

در فایل داده اطلاعات پزشکی، متغیر قند خون را متغیر پاسخ در نظر بگیرید.

الف- مدلی با تمامی متغیرها برازش دهید.

ب- اطلاعات فرد جدیدی را در نظر بگیرید و پیش بینی‌های نقطه‌ای و فاصله‌ای برای قند خون این فرد ارائه دهید.

پ- نقاط اهرم‌گون، دور افتاده و نیز تاثیرگذار را مشخص کنید. بیان کنید که هر نوع از این نقاط چه مشکلاتی می‌توانند ایجاد کنند.

ت- مهمترین نقطه تاثیرگذار را حذف کنید و دوباره مدل را برازش دهید.

ث برای مدل قسمت الف، مساله همخطی چندگانه را بررسی کنید. بیان کنید که این مساله چه تاثیری بر روی مدل دارد. سعی کنید این مشکل را برطرف کنید و دوباره مدل را برازش دهید.

ج- مدل قسمت الف را در نظر بگیرید. با استفاده از الگوریتم‌های forward، backward و stepwise مدل مناسب را برازش دهید.

چ- بیان کنید که چرا مدلی که از طریق این الگوریتم‌ها به دست می‌آید ممکن است به خوبی مدلی که با دخالت شما و مرحله به مرحله به دست می‌آید، نباشد.

ح- فرضیات مدل مانند نرمال بودن توزیع خطاها و ثبات واریانس را بررسی کنید.

خ- مدل قسمت الف را در نظر بگیرید. آزمون‌های زیر را در سطح معناداری ۵ درصد انجام دهید.

۱- آزمون معناداری کلی مدل رگرسیونی را انجام دهید.

۲- آزمون ضرایب رگرسیون جزئی را انجام دهید.

۳- آیا ضریب رگرسیونی متغیر جنسیت برابر ۲ است؟

۴- آیا ضریب متغیر hdl با ضریب متغیر ldl برابر است؟

۵- آیا ضریب متغیر hdl با ضریب متغیر ldl برابر و ضریب متغیر جنسیت برابر ۲ است؟