

הגישה במאמר – רקע כללי

- הגישה המוצגת במאמר מבוססת על התאמת רגרסיה עבור התוחלת ועבור סטיית התקן של המדדים השונים כתלות בגיל העובר.
- **הנחה:** בכל שבוע היריון, המשתנה שאנחנו מודדים מתפלג נורמלית עם **תוחלת ו־סטיית תקן**, כאשר שתיהן תלויות בגיל ההיריון.
- **מוטיבציה:** לחשב עקומת אחוזון. שכן ידוע:
$$\text{centile} = \text{mean} + K * SD$$
כאשר K הוא ה-z-score של האחוזון הרצוי בהתפלגות נורמלית סטנדרטית.
- **הערה:** אם רואים שסטיית התקן **גדלה מאוד עם הגיל**, זה יכול להקשות על ההתאמה (הפיזור משתנה בצורה לא ליניארית). במקרה כזה, במאמר מציעים לנסות טרנספורמציה לוגריתמית על המשתנה המוסבר שלנו. אחרי שמחשבים את עקומות האחוזונים על סקאלת הלוג, פשוט עושים **אנטי־לוג** כדי לחזור ליחידות המקוריות.
- **אבל,** אם הטרנספורמציה יוצרת **עיוות בהתפלגות**, למשל "זנב" ארוך לערכים נמוכים, אז עדיף לוותר על הלוג ולעבוד עם הערכים המקוריים.

אז איך בונים את מודלי הרגרסיה?

1. מקשרים בין המשתנה המוסבר לגיל העובר בעזרת פולינום ממעלה שלישית:

$$Y = a + b \cdot \text{age} + c \cdot \text{age}^2 + d \cdot \text{age}^3 + \text{error}$$

כאשר מוצאים את המקדמים a, b, c, d בעזרת שיטת הריבועים הפחותים.

2. בודקים האם המקדם של age^3 , במקרה שלנו d , מובהק סטטיסטית, כלומר האם יש לו תרומה אמיתית. הקריטריון לכך:

$$\frac{|d|}{SE(d)} < 2$$

3. אם אי השוויון הנ"ל מתקיים, אז ל- d **אין** תרומה אמיתית. אז מסירים אותו. עושים התאמה חדשה, הפעם עם פולינום ריבועי:

$$Y = a + b \cdot \text{age} + c \cdot \text{age}^2$$

4. שוב בודקים עם c מובהק. אם לא אז נסיר גם אותו ונישאר עם קו ישר.

דיברנו רק על התוחלת, מה עם המודל לסטיית תקן?

1. לוקחים כל תצפית ומחסרים ממנה את הערך הצפוי בהתאם לשבוע הנתון

$$r_i = Y_i - \mu(GA_i)$$

כאשר Y_i הוא מדד כלשהו נצפה (היקף בטן, היקף ראש וכו') ו- $\mu(GA_i)$ היא הצבה בנוסחא שהתקבלה בעזרת השלב הקודם.

2. לוקחים ערך מוחלט ומכפילים ב-1.25

$$s_i = 1.25 \times |r_i|$$

3. נשאר לבדוק האם הפיזור משתנה עם הגיל?

נצייר גרף של s מול GA

אם הנקודות מפוזרות בלי מגמה ברורה אז סימן שהפיזור קבוע ולכן SD קבוע.

אם הנקודות עולות עם הגיל אז סימן שסטיית התקן גדלה עם הגיל ולכן צריך מודל שמנבא את SD לפי GA

4. אם התשובה לסעיף 3 חיובית אז נעשה רגרסיה של s על GA

$$s_i = \alpha + \beta \times GA_i + \varepsilon_i$$

איך בודקים שהמודלים באמת מתאימים לנתונים?

ראשית, נחשב ציוני Z לכל תצפית. לאחר מכן, ישנן שתי דרכים לבדוק האם המודלים מתאימים לנתונים

1. **Normal Probability Plot**: משרטטים גרף שבו על הציר האופקי שמות הערכים הצפויים מהתפלגות נורמלית, ועל הציר האנכי את ציוני ה- Z . אם הנקודות יושבות פחות או יותר על קו ישר סימן שהתפלגות ה- Z שלנו אכן נורמלית והמודל מתאים טוב

2. **ציוני Z מול גיל הריון**: נצפה לפיזור אקראי סביב 0