

چگونه نرم افزار تحلیل آماری خود را انتخاب کنیم؟

با توجه به تنوع روش‌ها و پیچیدگی محاسبات در تحلیل‌های آماری، استفاده از نرم‌افزارهای رایانه‌ای، گریز ناپذیر است. از طرفی نرم‌افزارهای مختلفی برای انجام چنین تحلیل‌هایی توسعه یافته‌اند. بنابراین انتخاب نرم‌افزار مناسب که مطابق با نیازها و احتیاج ما باشد، امری ضروری به نظر می‌رسد. در این نوشتار سعی شده که با ویژگی‌ها و امکاناتی که نرم‌افزارهای پرفرمدار در این زمینه دارند آشنا شده و آن‌ها را با یکدیگر مقایسه کرده تا شاید پس از مطالعه، دست به انتخاب بهتری بزنید. نرم‌افزارهای انتخابی برای مقایسه در اینجا Minitab، SAS، R و همچنین SPSS هستند که همگی از پرفرمدارترین و به روزترین برنامه‌های محاسبات و تحلیل‌های آماری محسوب می‌شوند.

از لحاظ سابقه، می‌توان SAS را قدیمی‌ترین نرم‌افزار در این حوزه نامید. SAS در سال ۱۹۶۶ در دانشگاه کارولینای شمالی متولد شد و توسعه یافت. چندی بعد، یعنی در سال ۱۹۶۸ شرکت SPSS بسته محاسبات آماری خود را ارائه داد. Minitab نیز یکی از نرم‌افزارهایی است که توسط دانشگاه پسنیلوانیا در سال ۱۹۷۲ معرفی شد. Stata یک برنامه آماری چند منظوره است که اولین نسخه آن در سال ۱۹۸۵ عرضه شد ولی R یکی از جدیدترین زبان‌های محاسبات آماری است که در سال ۱۹۹۲ توسط دانشگاه اوکلند (Auckland) در نیوزیلند مورد بهره‌برداری قرار گرفت. در حقیقت R یک زبان برنامه‌نویسی است که برای انجام محاسبات آماری طراحی گردیده است و فرزند خوبی برای زبان محاسبات آماری S محسوب می‌شود.



Minitab® 18



STATA®

ویژگی‌های اصلی نرم‌افزارهای تحلیل آماری

ویژگی‌های اصلی نرم‌افزارهای آماری را می‌توان به چند گروه تقسیم کرد:

۱. خصوصیات مربوط به زبان توسعه و رابط کاربران
۲. خصوصیات مربوط به امکان ترسیم نمودارها
۳. خصوصیات مربوط به تحلیل واریانس
۴. خصوصیات مربوط به محاسبه انواع روش‌های رگرسیونی
۵. خصوصیات مربوط به تحلیل‌های سری زمانی
۶. خصوصیات مربوط به محاسبات آمار توصیفی و تحلیل‌های اختصاصی

در قسمت اول به بررسی ویژگی‌های گروه اول می‌پردازیم.

۱- خصوصیات زبان توسعه و رابط کاربران

در این قسمت به بررسی ویژگی‌های محیطی اصلی تعدادی از نرم‌افزارهای تحلیل و محاسبات آماری می‌پردازیم. در جدول زیر بعضی از این خصوصیات فهرست و مقایسه شده‌اند.

نام نرم‌افزار	شرکت توسعه دهنده	سیستم عامل	متن‌باز	رابط کاربری	زبان توسعه نرم‌افزار	زبان Scripting
Minitab	Minitab Inc.	ویندوز	خیر	تصویری / خط فرمان	C, C++ Java	---
R	R Foundation	Mac OS/ ویندوز/ لینوکس	بله	تصویری / خط فرمان	C, Fortran, R	R language, Python
SAS	SAS Institute	ویندوز/ لینوکس	خیر	تصویری / خط فرمان	C	SAS language
SPSS	IBM	Mac OS/ ویندوز/ لینوکس	خیر	تصویری / خط فرمان	Java	R, Python, SaxBasic
Stata	StataCorp	Mac OS/ ویندوز/ لینوکس	خیر	تصویری / خط فرمان	C	

همانطور که دیده می‌شود، زبان برنامه‌نویسی آماری R از لحاظ پشتیبانی از سیستم عامل‌های مختلف بهتر عمل کرده است. از طرفی متن‌باز (Open Source) بودن آن نیز به جاذبه‌هایش می‌افزاید.

۲- خصوصیات مربوط به امکان ترسیم نمودارها

رسم نمودارهای توصیفی و تحلیلی یکی از ویژگی‌های اصلی برای نرم‌افزارهای آماری محسوب می‌شود. جدول زیر نیز به مقایسه این امر برای نرم‌افزارهای آماری زیر پرداخته است.

نام نرم‌افزار	نمودار میله‌ای (Bar Chart)	نمودار جعبه‌ای (Boxplot)	نمودار همبستگی (Corrologram)	نمودار هیستوگرام	نمودار خطی (Line Chart)	نمودار پراکندگی (ScatterPlot)
Minitab	بله	بله	بله	بله	بله	بله
R	بله	بله	بله	بله	بله	بله
SAS	بله	بله	بله	بله	بله	بله
SPSS	بله	بله	بله	بله	بله	بله
Stata	بله	بله		بله		بله

همانطور که دیده می‌شود از لحاظ ترسیم نمودارها، این ۵ نرم‌افزار با یکدیگر تفاوتی ندارند. هرچند نمودارهای ترسیمی توسط نرم‌افزار R از دقت و کیفیت بالاتری برخوردار است.

۳- خصوصیات مربوط به تحلیل واریانس

تحلیل واریانس یکی از کاربردی‌ترین روش‌های آماری است که بواسطه آن بسیاری از مقایسه‌ها برای میانگین جامعه یا سنجش مناسب بودن مدل‌های آماری صورت می‌گیرد. در نتیجه وجود چنین امکانی در نرم‌افزارهای محاسبات آماری ضروری است. جدول زیر به بررسی این ویژگی در بین نرم‌افزارها می‌پردازد.

نام نرم‌افزار	آنالیز واریانس یک طرفه-One way ANOVA)	آنالیز واریانس دو طرفه-Two way ANOVA)	آنالیز واریانس چندمتغیره (MANOVA)	مدل‌های خطی تعمیم یافته (GLM)	پس آزمون (Post-hoc)	مربعات لاتین (Latin) (Square
Minitab	بله	بله	بله	بله	بله	بله
R	بله	بله	بله	بله	بله	بله
SAS	بله	بله	بله	بله	بله	بله
SPSS	بله	بله	بله	بله	بله	بله
Stata	بله	بله	بله	بله	بله	بله

با توجه به این جدول، مشخص است که همه نرم‌افزارهای معرفی شده، از امکانات محاسباتی مربوط به تحلیل واریانس برخوردار هستند.

۴- خصوصیات مربوط به محاسبه انواع روش‌های رگرسیونی

در زمینه‌های هوش مصنوعی و مدل سازی، روش‌های برازش مدل‌های خطی، بسیار به کار می‌رود. در نتیجه امکان انجام محاسبات مربوط به روش‌های رگرسیونی، اهمیت زیادی در انتخاب نرم‌افزار مناسب دارد. در جدول زیر به مقایسه نرم‌افزارهای آماری پرتعداد در این زمینه پرداخته‌ایم.

نام نرم‌افزار	رگرسیون خطی (OLS)	رگرسیون وزنی (WLS)	رگرسیون لجستیک	رگرسیون گام به گام	رگرسیون چندکی	رگرسیون کاکس (COX)
Minitab	بله	بله	بله	بله	خیر	خیر
R	بله	بله	بله	بله	بله	بله
SAS	بله	بله	بله	بله	بله	بله
SPSS	بله	بله	بله	بله	بله	بله
Stata	بله	بله	بله	بله	بله	بله

در این بین به نظر می‌رسد که نرم‌افزار Minitab از لحاظ انجام محاسبات رگرسیونی دچار ضعف است.

۵- خصوصیات مربوط به تحلیل‌های سری زمانی

یکی از کاربردهای فنون آماری، پیش‌بینی پدیده‌هایی است که به زمان مرتبط هستند. هرچند این محاسبات بسیار طولانی و پیچیده‌اند ولی نرم‌افزارهای آماری این کارها را به سادگی و سرعت انجام می‌دهند. جدول زیر نیز به مقایسه ویژگی‌های مربوط به تحلیل‌های سری زمانی می‌پردازد.

نام نرم‌افزار	مدل ARIMA	مدل GARCH	مدل GARCH چند متغیره	آزمون‌های ایستایی سری زمانی (unit root test)	خودرگرسیون برداری (Vector Autoregression)	آزمون هم‌افزایی (Co-integration)
Minitab	بله	خیر	خیر	خیر	خیر	خیر
R	بله	بله	بله	بله	بله	بله
SAS	بله	بله	بله	بله	بله	بله
SPSS	بله	بله	خیر	خیر	خیر	خیر
Stata	بله	بله			بله	

مشخص است که هم SPSS و هم Minitab در زمینه تحلیل سری‌های زمانی دچار ضعف هستند.

۶- خصوصیات مربوط به محاسبات آمار توصیفی و تحلیل‌های اختصاصی

همیشه همراه هر گزارش تحلیلی، آمارهای توصیفی و نمودارهای مربوطه نیز باید ارائه شود. بنابراین انجام این گونه محاسبات و نمودارها در همه نرم‌افزارهای آماری ضروری به نظر می‌رسد. این ویژگی‌ها در جدول زیر مقایسه شده‌اند.

نام نرم‌افزار	آمار توصیفی	آماره‌های ناپارامتری (Nonparametric Tests)	کنترل کیفیت آماري (Quality Control)	تحليل بقا (Survival Analysis)	خوشه‌بندی (Clustering)	آنالیز تشخیص (Discriminant Analysis)
Minitab	بله	بله	بله	بله	بله	بله
R	بله	بله	بله	بله	بله	بله
SAS	بله	بله	بله	بله	بله	بله
SPSS	بله	بله	بله	بله	بله	بله
Stata	بله	بله	بله	بله	بله	بله

همانطور که دیده می‌شود از لحاظ ویژگی‌های ذکر شده، این ۵ نرم‌افزار با یکدیگر تفاوتی ندارند ولی دسترسی دستورات کنترل کیفیت آماری در Minitab به مراتب بیشتر از دیگر نرم‌افزارهای آماری است.

منبع:

<https://blog.faradars.org/choosing-best-statistical-packages/>