



Exploratory Analysis of Longitudinal and Clustered Psychological Data

تحلیل اکتشافی داده‌های روانشناختی طولی و خوشه‌ای

نام و نام خانوادگی دانشجو: مبینا امینی پارسا

رشته: علم داده‌ها

دانشکده: علوم ریاضی

دانشگاه: شهید بهشتی

درس: تحلیل اکتشافی داده‌ها

استاد راهنما: دکتر احسان بهرامی

زمستان ۱۴۰۴

چکیده

هدف این پژوهش، به کارگیری رویکردهای تحلیل اکتشافی داده‌ها برای بررسی الگوهای زمانی و بین فردی در داده‌های طولی و خوشه‌ای است. داده‌های مورد استفاده شامل مشاهدات روزانه ۷۲ فرد در طول ۳۰ روز متوالی (در مجموع ۲۰۷۰ مشاهده) بوده و متغیرهایی نظیر استرس روزانه، راهبردهای مقابله‌ای، عاطفه منفی، استرس ادراک شده، اضطراب صفتی و تاب‌آوری را در بر می‌گیرد. در گام نخست، با استفاده از تحلیل اکتشافی طولی، رفتار متغیرها در طول زمان از طریق نمودارهای مسیر فردی، روندهای میانگین جمعیتی، نوسانات زمانی، و توزیع‌های متغیرها بررسی شد. این مرحله امکان شناسایی الگوهای کلی، تغییرپذیری درون فردی و ناهمگنی بین افراد را فراهم کرد. در ادامه، هر فرد به عنوان یک واحد رفتاری مستقل در نظر گرفته شد و به منظور مقایسه افراد، داده‌های طولی هر فرد به مجموعه‌ای از شاخص‌های خلاصه شده شامل میانگین، انحراف معیار و شیب تغییرات زمانی تبدیل گردید. بر اساس این ویژگی‌ها، تحلیل خوشه‌ای به منظور شناسایی گروه‌های همگن از افراد با الگوهای رفتاری مشابه انجام شد و افراد در سه خوشه متمایز دسته‌بندی شدند.

نتایج نشان داد که خوشه‌های شناسایی شده از نظر سطح کلی استرس، میزان نوسان هیجانی و روند تغییرات زمانی تفاوت‌های معناداری دارند و می‌توان آن‌ها را به عنوان پروفایل‌های رفتاری متمایز تفسیر کرد. این پژوهش نشان می‌دهد که ترکیب تحلیل اکتشافی داده‌های طولی با خوشه‌بندی افراد، چارچوبی کارآمد برای درک الگوهای پیچیده رفتاری در داده‌های تکرارشونده فراهم می‌کند و می‌تواند مبنایی برای تحلیل‌های پیشرفته‌تر در مطالعات آینده باشد.

واژگان کلیدی: تحلیل اکتشافی داده‌ها، داده‌های طولی، داده‌های خوشه‌ای، خوشه‌بندی، پروفایل رفتاری

فهرست مطالب

پیشگفتار

فصل ۱. مفاهیم و تعاریف اساسی و معرفی داده ها

۱.۱. تعریف کوتاه و علمی همه اصطلاحات و روش های استفاده شده

۲.۱. معرفی داده ها

۱.۲.۱. بررسی ویژگی ها در هر ستون

فصل ۲. روش شناسی علمی تحلیل اکتشافی داده های طولی و خوشه ای

۱.۲. هدف کلی تحلیل

۲.۲. بررسی اولیه ساختار داده ها

۳.۲. شناسایی ساختار طولی و خوشه ای

۴.۲. آماره های توصیفی و شناخت اولیه متغیر ها

۵.۲. تحلیل اکتشافی طولی در سطح جمعیت

۶.۲. بررسی تغییرپذیری توزیعی در طول زمان

۷.۲. تحلیل اکتشافی خوشه ای در سطح افراد

۸.۲. استخراج ویژگی های فردی از داده های طولی

۹.۲. تحلیل اکتشافی خوشه ای افراد

۱۰.۲. تعیین تعداد خوشه ها

۱۱.۲. خوشه بندی نهایی افراد

۱۲.۲. تحلیل ترکیبی طولی و خوشه ای

۱۳.۲. چارچوب یا استراتژی تحلیلی پژوهش

فصل ۳. کاربرد و نتایج تحلیل اکتشافی

۱.۳. آماده سازی و پاکسازی داده ها

۲.۳. بررسی توزیع متغیر ها

۳.۳. تحلیل اکتشافی طولی

۱.۳.۳. بررسی مسیرهای زمانی افراد

۲.۳.۳. بررسی روند کلی جمعیت

۳.۳.۳. بررسی تغییرات توزیع در طول زمان

۴.۳.۳. بررسی وابستگی زمانی

۴.۳. بررسی تفاوت های فردی

۵.۳. ساخت شاخص های رفتاری افراد

۶.۳. خوشه بندی افراد

۷.۳. تحلیل الگو های رفتاری خوشه ها

۱.۷.۳. مقایسه سطح متغیرها در خوشه ها

۲.۷.۳. تحلیل مسیر زمانی خوشه ها

۳.۷.۳. بررسی تفاوت افراد داخل خوشه

۴.۷.۳. نمایش کلی رفتار افراد

۵.۷.۳. بررسی رابطه متغیرها

۸.۳. مقایسه ویژگی های خوشه ها

۹.۳. بررسی روابط بین ویژگی ها با شاخص های آن ها

فصل ۴. نتیجه گیری

منابع

پیشگفتار

در مطالعات روان‌شناختی، داده‌های طولی امکان بررسی تغییرات درون‌فردی را در طول زمان فراهم می‌کنند و به پژوهشگران اجازه می‌دهند پویایی متغیرهای رفتاری و هیجانی را با دقت بیشتری مطالعه کنند. هنگامی که چنین داده‌هایی به صورت تکرارشونده از چندین فرد جمع‌آوری می‌شوند، ساختاری خوشه‌ای شکل می‌گیرد که در آن مشاهدات مربوط به هر فرد به یکدیگر وابسته بوده و نمی‌توان آن‌ها را مستقل در نظر گرفت.

در این میان، تحلیل اکتشافی داده‌ها به عنوان گامی اساسی پیش از هرگونه مدل‌سازی آماری، نقش مهمی در درک ساختار داده‌ها، شناسایی الگوهای پنهان، بررسی تغییرپذیری درون‌فردی و بین‌فردی، و کشف روابط اولیه میان متغیرها ایفا می‌کند. استفاده از نمودارها و شاخص‌های توصیفی در این مرحله، امکان مشاهده مستقیم ویژگی‌های داده و ارزیابی کیفیت آن‌ها را فراهم می‌سازد.

هدف اصلی این پروژه، به‌کارگیری رویکردهای تحلیل اکتشافی برای بررسی داده‌های طولی و خوشه‌ای در یک مطالعه روان‌شناختی است. در این راستا، تمرکز بر تحلیل نموداری رفتار متغیرها در طول زمان، بررسی ناهمگنی بین افراد، و شناسایی پروفایل‌های رفتاری مبتنی بر خلاصه‌سازی ویژگی‌های طولی هر فرد قرار دارد. این پژوهش بدون استفاده از مدل‌های پیش‌بینی یا استنباطی، چارچوبی توصیفی و اکتشافی برای فهم ساختار داده‌های تکرارشونده ارائه می‌دهد.

در ادامه گزارش، ابتدا مفاهیم و تعاریف اساسی مرتبط با داده‌های طولی و خوشه‌ای معرفی می‌شوند، سپس روش‌شناسی علمی و مراحل تحلیل داده‌ها تشریح شده و در نهایت نتایج حاصل از تحلیل اکتشافی و خوشه‌بندی افراد ارائه و تفسیر می‌گردد.

توصیه‌های سیاستی

یافته‌های حاصل از تحلیل اکتشافی داده‌های طولی می‌توانند به تصمیم‌گیری‌های مبتنی بر شواهد در حوزه سلامت روان کمک کنند. شناسایی الگوهای پایدار یا نوسانی استرس و عاطفه منفی در طول زمان، امکان طراحی مداخلات هدفمندتر برای گروه‌های مختلف افراد را فراهم می‌سازد. همچنین، مشاهده تفاوت‌های بین‌فردی در پاسخ‌های روان‌شناختی می‌تواند سیاست‌گذاران را به سمت اتخاذ رویکردهای شخصی‌سازی‌شده به‌جای راهکارهای یکسان برای همه افراد سوق دهد.

فصل ۱

مفاهیم و تعاریف اساسی و معرفی داده ها

۱.۱. تعریف کوتاه و علمی همه اصطلاحات و روش های استفاده شده

داده طولی (Longitudinal Data) : داده‌های طولی شامل اندازه‌گیری‌های مکرر یک یا چند متغیر برای واحدهای آماری در طول زمان هستند. این نوع داده‌ها امکان بررسی دو نوع تغییرپذیری را فراهم می‌کنند: تغییرات درون‌فردی (within-subject) و تفاوت‌های بین‌فردی (between-subject).

داده خوشه‌ای (Clustered Data) : داده‌های خوشه‌ای به داده‌هایی اطلاق می‌شوند که در آن‌ها مشاهدات درون یک خوشه (مانند یک فرد) مستقل از یکدیگر نیستند.

تحلیل اکتشافی داده‌ها (EDA) : مجموعه‌ای از روش‌های آماری و بصری است که برای شناخت اولیه داده‌ها، کشف الگوها و شناسایی روابط بین متغیرها استفاده می‌شود.

Spaghetti Plot : نموداری که مسیر تغییرات هر فرد را در طول زمان نمایش می‌دهد و برای بررسی ناهمگنی رفتار افراد استفاده می‌شود.

Mean Trend Plot : نموداری که میانگین مقدار یک متغیر را در طول زمان نمایش می‌دهد و روند کلی جمعیت را نشان می‌دهد.

Boxplot زمانی : نموداری که تغییر توزیع داده‌ها را در زمان نشان می‌دهد و برای بررسی پراکندگی و تغییرات آماری استفاده می‌شود.

همبستگی (Correlation) : معیاری آماری است که میزان ارتباط خطی بین دو متغیر را اندازه‌گیری می‌کند.

نرمال‌سازی داده‌ها (Standardization) : فرآیندی که در آن داده‌ها به مقیاس استاندارد تبدیل می‌شوند تا اثر تفاوت واحد اندازه‌گیری متغیرها کاهش یابد.

خوشه‌بندی (Clustering) : روشی بدون ناظر در یادگیری ماشین است که داده‌ها را بر اساس شباهت ویژگی‌ها در گروه‌های مختلف تقسیم می‌کند.

الگوریتم K-Means : یکی از روش‌های متداول خوشه‌بندی است که داده‌ها را به k گروه تقسیم می‌کند به طوری که شباهت درون خوشه‌ای بیشینه و بین خوشه‌ای کمینه باشد.

روش Elbow : روشی برای تعیین تعداد مناسب خوشه‌ها است که با بررسی تغییر خطای درون خوشه‌ای انجام می‌شود.

Inertia : معیاری از میزان پراکندگی داده‌ها در داخل هر خوشه است که در الگوریتم K-Means استفاده می‌شود.

تحلیل ترکیبی طولی و خوشه‌ای : روشی که در آن ابتدا افراد خوشه‌بندی می‌شوند و سپس تغییرات زمانی هر خوشه بررسی می‌شود.

۲.۱. معرفی داده ها

داده‌ی مورد استفاده شامل ۷۲ فرد است که طی ۳۰ روز متوالی متغیرهای روان‌شناختی مختلفی برای آن‌ها اندازه‌گیری شده است. این ساختار، داده‌ای طولی با خوشه‌بندی طبیعی بر اساس افراد ایجاد می‌کند که برای تحلیل اکتشافی چندسطحی مناسب است.

در این پژوهش داده‌ها دارای ساختار طولی هستند، به‌طوری‌که برای هر فرد اندازه‌گیری‌های مکرر در طول ۳۰ روز ثبت شده است. در مرحله آماده‌سازی داده‌ها، هر فرد به عنوان یک واحد تحلیلی مستقل در نظر گرفته شد تا امکان بررسی رفتار زمانی افراد فراهم گردد. سپس با استفاده از شاخص‌های خلاصه رفتاری استخراج‌شده از داده‌های طولی، افراد بر اساس الگوهای رفتاری مشابه خوشه‌بندی شدند. به عبارتی،

در این پژوهش فرد واحد اصلی تحلیل بوده و خوشه‌بندی در سطح افراد انجام شده است.

ادامه گزارش با بیان یک پرسش آغاز می‌شود که آیا در تحلیل اکتشافی متغیر هدف داریم؟ خیر.

در تحلیل اکتشافی داده‌ها متغیر هدف (Target / Response) به معنای مدل سازی وجود ندارد.

تنها یک یا چند متغیر اصلی مورد بررسی (measurement variable) وجود دارد. که:

- روی زمان معنا داشته باشد.
- بین افراد نوسان نشان دهد.
- بتوان الگو کشف کرد.

بنابراین، در تحلیل اکتشافی داده‌ها، هیچ متغیر پاسخی تعریف نمی‌شود. تمرکز بر درک ساختار، الگوها و تغییرپذیری داده‌ها است.

۱.۲.۱. بررسی ویژگی ها در هر ستون

ستون	نقش مفهومی در تحلیل
ID	شناسه فرد، ۷۲ نفر
day	زمان مشاهده، ۱ تا ۳۰ روز —> طولی
Daily Stressor	استرس روزانه - متغیر اصلی پویا
EF coping	راهبرد مقابله هیجان محور
NeA	عاطفه منفی
PSS10	استرس ادراک شده
Trait Anxiety	اضطراب صفتی - تقریباً ثابت
RES	تاب آوری

ID

- نوع داده: شناسه منحصر به فرد
- توضیح: هر ردیف داده متعلق به یک شرکت کننده یا واحد نمونه است.
- نقش در تحلیل: شناسه کلیدی برای داده‌های طولی، چون ما ممکن است چند رکورد در روزهای مختلف برای یک فرد داشته باشیم. برای خوشه‌بندی معمولاً این ستون حذف می‌شود چون صرفاً شناسه است.

Day

- نوع داده: تاریخ یا زمان
- توضیح: نشان‌دهنده روزی است که داده‌ها ثبت شده‌اند. مثلاً روز اول، دوم و...
- نقش در تحلیل طولی: این ستون **متغیر زمان** است و برای رسم تغییرات روزانه، تحلیل روند و مدل‌های داده طولی لازم است.

Daily Stressor

- نوع داده: عددی
- توضیح: مقدار **استرس روزانه** شرکت کننده. یک نمره یا مقیاس ۰-۲۰ است.
- نقش در تحلیل:
- تحلیل طولی: بررسی روند استرس روزانه هر فرد
- خوشه بندی: می‌توان از آن برای دسته‌بندی افراد بر اساس الگوی استرس روزانه استفاده کرد.

EFcoping

- نوع داده: عددی
- توضیح: توانایی یا استراتژی‌های کنترل و مقابله اجرایی (Executive Function) (coping) در مقابل استرس.
- نقش در تحلیل: می‌توان رفتار مقابله‌ای را بین افراد مقایسه کرد و برای خوشه‌بندی استفاده کرد.

NeA

- نوع داده: عددی
- توضیح: احتمالاً مخفف Negative Affect یا Neuroticism Affect است. یعنی میزان حالت منفی یا حساسیت به عواطف منفی.
- نقش در تحلیل:
- طولی: بررسی تغییرات حالات منفی در طول روزها
- خوشه بندی: ترکیب با دیگر ویژگی‌ها برای پیدا کردن گروه‌های مشابه از نظر حالات روانی

PSS10

- نوع داده: عددی
- توضیح: نمره Perceived Stress Scale 10-item، یعنی مقیاس ۱۰ سوالی ارزیابی استرس ادراک شده.
- نقش در تحلیل:
- طولی: اگر هر روز یا هر بار داده گرفته شود، روند استرس را نشان می‌دهد.
- خوشه بندی: شاخص کلی استرس افراد برای گروه‌بندی.

Trait Anxiety

- نوع داده: عددی
- توضیح: نمره اضطراب صفاتی، یعنی گرایش پایدار فرد به اضطراب
- نقش: شاخص پایدار فرد که معمولاً متغیر ثابت طولی است و می‌تواند به خوشه‌بندی افراد کمک کند.

RES

- نوع داده: عددی
- توضیح: احتمالاً مخفف Resilience، یعنی میزان تاب‌آوری روانی فرد
- نقش:
- طولی: می‌توان دید آیا تاب‌آوری باعث کاهش نوسانات استرس روزانه می‌شود.
- خوشه‌بندی: افراد با تاب‌آوری بالا و پایین ممکن است در خوشه‌های جداگانه قرار گیرند.

متغیر کانونی اصلی:

NeA (Negative Affect) ✓

- کشف الگو
- خوشه بندی مسیر های هیجانی

متغیر کانونی دوم (برای مقایسه):

Daily Stressor ✓

- Spaghetti plot
- Mean trend
- خوشه بندی الگوهای زمانی

یا

PSS10 ✓

- ادراک استرس
- قابل تفسیر روانشناختی

متغیر های تفسیری (نه هدف) :

- EFcoping
- TraitAnxiety
- RES

فصل ۲

روش شناسی علمی تحلیل اکتشافی داده های طولی و خوشه ای

در این پروژه، روش شناسی مبتنی بر تحلیل اکتشافی داده ها با تمرکز بر نمودارها و آمارهای توصیفی اتخاذ شده است. هدف از این مرحله، درک ساختار داده ها و کشف الگوهای اولیه پیش از هرگونه مدل سازی رسمی بوده است.

ابزارها

تحلیل ها با استفاده از زبان برنامه نویسی Python و کتابخانه های pandas، matplotlib و seaborn انجام شده اند.

۱.۲. هدف کلی تحلیل

هدف اصلی این فصل، تشریح فرآیند تحلیل اکتشافی داده های طولی و خوشه ای به گونه ای شفاف و گام به گام است. تمرکز این تحلیل بر درک رفتار زمانی متغیرهای روان شناختی، بررسی ناهمگنی بین افراد، و شناسایی الگوهای رفتاری متمایز در میان افراد بدون استفاده از مدل های پیش بینی یا استنباطی بوده است.

به طور خاص، هدف آن بود که مشخص شود آیا افراد، با وجود قرار گرفتن در یک جمعیت مشترک، الگوهای زمانی متفاوتی در طول دوره ۳۰ روزه از خود نشان می دهند یا خیر.

۲.۲. بررسی اولیه ساختار داده ها

در نخستین گام، ساختار داده‌ها از نظر تعداد افراد، تعداد مشاهدات در طول زمان و میزان کامل بودن داده‌ها بررسی شد. داده‌ها شامل ۷۲ فرد بودند که هر یک به‌طور متوسط دارای ۳۰ مشاهده روزانه بودند و در مجموع ۲۰۷۰ مشاهده ثبت شده بود. این بررسی نشان داد که داده‌ها ساختاری متوازن دارند و برای تحلیل طولی مناسب هستند. شناسایی این ساختار پیش از ادامه تحلیل ضروری بود، زیرا اعتبار الگوهای زمانی به کامل بودن داده‌ها وابسته است.

۳.۲. شناسایی ساختار طولی و خوشه ای

با توجه به اینکه مشاهدات به‌صورت تکرارشونده برای هر فرد ثبت شده‌اند، داده‌ها به‌طور طبیعی دارای ساختار خوشه‌ای هستند؛ به‌گونه‌ای که مشاهدات درون هر فرد به یکدیگر وابسته‌اند. در این پروژه، هر فرد به‌عنوان یک خوشه طبیعی در نظر گرفته شد. این تصمیم مفهومی مبتنی بر ماهیت داده‌ها بود و به جلوگیری از تحلیل نادرست مشاهدات روزانه به‌عنوان واحدهای مستقل کمک کرد.

۴.۲. آماره های توصیفی و شناخت اولیه متغیر ها

در ادامه، آماره‌های توصیفی شامل میانگین، انحراف معیار، حداقل و حداکثر برای هر یک از متغیرهای روان‌شناختی محاسبه شد. هدف از این مرحله، درک دامنه تغییرات، مقیاس اندازه‌گیری و پراکندگی هر متغیر بود. این مرحله به شناسایی تفاوت‌های کلی بین متغیرها و بررسی وجود مقادیر غیرعادی کمک کرد و مبنایی برای تفسیر نمودارهای بعدی فراهم ساخت.

۵.۲. تحلیل اکتشافی طولی در سطح جمعیت

برای بررسی الگوهای زمانی کلی، میانگین هر متغیر در هر روز محاسبه و نمودارهای روند میانگین ترسیم شدند. این نمودارها امکان بررسی تغییرات تدریجی یا پایداری متغیرها در سطح جمعیت را فراهم کردند. هدف از این مرحله، شناسایی الگوهای عمومی زمانی بدون در نظر گرفتن تفاوت‌های فردی بود. نتایج نشان دادند که برخی متغیرها تغییرات قابل توجهی در طول زمان دارند، در حالی که برخی دیگر نسبتاً پایدار باقی مانده‌اند.

۶.۲. بررسی تغییرپذیری توزیعی در طول زمان

برای تکمیل تحلیل میانگین‌ها، از نمودارهای جعبه‌ای در طول زمان استفاده شد. این نمودارها اطلاعات بیشتری درباره پراکندگی، چولگی و نوسانات روزانه متغیرها ارائه دادند. این مرحله اهمیت داشت زیرا تغییرات میانگین به تنهایی قادر به نمایش تغییرات کامل توزیع داده‌ها نیست و ممکن است تغییرپذیری بین افراد را پنهان کند.

۷.۲. تحلیل اکتشافی خوشه‌ای در سطح افراد

پس از بررسی روندهای کلی، تمرکز تحلیل به تفاوت‌های بین افراد معطوف شد. برای این منظور، مسیر زمانی هر فرد به صورت جداگانه در قالب نمودارهای مسیر فردی (Spaghetti plots) ترسیم شد. این نمودارها ناهمگنی قابل توجهی در سطح پایه، دامنه نوسانات و جهت تغییرات زمانی بین افراد را نشان دادند. مشاهده این تفاوت‌ها نشان داد که تحلیل در سطح میانگین جمعیت نمی‌تواند رفتار تمامی افراد را به طور کامل توصیف کند.

۸.۲. استخراج ویژگی های فردی از داده های طولی

به منظور مقایسه سیستماتیک افراد، داده های طولی هر فرد به مجموعه ای از ویژگی های خلاصه تبدیل شدند. این ویژگی ها شامل میانگین، انحراف معیار و شیب تغییرات زمانی متغیرهای کلیدی بودند.

این مرحله پلی میان تحلیل طولی و تحلیل خوشه ای ایجاد کرد و امکان بررسی الگوهای رفتاری افراد به عنوان واحدهای مستقل را فراهم ساخت.

۹.۲. تحلیل اکتشافی خوشه ای افراد

با استفاده از ویژگی های خلاصه شده، افراد از نظر شباهت رفتاری مورد بررسی قرار گرفتند. پیش از اعمال هر الگوریتم خوشه بندی، نمودارهای پراکندگی بین ویژگی های کلیدی ترسیم شد تا وجود الگوهای بالقوه و تعداد تقریبی خوشه ها به صورت بصری ارزیابی شود. این تحلیل بصری نشان داد که افراد تمایل دارند در چند گروه رفتاری متمایز قرار گیرند که از نظر سطح استرس، میزان نوسان و تاب آوری تفاوت دارند.

۱۰.۲. تعیین تعداد خوشه ها

برای توجیه انتخاب تعداد خوشه ها، از روش آرنج (Elbow method) و تحلیل سیلوئت استفاده شد. این روش ها کاهش واریانس درون خوشه ای و میزان تفکیک خوشه ها را ارزیابی می کنند. نتایج این تحلیل ها نشان دادند که انتخاب سه خوشه می تواند تعادلی مناسب بین سادگی و قدرت تفکیک ایجاد کند.

۱۱.۲. خوشه بندی نهایی افراد

در مرحله نهایی، خوشه بندی افراد بر اساس ویژگی های خلاصه شده انجام شد. تأکید می شود که این خوشه بندی با هدف توصیف الگوهای رفتاری و نه پیش بینی یا استنباط آماری انجام شده است. خوشه های حاصل نشان دهنده پروفایل های رفتاری متمایزی بودند که از نظر سطح کلی استرس، تغییر پذیری و تاب آوری تفاوت داشتند.

۱۲.۲. تحلیل ترکیبی طولی و خوشه ای

در گام پایانی، برچسب خوشه ها به داده های طولی بازگردانده شد و مسیرهای زمانی میانگین برای هر خوشه ترسیم گردید. این تحلیل ترکیبی امکان مقایسه مستقیم الگوهای زمانی بین گروه های رفتاری مختلف را فراهم کرد. نتایج نشان دادند که خوشه های شناسایی شده نه تنها از نظر سطح کلی متغیرها، بلکه از نظر الگوی تغییر در طول زمان نیز تفاوت دارند.

EDA پروژه ۶ بخش کامل و استاندارد را پوشش می دهد:

بخش EDA	روش و نمودار مناسب
۱. ساختار داده ها	Info , nunique , balance
۲. توصیف آماری	Describe()
۳. EDA طولی	Mean trajectory , boxplot
۴. EDA بین فردی – خوشه ای	Spaghetti plot
۵. روابط بین متغیر ها	Correlation , Scatter
۶. EDA ترکیبی طولی + خوشه ای	Grouped trajectories

این رویکرد ها نشان داد که چگونه می توان بدون تکیه بر مدل های پیچیده، از داده های طولی و خوشه ای اطلاعات ارزشمندی استخراج کرد.

۱۳.۲. استراتژی تحلیلی پژوهش

در این پژوهش به منظور شناسایی الگوهای رفتاری افراد، یک چارچوب تحلیلی چندمرحله‌ای مبتنی بر تحلیل اکتشافی داده‌های طولی و تحلیل خوشه‌ای طراحی شد. این رویکرد با هدف درک تغییرات زمانی متغیرهای روان‌شناختی و طبقه‌بندی افراد بر اساس الگوهای رفتاری آن‌ها اجرا گردید.

در گام نخست، تحلیل اکتشافی داده‌های طولی برای بررسی رفتار هر فرد در طول دوره مطالعه انجام شد. در این مرحله مسیر تغییرات متغیرهای مورد مطالعه برای هر فرد به صورت جداگانه بررسی شد. همچنین روند میانگین متغیرها در طول زمان و میزان نوسانات آن‌ها مورد ارزیابی قرار گرفت. نتایج این مرحله نشان داد که افراد از نظر شدت و پایداری تجربه استرس و حالات هیجانی الگوهای متفاوتی دارند؛ به‌طوری‌که برخی افراد دارای سطح بالای پایدار استرس، برخی دارای نوسانات زیاد و برخی دارای سطح پایین و پایدار بودند.

در گام دوم، با توجه به ماهیت داده‌های طولی، هر فرد به عنوان یک واحد تحلیلی مستقل در نظر گرفته شد. دلیل این رویکرد آن است که اندازه‌گیری‌های روزانه در داخل هر فرد به یکدیگر وابسته بوده و تحلیل در سطح روزها می‌تواند ساختار وابستگی داده‌ها را نادیده بگیرد. بنابراین هدف تحلیل خوشه‌ای در این پژوهش، گروه‌بندی افراد بر اساس الگوهای رفتاری آن‌ها بود.

در گام سوم، به منظور مقایسه افراد و انجام تحلیل خوشه‌ای، داده‌های طولی هر فرد به مجموعه‌ای از شاخص‌های خلاصه تبدیل شد. این شاخص‌ها شامل میانگین متغیرهای استرس روزانه به عنوان نمایانگر سطح کلی استرس، انحراف معیار به عنوان شاخص نوسان رفتاری، شیب تغییرات زمانی به عنوان نشان‌دهنده روند تغییرات، میانگین عاطفه منفی به عنوان شاخص وضعیت هیجانی و میزان تاب‌آوری به عنوان عامل محافظ روان‌شناختی بودند. استفاده از این شاخص‌ها امکان مقایسه الگوهای رفتاری افراد و شناسایی گروه‌های همگن را فراهم ساخت.

این چارچوب تحلیلی امکان بررسی همزمان پویایی زمانی متغیرها و ناهمگنی بین افراد را فراهم کرده و به شناسایی الگوهای رفتاری متمایز در جامعه مورد مطالعه کمک نمود.

فصل ۳

کاربرد و نتایج تحلیل اکتشافی

۱.۳. آماده سازی و پاکسازی داده ها

در ابتدا داده‌های مطالعه شامل ۷۲ فرد بود که متغیرهای روان‌شناختی آنان طی ۳۰ روز اندازه‌گیری شده بود. در مجموع داده‌ها شامل ۲۰۷۰ مشاهده بود.

برای افزایش دقت تحلیل‌ها، کیفیت داده‌ها بررسی شد. ابتدا داده‌های تکراری و داده‌های گمشده شناسایی شدند. سپس مقادیر نامعتبر مانند کد ۹۹۹ به عنوان داده گمشده در نظر گرفته شد. برای حفظ ساختار طولی داده‌ها، مقادیر گمشده با استفاده از میانگین هر فرد جایگزین شدند. این روش باعث حفظ ویژگی‌های رفتاری هر فرد شد.

در نهایت داده‌ها بر اساس شناسه فرد و زمان مرتب شدند.

پنج سطر ابتدایی داده‌ها:

	ID	day	DailyStressor	EFcoping	NeA	PSS10	TraitAnxiety	RES
0	1	1	13	8	2.2	23	38	.06
1	1	2	13	8	3.7	25	38	.06
2	1	3	15	10	2.1	21	38	.06
3	1	4	13	10	3.1	21	38	.06
4	1	5	13	10	2.6	22	38	.06

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2071 entries, 0 to 2070
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   ID              2071 non-null  object
1   day             2071 non-null  object
2   DailyStressor   2071 non-null  object
3   Efcoping        2071 non-null  object
4   NeA             2071 non-null  object
5   PSS10          2071 non-null  object
6   TraitAnxiety    2071 non-null  object
7   RES            2071 non-null  object
dtypes: object(8)
```

هشت ستون متغیر مجموعه داده را مشاهده می کنیم که ۲۰۷۱ مشاهده غیر خالی داریم و متوجه می شویم مقدار متغیر ها را با توجه به اینکه عددی هستند به صورت **object** تعریف شده است بنابراین باید نوع داده ها را تغییر بدیم.

	ID	day	DailyStressor	Efcoping	NeA	PSS10	TraitAnxiety	RES
count	2071	2071	2071	2071	2071	2071	2071	2071
unique	70	31	15	12	39	39	28	58
top	1	7	11	6	1	20	45	.06
freq	30	70	297	427	251	237	210	90

count تعداد کل مقادیر موجود در هر ستون را نشان می دهد.

unique تعداد مقادیر متفاوت در آن ستون است.

top پرتکرارترین مقدار در آن ستون را معرفی می کند.

freq تعداد دفعاتی است که آن مقدار پرتکرار ظاهر شده.

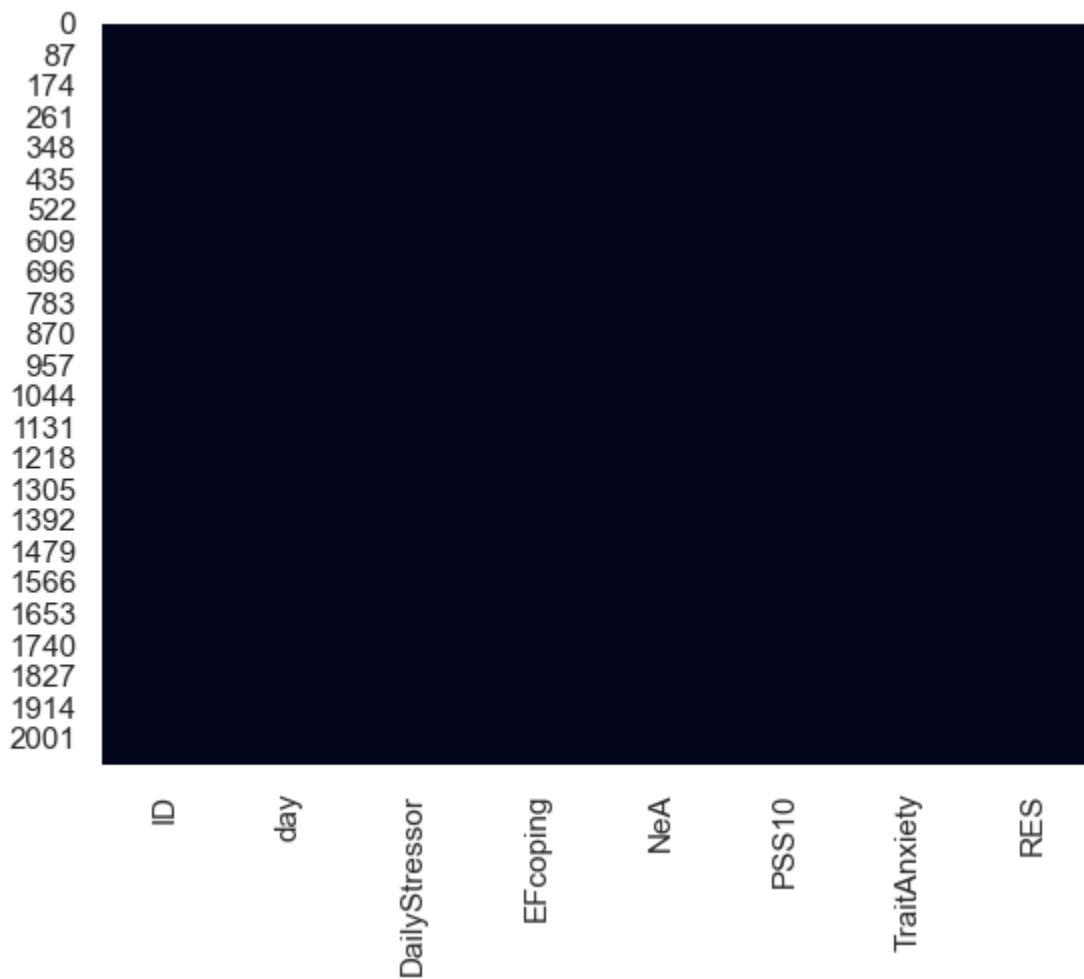
Total duplicated rows: 0

تعداد ردیف های تکراری را بررسی کردیم که با توجه به خروجی ردیف تکراری نداریم.

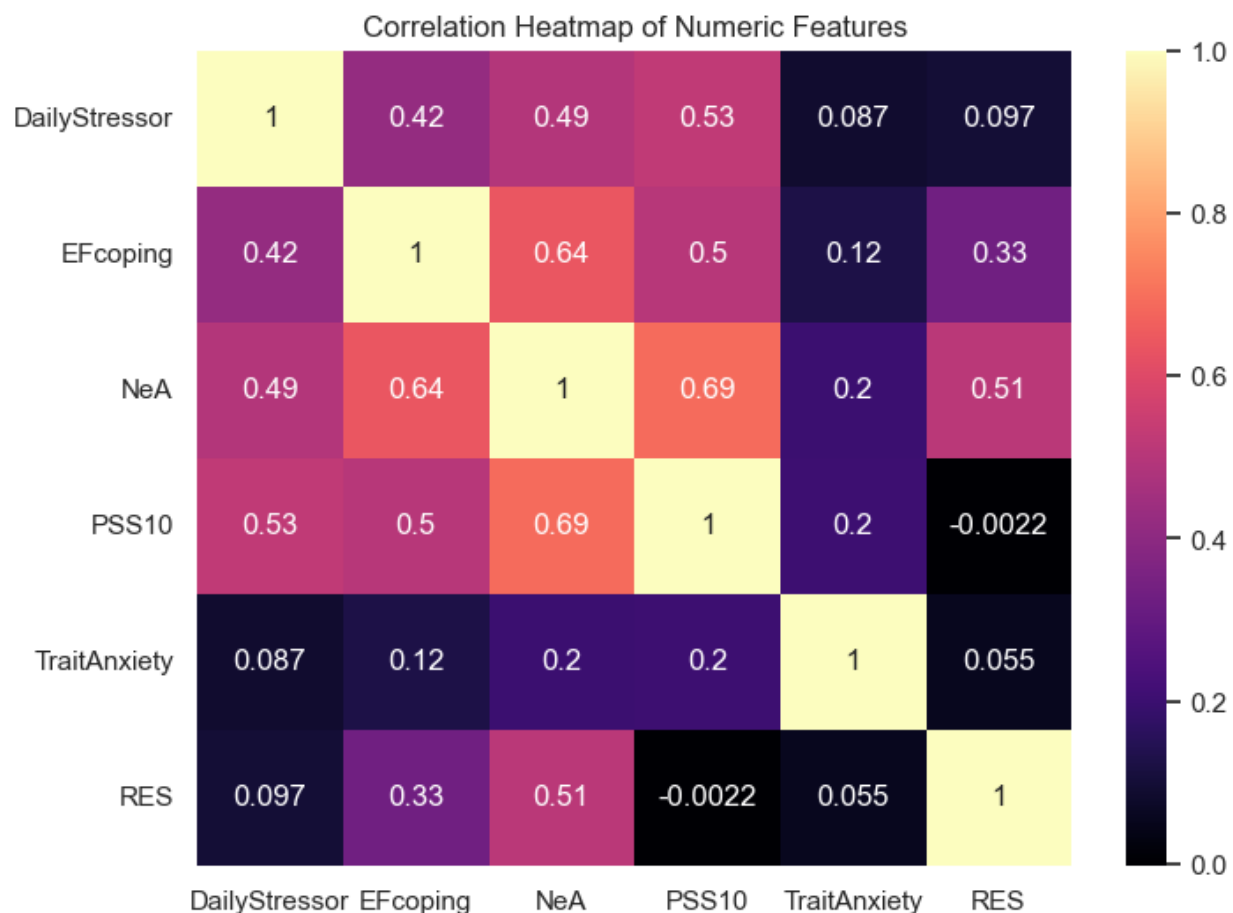
ID	0
day	0
DailyStressor	0
EFcoping	0
NeA	0
PSS10	0
TraitAnxiety	0
RES	0

مجموعه داده هیچ مقدار گم شده ای ندارد.

Heatmap داده های گم شده که نشان می دهد داده ناقص وجود ندارد:



نمودار حرارتی متغیرهای عددی:

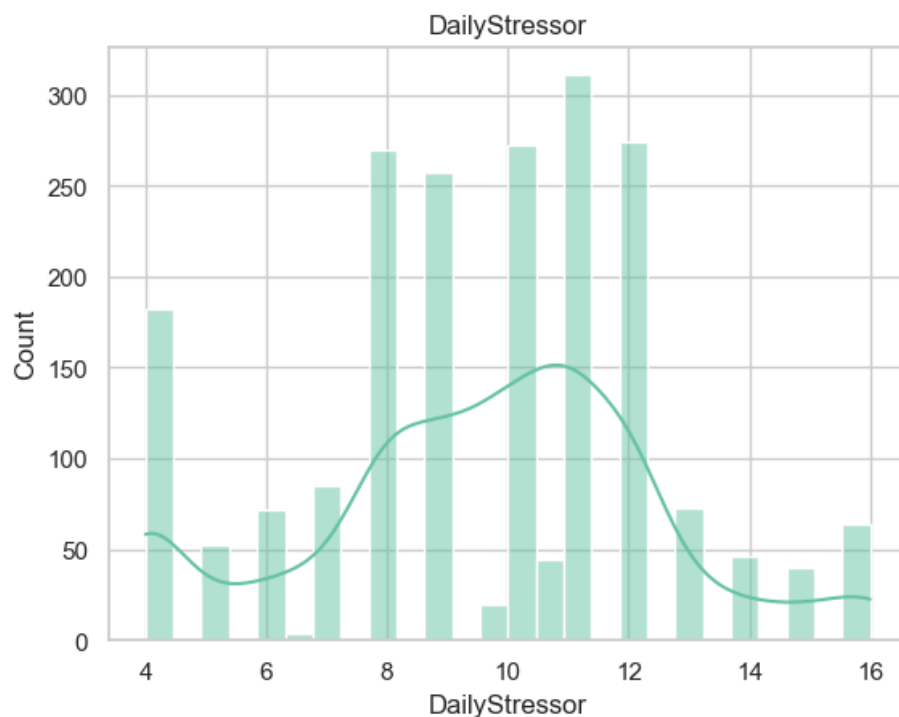


این هیت‌مپ نشان می‌دهد که متغیرهای روزانه مثل NeA، PSS10 و EFcoping همبستگی‌های نسبتاً قوی و مثبت با یکدیگر دارند، که بیانگر هم‌تغییری هیجان منفی، استرس ادراک‌شده و مقابله است. DailyStressor نیز با این سه متغیر همبستگی متوسط دارد و نشان می‌دهد افزایش استرس‌ورهای روزانه معمولاً با افزایش هیجان منفی و استرس همراه است. در مقابل، TraitAnxiety و RES همبستگی‌های بسیار ضعیف‌تری با سایر متغیرها دارند و نقش بین‌فردی بیشتری نسبت به نقش روزانه نشان می‌دهند.

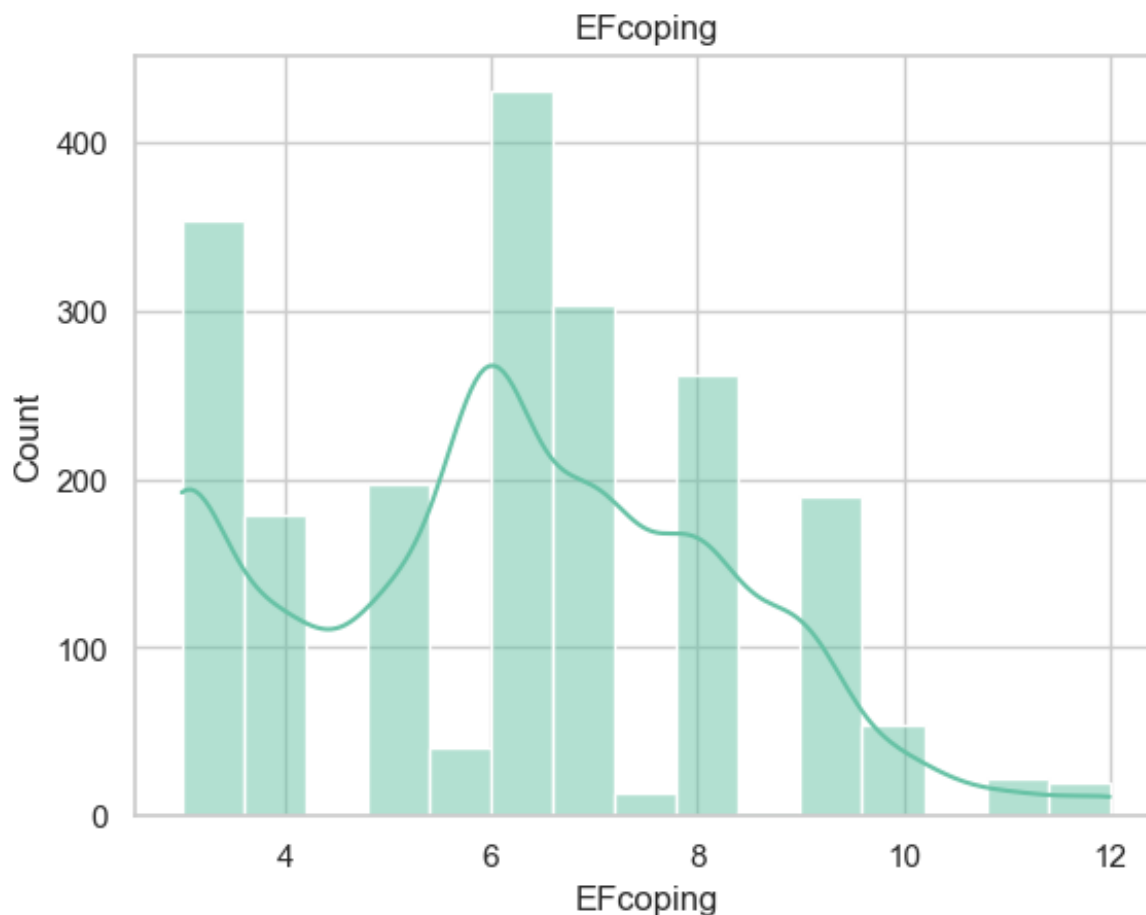
۲.۳. بررسی توزیع متغیرها (EDA تک متغیره)

در این مرحله توزیع متغیرهای اصلی شامل استرس روزانه، مقابله هیجانی، عاطفه منفی، استرس ادراک شده و تاب‌آوری بررسی شد. پیام علمی آن پراکندگی متغیرها، وجود چولگی و دامنه تغییرات است.

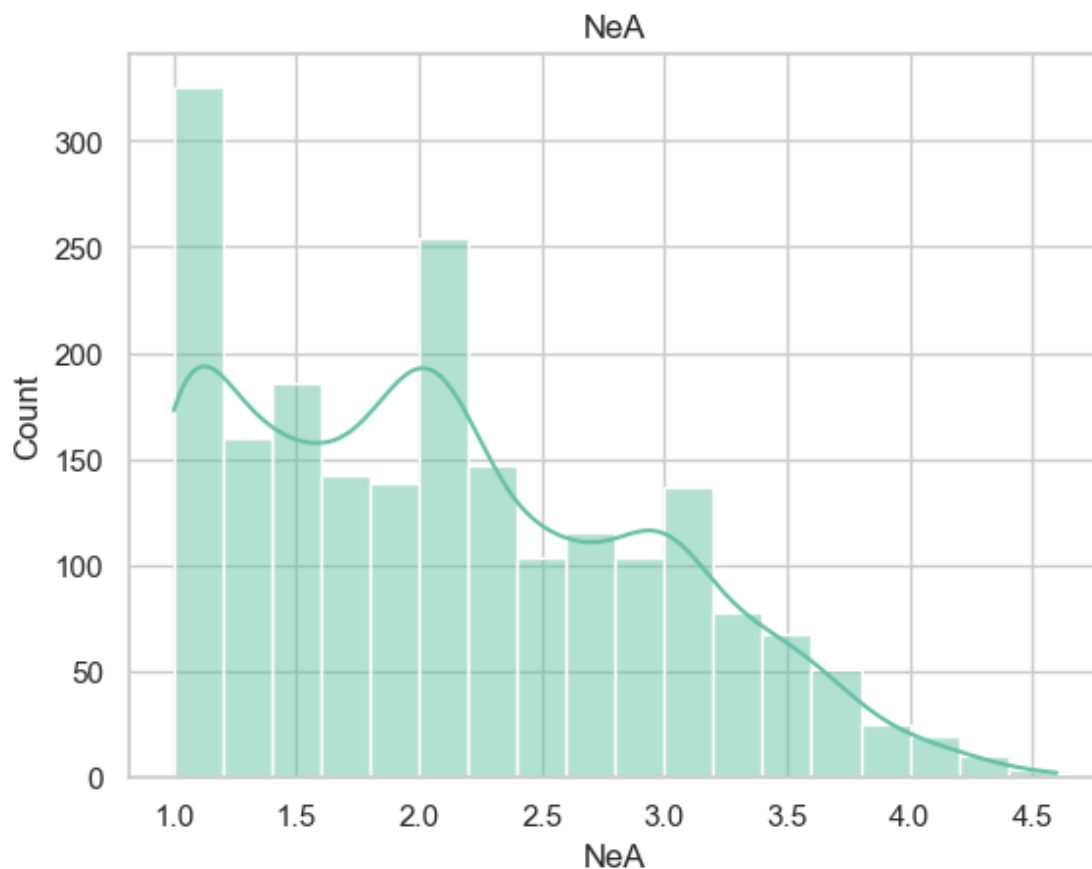
نتایج نشان داد که متغیرها دارای پراکندگی قابل توجهی هستند که نشان‌دهنده وجود تفاوت‌های فردی در پاسخ‌های روانی افراد است.



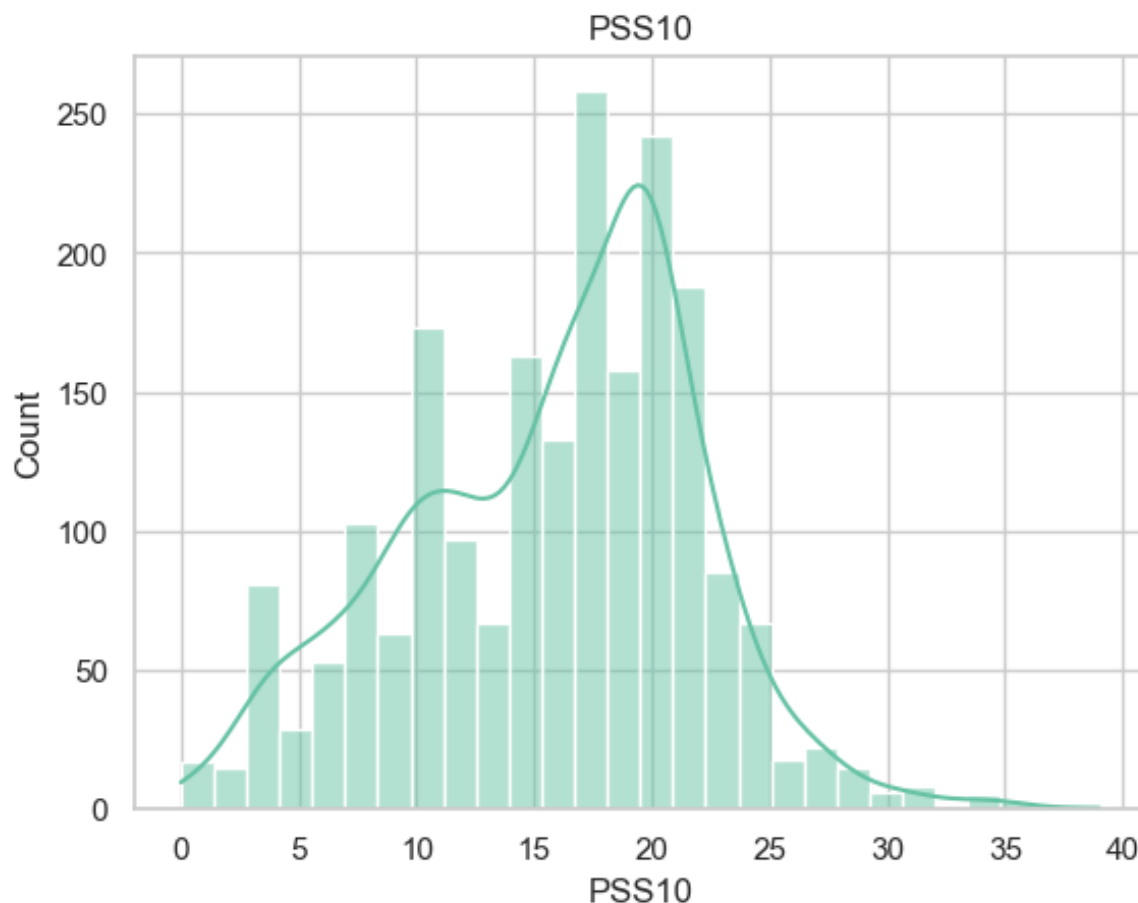
این نمودار نشان می‌دهد که توزیع DailyStressor حول یک مقدار مرکزی متمرکز است و بیشترین فراوانی در محدوده‌ی حدود ۱۰ تا ۱۳ دیده می‌شود. شکل توزیع نسبتاً متقارن است اما کمی تمایل به سمت مقادیر بالاتر دارد. وجود منحنی چگالی نشان می‌دهد که داده‌ها پراکندگی زیادی ندارند و بیشتر در یک بازه‌ی محدود تجمع کرده‌اند. مقادیر خیلی پایین یا خیلی بالا کمتر مشاهده می‌شوند، که نشان‌دهنده‌ی ثبات نسبی استرس‌سورهای روزانه در نمونه است.



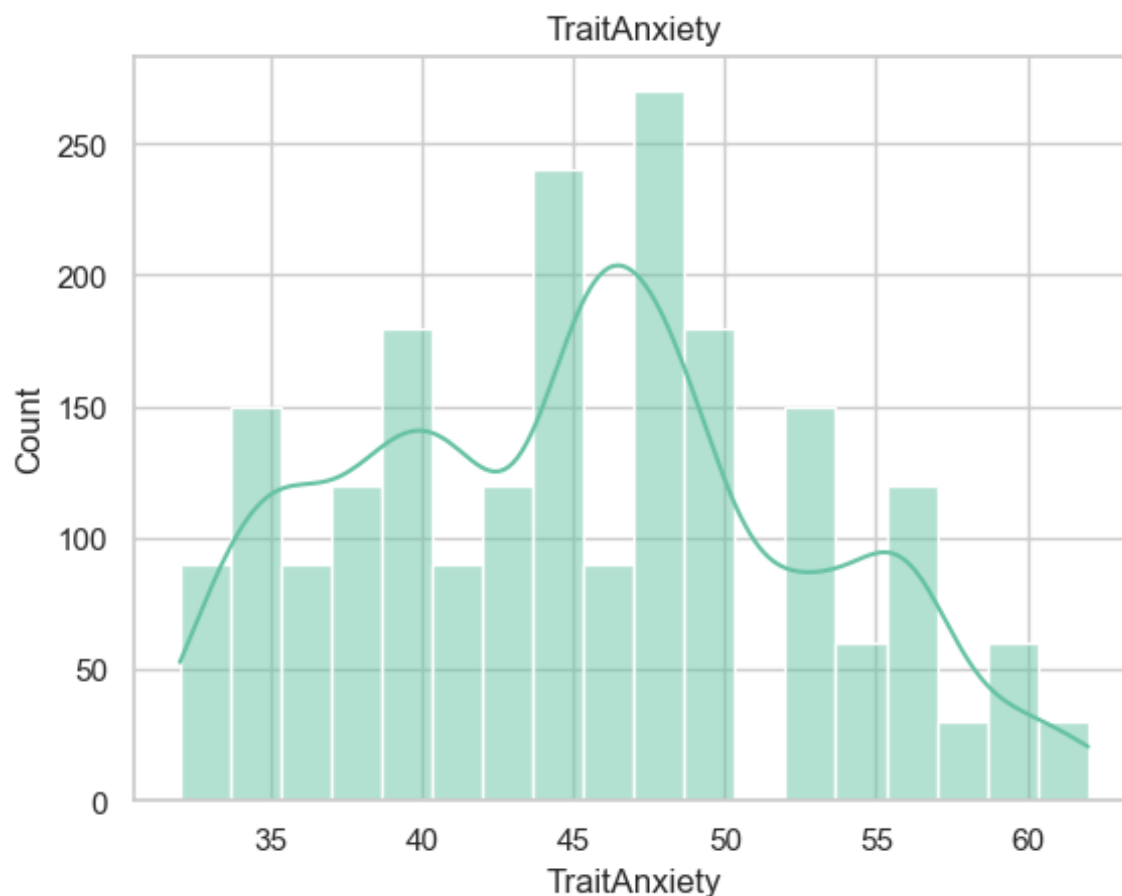
این توزیع EFcoping پراکندگی نسبتاً محدودی دارد و بیشتر مقادیر در یک بازه‌ی فشرده تجمع کرده‌اند، که نشان‌دهنده‌ی ثبات نسبی این متغیر در طول روزهاست. دامنه‌ی تغییرات از حدود ۳ تا ۱۲ است و بیشترین تراکم در حوالی مقدار ۶ دیده می‌شود. شکل توزیع کمی چولگی به سمت راست دارد، یعنی مقادیر بالاتر نسبت به مقادیر پایین کمی بیشتر کشیده شده‌اند. این الگو معمولاً بیانگر آن است که سطح مقابله‌ی کارکردی برای اکثر افراد متوسط است و مقادیر خیلی بالا کمتر رخ می‌دهد.



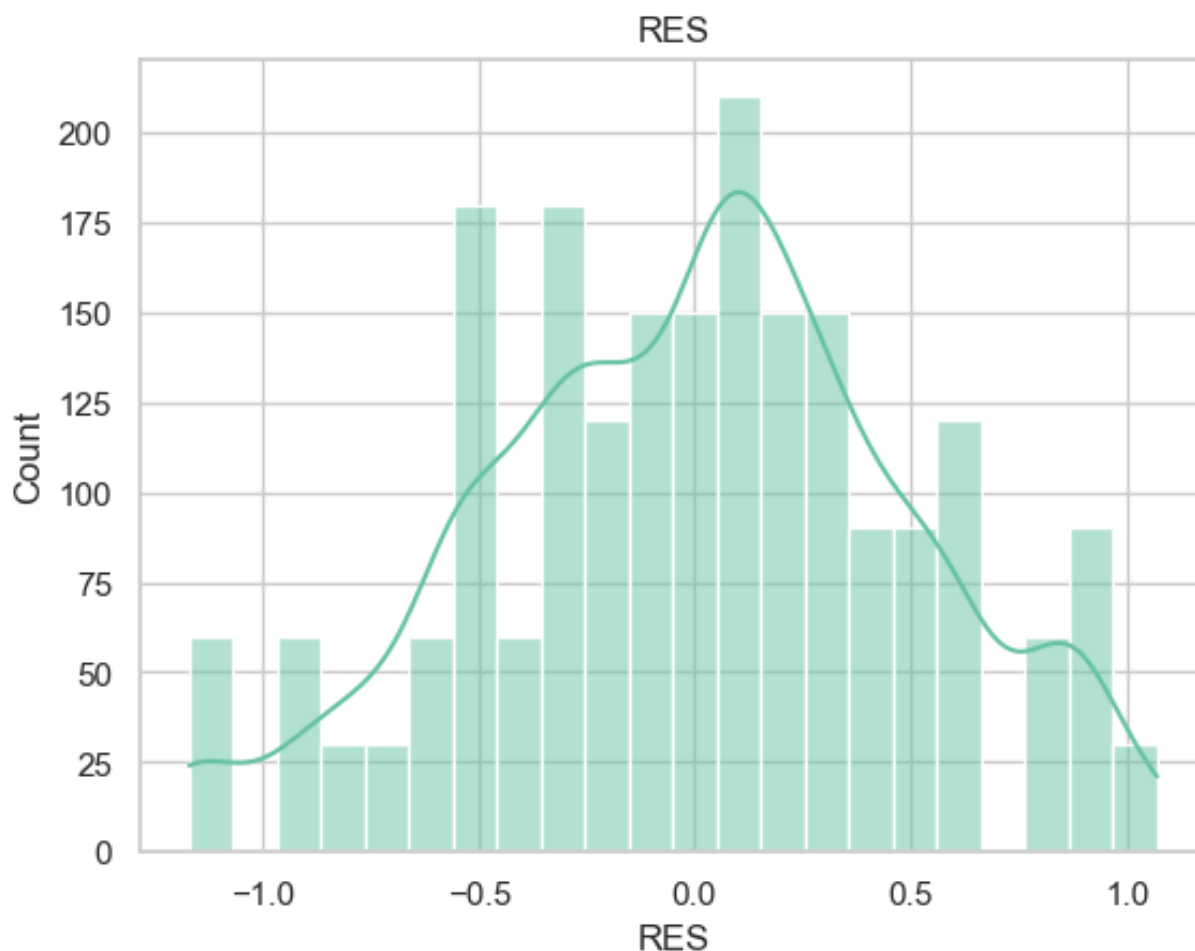
این توزیع نشان می‌دهد که مقدار **NeA** بیشتر در سطوح پایین متمرکز است و با افزایش مقدار، فراوانی به تدریج کاهش می‌یابد. شکل نمودار چولگی به راست دارد، یعنی مقادیر پایین بسیار رایج‌تر از مقادیر بالا هستند. دامنه تغییرات تقریباً از ۱ تا حدود ۴.۵ است و این نشان می‌دهد که هیچ‌ان منفی در بیشتر روزها در سطح پایین تا متوسط قرار داشته. پراکندگی کلی متوسط است و منحنی چگالی این الگوی کاهش تدریجی را تأیید می‌کند.



این توزیع نشان می‌دهد که مقدار PSS10 بیشتر در حوالی ۲۰ متمرکز است و این نقطه بیشترین فراوانی را دارد. شکل توزیع چولگی به راست دارد، یعنی مقادیر پایین‌تر رایج‌ترند و با افزایش مقدار، فراوانی کاهش پیدا می‌کند. دامنه تغییرات گسترده است و از حدود ۰ تا نزدیک ۴۰ امتداد دارد، که نشان‌دهنده تنوع زیاد در سطح استرس ادراک‌شده بین افراد و روزهاست. منحنی چگالی نیز این الگوی تمرکز مرکزی و کاهش تدریجی در مقادیر بالاتر را تأیید می‌کند.



این توزیع نشان می‌دهد که **TraitAnxiety** در یک بازه‌ی نسبتاً محدود (حدود ۳۰ تا ۶۵) تغییر می‌کند و بیشترین تراکم در حوالی مقدار ۴۸ دیده می‌شود. پراکندگی متوسط است و داده‌ها بیشتر در بخش میانی توزیع تجمع کرده‌اند. شکل نمودار کمی چولگی به راست دارد، یعنی مقادیر پایین‌تر کمی رایج‌تر از مقادیر خیلی بالا هستند. این الگو معمولاً نشان می‌دهد که اضطراب صفتی در نمونه نسبتاً پایدار است و افراد عمدتاً در سطح متوسط تا کمی بالا قرار دارند.

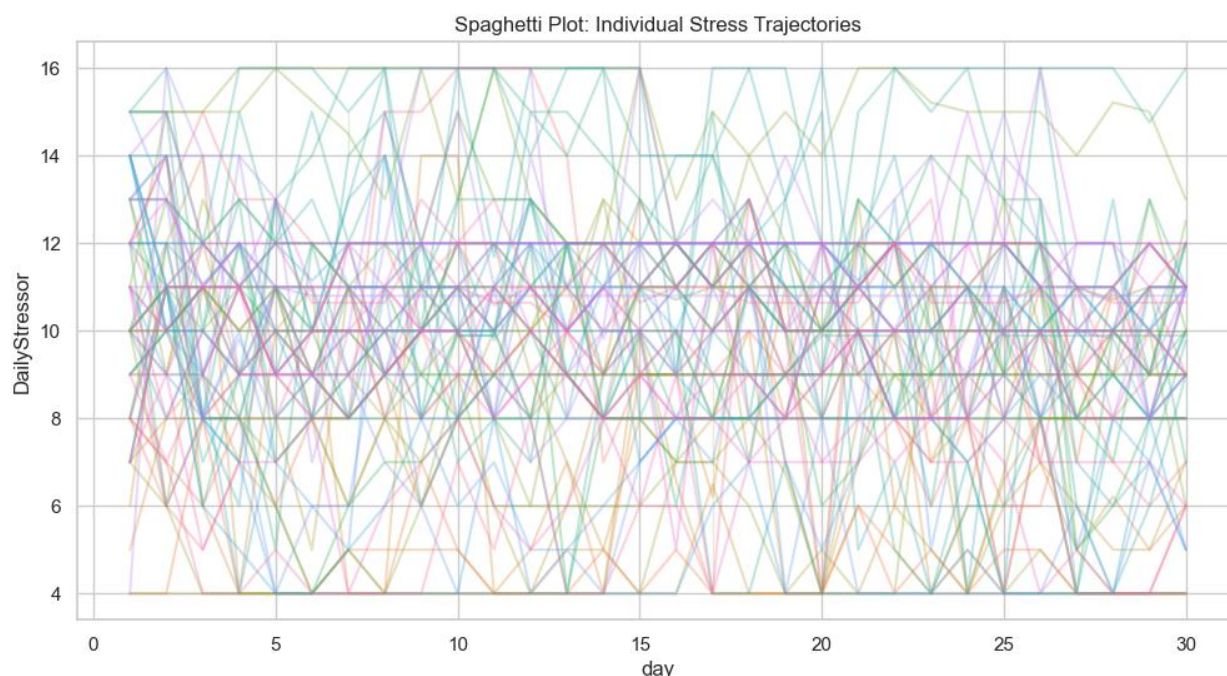


این توزیع نشان می‌دهد که مقدار RES حول صفر متمرکز است و شکل کلی آن تقریباً شبیه یک توزیع نرمال است. پراکندگی کم است و داده‌ها در یک بازه‌ی محدود قرار گرفته‌اند، که نشان می‌دهد تاب‌آوری بین افراد تفاوت زیادی ندارد. چولگی تقریباً وجود ندارد و توزیع متقارن به نظر می‌رسد. این الگو معمولاً بیانگر یک ویژگی نسبتاً پایدار و کم‌نوسان در سطح بین‌فردی است.

۳.۳. تحلیل اکتشافی طولی (هسته اصلی پروژه)

۱.۳.۳. بررسی مسیرهای زمانی افراد

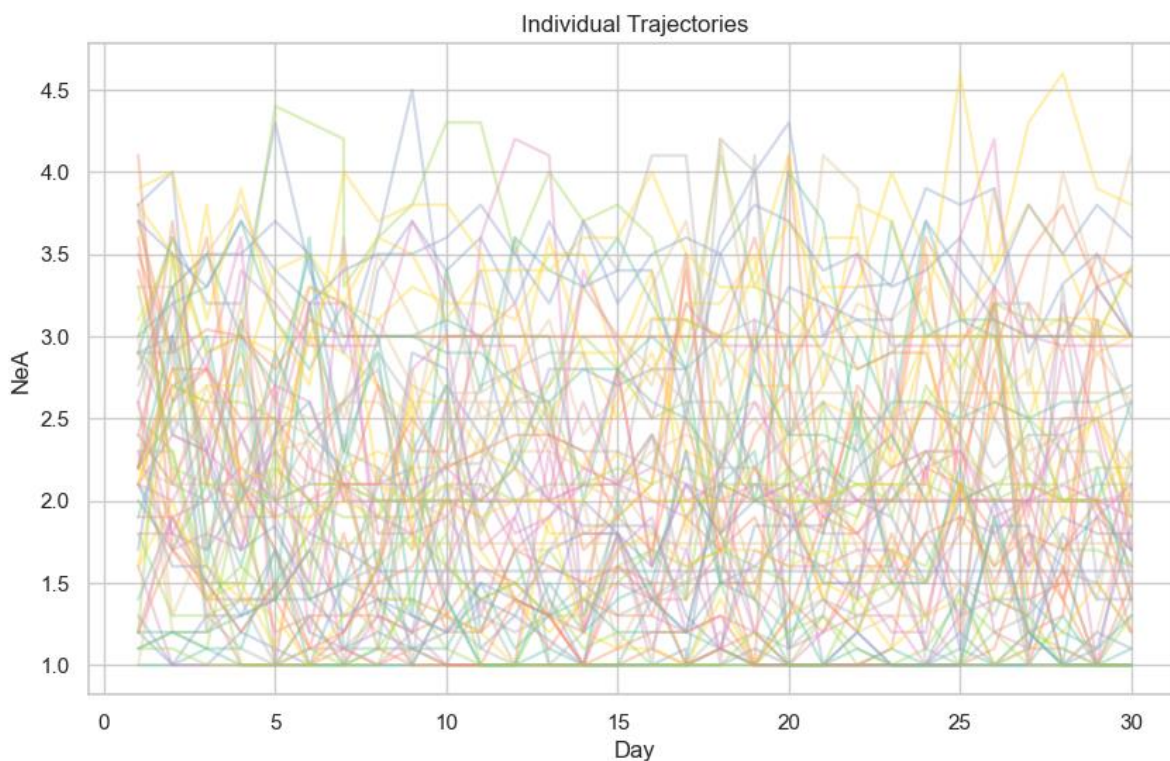
با استفاده از نمودارهای مسیر زمانی، رفتار هر فرد در طول دوره ۳۰ روزه بررسی شد. این تحلیل نشان داد که افراد الگوهای متفاوتی از استرس و عاطفه منفی دارند. برخی افراد دارای سطح استرس بالا و پایدار بودند، در حالی که برخی دیگر نوسانات قابل توجهی را نشان دادند.



در این نمودار، مسیرهای زمانی افراد نشان می‌دهد که الگوهای استرس روزانه بسیار متنوع و نوسانی هستند و هر فرد روند خاص خود را در طول ۳۰ روز دنبال می‌کند. تراکم خطوط بیانگر آن است که بخش زیادی از افراد در بازه‌ی میانی استرس قرار دارند و تنها تعداد کمی الگوهای بسیار بالا یا بسیار پایین دارند. نبود یک روند مشترک بین همه‌ی خطوط نشان می‌دهد که تغییرات استرس

بیشتر ماهیت درون فردی دارد تا بین فردی. این الگو برای داده‌های طولی روان‌شناختی کاملاً طبیعی است و نشان‌دهنده‌ی تفاوت‌های فردی پایدار در تجربه استرس است.

NeA فردی Trajectory :

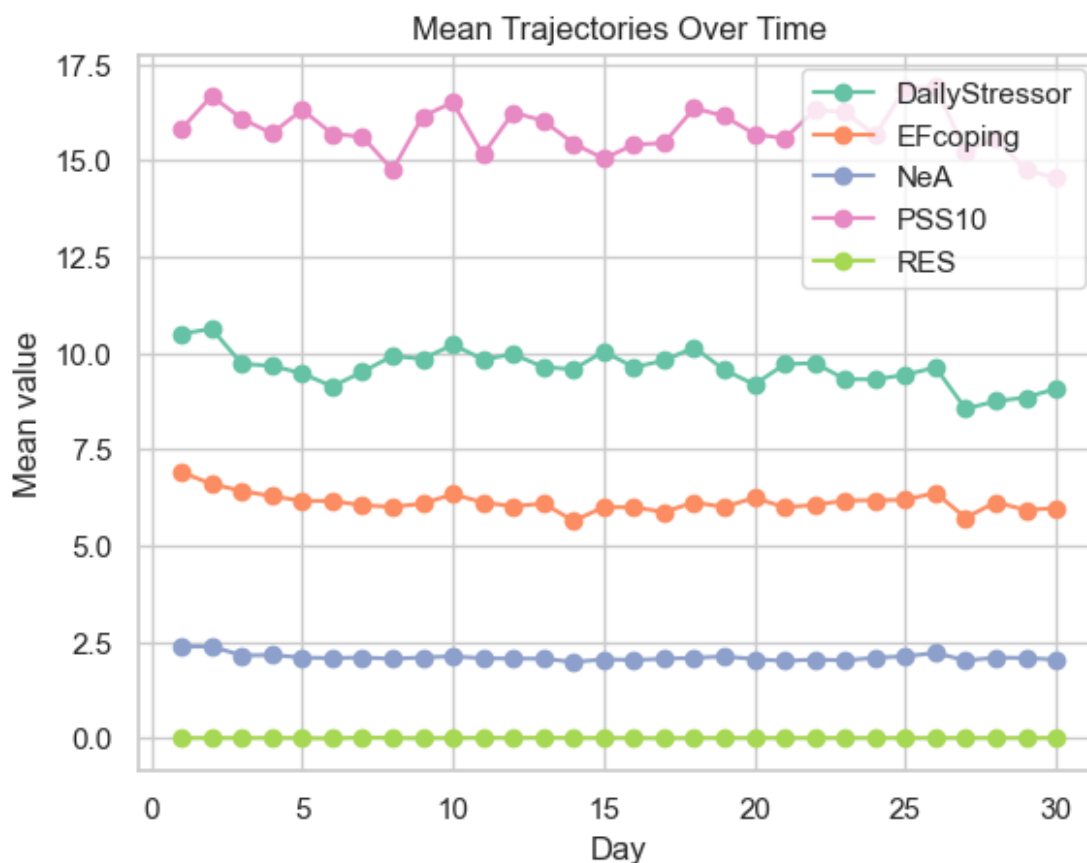


این نمودار نشان می‌دهد که مسیرهای زمانی NeA بین افراد بسیار متغیر و نوسانی است و هر فرد الگوی خاص خود را در طول ۳۰ روز دنبال می‌کند. بخش زیادی از خطوط در محدوده‌ی پایین تا متوسط قرار دارند، که نشان می‌دهد هیجان منفی برای بیشتر افراد در سطح نسبتاً پایین تجربه شده. نبود یک روند مشترک بین همه‌ی خطوط بیانگر آن است که تغییرات NeA بیشتر درون فردی است تا بین فردی. هم‌پوشانی زیاد خطوط نیز نشان می‌دهد که دامنه‌ی تغییرات محدود است اما نوسانات روزانه قابل توجه‌اند.

۲.۳.۳. بررسی روند کلی جمعیت

نمودار میانگین متغیرها در طول زمان نشان داد که تغییرات جمعیت دارای روند مشخصی است. همچنین باند انحراف معیار نشان داد که پراکندگی پاسخ افراد در طول زمان متغیر است.

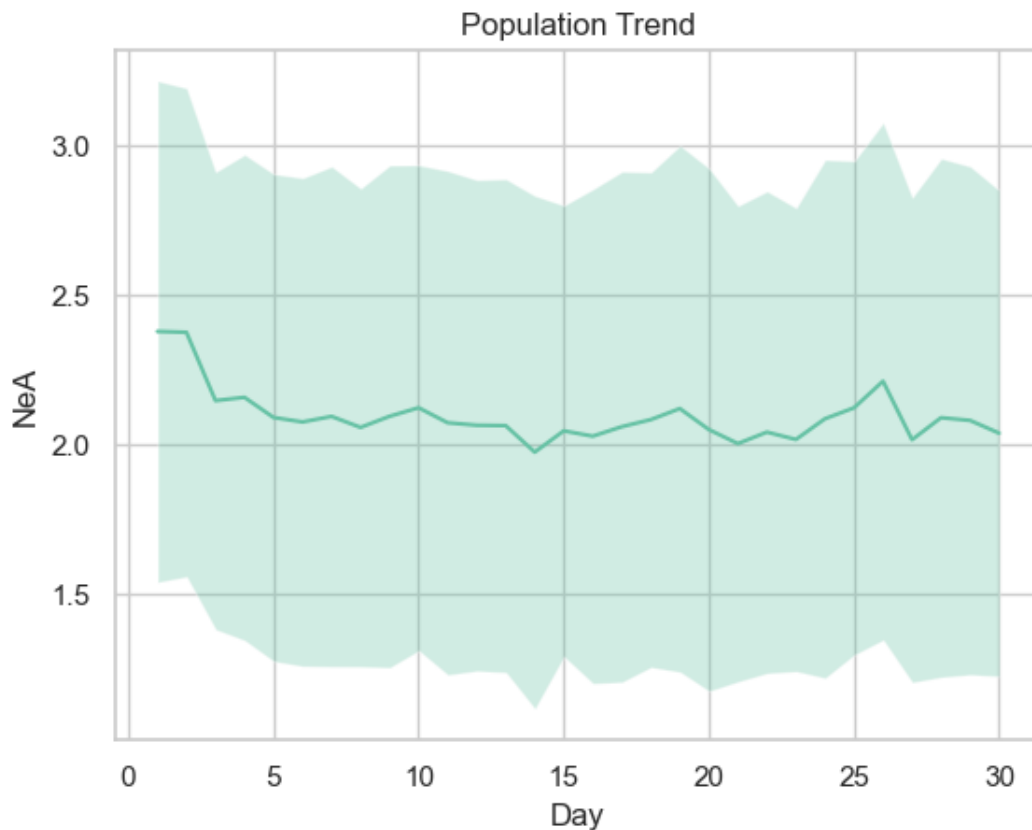
میانگین متغیرها در زمان:



این نمودار میانگین متغیرها در طول ۳۰ روز را نشان می‌دهد و مشخص است که هر متغیر الگوی زمانی نسبتاً پایدار اما با نوسانات جزئی دارد. PSS10 بالاترین میانگین را دارد و در تمام روزها در سطح بالاتری نسبت به سایر متغیرها قرار می‌گیرد. DailyStressor و EFcoping در سطح میانی قرار دارند و تغییرات روزانه‌ی ملایمی را نشان می‌دهند. NeA و RES پایین‌ترین میانگین‌ها را دارند و روند آن‌ها نسبتاً ثابت و کم‌نوسان است. این الگو بیانگر تفاوت سطح پایه‌ی متغیرها و ثبات نسبی آن‌ها در سطح گروهی است.

Population Trend Plot

Mean + SD band:



نمودار روند میانگین با نوار انحراف معیار میانگین NeA در سطح جمعیت را در طول ۳۰ روز نشان می‌دهد و ناحیه‌ی اطراف خط، باند $\text{Mean} \pm \text{SD}$ است. این باند بیان می‌کند که مقدار NeA برای بیشتر افراد در چه محدوده‌ای نوسان می‌کند. ثبات نسبی پهنای باند نشان می‌دهد که پراکندگی بین فردی در طول زمان تقریباً ثابت مانده. همچنین خط میانگین تقریباً افقی است، یعنی روند جمعیتی NeA تغییر قابل توجهی ندارد و بیشتر نوسانات در سطح درون فردی رخ می‌دهد.

Mean NeA Trend:

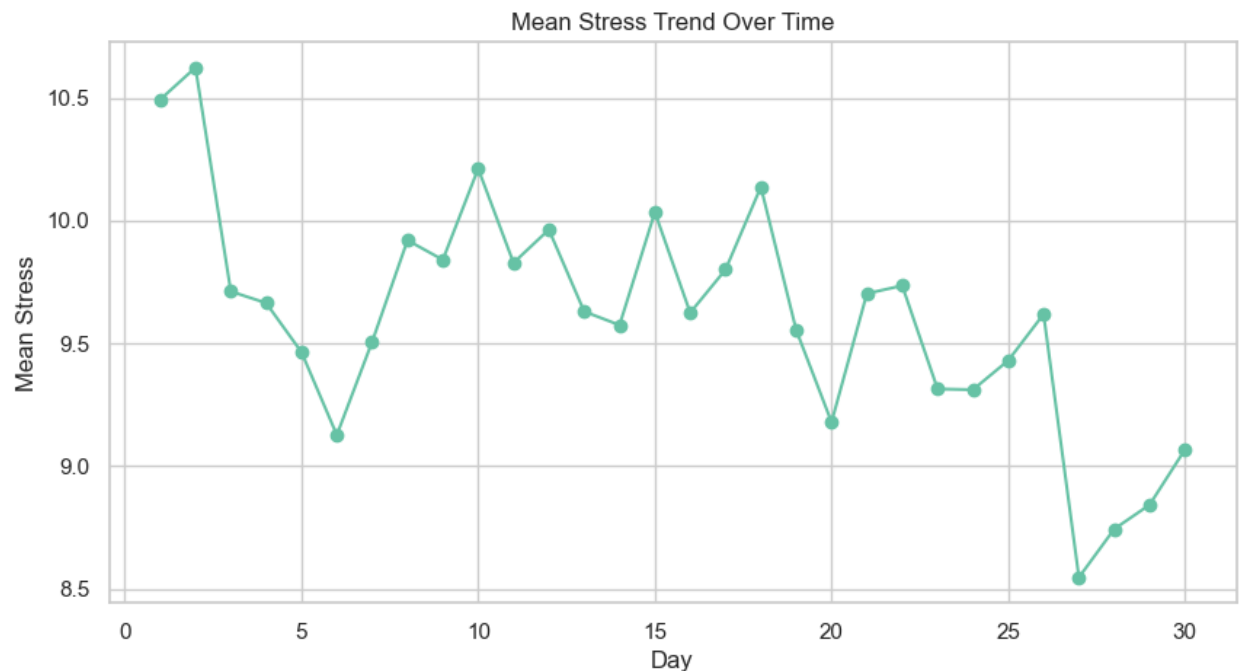


این نمودار روند میانگین NeA در سطح جمعیت را در طول ۳۰ روز نشان می‌دهد. مقدار میانگین در روزهای ابتدایی کمی بالاتر است اما خیلی سریع به محدوده‌ی پایدار نزدیک ۲.۰ می‌رسد. پس از آن، نوسانات کوچک و کوتاه‌مدت دیده می‌شود، اما هیچ روند افزایشی یا کاهشی معناداری در طول زمان شکل نمی‌گیرد. یک جهش کوتاه در حوالی روز ۲۵ دیده می‌شود، اما پایدار نیست و دوباره به سطح معمول بازمی‌گردد. این الگو نشان می‌دهد که هیجان منفی در سطح جمعیت تقریباً ثابت و کم‌نوسان است. پیام علمی:

- نمودار قبلی: هیجان منفی درون افراد متغیر است.
- نمودار جدید: هیجان منفی در سطح جمعیت پایدار است.

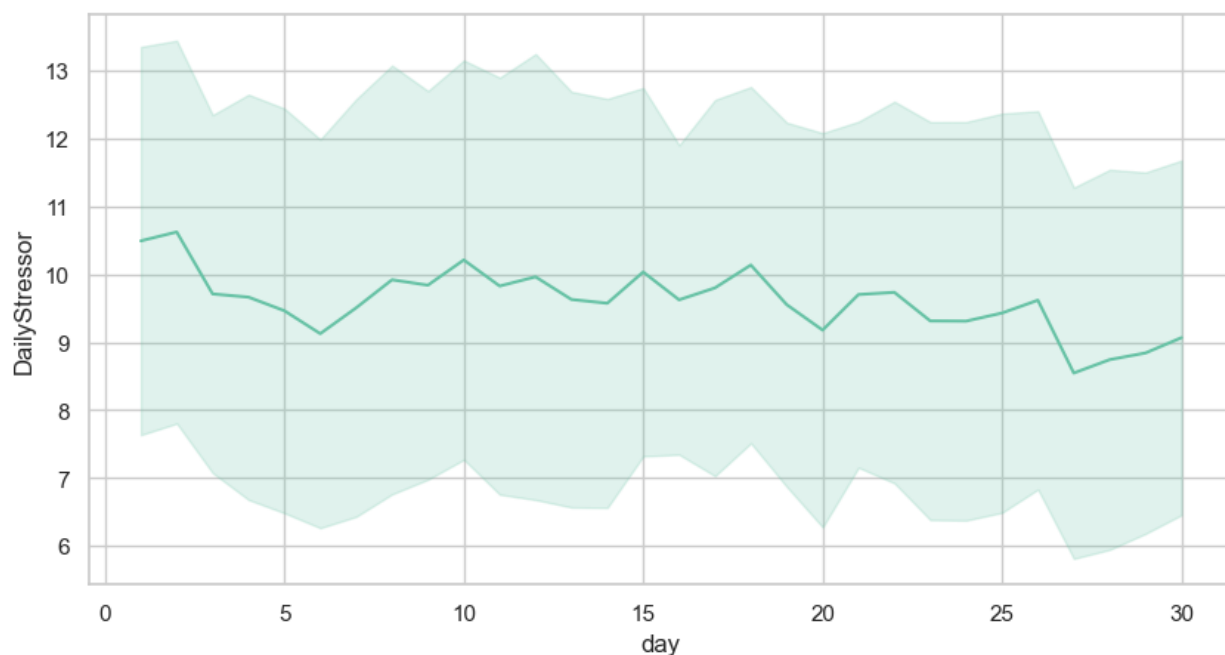
این دقیقاً همان تفاوت کلاسیک **Within-person vs. Between-person** در داده‌های طولی است.

Mean Stress Trend:



این نمودار روند میانگین استرس روزانه را در طول ۳۰ روز نشان می‌دهد و چند نکته علمی از آن قابل برداشت است:

- میانگین استرس در یک بازه‌ی نسبتاً محدود نوسان می‌کند و تغییرات شدید دیده نمی‌شود، که نشان‌دهنده‌ی ثبات نسبی سطح استرس در سطح جمعیت است.
- الگو دارای افزایش‌ها و کاهش‌های کوتاه‌مدت است؛ مثلاً چند قله و چند افت مشخص دیده می‌شود، اما هیچ روند صعودی یا نزولی پایدار شکل نمی‌گیرد.
- این نوسانات کوچک معمولاً نشان‌دهنده‌ی تغییرات موقعیتی روزانه هستند، نه تغییرات ساختاری در استرس.
- سطح میانگین استرس در اکثر روزها نزدیک به یک مقدار مرکزی ثابت باقی می‌ماند، که بیانگر پایداری بین‌فردی است.



این نمودار روند میانگین Daily Stressor را در سطح جمعیت نشان می‌دهد و چند تفاوت مهم با نمودارهای قبلی NeA دارد:

میانگین استرس روزانه در یک بازه‌ی بزرگ‌تر و پرنوسان‌تر نسبت به NeA حرکت می‌کند. قله‌ها و افت‌های مشخص‌تری دارد، یعنی استرس روزانه نسبت به هیجان منفی واکنش‌پذیرتر و موقعیتی‌تر است.

برخلاف NeA که میانگین آن تقریباً صاف و پایدار بود، این نمودار نشان می‌دهد که استرس روزانه نوسانات دوره‌ای و موج‌دار دارد. این یعنی استرس بیشتر تحت تأثیر شرایط روزمره تغییر می‌کند. پیام علمی:

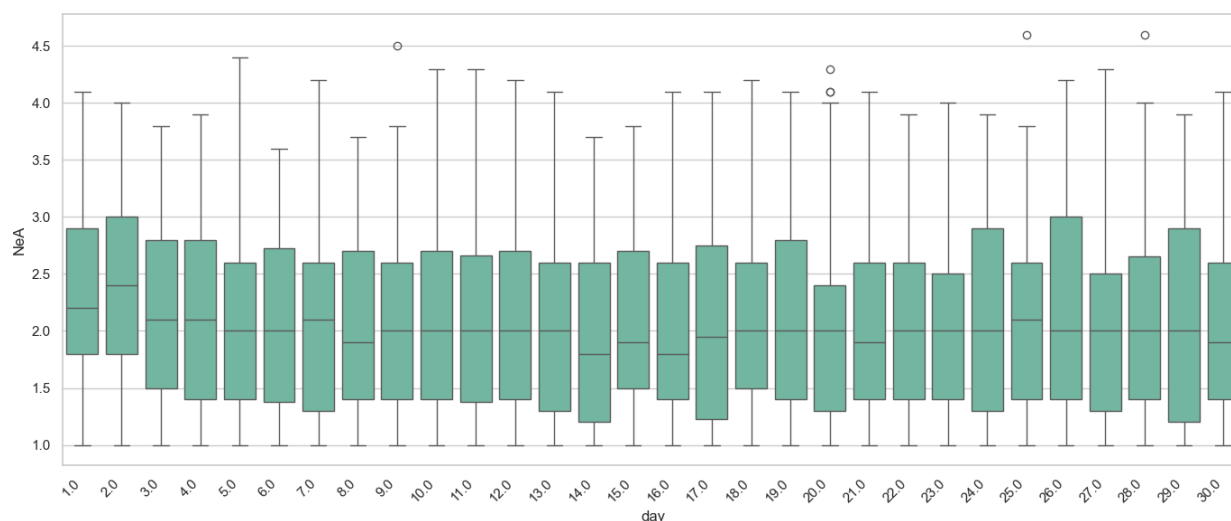
در سطح جمعیت، هیجان منفی پایدارتر است اما استرس روزانه متغیرتر. این تفاوت معمولاً در داده‌های روان‌شناختی دیده می‌شود:

- هیجان منفی بیشتر یک ویژگی نسبتاً پایدار است.
- استرس روزانه بیشتر یک تجربه‌ی موقعیتی و وابسته به رویدادهای روزمره است.

۳.۳.۳. بررسی تغییرات توزیع در طول زمان

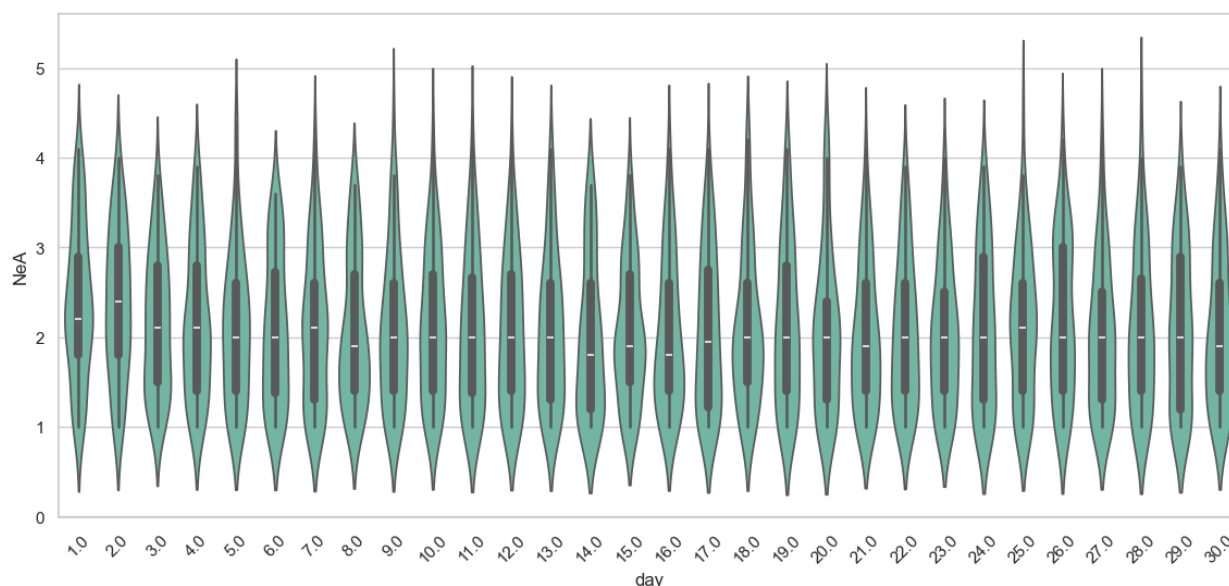
نمودارهای جعبه‌ای و ویولن نشان دادند که توزیع متغیرهای روانی در روزهای مختلف متفاوت است و این موضوع بیانگر تغییرات رفتاری افراد در طول زمان است.

Boxplot زمان برای NeA :



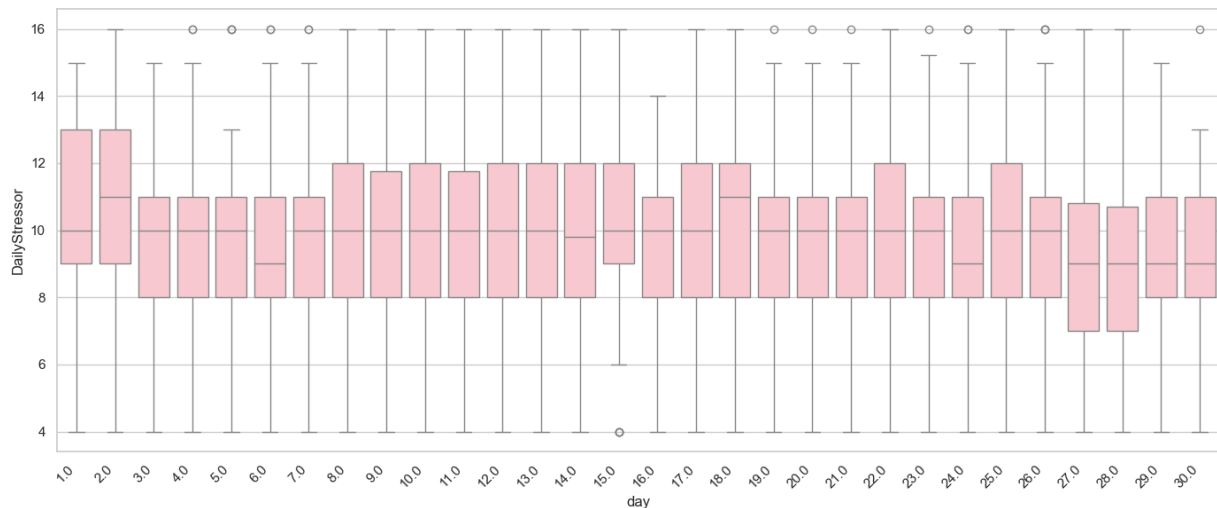
این نمودار توزیع روزانه‌ی متغیر NeA را با استفاده از نمودار جعبه‌ای برای هر روز نشان می‌دهد و امکان مقایسه‌ی ساختار توزیع در طول زمان را فراهم می‌کند. میانه‌ی NeA در اغلب روزها در سطح پایین قرار دارد و دامنه‌ی بین چارکی نسبتاً ثابت است، که بیانگر ثبات نسبی هیجان منفی در سطح جمعیت است. تغییرات اندازه‌ی جعبه‌ها نشان می‌دهد که میزان پراکندگی بین فردی در برخی روزها بیشتر و در برخی روزها کمتر است. وجود نقاط پرت در چند روز نشان‌دهنده‌ی تجربه‌ی هیجان منفی غیرمعمول در تعداد کمی از افراد است. این نمودار برخلاف نمودارهای روند میانگین، اطلاعات دقیق‌تری درباره‌ی پراکندگی و ناهمگونی روزانه ارائه می‌دهد. در مجموع، الگو نشان می‌دهد که هیجان منفی در سطح گروهی پایدار است اما تفاوت‌های بین فردی در طول روزها قابل مشاهده است.

Violin زمان برای NeA :



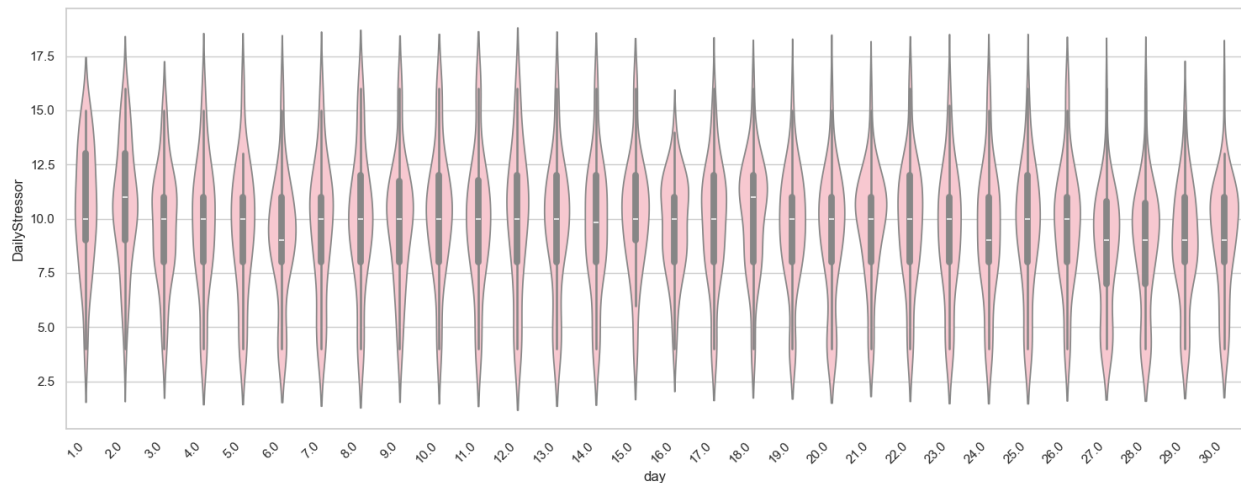
این نمودار توزیع روزانه‌ی متغیر NeA را با استفاده از ویولن پلات نشان می‌دهد و امکان مشاهده‌ی شکل توزیع، تراکم داده‌ها و میانه‌ی هر روز را فراهم می‌کند. پهنای هر ویولن بیانگر میزان تراکم داده‌ها در سطوح مختلف NeA است و جعبه‌ی داخلی نیز میانه و دامنه‌ی بین‌چارکی را مشخص می‌کند. الگوهای مشاهده‌شده نشان می‌دهد که توزیع NeA در بیشتر روزها نسبتاً فشرده و در سطوح پایین متمرکز است. تفاوت پهنای ویولن‌ها در روزهای مختلف بیانگر تغییرات پراکندگی بین‌فردی در طول زمان است. وجود بخش‌های پهن‌تر در برخی روزها نشان‌دهنده‌ی ناهمگونی بیشتر بین افراد است. در مجموع، این نمودار تصویری دقیق‌تر از ساختار توزیع روزانه ارائه می‌دهد و مکمل تحلیل‌های مبتنی بر میانگین است.

Boxplot زمان برای Stress :



این نمودار توزیع روزانه‌ی متغیر Daily Stressor را با استفاده از نمودار جعبه‌ای برای هر روز نمایش می‌دهد و امکان بررسی تغییرات پراکندگی و میانه‌ی استرس در طول زمان را فراهم می‌کند. اندازه‌ی جعبه‌ها نشان‌دهنده‌ی دامنه‌ی بین‌چارکی است و بیان می‌کند که بخش اصلی داده‌ها در هر روز در چه محدوده‌ای قرار داشته‌اند. میانه‌ی استرس در بیشتر روزها در سطح میانی قرار دارد و تغییرات آن نسبتاً محدود است. تفاوت اندازه‌ی جعبه‌ها و طول خط‌های بالا و پایین نشان می‌دهد که میزان پراکندگی بین‌فردی در برخی روزها بیشتر بوده است. وجود نقاط پرت در چند روز بیانگر تجربه‌ی استرس غیرمعمول توسط تعداد کمی از افراد است. این نمودار تصویری دقیق از ناهمگونی روزانه ارائه می‌دهد و مکمل تحلیل‌های مبتنی بر میانگین است.

Violin زمان برای Stress :



این نمودار توزیع روزانه‌ی متغیر Daily Stressor را با استفاده از ویولن پلات نمایش می‌دهد و امکان بررسی شکل توزیع، تراکم داده‌ها و میانه‌ی هر روز را فراهم می‌سازد. پهنای هر ویولن نشان‌دهنده‌ی میزان تراکم داده‌ها در سطوح مختلف استرس است و جعبه‌ی داخلی نیز میانه و دامنه‌ی بین‌چارکی را مشخص می‌کند. الگوهای مشاهده‌شده نشان می‌دهد که استرس روزانه در بیشتر روزها در یک بازه‌ی مشخص متمرکز است، اما میزان پراکندگی بین‌فردی در برخی روزها بیشتر است. تفاوت پهنای ویولن‌ها بیانگر تغییرات ناهمگونی افراد در تجربه‌ی استرس روزانه است. وجود بخش‌های پهن‌تر یا باریک‌تر در روزهای مختلف نشان می‌دهد که شدت و پراکندگی استرس تحت تأثیر شرایط روزانه تغییر می‌کند. این نمودار تصویری جامع‌تر از ساختار توزیع ارائه می‌دهد و مکمل تحلیل‌های مبتنی بر میانگین است.

توزیع متغیرها در طول زمان تغییر می‌کند.

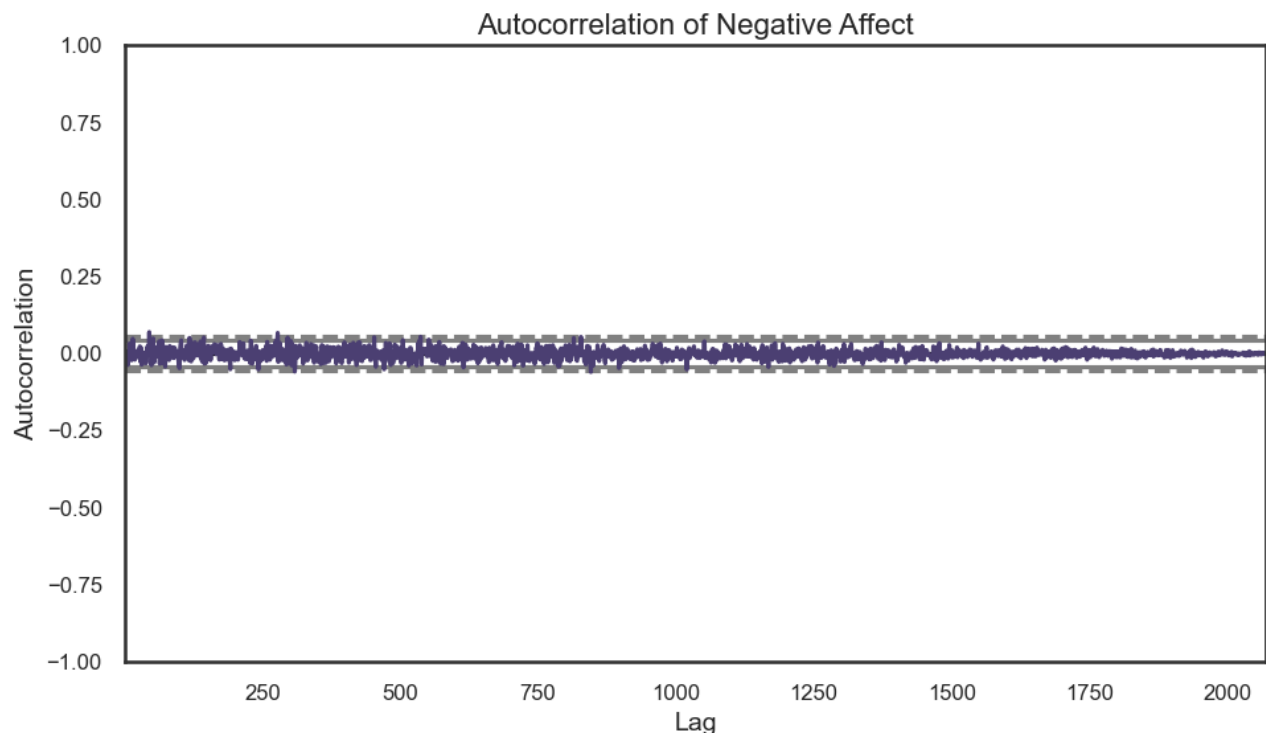
فرقش با نمودارهای قبلی چیه؟

- نمودارهای قبلی فقط میانگین کل افراد را نشان می‌دادند.
- اما این نمودار نشان می‌دهد افراد در هر روز چقدر با هم فرق داشتند.

۴.۳.۳. بررسی وابستگی زمانی

تحلیل خودهمبستگی و نمودارهای تأخیری نشان داد که مقادیر متغیرهای روانی در روزهای متوالی به یکدیگر وابسته هستند. این نتیجه اهمیت تحلیل طولی داده‌ها را تأیید می‌کند.

Autocorrelation plot:

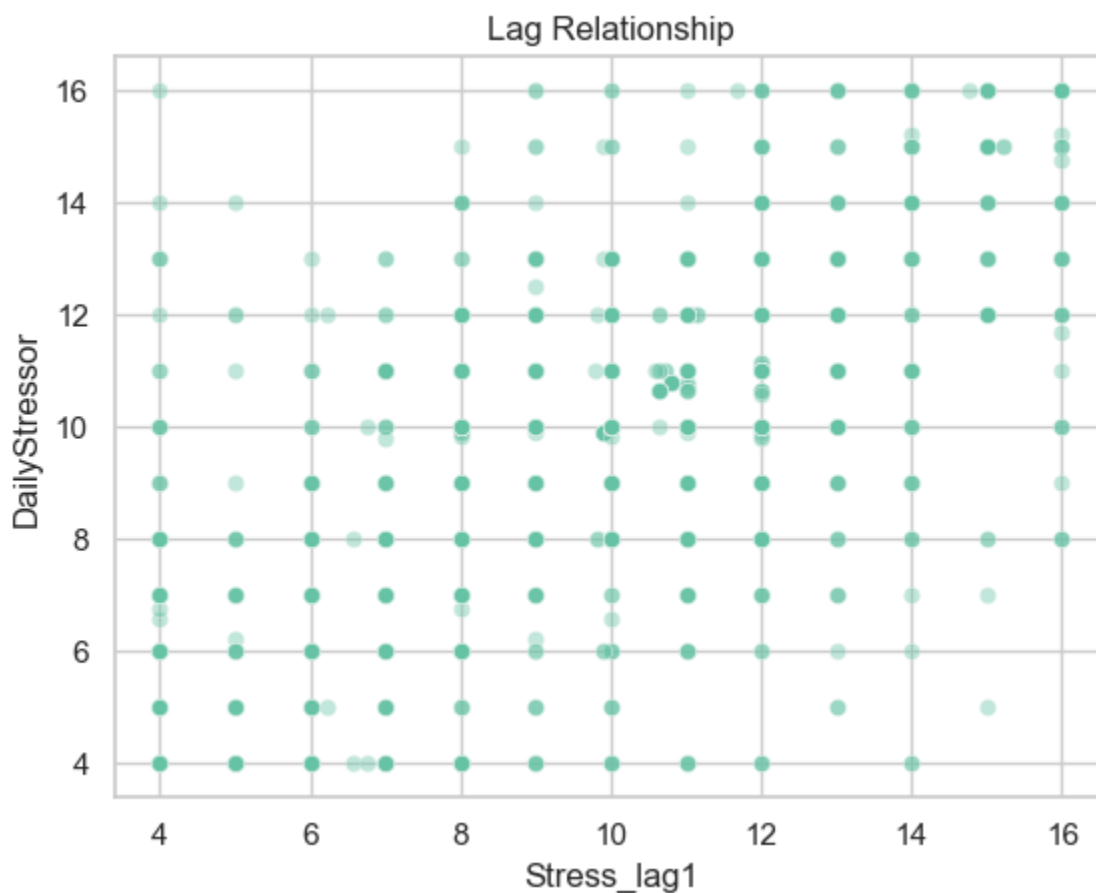


- مقدار اتوکورلیشن در همه‌ی لگ‌ها تقریباً نزدیک صفر است.
- هیچ قله‌ی معناداری خارج از خطوط اطمینان دیده نمی‌شود.

نمودار اتوکورلیشن نشان می‌دهد که مقدار *Negative Affect* در هیچ‌یک از لگ‌های زمانی دارای همبستگی معنادار با گذشته‌ی خود نیست و مقادیر تقریباً حول صفر نوسان می‌کنند. این الگو بیانگر آن است که هیجان منفی در این مجموعه داده وابستگی زمانی قابل توجهی ندارد و مقدار امروز

پیش‌بینی‌کننده‌ی مقدار روزهای قبل یا بعد نیست. با این حال، نبودِ اتوکورلیشن به معنای نامناسب بودن داده‌ها برای تحلیل طولی نیست؛ زیرا داده‌ها همچنان ماهیت تکرارشونده‌ی روزانه دارند و امکان بررسی روندها، نوسانات درون‌فردی، تفاوت‌های بین‌فردی و الگوهای توزیع روزانه را فراهم می‌کنند. علاوه بر این، چنین ساختاری مانع انجام تحلیل‌های اکتشافی طولی یا خوشه‌بندی نمی‌شود و حتی تفسیر نتایج را ساده‌تر می‌کند. بنابراین، داده‌ها کاملاً برای اهداف پروژه مناسب هستند و می‌توان از آن‌ها برای تحلیل‌های توصیفی و الگوهای گروهی استفاده کرد.

Lag plot :



این نمودار رابطه‌ی بین استرس روز قبل (Stress_lag1) و استرس امروز (DailyStressor) را نشان می‌دهد. پراکندگی نقاط نشان می‌دهد که هیچ الگوی مشخص، خطی یا قابل‌پیش‌بینی بین این دو متغیر وجود ندارد. به بیان ساده: استرس امروز به استرس دیروز وابسته نیست و مقدار روز قبل نمی‌تواند مقدار امروز را پیش‌بینی کند.

این الگو تأیید می‌کند که در این داده‌ها وابستگی زمانی قوی وجود ندارد؛ یعنی تغییرات استرس بیشتر تحت تأثیر شرایط همان روز است، نه روزهای قبل. این نتیجه با نمودار اتوکورلیشن هم سازگار است.

آیا این برای پروژه مشکل ایجاد می‌کند؟

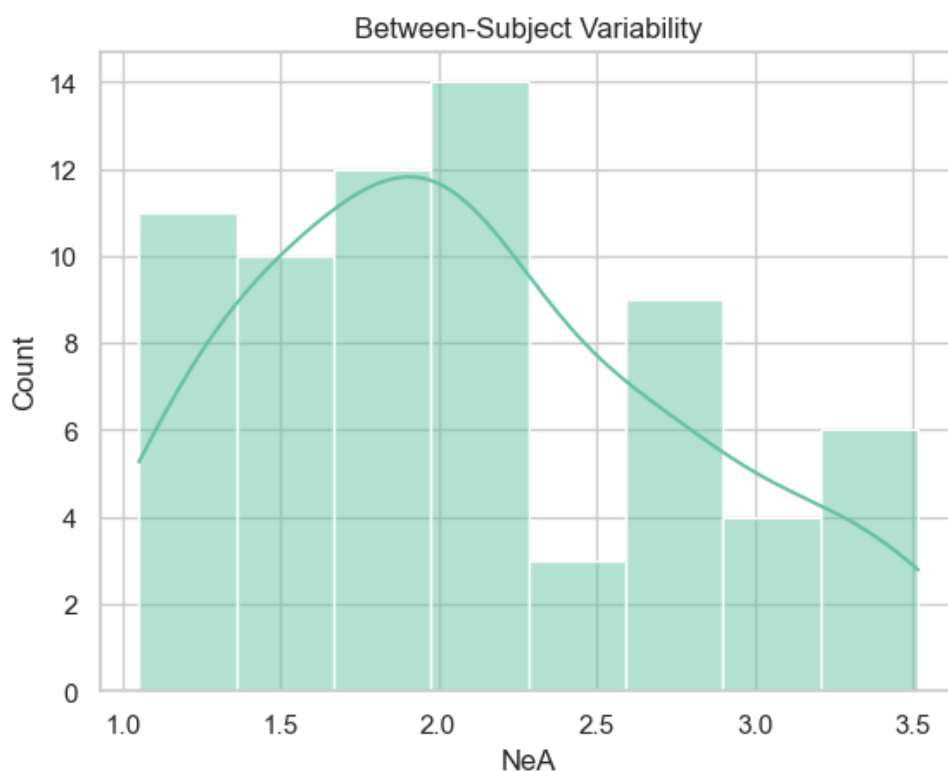
نه، اصلاً. نبود وابستگی زمانی به معنای نامناسب بودن داده‌ها نیست.

در نتیجه نمودار پراکندگی نشان می‌دهد که رابطه‌ی معناداری بین استرس روز قبل و استرس امروز وجود ندارد و داده‌ها فاقد وابستگی زمانی قوی هستند. با این حال، این ویژگی مانعی برای تحلیل طولی نیست، زیرا داده‌ها همچنان ساختار تکرار روزانه دارند و امکان بررسی روندها، نوسانات فردی و تفاوت‌های بین فردی را فراهم می‌کنند. بنابراین، داده‌ها برای تحلیل اکتشافی طولی و روش‌های خوشه‌بندی کاملاً مناسب هستند.

۴.۳. بررسی تفاوت های فردی

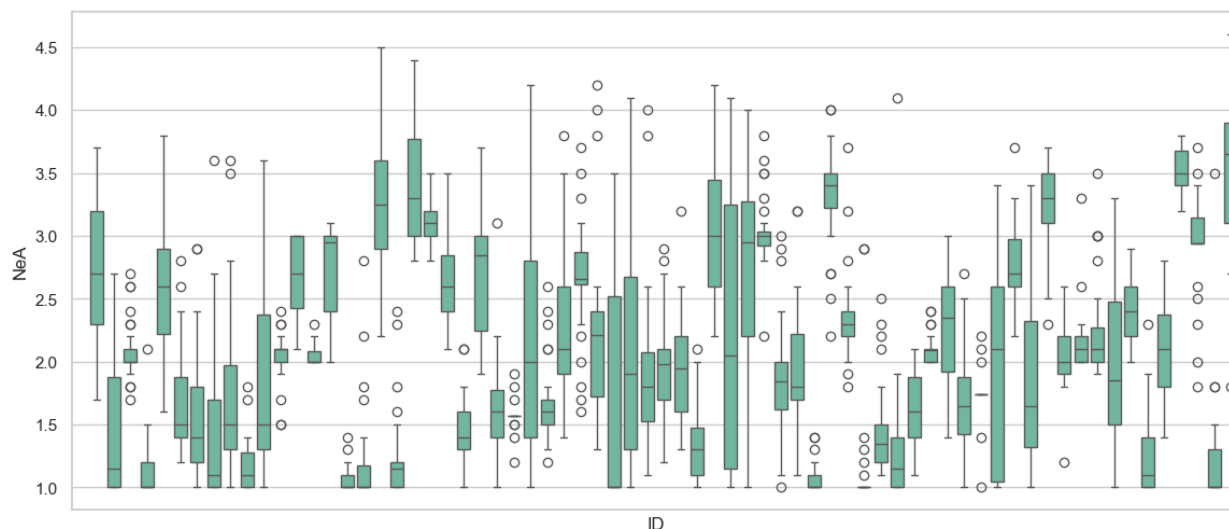
برای بررسی تفاوت بین افراد، میانگین متغیرهای روانی برای هر فرد محاسبه شد. نتایج نشان داد که افراد دارای سطوح متفاوتی از عاطفه منفی و استرس هستند. همچنین توزیع شیب تغییرات زمانی نشان داد که برخی افراد روند افزایشی و برخی روند کاهشی در متغیرهای روانی داشتند.

Distribution میانگین افراد:



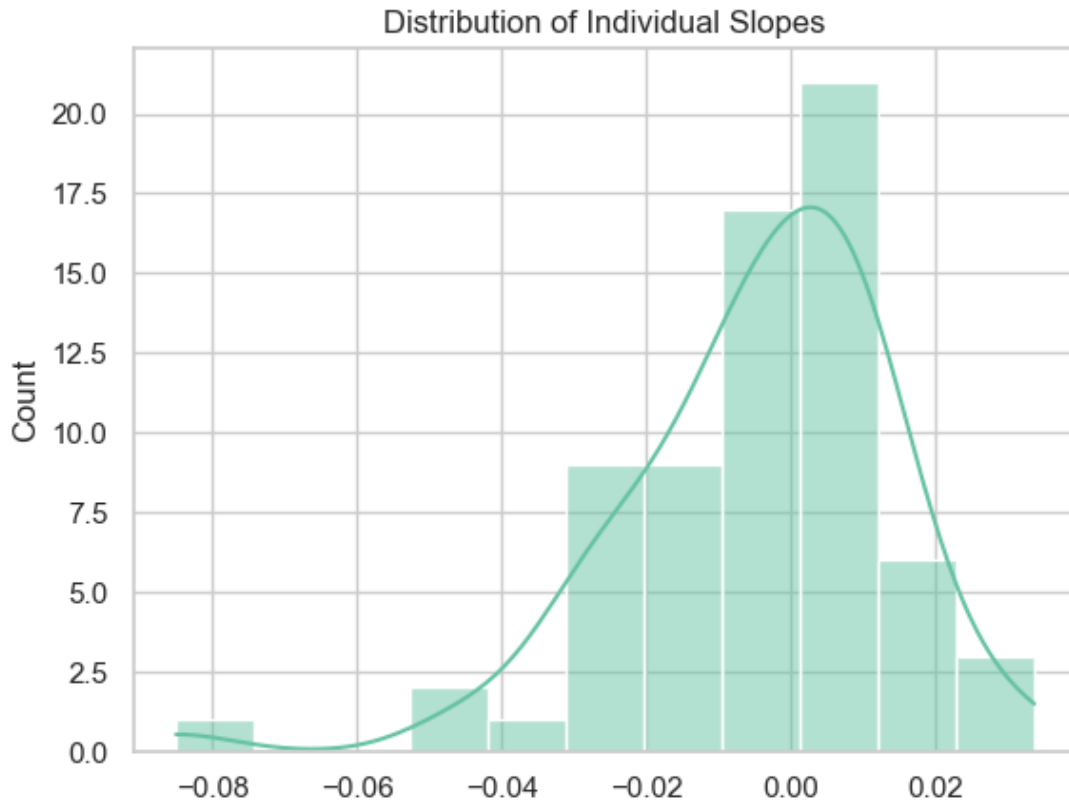
این نمودار توزیع مقادیر NeA را در سطح بین فردی نشان می دهد؛ یعنی اینکه افراد مختلف به طور کلی چه مقدار NeA داشته اند. هیستوگرام شکل کلی توزیع را نمایش می دهد و منحنی چگالی نشان می دهد که بیشترین تراکم داده ها در حدود مقدار ۲ قرار دارد و سپس به تدریج کاهش می یابد. این الگو نشان می دهد که افراد از نظر NeA با هم تفاوت دارند و توزیع نسبتاً متمرکز اما ناهمگن است.

Boxplot NeA بر حسب ID :



این نمودار جعبه‌ای مقدار NeA را برای هر فرد نشان می‌دهد و تفاوت‌های بین فردی را به خوبی مشخص می‌کند. برای هر ID، میانه، دامنه‌ی بین چارکی و نقاط پرت قابل مشاهده است و نشان می‌دهد که سطح کلی NeA بین افراد یکسان نیست. برخی افراد مقدارهای پایین‌تری دارند و برخی مقدارهای بالاتر، که بیانگر ناهمگونی پایدار بین افراد است. این الگو طبیعی و رایج در داده‌های روان‌شناختی است و نشان می‌دهد افراد در تجربه‌ی هیجان منفی با هم تفاوت دارند. چنین ساختاری برای تحلیل طولی، بررسی روندها و حتی خوشه‌بندی کاملاً مناسب است. این نمودار بخشی از تحلیل بین فردی پروژه را تکمیل می‌کند و به خوبی تفاوت‌های فردی را نشان می‌دهد.

Distribution شیب افراد:



این نمودار توزیع شیب‌های فردی را نشان می‌دهد و بیان می‌کند که تغییرات NeA در طول زمان برای هر فرد چه روندی داشته است. بیشتر شیب‌ها نزدیک صفر قرار گرفته‌اند، یعنی افراد به‌طور متوسط تغییر زمانی قابل‌توجهی در NeA تجربه نکرده‌اند. توزیع تقریباً نرمال است و نشان می‌دهد که افزایش یا کاهش شدید در NeA بین افراد نادر بوده است. این الگو تأیید می‌کند که NeA در سطح زمانی برای افراد پایدار نیست و تغییرات روزانه بیشتر تصادفی یا موقعیتی‌اند.

افراد الگوهای متفاوت دارند.

۵.۳. ساخت شاخص های رفتاری افراد

برای مقایسه افراد، داده های روزانه به مجموعه ای از شاخص های خلاصه تبدیل شدند. این شاخص ها شامل:

- میانگین استرس روزانه
- میزان نوسان استرس
- میانگین عاطفه منفی
- میانگین استرس ادراک شده
- میانگین تاب آوری

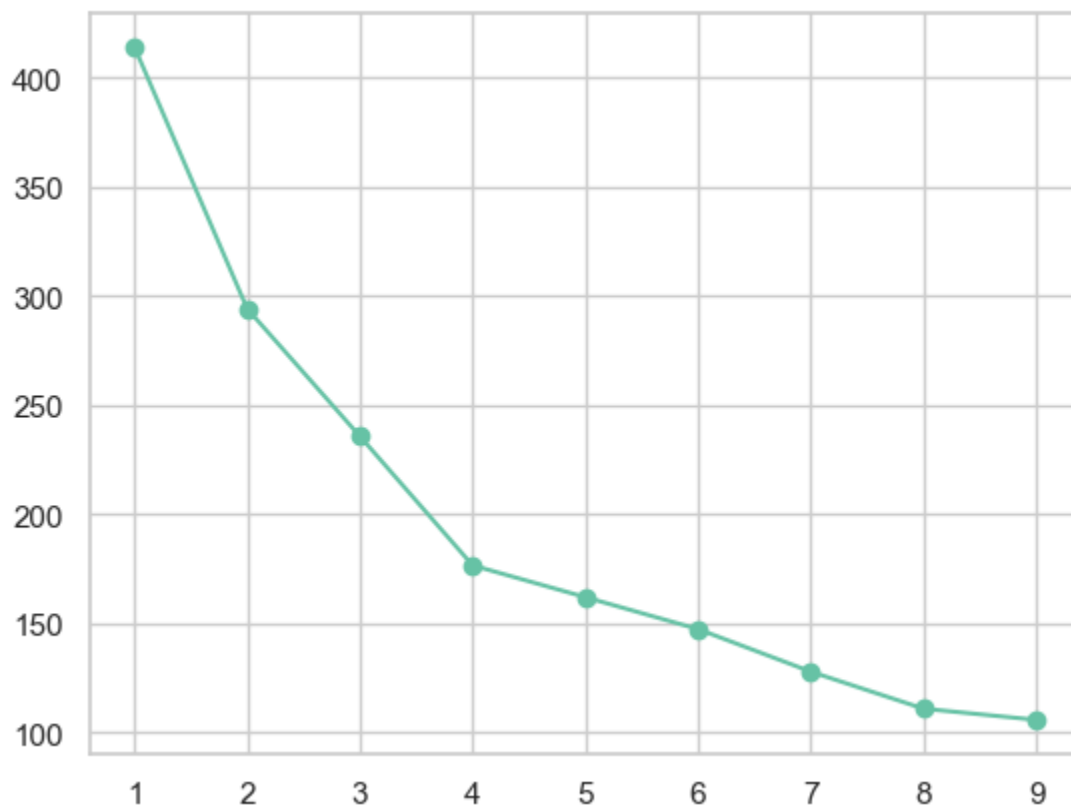
این شاخص ها نمایانگر پروفایل رفتاری هر فرد بودند.

۶.۳. خوشه بندی افراد

برای شناسایی گروه های رفتاری مشابه، افراد بر اساس شاخص های رفتاری خوشه بندی شدند. تعداد خوشه ها با استفاده از نمودار آرنج تعیین شد و سه خوشه به عنوان ساختار بهینه انتخاب شد.

Elbow Plot برای تعیین تعداد بهینه ی خوشه ها در روش هایی مثل k-means استفاده می شود. در این نمودار مقدار خطای درون خوشه ای (WSS) برای تعداد خوشه های مختلف محاسبه و رسم می شود. با افزایش تعداد خوشه ها، WSS کاهش می یابد، اما از یک نقطه به بعد سرعت کاهش کند می شود. این نقطه ی تغییر شیب که شبیه «آرنج» نمودار است، تعداد خوشه های مناسب را نشان می دهد. انتخاب این نقطه باعث می شود مدل نه بیش از حد ساده باشد و نه بیش از حد پیچیده. بنابراین Elbow Plot ابزاری استاندارد برای انتخاب تعداد خوشه ها در تحلیل خوشه ای است.

Elbow Plot:



سه خوشه مناسب ترین ساختار است.

چرا ۳ خوشه انتخاب خوبی است؟

- در Elbow Plot معمولاً دنبال نقطه‌ای می‌گردیم که کاهش خطا بعد از آن کند می‌شود.
 - در نمودار، شیب از $k=1$ به $k=2$ و از $k=2$ به $k=3$ زیاد است (یعنی خوشه‌بندی واقعاً بهتر می‌شود).
 - اما از $k=3$ به بعد، کاهش WSS خیلی کم می‌شود و نمودار تقریباً صاف می‌شود.
 - این دقیقاً همان جایی است که به آن می‌گوییم Elbow یا آرنج.
- پس انتخاب ۳ خوشه نه زیاد است، نه کم — یک نقطه‌ی بهینه است.

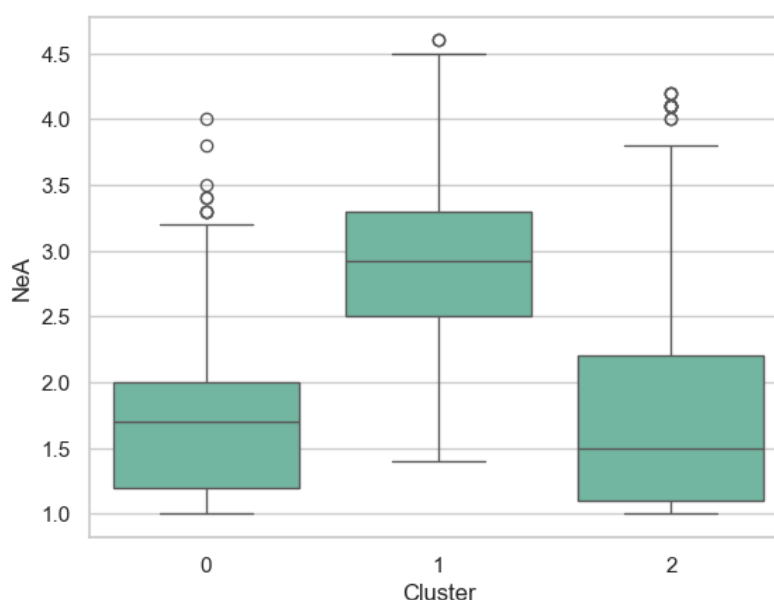
۷.۳. تحلیل الگوهای رفتاری خوشه ها (نتیجه نهایی پروژه)

نتایج نشان داد که خوشه‌ها دارای الگوهای زمانی متفاوتی هستند. برخی خوشه‌ها سطح بالاتری از عاطفه منفی داشتند و برخی دیگر سطح پایین‌تر و پایدارتر بودند.

همچنین نمودارهای حرارتی نشان دادند که الگوهای زمانی افراد در هر خوشه متفاوت است.

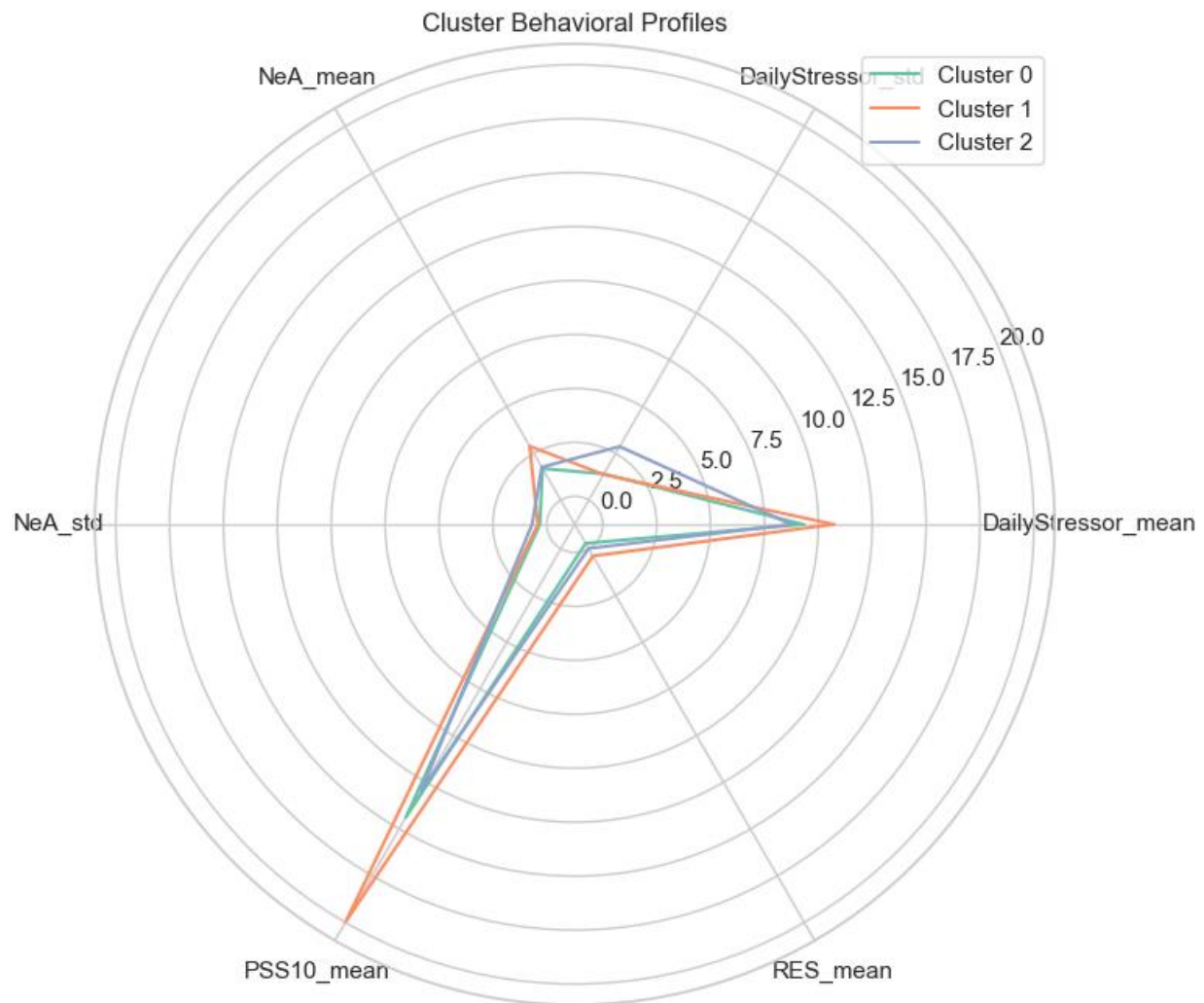
۱.۷.۳. مقایسه سطح متغیرها در خوشه ها

Boxplot خوشه‌ها :



این نمودار توزیع NeA را در سه خوشه‌ی استخراج‌شده نشان می‌دهد و تفاوت سطح هیجان منفی بین خوشه‌ها را مشخص می‌کند. خوشه‌ی ۱ بالاترین میان‌ه‌ی NeA را دارد و نشان می‌دهد افراد این گروه به‌طور کلی هیجان منفی بیشتری تجربه کرده‌اند. خوشه‌ی ۲ پایین‌ترین سطح NeA را دارد و پراکندگی آن نسبتاً گسترده است. خوشه‌ی ۰ در سطح میانی قرار می‌گیرد و چند نقطه‌ی پرت در مقادیر بالاتر دارد. این الگو تأیید می‌کند که خوشه‌بندی توانسته تفاوت‌های بین‌فردی در NeA را به‌خوبی تفکیک کند. بنابراین سه خوشه انتخاب مناسبی بوده و ساختار معناداری در داده‌ها ایجاد کرده است.

Radar Plot:



این نمودار چطور کشیده می شود؟

الف) انتخاب متغیرها

اول میانگین و انحراف معیار چند متغیر رفتاری برای هر خوشه محاسبه می شود، مثل:

• NeA_mean

• NeA_std

• PSS10_mean

• RES_mean

• DailyStressor_mean

• DailyStressor_std

این‌ها ویژگی‌های خلاصه‌شده‌ی هر خوشه هستند.

ب) نرمال‌سازی یا مقیاس‌بندی

چون مقیاس متغیرها متفاوت است، معمولاً داده‌ها مقیاس‌بندی می‌شوند تا همه روی یک بازه قابل مقایسه قرار بگیرند.

ج) رسم نمودار رادار

• هر محور یک متغیر است.

• مقدار هر خوشه روی هر محور مشخص می‌شود.

• نقاط هر خوشه به صورت یک چندضلعی بسته به هم وصل می‌شوند.

• هر خوشه یک رنگ جداگانه دارد.

این باعث می‌شود شکل کلی رفتار هر خوشه قابل مشاهده باشد.

تیپ رفتاری هر خوشه بر اساس نمودار رادار

خوشه ۱ - تیپ هیجان منفی بالا و استرس پذیر

- بالاترین مقدار NeA_mean = افراد این خوشه به طور کلی هیجان منفی بیشتری تجربه می کنند.
- NeA_std نسبتاً بالا = نوسانات هیجان منفی زیاد است.
- $PSS10_mean$ بالا = استرس ادراک شده ی بیشتری دارند.
- RES_mean پایین تر = از بقیه تاب آوری کمتر.
- $DailyStressor_mean$ بالا = رویدادهای استرس زای روزانه بیشتر.
- $DailyStressor_std$ بالا = نوسان زیاد در استرس روزانه.

تیپ رفتاری: افرادی با هیجان منفی بالا، استرس پذیر، کم تاب آور و دارای نوسانات زیاد در تجربه ی استرس. این خوشه «آسیب پذیرترین» گروه است.

خوشه ۰ - تیپ متوسط و متعادل

- NeA_mean در حد میانی / نه خیلی بالا، نه خیلی پایین.
- NeA_std متوسط / نوسانات هیجان منفی کنترل شده.
- $PSS10_mean$ متوسط / استرس ادراک شده معمولی.
- RES_mean متوسط / تاب آوری نه بالا، نه پایین.
- $DailyStressor_mean$ متوسط / میزان استرس روزانه معمولی.

• **DailyStressor_std** متوسط / نوسان استرس روزانه قابل پیش‌بینی.

تیپ رفتاری: افرادی با الگوی هیجانی و استرسی متعادل. نه خیلی آسیب‌پذیر، نه خیلی مقاوم. این خوشه «گروه میانی» است.

خوشه ۲ - تیپ هیجان منفی پایین و تاب‌آور

• **NeA_mean** پایین‌ترین مقدار / افراد این خوشه هیجان منفی کمی دارند.

• **NeA_std** پایین / نوسانات هیجان منفی کم و پایدار.

• **PSS10_mean** پایین / استرس ادراک شده کم.

• **RES_mean** بالاتر از بقیه / تاب‌آوری بالا.

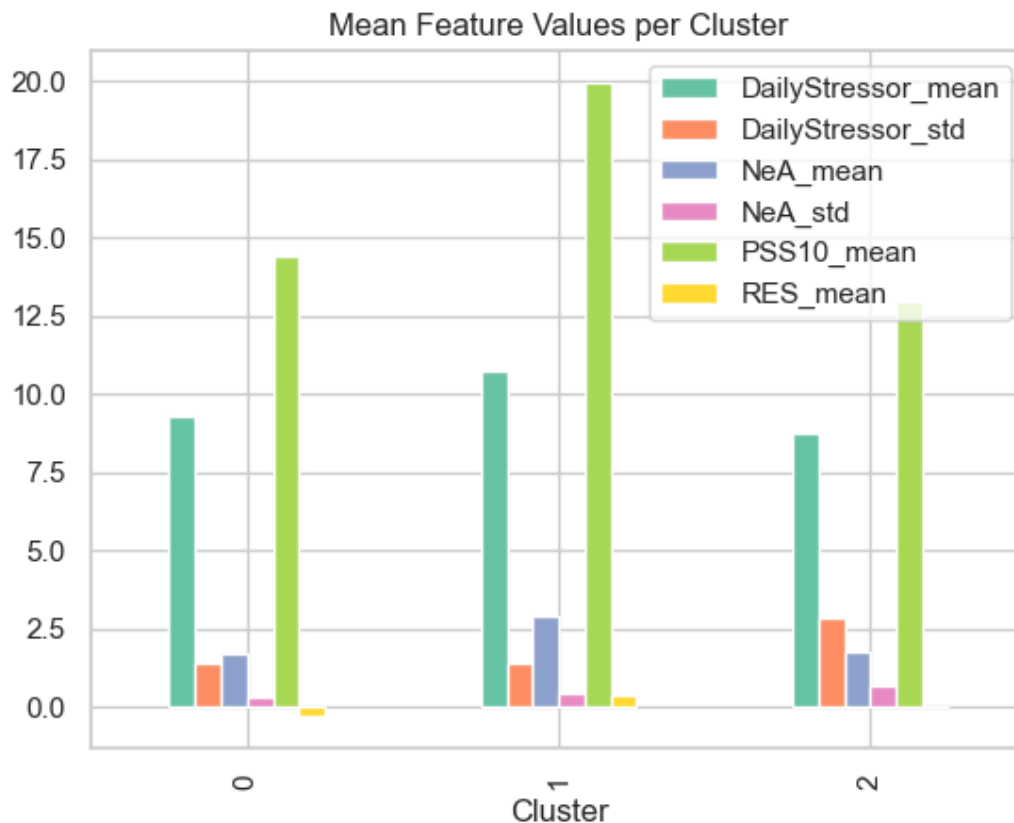
• **DailyStressor_mean** پایین / رویدادهای استرس‌زای روزانه کمتر.

• **DailyStressor_std** پایین / نوسان کم در استرس روزانه.

تیپ رفتاری: افرادی آرام، پایدار، کم‌استرس و تاب‌آور. این خوشه سالم‌ترین و مقاوم‌ترین گروه است.

نمودار رادار نشان می‌دهد که خوشه‌بندی بر اساس **NeA** توانسته سه تیپ رفتاری متمایز ایجاد کند. خوشه‌ی ۱ نمایانگر افراد با هیجان منفی بالا، استرس ادراک شده زیاد و تاب‌آوری پایین است. خوشه‌ی ۰ الگوی متوسط و متعادلی دارد و در هیچ شاخصی به‌طور افراطی بالا یا پایین نیست. خوشه‌ی ۲ کمترین سطح هیجان منفی، پایین‌ترین استرس و بالاترین تاب‌آوری را نشان می‌دهد. این الگو تأیید می‌کند که خوشه‌ها نه تنها از نظر **NeA**، بلکه از نظر شاخص‌های روان‌شناختی دیگر نیز پروفایل‌های رفتاری متفاوت و قابل تفسیر دارند.

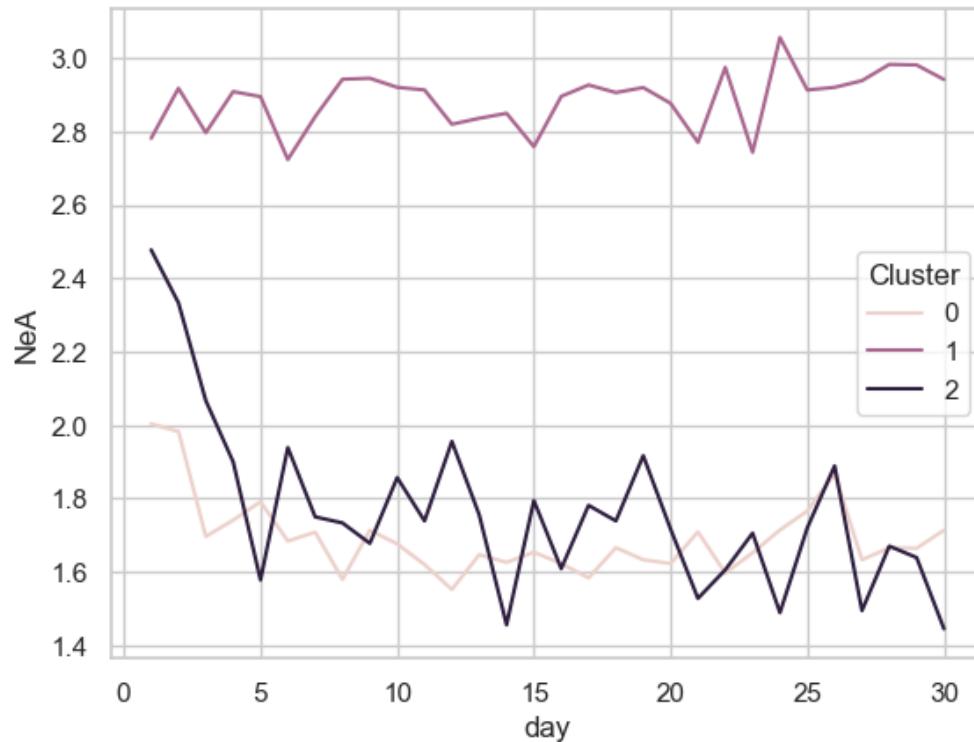
Grouped Bar Chart:



- خوشه ۱ تقریباً در تمامی ویژگی‌ها بالاترین مقدار را دارد، مخصوصاً در PSS10_mean و NeA_mean. این یعنی افراد این خوشه استرس ادراک شده و هیجان منفی بیشتری دارند و از نظر روان‌شناختی آسیب‌پذیرترند.
- خوشه ۰ پایین‌ترین مقادیر را در اکثر ویژگی‌ها دارد، به‌خصوص در NeA_mean و DailyStressor_mean. این خوشه نمایانگر افراد آرام‌تر، کم‌استرس‌تر و پایدارتر است.
- خوشه ۲ معمولاً بین این دو قرار می‌گیرد؛ نه به اندازه‌ی خوشه ۱ بالا و نه به اندازه‌ی خوشه ۰ پایین. این خوشه تیپ «متوسط و نوسانی» را نشان می‌دهد. یک خوشه‌ی پرخطر (۱)، یک خوشه‌ی کم‌خطر و آرام (۰)، و یک خوشه‌ی متوسط (۲)

۲.۷.۳. تحلیل مسیر زمانی خوشه ها

Cluster trajectory:



تیپ زمانی هر خوشه بر اساس روند ۳۰ روزه NeA

خوشه ۱ - هیجان منفی بالا و پایدار

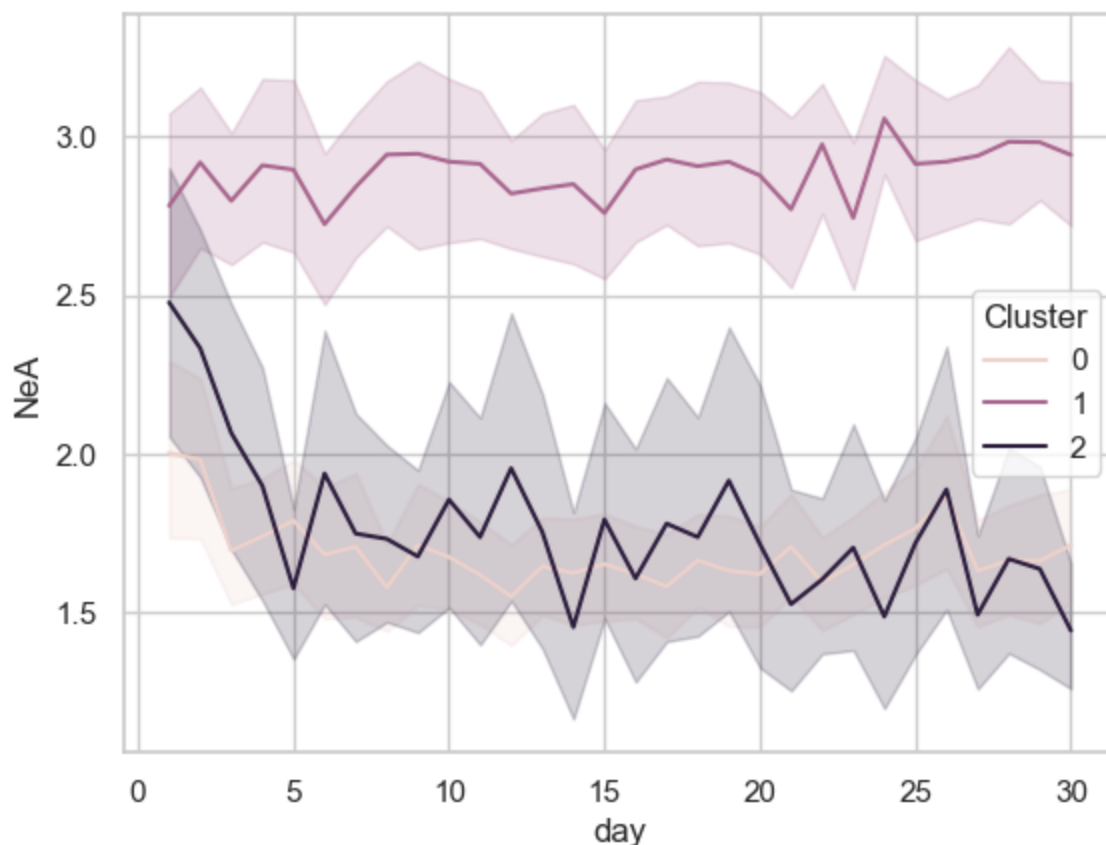
- بالاترین سطح NeA را در تمام ۳۰ روز دارد.
 - تقریباً هیچ افت و خیز جدی ندارد؛ خط روند صاف و پایدار است.
 - یعنی افراد این خوشه به طور مزمن هیجان منفی بالایی دارند و این وضعیت در طول زمان تغییر نمی کند.
- تیپ زمانی: افراد با هیجان منفی مزمن، پایدار و مقاوم به تغییر. این گروه همان خوشه ی «آسیب پذیر» است که قبلاً هم در رادار دیدیم.

خوشه ۰ - هیجان منفی پایین و نسبتاً پایدار

- سطح NeA پایین تر از بقیه خوشه‌ها است.
- روند نسبتاً صاف و بدون نوسان شدید دارد.
- یعنی افراد این خوشه هیجان منفی کمی دارند و این وضعیت ثابت می‌ماند.
- تیپ زمانی: افراد آرام، کم‌هیجان منفی و پایدار. این خوشه همان گروه «متعادل» است.

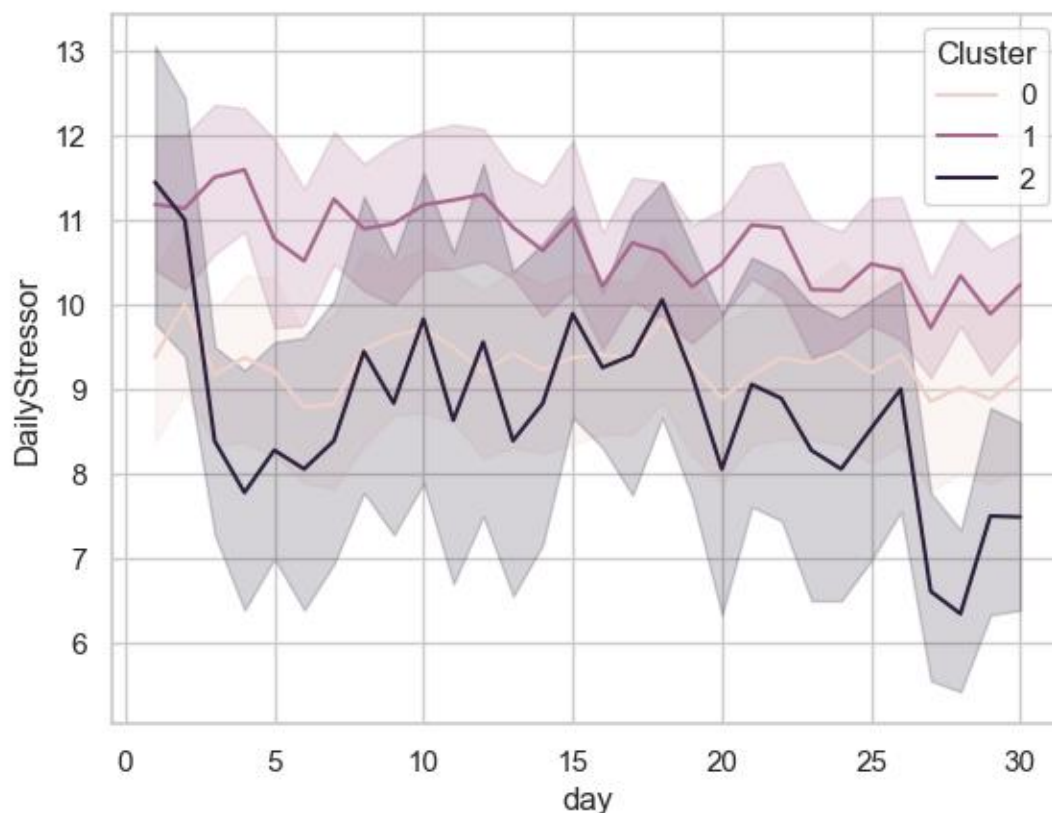
خوشه ۲ - هیجان منفی متوسط با نوسان و کاهش تدریجی

- در روزهای اول مقدار NeA کمی بالاتر است.
- سپس یک کاهش تدریجی دیده می‌شود.
- نوسانات بیشتری نسبت به خوشه ۰ و ۱ دارد.
- یعنی افراد این خوشه در شروع هیجان منفی بیشتری دارند اما در طول زمان بهتر می‌شوند.
- تیپ زمانی: افراد با هیجان منفی متوسط، نوسانی و رو به بهبود. این خوشه همان گروه «تاب‌آورتر» است که در رادار هم RES بالاتری داشت.



نمودار اول روند میانگین NeA را برای هر خوشه در طول ۳۰ روز نشان می‌دهد، در حالی که این نمودار علاوه بر روند، میزان نوسان و پراکندگی درون خوشه‌ای را نیز نمایش می‌دهد. باندهای سایه‌دار در نمودار دوم امکان مقایسه‌ی ثبات و تغییرپذیری خوشه‌ها را فراهم می‌کنند. بنابراین نمودار دوم اطلاعات دقیق‌تری درباره‌ی پایداری زمانی و ناهمگونی افراد در هر خوشه ارائه می‌دهد.

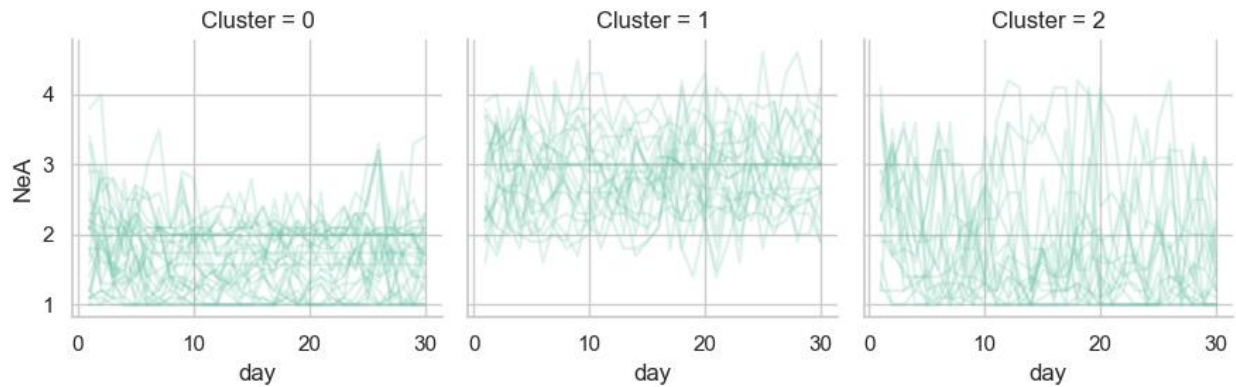
- خوشه ۲ فقط نوسانی نیست، نوسانش زیاد است.
- خوشه ۰ فقط پایدار نیست، پراکندگی‌اش هم کم است.
- خوشه ۱ فقط بالا نیست، ثبات درونی‌اش هم زیاد است.



در نمودار DailyStressor :

- خوشه ۱ استرس بالاتر و پایدار دارد.
 - خوشه ۰ کمترین استرس و پایدارترین روند را دارد.
 - خوشه ۲ نوسان زیاد دارد و بعد از روز اول افت می‌کند.
- این یعنی هیجان منفی و استرس روزانه همبسته‌اند، اما کاملاً هم‌مسیر نیستند.
و این دقیقاً همان چیزی است که در روان‌شناسی انتظار داریم.

۳.۷.۳. بررسی تفاوت افراد داخل خوشه



✓ این نمودار time series خام افراد را نشان می‌دهد یعنی:

- هر خط = یک فرد
 - هر پنل = یک خوشه
 - محور زمان = ۳۰ روز
 - مقدار NeA = هیجان منفی روزانه
- پس این نمودار دارد نشان می‌دهد افراد داخل هر خوشه چطور در طول زمان رفتار می‌کنند.

✓ این نمودار چه چیزی را درباره‌ی خوشه‌ها می‌فهماند؟

♦ خوشه

- خطوط نزدیک به هم
- نوسان کم
- سطح NeA پایین یعنی افراد این خوشه شبیه هم‌اند و هیجان منفی پایدار و کم دارند.

خوشه ۱

- خطوط بالا
- نوسان کم
- سطح NeA بالا یعنی افراد این خوشه همگی هیجان منفی مزمن و پایدار دارند.

خوشه ۲

- خطوط با نوسان زیاد
- سطح متوسط
- الگوهای متنوع یعنی این خوشه ناهمگن تر است و افراد رفتارهای نوسانی تری دارند

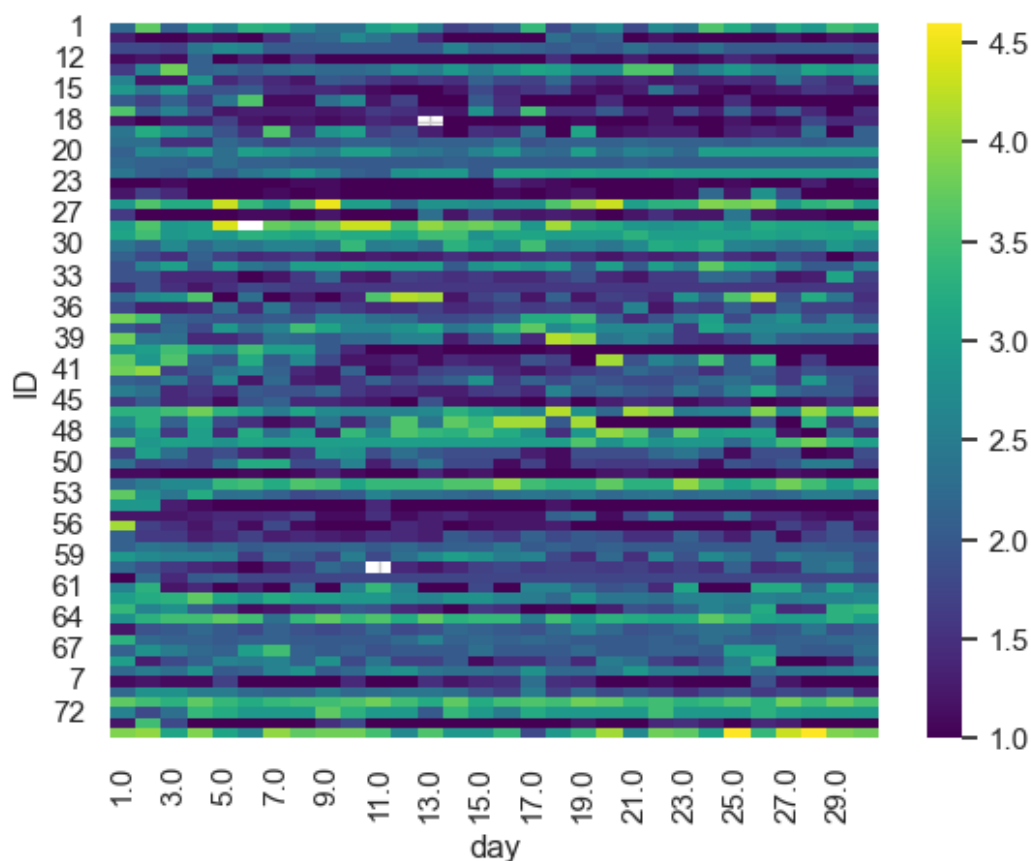
✓ چرا این نمودار مهم تر از بقیه است؟

چون نشان می دهد:

- خوشه بندی فقط یک میانگین مصنوعی نیست.
- افراد واقعاً الگوهای مشابهی دارند.
- تفاوت خوشه ها درون فردی و بین فردی قابل مشاهده است.
- خوشه ۲ واقعاً نوسانی تر است.
- خوشه ۱ واقعاً پایدار و بالا است.
- خوشه ۰ واقعاً آرام و کم نوسان است.

این نمودار اعتبار خوشه بندی را ثابت می کند.

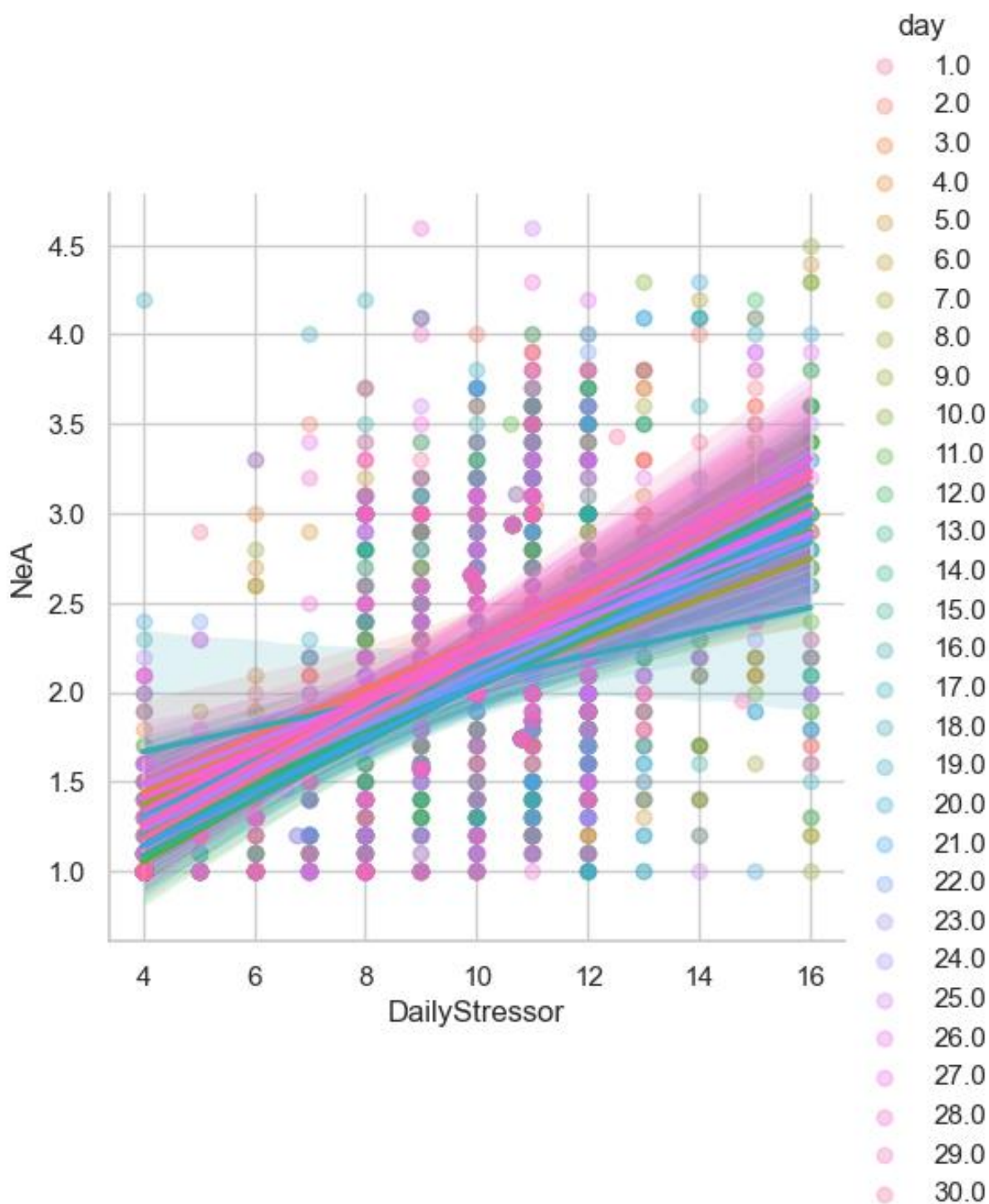
۴.۷.۳. نمایش کلی رفتار افراد



این هیت‌مپ الگوی خام NeA را برای هر فرد در طول روزها نشان می‌دهد و برخلاف نمودارهای میانگین، رفتار واقعی و فرد به فرد را نمایش می‌دهد. رنگ‌های روشن‌تر بیانگر هیجان منفی بالاتر و رنگ‌های تیره‌تر نشان‌دهنده مقادیر پایین‌تر هستند. تفاوت رنگ‌ها در هر ردیف نشان می‌دهد هر فرد چقدر نوسان یا ثبات هیجانی دارد. الگوهای عمودی مشخص می‌کنند در چه روزهایی هیجان منفی در کل نمونه بیشتر یا کمتر بوده است. این نمودار به خوبی ناهمگونی بین افراد و تفاوت‌های زمانی را آشکار می‌کند و نشان می‌دهد چرا خوشه‌بندی بر اساس الگوهای زمانی معنا داشته است.

۵.۷.۳. بررسی رابطه متغیرها

Lmplot:



این نمودار یک Scatter Plot با رگرسیون روزانه است یعنی رابطه‌ی بین دو متغیر را نشان می‌دهد، اما برای هر روز یک خط رگرسیون جداگانه رسم شده است .
یک نمودار پراکندگی است که روی آن ۳۰ خط رگرسیون (برای ۳۰ روز) قرار گرفته . سپس رابطه‌ی کلی بین DailyStressor و NeA را می‌بینی، هم اینکه این رابطه در هر روز چطور تغییر می‌کند.

چطور ساخته شده است؟

- محور افقی DailyStressor :
- محور عمودی NeA :
- هر نقطه: یک مشاهده‌ی روزانه از یک فرد
- برای هر روز یک مدل رگرسیون جداگانه برازش شده
- خطوط رنگی = خط رگرسیون مخصوص همان روز
- نوارهای سایه‌دار = عدم قطعیت یا پراکندگی داده‌ها در آن روز

رنگ‌ها چه معنایی دارند؟

از روز ۱ تا روز ۳۰، هر روز یک رنگ متفاوت دارد . این باعث می‌شود بفهمیم آیا رابطه‌ی استرس و هیجان منفی در روزهای مختلف قوی‌تر، ضعیف‌تر یا متفاوت است.

پیام اصلی نمودار چیست؟

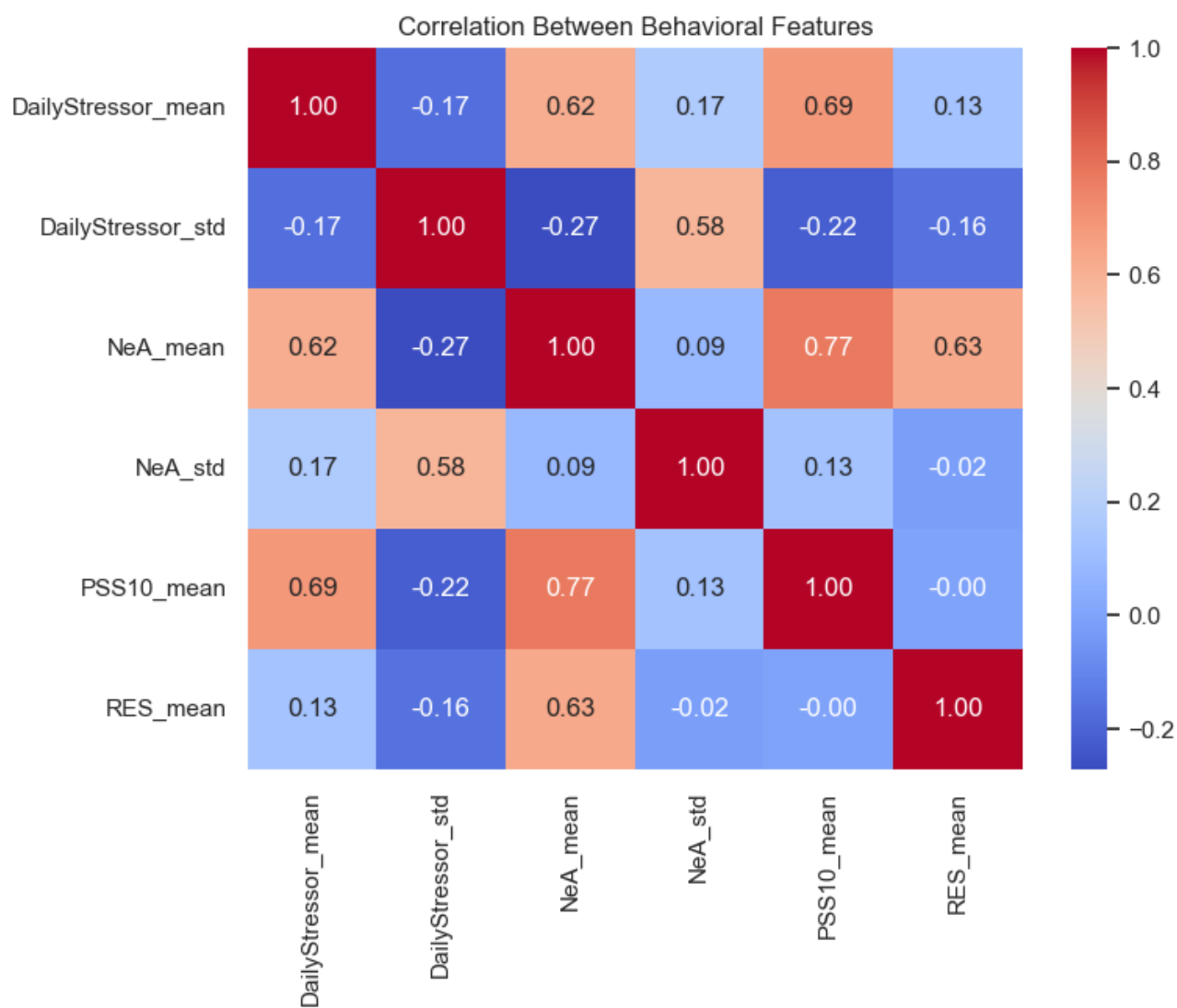
- رابطه‌ی کلی مثبت است: هرچه استرس روزانه بیشتر باشد، NeA هم بالاتر می‌رود.
- اما شدت این رابطه ثابت نیست؛ بعضی روزها شیب تندتر است، یعنی استرس اثر بیشتری روی هیجان منفی دارد.
- پراکندگی زیاد نشان می‌دهد افراد واکنش‌های متفاوتی به استرس دارند.
- باندهای پهن‌تر یعنی رابطه در آن روز کم‌ثبات‌تر بوده.

نتیجه‌ی رفتاری مهم

این نمودار نشان می‌دهد که:

- DailyStressor یک پیش‌بینی‌کننده‌ی مهم NeA است.
- اما اثرش روزبه‌روز تغییر می‌کند.
- و افراد واکنش‌های یکسانی به استرس ندارند.

Correlation heatmap:



هیت‌مپ همبستگی بین ویژگی‌های رفتاری

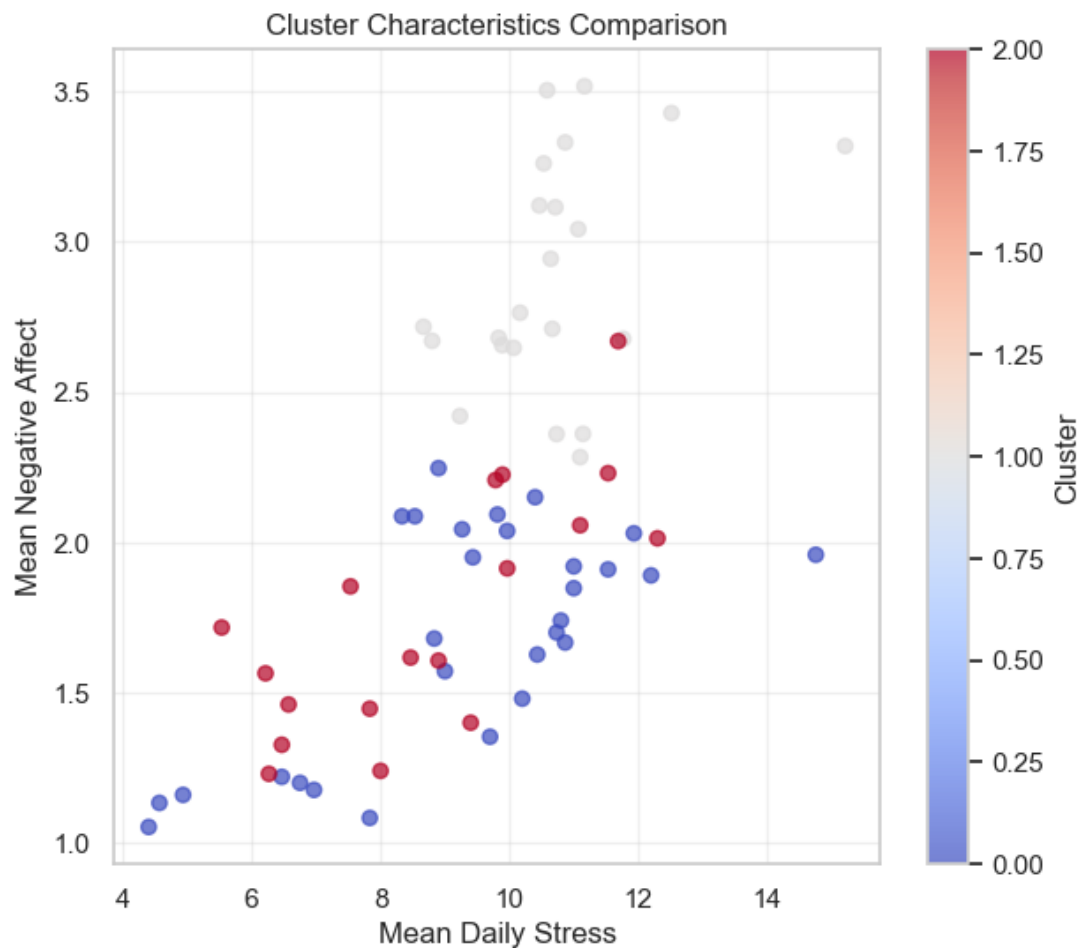
این هیت‌مپ میزان همبستگی بین میانگین و نوسان متغیرهای رفتاری را نشان می‌دهد و مشخص می‌کند کدام ویژگی‌ها با یکدیگر هم‌جهت یا ناهم‌جهت تغییر می‌کنند .

DailyStressor_mean با NeA_mean و PSS10_mean همبستگی مثبت و نسبتاً قوی دارد، به این معنا که افزایش استرس روزانه با افزایش هیجان منفی و استرس ادراک‌شده همراه است. در مقابل، DailyStressor_std با NeA_std رابطه مثبت دارد اما با میانگین‌ها رابطه منفی نشان می‌دهد، که بیانگر این است که نوسان استرس بیشتر با بی‌ثباتی هیجانی مرتبط است .

NeA_mean نیز با PSS10_mean و RES_mean همبستگی بالا دارد و نشان می‌دهد افراد با هیجان منفی بیشتر، استرس ادراک‌شده و تاب‌آوری بالاتری گزارش کرده‌اند. در مجموع، این الگوها ساختار روان‌شناختی داده‌ها را روشن می‌کنند و نشان می‌دهند که میانگین‌ها و نوسانات هر متغیر روابط متفاوت اما معناداری با یکدیگر دارند .

پیامش روابط روان‌شناختی متغیرها است.

Cluster Characteristics Comparison:



این نمودار میانگین استرس روزانه را در محور افقی و میانگین هیجان منفی را در محور عمودی نشان می‌دهد و هر نقطه یک فرد است. رنگ نقاط مشخص می‌کند فرد در کدام خوشه قرار گرفته و به‌خوبی دیده می‌شود که خوشه‌ها در این فضای دوبعدی از هم جدا هستند. خوشه ۱ در ناحیه‌ی استرس بالا و هیجان منفی بالا قرار دارد، خوشه ۰ در ناحیه‌ی پایین‌تر و خوشه ۲ بین این دو دیده می‌شود. این نمودار نشان می‌دهد خوشه‌بندی واقعاً الگوهای رفتاری متفاوتی را گرفته و سه تیپ روان‌شناختی متمایز شکل داده است.

جمع بندی نهایی نمودار های مورد استفاده در هر بخش:

نمودار ها	بخش
Missing heatmap	آماده سازی داده ها
Histogram	توزیع متغیرها
Spaghetti + Trajectory	مسیر زمانی افراد
Mean plots	روند جمعیت
Box + Violin	توزیع در زمان
Autocorrelation + Lag	وابستگی زمانی
Mean distribution + Slope	تفاوت افراد
Elbow	تعیین خوشه
Box + Radar	مقایسه خوشه ها
Cluster trajectory	رفتار زمانی خوشه ها
FacetGrid	تنوع داخل خوشه
Heatmap	نمایش کلی
Implot + Corr heatmap	روابط متغیرها

فصل ۴

نتیجه گیری

هدف اصلی این پژوهش، به کارگیری تحلیل اکتشافی داده‌ها برای بررسی الگوهای رفتاری در داده‌های طولی و خوشه‌ای و شناسایی تیپ‌های رفتاری متمایز در سطح فردی بود. داده‌های مورد استفاده شامل اندازه‌گیری‌های روزانه چندین شاخص روان‌شناختی برای ۷۲ فرد طی یک بازه ۳۰ روزه بود که ماهیتی ذاتاً طولی و وابسته درون‌فردی داشت.

در گام نخست، ساختار و کیفیت داده‌ها مورد بررسی قرار گرفت. نتایج تحلیل‌های اولیه نشان داد که داده‌ها از نظر ساختاری سازگار بوده و مقادیر گمشده به صورت پراکنده و بدون الگوی سیستماتیک در میان افراد و زمان توزیع شده‌اند. این امر امکان انجام تحلیل‌های بعدی و استفاده از روش‌های جایگزینی مناسب را بدون ایجاد سوگیری قابل توجه فراهم ساخت.

در ادامه، تحلیل توزیعی متغیرها نشان داد که شاخص‌های مورد مطالعه از نظر شکل توزیع و میزان پراکندگی تفاوت‌های معناداری دارند. متغیرهایی نظیر استرس روزانه و عاطفه منفی دارای نوسان بالا و رفتار غیرنرمال بودند، در حالی که متغیرهایی مانند تاب‌آوری و اضطراب صفتی الگوهای پایدارتر و متمرکزتری از خود نشان دادند. این یافته‌ها بیانگر آن بود که تحلیل صرفاً مبتنی بر میانگین‌های جمعیتی قادر به بازنمایی پیچیدگی رفتارهای فردی نیست.

تحلیل اکتشافی طولی در سطح فردی از طریق نمودارهای مسیر زمانی نشان داد که افراد الگوهای رفتاری بسیار متفاوتی را در طول زمان تجربه می‌کنند. برخی افراد دارای سطوح بالای پایدار از استرس و عاطفه منفی بودند، در حالی که برخی دیگر الگوهای نوسانی یا سطوح پایین و پایدار را

نشان دادند. در مقابل، روندهای میانگین جمعیتی تصویری نسبتاً هموار ارائه می‌دادند که این موضوع بر محدودیت تحلیل‌های جمعیتی در داده‌های ناهمگن تأکید می‌کند.

بررسی تغییرات توزیع متغیرها در طول زمان نشان داد که اگرچه تغییرات سیستماتیک شدیدی در سطح جمعیت مشاهده نمی‌شود، اما پراکندگی و شکل توزیع متغیرها در روزهای مختلف متفاوت است. این یافته حاکی از آن است که نوسانات مشاهده‌شده بیشتر ناشی از تفاوت‌های بین‌فردی است تا تغییرات زمانی مشترک برای همه افراد.

تحلیل وابستگی زمانی نیز نشان داد که مقادیر متغیرها در روزهای متوالی به یکدیگر وابسته نیستند. با توجه به وابستگی درون‌فردی داده‌ها، در این پژوهش هر فرد به‌عنوان یک واحد رفتاری مستقل در نظر گرفته شد. به‌منظور مقایسه افراد، داده‌های روزانه هر فرد با استفاده از شاخص‌هایی نظیر میانگین، انحراف معیار و روند تغییرات خلاصه‌سازی شد و یک پروفایل رفتاری برای هر فرد تشکیل گردید. این رویکرد امکان تبدیل داده‌های طولی به فضای ویژگی‌های فردمحور را فراهم ساخت.

در ادامه، تحلیل خوشه‌ای بر روی پروفایل‌های رفتاری افراد انجام شد. نتایج روش آرنجی نشان داد که تقسیم افراد به سه خوشه تعادل مناسبی میان سادگی مدل و توان تفکیک داده‌ها ایجاد می‌کند. تحلیل ویژگی‌های خوشه‌ها نشان داد که این سه گروه از نظر سطح استرس، عاطفه منفی، میزان نوسان و تاب‌آوری تفاوت‌های معناداری دارند.

بر این اساس، **خوشه نخست** شامل افرادی بود که سطوح بالای استرس و عاطفه منفی را به‌صورت پایدار تجربه می‌کردند و می‌توان آن‌ها را به‌عنوان گروه «**پرریسک و هیجانی**» توصیف کرد. **خوشه دوم** شامل افرادی با الگوهای نوسانی و تغییرپذیر بود که رفتار هیجانی آن‌ها در طول زمان ثبات کمتری داشت و به‌عنوان گروه «**نوسانی**» قابل تفسیر است. **خوشه سوم** شامل افرادی با سطوح پایین‌تر استرس و عاطفه منفی و میزان بالاتری از تاب‌آوری بود که رفتار نسبتاً پایدار و سازگارانه‌ای از خود نشان دادند و می‌توان آن‌ها را گروه «**تاب‌آور و پایدار**» نامید.

به طور کلی، نتایج این پژوهش نشان داد که ترکیب تحلیل اکتشافی طولی و تحلیل خوشه‌ای می‌تواند بدون اتکا به مدل‌های پیچیده آماری، الگوهای رفتاری معنادار و قابل تفسیر را در داده‌های روان‌شناختی آشکار سازد. این رویکرد نه تنها تفاوت‌های بین فردی را برجسته می‌کند، بلکه پویایی زمانی رفتار افراد را نیز به صورت شهودی و قابل فهم نمایش می‌دهد.

در نهایت، می‌توان نتیجه گرفت که تحلیل اکتشافی داده‌ها نقش مهمی در درک ساختار داده‌های طولی و خوشه‌ای ایفا می‌کند و می‌تواند به عنوان گامی مقدماتی و ضروری پیش از مدل‌سازی‌های پیشرفته‌تر مورد استفاده قرار گیرد. یافته‌های این پژوهش می‌توانند مبنایی برای مطالعات آینده در زمینه مدل‌سازی پیش‌بینانه، تحلیل‌های علی و طراحی مداخلات مبتنی بر تیپ‌های رفتاری شناسایی شده فراهم سازند.

بهرامی سامانی، احسان. (۱۴۰۲). روش‌های نموداری در تحلیل اکتشافی داده‌ها. روش‌های نموداری در تحلیل اکتشافی داده‌ها.

تهران: انتشارات [دانشگاه شهید بهشتی]

Lu, X., et al. (2024).

Clustering longitudinal data: A review of methods and software packages. *International Statistical Review*, 92(1), 1–25.

Genolini, C., & Falissard, B. (2011).

Clustering of longitudinal data: A tutorial on a variety of approaches. *Advances in Data Analysis and Classification*, 5(4), 249–272.

Bouveyron, C., et al. (2019).

Longitudinal data clustering methods: A systematic review. *Pattern Recognition*, 91, 61–79.

Tukey, J. W. (1977).

Exploratory Data Analysis.
Reading, MA: Addison-Wesley.

Daily stress and affect longitudinal dataset.(2023).

Unpublished dataset used for educational and research purposes.

Pedregosa, F., et al. (2011).

Scikit-learn: Machine learning in Python.

Journal of Machine Learning Research, 12, 2825–2830.

Waskom, M. L. (2021).

Seaborn: Statistical data visualization.

Journal of Open Source Software, 6(60), 3021.