دراسة في المتغيرات المضببة والأنحدارالمتعدد المضبب

- د. محمد طه أحمد الغنام
- م.م. هبة علي طه الصباغ

الخلاصة

يعتبر الإحصاء المضبب اتجاه ناشئ وذو أهمية في محور الإحصاء والمعلومات وله مجال عمل واسع لاعتماده على نظرية المجموعات المضببة . تناول هذا البحث تعريف بالمجاميع المضببة ودراسة المتغيرات المضببة، والمتغيرات العشوائية المضببة ، ثم الانحدار المتعدد المضبب وقد استعملنا الأخير لبيانات تجربة إحيائية عن مكونات دم الإنسان ، ولافتراضنا بوجود ضبابية في قياساتها ، فتم تقدير نموذج الانحدار المتعدد المضبب لهذه البيانات باستخدام طريقة المربعات الصغرى المضببة .

ABSTRACT

The trend of fuzzy statistics is new in sciences, which is wide and important that depend on fuzzy set theory . In this paper we define fuzzy sets and study fuzzy variable , fuzzy random variable , and fuzzy multiple regression . we estimate a fuzzy multiple regression model to experiment of five component human blood with the age that represent a fuzzy data , using fuzzy least squares method .

1 - المقدمة:

يعالج الإحصاء التقليدي الظواهر الحياتية والمتغيرات التي تعتمد العشوائية منهاجا في التحليل والنتائج ، والقرار عليها ، هذه الظواهر تكون ذات قياسات جازمة ، محددة ومعلومة (القياسات النقطية) والأخطاء الناجمة عنها تكون متغيرات عشوائية يمكن السيطرة عليها من خلال دراسة سلوكها فقد تكون مثلا تتبع احد التوزيعات المعروفة . إما الإحصاء المضبب Fuzzy statistic فقد ظهر حديثا بعد نشوء نظرية المجاميع المضببة ليهتم بالظواهر التي لا يمكن قياس متغيراتها نقطيا وأنما تقاس بفترات ، اوما يوصف بالحالات اللامؤكدة أو الحالات ذات البيانات الضبابية لما تتصف به من صفات تجعلها غير واضحة كالمتغيرات التي تنتمي بنسب معينة إلى مجاميعها وليس لها انتماء تام ،وكذلك المتغيرات اللغوية التي لا يمكن قياسها عدديا ، وهناك متغيرات تقاس بشكل تقريبي الا أنها في الحقيقة فيها ضبابية .

ونشير إلى أن هناك دراسات عديدة في مجالات الإحصاء المختلفة في ظروف الضبابية وضمن إطار المجموعات المضببة ، وهو مجال واسع يشمل الأشياء الممكنة الحدوث possibility ، ذلك لأن الإحصاء التقليدي يدرس الأشياء المحتملة الحدوث probability ضمن إطار المجموعات التقليدية . وقد تناول الباحثين مجالات مختلفة للموضوع ، وبشكل موجز فقد بحث بعضهم النمذجة المضببة الخطية كالانحدار الخطي ، والنمذجة الرياضية ، والمتغيرات العشوائية المضببة . وفي مجال تقدير المعالم لبيانات مضببة قام آخرون بتوسيع طريقة الإمكان الأعظم لتشمل الحالة المضببة للتقدير النقطي وصفاته والنظريات ذات العلاقة والقيمة المتوقعة المضببة ، وكذلك دراسة التقدير بفترة ومناطق الثقة ، وفي اختبار الفرضيات تم اعتماد توسيع نظرية نيمان بيرسون لتلاءم حالة الضبابية ، فضلا عن بحوث كثيرة لتعميم أساليب تحليل بيز التقليدي إلى حالة البيانات المضببة ونظرية القرار والتوزيعات السابقة ، ودراسات في مجالات اخرى عديدة (Taheri , 2003).

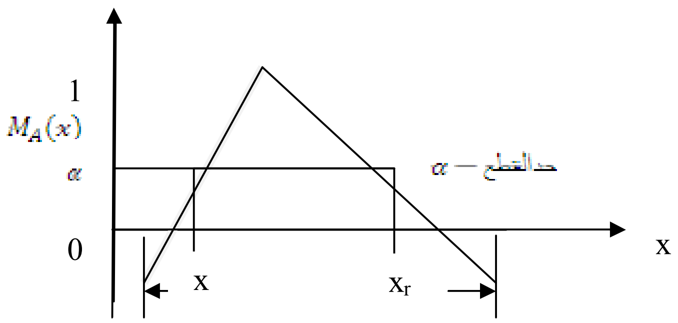
- يهدف هذا البحث الى التعريف بالمتغيرات المضببة والمتغيرات العشوائية المضببة كونها تمثل جزءا مهما في الإحصاء المضبب واستعراض الانحدار المتعدد المضبب والتطبيق عليه باستخدام المربعات الصغرى المضببة لتجربة متغيراتها هي مكونات دم الإنسان (عددها خمسة) مع العمر ، مفترضين تضمنها لنسبة من الضبابية بسبب صعوبة الحصول على بيانات جازمة تماما.

2 - المجموعة المضببة Fuzzy Set: هي المجموعة التي تمتلك عناصرها نسبة انتماء يسمى درجة الانتماء أو درجة العضوية membership degree والتي تكون إعداد حقيقية ضمن الفترة المغلقة [0,1] ،ويعبر عن درجة العضوية بدالة ألانتماء $M_A(x)$ التي تمثل درجة انتماء العنصر من المتغير x إلى المجموعة المضببة A وتكتب بالشكل العام الأتي :-

$$M_A (x): x \rightarrow [0,1]$$

يتغير الانتماء للعناصر من العضوية التامة الى عدم العضوية ،فهي إما أن تمتلك عضوية تامة، أو عضوية جزئية أو لا تمتلك عضوية ، عليه فان عناصر المجموعات الاعتيادية لها انتماء تام أي إن $(M_A(x)=1)$ أو ما يدعى بالعناصر الجازمة Crisp .

من المفاهيم المتعلقة بالمجموعة المضببة هو مجموعة القطع α - cut فان $^{\alpha}A$ هي مجموعة تحوي عناصر x لها درجة انتماء لا تقل عن النسبة α لتعبر عن درجة انتماء العناصر المهمة والتي تدعى مجموعة المستوى - α (α -level) ينحصر الانتماء المهم بين قيمتين يمنى ويسرى (x_l , x_r) على خط ارتكاز المجموعة A (Support A) وماعدا ذلك فأنه قليل الأهمية ويترك غالبا ليكون خارج نطاق العمل (cut out) ونستطيع توضيح هذه المفاهيم بالشكل الآتي :-



شكل رقم (1)

،وتسمى $A^{+\alpha}$ مجموعة القطع القوي التي لها درجة انتماء اكبر من α وان اكبر درجة انتماء يسمى ارتفاع المجموعة المضببة ، أما جوهر المجموعة A (core(A)) فهو مجموعة العناصر التي لها درجة انتماء تساوي واحد .

تعريف : اذا كانت B ،A مجموعتين مضببتين في X فان :-

1 – $A=B$ iff $M_A(X) = M_B(X)$

$2 - \quad A \subset B \quad \text{iff} \quad M_A(x) < M_B(x)$

يمكن تمثيل المجموعة المضببة بعدة صيغ أهمها :

1 - شكل الأزواج المرتبة : يكتب فيها العنصر مع درجة انتماؤه :

$$\tilde{A}=(x, M_A(x))$$

مثال :مجموعة مضببة تتضمن الأزواج الآتية : $\tilde{B}=\{(2,\ 0.4),(3,\ 0.6)\}$

أي أنها تحتوي العنصر 3 الذي درجة انتماؤه 0.6 والعنصر 2 ودرجة انتماؤه 0.4 او بتعبير آخر قيم المتغير X ودرجة انتماؤه (x=3 بدرجة انتماء قدرها 0.6) وهكذا

.

2 - دوال الانتماء : يمكن ان نعبر عن المجموعة المضببة بذكر دالة الانتماء لعناصرها أو متغيراتها ، وهناك عدة أشكال لدوال الانتماء منها ما يأتي :

1- دالة الانتماء المثلثية : وهي دالة خطية عند رسمها تعطي شكل المثلث قاعدته الفترة المحددة في الصيغة ورأسه مركز العدد المضبب عادة (الشكل رقم (1) يمثل دالة مثلثية)، ويمكن كتابة صيغتها الرياضية كما يأتي :-

$$M(x)=1-\frac{a-x}{c} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad a-c\leq x\leq a+c$$

ويكون لكل قيمة من قيم المتغير x درجة انتماء متغيرة تحددها هذه الدالة .

2- الدالة القياسية S : وهي دالة خطية متزايدة ، صيغتها :

$$S_{(x)}=\begin{cases} 2\left[\frac{x-b}{c-b}\right] & \ldots\ldots\ldots\ldots \quad a<x<b \\ 1-2\left[\frac{x-c}{c-a}\right] & \ldots\ldots\ldots\ldots \quad b<x<c \\ 1 & \ldots\ldots\ldots\ldots \quad x>c \end{cases}$$

3 - دالة الشكل ألجرسي : دالة أسية غير خطية تعبر عن شكل المنحنى الطبيعي وصيغتها :

$$M_{(x)}=e^{-\frac{(x-a)}{b}} \quad \ldots\ldots\ldots\ldots \quad -\infty<x<\infty$$

وهناك دوال أخرى مثل دالة شبه المنحرف ودالة اللوجستك ودالة مماس القطع الزائد وغيرها

.

أما العمليات الحسابية على الأعداد المضببة فهي عمليات على فترات مغلقة كل فترة تعبر عن عدد مضبب ، فلو كانت لدينا الفترتان المغلقتان [a,b],[c,d] ستكون العمليات الحسابية الأربعة كما يأتي :

1-[a,b]+[c,d]=[a+c,b+d]

2-[a,b]-[c,d]=[a-d,b-c]

3-[a,b].[c,d]=[min(ac,ad,bc,bd),max(ac,ad,bc,bd)]

4-[a,b]/[c,d]=[a,b].[1/d,1/c],0

3- المتغيرات العشوائية المضببة Fuzzy Random Variables (FRV) :-

يمكن ان نعرف المتغير العشوائي المضبب بانه " متغير عشوائي قيمه ليست حقيقية بل أعداد مضببة" (Shapiro, 2008) ذلك لأنه يعبر عن ظواهر غامضة أو غير مؤكدة ذات قياسات ليست جازمة. مثلا لو كان لدينا وعاء فيه 60 كرة مختلفة الأحجام ، ما هو احتمال سحب كرة كبيرة الحجم ؟ الجواب يكون عددا مضببا لأن كلمة كبيرة تتضمن ضبابية ، هذا يمثل متغير عشوائي مضبب دالته من الفضاء ألاحتمالي إلى مجموعة المتغيرات المضببة.

ويمكن تعريف المتغير العشوائي المضبب كذلك عن طريق توسيع بديهيات فكرة الاحتمال التقليدي لكولموكروف (Bernard et.al,) ، فإذا كانت أبعاد الفضاء الاحتمالي الثلاثي التقليدي (Ω, f, p) ضبابية،سيكون لدينا قياس احتمالي غير مؤكد يعرف علىالفضاء الأقليدي R^n ، لنصف الفضاء الاحتمالي المضبب بالرموز $(\tilde{\Omega}, \tilde{f}, \tilde{p})$ فيكون المتغير العشوائي المضبب $\tilde{x}$ هو نتيجة مضببة للراسم غير ألأكيد الآتي:-

$$\tilde{\Omega} \tilde{\rightarrow} F(R^n)$$

حيث إن $F(R^n)$ مجموعة كل الأعداد المضببة في R^n .

أن كل متغير عشوائي X (بدون ضبابية) معرف على مجال X محتوى تماما في $\tilde{X}$ ، والمتغير العشوائي المضبب ايضا هو مجموعة مضببة لكل المراكز Original الممكنة الى X ومحتواة في $\tilde{X}$. ويمكن التعبير عنه رياضيا بالدالة التوزيعية المضببة $\tilde{F}(x)$ التي

توصف بأنها مجموعة كل الدوال التوزيعية للمراكز X_i في $\tilde{X}$ بقيم انتماء قدرها $M(F_{(x)})$.

وفي وصف آخر للمتغير المضبب [المصدر رقم 9](Kwakernak et.al, 1978) أنه هو الراسم الاتي :

$$\xi = \Omega \longrightarrow F(R) \quad such\ that\ for \quad \alpha \in [o, 1], w \in \Omega$$

وتكون القيم الحقيقية للراسم كما يأتي :

$$inf\,\xi_\alpha : \Omega \to R \quad \text{satisfying} \quad inf\,\xi_{\alpha(w)} = \inf \xi_{(w)\,\alpha}$$

$$sup\xi_\alpha : \Omega \to R \quad \text{satisfying} \quad sup\xi_{\alpha(w)} = sup\xi_{(w)\alpha}$$

تمثل متغيرات عشوائية ذات قيم حقيقية وهذا ما يسمى قياس بوريل لدوال بقيم حقيقية . من جانب آخر وضع (Puri &Ralescu, 1986) [المصدر رقم 9] تحديدات على الوصف السابق : انه متغير قيمه مجاميع جزئية مضببة من R^n (فضاء بناخ Banach space)عندما تربط مع مفهوم المجموعة العشوائية .

ويمكن أيضا أن يعبر عن المتغير العشوائي المضبب بحوادث مضببة في فضاء العينة ، بموجب الدراسة التي قام بها الباحث (Shapiro, 2008) عن نوع تمهيدي منها (Primer FRV) لتوضيح الفكرة.

مثال : إذا كان لدينا متغيرين FRV هما $\tilde{X}, \tilde{Y}$ لهما دالة انتماء مثلثية ، والحوادث الآتية مع احتمالاتها وكما يأتي :

$$\tilde{X} = \begin{cases} (a_1, a_2, a_3) \;\dots\dots\; p = 0.7 \\ (b_1,\ b_2, b_3) \;\dots\dots\; p = 0.3 \end{cases}$$

$$\tilde{Y} = \begin{cases} (c_1, c_2, c_3) \;\dots\dots\; p = 0.5 \\ (d_1, d_2, d_3) \;\dots\dots\; p = 0.5 \end{cases}$$

واذا كان $\tilde{Z}$ متغير مضبب حيث $\tilde{Z} = \tilde{X} + \tilde{Y}$ ، فأنه سيأخذ الحوادث الآتية واحتمالاتها :

$$\tilde{Z} = \begin{cases} (a_1 + c_1; a_2 + c_2; a_3 + c_3) \;\dots\dots\dots\; p = 0.35 \\ (a_1 + d_1; a_2 + d_2; a_3 + d_3) \;\dots\dots\dots\; p = 0.35 \\ (b_1 + c_1, b_2 + c_2; b_3 + c_3) \;\dots\dots\dots\; p = 0.15 \\ (b_1 + d_1, b_2 + d_2; b_3 + d_3) \;\dots\dots\dots\; p = 0.15 \end{cases}$$

وان ما يعبر عن الضبابية هو إمكانية حدوث الأشياء أو الحوادث الممكنة possibility events بخلاف الحوادث المحتملة التي ترتبط كثيرا بحالة التأكد .

إما التوزيع الاحتمالي للمتغيرات العشوائية المضببة فقد وصفته بعض الدراسات بأنه ممكن أن يسلك على نفس التوزيعات التقليدية بمعالم مضببة ، احد الاتجاهات (Shapiro , 2008) استنتج التوزيع على وفق أسلوب بيز ، بافتراض أن التوزيع السابق هو دوال الانتماء للمعالم ، مع دالة توزيع احتمالي للمتغير الحالي .

من جانب آخر بين كل من (Xiang, et al ,2007) ؛ انه إذا كان المتغير العشوائي X ذو بيانات جازمة ويتبع التوزيع الطبيعي $X \sim N(\mu,\sigma)$ تكون معالمه نقطية ، ولكن في حالة كون المتغير frv قيمه ضمن فترة (x_l , x_r) فيمكن وصف توزيعه بأسلوب الدالة التوزيعية بالفترة $[F_l , F_r]$ لكل القيم الممكنة وسيكون مجال صندوقي لهذه الدوال أبعاده (p- box), p - وبينوا أن معلمة الموقع (الوسط) ممكن تحديدها ضمن الفترة $(\underline{\mu} , \overline{\mu})$ الا ان المشكلة تكون في معلمة التباين وحيث يمكن التعبير عنها عنها بالفترة $(\underline{\sigma} , \overline{\sigma})$ الا ان ايجادها صعبا ، ثم استخدموا النسبة r حيث ان $r = \frac{x-\mu}{\sigma}$ ، والفترة المترتبة عليها $(\underline{r} , \overline{r})$ لدراسة وتحديد الدالة التوزيعية الطبيعية للمتغير x .

4- الانحدار المضبب :-

أن أول من اقترح الانحدار الخطي المضبب هو (تنكا وآخرون , Tanaka et al (1980,1982) مستخدما البرمجة الخطية في تقدير النموذج ومعلماته . وأستخدم (Celmins, 1987) مقياس للتوافق (compatibility measures) بين معالم النموذج المضبب ، وفي عام 1988 اقترح (Dimond) طريقة المربعات الصغرى المضببة باستخدام معيار (L^2-metric) التي تجاوزت عدد من مشاكل نموذج Tanaka في هذا المجال ، وفي عام 2004 قدم الباحثان (Sanli K., Apaydin ,A) طريقة لتحليل الانحدار المضبب الحصين اعتمادا على أسلوب الانحدار التقليدي الحصين Robust regression) نفسه ، ومعرفان مصفوفة أوزان تعتمد على دالة الانتماء ذات العلاقة ، واستخدم (Donoso, 2006) البرمجة التربيعية لإيجاد نماذج الانحدار المضبب ثم إيجاد انحدار الحرف المضبب (Fuzzy Ridge regression) فضلا عن دراسات أخرى تناولت حالات الانحدار بمتغيرات مضببة .

يمكننا تعريف الانحدار المضبب بأنه وضع صيغة نموذج يعبر عن العلاقة الدالية بين متغير الاستجابة ومجموعة المتغيرات التوضيحية (واحد أو أكثر)في محيط مضبب الذي ينتج عن أن طبيعة العلاقة بين متغيرات النموذج (المتغير المعتمد والمتغيرات التوضيحية)هي ضبابية وأن المتغيرات نفسها هي ذات قياسات اعتيادية (جازمة Crisp) ، أو أن متغيرات النموذج نفسها ضبابية تعبر عن حوادث غير أكيدة .

وبناء على ذلك فسيكون هناك ثلاث أنواع من الانحدار المضبب :-

نموذج تكون فيه المتغيرات جازمة (كما هو الحال في الانحدار التقليدي) والمعالم ضبابية.

أوالمعالم جازمة والمتغيرات ضبابية ،أو أن كل من المعالم والمتغيرات ضبابية .

ومن الممكن أن يصنف العمل بالانحدار المضبب إلى صنفين :-

١ – الانحدار الممكن Possibilistic regression اعتمادا على فكرة الشيء الممكن ، وقد استخدمت البرمجة الرياضية لإيجاده.

٢ – الانحدار المحتمل probabilistic regression و تستخدم فيه طريقة المربعات الصغرى لتقدير النموذج .

ومن جانب آخر وصف (Shapiro, 2008) أنه لو قدرنا نقاط الانتشار العلوية للبيانات بخط مستقيم نسميه Y^U وكذلك نقاط الانتشار السفلية بالخط المستقيم Y^L وعند رسم خط وسطي $Y^{h=1}=(Y^U+Y^L)/2$ يمثل منتصف المسافة بينهما ، فان هذه الخطوط الثلاثة تصف نموذج الانحدار المضبب ،حيث يمثل h نقطة على المحور العمودي تسمى معامل التأكد h-certain الذي يقيم ارتكاز دالة الانتماء و يعبر عن النسبة المهمة لتمركز البيانات، وله تأثير في حدود انتشارات مكونات نموذج الانحدار .

ويمكن كتابة نموذج الانحدار الضبابي الخطي المتعدد بالصيغة الآتية :-

$$\ddot{Y}_i = \ddot{A}_0 + \ddot{A}_1 \ddot{X}_{1i} + \ddot{A}_2 \ddot{X}_{2i} + \cdots + \ddot{A}_p \ddot{X}_{pi} + \epsilon_i \quad \ldots \quad (1)$$

إذا أن :- $\ddot{X}$ المتغيرات التوضيحية الضبابية ، تكون قيمها أعداد ضبابية ضمن انتشار معين ، بحد أيمن هو r_x وحد أيسر هو l_x وقيمة مركزية ويمكن أن تكتب $X=(x,l_x,r_x)$.

$\ddot{Y}_i$:- المتغير المعتمد المضبب ، تكون قيمه محصورة بين حد ايمن r_y وحد أيسر l_y وقيمة مركزية، $\ddot{Y}_i=(y_i, l_y, r_y)$ ويجب أن تكون هناك دالة انتماء لهذه المتغيرات ، وفي

هذا البحث نفترضها دالة مثلثية . $A_0, A_1, \dots . A_p,$ معالم النموذج حيث لكل واحدة منها قيمة مركزية وحد ايمن وحد ايسر لمدى الانتشار بالإضافة إلى دالة انتماء .

ϵ_i: حد الخطأ للنموذج .

يمكن تقدير معالم النموذج المذكور باستخدام طريقة المربعات الصغرى الضبابية FLR التي وصفها Diamond معتمدا في حساب المسافات التربيعية على القياسات الفترويه فإذا كان المتغيرين X, Y بالوصف أعلاه وبدالة انتماء مثلثية ، فأن معيار المسافة يكتب كما يأتي:

$$d(X,Y)^2 = [\, x-y-(l_x-l_y)]^2 + [x-y + (r_x-r_y)]^2 + (x-y)^2$$

أن أفضل مطابقة لنموذج ألانحدار المتعدد تتم بتصغير المسافة التربيعية بين المدخلات والمخرجات ، وقد وصفها (Sanli & Apaydin,2004) وكما يأتي :-

$$Min\ Q = Min\ r(A_0, A_1, \dots \dots . A_p)$$
$$= \sum d(\tilde{A}_0 + \tilde{A}_1\tilde{X}_{1i} + \tilde{A}_2\tilde{X}_{2i} + \cdots + \tilde{A}_p\tilde{X}_{pi}\ ,\ Y_i)^2$$

إذ أن d يمكن كتابتها بالشكل الأتي :-

$$\begin{aligned} d(\tilde{A}_0 + \tilde{A}_1\tilde{X}_{1i} + \cdots + \tilde{A}_p\tilde{X}_{pi}, Y_i)^2 \\ &= [\tilde{A}_0 + \tilde{A}_1\tilde{X}_{1i} + \cdots + \tilde{A}_p\tilde{X}_{pi} - \tilde{y}_i \\ &- (\tilde{A}_1 l_{x1} + \tilde{A}_2 l_{x2} + \cdots + \tilde{A}_p l_{xp} - l_y)]^2 + [\tilde{A}_0 + \tilde{A}_0\tilde{X}_{1i} \\ &+ \cdots + \tilde{A}_p\tilde{X}_{pi} - \tilde{y}_i \\ &+ (\tilde{A}_1 r_{x1} + \tilde{A}_2 r_{x2} + \cdots + \tilde{A}_p r_{xp} - r_y)]^2 + (\tilde{A}_0 + \tilde{A}_1\tilde{X}_{1i} \\ &+ \tilde{A}_2\tilde{X}_{2i} + \cdots + \tilde{A}_p\tilde{X}_{pi} - \tilde{y}_i)^2 \quad \dots\dots\dots\dots\dots\dots (2) \end{aligned}$$

وبأخذ التفاضل الجزئي للمعادلة المذكورة نسبة إلى عدد المعالم $\tilde{A}_0, \tilde{A}_1, \dots . \tilde{A}_p$ المراد تقديرها نحصل على معادلات عددها p+1 من المعادلات التفاضلية ، وبمساواتها بالصفر(عندما تكون ذات قيمة عظمى) وحلها آنيا نحصل على تقديرات المربعات الصغرى .

أن معادلة الحل ، وباستخدام المصفوفات للتعبير عن نموذج الانحدار المتعدد يمكن أن تكتب بالشكل الآتي :-

$$\tilde{\tilde{A}} = (C^tC + X^tX + D^tD)^{-1}($$

$$C^tE + X^tY + D^tF) \ldots\ldots\ldots\ldots(3)$$

إذ أن :-

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1} \\ \vdots & x_{12} & \ddots & \vdots \\ . & . & . & . \\ 1 & x_{1n} & \cdots & x_{pn} \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ . \\ y_n \end{bmatrix}$$

$$E = \begin{bmatrix} y_1 - l_y \\ y_2 - l_y \\ \vdots \\ y_n - l_y \end{bmatrix} \qquad F = \begin{bmatrix} y_1 + r_y \\ y_2 + r_y \\ \vdots \\ y_n + r_y \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & x_{11} - l_{x11} & \cdots & x_{p1} - l_{xp1} \\ 1 & x_{12} - l_{x12} & . & x_{p2} - l_{xp2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} - l_{x1n} & \cdots & x_{pn} - l_{xpn} \end{bmatrix}$$

$$D = \begin{bmatrix} 1 & x_{11} + r_{x_{11}} & \cdots & x_{p1} + r_{x_{p1}} \\ 1 & x_{12} + r_{x_{12}} & . & x_{p2} + r_{x_{p2}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{11} + r_{x_{11}} & \cdots & x_{pn} + r_{x_{pn}} \end{bmatrix}$$

حيث يمكن تصور الحالة أن النموذج العام يكتب باستخدام المصفوفات كالآتي :

$$Y = \tilde{A}\,X$$

وتقسم مصفوفات متغيرات النموذج إلى ثلاث أجزاء احدها يمثل المركز والجزئين الآخرين لقيمتي الانتشار ثم باستخدام صيغة المربعات الصغرى الاعتيادية لهذا النموذج نحصل على معادلة الحل المذكورة أعلاه رقم (3) .

إن أسلوب المربعات الصغرى المضببة قد تغلب على عدد من المعوقات التي يعاني منها تحليل تنكة (Tanaka et al,1982) مثل مشاكل التعدد الخطي ومميزات مثل قياس جودة النموذج ، ومقاومة الشواذ ، وغيرها (Shapiro 2008).

<u>5-الارتباط المضبب :-</u>

يمكن كتابة الارتباط المضبب بموجب الصيغة الآتية (Pashayev et al ,2006) :

$$\tilde{r}_{xy} = \frac{\sum \tilde{x} \otimes \tilde{y} - \frac{1}{n} \sum \tilde{x} \otimes \sum \tilde{y}}{\tilde{R}_x \otimes \tilde{R}_y}$$

اذ أن :   علامة ضرب مضبب : $\otimes$

$$\tilde{R}_x = \sqrt{\sum \tilde{x} - \frac{1}{n}\left(\sum \tilde{x}\right)^2}$$

$$\tilde{R}_y = \sqrt{\sum \tilde{y} - \frac{1}{n}\left(\sum \tilde{y}\right)^2}$$

هذه العلاقة المضببة يمكن وصفها كما يأتي : أذا كانت هناك اعتمادية دالية خطية تامة بين المتغيرين ($\tilde{y}, \tilde{x}$) فأن :- ($\tilde{r}_{xy} \cong \tilde{1}$ (*fuzzy one*)) وإذا العلاقة مستقلة فأن :- ($\tilde{r}_{xy} \cong \tilde{0}$ (*fuzzy zero*)) وأن لم تكن الاعتمادية بين المتغيرين تامة كان الارتباط المضبب ($\tilde{r}_{xy} \lesssim \tilde{1}$) .

6- الجانب التطبيقي:-

البيانات في الجدول الآتي تمثل خمسة متغيرات من مكونات دم الإنسان مع العمر إلى 51 شخص لتجربة في مستشفيات كركوك قامت بها الباحثة (*) للمراجعين من المرضى باليرقان أو الأصحاء .

جدول رقم (1) مكونات دم الإنسان مع العمر بوحدات القياس المناسبة

مكون الدم العمر Y	الكريات الحمر x1	الهيمكلوبين x2	حجم الكريات الحمر x3	الخلايا البيض x4	ترسيب الكريات الحمر x5
18	0.22	0.47	1.55	0.76	6.45
8	0.82	1.67	4.03	1.04	5.55
6	0.24	0.94	2.91	2.48	4.58
5	1.04	1.48	5.17	0.96	1.75
12	0.9	1.02	3.95	1.43	4.74
13	0.92	1.04	4.46	1.62	4.82
17	1.01	1.33	4.96	2.38	4.93
7	0.31	0.64	2.93	0.79	2.01
10	0.83	1.07	4.02	1.93	3.85
19	1.04	1.1	5.11	2.47	5.31
20	0.42	0.53	1.77	1	5.06
27	0.58	1.16	1.96	1.54	7.61

2.43	2.12	2.67	0.81	0.57	33
2.48	1.43	5.96	1.79	0.65	38
5.52	1.51	3.52	0.97	0.54	30
6.32	1.2	2.34	0.71	0.47	25
5.73	1.62	3.73	1.02	0.53	31
5.34	1.48	3.12	0.93	0.51	29
3.82	1.92	4.07	1.22	0.56	34
8.14	2.04	3.26	0.93	0.51	36
6.22	1.62	4.25	1.03	0.42	35
4.85	1.96	2.16	1.58	0.62	28
5.56	2.43	2.13	1.38	0.26	14
3.94	1.12	3.18	0.96	0.34	7
2.91	0.82	4.81	0.88	1.02	5
8.52	1.18	3.01	0.91	0.44	40
8.87	1.37	3.28	1.13	0.47	45
2.64	1.52	2	0.66	0.46	49
1.88	0.59	6.43	0.97	0.67	57
1.94	0.6	2.66	0.93	0.46	41
4.87	0.94	4.85	0.95	0.5	52
5.02	1.26	5.12	0.98	0.56	54
2.25	0.82	3.24	0.72	0.45	43
3.14	0.86	3.52	0.76	0.48	44
7.8	1.34	5.73	1.03	0.61	55
3.55	1.07	3.81	0.85	0.49	40
4.06	0.9	4.23	1.02	0.48	46
8.13	1.18	5.51	1.16	0.54	52
7.56	3.1	2.02	0.61	0.49	62
9.04	1.77	3.33	0.95	0.47	66
4.32	3.38	5.56	1.48	0.46	75
1.97	0.44	5.39	0.9	0.67	60
3.81	1.16	3.46	0.88	0.66	61
8.6	3.22	3.82	0.75	0.48	72
7.83	2.35	4.46	1.14	0.58	68
8.26	3.14	4.15	1.09	0.5	71
9.21	3.42	3.38	0.92	0.46	76
7.31	1.84	3.82	0.83	0.65	64
3.92	1.86	3.42	0.92	0.69	63
7.44	3.18	4.29	1.04	0.48	70
8.62	2.81	3.92	0.96	0.54	67

(*) المصدر: (فاتن عبد الواحد مجيد ، (2005) ،" دراسة مقارنة لأنواع اليرقان وتأثيرها على بعض مكونات الدم في مدينة " كركوك رسالة ماجستير علوم حياة - كلية التربية للبنات ، جامعة تكريت .)

وبالطبع فأن هناك خطأ في أجهزة القياس كان مقدارها تقريبا 1% اعتبرناها نسبة للضبابية أما العمر ففيه نسبة غير محددة من الضبابية . وهدفنا هو تطبيق بيانات حقيقية للانحدار المتعدد المضبب ، باستخدام طريقة المربعات الصغرى الضبابية وفق العلاقة (3) الفقرة (4)، فتم تشكيل المصفوفات (C , D , E , F) التي تمثل انتشار القيم لليمين واليسار لتطبيق العلاقة المذكورة وكانت كما يأتي :

$$E = \begin{bmatrix} 17 \\ 8 \\ . \\ . \\ 66 \end{bmatrix} \qquad \mathrm{F} = \begin{bmatrix} 19 \\ 9 \\ . \\ . \\ 68 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0.2178 & 0.4653 & 1.5345 & 0.7524 & 6.3855 \\ 1 & 0.8118 & 1.6533 & 3.9897 & 1.0296 & 5.4945 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & 0.5346 & 0.9504 & 3.8808 & 2.7819 & 8.5338 \end{bmatrix}$$

$$D = \begin{bmatrix} 1 & 0.2222 & 0.4747 & 1.5655 & 0.7676 & 6.5145 \\ 1 & 0.8282 & 1.6867 & 4.0703 & 1.0504 & 5.6055 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & 0.5454 & 0.9696 & 3.9592 & 2.8381 & 8.7062 \end{bmatrix}$$

وباستخدام دالة الانتماء المثلثية ، فكان لدينا فيما يخص المتغير المعتمد Y وهو العمر فقد تم تضبيبه باستخدام الفترة ± 1 سنة (فرضية الباحثين) أي أن الانتشار حول قيمة المركز سنة واحدة فإذا كان على سبيل المثال العمر 18 سنة فان قيمة الحد الأعلى تكون 19 سنة وقيمة الحد الأدنى تكون 17 سنة. أما بالنسبة للمتغيرات المفسرة (مكونات الدم) فعلى اعتبار هناك خطا في الأجهزة المختبرية أو ربما خطأ ناشئ من الفنيين الذين يعملون على الأجهزة حيث تكون نسبة الخطأ 1% المصرح بها من الأجهزة لتعبر عن نسبة الضبابية لقياسات المتغير X ، وأيضا باستخدام دالة الانتماء المثلثية.

وبناء على ذلك فقد تم استخدام نسبة الضبابية 1 % و 2% ، وقد استخرجنا معلمات نموذج الانحدار المضبب المتعدد للتجربة المذكورة ، المعبرة عن انحدار العمر كمتغير مضبب إلى مكونات الدم المضببة باستخدام البرنامج الجاهز Mat LAB كما في الجدول الآتي :

جدول رقم (2) المعلمات لنموذج الانحدار مع نسبة الضبابية للمتغيرات التفسيرية

نسبة الضبابية	a0	a1	a2	a3	a4	a5
1%	18.6964	-39.4720	-24.1713	10.8967	9.0094	2.1041

2.1071	9.0087	10.8983	-24.1443	-39.4400	18.6306	2%

ويلاحظ أن بزيادة نسبة الضبابية تقل القيم للمعلمات بنسبة قليلة. وبذلك يمكن كتابة أنموذج الانحدار المضبب للحالة الأولى كما يأتي :-

$$\tilde{Y}_i = 18.6964 - 39.4720\tilde{X}_1 - 24.1713\tilde{X}_2 + 10.8967\tilde{X}_3 + 9.0094\tilde{X}_4 + 2.1041\tilde{X}_5$$

والأنموذج للحالة الثانية :-

$$\tilde{Y}_i = 18.6306 - 39.4400\tilde{X}_1 - 24.1443\tilde{X}_2 + 10.8983\tilde{X}_3 + 9.0087\tilde{X}_4 + 2.1071\tilde{X}_5 .....$$

أما نموذج ألانحدار التقليدي فكان كالآتي :-

$$Y_i = 18.8407 - 39.5248\,X_1 - 24.2192\,X_2 + 10.8897X_3 + 9.0078X_4 + 2.0972X_5$$

ويلاحظ ان النموذج المضبب قريب من النموذج التقليدي ، وقد يدل هذا على قلة ضبابية البيانات وان التحليل الاعتيادي ببيانات جازمة التي تمثل مركز انتشار القيم ، ملائم لها .

المصادر

1 - هندوش ، رنا وليد بهنام ، (2003) ، دراسة عن النمذجة المضببة مع تطبيقات ، رسالة ماجستير رياضيات ،كلية الحاسبات والرياضيات ،جامعة الموصل ،2003.

2 – محمد ، جاسم محمد ، التقديرات الحصينة للانحدار الضبابي ، (2006) ، أطروحة دكتوراه إحصاء ، كلية الإدارة والاقتصاد – جامعة بغداد .

3- Bernd Moller ,Wolf gang Grat ,Michael Beer,and Jan-Uwesickcrt ,(2002) ,"Fuzzy Randomness-Word congress on Computational Mechanics" , Vienna,Austria.

4-Diamond ,Pill ," Fuzzy least squares",(1988) Information Sciences 46 , (141-157)

5-Gang Xiang,Valdik kreinovich , Scott Ferson, (2007)," Fitting anormal distribution to interval and fuzzy data" , University of Texas at EL Paso , USA.

6- Pashayev,A.M., Ardil C. , Askerov D.D., Saiqov ,R. A. , and Abdullayev , P.,S.,(2006), "Aircraft gas turbine engines "Technical condition identification system" (using fuzzy statistical analysis) ,International journal of applied mathematics and computer sciences ,vol 2,number 4.

7-Sanli ,Kamile ,and Apaydin ,Aysen , (2004)," The fuzzy robust regression analysis, the case of fuzzy data set has outlier", Journal of science, 17 (3):71-84.

8-Shapiro , Arnold F. (2005),"fuzzy regression and the term structure of Interest rates revisited " ,University Park ,PA16802, USA

9- Shapiro , Arnold F. (2008),"A fuzzy random variable primer", Smeal college of Business,Penn State University,University Park,PA 16802, USA.

10- Taheri ,S., Mahmoud, (2003) ,"Trends in fuzzy statistics " Austrian journal of statistics , vol 32,Number 3, 239-257 .

11- Tanaka, H., Uejima, S.and Asai, K., (1982), Linear Regression Analysis with Fuzzy Model. IEEE Trans. Systems, Man, Cybernet. 12, 903–907.

12- Celmins,A .,(1987) , "Least squares model fiting to fuzzy vector data ", Fuzzy Sets and Systems 22, 245-269 .