<u>רגרסיה ומודלים סטטיסטיים- תרגול 4</u>

התפלגויות רב מימדיות:

יהיו $Z_1, \dots, Z_n$ משתנים מקריים המפולגים במשותף (ללא שום הנחות נוספות) בהתפלגות $f_{Z_1,\dots,Z_n}(z_1, \dots, z_n)$. נגדיר $Z \in R^n = (Z_1, \dots Z_n)$ ונגדיר $E(Z) = (E[Z_1], \dots, E[Z_n])$.

באותו האופן, אם A היא **מטריצה מקרית**, כלומר A_{ij} הוא מ"מ: $E[A] = \begin{pmatrix} E[A_{11}] & \dots & E[A_{1m}] \\ \dots & \dots & \dots \\ E[A_{n1}] & \dots & E[A_{nm}] \end{pmatrix}$

תכונות של תוחלות של מטריצות (ובפרט וקטורים) מקריים:

עבור Z, W מטריצות מקריות ו-A, B מטריצות קבועות, לא מקריות (דטרמיניסטיות):

1. $\mathbb{E}[\boldsymbol{Z} + \boldsymbol{W}] = \mathbb{E}[\boldsymbol{Z}] + \mathbb{E}[\boldsymbol{W}]$
2. $\mathbb{E}[\boldsymbol{AZB}] = \boldsymbol{A}\mathbb{E}[\boldsymbol{Z}]\boldsymbol{B}$
3. $\mathbb{E}[\boldsymbol{AU} + \boldsymbol{C}] = \boldsymbol{A}\mathbb{E}[\boldsymbol{U}] + \boldsymbol{C}$ (from 1+2)

עבור זוג וקטורים מקריים Z, W נגדיר את מטריצת השונות המשותפת בין Z, W:

$$Cov(Z, W) \coloneqq E([Z - E(Z)][W - E(W)]^T)$$

ובאופן דומה את מטריצת השונויות של הוקטור Z:

$$Var(Z) := Cov(Z, Z) \coloneqq E([Z - E(Z)][Z - E(Z)]^T)$$

טענה: $Var(Z)_{ij} = cov(Z_i, Z_j) = cov(Z_j, Z_i) = Var(Z)_{ji}$

"הוכחה":

$$Var(Z)_{ij} = E([Z - E(Z)][Z - E(Z)]^T)_{ij}$$
$$= E\big([Z - E(Z)]_i [Z - E(Z)]_j\big) = E([Z_i - E(Z_i)][Z_j - E(Z_j)]) = cov(Z_i, Z_j)$$

בתרגיל תוכיחו את התכונות הבאות:

Properties of covariance matrix. $\boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{R}$ random vectors; $\boldsymbol{a}$ fixed vector. Then the following properties hold:

1. $\text{cov}(\boldsymbol{Z}, \boldsymbol{W}) = \text{cov}(\boldsymbol{W}, \boldsymbol{Z})^\top$
2. $\text{cov}(\boldsymbol{Z} + \boldsymbol{R}, \boldsymbol{W}) = \text{cov}(\boldsymbol{Z}, \boldsymbol{W}) + \text{cov}(\boldsymbol{R}, \boldsymbol{W})$
3. $\text{cov}(\boldsymbol{AZ}, \boldsymbol{BW}) = \boldsymbol{A}\text{cov}(\boldsymbol{Z}, \boldsymbol{W})\boldsymbol{B}^\top$
4. $\text{cov}(\boldsymbol{AZ}) = \boldsymbol{A}\text{cov}(\boldsymbol{Z})\boldsymbol{A}^\top$ (from 3)
5. $V(\boldsymbol{a}^\top \boldsymbol{Z}) = \boldsymbol{a}^\top \text{cov}(\boldsymbol{Z})\boldsymbol{a}$ (from 4)
6. $\text{cov}(\boldsymbol{Z})$ is a nonnegative definite matrix (from 5)

<u>שאלה:</u>

יהיו המשתנים המקריים הבאים: $Z_1, \dots, Z_5 \sim N(0,1)\ iid.$.

א. מצאו את וקטור התוחלות ומטריצת השונויות של הוקטור המקרי $Z = (Z_1, \dots, Z_5)^T$.

ב. נגדיר את ההעתקה: $A: R^5 \to R^3$: $A(v) = (2v_1 - 3v_2 + 4v_3 + 7, 8v_2 - v_4 + 2v_5 + 3, 6v_3 - 1)^T$. האם ההעתקה ליניארית?

על בסיס הסעיף הקודם חשבו את וקטור התוחלות ומטריצת השונויות של $A(Z)$.

ג. האם תשובתכם הייתה משתנה אילו היה ידוע כי $Z_1, \dots Z_5$ היו נדגמים כך ש- $Z_i \sim unif\left(-\sqrt{3}, \sqrt{3}\right)$?

פתרון:

א. עבור וקטור התוחלות אנו צריכים לחשב את התוחלת של כל אחת מ-5 הכניסות:

$$E(Z)^t = \left(E(Z_1), \dots, E(Z_5)\right)^T = (0, \dots, 0)^T$$

בעבור מטריצת השונויות: כפי שראינו, עלינו לחשב כל אחת מהשונויות המשותפות:
כיוון שבלתי תלויים אז בלתי מתואמים ולכן $Cov\left(Z_i, Z_j\right) = 0\ , \forall i \neq j$.
ואילו $Cov(Z_i, Z_i) = Var(Z_i) = 1, \forall i$. לכן- מטריצת השונויות המשותפות: $cov(Z) = I_5$

ב. זו איננה העתקה ליניארית כיוון שהוכחנו שכל העתקה ליניארית מקיימת $T(0) = 0$. שימו לב שנוכל לכתוב את A כך:

$$A(v) = Bv + (7,3,-1)^T \coloneqq Bv + \mu$$

עבור $B = \begin{pmatrix} 2 & -3 & 4 & 0 & 0 \\ 0 & 8 & 0 & -1 & 2 \\ 0 & 0 & 6 & 0 & 0 \end{pmatrix}$. נקבל:

$$E(AZ) = BE(Z) + \mu = \mu$$
$$cov(AZ) = cov(BZ + \mu) = cov(BZ) = Bcov(Z)B^T = BIB^T = BB^T$$

יהיו $Z, W \in R^p$ וקטורים מקריים. הראו שהבאים שקולים:

(1) $\forall v \in R^p, Var(v^T Z) \geq Var(v^T W)$

(2) $B \coloneqq Var(Z) - Var(W)$ היא מטריצה חיובית למחצה.

(3) קיימת המטריצה $B^{\frac{1}{2}}$.

הוכחה: תרגיל.

המודל הלינארי:

נתונים: $(\boldsymbol{X}_i, Y_i), \quad i = 1, \dots, n$

מודל ליניארי:

$$Y_i = \sum_{j=0}^{p} \beta_j X_{ij} + \epsilon_i, \qquad \mathrm{Cov}(\epsilon_i, \epsilon_{i'}) = \begin{cases} \sigma^2, & i = i' \\ 0, & i \neq i' \end{cases}$$

כאשר $\boldsymbol{X}_i = (1, X_{i1}, \dots, X_{ip})^\top$ הם ממימד $p+1$,

וכאשר $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ ו-σ^2 הם קבועים לא ידועים

אפשר לקבל ייצוג קומפקטי בעזרת כתיב מטריצות. נסמן:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \qquad \boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{bmatrix}, \qquad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

אז את המודל הליניארי אפשר לכתוב:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}, \qquad \mathbb{E}[\boldsymbol{\epsilon}] = \boldsymbol{0}, \ \ \mathrm{cov}[\boldsymbol{\epsilon}] = \sigma^2 \boldsymbol{I}$$

כמו כן: $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}, \qquad \boldsymbol{e} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$

(הערה: אם לא נציין אחרת, $\hat{\boldsymbol{\beta}}$ זה אומד הריבועים הפחותים)

<u>שאלה</u>

לפניכם מספר טענות. התאימו לכל טענה האם מדובר בהנחה או בתוצאה מתמטית:

1. $\hat{\beta} = argmin_{\beta}||Y - X\beta||^2$
2. $X\beta = E(Y|X)$
3. $0 = E(e_i)$
4. $0 = E(\bar{e})$ (עבור רגרסיה עם חותך).
5. $X^T e = 0$
6. $Cov(Y) = \sigma^2 I$

פתרון:

1. תוצאה מתמטית. הראינו במספר דרכים. זהו פתרון בעיית אופטימיזציה.
2. הנחה: $Y = X\beta + \epsilon$ וכן $E(\epsilon) = EE(\epsilon|X) = E(0) = 0$.
3. הנחה:

$$E(e_i) = E\left(Y_i - X_i^T\hat{\beta}\right) = E(Y_i) - E(X[X^TX]^{-1}X_i^TY) = X_i^T\beta - X[X^TX]^{-1}X_i^TX\beta = X_i^T\beta - X_i^T\beta = 0$$

4. תוצאה: מתנאי סדר ראשון מתקבלת המשוואה הנורמלית הראשונה: $\sum_{i=1}^n e_i = 0$. דרך נוספת:

$$\sum_{i=1}^n e_i = \sum_{i=1}^n Y_i - \hat{Y}_i = \sum_{i=1}^n (I - P_X)Y = 1_n^T(I - P_X)Y = 1_n^T(I - P_X)^TY = ((I - P_X)1)_n^TY$$
$$= 0^TY = 0$$

וזאת כיוון שברגרסיה עם חותך, עמודת האחדות שייכת ל-$IM(X)$.

5. תוצאה: $X^Te = X^T(Y - P_XY) = X^T(I - P_X)Y = 0$
6. הנחה: $cov(Y) = cov(X\beta + \epsilon) = cov(\epsilon) = \sigma^2 I$

וכאן השתמשנו בהנחת הליניאריות במעבר הראשון ובשיוויון השונויות במעבר האחרות.

שאלה:

לפניכם מתוארים מספר מקרים. עבור כל אחד מהם פרטו את ההתפלגויות של $X,Y,Y|X,\epsilon$ וכתבו אילו הנחות של המודל הלינארי כל אחד מהם מקיים:

1. $X_i \in R^p$ **קבועים מראש**, $Y_i = X_i^T\beta + \epsilon_i$ כאשר $\epsilon_i \sim N(0,1)$ ב"ת.
2. $X_i \in R^p$ מ"מ נורמליים סטנדרטיים וב"ת, $Y_i = X_i^T\beta + \epsilon_i$ כאשר $\epsilon_i \sim N(0,1)$ ב"ת.
3. $X_i \in R^1$ מ"מ נורמליים סטנדרטיים וב"ת, $Y_i = X_i^2\beta + \epsilon_i$ כאשר $\epsilon_i \sim U(-1,1)$ ב"ת.
4. $X_{it} \in R^p$ **קבועים מראש**- כאשר X_{itj} היא הדגימה של האדם ה- i בתקופה ה-t, $Y_{it} = X_{it}\beta + \epsilon_{it}$ כאשר $\epsilon_{it} \sim N(0,1)$ והנדגמים ב"ת.

פתרון:

1. כל ההנחות מתקיימות.
2. כנ"ל.
3. כנ"ל. באופן כללי, נוכל להגדיר את X כמטריצת $Vandermonde$: בה האיבר ה-ij הוא X_i^{j-1} כאשר במקרה כזה האמידה של וקטור המקדמים תהיה בעצם אמידה של מקדמי פולינום.
4. כאן סביר להניח שהנחת חוסר המתאם בין $\epsilon_{it_k},\epsilon_{it_m}$ אינה מתקיימת. דוגמה למודל שכזה:

$$Y_i = X_i\beta + \epsilon_i + \epsilon_{i+1}\cdot \mathbf{1}_{\{i\ is\ odd\}}$$

שאלה

1. יהי $0 \neq v \in R^n$. הראו כי $\frac{vv^T}{||v||^2}$ היא מטריצת הטלה אורתוגונלית. מה דרגת המטריצה?
2. יהיו $Y_1,\dots,Y_n \sim (\mu,\sigma^2)$ מ"מ ב"ת ש"ה, כאשר שני הפרמטרים לא ידועים. הראו כי $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i-\bar{Y})^2$ הוא אומד ח"ה ל-σ^2.
3. הניחו כעת כי $Y_1,\dots,Y_n \sim N(\mu,\sigma^2)$. הוכיחו את התוצאה, שראינו בעבר $(n-1)S_n^2 \sim \sigma^2\chi_{n-1}^2$.

פתרון:

1. $(\frac{vv^T}{||v||^2})^T = \frac{1}{||v||^2}v^{T^T}v^T = \frac{vv^T}{||v||^2}$

$$\left(\frac{vv^T}{||v||^2}\right)^2 = \frac{vv^T}{||v||^2}\frac{vv^T}{||v||^2} = \frac{||v||^2}{||v||^4}vv^T = \frac{vv^T}{||v||^2}$$

דרגת המטריצה היא 1 (אפשר לדרג ולראות, או לשים לב שמדובר ב-P_X עבור $X=v$.

2. תרגיל (שפתרנו הרבה פעמים בעבר).
3. ראשית, ניתן לכתוב כל משתנה נורמלי כ-$\mu + \sigma Z_i$ עבור $Z_i \sim N(0,1)$:

$$\sum_{i=1}^n (Z_i - \bar{Z})^2 = \left|\left|\left(I - \frac{1}{n}\cdot 1_n 1_n^T\right)Z\right|\right|^2 = Z^T\left(I - \frac{1}{n}\cdot 1_n 1_n^T\right)Z$$

כאשר $Z=(Z_1,\dots,Z_n)^T$, לקחנו את v מסעיף א' להיות הוקטור 1_n וכיוון שזו מטריצת הטלה אז גם $(I-\frac{1}{n}\cdot 1_n 1_n^T)$ המטילה למרחב המשלים שמימדו $n-r=n-1$. כמו כן סימטרית ואיידמפוטנטית ומכאן המעבר האחרון. כעת כיוון שמטריצת הטלה אז ניתנת לליכסון כך שישנם $n-1$ ע"ע שהם 1 וע"ע אחד שהוא 0. לכן ממשפט הפירוק הספקטרלי:

$$Z^T\sum_{i=1}^n \lambda_i U_i U_i^T Z = \sum_{i=1}^{n-1}(U_i^TZ)^T(U_i^TZ) = \sum_{i=1}^{n-1}(U_i^TZ)^2$$

נתבונן בהתפלגות של $U_i^T Z = \sum_j U_i^j Z_j$. זו קומבינציה לינארית של מ״מ נורמליים ולכן מתפלגת נורמלית. התוחלת היא: $\sum_j U_i^j E(Z_j) = 0$ והשונות היא: $\sum_j {U_i^j}^2 = ||U_i||^2 = 1$. כלומר התפלגות נורמלית סטנדרטית, כמו במקור. כיוון שזו פונקציה של מ״מ ב״ת אז גם הפונקציות ב״ת. מכאן שנותרנו עם סכום ריבועים של $n-1$ מ״מ נורמליים סטנדרטיים וב״ת אשר בהגדרה (עד שתראו זאת בהסתברות לסטטיסטיקאים) מפולגים χ^2_{n-1}.
כעת כל שנותר הוא להכפיל ב-σ לקבל את המבוקש.

<u>שאלה</u>

2.נתון $Z \in \mathbb{R}^m$ וקטור מקרי. הראו כי מתקיים ש

$$\mathbb{E}\left(||Z||^2\right) = tr(\mathbb{E}[ZZ^T])$$

הסיקו מכך כי אם $\mathbb{E}[Z] = 0$ אזי מתקיים כי

$$\mathbb{E}\left(||Z||^2\right) = tr(cov[Z])$$

פרטו והצדיקו כל שלב בהוכחה.

2 הוכיחו כי עבור U, V וקטורים מקריים שידוע שהתוחלת של לפחות אחד מהם היא וקטור ה-0, מתקיים:

$$E(U^T V) = tr\left(cov(VU^T)\right)$$

הוכחה:

1. $E\left(||Z||^2\right) = E(Z^T Z) = E(tr[Z^T Z]) = E(tr(ZZ^T))$
2. $E\left(||Z||^2\right) = tr\left(E(ZZ^T)\right) = tr\left(E\left(Z - E(Z)\right)\left(E\left(Z - E(Z)\right)^T\right) = tr(Cov(Z))\right.$

שאלה- מועד א׳ תשפ״ד

הניחו שהנחות המודל הלינארי מתקיימות.

ד. (15 נק׳) נסמן $M := \text{Im}(\boldsymbol{X})$. נסמן ב-$\tilde{\boldsymbol{X}}$ את המטריצה המתקבלת מ-$\boldsymbol{X}$ ע״י מחיקת חלק מהעמודות, ונסמן $L := \text{Im}(\tilde{\boldsymbol{X}})$. הראו שמתקיים $P_L \boldsymbol{Y} = P_L \hat{\boldsymbol{Y}}$ (כאשר $\hat{\boldsymbol{Y}}$ זה וקטור הערכים החזויים במודל עם מטריצת ה-$\boldsymbol{X}$ המקורית). האם נדרשות הנחות המודל הלינארי או המודל הלינארי הנורמלי כדי שהטענה תתקיים?

ב. מצאו את $E\left(||e||^2\right)$

ב. **(10 נק׳)** חשבו את הגדלים הבאים, <u>וציינו את המימד</u> שלהם: $\text{cov}(\boldsymbol{e})$, $\text{cov}(\boldsymbol{\epsilon}, \boldsymbol{e})$, $\text{cov}(\hat{\boldsymbol{\beta}}, \boldsymbol{\epsilon})$.

ג. מצאו את $cov(Y_i, \hat{Y}_i)$.

פתרון:

1. ראינו כי לכל L, M תתי מרחבים כך ש-$M \subseteq L$ מתקיים $P_M P_L = P_L P_M = P_L$ שכן (תזכורת להוכחה- אפשר היה פשוט לצטט את המשפט):

$$\forall v \in R^n, \quad P_M P_L v = P_L v \ \ since \ P_L v \in M$$
$$\Rightarrow P_L P_M v = P_L^T P_M^T v = (P_M P_L)^T v = P_L^T v = P_L v$$

ואז:

$$P_L \hat{Y} = P_L P_M Y = P_L Y$$

רק נותר להראות $IM(\tilde{X}) \subseteq IM(X)$: נניח בה"כ שמדובר ב-$p$ < m העמודות האחרונות. אז לכל
$w \in IM(\tilde{X}) \Rightarrow w = \sum_{i=1}^{p+1-m} X^i k_i = \sum_{i=1}^{p+1-m} X^i k_i + \sum_{i=p-m}^{p+1} X^i \cdot 0 \Rightarrow w \in IM(X)$

כל אלו נכונות ללא שום הנחה.

2.

$$E(e) = E(Y - \hat{Y}) = E(Y) - E(P_X Y) = X\beta - P_X X\beta = 0$$

כאשר השתמשנו בהנחת הליניאריות ובכך ש-$P_X X = X$.
נפעיל את הטענה מהשאלה הקודמת ונקבל:

$$E\left(||e||^2\right) = tr\left(cov(e)\right) = tr(cov(I - P_X)Y)) = tr((I - P_X)cov(Y)(I - P_X)^T)$$
$$= tr(\sigma^2 (I - P_X)(I - P_X)^T) = tr(\sigma^2 (I - P_X)) = \sigma^2 \cdot rank(I - P_X)$$
$$= \sigma^2 \cdot (n - p - 1)$$

3. תרגיל

שאלה- בוחן אמצע תשפ"ב

נתון מודל ליניארי, $Y = X\beta + \varepsilon$ כאשר $\varepsilon \sim (0, \sigma^2 I_n)$.

עבור $u, v \in \mathbb{R}^n$ כלשהם נגדיר את נורמת מהלנוביס (Mahalanobis) באופן הבא:

$$||u - v||_\Sigma = \sqrt{(u - v)^T \Sigma^{-1} (u - v)}$$

כאשר $\Sigma \in \mathbb{R}^{n \times n}$ מטריצה סימטרית חיובית.

כמו כן נגדיר את $\hat{\beta}^\Sigma$ באופן הבא:

$$\hat{\beta}^\Sigma = \underset{\beta \in \mathbb{R}^{p+1}}{argmin} \, ||Y - X\beta||_\Sigma^2$$

א. הראו שמתקיים

$$\hat{\beta}^\Sigma = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$$

והסבירו את האומד שמתקבל במקרה $\Sigma = I_n$.

הדרכה: אפשר (לא חייב ש) להעזר בעובדה הבאה: $A \in \mathbb{R}^{m\times m}$ מטריצה סימטרית חיובית, אזי קיימת מטריצה $B \in \mathbb{R}^{m\times m}$ סימטרית חיובית כך ש: $B^TB = A$. ניתן להפעיל מסקנה זו על Σ כך שיהיה ניתן לייצג את Σ^{-1} באופן הבא: $\Sigma^{-1} = C^TC$ עבור מטריצה C ריבועית כלשהי. מצאו מי היא C, הסבירו מדוע היא מוגדרת היטב, ואז הגדירו משתנים חדשים

$\tilde{X} = \boldsymbol{C}X \qquad \tilde{Y} = \boldsymbol{C}Y,$ ופתרו את הבעיה (שימו לב שעבור ה **C** הנכון הפתרון מיידי, חישבו מדוע).

ב. הראו כי $\hat{\beta}^{\Sigma}$ הינו אומד חסר הטיה לβ, וכן מצאו את $cov(\hat{\beta}^{\Sigma})$.

פתרון: תרגיל.