

שיטות מחקר תרגיל 8

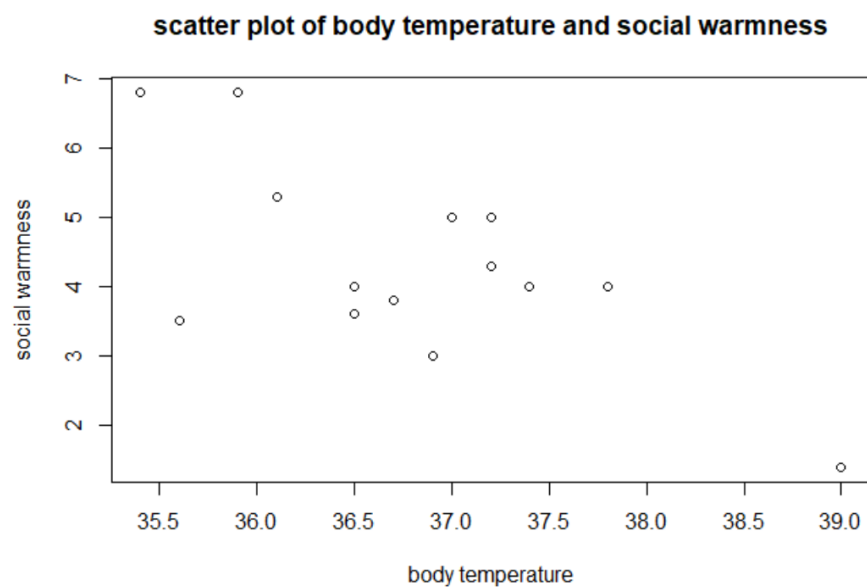
מגישות:

נועה בן דרור 316163260

מיכל דגן 315657064

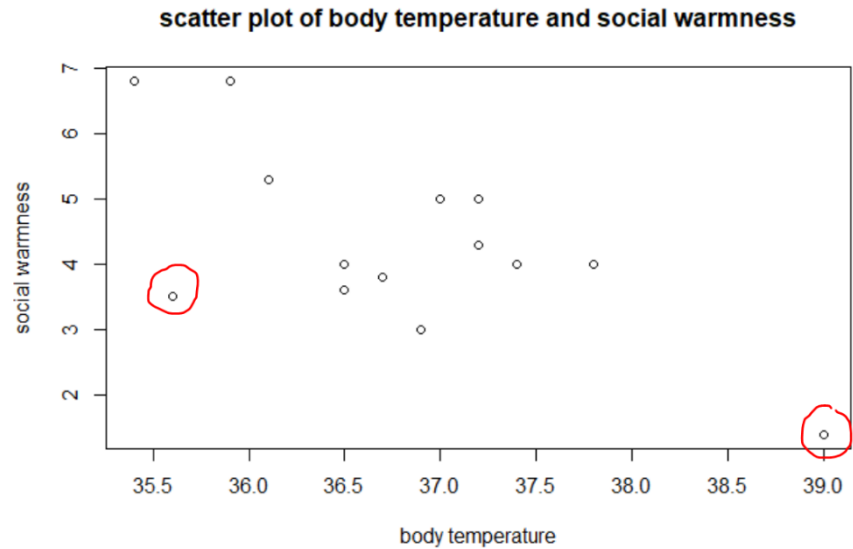
שאלה 1

א.



נוכל להבחין כי ככל הנראה קיים קשר שלילי בין טמפרטורת גוף לבין "חמימות חברתית". כלומר, ככל שטמפרטורת הגוף של אדם כלשהו נמוכה, כך "החמימות החברתית" שלו גבוהה.

ניתן לזהות תצפיות חריגות (מסומנות באדום). התצפית החריגה מימין אומרת שיש אדם בעל טמפרטורת גוף גבוהה מאוד, ו"חמימות חברתית" נמוכה מאוד. התצפית החריגה משמאל אומרת שיש אדם בעל טמפרטורת גוף נמוכה, ו"חמימות חברתית" גבוהה.



```
> R_2
[1] 0.4562862
```

ב.

אחוז השונות המוסברת הוא 45%. המשמעות של המספר הזה היא ש45% מ"החמימות החברתית" של אדם כלשהו, נוכל להסביר על ידי טמפרטורת הגוף שלו.

ג. קיבלנו כי מקדם מתאם פירסון הוא -0.6754896.

```
Pearson's product-moment correlation
data: body_temp_x and social_warmness_y
t = -3.1734, df = 12, p-value = 0.008019
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.8878594 -0.2258577
sample estimates:
cor
-0.6754896
```

ד. p-value שהתקבל:

```
> corr$p.value
[1] 0.008018923
```

קיבלנו ערך קטן מ-0.05 (אלפא) ולכן נוכל לדחות את השערת האפס (לפיה אין קורלציה בין "חמימות חברתית" לבין טמפרטורת גוף).

ה.

```

~ "
> regression(body_temp_x, social_warmness_y, 1)
[1] 40.90241
~

```

קיבלנו 40.90241.

1.

```

> linear1 = lm(social_warmness_y ~ body_temp_x)
> summary(linear1)

```

Call:

```
lm(formula = social_warmness_y ~ body_temp_x)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-2.0476	-0.6621	0.2775	0.8373	1.5589

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	41.924	11.853	3.537	0.00409 **
body_temp_x	-1.022	0.322	-3.173	0.00802 **

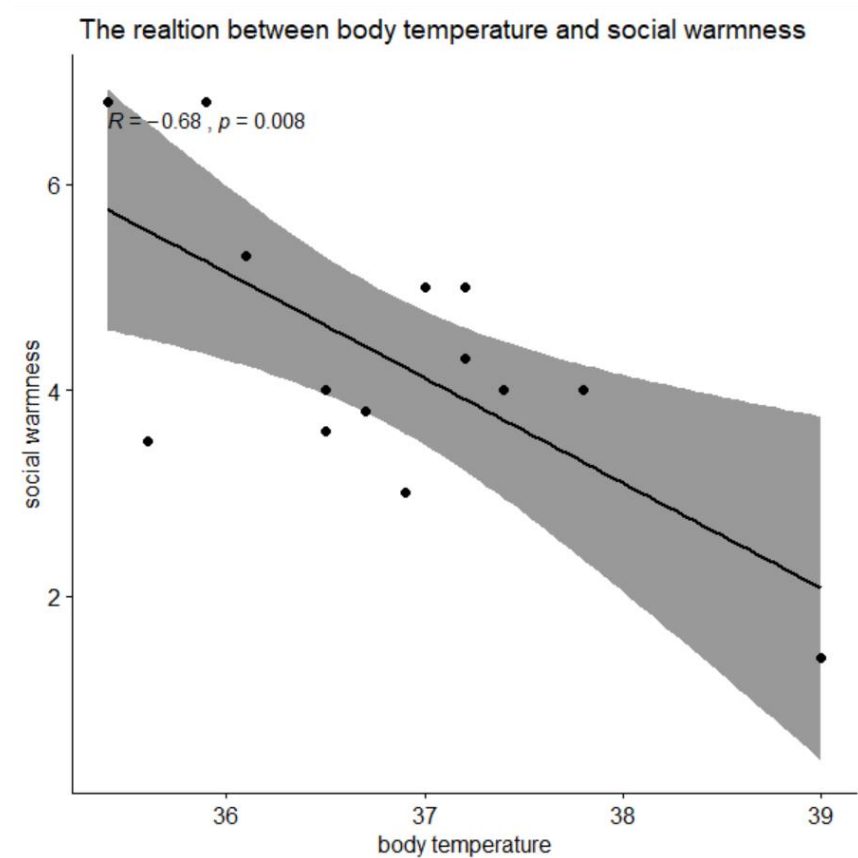
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.09 on 12 degrees of freedom

Multiple R-squared: 0.4563, Adjusted R-squared: 0.411

F-statistic: 10.07 on 1 and 12 DF, p-value: 0.008019

2.



ח. נערוך מבחן t כדי לבדוק את ההבדל ברמת "החמימות החברתית" בין שתי הקבוצות. מעניין אותנו לבחון את ההפרש בין שתי הקבוצות.

עלינו לבדוק שהשערות מבחן t אכן מתקיימות.

באמצעות מבחן שפירא, נוודא שהמשתנה מתפלג נורמלית.

```
> shapiro.test(social_warmness_y)
```

shapiro-wilk normality test

```
data: social_warmness_y  
W = 0.93895, p-value = 0.405
```

קיבלנו p-value גדול מ-0.05, ולכן לא נדחה את השערת האפס – כלומר לא נדחה את ההשערה לפיה המשתנה מתפלג נורמלית.

כמו כן, נצטרך לבדוק שיוויון שונויות בין הקבוצות. נעשה זאת באמצעות מבחן F.

F test to compare two variances

```
data: subset(my_data_frame, body_temp < my_med)$social_warmness and subset(my_data_frame, body_temp >  
my_med)$social_warmness  
F = 1.3544, num df = 6, denom df = 6, p-value = 0.722  
alternative hypothesis: true ratio of variances is not equal to 1  
95 percent confidence interval:  
0.2327323 7.8825427  
sample estimates:  
ratio of variances  
1.354445
```

קיבלנו p-value גדול מ-0.05, ולכן לא נדחה את השערת האפס – כלומר לא נדחה את ההשערה לפיה יש שיוויון שונויות.

כעת נבצע את מבחן t.

```
> test1
```

Two Sample t-test

```
data: subset(my_data_frame, body_temp < my_med)$social_warmness and subset(my_data_frame, body_temp >  
my_med)$social_warmness  
t = 1.382, df = 12, p-value = 0.1922  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-0.5847899 2.6133613  
sample estimates:  
mean of x mean of y  
4.828571 3.814286
```

קיבלנו p-value גדול מ-0.05, לכן לא נדחה את השערת האפס. כלומר, לא קיים הבדל בין שתי הקבוצות מבחינת "חמימות חברתית".

ההבדל בין התוצאות שקיבלנו פה, לעומת התוצאות שקיבלנו בסעיף ג', הוא שבסעיף זה לא דחינו את השערת האפס, כלומר הגענו למסקנה שמבחינת "חמימות חברתית" אין הבדל בין קבוצה של אנשים בעלי חום גוף גבוה לבין קבוצה של אנשים בעלי חום גוף נמוך. לעומת זאת, בסעיף ג', דחינו את ההשערה לפיה אין קשר בין טמפרטורת הגוף לחמימות החברתית (למעשה הגענו לשתי מסקנות שונות).

שאלה 2

א. הנחות הרגרסיה:

קיים קשר ליניארי בין המשתנה המנבא (X) למשתנה המנובא (Y)

הטעויות בלתי תלויות בין כל תצפית ומתפלגות נורמלית ובצורה שווה לאורך המשתנה התלוי

אין תלות בין ערך X לטעויות בניבוי של Y

(לכל X יש התפלגות נורמלית של טעויות סביב Y

ב. תוצאות:

```
Call:
lm(formula = annoyance ~ fluc)

Residuals:
    Min       1Q   Median       3Q      Max
-1.78571 -0.73768 -0.03562  0.79029  2.11928

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.8826    0.7226  -2.605   0.0122 *
fluc         108.0673   10.0253  10.779 2.05e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

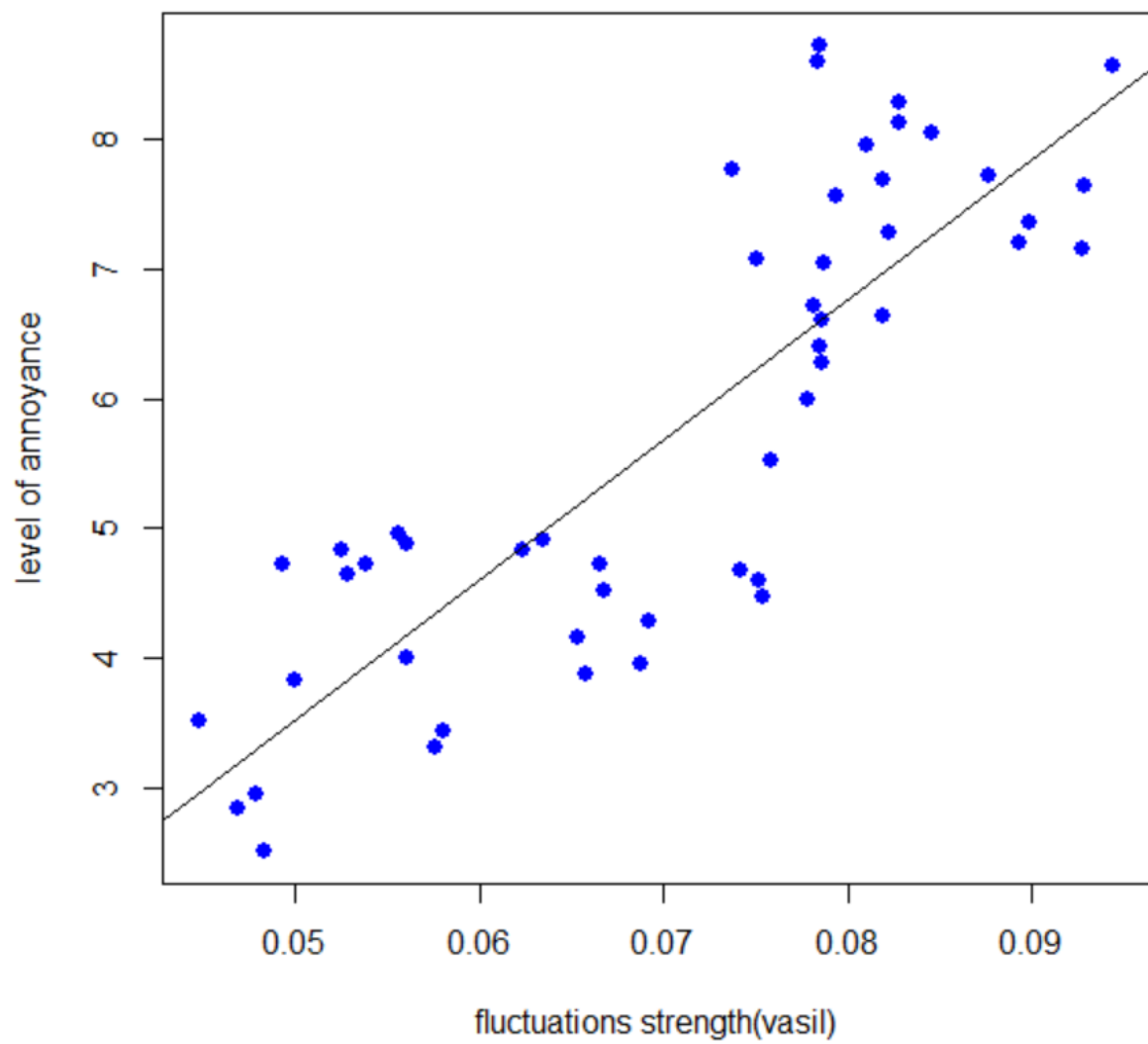
Residual standard error: 0.9745 on 48 degrees of freedom
Multiple R-squared:  0.7077,    Adjusted R-squared:  0.7016
F-statistic: 116.2 on 1 and 48 DF,  p-value: 2.05e-14
```

116.2=F דרגות חופש 1.48	F
116.2=F(1,48)	
10.779	t
108.0673	b
2.05e-14***	pvalue
0.7077	R^2

עפ"י התוצאות ניתן לומר כי ניתן לנבא את המשתנה של רמת העצבנות ע"י חוזק התנודות (המשתנה הבלתי תלוי)

ג.

**The relation between the annoyance level
and the fluctuations strength in vasil**



```

Call:
lm(formula = annoyance ~ measurement_time, data = my_data)

Residuals:
    Min       1Q   Median       3Q      Max
-3.1968 -1.2764 -0.4968  1.6232  3.0032

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.7168     0.3604  15.864  <2e-16 ***
measurement_time  0.0928     0.5096   0.182    0.856
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.802 on 48 degrees of freedom
Multiple R-squared:  0.0006903, Adjusted R-squared:  -0.02013
F-statistic: 0.03316 on 1 and 48 DF,  p-value: 0.8563

```

האם המשתנה הנומרי של זמן הדגימה יכול לנבא את רמת העצבנות ?
השערת ה-0 תהיה שאין קשר בין רמת העצבנות לזמן הדגימה.
לפי תוצאות מבחן T שהשווה בין הקבוצות נראה שערכי T קרובים ל-0 ובנוסף ערך ה-P גדול מ-0.05 ולכן לא ניתן באופן מובהק סטטיסטית לדחות א השערת ה-0. ניתן להסיק שעל סמך משתנה זה בלבד לא נוכל לנבא את רמת העצבנות.

ה. משמעות מקדם b (שיפוע) ומשמעות a (חותך) במודל זה :
מאחר וקבענו שהדגימות שנאספו בשעות הבוקר יקודדו ב-0, כלומר $X=0$, המשמעות של a תהיה הממוצע של קבוצת הדגימות שנאספו בשעות הבוקר. $(a + bX = Y)$
משמעות הערך b תהיה ההפרש בין התוחלת של הקבוצות מכיון שערכי ה- X הם 0 או 1 כך ההפרש בין ערכי ה- X הוא 1.