

داده‌های مورد استفاده در تکالیف رگرسیون ۱ را در نظر بگیرید و به سوالات زیر، با ذکر تمامی فرمول‌های مورد استفاده، پاسخ دهید:

بخش اول : ابتدا در نظر بگیرید، تمامی فرضیات مدل رگرسیون خطی برقرار است:

الف- رگرسیون خطی با تمام متغیرهای توضیحی موجود به داده‌ها برازش دهید.

ب- تفسیر عرض از مبدا و دو تا از ضرایب رگرسیونی را بیان کنید.

پ- آزمون معناداری کلی مدل رگرسیونی را انجام دهید.

ت- آزمون ضرایب رگرسیون جزئی را انجام دهید و نتایج آزمون را تفسیر کنید.

ث- ضرایب رگرسیونی استاندارد شده را به دست آورید. چه استفاده‌ای از این ضرایب می‌توانید داشته باشید؟

ج- آیا مشکل همخطی چندگانه برقرار است؟ توضیح دهید که، همخطی چندگانه چه مساله‌ای برای مدل‌ها ایجاد می‌کند و

چگونه این مشکل را می‌توان برطرف کرد؟

بخش دوم: انتخاب متغیر

الف- با استفاده از هر سه الگوریتم انتخاب پیشرو، حذف پسرو و گام به گام، مدل‌های مناسبی به داده‌ها برازش دهید. مراحل انجام شده در هر الگوریتم را با توجه به خروجی‌ها توضیح دهید.

ب- از بین مدل‌های پیشنهاد شده در قسمت الف همین بخش، کدام مدل را انتخاب می‌کنید؟ چرا؟

پ- ایراد الگوریتم‌های بالا در انتخاب متغیر (مدل) چیست؟

بخش سوم: متغیرهای نشانگر

یکی از متغیرهای توضیحی رسته‌ای مدل را در نظر بگیرید (اگر چنین متغیریدر داده‌ها ندارید با رسته‌ای کردن یکی از متغیرهای توضیحی آن را ایجاد کنید). همچنین دو تا از متغیرهای کمی مدل را در نظر بگیرید. سپس بررسی کنید که آیا ابرصفحه‌هایی که بین پاسخ و آن دو متغیر کمی برازش داده می‌شوند در زیر جامعه‌های متناظر با سطوح متغیر رسته‌ای شما با هم برابر هستند یا خیر.

بخش چهارم: بررسی فرضیات مدل

مدل بخش یک را در نظر بگیرید.

الف- آیا فرض ثبات واریانس برقرار است؟ در صورت عدم برقراری این فرض، چه مشکلاتی به وجود خواهد آمد. مشکل به وجود آمده را برطرف کنید.

ب- آیا فرضیه نرمال بودن توزیع خطا برقرار است؟ در صورت عدم برقراری این فرضیه، چه مشکلاتی به وجود خواهد آمد.

بخش پنجم: داده‌های دورافتاده، تاثیرگذار و اهرم‌گون

مدل برآزش داده شده در بخش اول را در نظر بگیرید.

الف- داده‌های دورافتاده را پیدا کنید، شماره داده‌ها را ذکر کنید. بیان کنید که حضور داده‌های دور افتاده چه تاثیری در برآزش مدل دارد؟

ب- داده‌های اهرم‌گون را پیدا کنید، شماره داده‌ها را ذکر کنید. بیان کنید که حضور این نوع داده‌ها چه تاثیری در برآزش مدل دارد؟

پ- داده‌های تاثیرگذار را پیدا کنید (حداقل سه خصوصیت مهم مدل را جهت پیدا کردن چنین داده‌هایی در نظر بگیرید)، شماره داده‌ها را ذکر کنید. بیان کنید که حضور این نوع داده‌ها چه تاثیری در برآزش مدل دارد؟ سعی کنید تا جایی که امکان دارد، مشکل را برطرف کنید.

موفق باشید.