موسسه آموزش عالی آزاد توسعه
https://tihe.ac.ir
۰۲۱-۸۶۷۴۱

به نام خدا

سرفصل های دوره آمار برای علم داده

آمار توصیفی (۶ ساعت)

- انواع داده
- تفاوت نمونه و جامعه
- شاخص های تمرکز و پراکندگی شامل :

میانگین، میانه، مد، دامنه، واریانس، انحراف معیار، چولگی، کشیدگی، دامنه میان چارکی

- ترسیم گرافیکی داده ها شامل نمودارهای:

هیستوگرام، نمودار جعبه ای یا Box Plot ، نمودار نقطه ای یا Dot Plot ، Bar Chart، Pie chart ، Heatmap ، نمودار پراکندگی یا Scatter Plot ، 3D scatterplot ، Contour Plot و Bubble plot

- نحوه بررسی توزیع انواع داده ها و ارتبط متغیرها به وسیله نمودار های مختلف

به همراه بررسی و تحلیل داده‌های مطالعه موردی در حوزه‌های مختلف با نرم افزار minitab و زبان برنامه نویسی R

احتمال (۸ ساعت)

- تعاریف احتمال، مقدمه ای بر جبر مجموعه ها، احتمال شرطی، قانون ضرب در احتمال قانون احتمال کل، پیشامدهای مستقل و وابسته، آزمایش های مستقل و وابسته
- کاربرد احتمالات در علم داده شامل:

قوانین انجمنی یا association rules ، روش دسته بندی naïve bayes و فرایندهای مارکوف

به همراه بررسی و تحلیل داده‌های مطالعه موردی در حوزه‌های مختلف با نرم افزار minitab و زبان برنامه نویسی R

موسسه آموزش عالی آزاد توسعه
https://tihe.ac.ir
۰۲۱-۸۶۷۴۱

متغیرهای تصادفی(۵ ساعت)

- تعریف متغیر تصادفی، متغیرهای پیوسته، گسسته و آمیخته، تابع جرم احتمال و تابع چگالی احتمال، بررسی رابطه هیستوگرام وتابع احتمال و تابع چگالی احتمال
- مفهوم واریانس و امید ریاضی جامعه و نمونه
- معرفی توزیع های برنولی، دوجمله‌ای، پواسون و نرمال
- توزیع تابعی از یک متغیر تصادفی
- بررسی نوع توزیع متغیر تصادفی به وسیله آزمون اندرسون دارلینگ Anderson Darling
- تبدیل های متغیر Box-Coc و Johnson برای نرمال سازی متغیر های تصادفی

به همراه بررسی با نرم افزارminitab و زبان برنامه نویسی R

توزیع های توأم (۵ ساعت)

- توزیع های توأم گسسته و پیوسته، تابع جرم احتمال و چگالی توأم، توزیع حاشیه‌ای
- توزیع و امیدریاضی تابعی از چند متغیر تصادفی، توزیع مجموع متغیرهای تصادفی، قضیه حد مرکزی
- کواریانس و همبستگی، ضریب همبستگی پیرسن و اسپیرمن، ماتریس کوواریانس و همبستگی
- برخی کاربرد های ماتریس کوواریانس و همبستگی مانند PCA یا تحلیل مولفه های اصلی

به همراه بررسی با نرم minitab

تعیین و تحلیل نقاط دورافتاده (۲ ساعت)

- روش های تحلیل نقاط دور افتاده در فضای تک متغیره
- روش های تحلیل نقاط دور افتاده در فضای چند متغیره

به همراه بررسی با نرم minitab

موسسه آموزش عالی آزاد توسعه
https://tihe.ac.ir
۰۲۱-۸۶۷۴۱

دسته بندی یا Classification (۸ ساعت)

- برخی روش های دسته بندی شامل:
  naïve bayes ، k نزدیکترین همسایه یا KNN، درخت تصمیم یا Decision Tree و جنگل تصادفی یا Random forests
- بررسی بیش برازش و کم برازش یا Over Fitting و Under Fitting
- بیش نمونه گیری و کم نمونه گیری Over Sampling و Under Sampling
- معیارهای های ارزیابی روش های دسته بندی شامل:
  Confusion Matrix، Accuracy، Sensitivity، Precision، Specificity، Lift، ROC، AUC
- اعتبار سنجی یا cross validation مدل های دسته بندی

به همراه بررسی و تحلیل داده‌های مطالعه موردی در حوزه‌های مختلف با نرم minitab و زبان برنامه نویسی R

توزیع های نمونه ای(۴ ساعت)

- توزیع های کای دو، Fisher ، T-Student ، توزیع آماره های X Bar ، S ، توزیع تفاضل میانگین نمونه ها و توزیع نسبت واریانس نمونه ها
- باز نمونه گیری و bootstrapping

به همراه بررسی با نرم minitab

نظریه برآورد نقطه ای و فاصله ای (۵ ساعت)

- روش MLE ، روش گشتاورها ، برآورد فاصله ای

به همراه بررسی با نرم minitab

آزمون فرض (۶ ساعت)

- مفاهیم ، خطای نوع ۱ و۲ و تابع توان و مقدار احتمال یا P-value
- آزمون برابری پارامترهای توزیع های مختلف، آزمون مقایسه پارامترهای چند جامعه

- آزمون نیکویی برازش
- آزمون کولموگروف Kolmogrov ، آزمون اندرسون دارلینگ Anderson Darling ، آزمون رایان جوینر Ryan Joiner برای تشخیص نوع توزیع
- آزمون نسبت درست نمایی

به همراه بررسی و تحلیل داده‌های مطالعه موردی در حوزه‌های مختلف با نرم minitab

آنالیز واریانس (۵ ساعت)

- آنالیز واریانس یک عاملی ، آزمون توکی Tukey ، آزمون دانت Dunnet ، آزمون فیشر Fisher ، تحلیل باقی مانده ها
- آنالیز واریانس دو عاملی بدون اثر متقابل ، آنالیز واریانس دوعاملی با اثر متقابل ،آنالیز واریانس چند عاملی

به همراه بررسی و تحلیل داده‌های مطالعه موردی در حوزه‌های مختلف با نرم افزار minitab

رگرسیون (۲۱ ساعت)

- رگرسیون خطی ساده
- رگرسیون چندگانه
- رگرسیون چند جمله ای
- رگرسیون قدم به قدم Stepwise، backward و forward
- بررسی کفایت مدل شامل

بررسی همگنی یا ناهمگنی واریانس خطاهای پیش بینی (Heteroscedasticity)

بررسی نرمال بودن خطاهای پیش بینی (Normality of residuals)

بررسی وجود خودهمبستگی خطاهای پیش بینی (Autocorrelation)

بررسی وجود هم‌خطی چندگانه(Multicolinearity)

بررسی نقاط پرت(Outlier Detection)

معیارهای ارزیابی مدل شامل:

R^2، Adjusted R^2 ، AIC ، BIC و Mallows's Cp

- اعتبار سنجی یا cross validation مدل های رگرسیون
- رگرسیون لجستیک
- رگرسیون پواسون

به همراه بررسی و تحلیل داده‌های مطالعه موردی در حوزه‌های مختلف با نرم افزار minitab و زبان برنامه نویسی R

سری های زمانی (۸ ساعت)

- مقدمات سری های زمانی شامل:

Time Series Plot ، توابع اتوکوواریانس و خودهمبستگی، خودهمبستگی جزئی، سری های زمانی ایستا و غیر ایستا، روش های رفع ناایستایی در میانگین و واریانس

- روش های کلاسیک تجزیه سری زمانی به روند و فصلی
- مدل سازی سری های زمانی شامل:

مدل میانگین متحرک Moving Average ، مدل خود همبسته Autoregressive ، مدل Arma ، مدل Arima

معیار های ارزیابی مدل های سری های زمانی شامل:

R^2، Adjusted R^2 ، AIC ، BIC و RMSE

- اعتبار سنجی یا cross validation مدل های سری زمانی

به همراه بررسی و تحلیل داده‌های مطالعه موردی در حوزه‌های مختلف با نرم افزار minitab و زبان برنامه نویسی R

تحلیل های آماری چند متغیره (۴ ساعت)

- معرفی توزیع نرمال چند متغیره
- تجزیه‌وتحلیل مولفه های اصلی
- تجزیه‌وتحلیل عاملی

به همراه بررسی و تحلیل داده‌های مطالعه موردی در حوزه‌های مختلف با نرم افزار minitab

موسسه آموزش عالی آزاد توسعه
https://tihe.ac.ir
۰۲۱-۸۶۷۴۱

آمار ناپارامتری (۳ ساعت)

- آزمون های ناپارامتری شامل:

آزمون Wilcoxon ، آزمون Sign ، آزمون Mann-Whitney ، آزمون Moods-Median ،
آزمونFriedman ، آزمون Run ، آزمون Kruskal-wallis

به همراه بررسی و تحلیل داده‌های مطالعه موردی در حوزه‌های مختلف با نرم افزار minitab