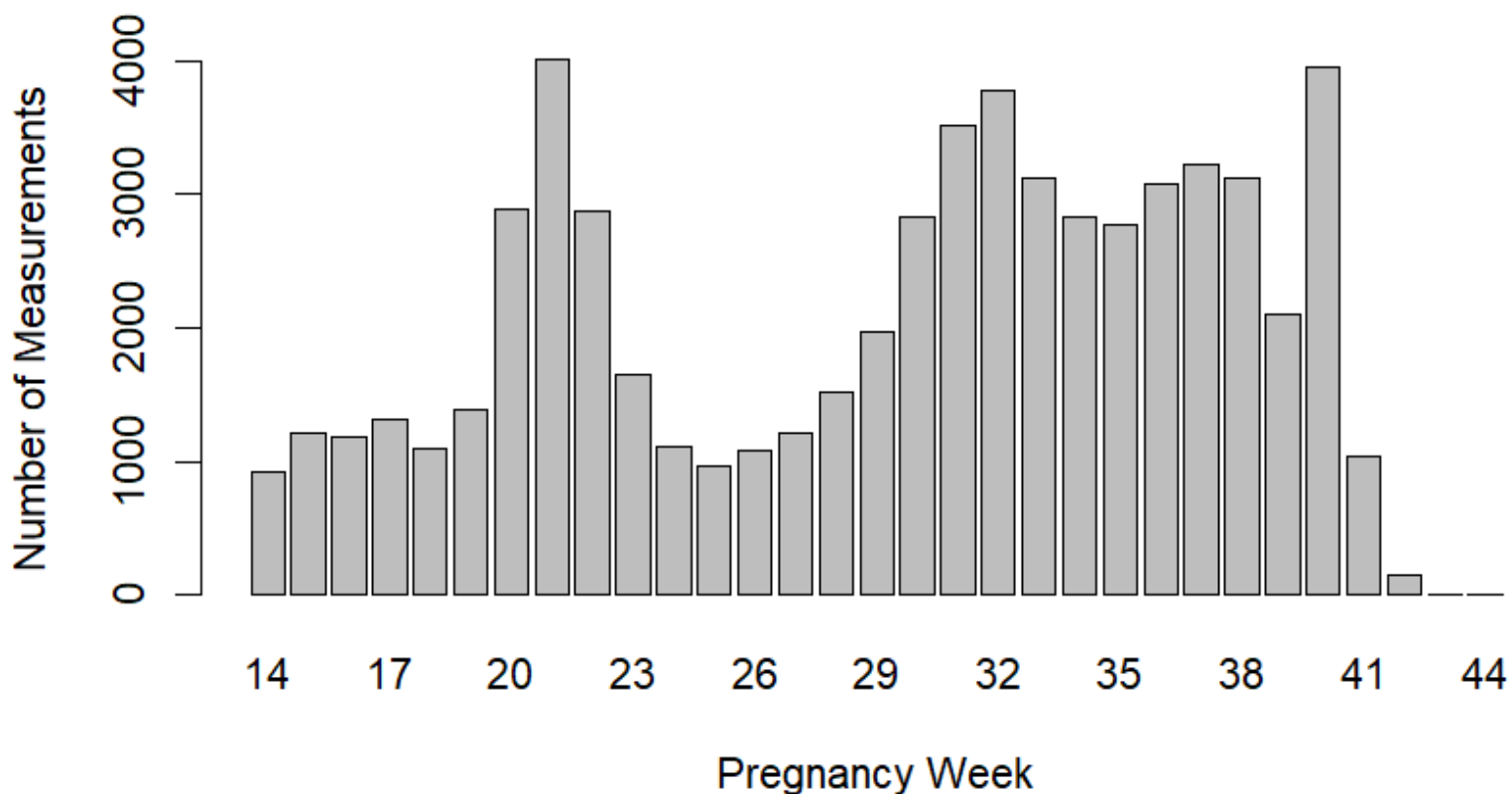


# יצירת עקומות גדילה לעוברים

קבוצה AC3  
מטלה 1

# הגבלת הנתונים ומספר התצפיות

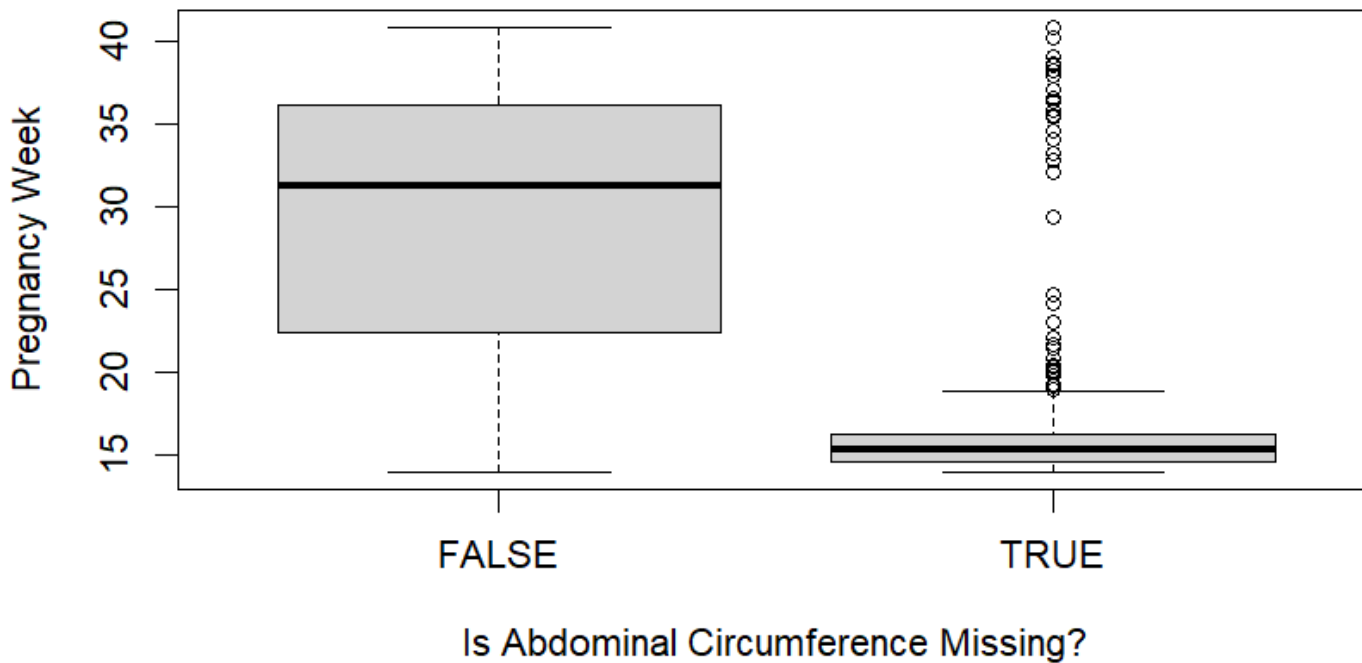
Number of Measurements per Pregnancy Week



- ❖ בתרשים ניתן לראות את מספר המדידות עבור כל שבוע בהריון.
- ❖ הגבלנו את הנתונים עד לשבוע 40 (כולל), כפי שהתבקשנו לעשות, והורדנו את המדידות שהתקיימו מהשבוע ה-41 (כולל) ואילך – סך הכל הסרנו 1,208 תצפיות.
- ❖ בהתמקדות בנתונים עד השבוע ה-40 (כולל), אנחנו מסתכלים על 15,451 נשים ועל 24,138 הריונות.

# מדידות חסרות

**Pregnancy Week by Missingness in Abdominal Circumference**

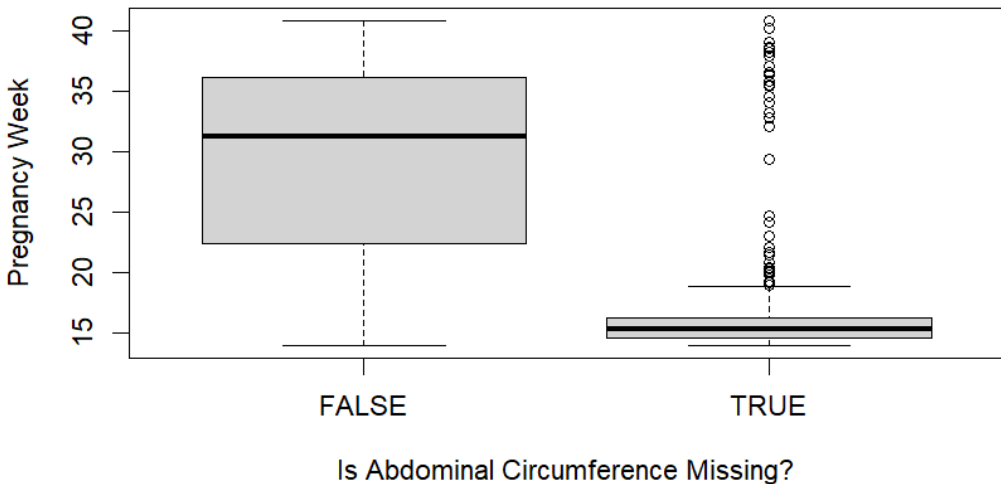


- ❖ בתרשים מוצגים החוסרים במדידות היקף הבטן של העוברים לפי שבוע ההריון
- ❖ לפני שנזרוק ערכים חסרים, נבדוק האם יש קשר בין המשתנה החסר (AC) לבין שבוע ההריון?
- ❖ מהבוקס-פלוט אפשר לראות שיש קשר משמעותי בין השניים:
- כאשר ה־ `ab_cir` לא חסר — (FALSE) התצפיות משתרעות על פני כל טווח ההריון.
- כאשר ה־ `ab_cir` חסר — (TRUE) רוב התצפיות הן בהריונות מוקדמים מאוד.
- ❖ ננסה להבין, מהם בדיוק השבועות של ההריונות המוקדמים? כי הבוקס-פלוט נותן מידע מוגבל בעמודת TRUE

# מדידות חסרות

```
> bp$stats  
      [,1]      [,2]  
[1,] 14.00000 14.00000  
[2,] 22.42857 14.57143  
[3,] 31.28571 15.28571  
[4,] 36.14286 16.28571  
[5,] 40.85714 18.85714
```

Pregnancy Week by Missingness in Abdominal Circumference



❖ עמודה 1 בטבלה מייצגת את FALSE ועמודה 2 את TRUE, כאשר: שורה 1 מייצגת את הערך המינימלי; שורה 2 את הרבעון הראשון; שורה 3 את החציון; שורה 4 את הרבעון השלישי; ושורה 5 את הגבול העליון לפני הערכים החריגים.

❖ ניתן לראות כי המדידה הנמוכה ביותר של שבוע היריון בתצפיות החסרות היא שבוע 14.

❖ חצי מהתצפיות החסרות הן עד שבוע 15.3

❖ השפיץ העליון הוא 18.86, כלומר הערך הגבוה ביותר לפני שמתחילים החריגים הוא בערך שבוע 19.

❖ השפיץ העליון הוא הערך הגדול ביותר שעדיין נחשב "לא חריג", כלומר כל ערך שהוא קטן או שווה ל-:  $Q3 + 1.5 * IQR$  כאשר  $IQR = Q3 - Q1$

❖ כל ערך שמעל הגבול הזה נחשב חריג (outlier) ומצויר בתור עיגול.

❖ מאחר שהחסרים ב־ AC מتركזים כמעט אך ורק בשבועות 14–18, והמדידות הופכות עקביות ורציפות רק החל משבוע 19, בחרנו לנתח את הנתונים החל משבוע זה.

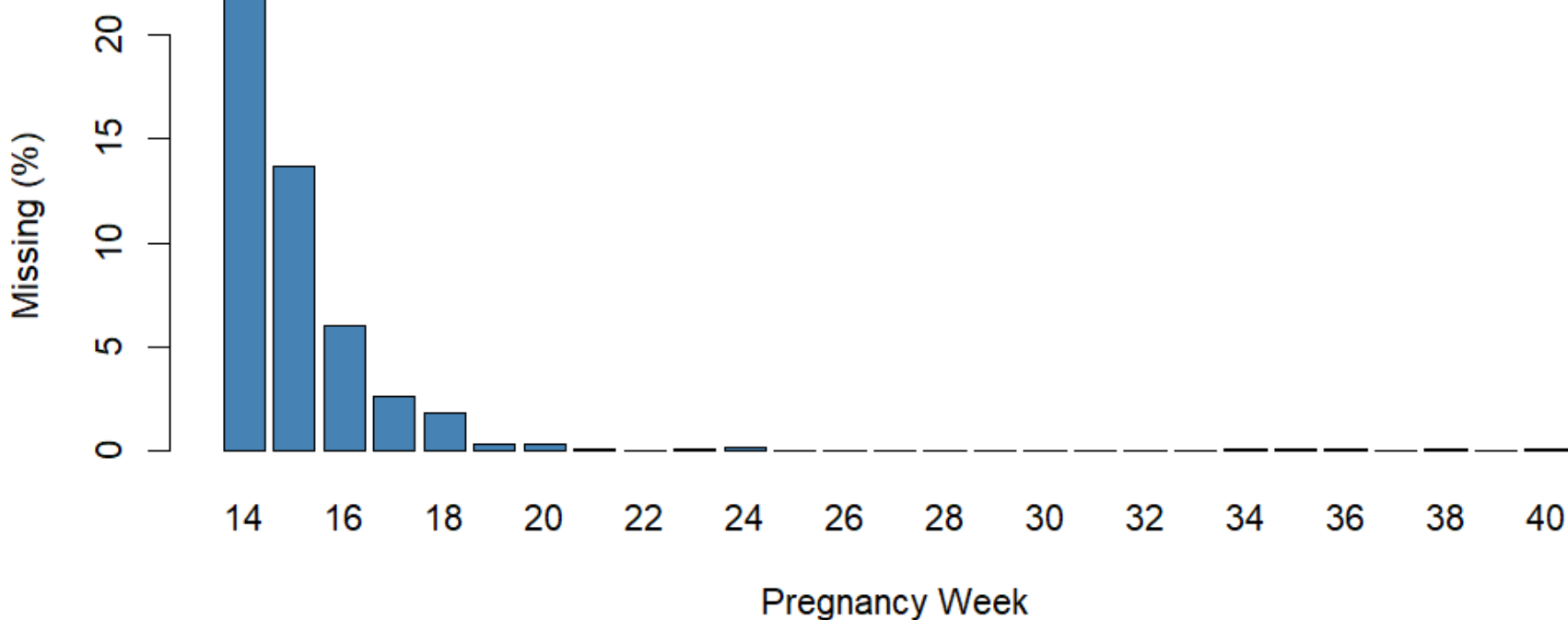
# התמודדות עם מדידות חסרות

	Week	Total	NonMissing	Missing	MissingPercent
14	14	929	727	202	21.74
15	15	1207	1042	165	13.67
16	16	1179	1108	71	6.02
17	17	1320	1285	35	2.65
18	18	1103	1083	20	1.81
19	19	1383	1378	5	0.36
20	20	2894	2885	9	0.31
21	21	4016	4012	4	0.10
22	22	2872	2871	1	0.03
23	23	1654	1653	1	0.06
24	24	1111	1109	2	0.18
25	25	972	972	0	0.00
26	26	1084	1084	0	0.00
27	27	1220	1220	0	0.00
28	28	1516	1516	0	0.00
29	29	1973	1972	1	0.05
30	30	2835	2835	0	0.00
31	31	3512	3512	0	0.00
32	32	3782	3780	2	0.05
33	33	3130	3129	1	0.03
34	34	2826	2824	2	0.07
35	35	2781	2778	3	0.11
36	36	3078	3076	2	0.06
37	37	3233	3232	1	0.03
38	38	3119	3115	4	0.13
39	39	2100	2099	1	0.05
40	40	3962	3958	4	0.10

- ❖ בטבלה מוצגים נתוני המדידות הקיימות והחסרות של היקף בטן העובר עבור כל שבוע, מהשבוע ה-14 ועד השבוע ה-40
- ❖ בנוסף, בעמודה הימנית, מוצג אחוז המדידות החסרות עבור שבוע ההריון המתאים
- ❖ ניתן לראות כי בשבועות 14-18 אחוז המדידות החסרות הוא גבוה, והחל משבוע 19 אחוז המדידות החסרות נמוך יותר
- ❖ לכן החלטנו להתחיל את הניתוח משבוע 19 (כולל)

# התמודדות עם מדידות חסרות

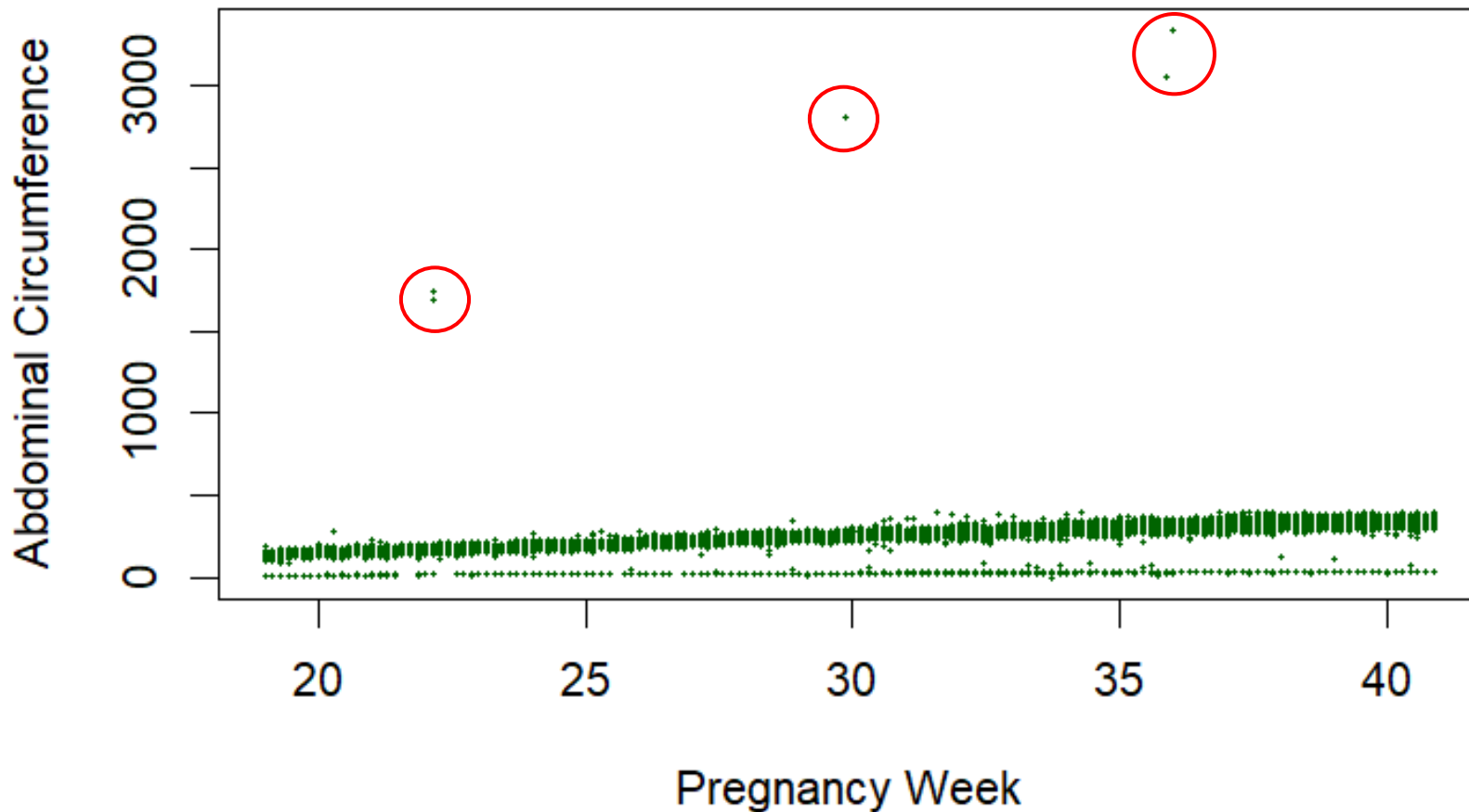
Percentage of Missing Abdominal Circumference Values by Pregnancy Week



❖ בתרשים מוצג אחוז החוסרים במדידות היקף בטן העובר עבור כל שבוע בהריון  
❖ המסקנה העולה מהתרשים  
הזה זהה למסקנה מהתרשים  
בשקף הקודם

# מדידות חריגות ויחידות מידה

Abdominal Circumference vs. Pregnancy Week ( $\geq 19$ )



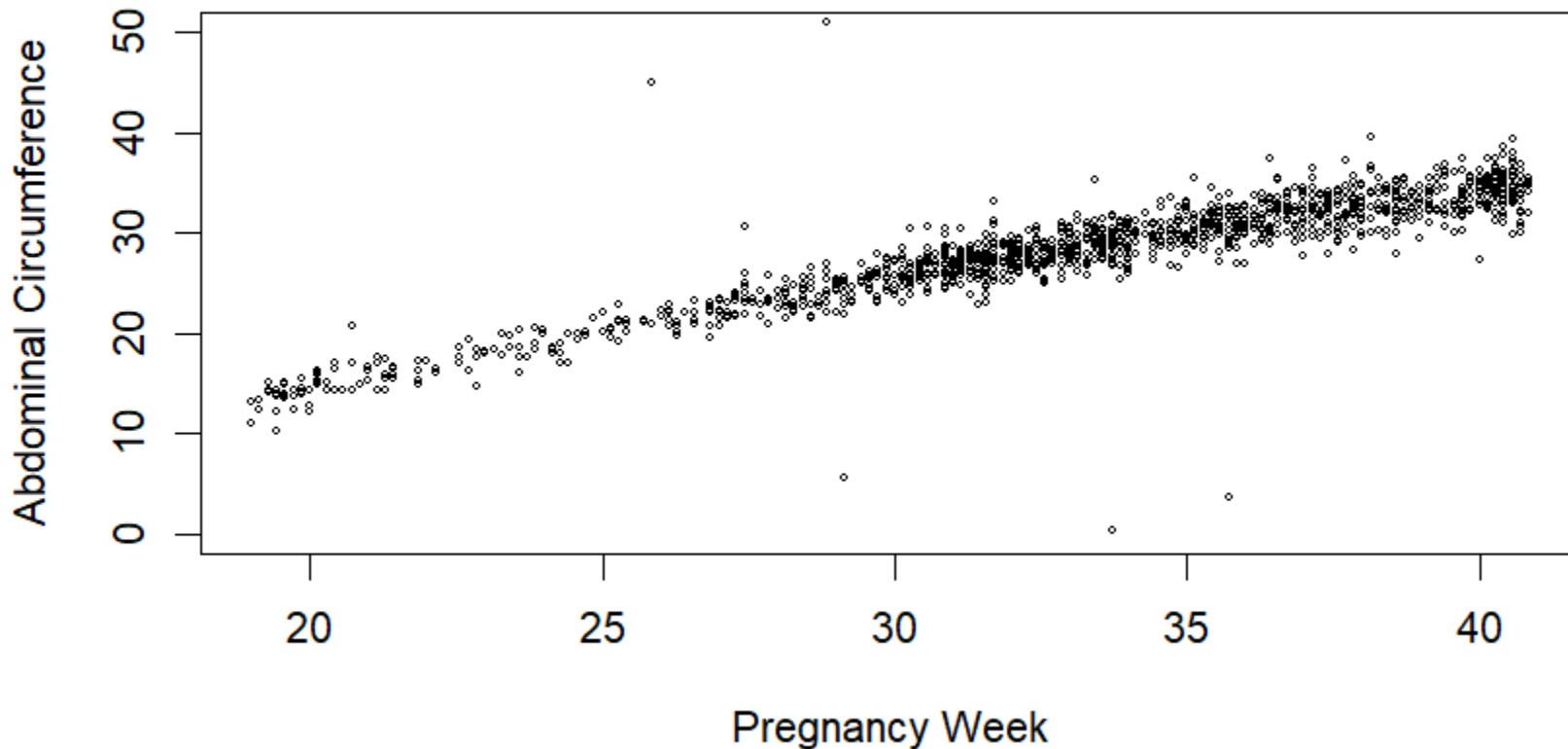
❖ בתרשים מוצגות מדידות היקף הבטן של העובר כנגד שבועות ההריון.

❖ יש מספר קטן של תצפיות חריגות (outliers), שמסומנות באדום, בהן מדידות היקף בטן העובר הוא מעל 1000. החלטנו להסירן.

❖ לא הוספנו יחידות מידה בגלל שה"עננים" השונים שמוצגים בתרשים מדגימים את הפער הקיים ביחידות המידה.

# מדידות חריגות ויחידות מידה

Abdominal Circumference vs Pregnancy Week (Zoomed In: 0–50)

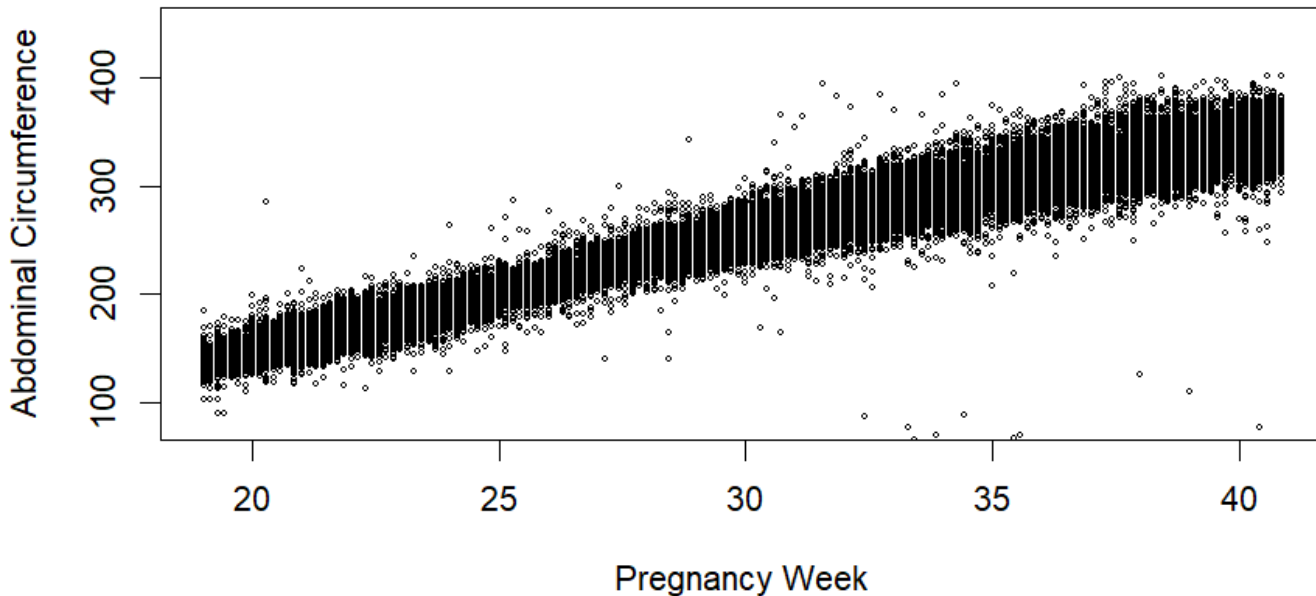


- ❖ בתרשים מוצגות מדידות היקף בטן העובר לאורך שבועות ההריון.
- ❖ בגלל התצפיות החריגות שמוצגות בתרשים בשקף הקודם, לא יכולנו לראות מה יחידות המידה.
- ❖ עשינו זום-אין כדי לראות את הענן הקטן מהתרשים הקודם בבירור.



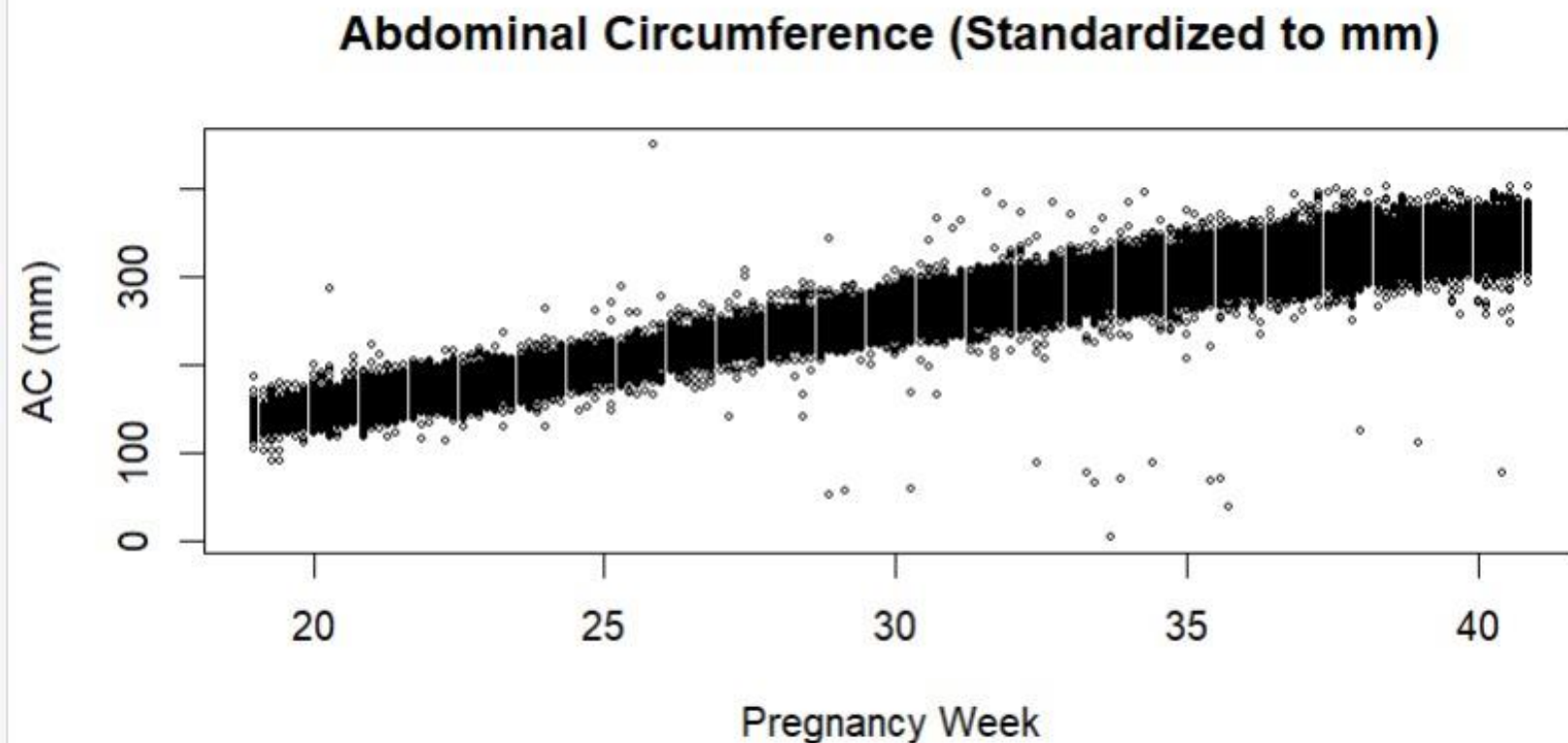
# מדידות חריגות ויחידות מידה

Abdominal Circumference vs Pregnancy Week (Zoomed In: 80–450)



- ❖ בתרשים מוצגות מדידות היקף בטן העובר לאורך שבועות ההריון.
- ❖ בגלל התצפיות החריגות שמוצגות בתרשים בשקף 7, לא יכולנו לראות מה יחידות המידה.
- ❖ עשינו זום-אין כדי לראות את הענן הגדול יותר מהתרשים הקודם בבירור.
- ❖ אחרי שבחנו את התרשימים בשקף הזה ובשקף הקודם, הסקנו כי בחלק מהמדידות השתמשו בס"מ ובחלק מהמדידות במ"מ.
- ❖ כדי לגשר על הפער, המרנו את כלל המדידות למ"מ.

# התמודדות עם מדידות חריגות



- ❖ בתרשים מוצגות מדידות היקף בטן העובר (מ"מ) לאורך שבועות ההריון.
- ❖ בדקנו כמה ערכים חריגים נשארו לנו, כלומר בכמה מהמדידות היקף בטן העובר קטן מ-100, או לחילופין גדול מ-400.
- ❖ סה"כ קיבלנו ש-19 מהתצפיות הן חריגות – מספר זניח מתוך 55,005 תצפיות סך הכל.

# בדיקת הנוסחה של הדלוק למדידת היקף בטן העובר

$$AC = -13.3 + 1.61 \times GA - 0.00998 \times GA^2$$

Input Unit: w (week)

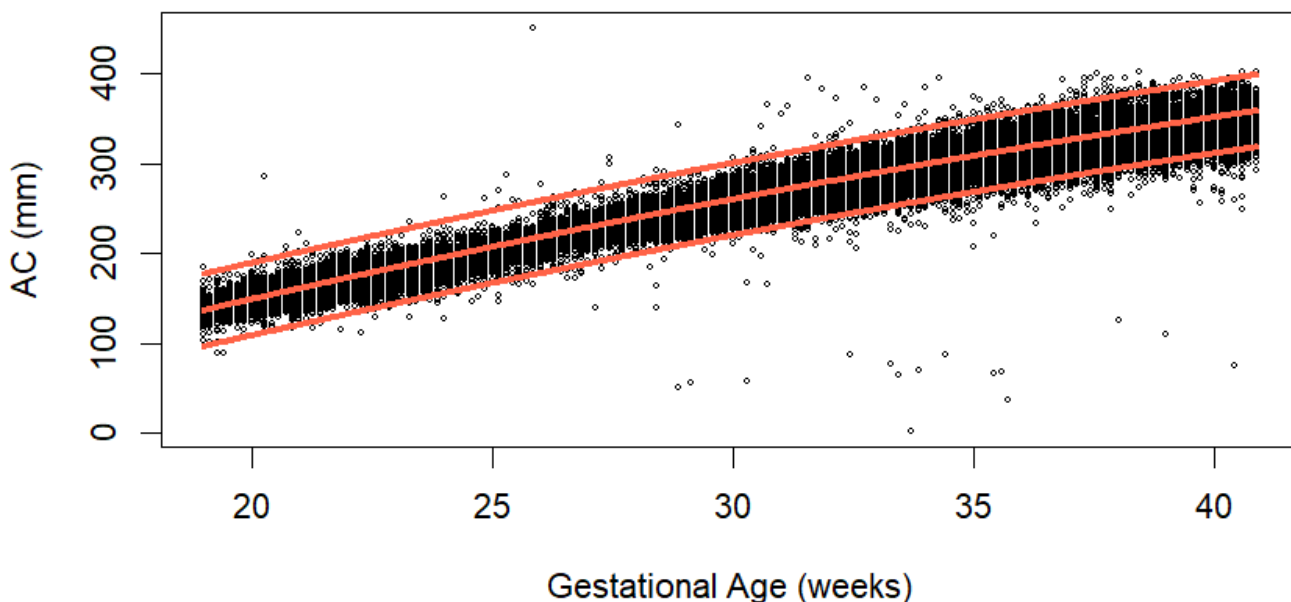
Output Unit: cm

Min Range: 12.0 weeks

Max Range: 40.0 weeks

Standard Deviation ( $\pm$ ): 1SD = 1.34 cm

Observed AC (mm) vs Hadlock  $\pm 3SD$



- ❖ בתרשים העליון מוצגת נוסחת הדלוק לחישוב הערך הצפוי של היקף בטן לפי גיל ההריון.
- ❖ יחידות המידה של הקלט הן שבועות, יחידות המידה של הפלט הן ס"מ. סטיית התקן היא 1.34 ס"מ.
- ❖ אנחנו עבדנו במ"מ ולכן בנוסחה שעשינו הכפלנו ב-10.
- ❖ כעת נתייחס לקווים האדומים שמופיעים בתרשים הפיזור שלנו: הקו האמצעי הוא הערך הצפוי לפי הנוסחה, הקווים החיצוניים מייצגים את החום של  $\pm 3$  סטיות תקן.
- ❖ ניתן לראות שנוסחת הדלוק ככל הנראה לא מתאימה לתצפיות שלנו – נבחן זאת לעומק.

# בדיקת התאמה להתפלגות נורמלית סביב מודל הדלוק

	Tail	Observed	Expected	Lower_CI	Upper_CI
1	Below 3%	0.08139260	0.030	0.02854529	0.03145471
2	Below 5%	0.11324425	0.050	0.04814144	0.05185856
3	Below 10%	0.18251068	0.100	0.09744171	0.10255829
4	Above 95%	0.02757931	0.050	0.04814144	0.05185856
5	Above 97.5%	0.01641669	0.025	0.02366862	0.02633138

❖ ביצענו ניתוח של שיעור התצפיות שנמצאות ב"זנבות" של ההתפלגות – כלומר, חריגות מהצפי של מודל הדלוק לפי סטיית התקן שניתנה (1.34 ס"מ).

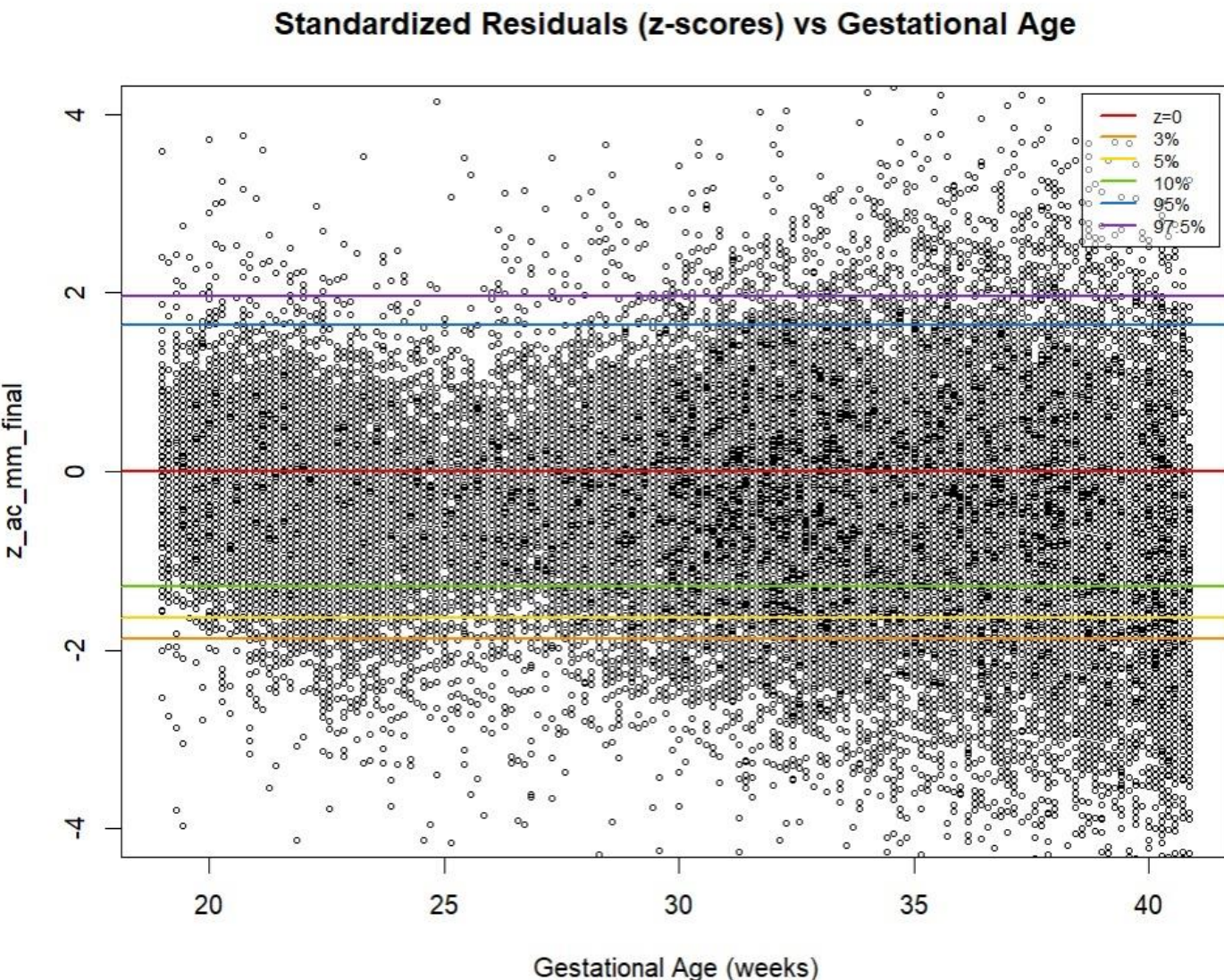
❖ בטבלה מוצגות: Observed - שיעור תצפיות בפועל שנמצאות מתחת או מעל אחוזונים מסוימים של ההתפלגות הנורמלית; Expected - שיעור צפוי תיאורטי (לפי ההתפלגות נורמלית תקנית; רווח סמך (95%) סביב הערך הצפוי, תחת ההנחה שהמודל תקף.

❖ מתחת לאחוזון 3 אנו מצפים לראות 3% מהעוברים. בפועל קיבלנו 8.14% - יותר מפי 2.5 מהצפוי, וגם מחוץ לרווח הסמך. גם ב- Below 5% ו- Below 10% יש עודף משמעותי של תצפיות נמוכות ממה שהמודל מנבא.

❖ הנתונים מראים הטיה שמאלית באוכלוסייה – כלומר, מדידות היקף הבטן בישראל נוטות להיות קטנות יותר ממה שמודל הדלוק מנבא.



# בדיקת שאריות סטנדרטיות מודל הדלוק



- ❖ בשלב זה חישבנו את השאריות המתוקננות (Z-scores) של כל מדידת AC תוך השוואה לערך הצפוי לפי נוסחת הדלוק ולפי סטיית התקן.
- ❖ כל נקודה מייצגת מדידת AC של עובר מסוים, לאחר חישוב Z-score
- ❖ הציר האופקי מייצג את שבוע ההיריון, והציר האנכי את השארית המתוקננת.
- ❖ הקווים האופקיים מייצגים אחוזונים מרכזיים בהתפלגות תקינה: קו אדום הוא הצפי לפי המודל; קווים צבעוניים נוספים: אחוזונים 3%, 5%, 10%, 90%, 95%, 97.5%.
- ❖ ניכרת אסימטריה ברורה: יש עודף ערכים נמוכים מאוד ( $Z < -2$ ) כמעט בכל שבועות ההיריון.
- ❖ חריגות גבוהות ( $Z > +2$ ) קיימות, אך שכיחות הרבה פחות.
- ❖ התקנון מאפשר לזהות חריגות מערכתיות ממודל הדלוק ומראה כי קיימת נטייה למדידות נמוכות יותר מהצפוי באוכלוסייה הישראלית – ממצא עקבי לאורך שבועות ההיריון.

# Kurmanavicius ניתוח זנבות במודל

Abdominal circumference (AC) - היקף בטן:

$$\text{mean} = -89.39 + 12.03w - 0.000863w^3$$

$$\text{SD} = 1.179 + 0.4753w$$

	Tail	Observed	Expected	Lower_CI	Upper_CI
1	Below 3%	0.009199164	0.030	0.02854529	0.03145471
2	Below 5%	0.014689574	0.050	0.04814144	0.05185856
3	Below 10%	0.031815290	0.100	0.09744171	0.10255829
4	Above 95%	0.069884556	0.050	0.04814144	0.05185856
5	Above 97.5%	0.036396691	0.025	0.02366862	0.02633138

❖ ביצענו ניתוח של שיעור התצפיות שנפלו בזנבות של התפלגות השאריות (Z-scores) לפי מודל קורמנביציוס. המטרה הייתה לבדוק האם הנתונים האמפיריים תואמים את הציפיות מהתפלגות נורמלית תקנית.

❖ Below 3%: שיעור התצפיות בפועל הוא 0.92%, לעומת ציפייה ל-3%. מובהק סטטיסטית – נמצא מתחת לרווח הסמך (3.14%-2.85%).

❖ Below 5%: שיעור בפועל ~1.47%, ציפייה: 5%. גם כאן, נמוך משמעותית ומחוץ לרווח הסמך.

❖ Below 10%: בפועל ~3.18%, ציפייה: 10%. פער גדול במיוחד, מחוץ לרווח הסמך (9.74%-10.26%).

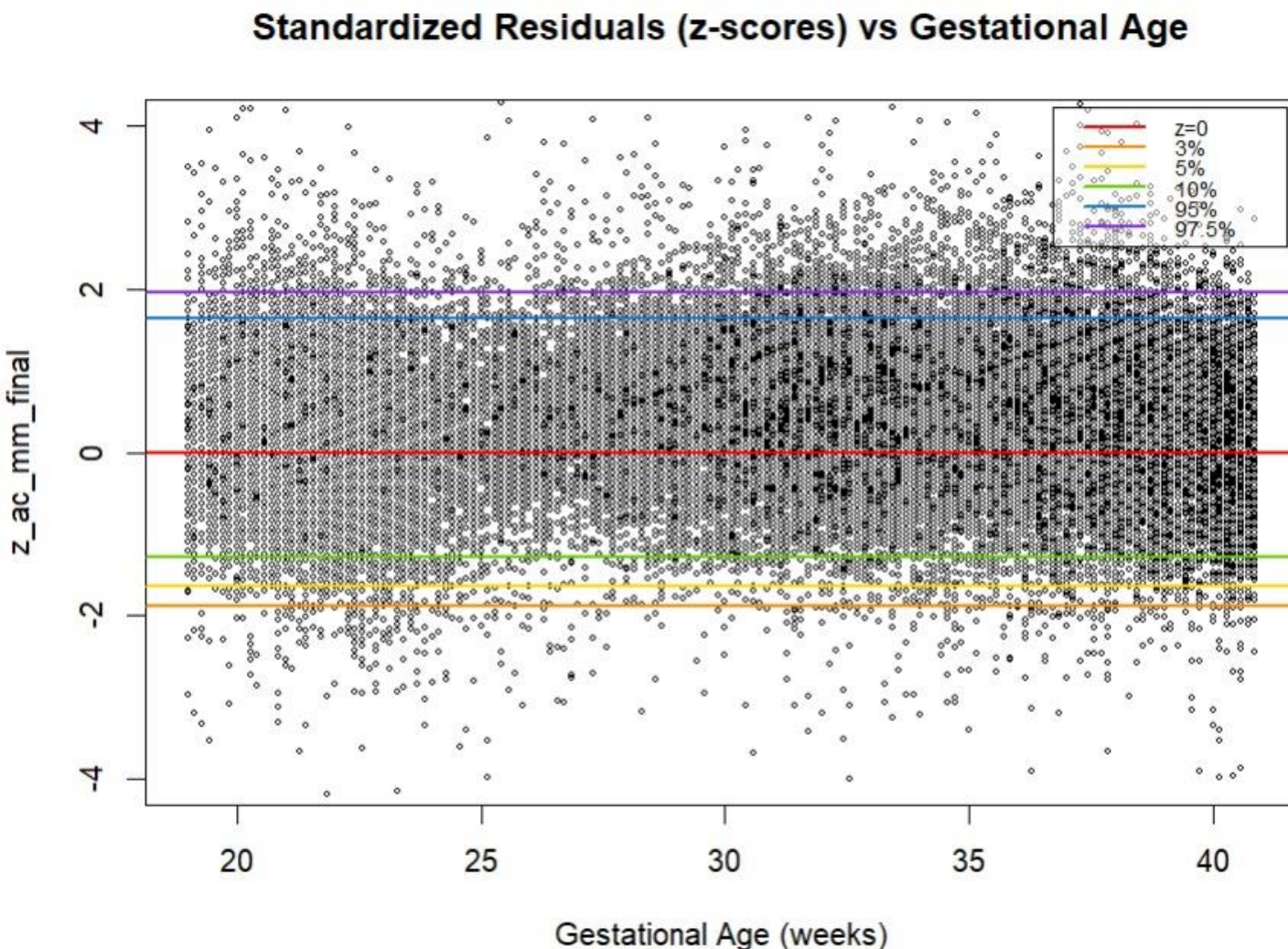
❖ Above 95%: שיעור בפועל ~6.99%, גבוה מהצפוי (5%) ומעל הרף העליון של רווח הסמך (5.19%) – כלומר גם עודף מובהק של ערכים גבוהים.

❖ Above 97.5%: שיעור בפועל ~3.64%, בעוד שהצפוי הוא 2.5%, ורווח הסמך נגמר ב-2.63%. גם כאן, מובהקות סטטיסטית כלפי מעלה.

❖ בשונה ממודל הדלוק שבו נמצאו יותר מדי ערכים נמוכים מהצפוי, במודל קורמנביציוס מתקבל דפוס הפוך: יש פחות מדי תצפיות נמוכות, ויותר מדי תצפיות גבוהות – באופן מובהק סטטיסטית.



# בדיקת שאריות סטנדרטיות מודל Kurmanavicius



❖ בשלב זה חישבנו את השאריות המתוקנות ( $Z$ -scores) של כל מדידת AC תוך השוואה לערך הצפוי לפי נוסחת Kurmanavicius ולפי סטיית התקן המשוערת לכל שבוע היריון.

❖ כל נקודה מייצגת מדידת AC של עובר מסוים לאחר חישוב  $Z$ -score הציר האופקי מייצג את שבוע ההיריון, והציר האנכי את השארית המתוקנת. הקווים האופקיים מייצגים אחוזונים מרכזיים בהתפלגות תקינה: קו אדום הוא הצפי לפי המודל, והקווים הצבעוניים מציינים את האחוזונים 97.5%, 95%, 10%, 5%, 3%.

❖ בניגוד למודל הדלוק, כאן ניכרת אסימטריה הפוכה: יש עודף ערכים חיוביים לאורך כל טווחי ההיריון, בעוד שמספר הערכים הנמוכים קטן מהצפוי.

❖ הדפוס מעיד כי מדידות AC בנתונים האמיתיים נוטות להיות גבוהות יותר מהתחזיות של קורמנביציוס, כלומר המודל מעריך בחסר את היקף הבטן של העוברים באוכלוסייה הישראלית.