**ניתוח בייסיאני של קשרים בין מאפיינים דמוגרפיים ואקדמיים לבין רמות לחץ, חרדה ודיכאון בקרב סטודנטים:**

**מבוא:**

בעידן המודרני, חוויות האקדמיה והלחץ המנטלי הכרוך בכך משפיעות רבות על רמת הלחץ, חרדה ודיכאון בקרב סטודנטים. מטרת המחקר הנוכחי היא לבחון את הקשר בין מאפיינים דמוגרפיים ואקדמיים של הסטודנטים לבין רמות הלחץ, החרדה והדיכאון שהם חווים במהלך לימודיהם.

הנתונים המשמשים במחקר זה נאספו באמצעות סקר שנערך בקרב סטודנטים מאוניברסיטאות שונות במדינת בנגלדש. הנתונים כוללים משתנים דמוגרפיים כגון גיל, מגדר, מחלקה אקדמית, שנת לימוד, ממוצע ציונים, (CGPA) קבלת מלגות וכן מדדים של רמות לחץ, חרדה ודיכאון.

בעזרת ניתוחים בייסיאניים, המחקר שואף לבדוק שאלות מחקר מרכזיות כגון: האם ישנם הבדלים משמעותיים ברמות הדיכאון בין נשים וגברים? האם לגורמים כמו גיל, מחלקה אקדמית ושנת הלימודים יש השפעה על רמות הלחץ והחרדה? בנוסף, ייבדק הקשר בין רמות הלחץ והחרדה לבין רמות הדיכאון בעזרת השוואת מודלים, במטרה לזהות משתנים אשר עשויים להוות גורמים מנבאים לתופעות אלו בקרב הסטודנטים.

הנתונים נותחו באמצעות שיטות מתקדמות של סטטיסטיקה בייסיאנית במטרה להשיג הערכות מדויקות ולהבין את הדינמיקה המורכבת של בריאות הנפשית בקרב הסטודנטים.

**מקור הנתונים:**

[**https://www.kaggle.com/datasets/mohsenzergani/bangladeshi-university-students-mental-health**](https://www.kaggle.com/datasets/mohsenzergani/bangladeshi-university-students-mental-health)   
**הצדקת השימוש בפונקציית הנראות בכל השאלות:**

במהלך הניתוחים הבייסיאניים שנעשו, בכל שלושת השאלות במחקר זה, נבחרה פונקציית הנראות הנורמלית (Normal Likelihood Function) כבסיס למודלים הסטטיסטיים. הבחירה בפונקציה זו נבעה ממספר שיקולים מרכזיים המשותפים לכל השאלות שנבחנו:

1. התפלגות נורמלית: המדדים הפסיכולוגיים שנבחנו, כגון דיכאון, לחץ וחרדה, נוטים להתפלג בצורה נורמלית באוכלוסיית הסטודנטים. משתנים אלה נמדדים על פי רוב בצורה רציפה, והם מתאפיינים בפיזור סביב ממוצע מסוים. מכאן, שהשימוש בהתפלגות נורמלית לניבוי ערכים אלו הוא טבעי ומתאים להנחות הבסיסיות הקיימות במחקר הפסיכולוגי.
2. פשטות ואינטואיטיביות: פונקציית הנראות הנורמלית נבחרה גם בשל פשטותה והיכולת שלה לספק תוצאות אינטואיטיביות וברורות. היא מאפשרת ניתוח ישיר ומובן של הקשרים בין המשתנים הבלתי תלויים לבין המשתנים התלויים, תוך שמירה על בהירות ההבנה של התוצאות.
3. התאמה לנתונים :הבחירה בפונקציית הנראות הנורמלית הוכיחה את עצמה במהלך הבדיקות הפוסטריוריות שבוצעו. בכל השאלות נבדקה ההתאמה בין הנתונים הנצפים בפועל לבין הנתונים המדומים על פי המודלים, והמודלים הנורמליים הציגו התאמה טובה לנתונים האמיתיים. התאמה זו מחזקת את ההנחה כי השימוש בפונקציית הנראות הנורמלית מתאים באופן הטוב ביותר לניתוח המשתנים שבדקנו**.**

**שאלה 1: האם קיים קשר בין מגדר לרמת דיכאון בקרב הסטודנטים?**

**שיטה:**

כדי לענות על שאלה זו, ביצענו ניתוח בייסיאני של נתוני הסטודנטים באמצעות מודל רגרסיה ליניארית. במודל זה, רמת הדיכאון (המשתנה התלוי) נמדדה כמשתנה רציף, ומגדר (המשתנה הבלתי תלוי) נמדד כמשתנה בינארי (1 עבור נשים ו-0 עבור גברים). השימוש בפונקציית נראות נורמלית היה מתאים, שכן ההנחה היא שרמת הדיכאון מתפלגת בצורה נורמלית באוכלוסיית הסטודנטים.

המודל שנבנה כלל פרמטרים של מקדם המגדר ,(beta\_gender) פרמטר החותך ,(alpha) וסטיית התקן של ההתפלגות .(sigma) הנתונים נבדקו על ידי שימוש ב-4 שרשראות מרקוב (MCMC) עם 2000 איטרציות בכל אחת, כאשר 1000 מתוכן שימשו כ"חימום."

**הצדקת הפריורים:**

* **:** alpha **∼ N(0,10)** הפרמטר α מייצג את החותך של המודל, כלומר את רמת הדיכאון הבסיסית כאשר יתר המשתנים הם באפס. הבחירה בפריור זה משקפת את ההנחה שרמת הדיכאון בקבוצת הבסיס היא ממוצעת.
* beta\_gender **∼ N(5,2) :** הפרמטרbeta\_gender מייצג את השפעת המגדר על רמת הדיכאון. פריור זה עם תוחלת של 5 וסטיית תקן של 2 מבוסס על האמונה שנשים מדווחות על רמות דיכאון גבוהות יותר מגברים, תוך שמירה על גמישות מסוימת במודל.
* sigma **∼ N(0,5) :** הפרמטר σ מייצג את סטיית התקן של רמות הדיכאון המנובאות, ומשקף את השונות בנתונים. הפריור עם תוחלת 0 וסטיית תקן של 5 נבחר כדי לאפשר למודל גמישות בהערכת שונות הנתונים, במיוחד במדדים פסיכולוגיים.

**בדיקת רגישות:**

במהלך הניתוח נבדקה הרגישות של המודל לשינויים בפריורים שהונחו עבור הפרמטרים. בתחילה, הונחו הפריורים הבאים:

* alpha **∼ N(0,10)**, beta\_gender **∼ N(5,2)**, sigma **∼ N(0,5)**

לאחר מכן, בוצעה בדיקת רגישות על ידי שינוי הפריורים:

* alpha **∼ N(0,10)**, beta\_gender **∼ N(5,5)**, sigma **∼ N(0,2)**

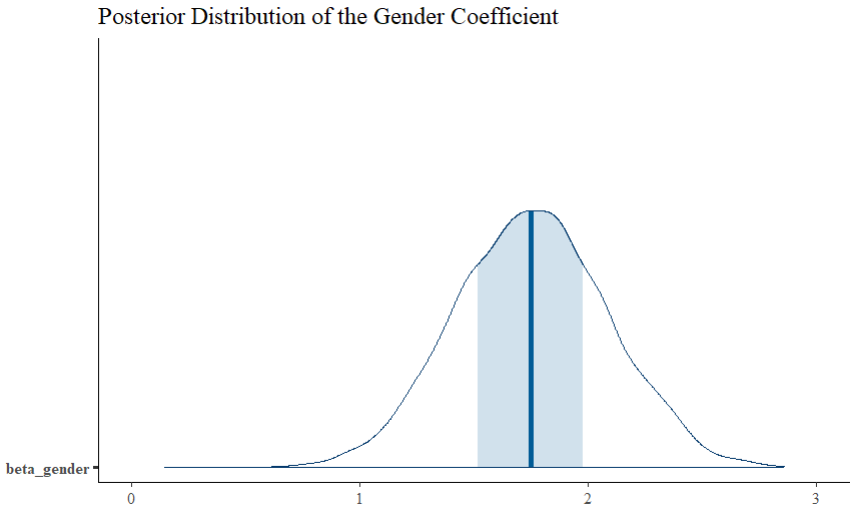
השפעת השינויים בפריורים על התוצאות:

* הממוצע של beta\_gender​ ירד מעט מ-1.7549 ל-1.6790. סטיית התקן של beta\_gender עלתה קלות מ-0.3344 ל-0.3477. רווח הסמך של beta\_gender השתנה מעט, אך נותר יציב יחסית, נע בין 1.1027 ל-2.4347 במודל המקורי, ובין 1.0078 ל-2.3590 לאחר השינוי. ערכי EES ו Rhat-נותרו דומים, עם Rhat קרוב ל-1 ו EES-גבוה, מה שמעיד על התכנסות טובה של המודל.

**מסקנה:**

השינוי בפריורים לא השפיע בצורה משמעותית על התוצאות, מה שמראה שהמודל יציב ואמין גם תחת שינויים בפריורים. הממצאים אינם תלויים יתר על המידה בהנחות הראשוניות, ולכן הפריורים שנבחרו בתחילה מתאימים לשאלת המחקר ומספקים תוצאות אמינות.

תמונה שמכילה עיצוב

התיאור נוצר באופן אוטומטי ברמת מהימנות בינונית

**הסברים על הגרפים:**

1. **Posterior Distribution of the Gender Coefficient** - גרף זה מציג את ההתפלגות הפוסטריורית של מקדם המגדר .(beta\_gender​) ניתן לראות את ההסתברות שרמת הדיכאון גבוהה יותר עבור נשים, כאשר ההתפלגות מרוכזת סביב ערך חיובי עם רווח סמך של 95%.
2. **Posterior Predictive Check** - גרף זה משווה בין ההתפלגות של הנתונים המדווחים בפועל (קו כחול כהה) לבין ההתפלגות המדומה של הנתונים על פי המודל (כחול בהיר). ההתאמה בין ההתפלגויות מצביעה על כך שהמודל מצליח לנבא את הנתונים בצורה טובה.

**מסקנות סופיות מהשאלה:** נמצא כי יש קשר חיובי ומשמעותי בין מגדר לרמת הדיכאון, כאשר נשים נוטות לדווח על רמות דיכאון גבוהות יותר מגברים. המודל שהותאם בעזרת ניתוח בייסיאני הציג התאמה טובה לנתונים האמיתיים, והפריורים שנבחרו התבררו כמתאימים ויציבים בבדיקת הרגישות.

**שאלה 2: האם יש קשר בין משתנים דמוגרפיים ואקדמיים (גיל, הנדסה, שנת לימוד) לבין רמות הלחץ והחרדה בקרב הסטודנטים?**

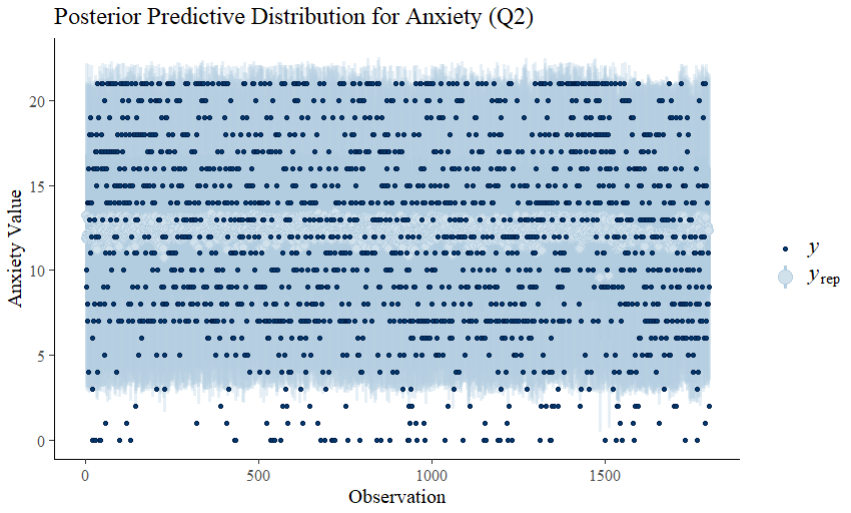
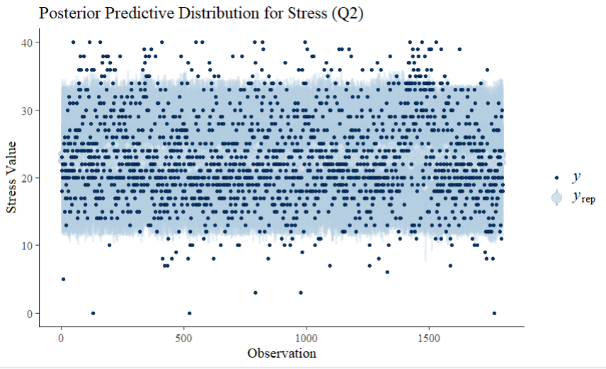
שיטה: כדי לענות על שאלה זו, ביצענו ניתוח בייסיאני באמצעות מודל רגרסיה ליניארית למספר משתנים תלויים. במודל זה, רמות הלחץ והחרדה (המשתנים התלויים) נמדדו כמשתנים רציפים. המודל כלל את הגיל, השתייכות הסטודנט למקצועות ההנדסה ושנת הלימוד (האם בשנתיו הראשונות או האחרונות) כמשתנים בלתי תלויים. עבור כל משתנה תלוי, נבנה מודל נפרד הכולל פרמטרים של חותך ,(alpha) מקדמים עבור המשתנים הבלתי תלויים (beta\_age, beta\_engineer, beta\_senior), וסטיית התקן של ההתפלגות .(sigma) הניתוח בוצע באמצעות שרשראות מרקוב (MCMC) בדומה לשאלה 1.

**הצדקת הפריורים:**

* **- α\_stress ∼ N(0,10), α\_anxiety ∼ N(0,10):** פרמטרים אלה מייצגים את החותך של המודל, כלומר את רמת הלחץ או החרדה הבסיסית. הבחירה בפריור נורמלי עם תוחלת 0 וסטיית תקן 10 משקפת את הציפייה לערכים ממוצעים, אך גם לפוטנציאל לערכים גבוהים במידה מסוימת.
* **β\_age\_stress ∼ N(-5,2), β\_age\_anxiety ∼ N(-5,2):** - פרמטרים אלה מייצגים את השפעת הגיל על רמות הלחץ והחרדה. הבחירה בתוחלת שלילית (-5) משקפת את ההנחה שגיל מבוגר יותר מקושר לרמות נמוכות יותר של לחץ וחרדה, תוך שמירה על גמישות מסוימת במודל.
* **β\_engineer\_stress ∼ N(5,2), β\_engineer\_anxiety ∼ N(5,2):** - פרמטרים אלו מייצגים את השפעת תחום ההנדסה על רמות הלחץ והחרדה. הבחירה בתוחלת חיובית (5) מבוססת על ההנחה שסטודנטים להנדסה חווים יותר לחץ וחרדה עקב עומס לימודים גבוה יותר.
* **β\_senior\_stress ∼ N(5,2), β\_senior\_anxiety ∼ N(5,2):** - פרמטרים אלו מייצגים את השפעת התקדמות בשנות הלימוד על רמות הלחץ והחרדה. הבחירה בתוחלת חיובית (5) משקפת את ההנחה שסטודנטים בשנים מתקדמות חווים לחץ וחרדה גבוהים יותר.
* **σ\_stress ∼ N(0,5), σ\_anxiety ∼ N(0,5):** - פרמטרי סטיית התקן מייצגים את השונות ברמות הלחץ והחרדה. הפריורים נבחרו כדי לאפשר למודל גמישות רחבה בהערכת שונות הנתונים, מבלי לכפות הערכות ספציפיות מראש.

**בדיקת רגישות:**

הפריורים הראשוניים שנבחרו עבור המודל שיקפו את האמונות שלנו לגבי ההשפעות הצפויות של המשתנים. הערכים של הפריורים שהכונסו היו מוטים בהתאם למפורט לעיל. לאחר מכן, שינינו את הפריורים לפריורים רחבים יותר ושאינם מוטים כדי לבדוק את מידת ההשפעה של ההטיות הראשוניות על המודל. הממצאים הראו כי למרות השינוי בפריורים, המודל שמר על יציבותו, מה שמעיד על כך שהמסקנות המרכזיות של המודל אינן תלויות במידת ההטייה בפריורים.

  
**הסברים על הגרפים:**

1. **- Posterior Predictive Distribution for Stress (Q2)**: הגרף הזה מראה איך הנתונים האמיתיים של רמות הלחץ (הנקודות הכהות) משווים לנתונים המדומים שנוצרו על ידי המודל (הצללים הבהירים). אפשר לראות שהמודל מצליח לתפוס את המגמות הכלליות, אבל יש אזורים שבהם הפיזור רחב יותר, מה שעשוי להצביע על כך שיש שונות גבוהה בנתונים האמיתיים.
2. **Posterior Predictive Distribution for Anxiety (Q2)**: - הגרף הזה דומה לראשון, אבל מתמקד ברמות החרדה. שוב, רואים שהמודל מתאים באופן סביר את הנתונים האמיתיים (הנקודות הכהות) לנתונים המדומים (הצללים הבהירים).

**מסקנות סופיות מהשאלה:**

תוצאות המודל שנבדק לשאלה 2 מראות ערכי Rhat הקרובים ל-1 דבר אשר מעיד על התכנסות טובה של השרשראות, כאשר גודל המדגם האפקטיבי (n\_eff) היה גבוה, מה שמרמז על דגימה יעילה.

במונחים של השפעת המשתנים, β\_age\_stress (0.62-) ו-β\_age\_anxiety (0.34-) מראים שלגיל יש השפעה שלילית על רמות הלחץ והחרדה, כלומר, ככל שהגיל נמוך יותר, רמות הלחץ והחרדה נמוכות יותר. β\_engineer\_stress (1.05-) ו-β\_engineer\_anxiety (0.04-) מצביעים על כך שלהיות סטודנט בהנדסה קשור להפחתת רמות הלחץ והחרדה, במיוחד ללחץ. לעומת זאת, β\_senior\_stress (0.89) ו-β\_senior\_anxiety (1.01) מצביעים על כך שהיותו של סטודנט בשנים מתקדמות מעלה את רמות הלחץ והחרדה. תוצאות אלו מבוססות על מובהקות סטטיסטית ברווחי הסמך של הפרמטרים, המעידים על השפעה מובהקת של המשתנים הנבדקים.

**שאלה 3: האם הוספת מדדי לחץ וחרדה למודל דמוגרפי ואקדמי (גיל, ממוצע ציונים, שנת לימודים) משפרת את הניבוי של רמת הדיכאון בקרב הסטודנטים?**

**שיטה:**  כדי לענות על שאלה זו, בנינו שני מודלים בייסיאניים שונים וביצענו השוואה ביניהם. במודל הראשון, רמת הדיכאון (המשתנה התלוי) נמדדה כמשתנה רציף, והמשתנים הבלתי תלויים היו גיל, ממוצע ציונים ,(CGPA) ושנת הלימודים של הסטודנטים. במודל השני, נוספו מדדי חרדה ולחץ כמשתנים בלתי תלויים נוספים. המטרה היא לבדוק האם הוספת מדדים אלו משפרת את יכולת הניבוי של המודל לגבי רמת הדיכאון.

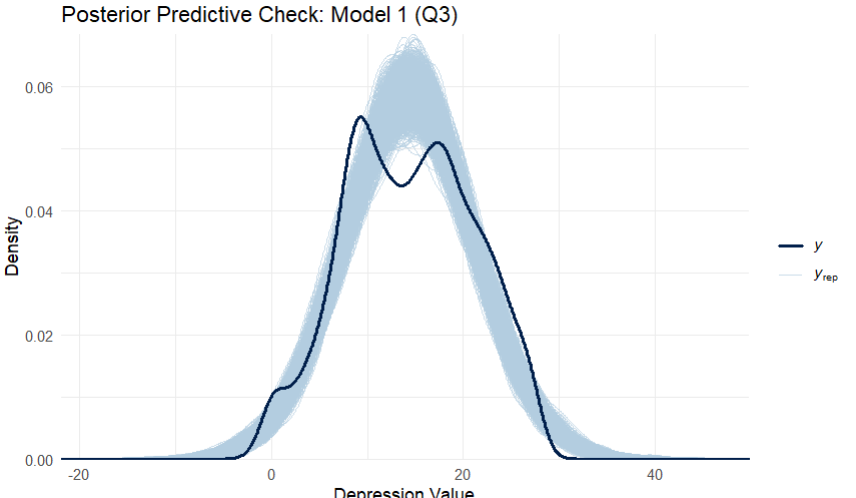
**המודלים שנבנו כוללים את הפרמטרים הבאים:**

* **B0 - חותך:** מייצג את ערך הבסיס של רמת הדיכאון כאשר יתר המשתנים הם באפס.
* **B\_Age:** השפעת הגיל על רמת הדיכאון.
* **B\_CGPA:** השפעת ממוצע הציונים על רמת הדיכאון.
* **B\_Year:** השפעת שנת הלימודים על רמת הדיכאון.
* **B\_Anxiety\_Value:** השפעת רמת החרדה על רמת הדיכאון (במודל השני בלבד).
* **B\_Stress\_Value:** השפעת רמת הלחץ על רמת הדיכאון (במודל השני בלבד).
* **- σסטיית תקן:** מייצג את השונות של רמות הדיכאון המנובאות.

**הצדקת הפריורים:**

1. **B0 ∼ N(10, 3)** **-** החותך נבחר עם תוחלת של 10 וסטיית תקן של 3, מה שמשקף את ההנחה שסטודנטים חווים רמת דיכאון בסיסית מסוימת, אך לא גבוהה מדי.
2. **- B\_Age ∼ N(5, 2)** פרמטר זה מייצג את השפעת הגיל על רמת הדיכאון, עם ההנחה שגיל מבוגר יותר עשוי להיות קשור לרמות דיכאון גבוהות יותר.
3. **- B\_CGPA ∼ N(-5, 2)** פרמטר זה נבחר עם תוחלת שלילית של -5, המשקפת את ההנחה שממוצע ציונים גבוה יותר קשור לרמות דיכאון נמוכות יותר.
4. **- B\_Year ∼ N(5, 2)** פרמטר זה מייצג את השפעת שנת הלימודים על רמת הדיכאון, עם ההנחה שסטודנטים בשנים מתקדמות עשויים לחוות יותר לחץ ודיכאון.
5. **B\_Anxiety\_Value ∼ N(5, 2) ו- B\_Stress\_Value ∼ N(5, 2) -** פרמטרים אלו משקפים את ההנחה שחרדה ולחץ גבוהים תורמים לרמות דיכאון גבוהות יותר.
6. **σ ∼ Exponential(1)** **-** סטיית התקן נבחרה בפריור אקספוננציאלי כדי לתת למודל גמישות רחבה בהערכת השונות בנתונים.

**בדיקת רגישות:**

שינינו את הפריורים לפריורים לא מוטים, והתוצאות נותרו דומות. גיל, ממוצע ציונים, שנת לימודים, לחץ וחרדה המשיכו להשפיע באופן דומה על רמות הדיכאון. זה מחזק את אמינות המודל תחת הנחות שונות. המודל הכולל מדדי חרדה ולחץ הראה שיפור משמעותי בניבוי. תמונה שמכילה טקסט, עלילה, תרשים, צילום מסך

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, צילום מסך, תרשים, עלילה

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, תרשים, צילום מסך, עלילה

התיאור נוצר באופן אוטומטי

**Posterior Predictive Check: Model 1 (Q3)**:

* הגרף משווה בין הערכים המדווחים בפועל של רמות הדיכאון (קו כחול כהה) לבין הערכים המדומים על פי המודל הראשון (צללים כחולים בהירים). המודל מצליח לתפוס את המבנה הכללי של הנתונים, אך יש פיזור רחב יותר בקצוות ההתפלגות, מה שעשוי להעיד על כך שהמודל פחות מדויק בערכים הקיצוניים של הדיכאון.

: **Posterior Predictive Check: Model 2 (Q3)**

* הגרף דומה לקודם, אך עבור המודל השני, הכולל את משתני הלחץ והחרדה. המודל השני מדויק יותר בתפיסת ההתפלגות הכללית, אך גם כאן יש סטיות קלות בקצוות ההתפלגות של הדיכאון.

**:Posterior Predictive Mean Comparison: Model 1 (Q3)**

* הגרף משווה בין ממוצע הערכים המדווחים בפועל לבין ממוצע הערכים המדומים במודל הראשון. התפלגות הממוצעים יחסית צרה וממוקדת סביב הממוצע האמיתי, אך ישנה סטייה קלה שמראה שהמודל מצליח לנבא את הממוצע הכללי, אם כי יש מקום לשיפור.

:**Posterior Predictive Mean Comparison: Model 2 (Q3)**

* הגרף מציג את אותו ניתוח עבור מודל 2, הכולל את משתני הלחץ והחרדה. התפלגות הממוצעים דומה מאוד לזו של מודל 1, אך המודל השני מראה התאמה מעט טובה יותר לממוצע האמיתי, עם התפלגות צרה יותר.

**סיכום כללי לשאלה 3:**

* ההשוואה בין המודלים מראה שמודל 2, הכולל את משתני הלחץ והחרדה, מצליח לתפוס בצורה מדויקת יותר את המבנה הכללי של הנתונים ואת ממוצע הדיכאון בקרב הסטודנטים, בהשוואה למודל הראשון. עם זאת, בשני המודלים יש סטיות בקצוות ההתפלגות, מה שמצביע על שונות גדולה יותר בקבוצות מסוימות באוכלוסייה. שילוב משתנים כמו לחץ וחרדה משפר את התאמת המודל**.**

**דיון כללי:**

* המחקר מצביע על כך שקיימים קשרים בין מאפיינים דמוגרפיים ואקדמיים לרמות הלחץ, החרדה והדיכאון בקרב סטודנטים. נמצא שגיל מבוגר ושנות לימודים מתקדמות מעלים את רמות הלחץ והחרדה, בעוד שסטודנטים להנדסה חווים פחות לחץ וחרדה בהשוואה לאחרים. בנוסף, נשים נוטות לדווח על רמות דיכאון גבוהות יותר מגברים, מה שמדגיש את החשיבות של התייחסות למגדר בניבוי דיכאון. שילוב משתני לחץ וחרדה במודל שיפר את הדיוק בניבוי רמות הדיכאון. המודלים הבייסיאניים שהשתמשנו בהם הוכיחו את עצמם ככלים יעילים לזיהוי דפוסים ולהבנת השפעות מורכבות.

מגישים: אליהו אלימלך - 207975699 & שגב כהן - 206947400