מבוא להסתברות וסטטיסטיקה – 094480, אביב תשס"ט

(נכתב ע"י מיכל מנדל ואסף ליבוביץ)

דף עזר 1

ערכים מסכמים למ"מ בדיד:

מדדי מיקום: **ממוצע** – $\overline{X}=\frac{\sum_{i=1}^{n}x_i}{n}$

חציון – הערך שמחצית המקרים קטנים ממנו או שווים לו :

- עבור n אי-זוגי : $MED=X_{\left(\frac{n+1}{2}\right)}$ - עבור n זוגי : $MED=\frac{X_{\left(\frac{n}{2}\right)}+X_{\left(\frac{n}{2}+1\right)}}{2}$

רבעון ראשון/תחתון – הערך שרבע מהמקרים קטנים ממנו או שווים לו : $Q_1=X_{\left(\left\lfloor\frac{n}{4}\right\rfloor+1\right)}$. **רבעון שני** – חציון. **רבעון שלישי/עליון** – הערך ששלושה רבעים מהמקרים קטנים ממנו או שווים לו : $Q_3=X_{\left(\left\lfloor\frac{3n}{4}\right\rfloor+1\right)}$. $\lfloor\ \rfloor$ - ערך תחתון שלם, מעגלים כלפי מטה.

שכיח – הערך של המשתנה הנחקר בעל התדירות הגבוהה ביותר (ערך הנפוץ ביותר).

מדדי פיזור:

שונות–מודדת את הפיזור של ערכי הסדרה הסטטיסטית סביב הממוצע.

שונות מדגמית : $S^2=\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}{n-1}=\frac{\sum_{i=1}^{n}x_i^2-n\overline{x}^2}{n-1}$

שונות האוכלוסייה : $\sigma^2=\frac{\sum_{i=1}^{n}\left(x_i-\mu\right)^2}{n}$ **סטיית תקן** : $S=\sqrt{S^2}$, $\sigma=\sqrt{\sigma^2}$

מקדם פיזור : $C.V.=\frac{S}{\overline{X}}\cdot 100$

תחום בין-רבעוני : בו מרוכזות 50% מהתצפיות המרכזיות : $IQR=Q_3-Q_1$.

UIF, LIF : $\boxed{step=1.5\cdot(Q1-Q3)}$ $\boxed{LIF=Q_1-step}$ $\boxed{UIF=Q_3+step}$

בניית טבלת שכיחויות

1. נמנה את השכיחות (מספר) ערכי X עבור כל ערך במדגם $\left(\sum_i n_i=n\right)$.
2. חישוב שכיחות יחסית של מקרים בכל קטגוריה : $f_i=\frac{n_i}{n}$ - שכיחות יחסית.
3. חישוב שכיחות מצטברת ושכיחות יחסית מצטברת עבור כל קטגוריה.

חזרה על קומבינטוריקה בסיסית.

חוק הסכום :
אם אפשר לבחור איבר מסוג 1 ב- n_1 אופנים, איבר מסוג 2 ב- n_2 אופנים, ... , איבר מסוג k ב- n_k אופנים, והקבוצות זרות, אזי לבחירת בדיוק איבר אחד מבין k סוגי העצמים יש $n_1+n_2+...+n_k$ אופנים.

חוק הכפל:
נניח כי : מספר האפשרויות לבחור איבר מסוג 1 הוא n_1 .
מספר האפשרויות לבחור איבר מסוג 2 הוא n_2 . ...
מספר האפשרויות לבחור איבר מסוג k הוא n_k .
והקבוצות זרות, אזי מספר האפשרויות לבחירת קבוצה בת k איברים, כך שיהיה איבר אחד מכל סוג, הוא $n_1*n_2*...*n_k$.

חליפות:
נתונה קבוצה של n איברים שונים (הניתנים להבדלה). וקטור סדור באורך r, שאבריו מתוך n האיברים הנ"ל ושכל רכיביו שונים זה מזה, נקרא חליפה של r איברים מתוך n (r< n).

מספר החליפות : $\boxed{P(n,r)=(n)_r=\frac{n!}{(n-r)!}}$

תמורות:
חליפה של n מתוך n נקראת תמורה (פרמוטציה) של n איברים.

מספר התמורות : $\boxed{P(n,n)=n!}$

צירופים:
נתונה קבוצה של n איברים. תת קבוצה של קבוצה זו בעלת r איברים (0 < r < n) (אין חשיבות לסדר האיברים בתת קבוצה) נקראת צירוף של r מתוך n איברים.
מספר הצירופים :

$$\boxed{C(n,r)=C_n^r=\binom{n}{r}=\frac{n!}{r!(n-r)!}}$$

תמורות עם חזרות:
נתונים : q1 עצמים זהים מסוג 1, q2 עצמים זהים מסוג 2, ...
qk עצמים זהים מסוג k, כך ש : n = q 1 + q 2 + q 3 + ... + q k,
אזי, מספר התמורות של העצמים האלו הוא : $\boxed{\frac{n!}{q1!\cdot q2!\cdot\ldots\cdot qk!}}$

חליפות עם חזרות:
נתונים n סוגי עצמים, ללא הגבלה בכמות העצמים מכל סוג שהוא. רוצים לבחור r מתוכם, כשיש חשיבות הבחירה. (שימו לב כי ייתכן r > n). אזי מספר האפשרויות הוא : $\boxed{n^r}$

צירופים עם חזרות:
נתונים n סוגי עצמים, ללא הגבלה בכמות העצמים מכל סוג שהוא. רוצים לבחור r עצמים מתוכם, כאשר אין חשיבות לסדר הבחירה. מספר האפשרויות הוא :

$$\boxed{\binom{n+r-1}{n-1}=\binom{n+r-1}{r}}$$

שימו לב, כי הבעיה שקולה לבעיית חלוקת r כדורים זהים ל- n תאים שונים.

דף עזר 2

תורת הסתברות

הגדרות:
ניסוי מקרי - הוא ניסוי שאין לדעת בוודאות מראש את תוצאתו אלא קיים מרחב של תוצאות אפשריות.

מרחב מדגם Ω - הקבוצה הכוללת – אוסף כל התוצאות האפשריות של הניסוי המקרי.

מאורע פשוט ω_i - תוצאה בודדת של מרחב המדגם (איבר ב- Ω).

מאורע A (B, C, ...) - אוסף של מאורעות פשוטים (קבוצה חלקית ל- Ω).

קבוצה חלקית (מוכלת) - $A\subset B$ - A תת-קבוצה של B (A מוכלת ב-B) אם כל איבר של A הוא איבר של B.
אנו אומרים שמאורע A קורה אם תוצאת הניסוי (שהיא אחת מנקודות המרחב) שייכת ל-A .

מאורע משלים למאורע A - A^c - כל האיברים ב- Ω שאינם שייכים ל-A .

מאורע ריק $\varnothing$ (קבוצה ריקה) - אינו כולל אף איבר.

פעולות בין מאורעות:

איחוד - $A\cup B$ - כל האיברים שהם או ב-A או ב-B או בשתי הקבוצות.

חיתוך - $A\cap B$ - כל האיברים שהם גם ב-A וגם ב-B .

* אם החיתוך בין הקבוצות הוא ריק, כלומר אין להן איברים משותפים, אזי הקבוצות נקראות זרות.

מאורע A והמאורע המשלים לו A^c הם מאורעות זרים : $A\cap A^c=\varnothing$

האיחוד בין מאורעות A ו- A^c הוא מרחב המדגם : $A\cup A^c=\Omega$

הפרש - $A\setminus B$ - כל האיברים שהם ב-A אך לא ב-B : $A\setminus B=A\cap B^c$

כללי פעולות בין הקבוצות

חוק החילוף: $\boxed{A\cup B=B\cup A}$ $\boxed{A\cap B=B\cap A}$

חוק הפילוג: $A\cap\left(B\cup C\right)=\left(A\cap B\right)\cup\left(A\cap C\right)$

$A\cup\left(B\cap C\right)=\left(A\cup B\right)\cap\left(A\cup C\right)$

חוקי דה-מורגן: $\boxed{\left(A\cap B\right)^C=A^C\cup B^C}$ $\boxed{\left(A\cup B\right)^c=A^c\cap B^C}$

הסתברות מותנית

יהיה A מאורע, כך ש- $P(A)>0$. ההסתברות המותנית של מאורע B בהינתן A נתונה ע"י : $\boxed{P(B/A)=\frac{P(A\cap B)}{P(A)}}$. באופן דומה : $\boxed{P(\overline{B}/A)=\frac{P(\overline{B}\cap A)}{P(A)}}$

נוסחת הכפל

יהיו A ו- B מאורעות כך ש- $P(A)>0$.

ידוע כי $P(B/A)=\frac{P(B\cap A)}{P(A)}$ $\Rightarrow$ $P(B\cap A)=P(B/A)P(A)$.

בדומה עבור 3 מאורעות :

$$P(A\cap B\cap C)=P(C/B\cap A)P(B\cap A)=P(C/B\cap A)P(B/A)P(A)$$

* הערה : כאשר A ו- B מאורעות זרים $\boxed{P(A\cap B)=0}$

נוסחת ההסתברות השלמה

יהיו $A_1,...,A_n$ מאורעות זרים במרחב הסתברות ובעלי הסתברות חיובית כ ך ש- $\boxed{\bigcup_{i=1}^{n}A_i=\Omega}$, ויהיה B מאורע כלשהו במרחב המדגם, אזי :

$$P(B)=\sum_{i=1}^{n}P(B/A_i)P(A_i)=\sum_{i=1}^{n}P(B\cap A_i)\ .$$

באותו אופן :

$$P(\overline{B})=\sum_{i=1}^{n}P(\overline{B}/A_i)P(A_i)=\sum_{i=1}^{n}P(\overline{B}\cap A_i)$$

עבור 2 מאורעות : $P(\overline{B})=P(\overline{B}/A)P(A)+P(\overline{B}/\overline{A})P(\overline{A})$.

נוסחת בייס

באותם התנאים של נוסחת ההסתברות השלמה, יהיה B מאורע כלשהו בעל הסתברות חיובית, אזי :

$$P(A_j/B)=\frac{P(A_j\cap B)}{P(B)}=\frac{P(B/A_j)P(A_j)}{\sum_{i=1}^{n}P(B/A_i)P(A_i)}\ .$$

עבור 2 מאורעות : $P(A/B)=\frac{P(B/A)P(A)}{P(B)}$.

באופן דומה : $P(A/\overline{B})=\frac{P(A\cap\overline{B})}{P(\overline{B})}$

דיאגרמת עץ
* כל פיצול מתייחס לשלב ניסוי מסוים.
* סכום ההסתברויות בכל פיצול שווה ל 1.
* ההסתברות המתאימה לכל ענף היא ההסתברות המותנית בהינתן המאורעות המתאימים לענפים הקודמים

פונקצית הסתברות

יהי מרחב המדגם Ω - אוסף המאורעות הפשוטים : $\Omega=\left\{\omega_1,\omega_2,...\right\}$

לכל מאורע פשוט נצמיד מספר $P\left(\omega\right)$ הנקרא ההסתברות של מאורע ω :

$0\le P\left(\omega\right)\le 1$ לכל ω .

$0\le P\left(A\right)\le 1$, כאשר A הוא מאורע כלשהו במרחב המדגם (אוסף המאורעות הפשוטים). $P\left(\Omega\right)=1$, $P\left(\varnothing\right)=0$.

מרחב מדגם סופי ושווה-הסתברות (מרחב סימטרי בדיד)
מרחב מדגם סופי נקרא שווה-הסתברות כאשר לכל המאורעות הפשוטים במרחב זה יש אותה הסתברות להתרחש :

$$P\left(\omega_1\right)=P\left(\omega_2\right)=...P\left(\omega_n\right)=\frac{1}{n}=\frac{1}{\#\left(\Omega\right)}$$

במרחב סימטרי בדיד מתקיים : $P\left(A\right)=\frac{\#\left(A\right)}{\#\left(\Omega\right)}$

מרחב מדגם סימטרי רציף
מרחב מדגם סימטרי רציף – מרחב רציף שבו לכל הקטעים\השטחים השווים בגודלם הסתברות שווה להתקבל. במרחב סימטרי רציף מתקיים : $P(A)=\frac{S\left(A\right)}{S\left(\Omega\right)}$

חוקים בחישוב ההסתברות
הסתברות האיחוד של שני מאורעות זרים :

$$P(A\cup B)=P(A)+P(B)$$

מקרה פרטי : חישוב הסתברות האיחוד של מאורע A והמאורע המשלים לו A^C .

$$P(A)=1-P(A^C)$$

הסתברות האיחוד של שני מאורעות לא זרים :

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

הסתברות האיחוד של שלושה מאורעות לא זרים :

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(B\cap C)-P(A\cap C)+P(A\cap B\cap C)$$

דף עזר 3

אי תלות של מאורעות
יהיו A, B מאורעות במרחב ההסתברות. A ו-B יקראו "מאורעות-בלתי-תלויים" אם קיים השוויון : $\boxed{P(A\cap B)=P(A)\cdot P(B)}$. מושג אי תלות בין שני מאורעות הוא סימטרי – A בלתי תלוי ב-B וגם B בלתי תלוי ב-A.

אי תלות עבור 3 מאורעות:
A, B, C מאורעות בלתי תלויים אם מתקיימים התנאים הבאים :

$\boxed{P(A\cap B)=P(A)\cdot P(B)}$ $\boxed{P(A\cap C)=P(A)\cdot P(C)}$

$\boxed{P(B\cap C)=P(B)\cdot P(C)}$ $\boxed{P(A\cap B\cap C)=P(A)\cdot P(B)\cdot P(C)}$

הבדל בין זרים לבלתי תלויים:
A ו-B זרים : $\boxed{P(B\cap A)=0}$ A ו-B ב"ת : $\boxed{P(A\cap B)=P(A)\cdot P(B)}$

טבלה מסכמת:

זרים	תלויים	בת"ל
$A\cap B=\varnothing$ $P(A\mid B)=0$ $P(A\cap B)=0$	$P(A\mid B)=\frac{P(A\cap B)}{P(B)}$ $P(A\cap B)=P(B)P(A\mid B)$	$P(A\mid B)=P(A)$ $P(A\cap B)=P(A)P(B)$

משתנים מקריים
משתנה מקרי – נתון מרחב ההסתברות. הפונקציה X = x(w) המתאימה לכל נקודה w במרחב המדגם Ω נקראת מ"מ (משתנה מקרי)
משתנה מקרי בדיד - מ"מ X יקרא **מ"מ בדיד** אם הוא מקבל סדרה סופית או בת מניה של ערכים.

פונקצית הסתברות עבור מ"מ בדיד:
הפונקציה $P_X(x)=P(X=x),\ x\in R$ נקראת פונקצית הסתברות עבור מ"מ X.

הערה : $\sum_x P_X(x)=1$

פונקצית התפלגות מצטברת עבור מ"מ בדיד:
הפונקציה $F_x(x)$, המוגדרת על ידי : $F_X(x)=\sum_{i:X_i\le x}P_X(x_i)=P(X\le x)$

(הסתברות שמ"מ X קטן או שווה למספר x), נקראת פונקצית התפלגות מצטברת של מ"מ X בנק' x.

תכונות פונקצית ההתפלגות:

1. $0\le F_X(x)\le 1$ לכל $x\in\Re$.
2. הפונקציה $F_X(x)$ מונוטונית לא יורדת ב- x, כלומר :
   אם $x_1<x_2$, אז $F_X(x_1)\le F_X(x_2)$.
3. $\lim_{x_0\to-\infty}F_X(x_0)=0$, $\lim_{x_0\to\infty}F_X(x_0)=1$.

הערה : אם נתונה $P_X(x)$ ניתן לחשב ממנה את $F_X(x)$.

אם נתונה $F_X(x)$: $P_X(x)=F_X(x)-F_X(x-1)$, כלומר, פונקצית ההסתברות שווה לפונקצית ההתפלגות פחות פונקצית ההתפלגות בנקודה שלפני.

תוחלת של מ"מ בדיד
יהי X מ"מ המקבל ערכים $x_1,x_2,\ldots x_n$ בהסתברויות $P_X(x_1),P_X(x_2),\ldots,P_X(x_n)$ בהתאמה, אזי התוחלת של X מוגדרת ע"י :

$E\left(X\right)=\sum_{x\in\Omega}x_iP_X\left(x\right)$, כאשר Ω היא מרחב המצבים.

תכונות התוחלת:

1. עבור קבוע a , $E\left(a\right)=a$ 2. $E\left(aX\right)=aE\left(X\right)$ 3. $E\left(X+a\right)=E\left(X\right)+a$ 4. $E\left(aX+b\right)=aE\left(X\right)+b$	5. $Eg(X)=\sum g(X)\cdot P(X=x)$ 6. אם $X_1,\ldots X_K$ סדרה של מ"מ כלשהם אזי : $E\left(X_1+X_2+\ldots X_K\right)=E\left(X_1\right)+\ldots+E\left(X_n\right)$

שונות
שונות הִנה מדד לפיזור. שונות של מ"מ מסומנת על ידי $Var(X)$ או על ידי $V(X)$.

הגדרה תיאורטית של שונות : $\mathrm{var}(X)=E\left[X-E(X)\right]^2$.

נוסחת העבודה : $Var(X)=E(X^2)-E^2(X)$

סטיית תקן : $\sigma(X)=\sqrt{Var(X)}$

תכונות של שונות:

1. $Var(X)\ge 0$ 2. $Var(a)=0$ 3. $Var(a\cdot X)=a^2\cdot Var(X)$ 4. $Var(a+X)=Var(X)$	5. אם X ו- Y מ"מ **בלתי תלויים** אזי $Var(X+Y)=Var(X)+Var(Y)$ 6. אם X ו- Y מ"מ **בלתי תלויים** אזי $Var(X-Y)=Var(X)+Var(Y)$

דף עזר 4

פונקצית התפלגות ברנולי עם פרמטר p
מבצעים ניסוי ברנולי עם הסתברות p להצלחה. X מקבל 1 אם הניסוי הצליח ו- 0 אחרת. אזי :

$$p_X(x)=\begin{cases}p & x=1\\ q & x=0\\ 0 & otherwise\end{cases}$$

ומסמנים : X~Ber(p)

תוחלת : $E(X)=p$ ושונות : $Var(X)=p\cdot q$

פונקצית התפלגות בינומית עם פרמטרים n,p
X מ"מ שסופר את מספר ההצלחות ב- n ניסויי ברנולי בלתי תלויים עם הסתברות p להצלחה, אזי :

$$p_X(x)=\begin{cases}\binom{n}{x}p^xq^{n-x} & x=0,1,...,n\\ 0 & otherwise\end{cases}$$

ומסמנים : $X\sim Bin(n,p)$

אם X~Bin(n,p) (בינומי) אז התוחלת שלו היא : $E(X)=n\cdot p$

השונות שלו היא : $Var(X)=n\cdot p\cdot q$

<u>פונקצית התפלגות גיאומטרית עם פרמטר p</u>

X מ״מ שסופר את מספר הניסויים עד להצלחה הראשונה בסדרת ניסויי ברנולי בלתי תלויים עם הסתברות p להצלחה, אזי:

$$p_X(x)=\begin{cases} q^{x-1}p & x=1,2,3,\dots \\ 0 & otherwise \end{cases}$$ ומסמנים: X~Geo(p). אם X~Geo(p)

(גיאומטרי) אז: $E(X)=\frac{1}{p}$, $Var(X)=\frac{q}{p^2}$. $\boxed{P(x=0)=0}$

$\boxed{P(x>n)=q^n}$ $\boxed{P(x\le n)=1-q^n}$, תכונת חוסר הזכרון:

$\boxed{P(x>n+k/x>k)=P(x>n)=q^n}$ $\boxed{P(x=n+k/x>k)=P(x=n)=pq^{n-1}}$

<u>פונקצית התפלגות פואסונית עם פרמטר λ</u>

X - מ״מ פואסוני מונה אירועים לאורך זמן:

$$p_X(x)=\begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & x=0,1,2,\dots \\ 0 & otherwise \end{cases}$$ ומסמנים: $X\sim Pois(\lambda)$

אם $X\sim Pois(\lambda)$ (פואסוני) אז: $Var(X)=E(X)=\lambda$. <u>עבור t-יחידת זמן</u>:

$\boxed{P(X=x)=\frac{(\lambda t)^x\cdot e^{-\lambda t}}{x!}}$ $\boxed{E(x)=V(x)=\lambda t}$

<u>התפלגות בינומית שלילית</u>

מס׳ ניסויים עד הצלחה m. בשלב k אנו מקבלים הצלחה מס׳ m.

נתון מרחב המדגם בן-מנייה $\Omega=\{m,m+1,\dots\}$ ועליו מוגדרת מידת

ההסתברות: (עבור m מסוים)
$$P(k)=\begin{cases} \binom{k-1}{m-1}p^m q^{k-m}, & k\ge m \\ 0, & otherwise \end{cases}$$

כאשר: $q=1-p$, $0<p\le 1$. $\boxed{E(x)=\frac{m}{p}}$ $\boxed{V(x)=\frac{mq}{p^2}}$

$\boxed{X\sim NB(n,p)}$ p - הסתברות להצלחה, n – מספר הצלחות

<u>מודל בינומי שלילי</u>

מבצעים סידרה של ניסויי ברנולי ב״ת עד להצלחה ה- m-ית.
ההסתברות שיהיו בדיוק k ניסוים, ז״א ההסתברות לכך שההצלחה ה- m-ית תתקבל בדיוק בניסיון ה-k היא P(k) .

<u>התפלגות היפר גיאומטרית</u>

בכד נמצאים N כדורים, מתוכם D שחורים ו- N-D לבנים. מוציאים n כדורים באקראי וללא החזרה. ההסתברויות שבין הכדורים שהוצאו נמצאים k כדורים שחורים היא:

$$\frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}$$

פרופורציה מיוחדים: $\boxed{p=\frac{R}{N}}$ $\boxed{E(x)=np}$

$\boxed{V(x)=npq(1-\frac{n-1}{N-1})}$

<u>התפלגות מולטינומית</u>

הגדרה: מבצעים n ניסויים ב״ת . לכל ניסוי קיימות k תוצאות אפשריות כך

שההסתברות לתוצאה i היא p_i ומתקיים: $\sum_{i=1}^{k} p_i=1$

Xi – מציין את מספר התוצאות מסוג i שהתקבלו ב- n ניסויים X=(X1,X2,…,Xk)
אזי: X~Multi(n,p1,p2,…,pk)

$$P(X_1=n_1,X_2=n_2,\dots,X_k=n_k)=$$
$$\begin{cases} \frac{n!}{n_1!\cdot n_2!\cdot\dots\cdot n_k!}\cdot p_1^{n1}\cdot p_2^{n2}\cdot\dots\cdot p_k^{nk} & ,if \ \sum_{i=1}^{k} n_i=n \\ 0 & ,else \end{cases}$$

<u>התפלגות אחידה בדידה</u>:

סימון: X~Uni[a,b] (a<b)

$$p_X(x)=\begin{cases} \frac{1}{b-a+1} & x=a,..,b \\ 0 & otherwise \end{cases}$$

$$E(X)=\frac{a+b}{2} \qquad Var(X)=\frac{(b-a+1)^2-1}{12}$$

<u>דף עזר 5</u>

<u>משתנה מקרי רציף:</u>
<u>פונקצית צפיפות</u>

משתנה מקרי רציף מאופיין ע״י פונקצית צפיפות $f_X(x)$ המקיימת:

1. $f_X(x)\ge 0$ 2. $\int_{-\infty}^{\infty} f_X(t)dt=1$

<u>עבור משתנה מקרי רציף</u>

$P(X=x)=0$ (כלומר, ההסתברות לקבל ערך נקודתי מסוים של מ״מ רציף שווה ל- 0)

$$P(a\le X\le b)=\int_a^b f_X(x)dx$$

<u>פונקצית התפלגות מצטברת</u>

יהי X מ״מ. פונק׳ התפלגות מצטברת של X היא פונקציה $F_X:R\rightarrow[0,1]$ המוגדרת

ע״י: $F_X(x)=P(X\le x)$

עבור X רציף: $F_X(x)=P(X<x)=\int_{-\infty}^{x} f_X(t)dt$

חישוב הסתברות בעזרת פונקציית התפלגות מצטברת:
$$P(a<X<b)=F_X(b)-F_X(a)$$

<u>תוחלת של מ״מ רציף:</u>

$$E(X)=\int_{-\infty}^{\infty} x\cdot f_X(x)dx$$

תכונות התוחלת של מ״מ בדיד נכונות גם עבור מ״מ רציף פרט ל:

$$E(g(X))=\int_{-\infty}^{x} g(x)\cdot f_X(x)dx$$

<u>שונות של מ״מ רציף:</u> $Var(X)=E(X^2)-E^2(X)$

<u>משתנים מקריים רציפים מיוחדים</u>:

<u>התפלגות אחידה רציפה</u>:
סימון: X~U(a,b) (a<b)

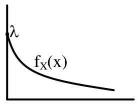
$$f_X(x)=\begin{cases} \frac{1}{b-a} & a<x<b \\ 0 & otherwise \end{cases}$$

$$E(X)=\frac{a+b}{2} \qquad Var(X)=\frac{(b-a)^2}{12}$$

פונקציית התפלגות מצטברת: $\boxed{F(x)=\int_a^x f(x)dx=\int_a^x \frac{1}{b-a}dx=\frac{x-a}{b-a}}$

<u>התפלגות מעריכית (אקספוננציאלית)</u>:

סימון: X~exp(λ) (λ נקרא ״הקצב״ של התפלגות)



$$f_X(x)=\begin{cases} \lambda e^{-\lambda x} & x\ge 0 \\ 0 & otherwise \end{cases}$$

$$E(X)=\frac{1}{\lambda} \qquad Var(X)=\frac{1}{\lambda^2}$$

$$P(X>a)=\int_a^{\infty} f_X(x)dx=\int_a^{\infty}\lambda e^{-\lambda x}dx=e^{-\lambda a}$$

$$F_x(X)=P(X<a)=1-e^{-\lambda a}$$

* בעל תכונה של חוסר זכרון.

$\boxed{P(a\le x\le b)=F(b)-F(a)=e^{-\lambda a}-e^{-\lambda b}}$

<u>התפלגות נורמלית</u>:

$X\sim N(\mu,\sigma^2)$

$\boxed{f_X(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty<x<\infty}$ $\boxed{\begin{matrix} E(X)=\mu \\ Var(X)=\sigma^2 \end{matrix}}$

מקרה פרטי של התפלגות נורמלית – מ״מ נורמלי סטנדרטי:

$Z\sim N(0,1)$ - מ״מ מתפלג נורמלית עם תוחלת $\mu=0$ ושונות $\sigma^2=1$

<u>חישוב הסתברות בהתפלגות נורמלית (תקנון משתנה נורמלי – מעבר להתפלגות נורמלית סטנדרטית):</u>

אם $X\sim N(\mu,\sigma^2)$, אזי $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$

$$F_x(X)=P(X<a)=P\left(Z<\frac{a-\mu}{\sigma}\right)=\varphi\left(\frac{a-\mu}{\sigma}\right)$$

פונקציית צפיפות: $\boxed{f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}}$

<u>חישוב הסתברות בהתפלגות נורמלית סטנדרטית</u>:

$P(Z<z)=\varphi(z)$, מאחר שההסתברות סימטרית סביב ה-0:

$$P(Z>z)=1-P(Z<z)=1-\varphi(z)=\varphi(-z)=P(Z<-z)$$

$\boxed{P(z_1\le Z\le z_2)=\phi(z_2)-\phi(z_1)}$ $\boxed{\phi(Z\ge -z)=\phi(z)}$

<u>סימון:</u>

z_t - ערך של Z שמשמאלו שטח (הסתברות) של t.

<u>הערה</u>: $z_t=-z_{1-t}$ מתקיים בגלל הסימטריה של התפלגות נורמלית סטנדרטית

* שיברון p של x:
$$x_p=\begin{cases} \mu+z_p\cdot\sigma & :P>0.5 \\ \mu-z_{1-p}\cdot\sigma & :P<0.5 \end{cases}$$

<u>חישוב ערך X</u> כאשר t ידוע ו- $X\sim N(\mu,\sigma^2)$:

$$X\sim N(\mu,\sigma^2),\quad Z=\frac{X-\mu}{\sigma}\sim N(0,1),$$
$$X_t=\mu+z_t*\sigma$$

<u>דף עזר 6</u>

<u>המשך התפלגות נורמאלית</u>:

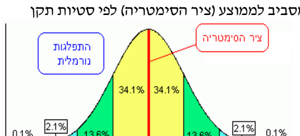
אחוזי ההתפלגות הנורמאלית מסביב לממוצע (ציר הסימטריה) לפי סטיות תקן

$\left[\mu\pm\sigma\right]=68.2\%$

$\left[\mu\pm 2\sigma\right]=95.4\%$

$\left[\mu\pm 3\sigma\right]=99.6\%$

$\left[\mu\pm 3.5\sigma\right]=99.8\%$



* אם כל x מתפלג נורמלית, אז משפט הגבול המרכזי מתקיים עבור כל n.

<u>משפט הגבול המרכזי</u>

• יהיו: X1,X2,…,Xn מ״מ בלתי תלויים שווי התפלגות עם תוחלת μ ושונות σ^2.

נגדיר $\overline{X}n=\frac{1}{n}\sum_{i=1}^{n}X_i$, אזי: $\frac{\overline{X}n-\mu}{\sigma/\sqrt{n}} \underset{n\to\infty}{\Rightarrow} N(0,1)$, זאת כיוון ש- $\overline{X_n}$

בקירוב מתפלג כמו משתנה מקרי מהתפלגות

כלומר $\boxed{N(\mu,\frac{\sigma^2}{n})}$

$\overline{X_n}$ - ממוצע
S_n - סכום

(הערה: בפועל $P(\frac{\overline{X}n-\mu}{\sigma/\sqrt{n}}\le a)\xrightarrow[n\to\infty]{}\Phi(a)$

נסתפק ב-n>=30)

(כאשר $\Phi(a)$ פונקצית התפלגות מצטברת של מ״מ נורמאלי-סטנדרטי וערכה ייקבע לפי הטבלה)

• באופן שקול אם נגדיר $S_n=\sum_{i=1}^{n}X_i$, אזי: $\frac{S_n-n\mu}{\sqrt{n}\sigma} \underset{n\to\infty}{\Rightarrow} N(0,1)$

S_n מתפלג בקירוב כמו משתנה מקרי נורמאלי עם הפרמטרים: $\boxed{N(n\mu,n\sigma^2)}$

<u>דף עזר 7</u>

<u>קירוב נורמלי להתפלגות בינומית</u>

אם $X\sim Bin(n,p)$, אזי לפי משפט הגבול המרכזי, עבור n מספיק גדול (np>5):

$$X\sim N(np,npq)\Rightarrow Z=\frac{X-np}{\sqrt{npq}}$$

$\hat{p}=\frac{x}{n}$ - פרופורציה מדגמית. $\hat{p}\sim N(p,\frac{p(1-p)}{n})$

<u>משתנה מקרי דו מימדי</u>

יהי נתון מרחב מדגם Ω . הזוג (X,Y) יקרא משתנה מקרי דו מימדי על מרחב המדגם Ω .

יהי פונקצית ההסתברות המשותפת לכל x, y $P(X=x,Y=y)$ של מ״מ דו מימדי (X,Y).

<u>הערה</u>: נהוג להציג את ההתפלגות המשותפת של מ״מ דו מימדי (X,Y) בעזרת טבלה דו-מימדית כדלקמן:

	Y1	Y2	…	Yj	…	P(X)
X1						P(X1)
X2						P(X2)
…						
Xi				P(X=Xi,Y=Yj)		P(Xi)
…						
P(Y)	P(Y1)	P(Y2)		P(Yj)		1

* אם יש 0 אחד לפחות בתוך הטבלה, אז הם תלויים.

$P_X(x)$ – הסתברות (שולית) של X, $P_Y(y)$ - הסתברות (שולית) של Y

$$P_Y(y)=\sum_x P(x,y) \qquad P_X(x)=\sum_y P(x,y)$$

<u>פונקצית ההסתברות המותנית של X בהינתן Y=y:</u>

$$P(X=x/Y=y)=\frac{P(X=x,Y=y)}{P(Y=y)}$$

<u>פונקצית ההסתברות המותנית של Y בהינתן X=x:</u>

$$P(Y=y/X=x)=\frac{P(X=x,Y=y)}{P(X=x)}$$

<u>חוק הכפל</u>

$$P(X=x,Y=y)=P_X(x)*P(Y=y/X=x)=P_Y(y)*P(X=x/Y=y)$$

<u>שונות משותפת של מ״מ דו-מימדי</u>

יהי X ו-Y מ״מ המוגדרים על אותו מרחב מדגם Ω .
שונות משותפת של X ו-Y מוגדרת על ידי:

$$Cov(X,Y)=E((X-E(X)*(Y-E(Y)))$$

או $\boxed{Cov(X,Y)=E(X*Y)-E(X)*E(Y)}$

במדגם: $\boxed{\widehat{Cov}(x,y)=\frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{n-1}=\frac{\sum x_i y_i-n\overline{x}\overline{y}}{n-1}}$

<u>תכונות שונות משותפת</u>:

1. שונות המשותפת היא סימטרית: $Cov(X,Y)=Cov(Y,X)$
2. $Cov(X,X)=Var(X)$
3. אם a,b,c,d מספרים קבועים אז: $Cov(aX+b,cY+d)=acCov(X,Y)$
4. אם X,Y,Z מ״מ המוגדרים על אותו מרחב Ω אז:
$$Cov(X,Y+Z)=Cov(X,Y)+Cov(X,Z)$$
$$Cov(X+Y,Z)=Cov(X,Z)+Cov(Y,Z)$$

<u>תוחלת ושונות של סכום והפרש של מ״מ</u>:

יהי Y,X מ״מ מוגדרים על אותו מרחב Ω אז:

$$E(X\pm Y)=E(X)\pm E(Y)$$
$$Var(X\pm Y)=Var(X)+Var(Y)\pm 2Cov(X,Y)$$

<u>מקדם המתאם</u>:

כמו שראינו ברגרסיה ליניארית: $\rho(X,Y)=\frac{Cov(X,Y)}{\sigma(x)\sigma(y)}$ באוכלוסיה

במדגם: $\boxed{r=\frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum(x_i-\overline{x})^2}\cdot\sqrt{\sum(y_i-\overline{y})^2}}}$

<u>משתנים מקריים בדידים X ו-Y הם ב״ת אם מתקיים לכל x,y:</u>

$$P(X=x,Y=y)=P(X=x)*P(Y=y)$$

<u>משתנים מקריים בלתי מתואמים</u>:

<u>הגדרה</u>: יהי Y,X מ״מ מוגדרים על אותו מרחב Ω .
Y,X <u>בלתי מתואמים</u> אם לא קיים בניהם קשר ליניארי.

כלומר: $\boxed{\begin{matrix} Cov(X,Y)=0 \quad or \\ E(X*Y)=E(X)*E(Y) \\ Var(X+Y)=Var(X-Y)=VaR(X)+Var(Y) \end{matrix}}$ עבור ב״ת ובלתי מתואמים.

* אם x,y ב״ת, הם בהכרח בלתי מתואמים.

עבור x ו-y תלויים: $\boxed{\begin{matrix} V(x+y)=V(x)+V(y)+2Cov(x,y) \\ V(x-y)=V(x)+V(y)-2Cov(x,y) \end{matrix}}$

<u>הערות</u>:

1. $\left|\rho(X,Y)\right|=1$ אם X ו-Y <u>תליים ליניארית</u>
2. השונות המשותפת קובעת את סימון מקדם המתאם:
$$\rho(X,Y)>0 \Leftrightarrow Cov(X,Y)>0$$
$$\rho(X,Y)<0 \Leftrightarrow Cov(X,Y)<0$$

<u>דף עזר 8</u>

<u>אמידה נקודתית</u>

<u>אוכלוסייה</u> – אוסף פריטים עליהם מעוניינים במידע מסוים.
<u>מדגם</u> – אוסף חלקי של פריטים מהאוכלוסייה.
<u>פרמטר</u> – מספר קבוע המאפיין את האוכלוסייה (θ) .
<u>סטטיסטי</u> – ערך הניתן לחישוב מנתוני המדגם.
<u>אמד</u> – סטטיסטי המשמש לאמידת פרמטר $(\hat{\theta})$.
<u>אומדן</u> – הערך המספרי של האמד.

אמד $\hat{\theta}$ ייקרא **<u>אמד חסר הטיה</u>** לפרמטר θ , אם קיים: $E(\hat{\theta})=\theta$.

אחרת, הוא **<u>אמד מוטה</u>** והטייתו ניתנת ע״י: $Bias(\hat{\theta})=E(\hat{\theta})-\theta$.

<u>נעדיף אמד חסר הטיה על אמד מוטה.</u>

שגיאה ריבועית ממוצעת

(MSE): $MSE(\hat{\theta})=E(\hat{\theta}-\theta)^2=Var(\hat{\theta})+[Bias(\hat{\theta})]^2$.

נעדיף את האמד בעל MSE מינימלי.

הערה : אם האמד מוטה ובעל MSE נמוך יותר הוא עדיף על אמד שאינו מוטה ובעל MSE גבוה יותר.

הערה: אם $\hat{\theta}$ הוא אמד חסר הטיה לפרמטר θ , מתקיים:

• $MSE(\hat{\theta}) = Var(\hat{\theta}) \Leftarrow Bias(\hat{\theta}) = 0$

. $MSE(\hat{\theta}) = E[(\hat{\theta}-\theta)^2] = Var(\hat{\theta}) + B^2$

סטיית התקן של הממוצע כאשר σ_X ידועה: $SE_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$ $\boxed{SE_{\bar{x}} = \sqrt{MSE_{\bar{x}}}}$

סטיית התקן של הממוצע כאשר σ_X לא ידועה: $SE_{\bar{X}} = \frac{S}{\sqrt{n}}$

$$S^2 = \frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{X}^2}{n-1}$$

אמדים:

$\hat{\mu} = \bar{x}$

$\hat{\sigma}^2 = S^2$

$\hat{P} = \frac{x}{n} \to X \sim Bin(n,p)$

$\hat{\lambda} = \bar{x} \to X \sim Pois(\lambda)$

$\hat{P} = \frac{1}{\bar{x}} \to X \sim Geo(p)$

* עקיבות: אמד $\hat{\theta}$ עקיב ביחס ל θ , אם בהסתברות גבוהה $\hat{\theta}_n \to \theta$

* אם $MSE_{\hat{\theta}}(\theta) \to 0$, אזי $\hat{\theta}$ עקיב ביחס ל θ .

* אמד טוב הוא אמד חסר הטייה, עקיב אם MSE נמוך.

התפלגויות דגימה

1. התפלגות T : התפלגות סימטרית בעלת k דרגות חופש .

עבור מ"מ המתפלג $T \sim t(6)$ מצא את השיברון ה- 0.1 וה - 0.995

2. התפלגות חי בריבוע: התפלגות לא סימטרית בעלת n דרגות חופש. מ"מ בעל התפלגות הינו חיובי. עבור משתנה מיקרי $W \sim \chi^2(12)$ מצא את האחוזון ה 0.9 וה- 0.1.

3. התפלגות F : איננה סימטרית ובעלת k דרגות חופש במונה ו-m דרגות חופש במכנה.

תכונת התפלגות F :

$F(k,m) = \frac{1}{F(m,k)}$ ולכן לכל $0 \le p \le 1$ מתקיים $\boxed{x_{1-p}(m,k) = \frac{1}{x_p(k,m)}}$.

שאלה: מצא את השיברונים ה-0.05 וה-0.95 עבור F(15,10) ועבור F(10,15).

תשובה: $\boxed{F_{0.95}^{(15,10)} = 2.85}$ $\boxed{F_{0.05}^{(15,10)} = \frac{1}{F_{0.95}^{(10,15)}} = \frac{1}{2.54}}$

דף עזר 9 – רווחי סמך

רמת בטחון = רמת סמך = $1-\alpha$. רמת מובהקות = α

רווח סמך: $\boxed{parameter \in [amad \pm shivron_{1-\frac{\alpha}{2}} \cdot SE_{amad}]}$

(1) רווח בר-סמך לתוחלת μ של אוכלוסייה נורמלית כאשר שונות σ^2 ידועה:

$\bar{X}$ אמד נקודתי לתוחלת μ .

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

ולכן, $\boxed{\mu \in \left[\bar{X} \pm Z_{1-\alpha/2}\, \sigma/\sqrt{n}\right]}$

זהו רווח בר-סמך לתוחלת μ ברמת ביטחון $1-\alpha$.

d – חצי אורך של רווח הסמך : $\boxed{d = Z_{1-\alpha/2}\, \sigma/\sqrt{n}}$ (2d – אורך רווח סמך)

על מנת למצוא את גודל המדגם שיביא לסטייה של d מהתוחלת: $n \ge \left(\frac{Z_{1-\alpha/2} \cdot \sigma}{d}\right)^2$

(2) רווח בר-סמך לתוחלת μ של אוכלוסייה נורמלית כאשר שונות σ^2 אינה ידועה: נאמוד את השונות בעזרת סטטיסטי S^2 המחושב מהמדגם:

$S^2 = \frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{X}^2}{n-1}$. מתקיים: $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$,

כאשר $(n-1)$ - מספר דרגות חופש – פרמטר בהתפלגות $t_{(n-1)}$; n הוא גודל המדגם שעליו מבוסס הממוצע. ולכן, $\mu \in \left[\bar{X} \pm t_{n-1}^{1-\alpha/2}\, S/\sqrt{n}\right]$

זהו רווח בר-סמך לתוחלת μ ברמת ביטחון $1-\alpha$.

d – חצי אורך של רווח הסמך : $d = t_{(n-1)}^{1-\frac{\alpha}{2}}\, S/\sqrt{n}$ (2d – אורך רווח סמך)

על מנת למצוא את גודל המדגם שיביא לסטייה של d מהתוחלת במקרה זה:

$n \to \left(\frac{Z_{1-\alpha/2} \cdot S}{d}\right)^2$ הערה: יש לשים לב ש- n שואף לערך זה מכוון שאנו משתמשים בהתפלגות Z ולא ב-t.

(3) רווח בר-סמך לשונות σ^2 של אוכלוסייה נורמלית כאשר התוחלת μ ידועה:

$\frac{\sum(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$, מכאן הרווח סמך עבור σ^2 ברמת סמך $1-\alpha$:

$$\left[\frac{\sum_{i}^{n}\left(x_i-\mu\right)^2}{\chi^{2(n)}_{1-\alpha/2}}, \frac{\sum_{i}^{n}\left(x_i-\mu\right)^2}{\chi^{2(n)}_{\alpha/2}}\right]$$

(4) רווח בר-סמך לשונות σ^2 של אוכלוסייה נורמלית כאשר התוחלת μ אינה ידועה: S^2 אמד נקודתי לשונות. $\frac{S^2(n-1)}{\sigma^2} = \frac{\sum(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$

מכאן הרווח סמך עבור σ^2 ברמת סמך $1-\alpha$:

$$\left[\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\chi^{2(n-1)}_{1-\alpha/2}}, \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\chi^{2(n-1)}_{\alpha/2}}\right] = \left[\frac{S^2\cdot(n-1)}{\chi^{2(n-1)}_{1-\alpha/2}}, \frac{S^2\cdot(n-1)}{\chi^{2(n-1)}_{\alpha/2}}\right]$$

הרווח סמך עבור σ (סטיית תקן) ברמת סמך $1-\alpha$:

$$\left[S\cdot\sqrt{\frac{(n-1)}{\chi^{2(n-1)}_{1-\alpha/2}}},\ S\cdot\sqrt{\frac{(n-1)}{\chi^{2(n-1)}_{\alpha/2}}}\right]$$

(5) רווח סמך ל – p - פרופורציה באוכלוסייה:

$\hat{p} = \frac{x}{n}$ אמד נקודתי ל-p הפרופורציה במדגם. x - הנו מספר בעלי התכונה במדגם בגודל n. עבור np $np \ge 5$ $nq \ge 5$ מתקיים בקירוב

$$\hat{p} \sim N\left(p, \frac{p\cdot(1-p)}{n}\right) \Rightarrow \frac{\hat{p}-p}{\sqrt{\frac{p\cdot(1-p)}{n}}} \sim N(0,1)$$

ולכן, $-Z_{1-\alpha/2} \le \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} \le Z_{1-\alpha/2}$ $\boxed{p \in [\hat{p} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}}]}$

כעת ניתן לפתור בכמה דרכים:

א. **הצבת אמד במכנה**: בקירוב עבור n די גדול מתקבל $\frac{\hat{p}\hat{q}}{n} \approx \frac{pq}{n}$. לכן נציב במקום p את האמד $\hat{p}$, ובמקום q $1-\hat{p}$, ונקבל את רווח הסמך:

$$p \in \left[\hat{p} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}\right]$$

ב. **שיטת המשוואה הריבועית:**

יש לפתור את המשוואה הריבועית הבאה:

$$p^2(1+{Z_{1-\alpha/2}}^2 * \frac{1}{n}) + p(-{Z_{1-\alpha/2}}^2 * \frac{1}{n} - 2\hat{p}) + \hat{p}^2 \le 0$$

כעת נותר להציב את $\hat{p}$ α n ולמצוא את שורשי המשוואה $p_1,\ p_2$.

ונקבל את רווח הסמך: $p \in \left[p_1, p_2\right]$

ג. רווח סמך שמרני:

נציב $p = q = \frac{1}{2}$ ונקבל את רווח הסמך: $p \in \left[\hat{p} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\frac{1}{2}*\frac{1}{2}}{n}}\right]$

על מנת למצוא את גודל המדגם שיביא לסטייה של d מ-p:

$n \ge \left(\frac{Z_{1-\alpha/2}}{2d}\right)^2$ (d – חצי אורך רווח סמך)

דף עזר 10

אמידה מרווחית (המשך)

(6) רווח סמך להפרש תוחלות במדגמים נורמליים בלתי-תלויים:

$$\bar{X} \sim N\left(\mu_x, \frac{\sigma_x^2}{n_x}\right), \quad \bar{Y} \sim N\left(\mu_y, \frac{\sigma_y^2}{n_y}\right)$$

א. כאשר השונויות של שתי האוכלוסיות ידועות:

$$\bar{X}-\bar{Y} \sim N\left(\mu_x-\mu_y, \frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}\right) \Rightarrow \frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_x-\mu_y\right)}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} \sim N(0,1).$$

$\bar{X}-\bar{Y}$ - אמד נקודתי להפרש תוחלות $(\mu_x-\mu_y)$

ולכן, $\mu_x-\mu_y \in \left[\bar{X}-\bar{Y} \pm Z_{1-\frac{\alpha}{2}}\sqrt{\left(\left(\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}\right)\right)}\right]$

ב. כאשר שונויות לא ידועות ומניחים ששוות :

את השונות נאמוד על-ידי:

$$S_p^2 = \frac{\left(n_x-1\right)S_x^2+\left(n_y-1\right)S_y^2}{n_x+n_y-2}, = \frac{\sum(x_i-\bar{x})^2+\sum(y_i-\bar{y})^2}{n_x+n_y-2}$$

כאשר: $S_x^2 = \frac{\sum_{i=1}^{n_x}\left(x_i-\bar{X}\right)^2}{n_x-1}$, $S_y^2 = \frac{\sum_{i=1}^{n_y}\left(y_i-\bar{Y}\right)^2}{n_y-1}$ n_x+n_y-2 - מס' דרגות חופש

מתקיים:

$$\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_x-\mu_y\right)}{\sqrt{S_p^2\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}} \sim t_{(n_x+n_y-2)}.$$

ולכן, $\mu_x-\mu_y \in \left[\bar{X}-\bar{Y} \pm t^{1-\alpha/2}_{(n_x+n_y-2)}\sqrt{S_p^2\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}\right]$

זהו רווח בר-סמך להפרש תוחלות ברמת ביטחון $1-\alpha$.

ג. שונויות לא ידועות ולא מניחים שהן שוות:

$\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_x-\mu_y\right)}{\sqrt{\left(\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}\right)}} \sim t(f)$, כאשר: $f = \frac{\left[\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right]^2}{\frac{S_1^4}{n_1^2(n_1-1)}+\frac{S_2^4}{n_2^2(n_2-1)}}$

f - מס' דרגות חופש.

לכן ר"ס להפרש תוחלות ברמת סמך/ביטחון $1-\alpha$:

$$\mu_x-\mu_y \in \left[\bar{X}-\bar{Y} \pm t_{(f)}^{1-\alpha/2}\sqrt{\left(\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}\right)}\right]$$

ד. כאשר המדגמים מזווגים: (כלומר תלויים)

$d_i = x_i - y_i$

נגדיר: $\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$, רווח הסמך המתקבל: $\mu_d \in \left[\bar{d} \pm t_{1-\frac{\alpha}{2}}^{(n-1)}\frac{S_d}{\sqrt{n}}\right]$

$\mu_d = \mu_x - \mu_y$

כאשר: $s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i-\bar{d})^2}{n-1}} = \frac{\sum d_i^2 - n\bar{d}^2}{n-1}$

אם יודעים את השונות σ_d^2 : $\mu_d \in [\bar{d} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_d}{\sqrt{n}}]$

(7) רווח סמך להפרש פרופורציות:

באותו אופן כמו בהפרש תוחלות:

$$p_1-p_2 \in [\hat{p}_1-\hat{p}_2 \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}]$$

(8) רווח סמך ליחס שונויות :

$$\frac{S_X^2}{S_Y^2}\frac{1}{F_{1-\alpha/2}^{(n_x-1,n_Y-1)}} \le \frac{\sigma_X^2}{\sigma_Y^2} \le \frac{S_X^2}{S_Y^2}F_{1-\alpha/2}^{(n_Y-1,n_X-1)}$$

דף עזר 11

בדיקת השערות:

המטרה: רוצים להחליט על נכונותה או אי – נכונותה של השערה מסוימת לגבי פרמטרים של התפלגויות.

ההחלטה אם לדחות או לא לדחות את ההשערה נעשית על סמך המדגם , ובעזרת כלל הכרעה הקובע לכל תוצאת מדגם אפשרית לדחות או לא לדחות את ההשערה.

השערות סטטיסטיות:

H_0 : השערת האפס – זוהי ההשערה שעומדת לבדיקה סטטיסטית. עורכים את המחקר כדי לבדוק האם ניתן לסטות מהנחה זו (בד"כ מה שידוע ומקובל על הפרמטר).

H_1 : השערה אלטרנטיבית – הטענה החדשה שאותה רוצים לאמת במחקר – השערת החוקר.

סטטיסטי מבחן : ערך המחושב מהמדגם שבעזרתו מחליטים אם לדחות או לא לדחות את השערת האפס.

סימונים:

α - הסתברות לטעות מסוג ראשון או רמת מובהקות

$\alpha = P(\text{reject the null hypothesis} \mid \text{the null hypothesis is correct})$

β - הסתברות לטעות מסוג שני

$\beta = P(\text{don't reject the null hypothesis} \mid \text{the alternative hypothesis is correct})$

$\pi = 1-\beta$ - עוצמת המבחן

* מקבלים H_0 = תוצאת המדגם לא מובהקת.
* דוחים H_0 = תוצאת המדגם מובהקת.

סוגי השערות:

השערה פשוטה – מגדירה באופן חד ערכי את ההתפלגות

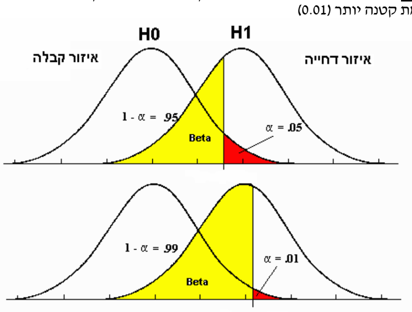
השערה מורכבת – כוללת הרבה אפשרויות

השערה מורכבת חד-צדדית (דוגמה: $H_0: \mu = 800$, $H_1: \mu < 800$)

השערה מורכבת דו-צדדית (דוגמה: $H_0: \mu = 800$, $H_1: \mu \ne 800$)

- לכל סוג של השערה אלטרנטיבית נקבע **אזור הדחייה** שאם הממוצע המדגמי נופל בו, המסקנה תהיה **לדחות את H_0**. נהוג לסמן אזור דחייה ב- C.
- האזור המשלים נקרא אזור הקבלה. אם הממוצע המדגמי נופל בו המסקנה תהיה לא לדחות את H0.

דוגמא: עבור סט השערות אחד – הבדל בין טעות מסוג ראשון גדולה יותר (0.05) לעומת קטנה יותר (0.01)



- עבור השערה חד-צדדית ימנית אנו נדחה עבור ערכים גדולים של הממוצע.
- עבור השערה חד-צדדית שמאלית אנו נדחה עבור ערכים קטנים של הממוצע.
- עבור השערה דו-צדדית אנו נדחה עבור ערכים קטנים או גדולים של הממוצע.

	כלל הכרעה	סטטיסטי מבחן	P-value
σ^2 **ידועה**	$C_\alpha = \mu_0 + z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}} \quad :\mu_1 > \mu_0$ $C_\alpha = \mu_0 - z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}} \quad :\mu_1 < \mu_0$ $C_\alpha = \mu_0 \pm z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}} \quad :\mu_1 \neq \mu_0$ * משווים ל- $\bar{x}$	$z_{calc} = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ * משווים עם השברון המתאים	$P.V. = P(\bar{x} > \bar{x}_n) =$ $= P(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > \frac{\bar{x}_n-\mu_0}{\sigma/\sqrt{n}}) =$ $= P(z > z_{calc}) = 1-\phi(z_{calc})$ $P.V. < \alpha \rightarrow$ דוחים H_0 $P.V. > \alpha \rightarrow$ מקבלים H_0 * עבור מבחן דו-צדדי: $\frac{1}{2}P.V. = 1-\phi(\left\lvert\frac{\bar{x}_n-\mu_0}{\sigma/\sqrt{n}}\right\rvert)$
σ^2 **לא ידועה**	$C_\alpha = \mu_0 + t_{1-\alpha}^{n-1}\cdot\frac{S}{\sqrt{n}} \quad :\mu_1 > \mu_0$ $C_\alpha = \mu_0 - t_{1-\alpha}^{n-1}\cdot\frac{S}{\sqrt{n}} \quad :\mu_1 < \mu_0$ * משווים ל- $\bar{x}$	$t_{calc} = \frac{\bar{x}-\mu_0}{S/\sqrt{n}}$ * משווים עם השברון המתאים	
פרופורציה	$C_\alpha = p_0 + z_{1-\alpha}\cdot\sqrt{\frac{p_0q_0}{n}} \quad :p_1 > p_0$ $C_\alpha = p_0 - z_{1-\alpha}\cdot\sqrt{\frac{p_0q_0}{n}} \quad :p_1 < p_0$ $C_{\alpha(1,2)} = p_0 \pm z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{p_0q_0}{n}} \quad :p_1 \neq p_0$	$z_{calc} = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0q_0}{n}}}$	$P.V. = P(z > z_{calc}) = P(z > \frac{\hat{p}-p_0}{\sqrt{\frac{p_0q_0}{n}}})$

דף עזר 13

מבחני טיב התאמה:

המטרה לבדוק האם המדגם שייך להתפלגות מסוימת.

$$H_0: X\sim F$$
$$H_1: otherwise$$

סטטיסטי המבחן עבור מבחני טיב התאמה הינו :

$$\chi_p^2 = \sum_{i=1}^{k}\frac{(n_i - np_{i0})^2}{np_{i0}} = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i}$$ - סטטיסטי פירסון

* n_i - מספר התצפיות בקטגוריה i
* k – מספר הקטגוריות

אזור דחייה ברמת מובהקות α

נדחה את H_0 אם $\chi_p^2 > \chi_{1-\alpha}^2(k-l-1)$, l - מספר הפרמטרים שנאמדו .

<u>שכיחות במדגם</u>: $x_i = n_i = O_i$ $\quad X_i \sim Bin(n, p_i)$ $\quad$ <u>שכיחות צפויה</u>: $E_i = n\cdot p_{i,0}$

* צריך לוודא כי לכל i מתקיים: $n\cdot p_{i,0} \geq 5$, ואם לא אז צריך לאחד.

רגרסיה ליניארית

Y – משתנה תלוי / משתנה מוסבר.
X – משתנה בלתי-תלוי / משתנה מסביר.
לפעמים יש מספר משתנים מסבירים (p משתנים מסבירים) ואז הם מסומנים: $X_1, X_2, ..., X_p$. **המטרה**: רוצים להסביר את Y בעזרת המשתנה/ים המסביר/ים ולתת תחזיות לציון Y על סמך ציון ה – X.

שלבים:

1. ציור דיאגרמת פיזור: מדיאגרמת הפיזור נלמד האם קיים קשר לינארי בין X ל– Y.
2. מציאת קו הרגרסיה: $\boxed{Y = \beta X + \alpha}$ כאשר β - שיפוע הקו, ו- α - חותך הקו
3. חיזוי Y על-פי X ועל-פי קו הרגרסיה.

* מחפשים פרמטר כך ש: $\sum e_i^2 = \min$

שיפוע הקו:

$$\beta = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \frac{\sum xy - n\overline{xy}}{\sum(x^2) - n(\bar{x})^2}$$
$$\beta = r\frac{\sqrt{\sum(y^2)-n(\bar{y})^2}}{\sqrt{\sum(x^2)-n(\bar{x})^2}} = r\cdot\frac{\sqrt{\sum(y_i-\bar{y})^2}}{\sqrt{\sum(x_i-\bar{x})^2}}$$

החותך:

$$\alpha = \bar{Y} - \beta\bar{X}$$

שונות הרגרסיה:

$$\hat{\sigma}^2 = \frac{\sum(Y_i - (\beta X_i + \alpha))^2}{n-2} == \frac{\left(\sum_{i=1}^{n}(y_i-\bar{y})^2\right) - \beta^2\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)}{n-2}$$

מקדם המתאם:

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\left(\sum_{i=1}^{n}(y_i-\bar{y})^2\right)}} = \frac{\sum_{i=1}^{n}(x_i\cdot y_i) - n\cdot\bar{x}\cdot\bar{y}}{\sqrt{\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n}y_i^2 - n\bar{y}^2\right)}}$$

תכונות מקדם המתאם:

- $-1 \leq r \leq 1$
- $r = 0$: אין קשר לינארי בין x ל – y .
- $0 < r \leq 1$: קשר לינארי חיובי בין x ל – y . (כש – x עולה , y עולה).
- $-1 \leq r < 0$: קשר לינארי שלילי בין x ל – y . (כש – x עולה , y יורד).

r^2 - פרופורצית ה"שונות המוסברת" על-ידי x (מקדם דטרמינציה) $0 \leq r^2 \leq 1$

דף עזר 14 – רגרסיה ליניארית

משוואת הרגרסיה: $\boxed{\hat{y} = \hat{\beta}x + \hat{\alpha}}$

סימונים:

$$S_{XX} = \sum_{i=1}^{n}(X_i-\bar{X})^2 = \sum_{i=1}^{n}X_i^2 - n\bar{X}^2 = \sum(x_i-\bar{x})x_i$$
$$S_{YY} = \sum_{i=1}^{n}(Y_i-\bar{Y})^2$$
$$S_{XY} = \sum_{i=1}^{n}Y_i(X_i-\bar{X}) = \sum_{i=1}^{n}X_i(Y_i-\bar{Y}) = \sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}) = \sum x_iy_i - n\bar{x}\bar{y}$$

שיפוע הקו:

$$\boxed{\hat{\beta} = \frac{S_{XY}}{S_{XX}} = r\sqrt{\frac{S_{YY}}{S_{XX}}} = r\sqrt{\frac{(n-1)S_Y^2}{(n-1)S_X^2}} = r\frac{S_Y}{S_X}} \qquad \boxed{Var(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}}$$

כאשר r הוא מקדם המתאם (מופיע בהמשך). כמו כן, יש לשים לב כי S_X ו- S_Y הינם אמדים לסטיות התקן של X ו – Y ושונים מהסימונים S_{XX} ו- S_{YY} שהגדרנו לעיל.

החותך: $\boxed{\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}}$ $\qquad \boxed{Var(\hat{\alpha}) = \frac{\sigma^2}{n} + \bar{x}^2\cdot\frac{\sigma^2}{S_{xx}} = \sigma^2(\frac{\sum x_i^2}{nS_{xx}})}$

אמד לשונות הרגרסיה:

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = \frac{\sum(Y_i - (\hat{\alpha}+\hat{\beta}X_i))^2}{n-2} = \frac{S_{YY} - \hat{\beta}^2S_{XX}}{n-2} = \frac{SS_e}{n-2} = MS_e$$

האמד לשונות של β :

$$\boxed{\hat{\sigma}_{\hat{\beta}}^2 = \frac{\hat{\sigma}^2}{S_{xx}}}$$

האמד לשונות של אלפא:

$$\hat{\sigma}_{\hat{\alpha}}^2 = \frac{\hat{\sigma}^2\sum X_i^2}{nS_{xx}}$$

מקדם המתאם:

$$r^2 = \frac{\hat{\beta}^2S_{xx}}{S_{yy}} = \frac{SS_B}{SS_T} = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$ - אחוז השונות המוסברת

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\left(\sum_{i=1}^{n}(y_i-\bar{y})^2\right)}} = \frac{\sum_{i=1}^{n}(x_i\cdot y_i) - n\cdot\bar{x}\cdot\bar{y}}{\sqrt{\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n}y_i^2 - n\bar{y}^2\right)}} = \hat{\beta}\sqrt{\frac{S_{xx}}{S_{yy}}}$$

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

נוסחא חשובה:

$$r\cdot\frac{S_y}{S_x} = \frac{\sum(x_i-\bar{x}/y_i-\bar{y})\cdot\sqrt{\frac{\sum(y_i-\bar{y})^2}{n-1}}}{\sqrt{\sum(x_i-\bar{x})^2}\cdot\sqrt{\sum(y_i-\bar{y})^2}\cdot\sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}} = \hat{\beta}$$

בדיקת השערות לשיפוע הרגרסיה

H_0: β=0
H_1: β≠0

$$T = \frac{\hat{\beta}-\beta}{\hat{\sigma}_{\hat{\beta}}} = \frac{\hat{\beta}}{SE(\hat{\beta})} = \frac{\hat{\beta}}{\hat{\sigma}}\sqrt{S_{XX}} \sim t_{n-2}$$

אזור הדחייה הוא: $T > t_{1-\frac{\alpha}{2}}(n-2)$, $T < -t_{1-\frac{\alpha}{2}}(n-2)$

$$T^2 = \left[\frac{\hat{\beta}}{SE(\hat{\beta})}\right]^2 = \left[\frac{\hat{\beta}}{\hat{\sigma}}\right]^2S_{XX} = \frac{MS_B}{MS_E} \sim F_{1,n-2}^{1-\alpha}$$

אזור הדחייה הוא: $F > F_{1-\alpha}(1, n-2)$

ניתן לחשב את T גם באמצעות: $T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$

רווח סמך עבור שיפוע הקו

$$\hat{\beta} - t_{n-2}^{1-\alpha/2}\hat{\sigma}_{\hat{\beta}} \leq \beta \leq \hat{\beta} + t_{n-2}^{1-\alpha/2}\hat{\sigma}_{\hat{\beta}}$$

אם 0 מופיע בר"ס לא נדחה את השערת האפס.

לוח ניתוח שונות

את תוצאות הרגרסיה נהוג לסכם בלוח הנקרא לוח ניתוח השונות כמפורט להלן:

	SS	df	MS	F
רגרסיה (explained)	SS_B	1	$MS_B = SS_B/df_B$	MS_B/MS_E
שארית / טעות (error)	SS_E	n-2	$MS_E = SS_E/df_E$	
סה"כ(total)	SS_T	n-1		

$$SS_T = SS_B + SS_E = S_{YY} = \hat{\beta}^2S_{XX} + \sum e_i^2 = S_Y^2(n-1) = \sum(Y_i-\bar{Y})^2$$
$$SS_B = SS_{reg} = \hat{\beta}^2S_{XX} = S_{YY}r^2$$
$$SS_E = \sum e_i^2 = S_{YY} - \hat{\beta}^2S_{XX} = S_{YY}(1-r^2)$$
$$MS_E = \frac{SS_E}{n-2} = \frac{\sum e_i^2}{n-2} = \hat{\sigma}^2$$
$$\frac{SS_B}{SS_T} + \frac{SS_E}{SS_T} = 1 = r^2 + (1-r^2) = \frac{\hat{\beta}^2S_{XX}}{S_{YY}} + \frac{\sum e_i^2}{S_{YY}}$$
$$F = \frac{MS_B}{MS_E} = \frac{\hat{\beta}^2S_{XX}}{\hat{\sigma}^2} = T^2 = \left[\frac{\hat{\beta}}{SE(\hat{\beta})}\right]^2 = \frac{r^2(n-2)}{1-r^2}$$

סטטיסטי F תמיד בעל שתי דרגות חופש: של המונה ושל המכנה. ברגרסיה פשוטה דרגות החופש של המונה תמיד 1, ושל המכנה תמיד n-2. בטבלה מחפשים את הערך של $F_{(1,n-2)}^{1-\alpha}$

סטטיסטי מבחן:

-EOF-