



دانشکده بهداشت

دفتر توسعه آموزش

طرح دوره (Course Plan)

نام درس: تحلیل داده های طبقه بندی شده	کد درس:
مقطع تحصیلی: دکترای	رشته تحصیلی: آمارزیستی
گروه آموزشی: آمارزیستی	پیش نیاز:
تعداد واحد: ۳ واحد	محل برگزاری: دانشکده بهداشت
روز و ساعت برگزاری: یکشنبه ۱۰-۸ و سه شنبه ۱۲-۱۰	مسئول درس: دکتر آزاده ساکی
طراحی اولیه ☐ بازنگری ☒	
هدف کلی دوره: تحلیل داده های رسته ای در علوم پزشکی مبتنی بر روش و هدف مطالعه	
اهداف اختصاصی دوره: 1- مفاهیم و تعاریف داده های گسسته و انواع آن را تعریف کنند. 2- توزیع احتمال مربوط به انواع داده های گسسته شامل دوجمله ای، چندجمله ای و پواسن و دوجمله ای منفی در انواع مطالعات را تشخیص دهند. 3- ارتباط سنجی و آزمون های مرتبط با پاسخ های رسته ای را بدون استفاده از مدل سازی مطابق با هدف مطالعه انجام دهند. 4- مدل های خطی تعمیم یافته GLM را تعریف کنند و توابع پیوند مناسب با هر نوع داده را استخراج و مدلسازی با کد نویسی در نرم افزار انجام دهند. 5- مدل های متداول برای پاسخ های دودویی شامل لجستیک، پروبیت، خطی را تعریف و بتوانند با استفاده از نرم افزار مدلسازی را انجام دهند. 6- مدل های جمعی تعمیم یافته GAM را برازش و از آنها جهت بهبود مدل های GLM استفاده کنند. 7- روش های جایگزین برای مدل های GLM همچون درخت تصمیم، یادگیری ماشین و مدل های فازی را بدانند و با نرم افزار اجرا کنند. 8- مدل رگرسیون لجستیک را بر روی داده های مثال برازش داده و ضرائب آن را تفسیر و توضیح دهند. 9- مدل لوگ خطی را بر روی داده های شمارشی بر ارزش داده و تفسیر کنند و مفاهیم آن را توضیح دهند. 10- بیش پراکنی در مدل های گسسته را شناسایی و علل آن را توضیح دهند. 11- بدون استفاده از مدل مطالعات همسان شده را تحلیل و نسبت شانس را محاسبه کنند. 12- مدل های حاشیه ای، GEE برای پاسخ های دودویی همبسته را تعریف و مدلسازی را با استفاده از نرم افزار انجام دهند. 13- مدل های آمیخته و اثرات تصادفی را برای پاسخ های دودویی همبسته تعریف و مدلسازی را با استفاده از نرم افزار انجام دهند. ۱۴- ارزیابی روایی و پایایی ابزارهای سنجش (اندازه گیری و پرسشنامه ها)، ارزش تشخیصی، روایی ملاکی، پایایی درونی و آزمون-بازآزمون را با استفاده از مدل های آمیخته انجام دهند. ۱۵- مدل های چند سطحی، تحلیل کلاس پنهان، تحلیل کلاس پنهان چند سطحی و رگرسیون کلاس پنهان را با استفاده از نرم افزار اجرا کنند. 16- تحلیل داده های با بعد بالا بدون استفاده از مدل و با کنترل نرخ کشف کاذب. 17- روش های تنظیم کنندگی در مدلسازی داده های با بعد بالا و معرفی مدل های مناسب جهت این نوع داده ها.	
ط	



دانشکده بهداشت

دفتر توسعه آموزش

طرح دوره (Course Plan)

جدول ترتیب ارائه محتوای آموزشی در طول دوره	
جلسه	رئوس مطالب
اول	تعریف داده های گسسته و توزیع های احتمال آن ها - استنباط آماری برای سهم ها و توابع درستنمایی و آزمون های مرتبط با آن - معرفی آزمون های دقیق ، آزمون جایگشت، آزمون فیشر، و تکنیک Bootstrap
دوم	جداول توافقی و ساختار احتمال آن ها - توزیعهای احتمال متناسب با جداول توافقی در انواع مطالعات مشاهده ای- کاربرد جداول توافقی در بررسی ارتباط دو متغیر کیفی-محاسبه نسبت شانس و محاسبه حساسیت و اختصاصیت- آزمون مقایسه نسبت در دو جامعه مستقل-
سوم	بررسی مخدوشگرها در جداول توافقی ، نسبت شانس شرطی و حاشیه ای، همگنی نسبت شانس در جداول توافقی- آزمون منل هنزل و پارادوکس سیمپسون
چهارم	انجام تحلیل روی داده های واقعی به صورت مطالعه موردی case study
پنجم	معرفی مدل های خطی تعمیم یافته مفاهیم ومولفه های مدل های خطی تعمیم یافته - تعریف تابع پیوند و انواع آن-معرفی مدل های مناسب برای داده های دو حالتی ، چند حالتی، رتبه ای ، و شمارشی
ششم	معرفی مدل رگرسیون لجستیک و رگرسیون لجستیک چندگانه و برازش آن روی داده های حقیقی و تفسیر پارامترهای مدل
هفتم	استنباط آماری برای مدل رگرسیون لجستیک و بررسی مناسب مدل و انتخاب بهترین مدل بر اساس معیارهای AIC و AICC و بررسی بیش پراکنش روش های ساخت و به کارگیری یک مدل رگرسیون لجستیک مناسب روش پیشرو، پس رو و گام به گام -آزمون های مناسب بر اساس شاخصهای LR , deviance
هشتم	معرفی مدل های GAM، و چک کردن مناسبیت مدل رگرسیون لجستیک، بررسی قدرت پیشگویی مدل بر اساس جداول طبقه بندی، منحنی راک و همبستگی روش های تعیین حجم نمونه مناسب برای برازش مدل لجستیک
نهم	روشهای جایگزین برای مدل لجستیک همچون درخت تصمیم، یادگیری ماشین و مدل های فازی را بدانند و با نرم افزار اجرا کنند.
دهم	برازش مدل رگرسیون لجستیک و یا سایر مدل های مناسب روی داده های واقعی به صورت مطالعه موردی case study
یازدهم	تحلیل مدل رگرسیون لجستیک چندگانه و بررسی مناسبیت مدل با R و انجام تمرین های منتخب
دوازدهم	مدل های لجستیک برای داده های کیفی چند جمله ای و معرفی مدل های لجستیک تجمعی و لجستیک رتبه ای برای داده های کیفی رتبه ای و برازش و تفسیر نتایج مدل
سیزدهم	برازش مدل رگرسیون لجستیک چندگانه و رتبه ای و یا سایر مدل های مناسب روی داده های واقعی به صورت مطالعه موردی case study
چهاردهم	معرفی مدل لوگ خطی برای داده های شمارشی و جداول توافقی دو راهه و سه راهه
پانزدهم	استنباط در مورد پارامترهای مدل و بررسی ارتباطات شرطی و غیر شرطی. ادغام جداول و استفاده از رگرسیون لجستیک. بررسی مناسبیت مدل و بیش پراکنشی در این مدل ها
شانزدهم	برازش مدل های لوگ خطی یا سایر مدل های مناسب روی داده های واقعی به صورت مطالعه موردی case study



دانشکده بهداشت

دفتر توسعه آموزش

طرح دوره (Course Plan)

هفدهم	تحلیل مطالعات همسان سازی شده، و آزمون مک نمار برای بررسی استقلال در مطالعات همسان سازی شده محاسبه odds ratio در مطالعات همسان سازی شده
هجدهم	مدل های حاشیه ای ، GEE برای پاسخ های دودویی
نوزدهم	مدل های آمیخته و اثرات تصادفی را برای پاسخ های دودویی همبسته-
بیستم	ارزیابی روایی و پایایی ابزارهای سنجش (اندازه گیری و پرسشنامه ها)
بیستم و یکم	مدلهای چند سطحی، تحلیل کلاس پنهان، تحلیل کلاس پنهان چند سطحی و رگرسیون کلاس پنهان
بیست و دوم	تحلیل یک مطالعه همسان شده بدون مدل و با مدل های مناسب روی داده های واقعی به صورت مطالعه موردی case study
بیست و سوم	تحلیل داده های با بعد بالا بدون استفاده از مدل و با کنترل نرخ کشف کاذب.
بیست و چهارم	روشهای تنظیم کنندگی در مدلسازی داده های با بعد بالا و معرفی مدل های مناسب جهت این نوع داده ها
بیست و پنجم	تحلیل مجموعه داده های با بعد بالای ژنتیکی بدون مدل و با مدل های مناسب روی داده های واقعی به صورت مطالعه موردی case study
روش های آموزشی (نحوه ارائه درس): PBL، بحث گروهی، سخنرانی و اسلاید، ارایه سناریو - تفکر انتقادی در حل تمرینها	
وسایل آموزشی: کامپیوتر- دیتا- کتاب	
وظایف و تکالیف دانشجوی: مطالعه مباحث و جستجو در پایگاههای معتبر برای هر جلسه. حل تمرینهای منتخب و ارایه در کلاس	
روش های ارزیابی دانشجوی (لطفا سهم هر روش بر حسب درصد نوشته شود):	
روش های ارزیابی در طول دوره: پاسخ به پرسشهای مطرح شده در کلاس ۲۰٪- حل تمرینهای منتخب ۲۰٪- ارایه مقاله و بحث پیرامون آن ۱۰٪-	
روش های ارزیابی در انتهای دوره: آزمون پایان دوره ۵۰٪	
منابع اصلی درس:	
1- Agresti, A. An introduction to categorical data analysis. Third Ed.2019	
2- Collett,D Modeling Binary Data. Second Ed.2002	
3- Hosmer,D. W. and Lemeshow, S. Applied Logistic Regression. Second ED.2000	
4- Fleiss,j.l. statistical methods for rates and proportions, second Ed.1981	