<u>סטטיסטיקה / תרגיל 4#</u>

אריאל סטולרמן
קבוצה 03

ההתפלגות הנורמלית

(1)

- $\Phi(2) = \mathbf{0.9772}$
- $\Phi(2.2) = \mathbf{0.9861}$
- $\Phi(2.22) = \mathbf{0.9868}$
- $\Phi(-1.5) = 1 - \Phi(1.5) = \mathbf{0.0668}$
- $\Phi(0) = \mathbf{0.5}$
- $\Phi(-5) = 1 - \Phi(5) < 1 - \Phi(4.417) = \mathbf{0.000005}$
- $\Phi(-6) = 1 - \Phi(6) < 1 - \Phi(4.417) = \mathbf{0.000005}$

נתון: $Z \sim N(\mu = 0, \sigma = 1)$

a) להלן חישוב ההסתברויות:

- $P(-2.22 \leq Z \leq 1.55) = \Phi(1.55) - \Phi(-2.22) = \Phi(1.55) - 1 + \Phi(2.22) =$
  $= 0.9394 - 1 + 0.9868 = \mathbf{0.9262}$
- $P(-1 \leq Z \leq -0.5) = \Phi(-0.5) - \Phi(-1) = 1 - \Phi(0.5) - 1 + \Phi(1) =$
  $= -0.6915 + 0.8413 = \mathbf{0.1498}$
- $P(2 \leq Z \leq 3) = \Phi(3) - \Phi(2) = 0.9987 - 0.9772 = \mathbf{0.0215}$
- $P(Z \geq -2.5) = P(Z \leq 2.5) = \Phi(2.5) = \mathbf{0.9938}$
- $P(Z \geq -2.33) = P(Z \leq 2.33) = \Phi(2.33) = \mathbf{0.9901}$
- $P(Z \geq 3) = P(Z \leq -3) = \Phi(-3) = 1 - \Phi(3) = 1 - 0.9987 = \mathbf{0.0013}$

b) להלן חישוב האחוזונים:

- $Z_{0.25} = \mathbf{-0.6744898}$:
```
> qnorm(0.25)
[1] -0.6744898
```
- $Z_{0.75} = -Z_{1-0.25} = \mathbf{0.6744898}$
- $Z_{0.95} = \mathbf{1.645}$
- $Z_{0.975} = \mathbf{1.960}$
- $Z_{0.9} = \mathbf{1.282}$
- $Z_{0.025} = -Z_{1-0.975} = \mathbf{-1.960}$

(2)

$$X \sim N(\mu = 10, \sigma^2 = 100), \qquad Z \sim N(\mu = 0, \sigma = 1)$$

(a)

$$P(X \le 0) = P\left(\frac{X-10}{10} \le \frac{0-10}{10}\right) = P(Z \le -1) = \Phi(-1) = 1 - \Phi(1) = 1 - 0.8413 = \mathbf{0.1587}$$

(b)

$$P(X \le x) = 0.1586553 = P\left(\frac{X-10}{10} \le \frac{x-10}{10}\right) \Leftrightarrow \frac{x-10}{10} = Z_{0.1586553} \Leftrightarrow x = 10Z_{0.1586553} + 10 \Leftrightarrow$$

$$x = -10Z_{1-0.1586553} + 10 = -10Z_{0.8413447} + 10 = -10 \times 0.9999998 + 10 = \mathbf{0.000002}$$

```
> qnorm(0.8413447)
[1] 0.9999998
```

(c)

- $P\left(\frac{X-10}{10} \le \frac{x-10}{10}\right) = 0.5 \Leftrightarrow \frac{x-10}{10} = Z_{0.5} \Leftrightarrow \boldsymbol{x} = \mathbf{10}$
- $P\left(\frac{X-10}{10} \le \frac{x-10}{10}\right) = 0 \Leftrightarrow \frac{x-10}{10} = Z_0 \Leftrightarrow \frac{x-10}{10} = -Z_1 \Leftrightarrow \frac{x-10}{10} = -\infty \Leftrightarrow \boldsymbol{x} = -\infty$
- $P\left(\frac{X-10}{10} \le \frac{x-10}{10}\right) = 1 \Leftrightarrow \frac{x-10}{10} = Z_1 \Leftrightarrow \frac{x-10}{10} = \infty \Leftrightarrow \boldsymbol{x} = \infty$

(d)

$$P\left(\frac{X-10}{10} \le \frac{x-10}{10}\right) = 0.1 \Leftrightarrow \frac{x-10}{10} = Z_{0.1} \Leftrightarrow \frac{x-10}{10} = -Z_{0.9} \Leftrightarrow \frac{x-10}{10} = -1.282 \Leftrightarrow \boldsymbol{x} = \mathbf{-2.82}$$

(3)

תחת הנחה שדגימת הגבהים באוכלוסיה נעשית באופן ב״ת ואקראי לחלוטין עם סיכוי שווה לשלוף כל דגימה, ואין קשר בין גובה אחד למשנהו בפרטים באוכלוסיה, וכמובן שעבור אוסף דגימות ניתן לחשב תוחלת ושונות קבועים, והרי שגבהים הם נתונים מספריים ממשיים (ואף חיוביים), לפי משפט הגבול המרכזי הדגימות מתכנסות להתפלגות נורמלית (סטנדרטית). כמו כן, היות והגבהים הנדגמים מאפיינים זן כלשהו בטבע בעל תכונות ורקע דומים (המין האנושי), הגיוני יהיה לצפות להתרכזות רוב סביב ממוצע הגבהים, עם ירידה בכמות הדגימות לכיוון גבהים גדולים או קטנים מהממוצע.

(4)

- נניח כי בארץ יש בערך 40% השתמטות משירות צבאי, ומספר הבחורים הצעירים שווה בערך למספר הבחורות הצעירות. התפלגות שנות השירות הסדיר בקרב אוכלוסיית הבחורים והבחורות בארץ – לא סביר שהתפלגות זו תהיה נורמלית, שכן יהיו הרבה פרטים ב-0 שנות שירות, והרוב באיזור 2-3 שנות שירות, כאשר הממוצע יסבוב את ה-1.5 שנות שירות.
- התפלגות ציונים במבחן במתמטיקה עבור אוכלוסית תלמידים המורכבת משתי כיתות שוות גודל – כיתת מחוננים וכיתה טיפולית - תחת הנחה שתלמידי כיתה טיפולית גרועים במתמטיקה, ותלמידי כיתת מחוננים מצטיינים במתמטיקה, לא סביר שהתפלגות זו תהיה נורמלית – סביר שכמעט אף אחד לא יקבל ציון ממוצע, והתפלגות הציונים תהיה בערך מחצית באזור הציונים הגבוהים מאוד, ומחצית באזור הציונים הנמוכים מאוד.
- התפלגות הכנסה שנתית ממוצעת למשק בית בארה״ב בשנת 2005 – מנתונים סטטיסטיים עולה כי ממוצע ההכנסה באותה שנה עמד על $63,344, בעוד הרוב המוחלט סבב את ה-$15,000, ומעט מאוד היו בטווח מעל $150,000. גרף ההתפלגויות הוא בעל נטיה ימינה ושיא שאינו נופל באזור הממוצע אלא נמוך ממנו משמעותית.

(5)

נסמן מ"מ לציון בבחינה X, ונתון: $X \sim N(\mu = 60, \sigma = 10)$, כמו כן משתנה המתפלג נורמלית סטנדרטית Z כמוגדר לעיל.

(a)

$$P(X \geq 60) = P\left(\frac{X-60}{10} \geq \frac{60-60}{10}\right) = P(Z \geq 0) = P(Z \leq 0) = \Phi(0) = \mathbf{0.5}$$

(b)

אם ידוע לנו אך ורק שההתפלגות היא סימטרית עם ממוצע 60, ולא ידוע אם היא נורמלית או לא, עדיין ניתן היה לחשב את הסעיף הקודם, שכן הסימטריה מעידה על כך ש-60 הוא החציון, ולכן ישנו סיכוי שווה (0.5) לעבור את הבחינה או להיכשל בה (אם כוונת השאלה היא שידוע לנו אך ורק סימטריה, בלי ידיעת הממוצע, אז לא יהיה ניתן לחשב).

ננסה לחסום ע"י א"ש Chebyshev:

$$P(X - \mu \geq \epsilon) \leq \frac{\sigma^2}{\sigma^2 + \epsilon^2}$$

נציב את הנתונים:

$$P(X \geq 60 + \epsilon) \leq \frac{100}{100 + \epsilon^2} \rightarrow_{\epsilon \to 0} 1$$

קיבלנו חסם טרוויאלי לכל הסתברות שהיא, אך כיוון שיכולנו לחשב במדוייק את ההסתברות, ניתן פשוט לקחת ערך זה (0.5) כחסם.

(c)

- $P(X \geq 80) = P\left(\frac{X-60}{10} \geq \frac{80-60}{10}\right) = P(Z \geq 2) = 1 - P(Z \leq 2) = 1 - \Phi(2) = 1 - 0.9772 = \mathbf{0.0228}$
- $P(X \geq 90) = P\left(\frac{x-60}{10} \geq \frac{90-60}{10}\right) = P(Z \geq 3) = 1 - P(Z \leq 3) = 1 - \Phi(3) = 1 - 0.9987 = \mathbf{0.0013}$

(d)

$$P(X \leq 20) = P\left(\frac{x-60}{10} \leq \frac{20-60}{10}\right) = P(Z \leq -4) \cong 0.0000316$$

```
> pnorm(-4)
[1] 3.167124e-05
```

(e)

חישוב העשירון העליון:

$$P(X \leq x) = 0.9 \Leftrightarrow P\left(\frac{X-60}{10} \leq \frac{x-60}{10}\right) = 0.9 \Leftrightarrow \frac{x-60}{10} = Z_{0.9} = 1.282 \Leftrightarrow \boldsymbol{x} = \mathbf{72.82}$$

חישוב העשירון התחתון:

$$P(X \leq x) = 0.1 \Leftrightarrow P\left(\frac{X-60}{10} \leq \frac{x-60}{10}\right) = 0.1 \Leftrightarrow \frac{x-60}{10} = Z_{0.1} = -Z_{0.9} = -1.282 \Leftrightarrow \boldsymbol{x} = \mathbf{47.18}$$

(f)

מתוך 100 תלמידים:

- האחוז שעברו את הבחינה הוא האחוז שקיבלו ציון גדול או שווה ל-60 שהוא 50%, לכן סה"כ עברו את הבחינה 50 תלמידים.
- האחוז שקיבלנו ציון מעל 70 (נניח גדול או שווה ל-70):

$$P(X \geq 70) = P\left(\frac{X-60}{10} \geq \frac{70-60}{10}\right) = P(Z \geq 1) = 1 - P(Z \leq 1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

כלומר האחוז המבוקש הוא 15.87%, ומכאן שסה"כ 15 תלמידים קיבלנו ציון מעל 70.

(g) <u>לאחר פקטור של 10% לציונים, נקבל את השינויים הבאים בתוחלת ובשונות</u>:

$$\mu_{new} = E(X \times 1.1) = 1.1E(X) = 1.1\mu = 66, \quad \sigma_{new} = \sqrt{Var(X \times 1.1)} = 1.049\sigma = 10.49$$

מכאן ש-X החדש מתפלג נורמלית עם הפרמטרים: $X \sim N(\mu = 66, \sigma = 10.49)$. נחשב כעת את הסעיפים הקודמים:

- $P(X \geq 60) = P\left(\frac{X-66}{10.49} \geq \frac{60-66}{10.49}\right) = P(Z \geq -0.5719) = P(Z \leq 0.5719) \cong \mathbf{0.7157}$

התוצאה שהתקבלה **גדולה** מהתוצאה הקודמת, שכן ברור כי עם הפקטור יותר תלמידים יקבלו ציון מעל 60.

בסעיף b התשובה בהתאם לנתונים החדשים היא ש**לא** יכולנו לחשב את הסעיף הקודם רק עם הידיעה שההתפלגות סימטרית (עם ממוצע 66 וס"ת 10.49), כיוון שלא יכולנו לדעת כיצד מתפלגים הציונים סביב הממוצע שכעת אינו שווה לציון המעבר. למשל, אם מחצית הכיתה קיבלו 67, והמחצית השניה קיבלו 65, אזי כולם עברו את הבחינה. אבל, אם מחצית הכיתה קיבלו 100, והמחצית השניה קיבלו 32, אזי רק מחצית הכיתה עברה את הבחינה.

- $P(X \geq 80) = P\left(\frac{X-66}{10.49} \geq \frac{80-66}{10.49}\right) = P(Z \geq 1.334) = P(Z \leq -1.334) \cong 1 - 0.9082 = \mathbf{0.0918}$

  התוצאה שהתקבלה **גדולה** מהתוצאה הקודמת, כיוון שברור שהסיכוי לקבל ציון גבוה מ-80 גדל.

- $P(X \geq 90) = P\left(\frac{X-66}{10.49} \geq \frac{90-66}{10.49}\right) = P(Z \geq 2.287) = P(Z \leq -2.287) \cong 1 - 0.989 = \mathbf{0.011}$

  התוצאה שהתקבלה **גדולה** מהתוצאה הקודמת מאותה סיבה.

- $P(X \leq 20) = P\left(\frac{X-66}{10.49} \leq \frac{20-66}{10.49}\right) = P(Z \leq -4.385) \cong 0.0000058$

```
> pnorm(-4.385)
[1] 5.799289e-06
```

  התוצאה שהתקבלה **קטנה** מהתוצאה הקודמת, כיוון שברור שהסיכוי לקבל ציון נמוך מ-20 קטן.

- $P(X \leq x) = 0.9 \Leftrightarrow P\left(\frac{X-66}{10.49} \leq \frac{x-66}{10.49}\right) = 0.9 \Leftrightarrow \frac{x-66}{10.49} = Z_{0.9} = 1.282 \Leftrightarrow \boldsymbol{x = 79.44}$

  התוצאה שהתקבלה **גדולה** מהתוצאה הקודמת, כיוון שכל הציונים גבוהים יותר, בפרט האחוזון המבוקש.

- $P(X \leq x) = 0.1 \Leftrightarrow P\left(\frac{X-66}{10.49} \leq \frac{x-66}{10.49}\right) = 0.1 \Leftrightarrow \frac{x-66}{10.49} = Z_{0.1} = -Z_{0.9} = -1.282 \Leftrightarrow \boldsymbol{x = 52.55}$

  התוצאה שהתקבלה **גדולה** מהתוצאה הקודמת מאותה סיבה.

מתוך 100 תלמידים:

- 71.57% קיבלו מעל 60, כלומר **71 תלמידים** עברו את הבחינה.
- תלמידים שקיבלו מעל 70 (ציון גדול או שווה ל-70):

$$P(X \geq 70) = P\left(\frac{X-66}{10.49} \geq \frac{70-66}{10.49}\right) = P(Z \geq 0.381) = 1 - P(Z \leq 0.381) \cong 1 - 0.6480 = \mathbf{0.352}$$

35.2% מהתלמידים קיבלו מעל 70, כלומר **35 תלמידים**.

שתי התוצאות **גדולות** מהתוצאות הקודמות, כיוון שהסיכויים לקבלת ציון גבוה מ-60 או 70 גדל, ולכן מס' התלמידים גדל.

(6) התפלגות הטמפ' בישראל לשנת 2006: $T \sim N(\mu = 15, \sigma = 6)$

(a)

$$P(T \leq t) = 0.3 \Leftrightarrow P\left(\frac{T-15}{6} \leq \frac{t-15}{6}\right) = 0.3 \Leftrightarrow \frac{t-15}{6} = Z_{0.3} = -0.5244 \Leftrightarrow \boldsymbol{t} \cong \mathbf{11.85}$$

```
> qnorm(0.3)
[1] -0.5244005
```

הטמפ' הגבוהה ביותר מבין 30% הימים הקרים בשנה היתה 11.85 מעלות.

(b)

$$P(T \geq 20) = P\left(\frac{T-15}{6} \geq \frac{20-15}{6}\right) = P(Z \geq 0.8333) = 1 - P(Z \leq 0.8333) \cong 1 - 0.7967 = \mathbf{0.2033}$$

ב-20.33% מהימים היה חם יותר מ-20 מעלות.

(c)

$$T_c = \frac{5}{9}\left(T_f - 32\right) \Leftrightarrow T_f = \frac{9}{5}T_c + 32$$

$$\Rightarrow E\left(T_f\right) = E\left(\frac{9}{5}T_c + 32\right) = \frac{9}{5}E(T_c) + 32 = \mathbf{59}$$

אין צורך להכיר את תכונות ההתפלגות הנורמלית בשביל הפתרון (מספיק לפי תכונות התוחלת).

(d)

$$Var\left(T_f\right) = Var\left(\frac{9}{5}T_c + 32\right) = \frac{81}{25}Var(T_c) = \frac{81}{25}\sigma^{\mathbf{2}}_{T_c} = \mathbf{116.64} \quad (\Rightarrow \sigma = 10.8)$$

אין צורך להכיר את תכונות ההתפלגות הנורמלית בשביל הפתרון (מספיק לפי תכונות השונות).

(e) מ"מ הטמפ' בפרנהייט: $T_f \sim N(\mu = 59, \sigma = 10.8)$

$$P\left(T_f \leq t\right) = 0.3 \Leftrightarrow P\left(\frac{T_f - 59}{10.8} \leq \frac{t-59}{10.8}\right) = 0.3 \Leftrightarrow \frac{t-59}{10.8} = Z_{0.3} = -0.5244 \Leftrightarrow \boldsymbol{t} \cong \mathbf{53.33}$$

או ניתן בחישוב ישיר:

$$P\left(T_f \leq t\right) = 0.3 \Leftrightarrow t = \frac{9}{5} \times 11.85 + 32 = \mathbf{53.33}$$

אין צורך להכיר את תכונות ההתפלגות הנורמלית בשביל הפתרון, **בהינתן הפתרון לסעיף a** (עבורו צריך תכונות נורמלית).

(f)

$$P(T_f \le t) = 0.44 \Leftrightarrow \frac{t-59}{10.8} = Z_{0.44} = -Z_{0.56} \cong -0.15 \Leftrightarrow \boldsymbol{t} \cong \mathbf{57.38}$$

(7) נתון $X \sim B(n,p)$

(a)

n=10, p=1/10 :

$$P(X=np) = P(X=1) = \binom{10}{1}\left(\frac{1}{10}\right)^1\left(\frac{9}{10}\right)^9 \cong 10 \times \left(\frac{1}{10}\right) \times 0.3874 = \mathbf{0.3874}$$

(b)

$$X \sim B(n,p) \Rightarrow X \approx N\big(np, np(1-p)\big) \Leftrightarrow X \approx N\left(\mu = 1, \sigma^2 = \frac{9}{10}\right) \Rightarrow$$

$$P_B(X=1) = P_N(1 \le X \le 2) = P_N\left(\frac{1-1}{\sqrt{\frac{9}{10}}} \le \frac{X-1}{\sqrt{\frac{9}{10}}} \le \frac{2-1}{\sqrt{\frac{9}{10}}}\right) = P_N\left(0 \le Z \le \sqrt{\frac{10}{9}}\right)$$

$$= \Phi(1.054) - \Phi(0) = 0.854 - 0.5 = \mathbf{0.354}$$

(c)

n=100, p=1/4 :

$$P(X=np) = P(X=25) = \binom{100}{25}\left(\frac{1}{4}\right)^{25}\left(\frac{3}{4}\right)^{75} = \mathbf{0.0917}$$

(d)

$$X \sim B(n,p) \Rightarrow X \approx N\big(np, np(1-p)\big) \Leftrightarrow X \approx N(\mu = 25, \sigma^2 = 18.75) \Rightarrow$$

$$P_B(X=25) = P_N(25 \le X \le 26) = P_N\left(\frac{25-25}{4.33} \le \frac{X-25}{4.33} \le \frac{26-25}{4.33}\right) = P_N(0 \le Z \le 0.2309)$$

$$= \Phi(0.2309) - \Phi(0) = 0.5913 - 0.5 = \mathbf{0.0913}$$

(e)

ברור כי טיב הקירוב במקרה השני (99.56%) גבוה יותר מאשר המקרה הראשון (91.38%), וזאת כיוון ש-n גדול יותר במקרה השני.

(8)

התפלגות הגבהים במדינת ישראל: $X \sim N(\mu,\sigma)$

(a)

- $P(X \ge \mu+\sigma) = P\left(\frac{X-\mu}{\sigma} \ge \frac{\mu+\sigma-\mu}{\sigma}\right) = P(Z \ge 1) = 1 - \Phi(1) = \mathbf{0.1587}$
- $P(X \le \mu-\sigma) = P\left(\frac{X-\mu}{\sigma} \le \frac{\mu-\sigma-\mu}{\sigma}\right) = P(Z \le -1) = \Phi(-1) = 1 - \Phi(1) = \mathbf{0.1587}$

(b)

$$P(X \geq \mu + \sigma \ \vee X \leq \mu - \sigma) = P(X \geq \mu + \sigma) + P(X \leq \mu - \sigma) = 2 \times 0.1587 = \mathbf{0.3174}$$

(c)

- $P(X \geq \mu + 2\sigma \ \vee \ X \leq \mu - 2\sigma) = P(X \geq \mu + 2\sigma) + P(X \leq \mu - 2\sigma) = P\left(\frac{X-\mu}{\sigma} \geq \frac{\mu+2\sigma-\mu}{\sigma}\right) + P\left(\frac{X-\mu}{\sigma} \leq \frac{\mu-2\sigma-\mu}{\sigma}\right) = P(Z \geq 2) + P(Z \leq -2) = 2 - 2\Phi(2) = 2 - 2 \times 0.9772 = \mathbf{0.0456}$
- $P(X \geq \mu + 3\sigma \ \vee \ X \leq \mu - 3\sigma) = P(X \geq \mu + 3\sigma) + P(X \leq \mu - 3\sigma) = P\left(\frac{X-\mu}{\sigma} \geq \frac{\mu+3\sigma-\mu}{\sigma}\right) + P\left(\frac{X-\mu}{\sigma} \leq \frac{\mu-3\sigma-\mu}{\sigma}\right) = P(Z \geq 3) + P(Z \leq -3) = 2 - 2\Phi(3) = 2 - 2 \times 0.9987 = \mathbf{0.0026}$

(d)

התפלגות הגבהים בשוודיה: $Y \sim N(\mu^*, \sigma^*)$

$$P(Y \geq \mu^* + \sigma^*) = P\left(\frac{Y - \mu^*}{\sigma^*} \geq \frac{\mu^* + \sigma^* - \mu^*}{\sigma^*}\right) = P(Z \geq 1) = 1 - \Phi(1) = \mathbf{0.1587}$$

(e)

אם נניח הנחה סבירה שממוצע הגבהים בארץ הוא 1.75 מ׳, וכי ממוצע זה מתייחס לאוכלוסייה בוגרת (נניח גילאים 16-65), אז אם סטיית הגבהים בישראל היא חצי מטר, לפי החישובים לעיל כ-31.7% מהאוכלוסיה הבוגרת בארץ תהיה בגובה 2.25 מ׳ ומעלה או 1.25 מ׳ ומטה, ונתון זה לא נראה סביר. על כן לא נראה סביר שסטיית התקן תהיה חצי מטר.

דגימה

(a)

i. **אוכלוסיה**.

ii. **מדגם מקרי** (אוכלוסיה: עוגות אותו יום).

iii. **מדגם** (ישנו קשר בין הדגימות: כולם מאותו יום).

iv. **מדגם מקרי** (אוכלוסיה: כלל העוגות במאפייה שנוצרו אי פעם).

v. **מדגם** (ישנה תלות בין הדגימות: כולן בטעם כבש).

vi. יש בשאלה זו עניין של הבנת הנקרא, אני מבין זאת כ**מדגם מקרי** – כלומר מתוך כלל עוגות הכבש במאפייה, אלו היחידות שנבחרו ע״מ לייצג את אוכלוסיית כלל עוגות הכבש שיוצרו אי פעם במאפייה (בהנחה שנלקחו בימים שונים). ניתן להבין זאת גם כך: אלו כל עוגות הכבש במאפייה, והן נבחרו לייצג את כל עוגות הכבש במאפייה, כלומר הן אוכלוסיה (שנלקחה כולה כמדגם).

(b) יהי X מ״מ המייצג את מספר הצימוקים בעוגה:

$$E(X) = \frac{1 \times 54 + 1 \times 60 + 1 \times 44 + 1 \times 56}{4} = \mathbf{53.5}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{54^2 + 60^2 + 44^2 + 56^2}{4} - 53.5^2 = \mathbf{34.75}$$