التحليل الإحصائي باستخدام لغةR

في ظل الثورة المعلوماتية الهائلة التي نعيشها، ومع تزايد عدد البيانات الضخمة التي تحتاج المؤسسات إلى تحليلها والتعامل معها، سواء في مجالات البحث والعلوم أو في ميادين الاقتصاد أو غيرهما. تبرز الحاجة إلى تقنيات وأدوات تساعد في عملية تحليل البيانات وتسهل التعامل معها. ومن هنا ظهرت العديد من البرامج الإحصائية المتطورة والتي تستهدف مساعدة الإحصائيين أو الباحثين في الوصول إلى أدق النتائج بسرعة وسهولة. ومن أهم الآليات التي تصاعدت في الآونة الأخيرة والتي أضحت تستعمل بكثافة في التحليل الإحصائي في البحث Rلغة البرمجة وهي التي نتناولها في هذا المقال. .العلمي

ماهي لغة البرمجة R؟

تعتبر لغة البرمجة R من اللغات التي ذاعت حديثاً، حيث يكثر الكلام عليها مؤخراً في أوساط المهتمين بالبرمجة العلمية والإحصاء، خاصة في ميادين الصحة العامة والمعلوماتية الحيوية (Bioinformatics) والبيولوجيا الجزيئية (Molecular Biology)وغيرهم من العلوم الحديثة نسبياً. ففي السنوات القليلة الماضية أصبحت لغة البرمجة R واحدة من أهم الوسائل والأدوات التي يعتمد عليها الباحثون في نشر البحوث العلمية في المجلات الدولية، فقد أضحت لغة البرمجة R لغة معتمدة في العديد من الجامعات المرموقة والمراكز البحثية العلمية. وتعد لغة البرمجة R لغة حديثة نسبياً فقد تم إنشاؤها في جامعة أوكلاند في نيوزيلندا على يد كلٍ من: وروبيرت جينتلمان، وروس ايهاكا عام ١٩٩٥م. وكما ترى

فاسم كل منهما يبدأ بحرف R ومنه اشتق اسم اللغة، وقد ظهرت النسخة الأولية للغة البرمجة R في عام ١٩٩٥م قبل أن تصدر أول نسخة بيتا بعد ذلك بخمس سنوات. وتعد لغة البرمجة R من اللغات مفتوحة المصدر مما يتيح للمبرمجين الاطلاع على الشفرة البرمجية للبرامج، وتحديثها أو تعديلها وإضافة مزايا جديدة لها.

توفر شركة دراسة للباحثين خدمة التحليل الإحصائي ومناقشة النتائج من خلال البرنامج SPSS لضمان خروج الأبحاث بكفاءة ودقة.

مميزات لغة R وكيفية استخدامها في التحليل الإحصائي

ما سبق وأشرنا فإن لغة البرمجة R تعتبر من لغات البرمجة المشهورة مفتوحة المصدر، مما يعني توفر العديد من المصادر والدورات التدريبية والدروس التعليمية والكتب الإلكترونية، التي يسهل إتقانها على الراغبين. فهي تتميز بالسهولة والبساطة وعدم التعقيد، مما أكسبها العديد من المزايا حتى على حساب البرامج الإحصائية المشهورة في أوساط البحث العلمي مثل: برنامجي SAS، وSPSS. فلغة البرمجة R تمتاز بتوافر العديد من الخوارزميات والأساليب الحديثة، مع كونها متاحة مجاناً على عكس البرامج الأخرى غالية الثمن. مما دفع العديد من الباحثين وطلاب الدراسات العليا إلى الاعتماد عليها بدلاً من البرامج الجاهزة المستخدمة في البرمجة الرياضية العلمية والتحليل الإحصائي. وبالإضافة إلى ذلك فإن البرامج الأخرى تفتقر إلى التحديث باستمرار، والذي عادة ما يتسم بالبطء؛ إذ أن إضافة التحديثات غالباً ما تخضع لدورات تجارية تتحكم بها الشركات المنتجة. وتحتوي لغة R على عدد من أنواع البيانات وهي:

١. بيانات منطقية (Logical): وتضم متغيري الصح والخطأ (true, false) المعروفة في سائر لغات البرمجة.

٢. البيانات الرقمية (Numeric): ويُمكن أن تشمل أي رقم مهما بلغت سعته وبوجود أي عدد من المراتب العشرية.
٣. الأعداد الصحيحة (integer) وتشمل الأعداد الصحيحة.
٤. المركبة (complex): وتشتمل على الأعداد الحقيقية.
٥. الرموز (character): وتشتمل على كافة الرموز، ويتم تخزين الرموز على هيئتها دون اعتبار حقيقة محتواها كالأرقام وعلامات الجمع والطرح كما في لغات البرمجة الأخرى.
٦. الخام (row): وتقوم بتخزين الرموز بالارقام من نظام العد الست عشري.

وتعتبر لغة البرمجة آر Rبيئة عمل برمجية تتيح للمستخدمين بناء البرامج الإحصائية وإجراء التطبيقات الإحصائية، ويمكن للمبرمج إدخال البيانات إليها كأشعة أو مصفوفات أو سلاسل زمنية، كما ان من الممكن أن تكون رسومات بيانية أو توابع. كما تتميز لغة R بأنها تعتمد على واجهة مكونة من لغة HTML مما يساهم في تنفيذ الأوامر البرمجية بسلاسة ويسر.
قد أشرنا في أكثر من مقال سابق إلى العديد من التقنيات والبرامج المستخدمة في التحليل الإحصائي في البحث العلمي مثل: التحليل الإحصائي باستخدام برنامج SPSS، وتحليل البيانات باستخدام برنامج SAS ، وبرنامج AMOSوغيرهم من التقنيات التي تساهم في رفع كفاءة تحليل البيانات وزيادة دقة النتائج. واستكمالاً لما سبق يأتي هذا المقال لنعرض فيه تقنية مختلفة تساعد الباحثين في التعامل مع البيانات بكفاءة من خلال لغة البرمجة R.ونرجو أن يكون هذا المقال وافياً بمقصودة نافعاً لقارئه في توجيه الأنظار إلى أهمية لغة البرمجة Rومميزاتها.