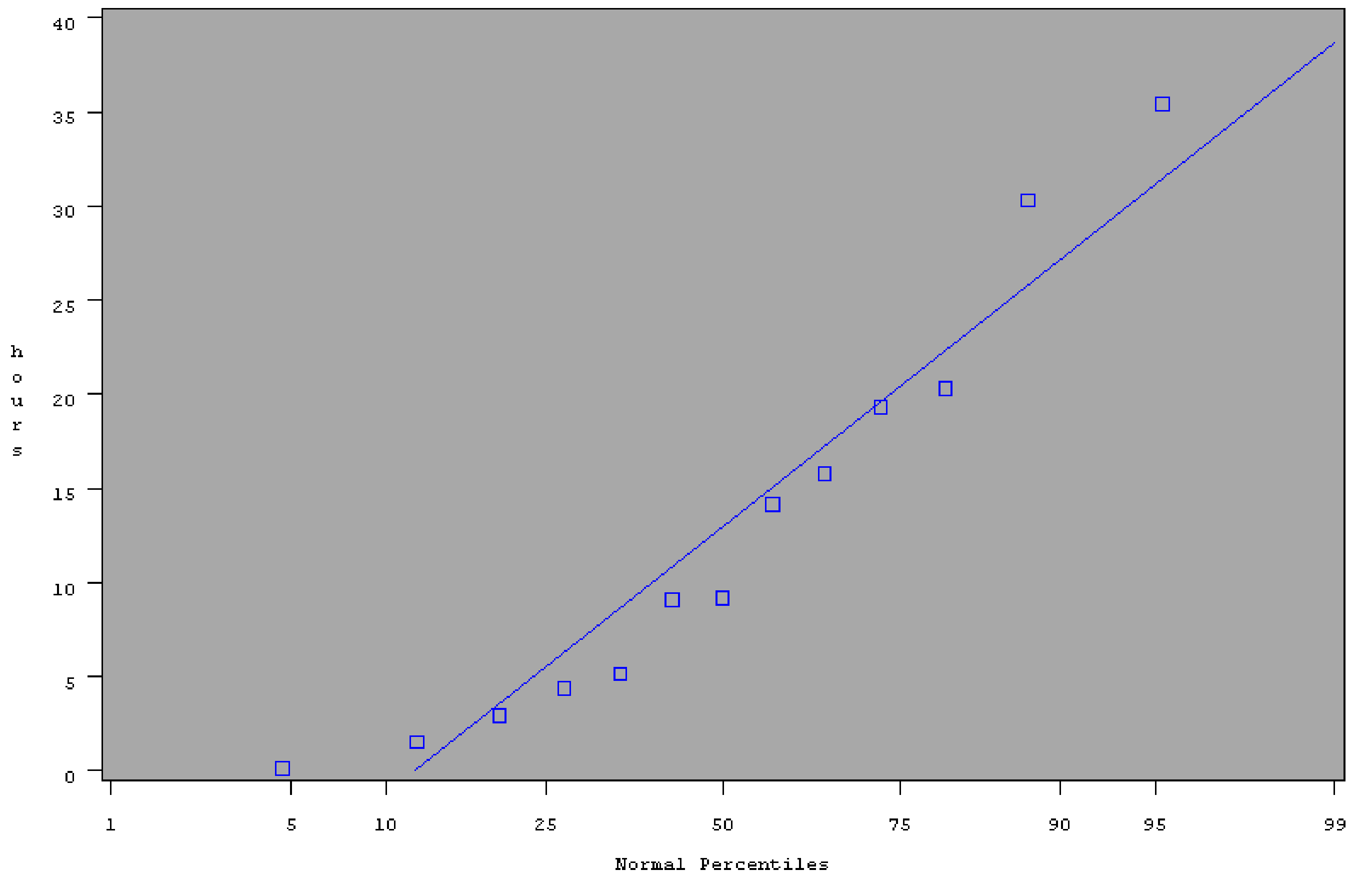
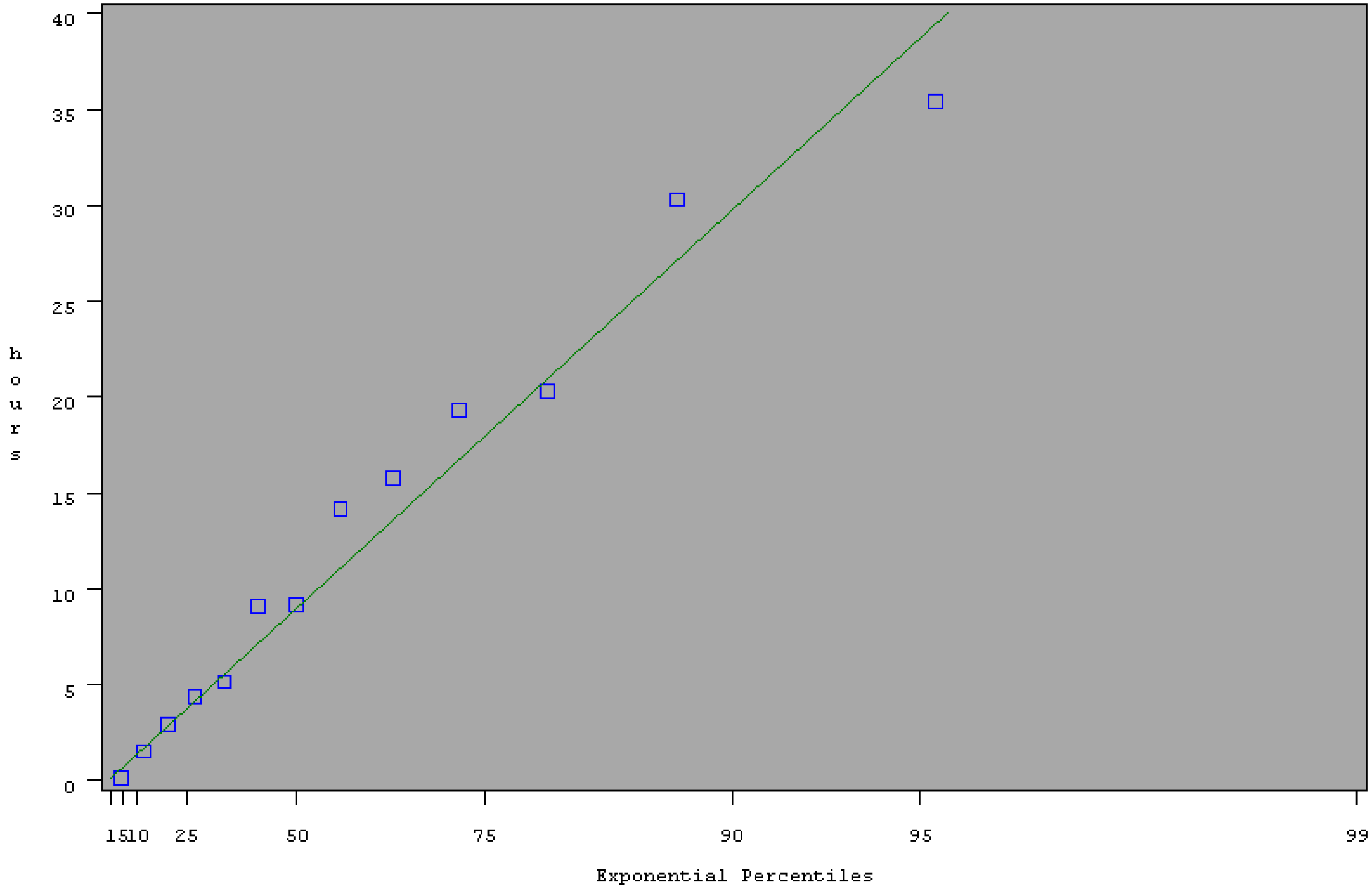
פרק א-3

שימוש בשיטה א-פרמטרית להשוואת תוחלות של שתי אוכלוסיות.

לפעמים הנתונים אינם מהתפלגות נורמלית



בדוגמא הזו, הנתונים מהתפלגות exp.



מה ניתן לעשות?

- להתעלם.
- לבצע טרנספורמציה.
- להשתמש בשיטה ספציפית לפילוג הנתונים.
- להשתמש בשיטה א-פרמטרית.

בשביל זה צריך לדעת כיצד מפולגים הנתונים

מומלץ כאשר לא רוצים או לא ניתן להניח הנחות לגבי פילוג הנתונים

מהן שיטות "א-פרמטריות"...

- מגוון רחב של שיטות סטטיסטיות אשר אינן עושות שימוש בהנחות התפלגותיות לגבי הנתונים וכאן חוזקן.
- לפעמים אי-השימוש בהנחות התפלגותיות תגרור עוצמת מבחן חלשה יותר (ולכן במידה וניתן להשתמש בשיטה פרמטרית, אז לרוב עדיף).
- בקורס זה נפגוש 2 דוגמאות לצורך המחשה:
  - היום: מבחן Mann-Whitney (Wilcoxon).
  - בהמשך: מבחן Kruskal –Wallis.
- בSAS: PROC NPAR1WAY

החלפת הנתונים בדרגות (ranks)

- מהן דרגות?

12, 2.5, 4.7, 22 $\longrightarrow$ 3,1,2, 4

- דרגות הנתונים אינם משתנים כאשר מפעילים טרנספורמציות מונוטוניות על הנתונים.
- השפעת תצפיות חריגות (outliers) מופחתת.
- כאשר מנתחים את דרגות הנתונים (ולא את הנתונים המקוריים) אז לא ניתן לנתח גדלים כגון תוחלת או שונות הנתונים על סמך הדרגות.

השוואת שתי אוכלוסיות ב"ת...

- נניח שיש ברשותנו $n1+n2$ יחידות ניסוי.
- $n1$ יחידות מוקצות לקבוצת טיפול.
- $n2$ יחידות מוקצות לקבוצת הביקורת.
- ההקצאה של יחידות לטיפול וביקורת היא אקראית.
- השארת האפס: אין לטיפול השפעה.
- תחת השארת האפס, כל חריגה בין תוצאות קבוצת הטיפול לתוצאות קבוצת הביקורת היא בגלל רנדומיזציה.

חישוב סטטיסטי המבחן...

- נקבץ את כל $n1+n2$ התצפיות לקבוצה אחת.
- נמיין את התצפיות. ועבור כל תצפית נצמיד את הדרגה שלה (הסדר שלה במיון).
- נחשב את סכום הדרגות של התצפיות מקבוצת הבקרה: S.
- בינתיים נניח שאין שתי תצפיות זהות.

תחת H0 לא סביר שסכום הדרגות יהיה "מאוד" קטן או "מאוד" גדול

דוגמונת:

Sum=3 S=Sum=7

קבוצת ביקורת	קבוצת טיפול
6.77	1.34
4.23	3.45

(1)

(2)

(3)

(4)

האם זה סביר תחת H0?

קומבינטוריקה של הדרגות...

- תחת H0 כל הקומבינציות של דרגות עבור תצפיות זהות. לכן כל בחירה אפשרית של 2 דרגות עבור קבוצת הביקורת היא שווה הסתברות.
- הבחירות האפשריות:

דרגות	S
{1,2}	3
{1,3}	4
{1,4}	5
{2,3}	5
{2,4}	6
{3,4}	7

- ולכן תחת H0 פילוג S הוא:

$$P(S=s)=\begin{cases} 1/6 & s=3 \\ 1/6 & s=4 \\ 1/3 & s=5 \\ 1/6 & s=6 \\ 1/6 & s=7 \end{cases}$$

P-value

- קיבלנו S=7. תחת H0, הסיכוי ש S יקבל ערך קיצוני זה הוא 1/6=0.166.
- ז"א במידה וההשארות שלנו היו:
  כאשר קבוצת הטיפול היא מס' 1,

  $$H_0 : \mu_1 < \mu_2$$
  $$H_1 : \mu_1 \geq \mu_2$$

  אז נדחה את H0 עבור כל רמת מובהקות גדולה מ 0.166
- הערה: ב"דוגמונת" זו, גודל המדגם כ"כ קטן כך שלא ניתן לבצע את המבחן ברמת מובהקות יותר קטנה מ 0.166.

טיפול בתצפיות בעלות ערכים זהים:

- נניח שיש 4 תצפיות בעלות ערך זהה. נניח שדרגות התצפיות הנ"ל הינם 3,4,5,6.
- אז לצורך חישוב סטטיסטי המבחן יש לתת לכל אחת מהתצפיות את הערך:

$$\frac{3+4+5+6}{4}=4.5$$

קרוב למבחן המדויק...

- לפעמים חישוב פילוג סטטיסטי המבחן תחת H0 לוקח זמן חישוב רב...
- במקרים כאלו ניתן לקרב את פילוג הסטטיסטי ע"י התפלגות רציפה.
- במקרה שלנו (Mann-Whitney), נשתמש בקרוב ע"י התפלגות נורמלית שצידוקו נובע ממשפט הגבול המרכזי.
- כמובן שזהו רק קרוב, דיוק הקרוב תלוי בגודל המדגם.

דוגמת פלט ב – SAS:

- נסתכל על דוגמה המשווה את אורך החיים של שתי סוגי נורות.
- Wilcoxon Scores:
  - N
  - Sum of Scores
  - Expected Under H0
  - StdDev Under H0
  - MeanScore
- Wilcoxon Two Sample Test
  - Statistic
  - Normal Approximation
  - T Approximation
  - Exact Test