



**מספר קורס:** 4040243

**שם הקורס בעברית:** מבוא לסטטיסטיקה למדעי הרוח

**שם הקורס באנגלית:** Introduction to Statistics for the Humanities

**שנה"ל:** תשפ"ד

**שם המרצה:** ד"ר גלעד יעקובסון

**שם תכנית הלימודים:** התכנית ללימודים הומניסטיים

**סיווג הקורס:** שיעור

**משך והיקף הקורס:** סמסטר ב', 2 ש"ס

**נק' זכות אקדמיות:** 2 נ"ז

**דרישות קדם:** ציון סופי של 75 ומעלה בקורס: מבוא לחשיבה כמותית למדעי הרוח

**שעות הקבלה:** ימי שלישי בין השעות 11:30-12:30 (שעות אחרות באותו יום / ימים אחרים

בתיאום מראש)

**פרטי קשר:** אימייל: [giladj@gmail.com](mailto:giladj@gmail.com)

#### **תיאור הקורס:**

הקורס יכלול הן הוראה פרונטלית, והן "מיני סדנאות" במהלך השיעור, בהן נעבוד עם נתוני אמת על המחשב. מיני-סדנאות אלה יסייעו בידינו לפתח אינטואיציה שאותה נחדד אחר כך בכלים מתמטיים.

הקורס יכסה את הנושאים הבאים:

- מבוא להסתברות (המשך מקורס תנאי הקדם) + הדגמה בכיתה.
- סטטיסטיקה תיאורית: מדדי מיקום מרכזי ומדדי פיזור (ממוצע, סטיית תקן, שגיאת תקן, חציון, אחוזונים, פסילת נתוני קיצון ועוד).
- מהו מדגם? מה הקשר בין תכונות המדגם לבין תכונות של ההסתברות התיאורטית? כמה מרוויחים מכך שהמדגם גדל?
- הצגה גרפית של נתונים.
- כשלים לוגיים ומתודיים באיסוף ובניתוח נתונים סטטיסטיים.
- הסקה סטטיסטית: מהם מבחנים סטטיסטיים? כיצד משתמשים בהם באופן מושכל?
- יישומים סטטיסטיים במדעי הרוח: מי חיבר את מסמכי הפדרליסט? איך משתמשים בחכמת המונים לתעתוק טקסטים מהגניזה הקהירית? איך מזהים את השפה בה כתוב מסמך? ועוד.
- יישומים סטטיסטיים בעידן הרשתות החברתיות וה-Deep Learning: ההסתמכות של אלגוריתמים מתאימי-תוכן על נתונים סטטיסטיים. ביג-דאטה וסטטיסטיקה. למידת מכונה. יישומים פשוטים יותר (דחיסת קבצים, דחיסת דיבור בתקשורת ניידת...).

הלמידה תכלול ניתוח נתונים אמיתיים, ומעקב אחרי הישגים וכשלים סטטיסטיים בספרות המדעית / רפואית / מחקרית (כולל מדעי הרוח). כמו-כן יינתנו תרגילי בית שיחייבו התמודדות בסיסית עם

טבלאות נתונים והצגה גרפית שלהם. התכנה שתשרת את הסטודנטיות והסטודנטים בקורס תהיה EXCEL.

### **מטרות הקורס:**

- לפתח אצל הסטודנטים אינטואיציה עמוקה לגבי תכונות של מדגמים, והדרך שבה הן יכולות ללמד אותנו על תכונותיהן של הקבוצות הנדגמות. בפרט, יושם דגש רב על הבנת כשלים לוגיים וטכניים אופיינים בסטטיסטיקה.
- לפתח אצל הסטודנטים כלים לחישוב תכונות בסיסיות של מדגמים (ממוצע, סטיית/שגיאת תקן, אחוזונים...) והבנה של משמעות מדדים אלו, ורגישותם לאופן בו נתונים נדגמים ולגודל המדגם.
- להנחיל כלים להצגה גרפית יעילה ובהירה של נתונים.
- לאפשר התנסות בכלים סטטיסטיים בהקשרים של מדעי הרוח ("מדעי הרוח הדיגיטליים").

המוטו של הקורס, והפילוסופיה שתלווה אותו לכל אורך הדרך, היא: כיצד לדווח את מרב האינפורמציה על נתונים, תוך שימוש בתיאור הקצר ביותר (דהיינו, על ידי מספרים ספורים).

### **ספרי עזר:**

הקורס נבנה באופן ייחודי עבור הסטודנטים של המרכז האקדמי שלם, ומעניק, מעבר לחומר הלימוד המקובל בקורסי סטטיסטיקה, דיון ביישומים מודרניים של סטטיסטיקה במדעי הרוח ובתעשיית ההיי-טק. הקדמה משמעותית לתורת ההסתברות ניתנה בקורס תנאי הקדם שעסק בחשיבה כמותית. ספר לימוד שימושי כליווי לקורס הוא:

זמיר ובייט-מרום, מבוא לסטטיסטיקה לתלמידי מדעי החברה א' (או"פ 1993).

באנגלית קיים מגוון עצום של ספרי לימוד שצר המקום מלסקור. לבעלי רקע מתמטי נרחב (חדו"א ולאגברה לינארית), המעוניינים להעמיק בנושא **מעבר** לדרישות הקורס, מומלץ מאד הספר: Casella G., Berger, R.L. *Statistical Inference*, 2<sup>nd</sup> Ed. (2001)

### **חובות הקורס והרכב הציון בקורס:**

תרגילי בית (פתרון שאלות מתמטיות על החומר הנלמד ועבודה עם EXCEL)

משקל: 50%

חובת הגשה: 6 מתוך 8 תרגילים. הממוצע יחושב מתוך 6 התרגילים הטובים ביותר. אם יוגשו 5 תרגילים יחושב הממוצע כאילו הציון בתרגיל החסר היה 0. במידה ויוגשו פחות מ-5 תרגילים, לא ניתן יהיה להגיש את העבודה המסכמת ולעבור את הקורס.

### **עבודת סיום:**

משקל: 50%

עבודת הסיום של הקורס היא מעין "מבחן בית", והיא כוללת אתגרים משני סוגים: שאלות מתמטיות בתחומים שנלמדו בקורס (המחייבות מענה מתמטי כתוב, ללא צורך במחשב), ושאלות הנוגעות לקובצי נתונים המצורפים, ומחייבות הפגנת מיומנות ה-EXCEL שנרכשו במהלך הקורס, כמו גם הפגנת יכולת התמודדות עם נתוני אמת.

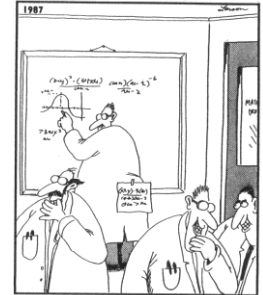
**ציון עובר בקורס: 60**

**[MP1] עם הערות:** גלעד, האם זה מבחן בית או עבודה? אם זה מבחן בית, הוא יהיה מוגבל בזמן של עד 24 או 48 שעות (אולי 72). ואם זו עבודה, יהיה לסטודנטים כחודש עד להגשתה בחדר עם שאר העבודות. אין לכתוב "מעין מבחן בית", צריכים להיות ברורים לגבי המטלות.

מבחינת המערכת, משיקולים טכניים ולא פדגוגיים, במנהל מעדיפים עבודות, ולא מבחני בית - מכיוון שבקורס שלך יהיו סטודנטים משני המחזורים, יהיה מאתגר לשבץ את מבחן הבית בלוח המבחנים של מועדי א' ו-ב'. אך כמו שאמרת, ההחלטה היא פדגוגית שלך.

**פירוט השיעורים, הנושאים הנלמדים בכל שיעור וקריאת החובה והרשות בכל שיעור:**

**שיעור 1: מנין נובע הצורך בסטטיסטיקה וכיצד היא פועלת**



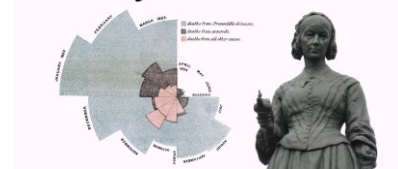
בשיעור מבוא זה נתאר את הצורך בתיאור סטטיסטי של (היבטים מסויימים של) העולם, נספק רקע היסטורי להתפתחותה של הסטטיסטיקה ונתאר את הדרך בה נעשה מחקר סטטיסטי. בשיעור יוצגו דוגמאות של שאלות בעלות אופי סטטיסטי ואת עקרונות ההסקה הסטטיסטית ומגבלותיה ממעוף הציפור.

**קריאה מומלצת:**

זמיר ובייט-מרום: מבוא לסטטיסטיקה לתלמידי מדעי החברה א', או"פ 1993  
פרקים 1.2-1.6.

**שיעור 2: סטטיסטיקה תיאורית: משכחויות למדדי פיזור**

*The Lady with the Pie Chart*



בשיעור זה נדון במדדים, בדגימה, ובדרך לסכם נתונים נדגמים. נדון בתיאור שכיחויות ובדרכים להצגתן הגרפית (גרף שכיחויות, תרשים עוגה, box & whisker plot), ונמשיך בהגדרת מדדי מיקום מרכזי ופיזור פשוטים, כגון חציון, רבעון ואחוזון, ותכונותיהם.

**קריאה מומלצת:**

זמיר ובייט-מרום: מבוא לסטטיסטיקה לתלמידי מדעי החברה א', או"פ 1993  
פרקים 2.2-2.4.

### שיעורים 3-6: הבסיס ההסתברותי לסטטיסטיקה



בשיעורים אלו נדון בבסיס ההסתברותי של הסטטיסטיקה. נלמד מהו משתנה מקרי, ונדגים את תכונותיו הבסיסיות דרך כמה התפלגויות שכיחות: התפלגות ברנולי והתפלגות אחידה (כמו למשל הטלת קובייה או מטבע הוגנים), התפלגות בינומית והתפלגות גיאומטרית. נראה כיצד אפשר להציג גרפית את פונקציית ההתפלגות ואת פונקציית ההסתברות ונבין את הקשר ביניהן. נפתח את מושגי התוחלת והשונות ונחקור את תכונותיהם. לסיים נראה כיצד ניתן לחשוב על ההתפלגות כעל גבול השכיחות של המדגם.

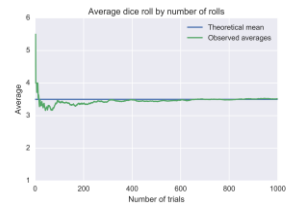
#### קריאת חובה:

זמיר ובייט-מרום: מבוא לסטטיסטיקה לתלמידי מדעי החברה א', או"פ 1993, יחידה 6 (שימו לב כי 6.1-6.2 מהווים חזרה על חומר מסמסטר א') ופרקים 8.1-8.5 (8.1-8.3 מהווים חזרה).

#### קריאה מומלצת:

זמיר ובייט-מרום: מבוא לסטטיסטיקה לתלמידי מדעי החברה א', או"פ 1993 פרק 7.5.

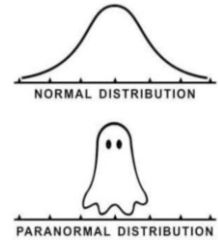
### שיעור 7: התכנסות של מדדי פיזור כתלות בגודל המדגם



בשיעור זה נדון בהתכנסות התוחלת לממוצע: התנאים בהם ישנה התכנסות, בשגיאת התקן ובקצב ההתכנסות. נדון בקשר בין השונות וסטיית התקן התיאורטיות ואלו של המדגם. נתאר את חוק המספרים הגדולים החלש. דבר זה יתן לנו כלים משמעותיים כדי להבין כיצד ראוי לבנות שאלונים, ובעיקר – איזה כח סטטיסטי צפוי להיות להם כתלות במספר הנשאלים.

## ההתפלגות הנורמלית ותכונותיה

שיעור 8:



בשיעור זה נציג את ההתפלגות הנורמלית. זוהי התפלגות חשובה ביותר, שכן היא מופיעה באינספור מצבים – ציוני פסיכומטרי, התפלגות גובה ומשקל, ועוד. נציג את ההתפלגות הנורמלית כמקרה גבול של התפלגות בינומית, ואת משפט הגבול המרכזי. נדון בטבלאות של פונקציית השגיאה ובהבדלים שבין משתנים מקרים בדידים (בהם דנו עד כה) ורציפים.

קריאת חובה:

זמיר ובייט-מרום: מבוא לסטטיסטיקה לתלמידי מדעי החברה א', או"פ 1993  
פרק 4.2.

## הסתברות מותנה ושימושיה

שיעור 9:

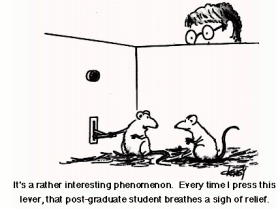


בשיעור זה נדון במקרים בהם אנחנו חוקרים את היחס בין זוגות של משתנים מקריים. נציג את מושג אי-התלות הסטטיסטית ונפתח את מושג ההסתברות המותנה. תוך שימוש במושגים אלו נפתח את נוסחת בייס. יוצגו כמה דוגמאות בהן דרך התיאור הטבעית של תהליכים נעשית פשוטה יותר כאשר משתמשים בהסתברויות מותנות.

קריאת חובה:

זמיר ובייט-מרום: מבוא לסטטיסטיקה לתלמידי מדעי החברה א', או"פ 1993  
פרק 7.1-7.4.

**שיעור 10: מה הקשר? על הדרך בה הסטטיסטיקה יכולה לחקור קשרים**

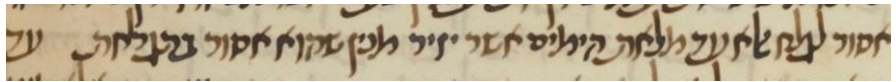


בשיעור זה נפתח את מושג המתאם (קורלציה) של זוגות משתנים מקריים. נדון בכמה הבדלים כשלים אופייניים ומהותיים בהבנת מושג המתאם: (1) היעדר הקשר שבין מתאם לסיבתי; (2) היעדר הקשר בין מתאם מובהק לחוזק הקשר; (2) ההבדלים המהותיים בין אי-תלות סטטיסטית לבין מתאם נמוך.

**קריאת חובה:**

זמיר ובייט-מרום: מבוא לסטטיסטיקה לתלמידי מדעי החברה א', או"פ 1993, פרקים 5.1-5.4.

**שיעור 11: מדעי הרוח הדיגיטלים**

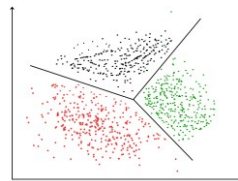


בשיעור זה נשמע על שימוש בכלים סטטיסטיים וכלים כמותיים אחרים כדי לענות על שלל שאלות מרתקות מתחום מדעי הרוח הדיגיטלים. כיצד להפוך כתב יד עתיק לטקסט בו ניתן לחפש תוך שילוב למידת מכונה וחכמת המונים (פרויקט תיקון סופרים של הגניזה הקהירית)? כיצד ניתן להשתמש בשיטות סטטיסטיות, חלקן מחקר הגנטיקה, כדי להשוות גרסאות שונות של טקסט וליצור "עץ משפחה" של השתלשלותו?

לינק לאתר תיקון סופרים כהכנה לשיעור:

<https://tikkoun-sofrim.firebaseio.com/en>

**שיעור 12-13: הסקה סטטיסטית ולמידה**



IT WAS GETTING HARDER AND HARDER TO FIND A TRULY MEANINGFUL RELATIONSHIP AT THE MEDICAL JOURNAL HAPPY HOUR.

בשיעורים אלה נפתח את עקרונות ההסקה הסטטיסטית. נפתח את מושגי האמידה הנקודתית (בה כבר דנו בהקשר של ממוצע), אמידת טווח (רווח בר סמך) ובדיקת השערות. תוך כדי כך נדון בשגיאות מסוג I ו-II ובמושג העצמה הסטטיסטית.

נבין אילו הנחות יסוד נמצאות בבסיס קבלת או דחיית השערה, מהו ערך  $p$  ומהי ההתפלגות העומדת בבסיסו. נלמד כיצד יש לוודא לפני הפעלת מבחן סטטיסטי האם הנתונים עומדים בתנאים הדרושים לקיומו, ובבעיות הכרוכות בפסילת נתונים.

כמו-כן נציג בשיעורים אלה גישות רווחות לטיפול בנתונים בעלי אופי סטטיסטי. נסביר מהי הורדת ממד (dimensionality reduction) ונדגים את חשיבות השיטה לניתוח תוצאות מבחנים שונים ולהצגה נגישה של מידע רב-ממדי. נסביר את עקרונות ניתוח האשכולות (clustering) ונדגים את חשיבותה בניתוח קטגורי של תופעות שונות מתחומי ידע מגוונים. חלק מהשיעור האחרון יוקדש למיני-סדנה בה נעבד נתונים ממאגרי נתונים אמיתיים וננסה להפעיל עליהם שיטות שרכשנו במהלך הקורס.