

# צוּחַק מִי שְׂצוּחַק אַחֲרוֹן

פרוייקט במסגרת הקורס "סטטיסטיקה למדעי המחשב"

סמסטר א' 9-2008

מרצה: ד"ר סהרון רוסט

מתרגל: יונתן רוזנבלט

מגישים:

כלב אלפרנס

קיריל סולוביי

ודים סטוטלנד

אריאל סטולרמן

## תוכן עניינים

3	מבוא
3	פרק 1 : נושא ושאלות המחקר
4	פרק 2 : הנתונים
4	• איסוף הנתונים
5	• תיאור הנתונים
7	פרק 3 : ניתוח הנתונים
7	• שאלה 1 : האם יש קשר בין נתונים אובייקטיביים שנאספו וחוש ההומור?
10	• שאלה 2 : איזו פקולטה צוחקת יותר?
12	• שאלה 3 : האם תפיסת אנשים את חוש ההומור שלהם תואמת בפועל את ההנאה שהפיקו מהבדיחות?
14	פרק 4 : מסקנות
15	נספחים
15	• נספח 1 : שאלון הסקר
16	• נספח 2 : הצגות גרפיות של תמציות הנתונים
17	• נספח 3 : קוד ה R ששימש לניתוח הנתונים

## מבוא

"הומור הוא תכונה או יכולת של בני אדם, עצמים או מצבים להיות או לעורר תחושות של הנאה, צחוק ובדיחות הדעת אצל אנשים אחרים. הומור יכול להתבטא בכל צורה של תקשורת – מילולית, חזותית ושמיעתית – הגורמת לתגובות כאלו אצל בני אדם." -- וויקיפדיה

דבר ידוע הוא שהומור שונה מאדם לאדם. בעוד שאחד יצחק ממושכות מבדיחה מסוימת, אחר עלול אף לא להבין מה היה מצחיק בה. עם זאת, הסיבות להבדלים בין חוש ההומור של אנשים שונים אינן מובנות מאליהן. במחקר זה רצינו לבדוק האם קיימים קשרים בין חוש ההומור לבין אפיונים שונים של סטודנטים באוניברסיטת תל אביב, ביניהם מין, מוצא, פקולטה וכו'. על מנת לבדוק קיום קשרים אלה, חולק שאלון בקרב כ-250 סטודנטים מפקולטות שונות.

### פרק 1: נושא ושאלות המחקר

#### **שאלה 1: האם יש קשר בין נתונים אובייקטיביים שנאספו (גיל, פקולטה וכו') וחוש ההומור?**

כלומר: האם ובאיזו מידה מידת ההנאה של אנשים ממכלול הבדיחות שנבחרו או מהז'אנרים שהוצגו בשאלון מושפעת מאחד או יותר מהמאפיינים לפיהם ניתן לחלק את האוכלוסיה הנדגמת (שנבדקו בשאלון). להלן רשימת הקשרים המרכזיים שנרצה לבדוק:

- 1.1 הקשר בין מוצא ובין הנאה מבדיחות עדתיות: האם אנשים ממוצא אשכנזי נהנים פחות מבדיחות על אשכנזים מאשר אנשים ממוצא ספרדי? ולהיפך?
- 1.2 הקשר בין מגדר ובין הנאה מבדיחות מגדר: האם נשים נהנות יותר מגברים מבדיחות פמניסטיות? האם גברים נהנים יותר מנשים מבדיחות שוביניסטיות?
- 1.3 הקשר בין מצב זוגי ובין הנאה מבדיחות גסות: האם רווקים נהנים יותר מבדיחות גסות מאשר אנשים הנמצאים במערכת יחסים זוגית?
- 1.4 הקשר בין מוצא ובין הנאה מבדיחות שואה: האם אנשים ממוצא אשכנזי נהנים פחות מבדיחות בנושא שואה?

#### **שאלה 2: איזו פקולטה צוחקת יותר?**

מהתבוננות בנתונים נעלה סברה לגבי איזו מהפקולטות נהנת במידה הגבוהה ביותר מכלל הבדיחות והז'אנרים שנבדקו בסקר, ואיזו נהנית הכי פחות. לאור סברה זו נשאל:

- 2.1 האם הפקולטה שנהנית במידה הגבוהה ביותר אכן נהנית מבדיחות יותר משאר הפקולטות?
- 2.2 האם הפקולטה שנהנית במידה הנמוכה ביותר אכן נהנית מבדיחות פחות משאר הפקולטות?

#### **שאלה 3: האם תפיסת אנשים את חוש ההומור שלהם תואמת בפועל את ההנאה שהפיקו מהבדיחות?**

כלומר, האם התשובות לשאלה: "האם סוג זה של בדיחות מצחיק אותך?" תואמות את הציון שניתן לסוגה זו של בדיחות בפועל (עבור הבדיחות שנבחרו לייצגה בשאלון). נבדוק שאלה זו פעמיים:

- 3.1 האם תפיסת כלל הסטודנטים את חוש ההומור שלהם עומדת בקנה אחד עם מידת ההנאה שהם הפיקו מהבדיחות?
- 3.2 האם תפיסת הסטודנטים, בפילוח לפקולטות, עומדת בקנה אחד עם מידת ההנאה שהם הפיקו מהבדיחות?

## פרק 2: הנתונים

### איסוף הנתונים

#### איסוף הנתונים:

- הנתונים נאספו באמצעות סקר (נספח 1) שחולק לסטודנטים מכל התארים והפקולטות באוניברסיטת תל אביב.
- שאלונים חולקו בקרב העוברים ושבים (והיושבים במדשאות) בכל רחבי האוניברסיטה.
- השאלון הכיל סדרה של 12 בדיחות מ-6 סוגות שונות, אותן התבקשו הנסקרים לדרג בסולם מ-0 עד 5, וכן התבקשו לציין עבור כל בדיחה האם סוג זה של בדיחות מצחיק אותם באופן כללי. כמו כן השאלון הכיל סדרה של פרטים אישיים (גיל, מין, פקולטה, איזור מגורים וכו').
- הקטגוריות שנבחרו נועדו לייצג מגוון גדול ככל האפשר (במסגרת מגבלות אורך הפרוייקט) של סוגי בדיחות שונים, ולייצג את הסוגים הפופולריים ביותר: בדיחות קרש (סתמייות), בדיחות שחורות, בדיחות עדתיות, בדיחות מגדר (שוביניסטייות / פמיניסטייות), בדיחות שואה ובדיחות גסות.
- נציין כי כל סטודנט הלומד בשתי פקולטות שונות נחשב בניתוח הנתונים פעמיים – פעם לכל פקולטה בה לומד.

#### טעויות אפשריות באיסוף הנתונים ואופן המנעותנו מהן:

- *Selection Bias*, דגימת מכסה: בכדי להימנע מדגימה מוטה מתוך אוכלוסית הסטודנטים, דאגנו לדגום ברחבי כל האוניברסיטה כאשר לא נקבע מראש זמן קבוע או "מכסת" דגימות לכל אזור.
- הצגת השאלה: השאלות שהוצגו בפני הנסקרים היו שאלות סגורות בלבד, זאת על מנת לשמור על אחידות בתשובות, ולצמצם למינימום את השפעת ניסוח השאלה על התשובה.
- היבטים נוספים באופן איסוף הנתונים:
  - הסקר בוצע ע"י 4 סוקרים שונים: מתוך הנחה כי למרות הניסיון לדגום אנשים בצורה ראנדומלית, לכל אדם תהיה נטיה נסתרת לבחירת נסקר כזה או אחר.
  - הסקר חולק על פני מספר ימים שונים ושעות שונות במהלך היום.
  - בחירת הבדיחות שייצגו נאמנה את הקטגוריות השונות נעשתה באופן האקראי ביותר הניתן: נעשה ניסיון לסנן בדיחות "תפלות" ובדיחות קיצוניות (פוגעניות יתר על המידה). כמו כן הבדיחות שוקללו מתוך בחירותיהם של 4 אנשים שונים.
  - סדר הופעת הבדיחות מהסוגים השונים בשאלון נקבע כך שלא יהיו מקבצים של בדיחות מאותו סוג, על מנת ליצור אקראיות בסדר הופעת הבדיחות. ניתן להניח כי אם הבדיחות היו מופיעות לפי סוגן, סדר זה היה משפיע על תגובת הנסקר ודירוגו את הבדיחה השניה בכל סוג.
- כמו כן, ניסינו למנוע אנטגוניזם בקרב חלקים באוכלוסיה שהיו סולדים מז'אנרים מסוימים בכך שהתחלנו עם בדיחת "פרווה", ואת הבדיחות הקשות יותר הכנסנו בהמשך.

#### טעויות שנעשו במהלך איסוף הנתונים:

- רוב הנשאלים לא קראו כהלכה את הוראות הסקר, וישר פנו למילוי דירוג הבדיחות. אופן בניית השאלון גרמה ככל הנראה לבלבול בקרב חלק מהנסקרים, אשר לא מילאו כתוצאה מכך את עמודת ה"כן/לא". יצויין כי כאשר שמנו לב לתופעה, דאגנו להסביר בקצרה לנסקרים את אופן מילוי השאלון.

**תיאור הנתונים**

הנתונים האישיים שנאספו מהנסקרים הם: גיל, מין, פקולטה, חוג לימודים, תואר נלמד, השקפה דתית, מוצא, מצב משפחתי (ומספר ילדים), איזור מגורים, מצב מגורים (עם מי הנסקר מתגורר) והשקפה פוליטית.

להלן תיאור תמציות הנתונים הסופיים בהם השתמשנו כבסיס לשאלות המחקר; הצגות גרפיות של הנתונים בנספח 2.

גיל:

תמצית	ממוצע	ממוצע קצוץ $\alpha = 0.05$	חציון	סטיית תקן	MAD	YULE
ערך	23.96	23.86	24	3.006346	2.9652	-0.0212766
משמעויות	<ul style="list-style-type: none"> <li>לא היו הרבה דגימות קיצוניות מבחינת גיל.</li> <li>זנב קטן לכיוון גילאים צעירים יותר.</li> </ul>					

מין:

השקפה דתית:

השקפה	דתי	מסורתי	חילוני	אתאיסט
מספר הנדגמים	7	21	172	53
אחוז מהמדגם	2.76%	8.3%	67.98%	20.94%

	אחוז מהמדגם	מספר בפועל
גברים	51%	125
נשים	49%	130

פקולטה:

פקולטה	אומנויות	הנדסה	מדוייקים	משפטים	חיים	ניהול	רפואה	חברה	רוח
מספר הנדגמים	29	26	66	18	27	6	32	23	27
אחוז מהמדגם	11.41%	10.23%	25.98%	7.08%	10.62%	2.36%	12.59%	9.05%	10.62%

מוצא:

מצב משפחתי:

מוצא	אשכנזי	ספרדי	מעורב	ערבי
מספר הנדגמים	185	30	34	6
אחוז מהמדגם	72.54%	11.76%	13.33%	2.35%

מצב משפחתי	רווק	יש בן / בת זוג	נשוי
מספר הנדגמים	174	65	16
אחוז מהמדגם	68.23%	25.49%	6.27%

אזור מגורים:

מצב מגורים:

אזור	מרכז	י-ם	צפון	שרון	דרום
מספר הנדגמים	192	6	10	39	8
אחוז מהמדגם	75.29%	2.35%	3.92%	15.29%	3.13%

מצב מגורים	לבד	בן / בת זוג	הורים	שותפים
מספר הנדגמים	22	52	133	48
אחוז מהמדגם	8.62%	20.39%	52.15%	18.82%

השקפה פוליטית:

השקפה	ימין	מרכז-ימין	מרכז	מרכז-שמאל	שמאל
מספר הנדגמים	29	23	49	72	68
אחוז מהמדגם	12.03%	9.54%	20.33%	29.87%	28.21%

**תוצאות דירוג הבדיחות והסגנונות :**

מגדר 2 (פמני)	מגדר 1 (שובני)	גסה 2 (חומוס)	גסה 1 (דבורים)	עדות 2 (מרוקאית)	עדות 1 (פולניה)	שואה 2 (קצין)	שואה 1 (מדרגה)	שחורה 2 (תנוק)	שחורה 1 (סרטן)	קרש 2 (קבלן)	קרש 1 (חולצה)	בדיחה / תמצית
2.614	2.2	2.492	2.780	2.226	2.393	2.102	2.2	1.777	2.674	2.043	2.219	ממוצע
3	2	3	3	2	3	2	2	1	3	2	2	חציון
1.509	1.63	1.746	1.645	1.633	1.623	1.663	1.601	1.725	1.702	1.45	1.444	סטיית תקן
1.482	2.965	1.482	1.482	1.482	1.482	1.482	1.482	1.482	1.482	1.482	1.482	MAD
-0.142	0.2	-0.2	-0.2	0.2	-0.2	0	0	0.333	-0.2	0	0	YULE

מגדר	גסות	עדות	שואה	שחורות	קרש	סוג הבדיחות / האם סוג זה מצחיק
247	282	250	242	229	225	מספר ה"כן"
217	181	207	219	230	235	מספר ה"לא"
53.23%	60.9%	54.7%	52.49%	49.89%	48.91%	אחוז ה"כן"
46.77%	39.1%	45.3%	47.51%	50.11%	51.09%	אחוז ה"לא"

**משמעויות :**

- ניתן לראות כי ממוצע הציון של הרוב המוחלט של הבדיחות הינו קרוב לממוצע הדירוג האפשרי (שהוא 2.5). ממוצע ממוצעי הציונים לכלל הבדיחות הוא 2.3 (בקירוב).
- פרט לבדיחות הגסות, מרבית הז'אנרים מתחלקים בערך באופן שווה מבחינת ההנאה מהם.

### פרק 3: ניתוח הנתונים

**שאלה 1: האם יש קשר בין נתונים אובייקטיביים שנאספו (גיל, פקולטה וכו') וחוש ההומור?**

1.1 הקשר בין מוצא ובין הנאה מבדיחות עדתיות:

השערת האפס הכללית: אין קשר בין מוצא (אשכנזי / ספרדי / מעורב) ובין מידת ההנאה מבדיחות עדתיות. ההשערה האלטרנטיבית הכללית: קיים קשר כלשהו בין מוצא ובין מידת ההנאה מבדיחות עדתיות (השערה דו כיוונית). ניקח את סטטיסטי המבחן להיות הפרש תוחלות הציונים של בדיחה על עדה מסוימת בין אשכנזים וספרדים. נבדוק שני קשרים שונים:

- הקשר בין ממוצע ציון שניתן ע"י אשכנזים לבדיחה על אשכנזים, ובין זה שניתן ע"י ספרדים לאותה בדיחה.
  - הקשר בין ממוצע ציון שניתן ע"י אשכנזים לבדיחה על ספרדים, ובין זה שניתן ע"י ספרדים לאותה בדיחה.
- סימונים:

1.  $\bar{X}_1$  יהיה הציון הממוצע של בדיחת האשכנזים בקרב האשכנזים.
2.  $\bar{X}_2$  יהיה הציון הממוצע של בדיחת האשכנזים בקרב הספרדים.
3.  $\bar{X}_3$  יהיה הציון הממוצע של בדיחת הספרדים בקרב האשכנזים.
4.  $\bar{X}_4$  יהיה הציון הממוצע של בדיחת הספרדים בקרב הספרדים.

בכל אחד מהמקרים נבדוק את הקשר בין שני משתנים מקריים בלתי תלויים. מהתבוננות בנתונים בחרנו להניח שוויון שונויות. נגדיר משתנים מקריים (נסמן את מספר האשכנזים שדרג את הבדיחה ב- $n$ , ואת הספרדים ב- $m$ ):

$$\bar{X}_i \sim N\left(\mu_{X_i}, \frac{\sigma^2}{n}\right), \bar{X}_j \sim N\left(\mu_{X_j}, \frac{\sigma^2}{m}\right)$$

$$\bar{X}_i - \bar{X}_j \sim N\left(\mu_{X_i} - \mu_{X_j}, \frac{n+m}{n \cdot m} \cdot \sigma^2\right)$$

נשתמש באומד חסר הטייה לשוונות בקרב האשכנזים והספרדים ביחס לציון שניתן לבדיחה הרלוונטית:

$$\widehat{\sigma^2} = \frac{(n-1) \cdot \widehat{\sigma_{\bar{X}_i}^2} + (m-1) \cdot \widehat{\sigma_{\bar{X}_j}^2}}{n+m-2}$$

עבור בדיחת האשכנזים:

$$H_0: \mu_{\bar{X}_1 - \bar{X}_2} = 0, H_1: \mu_{\bar{X}_1 - \bar{X}_2} \neq 0$$

נציב את הנתונים ונקבל:

$$\widehat{\sigma^2} = 2.576, \widehat{\mu_{\bar{X}_1 - \bar{X}_2}} = -0.653$$

$$\Rightarrow C_\alpha = \left( -\infty, \mu_0 - t_{n+m-2, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{n+m}{n \cdot m} \cdot \sigma^2} \right) \cup \left[ \mu_0 + t_{n+m-2, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{n+m}{n \cdot m} \cdot \sigma^2}, +\infty \right) =$$

$$\left( -\infty, -t_{185+30-2, 0.975} \cdot \sqrt{\frac{185+30}{185 \times 30} \cdot 2.576} \right) \cup [\dots] = (-\infty, -0.622] \cup [0.622, +\infty)$$

מכאן **שנדחה את השערת האפס**, כלומר ישנו קשר בין מוצא אשכנזי או ספרדי ובין מידת ההנאה מבדיחות על אשכנזים. בפרט, הקשר הוא: ספרדים נהנים יותר מבדיחות על אשכנזים מאשר אשכנזים ברמת ביטחון של 95%.

עבור בדיחת הספרדים:

$$H_0: \mu_{\bar{X}_3 - \bar{X}_4} = 0, H_1: \mu_{\bar{X}_3 - \bar{X}_4} \neq 0$$

נציב את הנתונים ונקבל:

$$\widehat{\sigma^2} = 2.753, \widehat{\mu_{\bar{X}_3 - \bar{X}_4}} = -0.425$$

$$\Rightarrow C_\alpha = (-\infty, -0.644] \cup [0.644, +\infty)$$

מכאן **שלא נדחה את השערת האפס**, כלומר **אין קשר** בין מוצא למידת ההנאה מבדיחות על ספרדים, ברמת ביטחון של 95%.

1.2 הקשר בין מגדר ובין הנאה מבדיחות מגדר:

השערת האפס הכללית: אין קשר בין מגדר ובין מידת ההנאה מבדיחות מגדר.

ההשערה האלטרנטיבית הכללית: קיים קשר כלשהו בין מגדר ובין מידת ההנאה מבדיחות מגדר (השערה דו כיוונית).

ניקח את סטטיסטי המבחן להיות הפרש תוחלות הציונים של בדיחה על מגדר מסויים בין נשים וגברים. נבדוק שני קשרים:

- הקשר בין ממוצע ציון שניתן ע"י גברים לבדיחה על גברים (פמניסטית), ובין זה שניתן ע"י נשים לאותה בדיחה.
- הקשר בין ממוצע ציון שניתן ע"י נשים לבדיחה על נשים (שוביניסטית), ובין זה שניתן ע"י גברים לאותה בדיחה.

סימונים:

1.  $\bar{X}_1$  יהיה הציון הממוצע של בדיחת הגברים בקרב הגברים.

2.  $\bar{X}_2$  יהיה הציון הממוצע של בדיחת הגברים בקרב הנשים.

3.  $\bar{X}_3$  יהיה הציון הממוצע של בדיחת הנשים בקרב הגברים.

4.  $\bar{X}_4$  יהיה הציון הממוצע של בדיחת הנשים בקרב הנשים.

בכל אחד מהמקרים נבדוק את הקשר בין שני משתנים מקריים בלתי תלויים. גם כאן נניח שוויון שונויות. נשתמש בסימונים של השאלה הקודמת.

עבור הבדיחה הפמניסטית:

$$H_0: \mu_{\bar{X}_1 - \bar{X}_2} = 0, H_1: \mu_{\bar{X}_1 - \bar{X}_2} \neq 0$$

נציב את הנתונים ונקבל:

$$\widehat{\sigma^2} = 2.271, \widehat{\mu_{\bar{X}_1 - \bar{X}_2}} = -0.239$$

$$\Rightarrow C_\alpha = (-\infty, -0.372] \cup [0.372, +\infty)$$

מכאן **שלא נדחה את השערת האפס**, כלומר נשים וגברים **נהנים באותה מידה** מבדיחות על גברים ברמת ביטחון של 95%.

עבור הבדיחה הפמניסטית:

$$H_0: \mu_{\bar{X}_3 - \bar{X}_4} = 0, H_1: \mu_{\bar{X}_3 - \bar{X}_4} \neq 0$$

נציב את הנתונים ונקבל:

$$\widehat{\sigma^2} = 2.621, \widehat{\mu_{\bar{X}_3 - \bar{X}_4}} = 0.439$$

$$\Rightarrow C_\alpha = (-\infty, -0.4] \cup [0.4, +\infty)$$

מכאן **שנדחה את השערת האפס**, כלומר נשים וגברים **לא נהנים באותה מידה** מבדיחות על נשים ברמת ביטחון של 95%.

בפרט, גברים נהנים יותר מבדיחות על נשים מאשר נשים.



1.3 הקשר בין מצב זוגי ובין הנאה מבדיחות גסות :

השערת האפס הכללית: אין קשר בין מצב זוגי ובין מידת ההנאה מבדיחות גסות. ההשערה האלטרנטיבית הכללית: קיים קשר כלשהו בין מצב זוגי ובין מידת ההנאה מבדיחות גסות (השערה דו כיוונית). לכל תצפית ניקח כמדד לציון הבדיחות הגסות את הציון הממוצע בין 2 הבדיחות עבור אותו מדרג. ניקח את סטטיסטי המבחן להיות הפרש תוחלות הציונים הנ"ל בין רווקים ובין אנשים בקשר זוגי (נשואים או בעלי בן / בת זוג).

סימונים:

1.  $\bar{x}_1$  יהיה הציון הממוצע כנ"ל בקרב רווקים.

2.  $\bar{x}_2$  יהיה הציון הממוצע כנ"ל בקרב אנשים בקשר זוגי.

נבדוק את הקשר בין שני משתנים מקריים בלתי תלויים. גם כאן נניח שוויון שונויות. נשתמש בסימונים של השאלה הקודמת.

$$H_0: \mu_{\bar{x}_1 - \bar{x}_2} = 0, H_1: \mu_{\bar{x}_1 - \bar{x}_2} \neq 0$$

נציב את הנתונים ונקבל:

$$\widehat{\sigma^2} = 2.111, \widehat{\mu_{\bar{x}_1 - \bar{x}_2}} = 0.199$$

$$\Rightarrow C_\alpha = (-\infty, -0.385] \cup [0.385, +\infty)$$

מכאן **שלא** נדחה את השערת האפס, כלומר רווקים ואנשים הנמצאים בקשר זוגי **נהנים באותה מידה** מבדיחות גסות ברמת ביטחון של 95%.

1.4 הקשר בין מוצא ובין הנאה מבדיחות שואה :

השערת האפס הכללית: אין קשר בין מוצא ובין מידת ההנאה מבדיחות שואה. ההשערה האלטרנטיבית הכללית: קיים קשר כלשהו בין מוצא ובין מידת ההנאה מבדיחות שואה (השערה דו כיוונית). לכל תצפית ניקח כמדד לציון הבדיחות השואה את הציון הממוצע בין 2 הבדיחות עבור אותו מדרג. ניקח את סטטיסטי המבחן להיות הפרש תוחלות הציונים הנ"ל בין אשכנזים ובין ספרדים. סימונים:

1.  $\bar{x}_1$  יהיה הציון הממוצע כנ"ל בקרב אשכנזים.

2.  $\bar{x}_2$  יהיה הציון הממוצע כנ"ל בקרב ספרדים.

נבדוק את הקשר בין שני משתנים מקריים בלתי תלויים. גם כאן נניח שוויון שונויות. נשתמש בסימונים של השאלה הקודמת.

$$H_0: \mu_{\bar{x}_1 - \bar{x}_2} = 0, H_1: \mu_{\bar{x}_1 - \bar{x}_2} \neq 0$$

נציב את הנתונים ונקבל:

$$\widehat{\sigma^2} = 1.947, \widehat{\mu_{\bar{x}_1 - \bar{x}_2}} = -0.044$$

$$\Rightarrow C_\alpha = (-\infty, -0.541] \cup [0.541, +\infty)$$

מכאן **שלא נדחה את השערת האפס**, כלומר אשכנזים וספרדים **נהנים באותה מידה** מבדיחות שואה ברמת ביטחון של 95%. כמו כן, מערכו הנמוך של סטטיסטי המבחן, ניתן להסיק שמידות ההנאה של שתי העדות קרובות מאוד.

## שאלה 2: איזו פקולטה צוחקת יותר?

איזו מהפקולטות נהנת במידה הגבוהה ביותר מכלל הבדיחות והז'אנרים שנבדקו בסקר:

נתבונן עבור כל פקולטה בממוצע ציוני הבדיחות שלה, וכן באומד לשונות של ציוני הבדיחות שלה. עבור כל פקולטה בידינו מספר רב של בדיחות, ולכן ניתן לומר כי  $\bar{X}_i \sim N\left(\mu_{X_i}, \frac{\sigma^2}{n}\right)$ . כלומר, הציון הממוצע של כלל הבדיחות מתפלג בקירוב בהתפלגות נורמאלית. הטבלה שלפנינו מציגה את האומד לתוחלת, הלוא הוא בעצם הממוצע של ציוני כלל הבדיחות בפקולטה מסוימת, וכן ואומד לשונות, כלומר את  $\widehat{\sigma^2}$ :

פקולטה	אומנות	הנדסה	מדוייקים	משפטים	חיים	ניהול	רפואה	חברה	רוח
אומד לתוחלת	<b>2.5907</b>	2.3806	2.3012	2.1342	2.1327	2.4084	<b>2.1243</b>	2.1956	2.5541
אומד לשונות	<b>2.8667</b>	2.7996	2.6340	<b>2.3493</b>	2.4869	2.6450	2.7510	2.6888	2.6329

2.1 האם הפקולטה לאומנויות אכן נהנית מבדיחות יותר משאר הפקולטות:

נבדוק את הקשר בין שני משתנים מקריים בלתי תלויים, כאשר  $\bar{X}_1$  הוא ממוצע ציוני הבדיחות בקרב הסטודנטים בפקולטה לאומנות,  $\bar{X}_2$  הוא ממוצע ציוני הבדיחות בקרב כל שאר הסטודנטים. גם כאן בחרנו להניח שוויון שונות. נגדיר משתנים מקריים (נסמן את מספר הסטודנטים בפקולטה לאומנות ב- $n$ , ואת מספר הסטודנטים הנותרים ב- $m$ ):

$$\bar{X}_1 \sim N\left(\mu_{X_1}, \frac{\sigma^2}{n}\right), \bar{X}_2 \sim N\left(\mu_{X_2}, \frac{\sigma^2}{m}\right)$$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_{X_1} - \mu_{X_2}, \frac{n+m}{n \cdot m} \cdot \sigma^2\right)$$

נגדיר את השערת האפס והאלטרנטיבה:

$$H_0: \mu_{\bar{X}_1 - \bar{X}_2} = 0, H_1: \mu_{\bar{X}_1 - \bar{X}_2} > 0$$

נציב את הנתונים ונקבל:

$$\widehat{\sigma^2} = 2.674, \widehat{\mu_{\bar{X}_1 - \bar{X}_2}} = 0.315$$

$$\Rightarrow C_\alpha = [0.153, +\infty)$$

מכאן **שנדחה את השערת האפס**, כלומר סטודנטים לאומנות **נהנים יותר** מבדיחות מאשר שאר הסטודנטים באוניברסיטה, ברמת ביטחון של 95%.

2.2 האם הפקולטה לרפואה אכן נהנית מבדיחות פחות משאר הפקולטות:

נבדוק את הקשר בין שני משתנים מקריים בלתי תלויים, כאשר  $\bar{X}_1$  הוא ממוצע ציוני הבדיחות בקרב הסטודנטים בפקולטה לרפואה,  $\bar{X}_2$  הוא ממוצע ציוני הבדיחות בקרב כל שאר הסטודנטים. גם כאן בחרנו להניח שוויון שוניות. נגדיר משתנים מקריים (נסמן את מספר הסטודנטים בפקולטה לרפואה ב- $n$ , ואת מספר הסטודנטים הנוותרים ב- $m$ ):

$$\bar{X}_1 \sim N\left(\mu_{X_1}, \frac{\sigma^2}{n}\right), \bar{X}_2 \sim N\left(\mu_{X_2}, \frac{\sigma^2}{m}\right)$$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_{X_1} - \mu_{X_2}, \frac{n+m}{n \cdot m} \cdot \sigma^2\right)$$

נגדיר את השערת האפס והאלטרנטיבה:

$$H_0: \mu_{\bar{X}_1 - \bar{X}_2} = 0, H_1: \mu_{\bar{X}_1 - \bar{X}_2} < 0$$

נציב את הנתונים ונקבל:

$$\widehat{\sigma^2} = 2.679, \widehat{\mu_{\bar{X}_1 - \bar{X}_2}} = -0.213$$

$$\Rightarrow C_\alpha = (-\infty, -0.148]$$

מכאן **שנדחה את השערת האפס**, כלומר סטודנטים לרפואה **נהנים מבדיחות פחות** משאר הסטודנטים באוניברסיטה, ברמת ביטחון של 95%.

### שאלה 3: האם תפיסת אנשים את חוש ההומור שלהם תואמת בפועל את ההנאה שהפיקו מהבדיחות?

השערת האפס והאלטרנטיבה הבאות תקפות הן ל 3.1 ו ל 3.2.

השערת האפס הכללית: קיימת התאמה בין תפיסת הסטודנטים את חוש ההומור שלהם לבין מידת ההנאה שהם הפיקו בפועל. ההשערה האלטרנטיבית הכללית: לא קיימת התאמה כזו (השערה דו כיוונית).

סטטיסטי המבחן: ההפרש בין פרופורציית ה"כן", לבין ממוצע ציוני הבדיחות המתוקנן (בין 0 ל-1), היינו  $X_N := \bar{X}/5$ . מספר ה"כן" הוא משתנה מקרי המתפלג בינומית, ונתוני המדגם מאפשרים לנו להשתמש בקירוב הנורמלי לפרופורציה זו.

#### 3.1 האם תפיסת כלל הסטודנטים את חוש ההומור שלהם עומדת בקנה אחד עם מידת ההנאה שהם הפיקו מהבדיחות:

נרצה לחשב תחילה כיצד מתפלגת פרופורציית ה"כן" לפי הקירוב הנורמלי. נסמן את האומדן לפרופורציית ה"כן" מכלל דירוגי "כן" או "לא" ב- $\hat{p}$ . נסמן ב- $Y$  את מספר ה"כן" הכולל, וכיוון שאנו רוצים תיקון לערך בין 0 ל-1, נקבל כי:  $Y_N \sim N(\hat{p}, 9 \times 10^{-5})$ , כאשר  $Y_N$  הוא סימון ל- $Y$  המתוקנן (והוא בערך  $\hat{p}$ ).

$$H_0: X_N - Y_N = 0, H_1: X_N - Y_N \neq 0$$

נניח שוויון שונויות ונחשב אומדן לשונות ההפרש; נסמן ב- $n$  את מספר ציוני הבדיחות הכולל וב- $m$  את מספר דירוגי ה"כן"/"לא" הכולל:

$$\hat{\sigma}^2 = \frac{(n-1) \cdot \hat{\sigma}_{X_N}^2 + (m-1) \cdot \hat{\sigma}_{Y_N}^2}{n+m-2} = \frac{(3047-1) \cdot 2.683 + (2764-1) \cdot 9 \cdot 10^{-5}}{3047+2764-2} = 1.407$$

כמו כן מהנתונים עולה כי  $X_N - Y_N = -0.0714$ . נחשב איזור דחיה (דו כיווני) ברמה  $\alpha = 0.05$ :

$$C_\alpha = \left( -\infty, \mu_0 - t_{n+m-2, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{n+m}{n \cdot m} \cdot \sigma^2} \right) \cup \left( \mu_0 + t_{n+m-2, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{n+m}{n \cdot m} \cdot \sigma^2}, +\infty \right) =$$

$$\left( -\infty, -t_{3047+2764-2, 0.975} \cdot \sqrt{\frac{3047+2764}{3047 \times 2764} \cdot 1.407} \right) \cup [\dots] = (-\infty, -0.0610] \cup [0.0610, +\infty)$$

מכאן **שנדחה את השערת האפס**, כלומר שניתן להגיד ברמת ביטחון של 95% שתפיסת כלל הסטודנטים את חוש ההומור שלהם אינה עומדת בקנה אחד עם מידת ההנאה שהם הפיקו מהבדיחות. בפרט, מידת ההנאה שהפיקו כלל הסטודנטים מהבדיחות נמוכה מאשר ההנאה המצופה של אותם סטודנטים, בהתאם להגדרתם את חוש ההומור שלהם.

#### 3.2 האם תפיסת הסטודנטים, בפילוח לפקולטות, עומדת בקנה אחד עם מידת ההנאה שהם הפיקו מהבדיחות:

נבדוק עבור הפקולטות הבאות: מדעים מדויקים, מדעי החברה ואומנויות. גם כאן:

$$H_0: X_N - Y_N = 0, H_1: X_N - Y_N \neq 0$$

כאשר  $X_N$  הוא תקנון הציון הממוצע של הבדיחות בקרב הפקולטה הנבדקת,  $Y_N$  הוא תקנון למספר ה"כן" באותה פקולטה. גם בשאלות אלו נניח שוויון שונויות.

מדעים מדויקים :

נציב את הנתונים ונקבל :

$$\widehat{\sigma^2} = 1.3719, \widehat{X_N - Y_N} = -0.0843$$

$$C_\alpha = (-\infty, -0.1179] \cup [0.1179, +\infty)$$

מכאן **שלא נדחה את השערת האפס**, כלומר : תפיסת הסטודנטים בפקולטה למדעים מדויקים את חוש ההומור שלהם אכן **עומדת בקנה אחד** עם מידת ההנאה שהפיקו מהבדיחות, ברמת ביטחון של 95%.

מדעי החברה :

נציב את הנתונים ונקבל :

$$\widehat{\sigma^2} = 1.3696, \widehat{X_N - Y_N} = -0.1093$$

$$C_\alpha = (-\infty, -0.1971] \cup [0.1971, +\infty)$$

מכאן **שלא נדחה את השערת האפס**, כלומר : תפיסת הסטודנטים בפקולטה למדעי החברה את חוש ההומור שלהם אכן **עומדת בקנה אחד** עם מידת ההנאה שהפיקו מהבדיחות, ברמת ביטחון של 95%.

אומנויות :

נציב את הנתונים ונקבל :

$$\widehat{\sigma^2} = 1.5030, \widehat{X_N - Y_N} = 0.0228$$

$$C_\alpha = (-\infty, -0.1870] \cup [0.1870, +\infty)$$

מכאן **שלא נדחה את השערת האפס**, כלומר : תפיסת הסטודנטים בפקולטה לאומנויות את חוש ההומור שלהם אכן **עומדת בקנה אחד** עם מידת ההנאה שהפיקו מהבדיחות, ברמת ביטחון של 95%.

פקולטות אחרות :

מבחן זה נעשה גם עבור כל שאר הפקולטות, ובכולן הגענו למסקנה **שלא נדחה את השערת האפס** ברמת ביטחון של 95%. כלומר, אם נתנה בפקולטה, נקבל דמיון בין תפיסת הסטודנטים את חוש ההומור שלהם ובין מידת הנאתם מהבדיחות (לעומת 3.1, שמסקנתה היתה שישנו שוני מובהק בין השתיים).

הסבר אפשרי לתופעה הוא גודל המדגם של כל אחת מהפקולטות בהשוואה לגודל המדגם של כל הפקולטות יחדיו. מאחר וגודל המדגם הכללי גדול משמעותית מגודל המדגם בכל אחת מהפקולטות, אזור הדחייה במקרה הראשון גדל לעומת אזור הדחייה במקרה השני. בנוסף, התוחלת המשוערת במקרה של השערת האפס זהה ( $\mu_0 = 0$ ), עובדה זו מבטיחה כי אזור הדחייה המצומצם יותר במקרה של כל פקולטה יוכל באזור הדחייה של כל הפקולטות יחד.

מסיבה זו תוצאה שגורמת לדחייה של השערת האפס במקרה המצומצם תגרום גם לדחייה במקרה הכללי, אך מצד שני תוצאה שנדחת במקרה הכללי לא בהכרח תדחה במקרה המצומצם, וזה בדיוק תואם את המקרה שלנו שבו החלטנו לא לדחות את השערת האפס במקרה של התוצאות  $-0.0843, -0.1093$  בהתניה על הפקולטות, בעוד שבמקרה הכללי היינו דוחים את השערת האפס.

אומנם זה הגיוני ולגיטימי לחלוטין לקבל את התוצאות הנ"ל, אך במקרה ונרצה לבדוק את השאלה השלישית באופן יותר מדויק, נציע לדגום באופן הבא : לדגום קבוצת סטודנטים מכל הפקולטות בגודל כלשהו ולאחר מכן לדגום באופן בלתי תלוי קבוצות בגודל זהה עבור כל פקולטה בנפרד. דגימה זו תבטיח לנו בדיקה מדויקת יותר של ההשערות, בגלל שוויון גודל המדגמים.

## פרק 4: מסקנות

להלן מסקנותינו מניתוח הנתונים ומהתוצאות שהתקבלו, וכן ניסיון להסביר את התוצאות הנ"ל:

### שאלה 1:

- 1.1 בעוד שאשכנזים וספרדים נהנים באותה המידה מבדיחות על ספרדים, אשכנזים נהנים פחות מספרדים מבדיחות על אשכנזים.  
לדעתנו הדבר נובע מכך שלפחות במידת מה, סטיגמות על "פולנים" תואמות את המציאות, והם באמת לוקחים ללב. כמו כן לדעתנו זה מעיד על מידה גבוהה יותר של הומור עצמי בקרב ספרדים.
- 1.2 בעוד שנשים וגברים נהנים באותה מידה מבדיחות על גברים, נשים נהנות פחות מאשר גברים מבדיחות על נשים. לדעתנו הדבר נובע מכך שאפליית נשים הינה תופעה רווחת יותר מאפליית גברים ולכן בדיחות שוביניסטיות (על נשים) מעוררות אנטגוניזם בקרב חלק מהנשים.
- 1.3 רווקים ואנשים הנמצאים בקשר זוגי נהנים באותה מידה מבדיחות גסות.
- 1.4 אשכנזים וספרדים נהנים באותה מידה מבדיחות שואה.  
לדעתנו הדבר נובע מכך שהאוכלוסיה שנבדקה התחנכה באותה מערכת החינוך הישראלית. ולכן אם קיימים הבדלים ביחס לשואה בין אשכנזים וספרדים הם ינבעו מחינוך מבית. עם זאת, מאחר ורוב הנסקרים הם דור שלישי או רביעי לשואה, סביר להניח שנושא השואה לא היה מרכיב דומיננטי יותר בחינוך בקרב אשכנזים מאשר ספרדים.

### שאלה 2:

- 2.1 סטודנטים לאומנות צוחקים יותר מאשר חבריהם מפקולטות אחרות.  
לדעתנו הדבר נובע מכך שסטודנטים לאומנות הם מטבעם אנשים יצירתיים, משוחררים ובעלי "ראש פתוח" יותר מאשר סטודנטים בפקולטות אחרות. על כן סביר להניח שיהנו יותר ממגוון רחב יותר של בדיחות.
- 2.2 סטודנטים לרפואה צוחקים פחות מאשר חבריהם מפקולטות אחרות.  
לדעתנו הדבר נובע מכך שסטודנטים לרפואה ממויינים מראש לפי תכונות אופי הנדרשות מרופאים, ביניהן אחריות ורצינות (וחיבה לגופות). כמו כן לימודי הרפואה מפרכים, מה שמשפיע על מצב הרוח של הסטודנטים ובאופן ישיר על מידת הנאתם מבדיחות.

### שאלה 3:

- 3.1 נמצאה אי התאמה בין הערכת אנשים את חוש ההומור שלהם, לבין התוצאות דה-פאקטו. בפרט, אנשים העריכו את חוש ההומור שלהם להיות גבוה יותר ממה שהוא היה בפועל.
  - 3.2 לא נמצאה אי התאמה כנ"ל.
- הסבר לחוסר התאימות בין המסקנות ל-3.1 ו-3.2 מופיע בסוף שאלה 3 בפרק 3.

## צוחק מי שצוחק אחרון

השאלון מנוסח בלשון זכר אך מתייחס לשני המינים.

נשמח אם תוכל להקדיש מספר דקות מזמנך כדי להשתתף בסקר קצר ומשעשע...

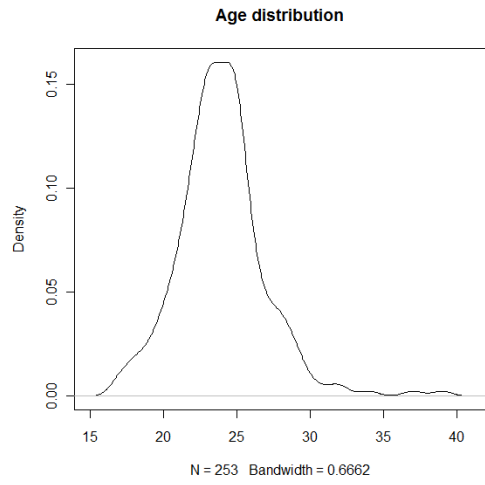
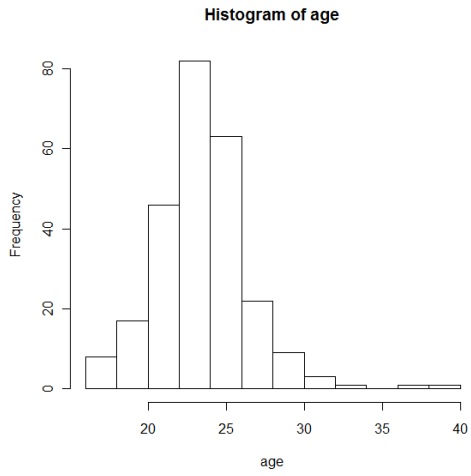
להלן רצף בדיחות, אנא קראי ודרג'י למיטב הנאתך (0 – לא מצחיק בכלל, 1 – מצחיק בקושי, ..., 5 – מצחיק מאוד)

האם סוג זה של בדיחות מצחיק אותך?		דירוג					בדיחה	
לא	כן	5	4	3	2	1	0	
לא	כן	5	4	3	2	1	0	1 ש: מה קרה לטי-שירט שנתקעה מתחת לסלע? ת: היא חולצה!
לא	כן	5	4	3	2	1	0	2 משאית דורסת אישה במעבר חציה, מי אשם? האישה - מה היא עושה מחוץ למטבח? אופנוע דורס אישה, מי אשם? האופנוע - מה הוא עושה במטבח?
לא	כן	5	4	3	2	1	0	3 ש: מה כתוב בכניסה לתאי הגזים? ת: "זהירות מדרגה"
לא	כן	5	4	3	2	1	0	4 ש: איך מזהים פולנייה באיצטדיון מלא אנשים? ת: משחררים נשר, הוא כבר ימצא את הנבלה...
לא	כן	5	4	3	2	1	0	5 בעל מבקש מאישתו: "אני רוצה שנקיים יחסי מין כמו הדבורים." שואלת האשה: "אתה רוצה למרוח דבש ולקק?" עונה לה הבעל: "לא, אני רוצה שתמצצי ותעופי."
לא	כן	5	4	3	2	1	0	6 הבן: "אמא, איזה דג אמרת שיש לי?" האמא: "לא דג חמוד, סרטן"
לא	כן	5	4	3	2	1	0	7 גבר לחברו: "אצלינו בבית, תמיד אני אומר את המילה האחרונה." חברו: "ומה המילה?" הגבר: "סליחה."
לא	כן	5	4	3	2	1	0	8 ש: מהם 70 הדברים שאישה מרוקאית יודעת לעשות? ת: 69 וקוסקוס
לא	כן	5	4	3	2	1	0	9 אני לעולם לא אסלח לנאצים על מה שהם עשו לסבא שלי בשואה. 6 שנים ב-SS ולא קידמו אותו לקצין!
לא	כן	5	4	3	2	1	0	10 ש: מה ההבדל בין משאית מלאה כדורי באולינג למשאית מלאה בתינוקות? ת: משאית כדורי באולינג אי אפשר להעמיס עם קילשון
לא	כן	5	4	3	2	1	0	11 ש: מה ההבדל בין קבלן לרופא? ת: קבלן מכסה את הטעויות שלו בבטון, הרופא מכסה באדמה...
לא	כן	5	4	3	2	1	0	12 אישה אחת יורדת לחבר שלה, עד שהוא גומר לה על הפנים. האישה שואלת אותו: "יש לך משהו לנגב?" הוא עונה: "יש חומוס במקרר"

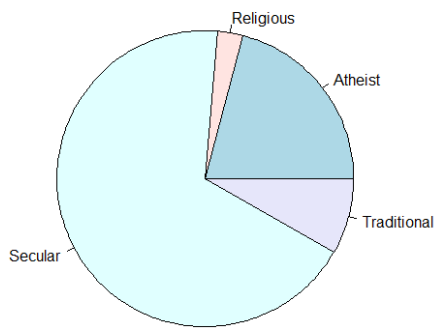
פרטים אישיים:

- גיל: \_\_\_\_\_
- מין: ז / נ
- פקולטה: \_\_\_\_\_ חוג: \_\_\_\_\_
- תואר: תעודה / ראשון / שני / שלישי
- השקפה דתית: דתי / מסורתי / חילוני / אתאיסט
- מוצא: אשכנזי / ספרדי / ערבי / אחר: \_\_\_\_\_
- מצב משפחתי: רווק / יש בן/בת זוג / נשוי / אחר: \_\_\_\_\_ מספר ילדים: \_\_\_\_\_
- אזור מגורים: צפון / שרון / מרכז / ירושלים והסביבה / דרום
- מצב מגורים: הורים / בן/בת זוג / שותפים / לבד
- השקפה פוליטית: ימני / מרכז-ימין / מרכז / מרכז-שמאל / שמאלני

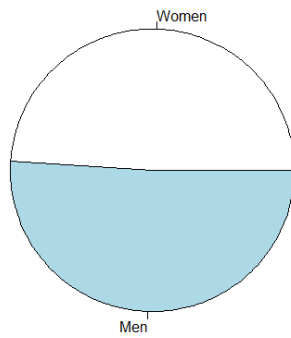
נספח 2: הצגות גרפיות לתמציות הנתונים



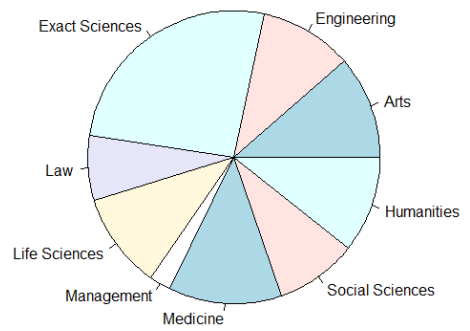
**Religious Views**



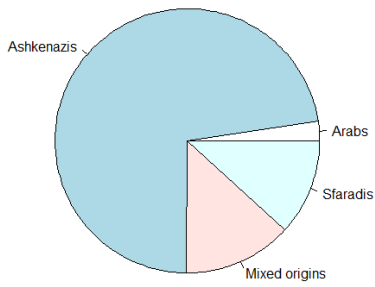
**Gender**



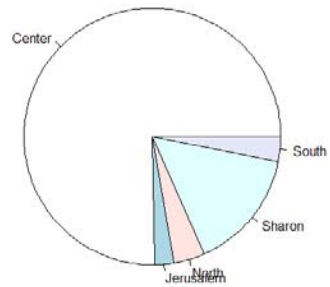
**Faculties**



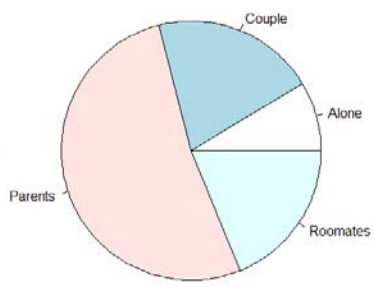
**Ethnic origin**



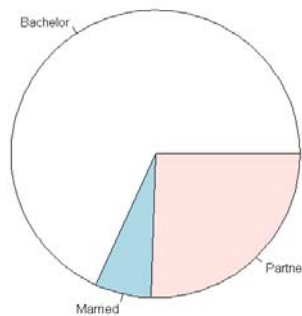
**Living area**



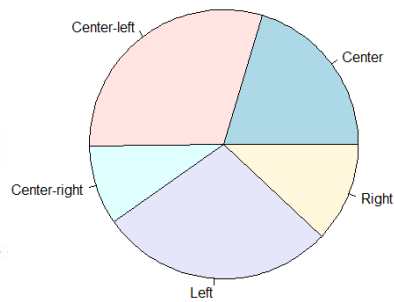
**Living status**



**Family Status**



**Political views**





```

# Data retrieving #
#####
setwd("C:/Users/Kiril/Documents/Studies/Courses/Statistics For CS/Project")
getwd()
data=read.csv("survey_res.csv")
data

#####
#           #
#   Chapter 2   #
#           #
#####

# general var declerations #
#####

alpha=0.05

# data summaries #
#####

age=data$age
age=age[!is.na(age)]

# location summaries:
age_mean_a_cut=mean(age, trim=alpha)
age_mean_a_cut
age_mean=mean(age)
age_mean
age_median=median(age)
age_median

# deviation summaries:
age_MAD=mad(age)
age_MAD
age_sd=sd(age)
age_sd

# skewness summaries:
quan_mean=(quantile(age,0.25)+quantile(age,0.75))/2
age_YULE=(quan_mean-age_median)/quan_mean
age_YULE      # ignore the "25%"

sex=data$sex
f_ratio=0
for (i in 1:length(sex)){
  if (sex[i]=="f") f_ratio=f_ratio+1
}
f_ratio=f_ratio/length(sex)
m_ratio=1-f_ratio

f_ratio
m_ratio

faculties=data$faculty
faculties=faculties[faculties[!==""]]
faculties=table(faculties)
faculties
prop.table(faculties)

religion=data$religion
religion=religion[religion[!==""]]
religion=table(religion)
religion
prop.table(religion)

ethnic=data$ethnic
ethnic=ethnic[ethnic[!==""]]
ethnic=table(ethnic)
ethnic
prop.table(ethnic)

fam_stat=data$fam_stat
fam_stat=fam_stat[fam_stat[!==""]]

```

```

fam_stat=table(fam_stat)
fam_stat
prop.table(fam_stat)

living_stat=data$living_stat
living_stat=living_stat[living_stat[!==""]]
living_stat=table(living_stat)
living_stat
prop.table(living_stat)

living_area=data$living_area
living_area=living_area[living_area[!==""]]
living_area=table(living_area)
living_area
prop.table(living_area)

political=data$political
political=political[political[!==""]]
political=table(political)
political
prop.table(political)

# results summaries #
#####

keresh1=data$j01
# keresh1=keresh1[!is.na(keresh1)]
length(keresh1)

keresh2=data$j11
# keresh2=keresh2[!is.na(keresh2)]
length(keresh2)

black1=data$j06
# black1=black1[!is.na(black1)]
length(black1)

black2=data$j10
# black2=black2[!is.na(black2)]
length(black2)

holoc1=data$j03
# holoc1=holoc1[!is.na(holoc1)]
length(holoc1)

holoc2=data$j09
# holoc2=holoc2[!is.na(holoc2)]
length(holoc2)

ethnic1=data$j04
# ethnic1=ethnic1[!is.na(ethnic1)]
length(ethnic1)

ethnic2=data$j08
# ethnic2=ethnic2[!is.na(ethnic2)]
length(ethnic2)

volgar1=data$j05
# volgar1=volgar1[!is.na(volgar1)]
length(volgar1)

volgar2=data$j12
# volgar2=volgar2[!is.na(volgar2)]
length(volgar2)

gender1=data$j02
# gender1=gender1[!is.na(gender1)]
length(gender1)

gender2=data$j07
# gender2=gender2[!is.na(gender2)]
length(gender2)

jokes=matrix(c(keresh1, keresh2, black1, black2, holoc1, holoc2, ethnic1, ethnic2, volgar1, volgar2, gender1,
gender2),255,12)

```

```

jokes

# jokes summaries: mean, median, sd, MAD, YULE
jokes_summaries=matrix(,5,12)
for (i in 1:12){
  # mean
  temp=jokes[,i]
  jokes_summaries[1,i]=mean(temp[!is.na(temp)])
}
for (i in 1:12){
  # median
  temp=jokes[,i]
  jokes_summaries[2,i]=median(temp[!is.na(temp)])
}
for (i in 1:12){
  # sd
  temp=jokes[,i]
  jokes_summaries[3,i]=sd(temp[!is.na(temp)])
}
for (i in 1:12){
  # MAD
  temp=jokes[,i]
  jokes_summaries[4,i]=mad(temp[!is.na(temp)])
}
for (i in 1:12){
  # YULE
  temp=jokes[,i]
  temp=temp[!is.na(temp)]
  quan_mean=(quantile(temp,0.25)+quantile(temp,0.75))/2
  temp_YULE=(quan_mean-median(temp))/quan_mean
  jokes_summaries[5,i]=temp_YULE
}

jokes_summaries

genre1=append(as.character(data$s01), as.character(data$s11))
genre1=table(genre1)
genre1

genre2=append(as.character(data$s06), as.character(data$s10))
genre2=table(genre2)
genre2

genre3=append(as.character(data$s03), as.character(data$s09))
genre3=table(genre3)
genre3

genre4=append(as.character(data$s04), as.character(data$s08))
genre4=table(genre4)
genre4

genre5=append(as.character(data$s05), as.character(data$s12))
genre5=table(genre5)
genre5

genre6=append(as.character(data$s02), as.character(data$s07))
genre6=table(genre6)
genre6

# summaries graphs #
#####

plot(density(age), main='Age distribution')
hist(age)

sext=table(sex)
pie(sext, labels=c("Women", "Men"), main='Gender')

pie(faculties,
    labels=c("","Arts","Engineering","Exact Sciences","Law","Life Sciences","Management","Medicine","Social Sciences","Humanities"), main='Faculties')

pie(religion, labels=c("","Atheist","Religious","Secular","Traditional"), main='Religious Views')

pie(ethnic, labels=c("Arabs","Ashkenazis","Mixed origins","Sfaradis"), main='Ethnic origin')

pie(fam_stat, labels=c("Bachelor","Married","Partner"), main="Family Status")

pie(living_area, labels=c("Center","Jerusalem","North","Sharon","South"), main='Living area')

```

```

pie(living_stat, labels=c("Alone", "Couple", "Parents", "Roomates"), main='Living status')

pie(political, labels=c("", "Center", "Center-left", "Center-right", "Left", "Right"), main='Political views')

#####
#                               #
#       Chapter 3               #
#                               #
#####

### 1.1 ###

origin_vs_ejokes=data.frame(data$ethnic,data$j04,data$j08)
origin_vs_ejokes

# X1_bar, X2_bar:

X1=origin_vs_ejokes$data.j04[origin_vs_ejokes$data.ethnic=="ashkenaz"]
X1_mean=mean(X1)
X1_var_hat=var(X1)

X2=origin_vs_ejokes$data.j04[origin_vs_ejokes$data.ethnic=="sfarad"]
X2_mean=mean(X2)
X2_var_hat=var(X2)

X1X2_var_hat=((length(X1)-1)*X1_var_hat+(length(X2)-1)*X2_var_hat)/(length(X1)+length(X2)-2)
X1X2_var_hat
X1_mean-X2_mean

C1_alpha_crit_low=0-qt((1-alpha/2),(length(X1)+length(X2)-2))*sqrt((length(X1)+length(X2))*X1X2_var_hat/(length(X1)*length(X2)))
C1_alpha_crit_low
C1_alpha_crit_hi=C1_alpha_crit_low*(-1)
C1_alpha_crit_hi

length(X1)
length(X2)

# X3_bar, X4_bar
X3=origin_vs_ejokes$data.j08[origin_vs_ejokes$data.ethnic=="ashkenaz"]
X3=X3[!is.na(X3)]
X3_mean=mean(X3)
X3_var_hat=var(X3)

X4=origin_vs_ejokes$data.j08[origin_vs_ejokes$data.ethnic=="sfarad"]
X4=X4[!is.na(X4)]
X4_mean=mean(X4)
X4_var_hat=var(X4)

X3X4_var_hat=((length(X3)-1)*X3_var_hat+(length(X4)-1)*X4_var_hat)/(length(X3)+length(X4)-2)
X3X4_var_hat
X3_mean-X4_mean

C2_alpha_crit_low=0-qt((1-alpha/2),(length(X3)+length(X4)-2))*sqrt((length(X3)+length(X4))*X3X4_var_hat/(length(X3)*length(X4)))
C2_alpha_crit_low
C2_alpha_crit_hi=C2_alpha_crit_low*(-1)
C2_alpha_crit_hi

length(X3)
length(X4)

### 1.2 ###

gender_vs_gjokes=data.frame(data$sex,data$j07,data$j02)
gender_vs_gjokes

# X5_bar, X6_bar:

X5=gender_vs_gjokes$data.j07[gender_vs_gjokes$data.sex=="m"]
X5=X5[!is.na(X5)]
X5_mean=mean(X5)
X5_var_hat=var(X5)

```

```

X6=gender_vs_gjokes$data.j07[gender_vs_gjokes$data.sex=="f"]
X6=X6[!is.na(X6)]
X6_mean=mean(X6)
X6_var_hat=var(X6)

X5X6_var_hat=((length(X5)-1)*X5_var_hat+(length(X6)-1)*X6_var_hat)/(length(X5)+length(X6)-2)
X5X6_var_hat
X5_mean-X6_mean

C3_alpha_crit_low=0-qt((1-alpha/2),(length(X5)+length(X6)-2))*sqrt((length(X5)+length(X6))*X5X6_var_hat/(length(X5)*length(X6)))
C3_alpha_crit_low
C3_alpha_crit_hi=C3_alpha_crit_low*(-1)
C3_alpha_crit_hi

length(X5)
length(X6)

# X7_bar, X8_bar:

X7=gender_vs_gjokes$data.j02[gender_vs_gjokes$data.sex=="m"]
X7=X7[!is.na(X7)]
X7_mean=mean(X7)
X7_var_hat=var(X7)

X8=gender_vs_gjokes$data.j02[gender_vs_gjokes$data.sex=="f"]
X8=X8[!is.na(X8)]
X8_mean=mean(X8)
X8_var_hat=var(X8)

X7X8_var_hat=((length(X7)-1)*X7_var_hat+(length(X8)-1)*X8_var_hat)/(length(X7)+length(X8)-2)
X7X8_var_hat
X7_mean-X8_mean

C4_alpha_crit_low=0-qt((1-alpha/2),(length(X7)+length(X8)-2))*sqrt((length(X7)+length(X8))*X7X8_var_hat/(length(X7)*length(X8)))
C4_alpha_crit_low
C4_alpha_crit_hi=C4_alpha_crit_low*(-1)
C4_alpha_crit_hi

length(X7)
length(X8)

### 1.3 ###

a1=data$j05
a2=data$j12
a3=(a1+a2)/2
fam_vs_vjokes=data.frame(data$fam_stat,a3)
fam_vs_vjokes

# X9_bar, X10_bar:

X9=fam_vs_vjokes$a3[fam_vs_vjokes$data.fam_stat=="bachelor"]
X9=X9[!is.na(X9)]
X9
X9_mean=mean(X9)
X9_var_hat=var(X9)

X10=fam_vs_vjokes$a3[fam_vs_vjokes$data.fam_stat!="bachelor"]
X10=X10[!is.na(X10)]
X10
X10_mean=mean(X10)
X10_var_hat=var(X10)

X9X10_var_hat=((length(X9)-1)*X9_var_hat+(length(X10)-1)*X10_var_hat)/(length(X9)+length(X10)-2)
X9X10_var_hat
X9_mean-X10_mean

C5_alpha_crit_low=0-qt((1-alpha/2),(length(X9)+length(X10)-2))*sqrt((length(X9)+length(X10))*X9X10_var_hat/(length(X9)*length(X10)))
C5_alpha_crit_low
C5_alpha_crit_hi=C5_alpha_crit_low*(-1)
C5_alpha_crit_hi

length(X9)

```

```

length(X10)

### 1.4 ###

b1=data$j03
b2=data$j09
b3=(b1+b2)/2
ethnic_vs_hjokes=data.frame(data$ethnic,b3)
ethnic_vs_hjokes

# X11_bar, X12_bar:

X11=ethnic_vs_hjokes$b3[ethnic_vs_hjokes$data.ethnic=="ashkenaz"]
X11=X11[!is.na(X11)]
X11
X11_mean=mean(X11)
X11_var_hat=var(X11)

X12=ethnic_vs_hjokes$b3[ethnic_vs_hjokes$data.ethnic=="sfarad"]
X12=X12[!is.na(X12)]
X12
X12_mean=mean(X12)
X12_var_hat=var(X12)

X11X12_var_hat=((length(X11)-1)*X11_var_hat+(length(X12)-1)*X12_var_hat)/(length(X11)+length(X12)-2)
X11X12_var_hat
X11_mean-X12_mean

C6_alpha_crit_low=0-qt((1-alpha/2),(length(X11)+length(X12)-2))*sqrt((length(X11)+length(X12))*X11X12_var_hat/(length(X11)*length(X12)))
C6_alpha_crit_low
C6_alpha_crit_hi=C6_alpha_crit_low*(-1)
C6_alpha_crit_hi

length(X11)
length(X12)

### 2 general ###

fac=c(as.character(data$faculty))
for (i in 1:11){
  fac=c(as.character(fac),as.character(data$faculty))
}
fac_jokes=c(data$j01,data$j02,data$j03,data$j04,data$j05,data$j06,data$j07,data$j08,data$j09,data$j10,data$j11,data$j12)
faculty_jokes=data.frame(fac,fac_jokes)
faculty_jokes

# Arts faculty

Xarts=faculty_jokes$fac_jokes[faculty_jokes$fac=="arts"]
Xarts=Xarts[!is.na(Xarts)]
Xarts
Xarts_mean=mean(Xarts)
Xarts_var_hat=var(Xarts)
Xarts_mean
Xarts_var_hat

# Engeneering faculty

Xengi=faculty_jokes$fac_jokes[faculty_jokes$fac=="engineering"]
Xengi=Xengi[!is.na(Xengi)]
Xengi
Xengi_mean=mean(Xengi)
Xengi_var_hat=var(Xengi)
Xengi_mean
Xengi_var_hat

# Exact Sciences faculty

Xexac=faculty_jokes$fac_jokes[faculty_jokes$fac=="exact"]
Xexac=Xexac[!is.na(Xexac)]
Xexac
Xexac_mean=mean(Xexac)
Xexac_var_hat=var(Xexac)

```

```

Xexac_mean
Xexac_var_hat

# Law faculty

Xlaw=faculty_jokes$fac_jokes[faculty_jokes$fac=="law"]
Xlaw=Xlaw[!is.na(Xlaw)]
Xlaw
Xlaw_mean=mean(Xlaw)
Xlaw_var_hat=var(Xlaw)
Xlaw_mean
Xlaw_var_hat

# Life sciences faculty

Xlife=faculty_jokes$fac_jokes[faculty_jokes$fac=="life"]
Xlife=Xlife[!is.na(Xlife)]
Xlife
Xlife_mean=mean(Xlife)
Xlife_var_hat=var(Xlife)
Xlife_mean
Xlife_var_hat

# Management faculty

Xmana=faculty_jokes$fac_jokes[faculty_jokes$fac=="management"]
Xmana=Xmana[!is.na(Xmana)]
Xmana
Xmana_mean=mean(Xmana)
Xmana_var_hat=var(Xmana)
Xmana_mean
Xmana_var_hat

# Medicine faculty

Xmed=faculty_jokes$fac_jokes[faculty_jokes$fac=="medical"]
Xmed=Xmed[!is.na(Xmed)]
Xmed
Xmed_mean=mean(Xmed)
Xmed_var_hat=var(Xmed)
Xmed_mean
Xmed_var_hat

# Social sciences faculty

Xsoc=faculty_jokes$fac_jokes[faculty_jokes$fac=="social"]
Xsoc=Xsoc[!is.na(Xsoc)]
Xsoc
Xsoc_mean=mean(Xsoc)
Xsoc_var_hat=var(Xsoc)
Xsoc_mean
Xsoc_var_hat

# Wind faculty

Xwind=faculty_jokes$fac_jokes[faculty_jokes$fac=="wind"]
Xwind=Xwind[!is.na(Xwind)]
Xwind
Xwind_mean=mean(Xwind)
Xwind_var_hat=var(Xwind)
Xwind_mean
Xwind_var_hat

### irrelevant ###
# graphs:

#fac_means=c(Xarts_mean,Xengi_mean,Xexac_mean,Xlaw_mean,Xlife_mean,Xmana_mean,Xmed_mean,Xsoc_mean,Xwind_mean)
#fac_means=sort(fac_means)
#fac_means      # vector of sorted means by faculty

#fac_var=c(Xarts_var_hat,Xengi_var_hat,Xexac_var_hat,Xlaw_var_hat,Xlife_var_hat,Xmana_var_hat,Xmed_var_hat,Xsoc_var_hat,Xwind_var_hat)
#fac_var=sort(fac_var)
#fac_var      # vector of sorted vars by faculty

```

```
### 2.1 ###
```

```
# Not Arts faculty
```

```
Xnarts=faculty_jokes$fac_jokes[faculty_jokes$fac!="arts"]
Xnarts=Xnarts[!is.na(Xnarts)]
Xnarts
Xnarts_mean=mean(Xnarts)
Xnarts_var_hat=var(Xnarts)
Xnarts_mean
Xnarts_var_hat
```

```
# calculation
```

```
XartsXnarts_var_hat=((length(Xarts)-1)*Xarts_var_hat+(length(Xnarts)-
1)*Xnarts_var_hat)/(length(Xarts)+length(Xnarts)-2)
XartsXnarts_var_hat
Xarts_mean-Xnarts_mean
```

```
C7_alpha_crit_hi=0+qt((1-alpha),(length(Xarts)+length(Xnarts)-
2))*sqrt((length(Xarts)+length(Xnarts))*XartsXnarts_var_hat/(length(Xarts)*length(Xnarts)))
C7_alpha_crit_hi
```

```
### 2.2 ###
```

```
# Not Medical faculty
```

```
Xnmed=faculty_jokes$fac_jokes[faculty_jokes$fac!="medical"]
Xnmed=Xnmed[!is.na(Xnmed)]
Xnmed
Xnmed_mean=mean(Xnmed)
Xnmed_var_hat=var(Xnmed)
Xnmed_mean
Xnmed_var_hat
```

```
# calculation
```

```
XmedXnmed_var_hat=((length(Xmed)-1)*Xmed_var_hat+(length(Xnmed)-1)*Xnmed_var_hat)/(length(Xmed)+length(Xnmed)-2)
XmedXnmed_var_hat
Xmed_mean-Xnmed_mean
```

```
C8_alpha_crit_hi=0-qt((1-alpha),(length(Xmed)+length(Xnmed)-
2))*sqrt((length(Xmed)+length(Xnmed))*XmedXnmed_var_hat/(length(Xmed)*length(Xnmed)))
C8_alpha_crit_hi
```

```
### 3.1 ###
```

```
all_yes_no=c(as.character(data$s01),as.character(data$s02),as.character(data$s03),as.character(data$s04),as.char
acter(data$s05),as.character(data$s06),as.character(data$s07),as.character(data$s08),as.character(data$s09),as.c
haracter(data$s10),as.character(data$s11),as.character(data$s12))
all_yes_no=all_yes_no[all_yes_no!=""]
prop=table(all_yes_no)# 1289 nos, 1475 yeses
yes_prop=1475/(1475+1289)
yes_prop
yes_no_norm_var_hat=(yes_prop)*(1-yes_prop)/length(all_yes_no)
yes_no_norm_var_hat
```

```
all_jokes=fac_jokes # already calculated in 2.1
all_jokes=all_jokes[!is.na(all_jokes)]
```

```
all_jokes_mean=mean(all_jokes)
all_jokes_mean
all_jokes_var_hat=var(all_jokes)
all_jokes_var_hat
jokes_mean_prop=all_jokes_mean/5
jokes_mean_prop
```

```
yes_vs_jokes_mean=jokes_mean_prop-yes_prop
yes_vs_jokes_mean
```

```
var_hat=((length(all_yes_no)-
1)*yes_no_norm_var_hat+length(all_jokes)*all_jokes_var_hat)/(length(all_yes_no)+length(all_jokes)-2)
var_hat
```



```

C_alpha_crit_low=0-qt((1-alpha/2),(length(all_yes_no)+length(all_jokes)-
2))*sqrt((length(all_yes_no)+length(all_jokes))*var_hat/(length(all_yes_no)*length(all_jokes)))
C_alpha_crit_low
C_alpha_crit_hi=C_alpha_crit_low*(-1)
C_alpha_crit_hi

### 3.2 ###

all_yes_no=c(as.character(data$s01),as.character(data$s02),as.character(data$s03),as.character(data$s04),as.char
acter(data$s05),as.character(data$s06),as.character(data$s07),as.character(data$s08),as.character(data$s09),as.c
haracter(data$s10),as.character(data$s11),as.character(data$s12))
faculty_jokes_yesno=data.frame(fac,fac_jokes,all_yes_no)
faculty_jokes_yesno

# exact sciences:

exact_jokes=faculty_jokes_yesno$fac_jokes[faculty_jokes$fac=="exact"]
exact_jokes=exact_jokes[!is.na(exact_jokes)]
exact_jokes
exact_yes_no=faculty_jokes_yesno$all_yes_no[faculty_jokes$fac=="exact"]
exact_yes_no=exact_yes_no[exact_yes_no!=""]
exact_yes_no

tmp=0
for (i in 1:length(exact_yes_no)){
  if (exact_yes_no[i]=="y") tmp=tmp+1
}

exact_yes_prop=tmp/length(exact_yes_no)
exact_yes_prop
exact_yes_no_norm_var_hat=exact_yes_prop*(1-exact_yes_prop)/length(exact_yes_no)
exact_yes_no_norm_var_hat

exact_jokes_mean=mean(exact_jokes)
exact_jokes_mean_prop=exact_jokes_mean/5
exact_jokes_mean_prop
exact_jokes_var_hat=var(exact_jokes)
exact_jokes_var_hat

exact_yes_vs_jokes_mean=exact_jokes_mean_prop-exact_yes_prop
exact_yes_vs_jokes_mean

exact_var_hat=((length(exact_yes_no)-
1)*exact_yes_no_norm_var_hat+length(exact_jokes)*exact_jokes_var_hat)/(length(exact_yes_no)+length(exact_jokes)-
2)
exact_var_hat

Cexact_alpha_crit_low=0-qt((1-alpha/2),(length(exact_yes_no)+length(exact_jokes)-
2))*sqrt((length(exact_yes_no)+length(exact_jokes))*exact_var_hat/(length(exact_yes_no)*length(exact_jokes)))
Cexact_alpha_crit_low
Cexact_alpha_crit_hi=Cexact_alpha_crit_low*(-1)
Cexact_alpha_crit_hi

# social sciences:

social_jokes=faculty_jokes_yesno$fac_jokes[faculty_jokes$fac=="social"]
social_jokes=social_jokes[!is.na(social_jokes)]
social_jokes
social_yes_no=faculty_jokes_yesno$all_yes_no[faculty_jokes$fac=="social"]
social_yes_no=social_yes_no[social_yes_no!=""]
social_yes_no

tmp=0
for (i in 1:length(social_yes_no)){
  if (social_yes_no[i]=="y") tmp=tmp+1
}

social_yes_prop=tmp/length(social_yes_no)
social_yes_prop
social_yes_no_norm_var_hat=social_yes_prop*(1-social_yes_prop)/length(social_yes_no)
social_yes_no_norm_var_hat

social_jokes_mean=mean(social_jokes)
social_jokes_mean_prop=social_jokes_mean/5
social_jokes_mean_prop
social_jokes_var_hat=var(social_jokes)

```

```

social_jokes_var_hat

social_yes_vs_jokes_mean=social_jokes_mean_prop-social_yes_prop
social_yes_vs_jokes_mean

social_var_hat=((length(social_yes_no)-
1)*social_yes_no_norm_var_hat+length(social_jokes)*social_jokes_var_hat)/(length(social_yes_no)+length(social_jokes)-2)
social_var_hat

Csocial_alpha_crit_low=0-qt((1-alpha/2),(length(social_yes_no)+length(social_jokes)-2))*sqrt((length(social_yes_no)+length(social_jokes))*social_var_hat/(length(social_yes_no)*length(social_jokes)))
Csocial_alpha_crit_low
Csocial_alpha_crit_hi=Csocial_alpha_crit_low*(-1)
Csocial_alpha_crit_hi

# arts:

arts_jokes=faculty_jokes_yesno$fac_jokes[faculty_jokes$fac=="arts"]
arts_jokes=arts_jokes[!is.na(arts_jokes)]
arts_jokes
arts_yes_no=faculty_jokes_yesno$all_yes_no[faculty_jokes$fac=="arts"]
arts_yes_no=arts_yes_no[arts_yes_no!=""]
arts_yes_no

tmp=0
for (i in 1:length(arts_yes_no)){
  if (arts_yes_no[i]=="y") tmp=tmp+1
}

arts_yes_prop=tmp/length(arts_yes_no)
arts_yes_prop
arts_yes_no_norm_var_hat=arts_yes_prop*(1-arts_yes_prop)/length(arts_yes_no)
arts_yes_no_norm_var_hat

arts_jokes_mean=mean(arts_jokes)
arts_jokes_mean_prop=arts_jokes_mean/5
arts_jokes_mean_prop
arts_jokes_var_hat=var(arts_jokes)
arts_jokes_var_hat

arts_yes_vs_jokes_mean=arts_jokes_mean_prop-arts_yes_prop
arts_yes_vs_jokes_mean

arts_var_hat=((length(arts_yes_no)-
1)*arts_yes_no_norm_var_hat+length(arts_jokes)*arts_jokes_var_hat)/(length(arts_yes_no)+length(arts_jokes)-2)
arts_var_hat

Carts_alpha_crit_low=0-qt((1-alpha/2),(length(arts_yes_no)+length(arts_jokes)-2))*sqrt((length(arts_yes_no)+length(arts_jokes))*arts_var_hat/(length(arts_yes_no)*length(arts_jokes)))
Carts_alpha_crit_low
Carts_alpha_crit_hi=Carts_alpha_crit_low*(-1)
Carts_alpha_crit_hi

```