



دورة لغة R لتحليل البيانات: RStudio وحزم tidyverse والرسوم البيانية ggplot2 وتقارير R Markdown



دورة لغة R لتحليل البيانات: RStudio وحزم tidyverse والرسوم البيانية ggplot2 وتقارير R Markdown

مقدمة:

دورة لغة R لتحليل البيانات باستخدام RStudio وحزم tidyverse و ggplot2 و R Markdown دورة تدريبية مدتها خمسة أيام لفرق التحليل والإحصاء والبحث، وتنتهي بإعداد تقرير تحليلي قابل لإعادة الإنتاج بلغة R لمجموعة بيانات محل الدراسة. ما زالت وحدات كثيرة تنقل الأرقام يدويا بين جداول البيانات وبرامج الإحصاء القائمة على القوائم، فلا يمكن إعادة تشغيل الجداول، ولا يستطيع المراجعون تتبع طريقة الوصول إلى النتيجة. يعد المرشحون فعلا جداول إحصائية أو يجرون تحليلات أساسية أو يكتبون نتائج بحثية، وتجري كل جلسة في مختبر تطبيقي على RStudio. وتقدم كور كونسبت هذه الدورة في البرمجة بلغة R.

أهداف الدورة:

- إعداد مشاريع RStudio وتشبيت حزم CRAN وكتابة نصوص برمجية منظمة بلغة R تعتمد على المتجهات والعوامل والقوائم وإطارات البيانات
- استيراد ملفات CSV و Excel بحزمتي readr و readxl، ثم إعادة تشكيلها وتلخيصها بأفعال dplyr و tidyr
- بناء رسوم بيانية متعددة الطبقات بحزمة ggplot2 تناسب السؤال التحليلي وتصديرها بجودة النشر
- إجراء الإحصاء الوصفي واختبارات t ومربع كاي وتحليل التباين ANOVA في R وقراءة المخرجات بدقة
- بناء نماذج الانحدار الخطي واللوجستي وفحص تشخيصاتها وتحويل مخرجاتها إلى جداول مرتبة بحزمة broom
- كتابة دوال R قابلة لإعادة الاستخدام وإصدار تقارير R Markdown يعيد فريق العمل تشغيلها على بيانات جديدة

الفئة المستهدفة:

- فرق التحليل المسؤولة عن الجداول الإحصائية الدورية والمؤشرات ومستخرجات البيانات
- وحدات البحث المسؤولة عن بيانات المسوح والدراسات والتقييم
- فرق المتابعة والتقييم المسؤولة عن تحليل المؤشرات وإعداد تقارير النتائج الدورية
- فرق الإحصاء والتخطيط المسؤولة عن الأرقام الرسمية والإسقاطات وفحص جودة البيانات
- فرق بيانات الجودة والمختبرات المسؤولة عن نتائج القياس ومراجعة الاتجاهات
- فرق ذكاء الأعمال التي تضيف النماذج الإحصائية إلى لودات المعلومات

محاوِر الدورة:

اليوم 1: أساسيات لغة R وبيئة العمل في RStudio

- نوافذ RStudio والمشاريع وتحديد مجلد العمل
- تثبيت حزم CRAN باستخدام `install.packages` library
- المتجهات الذرية في R وتحويل الأنواع والقيم المفقودة NA
- العوامل والقوائم وإطارات البيانات بوصفها حاويات للتحليل
- الفهرسة في R بالأسفاس وعلامة الدولار والاختبارات المنطقية

اليوم 2: استيراد البيانات وقواعد تهيتها بحزم tidyverse

- استيراد ملفات CSV وExcel بحزمتي `readr` و `readxl`
- مبادئ البيانات المرتبة Tidy Data للمتغيرات والمشاهدات والقيم
- أفعال dplyr للتصفية والاختيار والاشتقاق والترتيب
- التجميع والتلخيص بدالتي `group_by` و `summarise` لجداول المؤشرات
- إعادة التشكيل بدالتي `pivot_longer` و `pivot_wider` والدمج بدالة `left_join`

اليوم 3: الرسوم البيانية بحزمة ggplot2 والاختبارات الإحصائية في R

- الخصائص الجمالية والأشكال الهندسية وقواعد الرسم متعدد الطبقات في ggplot2
- التقسيم إلى لوحات والمقاييس والسمات والتصدير بدالة `ggsave`
- الإحصاء الوصفي بدوال `summary` و `quantile` وملخصات `skimr`
- اختبار الفرضيات بدوال `t.test` و `chisq.test` و `aov`
- عرض حجم الأثر وفترة الثقة والقيمة الاحتمالية في R

اليوم 4: نماذج الانحدار والتواريخ والنصوص وجودة الشيفرة

- الانحدار الخطي بدالة `lm` ورسوم البواقي التشخيصية
- الانحدار اللوجستي بدالة `glm` وتفسير نسبة الأرجحية
- جداول مخرجات النماذج بدالتي `tidy` و `glance` من حزمة broom
- تحليل التواريخ بحزمة lubridate وتنظيف النصوص بحزمة stringr
- الدوال المخصصة في R ورسائل الخطأ وتتبع الأخطاء البرمجية

اليوم 5: مختبر RStudio وإعداد تقرير التحليل القابل لإعادة الإنتاج بلغة R

- هيكل تقرير R Markdown ومقاطع الشيفرة والإخراج بحزمة knitr
- مختبر استيراد بيانات الحالة وتنظيفها وتهيئتها
- مختبر الرسوم البيانية بحزمة ggplot2 والاختبار الإحصائي لبيانات الحالة
- نموذج الانحدار لبيانات الحالة وجدول النتائج بحزمة broom
- إخراج تقرير التحليل القابل لإعادة الإنتاج ومراجعة الشيفرة مع النظراء

المهارات التي ستكتسبها:

- البرمجة بلغة R
- استيراد البيانات وتهيئتها
- تصميم الرسوم الإحصائية
- اختيار الاختبار الإحصائي المناسب
- النمذجة بالانحدار
- معالجة التواريخ والنصوص
- كتابة الدوال البرمجية
- إعداد التقارير القابلة لإعادة الإنتاج

لماذا تحضر هذه الدورة:

- تسليم تقرير تحليلي قابل لإعادة الإنتاج بلغة R لمجموعة بيانات محل الدراسة إلى رئيس وحدة البحث أو التحليل والفريق الذي يراجع الأرقام المنشورة
- المفاضلة بين اختبار t ومربع كاي وتحليل التباين ونموذج الانحدار لسؤال بحثي محدد، وتبرير الاختيار من مخرجات R
- تجنب التعديلات غير القابلة للتتبع في جداول البيانات وتعارض أرقام الرسوم مع الجداول وخطوات النسخ واللصق اليدوية التي تؤخر التقارير الدورية
- نقل النصوص البرمجية والدوال المخصصة وسمات ggplot2 وقوالب R Markdown إلى المحللين في الوحدة

الخاتمة:

بعد العودة إلى العمل يسلم المشاركون رئيس وحدة البحث أو التحليل تقريراً تحليلياً قابلاً لإعادة الإنتاج بلغة R، تتولد فيه خطوات الاستيراد والتنظيف والرسوم والاختبارات ومخرجات الانحدار من نص برمجي واحد. ويستخدمه المراجعون للتحقق من طريقة إنتاج كل رقم منشور، وتستخدم الوحدة قالباً لتحديث الجداول الإحصائية الدورية عند وصول بيانات جديدة. وبعد أول إعادة استخدام تراجع الوحدة المدة التي استغرقها التشغيل، ومشكلات البيانات التي كشفها النص البرمجي، والدوال وسمات ggplot2 التي ينبغي اعتمادها معياراً مشتركاً للفريق.

الأسئلة الشائعة:

ما الذي يحتاج المشاركون إلى معرفته قبل دورة لغة R لتحليل البيانات؟

يكفي أن يتعامل المشاركون مع جداول البيانات ويفهم المتوسطات والنسب والمفاهيم الإحصائية الأساسية، ولا تشترط خبرة سابقة في البرمجة. ويساعد تثبيت R و RStudio مسبقاً وإحضار مجموعة بيانات مجهولة الهوية من عمله على تطبيق المختبرات مباشرة.

بم تختلف دورة لغة R لتحليل البيانات عن دورات تحليلات Python أو برامج الإحصاء القائمة على القوائم؟

تعمل الدورة كلياً في R و RStudio مع حزم tidyverse و ggplot2 و R Markdown، وتركز على الاختبارات الإحصائية ومخرجات الانحدار والتقارير القابلة لإعادة الإنتاج. أما دورات Python فتتميل إلى مسارات تعلم الآلة، وتنفذ برامج الإحصاء القائمة على القوائم الاختبارات عبر النوافذ بدلاً من نصوص برمجية يعاد تشغيلها.

لماذا يعتمد المحللون على حزم tidyverse في تحليل البيانات بلغة R؟

تشترك حزم tidyverse في قواعد صياغة موحدة، فيصبح الاستيراد بحزمة readr وإعادة التشكيل بحزمي dplyr و tidyr والرسوم بحزمة ggplot2 مساراً واحداً مقروءاً. ويسهل ذلك مراجعة الشيفرة وتصحيحها وإعادة استخدامها مقارنةً بأوامر base R المتداخلة.

ماذا يعود به المشاركون من دورة لغة R لتحليل البيانات؟

يعود المشاركون بتقرير تحليلي قابل لإعادة الإنتاج بلغة R لمجموعة بيانات محل الدراسة، يضم ملف R Markdown المصدري ونصوص الاستيراد والتنظيف والدوال المخصصة ورسوم ggplot2 ونتائج الاختبارات وجداول الانحدار بحزمة broom، جاهزاً للتكليف مع بيانات وحدته.