פרק ג-1

יסודות ניתוח שונות גורמים מרובים.

מעבר לניתוח רב-גורמי

- בחלק זה של הקורס (חלק ג') נדון בהסקה לגבי ההשפעה המשותפת של מספר **גורמים**. (בחלק הקודם דנו בהסקה לגבי ההשפעה של סוג גורם יחיד, בעל a **רמות**).
- רוב הדיון יתמקד במקרה של 2 גורמים – ניתוח שונות דו-כווני. כאשר ישנם a סוגי **טיפולים** מהגורם הראשון ו – b סוגי טיפולים מהגורם השני.
- כאשר דנו בניתוח שונות חד-כווני, יכולנו להתייחס לטיפולים (1,…,a) גם כאוכלוסיות. בחרנו להשתמש במונח טיפול או אוכלוסיה לפי ההקשר.
- לעומת זאת, בניתוח שונות דו כווני ישנם 2 סוגים של טיפולים (טיפולים עבור הגורם הראשון, 1,…,a, וטיפולים עבור הגורם השני, 1,…,b). אם כך יש לנו בעצם a*b "אוכלוסיות" (כל אוכלוסיה מקבלת קומבינציה כלשהי של טיפול מהגורם הראשון וטיפול מהגורם השני), ולכן המונחים **טיפול ואוכלוסיה** הם בעלי משמעות שונה.

דוגמאות לניתוח שונות דו-כווני.

- דוגמא 1: רשת סופרמרקט מעוניינת לבדוק כיצד הסידור של מוצר מסוים במדף משפיעה על התצרוכת שלו. ניתן לסדר מוצר במדף תחתון (B), אמצעי בגובה (M) או עליון (T). בנוסף ניתן לסדר את המוצר במדפים בעלי רוחב רגיל (R) או רוחב רחב במיוחד (W).
  - הגורמים הינם:
    - גובה המדף בעל 3 רמות של טיפולים (B,M,T).
    - רוחב המדף בעל 2 רמות של טיפולים (R,W).

דוגמאות המשך:

- דוגמא 2: מעוניינים לבדוק את ההשפעה ארוכת הטווח של צריכה של אלכוהול ושימוש בכדורי שינה. בפרט מעוניינים לבדוק מהי רמת הריכוז של אנשים אשר צורכים אחד, שניים או לא צורכים חומרים אלו בכלל.
  - מהם הגורמים? מהם הטיפולים?

דוגמאות המשך...

- דוגמא 3: מהנדס מתכנן סוג חדש של סוללה ומתלבט האם להשתמש בחומר מסוג 1, 2 או 3. מטרתו היא לבחור את החומר אשר ייתן לסוללה אורך חיים ארוך. הסוללה אמורה לתפקד בתנאי סביבה שונים (טמפ' מינימלית 10- מעלות, טמפ' טיפוסית 20 מעלות, טמפ' מקסימלית 50 מעלות). מעבר לדרישה של אורך חיים ארוך המהנדס מעוניין בבחירת החומר אשר יגרום לכך שאורך חיי הסוללה לא יהיה מושפע מטמפ' הסביבה שבה עובדת הסוללה.

איזה נתונים נאסוף?

כאשר עבור כל קומבינציה של טיפולים יש תצפית אחת:

y_{11}	y_{12}	$\cdots$	y_{1b}
y_{21}	$\ddots$	$\cdots$	$\vdots$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
y_{a1}	$\cdots$	$\cdots$	y_{ab}

כאשר יש n תצפיות (עבור כל קומבינציה של טיפולים):

$y_{111}, y_{112}, \ldots, y_{11n}$	$y_{121}, y_{122}, \ldots, y_{12n}$	$\cdots$	$y_{1b1}, y_{1b2}, \ldots, y_{1bn}$
$y_{211}, y_{212}, \ldots, y_{21n}$	$\ddots$	$\cdots$	$\vdots$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$y_{a11}, y_{a12}, \ldots, y_{a1n}$	$\cdots$	$\cdots$	$y_{ab1}, y_{ab2}, \ldots, y_{abn}$

כאשר יש n_{ik} תצפיות:

$y_{111}, y_{112}, \ldots, y_{11n_{11}}$	$y_{121}, y_{122}, \ldots, y_{12n_{12}}$	$\cdots$	$y_{1b1}, y_{1b2}, \ldots, y_{1bn_{1b}}$
$y_{211}, y_{212}, \ldots, y_{21n_{21}}$	$\ddots$	$\cdots$	$\vdots$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$y_{a11}, y_{a12}, \ldots, y_{a1n_{a1}}$	$\cdots$	$\cdots$	$y_{ab1}, y_{ab2}, \ldots, y_{abn_{ab}}$

מדוע לארגן/לאסוף את הנתונים באופן הבא.

- ארגון הנתונים אשר בחרנו מעיד על כך שאנו לא משנים גורם אחד כל פעם בנפרד אלה מנסים את כל הקומבינציות האפשריות. (factorial design).
- ניתן לאסוף עבור כל גורם נתונים נפרדים ולבצע שני ניתוחי שונות חד-כוונים. אבל....
  - איסוף וניתוח נתונים שכזה מתעלם מהאפשרות לאינטרקציה (נתייחס בפרק הבא). וזאת יכולה להיות שגיאה חמורה.
  - באיסוף שכזה צריך לאסוף יותר תצפיות... (ראה חישובים בשקפים הבאים).

היתרון של ניסוי רב-גורמי על ניסוי ובו משנים גורם אחד בכל דגימה.

- ניקח כדוגמא ניסוי עם 2 גורמים ו 2 טיפולים בכל גורם:
  - גורם A, טיפולים A1 ו A2.
  - גורם B, טיפולים B1 ו B2.
- נניח ואנו אוספים כפי שהצגנו דגימה אחת מכל קומבינציה. ז"א אנו אוספים 4 תצפיות:

$y_{A_1B_1}$	$y_{A_1B_2}$
$y_{A_2B_1}$	$y_{A_2B_2}$

- חלופה אחרת היא לשנות בכל דגימה גורם יחיד ולאסוף 3 תצפיות:

$y_{A_1B_1}$	$y_{A_1B_2}$
$y_{A_2B_1}$	

- אבל בשביל להשוות את הדיוק הסטטסיטי שלנו, היינו רוצים שמספר התצפיות המינימאלי עבור כל טיפול יהיה זהה בשני השיטות ולכן בעצם בשיטה השנייה נאסוף (סה"כ 6 תצפיות):

$y_{A_1B_1}1, y_{A_1B_1}2$	$y_{A_1B_2}1, y_{A_1B_2}2$
$y_{A_2B_1}1, y_{A_2B_1}2$	

היתרון של ניסוי רב-גורמי (המשך)...

איסוף בשיטת
“factorial design”
(השיטה שלנו):

	ממוצע טיפול B1	ממוצע טיפול B2
ממוצע טיפול A1	$y_{A_1B_1}$	$y_{A_1B_2}$
ממוצע טיפול A2	$y_{A_2B_1}$	$y_{A_2B_2}$

מס' תצפיות עבור כל טיפול: 2

סה"כ תצפיות: 4

איסוף ע"י שינוי גורם
יחיד בכל דגימה:

	ממוצע טיפול B1	ממוצע טיפול B2
ממוצע טיפול A1	$y_{A_1B_1}1, y_{A_1B_1}2$	$y_{A_1B_2}1, y_{A_1B_2}2$
ממוצע טיפול A2	$y_{A_2B_1}1, y_{A_2B_1}2$	

מס' תצפיות מינימאלי עבור כל טיפול: 2

סה"כ תצפיות: 6

יעילות השיטה שלנו: $\frac{6}{4}=1.5$

עבור a טיפולים (עבור כל סוג גורם):

סה"כ תצפיות: a^2

סה"כ תצפיות: $(2a-1)a$

יעילות השיטה שלנו: $\frac{(2a-1)a}{a^2}=2-\frac{1}{a}\approx 2$

היתרון של ניסוי רב-גורמי (עבור ניתוח שונות L כווני כאשר יש 2 טיפולים לכל גורם)...

איסוף ע"י שינוי גורם יחיד בכל דגימה:	איסוף בשיטת "factorial design" (השיטה שלנו):
$N=6$ $\begin{array}{c\|c} y_{A_1B_1}1, y_{A_1B_1}2 & y_{A_1B_2}1, y_{A_1B_2}2 \\ \hline y_{A_2B_1}1, y_{A_2B_1}2 & \end{array}$	$L=2$ $N=4$ $\begin{array}{c\|c} y_{A_1B_1} & y_{A_1B_2} \\ \hline y_{A_2B_1} & y_{A_2B_2} \end{array}$
$N=16$ $\begin{array}{c\|c} y_{A_1B_1C_1}1, y_{A_1B_1C_1}2, y_{A_1B_1C_1}3, y_{A_1B_1C_1}4 & y_{A_1B_2C_1}1, y_{A_1B_2C_1}2, y_{A_1B_2C_1}3, y_{A_1B_2C_1}4 \\ \hline y_{A_2B_1C_1}1, y_{A_2B_1C_1}2, y_{A_2B_1C_1}3, y_{A_2B_1C_1}4 & \end{array}$ וגם $y_{A_1B_1C_2}1, y_{A_1B_1C_2}2, y_{A_1B_1C_2}3, y_{A_1B_1C_2}4$ מס' תצפיות (מינימאלי) עבור כל טיפול: 4	$L=3$ $N=8$ $\begin{array}{c\|c} y_{A_1B_1C_2} & y_{A_1B_2C_2} \\ \hline y_{A_2B_1C_2} & y_{A_2B_2C_2} \end{array}$ $\begin{array}{c\|c} y_{A_1B_1C_1} & y_{A_1B_2C_1} \\ \hline y_{A_2B_1C_1} & y_{A_2B_2C_1} \end{array}$ מס' תצפיות עבור כל טיפול: 4
$N=(1+L)2^{L-1}$	L כללי (2 טיפולים עבור כל גורם): $N=2^L$ מס' תצפיות עבור כל טיפול: 2^{L-1}

יעילות השיטה שלנו: $\dfrac{(1+L)2^{L-1}}{2^L}=\dfrac{1}{2}+\dfrac{1}{2}L$

מודל ניתוח שונות דו-כווני

$$y_{ikj} = \underbrace{\mu + \tau_i + \beta_k + (\tau\beta)_{ik}}_{\mu_{ik}} + \tilde{\varepsilon}_{ikj}$$

$i = 1,\dots,a$ ← a טיפולים עבור הגורם הראשון (אינדקס i).

$k = 1,\dots,b$ ← b טיפולים עבור הגורם השני (אינדקס k).

$j = 1,\dots,n$ ← n תצפיות עבור כל קומבינציה (בניסוי מאוזן)

$$\sum_{i=1}^{a} \tau_i = 0 \ , \ \sum_{k=1}^{b} \beta_k = 0$$

$$\sum_{i=1}^{a} (\tau\beta)_{ik} = 0 \ , \ \sum_{k=1}^{b} (\tau\beta)_{ik} = 0$$

$$\tilde{\varepsilon}_{ikj} \sim NID(0, \sigma^2)$$

נרצה לבדוק:

"אין השפעה לגורם הראשון"

$H_0 : \tau_1 = \dots = \tau_a = 0$

$H_1 : otherwise$

"אין השפעה לגורם השני"

$H_0 : \beta_1 = \dots = \beta_b = 0$

$H_1 : otherwise$

"אין אינטרקציה בין הגורם הראשון לשני"

$H_0 : (\tau\beta)_{ik} = 0 \quad \forall i, k$

$H_1 : otherwise$