בוחן קורס מעבדה - סמסטר ב׳ תשפ״ד

21.7.2024

מספר ת״ז:

הנחיות:

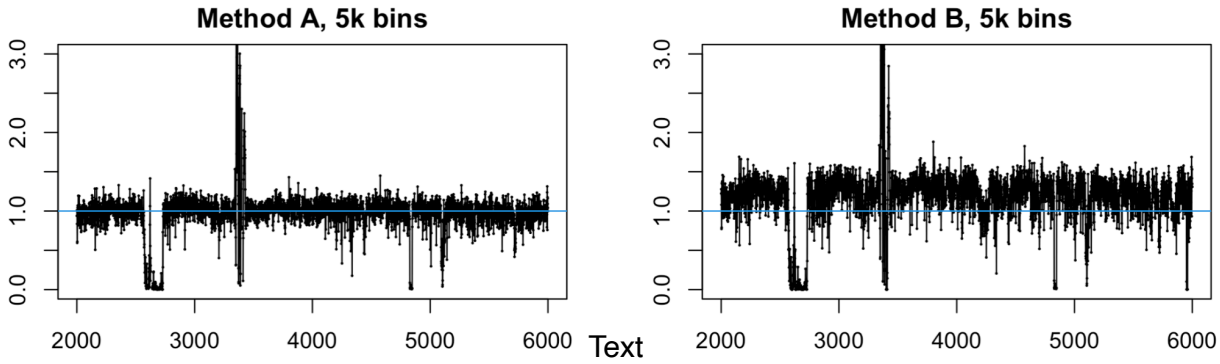
- **משך הבוחן שעה**
- **בבוחן 2 שאלות (סהכ 6 סעיפים). הניקוד לכל סעיף מופיע עליו.**
- **בבקשה לענות תשובות על גבי הטופס. אם חסר מקום, ניתן לענות בצד השני של הדף.**

אני מבטיח לשמור על מהימנות הבוחן.

מספר ת״ז _____

שאלה 1

מוצגים שני גרפים של אומדי מספר עותקים.
בציר הx מיקום בכרומוזום. בציר הy מספר עותקים ביחס למצב הבריא.



א. בגרפים אלו קיים קו מחבר בין כל שתי נקודות עוקבות.
רשמו יתרון אחד וחסרון אחד של הוספת הקווים המחברים על פני גרף נקודות בלבד. **[20 נק׳]**

ככלל - הקוים מדגישים את ההפרשים בין הנקודות ולא את הנקודות עצמן.

אם כן הם מדגישים אזורים בהם הרעש נמוך יחסית למגמה הרציפה (לדוגמה באזור התא ה2600 פה) [יתרון]
מהווה יתרון גם במקרה שרוצים להדגיש נקודות יוצאות דופן.

חסרון: כאשר הרעש יחסית גדול ההפרשים רועשים עוד יותר. כך הקוים מסתירים את הנקודות עצמן, ומדגישות אלמנטים פחות חשובים.

(סעיפים ב׳ וג׳ אנא הסבירו במילים. ניתן גם להראות על גבי הגרפים)

ב. הציעו תוספת לגרף <u>שתאפשר להשוות באופן וויזואלי</u> את <u>הפיזור</u> של שני התיקונים. **[20 נק׳]**

אפשר להוסיף קוים אופקיים המעידים על פיזור, לדוגמה רבעון 1 + 3 לכל גרף, או +- סטיית תקן מהממוצע.
אפשר גם להוסיף באופן דומה קוים המעידים על פיזור מקומי.

אפשר נוספת היא להוסיף גרף עם ההתפלגות השולית (לדוגמה תרשימי צפיפות) .

ציונים: הורדנו נקודות עבור הצעות שאינן קשורות בפיזור, לדוגמה קו מגמה שאיננו באחוזונים, וכו׳.

ג. הציעו שיפור אחד נוסף לגרפים **[10 נק׳]**

ניתן להוסיף קו מגמה (ממוצע רץ).
להרחיב את ציר הx
להוסיף ליד גרף משווה בין שני הגרפים (לדוגמה פיזור של שני האומדים, או השוואה של ההתפלגות השולית).
להוסיף תרשים בצד הגרף שיראה את ההתפלגות השולית (עדיף כך שוכב עם ציר y מתואם)

כמעט ולא הורדנו ציונים על הצעות כאן

שאלה 2

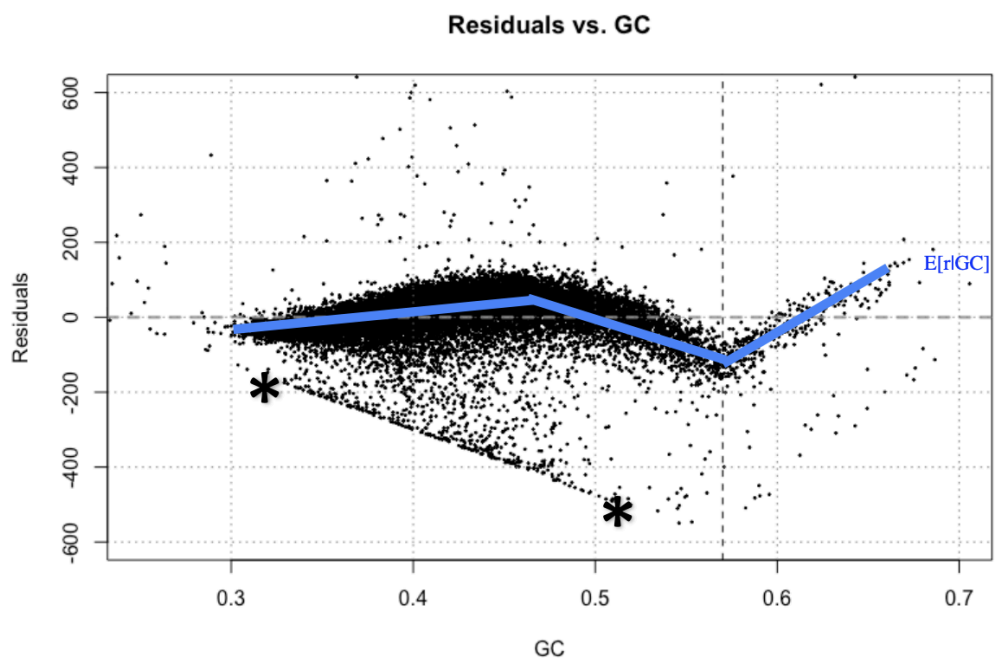
לצורך אמידת הקשר בין GC לבין כיסוי (מספר פרגמנטים בתא), יובל אומד קו מגמה בעזרת רגרסית spline מדרגה 1 עם knot בודד ב0.57=GC.
בגרף רואים את השאריות R כנגד GC. עליו יובל שרטט בקירוב את קו המגמה, כלומר את $E[R|GC]$.

```
Call:
lm(formula = reads ~ GC + I((GC - 0.57) * (GC > 0.57)), data = train_dat,
subset = (reads > 50) & (reads < 700))

Residuals:
Min       1Q  Median       3Q      Max
432.58  34.49   3.50    26.18-  502.56-

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)                    -399.991      2.898 -138.04   <2e-16 ***
GC                             1749.449      6.918  252.88   <2e-16 ***
I((GC - 0.57) * (GC > 0.57)) -5584.674     76.869  -72.65   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 57.23 on 22076 degrees of freedom
Adjusted R-squared:  0.7439   Multiple R-squared:  0.7439,
F-statistic: 3.207e+04 on 2 and 22076 DF,  p-value: < 2.2e-16
```



א. עבור ערכי GC>0.6, בכמה עולה התחזית של הכיסוי $\hat{f}(GC)$ עבור כל תוספת של 1% (0.01) בGC ?
[20 נק׳]

השיפוע של הגרף מודד שינוי של fhat עבור כל תוספת של 1 GC כלומר 100%.
לכן התשובה תהיה השיפוע חלקי 100.
בפועל, השיפוע בGC>0.6 מורכב מחיתוך של המקדם הלינארי (GC) והמקדם הלינארי הנוסף עבור GC>0.57
כלומר 1749 + 5584-
שווה ל 3835-
נחלק ב100 ונקבל 38.35

ב. לפי גרף השאריות, לאילו מהפונקציות נראה שיש שיפוע גבוה יותר עבור ערכי GC>0.6 ?
[15 נק׳]

א. התחזית $\hat{f}(GC)$
ב. הנצפית $f(GC)$
ג. לא ניתן לדעת

הסבר: גרף השאריות הוא y-yhat . אנחנו רואים שהשיפוע של השאריות גדול מ0 באזור של GC>0.6
(ראו את קו המגמה). זה סימן של f יש שיפוע גבוה יותר (חיובי יותר) מלf

ג. בתחתית גרף השאריות ניתן לראות נקודות מסודרות באלכסון (מסומן ב*). הסבירו מהן נקודות אלו.
[15 נק׳]

אלו השאריות המתאימות לתאים בהם y=0.

מי שרק אמרו שמדובר ב"ארטיפקט" אבל לא הסבירו מהו קיבלו ציון חלקי.