



تمرین : سری ۴

مدّرس: دکتر محسن شریفی تبار

سوال ۱

به سوالات زیر پاسخ دهید :

- الف) نشان دهید با فرض برقراری مفروضات سه گانه، تخمین های کمینه ی مربعات در رگرسیون ساده نااریب هستند. به عبارت دیگر نشان دهید : $E(\hat{\beta}_i) = \beta_i$
- ب) چنانچه β_0 و β_1 تخمین های ضرایب رگرسیون ساده باشند، مقدار $Cov(\hat{\beta}_0, \hat{\beta}_1)$ را به دست آورید.

سوال ۲

مدل $y_i = \beta_0 + \beta_1 x_i + e_i$ را در نظر بگیرید که در آن خطاهای e_i مستقل اند و دارای میانگین صفر هستند، اما $Var(e_i) = \rho_i^2 \sigma^2$ که در آن ρ_i مقادیر ثابت معینی هستند. بنابراین خطاها واریانس یکسانی ندارند. این شرایط از آن جا ناشی می شود که y_i ها میانگین مشاهدات زیادی از x_i باشد. در این حالت اگر y_i میانگین n_i مشاهده ی مستقل باشد، صورت مسئله به صورت مقابل درمی آید:

$$\rho_i^{-1} y_i = \rho_i^{-1} \beta_0 + \rho_i^{-1} \beta_1 x_i + \rho_i^{-1} e_i$$

یا

$$z_i = u_i \beta_0 + v_i \beta_1 + \delta_i$$

که در آن $\delta_i = \rho_i^{-1} e_i$ و $v_i = \rho_i^{-1} x_i$ ، $u_i = \rho_i^{-1}$

الف) ابتدا نشان دهید: $\rho_i^2 = \frac{1}{n_i}$

ب) نشان دهید مفروضات سه گانه برای مدل جدید نیز صادق است.

ج) کمترین مربعات β_0 و β_1 را بیابید.

د) نشان دهید اجرای روش کمترین مربعات روی مدل جدید که در قسمت قبل انجام شد معادل با کمینه کردن عبارت زیر است:

$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \rho_i^{-2}$$

ه) واریانس تخمین های قسمت (ج) را به دست آورید.

سوال ۳

مدل زیر را برای بررسی وجود یا عدم وجود تبعیض جنسیتی در پرداخت حقوق کارمندان یک شرکت در نظر بگیرید:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_2^2 + \xi$$

که در آن Y حقوق سالانه بر حسب هزار دلار است و

$$x_1 = \begin{cases} 1 & \text{زن} \\ 0 & \text{مرد} \end{cases} \quad (۱)$$

و x_2 نیز برابر با تجربه ی کار بر حسب سال می باشد. حال این مدل را بر روی 200 داده از کارمندان fit می کنیم و به دست می آید: $SSE = 783.90$. حال مدل کاهش یافته ی $Y = \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \xi$ را نیز روی همان داده ها fit می کنیم و به دست می آید: $SSE = 795.23$. آیا این داده ها شواهد کافی برای تایید این ادعا که میانگین حقوق کارمندان متأثر از جنسیت آن ها است، ارائه می دهد؟ ($\alpha = 0.05$)

سوال ۴

فرض کنید که هم اتاقی شما دانشجویی باهوش است، اما هیچ اقتصاد نخوانده، و کمی آمار خوانده است. (مشخصا، او رگرسیون ساده اما نه چندگانه را می فهمد). او برای تبیین اینکه چه چیزی بر سطح قیمت ها در ایالت های متحده تاثیر می گذارد، رگرسیون قیمت ها در ایالت های متحده برحسب 100 متغیر تصادفی مختلف را هر بار با یکی (یعنی در 100 رگرسیون ساده) حساب کرده است. افزون بر این، او ظاهرا این متغیرها را به طریقی کاملا اتفاقی بدون هیچ تصویری از روابط علت و معلولی بالقوه انتخاب کرده است. او پنج متغیر کشف کرده است که از لحاظ آماری در سطح $\alpha = 0.05$ تمیزپذیرند و نتیجه گرفته است که هر یک از آن ها بر قیمت ها در ایالت های متحده تاثیر می گذارند.

الف) به او توضیح دهید که چه تردید هایی درباره ی نتیجه گیری اش دارند.

ب) اگر 20 متغیر تمیزپذیر آماری کشف کرده بود، اکنون چه تردید هایی داشتید؟ او چگونه می تواند تحلیلش را بهبود بخشد؟

سوال ۵

در سال 1970، نمونه ای تصادفی مرکب از 50 مرد آمریکایی با سن 35 تا 54، رابطه ی زیر را بین درآمد سالانه ی Y (به دلار) و تحصیلات X (به سال) نشان دادند:

$$\hat{Y} = 1200 + 800x$$

متوسط درآمد $\bar{Y} = 10000$ دلار و متوسط تحصیلات $\bar{X} = 11$ سال با $\sum x^2 = 900$ بود. انحراف معیار مانده های پیرامون خط برازنده شده $s = 7300$ دلار بود.

الف) یک بازه ی اطمینان 95 درصدی برای شیب جامعه ای بسازید.

ب) آیا رابطه ی درآمد با تحصیلات در سطح 5 درصد از لحاظ آماری تمیزپذیر است؟

ج) درآمد مردی را که 2 سال از دبیرستان را تمام کرده است ($X = 10$)، پیشگویی کنید. بازه ای به قدر کافی پهن را ارائه کنید که روی آن با بخت 95 به 5 شرط بندی خواهید کرد.

د) آیا منصفانه خواهد بود که بگوییم هر یک سال تحصیل بیشتر از 800 دلار ارزش دارد؟ چرا؟