**מעבדה לסטטיסטיקה: אמידת מספר עותקים מדגימה בודדת (א)**

אנחנו רוצים להשתמש בפונקציית הGC שלמדנו ובערכים הנצפים כדי לאמוד את מספר-העותקים (Copy number) של הגנום.

בסיכום הזה נציג שיטת אמידה שדורשת רק אמידה של פונקציית ההשפעה של GC. לעיתים, קוראים לשיטות אמידה אלו ״תיקון״, משום שהן מתקנות את ההטיה הנוצרת מאפקט הGC.

**מודל המתאר קשר בין מספר עותקים למספר פרגמנטים**

נשתמש ב k=1…K לתאר את מספר התא, וב לתאר את הכיסוי הנצפה בתא.   
נסמן את התוחלת של התא הk ב

המודל עבור התוחלת צריך לסמן את הרכיבים שמרכיבים את התוחלת, ולאפשר לנו לתאר איך היינו מייצרים נתונים כאלו. נשים לב ל3 רכיבים שאנחנו יודעים שמשפיעים על הכיסוי:

* כמות העותקים הוא מספר בכפולות שלמות של חצי 0, ½, 1, 1 ½ , 2)...(. תוחלת הכיסוי צריכה להיות פרופורציונלית למספר העותקים.
* סה״כ כמות הפרגמנטים )ע״פ כל התאים) יכולה להשתנות מניסוי לניסוי, ואינה קשורה למספר העותקים אלא למכונות הריצוף, נסמנה בJ.
* לכמות הGC בתא (או נגזרות כמו כמות הGC בתתי התאים) יש השפעה על הכיסוי. תיארנו תלות זו בעזרת מודלים של רגרסיה.

אם כן, נוח לתאר את התוחלת כפונקציה כפלית של שלושה גורמים בלתי תלויים,  
כפול גורם לא ידוע :

לצורך זיהוי הפרמטרים צריך להניח הנחות לגבי הממוצעים של, ושל . נבחר לדרוש מהחציון של להיות 1, והממוצעים של הגורמים האחרים להיות 1:

ואז N הקבוע שמתקן בהתאם, כלומר

**אמידת מספר עותקים מהמודל**

כעת נניח שיש לנו אומדן לפונקציה ואת מספרי הפרגמנטים הנצפים , ואנחנו רוצים לאמוד את מספר העותקים .

(הפונקציה כפי שאמדנו בכיתה לא תקיים אלא  
 ולכן נצטרך לשים לב לקבועים בהמשך.)

ראשית נחלץ את מהמודל:

וכעת נשתמש באומד ההצבה כאשר נציב במקום ואת במקום .

**הערות**

1. ניתן להכפיל בעוד קבוע M[[1]](#footnote-1) שידאג שהחציון של האומדים יהיה 1:
2. ייתכן שיש עוד גורמים שתלויים ברצף הגנומי שמשפיעים על , כמו לדוגמה הסיכוי למפות פרגמנט, ואולי דרכי מדידה יותר עדינות ממספר הgc בתא. כדי לסמן מצב כזה, ניתן להשתמש ב בתיאור המודל במקום . ככל שנכניס יותר גורמים ל, כך המשמעות של תשתנה ובתקווה השונות שלו תקטן.

1. מגלם גם את 1/N וגם את הממוצע של [↑](#footnote-ref-1)