מבחני מובהקות

הסקה סטטיסטית - סמסטר ב תשפ"א

שי מינצר

מבחן Z לממוצע יחיד

כאשר השונות באוכלוסייה ידועה וסטטיסטי המבחן מתפלג נורמלי או נורמלי בקירוב

נתונים (להוציא מתוך השאלה)

σ^2 - שונות האוכלוסייה

μ - תוחלת האוכלוסייה

$\bar{X}$- ממוצע המדגם

N - גודל המדגם

השערות

השערה דו צדדית	השערה חד צדדית ימנית	השערה חד צדדית שמאלית
$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$	$H_1: \mu > \mu_0$	$H_1: \mu < \mu_0$

הנחות

- דגימה מקרית ובלתי תלויה של התצפיות
- המשתנה מתפלג נורמלי באוכלוסייה / התפלגות הדגימה של הממוצע מתפלגת נורמלית בקירוב (לפי משפט הגבול המרכזי, כאשר N < 30)

קביעת רמת מובהקות ואזורי דחייה וקבלה

- α, סוג מבחן (דו זנבי / חד זנבי ימני / חד זנבי שמאלי)
- $Z_C(\alpha)$
- איזור דחייה
- איזור קבלה

מבחן סטטיסטי

טעות התקן (סטיית התקן של התפלגות הדגימה של הממוצע)

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}$$

סטטיסטי המבחן

$$Z_{\bar{x}} = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

החלטה

מבחן t לממוצע יחיד

כאשר השונות באוכלוסייה לא ידועה וסטטיסטי המבחן מתפלג t עם N-1 דרגות חופש

נתונים (להוציא מתוך השאלה)

μ - תוחלת האוכלוסייה

$\bar{X}$- ממוצע המדגם

N - גודל המדגם

S^2 - שונות המדגם

השערות

השערה דו צדדית	השערה חד צדדית ימנית	השערה חד צדדית שמאלית
$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$	$H_1: \mu > \mu_0$	$H_1: \mu < \mu_0$

הנחות

- דגימה מקרית ובלתי תלויה של התצפיות
- המשתנה מתפלג נורמלי באוכלוסייה

קביעת רמת מובהקות ואזורי דחייה וקבלה

- α, סוג מבחן (דו זנבי / חד זנבי ימני / חד זנבי שמאלי)
- $df = N - 1$
- $t_C(df, \alpha)$
- איזור דחייה
- איזור קבלה

מבחן סטטיסטי

האומד חסר ההטיה לסטיית התקן של האוכלוסייה

$$\hat{S}_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N-1}} = \sqrt{\frac{S_x^{\,2} \cdot N}{N-1}}$$

טעות התקן (סטיית התקן של התפלגות הדגימה של הממוצע)

$$\hat{S}_{\bar{x}} = \frac{\hat{S}_x}{\sqrt{N}}$$

סטטיסטי המבחן

$$t_{\bar{x}} = \frac{\bar{X} - \mu_{\bar{X}}}{\hat{S}_{\bar{x}}}$$

החלטה

מבחן Z לפרופורציה יחידה

כאשר המשתנה מתפלג בינומית

השערות

השערה דו צדדית	השערה חד צדדית ימנית	השערה חד צדדית שמאלית
$H_0: p = p_0$	$H_0: p = p_0$	$H_0: p = p_0$
$H_1: p \neq p_0$	$H_1: p > p_0$	$H_1: p < p_0$

הנחות

- דגימה מקרית ובלתי תלויה של התצפיות
- התפלגות הדגימה מתפלגת נורמלית בקירוב - לפי הקירוב הנורמלי של הבינום

(**נדרש:** $N \cdot \hat{q} > 5$ **וגם:** $N \cdot \hat{p} > 5$)

קביעת רמת מובהקות ואזורי דחייה וקבלה

- α
- $Z_C(\alpha)$
- איזור דחייה
- איזור קבלה

מבחן סטטיסטי

טעות התקן

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{q} \cdot \hat{p}}{N}}$$

סטטיסטי המבחן

$$Z_{\hat{p}} = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}}$$

החלטה

מבחן Z להפרש תוחלות במדגמים בלתי תלויים

כאשר השונויות באוכלוסיות ידועות ושוות

נתונים (להוציא מתוך השאלה)

שונויות האוכלוסיות - σ^2_1 , σ^2_2

ממוצעי המדגמים - $\bar{x}_1 , \bar{x}_2$

גדלי המדגמים - N_1 , N_2

השערות

השערה חד צדדית שמאלית	השערה חד צדדית ימנית	השערה דו צדדית
$H_0: \mu_1 - \mu_2 = d_0$	$H_0: \mu_1 - \mu_2 = d_0$	$H_0: \mu_1 - \mu_2 = d_0$
$H_1: \mu_1 - \mu_2 < d_0$	$H_1: \mu_1 - \mu_2 > d_0$	$H_1: \mu_1 - \mu_2 \neq d_0$

הנחות

- דגימה מקרית ובלתי תלויה של התצפיות
- המשתנה מתפלג נורמלי בשתי האוכלוסיות

קביעת רמת מובהקות ואזורי דחייה וקבלה

- α, סוג מבחן (דו זנבי / חד זנבי ימני / חד זנבי שמאלי)
- $Z_C(\alpha)$
- איזור דחייה
- איזור קבלה

מבחן סטטיסטי

טעות התקן

$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma^2_1}{N_1} + \frac{\sigma^2_2}{N_2}}$$

סטטיסטי המבחן

$$Z_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sigma_{(\bar{x}_1 - \bar{x}_2)}}$$

החלטה

מבחן t להפרש תוחלות במדגמים בלתי תלויים

כאשר השונויות באוכלוסיות לא ידועות

נתונים (להוציא מתוך השאלה)

$\bar{x}_1$, $\bar{x}_2$ - ממוצעי המדגמים

N_1 , N_2 - גדלי המדגמים

S^2_1 , S^2_2 - שונויות המדגמים

השערות

השערה חד צדדית שמאלית	השערה חד צדדית ימנית	השערה דו צדדית
$H_0: \mu_1 - \mu_2 = d_0$	$H_0: \mu_1 - \mu_2 = d_0$	$H_0: \mu_1 - \mu_2 = d_0$
$H_1: \mu_1 - \mu_2 < d_0$	$H_1: \mu_1 - \mu_2 > d_0$	$H_1: \mu_1 - \mu_2 \neq d_0$

הנחות

- דגימה מקרית ובלתי תלויה של התצפיות
- המשתנה מתפלג נורמלי בשתי האוכלוסיות
- שוויון שונויות בשתי האוכלוסיות

קביעת רמת מובהקות ואזורי דחייה וקבלה

- α, סוג מבחן (דו זנבי / חד זנבי ימני / חד זנבי שמאלי)
- $df = N_1 + N_2 - 2$
- $t_C(df, \alpha)$
- איזור דחייה
- איזור קבלה

מבחן סטטיסטי

האומד חסר ההטיה לסטיית התקן של האוכלוסייה

$$\hat{S}_x = \sqrt{\frac{\Sigma(x_1 - \bar{x}_1)^2 + \Sigma(x_2 - \bar{x}_2)^2}{N_1 + N_2 - 2}} = \sqrt{\frac{S^2_{x1\_} \cdot N_1 + S^2_{x2\_} \cdot N_2}{N_1 + N_2 - 2}}$$

טעות התקן (סטיית התקן של התפלגות הדגימה של הממוצע)

$$\hat{S}_{(\bar{x}_1 - \bar{x}_2)} = \hat{S}_x \cdot \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

סטטיסטי המבחן

$$t_{(\bar{x}_1 - \bar{x}_2)} = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\hat{S}_{(\bar{x}_1 - \bar{x}_2)}}$$

החלטה

מבחן t לתוחלת ההפרשים במדגמים תלויים / מזווגים

כאשר השונויות באוכלוסיות לא ידועות

נתונים (להוציא מתוך השאלה)

$\overline{x}_1$, $\overline{x}_2$ - ממוצעי המדגמים

N - גודל המדגמים

S^2 - שונות המדגמים

השערות

השערה חד צדדית שמאלית	השערה חד צדדית ימנית	השערה דו צדדית
$H_0: \mu_d = d_0$	$H_0: \mu_d = d_0$	$H_0: \mu_d = d_0$
$H_1: \mu_d < d_0$	$H_1: \mu_d > d_0$	$H_1: \mu_d \neq d_0$

הנחות

- דגימה מקרית ובלתי תלויה של התצפיות
- המשתנה (d) מתפלג נורמלית באוכלוסייה

קביעת רמת מובהקות ואזורי דחייה וקבלה

- α, סוג מבחן (דו זנבי / חד זנבי ימני / חד זנבי שמאלי)
- $df = N - 1$
- $t_C(df, \alpha)$
- איזור דחייה
- איזור קבלה

מבחן סטטיסטי

האומד חסר ההטיה לסטיית התקן של האוכלוסייה (אוכלוסיית ההפרשים)

$$\hat{S}_d = \sqrt{\frac{\Sigma(d - \overline{d})^2}{N - 1}}$$

טעות התקן (סטיית התקן של התפלגות הדגימה של הממוצע)

$$\hat{S}_{\overline{d}} = \frac{\hat{S}_d}{\sqrt{N}}$$

סטטיסטי המבחן

$$t_{\overline{d}} = \frac{\overline{d} - d_0}{\hat{S}_{\overline{d}}}$$

החלטה

מבחן F למובהקות שוויון שונויות

השערות

$H_0: \sigma^2_1 = \sigma^2_2$

$H_1: \sigma^2_1 \neq \sigma^2_2$

הנחות

- דגימה מקרית ובלתי תלויה של התצפיות
- המשתנה מתפלג נורמלי בשתי האוכלוסיות

קביעת רמת מובהקות ואזורי דחייה וקבלה

- α
- $df_1 = n_1 - 1 \quad ; \quad df_2 = n_2 - 1$
- $F_C(df_1, df_2, \alpha)$
- איזור דחייה
- איזור קבלה

מבחן סטטיסטי

האומד חסר ההטיה לשונות האוכלוסייה

$$\hat{S}_1^2 = \frac{\Sigma(x-\bar{x}_1)^2}{n_1 - 1} = \frac{S_1^2 \cdot N}{n_1 - 1}$$

$$\hat{S}_2^2 = \frac{\Sigma(x-\bar{x}_2)^2}{n_2 - 1} = \frac{S_2^2 \cdot N}{n_2 - 1}$$

סטטיסטי המבחן

$$F = \frac{\hat{S}_1^2}{\hat{S}_2^2}$$

החלטה

מבחן t למובהקות מתאם פירסון

השערות

$H_0: \rho = 0$

$H_1: \rho \neq 0$

הנחות

- דגימה מקרית ובלתי תלויה של התצפיות
- התפלגות האוכלוסיה היא נורמלית דו-משתנית

קביעת רמת מובהקות ואזורי דחייה וקבלה

- α
- $df = N - 2$
- $t_C(df, \alpha)$
- איזור דחייה
- איזור קבלה

מבחן סטטיסטי

$$t = r\sqrt{\frac{N-2}{1-r^2}}$$

החלטה

ניתוח שונות פשוט (חד גורמי)

השערות

$H_0: \mu_1 = \mu_2 = ... = \mu_n$

H_1: לפחות אחת מהתוחלות שונה

הנחות

- דגימה מקרית ובלתי תלויה של התצפיות במדגמים
- המשתנה מתפלג נורמלית באוכלוסיות
- שוויון / הומוגניות שונויות
- המדגמים בלתי תלויים

קביעת רמת מובהקות ואזורי דחייה וקבלה

- α, מבחן חד זנבי ימני
- $F_C(df_B, df_W, \alpha)$
- איזור קבלה
- איזור דחייה

מבחן סטטיסטי

	SS	DF	MS
B	$SSB = \sum n_j(\bar{x}_j - \bar{\bar{x}})^2$ $SSB = df_B \cdot MSB$ $SSB = SST - SSW$	$df_B = J - 1$ $df_B = SSB/MSB$ $df_B = df_T - df_B$	$MSB = SSB/df_B$
W	$SSW = \sum_j \sum_i (x_{ij} - \bar{x}_j)^2$ $SSW = df_W \cdot MSW$ $SSW = SST - SSB$	$df_W = N - J$ $df_W = SSW/MSW$ $df_W = df_T - df_B$	$MSW = SSW/df_W$
T	$SST = \sum_j \sum_i (x_{ij} - \bar{\bar{x}}_j)^2$ $SST = df_T \cdot MST$ $SST = SSB + SSW$	$df_T = N - 1$ $df_T = SST/MST$ $df_T = df_B + df_w$	$MST = SST/df_T$

$$F = \frac{MSB}{MSW}$$

החלטה

ניתוח המשך - קונטרסט

K-1 השוואות אורתוגונליות - סידור אפשרי

	Group 1	Group 2	Group 3	Group 4	Group 5
C_1	-1	1	0	0	0
C_2	-1	-1	2	0	0
C_3	-1	-1	-1	3	0
C_4	-1	-1	-1	-1	4

בדיקת אורתוגונליות

עבור שלוש קבוצות

$$\Sigma C_{j1}\cdot C_{j2}=(-1\cdot-1)+(1\cdot-1)+(0\cdot2)=0$$

עבור ארבע קבוצות

$$\Sigma C_{j1}\cdot C_{j2}=(-1\cdot-1)+(1\cdot-1)+(0\cdot2)+(0\cdot0)=0$$
$$\Sigma Cj_{1}\cdot C_{j3}=(-1\cdot-1)+(1\cdot-1)+(0\cdot-1)+(0\cdot3)=0$$
$$\Sigma C_{j2}\cdot C_{j3}=(-1\cdot-1)+(-1\cdot-1)+(2\cdot-1)+(0\cdot3)=0$$

עבור חמש קבוצות

$$\Sigma C_{j1}\cdot C_{j2}=(-1\cdot-1)+(1\cdot-1)+(0\cdot2)+(0\cdot0)+(0\cdot0)=0$$
$$\Sigma C_{j1}\cdot C_{j3}=(-1\cdot-1)+(1\cdot-1)+(0\cdot-1)+(0\cdot3)+(0\cdot0)=0$$
$$\Sigma C_{j1}\cdot C_{j4}=(-1\cdot-1)+(1\cdot-1)+(0\cdot-1)+(0\cdot-1)+(0\cdot4)=0$$
$$\Sigma C_{j2}\cdot C_{j3}=(-1\cdot-1)+(-1\cdot-1)+(2\cdot-1)+(0\cdot3)+(0\cdot0)=0$$
$$\Sigma C_{j2}\cdot C_{j4}=(-1\cdot-1)+(-1\cdot-1)+(2\cdot-1)+(0\cdot-1)+(0\cdot4)=0$$

מבחן סטטיסטי - עבור כל סט משקולות

$$df_{comp}=1\quad;\quad SS_{COMP}=\frac{n[\Sigma C_J\cdot\bar{X}_J]}{\Sigma C_J^{\,2}}$$

$$MS_{COMP}=\frac{SS_{COMP}}{df_{COMP}}=\frac{SS_{COMP}}{1}=SS_{COMP}$$

$$F_C(df_{comp},df_W,\alpha)$$

$$F=\frac{MS_{COMP}}{MS_W}$$

החלטה

מבחן חי בריבוע לטיב התאמה

נתונים (להוציא מתוך השאלה)

$P_1, P_2, P_3 ... P_n$ - השכיחות היחסית באוכלוסייה של כל קטגוריה

$O_1, O_2, O_3 ... O_n$- השכיחות של כל קטגוריה כפי שהתקבלה במדגם

N - גודל המדגם

השערות

H_0: התפלגות כמו באוכלוסייה

H_1: התפלגות לא כמו באוכלוסייה

הנחות

- דגימה מקרית ובלתי תלויה של התצפיות
- הקטגוריות מוציאות וממצות - כל תצפית נופלת בקטגוריה אחת בלבד
- המדגם מספיק גדול
  - אם df = 1: בכל קטגוריה $Ex \geq 10$
  - אם df > 1: בכל קטגוריה $Ex \geq 5$

קביעת רמת מובהקות ואזורי דחייה וקבלה

- α, מבחן חד זנבי ימני
- $df = j - 1$
- $\chi^2_C(df, \alpha)$
- איזור קבלה
- איזור דחייה

מבחן סטטיסטי

	קטגוריה 1	קטגוריה 2	קטגוריה 3
Expected	$E_1 = P_1 \cdot N$	$E_2 = P_3 \cdot N$	$E_3 = P_3 \cdot N$
Observed	O_1	O_2	O_3

$$\chi^2 = \sum_j \left[\frac{(O_j - E_j)^2}{E_j}\right]$$

החלטה

מבחן חי בריבוע לאי תלות

השערות

H_0: אין תלות/קשר בין <משתנה 1> לבין <משתנה 2>

H_1: יש תלות/קשר בין <משתנה 1> לבין <משתנה 2>

הנחות

- דגימה מקרית ובלתי תלויה של התצפיות
- הקטגוריות מוציאות וממצות - כל תצפית נופלת בקטגוריה אחת בלבד
- המדגם מספיק גדול
  - אם df = 1: בכל קטגוריה $Ex \geq 10$
  - אם df > 1: בכל קטגוריה $Ex \geq 5$

קביעת רמת מובהקות ואזורי דחייה וקבלה

- α, מבחן חד זנבי ימני
- $df = (Rows - 1) \cdot (Columns - 1)$
- $\chi^2_C(df, \alpha)$
- איזור קבלה
- איזור דחייה

מבחן סטטיסטי

Observed

	קטגוריה 1	קטגוריה 2	קטגוריה 3	סה"כ
ערך 1	O_{11}	O_{12}	O_{13}	ΣO_{1k}
ערך 2	O_{21}	O_{22}	O_{23}	ΣO_{2k}
ערך 3	O_{31}	O_{32}	O_{33}	ΣO_{3k}
סה"כ	ΣO_{j1}	ΣO_{j2}	ΣO_{j3}	$N = \Sigma\Sigma O_{jk}$

Expected

	קטגוריה 1	קטגוריה 2	קטגוריה 3
ערך 1	$E_{11} = \frac{\Sigma O_{1k} \cdot \Sigma O_{j1}}{N}$	$E_{12} = \frac{\Sigma O_{1k} \cdot \Sigma O_{j2}}{N}$	$E_{13} = \frac{\Sigma O_{1k} \cdot \Sigma O_{j3}}{N}$
ערך 2	$E_{21} = \frac{\Sigma O_{2k} \cdot \Sigma O_{j1}}{N}$	$E_{22} = \frac{\Sigma O_{2k} \cdot \Sigma O_{j2}}{N}$	$E_{23} = \frac{\Sigma O_{2k} \cdot \Sigma O_{j3}}{N}$
ערך 3	$E_{31} = \frac{\Sigma O_{3k} \cdot \Sigma O_{j1}}{N}$	$E_{32} = \frac{\Sigma O_{3k} \cdot \Sigma O_{j1}}{N}$	$E_{33} = \frac{\Sigma O_{3k} \cdot \Sigma O_{j3}}{N}$

$$\chi^2 = \sum_j \sum_k \left[\frac{(O_{jk} - E_{jk})^2}{E_{jk}}\right]$$

החלטה