



خزائن للعلوم الاقتصادية والإدارية  
KHAZAYIN OF ECONOMIC AND  
ADMINISTRATIVE SCIENCES

ISSN: 2960-1363 (Print)

ISSN: 3007-9020 (Online)



## Using Logistic Regression and Bayesian Regression in The analysis of Chronic Disease Data

Prof. Dr. Haitham Yaqoub Youssef      Researcher Enaam Haitham Yaqoub

<sup>1</sup> College of Administration and Economics, University of ERMOOK, Iraq.

<sup>2</sup> College of Administration and Economics, University of Diyala, Iraq.

[haitham@uodiyala.edu.iq](mailto:haitham@uodiyala.edu.iq)

**ABSTRACT:** This study aims to evaluate the effectiveness of Bayesian modeling in analyzing large-scale medical data by employing the Bayesian logistic regression model and comparing its results with those of the classical logistic regression model. The research relied on a dataset of 50,000 public health records containing demographic and medical information of patients, including those suffering from chronic diseases such as diabetes and heart disease. The dependent variable (Y) represented the presence or absence of a chronic disease, while the independent variables included age, body mass index (BMI), blood pressure, and blood sugar.

The results revealed that the Bayesian model achieved a classification accuracy of 86.7%, compared to 63.2% for the classical model. Furthermore, the Bayesian model recorded a higher AUC value (0.91) versus 0.87 for the classical approach, highlighting the superiority of Bayesian inference in handling large datasets and expressing uncertainty through probability distributions. These findings emphasize that Bayesian modeling offers a more flexible and accurate statistical framework compared to traditional methods, making it an effective tool in supporting future medical decision-making.

**Keywords:** Bayesian regression, logistic regression, medical data, chronic diseases, AUC, R.

DOI: [10.69938/Keas.25020311](https://doi.org/10.69938/Keas.25020311)

### استخدام الانحدار اللوجستي والانحدار البيزي في تحليل بيانات الأمراض المزمنة

الباحثة إنعام هيثم يعقوب  
كلية الادارة والاقتصاد , جامعة ديالى , العراق

أ.د. هيثم يعقوب يوسف  
كلية اليرموك الجامعة , ديالى , العراق  
[haitham@uodiyala.edu.iq](mailto:haitham@uodiyala.edu.iq)

#### المستخلص

تهدف هذه الدراسة إلى تقييم فعالية النمذجة البيزية في تحليل البيانات الطبية الضخمة من خلال استخدام نموذج الانحدار اللوجستي البيزي لمقارنة نتائجه مع نموذج الانحدار اللوجستي الكلاسيكي. اعتمد البحث على بيانات صحية عامة شملت (50,000) سجل طبي يتضمن معلومات ديموغرافية وطبية للمرضى، من بينهم مصابون بأمراض مزمنة مثل داء السكري وأمراض القلب. تم استخدام المتغير التابع (Y) لتمثيل وجود أو عدم وجود المرض المزمن، فيما شملت المتغيرات المستقلة العمر، مؤشر كتلة الجسم (BMI)، ضغط الدم، وسكر الدم.

أظهرت نتائج التحليل باستخدام النموذج البيزي دقة تصنيف بلغت (86.7%) مقارنة بـ (63.2%) للنموذج الكلاسيكي، كما حقق النموذج البيزي قيمة أعلى لمعيار (AUC=0.91) مقابل (0.87) في النموذج الكلاسيكي، مما يعكس تفوق المنهج البيزي في التعامل مع البيانات الضخمة والقدرة على التعبير عن عدم اليقين من خلال التوزيعات الاحتمالية. وتشير النتائج إلى أن النموذج البيزي يوفر إطاراً إحصائياً أكثر مرونة ودقة مقارنة بالنهج التقليدي، مما يعزز من فعاليته في دعم القرارات الطبية المستقبلية.

**الكلمات المفتاحية:** الانحدار البيزي، الانحدار اللوجستي، البيانات الطبية، الأمراض المزمنة، AUC، R.

## 1. المقدمة

شهدت العقود الأخيرة تطوراً هائلاً في حجم وتنوع البيانات الناتجة عن الأنشطة الإنسانية، خاصة في المجالات الصحية، حيث أصبحت السجلات الطبية الإلكترونية (Electronic Health Records) تمثل مصدراً غنياً للمعلومات التشخيصية والديموغرافية والسريرية. هذا التوسع في البيانات وضع الإحصاء التطبيقي أمام تحديات جديدة تتطلب أساليب تحليلية أكثر مرونة وقدرة على التعامل مع البيانات غير المتجانسة والمعقدة (He & Chen, 2021). في هذا السياق، برزت النمذجة البيزية كأحد أهم الأساليب الحديثة في التحليل الإحصائي، إذ تتيح دمج المعرفة السابقة (Prior Information) مع البيانات الحالية للوصول إلى استنتاجات احتمالية أكثر دقة وموضوعية (Gelman et al., 2013). بخلاف النماذج الكلاسيكية التي تقتصر على تقديرات نقطية، يوفر الانحدار البيزي توزيعات احتمالية كاملة للمعلمات، وهو ما يعزز من قدرة الباحثين على تمثيل عدم اليقين في التقدير، والتعامل مع القيم المتطرفة والفجوات في البيانات (Carpenter et al., 2017). أما الانحدار اللوجستي الكلاسيكي، فلا يزال يحظى بمكانة واسعة في التطبيقات الطبية كونه يتيح تقدير احتمالية حدوث مرض معين بناءً على مجموعة من المتغيرات التفسيرية. غير أن محدوديته في التعامل مع الضوضاء وضعف مرونته في استيعاب البيانات الضخمة جعلها الحاجة ملحة إلى مقارنته بالنهج البيزي الأكثر تطوراً (Hosmer et al., 2013). لقد تناولت العديد من الدراسات أهمية إدخال النماذج البيزية في التحليل الطبي. على سبيل المثال، أظهرت دراسة الزهراني (2021) أن النماذج البيزية في التنبؤ بالأمراض المزمنة باستخدام السجلات الصحية في المملكة العربية السعودية حققت دقة أعلى مقارنة بالنماذج الكلاسيكية. كما أشار الخطيب والشيخ (2020) في دراستهما حول السلوك الاستهلاكي الإلكتروني إلى أن الأساليب البيزية توفر تفسيراً أكثر عمقاً عند التعامل مع بيانات كبيرة ومتنوعة. وفي السياق العراقي، طبق العبيدي (2022) نماذج الانحدار اللوجستي البيزي على بيانات انتخابية، مؤكداً كفاءتها في التنبؤ بنتائج معقدة متعددة العوامل. وعلى الصعيد العالمي، توصل (Ahmadi & Javaheri, 2022) إلى أن استخدام الانحدار البيزي في تحليل البيانات الطبية الضخمة أسهم في تحسين دقة التشخيص المبكر للأمراض القلبية، فيما أظهر (Neal, 2011) أن خوارزميات الاستدلال العددي مثل Hamiltonian Monte Carlo توفر حلولاً عملية وفعالة لتقدير النماذج في فضاءات متعددة الأبعاد. كذلك، أكد (Carpenter et al., 2017) على أهمية استخدام منصات برمجية متقدمة مثل Stan لتبسيط عملية تنفيذ النماذج البيزية. ورغم تنوع البيانات التطبيقية، تتفق معظم الدراسات السابقة سواء العربية أو الأجنبية، على أن النهج البيزي يقدم مرونة وقدرة أعلى على التعامل مع الخصائص المعقدة للبيانات الضخمة، مقارنة بالأساليب الكلاسيكية. وتبين أن التوزيعات القبلية المناسبة وتقنيات الاستدلال مثل MCMC و HMC تلعب دوراً جوهرياً في تحسين دقة النماذج.

بناءً على ما تقدم، يتضح أن تطبيق الانحدار البيزي إلى جانب الانحدار اللوجستي الكلاسيكي في تحليل بيانات الأمراض المزمنة (مثل السكري وأمراض القلب) لا يقتصر على المقارنة بين منهجين إحصائيين، بل يتعدى ذلك إلى تقديم إطار تطبيقي يعكس كيف يمكن للمنهجيات الحديثة أن تدعم القرارات الطبية المبنية على البيانات. ومن هنا تنبع أهمية هذا البحث الذي يهدف إلى استكشاف كفاءة الانحدار البيزي في مواجهة التحديات التطبيقية، وإبراز أوجه القوة والقصور عند مقارنته بالانحدار اللوجستي التقليدي، باستخدام بيانات طبية عامة واسعة النطاق.

## 2. منهجية البحث

اعتمد هذا البحث على المنهج الوصفي التحليلي بهدف دراسة العلاقة بين بعض المتغيرات السريرية والديموغرافية واحتمالية الإصابة بمرضين مزمنين هما السكري من النوع الثاني وأمراض القلب. وقد تم استخدام طريقتين للتحليل: الانحدار اللوجستي التقليدي والانحدار البيزي، من أجل مقارنة الأداء الإحصائي للنموذجين.

### 2.1 البيانات المستخدمة

اعتمد البحث على بيانات صحية عامة تتضمن (50,000) سجل طبي لمرضى من أعمار مختلفة، تشمل معلومات ديموغرافية وطبية. المتغير التابع (Y) ثنائي القيمة (0 = لا يوجد مرض مزمن، 1 = يوجد مرض مزمن)، فيما شملت المتغيرات المستقلة:

- العمر (Age).
- مؤشر كتلة الجسم (BMI).
- ضغط الدم (Blood Pressure).
- سكر الدم (Blood Sugar).
- الجنس (Gender).

و.الجدول (1) يبين المتغيرات العشوائية الديموغرافية والسريرية المستخدمة في البحث.

جدول 1: المتغيرات العشوائية المستخدمة في البحث

| المتغير | النوع | الوصف                                                  |
|---------|-------|--------------------------------------------------------|
| Y       | تابع  | متغير ثنائي (0 : لا يوجد مرض مزمن , 1 : يوجد مرض مزمن) |
| X1      | مستقل | العمر                                                  |
| X2      | مستقل | مؤشر كتلة الجسم (BMI)                                  |

|    |       |                    |
|----|-------|--------------------|
| X3 | مستقل | ضغط الدم الانقباضي |
| X4 | مستقل | سكر الدم الصيامي   |
| X5 | مستقل | الجنس (ذكر / انثى) |

والجدول (2) يبين البيانات الوصفية للمتغيرات العشوائية:

جدول 2 : الخصائص الوصفية للمتغيرات العشوائية لعينة البحث

| المتغير            | المتوسط | الانحراف المعياري | الحد الأدنى | الحد الأقصى |
|--------------------|---------|-------------------|-------------|-------------|
| العمر ( $X_1$ )    | 52.3    | 11.6              | 18          | 89          |
| BMI ( $X_2$ )      | 27.8    | 5.4               | 16.0        | 45.5        |
| ضغط الدم ( $X_3$ ) | 128.4   | 14.9              | 90          | 180         |
| سكر الدم ( $X_4$ ) | 108.1   | 23.6              | 60          | 240         |

## 2.2 تجهيز البيانات

قبل البدء بالنمذجة، خضعت البيانات إلى سلسلة من خطوات المعالجة الإحصائية:

- معالجة القيم الناقصة باستخدام أسلوب التقدير (Mean/Mode Imputation).
- التحقق من القيم المتطرفة عبر اختبارات IQR والتأكد من عدم تأثيرها الكبير على النماذج.

## 2.3 الأدوات الإحصائية

تم استخدام برنامج R (إصدار 4.3) مع مكتبة rstanarm لتنفيذ النموذج البيزي، بينما تم اعتماد برنامج EViews-10 لتقدير النموذج الكلاسيكي، ضمناً لدقة المقارنة.

## 2.4 نموذج الانحدار اللوجستي الكلاسيكي

تمثل دالة اللوغيت (Logit Function)، حيث تم أولاً بناء نموذج انحدار لوجستي تقليدي باستخدام طريقة الاحتمالية العظمى .  
(Maximum Likelihood Estimation), MLE ويُعبر عن النموذج بالمعادلة:

$$\text{Logit}(P(Y=1 | X)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

حيث :-

•  $\text{Logit}(P) = \ln(P / (1 - P))$  : وهو التحويل اللوغاريتمي لنسبة النجاح إلى الفشل.

•  $Y$  : المتغير التابع (وجود المرض =1، عدم وجود المرض = 0) .

•  $X_1, X_2, \dots, X_k$  : المتغيرات المستقلة (مثل العمر، ضغط الدم، سكر الدم....).

•  $\beta_0$  : الثابت.

•  $\beta_1, \dots, \beta_k$  : معاملات الانحدار التي تعكس تأثير كل متغير على احتمال الإصابة

ولغرض توضيح الأسس الرياضية وآلية التقدير بشكل أعمق، يمكن تناول النموذج اللوجستي التقليدي من زاوية نظرية أكثر تفصيلاً كما يأتي:

رغم شيوع الانحدار الخطي في التطبيقات الإحصائية، إلا أنه غير مناسب للتعامل مع المتغيرات التابعة الثنائية مثل حالة الإصابة بالمرض (نعم/لا). لذلك يُعتمد الانحدار اللوجستي باعتباره الأسلوب الأمثل لتمثيل الاحتمالية ضمن المجال  $[0,1]$ . الفكرة الرئيسة تقوم على استخدام لوغاريتم الأرجحية (Log-Odds) بدلاً من الاحتمالية المباشرة، بحيث يصبح ارتباط المتغيرات المستقلة بالمتغير التابع أكثر واقعية من الناحية الرياضية.

ويُعرّف لوغاريتم الأرجحية كما يلي:

$$\text{Logit}(p) = \ln \left( \frac{p}{1-p} \right)$$

وبذلك يمكن صياغة النموذج:

$$\ln \left( \frac{P(Y=1 | X)}{1 - P(Y=1 | X)} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

ويمكن العودة إلى صيغة الاحتمال:

$$P(Y=1 | X) = 1 / [ 1 + \exp(-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)) ]$$

تُقدَّر معاملات النموذج باستخدام طريقة الاحتمالية العظمى (Maximum Likelihood Estimation) MLE ، التي تبحث عن القيم التي تجعل البيانات المرصودة أكثر احتمالاً. وتُعتبر هذه الطريقة فعالة في العينات الكبيرة، لكنها تواجه قيوداً عملية أهمها:

- ضعف الاستقرار عند العينات الصغيرة.
  - مشكلة التعدد الخطي بين المتغيرات المستقلة.
  - محدودية المرونة في التعامل مع القيم المتطرفة والبيانات الضخمة ذات الأبعاد العالية.
- ورغم ذلك، يبقى الانحدار اللوجستي التقليدي أداة قياسية مهمة في الأبحاث الطبية والإحصائية لما يتميز به من بساطة في التنفيذ وسهولة في التفسير.

## 2.5 نموذج الانحدار اللوجستي البيزي

بعد ذلك تم بناء نموذج انحدار لوجستي بيبي باستخدام بيئة R/Stan ، وذلك بتعريف توزيعات قبلية (Priors) مناسبة، واستعمال خوارزمية Hamiltonian Monte Carlo (HMC) لتقدير المعلمات. الصياغة البيزية للانحدار اللوجستي تبدأ من العلاقة الاحتمالية:

$$P(Y=1|X, \beta) = 1 / (1 + \exp(-(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)))$$

حيث إن رموز المتغيرات والمعاملات هي ذاتها المعرفة سابقاً في معادلة الانحدار اللوجستي الكلاسيكي

ولغرض توضيح الأسس الرياضية وآلية التقدير بشكل أعمق، يمكن تناول النموذج اللوجستي التقليدي من زاوية نظرية أكثر تفصيلاً كما يأتي: على عكس التقدير الكلاسيكي، يقوم الانحدار اللوجستي البيبي على دمج المعلومات السابقة (Priors) مع البيانات الحالية من خلال مبرهنة بايز، مما يؤدي إلى تكوين توزيع بعدي (Posterior Distribution) يعكس عدم اليقين بدرجة أكبر. وتتمثل هذه الصياغة في:

$$P(\beta | D) \propto P(D | \beta) \cdot P(\beta)$$

هذه تُعبر أن التوزيع البعدي للمعلمات (  $\beta$  ) يتناسب مع حاصل ضرب دالة الاحتمالية بالقبلية

حيث تمثل  $P(\beta)$  التوزيعات القبلية للمعلمات، و  $P(D | \beta)$  دالة الاحتمالية المستخلصة من العينة، و  $P(\beta | D)$  هو التوزيع البعدي الذي يدمج بينهما.

الميزة الأساسية في هذه المقاربة هي أن معاملات النموذج لا يتم التعبير عنها كتقديرات نقطية فقط، وإنما كتوزيعات احتمالية، تتضمن الوسط الحسابي والانحراف المعياري والفواصل الاحتمالية (Credible Intervals). وبذلك يمكن للباحث أن يحصل على صورة أوضح لمدى عدم اليقين في التقديرات، وهو أمر بالغ الأهمية في تحليل البيانات الطبية.

ولغرض التقدير، استُخدمت خوارزمية Hamiltonian Monte Carlo (HMC) لتوليد 4000 عينة بعد فترة حرق (Burn-in) بلغت 1000 عينة (Carpenter et al., 2017; Neal, 2011).

من الناحية التطبيقية، يوفر الانحدار البيبي مزايا عملية مهمة، منها:

- تحسين دقة التقديرات في حال العينات الصغيرة أو غير المتوازنة.
- إمكانية دمج مصادر متعددة للمعلومات، بما في ذلك الخبرة السريرية أو نتائج الدراسات السابقة.
- مرونة أكبر في التعامل مع القيم المتطرفة وتباين التوزيعات، خصوصاً مع البيانات الطبية الكبيرة والمعقدة.

## 2.6 إجراءات التقدير والتحقق

- جرى التحقق من جودة النموذج عبر مؤشر  $\hat{R}$  (R-hat) ، حيث اعتُبر النموذج مقارباً إذا كان  $\hat{R} < 1.01$ .
- تمت مقارنة دقة التصنيف (Accuracy) ومنحنى ROC (AUC) بين النموذج البيبي والنموذج الكلاسيكي باستخدام عينة الاختبار

## 2.7 خطوات تنفيذ النموذج

تم اتباع سلسلة من الخطوات المنهجية الدقيقة لتنفيذ نموذج الانحدار البيبي على البيانات الطبية الضخمة، وذلك كما يلي:

- تنظيف البيانات
- حذف القيم المفقودة والبيانات غير المنطقية لضمان جودة التحليل.
- تحليل استكشافي للبيانات
- توصيف المتغيرات وتحليل توزيعاتها للكشف عن الأنماط والقيم الشاذة.
- تحديد التوزيعات القبلية المناسبة
- اختيار التوزيعات المسبقة بناءً على أدبيات علمية أو معرفة ميدانية متخصصة.
- بناء النموذج البيبي باستخدام خوارزميات MCMC
- تنفيذ النموذج في بيئة Stan باستخدام لغة R لتقدير المعلمات الاحتمالية.
- تقييم أداء النموذج

عبر مؤشرات مثل منحنى ROC (AUC) ، دوال الخسارة، واختبارات التشخيص الإحصائي

• المقارنة مع نموذج انحدار كلاسيكي (Logistic OLS) :

لبناء مرجعية تقييمية تحدد مدى تفوق النموذج البيزي.

### 3. التحليل والنتائج

#### 3.1 المقدمة

يهدف هذا القسم إلى عرض نتائج النمذجة باستخدام كل من الانحدار اللوجستي الكلاسيكي و الانحدار البيزي، ومن ثم إجراء مقارنة كمية ونوعية بين النموذجين لتقييم مدى دقة وموثوقية التقديرات.

تم استخدام مكتبة rstanarm في لغة R لبناء نموذج انحدار لوجستي بيبي، وفق التوزيعات القبلية التالية :

$$\beta_j \sim N(0, 10^2) \text{ لجميع المعاملات.}$$

#### 3.2 النتائج الرئيسية للنموذج الكلاسيكي

جدول 3 : نتائج التقدير البعدي للنموذج البيبي

| المتغير                | التقدير<br>(Posterior Mean) | الانحراف المعياري | حدود الثقة     |
|------------------------|-----------------------------|-------------------|----------------|
| المعتزض ( $\beta_0$ )  | -2.14                       | 0.21              | [-2.56, -1.74] |
| العمر ( $\beta_1$ )    | 0.032                       | 0.004             | [0.024, 0.041] |
| BMI ( $\beta_2$ )      | 0.058                       | 0.009             | [0.041, 0.076] |
| ضغط الدم ( $\beta_3$ ) | 0.015                       | 0.003             | [0.010, 0.021] |
| السكر ( $\beta_4$ )    | 0.018                       | 0.005             | [0.009, 0.027] |
| الجنس ( $\beta_5$ )    | 0.41                        | 0.06              | [0.30, 0.52]   |

أظهرت نتائج الانحدار اللوجستي التقليدي أن معظم المتغيرات التفسيرية كانت دالة إحصائياً عند مستوى (0.05). وقد اتضح أن:

- العمر ومؤشر كتلة الجسم BMI و ضغط الدم و سكر الدم يرتبطون بشكل إيجابي مع احتمالية الإصابة بمرض مزمن.
- متغير الجنس أظهر فرقاً معنوياً بين الذكور والإناث لصالح ارتفاع احتمالية الإصابة عند الذكور.

لكن المخرجات أظهرت أن الفواصل حدود الثقة (Confidence Intervals) كانت واسعة نسبياً، مما يشير إلى محدودية الدقة في التقديرات. كما بلغت دقة التصنيف الكلية للنموذج حوالي 63.2%، وقيمة المساحة تحت المنحنى AUC وصلت إلى 0.87.

#### 3.3 النتائج الرئيسية للنموذج البيبي

أما بالنسبة للانحدار البيبي، فقد قدم النموذج توزيعات بعدية (Posterior Distributions) أكثر استقراراً للمعاملات. وُجد أن:

- جميع المتغيرات المدخلة احتفظت باتجاه تأثير مشابه للنموذج الكلاسيكي (ارتباط موجب مع احتمالية الإصابة).
- إلا أن الفواصل الاحتمالية (Credible Intervals) كانت أضيق من فواصل الثقة الكلاسيكية، ما يعكس زيادة دقة التقديرات.

من حيث دقة الأداء:

- بلغت دقة التصنيف 86.7% (Accuracy)، وهي أعلى بكثير من النموذج الكلاسيكي.
- بلغت قيمة  $AUC = 0.91$ ، ما يدل على قدرة عالية على التمييز بين المرضى المصابين وغير المصابين.

أظهرت التقديرات البعدية للمتغيرات المستقلة دلالة إحصائية واضحة كما هو مبين في الجدول (4) لاحقاً

#### 3.4 المقارنة التحليلية بين النموذجين

ولأغراض المقارنة بين المنهج البيبي واللوجستي التقليدي، يوضح الجدول (4) أبرز الفروق في الأداء بين النموذجين

جدول 4: مقارنة الأداء بين النموذج البيبي والنموذج الكلاسيكي

| المعيار                  | النموذج البيبي   | النموذج الكلاسيكي |
|--------------------------|------------------|-------------------|
| دقة التصنيف (Accuracy)   | 86.7%            | 63.2%             |
| AUC (منحني ROC)          | 0.91             | 0.87              |
| التفسير الاحصائي         | توزيعات احتمالية | نقطة تقدير فقط    |
| تضمين المعلومات القبلية  | نعم              | لا                |
| مرونة التعامل مع الضوضاء | مرتفعة           | منخفضة            |

#### تحليل النتائج:

- يظهر بوضوح أن النموذج البيبي أكثر كفاءة في التعامل مع البيانات الطبية الضخمة، وذلك بفضل استخدامه للتوزيعات القبلية وقدرته على التعبير عن عدم اليقين بشكل كامل.
- في المقابل، يبقى النموذج الكلاسيكي مناسباً في حال العينات الصغيرة أو عند الرغبة في تفسير بسيط وسريع، لكنه أقل دقة ومرونة.



• النتائج تدعم توجه نحو تبني النمذجة البيزية خصوصاً في المجالات الطبية، حيث تُعد دقة التنبؤ عاملاً أساسياً لاتخاذ القرارات السريرية.

### 3.5 التحقق من النموذج (Model Diagnostics)

• أظهرت مؤشرات R-hat لجميع المعاملات أنها أقل من (1.01)، مما يؤكد تقارب السلاسل المولدة واستقرار التقدير.  
• أُعيد اختبار النموذج البيزي باستخدام عينة مستقلة (Test Set)، وحقق دقة تصنيف بلغت 87.1%، ما يعزز قوة النتائج ويؤكد موثوقيتها

### 3.6 المناقشة (Discussion)

أظهرت نتائج هذا البحث أنّ النموذج البيزي قدّم أداءً متفوقاً على النموذج الكلاسيكي (اللوجستي) في تحليل السجلات الطبية الكبيرة. إذ تجاوزت دقته في التصنيف (86.7%)، بينما لم تتعدّ دقة النموذج الكلاسيكي (63.2%). كما سجّل النموذج البيزي قيمة أعلى لمنحنى AUC بلغت (0.91) مقارنة بـ (0.87) في النموذج الكلاسيكي، ما يعكس قدرة الأسلوب البيزي على تمثيل العلاقات المعقدة بين المتغيرات الطبية والتعامل مع القيم المتطرفة بكفاءة عالية (Gelman et al., 2013; Neal, 2011). ويُعزى هذا التفوق إلى عدة عوامل رئيسية:

1. تضمين المعلومات القبلية (Priors) إذ ساعد إدراج المعرفة السابقة في استقرار التقديرات الإحصائية، خصوصاً مع وجود متغيرات متداخلة مثل ضغط الدم وسكر الدم. (Carpenter et al., 2017)
2. التعبير عن عدم اليقين: على خلاف الانحدار الكلاسيكي الذي يقدّم تقديرات نقطية فقط، قدّم النموذج البيزي توزيعات احتمالية كاملة للمعاملات، وهو ما أتاح تفسيراً أوسع وأكثر واقعية للنتائج. (Kruschke, 2015)
3. التعامل مع البيانات الضخمة: رغم أنّ التحليل ركّز على أمراض مزمنة محددة (السكري وأمراض القلب)، فإن خصائص المنهج البيزي أظهرت ملاءمة واضحة لمعالجة أحجام كبيرة من السجلات الطبية، وهو ما يتماشى مع نتائج دراسات مشابهة في المجال الصحي (Zahrani, 2021)؛ (Ahmadi & Javaheri, 2022).

### 3.7 مقارنة النتائج مع الدراسات السابقة

تتفق نتائج هذا البحث مع ما توصلت إليه الزهراني (2021)، الذي أكّد على تفوق النماذج البيزية في التنبؤ بالأمراض المزمنة باستخدام السجلات الصحية الضخمة في المملكة العربية السعودية. فقد أظهرت تلك الدراسة أنّ إدراج التوزيعات القبلية مكّن من بناء نموذج أكثر دقة ومرونة، وهو ما يتقاطع مع ما تم إثباته في هذه الدراسة.

كما دعمت دراسة العبيدي (2022) النتائج الحالية، إذ استخدم نموذج الانحدار اللوجستي البيزي في تحليل بيانات انتخابية عراقية، وأثبت قدرة المنهج البيزي على التعامل مع البيانات ذات الأبعاد العالية والتباين الكبير بين الملاحظات. وعلى الرغم من اختلاف موضوع التطبيق، فإن المنهجية الإحصائية أكدت فائدتها في سياقات متنوعة.

من جهة أخرى، أشار (Ahmadi & Javaheri (2022 إلى أن النماذج البيزية أكثر قدرة على استيعاب التعقيدات الكامنة في بيانات الرعاية الصحية، خاصة مع إدخال متغيرات سريرية متعددة مثل ضغط الدم ومستويات السكر في الدم. تتقاطع هذه النتائج بشكل مباشر مع ما خرج به هذا البحث.

أما على المستوى الدولي، فقد أوضحت دراسة (Wang et al. (2021 أنّ ارتفاع مؤشر كتلة الجسم BMI يعد أحد أقوى المؤشرات المرتبطة بزيادة احتمالية الإصابة بالسكري من النوع الثاني، وهو ما أكدته نتائجنا بعد ثبوت دلالة هذا المتغير في النموذج البيزي.

كما أنّ ما ورد في تقرير منظمة الصحة العالمية (WHO, 2020) بشأن الترابط الوثيق بين العوامل الديموغرافية (مثل العمر والجنس) والمؤشرات الطبية (ضغط الدم وسكر الدم) والإصابة بالأمراض المزمنة، جاء منسجماً مع التقديرات التي قدّمها النموذج البيزي في هذه الدراسة.

### 3.8 أهمية النتائج الحالية

تدل هذه النتائج على أنّ استخدام النماذج البيزية لا يقتصر على تعزيز دقة التنبؤ فحسب، بل يمتد أيضاً إلى تحسين موثوقية النتائج في ظل ظروف البيانات غير المتجانسة. كما أن النموذج البيزي أتاح إمكانية دمج المعرفة السابقة مع الأدلة المستخلصة من السجلات الطبية، ما يجعل التوصيات المستندة إلى نتائجه أكثر قوة وواقعية (Gelman et al., 2013).

### 4. الاستنتاجات والتوصيات

#### 4.1 الاستنتاجات ( Conclusion )

من خلال تطبيق كلّ من نموذج الانحدار اللوجستي التقليدي والانحدار البيزي على السجلات الطبية الخاصة بالأمراض المزمنة (السكري وأمراض القلب)، يمكن تلخيص الاستنتاجات الرئيسية بما يلي:

1. النماذج البيزية توفر إطاراً أكثر مرونة ودقة مقارنة بالنماذج الكلاسيكية.
2. يمكن للنمذجة البيزية التعامل مع القيم المتطرفة والفجوات في البيانات بكفاءة أعلى.
3. الدمج بين المعرفة السابقة والبيانات الكبيرة يعزز من التفسير الإحصائي.

4. على الرغم من التحديات الحسابية، إلا أن خوارزميات مثل HMC جعلت التطبيق العملي ممكناً.
- 4.2 التوصيات (Recommendations)  
استناداً إلى ما سبق، يقترح الباحثان مجموعة من التوصيات العملية والعلمية:
  - تشجيع توظيف النمذجة البيزية في الدراسات الطبية والمالية التي تتعامل مع بيانات ضخمة.
  - توفير تدريب متخصص للباحثين في استخدام حزم برمجية بيزية مثل Stan و PyMC.
  - التوسع في دمج النماذج البيزية مع تقنيات الذكاء الاصطناعي لدعم اتخاذ القرار الطبي.

#### 4.3 قائمة المراجع (References)

##### المراجع العربية :

1. الزهراني، عبد الرحمن. (2021). "توظيف الانحدار البيزي في التنبؤ بالأمراض المزمنة باستخدام البيانات الضخمة: دراسة تطبيقية على السجلات الصحية بالملكة".  
مجلة جامعة الملك سعود للعلوم، 33(4)، 55-72.
2. الخطيب، سامر، والشيخ، نضال. (2020) "تحليل سلوك المستهلك الإلكتروني باستخدام الانحدار البيزي والبيانات الكبيرة".  
مجلة العلوم الاقتصادية والإدارية، 36(1)، 112-129. جامعة اليرموك.
3. العبيدي، محمد. (2022) "تطبيق نماذج الانحدار اللوجستي البيزي على بيانات الانتخابات البرلمانية العراقية". أطروحة ماجستير غير منشورة، جامعة بغداد، كلية العلوم، قسم الاحصاء

##### المراجع الأجنبية :

4. Ali, A., and Noor, S. "Bayesian Regression under Multicollinearity: An Application to High-Dimensional Data." *Statistical Modeling*, vol. 22, no. 6, 2022, pp. 555–570.
5. Ahmed, K., Zhao, M., and Lee, C. "Bayesian Modeling for Macroeconomic Forecasting with Big Data." *Journal of Applied Econometrics*, vol. 39, no. 1, 2024, pp. 22–41.
6. Bhadra, A., Datta, J., Polson, N. G., and Willard, B. "The Horseshoe+ Estimator of Ultra-Sparse Signals." *Bayesian Analysis*, vol. 11, no. 4, 2016, pp. 1105–1131.
7. Brown, T., and Smith, L. "Big Data Meets Bayesian Inference: Challenges and Strategies." *Statistical Science Advances*, vol. 11, no. 2, 2023, pp. 189–203.
8. Carpenter, Bob, et al. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software*, vol. 76, no. 1, 2017, pp. 1–32. <https://doi.org/10.18637/jss.v076.i01>.
9. Chen, W., and Zhao, L. "Real-Time Bayesian Modeling Using Streaming Data." *Applied Soft Computing*, vol. 141, 2024, article no. 110121.
10. Fernandez, R., and Martinez, A. "Prior Elicitation Techniques for Bayesian Analysis." *International Statistical Review*, vol. 90, no. 4, 2022, pp. 599–617. <https://doi.org/10.1111/insr.12496>
11. Gelman, A., et al. *Bayesian Data Analysis*. 3rd ed., Chapman and Hall/CRC, 2013.
12. Gelman, A., et al. "Advances in Bayesian Computation with Stan." *Bayesian Analysis*, vol. 18, no. 2, 2023, pp. 415–438.
13. Kitchin, R. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. SAGE Publications, 2014.
14. Kumar, N., and Patel, R. "Integrating Prior Knowledge in Big Data Models." *Big Data Research*, vol. 24, 2023, article no. 100342. <https://doi.org/10.1016/j.bdr.2023.100342>
15. Lee, S., and Choi, H. "Bayesian Logistic Regression in Imbalanced Big Data." *Data Science and Analytics Journal*, vol. 14, no. 3, 2023, pp. 312–328.
16. Lin, C., and Zhou, Y. "Bayesian Variable Selection in Large-Scale Regression." *Journal of Computational and Graphical Statistics*, vol. 32, no. 1, 2023, pp. 55–72.
17. Neal, Radford M. "MCMC Using Hamiltonian Dynamics." *Handbook of Markov Chain Monte Carlo*, edited by Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng, Chapman and Hall/CRC, 2011, pp. 113–162.
18. Nguyen, M., and Tran, D. "Scalable Variational Inference for Healthcare Applications." *Journal of Machine Learning Research*, vol. 25, no. 1, 2024, pp. 1001–1023. <http://jmlr.org/papers/v25/23-056.html>

19. Park, J., and Kim, H. "Hierarchical Bayesian Models for Big Data Analytics." *Computational Statistics & Data Analysis*, vol. 173, 2022, article no. 107531. <https://doi.org/10.1016/j.csda.2022.107531>
20. Polson, N. G., Scott, J. G., and Windle, J. "Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables." *Journal of the American Statistical Association*, vol. 108, no. 504, 2013, pp. 1339–1349.
21. Singh, A., and Roy, D. "Evaluation of Bayesian Regression in Financial Analytics." *Journal of Finance and Data Science*, vol. 8, no. 2, 2022, pp. 100–112. <https://doi.org/10.1016/j.jfds.2022.05.002>
22. Wang, T., Yu, L., and Zhang, X. "Scalable Bayesian Regression for Electronic Health Records." *Statistics in Medicine*, vol. 42, no. 4, 2023, pp. 551–568. <https://doi.org/10.1002/sim.9602>
23. Zhang, Y., and Liu, S. "Bayesian Methods for Consumer Behavior Analysis Using Big Data." *Journal of Business Research*, vol. 149, 2022, pp. 231–240. <https://doi.org/10.1016/j.jbusres.2022.05.003>
24. Zhou, Z., Chen, C., Li, Y., and Xu, J. "Bayesian Deep Learning for Real-Time Data Streams with Concept Drift." *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 1107–1116. <https://doi.org/10.1002/jae.2954>
25. Yousef, M., and Darwish, H. "Predictive Modeling in Healthcare with Bayesian Networks." *Health Informatics Journal*, vol. 29, no. 1, 2023, pp. 58–74. <https://doi.org/10.1177/14604582221132141>