

מעבדה לסטטיסטיקה 2024, מטלה 3

הצגה ב 4.6

חלק א' השוואה למודל פוואסון אחיד

א. עבור בסיסים יחידים: חשבו את הכיסוי הממוצע (עבור בסיס יחיד) באזור שבחרתם במטלה 2. חשבו את ההתפלגות המצופה עבור מודל פוואסון אחיד. השוו את ההתפלגות המצופה להתפלגות הנצפית בעזרת גרף וכמו כן בעזרת טבלה. [השתמשו בטבלה בערכים 0-5 ו"גדול מ-5". דאגו לעגל ל-2 הספרות המשמעותיות].

ב. כעת, בחרו גודל תא (מעל 5000) וסכמו את מספר הפרגמנטים בכל תא (עבור כל כרומ' 1).

a. אמדו את המרכז והפיזור באופן חסין (robust). השוו את הפיזור הנצפה לפיזור הצפוי לפי מודל הפוואסון האחיד.

b. הציגו היסטוגרמה של מספר הפרגמנטים בתא. השוו ויזואלית למודל הפוואסון. האם מרכז ההתפלגות נראה כמו משתנה פוואסוני אחיד?

c. הציעו מדד מספרי שמתארת כמה שונה ההתפלגות הנצפית מפוואסון אחיד

הדרכה עבור ב'

- כדאי לנפות קודם תאים קיצוניים. ניתן לייצר קו "לוגי" שרושם TRUE אם הערך הנצפה בטווח הרצוי, FALSE עם לא. אז ניתן לאמוד היסטוגרמה רק עבור התאים הרצויים, ולבקש את משתנה density (freq=FALSE).
- כדי להשיג את ההתפלגות הצפויה, השתמשו בפקודה dpois עבור כל אחד מהגדלים. אח"כ, אחדו את ההסתברות של כל קבוצת ערכים שיתאימו להסטוגרמה.
- נסו למצוא דרך טובה ויזואלית להשוות את ההסטוגרמה להתפלגות התיאורטית.

חלק ב' השוואת GC לכיסוי (מספר reads) בתא

ב. בחרו אזור של כ-50 מיליון בסיסים, וגודל תא כמו במעבדה הקודמת.

- סכמו את מספר בסיסי ה-C+G בכל תא (בדומה למעבדה 1)
- מצאו את הכיסוי של כל תא (סכמו את מספר ה-reads כמו מעבדה 2)
- בדקו האם קיימת קורלציה בין מספר ה-GC בתא לבין הכיסוי. האם היחס חיובי? שלילי?
- ציירו על גבי גרף את ה-GC כנגד מספר ה-reads. מה אתם רואים? במה דומה ובמה שונה מהגרפים במאמר של Dohm? הוסיפו לגרף קו מגמה המאפיין את הקשר בין GC למספר ה-reads.
- חשבו: מה הקשר בין התוצאות שלכם בחלק ב' לבדיקות שעשינו בחלק א'?