

סטטיסטיקה רב משתנית ב

פרק 2 - מבחנים אפרמטריים למדגמים בלתי תלויים

תוכן העניינים

1. מבחן מאן וויטני - שימוש בפלטים.....1
2. מבחן וילקוקסון למדגמים בלתי תלויים.....5

מבחנים אפרמטריים למדגמים בלתי תלויים

ניתוח פלטים במבחן מן-וויטני – רקע

מבחן מן-וויטני מיועד לבדוק האם לשתי אוכלוסיות התפלגות שווה. המבחן בוחן באופן רוחבי את כל תחום הערכים ולא מתמקד בערך מרכזי אחד. נשתמש במבחן זה כאשר יש שני מדגמים בלתי תלויים והמשתנה הכמותי הנחקר אינו מתפלג נורמלית או שמדובר במשתנה מסולם סדר. המבחן מתבסס על דירוג כל התצפיות. בעצם, מבחן זה הוא המענה האפרמטרי למבחן הפרמטרי להפרש תוחלות במדגמים בלתי תלויים.

דוגמה (פתרון בהקלטה) :

להלן תוצאות הערכות שקיבלו שני מורים : ד"ר A ופרופסור B. סטודנטים נתנו משווא כללי על המורים בסקלה של 1 (גרוע) עד 5 (מצוין). הטענה היא שד"ר A הוא מרצה טוב יותר מאשר פרופסור B.

Mann-Whitney Test

Ranks				
	teacher	N	Mean Rank	Sum of Ranks
grade	dr A	17	25.00	425.00
	prof B	20	13.90	278.00
	Total	37		

Test Statistics ^a	
	grade
Mann-Whitney U	68.000
Wilcoxon W	278.000
Z	-3.249
Asymp. Sig. (2-tailed)	.001
Exact Sig. [2*(1-tailed Sig.)]	.001 ^b

a. Grouping Variable: teacher

b. Not corrected for ties.

- הסבירו מדוע נעשה כאן מבחן מן-וויטני.
- מה המסקנה ברמת מובהקות של 1%?

שאלות

1) מחקרים טוענים שקיים הבדל בין שעות השינה של גברים לשעות השינה של נשים. כיוון שלא ניתן להוכיח ששעות שינה הינו משתנה המתפלג נורמלית ביצעו מבחן מן-וויטני בו לקחו נשים וגברים אקראיים ובדקו את שעות השינה שלהם.

Ranks				
	gender	N	Mean Rank	Sum of Ranks
sleeptime	male	10	???	135.00
	female	???	7.50	75.00
	Total	20		

Mann-Whitney Test

Test Statistics ^a	
	sleeptime
Mann-Whitney U	20.000
Wilcoxon W	75.000
Z	-2.319
Asymp. Sig. (2-tailed)	.020

- א. השלימו את סימני השאלה החסרים בפלט.
- ב. מהי המסקנה ברמת מובהקות של 5%?
- ג. מה הייתה מובהקות התוצאה אם טענת המחקר הייתה שגברים ישנים יותר מנשים?

2) שני אנשים נתבקשו לבדוק את מספר תאונות הדרכים בשבוע בשני קטעי כביש שונים. כל אחד בחר את השבועות באופן אקראי ובלתי תלוי באחר וספר כמה תאונות היו בכל כביש בשבוע. הפלטים שהתקבלו:

מספר תאונות

	Frequency
.00	2
1.00	1
2.00	2
3.00	1
Total	6

a. road = 1.00

מספר תאונות

	Frequency
.00	2
1.00	1
2.00	1
3.00	1
4.00	1
5.00	1
Total	7

a. road = 2.00

Ranks

	road	N	Mean Rank	Sum of Ranks
מספר תאונות	1.00	A	B	C
	2.00	D	E	F
	Total	G		

Test Statistics^a

	VAR00002
Asymp. Sig. (2-tailed)	.465

- א. השלימו בטבלה השלישית את המספרים החסרים במקום האותיות.
- ב. מהי מובהקות התוצאה לבדיקת ההשערה שאין הבדל בין הכבישים מבחינת התפלגות תאונות הדרכים?
- ג. כיצד הייתה משתנה התשובה של הסעיף הקודם אם כל חוקר היה מוסיף נתונים על שבוע נוסף לכל כביש?
- ד. מה המסקנה ברמת מובהקות של 5%?
- ה. מהי מובהקות התוצאה לבדיקת ההשערה שכביש מספר 1 עם התפלגות תאונות גבוהה יותר מאשר כביש מספר 2?

תשובות סופיות

- (1) א. 10, 13.5 ב. נדחה H_0 ג. 0.01
- (2) א. 6, 6.17, 7, 7.71, 13, 54 ב. 0.465
- ג. לא ניתן לדעת. ד. לא נדחה H_0 ה. 0.7675

מבחן ווילקוקסון למדגמים בלתי תלויים – רקע

מבחן ווילקוקסון למדגמים בלתי תלויים נכנס לקטגוריות המבחנים האפרמטריים. מבחן זה רלבנטי כאשר רוצים להשוות בין שתי אוכלוסיות על סמך שני מדגמים בלתי תלויים. המשתנה התלוי הוא משתנה כמותי שאינו מתפלג נורמלית או משתנה מסולם סדר. מבחן זה הוא החלופה האפרמטרית למבחן הפרמטרי להשוואת תוחלות על סמך שני מדגמים בלתי תלויים.

דוגמה:

מחקר חינוכי מעוניין להשוות בין 2 שיטות חינוך. המחקר רוצה לבדוק האם קיים הבדל ברמת ביטחון העצמי של הילדים בשיטות החינוך השונות. נבחרו באקראי 5 ילדים שחונכו בשיטת A. כמו כן נדגמו באקראי 5 ילדים שחונכו בשיטת B. פסיכולוגים בחנו את 10 הילדים ונתנו ציון לביטחון העצמי בסקאלה של 1-20. מהן ההשערות ומהו המבחן הסטטיסטי המתאים?

שיטה A	שיטה B
16	14
17	14
20	19
10	9
18	8

כדי לבצע את המבחן יש לחשב על סמך תוצאות המדגם את סטטיסטי המבחן שנסמן באות U.

השלבים לחישוב סטטיסטי המבחן:

- נסדר את כלל התצפיות של המחקר בסדר עולה מהנמוך ביותר לגבוה ביותר אך יש לדעת כל תצפית מאיזה מדגם היא באה.
- נדרג את כלל התצפיות של המחקר (אם יש תצפיות עם ערכים זהים הדירוג שלהן יהיה ממוצע המקומות שהם תופסים)
- נחשב את W_1 - סכום הדירוגים של התצפיות השייכות למדגם 1,
- W_2 - סכום הדירוגים של התצפיות השייכות למדגם 2.
- נחשב את הגדלים הבאים:

$$U_1 = W_1 - \frac{n_1(n_1+1)}{2} \quad U_2 = W_2 - \frac{n_2(n_2+1)}{2}$$

- הסטטיסטי U: במבחן דו צדדי $U = \min(U_1, U_2)$ במבחן חד צדדי U יהיה ה- U_i , שאמור להיות יותר קטן לפי השערת המחקר.

כדי להגיע למסקנה יש שני סוגים של טבלאות סטטיסטיות.
 טבלה מהסוג הראשון: עוזרת לנו לחשב את מובהקות התוצאה לאחר שחישבנו את
 ה- U הסטטיסטי.

טבלה מהסוג השני שקובעת מראש את הערך הקריטי של U שנסמן ב U_c .

טבלה מהסוג הראשון	טבלה מהסוג השני
$n_1 - n_2 \leq 8$	n_1 או $n_2 \geq 9$
עוזרת לחשב על סמך תוצאות המדגם את P_v אם $P_v \leq \alpha$ דוחים את H_0 .	נותנת את הערך הקריטי U_c כל הכרעה: נדחה את H_0 אם $U \leq U_c$

דוגמה:

מה המסקנה ברמת מובהקות של 5%?

טבלאות למציאת מובהקות התוצאה במבחן ווילקוקסון למדגמים בלתי תלויים

 $n_2 = 3$

u	n_1		
	1	2	3
0	0.250	0.100	0.050
12	0.500	0.200	0.100
3	0.750	0.400	0.200
4		0.600	0.350
5			0.500
			0.650

 $n_2 = 4$

u	n_1			
	1	2	3	4
0	0.200	0.067	0.028	0.014
1	0.400	0.133	0.057	0.029
2	0.600	0.267	0.114	0.057
3		0.400	0.200	0.100
4		0.600	0.314	0.171
5			0.429	0.243
6			0.571	0.343
7				0.443
8				0.557

 $n_2 = 5$

u	n_1				
	1	2	3	4	5
0	0.167	0.047	0.018	0.008	0.004
1	0.333	0.095	0.036	0.016	0.008
2	0.500	0.190	0.071	0.032	0.016
3	0.667	0.286	0.125	0.056	0.028
4		0.429	0.196	0.095	0.048
5		0.571	0.286	0.143	0.075
6			0.393	0.206	0.111
7			0.500	0.278	0.155
8			0.607	0.365	0.210
9				0.452	0.274
10				0.548	0.345
11					0.421
12					0.500
13					0.579

$$n_2 = 6$$

u	n_1					
	1	2	3	4	5	6
0	0.143	0.036	0.012	0.005	0.002	0.001
1	0.286	0.071	0.024	0.010	0.004	0.002
2	0.428	0.143	0.048	0.019	0.009	0.004
3	0.571	0.214	0.083	0.033	0.015	0.008
4		0.321	0.131	0.057	0.026	0.013
5		0.429	0.190	0.086	0.041	0.021
6		0.571	0.274	0.129	0.063	0.032
7			0.357	0.176	0.089	0.047
8			0.452	0.238	0.123	0.066
9			0.548	0.305	0.165	0.090
10				0.381	0.214	0.120
11				0.457	0.268	0.155
12				0.545	0.331	0.197
13					0.396	0.242
14					0.465	0.294
15					0.535	0.350
16						0.409
17						0.469
18						0.531

$$n_2 = 7$$

u	n_1						
	1	2	3	4	5	6	7
0	0.125	0.028	0.008	0.003	0.001	0.001	0.000
1	0.250	0.056	0.017	0.006	0.003	0.001	0.001
2	0.375	0.111	0.033	0.012	0.005	0.002	0.001
3	0.500	0.167	0.058	0.021	0.009	0.004	0.002
4	0.625	0.250	0.092	0.036	0.015	0.007	0.003
5		0.333	0.133	0.055	0.024	0.011	0.006
6		0.444	0.192	0.082	0.037	0.017	0.009
7		0.556	0.258	0.115	0.053	0.026	0.013
8			0.333	0.158	0.074	0.037	0.019
9			0.417	0.206	0.101	0.051	0.027
10			0.500	0.264	0.134	0.069	0.036
11			0.583	0.324	0.172	0.090	0.049
12				0.394	0.216	0.117	0.064
13				0.464	0.265	0.147	0.082
14				0.538	0.319	0.183	0.104
15					0.378	0.223	0.130
16					0.438	0.267	0.159
17					0.500	0.314	0.191
18					0.562	0.365	0.228
19						0.418	0.267
20						0.473	0.310
21						0.527	0.355
22							0.402
23							0.451
24							0.500
25							0.549

$$n_2 = 8$$

u	n_1							
	1	2	3	4	5	6	7	8
0	0.111	0.022	0.006	0.002	0.001	0.000	0.000	0.000
1	0.222	0.044	0.012	0.004	0.002	0.001	0.000	0.000
2	0.333	0.089	0.024	0.008	0.003	0.001	0.001	0.000
3	0.444	0.133	0.042	0.014	0.005	0.002	0.001	0.001
4	0.556	0.200	0.067	0.024	0.009	0.004	0.002	0.001
5		0.267	0.097	0.036	0.015	0.006	0.003	0.001
6		0.356	0.139	0.055	0.023	0.010	0.005	0.002
7		0.444	0.188	0.077	0.033	0.015	0.007	0.003
8		0.556	0.248	0.107	0.047	0.021	0.010	0.005
9			0.315	0.141	0.064	0.030	0.014	0.007
10			0.387	0.184	0.085	0.041	0.020	0.010
11			0.461	0.230	0.111	0.054	0.027	0.014
12			0.539	0.285	0.142	0.071	0.036	0.019
13				0.341	0.177	0.091	0.047	0.025
14				0.404	0.217	0.114	0.060	0.032
15				0.467	0.262	0.141	0.076	0.041
16				0.533	0.311	0.172	0.095	0.052
17					0.362	0.207	0.116	0.065
18					0.416	0.245	0.140	0.080
19					0.472	0.286	0.168	0.097
20					0.528	0.331	0.198	0.117
21						0.377	0.232	0.139
22						0.426	0.268	0.164
23						0.475	0.306	0.101
24						0.525	0.347	0.221
25							0.389	0.253
26							0.433	0.287
27							0.478	0.323
28							0.522	0.360
29								0.399
30								0.439
31								0.480
32								0.520

טבלה למציאת U_c

ברמת מובהקות של 5% למבחן חד צדדי או ברמת מובהקות של 10% למבחן דו צדדי במבחן ווילקוקסון למדגמים בלתי תלויים.

n_1	n_2											
	9	10	11	12	13	14	15	16	17	18	19	20
1											0	0
2	1	1	1	2	2	2	3	3	3	4	4	4
3	3	4	5	5	6	7	7	8	9	9	10	11
4	6	7	8	9	10	11	12	14	15	16	17	18
5	9	11	12	13	15	16	18	19	20	22	23	25
6	12	14	16	17	19	21	23	25	26	28	30	32
7	15	17	19	21	24	26	28	30	33	35	37	39
8	18	20	23	26	28	31	33	36	39	41	44	47
9	21	24	27	30	33	36	39	42	45	48	51	54
10	24	27	31	34	37	41	44	48	51	55	58	62
11	27	31	34	38	42	46	50	54	57	61	65	69
12	30	34	38	42	47	51	55	60	64	68	72	77
13	33	37	42	47	51	56	61	65	70	75	80	84
14	36	41	46	51	56	61	66	71	77	82	87	92
15	39	44	50	55	61	66	72	77	83	88	94	100
16	42	48	54	60	65	71	77	83	89	95	101	107
17	45	51	57	64	70	77	83	89	96	102	109	115
18	48	55	61	68	75	82	88	95	102	109	116	123
19	51	58	65	72	80	87	94	101	109	116	123	130
20	54	62	69	77	84	92	100	107	115	123	130	138

שאלות

- (1) מעוניינים להשוות בין שתי קבוצות כדורסל. נלקחו 5 משחקים מקבוצה א' ושישה משחקים מקבוצה ב'. נבדק בכל משחק ועבור כל קבוצה מספר הנקודות שצברה במשחק. בדקו ברמת מובהקות של 5% האם קיים הבדל בין הקבוצות מבחינת הניקוד שצברה במשחק.

קבוצה א	קבוצה ב
68	82
82	74
78	82
94	64
87	67
	65

- (2) מעוניינים לבדוק האם קורס קיץ באנגלית משפר את יכולות האנגלית לתלמידי חטיבת ביניים. נלקחו 20 ילדים בגיל חטיבת הביניים ברמת אנגלית דומה. 12 מהם נשלחו לקורס קיץ והיתר לא. בסוף הקיץ כולם נבחנו במבחן באנגלית הציון הגבוה ביותר התקבל בקרב אחד שלא עשה את הקורס ושבעת הציונים הנמוכים ביותר היו גם בקרב תלמידים שלא עשו את הקורס. מה המסקנה ברמת מובהקות של 5%?
- (3) במחקר לבדיקת יעילות ויטמין C נבחרו 15 מתנדבים מבין עובדי המפעל. תשעה מהם נבחרו מקרית וקיבלו טיפול שוטף בוויטמין C, ואילו שאר המתנדבים (קבוצת הביקורת) קבלו גלולת סוכר. במשך שלוש שנות המחקר היו מספר ימי ההיעדרות בגלל ההצטננות:
קבוצת הטיפול: 1, 3, 9, 3, 4, 0, 8, 12, 16.
קבוצת הביקורת: 19, 7, 28, 13, 23, 12.
 בדקו ברמת מובהקות של 5% שמספר ימי המחלה במשך שלוש שנים מצטמצם ביותר מ-4 ימים עם לקיחת ויטמין C.

תשובות סופיות

- (1) לא נדחה את H_0 .
- (2) נדחה את H_0 .
- (3) לא נדחה את H_0 .

\\