

۱- داده های Daroo.sav را در نظر بگیرید. در این داده‌ها، y_i متوسط فروش کارمندان i -امین شرکت‌های پخش دارو است. n_i تعداد کارمندان واحد فروش شرکت i -ام و متغیر x_{i1} نشان دهنده نوع شرکت (دولتی (۱)، خصوصی (۰) و یا تعاونی (۳)) و x_{i2} سابقه فعالیت شرکت است.

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

الف*) رگرسیون خطی به مدل برازش دهید.

ب*) تفسیر ضرایب 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*) آیا متوسط فروش شرکت‌های تازه تاسیس بخش دولتی با سایر بخش‌ها متفاوت است؟ آزمون کنید.

ج*) آیا تاثیر سابقه فعالیت شرکت بر متغیر پاسخ به نوع شرکت بستگی دارد؟

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: چک کردن فرضیت مدل:

الف) آیا فرضیه ثبات و اریانس برقرار است؟ در صورت عدم برقراری این فرضیه، 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 سوم: داده‌های دورافتاده، تاثیرگذار و اهرم‌گون

مدل برازش داده شده در قسمت الف 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 برازش مدل دارد؟

ب) داده‌های اهرم‌گون را پیدا کنید، شماره داده‌ها را ذکر کنید. بیان کنید که حضور این نوع داده‌ها چه تاثیری در برازش 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 برازش مدل دارد؟ سعی کنید تا جایی که امکان دارد، مشکل را برطرف کنید.