

کارگاه های دوم و سوم مرکز تحقیقات مدل سازی در سلامت، تحت

عناوین **Decision Tree Modeling** و **Imputation** of

**Missing Data** از سری ماموریت های توسعه دانش مدل سازی در

علوم سلامت، واگذار شده از طرح وزارت بهداشت درمان و آموزش

پزشکی به منطقه هشت آمایشی، در تاریخ های ۱۴ و ۱۵ مهرماه برگزار

شد.

هر دو کارگاه ساعت ۸ صبح در سالن کنفرانس پژوهشکده آینده پژوهی

در سلامت، با ارائه جناب آقای دکتر محمد رضا بانوشی برگزار شد.

حدود پنجاه نفر به شکل حضوری در کارگاه حضور داشتند، حدود ده

نفر آنها عضو هیات علمی بودند. همزمان شرکت کنندگان بصورت

وبینار و ویدئو کنفرانس بصورت آنلاین از ۱۰ دانشگاه کشور مانند

شیراز، همدان، کاشان، سبزوار، شهید بهشتی، کردستان، تربت حیدریه،

بیرجند، ایران و دانشگاه تهران حضور داشته و بصورت مجازی از تمام

مطالب کارگاه ها بصورت فیلم و صوت بهره می بردند. شایان ذکر است، ۵ نفر از دیگر دانشگاه های کشور از جمله دانشگاه علوم پزشکی ایران، جیرفت، اصفهان، مشهد، بصورت حضوری در این دوره از کارگاه ها حضور داشتند.

برای پیش بینی یک متغیر وابسته بر اساس چندین متغیر مستقل معمولاً از رگرسیون استفاده میشود. این مدلها منجر به یک فرمول آماری میشوند که میتوان از آنها برای پیش بینی مقدار متوسط متغیر مستقل برای افراد جدید استفاده کرد. ولی در شرایطی که پیش نیازهای مدل های رگرسیون کلاسیک برقرار نبوده، تعداد متغیرهای مستقل زیاد باشد، همبستگی بین متغیرهای مستقل وجود داشته باشد و یا ارتباط متغیرهای مستقل با پیامد، غیر خطی باشد و یا اثرات متقابل پیچیده بین متغیرهای مستقل وجود داشته باشد، میتوان از مدلهای درختی به عنوان مدلی مکمل و قوی استفاده کرد.

مدل درختی با قرار دادن همه افراد در یک گروه شروع میشود. سپس بر اساس یکی از متغیرهای مستقل و یک نقطه برش نمونه ها به دو دسته تقسیم میشوند. تقسیم به شکلی صورت میگیرد که در هر کدام از دو دسته حداکثر همگنی و بین آنها حداکثر ناهمگنی وجود داشته باشد. سپس هر کدام از این دسته ها مجدداً به دو دسته تقسیم میشوند و کار به همین شکل پیش میرود. در این کارگاه در مورد معیارهای تقسیم و همچنین زمان مناسب برای توقف درخت صحبت شد. بعلاوه در مورد هرس کردن درخت به منظور جلوگیری از بیش برآزشی صحبت شد. بر اساس درخت نهایی خیلی ساده میتوان برای افراد جدید کار پیش بینی را انجام داد.

موضوع کارگاه دوم داده های گمشده بود. به ندرت پیش می آید که با مجموعه داده ای مواجه شویم که اطلاعات همه نمونه ها، به طور کامل وجود داشته باشد. در اکثر موارد اطلاعات افرادی که داده گمشده دارند

به طور کامل حذف میشود. مثلاً فرض کنید، از پرونده بیماران اطلاعات متغیرهای دموگرافیک و پاتولوژیک را استخراج کرده باشیم. بطور معمول اگر حتی یکی از این متغیرها برای یک بیمار اندازه گیری نشده باشد، آن بیمار از کلیه آنالیزها کنار میرود. این کار باعث از دست دادن داده ها و کاهش توان آزمون میگردد. در مورد مکانیزم و الگوی گمشدگیبه طور کامل بحث شد. سپس روشهای رایج برآورد داده های گمشده و نقاط قوت و ضعف هر کدام مرور گردید. در نهایت توصیه هایی در مورد انتخاب روش برآورد مناسب ارائه گردید.

در هر دو کارگاه علاوه بر بیان مطالب تئوری مدل ها به ارائه مثال عملی در نرم افزار spss پرداخته شد. همچنین از شرکت کنندگان خواسته شد، در کارگروه هایی به تمرین مباحث نرم افزاری پرداخته تا در صورت وجود هر گونه اشکال، توضیحات لازم ارائه و مشکلات آنها رفع گردد. لازم به ذکر است که عمق مباحث مطرح شده در این کارگاه

ها به طور معمول در هیچ یک از دوره های تخصصی اپیدمیولوژی و آمار زیستی تدریس نمی شود. در واقع هدف مرکز این است که در سطحی فراتر از کوریکولوم های آموزشی کارگاهها را برگزار کند.

مالتی مدیا آموزشی این کارگاه ها نیز در حال آماده سازی بوده، تا در اختیار همه علاقمندان قرار گیرد. عکس هایی از کارگاه های مربوطه به پیوست می باشد. مزید امتنان خواهد بود، در صورت صلاحدید، موارد به معاونتهای آموزشی و پژوهشی کل کشور ابلاغ گردد تا هم در جریان کار قرار گرفته و هم با انتشار در زیر مجموعه های خود امکان استفاده سایر محققین از این کارگاهها فراهم گردد.