الجمهورية العربية السورية

جامعة دمشق

كلية العلوم

قسم الإحصاء الرياضي

التَّصنيفُ الأتومَاتِيكِيُّ لِلإشَارَةِ القَلبِيَّة

Automatic Classification Of ECG

رسالة مقدمة لنيل درجة الماجستير في الإحصاء الرياضي

إعداد الطالبة :

علا الزعبي

باشراف:

الدكتور عزّات قاسم

أستاذ مساعد في قسم الإحصاء الرياضي

كلية العلوم - جامعة دمشق

الدكتور المهندس جورج كرّاز

مدرس في قسم الذكاء الصنعي

كلية الهندسة المعلوماتية - جامعة دمشق

2014-2015

الملخص:

يندرج موضوع هذه الرسالة في مجال ضبط جودة قياس الأجهزة التي تُستخدم لرسم الصورة البيانية الكهربائية للقلب ECG، ومدى تأثير عدد من المتغيرات المدروسة عليها، سنعتمد قاعدة البيانات MIT-BIH العالمية لاضطراب النظم القلبي (Arrythmia) نستخدم هنا في مجال التصنيف الأتوماتيكي الإشارة القلبية خوارزمية (G.Karraz,G.Magenes) بالاعتماد على برنامج MATLAB. حيث يتم ضمنها استخدام الشبكات العصبونية لفصل مكونات الإشارة القلبية، وبعدها يتم استخلاص المتغيرات منها وفصلهاعن بعض وهي (عدد النبضات الكلي، مجموعة النقاط $\{\mathrm{Q_i}\}$، مجموعة النقاط $\{S_i\}$، مجموعة النقاط $\{R_i\}$). $i = 1, \cdots ,48$.

ونطبق مبرهنة النهاية المركزية على كل مجموعة نقاط. وبهذا يتم عزل النقاط المتطرفة لكل من المجموعات : $\{\mathrm{Q_i}\}$ ، $\{S_i\}$ ، $\{R_i\}$ الواقعة خارج المجالات المفروضة. بعد تطبيق الخوارزمية على قاعدة البيانات لاضطراب النظم القلبي MIT-BIH العالمية نجد بأن القيم لا تتوزع جميعها ضمن المجالات المفروضة، أي يوجدعدد من النقاط المتطرفة، ليس فقط عند تسجيلات المرضى بل الأشخاص السليمين أيضاً. يتم تحديد المتغيرات المهمة للدراسة. والمتغير التابع الذي هوعبارة عن جودة الإشارة القلبية المرسومة (من ناحية نسبة الضجيج الموجود فيها) ثم تُطرح الأسئلة التالية : هل يوجد علاقة رياضية تربط بين هذه المتغيرات؟. وهل يمكن من خلال هذه العلاقة الحصول على تنبؤ مفيد في تحديد دقة قياس الأجهزة ومدى قدرة الضجيج على التأثير على صحة التشخيص؟، ثم تتم عملية النمذجة باستخدام برنامج SPSS، واختيار أفضل نموذج (انحدار خطي بمتغير تابع ومتغيرين مستقلين) يلائم هذه البيانات التي بحوزتنا.

SUMMARY

This thesis presents a study in quality control of ECG device, and how some studied variables can affect the signal. We will depend on the MIT-BIH database. we use here (G.Karraz,G.Magenes) algorithm in ECG analysis, which utilize neural networks for separating ECG features.Then we extract the parameters (the number of total pulses N_a,the points set $\{\mathrm{Q_i}\}$, the points set $\{S_i\}$, the points set $\{R_i\}$). After that we plot each points sets in a time series, and apply the central limit theorem on thepoints sets, so that we segregate the irregular points for every set that lie outside the assumed intervals. After applying the algorithm on the MIT-BIH dataset,we will find a number of irregular points not only in patients signals, but also in normal signals. Then we determine the important variables for the study. And the dependent variable that is the quality of the signal ECG (regarding noise ratio). Then the question asked : is there a relationship between these variables?. And can we obtain a useful forecasting in determining the accuracy degree of the device and its ability to influence the diagnostic, through this relationship ?. Finally, modeling process using spss was achieved. And we have chosen the best model (multiple linear regression model with one dependent variable and two predictors variables) that fits the dataset we have.

الفهرس

فهرس الأشكال 7

فهرس الجداول.. 8

جدول الاختصارات.. 8

مقدمة 9

الفصل الأول: مقدمة طبية

1.1 فيزيولوجيا وتشريح القلب.. 13

2.1 مرض اضطراب النظم القلبي(Arrhythmia) 15

الفصل الثاني : التحليل الأوتوماتيكي للإشارة القلبية

1.2 خوارزميات الكشف عن المعقد QRS 16

1.1.2 خوارزمية Pan and Tompkins.. 17

2.1.2 خوارزمية Suppappola and Sun 23

3.1.2 خوارزمية So H and Chan K L 25

4.1.2 خوارزمية Antti 27

5.1.2 الخوارزمية المطورة G.Karraz,G.Magenes............................ 32

2.2 اختيار الخوارزمية الأنسب للفصل بين مكونات الإشارة....................... 38

1.2.2 اختبار وتقدير خوارزميات الكشف عن QRS................................ 38

3.2 القسم البرمجي.. 40

4.2 علاقة المتغيرات مع اضطراب النظم الخطي.................................... 51

5.2 مبرهنة النهاية المركزية.. 51

الفصل الثالث : استراتيجيات البحث عن النموذج الملائم (المصنّف)

الطرق المستخدمة في الحيز الزمني 54

1.3 الشبكات العصبونية... 54

1.1.3 الشبكة ذات الطبقة المفردة.. 57

2.1.3 الشبكة العصبونية متعددة الطبقات.. 58

3.1.3 خوارزمية الانتشار العكسي 59

4.1.3 القيم المحلية الأفضل الأدوار... 61

5.1.3 فرط الملاءمة وخيار الأدوات التدريبية 62

6.1.3 خوارزميات Quasi-Newton 62

7.1.3 خوارزمية قاطع ذات خطوة واحدة.. 63

8.1.3 خوارزمية Levenberg-Marquardt................................ 63

9.1.3 خوارزمية Bayesian 64

الفصل الرابع : الانحدار الخطي

1.4 الانحدار الخطي البسيط.. 66

1.1.4الانحدار الخطي بمتغير مستقل واحد...................................... 66

2.1.4 العلاقات بين المتغيرات... 66

3.1.4 نماذج الانحدار واستخداماتها... 66

4.1.4 نموذج انحدار بتوزيع غير معروف لحد الخطأ.......................... 69

5.1.4 بيانات تحليل الانحدار... 71

6.1.4 نظرةعامة على تحليل الانحدار.. 72

7.1.4 تقدير دالة الانحدار.. 73

8.1.4 تقدير تباين حدود الأخطاء.. 80

9.1.4 نموذج انحداربخطأ طبيعي 83

2.4 الانحدار المتعدد 88

1.2.4 نماذج الانحدار المتعدد 88

2.2.4 نموذج انحدار خطي عام بدلالة المصفوفات........................... 94

3.2.4 مقدرات المربعات الصغرى 96

4.2.4 القيم التوفيقية والرواسب 96

5.2.4 نتائج تحليل التباين.. 97

6.2.4 استدلالات حول وسطاء الانحدار.. 101

7.2.4 استدلالات حول متوسط الاستجابة....................................... 103

8.2.4 رسومات الرواسب، تشخيصات أخرى وتدابير علاجية................ 104

9.2.4 الخطية المتعددة وتأثيراتها 106

الفصل الخامس : اختبار النتائج إحصائياً

1.5 بناء النموذج.. 108

2.5 النتائج ومناقشتها 112

3.5 الاستنتاجات 121

4.5 المقترحات والتوصيات 122

الملحقات 123

1) الجداول الإحصائية... 124
2) قائمة بالمصطلحات العلمية (انكليزي –عربي).......................... 125
3) قائمة بالمراجع العلمية 133

فهرس الأشكال

الصفحة	الرقم	العنوان
10	(1)	الإشارة القلبية ECG ومكوناتها الأساسية
13	(1.1)	موقع القلب
18	(1.2)	استجابة السعة لمرشّح المشتق الرقمي
19	(2.2)	العلاقة بين QRS وموجة التكامل المتحرك
22	(3.2)	أداء خوارزمية Pan and Tompkins
24	(4.2)	أداء خوارزمية Suppappola and Sun
26	(5.2)	أداء خوارزمية So H and Chan K L
29	(6.2)	طبولوجيا المرشح النظير
31	(7.2)	أداء خوارزمية Antti
33	(8.2)	أداء خوارزمية G.Karraz,G.Magenes
34	(9.2)	أداء مرحلة ماقبل المعالجة
38	(10.2)	أداء المعامل غير الخطي
40	(11.2)	مقارنة بين أداء الخوارزميات الخمس
40	(12.2)	مقارنة بين أداء الخوارزميات الخمس
46	(13.2)	رسم للإشارة(121) بالتوصيلتين
47	(14.2)	مراحل عمل الخوازمية على التوصيلة الأولى
48	(15.2)	مراحل عمل الخوازمية على التوصيلة الثانية
49	(16.2)	رسم مجموعات النقاط $\{\mathrm{Q_i}\}$، $\{S_i\}$، $\{R_i\}$ بالنسبة للزمن
50	(17.2)	رسم المجالات RR ضمن العينة
55	(1.3)	الشكل الرياضي لعصبون
57	(2.3)	بنية الشبكات العصبونية متعددة الطبقات
74	(1.4)	منهجية نموذجية لبناء نموذج الانحدار
110	(1.5)	الإشارة 200
118	(2.5)	المخطط التبعثري لقيم الرواسب مقابل المتغير التابع
119	(3.5)	المدرج التكراري للرواسب

فهرس الجداول

الصفحة	الرقم	العنوان
56	(1.3)	قائمة كاملة لدوال انتقال
98	(1.4)	جدول تحليل التباين
111	(1.5)	خرج اختبار كولموغورف-سميرنوف لـ Rn
112	(2.5)	خرج اختبار كولموغورف-سميرنوف لـ err
112	(3.5)	خرج الإحصاء الوصفي
113	(4.5)	خرج المتغيرات المدخلة
113	(5.5)	خرج ملخص عن النموذج
114	(6.5)	خرج مجاميع المربعات
115	(7.5)	خرج وسطاء النموذج
116	(8.5)	خرج تشخيص الخطية
117	(9.5)	خرج معامل ارتباط بيرسون
120	(10.5)	خرج اختبار كولموغورف- سميرنوف للرواسب

قائمة الاختصارات والرموز

ANN	الشبكات العصبونية الصنعية
ECG	الصورة البيانية الكهربائية للقلب
p-value	قيمة المعنوية المشاهدة
VIF	معامل تضخم التباين

مقدمة

1- الإحصاء الحيوي

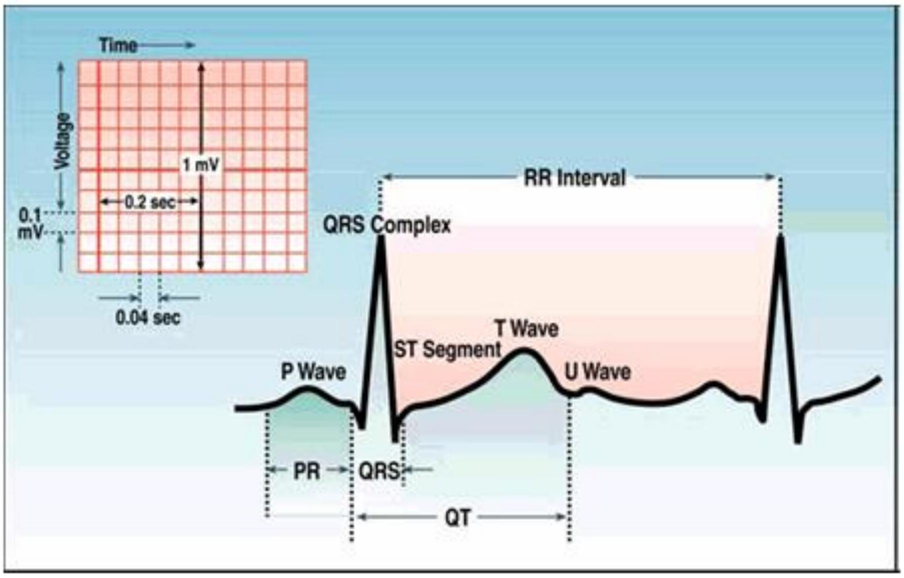
إن الإحصاء الحيوي هو فرع من فروع الإحصاء الرياضي والتطبيقي، ويهتم بالتفسيرات المناسبة لمعطيات بيولوجية أو طبية أو صحية.... إلخ. ضمن هذه العلوم، تُبدي مواد الدراسة وهي (المرضى، الفئران، الخلايا ....إلخ) تغيراً ملحوظاً باستجابتها للمنبهات، وهذا التغيُّر من الممكن أن يكون سببه المعالجات المختلفة، الصدفة، أخطاء القياس، أو ميّزات أخرى شخصية للأفراد. ويأخذ التصنيف الأتوماتيكي بجميع أشكاله (تحليل الإشارات، استخلاص السمات لها والتشخيص اعتماداً على هذا، جودة قياس الأجهزة) جزءً كبيراً ومهماً من الإحصاء الحيوي.

يُعنى الإحصاء الحيوي كباقي فروع الإحصاء بإيجاد المصادر المختلفة للتغيُّر، ويسعى للتمييز بين الارتباط والسبب. ولوضع استدلالات صالحة من عينات معروفة حول المجتمع الذي أُخذت منه. إن الإحصاء الحيوي هو نظام واسع يشمل تطبيق المفاهيم الإحصائية على مشكلات الحياة الواقعية، التدرُّب على تصميم وإجراء التجارب الطبية البيولوجية والتجارب السريرية، دراسة الخوارزميات الحسابية المرتبطة بذلك وعرض البيانات، وتطوير المفاهيم الإحصائية .

2- تمهيد

الصورة البيانية الكهربائية للقلب Electrocardiogram (ECG)

إن ECG هي تمثيل بياني للنشاط الكهربائي لعضلة القلب، وتعتبر تقنية غير جراحية مستخدمة كأداة جوهرية تشخيصية للكشف عن الأمراض القلبية الوعائية. أحد أهم هذه الأمراض اضطراب النظم القلبي Arryhthmia سنقوم في هذه الدراسة باستخدام قاعدة البيانات العالمية MIT-BIH لاضطراب النظم القلبي، حيث ستقوم الخوارزمية المستخدمة باستخلاص بعض المعالم الهامة من إشارة ECG .



الشكل (1) :الإشارة القلبية ومكوناتها الأساسية

الموجة P الدالة على النشاط الكهربائي الأذيني

والمعقد QRS والموجة T الدالتان على النشاط الكهربائي البطيني

والموجة U والقطعة ST

3-استخدام مجموعة بيانات قياسية عالمية لإشارة القلب (MIT-BIH)

منذ عام 1975 دعمت المخابر في مشفى Boston' s Beth (الآن مركز Beth الطبي الشمَّاسي) وفي MIT، الأبحاث لتحليل اضطراب النظم القلبي Arrhythmia ومواضيع مرتبطة به ، لقد كانت قاعدة بيانات اضطراب النظم القلبي MIT-BIH واحدة من النتاجات العظمى الأولى لهذه الجهود، والتي تم إكمالها وبدأ توزيعها في عام 1980، كانت قاعدة البيانات بشكل عام أول مجموعة متاحة لمادة الاختبار القياسية من أجل تقدير أدوات الكشف عن اضطراب النظم القلبي Arrhythmia، وتم استخدامها لهذا الهدف، بالإضافة إلى الأبحاث القياسية في الديناميكيات القلبية لأكثر من 500 موقع عالمياً. أولاً، تم توزيع البيانات على شريط رقمي نصف إنش بتسعة مسارات عند 800 و 1600 bpi، وعلى شريط نظير FM ربع إنش تصميم IRIG. في أغسطس 1989 ،أُنتجت نسخة CD-ROM لقاعدة البيانات. تحتوي قاعدة بيانات اضطراب النظم القلبي MIT-BIH على 48 جزءاً مقتطعاً لمدة نصف ساعة من تسجيلات ECG جوّالة بتوصيلتين، تم الحصول عليها من 47 مادة مدروسة بواسطة مختبر BIH لاضطراب النظم القلبي بين عامي 1975 و 1979، واختير 23 تسجيلاً بعشوائية من مجموعة من 4000 تسجيل جوَّال لـ ECG مدته 24 ساعة مختار من مجتمع مختلط من المرضى المقيمين

(حوالي 60%)، ومرضى خارجيين (حوالي 40%) من مشفى Boston' s Beth،

تم اختيار الخمسة وعشرين تسجيلاً المتبقّية من نفس المجموعة وذلك لتشمل اضطرابات النظم القلبي الأقل شيوعاً ولكن الأهم سريرياً، والذي لا يكون مُمثَّلاً بشكل جيد في عينة عشوائية صغيرة. تم ترقيم التسجيلات عند 360 عنصر في الثانية لكل قناة مع مستوى دقة 11 بت على مجال 10 MV . لقد علّق طبيبا قلب أو أكثر على كل تسجيل ، وتم حل الخلافات للحصول على تعليقات توضيحية مرجعية قابلة للقراءة بواسطة الحاسوب لكل نبضة (110000 تعليق توضيحي لدى الكل تقريبياً) مشمولة مع قاعدة البيانات.يحتوي الدليل الموجود في PhysioNet على قاعدة بيانات اضطراب النظم القلبي MIT-BIH الكلية. حوالي نصف(25 من 48 تسجيلاً كاملاً، وملفات تعليقات توضيحية مرجعية لجميع الثماني وأربعين تسجيلاً)لقاعدة البيانات هذه، أصبحت متاحة منذ بداية PhysioNet's في سبتمبر عام 1999. إن ملفات الإشارة الثلاثة والعشرين المتبقية، والتي كانت متوفرة على CD-ROM لـ MIT-BIH قاعدة بيانات Arrhythmia فقط ، نُشرَت في فبراير 2005.

أهمية البحث

لقد شهدت السنوات الأخيرة تطوراً ملحوظاً في حوسبة جهاز تخطيط القلب حيث تم وضع العديد من الخوارزميات بهدف التصنيف الأوتوماتيكي للمكونات الأساسية لإشارة القلب الكهربائية .

تلعب المتغيرات المستخلصة من الإشارة ECG دوراً جوهرياً في تشخيص الأمراض القلبية، إن هدف استخلاص المتغيرات هو إيجاد خصائص قليلة قدر الإمكان خلال إشارة ECG بحيث تؤمن كشف ناجح عن الشذوذ وتنبؤ فعّال. وهنا ستساهم هذه الخصائص في موضوع تحديد جودة أجهزة قياس الـ ECG، والتنبؤ فيما بعد بقيم جودة الإشارة المرسومة .وبهذا سيكون لها دور في دقة التشخيص بواسطة الجهاز المستخدم .

أهداف البحث

تعود المشاكل الأساسية التي تواجه الدراسات في مجال التصنيف الأوتوماتيكي لإشارة القلب الكهربائية، إلى صغر مجال التردد للإشارة القلبية، وقابليتها للتداخل مع الضجيج الناجم عن مؤثرات خارجية المنشأ كالتيار المتناوب، أو داخلية المنشأ أي جسم الإنسان (حركة المريض، الجهاز التنفسي ....).

يؤثر الضجيج على قراءة الإشارة، وإن محاولة إزالته يمكن أن يتسبب بضياع معلومات أساسية من الإشارة .

ولذلك سيكون هدف البحث هو تقدير نسبة وجود الضجيج في الإشارة(جودة الإشارة) من خلال المتغيرات التي أمكننا الحصول عليها، ويساهم ذلك في معرفة مدى دقة التشخيص، بحيث يصبح اللجوء إلى طرق تشخيصية أخرى ضرورياً في حال أثَّروجود الضجيج على التصنيف.

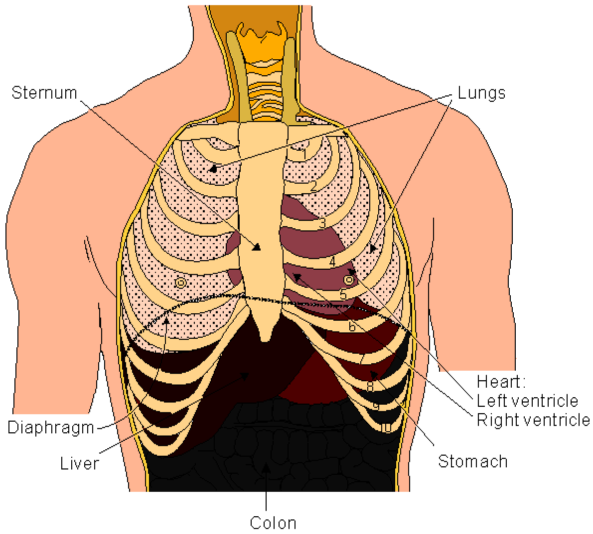
الفصل الأول:

مقدمة طبية

(G.Karraz,Doctoral Thesis (2007))

1.1 فيزيولوجيا تشريح القلب:

يقع القلب في الصدر خلف عظام القص وفوق الحجاب. إن القلب محاط بالتامور(وهو كيس رقيق يحيط بالقلب ويسمح له بالتحرك بسهولة)، وحجم القلب بحجم قبضة يد الشخص تقريباً، يتراوح وزن القلب بين250---300 غرام. يقع مركز القلب على بعد 1.5 سم من المستوي السهمي النصفي، يوجد فوق القلب الأوعية الكبيرة (الوريد الأجوف العلوي والسفلي)، الشريان الرئوي والوريد، بالإضافة إلى الأبهر. يقع القوس الأبهر خلف القلب، يقع المري والشوكة خلف القلب ولكن على مسافة أبعد. انظر إلى الشكل (1.1)



الشكل(1.1): موقع القلب

تشريح القلب :

تتكون جدران القلب من عضلة قلبية تدعى Myocardium، وتتكون أيضاً من تخطيطات كالتي في العضلة الهيكلية. يتألف القلب من أربعة أجزاء مستقلة الأّذينتين والبطينين (الأيمن والأيسر). إن القلب موجه بحيث يكون البطين الأيمن هو المنظر الأمامي، بينما يبين المنظر الخلفي الأّذينة اليسرى، تشكل الأذينتان مجموعة متكاملة واحدة، والبطينان مجموعة أخرى ولهذا الأمر أهمية خاصة من أجل الوظيفة الكهربائية للقلب .

يضخ البطين الأيسر الدم إلى الدوران الكبير(المجموعي) -حيث يكون الضغط أعلى من ضغط الدوران الصغير(الرئوي) بشكل ملحوظ- وهذا الدم ينشأ من التدفق البطيني الأيمن.
إن ألياف عضلة القلب موجهة بشكل لولبي (حلزوني) ومقسمة لأربع مجموعات، مجموعتان من الألياف تمتدان بشكل حلزوني على طول السطح الخارجي للبطينين، وتحت هذه الألياف تمتد مجموعة ثالثة حول البطينين، وتحت هذه الألياف تتوضع مجموعة رابعة ولكن فقط على طول البطين الأيسر. إن حقيقة أن خلايا العضلة القلبية موجهة بشكل مماسي أكثر مما هو شعاعي ومقاومة العضلة يكون أقل في جهة الليف، له أهمية في التخطيط الكهربائي للقلب. للقلب أربعة صمامات، يقع الصمام ثلاثي الشرف بين الأذينة اليمنى والبطين الأيمن، يقع الصمام التاجي بين الأذينة والبطين الأيسر. يقع الصمام الرئوي بين البطين الأيمن والشريان الرئوي، بينما يقع الصمام الأبهر في مجرى تدفق البطين الأيسر (متحكماً بالتدفق للأبهر). يعود الدم من الدوارن الكبير إلى الأذينة اليمنى، ومن هناك يذهب من خلال الصمام ثلاثي الشرف إلى البطين الأيمن. ينبعث الدم من البطين الأيمن عبر الصمام الرئوي إلى الرئتين. يعود الدم المؤكسج من الرئتين إلى الأذينة اليسرى، ومن هناك وعبرالصمام التاجي إلى البطين الأيسر.وأخيراً يكون الدم موجوداً ضمن الصمام الأبهر إلى الشريان الأبهر والدوران الكبير (المجموعي).

2.1 مرض اضطراب النظم القلبي Arrythmia

هو عبارة عن إيقاع شاذ لنبضات القلب، يمكن أن يحصل بسبب آلية عمل غير طبيعية للعقدة الجيبية، أو يمكن أن ينشأ من مناطق أخرى (والتي لا تبدأ فيها عادة الدفعات الكهربائية).

الأنماط الشائعة

يمكن تصنيف المرض :

- كضربات سريعة ومنتظمة للقلب (تسرع القلب Tachycardia)

- ضربات سريعة وغير منتظمة (الرجفان Fibrillation)

- ضربات بطيئة للقلب (بطء نظم القلب Bradycardia)

- وضربات إضافية شاذة تحدث قبل الضربات الطبيعية المتوقعة وبطريقة دورية (تقلص مبتسر Premature Contractions)

تدعى اضطرابات النظم القلبي الناشئة من حجرات القلب العلوية-فوق البطينين- (اضطرابات النظم القلبي فوق البطينية Supraventricular). وتلك التي تنشأ من الحجرات القلبية السفلية - البطينين- تدعى (اضطرابات النظم البطينية). تعتمد الأعراض التي يسببها المرض على نمط و شدة ومدة وتكرار المرض. وترتبط شدة الأعراض أيضاً بالوظيفة الأساسية للقلب. يمكن للشخص اليافع الذي يملك قلباً طبيعياً أن يتحمل اضطراب نظم القلب بشكل جيد، ولكن لن يستطيع شخص مسن ولديه أمراض قلبية أخرى مرتبطة أن يتحمل اضطراب النظم بنفس المقدار. إن الأعراض المرتبطة باضطراب النظم القلبي ليست ثابتة وتتضمن : الخفقان (الشعور بضربات سريعة للقلب)، ضعف، تعب، إزعاج في الصدر، ضيق النفس، فترات دوار، وفي أشد أشكالها تكون فقدان حقيقي للوعي وتوقف القلب.

الفصل الثاني :

التحليل الأوتوماتيكي للإشارة القلبية

1.2 خوارزميات الكشف عن المعقد QRS

مقدمة:

الكشف عن QRS

يُعتبر تحليل الصورة البيانية للقلب ECG أداه مهمة في تنظيم أمراض القلب، إنَّ من أهم مهمات التحليل الأوتوماتيكي لـ ECG هوالكشف ووضع سمات لكل موجه، وخاصةً المعقَّدQRS. إن المعقد QRS هو أكثر شكل ملحوظ للموجات خلال الـ ECG، بما أنها تُعبِّرعن النشاط الكهربائي داخل القلب خلال الانقباض البطيني، يعطي- توقيت حدوثها بالإضافة إلى شكلها- معلومات أكثر عن الوضع الحالي للقلب. يتضمن المعقد QRS عناصر إشارة بحزمة إشارة واسعة نسبياً من حوالي 2-HZ100 مع قمة عند 10-HZ15. ويعتبر المعقد QRS بسبب شكله المميّز أساس القرار المتحرك لمعدل القلب كنقطة إدخال لنماذج تصنيف الدورة القلبية، وغالباً يستخدم في خوارزميات ضغط بيانات ECG. تبيّن التجربة-من خلال عدة سنوات-أن الاستراتيجيات المقترحة لتحليلECG وبشكل خاص الكشف عنQRS المعتمدة على تقنيات معالجة الإشارة قد وصلت لأداء كشف قريب، وذلك بشكل رئيسي سببه تعدد الحالات التي تواجهنا في البيئات المرضية. لقد تم اقتراح عدد من خوارزميات الكشف عن QRS، سنذكرها في هذا الفصل مع الخوارزمية المطوّرة(G.Karraz,G.Magenes) المستخدمة في الدراسة بالإضافة إلى مقارنة لهذه الخوارزمية مع الخوارزميات الأربع الأخرى، لقد تم اختيار أداة الكشف في الاختبار، بأخذ اعتبار التعقيد والفعالية مع تفضيل الأقل تعقيداً مقابل الفعالية. يُجرى التقييم هذا على التسجيلات المتوفرة من قاعدة بيانات MIT-BIH .

الخوارزميات

1.1.2 خوارزمية Pan and Tompkins للكشف عن QRS

(Pan.J and Tompkins.W.J 1985)

تُعرّف هذه الخوارزمية معقدات QRS اعتماداً على التحليل الرقمي لميل وسعة وعرض بيانات ECG. تُطبّق الخوارزمية مرشح تمريرحزمة رقمي خاص، بإمكانه التخفيض من الكشف الخاطئ الذي سببته الأنواع المختلفة من الضجيج الموجود في إشارة ECG. تسمح هذه التقنية باستعمال العتبات المنخفضة وبهذا نزيد من حساسية الكشف.تعدّل الخوارزمية تلقائياً العتبات والوسطاء بشكل دوري لتتكيف مع التغييرات في تكوين QRS ومعدل القلب. وباختصار تتكون الخوارزمية من الخطوات التالية:

1- التنقية بطريقة ممررالحزمة:

إن هذا المرشح هو من 5-12 HZ لكنه مكون من مرشحات low pass و high pass.
إن دالة الاشتقاق لهذه المرشحات (1.2) و (2.2) على الترتيب:

$$Y(nT) = 2y(nT - T) - y(nT - 2T) + x(nT) - 2x(nT - 5T) + x(nT - 12T) \quad (1.2)$$

$$y(nT) = y(nT - T) - \frac{x(nT)}{32} + x(nT - 16T) - x(nT - 17T) + \frac{x(nT - 32T)}{32} \quad (2.2)$$

حيث أنه في هذا الفصل كله ستكون الرموز كالتالي :

x هي سعة بيانات ECG عند العنصر المفرد n

T هو متجه المعطيات

أي x(nT) هي العينة المدخلة

وستكون y(nT) هي العينة المخرجة

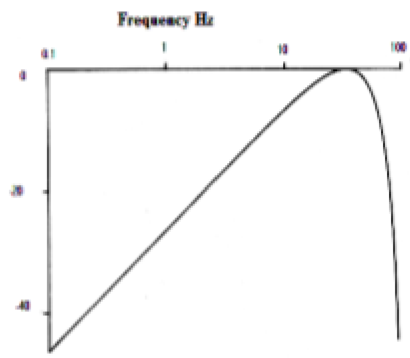
2- الاشتقاق:

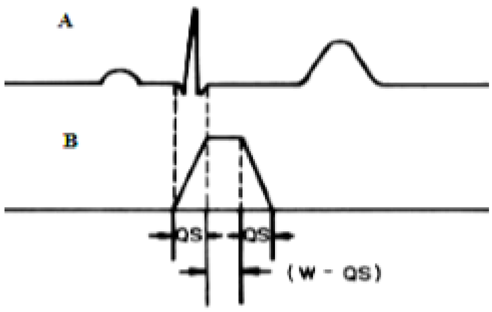
بعد التنقية تُفاضل الإشارة للتزويد بالمعلومات عن ميل المعقد QRS. تم استخدام مشتق لـ 5 نقاط مع المعادلة التالية:

$$y(nT) = \left(\frac{1}{8}\right)[2x(nT) + x(nT - T) - x(nT - 3T) - 2x(nT - 4T)] \quad (3.2)$$

يبين الشكل (1.2) استجابة التردد لهذا المشتق تكون تقريباً خطية بينdc و HZ30

الشكل(1.2) :استجابة السعة لمرشح المشتق الرقمي





الشكل(2.2): العلاقة بين معقد QRS وشكل موجة التكامل المتحرك

(A) إشارة ECG

(B) خرج مكامِل النافذة المتحركة

QS هي عرض QRS ، Wهو عرض نافذة التكامل

3- دالة التربيع:

بعد المفاضلة يتم تربيع الإشارة نقطة بعد نقطة وهذا يجعل كل نقاط البيانات موجبة، ويصنع تكبير غير خطي لنتيجة المشتق مؤكداً الترددات العالية أي بالدرجة الأولى ترددات ECG.

انتهت نتيجة دالة التربيع عند قيمة عظمى هي 255 .

4- تكامل النافذة المتحركة:

إن الهدف من تكامل النافذة المتحركة هو الحصول على معلومات عن سمة شكل الموجة بالإضافة إلى ميل الموجة R. يُحسب من العلاقة:

$$y(nT) = \left(\frac{1}{N}\right)[x\{nT-(N-1)T\} + x\{nT-(N-2)T\} + \cdots + x(nT)] \quad (4.2)$$

حيث N عدد العينات في عرض نافذة التكامل.

تعليق : يبين الشكل(2.2) العلاقة بين شكل موجة تكامل النافذة المفتوحة ومعقدQRS.

إن عدد العينات N هام في النافذة المتحركة بشكل عام يجب أن يكون عرض النافذة المتحركة مثل أعرض معقد QRS ممكن، إذا كانت النافذة عريضة جداً فإن النموذج سوف يدمج المعقّد T و QRS معاً، وإذا كانت ضيقة جداً فسوف تنتج بعض معقدات QRS ذرى متعددة في تموج التكامل، ومن الممكن أن يسبب هذا صعوبة في عملية الكشف اللاحقة عن QRS. يُحدّد عرض النافذة لمعدل الاختيار تجريبياً لـ 200 عينة سيكون عرض النافذة 30 عينة 50 ms.

5- تحديد نقطة الثقة وتعديل العتبات الحديّة:

يسعى المعقد QRS إلى الحافة الصاعدة من تموج التكامل، تكون هذه الحافة الصاعدة مساوية لعرض المعقد QRS، يمكن تحديد نقطة الثقة للوقع المؤقت للمعقد QRS من الحافة الصاعدة بحسب وسيط النموذج المراد وضعها كميل أعظم أو كقمة لموجة R، تُعدّل العتبات لتقوم فوق الضجيج أتوماتيكياً.

من الممكن الحصول على عتبات قليلة، بسبب تطور نسبة الإشارة إلى الضجيج بواسطة ترشيح ممرر الحزمة، تستخدم العتبة العليا بين العتبتين في كل من المجموعتين لأول تحليل للإشارة.

تستخدم الصغرى إذا لم يتم الكشف عنQRS في مجال زمني معين بحيث تصبح تقنية البحث للوراء ضرورية للنظر بالزمن للوراء بحثاً عن المعقد QRS.

إن مجموعة العينات المطبقة مبدئياً على تموّج التكامل، تحسب من خلال:

$SPKI=0.125\ PEAKI+0.875\ SPKI$ (if PEAKI is the signal peak)

$NPKI=0.125\ PEAKI+0.875\ NPKI$ (if PEAKI is the noise peak)

$THRESHOLD\ I1=NPKI+0.25(SPKI-NPKI), THRESHOLD\ I2=0.2\ THRESHOLD\ I1$

إذا كانت PEAKI هي ذروة الإشارة حيث تشير كل المتغيرات إلى تموّج التكامل

PEAKI هي الذروة الكلية.

SPKI هو التقدير الجار لذروة الإشارة.

NPKI هو التقدير الجار لذروة الضجيج.

THRESHOLD I1 هي العتبة الأولى المُطبَّقة.

THRESHOLD I2 هي العتبة الثانية المُطبَّقة.

ملاحظة :

إن الذروة هي قيمة عظمى محلية محددة من خلال مراقبة اتجاه تغير الإشارة خلال مجال زمني معرّف سابقاً، إن ذروة الإشارة SPKI هي ذروة منشأة سابقاً من الخوارزمية لتكون هي المعقد QRS. ذروة الضجيج هي أي ذروة غير مرتبطة بQRS (مثلاً موجة T). تعتمد العتبات على التقديرات الجارية لـ SPKI و NPKI أي(بحيث) تحسب قيم جديدة من هذه المتغيرات بجزء من قيمتها السابقة. عندما يتم الكشف عن ذروة جديدة يجب أولاً أن تُصنّف ذروة ضجيج أو ذروة إشارة. يجب أن تتجاوز الذروة THRESHOLD I1 لتكون ذروة للإشارة عندما تُحلل الإشارة للمرة الأولى، أو THRESHOLD I2 إذا كان البحث للوراء لازم لإيجاد QRS. عندما يتم إيجاد المعقد QRS باستخدام العتبة الثانية:

SPKI=0.25PEAKI+0.75SPKI

تُحدد مجموعة العينات المطبقة لتنقية ECG من خلال:

SPKF=0.125PEAKF+0.875SPKI(if PEAKF is the signal peak)

NPKF=0.125PEAKF+0.875NPKF(if PEAKF is the noise peak)

THRESHOLD F1=NPKF+0.25(SPKF-NPKF)

,THRESHOLD F2=0.5 THRESHOLD F1

حيث تشير جميع المتغيرات إلى الإشارة المرشّحة

PEAKF هي الذروة الشاملة.

SPKI هي التقدير الجاري لذروة الإشارة.

NPKF هي التقدير الجاري لذروة الضجيج.

THRESHOLD F1 هي العتبة الأولى المطبقة.

THRESHOLD F2 هي العتبة الثانية المطبقة.

إذا وُجدت QRS باستخدام العتبة الثانية:

SPKF=O.25PEAKF+0.75SPKF

لمعدلات القلب الشاذة، يتم خفض العتبة الأولى من كل مجموعة للنصف لنرفع من حساسية الكشف ولنتجنب النبضات المفقودة:

THRESHOLD I1=0.5THRESHOLD I1 و THRESHOLD F1=0.5 THRESHOLD F1

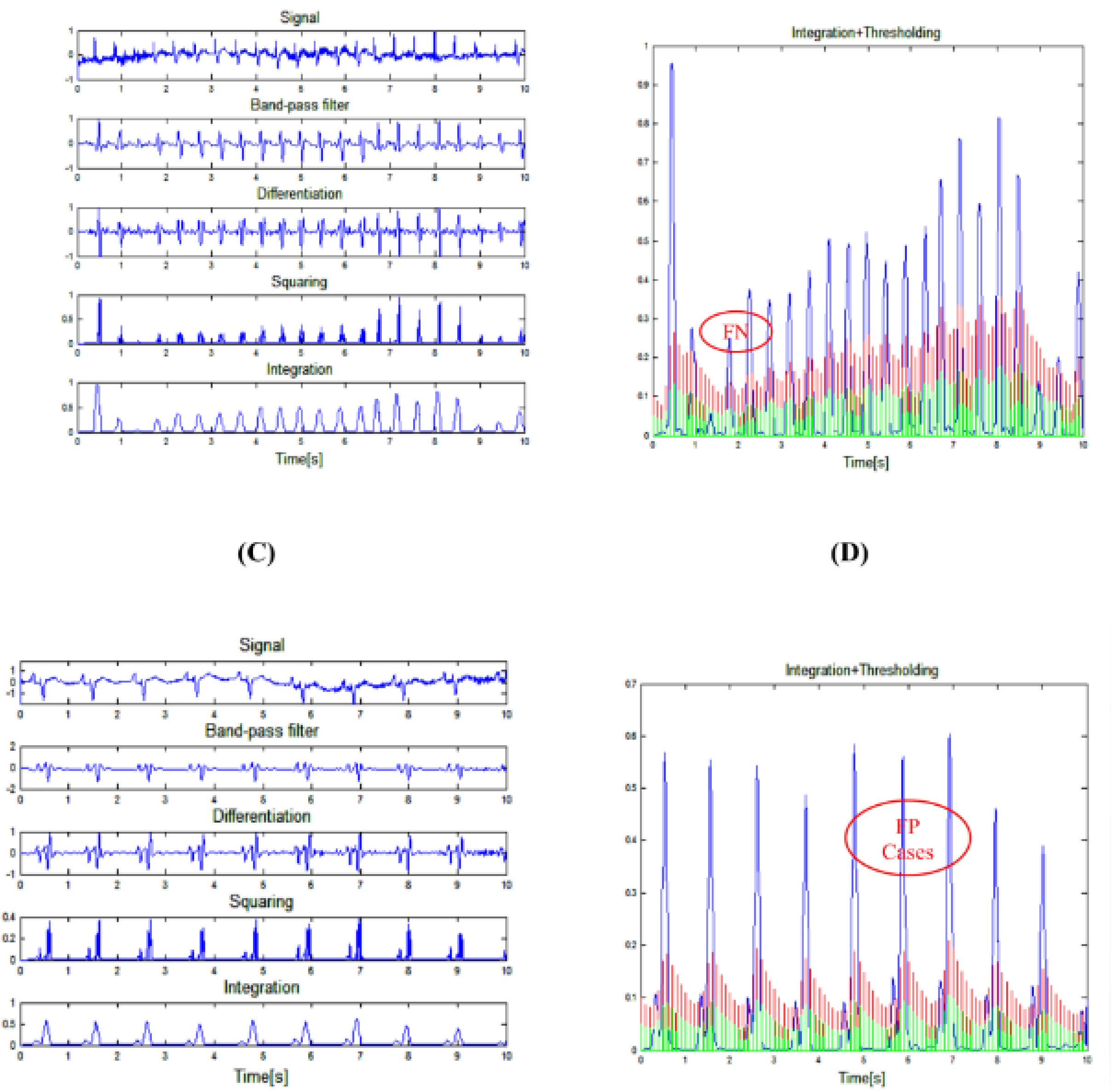
أُختبرت الخوارزمية الحالية على التسجيلات المتاحة من بيانات MITBIH وبيانات ECG

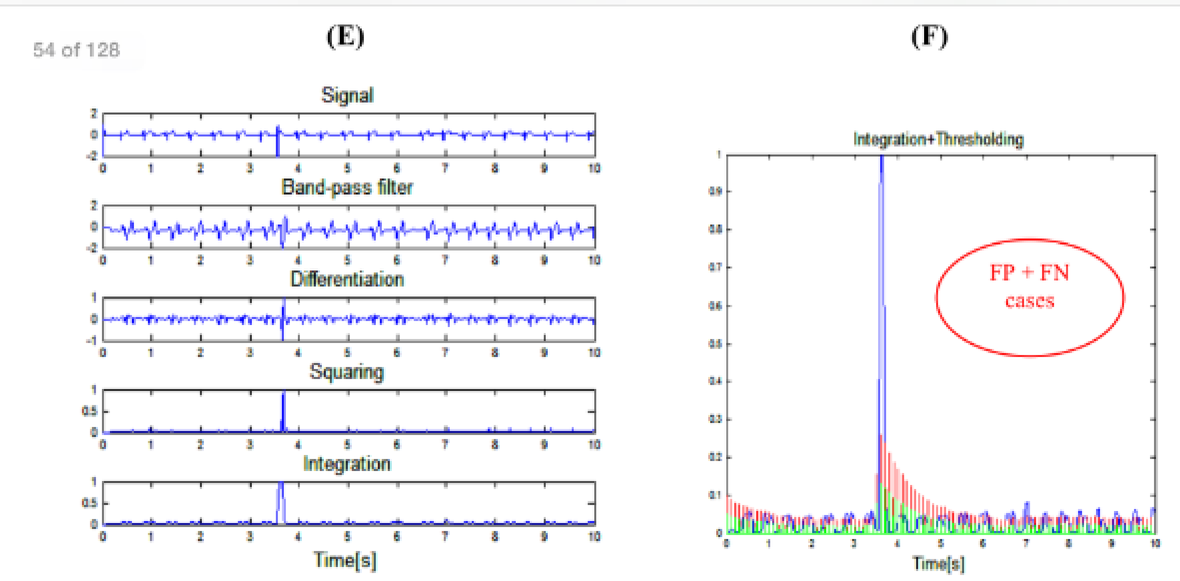
يوضّح الشكل (3.2) أداء خوارزمية Pan and Tompkins على مسارات عشوائية مأخوذة من التسجيلات 104 و 108 من بيانات MITBIH

والتسجيل E0110 من ST-T بيانات EUROPE

(A) (B)

(C) (D)





الشكل (3.2): أداء خوارزمية Pan &Tompkin:

A,B : مسار مأخوذ من التسجيل 104من قاعدة البيانات MIT-BIH، لاحظ أنه حدث نبض سلبي خاطئ.
C,D : مسار مأخوذ من التسجيل 108من قاعدة البيانات MIT-BIH ، لاحظ أنه حدث ثلاثة نبضات إيجابية خاطئة.
E,F : مسار مأخوذ من التسجيلE0110 من قاعدة البيانات الأوروبية، لاحظ أنه حدث نبضات إيجابية وسلبية خاطئة مختلفة. الخطوط المرسومة تُمثلTHRESHOLD I1 و THRESHOLD I2على الترتيب.

2.1.2 خوارزمية Suppappola and Sun للكشف عن QRS

(Suppappola and Sun 1991)

تستخدم هذه الخوارزمية تحويل غير خطي يدعى MOBD

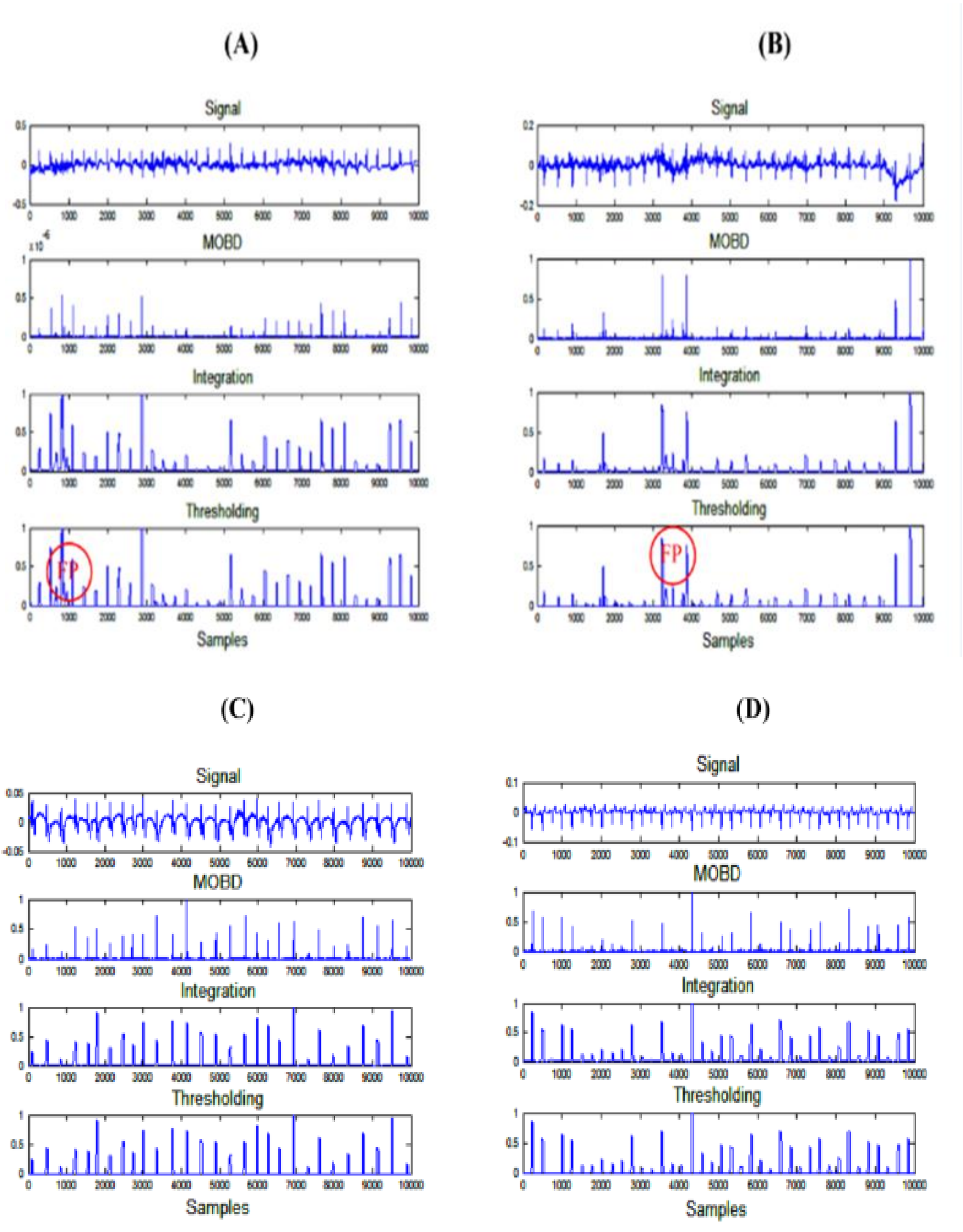
{ضرب الفروق العكسي(التراجعي)} والمشتق من الصيغة:

$$y(n) = \prod_{k=0}^{4} |x(n-k) - x(n-k-1)| \qquad (5.2)$$

يُطبَّق على نافذة مُعدَّل تكامل 20 نقطة وبعدها تُؤدّى عملية الكشف عن QRS بمجموعة من العتبات المعدّلة،

وبسبب عدم وجود عمليات أخرى لما قبل معالجة الإشارة، يجب الكشف أيضاً عن الضجيج والأخطاء اليدوية في الإشارة المعلمية الناتجة.

يوضّح الشكل (4.2) أداء الخوارزمية الحالية على مسارات عشوائية مأخوذة من التسجيلات 104، 108 لبيانات MITBIH والتسجيلات E0110 و E0112 من البيانات الأوروبية.



الشكل (4.2) أداء خوارزمية Suppappola and Sun

تعليق على الشكل (4.2) **:** تُمثل A,B,C,D أداء الخوارزمية على مسارات عشوائية مأخوذة من:
التسجيلات 104 و108 من قاعدة البيانات MIT-BIH والتسجيلات E0112 و E0110 من قاعدة البيانات الأوروبية على الترتيب. لاحظ أن الحالات الإيجابية الخاطئة في : (التسجيل 104 من MIT-BIH) و(التسجيل 108 من MIT-BIH).
يمكن أن تحدث هذه الحالات الإيجابية الخاطئة ، عندما تُقاطع الإشارة من أنماط متعددة من الضجيج. ليست كافية لتخفيض أو إلغاء الضجيج و الأخطاء اليدوية.

3.1.2 خوارزمية So H and Chan K L للكشف عنQRS

(So H and Chan K L,1997)

تعتمد هذه الخوارزمية على الميل الأعظم لإيجاد العتبات المعدّلة لبيانات ECG، وهي تسعى لتنفيذها على شاشة ECG متحركة، أُبقيت متطلبات الحاسب عند مستوى منطقي،وبدون تخفيض الثقة، في المقطع التالي نشرح هذه الخوارزمية:

لتكن x تمثل سعة بيانات ECG عند العنصر المفرد n. نحصل على ميل موجة ECG من خلال:

$$Slope(n)=-2x(n-2)-x(n-1)+x(n+1)+2x(n-2) \qquad (6.2)$$

تعطى عتبة الميل(الانحدار) بالعلاقة:

$$slope\_thresh = \frac{thresh_{param}}{16} maxi \qquad (7.2)$$

عندما تحقق اثنتان من بيانات ECG المتعاقبة الشرط :

$$slope(n) > slope\_thresh$$

فإن بداية المعقد QRS قد تم الكشف عنها، يمكن وضع الوسيط thresh_param كقيم 2،4،8 أو 32

بعد الكشف عن بداية المعقد QRS فإن البحث عن النقطة العظمى قد تم ،ونعتبرها النقطة R عندها تُعدل MAXI عن طريق:

$$maxi = \frac{first\_max - maxi}{filter\_param} + maxi \qquad (8.2)$$

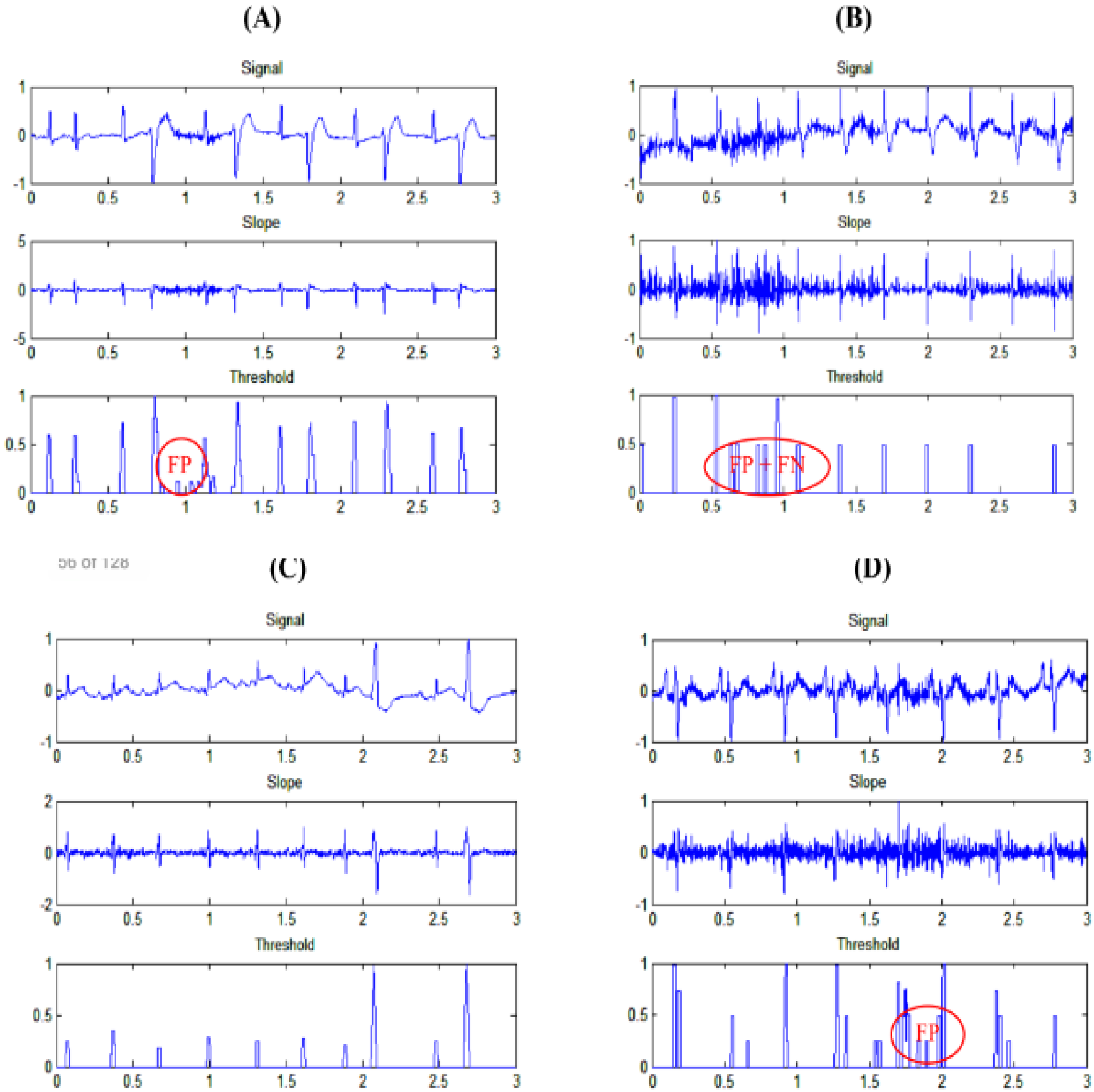
حيث أن first_max=ارتفاع نقطة R-ارتفاع بداية QRS

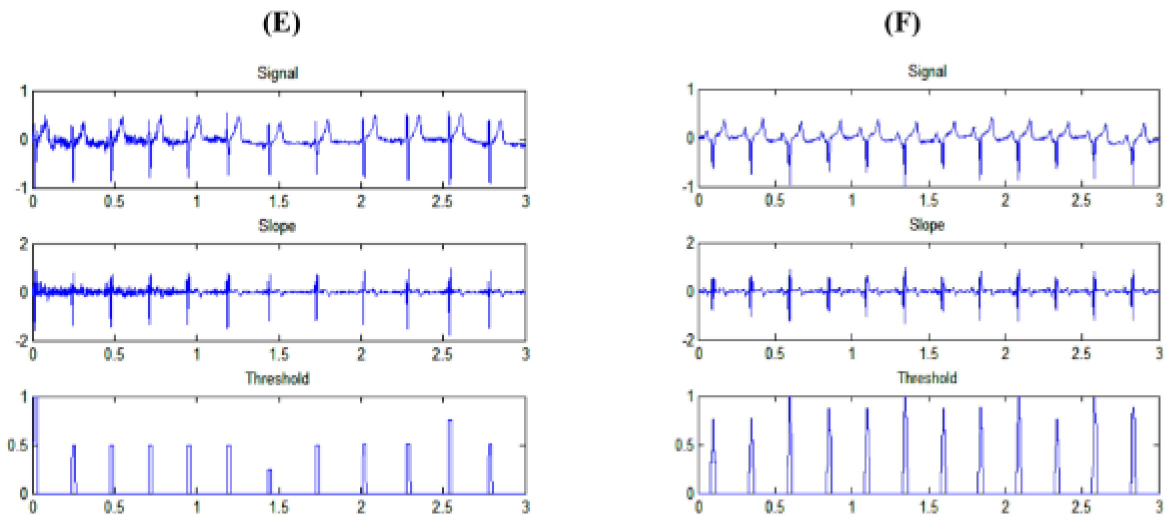
ويمكن وضع filter_param 8,4,2 أو 16

إن maxi الابتدائية الميل الأكبر لأول نقطة بيانات(عينات) في ملف ECG.

تم اختبار الخوارزمية الحالية على بيانات MITBIH و ST-T EUROPEAN

يبين الشكل(5.2) أدائها على بعض المسارات.





الشكل(5.2): {A,B,C,D,E,F}

أداء خوارزمية (So.H And Chan .K L)

تُمثِّل الأشكال A ,B ,C, D , E, F أداء الخوارزمية على مسارات عشوائية مأخوذه من التسجيلات 108،228،104،200 من قاعدة البيانات MIT-BIH

التسجيلات E0104 ،E0106 من قاعدة البيانات الأوروبية على الترتيب،

لاحظ الحالات الإيجابية الخاطئة في:

A: (التسجيل200 من MIT-BIH)،

B: (التسجيل 104من MIT-BIH)،

D: (التسجيل 108من MIT-BIH).

4.1.2 خوارزمية Antti للكشف عن QRS

(Antti et al 1997)

هي خوارزمية للكشف عن QRS ،تُوظِّف معالجة الإشارة في مناطق المرشح الرقمي، كما نُبيِّن في المخطط التالي:

ECG ← مرشح notch ← مرشح تمرير حزمة ← تطابق ← كشف

50 أو 60 40-15

1 - مرشَّح notch

يزيل أي تدخل لخط قوى الرواسب من الإشارة العينة،إن مرشح notch معرّف باستخدام طرائق التنقية الرقمية .

2-مرشح ممرر الحزمة

يُخفّض من الضجيج ذو التردد المنخفض، والذي من الممكن أن تسببه الأخطاء اليدوية الحركية، إن نتيجة مرشح ممرر الحزمة المختارة مع الضجيج الأبيض تسمح بدالة ملائمة لترشيح مطابق أفضل يعمل في المرحلة اللاحقة. يستخدم مرشح ممرر الحزمة الحالي عينتي إدخال مؤجلات، وعينتي إخراج مؤجلات لحساب النتيجة (الخرج)

$$y(nT) = x(nT) - 2x(nT - 16T) + x(nT - 32T) - 2y(nT - 8T) - y(nT - 16T) \tag{9.2}$$

3-المرشح النظير:

يؤمن هذا المرشح النسبة (الإشارة إلى الضجيج) الأفضل SNR (signal noise ratio) وفوق كل هذا تموج نبض مخرج متناظر، نحسب مخرجات المرشح النظير لطول استجابة نبض مرشح (n=30) كالتالي:

$$y(nT) = \sum_{i=0}^{30} k_i x(nT - iT) \tag{10.2}$$

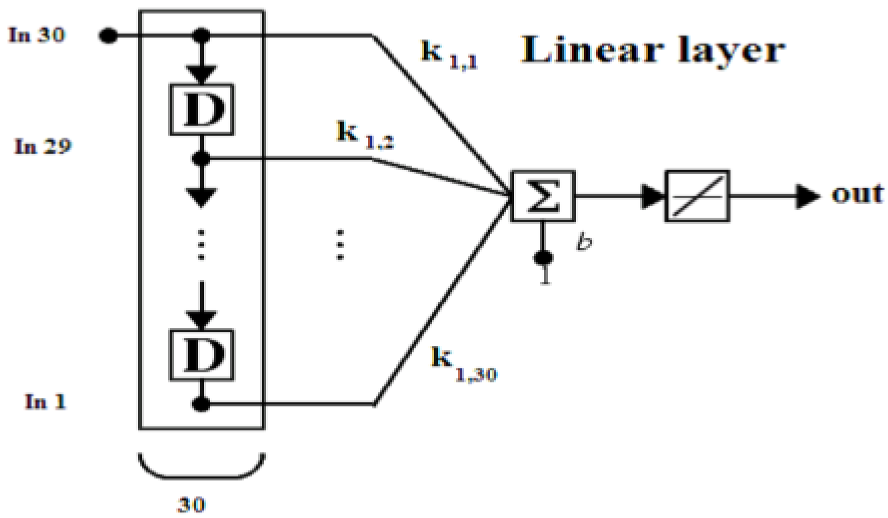
حيث أن $x(nT)$ هي العينة المدخلة

$y(nT)$ العينة المخرجة

k هو العامل المرشح

يتم إيجاد استجابة مع الترشيح النظير الأفضل لكل مريض، عن طريق اختيار معاملات المرشح في بداية القياس، وذلك باختيار تمثيل جيد للمعقد QRS المرشح بواسطة ممرر الحزمة.

لقد تم استخدام طريقة المربعات الصغرى كطريقة للحصول على الحل الأمثل لتصغير الخطأ بين المدخلات المختارة التي تمثل الإشارة المرشحة بممرر الحزمة مضروبة بمعامل يساوي 0.1 ، والإشارة المرشحة بممرر الحزمة كهدف.تتحقق هذه الطريقة باستخدام مرشح خطي معدل بنافذة مؤجلة مختارة مؤلفة من 30 عينة، انظر الشكل (6.2)



الشكل(6.2) طبولوجيا المرشح النظير

إن أشكال الموجة المرشَّحة النموذجية غالباً خالية من الضجيج،إشارات ECG المشوشة مبينة في الشكل (7.2)

تكون موجة النبض المخرج المتناظر للمرشح النظير واضحة، ولكي نؤمن فصل أفضل للزمن يتم تقدير الإشارة التي حصلنا عليها من خرج المرشح النظير كقيمة مساوية لثلاث أضعاف نسبة أخذ العينات الأصلي .

4-إيجاد العتبات:

يتم الكشف عن *QRS* من الإشارة المُنقيّة فيما بعد بواسطة كاشف عتبة معدّل، والذي يستخدم عتبة بحوالي 40% من القيمة العظمى الحاصلة في نتيجة (خرج) مراحل الترشيح خلال 1.5 ثانية الأخيرة من أجل 200 ms بعد كل كشف عن *QRS*.

تُرفع العتبة إلى 90% من القيمة العظمى المذكورة سابقاً لمنع الكشف الخاطئ

بسبب موجة T

يمكن وصف خوارزمية تعديل العتبة والكشف مستخدمين شيفرة مزيفة كالتالي:

تعريف المتغيرات

ENV : مغلف نتيجة المرشح النظير

THR : ثابت العتبة (0.6-0.4)

THR*s*: معامل العتبة

DET : إشارة كشف القيمة الثنائية

ETR : ثابت معدل ارتفاع المغلف

ETD : ثابت نسبة تلاشي المغلف

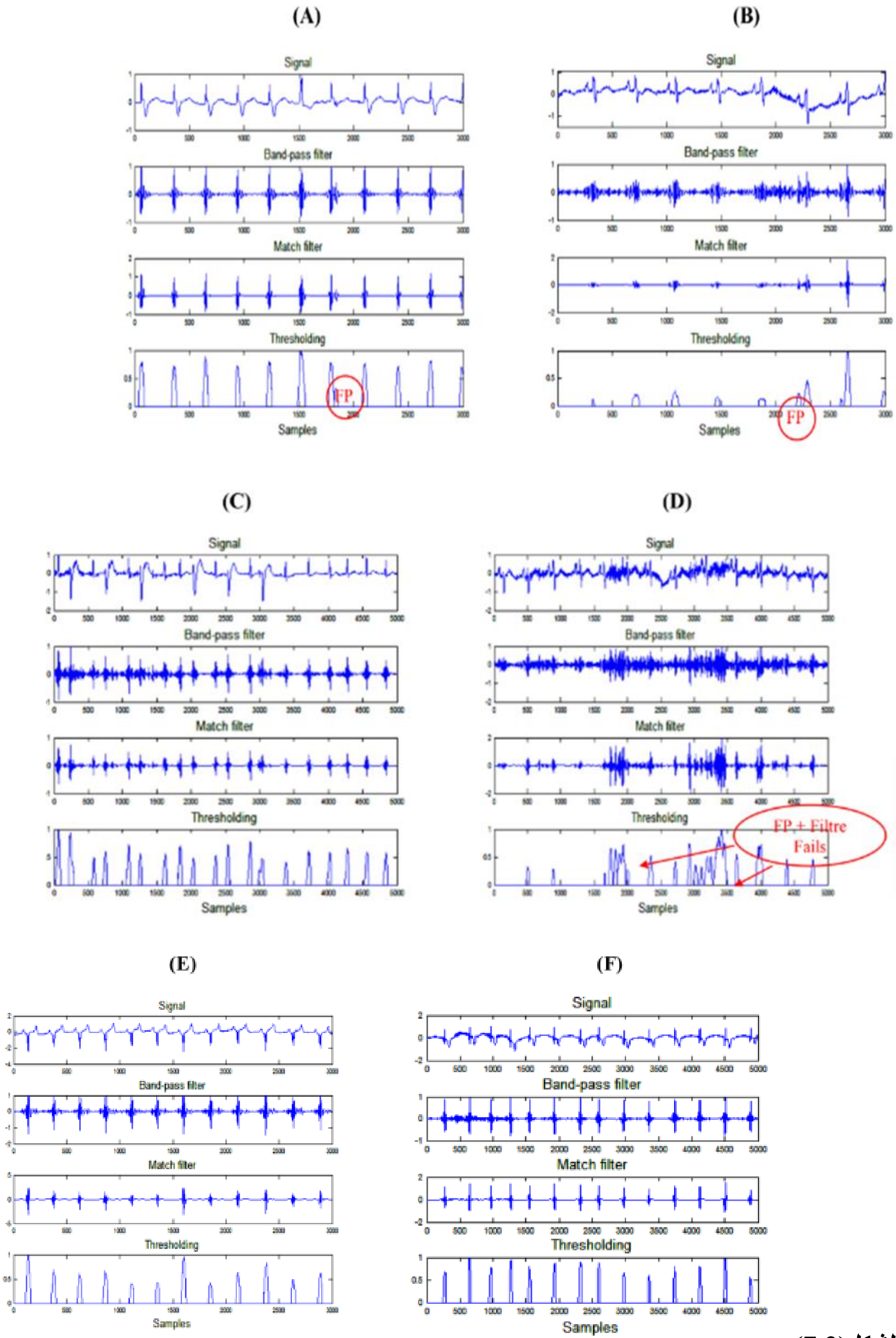
EHC : ثابت زمن إعاقة المغلف

TLAST-p : زمن بقاء عتبة الكشف (200)

تم اختبار هذه الخوارزمية على كل التسجيلات المتوفرة لبيانات MITBIH

وبعض التسجيلات ST-T لبيانات EUR،

يوضح الشكل(7.2) أداء الخوارزمية على مسارات عشوائية. لقد تم ملاحظة عدم كفاءة في أداء المرشحات المختارة لاستخلاص المعقدات *QRS* من العناصر السلبية الأخرى للإشارة المشوشة في إشارة مشوشة ما، وهذا لايقلل تأثيرات الأخطاء اليدوية.


(A)
(B)
Signal
Band-pass filter
Match filter
Thresholding
FP
Samples
(C)
(D)
FP + Filtre Fails
(E)
(F)

الشكل(7.2)

الشكل(7.2) : أداء خوارزمية Antti

تمثل أداء الخوارزمية على مسارات عشوائية مأخوذة من التسجيلات 104,108,200 من قاعدة البيانات MIT-BIH

والتسجيلات E0106,E0110,E0112 (في الشكل هي C,E,F) من قاعدة البيانات الأوروبية على الترتيب .

لاحظ الحالات الإيجابية الخاطئة في :

A (التسجيل 200من MIT-BIH)

B (التسجيل 104من MIT-BIH)

D (التسجيل 106من قاعدة البيانات الأوروبية).

هنا يفشل المرشّح بفصل QRS عن العناصر الأخرى في الإشارة المشوشة .

5.1.2 الخوارزمية المطوّرة(G.Karraz,G.Magenes)

لقد تم شرح جميع خوارزميات الكشف عن QRS ،وهناك أخرى مقترحة في المحاضرة:

Gritziali, 1988;Polli et ., 1995;Kadambe *et* al 1999;Daskalov) (and Stoyanov ,2004

تُبيّن أداء جيد عندما تكون *ECG* نقية،غير مشوشة بالضجيج، في حالات الإشارات المشوشة يكون لبعض هذه الخوارزميات أداء مقبول كخوارزمية Pan & Tompkin وبينما تفشل خوارزميات أخرى بشكل كامل وتُظهر أداء سيء مثل MOBD ولذا فإن هذا القسم سيبين حل المشاكل الملاحَظة في خوارزميات الكشف عن QRS عن طريق تطوير:

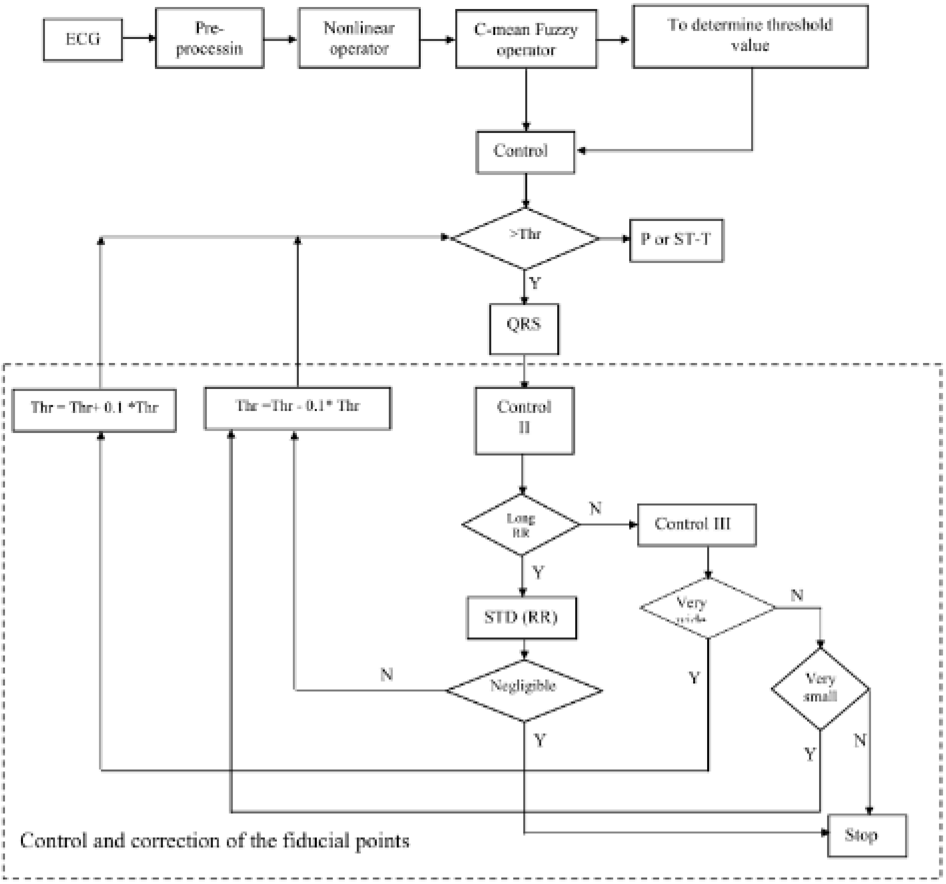
1- مرحلة ما قبل المعالجة: تضمن تخفيض ملحوظ أو إنهاء للضجيج المفروض على إشارة ECG.

2- مرحلة إيجاد العتبة: تضمن تفكيك عالي أو قدرة على التعرُّف على شكل موجة QRS من العناصر السلبية الأخرى.

تُطوّر هذه الخوارزمية الطريقة المقترحة بواسطة

(M.Pauletti and C.Marchesi ,2001)

ونوضِّح في الشكل(8.2) مخطط يُمثل مراحل خوارزمية الكشف عن *QRS* المطوّرة.



الشكل(8.2): الخوارزمية المطوَّرة للكشف عنQRS ممثّلة بشكل مخطط

1- قبل المعالجة

تهدف إلى تنقية وتخفيض الضجيج المطبق على ECG، تؤسس هذه المرحلة مهمة الكشف عن *QRS*.

أولاً: أُستخدم مرشح ممرر حزمة FIR لإزالة التنقلات المستمرة ومشروع الخط الأساسي،لقد تم تصحيح معاملاته بـ sptool.

ثانياً: أُستخدم مرشح notch لتخفيض تداخل خط القوة ذوالتردد50-60 HZ، وأخيراً تم إعادة اختيار العينة للإشارة للحصول على تردد عيني 1000، يوضح الشكل (9.2) أداء خطوة ما قبل المعالجة خلال مسار ECG مشوش.

2 -المعامل غير الخطي :

لقد تم تقديم معامل مبتكر غير خطي بواسطة (M.Pauletti and C.Marchesi.2001) إنه يعتمد على تقدير مستمر لطول منحنى مثالي مولد لعينات الإشارة.

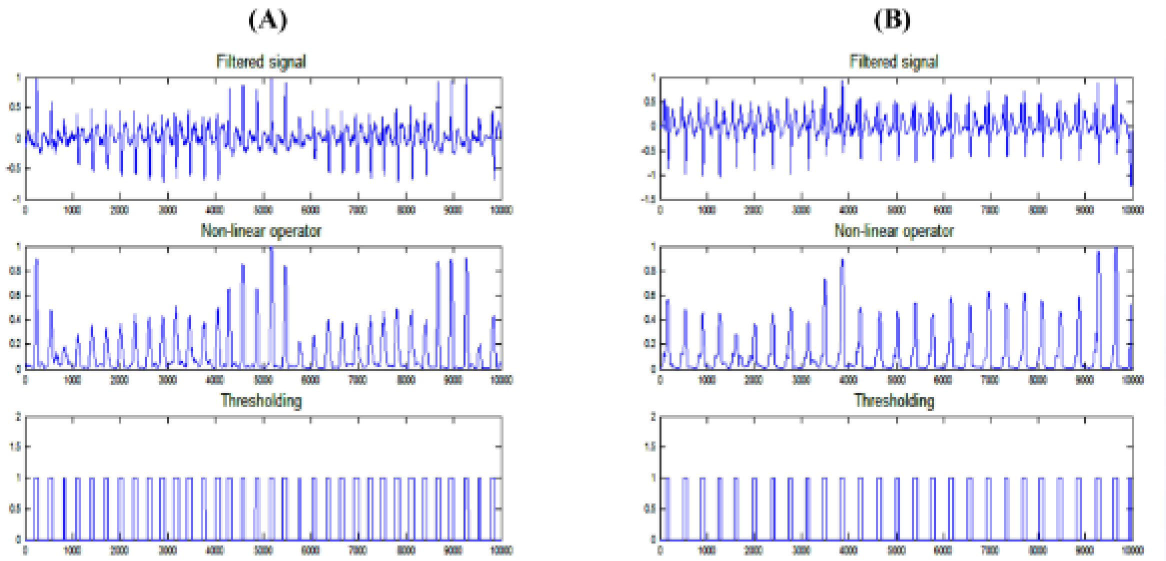
نستطيع في النطاق الزمني المنقطع تقريب طول القوس المتعلق بالعينة ذات الترتيب i لطول المستقيم فنحصل على:

$$U_i = D_i \sum_{w} \sqrt{1 + \frac{{D_y}^2}{{D_t}^2}} \qquad (11.2)$$

U هو متجه المعطيات

D_t هو المجال المختبر، D_y يمثل الزيادة رقم i

w هو تقدير تقريبي لمدة المرحلة (أو شكل الموجة) للكشف.



الشكل(9.2):أداء مرحلة قبل المعالجة

A,B هي مسارات عشوائية من الإشارات المشوشة مأخوذة من:

التسجيلات 104,108 من قاعدة البيانات MIT-BIH على الترتيب .

إن هذين التسجيلين مشهوران كإشارات مشوشة لاحتوائهما على أنماط مختلفة من الضجيج.

الآن كون D_t ثابت وبأخذ التربيع لتعزيز الأطراف أكثر (بمعنى مدة الخط تشير إلى الزمن) نحصل على:

$$\mathrm{U} = \sum_{w} \mathrm{D_y}^2 \quad \text{or even} \quad \mathrm{U_i} = \sum_{k=i-n}^{i+n} \Delta\emptyset_k^2 \qquad (12.2)$$

إذا كانت Δ هو معامل فرق منته و $w = 2n + 1$ نافذة المرشح، فيكون لدينا كشرط تمهيدي:

$$\mathrm{U_n} = \sum_{k=0}^{2n} (\mathrm{y_k} - \mathrm{y_{k-2}})^2 \qquad (13.2)$$

المعادلة التكرارية:

$$U_i = U_{i-1} - (y_{i-n-2} - y_{i-n})^2 + (y_{i+n-2} - y_{i+n})^2 \quad (14.2)$$

3- إيجاد العتبة

كما نرى هناك عنصران رئيسيان، الأول يسعى لمعقدات QRS والثاني يسعى لموجتي P,T والذي يزعج عملية الكشف عن QRS.

وللحد من ذلك ولإيجاد عتبة المعقدات QRS، نقترح طريقة عتبة جديدة تعتمد على خوارزمية Fuzzy clustering algorithm (S.Osowski et,2000)

لتكن U نتيجة المعامل غير الخطي، نريد تقسيمها إلى j عنقود (cluster)، وكل عنقود ممثل بمركزه $cent_j$ أي يعتبر هنا تصنيف عنقودي.

لتكن C تدل على مصفوفة التقسيم المؤلفة من الأسطر C_j حيث:

$$[C_j] = [c, \dots \dots, \mathrm{c_{jn}}]$$

وكل عنصر يمثل درجة انتماء متجه المعطيات U في العنقود ذو الترتيب j_n

$(i=1,..n)$ و $(j=1,\cdots J)$

تبحث خوارزمية fuzzy clustering عن مصفوفة التقسيم ومراكز العناقيد بحيث تم تصغير الدالة الهدف E :

$$E = \sum_{J=1}^{J} C_j^m \, d^2(U, cent_j) \qquad (15.2)$$

$$\sum_{j=1}^{J} C_j = 1$$

يتحكم الوسيط m بالعناقيد (عادة m=2)،

تقيس الدالة $dj = d(U, cent_j)$ المسافة بين متجه المعطيات U و j

مركز العنقود ذو الترتيب j

$$d^2(U, cent_j) = (U - cent_j)^t \, M_j (U - cent_j) \qquad (16.2)$$

حيث أن M_j قيمة معرفة موجبة معدلة للأشكال الفعلية للعنقود ذو الترتيب j. تُعرَّف M_j كـ:

$$M_j = \sqrt[n]{F_j} \; F^{-1} \qquad (17.2)$$

$$F_j = C_j^m (U - cent_j)(U - cent_j)^T \qquad (18.2)$$

حيث أن F_j هو تباين العنقود

بإعطاء متجه البيانات، نختار عدد العناقيد J (في حال 2 أو 3) معاملات الوزنm،

والاستطاعة النهائية ع. يتم تهيئة المصفوفة المقسمة C عشوائياً بطريقة يكون فيها الشرط 3 محقق، بعدئذ تتكرر الخوارزمية بالخطوات التالية:

1- تحديد نموذج العنقود

$$j = 1 \qquad cent_j = \frac{\sum C_j^m U}{\sum C_j^m} \qquad (19.2)$$

2- حساب F_j بحسب المعادلة (18.2)

3- حساب المسافة مربع dj بحسب (16.2)

4- تحديث مصفوفة التقسيم C_j

$$C_j = \frac{1}{(d_J)^{\frac{2}{m-1}}} \quad \text{if } d_J(i) = 0 \text{ then take } C_j(i) = 1$$

5- تكرار حتى

$$\|C - C^{-1}\| \leq \varepsilon$$

6- أخذ المتجه C_j الذي يسعى إلى المركز الأعلى للعناقيد، وإيجاد عتبة له عن طريق اختيار قيمة العتبة بحسب قيمة الانحراف المعياري (STD) لمتجه الإشارة:

$$if \;\; C_j(i) > std * \max(C_j) \rightarrow C_j(\mathrm{i}){=}1 \rightarrow QRS\ complex$$

$$else\ if \;\; C_j(i) \leq std * \max(C_j) \rightarrow C_j(\mathrm{i}){=}0 \rightarrow P\&T\ waves$$

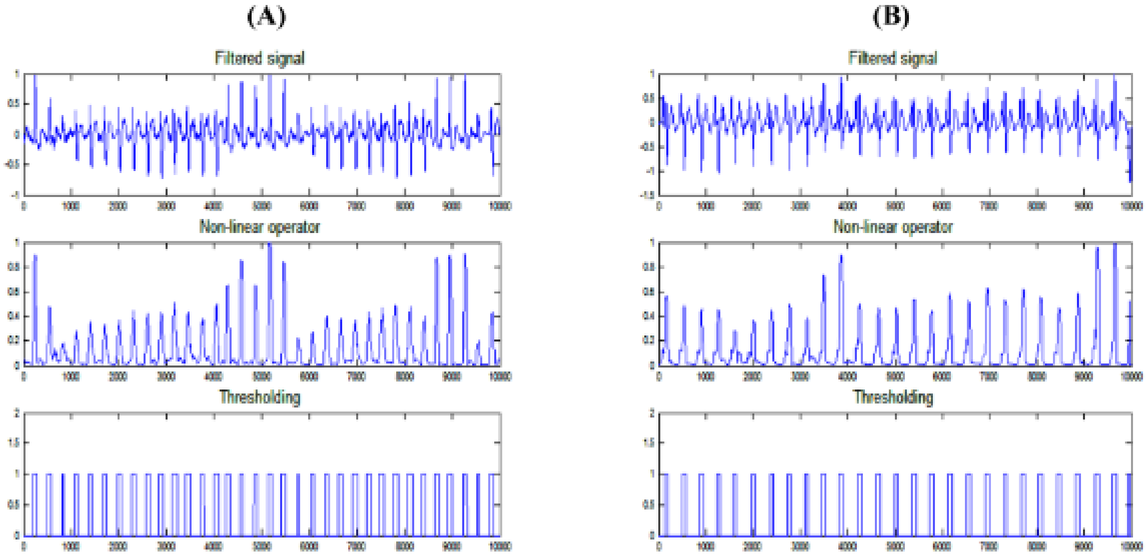
يُوضِّح الشكل (10.2) أداء هذه الطريقة على نفس الإشارات المذكورة في الشكل(9.2)

لقد تم إضافة إجرائية تحكّم إلى عملية الكشف عن QRS لـ:

1- التحكم ضمن الفترة التي تسعى إلى المسافة بين معقدات QRS التي تم الكشف عنها

(فترات الـ RR)،في حالة مسافة طويلة جداً بالنسبة إلى سابقتها(بالاعتماد علىSTD للإشارة) هذا يعني أنه يوجد موجات QRS مفقودة ولذا تبحث الخوارزمية عنها مستخدمة قيمة عتبة معدلة أخرى، أقل من السابقة ثم تعيد الإجرائية هذه الخطوة حتى يتم الكشف عن جميع معقدات QRS المفقودة.

2- تصحيح عرض كل المعقدات QRS التي تم الكشف عنها، لكي نتجاهل الصغيرات جداً بالنسبة لمتوسط الأخريات، يمكن للسعة العالية للموجة T أن تسبب QRS إيجابي خاطئ.



الشكل(10.2):أداء المعامل غير الخطي وعملية إيجاد العتبات على نفس مسارات ECG المرشحة المذكورة في الشكل **(9.2)**

2.2 اختيار الخوارزمية الأنسب للفصل بين مكونات الإشارة

1.2.2 اختبار وتقدير خوارزميات الكشف عن QRS

كجزء من الدراسة ننظر إلى أداء الخوارزميات المختلفة الخمسة للكشف عن QRS،والمطورة هي واحدة منها وقد تم ذكر شرح الأخريات فيما سبق.تم اختبار الخوارزميات الخمس باستخدام بيانات MITBIH.

إن البيانات الخام المنتجة بواسطة مخطط تصنيف خلال الاختبار هي تعدادات للتصنيفات الصحيحة والخاطئة من كل صنف.

وبعدها فإن هذه المعلومات تعرض بشكل طبيعي في المصفوفة المختلطة،إن المصفوفة المختلطة هي شكل من أشكال جدول الاحتمال، يعرض الفروق بين الأصناف الحقيقية والمتنبأ بها لمجموعة من البيانات المبوبة.

Tp,Tn هي عدد الحالات الإيجابية الصحيحة والسلبية الصحيحة على الترتيب.

Cn,Cp عدد الأمثلة الإيجابية والسلبية الحقيقية.

مجاميع الأعمدة Rn , Rp هي عدد الأمثلة الإيجابية والسلبية المتنبأ بها.

N تكون العدد الاجمالي للأمثلة $(N = C_n + C_p = R_n + R_p)$

مع أن المصفوفة المختلطة تعرض جميع المعلومات عن أداء المصنّف،يمكن استخلاص قياسات معنوية منها لتبين معايير أداء معينة مثلاً:

1) الثقة = (1- الخطأ)

$$P(C) = \frac{T_p + T_n}{C_p + C_n}$$

2) الحساسية (1-β)

$$P(T_p) = \frac{T_p}{C_p}$$

3) النوعية(1-α)

$$P(T_n) = \frac{T_n}{C_n}$$

4) قيمة تنبؤ موجبة $\frac{T_p}{R_p}$

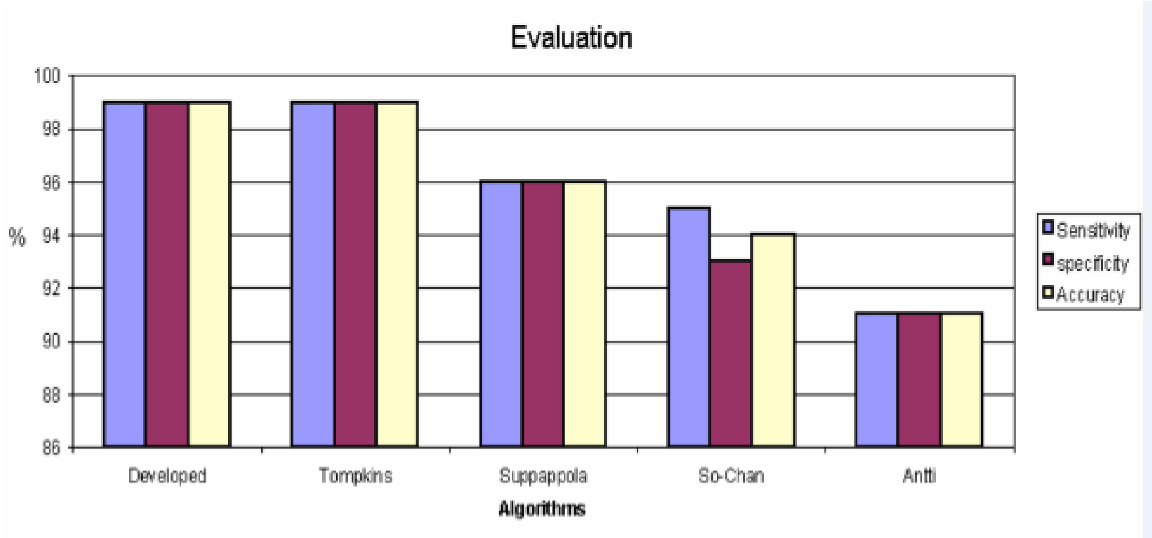
5) قيمة تنبؤ سالبة $\frac{T_n}{R_n}$

إن كل مقاييس الأداء هذه صالحة فقط لنقطة تشغيل واحدة، عادة ما تختار نقطة التشغيل لتقليل احتمال الخطأ أي إنتاج مصنّف بثقة كلية عليا، بشكل عام لا نريد تخفيض خطأ التصنيف بل تكلفة خطأ التصنيف،يُعرّف خطأ التصنيف عادة كالتالي:

$$cost = F_P C_{FP} + F_n C_{Fn} \qquad (20.2)$$

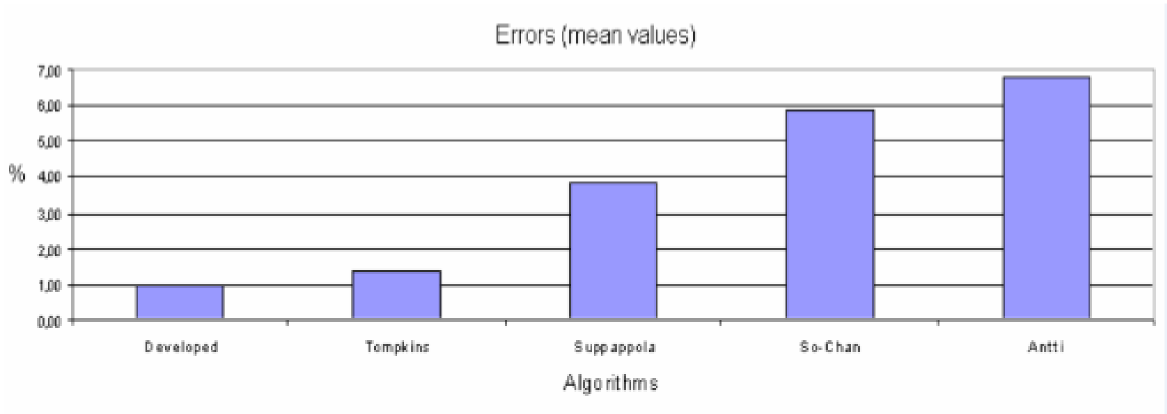
تُوضِّح الجداول النتائج مُمثِّلة المصفوفة المختلطة للخوارزميات المدروسة للكشف عن QRS

تُبيِّن الأشكال (11.2)و(12.2) مقارنة بين القيمة المتوسطة للحساسية، النوعية، الثقة والخطأ.



الشكل(11.2)

تقدير للحساسية والنوعية والثقة للخوارزميات الكشف عن QRS الخمسة على تسجيلات قاعدة البياناتMIT-BIH ،مُمثَّلة كقيم متوسطة.



الشكل(12.2)

يمثل الشكل(12.2) الخطأ المتوسط لخوارزميات الكشف عنQRS الخمس مقدراً على تسجيلات قاعدة البيانات MIT-BIH

3.2 القسم البرمجي

قمنا باختيار الخوارزمية الخامسة(G.Karraz,G.Magenes) لتطبيقها في الأطروحة على قاعدة البيانات MIT-BIH.وسيكون القسم البرمجي عبارة عن دالتين :

الأولى باسم peak_detection

تقوم بتقطيع الإشارة ورسمها بالتوصيلتين المتاحتين، بالإضافة إلى مراحل عمل الشبكة العصبونية المستخدمة.

```
function[Impulse1,Impulse2]=peak_detection(signal,fs,w,sn);
Time=signal((1:sn),1)./fs;
INT1(1:length(Time))=0;
INT2(1:length(Time))=0;
S1=signal((1:sn),2);
S2=signal((1:sn),3);
S1=((S1)./max(S1));
S2=((S2)./max(S2));
figure
subplot(2,1,1)
hold on
title('signal Lead I');
xlabel('Time (sec)');
ylabel('signal (Mv)');
plot(Time,S1);
grid
hold off
subplot(2,1,2)
hold on
title('signal Lead II');
xlabel('Time (sec)');
ylabel('signal (Mv)');
plot(Time,S2);
grid
hold off

%pause;
D1=diff(S1);
D2=diff(S2);
DD1=D1.^2;
DD2=D2.^2;
```

```
beg=round(w/2);
for i=(1+beg):(length(DD1)-beg)
   INT1(i)=sum(DD1(i-beg:i+beg));
   INT1(i)=INT1(i)./w;
end
INT1=((INT1-mean(INT1))./max(INT1));
for i=(1+beg):(length(DD2)-beg)
   INT2(i)=sum(DD2(i-beg:i+beg));
   INT2(i)=INT2(i)./w;
end
INT2=((INT2-mean(INT2))./max(INT2));
Impulse1=(INT1> (0.3*max(INT1)));
Impulse2=(INT2> (0.3*max(INT2)));


figure
hold on
subplot(3,1,1)
hold on
title('signal Lead I');
plot(Time,S1)
grid
hold off
subplot(3,1,2)
hold on
title('Derivation');
plot(Time,[0 D1'])
grid
hold off
subplot(3,1,3)
hold on
title('Squaring');
plot(Time,[0 DD1'])
grid
hold off
```

الدالة الثانية تحمل اسم PEX :

في دراستنا هذه ستعتمد فعالية الكشف الملائم على الاستخلاص المناسب للنقاط Q,R,S والمجال RR، لذا ستقوم هذه الدالة بالاستفادة من خرج الشبكة العصبونية، لفصل مكونات الإشارة التي ستفيدنا في دراستنا، حيث يقوم باستخلاص وفصل النقاط $\{Q_i\}$،$\{S_i\}$، $\{R_i\}$،من المعقدات QRS ثم رسم كل منها في سلسلة زمنية، وتطبيق مبرهنة النهاية المركزية على مجموعات النقاط لنحصل على النقاط المتطرفة خارج المجالات والتي سنستخدمها في الدراسة. ثم تقوم الدالة بإعطاء الخرج وهوعدد تلك النقاط. بالإضافة إلى عدد النبضات الكلية التي استطاعت الخوارزمية كشفها ضمن التردد العيني المختار.

```
function [R,S,Q,RR]=PEX(signal,imp,fs);
j=0;
k=0;
F=0;
for i=1:(length(imp)-1)
   if (imp(i)==0)&(imp(i+1)==1)
      j=j+1;
indQ(j)=i+1
   elseif (imp(i)==1)&(imp(i+1)==0)
      k=k+1;
      indS(k)=i;
   end
end
x1=length(indQ);
x2=length(indS);

if x1==x2 & indQ(1)<indS(1)
   for j=1:x1

indr=find(signal(indQ(j):indS(j))==max(abs(signal(indQ(j):indS(j)))))
;
     F=F+1;

     indR(F)=indQ(j)+indr(1);
   end
elseif ((x1~=x2)&(indQ(1)<indS(1)))|((x1~=x2) &(indQ(1)>indS(1)))
   indqq=indQ(2:(length(indQ)-2));
```

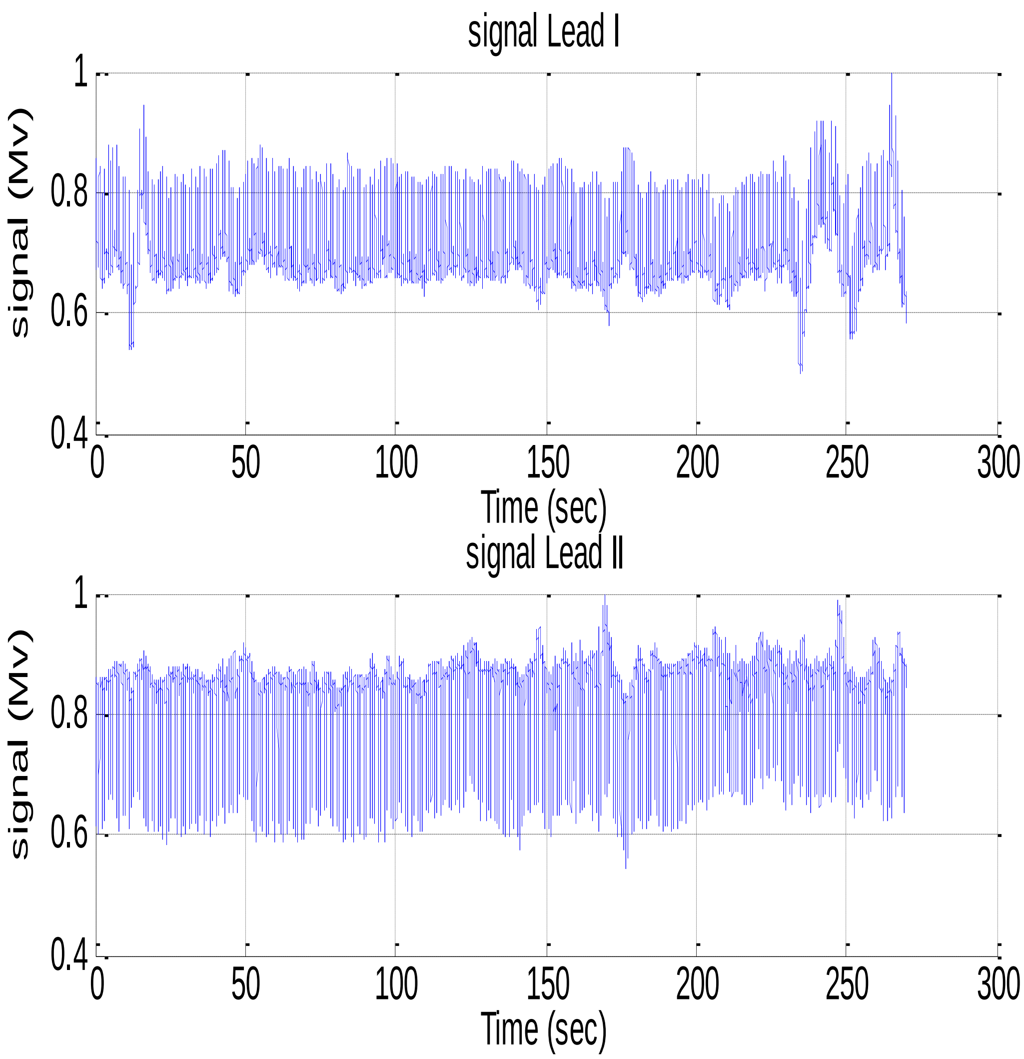
```
    indss=indS(2:(length(indS)-2));
    x1=length(indqq);
    for j=1:x1
      indr=find(signal(indqq(j):indss(j))==
max(abs(signal(indqq(j):indss(j)))));
      F=F+1;
      indR(F)=indQ(j)+indr(1);
     end
end

Q=signal(indQ);
S=signal(indS);
R=signal(indR);
RR=diff(indR);
TimeQ=indQ./fs;
TimeS=indS./fs;
TimeR=indR./fs;
'number of plulses is'
length(R)
mq=mean(Q);
stq=std(Q);
indirq=find((Q<(mq-(3*stq)))| (Q>(mq+(3*stq))));
'irrigular Q cases='
length(indirq)
ms=mean(S);
sts=std(S);
indirs=find((S<(ms-(3*sts)))| (S>(ms+(3*sts))));
'irrigular S cases='
length(indirs)
mr=mean(R);
str=std(R);
indirr=find((R<(mr-(3*str)))| (R>(mr+(3*str))));
'irrigular R cases='
length(indirr)
figure
hold on
subplot(3,1,1)
```

```
hold on
title('Q');
plot(TimeQ,Q)
grid
hold off
subplot(3,1,2)
hold on
title('S');
plot(TimeS,S)
grid
hold off
subplot(3,1,3)
hold on
title('R');
plot(TimeR,R)
grid
hold off
figure
hold on
title('RR');
plot(TimeR(2:length(R)),RR)
grid
hold off
end
```

مثال عن خرج الدالتين البياني لإحدى الإشارات :

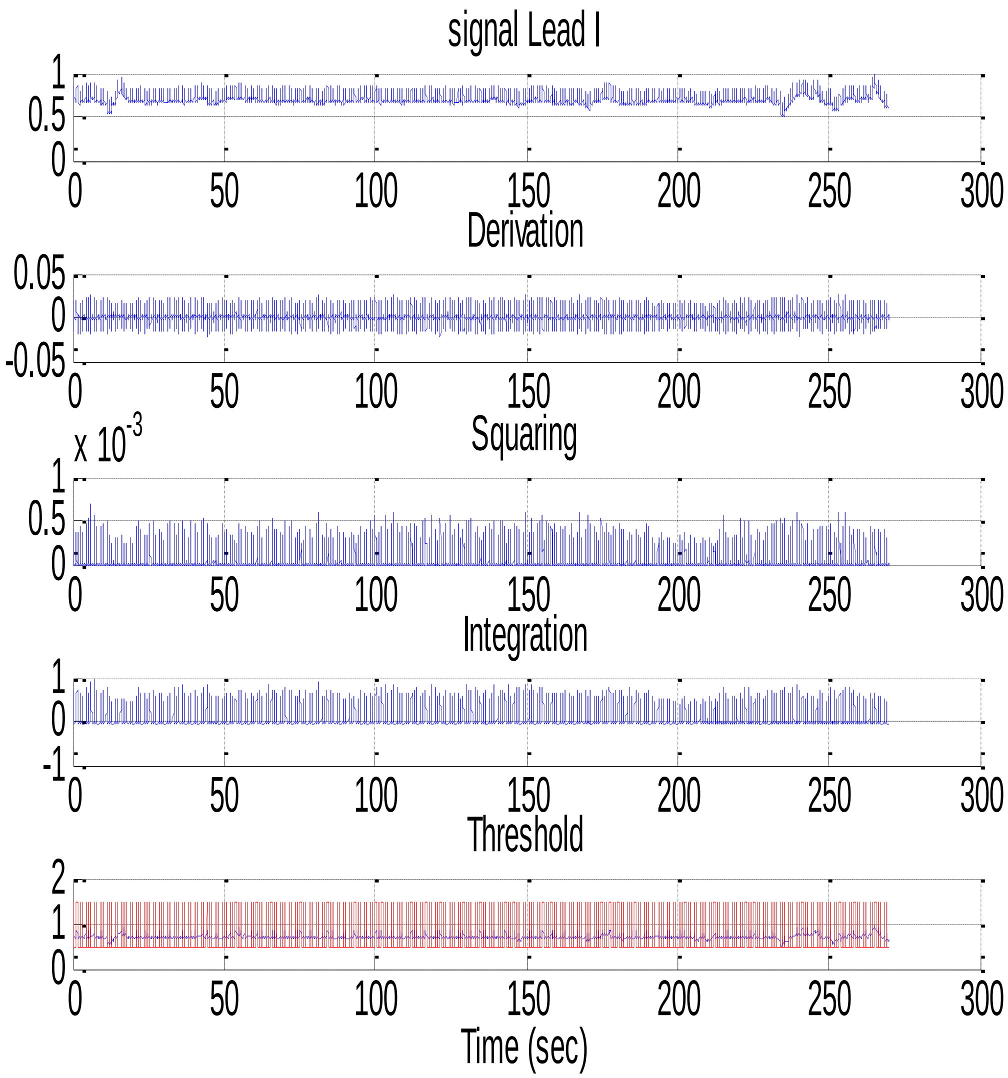
الإشارة (121)



الشكل (13.2) :

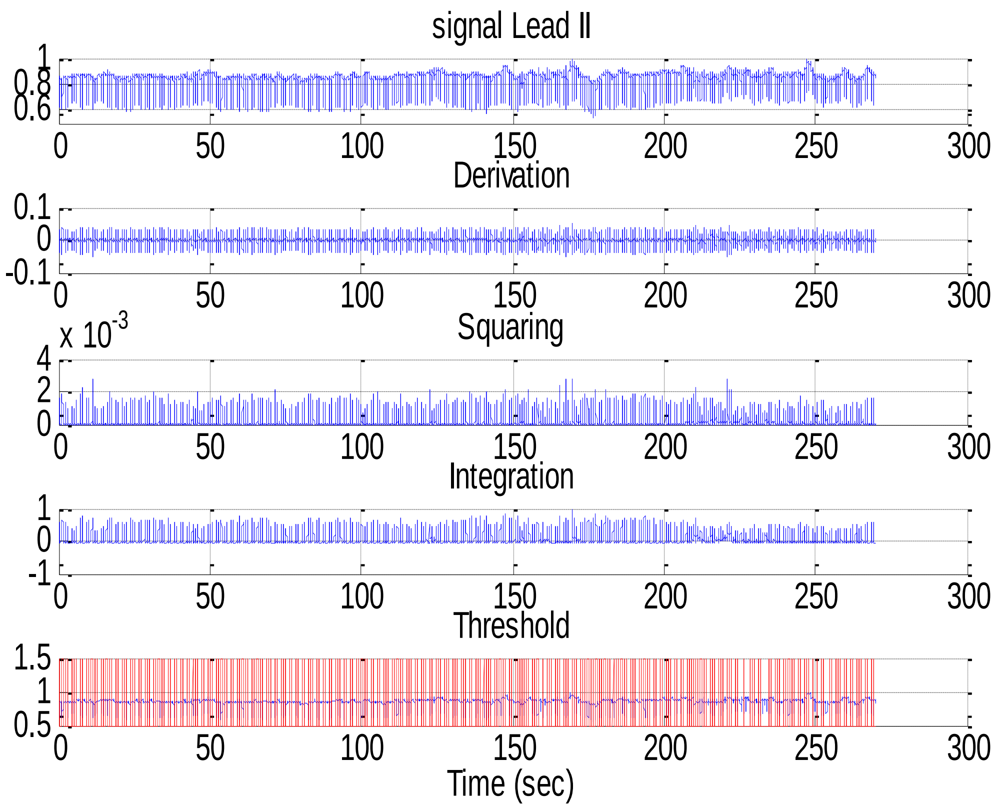
يوضح الشكل (13.2) رسم الإشارة كاملة على التوصيلتين الأولى والثانية

(lead I, lead II)



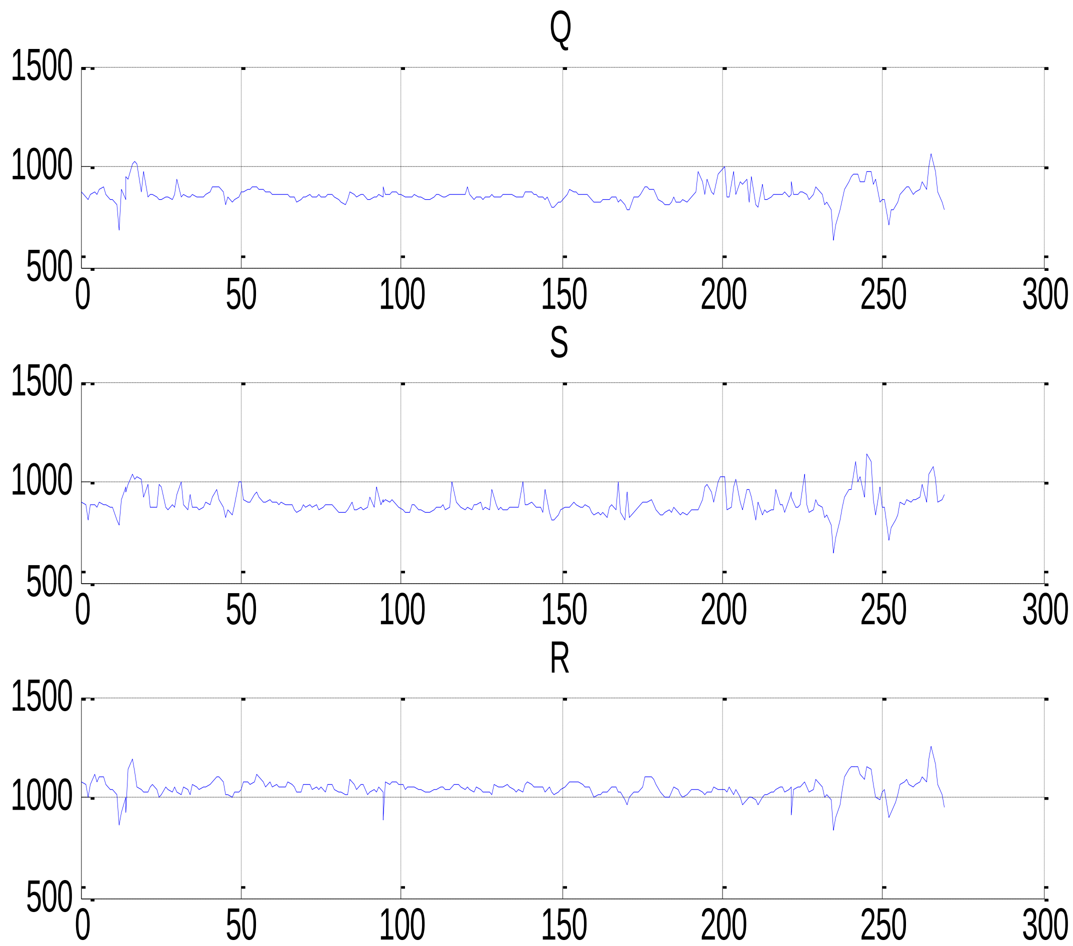
الشكل (14.2)

مراحل عمل الخوارزمية على الوصلة الأولى(lead I)



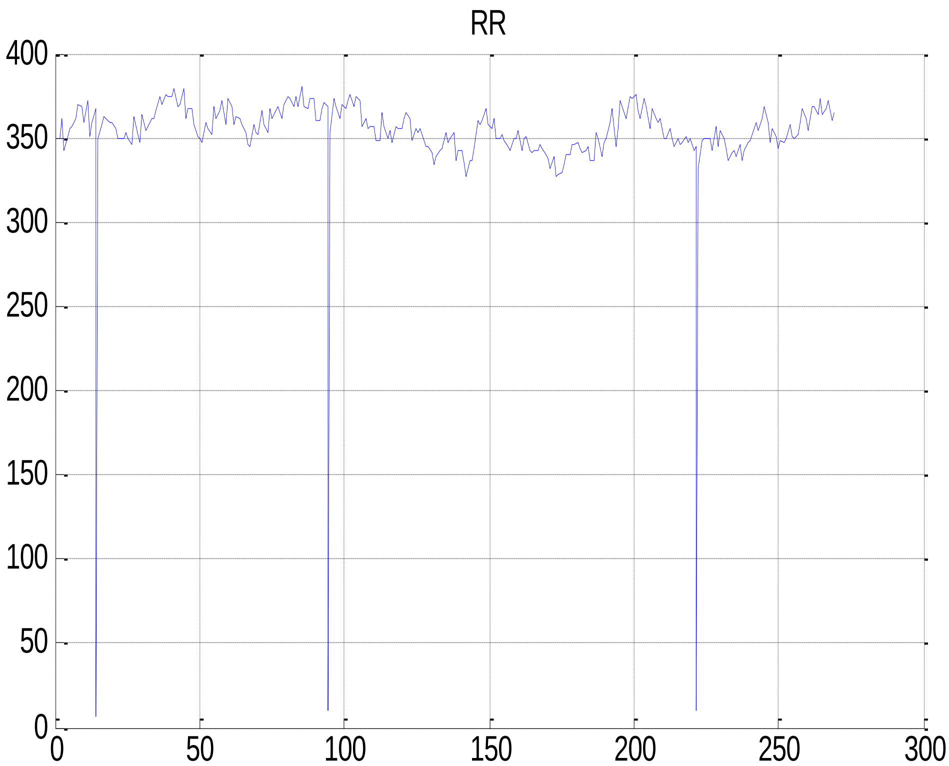
الشكل (15.2)

مراحل عمل الخوارزمية على الوصلة الثانية (lead II)



الشكل (16.2)

رسم مجموعات النقاط $\{Q_i\}$، $\{S_i\}$، $\{R_i\}$ ضمن العينة المأخوذة بالنسبة لمحور الزمن



الشكل (17.2) :

رسم بياني للمجالات RR ضمن العينة

4.2 علاقة المتغيرات مع اضطراب النظم الخطي

يؤثر اضطراب النظم القلبي Arrhythmia على القيم المقاسة (المتغيرات).

سنعتمد قاعدة بيانات اضطراب النظم القلبي MIT-BIH بالحالات السليمة والمرضية .

5.2 مبرهنة النهاية المركزية

(د.محمد صبح، د.عدنان عمورة، د.عزات قاسم (2001). نظرية الاحتمالات)

أهمية مبرهنة النهاية المركزية :

تعرض مبرهنة النهاية المركزية وتحت شروط عامة جداً أن كلاً من مجموع ومتوسط عينة عشوائية، مسحوبة من مجتمع ما، عند تكرار هذه العينات عدداً كبيراً من المرات توزيعا له على وجه التقريب شكل الجرس. فالفكرة الأساسية لمبرهنة النهاية المركزية تكمن في أنه إذا سحبنا عينات عشوائية حجم كل منها n، من مجتمع توقعه μ وانحرافه المعياريσ محدودان، فإن توزيع متوسط العينة $\bar{X}$ يتطابق تقريباً مع التوزيع الطبيعي بمتوسط يساوي μ وانحراف معياري $\frac{\sigma}{\sqrt{n}}$ وتزداد دقة التقريب كلما ازداد n . وبشكل مكافئ، يمكن أن نقول أن $\sum_{i=1}^{n} X_i$ يتوزع طبيعياً على وجه التقريب بمتوسط $n\mu$ وانحراف معياري $\sigma\sqrt{n}$ وذلك عندما يصبح n كبيراً جداً. وتبدو أهمية مبرهنة النهاية المركزية من جهتين .

فمن جهة أولى نلاحظ نزوع العديد من المتغيرات العشوائية إلى أن يكون توزيعها بصورة تقريبية، هو التوزيع الطبيعي، إذ يمكن مثلاً أن نتصور طول الإنسان حصيلة عدد كبير من المتغيرات العشوائية مثل طول الأب وطول الأم والمورثات (وعددها كبير) ،ونشاط الغدة أو الغدد ذات العلاقة بالطول والبيئة أو المحيط بأنواعه والتغذية والنشاط الرياضي. وإذا كانت هذه الآثار لهذه العوامل، تضاف بعضها إلى بعض، لتنتج واقعاً معيناً بالنسبة إلى طول الإنسان فعندئذ يمكن عد الطول حصيلة لعدد كبير من المتغيرات العشوائية، وهكذا تُطبّق مبرهنة النهاية المركزية ويكون توزيع متغير الطول على وجه التقريب هو التوزيع الطبيعي وذلك بصرف النظر عن توزيع أي من المتغيرات العشوائية التي تؤثر في تحديد الطول وهذه بالطبع محاولة للتعليل ليس أكثر إذ إن ما يجري في الواقع غير معروف لنا بصورة دقيقة، ولكن ما يمكن قوله على كل حال هو أن مبرهنة النهاية المركزية توضح سبب وجود العديد من المتغيرات العشوائية التي نصادفها في حياتنا العامة والتي يُعتبر توزعها الاحتمالي هو التوزيع الطبيعي. ومن جهة ثانية، إن لمبرهنة النهاية المركزية أهمية كبيرة

في مسألة الاستقراء الرياضي فالعديد من الإحصاءات التي تُستخدم للقيام باستقراءات حول وسطاء توزيع (والتي تحتل بدورها خصائص مهمة لمجتمع القياسات مثل P(احتمال النجاح في التوزيع الحداني أو μ متوسط التوزيع الطبيعي) هذه الإحصاءات تأخذ شكل مجموع لقياسات العينة أو متوسط هذه القياسات، وإذا كان الحال كذلك، وكانت n كبيرة كفاية فيمكننا عد التوزيع الطبيعي تقريباً جيد للتوزيع الاحتمالي لذلك الإحصاء. ولكن السؤال المهم هنا كم يجب أن يبلغ حجم العينة n حتى يصبح التقريب الناشئ عن تطبيق مبرهنة النهاية المركزية تقريباً جيد، ولسوء الحظ لا يوجد جواب عام ومحدد تماماً لهذا السؤال لأن الأمر يتعلق بالتوزيع الاحتمالي الموافق للمجتمع الذي جاءت منه العينة ، ويتعلق بالغاية من استخدام التقريب، ولكن يمكن القول إنه كلما كانت درجة التناظر في التوزيع الذي نعاينه عالية، كان التقريب جيد حتى في عينات صغيرة الحجم لذلك أحياناً نجد في بعض المراجع نستخدم $n \geq 12$ أو $n \geq 30$ ولكن بشكل عام تعد عتبة الـ 30 لحجم عينة هي بداية تطبيق مبرهنة النهاية المركزية.

نص المبرهنة :

لتكن $X_1, X_2, \dots \dots X_n$ عينة عشوائية لمتغير عشوائي Xتوقعه μ وتباينه σ^2 ولنأخذ $S_n = \sum_{i=1}^{n} X_i$

ثم نأخذ $S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ ، عندئذ من أجل n كبيرة كفاية فإن : $S_n^* \approx N(0,1)$

وكذلك فإن : $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0,1)$

القاعدة التجريبية

إن قياسات عينة طبيعية مسحوبة من مجتمع طبيعي توقعه μ وتباينه σ^2 تكون موزعة بالشكل التالي :

%68.26 من القياسات تقع ضمن المجال $\mu \pm \sigma$

%95.44 من القياسات تقع ضمن المجال $\mu \pm 2\sigma$

%99.74 من القياسات تقع ضمن المجال $\mu \pm 3\sigma$

ولإثبات ذلك نأخذ أولاً :

$$P\left[\mu-\sigma \le X \le \mu+\sigma\right] = P\left[\frac{\mu-\sigma-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{\mu+\sigma-\mu}{\sigma}\right]$$

$$P[-1 \le Z \le 1] = P(Z \le 1) - P(Z \le -1) =$$

$$0.8413 - 0.1587 = 0.6826$$

بالأسلوب نفسه نثبت أن:

$$P\left[\mu-2\sigma \le X \le \mu+2\sigma\right] = 0.9544$$

$$P\left[\mu-3\sigma \le X \le \mu+3\sigma\right] = 0.9974$$

ملاحظة :

من S_n و ${S_n}^*$ نجد أن :

$$E(S_n) = n\mu$$

$$V(S_n) = n\sigma^2$$

وبالتالي بملاحظة أن : $\bar{X} = \frac{S_n}{n}$ فإن : $V(\bar{X}) = \frac{\sigma^2}{n}$

$$E(\bar{X}) = \mu$$

وهكذا نكون قد بيّنّا أن :

$${S_n}^* \approx N(0,1) \Leftrightarrow \frac{S_n - n\mu}{\sigma\sqrt{n}} \approx N(0,1) \Leftrightarrow S_n \approx N(n\mu, n\sigma^2)$$

$$\Leftrightarrow \frac{S_n}{n} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0,1)$$

الفصل الثالث :

استراتيجيات البحث عن النموذج الملائم (المُصنِّف)

سنتحدث في هذا الفصل عن طريقتين من الطرق المستخدمة في الحيِّز الزمني.

1.3 الشبكات العصبونية الصُّنعية

(G.Karraz,Doctoral Thesis (2007))

مقدمة

نقدم في هذا الفصل نظرة عامة عن الشبكات العصبونية الصنعية ANN كنموذج رئيسي لأنظمة التشخيص الذكية، لكي نتمكن من تفسير إشارة ECG بشكل أتوماتيكي.

لقد مرَّت الشبكات العصبونية عبر فترتي تطور أساسيتين، وهما بداية الستينات ومنتصف الثمانينيات. شكّلت هاتان الفترتان تطوراً جوهرياً في مجال قدرة الآلة على التعلم.

تمّ استيحاء ANNعن طريق اكتشافات مرتبطة بسلوك الدماغ ،بحيث تُدعى شبكة من الوحدات (عصبون).

يملك الدماغ البشري تقديرياً حوالي 10 بليون عصبون وكل منها متصل بمعدل وسطي

بـ 10 آلاف عصبون آخر، يتلقى كل عصبون إشارات عبر مشابك والتي تتحكم بآثار الإشارات على العصبون.

لقد ساد اعتقاد أن هذه الاتصالات المشبكية تلعب دوراً أساسياً في السلوك الدماغي،

إن كتلة البناء الأساسية في شبكة عصبونية صنعية هي النموذج الرياضي لعصبون،كما هو موضح في الشكل(1.3)

إن العناصر الثلاثة الأساسية للعصبونات الشبكية هي :

1) المشابك أو حلقات الربط التي تؤمن الأوزان ω_{j} - للقيم المدخلة-

x_j من أجل j=1……m

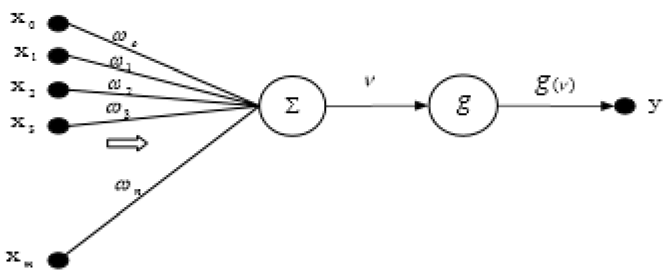
2) المضيف : والذي يجمع قيم الدخل الموزونة لحساب الدخل لدالة التفعيل

$$v = \omega_0 + \sum_{j=1}^{m} \omega_j \, x_j$$

حيث ω_0 يسمى الانحياز وهو قيمة عددية مرتبطة بالعصبون. من المناسب اعتبار الانحياز وزن للدخل x_0 والذي قيمته دائماً مساوية للواحد بحيث :

$$v = \sum_{j=1}^{m} \omega_j \, x_j$$

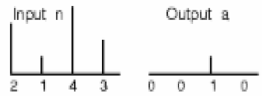
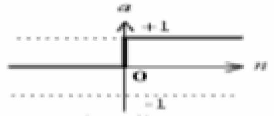
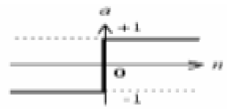
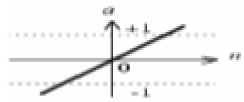
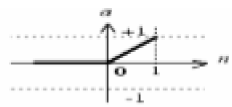
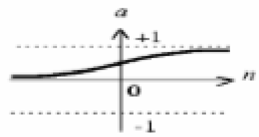
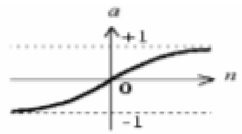
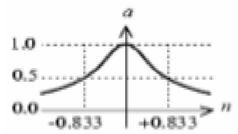
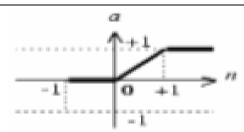
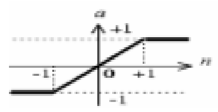
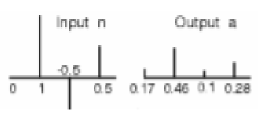
3) دالة تفعيل (g وتسمى أيضاً إزالة الأوزان وتعديل أوزان جديدة أو دالة انتقال)التي تصل بيانياً لقيمة الخرج للعصبون. إن هذه الدالة هي دالة مطّردة.
الجدول (1.3) يبين قائمة كاملة بدوال انتقال.



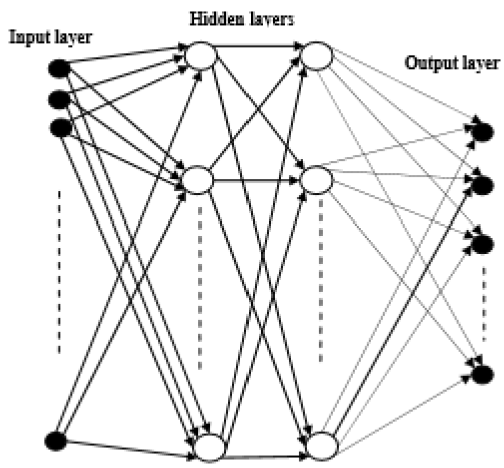
الشكل(1.3): الشكل الرياضي لعصبون

بينما تم دراسة عدة بنى مختلفة للشبكات العصبونية الصنعية بواسطة الباحثين، كان التطبيق الأنجح بتعريف بيانات الشبكات العصبونية هو شبكات الارتجاع المسبق متعددة الطبقات .
يوجد شبكات والتي ضمنها يوجد طبقة دخل مكونة من عقد ببساطة تقبل القيم المدخلة، وطبقات متعاقبة من العقد والتي هي عصبونات كما هو ممثل بالجدول (1.3)

الجدول (1.3)

التمثيل	الوصف	الدالة
	تُخرِج المتجهات بعناصر بين 0 و 1 ولكن بعلاقات سليمة بالنسبة لحجمها، لا تملك softmax دالة اشتقاق .	Compet
	تحدد العصبون المخرج إما ب 0 عندما يكون معامل دخل الشبكة n أقل من 0 أو 1عندما يكون n أكبر أو تساوي 0 ،تُستخدم هذه الدالة لتوليد عصبونات تصنع قرارات التصنيف	Hard-limit
	$=\begin{cases}(1)\ \dots..if \dots.v \geq 0\\(-1)\ \dots.\text{otherwise}\end{cases}$	Symmetric Hard-limit
	تطبق علاقة خطية،تُستخدم العصبونات من هذا النوع كتقريبات خطية في المرشّحات الخطية.	linear
	تطبّق علاقة خطية موجبة.	Positive linear
	تأخذ دخلاً والذي يمكن أن يكون له أي قيمة بين ∞-و∞+ وتُخفّض الخرج إلى قيمة ضمن المجال [0,1]	Log-sigmoid
	تأخذ دخلاً والذي يمكن أن يكون له أي قيمة بين ∞-و∞+ وتُخفّض الخرج ضمن المجال [-1,1].	Tangent Sigmoid
	تملك هذه الدالة قيمة عظمى 1 عندما يكون دخل، يتصرف عصبون Radial basis كأداة كشف، يُخرج 1 كلما كان الدخل p مطابق لمتجه وزنه p.	Radial Basis
	تأخذ دخل Nوتُخرِج قيم لـ N مشذبة ضمن المجال [0,1]	Satlin
	تأخذ دخل N وتُخرِج قيم N مشذبة ضمن المجال [-1,1]	Symmetrical Satlin
	تأخذ معامل دخل واحد وتكون المتجهات المخرجة بعناصر بين 0 و 1 ولكن مع الحفاظ على العلاقات المتعلقة بحجمها سليمة	Softmax
	تأخذ دخل واحد N وتُخرِج كل عنصر من N ماراً بدالة متجهية أساسية	Triangular Basis

إن خرج العصبونات في طبقة هي دخل عصبونات في الطبقة التالية. تسمى الطبقة الأخيرة طبقة الخرج. تُعرَف الطبقات بين طبقات الدخل والخرج بالطبقات المخفية، الشكل(2.3) هو مخطط لهذه البنية. ضمن عملية المراقبة للشبكة، حيث تُستخدم الشبكة العصبونية للتنبؤ بكمية عددية، يوجد عصبون واحد في طبقة الخرج وخرجه هو التنبؤ، عندما يتم استخدام الشبكة في التصنيف يكون لدى طبقة الخرج بشكل نموذجي عدد عقد كعدد الصفوف، وتعطي عقدة طبقة الخرج التي لها أكبر قيمة خرج تقدير الشبكة للصف من أجل دخل معطى. في الحالة الخاصة لصفين، من الشائع أن يكون لدينا عقدة واحدة في طبقة الخرج. لقد تم إيجاد التصنيف بين الصفين بتطبيق عزل لقيمة الخرج عند العقدة.



الشكل(2.3) :بنية الشبكات العصبونية متعددة الطبقات

1.1.3 الشبكة ذات الطبقة المفردة :

لنبدأ باختبار الشبكات العصبونية بطبقة عصبونية واحدة فقط (طبقة خرج فقط لا يوجد طبقة مخفية). تتألف أبسط شبكة من عصبون واحد بالدالة g التي تم اختيارها لتكون دالة التعريف $g(v) = v$ من أجل جميع قيم v. لاحظ في هذه الحالة أن خرج الشبكة يكون $\sum_{j=0}^{m} \omega_j x_j$ دالة خطية بعناصرمتجه الدخل x.عندما نضع نموذج للمتغير التابع y باستخدام الانحدار الخطي المضاعف، نستطيع تفسير الشبكة العصبونية كبنية تتنبأ بقيمة $\hat{y}$ من أجل متجه x معطى بأوزان تمثل المعاملات. إذا اخترنا هذه الأوزان لتصغير مربع متوسط الخطأ مستخدمين المشاهدات في مجموعة بيانات تدريبية. على كل حال يوجد توجه مختلف في حالة الشبكات العصبونية : تكون الأوزان "متعلمة".إن الشبكة معروضة مع بيانات تدريبية الواحدة تلو الأخرى وتُصحَّح الأوزان بعد كل حالة في محاولة لتصغير مربع متوسط الخطأ. تعتمد عملية التعديل المتزايدة للأوزان على الخطأ المصنوع للحالات التدريبية وتعرف بـ (التدريب) للشبكة العصبونية. إن أكثر خوارزمية تعديل ديناميكية

مستخدمة عالمياً من أجل نسخة الشبكة العصبونية للانحدار الخطي تُعرَف بقاعدة Widrow-Hoff أوخوارزمية مربع المتوسطات الأصغري(LMS)، تنص ببساطة:

لتكن $x(i)$ رمز لمتجه الدخل من أجل الحالة ذات الترتيب i المستخدمة لتدريب الشبكة، وعُرِضَت الأوزان قبل هذه الحالة على الشبكة بواسطة المتجه $\omega(i)$.

تكون قاعدة التعديل:

$$w(i+1) = w(i) + \eta\left(y(i) - \hat{y}(i)\right)x(i) \qquad with\; w(0) = 0$$

من الظاهر أنه إذا تم تدريب الشبكة بهذه الطريقة بواسطة عرض مشاهدات بيانات الاختبار بشكل تكراري واحدة تلو الأخرى ثم من أجل قيم لـ η صغيرة مناسبة (η هي وسيط ضبط هام تم اختياره من خلال طريقة التجربة والخطأ)، ستتعلم الشبكة (التقارب إلى) القيم w. لاحظ أنه يجب تقديم البيانات التدريبية عدة مرات $w(i)$ لتعريف الأفضل w. إن فائدة التعديل الديناميكي هي أن الشبكة تتعقب الاتجاهات الزمنية المتوسطة في النموذج الخطي الأساسي بشكل فعال جداً. إذا استخدمنا الشبكة العصبونية ذات الطبقة الواحدة لتصنيف c صفاً، سوف نستخدم c عقدة في طبقة الخرج. إذا فكرنا بتحليل الفرق التقليدي في حدود الشبكة العصبونية، فإن المعاملات في توابع تصنيف فيشر تعطينا الأوزان للشبكة أفضل، إذا أتت متجهات الدخل من توزيعات طبيعية متعددة المتغيرات (MND) بمصفوفة تغاير مشتركة. للتصنيف ضمن صفين، فإن طريقة البحث الخطية عن الحل الأمثل التي اختبرناها في صف، يمكن أن تظهر كاختيار أوزان مثلى في شبكة عصبونية بطبقة واحدة باستخدام دالة الهدف الملائمة. ويمكن اعتبار معاملات الانحدار اللوجستي التي نحصل عليها بطريقة الاحتمالية العظمى أيضاً كأوزان في شبكة عصبونية لتصغير دالة الرواسب المسماة بـ (الانحراف)، في هذه الحالة الدالة اللوجستية $g(v) = \frac{e^v}{1+e^v}$ هي دالة التفعيل للعقدة الخرج.

2.1.3 الشبكة العصبونية متعددة الطبقات

إن الشبكات العصبونية متعددة الطبقات هي بلا شك أكثر الشبكات شيوعاً في استخدامها في التطبيقات، بينما من الممكن الأخذ بعين الاعتبار عدة دوال تفعيل.

وجد بالتجريب أن الدالة اللوجستية (والمسمية أيضاً Sigmoid)

$g(v) = \frac{e^v}{1+e^v}$ كدالة تفعيل (أو البدائل الثانوية كدالة tanh) تعمل كأفضل دالة.

في الحقيقة لقد أعادت النجاحات في تدريب الشبكات العصبونية باستخدام هذه الدالة إحياء الاهتمام بالشبكات العصبونية بدلاً من دالة الخطوة(perceptron) المستوحى بيولوجياً عبر التاريخ،لاحظ أن استخدام دالة خطية لا ينجز أي شيء في الشبكات متعددة الطبقات، وهذا مختلف عما يمكن فعله بالشبكات ذات الطبقة الواحدة بدالة تفعيل خطية.

إن ما يرفع قيمة الدالة اللوجستية التطبيقية هو حقيقةً أنها خطية تقريباً في المجال

حيث تكون g بين 0.1 و 0.9 ولكن له أثر محو على قيم v الصغيرة جداً أو الكبيرة جداً.

نظرياً، من الكاف أن نعتبر الشبكات بطبقتين، وهذه بالتأكيد الحال بالنسبة لكثير من التطبيقات.

على كل حال هناك عدد من الحالات حيث تكون ثلاث وأحياناً أربع وخمس طبقات ذات فعالية أكبر.

للتنبؤ بعقدة الخرج ،غالباً تعطى دالة تفعيل خطي لتؤمن التنبوئات والتي لا تنتهي إلى الصفر بمجال واحد، وكبديل ندرج الخرج في القسم الخطي (0.1-0.9) للدالة اللوجستية.

لسوء الحظ لا يوجد نظرية واضحة لترشدنا في اختيار عدد العقد في كل طبقة مخفية في عدد الطبقات بالحقيقة.

إن الممارسة الشائعة هي استخدام التجربة والخطأ، مع أنه يوجد مخططات لدمج طرق البحث عن الحل الأمثل، كمثال خوارزميات وراثية لتدريب الشبكة من أجل هذه الوسطاء.

بما أن التجربة والخطأ هما جزء من تطبيقات الشبكة العصبونية.من المهم فهم الطريقة المعيارية المستخدمة لتدريب الشبكة متعددة الطبقات :الانتشار العكسي.

ليس من المبالغة القول أن سرعة خوارزمية الانتشار العكسي المصنوعة من الشبكات العصبونية هي أداة عملية، بالطريقة التي فيها النظام الأبسط الخطي للبحث عن الحل الأمثل هو أداة عملية،كانت فعالية خوارزمية الانتشار العكسي سبباً كبيراً في إعادة الاهتمام القوي بالشبكات العصبونية في أواسط الثمانينات.

3.1.3 خوارزمية الانتشار العكسي:

سنناقش خوارزمية الانتشار العكسي لتصنيف المشكلات. هناك تعديل بسيط للتنبؤ بالمشكلات، عندما نحاول التنبؤ بقيمة عددية مستمرة. في هذه الحالة نُغيّر دالة التفعيل لعصبونات طبقة الخرج إلى دالة التعريف، والتي تملك قيمة خرج تساوي قيمة دخل. (يمكن كبديل أن نعيد قياس و تمركز الدالة اللوجستية للسماح للخرج أن يكون تقريباً خطي في

مجال قيم المتغيرات التابعة). تُنجَز خوارزمية الانتشار العكسي في عبورين ملحوظين، عبور أمامي، يتبعه عبور عكسي عبر طبقات الشبكة. تتبدّل الخوارزمية بين هذين العبورين مرات عديدة، بينما تمسح(تفحص) البيانات التدريبية.

بشكل نموذجي، يجب أن تُفحَص البيانات التدريبية عدة مرات قبل أن "تتعلم" الشبكة عمل تصنيف جيد.

1- العبور الأمامي : حساب خرج كل العصبونات في الشبكة :

تبدأ الخوارزمية بأول طبقة مخفية مستخدمة كقيم إدخال المتغيرات المستقلة لحالة (عادة يسمى "نموذج" في مجتمع تعليم الآلة) من مجموعة البيانات التدريبية.

يتم حساب خرج العصبونات من أجل جميع العصبونات في الطبقة الأولى المخفية بإجراء تقدير لدالة التفعيل والجمع المرتبط. إن هذه المخرجات هي دخل للعصبونات في الطبقة المخفية الثانية. وفضلاً عن ذلك، يتم إجراء حسابات دالة التفعيل والجمع المرتبط لحساب خرج عصبونات الطبقة الثانية .ويستمر هذا طبقة بعد طبقة حتى نصل لطبقة الخرج، ونحسب الخرج لهذه الطبقة.تشكل قيم الخرج هذه تقديراً للشبكة العصبونية عند قيمة المتغير التابع.عند استخدامنا الشبكة العصبونية للتصنيف ،(إذا كان C=2، نستخدم عقدة خرج واحدة فقط مع قيمة فصل لنرسم قيمة خرج عددية لواحد من هذين الصفين). لنرمز بـ ω_{ij} لوصلة من العقدة i إلى العقدة j . تبدأ قيم ω_{ij} بأعداد صغيرة (عشوائية بشكل عام) في المجال 0.00 ± 0.05

يتم تعديل هذه الأوزان إلى قيم جديدة في العبور العكسي كما هو موصوف أدناه.

2-العبور العكسي :انتشار الخطأ وتعديل الأوزان :

يبدأ هذا الطور بحساب خطأ كل عصبون في طبقة الخرج.هناك دالة خطأ شائعة وهي الفرق التربيعي بين O_k خرج العقدة k و Y_k هدف هذه العقدة تكون قيمة الهدف فقط 1 لعقدة الخرج بالتوافق مع صف النموذج، وصفر لعقد الخرج الأخرى .

(وُجِد خلال التطبيق أنه من الأفضل استخدام قيم 0.9 و 0.1على الترتيب).تُحسب حدود الخطأ لكل عقدة طبقة خرج كالتالي :

$$\delta_k = O_K(1\text{-}O_K)(Y_k - O_K) \qquad (1.3)$$

تُستخدم هذه الأخطاء لتعديل أوزان الوصلات بين الطبقة الأخيرة -ولكن- الوحيدة للشبكة والطبقة الخرج.تُعطى القيمة الجديدة لوزن الوصلة من العقدة j إلى k العقدة بـ:

$$\omega_{ij}^{new} = \omega_{ij}^{old} + \eta O_j \delta_k \qquad (2.3)$$

هنا η هي وسيط ضبط هام تم اختياره من التجربة والخطأ، بواسطة التشغيلات المتكررة على البيانات التدريبية. تكون القيم النموذجية لـ η في المجال 0.1 إلى 0.9.
تؤدي القيم الصغيرة لتعلّم بطيء ولكن موثوق، تؤدي القيم الكبيرة لتعلّم شاذ(مذبذب)، ويمكن أن تقود لشبكة غير مستقرة. تعاد العملية للوصلات بين العقد في الطبقة المخفية الأخيرة والطبقة المخفية الأخيرة - ولكن - الوحيدة. إن وزن الوصلة بين العقد i و j وفق(2.3):

$$\omega_{ij}^{new} = \omega_{ij}^{old} + \eta O_j \delta_k$$

حيث $$\delta_j = O_j(1\text{-}O_j) \sum_k \omega_{jk} \delta_k$$

من أجل كل عقدة j في الطبقة المخفية الأخيرة.
يستمر الانتشار العكسي لتعديل الأوزان عبر هذه الخطوط حتى الوصول إلى طبقة الدخل.
في هذا الوقت ، يكون أصبح لدينا مجموعة جديدة من الأوزان والتي يمكن أن نعمل عليها عبور جديد عندما تُعرض مع مشاهدات بيانات تدريبية.

4.1.3 القيم المحلية الأفضل، الأدوار :

إن خوارزمية الانتشار العكسي هي نسخة عن طريقة الهبوط الحاد للبحث عن الحل الأمثل، والمطبّقة على مشكلة إيجاد الأوزان التي تُصغّر دالة الخطأ لخرج الشبكة. وبسبب تعقيد الدالة والأعداد الكبيرة للأوزان والتي تم "تدريبها" خلال "تعلّم" الشبكة ، لا يوجد هناك تأكيد أن خوارزمية الانتشار العكسي (وحقيقةً أي خوارزمية تطبيقية) ستجد الأوزان الأمثل التي تُخفّض الخطأ. من الممكن أن تتوقف هذه الإجرائية عن العمل مؤقتاً عند قيمة صغرى محلية. وٌجد أنه من المفيد أن نختار عشوائياً ترتيب تقديم الحالات في مجموعة تدريبية بين تدقيقات(فحوصات) مختلفة. يمكن تسريع الخوارزمية عبر حشد البيانات، والذي معناه تحديث الأوزان لنماذج عدة في عبور. وعلى كل حال ،فقد وجد أن الحالة القصوى على الأقل من استخدام مجموعة البيانات التدريبية كاملة في كل تحديث يتوقف كثيراً عند قيمة صغرى محلية قليلة. يدعى الفحص الواحد لكل الحالات في البيانات التدريبية (دور). تتطلب العديد من شبكات الارتجاع المسبق والانتشار العكسي أدواراً عديدة، قبل أن تصبح الأخطاء صغيرة نسبياً. تم اقتراح عدد من التعديلات لتقليل الأدوار التي نحتاجها لتدريب شبكة عصبونية. هناك فكرة مستخدمة عموماً وهي إشراك حد العزم والذي يُدخِل عطالة ما في تعديل الوزن بالعبور العكسي.

يتم ذلك عن طريق إضافة حد لعبارة تعديل الوزن لتلك الوصلة. يسمى هذا الكسر وسيط التحكم بالعزم. ستُجبِر القيم الكبيرة لوسيط العزم تعديلات الوزن المتعاقبة أن تكون بنفس الاتجاه. وهناك فكرة أخرى وهي تغيير وسيط التعديل δ بحيث ينخفض كلما زاد عدد الأدوار .

بديهياً، إن هذا مفيد لأنه يُجنّبنا فرط الملاءمة والذي يحدث على الأرجح في أدوار أخيرة أكثر من أدوار مبكرة.

5.1.3 فرط الملاءمة وخيار الأدوار التدريبية:

إن ضعف الشبكات العصبونية هو أنها يمكن بسهولة أن تتلاءم بشكل زائد، مسببةً نسبة خطأ في بيانات فاعلة أعلى من نسبة الخطأ في بيانات تدريبية. ولذلك من المهم ألّا نفرط في تدريب البيانات. هناك طريقة جيدة لاختيار عدد الأدوار التدريبية باستخدام مجموعة البيانات الفاعلة بشكل دوري لحساب معدل الخطأ لها خلال تدريب الشبكة. ينقص خطأ الفاعلية في الأدوار المبكرة للانتشار العكسي، ولكنه بعد فترة يبدأ بالزيادة. إن نقطة خطأ الفاعلية الأصغري هي مؤشر جيد على العدد الأفضل للأدوار من أجل التدريب وعلى الأرجح فإن الأوزان في تلك المرحلة تؤمن نسبة الخطأ الأفضل في بيانات جديدة.

6.1.3 خوارزميات Quasi-Newton

إن طريقة نيوتن هي بديل لطرائق الميل المتحدة للوصول السريع للحل الأمثل. إن الخطوة الأساسية لطريقة نيوتن هي :

$$\omega_{ij}^{new} = \omega_{ij}^{old} - A_{ij}^{-1} g_{ij} \qquad (3.3)$$

حيث A_{ij} هي مصفوفة Hessian (مشتقات درجة ثانية) لمؤشر الأداء عند القيم الحالية للأوزان و الانحيازات .

تتقارب طريقة نيوتن عادة بشكل أسرع من طرائق الميل المتحدة. لسوء الحظ أنه من المكلف والمعقّد حساب مصفوفة لشبكات ارتجاع مسبق عصبونية.

هناك صف من الخوارزميات معتمدة على طريقة نيوتن، ولكنها لا تتطلب حساب المشتقات من الدرجة الثانية، وتدعى هذه بـ طرائق Quasi-Newton أو (القاطع).

تُحدّث هذه الطرائق مصفوفة Hessian تقريبية عند كل تكرار للخوارزمية. يُحسب التحديث كأنه دالة للميل. إن طريقة Quasi-Newton الناجحة جداً في الدراسات المنشورة هي Broyden,Fletcher,Goldfarb و تعديل Shanno(BFGS)

7.1.3 خوارزمية قاطع ذات خطوة واحدة :

بما أن خوارزمية BFGS تتطلب تخزين و حساب أكثر في كل تكرار، فإنها تحتاج إلى تقريب قاطع بمتطلبات حساب وتخزين أصغر. لايُخزّن القاطع بخطوة واحدة (OSS)مصفوفة Hessian الكاملة، يَفترض عند كل تكرار أن مصفوفة Hessian السابقة كانت مصفوفة التعريف. ولهذا ميزة إضافية ، وهي أن جهة البحث الجديدة يمكن إيجادها بدون حساب معكوس مصفوفة.

8.1.3 خوارزمية Levenberg-Marquardt

تم تصميم خوارزمية Levenberg-Marquardt من أجل الوصول لسرعة تدريب من الدرجة الثانية بدون الحاجة لحساب مصفوفة Hessian، وعندما يكون لدالة الأداء شكل مجموع مربعات (كما هو الحال بشكل نموذجي في تدريب شبكات الارتجاع العكسي)، فإنه يمكن تقريب مصفوفة Hessian كالتالي :

$$H = J^T J \qquad (4.3)$$

ويمكن حساب الميل كالتالي :

$$g = J^T e \qquad (5.3)$$

حيث أن J هي مصفوفة يعقوبيان، والتي تحوي المشتقات الأولى لأخطاء الشبكة بالنسبة للأوزان والانحيازات .

e هو متجه لأخطاء شبكة. يمكن حساب مصفوفة يعقوبيان عن طريق تقنية انتشار عكسي معيارية، والتي تكون أقل تعقيداً من حساب مصفوفة Hessian.

تستخدم خوارزمية Levenberg-Marquardt هذا التقريب لمصفوفة Hessian في التعديل شبيه نيوتن التالي :

$$\omega_{ij}^{new} = \omega_{ij}^{old} - [J^T J + \mu I]^{-1} J^T e \qquad (6.3)$$

عندما تكون الكمية العددية μ غير الموجهة تساوي 0، عندها فإن هذه هي طريقة نيوتن، باستخدام مصفوفة Hessian التقريبية. عندما تكون μ كبيرة، يصبح هذا هبوط ميل بقياس خطوة صغيرة. إن طريقة نيوتن أسرع وأكثر دقة قرب قيمة صغرى للخطأ، لذا فإن هدفها الانتقال نحو طريقة نيوتن بأسرع وقت ممكن. وهكذا فإن μ ينقص بعد كل خطوة

ناجحة(نقصان دالة الأداء)، ويزداد فقط عندما تزيد خطوة تجريبية دالة الأداء. وبهذه الطريقة، ستنقص دالة الأداء دائماً عند كل تكرار للخوارزمية.

9.1.3 خوارزمية Bayesian

سوف نشرح في هذا المقطع كيفية فهم نموذج Bayesian صُنعي بشكل واضح.

لنفرض أنه لدينا عينة تدريبية D تتضمن N زوجاً دخل – خرج:

$$D = [(x^n, y^n)|n = 1,2 \dots \dots . N] \qquad (3.7)$$

حيث أن x هو متجه دخل يتألف من I عنصر، و y هو علامة الصف الموافق المؤلف من K صف، الهدف هو استخدام ANN لنمذجة علاقة الدخل- الخرج $(y = k|x)$ ،

ليكن لدينا 6 صفوف $\text{k} \in \{k_1, k_2, k_3, k_4, k_5, k_6\}$

لفهم نموذج الانحدار اللوجستي معتمد على طريقة Bayesian، حيث قدّرنا احتمالية الصف للدخل المعطى بواسطة : $P(Y = k|X = x)$

ولنفرض أن خرج عملية الجمع ودالة التفعيل Sigmoid - في العصبونات المخفية والعصبونات الخرج- يرمز لها بـ S_j, S_k علىالترتيب يمكن كتابتها كالتالي :

$$S_j = \tanh\left(\sum_i \omega_{ij} X_i + \omega_{j0}\right) \qquad (8.3)$$

$$S_k = \left(\sum_k \omega_{kj} S_j + \omega_{k0}\right) \qquad (9.3)$$

حيث أن tanh هي دالة المماس للقطع الزائد.

ترمز ω_{ij} لمصفوفة الوزن في طبقة الدخل و ω_{kj} لمصفوفة الوزن في طبقة الخرج.

لنضمن إمكانية تفسير الخرج كاحتمالات. يُستخدم الانحدار اللوجستي لنمذجة خطورة (أو احتمالية) حدوث اضطراب النظم القلبي .

لتكن$P(Y = k|X)$ احتمال الحدث $Y = 1$ بفرض أن الدخل X معطى .

تم نمذجة هذا بواسطة دالة لخرج الشبكة y كالتالي :

$$P(Y=k|X)=\frac{1}{1+\exp(-S_k)} \qquad (10.3)$$

إن نموذج الانحدار اللوجستي هو ببساطة تحويل غير خطي للانحدار الخطي .

التوزيع اللوجستي هو دالة توزيع له شكلS المشابه للتوزيع الطبيعي المعياري (والذي يُنتِج نموذج انحدار وحدة احتمالية) ولكنه أسهل للعمل ضمن معظم التطبيقات. تجبر الدالة اللوجستية الاحتمالات المقدرة أن تتوضع بين0 و1 .

إن آخر جزء في النظام هو الشبكة العصبونية متعددة الطبقات (Perceptron)، مدربة باستخدام خوارزمية BNN ، تُحسّن الشبكة باستخدام دالة الكلفة الاحتمالية اللوغاريتمية،المعطاة بـ :

$$C(\omega)=-\frac{1}{k}\sum_{k}\sum_{i}Y_i(k)\ln P[(Y=k|X)] \qquad (11.3)$$

حيث$\omega=(\omega_{ij},\omega_{jk})$ هو متجه أوزان الشبكة .

لتصغير دالة التكلفة بين الخرج الحقيقي والمرغوب للشبكة، تمرِّر خوارزمية BNN المعلومات من العصبون الخرج عكسياً إلى جميع الوحدات المخفية لتشكيل حدود الخطأ والتي تستخدم لتحديث أوزان الشبكة متعددة الطبقات.

ملاحظة: (المتغير العشوائيY المستخدم في الفصل الأخير سيكون مختلف عن أي متغير بالرمز Y تم ذكره قبل الفصل الأخير)

الفصل الرابع :

الانحدار الخطي

1.4 الانحدار الخطي البسيط

جون نتر ،وليام وازرمان،ميخائيل كتنر(1995)

Kutner ,M. H.,Nachtsheim,C.J.,Neter , J., Li ,W.(2005)

1.1.4 الانحدار الخطي بمتغير مستقل واحد

إن تحليل الانحدار هو أداة إحصائية تستفيد من العلاقة بين متغيرين كميين أو أكثر للتنبؤ بأحد المتغيرات استناداً إلى قيم المتغير أو المتغيرات الأخرى .

2.1.4 العلاقات بين المتغيرات

سوف نميز بين علاقة دالية وعلاقة إحصائية .

العلاقات الدالية بين متغيرين : يعبر عن العلاقة الدالية بين متغيرين بصيغة رياضية ،فإذا كان X المتغير المستقل و Y المتغير التابع فإن العلاقة الدالية تكون من الشكل:

$$Y = f(X)$$

وإذا أُعطيت قيمة معينة لـ X فإن الدالة f تشير إلى قيمة Y المقابلة .

- **العلاقة الإحصائية بين متغيرين**

على عكس العلاقة الدالية فالعلاقة الإحصائية ليست تماماً دالة . وبوجه عام ، لاتقع مشاهدات الدالة الإحصائية تماماً على منحني العلاقة بينهما.

3.1.4 نماذج الانحدار واستخداماتها

مفاهيم أساسية

إن نموذج الانحدار ماهو إلا وسيلة رسمية للتعبير عن عنصريين أساسيين من عناصر العلاقة الإحصائية :

1- نزوع المتغير التابع Y للتغير مع االمتغير المستقل X بصورة نمطية.

2- تبعثر النقاط حول منحني العلاقة الإحصائية .

وقد تجسدت هاتان الخاصتان في نموذج الانحدار من خلال الافتراضين التاليين :

1) يوجد توزيع احتمالي للمتغير Y عند كل مستو من مستويات X .

2) تتغير متوسطات هذه التوزيعات الاحتمالية بصورة نمطية مع تغير X .

ملاحظة

التعبير عن X على أنه" متغير مستقل " أو "متغير تنبؤ"، والتعبير عن Y على أنه "متغير تابع " أو "متغير استجابة "هي ألقاب اصطلاحية. ولاتتضمن في حالات معينة، أن Y تعتمد اعتمادا ً سببياً على X . وبصرف النظر عن قوة العلاقة الإحصائية فإن النموذج الإحصائي لاينطوي بالضرورة على ثنائية السبب والنتيجة. وفي بعض التطبيقات يعتمد المتغير المستقل اعتماداً سببياً على متغير الاستجابة، كما هو الحال عندما نُقدِّر درجة الحرارة (الاستجابة) من ارتفاع الزئبق (المتغير المستقل) في ميزان الحرارة .

- **نماذج انحدار بأكثر من متغير مستقل واحد :**

قد تحتوي نماذج الانحدار على أكثر من متغير مستقل واحد.

بناء نماذج الانحدار :

- **اختيار المتغيرات المستقلة:** بما أنه ينبغي لنا، عند بنائنا النماذج اختزال الواقع الفعلي إلى جزء طيّع منه، يمكن التعامل معه، فينبغي أن يقتصر نموذج الانحدار لأي مسالة ندرسها على عدد محدود من المتغيرات المستقلة. ولذلك فإن المشكلة الأساسية تكمن في اختيار مجموعة من المتغيرات المستقلة لنموذج الانحدار يمكن أن توصف بمعنى ما، أنَّها ولأغراض التحليل متغيرات جيدة. والعامل الرئيس في اختيار متغير مستقل هومدى مساهمته في تخفيض ما تبقى من التغير في Y. بعد أن تكون مساهمات متغيرات اخرى، تم مبدئياً اختيارها إلى النموذج، قد أُخذت في الاعتبار. ومن العوامل الأخرى تأتي أهمية المتغير كعامل سببي في العملية موضع التحليل. ودرجة الدقة، وسرعة الحصول على مشاهدات المتغير، وتكلفتها، مقارنة بمتغيرات أخرى منافسة. وكذلك إمكانية وضع المتغير تحت إدارة المجرب .

• الشكل الدالي لعلاقة انحدار :

يرتبط اختبار الشكل الدالي لمعادلة الانحدار باختيار المتغيرات المستقلة. ففي بعض الأحيان يمكن أن تشير المعرفة النظرية المتيسرة إلى الشكل الدالي. فعلى سبيل المثال، قد تشير معرفتنا وتجربتنا العملية إلى أن دالة الانحدار التي تربط تكلفة وحدة إنتاج بعدد مرات إنتاجها سابقاً، ينبغي لها أن تتخذ شكلاً محدداً بخواص مقاربة معينة .

وفي الغالب-على كل حال- لايكون الشكل الدّالي لعلاقة الانحدار معروفاً سلفاً، ولابد من اتخاذ قرار بشأنها حالما يتم جمع البيانات وتحليلها. وهكذا أُستخدمت، في الغالب، علاقات انحدار خطية وتربيعية كتقريبات أوليَّة مرضية لدوال انحدار من طبيعة غير معروفة. وفي الحقيقة يمكن استخدام هذه الأنواع البسيطة من دوال الانحدار عندما يكون الشكل معقداً جداً، ولكن يمكن تقريبه بشكل معقول بدالة انحدار خطية أو تربيعية .

• مجال النموذج

عند صياغة نموذج انحدار، نحتاج عادة إلى تقييد تغطية النموذج بحيث تقتصر على فترة أو منطقة من القيم للمتغير أو المتغيرات المستقلة. ويتحدّد المجال من خلال تصميم الدراسة أو من خلال مدى البيانات المتوفرة.

• استخدامات تحليل الانحدار

يخدم تحليل الانحدار ثلاثة أغراض رئيسة:

1) الوصف : أي يصف سلوك المتغير التابع (متغير الاستجابة) والتغير في متوسطه.

2) السيطرة: أي عندما يكون للمتغيرات المستقلة تأثير متفاوت على المتغير التابع.

3) التنبؤ : يقوم بالتنبؤ بقيم متغير الاستجابة اعتماداً على القيم المقدرة لمعاملات الانحدار.

وتتداخل الأغراض المتعدِّدة لتحليل الانحدار في الواقع العملي.

4.1.4 نموذج انحدار بتوزيع غير معروف لحد الخطأ

• شكل النموذج

وهو نموذج انحدار أساسي بمتغير مستقل واحد ودالة انحدار خطية ويمكن عرضه كما يلي :

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (1.4)$$

حيث :

Y_i قيمة متغير الاستجابة في المشاهدة i

β_0 و β_1 وسيطا النموذج (ثوابت الانحدار)

X_i ثابت معلوم، ونقصد قيمة المتغير المستقل من أجل المشاهدة i

ε_i حدود الخطأ العشوائي غير مترابطة بمتوسط معدوم وتباين σ^2 ثابت، أي أن:

$E(\varepsilon_i) = 0$ و $var\{\varepsilon_i\} = \sigma^2$ و $cov\{\varepsilon_i, \varepsilon_j\} = 0$ لكل $i \neq j$ ، $i = 1, \cdots n$

يقال عن نموذج الانحدار (1.4) أنه بسيط وخطي في الوسطاء وخطي في المتغير المستقل. فهو بسيط لأنه يستخدم متغيراً مستقلاً واحداً فقط. وخطِّي الوسطاء لأنه لايظهر أي وسيط كأس أو مضروب بوسيط آخر أو مقسومة على وسيط آخر، وخطي في المتغير المستقل لأن هذا المتغير لايظهر إلا مرفوعاً للأس واحد، ويسمى النموذج الخطي في الوسطاء، والخطي في المتغير المستقل نموذجا من المرتبة الأولى أيضاً .

• سمات مهمة للنموذج

1) القيمة المشاهدة لـ Y في التجربة i هي مجموع مركبتين :

أ- الحد الثابت $\beta_0 + \beta_1 X_i$

ب- الحد العشوائي ε_i. وبالتالي يكون Y_iمتغيراً عشوائياً .

2) بما أنه $E(\varepsilon_i) = 0$ فإنه:

$$E\{Y_i\} = E(\beta_0 + \beta_1 X_i + \varepsilon_i\,) = \beta_0 + \beta_1 X_i + E(\varepsilon_i) = \beta_0 + \beta_1 X_i$$

لاحظ أن $\beta_0 + \beta_1 X_i$ يلعب دور الثابت، وهكذا تأتي الاستجابة عندما يكون مستوى في التجربة i هو X_i من توزيع احتمالي توقعه :

$$E\{Y_i\} = \beta_0 + \beta_1 X_i \qquad (2.4)$$

ومن ثم نعلم أن دالة الانحدار للنموذج (1.4) هي:

$$E\{Y\} = \beta_0 + \beta_1 X \qquad (3.4)$$

ذلك لأن دالة الانحدار تربط متوسطات توزيعات Y الاحتمالية الموافقة لقيم معطاة لـ X . بمستوى X .

3) تتجاوز قيمة Y التي شوهدت في التكرار i قيمة دالة الانحدار، أو تقل عنها، بحد خطأ قدره ε_i

4) يفترض لحد الخطأ ε_i تباين ثابت σ^2 . وينتج عن ذلك أن للاستجابات التباين الثابت نفسه:

$$var\{Y_i\} = \sigma^2 \qquad (4.4)$$

إذاً لدينا :

$$var(\beta_0 + \beta_1 X_i + \varepsilon_i) = var\{\varepsilon_i\} = \sigma^2$$

وهكذا يفترض نموذج الانحدار (1.4) أن لتوزيعات Y الاحتمالية التباين σ^2 نفسه، وذلك بغض النظر عن مستوى المتغير المستقل X .

5) يفترض أن حدود الخطأ غير مرتبطة. لذا لاتؤثر نتيجة أي تكرار للتجربة على حد الخطأ لتكرار آخر، من حيث كونه موجباً أم سالباً، صغيراً أم كبيراً. وبما أن حدي الخطأ ε_j و ε_i غير مرتبطين فإن الاستجابتين Y_j و Y_i غير مرتبطتين .

6) والخلاصة يتضمن نموذج الانحدار (1.4) أن مشاهدات متغير الاستجابة تأتي من توزيعات احتمالية توقعاتها $E(Y_i) = \beta_0 + \beta_1 X_i$ وتبايناتها σ^2 هي نفسها لكل مستويات X، بالإضافة إلى كون أي مشاهدتين Y_j و Y_i غير مرتبطتين.

• معاني وسطاء الانحدار

تسمى الوسطاء β_0 و β_1 في نموذج الانحدار(1.4) معاملات الانحدار. وتشير β_1 وهي ميل خط الانحدار، إلى التغير في متوسط توزيع Y الاحتمالي لكل وحدة زيادة في X،

والوسيط β_0 هو التقاطع الصادي لخط الانحدار. إذا تضمن مدى النموذج القيمة $X = 0$،
فإن β_0 تعطي متوسط توزيع Y الاحتمالي عند $X = 0$، وليس للوسيط β_0 أي تفسير
خاص بها كحد منفصل في نموذج الانحدار إذا لم يتضمن مجاله القيمة $X = 0$.

5.1.4 بيانات تحليل الانحدار

عادة، لا نعرف قيم وسطاء الانحدار في نموذج الانحدار (1.4). ونحتاج إلى تقديرهما من بيانات مناسبة. وفي الحقيقة، وكما ذكرنا سابقاً، ليس لدينا في معظم الأحيان، معرفة مسبقة وكافية عن المتغيرات المستقلة المناسبة، وعن الشكل الدالي لعلاقة لانحدار، (مثلاً خطية أو منحنية) ونحتاج إلى الاعتماد على خواص مميزة للبيانات كي نطوِّر نموذج الانحدار المناسب. ويمكن الحصول على بيانات لتحليل الانحدار بطرق غير تجريبية وتجريبية وسوف نتطرق لكل منهما بدوره .

• بيانات المشاهدة

بيانات المشاهدة هي بيانات غير تجريبية . حيث نحصل على بيانات من هذا القبيل بدون السيطرة على المتغير (المتغيرات) المستقل (المستقلة) موضع الاهتمام. فمثلاً رغب مسؤولو شركة دراسة العلاقة بين عمر المستخدم X وعدد أيام المرض Y خلال السنة الماضية . فاستخدموا لتحليل الانحدار بيانات تمَّ الحصول عليها من الملفات الشخصية، وبيانات كهذه هي بيانات مشاهدة. إذ لا يمكن السيطرة على المتغير المستقل. وفي معظم الأحيان يستند تحليل الانحدار على بيانات مشاهدة لأنه لا يمكننا في الغالب القيام بتجارب نتحكم في مسارها. ومن العيوب الرئيسة لبيانات المشاهدة أنها لاتزوِّدنا بمعلومات كافية عن العلاقات بين السبب والتأثير. فمثلاَ قد لا تعني العلاقة الإيجابية بين عمر المستخدم وعدد أيام المرض في مثال أفراد الشركة، أن عدد أيام المرض هو النتاج المباشر للعمر. فلربَّما كان مستخدمو الشركة الشباب يعملون أساساَ في الخارج في حين يعمل المستخدمون الكبار، عادةَ، في الداخل، ويتحمل مكان العمل المسؤولية الأهم في عدد أيام المرض. وحينما يعتمد تحليل انحدار، نقوم به لأغراض وصفية، على بيانات مشاهدة فإنه ينبغي تقصِّي ما إذا كانت هناك متغيرات مستقلة غير المتغيرات المعتمَدة في النموذج، يمكنها أن تُفسِّر بصورة أفضل علاقات سبب وتأثير .

• بيانات تجريبية

وكثيراَ ما يمكن إجراء تجربة نتحكم فيها ،لتزوِّدنا ببيانات تُمكِّننا من تقدير وسطاء الانحدار.افترض ،مثلاَ، أن شركة تأمين ترغب في دراسة العلاقة بين إنتاجية محلليها في معالجة الدعاوى وبين مدة التدريب. وتستخدم الدراسة ثمانية محللين. نختار ثلاثة منهم

بصورة عشوائية ليتمرَّنوا لمدة أسبوعين، واثنان منهم لثلاثة أسابيع، وثلاثة منهم لخمسة أسابيع. ثم نشاهد إنتاجيتهم خلال الأسابيع العشرة التالية .والبيانات التي نحصل عليها ستكون بيانات تجريبية لأن نوعاَ من السيطرة قد مورس على المتغير المستقل وهو طول فترة التدريب. وعندما تمارَس القيود على المتغير (المتغيرات) المستقل (المستقلة) من خلال تخصيص عشوائي، فهنا البيانات التجريبية الناتجة تزودنا بمعلومات أقوى بكثير عن علاقات السبب والتأثير من تلك التي توفرها بيانات المشاهدة. والسبب في ذلك يعود إلى أن العشوائية تتولى موازنة تأثيرات المتغيرات الأخرى التي يمكن أن تؤثر في المتغير التابع، مثل تأثير استعدادات المستخدم على الإنتاجية. وفي مصطلحات تصميم التجارب، تسمى فترة التدريب المخصصة للمحلِّل في مثال دراسة الإنتاجية معالجة. ويسمى المحللون المشاركون في الدراسة الوحدات التجريبية. وعندئذٍ تتمثَّل السيطرة على المتغيرات المستقلة بتخصيص معالجة لكل وحدة تجريبية بطرق عشوائية .

• تصميم تام العشوائية

التصميم تام العشوائية هو النوع الأساسي في التصميم الإحصائي المتعلِّق بعملية تخصيص المعالجات عشوائياً للوحدات التجريبية (والعكس بالعكس). ووفق هذا التصميم يتم التخصيص بالكامل عشوائياً. وتمنح هذه العشوائية التامة كل وحدة تجريبية الفرصة نفسها في تلقي أي من المعالجات، أو بصورة مكافئة، يكون لجميع الاختيارات الممكنة من الوحدات التجريبية التي خُصِّصت لمعالجات مختلفة، الفرصة نفسها. وعلى وجه الخصوص يكون التصميم تام العشوائية مفيداً عندما تكون الوحدات التجريبية متجانسة إلى حد كبير. وهذا التصميم مرن جداً ويلائم أي عدد من المعالجات، ويسمح بأحجام عينات مختلفة، وعيبه الرئيس هو أنه عندما تكون الوحدات التجريبية غير متجانسة، فلايكون هذا التصميم فعالاً بالمقارنة مع تصاميم إحصائية أخرى .

6.1.4 نظرة عامة على تحليل الانحدار

يمكن الإفادة من تحليل الانحدار في بيانات مشاهدة أو بيانات تجريبية وفق تصميم تام العشوائية (ونستطيع الانتفاع من تحليل الانحدار في بيانات من أنواع أخرى لتصميم التجارب) وسواء أكان البيان بيان مشاهدات أو بياناً تجريبياً، فمن الضروري أن تكون شروط نموذج الانحدار مناسبة للبيانات التي في حوزتنا .وسوف نبدأ دراستنا لتحليل الانحدار بالاستقراء عن وسطاء الانحدار في نموذج الانحدار الخطي البسيط (1.4) وفي الحالة النادرة التي يتوفر فيها معلومات سابقة أو نظرية تحدِّد لنا، بمفردها ، نموذج الانحدار المناسب تكون الاستقراءات المبنيَّة على نموذج الانحدار هذا ،هي الخطوة الأولى في تحليل الانحدار. وعلى كل حال، في الحالات الاعتيادية حيث لا نملك المعلومات الكافية لتحديد

نموذج الانحدار المناسب سلفاً، تكون الدراسة الاستكشافية للبيانات الخطوة الأولى كما هو موضَّح في مخطط التدفق في الشكل (1.4). وبالاستناد على هذا التحليل الاستكشافي المبدئي يُستحدث نموذج أولي أو أكثر للانحدار. وتُفحص نماذج الانحدار هذه من حيث صلاحيتها للبيانات التي في حوزتنا ثم تُنقَّح أو تُستحدث نماذج جديدة ،حتى يقتنع الدارس بأن نموذجاً بالذات من بينها هو النموذج المناسب. وعندئذ فقط تتم الاستقراءات بالاستناد إلى نموذج الانحدار هذا، كالاستقراءات حول وسطاء الانحدار للنموذج أوتنبوءات بمشاهدات جديدة .

7.1.4 تقدير دالة الانحدار

• طريقة المربعات الصغرى

لإيجاد مقدرات "جيدة" لوسطاء الانحدار β_0 و β_1 سوف نستخدم طريقة المربعات الصغرى. و لكل مشاهدة عينة (X_i, Y_i)، تأخذ طريقة المربعات الصغرى انحراف Y_iعن قيمته المتوقعة.

$$Y_i - (\beta_0 + \beta_1 X_i) \qquad (5.4)$$

على وجه الخصوص تتطلَّب طريقة المربعات الصغرى اعتبار مجموع مربعات الانحرافات الـn ويرمز لهذا المعيار بـ Q :

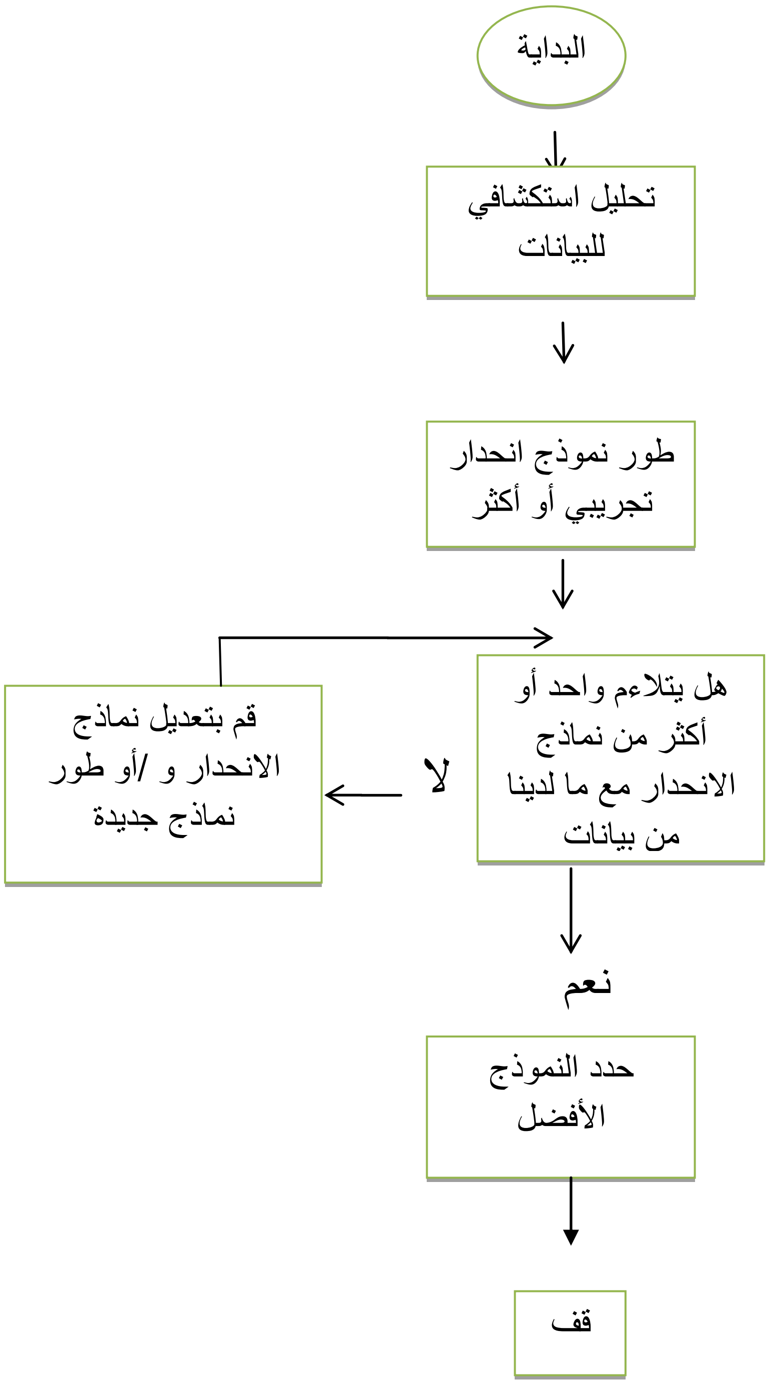
$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^{2} \qquad (6.4)$$

وطبقاً لطرق المربعات الصغرى تكون تقديرات β_0 و β_1 تلك القيم b_0 و b_1 على الترتيب، التي تجعل المعيار أصغر ما يمكن، وذلك من أجل مشاهدات العينة المعطاة(X_i, Y_i)

والهدف من طريقة المربعات الصغرى، هو إيجاد تقديرات b_0 و b_1 لـ β_0 و β_1 على الترتيب، يكون من أجلها أصغر ما يمكن. وبمعنى، ستأتي مناقشته بعد قليل، تزودنا هذه التقديرات بتوفيق "جيد" لدالة انحدار خطي .

حيث : $i = 1,2 \cdots, n$

شكل (1.4) :منهجية نموذجية لبناء نموذج الانحدار



• مقدرات المربعات الصغرى :

يمكن الحصول على المقدرات b_0 و b_1 التي تحقق قاعدة المربعات الصغرى بطريقتين أساسيتين. في الأولى يمكن استخدام طرق البحث العددية التي تحسب بصورة متناسقة معيار المربعات الصغرى Q لتقديرات b_0 و b_1 مختلفة حتى نعثر على تلك التي تجعل Q أصغر ما يمكن. وفي الطريقة الثانية نجد بأسلوب تحليلي القيم b_0, b_1 التي تجعل Q أصغر ما يمكن. وتكون الطريقة التحليلية ممكنة عندما لايكون النموذج معقداً من الناحية الرياضية كما هو الحال هنا.

ويمكن إثبات أن القيم b_0 و b_1 التي تجعل Q أصغرياً لأي مجموعة محددة من بيانات عينة حجمها n تُعطى بالمعادلتين الآتيتين :

$$\sum_{i=1}^{n} Y_i = nb_0 + b_1 \sum_{i=1}^{n} X_i \qquad (7a.4)$$

$$\sum_{i=1}^{n} X_i Y_i = b_0 \sum_{i=1}^{n} X_i + b_1 \sum_{i=1}^{n} {X_i}^2 \qquad (7b.4)$$

وتسمى المعادلتان (7.4) المعادلتين الناظميتين، وتسمى b_0 و b_1 المقدرات النقطية لـ β_0 و β_1 على الترتيب. وتُحسب الكميات $\sum X_i$ و $\sum Y_i$ وشبيهاتها في (7.4) من مشاهدات العينة (X_i, Y_i). ومن ثم يمكن حل المعادلتين معاً من أجل b_0 و b_1 ، وكبديل آخر يمكن الحصول على b_0 و b_1 مباشرة كما يلي:

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{(\sum_{i=1}^{n} {X_i}^2) - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$(8a.4)$$

$$b_0 = \frac{1}{n}\left(\sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i\right) = \bar{Y} - b_1 \bar{X} \qquad (8b.4)$$

حيث $\bar{X}$ و $\bar{Y}$ متوسطا المشاهدات X_i و Y_iعلى الترتيب، $i = 1,2 \cdots , n$

ملاحظة : يمكن اشتقاق طرفي المعادلتين الناظميتين (7.4) باستخدام حساب التفاضل بالنسبة للوسيطين β_0 وβ_1 . فمن أجل مشاهدات (X_i, Y_i) معطاة تكون الكمية Q دالة في β_0 وβ_1

ويمكن الحصول على قيم β_0 وβ_1 التي تجعل Q أصغر ما يمكن باشتقاق Q جزئياً

بالنسبة لـ β_0 وβ_1 لنجد :

$$\frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)$$

$$\frac{\partial Q}{\partial \beta_1} = -2X_i \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)$$

ثم نضع المشتقتين الجزئيتين مساويتين للصفر، مستخدمينb_0 وb_1 كرمزين لقيم β_0 و β_1 على الترتيب، التين تجعلان Q أصغر ما يمكن لنحصل على :

$$-2 \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i) = 0$$

$$-2X_i \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i) = 0$$

ونحصل بالتبسيط على:

$$\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i) = 0$$

$$\sum_{i=1}^{n} X_i (Y_i - b_0 - b_1 X_i) = 0$$

وبفك القوسين نجد :

$$\sum_{i=1}^{n} Y_i - nb_0 - b_1 \sum_{i=1}^{n} X_i = 0$$

$$\sum_{i=1}^{n} X_i Y_i - b_0 \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} {X_i}^2 = 0$$

حيث : $i = 1,2 \cdots, n$

ومنها نحصل على المعادلتين الناظميتين (7.4) وذلك بإعادة ترتيب الحدود. وباختبار المشتقات الجزئية الثانية نرى أن القيمة الصغرى تتحقق عند مقدري المربعات الصغرى b_0 و b_1 .

- **خواص مقدرات المربعات الصغرى**

تعرض مبرهنة مهمة تسمى مبرهنة غاوص-ماركوف مايلي :

تعليق(9.4): تحت شروط نموذج الانحدار (1.4) تكون مقدرات المربعات الصغرى b_0 و b_1 في (8.4) غير منحازة ولها أصغر تباين بين كافة المقدرات الخطية غير المنحازة.
وتُصرِّح هذه المبرهنة، أولاً أن كل من b_0 و b_1 مقدران غير منحازين. و هكذا نكتب :

$$E\{b_0\} = \beta_0$$

$$E\{b_1\} = \beta_1$$

ولذلك لايميل المقدر ميلاً منتظماً، إلى التقدير بالزيادة أو التقدير بالنقصان .

ثانيا :تعرض المبرهنة أن تباين توزيع المعاينة لـ b_0 و b_1 أقل من تباين أي من المقدرات الأخرى التي تنتمي إلى صف خاص من المقدرات . و هكذا، تكون مقدرات المربعات الصغرى أكثر إحكاماً من أي من هذه المقدرات .ويتألف صف المقدرات، الذي تتصدره مقدرات المربعات الصغرى كأفضل ما فيه، من جميع المقدرات غير المنحازة والتي هي دوال خطية في المشاهدات Y_1،…، Y_n والمقدران b_0 و b_1 هما دالتان من هذا الصف، أي دالتان خطيتان في الـ Y_i. اعتبر ، مثلاً، b_1 فلدينا من (8a.4):

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

ويمكن كتابة المقدار b_1 على الشكل :

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \sum_{i=1}^{n} k_i\, (Y_i - \bar{Y})$$

حيث
$$k_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

وحيث إن الـ k_i ثوابت معروفة (لأن الـX_i ثوابت معروفة). يكون b_1 تركيباً خطياً في الـ Y_i، وبالتالي فهو مقدر خطي. وبالطريقة نفسها يمكن إثبات أن b_0 مقدر خطي. ومن بين جميع المقدرات الخطية غير المنحازة، فإن b_0 و b_1 هما الأقل تبايناً عند سحب عينات متكررة يبقى مستوى X فيها ثابتاً .

• تقدير نقطي لمتوسط الاستجابة

دالة الانحدار المقدَّرة :

إذا عُرفت مقدرات العينة b_0 و b_1 للمعلمتين في دالة الانحدار (3.4) :

$$E\{Y\} = \beta_0 + \beta_1 X$$

فإننا نقدر دالة الاحتمال كما يلي :

$$\hat{Y} = b_0 + b_1 X \qquad (10.4)$$

حيث $\hat{Y}$ هي القيمة المقدرة لدالة الانحدار عند المستوى X للمتغير المستقل .

سوف نسمي قيمة متغير الاستجابة، " استجابة"، ونسمي $E\{Y\}$ متوسط الاستجابة. وهكذا يكون متوسط الاستجابة هو متوسط التوزيع الاحتمالي لـ Y المقابل للمستوى للمتغير المستقل X ، وعندئذ يكون $\hat{Y}$ مقدراً نقطياً لمتوسط الاستجابة عندما يكون مستوى المتغير المستقل X. وكامتداد لمبرهنة غاوص –ماركوف (9.4)

يمكن إثبات أن $\hat{Y}$ مقدر غير منحاز لـ Y بتباين أصغري في صف المقدرات الخطية وغير المنحازة لـ$E\{Y\}$

وفي الحالات قيد الدراسة، سوف نسمي $\hat{Y}_i$ حيث :

$$\hat{Y}_i = b_0 + b_1 X_i \qquad (11.4)$$

القيمة التوفيقية للمشاهدة i. وهكذا ننظر إلى القيمة التوفيقية $\hat{Y}$ كشيئ مختلف عن القيمة الملحوظة Y_i.

• الرواسب

الراسب الـ i هو الفرق بين القيمة الملحوظة Y_i والقيمة التوفيقية المقابلة $\hat{Y}_i$ وإذا رمزنا بـ e_i لهذا الراسب يمكننا كتابة :

$$e_i = Y_i - \hat{Y}_i = Y_i - b_0 - b_1 X_i \qquad (12.4)$$

وينبغي لنا التمييز بين قيمة حد الخطأ في النموذج $\varepsilon_i = Y_i - E\{Y_i\}$ والراسب $e_i = Y_i - \hat{Y}_i$.

فالأول يعني انحراف Y_i الرأسي عن خط الانحدار الحقيقي غير المعروف، وبالتالي فهو غير معروف. ومن جهة أخرى فإن الراسب هو الانحراف الرأسي لـ Y_i عن القيمة التوفيقية $\hat{Y}_i$ ، على خط الانحدار المقدر .

والرواسب مفيدة جداً في دراسة ما إذا كان نموذج انحدار معين مناسباً للبيانات التي في حوزتنا .

• خواص خط الانحدار التوفيقي

لخط الانحدار الموفق بطريقة المربعات الصغرى عدد من الخواص التي تستحق الذكر :

1- مجموع الرواسب يساوي صفراً $\sum_{i=1}^{n} e_i = 0$، وبالطبع قد تحدث أخطاء نتيجة تدوير الأرقام العشرية لأي مشاهدة معينة مما يجعل مجموع الرواسب غير مساو للصفر تماماً.

2- يكون مجموع مربعات الرواسب $\sum_{i=1}^{n} {e_i}^2$ أصغرياً. وهذا هو المتطلَّب الذي كان ينبغي تحققه عند استنباط مقدرات المربعات الصغرى لوسطاء الانحدار .

3- مجموع القيم الملحوظة Y_i يساوي مجموع القيم التوفيقية $\hat{Y}_i$:

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$$

4- يكون مجموعة الرواسب الموزونة صفراً عندما يوزن راسب المشاهدة i بمستوى المتغير المستقل في تلك المشاهدة :

$$\sum_{i=1}^{n} X_i e_i = 0$$

5- يكون مجموع الرواسب المرجَّحة صفراً عندما يوزن راسب المشاهدة i بالقيمة التوفيقية لمتغير الاستجابة للمشاهدة i:

$$\sum_{i=1}^{n} \hat{Y}_i e_i = 0$$

6- يمر خط الانحدار دائماً بالنقطة $(\bar{X}, \bar{Y})$.

- **تعليقات**

1 -يمكن استنتاج الخواص الست للرواسب مباشرة من المعادلة الناظمية للمربعات الصغرى .

2- تنطبق خواص الرواسب التي لاحظناها هنا على نموذج الانحدار (1.4) ولا تنطبق هذه الخواص على جميع نماذج الانحدار .

8.1.4 تقدير تباين حدود الأخطاء σ^2

ولأغراض متعددة نحتاج إلى تقدير التباين σ^2 لحدود الأخطاء ε_i في نموذج الانحدار (1.4) فكثيراً مانرغب في الحصول على مؤشر عن تباين التوزيع الاحتمالي لـ Y وبالإضافة إلى ذلك ، يتطلب العديد من الاستقراءات حول دالة الانحدار والتنبؤ عن Y ، تقدير σ^2 .

• تقدير نقطي لـ σ^2

• مجتمع بمفرده :

كي نضع الأساس لتطوير مقدر لـ σ^2 في نموذج الانحدار (1.4) لنعتبر لهنيهة المسألة الأبسط، وهي المعاينة من مجتمع بمفرده. وللحصول على تباين العينة s^2 نبدأ باعتبار انحراف المشاهدة Y_i عن المتوسط المقدر $\bar{Y}$ فنربعه، ثم نجمع جميع هذه الانحرافات المربعة:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

ويسمى مجموع كهذا، مجموع مربعات . وبعد ذلك نُقسِّم مجموع المربعات على درجات الحرية المرتبطة به. والعدد هنا هو $(n-1)$ فقد خسرنا درجة حرية واحدة نظراً لاستخدام المقدر $\bar{Y}$ بدلاً من متوسط المجتمع . ويكون تباين العينة المعتاد هو المقدر الناتج :

$$s^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}$$

وهو مقدر غير منحاز للتباين σ^2 لمجتمع لانهائي. وغالباً ما يسمى تباين العينة متوسط مربعات ، لأن مجموع المربعات قُسّم على عدد درجات الحرية المناسبة .

• نموذج الانحدار :

لايختلف المنطق في تطوير مقدر لـσ^2 في نموذج الانحدار عنه عند المعاينة من مجتمع منفرد . ولنتذكر لهذا الغرض، بما أن تباين كل مشاهدة Y_i هو σ^2 (سمات نموذج الانحدار) وهو نفسه لكل حد خطأ ε_i. ونحتاج مرة ثانية حساب مجموع مربعات الانحرافات ، ولكن ينبغي إدراك أن الـY_i ، جاءت من توزيعات احتمالية مختلفة لها متوسطات تختلف باختلاف المستوى X_i. وهكذا ينبغي حساب انحراف مشاهدة Y_i عن متوسطها المقدر الخاص بها $\hat{Y}_i$. ومن ثم تكون الانحرافات هي الرواسب: $e_i = Y_i - \hat{Y}_i$

ويكون مجموع المربعات المناسب ويرمز له بـ SSE

$$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)^2 = \sum_{i=1}^{n} {e_i}^2 \qquad (13.4)$$

حيث يرمز SSE لمجموع مربعات الخطأ أو لمجموع مربعات الرواسب .

ويرتبط بمجموع المربعات SSE ، $(n-2)$ درجات من الحرية. وخسرنا درجتي حرية لأنه كان علينا تقدير كل من β_0 و β_1 من أجل الحصول على المتوسطات المقدرة $\hat{Y}_i$.
وبالتالي فإن متوسط المربعات المناسب ويرمز له بـ MSE هو :

$$MSE = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2} = \frac{\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)^2}{n-2}$$
$$= \frac{\sum_{i=1}^{n} {e_i}^2}{n-2} \qquad (14.4)$$

حيث يرمز MSE لمتوسط مربعات الخطأ أو متوسط مربعات الرواسب. ويمكن إثبات أن MSE مقدر غير منحاز لـ σ^2 في نموذج الانحدار (1.4) :

$$E\{MSE\} = \sigma^2 \qquad (15.4)$$

ويكون مقدر الانحراف المعياري ببساطة الجذر التربيعي الموجب لـ MSE .

• صيغ حسابية بديلة

يوجد عدد من الصيغ الحسابية البديلة لـSSE . وفيما يلي ثلاث صيغ منها :

$$SSE = \sum_{i=1}^{n} {Y_i}^2 - b_0\sum_{i=1}^{n} Y_i - b_1\sum_{i=1}^{n} X_iY_i \qquad (16a.4)$$

$$SSE = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 - \frac{[\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad (16b.4)$$

$$SSE = \left[\sum_{i=1}^{n} {Y_i}^2 - \frac{(\sum_{i=1}^{n} Y_i)^2}{n}\right] - \frac{\left[\sum_{i=1}^{n} X_iY_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_i}{n}\right]^2}{\sum_{i=1}^{n} {X_i}^2 - \frac{\sum_{i=1}^{n}(X_i)^2}{n}}$$

$$(16c.4)$$

- **تعليقات**

1- تكون الصيغة $(16a.4)$ مفيدة إذا تم حساب b_0 و b_1

خلاف ذلك تكون $(16c.4)$،$(16b.4)$ هما الأكثر مباشرة .

2- في الصيغة $(16a.4)$ يجب حساب المقدرات b_0 و b_1 إلى عدد كبير من الأرقام العشرية كي نحصل على نتائج موثوقة لـSSE .

3- لاتوفر أي من الصيغ البديلة الثلاث الرواسب e_i بصورة صريحة. وكما ذكرنا سابقاً، فإن الرواسب مفيدة في دراسة صلاحية النموذج أو مصداقيته .

9.1.4 نموذج انحدار بخطأ طبيعي

مهما يكن الشكل الدالي لتوزيع ε_i (ومن ثَمَّ لِـ Y_i)، توفر طريقة المربعات الصغرى مقدرات نقطية غير منحازة لـ β_0 و β_1 لها تباين أصغري بين جميع المقدرات الخطية غير المنحازة. ولصياغة تقديرات بفترة أو القيام باختبار فرضيات حولها سنحتاج، على أي حال، لوضع افتراض حول الشكل الدالي لتوزيع الـ ε_i، والافتراض المعتاد هو أن حدود الخطأ موزعة طبيعياً، وسنتبنى هنا هذا الافتراض. ولا يتوقف تغيرها العشوائي على المتغير المستقل X، ويُسهِّل حد الخطأ الطبيعي مفاهيم تحليل الانحدار تسهيلاً كبيراً، وله ما يبرره في العديد من الحالات في واقع الحياة التي يُطبَّق فيها تحليل الانحدار.

- **النموذج :**

يُعرَّف نموذج الانحدار الطبيعي كما يلي:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (17.4)$$

حيث :

Y_i هي الاستجابة الملحوظة في المشاهدة i ، $i = 1,2 \cdots, n$

X_i ثابت معروف، مستوى المتغير المستقل في المحاولة i .

β_0 و β_1 وسيطان يمثلان ثوابت الانحدار

ε_i مستقلة ولها التوزيع الطبيعي بمتوسط 0 وتباين σ^2

$i = 1,2 \cdots, n$

• تعليقات

1- يمثل الرمز $N(0,\sigma^2)$ "موزعة طبيعيا"، بمتوسط 0 وتباين σ^2 .

2- نموذج الخطأ الطبيعي (17.4) هو ذاته نموذج الانحدار (1.4) بتوزيع خطأ غير محدد، ماعدا أن النموذج (17.4) يفترض أن الأخطاء ε_i موزعة طبيعياً .

3- حيث يفترض نموذج الانحدار (17.4) أن الأخطاء موزعة طبيعياً فإن فرضية عدم الارتباط للأخطاء ε_i في نموذج الانحدار (1.4) تصبح فرضية استقلال في نموذج الخطأ الطبيعي.

4- يتضمن نموذج الانحدار (17.4) أن الـ Y_i متغيرات عشوائية مستقلة وطبيعية بتوقع $E\{Y_i\} = \beta_0 + \beta_1 \mathrm{X_i}$ وتباين σ^2،

5- والسبب الرئيس لتبرير فرضية طبيعية الخطأ في العديد من الحالات هو أن حدود الخطأ تمثل أكثر ما تمثل، تأثيرات لعديد من العوامل التي لا يذكرها النموذج صراحة، وهي تؤثر إلى حد ما في متغير الاستجابة، ولا يتوقف تغيرها العشوائي على المتغير المستقل X . وكذلك قد تكون هناك أخطاء قياس عشوائية في تسجيل Y. ومما تقدم وحيث إن لهذه التأثيرات العشوائية درجة من الاستقلال فيما بينها، فإن حد الخطأ المركب ε_i الذي يمثل جميع هذه العوامل يميل للإذعان لمبرهنة النهاية المركزية فيتقارب توزيع حد الخطأ إلى الطبيعي عندما يصبح عدد العوامل المؤثرة كبيراً. وسبب آخر للتبرير المتواتر لافتراض طبيعية حدود الخطأ، هو استناد طرق التقدير والاختبارات على توزيع t وهو توزيع لايتأثر بحيدان معتدل عن الطبيعية. وهكذا فإنه إذا لم يكن الحيدان عن الطبيعية خطيراً، وخصوصاً فيما يتعلق بالالتواء فإن معامل الثقة الحقيقي ومخاطر الوقوع بأخطاء ستكون قريبة من المستويات الموافقة لتوزيع طبيعي.

• تقدير الوسطاء بطرق الإمكانية العظمى

عندما يتحدد الشكل الدالي للتوزيع الاحتمالي لحدود الأخطاء، يمكن الحصول على مقدرات للوسطاء β_0 و β_1 و σ^2 بطريقة الإمكانية العظمى. وتستخدم هذه الطريقة التوزيع الاحتمالي المشترك لمشاهدات العينة. وعندما ينظر إلى هذا التوزيع الاحتمالي المشترك

كدالة في الوسطاء علماً أن مشاهدات العينة معروفة، فتدعى عندئذ دالة الإمكانية. وتكون دالة الإمكانية من أجل الانحدار بأخطاء طبيعية،علماً أن مشاهدات العينة هي $Y_1 \cdots$ ، Y_n كما يلي :

$$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{1}{2\sigma^2}(Y_i - \beta_0 - \beta_1 \mathrm{X_i})^2\right]$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 \mathrm{X_i})^2] \qquad (18.4)$$

وقيم β_0 و β_1 و σ^2 التي تجعل دالة الإمكانية هذه أعظم ما يمكن هي مقدرات الإمكانية العظمى

الوسيط	مقدر الإمكانية العظمى	
β_0	b_0 مثل $(8b.4)$	
β_1	b_1 مثل $(8a.4)$	(19.4)
σ^2	$\hat{\sigma}^2 = \frac{\sum(Y_i - \hat{Y}_i)^2}{n}$	

وهكذا تكون مقدرات الإمكانية العظمى لـ β_0 و β_1 هي المقدرات نفسها التي توفرها طريقة المربعات الصغرى. ومقدر الإمكانية العظمى $\hat{\sigma}^2$ منحاز. ومن المعتاد استخدام المقدر MSE غير المنحاز. ونلاحظ أن المقدر غير المنحاز MSE يختلف اختلافاً طفيفاً عن مقدر الإمكانية العظمى $\hat{\sigma}^2$. خاصة عندما لا يكون n صغيراً :

$$MSE = \frac{n}{n-2}\,\hat{\sigma}^2 \qquad (20.4)$$

• تعليقات

1- حيث إن مقدرات الإمكانية العظمى b_0 و b_1 هي بالذات مقدرات المربعات الصغرى، فإن لها جميع خواص مقدرات المربعات الصغرى :

(أ) إنها غير منحازة.

(ب) لها أقل تباين بين جميع المقدرات الخطية غير المنحازة.

وبالإضافة إلى ذلك تمتلك مقدرات الإمكانية العظمى b_0 و b_1 لنموذج الانحدار بأخطاء طبيعية (17.4) خواص أخرى مرغوبة هي :

(جـ) إنها متسقة .

(د) إنها كافية .

(هـ) إنها غير منحازة أصغرية التباين، أي لها التباين ضمن صف جميع المقدرات غير المنحازة (خطية وغيرها).

وهكذا من أجل نماذج الخطأ الطبيعي، يمتلك المقدران b_0 و b_1 العديد من الخواص المرغوبة.

2-نحصل على القيم β_0 و β_1 و σ^2 التي تجعل دالة الإمكانية L في (18.4) أعظم ما يمكن بأخذ المشتقات الجزئية لـ L بالنسبة لـ β_0 و β_1 و σ^2

ومساواة كل منها بالصفر ثم حل نظام المعادلات الناتج. ويمكن استخدام $\log_e L$

بدلاً من L، لأن كلاً من $\log_e L$ و L تبلغ قيمتها العظمى عند القيم نفسها لـ β_0

و β_1 و σ^2 .

$$\text{Log}_e L = -\frac{n}{2}\log_e 2\pi - \frac{n}{2}\log_e \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 \text{X}_\text{i})^2 \qquad (21.4)$$

وتكون المشتقات الجزئية لهذه الإمكانية اللوغاريتمية أسهل كثيراً وهي تعطي :

$$\frac{\partial(\log_e L)}{\partial\beta_0} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 \mathrm{X_i})$$

$$\frac{\partial(\log_e L)}{\partial\beta_1} = \frac{1}{\sigma^2}\sum_{i=1}^{n}X_i(Y_i - \beta_0 - \beta_1 \mathrm{X_i})$$

$$\frac{\partial(\log_e L)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 \mathrm{X_i})^2$$

نجعل الآن هذه المشتقات الجزئية مساوية للصفر ونضع b_0 و b_1 و $\hat{\sigma}^2$ بدلاً من β_0 و β_1 و σ^2 ونحصل بعد بعض التبسيطات على :

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1 \mathrm{X_i}) = 0 \qquad (22a.4)$$

$$\sum_{i=1}^{n}X_i(Y_i - b_0 - b_1 \mathrm{X_i}) = 0 \qquad (22b.4)$$

$$\frac{\sum_{i=1}^{n}(Y_i - b_0 - b_1 \mathrm{X_i})^2}{n} = \hat{\sigma}^2 \qquad (22c.4)$$

تتطابق العلاقتان $(22a.4)$و $(22b.4)$

مع معادلتي المربعات الصغرى الناظميتين المعطاتين سابقاً في (7.4)

أما $(22c.4)$ فهو مقدر σ^2 المنحاز المعطى سابقاً في (19.4).

2.4 الانحدار المتعدد

(جون نتر، وليام وازرمان، ميخائيل كتنر(1995))

1.2.4 نماذج الانحدار المتعدد

- **الحاجة لعدة متغيرات مستقلة:**

في أمثلة انحدار خطي عديدة في الحياة، سوف لا يزودنا متغير مستقل واحد بوصف ملائم طالما أن عدداً من المتغيرات الرئيسة تؤثر بطرق مهمة ومتميزة في متغير الاستجابة. وفضلاً عن ذلك، فكثيراً ما سيجد المرء في حالات من هذا النوع، أن التنبؤات بقيم متغير الاستجابة المستندة إلى نموذج يتضمن متغيراً مستقلاً واحداً فقط، هي من عدم الدقة بحيث تصبح عديمة الفائدة. والنموذج الأكثر تعقيداً المتضمن لمتغيرات مستقلة إضافية، هو عادةً أكثر عوناً في تقديم تنبؤات دقيقة كفاية لمتغير الاستجابة.

ويمكن الاستفادة من نماذج الانحدار المتعدد التي سنصفها الآن في كل من بيانات المشاهدة والبيانات التجريبية الناتجة عن تصميم تام العشوائية .

- **نموذج من المرتبة الأولى مع متغيرين مستقلين :**

عندما يوجد متغيران مستقلان X_1, X_2 يدعى نموذج الانحدار :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad (23.4)$$

حيث : $i = 1,2 \cdots, n$

نموذجاً من المرتبة الأولى مع متغيرين مستقلين.وإن النموذج من المرتبة الأولى هو نموذج خطي في الوسطاء وخطي في المتغيرات المستقلة ويرمز Y_i كالعادة للاستجابة في التكرار i

X_{i1} وX_{i2} هما قيمتا المتغيرين المستقلين في التكرار i. ووسطاء النموذج هي β_0 ، β_1 و β_2 ، ε_i هو حد الخطأ.

وبافتراض $\mathrm{E}\{\varepsilon_i\} = 0$ تكون دالة انحدار النموذج (23.4) هي :

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \qquad (24.4)$$

وبصورة مشابهة للانحدار الخطي البسيط حيث تكون دالة الانحدار

$E\{Y\} = \beta_0 + \beta_1 X$ خطاً مستقيماً فإن دالة الانحدار (24.4) هي مستوي.

وكثيرا ماتدعى دالة الانحدار في الانحدار المتعدد سطح الانحدار أو سطح الاستجابة .

• معنى وسطاء الانحدار

لنعتبر الآن معنى وسطاء الانحدار في دالة الانحدار المتعدد (24.4)، فالوسيط β_0 هو تقاطع Y مع مستوي الانحدار. وإذا امتد مجال النموذج ليشمل $X_1 = 0$ و $X_2 = 0$

فإن β_0 يعطي متوسط الاستجابة عند $X_1 = 0$ و $X_2 = 0$ وفيما عدا ذلك لايكون لـ β_0 كحد منفصل في نموذج الانحدار أي معنى محدد. ويشير الوسيط β_1 إلى التغير في متوسط الاستجابة لكل زيادة بمقدار الواحد في X_1 وذلك عندما يبقى X_2 ثابتاً. وبالمثل، تشير β_2 إلى التغير في متوسط الاستجابة لكل زيادة بمقدار واحد في X_2 وذلك عندما يبقى X_1 ثابتاً. وعندما لايعتمد تأثير X_1 في متوسط الاستجابة على مستوى X_2، وفي المقابل لايعتمد تأثير X_2 على مستوى X_1، يقال إن للمتغيرين المستقلين تأثيرات تجميعية أو إنهما لا يتفاعلان.
وهكذا فإن نموذج الانحدار من المرتبة الأولى (23.4) مصمم لمتغيرين مستقلين تأثيراهما على متوسط الاستجابة تجميعيان أولا يتفاعلان وكثيراً ما يدعى الوسيطان β_1 و β_2 وسيطي انحدار جزئيين، لأنهما يعكسان التأثير الجزئي لمتغير مستقل عندما يكون المتغير المستقل الآخر مشمولاً في النموذج مع بقائه ثابتاً.

• نموذج من المرتبة الأولى بأكثر من متغيرين مستقلين

نعتبر الآن الحالة التي فيها($P-1$) من المتغيرات المستقلة $X_1 \cdots، X_{P-1}$ ويدعى نموذج الانحدار:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \qquad (25.4)$$

نموذج من المرتبة الأولى مع $(P-1)$ متغيراً مستقلاً ويمكن أيضاً كتابته على :

$$Y_i = \beta_0 + \sum_{k=1}^{p-1} \beta_k X_{ik} + \varepsilon_i \qquad (25a.4)$$

أو يمكن كتابته إذا جعلنا $X_{i0} \equiv 1$ على الشكل :

$$Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i \qquad (25b.4)$$

حيث: $X_{i0} \equiv 1$

وبفرض $E\{ \varepsilon_i\} = 0$ تكون دالة الاستجابة لنموذج الانحدار (25.4) هي :

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1} \qquad (26.4)$$

ودالة الاستجابة هذه هي فوق مستوي وهو مستو بأكثر من بعدين. ولم يعد من الممكن تصوير سطح الاستجابة هذا كما كنا قادرين على ذلك في حالة متغيرين مستقلين. ويشير الوسيط β_k إلى التغير في متوسط الاستجابة $E\{Y\}$ عند زيادة بمقدار الواحد في المتغير المستقل X_k، مع بقاء جميع المتغيرات المستقلة الأخرى في نموذج الانحدار ثابتة. ولنلاحظ أن تأثير أي متغير مستقل على متوسط الاستجابة في نموذج الانحدار (25.4) يبقى نفسه بصرف النظر عن المستويات التي ثبتنا عندها المتغيرات المستقلة الأخرى، وهكذا فإن نموذج الانحدار من المرتبة الأولى (25.4) مصمم لمتغيرات مستقلة تكون تأثيراتها على متوسط الاستجابة تجميعية وبالتالي فهي لاتتفاعل.

• نموذج الانحدار الخطي العام

وبصورة عامة ، ليس من الضروري أن تمثل المتغيرات $X_{P-1} \cdots$ ،X_1 في نموذج الانحدار متغيرات مستقلة مختلفة، ولذلك نعرف نموذج الانحدار الخطي العام ببساطة بدلالة متغيرات X مع حدود خطأ طبيعية، كما يلي :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{P-1} X_{i,P-1} + \varepsilon_i \qquad (27.4)$$

حيث: β_0، β_1، ....، β_{P-1} هي وسطاء

X_{i1} ، ....، $X_{i,P-1}$ ثوابت معروفة

ε_i مستقلة ولها التوزيع الطبيعي $N(0, \sigma^2)$ ، $i = 1,2 \cdots, n$

و إذا جعلنا $X_{i0} \equiv 1$ فيمكن كتابة نموذج الانحدار (27.4) كما يلي :

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{P-1} X_{i,P-1} + \varepsilon_i \qquad (27a.4)$$

حيث: $X_{i0} \equiv 1$ ، $i = 1,2 \cdots, n$

أو

$$Y_i = \sum_{k=0}^{P-1} \beta_k X_{ik} + \varepsilon_i \qquad (27b.4)$$

$$k = 1,2, \ldots .. P-1$$

حيث: $X_{i0} \equiv 1$

وباعتبار $E\{\varepsilon_i\} = 0$ فإن دالة الاستجابة لنموذج الانحدار (27.4) هي :

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{P-1} X_{P-1} \qquad (28.4)$$

وهكذا فإن نموذج الانحدار الخطي العام بحدود خطأ طبيعية يتضمن أن المشاهدات Y_i هي متغيرات طبيعية مستقلة

بمتوسط $E\{Y_i\}$ كما هو معطى في (28.4) و بتباين ثابت σ^2 .

ويحيط هذا النموذج الخطي العام بتشكيلة واسعة من الحالات سنذكر قليلاً منها الآن :

$(P-1)$ متغيراً مستقلاً :

عندما تمثل المتغيرات X_1، $\cdots X_{P-1}$

($P-1$) من المتغيرات المستقلة المختلفة فإن نموذج الانحدار الخطي العام (27.4) هو كما رأينا، نموذج من المرتبة الأولى لايتضمن تأثيرات تفاعل بين المتغيرات المستقلة .

• انحدار كثيرات الحدود:

لنعتبر نموذج الانحدار المنحني متغير مستقل واحد :

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 {X_i}^2 + \varepsilon_i \qquad (29.4)$$

إذا فرضنا $X_{i1} = X_i$ و $X_{i2} = X_i{}^2$ فيمكن كتابة (29.4) كما يلي :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

أي أن النموذج (29.4) هو حالة خاصة من نموذج الانحدار الخطي العام (27.4)

وبينما يوضح النموذج (29.4). نموذج انحدار منحن، دالة الاستجابة فيه تربيعية، فإن دوال استجابة على شكل كثيرات حدود من درجة أعلى هي أيضاً حالات خاصة من نموذج انحدار خطي عام

• متغيرات محوّلة :

لنعتبر النموذج :

$$logY_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad (30.4)$$

فسطح الاستجابة هنا معقَّد، ومع ذلك يمكن التعامل مع النموذج (30.4) كنموذج انحدار خطي عام. إذ لو جعلنا $\grave{Y}_i = logY_i$ فيمكن كتابة النموذج (30.4) كما يلي :

$$\grave{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

و هو في صيغة نموذج الانحدار الخطي العام (30.4) لقد اتفق أن المتغير التابع هو لوغاريتم Y.

ويمكن تحويل العديد من النماذج إلى نماذج انحدار خطية عامة و هكذا يمكن تحويل النموذج :

$$Y_i = \frac{1}{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i} \qquad (31.4)$$

إلى نموذج خطي عام بجعل $\grave{Y}_i = 1/Y_i$ ونجد عندئذ :

$$\grave{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

• تأثيرات تفاعل :

لنعتبر نموذج الانحدار بمتغيرين مستقلين X_1 و X_2 :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i \qquad (32.4)$$

فمعنى β_1 و β_2 هنا يختلف عن معناهما المعطى سابقاً بسبب وجود الحد الجدائي $\beta_3 X_{i1} X_{i2}$ ويمكن تبيان أن المتغير في متوسط الاستجابة المقابل لزيادة قدرها الواحد في X_1 مع بقاء X_2 ثابتاً هو:

$$\beta_1 + \beta_3 X_2 \qquad (33.4)$$

وبصورة مماثلة ، فإن التغير في متوسط الاستجابة المقابل لزيادة قدرها الواحد في X_2 مع بقاء X_1 ثابتاً هو :

$$\beta_2 + \beta_3 X_1 \qquad (34.4)$$

وبالتالي فإن كلاً من تأثير X_1 من أجل مستوى معطى لـ X_2 وتأثير X_2 من أجل مستوى معطى لـ X_1 يعتمد، في نموذج الانحدار (34.4)، على مستوى المتغير المستقل الآخر .وهكذا فإن β_1 في نموذج الانحدار المتضمن لحد جدائي، لم يعد يشير إلى التغير في متوسط الاستجابة المقابل لزيادة مقدارها الواحد في X_1 من أجل أي مستو معطى لـ X_2. ففي هذا النموذج يعتمد ذلك التأثير على مستوى X_2 ونموذج الانحدار(32.4)المتضمن لحد جدائي هو إذا مصمم لمتغيرات مستقلة تتفاعل تأثيراتها على المتغير التابع ويدعى الحد الجدائي $\beta_3 X_{i1} X_{i2}$ **حد التفاعل**. وبينما يبقى متوسط الاستجابة في نموذج الانحدار (32.4) دالة خطية في X_1 عندما يكون X_2 ثابتاً، إلا أن كلاً من الجزء المقطوع لدالة الاستجابة وميلها يتغيران مع تغير القيمة التي ثبتنا عندها مستوى X_2. ويصح الشيء نفسه عند اعتبار متوسط الاستجابة كدالة في X_2 مع بقاء X_1 ثابتاً.

وبالرغم من التعقيدات في نموذج الانحدار (32.4) فلا يزال من الممكن اعتباره كنموذج انحدار خطي عام. لتكن $X_{i3} = X_{i1} X_{i2}$ فيمكن أن نكتب (32.4) الآن كما يلي :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

وهوفي صيغة نموذج الانحدار الخطي العام (27.4).

• مركب من الحالات :

قد يضم نموذج انحدار عدداً من العناصر التي ذكرناها آنفاً ، ونبقى قادرين مع ذلك على معالجته كنموذج انحدار خطي عام . فلنعتبر نموذج انحدار بمتغيرين مستقلين كل منهما في صيغة تربيعية مع حد تفاعل:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 {X_{i1}}^2 + \beta_3 X_{i2}$$
$$+ \beta_4 {X_{i2}}^2 + \beta_5 X_{i1} X_{i2} \varepsilon_i \qquad (35.4)$$

ولنعرف $\mathrm{Z}_{i5} = X_{i1}X_{i2}$ $\quad \mathrm{Z}_{i4} = {X_{i2}}^2$ $\quad \mathrm{Z}_{i3} = X_{i2}$ $\quad \mathrm{Z}_{i2} = {X_{i1}}^2$
$\mathrm{Z}_{i1} = X_{i1}$

فيمكن عندئذ كتابة نموذج الانحدار (35.4) كما يلي :

$$Y_i = \beta_0 + \beta_1 \mathrm{Z}_{i1} + \beta_2 \mathrm{Z}_{i2} + \beta_3 \mathrm{Z}_{i3} + \beta_4 \mathrm{Z}_{i4} + \beta_5 \mathrm{Z}_{i5} \varepsilon_i$$

و هو في صيغة نموذج الانحدار الخطي العام (27.4) .

- **تعليقات:**

1- ينبغي أن يكون واضحاً أن نموذج الانحدار الخطي العام (27.4) غير مقصور على سطوح استجابة خطية. ويشير مصطلح "النموذج الخطي " إلى حقيقة أن (27.4) خطي في الوسطاء ولايشير إلى شكل سطح الاستجابة .

2- هناك بعض سطوح الاستجابة المعقَّدة عندما يكون لدينا متغيرين مستقلين بحيث يمكن تمثيلهما عن طريق نموذج الانحدار الخطي العام (27.4) .

2.2.4 نموذج انحدار خطي عام بدلالة المصفوفات :

سنقدم الآن النتائج الرئيسية لنموذج الانحدار الخطي العام (27.4) بدلالة المصفوفات. وكما لاحظنا فإن هذا النموذج يحيط بتشكيلة واسعة من الحالات الخاصة والنتائج التي سنقدمها قابلة للتطبيق على جميع هذه الحالات .

للتعبير عن نموذج الانحدار الخطي العام (27.4) :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

بدلالة المصفوفات نحتاج إلى تعريف المصفوفات التالية:

$$\boldsymbol{Y}_{n\times1}=\begin{bmatrix}\boldsymbol{Y_1}\\ \boldsymbol{Y_2}\\ \vdots\\ \boldsymbol{Y_n}\end{bmatrix} \quad , \quad \boldsymbol{X}_{n\times p}=\begin{bmatrix}1 & X_{11} & X_{12} & \dots & X_{1,p-1}\\ 1 & \mathrm{X}_{21} & \mathrm{X}_{22} & \dots & \mathrm{X}_{2,p-1}\\ \vdots & & & & \\ \vdots & & & & \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1}\end{bmatrix}$$

$$\boldsymbol{\beta}_{p\times1}=\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_{p-1}\end{bmatrix} \quad , \quad \boldsymbol{\varepsilon}_{n\times1}=\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix} \qquad (36.4)$$

يصبح نموذج الانحدار الخطي العام (27.4):

$$\mathbf{Y}_{\mathrm{n}\times1}=\mathbf{X}_{\mathrm{n}\times\mathrm{p}}\boldsymbol{\beta}_{\mathrm{p}\times1}+\boldsymbol{\varepsilon}_{\mathrm{n}\times1} \qquad (37.4)$$

حيث :

Y متجه الاستجابات

β متجه الوسطاء

X مصفوفة من الثوابت، قيم متغيرات التنبؤ

ε متجه من المتغيرات العشوائية الطبيعية المستقلة بتوقع $E\{\boldsymbol{\varepsilon}\}=0$ ومصفوفة تغاير $\boldsymbol{var}\{\boldsymbol{\varepsilon}\}=\boldsymbol{\sigma}^2\mathbf{I}$

وبالتالي فإن للمتجه العشوائي $\boldsymbol{Y}$ توقعاً :

$$\boldsymbol{E}_{\mathrm{n}\times1}\{\mathbf{Y}\}=\mathbf{X}\,\boldsymbol{\beta} \qquad (a)(37.4)$$

ومصفوفة تغاير **Y** هي :

$$\boldsymbol{var}_{\mathrm{n}\times\mathrm{n}}\{\mathbf{Y}\}=\boldsymbol{\sigma}^2\mathbf{I} \qquad (\mathrm{b})\ (37.4)$$

3.2.4 مقدرات المربعات الصغرى :

لنرمز بـ **b** لمتجه وسطاء الانحدار المقدرة b_0، b_1، ...... ، b_{p-1} :

$$\boldsymbol{b}_p = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ \vdots \\ b_{p-1} \end{bmatrix} \qquad (38.4)$$

معادلات المربعات الصغرى الناظمية لنموذج الانحدار الخطي العام (37.4) هي :

$$\mathbf{X}'\boldsymbol{X}\boldsymbol{b} = \mathbf{X}'\boldsymbol{Y} \qquad (39.4)$$

ومقدرات المربعات الصغرى هي :

$$\boldsymbol{b}_{p\times1} = (\mathbf{X}'\boldsymbol{X})^{-1}_{p\times p}\mathbf{X}'\boldsymbol{Y}_{p\times1} \qquad (40.4)$$

وفي حالة نموذج الانحدار (37.4) تكون مقدرات المربعات الصغرى هذه مقدِّرات الإمكانية العظمى أيضاً ولها جميع الخواص :

فهي غير منحازة، وغير منحازة بتباين أصغري، ومتسقة وكافية .

4.2.4 القيم التوفيقية والرواسب

لنرمز بـ $\widehat{\boldsymbol{Y}}$ لمتجه القيم التوفيقية $\widehat{Y}_i$ و $\boldsymbol{e}$ لمتجه عناصره الرواسب e_i

$$e_i = Y_i - \widehat{Y}_i \quad ، \quad i = 1,2 \cdots, n$$

$$\widehat{\boldsymbol{Y}}_{n\times1} = \begin{bmatrix} \widehat{\boldsymbol{Y}}_1 \\ \widehat{\boldsymbol{Y}}_2 \\ \vdots \\ \widehat{\boldsymbol{Y}}_n \end{bmatrix} \qquad \boldsymbol{e}_{n\times1} = \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_n \end{bmatrix} \qquad (41.4)$$

فيمكن تمثيل القيم التوفيقية على الشكل :

$$\widehat{\boldsymbol{Y}}_{n\times1} = \boldsymbol{X}\boldsymbol{b} \qquad (42.4)$$

ومتجه الرواسب على الشكل :

$$\boldsymbol{e}_{\mathrm{n}\times 1} = \mathbf{Y} - \widehat{\boldsymbol{Y}} = \mathbf{Y} - \boldsymbol{Xb} \qquad (43.4)$$

ويمكن التعبير عن متجه القيم التوفيقية $\widehat{\boldsymbol{Y}}$ بدلالة مصفوفة القبعة $\mathbf{H}$ (مصفوفة Hessian) كما يلي :

$$\widehat{\boldsymbol{Y}}_{n\times 1} = \mathbf{H}\,\mathbf{Y} \qquad (44.4)$$

حيث :

$$\mathbf{H}_{n\times n} = \boldsymbol{X}(\mathbf{X}'\boldsymbol{X})^{-1}\,\mathbf{X}' \qquad (44a.4)$$

وبصورة مماثلة يمكن التعبير عن متجه الرواسب كما يلي:

$$\boldsymbol{e}_{\mathrm{n}\times 1} = (\mathbf{I} - \mathbf{H})\mathbf{Y} \qquad (45.4)$$

ومصفوفة تغاير تباين الرواسب هي :

$$\boldsymbol{\sigma}^{2}_{\mathrm{n}\times\mathrm{n}}\{\mathbf{e}\} = \boldsymbol{\sigma}^{2}(\mathbf{I} - \mathbf{H}) \qquad (46.4)$$

وهي تقدر بما يلي :

$$\boldsymbol{s}^{2}_{\mathrm{n}\times\mathrm{n}}\{\mathbf{e}\} = \mathrm{MSE}(\mathbf{I} - \mathbf{H}) \qquad (47.4)$$

5.2.4 نتائج تحليل التباين

• مجموع مربعات ومتوسط مربعات

ومجاميع المربعات في تحليل التباين بدلالة المصفوفات هي:

$$SSTO = \boldsymbol{Y}'\,\mathbf{Y} - \left(\frac{1}{n}\right)\boldsymbol{YJY}' = \boldsymbol{Y}'\,\left[\mathbf{I} - \left(\frac{1}{n}\right)\,\mathbf{J}\right]\mathbf{Y} \qquad (48.4)$$

$$\begin{aligned} SSE = \boldsymbol{\acute{e}e} &= (\boldsymbol{Y} - \boldsymbol{Xb})'(\boldsymbol{Y} - \boldsymbol{Xb}) = \boldsymbol{Y}'\,\mathbf{Y} - \boldsymbol{b}'\mathbf{X}'\boldsymbol{Y} \\ &= \boldsymbol{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} \qquad (49.4) \end{aligned}$$

$$\mathrm{SSR} = \boldsymbol{b'}\mathbf{X'}\boldsymbol{Y} - \left(\frac{1}{n}\right)\boldsymbol{YJY'} = \boldsymbol{Y'}\left[\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \qquad (50.4)$$

حيث **J** هو $\mathrm{n} \times \mathrm{n}$ مصفوفة من المقادير 1 و **H** مصفوفة القبعة كماعرفناها في $(44a.4)$.

$SSTO$ كالمعتاد لها $(\mathrm{n}-1)$ درجة من الحرية تترافق معها، وSSE لها $\mathrm{n}-P$ من درجات الحرية تترافق معها باعتبار أننا نحتاج إلى تقدير P من الوسطاء في دالة الانحدار للنموذج (37.4). وأخيراً SSR لها $(P-1)$ درجة من الحرية تترافق معها ممثلة لعدد المتغيرات X وهي: $X_1, \dots, X_{P-1}$. ويبين الجدول (1.4) هذه النتائج لتحليل التباين بالإضافة إلى متوسطي المربعات $\mathrm{MS}R$ و MSE :

$$\mathrm{MS}R = \frac{SSR}{P-1} \qquad (51.4)$$

$$\mathrm{MSE} = \frac{SSE}{n-P} \qquad (52.4)$$

مصدر التغير	SS	df	MS
الانحدار	$\mathrm{SSR} = \boldsymbol{b'}\mathbf{X'}\boldsymbol{Y} - \left(\frac{1}{n}\right)\mathbf{YJ'Y}$	$P-1$	$\mathrm{MS}R = \frac{SSR}{P-1}$
الخطأ	$SSE = \boldsymbol{Y'}\,\mathbf{Y} - \boldsymbol{b'}\mathbf{X'}\boldsymbol{Y}$	$\mathrm{n}-\mathrm{P}$	$\mathrm{MSE} = \frac{SSE}{n-P}$
المجموع	$SSTO = \boldsymbol{Y'}\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{YJ'Y}$	$\mathrm{n}-1$	

الجدول (1.4)

وتوقع MSE هو σ^2 وتوقع $\mathrm{MS}R$ هو σ^2 مضافاً إليه كمية غير سالبة

• الاختبار F لعلاقة انحدار :

ولاختبار ما إذا كانت توجد علاقة انحدار بين المتغير التابع Y ومجموعة المتغيرات X وهي $X_1 \cdots ، X_{P-1}$ أي نختبر صحة الفرضية :

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_{P-1} = 0 \qquad (53a.4)$$

مقابل الفرضية :

$$H_a: \text{ليست كل } \beta_k \ (k = 1, \cdots, p-1) \text{ تساوي الصفر}$$

نستخدم إحصاءة الاختبار:

$$F^* = \frac{\text{MS}R}{\text{MSE}} \qquad (53b.4)$$

وقاعدة القرار عند ضبط الخطأ من النوع الأول عند α هي :

$$\text{إذا كان } F^* \leq F(1-\alpha; p-1, n-p) \text{ نقبل } H_0 \text{ ونرفض } H_a \qquad (53c.4)$$

$$\text{إذا كان } F^* > F(1-\alpha; p-1, n-p) \text{ نرفض } H_0 \text{ ونقبل } H_a$$

ووجود علاقة انحدار لذاتها لا يؤكد بالطبع إمكانية الوصول إلى تنبؤات مفيدة باستخدام هذه العلاقة .

• معامل التحديد المتعدد:

يعرف معامل التحديد المتعدد ونرمز له بـ R^2 كما يلي:

$$R^2 = \frac{\text{SSR}}{\text{SSTO}} = 1 - \frac{SSE}{SSTO} \qquad (4.54)$$

وهو يقيس التخفيض النسبي في التغير الكلي لـ Y الذي يترافق مع استخدام مجموعة المتغيرات X وهي $X_1 ، \cdots . X_{P-1}$ ويُختزل معامل التحديد المتعدد R^2 إلى معامل التحديد r^2 الخاص بانحدار خطي بسيط عندما يكون $p - 1 = 1$ أي عندما يوجد متغير مستقل واحد في نموذج الانحدار (37.4) ولدينا تماماً كما في حالة r^2 :

$$0 \leq R^2 \leq 1 \qquad (55.4)$$

ويفترض R^2 القيمة 0 عندما تكون جميع المقادير: $b_k = 0$, $(k = 1, \cdots, p - 1)$.
ويأخذ R^2 القيمة 1 عندما تقع جميع المشاهدات Y على سطح الاستجابة التوفيقي مباشرة أي عندما يكون $Y_i = \hat{Y}_i$ من أجل جميع قيم i

• تعليقات :

1) للتمييز بين معاملي التحديد في حالتي انحدار بسيط وانحدار متعدد سندعو r^2 من الآن فصاعداً معامل التحديد البسيط .

2) يمكن تبيان إمكانية النظر إلى معامل التحديد المتعدد R^2 كمعامل تحديد بسيط r^2 بين الاستجابة Y_i والقيم التوفيقية $\hat{Y}_i$.

3)القيم الكبيرة لـ R^2 لاتضمن بالضرورة أن النموذج الذي تم توفيقه هو نموذج مفيد. وعلى سبيل المثال، يمكن أن تكون المشاهدات قد أُخذت عند مستويات قليلة فقط للمتغيرات االمستقلة. وبالرغم من ارتفاع R^2 في هذ الحالة فقد لايكون النموذج مفيداً لأن معظم التنبؤات ستحتاج إلى توسع في الاستقراء خارج منطقة المشاهدات، وأيضاً ،حتى عندما يكون R^2 كبير فقد يكون MSE كبيراً إلى حد لاتكون معه الاستقراءات مفيدة إذا أردنا لدقة هذه الاستقراءات أن تكون عالية .

4)إضافة المزيد من المتغيرات المستقلة إلى النموذج يمكن أن يؤدي فقط لزيادة R^2 ولا يخفضها، لأن SSE لايمكن أن تصبح أبداً أكبر مع مزيد من المتغيرات المستقلة، ولأن $SSTO$ تبقى دائماً نفسها من أجل مجموعة معطاة من الاستجابات .وبما أنه يمكننا في الغالب، جعل R^2 كبيرة باعتماد عدد كبير من المتغيرات المستقلة، فيقترح أحياناً استخدام مقياس معدل يأخذ في الاعتبار عدد المتغيرات المستقلة في النموذج . ومعامل التحديد المتعدد المعدل ويرمزله ب R_a^2 ، يعدل R^2 بتقسيم كل مجموع مربعات على عدد درجاته من الحرية وهكذا نجد :

$$R_a^2 = 1 - \frac{\frac{SSE}{n-p}}{\frac{SSTO}{n-1}} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SSTO} \qquad (56.4)$$

ويمكن أن يصبح معامل التحديد المتعدد هذا أصغر عند إدخال متغير مستقل آخر إلى النموذج، لأن النقص في SSE يمكن أن يكون أكثر من أن يعوض عن نقص درجة حرية في االمقام $n - p$.

- **معامل الارتباط المتعدد :**

معامل الارتباط المتعدد R هو الجذر التربيعي الموجب لـ R^2

$$R = \sqrt{R^2} \qquad (57.4)$$

وهويساوي في القيمة المطلقة معامل الارتباط rلارتباط بسيط عندما يكون

$P - 1 = 1$ ، أي عندما يوجد متغير مستقل واحد في نموذج الانحدار (37.4)

6.2.4 استدلالات حول وسطاء الانحدار

مقدرات المربعات الصغرى في $\boldsymbol{b}$ غير منحازة :

$$E\{\boldsymbol{b}\} = \boldsymbol{\beta} \qquad (58.4)$$

ومصفوفة التغاير $\boldsymbol{var}\{\boldsymbol{b}\}$:

$$\boldsymbol{\sigma}^2_{p\times p}\{\boldsymbol{b}\} = \begin{bmatrix} \sigma^2\{b_0\} & \sigma\{b_0,b_1\} & \dots & \sigma\{b_0,b_{p-1}\} \\ \sigma\{b_0,b_1\} & \sigma^2\{b_0\} & \dots & \sigma\{b_0,b_{p-1}\} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \sigma\{b_{p-1},b_0\} & \sigma\{b_{p-1},b_0\} & \dots & \sigma^2\{b_{p-1}\} \end{bmatrix} \qquad (59.4)$$

معطاة بالعلاقة:

$$\boldsymbol{var}_{p\times p}\{\boldsymbol{b}\} = \boldsymbol{\sigma}^2(\mathbf{X}'\boldsymbol{X})^{-1} \qquad (60.4)$$

ومصفوفة التغاير المقدرة $\boldsymbol{s}^2\{\boldsymbol{b}\}$:

$$\boldsymbol{s}^2_{p\times p}\{\boldsymbol{b}\} = \begin{bmatrix} s^2\{b_0\} & s\{b_0,b_1\} & \dots & s\{b_0,b_{p-1}\} \\ s\{b_0,b_1\} & s^2\{b_0\} & \dots & s\{b_0,b_{p-1}\} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ s\{b_{p-1},b_0\} & s\{b_{p-1},b_0\} & \dots & s^2\{b_{p-1}\} \end{bmatrix} \qquad (61.4)$$

معطاة بالعلاقة:

$$s^2_{p\times p}\{\boldsymbol{b}\} = MSE(\mathbf{X}'\boldsymbol{X})^{-1} \qquad (62.4)$$

ويمكن أن نحصل من $\boldsymbol{s^2}\{\boldsymbol{b}\}$ على $s^2\{b_0\}$ و $s^2\{b_0\}$ أو أي تباين آخر نحتاجه، أو أية تغايرات نحتاجها.

• التقدير بفترة لـβ_k

في نموذج الانحدار (37.4) ذي الخطأ الطبيعي لدينا :

$$\frac{b_k - \beta_k}{s\{b_k\}} \sim t(n-p) \qquad k = 0,1,\cdots,p-1 \qquad (63.4)$$

وبالتالي فإن حدي الثقة لـ β_k بمعامل ثقة $1-\alpha$ هما :

$$b_k \mp t(1-\alpha/2\,;n-p)s\{b_k\} \qquad (64.4)$$

• اختبارات β_k

تجري اختبارات β_k بالطريقة المعتادة. فالاختبار :

$$H_0: \beta_k = 0$$

$$H_a: \beta_k \neq 0 \qquad (65a.4)$$

يمكن استخدام إحصاءة الاختبار :

$$t^* = \frac{b_k}{s\{b_k\}} \qquad (65b.4)$$

وقاعدة القرار هي:

إذا كان $|t^*| \leq t(1-\alpha\,/\,2;\mathrm{n}-\mathrm{P})$ نقبل H_0 ونرفض H_a

فيما عدا ذلك نقبل H_a ونرفض H_0 $\qquad (65c.4)$

7.2.4 استدلالات حول متوسط الاستجابة

• اختبار F حول نقص التوفيق:

لاختبار ما إذا كانت دالة استجابة الانحدار المتعدد :

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1}$$

تمثل سطح استجابة مناسب لمجموعة من البيانات فإننا نحتاج، إلى مشاهدات مكررة، والمشاهدات المكررة في الانحدار المتعدد هي مشاهدات متكررة لقيمة Y المقابلة لمستويات كل من المتغيرات X التي تبقى ثابتة من تكرار لآخر. وهكذا فإن المشاهدات المكررة في حالة متغيرين مستقلين يتطلب بقاء كل من X_1, X_2 عند مستو ثابت من مشاهدة لقيمة Y إلى مشاهدة أخرى. الإجراءات التي ذكرناها في الانحدار البسيط المتعلقة باختبار F حول نقص التوفيق هي اجراءات قابلة للتطبيق في الانحدار المتعدد وحالما نحصل على جدول تحليل التباين، نحلل SSE إلى مركبتي خطأ بحث ونقص توفيق، ونحصل على مجموع مربعات الخطأ البحت $SSPE$ بأن نحسب أولاً، ولكل زمرة من المشاهدات المكررة مجموع مربعات انحرافات المشاهدات Y عن متوسط الزمرة، حيث تبقى قيم المتغيرات X نفسها في كل من زمر التكرارات. فلنفرض وجود c من زمر التكرارات. بمجموعات متميزة من مستويات المتغيرات X، ولنرمز لمتوسط المشاهدات Y في الزمرة j بـ $\bar{Y}_j$. فمجموع المربعات للزمرة j يعطى بالعلاقة $\sum_i \left(Y_{ij} - \bar{Y}_j\right)^2$ ، ومجموع مربعات الخطأ البحت هو مجموع هذه المجاميع من المربعات كما هومعطى في

$SSE(F) = \sum_j \sum_i (Y_{ij} - \bar{Y}_j)^2 = SSPE$. ومجموع مربعات نقص التوفيق $SSLF$ يساوي الفرق :

$SSLF = SSE - SSPE$. وعدد درجات الحرية المرافق لـ $SSPE$ هو $(n-c)$، وعدد درجات الحرية المرافق لـ$SSLF$ هو $(n-P)-(n-c) = c-P$.

ويجري الاختبار F كما وصفنا سابقاً في الانحدار البسيط، ولكن بدرجات من الحرية معدلة عن تلك المعروضة هناك. يهدف هذا الاختبار إلى التحقق مما إذا كانت دالة انحدار خطية توفيقاً جيداً للبيانات .

• الفرضيات

يفترض اختبار نقص التوفيق أن المشاهدات Y المقابلة لـ X معطاة هي :

(1) مستقلة

(2) متوزعة طبيعياً

(3) لتوزيعات الـ Y التباين $\boldsymbol{\sigma^2}$ نفسه .

وكذلك يتطلب اختبار نقص التوفيق تعدد المشاهدات عند مستوى واحد أو أكثر لـ X .

ولاختبار البدائل:

$$H_0: E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1}$$

$$H_a: E\{Y\} \neq \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1} \qquad (66a.4)$$

تكون إحصاءة الاختبار المناسبة :

$$F^* = \frac{SSLF}{c-p} \div \frac{SSPE}{n-c} = \frac{MSLF}{MSPE} \qquad (66b.4)$$

وقاعدة القرار المناسبة هي :

إذا كانت $F^* \leq F(1-\alpha; c-p, n-c)$ نقبل H_0 ونرفض H_a

إذا كانت $F^* > F(1-\alpha; c-p, n-c)$ نرفض H_0 ونقبل H_a $\qquad (66c.4)$

8.2.4 رسومات الرواسب، تشخيصات أخرى، تدابير علاجية

إن رسومات الصناديق ورسومات الزمن ورسومات الجذع والورقة، والرسومات النقطية لكل من المتغيرات المستقلة يمكن أن يقدم معلومات مساعدة وتمهيدية حول هذه المتغيرات وبصورة مماثلة فإن رسومات الانتشار للمتغير التابع في مقابل كل من المتغيرات المستقلة يمكن أن تعين في تحديد طبيعة وقوة العلاقة بين المتغير المستقل

والمتغير التابع، وفي التعرف على ثغرات في نقاط تمثيل البيانات بالإضافة إلى التعرف على نقاط البيانات القاصية. ورسومات الانتشار لكل متغير مستقل مقابل كل من المتغيرات المستقلة الأخرى هي رسومات معينة في دراسة العلاقات بين المتغيرات المستقلة وفي إيجاد ثغرات، وتحرِّي المشاهد القاصية. ورسم الرواسب في مقابل القيم التوفيقية مفيد لتثمين صلاحية دالة الانحدار وثبات تباين حدود الخطأ، بالإضافة إلى تقديم معلومات عن المشاهد القاصية. وبصورة مماثلة يمكن أن يقدم رسم الرواسب في مقابل الزمن معلومات تشخيصية حول ارتباطات ممكنة بين حدود الخطأ وتفيد رسومات Box ورسومات الاحتمال الطبيعي للرواسب في تفحص ما إذا كانت حدود الخطأ تتوزع بصورة معقولة وفق التوزيع الطبيعي وبالإضافة إلى ذلك ينبغي رسم الرواسب في مقابل كل من المتغيرات المستقلة. ويمكن أن يقدم كل من هذه الرسومات معلومات إضافية حول صلاحية دالة الانحدار بالنسبة لذلك المتغير المستقل (مثلاً ما إذا كنا نحتاج تمثيلاً منحنياً لتأثير ذلك المتغير) وحول تغيرات ممكنة في مقدار تباين الخطأ فيما يتعلق بذلك المتغير المستقل. وأخيراً ينبغي رسم الرواسب في مقابل متغيرات مستقلة مهمة حذفت من النموذج، للرؤية ما إذا كان للمتغيرات المحذوفة تأثيرات إضافية مهمة على المتغير التابع لم نتعرف عليها بعد من خلال نموذج الانحدار. وينبغي أيضاً رسم الرواسب في مقابل حدود التفاعل غير المشمولة في نموذج الانحدار مثل X_1X_2، X_1X_3 وX_2X_3 إذا كنا نحتاج في النموذج لبعض حدود التفاعل هذه أولها جميعاً. والتدابير العلاجية الموصوفة للانحدار الخطي البسيط، هي أيضاً تدابير قابلة للتطبيق في الانحدار المتعدد وإذا تطلب الأمر نموذجا أكثر تعقيداً يسلم بوجود تأثيرات منحنية أو تأثيرات تفاعل، فيمكن توسيع نموذج الانحدار المتعدد ليشمل هذه التأثيرات. وعلى سبيل المثال يمكن إضافة $X_2{}^2$ كمتغير ليأخذ في الاعتبار تأثيراً منحنياً لـX_2 أو يمكن إضافة X_1X_3 كمتغير اعترافاً بوجود تأثير تفاعل بين X_3 و X_1 على المتغير التابع. وبصورة بديلة، يمكن القيام بتحويلات على المتغير التابع و/أو المتغيرات المستقلة، متبعين في ذلك المبادئ الأساسية لعلاج أي عيوب في النموذج. ويمكن أن تكون التحوبلات في المتغير التابع مساعدة عندما تكون توزيعات حدود الخطأ ملتوية بوضوح وتباين حدود الخطأ غير ثابت. كما يمكن أن تكون التحويلات في بعض المتغيرات المستقلة مساعدة عندما تكون تأثيرات هذه المتغيرات تأثيرات منحنية. وبالإضافة إلى ذلك : يمكن أن تكون التحويلات على Y و/أو المتغيرات المستقلة مساعدة في حذف تأثيرات التفاعل أو تخفيضها تخفيضاً هائلاً. وكما في الانحدار الخطي البسيط، نحتاج طبعاً إلى الاطمئنان إلى فائدة التحويلات باستخدام رسومات الرواسب وأدوات التشخيص الأخرى، وذلك لتحديد ما إذا كان نموذج الانحدار المتعدد مناسباً للبيانات بعد التحويل.

9.2.4 الخطية المتعددة وتأثيراتها

في تحليل الانحدار المتعدد، يهتم المرء غالباً بطبيعة وأهمية العلاقة بين المتغيرات المستقلة والمتغير التابع . ومن الأسئلة التي كثيراً ما تطرح:

1-ماهي الأهمية النسبية لتأثيرات المتغيرات المستقلة المختلفة ؟

2-ما هو مقدار تأثير متغير مستقل بعينه على المتغير التابع ؟

3-هل يمكن شطب أي من المتغيرات المستقلة من النموذج لأن تأثيره على المتغير التابع هو تأثير طفيف؟

4-هل ينبغي النظر في إمكانية ضم أية متغيرات مستقلة ، لم يشملها النموذج بعد ، إلى النموذج؟

وإذا كانت المتغيرات المستقلة التي يشملها النموذج

(1)غير مرتبطة فيما بينها .

(2)غير مرتبطة مع أية متغيرات مستقلة أخرى تتصل بالمتغير التابع ولكنها ملغاة من النموذج ، فيمكن إعطاء أجوبة بسيطة نسبياً على هذه الأسئلة.ومن سوء الطالع تميل المتغيرات المستقلة في العديد من الدراسات غير التجريبية في الأعمال، الاقتصاد، والعلوم الاجتماعية والبيولوجية، إلى أن تكون مرتبطة فيما بينها ومرتبطة مع متغيرات أخرى ذات صلة بالمتغير التابع وغير مشمولة في النموذج . وعلى سبيل المثال، في انحدار نفقات الطعام لأسرة على المتغيرات المستقلة: دخل الأسرة ،توفيرات الأسرة وعمر رب الأسرة، ستكون المتغيرات المستقلة مرتبطة فيما بينها. وأكثر من ذلك، ستكون المتغيرات المستقلة مرتبطة أيضا بمتغيرات اجتماعية-اقتصادية غير مشمولة في النموذج ولها تأثيرها على نفقات طعام الأسرة، مثل حجم الأسرة.

وعندما تكون المتغيرات المستقلة مرتبطة فيما بينها يقال أنه يوجد ارتباط داخلي أو خطية متعددة فيما بينها .(وأحياناً نحتفظ بالمصطلح الأخير لتلك الحالات التي يكون الارتباط فيها، بين المتغيرات المستقلة، عالياً جداً).

• تأثيرات الخطية المتعددة

في الواقع العملي ،نادراً ما نجد متغيرات مستقلة على علاقة تامة ببعضها ،أو بيانات لا تتضمن مركبة خطأ عشوائي

1- بصورة عامة، لاتكبح حقيقة أن تكون بعض المتغيرات المستقلة أو جميعها مرتبطة فيما بينها، قدرتنا على الحصول على توفيق جيد. ولا تنزع إلى التأثير في استقراءات حول متوسط الاستجابة، أو تنبؤات بمشاهدات جديدة، شريطة أن تتم هذه الاستقراءات ضمن منطقة المشاهدات .

2-العديد من دوال الانحدار المختلفة التي تقدم توفيقات من البيانات على مستوى جيد من الجودة، يقابلها في الحياة العملية أن وسطاء الانحدار المقدرة تميل إلى يكون لها تشتت معاينة كبير، عندما تكون المتغيرات المستقلة على درجة عالية من الارتباط. وهكذا تميل معاملات الانحدار المقدرة إلى أن تتغير تغيراً واسعاً من عينة إلى عينة، وذلك عندما تكون المتغيرات المستقلة عالية الارتباط. وكنتيجة لذلك، لايمكن أن تتوافر لنا إلا معلومات غير دقيقة عن حقيقة وسطاء الانحدار المقدَّرة بمفردها غير مهمة إحصائياً مع أن هناك بالتأكيد علاقة إحصائية قائمة بين المتغير التابع ومجموعة المتغيرات المستقلة.

3- التفسير الشائع لوسطاء الانحدار كقياس للتغير في القيمة المتوقعة للمتغير التابع عندما يزداد المتغير المستقل المقابل بوحدة واحدة، مع بقاء جميع المتغيرت المستقلة الأخرى ثابتة، يصبح تفسيراً غير قابل للتطبيق تماما عند تواجد الخطية المتعددة. فبينما قد يكون من الممكن نظرياً تغيير متغير مستقل واحد وإبقاء المتغيرات الأخرى ثابتة، إلا أنه قد لا يكون ذلك ممكناً عملياً في حالات متغيرات مستقلة عالية الارتباط .

الفصل الخامس:

اختبار النتائج

1.5 بناء النموذج

نهتم بدراستنا فيما إذا كان جودة الإشارة القلبية ECG المرسومة مرتبطاً بـ err (خطأ التصنيف و(Rn)مجموع النقاط المتطرفة $\grave{Q}_\iota + \grave{S}_\iota + \grave{R}_\iota$)، وقد تم قياس المتغيرات:

1) متغير الاستجابة $quality$ (المتغير التابع): قيمة جودة الإشارة القلبية، وهو عبارة عن متغير مستمر يأخذ قيمه في المجال [0,1]

2) المتغير التنبؤي Rn (متغير مستقل) : ويمثل مجموع عدد النقاط المتطرفة

$$Rn = \grave{Q}_\iota + \grave{S}_\iota + \grave{R}_\iota \quad \text{،} \quad i = 1, \cdots, 48$$

3) المتغير التنبؤيerr (متغير مستقل) : وهو الخطأ المرتكب بالتصنيف

$$\text{خطأ التصنيف} = 1\text{- الثقة} = 1 - \frac{\text{عدد النبضات من الخوارزمية } N_a}{\text{عدد النبضات الحقيقي } C}$$

إن المتغير $quality$ يمكن لخبير وضع قيمه بالنظر للإشارة، والأخذ بعين الاعتبار مدى الضجيج وتأثيره عليها، وسنعرض فيما يلي طريقة تجريبية لحساب قيمة تكون أقرب مايكون لواقع الإشارة المشاهَد .

سنقسم قيمة متغيرجودة الإشارة $quality$ إلى قسمين، يأخذ كل قسم 0.50 من قيمة المتغير:

- الـ 0.50 الأولى تحسب كالتالي :

إن خوارزميات الكشف عن وسطاء الإشارة يتأثر أداؤها بالتشويش، وهذا موضح في الفصل الثاني، ولذا سنحسب أولاً الخطأ المتوفر لدينا ضمن الخوارزمية، وهو $\frac{\text{مجموع النقاط المتطرفة}}{\text{عدد النبضات من الخوارزمية}}$ ، ويمثل أكبر قيمة ممكنة للخطأ المحسوب (لأن كل نقطة شاذة هي على الأكثر نبضة شاذة) ،ثم نقوم بحساب الثقة المرافقة له، وهي(1- الخطأ) وسنسميها acc_a

$$acc_a = 1 - \frac{Rn}{N_a}$$

ثم نقوم بجعل هذا المقدار نسبة من 50% ، أي :

$(1) \quad acc_a \times \frac{50}{100}$ ،

- الـ 0.50 الثانية تُحسب كالتالي :

ننظر إلى القيمة السابقة (1) فإذا كانت أقل من 0.47 لا نضيف للقسم الأول أي قيمة أخرى ،أي نعتبر أن هذا الجزء يساوي صفر ونعتمد (1)كقيم دالة على جودة الإشارة الموافقة .

إذا كان (1)أكبر أو يساوي 0.47 سننظر إلى الإشارة المرسومة الموافقة ،ونعطيها نسبة للجودة (النقاء أو خلوها من الضجيج) من $\frac{50}{100}$ ونضيفه للمقدار الأول (1)، والمجموع هو نسبة من مئة تعبر عن قيمة متغير جودة هذه الإشارة $quality$.

ملاحظة:تم استثناء 5 إشارات من المرحلة التجريبية، وإعطائها قيم للجودة قريبة من الصفر وسنوضح فيما يأتي هذه الإشارات والأسباب، وسنذكر الملاحظة المتعلقة بالإشارة الواردة في MIT-BIH

حرفياً، والتي على أساسها تم **الاستثناء** مع بعض التوضيحات.

ملاحظة: تم اختيار القيمة 0.47 بالتجريب، وذلك عن طريق المقارنة مع قيم جودة للإشارات وضعت بشكل مستقل.

1) **الإشارة 104:**

"هناك حدوث لعدة فترات ضجيج عالي التردد في الإشارة سببها ضجيج ناجم من العضلات، ولكن الإشارة جيدة "

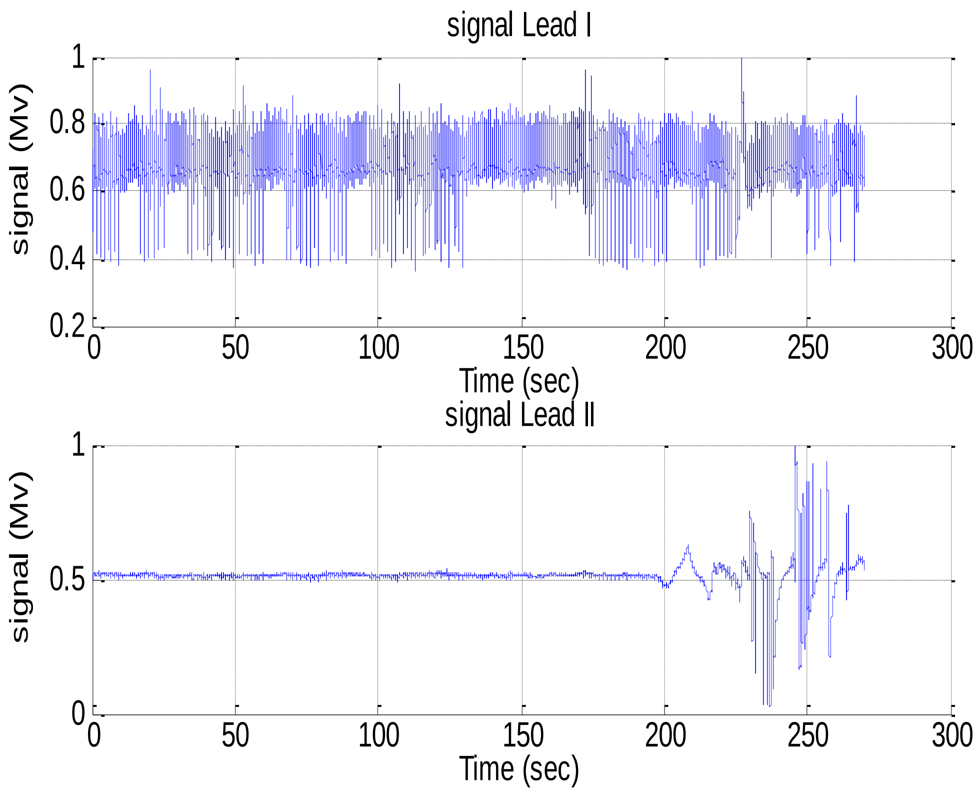
أي أن هناك سبب واضح(العضلات) للضجيج الموجود في الإشارة، وليست متغيرات الدراسة

وتم اختيار هذه الإشارة بسبب وجود إشارة مماثلة كانت العضلات سبب للضجيج فيها ،ولكن بقيت الإشارة بحالة ممتازة .

2) **الإشارة 200:**

"يوجد فترات لضجيج عال التردد في التوصيلة العلوية، وضجيج كبير وأخطاء يدوية في التوصيلة السفلية".

هذه الإشارة موضحة في الشكل (1.5)



.

الشكل (1.5)

3) **الإشارة 203:**

"هناك تغيرات في للمعقدQRS في التوصيلة الثانية بسبب إزاحات للمحاور. هناك ضجيج ملحوظ في كلا التوصيلتين، يتضمن الأخطاء الناجمة عن العضلات، وإنزياحات للخط الأساسي. إنه تسجيل صعب جداً، حتى على البشر!"

4) **الإشارة 207:**

"هذا تسجيل صعب للغاية أو لأقصى حد"

5) **الإشارة 228:**

"هناك حدوث لثلاث إفلاتات (إنزلاقات) لشريط التسجيل بمدة أقصاها 2.2 ثانية"

باستخدام برنامج **SPSS** سنقوم بإثبات أن للمتغيرين المستقلين التوزيع الطبيعي، وسيكون عندها المتغير التابع هو عبارة عن مجموع متغيرين طبيعيين، فهو إذاً متغيرطبيعي.

نتحقق من أن للمتغيرين err و Rn التوزيع الطبيعي:

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Rn is normal with mean 12.000 and standard deviation 10.68.	One-Sample Kolmogorov-Smirnov Test	.386	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

الخرج(1.5)

إن هذا الخرج هونتيجة لاختبار كولموغورف – سميرنوف للعينة الواحدة

لدراسة جودة التطابق مع التوزيع الطبيعي وتكون الفرضيات:

فرضية العدم H_0: تتوزع قيم المتغير Rn وفق التوزيع الطبيعي.

الفرضية البديلة H_1: لا تتوزع قيم المتغير Rn وفق التوزيع الطبيعي .

حيث sig أو (p-value)مستوى المعنوية المشاهد، نقوم بالمقارنة بينه وبين مستوى الأهمية

$\alpha = 0.05$ فنجد :

$\alpha < 0.386$ إذاً نقبل فرضية العدم H_0 ونرفض الفرضية البديلة H_1
عند مستوى المعنوية $\alpha = 0.05$ يكون للمتغير Rn التوزيع الطبيعي
$Rn \sim N(12, (10.68)^2)$

(د.عدنان عمورة.2010.الإحصاء(4)

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of err is normal with mean 0.224 and standard deviation 0.25.	One-Sample Kolmogorov-Smirnov Test	.006	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

الخرج(2.5)

عند مستوى المعنوية $\alpha = 0.001$ يكون للمتغير err التوزيع الطبيعي أي $err \sim N(0.224,(0.25)^2)$

وبذلك نكون قد تحققنا من أن جميع المتغيرات التي ستستخدم لبناء النموذج لها التوزيع الطبيعي .

2.5 النتائج ومناقشتها

الآن من القائمة analyze نختار:

analyze→ regression→ linear

ونقوم باختيار الاختبارات والرسوم البيانية المناسبة، فتكون لدينا النتائج التالية :

Descriptive Statistics الخرج (3.5)

	Mean	Std. Deviation	N
quality	.51	.20	48
Rn	12.00	10.681	48
err	.22	.2496	48

يبين الخرج (3.5) **Descriptive Statistics** (الإحصاء الوصفي) قيم المتوسط والانحراف المعياري للمتغيرات الثلاثة.

Variables Entered/Removed[a] الخرج(4.5)

Model	Variables Entered	Variables Removed	Method
1	err, Rn[b]	.	Enter

a. Dependent Variable: quality
b. All requested variables entered.

يظهر الخرج(4.5) **Variables Entered/Removed[a]** المتغيرات المستقلة التي يحويها النموذج err, Rn وطريقة إدخال البيانات للنموذج وهي(Enter)

Model Summary[b] الخرج (5.5)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.886[a]	.785	.775	.0958

a. Predictors: (Constant), err, Rn
b. Dependent Variable: quality

Model Summary[b]

Model	Change Statistics					Durbin-Watson
	R Square Change	F Change	df1	df2	Sig. F Change	
1	.785	81.911	2	45	.000	2.437

b. Dependent Variable: quality

يظهرالخرج (5.5) **Model Summary** وهو يمثل ملخص عن النموذج، النتائج التالية:
اختبار F للانحدار الخطي،إن $\text{sig}=.000<\alpha$ ولذا نفترض أن النموذج المفروض يلائم البيانات.

ونجد أيضاً أن قيمة معامل التحديد المتعدد $R^2=0.785$ أي أن المتغيرات المستقلة

(err و Rn) استطاعت أن تفسر 79% من التغيرات الحاصلة في المتغير التابع $quality$

وقيمة معامل التحديد المعدّل (Adjusted R Square) هي 0.775

ننتقل للإحصاءة D لاختبار دوربن- واتسون وقيمتها 2.437

عند مستوى المعنوية لعدد $p-1=2$ من المتغيرات المستقلة حيث $n=50$، ومن جدول حدي اختبار دوربن-واتسون(1.1) نجد :

$$d_U = 1.63 \quad \text{و} \quad d_L = 1.46$$

$D > d_U$ ومنه نقبل H_0، وبذلك تكون حدود الخطأ ε_i مستقلة. $i = 1, \cdots, 48$

الخرج(6.5)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.502	2	.751	81.911	.000[b]
	Residual	.413	45	.009		
	Total	1.914	47			

a. Dependent Variable: quality
b. Predictors: (Constant), err, Rn

إن الخرج(6.5) هو عبارة عن جدول لقيم مجاميع المربعات اللازمة لاختبار F، وهي :
{ SSR و $SSTO$ و SSE }

إن $\alpha >$ sig= .000 ولذا نفترض وجود علاقة خطية بين المتغيرات في نموذجنا.

أي نقبل بوجود علاقة خطية بين المتغير التابع quality والمتغيرات التنبؤية err, Rn

Coefficients[a] **الخرج (7.5)**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.776	.028		27.402	.000
	Rn	-.008	.001	-.418	-5.670	.000
	err	-.759	.060	-.938	-12.726	.000

a. Dependent Variable: quality

Coefficients[a]

Model		95.0% Confidence Interval for B		Collinearity Statistics	
		Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	.719	.833		
	Rn	-.011	-.005	.881	1.135
	err	-.879	-.638	.881	1.135

a. Dependent Variable: quality

يبين الخرج (7.5) **Coefficients[a]** تقديرات الانحدار الخطي المتعدد للوسطاء، متضمنة حد التقاطع ومستوى الأهمية.

من النتائج نجد أن $sig = .000$ لجميع وسطاء النموذج وهي أصغر من مستوى الأهمية $\alpha = 0.05$

أي أن المتغيرات هي متنبئات معنوية ويبين الجدول أيضاً فترات الثقة الخاصة بكل معامل.

ويحوي هذا الجدول بالإضافة إلى ماسبق ، إحصاءات متعلقة بالخطية، وهي قيمة السماحية (Tolerance) ومعامل تضخم التباين VIF(variance inflation factor)

نلاحظ أن قيم السماحية للمتغيرين التنبؤيين أكبر من 0.2 ، حيث تكون قيم التسامح الأصغر من 0.2 دالة بشكل أكيد على مشكلة جدية في الارتباط الخطي (Menard 1995)

ويقترح (MYER(1990 أن قيم VIF التي تكون أكبر من 10 هي سبب للاهتمام، ومن أجل بياناتنا هنا ، فإن القيم VIF للمتغيرين التنبؤيين هي أقل من 10.

Collinearity Diagnostics[a] **الخرج (8.5)**

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	Rn	err
1	1	2.188	1.000	.05	.05	.06
	2	.667	1.811	.00	.25	.41
	3	.145	3.889	.95	.70	.53

a. Dependent Variable: quality

إن الخرج **(8.5) Collinearity Diagnostics[a]** هوجدول لتشخيص الخطية،

(Kutner ,M. H.,Nachtsheim,C.J.,Neter , J., Li ,W.(2005))

ويحوي الجدول القيم الخاصة لمصفوفة الجداء المتصالب اللامركزي، دليل الحالة Condition Index و نسب التباين Variance Proportions لكل متغير تنبؤي.

إذا كانت أي من القيم الخاصة في هذا الجدول أكبر من بقية القيم فإن مصفوفة الجداء المتصالب اللامركزي تكون ill-conditioned وهذا يعني أن وسطاء الانحدار يمكن أن تتأثر بشكل كبير بالتغيرات الصغيرة في المتغيرات التنبؤية.

أي بمعنى آخر تعطينا فكرة عن مدى دقة نموذج الانحدار، فإذا كانت القيم الذاتية متشابهة، فإن النموذج لن يتأثر بالتغيرات الصغيرة في المتغيرات المقاسة.

أما أدلة الحالة فهي طريقة أخرى للتعبيرعن القيم الذاتية، وتساوي الجذر التربيعي لنسبة القيمة الذاتية الكبرى إلى القيمة الذاتية المدروسة. ومن الجدول نجد أن لبياناتنا أدلة حالة متقاربة، وبالتالي لاوجود لمشاكل الارتباط الخطي. تفيدنا نسب التباين في معرفة نسبة التباين لكل وسيط انحدار للمتغير التنبؤي. وفيما يتعلق بالارتباط الخطي، فإننا نبحث عن المتغيرات التنبؤية ذات النسب المرتفعة من أجل نفس القيمة الذاتية الصغيرة، لأن ذلك يشير إلى أن التباينات لوسطاء الانحدار الخاصة بها هي مستقلة. ولذلك فإننا نهتم بشكل أساسي بالأسطر القليلة الأخيرة في الجدول (التي تمثل القيم الذاتية الصغيرة).

وبالنسبة لدراستنا نجد أن 70%من التباين في وسيط الانحدارللمتغير Rn مرتبط بالقيمة الذاتية رقم 3 (القيمة الذاتية الأصغر)، كما أن 53%من التباين في وسيط الانحدار للمتغير err مرتبط بالقيمة الذاتية رقم 3 ، مما يدل على عدم التبعية بين هذه المتغيرات، والنتيجة هي أن لايوجد ارتباط خطي بين متغيري الدراسة.

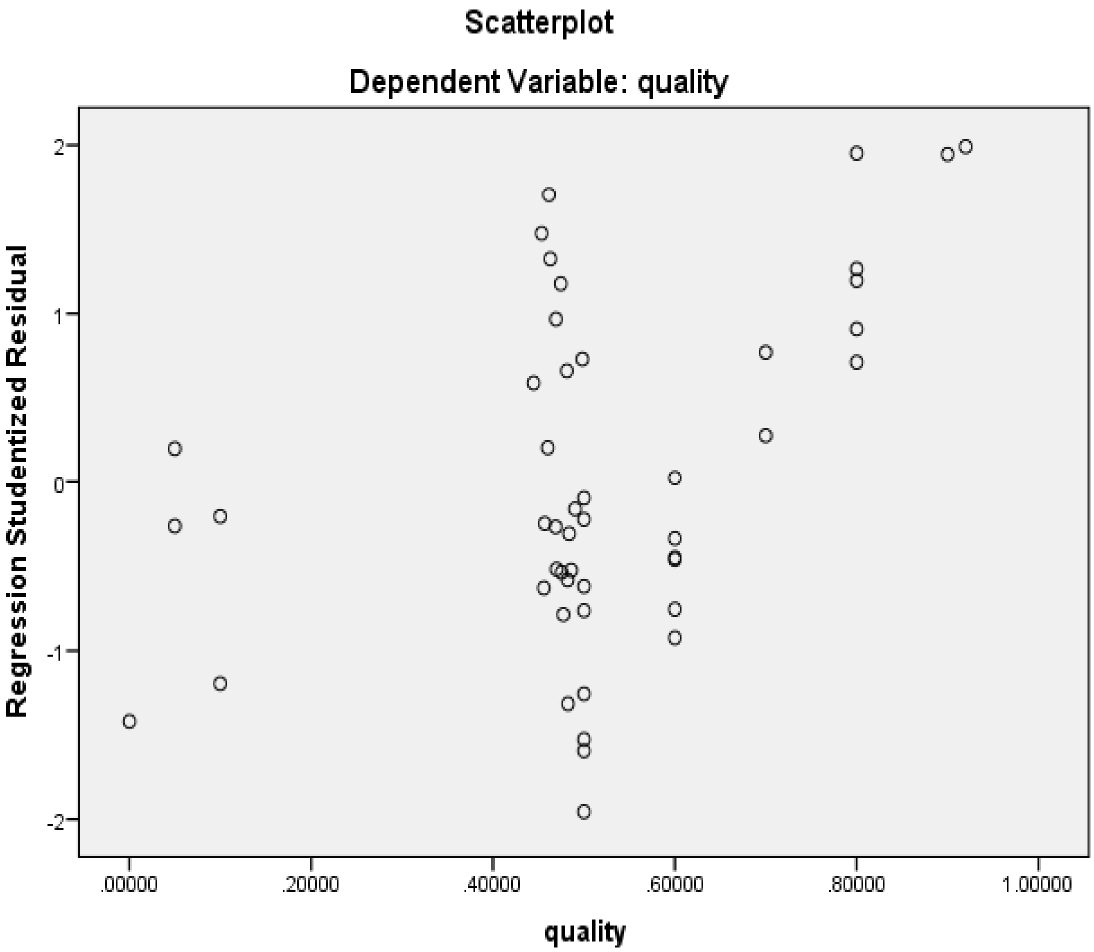
Correlations (9.5) الخرج

		quality	Rn	err
Pearson Correlation	quality	1.000	-.095	-.794
	Rn	-.095	1.000	-.345
	err	-.794	-.345	1.000
Sig. (1-tailed)	quality	.	.261	.000
	Rn	.261	.	.008
	err	.000	.008	.
N	quality	48	48	48
	Rn	48	48	48
	err	48	48	48

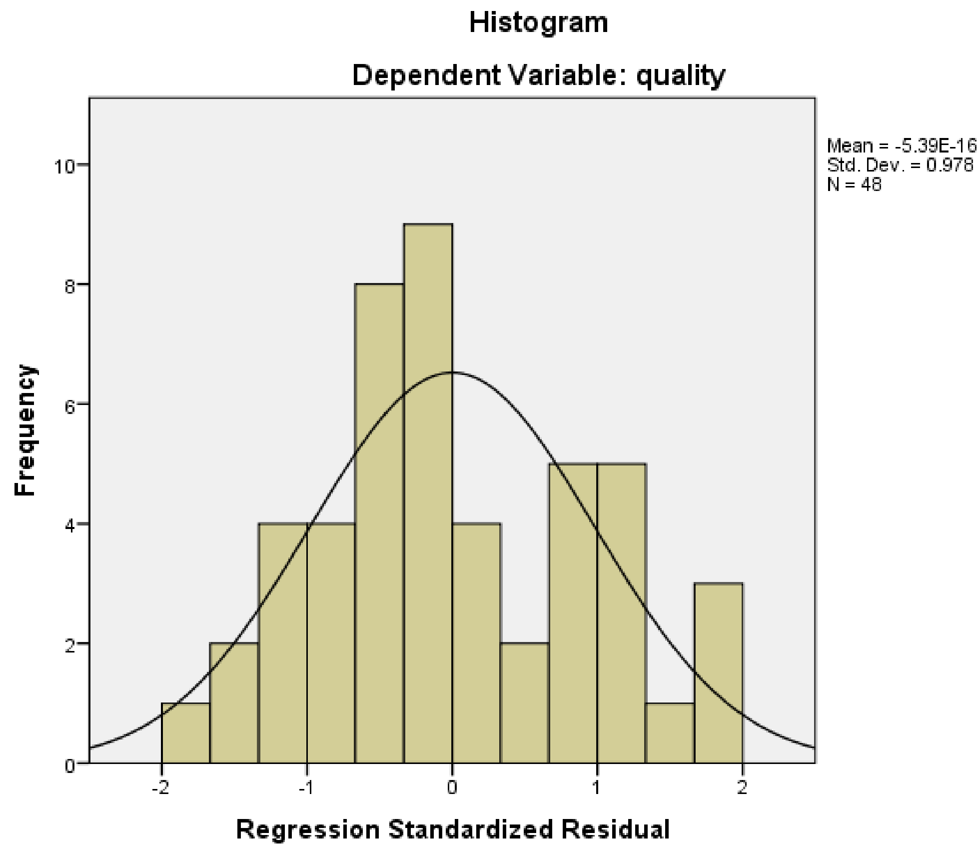
يبين الخرج (9.5) قيم معامل بيرسون للارتباط بين جميع المتغيرات في تحليل الانحدار، يمكن استنتاج أنه لاارتباط خطي معنوي بين المتغيرين المستقلين حيث أن $r = -0.35$.

• دراسة الرواسب

ومن أجل التحقق من عشوائية اختيار العينات، نأخذعينة أخرى مختلفة عن تلك التي استخدمناها لبناء النموذج وتكون العينة الثانية عبارة عن 100000 تردد عيني للإشارة مختلف عن السابق، حيث نطبق الخوارزمية عليها فتعطينا قيم جديدة للمتغيرات err و Rn ثم باستخدام دالة الاستجابة المدروسة، نحصل على تنبؤ لقيم متغير الاستجابة $\widehat{Y}_i$ (جودة الإشارة) ومنه نحسب الرواسب: $e_i = Y_i - \widehat{Y}_i$.



الشكل (2.5) :المخطط التبعثري لرواسب الانحدار المعيارية مقابل المتغير التابع



الشكل (3.5): المدرج التكراري للرواسب

الخرج (10.5)

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of ei is normal with mean -0.036 and standard deviation 0.15.	One-Sample Kolmogorov-Smirnov Test	.734	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

قمنا بإجراء اختبار كولموغوروف-سميرنوف للرواسب e_i، ويبين الخرج (10.5) أن $sig = 0.734 > \alpha$ ، أي للرواسب التوزيع الطبيعي عند مستوى المعنوية $\alpha = 0.05$.

3.5 الاستنتاجات:

1- عند تطبيق خوارزمية (G.Karraz,G.Magenes) على قاعدة البيانات العالمية -MIT BIH ورسم الإشارات المتوفرة ،حصلنا على عدة متغيرات مهمة ،تم استنباطها اعتماداً على السمات الواسمة للإشارة في الحيز الزمني.

2- اعتماد المتغيرين err و Rn كمتغيري تنبؤ معنويين لبناء هذا النموذج للأسباب التالية:

أ) إن قيمة معامل التحديد $R^2 = 0.785$ أي استطعنا تفسير 79% من أسباب التغيرالحاصل في جودة الإشارة $quality$ (كمية الضجيج المتراكب عليها) بواسطة المتغيرين err و Rn ، أي أن21% من سبب التغير في جودة الإشارة لازال غير معروف (ويعزى ذلك إلى أسباب أخرى كالتيار المتناوب، حركة المريض، العضلات، الجهاز التنفسي ....).

ب) إن اختبار F لنقص التوفيق، هولاختبار ما إذا كانت دالة استجابة الانحدار المتعدد تمثل سطح استجابة مناسب لمجموعة من البيانات، إن قيمة هذه الإحصاءة كانت 81.9، ويهدف هذا الاختبار إلى التحقق مما إذا كانت دالة انحدار خطية توفيقاً جيداً للبيانات. أي أن نموذجنا يتوقع بيانات العالم الحقيقي بشكل جيد .

ج) تشير قيم الدلالة للاستقراءات حول وسطاء الانحدار أن المتغيرات err و Rnمعنوية، وكل من قيم الوسطاء تنتمي إلى مجال الثقة الخاص بها.

كما توصلنا إلى أنه لا وجود لحد تفاعل أو حد تربيعي في نموذج الانحدار، وأنه يحقق أفضل معاييرفرضيات الانحدار الخطي المتعدد بصيغته هذه وذلك بعد تجربة عدة تحويلات للمتغيرات المستقلة وللمتغير التابع، وبالتالي يكون النموذج :

$$Y_i = 0.776 - 0.008\, Rn_i - 0.759\, err_i + \varepsilon_i$$

وبافتراض $\mathrm{E}\{\varepsilon_i\} = 0$ تكون دالة انحدار النموذج هي :

$$E\{Y\} = 0.776 - 0.008\, Rn - 0.759\, err$$

4.5 المقترحات والتوصيات

نقترح **صلاحية** النموذج النهائي للتعميم لتوفر الشروط:

1) عدم وجود خطية متعددة ، حيث قيمة إحصاءة التسامح (Tolerance) لكل من متغيري التنبؤ أكبر من 0.2 ،كما أن قيمةVIF لكل من المتغيرين أيضاً هي أصغر من 10 . وإن 70%من التباين في وسيط الانحدار للمتغير Rnمرتبط بالقيمة الذاتية رقم 3 (القيمة الذاتية الأصغر)،كما أن 53%من التباين في وسيط الانحدار للمتغير err مرتبط بالقيمة الذاتية رقم 3 ، مما يدل على عدم التبعية بين هذه المتغيرات.

2) تم حساب $\widehat{Y_i}$من عينة مختلفة للتأكد من عشوائية اختيار العينات في بناء النموذج، وبعد حساب الرواسب e_i وجدنا أنها تتوزع وفق التوزيع الطبيعي $e_i \sim N(-0.036, (0.15)^2)$.

3) تساعد القيم المرتفعة لجودة الإشارة $quality$ التي نتنبأ بها باستخدام النموذج في الحصول على تشخيص موثوق معتمدين على ECG، بينما تدل القيم المنخفضة على احتمال تشخيصات غير دقيقة باستخدام جهاز التخطيط ECG بسبب الضجيج المتراكب على الإشارة، مما يستدعي اللجوء إلى إجراءات طبية أخرى تساهم في إعطاء تشخيص صحيح .

الملحقات

1)الجداول الإحصائية

(1.1)حدا اختبار دوربن-واتسون

2)المصطلحات العلمية

3)المراجع العلمية

جدول (1.1) : حدا اختبار دوربن-واتسون

مستوى الأهمية $\alpha = 0.05$

n	p-1=1		p-1=2		p-1=3		p-1=4		p-1=5	
	d_L	d_u	d_L	d_u	d_L	d_u	d_L	d_u	d_L	d_u
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

المصطلحات العلمية : انكليزي-عربي

-A-

Accuracy	الثقة
Alternative Hypothesis	الفرض البديل
Analysis Of Variance	تحليل التباين
Biostatistic	الإحصاء الحيوي
Arrythmia	اضطراب النظم القلبي
Artifacts	الأخطاء اليدوية
Associated Level Of Significance (P.V)	مستوى المعنوية المشاهد
Automatic Analysis	التحليل الأوتوماتيكي

-B-

Backward Propogation	الانتشار العكسي

Baseline	الخط القاعدي

-C-

Central Limit Theorem	مبرهنة النهاية المركزية
Continuous Variable	متغير مستمر

-D-

Data	بيانات (معطيات)
Dependant Variable	متغير تابع
Descriptive Statistic	الإحصاء الوصفي
Drift	الانحراف

-E-

Effect	تأثير

Elementary Hypothesis	فرضية ابتدائية

-F-

Factor	عامل
Factor Analysis	تحليل العوامل
Filter	المرشَّح
Frequency Domain	الحيّز الترددي
Frequency Sampling	تردد عيني

-H-

Histogram	مدرج تكراري
Homogeneous	متجانس
Hypothesis testing	اختبار الفرضيات

-I-

Independent Variable	متغير مستقل
Interference	تداخل
Interval Estimation	تقدير مجالي

-K-

Kolmogorov-Smirnov Test	اختبار كولموغوروف-سميرنوف

-L-

Large Sample	عينة كبيرة
Least Squares Method	طريقة المربعات الصغرى
Linear Correlation	ارتباط خطي
Linear Regression	انحدار خطي

-M-

Mean	متوسط
Mean Squares Error	متوسط مربعات الخطأ
Maximum Likelihood Estimation	تقدير المعقولية العظمى

-N-

Noise	الضجيج (التشويش)
Non Parametric Methods	طرق لاوسيطية
Null Hypothesis	فرضية العدم

-O-

Observed Value	القيمة المشاهدة
Over Fitting	فرط الملاءمة

-P-

Parameter	وسيط المجتمع
Pearson Coefficient	معامل بيرسون

-Q-

QRS Detection	الكشف عن المعقد (QRS)

-R-

Residuals	الرواسب
Randomization	العشوائية
Random Error	خطأ عشوائي
Random Sample	عينة عشوائية
Random Variable	متغير عشوائي
Ratio	نسبة

-S-

Sample	عينة
Sample Statistic	إحصاء العينة
Sample Design	تصميم العينة
Sampling	معاينة
Sample Size	حجم العينة
Sensitivity	الحساسية
Signal Characteristics	السمات الواسمة للإشارة
Signal Features	معالم الإشارة
Specificity	النوعية
Standard Deviation	انحراف معياري
Standard Error	خطأ معياري

Statistical Inference	استدلال إحصائي
Statistical Hypothesis	الفرضية الإحصائية
Statistical Methods	طرق إحصائية

-T-

Table	جدول
Test	اختبار
T-distribution	التوزيع-t

-V-

Variable	متغير
Variance	تباين
Variance Analysis	تحليل التباين

-W-

Wave	موجة
Weights	أوزان

المراجع العربية

[1] جون نتر، وليام وازرمان، ميخائيل كتنر(1995): نماذج خطية إحصائية تطبيقية: الجزء الأول (الانحدار).

[2] د.عدنان عمورة (2010). الإحصاء(4). قسم الإحصاء الرياضي. كلية العلوم. جامعة دمشق

[3] د.محمد جاسم المحمد (2011). تحليل البيانات. قسم علم النبات. كلية العلوم. جامعة دمشق.

[4] د.محمد صبح، د.عدنان عمورة، د.عزات قاسم (2001). نظرية الاحتمالات. قسم الإحصاء الرياضي. كلية العلوم. جامعة دمشق.

المراجع الأجنبية

[5] Antti Ruha, Sami Sallinen, Seppo Nissila.(1997), "A Real Time Microprocessor QRS Detector System with a 1ms Timing Accuracy for the Measurment of Ambulatory HRV " IEEE Transaction on biomedical engineering, Volume 44, No. 3, PP.159-167.

[6] G.Clifford,(2000)

"Matlab software"

[7] G.Karraz,Doctoral Thesis,2007,University of Pavia,Italy

[8] G.Karraz.,G.Magenes(2006)

"Automatic classification of heart beats using neural network classifier based on a Bayesian framework") Proceeding of the 28th IEEEEMBS Annual international Conference,New YorkCity,USA,FrEp8.10,pp4016-4019.

[9] J.Pan&W.J,Tompkins,

" A Real Time QRS Detection Algorithm"

[10] Kleinbaum,D.G,kupper,LL.And muller.K.E.(1988):Applied Regression Analysis and other Multivariable methods DWS-KENT publishing company, a division of wadsworth p317.

[11] Kutner ,M. H.,Nachtsheim,C.J.,Neter , J., Li ,W.(2005).

"Appplied Linear Statistical Models",5th Edition.Irwin,New York,NY,10020.

[12] Lippman,R.P.(1987),

"An introduction to computing with neural nets"

[13] MIT-BIH Arrhythmia Database Directory(Records).Available:http//www.Phsionet.com.

[14] Nelson, C.V &Geselowitz, D.B(1976):

"The theoretical basis of Electrocardiology"

[15] Rawling,J.O.,Pantula,S.G.,Dicky,D.A.(1998).

"Applied Regression Analysis",2end Edition.Raleigh,NC27695,USA.

[16] So.H. H and Chan K. L(1997)"Development of QRS Detection Method for Real Time Ambulatory Cardiac Monitor", Proc 19 Annu Int Conf IEEE EMBS,Chicago, USA,PP.289-292.

[17] Suppappola S, Sun Y. A.(1991):

"Comparison of three QRS detection Algorithm using the AHA ECG database"

المراجع العلمية للاستزادة والاطلاع

[18] D.D.dorfmann E.Alf.(1969)

"Maximum likelihood Estimation of Parameters of signal detection Theory & determination of confidence intervals rating methods data"

[19] J.A.Hanely,B.J.Mcneil(1982) .

"The meaning & use of the area under a receiver operating characteristic(ROC) curve"

[20] Kadambe,S.Murray, R & Boudreaux-Bartels,F.(1999),

"Wavelet transform-based QRS complex detector"

[21] Kay, S. M& Marple ,S. L(1981),

"Spectrum analysis- a Modern perspective".

[22] M.Pauletti,C.Marchesi(2001):"Model Based Signal Characterisation for Long Term Personal Monitoring",IEEE Computer in

Cardiology;28:pp.413-416

[23] Poli,R.,Cagnoni,S,.& Valli ,G(1995)

"Genetic design of optimum linear & nonlinear QRS detectors"

[24] RH.Clayton,SW.Lord,JM.McComb,A.Murray(1997):

"Comparison of autoreggrissiv fourier transform based techniques for estimating RR interval spectra"

[25] Rissanen. J (1984)

"Universal coding ,information prediction& estimation "

[26] Scheidt S (1983) :

"Basic Electrocardiography Leads,Axis,Arrhythmia"

[27] Scheidt , S . (1984) :

"Basic Electrocardiography Abnormalities of Electrocardiographic Patterns"

[28] Zgallai,W.(2000):

"The application of higher order statistics to non- invasive fetal electrocardiogram ECG detection"

Damascus University

Faculty of Science

Department of Mathematical

Statistics

Automatic Classification Of ECG

Thesis Submitted For Master Degree

in Mathematical Statistics

Prepared by

Ola Alzoubi

Supervised By

Dr. Izzat Kassem

Assistant Professor
Department of Mathematical Statistics
Faculty of Science
Damascus University

Dr. George Karraz

Associated Professor
Department of Artificial Intelligence
Faculty of Technology Engineering
Damascus University

2014-2015