פרק ד-1

# אפקטים אקראיים בניתוח שונות חד-כווני.

# השוואה בין $a=\infty$ טיפולים.

- ניתוח שונות חד-כווני אפשר לנו להשוות בין מספר סופי של טיפולים (a).
- לפעמים נרצה להשוות בין מספר גדול מאוד (נמדל כאין סופי) של טיפולים.
- במקרה זה לא נדגום מכל קבוצת טיפול (יש יותר מדי וזה יהיה יקר מדי) אלה רק ממספר סופי של קבוצות טיפול (c).
- דוגמא:
  - לרשת חנויות "מקדונלדס" יש בסביבות 10,000 סניפים בארה"ב.
  - התזונאי הראשי של הרשת מעוניין לבדוק האם זמן הטיגון של ציפס הוא זהה בכל הסניפים.
  - לצורך כך נלקח מדגם **אקראי** של (c=4) סניפים ובכל סניף נלקח מדגם **אקראי** של 4 משמרות במהלך החודש ובכל משמרת נבדק זמן הטיגון הממוצע של ציפס במהלך משמרת (בשניות).
  - H0: אין שוני בין זמן הטיגון בין כל הסניפים בארה"ב.
  - H1: קיים שוני בין הסניפים.

האם עבור נתונים אלו מתקיימות הנחות נורמאליות?

# מודל "קצת" אחר...

c זה לא מספר
הרמות של הגורם
שיש אלה מספר
הרמות אשר אנו
דוגמים.

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

האפקט הוא **אקראי**

$$i = 1, \ldots, c$$

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

$$\tilde{\tau}_i \sim NID(0, \sigma_\tau^{\ 2})$$

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

---

$$\boxed{\begin{aligned} H_0 &: \ \sigma_\tau^{\ 2} = 0 \\ H_1 &: \ \sigma_\tau^{\ 2} > 0 \end{aligned}}$$

מהו H0?

$$\boxed{Var(y_{ij}) = \sigma_\tau^{\ 2} + \sigma^2}$$

## סטטיסטי מבחן זהה לזה של לניתוח שונות חד-כווני עם אפקטים קבועים...

השונות של כל תצפית:

$$Var(y_{ij}) = \sigma_\tau^{\ 2} + \sigma^2$$

הקשר האלגברי של פרוק
סכום הריבועים עדיין מתקיים:

$$\underbrace{SS_{Total}}_{N-1} = \underbrace{SS_{Treatments}}_{c-1} + \underbrace{SS_E}_{N-c}$$

תמיד מתקיים (גם לא תחת H0):

$$\frac{SS_E}{\sigma^2} \sim \chi^2_{N-c}$$

תחת H0:

$$\frac{SS_{Treatments}}{\sigma^2} \sim \chi^2_{c-1}$$

ולכן תחת H0:

$$F_0 = \frac{SS_{Treatments}/(c-1)}{SS_E/(N-c)} = \frac{MS_{Treatments}}{MS_E} \sim F_{c-1,N-c}$$

# ביצוע ניתוח שונות

- למרות שהמודל **שונה** קבלנו סטטיסטי מבחן זהה לסטטיסטי המבחן של ניתוח שונות חד-כווני אם אפקטים קבועים (לא אקראיים).
- לכן ניתוח השונות הוא זהה.

# דוגמא (מקדונלד):

| | branch | fryTime |
|---|---|---|
| 1 | new york | 98 |
| 2 | new york | 97 |
| 3 | new york | 99 |
| 4 | new york | 96 |
| 5 | urbana | 91 |
| 6 | urbana | 90 |
| 7 | urbana | 93 |
| 8 | urbana | 92 |
| 9 | chicago | 96 |
| 10 | chicago | 95 |
| 11 | chicago | 97 |
| 12 | chicago | 95 |
| 13 | denver | 95 |
| 14 | denver | 96 |
| 15 | denver | 99 |
| 16 | denver | 98 |

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 3 | 89.1875000 | 29.7291667 | 15.68 | 0.0002 |
| Error | 12 | 22.7500000 | 1.8958333 | | |
| Corrected Total | 15 | 111.9375000 | | | |

| R-Square | Coeff Var | Root MSE | fryTime Mean |
|---|---|---|---|
| 0.796762 | 1.442717 | 1.376893 | 95.43750 |

| Source | DF | Anova SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| branch | 3 | 89.18750000 | 29.72916667 | 15.68 | 0.0002 |

# מה לגבי מבחני PostHoc?

- אין משמעות להשוואה בין טיפולים כי דגמנו מספר קטן של טיפולים מתוך כלל אוכלוסיית הטיפולים.
- לכן לא מבצעים מבחנים PostHoc כאשר הגורם הוא אקראי.
- מה שכן,... נתעניין באמידה של פרמטרי השונות של המודל...

# תוחלות של ממוצעי ריבועים...

$$E[MS_{Treatments}] = \frac{1}{c-1} E[SS_{Treatments}] = \ldots = \sigma^2 + n\sigma_\tau^{\ 2}$$

חסר הטייה
תחת H0

$$E[MS_E] = \frac{1}{N-c} E[SS_E] = \ldots = \sigma^2$$

תמיד חסר
הטייה

# אמידה של מרכיבי השונות "שיטת ניתוח שונות לאמידה"

$$E[MS_{Treatments}] = \sigma^2 + n\sigma^2_{\tau}$$
$$E[MS_E] = \sigma^2$$

ננסח משוואות עבור
אמדי פרמטרי השונות

$$MS_{Treatments} = \hat{\sigma}^2 + n\hat{\sigma}^2_{\tau}$$
$$MS_E = \hat{\sigma}^2$$

$$\hat{\sigma}^2 = MS_E$$
$$\hat{\sigma}^2_{\tau} = \frac{MS_{Treatments} - MS_E}{n}$$

יכול להיות שלילי!!!

עבור מדגמים לא מאוזנים
נשתמש בערך הבא עבור n:

$$n_0 = \frac{1}{c-1}\left[\sum_{i=1}^{n} n_i - \frac{\sum_{i=1}^{c} n_i^2}{\sum_{i=1}^{c} n_i}\right]$$

# דוגמא

- בחברה לייצור פנימיות של אופניים ישנם 130 מכונות מאותו סוג ליצור פנימיות. משקלה של פנימית אמור להיות 80 גרם. מכונה מייצרת כ – 500 פנימיות ליום.
- דגימה של 40 פנימיות הראתה שסטיית התקן של משקל פנימית הוא 15 גרם. המשמעות היא שישנן מספר לא מבוטל של פנימיות השוקלות יותר מדי, ומספר לא מבוטל של פנימיות חלשות מדי (אין בהם מספיק חומר) .
- מהנדס המפעל מעוניין להבין מה גורם לשוני כה גדול בין פנימית לפנימית. האם זה השוני בין פנימיות המיוצרות בכל מכונה או האם אוכלוסיית המכונות אינה הומוגנית (ז"א שונות משקל הפנימית המיוצר ע"י מכונות שונות היא גדולה).
- לצורך כך דגם המהנדס 6 מכונות, על כל מכונה 4 פנימיות.

# המשך דוגמא...

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 5 | 674.5000000 | 134.9000000 | 17.34 | <.0001 |
| Error | 18 | 140.0000000 | 7.7777778 | | |
| Corrected Total | 23 | 814.5000000 | | | |

| R-Square | Coeff Var | Root MSE | weight Mean |
|---|---|---|---|
| 0.828115 | 3.497012 | 2.788867 | 79.75000 |

| | machine | weight |
|---|---|---|
| 1 | a | 80 |
| 2 | a | 79 |
| 3 | a | 77 |
| 4 | a | 75 |
| 5 | b | 82 |
| 6 | b | 83 |
| 7 | b | 84 |
| 8 | b | 79 |
| 9 | c | 79 |
| 10 | c | 74 |
| 11 | c | 81 |
| 12 | c | 83 |
| 13 | d | 91 |
| 14 | d | 93 |
| 15 | d | 87 |
| 16 | d | 86 |
| 17 | e | 72 |
| 18 | e | 71 |
| 19 | e | 69 |
| 20 | e | 74 |
| 21 | f | 81 |
| 22 | f | 79 |
| 23 | f | 80 |
| 24 | f | 75 |

$$\hat{\sigma}^2 = MS_E = 7.77$$

$$\hat{\sigma}_\tau^2 = \frac{MS_{Treatments} - MS_E}{n} = 31.78$$

אם כך, הנתונים מעידים על כך שבכדי לייצב את משקל הפנימיות המיוצרות מומלץ ראשית לחפש דרכים להפוך את המכונות המייצרות ליותר הומוגניות

# רווח סמך עבור $\sigma^2$

תחת הנחת הנורמאליות

$$\frac{(N-c)MS_E}{\sigma^2} \sim \chi^2_{N-c}$$

$$P(\chi^2_{\alpha/2,N-c} \le \frac{(N-c)MS_E}{\sigma^2} \le \chi^2_{1-\alpha/2,N-c}) = 1-\alpha$$

$$P(\frac{(N-c)MS_E}{\chi^2_{1-\alpha/2,N-c}} \le \sigma^2 \le \frac{(N-c)MS_E}{\chi^2_{\alpha/2,N-c}}) = 1-\alpha$$

# רווח סמך עבור ${\sigma_\tau}^2$

אין ביטוי סגור לרווח סמך עבור ${\sigma_\tau}^2$

אבל ניתן למצוא רווח סמך עבור $\frac{{\sigma_\tau}^2}{{\sigma_\tau}^2+\sigma^2}$

# רווח סמך עבור $\frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2}$

$$L=\frac{1}{n}\left(\frac{MS_{Treatments}}{MS_E}\frac{1}{F_{1-\alpha/2,c-1,N-c}}-1\right)$$

$$U=\frac{1}{n}\left(\frac{MS_{Treatments}}{MS_E}\frac{1}{F_{\alpha/2,c-1,N-c}}-1\right)$$

$$P\left(\frac{L}{1+L}\le\frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2}\le\frac{U}{1+U}\right)=1-\alpha$$