

ניתוח סטטיסטי של מאגרי נתונים טבלאיים

מדד החופש הכלכלי לשנת 2019

חלק ב'



מגישים:

רותם גרבר 311375455

יער אור 203506662

שקד חיימי 313235160

תוכן עניינים

4	תקציר מנהלים
4	עיבוד מקדים
4	הגדרת משתנים
5	הסרת משתנים
7	התאמת המשתנים
8	הגדרת משתני דמה
8	הוספת משתני אינטראקציה עבור משתני דמה
10	התאמת המודל ובדיקת הנחות המודל
13	בדיקת הנחות המודל
13	הנחת הנורמליות של השגיאות (נספח 6)
13	הנחת שיוויון שונות והנחת הלינאריות (נספח 7)
14	דוגמה לשימוש במודל הנבחר
14	בדיקת השערה המבוססת על תוצאות המודל
15	שיפור המודל
16	מסקנות והמלצות
18	נספחים
18	נספח 1: פלט R מובהקות- כל משתנה מסביר רציף בנפרד ביחס למוסבר :
20	נספח 2- כלל המשתנים המסבירים הרציפים יחדיו ביחס למוסבר :
21	נספח 3 – פיזור התצפיות של המשתנים החשודים להסרה :
22	נספח 4 – התאמת משתנים
23	נספח 5 – גרפים של משתני אינטראקציה
25	נספח 6 – בדיקת הנחות - הנחת הנורמליות של השגיאות
25	נספח 7 – הנחת שוויון שונות והנחת לינאריות
26	טבלאות
26	טבלה 1 – מובהקות המשתנים המסבירים מול המשתנה המוסבר
26	טבלה 2 – טבלה מסכמת של מדד פירסון (פלט R הוצג בסעיף 2)
27	פלט
27	חלק 1 – המודל המקורי
28	חלק 2 – רגרסיה לאחר
28	AIC
31	BIC
33	חלק 3 – רגרסיה לפנים
33	AIC

37BIC
40 חלק אחרון – רגרסיה בצעדים
40AIC
45BIC
49 המודל הסופי – לפני שיפור
51 בחינת המודל הסופי – לאחר שיפור
52 חלק 5 - רגרסיה לאחר
52AIC
53BIC
54 חלק 6 - רגרסיה לפנים
54AIC
58BIC
61 חלק 7 - רגרסיה בצעדים
61AIC
66BIC

תקציר מנהלים

פרויקט זה בוחן את הגורמים המשפיעים על דירוג החופש הכלכלי שנמדד במדינות רבות בעולם ע"י Heritage foundation. לשם ביצוע חקר מעמיק בנושא אספנו נתונים על 180 מדינות, מידע על פקטורים שלדעתנו עשויים להשפיע

על המשתנה המוסבר שלנו. לאחר מחשבה, החלטנו להוציא אחת מן המדינות (צפון קוריה) מן הנתונים, בשל הנתונים הקיצוניים בכל המשתנים המסבירים. במהלך החקר עלו מספר גורמים העשויים להשפיע: משתנים רציפים: זכויות רכוש, יעילות משפטית, יושרה ממשלתית, נטל המס, הוצאות ממשלתיות, בריאות פיסקלית, חופש עיסקי, חופש עבודה, חופש ממנוי וחופש מסחר. משתנים קטגוריאליים: חופש השקעות, חופש פיננסי, יבשת וליגלציה בנושא אלימות. לשם חיזוי יעיל ככל האפשר, בנינו מודל תוך כדי שימוש בכלי רגרסיה ליניארית מרובת משתנים ובעזרתו אמדנו את השפעת המשתנים המסבירים על מדד החופש הכלכלי וכמו כן, ביצענו חיזוי נתונים עתידיים. בעזרת כלים סטטיסטיים כגון מבחני השערות ושיטות לבחינת טיב המודל, נמדדה מידת השפעתו של כל משתנה הן כיחיד והן כחלק מהכלל. בכדי ליצור מודל אידיאלי ככל האפשר, בחנו את המשתנים המשפיעים על המשתנה המוסבר שלנו ואת אלו שנראו כי לא משפיעים או משפיעים במידה מועטה הסרנו מהמודל. לאחר מכן בדקנו האם יש צורך בהגדרה מחדש של משתנים, עבור משתנים קטגוריאליים ניתן לבצע איחוד קטגוריות במידה וקיימות קטגוריות בעלות מספר מופעים קטן. לאחר בחינה במספר דרכים – החלטנו להוציא את המשתנה הוצאות ממשלתיות בגלל שהקשר בינו לבין המשתנה המוסבר חלש מאוד. בדקנו משתני דמה והוספנו משתני אינטראקציה, בחלק זה רצינו למצוא השפעה של משתנים קטגוריאליים על שיפוע קו הרגרסיה במודל שלנו עבור שילוב המשתנה הקטגוריאלי עם משתנים מסבירים שונים. המודלים נבדקו ונבחנו בהתאם לשיטות שלמדנו. בעזרת תוכנת R ומספר מדדים, נערכו השוואות בין המודלים השונים, כאשר המדד המרכזי היה ה- R_{adj}^2 . לאחר ביצוע טרנספורמציות על המשתנים המסבירים וביצוע טרנספורמציה על המשתנה המוסבר, קיבלנו כי המודל הסופי שלנו הוא:

$$\begin{aligned} \hat{Y} = & \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 C_5 + \hat{\beta}_6 X_6 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8 + \hat{\beta}_9 X_9 + \hat{\beta}_{10} X_{10} \\ & + \hat{\beta}_{11} C_1 + \hat{\beta}_{12} C_2 + \hat{\beta}_{13} C_3 + \hat{\beta}_{14} C_4 + \hat{\beta}_{15} FF_1 + \hat{\beta}_{16} FF_2 + \hat{\beta}_{17} FF_3 + \hat{\beta}_{18} IF_1 \\ & + \hat{\beta}_{19} IF_2 + \hat{\beta}_{20} IF_3 + \sum_{i=1}^3 \hat{\beta}_{20+i-1} X_3 * FF_i + \sum_{i=1}^5 \hat{\beta}_{23+j-1} X_9 * C_i \end{aligned}$$

עיבוד מקדים

הגדרת משתנים

בחלק זה התבקשנו להוסיף שני משתנים קטגוריאליים.

בחרנו להוסיף את המשתנים:

X₁₃ - זהו משתנה בינארי (0 או 1) אשר מציין האם היית קיימת חקיקה בתחום אלימות במשפחה בשנת 2018, שנה לפני שנת אסיפת הנתונים.

X₁₄ - זהו משתנה המתאר את היבשת של אותה מדינה. ישנם סה"כ 8 אפשרויות - Asia, Europe, North America, Africa, South America, Oceania, Central America, Middle East

סימון	תיאור המשתנה	מסבילמוס בר	סוג משתנה	יחידות מידה	תחום ערכים
Y	Overall Score ציון	מוסבר	רציף	מספרי	ערכים בין 0-100. ככל שהמספר גדול יותר, מדד החופש הכלכלי במדינה גבוה יותר.
X1	זכויות רכוש Property Rights	מסביר	רציף	מספרי	ערכים בין 0-100. ככל שההגנה החוקית על רכוש מאובטחת יותר, ציון המדינה גבוה יותר.
X2	יעילות משפטית Judicial Effectiveness	מסביר	רציף	מספרי	ערכים בין 0-100. ככל שהיעילות המשפטית נחשבת טובה יותר, ציון המדינה גבוה יותר.
X3	ישרה ממשלתית Government Integrity	מסביר	רציף	מספרי	ערכים בין 0-100. ככל שהממשלה נחשבת יותר מושחתת, ציון המדינה נמוך יותר.
X4	נטל מס Tax Burden on	מסביר	רציף	מספרי	ערכים בין 0-100. ככל שנטל המס נמוך יותר, ציון המדינה גבוה יותר.
X5	הוצאות ממשלתיות Government Spending	מסביר	רציף	מספרי	ערכים בין 0-100. ככל שההוצאות הממשלתיות גבוהות יותר, ציון המדינה יהיה נמוך יותר.
X6	בריאות פסקית Fiscal Health	מסביר	רציף	מספרי	ערכים בין 0-100. ככל שהיכולת של הממשלה גבוהה יותר, ציון המדינה יהיה גבוה יותר.
X7	חופש עסקי Business Freedom	מסביר	רציף	מספרי	ערכים בין 0-100. ככל שיש יותר חופש עסקי, ציון המדינה יהיה גבוה יותר.
X8	חופש העבודה Labor Freedom	מסביר	רציף	מספרי	ערכים בין 0-100. ככל שיש יותר חופש עבודה, ציון המדינה יהיה גבוה יותר.
X9	חופש ממוני Monetary Freedom	מסביר	רציף	מספרי	ערכים בין 0-100. ככל שיש יותר חופש ממוני, ציון המדינה יהיה גבוה יותר.
X10	חופש מסחר Trade Freedom	מסביר	רציף	מספרי	ערכים בין 0-100. ככל שיש יותר חופש מסחר, ציון המדינה יהיה גבוה יותר.
X11	חופש השקעות Investment Freedom	מסביר	קטגוריאל	מספרי	טווח ערכים 0-100 בקפיצות של 5. ככל שיש יותר חופש השקעות, ציון המדינה יהיה גבוה יותר.
X12	חופש פיננסי Financial Freedom	מסביר	קטגוריאל	מספרי	טווח ערכים 0-100 בקפיצות של 10. ככל שיש יותר חופש פיננסי, ציון המדינה יהיה גבוה יותר.
X13	Legislation exists on domestic violence in 2018 אלימות במשפחה בשנה לפני (2018)	מסביר	קטגוריאל	בינארי	0 - לא הייתה קיימת חקיקה, 1 - הייתה קיימת חקיקה
X14	Continent יבשת	מסביר	קטגוריאל	מלל	8 ערכים אפשריים: Asia, Europe, North America, Africa, South America, Central America, Oceania, Middle East

הסרת משתנים

החלטנו לבחון הסרת משתנים על פי רמת המובהקות שלהם ביחס למשתנה המוסבר כאשר רק הם במודל, רמת המובהקות שלהם ביחס למוסבר במודל המלא, מתאם פירסון ותרשימי פיזור. תחילה בדקנו את מקדם המתאם של פירסון עבור כל משתנה מסביר רציף עם המשתנה המוסבר וכתוצאה מכך הגענו אל מדד טיב ההתאמה של המשתנים. מדד ההתאמה מאפשר לבחון עד כמה המודל מדויק, ועד כמה המשתנה המסביר אכן מסביר את המשתנה המוסבר. המדד אינו רלוונטי עבור המשתנים קטגוריאליים. ניתן לראות בהמשך כי כלל המשתנים אינם בעלי מתאם גבוה ע"פ מדד פירסון, על כן בבדיקתנו אלו משתנים נרצה להשאיר ואילו להסיר נתחשב בכלל המדדים לבדיקת מובהקות.

מובהקות כל משתנה מסביר בנפרד אל מול המשתנה המוסבר ([נספח 1](#), [טבלה 1](#)).

על פי טבלה זו נראה כי לכלל המשתנים ישנו ערך מובהקות גבוה ביחס למוסבר, מלבד ל- X_4 (Tax Burden).

מובהקות כל המשתנים יחד ביחס למשתנה המוסבר ([נספח 2](#)).

על פי טבלה זו נראה שכאשר המשתנים משולבים יחדיו, ישנו ערך מובהקות גבוהה מאד אצל כלל המשתנים המסבירים ביחס למשתנה המוסבר.

מתאם בין מסבירים (מולטיקולינאריות) – המתאם שבין המשתנים המסבירים לבין עצמם. בדיקה זו בצענו באמצעות מתאם פירסון. ערך המדד נע בין הערכים 1 ל- (-1). ככל שערכו המוחלט של המדד קרוב יותר ל-1, כך המדד יצביע על קשר לינארי חזק יותר בין המשתנים. כמו כן, מקדם מתאם פירסון נמוך של משתנה מסביר ביחס למשתנה מוסבר, מעיד על כך שמשתנה מסביר זה איננו מסביר טוב (יחסית) את המשתנה המוסבר ולכן גם אותו נבחר להסיר. קיבלנו שבין המשתנים:

Property_Rights - Government_Integrity - Judicial_Effectiveness

יש מתאם גבוה יחסית כלומר משפיעים בצורה דומה על המוסבר. (טבלה 2-טבלה מסכמת)

	Overall_Score	Property_Rights	Judicial_Effectiveness	Government_Integrity	Tax_Burden	Government_Spending	Fiscal_Health	Business_Freedom	Labor_Freedom	Monetary_Freedom	Trade_Freedom
Overall_Score	1.00000000	0.8304095	0.8042531	0.7637149	0.1372788	0.04546061	0.56317102	0.81252325	0.57963462	0.64507606	0.7530860
Property_Rights	0.8304095	1.0000000	0.8204186	0.8536377	-0.15322541	-0.27820102	0.32082636	0.77061720	0.42640167	0.45586736	0.6524350
Judicial_Effectiveness	0.8042531	0.8204186	1.0000000	0.8759106	-0.11376091	-0.15846872	0.30781660	0.72036601	0.45643962	0.43026649	0.5626601
Government_Integrity	0.7637149	0.8536377	0.8759106	1.0000000	-0.21862805	-0.33257065	0.29298057	0.70668191	0.40731437	0.40090994	0.5400178
Tax_Burden	0.1372788	-0.1532254	-0.1137609	-0.2186281	1.0000000	0.39391182	0.01627909	0.09312759	0.25724819	0.12743066	0.1771228
Government_Spending	0.04546061	-0.2782010	-0.1584687	-0.3325707	0.39391182	1.0000000	0.06898491	-0.17044711	0.04980025	-0.03556782	-0.1306426
Fiscal_Health	0.56317102	0.3208264	0.3078166	0.2929806	0.01627909	0.06898491	1.0000000	0.32395826	0.16912528	0.30886298	0.3140183
Business_Freedom	0.81252325	0.7706172	0.7203660	0.7066819	0.09312759	-0.17044711	0.32395826	1.0000000	0.51211245	0.47064838	0.6557584
Labor_Freedom	0.57963462	0.4264017	0.4564396	0.4073144	0.25724819	0.04980025	0.16912528	0.51211245	1.0000000	0.35858312	0.3990649
Monetary_Freedom	0.64507606	0.4558674	0.4302665	0.4009099	0.12743066	-0.03556782	0.30886298	0.47064838	0.35858312	1.0000000	0.5190576
Trade_Freedom	0.75308603	0.6524350	0.5626601	0.5400178	0.17712281	-0.13064264	0.31401831	0.65575838	0.39906489	0.51905760	1.0000000

כתוצאה מניתוח זה העלנו מספר משתנים החשודים להסרה:

X1- Property_Rights , X2- Government_Integrity , X3- Judicial_Effectiveness

תחילה, על מנת לבחון באופן ויזואלי את הקשר הלינארי בין כל משתנה מסביר רציף לבין המשתנה המוסבר, יצרנו תרשימי פיזור (נספח 4).

במבט ראשוני נראה כי הפיזור דיי דומה, עם שיפוע חיובי (יש קשר לינארי מסיים בין המשתנים למוסבר).

בנוסף, משתנים אלו בעלי ערך מדד פירסון גבוה ביחס למשתנה המוסבר:

	Overall_Score	Property_Rights	Judicial_Effectiveness	Government_Integrity
Overall_Score	1.00000000	0.8304095	0.8042531	0.7637149

אלמנט נוסף הוא שערך המובהקות של משתנים אלו גבוה, גם במודל המלא וגם כל משתנה בנפרד ביחס למוסבר. דבר המצביע על מובהקות המשתנה.

ועל כן החלטנו לא להסיר משתנים אלו- הם יוכלו לספר לנו מידע רב על המוסבר.

x5 – Government Spending משתנה נוסף שחשוד להסרה הינו משתנה:

תחילה נרשם מפיזור הנתונים ביחס למשתנה המוסבר. ניתן לא קיים קשר לינארי משמעותי הגורם לנו

להתרשם כי משתנה זה יכול לתרום למודל. כעת,

נרצה לראות מספרית ובאופן מדויק יותר את הקשר

בין המשתנה X5 למשתנה המוסבר.

ערך מדד פירסון של משתנה זה הינו **0.04546061**

ערך רמת המובהקות של משתנה זה בלבד ביחס ל Y

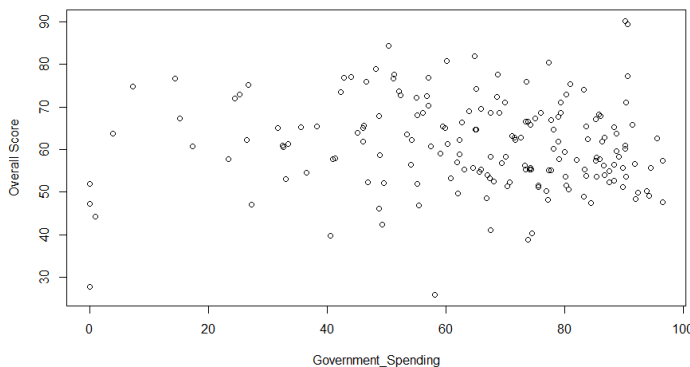
היא: **0.5445**

ערך רמת המובהקות של משתנה זה בצירוף שאר

המשתנים ביחס ל Y היא: **2e-16 *****

שילוב ערכים אלו עם ההתרשמות מגרף הפיזור

הביא אותנו למסקנה כי עלינו להסיר משתנה זה.



תרשים פיזור של משתנה 'הוצאות ממשלתיות' ביחס למשתנה המוסבר

התאמת המשתנים

בסעיף זה התבקשנו לבדוק האם יש צורך להגדיר מחדש משתנים.

תחילה נרצה לתקן את המשתנים הבדידים מחלק א'. את המשתנים הללו (X_{11} , X_{12}) ניתחנו בצורה שגויה ועל כן ראינו לנכון לתקנם ולנתחם מחדש. תחילה נראה את הפיזור של משתנים אלו ([נספח 4, גרפים 4-7](#)).

הפיכת משתנה בדיד לקטגוריאלי:

X_{11} - Investment Freedom

החלטנו לחלק את משתנה זה מחדש על פי דמיון יחסי בפיזור התצפיות (מדינות).

קבוצה 1- $0 \leq X_{11} < 20$, קבוצה 2- $20 \leq X_{11} < 60$, קבוצה 3- $60 \leq X_{11} < 80$, קבוצה 4- $80 \leq X_{11} < 100$

הסבר:

בקבוצה 1 ישנם 2 תוצאות חריגות של מדינות וונצואלה וקובה. אלו מדינות בעלות ציון נמוך משאר מדינות העולם באופן חריג. ועל כן ישנה שונות יחסית גבוהה בתוצאות קטגוריה 1. מלבד זה התוצאות הראשיות מכילות שונות גבוהה בתוך הקטגוריות עצמן. ולכן נצפה לשונות מעט גבוה יותר לאחר האיחוד.

קטגוריה	1	2	3	4
סטיית תקן	11.01551	6.351183	6.450761	7.950399

X_{12} - Financial Freedom

החלטנו לחלק את משתנה זה על פי דמיון יחסי בפיזור התצפיות (מדינות).

קבוצה 1- $0 \leq X_{12} < 20$, קבוצה 2- $20 \leq X_{12} < 60$, קבוצה 3- $60 \leq X_{12} < 80$, קבוצה 4- $80 \leq X_{12} < 100$

הסבר:

בקבוצה 1 ישנם 2 תוצאות חריגות של מדינות וונצואלה וקובה. אלו מדינות בעלות ציון נמוך משאר מדינות העולם באופן חריג. ועל כן ישנה שונות יחסית גבוהה בתוצאות קטגוריה 1. מלבד זה התוצאות הראשיות מכילות שונות גבוהה בתוך הקטגוריות עצמן. ולכן נצפה לשונות מעט גבוה יותר לאחר האיחוד.

קטגוריה	1	2	3	4
סטיית תקן	9.31205	6.602309	5.837354	6.078988

סיבה לשלילת איחוד קטגוריות במשתנה "יבשת": בנוסף החלטנו לשלול איחוד קטגוריות המשתנה "יבשת" משתי סיבות עיקריות. הראשונה כתוצאה ממשמעות המשתנה- להערכתנו מיקום גאוגרפי הינו היבט מכריע. ניתן להעריך כי ציון המדינות ביבשת אפריקה יהיה נמוך ממדינות אירופה למשל. כל גם מדינות דרום אמריקה אל מול צפון אמריקה. הסיבה השניה היא כמותית. ניתן להסתכל על פיזור התצפיות ([נספח 4, גרף 8](#)) ולהצדיק את טענתנו. גם אם נראה כי קטגוריות 5, 6 ניתנות לאיחוד שכן ממוצע התוצאות יחסית דומה, ואם נרצה לאחד אותם על סמך כמות תצפיות גם כן נענה בחיוב. אך כשמסתכלים על משמעות האיחוד- יבשת דרום אמריקה ויבשת אוקיינוס, ניתן להבין כי איחוד הקטגוריות יכול להיות שגוי. מה גם ששונות הקבוצה החדשה תהיה גבוהה מאד. מדינות כמו אוסטרליה וניו זילנד השייכות לקטגוריה 5 ימשכו את פריסת התוצאות כלפי מעלה, ובניגוד אליהן, מדינות כמו וונצואלה ובוליביה- ימשכו מטה.

סיבה לאיחוד קטגוריות במשתנה "יבשת":

כאשר מסתכלים על קטגוריות 7 (מרכז אמריקה) ו-6 (דרום אמריקה) גם המשמעות (מיקום גיאוגרפי דומה וברובן המכריע מדינות בעלות אותו סוג כלכלה ואותה רמת פיתוח) וגם פיזור הנתונים מצדיק את איחודם.

בנוסף, קטגוריות 3 (אירופה) ו-4 (צפון אמריקה) גם דומות במשמעותן (מדינות מפותחות) וגם פיזור הנתונים מצדיק איחוד ביניהן. בנוסף, קטגוריית צפון אמריקה כוללת רק שתי מדינות – ארה"ב וקנדה. לכן כדאי לאחד אותה עם קטגוריית אירופה שבה יש יותר רשומות.

לסיכום, החלטנו לאחד את קטגוריות 3 ו-4 ואת קטגוריות 6 ו-7.

הגדרת משתני דמה

משתנה דמה הינו משתנה המביע את התרומה השולית של המשתנים (אשר לא נמצאים בקבוצת הבסיס) על החותך, נשתמש במשתנה זה בכדי לשלב את המשתנים הקטגוריאליים בהגדרת משתנים מסבירים. עבור שלושת המשתנים הקטגוריאליים אשר הוגדרו והותאמו מחדש, נגדיר משתני דמה חדשים, על מנת להבחין בקשר אשר כל אחת מן הקטגוריות מביעה. נבצע זאת באמצעות הוספת משתנים בינאריים כמספר הקטגוריות בכל משתנה פחות 1, זאת כדי שלא נקבל מולטי-קולינאריות מלאה יחד עם החותך.

חופש פיננסי – משתנה זה הינו משתנה קטגוריאלי. במקור היו לו 10 קטגוריות, אך לאחר איחוד חלק מן הקבוצות חילקנו את המשתנה ל-4 קטגוריות. קבוצת הבסיס עבור משתנה זה תהיה חופש נמוך (0-20)

$$FF_1 = \begin{cases} 1, & \text{if Financial_Freedom}_i \in [30 - 50] \\ 0, & \text{else} \end{cases} \quad FF_2 = \begin{cases} 1, & \text{if Financial_Freedom}_i \in [60 - 70] \\ 0, & \text{else} \end{cases}$$
$$FF_3 = \begin{cases} 1, & \text{if Financial_Freedom}_i \in [80 - 100] \\ 0, & \text{else} \end{cases}$$

חופש השקעה – משתנה זה הינו משתנה קטגוריאלי. במקור היו לו 20 קטגוריות, אך לאחר איחוד חלק מן הקבוצות חילקנו את המשתנה ל-4 קטגוריות. קבוצת הבסיס עבור משתנה זה תהיה 0-20.

$$IF_1 = \begin{cases} 1, & 25 - 55 \\ 0, & \text{else} \end{cases} \quad IF_2 = \begin{cases} 1, & 60 - 75 \\ 0, & \text{else} \end{cases} \quad IF_3 = \begin{cases} 1, & 80 - 100 \\ 0, & \text{else} \end{cases}$$

יבשת – משתנה זה הוא משתנה קטגוריאלי. קבוצת הבסיס עבור משתנה זה תהיה **Africa**.

$$C_1 = \begin{cases} 1, & \text{if Continent} = "Asia" \\ 0, & \text{else} \end{cases} \quad C_2 = \begin{cases} 1, & \text{if Continent} = "Middle East" \\ 0, & \text{else} \end{cases}$$

$$C_3 = \begin{cases} 1, & \text{if Continent} = "Central America" \text{ or } "South America" \\ 0, & \text{else} \end{cases}$$

$$C_4 = \begin{cases} 1, & \text{if Continent} = "Oceania" \\ 0, & \text{else} \end{cases} \quad C_5 = \begin{cases} 1, & \text{if Continent} = "North America" \text{ or } "Europe" \\ 0, & \text{else} \end{cases}$$

חקיקה נגד אלימות - משתנה **x13** הוא משתנה בינארי בפני עצמו. נשתמש בו כמשתנה קטגוריאלי תחת התנאים הבאים:

$$\text{Legislation_Violence}_i = X13_i = \begin{cases} 1, & \text{if Legislation exists violence in 2018 in country } i \\ 0, & \text{else} \end{cases}$$

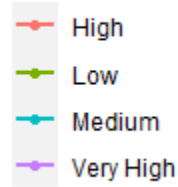
הוספת משתני אינטראקציה עבור משתני דמה

משתנה אינטראקציה הוא משתנה המציין תרומה שולית על השיפוע, עבור שילוב של שני משתנים מסבירים, ע"י מכפלה, על המשתנה המוסבר. בחלק זה, נרצה למצוא השפעה של משתני הדמה על שיפוע קו ההרגרסיה

במודל שלנו עבור כל משתנה דמה מול המשתנים המסבירים ע"י תרשימי פיזור. בבדיקה זו בחנו את האינטרקציה בין המשתנים הקטגוריאליים והרציפים ואת ההבדל בהשפעה שלהם על המשתנה המוסבר (גרפים [בנספח 5](#)).

להלן החלוקה שביצענו עבור המשתנה הקטגוריאלי חופש פיננסי –

Financial Freedom



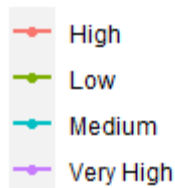
יעילות משפטית – עפ"י הגרף ניתן לראות כי בכל הקטגוריות ישנו שיפוע חיובי, כלומר ככל שהיעילות המשפטית גדלה, כך גם מדד החופש הכלכלי גדל. בנוסף, ניתן לראות גם כי כל שהחופש הפיננסי עולה, כך מדד החופש הכלכלי עולה לפי שינוי הצבעים בגרף. עבור הקטגוריה "High" ניתן לראות כי השיפוע תלול יותר. ניתן לשער כי במדינות בהן יש חופש פיננסי, ישנה גם מערכת משפטית יעילה יותר המספקת הגנה לתושבים בנושאים הפיננסיים השונים. בחרנו לא להוסיף משתנה אינטררציה מכיוון ונראה לנו כי ההשפעה על המשתנה המוסבר היא שולית.

יושרה ממשלתית –

עפ"י הגרף ניתן לראות כי בכל הקטגוריות ישנו שיפוע חיובי, כלומר ככל שהיושרה הממשלתית נחשבת ברמה גבוהה יותר, כך גם מדד החופש הכלכלי גדל בהתאם. בנוסף, ניתן לראות גם כי כל שהחופש הפיננסי עולה, כך היושרה הממשלתית ומדד החופש הכלכלי עולים שניהם, לפי שינוי הצבעים בגרף. ניתן לשער כי כאשר יש שחיתות בממשלה, קיימת יותר התערבות מצידה בתחומים פיננסיים לטובת אינטרסים אישיים, וכך היא מונעת את החופש הפיננסי מתושבי המדינה. ניתן לראות כי בקטגוריה "High" קיים שיפוע תלול יותר משאר הקטגוריות. לכן נוסיף את משתנה אינטרקציה זה למודל ונגדיר אותו באופן הבא:

$$X_3 * FF_2 = \begin{cases} X_3 * FF_2, & \text{If the financial freedom is "High"} \\ 0, & \text{else} \end{cases}$$

Investment Freedom



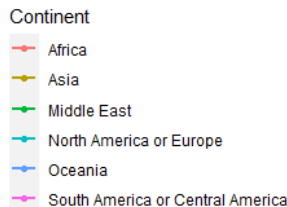
להלן החלוקה שביצענו עבור המשתנה הקטגוריאלי חופש השקעות –

חופש עיסקי –

עפ"י הגרף ניתן לראות כי בכל הקטגוריות ישנו שיפוע חיובי, כלומר ככל שהחופש העיסקי עולה, כך גם מדד החופש הכלכלי גדל בהתאם. בנוסף, ניתן לראות גם כי כל שהחופש הפיננסי עולה בהתאמה עם החופש העיסקי וגם עם מדד החופש הכלכלי. זאת מכיוון שכאשר יש חופש עיסקי גבוה במדינה, חופש ההשקעה חייב להיות גם הוא גבוה בהתאם, על מנת לאפשר השקעות בעסקים חדשים היאפשרו להם לגדול ולצמוח. ללא חופש ההשקעות, לעסקים חדשים יהיה קשה לקום להרוויח בתחילתם ולכן הם חייבים לבוא יחדיו. החלטנו שלא להוסיף משתנה אינטראקציה זה למודל, שכן אנו סבורים כי תרומתו השולית לשיפוע אינה משמעותית.

יעילות משפטית – עפ"י הגרף ניתן לראות כי בכל הקטגוריות ישנו שיפוע חיובי, כלומר ככל שהיעילות המשפטית גדלה, כך גם מדד החופש הכלכלי גדל. בנוסף, ניתן לראות גם כי כל שהחופש הפיננסי עולה, כך היעילות המשפטית וגם מדד החופש הכלכלי עולים לפי שינוי הצבעים בגרף. הדבר הגיוני מכיוון ותושבי המדינה ירגישו ביטחון רב יותר ויבצעו השקעות רבות יותר עם הידיעה כי קיימת מערכת משפט יעילה שתגן על האינטרסים של התשובים ועל סכומי הכסף שהשקיעו.

להלן החלוקה שביצענו עבור המשתנה הקטגוריאלי יבשת –



חופש ממוני – ניתן לראות כי לרוב היבשות רמת החופש הממוני די נמוכה, ועולה במתינות עם העלייה במדד החופש הכלכלי. יוצאת דופן היא הקטגוריה "South America or Central America" בה ניתן לראות שיפוע הרבה יותר תלול. זאת מכיוון שבמדינות דרום ומרכז אמריקה ישנה התערבות ממשלתית גדולה יותר בכלכלה ואינפלציה גבוהה יותר. ניתן לראות כי קיים נתון נמוך בצורה חריגה בשני המדדים המייצג את ונצואלה, מדינה בה ההתערבות הממשלתית והאינפלציה גבוהות מאוד. נוסיף משתנה אינטרקציה זה למודל ונגדיר אותו באופן הבא:

$$X_9 * C_2 = \begin{cases} X_9 * C_2, & \text{If the Continent is "South America" or "Central America"} \\ 0, & \text{else} \end{cases}$$

המודל שהתקבל הוא :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 C_5 + \hat{\beta}_6 X_6 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8 + \hat{\beta}_9 X_9 + \hat{\beta}_{10} X_{10} + \hat{\beta}_{11} C_1 + \hat{\beta}_{12} C_2 + \hat{\beta}_{13} C_3 + \hat{\beta}_{14} C_4 + \hat{\beta}_{15} FF_1 + \hat{\beta}_{16} FF_2 + \hat{\beta}_{17} FF_3 + \hat{\beta}_{18} IF_1 + \hat{\beta}_{19} IF_2 + \hat{\beta}_{20} IF_3 + \hat{\beta}_{21} X_{13} + \sum_{i=1}^3 \hat{\beta}_{22+i-1} X_3 * FF_i + \sum_{i=1}^5 \hat{\beta}_{25+j-1} X_9 * C_i$$

התאמת המודל ובדיקת הנחות המודל

בחירת משתני המודל

לצורך בחירת משתני המודל, נשתמש בשלושה אלגוריתמים ע"פ השיטות שנלמדו בהרצאות :

1. גרסיה לפניס – Forward Selection
2. גרסיה לאחור – Backward Elimination
3. גרסיה בצעדים – Stepwise Regression

נבחן את האלגוריתמים בעזרת המבחנים הבאים :

1. מבחן AIC - Akaike Information Criterion – נחשב את ערכי המדד AIC עבור כל משתנה בנפרד בכל איטרציה, כך שנבחר במודל בעל ערך ה-AIC המינימלי ביותר אשר קטן ממדד ה-AIC של המודל מהאיטרציה האחרונה. נעצור ונסיים כאשר ה-AIC המינימלי ביותר המתקבל גדול או שווה ל-AIC של המודל מהאיטרציה האחרונה.

$$\text{AIC} = n [\ln(\text{SSE}/n) + \ln(2\pi) + 1] + 2(k + 2) \text{ : מדד AIC מחושב בתוכנת ה-R באופן הבא}$$

2. מבחן BIC – Bayesian Information Criterion – נחשב את ערכי המדד BIC עבור כל משתנה בנפרד בכל איטרציה, כך שנבחר במודל בעל ערך ה-BIC המינימלי ביותר אשר קטן ממדד ה-BIC של המודל מהאיטרציה האחרונה. נעצור ונסיים כאשר ה-BIC המינימלי ביותר המתקבל גדול או שווה ל-BIC של המודל מהאיטרציה האחרונה.

$$\text{BIC} = n [\ln(\text{SSE}/n) + \ln(2\pi) + 1] + \ln(n)(k + 2) \text{ : מדד BIC מחושב בתוכנת ה-R באופן הבא}$$

Forward Selection – גרסיה לפניס

באלגוריתם זה נתחיל עם מודל ללא משתנים, כך שבכל איטרציה נוסיף למודל הרגרסיה את המשתנה המובהק ביותר.

מבחן AIC – בכל איטרציה, נחשב את מדד ה-AIC, עבור כל משתנה שמועמד להיכנס למודל, כך שבסוף נבחר את המשתנה בעל ערך ה-AIC המינימלי. אם עבור אחד המשתנים התקבל מדד AIC נמוך ממדד ה-AIC שהתקבל עבור המודל הקודם, נוסיף את המשתנה (שכן הוא משפר את המודל) ונמשיך בביצוע האיטרציות. במידה ולא, נעצור את ביצוע האיטרציות ונבחר במודל האחרון שהתקבל באיטרציה הקודמת כמודל הסופי.

לאחר ביצוע מבחן זה עבור גרסיה לפניס ([פלט ר](#)), התקבל המודל הבא :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_4 X_4 + \hat{\beta}_3 C_5 + \hat{\beta}_6 X_6 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8 + \hat{\beta}_9 X_9 + \hat{\beta}_{10} X_{10} + \hat{\beta}_{11} C_1 + \hat{\beta}_{12} C_2 + \hat{\beta}_{13} C_3 + \hat{\beta}_{14} C_4 + \hat{\beta}_{15} FF_1 + \hat{\beta}_{16} FF_2 + \hat{\beta}_{17} FF_3 + \hat{\beta}_{18} IF_1 + \hat{\beta}_{19} IF_2 + \hat{\beta}_{20} IF_3 + \sum_{i=1}^5 \hat{\beta}_{25+j-1} X_9 * C_i$$

מבחן BIC – בכל איטרציה, נחשב את מדד ה-BIC, עבור כל משתנה שמועמד להיכנס למודל, כך שבסוף נבחר את המשתנה בעל ערך ה-BIC המינימלי. אם עבור אחד המשתנים התקבל מדד BIC נמוך ממדד ה-BIC שהתקבל עבור המודל הקודם, נוסיף את המשתנה (שכן הוא משפר את המודל) ונמשיך בביצוע האיטרציות.

במידה ולא, נעצור את ביצוע האיטרציות ונבחר במודל האחרון שהתקבל באיטרציה הקודמת כמודל הסופי.

לאחר ביצוע מבחן זה עבור רגרסיה לפנים (פלטנים), התקבל המודל הבא :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_4 X_4 + \hat{\beta}_5 C_5 + \hat{\beta}_6 X_6 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8 + \hat{\beta}_9 X_9 + \hat{\beta}_{10} X_{10} + \hat{\beta}_{11} C_1 + \hat{\beta}_{12} C_2 + \hat{\beta}_{13} C_3 + \hat{\beta}_{14} C_4 + \hat{\beta}_{15} FF_1 + \hat{\beta}_{16} FF_2 + \hat{\beta}_{17} FF_3 + \hat{\beta}_{18} IF_1 + \hat{\beta}_{19} IF_2 + \hat{\beta}_{20} IF_3$$

Backward Selection – רגרסיה לאחור

באלגוריתם זה נתחיל עם מודל מלא המכיל את כלל המשתנים, כך שבכל איטרציה נבדוק את מובהקותו של כל אחד מהם ונשמיט את המשתנה בעל המובהקות הנמוכה ביותר.

מבחן AIC – בכל איטרציה, נחשב את מדד ה-AIC, עבור כל משתנה שמועמד להיכנס למודל, כך שבסוף נבחר את המשתנה בעל ערך ה-AIC המקסימלי. אם עבור אחד המשתנים התקבל מדד AIC גבוה ממדד ה-AIC שהתקבל עבור המודל הקודם, נוריד את המשתנה (שכן הוא גורם למודל להיות פחות טוב) ונמשיך בביצוע האיטרציות. במידה ולא, נעצור את ביצוע האיטרציות ונבחר במודל האחרון שהתקבל באיטרציה הקודמת כמודל הסופי.

לאחר ביצוע מבחן זה עבור רגרסיה לאחור (פלטנים), התקבל המודל הבא :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 C_5 + \hat{\beta}_6 X_6 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8 + \hat{\beta}_9 X_9 + \hat{\beta}_{10} X_{10} + \hat{\beta}_{11} C_1 + \hat{\beta}_{12} C_2 + \hat{\beta}_{13} C_3 + \hat{\beta}_{14} C_4 + \hat{\beta}_{15} FF_1 + \hat{\beta}_{16} FF_2 + \hat{\beta}_{17} FF_3 + \hat{\beta}_{18} IF_1 + \hat{\beta}_{19} IF_2 + \hat{\beta}_{20} IF_3 + \sum_{i=1}^3 \hat{\beta}_{21+i-1} X_3 * FF_i + \sum_{j=1}^5 \hat{\beta}_{24+j-1} X_9 * C_j$$

מבחן BIC – בכל איטרציה, נחשב את מדד ה-BIC, עבור כל משתנה שמועמד להיכנס למודל, כך שבסוף נבחר את המשתנה בעל ערך ה-BIC המקסימלי. אם עבור אחד המשתנים התקבל מדד BIC גבוה ממדד ה-BIC שהתקבל עבור המודל הקודם, נוריד את המשתנה (שכן הוא גורם למודל להיות פחות טוב) ונמשיך בביצוע האיטרציות. במידה ולא, נעצור את ביצוע האיטרציות ונבחר במודל האחרון שהתקבל באיטרציה הקודמת כמודל הסופי.

לאחר ביצוע מבחן זה עבור רגרסיה לאחור (פלטנים), התקבל המודל הבא :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_4 X_4 + \hat{\beta}_6 X_6 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8 + \hat{\beta}_9 X_9 + \hat{\beta}_{10} X_{10} + \hat{\beta}_{15} FF_1 + \hat{\beta}_{16} FF_2 + \hat{\beta}_{17} FF_3 + \hat{\beta}_{18} IF_1 + \hat{\beta}_{19} IF_2 + \hat{\beta}_{20} IF_3 + \hat{\beta}_5 * C_5 + \hat{\beta}_3 * C_3 + \hat{\beta}_{21} * C_1 + \hat{\beta}_{22} * C_2 + \hat{\beta}_{23} * C_4$$

Regression – רגרסיה בצעדים

אלגוריתם זה משלב את שתי השיטות הקודמות – רגרסיה לפנים ורגרסיה לאחור. נתחיל עם מודל ללא משתנים, כך שבכל איטרציה נבצע שתי פעולות – הוספת משתנה מסביר ובדיקת המודל החדש, והוצאת משתנים אשר לפי בדיקה אינם משפרים את המודל. לאחר השלמת האיטרציות, נקבל את המודל בעל המשתנים המתאימים ביותר עבור המשתנה המוסבר.

מבחן AIC – לאחר ביצוע מבחן זה עבור רגרסיה בצעדים (פלטנים), התקבל המודל הבא :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_4 X_4 + \hat{\beta}_5 C_5 + \hat{\beta}_6 X_6 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8 + \hat{\beta}_9 X_9 + \hat{\beta}_{10} X_{10} + \hat{\beta}_{11} C_1 + \hat{\beta}_{12} C_2 + \hat{\beta}_{13} C_3 + \hat{\beta}_{14} C_4 + \hat{\beta}_{15} FF_1 + \hat{\beta}_{16} FF_2 + \hat{\beta}_{17} FF_3 + \hat{\beta}_{18} IF_1 + \hat{\beta}_{19} IF_2 + \hat{\beta}_{20} IF_3 + \sum_{i=1}^5 \hat{\beta}_{25+i-1} X_9 * C_i$$

מבחן BIC – לאחר ביצוע מבחן זה עבור רגרסיה בצעדים (פלטנים), התקבל המודל הבא :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_4 X_4 + \hat{\beta}_6 X_6 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8 + \hat{\beta}_9 X_9 + \hat{\beta}_{10} X_{10} + \hat{\beta}_{15} FF_1 + \hat{\beta}_{16} FF_2 + \hat{\beta}_{17} FF_3 + \hat{\beta}_{18} IF_1 + \hat{\beta}_{19} IF_2 + \hat{\beta}_{20} IF_3 + \hat{\beta}_5 * C_5 + \hat{\beta}_3 * C_3 + \hat{\beta}_{21} * C_1 + \hat{\beta}_{22} * C_2 + \hat{\beta}_{23} * C_4$$

ניתוח תוצאות האלגוריתמים

מדד R^2 גדל עם עלייה במספר המשתנים המסבירים. לכן, את המודלים שקיבלנו נשווה על ידי מדד R^2_{adj} אשר מוטלים בו קנסות על הוספת משתנים מיותרים למודל. מדד R^2_{adj} מייצג את אחוז השונות המוסברת במודל הרגרסיה, תוך התמודדות עם השפעתה המלאכותית של כמות המשתנים המסבירים על ערך המדד.

נבחר במודל שערך מדד ה- R^2_{adj} שלו הוא מקסימלי.

דרכי חישוב המדדים –

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad \bullet \text{ מדד } R^2$$

$$R^2_{adj} = 1 - \frac{SSE / (n - k - 1)}{SST / (n - 1)} \quad \bullet \text{ מדד } R^2_{adj}$$

Stepwise Regression	Backward	Forward	מבחן
0.9769	0.9772	0.9769	AIC
0.9748	0.9748	0.9748	BIC

- **מבחן AIC:** לפי מבחן זה, מדדי ה- R^2_{adj} המתקבלים בשיטת הרגרסיה לפני (0.9769) וכן בשיטת הרגרסיה בצעדים (0.9769) קטנים מהערך המתקבל בשיטה רגרסיה לאחר (0.9772), זאת מכיוון ששני האלגוריתמים מחזירים את אותו המודל הסופי.
- **מבחן BIC:** לפי מבחן זה, מדד ה- R^2_{adj} המתקבל בשיטת הרגרסיה לאחר שווה לערך המתקבל בשיטה רגרסיה לאחר ורגרסיה בצעדים (0.9748) זאת מכיוון ששלושת האלגוריתמים מחזירים את אותו המודל הסופי.

המודל הסופי

המודל הסופי בו נבחר הינו המודל בעל מדד ה- R^2_{adj} הגדול ביותר שהתקבל מכלל השיטות. מודל זה התקבל עבור האלגוריתם רגרסיה לאחר לפי מבחן AIC – וערכו הוא 0.9772. המודל המתאים הוא:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 C_5 + \hat{\beta}_6 X_6 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8 + \hat{\beta}_9 X_9 + \hat{\beta}_{10} X_{10} + \hat{\beta}_{11} C_1 + \hat{\beta}_{12} C_2 + \hat{\beta}_{13} C_3 + \hat{\beta}_{14} C_4 + \hat{\beta}_{15} FF_1 + \hat{\beta}_{16} FF_2 + \hat{\beta}_{17} FF_3 + \hat{\beta}_{18} IF_1 + \hat{\beta}_{19} IF_2 + \hat{\beta}_3 IF_3 + \sum_{i=1}^3 \hat{\beta}_{20+i-1} X_3 * FF_i + \sum_{i=1}^5 \hat{\beta}_{23+j-1} X_9 * C_i$$

$$\begin{aligned}\hat{Y} = & 7.760369 + 0.117892 * X_1 + 0.129485 * X_2 - 0.052628 * X_3 + 0.119008 * X_4 \\ & - 10.968077 * C_5 + 0.098241 * X_6 + 0.080912 * X_7 + 0.092126 * X_8 \\ & + 0.045169 * X_9 + 0.088162 * X_{10} - 8.308124 * C_1 + 1.843556 * C_2 \\ & - 5.329688 * C_3 + 11.102173 * C_4 + 1.767749 * FF_1 + 1.733506 * FF_2 \\ & + 4.389707 * FF_3 + 2.444474 * IF_1 + 4.834075 * IF_2 + 6.140582 * IF_3 \\ & + 0.03228 * X_3 * FF_1 + 0.076302 * X_3 * FF_2 + 0.066511 * X_3 * FF_3 \\ & + 0.09913 * X_9 * C_1 - 0.044716 * X_9 * C_2 + 0.068376 * X_9 * C_3 - 0.175946 \\ & * X_9 * C_4 + 0.105151 * X_9 * C_5\end{aligned}$$

בדיקת הנחות המודל

בבדיקת הנחות המודל עלינו לנתח את הגרפים המתאימים ולהראות באמצעות המבחן המתאים האם נדחה את השערת האפס.

הנחת הנורמליות של השגיאות (נספח 6)

בכדי לבדוק הנחה זו נשתמש בתרשימים ומבחני התאמה.

בחנו שני תרשימים- היסטוגרמה וQ-Q Plot, שניהם של השגיאות המתוקננות.

בתרשים ה Q-Q Plot, ניתן לראות כי רובן המכריע של התצפיות נמצאות על הקו, עם מעט הסתייגות בצדדים. בהיסטוגרמה ניתן לראות כי פונקציית הצפיפות דומה מאוד לפעמון גאוס עם הטיה שמאלה. שני הדברים הללו מצביעים על כך שהשגיאות המתוקננות מגיעות מהתפלגות נורמלית.

one-sample Kolmogorov-Smirnov test

ממבחן קולמוגורב- סמירוב (KS):

data: stdResiduals
D = 0.064366, p-value = 0.4485
alternative hypothesis: two-sided

ניתן לראות כי $Pvalue > 0.05$. לכן, לא נדחה את השערת האפס, ונוכל להצהיר שהנתונים מגיעים מהתפלגות נורמלית.

הנחת שיוויון שוניות והנחת הלינאריות (נספח 7)

בכדי לבצע בדיקה זו, יצרנו גרף של פיזור השגיאות המתוקננות כפונקציה של Y (כובע):

מפיזור הנתונים ניתן להסיק כי הנחת שוויון השוניות מתקיימת. הנקודות מפוזרות מעל ומתחת קו ה"0" באופן יחסית שוויוני (רנדומלי- "רעש לבן"), מה גם שלא ניתן להצביע על מגמתיות כזו או אחרת. בכדי לקבוע זאת, נבצע מבחן לבדיקת שוויון השוניות, ונראה כי $Pvalue > 0.05$, לכן לא נדחה את השערת האפס ונגיד שהנחת שוויון השוניות מתקיימת. בבדיקת הנחת הלינאריות- בדיקה זו נעשתה גם כן על פי גרף הפיזור. ע"פ הגרף ניתן לראות כי השגיאות מפוזרות בצורה אקראית ויחסית סימטרית סביב התוחלת, לכן נסיק כי הנחת הלינאריות מתקיימת.

פלט מבחן Goldfeld-Quandt:

Goldfeld-Quandt test

data: FinalModel
GQ = 0.60015, df1 = 61, df2 = 60, p-value = 0.04896
alternative hypothesis: variance changes from segment 1 to 2

דוגמה לשימוש במודל הנבחר

נדגים שימוש במודל לחיזוי תצפיות חדשות שאינן בבסיס הנתונים לקבלת אומד לדירוג החופש הכלכלי לשנת 2020. לטובת הדוגמה נשתמש בנתונים מהמדינות אוסטריה ומקסיקו לשנת 2020.

מדינה	זכויות רכוש	יעילות משפטית	יושרה ממשלתית	נטל המס	פסיקלית בריאות	חופש עיסקי	חופש עבודה	חופש ממוני	חופש מסחר	חופש השקעות	חופש פיננסי	יבשת
אוסטריה	87.3	73.2	84.0	51.3	87.9	73.0	68.3	81.0	86.4	90	70	אירופה
מקסיקו	58.3	34.7	36.7	76.1	87.5	67.0	58.4	70.9	87.6	75	60	מרכז אמריקה

לקבלת אומדן למדד החופש הכלכלי לשנת 2020, נציב את הנתונים במשוואת קו הרגרסיה (הצבה מלאה בנספח 8).

עבור אוסטריה –

$$\begin{aligned}\hat{Y}_{Austria} = & 7.760369 + 0.117892 * 87.3 + 0.129485 * 73.2 - 0.052628 * 84 + 0.119008 \\ & * 51.3 - 10.968077 * 1 + 0.098241 * 87.9 + 0.080912 * 73 + 0.092126 \\ & * 68.3 + 0.045169 * 81 + 0.088162 * 86.4 - 8.308124 * 0 + 1.843556 * 0 \\ & - 5.329688 * 0 + 11.102173 * 0 + 1.767749 * 0 + 1.733506 * 1 + 4.389707 \\ & * 0 + 2.444474 * 0 + 4.834075 * 0 + 6.140582 * 1 + 0.03228 * 84 * 0 \\ & + 0.076302 * 84 * 1 + 0.066511 * 84 * 0 + 0.09913 * 81 * 0 - 0.044716 \\ & * 81 * 0 + 0.068376 * 81 * 0 - 0.175946 * 81 * 0 + 0.105151 * 81 * 1 =\end{aligned}$$

התקבל כי האומדן למדד החופש הכלכלי הוא 73.1576625. מדד החופש הכלכלי לשנת 2020 בפועל הוא 73.2. כלומר השגיאה היא 0.0423375.

עבור מקסיקו –

$$\begin{aligned}\hat{Y}_{Mexico} = & 7.760369 + 0.117892 * 58.3 + 0.129485 * 34.7 - 0.052628 * 36.7 + 0.119008 \\ & * 76.1 - 10.968077 * 0 + 0.098241 * 87.5 + 0.080912 * 67 + 0.092126 \\ & * 58.4 + 0.045169 * 70.9 + 0.088162 * 87.6 - 8.308124 * 0 + 1.843556 * 0 \\ & - 5.329688 * 1 + 11.102173 * 0 + 1.767749 * 0 + 1.733506 * 1 + 4.389707 \\ & * 0 + 2.444474 * 0 + 4.834075 * 1 + 6.140582 * 0 + 0.03228 * 36.7 * 0 \\ & + 0.076302 * 36.7 * 1 + 0.066511 * 36.7 * 0 + 0.09913 * 70.9 * 0 \\ & - 0.044716 * 70.9 * 0 + 0.068376 * 70.9 * 1 - 0.175946 * 70.9 * 0 \\ & + 0.105151 * 70.9 * 0 = 65.46052\end{aligned}$$

התקבל כי האומדן למדד החופש הכלכלי הוא 65.46052. מדד החופש הכלכלי לשנת 2020 בפועל הוא 66.0. כלומר השגיאה היא 0.53948.

בדיקת השערה המבוססת על תוצאות המודל

נבצע בדיקת השערה באמצעות מבחן F על הקשר בין המשתנה המוסבר, מספר בעיטות ממוצע לשער למשחק, למשתנים המסבירים. על ידי מבחן זה נבחן את מובהקות הרגרסיה, כלומר האם קיים קשר ליניארי בין המשתנה התלוי לבין לפחות אחד מהמשתנים המסבירים שהשארנו במודל. השערת האפס הינה שכל מקדמי הרגרסיה שווים לאפס, כלומר שאין קשר בין המשתנה המוסבר לבין אף אחד מהמשתנים המסבירים. ניסוח ההשערות (עבור $\alpha=0.05$):

$$H_0: \beta_i = 0 \text{ (All } \beta \text{ equal to 0)}$$

$$H_1: \text{else (one of the } \beta \text{ or more are not equal to 0)}$$

$$\text{כלל ההחלטה - נדחה את השערת האפס כאשר מתקיים: } F_{st} \geq F_{k,n-k-1}^{1-\alpha}$$

טבלת ANOVA לבדיקת השערה המבוססת על התוצאות:

```
> anova(FinalModel)
Analysis of Variance Table
```

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
x1	1	14615.6	14615.6	5804.9621	< 2.2e-16	***
x2	1	616.2	616.2	244.7212	< 2.2e-16	***
x3	1	2.4	2.4	0.9424	0.333219	
x4	1	613.0	613.0	243.4831	< 2.2e-16	***
x6	1	1566.4	1566.4	622.1219	< 2.2e-16	***
x7	1	198.7	198.7	78.9125	1.865e-15	***
x8	1	205.4	205.4	81.5726	7.761e-16	***
x9	1	336.0	336.0	133.4336	< 2.2e-16	***
x10	1	165.9	165.9	65.8943	1.597e-13	***
factor(dataset\$Continent)	5	293.3	58.7	23.3019	< 2.2e-16	***
factor(x12)	3	365.5	121.8	48.3861	< 2.2e-16	***
factor(x11)	3	230.3	76.8	30.4942	1.890e-15	***
x3:factor(x12)	3	11.1	3.7	1.4704	0.224943	
x9:factor(x14)	5	53.0	10.6	4.2085	0.001313	**
Residuals	150	377.7	2.5			

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

להלן פירוט חישוב ה- F_{st} :

$$F_{st} = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} = \frac{\frac{19272.8}{28}}{\frac{377.7}{179-28-1}} = \frac{688.314286}{2.518} = 273.35754$$

$$F_{cr} = F_{28,150}^{0.95} = 1.551506$$

מכיוון שמתקיים התנאי $F_{st} > F_{cr}$ נקבע כי קיימים מספיק נתונים על מנת לדחות את השערת האפס ברמת מובהקות 5%, ולכן לפחות אחד מהמשתנים המסבירים במודל מובהק בהשפעתו על המשתנה המוסבר.

שיפור המודל

לאחר ביצוע התהליכים הקודמים בבניית המודל הסופי, נבחן האם הוספת משתני טרנספורמציה או החלפת משתנה קיים במשתנה טרנספורמציה כדאית לנו, ותשפר את המודל. עבור כל משתנה מסביר X_i אשר נכנס למודל הסופי בדקנו את שינויו ע"י \sqrt{X} , X^2 . (למעט משתנים קטגוריאליים שטרנספורמציות אלו לא משפיעות עליהם). את הטרנספורמציה בנינו על המודל החדש שקיבלנו ולא על המודל הראשוני, זאת כיוון שמדד פירסון עבור משתנים אלו נמוך בהרבה ביחס למשתנה המוסבר.

הערה: בדקנו גם פונקציית \ln אך אף משתנה לא שופר כתוצאה מטרנספורמציה זו.

X_i	Variable	Pearson
X_1	Property Rights	0.8304095
X_2	Judicial Effectiveness	0.8042531
X_4	Tax Burden	0.13727880
X_6	Fiscal Health	0.56317102
X_7	Business Freedom	0.81252325
X_8	Labor Freedom	0.57963462
X_9	Monetary Freedom	0.64507606
X_{10}	Trade Freedom	0.7530860

עפ"י טבלת המתאם לפי פירסון, בדקנו האם יש משתנים עבורם משתני הטרנספורמציה הגדילו את מדד פירסון.

טרנספורמציות שורש:

ניתן לראות כי המשתנה " Property Rights " וגם המשתנה " Judicial Effectiveness "

שיפרו את מדד פירסון המקורי שלהם כלומר הם מסבירים באופן מובהק יותר את המשתנה המוסבר ולכן לפי טרנספורמציות שורש נשקול להכניס אותם למודל.

טרנספורמציה ריבועית:

	Overall_Score	Property_Rights	Judicial_Effectiveness	Tax_Burden	Fiscal_Health	Business_Freedom	Labor_Freedom	Monetary_Freedom	Trade_Freedom
Overall_Score	1.00000000	0.8471376	0.8118034	-0.03342846	0.53888434	0.77937753	0.5357025	0.50264829	0.6944866
Property_Rights	0.84713757	1.00000000	0.7929784	-0.19554682	0.28410302	0.4109965	0.44935460	0.44935460	0.6650095
Judicial_Effectiveness	0.81180344	0.7929784	1.00000000	-0.19770514	0.29702941	0.69763250	0.4315658	0.36305514	0.5404999
Tax_Burden	-0.03342846	-0.1955468	-0.1977051	1.00000000	-0.07215093	-0.01654575	0.1500015	-0.08595376	-0.0315409
Fiscal_Health	0.53888434	0.2841030	0.2970294	-0.07215093	1.00000000	0.30858878	0.1368622	0.22999559	0.2512915
Business_Freedom	0.77937753	0.7492332	0.6976325	-0.01654575	0.30858878	1.00000000	0.4742259	0.33724170	0.5977973
Labor_Freedom	0.53570250	0.4109965	0.4315658	0.15000150	0.13686216	0.47422588	1.00000000	0.26639814	0.3045293
Monetary_Freedom	0.50264829	0.4493546	0.3630551	-0.08595376	0.22999559	0.33724170	0.2663981	1.00000000	0.2961373
Trade_Freedom	0.69448663	0.6650095	0.5404999	-0.03154090	0.25129147	0.59779728	0.3045293	0.29613729	1.00000000

אף משתנה לא משפר את ערך מדד הפירסון ולכן לא נרצה להוסיף אף משתנה עם טרנספורמציות ריבועיות. ניתן לראות כי משתנה " Property Rights " השתפר יותר מאשר בטרנספורמציות השורש.

גרסיה לאחר:

	Overall_Score	Property_Rights	Judicial_Effectiveness	Tax_Burden	Fiscal_Health	Business_Freedom	Labor_Freedom	Monetary_Freedom	Trade_Freedom
Overall_Score	1.00000000	0.8543113	0.7821076	-0.01537644	0.53299297	0.80183587	0.5224080	0.5885191	0.72984882
Property_Rights	0.85431129	1.00000000	0.8449493	-0.22676450	0.33069061	0.80784229	0.4236770	0.4957405	0.7084998
Judicial_Effectiveness	0.78210758	0.8449493	1.00000000	-0.21290603	0.26492093	0.71955027	0.4367850	0.4127502	0.53686308
Tax_Burden	-0.01537644	-0.2267645	-0.2129060	1.00000000	-0.03185099	-0.05834893	0.1599391	-0.1455859	-0.02391423
Fiscal_Health	0.53299297	0.3306906	0.2649209	-0.03185099	1.00000000	0.28635224	0.1066248	0.2608392	0.30094769
Business_Freedom	0.80183587	0.8078423	0.7195503	-0.05834893	0.28635224	1.00000000	0.4648666	0.4231602	0.63603573
Labor_Freedom	0.52240799	0.4236770	0.4367850	0.15993909	0.10662484	0.46486659	1.00000000	0.2342746	0.33139077
Monetary_Freedom	0.58851913	0.4957405	0.4127502	-0.14558591	0.26083916	0.42316016	0.2342746	1.00000000	0.41880180
Trade_Freedom	0.72984882	0.7084998	0.5368631	-0.02391423	0.30094769	0.63603573	0.3313908	0.4188018	1.00000000

על פי תוצאות של רגרסיה לפנים, לאחר ורגרסיה בצעדים ולפי מדדים AIC , BIC ו- R^2_{adj} , לאחר הכנסת המשתנים המסבירים שעברו טרנספורמציה הגענו למסקנה כי רגרסיה לאחר לפי מדד AIC הוא המודל המתאים (המדד בו בחרנו להתמקד לאורך העבודה). כמו כן מדד R^2_{adj} (0.9777) השתפר מהרצת המודל הראשוני. ניתן לראות כי ערך מדד AIC שופר גם כן (697.9275)

ולכן במידה והמודל החדש שיצא בעת התאמת המודל לפי השיטות הנ"ל יעמוד בכל הנחות הרגרסיה, נבחר להשתמש בו. ניתן לראות סימוכין ב**נספח זה**.

מסקנות והמלצות

בפרויקט זה בחנו כיצד משתנים מסבירים שונים משפיעים על מדד החופש הכלכלי. תחילה שיערנו כי כל המשתנים המסבירים עשויים להשפיע על המודל. ביצענו הליך של בדיקת המודל, התאמתו ושיפורו במספר שלבים. כשלב מקדים, שיערנו מי הם המשתנים החשודים להוצאה מהמודל, והחלטנו על משתנה אחד כזה (x5). הבחירה נעשתה באמצעות בדיקה של רמת המובהקות בין המשתנים המסבירים למשתנה המוסבר, מדד פירסון ובדיקת התאמות בין המסבירים לבין עצמם. לאחר מכן, הוספנו למודל את משתני הדמה והאינטראקציה והתקבל מודל הרגרסיה הראשוני הבא:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 C_5 + \hat{\beta}_6 X_6 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8 + \hat{\beta}_9 X_9 + \hat{\beta}_{10} X_{10} + \hat{\beta}_{11} C_1 + \hat{\beta}_{12} C_2 + \hat{\beta}_{13} C_3 + \hat{\beta}_{14} C_4 + \hat{\beta}_{15} FF_1 + \hat{\beta}_{16} FF_2 + \hat{\beta}_{17} FF_3 + \hat{\beta}_{18} IF_1 + \hat{\beta}_{19} IF_2 + \hat{\beta}_{20} IF_3 + \hat{\beta}_{21} X_{13} + \sum_{i=1}^3 \hat{\beta}_{22+i-1} X_3 * FF_i + \sum_{j=1}^5 \hat{\beta}_{25+j-1} X_9 * C_j$$

בשלב הבא על מנת למצוא את המודל המתאים ביותר ביצענו רגרסיה לפנים, לאחר ובצעדים ובחנו את התוצאות ע"י השוואה למצב הקיים במדדים BIC, AIC ו R_{adj}^2 . השיטה שנבחרה הינה רגרסיה לאחר ע"פ מדד AIC שהניבה את ערך R_{adj}^2 המקסימלי וערך AIC המינימלי. בנוסף, לאחר שביצענו בירור ידני – בדקנו אפשרות של ניפוי משתנה x_3 בהתייעצות עם סגל הקורס, אך החלטנו שלא לנפות משתנה זה. לאחר בחירת משתני המודל, התקבל המודל הבא:

$$\begin{aligned}\hat{Y} = & 7.760369 + 0.117892X_1 + 0.129485X_2 - 0.052628X_3 + 0.119008X_4 - 10.968077C_5 \\ & + 0.098241X_6 + 0.080912X_7 + 0.092126X_8 + 0.045169X_9 + 0.088162X_{10} \\ & - 8.308124C_1 + 1.843556C_2 - 5.329688C_3 + 11.102173C_4 + 1.767749FF_1 \\ & + 1.733506FF_2 + 4.389707FF_3 + 2.444474IF_1 + 4.834075IF_2 \\ & + 6.140582IF_3 + 0.03228X_3 * FF_1 + 0.076302X_3 * FF_2 + 0.066511X_3 * FF_3 \\ & + 0.09913X_9 * C_1 - 0.044716X_9 * C_2 + 0.068376X_9 * C_3 - 0.175946X_9 * C_4 \\ & + 0.105151X_9 * C_5\end{aligned}$$

ניתן לראות כי משתנים רבים אינם נמצאים במודל כיוון שנמצאו כלא תורמים מספיק לערך המוסבר. לאחר מכן, בדקנו האם שלושת הנחות המודל מתקיימות- לינאריות, נורמליות ושוויון שונות ומצאנו שהן אכן מתקיימות. לבסוף, בצענו טרנספורמציה על המשתנים המסבירים ומצאנו כי התקבל מודל חדש ע"פ רגרסיה לאחר לפני מדד AIC אשר ממזער את מדד AIC, ומגדיל את מדד R_{adj}^2 גם. על כן, במידה ומודל זה יעמוד בכלל הנחות הרגרסיה נבחר במודל זה:

$$\begin{aligned}\hat{Y} = & 2.296594 + 0.117645X_1 + 1.656639\sqrt{X_2} + 0.119184X_4 - 10.592899C_5 + 0.097407X_6 \\ & + 0.077048X_7 + 0.092739X_8 + 0.048258X_9 + 0.088957X_{10} - 8.680377C_1 \\ & + 1.419713C_2 - 4.232064C_3 + 10.732144C_4 + 2.065099FF_1 + 1.570213FF_2 \\ & + 3.500643FF_3 + 2.260452IF_1 + 4.686984IF_2 + 5.999718IF_3 - 0.035409X_3 \\ & * FF_0 - 0.01788X_3 * FF_1 + 0.036688X_3 * FF_2 + 0.037795X_3 * FF_3 \\ & + 0.10502X_9 * C_1 - 0.039334X_9 * C_2 + 0.05601X_9 * C_3 - 0.171259X_9 * C_4 \\ & + 0.100301X_9 * C_5\end{aligned}$$

כמסקנה מהפרויקט הבנו שעל מנת לבחור מודל מיטבי המתאר את המשתנה המוסבר בצורה הטובה ביותר, יש לבחון כמה מסבירים בכמה שיטות שונות. עלינו לנסות מסי' מניפולציות על המודל, בכמה דרכים שונות ולנסות ולשפר בכל פעם את מדדי טיב ההתאמה. בסופו של דבר הצלחנו לשפר בצורה מינורית את המדדים שלנו לעומת המודל הראשוני, ונציין כי מדד R_{adj}^2 שלנו גבוה בצורה משמעותית. כלומר, ככל הנראה קיים קשר מאוד חזק בין המשתנים המסבירים שבחרנו לבין המוסבר, בהתאמה למה שחשבנו בתחילת הפרויקט, אך המשתנים בהם חשדנו כי אין ביניהם קשר לבין המוסבר. עובדה מעניינת היא שכל אבעת המשתנים החדשים שהוספנו למודל בחלק זה אכן נמצאים במודל הסופי. כלומר, ההוספה שלהם למודל תרמה מאוד להסברת המוסבר. לסיכום נמליץ לבצע תצפיות נוספות, ולנסות להכניס משתנים מסבירים חדשים למודל על מנת לקבל מדדים טובים יותר על מנת להיות מסוגלים לחזות מספר צפיות ביו טיוב לשיר בצורה מדויקת יותר.

נספחים

נספח 1: פלט R מובהקות- כל משתנה מסביר רציף בנפרד ביחס למוסבר:
זכויות רכוש:

```
Call:
lm(formula = y ~ x1)

Residuals:
    Min       1Q   Median       3Q      Max
-44.496  -3.127   0.270   3.904  13.119

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 35.10344    1.37314   25.56  <2e-16 ***
x1          0.48395    0.02434   19.89  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.289 on 178 degrees of freedom
Multiple R-squared:  0.6896,    Adjusted R-squared:  0.6878
F-statistic: 395.4 on 1 and 178 DF,  p-value: < 2.2e-16
```

יעילות משפטית:

```
Call:
lm(formula = y ~ x2)

Residuals:
    Min       1Q   Median       3Q      Max
-34.392  -3.685  -0.070   4.415  14.397

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 37.76614    1.36858   27.59  <2e-16 ***
x2          0.50514    0.02798   18.05  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.708 on 178 degrees of freedom
Multiple R-squared:  0.6468,    Adjusted R-squared:  0.6448
F-statistic: 326 on 1 and 178 DF,  p-value: < 2.2e-16
```

יושרה ממשלתית:

```
Call:
lm(formula = y ~ x3)

Residuals:
    Min       1Q   Median       3Q      Max
-47.089  -3.512   0.008   4.440  14.155

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 42.29558    1.29024   32.78  <2e-16 ***
x3          0.43824    0.02777   15.78  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.287 on 178 degrees of freedom
Multiple R-squared:  0.5833,    Adjusted R-squared:  0.5809
F-statistic: 249.1 on 1 and 178 DF,  p-value: < 2.2e-16
```

נטל המס:

```
Call:
lm(formula = y ~ x4)

Residuals:
    Min       1Q   Median       3Q      Max
-45.836  -6.948  -0.882   6.887  27.573

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 51.73561    4.95567   10.440  <2e-16 ***
x4          0.11698    0.06327   1.849   0.0661 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.18 on 178 degrees of freedom
Multiple R-squared:  0.01885,    Adjusted R-squared:  0.01333
F-statistic: 3.419 on 1 and 178 DF,  p-value: 0.06611
```

הוצאות ממשלתיות:

```
Call:
lm(formula = y ~ x5)

Residuals:
    Min       1Q   Median       3Q      Max
-53.419  -7.037   0.269   7.136  28.853

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  59.31923    2.53039   23.443  <2e-16 ***
x5           0.02246    0.03699    0.607   0.545
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.28 on 178 degrees of freedom
Multiple R-squared:  0.002067, Adjusted R-squared:  -0.00354
F-statistic: 0.3686 on 1 and 178 DF, p-value: 0.5445
```

בריאות פיסקלית:

```
Call:
lm(formula = y ~ x6)

Residuals:
    Min       1Q   Median       3Q      Max
-41.273  -6.150   0.334   6.170  25.972

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  47.17337    1.64888   28.609  <2e-16 ***
x6           0.20318    0.02235    9.093  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.327 on 178 degrees of freedom
Multiple R-squared:  0.3172, Adjusted R-squared:  0.3133
F-statistic: 82.68 on 1 and 178 DF, p-value: < 2.2e-16
```

חופש עסקי:

```
Call:
lm(formula = y ~ x7)

Residuals:
    Min       1Q   Median       3Q      Max
-20.3073  -4.1367   0.3745   4.1578  17.0192

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  23.28095    2.07463   11.22  <2e-16 ***
x7           0.58528    0.03147   18.60  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.58 on 178 degrees of freedom
Multiple R-squared:  0.6602, Adjusted R-squared:  0.6583
F-statistic: 345.8 on 1 and 178 DF, p-value: < 2.2e-16
```

חופש עבודה:

```
Call:
lm(formula = y ~ x8)

Residuals:
    Min       1Q   Median       3Q      Max
-30.2678  -6.5828  -0.1336   6.0467  21.2987

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  33.91433    2.91154   11.65  <2e-16 ***
x8           0.45070    0.04749    9.49  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.198 on 178 degrees of freedom
Multiple R-squared:  0.336, Adjusted R-squared:  0.3322
F-statistic: 90.06 on 1 and 178 DF, p-value: < 2.2e-16
```

חופש ממוני :

```
Call:
lm(formula = y ~ x9)

Residuals:
    Min       1Q   Median       3Q      Max
-26.4995  -4.2900   0.1807   5.4662  22.1521

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.93925    4.47054   2.447  0.0154 *
x9           0.66098    0.05869  11.263 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.625 on 178 degrees of freedom
Multiple R-squared:  0.4161,    Adjusted R-squared:  0.4128
F-statistic: 126.9 on 1 and 178 DF,  p-value: < 2.2e-16
```

חופש מסחר :

```
Call:
lm(formula = y ~ x10)

Residuals:
    Min       1Q   Median       3Q      Max
-25.6974  -4.1552   0.1947   5.0559  20.6976

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.87500    3.44285   2.578  0.0108 *
x10          0.69723    0.04566  15.271 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.426 on 178 degrees of freedom
Multiple R-squared:  0.5671,    Adjusted R-squared:  0.5647
F-statistic: 233.2 on 1 and 178 DF,  p-value: < 2.2e-16
```

נספח 2- כלל המשתנים המסבירים הרציפים יחדיו ביחס למוסבר :

```
Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 +
  x10)

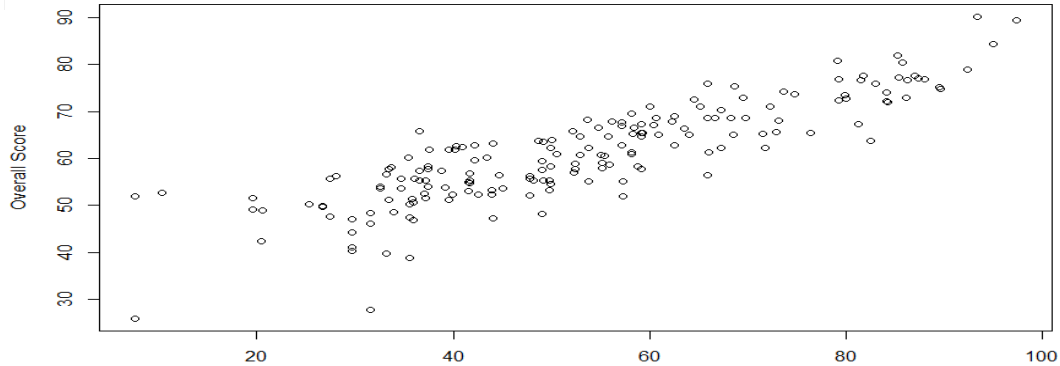
Residuals:
    Min       1Q   Median       3Q      Max
-5.2675  -0.9639   0.2243   1.3896   5.6942

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.453373    1.383061  -3.220  0.001538 **
x1           0.156515    0.018272   8.566  6.44e-15 ***
x2           0.060860    0.019559   3.112  0.002185 **
x3           0.101990    0.019669   5.185  6.10e-07 ***
x4           0.050189    0.014524   3.456  0.000695 ***
x5           0.099993    0.007988  12.518 < 2e-16 ***
x6           0.079082    0.005337  14.818 < 2e-16 ***
x7           0.099452    0.017482   5.689  5.53e-08 ***
x8           0.072392    0.012996   5.570  9.85e-08 ***
x9           0.161565    0.016862   9.582 < 2e-16 ***
x10          0.152766    0.018840   8.109  9.94e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

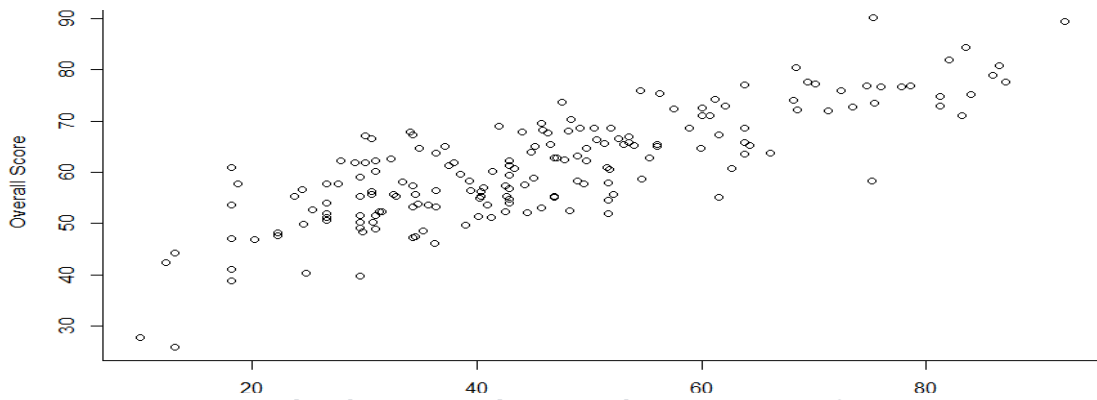
Residual standard error: 2.022 on 169 degrees of freedom
Multiple R-squared:  0.9695,    Adjusted R-squared:  0.9677
F-statistic: 538 on 10 and 169 DF,  p-value: < 2.2e-16
```

נספח 3 – פיזור התצפיות של המשתנים החשודים להסרה :

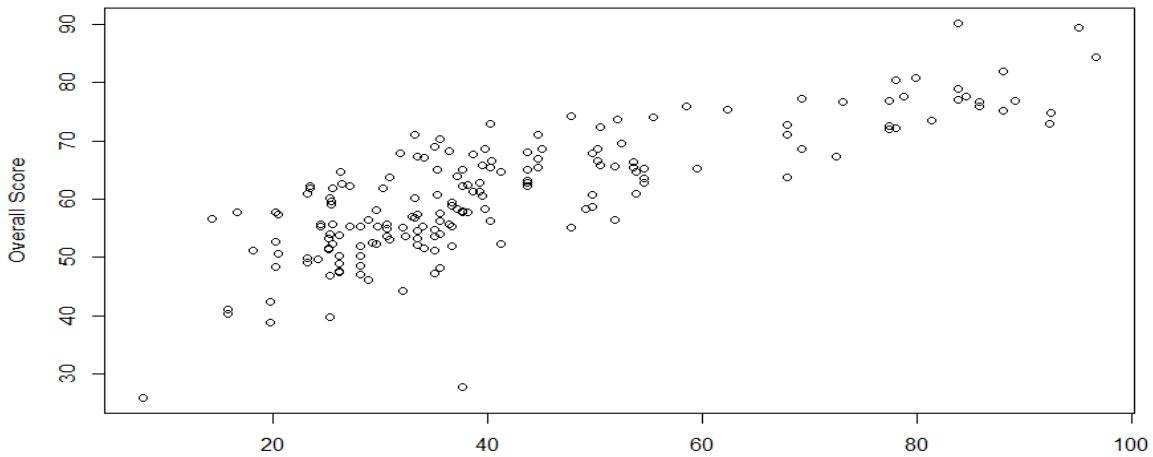
```
plot(y=y, x=x1, ylab='Overall Score', xlab='Property_Rights')  
plot(y=y, x=x2, ylab='Overall Score', xlab='Government_Integrity')  
plot(y=y, x=x3, ylab='Overall Score', xlab='Judicial_Effectiveness')  
plot(y=y, x=x5, ylab='Overall Score', xlab='Government_Spending')
```



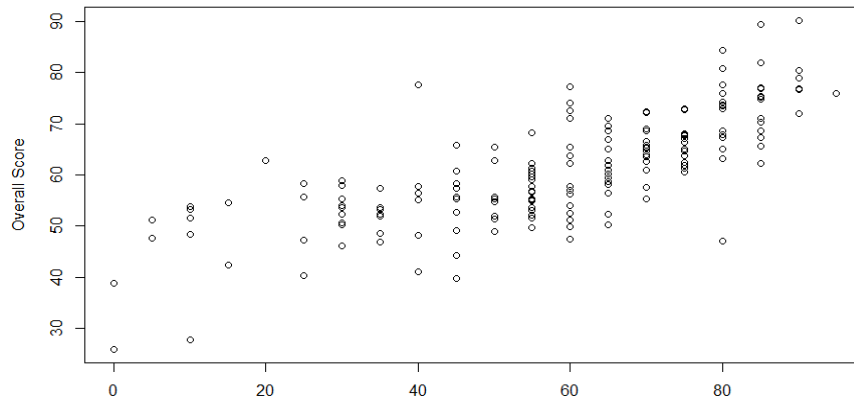
גרף 1: תרשים פיזור של המשתנה 'זכויות רכוש'



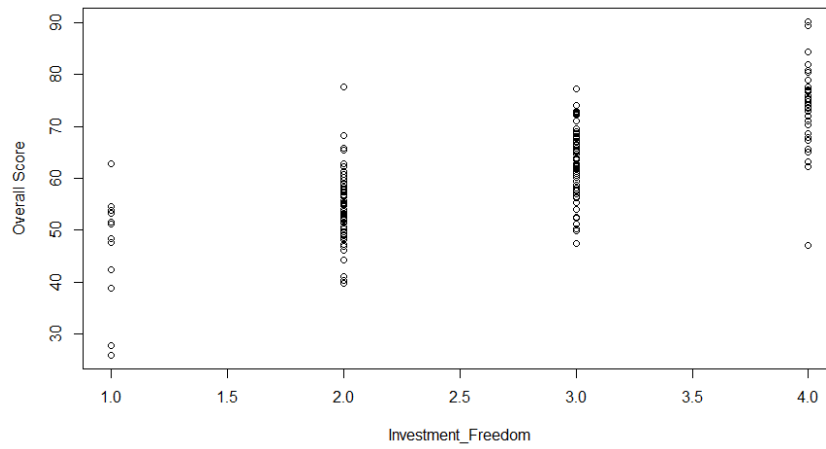
גרף 2: תרשים פיזור של המשתנה 'יחסינות ממשלתית'



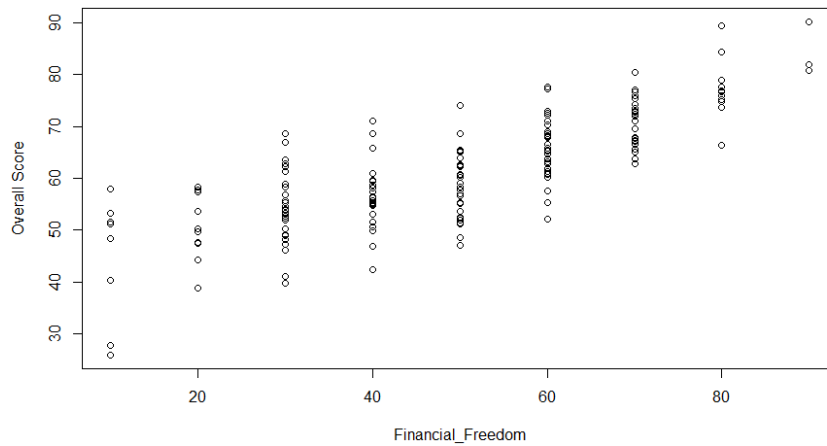
גרף 3: תרשים פיזור של המשתנה 'יעילות משפטית'



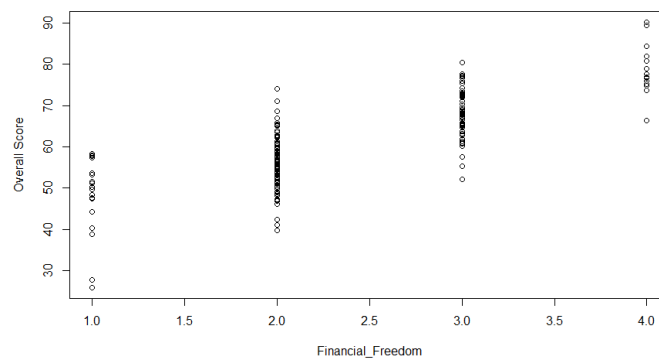
גרף 4: פיזור לפי קטגוריות של המשתנה 'חופש השקעות' לפני איחוד



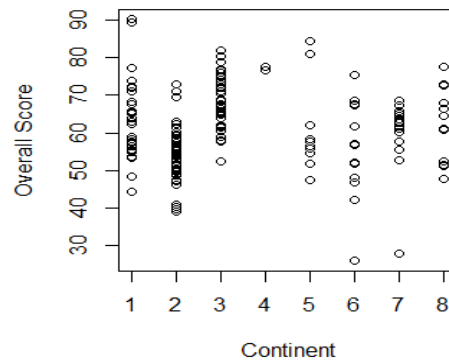
גרף 5: פיזור לפי קטגוריות של המשתנה 'חופש השקעות' לאחר איחוד



גרף 6: פיזור לפי קטגוריות של המשתנה 'חופש פיננסי' לפני איחוד

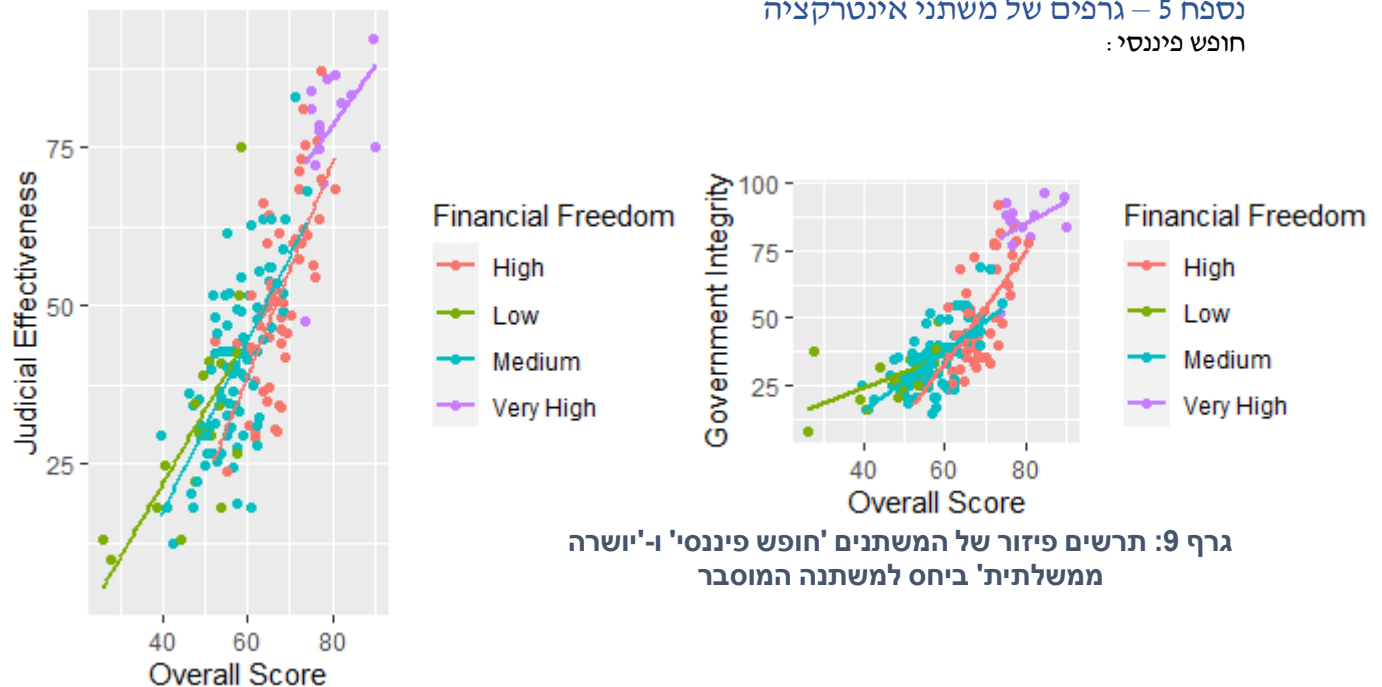


גרף 7: פיזור לפי קטגוריות של המשתנה 'חופש פיננסי' לאחר איחוד



גרף 8: פיזור לפי קטגוריות של המשתנה 'יבשת' לפני איחוד

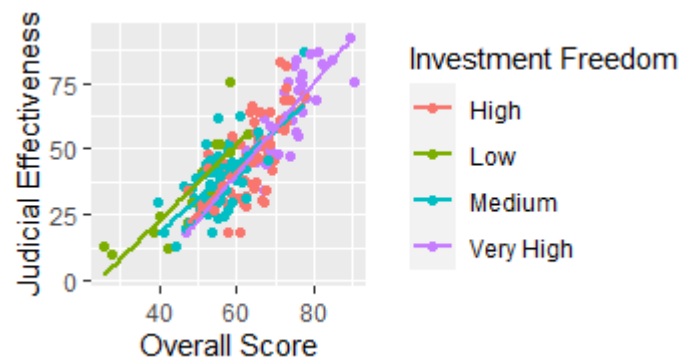
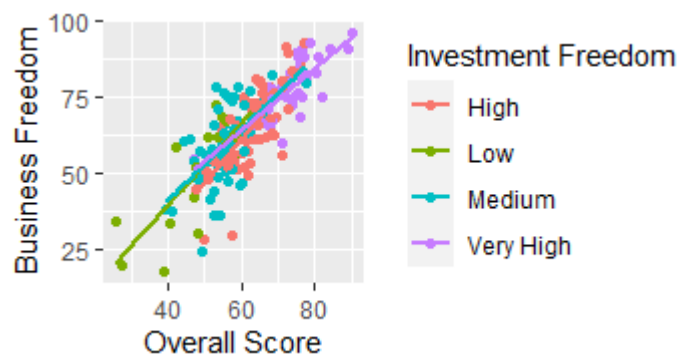
נספח 5 – גרפים של משתני אינטרקציה
חופש פיננסי:



גרף 9: תרשים פיזור של המשתנים 'חופש פיננסי' ו-'יחוס ממשלתי' ביחס למשתנה המוסבר

גרף 10: תרשים פיזור של המשתנים 'חופש פיננסי' ו-'יעילות משפטית' ביחס למשתנה המוסבר

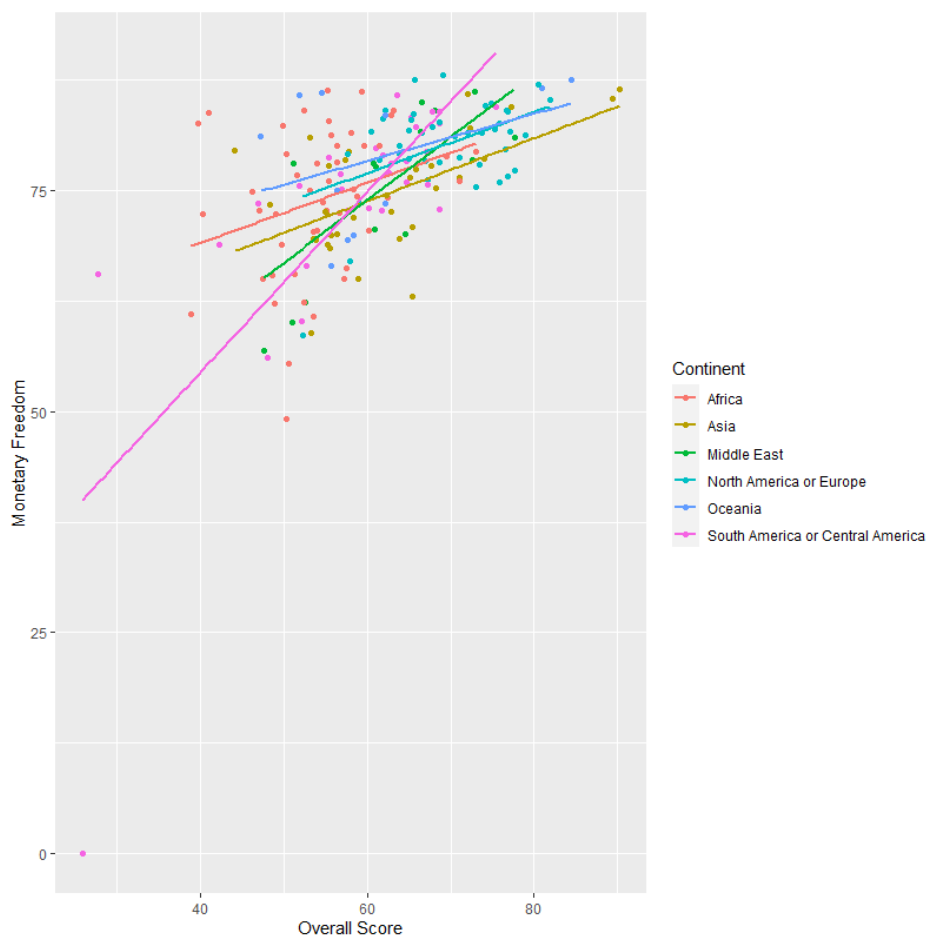
חופש השקעות:



גרף 12: תרשים פיזור של המשתנים 'חופש השקעות' ו- 'חופש עיסקי' ביחס למשתנה המוסבר

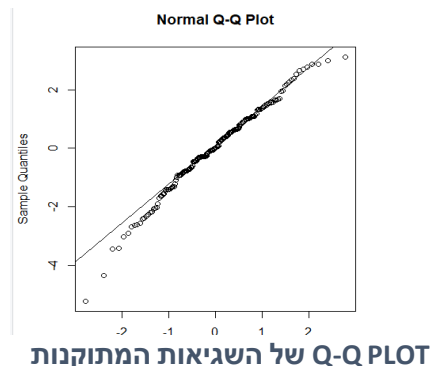
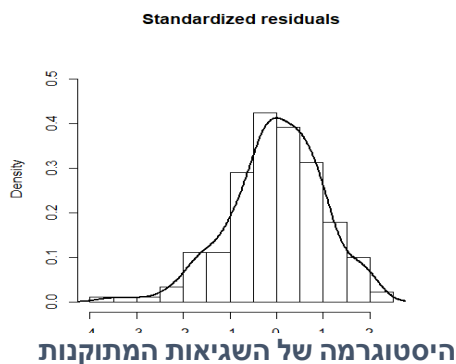
גרף 11: תרשים פיזור של המשתנים 'חופש השקעות' ו- 'יעילות משפטית' ביחס למשתנה המוסבר

יבשת:

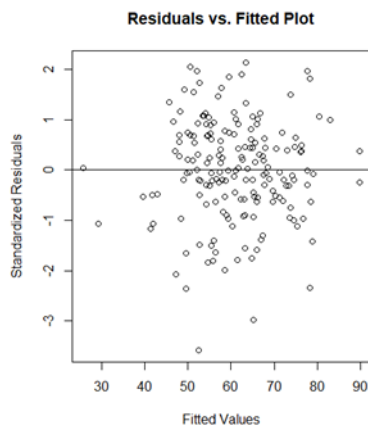


גרף 13: תרשים פיזור של המשתנים 'יבשת' ו-'חופש ממוני' ביחס למשתנה המוסבר

נספח 6 – בדיקת הנחות - הנחת הנורמליות של השגיאות



נספח 7 – הנחת שוויון שוניות והנחת לינאריות



גרף פיזור של השגיאות המתוקנות

$$\begin{aligned}
 \hat{Y}_{Austria} = & -0.547755 + 0.117892 * 87.3 + 0.129485 * 73.2 - 0.052628 * 84 + 0.119008 \\
 & * 51.3 + 0.098241 * 87.9 + 0.080912 * 73 + 0.092126 * 68.3 + 0.144299 \\
 & * 81 + 0.088162 * 86.4 + 8.308124 * 0 + 2.978436 * 0 + 19.410297 * 0 \\
 & + 10.151680 * 0 - 2.659953 * 1 + 1.767749 * 0 + 1.733506 * 1 + 4.389707 \\
 & * 0 + 2.444474 * 0 + 4.834075 * 0 + 6.140582 * 1 + 0.032280 * 84 * 0 \\
 & + 0.076302 * 84 * 1 + 0.066511 * 84 * 0 - 0.099130 * 81 * 0 - 0.030754 \\
 & * 81 * 0 - 0.275076 * 81 * 0 - 0.143846 * 81 * 0 + 0.006022 * 81 * 1 \\
 = & 73.1577
 \end{aligned}$$

טבלאות

טבלה 1 – מובהקות המשתנים המסבירים מול המשתנה המוסבר

X_i	Variable	מובהקות
X_1	<i>Property Rights</i> זכויות רכוש	2.2e-16 ***
X_2	<i>Judicial Effectiveness</i> יעילות משפטית	2.2e-16 ***
X_3	<i>Government Integrity</i> יושרה ממשלתית	2.2e-16 ***
X_4	<i>Tax Burden</i> נטל מס	0.0661
X_5	<i>Government Spending</i> הוצאות ממשלתיות	0.5445
X_6	<i>Fiscal Health</i> בריאות פיסקלית	2.2e-16 ***
X_7	<i>Business Freedom</i> חופש עיסקי	2.2e-16 ***
X_8	<i>Labor Freedom</i> חופש עבודה	2.2e-16 ***
X_9	<i>Monetary Freedom</i> חופש ממוני	2.2e-16 ***
X_{10}	<i>Trade Freedom</i> חופש מסחר	2.2e-16 ***

טבלה 2 – טבלה מסכמת של מדד פירסון (פלט R הוצג בסעיף 2)

X_i	Variable	Pearson
X_1	<i>Property Rights</i> זכויות רכוש	0.8304095
X_2	<i>Judicial Effectiveness</i> יעילות משפטית	0.8042531
X_3	<i>Government Integrity</i> יושרה ממשלתית	0.7637149
X_4	<i>Tax Burden</i> נטל מס	0.13727880

X_5	<i>Government Spending</i> הוצאות ממשלתיות	0.04546061
X_6	<i>Fiscal Health</i> בריאות פיסקלית	0.56317102
X_7	<i>Business Freedom</i> חופש עיסקי	0.81252325
X_8	<i>Labor Freedom</i> חופש עבודה	0.57963462
X_9	<i>Monetary Freedom</i> חופש ממוני	0.64507606
X_{10}	<i>Trade Freedom</i> חופש מסחר	0.7530860

פלטרים

חלק 1 – המודל המקורי

```
> #-----Choosing Variables-----
>
> TestModel<-lm(formula = y~x1+x2+x3+x4+x6+x7+x8+x9+x10+factor(x13)+factor(dataset$Continent)+factor(x12)
+factor(x11)+factor(x12):x3+factor(dataset$Continent):x9, data = dataset)
>
> anova(TestModel)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq  F value    Pr(>F)
x1          1 14615.6 14615.6 5766.2926 < 2.2e-16 ***
x2          1  616.2   616.2  243.0910 < 2.2e-16 ***
x3          1    2.4    2.4   0.9361 0.334840
x4          1  613.0   613.0  241.8611 < 2.2e-16 ***
x6          1 1566.4  1566.4  617.9777 < 2.2e-16 ***
x7          1  198.7   198.7   78.3868 2.312e-15 ***
x8          1  205.4   205.4   81.0292 9.674e-16 ***
x9          1  336.0   336.0  132.5448 < 2.2e-16 ***
x10         1  165.9   165.9   65.4554 1.922e-13 ***
factor(x13)  1    0.0    0.0   0.0082 0.927970
factor(dataset$Continent) 5   293.4   58.7  23.1543 < 2.2e-16 ***
factor(x12)  3   365.5  121.8  48.0665 < 2.2e-16 ***
factor(x11)  3   230.6   76.9  30.3247 2.317e-15 ***
x3: factor(x12)  3    10.7    3.6   1.4106 0.241995
x9: factor(dataset$Continent) 5   53.0   10.6   4.1795 0.001392 **
Residuals    149   377.7    2.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(TestModel)
```

Call:

```
lm(formula = y ~ x1 + x2 + x3 + x4 + x6 + x7 + x8 + x9 + x10 +
  factor(x13) + factor(dataset$Continent) + factor(x12) + factor(x11) +
  factor(x12):x3 + factor(dataset$Continent):x9, data = dataset)
```

Residuals:

```
Min    1Q  Median    3Q   Max
```

-5.2221 -0.7926 0.0211 1.0042 3.1173

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept)      7.745764  3.001257  2.581 0.010821 *
x1                0.117798  0.016916  6.963 9.89e-11 ***
x2                0.129465  0.016134  8.024 2.80e-13 ***
x3               -0.052766  0.050099 -1.053 0.293940
x4                0.119031  0.013052  9.120 4.82e-16 ***
x6                0.098259  0.004716 20.837 < 2e-16 ***
x7                0.080944  0.014827  5.459 1.95e-07 ***
x8                0.092114  0.011212  8.216 9.38e-14 ***
x9                0.045353  0.031290  1.449 0.149317
x10               0.088168  0.017959  4.909 2.37e-06 ***
factor(x13)1      0.010427  0.373030  0.028 0.977738
factor(dataset$Continent)1 -8.290561  4.211403 -1.969 0.050855 .
factor(dataset$Continent)2  1.877941  4.701289  0.399 0.690132
factor(dataset$Continent)3 -5.315717  2.728072 -1.949 0.053231 .
factor(dataset$Continent)4 11.122223  5.807215  1.915 0.057377 .
factor(dataset$Continent)5 -10.964239  4.598899 -2.384 0.018380 *
factor(x12)2      1.761836  1.652794  1.066 0.288159
factor(x12)3      1.723754  1.746063  0.987 0.325134
factor(x12)4      4.381066  3.291980  1.331 0.185278
factor(x11)2      2.444980  0.653084  3.744 0.000258 ***
factor(x11)3      4.834399  0.698916  6.917 1.27e-10 ***
factor(x11)4      6.140822  0.821835  7.472 6.20e-12 ***
x3: factor(x12)2   0.032443  0.050938  0.637 0.525161
x3: factor(x12)3   0.076560  0.050495  1.516 0.131592
x3: factor(x12)4   0.066727  0.059762  1.117 0.265985
x9: factor(dataset$Continent)1 0.098858  0.056801  1.740 0.083847 .
x9: factor(dataset$Continent)2 -0.045168  0.063016 -0.717 0.474632
x9: factor(dataset$Continent)3  0.068138  0.036912  1.846 0.066880 .
x9: factor(dataset$Continent)4 -0.176243  0.074571 -2.363 0.019398 *
x9: factor(dataset$Continent)5  0.105044  0.058469  1.797 0.074429 .
```

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

Residual standard error: 1.592 on 149 degrees of freedom

Multiple R-squared: 0.9808, Adjusted R-squared: 0.977

F-statistic: 262.2 on 29 and 149 DF, p-value: < 2.2e-16

```
> AIC<-AIC(TestModel)
> BIC<-AIC(TestModel, k=log(179))
> print(AIC)
[1] 703.6254
> print(BIC)
[1] 802.4344
```

חלק 2 – רגרסיה לאחר

```
> #-----Back-----
>
> #backward regression- start with full data
```

AIC

```
> #AIC
> BackAIC<-step(TestModel,direction = "backward",test="F")
Start: AIC=193.65
y ~ x1 + x2 + x3 + x4 + x6 + x7 + x8 + x9 + x10 + factor(x13) +
  factor(dataset$Continent) + factor(x12) + factor(x11) + factor(x12): x3 +
  factor(dataset$Continent): x9
```

Df Sum of Sq RSS AIC F value Pr(>F)

```

- factor(x13)          1   0.00 377.67 191.65 0.0008 0.977738
<none>                 377.67 193.65
- x3: factor(x12)       3   13.23 390.90 193.81 1.7399 0.161309
- x9: factor(dataset$Continent) 5   52.97 430.63 207.14 4.1795 0.001392
- x10                   1   61.09 438.76 218.48 24.1031 2.372e-06
- x7                    1   75.55 453.21 224.29 29.8051 1.952e-07
- x1                    1  122.91 500.57 242.08 48.4901 9.894e-11
- x2                    1  163.20 540.87 255.94 64.3885 2.802e-13
- x8                    1  171.10 548.76 258.53 67.5021 9.376e-14
- factor(x11)           3   205.63 583.29 265.45 27.0418 5.076e-14
- x4                    1  210.80 588.47 271.03 83.1678 4.816e-16
- x6                    1 1100.52 1478.18 435.90 434.1856 < 2.2e-16

```

```

- factor(x13)
<none>
- x3: factor(x12)
- x9: factor(dataset$Continent) **
- x10          ***
- x7           ***
- x1           ***
- x2           ***
- x8           ***
- factor(x11)  ***
- x4           ***
- x6           ***
---

```

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

Step: AIC=191.65

$y \sim x1 + x2 + x3 + x4 + x6 + x7 + x8 + x9 + x10 + \text{factor}(\text{dataset}\$Continent) + \text{factor}(x12) + \text{factor}(x11) + x3: \text{factor}(x12) + x9: \text{factor}(\text{dataset}\$Continent)$

```

              Df Sum of Sq  RSS   AIC F value    Pr(>F)
<none>                 377.67 191.65
- x3: factor(x12)       3   13.76 391.42 192.05  1.8214 0.145695
- x9: factor(dataset$Continent) 5   52.98 430.65 205.14  4.2085 0.001313
- x10                   1   61.09 438.76 216.49 24.2646 2.194e-06
- x7                    1   75.95 453.62 222.45 30.1670 1.657e-07
- x1                    1  128.22 505.89 241.97 50.9274 3.820e-11
- x2                    1  163.58 541.25 254.06 64.9706 2.213e-13
- x8                    1  171.40 549.07 256.63 68.0773 7.429e-14
- factor(x11)           3   205.64 583.30 263.46 27.2244 4.093e-14
- x4                    1  211.60 589.27 269.28 84.0437 3.469e-16
- x6                    1 1121.26 1498.93 436.40 445.3365 < 2.2e-16

```

```

<none>
- x3: factor(x12)
- x9: factor(dataset$Continent) **
- x10          ***
- x7           ***
- x1           ***
- x2           ***
- x8           ***
- factor(x11)  ***
- x4           ***
- x6           ***
---

```

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

> AIC<-AIC(BackAIC)

> anova(BackAIC)

Analysis of Variance Table

Response: y

```

              Df Sum Sq Mean Sq  F value    Pr(>F)
x1              1 14615.6 14615.6 5804.9621 < 2.2e-16 ***

```



```

x2      1  616.2  616.2 244.7212 < 2.2e-16 ***
x3      1   2.4   2.4  0.9424 0.333219
x4      1  613.0  613.0 243.4831 < 2.2e-16 ***
x6      1 1566.4 1566.4 622.1219 < 2.2e-16 ***
x7      1  198.7  198.7 78.9125 1.865e-15 ***
x8      1  205.4  205.4 81.5726 7.761e-16 ***
x9      1  336.0  336.0 133.4336 < 2.2e-16 ***
x10     1  165.9  165.9 65.8943 1.597e-13 ***
factor(dataset$Continent) 5 293.3 58.7 23.3019 < 2.2e-16 ***
factor(x12) 3 365.5 121.8 48.3861 < 2.2e-16 ***
factor(x11) 3 230.3 76.8 30.4942 1.890e-15 ***
x3: factor(x12) 3 11.1 3.7 1.4704 0.224943
x9: factor(dataset$Continent) 5 53.0 10.6 4.2085 0.001313 **
Residuals      150 377.7 2.5

```

```

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

```

```
> summary(BackAIC)
```

Call:

```
lm(formula = y ~ x1 + x2 + x3 + x4 + x6 + x7 + x8 + x9 + x10 +
  factor(dataset$Continent) + factor(x12) + factor(x11) + x3: factor(x12) +
  x9: factor(dataset$Continent), data = dataset)
```

Residuals:

```

Min      1Q  Median      3Q      Max
-5.2203 -0.7927  0.0209  1.0027  3.1184

```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    7.760369  2.945561  2.635 0.009308 **
x1              0.117892  0.016520  7.136 3.82e-11 ***
x2              0.129485  0.016064  8.060 2.21e-13 ***
x3             -0.052628  0.049690 -1.059 0.291240
x4              0.119008  0.012981  9.168 3.47e-16 ***
x6              0.098241  0.004655 21.103 < 2e-16 ***
x7              0.080912  0.014731  5.492 1.66e-07 ***
x8              0.092126  0.011166  8.251 7.43e-14 ***
x9              0.045169  0.030494  1.481 0.140634
x10             0.088162  0.017898  4.926 2.19e-06 ***
factor(dataset$Continent)1 -8.308124  4.150375 -2.002 0.047112 *
factor(dataset$Continent)2  1.843556  4.522361  0.408 0.684109
factor(dataset$Continent)3 -5.329688  2.672951 -1.994 0.047972 *
factor(dataset$Continent)4 11.102173  5.743520  1.933 0.055121 .
factor(dataset$Continent)5 -10.968077  4.581512 -2.394 0.017902 *
factor(x12)2      1.767749  1.633732  1.082 0.280975
factor(x12)3      1.733506  1.705142  1.017 0.310964
factor(x12)4      4.389707  3.266500  1.344 0.181024
factor(x11)2      2.444474  0.650655  3.757 0.000246 ***
factor(x11)3      4.834075  0.696488  6.941 1.10e-10 ***
factor(x11)4      6.140582  0.819048  7.497 5.27e-12 ***
x3: factor(x12)2    0.032280  0.050432  0.640 0.523114
x3: factor(x12)3    0.076302  0.049478  1.542 0.125148
x3: factor(x12)4    0.066511  0.059060  1.126 0.261898
x9: factor(dataset$Continent)1  0.099130  0.055775  1.777 0.077542 .
x9: factor(dataset$Continent)2 -0.044716  0.060702 -0.737 0.462483
x9: factor(dataset$Continent)3  0.068376  0.035802  1.910 0.058063 .
x9: factor(dataset$Continent)4 -0.175946  0.073564 -2.392 0.018007 *
x9: factor(dataset$Continent)5  0.105151  0.058149  1.808 0.072561 .

```

```

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

```

Residual standard error: 1.587 on 150 degrees of freedom

Multiple R-squared: 0.9808, Adjusted R-squared: 0.9772

F-statistic: 273.4 on 28 and 150 DF, p-value: < 2.2e-16

```
> BIC<-AIC(BackAIC, k=log(179))
> print(AIC)
[1] 701.6263
> print(BIC)
[1] 797.2479
```

BIC

#BIC

```
> BackBIC<-step(TestModel, direction = "backward",k=log(179))
```

Start: AIC=289.27

```
y ~ x1 + x2 + x3 + x4 + x6 + x7 + x8 + x9 + x10 + factor(x13) +
  factor(dataset$Continent) + factor(x12) + factor(x11) + factor(x12):x3 +
  factor(dataset$Continent):x9
```

	Df	Sum of Sq	RSS	AIC
- x3: factor(x12)	3	13.23	390.90	279.87
- factor(x13)	1	0.00	377.67	284.08
- x9: factor(dataset\$Continent)	5	52.97	430.63	286.82
<none>			377.67	289.27
- x10	1	61.09	438.76	310.92
- x7	1	75.55	453.21	316.72
- x1	1	122.91	500.57	334.51
- x2	1	163.20	540.87	348.37
- x8	1	171.10	548.76	350.96
- factor(x11)	3	205.63	583.29	351.51
- x4	1	210.80	588.47	363.47
- x6	1	1100.52	1478.18	528.34

Step: AIC=279.87

```
y ~ x1 + x2 + x4 + x6 + x7 + x8 + x9 + x10 + factor(x13) +
  factor(dataset$Continent) + factor(x12) + factor(x11) + x9: factor(dataset$Continent)
```

	Df	Sum of Sq	RSS	AIC
- factor(x13)	1	0.53	391.42	274.92
- x3	1	0.67	391.56	274.99
- x9: factor(dataset\$Continent)	5	50.46	441.36	275.67
<none>			390.90	279.87
- x10	1	56.93	447.83	299.02
- x7	1	72.86	463.76	305.28
- x1	1	124.45	515.34	324.16
- factor(x12)	3	187.92	578.82	334.57
- x2	1	157.01	547.90	335.12
- factor(x11)	3	198.14	589.04	337.70
- x8	1	166.95	557.85	338.34
- x4	1	208.42	599.32	351.18
- x6	1	1125.57	1516.46	517.35

Step: AIC=274.92

```
y ~ x1 + x2 + x3 + x4 + x6 + x7 + x8 + x9 + x10 + factor(dataset$Continent) +
  factor(x12) + factor(x11) + x9: factor(dataset$Continent)
```

	Df	Sum of Sq	RSS	AIC
- x3	1	0.92	392.34	270.16
- x9: factor(dataset\$Continent)	5	50.33	441.75	270.64
<none>			391.42	274.92
- x10	1	56.87	448.30	294.02
- x7	1	74.14	465.57	300.79
- x1	1	126.22	517.64	319.76
- factor(x12)	3	187.41	578.83	329.39
- x2	1	156.49	547.91	329.94
- factor(x11)	3	197.79	589.21	332.57
- x8	1	166.43	557.85	333.16
- x4	1	210.35	601.78	346.72
- x6	1	1156.67	1548.10	515.86

Step: AIC=270.16

```
y ~ x1 + x2 + x4 + x6 + x7 + x8 + x9 + x10 + factor(dataset$Continent) +
  factor(x12) + factor(x11) + x9: factor(dataset$Continent)
```

	Df	Sum of Sq	RSS	AIC
- x9: factor(dataset\$Continent)	5	50.01	442.35	265.69
<none>		392.34	270.16	
- x10	1	56.07	448.41	288.88
- x7	1	78.90	471.24	297.77
- x1	1	142.01	534.35	320.26
- factor(x12)	3	190.10	582.44	325.31
- x8	1	166.68	559.02	328.34
- factor(x11)	3	202.54	594.88	329.10
- x4	1	211.64	603.99	342.19
- x2	1	246.17	638.51	352.14
- x6	1	1155.75	1548.10	510.67

Step: AIC=265.69

```
y ~ x1 + x2 + x4 + x6 + x7 + x8 + x9 + x10 + factor(dataset$Continent) +
  factor(x12) + factor(x11)
```

	Df	Sum of Sq	RSS	AIC
<none>		442.35	265.69	
- x10	1	53.66	496.01	281.00
- x9	1	57.83	500.18	282.50
- factor(dataset\$Continent)	5	132.06	574.41	286.52
- x7	1	79.58	521.93	290.12
- factor(x12)	3	172.15	614.50	308.97
- x1	1	168.63	610.98	318.31
- x8	1	186.42	628.77	323.45
- factor(x11)	3	234.26	676.61	326.20
- x4	1	216.77	659.12	331.89
- x2	1	234.69	677.04	336.69
- x6	1	1191.91	1634.26	494.43

```
> AIC<-AIC(BackBIC)
```

```
> BIC<-AIC(BackBIC, k=log(179))
```

```
> summary(BackBIC)
```

Call:

```
lm(formula = y ~ x1 + x2 + x4 + x6 + x7 + x8 + x9 + x10 + factor(dataset$Continent) +
  factor(x12) + factor(x11), data = dataset)
```

Residuals:

```
Min      1Q  Median      3Q      Max
-5.8992 -0.9168  0.1411  1.0934  4.4043
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.441452	1.713048	2.593	0.0104 *
x1	0.126948	0.016306	7.785	8.37e-13 ***
x2	0.125875	0.013705	9.185	< 2e-16 ***
x4	0.115682	0.013105	8.827	1.85e-15 ***
x6	0.098250	0.004747	20.698	< 2e-16 ***
x7	0.079284	0.014824	5.348	3.04e-07 ***
x8	0.093167	0.011381	8.186	8.25e-14 ***
x9	0.078110	0.017132	4.559	1.02e-05 ***
x10	0.079033	0.017996	4.392	2.05e-05 ***
factor(dataset\$Continent)1	-0.973881	0.468368	-2.079	0.0392 *
factor(dataset\$Continent)2	-1.367764	0.599009	-2.283	0.0237 *
factor(dataset\$Continent)3	-0.422594	0.420842	-1.004	0.3168
factor(dataset\$Continent)4	-2.978579	0.621956	-4.789	3.81e-06 ***
factor(dataset\$Continent)5	-2.745970	0.492190	-5.579	1.02e-07 ***
factor(x12)2	2.358283	0.516314	4.568	9.84e-06 ***
factor(x12)3	4.105031	0.615637	6.668	4.06e-10 ***
factor(x12)4	6.441983	0.851448	7.566	2.92e-12 ***
factor(x11)2	2.745054	0.625313	4.390	2.06e-05 ***

```

factor(x11)3      5.119219  0.693433  7.382 8.19e-12 ***
factor(x11)4      6.647644  0.839982  7.914 4.00e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.668 on 159 degrees of freedom
Multiple R-squared:  0.9775    Adjusted R-squared:  0.9748 
F-statistic: 363.4 on 19 and 159 DF, p-value: < 2.2e-16

```

```

> print(AIC)
[1] 711.9244
> print(BIC)
[1] 778.8595

```

חלק 3 – רגרסיה לפני

```

> #-----Forward-----
> #forward regression- start with empty data
> forwardRegression<-lm(y~1, data=dataset)

```

AIC

```

> #-----Forward-----
>
> #forward regression- start with empty data
> forwardRegression<-lm(y~1, data=dataset)
> #AIC
> forwardAIC<-step(forwardRegression, direction = "forward",test="F",scope = formula(TestModel))
Start: AIC=843.03
y ~ 1

```

```

      Df Sum of Sq  RSS   AIC F value    Pr(>F)
+ x1      1 14615.6 5034.8 601.28 513.8199 < 2.2e-16 ***
+ x2      1 12865.7  6784.7 654.67 335.6454 < 2.2e-16 ***
+ x3      1 12439.7  7210.7 665.57 305.3566 < 2.2e-16 ***
+ x7      1 12395.2  7255.2 666.67 302.3988 < 2.2e-16 ***
+ factor(x12)      3 11862.1  7788.3 683.37  88.8458 < 2.2e-16 ***
+ factor(x11)      3 10818.2  8832.2 705.88  71.4502 < 2.2e-16 ***
+ x10      1  9845.3  9805.0 720.58 177.7275 < 2.2e-16 ***
+ x9       1  6444.1 13206.3 773.89  86.3684 < 2.2e-16 ***
+ x6       1  5923.6 13726.8 780.81  76.3819 1.753e-15 ***
+ x8       1  5593.1 14057.3 785.07  70.4239 1.480e-14 ***
+ factor(dataset$Continent) 5  5374.1 14276.3 795.84 13.0247 9.201e-11 ***
+ factor(x13)      1 1701.5 17948.9 828.81 16.7795 6.381e-05 ***
<none>                19650.4 843.03
+ x4       1  14.6 19635.8 844.89  0.1316  0.7172
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Step: AIC=601.28
y ~ x1

```

```

      Df Sum of Sq  RSS   AIC F value    Pr(>F)
+ x6      1 1700.82 3334.0 529.49 89.7863 < 2.2e-16 ***
+ factor(x12)      3 1055.62 3979.2 565.16 15.3867 6.359e-09 ***
+ factor(x11)      3 1015.20 4019.6 566.97 14.6488 1.504e-08 ***
+ x7       1  728.51 4306.3 575.30 29.7746 1.628e-07 ***
+ x8       1  690.66 4344.1 576.87 27.9817 3.604e-07 ***
+ x9       1  622.30 4412.5 579.66 24.8214 1.495e-06 ***
+ x2       1  616.15 4418.6 579.91 24.5423 1.697e-06 ***
+ x4       1  496.47 4538.3 584.69 19.2536 1.968e-05 ***
+ x10      1  454.39 4580.4 586.35 17.4599 4.614e-05 ***

```

```

+ x3          1  260.19 4774.6 593.78 9.5909 0.002276 **
<none>          5034.8 601.28
+ factor(x13)   1   8.50 5026.3 602.97 0.2978 0.585953
+ factor(dataset$Continent) 5  42.94 4991.8 609.74 0.2959 0.914695
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Step: AIC=529.49

y ~ x1 + x6

```

              Df Sum of Sq  RSS   AIC F value    Pr(>F)
+ factor(x11)    3  910.21 2423.7 478.42 21.6561 5.881e-12 ***
+ factor(x12)    3  863.40 2470.6 481.84 20.1531 3.000e-11 ***
+ x8             1  693.43 2640.5 489.75 45.9566 1.782e-10 ***
+ x7             1  557.33 2776.6 498.75 35.1261 1.605e-08 ***
+ x2             1  503.69 2830.3 502.17 31.1440 8.980e-08 ***
+ x4             1  486.35 2847.6 503.27 29.8884 1.558e-07 ***
+ x9             1  375.45 2958.5 510.10 22.2086 4.969e-06 ***
+ x10            1  321.72 3012.2 513.33 18.6906 2.575e-05 ***
+ x3             1  210.01 3123.9 519.84 11.7646 0.000753 ***
+ factor(dataset$Continent) 5  229.18 3104.8 526.74 2.5245 0.031096 *
<none>          3334.0 529.49
+ factor(x13)    1   0.02 3333.9 531.49 0.0009 0.976619
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Step: AIC=478.42

y ~ x1 + x6 + factor(x11)

```

              Df Sum of Sq  RSS   AIC F value    Pr(>F)
+ x8             1  706.58 1717.2 418.73 70.7749 1.490e-14 ***
+ x4             1  605.28 1818.5 428.99 57.2501 2.220e-12 ***
+ x7             1  488.78 1935.0 440.10 43.4477 5.128e-10 ***
+ x2             1  483.34 1940.4 440.60 42.8436 6.563e-10 ***
+ factor(dataset$Continent) 5  407.66 2016.1 455.45 6.7941 8.516e-06 ***
+ factor(x12)    3  335.13 2088.6 457.78 9.0924 1.291e-05 ***
+ x3             1  173.91 2249.8 467.09 13.2955 0.0003523 ***
+ x10            1  147.76 2276.0 469.16 11.1663 0.0010219 **
+ x9             1  100.35 2323.4 472.85 7.4289 0.0070825 **
<none>          2423.7 478.42
+ factor(x13)    1   5.55 2418.2 480.01 0.3948 0.5305966
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Step: AIC=418.73

y ~ x1 + x6 + factor(x11) + x8

```

              Df Sum of Sq  RSS   AIC F value    Pr(>F)
+ x2             1  330.48 1386.7 382.46 40.7541 1.568e-09 ***
+ x4             1  306.18 1411.0 385.57 37.1072 7.177e-09 ***
+ factor(x12)    3  317.99 1399.2 388.07 12.8030 1.402e-07 ***
+ x7             1  259.09 1458.1 391.45 30.3850 1.286e-07 ***
+ factor(dataset$Continent) 5  305.10 1412.1 393.71 7.2167 3.839e-06 ***
+ x10            1  123.03 1594.1 407.42 13.1968 0.0003705 ***
+ x3             1  116.84 1600.3 408.11 12.4844 0.0005278 ***
+ x9             1   63.35 1653.8 414.00 6.5498 0.0113556 *
<none>          1717.2 418.73
+ factor(x13)    1  17.22 1699.9 418.92 1.7324 0.1898627
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Step: AIC=382.46

y ~ x1 + x6 + factor(x11) + x8 + x2

```

              Df Sum of Sq  RSS   AIC F value    Pr(>F)

```

```

+ x4          1  382.29 1004.4 326.73 64.7057 1.423e-13 ***
+ factor(x12)  3  285.31 1101.4 347.23 14.5065 1.902e-08 ***
+ x7          1  180.17 1206.5 359.55 25.3861 1.191e-06 ***
+ factor(dataset$Continent) 5  220.61 1166.1 361.45 6.2812 2.299e-05 ***
+ x10         1  146.09 1240.6 364.54 20.0186 1.399e-05 ***
+ x9          1  64.85 1321.8 375.89 8.3404 0.004382 **
<none>              1386.7 382.46
+ factor(x13)    1   3.80 1382.9 383.97 0.4667 0.495443
+ x3            1   0.09 1386.6 384.45 0.0109 0.916888
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Step: AIC=326.73
 $y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4$

```

              Df Sum of Sq  RSS   AIC F value  Pr(>F)
+ factor(x12)    3  262.797 741.59 278.43 19.7266 5.384e-11 ***
+ x7            1  103.439 900.95 309.28 19.4030 1.873e-05 ***
+ x9            1   83.956 920.43 313.10 15.4151 0.0001255 ***
+ x10           1   75.652 928.73 314.71 13.7663 0.0002806 ***
+ factor(dataset$Continent) 5  101.659 902.73 317.63 3.7162 0.0032552 **
<none>              1004.39 326.73
+ x3            1   7.268 997.12 327.43 1.2319 0.2686185
+ factor(x13)    1   0.918 1003.47 328.57 0.1545 0.6947418
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Step: AIC=278.43
 $y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12)$

```

              Df Sum of Sq  RSS   AIC F value  Pr(>F)
+ x7            1   76.454 665.14 260.96 19.0808 2.2e-05 ***
+ factor(dataset$Continent) 5   94.763 646.83 263.96 4.7468 0.0004491 ***
+ x9            1   50.871 690.72 267.71 12.2258 0.0006047 ***
+ x10           1   50.545 691.04 267.80 12.1418 0.0006307 ***
<none>              741.59 278.43
+ x3            1   2.274 739.32 279.88 0.5106 0.4758727
+ factor(x13)    1   1.792 739.80 280.00 0.4022 0.5268520
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Step: AIC=260.96
 $y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7$

```

              Df Sum of Sq  RSS   AIC F value  Pr(>F)
+ factor(dataset$Continent) 5  107.899 557.24 239.27 6.2349 2.587e-05 ***
+ x9            1   55.600 609.54 247.33 15.0509 0.000151 ***
+ x10           1   36.300 628.84 252.91 9.5247 0.002378 **
<none>              665.14 260.96
+ factor(x13)    1   1.319 663.82 262.60 0.3278 0.567723
+ x3            1   0.129 665.01 262.92 0.0319 0.858390
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Step: AIC=239.27
 $y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7 + \text{factor}(\text{dataset}\$Continent)$

```

              Df Sum of Sq  RSS   AIC F value  Pr(>F)
+ x9            1  61.230 496.01 220.44 19.7515 1.641e-05 ***
+ x10           1  57.055 500.18 221.94 18.2511 3.312e-05 ***
<none>              557.24 239.27
+ x3            1   0.885 556.35 240.99 0.2546 0.6146
+ factor(x13)  1   0.000 557.24 241.27 0.0000 0.9976
---

```

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

Step: AIC=220.44

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7 + \text{factor}(\text{dataset}\$Continent) + x9$

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
+ x10	1	53.656	442.35	201.94	19.2863	2.045e-05 ***
+ x9: factor(dataset\$Continent)	5	47.594	448.41	212.38	3.2903	0.007485 **
<none>		496.01	220.44			
+ x3	1	0.141	495.87	222.39	0.0453	0.831715
+ factor(x13)	1	0.128	495.88	222.39	0.0411	0.839595

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

Step: AIC=201.94

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7 + \text{factor}(\text{dataset}\$Continent) + x9 + x10$

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
+ x9: factor(dataset\$Continent)	5	50.007	392.34	190.47	3.9257	0.002231 **
<none>		442.35	201.94			
+ x3	1	0.597	441.75	203.70	0.2135	0.644651
+ factor(x13)	1	0.546	441.81	203.72	0.1951	0.659319

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

Step: AIC=190.47

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7 + \text{factor}(\text{dataset}\$Continent) + x9 + x10 + \text{factor}(\text{dataset}\$Continent):x9$

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>		392.34	190.47			
+ x3	1	0.91877	391.42	192.05	0.3591	0.5499
+ factor(x13)	1	0.77978	391.56	192.11	0.3047	0.5818

> AIC<-AIC(forwardAIC)
> BIC<-AIC(forwardAIC, k=log(179))
> summary(forwardAIC)

Call:

`lm(formula = y ~ x1 + x6 + factor(x11) + x8 + x2 + x4 + factor(x12) + x7 + factor(dataset$Continent) + x9 + x10 + factor(dataset$Continent):x9, data = dataset)`

Residuals:

	Min	1Q	Median	3Q	Max
	-5.3644	-0.7512	0.0768	0.9322	3.4615

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.542431	2.725324	2.768	0.006341 **
x1	0.119290	0.015978	7.466	5.72e-12 ***
x6	0.098873	0.004642	21.299	< 2e-16 ***
factor(x11)2	2.262405	0.632204	3.579	0.000462 ***
factor(x11)3	4.569475	0.687618	6.645	4.93e-10 ***
factor(x11)4	6.059266	0.816999	7.416	7.53e-12 ***
x8	0.089450	0.011059	8.088	1.68e-13 ***
x2	0.130075	0.013233	9.830	< 2e-16 ***
x4	0.115381	0.012659	9.114	4.01e-16 ***
factor(x12)2	2.757174	0.508657	5.420	2.25e-07 ***
factor(x12)3	4.531098	0.605302	7.486	5.12e-12 ***
factor(x12)4	7.018784	0.839599	8.360	3.48e-14 ***
x7	0.081320	0.014613	5.565	1.14e-07 ***
factor(dataset\$Continent)1	-10.153438	4.085604	-2.485	0.014018 *
factor(dataset\$Continent)2	-0.013171	4.441321	-0.003	0.997638


```

factor(dataset$Continent)3 -5.149807 2.664937 -1.932 0.055142 .
factor(dataset$Continent)4 10.096763 5.714780 1.767 0.079246 .
factor(dataset$Continent)5 -11.092626 4.542589 -2.442 0.015742 *
x9 0.037543 0.030316 1.238 0.217459
x10 0.083233 0.017742 4.691 5.95e-06 ***
factor(dataset$Continent)1:x9 0.123186 0.054984 2.240 0.026496 *
factor(dataset$Continent)2:x9 -0.018243 0.059431 -0.307 0.759295
factor(dataset$Continent)3:x9 0.063569 0.035633 1.784 0.076395 .
factor(dataset$Continent)4:x9 -0.164037 0.073206 -2.241 0.026472 *
factor(dataset$Continent)5:x9 0.107418 0.057813 1.858 0.065073 .

```

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

Residual standard error: 1.596 on 154 degrees of freedom

Multiple R-squared: 0.98, Adjusted R-squared: 0.9769

F-statistic: 315 on 24 and 154 DF, p-value: < 2.2e-16

```
> print(AIC)
```

```
[1] 700.4506
```

```
> print(BIC)
```

```
[1] 783.3226
```

BIC

```
> #BIC
```

```
> forwardBIC<-step(forwardRegression, direction = "forward",scope = formula(TestModel),k=log(179))
```

```
Start: AIC=846.21
```

```
y ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ x1	1	14615.6	5034.8	607.65
+ x2	1	12865.7	6784.7	661.05
+ x3	1	12439.7	7210.7	671.95
+ x7	1	12395.2	7255.2	673.05
+ factor(x12)	3	11862.1	7788.3	696.11
+ factor(x11)	3	10818.2	8832.2	718.63
+ x10	1	9845.3	9805.0	726.96
+ x9	1	6444.1	13206.3	780.27
+ x6	1	5923.6	13726.8	787.18
+ x8	1	5593.1	14057.3	791.44
+ factor(dataset\$Continent) 5		5374.1	14276.3	814.96
+ factor(x13)	1	1701.5	17948.9	835.19
<none>			19650.4	846.21
+ x4	1	14.6	19635.8	851.27

Step: AIC=607.65

```
y ~ x1
```

	Df	Sum of Sq	RSS	AIC
+ x6	1	1700.82	3334.0	539.05
+ factor(x12)	3	1055.62	3979.2	581.09
+ factor(x11)	3	1015.20	4019.6	582.90
+ x7	1	728.51	4306.3	584.86
+ x8	1	690.66	4344.1	586.43
+ x9	1	622.30	4412.5	589.22
+ x2	1	616.15	4418.6	589.47
+ x4	1	496.47	4538.3	594.26
+ x10	1	454.39	4580.4	595.91
+ x3	1	260.19	4774.6	603.34
<none>			5034.8	607.65
+ factor(x13)	1	8.50	5026.3	612.54
+ factor(dataset\$Continent) 5		42.94	4991.8	632.05

Step: AIC=539.05

```
y ~ x1 + x6
```

	Df	Sum of Sq	RSS	AIC
+ factor(x11)	3	910.21	2423.7	497.54
+ factor(x12)	3	863.40	2470.6	500.97
+ x8	1	693.43	2640.5	502.50
+ x7	1	557.33	2776.6	511.50
+ x2	1	503.69	2830.3	514.92
+ x4	1	486.35	2847.6	516.02
+ x9	1	375.45	2958.5	522.85
+ x10	1	321.72	3012.2	526.08
+ x3	1	210.01	3123.9	532.59
<none>			3334.0	539.05
+ factor(x13)	1	0.02	3333.9	544.24
+ factor(dataset\$Continent)	5	229.18	3104.8	552.24

Step: AIC=497.54
 $y \sim x1 + x6 + \text{factor}(x11)$

	Df	Sum of Sq	RSS	AIC
+ x8	1	706.58	1717.2	441.04
+ x4	1	605.28	1818.5	451.30
+ x7	1	488.78	1935.0	462.41
+ x2	1	483.34	1940.4	462.92
+ factor(x12)	3	335.13	2088.6	486.47
+ x3	1	173.91	2249.8	489.40
+ factor(dataset\$Continent)	5	407.66	2016.1	490.51
+ x10	1	147.76	2276.0	491.47
+ x9	1	100.35	2323.4	495.16
<none>			2423.7	497.54
+ factor(x13)	1	5.55	2418.2	502.32

Step: AIC=441.04
 $y \sim x1 + x6 + \text{factor}(x11) + x8$

	Df	Sum of Sq	RSS	AIC
+ x2	1	330.48	1386.7	407.96
+ x4	1	306.18	1411.0	411.07
+ x7	1	259.09	1458.1	416.95
+ factor(x12)	3	317.99	1399.2	419.94
+ factor(dataset\$Continent)	5	305.10	1412.1	431.96
+ x10	1	123.03	1594.1	432.92
+ x3	1	116.84	1600.3	433.61
+ x9	1	63.35	1653.8	439.50
<none>			1717.2	441.04
+ factor(x13)	1	17.22	1699.9	444.42

Step: AIC=407.96
 $y \sim x1 + x6 + \text{factor}(x11) + x8 + x2$

	Df	Sum of Sq	RSS	AIC
+ x4	1	382.29	1004.4	355.42
+ factor(x12)	3	285.31	1101.4	382.29
+ x7	1	180.17	1206.5	388.24
+ x10	1	146.09	1240.6	393.22
+ factor(dataset\$Continent)	5	220.61	1166.1	402.88
+ x9	1	64.85	1321.8	404.58
<none>			1386.7	407.96
+ factor(x13)	1	3.80	1382.9	412.66
+ x3	1	0.09	1386.6	413.14

Step: AIC=355.42
 $y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4$

	Df	Sum of Sq	RSS	AIC
+ factor(x12)	3	262.797	741.59	316.68
+ x7	1	103.439	900.95	341.15

```

+ x9          1  83.956 920.43 344.98
+ x10         1  75.652 928.73 346.59
<none>              1004.39 355.42
+ x3          1  7.268 997.12 359.30
+ factor(x13)   1  0.918 1003.47 360.44
+ factor(dataset$Continent) 5 101.659 902.73 362.25

```

Step: AIC=316.68

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12)$

```

          Df Sum of Sq  RSS  AIC
+ x7          1  76.454 665.14 302.39
+ x9          1  50.871 690.72 309.15
+ x10         1  50.545 691.04 309.23
<none>              741.59 316.68
+ factor(dataset$Continent) 5  94.763 646.83 318.15
+ x3          1  2.274 739.32 321.32
+ factor(x13)   1  1.792 739.80 321.44

```

Step: AIC=302.39

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7$

```

          Df Sum of Sq  RSS  AIC
+ x9          1  55.600 609.54 291.95
+ factor(dataset$Continent) 5 107.899 557.24 296.65
+ x10         1  36.300 628.84 297.53
<none>              665.14 302.39
+ factor(x13)   1  1.319 663.82 307.23
+ x3          1  0.129 665.01 307.55

```

Step: AIC=291.95

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7 + x9$

```

          Df Sum of Sq  RSS  AIC
+ factor(dataset$Continent) 5 113.529 496.01 281.00
+ x10         1  35.125 574.41 286.52
<none>              609.54 291.95
+ factor(x13)   1  1.253 608.28 296.77
+ x3          1  0.060 609.48 297.12

```

Step: AIC=281

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7 + x9 + \text{factor}(\text{dataset}\$Continent)$

```

          Df Sum of Sq  RSS  AIC
+ x10         1  53.656 442.35 265.69
<none>              496.01 281.00
+ x3          1  0.141 495.87 286.13
+ factor(x13)   1  0.128 495.88 286.14
+ x9: factor(dataset$Continent) 5  47.594 448.41 288.88

```

Step: AIC=265.69

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7 + x9 + \text{factor}(\text{dataset}\$Continent) + x10$

```

          Df Sum of Sq  RSS  AIC
<none>              442.35 265.69
+ x9: factor(dataset$Continent) 5  50.007 392.34 270.15
+ x3          1  0.597 441.75 270.64
+ factor(x13)   1  0.546 441.81 270.66
> AIC<-AIC(forwardBIC)
> BIC<-AIC(forwardBIC, k=log(179))
> summary(forwardBIC)

```

Call:
lm(formula = y ~ x1 + x6 + factor(x11) + x8 + x2 + x4 + factor(x12) +
x7 + x9 + factor(dataset\$Continent) + x10, data = dataset)

Residuals:

Min 1Q Median 3Q Max
-5.8992 -0.9168 0.1411 1.0934 4.4043

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.441452	1.713048	2.593	0.0104 *
x1	0.126948	0.016306	7.785	8.37e-13 ***
x6	0.098250	0.004747	20.698	< 2e-16 ***
factor(x11)2	2.745054	0.625313	4.390	2.06e-05 ***
factor(x11)3	5.119219	0.693433	7.382	8.19e-12 ***
factor(x11)4	6.647644	0.839982	7.914	4.00e-13 ***
x8	0.093167	0.011381	8.186	8.25e-14 ***
x2	0.125875	0.013705	9.185	< 2e-16 ***
x4	0.115682	0.013105	8.827	1.85e-15 ***
factor(x12)2	2.358283	0.516314	4.568	9.84e-06 ***
factor(x12)3	4.105031	0.615637	6.668	4.06e-10 ***
factor(x12)4	6.441983	0.851448	7.566	2.92e-12 ***
x7	0.079284	0.014824	5.348	3.04e-07 ***
x9	0.078110	0.017132	4.559	1.02e-05 ***
factor(dataset\$Continent)1	-0.973881	0.468368	-2.079	0.0392 *
factor(dataset\$Continent)2	-1.367764	0.599009	-2.283	0.0237 *
factor(dataset\$Continent)3	-0.422594	0.420842	-1.004	0.3168
factor(dataset\$Continent)4	-2.978579	0.621956	-4.789	3.81e-06 ***
factor(dataset\$Continent)5	-2.745970	0.492190	-5.579	1.02e-07 ***
x10	0.079033	0.017996	4.392	2.05e-05 ***

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

Residual standard error: 1.668 on 159 degrees of freedom

Multiple R-squared: 0.9775, Adjusted R-squared: 0.9748

F-statistic: 363.4 on 19 and 159 DF, p-value: < 2.2e-16

> print(AIC)

[1] 711.9244

> print(BIC)

[1] 778.8595

חלק אחרון – רגרסיה בצעדים

> #-----steps-----

>

> #stepwise regression

AIC

> #stepwise regression

> #AIC

> StepwiseAIC<-step(forwardRegression,direction = "both",scope = formula(TestModel))

Start: AIC=843.03

y ~ 1

	Df	Sum of Sq	RSS	AIC
+ x1	1	14615.6	5034.8	601.28
+ x2	1	12865.7	6784.7	654.67
+ x3	1	12439.7	7210.7	665.57
+ x7	1	12395.2	7255.2	666.67
+ factor(x12)	3	11862.1	7788.3	683.37

```

+ factor(x11)      3 10818.2 8832.2 705.88
+ x10              1  9845.3 9805.0 720.58
+ x9              1  6444.1 13206.3 773.89
+ x6              1  5923.6 13726.8 780.81
+ x8              1  5593.1 14057.3 785.07
+ factor(dataset$Continent) 5  5374.1 14276.3 795.84
+ factor(x13)      1  1701.5 17948.9 828.81
<none>            19650.4 843.03
+ x4              1   14.6 19635.8 844.89

```

Step: AIC=601.28
y ~ x1

	Df	Sum of Sq	RSS	AIC
+ x6	1	1700.8	3334.0	529.49
+ factor(x12)	3	1055.6	3979.2	565.16
+ factor(x11)	3	1015.2	4019.6	566.97
+ x7	1	728.5	4306.3	575.30
+ x8	1	690.7	4344.1	576.87
+ x9	1	622.3	4412.5	579.66
+ x2	1	616.2	4418.6	579.91
+ x4	1	496.5	4538.3	584.69
+ x10	1	454.4	4580.4	586.35
+ x3	1	260.2	4774.6	593.78
<none>			5034.8	601.28
+ factor(x13)	1	8.5	5026.3	602.97
+ factor(dataset\$Continent)	5	42.9	4991.8	609.74
- x1	1	14615.6	19650.4	843.03

Step: AIC=529.49
y ~ x1 + x6

	Df	Sum of Sq	RSS	AIC
+ factor(x11)	3	910.2	2423.7	478.42
+ factor(x12)	3	863.4	2470.6	481.84
+ x8	1	693.4	2640.5	489.75
+ x7	1	557.3	2776.6	498.75
+ x2	1	503.7	2830.3	502.17
+ x4	1	486.3	2847.6	503.27
+ x9	1	375.5	2958.5	510.10
+ x10	1	321.7	3012.2	513.33
+ x3	1	210.0	3123.9	519.84
+ factor(dataset\$Continent)	5	229.2	3104.8	526.74
<none>			3334.0	529.49
+ factor(x13)	1	0.0	3333.9	531.49
- x6	1	1700.8	5034.8	601.28
- x1	1	10392.8	13726.8	780.81

Step: AIC=478.42
y ~ x1 + x6 + factor(x11)

	Df	Sum of Sq	RSS	AIC
+ x8	1	706.6	1717.2	418.73
+ x4	1	605.3	1818.5	428.99
+ x7	1	488.8	1935.0	440.10
+ x2	1	483.3	1940.4	440.60
+ factor(dataset\$Continent)	5	407.7	2016.1	455.45
+ factor(x12)	3	335.1	2088.6	457.78
+ x3	1	173.9	2249.8	467.09
+ x10	1	147.8	2276.0	469.16
+ x9	1	100.4	2323.4	472.85
<none>			2423.7	478.42
+ factor(x13)	1	5.6	2418.2	480.01
- factor(x11)	3	910.2	3334.0	529.49
- x6	1	1595.8	4019.6	566.97

- x1 1 3802.5 6226.2 645.30

Step: AIC=418.73

y ~ x1 + x6 + factor(x11) + x8

	Df	Sum of Sq	RSS	AIC
+ x2	1	330.48	1386.7	382.46
+ x4	1	306.18	1411.0	385.57
+ factor(x12)	3	317.99	1399.2	388.07
+ x7	1	259.09	1458.1	391.45
+ factor(dataset\$Continent)	5	305.10	1412.1	393.71
+ x10	1	123.03	1594.1	407.42
+ x3	1	116.84	1600.3	408.11
+ x9	1	63.35	1653.8	414.00
<none>			1717.2	418.73
+ factor(x13)	1	17.22	1699.9	418.92
- x8	1	706.58	2423.7	478.42
- factor(x11)	3	923.37	2640.5	489.75
- x6	1	1585.66	3302.8	533.81
- x1	1	2477.15	4194.3	576.58

Step: AIC=382.46

y ~ x1 + x6 + factor(x11) + x8 + x2

	Df	Sum of Sq	RSS	AIC
+ x4	1	382.29	1004.4	326.73
+ factor(x12)	3	285.31	1101.4	347.23
+ x7	1	180.17	1206.5	359.55
+ factor(dataset\$Continent)	5	220.61	1166.1	361.45
+ x10	1	146.09	1240.6	364.54
+ x9	1	64.85	1321.8	375.89
<none>			1386.7	382.46
+ factor(x13)	1	3.80	1382.9	383.97
+ x3	1	0.09	1386.6	384.45
- x2	1	330.48	1717.2	418.73
- x8	1	553.73	1940.4	440.60
- x1	1	580.21	1966.9	443.03
- factor(x11)	3	918.59	2305.3	467.45
- x6	1	1491.77	2878.4	511.19

Step: AIC=326.73

y ~ x1 + x6 + factor(x11) + x8 + x2 + x4

	Df	Sum of Sq	RSS	AIC
+ factor(x12)	3	262.80	741.59	278.43
+ x7	1	103.44	900.95	309.28
+ x9	1	83.96	920.43	313.10
+ x10	1	75.65	928.73	314.71
+ factor(dataset\$Continent)	5	101.66	902.73	317.63
<none>			1004.39	326.73
+ x3	1	7.27	997.12	327.43
+ factor(x13)	1	0.92	1003.47	328.57
- x8	1	266.54	1270.93	366.86
- x4	1	382.29	1386.68	382.46
- x2	1	406.59	1410.98	385.57
- x1	1	624.75	1629.14	411.31
- factor(x11)	3	985.60	1989.99	443.12
- x6	1	1442.09	2446.47	484.09

Step: AIC=278.43

y ~ x1 + x6 + factor(x11) + x8 + x2 + x4 + factor(x12)

	Df	Sum of Sq	RSS	AIC
+ x7	1	76.45	665.14	260.96
+ factor(dataset\$Continent)	5	94.76	646.83	263.96

```

+ x9          1  50.87 690.72 267.71
+ x10         1  50.55 691.04 267.80
<none>              741.59 278.43
+ x3          1  2.27 739.32 279.88
+ factor(x13)   1  1.79 739.80 280.00
- factor(x12)   3  262.80 1004.39 326.73
- x8           1  255.24 996.83 329.38
- x4           1  359.78 1101.37 347.23
- x2           1  360.64 1102.23 347.37
- x1           1  375.37 1116.96 349.74
- factor(x11)   3  431.04 1172.63 354.45
- x6           1  1305.89 2047.48 458.22

```

Step: AIC=260.96

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7$

```

              Df Sum of Sq  RSS  AIC
+ factor(dataset$Continent) 5  107.90 557.24 239.27
+ x9          1  55.60 609.54 247.33
+ x10         1  36.30 628.84 252.91
<none>              665.14 260.96
+ factor(x13)   1  1.32 663.82 262.60
+ x3          1  0.13 665.01 262.92
- x7          1  76.45 741.59 278.43
- x1          1  199.41 864.55 305.89
- x8          1  205.75 870.89 307.20
- factor(x12)   3  235.81 900.95 309.28
- x2          1  292.11 957.25 324.13
- x4          1  295.53 960.66 324.76
- factor(x11)   3  431.75 1096.88 344.50
- x6          1  1253.54 1918.67 448.59

```

Step: AIC=239.27

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7 + \text{factor}(\text{dataset}\$Continent)$

```

              Df Sum of Sq  RSS  AIC
+ x9          1  61.23 496.01 220.44
+ x10         1  57.06 500.18 221.94
<none>              557.24 239.27
+ x3          1  0.89 556.35 240.99
+ factor(x13)   1  0.00 557.24 241.27
- factor(dataset$Continent) 5  107.90 665.14 260.96
- x7          1  89.59 646.83 263.96
- x8          1  206.04 763.28 293.59
- factor(x12)   3  229.83 787.07 295.09
- x2          1  220.70 777.94 297.00
- x4          1  230.57 787.81 299.25
- x1          1  256.63 813.87 305.08
- factor(x11)   3  355.47 912.71 321.60
- x6          1  1323.58 1880.82 455.02

```

Step: AIC=220.44

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7 + \text{factor}(\text{dataset}\$Continent) + x9$

```

              Df Sum of Sq  RSS  AIC
+ x10         1  53.66 442.35 201.94
+ x9:factor(dataset$Continent) 5  47.59 448.41 212.38
<none>              496.01 220.44
+ x3          1  0.14 495.87 222.39
+ factor(x13)   1  0.13 495.88 222.39
- x9          1  61.23 557.24 239.27
- factor(dataset$Continent) 5  113.53 609.54 247.33
- x7          1  92.70 588.71 249.11

```

```

- factor(x12)      3  197.97 693.98 274.55
- x8               1  196.96 692.97 278.29
- x1               1  216.32 712.32 283.23
- x2               1  226.16 722.17 285.68
- x4               1  238.76 734.77 288.78
- factor(x11)      3  272.54 768.55 292.82
- x6               1 1183.83 1679.84 436.79

```

Step: AIC=201.94

```

y ~ x1 + x6 + factor(x11) + x8 + x2 + x4 + factor(x12) + x7 +
  factor(dataset$Continent) + x9 + x10

```

	Df	Sum of Sq	RSS	AIC
+ x9: factor(dataset\$Continent)	5	50.01	392.34	190.47
<none>			442.35	201.94
+ x3	1	0.60	441.75	203.70
+ factor(x13)	1	0.55	441.81	203.72
- x10	1	53.66	496.01	220.44
- x9	1	57.83	500.18	221.94
- x7	1	79.58	521.93	229.56
- factor(dataset\$Continent)	5	132.06	574.41	238.71
- factor(x12)	3	172.15	614.50	254.78
- x1	1	168.63	610.98	257.75
- x8	1	186.42	628.77	262.89
- x4	1	216.77	659.12	271.33
- factor(x11)	3	234.26	676.61	272.02
- x2	1	234.69	677.04	276.13
- x6	1	1191.91	1634.26	433.87

Step: AIC=190.47

```

y ~ x1 + x6 + factor(x11) + x8 + x2 + x4 + factor(x12) + x7 +
  factor(dataset$Continent) + x9 + x10 + factor(dataset$Continent): x9

```

	Df	Sum of Sq	RSS	AIC
<none>			392.34	190.47
+ x3	1	0.92	391.42	192.05
+ factor(x13)	1	0.78	391.56	192.11
- factor(dataset\$Continent): x9	5	50.01	442.35	201.94
- x10	1	56.07	448.41	212.38
- x7	1	78.90	471.24	221.27
- x1	1	142.01	534.35	243.77
- x8	1	166.68	559.02	251.84
- factor(x12)	3	190.10	582.44	255.19
- factor(x11)	3	202.54	594.88	258.98
- x4	1	211.64	603.99	265.69
- x2	1	246.17	638.51	275.64
- x6	1	1155.75	1548.10	434.17

```

> AIC<-AIC(StepwiseAIC)
> BIC<-AIC(StepwiseAIC, k=log(179))
> summary(StepwiseAIC)

```

Call:

```

lm(formula = y ~ x1 + x6 + factor(x11) + x8 + x2 + x4 + factor(x12) +
  x7 + factor(dataset$Continent) + x9 + x10 + factor(dataset$Continent): x9,
  data = dataset)

```

Residuals:

```

      Min       1Q   Median       3Q      Max
-5.3644 -0.7512  0.0768  0.9322  3.4615

```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.542431	2.725324	2.768	0.006341 **
x1	0.119290	0.015978	7.466	5.72e-12 ***
x6	0.098873	0.004642	21.299	< 2e-16 ***


```

factor(x11)2      2.262405  0.632204  3.579 0.000462 ***
factor(x11)3      4.569475  0.687618  6.645 4.93e-10 ***
factor(x11)4      6.059266  0.816999  7.416 7.53e-12 ***
x8                0.089450  0.011059  8.088 1.68e-13 ***
x2                0.130075  0.013233  9.830 < 2e-16 ***
x4                0.115381  0.012659  9.114 4.01e-16 ***
factor(x12)2      2.757174  0.508657  5.420 2.25e-07 ***
factor(x12)3      4.531098  0.605302  7.486 5.12e-12 ***
factor(x12)4      7.018784  0.839599  8.360 3.48e-14 ***
x7                0.081320  0.014613  5.565 1.14e-07 ***
factor(dataset$Continent)1 -10.153438  4.085604 -2.485 0.014018 *
factor(dataset$Continent)2 -0.013171  4.441321 -0.003 0.997638
factor(dataset$Continent)3 -5.149807  2.664937 -1.932 0.055142 .
factor(dataset$Continent)4  10.096763  5.714780  1.767 0.079246 .
factor(dataset$Continent)5 -11.092626  4.542589 -2.442 0.015742 *
x9                0.037543  0.030316  1.238 0.217459
x10               0.083233  0.017742  4.691 5.95e-06 ***
factor(dataset$Continent)1:x9 0.123186  0.054984  2.240 0.026496 *
factor(dataset$Continent)2:x9 -0.018243  0.059431 -0.307 0.759295
factor(dataset$Continent)3:x9 0.063569  0.035633  1.784 0.076395 .
factor(dataset$Continent)4:x9 -0.164037  0.073206 -2.241 0.026472 *
factor(dataset$Continent)5:x9 0.107418  0.057813  1.858 0.065073 .

```

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

Residual standard error: 1.596 on 154 degrees of freedom

Multiple R-squared: 0.98, Adjusted R-squared: 0.9769

F-statistic: 315 on 24 and 154 DF, p-value: < 2.2e-16

```
> print(AIC)
```

```
[1] 700.4506
```

```
> print(BIC)
```

```
[1] 783.3226
```

BIC

```
#BIC
```

```
> StepwiseBIC<-step(forwardRegression,direction = "both",scope=formula(TestModel), k=log(179))
```

```
Start: AIC=846.21
```

```
y ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ x1	1	14615.6	5034.8	607.65
+ x2	1	12865.7	6784.7	661.05
+ x3	1	12439.7	7210.7	671.95
+ x7	1	12395.2	7255.2	673.05
+ factor(x12)	3	11862.1	7788.3	696.11
+ factor(x11)	3	10818.2	8832.2	718.63
+ x10	1	9845.3	9805.0	726.96
+ x9	1	6444.1	13206.3	780.27
+ x6	1	5923.6	13726.8	787.18
+ x8	1	5593.1	14057.3	791.44
+ factor(dataset\$Continent)	5	5374.1	14276.3	814.96
+ factor(x13)	1	1701.5	17948.9	835.19
<none>			19650.4	846.21
+ x4	1	14.6	19635.8	851.27

```
Step: AIC=607.65
```

```
y ~ x1
```

	Df	Sum of Sq	RSS	AIC
+ x6	1	1700.8	3334.0	539.05
+ factor(x12)	3	1055.6	3979.2	581.09
+ factor(x11)	3	1015.2	4019.6	582.90
+ x7	1	728.5	4306.3	584.86

```

+ x8          1  690.7 4344.1 586.43
+ x9          1  622.3 4412.5 589.22
+ x2          1  616.2 4418.6 589.47
+ x4          1  496.5 4538.3 594.26
+ x10         1  454.4 4580.4 595.91
+ x3          1  260.2 4774.6 603.34
<none>              5034.8 607.65
+ factor(x13)      1    8.5 5026.3 612.54
+ factor(dataset$Continent) 5   42.9 4991.8 632.05
- x1            1 14615.6 19650.4 846.21

```

Step: AIC=539.05

y ~ x1 + x6

	Df	Sum of Sq	RSS	AIC
+ factor(x11)	3	910.2	2423.7	497.54
+ factor(x12)	3	863.4	2470.6	500.97
+ x8	1	693.4	2640.5	502.50
+ x7	1	557.3	2776.6	511.50
+ x2	1	503.7	2830.3	514.92
+ x4	1	486.3	2847.6	516.02
+ x9	1	375.5	2958.5	522.85
+ x10	1	321.7	3012.2	526.08
+ x3	1	210.0	3123.9	532.59
<none>			3334.0	539.05
+ factor(x13)	1	0.0	3333.9	544.24
+ factor(dataset\$Continent)	5	229.2	3104.8	552.24
- x6	1	1700.8	5034.8	607.65
- x1	1	10392.8	13726.8	787.18

Step: AIC=497.54

y ~ x1 + x6 + factor(x11)

	Df	Sum of Sq	RSS	AIC
+ x8	1	706.6	1717.2	441.04
+ x4	1	605.3	1818.5	451.30
+ x7	1	488.8	1935.0	462.41
+ x2	1	483.3	1940.4	462.92
+ factor(x12)	3	335.1	2088.6	486.47
+ x3	1	173.9	2249.8	489.40
+ factor(dataset\$Continent)	5	407.7	2016.1	490.51
+ x10	1	147.8	2276.0	491.47
+ x9	1	100.4	2323.4	495.16
<none>			2423.7	497.54
+ factor(x13)	1	5.6	2418.2	502.32
- factor(x11)	3	910.2	3334.0	539.05
- x6	1	1595.8	4019.6	582.90
- x1	1	3802.5	6226.2	661.23

Step: AIC=441.04

y ~ x1 + x6 + factor(x11) + x8

	Df	Sum of Sq	RSS	AIC
+ x2	1	330.48	1386.7	407.96
+ x4	1	306.18	1411.0	411.07
+ x7	1	259.09	1458.1	416.95
+ factor(x12)	3	317.99	1399.2	419.94
+ factor(dataset\$Continent)	5	305.10	1412.1	431.96
+ x10	1	123.03	1594.1	432.92
+ x3	1	116.84	1600.3	433.61
+ x9	1	63.35	1653.8	439.50
<none>			1717.2	441.04
+ factor(x13)	1	17.22	1699.9	444.42
- x8	1	706.58	2423.7	497.54
- factor(x11)	3	923.37	2640.5	502.50

```
- x6          1 1585.66 3302.8 552.94
- x1          1 2477.15 4194.3 595.71
```

Step: AIC=407.96

y ~ x1 + x6 + factor(x11) + x8 + x2

	Df	Sum of Sq	RSS	AIC
+ x4	1	382.29	1004.4	355.42
+ factor(x12)	3	285.31	1101.4	382.29
+ x7	1	180.17	1206.5	388.24
+ x10	1	146.09	1240.6	393.22
+ factor(dataset\$Continent)	5	220.61	1166.1	402.88
+ x9	1	64.85	1321.8	404.58
<none>		1386.7	407.96	
+ factor(x13)	1	3.80	1382.9	412.66
+ x3	1	0.09	1386.6	413.14
- x2	1	330.48	1717.2	441.04
- x8	1	553.73	1940.4	462.92
- x1	1	580.21	1966.9	465.34
- factor(x11)	3	918.59	2305.3	483.38
- x6	1	1491.77	2878.4	533.51

Step: AIC=355.42

y ~ x1 + x6 + factor(x11) + x8 + x2 + x4

	Df	Sum of Sq	RSS	AIC
+ factor(x12)	3	262.80	741.59	316.68
+ x7	1	103.44	900.95	341.15
+ x9	1	83.96	920.43	344.98
+ x10	1	75.65	928.73	346.59
<none>		1004.39	355.42	
+ x3	1	7.27	997.12	359.30
+ factor(x13)	1	0.92	1003.47	360.44
+ factor(dataset\$Continent)	5	101.66	902.73	362.25
- x8	1	266.54	1270.93	392.36
- x4	1	382.29	1386.68	407.96
- x2	1	406.59	1410.98	411.07
- x1	1	624.75	1629.14	436.81
- factor(x11)	3	985.60	1989.99	462.25
- x6	1	1442.09	2446.47	509.59

Step: AIC=316.68

y ~ x1 + x6 + factor(x11) + x8 + x2 + x4 + factor(x12)

	Df	Sum of Sq	RSS	AIC
+ x7	1	76.45	665.14	302.39
+ x9	1	50.87	690.72	309.15
+ x10	1	50.55	691.04	309.23
<none>		741.59	316.68	
+ factor(dataset\$Continent)	5	94.76	646.83	318.15
+ x3	1	2.27	739.32	321.32
+ factor(x13)	1	1.79	739.80	321.44
- factor(x12)	3	262.80	1004.39	355.42
- x8	1	255.24	996.83	364.44
- x4	1	359.78	1101.37	382.29
- x2	1	360.64	1102.23	382.43
- factor(x11)	3	431.04	1172.63	383.14
- x1	1	375.37	1116.96	384.81
- x6	1	1305.89	2047.48	493.28

Step: AIC=302.39

y ~ x1 + x6 + factor(x11) + x8 + x2 + x4 + factor(x12) + x7

	Df	Sum of Sq	RSS	AIC
+ x9	1	55.60	609.54	291.95

```

+ factor(dataset$Continent) 5 107.90 557.24 296.65
+ x10 1 36.30 628.84 297.53
<none> 665.14 302.39
+ factor(x13) 1 1.32 663.82 307.22
+ x3 1 0.13 665.01 307.55
- x7 1 76.45 741.59 316.68
- factor(x12) 3 235.81 900.95 341.15
- x1 1 199.41 864.55 344.14
- x8 1 205.75 870.89 345.45
- x2 1 292.11 957.25 362.37
- x4 1 295.53 960.66 363.01
- factor(x11) 3 431.75 1096.88 376.37
- x6 1 1253.54 1918.67 486.84

```

Step: AIC=291.95

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7 + x9$

	Df	Sum of Sq	RSS	AIC
+ factor(dataset\$Continent)	5	113.53	496.01	281.00
+ x10	1	35.12	574.41	286.52
<none>			609.54	291.95
+ factor(x13)	1	1.25	608.28	296.77
+ x3	1	0.06	609.48	297.12
- x9	1	55.60	665.14	302.39
- x7	1	81.18	690.72	309.15
- factor(x12)	3	204.51	814.05	328.18
- x1	1	164.28	773.82	329.48
- x8	1	193.67	803.21	336.16
- factor(x11)	3	345.67	955.21	356.80
- x2	1	293.25	902.79	357.08
- x4	1	300.12	909.65	358.43
- x6	1	1118.54	1728.08	473.30

Step: AIC=281

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7 + x9 + \text{factor}(\text{dataset\$Continent})$

	Df	Sum of Sq	RSS	AIC
+ x10	1	53.66	442.35	265.69
<none>			496.01	281.00
+ x3	1	0.14	495.87	286.13
+ factor(x13)	1	0.13	495.88	286.14
+ x9: factor(dataset\$Continent)	5	47.59	448.41	288.88
- factor(dataset\$Continent)	5	113.53	609.54	291.95
- x9	1	61.23	557.24	296.65
- x7	1	92.70	588.71	306.48
- factor(x12)	3	197.97	693.98	325.55
- x8	1	196.96	692.97	335.67
- x1	1	216.32	712.32	340.60
- x2	1	226.16	722.17	343.06
- factor(x11)	3	272.54	768.55	343.82
- x4	1	238.76	734.77	346.15
- x6	1	1183.83	1679.84	494.17

Step: AIC=265.69

$y \sim x1 + x6 + \text{factor}(x11) + x8 + x2 + x4 + \text{factor}(x12) + x7 + x9 + \text{factor}(\text{dataset\$Continent}) + x10$

	Df	Sum of Sq	RSS	AIC
<none>			442.35	265.69
+ x9: factor(dataset\$Continent)	5	50.01	392.34	270.16
+ x3	1	0.60	441.75	270.64
+ factor(x13)	1	0.55	441.81	270.66
- x10	1	53.66	496.01	281.00

```

- x9          1  57.83 500.18 282.50
- factor(dataset$Continent) 5 132.06 574.41 286.52
- x7          1  79.58 521.93 290.12
- factor(x12)   3 172.15 614.50 308.97
- x1          1 168.63 610.98 318.31
- x8          1 186.42 628.77 323.45
- factor(x11)   3 234.26 676.61 326.20
- x4          1 216.77 659.12 331.89
- x2          1 234.69 677.04 336.69
- x6          1 1191.91 1634.26 494.43
> AIC<-AIC(StepwiseBIC)
> BIC<-AIC(StepwiseBIC, k=log(179))
> summary(StepwiseBIC)

```

Call:

```
lm(formula = y ~ x1 + x6 + factor(x11) + x8 + x2 + x4 + factor(x12) +
  x7 + x9 + factor(dataset$Continent) + x10, data = dataset)
```

Residuals:

```

      Min       1Q   Median       3Q      Max
-5.8992 -0.9168  0.1411  1.0934  4.4043

```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.441452	1.713048	2.593	0.0104 *
x1	0.126948	0.016306	7.785	8.37e-13 ***
x6	0.098250	0.004747	20.698	< 2e-16 ***
factor(x11)2	2.745054	0.625313	4.390	2.06e-05 ***
factor(x11)3	5.119219	0.693433	7.382	8.19e-12 ***
factor(x11)4	6.647644	0.839982	7.914	4.00e-13 ***
x8	0.093167	0.011381	8.186	8.25e-14 ***
x2	0.125875	0.013705	9.185	< 2e-16 ***
x4	0.115682	0.013105	8.827	1.85e-15 ***
factor(x12)2	2.358283	0.516314	4.568	9.84e-06 ***
factor(x12)3	4.105031	0.615637	6.668	4.06e-10 ***
factor(x12)4	6.441983	0.851448	7.566	2.92e-12 ***
x7	0.079284	0.014824	5.348	3.04e-07 ***
x9	0.078110	0.017132	4.559	1.02e-05 ***
factor(dataset\$Continent)1	-0.973881	0.468368	-2.079	0.0392 *
factor(dataset\$Continent)2	-1.367764	0.599009	-2.283	0.0237 *
factor(dataset\$Continent)3	-0.422594	0.420842	-1.004	0.3168
factor(dataset\$Continent)4	-2.978579	0.621956	-4.789	3.81e-06 ***
factor(dataset\$Continent)5	-2.745970	0.492190	-5.579	1.02e-07 ***
x10	0.079033	0.017996	4.392	2.05e-05 ***

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

Residual standard error: 1.668 on 159 degrees of freedom

Multiple R-squared: 0.9775, Adjusted R-squared: 0.9748

F-statistic: 363.4 on 19 and 159 DF, p-value: < 2.2e-16

```

> print(AIC)
[1] 711.9244
> print(BIC)
[1] 778.8595

```

המודל הסופי – לפני שיפור

#-----Choosing the Model-----
>

```
> FinalModel<-lm(formula = y ~ x1 + x2 + x3 + x4 + x6 + x7 + x8 + x9 + x10 +
+ factor(dataset$Continent) + factor(x12) + factor(x11) + x3: factor(x12) +
+ x9: factor(dataset$Continent))
> anova(FinalModel)
Analysis of Variance Table
```

```
Response: y
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
x1	1	14615.6	14615.6	5804.9621	< 2.2e-16 ***
x2	1	616.2	616.2	244.7212	< 2.2e-16 ***
x3	1	2.4	2.4	0.9424	0.333219
x4	1	613.0	613.0	243.4831	< 2.2e-16 ***
x6	1	1566.4	1566.4	622.1219	< 2.2e-16 ***
x7	1	198.7	198.7	78.9125	1.865e-15 ***
x8	1	205.4	205.4	81.5726	7.761e-16 ***
x9	1	336.0	336.0	133.4336	< 2.2e-16 ***
x10	1	165.9	165.9	65.8943	1.597e-13 ***
factor(dataset\$Continent)	5	293.3	58.7	23.3019	< 2.2e-16 ***
factor(x12)	3	365.5	121.8	48.3861	< 2.2e-16 ***
factor(x11)	3	230.3	76.8	30.4942	1.890e-15 ***
x3: factor(x12)	3	11.1	3.7	1.4704	0.224943
x9: factor(dataset\$Continent)	5	53.0	10.6	4.2085	0.001313 **
Residuals	150	377.7	2.5		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> print(FinalModel)
```

```
Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x6 + x7 + x8 + x9 + x10 +
    factor(dataset$Continent) + factor(x12) + factor(x11) + x3: factor(x12) +
    x9: factor(dataset$Continent))
```

Coefficients:

(Intercept)	x1
7.76037	0.11789
x2	x3
0.12949	-0.05263
x4	x6
0.11901	0.09824
x7	x8
0.08091	0.09213
x9	x10
0.04517	0.08816
factor(dataset\$Continent)1	factor(dataset\$Continent)2
-8.30812	1.84356
factor(dataset\$Continent)3	factor(dataset\$Continent)4
-5.32969	11.10217
factor(dataset\$Continent)5	factor(x12)2
-10.96808	1.76775
factor(x12)3	factor(x12)4
1.73351	4.38971
factor(x11)2	factor(x11)3
2.44447	4.83407
factor(x11)4	x3: factor(x12)2
6.14058	0.03228
x3: factor(x12)3	x3: factor(x12)4
0.07630	0.06651
x9: factor(dataset\$Continent)1	x9: factor(dataset\$Continent)2
0.09913	-0.04472
x9: factor(dataset\$Continent)3	x9: factor(dataset\$Continent)4
0.06838	-0.17595
x9: factor(dataset\$Continent)5	
0.10515	

```
> summary(FinalModel)
```

Call:

```
lm(formula = y ~ x1 + x2 + x3 + x4 + x6 + x7 + x8 + x9 + x10 +  
  factor(dataset$Continent) + factor(x12) + factor(x11) + x3: factor(x12) +  
  x9: factor(dataset$Continent))
```

Residuals:

```
Min    1Q  Median    3Q   Max  
-5.2203 -0.7927  0.0209  1.0027  3.1184
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)  
(Intercept)      7.760369  2.945561  2.635 0.009308 **  
x1                0.117892  0.016520  7.136 3.82e-11 ***  
x2                0.129485  0.016064  8.060 2.21e-13 ***  
x3               -0.052628  0.049690 -1.059 0.291240  
x4                0.119008  0.012981  9.168 3.47e-16 ***  
x6                0.098241  0.004655 21.103 < 2e-16 ***  
x7                0.080912  0.014731  5.492 1.66e-07 ***  
x8                0.092126  0.011166  8.251 7.43e-14 ***  
x9                0.045169  0.030494  1.481 0.140634  
x10               0.088162  0.017898  4.926 2.19e-06 ***  
factor(dataset$Continent)1 -8.308124  4.150375 -2.002 0.047112 *  
factor(dataset$Continent)2  1.843556  4.522361  0.408 0.684109  
factor(dataset$Continent)3 -5.329688  2.672951 -1.994 0.047972 *  
factor(dataset$Continent)4 11.102173  5.743520  1.933 0.055121 .  
factor(dataset$Continent)5 -10.968077  4.581512 -2.394 0.017902 *  
factor(x12)2       1.767749  1.633732  1.082 0.280975  
factor(x12)3       1.733506  1.705142  1.017 0.310964  
factor(x12)4       4.389707  3.266500  1.344 0.181024  
factor(x11)2       2.444474  0.650655  3.757 0.000246 ***  
factor(x11)3       4.834075  0.696488  6.941 1.10e-10 ***  
factor(x11)4       6.140582  0.819048  7.497 5.27e-12 ***  
x3: factor(x12)2    0.032280  0.050432  0.640 0.523114  
x3: factor(x12)3    0.076302  0.049478  1.542 0.125148  
x3: factor(x12)4    0.066511  0.059060  1.126 0.261898  
x9: factor(dataset$Continent)1 0.099130  0.055775  1.777 0.077542 .  
x9: factor(dataset$Continent)2 -0.044716  0.060702 -0.737 0.462483  
x9: factor(dataset$Continent)3 0.068376  0.035802  1.910 0.058063 .  
x9: factor(dataset$Continent)4 -0.175946  0.073564 -2.392 0.018007 *  
x9: factor(dataset$Continent)5 0.105151  0.058149  1.808 0.072561 .  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 1.587 on 150 degrees of freedom

Multiple R-squared: 0.9808, Adjusted R-squared: 0.9772

F-statistic: 273.4 on 28 and 150 DF, p-value: < 2.2e-16

```
> AIC<-AIC(FinalModel)  
> BIC<-AIC(FinalModel, k=log(179))  
> print(AIC)  
[1] 701.6263  
> print(BIC)  
[1] 797.2479
```

בחינת המודל הסופי – לאחר שיפור

```
NewFM<-lm(formula = y ~ x1^2 + sqrt(x2) + x4 + x6 + x7 + x8 + x9 + x10 +  
  factor(dataset$Continent) +  
  + factor(dataset$Financial_Freedom) + factor(dataset$Investment_Freedom) +  
  + x3: factor(dataset$Financial_Freedom) + x9: factor(dataset$Continent),  
  + data = dataset)
```

חלק 5 - רגרסיה לאחור

AIC

```
> #-----Backward-----
> #-----AIC-----
> NewBackAIC = step(NewFM , direction = "backward")
Start: AIC=187.95
y ~ x1^2 + sqrt(x2) + x4 + x6 + x7 + x8 + x9 + x10 + factor(dataset$Continent) +
  factor(dataset$Financial_Freedom) + factor(dataset$Investment_Freedom) +
  x3: factor(dataset$Financial_Freedom) + x9: factor(dataset$Continent)
```

```

              Df Sum of Sq  RSS   AIC
<none>                 369.94 187.95
- factor(dataset$Financial_Freedom): x3  4    23.68  393.62 191.05
- x9: factor(dataset$Continent)         5    47.76  417.70 199.68
- x10                                    1    62.19  432.13 213.76
- x7                                    1    68.54  438.49 216.37
- x1                                    1   127.88  497.82 239.09
- sqrt(x2)                             1   171.31  541.25 254.06
- x8                                    1   173.94  543.89 254.93
- factor(dataset$Investment_Freedom)  3   203.68  573.62 260.46
- x4                                    1   212.23  582.17 267.11
- x6                                    1  1096.63 1466.58 432.49
> summary(NewBackAIC)
```

Call:

```
lm(formula = y ~ x1^2 + sqrt(x2) + x4 + x6 + x7 + x8 + x9 + x10 +
  factor(dataset$Continent) + factor(dataset$Financial_Freedom) +
  factor(dataset$Investment_Freedom) + x3: factor(dataset$Financial_Freedom) +
  x9: factor(dataset$Continent), data = dataset)
```

Residuals:

```

      Min       1Q   Median       3Q      Max
-5.3093 -0.8114  0.0792  0.8807  3.1644
```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.296594   3.026366   0.759 0.449126
x1                0.117645   0.016338   7.201 2.69e-11 ***
sqrt(x2)         1.656639   0.198776   8.334 4.60e-14 ***
x4               0.119184   0.012848   9.276 < 2e-16 ***
x6               0.097407   0.004619  21.087 < 2e-16 ***
x7               0.077048   0.014615   5.272 4.63e-07 ***
x8               0.092739   0.011043   8.398 3.18e-14 ***
x9               0.048258   0.030203   1.598 0.112194
x10              0.088957   0.017716   5.021 1.44e-06 ***
factor(dataset$Continent)1  -8.680377   4.109472  -2.112 0.036319 *
factor(dataset$Continent)2   1.419713   4.476486   0.317 0.751572
factor(dataset$Continent)3  -4.232064   2.657221  -1.593 0.113340
factor(dataset$Continent)4   10.732144   5.680222   1.889 0.060770 .
factor(dataset$Continent)5  -10.592899   4.535764  -2.335 0.020847 *
factor(dataset$Financial_Freedom)2  2.065099   1.618526   1.276 0.203958
factor(dataset$Financial_Freedom)3  1.570213   1.688393   0.930 0.353863
factor(dataset$Financial_Freedom)4  3.500643   3.230394   1.084 0.280255
factor(dataset$Investment_Freedom)2  2.260452   0.646697   3.495 0.000623 ***
factor(dataset$Investment_Freedom)3  4.686984   0.690374   6.789 2.46e-10 ***
factor(dataset$Investment_Freedom)4  5.999718   0.811381   7.394 9.31e-12 ***
factor(dataset$Financial_Freedom)1: x3 -0.035409   0.048835  -0.725 0.469529
factor(dataset$Financial_Freedom)2: x3 -0.017880   0.023274  -0.768 0.443558
factor(dataset$Financial_Freedom)3: x3  0.036688   0.017380   2.111 0.036435 *
factor(dataset$Financial_Freedom)4: x3  0.037795   0.036115   1.047 0.297000
x9: factor(dataset$Continent)1    0.105020   0.055229   1.902 0.059149 .
x9: factor(dataset$Continent)2   -0.039334   0.060078  -0.655 0.513645
x9: factor(dataset$Continent)3    0.056010   0.035534   1.576 0.117079
x9: factor(dataset$Continent)4   -0.171259   0.072766  -2.354 0.019891 *
```



```
x9: factor(dataset$Continent)5      0.100301  0.057564  1.742 0.083484 .
```

```
---
```

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

```
Residual standard error: 1.57 on 150 degrees of freedom
```

```
Multiple R-squared:  0.9812,    Adjusted R-squared:  0.9777
```

```
F-statistic: 279.2 on 28 and 150 DF,  p-value: < 2.2e-16
```

```
> AIC(NewBackAIC)
```

```
[1] 697.9275
```

```
> BIC(NewBackAIC)
```

```
[1] 793.549
```

BIC

```
> #-----BIC-----
```

```
> NewBackBIC = step(NewFM, direction = "backward", k=log(179))
```

```
Start: AIC=280.38
```

```
y ~ x1^2 + sqrt(x2) + x4 + x6 + x7 + x8 + x9 + x10 + factor(dataset$Continent) +  
  factor(dataset$Financial_Freedom) + factor(dataset$Investment_Freedom) +  
  x3: factor(dataset$Financial_Freedom) + x9: factor(dataset$Continent)
```

	Df	Sum of Sq	RSS	AIC
- factor(dataset\$Financial_Freedom): x3	4	23.68	393.62	270.74
- x9: factor(dataset\$Continent)	5	47.76	417.70	276.18
<none>			369.94	280.38
- x10	1	62.19	432.13	303.01
- x7	1	68.54	438.49	305.62
- x1	1	127.88	497.82	328.34
- sqrt(x2)	1	171.31	541.25	343.31
- factor(dataset\$Investment_Freedom)	3	203.68	573.62	343.33
- x8	1	173.94	543.89	344.18
- x4	1	212.23	582.17	356.36
- x6	1	1096.63	1466.58	521.74

```
Step: AIC=270.74
```

```
y ~ x1 + sqrt(x2) + x4 + x6 + x7 + x8 + x9 + x10 + factor(dataset$Continent) +  
  factor(dataset$Financial_Freedom) + factor(dataset$Investment_Freedom) +  
  x9: factor(dataset$Continent)
```

	Df	Sum of Sq	RSS	AIC
- x9: factor(dataset\$Continent)	5	46.40	440.02	264.75
<none>			393.62	270.74
- x10	1	54.39	448.01	288.72
- x7	1	74.76	468.38	296.68
- x1	1	155.19	548.81	325.04
- factor(dataset\$Financial_Freedom)	3	192.40	586.02	326.41
- factor(dataset\$Investment_Freedom)	3	201.36	594.98	329.13
- x8	1	167.92	561.54	329.15
- x4	1	201.61	595.23	339.58
- sqrt(x2)	1	244.89	638.51	352.14
- x6	1	1142.69	1536.31	509.30

```
Step: AIC=264.75
```

```
y ~ x1 + sqrt(x2) + x4 + x6 + x7 + x8 + x9 + x10 + factor(dataset$Continent) +  
  factor(dataset$Financial_Freedom) + factor(dataset$Investment_Freedom)
```

	Df	Sum of Sq	RSS	AIC
<none>			440.02	264.75
- x10	1	54.08	494.10	280.31
- x9	1	56.34	496.36	281.12
- x7	1	73.45	513.48	287.19
- factor(dataset\$Continent)	5	138.27	578.29	287.72
- factor(dataset\$Financial_Freedom)	3	180.75	620.78	310.79
- x1	1	182.73	622.75	321.73

```

- x8                1  188.04 628.07 323.25
- factor(dataset$Investment_Freedom) 3  230.93 670.95 324.70
- x4                1  206.01 646.03 328.30
- sqrt(x2)          1  237.02 677.04 336.69
- x6                1 1163.11 1603.13 490.99
> summary(NewBackBIC)

```

Call:

```

lm(formula = y ~ x1 + sqrt(x2) + x4 + x6 + x7 + x8 + x9 + x10 +
  factor(dataset$Continent) + factor(dataset$Financial_Freedom) +
  factor(dataset$Investment_Freedom), data = dataset)

```

Residuals:

Min	1Q	Median	3Q	Max
-5.9833	-0.9779	0.1600	1.0142	4.3746

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.127417	1.813711	-0.070	0.9441
x1	0.130524	0.016063	8.126	1.17e-13 ***
sqrt(x2)	1.644876	0.177737	9.255	< 2e-16 ***
x4	0.112249	0.013010	8.628	6.09e-15 ***
x6	0.097336	0.004748	20.501	< 2e-16 ***
x7	0.076487	0.014846	5.152	7.55e-07 ***
x8	0.093523	0.011346	8.243	5.90e-14 ***
x9	0.077083	0.017085	4.512	1.24e-05 ***
x10	0.079348	0.017950	4.420	1.82e-05 ***
factor(dataset\$Continent)1	-0.903756	0.467443	-1.933	0.0550 .
factor(dataset\$Continent)2	-1.333422	0.596941	-2.234	0.0269 *
factor(dataset\$Continent)3	-0.267762	0.422037	-0.634	0.5267
factor(dataset\$Continent)4	-2.935489	0.620930	-4.728	4.97e-06 ***
factor(dataset\$Continent)5	-2.753005	0.490751	-5.610	8.79e-08 ***
factor(dataset\$Financial_Freedom)2	2.177123	0.515267	4.225	4.01e-05 ***
factor(dataset\$Financial_Freedom)3	3.921454	0.613551	6.391	1.74e-09 ***
factor(dataset\$Financial_Freedom)4	6.668367	0.847143	7.872	5.10e-13 ***
factor(dataset\$Investment_Freedom)2	2.568855	0.625514	4.107	6.40e-05 ***
factor(dataset\$Investment_Freedom)3	4.949718	0.692775	7.145	3.07e-11 ***
factor(dataset\$Investment_Freedom)4	6.541614	0.838572	7.801	7.66e-13 ***

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

Residual standard error: 1.664 on 159 degrees of freedom
 Multiple R-squared: 0.9776, Adjusted R-squared: 0.9749
 F-statistic: 365.3 on 19 and 159 DF, p-value: < 2.2e-16

```

> BIC(NewBackBIC)
[1] 777.9145
> AIC(NewBackBIC)
[1] 710.9794

```

חלק 6 - גרסיה לפני
 AIC

```

> #-----Forward-----
> #-----AIC-----
> NewforwardAIC <- stepAIC(y~1, data=dataset, direction = "forward",scope = ~x1^2 + sqrt(x2) + x4 + x6 + x7 + x8
+ x9 + x10 + factor(dataset$Continent) +
+ factor(dataset$Financial_Freedom) + factor(dataset$Investment_Freedom) +
+ x3: factor(dataset$Financial_Freedom) + x9: factor(dataset$Continent))
Start: AIC=843.03
y ~ 1

```

Df	Sum of Sq	RSS	AIC
----	-----------	-----	-----

```

+ x1                1  14615.6 5034.8 601.28
+ sqrt(x2)          1  12950.1 6700.3 652.43
+ x7                1  12395.2 7255.2 666.67
+ factor(dataset$Financial_Freedom) 3  11862.1 7788.3 683.37
+ factor(dataset$Investment_Freedom) 3  10818.2 8832.2 705.88
+ x10               1  9845.3 9805.0 720.58
+ x9                1  6444.1 13206.3 773.89
+ x6                1  5923.6 13726.8 780.81
+ x8                1  5593.1 14057.3 785.07
+ factor(dataset$Continent) 5  5374.1 14276.3 795.84
<none>              19650.4 843.03
+ x4                1  14.6 19635.8 844.89

```

Step: AIC=601.28

y ~ x1

```

              Df Sum of Sq  RSS  AIC
+ x6          1  1700.82 3334.0 529.49
+ factor(dataset$Financial_Freedom) 3  1055.62 3979.2 565.16
+ factor(dataset$Investment_Freedom) 3  1015.20 4019.6 566.97
+ sqrt(x2)     1   737.07 4297.7 574.94
+ x7           1   728.51 4306.3 575.30
+ x8           1   690.66 4344.1 576.87
+ x9           1   622.30 4412.5 579.66
+ x4           1   496.47 4538.3 584.69
+ x10          1   454.39 4580.4 586.35
<none>        5034.8 601.28
+ factor(dataset$Continent) 5  42.94 4991.8 609.74

```

Step: AIC=529.49

y ~ x1 + x6

```

              Df Sum of Sq  RSS  AIC
+ factor(dataset$Investment_Freedom) 3  910.21 2423.7 478.42
+ factor(dataset$Financial_Freedom) 3  863.40 2470.6 481.84
+ x8           1   693.43 2640.5 489.75
+ sqrt(x2)     1   580.51 2753.4 497.25
+ x7           1   557.33 2776.6 498.75
+ x4           1   486.35 2847.6 503.27
+ x9           1   375.45 2958.5 510.10
+ x10          1   321.72 3012.2 513.33
+ factor(dataset$Continent) 5  229.18 3104.8 526.74
<none>        3334.0 529.49

```

Step: AIC=478.42

y ~ x1 + x6 + factor(dataset\$Investment_Freedom)

```

              Df Sum of Sq  RSS  AIC
+ x8           1   706.58 1717.2 418.73
+ x4           1   605.28 1818.5 428.99
+ sqrt(x2)     1   510.95 1912.8 438.04
+ x7           1   488.78 1935.0 440.10
+ factor(dataset$Continent) 5  407.66 2016.1 455.45
+ factor(dataset$Financial_Freedom) 3  335.13 2088.6 457.78
+ x10          1   147.76 2276.0 469.16
+ x9           1   100.35 2323.4 472.85
<none>        2423.7 478.42

```

Step: AIC=418.73

y ~ x1 + x6 + factor(dataset\$Investment_Freedom) + x8

```

              Df Sum of Sq  RSS  AIC
+ sqrt(x2)     1   349.80 1367.4 379.95
+ x4           1   306.18 1411.0 385.57
+ factor(dataset$Financial_Freedom) 3  317.99 1399.2 388.07

```

	Df	Sum of Sq	RSS	AIC
+ x7	1	259.09	1458.1	391.45
+ factor(dataset\$Continent)	5	305.10	1412.1	393.71
+ x10	1	123.03	1594.1	407.42
+ x9	1	63.35	1653.8	414.00
<none>			1717.2	418.73

Step: AIC=379.95

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2)$

	Df	Sum of Sq	RSS	AIC
+ x4	1	362.01	1005.4	326.90
+ factor(dataset\$Financial_Freedom)	3	287.59	1079.8	343.69
+ factor(dataset\$Continent)	5	222.93	1144.4	358.10
+ x7	1	168.29	1199.1	358.44
+ x10	1	144.89	1222.5	361.90
+ x9	1	60.09	1307.3	373.91
<none>			1367.4	379.95

Step: AIC=326.9

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2) + x4$

	Df	Sum of Sq	RSS	AIC
+ factor(dataset\$Financial_Freedom)	3	266.587	738.77	277.75
+ x7	1	97.048	908.31	310.73
+ x9	1	77.453	927.91	314.55
+ x10	1	76.045	929.31	314.82
+ factor(dataset\$Continent)	5	109.225	896.13	316.32
<none>			1005.36	326.90

Step: AIC=277.75

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2) + x4 + \text{factor}(\text{dataset}\$Financial_Freedom)$

	Df	Sum of Sq	RSS	AIC
+ factor(dataset\$Continent)	5	101.810	636.96	261.21
+ x7	1	72.410	666.36	261.29
+ x10	1	50.122	688.65	267.18
+ x9	1	48.279	690.49	267.65
<none>			738.77	277.75
+ factor(dataset\$Financial_Freedom): x3	4	17.905	720.87	281.36

Step: AIC=261.21

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2) + x4 + \text{factor}(\text{dataset}\$Financial_Freedom) + \text{factor}(\text{dataset}\$Continent)$

	Df	Sum of Sq	RSS	AIC
+ x7	1	83.131	553.83	238.18
+ x10	1	70.243	566.72	242.29
+ x9	1	56.643	580.32	246.54
<none>			636.96	261.21
+ factor(dataset\$Financial_Freedom): x3	4	21.507	615.45	263.06

Step: AIC=238.18

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2) + x4 + \text{factor}(\text{dataset}\$Financial_Freedom) + \text{factor}(\text{dataset}\$Continent) + x7$

	Df	Sum of Sq	RSS	AIC
+ x9	1	59.732	494.10	219.75
+ x10	1	57.473	496.36	220.56
<none>			553.83	238.18
+ factor(dataset\$Financial_Freedom): x3	4	16.035	537.79	240.92

Step: AIC=219.75

```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom) + factor(dataset$Continent) +
  x7 + x9
```

	Df	Sum of Sq	RSS	AIC
+ x10	1	54.077	440.02	201.00
+ x9: factor(dataset\$Continent)	5	46.088	448.01	212.22
<none>		494.10	219.75	
+ factor(dataset\$Financial_Freedom): x3	4	17.591	476.51	221.26

Step: AIC=201

```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom) + factor(dataset$Continent) +
  x7 + x9 + x10
```

	Df	Sum of Sq	RSS	AIC
+ x9: factor(dataset\$Continent)	5	46.401	393.62	191.05
+ factor(dataset\$Financial_Freedom): x3	4	22.321	417.70	199.68
<none>		440.02	201.00	

Step: AIC=191.05

```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom) + factor(dataset$Continent) +
  x7 + x9 + x10 + factor(dataset$Continent): x9
```

	Df	Sum of Sq	RSS	AIC
+ factor(dataset\$Financial_Freedom): x3	4	23.677	369.94	187.95
<none>		393.62	191.05	

Step: AIC=187.95

```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom) + factor(dataset$Continent) +
  x7 + x9 + x10 + factor(dataset$Continent): x9 + factor(dataset$Financial_Freedom): x3
```

> summary(NewforwardAIC)

Call:

```
lm(formula = y ~ x1 + x6 + factor(dataset$Investment_Freedom) +
  x8 + sqrt(x2) + x4 + factor(dataset$Financial_Freedom) +
  factor(dataset$Continent) + x7 + x9 + x10 + factor(dataset$Continent): x9 +
  factor(dataset$Financial_Freedom): x3, data = dataset)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.3093	-0.8114	0.0792	0.8807	3.1644

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.296594	3.026366	0.759	0.449126
x1	0.117645	0.016338	7.201	2.69e-11 ***
x6	0.097407	0.004619	21.087	< 2e-16 ***
factor(dataset\$Investment_Freedom)2	2.260452	0.646697	3.495	0.000623 ***
factor(dataset\$Investment_Freedom)3	4.686984	0.690374	6.789	2.46e-10 ***
factor(dataset\$Investment_Freedom)4	5.999718	0.811381	7.394	9.31e-12 ***
x8	0.092739	0.011043	8.398	3.18e-14 ***
sqrt(x2)	1.656639	0.198776	8.334	4.60e-14 ***
x4	0.119184	0.012848	9.276	< 2e-16 ***
factor(dataset\$Financial_Freedom)2	2.065099	1.618526	1.276	0.203958
factor(dataset\$Financial_Freedom)3	1.570213	1.688393	0.930	0.353863
factor(dataset\$Financial_Freedom)4	3.500643	3.230394	1.084	0.280255
factor(dataset\$Continent)1	-8.680377	4.109472	-2.112	0.036319 *
factor(dataset\$Continent)2	1.419713	4.476486	0.317	0.751572
factor(dataset\$Continent)3	-4.232064	2.657221	-1.593	0.113340
factor(dataset\$Continent)4	10.732144	5.680222	1.889	0.060770 .
factor(dataset\$Continent)5	-10.592899	4.535764	-2.335	0.020847 *

```

x7                0.077048  0.014615  5.272 4.63e-07 ***
x9                0.048258  0.030203  1.598 0.112194
x10              0.088957  0.017716  5.021 1.44e-06 ***
factor(dataset$Continent)1: x9    0.105020  0.055229  1.902 0.059149 .
factor(dataset$Continent)2: x9   -0.039334  0.060078 -0.655 0.513645
factor(dataset$Continent)3: x9    0.056010  0.035534  1.576 0.117079
factor(dataset$Continent)4: x9   -0.171259  0.072766 -2.354 0.019891 *
factor(dataset$Continent)5: x9    0.100301  0.057564  1.742 0.083484 .
factor(dataset$Financial_Freedom)1: x3 -0.035409  0.048835 -0.725 0.469529
factor(dataset$Financial_Freedom)2: x3 -0.017880  0.023274 -0.768 0.443558
factor(dataset$Financial_Freedom)3: x3  0.036688  0.017380  2.111 0.036435 *
factor(dataset$Financial_Freedom)4: x3  0.037795  0.036115  1.047 0.297000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 1.57 on 150 degrees of freedom

Multiple R-squared: 0.9812, Adjusted R-squared: 0.9777

F-statistic: 279.2 on 28 and 150 DF, p-value: < 2.2e-16

> AIC(NewforwardAIC)

[1] 697.9275

> BIC(NewforwardAIC)

[1] 793.549

BIC

> #-----BIC-----

```

> NewforwardBIC = step(lm(y~1, data=dataset), direction = "forward", scope = ~ x1^2 + sqrt(x2) + x4 + x6 + x7 + x8
+ x9 + x10 + factor(dataset$Continent) +
+ factor(dataset$Financial_Freedom) + factor(dataset$Investment_Freedom) +
+ x3: factor(dataset$Financial_Freedom) + x9: factor(dataset$Continent), k=log(179))
Start: AIC=846.21
y ~ 1

```

	Df	Sum of Sq	RSS	AIC
+ x1	1	14615.6	5034.8	607.65
+ sqrt(x2)	1	12950.1	6700.3	658.81
+ x7	1	12395.2	7255.2	673.05
+ factor(dataset\$Financial_Freedom)	3	11862.1	7788.3	696.11
+ factor(dataset\$Investment_Freedom)	3	10818.2	8832.2	718.63
+ x10	1	9845.3	9805.0	726.96
+ x9	1	6444.1	13206.3	780.27
+ x6	1	5923.6	13726.8	787.18
+ x8	1	5593.1	14057.3	791.44
+ factor(dataset\$Continent)	5	5374.1	14276.3	814.96
<none>			19650.4	846.21
+ x4	1	14.6	19635.8	851.27

Step: AIC=607.65

y ~ x1

	Df	Sum of Sq	RSS	AIC
+ x6	1	1700.82	3334.0	539.05
+ factor(dataset\$Financial_Freedom)	3	1055.62	3979.2	581.09
+ factor(dataset\$Investment_Freedom)	3	1015.20	4019.6	582.90
+ sqrt(x2)	1	737.07	4297.7	584.50
+ x7	1	728.51	4306.3	584.86
+ x8	1	690.66	4344.1	586.43
+ x9	1	622.30	4412.5	589.22
+ x4	1	496.47	4538.3	594.26
+ x10	1	454.39	4580.4	595.91
<none>			5034.8	607.65
+ factor(dataset\$Continent)	5	42.94	4991.8	632.05

Step: AIC=539.05

y ~ x1 + x6

	Df	Sum of Sq	RSS	AIC
+ factor(dataset\$Investment_Freedom)	3	910.21	2423.7	497.54
+ factor(dataset\$Financial_Freedom)	3	863.40	2470.6	500.97
+ x8	1	693.43	2640.5	502.50
+ sqrt(x2)	1	580.51	2753.4	510.00
+ x7	1	557.33	2776.6	511.50
+ x4	1	486.35	2847.6	516.02
+ x9	1	375.45	2958.5	522.85
+ x10	1	321.72	3012.2	526.08
<none>		3334.0	539.05	
+ factor(dataset\$Continent)	5	229.18	3104.8	552.24

Step: AIC=497.54

y ~ x1 + x6 + factor(dataset\$Investment_Freedom)

	Df	Sum of Sq	RSS	AIC
+ x8	1	706.58	1717.2	441.04
+ x4	1	605.28	1818.5	451.30
+ sqrt(x2)	1	510.95	1912.8	460.35
+ x7	1	488.78	1935.0	462.41
+ factor(dataset\$Financial_Freedom)	3	335.13	2088.6	486.47
+ factor(dataset\$Continent)	5	407.66	2016.1	490.51
+ x10	1	147.76	2276.0	491.47
+ x9	1	100.35	2323.4	495.16
<none>		2423.7	497.54	

Step: AIC=441.04

y ~ x1 + x6 + factor(dataset\$Investment_Freedom) + x8

	Df	Sum of Sq	RSS	AIC
+ sqrt(x2)	1	349.80	1367.4	405.45
+ x4	1	306.18	1411.0	411.07
+ x7	1	259.09	1458.1	416.95
+ factor(dataset\$Financial_Freedom)	3	317.99	1399.2	419.94
+ factor(dataset\$Continent)	5	305.10	1412.1	431.96
+ x10	1	123.03	1594.1	432.92
+ x9	1	63.35	1653.8	439.50
<none>		1717.2	441.04	

Step: AIC=405.45

y ~ x1 + x6 + factor(dataset\$Investment_Freedom) + x8 + sqrt(x2)

	Df	Sum of Sq	RSS	AIC
+ x4	1	362.01	1005.4	355.59
+ factor(dataset\$Financial_Freedom)	3	287.59	1079.8	378.75
+ x7	1	168.29	1199.1	387.13
+ x10	1	144.89	1222.5	390.59
+ factor(dataset\$Continent)	5	222.93	1144.4	399.53
+ x9	1	60.09	1307.3	402.59
<none>		1367.4	405.45	

Step: AIC=355.59

y ~ x1 + x6 + factor(dataset\$Investment_Freedom) + x8 + sqrt(x2) + x4

	Df	Sum of Sq	RSS	AIC
+ factor(dataset\$Financial_Freedom)	3	266.587	738.77	316.00
+ x7	1	97.048	908.31	342.61
+ x9	1	77.453	927.91	346.43
+ x10	1	76.045	929.31	346.70
<none>		1005.36	355.59	
+ factor(dataset\$Continent)	5	109.225	896.13	360.94

Step: AIC=316

```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom)
```

	Df	Sum of Sq	RSS	AIC
+ x7	1	72.410	666.36	302.72
+ x10	1	50.122	688.65	308.61
+ x9	1	48.279	690.49	309.09
+ factor(dataset\$Continent)	5	101.810	636.96	315.39
<none>		738.77	316.00	
+ factor(dataset\$Financial_Freedom): x3 4		17.905	720.87	332.36

Step: AIC=302.72

```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom) + x7
```

	Df	Sum of Sq	RSS	AIC
+ x9	1	52.999	613.36	293.07
+ factor(dataset\$Continent)	5	112.531	553.83	295.55
+ x10	1	36.268	630.09	297.89
<none>		666.36	302.72	
+ factor(dataset\$Financial_Freedom): x3 4		13.860	652.50	319.71

Step: AIC=293.07

```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom) + x7 + x9
```

	Df	Sum of Sq	RSS	AIC
+ factor(dataset\$Continent)	5	119.263	494.10	280.31
+ x10	1	35.070	578.29	287.72
<none>		613.36	293.07	
+ factor(dataset\$Financial_Freedom): x3 4		15.559	597.80	309.23

Step: AIC=280.31

```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom) + x7 + x9 + factor(dataset$Continent)
```

	Df	Sum of Sq	RSS	AIC
+ x10	1	54.077	440.02	264.75
<none>		494.10	280.31	
+ x9: factor(dataset\$Continent)	5	46.088	448.01	288.72
+ factor(dataset\$Financial_Freedom): x3 4		17.591	476.51	294.57

Step: AIC=264.75

```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom) + x7 + x9 + factor(dataset$Continent) +
  x10
```

	Df	Sum of Sq	RSS	AIC
<none>		440.02	264.75	
+ x9: factor(dataset\$Continent)	5	46.401	393.62	270.74
+ factor(dataset\$Financial_Freedom): x3 4		22.321	417.70	276.18

> summary(NewforwardBIC)

Call:

```
lm(formula = y ~ x1 + x6 + factor(dataset$Investment_Freedom) +
  x8 + sqrt(x2) + x4 + factor(dataset$Financial_Freedom) +
  x7 + x9 + factor(dataset$Continent) + x10, data = dataset)
```

Residuals:

```
Min      1Q  Median      3Q      Max
-5.9833 -0.9779  0.1600  1.0142  4.3746
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept)      -0.127417   1.813711  -0.070   0.9441
```



```

x1                0.130524  0.016063  8.126 1.17e-13 ***
x6                0.097336  0.004748 20.501 < 2e-16 ***
factor(dataset$Investment_Freedom)2 2.568855 0.625514 4.107 6.40e-05 ***
factor(dataset$Investment_Freedom)3 4.949718 0.692775 7.145 3.07e-11 ***
factor(dataset$Investment_Freedom)4 6.541614 0.838572 7.801 7.66e-13 ***
x8                0.093523  0.011346  8.243 5.90e-14 ***
sqrt(x2)          1.644876  0.177737  9.255 < 2e-16 ***
x4                0.112249  0.013010  8.628 6.09e-15 ***
factor(dataset$Financial_Freedom)2 2.177123 0.515267 4.225 4.01e-05 ***
factor(dataset$Financial_Freedom)3 3.921454 0.613551 6.391 1.74e-09 ***
factor(dataset$Financial_Freedom)4 6.668367 0.847143 7.872 5.10e-13 ***
x7                0.076487  0.014846  5.152 7.55e-07 ***
x9                0.077083  0.017085  4.512 1.24e-05 ***
factor(dataset$Continent)1          -0.903756  0.467443 -1.933 0.0550 .
factor(dataset$Continent)2          -1.333422  0.596941 -2.234 0.0269 *
factor(dataset$Continent)3          -0.267762  0.422037 -0.634 0.5267
factor(dataset$Continent)4          -2.935489  0.620930 -4.728 4.97e-06 ***
factor(dataset$Continent)5          -2.753005  0.490751 -5.610 8.79e-08 ***
x10               0.079348  0.017950  4.420 1.82e-05 ***

```

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

Residual standard error: 1.664 on 159 degrees of freedom

Multiple R-squared: 0.9776, Adjusted R-squared: 0.9749

F-statistic: 365.3 on 19 and 159 DF, p-value: < 2.2e-16

> BIC(NewforwardBIC)

[1] 777.9145

> AIC(NewforwardBIC)

[1] 710.9794

חלק 7 - רגרסיה בצעדים AIC

```

> #-----stepwise-----
> #-----AIC-----
> NewstepwiseAIC = step(lm(y~1, data=dataset), direction = "both",scope = ~x1^2 + sqrt(x2) + x4 + x6 + x7 + x8 + x
9 + x10 + factor(dataset$Continent) +
+ factor(dataset$Financial_Freedom) + factor(dataset$Investment_Freedom) +
+ x3: factor(dataset$Financial_Freedom) + x9: factor(dataset$Continent))
Start: AIC=843.03
y ~ 1

```

	Df	Sum of Sq	RSS	AIC
+ x1	1	14615.6	5034.8	601.28
+ sqrt(x2)	1	12950.1	6700.3	652.43
+ x7	1	12395.2	7255.2	666.67
+ factor(dataset\$Financial_Freedom)	3	11862.1	7788.3	683.37
+ factor(dataset\$Investment_Freedom)	3	10818.2	8832.2	705.88
+ x10	1	9845.3	9805.0	720.58
+ x9	1	6444.1	13206.3	773.89
+ x6	1	5923.6	13726.8	780.81
+ x8	1	5593.1	14057.3	785.07
+ factor(dataset\$Continent)	5	5374.1	14276.3	795.84
<none>			19650.4	843.03
+ x4	1	14.6	19635.8	844.89

Step: AIC=601.28

y ~ x1

	Df	Sum of Sq	RSS	AIC
+ x6	1	1700.8	3334.0	529.49

```

+ factor(dataset$Financial_Freedom) 3 1055.6 3979.2 565.16
+ factor(dataset$Investment_Freedom) 3 1015.2 4019.6 566.97
+ sqrt(x2) 1 737.1 4297.7 574.94
+ x7 1 728.5 4306.3 575.30
+ x8 1 690.7 4344.1 576.87
+ x9 1 622.3 4412.5 579.66
+ x4 1 496.5 4538.3 584.69
+ x10 1 454.4 4580.4 586.35
<none> 5034.8 601.28
+ factor(dataset$Continent) 5 42.9 4991.8 609.74
- x1 1 14615.6 19650.4 843.03

```

Step: AIC=529.49

y ~ x1 + x6

```

      Df Sum of Sq  RSS   AIC
+ factor(dataset$Investment_Freedom) 3  910.2 2423.7 478.42
+ factor(dataset$Financial_Freedom) 3  863.4 2470.6 481.84
+ x8 1  693.4 2640.5 489.75
+ sqrt(x2) 1  580.5 2753.4 497.25
+ x7 1  557.3 2776.6 498.75
+ x4 1  486.3 2847.6 503.27
+ x9 1  375.5 2958.5 510.10
+ x10 1  321.7 3012.2 513.33
+ factor(dataset$Continent) 5  229.2 3104.8 526.74
<none> 3334.0 529.49
- x6 1 1700.8 5034.8 601.28
- x1 1 10392.8 13726.8 780.81

```

Step: AIC=478.42

y ~ x1 + x6 + factor(dataset\$Investment_Freedom)

```

      Df Sum of Sq  RSS   AIC
+ x8 1  706.6 1717.2 418.73
+ x4 1  605.3 1818.5 428.99
+ sqrt(x2) 1  510.9 1912.8 438.04
+ x7 1  488.8 1935.0 440.10
+ factor(dataset$Continent) 5  407.7 2016.1 455.45
+ factor(dataset$Financial_Freedom) 3  335.1 2088.6 457.78
+ x10 1  147.8 2276.0 469.16
+ x9 1  100.4 2323.4 472.85
<none> 2423.7 478.42
- factor(dataset$Investment_Freedom) 3  910.2 3334.0 529.49
- x6 1 1595.8 4019.6 566.97
- x1 1  3802.5 6226.2 645.30

```

Step: AIC=418.73

y ~ x1 + x6 + factor(dataset\$Investment_Freedom) + x8

```

      Df Sum of Sq  RSS   AIC
+ sqrt(x2) 1  349.80 1367.4 379.95
+ x4 1  306.18 1411.0 385.57
+ factor(dataset$Financial_Freedom) 3  317.99 1399.2 388.07
+ x7 1  259.09 1458.1 391.45
+ factor(dataset$Continent) 5  305.10 1412.1 393.71
+ x10 1  123.03 1594.1 407.42
+ x9 1  63.35 1653.8 414.00
<none> 1717.2 418.73
- x8 1  706.58 2423.7 478.42
- factor(dataset$Investment_Freedom) 3  923.37 2640.5 489.75
- x6 1 1585.66 3302.8 533.81
- x1 1  2477.15 4194.3 576.58

```

Step: AIC=379.95

y ~ x1 + x6 + factor(dataset\$Investment_Freedom) + x8 + sqrt(x2)

	Df	Sum of Sq	RSS	AIC
+ x4	1	362.01	1005.4	326.90
+ factor(dataset\$Financial_Freedom)	3	287.59	1079.8	343.69
+ factor(dataset\$Continent)	5	222.93	1144.4	358.10
+ x7	1	168.29	1199.1	358.44
+ x10	1	144.89	1222.5	361.90
+ x9	1	60.09	1307.3	373.91
<none>		1367.4	379.95	
- sqrt(x2)	1	349.80	1717.2	418.73
- x8	1	545.43	1912.8	438.04
- x1	1	607.06	1974.4	443.72
- factor(dataset\$Investment_Freedom)	3	877.33	2244.7	462.68
- x6	1	1445.51	2812.9	507.07

Step: AIC=326.9

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2) + x4$

	Df	Sum of Sq	RSS	AIC
+ factor(dataset\$Financial_Freedom)	3	266.59	738.77	277.75
+ x7	1	97.05	908.31	310.73
+ x9	1	77.45	927.91	314.55
+ x10	1	76.04	929.31	314.82
+ factor(dataset\$Continent)	5	109.22	896.13	316.32
<none>		1005.36	326.90	
- x8	1	270.12	1275.48	367.50
- x4	1	362.01	1367.37	379.95
- sqrt(x2)	1	405.62	1410.98	385.57
- x1	1	671.44	1676.80	416.47
- factor(dataset\$Investment_Freedom)	3	945.29	1950.65	439.55
- x6	1	1397.60	2402.96	480.88

Step: AIC=277.75

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2) + x4 + \text{factor}(\text{dataset}\$Financial_Freedom)$

	Df	Sum of Sq	RSS	AIC
+ factor(dataset\$Continent)	5	101.81	636.96	261.21
+ x7	1	72.41	666.36	261.29
+ x10	1	50.12	688.65	267.17
+ x9	1	48.28	690.49	267.65
<none>		738.77	277.75	
+ factor(dataset\$Financial_Freedom): x3	4	17.91	720.87	281.36
- factor(dataset\$Financial_Freedom)	3	266.59	1005.36	326.90
- x8	1	257.22	995.99	329.23
- x4	1	341.01	1079.78	343.69
- sqrt(x2)	1	363.46	1102.23	347.37
- factor(dataset\$Investment_Freedom)	3	417.75	1156.52	351.98
- x1	1	396.93	1135.70	352.72
- x6	1	1269.03	2007.80	454.72

Step: AIC=261.21

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2) + x4 + \text{factor}(\text{dataset}\$Financial_Freedom) + \text{factor}(\text{dataset}\$Continent)$

	Df	Sum of Sq	RSS	AIC
+ x7	1	83.13	553.83	238.18
+ x10	1	70.24	566.72	242.29
+ x9	1	56.64	580.32	246.54
<none>		636.96	261.21	
+ factor(dataset\$Financial_Freedom): x3	4	21.51	615.45	263.06
- factor(dataset\$Continent)	5	101.81	738.77	277.75
- factor(dataset\$Financial_Freedom)	3	259.17	896.13	316.32
- x8	1	254.80	891.76	319.44

```

- x4          1  256.60 893.56 319.80
- sqrt(x2)    1  302.97 939.93 328.86
- factor(dataset$Investment_Freedom) 3  360.13 997.09 335.42
- x1          1  428.27 1065.23 351.26
- x6          1 1304.38 1941.34 458.69

```

Step: AIC=238.18

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2) + x4 + \text{factor}(\text{dataset}\$Financial_Freedom) + \text{factor}(\text{dataset}\$Continent) + x7$

```

          Df Sum of Sq  RSS  AIC
+ x9          1   59.73 494.10 219.75
+ x10         1   57.47 496.36 220.56
<none>                553.83 238.18
+ factor(dataset$Financial_Freedom): x3 4   16.04 537.79 240.92
- x7          1   83.13 636.96 261.21
- factor(dataset$Continent)      5  112.53 666.36 261.29
- x8          1  207.49 761.32 293.13
- factor(dataset$Financial_Freedom) 3  235.60 789.43 295.62
- x4          1  220.26 774.09 296.11
- sqrt(x2)    1  224.11 777.94 297.00
- x1          1  274.38 828.21 308.21
- factor(dataset$Investment_Freedom) 3  348.54 902.37 319.56
- x6          1 1290.94 1844.77 451.56

```

Step: AIC=219.75

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2) + x4 + \text{factor}(\text{dataset}\$Financial_Freedom) + \text{factor}(\text{dataset}\$Continent) + x7 + x9$

```

          Df Sum of Sq  RSS  AIC
+ x10          1   54.08 440.02 201.00
+ x9: factor(dataset$Continent)      5   46.09 448.01 212.22
<none>                494.10 219.75
+ factor(dataset$Financial_Freedom): x3 4   17.59 476.51 221.26
- x9          1   59.73 553.83 238.18
- x7          1   86.22 580.32 246.54
- factor(dataset$Continent)      5  119.26 613.36 248.45
- factor(dataset$Financial_Freedom) 3  207.46 701.56 276.50
- x8          1  198.69 692.79 278.25
- x4          1  227.84 721.94 285.63
- sqrt(x2)    1  228.07 722.17 285.68
- x1          1  233.44 727.54 287.01
- factor(dataset$Investment_Freedom) 3  268.20 762.30 291.36
- x6          1 1155.61 1649.71 433.55

```

Step: AIC=201

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2) + x4 + \text{factor}(\text{dataset}\$Financial_Freedom) + \text{factor}(\text{dataset}\$Continent) + x7 + x9 + x10$

```

          Df Sum of Sq  RSS  AIC
+ x9: factor(dataset$Continent)      5   46.40 393.62 191.05
+ factor(dataset$Financial_Freedom): x3 4   22.32 417.70 199.68
<none>                440.02 201.00
- x10          1   54.08 494.10 219.75
- x9          1   56.34 496.36 220.56
- x7          1   73.45 513.48 226.63
- factor(dataset$Continent)      5  138.27 578.29 239.91
- factor(dataset$Financial_Freedom) 3  180.75 620.78 256.60
- x1          1  182.73 622.75 261.17
- x8          1  188.04 628.07 262.69
- x4          1  206.01 646.03 267.74
- factor(dataset$Investment_Freedom) 3  230.93 670.95 270.51

```

```
- sqrt(x2)          1  237.02 677.04 276.13
- x6                1 1163.11 1603.13 430.43
```

Step: AIC=191.05

```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom) + factor(dataset$Continent) +
  x7 + x9 + x10 + factor(dataset$Continent):x9
```

	Df	Sum of Sq	RSS	AIC
+ factor(dataset\$Financial_Freedom):x3	4	23.68	369.94	187.95
<none>		393.62	191.05	
- factor(dataset\$Continent):x9	5	46.40	440.02	201.00
- x10	1	54.39	448.01	212.22
- x7	1	74.76	468.38	220.18
- x1	1	155.19	548.81	248.54
- x8	1	167.92	561.54	252.65
- factor(dataset\$Financial_Freedom)	3	192.40	586.02	256.29
- factor(dataset\$Investment_Freedom)	3	201.36	594.98	259.00
- x4	1	201.61	595.23	263.08
- sqrt(x2)	1	244.89	638.51	275.64
- x6	1	1142.69	1536.31	432.81

Step: AIC=187.95

```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom) + factor(dataset$Continent) +
  x7 + x9 + x10 + factor(dataset$Continent):x9 + factor(dataset$Financial_Freedom):x3
```

	Df	Sum of Sq	RSS	AIC
<none>		369.94	187.95	
- factor(dataset\$Financial_Freedom):x3	4	23.68	393.62	191.05
- factor(dataset\$Continent):x9	5	47.76	417.70	199.68
- x10	1	62.19	432.13	213.76
- x7	1	68.54	438.49	216.37
- x1	1	127.88	497.82	239.09
- sqrt(x2)	1	171.31	541.25	254.06
- x8	1	173.94	543.89	254.93
- factor(dataset\$Investment_Freedom)	3	203.68	573.62	260.46
- x4	1	212.23	582.17	267.11
- x6	1	1096.63	1466.58	432.49

> summary(NewstepwiseAIC)

Call:

```
lm(formula = y ~ x1 + x6 + factor(dataset$Investment_Freedom) +
  x8 + sqrt(x2) + x4 + factor(dataset$Financial_Freedom) +
  factor(dataset$Continent) + x7 + x9 + x10 + factor(dataset$Continent):x9 +
  factor(dataset$Financial_Freedom):x3, data = dataset)
```

Residuals:

```
Min      1Q  Median      3Q      Max
-5.3093 -0.8114  0.0792  0.8807  3.1644
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.296594	3.026366	0.759	0.449126
x1	0.117645	0.016338	7.201	2.69e-11 ***
x6	0.097407	0.004619	21.087	< 2e-16 ***
factor(dataset\$Investment_Freedom)2	2.260452	0.646697	3.495	0.000623 ***
factor(dataset\$Investment_Freedom)3	4.686984	0.690374	6.789	2.46e-10 ***
factor(dataset\$Investment_Freedom)4	5.999718	0.811381	7.394	9.31e-12 ***
x8	0.092739	0.011043	8.398	3.18e-14 ***
sqrt(x2)	1.656639	0.198776	8.334	4.60e-14 ***
x4	0.119184	0.012848	9.276	< 2e-16 ***
factor(dataset\$Financial_Freedom)2	2.065099	1.618526	1.276	0.203958
factor(dataset\$Financial_Freedom)3	1.570213	1.688393	0.930	0.353863
factor(dataset\$Financial_Freedom)4	3.500643	3.230394	1.084	0.280255

```

factor(dataset$Continent)1      -8.680377  4.109472 -2.112 0.036319 *
factor(dataset$Continent)2      1.419713  4.476486  0.317 0.751572
factor(dataset$Continent)3     -4.232064  2.657221 -1.593 0.113340
factor(dataset$Continent)4     10.732144  5.680222  1.889 0.060770 .
factor(dataset$Continent)5    -10.592899  4.535764 -2.335 0.020847 *
x7              0.077048  0.014615  5.272 4.63e-07 ***
x9              0.048258  0.030203  1.598 0.112194
x10            0.088957  0.017716  5.021 1.44e-06 ***
factor(dataset$Continent)1:x9    0.105020  0.055229  1.902 0.059149 .
factor(dataset$Continent)2:x9   -0.039334  0.060078 -0.655 0.513645
factor(dataset$Continent)3:x9    0.056010  0.035534  1.576 0.117079
factor(dataset$Continent)4:x9   -0.171259  0.072766 -2.354 0.019891 *
factor(dataset$Continent)5:x9    0.100301  0.057564  1.742 0.083484 .
factor(dataset$Financial_Freedom)1:x3 -0.035409  0.048835 -0.725 0.469529
factor(dataset$Financial_Freedom)2:x3 -0.017880  0.023274 -0.768 0.443558
factor(dataset$Financial_Freedom)3:x3  0.036688  0.017380  2.111 0.036435 *
factor(dataset$Financial_Freedom)4:x3  0.037795  0.036115  1.047 0.297000

```

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

Residual standard error: 1.57 on 150 degrees of freedom

Multiple R-squared: 0.9812, Adjusted R-squared: 0.9777

F-statistic: 279.2 on 28 and 150 DF, p-value: < 2.2e-16

> AIC(NewstepwiseAIC)

[1] 697.9275

> BIC(NewstepwiseAIC)

[1] 793.549

BIC

> #-----BIC-----

> NewstepwiseBIC = stepAIC(y~1, data=dataset, direction = "both",scope = ~x1^2 + sqrt(x2) + x4 + x6 + x7 + x8 + x9 + x10 + factor(dataset\$Continent) +

+ factor(dataset\$Financial_Freedom) + factor(dataset\$Investment_Freedom) +

+ x3: factor(dataset\$Financial_Freedom) + x9: factor(dataset\$Continent), k=log(179))

Start: AIC=846.21

y ~ 1

	Df	Sum of Sq	RSS	AIC
+ x1	1	14615.6	5034.8	607.65
+ sqrt(x2)	1	12950.1	6700.3	658.81
+ x7	1	12395.2	7255.2	673.05
+ factor(dataset\$Financial_Freedom)	3	11862.1	7788.3	696.11
+ factor(dataset\$Investment_Freedom)	3	10818.2	8832.2	718.63
+ x10	1	9845.3	9805.0	726.96
+ x9	1	6444.1	13206.3	780.27
+ x6	1	5923.6	13726.8	787.18
+ x8	1	5593.1	14057.3	791.44
+ factor(dataset\$Continent)	5	5374.1	14276.3	814.96
<none>			19650.4	846.21
+ x4	1	14.6	19635.8	851.27

Step: AIC=607.65

y ~ x1

	Df	Sum of Sq	RSS	AIC
+ x6	1	1700.8	3334.0	539.05
+ factor(dataset\$Financial_Freedom)	3	1055.6	3979.2	581.09
+ factor(dataset\$Investment_Freedom)	3	1015.2	4019.6	582.90
+ sqrt(x2)	1	737.1	4297.7	584.50
+ x7	1	728.5	4306.3	584.86
+ x8	1	690.7	4344.1	586.43
+ x9	1	622.3	4412.5	589.22
+ x4	1	496.5	4538.3	594.26

```

+ x10          1  454.4 4580.4 595.91
<none>          5034.8 607.65
+ factor(dataset$Continent) 5  42.9 4991.8 632.05
- x1           1 14615.6 19650.4 846.21

```

Step: AIC=539.05

y ~ x1 + x6

```

      Df Sum of Sq  RSS  AIC
+ factor(dataset$Investment_Freedom) 3  910.2 2423.7 497.54
+ factor(dataset$Financial_Freedom) 3  863.4 2470.6 500.97
+ x8          1  693.4 2640.5 502.50
+ sqrt(x2)    1  580.5 2753.4 510.00
+ x7          1  557.3 2776.6 511.50
+ x4          1  486.3 2847.6 516.02
+ x9          1  375.5 2958.5 522.85
+ x10         1  321.7 3012.2 526.08
<none>          3334.0 539.05
+ factor(dataset$Continent) 5  229.2 3104.8 552.24
- x6          1 1700.8 5034.8 607.65
- x1          1 10392.8 13726.8 787.18

```

Step: AIC=497.54

y ~ x1 + x6 + factor(dataset\$Investment_Freedom)

```

      Df Sum of Sq  RSS  AIC
+ x8          1  706.6 1717.2 441.04
+ x4          1  605.3 1818.5 451.30
+ sqrt(x2)    1  510.9 1912.8 460.35
+ x7          1  488.8 1935.0 462.41
+ factor(dataset$Financial_Freedom) 3  335.1 2088.6 486.47
+ factor(dataset$Continent) 5  407.7 2016.1 490.51
+ x10         1  147.8 2276.0 491.47
+ x9          1  100.4 2323.4 495.16
<none>          2423.7 497.54
- factor(dataset$Investment_Freedom) 3  910.2 3334.0 539.05
- x6          1 1595.8 4019.6 582.90
- x1          1 3802.5 6226.2 661.23

```

Step: AIC=441.04

y ~ x1 + x6 + factor(dataset\$Investment_Freedom) + x8

```

      Df Sum of Sq  RSS  AIC
+ sqrt(x2)    1  349.80 1367.4 405.45
+ x4          1  306.18 1411.0 411.07
+ x7          1  259.09 1458.1 416.95
+ factor(dataset$Financial_Freedom) 3  317.99 1399.2 419.94
+ factor(dataset$Continent) 5  305.10 1412.1 431.96
+ x10         1  123.03 1594.1 432.92
+ x9          1  63.35 1653.8 439.50
<none>          1717.2 441.04
- x8          1  706.58 2423.7 497.54
- factor(dataset$Investment_Freedom) 3  923.37 2640.5 502.50
- x6          1 1585.66 3302.8 552.94
- x1          1 2477.15 4194.3 595.71

```

Step: AIC=405.45

y ~ x1 + x6 + factor(dataset\$Investment_Freedom) + x8 + sqrt(x2)

```

      Df Sum of Sq  RSS  AIC
+ x4          1  362.01 1005.4 355.59
+ factor(dataset$Financial_Freedom) 3  287.59 1079.8 378.75
+ x7          1  168.29 1199.1 387.13
+ x10         1  144.89 1222.5 390.59
+ factor(dataset$Continent) 5  222.93 1144.4 399.53

```

```

+ x9                1  60.09 1307.3 402.59
<none>              1367.4 405.45
- sqrt(x2)          1  349.80 1717.2 441.04
- x8                1  545.43 1912.8 460.35
- x1                1  607.06 1974.4 466.03
- factor(dataset$Investment_Freedom) 3  877.33 2244.7 478.62
- x6                1  1445.51 2812.9 529.38

```

Step: AIC=355.59

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2) + x4$

```

              Df Sum of Sq  RSS  AIC
+ factor(dataset$Financial_Freedom) 3  266.59 738.77 316.00
+ x7                1  97.05 908.31 342.61
+ x9                1  77.45 927.91 346.43
+ x10               1  76.04 929.31 346.70
<none>              1005.36 355.59
+ factor(dataset$Continent) 5  109.22 896.13 360.94
- x8                1  270.12 1275.48 393.00
- x4                1  362.01 1367.37 405.45
- sqrt(x2)          1  405.62 1410.98 411.07
- x1                1  671.44 1676.80 441.97
- factor(dataset$Investment_Freedom) 3  945.29 1950.65 458.67
- x6                1  1397.60 2402.96 506.37

```

Step: AIC=316

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2) + x4 + \text{factor}(\text{dataset}\$Financial_Freedom)$

```

              Df Sum of Sq  RSS  AIC
+ x7                1  72.41 666.36 302.72
+ x10               1  50.12 688.65 308.61
+ x9                1  48.28 690.49 309.09
+ factor(dataset$Continent) 5  101.81 636.96 315.39
<none>              738.77 316.00
+ factor(dataset$Financial_Freedom):x3 4  17.91 720.87 332.36
- factor(dataset$Financial_Freedom) 3  266.59 1005.36 355.59
- x8                1  257.22 995.99 364.29
- x4                1  341.01 1079.78 378.75
- factor(dataset$Investment_Freedom) 3  417.75 1156.52 380.66
- sqrt(x2)          1  363.46 1102.23 382.43
- x1                1  396.93 1135.70 387.78
- x6                1  1269.03 2007.80 489.78

```

Step: AIC=302.72

$y \sim x1 + x6 + \text{factor}(\text{dataset}\$Investment_Freedom) + x8 + \text{sqrt}(x2) + x4 + \text{factor}(\text{dataset}\$Financial_Freedom) + x7$

```

              Df Sum of Sq  RSS  AIC
+ x9                1  53.00 613.36 293.07
+ factor(dataset$Continent) 5  112.53 553.83 295.55
+ x10               1  36.27 630.09 297.89
<none>              666.36 302.72
- x7                1  72.41 738.77 316.00
+ factor(dataset$Financial_Freedom):x3 4  13.86 652.50 319.71
- factor(dataset$Financial_Freedom) 3  241.95 908.31 342.61
- x8                1  208.84 875.20 346.33
- x1                1  215.75 882.11 347.74
- x4                1  281.50 947.87 360.61
- sqrt(x2)          1  290.89 957.25 362.37
- factor(dataset$Investment_Freedom) 3  419.95 1086.31 374.64
- x6                1  1222.87 1889.23 484.07

```

Step: AIC=293.07


```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom) + x7 + x9
```

	Df	Sum of Sq	RSS	AIC
+ factor(dataset\$Continent)	5	119.26	494.10	280.31
+ x10	1	35.07	578.29	287.72
<none>		613.36	293.07	
- x9	1	53.00	666.36	302.72
- x7	1	77.13	690.49	309.09
+ factor(dataset\$Financial_Freedom): x3	4	15.56	597.80	309.22
- factor(dataset\$Financial_Freedom)	3	214.15	827.51	331.12
- x1	1	180.48	793.85	334.06
- x8	1	197.19	810.56	337.79
- factor(dataset\$Investment_Freedom)	3	337.90	951.26	356.06
- x4	1	285.63	898.99	356.32
- sqrt(x2)	1	289.43	902.79	357.08
- x6	1	1093.13	1706.49	471.05

Step: AIC=280.31

```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom) + x7 + x9 + factor(dataset$Continent)
```

	Df	Sum of Sq	RSS	AIC
+ x10	1	54.08	440.02	264.75
<none>		494.10	280.31	
+ x9: factor(dataset\$Continent)	5	46.09	448.01	288.72
- factor(dataset\$Continent)	5	119.26	613.36	293.07
+ factor(dataset\$Financial_Freedom): x3	4	17.59	476.51	294.57
- x9	1	59.73	553.83	295.55
- x7	1	86.22	580.32	303.91
- factor(dataset\$Financial_Freedom)	3	207.46	701.56	327.50
- x8	1	198.69	692.79	335.62
- factor(dataset\$Investment_Freedom)	3	268.20	762.30	342.36
- x4	1	227.84	721.94	343.00
- sqrt(x2)	1	228.07	722.17	343.06
- x1	1	233.44	727.54	344.38
- x6	1	1155.61	1649.71	490.93

Step: AIC=264.75

```
y ~ x1 + x6 + factor(dataset$Investment_Freedom) + x8 + sqrt(x2) +
  x4 + factor(dataset$Financial_Freedom) + x7 + x9 + factor(dataset$Continent) +
  x10
```

	Df	Sum of Sq	RSS	AIC
<none>		440.02	264.75	
+ x9: factor(dataset\$Continent)	5	46.40	393.62	270.74
+ factor(dataset\$Financial_Freedom): x3	4	22.32	417.70	276.18
- x10	1	54.08	494.10	280.31
- x9	1	56.34	496.36	281.12
- x7	1	73.45	513.48	287.19
- factor(dataset\$Continent)	5	138.27	578.29	287.72
- factor(dataset\$Financial_Freedom)	3	180.75	620.78	310.79
- x1	1	182.73	622.75	321.73
- x8	1	188.04	628.07	323.25
- factor(dataset\$Investment_Freedom)	3	230.93	670.95	324.70
- x4	1	206.01	646.03	328.30
- sqrt(x2)	1	237.02	677.04	336.69
- x6	1	1163.11	1603.13	490.99

> summary(NewstepwiseBIC)

Call:

```
lm(formula = y ~ x1 + x6 + factor(dataset$Investment_Freedom) +
  x8 + sqrt(x2) + x4 + factor(dataset$Financial_Freedom) +
  x7 + x9 + factor(dataset$Continent) + x10, data = dataset)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.9833	-0.9779	0.1600	1.0142	4.3746

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.127417	1.813711	-0.070	0.9441
x1	0.130524	0.016063	8.126	1.17e-13 ***
x6	0.097336	0.004748	20.501	< 2e-16 ***
factor(dataset\$Investment_Freedom)2	2.568855	0.625514	4.107	6.40e-05 ***
factor(dataset\$Investment_Freedom)3	4.949718	0.692775	7.145	3.07e-11 ***
factor(dataset\$Investment_Freedom)4	6.541614	0.838572	7.801	7.66e-13 ***
x8	0.093523	0.011346	8.243	5.90e-14 ***
sqrt(x2)	1.644876	0.177737	9.255	< 2e-16 ***
x4	0.112249	0.013010	8.628	6.09e-15 ***
factor(dataset\$Financial_Freedom)2	2.177123	0.515267	4.225	4.01e-05 ***
factor(dataset\$Financial_Freedom)3	3.921454	0.613551	6.391	1.74e-09 ***
factor(dataset\$Financial_Freedom)4	6.668367	0.847143	7.872	5.10e-13 ***
x7	0.076487	0.014846	5.152	7.55e-07 ***
x9	0.077083	0.017085	4.512	1.24e-05 ***
factor(dataset\$Continent)1	-0.903756	0.467443	-1.933	0.0550 .
factor(dataset\$Continent)2	-1.333422	0.596941	-2.234	0.0269 *
factor(dataset\$Continent)3	-0.267762	0.422037	-0.634	0.5267
factor(dataset\$Continent)4	-2.935489	0.620930	-4.728	4.97e-06 ***
factor(dataset\$Continent)5	-2.753005	0.490751	-5.610	8.79e-08 ***
x10	0.079348	0.017950	4.420	1.82e-05 ***

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

Residual standard error: 1.664 on 159 degrees of freedom

Multiple R-squared: 0.9776, Adjusted R-squared: 0.9749

F-statistic: 365.3 on 19 and 159 DF, p-value: < 2.2e-16

> AIC(NewstepwiseBIC)

[1] 710.9794

> BIC(NewstepwiseBIC)

[1] 777.9145