

אקונומטריקה – עבודה מסכמת

מגישים:

יובל בקירוב – 209424803

איתן בקירוב - 209424811

חלק א' – הצגת שאלת המחקר וחשיבותה הכללית

1. מטרת מאמרם של מגהיר ופאלמה (2005) היא לבחון את השפעותיה של רפורמה חינוכית שהתרחשה בשוודיה בשנת 1940. מטרת הרפורמה הייתה לשפר את הנגישות לרמות השכלה גבוהות יותר על ידי הגדלת מספר שנות חינוך החובה ויישום תכנית לימודים אחידה בפריסה ארצית. המחברים מבקשים לחקור כיצד רפורמה זו השפיעה על התוצאות החינוכיות, תוך התמקדות באינטראקציה בין הרפורמה, יכולת התלמידים והרקע המשפחתי שלהם. המאמר נועד לספק ראיות אמפיריות על הקשר בין רפורמה חינוכית, מאפייני תלמידים והישגים חינוכיים.

ניתן לנסח את שאלת המחקר במאמרם של מגהיר ופאלמה כך :
"מהן ההשפעות של הרפורמה החינוכית, שמטרתה לשפר את הנגישות לרמות השכלה גבוהות יותר, על הישגי ההשכלה, וכיצד משתנות ההשפעות הללו בהתאם ליכולת התלמידים והרקע המשפחתי?"

2. המחברים מציעים לבדוק את שאלת המחקר על ידי שימוש בגישה מעין ניסויית המבוססת על יישום מדורג של הרפורמה החינוכית ברשויות השונות בשוודיה. הם משתמשים בנתוני סקר על אנשים עם נתוני לידה ייחודיים, במיוחד אלו שנולדו ב-1948 וב-1953. החוקרים בודקים עבור ילידי שני מחזורי לידה אלו עיריות בהן לא יושמה כלל הרפורמה, עיריות בהן יושמה הרפורמה רק בשנת 1953 (עיריות מתחלפות) ועיריות שבהן יישמו את הרפורמה כבר החל משנת 1948. על ידי השוואה בין שלוש הקבוצות הללו, הם שואפים להעריך את השפעת הרפורמה על הישגי ההשכלה.

המחברים משתמשים ב"הפרש ההפרשים" כאסטרטגיה האמפירית שלהם. גישת הפרש ההפרשים באה לידי ביטוי בהשוואת תוצאות בין אנשים שנולדו בשנים שונות, 1948 ו-1953, כדי להעריך את השפעות הרפורמה החינוכית. על ידי בחינת הפרש ההפרשים בתוצאות לאורך זמן בתוך שתי קבוצות אלו ועבור העיריות בהן יישמו או לא יישמו את הרפורמה, שואפים המחברים לבודד את ההשפעה הסיבתית של הרפורמה עצמה.

3. האסטרטגיה האמפירית של החוקרים מבוססת על מספר הנחות :
הנחת מגמות מקבילות : החוקרים מניחים שבהעדר הרפורמה המגמות בהישגי ההשכלה וההשתכרות היו דומות בין הקבוצות תחת הרפורמה ואלו שאינן תחת הרפורמה. הנחה זו מאפשרת להם לייחס כל הבדל בתוצאות בין שתי הקבוצות לרפורמה עצמה.

אקסוגניות של הקצאת רפורמה: החוקרים מניחים שהשיבוץ של יחידים לעיריות בהן ישנה רפורמה או אין רפורמות היה אקסוגני, כלומר לא היה קשור למאפיינים האישיים שעשויים להשפיע על תוצאות החינוך והרווחים. הנחה זו חיונית לביסוס קשר סיבתי בין הרפורמה לבין התוצאות הנצפות. אין השפעות זליגה: החוקרים מניחים כי השפעות הרפורמה על הישגי השכלה והשתכרות מונעות בעיקר משינויים בתוך העיריות שיישמו את הרפורמה וכי אין השפעות זליגה משמעותיות על אנשים המתגוררים בעיריות שכנות. הנחות אלו מהוות יסוד לאסטרטגיית הזיהוי של החוקרים ולפרשנות של ההשפעות המשוערות של הרפורמה. הפרות של הנחות אלו עלולות להכניס הטיות לתוצאות ולהגביל את תקפותן של הטענות הסיבתיות.

4. נראה שהגרף המוצג במאמר (FIGURE 1) מחזק את ההנחות שעליהן מתבססים החוקרים. הגרף משווה את הרווחים הריאליים לאורך זמן עבור קבוצות שונות בהתבסס על יישום הרפורמה בעיריות. החוקרים מנתחים את צמיחת הרווחים של אנשים עם אבות בעלי השכלה נמוכה, ומשווים את המתגוררים בעיריות שהטמיעו את הרפורמה (העיריות המטופלות תמיד והעיריות המתחלפות) לאלו המתגוררים בעיריות שלא יישמו את הרפורמה (שמעולם לא טופלו). הגרף מצביע על כך שאין הבדלים שיטתיים משמעותיים ברווחים בין הקבוצה שמעולם לא טופלה (עיגול) לבין המחליפים (משולש) בקבוצת 1948. לעומת זאת, העיריות בהן יושמה הרפורמה כבר בשנת 1948 נצפו גבוה יותר בגרף. זה מצביע על כך שניתן לייחס לרפורמה את ההבדלים שנצפו ברווחים. בנוסף, הגרף מראה כי עבור העיריות המתחלפות, לאחר יישום הרפורמה עבור קבוצת 1953, הרווחים עולים וחופפים לאלה שבקבוצה המטופלת תמיד. זה מצביע על כך שלרפורמה הייתה השפעה על הרווחים. עם זאת, החוקרים מציינים כי ברגע שהם מיישמים את אסטרטגיית הפרש-ההפרשים לצמיחה ברווחים לאורך זמן, ההבדל בצמיחת הרווחים הופך לבלתי מובהק סטטיסטית. מכאן עולה כי בעוד שהגרף מלמד בתחילה על השפעה חיובית של הרפורמה על צמיחת הרווחים, ניתוח נוסף מראה שההבדל אינו משמעותי. באופן כללי, הגרף מחזק את הנחות היסוד של החוקרים בכך שהוא מספק עדות חזותית להשפעת הרפורמה על הרווחים. עם זאת, הניתוח שלאחר מכן באמצעות הפרש-ההפרשים מוסיף ניואנסים לפרשנות של ממצאי הגרף.

5. תוצאות המאמר העיקריות הן: הישגים לימודיים: המחקר מגלה שהרפורמה במערכת החינוך השוודית הובילה לעלייה משמעותית בשיעור התלמידים שהגיעו לרמת דרישת החובה החדשה, במיוחד עבור ילדים עם אבות לא מלומדים ובעלי יכולת נמוכה. לימודים מעבר לרמת חובה: הרפורמה הביאה גם לעלייה בלימודים מעבר לרמת החובה החדשה לבעלי יכולת גבוהה יותר ואבות לא מלומדים. הדבר מצביע על כך שהרפורמה שיפרה את הגישה לחינוך מעבר לדרישת המינימום, ובכך עשויה להגביר את השינויים שחלו בהשוואה לדורות הקודמים בהיבט הכלכלי.

השתכרות בשוק העבודה: אנשים עם אבות לא מלומדים חוו עלייה משמעותית ברווחים בשוק העבודה בעקבות הרפורמה. עם זאת, ישנן אינדיקציות לכך שהרווחים ירדו עבור אנשים עם אבות בעלי כישורים גבוהים יותר, מה שמצביע על כך שביטול הסלקציה עשוי להפחית את איכות ההשכלה לקבוצה זו.

ניידות סלקטיבית: המחקר בוחן את ההשפעה הפוטנציאלית של ניידות סלקטיבית, שבה ייתכן שמשפחות עברו מגורים כדי לבחור בית ספר מתאים לילדיהם. המחברים מוצאים כי ניידות אינה מטה באופן משמעותי את התוצאות וכי השפעות הרפורמה נותרות חזקות גם כאשר לוקחים בחשבון ניידות סלקטיבית.

בסך הכל, מסקנת המאמר היא כי לרפורמה במערכת החינוך השוודית היו השפעות חשובות על הישגי השכלה והשתכרות, עם תוצאות חיוביות עבור חלק גדול מהאוכלוסייה. עם זאת, היו גם אינדיקציות לפשרות פוטנציאליות, במיוחד עבור אנשים עם אבות מלומדים יותר.

חלק ב' – ביצוע עבודה אמפירית

6. טבלת נתונים של פרטים שנולדו בשנת 1948 עבור 1985 :

1948	vars	changed	not_changed	Difference + sd_difference	p_value
1	llearn	6.7860	6.7391	0.046903 (0.018057)	0.00941723428
2	yearscho	11.469	10.9418	0.526768 (0.093587)	0.000000019057

7. הרשויות שיישמו את שיטת החינוך החדשה לא נבחרו באקראי ולכן התוצאות מפתיעות אותנו. מכיוון שהחוקרים בחרו בעצמם את הערים אנו מניחים שהם ניסו לנטרל כל הטיה אפשרית ולכן עצם קיום פער מובהק בין הערים מעיד על כך שהרפורמה השפיעה עליהן. שנתון 1948 לא היה אמור להיות מושפע מהמעבר לרפורמה שהיה צריך לקרות ב-1953. בתוצאות שקיבלנו ניתן לראות כי ילידי 1948 הושפעו מכך.

1953	vars	changed	not_changed	Difference + sd_difference	p_value
1	lnearn	6.7275	6.6493	0.078249 (0.01921)	0.0000469718382296443069 0607
2	yearschr	11.6781	10.7598	0.918221 (0.086919)	0.0000000000000000000000 00076612

בשונה מסעיף קודם, כאן נצפה לשינוי מובהק בין הקבוצות וזה אכן מתקיים. ברשויות שבהן הונהגה הרפורמה הייתה עלייה בשנות הלימוד הממוצעות ביחס לאלו שלא הונהגה בהן רפורמה. בנוסף, גם בשכר הייתה עלייה מובהקת, דבר המעיד על השפעת הרפורמה ברשויות הנבחרות.

9. ההשלכות מראות שיש צורך בשימוש באסטרטגיית "הפרש הפרשים", כי עבור שנות לידה שונות של הפרטים ישנן רשויות שונות בהן יושמה/לא יושמה הרפורמה. שיטת "הפרש הפרשים" תעזור לנו ליישב את הבעיה. עדויות אלה על ההפרשים מחזקות את אמינות האסטרטגיה.

10. בחרנו במשתנים המסבירים : hife, hiab3, treat_muni, yearschr כדי לבחון את הקשר בין שנות הלימוד של ילידי 1948 להכנסה שלהם, בשנת 1985. הרגרסיה המרובה היא :

$$\lnearn = b_0 + b_1 * yearschr + b_2 * treat_muni + b_3 * hiab3 + b_4 * hife + u$$

	אומדן	סטיית תקן	P Value
חותך	6.14844	0.03192	0.00...02
yearschr	0.05035	0.00282	0.00...02
treat_muni	0.01933	0.01732	0.26
hiab3	0.09084	0.01679	0.00..65
hife	-0.01662	0.02418	0.49

- Inearn - המשתנה המוסבר. משתנה זה יהיה בטרנספורמציות לוג על מנת לייצג את השפעת שנות הלימוד על השכר באחוזים, בדומה לנעשה בהוצאות ובשיעורי הבית.
- years - משתנה מסביר. משתנה זה מייצג שנות לימוד. בעזרת משתנה זה נוכל לאמוד את השפעת שנות הלימוד על ההכנסה.
- נוסיף משתנים מסבירים נוספים כדי לנטרל את ההשפעה של גורמים נוספים על השכר, שאינם שנות הלימוד :
- treat_muni - משתנה מסביר דמי. משתנה זה מייצג האם מדובר בעירייה שבה הונהגה הרפורמה החדשה ב - 1953 (1) או לא (0). בדומה למה שראינו בסעיף 6, הרפורמה עלולה להשפיע על הכנסת הפרטים.
- hiab3 - משתנה מסביר דמי. פרטים שציונם מעל הממוצע במבדקי היכולת (1) או לא (0). נרצה להפריד את השפעת היכולות במבדקים, דבר שיכול להשפיע על ההכנסה בעתיד. ציונים מעל הממוצע במבדקי היכולת מתואמים חיובית עם שנות השכלה והרווחים. השמטתו עלולה לגרום להטיה.
- hife - משתנה מסביר דמי. פרטים שאביהם בעל השכלה נוספת מעבר להשכלת החובה (1) או לא (0). השכלת האב יכולה להשפיע על הכנסת הפרטים. שנות השכלה של האב מתואמות חיובית עם שנות השכלת הבן והשמטת משתנה מסביר זה עלולה לגרום להטיה.
- b0 - חותך. לוג שכר בממוצע של פרטים שלא למדו (0 שנות לימוד), שלא הונהגה ברשותיהם רפורמה, שציונם במבדקי היכולת מתחת לממוצע ואלו שאביהם לא בעל השכלה מעל השכלת החובה הוא 6.14844. PV קטן מאוד לכן מקדם זה מובהק. לחותך זה אין משמעות כלכלית כי לא הגיוני שישנם פרטים עם 0 שנות השכלה עקב חובת לימודי חובה.
- b1 - עלייה של שנה אחת בשנות הלימוד של הפרט תעלה את שכרו ב-5.035 אחוזים בממוצע, בהינתן ששאר המשתנים המסבירים נותרים קבועים. PV קטן מאוד לכן מקדם זה מובהק. תוצאה זו הגיונית כיוון ששנות השכלה מעלות את הסיכוי של אדם להרוויח יותר.
- b2 - עבור פרטים שרשותיהם הנהיגו רפורמה בשנת 1953, תהיה עלייה בשכר של 1.933 אחוזים בממוצע, בהינתן ששאר המשתנים המסבירים קבועים. $PV = 0.26$ לכן מקדם זה אינו מובהק, כלומר, לא בהכרח משפיע על לוג השכר. קשה לדעת אם הגיוני או לא, תלוי בהשפעת הרפורמה.
- b3 - עבור פרטים שציונם מעל הממוצע בבדיקות היכולת, תהיה עלייה בשכר של 9.084 אחוזים בממוצע, בהינתן ששאר המשתנים המסבירים קבועים. PV קטן מאוד לכן מקדם זה מובהק. הגיוני שפרטים בעלי יכולות טובות יותר ישתכרו טוב יותר.

- b4 - עבור פרטים שאביהם בעל השכלה מעל השכלת החובה, תהיה ירידה בשכר של 1.662 אחוזים בממוצע, בהינתן ששאר המשתנים המסבירים קבועים. $PV = 0.49$ לכן מקדם זה אינו מובהק, כלומר, לא בהכרח משפיע על לוג השכר. אבות משכילים ככל הנראה ירוויחו יותר, ולכן בנים של אבות משכילים צפויים לפעול כמוהם ולבסוף להרוויח יותר.

$PV < 0.00...02$ לכן מודל רגרסיה זה מובהק, כלומר, אכן ניתן לתאר את לוג השכר בעזרת המשתנים המסבירים.

11. במודל של סעיף 10, האומד לשנות השכלה לא צפוי לתאר את הקשר הסיבתי שבין שנות השכלה לשכר, מכיוון שישנם משתנים נוספים שלא נלקחו בחשבון. על מנת ללמוד מהי ההשפעה הסיבתית של שנות השכלה על שכר היינו צריכים לבצע ניסוי מבוקר, כלומר, היינו צריכים להגריל את שנות הלימוד של הפרט. כך בעצם יש עוד גורמים אשר משפיעים על שכר הפרט (lnearn), אבל הם לא קשורים לכמות שנות הלימוד (years).
גורמים נוספים שמתאמים עם שנות לימודים ויכולים להשפיע על השכר: מצב סוציאקונומי, מצב בריאותי, ניסיון פרט וכו'...

12. נאמוד את הרגרסיה הבאה:

$$\text{lnearn} = b_0 + b_1 * \text{level_f} + b_2 * \text{treat_muni} + b_3 * \text{hiab3} + b_4 * \text{hife} + u$$

יצרנו משתנה פקטוריאלי level_f בעזרת משתנה מסביר level והרצנו מודל רגרסיה בעזרת כל המשתנים המסבירים מסעיף 10 עם משתנה level_f שמורכב ממשתני דמי לפי כל תת קטגוריה בתוכו.

	אומדן	סטיית תקן	P Value
חותך	6.55472	0.02063	0.00...02
level2	0.08566	0.03017	0.004543
level3	0.07307	0.02203	0.000917
level4	0.31109	0.02815	0.00...02
level5	0.21787	0.03062	0.00...125
level6	0.47193	0.02883	0.00...02
level7	0.69798	0.08627	0.00...72
treat_muni	0.01768	0.01730	0.30680
hiab3	0.08033	0.01682	0.00...182
hife	-0.02753	0.02416	0.25444

ניתן לרשום את הרגרסיה גם בדרך אחרת:

$$\text{lnearn} = b_0 + b_1 * \text{level_f2} + b_2 * \text{level_f3} + b_3 * \text{level_f4} + b_4 * \text{level_f5} + b_5 * \text{level_f6} + b_6 * \text{level_f7} + b_7 * \text{treat_muni} + b_8 * \text{hiab3} + b_9 * \text{hife} + u$$

חותך - לוג שכר בממוצע של פרטים שלמדו רק לימודי חובה שלא הונהגה ברשותיהם רפורמה, שציונם במבדקי היכולת מתחת לממוצע, ואלו שאביהם לא בעל השכלה מעל השכלת החובה הוא $PV = 6.5547$. קטן מאוד לכן מקדם זה מובהק.

b1 - עבור פרטים ברמת השכלה חטיבת ביניים לפני רפורמה, תהיה עלייה בשכר של 8.57 אחוזים בממוצע, בהינתן ששאר המשתנים המסבירים קבועים. PV קטן מאוד לכן מקדם זה מובהק.

b2 - עבור פרטים ברמת השכלה תיכון עד שנתיים כולל, תהיה עלייה בשכר של 7.31 אחוזים בממוצע, בהינתן ששאר המשתנים המסבירים קבועים. PV קטן מאוד לכן מקדם זה מובהק.

b3 - עבור פרטים ברמת השכלה תיכון מעל שנתיים, תהיה עלייה בשכר של 31.11 אחוזים בממוצע, בהינתן ששאר המשתנים המסבירים קבועים. PV קטן מאוד לכן מקדם זה מובהק.

b4 - עבור פרטים ברמת השכלה אחרי תיכון עד שנתיים כולל, תהיה עלייה בשכר של 21.79 אחוזים בממוצע, בהינתן ששאר המשתנים המסבירים קבועים. PV קטן מאוד לכן מקדם זה מובהק.

b5 - עבור פרטים ברמת השכלה אוניברסיטה או קולג', תהיה עלייה בשכר של 47.19 אחוזים בממוצע, בהינתן ששאר המשתנים המסבירים קבועים. PV קטן מאוד לכן מקדם זה מובהק.

b6 - עבור פרטים ברמת השכלה דוקטורט, תהיה עלייה בשכר של 69.8 אחוזים בממוצע, בהינתן ששאר המשתנים המסבירים קבועים. PV קטן מאוד לכן מקדם זה מובהק.

b7 - עבור פרטים שרשותיהם הנהיגו רפורמה בשנת 1953, תהיה עלייה בשכר של 1.77 אחוזים בממוצע, בהינתן ששאר המשתנים המסבירים קבועים. $PV = 0.30680$ לכן מקדם זה אינו מובהק, כלומר, לא בהכרח משפיע על לוג השכר.

b8 - עבור פרטים שציונם מעל הממוצע בבדיקות היכולת, תהיה עלייה בשכר של 8.03 אחוזים בממוצע, בהינתן ששאר המשתנים המסבירים קבועים. PV קטן מאוד לכן מקדם זה מובהק.

b9 - עבור פרטים שאביהם בעל השכלה מעל השכלת החובה, תהיה ירידה בשכר של 2.75 אחוזים בממוצע, בהינתן ששאר המשתנים המסבירים קבועים. $PV = 0.25444$ לכן מקדם זה אינו מובהק, כלומר, לא בהכרח משפיע על לוג השכר.

$PV < 0.00...02$ לכן מודל רגרסיה זה מובהק, כלומר, אכן ניתן לתאר את לוג השכר בעזרת המשתנים המסבירים.

13. ההנחה המובלעת בניסוח המודל בסעיף 10 היא שכל שנת לימודים משנה באופן זהה את ההכנסה. בנוסף, כל שנת לימודים בהכרח משנה את השכר.

ההנחה המובלעת במודל בסעיף 12 היא שבכל רמת השכלה (קטגוריה) טווח השנים אינו משפיע על השכר, כלומר, אדם שלמד שנה תואר ופרש ישנה את השכר באותו אופן כמו אדם שסיים תואר של 3 שנים.

מודל 12 מתאר טוב יותר את המתאם בין השכלה לבין שכר כיוון שכל קטגוריה בו משנה את השכר בכמות שונה בניגוד למודל שבסעיף 10, כך שאדם שעבר מתואר ראשון לשני יקבל תוספת שונה בשכר בהשוואה לאדם שסיים חטיבת ביניים ועלה לתיכון. נקבל תוצאות מדויקות יותר במודל שבסעיף 12, שמתחשב בסוג ההשכלה. כמו כן, $Adjusted-R^2$ של המודל בסעיף 12 (0.096) גבוה יותר משל המודל שבסעיף 10 (0.0884). כלומר, כל המשתנים המסבירים ברגרסיה של סעיף 12 מסבירים בצורה טובה יותר את השונות של לוג השכר בהשוואה למשתנים המסבירים של המודל בסעיף 10.

חלק ג' - נתרכז בשאלה האם הרפורמה במערכת החינוך השוודית השפיעה על ההשכלה והשכר של הפרטים?

14. Treat_muni – 1 אם העיר עברה לרפורמה החל משנת 1953 (אבל לא ב-1948), 0 אחרת.

D53 – 1 אם פרט נולד בשנת 1953, 0 אחרת.

treat_muni

d53		0	1	diff
	0	10.942	11.469	0.52677
	1	10.76	11.678	0.91822
	diff	-0.18195	0.20951	0.39145

ההשפעה הסיבתית של הרפורמה על שנות הלימוד היא 0.39145. כלומר, מעבר העירייה לרפורמה לילידי 1953 העלתה את שנות הלימוד ב-0.39145 שנים בממוצע.

15. נריץ את הרגרסיה: $years_{sch} = b_0 + b_1 * treat_muni + b_2 * d53 + b_3 * treat_muni:d53 + u$

	אומדן	סטיית תקן	P Value
חותך	10.9418	0.0732	0.00...02
treat_muni	0.5268	0.0858	0.00...84
d53	-0.1819	0.1135	0.1091
Treat_muni:d53	0.3915	0.1292	0.0024

B0 - חותך - קבוצת הבסיס - ממוצע שנות הלימוד של ילידי 1948 לפני הרפורמה הוא 10.9418 שנים $PV = 0.00...02$ קטן מ-0.05 כלומר מובהק.

B1 - פער שנות הלימוד בין קבוצת ילידי 1948 לפני הרפורמה לילידי 1948 אחרי הרפורמה הוא 0.5268 שנים בממוצע – $PV = 0.00...84$ קטן מ-0.05 כלומר מובהק.

B2 - פער שנות הלימוד בין ילידי 1953 לפני הרפורמה לילידי 1948 לפני הרפורמה הוא 0.1819 בממוצע. $PV = 0.1091$ גדול מ-0.05 כלומר אינו מובהק סטטיסטית.

B3 - הפרש ההפרשים - כמות שנות הלימוד בממוצע שהרפורמה הגדילה הוא 0.3915. $PV = 0.0024$ קטן מ-0.05 כלומר מובהק סטטיסטית.

מודל הפרש ההפרשים הוא בעצם אינטראקציה בין 2 משתני דמי (האם נולד בשנת 1953)*(האם חלה רפורמה) שהם $treat_muni * d53$

ניתן לחלק את הרגרסיה ל-4 מצבים אפשריים:

1- 1948 not treated: ($treat_muni = 0, d53 = 0$): $years_sch = B0$

2- 1948 treated: ($treat_muni = 1, d53 = 0$): $years_sch = B0 + B1$

3- 1953 not treated: ($treat_muni = 0, d53 = 1$): $years_sch = B0 + B2$

4- 1953 treated: ($treat_muni = 1, d53 = 1$): $years_sch = B0 + B1 + B2 + B3$

הפרש בין ילידי 1948 אחרי רפורמה לילידי 1948 לפני רפורמה הוא: $b1$

הפרש בין ילידי 1953 אחרי רפורמה לילידי 1953 לפני רפורמה הוא: $b1+b3$

הפרש ההפרשים בין ילידי 48 לפני ואחרי הרפורמה לבין ילידי 53 לפני ואחרי הרפורמה הוא: $b3$

treat_muni

d53			
		0	1
	0	B0	B0 + B1
	1	B0 + B2	B0 + B1 + B2 + B3
	diff	B2	B2 + B3
		B1	B1 + B3
		B3	

כפי שחישבנו קודם לכן, ההשפעה הסיבתית של הרפורמה על שנות הלימוד היא 0.39145 ($B3$). כלומר, מעבר העירייה לרפורמה לילידי 1953 העלתה את שנות הלימוד ב-0.39145 שנים בממוצע.

16. ההנחה המרכזית של המודל שאמדנו לצורך קבלה של השפעה סיבתית היא שלא חלו שינויים נוספים מלבד השינוי אותו נרצה לבחון, שבמקרה של המודל שלנו הוא הרפורמה בעיר. כלומר, השינויים בתוצאות הממוצעות בין שני מחזורי הילודה בערים שחלה בהם הרפורמה לערים שלא חלה בהם הרפורמה נותרו זהים. קבוצת הביקורת שלנו היא הערים שלא יושמה בהן הרפורמה, בעוד שקבוצת הטיפול שלנו היא הערים שכן יושמה בהן הרפורמה.

ההנחה סבירה, שכן בערים שונות, מלבד הרפורמה, לא צפויים לחול שינויים משמעותיים שעלולים לשנות את אופי העיר והפרטים בה. שינויים שיכולים לקרות באופן כזה שישפיע על הישגי הפרטים והצלחתם הם שינויים לא סבירים. לדוגמא: ראש עיר חדש שישפיע באופן לא סביר על הרכבי האוכלוסייה בעיר, שינוי יוצא דופן בתשתיות בעיר וכו'.

17. נבדוק אם ההנחה של שונות שווה סבירה במודל שאמדנו בסעיף 15:

לבדיקת שונות שווה נבצע מבחן BP:

נחשב את שאריות המודל $2^{\wedge}resids$, שיהיו המשתנה המוסבר במודל הרגרסיה של סעיף 15, ונבצע מבחן

סטטיסטי F למובהקות הרגרסיה כך ש:

$H_0: a_1 = a_2 = a_3 = 0$ (בריבוע - שונות שווה איקסים לא משפיעים על ריסידואלס)

$H_1: \text{else}$

תוצאות האמידה:

	אומדן	סטיית תקן	P Value
חותך	9.445	0.270	0.00...02
treat_muni	0.215	0.316	0.4960
d53	-1.922	0.418	0.00...43
Treat_muni:d53	-1.441	0.476	0.0025

ניתן לראות ש $PV = 0.02 \dots 0.0 < 0.05$, כלומר דוחים את H_0 במבחן מובהקות המודל - יש שונות שונה במודל.

ההנחה של שונות שווה לא סבירה במודל שאמדנו בסעיף 15.

בנוסף, ההשלכות מהתעלמות מבעיה זו הן שסטיות התקן של האומדים (שורש השונות) המופיעות בפלטים של

הרגרסיה של סעיף 15 ושל סעיף זה אינן נכונות ובנוסף לא ניתן להסתמך על מבחני ההשערות בפלט.

כלומר, התוצאה המספרית שיוצאת לנו של $1^{\wedge}V^{\wedge}ar(b)$ וגם של האומדי האחרים לשונות אינה נכונה!

תיקון White לשונות שונה:

ניתן לתוכנה פקודה אשר תתקן את שוניות האומדים, ונקבל:

	אומדן	סטיית תקן	P Value
חותך	10.9418	0.0792	0.00...02
treat_muni	0.5268	0.0931	0.00...16
d53	-0.1819	0.1153	0.1147
Treat_muni:d53	0.3915	0.1308	0.0028

תיקון ה-robust למעשה יגרום לתוכנה להתייחס לשונות הנכונה בהתאם לבעיית השונות השונה ולהשתמש

באומד מעט שונה לשונות ההפרעה האקראית.

לאחר התיקון:

$V^{\text{ar}}(b^i)$ לכל i תוקנה וניתן גם להשתמש בה למבחני t ולהסקת מסקנות. מתמטית - $V^{\text{ar}}(b^i)$ לכל i מוטה אך לפחות עקיבה (כי האומד לשונות ההפרעה האקראית מוטה ועקיב; לכן תיקון זה מתאים למדגמים גדולים).

$V^{\text{ar}}(b^i)$ לכל i אינה מינימלית - משום שהתיקון לא מתקן את בעיית השונות השונה של u וההנחות הקלאסיות לא מתקיימות (משפט גאוס מרקוב לא תקף על התיקון). מסיבה זו האומדים עצמם גם "לא יעילים" (כי הם לא בעלי שונות מינימלית).

18. נוסף למודל של סעיף 15 את המשתנים המפקחים: $hife$ ו- $female$ 3, $hiab$:

$$years_{sch} = b_0 + b_1 * treat_muni + b_2 * d53 + b_3 * treat_muni:d53 + female + hiab_3 + hife + u$$

$female$ - משתנה מסביר דמי. נשים (1) או גברים (0). ראינו במאמר כי אכן יש שונות בין המגדרים ולכן נרצה לבטל את ההשפעה השונה שיש לנשים מגברים.

$hiab_3$ - משתנה מסביר דמי. פרטים שציונם מעל הממוצע במבדקי היכולת (1) או לא (0). נרצה להפריד את השפעת היכולות במבדקים, דבר שיכול להשפיע על ההכנסה בעתיד. ציונים מעל הממוצע במבדקי היכולת מתואמים חיובית עם שנות השכלה והרווחים. השמטתו עלולה לגרום להטיה.

$hife$ - משתנה מסביר דמי. פרטים שאביהם בעל השכלה נוספת מעבר להשכלת החובה (1) או לא (0). השכלת האב יכולה להשפיע על הכנסת הפרטים. שנות השכלה של האב מתואמות חיובית עם שנות השכלת הבן והשמטת משתנה מסביר זה עלולה לגרום להטיה

	אומדן	סטיית תקן	P Value
חותך	9.7697	0.0723	0.00...02
treat_muni	0.3801	0.0760	0.00...59
d53	-0.2856	0.1006	0.0045
female	0.1891	0.0478	0.00007722
Hiab3	1.9733	0.0485	0.00...02
Hife	2.0015	0.0662	0.00...02
Treat_muni:d53	0.2065	0.1145	0.0712

נבדוק הומוסקדסטיות :

לבדיקת שונות שווה נבצע מבחן BP :

נחשב את שאריות המודל $resids2_{18}$, שיהיו המשתנה המוסבר במודל הרגרסיה של סעיף 18, ונבצע מבחן סטטיסטי F למובהקות הרגרסיה כך ש :

$H_0: a_1 = a_2 = a_3 = a_4 = a_5 = a_6 = 0$ (בריבוע - שונות שווה איקסים לא משפיעים על ריסידואלס)

$H_1: \text{else}$

ניתן לראות בפלט $PV < 0.05 = 0.02 \dots$, כלומר דוחים את H_0 במבחן מובהקות המודל - יש שונות שונה במודל.

תיקון White לשונות שונה :

ניתן לתוכנה פקודה אשר תתקן את שוניות האומדים, ונקבל :

	אומדן	סטיית תקן	P Value
חותך	9.7697	0.0757	0.00...02
treat_muni	0.3801	0.0826	0.00...43
d53	-0.2856	0.1044	0.0062
female	0.1891	0.0477	0.0000731
Hiab3	1.9733	0.0490	0.00...02
Hife	2.0015	0.0700	0.00...02
Treat_muni:d53	0.2065	0.1176	0.0791

בדומה לסעיף 17 תיקון white תיקן את האומדים לשונות של כל המשתנים המסבירים וניתן כעת להשתמש בהם למבחני t ולהסקת מסקנות.

19. לבדיקה האם המשתנים המסבירים שהוספנו בסעיף 18 מובהקים ביחד נעשה את מבחן F למובהקות. כלומר :

$H_0: B_4 = B_5 = B_6 = 0$

$H_1: \text{else}$

קיבלנו שסטטיסטי המבחן $F = 1019$.

תוצאת המבחן : $PV = 0.00 \dots > 0.05$

מסקנת המבחן : PV מדווח את מסקנת המבחן, כלומר נדחה את השערת האפס ב-95 אחוז לפחות. סה"כ, המשתנים המסבירים שהוספנו בסעיף 18 מובהקים ביחד, לכן היה מוצדק להוסיף אותם.

20. המקדם העיקרי שקיבלנו בסעיף 18 הוא המקדם של treat_muni:d 53 והוא 0.2065 והוא קטן יותר בהשוואה למקדם treat_muni:d 53 של המודל בסעיף 15 (0.3915). משמעותו של מקדם treat_muni:d 53 היא השינוי בממוצע של שנות הלימוד לאחר יישום הרפורמה בהינתן ששאר המשתנים קבועים. (זהו ההפרש של ההפרשים בין הפער של שנות לימוד בממוצע של לפני ואחרי הרפורמה של ילידי 1948 לפער בין הממוצע של לפני ואחרי הרפורמה עבור ילידי 1953). השינוי במקדם חל בעקבות הוספת משתנים מפקחים במודל של סעיף 18 שלקחתם בחשבון משפיעה על שנות הלימוד בממוצע. המודל של סעיף 18 מדויק יותר שכן מכיל משתנים מסבירים שיכולים להיות מתואמים עם שנות לימוד, לכן המודל בעל השפעה סיבתית מדויקת יותר. עם זאת, נשים לב שהאומדן החדש אינו מובהק ברמת ביטחון של 95% ($0.0791 > 0.05$) לכן אולי היה כדאי להישאר עם המודל הקודם.

21. ההנחה המרכזית של המודל שאמדנו לצורך קבלה של השפעה סיבתית, בדומה ל16, היא שלא חלו שינויים נוספים מלבד השינוי אותו נרצה לבחון, במקרה שלנו זוהי הרפורמה בעיר. ההנחה סבירה שכן בערים שונות מלבד הרפורמה לא צפויים לחול שינויים משמעותיים שעלולים לשנות את אופי העיר והפרטים בה. סבירות ההנחה של סעיף 18 חזקה יותר מהנחה של מודל 15, שנידונה בסעיף 16, שכן מודל זה מכיל יותר משתנים מפקחים ומבטל בסבירות גבוהה יותר פרמטרים שעלולים להיות מתואמים עם שנות לימוד.

22. המודל הנאמד :

$$\ln \text{earn}(it) = b_0 + b_1 * \text{treat_muni} + b_2 * d53 + b_3 * \text{treat_muni:d53} + b_4 * \text{female} + b_5 * \text{hiab3} + b_6 * \text{hife} + \text{year}(t) + u(it)$$

נשים לב שכעת אנו מסתכלים גם על שנת 1953 בנוסף ל1948. לכן, סביר להניח שיש שוני בין שנה אחת לשנייה ושימוש בנתוני פאנל (fixed effect) עבור שנים יפתור את הבעיה הזו. כלומר, אנו מפקחים על הזמן במודל זה.

	אומדן	סטיית תקן	P Value
treat_muni	0.03328	0.00239	0.00...25
d53	-0.11123	0.00313	0.00...11
Treat_muni:d53	0.04443	0.00474	0.00...13

מהטבלה ניתן להסיק ש :

- עבור פרטים שנולדו ב1948 השכר יהיה גבוה בכ3.328 אחוז אם נולדו בעירייה שהיה בה רפורמה לעומת ילידי 1948 אשר להם לא התבצעה רפורמה בעיר. נתון זה הגיוני גם כן כיוון שהרפורמה הגדילה את ההשכלה והשפיעה על השכר לטובה.
- עבור פרטים שנולדו ב1953 בעיריות שלא יושמה בהן הרפורמה השכר יהיה נמוך בכ11.23 אחוז בממוצע מאלו שנולדו ב-1948. זה מצב הגיוני כיוון שאלו פרטים בעלי פחות ניסיון.
- המקדם העיקרי במודל זה הוא מקדם האינטראקציה (treat_muni:d53) : עבור ילידי 1953 שנולדו בעיריות שבהן יושמה הרפורמה הייתה עלייה בשכר בכ4.44 יותר מאשר ילידי 1948 בעיריות אלו.

Dependent variable:						
	lnearn					
	(1)	(2)	(3)	(4)	(5)	(6)
treat_muni	0.023*** (0.004)	0.044*** (0.004)	0.104*** (0.014)	0.136*** (0.013)	0.010** (0.004)	0.030*** (0.004)
d53	-0.122*** (0.003)	-0.098*** (0.006)	-0.162*** (0.023)	-0.056*** (0.014)	-0.118*** (0.004)	-0.102*** (0.007)
treat_muni:d53	0.063*** (0.007)	0.024*** (0.005)	0.004 (0.028)	-0.080*** (0.013)	0.079*** (0.008)	0.041*** (0.005)
Observations	67,749	61,891	10,928	10,135	56,821	51,756
R2	0.214	0.287	0.209	0.270	0.204	0.286
Note:				*p<0.1; **p<0.05; ***p<0.01		

מודל 1 – כלל הגברים

מודל 2 – כלל הנשים

מודל 3 – הגברים עם אב בעל השכלה גבוהה

מודל 4 – הנשים עם אב בעל השכלה גבוהה

מודל 5 – הגברים עם אב בעל השכלה נמוכה

מודל 6 – הנשים עם אב בעל השכלה נמוכה

על מנת לבדוק את השפעת הרפורמה נתבונן על מקדם האינטראקציה כמו קודם בכל מודל :

- כלל הגברים – ניתן לראות כי השפעת הרפורמה מובהקת ואכן משמעותית בקרב כלל אוכלוסיית הגברים.
- כלל הנשים – גם כאן השפעה חיובית ומובהקת עבור כלל הנשים אך חלשה מעט ביחס לגברים.
- הגברים עם אב בעל השכלה גבוהה לעומת הנשים עם אב בעל השכלה גבוהה – ניתן לראות שבקבוצת הגברים השפעת הרפורמה חיובית לעומת השפעה שלילית לנשים – סיבה לכך יכולה להיות מצבן של הנשים לפני הרפורמה. אולי בגלל שלפני הרפורמה היה לנשים ייתרון ביחס לגברים ואחר העברת הרפורמה המצב התאזן קצת ביחס לגברים.
- הגברים עם אב בעל השכלה נמוכה והנשים עם אב בעל השכלה נמוכה – ניתן לראות שכאן קיימת השפעה יחסית שווה ומובהקת לשתי הקבוצות. זה מסתדר עם המציאות שאכן הרפורמה שיפרה את מצבם של אלו שהגיעו ממשפחות מוחלשות יותר ביחס לאלו שלא התבצעה רפורמה בעירייתם.

24. אחד האתגרים המרכזיים של המאמר הוא הצדקת ההנחה לקבלת אפקט סיבתי. על-מנת להצדיק את ההנחה לקבלת האפקט הסיבתי ולהבטיח שתהיה סבירה יותר, היינו מוסיפים עוד משתנים מסבירים כמו מוטיבציה, מצב סוציו-אקונומי של הפרט, שאיפות, הישגים עד כה, מצב חברתי שעלול להשפיע לטובה על הרצון ללמוד, מצב בריאותי שעלול להשפיע, רמת משמעת ועוד. נוסיף כל אלו כדי להבטיח קבוצות דומות ככל האפשר לפני יישום הרפורמה, ולזהות מאוחר יותר באופן ברור יותר את השפעות של הרפורמה. המטרה המרכזית של המאמר ושל עבודתנו כאן היא להבין באופן כמה שיותר מובהק את השפעות הרפורמה בעזרת הקשר הסיבתי שניסינו להבהיר.

```
> # setup -----
>
> # general setup
> rm(list=ls()) # clear environment
> gc() #cleans memory
      used (Mb) gc trigger (Mb) limit (Mb) max used (Mb)
Ncells 2758098 147.3  4757824 254.1    NA  4757824 254.1
Vcells 6553609  50.1  27098209 206.8   16384 27098209 206.8
> options(scipen=999) # tell R not to use Scientific notation
> options(digits = 5) # controls how many digits are printed by default
>
> # make sure to install before loading
> library(tidyverse)
> library(modelsummary) # produce regression's output to word.
> library(lmtest)
> library(sandwich)
> library(dplyr)
> library(car)
> library(plm)
> library(knitr)
> library(lfe)
> library(stargazer)
>
>
> # set\change working directory - make sure data is in same folder!
> path <- dirname(rstudioapi::getActiveDocumentContext())$path
> setwd(path)
>
> # reproducibility
> set.seed(1)
>
> options(na.action = "na.exclude")
>
> # PART B -----
> ## load the data
```



```

> mydata <- read.csv("term_paper_data_did.csv")
>
> # Question 6 -----
>
> # Filter the data for individuals born in 1948 and year 1985
> mydata_1948 <- subset(mydata, year == 85 & d53 == 0)
>
> # estimate model: llearn = b0 + b1 * treat_muni + u
> model1_1948 <- lm(llearn ~ treat_muni, data=mydata_1948)
> summary(model1_1948) # prints a summary of results

```

Call:

```
lm(formula = llearn ~ treat_muni, data = mydata_1948)
```

Residuals:

```

      Min       1Q   Median       3Q      Max
-6.786 -0.246  0.114  0.329  2.222

```

Coefficients:

```

              Estimate Std. Error t value      Pr(>|t|)
(Intercept)  6.7391     0.0154   437.2 <0.0000000000000002 *
treat_muni    0.0469     0.0181    2.6      0.0094 **

```

Signif. codes: 0 ‘*’ 0.001 ‘**’ 0.01 ‘.’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.598 on 5545 degrees of freedom

Multiple R-squared: 0.00122, Adjusted R-squared: 0.00104

F-statistic: 6.75 on 1 and 5545 DF, p-value: 0.00942

```
>
```

```

> # estimate model: yearsch = b0 + b1 * treat_muni + u
> model2_1948 <- lm(yearsch ~ treat_muni, data=mydata_1948)
> summary(model2_1948) # prints a summary of results

```

Call:

```
lm(formula = yearsch ~ treat_muni, data = mydata_1948)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.14	-1.85	-1.08	2.40	8.63

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.9418	0.0799	136.96	< 0.0000000000000002 *
treat_muni	0.5268	0.0936	5.63	0.000000019 *

Signif. codes: 0 ‘*’ 0.001 ‘**’ 0.01 ‘.’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.1 on 5545 degrees of freedom

Multiple R-squared: 0.00568, Adjusted R-squared: 0.0055

F-statistic: 31.7 on 1 and 5545 DF, p-value: 0.0000000191

>

>

> # Question 8 -----

>

> mydata_1953 <- subset(mydata, year == 85 & d53 == 1)

>

> # estimate model: llearn = b0 + b1 * treat_muni + u

> model1_1953 <- lm(llearn ~ treat_muni, data=mydata_1953)

> summary(model1_1953) # prints a summary of results

Call:

lm(formula = llearn ~ treat_muni, data = mydata_1953)

Residuals:

Min	1Q	Median	3Q	Max
-6.034	-0.226	0.129	0.325	1.653

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.6493	0.0173	385.14	< 0.0000000000000002 *

```
treat_muni 0.0782 0.0192 4.07 0.000047 *
```

```
---
```

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

Residual standard error: 0.565 on 5569 degrees of freedom

Multiple R-squared: 0.00297, Adjusted R-squared: 0.00279

F-statistic: 16.6 on 1 and 5569 DF, p-value: 0.000047

```
>
```

```
> # estimate model: yearsch = b0 + b1 * treat_muni + u
```

```
> model2_1953 <- lm(yearsch ~ treat_muni, data=mydata_1953)
```

```
> summary(model2_1953) # prints a summary of results
```

Call:

```
lm(formula = yearsch ~ treat_muni, data = mydata_1953)
```

Residuals:

```
Min 1Q Median 3Q Max
-4.35 -2.06 -1.29 2.19 8.81
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.7598 0.0781 137.7 <0.0000000000000002 *
treat_muni 0.9182 0.0869 10.6 <0.0000000000000002 *
```

```
---
```

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

Residual standard error: 2.56 on 5569 degrees of freedom

Multiple R-squared: 0.0196, Adjusted R-squared: 0.0195

F-statistic: 112 on 1 and 5569 DF, p-value: <0.0000000000000002

```
>
```

```
> # Question 10 -----
```

```
>
```

```
> # estimate model: llearn = b0 + b1 * yearsch + b2 * treat_muni + b3 * hiab3 + b4 * hife + u
```

```
> model10 <- lm(llearn ~ yearsch + treat_muni + hiab3 + hife, data=mydata_1948)
```

```
> summary(model10) # prints a summary of results
```

Call:

```
lm(formula = llearn ~ yearsch + treat_muni + hiab3 + hife, data = mydata_1948)
```

Residuals:

```
   Min    1Q  Median    3Q   Max
-6.652 -0.231  0.101  0.340  2.265
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.14844	0.03192	192.63	< 0.0000000000000002 *
years	0.05035	0.00282	17.83	< 0.0000000000000002 *
treat_muni	0.01933	0.01732	1.12	0.26
hiab3	0.09084	0.01679	5.41	0.000000065 *
hife	-0.01662	0.02418	-0.69	0.49

Signif. codes: 0 ‘*’ 0.001 ‘*’ 0.01 ‘.’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.571 on 5542 degrees of freedom

Multiple R-squared: 0.089, Adjusted R-squared: 0.0884

F-statistic: 135 on 4 and 5542 DF, p-value: <0.0000000000000002

```
>
```

```
> # Question 12 -----
```

```
>
```

```
> #Let's say we have a variable with many levels, e.g.: Level
```

```
> #We want to run regression with dummies for EACH ONE of the level types
```

```
> #The easiest approach is the factor variables approach
```

```
> class(mydata_1948$level)
```

```
[1] "integer"
```

```
> mydata_1948$level_f <- factor(mydata_1948$level)
```

```
> class(mydata_1948$level_f)
```

```
[1] "factor"
```

```
>
```

```
> # estimate model: llearn = b0 + b1 * level_f + b2 * treat_muni + b3 * hiab3 + b4 * hife + u
```

```
> model12 <-lm(lnearn ~ level_f + treat_muni + hiab3 + hife, data=mydata_1948)
> summary(model12) # prints a summary of results
```

Call:

```
lm(formula = lnearn ~ level_f + treat_muni + hiab3 + hife, data = mydata_1948)
```

Residuals:

```
Min    1Q  Median    3Q   Max
-6.658 -0.230  0.104  0.340  2.269
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.5547	0.0206	317.68	< 0.0000000000000002 *
level_f2	0.0857	0.0302	2.84	0.00454 **
level_f3	0.0731	0.0220	3.32	0.00092 *
level_f4	0.3111	0.0281	11.05	< 0.0000000000000002 *
level_f5	0.2179	0.0306	7.12	0.00000000000125854 *
level_f6	0.4719	0.0288	16.37	< 0.0000000000000002 *
level_f7	0.6980	0.0863	8.09	0.00000000000000072 *
treat_muni	0.0177	0.0173	1.02	0.30680
hiab3	0.0803	0.0168	4.78	0.00000182847252056 *
hife	-0.0275	0.0242	-1.14	0.25444

Signif. codes: 0 ‘*’ 0.001 ‘**’ 0.01 ‘.’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.569 on 5537 degrees of freedom

Multiple R-squared: 0.0975, Adjusted R-squared: 0.096

F-statistic: 66.5 on 9 and 5537 DF, p-value: <0.0000000000000002

```
>
```

```
> # PART C -----
```

```
> # Question 14 -----
```

```
>
```

```
> mydata_dd <- subset(mydata, year == 85)
```

```
>
```

```

> first_mean_48 <- mean(mydata_dd$yearschr[mydata_dd$treat_muni == 0 & mydata_dd$d53 ==
0])
> first_mean_48
[1] 10.942
> second_mean_48 <- mean(mydata_dd$yearschr[mydata_dd$treat_muni == 1 & mydata_dd$d53
== 0])
> second_mean_48
[1] 11.469
> first_mean_53 <- mean(mydata_dd$yearschr[mydata_dd$treat_muni == 0 & mydata_dd$d53 ==
1])
> first_mean_53
[1] 10.76
> second_mean_53 <- mean(mydata_dd$yearschr[mydata_dd$treat_muni == 1 & mydata_dd$d53
== 1])
> second_mean_53
[1] 11.678
>
> diff_48_1 <- second_mean_48 - first_mean_48
> diff_48_1
[1] 0.52677
> diff_53_1 <- second_mean_53 - first_mean_53
> diff_53_1
[1] 0.91822
> dd_14_1 <- diff_53_1 - diff_48_1
> dd_14_1
[1] 0.39145
>
> diff_48_2 <- first_mean_53 - first_mean_48
> diff_48_2
[1] -0.18195
> diff_53_2 <- second_mean_53 - second_mean_48
> diff_53_2
[1] 0.20951
> dd_14_2 <- diff_53_2 - diff_48_2
> dd_14_2
[1] 0.39145

```

```

>
>
>
> # Question 15 -----
>
> # estimate model: yearsch = b0 + b1 * treat_muni + b2 * d53 + b3 * treat_muni: d53 + u
> model_dd <- lm(yearsch ~ treat_muni + d53 + treat_muni: d53, data=mydata_dd)
> summary(model_dd) # prints a summary of results

```

Call:

```
lm(formula = yearsch ~ treat_muni + d53 + treat_muni: d53, data = mydata_dd)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.35	-1.85	-1.08	2.19	8.81

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.9418	0.0732	149.45	< 0.0000000000000002 *
treat_muni	0.5268	0.0858	6.14	0.000000000084 *
d53	-0.1819	0.1135	-1.60	0.1091
treat_muni: d53	0.3915	0.1292	3.03	0.0024 **

Signif. codes: 0 ‘*’ 0.001 ‘**’ 0.01 ‘.’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.84 on 11114 degrees of freedom

Multiple R-squared: 0.0123, Adjusted R-squared: 0.0121

F-statistic: 46.3 on 3 and 11114 DF, p-value: <0.0000000000000002

```

>
>
> # Question 17 -----
>
> #generate residuals
> mydata_dd$resids <- model_dd$residuals
> mydata_dd$resids2 <- mydata_dd$resids^2

```

```

>
> #generate yhat
> mydata_dd$yhat <- model_dd$fitted.values
>
> plot(y = mydata_dd$resids, x = mydata_dd$treat_muni)
> plot(y = mydata_dd$resids, x = mydata_dd$d53)
> plot(y = mydata_dd$resids, x = mydata_dd$yhat)
> plot(y = mydata_dd$resids^2, x = mydata_dd$yhat)
>
> #BP Test
> # estimate model: resids2 = b0 + b1 * treat_muni + b2 * d53 + b3 * treat_muni:d53 + u
> model_BP <- lm(resids2 ~ treat_muni + d53 + treat_muni:d53, data=mydata_dd)
> summary(model_BP) # prints a summary of results

```

Call:

```
lm(formula = resids2 ~ treat_muni + d53 + treat_muni:d53, data = mydata_dd)
```

Residuals:

```

  Min    1Q  Median    3Q   Max
-9.14 -7.39 -3.89  3.60 70.10

```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.445	0.270	35.04	< 0.0000000000000002 *
treat_muni	0.215	0.316	0.68	0.4960
d53	-1.922	0.418	-4.60	0.0000043 *
treat_muni:d53	-1.441	0.476	-3.03	0.0025 **

Signif. codes: 0 ‘*’ 0.001 ‘**’ 0.01 ‘.’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.5 on 11114 degrees of freedom

Multiple R-squared: 0.0221, Adjusted R-squared: 0.0219

F-statistic: 83.9 on 3 and 11114 DF, p-value: <0.0000000000000002

```

>

```

```
> #re-estimate the model with white's robust standard errors
```



```
> coeftest(model_dd, vcov = vcovHC(model_dd, type="HC1"))
```

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.9418	0.0792	138.09	< 0.0000000000000002 *
treat_muni	0.5268	0.0931	5.66	0.000000016 *
d53	-0.1819	0.1153	-1.58	0.1147
treat_muni: d53	0.3915	0.1308	2.99	0.0028 **

Signif. codes: 0 ‘*’ 0.001 ‘**’ 0.01 ‘.’ 0.05 ‘.’ 0.1 ‘ ’ 1

```
>
```

```
>
```

```
> # Question 18 -----
```

```
>
```

```
> # estimate model: yearsch = b0 + b1 * treat_muni + b2 * d53 + b3 * treat_muni: d53 + b4 * female
+ b5 * hiab3 + b6 * hife + u
```

```
> model_dd_18 <- lm(yearsch ~ treat_muni + d53 + treat_muni: d53 + female + hiab3 + hife,
data=mydata_dd)
```

```
> summary(model_dd_18) # prints a summary of results
```

Call:

```
lm(formula = yearsch ~ treat_muni + d53 + treat_muni: d53 + female +
    hiab3 + hife, data = mydata_dd)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.984	-1.871	-0.043	1.637	10.086

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.7697	0.0723	135.10	< 0.0000000000000002 *
treat_muni	0.3801	0.0760	5.00	0.000000059 *
d53	-0.2856	0.1006	-2.84	0.0045 **
female	0.1891	0.0478	3.95	0.00007722 *

```

hiab3      1.9733   0.0485  40.67 < 0.0000000000000002 *
hife       2.0015   0.0662  30.25 < 0.0000000000000002 *
treat_muni:d53 0.2065   0.1145   1.80      0.0712 .

```

Signif. codes: 0 ‘*’ 0.001 ‘*’ 0.01 ‘’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.52 on 11111 degrees of freedom

Multiple R-squared: 0.225, Adjusted R-squared: 0.225

F-statistic: 539 on 6 and 11111 DF, p-value: <0.0000000000000002

>

> #generate residuals

> mydata_dd\$resids_18 <- model_dd_18\$residuals

> mydata_dd\$resids2_18 <- mydata_dd\$resids_18^2

>

> #BP Test

> # estimate model: resids2_18 = b0 + b1 * treat_muni + b2 * d53 + b3 * treat_muni: d53 + b4 * female + b5 * hiab3 + b6 * hife + u

> model_BP_18 <- lm(resids2_18 ~ treat_muni + d53 + treat_muni: d53 + female + hiab3 + hife, data=mydata_dd)

> summary(model_BP_18) # prints a summary of results

Call:

lm(formula = resids2_18 ~ treat_muni + d53 + treat_muni: d53 + female + hiab3 + hife, data = mydata_dd)

Residuals:

```

      Min      1Q  Median      3Q      Max
-10.02  -5.52  -2.82   1.20   96.88

```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.198	0.248	24.95	< 0.0000000000000002 *
treat_muni	-0.219	0.261	-0.84	0.40140
d53	-1.353	0.346	-3.91	0.000091 *
female	-0.452	0.164	-2.75	0.00589 **

```

hiab3      3.213   0.167  19.28 < 0.0000000000000002 *
hife       0.627   0.227   2.76      0.00583 **
treat_muni:d53 -1.334   0.393  -3.39      0.00069 *

```

Signif. codes: 0 ‘*’ 0.001 ‘**’ 0.01 ‘.’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.64 on 11111 degrees of freedom

Multiple R-squared: 0.0525, Adjusted R-squared: 0.052

F-statistic: 103 on 6 and 11111 DF, p-value: <0.0000000000000002

>

> #re-estimate the model with white's robust standard errors

> coeftest(model_dd_18, vcov = vcovHC(model_dd_18, type="HC1"))

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.7697	0.0757	129.11	< 0.0000000000000002 *
treat_muni	0.3801	0.0826	4.60	0.0000043 *
d53	-0.2856	0.1044	-2.73	0.0062 **
female	0.1891	0.0477	3.97	0.0000731 *
hiab3	1.9733	0.0490	40.23	< 0.0000000000000002 *
hife	2.0015	0.0700	28.58	< 0.0000000000000002 *
treat_muni:d53	0.2065	0.1176	1.76	0.0791 .

Signif. codes: 0 ‘*’ 0.001 ‘**’ 0.01 ‘.’ 0.05 ‘.’ 0.1 ‘ ’ 1

>

> # Question 19 -----

>

> linearHypothesis(model_dd_18, c("hiab3= 0", "hife=0", "female=0"))

Linear hypothesis test

Hypothesis:

hiab3 = 0

hife = 0

female = 0

Model 1: restricted model

Model 2: $\text{years} \sim \text{treat_muni} + \text{d53} + \text{treat_muni}:\text{d53} + \text{female} + \text{hiab3} + \text{hife}$

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	11114	89659				
2	11111	70309	3	19350	1019	<0.0000000000000002 *

Signif. codes: 0 ‘*’ 0.001 ‘*’ 0.01 ‘’ 0.05 ‘.’ 0.1 ‘ ’ 1

>

> # Question 22 -----

>

> #fixed effect by female + hiab3 + hife + year, cluster by year

> model_dd_22 <- felm(lnearn ~ treat_muni*d53|female + hiab3 + hife + year|0|year,
data=mydata)

> summary(model_dd_22) # prints a summary of results

Call:

felm(formula = lnearn ~ treat_muni * d53|female + hiab3 + hife + year|0|year, data =
mydata)

Residuals:

Min	1Q	Median	3Q	Max
-7.497	-0.160	0.062	0.258	2.794

Coefficients:

	Estimate	Cluster s.e.	t value	Pr(> t)
treat_muni	0.03328	0.00239	13.90	0.0000000252501 *
d53	-0.11123	0.00313	-35.53	0.0000000000011 *
treat_muni:d53	0.04443	0.00474	9.38	0.0000013923143 *

Signif. codes: 0 ‘*’ 0.001 ‘*’ 0.01 ‘’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.485 on 129622 degrees of freedom

Multiple R-squared(full model): 0.319 Adjusted R-squared: 0.319

Multiple R-squared(proj model): 0.00797 Adjusted R-squared: 0.00784

F-statistic(full model, iid): 3.57e+03 on 17 and 129622 DF, p-value: <0.0000000000000002

F-statistic(proj model): 539 on 3 and 11 DF, p-value: 0.00000000000327

* Standard errors may be too high due to more than 2 groups and exactDOF=FALSE

>

> # Question 23 -----

>

> males <- subset(mydata, female == 0)

> females <- subset(mydata, female == 1)

>

> males_high_edu_father <- subset(males, hife == 1)

> females_high_edu_father <- subset(females, hife == 1)

>

> males_low_edu_father <- subset(males, hife == 0)

> females_low_edu_father <- subset(females, hife == 0)

>

>

> model_dd_23_males <- felm(lnearn ~ treat_muni*d53|hiab3 + hife + year|0|year, data=males)

> model_dd_23_females <- felm(lnearn ~ treat_muni*d53|hiab3 + hife + year|0|year,
data=females)

>

> model_dd_23_males_hef <- felm(lnearn ~ treat_muni*d53|hiab3 + year|0|year,
data=males_high_edu_father)

> model_dd_23_females_hef <- felm(lnearn ~ treat_muni*d53|hiab3 + year|0|year,
data=females_high_edu_father)

>

> model_dd_23_males_lef <- felm(lnearn ~ treat_muni*d53|hiab3 + year|0|year,
data=males_low_edu_father)

> model_dd_23_females_lef <- felm(lnearn ~ treat_muni*d53|hiab3 + year|0|year,
data=females_low_edu_father)

>

>

```
> model_list <- list(model_dd_23_males, model_dd_23_females, model_dd_23_males_hef,
model_dd_23_females_hef, model_dd_23_males_lef, model_dd_23_females_lef)
>
> stargazer(model_list, type = 'text', no.space=T, keep.stat = c('n', 'rsq'))
```

```
=====
=====
```

Dependent variable:

```
-----
                lnearn
      (1)   (2)   (3)   (4)   (5)   (6)
-----
treat_muni  0.023** 0.044* 0.104* 0.136* 0.010* 0.030***
            (0.004) (0.004) (0.014) (0.013) (0.004) (0.004)
d53         -0.122** -0.098* -0.162* -0.056* -0.118* -0.102**
            (0.003) (0.006) (0.023) (0.014) (0.004) (0.007)
treat_muni:d53 0.063** 0.024* 0.004 -0.080* 0.079* 0.041**
            (0.007) (0.005) (0.028) (0.013) (0.008) (0.005)
-----
Observations  67,749  61,891  10,928  10,135  56,821  51,756
R2            0.214   0.287   0.209   0.270   0.204   0.286
```

```
=====
=====
```

Note: p<0.1; **p<0.05; ***p<0.01

```
>
> stargazer_cols = c("(1) - males", "(2) - females", "(3) - males_high_edu_father", "(4) -
females_high_edu_father", "(5) - males_low_edu_father", "(6) - females_low_edu_father")
> print(stargazer_cols)
[1] "(1) - males"          "(2) - females"        "(3) - males_high_edu_father"
[4] "(4) - females_high_edu_father" "(5) - males_low_edu_father" "(6) -
females_low_edu_father"
```