

داده‌های auto.txt را در نظر بگیرید.

- 1- ابتدا بررسی نمایید که آیا توزیع قیمت خودرو به ازای هریک از سطوح متغیر استایل بدنه ماشین (style)، نرمال هست یا خیر.
 - 2- نموداری رسم کنید که متوسط قیمت خودروها را در مقابل استایل بدنه ماشین (style) نشان دهد. از این نمودار چه نتیجه‌ای می‌گیرید؟
 - 3- با فرض نرمال بودن توزیع قیمت خودرو به ازای هریک از سطوح متغیر (style)، آزمون کنید که آیا تغییر در استایل بدنه ماشین (style) بر متوسط قیمت خودرو تاثیرگذار است؟
 - 4- با فرض نرمال بودن توزیع قیمت خودرو به ازای هریک از سطوح متغیر (style)، با توجه به این مساله که می‌دانیم متغیرهای height و horsepower بر قیمت تاثیرگذار است، آزمون کنید که آیا تغییر در استایل بدنه ماشین (style) بر متوسط قیمت خودرو تاثیرگذار است؟
- 5- متغیرهای height، city_mpg، highway_mpg، length و horsepower را به عنوان متغیرهای توضیحی و قیمت خودرو را به عنوان متغیر پاسخ در نظر بگیرید. به سوالات زیر پاسخ دهید:

انتخاب متغیر:

- الف- ابتدا مدلی با تمام متغیرها برآزش دهید. تفسیر عرض از مبدا و ضریب fueltype و horsepower را بیان کنید.
- ب- آزمون معناداری کلی مدل را انجام دهید و آن را معنا کنید.
- ب- آزمون معناداری ضرایب رگرسیون جزئی را انجام دهید و آن را معنا کنید.
- ت- توضیح دهید چرا باید تا جایی که امکان دارد، متغیرهایی که حضورشان در مدل معنادار نیست را حذف کرد؟
- ث- همه متغیرها را وارد مدل کنید. آزمون کنید که آیا می‌توان همزمان متغیرهای length و height را از مدل حذف کرد؟ همچنین آزمون کنید که آیا ضریب متغیر height برابر با دو برابر ضریب متغیر fueltype به اضافه ضریب متغیر horsepower هست؟
- ج- با استفاده از روش stepwise مدل مناسبی را برآزش دهید. مراحل اجرای این الگوریتم را بر اساس خروجی‌ها توضیح دهید.

پیش بینی: (مدل قسمت ج را در نظر بگیرید)

ج- پیش بینی نقطه‌ای و فاصله‌ای برای قیمت ماشینی با خصوصیات زیر ارائه دهید:

fueltype=gas, height=50, city_mpg=22, highway_mpg=28, length=200, horsepower=170

ح- برآورد نقطه‌ای و فاصله‌ای برای تخمین متوسط قیمت ماشینی با الگوی قسمت ج برای متغیرهای توضیحی، ارائه دهید.

بررسی فرضیات مدل: (مدل قسمت ج را در نظر بگیرید)

- خ- آیا فرض نرمال بودن توزیع برقرار است؟ توضیح دهید که اگر این فرض برقرار نباشد، چه مشکلی پیش خواهد آمد.
- د- بررسی کنید که آیا فرض ثبات واریانس برقرار است؟ توضیح دهید که اگر به مساله عدم ثبات واریانس توجه نکنیم، چه مشکلی پیش می‌آید؟

ذ- وزن مناسب را با توجه به قسمت د به دست آورید و رگرسیون وزنی بزنید. با رسم نمودار $w_i^{1/2}e_i$ در مقابل $w_i^{1/2}\hat{y}_i$ بررسی کنید که آیا مشکل عدم ثبات واریانس با وزن انتخابی شما برطرف شده است؟

بررسی تاثیرگذاری مشاهدات: (مدل قسمت ج را در نظر بگیرید)

ر- داده‌های دورافتاده، اهرم گون و تاثیرگذار را مشخص کنید. توضیح دهید هر یک چه تاثیری بر مدل می‌گذارند.

ز- با استفاده از رسم شکل، نقاط اهرم گون، دورافتاده و برخی از نقاط تاثیرگذار را مشخص کنید.

بررسی مشکل همخطی چندگانه: (مدل قسمت ج را در نظر بگیرید)

ژ- بررسی کنید که آیا مشکل همخطی چندگانه وجود دارد یا خیر؟ توضیح دهید هم خطی چندگانه چه مشکلی را برای مدل ایجاد می‌کند.

س- با توجه به تمامی مراحل بالا، توضیح دهید اگر خودتان بخواهید مدل مناسبی برازش دهید، چه روندی را طی می‌کنید و چرا مدل شما می‌تواند مناسب تر از مدل ارائه شده توسط الگوریتم stepwise باشد.

موفق باشید.