

סילבוס: מודלים סטטיסטיים - ב סמסטר אביב – תשס"ז

מרצים: ארתור צ'ירגייב, יוני נצרתי
מתרגלים: ולדימיר גריסקין, פאולינה לרנר.
אתר הקורס: http://stat.haifa.ac.il/~yonin/stat_models_B_course_spring_07/stat_models_B.html

כללי:

קורס זה הוא קורס המשך לקורס מודלים סטטיסטיים א'. במודלים סטטיסטיים א' הסטודנט נפגש לראשונה עם מודלים של רגרסיה ליניארית. בנוסף הסטודנט רכש מיומנות בתפעול תוכנת SAS Enterprise ככלי לניתוחים סטטיסטיים והפנים את הטכניקות והתיאוריה הקשורות לבדיקת הנחות המודל. בקורס זה נלמדים מודלים סטטיסטיים נוספים: מודלים של ניתוח שונות, רגרסיה לוגיסטית, ומספר דוגמאות פשוטות לשיטות א-פרמטריות.

באופן דומה לקורס מודלים סטטיסטיים א', עבור כל מודל סטטיסטי, הסטודנט ירכוש ידע ומיומנות בשלושה היבטים שונים:

- א) **היבט התיאוריה** – כאן הסטודנט נדרש להבין את התיאוריה הבסיסית המצדיקה את החישובים והאנליזה המלווה כל מודל. חשוב מאוד שכל סטטיסטיקאי יבין את התיאוריה המלווה את העיבודים והניתוחים אשר הוא מבצע על הנתונים. הבנה זו מבטיחה שימוש מושכל ונכון בכלים הסטטיסטיים אשר ברשותו.
- ב) **היבט היישום** – הסטודנט נדרש להבין כיצד להתאים מודלים לסיטואציות שונות, ומהם השיקולים בכך. כיצד לנסח מבחני השערה, כיצד לקרוא פלטים של ניתוחים סטטיסטיים וכו'. בנוסף יוצגו מושגים בסיסיים הקשורים לתכנון ניסויים סטטיסטיים.
- ג) **היבט הביצוע** – הסטודנט נדרש לרכוש ניסיון בביצוע ניתוחים סטטיסטיים עבור כל מודל שנלמד. הביצוע יתבצע באמצעות תוכנת SAS Enterprise תוך הצגת פקודות שפת SAS ו STAT/SAS הנוצרות כאשר מתפעלים את SAS Enterprise (SAS code). בנוסף יתבצעו מספר חישובים ידניים לצורך המחשה.

הרכב הציון:

- 3 תרגילי בית ביצועיים (משתמשים ב SAS). הגשת התרגילים היא בזוגות. משקל כל תרגיל הוא 5% מהציון (סה"כ שלושת התרגילים 15%). הנתונים לצורך התרגיל אשר יינתנו לכל זוג יהיו ייחודיים עבור אותו זוג, לא ניתן להעתיק.
- בוחן ביניים. משקל הבוחן הוא 15% מהציון. כל הסטודנטים הרשומים לקורס (וכאלו שמסיבות כלשהן עדיין לא השלימו את הרישום, חייבים להבחן בבוחן). ציון עבור מבחן חסר הוא 0.
- מבחן סוף סמסטר בכיתת מחשב. חלק מהמבחן הוא ממוחשב (הסטודנט צריך לתפעל את SAS) חלק לא. משקל המבחן הסופי הוא 70%.

תוכנית לימוד מפורטת:

חלק א: כלים בסיסיים ומבוא להשוואות תוחלות			
פרק	כותרת	תוכן בהרצאה	תוכן בתרגול
א-0	מבוא לקורס.	<ul style="list-style-type: none"> • סילבוס. • ההיבט התיאוריה, היבט היישום, היבט הביצוע. 	-
א-1	מושגים סטטיסטיים, חזרה.	<ul style="list-style-type: none"> • מהו מבחן השערה? • P-values. • התפלגות נורמאלית, חי, t ו- F. 	<ul style="list-style-type: none"> • הצגה של חישוב p-value ואחוזונים ע"י שיטות נומריות, השוואה מול טבלאות. • Descriptive Analysis SAS Enterprise ב
א-2	השוואות תוחלות שתי אוכלוסיות ב"ת.	<ul style="list-style-type: none"> • תיאור היבט היישום. • תיאור היבט התיאוריה עבור מקרים שונים שונות ידועה/ לא ידועה ושווה/לא שווה, הנחת הנורמאליות במודל, מדגמים מאוזנים/לא מאוזנים. • ניתוח פלט SAS. 	<ul style="list-style-type: none"> • חזרה על היבט היישום. • ביצוע בSAS וניתוח פלט.
א-3	שימוש בשיטה א-פרמטרית: מבחן Mann-Whitney.	<ul style="list-style-type: none"> • שיטות א-פרמטריות, מדוע ולמה? • תיאור היבט התיאוריה והיבט היישום של המבחן. 	<ul style="list-style-type: none"> • חישוב ידני לדוגמא קטנה. • ביצוע בSAS וניתוח פלט.
א-4	שימוש בבלוקים: מבחני t מזווגים (תלויים)	<ul style="list-style-type: none"> • עקרון השימוש בבלוקים. • תיאור היבט היישום של מבחן t לאוכלוסיות תלויות. • תיאור היבט התיאוריה עבור מקרים שונים שונות ידועה/ לא ידועה ושווה/לא שווה, הנחת הנורמאליות במודל, מדגמים מאוזנים/לא מאוזנים. 	<ul style="list-style-type: none"> • חזרה על היבט היישום. • ביצוע בSAS וניתוח פלט. • ביצוע
א-5	קביעת גודל המדגם – חישובי עוצמה.	<ul style="list-style-type: none"> • עוצמת המבחן. • Operating Characteristic Curves. 	-
תרגיל בית ביצועי (SAS) ראשון: השוואת 2 אוכלוסיות			
חלק ב: השוואות תוחלות של מספר אוכלוסיות - יסודות של ניתוח שונות.			
פרק	כותרת	תוכן בהרצאה	תוכן בתרגול
ב-1	ניתוח שונות חד-כווני לטובת מה?	<ul style="list-style-type: none"> • הצגת המודל – השוואה למודל של מבחן t לא מזווג. • הבעיות בביצוע a-1 אוסף מבחני t. 	-
ב-2	החשבון של ניתוח שונות: פירוק סכומי ריבועים.	<ul style="list-style-type: none"> • פרוק סכום הריבועים עבור המקרה המאוזן. • המשמעות האלגברית של דרגות חופש. 	<ul style="list-style-type: none"> • פרוק סכום הריבועים עבור המקרה הלא מאוזן.
ב-3	הסטטיסטיקה של ניתוח שונות	<ul style="list-style-type: none"> • פילוג סכומי ריבועים ומנות של סכומי ריבועים. 	-

	<ul style="list-style-type: none"> • תוחלות של MS. • פילוג סטטיסטי המבחן תחת H_0. • הקשר למבחן t. 		
<ul style="list-style-type: none"> • ביצוע דוגמא ידנית. • ביצוע דוגמאות ב SAS. 	<ul style="list-style-type: none"> • טבלת ניתוח שונות. • נוסחאות חישוב (לצורך חישובים ידניים נוחים). • ביצוע דוגמא ידנית. • ביצוע הדוגמא ב SAS. 	ביצוע ניתוח שונות חד-כווני	ב-4
<ul style="list-style-type: none"> • תרגיל רישום המודל כמודל רגרסיה עבור מקרים פרטיים. 	<ul style="list-style-type: none"> • חזרה קצרצרה על משתני דמי במודל רגרסיה לינארית. • ניסוח ניתוח שונות דו-כווני כמודל רגרסיה. • הדגמא ב SAS לשוויון פלטים. 	הקשר למודל רגרסיה לינארית	ב-5
<ul style="list-style-type: none"> • ביצוע תרגילים לניסוח שאלות מעשיות כמבחני השערה לגבי קונטרסטים. • ביצוע דוגמאות ב SAS. 	<ul style="list-style-type: none"> • הצורך היישומי בהשוואות מרובות. • רווח סמך עבור השוואה אחת. החיסרון בשיטה הנאיבית (דעיכת רמת המובהקות). • שיטת LSD. • קונטרסטים וקונטרסטים אורתוגונאליים. • ביצוע דוגמא SAS עבור קונטרסטים. 	מבוא להשוואות מרובות.	ב-6
<ul style="list-style-type: none"> • ביצוע דוגמאות ב SAS. • שאלות תיאורטיות ממבחנים. 	<ul style="list-style-type: none"> • שיטת Scheffe להשוואת כל הקונטרסטים. • Bonferroni ותיקון Sidack • Tukey • Duncan • Dunett 	השוואות מרובות – שיטות נוספות.	ב-7
<ul style="list-style-type: none"> • תרגילים לניתוח וטרנספורמציות. • דוגמאות SAS. • תרגילי קריאת פלט. 	<ul style="list-style-type: none"> • תיאור קצר (בונה על קורס א'). • טרנספורמציות. 	בדיקת הנחות המודל, אבחון המודל וטרנספורמציות.	ב-8
<ul style="list-style-type: none"> • ביצוע דוגמא ידנית. • ביצוע דוגמא ב SAS. 	<ul style="list-style-type: none"> • סקירה קצרה כהכללה של מבחן Mann-Whitney. 	שיטות א-פרמטריות: Kruskal-Wallis	ב-9
הרצאת אורח (45דק): מר עידו קינרייך ממכון טכנוסטט			
תרגיל בית ביצועי (SAS) שני: ניתוח שונות חד-כווני			
חלק ג: ניתוח שונות רב-גורמי ואינטראקציות			
תוכן בתרגול	תוכן בהרצאה	כותרת	פרק
<ul style="list-style-type: none"> • המשך Descriptive Analysis ב SAS Enterprise (לנתונים של ניתוח שונות דו-כונית). 	<ul style="list-style-type: none"> • מה בודקים בניתוח שונות דו-כווני. • החיסכון בניסויי ניתוח שונות דו-כווני. 	יסודות ניתוח שונות גורמים מרובים ואינטראקציות.	ג-1
-	<ul style="list-style-type: none"> • דוגמאות למשמעות של אינטראקציות. • הצגות גראפיות שונות לצורך זיהוי אינטראקציות בנתונים ובמודל. 	עוד על אינטראקציות.	ג-2

3-ג	ביצוע ניתוח שונות דו-כווני	<ul style="list-style-type: none"> הצגה קצרה של פרוק סכום הריבועים. הצגה קצרה של התכונות הסטטיסטיות. טבלת ניתוח השונות. הצגה קצרה לנוסחאות החישוב. קריאת פלט דוגמא. 	<ul style="list-style-type: none"> ביצוע דוגמא ב SAS.
4-ג	סוגיות נוספות בניתוח שונות דו-כווני.	<ul style="list-style-type: none"> דיון קצר בנושא השוואות מרובות. דיון קצר לגבי בדיקת הנחות המודל. נתונים לא מאוזנים אבל פרופורציונאליים. שיטות מקורבות עבור נתונים לא מאוזנים. הצגה של מודל דו-כווני עם תצפית אחת בכל תא. מבחן Tuckey לבדיקת אי-קיום אינטראקציה במודל עם תצפית אחת. 	<ul style="list-style-type: none"> שאלות ממבחנים.
5-ג	מודל בלוקים פשוט.	<ul style="list-style-type: none"> תיאור של התפיסה של מודל בלוקים (הרחבה על מבחני t מזווגים). ניתוח מודל בלוקים כמודל ניתוח שונות דו-כווני ללא אינטראקציות עם תצפית אחת בכל תא. 	<ul style="list-style-type: none"> ביצוע דוגמא ב SAS + שאלות קריאות וניתוח פלטים.
6-ג	מספר גורמים כללי ומבוא לתיכון ניסויים.	<ul style="list-style-type: none"> הצגה קצרה של ניתוח שונות תלת כווני. סקירה של מבוא לתכנון ניסויים 	-
חלק ד: מודלים נוספים של ניתוח שונות			
פרק	כותרת	תוכן בהרצאה	תוכן בתרגול
1-ד	ניתוח שונות חד-כווני עם אפקטים אקראיים.	<ul style="list-style-type: none"> מודל האפקטים האקראיים. שימושים לאפקטים אקראיים. ביצוע דוגמא. רווחי סמך לשונניות. 	-
2-ד	ניתוח שונות דו-כווני עם אפקטים אקראיים ומעורבים.	<ul style="list-style-type: none"> ואריאציות של מודלים מעורבים. הדגמה ב SAS. 	<ul style="list-style-type: none"> ביצוע דוגמא ב SAS + שאלות הבנה הקשורות לקריאת הפלט.
3-ד	ניתוח שונות חד-כווני עם Repeated Measures.	<ul style="list-style-type: none"> תיאור הצורך ב Repeated Measures. ביצוע דוגמא ב SAS. 	<ul style="list-style-type: none"> פתרון שאלות ממבחנים ומספרים על חלקים ג' + ד'.
בוהן (ללא ביצוע ב SAS): מודלים של ניתוח שונות			
חלק ה: רגרסיה לוגיסטית			
פרק	כותרת	תוכן בהרצאה	תוכן בתרגול
1-ה	מודל הרגרסיה הלוגיסטית.	<ul style="list-style-type: none"> הצורך במשתנים מוסברים בדידים תיאור המודל. פרוש משמעות המודל. פתרון משוואות עבור MLE בשיטות נומריות. תיאור קצר של המקרה הרב-מימדי (ללא ישום). 	-
2-ה	ביצוע רגרסיה לוגיסטית.	<ul style="list-style-type: none"> ביצוע רגרסיה לוגיסטית ב SAS (חד-מימדי). 	<ul style="list-style-type: none"> ביצוע דוגמאות ב SAS.

ה-3	הסקה על המודל.	<ul style="list-style-type: none"> מבחני השערה ורווחי סמך לגבי הפרמטרים של המודל. 	<ul style="list-style-type: none"> קריאת פלטים של דוגמאות ושאלות ממבחנים ומספרים.
ה-4	טיב התאמת המודל	<ul style="list-style-type: none"> השטח מתחת ל ROC סקירה קצרה של שיטות נוספות. 	<ul style="list-style-type: none"> ביצוע דוגמאות ב SAS של ROC + שאלות ממבחנים ומספרים.
תרגיל בית ביצועי (SAS) שלישי: רגרסיה לוגיסטית			
מבחן סוף סמסטר			

הערה: בהנחה שלא יהיו שביתות או מלחמות, כל תוכנית הלימוד תושלם.

רשימה ביבליוגרפיות:

- D. C. Montgomery, *Design and Analysis of Experiments, Fourth Edition*, 1997.
- D. W. Hosmer, S. Lemeshow, *Applied Logistic Regression*, 2000.
- J. A. Rice, *Mathematical Statistics and Data Analysis*,
- R. Christensen, *Analysis of Variance, Design and Regression*, 1996.
- W. Mendenhall, T. Sincich, *A Second Course in Statistics, Regression Analysis, Fifth Edition*, 1996.
- Bowerman, O'Connell, *Linear Statistical Models, an Applied Approach*, 1990.

הערה: סטודנטים אשר מעוניינים להמשיך לתואר שני: מומלץ להתייחס ברצינות לספרות המומלצת.

בהצלחה.