



التحليل الإحصائي متعدد المتغيرات باستخدام حزم ساس (SAS)

تأليف

أ. د. هشام مصطفى النشواتي أ. د. خلف سلمان سلطان

أ. د. عبد الحميد عبد الله الزيد

قسم الإحصاء وبحوث العمليات – كلية العلوم

دار جامعة
الملك سعود للنشر
KING SAUD UNIVERSITY PRESS



ص. ب ٦٨٩٥٣ - الرياض ١١٥٣٧ المملكة العربية السعودية

ح دار جامعة الملك سعود للنشر، ١٤٣٨هـ (٢٠١٧م)

الطبعة الأولى: ١٤٣٢هـ (٢٠١١م)

الطبعة الثانية: ١٤٣٨هـ (٢٠١٧م)

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

النشواتي، هشام مصطفى

التحليل الإحصائي متعدد المتغيرات باستخدام ساس (SAS). / هشام مصطفى النشواتي؛

خلف سلمان سلطان؛ عبد الحميد عبد الله الزيد- ط٢- الرياض، ١٤٣٧هـ

٣٩٣ص؛ ١٧×٢٤سم

ردمك: ٦-٥٢٠-٥٠٧-٦٠٣-٩٧٨

١- التحليل الإحصائي ٢- البرمجة (حواسيب) أ. سلطان، خلف سلمان (مؤلف مشارك)

ب. الزيد، عبد الحميد عبد الله (مؤلف مشارك) ج. العنوان

١٤٣٧/٩٧٤٦

ديوي ٥، ٥١٩

رقم الإيداع: ١٤٣٧/٩٧٤٦

ردمك: ٦-٥٢٠-٥٠٧-٦٠٣-٩٧٨

نشر هذا الكتاب بناء على موافقة المجلس العلمي في اجتماعه التاسع عشر للعام الدراسي ١٤٣٦/١٤٣٧هـ المعقود بتاريخ ١٩/٨/١٤٣٧هـ الموافق ٢٦/٥/٢٠١٦م، بعد استيفائه شروط التحكيم العلمي بالجامعة.

جميع حقوق النشر محفوظة. لا يسمح بإعادة نشر أي جزء من الكتاب بأي شكل وبأي وسيلة سواء كانت إلكترونية أو آلية بما في ذلك التصوير والتسجيل أو الإدخال في أي نظام حفظ معلومات أو استعادتها بدون الحصول على موافقة كتابية من دار جامعة الملك سعود للنشر.

دار جامعة
الملك سعود للنشر
KING SAUD UNIVERSITY PRESS



مقدمة الطبعة الثانية

الحمد لله رب العالمين وسلام على عباده الذين اصطفى. تم اصدار الطبعة الاولى من هذا الكتاب في عام ١٤٣٢ وهو كتاب مقرر لمادة ٤٣٨ احص (طرق احصائية متعدد المتغيرات) لدرجة البكالوريوس.

يتميز عصرنا الراهن (عصر الانترنت والاقتصاد الرقمي والحواسيب المتطورة في التخزين والسرعة) بالسيل العظيم والانتشار واسع النطاق للبيانات حتى أضحى من المستحيل على المحللين استخلاص معلومات ذات معنى باللجوء فقط إلى المداخل التقليدية للتحليل التمهيدي للبيانات.

مع وجود كميات كبيرة من البيانات المخزنة في قواعد البيانات ومخازن البيانات ازدادت الحاجة إلى تطوير أدوات تمتاز بالقوة مثل التحليل الاحصائي لبيانات متعددة المتغيرات وخاصة طرقها وتقنياتها وتطبيقاتها في مجموعة متنوعة من التخصصات مثل الزراعة، والطب والأثروبولوجيا، والعلوم البيولوجية، والأعمال التجارية، والكيمياء، والاقتصاد القياسي، والتعليم، والهندسة، والتسويق، والطب، وعلم النفس، ومراقبة الجودة، وعلم الاجتماع والمناخ...إلخ.

ولهذا ظهرت علوم تعتمد الادوات او الطرق العلمية للتحليل الاحصائي متعدد المتغيرات للتنبؤ واتخاذ القرار كعلم البيانات Data Science وعلم تنقيب في البيانات Data Mining وهي تقنيات تهدف إلى استخراج المعرفة من كميات هائلة من البيانات ، وكذلك الجمع ما بين الإحصاء و تكنولوجيات الإعلام (قواعد البيانات)، الذكاء الاصطناعي، التعلم الآلي Machine Learning.

نحن نعتذر عن الأخطاء التي تم العثور عليها في الطبعة الأولى ، ونحن ممتنون للعديد من الذين وجدوا تلك الأخطاء وارسلوا لنا للتصويب. وبذلنا قصارى جهدنا لتجنب الأخطاء في هذه الطبعة الجديدة، لكن يبقى دائما الكمال لله. كما نود الإشارة الى ان العديد من طلاب البكالوريوس وطلاب الدراسات العليا والباحثين قدموا باقتراحات مفيدة ونحن نشكرهم على ذلك ، وفي النهاية نشكر جامعة الملك سعود وكلية العلوم وقسم الاحصاء وبحوث العمليات على الدعم والتشجيع الدائم.

هذا ونسال الله التوفيق لنا وللقارىء ، والحمد لله رب العالمين.

المؤلفون

المقدمة

الحمد لله رب العالمين والصلاة والسلام على أشرف المرسلين الذي بعثه الله رحمة للعالمين
وبعد:

نضع بين يدي القارئ العربي هذا الكتاب الذي أسميناه "التحليل الإحصائي متعدد المتغيرات باستخدام حزم ساس SAS" وهو من الكتب النادر في المكتبة العربية عن هذا الموضوع. يحتوي هذا الكتاب على الطرق الأساسية التي تساعد في التحليل الإحصائي متعدد المتغيرات والتي يدرس معظمها لطلبة بعض التخصصات في كلية العلوم وكلية العلوم الإدارية بجامعة الملك سعود. وقد تم استخدام حزم ساس لمعالجة هذه الطرق عددياً وبرمجياً. وقد روعي عرض وتقديم موضوعاته بأسلوب سهل وميسر ومتناسب في الخلفية العلمية لطلبة هذه التخصصات ومن في حكمهم ممن يهتمون في مجالات الإحصاء التطبيقي والنظري وبشكل خاص في مجالات الإحصاء الحيوي وعلم الأوبئة والبيئة والاقتصاد والزراعة والاقتصاد. وقد وزعت الموضوعات الرئيسة لهذا الكتاب على تسعة فصول وملاحق. كما يحتوي هذا الكتاب على تسع فصول بالإضافة إلى هذا الفصل (المقدمة). حيث المقدمة تحوي على عرض مختصر عن حزم SAS/IML وتوضيح كيفية التعامل مع هذه الحزم عن طريق النوافذ أو البرمجة.

في الفصل الثاني تم عرض ومناقشة بعض خواص المصفوفات التي تستخدم بكثرة في طرق التحليل الإحصائي متعدد المتغيرات ، حيث تم معالجة هذه الخواص باستخدام حزم SAS/IML وتعتبر هذه الحزم من أقوى الحزم الإحصائية التي تعالج المصفوفات.

وفي الفصل الثالث تمت دراسة التوزيع الطبيعي متعدد المتغيرات والتطرق لخواصه المختلفة باستخدام حزم SAS/IML. أما في الفصل الرابع فقد تمت فيه دراسة مناطق الثقة واختبارات الفرضيات حول متجهات معالم المجتمعات قيد الدراسة.

في الفصل الخامس تمت دراسة تحليل التباين متعدد المتغيرات في اتجاه واتجاهين مع عرض بعض الأمثلة التطبيقية. وفي الفصل السادس تمت دراسة وتحليل المركبات الأساسية وتطبيقاتها مع عرض أمثلة تطبيقية. وفي الفصل السابع تمت دراسة طرق التصنيف والتمييز وتطبيقاتها في مجالات متعددة مع إعطاء أمثلة كافية. أما في الفصل الثامن فقد تم فيه تحليل ودراسة خوارزميات طرق التجميع العنقدة وتطبيقاتها في مجال الوراثة والإحصاء الحيوي المعلوماتي. وأخيرا في الفصل التاسع تم استخدام الانحدار اللوجستيكي لنمذجة بيانات تحتوي على متغير استجابة وصفي ثنائي.

ولم يغفل الكتاب عن تقديم مجموعة متنوعة ومتكاملة من التمارين التي تغطي كافة الطرق الإحصائية المدروسة في التحليل الإحصائي متعدد المتغيرات ، جاء بعضها في مكانه المناسب داخل فصول الكتاب بغرض توضيح الأفكار والمفاهيم المطروحة وجاء كثير منها في نهاية الفصول بغرض الاستزادة والتعمق والتنوع.

ومع أننا بذلنا جهدا غير قليل لإخراج هذا الكتاب في وضعه الحالي فقد لا يكون هذا الوضع هو الأمثل. ولذا فانا نأمل من القراء أو مستخدمي هذا الكتاب أن يوافونا مشكورين بملاحظاتهم ومقترحاتهم التي قد تساهم في تحسين مضمون هذا الكتاب.

وبعد فإننا نأمل أن يكون قد قدمنا بعملمنا هذا فائدة مرجوة لمن يقرأ ويفهم العربية
كما أننا ندعو الله على أن يجعل هذا العمل في صحف حسناتنا يوم نلقاه. والله ولي
التوفيق.

المؤلفون

المحتويات

هـ مقدمة الطبعة الثانية

ز المقدمة

الفصل الأول: مدخل إلى التحليل الإحصائي متعدد المتغيرات

١ (١،١) مقدمة

٣ (١،٢) مقدمة عن حزم SAS/IML, SAS

الفصل الثاني: المصفوفات

٢٩ (٢،١) مقدمة

٢٩ (٢،٢) العمليات الأساسية على المصفوفات

٥٤ (٢،٣) بعض العمليات والخواص على المصفوفات

٦٠ (٢،٤) المتجهات العشوائية

٧٥ (٢،٥) تمارين

الفصل الثالث: التوزيع الطبيعي متعدد المتغيرات

- ٨١ (٣،١) مقدمة
- ٨٢ (٣،٢) التوزيع الطبيعي بمتغير واحد
- ٨٤ (٣،٣) التوزيع الطبيعي متعدد المتغيرات
- ٨٦ (٣،٤) خواص التوزيع الطبيعي متعدد المتغيرات
- ١٠٧ (٣،٥) تمارين

الفصل الرابع : الاستدلال الإحصائي حول متجه متوسطات

لمجتمع طبيعي متعدد المتغيرات

- ١٠٩ (٤،١) مقدمة
- ١٠٩ (٤،٢) اختبار فرضيات حول المتوسط - اختبار T^2 هوتلنج
- ١١٥ (٤،٣) مناطق الثقة
- ١٢١ (٤،٤) منطقة الثقة لعبارة خطية في المتوسطات (فترات ثقة متزامنة)
- ١٢٥ (٤،٥) فترات ثقة بنفروني
- ١٢٦ (٤،٦) المقارنات المزدوجة
- ١٣٣ (٤،٧) مقارنة بين متجهي متوسطي مجتمعين
- ١٣٦ (٤،٨) تمارين

الفصل الخامس: تحليل التباين متعدد المتغيرات

- ١٤٧ (٥،١) مقدمة
- ١٤٧ (٥،٢) تحليل التباين في اتجاه واحد بمتغير واحد
- ١٥١ (٥،٣) تحليل التباين في اتجاه واحد متعدد المتغيرات
- ١٦٢ (٥،٤) تحليل التباين في اتجاهين متغير واحد
- ١٦٥ (٥،٥) تحليل التباين في اتجاهين متعدد المتغيرات
- ١٧٨ (٥،٦) تحليل التباين في اتجاهين متعدد المتغيرات
- ١٨٥ (٥،٧) تمارين

الفصل السادس : تحليل المركبات الأساسية

- ١٩٣ (٦،١) مقدمة
- ١٩٣ (٦،٢) حساب المركبات الأساسية
- ١٩٧ (٦،٣) خواص المركبات الأساسية
- ٢٠٨ (٦،٤) التحليل العاملي
- ٢٠٩ (٦،٥) النموذج العاملي المتعامد
- ٢١٢ (٦،٦) طرق التقدير للتحليل العاملي

٢٢٢تمارين (٦،٧)

الفصل السابع : التمييز والتصنيف

٢٢٥مقدمة (٧،١)

٢٢٩ الاحتمال الكلي لأخطاء التصنيف (٧،٢)

٢٣٠ القيمة المتوقعة لتكلفة خطأ التصنيف (٧،٣)

٢٣٤ قاعدة التصنيف لمجتمعين لهما نفس التباين (والتغايرات) (٧،٤)

٢٣٩ قاعدة التصنيف المربعة لمجتمعين لهما تغاير مختلف أي $\Sigma_1 \neq \Sigma_2$ (٧،٥)

٢٤٢ قاعدة التصنيف بأسلوب بيز (٧،٦)

٢٤٢ دالة فيشر للتمييز بين عدة مجموعات والرسم البياني لها (٧،٧)

٢٤٦ معايير كفاءة دوال التصنيف (٧،٨)

٢٥١ اختيار أفضل نموذج في التمييز والتصنيف (٧،٩)

٢٥١ تمارين (٧،١٠)

الفصل الثامن: طرق التجميع

- (٨،١) مقدمة ٢٥٧
- (٨،٢) مقاييس التماثل وعدم التماثل ٢٥٧
- (٨،٣) طرق التجميع المتدرجة ٢٦١
- (٨،٤) طرق التجميع غير المتدرجة (طريقة k متوسط) ٢٨٠
- (٨،٥) تمارين ٢٨٧

الفصل التاسع: الانحدار اللوجستيكي

- (٩،١) مقدمة ٢٩١
- (٩،٢) متغير الاستجابة Y كمتغير مؤشر ٢٩٢
- (٩،٣) نموذج الانحدار الخطي اللوجستيكي الثنائي لمتغير الاستجابة ٢٩٣
- (٩،٤) مشكلات خاصة بنموذج الانحدار الخطي بمتغير استجابة ثنائي ٢٩٣
- (٩،٥) تقدير المعالم لنموذج الانحدار اللوجستيكي (اللوغاريتمي) ٣٠٦
- (٩،٦) اختبار الفرضيات ٣٠٧
- (٩،٧) معامل ونسبة الترجيح أو التفضيل لكشف العوامل المؤثرة وتعريف لوجيت ٣١٠ ٣١٠
- (٩،٨) بعض المعايير المهمة التي تعتمد على جدول التصنيف التوافقي ٣١٩
- (٩،٩) نموذج الانحدار اللوجستيكي (اللوغاريتمي) بالاعتماد على الأوزان ... ٣٢٣

٢٢٣ تقدير β_1, β_0 بطريقة المربعات الصغرى المرجحة أو الموزونة
٣٣٢ مقدمة عن الانحدار اللوجستيكي المتعدد
٣٣٥ الانحدار اللوجستيكي المتعدد لبيانات طولية
٣٤٩ تمارين
٣٥٩ ملحق (١)
٣٦٣ المراجع
٣٦٣ المراجع العربية
٣٦٣ المراجع الأجنبية
٣٦٧ ثبت المصطلحات
٣٦٧ أولاً: عربي - إنجليزي
٣٧٧ ثانياً: إنجليزي - عربي
٣٨٧ كشف الموضوعات