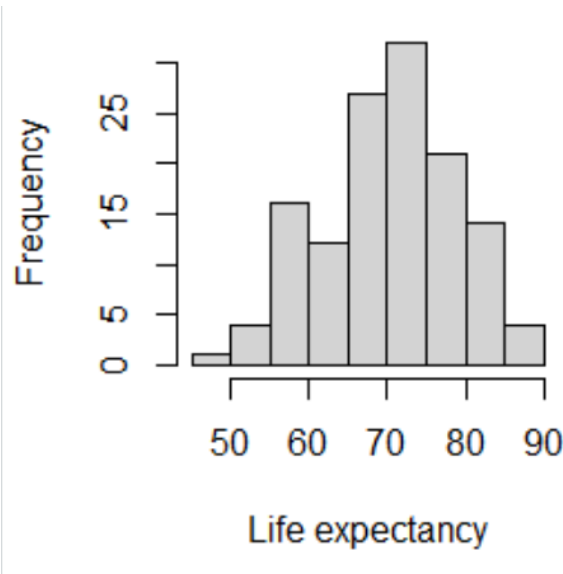


פרויקט ברגרסיה ליניארית

חלק א' - בחירת בסיס נתונים וניתוח סטטיסטי של הנתונים

ניתוח הגורמים השונים המשפיעים על תוחלת החיים בעולם



קבוצה 8

אופיר צרפתי 308244573
נעה מתתיהו 315474726
נוי כשר 314963810

תוכן עניינים

סעיף 1 - בחירת מאגר נתונים.....	שגיאה! הסימניה אינה מוגדרת.
סעיף 2 - יצירת טבלת משתנים.....	2
סעיף 3 - תיאור המשתנים.....	3
סעיף 4 - תיאור קשרים בין משתנים.....	6
סעיף 5 - ניתוח תיאורי של המשתנים.....	7
סעיף 6 - ניתוח חריגים.....	14
סעיף 7 - פונקציית צפיפות והתפלגות מצטברת.....	17
סעיף 8 - ייצוג קשרים בעזרת תרשימים.....	18
סעיף 9 - טבלאות שכיחות.....	20

סעיף 2 - יצירת טבלת משתנים

נגדיר את "תוחלת החיים" כמשתנה המוסבר ואת שאר המשתנים כמשתנים מסבירים, זאת משום שברצוננו לנתח את הגורמים המשפיעים על תוחלת החיים של מדינה ובכך לדעת כיצד לשפר אותה ביעילות.

סוג המשתנה - מוסבר/מסביר	סימון	יחידת מידה	סוג המשתנה - רציף/קטגורי	הסבר קצר על המשתנה
מוסבר	Y	שנים	רציף	משתנה המתאר את תוחלת החיים במדינה
מסביר	X ₁	null	קטגורי	סטטוס מדינה – 1-מפותחת או 2-מתפתחת
	X ₂	אחוז	רציף	שורש התמורה בקרב מבוגרים משני המינים (הסיכוי ל בגיל 15-60 מתוך אוכלוסייה בת 1000 אנשים)
	X ₃	אחוז	רציף	מקרי המוות בקרב תינוקות משני המינים מתוך אוכלוסייה בת 1000 אנשים
	X ₄	null	קטגורי	צריכת אלקהול לנפש מעל גיל 15 בליטרים – [3.046, 0.00518] [6.082, 3.046] [9.118, 6.082] [12.154, 9.118] 5 - (12.154, 9.118)
	X ₅	אחוז	רציף	הוצאות בריאות כאחוז מהתוצר המקומי הגולמי לנפש
	X ₆	אחוז	רציף	אחוז החיסונים נגד הפטיטיס B בקרב ילדים עד גיל 5
	X ₇	null	רציף	מקרי החצבת המדווה מתוך אוכלוסייה בת 1000 אנשים
	X ₈	null	רציף	מדד מסת גוף ממוצעת של כלל האוכלוסייה - BMI
	X ₉	null	רציף	מקרי המוות מתחת לגיל 5 מתוך אוכלוסייה בת 1000 אנשים
	X ₁₀	אחוז	רציף	כיסוי החיסונים נגד פוליו בקרב ילדים עד גיל שנה
	X ₁₁	אחוז	רציף	צאצאים ממשלתית כללית לבריאות כאחוז מסך ההוצאות הממשלתיות
	X ₁₂	אחוז	רציף	אחוז החיסונים נגד טטנוס ושעלת בקרב ילדים עד גיל 5
	X ₁₃	null	רציף	מקרי מוות בקרב לידות נשאי איידס בגילאים 0-4
	X ₁₄	דולר	רציף	תוצר מקומי גולמי לנפש
	X ₁₅	null	רציף	אוכלוסיית המדינה
	X ₁₆	אחוז	רציף	שכיחות הרזון בקרב ילדים ובני נוער בגילאי 10-19
	X ₁₇	אחוז	רציף	שכיחות הרזון בקרב ילדים בגילאי 5-9
	X ₁₉	שנים	רציף	מספר שנות הלימוד

סעיף 3 - תיאור המשתנים

בבסיס הנתונים, שסופק על ידי ארגון הבריאות העולמית (WHO), ישנם 20 משתני חיזוי, אשר מחולקים לארבע קבוצות עיקריות: גורמים הקשורים לחיסון, גורמי תמותה, גורמים כלכליים וגורמים חברתיים. (*) בכלל הגרפים, נסמן מדינות מפותחות בכתום, ומדינות מתפתחות בכחול.

תרומת Status ל life expectancy

משתנה המתאר האם המדינה הינה מדינה מפותחת (1) או מתפתחת (2). תוחלת החיים במדינות מתפתחות ירודה בשל המחסור בתרופות וברופאים ובכללי התברואה נמוכה. כמו כן, הנגישות למי שתייה נקיים לוקה בחסר, דבר אשר גם משפיע על התוחלת הנמוכה. לעומת זאת, תוחלת החיים במדינות מפותחות הינה גבוהה יותר.

תרומת Adult Mortality ל life expectancy

משתנה המתאר את מספר מקרי המוות של מבוגרים (בין הגילאים 15 עד 60) מתוך מדגם בגודל אלף. תמותה של מבוגרים מעידה על מצב הבריאות במהלך שנות הביניים – שהם רוב חיי האדם. לכן, **משתנה זה נמצא ביחס הפוך עם תוחלת החיים**.

תרומת infant deaths ל life expectancy

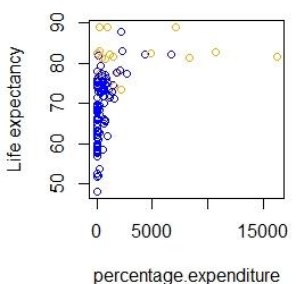
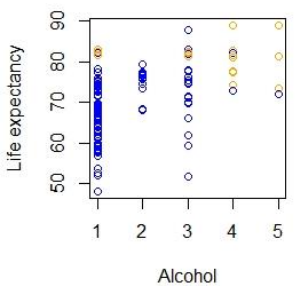
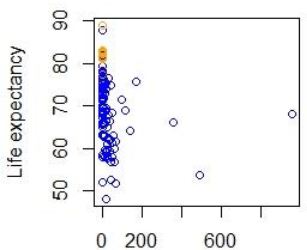
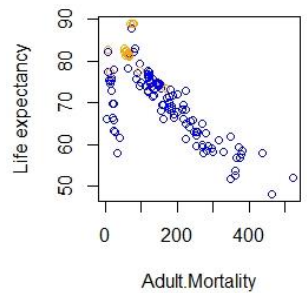
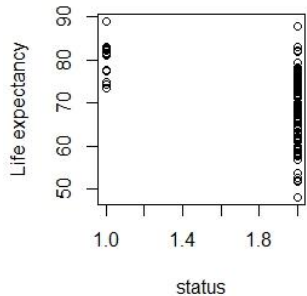
משתנה המתאר את מספר מקרי מוות בקרב תינוקות מתוך מדגם בגודל אלף. עד אמצע המאה ה-20, שיעור התמותה בקרב תינוקות היה בין 40-60% לערך מכלל התמותה של האוכלוסייה. כלומר, תמותה גבוהה של תינוקות גורמת לערכים נמוכים יותר של תוחלת החיים בלידה מאשר בגילאים אחרים. לכן, **משתנה זה נמצא ביחס הפוך עם תוחלת החיים**.

תרומת Alcohol ל life expectancy

משתנה המתאר את צריכת האלכוהול הנרשמת לנפש לשנה (מגיל 15 ומעלה) ביחידות של ליטר אלכוהול טהור. מחקרים מראים כי אנשים בעלי הפרעות שימוש באלכוהול בעלי תוחלת חיים קצרה יותר באופן משמעותי מהאוכלוסייה הכללית. על כן, **משתנה זה נמצא ביחס הפוך עם תוחלת החיים**. עם זאת, ניתן לצפות שהיחס במדינות מתפתחות יהיה חזק יותר מאשר במדינות מפותחות.

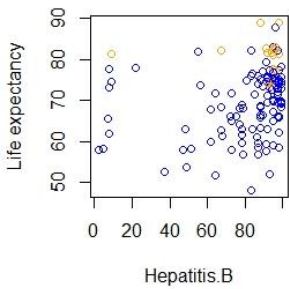
תרומת percentage expenditure ל life expectancy

משתנה המתאר את ההוצאה לבריאות מתוך התוצר המקומי הגולמי לנפש (GDP). תמ"ג הינו מונח כלכלי המציין מדידה של הערך הכולל של הסחורות והשירותים שיוצרו במדינה מסוימת במהלך תקופה נתונה במונחים של כסף. מחקרים הראו כי **משתנה זה נמצא ביחס ישר עם תוחלת החיים**, כלומר ככל שהוצאות עבור שירותי בריאות מתוך התמ"ג גדלות (באחוזים), כך ניתן לראות עלייה בתוחלת החיים, זאת לעומת מדינות בעלות תמ"ג זהה עם השקעה פחותה בבריאות.



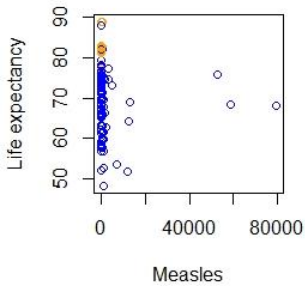
תרומת Hepatitis B ל life expectancy

משתנה המתאר כיסוי החיסונים לצהבת מסוג B בקרב ילדים בגיל שנה באחוזים. זיהום כרוני בצהבת מסוג B הוא גורם סיכון ידוע לתמותה מוגברת מסרטן HCC ומחלות כבד כרוניות (CLD). עם זאת נשאים שומרים על תוחלת חיים מספקת. על כן, ככל שאחוז החיסונים לצהבת מסוג B גבוה יותר, נצפה לראות תוחלת חיים גבוהה יותר. **כלומר, משתנה זה נמצא ביחס ישר עם תוחלת החיים.**



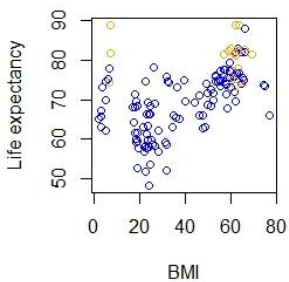
תרומת Measles ל life expectancy

משתנה המתאר את מספר מקרי החצבת המדווחים מתוך מדגם בגודל אלף. חצבת היא אחת המחלות הזיהומיות השכיחות בעולם, ונחשבת למחלה הנגיפית המדבקת ביותר. רוב האנשים שורדים את המחלה, ובמקרים מסוימים, בעיקר אצל ילדים, ישנם סיבוכים. על כן, ככל שמספר מקרי החצבת גבוה יותר, נצפה לראות תוחלת חיים נמוכה יותר אך לא באופן משמעותי. **כלומר, משתנה זה נמצא ביחס הפוך וחלש עם תוחלת החיים.**



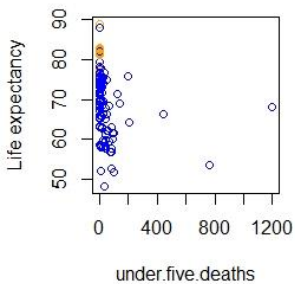
תרומת BMI ל life expectancy

משתנה המתאר את מדד מסת גוף ממוצע של כלל האוכלוסייה. BMI עודף קשור באופן מהותי לתוחלת חיים נמוכה יותר, עם רמת בריאות נמוכה יותר הכוללת מחלות כרוניות שונות. התמודדות עם השמנת יתר תוביל לאיכות חיים גבוהה יותר באוכלוסייה. כמו כן, התמודדות עם רזון תוביל גם היא לתוחלת חיים גבוהה יותר. מכאן, **משתנה זה נמצא ביחס הפוך עם תוחלת החיים כאשר הוא קיצוני לכל אחד מהצדדים.**



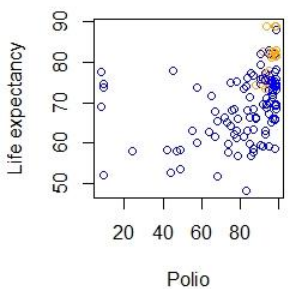
תרומת under-five deaths ל life expectancy

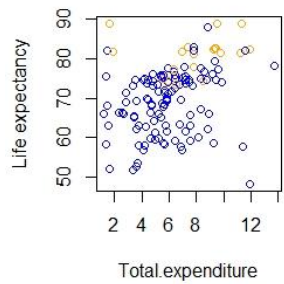
משתנה המתאר את מספר מקרי המוות מתחת לגיל חמש מתוך מדגם בגודל אלף. מחקרים הראו כי הנטל של מקרי מוות בקרב ילדים הוא הכבד ביותר באפריקה שמדרום לסהרה (53% בשנת 2019, עלייה של 22% משנת 1990 והוא צפוי לגדול אף יותר בעשורים הקרובים). מספר רב של מקרי מוות מתחת לגיל 5 במדינה מסוימת עשוי להעיד על תוחלת חיים נמוכה באותה המדינה. על כן, **משתנה זה נמצא ביחס הפוך עם תוחלת החיים.**



תרומת Polio ל life expectancy

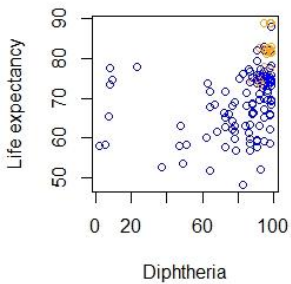
משתנה המתאר כיסוי חיסונים נגד פוליו (Pol3) בקרב ילדים בני שנה באחוזים. פוליו היא מחלה מסכנת חיים אשר ניתן למנוע באמצעות חיסון. תסמיני הפוליו משתנים מתסמיני דמויי שפעת קלים, לשיתוק ואף למוות. אנשים שסבלו מפוליו עשויים לחוות תופעות בשלב מאוחר יותר בחייהם. על כן, ככל שאחוז החיסונים לפוליו גבוה יותר, נצפה לראות תוחלת חיים גבוהה יותר. **כלומר, משתנה זה נמצא ביחס ישר עם תוחלת החיים.**





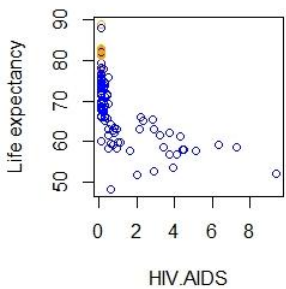
תרומת Total expenditure ל life expectancy:

משתנה המתאר את ההוצאה הממשלתית הכללית לבריאות כאחוז מסך ההוצאה הממשלתית. רמת מערכת הבריאות בכל מדינה נגזרת בין היתר מההוצאה הממשלתית הכללית לבריאות. כאשר ישנה השקעה גבוהה יותר (באחוזים) – נצפה לקבל תוחלת חיים גבוהה יותר. כלומר, **משתנה זה נמצא ביחס ישר עם תוחלת החיים.**



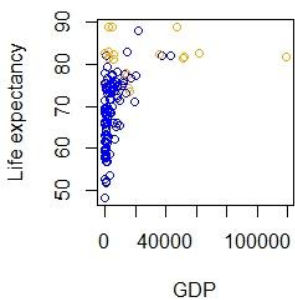
תרומת Diphtheria ל life expectancy:

משתנה המתאר את כיסוי חיסונים נגד טטנוס ושעלת (DTP3) בקרב ילדים בני שנה באחוזים. במקרים חמורים עלולה להתפתח חסימה בדרכי הנשימה. סיבוכים נוספים הינם דלקות בשריר הלב או דלקות בעצבים. מקרי המוות אשר נרשמו מהמחלה עומדים על 5-10%. על כן, ככל שאחוז החיסונים לטטנוס ושעלת גבוה יותר, נצפה לראות תוחלת חיים גבוהה יותר. כלומר, **משתנה זה נמצא ביחס ישר עם תוחלת החיים.**



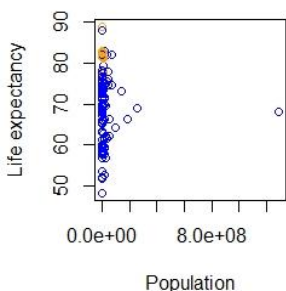
תרומת HIV/AIDS ל life expectancy:

משתנה המתאר את מספר מקרי מוות מאיידס למדגם בגודל אלף. לפני שלושים שנה לערך, אבחון כחולה HIV היה נחשב לגזר דין מוות. כיום, אנשים הנושאים את הנגיף יכולים לחיות חיים ארוכים ובריאים. איתור מוקדם וטיפול בזמן הם המפתח לניהול הנגיף, הארכת תוחלת החיים והפחתת הסיכון להעברה. **על כן, אינו צפוי קשר בין משתנה זה לתוחלת החיים באופן ישיר,** כלומר הדבר תלוי במערכת הבריאות של המדינה.



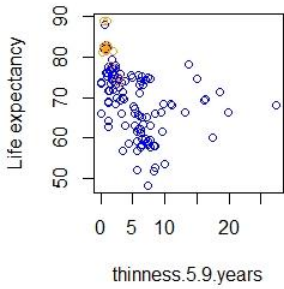
תרומת GDP ל life expectancy:

משתנה המתאר את התוצר המקומי הגולמי לנפש (בדולר אמריקאי). תמ"ג הינו מונח כלכלי המציין מדידה של הערך הכולל של הסחורות והשירותים שיוצרו במדינה מסוימת במהלך תקופה נתונה. מחקרים הראו שאנשים במדינות עשירות חיים זמן רב יותר מאנשים במדינות עניות. כלומר, **יש קשר ישר וחזק בין התוצר לתוחלת החיים.** ובכל זאת, כאשר התפוקה הכלכלית של מדינה – התוצר שלה – גבוה מהצפוי, שיעורי התמותה גם הם גבוהים מהצפוי (לדוגמה תמ"ג מאוד גבוה בעקבות אסון טבע, תמ"ג גבוה שגורם ליותר זיהום, יותר הרוגים בתאונות דרכים, שימוש בסמים ואלכוהול וכו'), ועל כן **במקרה זה משתנה זה יהיה ביחס הפוך לתוחלת החיים.**



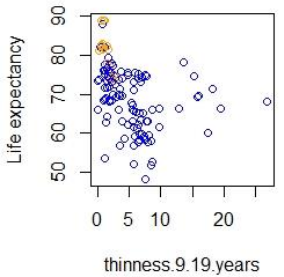
תרומת Population ל life expectancy:

משתנה המתאר את אוכלוסייה המדינה. **משתנה זה אינו משפיע באופן ישיר על תוחלת החיים** (בניגוד לצפיפות האוכלוסייה לדוגמה), אך משפיע על מדדים מסבירים אחרים ודרכם גם על המדד המוסבר.



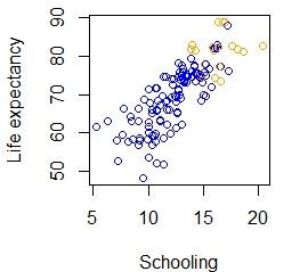
תרומת $life\ expectancy - l\ thinness\ 5-9\ years$

משתנה המתאר את שכיחות הרזון בקרב ילדים בגיל 5 עד 9 באחוזים. רזון מעיד על רעב ותזונה לקויה, גורמים שפוגעים בבריאות ומורידים את תוחלת החיים. על כן, **משתנה זה ביחס הפוך לתוחלת החיים.**



תרומת $life\ expectancy - l\ thinness\ 10-19\ years$

משתנה המתאר את שכיחות הרזון בקרב ילדים ובני נוער בגילאי 10 עד 19 באחוזים. רזון מעיד על רעב ותזונה לקויה, גורמים שפוגעים בבריאות ומורידים את תוחלת החיים. על כן, **משתנה זה ביחס הפוך לתוחלת החיים.**



תרומת $life\ expectancy - l\ Schooling$

משתנה המתאר את מספר שנות הלימוד הממוצע במדינה מסוימת. השקעה בחינוך מונעת מקרי אלימות ובזיזה, מעודדת אורח חיים בריא ולכן קשורה לתוחלת החיים. כאשר ישנו מספר גדול יותר של שנות לימוד, תוחלת החיים גבוהה יותר. לכן, **משתנה זה ביחס ישר לתוחלת החיים.**

סעיף 4 - תיאור קשרים בין משתנים

ניתוח הקשרים בין המשתנים המסבירים:

על מנת למצוא קשרים בין המשתנים המסבירים, השתמשנו במתאם פירסון. חישובנו בעזרת תוכנת R את הקשרים בין כל המשתנים הרציפים בבסיס הנתונים. בחרנו להסביר בהרחבה על הקשרים החזקים יותר שמתאם פירסון שלהם (P) נמצא בטווחים: $0.5 \leq P \leq 1.0$ ו- $-1.0 \leq P \leq -0.5$. מצורפת תמונה עם התוצאות שהתקבלו:

	Adult.Morl	infant.dea	percentag	Hepatitis.t	Measles	BMI	under.five	Polio	Total.exp	Diphtheria	HIV.AIDS	GDP	Population	thinness.9.1	thinness.5.9	Income.cc	Schooling
Adult.Morl	1	0.15429	-0.25189	-0.19436	0.04944	-0.39942	0.17907	-0.3407	-0.13952	-0.2325587	0.65088	-0.28426	0.02586886	0.2643751	0.2897276	-0.66756	-0.57088
infant.dea	0.15429	1	-0.09788	-0.08674	0.69649	-0.22082	0.99539	-0.12519	-0.15263	-0.0917612	0.06925	-0.09916	0.84893064	0.4837073	0.5469359	-0.19462	-0.21112
percentag	-0.25189	-0.09788	1	0.10717	-0.07111	0.15205	-0.10032	0.14077	0.23918	0.1365305	-0.14058	0.94742	-0.0379555	-0.2419179	-0.255424	0.44544	0.39796
Hepatitis.t	-0.19436	-0.08674	0.10717	1	-0.02207	0.07081	-0.10013	0.6491	0.10211	0.8918517	-0.1828	0.13952	-0.0309213	-0.0056216	-0.035887	0.21631	0.25517
Measles	0.04944	0.69649	-0.07111	-0.02207	1	-0.14642	0.66589	-0.03821	-0.09916	-0.0265581	-0.05473	-0.05235	0.6879521	0.2624452	0.3434262	-0.04128	-0.06264
BMI	-0.39942	-0.22082	0.15205	0.07081	-0.14642	1	-0.23067	0.13822	0.16039	0.1188557	-0.32155	0.19229	-0.1200018	-0.5348894	-0.575856	0.6288	0.55586
under.five	0.17907	0.99539	-0.10032	-0.10013	0.66589	-0.23067	1	-0.13957	-0.1566	-0.1084691	0.09445	-0.10285	0.8086743	0.4628805	0.5381046	-0.21579	-0.22897
Polio	-0.3407	-0.12519	0.14077	0.6491	-0.03821	0.13822	-0.13957	1	0.23763	0.7346604	-0.31395	0.17608	-0.0665209	-0.0558164	-0.080476	0.35637	0.36344
Total.exp	-0.13952	-0.15263	0.23918	0.10211	-0.09916	0.16039	-0.1566	0.23763	1	0.1930985	-0.09643	0.1664	-0.0804171	-0.2127369	-0.259477	0.26318	0.30204
Diphtheria	-0.23256	-0.09176	0.13653	0.89185	-0.02656	0.11886	-0.10847	0.73466	0.1931	1	-0.18764	0.16296	-0.0109919	-0.0579785	-0.089739	0.32688	0.34052
HIV.AIDS	0.65088	0.06925	-0.14058	-0.1828	-0.05473	-0.32155	0.09445	-0.31395	-0.09643	-0.187636	1	-0.16588	-0.0455103	0.1518572	0.1526601	-0.47666	-0.38571
GDP	-0.28426	-0.09916	0.94742	0.13952	-0.05235	0.19229	-0.10285	0.17608	0.1664	0.1629611	-0.16588	1	-0.0415363	-0.271914	-0.285268	0.50308	0.44948
Population	0.02587	0.84893	-0.03796	-0.03092	0.68795	-0.12	0.80867	-0.06652	-0.08042	-0.0109919	-0.04551	-0.04154	1	0.4528901	0.4427403	-0.04684	-0.0527
thinness.5	0.26438	0.48371	-0.24192	-0.00562	0.26245	-0.53489	0.46288	-0.05582	-0.21274	-0.0579785	0.15186	-0.27191	0.4528901	1	0.9394442	-0.49626	-0.48558
thinness.9	0.28973	0.54694	-0.25542	-0.03589	0.34343	-0.57586	0.5381	-0.08048	-0.25948	-0.0897386	0.15266	-0.28527	0.44274025	0.9394442	1	-0.50742	-0.49722
Income.cc	-0.66756	-0.19462	0.44544	0.21631	-0.04128	0.6288	-0.21579	0.35637	0.26318	0.3268842	-0.47666	0.50308	-0.0468367	-0.496263	-0.507418	1	0.91595
Schooling	-0.57088	-0.21112	0.39796	0.25517	-0.06264	0.55586	-0.22897	0.36344	0.30204	0.3405179	-0.38571	0.44948	-0.0527	-0.4855823	-0.49722	0.91595	1

קשרים סיבתיים:• הקשר בין infant deaths לבין under five deaths:

infant deaths מציג את מספר מקרי המוות של תינוקות במדגם בגודל 1000, ואילו under five deaths מציג את מספר מקרי המוות של ילדים מתחת לגיל 5 במדגם באותו הגודל. לכן, קל מאוד לראות את הקשר. בעוד שגיל של תינוק הוא בין 0-3, קבוצה זאת בעצם מוכלת בתוך הקבוצה של התמותה מתחת לגיל 5. ואכן קיבלנו קורלציה גבוהה מאוד – מקדם מתאם **0.995385655**

• הקשר בין GDP לבין percentage expenditure:

percentage expenditure הוא הוצאה לבריאות מתוך התוצר המקומי הגולמי לנפש, ואילו ה-GDP הוא התוצר המקומי הגולמי לנפש. נסביר את הקשר: ככל שלמדינה יש יותר כסף, היא יכולה להשקיע יותר ברווחה שלה, ופרט במערכת הבריאות. ולהפך – כאשר מדינה ענייה וה-GDP שלה נמוך, אין לה כסף רב להשקיע ברווחה והיא משקיעה פחות במערכת הבריאות שלה. לכן אנו רואים קורלציה גבוהה בין ה-GDP להוצאה לבריאות מתוך התוצר המקומי הגולמי לנפש – מקדם המתאם **0.947421497**.

• הקשר בין Measles לבין infant deaths:

קיבלנו קשר חזק בין מספר מקרי החצבת במדגם בגודל 1000, לבין מספר תמותות תינוקות במדגם באותו הגודל. התקבל מקדם מתאם **0.69649063**. ניתן להסביר את הקשר בכך שרוב מוחלט מהנדבקים בחצבת הם ילדים. לכן, כאשר ישנם יותר מקרי חצבת – ישנה תמותת תינוקות גדולה בהתאם, ולהפך.

קשרים מדגמיים:• הקשר בין Hepatitis B לבין Diphtheria:

נמצא קשר מדגמי חזק בין אחוז המתחסנים נגד צהבת לבין אחוז המתחסנים נגד שעלת/טטנוס, עם מקדם מתאם של **0.891851741**. סיבה **משוערת** יכולה להיות שבמדינות שבהן אין מודעת גבוהה לחיסונים, או להפך, מדובר בכמה סוגי חיסונים, בניהם צהבת וטטנוס.

• הקשר בין BMI לבין Schooling:

נמצא קשר מדגמי חזק בין מדד מסת הגוף הממוצע לבין מספר שנות הלימוד, עם מקדם מתאם של **0.555856207**.

סעיף 5 - ניתוח תיאורי של המשתניםהמשתנה המוסבר:

תוחלת חיים (Life expectancy) – משתנה רציף:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
48.10	64.65	72.00	70.52	75.80	89.00

סטיית תקן: 8.605226

א-סימטריה: -0.2192344

הסבר התוצאות:

תחילה, נראה כי ערך הממוצע (Mean) נמוך מערך החציון, כלומר קיימות תצפיות מסוימות המושכות את תוחלת החיים הכללית כלפי מטה. מכך נוכל להסיק כי התפלגות המשתנה נוטה להיות התפלגות א-סימטרית בעלת זנב שמאלי. הסבר נוסף המחזק טענה זו הינו מדד הא-סימטריה אשר ערכו שלילי ועומד על -0.2192344. נוכל להסביר זאת בכך שתצפיות רבות ממאגר הנתונים מהוות מדינות מתפתחות ועניות מאוד, ותוחלת החיים בהן נמוכה משמעותית מהממוצע במדינות מערביות. ולכן אנו רואים פונקציה א – סימטרית בעלת זנב שמאלי.

משתנים מסבירים:**תמותת מבוגרים (Adult.Mortality) – משתנה רציף:**

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.0	74.5	144.0	160.4	225.0	522.0

סטיית תקן: 110.1424

א-סימטריה: 0.738551

הסבר התוצאות:

נוכל לראות כי ערך הממוצע (Mean) גבוהה משמעותית מערך החציון, כלומר קיימות תצפיות מסוימות המושכות את המדד כלפי מעלה. מכך נוכל להסיק כי התפלגות המשתנה נוטה להיות התפלגות א-סימטרית בעלת זנב ימני. בנוסף, ניתן לראות כי הערך המקסימלי גבוה מאוד מהערך הממוצע ומהחציון, ונוכל להבין כי ערכים חריגים כאלו הם אלו שמושכים את המדד כלפי מעלה. ניתן להסביר את התופעה בכך שבמדינות רבות מהם נלקחו תצפיות מתרחשות מלחמות עקובות מדם, ובמדינות כאלו מספר המבוגרים המתים (מבוגר מוגדר בין הגילאים 15 – 60) הינו גבוה מאוד.

תמותת ילדים (infant.deaths) – משתנה רציף:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	0.00	3.00	28.56	20.00	957.00

סטיית תקן: 100.0954

א-סימטריה: 7.14421

הסבר התוצאות:

נוכל לראות כי הערך הממוצע (Mean) גבוהה משמעותית מערך החציון. למעשה, הערך הממוצע הוא פי 9.5 יותר מהחציון! לכן נוכל שבמקרה זה המדד הממוצע אינו אמין ואינו משקף את מה שקורה בפועל. כלומר, קיבלנו פונקציה א-סימטרית חריגה בעלת זנב ימני. נתון מעניין נוסף נמצא בקצוות – הערך המינימלי הוא 0 מתים ל-1000 אוכלוסייה. כלומר בדיוק 0% תמותת ילדים, ואילו הערך המקסימלי הוא 957 ילדים ל-1000 אוכלוסייה, כלומר כמעט 100% אחוז תמותה. לפי הקצוות היינו מצפים לקבל פונקציה סימטרית, אך האמור לעיל מפרך את השערה זו למרות הקצוות.

הוצאות בריאות כאחוז מהתוצר המקומי הגולמי לנפש (percentage.expenditure) - משתנה רציף:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.443	48.311	198.734	850.874	718.324	16255.162

סטיית תקן: 2071.444

א-סימטריה: 4.807182

הסבר התוצאות:

גם כאן נוכל לראות כי הערך הממוצע (Mean) גבוהה משמעותית מערך החציון. למעשה, הערך הממוצע הוא פי 4.3 יותר מהחציון. לכן נוכל שבמקרה זה המדד הממוצע אינו אמין ואינו משקף את מה שקורה בפועל. קיימות תצפיות מסוימות המושכות את המדד כלפי מעלה. קיבלנו פונקציה א-סימטרית בעלת זנב ימני.

כיסוי החיסונים נגד הפטיטיס b בקרבי ילדים עד גיל שנה (Hepatitis.B) - משתנה רציף:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.00	78.00	91.00	81.71	96.00	99.00

סטיית תקן: 23.76406

א-סימטריה: -2.080927

הסבר התוצאות:

ניתן לראות כי הערך הממוצע (Mean) נמוך מערך החציון. בנוסף, ערך האסימטריה שלילי וערכו -2.08. לכן קיבלנו פונקציה אסימטרית בעלת זנב שמאלי. נתון מעניין נוסף נמצא באחוזונים ה-25 וה-75. ב-25 כיסוי החיסונים הוא 78 אחוז, וב-75 הוא 99 אחוז. מדובר באחוזי חיסון מאוד גבוהים. מנתון זה ניתן לראות כי רוב אוכלוסיית העולם מחוסנת נגד צהבת B.

מספר מקרי החצבת המדווחים לכל 1000 אוכלוסיה (Measles):

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	0.0	10.0	2042.9	289.5	79563.0

סטיית תקן: 9842.341

א-סימטריה: 6.255131

הסבר התוצאות:

ניתן לראות כי הערך הממוצע (Mean) גבוהה משמעותית מערך החציון. למעשה, הערך הממוצע הוא פי 204 יותר מהחציון! לכן נוכל שבמקרה זה המדד הממוצע אינו אמין ואינו משקף את מה שקורה בפועל. ערך האסימטריה חיובי גבוהה ושווה ל-6.255131. קיבלנו פונקציה אסימטרית חריגה בעלת זנב ימני.

מדד מסת גוף ממוצעת - BMI (BMI):

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.00	22.85	45.90	40.48	59.45	77.10

סטיית תקן: 20.73367

א-סימטריה: -0.2207245

הסבר התוצאות:

ניתן לראות כי הערך הממוצע (Mean) נמוך מערך החציון. ערך האסימטריה שלילי ושווה ל-0.22. לכן קיבלנו פונקציה אסימטרית בעלת זנב שמאלי. נתון מעניין הוא שלפי מדד ה-BMI ערך הגבוה מ-30 מעיד על משקל יתר. נוכל לראות שגם החציון וגם הממוצע גבוהים מ-40, לכן ניתן להסיק שרוב העולם סובל ממשקל יתר.

מספר מקרי המוות מתחת לגיל 5 (under.five.deaths):

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	1.00	3.00	38.24	25.50	1200.00

סטיית תקן: 131.2931

א-סימטריה: 6.797799

הסבר התוצאות:

ניתן לראות כי הערך הממוצע (Mean) גבוהה משמעותית מערך החציון. למעשה, הערך הממוצע הוא פי 12 יותר מהחציון! לכן נוכל שבמקרה זה המדד הממוצע אינו אמין ואינו משקף את מה שקורה בפועל. ערך האסימטריה חיובי גבוהה ושווה ל-6.7977. לכן קיבלנו פונקציה אסימטרית בעלת זנב ימני.

כיסוי החיסונים נגד פוליו בקרבי ילדים עד גיל שנה (Polio):

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
8.0	78.0	92.0	83.5	97.0	99.0

סטיית תקן: 20.96641

א-סימטריה: -2.131426

הסבר התוצאות:

ניתן לראות כי הערך הממוצע (Mean) נמוך מערך החציון. ערך האסימטריה שלילי ושווה ל-2.13. לכן קיבלנו פונקציה אסימטרית בעלת זנב שמאלי. נתון מעניין נוסף הוא שערך האחוזון ה-75 שווה ל-97 והינו קרוב מאוד לערך המקסמלי שערכו 99. ניתן להסיק מכך שישנם מדינות רבות שמחוסנות כמעט לגמרי לפוליו, וככל הנראה מדובר במדינות המפותחות.

הוצאה ממשלתית כללית לבריאות כאחוז מסך ההוצאות הממשלתיות (Total.expenditure):

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.210	4.485	5.820	6.107	7.630	13.730

סטיית תקן: 2.533227

א-סימטריה: 0.3099319

הסבר התוצאות:

ניתן לראות כי הערך הממוצע כמעט ושווה ערך החציון. מכאן ניתן להסיק שהפונקציה נוטה להיות מאוזנת וסימטרית. ניתן לראות גם שהערכים של האחוזונים ה-75 וה-25 יחסית קרובים, ורחוקים מהערכים המינימליים

והמקסימלים. לכן ניתן גם להגיד שהפונקציה מרוכזת סביב אחוזונים אלו. כלומר, ההוצאת הממשלתית הכלליות לבריאות דומות ברוב מדינות העולם.

כיסוי החיסונים נגד טטנוס ושעלת בקרבי ילדים עד גיל שנה (Diphtheria):

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.00	80.00	92.00	83.89	97.00	99.00

סטיית תקן: 21.83985

א-סימטריה: -2.348706

הסבר התוצאות:

ניתן לראות כי הערך הממוצע (Mean) נמוך מערך החציון. ערך האסימטריה שלילי ושווה ל -2.34. לכן קיבלנו פונקציה אסימטרית בעלת זנב שמאלי. נתון מעניין נוסף הוא שערך האחוזון ה-75 שווה ל-97 והינו קרוב מאוד לערך המקסימלי שערכו 99. ניתן להסיק מכך שישנם מדינות רבות שמחוסנות כמעט לגמרי לשעלת וטטנוס, וככל הנראה מדובר במדינות המפותחות.

מספר מקרי מוות בקרב לידות נשאי איידס בגילאים 0-4 שנים (HIV.AIDS):

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.1000	0.1000	0.1000	0.8099	0.5000	9.4000

סטיית תקן: 1.562043

א-סימטריה: 2.88741

הסבר התוצאות:

ניתן לראות כי הערך הממוצע (Mean) גדול משמעותית מערך החציון. למעשה, הערך הממוצע גבוה פי 8 מערך החציון. לכן ניתן להסיק במקרה זה שהממוצע אינו אמין ואינו משקף את אשר קורה בפועל. ערך האסימטריה חיובי ושווה ל -2.88741. לכן ניתן להסיק כי הפונקציה אסימטרית בעלת זנב ימני. נתון מעניין שניתן להבחין בו הוא שהחציון שווה לאחוזון ה-25 ולערך המינימלי. כלומר רוב מוחלט של התצפיות מרוכזות בצד הימני של הפונקציה.

תוצר מקומי גולמי לנפש (GDP):

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
12.28	547.85	2522.80	7794.77	7497.76	119172.74

תקן: 14741.4

א-סימטריה: 4.517472

ניתן לראות כי הערך הממוצע (Mean) גדול משמעותית מערך החציון. למעשה, הערך הממוצע גבוה פי 3 מערך החציון. ערך האסימטריה חיובי גבוה ושווה ל -4.517472. לכן ניתן להסיק כי הפונקציה אסימטרית בעלת זנב ימני. נתון מעניין שניתן להבחין בו הוא הפערים בין כלכלות העולם – התוצר הלאומי לנפש במדינה בעלת הערך המקסימלי גדול פי 9931 (!!!) מהמדינה בעלת הערך המינימלי.

אוכלוסיית המדינה (Population):

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
12.28	554.92	2522.80	7256.85	7438.05	119172.74

סטיית תקן: 116699866

א-סימטריה: 9.964728

הסבר התוצאות:

ניתן לראות כי הערך הממוצע (Mean) גדול משמעותית מערך החציון. למעשה, הערך הממוצע גבוהה כמעט פי 3 מערך החציון. ערך האסימטריה חיובי גבוהה ושווה ל - 9.964728. לכן ניתן להסיק כי הפונקציה אסימטרית בעלת זנב ימני.

שכיחות הרזון בקרב ילדים ובני נוער בגילאי 10-19 (thinness.1.19.years):

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.100	1.500	3.300	4.648	6.650	26.800

סטיית תקן: 4.420587

א-סימטריה: 1.990179

הסבר התוצאות:

ניתן לראות כי הערך הממוצע (Mean) גדול מעט מערך החציון. ערך האסימטריה חיובי ושווה ל - 1.990179. לכן ניתן להסיק כי הפונקציה אסימטרית בעלת זנב ימני. כלומר, רוב ילדי העולם בגילאים הללו סובלים מרזון.

שכיחות הרזון בקרב ילדים בגילאי 5-9 (thinness.5.9.years):

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.100	1.550	3.500	4.886	6.800	27.400

סטיית תקן: 4.543478

א-סימטריה: 1.883708

הסבר התוצאות:

ניתן לראות כי הערך הממוצע (Mean) גדול מעט מערך החציון. ערך האסימטריה חיובי ושווה ל - 1.990179. לכן ניתן להסיק כי הפונקציה אסימטרית בעלת זנב ימני. כלומר, רוב ילדי העולם בגילאים הללו סובלים מרזון. נשים לב שהערכים במדד זה דומים מאוד לערכים במדד הרזון עד גיל 19, ולמעלה בשל הקירבה בנתונים היה אפשר לאחד את המדדים למדד אחד.

מספר שנות הלימוד (Schooling):

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5.30	10.75	12.70	12.68	14.70	20.40

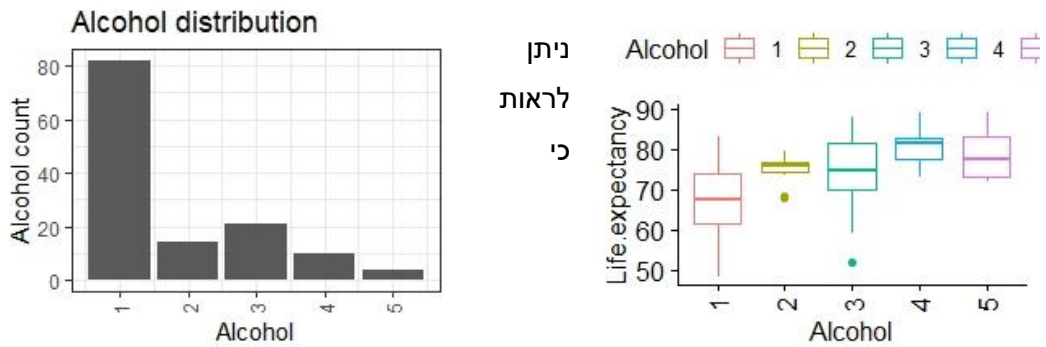
סטיית תקן: 2.75038

א-סימטריה: -0.03517039

הסבר התוצאות:

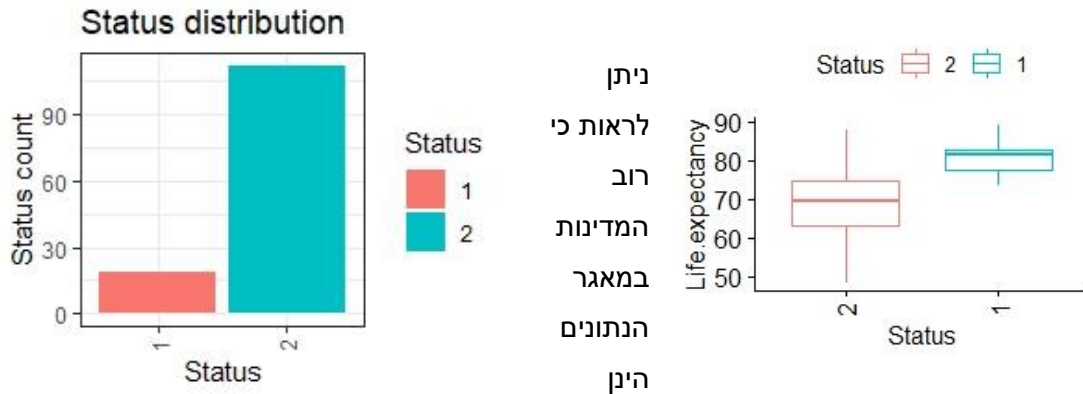
ניתן לראות כי הערך הממוצע כמעט ושווה ערך החציון. מכאן ניתן להסיק שהפונקציה נוטה להיות מאוזנת וסימטרית. למעלה, כאשר מסתכלים על ההפרשים בין האחוזונים ה-25 וה-75 לממוצע וההבדלים בין הערכים המינימליים והמקסימליים לאותם האחוזונים ניתן לראות כי הפונקציה מתפלגת נורמלית כמעט באופן מושלם.

אלכוהול – משתנה בדיד:



ההתפלגות הינה אסימטרית בעלת זנב שמאלי. למעשה, ברוב התצפיות שנלקחו (רוב המדינות) התקבל ערך בין 0.005 לבין 3.046 ליטרים של צריכת אלכוהול מזוקק לנפש. כלומר, במרבית המדינות אין בעיה מופרזת של שתיית אלכוהול. כלומר, במרבית התצפיות אלכוהול לא יסביר את תוחלת החיים, שכן ככל הנראה ההשוואה לרוב המדינות האחרות תהיה שווה.

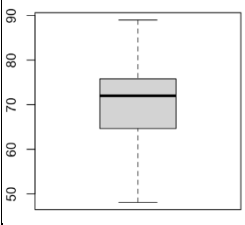
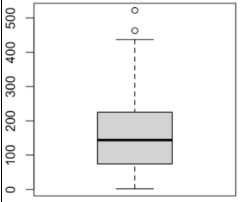
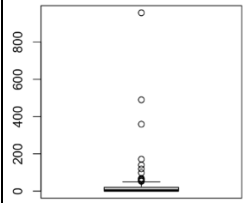
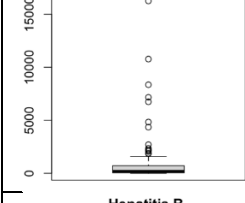
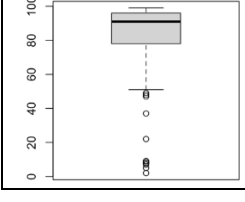
סטטוס – משתנה בדיד:

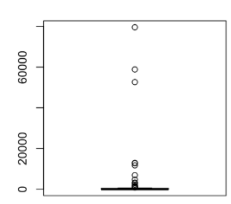
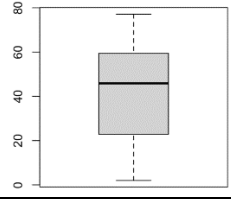
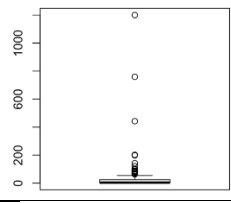
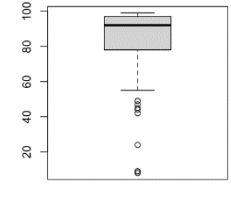
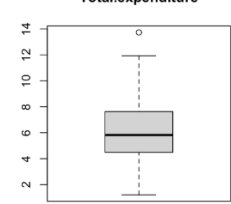
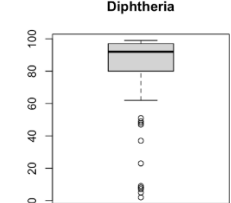
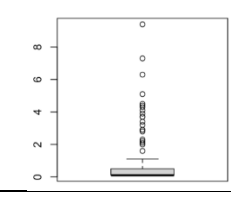


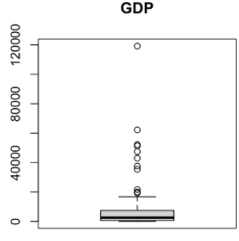
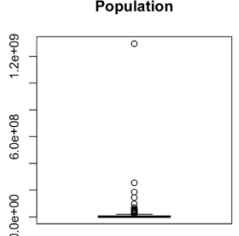
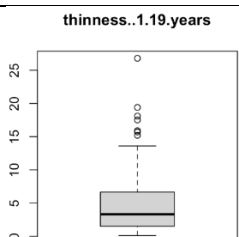
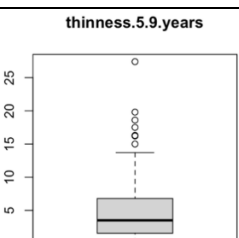
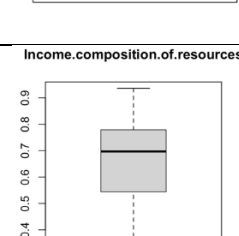
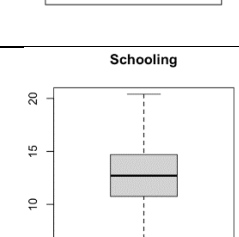
מדינות מתפתחות (status=2). בגרף המציג את הקשר בין סטטוס המדינה לבין מדד תוחלת החיים הממוצע במדינה זו, ניתן לראות קשר ישר וחזק בין השניים. כלומר, אם המדינה מפותחת מדד תוחלת החיים הממוצע בה יהיה בוודאות גבוה יותר מאשר במדינה מתפתחת.

סעיף 6 - ניתוח חריגים

נציין כי משמעות הורדת תצפיות היא הורדה של כלל ערכי המשתנים עבור אותה התצפית. כלומר, כאשר נוריד תצפית מסוימת עקב חריגה במשתנה אחד, נוותר על כלל המשתנים (העמודות) המבטאים ערכים עבור אותה התצפית (דהיינו, על הרשומה כולה). לכן נשקול בכובד ראש כל הורדה של תצפית ונבדוק האם החריגה מצדיקה הורדה.

משתנה	מספר חריגות	הסבר	תרשים הקופסא
תוחלת חיים (Life expectancy)	0	בתרשים ניתן לראות כי אין תצפיות חריגות.	
תמותת מבוגרים (Adult Mortality)	2	ניתן לראות כי ישנה חריגה של 2 תצפיות אשר חורגות מעל הגבול העליון של תרשים הקופסא. עם זאת, התצפיות קרובות אל הגבול העליון. כלומר, ערכן קרוב להיות "לא חריג". ולכן במקרה זה נחליט לא להוריד את התצפיות הללו.	
תמותת ילדים (infant deaths)	14	ניתן לראות כי ישנן 14 תצפיות חריגות מתוכן 3 במרחק חריג מהגבול העליון. חריגה זו יכולה להוות מחסום לניתוח איכותי של בסיס הנתונים בהמשך הדרך ולפגוע במהימנותם ועל כן נחליט להוריד את שלושת התצפיות הקיצוניות ביותר.	
הוצאות בריאות כאחוז מהתוצר המקומי הגולמי לנפש (percentage expenditure)	14	ניתן לראות כי חצי מן התצפיות (7) חורגות באופן גדול משאר התצפיות ולכן נבחר להוריד אותן.	
כיסוי החיסונים נגד הפטיטיס b בקרבי ילדים עד גיל שנה (Hepatitis.B)	13	ניתן לראות כי ישנם 10 תצפיות שמרחקן מהגבול התחתון גדול מאוד. למעכה, הערך של התצפית הגבוהה ביותר מתוך אותן 10 התצפיות הינו חצי מהגבול התחתון. לכן נבחר להוריד תצפיות אלו על מנת לשפר את איכות הניתוח.	

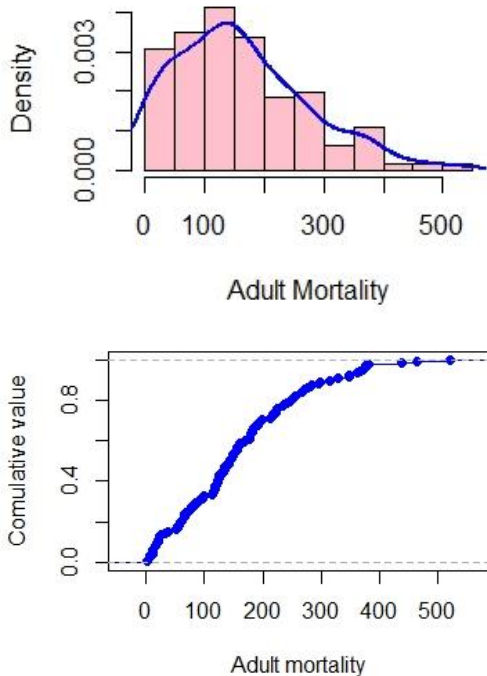
	<p>גם כאן ניתן לראות כי רוב החריגות אינן סטיות באופן רב מן הגבול העליון של תרשים הקופסא אך 3 מתוכם מבטאים חריגה רבה מן הגבול העליון. לאור מקרה זה נוצר פער בקצה ההתפלגות. פער זה הינו פער קיצוני ומבטא חוסר רצף. על כן נוריד את שלושת התצפיות החריגות.</p>	21	מספר מקרי החצבת המדווחים לכל 1000 אוכלוסיה (Measles)
	<p>בתרשים ניתן לראות כי אין חריגות</p>	0	מדד מסת גוף ממוצעת – BMI (BMI)
	<p>בדומה למשתנה החצבת, גם כאן ניתן לראות כי רוב החריגות אינן סטיות באופן רב מן הגבול העליון של תרשים הקופסא אך 3 מתוכם מבטאים חריגה רבה מן הגבול העליון. לאור מקרה זה נוצר פער בקצה ההתפלגות. פער זה הינו פער קיצוני ומבטא חוסר רצף. על כן נוריד את שלושת התצפיות החריגות.</p>	18	מספר מקרי המוות מתחת לגיל 5 (under.five.deaths)
	<p>נראה כי עבור משתנה זה קיימות 12 תצפיות חריגות שחורגות באופן רב מערך הגבול התחתון. אך אנו סבורים שמדד זה הינו מדד חשוב לבדיקת תוחלת החיים, ושכל הנראה החריגות מייצגות מדינות עולם שלישי שנרצה לבחון את נתונייהם בתכונות נוספות, ולכן בחרנו להשאיר את כל התצפיות החריגות.</p>	12	כיסוי החיסונים נגד פוליו בקרבי ילדים עד גיל שנה (Polio)
	<p>בתרשים ניתן לראות כי ישנה חריגה של תצפית אחת אשר חורגת מעל הגבול העליון של תרשים הקופסא. עם זאת, ערך החריגה קרוב לערך הגבול העליון, ולכן בחרנו להשאיר את התצפית.</p>	1	הוצאה ממשלתית כללית לבריאות כאחוז מסך ההוצאות הממשלתיות (Total.expenditure)
	<p>בדומה לכיסוי החיסונים נגד פוליו גם כאן ישנן 12 תצפיות חריגות שחורגות באופן רב מערך הרבעון אך כאמור אנו סבורים שהמדדים שנוגעים לכיסוי חיסונים חשובים מאוד לחישוב תוחלת החיים, ובנוסף שהמדינות החריגות הינן ממדינות עולם שלישי שחשוב לנו לבחון אותם בתכונות אחרות, ולכן בחרנו להשאיר את התצפיות החריגות.</p>	12	כיסוי החיסונים נגד טטנוס ושעלת בקרבי ילדים עד גיל שנה (Diphtheria)
	<p>במדד זה אנו רואים מספר רב של חריגות – 21. ניתן לראות כי למעט 3 חריגות, כל שאר החריגות מקובצות יחדיו קרוב לגבול העליון. לכן, למרות מספרן הרב של החריגות, בחרנו להשאיר 18 מהן ולהוריד את שלושת החריגות העליונות.</p>	21	מספר מקרי מוות בקרב לידות נשאי איידס בגילאים 0-4 שנים (HIV.AIDS)

 <p>GDP</p>	<p>במדד זה קיבלנו 13 תצפיות חריגות, כולם חריגות מהגבול העליון. ככל הנראה מדובר במדינות מערביות עשירות, וניתן לשים לב כי רובן לא רחוקות מהגבול העליון. מכיוון שהחריגה יחסית לא גדולה, ומכיוון שאנו מעריכים כי מדובר בקבוצה מסוימת של מדינות דומות שנרצה לקבל נתונים מהן, לא נמחק אף חריגה, גם לא את החריגה ביותר, כי ככל הנראה מדובר במדינה כמו ארה"ב או סין.</p>	13	תוצר מקומי גולמי (GDP) לנפש
 <p>Population</p>	<p>ניתן לראות כי התקבלו 11 חריגות – כולן מעל הגבול העליון, אך כולן חוץ מאחת מרוכזות סמוך מאוד לגבול העליון. לכן נבחר לא להוריד את החריגות האלו. לגבי החריגה הגדולה ביותר, אומנם מדובר בחריגה משמעותית, אך ככל הנראה היא מציינת את סין, שאוכלוסייתה מכילה כמעט חמישית מאוכלוסיית העולם. מכיוון שזוהי מדינה חשובה, נבחר לא להוריד גם את התצפית החריגה ביותר.</p>	11	אוכלוסיית המדינה (Population)
 <p>thinness..1.19.years</p>	<p>ניתן לראות כי ישנם 7 חריגות, כולן מעל הגבול העליון אך גם הן קרובות יחסית לגבול. מכיוון שרזון מעיד בדר"כ על רעב, בעיה ממנה סובלות המדינות העניות, נרצה לשמור את התצפיות הללו ובחר לא להוציא אף אחד מהחריגים.</p>	7	שכיחות הרזון בקרבי ילדים ובני נוער בגילאי 10-19 thinness.1.19.y (ears)
 <p>thinness.5.9.years</p>	<p>ניתן לראות כי מתוך 7 החריגות, ישנה חריגה אחת קיצונית יותר ושאר החריגות מקובצות בסמוך לגבול העליון. עם זאת, מכיוון שאנו רואים פונקציה לא סימטרית, אנו סבורים שגם חריגות אלו יכולות לפגוע בקבלת ניתוח נתונים איכותי. ולכן בחרנו להוריד את חמשת החריגות הגבוהות מבין ה-7.</p>	7	שכיחות הרזון בקרבי ילדים בגילאי 5-9 thinness.5.9.ye (ars)
 <p>Income.composition.of.resources</p>	<p>בתרשים ניתן לראות כי אין חריגות</p>	0	מדד התפתחות האדם מבחינת הרכב ההכנסות של המשאבים Income.compos ition.of.resource (s)
 <p>Schooling</p>	<p>בתרשים ניתן לראות כי אין חריגות</p>	0	מספר שנות הלימוד (Schooling)

סעיף 7 - פונקציית צפיפות והתפלגות מצטברת

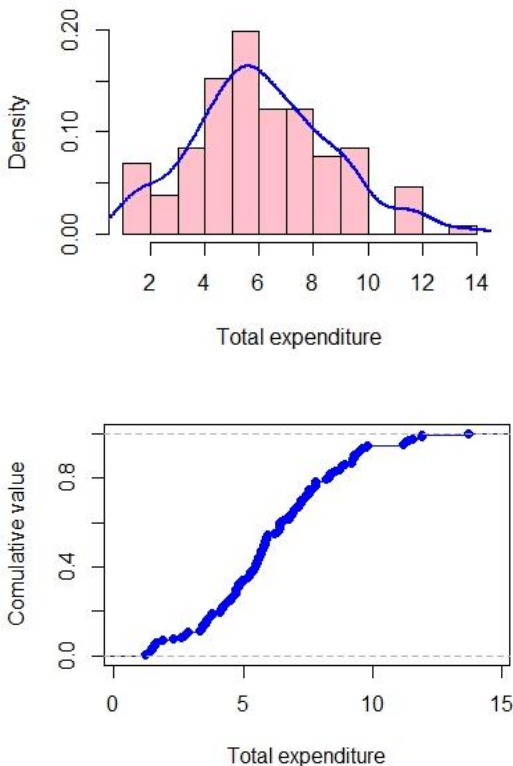
מספר מקרי המוות בקרב מבוגרים – משתנה רציף X_2

- עבור המשתנה המתאר את מספר מקרי המוות בקרב מבוגרים בגילאים 15-60 מתוך מדגם בגודל 1000, אנו עדים להתפלגות א-סימטרית בעלת זנב ימני. פונקציית הצפיפות ממחישה את מדד הא-סימטריה אשר ערכו חיובי. כמו כן, ניתן לראות כי מרבית התצפיות מרוכזות בין 50-150 מקרי מוות מתוך המדגם. כלומר, שיא ההתפלגות נמצא מתחת לחציון (אשר שווה ל – 144.0) ומתרכז באזור השמאלי של ההתפלגות. ניתוח זה עולה בקנה אחד עם ניתוח הנתונים שנעשה בסעיף 5.
- הצגה נוספת היכולה לאשש התפלגות א-סימטרית זו הינה פונקציית ההתפלגות המצטברת. ניתן לראות כי קיימת עלייה חדה עד 200 מקרי מוות בקרב מבוגרים לערך, ולאחר מכן העלייה נעשית מתונה יותר, כלומר משתטחת. התמתנות זו ממחישה את הפיזור על טווח ערכים גדול יותר בזנב הימני.



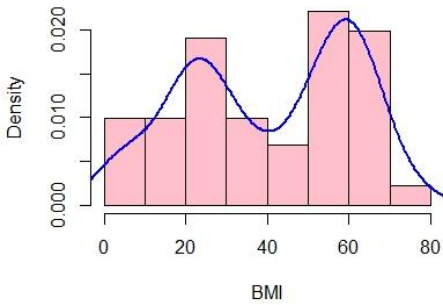
הוצאות ממשלתיות לבריאות כאחוז מסך ההוצאות הממשלתיות – משתנה רציף X_{11}

- עבור המשתנה המתאר את אחוז ההוצאות הממשלתיות על בריאות מתוך סך ההוצאות הממשלתיות במדינה מסוימת, ניתן לראות כי מדובר בהתפלגות הנוטה להיות מאוזנת וסימטרית המזכירה התפלגות גאוס. פונקציית הצפיפות ממחישה זאת על ידי כך ששיא ריכוז התצפיות נמצא במרכז ההתפלגות ונע בין ערך של 5-6. כלומר, ערך מדד הממוצע כמעט שווה לערך מדד החציון כפי שנטען בסעיף 5. עם זאת, ניתן להבחין בריכוז דל יותר של תצפיות בקצה הימני מאשר בקצה השמאלי, נתון המעיד על זנב ימני קטן.
- הצגה נוספת היכולה לאשש התפלגות סימטרית זו הינה פונקציית ההתפלגות המצטברת. ניתן להבחין בעלייה רציפה יחסית עד ערך 10, ולאחר מכן העלייה נעשית מתונה, דבר המעיד על הזנב הימני הקטן. באופן כולל, ניתן לומר כי התפלגות זו נוטה להזכיר צורת פעמון גאוס המעידה על התפלגות נורמלית וסימטרית, אך עם זאת חשוב לשים לב לזנב הימני שנוצר אשר מעיד על סתירה מסוימת למקרה התפלגות זה.



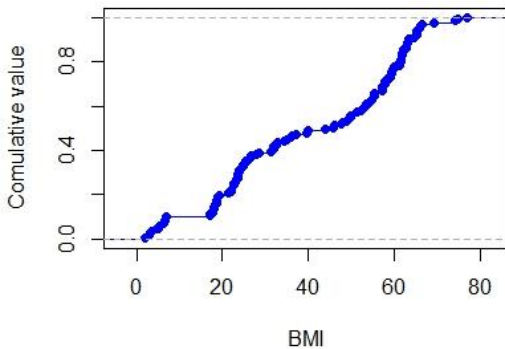
BMI – משתנה רציף X_8

- עבור המשתנה המתאר את מדד מסת הגוף הממוצעת של כלל האוכלוסייה במדינה מסוימת, ניתן לראות מספר



- דברים מעניינים. ראשית, ישנן שתי נקודות שיא של ריכוז התצפיות; האחת נמצאת בערך הנע בין 20-30, ואילו השנייה נמצאת בערך הנע בין 50-70. נקודת הקיצון הראשונה מתארת אוכלוסייה בעלת משקל תקין בממוצע עם סיכון נמוך באופן יחסי לחלות במחלות כל עוד אין מחלות רקע נלוות. נקודת הקיצון השנייה מתארת אוכלוסייה בעלת השמנה מדרגה 3 עם דרגת סיכון חמורה מאוד לחלות ממחלות. נתון זה עולה בקנה אחד עם הניתוח שנעשה בסעיף 5 אשר הצביע על ערכי מדדים של חציון ושל ממוצע הגבוהים מ-40 ורמזו על השמנת יתר בקרב מדינות רבות בעולם.

כמו כן, ניתן להבחין כי פונקציית הצפיפות הינה פונקציה א-סימטרית בעלת זנב שמאלי, דבר המעיד על ריכוז מועט של תצפיות באזורים המתארים אוכלוסיות הנמצאות בתת משקל ונמצאות בסיכון לתת תזונה.



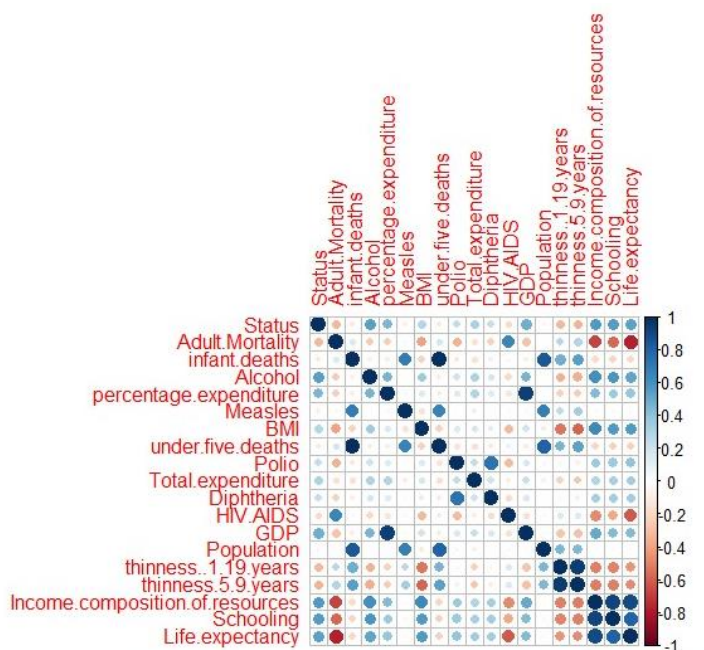
- הצגה נוספת היכולה לאשש התפלגות א-סימטרית זו הינה פונקציית ההתפלגות המצטברת. ניתן לראות עלייה מתונה עד ערך 20 המתאר תת משקל. לאחר מכן ניתן לראות עלייה חדה בין ערכים 20-30 המתארת את ריכוז התצפיות של המדינות בעלות ממוצע משקל תקין והתמתנות לאחריה. לבסוף, ניכרת עלייה חדה בין ערכים 50-70 המתארת את ריכוז התצפיות של המדינות בעלת ממוצע המשקל העודף. כל אלו מחזקים את ניתוח פונקציית הצפיפות, וכן עולים בקנה אחד עם ניתוח הנתונים שנעשה בסעיף 5.

סעיף 8 - ייצוג קשרים בעזרת תרשימים

ראשית, נציג את טבלת הקורלציות המציגה את עוצמת הקשרים בין המשתנים:

הערות:

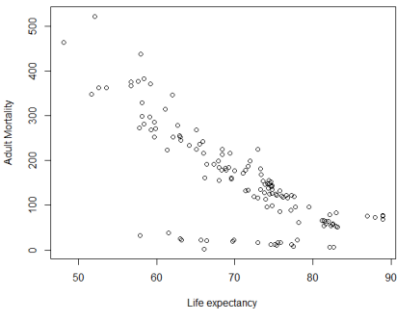
- גודל העיגול מעיד על עוצמת הקשר
- גוון העיגול מעיד על קיום הקשר. גוון כהה יותר – הקשר חזר יותר



הקשר בין life expectancy ל - adult mortality :

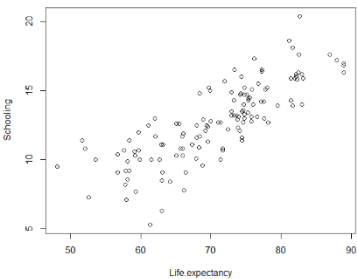
קיים קשר הפוך חזק בין מדד תמותת מבוגרים לבין המשתנה מוסבר המתאר את תוחלת החיים הממוצעת. קו המגמה הוא קו לינארי יורד, ומרבית התצפיות מרוכזות סביב קו זה. סוג הקשר הינו קשר סיבתי, זאת מכיוון שמדד תמותת המבוגרים רלוונטי לגילאי 15 עד 60 ואלו רוב שנות חייו של אדם. לכן, ככל שהתמותה בגילאים אלו גבוהה יותר, כך גם תוחלת החיים גבוהה יותר.

עם זאת, ניתן לראות כי בקצה העליון של תוחלת החיים, המגמה נעצרת – ניתן לראות כי ישנן תצפיות רבות של מדינות שתמותת המבוגרים בהם נמוכה מאוד, אך בכל זאת תוחלת החיים הממוצעת במדינות אלו לא עוברת את ה-85, ואילו דווקא במדינות שתמותת בהם קרובה ל-100, תוחלת החיים מגיעה ל-90.



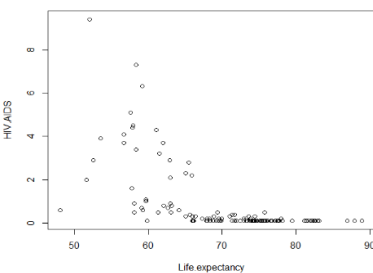
הקשר בין life expectancy ל - schooling :

קיים קשר ישר חזק בין מדד מספר שנות הלימוד הממוצע במדינה מסוימת לבין המשתנה המוסבר המתאר את תוחלת החיים הממוצעת. ככל ששנות הלימוד עולות – כך גם תוחלת החיים עולה. קו המגמה העולה מהגרף ברור לעין ומרבית התצפיות מרוכזות סביבו, עם מעט תצפיות חריגות. ניתן להסביר את הקשר בכמה דרכים: ראשית, ניתן לומר שככל שאדם יותר שנים במערכת החינוך, כך גדלה רווחתו ומכאן ישנה עליה בתוחלת החיים. אך הסבר זה הוא שברירי, ולדעתנו בסבירות גבוהה הקשר הוא מקרי: במרבית המדינות המפותחות בהן תוחלת החיים גבוהה יותר, מספר שנות הלימוד הממוצע גבוה יותר. יתכן שהקשר החזק נגזר משילוב של שנות הלימוד בבית הספר ושל מצב המדינה – מפותחת או לא מפותחת. לחיזוק ההשערה הזו, ניתן לראות כי ישנו קשר יחסית חזק (אם כי לא באופן חריג) בין status לבין תוחלת החיים.



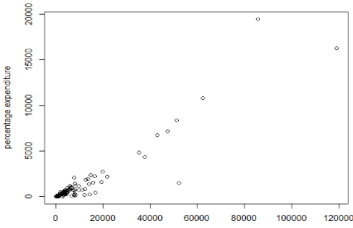
הקשר בין life expectancy ל - HIV :

קיים קשר הפוך בין מדד התמותה מאידס לבין המשתנה המוסבר המתאר את תוחלת החיים הממוצעת. קו המגמה הוא ברובו קו ישר, כלומר מדד זה בעל ערך של אפס ברוב המדינות. עם זאת, כאשר מדובר במדינה עם מדד תמותה מאידס גבוה מאפס, ניתן לראות כי ישנו פיזור גדול מאוד ואין קו מגמה ברור. כלומר, הקשר האיכותי בין המדדים הוא במדינות שבהן כמעט ואין תמותה מאידס. מכיוון שלא קיים חיסון לאידס, נוכל לתת שני הסברים אפשריים לקשר זה: הראשון, כיום אידס נחשבת למחלה כרונית ולא מחלה סופנית, קיים טיפול שמונע מהמחלה להתפרץ אצל נשאים. עם זאת, הטיפול קיים בעיקר במדינות מפותחות ועשירות – בהן תוחלת החיים גבוהה. שנית, ניתן להסביר זאת ע"י המודעות למחלה ולשימוש באמצעי מניעה במדינות בהן נמדדו התצפיות, ולקשר בין המודעות הכללית למניעת הידבקות לבין תוחלת החיים ולאו דווקא לבין אידס ותוחלת החיים.



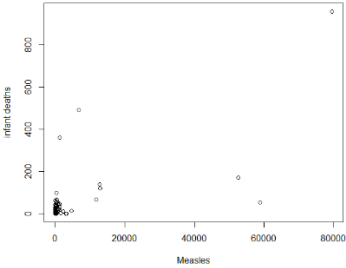
הקשר בין GDP ל - percentage expenditure:

ניתן לראות כי קיים קשר ישיר בין התוצר לאומי גולמי לנפש לבין ההוצאה לבריאות מתוך התוצר הלאומי הגולמי לנפש. קשר זה ניתן להסברה בקלות יחסית – ככל שהתוצר במדינה מסוימת גבוהה יותר, כך יש לה יותר כסף להשקיע ברווחת תושביה ובפרט במערכת הבריאות.



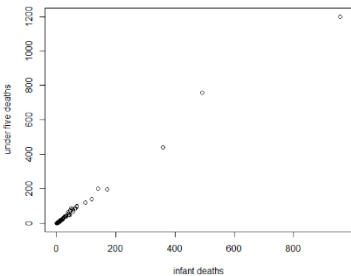
הקשר בין Measles לבין infant death:

ניתן לראות כי קיים קשר ישיר בין חצבת לבין מוות ילדים. אמנם מרבית הקשר מרוכז סביב האפס – דבר המעיד שחצבת היא מחלה שהולכת ונעלמת – אך בכל זאת ניתן לראות שככל שיותר אנשים חולים בחצבת במדינה מסוימת – כך גם מספר מקרי המוות מפעוטות גדול יותר. ניתן להסביר קשר זה בכך שחצבת היא מחלה שתוקפת בעיקר ילדים, והינה מחלה מסוכנת מאוד. ולכן לערכתנו, היא גורם ישיר לתמותת תינוקות.



הקשר בין infant death לבין under five death:

ניתן לראות כי קיים קשר ישיר בין תמותת תינוקות לתמותת ילדים מתחת לגיל 5, כאשר גם התצפיות החריגות נמצאות על אותו קו המגמה. ניתן להסביר זאת בכך שתמותת תינוקות – עד גיל שנתיים – מוכלת בתוך תמותת ילדים מתחת לגיל 5. אך גם עם זאת, ישנם את הילדים בין הגילאים 3-5 המדורגים במדד under five death, והקשר עדיין נשאר חזק. מכך אנו מסיקים כי אין הבדל משמעותי באחוזי תמותה בין פעוטות לבין ילדים מתחת לגיל 5, ושניתן לשקול להוריד את מדד תמותת התינוקות בפילוחים הבאים, בשל הקורלציה החזקה לתמותת ילדים מתחת לגיל 5.



סעיף 9 - טבלאות שכיחות

טבלאות שכיחות חד ממדיות:

א- טבלה המייצגת את מספר שנות הלימודים הממוצע בקרב מדינות בעולם - עד 6 שנים, בין 6 ל-12 שנים ומעל 12 שנים, כלומר השכלת בית ספר יסודי, תיכוני ואקדמי.

6	6-12	12+
1	51	79

ניתן לראות כי בקבוצה הראשונה (עד 6 שנות לימוד) ישנה רק תצפית אחת, וכל שאר הצפיות מרוכזות מעל 6 שנות לימוד, כאשר הרוב הוא 12 ומעלה. ניתן להסביר זאת בכך שישנו סטנדרט של 12 שנות לימוד ברוב המדינות המערביות ובמדינות העולם השני. ככל הנראה, מרבית התצפיות של 0-12 מגיעות ממדינות עניות.

ב- טבלת המייצגת את ערך ה-BMI הממוצע בקרב מדינות בעולם – עד 18.5, בין 18.5 ל-25 ומעל 25, כלומר תת מקל, טווח הנורמה וגבוה.

25+	18.5-25	-18.5
87	25	19

ניתן לראות שרק כחמישית מהתצפיות (המדינות) נמצאות בטווח הנורמה, ואילו שאר המדינות סובלות מעודף משקל או מתת משקל. ניתוח זה עולה בקנה אחד עם מדדי החציון והממוצע שבסעיף 5 וכן עם פונקציית

הצפיפות של משתנה זה בסעיף 7. לפי טבלה זו, ניתן לראות בבירור כי השמנת יתר היא בעיה גדולה בהרבה מתת משקל ולמעשה, פי 4.5 יותר גדולה.

טבלאות שכיחות דו ממדיות:

א- טבלה דו ממדית אשר מייצגת את כיסוי החיסונים נגד טטנוס ושעלת בקרבי ילדים עד גיל שנה לעומת מספר מקרי המוות מתחת לגיל 5.

מחזור החיסונים / מספר מקרי המוות	מתחת 50 אחוז	מעל 50 אחוז
0	1	29
מתחת 50	9	101
בין 50 ל-100	1	12
בין 100 ל-200	0	4
מעל 200	1	3

ראשית, ניתן לראות כי רוב המדינות נמצאות בתא אחד בטבלה – כאשר הן מחוסנות מעל ל-50 אחוז ויש מתחת ל-50 מקרי מוות מתחת לגיל 5. הנתון מראה שמדינות אלו אמנם מחוסנות, אך המחלה עודנה מחלה נגיפית שכיחה שנוטה להסתבך כאשר מדובר בילדים. זאת ועוד, כפי שראינו בסעיף 7

בהתפלגות הפעמון, ההוצאות הממשלתיות על בריאות מסך ההוצאות הממשלתיות ברוב המדינות אינו מאוד גבוה, דבר העולה בקנה אחד עם הטענה מקודם. בנוסף, ב-29 מדינות טטנוס ושעלת לא הורגות ילדים בכלל, וככל הנראה מדינות אלו הינן מדינות מפותחות בעלת תקציב בריאות גבוה וכלכלה חזקה.

ב- טבלה דו מימדית אשר מייצגת את סטטוס המדינה (מפותחת - 1 או מתפתחת - 2) לעומת צריכת אלוהול לנפש מעל גיל 15.

צריכת אלוהול לנפש / סטטוס	0.00518-3.046	3.046-6.082	6.082-9.118	9.118-12.154
Status 1	5	0	3	11
Status 2	77	14	18	3

מטבלה זו עולים כמה קשרים מעניינים. ניתן לראות כי ברוב המדינות מהמפותחות (57%) קיימת בעיה של שתייה מופרזת – בין 9.11 ל-12.15 ליטר מזוקק לנפש. לעומת זאת, במדינות המתפתחות המקרה הוא הפוך לקיצון השני; ברוב המדינות המתפתחות (68.7%) צריכת האלוהול היא נמוכה מאוד – בין 0.00518-3.046 ליטר מזוקק לנפש. ניתן להסביר זאת בכך שאלוהול הוא סמל סטטוס – משקאות יוקרתיים לעשירים או סם לבילוי במסיבות. בשל כך ובשל מחירו הגבוה של האלוהול, ככל הנראה קל יותר להשיג אלוהול במדינות מפותחות ולכן במדינות אלה ישנה בעיה מופרזת של שתיית אלוהול.