



كلية التجارة

قسم الإحصاء التطبيقي والتأمين

تقدير نموذج خطأ القياس غير المتجانس في حالة الانحدار شبه المعلمي

**Heteroscedastic measurement error model estimation in
semiparametric regression**

أ.م.د/ هشام محمد المنجي

أستاذ الإحصاء المساعد

كلية التجارة _ جامعة المنصورة

أ.د/ فاطمة علي عبد العاطي

أستاذ الإحصاء التطبيقي

كلية التجارة _ جامعة المنصورة

الباحثة/ سلوى أحمد مجاهد

المستخلص:

في عملية قياس المتغيرات قد تتم عملية القياس بدرجة عالية من الدقة وبالتالي يتمكن الباحث من الحصول علي القياسات الصحيحة للمتغيرات وأحياناً قد يواجه الباحث صعوبة في إجراء عملية القياس فيؤدي ذلك إلي قياسات غير دقيقة للمتغيرات و بالتالي ينتج عن ذلك ما يعرف إحصائياً بخطأ القياس (measurement error) وقد يكون خطأ القياس في متغير واحد أو في كل المتغيرات بناءً علي درجة دقة القياس. وأبسط تعريف لخطأ القياس هو الفرق ما بين القيمة المقاسة للمتغير والقيمة الحقيقية لنفس المتغير (جابر، ٢٠١٨)؛ وينقسم إلى نوعين خطأ قياس متجانس وخطأ قياس غير متجانس. وفي حالة وضع فروض خطأ قياس لا تناسب طبيعة البيانات فسوف يؤثر ذلك على دقة التقدير وكفاءة الاستنتاج الإحصائي.

وفي هذا البحث سيتم التركيز على فرض خطأ القياس غير المتجانس ومدى ملائمة للبيانات وأيضاً تأثيره على نتائج التقدير في حالة استخدام طريقة شرائح B (B-splines). أما بالنسبة للجانب التطبيقي سيتم استخدام نموذج الانحدار شبه المعلمي (الانحدار اللوجستي) لبيانات حجمها (429) مريض كورونا المستجد (COVID-19) لمعرفة هل يوجد علاقة بين ارتفاع ضغط الدم والإصابة بالفيروس أم لا.

الكلمات الافتتاحية:

خطأ القياس، عدم التجانس، الانحدار شبه المعلمي، الانحدار اللوجستي، المتغيرات الخالية من الأخطاء، شرائح B.

Abstract:

In the process of measuring variables, the measurement process could be carried out with a high degree of accuracy and thus the researcher can obtain the correct measurements of the variables and sometimes the researcher may face difficulty in conducting the measurement process which leads to inaccurate measurement of the variables and the thus results in what is known statistically as the measurement error. The measurement error may be in one variable or in all variables based on the degree of accuracy of the measurement. The simplest definition of measurement error is the difference between the measured value of the variable and the real value of the same variable (جابر، ٢٠١٨) and it is divided into two types: homoscedastic measurement error and heteroscedastic measurement error. In the case of incorrect assumptions of the measurement error and inappropriate to the nature of the data, this will affect the accuracy of the estimate and the efficiency of the statistical conclusion.

In this research, the focus will be on the hypothesis of the heteroscedastic measurement error and it's suitability to the data, also it's impact on the estimation results in the case of using the B-splines method. On the practical side, a semi-parametric regression model (logistic regression) will be used for the data of (429) corona patients (covid-19) to find out whether there is a relationship between high blood pressure and infection with the virus or not.

Key words

Measurement error, heteroscedasticity, semiparametric regression, logistic regression, error free covariate, B-splines.

المقدمة:

يواجه الباحث العديد من العقبات عند عملية التحليل الإحصائي لأي مجموعة بيانات ومن أهمها مشكلة خطأ القياس والفروض المتعلقة بها. فخطأ القياس له العديد من الأسباب ومنها الخطأ في قراءة قيمة متغير ما بسبب خطأ أو عطل في الألة المستخدمة في عملية القياس أو بسبب خطأ من الباحث وكذلك قد تكون تكلفة قياس المتغيرات عالية أو موقع المتغيرات يصعب الوصول إليه فيصعب علي الباحث الحصول علي قياسات دقيقة (2017) Yi .

ونماذج خطأ القياس (أخطاء في المتغيرات) هي نماذج بديلة لنماذج الانحدار الكلاسيكية. حيث أن نماذج خطأ القياس تقترض أن قياس المتغيرات التفسيرية تخضع للخطأ وبالتالي يؤثر هذا الافتراض بشكل كبير على الاستنتاجات الإحصائية وتقديرات المعلمات التي يتم تقديرها (Cheng & Riu, 2006).

ولمعالجة خطأ القياس في المتغيرات سيتم الاعتماد على نموذج انحدار شبه معلمي ويكون الشكل الرياضي له معبراً عن العلاقة بين بعض المتغيرات التفسيرية والمتغير التابع معروفة ويوجد متغير واحد على الأقل تكون العلاقة بينه وبين المتغير التابع غير معروفة (علي، 2017) و نموذج الانحدار شبه المعلمي المستخدم هو نموذج الانحدار اللوجستي ويستخدم في شكله الأساسي لنمذجة المتغير الثنائي التابع بقيمتين محتملتين (طه & حسن، 2020) ويكون النموذج الإحصائي به مجموعة متغيرات تفسيرية مقياسة بدقة ومتغير تفسيري مقياس في وجود أخطاء وهو ما يؤثر على نتائج التحليل.

وتستخدم شرائح - B (B-splines) لتقدير دالة تباين خطأ القياس وقد عرفها كلا من Eilers & Marx (1996) على أنها تتكون من قطع متعددة الحدود عند قيم معينة من المتغير التفسيري وهي ما تعرف بالعقد (knots) وتكون متصلة بطريقة معينة. حيث أن اختيار العقد أمر صعب حيث إن اختيار عدد كبير من العقد يؤدي إلى المبالغة في (توفيق) ملائمة (overfitting) البيانات؛ بينما يؤدي اختيار عدد قليل من العقد إلى عدم (توفيق) ملائمة (underfitting) البيانات. والتركيز على فروض خطأ القياس ومعرفة الفرض الملائم لكل مجموعة بيانات يعد من أهم الخطوات في عملية التحليل الإحصائي. وسيفترض البحث أن أخطاء القياس غير متجانسة وهو فرض ملائم لمجموعة البيانات المستخدمة. وسيتم مقارنة نتائج التحليل تحت ثلاثة فروض لخطأ القياس لمعرفة تأثير الفروض غير الصحيحة على النتائج وكذلك تحديد الفرض الذي يحقق أفضل نتائج. والفروض هي:

- (1) فرض عدم وجود أخطاء في قياس البيانات (naive) وهو ما يعرف بالفرض البديهي.
- (2) فرض وجود أخطاء متجانسة في البيانات (homoscedasticity) وهو ما يعرف بفرض التجانس.

(3) وأخيراً فرض عدم تجانس أخطاء القياس في البيانات (heteroscedasticity) وهو ما يعرف بفرض عدم التجانس.

(1) الدراسات السابقة:

عرضت دراسة (Sarkar et al (2014) تقدير قوي (robust estimation) لدالة الانحدار في وجود أخطاء قياس غير متجانسة؛ وذلك باستخدام المنهج شبه المعلمي البييزي حيث يعتمد على مزيج من شرائح-B ومزيج من عمليات ديريتشلي (Dirichlet process). وتمت إجراء عملية المحاكاة لمقارنة الطريقة المقدمة بطريقة (Berry et al (2002) وفيها يستخدم مزيج من الشرائح المقطوعة متعددة الحدود لتمثيل دالة الانحدار وأكدت نتيجة المحاكاة أن الطريقة المقدمة تتفوق من حيث خصائصها على طريقة (Berry et al, (2002). كذلك أوضحت دراسة جابر (٢٠١٨) مشكلة أخطاء القياس ومفهومها وأسباب حدوثها وكذلك تطرقت إلي النماذج شبه المعلمية وتأثير أخطاء القياس علي النتائج؛ وتناولت أيضاً بعض طرق التقدير لدالة الانحدار شبه المعلمية عند وجود أخطاء قياس في المتغيرات التوضيحية (explanatory) وهذه الطرق هي طريقة السيمكس (simex method)؛ طريقة المتغيرات المساعدة (instrument variable method) وأخيراً طريقة شبه الإمكان المعدلة (Quasi-likelihood method). وتم مقارنة نتائج الطرق الثلاث السابقة بناءً علي معيار متوسط مربعات الخطأ (AMSE). وكذلك قامت الدراسة باستخدام المحاكاة للمقارنة مابين الثلاث طرق وتم التوصل إلي أن طريقة المتغيرات المساعدة تتفوق في نتائجها علي باقي الطرق.

كما ركزت دراسة (Selim (2020) على نموذج الانحدار الخطي جزئياً. حيث قارنت بين بعض طرائق التقدير لتقدير هذا النموذج في كلا من الجزء المعلمي واللامعلمي مثل مقدر روبينسون (Robinson's estimator) ومقدر سبيكمان (Spackman estimator)؛ ومقدر التوفيق الخلفي (Back-fitting estimator)؛ ومقدر الشرائح الجزئية (Penalized splines estimator)؛ ومقدر الفروق (Difference-based estimator). وتم مقارنة المقدرات وفقاً لمعيار (AMSE) وتم استخدام محاكاة ماننت كارلو. وتم التوصل إلي أن مقدر الشرائح الجزئية (p-splines estimator) هو الأكثر كفاءة بين المقدرات.

وقد قدمت دراسة (Cao & Wong (2021) نموذج الخطأ الجمعي (additive measurement error) طبقاً لخصائص منظمة التحقيق الأوربي المستقبلية في السرطان والتغذية (EPIC). وتوصلت الدراسة إلي مقدر تقريبي لدالة الإمكان الأعظم (AMLE) للمتغيرات المقاسة في وجود خطأ القياس لنموذج الانحدار اللوجستي.

حيث أن هذه الطريقة يمكن النظر إليها على أنها نسخة معدلة من معايرة الانحدار (regression calibration) ويمكن من خلالها الحصول على تقدير تقريبي متنسق. وكذلك تم إجراء دراسات محاكاة لاختبار الطريقة المقدمة.

(2) النماذج محل الدراسة:

أولاً نموذج الانحدار شبه المعلمي (الانحدار اللوجستي) (Logistic regression model):

يتكون النموذج الإحصائي من متغير تفسيري غير ملاحظ (X_i) ويكون له بديل ملاحظ ومتجه متغيرات تفسيرية خالية من الأخطاء (Z_i) (Devanarayan & Stefanski (2002)). وتم استخدام النموذج شبه المعلمي نظراً لعدم معرفة العلاقة بين المتغير غير الملاحظ (X_i) والمتغير التابع (Y_i). وكذلك تم الاعتماد على النموذج اللوجستي نظراً لأن المتغير التابع ثنائي الاستجابة (0,1).

ويكون النموذج على الصورة التالية:

$$\text{Logit}\{\Pr(Y_i = 1|X_i, Y_i)\} = \beta_0 + \beta_X X_i + \beta_Z Z_i \quad (1)$$

$$\beta_Z = (\beta_{Z_1}, \beta_{Z_2}, \beta_{Z_3}, \beta_{Z_4}, \beta_{Z_5})^T \quad \text{متجه معاملات } Z_i \quad (2)$$

حيث إن:

Y_i : يمثل متغير الاستجابة (response variable).

X_i : يمثل المتغير غير الملاحظ (الكامن) (latent) (unobserved latent covariate).

Z_i : يمثل متجه المتغيرات الخالية من اخطاء القياس (error-free covariates vector).

ثانياً نموذج خطأ القياس غير المتجانس الذي يربط المتغير غير الملاحظ (X_i) مع البديل الملاحظ

له (W_i):

$$W_i = X_i + \sigma_{(X_i)} U_i, \quad i = 1, \dots, n \quad (3)$$

حيث إن:

W_i : يمثل البديل الملاحظ للمتغير (X_i) (observed surrogate of X_i).

• ومنه يمكن الحصول على صيغة رياضية للمتغير العشوائي :

$$U_i = (W_i - X_i) / \sigma(X_i) \quad (U_i \sim N(0,1)) \quad (4)$$

حيث إن:

➤ U_i يمثل متغير عشوائي وهو غير مقيد بتوزيع معين وسيتم افتراض أنه يتبع التوزيع الطبيعي بمتوسط صفر وتباين واحد؛ ويكون مستقل عن كلا من المتغيرات (X_i) , (Z_i) .

ثالثاً تقدير دالة الانحراف المعياري باستخدام (B-spline approximation):

حيث أن دالة الانحراف المعياري للمتغير غير الملاحظ (X_i) تكون غير معروفة وبالتالي نقدرها بشريحة B-Li et al (2019).

$$\hat{\sigma}(X_i) = \beta(X_i)^T \hat{\gamma} \quad (5)$$

ونظراً لأن $\hat{\gamma}$ ليست محل اهتمام البحث فيطلق عليها معلمة مزعجة (nuisance parameter) وسيتم تقديرها.

ويكون المدي لغعد شريحة B كالتالي:

$$\text{knots } t_{-r+1} = \dots = t_0 = 0 < t_1 < \dots < t_N < 1 = t_{N+1} = \dots = t_{N+r} \quad (6)$$

وشريحة B تكون من الرتبة (r) وتمثل (N) العُقد الداخلية (interior knots).

• ويمكن الحصول على صيغة $\hat{\gamma}$ بالتعويض بالمعادلة (3) في المعادلة (5):

$$\hat{\gamma} = \frac{\hat{\sigma}(X_i)}{\beta(X_i)^T} = \frac{W_i - X_i}{\beta(X_i)^T U_i} \quad (7)$$

$\sigma(X_i)$: دالة موجبة مزعجة غير معروفة (unknown positive nuisance function) والتي تصف عدم التجانس في خطأ القياس؛ وسيتم تقديرها باستخدام (B-spline approximation).

γ : تمثل أيضا معلمة مزعجة (nuisance parameter) سيتم تقديرها لاستخدامها في تقدير دالة التباين غير المعروفة.

التالي بعض خصائص المقدرات :

(1) كفاءة المقدر ($\hat{\beta}$) تعتمد علي التحديد الصحيح لتقدير الدالة الاحتمالية الشرطية $f_{x|z}$ حيث إذا كان $f_{x|z}^* = f_{x|z}$ فيكون $\hat{\beta}$ مقدر كفاء حيث يمكن تقديرها عن طريق إستخدام تقريب شرائح-B وهو ما يكون علي الصورة $B(X|Z)^T \xi$.

ونظراً لأنه سيكون لدينا متغير غير ملاحظ (X) وسيكون له بديل ملاحظ وهو المتغير (W) وكذلك سيكون هناك قياسات لمتغيرات مقياسة بدقة ونعبر عنها بالمتجه (Z) فبالتالي يمكن الحصول علي مقدر متسق للمتغير ξ عن طريق تعظيم لوغاريتم دالة الإمكان وتكون دالة الامكان كالتالي :

$$\prod_{i=1}^n \int f u \left\{ \frac{w_i - x}{\hat{\sigma}(x)} \right\} \hat{\sigma}(x)^{-2} B(X|Z_i)^T \xi dx \quad (8)$$

ويمكن الحصول علي تقدير مصفوفة التباين المتماثلة (v) كما يلي:

$$\hat{v} = \hat{Q}^{-1} \hat{\Sigma} (\hat{Q}^{-1})^T \quad (9)$$

ويكون تقدير دالة التباين كالتالي:

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n S_{eff}^*(Y_i, w_i, Z_i, \hat{\beta}_n, \hat{\gamma}_n, f_{x|z}^*) S_{eff}^*(Y_i, w_i, Z_i, \hat{\beta}_n, \hat{\gamma}_n, f_{x|z}^*)^T \quad (10)$$

ويكون تقدير مصفوفة Q علي الصورة التالية:

$$\hat{Q} = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \beta_n^T} S_{eff}^*(Y_i, w_i, Z_i, \hat{\beta}_n, \hat{\gamma}_n, f_{x|z}^*) \quad (11)$$

(2) سوف يؤثر تقدير المعالم المزعجة $(\sigma(x), f_{x|z})$ علي تقدير التباين للمعلمة (β) بحيث أن في حالة معرفة القيمة الحقيقية لدالة الانحراف المعياري للمتغير غير الملاحظ $\sigma(x)$ وتم إستخدامها في عملية تقدير النموذج الحقيقي للمتغير (X) في وجود المتغير (Z) فبالتالي تقدير تباين المعلمة (β) لن يتغير .

الحصول علي تقديرات لفترات الثقة للمعالم من أهم الأهداف التي تنتج عن عملية تصحيح خطأ القياس للمتغير . ومن المشكلات المتعلقة بفترات الثقة للمعالم عند وجود خطأ قياس هي عدم توازن المدي الخاص بالفترات أي قد تقع القيمة الحقيقية للمعلمة أسفل الحد الأدنى للفترة أو قد تقع فوق الحد الأعلى للفترة . ونظراً لأن إختبارات الفروض الإحصائية للمعلمة الحقيقية عادة يكون لها علاقة مباشرة مع تعديل وتصحيح فترة الثقة وبالتالي قد يرجع عدم التوازن لخصائص هذه الإختبارات (Lyles & Kupper (1999).

(3) الجانب التطبيقي:

يتناول الجانب التطبيقي عينة حجمها (429) مريض بفيروس كوفيد (19) وتم الحصول عليها من مستشفى كفر سعد المركزي بمحافظة دمياط وكانت البيانات تغطي الفترة الزمنية من بداية شهر أكتوبر (2020) وحتى منتصف شهر يناير (2021) وتم تجميع البيانات من قبل الباحث. وقد كان أول ظهور للفيروس في مدينة ووهان الصينية في عام (2019)، حيث انتشر حول العالم منذ ذلك الوقت مسبباً جائحة فيروس كورونا ومنذ بداية الجائحة حتي الآن تم الإبلاغ عن ملايين الإصابات بفيروس كورونا في كافة دول العالم وكذلك ملايين الوفيات.

ومن الأعراض الشائعة للمرض الحمى والسعال وضيق النفس. وعادة ما ينتشر هذا الفيروس بين الأشخاص أثناء الاتصال الوثيق بينهم غالباً عبر قطرات صغيرة من الرذاذ تنتج عن طريق السعال والعطس والتحدث. ويكون الفيروس أكثر قابلية للعدوى خلال الأيام الثلاثة الأولى بعد ظهور الأعراض، على الرغم من أن العدوى ممكن أن تحدث قبل ظهور هذه الأعراض ومن أشخاص قد لا تظهر عليهم أعراض المرض. بالإضافة إلى ذلك يوصى باستخدام غطاء الوجه لأولئك الذين يشكون في أن لديهم الفيروس والذين يقدمون الرعاية لهم.

تعريف المتغيرات التفسيرية والمتغير التابع للبيانات :

أولاً المتغير التابع:

Y : وهو نوع الضغط للمريض (Pressure kind) وهو متغير ثنائي الإستجابة وإستجابة المريض أما أن تكون:

(1) ضغط المريض طبيعي وبالتالي الإصابة بفيروس كوفيد(19) ليس له علاقة بإرتفاع الضغط ويأخذ القيمة (0).

(2) ضغط المريض عالي وبالتالي الإصابة بفيروس كوفيد(19) له علاقة بإرتفاع الضغط ويأخذ القيمة (1).

ثانياً المتغيرات التفسيرية :

X_1 : هو ضغط المريض (Pressure) وهو المتغير الكامن غير الملاحظ (latent variable) .

W_1 : هو متوسط الضغط (Average pressure) وهو البديل الملاحظ للمتغير الكامن حيث تم تجميع (4) قياسات ضغط لكل مريض وتم حساب المتوسط لكل مريض.

المتغيرات التالية هي المتغيرات الخالية من خطأ القياس وتكون كالتالي:

Z_1 : يمثل متغير العمر (Age).

Z_2 : يمثل متغير النوع وينقسم إلى: أما ذكر ويأخذ القيمة (0) وأما أنثى وتأخذ القيمة (1).

Z_3 : يمثل مدة إقامة المريض بالمستشفى (Duration) وتم حساب المدة لكل مريض من يوم دخول المستشفى حتى خروجه منها.

Z_4 : هو متوسط درجة الحرارة (Average temperature) وتم تجميع قياسات (4) درجات حرارة لكل مريض وتم حساب المتوسط للقياسات لكل مريض.

Z_5 : يمثل نبض المريض (pulse).

وقد تم تطبيق كلا من:

• نموذج الانحدار اللوجستي.

• شرائح B -

وفيما يلي عرض بالنتائج التي تم التوصل إليها بعد إجراء التحليل الإحصائي:

جدول (1): بعض الإحصائيات الوصفية لمتغيرات البيانات

Response Variable	Coding	Level	Freq.	%
Pressure kind	0	Normal Pressure	166	38.69
	1	High Pressure	263	61.31
Explanatory Variables	Coding	Level	Freq.	%
Sex	0	Male	213	49.66
	1	Female	216	50.34
	Min	Max	Mean	S.D
Average pressure	1.11	2.250	1.622	0.15
Age	15	90	56.32	14.63
Duration	1	26	5.59	3.84
average temperature	36	38	37.03	0.17
Pulse	63	150	84.73	8.43

يتضح من جدول (1) ما يلي:

- بالنظر في المتغير التابع (Pressure kind) نجد أن عدد الأشخاص مرتفعي الضغط (High Pressure) كان أكثر من عدد الأشخاص أصحاب الضغط الطبيعي (Normal Pressure).

- أيضاً تبين أن متغير الجنس (sex) قد حصلت الإثبات (female) على نسبة (50.34%) وهي أيضاً مقارنة للغاية لنسبة الذكور في مجموعة البيانات.
 - وأيضاً بالنظر إلى الانحراف المعياري للمتغيرات نستنتج أن البيانات لا تواجه تباين كبير ويتم توزيعها بشكل طبيعي.
- وقد نرغب في معرفة ما إذا كان هناك أي من مشاكل القياس في البيانات قبل استخدام نموذج الانحدار المناسب والملائم للبيانات. فقد نرغب في معرفة هل هناك علاقة خطية متعددة بين المتغيرات التفسيرية أم لا ولذلك تم استخدام ارتباط بيرسون .

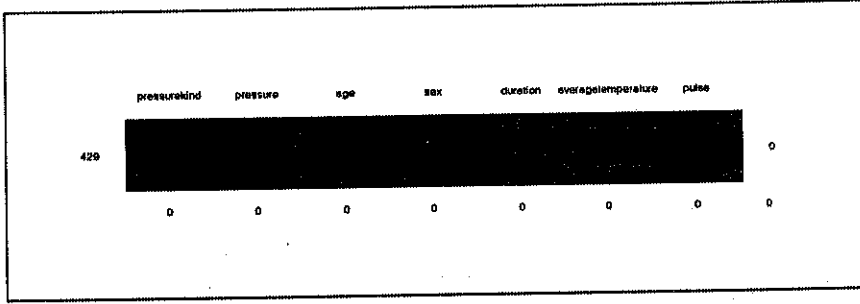
جدول (2): مصفوفة ارتباط بيرسون للمتغيرات التفسيرية

Variables	Ave.pressure	Age	Sex	Duration	Ave.temperature	Pulse
pressure	1					
Age	0.203	1				
Sex	-0.035	0.044	1			
Duration	-0.062	0.069	0.032	1		
.Ave Temp	-0.03	0.031	0.039	0.086	1	
Pulse	-0.075	0.028	0.024	0.074	-0.043	1
VIF	1.066	1.043	1.006	1.023	1.012	1.038

يوضح جدول (2) :

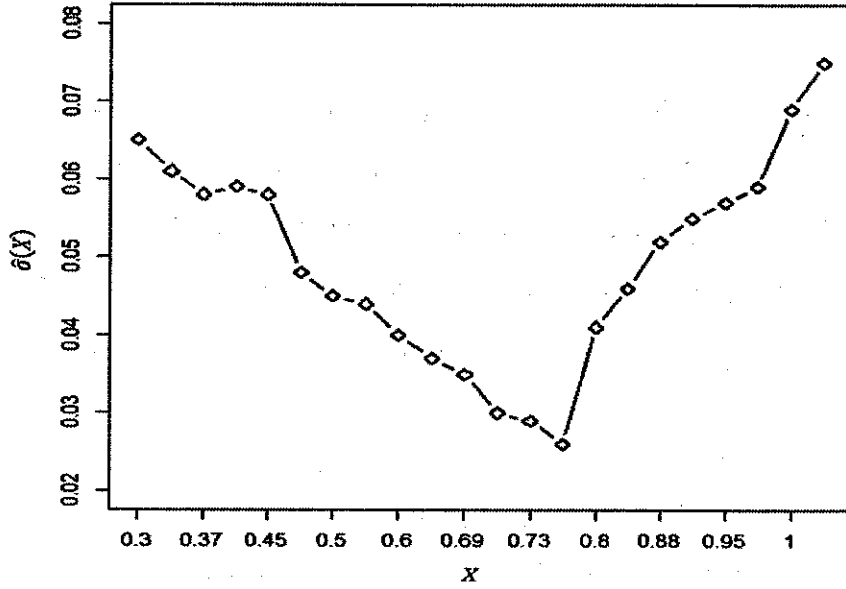
- مصفوفة معاملات الارتباطات البيئية للمتغيرات وقيم معامل تضخم التباين (Variance Inflation (VIF) Factor)، ومن الممكن ملاحظة أن بعض المتغيرات التفسيرية ضعيفة الارتباط مثل متغير النوع (sex) مع متغير متوسط الضغط (average pressure) حيث كانت قيمة معامل الارتباط بينهما = (-0.035).
- كما يتضح أيضاً من قيم معامل تضخم التباين (VIF) أن جميع القيم أقل من 5، وهذا دليل كافي على عدم وجود مشكلة الإزدواج الخطي في البيانات.

➤ وكذلك تم فحص مجموعة البيانات قيد الدراسة من مشكلة القيم المفقودة (missing values) ووجد أن مجموعة البيانات لا تعاني من مشكلة القيم المفقودة كما يتضح ذلك من الشكل رقم (1).



شكل رقم (1): التحقق من البيانات المفقودة باستخدام (test patterns)

- وكذلك تم تقدير دالة الانحراف المعياري المجهولة $\sigma(X)$ باستخدام شريحة B- (B-splines) وذلك في حالة عدم تجانس أخطاء القياس.
- حيث تم استخدام (6) عُقد (knots) لشريحة B.
- وكان تقدير دالة الانحراف المعياري باستخدام شريحة B- = (0.0515) وذلك في حالة تجانس خطأ القياس.



شكل رقم (2): تقدير دالة الانحراف المعياري للخطأ باستخدام طريقة تقدير B-Spline Estimator

وفقاً لشكل (2) السابق نجد أن:

- 1- تم تقدير دالة الانحراف المعياري للخطأ وذلك في حالة عدم تجانس أخطاء القياس ودائماً تتغير قيمتها فيما يتعلق بالمتغير (X).
- 2- وبالتالي فإن الافتراض الخاص بتجانس خطأ القياس قد يكون مقيداً للغاية لمجموعة البيانات المستخدمة.
- 3- يتضح من شكل (2) أنه عندما تكون قريبة من (1, 0) فيميل تباين خطأ القياس إلى أن يصبح أكبر ما يمكن وبالتالي يصبح قيمة الانحراف المعياري كبيرة .
- 4- بالتالي نستنتج أن التقدير في حالة عدم تجانس أخطاء القياس يكون أفضل لمجموعة البيانات المستخدمة.

جدول (3) نتائج تقدير نموذج الانحدار شبه المعلمي باستخدام طريقة (B-spline)

Variables	Intercept	aver.pressure	Age	Sex	Duration	average temperature	Pulse
No measurement error							
Estimate	0.4154	-1.2452	0.0160	1.1461	0.4774	-0.2809	0.4219
Std. Error	0.1861	0.5387	0.03 26	0.5083	0.2737	0.1942	0.3908
T-Statistic	2.2321	-2.3115	0.4906	2.2550	1.7442	-1.4467	1.0796
P-value	0.0165*	0.0185*	0.6183	0.0218*	0.1480	0.1093	0.2648
Homoscedastic measurement error							
Estimate	0.3154	-1.6015	0.1591	1.2432	0.4291	-0.2749	0.3183
Std. Error	0.1068	0.4987	0.0933	0.7358	3.1274	0.1678	0.5201
T-Statistic	2.9526	-3.2113	1.7064	1.6896	0.1372	-1.6382	0.6120
P-value	0.0084*	0.0026*	0.1369	0.0937	0.1024	0.0938	0.5195
Heteroscedastic measurement error							
Estimate	0.2075	-2.0601	1.3232	0.0161	0.4092	-0.2505	0.4234
Std. Error	0.0226	0.3264	0.3916	0.0303	0.2418	0.1078	0.1767
T-Statistic	9.1814	-6.3117	3.3791	0.5335	1.6921	-2.3232	2.3962
P-value	<0.001*	<0.001*	0.0014*	0.5164	0.0932	0.0217*	0.0172*

Note: The superscripts * indicate statistical significance at 0.05 level.

جدول (3) يوضح نتائج تقدير نموذج الانحدار شبه المعلمي باستخدام طريقة (B-spline) ومن الجدول يمكن ملاحظة التالي:

- ١- بالنسبة لفرض عدم وجود أخطاء قياس (no measurement error):
 - أ- نلاحظ معنوية كلا من متغير النوع (sex) ومتوسط الضغط للمريض (average pressure) وهو ما يدل على تأثير هذه المتغيرات بشكل كبير على المتغير التابع وذلك عند مستوى معنوية .

ب- بينما نلاحظ كلا من متغيرات العمر (age) و مدة الإقامة في المستشفى (duration) ومتوسط درجة حرارة الشخص (average temperature) ونبض المريض (pulse) ليست معنوية (Insignificant) أي ليست لها تأثير على المتغير التابع وذلك عند مستوى معنوية .

2- بالنسبة لفرض تجانس خطأ القياس (homoscedastic measurement error):

أ- نلاحظ أن متغير متوسط ضغط المريض (average pressure) هو المتغير الوحيد المعنوي (significant) وذلك عند مستوى معنوية .

ب- بينما باقي المتغيرات غير معنوية أي ليس لها تأثير على المتغير التابع .

3- بالنسبة لفرض عدم تجانس خطأ القياس (heteroscedastic measurement error):

أ- نلاحظ أن كلا من متغير متوسط ضغط المريض (average pressure) ومتغير العمر (age) ومتوسط درجة حرارة الشخص (average temperature) ونبض المريض (pulse) متغيرات معنوية وذلك عند مستوى معنوية .

ب- بينما نلاحظ أن كلا من متغير النوع (sex) والمدة (duration) متغيرات غير معنوية وذلك عند مستوى معنوية .

(4) النتائج:

يمكن إيجاز النقاط التي تم التوصل إليها في هذا البحث في كلاً من:

(1) بمقارنة نتائج التقدير للنموذج شبه المعلمي تحت الثلاث فروض الخاصة بخطأ القياس نجد أن هناك اختلاف ما بين النتائج وهذا الاختلاف يرجع لتجاهل خطأ القياس أو عدم وضع الفرض الصحيح الملائم لمجموعة البيانات .

(2) نتائج التقدير للمتغيرات عند استخدام فرض عدم تجانس خطأ القياس تكون معظم قيم المتغيرات عند قيمة وهو ما يدل على أنها أكثر معنوية ، وهذا يدل على أن فرض عدم تجانس خطأ القياس يناسب طبيعة هذه البيانات.

(5) البرامج المستخدمة:

برنامج R.

(6) التوصيات:

بناء على ما توصل إليه البحث من نتائج يوصي بالآتي:

- (1) التركيز على مشكلة أخطاء القياس وفروضها وتأثيرها على التحليل الإحصائي وخاصة مشكلة عدم تجانس أخطاء القياس.
- (2) تطبيق طرائق ونماذج إحصائية جديدة عليها مثل (P-splines , smoothing splines Nadarya- Watson estimator,...) ومعرفة أكثر عنها والمقارنة بين النتائج.
- (3) التركيز على طرق التقدير لنماذج الانحدار اللامعلمية والعمل على توضيح الطرق المختلفة للتقدير وذلك لأهميتها وكثرة الظواهر التي تحتاج إلى إستخدامها لتقدير النماذج .

(7) المراجع:

أولاً: المراجع العربية:

- جابر، أحمد مهدي (٢٠١٨) "تقدير دالة الانحدار شبه المعلمية بوجود خطأ القياس مع تطبيق عملي"، رسالة ماجستير، كلية الإدارة والاقتصاد-جامعة بغداد.
- طه، كوثر سعيد، حسن، دزخشان جلال (٢٠٢٠) "استخدام النموذج اللوجستي لدراسة أهم العوامل المؤثرة على مرض السكري حسب نوع المرض"، بحث مشترك، كلية الإدارة والاقتصاد - جامعة صلاح الدين.
- علي، زينب رزق البديوي (٢٠١٧) "مقارنة بعض أساليب تقدير دالة الانحدار اللامعلمي لبيانات طولية"، رسالة ماجستير، كلية التجارة-جامعة المنصورة.

ثانيا: المراجع الأجنبية:

- Berry, S. M., Carroll, R. J., & Ruppert, D. (2002). Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association*, 97(457), 160-169.
- Cao, Z., & Wong, M. Y. (2021). Approximate maximum likelihood estimation for logistic regression with covariate measurement error. *Biometrical Journal*, 63(1), 27-45.
- Cheng, C. L., & Riu, J. (2006). On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. *Technometrics*, 48(4), 511-519.
- Devanarayan, V., & Stefanski, L. A. (2002). Empirical simulation extrapolation for measurement error models with replicate measurements. *Statistics & Probability Letters*, 59(3), 219-225.
- Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical science*, 11(2), 89-121.
- Li, M., Ma, Y., & Li, R. (2019). Semiparametric regression for measurement error model with heteroscedastic error. *Journal of multivariate analysis*, 171, 320-338.
- Lyles, R. H., & Kupper, L.L. (1999). A note on confidence interval estimation in measurement error adjustment. *The American Statistician*, 53(3), 247-25.
- Sarkar, A., Mallick, B. K., & Carroll, R. J. (2014). Bayesian semiparametric regression in the presence of conditionally heteroscedastic measurement and regression errors. *Biometrics*, 70(4), 823-834.

Seliem, M. M. (2020). A Comparative Study of Some Estimation Methods for Partially Linear Model (Doctoral dissertation, Cairo University).

Yi, G. Y. (2017). Statistical analysis with measurement error or misclassification: strategy, method and application. Springer Series in Statistics.

