إستخدام الانحدار الحصين لإيجاد أنسب نموذج لتمثل بيانات الأنواء الجوية في مدينة أربيل خلال الفترة (2010-1998)

م. تارا احمد حسن*　　　م.م. مهدي صابر رضا**

المستخلص:

يلوج هذا البحث في كيفية توضيح استخدام الانحدار الحصين وتطبيقاته لتقدير معلمات الانحدار في ظل وجود المشاهدات الشاذة في البيانات ، بحيث يتم الكشف عنها بطريقة – احصائية **(Difference Fit) DIFFT , Leverage** و **Box-Whisker- Plot** و ثم تطبيق الطريقتين فقط من بين الطرق الحصينة لتقدير معلمات النموذج الخطي، و هي طريقة المربعات الوسيط الصغرى **(LMS)**، و طريقة المربعات المشرذمة الصغرى **(LTS)**، مقارنة بطريقة المربعات الصغرى الاعتيادية **(OLS)** . لهذا الغرض قام الباحثان بدراسة اهم العوامل المؤثرة في الرطوبة النسبية و المتمثلة بدرجات الحرارة العظمى والصغرى ، و الضغط الجوي ، و كمية الأمطار المتساقطة، وسرعة الرياح، و اختبارها للكشف عن وجود الشواذ فيها. و اخيرا لاختيار افضل طريقة لتمثل النموذج الملائم لتلك العلاقة من بين الطرق المستخدمة، و قام الباحثان باستخدام الإحصائية متوسط مربعات الخطأ (**MSE**) للمقارنة بين القيم المقدرة للطرائق الثلاثة.

Abstract:

In this paper we shall explain how to use Robust Regression methods and their application to estimate regression parameters in order to fit the best regression model in case of existence outlier observations in the data,
The outliers were detected using statistical methods, DIFFT (Difference Fit), Leverage, and Box-Whisker-Plot tests. Then two other statistical models were used to estimate linear regression model, Least Median Squares (LMS) and Least Trimmed Squares (LTS). The results of these two modes then were compared with results obtained using Ordinary Least Squares (OLS) model in order to find the best regression model. We have used meteorological data in this research. Finally the test we used for comparing these three models is MSE statistic.

* مدرس/جامعة صلاح الدين/كلية الإدارة والإقتصاد/قسم الإحصاء

** مدرس مساعد/هيئة التعليم التقني/قسم نظم المعلومات الهندسية

مقبول للنشر بتاريخ **2011/4/18**

1- المقدمة:

حظيت مشكلة الشواذ (Outliers) في البيانات باهتمام كبير في السنوات الأخيرة بسبب إدراك الكثير من الباحثين ان استخدام الأساليب التقليدية في تقدير المعالم عند ظهور هذه المشكلة إضافة الى ان البيانات لايمكن ان تخضع بشكل كامل الى الافتراضات الموضوعية، لذا كان من الضروري البحث عن طرائق بديلة تكون حصينة لظهور الشواذ في البيانات وخاصة في بعض الحالات يفضل عدم تغيير قيم المشاهدات الشاذة بل يجب ايجاد طرائق بديلة للتقدير تكون غير حساسة و تتكيف نحو وجود الشواذ في العينة. لانه كما هو معلوم عملية تقدير المعلمات النموذج المشخص تعتبر مؤشرا للحصول على تمثيل جيد للمجتمع الذي اخذت منه العينة قيد الدراسة.

من المعروف ان عملية الحصول على افضل تقدير للمعلمات هو هدف اي اسلوب تحليلي ليمثل المجتمع تمثيلا جيدا، و ان الطرائق التقليدية في التقدير تعد طرائق كفوءة في الحصول على مقدرات جيدة، و لكن ما يعاب على الوسط الحسابي انه يتاثر بنحو كبير في حالة وجود الشواذ (Outliers) بين البيانات الماخوذة.

بالرغم من وجود طرق حصينة مختلفة الا ان اغلبها تشترك في نقطتين أساسيتين هما: هو إعطاء وزن اقل للمشاهدة التي تنحرف لتقليل من تأثيرها، و الأخرى هي استخدام اسلوب التكرار لتقليل من مشكلة الارتباط الذاتي و التعددية الخطية.

2- الهدف من البحث:

في هذا البحث نحاول تطبيق طريقة الانحدار الحصين في حالة وجود الشواذ، وتحديد مستوياته، و التحري عنه، و كيفية التعامل معه في حالة وجوده في البيانات، ذلك بالاعتماد على إحصائية (DIFFTs, Leverage) ، و الطريقة المعروفة بطريقة بوكس-ويسكر-بلوت (Box -and -Whisker Plot) و ذلك للوصول الى ما تحاول الباحثين ان تثبته وهو ان استخدام الطرائق الحصينة افضلى' لتقدير معلمات النموذج الخطي سواء كانت البيانات تتوزع توزيعا طبيعيا ام لا (في ظل وجود الشواذ في البيانات ام لا). ولهذا الغرض تمت دراسة الرطوبة النسبية والعوامل المؤثرة فيها، في هذا البحث.

3- المفاهيم الأساسية:

3-1 مفهوم الانحدار ; The Concept of Regression

ان تحليل الانحدار عبارة عن وسيلة إحصائية يستخدم لتحليل العلاقة بين متغير مستقل واحد او اكثر و المتغير التابع. و يعتبر تحليل الإنحدار من اكثر الطرق الاحصائية استعمالا في مختلف العلوم؛ لانه يصف العلاقة بين المتغيرات على هيأة معادلة. فالمعادلة التي تظم متغيرا مستقلا واحدا تسمى الانحدار الخطي البسيط و تكتب كالاتي:-

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad ....(3\text{-}1)$$

حيث ان:-

$\beta_0 = Y$ نقطة تقاطع خط الانحدار بالمحور

β_1 = معلمات غير معروفة (معاملات الانحدار)

و ان:-

Y = المتغير المعتمد

X_i = المتغير المستقل

ε_i = قيمة الخطأ

بينما المعادلة التي تحوي على اكثر من متغير مستقل واحد فتسمى بمعادلة الانحدار الخطي المتعدد و تكتب كالاتي:-

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_n X_{in} + \varepsilon_i \qquad ....(3\text{-}2)$$

$i = 1, 2, 3, \ldots, n$ و $j = 0, 1, 2, 3, \ldots, k$

k:- عدد المتغيرات المفسرة

n:- عدد المشاهدات

عند تقدير معلمات الانحدار يمكن الاستدلال عن اهمية و قوة و اتجاه العلاقة بين المتغيرات [1].

3-2 اهمية التوزيع الطبيعي للخطأ العشوائي:
The Importance of the Normal Distribution of Error Term

ان فرضية التوزيع الطبيعي للأخطاء التي تشكل اهمية كبيرة من حيث تأثيرها في خصائص المربعات الصغرى، حيث بينت نظرية كاوس ماركوف انه اذا كانت الأخطاء تتبع توزيعاً معيناً بتباين ثابت، فان مقدرات المربعات الصغرى يكون لها اقل تباين من بين كل المقدرات الخطية غير المحيزة ل β_1 .[4]

كذلك فان نظرية الغاية المركزية مهمة؛ لانها تصنع اساساً لاختبار الفرضيات؛ فقد تبين ان مجموع عدد كبير من المتغيرات المستقلة و متماثلة التوزيع، بتباين ثابت تتجه لان تكون ذات توزيع طبيعي قياسي، و تبرز اهمية هذه النظرية عند التامل فيها؛ لان الخطأ في النموذج الخطي يستخدم لتوضيح تأثير عدد كبير من المتغيرات المحذوفة من النموذج و التي لا تؤثر بصورة كافية في Y لتكون موجودة في معادلة الانحدار، فاذا كانت تتبع توزيعات معينة بتباينات ثابتة فان ادخالها للمعادلة هو الذي يجعل الخطأ يمتلك التوزيع الطبيعي.

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

او

$$\underline{Y} = X\ \underline{\beta} + \underline{\varepsilon} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(3\text{-}3)$$

تبرز اهمية التوزيع الطبيعي للأخطاء في اختيار الفرضيات التي يتم بأستخدام أختباري t, F المبنية اساساً على التوزيع الطبيعي، حيث ان الانحرافات الفعلية في التوزيع الطبيعي تؤثر في صلاحية اختبار t ، و لكنه يكون غير حساس للانحرافات الطفيفة عن الحالة الطبيعية.[8]

كما ان للتباينات الثابتة اهمية في تحقيق افضلية المربعات الصغرى في التقدير؛ لان نظرية الغاية المركزية سوف لن تحقق اذا كانت المتغيرات المحزوفة ذات تاثيرات هامة، و الأخطاء ε_i 's حينها سوف لن تتوزع طبيعيا، ومن المحتمل ان تتحيز بتباين غير ثابت ، وبهذا فانه يمكن القول: إنّ فرضيتي الكفاءة للمربعات الصغرى و كذلك قابلية تطبيق نظرية التوزيع الطبيعي لاختبار الفرضيات سوف تكون مفقودة.[15]

3-3 طريقة المربعات الصغرى الاعتيادية: (OLSM)Ordinear Least Square Method

تعتبر من الطرق الشائعة للاستخدام في تقدير معلمات الانحدار الخطي البسيط و المتعدد، وذلك لما يتميز به هذه الطريقة من خواص مرغوبة في التقدير مثل عدم التحيز و صغر التباين.
فاذا اعدنا كتابة النموذج الخطي العام بصيغة المصفوفات المعادلة رقم (3-3)، بحيث:-

$$\underline{Y} = X\ \underline{\beta} + \underline{\varepsilon}$$

$\underline{Y}$ = يمثل متجه المشاهدات المتغير المعتمد.

X = مصفوفة الثوابت لـ K من المتغيرات المستقلة.

$\underline{\beta}$ = متجه معلمات الانحدار،

$\underline{\varepsilon}$ = متجه الأخطاء العشوائية.

و من النموذج السابق يمكن ايجاد :

$$\underline{\varepsilon} = \underline{Y} - X\ \underline{\beta}$$

و ان مجموع مربعات الخطأ هو:-

$$Q = \varepsilon'\ \varepsilon = (\underline{Y} - X\ \underline{\beta})'\,(\underline{Y} - X\ \underline{\beta})$$

و بأخذ المشتقة الجزئية بالنسبة لـ β و مساواتها بالصفر نحصل على قيم β كما يلي:-

$$\hat{\beta} = (X X)^{-1} X' Y \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3\text{-}4)$$

3-4 فرضيات نموذج الانحدار الخطي:[1] Linear Regression Assumptions of Model

هنالك عدد من الافتراضات التي نعتمد عليها عند دراسة طريقة المربعات الصغرى الاعتيادية في نموذج الانحدار الخطي وهي كالاتي:-

1- الافتراضات الخاصة بحد الخطا في النموذج الانحدار الخطي:

<u>اولا</u> : متوسط حد الخطا العشوائي ε_i يجب ان يساوي صفر،

$$E(\varepsilon_i) = 0 \qquad \text{for all } i = 1, 2, 3\ldots n$$

ثانيا: تباين حد الخطا العشوائي يجب ان تكون ثابتة خلال جميع الفترات الزمنية التي اخذت منها العينة المدروسة، ويسمى هذا الفرض بـ تجانس تباين الخطا اي خاصية (Homoscedasticity)،

$Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2$ **for all i= 1, 2, 3… n**

ثالثا: المتغير العشوائي يتوزع توزيعا طبيعيا بمتوسط صفر و تباين ثابت

$\varepsilon_i \sim N(0, \sigma^2 I_n)$ **for all i= 1, 2, 3… n**

على فرض ان $x'x$ ***مصفوفة (non singular) اي ان*** $|x'x| \neq 0$ ***عندئذ لها معكوس فأن***

$$\hat{\beta} = (x'x)^{-1}x'y$$

رابعا: التباين المشترك بين حدين عشوائيين

$\varepsilon_i \neq \varepsilon_j$ يجب ان يساوي صفرا،أي: يجب ان يكون ε_i مستقلا عن ε_j ، اي عدم وجود الارتباط الذاتي بينهما (Autocorrelation)،

$Cov(\varepsilon_i, \varepsilon_j) = 0$ **for all i=j= 1, 2, 3… n**

خامسا: وجود خاصية الاستقلالية بين المتغيرات المستقلةوالمتغير العشوائي وتسمى بمشكلة التعدد الخطي (Multicollinearity)

$E(\varepsilon_i, X_{ij}) = 0$ **for all i=j= 1, 2, 3… n**

2- الافتراضات الخاصة بتوزيع متغير المعتمد Y في نموذج الانحدار الخطي البسيط ويتضمن :

اولا: قيمة متوسط Y_i تقع على خط مستقيم.

$\bar{Y} = E(Y_i) = b_0 + b_1 X_i$ **for all i= 1, 2, 3… n**

حيث ان b_0 و b_1 معلمات مقدرة

ثانيا: ثبات قيمة التباين لاية قيمة من قيم i لـ Y_i.

$Var(Y_i) = \sigma^2$ **for all i= 1, 2, 3… n**

ثالثا: يتوزع المتغير المعتمد Y توزيعا طبيعيا بوسط معين و تباين ثابت.

$Y \sim N(\mu, \sigma^2 I_n)$

رابعا: التباين المشترك بين Y_i وY_j يساوي صفراً،

$Cov(Y_i, Y_j) = 0$ **for all i=j= 1, 2, 3… n and all $i \neq j$.**

خامسا: العلاقة بين $\hat{Y}$ و X تكون علاقة خطية تتمثل بمعادلة خط المستقيم.

$\hat{Y} = b_0 + b_1 X_i$ **for all i= 1, 2, 3… n**

3-5 خصائص مقدرات المربعات الصغرى الاعتيادية[1] :

Ordinary Least Square Method Properties of

اولا: ان b تقدير غير متحيز للمعلمات β ، (Unbiased)

$E(b) = \beta$

ثانيا: المقدرات لها اقل تباين من بين كل المقدرات الخطية غير المتحيزة مقارنة بتباينات المقدرات الخطية غير المتحيزة التي تعود الى طرق اخرى، اي انها افضل التقديرات الخطية غير المتحيزة اي يتمتع بخاصية المسمى بـ (*<u>B</u>est <u>L</u>inear <u>U</u>nbiased <u>E</u>stimations*)،

ثالثا: انها مقدرات متسقة (Consistent) ،

$\Pr[\,|b_{OLS} - \beta| < \varepsilon] = 1$, $n \to \infty$ **for all** $\varepsilon > 0$

رابعا: مقدرات كافية (Sufficient) اي ان مقدر b يسمى بتقدير كافٍ الى β اذا كانت كل المعلومات حول معلمة المجتمع β محتوية على' بيانات العينة.

خامسا: مصفوفة التباين و التباين المشترك لمقدر b تساوي

$V(b) = (X'X)^{-1}\sigma^2$

3-6 القيم الشاذة Outliers Value [2], [3], [7], [14]

هي المشاهدات التي تقع بعيدة عن معادلة الانحدار، ويكون لها خطأ كبير مقارنة ببقية المشاهدات الطبيعية الاخرى في مجموعة البيانات، و ذلك سيكون لها تاثير في النموذج الخطي و تقديراته.

3-7 النقاط عالية الخلل ; High-Leverage Points

هي تلك النقاط التي يكون فيها المتجه X لبعض البيانات بعيدا عن البقية، و بهذا يكون لها خلل عالٍ، و يمكن القول ان النقاط عالية الخلل ترتبط فقط بالمتغيرات المستقلة و ليس لها علاقة بمتغير الاستجابة.

3-8 المشاهدات المؤثرة ; Influential Observations

هي المشاهدات التي تؤثر بصورة فردية او مشتركة في معادلة الانحدار بالمقارنة مع بقية المشاهدات.

3-9 نقاط الجذب : Leverage points [6]

هي تلك النقاط التي تقع بعيدة عن معظم القيم في مصفوفة X الموجودة في النموذج (3-2) و التي تشمل واحدة او اكثر من المتغيرات التوضيحية في تحليل الانحدار، و لها تأثير قوي في مقدرات المربعات الصغرى الاعتيادية، أي عندما تكون x_i لها شاذة تسمى (x_u,y_u) بنقطة الجذب عندما تقع x_u بعيدة عن قيم x_i في العينة من دون الأخذ بنظر الاعتبار قيمة y_u و بالتالي ليس بالضرورة ان تكون نقطة (x_u,y_u) من شواذ الانحدار اذا كانت تقع قريبة من خط الانحدار المحدد بالجزء الاكبر من البيانات، و عند ذلك تعد نقطة جيدة الجذب *(Good Leverage Point)*، و هذه النقاط لها تاثير كبير في تقديرات المعلمات؛ و المقصود بنقطة جيدة الجذب هي ان تكون قيمة من القيم x_i شاذة الا ان قيم y_i المقابلة لها تطابق النموذج بشكل جيد جداو هذه النقطة تنجذب تقديرات المربعات الصغرى الاعتيادية نحوها.

3-10 نقطة الانهيار : Break Down Point [5]

ان نقطة الانهيار هي احدى معايير قياس حصانة المقدرات و قياس نسبة التلوث في البيانات ، فبالنسبة الى طريقةالمربعات الصغرى فان نقطة انهيارها هي 0%، اي لا تستطيع مقاومة وجود اية قيمة شاذة في بياناتها، اما نقطة الانهيار لطريقتي مربعات الوسيط الصغرى(LMS) ، و مربعات المشرذمة الصغرى (LTS) ، فهي 50%، اي انهما تتمتعان بتقديرات ذات نقطة انهيار عالية؛ لأنها في حالة وجود الشواذ تعتبر افضل نقطة انهيار يمكن توقعها . اما في حالة تجاوز نسبة التلوث 50% فيصبح من المتعذر التميز بين الجزء الجيد و الجزء غير الجيد بسبب الشواذ في العينة.

3-11 معايير دقة مقدرات معلمات الانحدار; Measures of Accuracy of Regression Parameters [2]

في حالات التقدير التطبيقية تعامل الدقة كمعيار لاختبار طريقة التقدير، حيث هناك عدة عوامل غالبا ما تنعكس في الدقة، وعلى سبيل المثال البيانات غير كافية، او استخدام اسلوب لا يطابق نوع البيانات، فذلك سينعكس على الدقة.

فالدقة هي المعيار الاكثر استخداما لتقويم الانجاز لطرق التقدير، وهي تعكس صحة التنبؤ، و لكن من الصعوبات المرفقة لمعيار الدقة في حالات التقدير هي غياب مقياس مفرد مقبول و جامع للدقة في آن واحد، لان طرق التقدير المختلفة تستخدم اساليب مختلفة تجعل المقارنة بمعيار واحد غير صحيحة غالبا.

- *متوسط مربعات الخطأ; Mean Squares Error (MSE)*
- *الخطأ النسبي; Percentage Error (PE)*
- *متوسط الخطأ النسبي المطلق; Mean Absolute Percentage Error (MAPE)*
- *متوسط مجموع لأاخطاء المطلقة; Mean Sum of Absolute Error (MSAE)*

و سنحاول في هذا البحث استخدام المعيار الاكثر استعمالا في مجال التطبيقات الإحصائية و هي طريقة متوسط مربعات الخطأ; Mean Squares Error (MSE)

و يمكن توضيح هذا المعيار كما يلي:-

$$MSe = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1} \quad \ldots\ldots\ldots\ldots(3\text{-}5)$$

4. مفهوم الانحدار الحصين[9]، The Concept of Robust Regression

وهي طريقة بديلة لتقدير معلمات النموذج الانحدار تستخدم في حالة و جود الشواذ في البيانات، لان تقدير معلمات النموذج الانحدار الاعتيادي في ظل وجود هذه الشواذ تكون غير كفوءة؛ لانه تحدث حالة عدم تطابق بين بيانات موضوع الدراسة و الفروض الاساسية الواجب توافرها في النموذج، و بذلك تفقد الطرائق التقليدية ، مثل، طريقة المربعات الصغرى، خصائصها الجيدة لتقدير معلمات النموذج المدروس، لذا تم ايجاد طرائق إحصائية بديلة تكون مقاومة لمعالجة هذه المشكلة و هي ماتسمى بالطرائق الحصينة، اما المقدرات الناتجة عن هذه الطريقة البديلة فتسمى بالمقدرات الحصينة و تكون غير حساسة تجاه الشواذ إضافتة الى انها

تعطي مقدرات كفؤة تعادل كفاءة طريقة المربعات الصغرى الاعتيادية في حالة وجود الشواذ و تعطي مقدرات اكثر كفاءة في حالة وجودها في البيانات.

4-1 احصائية DIFFT و Leverage[12] ,[13],:-

حيث تعرف بأنها الاحصائية التي تقيس مدى تاثر المعلمات المقدرة للتغيرات اذا تم استبعاد كل من المشاهدات في مجاميع من البيانات و تحتسب على نحو الاتي:-

| DFFITS| > cutoff فحينها تعتبر هذه المشاهدة هي مشاهدة مؤثرة، و ان قيمة cutoff تختلف من مصدر الى اخر فبعض منهم يقول بأنها تساوي 2 اما في الاخر تساوي $2\sqrt{\frac{K}{2}}$. بينما ***Leverage*** فهي الاحصائية التي تقيس مدى تاثير كل قيمة مؤثرة في تحديد معلمات النموذج المقدر ، في هذه الحالة ان معدل اي قيمة تكون لها قيمة الـ ***Leverage*** مساوية الى 0.5 و وجود بيانات اكبر بـثلاث مرات من معدل المستوى المطلوب يعتبر من الشواذ و تؤثر في معلمات النموذج المقدر.

4-2 بوكس- ويسكر-بلوت (Box - Whisker- Plot)[7] ,[14],

تعتبر من افضل الطرق للتحري و التعامل مع الشواذ في مجاميع من البيانات و التي تعتمد على المدى الربيعي (Quartile Range) حيث ان الربيعيات يمكن ان تستخدم لتكوين مقاييس اخر للتباينات ، و المعتمدة على مدى الربيعي الثاني او نصف المدى الربيعي اي (Q_2 **or Interquartile range**)، و المعرفة بالشكل الاتي:-

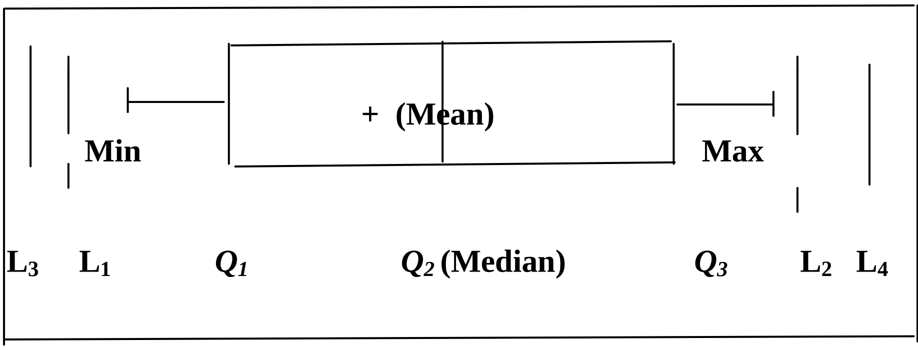
$$IQR= H = Q_3- Q_1$$

و الذي يقيس 50% من متوسط انحرافات للمشاهدات.

حيث تشير الاقيام الكبيرة من هذا المقياس على أن الربيعي الاول و الثالث متشتتان و الذي بدوره يشير الى مستوى عال من التباينات للمشاهدات.

و هذه الربيعيات تستخدم في بوكس-ويسكر-بلوت للكشف عن المشاهدات الشاذة و المعتمدة على خمس احصائيات: الحد الادنى و الحد الاعلى للمشاهدات، و الربيعي الاول و الثاني و الثالث. و يكون توزيعه على الشكل الخاص بهذه الإحصائية كما يلي:-

الخطوط الثلاثة العمودية منها تشير الى الربيعيات الثلاثةو الخطوط المددة من جهتي اليمنى واليسارى تسميان بالشوارب (***Whiskers***) بحيث اي قيمة تقع خارج هذا الشوارب تعتبر قيمة شاذة. كما هو موضح في الشكل رقم (1):-



المصدر :- تم اعداده من قبل الباحثان.

الشكل رقم (1)

توضح مكونات

BOX-WHISKER-PLOT

تمتد الشوارب الى الخارج حتى تصل الى اقل من 1.5 مرات من H او الى اقصى حد من المشاهدات المتطرفة.

بمعنى اخر، اذا وقع اي مشاهدة خارج مدى $Q_1 -(1.5\ IQR)$ و $Q_3 +(1.5\ IQR)$ و في هذه الحالة تعتبر القيم من نوع الشواذ الضمنية اما اذا وقع خارج حدود $Q_1 -(3\ IQR)$ و $Q_3 +(3\ IQR)$ حينها يعتبر وجود هذه المشاهدة الشاذة مشكلة في البيانات. و ايضا يمكن ان تستخدم كمقياس لمعرفة خاصية التوزيع الطبيعي للأخطاء العشوائية و ذلك عن طريق ملاحظة نوع و اتجاه الالتواء للشوارب (Whiskers).

4-3 طرق التقديرات الحصينة Robust Estimation Methods

بما ان تفسير اي بيانات مبني على اساس قواعد و طرائق خاصة ولكن في بعض الاحيان يواجه الباحث مشاكل في البيانات مثل: ابتعاد بيانات العينة عن التوزيع الطبيعي و ذلك بسبب وجود القيم الشاذة فيها، او اختلاف توزيع المجتمع قيد الدراسة عن التوزيع الطبيعي، و هذا يسبب انحرافات عن الافتراضات التي و ضعت للطرائق التقليدية منها: طريقة المربعات الصغرى الاعتيادية لتقدير معلمات النموذج الخطي، لهذا السبب لابد من وجود طرائق بديلة. و إضافة الى ذلك توصل كثير من الباحثين الى نتيجة بأن هذه الطرائق تكون غير كفوءة في حالة عدم تحقيق احدى الافتراضات الازمة او الشروط التي تعتمدها هذه الطرائق، لذا عمل كثير من الباحثين من اجل ايجاد طرائق اكثر كفاءة و لا تتاثر كثيرا بالانحرافات عن الافتراضات المحددة، فكانت هذه الطرائق المسمى بالطرائق الحصينة حيث اقل تاثرا في حالة اختراق البيانات لشرط من شروط التحليل المستخدم، فان المقدرات الحصينة لنموذج الانحدار قريبة من كفاءة طريقة المربعات الصغرى الاعتيادية في حالة تحقق هذه الشروط و تمتاز ايضا بانها مناسبة لفئة واسعة من التوزيعات في تقدير معلمات النموذج الخطي[9]. و في هذا البحث تم اختيار طريقتين لايجاد المقدرات في النموذج الانحدار الخطي و هما :-

4-3-1 طريقة مربعات الوسيط الصغرى; Least Median of Squares Method (LMS)[10]

ان طريقة مربعات الوسيط الصغرى (LMS) تم اقتراحه من قبل (Russeeuw, 1984) على اساس فكرة (Hample, 1975) لتزيد من مقدرات الحصانة للمعلمات في الموذج الانحدار الخطي، و ذلك باستبدال مجموع مربعات البواقي في طريقة (OLS) و وضع وسيط مربعات البواقي بدلا منها، ويجب للمقدرات (LMS) ان تحقق الامثلية التالية:

افرض ان

$X' = (x_{i1}, x_{i2}, \ldots, x_{ip})$ i =1, 2, …, n   j = 1,2, …, p

$Y_i = (y_1, y_2, \ldots, y_n)$

حيث ان X' و Y_i متجهات حقيقية معطاة و نفرض با ن $\frac{n}{2} \geq p$

P هو عدد المتغيرات الداخلة في النموذج

$\mathbf{X} = \lfloor x_{ij} \rfloor$

X هو مصفوفة ذات (nxp) تامة الرتب

افرض بان $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$

و بواسطة طريقة مربعات الوسيط الصغرى (LMS) يمكن ايجاد b_{LMS} حسب الصيغة الآتية:

$$b_{LMS} = \min_b \operatorname*{med}_{i=1}^{n} (Y_i - X'\beta)^2 \qquad \text{..........................(4-1)}$$

حيث ان

$$\varepsilon_i^2 = (Y_i - X'\beta)^2$$

أو

$$b_{LMS} = \min_b \operatorname*{med}_{i=1}^{n} (\varepsilon_i^2) \qquad \text{..........................(4-2)}$$

b_{LMS} مقدر يمثل منتصف اقصر نصف في العينة الجزئية و يسمى بمقدر مربعات الوسيط الصغرى ولايجاد مقدر b_{LMS} في حالة وجود متغير احادي ذي حجم n يجب الاخذ بنظر الاعتبار(n-h+1) من العينات الجزئية الآتية:

$$\{X_{(1)}, X_{(2)}, \ldots, X_{(h)}\}$$
$$\{X_{(2)}, X_{(3)}, \ldots, X_{(h+1)}\}$$
$$\{X_{(3)}, X_{(4)}, \ldots, X_{(h+2)}\} \qquad \text{..........................(4-3)}$$
.
.
$$\{X_{(n-h+1)}, X_{(n-h+2)}, \ldots, X_{(n)}\}$$

ثم ايجاد اقصر نصف(Shortest Half) في العينة الجزئية و هذا يكون بايجاد اصغر الفروقات كما يأتي:-

$X_{(h)} - X_{(1)}$
$X_{(h+1)} - X_{(2)}$
$X_{(h+2)} - X_{(3)}$ ……………………….(4-4)
.
$X_{(n)} - X_{(n-h+1)}$

حيث ان $h = \left[\frac{n}{2}\right] + 1$ و ان $X_{(i)}$ هي القيم المرتبة.

و العينة الجزئية التي تقابل اصغر فرق في (4-4) و تحتوي على h من القيم تسمى أقصر نصف (Shortest Half) لانها تمتلك المدى الاقصر (Shortest Range) من بين كل العينات الجزئية الممكنة ذات h من العناصر.

و مقدر b_{LMS} يكون مساويا لنقطة الوسط (Mid Pint) هذه العينة الجزئية التي تقابل اصغر فرق و في حالة عدة اصناف قصيرة متساوية لـ متوسط منتصفاتها.

4-3-2 طريقة المربعات المشرذمة Least Trimmed Squares Method [11]

وهي طريقة إحصائية لتقدير المعاملات المجهولة لنموذج الانحدار الخطي و تعد حصينة وبديلة للطرائق التقلية و قد اقتر (Rousseeuw, 1984) هذه الطريقة؛ لانها تتميز بكفاءة إحصائية و استقرار موضعي افضل من طريقة مربعات الوسيط الصغرى و لنفرض بان نموذج الانحدار الطي لعينة (Y_i, X_i) ، و متغير الاستجابة

$Y_i \in R$ متجه من المتغيرات التوضيحية $x \in R^p$ ، $\beta \in R^p$.

حيث P: تمثل عدد المعلمات المقدرة:

$$Y_i = X' \beta + \varepsilon_i \qquad \text{……………………….(4-5)}$$

حيث i = 1, 2, 3, …n

بواسطة طريقة المربعات المشرذمة الصغرى (LTS) يمكن الحصول على مقدر b_{LTS} كلاتي:

$$b_{LTS} = \min_b \sum_{i=1}^{h} (\varepsilon^2)_{nn} \qquad \text{……………………….(4-6)}$$

h: ثابت لديها مدى $(\frac{n}{2} < h < n)$

حيث ان $\left(\varepsilon^2\right)_{nn}$ هي مربع الاخطاء المرتبة (الاخطاء تربع ثم ترتب):

$$\left(\varepsilon^2\right)_{1n} \le \left(\varepsilon^2\right)_{2n} \le \ldots \le \left(\varepsilon^2\right)_{nn}$$

وقد اقترحت الباحثان مقدر b_{LTS} لموقع المتغير الاحادي ذي حجم n و الاخذ بنظر الاعتبار (n-h+1) من العينات الجزئية الاتية:

$\{X_{(1)}, X_{(2)}, \ldots, X_{(h)}\}$
$\{X_{(2)}, X_{(3)}, \ldots, X_{(h+1)}\}$
$\{X_{(3)}, X_{(4)}, \ldots, X_{(h+2)}\}$ ……………………….(4-7)
.
.
$\{X_{(n-h+1)}, X_{(n-h+2)}, \ldots, X_{(n)}\}$

كل عينة جزئية تحتوي على h من العناصر و سميت كل عينة جزئية بالنصف المتجاور او النصف المعدي (Contagious Half) و يتم حساب المتوسط لكل عينة جزئية كالآتي:

$$\bar{X}_{(1)} = \frac{1}{h} \sum_{i=1}^{h} X_{(i)}$$

$$\bar{X}_{(2)} = \frac{1}{h}\sum_{i=2}^{h+1} X_{(i)} \qquad \text{...........................(4-8)}$$

.

.

$$\bar{X}_{(n-h+1)} = \frac{1}{h}\sum_{i=n-h+1}^{n} X_{(i)}$$

و حساب مجموع المربعات لكل عينة جزئية:

$$SQ_{(1)} = \sum_{i=1}^{h}(X_{(i)} - \bar{X}_{(1)})^2$$

$$SQ_{(2)} = \sum_{i1}^{h}(X_{(i)} - \bar{X}_{(2)})^2 \qquad \text{...........................(4-9)}$$

.

.

$$SQ_{(n-h+1)} = \sum_{i=n-h+1}^{n}(X_{(i)} - \bar{X}_{(n-h+1)})^2$$

يكون مقدر المربعات المشرذمة الصغرى b_{LTS} ذلك المتوسط الذي يقابل اصغر مجموع مربعات في (4-9) .

5. الجانب التطبيقي

5-1 وصف البيانات

تم جمع البيانات من دائرة الانواء الجوية في اربيل، و تتألف العينة من بيانات شهرية مسجلة خلال السنوات (1998-2010) و بحجم 12 مشاهدة لكل سنة. و ذلك لدراسة العلاقة بين المتغير التابع و المتمثلة بالرطوبة النسبية، و المتغيرات المستقلة المتمثلة بدرجة الحرارة العضمى و الصغرى ، و الضغط الجوي ، و كمية الأمطار المتساقطة، و اخيرا سرعة الرياح. و من ثم تطبيق التحليل الإحصائي للبيانات و مناقشة النتائج كما ذكر من قبل في الجانب النظري من البحث و ذلك لغرض الوصول الى هدف البحث و المتمثل بان استخدام الطرائق الحصينة افضل لتقدير معلمات النموذج الخطي و ذلك لايجاد أنسب نموذج لتمثل العلاقة المدروسة؛ لأن العوامل المؤثرة في حالات الطقس لا تتوزع توزيعا طبيعيا اي (توجد فيها : للعوامل اقيام شاذة).

و اخيرا من الجدير بالذكر تم استخدام الحقيبة البرامجية الجاهزة ***Statgraphics, version 4(1999)*** و ***S- Plus, version 4.5 (2000)*** للحصول على النتائج.

5-2 اختبار البيانات

تم اختبار البيانات قيد الدراسة لاختبار التوزيع الطبيعي الذي تمكن ملاحظته من اشكال الـ (***Box - Whisker- Plot***) و الذي بدوره يؤثر بشكل كبير في تقدير المعلمات النموذج الخطي الاعتيادي " طريقة المربعات الصغرى الاعتيادية" و تحديد المشاهدات الشاذة و المؤثرة بطريقة الإحصائيتين المذكورتين ***DIFFT*** و ***Leverage*** و تحليلات بوكس- ويسكر-بلوت (***Box - Whisker- Plot***) بالنسبة للمشاهدات الشاذة و ذلك بالاعتماد على اختبار الفرضية التالية:-

H_0 : يوجد قيم شاذة

H_1 : لا يوجد قيم شاذة

و يتم تفسير نتائج الاحصائية ***DIFFT*** كالاتي:

إذا كانت اية قيمة اكبر من قيمة | **DFFITS| > cutoff** و قيمة الـ **cutoff** مساوية الى 2 او **2√(k/2),** فحينها تعتبر هذه المشاهدة مشاهدة مؤثرة، والتي تعتبر من الشواذ في البيانات النموذج؛ اما إحصائية ***Leverage*** تفسر بانها و جود اية مشاهدة او قيمة اكبر بـ(3) مرات من معدل مستوى المطلوب يعتبر تلك المشاهدة من الشواذ. بينما يتم تفسير احصائية بوكس- ويسكر-بلوت (***Box - Whisker-***

(*Plot* كالتالي:

اذا و قعت اية مشاهدة خارج مدى $Q_1-(1.5\ IQR)$ و $Q_3+(1.5\ IQR)$ و في هذه الحالة تعتبر القيم من الشواذ الضمنية، اما اذا وقعت خارج حدود $Q_1-(3\ IQR)$ و $Q_3+(3\ IQR)$ حينها تعتبر وجود هذه المشاهدة الشاذة في البيانات مشكلة. و كذلك عن طريق (*Box - Whisker- Plot*) تمكن معرفة نوع التوزيع لهذه البيانات أي: هل يتوزع البيانات توزيعا طبيعيا ام لا .

الجدول رقم (**1**)

يوضح المشاهدات الشاذةفي بيانات قيد الدراسة

Unusual Residuals

Row	Y	Predicted Y	Residual	Studentized Residual
13	88.0	68.175	19.825	2.86
16	72.0	57.637	14.363	2.05
22	23.0	41.6478	-18.6478	-2.68
45	18.0	33.7967	-15.7967	-2.25
53	42.0	26.7769	15.2231	2.29
86	50.0	64.9675	-14.9675	-2.13
87	35.0	60.8384	-25.8384	-3.80
88	26.0	50.0103	-24.0103	-3.48
93	53.0	30.6799	22.3201	3.23
94	63.0	41.1949	21.8051	3.15
156	48.0	67.0426	-19.0426	-2.73

المصدر:- تم اعداد الجدول من قبل الباحثان.

ان الجدول اعلاه يوضح لنا بانه يوجد 11 مشاهدة شاذة في البيانات و العائدة لبيانات الأشهر: كانون الثاني ، و نيسان، و تشرين الاول عام 1999 بينما وجود الشواذ عام 2001 تعود لشهر ايلول لعام2002 و لشهر ايار و ان البيانات لا تخلو من الاقيام الشاذة، كما هو واضح للأشهر: شباط، و اذار، و نيسان، و ايلول، و تشرين الاول عام 2005، يوجد شواذ في بيانات عام 2010 لشهر كاون الاول . و بالنتيجة يؤدي هذا الى عدم رفض الفرضية العدم القائلة بانه توجد اقيام شاذة في البيانات.

الجدول رقم (**2**)

يوضح المشاهدات المؤثرة في بيانات قيد الدراسة

Influential Points

Row	Leverage	DFITS
11	0.0536853	0.466953
13	0.0230776	0.439047
14	0.0637467	0.442994
22	0.0211637	-0.39345
53	0.119014	0.841039
79	0.531951	0.230998
87	0.0267876	-0.631161
90	0.205798	-0.243766
93	0.0187681	0.447028
94	0.0164977	0.407821

108	0.115227	-0.504892
121	0.071085	0.450473
122	0.161781	-0.0849514
146	0.144718	-0.577491

Average leverage of single data point = 0.0384615

المصدر:– تم اعداد الجدول من قبل الباحثان.

بينما الجدول رقم (2) يبين لنا اقيام كل من DIFFT و Leverage بالنسبة للاختبار الاول بحسب الإحصائية cutoff < |DFFITS| ، فيعتبر من النقاط المؤثرة، وكما معروف ان النقاط المؤثرة تجذب معادلة خط الانحدار نحوها. و تؤثر في معلمة الميل و بالتالي تؤثر في النتيجة، و هذا يعني اذا تم استبعاد هذه القيمة في البيانات فان قيمة المعلمات المقدرة سوف تتغير بالنسبة للعلاقة المدروسة و هنالك 6 قيمة لديها اقيام كبيرة غير معتادة من الـ DFFITS .

فاقيام لأشهر: كانون الثاني و تشرين الاول لعام 1999 و ايار عام 2002و اذار و ايلول و تشرين الاول عام 2005 ، يعدّ من المشاهدات الشاذة .

ويبين الجدول ايضا، بأن قيمة معدل *Leverage* لاقيام البيانات للظاهرة المدروسة تساوي 0.0384615 ؛ وان اية قيمة اكبر منه بثلاث مرات فتعد من الشواذ و تؤثر في النموذج المقدر، و هنا خمسة منها اكبر بثلاث مرات من قيمة معدل *Leverage* و اثنتان منها اكبر بخمس مرات منها ، و كالاتي:-

حيث ان القيمة العائدة لشهر تشرين الثاني عام 2003 ، و شهر شباط عام 1999 ، و تموز عام 2004 تليها قيمة شهر حزيران عام 2005 ، و كانون الاول لعام 2006، و الشهران: كانون الثاني و شباط عام 2008 ، أخيرا قيمة شهر شباط عام 2010 كلها تعدّ من المشاهدات الشاذة . فهذا يعني بانه من المؤكد سوف يغير النوذج في حالة عدم وجوده في البيانات.

الجدول رقم (3)

يوضح الوصف الاحصائي لبيانات قيد الدراسة

Summary Statistics

	Count	*Average*	*Median*	*Mode*	*Variance*
Relative Humidity	156	46.9231	46.5	23.0	381.62
Max Tempereture	156	27.6492	27.63	38.8	118.745
Min Tempereture	156	14.5803	14.66	5.1	68.1074
Atmospheric Pressure	156	962.618	964.3	0.0	43.9249
Rainfull	156	33.3292	14.5	0.0	1708.04
Wind Speed	156	2.45344	2.45	2.6	0.538993

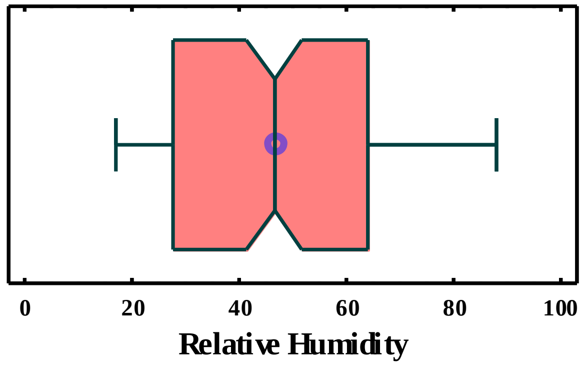
	SD	*Minimum*	*Maximum*	*Range*	*Q1*
Relative Humidity	19.5351	17.0	88.0	71.0	27.5
Max Tempereture	10.897	9.47	45.0	35.53	17.05
Min Tempereture	8.25272	0.632	29.9	29.268	6.8
Atmospheric Pressure	6.62759	921.3	972.4	51.1	957.89
Rainfull	41.3285	0.0	189.0	189.0	0.0
Wind Speed	0.734161	0.5	5.8	5.3	1.924

	Q3	*IQR*	*Stnd. skewness*	*Stnd. kurtosis*
Relative Humidity	64.0	36.5	0.989161	-3.4149
Max Tempereture	38.49	21.44	-0.0913549	-3.90142
Min Tempereture	22.45	15.65	0.591942	-3.5091
Atmospheric Pressure	968.0	10.11	-8.978	20.7861

Rainfull	**62.2**	**62.2**	**6.85502**	**3.37923**
Wind Speed	**3.0**	**1.076**	**2.31027**	**4.6822**

المصدر:- تم إعداد الجدول من قبل الباحثان.

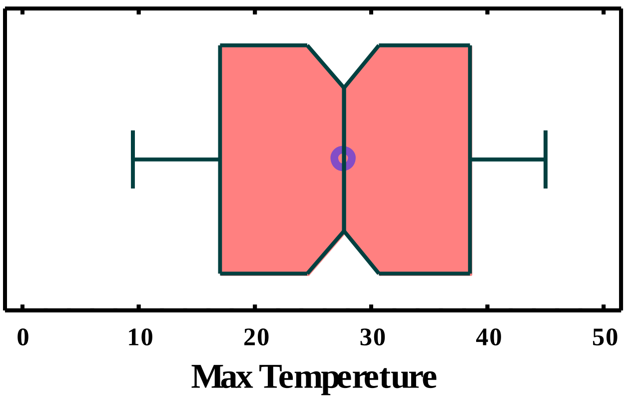
يوضح الجدول اعلاه، الجدول رقم (3)، الملخص الإحصائي لكل ظاهرة من الظواهر المدروسة. و يتضمن مقاييس النزعة المركزية، و مقايس التشتت و مقاييس الشكل. كما يوضح معامل الالتواء القياسي و التفلطح القياسي للبيانات، و اللتان بدورهما تستخدمان لتحديد فيما اذا كانت العيّنات المدروسة آتية من التوزيع الطبيعي ام لا. بما ان اقيام هذين الاحصائيين تقعان خارج مدى ±2 فهذا يشير الى معنوية ابتعاد البيانات من التوزيع الطبيعي و الذي بدوره يقود الى عدم تحقق اغلبية الفروض الواجب توافرها في البيانات.



الشكل رقم (2)

يوضح بوكس- ويسكر- بلوت لبيانات الرطوبة النسبية

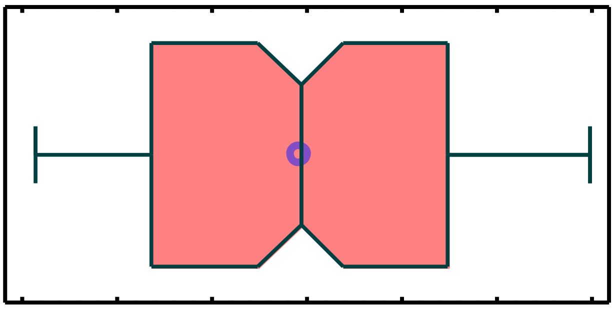
الشكل رقم (2)، يوضح بوكس- ويسكر- بلوت لبيانات الرطوبة النسبية، اما بالنسبة لتفسير المشاهدة و بالاعتماد على الجدول رقم (3): و بحسب تحليلات بوكس- ويسكر- بلوت للبيانات نرى بانه طالما لا توجد اية قيمة تقع خارج مدى (27.25- ,118.75) و لا خارج حدود (82- ,173.5) فهذا يعني انه لاتوجد قيمة شاذة ضمنية في هذه البيانات. و بالرغم من انّ قيمة تفلطحها القياسي تقع خارج الحد المسموح اي ±2 و لكنها تتوزع توزيعا طبيعيا بحسب الشكل اعلاه. و ممّا يؤكد ذلك ان قيمة الوسيط و المعدل متساوية.



الشكل رقم (3)

يوضح بوكس- ويسكر- بلوت لبيانات درجات الحرارة العظمى

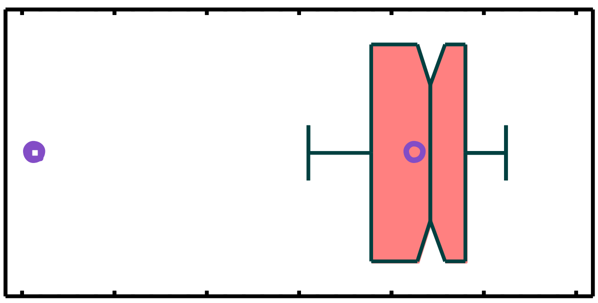
في حين يوضح الشكل رقم (3) بوكس- ويسكر- بلوت لبيانات درجات الحرارة العظمى بانه ليس هناك اقيام لتلك المشاهدة تقع خارج المدى (15.11- ,70.65) و لا خارج حدود (47.27- ,102.81) فهذا يعني بان هذه البيانات لا تحتوي على مشاهدات شاذة ضمنية التي بدورها تنجذب نحوها معادلة خط الانحدار و التي تؤثر في النتائج.



الشكل رقم (4)

يوضح بوكس– ويسكر– بلوت لبيانات درجات الحرارة الصغرى

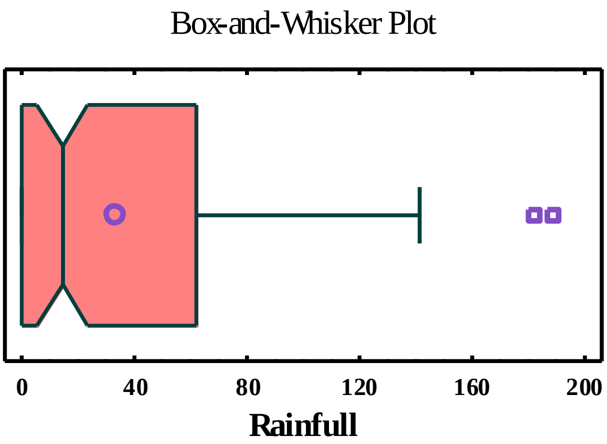
بينما يوضح الشكل رقم (4) بوكس- ويسكر- بلوت لبيانات درجات الحرارة الصغرى بانه هناك اقيام تلك المشاهدة تقع خارج المدى (45.925, -16.675) ، و لجميع الاعوام على المعدل ابتداءً من شهر ايار حتى شهر تشرين الاول، و من شهر كانون الثاني حتى شهر كانون الاول. و توجد اقيام تقع خارج حدود(69.5, -40.15) و العائدة لشهر تموز عام 2000، و شهري تموز و اب عام 2002، و الأشهر: نيسان ، و ايار، و حزيران، و تموز، و اب، و ايلول، و تشرين الاول، و تشرين الثاني، لعام 2005، و كذلك للأشهر: حزيران، و تموز، و اب، و ايلول، و تشرين الاول، و تشرين الثاني، و كانون الاول عام 2006 ، و كذلك الأ شهر: تموز، و اب، و ايلول، و تشرين الاول، و تشرين الثاني، و كانون الاول عام 2007، و كانون الثاني حتى كانون الاول عام 2008، و الأشهر: كانون الثاني، و شباط، و تموز، و اب لعام 2009 و اخيرا لشهر اب 2010 . فهذا يعني بان هذه البيانات تحتوي على مشاهدات شاذة ضمنية التي بدورها تنجذب نحوها معادلة خط الانحدار و التي تؤثر في النتائج.



الشكل رقم (5)

يوضح بوكس– ويسكر– بلوت لبيانات الضغط الجوي

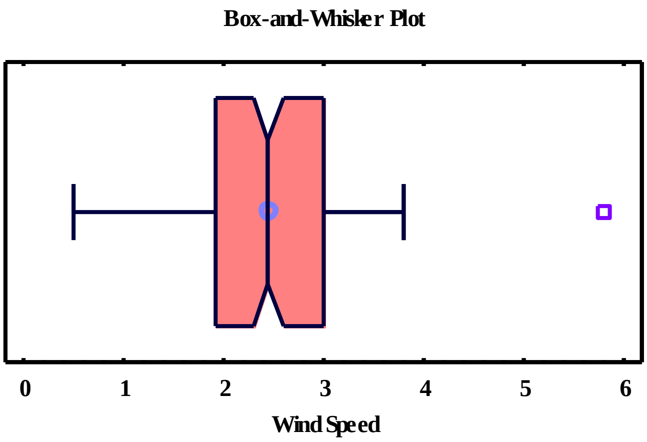
نلاحظ بانه من الشكل اعلاه، بوكس-ويسكر-بلوت للضغط الجوي، قيمة شاذة واضحة و مشاهدة مؤثرة ضمنية في اقيام بياناتها لانها تقع خارج المدى (942.725، 983.165) لجميع شهر تموز عام 2004 . اما بالنسبة للمشاهدات التي تعتبر مشكلة في البيانات فلا توجد اية قيمة خارج الحدود (998.33, 927.56). و كان هذا واضحا ايضا في الجدول رقم (3) الملخص الاحصائي فاقيام كل من الالتواء و التفلطح القياسي اثبتت ذلك .



الشكل رقم (6)

يوضح بوكس– ويسكر– بلوت لبيانات كمية الامطار المتساقطة

اما بالنسبة الى بيانات كمية الامطار المتساقطة الشكل رقم (6) ، هنـاك قيمتـان تقعان خـارج المدى (-155.5, 93.3) اي انـها قيمـة شـاذة ضـمنية بينمـا لا توجـد ايـة قيمـة شـاذة تقـع بيـن (-186.6, 248.8) تسبب فـي مشكلة للبيانـات المدروسة، كما هو واضح من الشكل ان هنالك قيمتـان متطرفتـان ومؤثرتـان و تعودان الـى شهر كانون الاول عام 2006 و شهر شباط عام 2010 .



الشكل رقم (7)

يوضح بوكس– ويسكر– بلوت لبيانات سرعةالرياح

واخيرا بالنسبة لبيانات سرعة الرياح، الشكل رقم (7) يوضح بان قيمة شهر شباط عام 2008 تقع خارج المدى (0.31، 4.614) و لكن لا توجد ايـة قيمـة تقـع خـارج المدى (-1.304, 6.228) فهذا يعـني بانـه توجد قيمـة شاذة و مؤثرة .

الجدول رقم (4)، يوضح نتـائج تحليل الانحـدار المتعدد و جدول تحليل التبـاين للعلاقـة بين الرطوبـة النسـبية و بقية المتغيرات المدروسة، عليه فان معادلة الانحدار الخطي لهذه العلاقة يمكن تمثيله بالصيغة الاتية :

$$Yi = -49.6527 - 1.41549X1 + 0.0970944X2 + 0.134376X3 + 0.0881755X4 + 0.817534X5$$

وان هذا النموذج يشير الى ان العلاقة بين الرطوبـة النسـبية و درجـة الحرارة العظمـى هي عكسية اي ان عندما تزداد درجات الحـرارة العظمى تقل الرطوبة النسـبية فـي حين ان الرطوبـة النسـبية تـزداد بزيـادة كل من درجات الحرارة الصغرى و الضغط الجوي و كمية الامطار المتساقطة و سرعة الرياح.

بما انه القيمة الاحتمالية لـ α اي P-value لجدول تحليل التباين اقل من 0.01 و حسب فرضية البحث يوجد علاقة احصائية معنوية بين المتغيرات.

H_0 : أنعدام المعنوية

H_1 : وجود علاقة معنويه

في حين يوضح قيمة R^2 المعدلة 86.4638% من معنوية هذه العلاقة اي 86.4638% تاثير كل من المتغيرات الخمسة على الرطوبة النسبية اي مدى القوة التفسيرية للمتغيرات المستقلة لهذه العلاقة على الرطوبة النسبية. و من الجدول نرى ايضا بان قيمة متوسط مربعات الاخطاء MSE تساوي 51.6567 ، و قيمة دربن واتسون D.W تساوي 1.07712 مقارنة بالقيمة القياسية المساوية الى 1.4 نلاحظ بانه توجد مشكلة الارتباط الذاتي في البيانات و هذه نتيجة منطقية بموجب طبيعة العلاقة بين البيانات المدروسة.

و لمعرفة فيما اذا كان بامكان تبسيط النموذج نلاحظ بانه اكبر قيمة لـ P-value تساوي 0.7441 و العائدة للمتغير المستقل درجة حرارة الصغرى (Min Temperature) تليها كبرى قيمة ثانية: 0.4119 العائدة لـ الضغط الجوي (Atmospheric Press)و ثالث واحد 0.3233 هو تاثيرات سرعة الرياح (Wind Speed) و كلها اكبر من 0.10 اي ليس لهذه المتغيرات تاثير معنوي في الرطوبة النسبية تحت مستوى المعنوية 90% او مستويات اعلى. و بالتالي يفضل استبعاد هذه المتغيرات من النموذج و ان هذه النتيجة معقولة في ظل وجود اقيام شاذة في بياناتها و لكن نظرا لاهمية تاثيرها في الرطوبة النسبية لا نستطيع استبعاده من النموذج.

الجدول رقم (4)

يوضح تقدير معالمات النموذج ومتوسط مربعات الخطأوR^2

Multiple Regression Analysis

Dependent variable: Relative Humidity

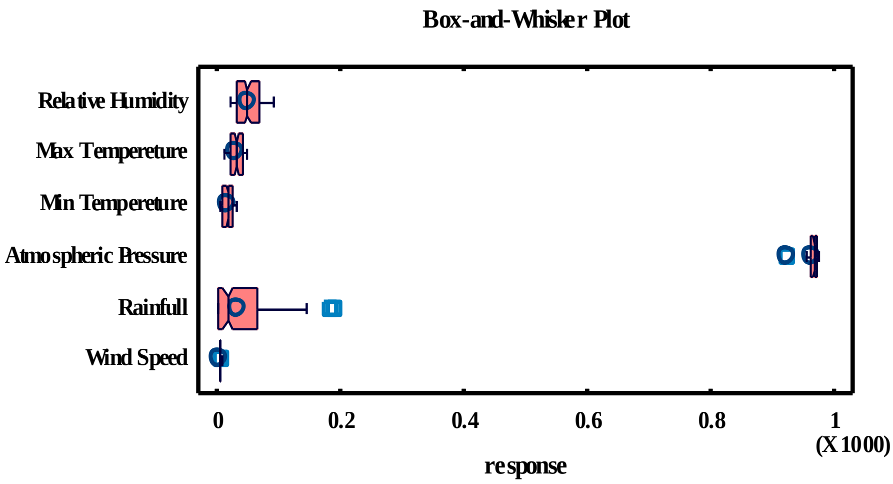
Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-49.6527	159.879	-0.310564	0.7566
Max Temperature	-1.41549	0.245534	-5.76495	0.0000
Min Temperature	0.0970944	0.296848	0.327084	0.7441
Atmospheric Press	0.134376	0.163298	0.82289	0.4119
Rain full	0.0881755	0.0214734	4.10627	0.0001
Wind Speed	0.817534	0.825072	0.990864	0.3233

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	51402.6	5	10280.5	199.02	0.0000
Residual	7748.5	150	51.6567		
Total (Corr.)	59151.1	155			

R-squared = 86.9005 percent
R-squared (adjusted for d.f.) = 86.4638 percent
Durbin-Watson statistic = 1.07712

المصدر:– تم اعداد الجدول من قبل الباحثان.



الشكل رقم (8)

يوضح بوكس– ويسكر– بلوت لبيانات تحت الدراسة معا

يتضح لنا من الشكل رقم (8) بانه بصورة عامة يوجد الشواذ في جميع المشاهدات المتعلقة بدراسة الانواء الجوية و هذا شيء طبيعي لانه لا يمكن التكهن بما سوف يكون عليه الطبيعة و بالتالي سوف لن نستطيع الاعتماد على طريقة المربعات الصغرى لتوفيق النموذج لتمثل العلاقة بين هذه المتغيرات.

الجدول رقم (5)

يوضح نتائج تحليل الانحدار الحصين باستخدام طريقة

LMS

Coefficients Value:

Intercept	**74.1487**
Max Temperatures	**-1.3544**
Min Temperatures	**0.1047**
Atmospheric Press	**0.0062**
Rainfall	**0.1525**
Wind Seed	**-1.0450**
MSE	**4.459**
Adj R^2	**% 60.09**

المصدر:– تم اعداد الجدول من قبل الباحثان.

يوضح لنا الجدول رقم (5) نتائج تحليل الانحدار المتعدد باستخدام طريقة LMS نرى بانه معادلة الاتجاه العام يكون بالصيغة الاتية:-

Yi=74.1487-1.3544 X1+0.1047 X2+0.0062 X3+0.1525 X4-1.0450X5

هذا النموذج يشير الى ان العلاقة بين الرطوبة النسبية و كل من درجة الحرارة العضمى و سرعة الرياح هي علاقة عكسية اي ان عندما تزداد كلاً من سرعة الرياح و درجات الحرارة العضمى تقل الرطوبة النسبية في حين ان الرطوبة النسبية تزداد بزيادة كل من درجات الحرارة الصغرى و الضغط الجوي و كمية الامطار المتساقطة .

في حين يوضح قيمة R^2 المعدلة بان كلاً من المتغيرات المستقلة تفسر 60.09% من تاثير المتغيرات في الرطوبة النسبية. في حين قيمة متوسط مربعات الأخطاء تساوي MSE = 4.459 التي بدوره تعتبر معيارا لقياس مدى معنوية المقدرات في النموذج.

الجدول رقم (6)

يوضح نتائج تحليل الانحدار الحصين باستخدام طريقة

LTS

Coefficients Value:

Intercept	-55.5875
Max Temperatures	-1.4570
Min Temperatures	0.2631
Atmospheric Press	0.1395
Rainfall	0.1274
Wind Seed	0.3975
MSE	4.963
Adj R^2	% 93.33

المصدر:– تم اعداد الجدول من قبل الباحثان.

الجدول رقم (6) يوضح نتائج تحليل الانحدار المتعدد باستخدام طريقة LTS و ان معادلة الاتجاه العام يتخذ الصيغة الاتية:-

$$Yi=-55.5875-1.4570X1+0.2631X2+0.1395X3+0.1274X4+0.3975X5$$

و ان هذا النموذج يشير الى ان العلاقة بين الرطوبة النسبية و درجة الحرارة العضمى هي عكسية اي ان عندما تزداد درجات الحرارة العضمى تقل الرطوبة النسبية في حين ان الرطوبة النسبية تزداد بزيادة كل من درجات الحرارة الصغرى و الضغط الجوي و كمية الامطار المتساقطة و سرعة الرياح.
في حين يوضح قيمة R^2 المعدلة بان كلاً من المتغيرات المستقلة تفسر 93.33 % من تاثير المتغيرات على درجات الحرارة. في حين قيمة متوسط مربعات الاخطاء تساوي MSE = 4.963 ومن هنا نلاحظ بانه ليس من الظروري ان تكون قيمة Adj R^2 كبيرة أو ان تكون دليلا لقوة العلاقة المدروسة.

الجدول رقم (7)

يوضح مقارنة بين الطرق المستخدمة لتقدير معلمات الانحدار باستخدام احصائية

MSE

الطرق المستخدمة	MSE
طريقة المربعات الصغرى (OLS)	51.6567
طريقة الوسيط الصغرى (LMS)	4.459
طريقة المتوسطات المشرذمة (LTS)	4.963

المصدر:– تم اعداد الجدول من قبل الباحثان.

6. الاستنتاجات و التوصيات

1- نستتج بأنّ بيانات المتغير المعتمد، الرطوبة النسبية *"Relative Humidity"* و المتغير المعتمد درجات الحرارة العضمى "Max Temperature"، يتوزع توزيعا طبيعيا و ذلك بسبب عدم وجود اقيام شاذة بين بياناتها.
2- بالرغم من ان القيمة الاحتمالية لـ α اي P-value بطريقة المربعات الصغرى للمتغيرات *" Min Temperature"* درجة الحرارة الصغرى و الضغط الجوي *"Atmospheric Pressure"* وسرعة الرياح *"Wind Speed"* ليست معنوية و لكن لا نبعدها من النموذج نظرا لاهمية هذه العوامل في

العلاقة المدروسة و يعود سبب كونها غير معنوية لوجود اقيام مؤثرة بين بياناتها من هنا ياتي دور استخدام الطرق الحصينة.

3- بما ان قيمة MSE لطريقة مربعات الوسيط الصغرى (LMS) اقل بقليل من طريقة المتوسطاتالمشرذمة من بقية الطرق ، انه يقيس دقة تقدير المعلمات المقدرة لمعلمات الانحدار، فهذا يعني المعلمات المقدرة بهذه الطريقة هي ادق من بقية الطرق.

4- و بالنتيجة، نختار الطريقة الحصينة: طريقة مربعات الوسيط الصغرى (LMS) لتمثل انسب نموذج للعلاقات بين العوامل المؤثرة في الرطوبة النسبية اي دراسة حالات الطقس في محافظة اربيل وفي ظل وجود الشواذ في البيانات.

5- نوصي باستخدام طرائق اخرى للكشف عن الشواذ في دراسات مستقبليةو مثل طريقة تقديرات S وL و R.

6- أخذ تاثير عوامل اضافية كمتغيرات مستقلة لمعرفة تاثيرها في الرطوبة النسبية مثل نسبة التلوث في الجو.

7- تطبيق الطرائق الحصينة في مثل هذه الدراسات، لان هذا النوع من العلاقات لا تخلو من الشواذ في بياناتها و الذي يؤثر بصورة مباشرة وغير مباشرة في نتيجة العلاقة.

المصادر

[1] الراوي، محمود خاشع، (1987)، المدخل الى تحليل الانحدار، وزارة التعليم العالي و البحث العلمي، جامعة الموصل، الموصل، ص. 65، ص. 127 ، ص ص 171-185.

[2] الطالب، بشار عبد العزيز،(1997)، *"مقارنة البرمجة الهدفية مع المربعات الصغرى و انحدار الحرف في تقدير المعلمات،"* رسالة ماجستير في الاحصاء(رسالة غير منشورة)، كلية الادارة والاقتصاد، جامعة الموصل، الموصل، ص 12، ص ص. 14-15 ، ص ص 41-43.

[3] Frank R. Hampel Elvezio M. Ronchetti , Peter J. Rousseeuw and Werner A. Stahel, (2005), Robust Statistics: The Approach Based on Influence Functions,John Wiley & Sons, Inc. pp.12-56

[4] Gungor, S. and Richard L. (2010). Testing linear factor pricing models in large cross-sections: a distribution-free approach, USA : Atlanta, pp. 4-35.

[5] Curran, P. J. , West S. G. and Finch J. F. (1996). "The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis," American Psychological Association, Inc. 1, No. l, pp. 16-29.

[6] Wiley (October 1987), Robust Regression and Outlier Detection (Wiley Series in Probability and Statistics). Retrieved January 28, 2011, from http://ebookee.org/Robust-Regression-and-Outlier-Detection-Wiley-Series-in-Probability-and-Statistics-_434180.html .

[7] Keller G. and Brian Warrack, (2000), Statistics For Management and Economics, 5th edition, Duxbury, Thomson Learning, U.S.A. pp. 119-123, pp. 120-1230 p p. 669-670.

[8] Neter Michael H. Kutner, Christopher J. Nachtsheim, John Neter, and William Li,(2004), Applied Linear Statistical Model, McGraw Hill, pp. 250-482.

[9] Rousscuw, Peter J. and Leroy, A., (2005), Robust Regression and Outlier Detection, John Wily and Sons.

[10] Rousscuw, Peter J. and Van Driessen K., (1999), *" A Fast Algorithm for the Minimum Covariance Determinant Estimator,"* Technometrics, 41, p20.

[11] Zaman, A., Rousseeuw, P.J., Orhan, M. (2001),"*Econometric applications of high-breakdown robust regression techniques*", EconometricsLetters, 71, pp1-8.

[12] Statgraphics Plus for Windows (2005). *Staticgraphic Users Guide*. Version 4, Microsoft Corp.

[13] Stigler, S. M.,(1973), *"Simon New Comb, Percy Daniel and the History of Robust Estimation (1885-1920),"* Jour. Amer. Stat. Assoc., 68, pp. 872-879.

[14] T. W. Anderson and J. D. Finn, (1996), The New Statistical Analysis f Data. Springer-Verlag , New York, Inc. pp. 123-127.

[15] Lawrence Erlbaum, (2002) Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 3rd Edition, pp. 25-64.