

סקירת הנתונים הזמינים

סדנת סקר מיומנויות
הלשכה המרכזית לסטטיסטיקה, ירושלים
ינואר 2016

במצגת זו

- הקבצים הזמינים כיום
- נתונים לשימוש ציבורי
- ✓ PIAAC Data Explorer (PDX)
- ✓ PIAAC Public-use files (PUF)
- ✓ פורמטים שבהם קיים ה-PUF
- הרשומות הכלולות
- ✓ הבדלים בין ה-PUF לקובץ ה-PDX
- המשתנים הכלולים
- ✓ הבדלים בין ה-PUF לקובץ ה-PDX
- ✓ משתני מפתח לניתוח בסדנה
- אופן הצגת ערכים חסרים
- שחרור הנתונים על ישראל

הקבצים הזמינים כיום

הקבצים הזמינים כיום

- הנתונים הזמינים לשימוש כיום הם הנתונים הבין-לאומיים של מדינות הגל הראשון
- אלה יוצגו בפירוט במצגת זו, וישמשו לעבודה בסדנה
- נתוני הגל השני יהיו זמינים החל ב-28.6

נתונים לשימוש ציבורי

בסיסי נתונים לשימוש ציבורי

- מבוססים על הנתונים הלאומיים המלאים
- כוללים את מדינות הגל הראשון
- שתי גרסאות (כמעט זהות)
- ✓ נתונים שבבסיס מחולל הנתונים **ETS PIAAC Data Explorer**
- ✓ קבצים לשימוש ציבורי (**PUF**) ובהם רשומות פרט

בסיסי נתונים זמינים

- גרסת ה-PUF שהותקנה לתרגיל ב-SPSS כוללת 23 מ-24 מדינות הגל הראשון
- ללא אוסטרליה (ניתן לקבל בנפרד)
- גרסאות מלאות יותר ניתן לבקש ישירות מהמדינות

מסמכים זמינים

- מסגרת תאורטית ותיאור רמות הביצוע
- שאלון הרקע
- דוחות בין-לאומיים, דוחות לאומיים
- הדוח הטכני
- משתנים וערכים של ה-PUF
- קוד לחישוב המשתנים המחושבים
- מדריכים לעבודה ב-SAS וב-STATA
- Webpackages – מכלול הלוחות שפורסמו בדוחות הבין-לאומיים
- Compendia – התפלגות בסיסית של המשתנים, לתיקוף
- <http://www.oecd.org/site/piaac/>
- תחת Public Data & Analysis ו-Publications ✓

פורמטים שבהם קיים ה-PUF

- הקבצים מגיעים בשני פורמטים
- SPSS (.sav) לגרסה 11 ומעלה
- ✓ שדות אלפא-נומריים מקודדים בפורמט Unicode (UTF8)
- ✓ מילון מפורט
- סוגים ופורמטים של המשתנים
- תוויות למשתנים
- תוויות לערכים (כולל לערכים החסרים)
- הגדרות לערכים החסרים
- הקבצות לערכים
- ✓ הערכים החסרים מקבלים ערך נומרי (ועל כך בהמשך)

פורמטים שבהם קיים ה-PUF

• SAS (.sas7bdat)

- ✓ שדות מקודדים בפורמט Unicode (UTF8)
- ✓ לקובץ הנתונים .sas7bdat מצורפת תכנית SAS עם הפורמטים למשתנים
- SAS אינו שומר את הפורמטים בקובץ עצמו באופן קבוע
- ✓ הערכים החסרים מוצגים ב-SAS בערכי מיוחדים (למשל ".V")

הרשומות הכלולות (ואלה שלא)

הרשומות הכלולות

- כדי להיכלל בקבצים לשימוש ציבורי, הרשומות צריכות...
 - ✓ לענות על הגדרת האוכלוסייה הבין-לאומית (16-65)
 - ✓ להיחשב ל"מקרים מלאים" שעברו תיקוף וקיבלו משקל
- נכללים מקרים עם מידע חלקי או מינימלי
 - ✓ מקרי אי-השבה הקשורה באוריינות (LRNR)
 - ✓ מקרי הפסקת ריאיון שבהם נאסף מידע מספיק למודל הפסיכומטרי

רשומות שאינן כלולות

- תוספות אוכלוסייה מקומיות
- מקרי סירוב ואי־השבה מסוג אחר ("נפל")
- זיופים
- משיבים ללא ערכים בפרטי החובה: גיל, מין, רמת השלכה ותכונות כוח העבודה
- מקרי LRNR ללא ערכים בשדות מין וגיל
- מקרים ספורים של חריגות או מקור/איכות לא ברורים

הרשומות הכלולות: הבדלים בין ה-PUF לקובץ ה-PDX

- מספר הרשומות בדרך כלל זהה
- מלבד ה-PUF של קנדה
- ✓ לא אמור להשפיע על הממצאים

המשתנים הכלולים

הבדלי מפתח בין ה-PUF לקובץ ה-PDX

- יש הבדלים במערך המשתנים בין שני המקורות
- ✓ יש משתנים שאינם שימושיים ל-PDX, אך נכללים ב-PUF לניתוח
- בכל PUF יש 1,328 משתנים
- רק 575 מתוכם כלולים ב-PDX
- רוב המשתנים הכלולים ב-PUF בלבד נוגעים לפריטים קוגניטיביים בודדים, ולמידע על תהליך הריאיון

משתנים מרכזיים שיש ב-PUF אך לא ב-PDX

שמות המשתנים	תיאור	קטגוריה
COMPUTEREXPERIENCE, NATIVESPEAKER, EDLEVEL3	משתנים שמחושבים אוטומטית בזמן הריאיון כדי לקבוע את המשך התזרים	משתני שאלון הרקע מחושבים לצורכי תזרים
CBA_CORE_STAGE*_SC ORE, CORESTAGE*_PASS, RANDOM_CBA_*, CBA_START, PPC_SCORE, RANDOM_PP	ציוני פרק הגרעין, ציון עובר, הקצאה אקראית של מודולים כדי שנקבעים בעת הריאיון	ציונים קוגניטיביים, הקצאת מספרים אקראית
ZZ*	תאור של המראיין את האינטראקציה עם המרואיין	מודול תצפית

משתני מפתח לניתוח

- משתני שאלון הרקע
 - ✓ חינוך והשכלה (B)
 - ✓ תעסוקה והיסטוריה תעסוקתית (C)
 - ✓ העבודה הנוכחית – כולל הכנסות (D)
 - ✓ העבודה האחרונה (E)
 - ✓ שימוש במיומנויות בעבודה וביום-יום (F, G, H)
 - ✓ רקע כללי (I, J)
- ציוני ביצוע שלוש המיומנויות הנמדדות (plausible values)

משתנים שאינם כלולים

- כל ההתאמות והתוספות המקומיות בשאלון הרקע
- מזהים של משיבים, סוקרים, סמלנים, מקלידי נתונים
- משתנים ארעיים (דגימה, משקל וכו')
- משתני טקסט חופשי הקיימים בגרסה מקודדת
- ערכי הציונים לפני תקנון בין-לאומי

השמטת נתונים

- בסיסי הנתונים ל-PUF ול-PDX עברו השמטות ברמת התא או העמודה למדינות מסוימות שדרשו זאת
- ברוב המקרים מדובר ב...
 - ✓ גיל מפורט
 - ✓ מידע מפורט על שפה, ארץ לידה או אזור
 - ✓ מידע מפורט על השכלה
 - ✓ מידע מפורט על משלח יד (ISCO) וענף כלכלי (ISIC) – למשל ברמה של 4 או 3 ספרות
 - ✓ הכנסה מפורטת
- ב-PUF ישנן יותר השמטות מאשר ב-PDX

תוצאות ההשמטות

- משתמשים עלולים שלא להצליח לשחזר במלואם לוחות או נתונים שפורסמו בדוחות הבין-לאומיים
- במקרה כזה אפשר לבדוק אם **PDX** מאפשר ליווחים שקובץ **PUF** מוגבל אינו מאפשר
- למשתנים מסוימים הופקו גרסאות "גסות" יותר (סיומת C_) כדי לאפשר שחרור המשתנה לשימוש ציבורי
- ✓ השתמשו באלה כאשר המטרה היא לכלול מספר גדול של מדינות
- ✓ גם בגרסתם ה"גסה", הממצאים הם תקפים סטטיסטית ומהותית
- גם כאן יש אפשרות לקבל נתונים מפורטים יותר על ידי פנייה למדינות

אופן הצגת ערכים חסרים

סוגי ערכים חסרים

- שלושה סוגים עיקריים
 - ✓ מובנים – היבט חשוב שיש לזכור בניתוח נתוני PIAAC!
 - ✓ כתוצאה מתהליך ההשבה
 - ✓ כתוצאה מהתאמה מקומית, לוגיסטיקה של הסקר, עיבוד או ניתוח
- לא נעשו זקיפות לערכים מלבד זקיפת הכנסות

ערכים חסרים המובנים בסקר

- לא רואיינו משיבים עם קשיי השבה הקשורים לאוריינות (LRNR)
- כתוצאה מכך, לא נזקפו להם ציוני ביצוע כלשהם
- מדינות מסוימות לא השתתפו בחלקים מהסקר
 - ✓ פתרון בעיות בסביבה טכנולוגית (קפריסין, צרפת, איטליה, ספרד)
 - ✓ רכיבי קריאה (פינלנד, צרפת, יפן)
- פריטים שונים בשאלון הרקע הוצגו מלכתחילה רק לתת-אוכלוסיות מסוימות (valid skip)

ערכים חסרים המובנים בסקר

- משיבים ללא ידע במחשב, שנכשלו במודול הגרעין או שסירבו להיבחן במחשב – נבחנו בנייר
- ✓ המשיבים בנייר לא נבחנו בפתרון בעיות בסביבה טכנולוגית, ולכן לא קיבלו ציון במיומנות זו
- המודולים הוקצו על בסיס אקראי, ובמבחן הממוחשב גם על פי הביצוע

ערכים חסרים כתוצאה מתהליך ההשבה

- משיבים מסוימים מפסיקים את הריאיון עקב שאלות רגישות, עייפות, חוסר בזמן...
- משיב עשוי לסרב באופן מפורש ("סירוב") או לא לדעת את התשובה ("לא ידוע")

ערכים חסרים מהתאמה מקומית, לוגיסטיקה, עיבוד

- שגיאות תזרים בהתאמות מקומיות לשאלון הרקע (נדיר)
- משתנים או ערכים מסוימים עשויים להיות מוסתרים עקב בעיות סודיות
- ערכים שקודדו או הוקלדו בטעות בבסיסי הנתונים המקומיים (מספר קטן של מקרים)

ערכים חסרים בנתונים לשימוש ציבורי

SPSS	SAS	תווית	תחולה	סוג הערך החסר
Numeric: 6, 96 ... String: "996," "9996"	Numeric: .V String: "996," "9996"	"Valid skip"	שאלון הרקע ומשתנים נגזרים, רכיבי קריאה	Valid skip
Numeric: 7, 97 ... String: "997," "9997"	Numeric: .D String: "997," "9997"	"Don't know"	שאלון הרקע ומשתנים נגזרים	לא ידוע
Numeric: 9, 98 ... String: "998," "9998"	Numeric: .R String: "998," "9998"	"Refused"	שאלון הרקע ומשתנים נגזרים	סירוב
Numeric: 9, 99 ... String: "999," "9999," "99999"	Numeric: .N String: "999," "9999," "99999"	"Not stated or inferred" (כללי) "Not reached/Not attempted" (פריטים קוגניטיביים)	כמעט כל המשתנים	לא צוין, לא ניתן לקידוד, הושמט, צונזר
Numeric: (.)	Numeric: (.)	אין	פריטים קוגניטיביים	לא נשאל / לא רלוונטי (מובנה בסקר)

שחרור קובצי הנתונים של ישראל

שחרור קובצי הנתונים על ישראל

- בתאריך 28.6 יעלה ה-OECD את ה-PDX עם נתוני הגל השני – ובהם ישראל
- למי שירצה לעשות עיבודים מורכבים יותר, יועלה לאתר גם קובץ PUF בגרסה שהיא בת השוואה בין-לאומית
- למי שירצה להשתמש במשתנים המקומיים – יונגש בלמ"ס קובץ PUF בגרסה מקומית מלאה

תודה!

שאלות?

לוח לדוגמה (A2.1)

Table A2.1 Percentage of adults scoring at each proficiency level in literacy

OECD	Below Level 1		Level 1		Level 2		Level 3		Level 4		Level 5		Missing	
	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.
National entities														
Australia	3.1	(0.3)	9.4	(0.5)	29.2	(0.7)	39.4	(0.9)	15.7	(0.7)	1.3	(0.2)	1.9	(0.2)
Austria	2.5	(0.3)	12.8	(0.7)	37.2	(0.9)	37.3	(0.9)	8.2	(0.5)	0.3	(0.1)	1.8	(0.2)
Canada	3.8	(0.2)	12.6	(0.5)	31.7	(0.7)	37.3	(0.7)	12.8	(0.5)	0.9	(0.1)	0.9	(0.1)
Czech Republic	1.5	(0.3)	10.3	(0.7)	37.5	(1.6)	41.4	(1.4)	8.3	(0.8)	0.4	(0.2)	0.6	(0.2)
Denmark	3.8	(0.3)	11.9	(0.6)	34.0	(0.9)	39.9	(0.8)	9.6	(0.5)	0.4	(0.1)	0.4	(0.1)
Estonia	2.0	(0.2)	11.0	(0.5)	34.3	(0.7)	40.6	(0.8)	11.0	(0.5)	0.8	(0.2)	0.4	(0.1)
Finland	2.7	(0.2)	8.0	(0.5)	26.5	(0.9)	40.7	(0.8)	20.0	(0.6)	2.2	(0.3)	0.0	(0.0)
France	5.3	(0.3)	16.2	(0.5)	35.9	(0.8)	34.0	(0.7)	7.4	(0.4)	0.3	(0.1)	0.8	(0.1)
Germany	3.3	(0.4)	14.2	(0.7)	33.9	(1.0)	36.4	(0.9)	10.2	(0.6)	0.5	(0.2)	1.5	(0.2)
Ireland	4.3	(0.4)	13.2	(0.8)	37.6	(0.9)	36.0	(0.9)	8.1	(0.5)	0.4	(0.1)	0.5	(0.1)
Italy	5.5	(0.6)	22.2	(1.0)	42.0	(1.0)	26.4	(1.0)	3.3	(0.4)	0.1	(0.0)	0.7	(0.2)
Japan	0.6	(0.2)	4.3	(0.4)	22.8	(0.8)	48.6	(1.0)	21.4	(0.7)	1.2	(0.2)	1.2	(0.1)
Korea	2.2	(0.2)	10.6	(0.5)	37.0	(0.9)	41.7	(0.9)	7.9	(0.5)	0.2	(0.1)	0.3	(0.1)
Netherlands	2.6	(0.3)	9.1	(0.5)	26.4	(0.7)	41.5	(0.8)	16.8	(0.6)	1.3	(0.2)	2.3	(0.2)
Norway	3.0	(0.3)	9.3	(0.6)	30.2	(0.8)	41.6	(0.8)	13.1	(0.6)	0.6	(0.1)	2.2	(0.2)
Poland	3.9	(0.3)	14.8	(0.6)	36.5	(0.9)	35.0	(0.9)	9.0	(0.5)	0.7	(0.1)	0.0	(0.0)
Slovak Republic	1.9	(0.2)	9.7	(0.5)	36.2	(1.0)	44.4	(0.9)	7.3	(0.5)	0.2	(0.1)	0.3	(0.1)
Spain	7.2	(0.5)	20.3	(0.8)	39.1	(0.7)	27.8	(0.7)	4.6	(0.4)	0.1	(0.1)	0.8	(0.1)
Sweden	3.7	(0.3)	9.6	(0.6)	29.1	(1.0)	41.6	(0.9)	14.9	(0.6)	1.2	(0.2)	0.0	(0.0)
United States	3.9	(0.5)	13.6	(0.7)	32.6	(1.2)	34.2	(1.0)	10.9	(0.7)	0.6	(0.2)	4.2	(0.6)
Sub-national entities														
Flanders (Belgium)	2.7	(0.3)	11.3	(0.5)	29.6	(0.8)	38.8	(0.9)	11.9	(0.5)	0.4	(0.2)	5.2	(0.2)
England (UK)	3.3	(0.4)	13.1	(0.7)	33.1	(1.0)	36.0	(1.0)	12.4	(0.7)	0.8	(0.2)	1.4	(0.2)
Northern Ireland (UK)	2.5	(0.5)	14.9	(0.9)	36.2	(1.5)	34.3	(1.6)	9.4	(0.6)	0.5	(0.2)	2.2	(0.3)
England/N. Ireland (UK)	3.3	(0.4)	13.1	(0.7)	33.2	(1.0)	35.9	(1.0)	12.3	(0.7)	0.8	(0.2)	1.4	(0.2)
Average	3.3	(0.1)	12.2	(0.1)	33.3	(0.2)	38.2	(0.2)	11.1	(0.1)	0.7	(0.0)	1.2	(0.0)
Partners														
Cyprus ¹	1.6	(0.2)	10.3	(0.5)	33.0	(0.9)	32.1	(0.9)	5.2	(0.4)	0.2	(0.1)	17.7	(0.4)
Russian Federation ²	1.6	(0.5)	11.5	(1.2)	34.9	(1.9)	41.2	(2.0)	10.4	(1.6)	0.4	(0.2)	0.0	(0.0)

