γίνεται με τη μέθοδο των ελαχίστων τετραγώνων (MET).

Υποθέσεις για τα τυχαία σφάλματα:

- $\mathbb{E}(\epsilon_i) = 0$, σφάλματα με μηδενική μέση τιμή.
- $\text{Var}(\epsilon_i) = \sigma^2$, ομοσκεδαστικότητα (ίση διασπορά).

- $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$, ασυσχέτιστα τυχαία σφάλματα (το σφάλμα σε οποιαδήποτε δοκιμή δεν επηρεάζει τα σφάλματα άλλων δοκιμών).

Το παρατηρούμενο επίπεδο σημαντικότητας (observed level of significance ή **p-value**) ορίζεται ως η πιθανότητα να πάρει η ελεγχουσυνάρτηση τιμή τόσο ή περισσότερο ακραία από την παρατηρούμενη στο δείγμα μας.

Η διαδικασία βηματικής επιλογής προς τα πίσω (backward stepwise) ξεκινάει με ένα πλήρες μοντέλο που περιλαμβάνει όλες τις ανεξάρτητες μεταβλητές και, στη συνέχεια, αφαιρεί σταδιακά μεταβλητές με σκοπό την ελαχιστοποίηση του κριτηρίου AIC (Akaike Information Criterion). Σε κάθε βήμα, αφαιρείται η μεταβλητή που έχει τη μικρότερη επίδραση στο μοντέλο (δηλαδή, αυτή με το μεγαλύτερο p-value), και ο νέος AIC επανυπολογίζεται. Το AIC ορίζεται ως:

$$AIC = 2k - 2\ln(L), \quad (2)$$

όπου k είναι ο αριθμός των παραμέτρων στο μοντέλο και L είναι η μέγιστη τιμή της συνάρτησης πιθανοφάνειας του μοντέλου. Η διαδικασία συνεχίζεται μέχρι να φτάσουμε σε ένα σημείο όπου η αφαίρεση οποιασδήποτε άλλης μεταβλητής θα αυξήσει το AIC. Το τελικό μοντέλο με το μικρότερο AIC θεωρείται ως το βέλτιστο.

Η εγκυρότητα του μοντέλου αξιολογήθηκε με βάση τους δείκτες R^2 , R^2_{adj} , F -statistic και τη στατιστική σημαντικότητα των συντελεστών. Ο δείκτης R^2 μετρά το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής που εξηγείται από το μοντέλο. Το F -statistic χρησιμοποιείται για τον έλεγχο της συνολικής σημασίας του μοντέλου. Το R^2_{adj} είναι ένα τροποποιημένο μέτρο του συντελεστή προσδιορισμού το οποίο λαμβάνει υπόψη τον αριθμό των ανεξάρτητων μεταβλητών στο μοντέλο σε σχέση με τον αριθμό των παρατηρήσεων. Χρησιμοποιείται για να δώσει μια πιο ακριβή εκτίμηση της ποιότητας του μοντέλου όταν προστίθενται περισσότερες μεταβλητές. Ο κύριος λόγος χρήσης του R^2_{adj} είναι ότι το απλό R^2 μπορεί να είναι παραπλανητικό όταν έχουμε πολλά ανεξάρτητα μεταβλητά. Το R^2 πάντα αυξάνεται ή παραμένει σταθερό όταν προστίθενται νέες μεταβλητές στο μοντέλο, ακόμη και αν αυτές οι νέες μεταβλητές δεν έχουν ουσιαστική συνεισφορά στο μοντέλο. Αυτό μπορεί να οδηγήσει σε υπερβολική προσαρμογή (overfitting), όπου το μοντέλο γίνεται πολύ περίπλοκο και αποτυγχάνει να γενικεύσει καλά σε νέα δεδομένα.

Ανάλυση Διακύμανσης (ANOVA)

Για την εκτίμηση των διαφορών στις μέσες τιμές των δεικτών παγκόσμιας ανάπτυξης ανάμεσα στις κατηγορίες γονιμότητας, χρησιμοποιήθηκε η Ανάλυση Διακύμανσης (ANOVA). Η ANOVA επιτρέπει τον έλεγχο της ύπαρξης στατιστικά σημαντικών διαφορών ανάμεσα στις ομάδες.

Πίνακας 8: Πίνακας ANOVA

Πηγή μεταβλητότητας	SS	df	MS	F
Παράγοντας A	SSA	$m - 1$	$\frac{SSA}{m-1}$	$\frac{MSA}{MSE}$
Παράγοντας B	SSB	$l - 1$	$\frac{SSB}{l-1}$	$\frac{MSB}{MSE}$
Παράγοντας AB	SS(AB)	$(m - 1)(l - 1)$	$\frac{SS(AB)}{(m-1)(l-1)}$	$\frac{MS(AB)}{MSE}$
Σφάλμα (Error)	SSE	$ml(c - 1)$	$\frac{SSE}{ml(c-1)}$	-
Σύνολο (Total)	SST	$mlc - 1$	-	-

Η βασική υπόθεση της ANOVA είναι ότι οι ομάδες έχουν την ίδια διακύμανση και ότι οι παρατηρήσεις είναι ανεξάρτητες και κανονικά κατανομημένες. Η μέθοδος βασίζεται στη σύγκριση της διακύμανσης μεταξύ των ομάδων με τη διακύμανση εντός των ομάδων. Το στατιστικό μέγεθος που χρησιμοποιείται είναι το F -statistic, το οποίο υπολογίζεται ως:

$$F = \frac{\text{Μεταβλητότητα μεταξύ των ομάδων}}{\text{Μεταβλητότητα εντός των ομάδων}}. \quad (3)$$

Αν το F -statistic είναι μεγαλύτερο από μια κρίσιμη τιμή, απορρίπτουμε τη μηδενική υπόθεση της ισότητας των μέσων όρων. Για την ανάλυση χρησιμοποιήθηκε η συνάρτηση `aov` στη γλώσσα R, και τα αποτελέσματα ερμηνεύθηκαν με βάση τις τιμές p -value και τα διαστήματα εμπιστοσύνης.

Η ανάλυση περιλαμβάνει την εκτίμηση των υποθέσεων της ανεξαρτησίας, της κανονικότητας των υπολοίπων και της ομοιογένειας των διακυμάνσεων. Η κανονικότητα ελέγχθηκε με το Q-Q plot και το Shapiro-Wilk test, ενώ η ομοιογένεια των διακυμάνσεων ελέγχθηκε με το Levene's test.

Σε περίπτωση σημαντικών διαφορών, διενεργήθηκε ανάλυση μετα-όρων (post-hoc analysis) με το τεστ του Tukey (Tukey's HSD test) για τον εντοπισμό των ζευγών των ομάδων που διαφέρουν σημαντικά.

$$HSD = q \sqrt{\frac{MS_{within}}{n}}, \quad (4)$$

όπου q είναι η σταθερά του Tukey, MS_{within} είναι η μέση τετραγωνική απόκλιση εντός των ομάδων, και n είναι το μέγεθος της ομάδας.

Με τις παραπάνω αναλύσεις επιδιώκεται η κατανόηση των σχέσεων και των διαφορών ανάμεσα στους δείκτες παγκόσμιας ανάπτυξης και τις κατηγορίες γονιμότητας και πληθωρισμού.

Αποτελέσματα

Σε αυτή την ενότητα, παρουσιάζονται τα αποτελέσματα της ανάλυσης γραμμικού μοντέλου και της ανάλυσης διακύμανσης (ANOVA).

Ανάλυση Γραμμικού Μοντέλου

Αρχικά, εκτιμήθηκε ένα πλήρες γραμμικό μοντέλο για την εξήγηση της μεταβλητής Έλεγχος Διαφθοράς (CorControl) από τις υπόλοιπες μεταβλητές, με εξαίρεση τις μεταβλητές Χώρα (Country), Πολιτική Σταθερότητα (PolStab) και Φωνή και Υπευθυνότητα (VoiceAcc). Ο παρακάτω πίνακας περιέχει τις μεταβλητές του γραμμικού μοντέλου και τους συντελεστές παλινδρόμησης.

Πίνακας 9: Συντελεστές του γραμμικού μοντέλου

Μεταβλητή	Συντελεστής
Intercept	-2.137×10^{-2}
ElectrAccess	-9.999×10^{-3}
CookAccess	1.065×10^{-3}
AgriLand	-2.645×10^{-3}
BirthRate	-5.516×10^{-2}
CO2	-1.750×10^{-2}
CompEdu	-4.110×10^{-2}
DeathRate	-1.034×10^{-2}
FoodExports	8.692×10^{-4}
Telephone	5.740×10^{-3}
Internet	1.248×10^{-2}
PopGrowth	-1.412×10^{-1}
BusinessTime	-1.464×10^{-3}
AdolFertRate	-1.233×10^{-2}
Precipitation	-4.492×10^{-8}
EmployerPerc	7.618×10^{-3}
FertRate	1.534×10^{-1}
GDPperc	2.725×10^{-3}
GDPdollars	2.670×10^{-5}
Inflation	-1.533×10^{-2}
WaterStress	-5.844×10^{-5}
LifeExp	1.342×10^{-2}
MortRate	4.353×10^{-3}
PopPerc14	6.442×10^{-2}
WomBusiness	7.241×10^{-3}
FertCat	2.863×10^{-1}
InflCat	4.377×10^{-2}

Το αποτέλεσμα του αρχικού μοντέλου είναι το εξής:

MSE (Mean Square Error)	0.4595
R^2	0.8228
R^2_{adj}	0.7733
F-statistic	16.61
Βαθμοί Ελευθερίας (DF)	26 and 93
p-value	< 2.2e-16

Πίνακας 10: Αποτελέσματα γραμμικού μοντέλου

Στη συνέχεια, εφαρμόστηκε η διαδικασία βηματικής επιλογής προς τα πίσω (backward stepwise) για την επιλογή του βέλτιστου μοντέλου. Το τελικό μοντέλο που προέκυψε έχει τα εξής αποτελέσματα:

Min	1Q	Median	3Q	Max
-0.94538	-0.34818	-0.01533	0.28638	1.26908

Πίνακας 11: Κατάλοιπα του γραμμικού μοντέλου

Μεταβλητή	Εκτιμώμενη Τιμή	Τυπικό Σφάλμα	p-value
(Intercept)	-8.472×10^{-1}	6.708×10^{-1}	0.20933
ElectrAccess	-1.114×10^{-2}	3.619×10^{-3}	0.00263
AgriLand	-3.633×10^{-3}	2.227×10^{-3}	0.10569
CO2	-2.564×10^{-2}	1.354×10^{-2}	0.06097
Internet	1.305×10^{-2}	4.690×10^{-3}	0.00634
PopGrowth	-1.236×10^{-1}	6.079×10^{-2}	0.04440
AdolFertRate	-1.346×10^{-2}	4.175×10^{-3}	0.00167
GDPdollars	2.933×10^{-5}	3.586×10^{-6}	5.74×10^{-13}
PopPerc14	2.624×10^{-2}	1.349×10^{-2}	0.05440
WomBusiness	6.641×10^{-3}	3.267×10^{-3}	0.04449
FertCat	2.630×10^{-1}	1.896×10^{-1}	0.16836

Πίνακας 12: Συντελεστές βέλτιστου γραμμικού μοντέλου με ανεξάρτητες μεταβλητές το ElectrAccess, AgriLand, CO2, Internet, PopGrowth, AdolFertRate, GDPdollars, PopPerc14, WomBusiness και FertCat.

Residual Standard Error (RSE)	0.4426
Multiple R-squared	0.8073
Adjusted R-squared	0.7896
F-statistic	45.66
Degrees of Freedom (DF)	10 and 109
p-value	$< 2.2e - 16$

Πίνακας 13: Στατιστικά του γραμμικού μοντέλου

Μοντέλο	AIC
Αρχικό μοντέλο	179.3194
Backward Stepwise	157.3954

Πίνακας 14: Τιμές AIC για τα δύο μοντέλα

Στατιστικό Μέτρο	Τιμή
R-squared	0.807295
Adjusted R-squared	0.7896156

Πίνακας 15: Τιμές R-squared και Adjusted R-squared

Σχολιασμός: Το μοντέλο που προέκυψε με τη χρήση της μεθόδου backward stepwise έχει χαμηλότερη τιμή AIC (157.3954) σε σύγκριση με το αρχικό μοντέλο (179.3194). Αυτό υποδηλώνει ότι το μοντέλο με τη μείωση των μεταβλητών παρουσιάζει καλύτερη προσαρμογή στα δεδομένα.

Στατιστικό Μέτρο	Κάτω Όριο	Άνω Όριο
(Intercept)	-2.6059	0.9115
ElectrAccess	-0.0206	-0.0017
AgriLand	-0.0095	0.0022
CO2	-0.0612	0.0099
Internet	0.0008	0.0253
PopGrowth	-0.2830	0.0357
AdolFertRate	-0.0244	-0.0025
GDPdollars	0.00002	0.00004
PopPerc14	-0.0091	0.0616
WomBusiness	-0.0019	0.0152
FertCat	-0.2342	0.7602

Πίνακας 16: 99% Διαστήματα Εμπιστοσύνης

Σχολιασμός: Με βάση τα αποτελέσματα της ανάλυσής μας, παρατηρούμε ότι οι μεταβλητές Intercept, AgriLand, CO2, PopGrowth, PopPerc14, WomBusiness και FertCat δεν είναι στατιστικά σημαντικές σε επίπεδο σημαντικότητας 1% καθώς τα αντίστοιχα διαστήματα εμπιστοσύνης περιέχουν το 0. Ως εκ τούτου, δεν προτείνεται η χρήση αυτών των μεταβλητών στο γραμμικό μοντέλο μας.

Έλεγχος Στατιστικής Σημαντικότητας των ανεξάρτητων μεταβλητών ElectrAccess και PopGrowth

Κάνουμε τον έλεγχο υποθέσεων:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

Υπολογίζουμε τις επικουρικές συναρτήσεις T για τις ανεξάρτητες μεταβλητές ως εξής:

$$T_1 = \frac{\hat{\beta}_1}{\hat{S}_{\beta_1}} \quad \text{και} \quad T_2 = \frac{\hat{\beta}_2}{\hat{S}_{\beta_2}}$$

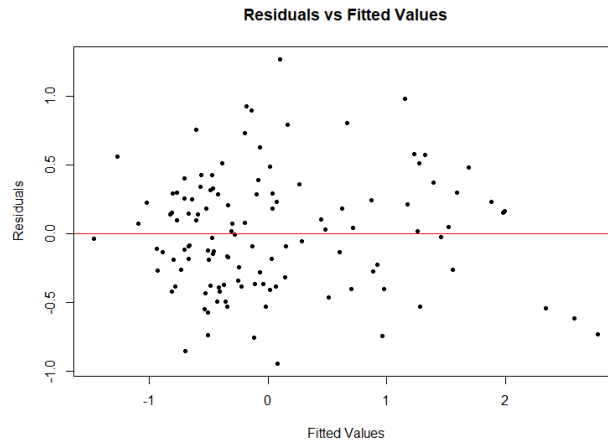
Οι τιμές T_1 και T_2 ακολουθούν την κατανομή t με 109 βαθμούς ελευθερίας. Υπολογίζουμε τις τιμές $T_1 = 2.033844$ και $T_2 = 3.078803$ και βλέπουμε ότι είναι μεγαλύτερες από την κρίσιμη τιμή $t_{\alpha/2;109} = 2.033844$ της κατανομής Student.

Άρα, απορρίπτουμε την μηδενική υπόθεση H_0 και συνεπώς οι ανεξάρτητες μεταβλητές ElectrAccess και PopGrowth είναι στατιστικά σημαντικές σε επίπεδο στατιστικής σημαντικότητας 5%.

Έλεγχος Υποθέσεων και Διαγνωστικά του Μοντέλου

Οι έλεγχοι για τις υποθέσεις του βέλτιστου μοντέλου περιλαμβάνουν:

- Ανεξαρτησία (Independance)
- Ομοσκεδαστικότητα (Homoscedasticity)
- Κανονικότητα (Normality)



Ανεξαρτησία: Για να εξετάσουμε την ανεξαρτησία των καταλοίπων φτιάχνουμε το γράφημα Residuals vs Fitted Values και παρατηρούμε ότι τα κατάλοιπα δεν έχουν κάποιο συγκεκριμένο μοτίβο και είναι τυχαία κατανεμημένα γύρω από το 0. Άρα ισχύει η υπόθεση της ανεξαρτησίας.

Ομοσκεδαστικότητα: Εκτελούμε το studentized Breusch-Pagan test, το οποίο μας δίνει αποτέλεσμα:

BP	df	p-value
15.321	10	0.1208

Με βάση το αποτέλεσμα του studentized Breusch-Pagan test, το οποίο έχει ένα p-value που είναι μεγαλύτερο του 0.05, δεν έχουμε αρκετά αποδεικτικά στοιχεία για να απορρίψουμε τη μηδενική υπόθεση. Συνεπώς, μπορούμε να συμπεράνουμε ότι δεν υπάρχουν στατιστικά σημαντικές αποκλίσεις στην διασπορά των καταλοίπων. Αυτό υποδηλώνει ότι το μοντέλο μας είναι ικανοποιητικό όσον αφορά την ομοσκεδαστικότητα, δηλαδή οι υποθέσεις της σταθερής διασποράς των σφαλμάτων ικανοποιούνται.

Κανονικότητα: Εκτελούμε το Shapiro-Wilk test για να ελέγξουμε αν τα κατάλοιπα είναι κανονικά κατανεμημένα, το οποίο μας δίνει αποτέλεσμα:

Shapiro-Wilk Test		Result
W statistic		0.99029
p-value		0.5603

Πίνακας 17: Shapiro-Wilk test για τα κατάλοιπα

Με βάση το αποτέλεσμα του Shapiro-Wilk test, το οποίο έχει ένα p-value που είναι μεγαλύτερο του 0.05, δεν έχουμε αρκετά αποδεικτικά στοιχεία για να απορρίψουμε τη μηδενική υπόθεση. Συνεπώς, μπορούμε να συμπεράνουμε ότι τα κατάλοιπα είναι κανονικά κατανεμημένα.

Συμπέρασμα: Οι υποθέσεις του γραμμικού μοντέλου μας ισχύουν.

Μέσα και Ατομικά Διαστήματα εμπιστοσύνης για την CorControl

Μετά την προσθήκη 8 νέων χωρών κατασκευάζουμε τα μέσα και τα ατομικά διαστήματα πρόβλεψης για τη μεταβλητή CorControl.

Χώρα	Προσαρμογή	Κάτω Όριο	Άνω Όριο
Κύπρος	0.8863817	0.7068971	1.0658664
Γεωργία	-0.3291076	-0.5043345	-0.1538807
Ελλάδα	0.4393572	0.2593801	0.6193342
Λουξεμβούργο	3.2028723	2.6563054	3.7494393
Μεξικό	-0.3314785	-0.5282404	-0.1347165
Φιλιππίνες	-0.4205245	-0.6668829	-0.1741661
Σλοβενία	0.7223933	0.5652714	0.8795153
Ουρουγουάη	0.1084105	-0.2202040	0.4370249

Πίνακας 18: Μέσα Διαστήματα Εμπιστοσύνης

Μεταβλητή	Προσαρμογή	Κάτω Όριο	Άνω Όριο
Κύπρος	0.8863817	-0.009041873	1.7818053
Γεωργία	-0.3291076	-1.223687467	0.5654723
Ελλάδα	0.4393572	-0.456165284	1.3348796
Λουξεμβούργο	3.2028723	2.169284339	4.2364603
Μεξικό	-0.3314785	-1.230524590	0.5675677
Φιλιππίνες	-0.4205245	-1.331711176	0.4906622
Σλοβενία	0.7223933	-0.168817063	1.6136037
Ουρουγουάη	0.1084105	-0.828369186	1.0451901

Πίνακας 19: Ατομικά Διαστήματα Πρόβλεψης

Ανάλυση ANOVA

Η ανάλυση διακύμανσης (ANOVA) πραγματοποιήθηκε για να εξεταστεί η επίδραση των κατηγοριών γονιμότητας (FertCat) και των κατηγοριών πληθωρισμού (InflCat) στην μεταβλητή Έλεγχος Διαφθοράς (CorControl). Τα αποτελέσματα της ανάλυσης ANOVA είναι:

	Df	Sum Sq	Mean Sq	F value	p-value
factor(FertCat)	2	30.27	15.135	24.081	1.89e-09
factor(InflCat)	1	5.22	5.217	8.301	0.00473
factor(FertCat):factor(InflCat)	2	3.68	1.840	2.927	0.05758
Residuals	114	71.65	0.628		

Πίνακας 20: Πίνακας ANOVA

Η αλληλεπίδραση μεταξύ των κατηγοριών γονιμότητας και πληθωρισμού δεν είναι στατιστικά σημαντική ($p\text{-value} > 0.05$). Ωστόσο, οι κύριες επιδράσεις των κατηγοριών γονιμότητας και πληθωρισμού είναι στατιστικά σημαντικές ($p\text{-value} < 0.05$).

Η ανάλυση post-hoc με τη δοκιμή Tukey έδειξε ότι οι διαφορές μεταξύ των κατηγοριών γονιμότητας είναι σημαντικές μεταξύ των κατηγοριών 1 και 3.

Σύγκριση	Διαφορά	Κάτω Όριο	Άνω Όριο	Προσαρμοσμένο p-value
2:0-1:0	-1.3789	-2.1821	-0.5757	0.0000337
3:0-1:0	-1.5786	-2.6837	-0.4735	0.0009247
1:1-1:0	-0.7000	-1.2421	-0.1579	0.0038025
2:1-1:0	-1.2873	-1.9155	-0.6591	0.0000005
3:1-1:0	-1.6739	-2.5823	-0.7655	0.0000070
3:0-2:0	-0.1996	-1.4391	1.0398	0.9971691
1:1-2:0	0.6789	-0.1014	1.4593	0.1263252
2:1-2:0	0.0916	-0.7508	0.9341	0.9995741
3:1-2:0	-0.2949	-1.3628	0.7729	0.9668524
1:1-3:0	0.8786	-0.2100	1.9672	0.1870854
2:1-3:0	0.2913	-0.8427	1.4252	0.9758164
3:1-3:0	-0.0953	-1.4054	1.2148	0.9999412
2:1-1:1	-0.5873	-1.1860	0.0114	0.0577981
3:1-1:1	-0.9739	-1.8621	-0.0857	0.0228767
3:1-2:1	-0.3866	-1.3299	0.5567	0.8417268

Πίνακας 21: Αποτελέσματα Post-hoc analysis TukeyHSD

Σχολιασμός:

- **Στατιστικά σημαντικές διαφορές ($p_{adj} < 0.05$):**
 - **2:0 vs 1:0:** Η ομάδα 2:0 έχει μέσο όρο 1.38 μονάδες χαμηλότερο από την ομάδα 1:0.
 - **3:0 vs 1:0:** Η ομάδα 3:0 έχει μέσο όρο 1.58 μονάδες χαμηλότερο από την ομάδα 1:0.
 - **1:1 vs 1:0:** Η ομάδα 1:1 έχει μέσο όρο 0.70 μονάδες χαμηλότερο από την ομάδα 1:0.
 - **2:1 vs 1:0:** Η ομάδα 2:1 έχει μέσο όρο 1.29 μονάδες χαμηλότερο από την ομάδα 1:0.
 - **3:1 vs 1:0:** Η ομάδα 3:1 έχει μέσο όρο 1.67 μονάδες χαμηλότερο από την ομάδα 1:0.
- **Μη στατιστικά σημαντικές διαφορές ($p_{adj} > 0.05$):**
 - **3:0 vs 2:0, 1:1 vs 2:0, 2:1 vs 2:0, 3:1 vs 2:0, 1:1 vs 3:0, 2:1 vs 3:0, 3:1 vs 3:0:**
Οι διαφορές μεταξύ αυτών των ομάδων δεν είναι στατιστικά σημαντικές.

Συζήτηση

Με βάση τα αποτελέσματα της ανάλυσης, προκύπτουν τα εξής συμπεράσματα:

- Η διαδικασία βηματικής επιλογής επέλεξε ένα πιο απλό μοντέλο με λιγότερες μεταβλητές, το οποίο είχε καλύτερη προσαρμογή (χαμηλότερο AIC) σε σχέση με το αρχικό μοντέλο.
- Οι συντελεστές ElectrAccess και PopGrowth του βέλτιστου γραμμικού μοντέλου επιβεβαιώσαμε ότι είναι στατιστικά σημαντικοί σε επίπεδο σημαντικότητας 5% μέσω δύο ελέγχων υποθέσεων.
- Η ανάλυση ANOVA έδειξε ότι οι κατηγορίες γονιμότητας και πληθωρισμού έχουν σημαντική επίδραση στον έλεγχο της διαφθοράς, ενώ η αλληλεπίδρασή τους δεν είναι στατιστικά σημαντική.

Τα αποτελέσματα ήταν σε μεγάλο βαθμό αναμενόμενα, δεδομένου ότι η βιβλιογραφία υποστηρίζει την επίδραση των κοινωνικοοικονομικών παραγόντων στον έλεγχο της διαφθοράς. Ωστόσο, υπάρχουν ορισμένοι περιορισμοί στις μεθόδους που χρησιμοποιήθηκαν. Η ανάλυση γραμμικού μοντέλου υποθέτει ανεξαρτησία, κανονικότητα και ομοσκεδαστικότητα των καταλοίπων, ενώ οι μετασχηματισμοί λογαρίθμου μπορεί να μην αποτυπώνουν πλήρως τη φύση των δεδομένων. Η ανάλυση ANOVA επίσης υποθέτει κανονικότητα των καταλοίπων και ομοσκεδαστικότητα, κάτι που μπορεί να επηρεάσει τα αποτελέσματα αν αυτές οι υποθέσεις δεν ισχύουν πλήρως.

Παρακάτω παρατίθενται οι 10 χώρες με τον μεγαλύτερο έλεγχο διαφθοράς:

Χώρα	Έλεγχος Διαφθοράς
Φιλανδία	2.171321
Δανία	2.156479
Νέα Ζηλανδία	2.135031
Σιγκαπούρη	2.134554
Σουηδία	2.109264
Νορβηγία	2.050117
Ελβετία	1.969518
Γερμανία	1.898325
Ολλανδία	1.888952
Ηνωμένο Βασίλειο	1.806572

Πίνακας 22: Οι 10 χώρες με τον μεγαλύτερο έλεγχο διαφθοράς

Σύμφωνα με τα προηγούμενα αποτελέσματα μας βρήκαμε μια σημειακή πρόβλεψη για τον Έλεγχο της Διαφθοράς στην Ελλάδα, η οποία είναι **0.4393572**. Παρατηρούμε ότι η τιμή αυτή είναι πολύ μικρότερη σε σχέση με τις 10 χώρες με τον μεγαλύτερο έλεγχο της διαφθοράς και συνεπώς υπάρχει μεγάλο περιθώριο βελτίωσης σε αυτόν τον τομέα.

Λαμβάνοντας υπόψη τη γραμμική εξάρτηση που παρατηρείται μεταξύ του ελέγχου διαφθοράς και των άλλων δεικτών ανάπτυξης, είναι λογικό να συμπεράνουμε ότι μπορούμε να επιτύχουμε βελτίωση στον έλεγχο διαφθοράς είτε με άμεση εφαρμογή μέτρων για την καταπολέμηση της διαφθοράς είτε με ενίσχυση άλλων δεικτών ανάπτυξης. Συγκεκριμένα, οι πρωτοβουλίες που προάγουν την οικονομική ανάπτυξη, την εκπαίδευση, την υγεία και την τεχνολογία, μπορούν να συμβάλουν έμμεσα αλλά ουσιαστικά στη βελτίωση του ελέγχου διαφθοράς. Έτσι, ένα ολοκληρωμένο και πολυδιάστατο πλάνο ανάπτυξης μπορεί να επιφέρει θετικά αποτελέσματα σε πολλαπλούς τομείς ταυτόχρονα, ενισχύοντας τη διαφάνεια και την ακεραιότητα σε εθνικό επίπεδο.

Αναφορές

References

- [1] Introduction to Linear Regression Analysis (hardcover), by D.C. Montgomery, E.A. Peck, and G.G. Vining (John Wiley and Sons).
- [2] Linear Regression Analysis (Wiley Series in Probability and Statistics) (hardcover), by G.A.F. Seber and A.J. Lee.

[3] Modern Regression Methods, T. P. Ryan, Wiley Series in Probability and Statistics