שקילות רגרסיה לוגיסטית

גיא עשירי-פרוסנר

האוניברסיטה העברית, מרכז ארצי לבחינות ולהערכה

11.02.2020

רקע: שקילות מודלים

- קבוצות אוכלוסיה נבדלות זו מזו בפרמטרים שונים:
  מין, גזע, גיל, תרבות, מצב סוציו-אקונומי...
- לכל אוכלוסיה ניתן להתאים מודל
- שקילות מודלים היא נושא חשוב מבחינה פסיכומטרית
- חשיבות מדעית (רוצים לדעת שהתופעה זהה בשתי קבוצות)
- בהנתן מודל קיים ואוכלוסיה חדשה, ייתכן שנרצה לשמור על המודל:
  - הוא מספק לנו הסבר טוב של התופעה
  - הוא מתאים לתאוריה
  - מומחים צברו מיומנות בשימוש בו
  - מחקרי המשך מתבססים עליו

רקע: שקילות מודלים

- החלפת מודל או ביצוע שינויים במודל הקיים עלולים לגרור:
  - אובדן כח מסבירני
  - התאמה פחות טובה לנתונים הישנים
  - פגיעה במהימנות המודל הישן והמחקרים הנסמכים עליו
- עבור השוואת מודלים של ניתוח גורמים, קיימת מערכת הכלים Measurement Invariance
- דוגמאות:
  - שקילות מודל 4 גורמי האישיות על בסיס מין (Byrne, 1988)
  - שקילות מודל 5 גורמי ההעדפות התעסוקתיות על בסיס גזע (Collins & Gleaves, 1998)
  - אי שקילות מודל גורמי של שאלון ערכי המשפחה (Byrne & van de Vijver, 2010)

רקע: רגרסיה לוגיסטית

מודל סטטיסטי המשמש לחיזוי ערכו של משתנה בינארי y_i על בסיס ערכיהם של משתנים רציפים $x_{1i}, \ldots, x_{pi}$, לדוגמא:

	x1	x2	x3	x4	x5	y_dyslexia
1	-2.05	-4.02	-3.52	-1.30	-0.89	0
2	-2.24	-2.79	-1.30	-0.63	-0.22	1
3	-1.75	-1.01	-3.35	-1.48	-0.79	0
4	-1.32	-2.54	-1.20	-1.50	-0.97	0
5	-0.73	0.24	-0.77	-1.60	-1.25	1
6	-3.08	-3.17	-3.56	-1.55	-3.01	1

לכל תצפית i מחושבים המנבא $\theta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}$

וההסתברות להימצאות לקות באופן

$$P(y_i = 1 | x_i) = \pi_i = \frac{e^{\theta_i}}{1 + e^{\theta_i}}$$

התחזית לכל תצפית נקבעת לפי נקודת חתך c, באופן $\hat{y}_i = \begin{cases} 1 & \hat{\pi}_i \geq c \\ 0 & \hat{\pi}_i < c \end{cases}$

רקע: רגרסיה לוגיסטית

Population A

$$\rightarrow \underbrace{\begin{pmatrix} 0.1 & 1.3 & 0 & \dots & 4.7 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ & & & & \end{pmatrix}_{n_A \times p}}_{X_A}, \underbrace{\begin{pmatrix} 1 \\ 0 \\ \vdots \end{pmatrix}_{n_A \times 1}}_{y_A} \rightarrow M^A: \underbrace{\begin{pmatrix} \hat{\beta}_0^A \\ \vdots \\ \hat{\beta}_p^A \end{pmatrix}}_{\hat{\beta}^A}, \hat{V}^A = Cov(\hat{\beta}^A), \hat{\theta}^A$$

Population B

$$\rightarrow \underbrace{\begin{pmatrix} 0.6 & 2.5 & 8 & \dots & 9.3 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ & & & & \end{pmatrix}_{n_B \times p}}_{X_B}, \underbrace{\begin{pmatrix} 1 \\ 1 \\ \vdots \end{pmatrix}_{n_B \times 1}}_{y_B} \rightarrow M^B: \underbrace{\begin{pmatrix} \hat{\beta}_0^B \\ \vdots \\ \hat{\beta}_p^B \end{pmatrix}}_{\hat{\beta}^B}, \hat{V}^B = Cov(\hat{\beta}^B), \hat{\theta}^B$$

רקע: השוואת מודלים של רגרסיה לוגיסטית

שלוש שאלות אפשריות בהשוואת מודלים של רגרסיה לוגיסטית:

1. כיצד המודל מתאר את התופעה?
   - נשתמש בוקטור מקדמי הרגרסיה $\hat{\beta}$
   - הוקטור מספק הצצה לאופן בו המודל מסביר את התופעה
2. באיזו מידה הפלט של המודל מתאים לתצפיות?
   - לכל תצפית x_i המודל מתאים את המנבא $\hat{\theta}_i$, ההסתברות $\hat{\pi}_i$ והתחזית $\hat{y}_i$
   - נבחר להשתמש ב-$\hat{\theta}_i$ כיוון שהוא על הישר הממשי
3. במצטבר, עד כמה מדויק המודל?
   - נשתמש בציון ברייר (Brier Score)

רקע: מבחני שקילות

- מרבית הכלים להשוואה בין מודלים מתבססים על מבחני מובהקות:
  - השערת האפס היא שלא קיים אפקט בין המודלים $(H_0: |\mu^A - \mu^B| = 0)$
  - האזור בו לא דוחים את השערת האפס מתכווץ ככל שהמדגם גדל
  - מצב זה מתמרץ שימוש במדגמים <u>קטנים</u> כדי להראות שאין הבדל
- הפתרון: שימוש במבחני שקילות
  - מגדירים רמת רגישות - גודל אפקט מינימלי להתייחסות $(H_0: |\mu^A - \mu^B| < \delta)$
  - הופכים את ההשערות - השערת האפס היא שקיים אפקט בגודל δ לפחות:

$$H_0: |\mu^A - \mu^B| \geq \delta, \qquad H_1: |\mu^A - \mu^B| < \delta$$

  - האזור בו דוחים את השערת האפס מתרחב ככל שהמדגם גדל
  - מצב זה מתמרץ שימוש במדגמים <u>גדולים</u> כדי להראות שאין הבדל

מטרת המחקר

- פיתוח שיטות להשוואת מודלים של רגרסיה לוגיסטית בין אוכלוסיות, בהסתמך על מאפיינים שונים של המודלים ועל ידי שימוש במבחני שקילות

שיטה: השוואת וקטור מקדמי הרגרסיה

- נניח כי קיימים לנו מודלים M^A,M^B עם וקטורי המקדמים $\hat{\beta}^A, \hat{\beta}^B$
- נגדיר את וקטור ההפרשים $\hat{q} = \hat{\beta}^A - \hat{\beta}^B$, מטריצת השונויות $Cov(\hat{q}) = S_q$
- נגדיר רמת רגישות בין מקדמים תואמים δ_β $\left(\left|\hat{\beta}_j^A - \hat{\beta}_j^B\right| \leq \delta_\beta\right)$
- עבור מודל עם p משתנים מנבאים, נבנה וקטור באורך p: δ_β^p
- מרחק Mahalanobis המותר בין הוקטורים: $\lambda_\beta^2 = {\delta_\beta^p}^T S_q^{-1} \delta_\beta^p$
- השערות:

$$H_0\colon \|\beta^A - \beta^B\|_\Sigma \geq \lambda_\beta, \qquad H_1\colon \|\beta^A - \beta^B\|_\Sigma < \lambda_\beta$$

- מבחן ברמה α לשקילות וקטורי המקדמים:

$$\left\{\hat{q}^T S_q^{-1} \hat{q} < \chi_{p,\alpha}^2\left({\delta_\beta^p}^T S_q^{-1} \delta_\beta^p\right)\right\}$$

שיטה: השוואת וקטור המנבאים

- נניח אוסף נתונים X_C בגודל k. עבור תצפית מסוימת $x_i \in X_C$, מודל M^A מפיק את המנבא $\hat{\theta}_i^A$ ואילו מודל M^B מפיק את המנבא $\hat{\theta}_i^B$
- יחס סיכויים: השינוי בהסתברות לקבל $\hat{y}_i = 1$ כאשר מחליפים מודל
- לוג יחס הסיכויים הוא $\hat{\xi}_i = \hat{\theta}_i^A - \hat{\theta}_i^B$. אנו מעוניינים בגודל השינוי ולא בכיוון: $\tilde{\xi}_i = \left|\hat{\theta}_i^A - \hat{\theta}_i^B\right|$
- נגדיר רמת רגישות לשינוי פר תצפית δ_θ
- השערות:

$$H_0: \mu_{\tilde{\xi}} \geq \delta_\theta, \qquad H_1: \mu_{\tilde{\xi}} < \delta_\theta$$

- מבחן ברמה α לשקילות וקטורי המנבאים:

$$\left\{ \frac{\sqrt{k}(\bar{\tilde{\xi}} - \delta_\theta)}{\sqrt{\widehat{Var}(\tilde{\xi})}} < t_{1-\frac{\alpha}{2},k-1} \right\}$$

שיטה: השוואת ציוני ברייר

- עבור תצפית מסוימת $x_i \in X_C$, מודל M^A מפיק את ההסתברות $\hat{\pi}_i^A$ ואילו מודל M^B מפיק את ההסתברות $\hat{\pi}_i^B$
- ציון ברייר (Brier Score) לביצועים של מודל M^A על אוכלוסיה X_C:

$$BS_{AC} = \frac{1}{k}\sum_{i=1}^{k}\left(\hat{\pi}_i^A - y_i\right)^2$$

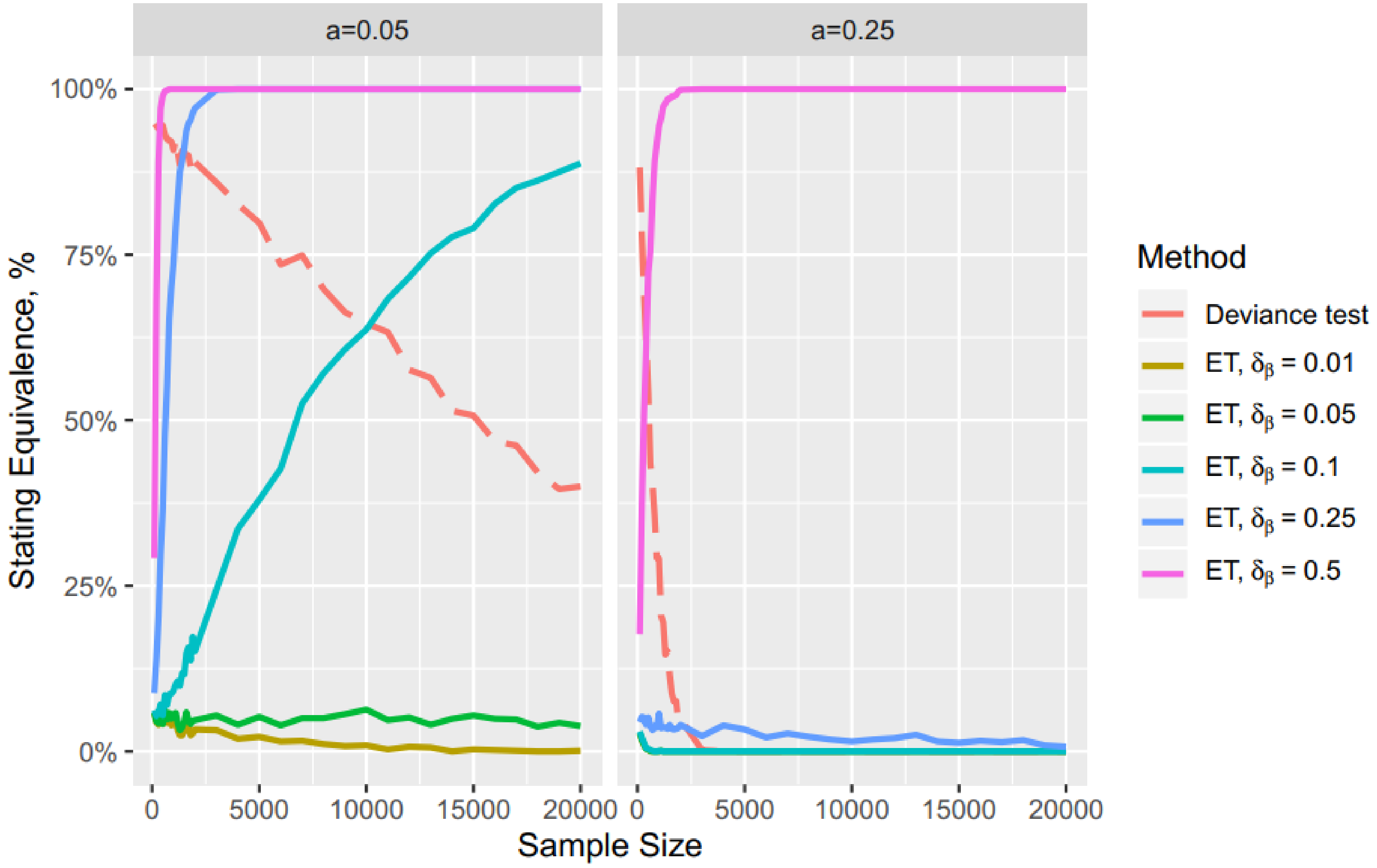
- ההפרש בין ביצועי המודלים: $D_{AB} = BS_{AC} - BS_{BC}$
- נגדיר רמת רגישות לשינוי בציון ברייר δ_B
- מבחן ברמה α לשקילות ציוני ברייר:

$$\left\{\frac{\sqrt{k}(D_{AB} + \delta_B)}{\sqrt{s_D^2}} > t_{1-\alpha,k-1} \quad \bigwedge \quad \frac{\sqrt{k}(D_{AB} - \delta_B)}{\sqrt{s_D^2}} < -t_{1-\alpha,k-1}\right\}$$

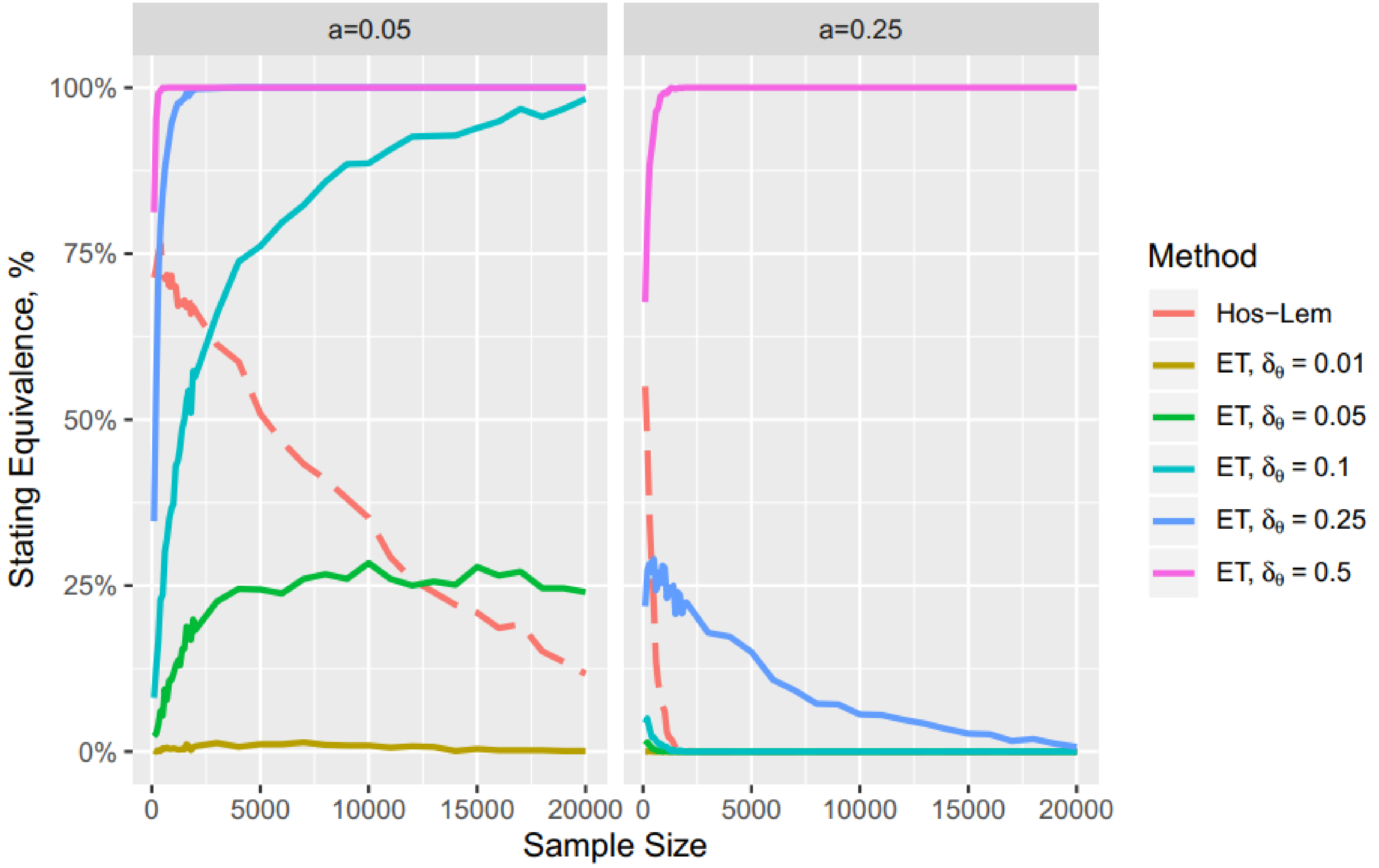
תוצאות: הגדרת הסימולציה

- בונים מדגמים x^A,x^B בגודל n מהתפלגות נורמלית סטנדרטית
- מגדירים אפקט בגודל a בין המדגמים: $\beta^A = \binom{0}{1}, \beta^B = \binom{a}{1-a}$
- מגרילים את וקטורי y^A,y^B לפי המודל הלוגיסטי עם $\theta_i = x_i^T \beta$
- בונים מודלים M^A,M^B
- שימוש בגדלי מדגמים $n = 100, \ldots, 2000, 3000, \ldots, 20000$
- גדלי אפקטים $a = 0.05, 0.25$
- לכל שילוב של גודל מדגם וגודל אפקט עורכים 1000 חזרות
- בכל חזרה משתמשים בשיטת השקילות ובשיטת השוואה קיימת

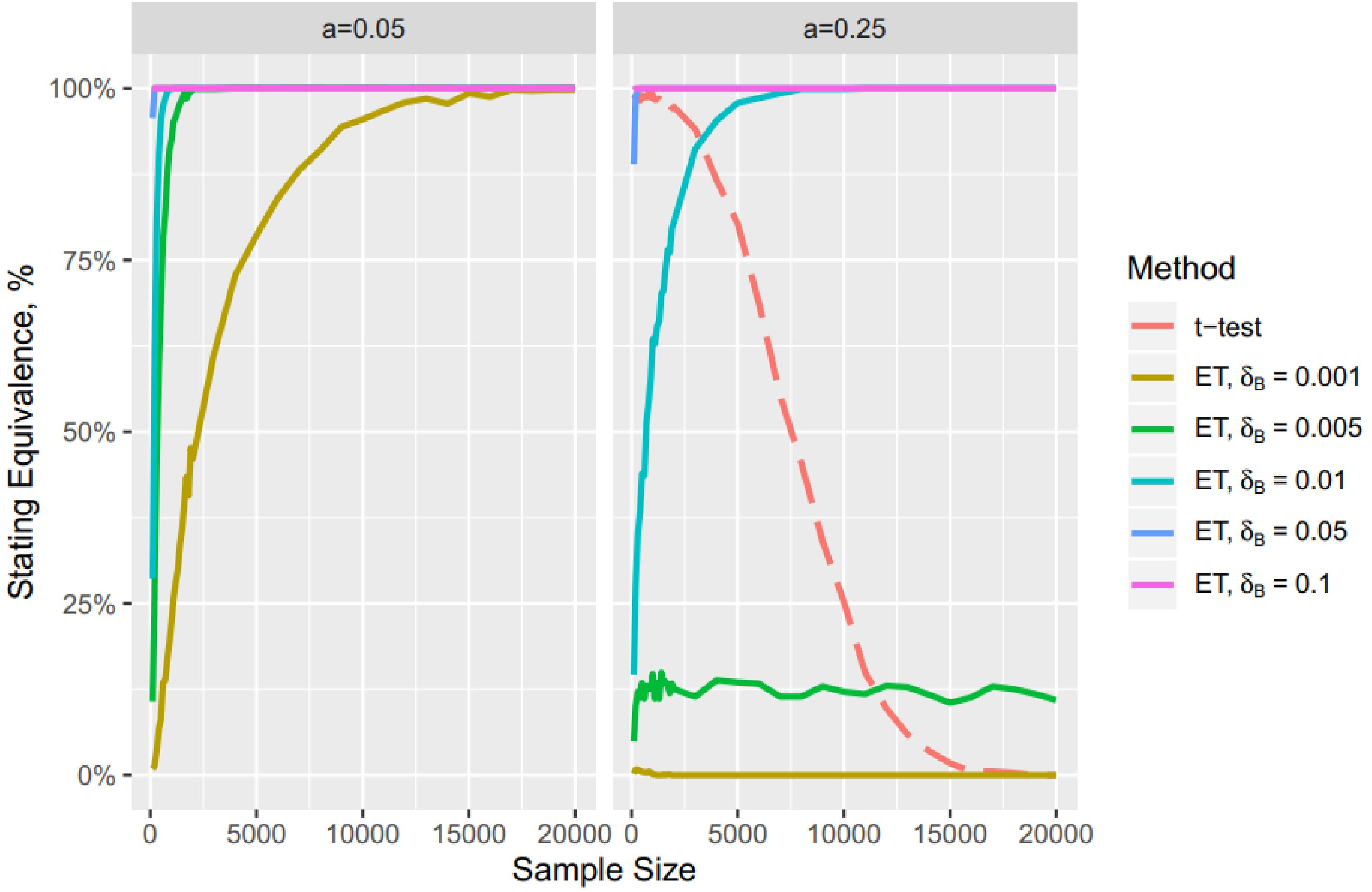
תוצאות: השוואת וקטור מקדמי הרגרסיה



תוצאות: השוואת וקטור המנבאים



תוצאות: השוואת ציוני ברייר



תוצאות: שימוש בנתוני מת"ל

- מת"ל היא מערכת לאבחון תפקודי למידה
- על בסיס 22 מטלות ושאלונים מופקות 4 אבחנות
- האבחנות הממוחשבות מופקות באמצעות רגרסיה לוגיסטית

- מדגם מחקרי: $n = 1046$, בחלוקה לפי מין: $n_f = 591, n_m = 455$
- לכל תת-קבוצה נבנו מודלים לניבוי דיסגרפיה ודיסקלקוליה

תוצאות: שימוש בנתוני מת"ל

- דיסגרפיה
  - שיטת שקילות מקדמים: המודלים **שקולים** בין המינים
  - שיטת שקילות מנבאים: המודלים **אינם שקולים** בין המינים
  - שיטת שקילות ציוני ברייר: המודלים **שקולים** בין המינים
  - מחקר fMRI (Berninger & O’Malley May, 2011): מרבית תפקודי הכתיבה שקולים בין המינים
- דיסקלקוליה
  - שיטת שקילות מקדמים: המודלים **אינם שקולים** בין המינים
  - שיטת שקילות מנבאים: המודלים **אינם שקולים** בין המינים
  - שיטת שקילות ציוני ברייר: המודלים **שקולים** בין המינים
  - מחקר (Devine et al, 2013): דיסקלקוליה התפתחותית שונה בין המינים

דיון

- כל שיטה מספקת תובנה שונה לגבי שקילות אפשרית בין המודלים
- שילוב השיטות מספק השוואה מקיפה
- שיטות פשוטות ליישום ושימוש

- ייתכן כי שיטת שקילות log-odds שמרנית מדי
- ייתכן כי שיטת שקילות ציוני ברייר מתירנית מדי, לעתים סותרת את השיטות האחרות
- בחירת רמת רגישות δ דורשת מיומנות רבה, זהו "קושי מתודולוגי עיקרי" של מבחני שקילות (Greene et al., 2008)

תודה על ההקשבה

שאלות?

guy@nite.org.il