

دانشگاه علوم پزشکی اصفهان

دانشکده بهداشت – گروه آمار زیستی و اپیدمیولوژی

نام درس: تحلیل داده‌های طولی **Longitudinal Data Analysis**

مدرس: دکتر سید محسن حسینی

مقطع: تحصیلات تکمیلی دکتری/ارشد آمار زیستی

تعداد واحد ۳: واحد نظری ارائه شده در ۱۷ جلسه ۳ ساعته

اهداف کلی دوره

- آشنایی عمیق با مبانی نظری و ریاضیاتی مدل‌سازی داده‌های طولی و مکرر.
- تسلط بر کاربرد نرم‌افزارهای آماری پیشرفته از جمله R، R در تحلیل داده‌های واقعی پزشکی و بهداشتی.
- توانایی طراحی، پیاده‌سازی، تحلیل و گزارش‌دهی نتایج حاصل از مطالعات طولی در قالب مقالات بین‌المللی.

منابع درسی اصلی

۱. Fitzmaurice GM, Laird NM, Ware JH. *Applied longitudinal analysis*, Second edition, 2012.

۲. Diggle PJ, Liang KY, Zeger SL. *Analysis of longitudinal data*, Second edition, 2002.

۳. Wang J, Wang X. *Structural equation modeling: Applications using Mplus*, First edition, 2012.

جدول برنامه‌ریزی جلسات ۱۷ گانه همراه با کار عملی

جلسه اول: تعاریف اولیه و مثال‌هایی از مطالعات طولی

- مفاهیم: تفاوت مطالعات مقطعی و طولی، مزایای مطالعات طولی، نمادگذاری داده‌ها و معرفی انواع مدل‌ها.
 - کار عملی: بررسی ساختار داده‌های طولی در نرم‌افزار و آماده‌سازی فایل داده. Wide vs. Long format
- جلسه دوم: ملاحظات طراحی و پروپوزال‌نویسی در مطالعات طولی

- مفاهیم: اربابی، کارایی، تعیین حجم نمونه در داده‌های طولی و متغیرهای کلیدی.
 - کار عملی: محاسبه حجم نمونه مقدماتی با استفاده از نرم‌افزارهای توان‌سنجی برای طرح‌های طولی.
- جلسه سوم: تحلیل اکتشافی **Exploratory Analysis** در داده‌های طولی

- مفاهیم: ترسیم پروفایل‌های فردی، هموارسازی ناپارامتری هسته Lowess، واسپلین، بررسی ساختار همبستگی و واریوگرام.
 - کار عملی: رسم نمودارهای اسپاگهتی و منحنی‌های هموارسازی برای داده‌های واقعی پیامد سلامت با استفاده از نرم‌افزار.
- جلسه چهارم: آشنایی پیشرفته با نرم‌افزار R در مدیریت داده‌های طولی

- مفاهیم: محیط R توابع ماتریسی، دستورات پردازش داده‌های ساختاریافته.
 - کار عملی: نوشتن برنامه‌های اولیه جبر ماتریسی و محاسبه شاخص‌های توصیفی چندمتغیره در R.
- جلسه پنجم: مرور مدل‌های خطی برای داده‌های طولی

- مفاهیم: مدل‌های خطی عمومی، همبستگی یکنواخت و نمایی، برآورد کمینه مربعات موزون و بیشینه درست‌نمایی ML و REML.
 - کار عملی: برازش مدل‌های خطی با ساختارهای کوواریانس مختلف با استفاده از پروسچرهای R.
- جلسه ششم: مدل‌های خطی تعمیم یافته **GLM** برای داده‌های غیرنرمال

- مفاهیم: مدل‌های رگرسیونی برای داده‌های دوحالتی، ترتیبی و شمارشی، مدیریت بیش‌پراکنش و صفرانباشستگی.
 - کار عملی: تحلیل داده‌های طولی شمارشی مانند تعداد دفعات بستری در محیط R.
- جلسه هفتم: مدل‌سازی پارامتری ساختار کوواریانس

- مفاهیم: بررسی مدل‌های همبستگی پیاپی، ترکیب خطاهای اندازه‌گیری و عرض از مبدأ تصادفی.
- کار عملی: مقایسه شاخص‌های برازش AIC, BIC در انتخاب بهترین ساختار کوواریانس با کمک منابع مرجع.

جلسه هشتم: مدل‌های حاشیه‌ای Marginal Models و معادلات برآوردی تعمیم‌یافته GEE

- مفاهیم: فرمول‌بندی مدل‌های حاشیه‌ای، استنباط مبتنی بر GEE و ساختارهای کاری کوواریانس.
- کار عملی: تحلیل داده‌های طولی دودویی با استفاده از دستور Proc Genmod در نرم‌افزار R.

جلسه نهم: مدل‌های اثرات مختلط خطی - LMM

- مفاهیم: رویکرد دورمرحله‌ای، اثرات تصادفی و شیب‌های تصادفی، پیش‌بینی اثرات تصادفی. BLUPs.
- کار عملی: برازش مدل‌های LMM با استفاده از دستور Proc Mixed در R و تفسیر ضریب تغییرات.

جلسه دهم: مدل‌های اثرات مختلط خطی تعمیم‌یافته GLLMM

- مفاهیم: تلفیق مدل‌های مختلط با GLM، روش بیشینه درست‌نمایی شرطی و مقایسه رویکرد حاشیه‌ای و شرطی.
- کار عملی: تحلیل داده‌های طولی پیامدهای رتبه‌ای.

جلسه یازدهم: روش‌های تقریبی برآورد پارامترها

- مفاهیم: روش‌های شبه درست‌نمایی تاوانیده، PQL حاشیه‌ای MQL و روش مربع‌سازی. Quadrature.
- کار عملی: بررسی مقایسه‌ای نتایج حاصل از روش‌های مختلف برآورد و تحلیل میزان اریبی.

جلسه دوازدهم: مدل‌های چندسطحی Multilevel Models در داده‌های طولی

- مفاهیم: ساختار داده‌های چندسطحی اندازه‌گیری‌ها در درون افراد، افراد در درون مراکز، مدل‌های دوسطحی و سه‌سطحی.
- کار عملی: پیاده‌سازی و تحلیل داده‌های ساختاریافته طولی-خوشه‌ای در نرم‌افزار.

جلسه سیزدهم: داده‌های گمشده و ریزش Missing Data & Dropout

- مفاهیم: مکانیسم‌های گمشدگی، MCAR, MAR, MNAR، روش‌های جانشین چندگانه Multiple Imputation و احتمال موزون وارون. IPW.
- کار عملی: مدیریت داده‌های گمشده و اجرای الگوریتم‌های جانشین در نرم‌افزار آماری.

جلسه چهاردهم: مدل‌های انتقال و زنجیره‌های مارکوف Transition Models

- مفاهیم: وابستگی مرتبه اول و بالاتر، مدل‌های مارکوف برای پیامدهای پیوسته، دودویی و ترتیبی.
- کار عملی: برازش مدل‌های انتقال مارکوف روی داده‌های طولی پزشکی با استفاده از R.

جلسه پانزدهم: مدل‌های رشد پنهان LGCM - Latent Growth Curve Models

- مفاهیم: رویکرد معادلات ساختاری برای داده‌های طولی، مدل‌های رشد خطی و غیرخطی، استفاده از متغیرهای پنهان عرض از مبدأ و شیب.
- کار عملی: برازش مدل‌های رشد پنهان خطی و محاسبه مسیرهای رشد با استفاده از نرم‌افزار R.

جلسه شانزدهم: مدل‌های رشد آمیخته و تحلیل خوشه‌ای طولی Mixture Models

- مفاهیم: مدل‌های رشد پنهان آمیخته، Latent Class Growth Models / Growth Mixture Modeling تعیین تعداد رده‌های پنهان و شاخص‌های نیکویی برازش.
- کار عملی: شناسایی زیرگروه‌های مختلف trajectory مسیرهای رشد ناهمگن در یک جمعیت بالینی با استفاده از نرم‌افزار R.

جلسه هفدهم: حل مثال‌های کاربردی پیشرفته و رفع اشکال جامع

- مفاهیم: مرور جامع مباحث، سناریوهای تحلیل داده‌های واقعی پژوهشی و نکات نگارش مقاله علمی.
- کار عملی: تحلیل پایانی یک مجموعه داده جامع طولی به عنوان پروژه عملیاتی دانشجویان و ارائه گزارش.

شیوه ارزشیابی و نمره دهی

۱. فعالیت کلاسی و ارائه مباحث تخصصی ۲:نمره
۲. ارزیابی میان‌ترم کوپیز و کوئیز نرم‌افزاری ۳:نمره
۳. امتحان پایان‌ترم کتبی و عملی ۱۵:نمره