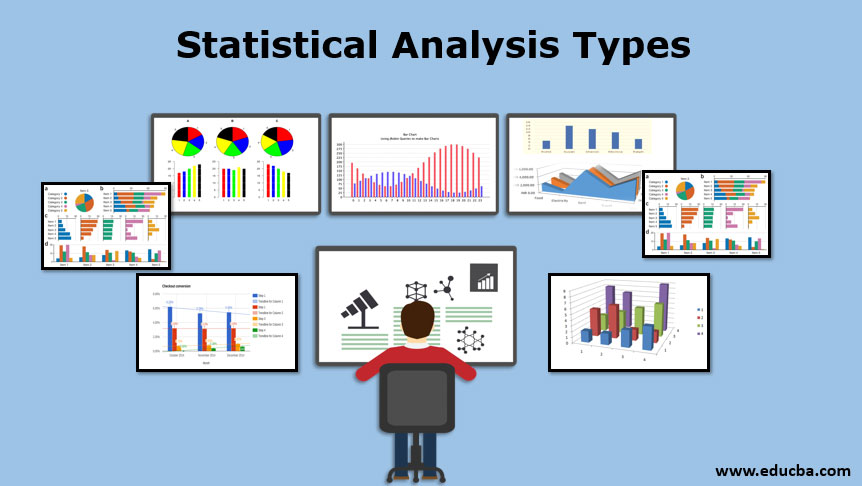
פרויקט במודלים של רגרסיה לינארית חלק ב'



מגישים :   
316320779  
316164177

קבוצה 41

**תוכן עניינים**

1.[**תקציר מנהלים** 3](#_Toc124885454)

[2. **עיבוד מקדים** 4](#_Toc124885455)

[2.1 הסרה של משתנים 4](#_Toc124885456)

[2.2 התאמת משתנים 6](#_Toc124885457)

[2.3 הגדרת משתנה דמה 7](#_Toc124885458)

[2.4 הגדרה והוספת משתני אינטראקציה 8](#_Toc124885459)

[3.**התאמת המודל ובדיקת הנחות המודל** 10](#_Toc124885460)

[3.1. בחירת משתני המודל: 10](#_Toc124885461)

[3.2 . בדיקת **הנחות** המודל 11](#_Toc124885462)

[4.**שיפור המודל** 13](#_Toc124885463)

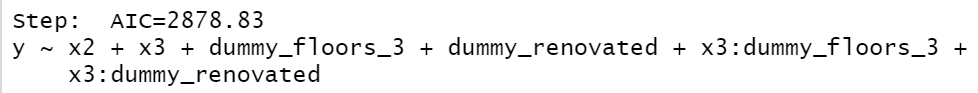
[**נספחים** 15](#_Toc124885464)

# תקציר מנהלים

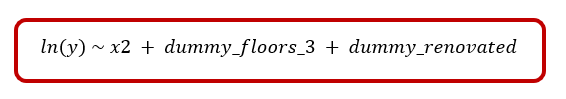
מטרת הפרויקט היא בניית מודל רגרסיה לינארית ושיפורו באמצעות כלים וידע תיאורטי אשר נלמדו בשיעורים תוך שימוש בתוכנת ה- R על מנת להגיע למודל שעוזר להסביר את המשתנה המוסבר בצורה הטובה ביותר. על ידי חקר משתנים שונים המשפיעים על מחיר בית שהינו המשתנה המוסבר במודל שלנו. במהלך חלק זה של הפרויקט בחנו התאמה ואי התאמה של משתנים מהמודל המקורי אותם ניתחנו והבנו לעומק בחלק א' של הפרויקט, תוך בחינת מידת השפעתם על המשתנה המוסבר. לשם כך נעזרנו במקדם המתאם של פירסון, בתרשימי פיזור, בהפיכת משתנים רציפים לקטגוריאליים ובאיחוד קטגוריות במידת הצורך. במהלך הבדיקה בחנו גם משתני אינטראקציה של משתנים קטגוריאליים שונים עם משתנים רציפים. וקיבלנו את המודל הבא:



לאחר מכן, הגדרנו את המודל ובדקנו באמצעות אלגוריתמי רגרסיה שונים את המודל האופטימלי. כלל האלגוריתמים השונים הצביעו על המודל הבא שנתן את ערך הAIC הנמוך ביותר ושהוסרו ממנו משתנים:

**  
  
עבורו בדקנו האם הנחות הרגרסיה הלינארית מתקיימות. לאחר שגילינו כי הנחת הלינאריות מתקיימת, בעוד שהנחת שוויון השונויות והנחת ההתפלגות הנורמלית של השגיאות לא מתקיימות, החלטנו לבצע בחינת מודל עם טרנספורמציה Box-cox כאשר 𝜆 שואף ל-0. ביצענו טרנפורמציות נוספות על מנת לשפר את המודל ולהגיע למודל שיסביר בצורה הטובה ביותר את המסביר. לבסוף המודל שבחרנו מקיים את שלוש הנחות הרגרסיה הלינארית ונותן ערך מדד AIC משמעותית קטן יותר מכל קודמיו.

והרי המודל הנבחר שלנו:



מודל זה מסייע לחזות את מחיר הנכס בהתבסס על האם הוא שופץ או לא, שטחו והאם יש בו 3 קומות.

טבלת משתנים מחלק א':

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| שם המשתנה | סוג המשתנה מוסבר/מסביר | סימון במודל | יחידת מידה | סוג המשתנה – רציף / קטגוריאלי | הסבר קצר על המשתנה |
| rooms | מסביר |  | מספרים | רציף | מספר החדרים בבית |
| sqft\_house | מסביר |  | מספרים | רציף | גודל הבית בSquare Feet |
| sqft\_lot | מסביר |  | מספרים | רציף | גודל המגרש בSquare Feet |
| sqft\_basement | מסביר |  | מספרים | רציף | גודל המרתף בSquare Feet |
| Floors | מסביר |  | מספרים | רציף | מספר הקומות בנכס |
| Condition | מסביר |  | 1-לא טוב 2- כמעט טוב 3- טוב 4- כמעט טוב מאוד 5-טוב מאוד | קטגוריאלי | מצב הנכס |
| yr\_built | מסביר |  | מספרים | רציף | השנה בה הנכס נבנה |
| Renovated | מסביר |  | 1-כן 0-לא | קטגוריאלי | האם הנכס שופץ (1-כן, 0-לא) |
| Price | מוסבר | Y | מספרים | רציף | מחיר הנכס |

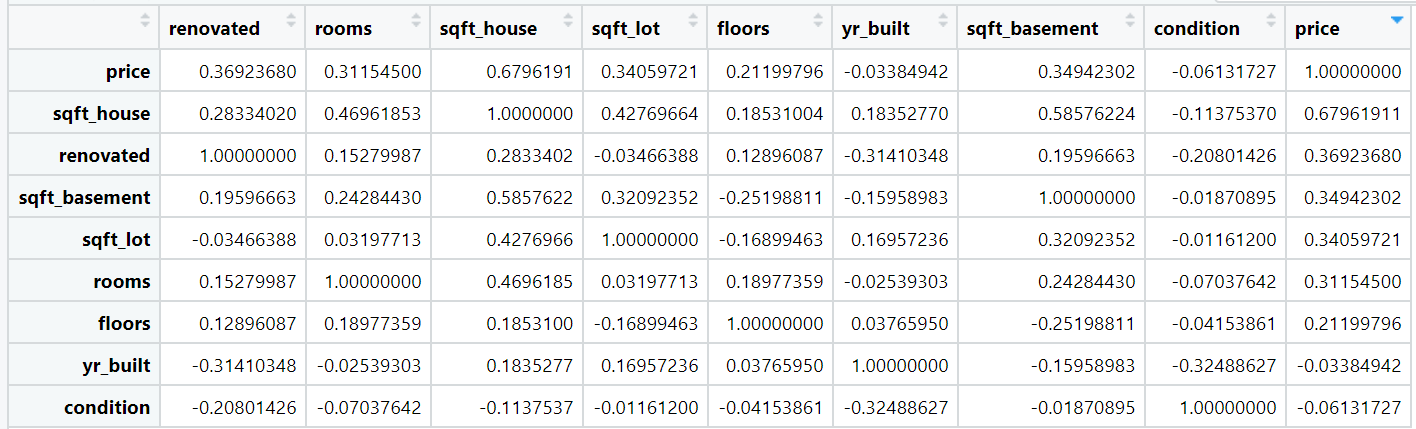
# 2. עיבוד מקדים

# 2.1 הסרה של משתנים

על מנת לבחון האם יש להסיר חלק מהמשתנים במודל נשתמש במקדם המתאם של פירסון ובמובהקות תוצאה P-Value עבור משתנים רציפים, ועבור משתנים קטגוריאלים נשתמש במובהקות תוצאה P-Value בתרשימי Boxplot. בנוסף נרצה לבחון הסרה של משתנה מסביר אם קיים קשר לינארי חזק בינו לבין משתנה מסביר נוסף כיוון שנרצה עבור כל משתנה במודל ערך שונה וייחודי לו.

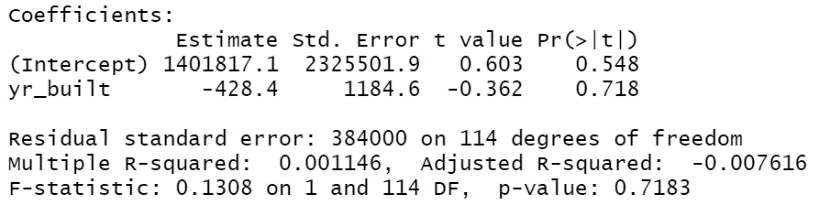
מקדם המתאם של פירסון מסייע בבחינת הקשר הלינארי בין שני משתנים רציפים. כשערכו שואף ל-1 בערך מוחלט, ניתן לומר כי קיים לינארי חזק בין המשתנים, ואילו כשערכו שואף לאפס ניתן לומר שהקשר בין המשתנים חלש.

מובהקות התוצאה P-Value הינה ההסתברות לקבל תוצאה זהה או קיצונית לפחות כמו זו שהתקבלה בניסוי תחת השערת האפס (אשר מחזקת את ההשערה האלטרנטיבית). במודל הרגרסיה המרובה נבדוק באיזו מידה המשתנה שנבחן מסביר את המוסבר.   
  
בטבלה הבאה מוצגות הקורלציות בין המשתנים המסבירים, ובינם לבין המשתנה המוסבר. ניכר כי בין המשתנים המסבירים אין מקדם מתאם גבוה אשר מצריך בחינה.  
בחרנו לבחון את כדאיות ההסרה של ארבעת המשתנים המסבירים אשר להם מקדם המתאם הנמוך ביותר, לאחר מכן, בחנו את ערך ה P-Value בכדי להחליט באופן סופי האם כדאי להסירם מהמודל או לא.

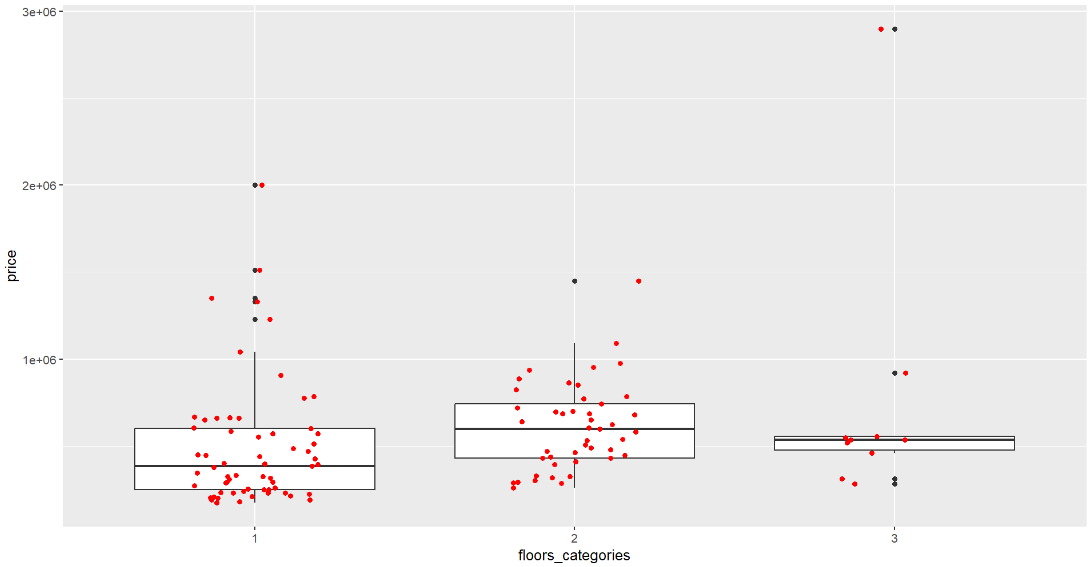


Year built:

בין המשתנה המוסבר למשתנה רציף זה יש קורלציה 0.033849- .מכך ומתוך הגרף כמעט ולא ניתן לזהות מתאם בין המשתנים. בהסתכלות על מודל הרגרסיה התקבל 0.7183=P-Value גבוה. נתונים אלו מחזקים את מה שציפינו לראות, כיוון שגם בשנים מוקדמות וגם בשנים מאוחרות נבנו בתים מכל מיני סוגים ובמחירים שונים. ההסתכלות על שנת הבניה בלבד אינה משקפת את מחיר הבית ועל כן גם לא מסבירה נכון את המשתנה המוסבר. לכן נבחר להסיר את משתנה זה מהמודל, מאחר והנתונים מעידים על כך שאינו מסביר טוב את המשתנה המוסבר ואין קשר סיבתי ניכר בין השניים.

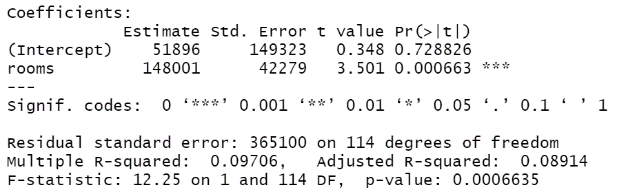


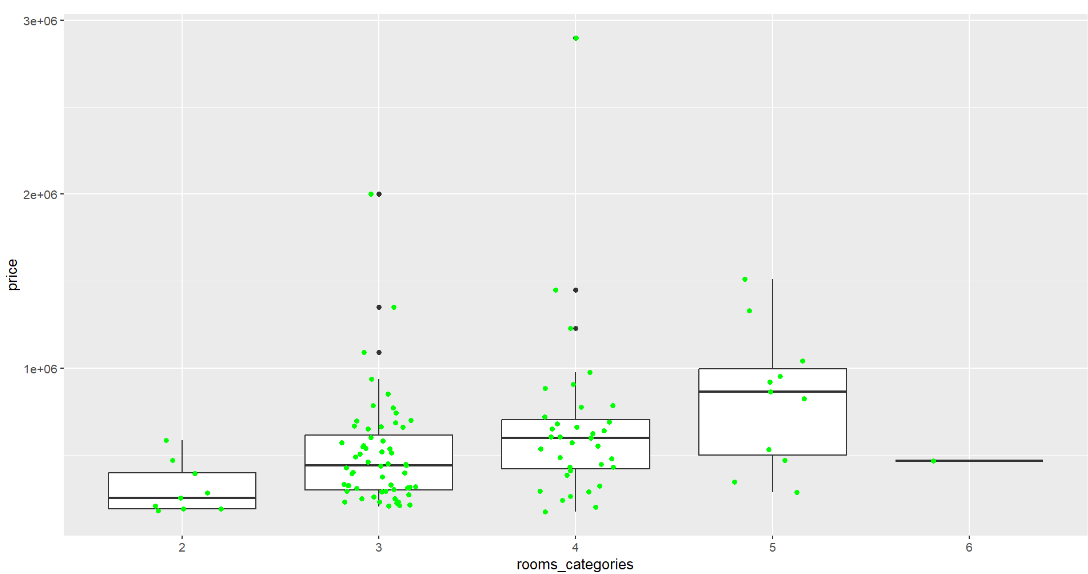
:Floors  
בין המשתנה המוסבר למשתנה רציף זה יש קורלציה 0.21199796. מכך ומתוך הגרף כמעט ולא ניתן לזהות מתאם בין המשתנים. בהסתכלות על מודל הרגרסיה התקבל P-value=0.02234 נמוך יחסית, דבר המחזק את הטענה שיתכן וקיים קשר בין המשתנים. ציפינו שיהיה קשר בין מספר הקומות בנכס לבין מחירו כי ייתכן ונכסים בעלי מספר רב יותר של קומות יכולים לאכלס כמות גבוהה יותר של דיירים ועל כן ניתן גם לגבות מחיר גבוה יותר. ועל אף שהקשר אינו מובהק כפי שציפינו, נבחר להשאיר את משתנה זה במודל.

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

:Rooms

בין המשתנה המוסבר למשתנה רציף זה יש קורלציה 0.31154500. מכך ומתוך הגרף כמעט ולא ניתן לזהות מתאם בין המשתנים. בהסתכלות על מודל הרגרסיה התקבל P-value=0.0006635 נמוך מאוד, דבר המחזק את הטענה שייתכן וקיים קשר בין המשתנים. ציפינו שיהיה קשר בין מספר החדרים בנכס לבין מחירו כי ייתכן ונכסים בעלי מספר רב יותר של חדרים יכולים לאכלס כמות גבוהה יותר של דיירים ועל כן ניתן גם לגבות מחיר גבוה יותר. ועל אף שהקשר אינו מובהק כפי שציפינו, נבחר להשאיר את משתנה זה במודל. בייחוד בשל העובדה שלרוב המשתנים המסבירים שלנו יש קורלציה יחסית דומה (0.3-0.4) ועל אף שקורלציה זו לא נחשבת גבוהה, נצטרך נתונים ומשתנים מסבירים בכדי לבסס את המשך הפרויקט ואת מסקנותינו.



# 2.2 התאמת משתנים

לאחר בחינה אילו משתנים נוכל לשנות ולהגדיר מחדש, ראינו כי יש צורך לבצע דיסקרטיזציה עבור המשתנים – floors ו-rooms מפני שבבסיס הנתונים שלנו כמות המופעים שלהם מוגבלת ומצומצמת ועל כן הם אינפורמטיביים יותר כאשר הם בתצורה של משתנים קטגוריאליים. בשל עובדה זו גם הצגנו אותם בBOXPLOT בסעיף הקודם. למשתנה rooms יהיו 4 קטגוריות 2-5 ומעלה, כך שלכל קטגוריה ישתייכו נכסים בעלי כמות החדרים המצוינת בקטגוריה. בחרנו לאחד את קטגוריות 5,6 מאחר וקיימת רק תצפית אחת של חדר בעל 6 חדרים ועל כן בחרנו לאחד את קטגוריות 5 ו-6 לקטגוריה אחת אליה ישתייכו נכסים בעלי 5 חדרים לפחות.  
למשתנה floors יהיו 3 קטגוריות 1-3 אשר כל קטגוריה מהווה את כמות הקומות בנכס. כך שלכל קטגוריה ישתייכו נכסים בעלי כמות הקומות המצוינת בקטגוריה. על אף שמספר התצפיות עבור בתים בעלי 3 קומות יחסית נמוך, בחרנו שלא לאחד קטגוריה. אנו רואים חשיבות באבחנה בין הקטגוריות הללו. חשוב לנו לציין כי בחרנו שלא לאחד קטגוריות במשתנה condition מחשש לזהם את הנתונים על אף שלערכים 1,2 יש תצפיות מועטות מאוד (נספח 2.2.1)

בנוסף, נרצה לבצע דיסקרטיזציה עבור משתנה sqft\_house. מצאנו לנכון לחלק את נתוניו ל-4 קטגוריות : , "mansion""villa", ""single family ,"studio". כך ש"studio" יהיה עבור נכסים בגודל של עד 1100, "single family" יהיה עבור נכסים בגודל של 1160 עד 2000 ,"villa" יהיה עבור נכסים בגודל של 2010 עד 3000 ו- "mansion" יהיה עבור נכסים בגודל של 3050 עד 5420. (יח' המידה עבור הגדלים שצוינו הן sqft).  
בחרנו לבצע חלוקה זו מאחר ורצינו לתת משמעות לגודל הנכס ולהבין באיזה סוג נכס מדובר, מעבר לגודל שלו. לגודל הנכס יש השפעה על מחיר הנכס, וכן כשמדובר בגודל בלבד, מאוד קשה להבין באיזה סוג נכס מדובר ועל כן הפיכת משתנה זה למשתנה קטגוריאלי, מאפשרת לקטלג את המחיר ולהתאימו לסוג הנכס. מהלך זה לא יפגע בנכונות המודל ואף יתרום להבנתו בצורה טובה יותר.

מעבר למשתנים אלה, לא מצאנו משתנים אשר איחודם או הפיכתם לקטגוריאליים יהיו רלוונטיים ולא יפגעו בנכונות המודל.

# 2.3 הגדרת משתנה דמה

נגדיר משתני דמה עבור המשתנים הקטגוריאליים במודל:

**rooms**

קבוצת הבסיס תהיה 2 rooms ונגדיר 3 משתני דמה :

**Sqft\_House**

קבוצת הבסיס תהיה -Studio ונגדיר 3 משתני דמה :

**floors**

קבוצת הבסיס תהיה 1 floors ונגדיר 2 משתני דמה :

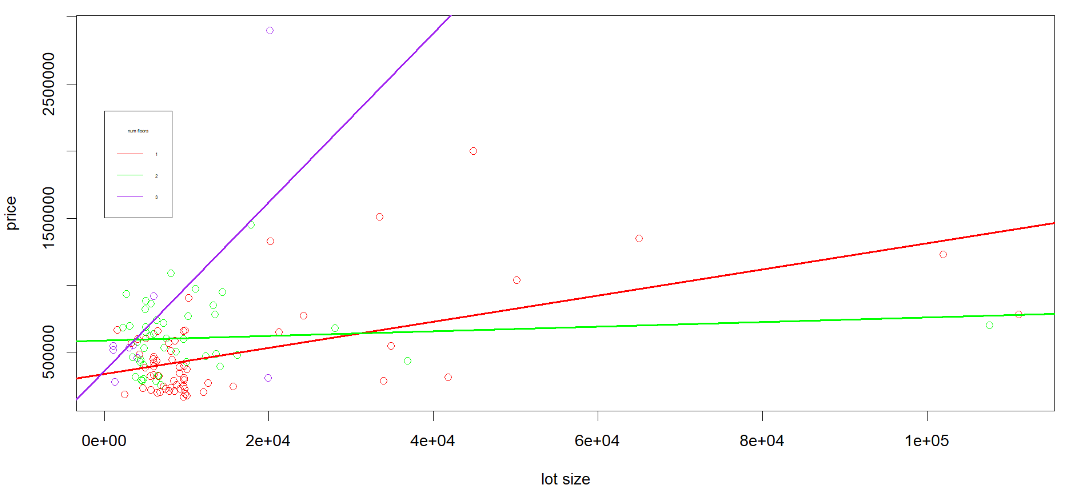
**condition**   
קבוצת הבסיס תהיה – 1 ונגדיר משתנה דמה אחד לכל אחת מהרמות.

**Renovated**   
קבוצת הבסיס תהיה -not renovated ונגדיר משתנה דמה אחד ל -:renovated

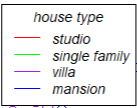
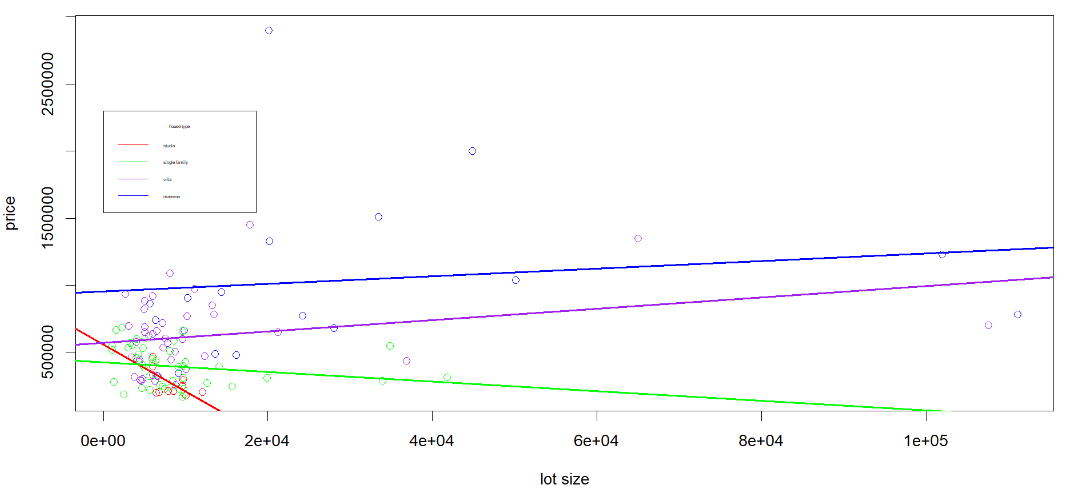
# 2.4 הגדרה והוספת משתני אינטראקציה

משתנה אינטראקציה הוא משתנה שמביע את תרומתה השולית של כל קטגוריה לשיפוע המודל. בעזרתו נוכל לדעת האם ישנה השפעה כלשהי לכל קטגוריה בנפרד יחד עם משתנה מסביר בנוסף למשתנה המוסבר. ננסה לבחון את ההשפעה שעשויה להיות למשתנים הקטגוריאליים בצירוף המשתנים הרציפים וכתוצאה מכך להבין את משמעות האינטראקציה:

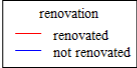
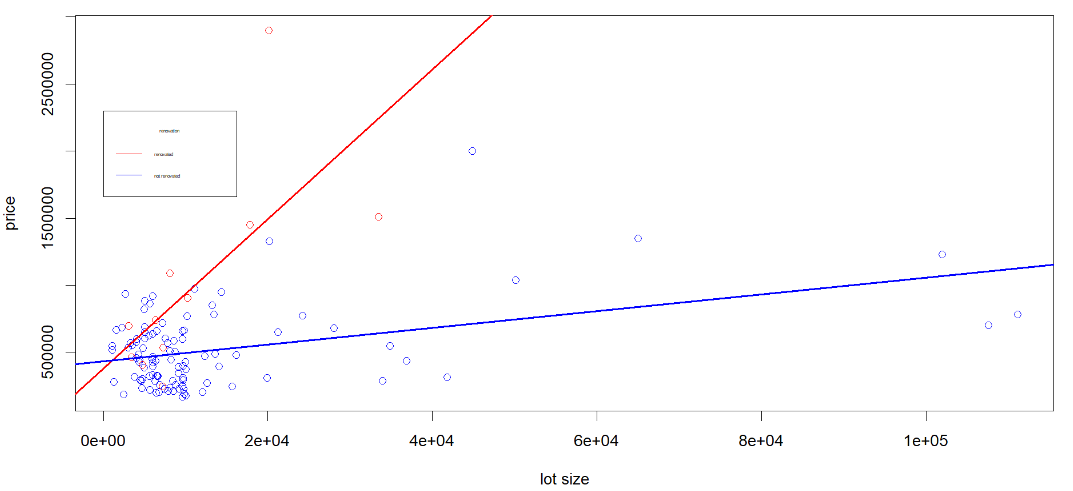
1. נבחן את השילוב של המשתנה הרציף sqft\_lot (*) יחד עם המשתנה הקטגוריאלי* floors (*). אנו משערים כי יש הבדל בין שטח המגרש עבור מבנים בעלי קומות שונות, כך שבנייה לגובה עשויה להקטין את שטח המגרש עליו בנוי הנכס לעומת נכס בעל קומה אחת אשר רוב גודל המגרש ישמש לשטח הבית הבנוי. מהגרף ניתן לראות כי השיפועים בשלוש הקטגוריות (מספר קומות) שונים מאוד וכן ניכר גם הבדל בחותכים שלהם. על כן נבחר להוסיף משתנה אינטראקציה זה למודל שלנו. בנספחים ניתן למצוא את סיכום האינטראקציה (נספח 2.4.1)*



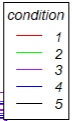
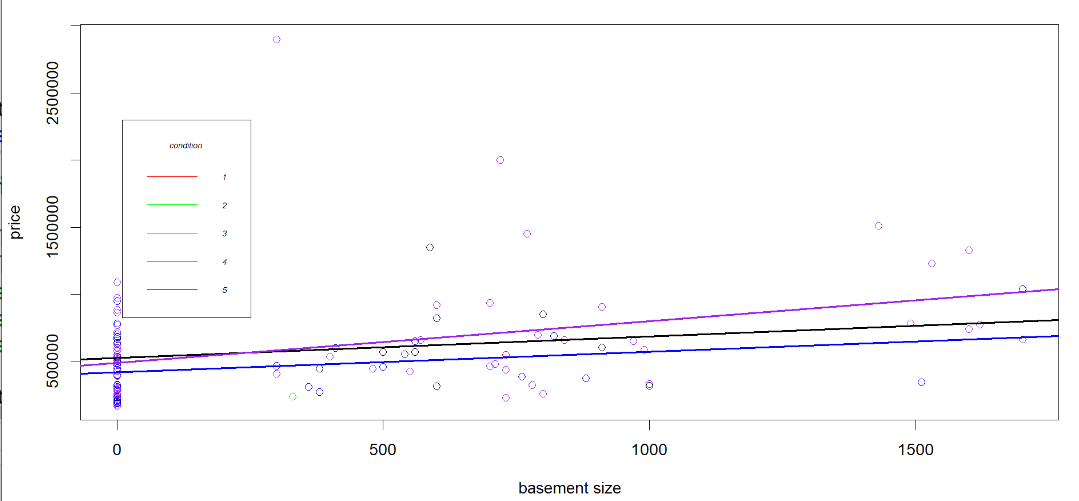
1. בנוסף נבחן את השילוב של המשתנה הרציף sqft\_lot (*) יחד עם המשתנה הקטגוריאלי שיצרנו house type המבוסס על sqft\_house ) אנו משערים כי יש הבדל בין סוג הבית המתבטא באופיים השונה, כך שלווילה לדוגמה, למשפחה יחידה יוקצה שטח קטן יותר מעבר לשטח הבית הבנוי.* מהגרף נוכל לראות כי זוגות הקווים מתנהגים בצורה דומה, דבר שעלול להעיד על צורך באיחוד קבוצות, אך לא נרצה לאחדם. בנוסף, ניתן לראות כי רק קבוצה אחת מייצגת קו חריג ביחס לשאר לכן נבחר לצרף למודל את משתנה האינטראקציה הנ"ל. (סיכום אינטראקציה בנספח 2.4.2)



1. נקיים בחינה נוספת עבור השילוב של המשתנה הרציף sqft\_lot (*) יחד עם המשתנה הקטגוריאלי*renovated (*). אנו משערים כי אחד השיפוצים השגורים ביותר הוא הרחבת שטח המגרש לטובת חנייה, בריכה, חצר ועוד. ניכר כי שיפוץ והגדלת שטח המגרש מעלים את מחיר הנכס. מהגרף ניתן לראות כי השיפועים בשתי הקטגוריות שונים מאוד וקיים גם הבדל בחותכים שלהם. על כן נבחר להוסיף משתנה אינטראקציה זה למודל שלנו.* (סיכום אינטראקציה בנספח 2.4.3)



1. נקיים בחינה נוספת עבור השילוב של המשתנה הרציף sqft\_basement (*) יחד עם המשתנה הקטגוריאלי* condition(*). אנו משערים* כי בתים אשר מכילים מרתף גדול יותר נבנו בשנה מאוחרת יותר על פי תקן בניה מחמיר ועדכני. דבר שעשוי להשליך על מצבו הכללי של הבית ועל ציונו בסקאלה. *מהגרף ניתן לראות כי השיפועים בשתי הקטגוריות דומים מאוד וכמעט ולא קיים הבדל בחותכים שלהם. על כן נבחר שלא להוסיף משתנה אינטראקציה זה למודל שלנו.* (סיכום אינטראקציה בנספח 2.4.4)  
   \* הערה: לא ניתן לראות בגרף ערכים עבור ציונים 1 ו-2 מאחר ואין מספיק תצפיות.



מעבר למשתנים אלו, לא מצאנו משתנים נוספים שהאינטראקציה ביניהם עשויה להוסיף מידע חיוני למודל שלנו.

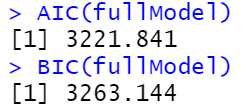
# 3.התאמת המודל ובדיקת הנחות המודל

# 3.1. בחירת משתני המודל:

על מנת לבחון ולבחור את משתני המודל שלנו נשתמש באלגוריתמים של רגרסיה לאחור, רגרסיה לפנים ורגרסיה בצעדים לפי מדד AIC ולפי מדד BIC. את המודל נבחר לפי ערך הAIC המינימלי המתקבל. במצב של שוויון במדד הAIC נשתמש במדד R^2 adj כדי לבחור את המודל, נבחר במודל בעל הערך המקסימלי מבין כולם.

המודל הנוכחי לאחר הוספת משתני דמה ומשתני אינטראקציה :

*כלומר מודל זה כולל את כל המשתנים המסבירים מלבד X7 אותו בחרנו להסיר בסעיף הקודם. בתוספת משתני האינטראקציות מספר קומה\*גודל המגרש, סוג הבית\*גודל המגרש ומצב הנכס\*גודל המגרש.*

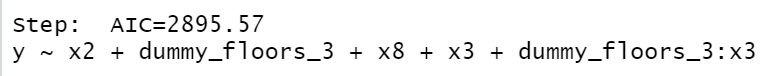
*חישבנו עבור המודל הנוכחי את המדדים: (נספח 2.8)*

*בחרנו לבחון שלוש שיטות שלמדנו בשיעור שיסייעו לנו לבחור את המודל המתאים לאחר שצירפנו למודל את משתני האינטראקציות שציינו מעלה:*

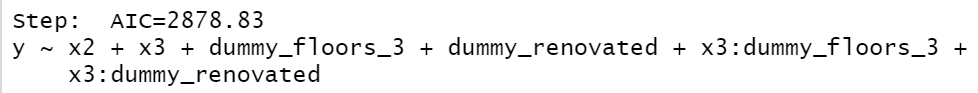
*בחינת המודל לפני שיטת Forward (נספח 3.1)*

*תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטיהקלט לאלגוריתם זה יהיה מודל ריק ללא משתנים מסבירים ועם חותך בלבד. מבדיקה זו התקבלו  
המדדים: (נספח 2.9)*

*בכל שלב נוסיף משתנה, עד שמתקבל המודל הבא:*

*מדד ה AIC שהתקבל לאחר בחירת משתנים רלוונטיים הוא AIC=2895.57 , מדד טוב יותר מאשר בתחילת הבדיקה.*

*בחינת המודל לפני שיטת Backward (נספח 3.2)  
הקלט הראשוני בבחינת שיטה זו הוא מודל הכולל את כל המשתנים, כולל משתני האינטראקציה. האלגוריתם מוריד בכל שלב את המשתנה בעל ערך ה AIC הנמוך ביותר, ובמקרה שלנו הסיר אחד אחרי השני את המשתנים. עד לקבלת המודל הבא: *

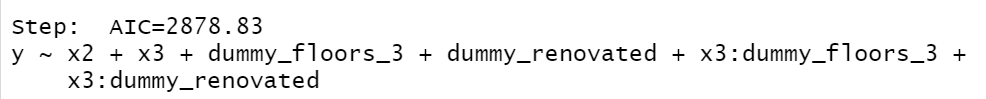
*מדד AIC , שהוא טוב יותר ממדד ה AIC שהתקבל בשיטת Forward ומערך המדד של המודל המקורי.*

*בחינת המודל לפני שיטת stepwise* (*נספח 3.3):*

*תמונה שמכילה טקסט

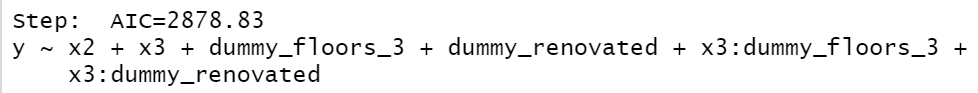
התיאור נוצר באופן אוטומטי*האלגוריתם מתחיל ללא משתנים מסבירים כמו בשיטת *Forward*, ומתחיל לבחון כניסת משתנים וכן לבדוק גם משתנים אשר נמצאים בתוך המודל האם הם צריכים להישאר חלק ממנו או לא.   
ערכי המדדים במצב הראשוני של האלגוריתם:

והמודל שמתקבל לאחר ביצוע הצעדים של אלגוריתם:



ניכר כי מדד הAIC זהה עבור *stepwise* ו- *Backward*בעוד שבשיטת *Forward* הוא יוצא גבוה יותר. נבחר במודל שמתקבל ע"י האלגוריתם *Backward* מאחר ובשלב הסופי הAIC שלו נמוך מכולם.

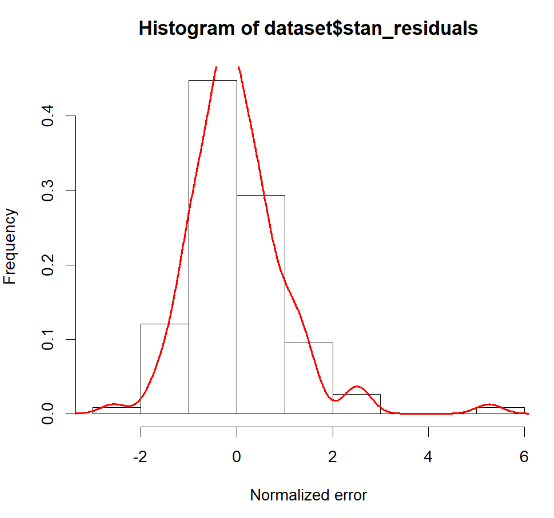
המודל הנבחר וערך הAIC שלו:

**

# 3.2 . בדיקת **הנחות** המודל

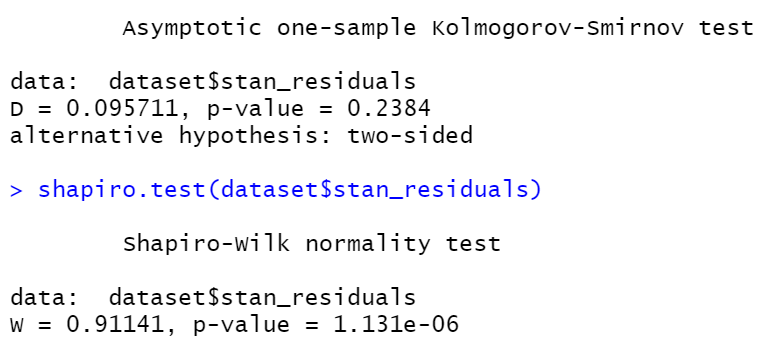
בדיקת **הנחת** הנורמאליות **של השגיאות:**

בהסתכלות על גרף תרשים Normal Q-Q ,המתאר את ההתנהגות המצופה מנתונים המגיעים מהתפלגות נורמלית בהשוואה לנתונים שלנו, ניתן לראות שרוב התצפיות לא נמצאות על הקו הישר, כלומר שההתפלגות לא נורמלית. בבואנו לבחון את ההיסטוגרמה ניכר כי קשה לקבוע באופן חד משמעי שמדובר בהתפלגות נורמלית.

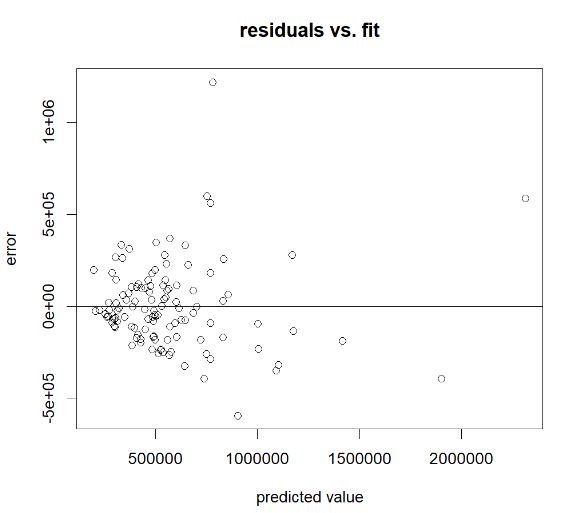


מכאן שעלינו לבצע מבחן KS על מנת לבחון את מובהקות התוצאה.  
לפי מבחן זה מתקבל כי:

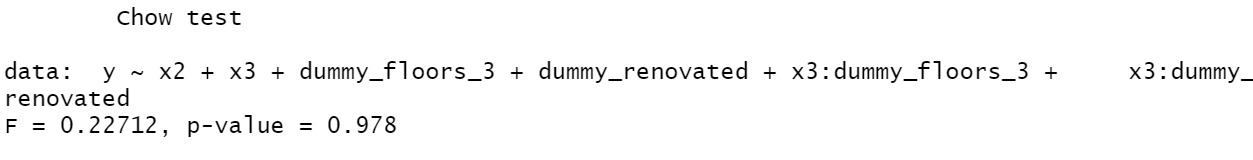
*היות וערך ה המתקבל אינו גבוה, נבצע מבחן נוסף- מבחן SW:*לפי מבחן זה מתקבל כי: *. לכן לפי מבחן זה נדחה את השערת האפס בר"מ 5% ונקבע כי השגיאות המתוקננות אינן מתפלגות נורמאלית.  
המבחנים:*



בדיקת הנחת שיוויון השונויות:  
בהסתכלות על תרשים פיזור השגיאות המתוקננות לעומת ערך החיזוי ניתן לראות פיזור מאוד לא אחיד סביב האפס, ככל שהמחיר עולה כך גם שונות השגיאות גדולה יותר, דבר המקנה צורת משפך. משום כך נסיק כי הנחת שיווין השונויות איננה מתקיימת.



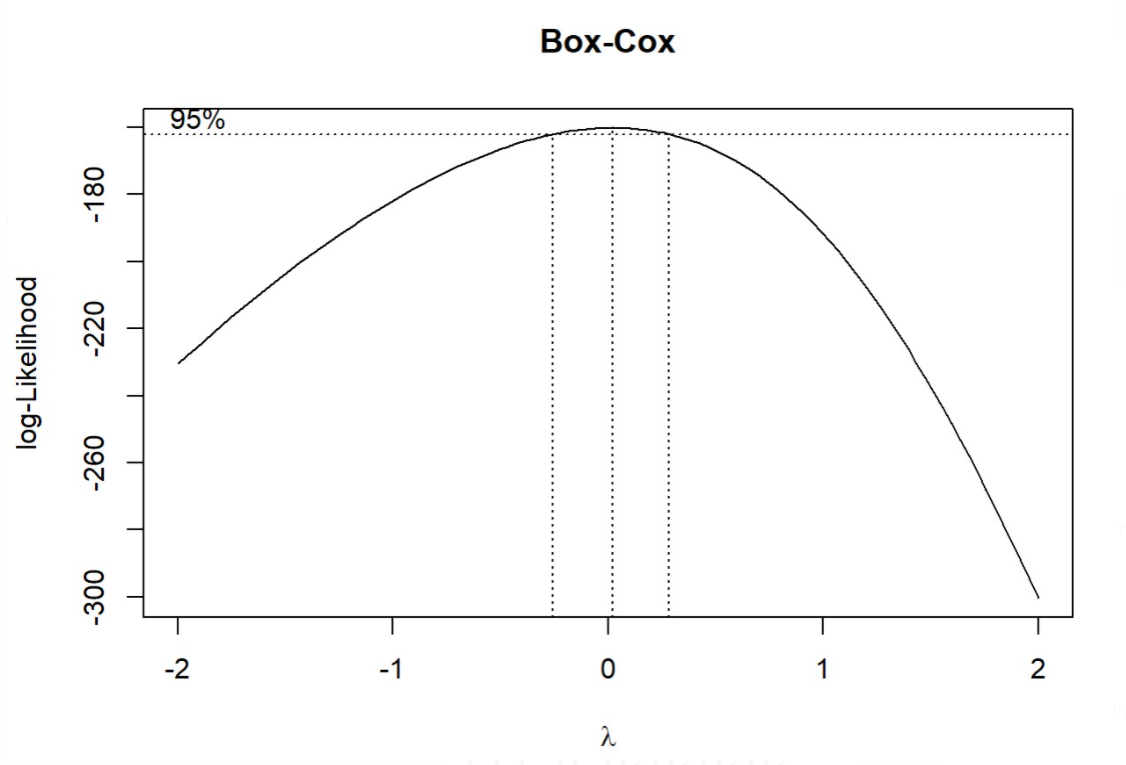
בדיקת הנחת הלינאריות:  
הנחה זו ניתנת למדידה ראשונית ע"י תרשים פיזור השגיאה המתוקננת הנ"ל. מתרשים השאריות ניכר כי לא ניתן לראות קו מגמה ברור בין הנקודות ולכן קשה יותר לקבוע האם הנחת הלינאריות מתקיימת או לא. בכדי לבחון הנחה זו נבצע מבחן סטטיסטי – מבחן Chow.

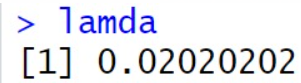
לפי מבחן זה מתקבל כי: *,* לכן נחליט שלא לדחות את השערת האפס ברמת מובהקות 5% ונאמר שהנחת הלינאריות מתקיימת.

# 4.שיפור המודל

עד כה, ביצענו פעולות כגון הסרת משתנים, קטגוריזציה עבור חלקם ואיחוד קטגוריות עבור חלקם האחר. יצרנו משתני דמה ואינטראקציה ובחנו את צירופם למודל שלנו. כך שהגענו למודל המסוגל לחזות את מחיר הנכס באמצעות שני משתני בסיס- שטח הבית ושטח המגרש בתוספת משתנה דמה עבור מספר הקומות, משתנה דמה עבור האם הנכס שופץ או לא ושני משתני אינטראקציה.

Text, letter

Description automatically generatedבסעיף הקודם בחנו את קיום הנחות המודל ונוכחנו לגלות כי בעוד הנחת הלינאריות מתקיימת, הנחת הנורמאליות והנחת שיוויון השונויות לא מתקיימת. על כן נרצה לבצע טרנספורמציה כלשהי במודל.   
אנו בחרנו לבצע טרנספורמציה חזקה –טרנספורמציית Box-Cox. על ידי שימוש בטרנספורמציה עלינו למצוא את האומד ללמדה האופטימלי. שיטה זו מתבססת על שימוש בפונקציית הנראות המקסימלית. מטרת סעיף זה היא לבחון האם הוספת משתנים, או שינוי במשתנים המסבירים או המוסברים משפרת את המודל. לאחר שימוש בהתמרת Box-Cox מ עבור ערך זה מתקבל ערך הLogliklihood הגבוה ביותר. בקירוב ניתן לקבוע כי , הפונקציה שלנו מתנהגת כפונקציה של ln(y). ניתן להשאיף את לאפס מאחר וגודל ה תלוי במדגם עצמו וסביר להניח שאם היינו לוקחים מדגם אחר, ערך ה- יכול היה להיות 0 בדיוק. לכן על מנת להסביר טוב יותר את המודל, אנו מניחים כי .



לאחר בדיקה עבור מודל ה ln(Y) , נראה שבעבור מודל זה שלוש ההנחות מתקיימות.

בדקנו מודלים נוספים וראינו כי אף מודל לא משתווה למודל זה, הן מבחינת קיום ההנחות והן מבחינת ערך הAIC. מתקבל AIC מינימלי ועל כן מודל זה הוא הטוב ביותר שנוכל לקבל.

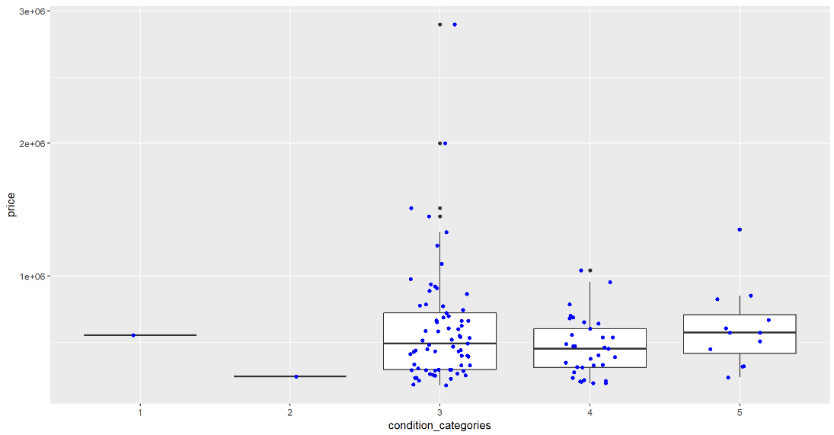
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| נספחים | שוויון שונויות | ליניאריות | נורמליות | AIC(forward) | טרנספורמציה |
| ללא | לא מתקיימת | מתקיימת | לא מתקיימת | 2878.83 | Y |
| 4.1 | מתקיימת | מתקיימת | מתקיימת |  | ln(Y) |
| 4.2 | לא מתקיימת | לא מתקיימת | לא מתקיימת |  |  |
| 4.3 | מתקיימת | לא מתקיימת | מתקיימת | תמונה שמכילה טקסט, כתום  התיאור נוצר באופן אוטומטי | sqrt(Y) |

נבחר במודל לאחר טרנפורמציית ln עבורו מתקבל ערך AIC מינימלי *הקטן אפילו מהמודל המקורי שלנו ועל כן נרצה לבחור בו, בייחוד בשל העובדה שכל ההנחות מתקיימות בו.*

# נספחים

2.2.1 – נספח המכיל BOXPLOT עבור משתנה condition

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, קבלה

התיאור נוצר באופן אוטומטינספח 2.4.1:

נספח 2.4.2:

תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטי

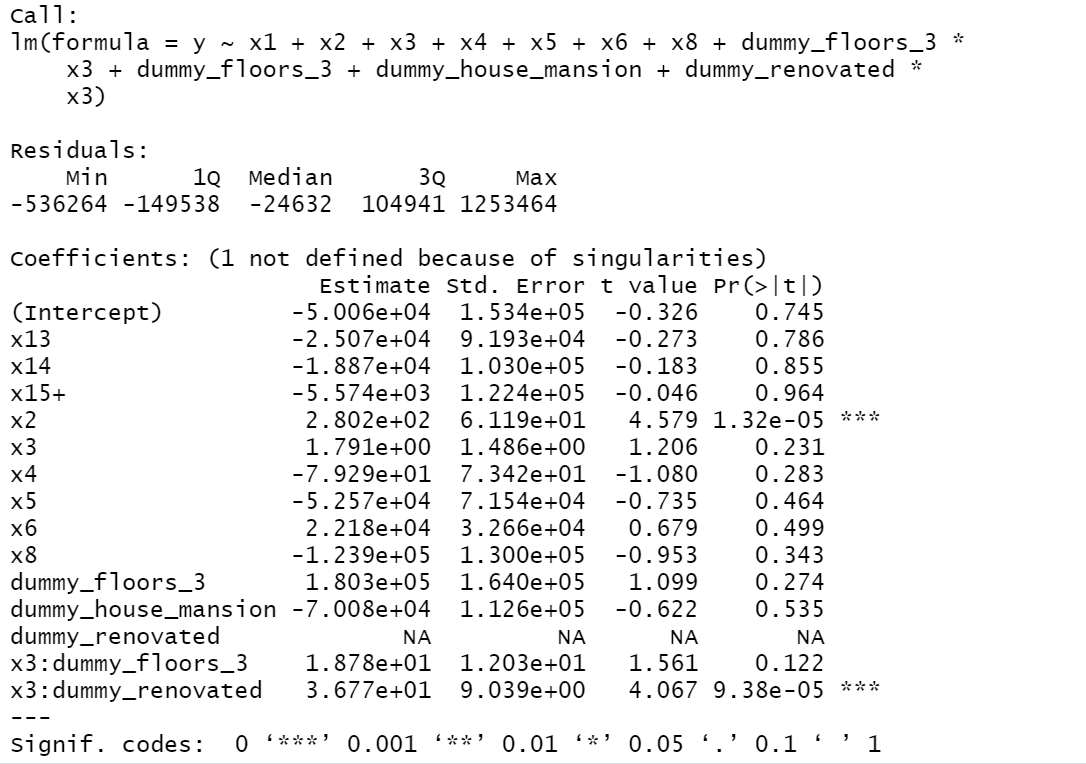
תמונה שמכילה טקסט, קבלה

התיאור נוצר באופן אוטומטינספח 2.4.3:

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטינספח 2.4.4:

נספח 2.8 המודל המלא:

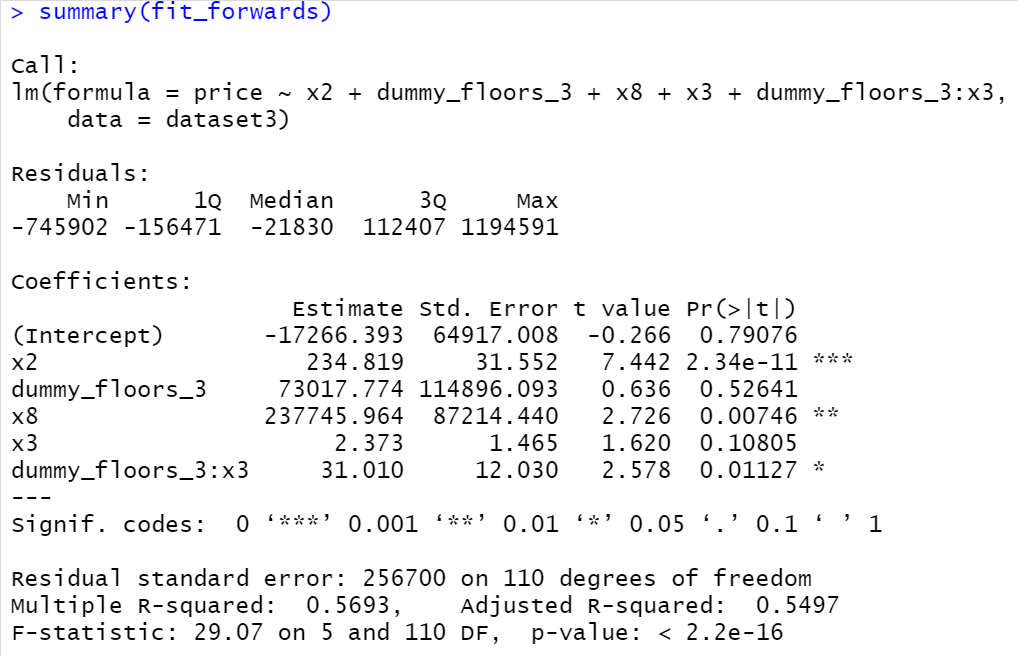


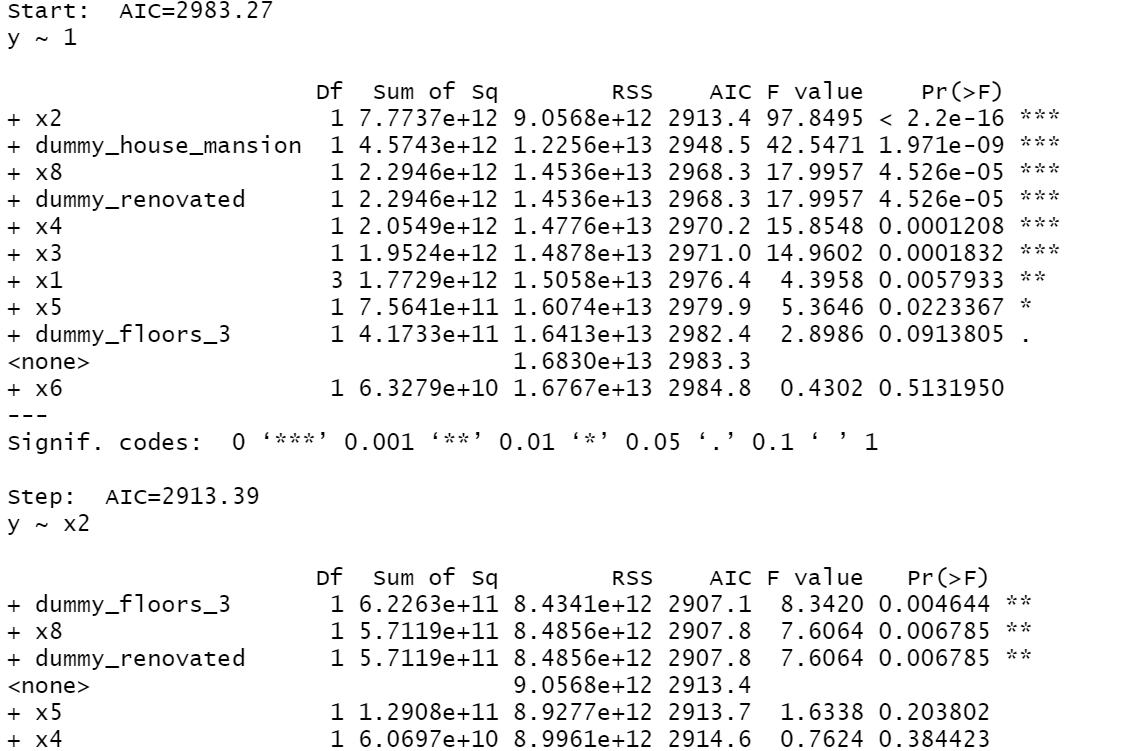
נספח 2.9 - המודל הריק:

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

נספח 3.1- forward:





תמונה שמכילה טקסט, קבלה

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, קבלה

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, קבלה

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, קבלה

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

נספח 3.2- backward:

תמונה שמכילה טקסט, קבלה

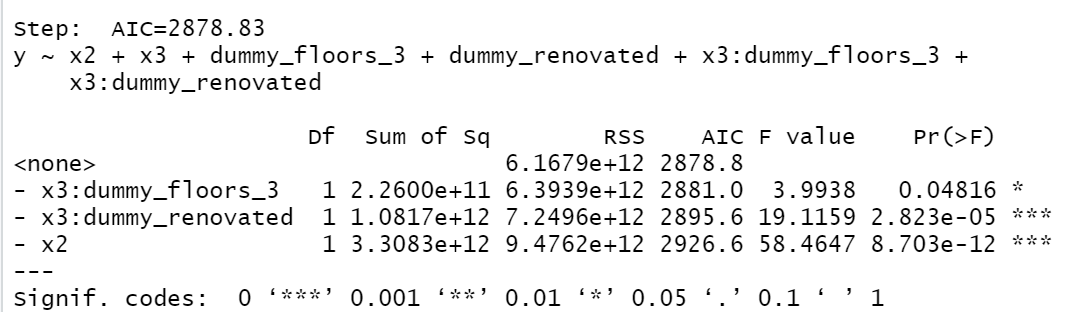
התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, קבלה

התיאור נוצר באופן אוטומטי

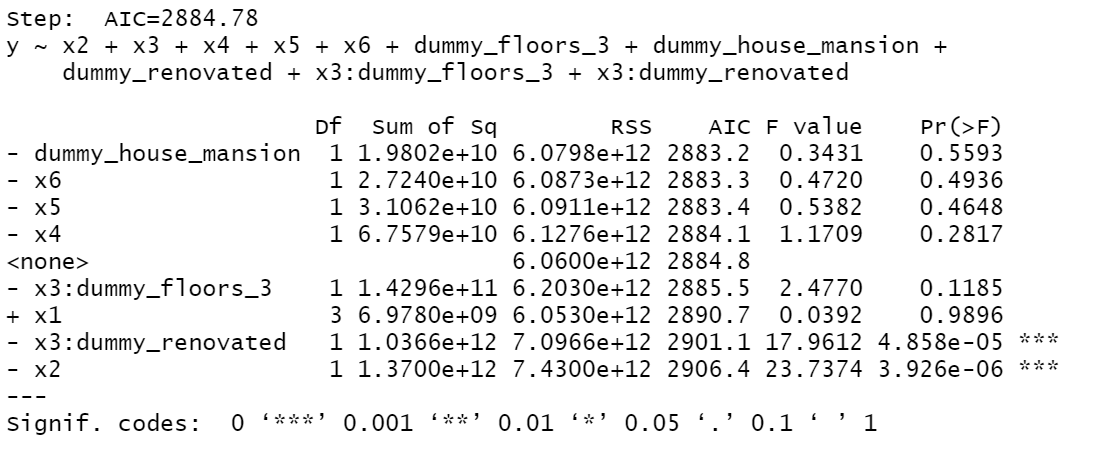
תמונה שמכילה טקסט, קבלה

התיאור נוצר באופן אוטומטי



נספח 3.3- stepside: תמונה שמכילה טקסט, קבלה, צילום מסך

התיאור נוצר באופן אוטומטי



תמונה שמכילה טקסט, קבלה

התיאור נוצר באופן אוטומטי תמונה שמכילה טקסט, קבלה

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, קבלה

התיאור נוצר באופן אוטומטי תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

נספח4.1 כלל הנספחים והגרפים לחישוב ל-ln(y):

Table

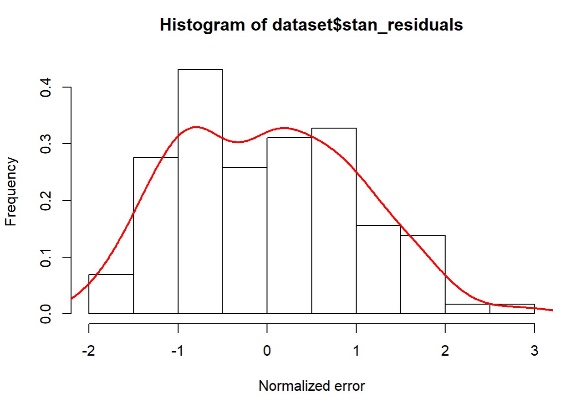
Description automatically generated

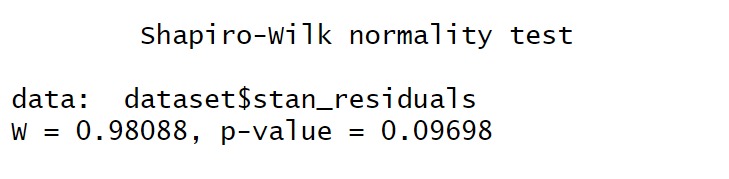
A black and white document

Description automatically generated with low confidence

Chart, scatter chart

Description automatically generated

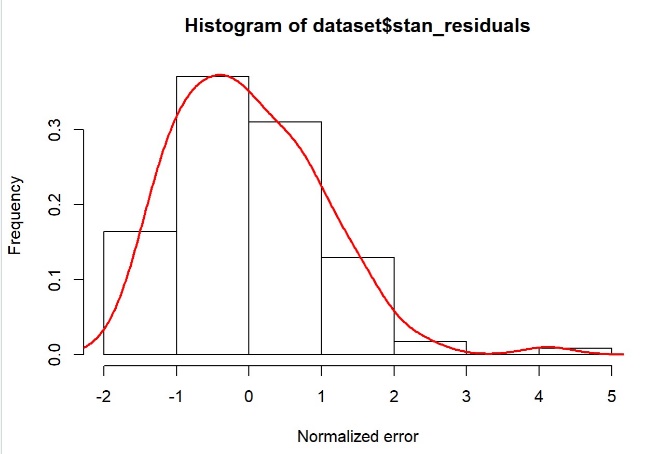
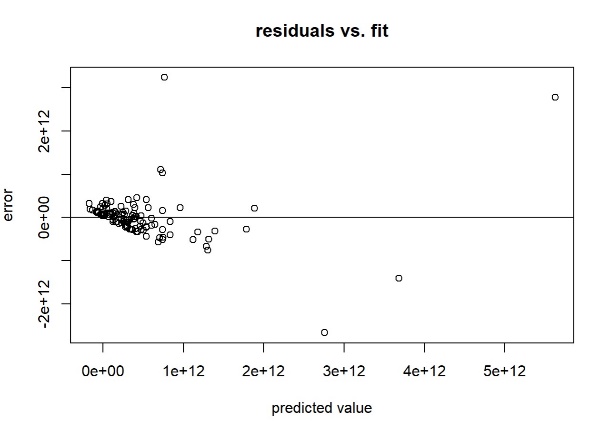
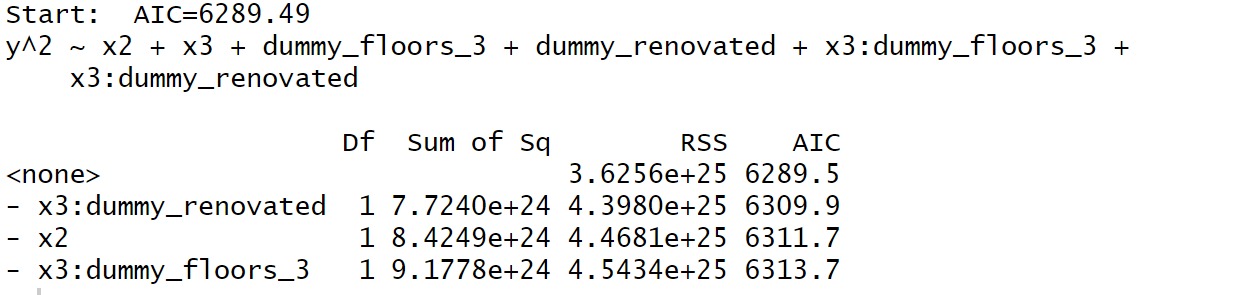
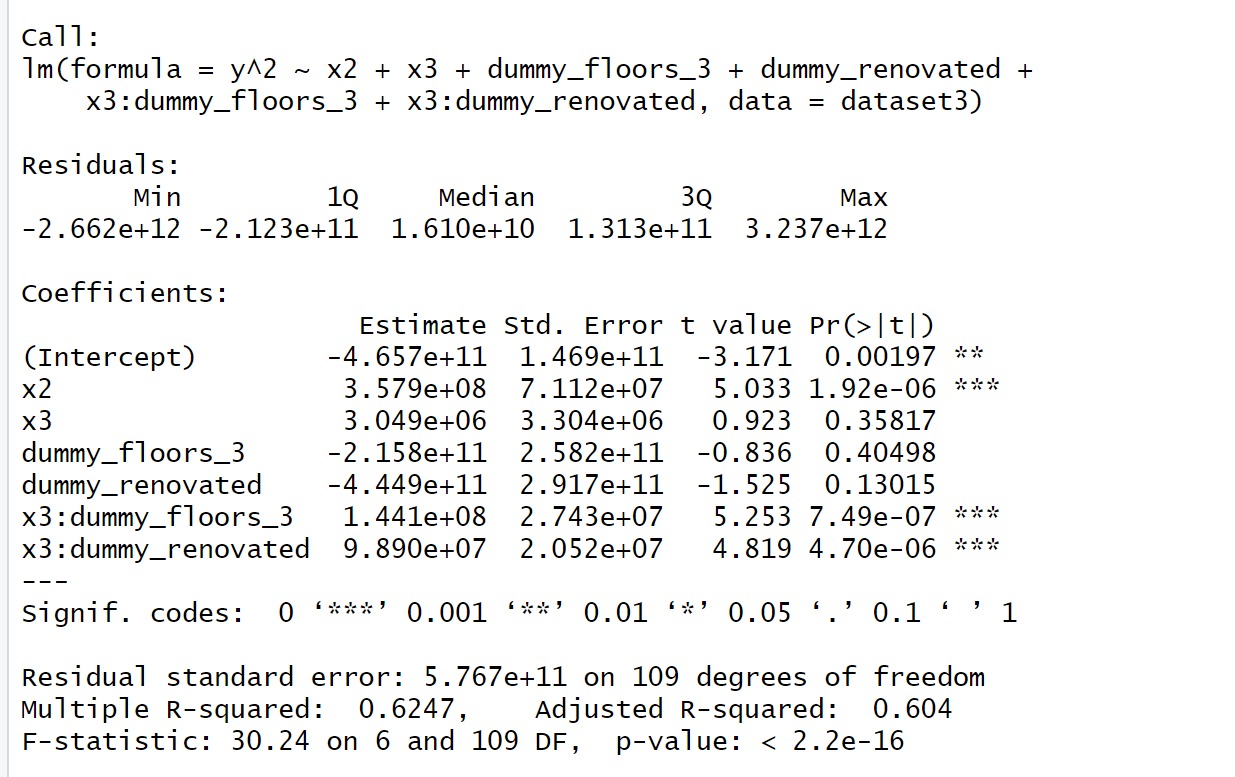
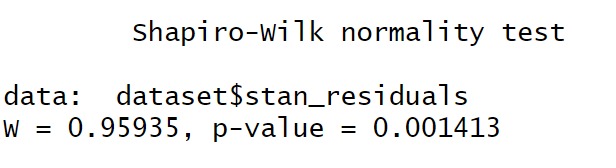
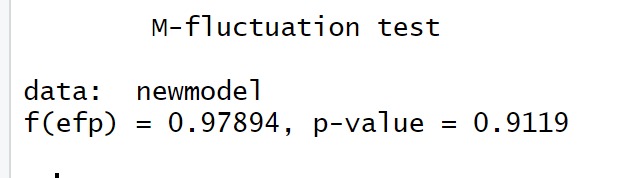


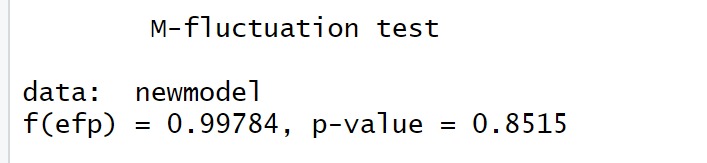
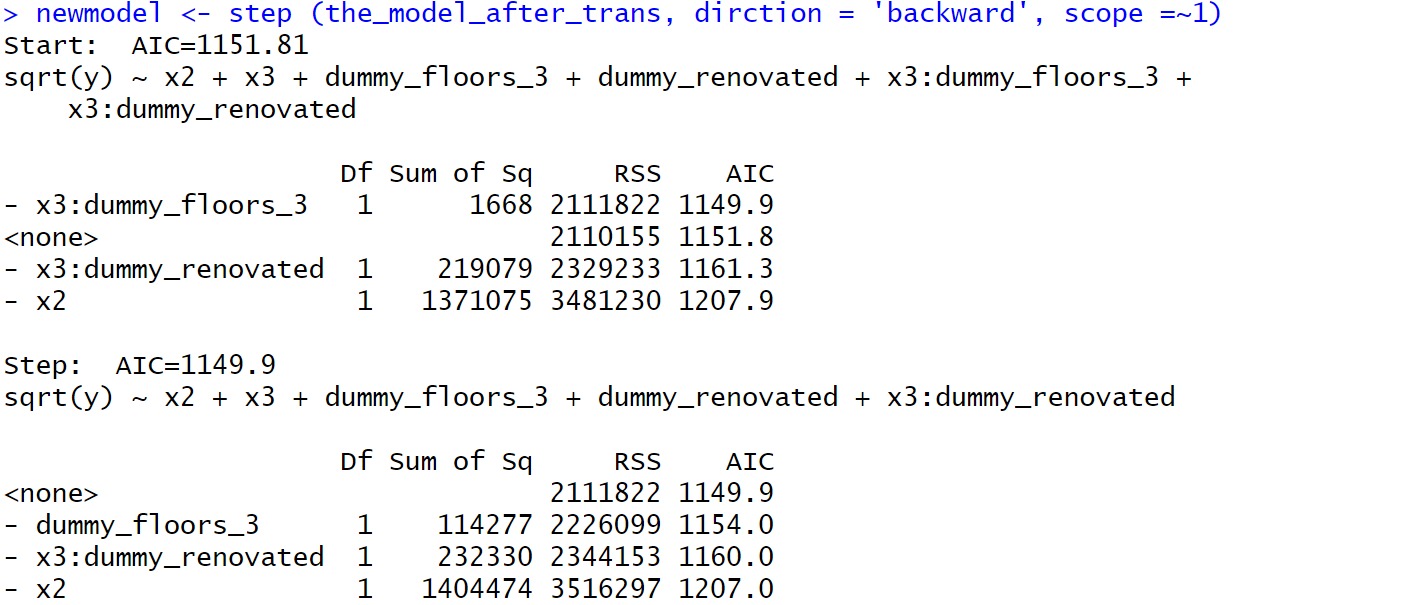
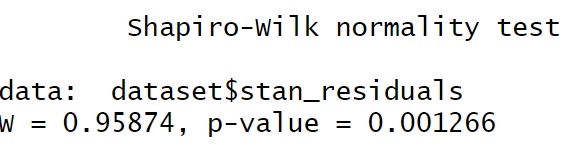
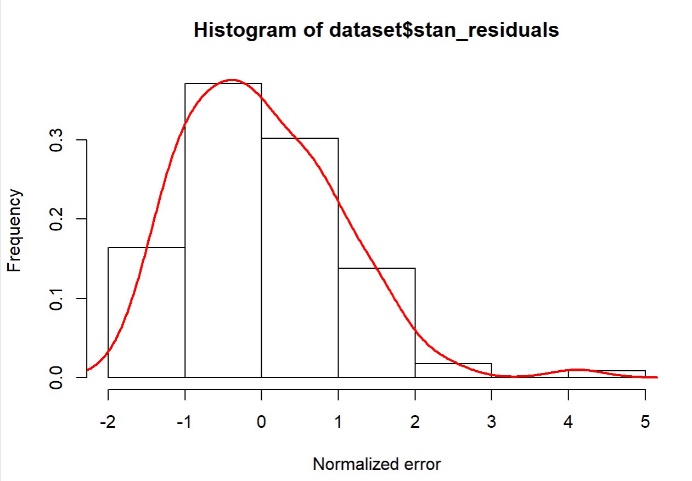
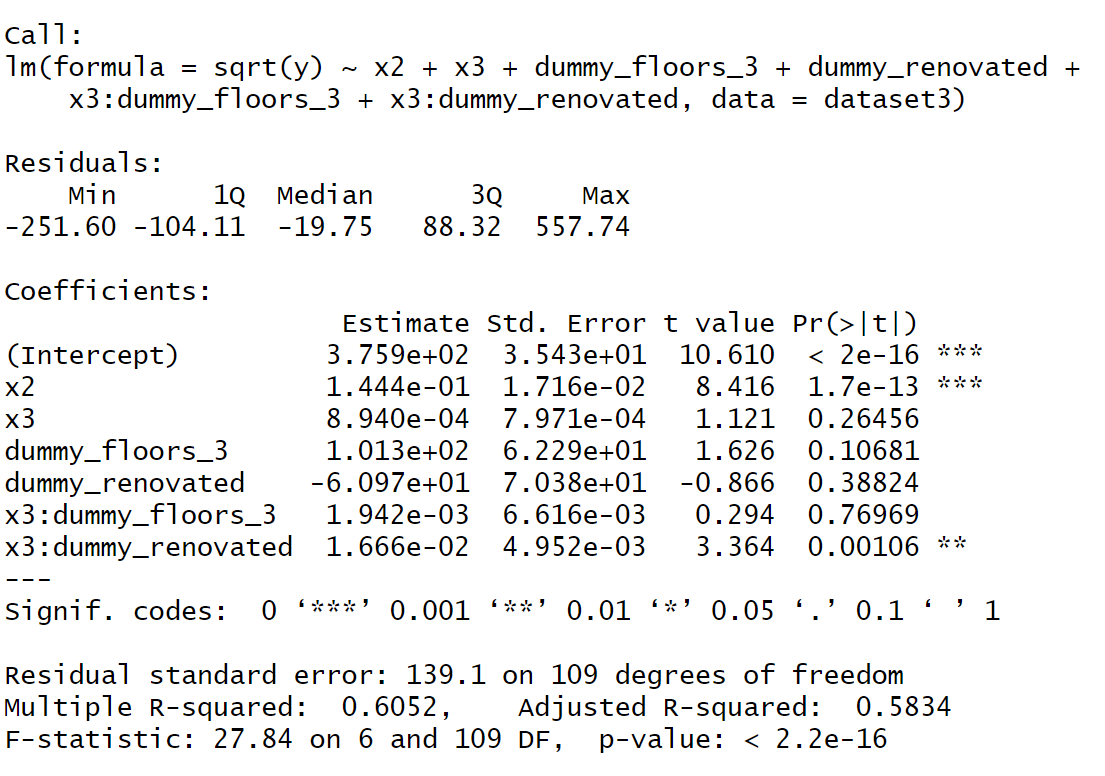


Graphical user interface, application

Description automatically generated

נספח4.2 כלל הנספחים והגרפים לחישוב ל-y^2



נספח4.3 כלל הנספחים והגרפים לחישוב ל- sqrt(y)

