

כללי	רווח סמך ו-SE	התפלגות האומד:	התקן	דחייה אם: $H_0: \theta = \theta_0$ $H_1: \theta \neq \theta_0$	דחייה אם: $H_0: \theta = \theta_0$ $H_1: \theta < \theta_0$	דחייה אם: $H_0: \theta = \theta_0$ $H_1: \theta > \theta_0$	התנאים	הפרמטר
$n \geq \left(\frac{(Z_{1-\alpha} + Z_{1-\beta}) \cdot \sigma}{\mu_0 - \mu_1} \right)^2$ $\bar{X} = \frac{A+B}{2}$ $n \geq \left(\frac{2 \cdot z_{\frac{\alpha}{2}} \cdot \sigma}{d_0} \right)^2$	$\bar{x} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ $S.E. = \frac{\sigma}{\sqrt{n}}$	$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$	$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$	או $\bar{X} > \mu_0 + Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ $\bar{X} < \mu_0 - Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	1. $x \sim N$ $n \geq 30$ 2. σ^2 ידועה	ממוצע יחיד- μ
	$\bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$ $S.E. \approx \frac{S}{\sqrt{n}}$	$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$	$Z = \frac{\bar{x} - \mu}{S/\sqrt{n}}$	או $\bar{X} > \mu_0 + Z_{1-\alpha/2} \cdot \frac{S}{\sqrt{n}}$ $\bar{X} < \mu_0 - Z_{1-\alpha/2} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{S}{\sqrt{n}}$	1. $n \geq 30$ 2. σ^2 אינה ידועה	μ
n-1=חופש=דרגות	$\bar{x} \pm t_{1-\frac{\alpha}{2}}^{(n-1)} \cdot \frac{S}{\sqrt{n}}$ $S.E. \approx \frac{S}{\sqrt{n}}$	$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$	$T = \frac{\bar{x} - \mu}{S/\sqrt{n}}$	או $\bar{X} > \mu_0 + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$ $\bar{X} < \mu_0 - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} < \mu_0 - t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} > \mu_0 + t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	1. $x \sim N$ $n < 30$ 2. σ^2 אינה ידועה	
$n \geq \left(\frac{Z_{1-\alpha} \sqrt{p_0 q_0} + Z_{1-\beta} \sqrt{p_1 q_1}}{p_0 - p_1} \right)^2$ $\hat{p} = \frac{A+B}{n}$	$\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$ $S.E. = \sqrt{\frac{\hat{p}\hat{q}}{n}} \approx \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq \sqrt{\frac{1}{4n}}$	$\hat{p} \sim N(p, \frac{pq}{n})$	$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$	או $\hat{p} < p_0 - Z_{1-\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$ $\hat{p} > p_0 + Z_{1-\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} < p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} > p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$np \geq 5$ $n(1-p) \geq 5$ ברווח סמך: $n \geq 30$	פרופורציה יחיד p
	$(\bar{x}_1 - \bar{x}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ $S.E. = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$	$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	או $\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ $\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	1. σ_1^2, σ_2^2 ידועות. 2. $x_1, x_2 \sim N$ $n_1, n_2 > 30$	הפרש ממוצעים במדגמים בלתי תלויים: $\mu_1 - \mu_2$
	$(\bar{x}_1 - \bar{x}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$	$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	או $\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	1. σ_1^2, σ_2^2 אינן ידועות $n_1 \geq 30$ $n_2 \geq 30$	$\mu_1 - \mu_2$
$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$ דרגות חופש = $n_1 + n_2 - 2$	$(\bar{x}_1 - \bar{x}_2) \pm t_{1-\frac{\alpha}{2}}^{n_1+n_2-2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ $S.E. \approx s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$\bar{x}_1 - \bar{x}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$	$T = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$\bar{x}_1 - \bar{x}_2 > c + t_{1-\alpha/2}^{n_1+n_2-2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ או $\bar{x}_1 - \bar{x}_2 < c - t_{1-\alpha/2}^{n_1+n_2-2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$\bar{x}_1 - \bar{x}_2 < c - t_{1-\alpha}^{n_1+n_2-2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$\bar{x}_1 - \bar{x}_2 > c + t_{1-\alpha}^{n_1+n_2-2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	1. $x_1, x_2 \sim N$ $n_1 < 30$ $n_2 < 30$ 3. סטיות התקן אינן ידועות אך $\sigma_1^2 = \sigma_2^2$	כאשר $H_0: \mu_1 - \mu_2 = C$
$\hat{p} = \frac{y_1 + y_2}{n + n} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$	$(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$ $S.E. \approx \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$	$\hat{p}_1 - \hat{p}_2 \sim N(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2})$	$Z_{\hat{p}_1 - \hat{p}_2} = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$ $Z_{\hat{p}_1 - \hat{p}_2} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$	או $\hat{p}_1 - \hat{p}_2 > 0 + Z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$ $\hat{p}_1 - \hat{p}_2 < 0 - Z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$	$\hat{p}_1 - \hat{p}_2 < 0 - Z_{1-\alpha} \cdot \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$	$\hat{p}_1 - \hat{p}_2 > 0 + Z_{1-\alpha} \cdot \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$		הפרש פרופורציות במדגמים בלתי תלויים: $H_0: p_1 - p_2 = 0$
n-1=חופש=דרגות	$\bar{D} \pm t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ $S.E. \approx \frac{S_D}{\sqrt{n}}$	$\bar{D} \sim N(\mu_D, \frac{\sigma_D^2}{n})$	$T_{\bar{D}} = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}}$	או $\bar{D} > C + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ $\bar{D} < C - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} < C - t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} > C + t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	1. המדגם מזווג 2. D מתפלג נורמלית.	תוחלת הפרש במדגמים מזווגים. $H_0: \mu_D = C$

התפלגות ממוצע המדגם \bar{X}

א. תכונות: $E(\bar{x}) = \mu$ - תוחלת ממוצע המדגם שווה לממוצע האוכלוסייה- כלומר ממוצע של כל ממוצעי המדגם האפשריים נותן את ממוצע האוכלוסייה האמיתי - הפרמטר.

$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}}$: סטיית התקן : $v(\bar{x}) = \frac{v(x)}{n} = \frac{\sigma^2}{n}$ - השונות של ממוצעי המדגם שווה לשונות האוכלוסייה לחלק ל n - גודל המדגם. סטיית התקן :

ב. משפט: אם האוכלוסייה מתפלגת נורמלית $x \sim N(\mu, \sigma^2)$ ונדגום n תצפיות אז ממוצע המדגם גם מתפלג נורמלית $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ ואז $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

ג. משפט הגבול המרכזי אם x מתפלג כלשהו וידוע $E(x) = \mu$ אזי עבור מדגם מספיק גדול (לפחות 30) $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ $V(x) = \sigma^2$

התפלגות סכום התצפיות במדגם **א. תכונות:** $V(\sum x_i) = n\sigma^2$ $E(\sum x_i) = n\mu$ אם $x \sim N(\mu, \sigma^2)$ אזי עבור מדגם מספיק גדול (לפחות 30) $\sum x_i \sim N(n\mu, n\sigma^2)$ ואז $Z = \frac{\sum x_i - n\mu}{\sqrt{n\sigma^2}}$

ג. משפט הגבול המרכזי: אם x מתפלג כלשהו וידוע $E(x) = \mu$ אזי עבור מדגם מספיק גדול (לפחות 30) $\sum x_i \sim N(n\mu, n\sigma^2)$

התפלגות מספר ההצלחות במדגם - התפלגות בינומית Y:

הסיכוי להצלחה p (גודל זה הוא פרמטר המתאר את האוכלוסייה) הסיכוי לכישלון q=1-p

$y \sim B(n, p)$

$E(Y) = np$
 $V(Y) = npq$
 $\sigma(Y) = \sqrt{npq}$
 $p(y = k) = \binom{n}{k} p^k q^{n-k}$
 $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

קירוב נורמלי להתפלגות הבינומית:

אם $y \sim B(n, p)$ והתנאי $np \geq 5$ וגם $n(1-p) \geq 5$ מתקיים אז $y \sim N(np, npq)$, $Z_y = \frac{y - np}{\sqrt{npq}}$ **לא לשכוח תיקון רציפות**

התפלגות פרופורציית ההצלחות במדגם: $\hat{p} = \frac{y}{n}$ אם באוכלוסייה פרופורציית ההצלחות היא P ו $np \geq 5$ וגם $n(1-p) \geq 5$ נאמר שהתפלגות פרופורציית ההצלחות במדגם \hat{p} היא נורמלית :

$E(\hat{P}) = p, V(\hat{P}) = \frac{pq}{n}$ $\sigma(\hat{P}) = \sqrt{\frac{p(1-p)}{n}} = SE = \sqrt{\frac{1}{4n}}$

אם באוכלוסייה פרופורציית ההצלחות היא P ו $np \geq 5$ וגם $n(1-p) \geq 5$ נאמר שהתפלגות פרופורציית ההצלחות במדגם \hat{p} היא נורמלית :

$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} - 1$ $\hat{p} \sim N(p, \frac{pq}{n})$

חוק המספרים הגדולים (כל מה שנאמר על פרופורציה נכון גם לגבי ממוצע):

חוק המספרים הגדולים מתייחס להשפעת גודל המדגם על הסיכוי של פרופורציית המדגם להיות קרובה או רחוקה מהפרופורציה האמיתית:

- ככל שהמדגם גדול יותר כך הסיכוי שפרופורציית המדגם (\hat{p}) תהיה בקרבת הפרופורציה באוכלוסייה (P) גבוהה יותר.
- ואילו הסיכוי לקבל ערך חריג הרחוק מהפרופורצייה של האוכלוסייה קטן יותר.
- אולם הסיכוי שפרופורציית המדגם תהיה בדיוק הפרופורצייה האמיתית הולכת וקטנה ככל שהמדגם גדול יותר.

$H_0 = P$ (נכונה) | לקבל את תוצאות המדגם וקיצוני P_v

אם $\alpha \geq P_v$ דוחים את H_0 .

הערה: במבחן דו צדדי יש להכפיל את הסיכוי פי 2 כדי לקבל P_v .

מובהקות התוצאה

$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$ - שונות המדגם (אומד חסר הטויה ל- σ^2)

הגדרות:

- פרמטר** הוא גודל קבוע המאפיין את האוכלוסייה (כמו μ) ואילו **סטטיסטי** הוא משתנה המחושב על תוצאות המדגם (כמו \bar{x})
- אומד חסר הטויה**: אם התוחלת של $\hat{\theta}$ תהיה שווה ל θ , כלומר $E(\hat{\theta}) = \theta$
- טעות התקן** של אומד היא סטיית התקן שלו, כלומר: $\sigma(\hat{\theta}) = S.E$
- שגיאת אמידה** הינו ההפרש בערך מוחלט בין האומד לפרמטר כלומר $|\hat{\theta} - \theta|$
- רמת סמך $1 - \alpha$** = הסיכוי שהפרמטר נופל ברווח הסמך כלומר: $P(A \leq \theta \leq B) = 1 - \alpha$
- טעות מסוג ראשון** - להכריע לדחות את H_0 למרות שבמציאות H_0 נכונה.
- טעות מסוג שני** - להכריע לקבל את H_0 למרות שבמציאות H_1 נכונה.
- תוצאה מובהקת** - תוצאה שיצאה באזור דחיית השערת האפס.
- אורך רווח הסמך $d = B - A = 2\varepsilon$**

$\beta = P(H_1 | \text{לקבל את } H_0) = P(\text{טעות מסוג שני})$
 $\pi = (1 - \beta) = P(H_0 | \text{לדחות את } H_0) = \text{עוצמה}$

$\alpha = P(H_0 | \text{לדחות את } H_0) = P(\text{טעות מסוג ראשון}) = \text{רמת מובהקות}$
 H_0 נכונה | לקבל את H_0 = **רמת בטחון $(1 - \alpha)$**

מבחן טיב התאמה: מבחן שמטרתו לבדוק האם האוכלוסייה מתפלגת לפי התאמה מסוימת.

השערות: ישנה התאמה מסוימת: H_0

אחרת: H_1

כלל הכרעה: $c-1$ דרגות החופש: כאשר c זה מספר הקטגוריות. הערך הקריטי הוא: $k = \chi^2_{1-\alpha, c-1}$ נידחה את השערת האפס אם הערך הסטטיסטי יהיה גדול מהערך הקריטי.

סטטיסטי המבחן: $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$ - שכיחות נצפית של קטגוריה i - O_i - הסתברות לצפוייה של קטגוריה i לפי השערת האפס. $E_i = np_i$ שכיחות צפויה של קטגוריה i.

מבחן לאי תלות בין משתנים: מבחן לאי תלות מטרותו לבדוק האם קיים קשר בין שני משתנים: שני המשתנים שנבדקים צריכים להיות מחולקים למספר קטגוריות. למשל, X - מין; Y - זכר ונקבה. דעה פוליטית: ימין, מרכז שמאל.

השערות: אין תלות בין X ל- Y H_0

יש תלות בין X ל- Y H_1

כלל הכרעה: נדחה את H_0 אם נקבל כי הערך הסטטיסטי גבוה מהערך הקריטי.

$df = (l-1)(m-1)$ $k = \chi^2_{1-\alpha, df} = 1$ מספר הקטגוריות בעמודות (למשל: ימין, מרכז, שמאל) 3 = m מספר הקטגוריות בשורות (למשל: נשים, גברים) 2 = l

הסטטיסטי: $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$ - שכיחות נצפית של משבצת i - O_i - הסתברות למשבצת i לפי השערת האפס. $E_i = \frac{f_{x'} \cdot f_{y'}}{n}$ שכיחות צפויה של משבצת i.