פרק ב-2

החשבון של ניתוח שונות, פירוק סכומי ריבועים.

מה נעשה בפרק זה?

- כפי שכבר הבנו בפרק הקודם, שיטות "ניתוח שונות" מאפשרות לנו להסיק על הבדלים בין תוחלות של אוכלוסיות.
- אז מה המשמעות של ניתוח שונות?
- נראה כיצד ניתן **<u>לפרק</u>** את כלל השונות של המדגם למרכיבי שונות שונים ובאמצעות מרכיבים אלו להסיק לגבי היחסים בין התוחלות באוכלוסיה.
- בנוסף נבין את המשמעות של מושג דרגות חופש.

המודל והמדגם...

המדגם

$$\begin{array}{ccccc|l}
y_{11} & y_{12} & y_{13} & \cdots & y_{1n} & \text{1 אוכלוסיה} \\
y_{21} & y_{22} & \cdots & \cdots & y_{2n} & \text{2 אוכלוסיה} \\
\vdots & \vdots & & & \vdots & \\
y_{a1} & y_{a2} & \cdots & & y_{an} & \text{a אוכלוסיה}
\end{array}$$

המודל

$$y_{ij} = \underbrace{\mu + \tau_i}_{\mu_i} + \tilde{\varepsilon}_{ij}$$

$$i = 1, \dots, a$$

$$j = 1, \dots, n$$

$$\sum_{i=1}^{a} \tau_i = 0$$

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

סכומים וממוצעים

$$\begin{matrix} y_{11} & y_{12} & y_{13} & \cdots & y_{1n} & \longrightarrow & \sum_{j=1}^{n} y_{1j} = y_{1\bullet} & \bar{y}_{1\bullet} = \frac{y_{1\bullet}}{n} \\ y_{21} & y_{22} & \cdots & \cdots & y_{2n} & \longrightarrow & \sum_{j=1}^{n} y_{2j} = y_{2\bullet} & \bar{y}_{2\bullet} = \frac{y_{2\bullet}}{n} \\ \vdots & \vdots & & & \vdots & & & \\ y_{a1} & y_{a2} & \cdots & & y_{an} & \longrightarrow & \sum_{j=1}^{n} y_{aj} = y_{a\bullet} & \bar{y}_{a\bullet} = \frac{y_{a\bullet}}{n} \end{matrix}$$

$$\sum_{i=1}^{a} y_{i\bullet} = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} = y_{\bullet\bullet}$$

$$N = a \cdot n$$

$$\bar{y}_{\bullet\bullet} = \frac{y_{\bullet\bullet}}{N} = \frac{\sum_{i=1}^{a} \bar{y}_{i\bullet}}{a}$$

סכום ריבועי הפרשים מהממוצע

$$\begin{matrix} y_{11} & y_{12} & y_{13} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & \cdots & y_{2n} \\ \vdots & \vdots & & & \vdots \\ y_{a1} & y_{a2} & \cdots & & y_{an} \end{matrix}$$

שונות המדגם הכוללת

$$S_{\bullet\bullet}^2 = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\bullet\bullet})^2}{\underbrace{N-1}_{an-1}}$$

$$SS_{Total} = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\bullet\bullet})^2 = (N-1)S_{\bullet\bullet}^2$$

זהו סך סכום הריבועים של הפרשים מהממוצע או בקצרה
"סכום הריבועים הכולל"

פרוק סכום הריבועים הכולל

$$SS_{Total}=\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{\bullet\bullet})^2=\ldots$$

קצת אלגברה

$$\ldots=\overbrace{n\sum_{i=1}^{a}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2}^{SS_{Treatments}}+\overbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2}^{SS_E}$$

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{\bullet\bullet})^2=\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet}+\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2=\sum_{i=1}^{a}\sum_{j=1}^{n}\left((y_{ij}-\bar{y}_{i\bullet})+(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})\right)^2=$$

$$\sum_{i=1}^{a}\sum_{j=1}^{n}\left((y_{ij}-\bar{y}_{i\bullet})^2+2(y_{ij}-\bar{y}_{i\bullet})(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})+(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2\right)=$$

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2+\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2+2\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})=$$

$$\sum_{j=1}^{n}\sum_{i=1}^{a}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2+\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2+2\sum_{i=1}^{a}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})\underbrace{\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})}_{0}=$$

סכום הפרשים ממוצע הוא 0

$$\underbrace{n\sum_{i=1}^{a}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2}_{SS_{Treatments}}+\underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2}_{SS_E}$$

דרגות החופש בכל סכום ריבועים

$$SS_{Total} = SS_{Treatments} + SS_E$$

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{\bullet\bullet})^2 \qquad n\sum_{i=1}^{a}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2 \qquad \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2$$

סכום הריבועים הכולל

סכום הריבועים בין הקבוצות

סכום הריבועים בתוך הקבוצות

דרגות חופש: $N-1=a\cdot n-1 \quad = \quad a-1 \quad + \quad a\cdot(n-1)=an-a$

פרוק סכום הריבועים מפרק גם את דרגות החופש

מקורות השונות תחת H0 ו- H1

$$SS_{Total} = SS_{Treatments} + SS_E$$

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{\bullet\bullet})^2 \qquad n\sum_{i=1}^{a}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2 \qquad \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2$$

סכום הריבועים הכולל

סכום הריבועים בין הקבוצות

סכום הריבועים בתוך הקבוצות

	SS_E	$SS_{Treatments}$	SS_{Total}
H0:	*עקב שגיאה מקרית*	*עקב שגיאה מקרית*	*עקב שגיאה מקרית*
H1:	*עקב שגיאה מקרית*	*עקב הפרשים בין תוחלות ושגיאה מקרית*	*עקב הפרשים בין תוחלות ושגיאה מקרית*

ממוצעי ריבועים...

$$S_{\bullet\bullet}^2 = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\bullet\bullet})^2}{N-1}$$

$$MS_{Treatments} = \frac{SS_{Treatments}}{a-1} = \frac{n\sum_{i=1}^{a}(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2}{a-1}$$

$$MS_E = \frac{SS_E}{N-a} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i\bullet})^2}{N-a}$$

$S_p^{\,2}$ הוא בעצם MSE

$$MS_E = \frac{SS_E}{N-a} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2}{N-a}$$

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2 = \sum_{i=1}^{a}\left[\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2\right]$$

$$S_i^2 = \frac{\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2}{n-1}$$

$$SS_E = \sum_{i=1}^{a}(n-1)\left[\frac{\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2}{n-1}\right] = \sum_{i=1}^{a}(n-1)S_i^2$$

$$MS_E = \frac{\sum_{i=1}^{a}(n-1)S_i^2}{N-a} = \frac{(n-1)S_1^2+(n-1)S_2^2+\ldots+(n-1)S_a^2}{(n-1)+\quad\ldots\quad+(n-1)}$$

הסטטיסטי F

$$F_0 = \frac{SS_{Treatments}/(a-1)}{SS_E/(N-a)} = \frac{MS_{Treatments}}{MS_E}$$

"נוסחאות עבודה"

$$SS_{Total} = \sum_{i=1}^{a}\sum_{j=1}^{n} {y_{ij}}^2 - \frac{{y_{\bullet\bullet}}^2}{N}$$

$$SS_{Treatments} = \frac{1}{n}\sum_{i=1}^{a} {y_{i\bullet}}^2 - \frac{{y_{\bullet\bullet}}^2}{N}$$

$$SS_E = SS_{Total} - SS_{Treatments} = \sum_{i=1}^{a}\sum_{j=1}^{n} {y_{ij}}^2 - \frac{1}{n}\sum_{i=1}^{a} {y_{i\bullet}}^2$$

בנוסף ניתן לקודד תצפיות (להפחית את הממוצע מכל התצפיות)

יתרונות במדגמים מאוזנים

- עד כה דיברנו על ניתוח שונות מאוזן (balanced).
- למדגמים מאוזנים יש יתרון בכך שפילוג סטטיסטי המבחן יחסית אינו רגיש לסטיות מההנחה של נורמאליות זהה (ללא הוכחה). לא כך עבור מדגמים לא מאוזנים.
- יתרון נוסף הוא מקסימיזציה של עוצמת המבחן (ע"י איזון). (גם ללא הוכחה).

מדגמים לא מאוזנים (unbalanced)

- לפעמים המדגמים אינם מאוזנים.
- הרבה פעמים (בנתונים ניסויים) עקב אובדן/פסילה של תצפיות.
- בניתוח שונות חד-כווני "הכל מסתדר".

$$y_{ij} = \underbrace{\mu + \tau_i}_{\mu_i} + \tilde{\varepsilon}_{ij}$$

$$i = 1, \ldots, a$$

$$j = 1, \ldots, n_i$$

לכל קבוצת טיפול גודל מדגם משלה

$$\sum_{i=1}^{a} \tau_i = 0$$

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

$$N = \sum_{i=1}^{a} n_i$$

$$\begin{array}{ccccc} y_{11} & y_{12} & y_{13} & \cdots & y_{1n_1} \\ y_{21} & y_{22} & \cdots & y_{2n_2} & \\ \vdots & \vdots & & & \vdots \\ y_{a1} & y_{a2} & \cdots & & y_{an_a} \end{array}$$

$$SS_{Total} = \sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{\bullet\bullet}^2}{N}$$

$$SS_{Treatments} = \sum_{i=1}^{a} \frac{y_{i\bullet}^2}{n_i} - \frac{y_{\bullet\bullet}^2}{N}$$