Eugene Kanzieper © All rights reserved 2004/05
חומר לימוד בקורס "הסתברות וסטטיסטיקה" מאת יוג'ין קנציפר © כל הזכויות שמורות 2004/05

הרצאה 9: בעיות אמידה קשורות להתפלגות נורמלית

בהרצאה זו אנחנו נלמד איך אפשר לאמוד <u>מדד מסוים לא ידוע</u> המאפיין את המשתנה המקרי הנבדק. נקרא למדדים כאלה <u>כפרמטרים.</u> נסמן אותם ב- θ. למשל:

א. על סמך מדגם מקרי של 100 נשים יש לאמוד את הגיל הממוצע של נישואין. פה, ה<u>פרמטר</u> θ הוא גיל הנישואין הממוצע.
ב. על סמך מדגם מקרי של חולים אשר קיבלו טיפול מסוים יש לאמוד את ההסתברות להצלחת הטיפול. ההסתברות הנ"ל היא ה<u>פרמטר</u> θ.

9.1. הגדרות פורמליות

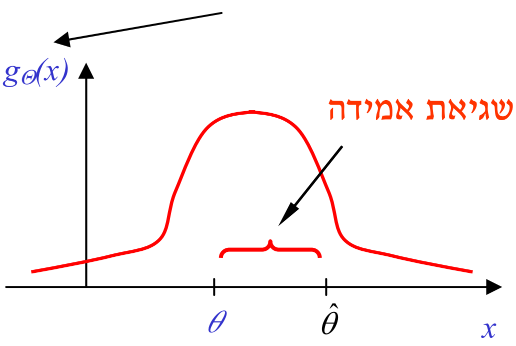
<u>הגדרה 1</u>: **סטטיסטי** הוא פונקציה של תצפיות המדגם $X_1, X_2, ..., X_n$. <u>סימון</u>: $\hat{\Theta}$.
<u>דוגמאות</u>: $X_1 \cdot X_2 \cdot X_3$, $X_1 + X_2 + X_3 + ... + X_n$ וכד' הם סטטיסטים.

<u>הגדרה 2</u>: **אומד נקודתי** הוא הסטטיסטי $\hat{\Theta}$ המשמש לשם אמידה של פרמטר θ.

<u>הגדרה 3</u>: **אומדן** הוא הערך הספציפי $\hat{\theta}$ שמקבל אומד נקודתי $\hat{\Theta}$ עבור מדגם מסוים.

<u>דוגמא</u>: נתונים התצפיות $X_1, X_2, ..., X_n$. אנחנו מעונינים באמידה של **הפרמטר** הנקרא **תוחלת** μ ($\theta = \mu$). מההרצאה הקודמת אפשר להניח כי **הסטטיסטי** $\bar{X} = \frac{1}{n}(X_1 + X_1 + ... + X_n)$ ($\hat{\Theta} = \bar{X}$) הוא **אומד נקודתי** לפרמטר שלנו – התוחלת. ברור ש<u>אומד נקודתי מהווה משתנה מקרי.</u> הערך שיקבל $\bar{X}$ עבור מדגם מסוים יהי' **אומדן** לתוחלת.

<u>הגדרה 4</u>: התפלגות של אומד נקודתי נקראת **התפלגות דגימה.**



התפלגות דגימה מסבירה למה תמיד קיימת **שגיאת אמידה.**

Terms of Use: Everyone is allowed to use these course materials for free provided the author's copyright notice is kept (Eugene Kanzieper © All rights reserved, 2004/05). No part of the lecture notes and other accompanying materials may be reproduced or stored in a retrieval system other than at **http://sciences.hait.ac.il/~ekanzieper**, copied or transmitted in any form or by any means without the prior permission of the author.

דוגמא: נתבונן במשתנה מקרי X = {תוצאה של הטלת קובייה תקינה}. ברור כי $X \sim U(1,N)$ כך שהתוחלת שלו שווה ל- $\mu = E[X] = \frac{N+1}{2} = \frac{6+1}{2} = 3.5$. נניח כי μ הוא **פרמטר** לא ידוע ואנחנו רוצים לאמוד אותו. נבחר בממוצע $\bar{X}$ כ**אומד נקודתי** ל- μ. נניח שבסדרה מסוימת של 5 הטלות קיבלנו את התוצאות (התצפיות) 3, 6, 1, 1, 4. האומד $\bar{X}$ מקבל הערך $\bar{X} = \frac{1}{5}(4+1+1+6+3) = 3$. זה אומר ש**אומדן** $\bar{X} = 3$. באותו רגע, הערך האמיתי של הפרמטר μ הוא $\mu = E[X] = 3.5$. ההפרש $3 - 3.5 = -0.5$ מהווה את **שגיאת האמידה** במקרה זה. ברור כי עבור סדרה אחרת של תצפיות אנחנו נקבל אומדן שונה ושגיאת אמידה שונה.

9.2. סוגים שונים של אומד נקודתי

בדרך כלל הערך האמיתי של הפרמטר θ לא ידוע. בגלל זה, רצוי כי לאומד נקודתי $\hat{\Theta}$ תהיה התפלגות דגימה כזאת ש- $E[\hat{\Theta}]$ תהיה שווה לערך האמיתי של הפרמטר θ.

<u>הגדרה 5</u>: אומד נקודתי $\hat{\Theta}$ ייקרא **אומד חסר הטיה** לפרמטר θ אם $E[\hat{\Theta}] = \theta$.

<u>הגדרה 6</u>: אומד נקודתי $\hat{\Theta}$ שעבורו $E[\hat{\Theta}] \neq \theta$ ייקרא **אומד מוטה**.

<u>שאלה 1</u>: משתנה מקרי X הוא תוצאה של הטלת קובייה תקינה. נא להוכיח כי $\hat{\Theta}_1 = \frac{1}{2}(X_1 + X_2)$ ו- $\hat{\Theta}_2 = \frac{1}{3}(2X_1 + X_2)$ הם אומדים חסרי הטיה עבור התוחלת $\mu = E[X]$, והאומד $\hat{\Theta}_3 = \frac{1}{3}(X_1 + X_2 + 3)$ הוא אומד מוטה.

<u>פתרון</u>: יש לבדוק כי עבור שני האומדים הראשונים מתקיים $E\left[\hat{\Theta}_1\right] = E\left[\hat{\Theta}_2\right] = \mu$ ועבור האומד השלישי $E\left[\hat{\Theta}_3\right] \neq \mu$.

9.3. האומדים $\bar{X}$, S^2 ו- $\hat{S}^2$

אם $X_1, X_2, ..., X_n$ הוא מדגם מקרי של n תצפיות בלתי תלויות מהתפלגות בעלת תוחלת μ ושונות σ^2, אזי:

א. ממוצע המדגם

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

הוא אומד חסר הטיה לתוחלת μ.

ב. אומד חסר הטיה עבור השונות σ^2 ניתן על ידי הנוסחא

Terms of Use: Everyone is allowed to use these course materials for free provided the author's copyright notice is kept (Eugene Kanzieper © All rights reserved, 2004/05). No part of the lecture notes and other accompanying materials may be reproduced or stored in a retrieval system other than at **http://sciences.hait.ac.il/~ekanzieper**, copied or transmitted in any form or by any means without the prior permission of the author.

$$\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

כאשר $n \geq 2$.

ג. אם ידועה התוחלת μ, גם האומד

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)^2$$

הוא אומד חסר הטיה עבור השונות σ^2.

<u>שאלה 2</u>: נא להוכיח את הטענה הנ"ל.
<u>פתרון</u>:

א. $E\left[\overline{X}\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right] = \frac{1}{n}\sum_{i=1}^{n}E\left[X_i\right] = \frac{1}{n}\cdot n\mu = \mu$.

ב.

$$E\left[\hat{S}^2\right] = E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}X_i^2 - 2\overline{X}\sum_{i=1}^{n}X_i + n\overline{X}^2\right]$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}E\left[X_i^2\right] - 2nE\left[\overline{X}^2\right] + nE\left[\overline{X}^2\right]\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n}E\left[X_i^2\right] - nE\left[\overline{X}^2\right]\right)$$

$$= \frac{1}{n-1}\left(n\sigma^2 + n\mu^2 - nE\left[\overline{X}^2\right]\right)$$

אבל $Var\left[\overline{X}\right] = E\left[\overline{X}^2\right] - \left(E\left[\overline{X}\right]\right)^2 = \frac{\sigma^2}{n}$ כך ש-

$$E\left[\hat{S}^2\right] = \frac{1}{n-1}\left(n\sigma^2 + n\mu^2 - nE\left[\overline{X}^2\right]\right) = \frac{1}{n-1}\left(n\sigma^2 + n\mu^2 - n\left(\mu^2 + \frac{\sigma^2}{n}\right)\right) = \sigma^2.$$

ג. פתרון דומה.

<u>שאלה 3</u>: אמוד את השונות בעזרת האומד $\hat{S}^2$ על סמך סדרת ההטלות

א. 3, 6, 1, 1, 4
ב. 4, 3, 6, 5, 5

ותשווה התוצאה עם שונות עצמה.

9.4. אמידה ורווח סמך ל- μ כאשר σ ידוע

הערך שיקבל הממוצע $\overline{X}$ במדגם מסוים יהי בדרך כלל שונה מן הערך האמיתי של הפרמטר μ (תוחלת). מה גודל של השגיאה זו?

יש להבדיל בין שתי שאלות שונות:

Terms of Use: Everyone is allowed to use these course materials for free provided the author's copyright notice is kept (Eugene Kanzieper © All rights reserved, 2004/05). No part of the lecture notes and other accompanying materials may be reproduced or stored in a retrieval system other than at **http://sciences.hait.ac.il/~ekanzieper**, copied or transmitted in any form or by any means without the prior permission of the author.

- מהי **מידת הדיוק** של אמידה?
- מה צריך להיות **גודלו של המדגם** על מנת שתובטח רמת דיוק רצויה?

9.4.1. מידת דיוק סטטיסטית ורווח סמך

משפט הגבול המרכזי טוען כי ממוצע $\overline{X}$ המחושב על סמך מדגם מקרי של תצפיות $X_1, X_2, ..., X_n$ עבור משתנה מקרי X בעל תוחלת μ ושונות σ^2 מתפלג נורמלית

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

עבור n מספיק גדול.

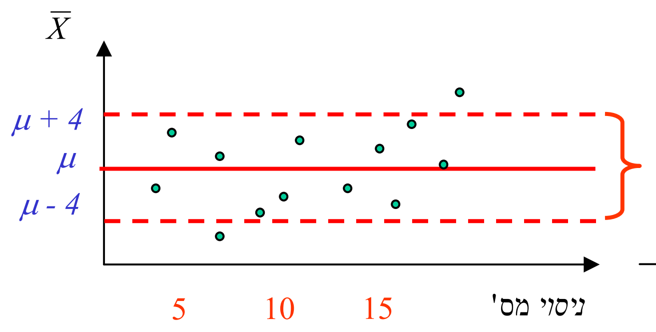
אנחנו נתייחס למדגם עם $n \geq 30$ כלמדגם גדול.

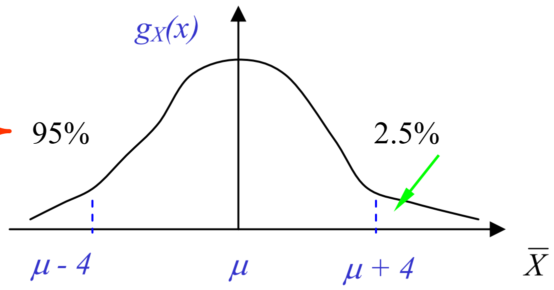
הערה: אם ידוע כי התצפיות מפולגות נורמלית – אין הגבלה עבור n.

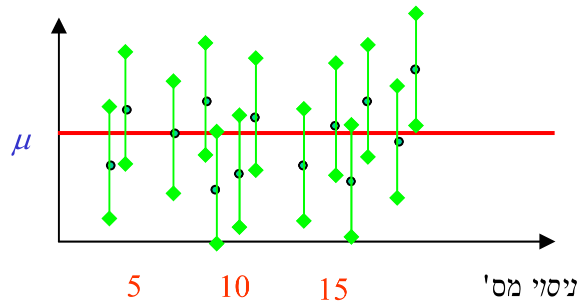
שאלה 4: במדגם מגודל $n = 25$ עבור משתנה מקרי X המפולג נורמלית עם תוחלת μ לא ידועה וסטיית התקן $\sigma = 10$, יש לחשב את ההסתברות לכך שממוצע המדגם $\overline{X}$ לא יסטה מהתוחלת μ ביותר מ-4 יחידות?

פתרון: הממוצע מפולג נורמלית עם התוחלת בלתי ידועה μ וסטיית התקן $\frac{\sigma}{\sqrt{n}} = {10}/{\sqrt{25}} = 2$. אזי $\overline{X} \sim N(\mu, 4)$. צריך לחשב ההסתברות $P(\mu - 4 < \overline{X} < \mu + 4)$. החישוב הפשוט מביא

$$P(\mu - 4 < \overline{X} < \mu + 4) = \Phi\left(\frac{\mu + 4 - \mu}{2}\right) - \Phi\left(\frac{\mu - 4 - \mu}{2}\right) = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 \approx 0.95$$







Terms of Use: Everyone is allowed to use these course materials for free provided the author's copyright notice is kept (Eugene Kanzieper © All rights reserved, 2004/05). No part of the lecture notes and other accompanying materials may be reproduced or stored in a retrieval system other than at **http://sciences.hait.ac.il/~ekanzieper**, copied or transmitted in any form or by any means without the prior permission of the author.

בהסתברות 0.95 יכיל הרווח המשתרע מ- $\bar{X}-4$ עד $\bar{X}+4$ את הערך האמיתי של התוחלת μ.

<u>ניסוח מילולי:</u> $\bar{X}-4<\mu<\bar{X}+4$ הוא **רווח סמך** עבור μ ברמה של 95%.

<u>הכללה</u>: באותה דרך אפשר לפתח נוסחא ל**רווח-סמך עבור** μ **ברמה של** $(1-\alpha)\cdot 100\%$ **אחוז** (כאשר σ ידוע). במקרה זה הרווח $\bar{X}-\varepsilon_\alpha<\mu<\bar{X}+\varepsilon_\alpha$ מתקבל מהמשוואה: $P\left(\left|\bar{X}-\mu\right|<\varepsilon_\alpha\right)=1-\alpha$. כיוון ש-

$$P\left(\left|\bar{X}-\mu\right|<\varepsilon_\alpha\right)=P\left(\mu-\varepsilon_\alpha<\bar{X}<\mu+\varepsilon_\alpha\right)=\Phi\left(\frac{\mu+\varepsilon_\alpha-\mu}{\sigma/\sqrt{n}}\right)-\Phi\left(\frac{\mu-\varepsilon_\alpha-\mu}{\sigma/\sqrt{n}}\right)$$

$$=\Phi\left(\frac{\varepsilon_\alpha\sqrt{n}}{\sigma}\right)-\Phi\left(-\frac{\varepsilon_\alpha\sqrt{n}}{\sigma}\right)=2\Phi\left(\frac{\varepsilon_\alpha\sqrt{n}}{\sigma}\right)-1=1-\alpha$$

מפה אפשר לנצל את ε_α: $\Phi\left(\frac{\varepsilon_\alpha\sqrt{n}}{\sigma}\right)=1-\frac{\alpha}{2}$. מקבלים:

$$\varepsilon_\alpha=\frac{z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}}$$

כאשר $z_{\frac{\alpha}{2}}$ מקיים:

$$\Phi\left(z_{\frac{\alpha}{2}}\right)=1-\frac{\alpha}{2}.$$

כך אנחנו מגיעים לנוסחא עבור **רווח-סמך ל-** μ **ברמה של** $(1-\alpha)\cdot 100\%$ **אחוז** (כאשר σ ידוע):

$$\bar{X}-z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}<\mu<\bar{X}+z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}.$$

<u>שאלה 5</u>: נתון שסטיית התקן של משתנה מקרי מסוים היא $\sigma=0.35$. במדגם שגודלו $n=49$ התקבל ממוצע המדגם $\bar{X}=3.8$. חשב רווח-סמך לתוחלת μ:

א. ברמה של 95%

ב. ברמה של 99%

<u>פתרון:</u>

א. רמת סמך של 95% מתאימה לערך $\alpha=0.05$. צריך למצוא $z_{\frac{\alpha}{2}}=z_{0.025}$ כזה שעבורו מתקיים $\Phi(z_{0.025})=1-0.025=0.975$. בטבלת ההתפלגות הנורמלית צריך למצוא את המספר $0.975-0.5=0.475$. עבור $z_{0.025}=1.96$. אזי $\bar{X}-z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}=3.8-1.96\cdot\frac{0.35}{\sqrt{49}}\approx 3.702$ ו- $\bar{X}+z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}=3.8+1.96\cdot\frac{0.35}{\sqrt{49}}\approx 3.898$. כלומר, $P(3.702<\mu<3.898)=0.95$ כך שרווח-סמך עבור μ ברמה של 95% הוא $3.702<\mu<3.898$.

Terms of Use: Everyone is allowed to use these course materials for free provided the author's copyright notice is kept (Eugene Kanzieper © All rights reserved, 2004/05). No part of the lecture notes and other accompanying materials may be reproduced or stored in a retrieval system other than at **http://sciences.hait.ac.il/~ekanzieper**, copied or transmitted in any form or by any means without the prior permission of the author.

ב. $z_{0.005} = 2.575$; $P(3.671 < \mu < 3.929) = 0.99$.

הערה: ככל שרמת הסמך גבוהה יותר, מתרחב רווח הסמך; ככל שהרווח צר יותר, רמת הסמך פוחתת.

9.4.2. מהו גודל המדגם עבור רמת הדיוק הרצויה?

אם ברצוננו להבטיח ששגיאת אמידה של μ על ידי $\overline{X}$ (כאשר σ ידוע) לא תעלה על ε ברמת סמך נתונה של $(1-\alpha)\cdot 100\%$, יש לקיים:

$$n \ge \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{\varepsilon}\right)^2 \quad \text{(שלם).}$$

הנוסחא נובעת מהסעיף הקודם.

שאלה 6: במכונה אוטומטית למשקאות קלים, כמות המשקה הנמזגת לכוס (בגרמים) מפולגת בקירוב לפי התפלגות נורמלית בעלת סטיית תקן $\sigma = 20$.

א. מהו רווח-סמך ברמה של 95% לתוחלת כמות הנוזל הנמזג, אם במדגם של 16 כוסות התקבל הממוצע של 280 גרם לכוס?

ב. איזה גודל מדגם יבטיח שברמת סמך זו לא יסטה ממוצע המדגם מן התוחלת ביותר מ-2 גרם?

פתרון:

א. $\sigma = 20$, $n = 16$, $\alpha = 1 - 0.95 = 0.05$; אזי $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$ ו-

$$\overline{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 280 - 1.96 \cdot \frac{20}{\sqrt{16}} \approx 270.2 \text{ ו-} \overline{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 280 + 1.96 \cdot \frac{20}{\sqrt{16}} \approx 289.8$$ כך

ש- $270.2 < \mu < 289.8$.

ב. $n \ge \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{\varepsilon}\right)^2 = \left(\frac{1.96 \cdot 20}{2}\right)^2 \approx 384.16$; משום ש- n חייב להיות מספר שלם, אנו מקבלים את התשובה: $n \ge 385$.

9.5. אמידה ורווח סמך ל- μ כאשר σ לא ידוע

בסעיף הקודם דנו באמידת μ על ידי ממוצע המדגם $\overline{X}$ כאשר

א. סטיית התקן σ של משתנה מקרי ידועה

ב. המדגם גדול (מעל 30) או שהמשתנה הנבדק מתפלג נורמלית

ברוב הבעיות σ אינו ידוע! איך להתמודד עם הבעיה?

Terms of Use: Everyone is allowed to use these course materials for free provided the author's copyright notice is kept (Eugene Kanzieper © All rights reserved, 2004/05). No part of the lecture notes and other accompanying materials may be reproduced or stored in a retrieval system other than at **http://sciences.hait.ac.il/~ekanzieper**, copied or transmitted in any form or by any means without the prior permission of the author.

מתברר כי עבור מדגמים גדולים ($n \geq 30$) מותר להחליף את σ בנוסחאות הקודמות באומדן $\hat{S} = \sqrt{\hat{S}^2}$:

$$\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

אזי, רווח-סמך ברמה של $(1-\alpha)\cdot 100\%$ אחוז עבור μ שאינו מסתמך על σ הוא:

$$\overline{X} - z_{\frac{\alpha}{2}} \cdot \frac{\hat{S}}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}} \cdot \frac{\hat{S}}{\sqrt{n}}$$

כאשר $z_{\frac{\alpha}{2}}$ עדיין מקיים:

$$\Phi\left(z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}.$$

<u>שאלה 7</u>: במדגם מקרי של 36 נבחנים בבחינת בגרות מסוימת, נמצאו ממוצע $\overline{X} = 6.5$, ואומד לסטיית התקן $\hat{S} = 0.75$. מצא רווח סמך של 95% ושל 99% לממוצע הארצי בבחינת בגרות זו.

<u>פתרון</u>:

א. $n = 36$, $\alpha = 1 - 0.95 = 0.05$; אזי $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$ ו-

$\overline{X} - z_{\frac{\alpha}{2}} \cdot \frac{\hat{S}}{\sqrt{n}} = 6.5 - 1.96 \cdot \frac{0.75}{\sqrt{36}} \approx 6.225$ ו- $\overline{X} + z_{\frac{\alpha}{2}} \cdot \frac{\hat{S}}{\sqrt{n}} = 6.5 + 1.96 \cdot \frac{0.75}{\sqrt{36}} \approx 6.745$ כך

ש- $6.225 < \mu < 6.745$.

ב. $n = 36$, $\alpha = 1 - 0.99 = 0.01$; אזי $z_{\frac{\alpha}{2}} = z_{0.005} = 2.575$ ו-

$\overline{X} - z_{\frac{\alpha}{2}} \cdot \frac{\hat{S}}{\sqrt{n}} = 6.5 - 2.575 \cdot \frac{0.75}{\sqrt{36}} \approx 6.178$ ו- $\overline{X} + z_{\frac{\alpha}{2}} \cdot \frac{\hat{S}}{\sqrt{n}} = 6.5 + 2.575 \cdot \frac{0.75}{\sqrt{36}} \approx 6.822$

כך ש- $6.178 < \mu < 6.822$.

<u>הערה</u>: אם ברצוננו להבטיח ששגיאת אמידה של μ על ידי $\overline{X}$ (כאשר σ לא ידוע) לא תעלה על ε ברמת סמך נתונה של $(1-\alpha)\cdot 100\%$, יש לקיים:

$$n \geq \left(\frac{z_{\frac{\alpha}{2}} \cdot \hat{S}}{\varepsilon}\right)^2 \quad \text{(שלם).}$$

Terms of Use: Everyone is allowed to use these course materials for free provided the author's copyright notice is kept (Eugene Kanzieper © All rights reserved, 2004/05). No part of the lecture notes and other accompanying materials may be reproduced or stored in a retrieval system other than at **http://sciences.hait.ac.il/~ekanzieper**, copied or transmitted in any form or by any means without the prior permission of the author.

9.6. אמידת פרופורציה

דוגמאות של בעיית אמידת פרופורציה: מהו אחוז המעשנים בישראל? מהו אחוז התומכים במועמד מסוים בקרב כלל הבוחרים? וכד'

<u>הגדרה</u>: **פרופורציה** באוכלוסיה מסומנת ב- p ונתונה על ידי הנוסחא:

$$p=\frac{\text{מספר האיברים באוכלוסיה שהם בעלי תכונה מסוימת}}{\text{מספר כל האיברים באוכלוסיה}}$$

איך מחשבים p ? בודקים איבר אחרי איבר וסופרים רק אלה שהם בעלי תחונה מסוימת. אם מצאנו איבר כזה – זו <u>"הצלחה"</u>. אם לא – <u>"כישלון"</u>.

במקרה של אוכלוסיה שמכילה n איברים, מדובר על n <u>ניסוי ברנולי</u>. אזי מספר Y האיברים בעלי תכונה מבוקשת הוא <u>מספר כולל של הצלחות</u>. לפיכך הוא <u>מתפלג בינומית</u> עם פרמטרים n ו- p :

$$Y \sim B(n,p).$$

כדי לאמוד פרופורציה, יש לבנות אומד נקודתי לפרופורציה p . האומד הטבעי הוא פרופורציה במדגם:

$$\hat{p}=\frac{Y}{n}.$$

<u>שאלה 8</u>: נא להוכיח כי $\hat{p}$ הוא **אומד חסר הטיה** ל- p .

<u>פתרון</u>: אנו למדנו כי למשתנה מקרי בינומי $Y \sim B(n,p)$ מתקיים $E[Y]=np$. אזי אנו מחשבים: $E[\hat{p}]=E\left[\frac{Y}{n}\right]=\frac{1}{n}E[Y]=\frac{np}{n}=p$. זה מוכיח כי $\hat{p}$ הוא אומד חסר הטיה.

איך לחשב רווח-סמך עבור p ?

כאשר n גדול מספיק, Y מתפלג נורמלית בקירוב (משפט דה מואבר-לפלס):

$$\frac{Y-np}{\sqrt{p(1-p)}} \sim N(0,1).$$

אבל $Y=n\cdot\hat{p}$ כך ש-

$$\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \sim N(0,1).$$

מזה נובע כי רווח סמך ברמה של $(1-\alpha)\cdot 100\%$ ניתן על ידי הנוסחא

$$P\left(\hat{p}-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{p(1-p)}{n}}<p<\hat{p}+z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{p(1-p)}{n}}\right)=1-\alpha.$$

Terms of Use: Everyone is allowed to use these course materials for free provided the author's copyright notice is kept (Eugene Kanzieper © All rights reserved, 2004/05). No part of the lecture notes and other accompanying materials may be reproduced or stored in a retrieval system other than at **http://sciences.hait.ac.il/~ekanzieper**, copied or transmitted in any form or by any means without the prior permission of the author.

הנוסחה הזו אינה מתאימה למטרה שלנו כי ערך אמיתי של הפרמטר p אינו ידוע. כאשר n גדול, מותר להחליף p באומד נקודתי $\hat{p}$ כך ש-

$$P\left(\hat{p}-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}<p<\hat{p}+z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)=1-\alpha.$$

מפה נובעת הנוסחה הדרושה: **רווח סמך** עבור הפרופורציה p ברמה של $(1-\alpha)\cdot 100\%$ שווה ל-

$$\hat{p}-z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}<p<\hat{p}+z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

שאלה 9: 600 מתוך 800 מרואיינים במדגם מקרי של ישראלים סבורים שאין סיכוי לרסן את האינפלציה בעתיד הנראה לעין. חשב רווח סמך ברמה של 96% לפרופורציית האוכלוסיה הסבורה כך.

פתרון: $1-\alpha=0.96 \Leftarrow \alpha=0.04$; מחפשים בטבלה של התפלגות נורמלית את המספר $1-\frac{\alpha}{2}-0.5=0.48$. מקבלים $z_{\frac{\alpha}{2}}=z_{0.02}\approx 2.05$. גם מחשבים $\hat{p}=\frac{600}{800}=0.75$. לכן, רווח סמך ברמה של 96% עבור p הוא

$$0.75-2.05\cdot\sqrt{\frac{0.75(1-0.75)}{800}}<p<0.75+2.05\cdot\sqrt{\frac{0.75(1-0.75)}{800}}.$$

לאחר החישוב מקבלים: $0.72<p<0.78$.

שאלה 10: 232 מתוך 1000 מכוניות שנבדקו במדגם מקרי נמצאו בלתי תקינות. מצא רווח סמך ברמה של 96% לפרופורציית המכוניות בלתי תקינות המופעלות בישראל.

פתרון: $\hat{p}=\frac{232}{1000}=0.232$. מחשבים:

$$0.232-2.05\cdot\sqrt{\frac{0.232(1-0.232)}{1000}}<p<0.232+2.05\cdot\sqrt{\frac{0.232(1-0.232)}{1000}}.$$

לאחר החישוב מקבלים: $0.205<p<0.259$.

מה צריך להיות גודלו של המדגם על מנת שתובטח רמת דיוק רצויה?

כדי להבטיח שבהסתברות $1-\alpha$ לא תעלה השגיאה $|\hat{p}-p|$ על קריטריון-קרבה נתון מראש ε, יש לבחור את n כך שיתקיים:

Terms of Use: Everyone is allowed to use these course materials for free provided the author's copyright notice is kept (Eugene Kanzieper © All rights reserved, 2004/05). No part of the lecture notes and other accompanying materials may be reproduced or stored in a retrieval system other than at **http://sciences.hait.ac.il/~ekanzieper**, copied or transmitted in any form or by any means without the prior permission of the author.

$$n \ge \frac{\left(z_{\frac{\alpha}{2}}\right)^2 \hat{p}(1-\hat{p})}{\varepsilon^2}$$ (שלם).

אם האומדן $\hat{p}$ לא ידוע, אפשר לאמוד n על פי הנוסחא

$$n \ge \left(\frac{z_{\frac{\alpha}{2}}}{2\varepsilon}\right)^2$$ (שלם).

<u>שאלה 11</u>: נא להוכיח את שתי הנוסחאות.

<u>שאלה 12</u>: מועמד לראשות עירייה מסוימת מעוניין לאמוד את פרופורציית התומכים בו כך שיוכל להיות בטוח בהסתברות של 98% ששגיאת האמידה לא תעלה על 0.03. איזה גודל מדגם יבטיח לו הסתברות כזאת?

<u>פתרון</u>: $1-\alpha = 0.98 \Leftarrow \alpha = 0.02$. לכן $z_{\frac{\alpha}{2}} = z_{0.01} = 2.33$. גם נתון כי $\varepsilon = 0.03$. אזי

$$n \ge \left(\frac{z_{\frac{\alpha}{2}}}{2\varepsilon}\right)^2 = \left(\frac{2.33}{2 \cdot 0.03}\right)^2 \approx 1508.03$$ התשובה: $n = 1509$.

<u>שאלה 13</u>: תחזור על אותה השאלה בתנאי שידוע האומדן $\hat{p} = 0.6$.

<u>פתרון</u>: $$n \ge \frac{\left(z_{\frac{\alpha}{2}}\right)^2 \hat{p}(1-\hat{p})}{\varepsilon^2} = \frac{(2.33)^2\, 0.6(1-0.6)}{(0.03)^2} \approx 1447.7$$ התשובה: $n = 1448$.

Terms of Use: Everyone is allowed to use these course materials for free provided the author's copyright notice is kept (Eugene Kanzieper © All rights reserved, 2004/05). No part of the lecture notes and other accompanying materials may be reproduced or stored in a retrieval system other than at **http://sciences.hait.ac.il/~ekanzieper**, copied or transmitted in any form or by any means without the prior permission of the author.