

رگرسیون غیرخطی بر روی تولید ناخالص و جمعیت

صورت مساله

مدل رگرسیون غیرخطی $Y = y_0 N^a + \epsilon$ ، که در آن $y_0 = 6611$ ، در کلاس بحث و بررسی شد. با در نظر گرفتن این مدل، مطالب کلاس و برای داده‌های جمعیت و تولید ناخالص شهری ۳۶۶ شهر آمریکا که در فایل gmp.dat قرار دارند، موارد زیر را انجام دهید:

(۱) با استفاده از دستور read.table داده‌های موجود در فایل gmp.dat را به R فراخوانی کنید و آن را به یک data.frame تبدیل کنید.

(۲) دو ستون آخر این مجموعه داده، برای هر شهر، کل تولید ناخالص شهری (NY) و تولید ناخالص شهری به ازای هر نفر (Y) هستند. با استفاده از آن‌ها برداری بسازید که حاوی جمعیت هر شهر، N ، می‌باشد و آن را به ساختار داده gmp با نام pop اضافه کنید.

(۳) MSE مدل قانون توانی را به ازای $a = 0/15$ محاسبه کنید.

(۴) یک دنباله از مقادیر a از $0/10$ تا $0/15$ با فاصله‌های $0/005$ بسازید. MSE را برای هر کدام از این مقادیر محاسبه و در یک بردار به نام mse ذخیره کنید.

(۵) نمودار ریشه مقادیر MSE به دست آمده در مرحله قبل را در مقابل مقادیر a رسم کنید. چه مقداری از a مقدار مناسب است. آن را با برآورد کمترین مربعات خطا که در کلاس به دست آمد مقایسه کنید و در یک نمودار نشان دهید.