

## بسم الله الرحمن الرحيم

تمرینات سری سوم درس طراحی و تحلیل آماری فرایندها

دستورالعمل. طبق دستورالعمل بارگذاری شده در تمرین دوم + به مهلت تمرین توجه شود که در روند نمره‌دهی تاثیر دارد.

سوال ۱. در دیتاست NHANES بررسی کنید آیا Age و BPSysAve پیش‌بینی‌کننده‌ای برای TotChol هستند؟

مقادیر R2, pvalue, RMSE را محاسبه و تفسیر کنید.

یک نمودار scatter برای نمایش رابطه Age و TotChol همراه با خط رگرسیون رسم کنید.

سن را به بازه‌های ۰ تا ۱۰۰ با گام ۲۰ تقسیم کنید و Boxplot کلاسترول در این گروه‌ها بکشید.

سوال ۲. در دیتاست airquality مقدار Ozone را بر اساس Temp (درجه حرارت) و Wind مدل کنید. حذف NA لازم است. مدل خود را بر حسب هر دو متغیر، هر بار فقط با یکی از متغیرها، و اثر متقابل متغیرها در نظر بگیرید. کدام مدل را انتخاب می‌کنید و دلیل این انتخاب چیست؟

در مدلی که هر دو متغیر وجود دارند، مقدار اوزون را برای دمای ۸۵ و wind=10 پیش‌بینی کنید.

سوال ۳. برای دیتاست Boston مقدار متوسط قیمت خانه (medv) را بر اساس میانگین اتاق (rm) و (nox) مدل درست را برازش دهید.

برازش مدل، تفسیر ضرایب (با واحدها)، و بازه‌ی اطمینان ۹۵ درصدی ضرایب، و R2 را گزارش دهید.

نمودار پراکنش medv مقابل rm را رسم کرده و خط رگرسیون + ۹۵% CI را اضافه کن؛ نقاط را بر اساس متغیر age (درصد خانه‌های ساخته شده قبل از ۱۹۴۰) رنگ‌بندی کنید.

cars و MASS (چیکو)

سوال ۴. برای دیتاست biopsy از چیکو MASS مدلی مناسب از تمام متغیرها در نظر بگیرید. ماتریس ابهام را در نظر بگیرید و تفسیر کنید چه تعداد درست حدس زده شده اند؟

برای ستون v1 و v3 نموداری در برابر یکدیگر رسم کرده و رنگ نمودار را بر اساس احتمال بدخیم بودن 1 در نظر بگیرید. تفسیر کنید این نمودار چه چیزی را نشان می‌دهد.

سوال ۵. در دیتاست NHANES بررسی کنید که آیا سن (Age) و کالری دریافتی روزانه (Calories)

می‌توانند وزن (Weight) را پیش‌بینی کنند.

مقادیر  $R^2$  pvalue و RMSe را حساب و تفسیر کرده و ضرایب را نیز تفسیر کنید. در صورتی که نیاز به اثر متقابل می بینید آن را نیز وارد کنید.

باکس پلات کالری برای وزن را در صورتی رسم کنید که داده های calories بر حسب 0,1000,1500,2000,2500,3000,3500 تقسیم شوند..