משרד החינוך
המינהל למדע ולטכנולוגיה

תכנית לימודים

שם התכנית: **ניהול תעשייתי ב**

מקצוע: **סטטיסטיקה ב**

כיתה: **י"ד**

מהדורת ניסוי

מעודכן למרץ 2013

התפיסה הרעיונית של התכנית

בעולם העסקים המודרני המתנהל באמצעות מאגרי מידע, מחשוב ותחרות כלל-עולמית, הסטטיסטיקה היא כלי יעיל ובעל ערך ניהולי. הבנה כמותית ומיומנות בסיסית בסטטיסטיקה, חיוניות לא רק לעוסקים במחקר אלא גם לאנשי טכנולוגיה, לאנשי מנהל ולאנשי מדעי הטבע והחברה, הזקוקים לכלים סטטיסטיים לצורך עבודתם או לימודיהם.

הסטטיסטיקה משמשת יסוד לנושאים כמו בקרת איכות, בקרת תהליכים, מחקרי שווקים וניהול משאבי אנוש. היא מתבססת על כלים וטכניקות להצגת נתונים כמותיים, ניתוחם ופירושם לצורך פתרון בעיות בנושאים שונים.

חשיבה סטטיסטית פירושה היכולת להבין ולנתח נתונים כמותיים ולתקשר בשפה כמותית. לסטטיסטיקה יש שני כיוונים עיקריים:

- **סטטיסטיקה תיאורית** – עוסקת בשיטות לארגון, לתיאור ולתמצות הנתונים שנאספו במחקר הסטטיסטי.
- **הסקה סטטיסטית** (שבה מתמקד מקצוע זה) – עוסקת בשיטות להכללה מתוצאות של מדגם מייצג לכלל האוכלוסייה.

המקצוע סטטיסטיקה ב מרחיב את הידע שנרכש במקצוע סטטיסטיקה א – שבו הוקנו הכלים והטכניקות להצגת נתונים, ניתוחם ופירושם. לימודי הסקה סטטיסטית הנלמדים במסגרת המקצוע "סטטיסטיקה ב" מַקנים כלים נוספים בנוגע להסקה ממדגם מייצג לגבי כלל האוכלוסייה ובדיקת השערות בתהליכי מחקר.

רוב המחקרים הסטטיסטיים מבוססים על מדגם שנבחר מתוך אוכלוסיית המחקר ולא על כלל האוכלוסייה שבה מעוניינים. תוצאות המדגם מוכללות על ידי החוקר לכלל האוכלוסייה. תהליך זה של הסקה סטטיסטית דורש מדגם המייצג את האוכלוסייה, כך שלכל מקרה באוכלוסייה תהיה הסתברות להיכלל במדגם. יש להביא בחשבון כי גם תהליך של הסקה ממדגם מייצג לאוכלוסייה אינו מדויק לחלוטין וכי יש בו רכיב של סיכון (סיכוי לטעות). המסקנה המחקרית חייבת אפוא להביא בחשבון את ההסתברות לטעות במסקנה.

המקצוע הסקה סטטיסטית העומד בלב תכנית לימודים זו, מתמקד בהצגת השיטות העיקריות להסקה מתוך מדגם לאוכלוסייה ולבדיקת השערות לגבי פרמטרים שונים. הדגש הוא בהיבט היישומי, במטרה להקנות ללומד כלים סטטיסטיים נוספים לפתרון בעיות בתחומי הניהול השונים.

מטרות כלליות

1. הכרת חשיבותה של ההסקה הסטטיסטית במחקרים מגוונים בתחומי הניהול.
2. שימוש בטכניקות מתחום הסטטיסטיקה ההיסקית לצורך ניתוח ממצאים והסקת מסקנות.
3. הקניית כלים כמותיים ומיומנויות ליישום טכניקות סטטיסטיות מתחום הסטטיסטיקה ההיסקית בתהליכי מחקר בעלי ערך ניהולי.

מטרות אופרטיביות

עם סיום לימודיהם, התלמידים:

1. יסבירו את חשיבותה של ההסקה הסטטיסטית בתהליכי מחקר בעלי ערך ניהולי .
2. יפרטו את הנושאים המרכזיים בתחום ההסקה הסטטיסטית.
3. יתארו את סוגי הטעויות הסטטיסטיות האפשריות בתהליך ההסקה הסטטיסטית.
4. יבחינו בין השיטות השונות להסקה מאומד למשתנה (סטטיסטי לפרמטר).
5. יציגו את המקרים השונים להסקה לגבי ממוצע האוכלוסייה μ.
6. יציגו את דרך ההסקה לגבי הפרופורציה p באוכלוסייה.
7. יבחנו את טיב ההתאמה ומידת התלות בין שני משתנים, תוך שימוש במבחני χ^2.
8. יפרשו ממצאי ניתוח שונות חד-כיווני ודו-כיווני באמצעות פלטי מחשב.
9. יבצעו יישומים סטטיסטיים ממוחשבים במסגרת תרגיל מעשי.

דרכי הוראה/למידה מומלצות

יש להשתמש בשיטות הוראה ובעזרי הוראה מגוונים. יש לשלב דרכי הוראה קונבנציונליות בחלופות ייחודיות התורמות להגברת המוטיבציה של תהליך הלמידה כגון:

1. ניתוח אירועים ונתונים ממקורות מידע שונים.
2. שימוש בפרסומי הלשכה המרכזית לסטטיסטיקה לצורך בניית תרגילים יישומיים.
3. התנסות ביישומים סטטיסטיים ממוחשבים.
4. עבודות ותרגילים הנשענים על בסיסי נתונים הניתנים לניתוח באמצעות הכלים הסטטיסטיים שנלמדו (רצוי מתחומי ניהול שונים: שיווק, אבטחת איכות, ייצור ועוד).
5. סיורים מקצועיים בלשכה המרכזית לסטטיסטיקה.

נושאי הלימוד ושעות לימוד מומלצות

נושאי לימוד	שעות לימוד מומלצות
1. מבוא להסקה סטטיסטית	6
2. שיטות הסקה אמד למשתנה (מסטטיסטי לפרמטר)	8
3. הסקה לגבי ממוצע האוכלוסייה μ כאשר שונות האוכלוסייה σ^2 ידועה	10
4. הסקה לגבי ממוצע האוכלוסייה μ כאשר שונות האוכלוסייה σ^2 אינה ידועה	8
5. הסקה לגבי הפרופורציה p באוכלוסייה	8
6. מבחני χ^2 לטיב התאמה ולאי תלות	12
7. ניתוח שונות חד- כיווני ודו - כיווני	20
סה"כ	**72**

פירוט התכנים וחלוקת השעות המוצעת

נושאי לימוד		שעות
פרק 1	**מבוא להסקה סטטיסטית**	6
1.1	מהי הסקה סטטיסטית	
1.2	מושגי יסוד בהסקה סטטיסטית	
1.3	שיטות דגימה וקריטריונים לשימוש (מדגם מקרי פשוט, מדגם שטחות, מדגם אשכולות, מדגם שיטתי)	
1.4	יתרונות ומגבלות בשימוש בשיטות הדגימה השונות	
פרק 2	**שיטות הסקה מאומד למשתנה (מסטטיסטי לפרמטר)**	8
2.1	אמידה	
	2.1.1 אמידה נקודתית	
	2.1.2 אמידה על ידי רווח/תחום	
2.2	בדיקת השערות - מושגי יסוד	
	2.2.1 סוגי השערות	
	2.2.2 סוגי טעויות סטטיסטיות	
	2.2.3 מובהקות	
	2.2.4 עצמת מבחן	
פרק 3	**הסקה לגבי ממוצע האוכלוסייה μ כאשר שונות האוכלוסייה σ^2 ידועה**	
3.1	רווח בר-סמך לממוצע אוכלוסייה μ	
	3.1.1 תכונות רווח הסמך-השפעות שינויים על רווח הסמך כגון שינוי גודל המדגם , רמת הביטחון ועוד.	
	3.1.2 אורך רווח הסמך הגורמים המשפיעים עליו כולל תרגילים עם הנעלם N	
	3.1.3 הסטייה של רווח הסמך-הגורמים המשפיעים על הסטייה כולל תרגילים עם הנעלם N	

נושאי לימוד		שעות
3.2	בדיקת השערות לממוצע האוכלוסייה μ תוך שימוש בהתפלגות נורמאלית ותכונותיה	
	3.2.1 מבחן חד-כיווני	
	3.2.2 מבחן דו-כיווני	
	3.2.3 הקשר בין רווח סֶמך ובדיקת השערות והסקת מסקנות בבדיקת השערות על סמך קשר זה	
	3.2.4 חישוב הסתברויות הטעויות הסטטיסטיות	
	3.2.5 קביעת גודל המדגם לפי הסתברויות הטעויות הסטטיסטיות	
	3.2.6 מציאת α מינימאלית לדחיית השערת $H\ 1$ (α')	
פרק 4	**הסקה לגבי ממוצע האוכלוסייה μ כאשר שונות האוכלוסייה σ^2 אינה ידועה**	8
4.1	התפלגות t ותכונותיה	
4.2	רווח בר-סמך ובדיקת השערות לממוצע האוכלוסייה μ	
פרק 5	**הסקה לגבי הפרופורציה p באוכלוסייה**	8
5.1	רווח בר-סמך לפרופורציה p	
5.2	בדיקת השערות לפרופורציה p	
	5.2.1 מבחן חד-כיווני ימני	
	5.2.2 מבחן חד-כיווני שמאלי	
	5.2.3 מבחן דו-כיווני	
	5.2.4 חישוב הסתברויות הטעויות הסטטיסטיות	
	5.2.5 קביעת גודל המדגם לפי הסתברויות הטעויות הסטטיסטיות	
	5.2.6 מציאת רמת α מינימאלית לדחיית השערת $H1$ (α')	
פרק 6	**מבחני χ^2 לטיב התאמה ולאי-תלות**	12
6.1	מבחן χ^2 לטיב התאמה	
6.2	מבחן χ^2 לאי-תלות	
6.3	שימוש בטבלת התפלגות χ^2	
6.4	מדד קשר בין שני משתנים (מדד קרמר)	

שעות	נושאי לימוד	
20	**ניתוח שונות חד- כיווני ודו - כיווני**	**פרק 7**
	מושגי יסוד: שונות בין קבוצות, שונות בתוך קבוצות, סך כל השונות	7.1
	סוגי מודלים במודל של ניתוח שונות חד-כיווני	7.2
	7.2.1 המודל הקבוע (Fixed Model)	
	7.2.2 המודל הרנדומלי (Random Model)	
	הסקת מסקנות	7.3
	ניתוח שונות דו-כיווני	7.4
	7.4.1 הגדרה ומטרות שימוש	
	ניתוח פלט של ניתוח שונות חד-כיווני ודו-כיווני	7.5
	7.5.1 הדגמת ממצאי ניתוח שונות חד כיווני ודו כיווני באמצעות פלטי מחשב והסקת מסקנות מתוך הפלט	
72	**סה"כ**	

דף נוסחאות בסטטיסטיקה א ו-ב

לתלמידי המגמה לתעשייה וניהול כיתות י"ג ו-י"ד

סטטיסטיקה תיאורית

א. שכיחות יחסית:

$$P = \frac{f}{n} = \frac{\text{שכיחות}}{\text{גודל\_המדגם}}$$

ב. מדדי המרכז

I. מדדי המרכז בסדרת נתונים:

(n = מספר הנתונים בסדרה)

($\bar{x}$) הממוצע:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\text{סכום\_המספרים}}{\text{מספר\_המספרים}}$$

(Me) החציון:

אם n אי-זוגי:

$$Me = x_{(n+1)/2}$$

אם n זוגי:

$$Me = \frac{x_{n/2} + x_{(n/2)+1}}{2}$$

II. מדדי המרכז בטבלת שכיחויות:

(k = מספר הקבוצות בטבלה)

($\bar{x}$) הממוצע:

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_k x_k}{n}$$

(Me) החציון:

$$Me = L + \frac{l}{f}\left(\frac{n}{2} - F\right)$$

כאשר:

L הוא הגבול התחתון של הקבוצה החציונית

l הוא רוחב הקבוצה החציונית

f מייצג את השכיחות של הקבוצה החציונית

F הוא השכיחות המצטברת עד (ולא כולל) את הקבוצה החציונית

($\tilde{x}$ או Mo) השכיח:

$$\tilde{x} = L + \frac{f_i' - f_{i+1}'}{(f_i' - f_{i-1}') + (f_i' - f_{i+1}')} \cdot l$$

כאשר:

L הוא הגבול התחתון של קבוצת השכיח

l הוא הרוחב של קבוצת השכיח

f_i' הוא הצפיפות של קבוצת השכיח

f_{i-1}' הוא הצפיפות של הקבוצה שלפני קבוצת השכיח

f_{i+1}' הצפיפות של הקבוצה שאחרי קבוצת השכיח

ג. מדדי הפיזור

I. מדדי הפיזור בסדרת נתונים:

(n = מספר הנתונים בסדרה)

השונות:

$$V = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2$$

סטיית התקן: $S = \sigma_n = \sqrt{V}$

התחום הבין רבעוני: $Q_3 - Q_1$

כאשר Q_1 הוא הרבעון הראשון ו- Q_3 הרבעון השלישי של הסדרה.

II. מדדי הפיזור בטבלת שכיחויות:

(k = מספר הקבוצות בטבלת השכיחויות)

$$V = \frac{f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \ldots + f_k(x_k - \bar{x})^2}{n} = \frac{\sum_{i=1}^{k} f_i x_i^2}{n} - \bar{x}^2$$

סטיית התקן: $S = \sigma_n = \sqrt{V}$

התחום הבין רבעוני: $Q_3 - Q_1$

כאשר: $Q_1 = L + \frac{l}{f}(\frac{n}{4} - F)$

בנוסחה זו:

n הוא מספר התצפיות

L הוא הגבול התחתון של הקבוצה הכוללת את Q_1

l רוחב הקבוצה שכוללת את Q_1

f שכיחות הקבוצה שכוללת את Q_1

F השכיחות המצטברת עד (ולא כולל) את הקבוצה שכוללת את Q_1

$$Q_3 = L + \frac{l}{f}(\frac{3n}{4} - F)$$

בנוסחה זו:

n הוא מספר התצפיות

L הוא הגבול התחתון של הקבוצה הכוללת את Q_3

l רוחב הקבוצה שכוללת את Q_3

f שכיחות הקבוצה שכוללת את Q_3

F השכיחות המצטברת עד (ולא כולל) את הקבוצה שכוללת את Q_3

רגרסיה ליניארית

א. r_{xy} = מקדם המתאם הליניארי בין המשתנים x ו-y

מקדם זה מקבל ערכים בין $+1$ ו- -1 בלבד. לכן: $-1 \le r_{xy} \le +1$

הנוסחה לחישוב מקדם המתאם הליניארי:

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$

כאשר:

$$S_y^2 = \frac{\sum_{i=1}^{n} y_i^2}{n} - \bar{y}^2 \qquad S_x^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2$$

$$S_{xy} = \mathrm{cov}(x,y) = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{x} \cdot \bar{y}$$

ב. קו הרגרסיה של y על-סמך x

משוואת הקו: $\hat{y} = a_{y/x} x + b_{y/x}$

הנוסחאות לחישוב המקדמים:

$$b_{y/x} = \bar{y} - a_{y/x}\bar{x} \qquad a_{y/x} = \frac{S_{xy}}{S_x^2}$$

ג. קו הרגרסיה של x על-סמך y

משוואת הקו: $\hat{x} = a_{x/y} y + b_{x/y}$

הנוסחאות לחישוב המקדמים: $b_{x/y} = \bar{x} - a_{x/y}\bar{y}$ $\qquad$ $a_{x/y} = \frac{S_{xy}}{S_y^2}$

יסודות ההסתברות

א. ההסתברות של המאורע A:

$$P(A)=\frac{\#(A)}{\#(\Omega)}=\frac{\text{מספר\_נקודות\_המדגם\_של\_המאורע\_}A}{\text{מספר\_נקודות\_המדגם\_של\_מרחב\_המדגם\_}\Omega}$$

ב. תכונות יסוד של ההסתברות של מאורע:

$0 \le P(A) \le 1$

$P(\Omega)=1$

$P(\varnothing)=0$

ג. משפטי יסוד של תורת ההסתברות

I. הסתברות המאורע המשלים: $P(\overline{A})=1-P(A)$

II. חוק הכפל של מאורעות בלתי תלויים:

אם A ו-B הם מאורעות בלתי תלויים אז: $P(A\cap B)=P(A)\cdot P(B)$

III. הסתברות האיחוד של שני מאורעות: $P(A\cup B)=P(A)+P(B)-P(A\cap B)$

ד. הסתברות מותנה

אם A ו-B הם מאורעות (לא ריקים) באותו מרחב מדגם, וידוע ש-A מתרחש, אז ההסתברות
ש- B יתרחש היא:

$$P(B/A)=\frac{P(A\cap B)}{P(A)}$$

ה. הסתברות החיתוך של שני מאורעות

אם A ו-B הם מאורעות לא ריקים באותו מרחב מדגם (לאו דווקא בלתי תלויים) אז:

$$P(A\cap B)=P(A)\cdot P(B/A)=P(B)\cdot P(A/B)$$

ו. נוסחת ההסתברות השלמה

אם המאורעות $A_1,A_2,\ldots,A_n$ מהווים חלוקה של מרחב המדגם Ω

(כלומר: $A_1\cup A_2\cup\ldots\cup A_n=\Omega$ ו-$A_i\cap A_j=\varnothing$ לכל $i\neq j$) ו- B הוא מאורע נוסף

באותו מרחב מדגם (אך שונה מ- $A_1,A_2,\ldots,A_n$) אז: $P(B)=\sum_{i=1}^{n}P(A_i)P(B/A_i)$

ז. נוסחת בייס

אם המאורעות $A_1,A_2,\ldots,A_n$ מהווים חלוקה של מרחב המדגם (כפי שהוסבר בסעיף שקדם לזה) ו- B הוא מאורע נוסף באותו מרחב מדגם (אך שונה מ- $A_1,A_2,\ldots,A_n$) אז:

$$P(A_k / B) = \frac{P(A_k)P(B / A_k)}{\sum_{i=1}^{n} P(A_i)P(B / A_i)} \qquad (k = 1,2,\ldots,n)$$

התפלגויות

א. התפלגות בינומית

משתנה מקרי בדיד X הוא בעל התפלגות בינומית אם:

$$P(X = k) = \binom{n}{k} p^k q^{n-k} \qquad (k = 0,1,2,\ldots,n)$$

כאשר:

n = גודל המדגם

k = מספר "ההצלחות" במדגם

p = ההסתברות של "הצלחה"

$q = 1 - p$ = ההסתברות של "כישלון"

את העובדה שלמשתנה בדיד X יש התפלגות בינומית עם פרמטרים n ו-p מקובל לסמן כך:

$X \sim B(n, p)$

נוסחת התוחלת של משתנה מקרי בינומי X : $\mu = np$

נוסחת השונות של X : $\sigma^2 = npq$

ב. התפלגות פואסונית

משתנה מקרי בדיד X הוא בעל התפלגות פואסונית אם:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad (k = 0,1,2,\ldots)$$

כאשר:

λ = התוחלת של X

$e \cong 2.71828\ldots$ = בסיס הלוגריתמים הטבעיים

התוחלת של X: $\mu = \lambda$

השונות של X : $\sigma^2 = \lambda$

ג. התפלגות נורמלית

אם X הוא משתנה מקרי רציף בעל התפלגות נורמלית, תוחלת μ וסטיית תקן σ

יסומן הדבר כך: $X \sim N(\mu,\sigma)$

משתנה מקרי נורמלי בעל תוחלת $\mu = 0$ וסטיית תקן $\sigma = 1$ מכונה משתנה נורמלי וסטנדרטי ויסומן ב- Z.

בהתאם לכך: $Z \sim N(0,1)$

כל משתנה מקרי נורמלי X הופך לסטנדרטי על-ידי הטרנספורמציה: $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$

הנוסחאות לחישוב ההסתברויות של משתנה נורמלי:

$$P(X < a) = \Phi(\frac{a-\mu}{\sigma})$$

$$P(X > a) = 1 - \Phi(\frac{a-\mu}{\sigma})$$

$$P(a < X < b) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$$

כאשר Φ היא פונקציית ההסתברות המצטברת של Z.

ד. משפט הגבול המרכזי

אם $X_1, X_2, \ldots, X_n$ הם משתנים מקריים בלתי תלויים ובעלי התפלגות זהה (לכל אחד תוחלת μ וסטיית תקן σ) אז עבור n גדול דיו $(n \geq 30)$:

$$\sum_{i=1}^{n} X_i = S_n \sim N(n\mu, \sigma\sqrt{n})$$

$$\frac{X_1 + X_2 + \ldots + X_n}{n} = \bar{x} = N(\mu, \frac{\sigma}{\sqrt{n}})$$

ה. קירוב נורמלי להתפלגות בינומית

אם $X \sim B(n,p)$, עבור n גדול דיו $(n \geq 30)$ מתקיים: $X \sim N(np, \sqrt{npq})$

הסקה סטטיסטית

א. רווח סמך לתוחלת של אוכלוסייה כאשר שונות האוכלוסייה ידועה

אם האוכלוסייה מפולגת התפלגות נורמלית או גודל המדגם הוא לפחות:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

כאשר:

μ = תוחלת האוכלוסייה

n = גודל המדגם

$\bar{x}$ = ממוצע המדגם

σ = סטיית התקן של האוכלוסייה

α = גודל השגיאה

$z_{\alpha/2}$ = ערכו של המשתנה נורמלי וסטנדרטי כך שהשטח מימינו הוא $\alpha/2$

ב. קביעת גודל המדגם n המבטיח ששגיאת האמידה של μ לא תעלה על מספר נתון ε:

$$n = \left(\frac{z_{\alpha/2}\sigma}{\varepsilon}\right)^2$$

ג. רווח סמך לתוחלת האוכלוסייה כאשר שונות האוכלוסייה אינה ידועה

I. אם האוכלוסייה מפולגת התפלגות נורמלית או שגודל המדגם הוא לפחות 30:

$$\bar{x} - z_{\alpha/2}\frac{\hat{S}}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\hat{S}}{\sqrt{n}}$$

כאשר:

$z_{\alpha/2}, \alpha, \bar{x}, n, \mu$ כפי שהוגדרו לעיל

$\hat{S}$ = סטיית התקן של המדגם

II. אם גודל המדגם הוא פחות מ-30:

$$\bar{x} - t_{\alpha/2}\frac{\hat{S}}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{\hat{S}}{\sqrt{n}}$$

כאשר:

$\hat{S}, \alpha, \bar{x}, n, \mu$ כפי שהוסבר לעיל

$t_{\alpha/2}$ = ערכו של t (בטבלת ההתפלגות המתאימה) עם n-1 דרגות חופש, כך שהשטח מימינו
הוא $\alpha/2$

ד. רווח סמך לפרופורציה p

I. הגדרת הפרופורציה:

$$p=\frac{\text{מספר\_האיברים\_באוכלוסיה\_בעלי\_תכונה\_מסוימת}}{\text{מספר\_האיברים\_של\_האוכלוסיה\_כולה}}$$

II. רווח סמך לפרופורציה (בהנחה שגודל המדגם הוא n>30) :

$$\hat{p}-z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}<p<\hat{p}+z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

כאשר:

$z_{\alpha/2},n$ כפי שהוסבר לעיל

$\hat{p}$ = פרופורציית האיברים מתוך המדגם שהם בעלי תכונה מסוימת

$\hat{q}=1-\hat{p}$ = פרופורציית האיברים מתוך המדגם שאינם בעלי התכונה המסוימת

III. קביעת גודל המדגם n המבטיח ששגיאת האמידה של p לא תעלה על מספר נתון ε:

$$n=\frac{z_{\alpha/2}^2}{4\varepsilon^2}$$

בדיקת השערות

א. בדיקת השערות עבור תוחלת האוכלוסייה μ על-סמך מדגמים גדולים (n>30)

הערך הקריטי	אזור הדחייה	ההשערה
(כאשר שונות האוכלוסייה ידועה) $C=\mu_0+z_\alpha\frac{\sigma}{\sqrt{n}}$ (כאשר שונות האוכלוסייה אינה ידועה) $C=\mu_0+z_\alpha\frac{\hat{S}}{\sqrt{n}}$	$\bar{x}\geq C$	$H_0:\mu=\mu_0$ $H_1:\mu>\mu_0$
(כאשר שונות האוכלוסייה ידועה) $C=\mu_0-z_\alpha\frac{\sigma}{\sqrt{n}}$ (כאשר שונות האוכלוסייה אינה ידועה) $C=\mu_0-z_\alpha\frac{\hat{S}}{\sqrt{n}}$	$\bar{x}\leq C$	$H_0:\mu=\mu_0$ $H_1:\mu<\mu_0$

ההשערה	אזור הדחייה	הערך הקריטי
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$\bar{x} \le C_1$ וגם $\bar{x} \ge C_2$	$C_1 = \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ (כאשר שונות האוכלוסייה ידועה) $C_2 = \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ (כאשר שונות האוכלוסייה ידועה) כאשר שונות האוכלוסייה אינה ידועה, יש להחליף את σ ב- $\hat{S}$

ב. בדיקת השערות עבור תוחלת האוכלוסייה μ על סמך מדגמים קטנים מתוך אוכלוסייה נורמלית

ההשערה	אזור הדחייה	הערך הקריטי
$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$\bar{x} \ge C$	$C = \mu_0 + t_\alpha \frac{\hat{S}}{\sqrt{n}}$ $(\nu = n-1)$
$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$\bar{x} \le C$	$C = \mu_0 - t_\alpha \frac{\hat{S}}{\sqrt{n}}$ $(\nu = n-1)$
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$\bar{x} \le C_1$ וגם $\bar{x} \ge C_2$	$C_1 = \mu_0 - t_{\alpha/2} \frac{\hat{S}}{\sqrt{n}}$ $(\nu = n-1)$ $C_2 = \mu_0 + t_{\alpha/2} \frac{\hat{S}}{\sqrt{n}}$ $(\nu = n-1)$

ג. בדיקת השערות עבור פרופורציה p על-סמך מדגמים גדולים (n>30)

ההשערה	אזור הדחייה	הערך הקריטי
$H_0 : p = p_0$ $H_1 : p < p_0$	$z < -z_\alpha$	$z = \frac{x - np_0}{\sqrt{np_0 q_0}}$
$H_0 : p = p_0$ $H_1 : p > p_0$	$z > z_\alpha$	$z = \frac{x - np_0}{\sqrt{np_0 q_0}}$

הערך הקריטי	אזור הדחייה	ההשערה
$z = \frac{x - np_0}{\sqrt{np_0 q_0}}$	$z < -z_{\alpha/2}$ וגם $z > z_{\alpha/2}$	$H_0 : p = p_0$ $H_1 : p \neq p_0$

ד. בדיקת טיב ההתאמה

השערת האפס: התוצאות שהתקבלו בניסוי שייכות להתפלגות מסוימת.

ההשערה האלטרנטיבית: התוצאות שהתקבלו בניסוי אינן שייכות להתפלגות זו.

תוצאות	מספר 1	מספר 2	....	מספר k
שכיחות שנצפתה (observed frequency)	O_1	O_2	....	O_k
שכיחות צפויה (expected frequency)	e_1	e_2	....	e_k

מחשבים את הסכום:

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$ (מחושב)

לסכום זה יש התפלגות χ^2 עם $v = k-1$ דרגות חופש. בטבלת ההתפלגות של χ^2 מחפשים את הערך של $\chi^2_a\,(v = n-1)$. אם χ^2 (מחושב) גדול מערך זה, דוחים את השערת האפס.

הערה: α היא רמת המובהקות של המבחן.

ה. בדיקת אי-תלות של שני משתנים

H_0 : המשתנים X ו-Y בלתי תלויים

H_1 : המשתנים X ו-Y תלויים

המשתנה X מקבל את הערכים: $x_1, x_2, \cdots, x_k$

המשתנה Y מקבל את הערכים: $y_1, y_2, \cdots, y_m$

כדי לבצע את הבדיקה, יש צורך בשתי טבלאות:

I. טבלת הערכים שנצפו

סה"כ	y_m	……	y_2	y_1	Y / X
$o_{1.}$	o_{1m}	……	o_{12}	o_{11}	x_1
$o_{2.}$	o_{2m}	……	o_{22}	o_{21}	x_2
⋮	⋮	⋮	⋮	⋮	⋮
$o_{k.}$	o_{km}	……	o_{k2}	o_{k1}	x_k
n	$o_{.m}$	……	$o_{.2}$	$o_{.1}$	סה"כ

כאשר o_{ij} מייצג את השכיחות שנצפתה של המשתנה שנחקר כאשר X מקבל את הערך x_i

ו-Y מקבל את הערך y_j .

$o_{i.}$ מייצג את סכום השכיחויות בשורה i . $(i = 1,2,\ldots,k)$

$o_{.j}$ מייצג את סכום השכיחויות בעמודה j . $(j = 1,2,\ldots,m)$

n = סה"כ מספר התצפיות.

II. טבלת הערכים הצפויים

y_m	……	y_2	y_1	Y / X
e_{1m}	……	e_{12}	e_{11}	x_1
e_{2m}	……	e_{22}	e_{21}	x_2
⋮	⋮	⋮	⋮	⋮
e_{km}	……	e_{k2}	e_{k1}	x_k

את איברי הטבלה הזאת מחשבים כך: $e_{ij} = \frac{o_{i.} \cdot o_{.j}}{n}$

עבור כל הערכים האפשריים של i ו- j : $i = 1,2,\ldots,k \;;\; j = 1,2,\ldots,m$

בהמשך מחשבים את הסכום: (מחושב) $\chi^2 = \sum_{i,j} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$

עבור כל הערכים האפשריים של i ו-j . אם: (מחושב) $\chi^2 > \chi^2_{1-\alpha,(k-1)(m-1)}$

דוחים את השערת האפס.

ו. ניתוח שונות (חד-כיוונית)

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$

H_1 : לפחות שתיים מהתוחלות אינן שוות

הנחות: k האוכלוסיות מפולגות התפלגות נורמלית ובעלות שונות משותפת σ^2.

מקור (SOURCE)	דרגות חופש $D.F.$	סכום הריבועים $S.S.$	ממוצע ריבועים $M.S.$	F (מחושב)
טיפול (בין הקבוצות) (BETWEEN)	$k-1$	SSB	$MSB = \frac{SSB}{k-1}$	$F = \frac{MSB}{MSW}$
שגיאה (בתוך הקבוצות) (WITHIN)	$n-k$	SSW	$MSW = \frac{SSW}{n-k}$	
כולל (TOTAL)	$n-1$	SST		

השערת האפס נדחית אם: $F > F_{\alpha,k-1,n-k}$ (מחושב)

$k =$ מספר האוכלוסיות

$n =$ גודל המדגם מכל אוכלוסייה

$\alpha =$ רמת המובהקות

מושגים עיקריים

המושג	הסבר
אומד (estimator)	הסטטיסטי שאנו משתמשים בו כדי לאמוד פרמטר של אוכלוסייה.
אומדן (estimate)	הערך המספרי שהאומד מקבל.
אומד נקודתי (point estimate)	אומד שנותן אומדן יחיד (בודד) להערכת פרמטר של אוכלוסייה.
אומד בלתי מוטה (unbiased estimator)	אומד שתוחלתו שווה לפרמטר שהוא משמש לו כאומד.
אומד במרווח (interval estimate)	אומד שנותן אומדן של פרמטר של אוכלוסייה באמצעות מרווח.
אזור הדחייה (rejection region)	קבוצת הערכים של הסטטיסטי שעבורם דוחים את השערת האפס.
אזור הקבלה (acceptance region)	קבוצת הערכים של הסטטיסטי שעבורם מקבלים את השערת האפס.
בדיקת השערות (hypothesis testing)	שיטה סטטיסטית לבדיקת השערות בעלות אופי סטטיסטי, שבמסגרתה אנו בודקים את השערת האפס לעומת ההשערה האלטרנטיבית.
גבולות סֶמך (confidence limits)	הקצוות של רווח הסמך.
הסקה סטטיסטית (statistical inference)	שיטות סטטיסטיות שנועדו להשיג מסקנות לגבי אוכלוסייה סטטיסטית על סְמך מדגמים מאותה אוכלוסייה.
השערה (hypothesis)	הנחה שאנו מנסחים כאשר אנו חוקרים תופעה כלשהי ואין בידינו די מידע כדי לקבל תמונה שלמה של התופעה הנחקרת.
השערה סטטיסטית (statistical hypothesis)	השערה הקשורה לאוכלוסייה סטטיסטית.
התפלגות הדגימה (sampling distribution)	התפלגות של סטטיסטי שמשמש אומד נקודתי של פרמטר של אוכלוסייה.
טעות התקן (standard error)	סטיית התקן של התפלגות הדגימה של אומד נקודתי.

המושג	הסבר
טעות מסוג ראשון (type I error)	הטעות שעושים כשדוחים את השערת האפס כאשר היא נכונה.
טעות מסוג שני (type II error)	הטעות שעושים כשמקבלים את השערת האפס כאשר היא אינה נכונה.
מבחן דו-צדדי (two-sided test)	מבחן סטטיסטי שבו ההשערה הנבדקת היא מן הסוג: "שווה לעומת שונה".
מבחן חד-צדדי (one-sided test)	מבחן סטטיסטי שבו ההשערה הנבדקת היא מן הסוג: "שווה לעומת גדול" או "שווה לעומת קטן".
מבחן טיב ההתאמה (goodness of fit test)	מבחן סטטיסטי שמאפשר לבדוק אם תוצאות סטטיסטיות, שהתקבלו בדרך ניסוי, שייכות להתפלגות סטטיסטית מסוימת.
מבחן לבדיקת אי-תלות (test for independence)	מבחן סטטיסטי שמאפשר לבדוק אם שני משתנים מקריים תלויים זה בזה או בלתי תלויים.
ניתוח שונות (analysis of variance)	שיטה סטטיסטית לבדיקת שוויון התוחלת של k אוכלוסיות (על סמך מדגמים שווים בגודלם), לעומת ההשערה החלופית שלפחות שתיים מן התוחלות האלה שונות זו מזו. הבדיקה נעשית בהנחה ש-k האוכלוסיות מפולגות נורמלית ויש להן אותה שונות.
נקודה קריטית (critical value)	הערך של הסטטיסטי המפריד בין אזור הקבלה לאזור הדחייה של השערת האפס.
סטטיסטי (statistic)	כל ערך מספרי המתאר גודל אופייני של מדגם.
עוצמה של מבחן (power of test)	ההסתברות שנדחה את השערת האפס ונקבל את ההשערה החלופית כאשר ההשערה החלופית נכונה. הסתברות זו מסומנת ב- $1-\beta$.
פרמטר (parameter)	כל ערך מספרי שמתאר גודל אופייני של אוכלוסייה סטטיסטית.
רווח סמך (confidence interval)	מרווח על ציר המספרים המאפשר לנו לדעת ברמת ביטחון מסוימת, שהפרמטר הנאמד נמצא בתוכו.
רמת סמך (degree of confidence)	ההסתברות להימצאות הפרמטר הנאמד בתוך רווח הסמך (מסומן ב- $1-\alpha$).
רמת המובהקות של המבחן (level of significance)	ההסתברות לעשות טעות מהסוג הראשון (מקובל לסמן אותה ב- α).
שגיאת האמידה	ההפרש (בערך מוחלט) בין הערך האמתי של הפרמטר

המושג	הסבר
(error of estimation)	ובין האומדן של אותו פרמטר (שמתקבל בעזרת האומד המתאים).

ביבליוגרפיה

1. האוניברסיטה הפתוחה, **מבוא לסטטיסטיקה לתלמידי מדעי החברה** (כרך ב), 1993.
2. רונית איזנבד, **סטטיסטיקה ללא סטטיסטיקאים**, הוצאת אקדמון, 1999.
3. שולה ישראלית, **סטטיסטיקה הלכה למעשה**, הוצאת לוגיק, 1999.