



סטטיסטיקה אי פרמטרית וניתוח משתנים בסולם סדר

1

סיכום קצר עד כה



- בדיקת השערות ורווחי סמך
 - התפלגות נורמלית עם שונות ידועה
 - התפלגות בינומית (פרופורציה) עם מדגם מספיק גדול
 - התפלגות כלשהי עם מדגם מספיק גדול
 - התפלגות נורמלית, שונות לא ידועה, מדגם קטן
- מתאם ורגרסיה – קשרים בין משתנים כמותיים
- ניתוח שונות – קשר בין משתנה כמותי ובין משתנים מסבירים איכותיים

2



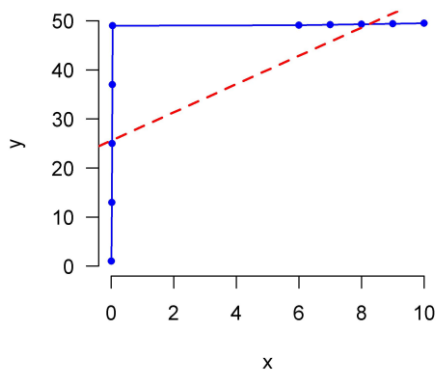
- בדיקת השערות כאשר ההתפלגות אינה נורמלית והמדגם לא מספיק גדול
- ניתוח משתנים בסולם סדר

מה הקשר?

גישה אפשרית כאשר נתון משתנה כמותי שאינו מהתפלגות נורמלית:
"לשנמך" אותו למשתנה בסולם סדר



A	B
x	y
0.01	1
0.02	13
0.03	25
0.04	37
0.05	49
6	49.1
7	49.2
8	49.3
9	49.4
10	49.5



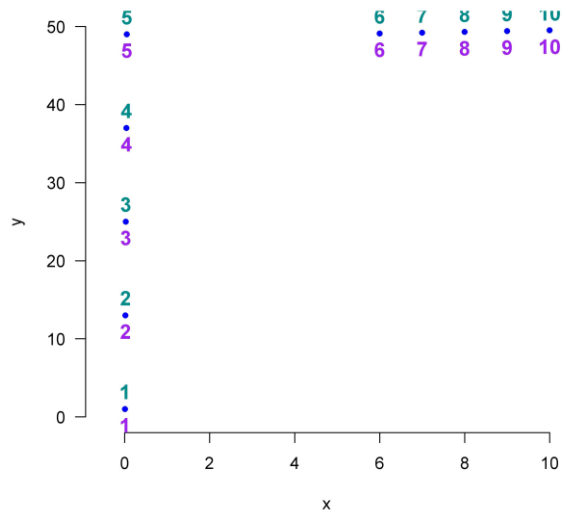
פתרון אפשרי

- הצעה: לשנמך את סולם המדידה של x ו- y מסולם מנה לסולם סדר
- המשמעות היא שלא נתייחס לערכים של המספרים אלא רק לסדר ביניהם
- אנחנו יכולים לדרג את הערכים מהנמוך לגבוה
- לבסוף: נתייחס לדרגות כאל משתנה כמותי!

5

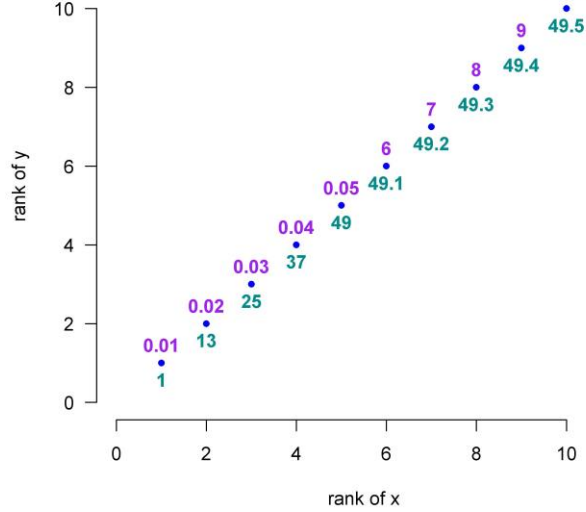
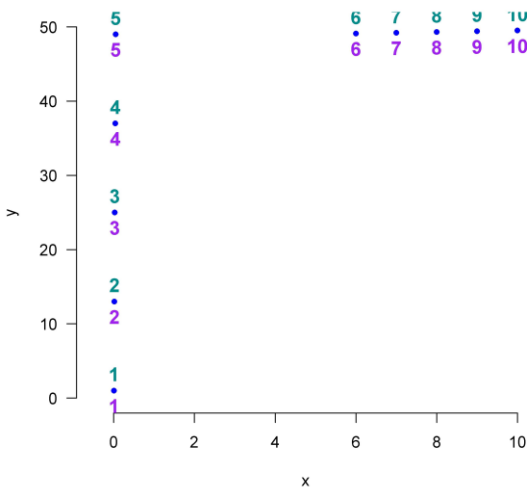
המשך הדוגמה

A	B	C	D	E
x	y		rank_x	rank_y
0.01	1		1	1
0.02	13		2	2
0.03	25		3	3
0.04	37		4	4
0.05	49		5	5
6	49.1		6	6
7	49.2		7	7
8	49.3		8	8
9	49.4		9	9
10	49.5		10	10



6

המשך הדוגמה



7

השוואת דיאטות

	weight.loss	diet.type	initial.weight	final.weight	weight.loss
Female	3.8	A	58	54.2	3.8
Female	6	A	60	54	6
Female	0.7	A	64	63.3	0.7
Female	2.9	A	64	61.1	2.9
Female	2.8	A	65	62.2	2.8
Female	2	A	66	64	2
Female	2	A	67	65	2
Female	8.5	A	69	60.5	8.5
Female	1.9	A	70	68.1	1.9
Female	3.1	A	70	66.9	3.1
Female	1.5	A	72	70.5	1.5
Female	3	A	72	69	3
Female	3.6	A	72	68.4	3.6
Female	0.9	A	82	81.1	0.9
Female	-2.1	B	58	60.1	-2.1
Female	2	B	58	56	2
Female	1.7	B	59	57.3	1.7
Female	4.3	B	61	56.7	4.3
Female	7	B	62	55	7
Female	0.6	B	63	62.4	0.6
Female	2.7	B	63	60.3	2.7
Female	3.6	B	63	59.4	3.6
Female	3	B	65	62	3

8

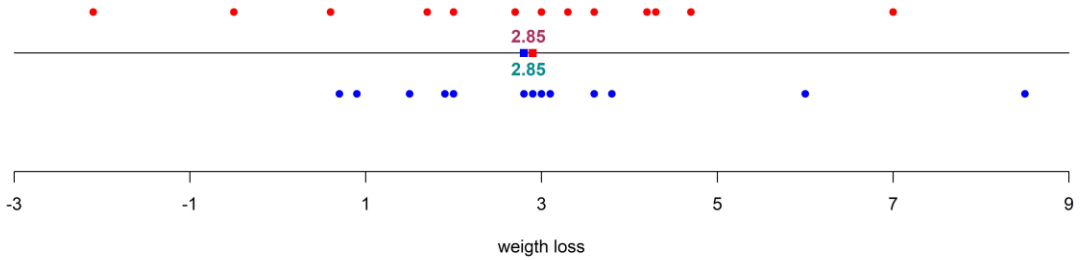
דיאטה A מול דיאטה B

2 קבוצות בלתי תלויות:

דיאטה A

דיאטה B

diet.type	weight.loss	rank
B	-2.1	1
B	-0.5	2
B	0.6	3
A	0.7	4
A	0.9	5
A	1.5	6



9

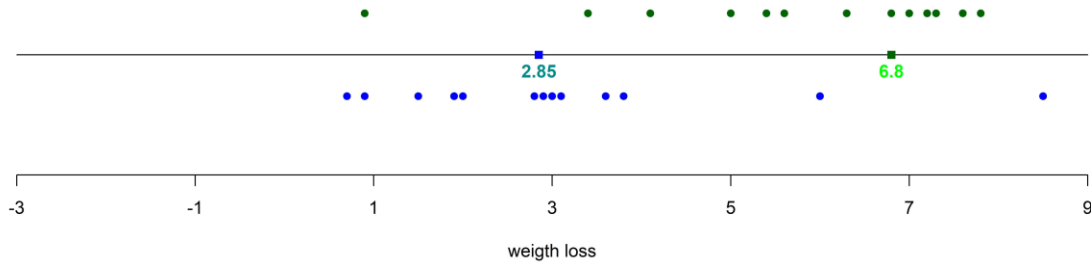
דיאטה A מול דיאטה C

2 קבוצות בלתי תלויות:

דיאטה A

דיאטה C

diet.type	weight.loss	rank
A	0.7	1
A	0.9	2
C	0.9	3
A	1.5	4
A	1.9	5
A	2	6



10



פרנק וילקוקסון
1892-1965

- השערת האפס: לשתי האוכלוסיות יש אותה התפלגות
- ההשערה האלטרנטיבית: לאוכלוסיות יש התפלגויות שונות, אך אינה נוקבת באיזה התפלגויות מדובר
- זאת בניגוד להשערות קודמות שדנו בהן, בהן לשתי האוכלוסיות הייתה אותה ההתפלגות והן נבדלו זו מזו רק בערך של פרמטר, למשל התוחלת. מסיבה זו מבחנים סטטיסטיים כגון מבחן וילקוקסון נקראים מבחנים אפרמטריים.
- מבחן וילקוקסון מיועד להשוואת ההתפלגויות של שתי קבוצות/אוכלוסיות בלתי תלויות
- המבחן מבוסס על דירוג הערכים של שתי הקבוצות יחד, ולאחר מכן על השוואת סך הדרגות של כל קבוצה
- אם יש שתי תצפיות זהות בארכן, משתמשים בדרגה הממוצעת
- יש הרחבה להשוואת יותר משתי קבוצות הנקראת מבחן קראסקל-ואליס

diet.type	weight.loss	rank
A	0.7	1
A	0.9	2
C	0.9	3
A	1.5	4
A	1.9	5
A	2	6
A	2	7
A	2.8	8
A	2.9	9
A	3	10
A	3.1	11
C	3.4	12
A	3.6	13
A	3.8	14
C	4.1	15
C	5	16
C	5.4	17
C	5.6	18
A	6	19
C	6.3	20
C	6.8	21

$S = 1 + 2 + 3 + \dots + 29 = \frac{29 \cdot 30}{2} = 435$ סכום כל הדרגות:

$T_A = 1 + 2 + 4 + 5 + \dots + 14 + 19 + 29 = 138$ סכום דרגות דיאטה A:

$T_C = 3 + 12 + 15 + \dots + 26 + 27 + 28 = 297$ סכום דרגות דיאטה C:

$T_C = S - T_A = 435 - 138 = 297$

```
wilcoxon rank sum test with continuity correction
data: ac$weight.loss[w1] and ac$weight.loss[w2]
W = 33.5, p-value = 0.001934
alternative hypothesis: true location shift is not equal to 0
```

בדיקת השערה על החציון – מבחן הסימן



- בדרך כלל החציון אינו פרמטר של ההתפלגות ולא תמיד ברור מהי התפלגות חציון המדגם
- בכל זאת, נרצה לבדוק השערה על חציון ההתפלגות, כגון: $H_0: \text{Median} = m$ $H_1: \text{Median} \neq m$
- רציונל: אם החציון שווה ל- m , ההסתברות כי תצפית המדגם גדולה מ- m שווה להסתברות כי היא רטנה מ- m .
- לכן השערות שקולות הן: $H_0: p = 0.5$ $H_1: p \neq 0.5$ כאשר p היא ההסתברות כי תצפית במדגם גדולה מ- m .
- באופן שקול, p היא ההסתברות כי הפרש בין תצפית במדגם ובין m היא חיובית $x_1 - m > 0$
- נדחה את השערת אם הפרופורציה של הפרשי התצפיות ו- m שונה מ-0.5 באופן מובהק

13

ציוני SAT



- מבחן SAT הנערך בארה"ב מקביל למבחן הפסיכומטרי הנערך בישראל.
- למבחן יש חלק מילולי וחלק מתמטי
- נתון קובץ נתונים ובו (בין היתר) ממוצעי הציונים בחלק המילולי עבור 50 מדינות ארה"ב וכן עבור וושינגטון די. סי.
- לרוע המזל הקובץ נפגם והצלחנו להציל נתונים רק עבור 15 מדינות
- אנו רוצים לבדוק השערה על החציון של התפלגות ציוני החלק המילוי

14

SAT ציוני

A	B	C
score	X-550	sign
549	-1	+
523	-27	+
507	-43	+
577	27	-
543	-7	+
536	-14	+
508	-42	+
489	-61	+
554	4	-
570	20	-
596	46	-
536	-14	+
564	14	-
574	24	-
497	-53	+

השערות $H_0: \text{median} = 550$ $H_1: \text{median} \neq 550$

יש 9 תצפיות עם סימן שלילי ו-6 עם סימן חיובי

Exact binomial test

data: 6 and 15
 number of successes = 6, number of trials = 15, p-value = 0.6072
 alternative hypothesis: true probability of success is not equal to 0.5
 95 percent confidence interval:
 0.1633643 0.6771302
 sample estimates:
 probability of success
 0.4

15

מבחן הסימן – נתונים מזווגים

A	B	C
aster	choise	Sign
1	Tropical Orange	-
2	Citrus Valley	+
3	Citrus Valley	+
4	Tropical Orange	-
5	Tropical Orange	-
6	Citrus Valley	+
7	Citrus Valley	+
8	Citrus Valley	+
9	Tropical Orange	-
10	Citrus Valley	+
11	No Preference	
12	Citrus Valley	+
13	Citrus Valley	+
14	Citrus Valley	+

השוואת מיצים:

14 צרכנים טעמו שני סוגים של מיץ תפוזים

כל צרכן ציין איזה מיץ הוא מעדיף

השערה: הצרכנים מעדיפים את המותג Citrus Valley

השערות פורמליות: $H_0: p = 0.5$ $H_1: p > 0.5$

9 מתוך 13 העדיפו את Citrus Valley

Exact binomial test

data: 9 and 13
 number of successes = 9, number of trials = 13, p-value = 0.1334
 alternative hypothesis: true probability of success is greater than 0.5
 95 percent confidence interval:
 0.4273807 1.0000000
 sample estimates:
 probability of success
 0.6923077

16