מבוא לסטטיסטיקה והסתברות / ד"ר דוד סרף

שעור מס' 8 : הסקה סטטיסטית – רב"ס לבדיקת השערות המשך.

הפרש ממוצעים שונות ידועה – ערך ערכשו טפלון באוכלוסיה אחת ומלחוני כוחו סמוך לבדיקת השערות כאשר השונות ידועה ולא ידועה. נטפל ערכשו במערכת הכוללת 2 אוכלוסיות ונדבר על הפרש ממוצעים בין 2 האוכלוסיות. הטפול היא בדה בכל כמו לאוכלוסיה אחת, אלא שהפרש הבטוי לטעות הדגימה משתנה.

$$A = z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

כאשר n_1, σ_1 מתיחסים לאוכלוסיה 1

n_2, σ_2 מתיחסים לאוכלוסיה 2

יש לשים לב שאם היתה רק אוכלוסיה אחת היינו מקבלים $A = z\sqrt{\frac{\sigma^2}{n}} = \frac{z\sigma}{\sqrt{n}}$ המתאים למה שלמדנו קודם.

תרגיל: יצרן שטיחים התקין בשער פועלי נמל חיפה היא 2000 ש"ח ובשער פועלי נמל אשדוד 2500 ש"ח. נלקח מדגם של 16 פועלי חיפה ונתקבל ממוצע שכר 22000, ומדגם של 25 מפועלי אשדוד ונתקבל ממוצע 23500. מצא ברמת בטחון 90%

א. רב"ס לשכר פועלי נמל חיפה.

ב. פועלי חיפה טוענים שהכרת נמוך מזה אשדוד. בדוק טענה זו.

א. לחישוב רב"ס לפועלי חיפה יש קבוצה אחת בלבד ולכן

$$A = \frac{z \cdot \sigma}{\sqrt{n}} = \frac{1.645 * 2000}{\sqrt{16}} = 822.5 \qquad \bar{x} = 22000 \mp 822.5$$

$$(21177.5, 22822.5)$$

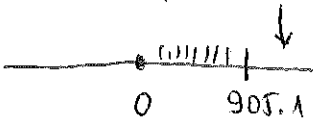
ב. לבדיקת ההשערה יש להתיחס לד 2 הקבוצות כלומר להפרש ממוצעים שונות ידועות

תש"ס-2003

$$A = z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 1.28\sqrt{\frac{2000^2}{16} + \frac{2500^2}{25}} = 905.1$$

לא כן H_0: $\mu_{\text{חיפה}} - \mu_{\text{אשדוד}} \geq 0$ לא כן H_1: $\mu_{\text{חיפה}} < \mu_{\text{אשדוד}}$

(23500 − 22000)



אם אין הבדל הפער הוא אפס ומתרחת טעות דגימה של 905 בין שהפער בפועל הוא 1500 (23500 − 22000) זהו פער מהותי ולפיכך נקבל את H_1 כלומר שכר פועלי חיפה נמוך יותר.

תרגיל נוסף – יודע שלהתפלגות הציונים בסטטיסטיקה שונות $\sigma^2=100$
ב 2 הקבוצות סטודנטים נלקחו מדגמים של 50 ונתקבלו
ממוצע של 80 בקבוצה הראשונה ו 70 בקבוצה השניה. ענה על השאלות הבאות
א. בנה רב"ס להפרש ממוצעי הציונים ב 2 הקבוצות ברמת בטחון 95%
ב. בדוק את ההשערה שברמת בטחון 95% אין הבדל משמעותי בציוני 2 הקבוצות
ג. מה רמת המובהקות המינימלית (רמת בטחון מקסימלית) שבה נדחה את הטענה
שממוצע הציונים בקבוצה הראשונה אינו גבוה מצו שבשניה.

פתרון: א.

קבוצה ראשונה	קבוצה שניה
$\sigma_1^2=100$	$\sigma_2^2=100$
$n_1=50$	$n_2=50$
$\bar{x}=80$	$\bar{x}=70$

רווח סמך בן-צדדי ולכן $z(0.975)=1.96$

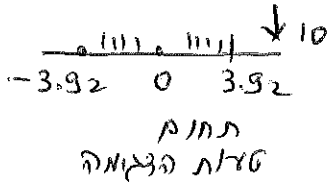
$$A=z\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}=1.96\sqrt{\frac{100}{50}+\frac{100}{50}}=3.92$$

הפרש הממוצעים הוא $80-70=10$ לכן רב"ס 10 ∓ 3.92 $(6.08,13.92)$

ב. $H_0: \mu_1=\mu_2$ (אין הבדל)
$H_1: \mu_1\neq\mu_2$ (יש הבדל)

בדיקת השערות בן-צדדית.

$$A=1.96\sqrt{\frac{100}{50}+\frac{100}{50}}=3.92$$



הפער גדול מתחום טעות הדגימה לכן נקבל את H_1 והטענה שיש הבדל משמעותי

ג. הפרם אינו מתפשרם את z בגודל $A=80-70$ שנה ההפרש בין הקבוצות

$$80-70=z\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}=z\cdot\sqrt{\frac{100}{50}+\frac{100}{50}}$$

$10=2z$
$5=z$

כאשר $z=5$ השטח הוא כל השטח בין שהטבלה
מסתיימת בסביבות 4 לפיכך המובהקות 0 ורמת
הבטחון 1.0 כלומר אנו יכולים לטעון שבודאות של 100%
יש הבדל בין 2 הקבוצות.

אם נניח שפער הציונים היה 3 נקודות ולא 10, אזי היינו מקבלים
$3=2z \Leftarrow z=1.5$, $\Phi(1.5)=0.9332$ וכיון שהבעיה בן-צדדית
תהיה רמת הבטחון 0.8664 (ראה תרגיל נוסף 3 בהרצאה 7)

הפרש ממוצעים שונות לא ידועות, אך שוות: באשר השונות של 2 הקבוצות לא ידועות ומקבלים אותן, מתוך המודגם ניתן להבחין ב 2 מקרים. מקרה אחד, שהן נבדלו מאוכלוסיות בעלות שונות ופתרונן דורש תהליך מסובך, שלא יטופל אצלנו, או המקרה שלנו, שבו ננים שהשונות אמנם לא ידועות אך שוות ב 2 האוכלוסיות (אבל ניתן באופן סטטיסטי לקבוע מתוך הפרש השונות שבמודגמים, אם הם נלקחו מאוכלוסיות בהן אמנם יש הבדל בינהם או לא. גם כאן לא נעשה נסות, ואם הבדיה תהיה שונות לא ידועות, ננים שהן באו מאוכלוסיות בעלות אותה שונות). במקרה זה כמובן נדרש להשתמש ב t ולא ב z וגם לקבוע מהי השונות השווה באוכלוסיה. לצורך זה נעשה שקלול מהשונות של 2 הקבוצות והערך המשוקלל יקרא S_p^2

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

מספר דרגות החופש עבור t יהיה $t(n_1+n_2-2$ כמות הטחון)

תרגיל: נלקח מודגם לקביעת שכר פועלי תיפה בגודל 16 ונתקבלו ממנו שכר ממוצע 22000 וסטיית תקן 2000. במודגם מקביל של פועלי אשדוד בגודל 25 נתקבל ממוצע 23500 וסטיית תקן 2500

א. מצאו רב"ס להפרש השכר בין 2 הנמלים. ב. בדוק את הטענה ששכר פועלי אשדוד גבוה יותר. 2 הסעיפים ברמת הטחון 90%

פ"ו בעיה של הפרש ממוצעים שונות לא ידועות, אך שוות.

נחשב שונות משוקללת S_p^2

$$S_p^2 = \frac{(16-1)\,2000^2 + (25-1)\cdot 2500^2}{16+25-2} = 5384615 \qquad S_p = 2320.48$$

שימו לב שהשונות המשוקללת S_p היא בין 2 השונות 2000 ו 2500 קרובה יותר ל 2500 (מדוע?) וזה משמש בבקרות שהתצלה הבינונית.

נחשב את A

$$A = t(16+25-2,\ 0.95)\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}} =$$

בין שמספר דרגות החופש 39 בדת"מ מתאר להשתמש בערך של Z

$$= 1.65\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 1.65 * 2320.48 \sqrt{\frac{1}{16} + \frac{1}{25}} = 1225.8 \qquad z(0.95) = 1.65$$

הרב"ס יהיה עבור הפרש הממוצעים $(23500 - 22000) \mp 1225.8$

$$(\ 274.2\ ,\ 2725.81\ )$$

ב. בדיקת הטענה ששכר פועלי אשפוז גבוה יותר.

לכן $H_0: \mu_{\text{אשפוז}} = \mu_{\text{חיפה}}$ $\quad H_1: \mu_{\text{אשפוז}} > \mu_{\text{חיפה}}$

אין $H_0: \mu_{\text{אשפוז}} - \mu_{\text{חיפה}} = 0$ $\quad H_1: \mu_{\text{אשפוז}} - \mu_{\text{חיפה}} > 0$

בעיה חד-צדדית

בגלל $df=39$ מותר לקחת z

$$A = t(39, 0.9)\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$$

$$= 1.28 * S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 1.28 * 2320.48 \sqrt{\frac{1}{16} + \frac{1}{25}} = 950.3$$

$(23500 - 22000)$

0 $\quad$ 950.3

הפער הפועל בין המוצאים גבוה מתחום טעות הדגימה לכן נקבל H_1 כלומר אכן שכר פועלי אשפוז גבוה מחיפה

<u>פרופורציה</u>: נניח שמעוניינים לאמוד את אחוז מחציקי הרכב בקרב הסטודנטים של שעקר. כדי בדגם בעיה בינומית שבה 2 תשובות: האמת אחוז מחציקי הרכב והשניה המשלים שלה כלומר האחוז שלא מחציק רכב. בעיות מן סוג הזה אנו מגדירים בעיות פרופורציה. טעות הדגימה במקרה זה תהיה

$$A = z\sqrt{\frac{pq}{n}}$$

כאשר p הוא הפרופורציה ו $q=1-p$. הסוג זה מצמצם את התוחלת בהתפלגות בינומית.

נפתור בעיה לדוגמא. כדי לאמוד את אחוז מחציקי הרכב בשעקר (פרופורציה) נקח מדגם של 100 סטודנטים ונמצא שלו 30 מהם היה רכב. בנה רווח סמך ברמת הסתמכות 90% לאחוז מחציקי הרכב בכלל אוכלוסיית הסטודנטים.

במקרה זה $p = \frac{30}{100} = 0.3$ $\quad q = 0.7$ $\quad n = 100$

$$A = z(0.95) * \sqrt{\frac{pq}{n}} = 1.645\sqrt{\frac{0.7*0.3}{100}} = 0.075$$

הרווח יהיה $p \mp A$ כלומר 0.3 ∓ 0.075 $(0.225, 0.375)$

תרגיל נוסף: מתוך מדגם נקבע רווח לפרופורציית מחציקי רכב בשעקר $(0.225, 0.375)$ ברמת הסתמכות 90%. מה גודל המדגם שנלקח לשם כך?

כדי שמאוזן יהיה ממוצע $p_{\text{ממוצע}} = \frac{0.375+0.225}{2} = 0.3$ ו $2A = 0.15$, $A = 0.075$

כדי למצוא את n

$$0.075 = A = 1.645\sqrt{\frac{0.3*0.7}{n}}$$

$$\sqrt{n} = \frac{1.645 * \sqrt{0.3*0.7}}{0.075} = 10.05 \Rightarrow n = 10.05^2 = 101.025$$

כלומר גודל המדגם 102

תרגיל נוסף: במדינה 2 מפלגות. מעוניינים למצוא את גודל המדגם כך שברמת בטחון 90% תהיה טעות הדגימה 4%.

במקרה זה $A=0.04$, רמת הבטחון דו-צדדית 90%. כיוון שאין מידע מוקדם נתונים לפרופורציה יוצא $p=0.5$ (בלומר סביר שווה להצביע או לא להצביע לאחת המפלגות)

$$0.04 = A = Z(0.95)\sqrt{\frac{pq}{n}} = 1.645 * \sqrt{\frac{0.5\times 0.5}{n}}$$

$$\sqrt{n} = \frac{1.645 * 0.5}{0.04} = 20.5625 \qquad n=422.81 \rightarrow n=423$$

בלומר נחוץ מדגם של 423 בוחרים.

אם ידוע מידע שבעבר רבוע אחת המפלגות הצביעו 30% ורבוע השניה 70% נוכל לנצל זאת ולרשום $p=0.3$

$$0.04 = 1.645\sqrt{\frac{0.7\times 0.3}{n}}$$

$$\sqrt{n} = \frac{1.645 * \sqrt{21}}{0.04} = 18.84 \qquad n=355$$

שימו לב שבאשר יש מידע מוקדם תמיד יהיה גודל המדגם קטן מאשר בשאין מידע ו $p=0.5$

<u>נתונים מצומדים</u>: כאן מצב של 2 קבוצות, אלא שנתוני 2 הקבוצות מראים תלות ביניהם. למשל ציוני מועד א' ומועד ב' רבוע אותם אנשים שנבחנו בשניהם. במקרה זה אין מטפלים בבעיה בשיטת הפרש ממוצעים בין 2 הקבוצות, אלא אם $A=Z\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ אלא מתייחסים אליהם כקבוצה אחת. לכן הם נקראים נתונים מצומדים.

נראה זאת בדוגמא הבאה: להלן ציוני מדגם סטודנטים במועד א': 60, 50, 45, 65. אותם תלמידים נבחנו במועד ב' וקבלו בהתאמה 65, 70, 50, 55. ענה על השאלות הבאות:

א. האם הציון הממוצע של כלל הנבחנים ב 2 המועדים שונה ברמת בטחון 95%?

ב. בנה רווח סמך לציון הממוצע של כלל הסטודנטים במועד ב'.

סעיף ג' מתייחס לבדיקת השערות. כמו שנאמר בנתונים מצומדים מתייחסים לב 2 הקבוצות כקבוצה אחת בשיטת הפרשים

מועד א'	60	50	45	65	
מועד ב'	65	70	50	55	
הפרש d	5	20	5	-10	

מוצאים את הממוצע וסטיית התקן של הבדלי הפרשים d

$$\bar{d} = 5 \qquad s^2 = \frac{(-10-5)^2+(5-5)^2+\dots}{3} = 150$$

$$S=\sqrt{150} = 12.25$$

יש לשים לב שבחישוב סטיית התקן מחלקים ב n-1 כי בפיזיקה מיודעים.
עתה מסתכלים על קבוצת ההפרשים d ומחפשים אותה, לפי ממוצע
שונות לא ידועה (שפן חיבון את השונות)

$$A = \frac{t(3,\ 0.95)\cdot S}{\sqrt{n}} = \frac{2.35 \cdot 12.25}{\sqrt{4}} = 14.4$$

$H_0: d \le 5$

$H_1: d > 5$ חל שפיר

בין שהשיעור הפנימי 5
הוא נמוך מטעות הדגימה
נקבל את H_0 בלומר הבדיל הכללי במועדה ב' לא שונה. רק אם השפיר
במיוחדים היה מעל 14.4 היינו יכולים לטעון שאמנם במועד ב' היו
ביונים גבוהים יותר. אפשר הדרישה לפחות כל כך גדול כמו 14 נק'.
כאלה מבין שהמיוחדים מאוד קטן (4 נבחנים). אלו היה מספר נבחנים
גדול יותר היה מספיק פחות קטן יותר.

נסכם עתה את הטיפול בבעיות הסקה סטטיסטית.
א. יש להחליט אם מדובר ברווח סמך (תמיד דו-צדדי) או בבדיקת
השערות שיכולה להיות חד-צדדית או דו-צדדית.
ב. יש לקבוע אם השונות ידועה (שונות האוכלוסיה) או לא ידועה (שונות
המדגם הנתונה או מחושבת על ידינו)
ג. לשייך את הבעיה לקטגוריה מתאימה ולהשתמש בנוסחה הדרושה
להלן הקטגוריות שלמדנו

1. ממוצע שונות ידועה $A = \frac{z\sigma}{\sqrt{n}}$
2. ממוצע שונות לא ידועה $A = \frac{t\cdot S}{\sqrt{n}}$
3. הפרש ממוצעים שונות ידועות $A = z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
4. הפרש ממוצעים שונות לא ידועות אך שוות.

$$A = t(n_1+n_2-2)\cdot\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}} = t\cdot S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

5. פרופורציה $A = z\sqrt{\frac{pq}{n}}$
6. נתונים מזווגים. בונים סידרת הפרשים ומשתמשים או בממוצע שונות ידוע
או בממוצע שונות לא ידוע
לפי הנתונים.