ניסויים השוואתיים

Comparative Experiments

ניסוי השוואתי בא להשוות בין שיטות או טיפולים.

ניסוי השוואתי יכול לכלול רק טיפולים חדשים או טיפולים חדשים ביחד עם *קבוצת ביקורת* של טיפול סטנדרטי או טיפול פלסבו.

Boy	F1	F2	
1	89.7	84.7	
2	81.4	86.1	
3	84.5	83.2	
4	84.8	91.9	
5	87.3	86.3	
6	79.7	79.3	
7	85.1	82.6	
8	81.7	89.1	
9	83.7	83.7	
10	84.5	88.5	
ממוצע	**84.24**	**85.54**	**1.3**
סטית תקן	2.9018	3.650327	

להלן נתוני הניסוי

משווים שני כמויות של דשנים

המשתנה Y – יבול ליחידת שטח

10 חלקות עם דשן F1

10 חלקות עם דשן F2

תכנוני ניסוי

(A) קיימות 20 חלקות במקומות שונים וחולקו מקרית ל שתי קבוצות

(B)קיימות 10 חלקות. כל חלקה חולקה לשניים – זוגות

(C)ישנן הרבה מאד חלקות המשתמשות ב F1 וגם הרבה מאד חלקות המשתמשות ב F2. נדגמו 10 חלקות מכל אוכלוסייה

(D) 10 חלקות – ראשית עם F1 ואח״כ עם F2

תכנוני הניסוי

(A) RANDOMIZED EXPERIMENT

(B) RANDOMIZED BLOCKING

(C) RANDOM SAMPLING

(D) נתונים הסטוריים: שימוש בהתפלגות התייחסות חיצונית

דוגמה: ניסוי החוטים

נחזור לניסוי במפעל הטקסטיל למצוא חוט שנקרע כמה שפחות במהלך תהליך האריגה.

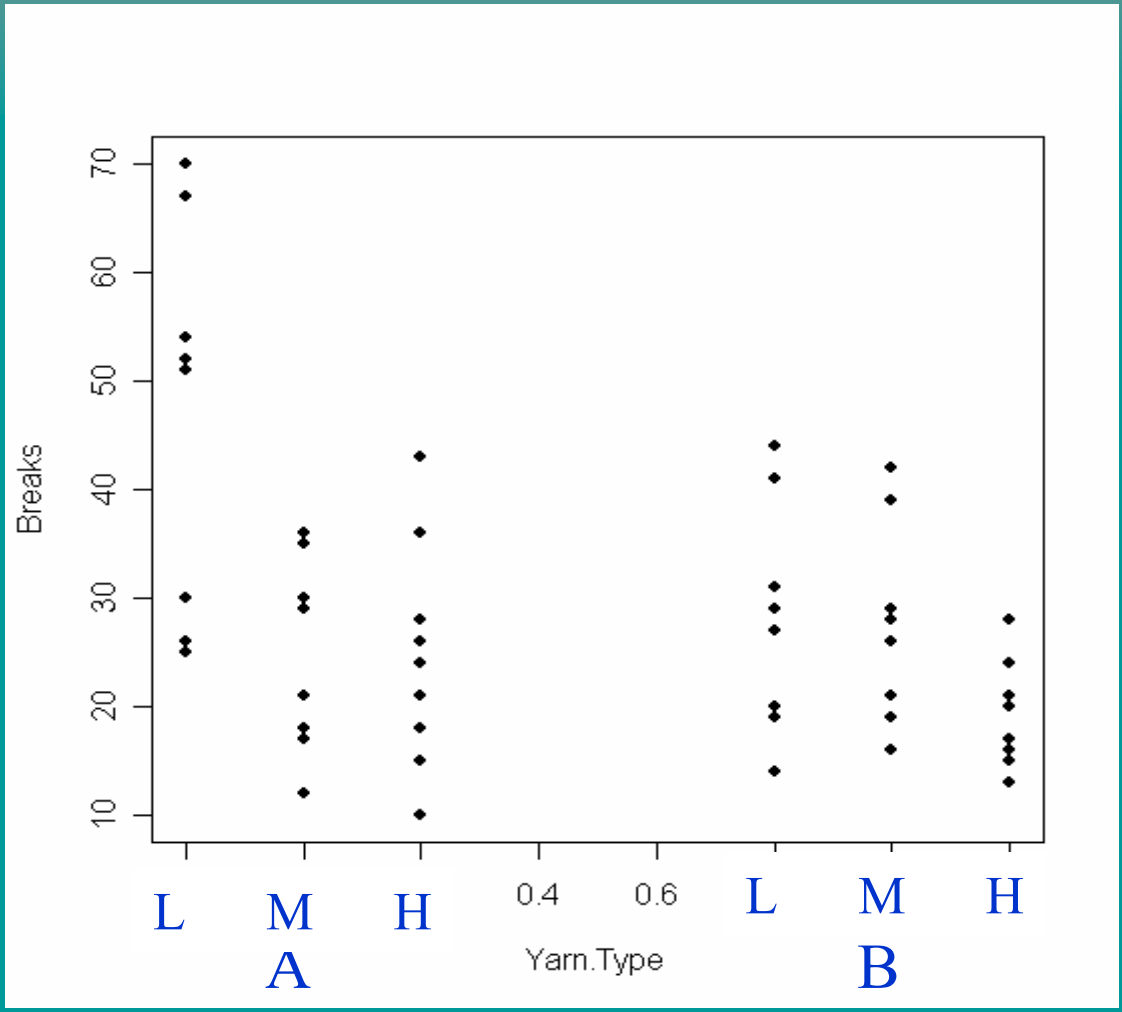
הניסוי כלל 6 סוגים של חוט, עם 2 סוגי כותנה כפול 3 רמות של צפיפות ליפוף. באקראי, חילקו 54 יחידות אריגה לששת הסוגים של חוט, 9 יחידות לכל סוג.

הביצועים נמדדו לפי מספר החוטים שנקרעו במהלך האריגה.

תוצאות הניסוי

A Low	A Med	A High	B Low	B Med	B High
26	18	36	27	42	20
30	21	21	14	26	21
54	29	24	29	19	24
25	17	18	19	16	17
70	12	10	29	39	13
52	18	43	31	28	15
51	35	28	41	21	15
26	30	15	20	39	16
67	36	26	44	29	28

מבט גרפי בנתונים



מעט מידע סיכומי על כל סוג של חוט

כותנה	צפיפות	ממוצע	חציון	SD
A	L	44.6	51	18.1
A	M	24.0	21	8.7
A	H	24.6	24	10.3
B	L	28.2	29	9.9
B	M	28.8	28	9.4
B	H	18.8	17	4.9

השוואת קבוצות

שאלות שמתבקשות:

- האם יש הבדל בין שני סוגי הכותנה?
- האם יש הבדל לפי רמת הצפיפות?
- האם ההבדל בין סוגי הכותנה משתנה בהתאם לצפיפות?

בדרך כלל, נתמקד בהבדלים הקשורים לתוחלת של התוצאה.
אך ניתן גם לחשוב על מצבים בהם חשוב למצוא גם
הבדלים במידת הפיזור, הופעה של ערכים קיצוניים, או
תכונות אחרות של התפלגות של התוצאות.

השוואת קבוצות

נבדוק באופן מפורט את השאלה:

- **האם יש הבדל בתוחלת הקרעים בין סוגי הכותנה בצפיפות האמצעית?**

בנוסף, ננסה "לכמת" מהו ההבדל בתוחלת מספר הקרעים.

לפי המבט הראשוני, מספר הקרעים מעט גבוה יותר עם סוג B, אך יש גם פיזור רב בין אריגה לאריגה.

השוואת קבוצות

גישה ראשונה: לנתח את הנתונים תוך הנחה שהם באים מהתפלגויות נורמליות.

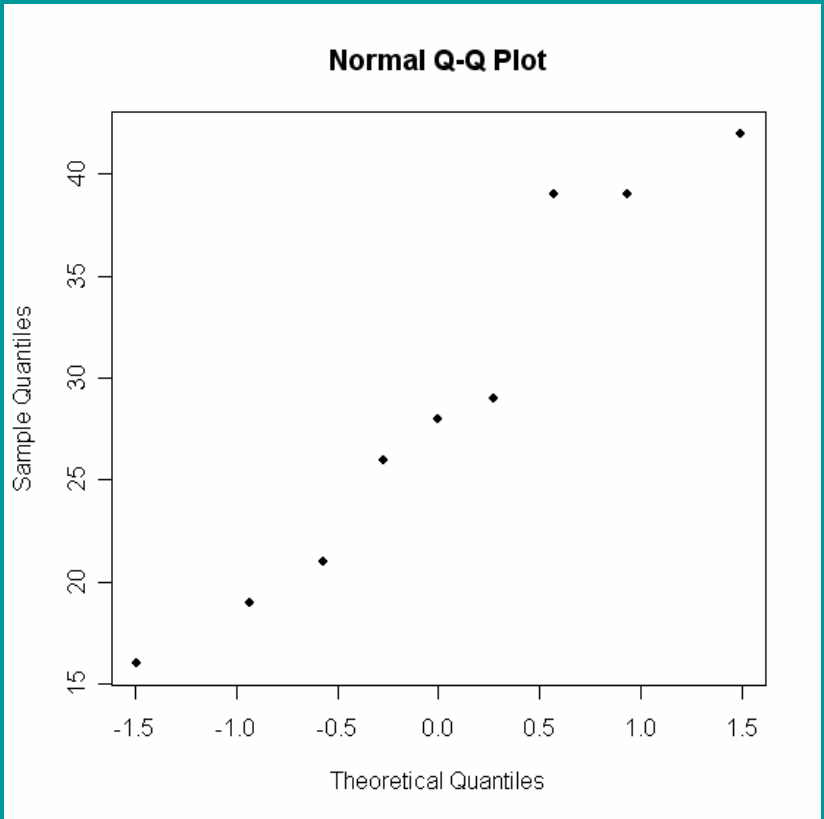
המודל הסטטיסטי שמכוון אותנו: התוצאות בכל אחת מן הקבוצות הן "כמו" מספרים אקראיים שנדגמו מהתפלגויות נורמליות.

לכל התפלגות תוחלת משלה.

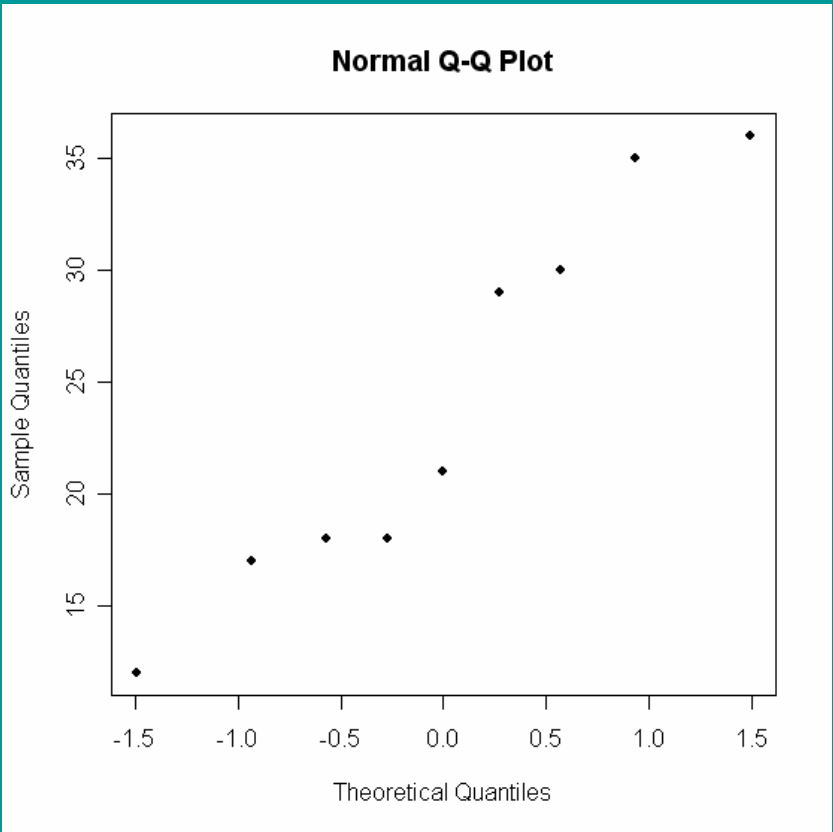
יתכן ולשתי הקבוצות אותה שונות, יתכן ולא.

תרשימי הסתברות נורמלים

סוג B



סוג A



סיכום כמותי ומבחן מובהקות

Standard Two-Sample t-Test

data: Breaks[(Cotton == "A") & (Density == "M")] and Breaks[(Cotton == "B") & (Density == "M")]

t = -1.1194, df = 16, p-value = 0.2795

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-13.825590 4.270034

sample estimates:

mean of x mean of y

24.00000 28.77778

כדי לבצע את מבחן t ב- EXCEL

Tools →

Data Analysis

בתפריט שנפתח:

T-Test: Two-Sample Assuming Equal Variances

בתפריט החדש:

למלא את המידע על מיקום הנתונים והפלט הדרוש.

השוואת קבוצות

מה לעשות אם הנתונים לא נראים מתאימים למודל הנורמלי?

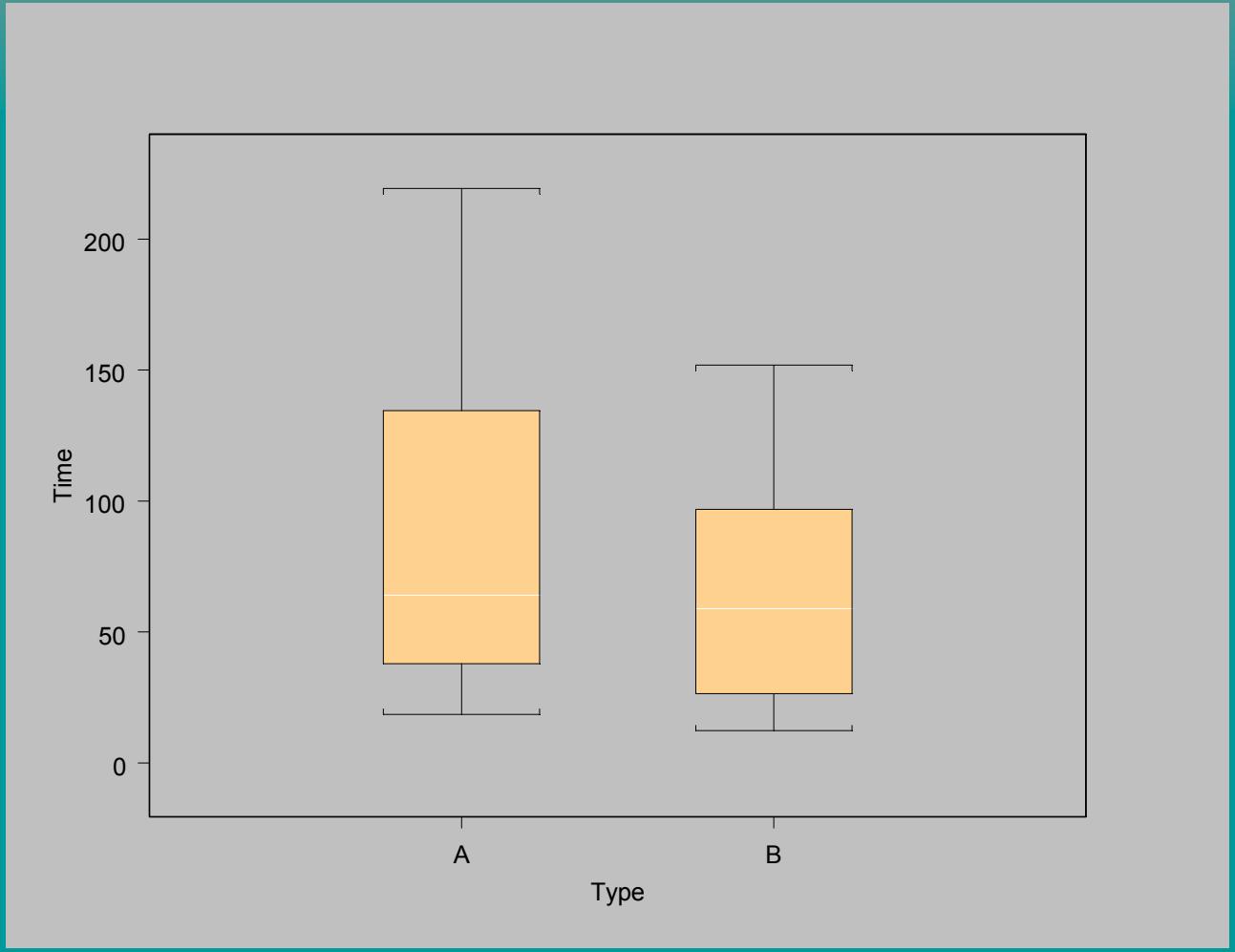
1. השונויות לא שוות – נציג גרסה חלופית של מבחן t.
2. ההתפלגויות אינן נורמליות – ניתן לחפש טרנספורמציה של הנתונים לסקלה אחרת (למשל log, שורש, או 1 חלקי) בה הנורמליות תקפה.
3. שתי הבעיות הנ"ל – ניתן לבצע מבחן Wilcoxon הלא-פרמטרי להשוואה בין הקבוצות.
4. אי-תלות הנתונים לא סבירה – להציע מודל שמשקף את התלות.

דוגמה: אורך חיים של חומר בידוד

Type A	Type B
219.3	21.8
79.4	70.7
86	24.4
150.2	138.6
21.7	151.9
18.5	75.3
121.9	12.3
40.5	95.5
147.1	98.1
35.1	43.2
42.3	28.6
48.7	46.9

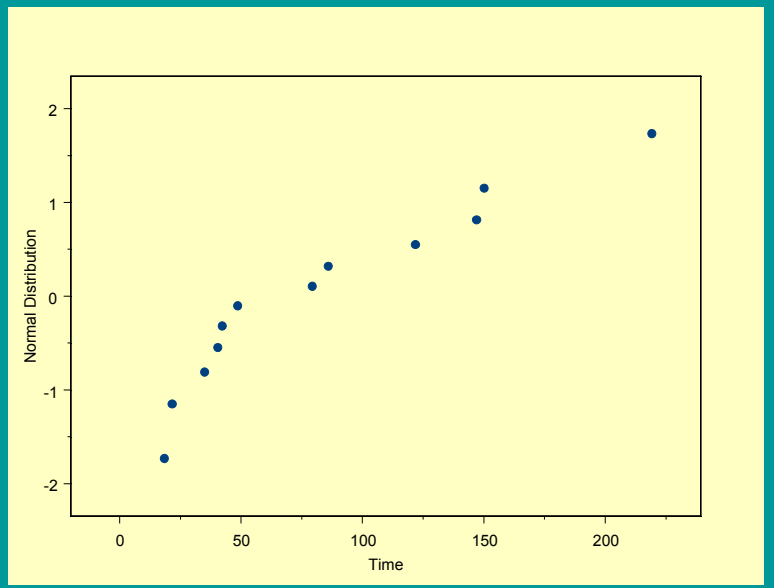
נערך ניסוי להשוות בין שני סוגים של חומר לבידוד חשמלי. לכל חומר נבדקו 12 יחידות ונמדד משך הזמן (בדקות) עד לכשל של החומר. ממול התוצאות.

מבט גרפי בנתונים

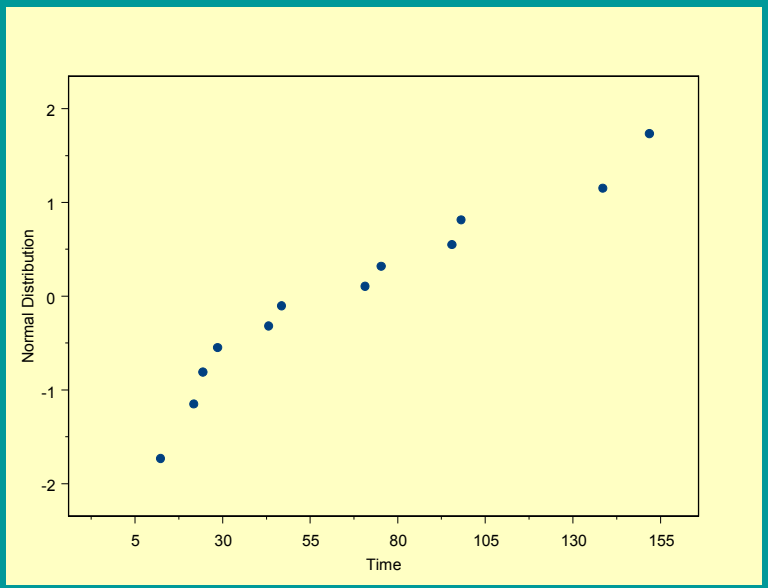


תרשימי הסתברות נורמלים

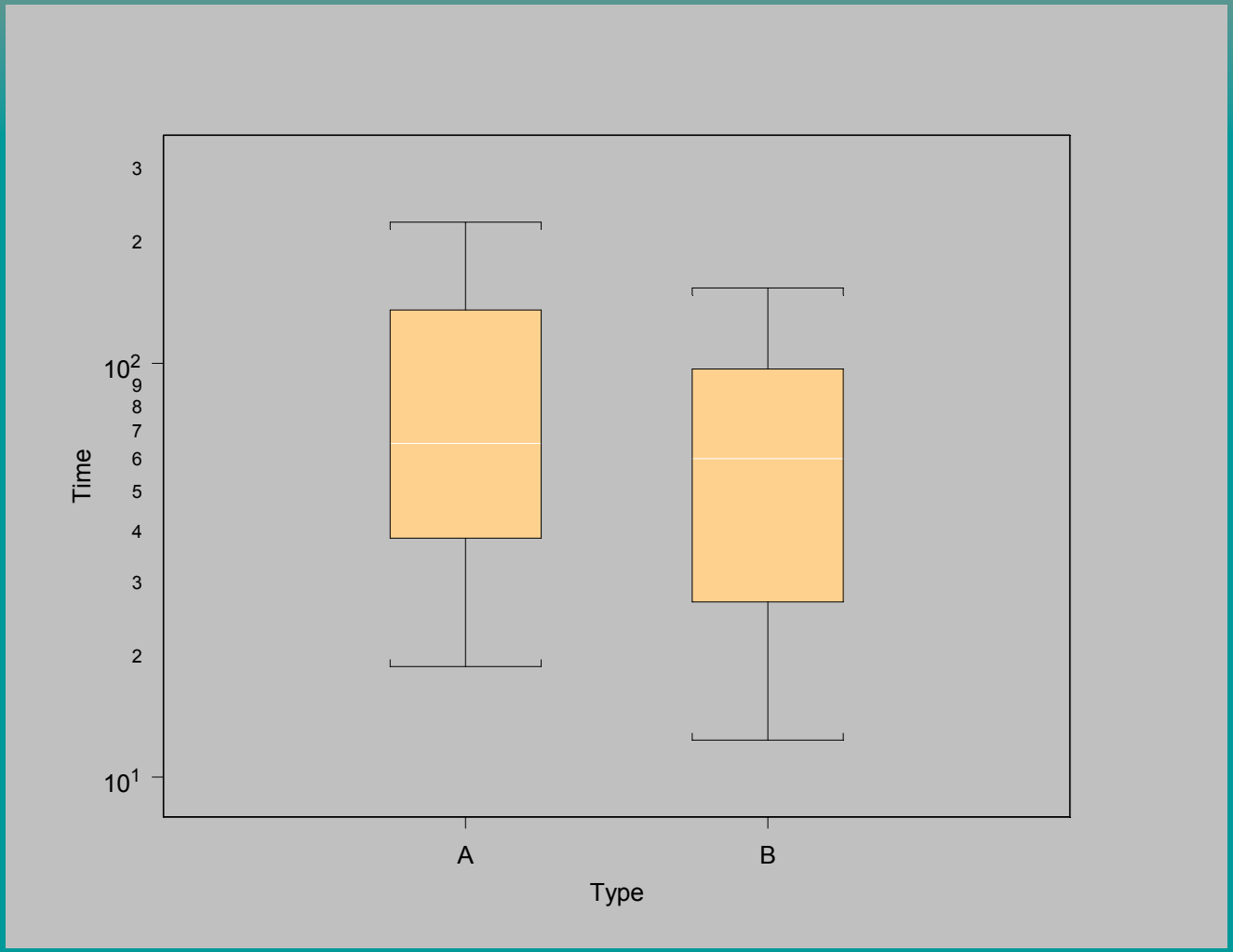
סוג A



סוג B

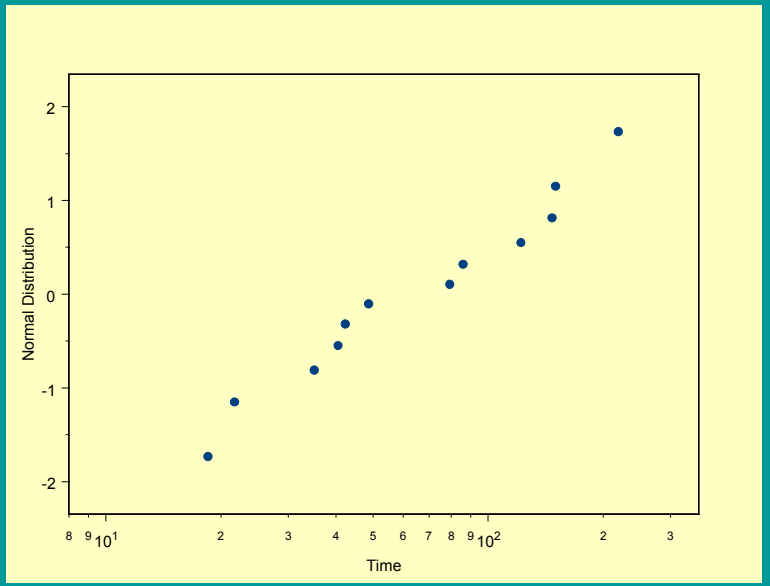


מבט גרפי בנתונים – מעבר לטרנספורמציה log

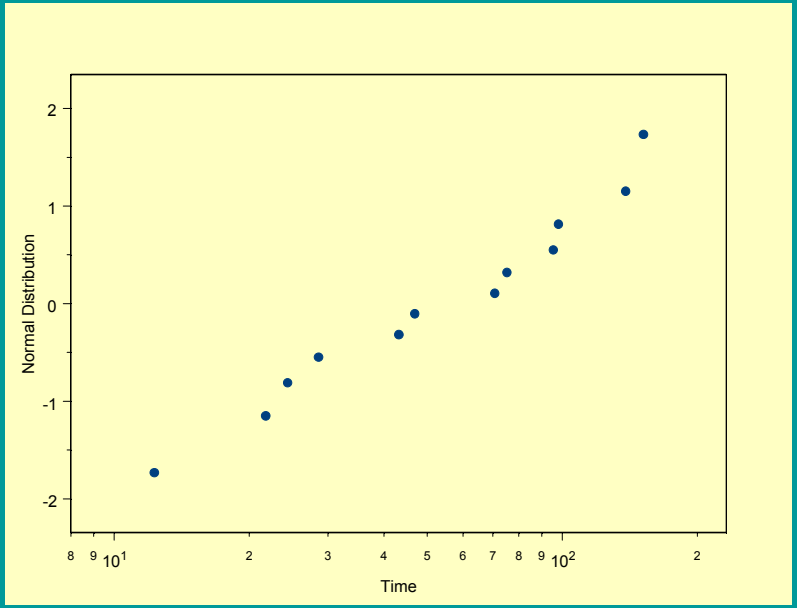


תרשימי הסתברות נורמלים

סוג A



סוג B



סיכום כמותי ומבחן מובהקות

	ממוצע	SD
סוג A	1.806	0.347
סוג B	1.717	0.345

סיכום כמותי ומבחן מובהקות

מבחן t נותן סטטיסטי t = 0.65 עם p-value = 0.53.

אין הבדל מובהק בין שני הסוגים של חומר בידוד.

רווח בר סמך (95%) להפרש התוחלות, בסקלת log לפי בסיס 10, הוא 0.20- עד 0.38.

שאלה: מה ניתן להסיק על ההבדל בזמן לכשל עצמו (ולא על הלוגריתם של הזמן)?

מבחן Wilcoxon

מבחן Wilcoxon בודק את ההשערה שלנתונים משתי הקבוצות אותה התפלגות. מבחן זה כן מניח שהנתונים בלתי תלויים, אבל אין הנחה של נורמליות כמו במבחן t.

המבחן מתבסס רק על הדירוג של הנתונים, מן הקטן ביותר (שמקבל דירוג 1), ועד לגדול ביותר.

מחשבים את הדירוג הממוצע של כל התצפיות מן הקבוצה הראשונה. תחת השערת האפס, הדירוגים בקבוצה זו הם בחירה אקראית מכלל הדירוגים האפשריים. שיקול זה מאפשר לחשב את ההתפלגות של הדירוג הממוצע ולחשב p-value מתוצאות הניסוי.

מבחן Wilcoxon

היות והמבחן מתבסס על הדירוגים, אין כל סיבה לבצע טרנספורמציה – כל טרנספורמציה מונוטונית תוביל לאותן מסקנות.

לנתונים על חומר הבידוד: הדירוג הממוצע הוא 13.17 וה-p-value = 0.67.

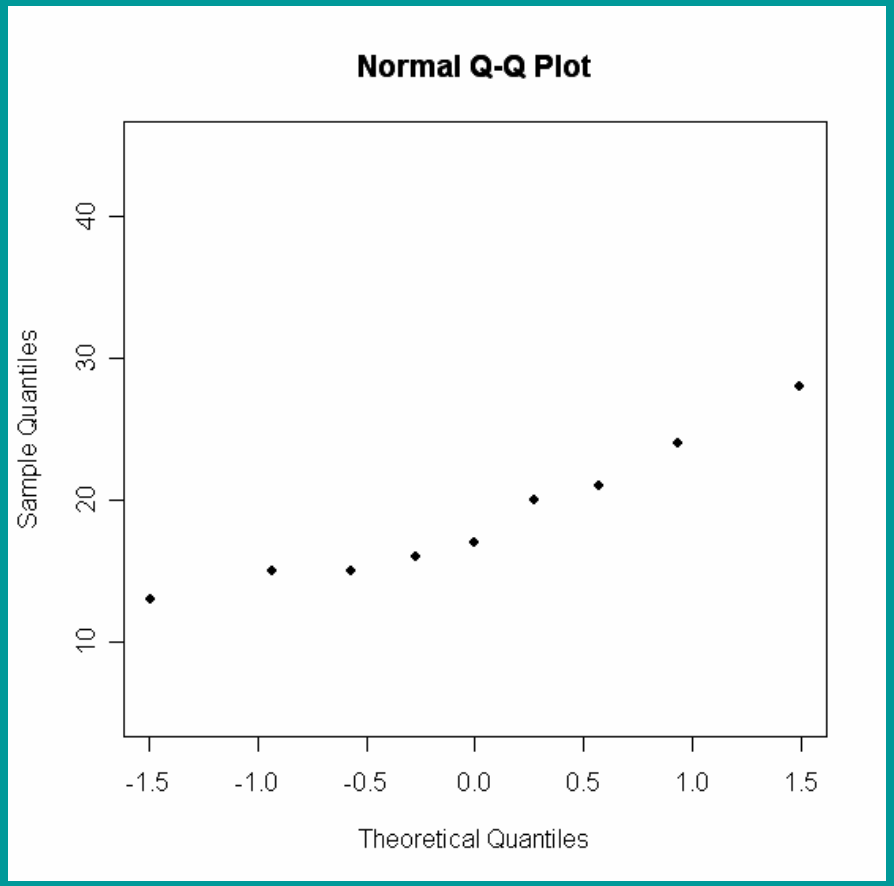
לביצוע המבחן ב- R, לבצע את הפקודה wilcox.test(x,y)

דוגמה: ניסוי החוטים בצפיפות גבוהה

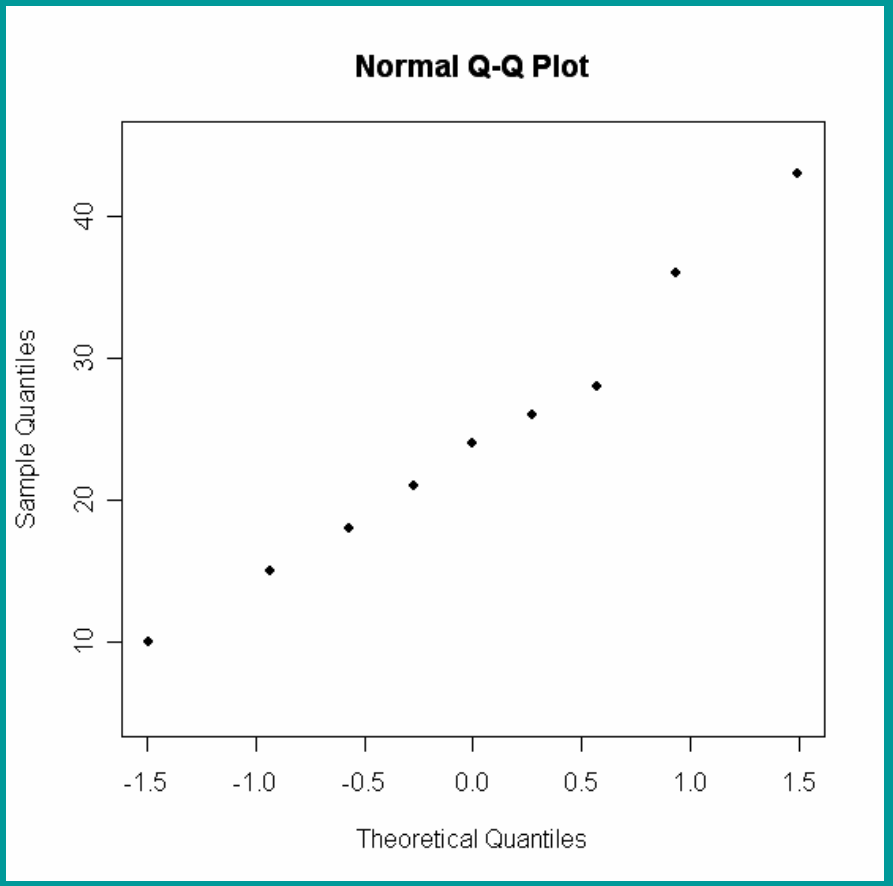
כותנה	צפיפות	ממוצע	חציון	SD
A	L	44.6	51	18.1
A	M	24.0	21	8.7
A	H	24.6	24	10.3
B	L	28.2	29	9.9
B	M	28.8	28	9.4
B	H	18.8	17	4.9

תרשימי הסתברות נורמלים

סוג B



סוג A



סיכום כמותי ומבחן מובהקות

בביצוע המבחן, מוסיפים: var.equal=F

מבחן t נותן סטטיסטי t = 1.52 עם 11.45 דרגות חופש ו- p-value = 0.155 (למבחן דו-צדדי).

לא נמצא הבדל מובהק בתוחלת מספר הקרעים בין שני סוגי הכותנה בצפיפות גבוהה.

רווח בר סמך (95%) להפרש התוחלות הוא 2.53- עד 14.09.

דוגמה: ניסוי קליני למניעת אירוע מוחי

- Aspirin and ticlopidine for prevention of recurrent stroke in Black patients: A randomized trial ***JAMA***; Chicago; Jun 11, 2003; Philip B Gorelick;DeJuran Richardson;Michael Kelly;Sean Ruland;et al.
- The primary outcome of recurrent stroke, myocardial infarction, or vascular death was reached by 133 (14.7%) of 902 patients assigned to ticlopidine and 112 (12.3 %) of 907 patients assigned to aspirin.

מבחן מובהקות

מבחן חי-בריבוע להשוואת הפרופורציות נותן סטטיסטי של 2.02 על דרגת חופש אחת, עם p-value = 0.16.

אין הבדל מובהק בין שני הטיפולים.

כדי לבצע את מבחן חי-בריבוע ב- R

א. לארגן את הנתונים עם שורה לכל נבדק, כאשר Tipul הוא המשתנה לטיפול שניתן ו- Status הוא משתנה השווה ל- 1 לאלה שעברו אירוע ללבי במהלך המעקב ול- 0 אם לא היה אירוע.

```
chisq.test(table(Tipul,Status))
```

הפקודה crosstabs לא עובדת ב- R.

ב. להגדיר מטריצה שמאחסנת את הטבלה 2 על 2 המתאימה לנתונים.

```
aa ← matrix(c(133,769,112,795),ncol=2)
chisq.test(aa)
```

דוגמה: ניסוי להשוואה בין חומר לסוליות נעליים

הניסוי משווה בין שני חומרים שונים להכנת סוליות.

השתתפו 10 ילדים בניסוי. כל ילד קיבל נעל אחת מחומר A ונעל אחת מחומר B. שיבוץ החומר לנעל ימין או שמאל בוצע באקראי.

הנתונים בניסוי זה אינם בלתי-תלויים!

יש תלות ברורה ומכוונת בין זוג הנתונים שמתקבלים מאותו ילד. ניסוי מסוג זה נקרא *ניסוי מזווג*.

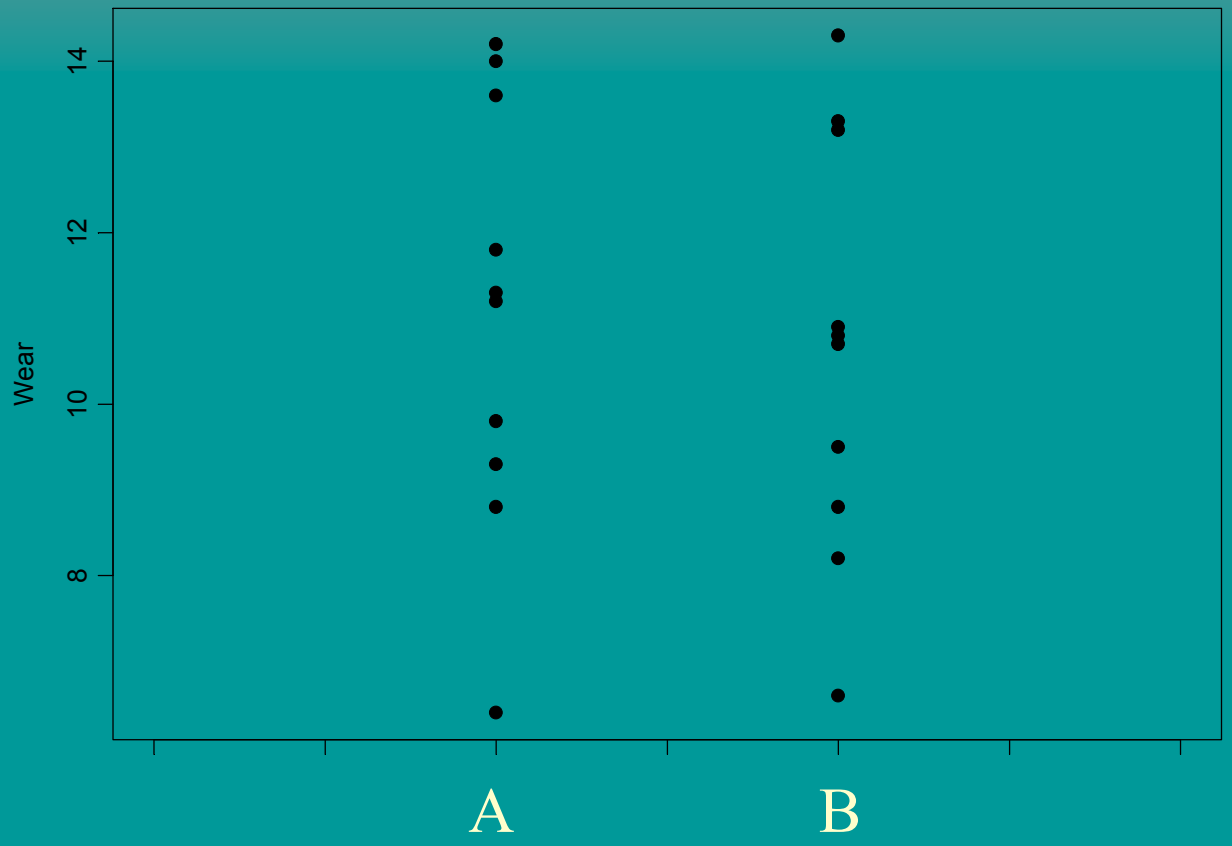
כל ילד מהווה *בלוק* ובכל בלוק תצפית אחת על כל סוג של חומר.

להלן נתוני הניסוי

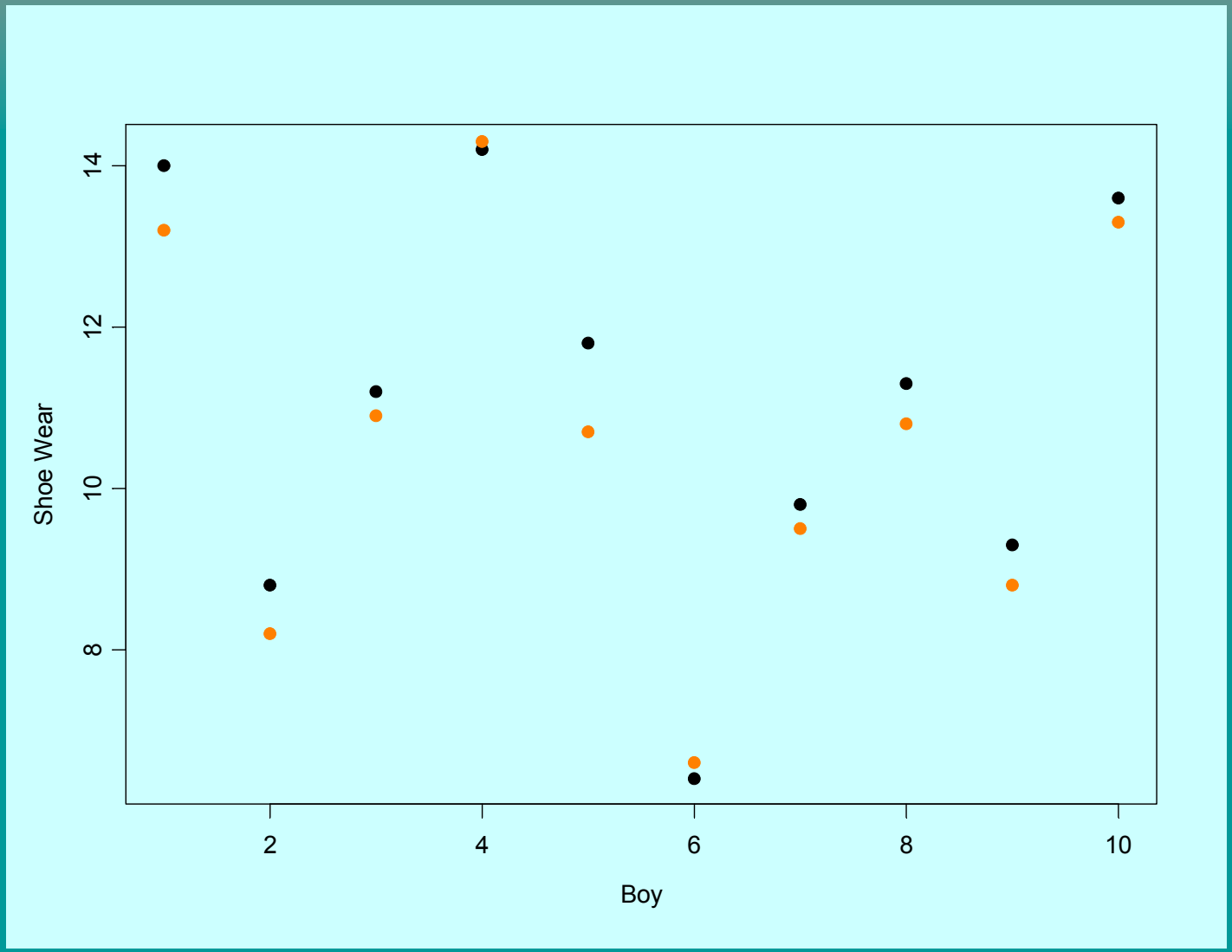
Boy	A	B	Right
1	14	13.2	A
2	8.8	8.2	A
3	11.2	10.9	B
4	14.2	14.3	A
5	11.8	10.7	B
6	6.4	6.6	A
7	9.8	9.5	A
8	11.3	10.8	A
9	9.3	8.8	B
10	13.6	13.3	A
	11.04	**10.63**	**0.41**
	2.52	**2.45**	
stdev(=2.5)*sqrt(2/10)			**1.12**

המשתנה הנמדד
הוא שחיקה
ביחידות של
מאיות אינץ'

מבט גרפי על הנתונים (בלי לנצל את הזוגות)



מבט גרפי על הנתונים עם הדגשת הזוגות



להלן נתוני הניסוי

המשתנה הנמדד הוא שחיקה ביחידות של מאיות אינץ'

Boy	A	B	Diff
1	14	13.2	0.8
2	8.8	8.2	0.6
3	11.2	10.9	0.3
4	14.2	14.3	-0.1
5	11.8	10.7	1.1
6	6.4	6.6	-0.2
7	9.8	9.5	0.3
8	11.3	10.8	0.5
9	9.3	8.8	0.5
10	13.6	13.3	0.3
	11.04	**10.63**	**0.41**
			0.39
stdev(=.39)*sqrt(1/10)			**0.12**
			3.35

ניתוח הנתונים המזווגים

Paired t-Test

data: x: A in shoes , and y: B in shoes

t = 3.3489, df = 9, p-value = 0.0085

alternative hypothesis: true mean of differences is not equal to 0

percent confidence interval: 95

0.6869539 0.1330461

sample estimates:

mean of x - y

0.41

מערך ניסוי עם השוואה ל- Baseline

מבנה אופייני לניסויים רבים הוא השוואה של תוצאה לאחר טיפול עם תוצאת Baseline שנמדדה לפני תחילת הטיפול.

בניסוי כזה, באופן טבעי, הנתונים מזווגים: לכל נבדק התוצאה אחרי הטיפול עומדת מול התוצאה שלפני הטיפול. ניתן לסכם את השפעת הטיפול לאותו נבדק על ידי ההפרש בין התוצאות, או השינוי היחסי.

מערך ניסוי עם השוואה ל- Baseline

ניסוי הכולל מדידות Baseline מאפשר כמה סוגים שונים של ניתוחים סטטיסטים.

1. ניתן לנתח את הנתונים המזווגים על הטיפול כדי לבחון איך השתנו הנבדקים.
2. כדי להפריך אפשרות שהשינוי היה נראה גם בלי הטיפול, נהוג לכלול קבוצת ביקורת, שלא עוברת את הטיפול או עוברת טיפול סטנדרטי. אז ניתן לבצע ניתוח סטטיסטי להשוואה בין מדגמים בלתי-תלויים, המתמקד במדד לשינוי אצל כל נבדק.

מערך ניסוי עם השוואה ל- Baseline

3. במקום לכלול קבוצה נפרדת של נבדקים לצורך קבוצת ביקורת, אפשר לבדוק גם את הטיפול הסטנדרטי על כל הנבדקים, כך שלכל אחד מקבלים מדד לשינוי בטיפול הנבדק וגם בטיפול הסטנדרטי. אז ניתן להשוות את שני הטיפולים תוך חישוב מדד סיכום השוואתי לכל נבדק.

דוגמה: ניסוי להערכת ההשפעה של עישון על חולי לב

מטרת הניסוי הייתה להעריך את ההשפעה של עישון על אנשים שעברו לאחרונה אוטם בלב ובמיוחד לבדוק את ההשפעה של CO, אחד החומרים החודרים לראיות דרך עישון.

השתתפו 12 אנשים. כל משתתף התבקש לבצע התעמלות ודיווח על הזמן (בשניות) עד לתחילת כאבים בחזה. לאחר מנוחה, הנבדק התבקש לעשן 5 סיגריות (ללא ניקוטין) ולחזור על מדידת הזמן עד לתחושת כאב.

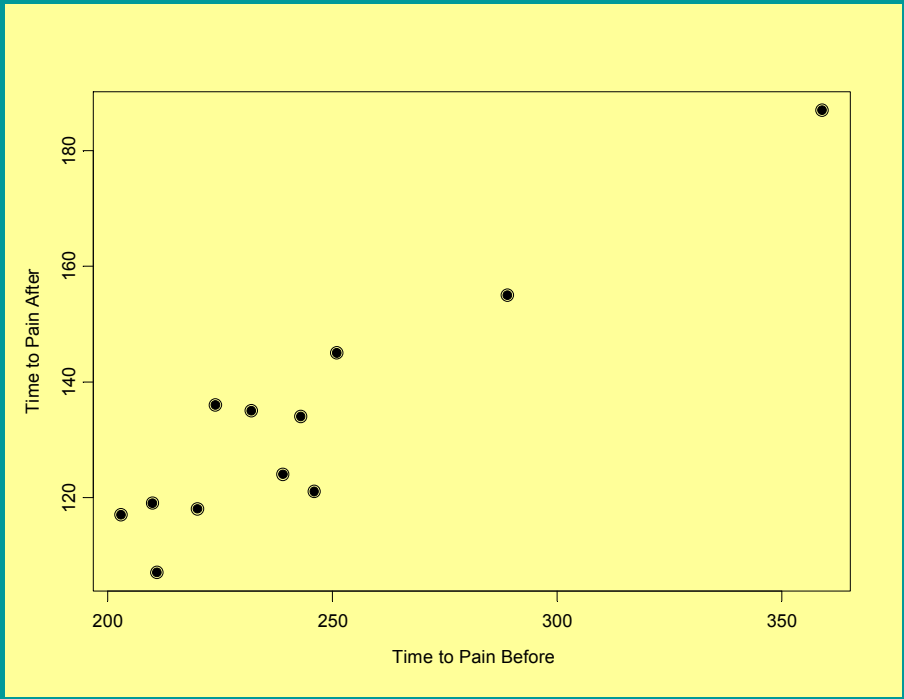
ביום נוסף של בדיקות, החוקרים חזרו על אותו מבנה, אך במקום העישון, הנבדקים נשמו כמות CO המקבילה לזו שמקבלים מ- 5 סיגריות.

בסיכום, יש לנו שני טיפולים – עישון ונשימת CO – וכל נבדק נמדד לפני ואחרי כל אחד מן הטיפולים.

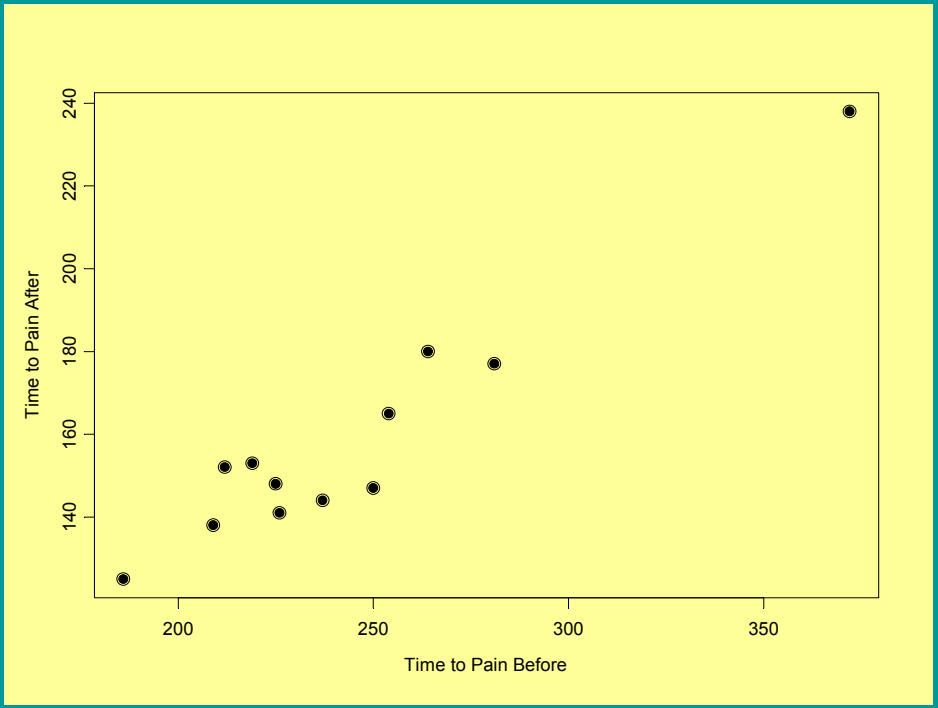
מערך הניסוי כאן – לכל נדבק ציון לשינוי בכל אחד מן הטיפולים.

השוואת הזמן אחרי והזמן לפני

עישון סיגריות

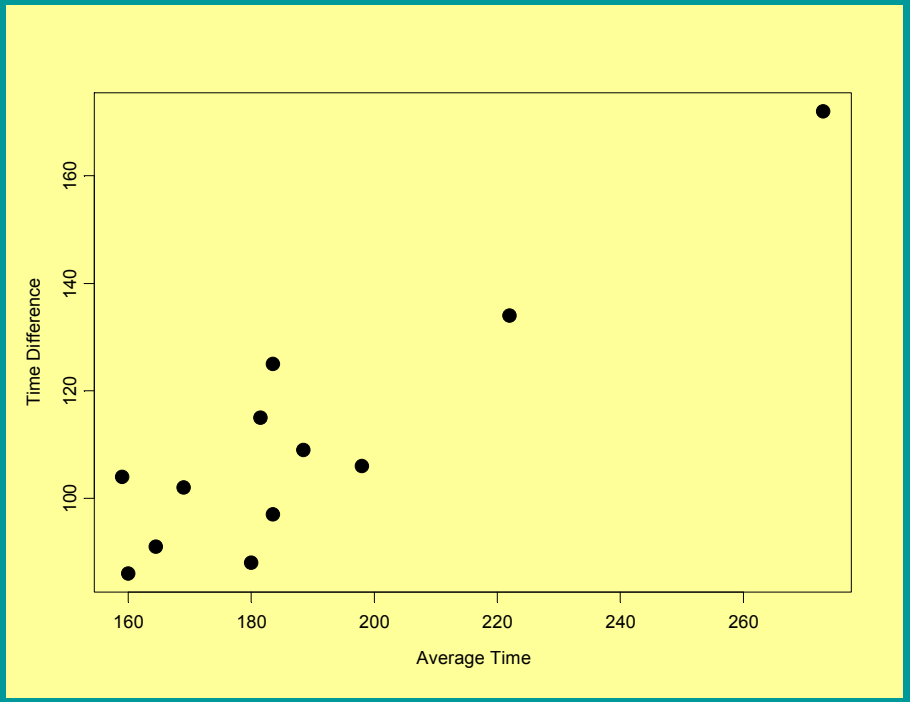


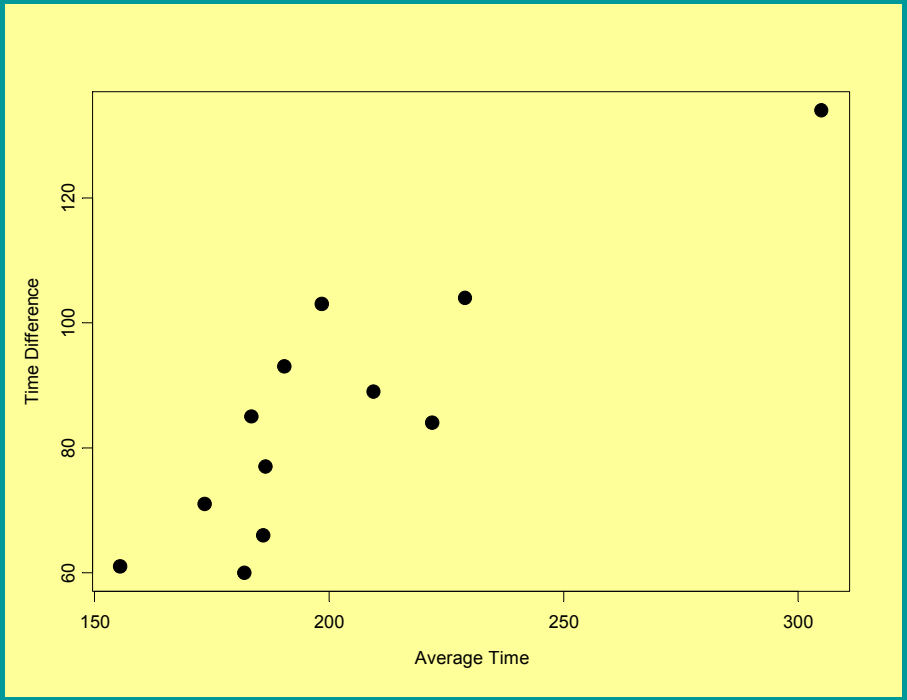
נשימת CO



איך כדאי לסכם כל זוג של מדידות לפני ואחרי?

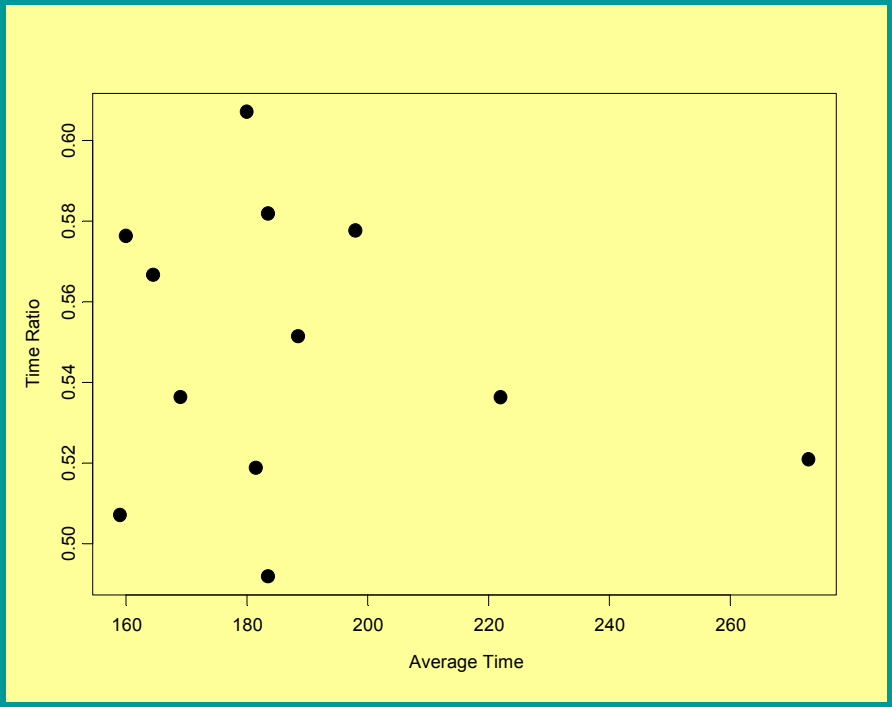
סיכום לפי הפרש הזמן בין לפני ואחרי.

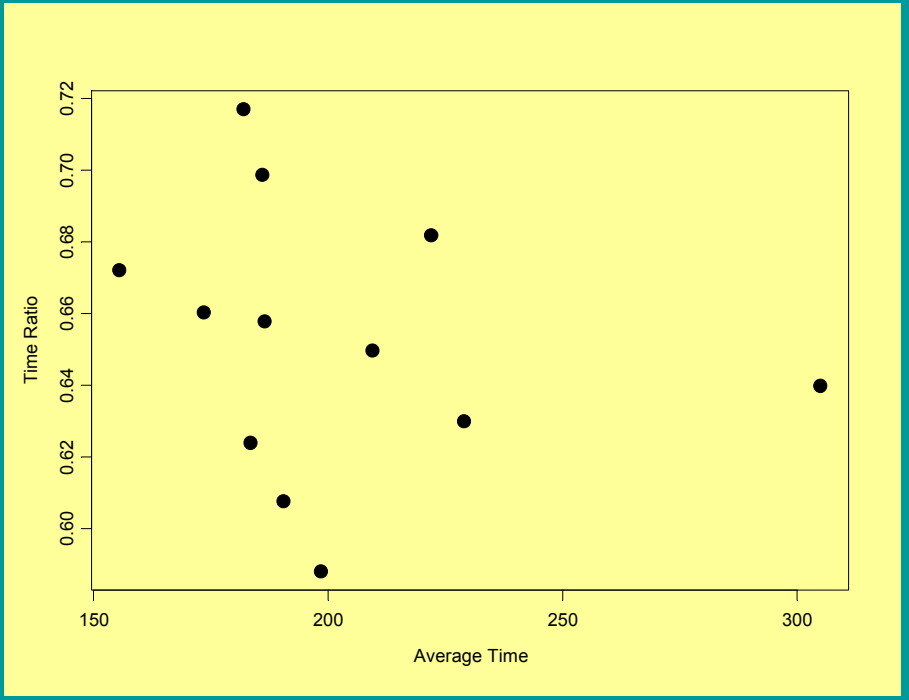




איך כדאי לסכם כל זוג של מדידות לפני ואחרי?

סיכום לפי השינוי היחסי.





סיכום לפי השינוי היחסי נראה יותר מתאים. סיכום זה מנטרל את הקשר בין מדד השינוי לבין הכושר הכללי של הנבדק.

	ממוצע	SD
יחס לאחר עישון	0.548	0.035
יחס לאחר נשימת CO	0.652	0.037

ניתן להעריך את ההשפעה של כל טיפול בנפרד על ידי מבחן t למדגם בודד, המופעל על המנה בין הזמן אחרה והזמן לפני. לשים לב שהשערת האפס של "חוסר השפעה" מתאימה ל- mu = 1.

```
t.test(Smoke.2/Smoke.1,mu=1)
```

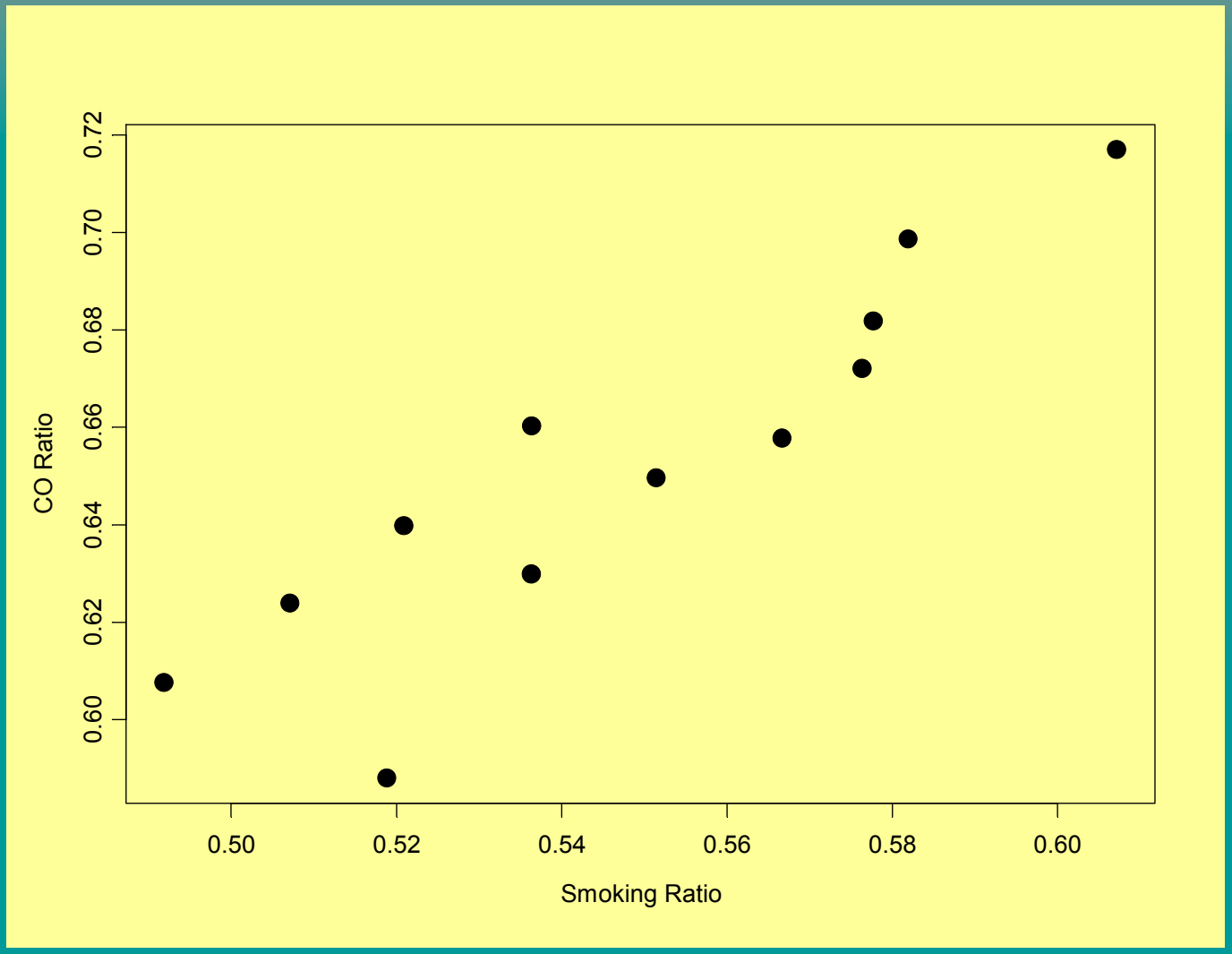
פלט ממבחן t להשפעת העישון

One-sample t-Test

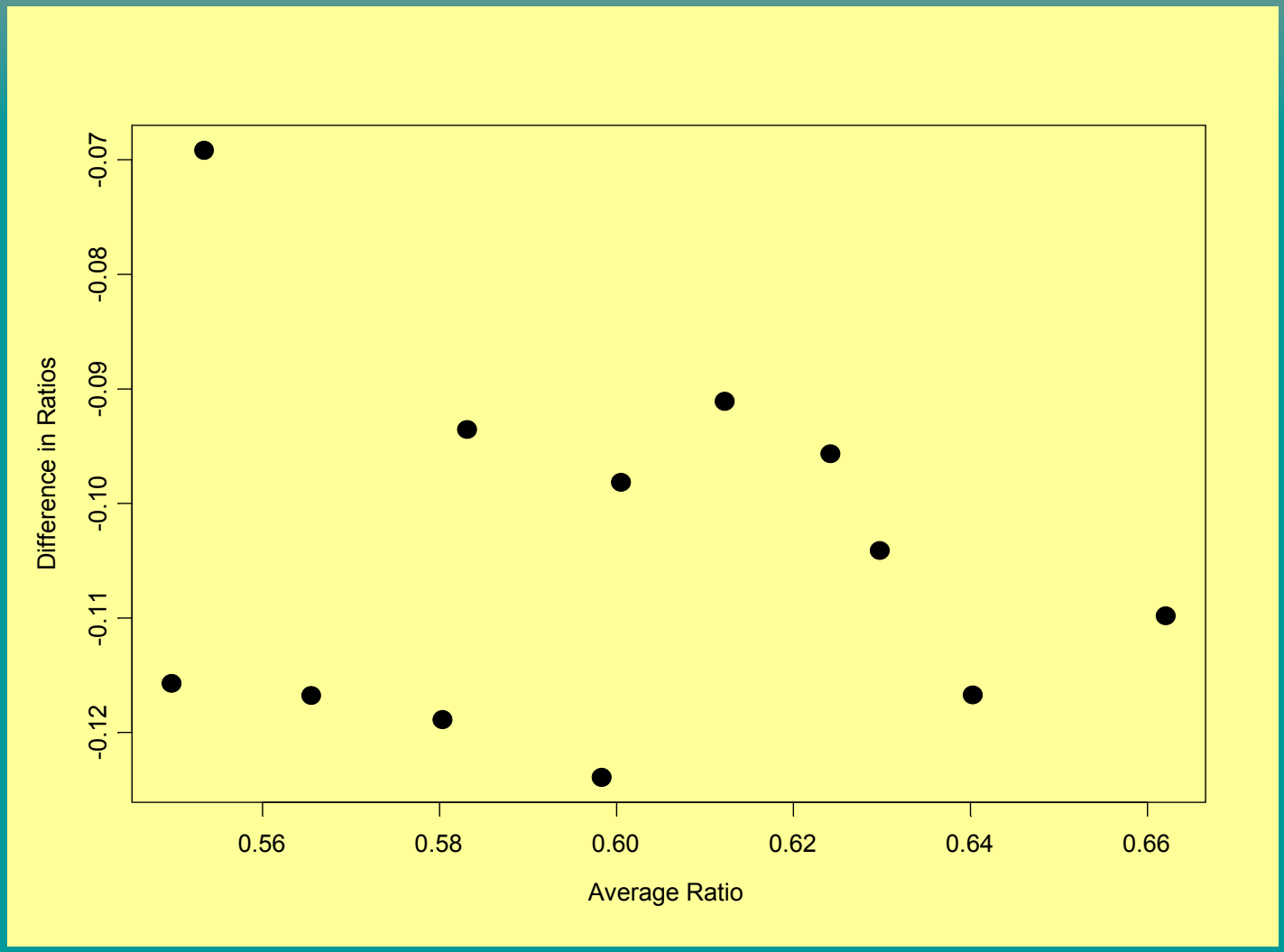
data: Smoke.2/Smoke.1
t = -44.9372, df = 11, p-value = 0
alternative hypothesis: true mean is not equal to 1
percent confidence interval: 95
0.5698679 0.5255629
sample estimates:
 mean of x
0.5477154

הניתוח של השפעת נשימת ה- CO מוביל למסקנות דומות, עם רווח סמך (95%) של 0.629 ל- 0.676.

השוואה בין העישון וה- CO



השוואה בין העישון וה- CO



השוואה בין העישון וה- CO

פלט ממבחן t

One-sample t-Test

data: Smk.Rat - Co.Rat

t = -23.0818, df = 11, p-value = 0

alternative hypothesis: true mean is not equal to 0

95percent confidence interval:

-0.09451463 -0.11443965

sample estimates:

mean of x

-0.1044771

ניסויים איטרטיביים

Iterative Experiments

דוגמה 5: מציאת מלפפון בעל יבול גבוה

תחנה לניסויים חקלאיים רוצה לערוך השואה בין 5 זנים של מלפפונים כדי לבדוק האם הם שונים מבחינת יבול.

הניסוי יתבצע על שדה ניסויי וכל חלקה בשדה תהיה זרועה באחד הזנים.

בניסוי כזה, יתרון ממוצע של 5% ביבול יכול להיות בעלת חשיבות כלכלית רבה. אבל יש גם פיזור רב – היבול בחלקה בודדת יכולה לסטות מן הממוצע ב-30% או יותר. לכן, הניסוי צריך להיות מספיק גדול לזהות הבדלים שהרבה מתחת לרמת הפיזור.

מספר סוגיות בתכנון הניסוי

איך כדאי לחלק את השדה לחלקות.

איך לקבוע איזה זן לזרוע בכל חלקה.

האם ניתן ליצור בלוקים של חלקות.

איך לדאוג שהחלקות לכל זן יהיו "ברי השוואה הוגנת".

מהו גודל המדגם הדרוש לניסוי.

דוגמה 6: מציאת מדיום אופטימלי לגידול תאים

חברה פרמאצבטית רוצה לפתח תרופה הדורשת גידול של תאים מסוג מסוים. התאים גדלים בתוך מדיום מועשר אך אין ידע תיאורטי היכול לכוון מה היא ההעשרה האופטימלית.

עובדי החברה הכינו רשימה של כ- 50 תוספים שונים העשויים לשפר את גידול התאים.

יתכן וכדאי להשתמש בכמה תוספים ביחד במדיום.

מספר סוגיות בתכנון הניסוי

יש מספיק זמן לכמה סבבים של ניסויים. איך כדאי לחלק את המאמץ בין הסבבים השונים.

איזה (וכמה) תוספים ניתן לכלול בניסוי.

איזה הרכבים של תוספים כדאי להפעיל ביחד.

דוגמה 7: חקירת חשיבה על בעיות הנדסה

ידוע שאנשים לפעמים טועים בגירויים קוגניטיביים כיוון שהם שמים לב להיבט לא רלוונטי של הגירוי.

ניסוי זה מציג שני מצולעים והנבדק צריך לסמן איזה מצולע בעל היקף גדול יותר. המצולעים שונים גם מבחינת שטח והבדל זה מספק את ההיבט הלא רלוונטי.

לחוקרים תיאוריה לפיה הזמן הדרוש לתת תשובה יהיה תלוי בהתאמה (או אי-התאמה) בסדר של השטחים וההיקפים.

מספר סוגיות בתכנון הניסוי

איך לקבוע כמה פריטים מקבל כל נבדק.

איך לקבוע כמה נבדקים לכלול.

מה יחידת הניסוי כאן? נבדק? מצולע? תשובה?

איך לקבוע את צורות המצולעים – אולי צורת המצולע מכניסה גורמים נוספים שעשויים להשפיע על תוצאות הניסוי.

התקדמות איטרטיבית

לשקול סבב של ניסויים במקום ניסוי אחד כוללני.