

**האוניברסיטה העברית בירושלים**  
**מחלקה לסטטיסטיקה**

**מודלים סטטיסטיים ויישומיים (52518)**  
**פרופ' דוד צוקר**  
**תשע"ה סמסטר א', מועד א'**

תאריך : כ"א בשבט תשע"ה, 10.2.15  
משך הבחינה : שעתיים וחצי

חומר עזר : מחשב CIS ודף רשימות (שני דפים בגודל A4, שני צדדים)  
 חל אייסור מוחלט להעתיק. תלמיד שייתפס יורחק לשנה מלימודיו.

בצלחה!!

**שאלה 1 (30 נקודות)**

נתיחס לניטוח שוונוט דו-כיווני לא מאוזן בלי אינטראקציה, דהיינו המודל

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

$$\epsilon_{ijk} \sim N(0, \sigma^2) \text{ independent}$$

$$i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, n_{ij}$$

נתיחס לكونרטטים מהצורה  $\psi(c) = \sum_{i=1}^I c_i \alpha_i$  (עם  $c = (\sum_{i=1}^I c_i)$ ).

[15] א. הוכיחו כי הערך של  $\psi(c)$  אינו תלוי במערכות המשקלות שימושים בהן לנתח את המודל.

רמז : זכרו את ההגדרה של  $\alpha_i$

[15] ב. פתחו נוסחה לר"ס עבור  $\psi$  כקונרטט יחיד. יש להגדיר את כל הסימונים לצורך מפורטת.

**שאלה 2 (20 נקודות)**

נתיחס למודל

$$Y_{ijk} = \mu + a_i + b_{ij} + \epsilon_{ijk}$$

$$a_i \sim N(0, \sigma_a^2), b_{ij} \sim N(0, \sigma_b^2), \epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$$

$$\text{all random variables independent}$$

$$i = 1, \dots, 20; j = 1, 2; k = 1, 2, 3$$

רשמו את המודל בצורה  $\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i$ . יש להגדיר את כל הסימונים לצורך מפורטת.

### שאלה 3 (25 נקודות)

להלן נתונים מסקר בנוגע מידת הה תלבות מפורדים. המשתנים הינם ה תלבות (נוכחה/בינוי/גבואה), גיל (עיר/מגור), ומין (זכר/נקבה). עבור המודל לפיו ה תלבות **בלתי תליה** **במיון בהינתן גיל**, רשמו את הערכיהם **המספריים** של אומדי ניראות המרבית עבור כל הפרמטרים מסווג. ניתן להסתמך על תוצאות שהוחחו בשיעורים או תרגילים.

הנתונים :

#### גיל

	עיר		מגור	מין	התלבות
נוכחה	20	5	זכר	נקבה	נוכחה
	3	4	נקבה		
בינוי	20	20	זכר	נקבה	בינוי
	15	16	נקבה		
גבואה	10	30	זכר	נקבה	גבואה
	12	20	נקבה		

### שאלה 4 (10 נקודות)

נתיחס לבעה של ניתוח שונות חד-כיווני.

[5] א. רשמו את השערת האפס ב מבחן לויין.

[5] ב. מה המשמעות של דחיפה ב מבחן לויין מבחן הניתוח (1-2 משפטים)?

### שאלה 4 (15 נקודות)

נתיחס לחומר בנספח עבור הרצה של רגרסיה לוגיסטיבית.

נסמן ב-  $\mathbf{x}_1$  את המצב term=24, amount=10000, age=30, land=0

וב-  $\mathbf{x}_2$  את המצב term=36, amount=15000, age=40, land=1

.  $p(\mathbf{x}) = \Pr(Y = 1 | \mathbf{X} = \mathbf{x})$  בניסוף, עבור כל  $\mathbf{x}$  אפשרי נסמן

[7] א. חשבו ר"ס ברמה 95% עבור  $p(\mathbf{x}_1)$ .

[8] ב. חשבו ר"ס ברמה 95% עבור  $\{p(\mathbf{x}_2) / (1 - p(\mathbf{x}_2))\} / \{p(\mathbf{x}_1) / (1 - p(\mathbf{x}_1))\}$

## נספח

Data on 1000 loan customers of a German bank.

The variables are as follows:

gc: 0/1 indicator of whether the person is a good risk for the requested loan (1) or not 0  
term: duration of requested loan  
amount: amount requested (in German Marks)  
age: age of the customer (years)  
land: 0/1 indicator of whether the customer owns land (1) or does not (0)

```
lterm = log(term)
amtlog = log(amount)
l.age = log(age)
```

### Results of Logistic Regression Analysis

	Coef	S.E.	Wald Z	Pr(> Z )
Intercept	-1.1179	1.1539	-0.97	0.3327
lterm	-0.9297	0.1714	-5.43	<0.0001
amtlog	0.1870	0.1230	1.52	0.1284
l.age	0.9560	0.2565	3.73	0.0002
land	-0.7341	0.1980	-3.71	0.0002

### Covariance Matrix of Estimated Coefficients

	Intercept	lterm	amtlog	l.age	land
Intercept	1.33159839	0.0137349459	-0.0752634941	-0.2258012175	0.0688971954
lterm	0.01373495	0.0293662476	-0.0134884945	0.0013452636	-0.0009395372
amtlog	-0.07526349	-0.0134884945	0.0151238319	-0.0007412548	-0.0036072797
l.age	-0.22580122	0.0013452636	-0.0007412548	0.0658113459	-0.0128218795
land	0.06889720	-0.0009395372	-0.0036072797	-0.0128218795	0.0392093907