|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **كلية الادارة والاقتصاد** | | | | College Name |
| **احصاء** | | | | Department |
| **لقاء علي محمد العلوي** | | | | Full Name as written in Passport |
|  | | | | e-mail |
| **Professor** | **Assistant Professor** | **Lecturer** | **Assistant Lecturer** | Career |
| PhD | | Master | |  |
| **مقارنة مقدرات التباين المشترك الحصينة في تحليل المركبات الرئيسة** | | | | Thesis Title |
| **1424 هـ بغـــــــــداد 2003 م** | | | | Year |
| **عندما نواجه بيانات بأبعاد عالية، غالبا ما نستخدم تحليل المركبات الرئيسة Principal Component Analysis (PCA) لتقليص تلك الابعاد، والذي يستخدم لتحليل عدد كبير من المتغيرات المرتبطة بهدف اختصارها الى عدد اقل من المركبات (العوامل) غير المرتبطة يعزى لها تباين تلك المتغيرات، وهي تتضمن حساب القيم الذاتية (Eigen Values) لمصفوفة التبايــن المشترك (Covariance Matrix) او مصفوفة الارتباط (Correlation Matrix) مع ان هذا التحليل يكون حساس بتطرف للمشاهدات الشاذة (Outliers) وبالتالي فأن النتائج المسحوبة لتحليل المركبات الرئيسة (PCA) سوف تكون مظللة.**  **بما ان المعدل التقليدي ومصفوفة التباين المشترك هي المفتاح المقوم لكل الطرائق الاحصائية في متعدد المتغيرات والتي تكون حساسة جدا للشواذ في البيانات لهذا اصبح تغيير المعدل التقليدي ومصفوفة التباين المشترك بالمقدرات الحصينة (Robust Estimator) هي ربما الاوسع شيوعا والاكثر حصانة في الاستخدام.**  **لاي اختيار لمقدر مصفوفة التباين المشترك الحصينة (أي طريقة تقدير لتلك المصفوفة) يقود الى تحليل آخر للمركبات الرئيسة والذي يمكننا من خلال حساب قيمها الذاتية قياس كمية المعلومات الموضحة بواسطة تلك المركبات.**  **ان وجود قيم شاذة ضمن مجموعة البيانات تؤثر بشكل كبير في نتائج التحليل الاحصائي لتلك البيانات ويظهر هذا في حقيقة ان اغلب الطرائق الاحصائية تؤكد على افتراض ان البيانات المستخدمة في التحليل هي بيانات متجانسة (Homogeneous Data) أي ان كل نقاط البيانات الموجودة في مجموعة معينة تتبع التوزيع المفترض نفسه.**  **في هذا البحث قمنا باشتقاق عدد من المقدرات الحصينة لمصفوفة التباين المشترك مع القيم والمتجهات الذاتية التابعة لها علما ان هذه المقدرات اشتقت في السنوات 1997 – 2002، فبداية العمل كانت مع مقدر القطع البيضوي الاصغر Minimum Volume Ellipsoid Estimator (MVE) حيث قدمت عدة خوارزميات لحساب هذا المقدر منذ عام (1985) حتى عام (2002) لكن بعضها يقود الى تعريف جديد لتلك المقدرات ويبقى مصرا على خصائص الحصانة (نقطة الانهيار العالية والكفاءة الاحصائية مضافا لها خاصية تساوي التغاير) والتي تجعل من تلك المقدرات جذابة اكثر ولكن يبقى شرط العمل بها هو التعامل مع احجام عينات صغيرة.**  **المقدر الثاني الذي اشتق ضمن هذا البحث هو مقدر اصغر محددة تباين مشترك معاد الاوزان Reweighted Minimum Covariance Determent Estimator (RMCD) حيث ان خطوة اعادة الوزن لهذا المقدر تقــود الى كفاءة احصائيــة افضل تزيــد مــن جاذبية مقـــدر (MCD). بعـــد ذلك تـــم اشتقاق مقــدر القطع البيضـــوي الاصغــر المتوسط (Averaged (MVE)) Averaged Minimum Volume Ellipsoid Estimator حيث ان مفتاح الفكرة هنا هو في اخذ المتوسطات لعدة محاولات قريبة الى الامثلية بدلا من اخذ محاولة واحدة مع قيمة دنيا لدالة الهدف (Object Function) والمقدر الناتج يحافظ على خصائص الحصانة مع امكانية التعامل مع جميع احجام العينات والتي تجعل منه مقدر مغري بالفعل.**  **رابع مقدر اشتق في هذا البحث هو مقدر (S) الحصين والذي يعد من اهم المقدرات الحصينة للموقع والتشتت حيث انها تعود الى اصناف تقديرات الانهيار العالية لمواقع متعددة المتغيرات والتشتت.**  **المقدرات الحصينة اعلاه تختلف في طريقة حسابها ولكنها تشترك بخاصية واحدة وهي انها جميعا تتعامل مع بيانات يكون فيها عدد المشاهدات (n) اكبر من عدد المتغيرات (p).**  **حالة غالبا ما نواجهها في الواقع العملي وهي العمل مع بيانات يكون فيها عدد المتغيرات اكبر من عدد المشاهدات (p > n) لذا كان هناك ضرورة ايجاد مقدر يستطيع التعامل مع هكذا بيانات.**  **مقدر Reflection – Based Algorithm for Principal Components Analysis (RAPCA) وجد للتعامل مع بيانات يزيد فيها عدد المتغيرات على عدد المشاهدات حيث انه يثبت كفاءة في التحليل ولكن تبقى نقطة الضعف فيه انه يضحي بنسبة من المعلومات من خلال الغاء عدد من المتغيرات. المقدر المقترح جاء ليتجاوز هذا الضعف في خوارزمية مقدر (RAPCA) حيث انه اثبت جدارة في التحليل مع التزامه بشروط الحصانة المقدمة من قبل بقية المقدرات.**  **سلوك جميع المقدرات اعلاه اتضح من خلال دراسة المحاكاة التي اجريت ضمن هذا البحث فضلا عن مجموعة بيانات حقيقية (واقعية)، والتي بينت ان كل المقدرات المستخدمة في هذا البحث تجمع بين الكفاءة العالية وخصائص الحصانة.**  **كون هذا البحث من اربعة فصول، ضم الفصل الاول المقدمة وهدف البحث فضلا عن الخلفية التاريخية، بينما عرض الفصل الثاني بعض الطرائق الحصينة المستخدمة لتقدير الموقع والتشتت في متعدد المتغيرات. الفصل الثالث المتضمن الاجراءات التطبيقية، ضم اولا (14) حالة لمجاميع البيانات ضمن دراسة المحاكاة ثم التعامل ثانيا مع مجموعة بيانات حقيقية.**  **الفصل الرابع قدم اهم الاستنتاجات والتوصيات التي توصلنا اليها.** | | | | Abstract |