רגרסיה ליניארית מרובה

- רווח סמך עבור $E(y_h)$:
  - x_h הינו וקטור התצפית הכולל את ערכי כל k המשתנים המסבירים עבור תצפית h

$$X_h = \begin{pmatrix} 1 \\ x_{1h} \\ x_{2h} \\ \dots \\ x_{kh} \end{pmatrix} \Rightarrow \boxed{E(y_h) = E(y \mid x_h) = ?}$$

  - ראשית יש לדעת כיצד $\hat{y}_h$ מתפלג:

$$\boxed{\hat{y}_h = X_h^T \times b} \Leftarrow$$ קומבינציה ליניארית ולכן מתפלג נורמלית

$$E(\hat{y}_h) = E(X_h^T \times b) = X_h^T \times E(b) = X_h^T \times \beta$$

$$V(\hat{y}_h) = \sum(\hat{y}_h) = \sum(X_h^T \times b) = X_h^T \times \sum(b) \times X_h$$

$$= X_h^T \times \sigma^2 \times (X^T \times X)^{-1} \times X_h = \sigma^2 \times X_h^T \times (X^T \times X)^{-1} \times X_h$$

$$\Rightarrow \boxed{\hat{y}_h \sim N\left(\underbrace{X_h^T}_{1\times(k+1)} \times \beta,\quad \sigma^2 \times \underbrace{X_h^T}_{1\times(k+1)} \times \underbrace{(X^T \times X)^{-1}}_{(k+1)\times(k+1)} \times \underbrace{X_h}_{(k+1)\times 1}\right)}$$

- נניח לדוגמא שיש לנו שני משתנים מסבירים כמו בדוגמא הקודמת:
  - אם יש אינסוף איזורים עם:

גודל אוכלוסייה: $x_{1h}=200$

רמת הכנסה ממוצעת: $x_{2h}=2,400$

- נרצה לחזות באיזה תחום ינוע מספר הקופסאות שימכרו בממוצע באינסוף אזורים שגודל האוכלוסייה בהם 200 אלף אנשים ורמת הכנסה שנתית של 2,400 אלף לנפש ברמת ביטחון של $1-\alpha$.
- רוצים למצוא רווח בר סמך לתוחלת מספר הקופסאות שנמכרו באינסוף אזורים המאופיינים בגודל אוכלוסייה של 200 אלף אנשים עם הכנסה שנתית של 2,400 אלף לנפש.

$$y_h = \beta_0 + \beta_1 x_{1h} + \beta_2 x_{2h} + E_h$$

אנו רוצים לבנות רב"ס עבור $E(y_h)$:

$$E(y_h) = E(\beta_0 + \beta_1 x_{1h} + \beta_2 x_{2h} + E_h) = \beta_0 + \beta_1 x_{1h} + \beta_2 x_{2h} + E(E_h)$$

ע"פ הנחות המודל שווה לאפס

$$E(y_h) = \beta_0 + \beta_1 x_{1h} + \beta_2 x_{2h}$$

כדי לבנות רב"ס אנו צריכים למצוא סטטיסטי שהתפלגותו ידועה וקשורה לפרמטר

$$\hat{y}_h = b_0 + b_1 x_{1h} + b_2 x_{2h}$$

$$E(\hat{y}_h) = \beta_0 + \beta_1 x_{1h} + \beta_2 x_{2h} = E(y_h)$$

שווה ל $E(y_h)$ ולכן $\hat{y}_h$ הינו אח"ה ל- $E(y_h)$

- רווח סמך עבור $E(y_h)$ (המשך):
  - כמו כן: $$E(y_h) = E(\hat{y}_h) = X_h^T \times \beta$$
  - הואיל ו- σ^2 אינו ידוע נשתמש ב- MSE:

$$\hat{\sigma}_{\hat{y}_h} = S(\hat{y}_h) = \sqrt{MSE \cdot \left(X_h^T \times \left(X^T \times X\right)^{-1} \times X_h\right)}$$

  - על כן: $$\frac{\hat{y}_h - E(y_h)}{S(\hat{y}_h)} \sim t_{n-k-1}$$
  - נקבל את רווח הסמך הבא עבור $E(y_h)$:

$$P\left(\hat{y}_h - t_{n\text{-}k\text{-}1,1-\frac{\alpha}{2}} S(\hat{y}_h) \le E(y_h) \le \hat{y}_h + t_{n\text{-}k\text{-}1,1-\frac{\alpha}{2}} S(\hat{y}_h)\right) = 1 - \alpha$$

- רווח סמך עבור $E(y_h)$, דוגמה חישובית -
  - להלן וקטור X_h:

$$X_h = \begin{pmatrix} 1 \\ 200 \\ 2400 \end{pmatrix}$$

  - נחשב את ערך $\hat{y}_h$ המתאים:

$$\begin{aligned} \hat{y}_h &= b_o + b_1 \cdot x_{1h} + b_2 \cdot x_{2h} \\ &= 3.563 + 0.496 \cdot 200 + 0.009 \cdot 2400 = 124.805 \end{aligned}$$

  - נחשב את השונות של $\hat{y}_h$:

$$S(\hat{y}_h) = \sqrt{MSE \cdot \left(X_h^T \times \left(X^T \times X\right)^{-1} \times X_h\right)} = 0.723$$

- רווח סמך עבור $E(y_h)$, דוגמה חישובית (המשך) -
  - לבסוף, נבנה את רווח הסמך:

$$P\left(\hat{y}_h - t_{12,0.975} \cdot S(\hat{y}_h) \le E(y_h) \le \hat{y}_h - t_{12,0.975} \cdot S(\hat{y}_h)\right) = 0.95$$

$$P\left(124.805 - 2.56 \cdot 0.723 \le E(y_h) \le 124.805 + 2.56 \cdot 0.723\right) = 0.95$$

$$P\left(122.953 \le E(y_h) \le 126.657\right) = 0.95$$

- רווח בר סמך עבור y_h (רווח חיזוי):
  - נתבונן במשתנה המקרי: $\boxed{\hat{y}_h - y_h = ?}$
  - ראשית יש לדעת כיצד משתנה מקרי זה מתפלג:

$$E(\hat{y}_h - y_h) = X_h^T \times \beta - X_h^T \times \beta = 0$$

$$V(\hat{y}_h - y_h) = V(\hat{y}_h) + V(y_h)$$

$$= \sigma^2 \times X_h^T \times (X^T \times X)^{-1} \times X_h + \sigma^2$$

$$\Rightarrow \boxed{\hat{y}_h - y_h \sim N\left\{0, \quad \sigma^2 \cdot \left[1 + X_h^T \times (X^T \times X)^{-1} \times X_h\right]\right\}}$$

  - הואיל ו- σ^2 אינו יודע נשתמש ב- MSE:

$$\boxed{\hat{\sigma}_{\hat{y}_h - y_h} = S(\hat{y}_h - y_h) = \sqrt{MSE \cdot \left(1 + X_h^T \times (X^T \times X)^{-1} \times X_h\right)}}$$

- רווח סמך עבור y_h (המשך):
  - על כן:

$$E\left(\hat{y}_h - y_h\right) = 0$$

$$\frac{\hat{y}_h - y_h}{S\left(\hat{y}_h - y_h\right)} \sim t_{n-k-1}$$

  - נקבל את רווח הסמך הבא עבור $E(y_h)$:

$$P\left(\hat{y}_h - t_{n\text{-}k\text{-}1,1\text{-}\frac{\alpha}{2}} S\left(\hat{y}_h - y_h\right) \le y_h \le \hat{y}_h + t_{n\text{-}k\text{-}1,1\text{-}\frac{\alpha}{2}} S\left(\hat{y}_h - y_h\right)\right) = 1 - \alpha$$

- רווח בר סמך עבור y_h (רווח חיזוי), דוגמה חישובית -
  - נשתמש באותו וקטור X_h:

$$X_h = \begin{pmatrix} 1 \\ 200 \\ 2400 \end{pmatrix}$$

ועל כן גם $\hat{y}_h$ נשאר זהה: $\hat{y}_h = 124.805$

  - נחשב את השונות של $y_h - \hat{y}_h$:

$$S(y_h - \hat{y}_h) = \sqrt{MSE \cdot \left(1 + X_h^T \times \left(X^T \times X\right)^{-1} \times X_h\right)} = 2.3$$

  - לבסוף, נבנה את רווח הסמך:

$$P\left(\hat{y}_h - t_{12,0.975} \cdot S(y_h - \hat{y}_h) \le y_h \le \hat{y}_h - t_{12,0.975} \cdot S(y_h - \hat{y}_h)\right) = 0.95$$

$$P\left(124.805 - 2.56 \cdot 2.3 \le y_h \le 124.805 + 2.56 \cdot 2.3\right) = 0.95$$

$$P\left(118.911 \le y_h \le 130.699\right) = 0.95$$

- בדיקת הנחות המודל –
  1. ליניאריות המודל
  2. השגיאות בלתי מתואמות
  3. הומוסכדסטיות – שונות קבועה של השגיאות בכל הטווח
  4. השגיאות מפולגות נורמלית (Ei)
  5. חוסר מולטיקוליניאריות – חוסר קורלציה (מתאם) בין המשתנים הבלתי-תלויים (המסבירים)

נבדוק את ההנחות הנ"ל באמצעות מבחנים סטטיסטיים שונים

1. ליניאריות המודל –

- לצורך בדיקת הנחה זו ניתן להשתמש במבחן F והסטטיסטי מטבלת ניתוח שונות
- כזכור, מבחן F בודק:

H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ → המודל אינו ליניארי

H_1: (לפחות אחד שונה מאפס) אחרת ← המודל ליניארי

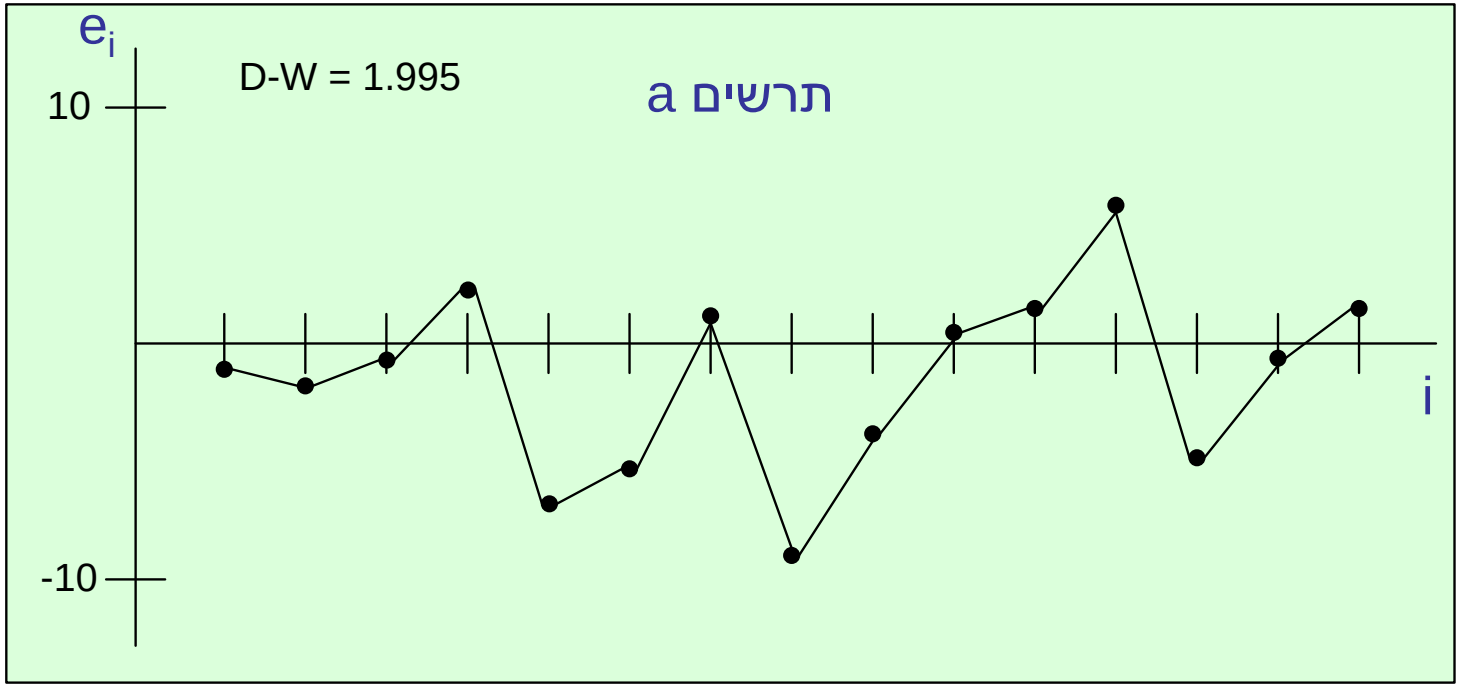
- אם לא מבצעים רגרסיה בשלבים, הוספה של משתנה בלתי תלוי אחד בכל פעם, אזי ייתכן שאחד המשתנים הבלתי תלויים אינו מקיים קשר ליניארי (אינו מובהק) למרות שמשוואת הרגרסיה בסה"כ מובהקת

❖ קיימים מבחנים נוספים לבדיקת ליניאריות, סטטיסטיים (מבחן Chow) ולא פורמאליים (גרף שאריות)

2. השגיאות בלתי מתואמות –

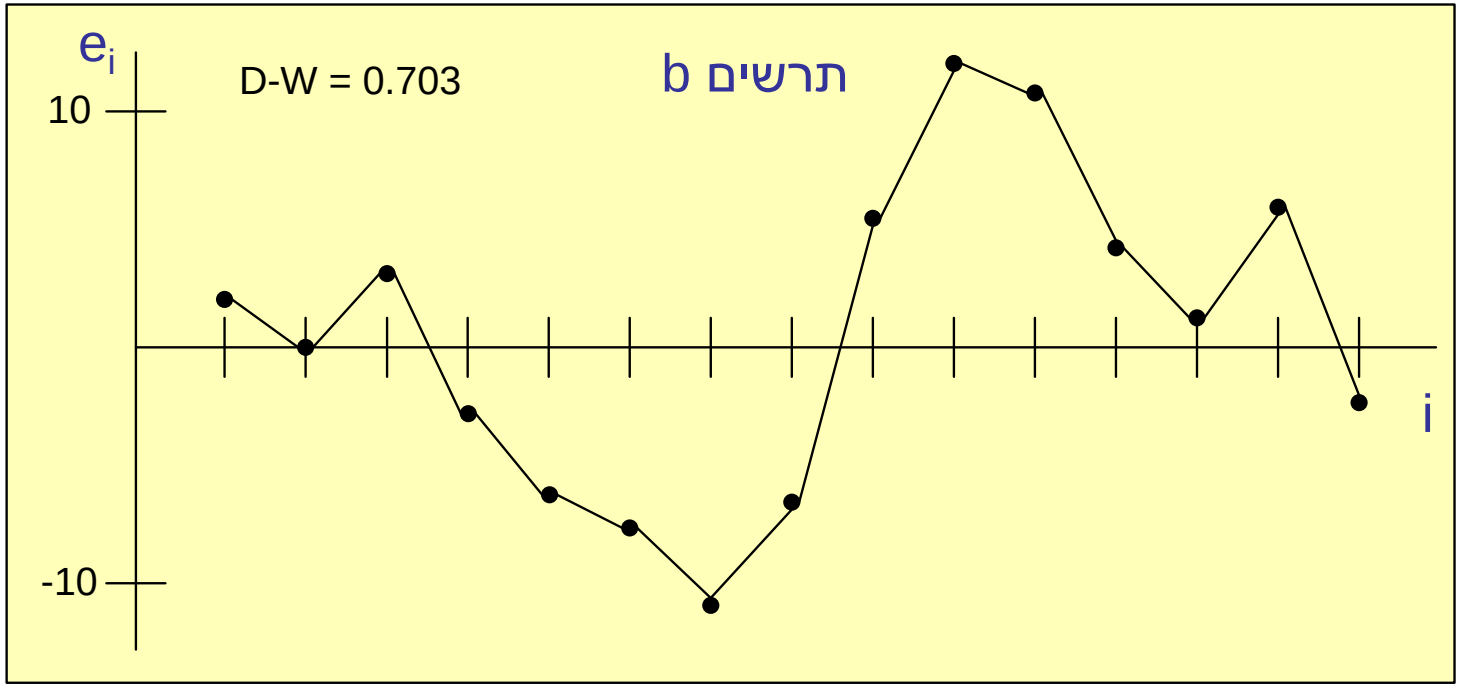
- ניתן להעלות על גרף את השאריות ולהתרשם מפיזורן
- נבחין בין שלושה מקרים שכיחים :

i. תרשים a - מדגים התפלגות שגיאות של תהליך אקראי שגיאות אקראיות



2. השגיאות בלתי מתואמות (המשך) –

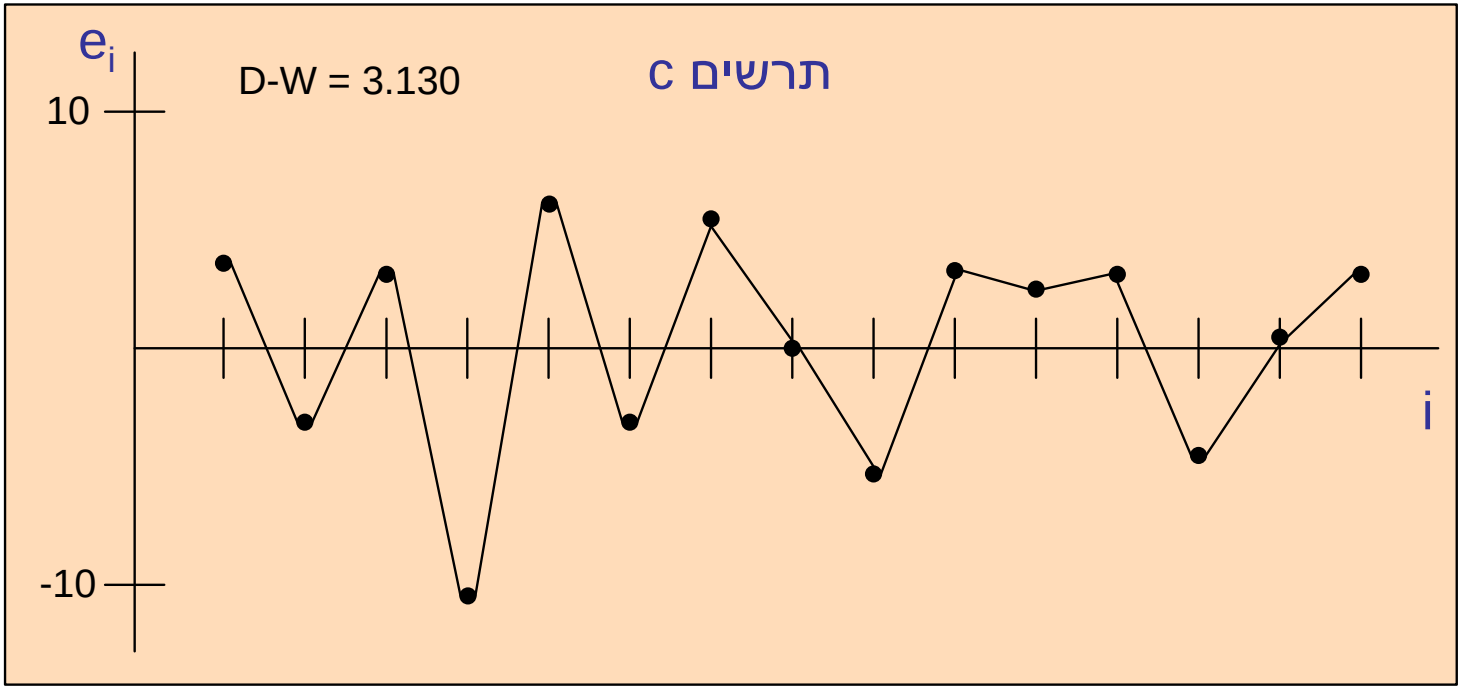
ii. תרשים b - מדגים קיום של מגמה בשגיאות (דפוס של תנועה איטית בשגיאות)



בתחילה, השגיאות נשארות שליליות יותר פעמים מהמצופה, ולאחר מכן נשארות חיוביות יותר פעמים מהמצופה

2. השגיאות בלתי מתואמות (המשך) –

iii. תרשים c - מדגים תנועת זיג-זג של השגיאות (דפוס של תנועה מהירה בשגיאות)



השגיאות קופצות למעלה ולמטה, פעם חיובית ופעם שלילית וחוזר חלילה

2. השגיאות בלתי מתואמות (המשך) –

- נשתמש במבחן הסטטיסטי בשם "דרבין ווטסון" (Durbin Watson):

$$DW = d = \frac{\sum_{i=2}^{n}\left(e_i - e_{i-1}\right)^2}{\sum_{i=1}^{n} e_i^2}$$

סכום ריבועים של הפרשי שגיאות עוקבות

סכום ריבועי השגיאות (SSE)

- הסטטיסטי של מבחן "דרבין ווטסון" רגיש לדפוסים שתוארו בתרשימים a – c
- סטטיסטי DW, בעל התפלגות סימטרית , נע בין 0 ל- 4 עם תוחלת 2

2. השגיאות בלתי מתואמות (המשך) – דוגמה חישובית לשימוש במבחן "דרבין ווטסון" :

TABLE 6-6 COMPUTATION OF THE DURBIN-WATSON STATISTIC FOR THREE SETS OF RESIDUALS (ERRORS)

	Set a			
Time	(1) Error e	(2) Squared Error	(3) $(e_t - e_{t-1})$	(4) $(e_t - e_{t-1})^2$
1	−0.60	0.36		
2	−0.90	0.81	−0.30	0.09
3	−0.30	0.09	0.60	0.36
4	3.20	10.24	3.50	12.25
5	−6.10	37.21	−9.30	86.49
6	−5.40	29.16	0.70	0.49
7	1.30	1.69	6.70	44.89
8	−8.10	65.61	−9.40	88.36
9	−2.20	4.84	5.90	34.81
10	0.30	0.09	2.50	6.25
11	1.30	1.69	1.00	1.00
12	4.90	24.01	3.60	12.96
13	−4.00	16.00	−8.90	79.21
14	−0.40	0.16	3.60	12.96
15	1.20	1.44	1.60	2.56
SS		193.40		382.68

$$\text{Durbin-Watson} = 1.995 = \frac{382.68}{193.40}$$

2. השגיאות בלתי מתואמות, מבחן "דרבין ווטסון" (המשך) –

- כאשר דפוס השגיאות הוא מהסוג של מגמה (תנועה איטית) כמו בתרשים b ⇐ ההפרשים בין שגיאות עוקבות נוטים להיות קטנים ⇐ סטטיסטי DW יהיה קטן
- כאשר דפוס השגיאות הוא מהסוג של זיג-זג (דפוס תנועה מהירה) כמו בתרשים c ⇐ ההפרשים בין שגיאות עוקבות נוטים להיות גדולים ⇐ סטטיסטי DW יהיה גדול
- כלומר, כאשר הסטטיסטי המחושב קטן או גדול מדי ⇐ זה מעיד על כך שהשגיאות מתואמות באיזו שהיא צורה

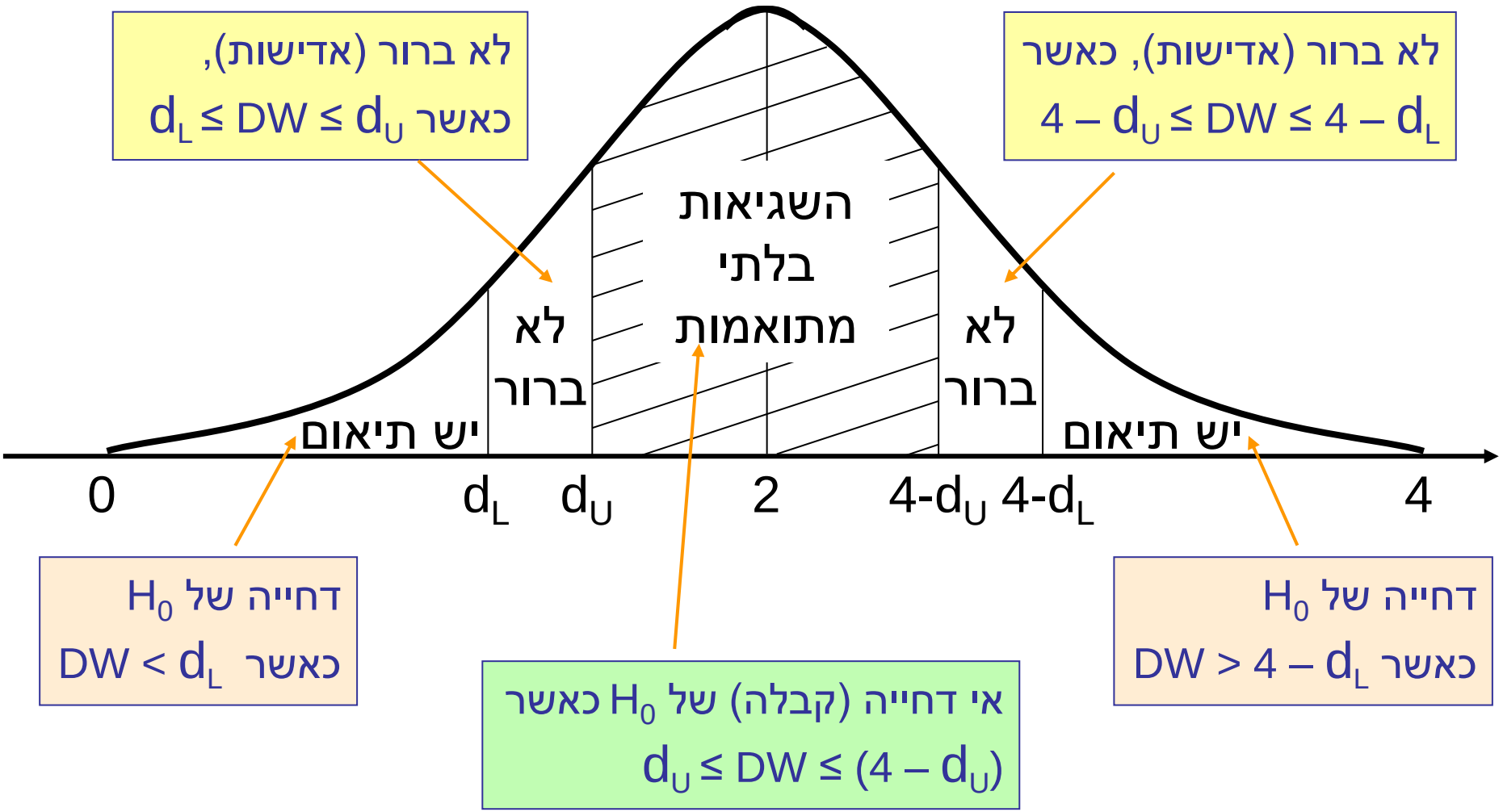
H_0: השגיאות בלתי מתואמות

H_1: השגיאות מתואמות

❖ מטרתנו במבחן זה הינה לא לדחות את H_0

2. השגיאות בלתי מתואמות, מבחן "דרבין ווטסון" (המשך) –

- ניתן לחלק את תחום הערכים של המבחן ל- 3 חלקים:



2. השגיאות בלתי מתואמות (המשך) , טבלת מבחן "דרבין ווטסון"

מס' משתנים מסבירים

מס' תצפיות

TABLE D VALUES OF THE DURBIN-WATSON STATISTIC

5 Percent Significance Points of d_l and d_u (for one-sided test)

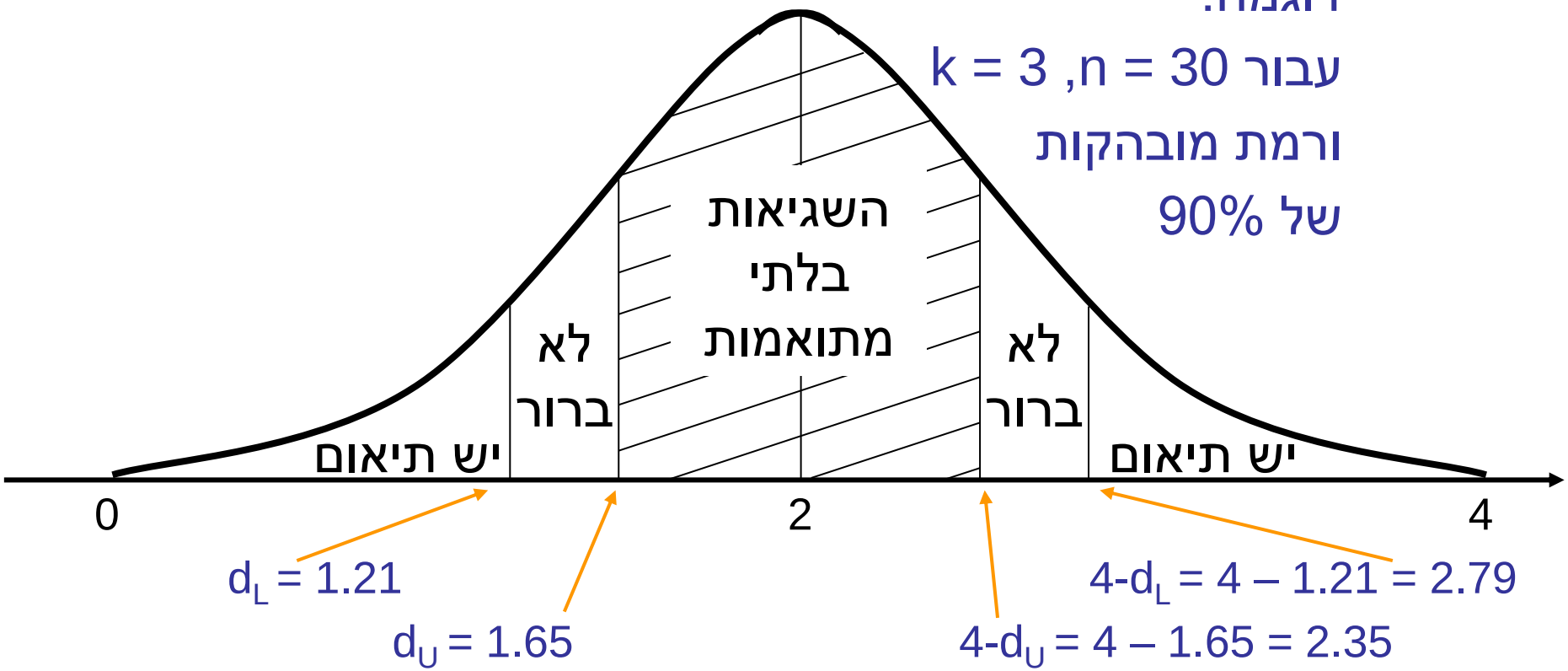
n	k =1		k =2		k =3		k =4		k =5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

Source: J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression," *Biometrika* **38** (1951), pp. 159–177. Reprinted by permission.

k = no. of regressors (independent variables)

2. השגיאות בלתי מתואמות, מבחן "דרבין ווטסון" (המשך) -

- דוגמה:
  עבור $n = 30$, $k = 3$
  ורמת מובהקות
  של 90%

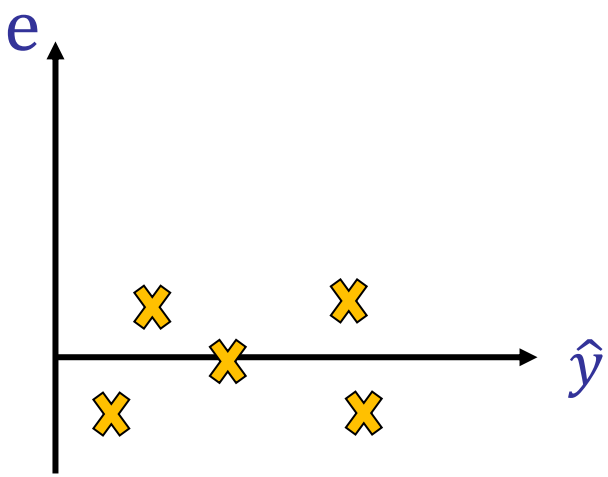


- על ידי שימוש בטבלת DW שנותנת את ערכי d_L ו- d_U למבחן חד צדדי ברמת מובהקות של 5% מושגת רמת מובהקות של 10% במבחן דו צדדי

- בדוגמא ביקשו רווח בר סמך של 10% ⇐ הולכים לטבלה של 5%
  אם מבקשים רווח סמך של 5% ⇐ הולכים לטבלה של 2.5%

3. הומוסכדטיות – הנחה של פיזור אחיד .

ברגרסיה מרובה:



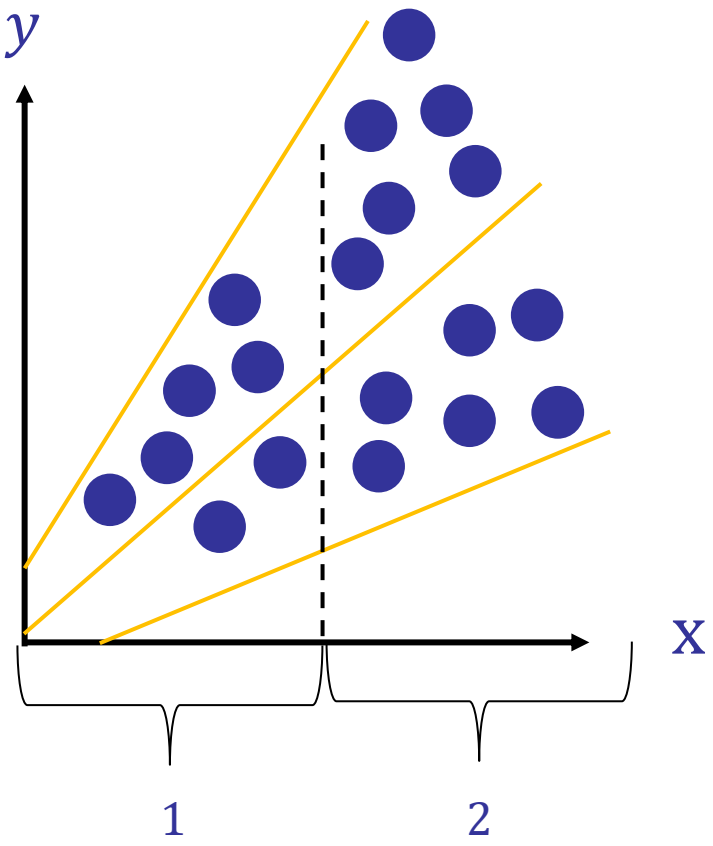
נבדוק את ההנחה לפי בדיקת השערות.

בפיזור חרוט נבדוק:

$$H_0: \sigma^2_{E_1} = \sigma^2_{E_2}$$

$$H_1: \sigma^2_{E_1} \neq \sigma^2_{E_2}$$

לפי מבחן F

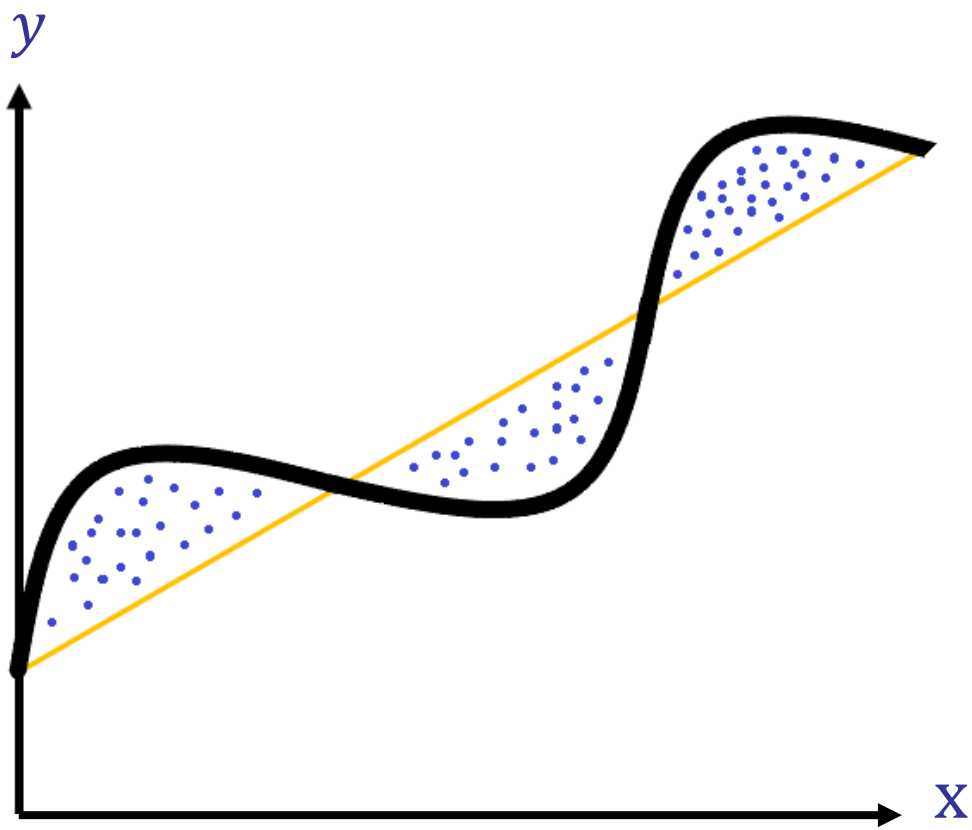


בפיזור של סינוס נבדוק השערה על תוחלות השגיאות.
(נבדוק תוחלת מעל הקו ותוחלת מתחת. התוחלות יצאו בסימונים הפוכים אך בערכים מוחלטים זהים)

$$H_0: \mu_{E_1} = \mu_{E_2}$$

$$H_1: \mu_{E_1} \neq \mu_{E_2}$$

לפי מבחן t



4. שגיאות מפולגות נורמלית

המבחן שנשתמש בו הוא χ^2

אנו מצפים שהתפלגות של השגיאות המנורמלות תהיה נורמלית סטנדרטית.

$$\frac{x_i - \mu}{\sigma} = \frac{e_i - 0}{\hat{\sigma}_e}$$

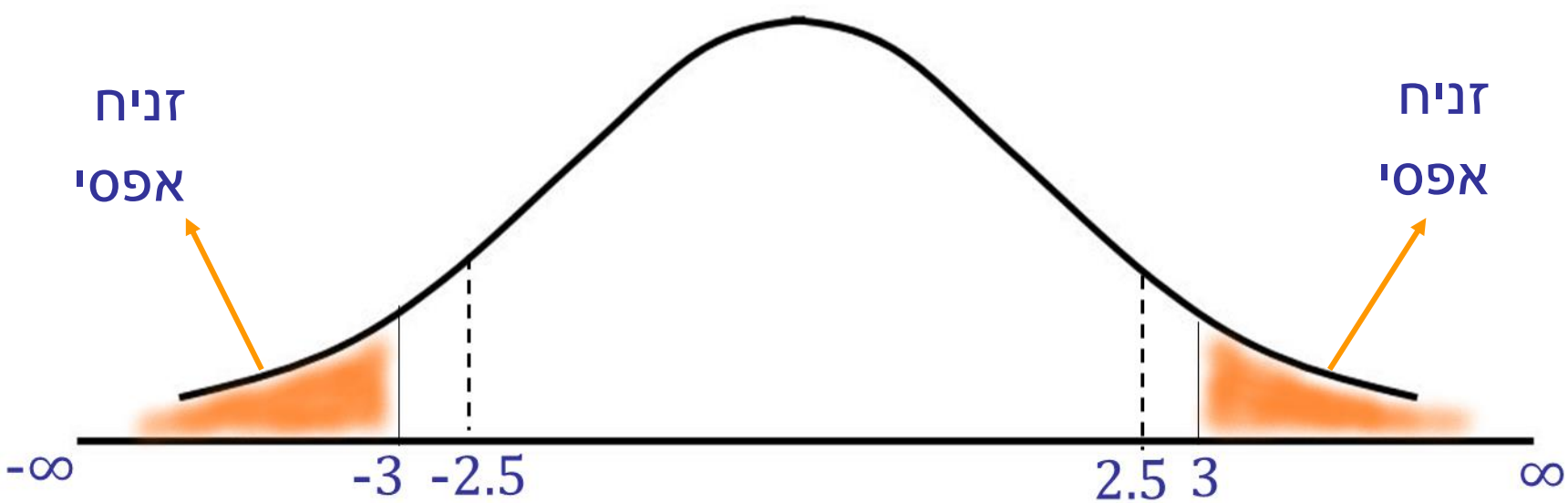
$$\hat{\sigma}_e = \sqrt{\frac{\sum (e_i)^2}{n-1}}$$

נבדוק כמה תצפיות צפויות ליפול בכל אינטרוול של התפלגות.

התפלגות נורמלית נעה מ- (∞-) עד ∞

$\Phi(3)-\Phi(2.5)$: ההסתברות ליפול בקטע [2.5 , 3]

נחלק מ-(3-) עד 3
בקפיצות של 0.5



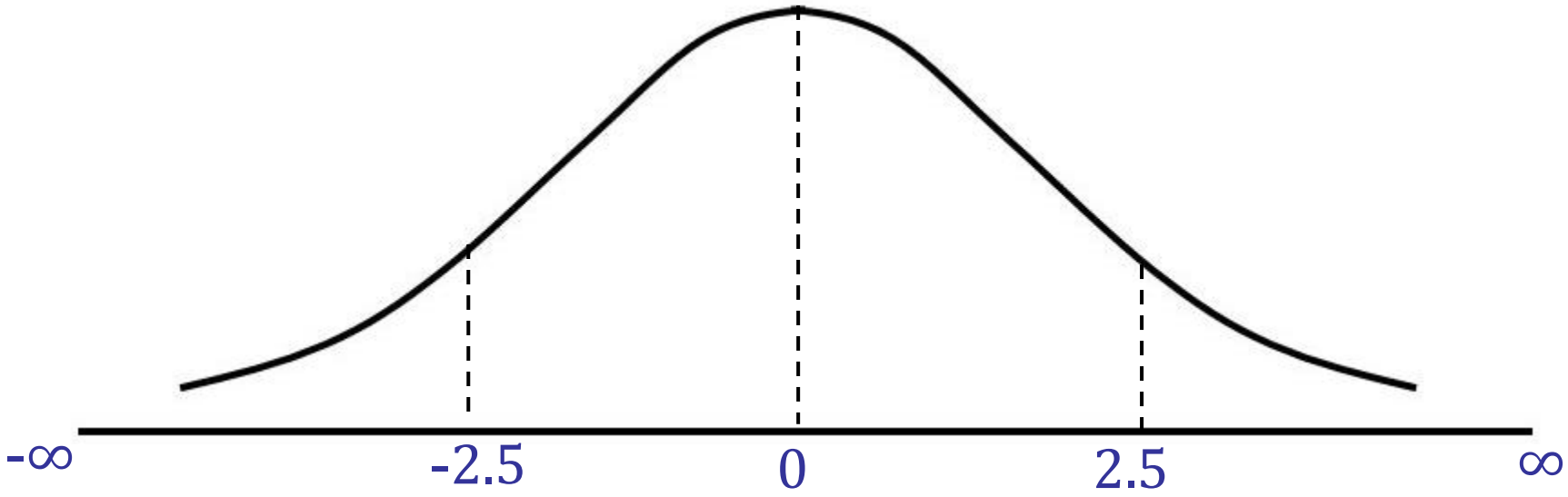
נבדוק כמה תצפיות צפויות ליפול בכל אינטרוול של התפלגות.

התפלגות נורמלית נעה מ- (∞-) עד ∞

כדי שההסתברות המצטברת בכל התחומים תהיה שווה ל-1,
אפשר להתחיל לצבור ממינוס אינסוף עד אינסוף.

ההסתברות לתחום התחתון תחושב כ: $\Phi(-2.5)$.

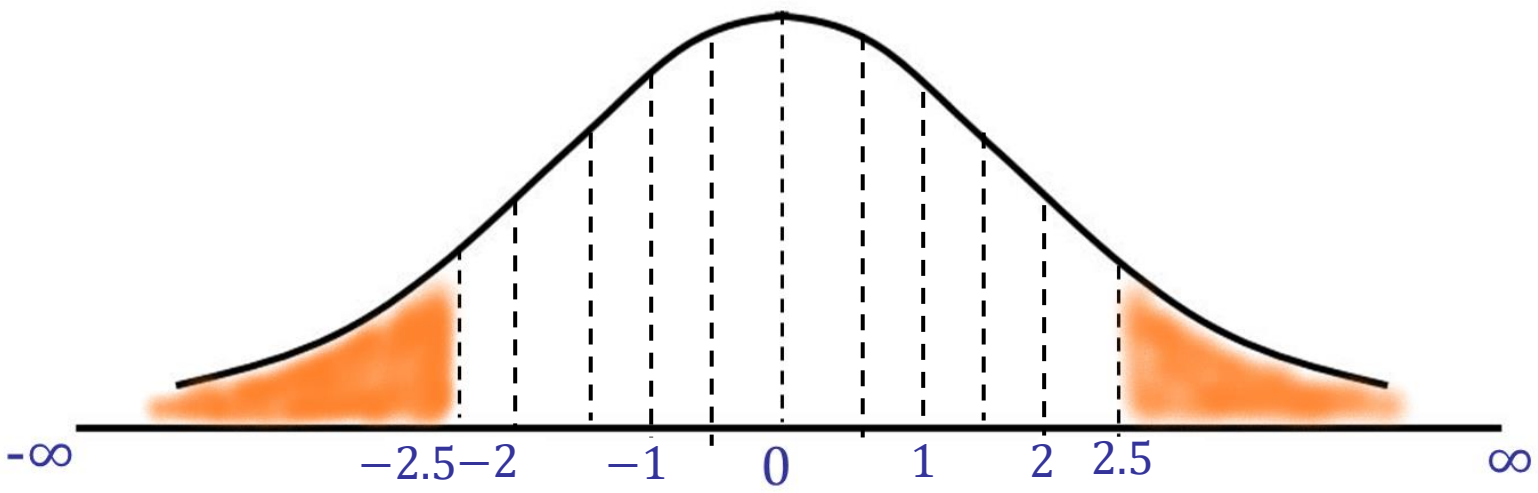
ההסתברות לתחום העליון תחושב כ: $1 - \Phi(2.5)$.



נבדוק כמה תצפיות צפויות ליפול בכל אינטרוול של התפלגות.

התפלגות נורמלית נעה מ- (∞-) עד ∞

כדי שההסתברות המצטברת בכל התחומים תהיה שווה ל-1,
אפשר להתחיל לצבור ממינוס אינסוף עד אינסוף.

ההסתברות לתחום התחתון תחושב כ: $\Phi(-2.5)$.

ההסתברות לתחום העליון תחושב כ: $1 - \Phi(2.5)$.