ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

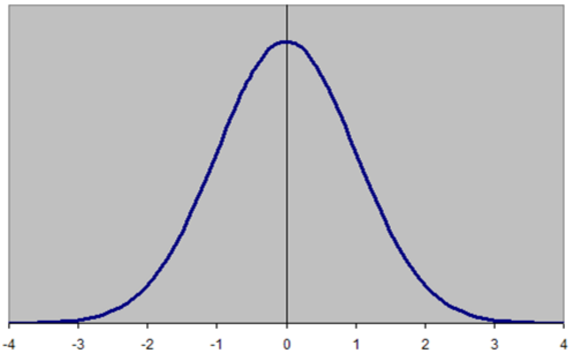
הרצאה 2-31.7.12

<u>התפלגות נורמאלית:</u>

צורת ההתפלגות בעלת חשיבות רבה בתיאוריה הסטט' של הסקה ממדגם לאוכלוסיה, נקראת גם התפלגות פעמון/גאוס. התפלגות נורמאלית היא התפלגות סימטרית, חד שיאית, דמוית פעמון, למשתנה רציף בגבולות $-\infty < x < \infty$. המשתנה הנורמאלי מאופיין ע"י שני פרמטרים:

1) תוחלת (ממוצע) של המשתנה המסומן בµ.
2) השונות-פיזור המסומן ע"י σ^2.

<u>סימון ההתפלגות</u>: $x \sim N(\mu, \sigma^2)$.



השטח מתחת לעקומה שווה 100%. הקו המרכזי בתמונה הוא הממוצע, שכיח והחציון. זהו מרכז ההתפלגות.

<u>ציון תקן</u>: $Z_x = \frac{X-\mu}{\sigma}$, כאשר Z_x מתפלג נורמאלית סטנדרטית, ז"א $Z_x \sim N(0,1)$, אפשר להיעזר בטבלת ההתפלגות הנורמאלית . השטח שמצוין בטבלה הוא תמיד השטח משמאל. (את הטבלה ניתן למצוא בנספחים)

<u>כללים שימושיים</u>:

- $p(Z > Z_1) = 1 - \Phi(Z_1)$
- $p(-Z_1) = 1 - \Phi(Z_1)$

דוגמה:

גובה של חיילים מתפלג נורמאלית עם ממוצע 175 ס"מ. וסטיית תקן של 10 ס"מ. נחבר חייל באופן מקרי.

א) מה ההסתברות שגובהו מעל 200 ס"מ?
ב) מה ההסתברות שגובהו מתחת ל165 ס"מ?
ג) מה גובה החיל ש90% מהחיילים נמוכים ממנו?
ד) מה גובה החייל ש10% נמוכים ממנו?

פתרון:

א) $x \sim N(175,100)$.

$$p(x > 200) = p\left(z > \frac{200-175}{10}\right) = p(z > 2.5) = 1 - p(z \le 2.5) = 1 - \Phi(2.5) = 1 - 0.9938 = 0.0062$$

ב) $$p(x < 165) = p\left(z < \frac{165-175}{10}\right) = p(z < -1) = \Phi(-1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

ג) $Z = \frac{x-175}{10} = 1.285$, x=187.85

ד) $Z = \frac{x-175}{10} = -1.285$, x=162.15

<u>התפלגות הממוצע</u>:

אם $x \sim N(\mu, \sigma^2)$ ז"א $Z_x = \frac{x-\mu}{\sigma}$. ואז $\bar{x}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ וע"פ הגדרת ציון התקן $Z_{\bar{x}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ ומתקבל כי:

$$\sum_n \sim N(n\cdot\mu, n\cdot\sigma^2) \text{ ואז } Z_{\Sigma} = \frac{\sum_x - n\cdot\mu}{\sigma\cdot\sqrt{n}}$$

דוגמה:

התפלגות משקל של ילדים הים נורמאלית עם ממוצע 40 ושונות של 49 ק"ג.

א) אם נבחר ילד באופן מקרי מה ההסתברות שמשקלו מעל 50 ק"ג?

ב) מה ההסתברות שממוצע משקלם של עשרה ילדים לא יעלה על 45 ק״ג?

ג) מה ההסתברות שסכום משקלם של שבעה ילדים יהיה לא פחות מ270 ק״ג?

פתרון:

א) $p(x > 50) = p\left(z > \frac{50-40}{7}\right) = 1 - p(z < 1.41) = 1 - 0.9207 = 0.0793$

ב) $p(\overline{X_{10}} \leq 45) = p\left(Z_{\bar{x}} < \frac{45-40}{7/\sqrt{10}}\right) = p(Z < 2.26) = \Phi(2.226) = 0.9881$

ג) $p(\sum 7 \geq 270) = p\left(Z_{\sum} \geq \frac{270-7\cdot 40}{7\cdot\sqrt{7}}\right) = p(Z_{\sum} \geq -0.53) = 1 - p(Z_{\sum} < -0.53) =$
$\Phi(0.53) = 0.7019$

<u>משפט הגבול המרכזי:</u>

אם לוקחים מדגם מקרי בגודל n $(n > 30)$ מתוך אוכלוסיה שלא מתפלגת נורמאלית או שהתפלגותה לא ידועה לא ידוע. אם התוחלת μ והשונות σ^2 ידועות באותה אוכלוסיה אז $\bar{x}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

<u>הסקה סטטיסטית:</u>

הסקה סטטיסטית עוסקת בשיטות להסקת מסקנות על כלל האוכלוסייה על סמך תוצאות של מדגם מייצג. רוב המחקרים מתבססים על תוצאות מדגם ולא על כל האוכלוסייה.

סימונים:	מדגם	אוכלוסייה
ממוצע	$\bar{x}$	μ
שונות	$\hat{s}^2$	σ^2
פרופורציה	$\hat{p}$	p

שני מושגים הקשורים להסקה סטטיסטית:

1. רווח בר סמך
2. בדיקת השערות

<u>רווח בר סמך:</u>

הממוצע הכללי של אוכלוסיה נמצא: $\bar{x} - \text{סטייה} \leq \mu \leq \bar{x} + \text{סטייה}$

כאשר אנו רוצים למצוא את הרווח אנחנו לוקחים בחשבון טעות שנקראת גם רמת מובהקות ומסומנת בα או רמת בטחון/סמך שמסומנת ב$1-\alpha$. רווח הסמך הוא תחום סימטרי שבו נמצא הממוצע האמיתי של האוכלוסייה כאשר המטרה שלנו היא למצוא את גבולות הרווח, כאשר: $p(a \leq \mu \leq b) = 1 - \alpha$.
לצורך בניית הגבולות נשתמש במשפט הגבול המרכזי.

מקרה I: רווח סמך לתוחלת כאשר השונות באוכלוסיה ידועה: (σ^2)

$\bar{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$. אורך רווח הסמך: $L = 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$, והסטייה בין ממוצע המדגם לאוכלוסייה $\frac{L}{2} = \epsilon = Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$. כאשר σ-סטיית התקן. n- גודל המדגם. z – ערך מטבלת ההתפלגות הנורמאלית.

תרגיל:

רופא רוצה לאמוד את ממוצע לחץ השם של אוכלוסיית מבוגרים. לשם כך לשח מדגם מקרי של 225 מבוגרים ונמצא ממוצע של 148. ידוע כי סטטית התקן באוכלוסייה היא 20.

א) חשבו רווח סמך ברמת בטחון של 95%

ב) מהו אורך רווח הסמך?

ג) מה הסטייה בין ממוצע המדגם לתוחלת האוכלוסייה?

ד) אם ניקח מדגם גדול פי ארבע, מה יקרה לאורך הרווח סמך?
ה) חשבו את סעיף א עבור רמת בטחון של 99%

פתרון:

א) $\alpha = 0.05$, נציב בנוסחא לרווח סמך כאשר σ=20. n=225, $\bar{x} = 148$
נציב בהתאם: $\bar{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$
$148 - Z_{1-\frac{0.05}{2}} \cdot \frac{20}{\sqrt{225}} \le \mu \le 148 + Z_{1-\frac{0.05}{2}} \cdot \frac{20}{\sqrt{225}}$ נבצע חישובים ונסתכל בטבלת ההתפלגות הנורמאלית כדי למצוא את Z. $Z_{1-\frac{0.05}{2}} = Z_{0.975} = 1.96$, שוב נציב. ונקבל כי:
$145.387 \le \mu \le 150.613$ ברמת בטחון של 95%.

ב) $L = 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 2 \cdot 1.96 \cdot \frac{20}{\sqrt{225}} = 5.226$

ג) $\frac{L}{2} = \epsilon = \frac{5.226}{2} = 2.613$

ד) $L = 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ ע"פ הנוסחא נקבל כי L יקטן פי 2.

ה) $= 0.01$, נציב בנוסחא לרווח סמך כאשר σ=20. n=225, $\bar{x} = 148$
נציב בהתאם: $\bar{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$
$148 - Z_{1-\frac{0.01}{2}} \cdot \frac{20}{\sqrt{225}} \le \mu \le 148 + Z_{1-\frac{0.01}{2}} \cdot \frac{20}{\sqrt{225}}$ נבצע חישובים ונסתכל בטבלת ההתפלגות הנורמאלית כדי למצוא את Z. $Z_{1-\frac{0.01}{2}} = Z_{0.995} = 2.575$, שוב נציב. ונקבל כי:
$144.57 \le \mu \le 151.43$ ברמת בטחון של 99%.

<u>הערה:</u>

$L \ge 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ נבצע כמה חישובים ונקבל כי $n \ge \left(\frac{2Z_{1-\frac{\alpha}{2}}\sigma}{L}\right)^2$ זוהי נוסחא לחישוב מדגם מינימאלי להגבלת אורך רווח הסמך.

דוגמה:

חוקר רוצה להעריך ברמת בטחון של 95% את ממוצע השכר בישראל. ידוע שסטיית התקן באוכלוסייה היא 2000 ₪. מהו גודל המדגם הנדרש כך ש א) אורך רווח השמך לא יעלה על 400 שח? ב) הסטייה בין המדגם לאוכלוסייה לא תעלה על 400 שח?

פתרון:

א) $\alpha = 0.05$, נציב בנוסחא שפיתחנו כאשר σ=2000. $Z_{1-\frac{0.05}{2}} = 1.96$. $n \ge \left(\frac{2 \cdot 1.96 \cdot 2000}{400}\right)^2$ מכן
$n \ge 384.16 \approx 385$ש

ב) $n \ge \left(\frac{2 \cdot 1.96 \cdot 2000}{800}\right)^2$ מכן ש$n \ge 96.04 \approx 97$