

رگرسیون مفصل

اصغر اکبری* (as.akbari54@gmail.com)

*گروه ریاضی، دانشگاه آزاد اسلامی، واحد مبارکه

چکیده

تحلیل رگرسیونی یکی از روش های متداول آماری در مدل سازی رابطه بین متغیرها است، اما این روش در شکل پایه ای خود یعنی کمترین مربعات معمولی^۱ روش مناسبی در مدل سازی مسایل مالی نیست زیرا در اغلب اینگونه مسایل رابطه ها غیر خطی بوده و توزیع احتمالاتی متغیر وابسته نیز غیر نرمال است. یک روش موفق برای فایق آمدن بر این مشکلات روش مدل خطی تعمیم یافته^۲ است که در آن نیز لازم است توزیع متغیر وابسته از خانواده توزیع های نمایی باشد. یک روش جایگزین برای دو روش پیش گفته روش رگرسیون مفصل^۳ است. مزیت اصلی روش رگرسیون مفصل این است که هیچ محدودیتی در نوع توزیع متغیر وابسته وجود ندارد. در این نوشتار ابتدا روش استفاده از رگرسیون مفصل را ارائه نموده و سپس چند مثال را ارائه کرده و در آن ها سه روش را با هم مقایسه می کنیم. در مثال ها از توزیع های با دم سنگین که در مدل سازی مسایل مالی کاربرد دارند استفاده می کنیم.

واژه های کلیدی: کمترین مربعات معمولی، مدل خطی تعمیم یافته، تابع مفصل، رگرسیون مفصل

۱- مقدمه

تجزیه و تحلیل رگرسیون یکی از روش های آماری با کاربردهای وسیع است که از نظر تاریخی به سال ۱۸۷۵ و فعالیت های پژوهشی فرانسیس گالتون^۴ باز می گردد. ایده اصلی این روش، تشخیص رابطه بین چند متغیر است. ساختار مدل رگرسیونی به این صورت است که یک متغیر وابسته یا پاسخ (Y) با چند متغیر مستقل (X_1, \dots, X_n) مرتبط در نظر گرفته می شود. مساله اصلی این است که تعیین کنیم توزیع متغیر پاسخ چگونه با متغیرهای مستقل در ارتباط است.

مدل پایه ای کمترین مربعات معمولی شکل خاصی را برای این ارتباط ارائه می دهد. در این مدل توزیع Y به شرط معلوم بودن متغیرهای مستقل، نرمال با واریانس ثابت (بدون وابستگی به متغیرهای مستقل) و میانگینی است که تابعی از متغیرهای مستقل به صورت $E[Y|X_1 = x_1, \dots, X_n = x_n] = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$ است. در این روش دو حالت معادل "حداکثر درست نمایی" و "حداقل مربعات" برای برآورد پارامترهای نامعلوم استفاده می شود. یک معیار مناسب بودن مدل کاهش واریانس Y در اثر کاربرد متغیرهای مستقل است.

داده های مالی به ندرت از توزیع نرمال پیروی می کنند و اغلب توزیع این داده توزیع های نامتقارن است، اضافه بر آن در نظر گرفتن رابطه خطی بین میانگین Y و متغیرهای مستقل به دلیل ساده کردن مساله از دیدگاه ریاضی است. در نظر گرفتن رابطه خطی در صورتی که توزیع متغیر مورد بررسی نرمال نباشد می تواند گمراه کننده باشد (مک نیل^۵ و استرامن^۶ ۲۰۰۲).

^۱ Ordinary Least Square (OLS)

^۲ Generalized Linear Models (GLM)

^۳ Copula Regression

^۴ Francis Galton

^۵ Mc Neil

^۶ Straumann