פרק א-2

השוואת תוחלות של שתי אוכלוסיות בלתי תלויות.

מבוא

- דוגמא: בניסוי רפואי מדגם אקראי של נבדקים מקבל טיפול מסוים, ומדגם אקראי נוסף מקבל "Control" Placebo.
- נניח שמדד מסוים עבור קבוצת הנבדקים המקבלת את הטיפול מתפלג F_1. ועבור קבוצת ה Control הפילוג הוא F_2.
- מעוניינים בהשוואה של $E(F_1)$ ושל $E(F_2)$.

דוגמאות נוספות:

- השוואות מאפיינים של שתי אוכלוסיות חיות/בני אדם.
- השוואות טיפולים מתחומי ביולוגיה, כימיה, פסיכולויגיה וכו'.

ייתכנו מקרים רבים...

- השונויות ידועות/לא ידועות.
- השונויות שוות/לא שוות.
- המדגם גדול/קטן.
- האוכלוסיות נורמאלית/לא נורמאלית.
- גדלי המדגם שווים/שונים.

המקרה הפשוט: אוכלוסיות נורמאליות, שוות שוניות אשר ידועות.

- נניח:

$$F_1 \equiv N(\mu_1, \sigma^2)$$
$$F_2 \equiv N(\mu_2, \sigma^2)$$

- המדגם שלנו הוא:

$$X_1, \ldots, X_{n_1} \sim NID(\mu_1, \sigma^2)$$
$$Y_1, \ldots, Y_{n_2} \sim NID(\mu_2, \sigma^2)$$

- אמד הגיוני עבור $\mu_1 - \mu_2$ הוא: $\overline{X} - \overline{Y}$

- מקבלים:

$$\overline{X} - \overline{Y} \sim N(\mu_1 - \mu_2, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$$

סטטיסטי...

- אז יש לנו: $\overline{X}-\overline{Y}\sim N(\mu_1-\mu_2,\sigma^2(\frac{1}{n_1}+\frac{1}{n_2}))$
- ולכן: $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1)$
- אם כך רווח סמך להפרש התוחלות הוא:

$$P\left(\overline{X}-\overline{Y}+Z_{\alpha/2}\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\le \quad \mu_1-\mu_2 \quad \le \overline{X}-\overline{Y}+Z_{1-\alpha/2}\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right) = 1-\alpha$$

- מכאן ניתן מייד לנסח מבחני השערה...

ניסוח מבחני השערה...

- דו-צדדי (נובע ישירות מרווח הסמך)
- חד-צדדי לכוון אחד.
- חד-צדדי לכוון שני.

בחיים לעיתים נדירות השונות ידוע...

- אמידה של שונות של מדגם:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}{n-1}$$

- עבור שני מדגמים:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

כאשר מניחים שונויות שוות

הסטטיסטי t

- משפט:

$$t=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{n_1+n_2-2}$$

- ז"א אם התוחלות שוות:

$$t=\frac{(\overline{X}-\overline{Y})}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{n_1+n_2-2}$$

עדיין מניחים שונויות שוות

עדיין מניחים אוכלוסיות נורמאליות

כאשר המדגמים אינם קטנים אז משפט הגבול המרכזי פותר אותנו מהצורך בהנחת נורמאליות

מכאן רווח סמך ומבחני השערה:

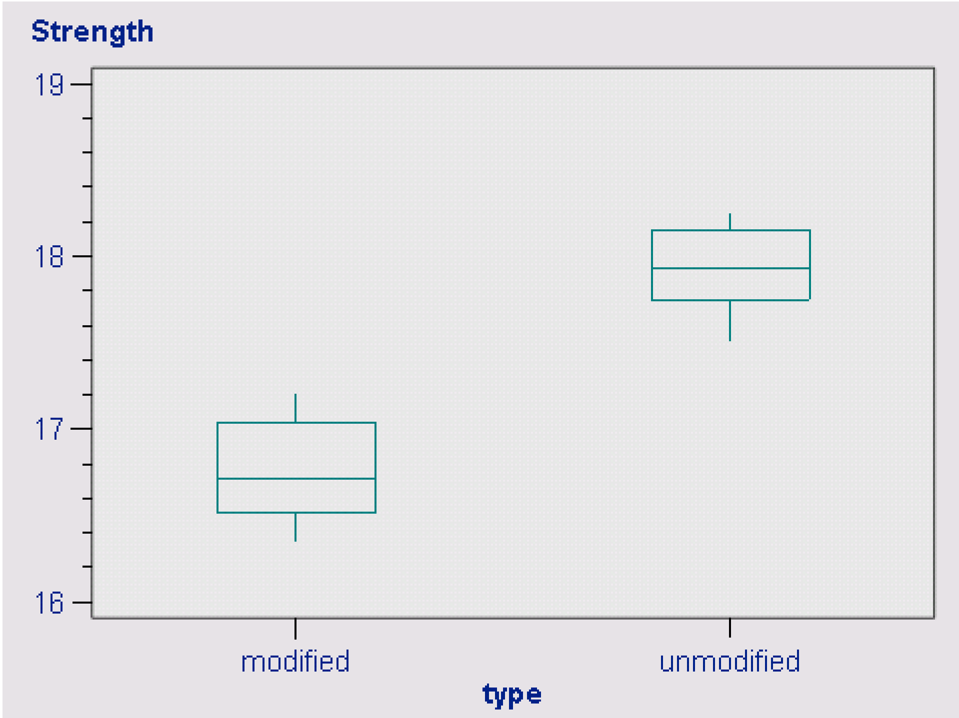
$$P\left(\overline{X}-\overline{Y}+t_{\alpha/2}^{n_1+n_2-2}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\le \quad \mu_1-\mu_2 \quad \le \overline{X}-\overline{Y}+t_{1-\alpha/2}^{n_1+n_2-2}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right) = 1-\alpha$$

- דו-צדדי.
- חד-צדדי לכוון אחד.
- חד-צדדי לכוון שני.

וואריאציות...

H0: שהפרש התוחלות הוא δ (לא בהכרח 0).

דוגמת חוזק הבטון (Montgomery).



ביצוע ב SAS

- PROC T-TEST.
- Analysis->ANOVA->T-test
  - בחירת "Two-Sample".
- אפשרויות:
  - Null Hypothesis
  - Confidence Level
  - Standard Deviation Confidence Intervals
  - Box and Whisker Plots
  - Means Plot (עם מספר אופציות)

עיקרי הפלט בSAS

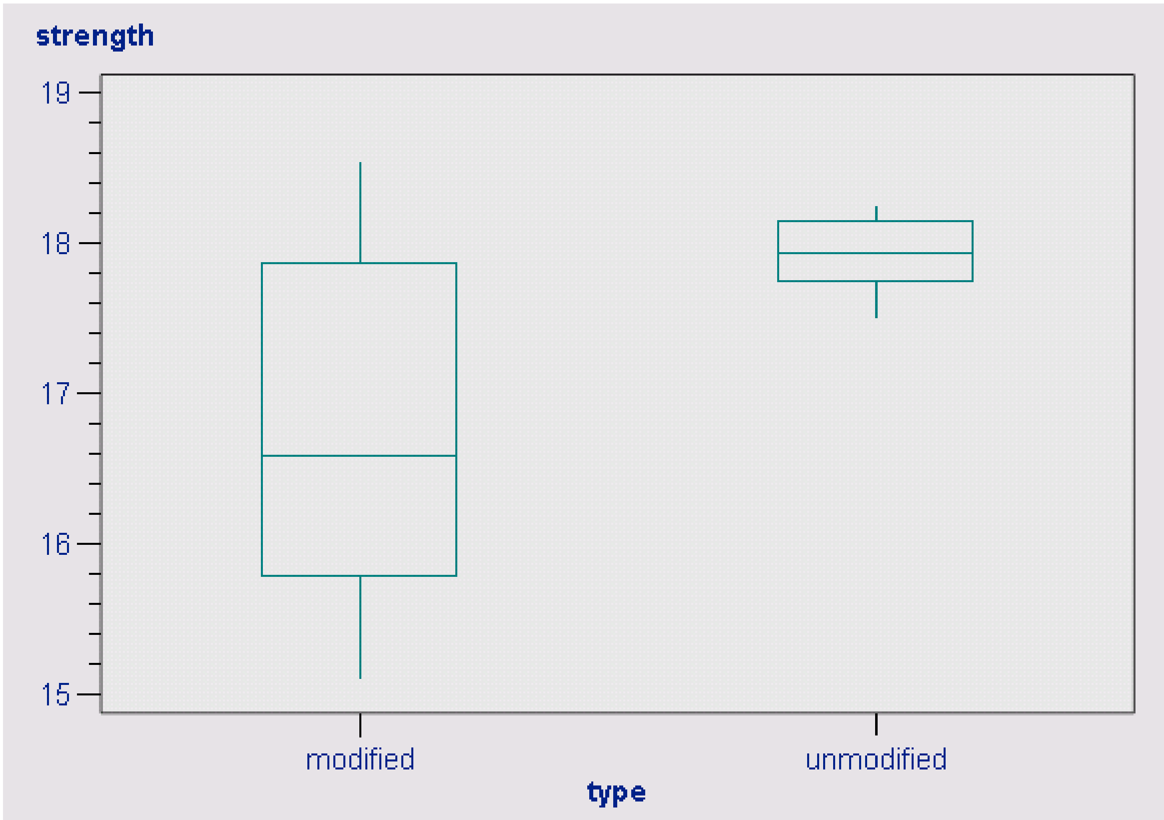
- משתנים:
  - טיפול 1,
  - טיפול 2,
  - ההפרש
- N
- (Lower CL mean, mean, Upper CL mean)
- (Lower CL std dev, std dev, Upper CL std dev)
- Std Err
- (Minimum, Maximum)

עיקרי הפלט בSAS (המשך...)

- מבחני t:
  - DF
  - T Value
  - Pr > |t|
  - שיטות (methods):
    - Pooled (שונויות שוות).
    - Satterthwaite (שונויות שונות)
- מבחני שוויון שונויות, Folded F.

טיפול מקורב בשונויות שונות...

- "נמתח" את דוגמת הבטון....



Satterthwaite תיקון ...שונות שונויות

- נתקן כאן את סטטיסטי המבחן:

$$\cancel{t=\frac{(\overline{X}-\overline{Y})}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}} \longrightarrow t'=\frac{(\overline{X}-\overline{Y})}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

- כעת הסטטיסטי 't כבר אינו מפולג בדיוק t, אבל בקירוב הוא מפולג t עם v דרגות חופש:

$$v=\left[\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1}+\frac{(S_2^2/n_2)^2}{n_2-1}}\right]$$

ניסוח כמודל ליניארי...

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

$$i = 1,2$$

$$j = 1,\dots,n_i$$

$$\sum_{i=1}^{2} \tau_i = 0$$

$$\tilde{\varepsilon}_{1j} \sim NID(0,\sigma_1^2)$$

$$\tilde{\varepsilon}_{2j} \sim NID(0,\sigma_2^2)$$

$$Assume\ \ \sigma_1^2 = \sigma_2^2$$