

סטטיסטיקה ב

פרק 21 - מבחני חי בריבוע

תוכן העניינים

1. מבחן טיב התאמה..... 1
2. מבחן לאי תלות..... 6

מבחני חי בריבוע

מבחן טיב התאמה – רקע

מבחן זה בא לבדוק האם אוכלוסייה מסוימת מתפלגת לפי התפלגות נתונה. המשתנה הנחקר מחולק למספר קטגוריות ויש לבדוק האם תוצאות המדגם תואמות להתפלגות הנתונה.

מבנה המבחן:

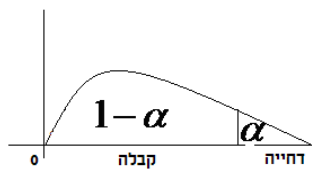
השערות:

המשתנה מתפלג לפי התפלגות מסוימת - H_0 .

אחרת - H_1 .

כלל הכרעה:

הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש. $d.f = K - 1$, כאשר K - מספר הקטגוריות.



הערך הקריטי הוא: $\chi^2_{1-\alpha, K-1}$, כלומר האחוזון ה- $1-\alpha$ בהתפלגות חי בריבוע שדרגות החופש הן $K-1$.

אם $\chi^2 > \chi^2_{1-\alpha, K-1}$, דוחים את השערת האפס.

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

סטטיסטי המבחן:

O_i - השכיחות שנצפתה במדגם בקטגוריה i .

p_i - הסתברות לקטגוריה i לפי השערת האפס.

$E_i = np_i$ - שכיחות צפויה במדגם לקטגוריה i בהנחת השערת האפס.

הערה:

תנאי כדי לבצע את המבחן הוא $E_i \geq 5$ לכל i . במידה ותנאי זה לא מתקיים יש אפשרות לאחד קטגוריות סמוכות עד שהתנאי יתקיים.

דוגמה (פתרון בהקלטה):

במדינה מסוימת שלוש מפלגות. בפרלמנט הנוכחי התפלגות מספר המושבים היא 30% למפלגה A, 60% למפלגה B ו-10% למפלגה C. לקראת הבחירות המתוכננות בשבוע הבא נעשה סקר שכלל 300 אזרחים. בסקר התקבל ש-40% יצביעו למפלגה A, 50% למפלגה B ו-10% למפלגה C. האם תוצאות הסקר תואמות להתפלגות המושבים בפרלמנט הנוכחי? בדקו ברמת מובהקות של 5%.

טבלת התפלגות חי-בריבוע – ערכי החלוקה



df	p												
	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.004393	0.004575	0.004982	0.005393	0.005988	0.006915	0.008332	0.01024	0.01279	0.01613	0.02036	0.02552	0.03179
2	0.01000	0.02001	0.05006	0.10303	0.21102	0.57513	1.39026	2.77055	4.61011	5.99148	7.38016	9.21074	10.59663
3	0.07171	0.11500	0.21600	0.35200	0.58400	1.21000	2.37000	4.11000	6.25000	7.81000	9.35000	11.30000	12.80000
4	0.20700	0.29700	0.48400	0.71100	1.06000	1.92000	3.36000	5.39000	7.78000	9.49000	11.10000	13.30000	14.90000
5	0.41200	0.55400	0.83100	1.15000	1.61000	2.67000	4.35000	6.63000	9.24000	11.10000	12.80000	15.10000	16.70000
6	0.67600	0.87200	1.24000	1.64000	2.20000	3.45000	5.35000	7.84000	10.60000	12.60000	14.40000	16.80000	18.50000
7	0.98900	1.24000	1.69000	2.17000	2.83000	4.25000	6.35000	9.04000	12.00000	14.10000	16.00000	18.50000	20.30000
8	1.34000	1.65000	2.18000	2.73000	3.49000	5.07000	7.34000	10.20000	13.40000	15.50000	17.50000	20.10000	22.00000
9	1.73000	2.09000	2.70000	3.33000	4.17000	5.90000	8.34000	11.40000	14.70000	16.90000	19.00000	21.70000	23.60000
10	2.16000	2.56000	3.25000	3.94000	4.87000	6.74000	9.34000	12.50000	16.00000	18.30000	20.50000	23.20000	25.20000
11	2.60000	3.05000	3.82000	4.57000	5.58000	7.58000	10.30000	13.70000	17.30000	19.70000	21.90000	24.70000	26.80000
12	3.07000	3.57000	4.40000	5.23000	6.30000	8.44000	11.30000	14.80000	18.50000	21.00000	23.30000	26.20000	28.30000
13	3.57000	4.11000	5.01000	5.89000	7.04000	9.30000	12.30000	16.00000	19.80000	22.40000	24.70000	27.70000	29.80000
14	4.07000	4.66000	5.63000	6.57000	7.79000	10.20000	13.30000	17.10000	21.10000	23.70000	26.10000	29.10000	31.30000
15	4.60000	5.23000	6.26000	7.26000	8.55000	11.00000	14.30000	18.20000	22.30000	25.00000	27.50000	30.60000	32.80000
16	5.14000	5.81000	6.91000	7.96000	9.31000	11.90000	15.30000	19.40000	23.50000	26.30000	28.80000	32.00000	34.30000
17	5.70000	6.41000	7.56000	8.67000	10.10000	12.80000	16.30000	20.50000	24.80000	27.60000	30.20000	33.40000	35.70000
18	6.26000	7.01000	8.23000	9.39000	10.90000	13.70000	17.30000	21.60000	26.00000	28.90000	31.50000	34.80000	37.20000
19	6.84000	7.63000	8.91000	10.10000	11.70000	14.60000	18.30000	22.70000	27.20000	30.10000	32.90000	36.20000	38.60000
20	7.43000	8.26000	9.59000	10.90000	12.40000	15.50000	19.30000	23.80000	28.40000	31.40000	34.20000	37.60000	40.00000
21	8.03000	8.90000	10.30000	11.60000	13.20000	16.30000	20.30000	24.90000	29.60000	32.70000	35.50000	38.90000	41.40000
22	8.64000	9.54000	11.00000	12.30000	14.00000	17.20000	21.30000	26.00000	30.80000	33.90000	36.80000	40.30000	42.80000
23	9.26000	10.20000	11.70000	13.10000	14.80000	18.10000	22.30000	27.10000	32.00000	35.20000	38.10000	41.60000	44.20000
24	9.89000	10.90000	12.40000	13.80000	15.70000	19.00000	23.30000	28.20000	33.20000	36.40000	39.40000	43.00000	45.60000
25	10.50000	11.50000	13.10000	14.60000	16.50000	19.90000	24.30000	29.30000	34.40000	37.70000	40.60000	44.30000	46.90000
26	11.20000	12.20000	13.80000	15.40000	17.30000	20.80000	25.30000	30.40000	35.60000	38.90000	41.90000	45.60000	48.30000
27	11.80000	12.90000	14.60000	16.20000	18.10000	21.70000	26.30000	31.50000	36.70000	40.10000	43.20000	47.00000	49.60000
28	12.50000	13.60000	15.30000	16.90000	18.90000	22.70000	27.30000	32.60000	37.90000	41.30000	44.50000	48.30000	51.00000
29	13.10000	14.30000	16.00000	17.70000	19.80000	23.60000	28.30000	33.70000	39.10000	42.60000	45.70000	49.60000	52.30000
30	13.80000	15.00000	16.80000	18.50000	20.60000	24.50000	29.30000	34.80000	40.30000	43.80000	47.00000	50.90000	53.70000

שאלות

- (1) במטרה לבדוק האם קובייה הוגנת, מטילים אותה 120 פעמים. התקבל 17 פעמים 1, 23 פעמים 2, 20 פעמים 3, 25 פעמים 4, 18 פעמים 5 ו-17 פעמים 6. מה מסקנתכם ברמת מובהקות של 5%?
- (2) מפעל מייצר סוכריות בצבעים כחול, אדום, ירוק וכתום. מעוניינים לבדוק שפרופורציית הסוכריות הכחולות גדולה פי 2 מכל צבע אחר. לצורך כך נדגמו באקראי 200 סוכריות והתקבל: 70 כחולות, 50 אדומות, 40 ירוקות והיתר כתומות. מה מסקנתכם ברמת מובהקות של 5%?
- (3) 200 איש נתבקשו לבחור ספרה באקראי והנה התוצאות שהתקבלו:
 18 איש בחרו בספרה 0, 24 איש בחרו בספרה 1, 17 איש בחרו בספרה 2, 19 איש בחרו בספרה 3, 20 איש בחרו בספרה 4, 18 איש בחרו בספרה 5, 22 איש בחרו בספרה 6 והיתר בחרו בספרות 7-9.
 א. על סמך התוצאות הללו האם בחירת הספרות אקראית?
 בדקו ברמת מובהקות של 2.5%.
 ב. תנו הערכה למובהקות התוצאה.
 ג. אם נגדיל את גודל המדגם פי 2 ונשמור על אותם יחסים של כמות האנשים במדגם שבחרו בספרות, כיצד הדבר ישפיע על ערכו של הסטטיסטי χ^2 ? מה תהיה המסקנה במקרה זה?
- (4) מעוניינים לבדוק האם קובייה היא הוגנת. הטילו את הקובייה פעמיים והתבוננו בסכום הוצאות. חזרו על התהליך 72 פעמים. להלן התוצאות שהתקבלו במדגם:
 מה המסקנה ברמת מובהקות של 5%?
- | מספר הטלות | סכום התוצאות |
|------------|--------------|
| 20 | 2-5 |
| 17 | 6-8 |
| 20 | 9-10 |
| 15 | 11-12 |
- (5) בפנס יש 4 סוללות. בבדיקה שנערכה ב-400 פנסים נמצאו סוללות פגומות לפי השכיחויות הבאות:
- | מספר הסוללות הפגומות | 0 | 1 | 2 | 3 ומעלה |
|----------------------|-----|-----|----|---------|
| שכיחות | 276 | 104 | 12 | 8 |
- מעוניינים לבדוק על סמך תוצאות מדגם אלה האם הסיכוי לסוללה פגומה הוא 20%. מהי רמת המובהקות המינימלית עבורה נכריע שהסיכוי לסוללה פגומה אינו 20%?

(6) להלן השערות מחקר: $H_0 : X \sim N(40, 2^2)$, $H_1 : else$

מספר הדגימות	X	מתחת 36	36-40	40-44	מעל 44
3A	50A	45A	2A		

תוצאות המדגם הן:

מהו ערכו המקסימלי של A עבורו נקבל את H_0 ברמת מובהקות של 5%?

תשובות סופיות

- (1) לא נדחה H_0 .
- (2) לא נדחה H_0 .
- (3) א. לא נדחה H_0 . ב. בין 0.95 ל-0.975. ג. יגדל פי 2; המסקנה לא תשתנה.
- (4) נכריע שהקובייה אינה הוגנת.
- (5) 0.005.
- (6) 14.

מבחן חי בריבוע לאי תלות בין משתנים – רקע

מבחן לאי תלות מטרתו לבדוק האם קיים קשר בין שני משתנים. שני המשתנים שנבדקים צריכים להיות מחולקים למספר קטגוריות.

מבנה המבחן:

השערות:

אין תלות בין המשתנים H_0 .

יש תלות בין המשתנים H_1 .

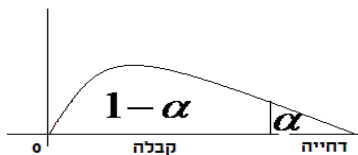
כלל הכרעה:

הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש $d.f = (r-1)(c-1)$. כאשר: r - מספר הקטגוריות של המשתנה שבשורות. c - מספר הקטגוריות של המשתנה שבעמודות.

הערך הקריטי הוא: $\chi^2_{1-\alpha, (r-1)(c-1)}$, כלומר האחוזון ה- $1-\alpha$ בהתפלגות חי בריבוע שדרגות החופש הן $(r-1)(c-1)$. אם $\chi^2 > \chi^2_{1-\alpha, (r-1)(c-1)}$ אז דוחים את השערת האפס.

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

כאשר:



O_i - השכיחות נצפית במדגם בתא i .

E_i - שכיחות צפויה במדגם בתא i בהנחת השערת האפס.

$$E_i = \frac{f(x) \cdot f(y)}{n}$$

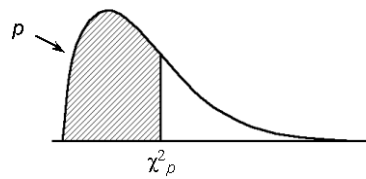
הערה:

תנאי כדי לבצע את המבחן הוא $E_i \geq 5$ לכל i . במידה ותנאי זה לא מתקיים יש אפשרות לאחד קטגוריות סמוכות עד שהתנאי יתקיים.
 תנאי חלופי: אין E קטן מ-1 וגם אין ביותר מ 20% מהתאים E קטן מ-5.

דוגמה (הפתרון בהקלטה):

האם יש תלות בין המגדר לבין דעה מסוימת?
 יש לבדוק ברמת מובהקות של 5% על סמך תוצאות הסקר:

המגדר / דעה	בעד	נגד	נמנע	סה"כ
גברים	50	40	10	
נשים	20	60	20	
סה"כ				

טבלת התפלגות חי-בריבוע – ערכי החלוקה χ^2_p 

df	p												
	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.00393	0.0157	0.03982	0.07393	0.158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7

שאלות

1) נבדקה התלות בין גודל הארגון לבין שביעות הרצון של העובדים.

להלן התוצאות:

גודל המפעל	שביעות רצון	נמוכה	בינונית	גבוהה	סה"כ
גדול	182	203	215	600	
קטן	154	110	136	400	
סה"כ	336	313	351	1000	

מה המסקנה ברמת מובהקות של 2.5%?

2) מפעל עובד בשלוש משמרות. להלן מספר המוצרים הפגומים והתקינים בכל

אחת מן המשמרות לפי מדגם שנעשה:

	לילה	ערב	יום
פגומים	70	60	50
תקינים	800	700	600

האם יש הבדל בין שיעורי הפגומים במשמרות השונות?

הסיקו עבור רמת מובהקות $\alpha = 0.05$.

3) נדגמו 50 מוצרים ממפעל מסוים מתוך 30 מוצרים שיוצרו ביום 17 נבחרו

לייצוא מתוך המוצרים שיוצרו בלילה 10 נבחרו לייצוא.

האם יש קשר בין היות מוצר לייצוא למועד שבו הוא יוצר?

בדקו ברמת בטחון של 95%.

תשובות סופיות

1) נסיק שיש קשר בין גודל הארגון לשביעות הרצון של העובדים.

2) נסיק שאין הבדל מובהק בין שיעור הפגומים במשמרות השונות.

3) נסיק שאין קשר בין היות מוצר לייצוא למועד שבו הוא יוצר.