

# סטטיסטיקה היסקית למנע"ס 2025

## מבנה המבחן:

20 שאלות אמריקאיות, על חלקן צריך קודם להריץ פלט אקסל כדי לענות.

יש גם שאלות תיאורטיות.

לא מגישים את קובץ האקסל עימו עובדים במהלך המבחן.

נקבל קובץ אקסל עם נתונים בתחילת המבחן.

אנו מצפים ששאלה אחת תהיה באקסל על בלתי תלויים,

שאלה נוספת באקסל על תלויים,

שאלה נוספת באקסל בנושא רגרסיה.

יכולה להיות גם שאלה באקסל

רק על מבחן F לשוויון שונות.

**פרופורציה**

**וחי בריבוע-**

**ירדו מהחומר**

## לוחות (טבלאות) שנקבל במבחן:

טבלת Z

טבלת T

טבלת F (מספר טבלאות...עבור רמות מובהקות שונות)

# 20% Off

**רועי עידן**  
איתך בכל מבחן

## 20% קופון על כל האתר:

יסודות המימון- מועד ב

סטטיסטיקה היסקית

SQL- מועד ב

כלכלה למנהלים מאקרו

כלים מתמטיים ב

[www.roy-idan.co.il](http://www.roy-idan.co.il)

קוד הקופון הוא:

**SAVE20**

המלצות חמות:

לא לפתור שאלות פתוחות

לפתור רק שאלות אמריקאיות

לא לגעת בחוברת התרגילים

לפתור רק מבחנים

## דף הנוסחאות העדכני שיצורף למבחן:

מדדים:

ממוצע

$$\bar{X} = \sum_{i=1}^c \frac{x_i \cdot f_i}{n}$$

סטית תקן

$$\hat{S} = \sqrt{\hat{S}^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^c (x_i - \bar{X})^2 * f_i}{n-1}} = \sqrt{\frac{\sum_{i=1}^n x_i^2 - n\bar{X}^2}{n-1}}$$

תוחלת אחת שונות האוכלוסייה ידועה

דגימה מהתפלגות נורמלית:

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

$$Z_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim Z \quad [Z \sim N(0,1)]$$

מציאת גודל המדגם n:

$$n \geq \left( \frac{Z * \sigma}{\bar{x} - \mu} \right)^2$$

## בדיקת השערות על תוחלת אחת שונות האוכלוסייה ידועה

דגימה מהתפלגות נורמלית:

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

$$Z_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

## בדיקת השערות על תוחלת אחת שונות האוכלוסייה אינה ידועה

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

$$t_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\hat{s}}{\sqrt{n}}}$$

## בדיקת השערות על הפרש תוחלות, מדגמים בלתי תלויים, שונויות האוכלוסיות אינן ידועות ושוות

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right)$$

$$\sigma_X^2 = \sigma_Y^2 = \sigma^2$$

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \sigma^2 \cdot \left(\frac{1}{n_X} + \frac{1}{n_Y}\right)\right)$$

$$\bar{t}_{\bar{x}-\bar{y}} = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\hat{s}_p^2}{n_x} + \frac{\hat{s}_p^2}{n_y}}} = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\hat{s}_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t_{n_x+n_y-2}$$

$$\hat{s}_p^2 = \frac{(n_x-1) \cdot \hat{s}_x^2 + (n_y-1) \cdot \hat{s}_y^2}{n_y+n_x-2}$$

$$F = \frac{\hat{S}_x^2}{\hat{S}_y^2} \stackrel{H_0}{\sim} F_{n_x-1, n_y-1}$$

**בדיקת השערות על הפרש תוחלות במדגמים תלויים ומזווגים:**

$$\bar{d} \sim N(\mu_d, \frac{\sigma_d^2}{n})$$

$$t_{\bar{d}} = \frac{\bar{d} - \mu_d}{\frac{\hat{S}_d}{\sqrt{n}}} \sim t_{n-1}$$

$$\frac{\hat{S}_d}{\sqrt{n}} = \sqrt{\frac{\hat{S}_x^2}{n_x} + \frac{\hat{S}_y^2}{n_y} - 2 \cdot r_{xy} \cdot \frac{\hat{S}_x \cdot \hat{S}_y}{\sqrt{n_x} \cdot \sqrt{n_y}}} \quad (n_x = n_y = n)$$

**בדיקת השערות על מובהקות משוואת הקו באוכלוסייה**

$$r_P = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\hat{S}_x \cdot \hat{S}_y \cdot (n-1)}$$

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

$$r_{(x,y)} = \frac{cov(x, y)}{\hat{S}_x \hat{S}_y}$$

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} \cdot X_i$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}$$

$$\hat{\beta} = r \cdot \frac{\hat{S}_y}{\hat{S}_x}$$

$$SSREG = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = r^2 * SSy$$

$$SSRES = \sum_{i=1}^n (y_i - \hat{y})^2 = (1 - r^2) * SSy$$

$$SSy = \sum_{i=1}^n (y_i - \bar{y})^2 = \hat{S}_y^2 * (n - 1)$$

$$MSREG = \frac{r^2 \cdot SSy}{1}$$

$$MSRES = \frac{SSRES}{n - 2} = \frac{(1 - r^2) \cdot (n - 1) \cdot \hat{S}_y^2}{(n - 2)}$$

# רשימת הנושאים למבחן:

בדיקת השערות כשנתונה סטיית התקן של האוכלוסייה (נעבוד עם לוח Z)

בדיקת השערות כשנתונה סטיית התקן של המדגם S (נעבוד עם לוח T)

טעויות (אלפא, בטא) ועוצמת המבחן (נושא זה קשור ללוח Z בלבד)

שני מדגמים בלתי תלויים (לוח T): ללא שימוש באקסל

שני מדגמים בלתי תלויים (לוח T): הפעם נשתמש באקסל

מבחן F ידני (לא באקסל) לשוויון שוניות (שוניות האוכלוסייה) חד או דו"צ

מבחן F (כן באקסל) לשוויון שוניות (שוניות האוכלוסייה) חד או דו"צ

שני מדגמים תלויים ומזווגים (לוח T), לא באקסל

שני מדגמים תלויים ומזווגים (לוח T), **הפעם באקסל**

**רגרסיה באקסל**

**רגרסיה לא באקסל**

## **האם צריך להכיר את הנחות בכל מודל?**

כן כי יכולה להיות שאלה לגבי מה הנחות מודל מסוים.  
עם זאת, לא להתחיל לרשום ביוזמתנו הנחות במהלך  
פתרון השאלות!

רק אם יש תשובה שאומרת  
"לא עמדנו בהנחות  
ולכן לא ניתן להתקדם עם הפתרון ולחשב"

כדאי יהיה לבדוק באמת  
האם עמדנו בהנחות והאילוצים

# בדיקת השערות מדגם יחיד מבחן Z:

זיהוי הנושא:

תהיה נתונה סטיית התקן של האוכלוסייה או תהיה נתונה שונות האוכלוסייה.  
ידובר על מדגם יחיד שגודלו N, קבוצה אחת, אוכלוסייה אחת.

כך הנושא נראה בדף נוסחאות:

## בדיקת השערות על תוחלת אחת שונות האוכלוסייה ידועה

דגימה מהתפלגות נורמלית:

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$
$$Z_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

ש 1

ביה"ס למנהל עסקים מעוניין לבדוק האם ממוצע הציונים גבוה מ-72 במקצוע מסוים כאשר השונות היא 169.  
נדגמו 30 סטודנטים באופן מקרי והתקבל  $p.value = 0.1038$ .  
נתון שהציונים מתפלגים נורמלית.  
מהו ממוצע המדגם שהתקבל?

א. 75

ב. 73

ג. 78

ד. 69

פתרון:

נוציא שורש לשונות ונקבל שסטיית התקן היא 13.

נתונה בידינו סטיית התקן של האוכלוסיה.

עם שטח של 0.1038 נביא  $Z = 1.26$

השערות:

השערת האפס: ממוצע הציונים לא גבוה מ-72

$H_1$  ממוצע הציונים גבוה מ-72

$$Z_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$1.26 = \frac{\bar{x} - 72}{\frac{13}{\sqrt{30}}}$$

$$\bar{x} = 75$$

## ש 2: אנו עדיין באותו נושא של מבחן Z מדגם יחיד

חברת ביטוח מפרסמת שהיא עדיפה על פני כלל חברות הביטוח, כי משך הטיפול הממוצע בתביעה, עד לקבלת התשלום המגיע למבוטחיה, קצר יותר. הועלה חשד כי אין אמת בפרסום זה, וחוקר נתבקש להחליט. חוקר בדק מדגם מקרי של 40 תביעות מחברה זו, ומצא כי משך הטיפול הממוצע עד לתשלום הוא 24.2 ימים.

לכלל חברות הביטוח, משך הזמן הממוצע הדרוש לטיפול בתביעה עד לתשלום, הוא 26 ימים עם סטיית תקן של 8 ימים.

### שאלה מספר 1:

מהי מסקנת הבדיקה?

- א. לא נדחה את  $H_0$  ברמת מובהקות 0.05, משך הטיפול הממוצע בתביעה אינו קטן יותר ממשך הטיפול הממוצע של כלל חברות הביטוח
- ב. לא נדחה את  $H_0$  ברמת מובהקות 0.1, משך הטיפול הממוצע בתביעה אינו קטן יותר ממשך הטיפול הממוצע של כלל חברות הביטוח
- ג. נדחה את  $H_0$  ברמת מובהקות 0.01, משך הטיפול הממוצע בתביעה קטן יותר ממשך הטיפול הממוצע של כלל חברות הביטוח
- ד. נדחה את  $H_0$  ברמת מובהקות 0.1, משך הטיפול הממוצע בתביעה אינו קטן יותר ממשך הטיפול הממוצע של כלל חברות הביטוח.

פתרון:

$$40=N$$

$$\text{ממוצע מדגם} = 24.2$$

$$\text{ממוצע אוכלוסיה (תוחלת)} = 26$$

$$\text{סטיית תקן אוכלוסיה} = 8$$

$H_0$  משך הטיפול לא קצר יותר

$H_1$  : משך הטיפול קצר יותר

$$\mu_1 / \mu < 26$$

חוק:

אם רמת מובהקות  $PV < \alpha$  דוחים  $H_0$  ומקבלים את  $H_1$

אם רמת מובהקות  $PV > \alpha$  מקבלים  $H_0$  ודוחים את  $H_1$

$$Z_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$Z = \frac{24.2 - 26}{\frac{8}{\sqrt{40}}} \quad Z = 1.42 \rightarrow \text{PV} \quad 0.0778$$

חוק:

אם רמת מובהקות  $PV < \alpha$  דוחים  $H_0$  ומקבלים את  $H_1$

אם רמת מובהקות  $PV > \alpha$  מקבלים  $H_0$  ודוחים את  $H_1$

$PV$   
0.0778  
נקבל  
 $H_0$

$>$  ר"מ  
0.01

$PV$   
0.0778  
נקבל  
 $H_0$

$>$  ר"מ  
0.05

$PV$   
0.0778  
נדחה  
 $H_0$

$<$  ר"מ  
0.10

תשובה א נכונה

# בדיקת השערות מבחן T מדגם יחיד:

## זיהוי הנושא:

תהיה נתונה סטיית התקן של המדגם S או תהיה נתונה שונות המדגם S בריבוע.  
ידובר על מדגם יחיד שגודלו N, קבוצה אחת, אוכלוסייה אחת.

לעיתים תהא נתונה רשימה מספרית ממדגם

ואז נחשב בעזרתה את S סטיית התקן מהמדגם (וגם את ממוצע המדגם).

החישוב יקרה באקסל (ולא עם נוסחאות ידניות) ע"י הפונקציות:

=STDEV.S (                      )

=AVERAGE(                      )

כך הנושא נראה בדף נוסחאות:

## בדיקת השערות על תוחלת אחת שונות האוכלוסייה אינה ידועה

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

$$t_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\hat{s}}{\sqrt{n}}}$$

הציונים בקורס "מבוא למימון" מתפלגים נורמלית עם תוחלת 60.

חוקר טען שחלה עלייה גדולה בתוחלת הציונים בקורס "יסודות המימון" עקב המבחן החרפתי שהתקיים ב 21.7.25 בעקבות השינוי בתנאי הקבלה.

בדגימה מקרית של 29 סטודנטים התקבלו ממוצע 65 וסטיית תקן 7.9.

החוקר רוצה לבדוק את הטענה ברמת מובהקות 0.02. מכאן ש-

א. ניתן להסיק שחלה עליה בתוחלת הציונים כי  $p.value = 0.001 < 0.02$

ב. ניתן להסיק שחלה עליה בתוחלת הציונים כי  $p.value = 0.002 < 0.02$

ג. ניתן להסיק שלא חלה עליה בתוחלת הציונים כי  $p.value = 0.05 > 0.02$

ד. ניתן להסיק שחלה עליה בתוחלת הציונים כי  $p.value = 0.0005 < 0.02$

### פתרון

נתונה לנו סטיית התקן מהמדגם  $S = 7.9$

$H_0$  אין עלייה בציונים

$H_1$  יש עלייה בציונים

מציאת PV:

$$t = \frac{65 - 60}{\frac{7.9}{\sqrt{28}}} = 3.4$$

דרגות חופש  
 $28 = N - 1$

0.001  
PV <  $\alpha = 0.02$

דוחים  $H_0$  ומקבלים את  $H_1$

תשובה א נכונה

## טעויות ועוצמת המבחן 1-B : קשור רק ל Z ולא ל T:

בשאלות חישוביות עובדים עם לוח Z בלבד (לא עם לוח T),  
לכן גם סטיית התקן של האוכלוסייה תהיה נתונה בשאלה.

## 2 טעויות:

1. טעות מסוג ראשון אלפא.  
מתי אנו בטעות מסוג ראשון? אם החלטנו שדוחים  $H_0$  (מקבלים  $H_1$ )
2. טעות מסוג שני בטא.  
מתי אנו בטעות מסוג שני? אם החלטנו שדוחים  $H_1$  (מקבלים  $H_0$ )

### איך נזהה שמבקשים שנמצא את אחת הטעויות?

פשוט ישאלו "מה ההסתברות לטעות במסקנה או מה הסיכוי להכרעה שגויה?"  
או שישאלו אותנו למשל  
"מהי ההסתברות להסיק **שהתוחלת עלתה** כשבמצב האמיתי היא **לא עלתה**?"  
כלומר רומזים לנו שטעינו בהחלטה: שמסקנתנו לא מתאימה למצב האמיתי!  
או  
שישאלו אותנו "מה הסיכוי **לא לגלות** את מה שקרה בפועל במציאות?"  
ואז אם לא גילינו...זה אומר שטעינו...

## עוצמת המבחן 1-B

עוצמת המבחן היא לא טעות!  
להפך: היא מספרת שצדקנו במסקנה אליה הגענו.  
מבחינה מספרית היא תמיד מספר ענק סביב ה 90%...

### איך נזהה שמבקשים את עוצמת המבחן?

המקרה הקל: פשוט ישאלו אותנו מהי עוצמת המבחן

או

שישאלו אותנו "מהי ההסתברות להיות **צודקים**"

או

שישאלו אותנו "מה הסיכוי **לגלות** את מה שאכן קרה בפועל במציאות?"

או

שישאלו אותנו "מהי ההסתברות להסיק **שהתוחלת עלתה** כשבמצב האמיתי היא **עלתה**?"

או

שישאלו אותנו "מהי ההסתברות להסיק **שהתוחלת ירדה** כשבמצב האמיתי היא **ירדה**?"

חברת ביטוח מפרסמת שהיא עדיפה על פני כלל חברות הביטוח, כי משך הטיפול הממוצע בתביעה, עד לקבלת התשלום המגיע למבוטחיה, קצר יותר. הועלה חשד כי אין אמת בפרסום זה, וחוקר נתבקש להחליט. חוקר בדק מדגם מקרי של 40 תביעות מחברה זו, ומצא כי משך הטיפול הממוצע עד לתשלום הוא 24.2 ימים.

לכלל חברות הביטוח, משך הזמן הממוצע הדרוש לטיפול בתביעה עד לתשלום, הוא 26 ימים עם סטיית תקן של 8 ימים.

מהי ההסתברות לטעות בהחלטת החוקר, אם אכן משך זמן הטיפול הממוצע של אותה חברה הוא לכל היותר 22 ימים, וההשערה נבדקה ברמת מובהקות 0.05? בחרו תשובה קרובה ביותר.

א. תיתכן טעות מסוג שני שהסתברותה  $\beta = 0.0655$

ב. תיתכן טעות מסוג שני שהסתברותה  $\alpha = 0.0655$

ג. תיתכן טעות מסוג ראשון שהסתברותה  $\alpha = 0.05$

ד. תיתכן טעות מסוג שני והסתברותה היא עוצמת המבחן 0.9655

### חוק:

אם קיבלנו תוחלת חדשה M חדש זה רק כדי שנמצא בעזרתה את בטא !

### השערות:

$H_0$ : משך הטיפול אינו קצר יותר

$H_1$ : המשך קצר יותר

$$N = 40$$

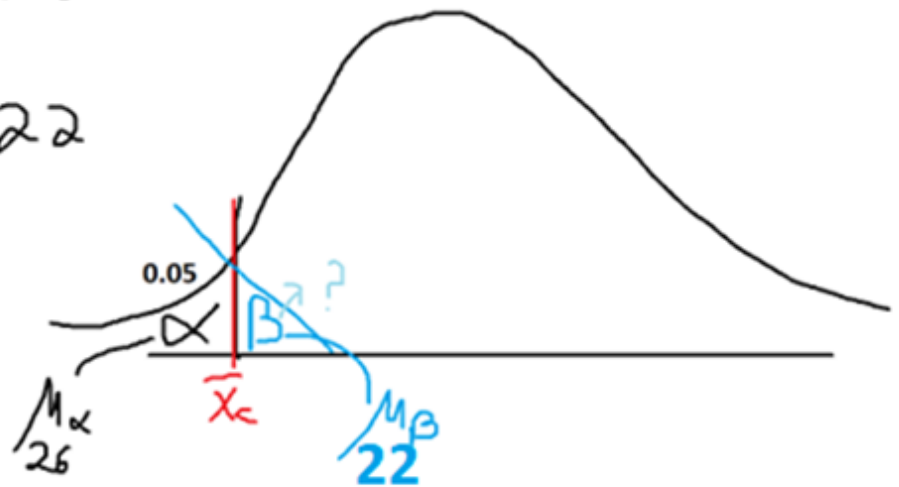
$$H_0: \mu \geq 26$$

$$\bar{X} = 24.2$$

$$H_1: \mu < 26$$

$$\mu_{\alpha} = 26$$

$$\mu_{\beta} = 22$$



שלב א הוא תמיד למצוא ממוצע קריטי

עם שטח אלפא של 0.05 מבפנים לבחון נביא  $Z = -1.64$   
הוספנו -

$$-1.64 = \frac{\bar{X}_c - \mu_{\alpha}}{\frac{s}{\sqrt{40}}}$$

$$\bar{X}_c = 23.92$$

שלב ב הוא כבר מציאת הבטא:

$$Z_{\beta} = \frac{23.92 - \mu_{\beta}}{\frac{s}{\sqrt{40}}}$$

$$Z_{\beta} = 1.51 \quad 0.0655$$

ממוצע הזמן היומי שאדם במדינה צופה בחדשות בטלוויזיה, הוא 47 דקות וסטיית תקן 12 דקות. עבור מדגם צופים בגודל 49, ההסתברות לטעון שבתקופת המלחמה, תוחלת זמן הצפייה היומי בחדשות עלתה ל- 52 דקות בממוצע ולהיות צודקים, שווה 0.9838. מהי ההסתברות של החוקר לטעות במסקנה שחלה עליה בזמן הצפייה הממוצע בחדשות? (בחרו את התשובה הקרובה ביותר).

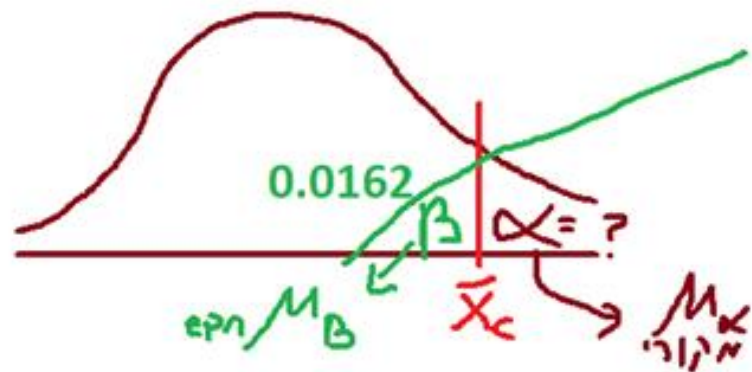
$$\alpha = ?$$

- א. 0.2177
- ב. 0.0162
- ג. 0.9838
- ד. 0.1

$$1 - \beta = 0.9838$$

$$0.0162 = \beta$$

H1 זמן הצפייה עלה



שלב א: מציאת ממוצע קריטי

בעזרת שטח בטא של 0.0162 נביא  $Z = -2.14$

$$-2.14 = \frac{\bar{X}_c - 52}{\frac{12}{\sqrt{49}}}$$

$$\bar{X}_c = 48.33$$

רק עתה נמצא את האלפא (שלב ב)

$$Z_{\alpha} = \frac{48.33 - \overset{M_{\alpha}}{47}}{\frac{12}{\sqrt{49}}}$$

$$Z_{\alpha} = 0.78$$

אלפא 21.77%

## **עוצמת המבחן וממוצע קריטי, מבחן Z מדגם יחיד:**

ש 5:

למשקל של תפוח יש תוחלת 260 גרם וס"ת תקן 38 גרם.

אגרונם המתמחה בגידול תפוחים טוען ששיטת השקיה חדשה שפיתח מעלה את משקל הממוצע של תפוח ל-275 גרם.

לבדיקה טענתו נלקח מדגם מקרי של 60 תפוחים אשר הושקו בשיטת ההשקיה החדשה.

הוחלט ששיטת ההשקיה החדשה תיכנס לשימוש אם משקל התפוחים הממוצע במדגם יהיה מעל 267 גרם.

לבסוף הוחלט להכניס לשימוש את שיטת ההשקיה החדשה

רק אם משקל התפוחים הממוצע במדגם יהיה מעל 273 גרם.

מה יקרה לעוצמת המבחן הסטטיסטי?

- א. עוצמת המבחן תקטן
- ב. עוצמת המבחן תגדל
- ג. חסרים נתונים כדי לדעת
- ד. העוצמה לא תשתנה

## שאלת הבנה חשובה ש5:

למשקל של תפוח יש תוחלת 260 גרם וס"ת תקן 38 גרם.

אגרונום המתמחה בגידול תפוחים טוען ששיטת השקיה חדשה שפיתח מעלה את משקל הממוצע של תפוח ל-275 גרם.



לבדיקה טענתו נלקח מדגם מקרי של 60 תפוחים אשר הושקו בשיטת ההשקיה החדשה.

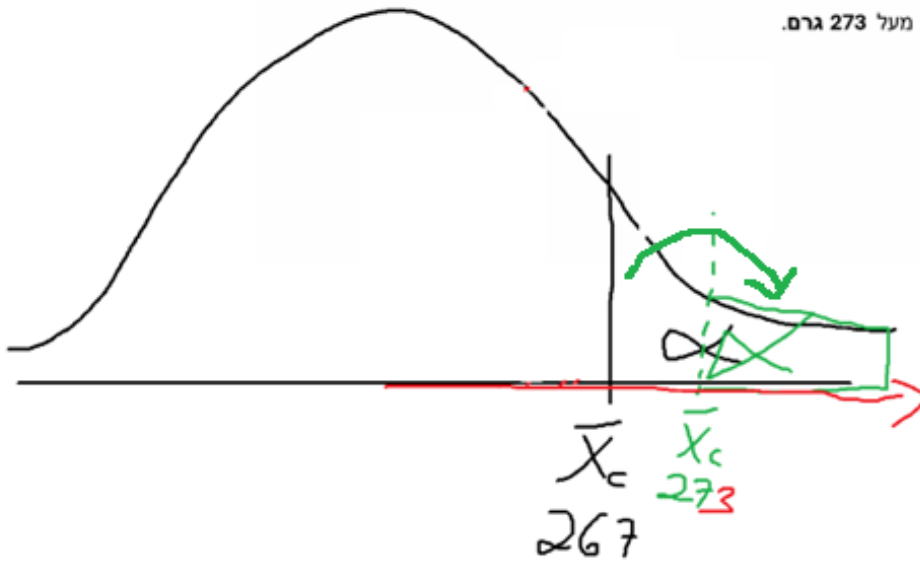
הוחלט ששיטת ההשקיה החדשה תיכנס לשימוש אם משקל התפוחים הממוצע במדגם יהיה מעל 267 גרם.

**ממוצע קריטי**

לבסוף הוחלט להכניס לשימוש את שיטת ההשקיה החדשה

רק אם משקל התפוחים הממוצע במדגם יהיה מעל 273 גרם.

מה יקרה לעוצמת המבחן הסטטיסטי?

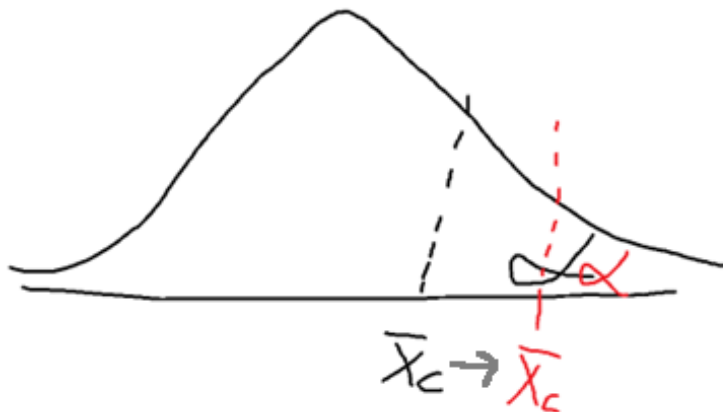


- ✓ א. עוצמת המבחן תקטן
- ב. עוצמת המבחן תגדל
- ג. חסרים נתונים כדי לדעת
- ד. העוצמה לא תשתנה

$\alpha \downarrow$

$\beta \uparrow$

$(1 - \beta) \downarrow$



**חוק:**

$\alpha \uparrow$   $\bar{x}_c \downarrow$

$\alpha \downarrow$   $\bar{x}_c$

חוק:

אם רמת המובהקות תקטן אז הממוצע הקריטי יגדל

להפך:

אם רמת המובהקות תגדל אז הממוצע הקריטי יקטן

אילו גורמים יכולים להגדיל את עוצמת המבחן  $\beta$ -1?

טעות מסוג שני תקטן

טעות מסוג ראשון תגדל

הגדלת גודל מדגם (כמות הנחקרים)

להקטין את סטיית התקן באוכלוסיה

להקטין את השונות באוכלוסיה

להגדיל את גודל האפקט

## מבחן T לשני מדגמים בלתי תלויים: לא באקסל!

בנושא זה יש שתי קבוצות שלא קשורות אחת לשנייה.

האנשים בקבוצה הראשונה לא מכירים את האנשים בקבוצה השנייה.

כשהיינו בנושא זה באקסל, לא ידענו האם השונויות שוות או שונות ולכן בדקנו זאת בעזרת מבחן F באקסל.

הפעם לא עובדים עם האקסל (מפעילים נוסחאות) ומניחים תמיד שהשונויות באוכלוסיות שוות.

לשמחתנו בדף הנוסחאות מזכירים לנו זאת בכותרת הנושא.

### כך נראה הנושא בדף הנוסחאות שנקבל:

בדיקת השערות על הפרש תוחלות, מדגמים בלתי תלויים,

שונויות האוכלוסיות אינן ידועות ושוות

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right)$$

$$\sigma_X^2 = \sigma_Y^2 = \sigma^2$$

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \sigma^2 \cdot \left(\frac{1}{n_X} + \frac{1}{n_Y}\right)\right)$$

$$\bar{t}_{\bar{x}-\bar{y}} = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\hat{s}_p^2}{n_x} + \frac{\hat{s}_p^2}{n_y}}} = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\hat{s}_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t_{n_x+n_y-2}$$
$$\hat{s}_p^2 = \frac{(n_x-1) \cdot \hat{s}_x^2 + (n_y-1) \cdot \hat{s}_y^2}{n_y+n_x-2}$$

חוקר דגם בדגימה מקרית, מבית חולים גדול בארץ, 19 עובדים זוטרים ו- 22 עובדים בכירים, ומדד את "החריגות": מספר השעות שעבד כל עובד מעבר לשעות המוקצבות לו בחוק.

המשתנה "חריגות" מתפלג נורמלית, והשונות של התפלגות החריגות בקרב הזוטרים והבכירים הינן שוות.

בקרב העובדים הזוטרים, ממוצע החריגות החודשי היה 65 שעות והאומדן לסטיית התקן הוא 12.

בקרב העובדים הבכירים, ממוצע החריגות היה 79 שעות והאומדן לסטיית התקן הוא 14.

החוקר רצה לבדוק אם החריגות בקרב העובדים הבכירים גדולות משמעותית

מאלו בקרב העובדים הזוטרים, ברמת מובהקות של 0.05.

סטטיסטי המבחן לבדיקת השערה זו הינו:

א. 3.408

ב. 1.684

ג. 2.021

ד. 12.361

| זוטרים         | בכירים         |
|----------------|----------------|
| $\bar{y} = 65$ | $\bar{x} = 79$ |
| $s_y = 12$     | $s_x = 14$     |
| $n_y = 19$     | $n_x = 22$     |

$$\bar{t}_{\bar{x}-\bar{y}} = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\hat{s}_p^2}{n_x} + \frac{\hat{s}_p^2}{n_y}}} =$$

$$\hat{s}_p^2 = \frac{(n_x-1) \cdot \hat{s}_x^2 + (n_y-1) \cdot \hat{s}_y^2}{n_y + n_x - 2}$$

השונות המשוקללת

$$\hat{s}_p^2 = 172$$

$$\bar{t}_{\bar{x}-\bar{y}} = \frac{(\bar{x} - \bar{y}) - \overbrace{(\mu_x - \mu_y)}^0}{\sqrt{\frac{\hat{s}_p^2}{n_x} + \frac{\hat{s}_p^2}{n_y}}} =$$

$$t = 3.408$$

סטטיסטי

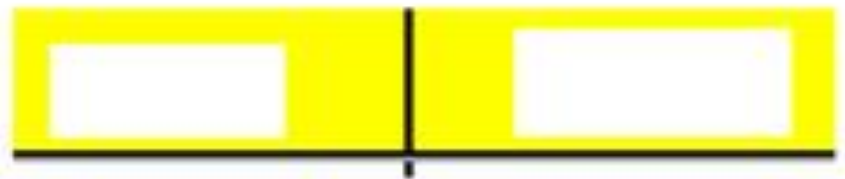
תשובה א נכונה

סעיף חדש, מהו PV ?

למציאת PV תמיד נפעיל את נוסחת הסטטיסטי של אותו נושא

$$t = 3.408$$

$$\text{דרגות חופש} = 22 + 19 - 2 = 39$$



0.0005

$PV > 1$

0.001

חוק

אם ההשערה דו"צ יהיה עלינו להכפיל את שני השטחים ורק כך נמצא את PV

אז אם אצלינו היו שואלים לגבי הבדל (שינוי, פער, השפעה) בין שתי הקבוצות היה עלינו להכפיל את שני השטחים כך ש PV הסופית היתה נראית:

$$0.001 < PV < 0.002$$

## מבחן T שני מדגמים בלתי תלויים: הפעם באקסל:

בנושא זה יש שתי קבוצות שלא קשורות אחת לשנייה.  
האנשים בקבוצה הראשונה לא מכירים את האנשים בקבוצה השנייה.  
השאלה כבר תפנה אותנו לעבוד עם האקסל.  
בנושא זה- אנו חייבים לזכור, גם אם לא ביקשו זאת בשאלה-  
שהפעולה הראשונה היא לבצע תמיד מבחן F באקסל עבור השונויות:  
F-Test Two-Sample for Variances

המטרה היא לבדוק באיזה מצב אנו נמצאים:

שונויות שוות equal Variances      או      שונויות שונות Unequal Variances

כך תמיד ייראו ההשערות (תמיד דו צדדיות בנושא זה):

H0 השונויות שוות equal Variances

H1 השונויות שונות Unequal Variances

הפלט ייתן לנו PV חד צדדי בלבד PV ONE TAIL

לכן נכפילו פי 2 ונקבל PV דו צדדי, אותו נשווה מול רמת המובהקות ונגלה האם השונויות שוות או שונות.  
נמחיש זאת בשאלה מספר 15 (יש לפתוח קובץ אקסל עם הנתונים שקיבלנו):

חוקר רצה לבדוק האם קיים פער השווה ל- 30 קוט"ש בצריכת החשמל בחודש ינואר, בין תושבי ירושלים (x) לבין תושבי תל-אביב (y).

בחודש ינואר האחרון, החוקר בדק מדגמים מקריים משתי הערים.

התוצאות מצורפות בקובץ האקסל, בלשונית "חשמל".

מה תהיה מסקנתו של החוקר לגבי שוויון שוניות צריכת החשמל ברמת מובהקות 0.02?

- א. שוניות צריכת החשמל שוות בשתי הערים
- ב. שוניות צריכת החשמל שונות בשתי הערים
- ג. המדגמים תלויים ומזווגים- במבחן הזה אין צורך לבדוק שיווין שוניות
- ד. חסרים נתונים לגבי דרגות החופש כדי לדעת מהי המסקנה

מה תהיה מסקנתו של החוקר, לגבי הפער (המופיע בפתיח) בצריכת החשמל בין תושבי הערים, ברמת מובהקות 0.02?

- א.  $p. value = 0.398$ , ולכן החוקר לא ידחה את  $H_0$
- ב.  $p. value = 0.199$ , ולכן החוקר ידחה את  $H_0$
- ג.  $p. value = 0.01$ , ולכן החוקר ידחה את  $H_0$
- ד.  $p. value = 0.199$ , ולכן החוקר לא ידחה את  $H_0$

הפתרון המלא מופיע בגיליון האקסל בלשונית "חשמל"

## **מבחן T לשני מדגמים תלויים ומזווגים, באקסל:**

שימו לב: דיבור על "שונויות שוות או שונויות שונות" לא רלבנטי בכלל בנושא זה!

זיהוי: יאמרו בשאלה שהפרשים מתפלגים נורמלית

מדובר על 2 קבוצות כאשר לכל אדם בקבוצה א יהיה את הפרטנר להשוואה שמתאים רק לו...שקשור רק אליו.....בקבוצה ב.

למשל:

קבוצת אבות מול קבוצת בניהם

קבוצת בעלים מול קבוצת נשותיהם

דוגמא קלאסית נוספת:

מדובר באותם האנשים אך פעם לפני דיאטה ופעם אחרי דיאטה שערך

או

שמדובר באותם אנשים אך פעם בלי תרופה ופעם עם תרופה

חוקר דגם 8 מפעלים, ורשם את מספר המועסקים במפעל בתחילת שנת 2023 ובסוף אותה השנה.

התוצאות מצורפות בקובץ האקסל, בלשונית "מועסקים".

בהנחה שהפרשי מספר המועסקים מתפלגים נורמלית, מהי מסקנת החוקר עבור רמת מובהקות 0.025?

- א.  $p. value < \alpha$ , החוקר יסיק כי חלה ירידה במספר המועסקים ב-2023
- ב.  $p. value < \alpha$ , החוקר יסיק כי לא חלה ירידה במספר המועסקים ב-2023
- ג.  $p. value > \alpha$ , החוקר יסיק כי חלה ירידה במספר המועסקים ב-2023
- ד.  $p. value > \alpha$ , החוקר יסיק כי לא חלה ירידה במספר המועסקים ב-2023

| סוף 23 | תחילת 23 |     |
|--------|----------|-----|
|        |          | דני |
|        |          | משה |
|        |          |     |

הפתרון המלא מופיע בגיליון האקסל בלשונית "חשמל"

## שני מדגמים תלויים ומזווגים (לוח T), לא באקסל:

שימו לב: דיבור על "שונויות שוות או שונויות שונות" לא רלבנטי בכלל בנושא זה!

(שונויות שוות או שונויות שונות קשור רק למבחן T בשני מדגמים בלתי תלויים)

זיהוי: יאמרו בשאלה שהפרשים מתפלגים נורמלית

כך הנושא נראה בדף הנוסחאות:

בדיקת השערות על הפרש תוחלות במדגמים תלויים ומזווגים:

$$\bar{d} \sim N(\mu_d, \frac{\sigma_d^2}{n})$$

$$t_{\bar{d}} = \frac{\bar{d} - \mu_d}{\frac{\hat{S}_d}{\sqrt{n}}} \sim t_{n-1}$$

$$\frac{\hat{S}_d}{\sqrt{n}} = \sqrt{\frac{\hat{S}_x^2}{n_x} + \frac{\hat{S}_y^2}{n_y} - 2 \cdot r_{xy} \cdot \frac{\hat{S}_x \cdot \hat{S}_y}{\sqrt{n_x} \cdot \sqrt{n_y}}} \quad (n_x = n_y = n)$$

נבדקו הבדלים בין שתי שיטות בדיקה להערכת נתחי שוק של שישה מותגי בירה בחנות בחודש מסוים. שיטה ראשונה מסורתית של ספירת מלאים, ושיטה שנייה חדשנית באמצעות מערכת אוטומטית שמנטרת מכירות. המשתנה הנבדק "הערכת נתח שוק" מתפלג נורמלית, ודגימת מותגי הבירה הייתה מקרית. האם יש הבדל מובהק בהערכת נתח שוק של ששת מותגי הבירה לפי המלאים, באמצעות שתי שיטות הבדיקה?

התקבלו הממצאים הבאים:

$$\bar{d} = 0.234$$

$$\sum_{i=1}^6 (d_i - \bar{d})^2 = 8.237$$

- א.  $P. value > 0.2$ , ולכן אין הבדל מובהק בין שתי השיטות להערכת נתח השוק ברמת מובהקות 0.1
- ב.  $P. value < 0.1$ , ולכן אין הבדל מובהק בין שתי שיטות להערכת נתח השוק ברמת מובהקות 0.1
- ג.  $P. value = 0.44$ , ולכן אין הבדל מובהק בין שתי השיטות להערכת נתח השוק ברמת מובהקות 0.05
- ד. חסרים נתונים על שוויון שונויות נתחי השוק בשתי שיטות הערכה כדי להסיק מסקנה

$$t_d = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} \sim t_{n-1}$$

$$\hat{S} = \sqrt{\hat{S}^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}}$$

$$S = \sqrt{\frac{8.23}{6-1}}$$

$$S_d = 1.28$$

3/7

H0 אין הבדל

H1 יש הבדל

$$t_d = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

$$t = \frac{0.234 - 0}{\frac{1.28}{\sqrt{6}}}$$

$$t = 0.44$$

$$S = 1.28$$

$$P V > 0.2$$

$$\boxed{>}$$

$$0.1$$

$$0.1$$

תשובה א נכונה

מקבלים H0 ודוחים את H1

$$P V > 0.1$$

$$P V > 0.2$$

## מבחן F ידני (הפעם לא באקסל) לשונויות האוכלוסיות:

בנושא הבלתי תלויים באקסל למדנו שחובה תמיד כבר בהתחלה לבצע באקסל מבחן F

ולגלות האם השונויות שוות או שונות, כלומר Unequal Variances **א** equal Variances

בנושא מבחן F ידני לא עובדים עם האקסל אך עדיין רוצים לגלות האם השונויות שוות או שונות.

פשוט ישאלו אותנו האם השונויות שוות או שונות והשאלה לא תפנה אותנו לאקסל!

נעבוד כאן עם לוח F שנקבל בבחינה.

כך הנושא נראה בדף נוסחאות שנקבל:

## בדיקת השערות על יחס בין שתי שונויות

$$F = \frac{\hat{S}_x^2}{\hat{S}_y^2} \stackrel{H_0}{\sim} F_{n_x-1, n_y-1}$$

1. חוקר רצה לבדוק האם פיזור זמני הצפייה בשעות בתוכנית "האח הגדול LIVE", גדול יותר מפיזור זמני הצפייה בשעות ב – "מהדורה המרכזית". זמן הצפייה מתפלג נורמלית בשתי התוכניות. לשם בדיקת ההשערה דגם מדגמים מקריים של צופים, ולהלן התוצאות שהתקבלו. זמן הצפייה מתפלג נורמלית בשתי התוכניות.
- עבור תכנית "האח הגדול":  $\hat{S}_x = 3.25$ ,  $\bar{x} = 6.5$ ,  $n=17$
  - עבור תכנית "המהדורה המרכזית":  $\hat{S}_y = 1.9$ ,  $\bar{y} = 7.5$ ,  $n=12$
- ניתן להסיק כי:
- א. פיזור זמני הצפייה בשעות בתוכנית "האח הגדול LIVE", גדול יותר מפיזור זמן הצפייה בשעות ב "מהדורה המרכזית" ברמת מובהקות 0.05
- ב. פיזור זמני הצפייה בשעות בתוכנית "האח הגדול LIVE", אינו גדול יותר מפיזור זמני הצפייה בשעות ב "מהדורה המרכזית" ברמת מובהקות 0.05
- ג. פיזור זמני הצפייה בשעות בתוכנית "האח הגדול LIVE", גדול יותר מפיזור זמני הצפייה בשעות ב "מהדורה המרכזית" ברמת מובהקות 0.01
- ד. הנחת שיוון השונויות לא מתקיימת, ולכן לא ניתן לענות על שאלה זו

בדיקת השערות על יחס בין שתי שונות

$$F = \frac{\hat{S}_x^2}{\hat{S}_y^2} \stackrel{H_0}{\sim} F_{n_x-1, n_y-1}$$

חוק

המילים פיזור או סטיית תקן או שונות מעידות שאנו במבחן F לבדיקת שוויון שונות

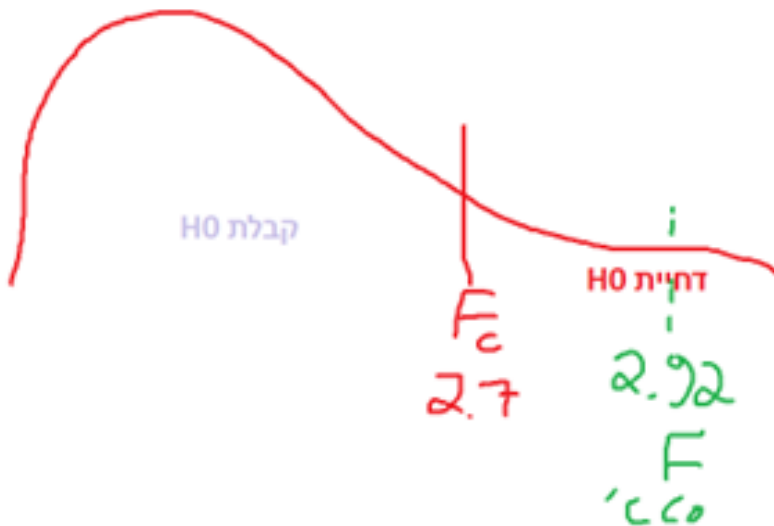
$$F_{\text{calculated}} = \frac{3.25^2}{1.9^2}$$

$$F_{\text{calculated}} = 2.92 \quad \square \quad F_c$$

$$n_x - 1 = 16$$

$$n_y - 1 = 11$$

$$F_{\text{calculated}} = 2.92 > 2.7$$



דוחים H0  
מקבלים H1

H0 פיזור האח לא גדול מבמהדורה  
H1 פיזור האח גדול מבמהדורה ✓

בנושא זה בודקים האם קיים קשר בין שני משתנים כמותיים X ל Y.  
למשל האם יש קשר בין גובה למשקל.  
בנושא זה בודקים האם X מנבא את Y (ניבוי המשתנה Y בעזרת המשתנה X).  
X הוא המשתנה המנבא את Y (המשתנה המנובא).

$$\hat{y} = \alpha + \beta \cdot X$$

## זיהוי הנושא:

לנושא זה מושגים ייחודים שלא יתאימו לאף נושא אחר כגון:  
מתאם קווי r פירסון (הקשר הקווי בין X ל Y)  
ניבוי של המשתנה המושפע Y (משוואה לניבוי Y באמצעות המשתנה X).

פרופורציית שונות מוסברת

פרופורציית שונות לא מוסברת

SSY SSRES SSREG

MSRES MSREG

השיפוע B

שונות הטעויות MSRES

כך הנושא נראה בדף נוסחאות:

## בדיקת השערות על מובהקות משוואת הקו באוכלוסייה

$$r_P = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\hat{S}_x \cdot \hat{S}_y \cdot (n-1)}$$

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{(n-1)}$$

$$r_{(x,y)} = \frac{\text{cov}(x, y)}{\hat{S}_x \hat{S}_y}$$

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} \cdot X_i$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}$$

$$\hat{\beta} = r \cdot \frac{\hat{S}_y}{\hat{S}_x}$$

$$SSREG = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = r^2 * SSy$$

$$SSRES = \sum_{i=1}^n (y_i - \hat{y})^2 = (1 - r^2) * SSy$$

$$SSy = \sum_{i=1}^n (y_i - \bar{y})^2 = \hat{s}_y^2 * (n - 1)$$

$$MSREG = \frac{r^2 \cdot SSy}{1}$$

$$MSRES = \frac{SSRES}{n - 2} = \frac{(1 - r^2) \cdot (n - 1) \cdot \hat{s}_y^2}{(n - 2)}$$

$$F = \frac{MSREG}{MSRES} = \frac{r^2 \cdot (n - 2)}{1 - r^2} \stackrel{H_0}{\sim} F_{1, n-2}$$

1. יש לפתוח את גיליון "שכירים" מקובץ האקסל. מהי פרופורציית השונות המוסברת של מספר החדרים בדירת המגורים לפי גיל העובדים השכירים?

- א. 11.8% מההבדלים במספר חדרים בדירה, ניתנים להסבר באמצעות גיל העובד
- ב. 34.4% מההבדלים במספר חדרים בדירה, ניתנים להסבר באמצעות גיל העובד
- ג. 88.2% מההבדלים במספר חדרים בדירה, ניתנים להסבר באמצעות גיל העובד
- ד. 54.56% מההבדלים במספר חדרים בדירה, ניתנים להסבר באמצעות גיל העובד

## הפתרון בגיליון האקסל המתאים

2.

מהו האומדן לשונות הטעויות במודל?

- א. 1.145
- ב. 6.43
- ג. 48.125
- ד. 54.56

## הפתרון בגיליון האקסל המתאים

3.

כמה חדרים תנבאו לשכירים בגיל 45 (יש לבחור תשובה קרובה ביותר), והאם הניבוי יהיה מדויק?

- א. 4 חדרים, הניבוי לא מדויק בשל המתאם הקווי הנמוך יחסית
- ב. 5 חדרים, הניבוי לא מדויק בשל המתאם הקווי הנמוך יחסית
- ג. 4 חדרים, הניבוי מדויק בשל המתאם הקווי הגבוה בין המשתנים
- ד. 3 חדרים, הניבוי מדויק בשל המתאם הקווי הגבוה בין המשתנים

## הפתרון בגיליון האקסל המתאים

10. על מנת לבדוק את השפעת ההשכלה (מספר שנות לימוד) על השכר החודשי, נדגמו 20 עובדים מהמגזר הפרטי בדגימה מקרית, ועבור כל אחד מהם נרשמו שנות ההשכלה והשכר (גיליון "שכר" בקובץ האקסל). האם ניתן להסיק שמספר שנות ההשכלה משפיע על השכר ברמת מובהקות 0.03?

- א. ניתן להסיק שיש קשר חיובי מובהק בין המשתנים כי השיפוע חיובי ו-  $0.029 < 0.03$
- ב. ניתן להסיק שהשיפוע לא מובהק
- ג. ניתן להסיק שהמתאם לא מובהק
- ד. ניתן להסיק שהמשוואה לא מובהקת

## הפתרון בגיליון האקסל המתאים

11. מהו השכר המנובא עבור עובד עם 19 שנות השכלה?

- א. 12,172
- ב. 13,146
- ג. 10,760
- ד. 11,953

## הפתרון בגיליון האקסל המתאים

# רגרסיה לא באקסל- תרגול עצמי

שאלה 10

כדי לבדוק האם יש קשר בין עישון למשקל, דגם החוקר מדגם מקרי של 61 מעשנים, ורשם עבור על אר מהם את מספר הסיגריות שמעשן ביום (x), ואת המשקל שלו בק"ג (y). החוקר מצא :

$$\hat{s}_y^2 = 167 \quad r = -0.3 \quad \hat{y}_i = 80 - 0.5x_i$$

מהו האומדן לשונות הטעויות במודל?

א. 154.54

ב. 0.91

ג. 166.66

ד. חסרים נתונים כדי לחשב

אומדן לשונות הטעויות  $MSRES =$

פשוט מציבים את הנתונים (יש לנו הכל)

ומגלים שתשובה א נכונה.

$$MSRES = \frac{(1 - r^2) \cdot (n - 1) \cdot \hat{s}_y^2}{(n - 2)}$$