



STATA vs SPSS ?

بعنوان یک پژوهشگر حرفه ای. کدام نرم افزار آماری برای
من کارایی بالاتری دارد؟

دکتر عباس کشت کار

عضو هیات علمی دانشگاه علوم پزشکی تهران



محصولی از آمارافزار

www.amarafzar.ir

اگر شما نیز از آن دسته از محققینی هستید که علی رغم تسلط کافی به نرم افزار SPSS، متوجه برخی کمبودهای این نرم افزار آماری در انجام تجزیه و تحلیلها یا برآورده کردن نیازهای آماری خود شده اید، این متن را تا انتها مطالعه فرمایید:

مقدمه: اگر شما نیز از کاربران ماهر یا مسلط نرم افزار SPSS هستید، حتما متوجه برخی مشکلات یا محدودیتهای این نرم افزار شده اید. لطفا سناریوهای زیر را با دقت بخوانید:

سناریو (۱): برآورد شیوع یا فراوانی پیامدها

شما قصد دارید داده های یک مطالعه مقطعی سه ساله که با هدف توزیع فراوانی یا شیوع برخی ناهنجاریهای لوله عصبی مادرزادی، طراحی گردیده را تحلیل نمایید. شما قبلا در دوره های آموزشی نرم افزار SPSS یاد گرفته اید که مشکل فقدان مولفه شیوع یا سهم (Proportion) در این نرم افزار را با تعریف کد ۱ برای وضعیت پیامد مثبت و کد صفر، برای وضعیت پیامد منفی، برطرف نموده و خوشبختانه این اقدام را نیز چندین بار برای موارد مشابه انجام داده و مطمئن هستید که برآورد نقطه ای و فاصله اطمینان ۹۵ درصد شیوع همان برآورد نقطه ای و فاصله ای میانگین متغیر مربوطه میباشد. اما در این مطالعه، سه نوع ناهنجاری دارید! اگر به هر یک از انواع ناهنجاریها، کدهای ۱ تا ۳ را اختصاص داده و به موارد سالم، کد صفر بدهیم، مشکل مشابه موارد قبلی حل میگردد؟ مطمئنا خیر. پس چه باید کرد؟

شما با کمی فکر کردن این راه حل به ذهنتان میرسد که برای هر ناهنجاری، یک متغیر تعریف نموده و سپس در هر متغیر، همان راهکار قبلی را بکار گیریم. به نظر شما حتی اگر این راهکار، مشکل را برطرف نماید، آیا فاصله اطمینان ۹۵ درصد برای این پیامد نادر (شیوع حدود ۳,۵ در هزار است) همانند یک پیامد شایع (بعنوان مثال برای بیماری دیابت با شیوع ۱۲ درصد) میتواند معتبر باشد؟

حال به این سناریو دقت فرمایید:

سناریو (۲): محاسبه شاخص کاپای وزنی

در مرکز تحقیقات شما، یکی از اساتید می‌خواهد در یک مطالعه پایلوت، توافق بین نظرات دو سیتولوژیست را بطور مستقل در تشخیص بدخیمی در آسپیراسیون سوزنی تیروئید، ارزیابی نماید. ایشان برای این ارزیابی از شما کمک می‌خواهد. شما با مراجعه به کتاب Epidemiology; Beyond the Basic در می‌یابید که شاخص کمی مناسب برای این شرایط، شاخص کاپا است، لیکن از آنجا که طبقه بندی هر سیتولوژیست برای هر نمونه آسپیراسیون، خوش خیم، مشکوک و بدخیم بوده، فلذا شاخص کاپای ساده نمیتواند بدرستی این توافق را ارزیابی نموده و لازم است، شاخص کاپای وزنی محاسبه گردد.

پس از تعیین شاخص آماری مناسب، سراغ نرم افزار SPSS رفته تا جایگاه محاسبه آنرا پیدا نمایید. متأسفانه متوجه میشوید که امکان محاسبه کاپای وزنی در این نرم افزار وجود ندارد!!!،

به این سناریو نیز دقت فرمایید:

سناریو (۳): رسم نمودار آلتمن-بلات

در یک مطالعه ارزش تشخیصی، قصد دارید اعتبار یک گلوکومتر خاص را در سنجش قند خون بیماران دیابتی، ارزیابی نمایید. بدین منظور از ۵۰ بیمار دیابتی در ابتدای صبح یک بار سنجش قند خون ناشتا (با استفاده از گلوکومتر) و همزمان یکبار نیز نمونه گیری خون وریدی (برای انجام آزمایش استاندارد سنجش قند خون) و ۵۰ بیمار دیابتی نیز سنجش قند خون عصر (مشابه گروه اول یعنی دو بار سنجش که یکبار با گلوکومتر و یکبار نیز با نمونه خون وریدی و سنجش استاندارد قند خون) انجام گرفت. از آنجا که هدف مطالعه ارزیابی روایی است و شما محدودیت ضریب همبستگی را بعنوان شاخص روایی میدانید، میخواهید با استفاده از نمودار آلتمن-بلات و آزمون مربوطه (Pittman test) اعتبار گلوکومتر مورد نظر را تعیین نمایید. متأسفانه در بین نمودارهای نرم افزار SPSS و همچنین منوی آنالیز، هیچگونه اثری از چگونگی رسم این نمودار و آزمون مربوط به آن وجود ندارد!!!!



آیا این سناریو برای شما آشنا بنظر نمی رسد؟

سناریو (۴): انجام آزمون آنالیز واریانس با اندازه گیری مکرر Repeated Measures ANOVA

برای تجزیه و تحلیل داده های یک مطالعه تجربی بر روی ۴ گروه از رتھا، پیامد مورد سنجش که یک پیامد کمی بوده، در مقاطع زمانی ساعت ۶، ۱۲، ۱۸ و ۲۴ پس از بکارگیری مداخله در گروهها، تحت اندازه گیری قرار گرفت. برای تجزیه و تحلیل این داده ها، به نحوی که اثر گروهها و مدت زمان سپری شده از مداخله، بصورت یکجا مورد ارزیابی قرار گیرد، قصد دارید از آزمون Repeated Measures ANOVA استفاده نمایید. در منوی مربوطه در نرم افزار SPSS، متوجه میشوید که ابتدا باید ساختار داده ها را مطابق با شرایط این آزمون تعریف نمایید. پس از این کار، برای رسیدن به اهداف آنالیز (ارزیابی اثر گروهها و اثر مدت زمان سپری شده) در بین گزینه های نسبتا زیاد و متعددی در منوی این آزمون دچار سردرگمی میشوید. بالاخره با پرس و جو و کمک گرفتن از دوستانتان، این آزمون را اجراء میکنید. با مشاهده خروجی این آزمون در فایل مربوطه، دوباره چهره شما در هم میرود! چرا که یافتن نتایج مرتبط با اهداف آنالیز شما در بین حجم زیادی از جداول و متون، کار دشواری بنظر میرسد

همانطور که در این سناریوها ملاحظه نمودید، نقائص نرم افزار SPSS را در انجام آزمونهای آماری و تجزیه و تحلیلها، به دو گروه قابل طبقه بندی است:

- دسته اول آزمونها یا روشهای آماری مختلفی است که از پایه و اساس در SPSS وجود نداشته و یا به عبارت دیگر، این نرم افزار قادر به اجراء یا انجام آن نیست.
- دسته دوم، آزمونها یا روشهایی است که هر چند نرم افزار SPSS قادر به انجام آن است، لیکن بنا به دلایل مختلف از قبیل دشوار بودن مراحل قبل از آنالیز (آماده سازی ساختار داده ها)، وجود گزینه های



STATA vs SPSS ?

متعدد و بیشمار در اجرای آنالیز، حجیم بودن خروجی یا نتیجه آنالیز و ... ، استفاده عملی از این روشها را بسیار محدود مینماید.

نمونه آزمونها یا روشهای آماری که در دسته دوم جای میگیرند، عبارتند از آنالیز واریانس با داده های مکرر، رگرسیون لجستیک، مدل‌های عمومیت یافته خطی (Generalized Linear Models)، محاسبه حساسیت، ویژگی، ارزش اخباری مثبت و منفی و

البته در بسیاری از ارزیابیهای نقادانه متخصصین بر روی نرم افزار SPSS، موضوع حجیم بودن خروجی یا یافته های آنالیز، بعنوان یک مساله مشترک در اکثریت آزمونها گزارش گردیده و بدیهی است که این مشکل در برخی شرایط (مانند مورد سناریو ۴) برجسته میگردد.

حتما تصدیق میفرمایید که یک یا چند مورد از سناریوهای فوق، برای شما یا همکارانتان رخ داده است. پس راه حل اساسی چیست؟؟

راه حل چیست؟

همانطور که حدس زده اید، اگر میخواهیم مشکلات فوق و یا سایر مشکلات و موانع مشابه، بطور اساسی برطرف گردد، باید از یک نرم افزار آماری جامعتر یا قویتر استفاده نماییم.

ممکن است برخی از پژوهشگران، در برابر نظر فوق، اینگونه پاسخ دهند که: «از کجا معلوم که نرم افزار قویتر یا کارآمدتر نیز، مشکلاتی دیگر به همراه نداشته باشد.»



STATA vs SPSS ?

همانطور که همه میدانیم، در دنیای تولید نرم افزارها، عنصر رقابت، یک عامل مهم در گسترش امکانات و تواناییهای نرم افزارهای تولیدی بوده و با گذشت زمان ویرایشهای جدید نرم افزار، علاوه بر اینکه تا حد امکان تواناییهای نرم افزارهای رقیب را دارا هستند، امکانات یا تواناییهای دیگری نیز به آنها اضافه میگردد تا بتوانند هر چند در طول یک بازه زمانی، سرآمد نرم افزارهای کاربردی آن حیطه باشند. همچنین یک عامل مهم در افزودن امکانات و تواناییهای جدید به نرم افزارها، تقاضا یا درخواستهای کاربران حرفه ای یا مشتریان آن نرم افزار در خصوص نقائص آن نرم افزار میباشد.

اگر بخواهیم این دیدگاه را در مقایسه بین SPSS با STATA و یا حتی بین دو نرم افزار آماری، مبنا قرار دهیم، قطعاً به این نتیجه میرسیم که برتری یک نرم افزار نسبت به دیگری فقط محدود به یک بازه زمانی مشخص بوده و در دراز مدت امکانات این دو نرم افزار مشابه یکدیگر است.

به دو دلیل نتیجه گیری فوق در مقایسه بین SPSS و STATA صحیح نمیشود:

۱- با توجه به تخصصی شدن علم آمار کاربردی برای گرایشهای مختلف علوم، حیطه علمی مورد استفاده در نرم افزار SPSS، علوم اجتماعی یا بصورت کلان تر، علوم انسانی و حیطه علمی نرم افزار STATA، علوم زیست پزشکی و گرایشهای وابسته به آن میباشد. پس بدیهی است که درجه انطباق امکانات نرم افزار STATA با نیازهای محققین علوم زیست پزشکی بیشتر از SPSS است.

۲- ماهیت برنامه نویسی در نرم افزار STATA (بر خلاف SPSS) ماهیت کدهای باز (Open Source) بوده و این ویژگی موجب گردیده که اجرای



STATA vs SPSS ?

آزمونها یا روشهای آماری مختلف در قالب یک ماژول جداگانه قابل انجام باشد و این ماژولها میتوانند توسط هر فردی که به زبان برنامه نویسی STATA تسلط یابد، تولید شوند. فلذا در عمل، امکانات و توانمندیهای مختلفی به برنامه STATA از طریق تولید این ماژولهای جدید اضافه گردیده و هر کاربری در جهان، میتواند از طریق شبکه جهانی وب، به این ماژولها دسترسی یافته و با نصب آنها، این امکانات و توانمندیها را کسب نماید. بدیهی است که این فرایند در نرم افزار SPSS وجود نداشته و توسعه امکانات و توانمندیهای نرم افزار فقط از طریق برنامه نویسی جدید (توسط شرکت تولید کننده نرم افزار) و ایجاد ویرایش جدید امکان پذیر میباشد. پس میتوان این چنین نتیجه گیری کرد که:

پژوهشگران علوم زیست پزشکی با یادگیری نرم افزار STATA و استفاده از آن در تجزیه و تحلیل داده ها، علاوه بر اینکه از نرم افزاری با درجه انطباق بیشتر با حیطه پژوهشهای خود که واجد امکانات بیشتری است، استفاده مینمایند، بصورت بالقوه به گنجینه ای از ماژولها یا روشهای آماری مختلف نیز دسترسی داشته و هر زمان که اراده نمایند، میتوانند با نصب آسان این ماژولها، توانایی و امکانات نرم افزار STATA خود را ارتقاء داده و در این محیط، میتوان ادعا نمود که به هیچ وجه، عدم توانایی در تجزیه و تحلیل وجود ندارد. در این شرایط تنها مانع فرا روی پژوهشگران در تجزیه و تحلیل داده ها، ضعف دانش آماری آنها است.



STATA vs SPSS ?

اینجانب با تجربه ۱۳ سال تدریس دوره ها و کارگاههای آموزشی مختلف مرتبط با روش شناسی پژوهش و آمار کاربردی در علوم زیست پزشکی، تا کنون بیش از ۶ دوره آموزش نرم افزار STATA در سطوح مختلف، از مقدماتی تا پیشرفته برگزار نموده و خوشبختانه ضمن موفق بودن ارزشیابی دوره ها توسط شرکت کنندگان، تعداد قابل توجهی از شرکت کنندگان به اندازه ای با این نرم افزار تسلط پیدا نموده اند که کلیه تجزیه و تحلیلهای مطالعات پژوهشی خود را با این نرم افزار انجام داده و موفق گردیده اند که نیازهای آماری خود را که تا قبل از آن با کمک مشاوران آماری و متخصصین برطرف مینمودند، بطور مستقل تامین نمایند.

دوره ها یا کارگاههای آموزشی STATA بگونه ای طراحی گردیده که به هر سه بعد مهم و مورد نیاز دانش آماری پایه، توانایی تفسیر و کاربرد آزمونهای مختلف و مهارتهای عملی کار با نرم افزار در شرکت کنندگان توجه داشته و از طریق تقویت همه جانبه این سه مولفه، اعتماد به نفس پژوهشگران و توانایی آماری آنها را بطور قابل توجهی ارتقاء میدهد.

مشخصات این دوره های آموزشی بشرح ذیل میباشد:

- دوره یا کارگاه STATA مقدماتی: ۲۵ ساعت

آشنایی با محیط نرم افزار STATA و ویژگیهای فنی و ویرایشهای متفاوت آن

بخشهای مختلف نرم افزار و شیوه اجرای آنالیزها (مبتنی بر منو و مبتنی بر تاپ فرامین)

مدیریت داده ها در STATA: تولید متغیرها، محاسبات ریاضی در متغیرها، کدگذاری مجدد و ...

مدیریت خروجی یا نتیجه آنالیزها

مدیریت سینتاکسها یا دستورالعملهای تجزیه و تحلیل، نحوه ذخیره سازی do file و ...



STATA vs SPSS ?

مدیریت تبدیل فایل‌های دیتا از فرمت‌های معمول و رایج (SPSS، Excel، Access و ...) به STATA

آشنایی با شکل ژنریک فرامین و توضیحات کافی در اجزاء مختلف این فرامین

آشنایی با راهنما یا Help نرم افزار و راه‌های مختلف آن

نحوه جستجو و یافتن ماژول‌های آنالیز داده‌ها (ado. Files) و چگونگی نصب و استفاده از آنها

آمار توصیفی و روش‌های آمار توصیفی و کاربرد آنها در آمار (با تاکید ویژه در تدوین مقالات و پایان نامه‌ها)

شاخص‌های مرکزی و پراکندگی و یادگیری هر آن چیزی که در موضوع شاخص‌های مرکزی و پراکندگی برای پژوهشگران مورد نیاز است.

شاخص‌های شکل توزیع آماری (skewness، kurtosis) و شاخص‌های صدک و چندکها و کاربرد آنها در پژوهش‌های زیست پزشکی

آشنایی با موضوع مهم و بنیادی تغییر متغیر (Transformation)، تغییر متغیرهای مهم و چرایی و چگونگی انجام آن

قضیه حد مرکزی و کلیه مباحث کاربردی مورد استفاده در این قضیه در آمار کاربردی

آشنایی با منطق برآورد (Estimation) و برآورد نقطه‌ای و فاصله‌ای برای دو شاخص اصلی میانگین و شیوع

آشنایی با احتمالات و قوانین پایه در آن

آشنایی با توزیع‌های آماری دو جمله‌ای و پواسون و کاربرد این توزیع‌ها در آمار

تولید داده‌هایی با توزیع دو جمله‌ای و پواسون با کمک نرم افزار

آشنایی با توزیع نرمال و قوانین توزیع نرمال و نرمال استاندارد



STATA vs SPSS ?

• دوره یا کارگاه STATA متوسطه: ۵۰ ساعت

• دوره یا کارگاه STATA پیشرفته: ۴۰ ساعت

