

ספר לימוד

SAS

Statistical Analysis System

גיא הוכמן ואלדד יחיעם



שמות מסחריים

שמות המוצרים והשירותים המוזכרים בספר הינם שמות מסחריים רשומים של החברות שלהם. הוצאת הוד-עמי והמחברים עשו כמיטב יכולתם למסור מידע אודות השמות המסחריים המוזכרים בספר זה ולציין את שמות החברות, המוצרים והשירותים. שמות מסחריים רשומים (registered trademarks) המוזכרים בספר צוינו בהתאמה.

הודעה חשובה

ספר זה מיועד לתת מידע אודות מוצרים שונים. נעשו מאמצים רבים לגרום לכך שהספר יהיה שלם ואמין ככל שניתן, אך אין משתמעת מכך אחריות כלשהי.

תוכן הספר וההפניות לספרים, לתוכנות, לאתרים ולמקורות מידע המוזכרים בו מסופקים "כמו שהם (as is)". השימוש בכל אלה הוא על אחריותו הבלעדית של המשתמש. הוצאת הוד-עמי והמחברים אינם אחראים כלפי יחיד או ארגון עבור כל אובדן או נזק ישיר או עקיף, אשר ייגרם, אם ייגרם, מהשימוש בספר ו/או בתוכנות ו/או באתרים ו/או כל מקור מידע או תוכנה המוזכרים בספר, ובכלל זה (רשימה חלקית): הפרעה במתן שירות, אובדן מידע, אובדן זמן, אובדן רווח וכד'. המשתמש רשאי להשתמש בתוכנות המוזכרות בספר ו/או לפנות לאתרים ו/או למקורות מידע אחרים על אחריותו. כל אלה הם בבעלות ובאחריות החברות המייצרות, משווקות ומציגות אותם. הוד-עמי והמחברים אינם גובים תשלום עבור השימוש בתוכנות ובמידע ממקורות אחרים המוזכרים בספר. הוד-עמי והמחברים אינם מספקים תמיכה בהתקנה ו/או ההפעלה של התוכנות ו/או בגישה לאתרים ומידע אחר. מחלקת התמיכה בהוצאת הוד-עמי תגיש עזרה רק עבור מקרים של אי בהירות של הסבר בספר או שיבוש דפוס. כל שאלה לגבי תוכנה ו/או אתר ו/או מקור מידע כלשהם יש להפנות אל מפתח/יוצרי/משווקי התוכנה ו/או אל בעלי האתרים ו/או מקורות המידע.

הוצאת הוד-עמי והמחברים עשו כל מאמץ שתוכן הספר יהיה אמין ושלם. עם זאת, ההוצאה והמחברים אינם טוענים לאמינות ולשלמות של התכנים המוצגים בספר זה, ובמיוחד דוחים כל אחריות, ובכלל זה טענה להתאמה של הנאמר בספר למקרה ספציפי כלשהו. לא ניתן ליצור או להרחיב אחריות על ידי מידע שיווקי ו/או פרסומי כלשהו. ייתכן שההצעות ו/או ההמלצות הניתנות בספר לא יתאימו לכל מצב ומקרה. הספר משווק ונמכר תוך הבנה שההוצאה והמחברים אינם מספקים שירותים שונים הכרוכים בשימוש בספר, אלא לשם הבנת הכתוב ותיקון שיבושי לשון. לקבלת שירות מקצועי יש לפנות אל בעלי המקצוע בתחום. הן ההוצאה והן המחברים אינם אחראים לכל אובדן או נזק ישיר או עקיף, אשר ייגרם, אם ייגרם, מהשימוש בספר ו/או בתוכנות ו/או באתרים ו/או כל מקור מידע או תוכנה המוזכרים בספר. אין בכוננת ההוצאה ו/או המחברים להמליץ או להעדיף תוכנה ו/או אתר ו/או מקור מידע כלשהם. רק המשתמש הוא שיחליט כיצד לנהוג על פי המוצג בספר. המשתמש צריך להיות ער לעובדה שאתרי האינטרנט הינם דינמיים ועלולים להיסגר, לשנות את התכנים שלהם וכד'. ההוצאה והמחברים אינם אחראים לשינויים אשר עלולים לחול באתרים המוזכרים בספר, ועל כן להיות שונים ממה שהוצג בספר. אין לעשות שימוש מסחרי ו/או להעתיק, לשכפל, לצלם, לתרגם, להקליט, לשדר, לקלוט ו/או לאחסן במאגר מידע בכל דרך ו/או אמצעי מכני, דיגיטלי, אופטי, מגנטי ו/או אחר - בחלק כלשהו מן המידע ו/או התמונות ו/או האיורים ו/או כל תוכן אחר הכלולים ו/או שצורפו לספר זה, בין אם לשימוש פנימי או לשימוש מסחרי. כל שימוש החורג מציטוט קטעים קצרים במסגרת של ביקורת ספרותית אסור בהחלט, אלא ברשות מפורשת בכתב מהמוציא לאור.

עריכה ועיצוב: גיא הוכמן
עיצוב עטיפה: גיא הוכמן
צילום עטיפה: יובל טבול

תודה לאריאל תלפז על הערות והארות

לשם שטף הקריאה כתוב ספר זה בלשון זכר בלבד. ספר זה מיועד לגברים ונשים כאחד ואין
בכוונתנו להפלות או לפגוע בציבור המשתמשים/ות.

(C)

כל הזכויות שמורות

הוצאת הוד-עמי בע"מ

ת.ד. 6108 הרצליה 46160

טלפון: 09-9564716 פקס: 09-9571582

info@hod-ami.co.il

www.hod-ami.co.il

הודפס בישראל נובמבר 2010

All Rights Reserved

HOD-AMI Ltd.

P.O.B. 6108, Herzliya

ISRAEL, 2010

מסת"ב 978-965-361-408-6 ISBN

את קבצי קוד המקור (התוכניות) ניתן להוריד מאתר האינטרנט של הוצאת הוד עמי:
www.hod-ami.co.il

מצא את הספר באתר ואת הלינק "קוד מקור" להורדת הקבצים. לחץ עליו ועקוב אחר ההוראות.
אם לא תגדיר אחרת, יועתקו הקבצים אוטומטית לדיסק שלך, לתיקה זו:

C:\HodAmiBooks\59444\

תוכל לבחור בעת ההתקנה בכל תיקיה אחרת.

בדרך כלל הפעולות שיש לבצע לאחר הלחיצה על הלינק: לחיצה על הפעל, לחיצה על הפעל, לחיצה על UnZip,
לחיצה על OK ו-Close.

כדי להשלים את כל התרגילים בספר תצטרך את הדברים הבאים:

<p>חשוב: הקובץ אינו מכיל את התוכנה SAS או כל תוכנה אחרת. את התוכנה יש לרכוש ולהתקין לפני תחילת השימוש בספר זה.</p>

לאחר שתוריד את קבצי קוד המקור מאתר הוד-עמי הם יימצאו (אם לא שינית) בתיקה 59444 שנמצאת תחת
HodAmiBooks שנמצאת בכונן הראשי C.

8	פרק 1 מבוא
15	פרק 2 יצירת קבצי נתונים
23	פרק 3 קריאת קובץ נתונים הקיים ב-SAS
29	פרק 4 תפעול קבצי נתונים
45	פרק 5 צירוף קבצים
53	פרק 6 פרוצדורות שירות I : מיון והפקת פלט
63	פרק 7 פרוצדורות שירות II : הגדרת משתנים וטיפול בתצפיות
78	פרק 8 פרוצדורות שירות III : טיפול בקבצי נתונים
92	פרק 9 פרוצדורות סטטיסטיות I : סטטיסטיקה תיאורית
141	פרק 10 פרוצדורות סטטיסטיות II : קשר בין משתנים
155	פרק 11 פרוצדורות סטטיסטיות III : מודלים ליניארים
191	פרק 12 פרוצדורות סטטיסטיות IV : מבחנים א-פרמטריים
200	פרק 13 פרוצדורות גראפיות : תרשימים וגרפים
221	פרק 14 פונקציות SAS מתקדמות
233	פתרון תרגילים

11	איור 1 – דוגמא לקובץ נתונים המכיל משתנה מחרוזת ושני משתנים נומריים
11	איור 2 – מנהל התצוגה של SAS
12	איור 3 – החלון Log
14	איור 4 – הרצת קוד SAS דרך התפריט הראשי
14	איור 5 – הרצת קוד SAS דרך סרגל הכלים
14	איור 6 – הרצת קוד SAS מתוך חלון ה- Editor
20	איור 7 – מבנה קובץ אקסל ש-SAS יכולה לקרוא מתוך ה-DATA STEP
21	איור 8 – כיצד להשיג את המחרוזת DDE triplet לקריאת קבצי אקסל
21	איור 9 – מחרוזת DDE triplet
56	איור 10 – דוגמא לפלט בסיסי של הפרוצדורה PRINT
58	איור 11 – כותרות אנכיות ב-PROC PRINT
60	איור 12 – פלט של PROC PRINT עם ההוראה BY, כולל ולא כולל ההוראה ID
61	איור 13 – פלט של PROC PRINT הכולל את ההוראות BY ו-SUM
111	איור 14 – פלט בסיסי של ההוראה HISTOGRAM ב-PROC UNIVARIATE
116	איור 15 – דוגמא להיסטוגרמה מותאמת אישית
120	איור 16 – פלט בסיסי של ההוראה PROBLOT ב-PROC UNIVARIATE
165	איור 17 – עקומת פיזור המופקת על ידי ההוראה PLOT ב-PROC REG
206	איור 18 – דיאגרמת פיזור ודיאגרמת BUBBLE ב-PROC GPLOT
208	איור 19 – דיאגרמת פיזור ב-PROC GPLOT הכוללת שני צירי Y
211	איור 20 – הצורה הבסיסית של כל התרשימים המופקים על ידי PROC CHART
217	איור 21 – הצורה הבסיסית של כל התרשימים המופקים על ידי PROC GPLOT
231	איור 22 – חלון המאפיינים של קיצור הדרך לתוכנת SAS
232	איור 23 – חלון המאפיינים של קיצור הדרך לתוכנת SAS לאחר הגדרת sasInitialFolder

9	טבלה 1 – סימנים מוסכמים ומשמעותם
32	טבלה 2 – אופרטורים בוליאניים ומשמעותם
68	טבלה 3 – הקשר בין ההוראה PICTURE ב-PROC FORMAT להגדרת בורר ספרה
79	טבלה 4 – סוגי קבצים הניתנים לייבוא על ידי PROC IMPORT
80	טבלה 5 – הוראות של PROC IMPORT
93	טבלה 6 – סטטיסטיים תיאוריים הזמינים ב-PROC MEANS
103	טבלה 7 – הסטטיסטיים התיאוריים ב-PROC UNIVARIATE
130	טבלה 8 – רשימת סטטיסטיים של ההוראה OUTPUT ב-PROC UNIVARIATE
134	טבלה 9 – קודי קיבוץ להפקת טבלאות ב-PROC FREQ
166	טבלה 10 – מילות מפתח למשתנה x ולמשתנה y בהוראה PLOT ב-PROC REG
185	טבלה 11 – רשימת סטטיסטיים של ההוראה OUTPUT ב-PROC GLM
203	טבלה 12 – קודי קיבוץ להפקת דיאגרמות פיזור על ידי PROC PLOT ו-PROC GPLOT
222	טבלה 13 – משתני מאקרו אוטומטיים

פרק 1

מבוא

אודות הספר

ספר ההדרכה הנוכחי מהווה מבוא מקיף לתכנות בסיסי ב-SAS (קליטת נתונים, יצירת נתונים חדשים) ולביצוע ניתוחים סטטיסטיים. הספר אינו מצריך ידע או ניסיון קודם בתכנות ב-SAS, והוא מאפשר למתכנתים מנוסים לנצל מאפיינים מתקדמים התכנות ועל יישומים בצורה אינטראקטיבית ונוחה. במקביל, הספר מאפשר למתכנתים מנוסים לנצל מאפיינים מתקדמים של SAS ולשפר את היכולת שלהם ליצור פלטי SAS מותאמים אישית. הספר מחולק לנושאים וכולל תרגילים ותוכניות לדוגמה. עם זאת, הספר כן מצריך ידע סטטיסטי, אם כי ניתן ליישם את החומר הנלמד בספר זה גם בלי ידע מוקדם זה.

כיצד להתמצא בספר

הספר הנוכחי כולל ארבעה נושאים:

- א. מבוא – כולל סקירה כללית של SAS
- ב. DATA STEP – כולל כיצד להכניס נתונים ל-SAS ולתפעל אותם
- ג. PROC STEP – כולל פרוצדורות נפוצות לניתוח סטטיסטיים ותפעול משתנים
- ד. פונקציות מתקדמות – כולל פעולות כלליות של SAS שנועדו להגביר יעילות

כל נושא כולל מספר פרקים המהווים כל אחד בפני עצמו מרכיב תפקודי עצמאי (component – חלק קוד SAS עצמאי המהווה תוכנית ברת הרצה) של SAS. כל פרק מתחיל בהצגה כללית של המרכיב, הכוללת את המבנה הכללי של הקוד להפעלת מרכיב זה. במקרה שזה רלוונטי, הפרק כולל גם את הפלט הבסיסי ש-SAS מפיקה מהרצת הקוד של המרכיב. בהמשך, הפרק כולל הוראות ואופציות מתקדמות יותר של כל מרכיב, שנועדו לאפשר למשתמשים להתאים את הפעולות של המרכיב לצרכים הספציפיים שלהם.

באופן כללי, הספר מתחיל עם רמת פירוט גבוהה הכוללת הסברים מפורטים, תוך מתן דוגמאות ספציפיות, לכל מרכיב תפקודי. כדי למנוע עומס, ככל שאנחנו מתקדמים בספר רמת הפירוט יורדת ומרכיבים מסוימים (במיוחד כאלה שיש להם מבנה דומה למבנים אחרים שנידונו בתת פרקים קודמים) מוצגים ברמה הכללית, ללא דוגמאות ספציפיות. במקרה בו יש חפיפה מלאה באופן הכתיבה או היישום של מרכיבים מסוימים, ניתנת בגוף הטקסט הפנייה לדוגמאות או הסברים רלוונטיים, שנידונו יותר בהרחבה בחלקים קודמים. לכן, מומלץ כי הקורא המתחיל יעבור על הספר לפי הסדר, במטרה לצבור את הידע והניסיון המתאימים כדי להתמודד עם מרכיבים תפקודיים מורכבים יותר המוצגים בחלקים מתקדמים יותר של הספר.

לדוגמה, למעט במקרים בהם מוגדר אחרת, כאשר רוצים להגדיר מספר משתנים עליהם ייושם מרכיב תפקודי, או כאשר רוצים להגדיר להוראה מסוימת של המרכיב מספר אפשרויות, יש לכתוב את המשתנים השונים אחד לאחר השני, כאשר המשתנים מופרדים על ידי רווחים. בתחילת הספר עובדה זו מפורטת, אולם בהמשך, אנחנו פשוט כותבים "רשימת משתנים" מתוך כוונה שהקורא כבר למד כי הדרך להכניס משתנים שונים (או אופציות שונות) היא על ידי כך שהמשתנים ברשימה יהיו מופרדים על ידי רווחים.

בנוסף, לאורך כל הספר, אנחנו משתמשים בסימנים מוסכמים המייצגים מונח או מושג מילולי. כדי להפיק את המירב מהספר, מומלץ ללמוד היטב סימנים אלה בטרם תתחילו לקרוא. הרשימה המלאה של הסימנים, כמו גם הפירוש שלהם, מוצגת בטבלה 1.

המלצות לקריאה יעילה



מאחר וחלק מההוראות והפקודות הזמינות ב-SAS עשויות להיות מאוד טכניות ומורכבות, למידתן והבנתן עשויות להיות משימה לא קלה בכלל, בעיקר למתכנתים חדשים, שרק מתחילים את דרכם בנבכי התוכנה. לכן, אנו ממליצים, בעיקר למשתמשים חדשים, לקרוא ספר הדרכה זה בשני שלבים, כפי שיפורט להלן.

בשלב הראשון, מומלץ להכיר את ההוראות הבסיסיות והפשוטות ביותר של SAS לקליטה של נתונים, המוצגות בהרחבה בפרקים הדניים ביצירת קבצי נתונים ב-SAS. בשלב זה לא מומלץ להיכנס לאפשרויות מתקדמות יותר של התוכנה, אלא להתמקד בעיקר בפעולות ההכרחיות והבסיסיות ביותר להתחלת העבודה. אותו הדין קיים גם לתפעול של קבצי נתונים, ולניתוח סטטיסטי של נתונים, שם מומלץ להתמקד במבנה הכללי של ההוראות, ולא להיכנס לכל האופציות הרבות והמורכבות הזמינות לכל סוג ניתוח. כדי להקל על משימה זו, הפרקים הדניים בפרוצדורות הסטטיסטיות של SAS מתחילים בהסבר כללי של הפרוצדורה ובדוגמא קונקרטית של קוד בסיסי ליישום הפרוצדורה, כמו גם את הפלט הבסיסי שנוצר כתוצאה מהרצה של קוד בסיסי זה. בשלב זה מומלץ גם להשתמש בתרגילים לתרגול עצמי הזמינים לקורא, בעיקר באלה שלא דורשים ידע מעמיק בתכנות, לשם יישום של החומר הנלמד.

בשלב הבא, ולאחר שנרכש כבר הידע הבסיסי הנדרש כדי לקלוט ולעבד קבצי נתונים, כמו גם לעבוד על קבצים אלה כדי להפיק נתונים סטטיסטיים בסיסיים שונים, מומלץ להתחיל להיכנס יותר אל הפרטים הטכניים שנועדו לספק ידע מקיף ונרחב יותר של האפשרויות הרבות הגלומות ב-SAS. בשלב זה מומלץ גם להתחיל להיכנס לתוך האופציות הרבות והמגוונות הזמינות למרבית הפרוצדורות הקיימות ב-SAS.

אנו מאמינים כי אופן עבודה זה יסייע לקורא "לשחות" בים הרחב שהוא עולם התכנות העשיר ש-SAS מציעה, בצורה הנוחה והיעילה ביותר, ויקל משמעותית את תהליך הלמידה של התוכנה.

כדי לסייע לקורא במשימה זו, כללנו בספר שני סוגים של טיפים:

1. טיפים לקריאה – טיפים אלה מסומנים על ידי הסימן , והם נועדו לכוון את הקריאה לפרקים בסיסיים יותר, ולהרחיק את הקורא הלא מנוסה מפרקים טכניים ומורכבים שעשויים לפגוע בתהליך הלמידה הראשוני.
2. טיפים ממומחה – טיפים אלה מסומנים על ידי הסימן , והם נועדו להפנות את תשומת הלב של הקורא להוראות או פקודות חשובות במיוחד, כמו גם להפנות את תשומת לבו לטעויות אופייניות בתכנות ב-SAS, בעיקר למתכנתים הנמצאים בתחילת דרכם.

קוראים שניהגו לפי טיפים אלה יוכלו למצוא בנקל את הדרך הנכונה והטובה ביותר לחלק את הקריאה של ספר זה לשני שלבים, ובכך יבטיחו למידה יעילה וטובה יותר של תכנות ב-SAS.

הסימן המוסכם	משמעות
<>	כל ביטוי המוכנס לתוך סוגריים משולשים מבטא מרכיב אופציונאלי בקוד, שאינו חובה, וניתן להשמיט אותו ללא פגיעה בתפקוד הבסיסי של המרכיב התפקודי
(קו אנכי)	סימן זה מייצג את המילה "או" והוא מפריד בין ביטויים הניתנים להחלפה בתוך הקוד
אותיות גדולות באנגלית	כל צעד (PROC) או הוראה (STATEMENT) בקוד SAS מיוצגת באותיות גדולות (להבדיל ממשתנים או אופציות של צעדים והוראות).

טבלה 1 – סימנים מוסכמים ומשמעותם

אודות התוכנה

SAS (Statistical Analysis System) היא מערכת משולבת של תוכניות מחשב שפותחה על ידי SAS Institute Inc.

SAS מאפשרת למשתמש לבצע, בין היתר:

א. ניהול נתונים :

1. קליטה וארגון של נתונים חדשים
2. טרנספורמציה של נתונים קיימים על מנת ליצור קבצי נתונים חדשים
3. שליפה של נתונים
4. יצירת דוחות

ב. ניתוח סטטיסטי :

1. הצגת נתונים בטבלאות שכיחות
2. הצגה גרפית
3. חישוב מתאמים
4. מבחנים סטטיסטיים
5. בדיקת השערות
6. משוואת מבניות (SEM)

ג. אחסון נתונים

ד. כריית מידע

1. מיצוי ידע מתוך קבצי נתונים גדולים
2. מידול

ה. OLAP (On Line Analytical Processing) – ניתוחים על קבצי נתונים רב ממדיים.

- ו. משחקים!!! למרות ש-SAS היא לא תוכנה גרפית, היא כוללת מספר משחקים מובנים. המשחקים כוללים סולייטר, בלאק ג'ק, פוקר, מכונת מזל, איקס-עיגול וסידור אותיות. כדי להגיע למשחקים ב-SAS, יש לבחור Solutions → Accessories → Games.

קובץ הנתונים

SAS עובדת עם קובץ נתונים ASCII, הבנוי ממשתנים ותצפיות. כברירת מחדל תצפיות בקובץ נתונים של SAS רשומות בשורות ומשתנים בעמודות. ניתן להפריד בין המשתנים על ידי רווחים, פסיקים, נקודות, וניתן גם לרשום משתנים (ותצפיות) ללא רווחים כלל (אך במצב כזה יש להגדיר לתוכנה באילו עמודות מופיע כל משתנה, כפי שיפורט בהמשך). לשם הנוחיות, ניתן לייבא קבצי נתונים שונים ל-SAS, כגון קובץ txt, excel, dat וכדומה).

בקובץ נתונים של SAS קיימים שני סוגים של משתנים :

- א. משתנים נומריים – משתנים בעלי ערכים מספריים.
- ב. מחרוזות (משתנים אלפאנומריים) – משתנים היכולים להכיל גם מספרים וגם אותיות (אך מטופלים כמשתנים שמיים בלבד).

איור 1 מציג קובץ נתונים בסיסי (המוצג ב-notepad) המכיל משתנה מחרוזת אחד ושני משתנים נומריים.

עבודה ב - SAS

העבודה ב-SAS נעשית בשתי דרכים :

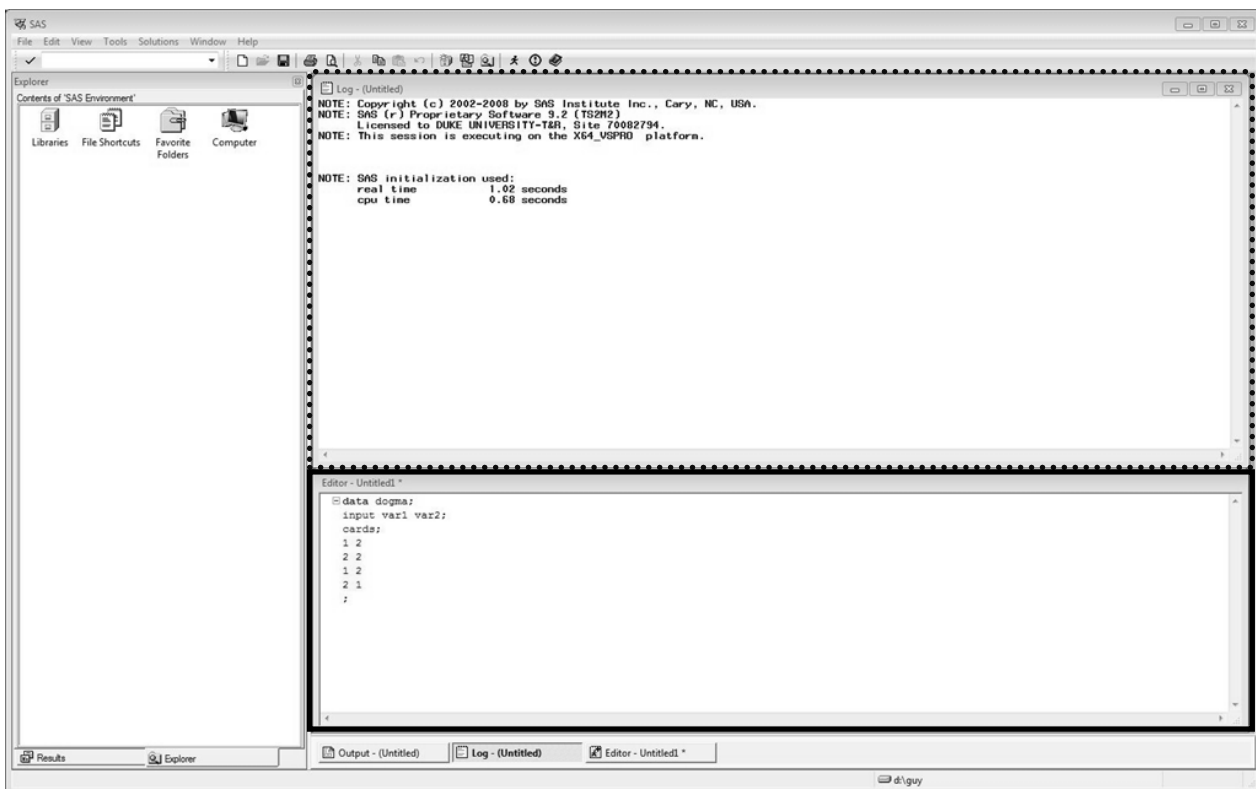
תכנות : בניגוד לתוכנות רבות אחרות, ב - SAS יש אפשרות לעבוד ללא שימוש בתפריטים, אלא לכתוב קוד עבור כל פעולה. תכונה זו של התוכנה מאפשרת גמישות רבה יותר ולפעמים אף מקצרת תהליכים בהשוואה לתוכנות מבוססות תפריטים, בהם כל הפרוצדורות קבועות מראש עפ"י מבנה התפריטים. ספר זה ייתמקד בעבודה ב-SAS באמצעות תכנות בלבד.

	trial	p_risk	p_risk2
1	t1	0.75	0.9
2	t2	0.65	0.45
3	t3	0.5	0.4
4	t4	0.7	0.6
5	t5	0.9	0.3
6	t6	0.75	0.25
7	t7	0.6	0.5
8	t8	0.75	0.55
9	t9	0.8	0.45
10	t10	0.75	0.45
11	t11	0.85	0.35
12	t12	0.9	0.4
13	t13	0.95	0.35
14	t14	0.85	0.35
15	t15	1	0.45
16	t16	0.8	0.4

איור 1 – דוגמא לקובץ נתונים המכיל משתנה מחרוזות ושני משתנים נומריים

תוכניות SAS בדרך כלל נכתבות, נבחנות ומורצות מתוך SAS Display Manager. מנהל התצוגה של SAS מורכב מחלונות. שלושת החלונות העיקריים הם:

1. החלון Editor (ראה מסגרת שחורה מלאה באיור 2)
2. החלון LOG (ראה מסגרת שחורה מקוקוות באיור 2)
3. החלון OUTPUT



איור 2 – מנהל התצוגה של SAS

כל אחד משלושת החלונות הללו נפתח כאשר מפעילים את SAS. כדי לעבור מחלון לחלון, יש ללחוץ על שם החלון מתוך שורת המשימות של התוכנה (ראה חלק תחתון באיור 2). לחילופין, ניתן לבחור בתפריט View או Window ולבחור את החלון הרצוי, או ללחוץ F5 בשביל החלון Editor, F6 בשביל החלון Log ו-F7 בשביל החלון Output.

עבודה עם תפריטים: SAS מאפשרת גם עבודה יותר אוטומטית עם תפריטים באמצעות SAS/ASSIST (על ידי בחירה בתפריט ASSIST → Solution). עם זאת, במסגרת ספר זה העבודה תתמקד בעבודות תכנות בלבד.

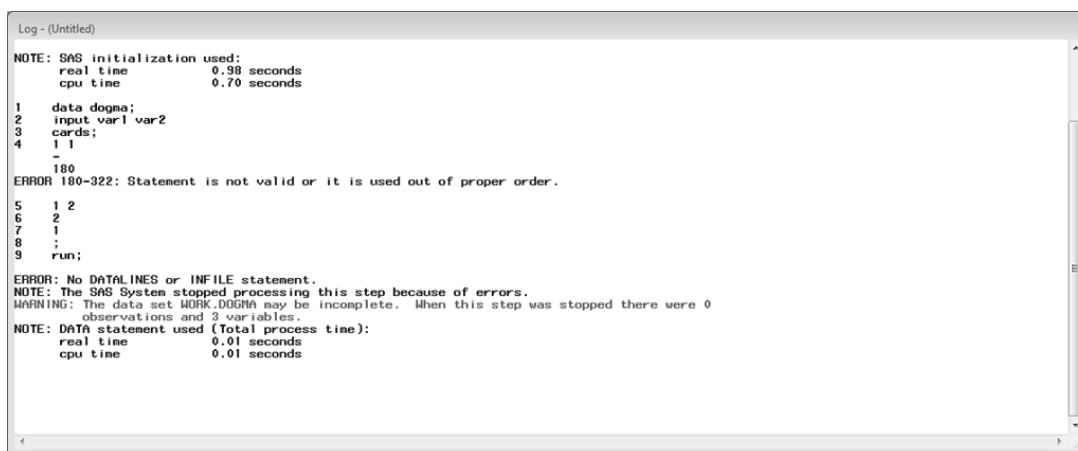
החלון EDITOR

במהלך עבודה ב-SAS, שורות קוד נכתבות לחלון Editor. כפי שיפורט בהמשך, בחלון Editor ניתן ליצור ולערוך קבצי נתונים, ולבצע פעולות (סטטיסטיות ואחרות) על נתונים באמצעות פרוצדורות מובנות ב-SAS, או באמצעות תכנות חופשי.

כדי לשמור את התוכנית הכתובה בחלון Editor, יש לוודא כי חלון זה הוא החלון הפעיל (לחיצה על F5). לאחר מכן, יש לבחור את התפריט File → Save As, ולתת שם לתוכנית. בדומה, כדי "לנקות" את התוכן של החלון Editor, יש לבחור את התפריט Edit → Clear All. לבסוף, כאשר רוצים לפתוח חלון Editor נוסף, יש לבחור את התפריט File → New Program.

החלון LOG

כאשר מריצים קוד ב-SAS, רשימה של הפעולות שנעשות על ידי התוכנה מוצגת בחלון Log. בנוסף, החלון Log מציג הערות לגבי קובץ הנתונים (כגון כמות המשתנים בהם נעשה שימוש בעיבוד וכמות התצפיות, משתנים בהם לא נעשה כלל שימוש וכדומה) ומידע כללי על תוכנת SAS (כגון גרסה נוכחית, פרטי הרישיון, זמן עיבוד הנתונים וכדומה). לבסוף, החלון Log מאפשר לעשות Debugging, שכן הוא מציג את השגיאות הקיימות בקוד.



```
Log - (Untitled)
NOTE: SAS initialization used:
      real time    0.98 seconds
      cpu time     0.70 seconds

1  data dogma;
2  input var1 var2
3  cards;
4  1 1
   -
   180
ERROR 180-322: Statement is not valid or it is used out of proper order.

5  1 2
6  2
7  1
8  ;
9  run;

ERROR: No DATA LINES or INFILE statement.
NOTE: The SAS System stopped processing this step because of errors.
WARNING: The data set WORK.DOGMA may be incomplete. When this step was stopped there were 0
observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time    0.01 seconds
      cpu time     0.01 seconds
```

איור 3 – החלון Log

הערות בקובץ Log מופיעות בטקסט כחול, אזהרות בטקסט ירוק ושגיאות בטקסט אדום. הקוד, הכולל את הפעולות שנעשו במהלך ההרצה, מופיע בטקסט שחור. איור 3 מציג דוגמא לסוגי הרשומות השונים המופיעים בחלון Log, ואת הצבעים האופייניים לכל סוג.

המידע המוצג בחלון Log הוא מצטבר. דהיינו, כל הרצה מוסיפה למידע שמוצג בחלון Log את המידע הרלוונטי להרצה הנוכחית, ולא מוחקת את המידע שהוצג על הרצות קודמות.

כדי לשמור את הרשומות המוצגות בקובץ Log, יש לוודא ראשית כי חלון זה הוא החלון הפעיל (F6). לאחר מכן, יש ללחוץ על התפריט File → Save As, ולתת שם לקובץ. את המוצג בחלון Log ניתן לשמור כקובץ Log ייעודי, קובץ RTF או קובץ DATA. כדי "לנקות" את תוכן החלון Log יש לבחור את התפריט Edit → Clear All או את התפריט File → New.

הפלט מההרצה של הקוד של SAS מוצג בתור קובץ נתונים בעל פורמאט ייחודי בחלון Output. אלא אם יישמר, הפלט המוצג בחלון Output נשמר כל עוד SAS מופעלת, אך הוא נמחק כאשר סוגרים את התוכנה.

בדומה לחלון Log, גם החלון Output מציג פלט מצטבר. כדי לשמור את הפלט המוצג בקובץ Output, יש לוודא ראשית כי חלון זה הוא החלון הפעיל (F7). לאחר מכן, יש ללחוץ על התפריט File → Save As, ולתת שם לקובץ. את המוצג בחלון Output ניתן לשמור כקובץ Output ייעודי, קובץ RTF או קובץ DATA. כדי "לנקות" את תוכן החלון Output יש לבחור את התפריט Edit → Clear All או את התפריט File → New. כמו כן, ניתן להדפיס את הפלט ישירות מהתוכנה, באמצעות בחירה בתפריט File → Print.

כתיבת תוכניות ב-SAS

כתיבת תוכניות ב-SAS נעשית באמצעות שפת תכנות פשוטה (SAS command language) מבוססת הוראות (Statements). כל הוראה אומרת למערכת SAS לבצע פעולה מסוימת, או מספקת מידע כלשהו.

מבנה הכתיבה (כמה כללי אצבע):

- ניתן לכתוב ב SAS תוך שימוש באותיות קטנות או גדולות (SAS איננה case sensitive).
- ניתן להתחיל לכתוב הוראות (שורות קוד) בכל מקום של השורה.
- הוראה יכולה להימשך אל מעבר לשורה אחת.
- ניתן לכתוב מספר שורות קוד באותה שורה, כאשר שורות הקוד השונות מופרדות על ידי נקודה פסיק.
- כל הוראה (שורת קוד) ב-SAS חייבת להסתיים בסימן ; (נקודה פסיק). אחרת, SAS מתייחסת לשורה כאל תחילתה של השורה הבאה.
- כדי להפוך שורות קוד ללא פעילות (למשל כדי להכניס הערות), יש לנקוט באחת משתי הפעולות הבאות:
 - להכניס את הסימן * (כוכבית) בתחילת השורה. במצב כזה, כל שורת הקוד (על לסימן ה - ; הקרוב) יצבע בצבע ירוק, והתוכנה תתעלם משורה זו.
 - להכניס את הסימן /* בתחילת קטע, ואת הסימן */ בסוף קטע. במצב כזה כל שורות הקוד בין שני הסימנים יצבעו בירוק, והתוכנה תתעלם מקטע זה.
- שם של משתנה חייב להתחיל באות.
- שם משתנה יכול לכלול אותיות, ספרות ואת הסימן _ (קו תחתון). שם המשנה יכול להכיל את תו הרווח, בתנאי שמגדירים זאת (כפי שיפורט בהמשך).
- נקודות מסמנות ערכים חסרים במשתנים נומריים, ורווחים מסמנים ערכים חסרים במשתנים אלפאנומריים. SAS הופכת אוטומטית ערכים חסרים (רווחים) במשתנים נומריים לנקודות.

מבנה התוכנית:

- תוכנית SAS היא סדרה של שורות קוד המורות למערכת לבצע משימות שונות או עיבודים שונים. ההוראות השונות מופיעות בשני סוגים של צעדים (steps): DATA STEP ו-PROC STEP (כפי שיפורט בהמשך). בסוף כל צעד צריכה להופיע הפקודה; RUN. עם זאת, ניתן לכלול כל צעד מספר פעמים בתוכנית SAS אחת.



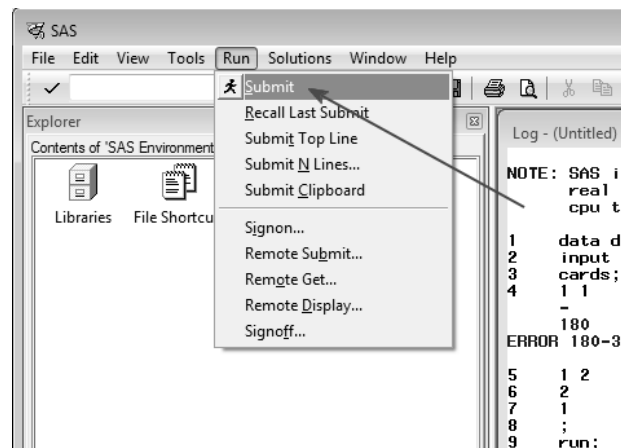
טיפ ממומחה: ב-SAS, למרכיבים שונים של הקוד יש צבעים שונים: הגדרת צעד (PROC או DATA) מופיע בקוד בצבע כחול כהה, הוראות ואופציות מופיעות בצבע כחול בהיר, מחרוזות מופיעות בצבע סגול, ומספרים מופיעים בצבע ירוק. לכן, בחינה של הקוד לפני ההרצה (לוודא שמה שאמור להיות בצבע כחול הוא כחול, ומה שאמור להיות ירוק הוא ירוק וכדומה) יכולה לעזור ולמצוא טעויות בהקלדה, ומהווה סימן טוב לבדיקה ראשונית שלו.

הרצת התוכנית :

כאשר מסיימים לכתוב תוכנית ב-SAS, יש להריץ אותה. ניתן להריץ תוכנית ב-SAS או מתוך חלון התוכנה הראשי או מתוך החלון Program editor, על פי הפירוט הבא :

מתוך חלון התוכנה :

- לחיצה על "Run" מתוך תפריט הפקודות ובחירה בפקודה "Submit" (ראה איור 4).
- לחיצה על המקש של האיש הרץ (ראה איור 5).
- לחיצה על המקש F8 (יעבוד רק כאשר החלון של SAS הוא החלון הפעיל).




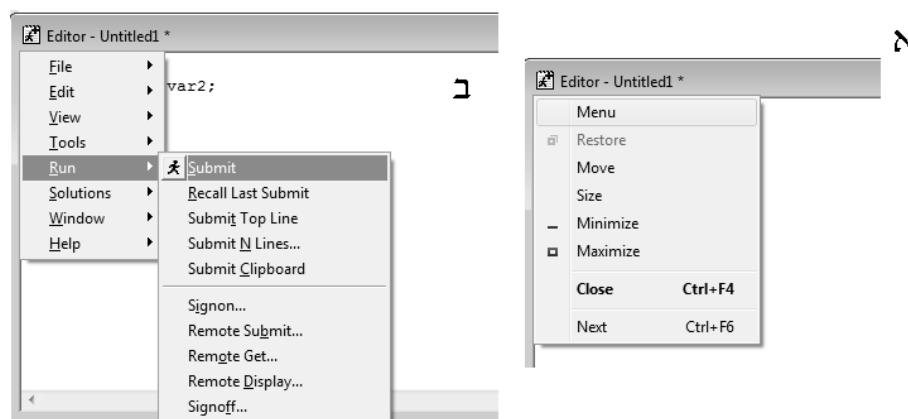
איור 4 – הרצת קוד SAS דרך התפריט הראשי



איור 5 – הרצת קוד SAS דרך סרגל הכלים

מתוך החלון Program editor :

- לחיצה על הפינה השמאלית העליונה של החלון Editor (על סימן ה- ) , ובחירה ב-Menu→Run→Submit (ראה איור 6).
- כאשר רוצים להריץ חלקים מסוימים בלבד של התוכנית : סימון החלקים הרלוונטיים עם העכבר ("צובע" את שורות הקוד בשחור), וחזרה על אחד מהשלבים להרצת תוכנית שתוארו לעיל.



איור 6 – הרצת קוד SAS מתוך החלון Editor

פרק 2

יצירת קבצי נתונים

ה-DATA STEP

הצעד DATA (DATA STEP) מכיל את כל ההוראות הדרושות ליצירת קובץ נתונים, טיפול בנתונים וכדומה. צעד זה כולל הוראה המציינת את שם קובץ הנתונים ומיקומו, את שם המשתנים ומיקומם בקובץ, את סוג המשתנים, כמו גם מאפיינים נוספים הכוללים ציון דרך הטיפול בנתונים חסרים, ביצוע שינויים (או טרנספורמציות) במשתנים קיימים, הגדרת משתנים חדשים, סינון תצפיות, בחירת משתנים וכדומה.

ה-DATA STEP יוצר מהנתונים הקיימים קובץ בפורמט SAS שהתוכנה יודעת לקרוא.
ה-DATA STEP מתחיל בהוראה DATA ומסתיים בהוראה RUN:

```
DATA שם קובץ הנתונים ;  
.....  
.....  
run;
```

ישנם שני סוגים של קבצי נתונים (dataset) ב-SAS:

1. קובץ נתונים זמני – קובץ נתונים הנשמר בספרייה work וקיים כל עוד תוכנת SAS רצה.
2. קובץ נתונים קבוע – קובץ נתונים הנשמר בספרייה מוגדרת על ידי המשתמש ונשמר גם אחרי ש-SAS איננה רצה.

ההוראה DATA

הוראת DATA יוצרת את קובץ הנתונים ב-SAS ונותנת לו שם המוגדר על ידי המשתמש.
אופן הכתיבה:

```
DATA שם כלשהו ;
```

דוגמא:

```
data dogma;
```

קריאת קובץ נתונים חיצוני

ההוראה INFILE

ההוראה INFILE מציינת את שם קובץ הנתונים החיצוני ואת מיקומו במחשב, והיא מאפשרת ל-SAS לקרוא את הנתונים הנמצאים בקובץ ולטעון אותם לזיכרון העבודה של התוכנה. ההוראה INFILE חייבת להופיע אחרי ההוראה DATA, והיא

חייבת לכלול את שם הקובץ ונתיבו (בתוך גרשיים). בנוסף, ההוראה INFILE יכולה לכלול גם אופציות שונות, כפי שיפורט בהמשך.
אופן הכתיבה:

<אופציות שונות> 'שם הקובץ ונתיבו' INFILE;

דוגמא:

```
data dogma; infile 'c:\course\data.txt';
```

אופציות של ההוראה INFILE

האופציות של ההוראה INFILE מאפשרות לציין מאפיינים שונים של קובץ הנתונים החיצוני או להגדיר פעולות שהוראת DATA תבצע עליו. כל האופציות נכתבות אחת אחרי השנייה לאחר שם הקובץ ומיקומו, כאשר האופציות מופרדות על ידי רווחים.

1. האופציה delimiter – אופציה זו מציינת את סוג התו המשמש כמפריד בין המשתנים (נקודה, פסיק, רווח וכדומה). כאשר אופציה זו לא מוגדרת, SAS מניח כי המשתנים מופרדים על ידי רווחים.
אופן הכתיבה:

'סוג התו המפריד' dlm =

דוגמא לקריאת קובץ המופרד על ידי טאבים:

```
data class1; infile 'c:\course\data.txt' dlm = '09'x;
```

דוגמא לקריאת קובץ המופרד על ידי נקודה פסיק:

```
data class1; infile 'c:\course\data.txt' dlm = ';';
```



טיפ ממומחה: הטעות הנפוצה ביותר של מתחילים ב SAS היא כתיבת פקודות ללא ; (נקודה פסיק) בסיומן

2. האופציה dsd – לאופציה זו יש שלושה תפקידים כאשר SAS קוראת קובץ חיצוני בעל משתנים מופרדים. התפקיד הראשון הוא "להפשיט" מרכאות המקיפות את ערכי המשתנים בתוך הקובץ (דבר האופייני מאוד לקבצים המופקים לדוגמא מתוכנת ויזואל בייסיק). התפקיד השני קשור לערכים חסרים. כאשר SAS נתקלת בתוך הקובץ החיצוני במפרידים עוקבים בתוך קובץ (לדוגמא ; ; בקובץ המופרד על ידי נקודה פסיק), ברירת המחדל של התוכנה תהיה להתייחס למפרידים אלה כאל יחידה אחת. אולם, בדרך כלל מפרידים עוקבים מציינים ערכים חסרים למשתנים. האופציה DSD אומרת ל-SAS להתייחס למפרידים עוקבים בנפרד, כך שערך החסר בין שני מפרידים עוקבים יוגדר ככזה על ידי התוכנה. התפקיד השלישי מניח שהמפריד בין ערכי המשתנים הוא פסיק. לכן, אם נשתמש באופציה DSD במצב בו התו המפריד הוא אכן ",", אין צורך להשתמש גם באופציה DLM. לעומת זאת, אם המפריד הוא לא התו ",", יש צורך להשתמש גם באופציה DLM במקביל.
אופן הכתיבה:

dsd