۱۲.۱ تمرین‌ها

۱. نشان دهید $\mathrm{Cov}(\bar{y}, \hat{\beta}_1) = 0$ و با به‌کارگیری آن برهانی دیگر برای رابطه‌ی زیر بیان کنید.

$$\mathrm{Var}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_x^2} \right)$$

۲. مجموعه داده‌های table.b1 را از بسته‌ی MPV در نرم‌افزار R در نظر بگیرید، که از جدول B.1 در پیوست کتاب مونتگمری (۱۹۹۲) برگرفته شده است. این داده‌ها دربردارنده‌ی عملکرد ۲۶ تیم لیگ ملی فوتبال آمریکا در سال ۱۹۷۶ است. انتظار می‌رود تعداد یاردها، نوعی امتیاز در فوتبال آمریکایی که در حمله‌ی حریف به‌دست می‌آید (x_8)، بر روی تعداد بازی‌های برده شده توسط تیم (y) اثر داشته باشد.

الف. یک مدل رگرسیون خطی ساده بین x_8 و y برازش دهید.

ب. مقدار مانده‌های مدل رگرسیونی بالا را به‌دست آورید.

ج. واریانس خطاها را براورد کنید.

۳. وزن و فشار خون ۲۶ مرد که به طور تصادفی از بین مردان ۲۵ تا ۳۰ سال انتخاب شده‌اند، اندازه‌گیری شده و در جدول ۷.۱ آمده است (مایرز، ۱۹۹۰).

الف. براورد ضریب همبستگی ρ میان وزن و فشارخون را به‌دست آورید.

ب. مدل رگرسیون خطی ساده را به این داده‌ها برای پیشگویی فشارخون بر اساس وزن برازش دهید.

جدول ۱.۷: داده‌های وزن و فشار خون

وزن	۱۶۵	۱۹۷	۱۸۰	۱۵۵	۲۱۲	۱۷۵	۱۹۰
فشارخون	۱۳۰	۱۳۳	۱۵۰	۱۲۸	۱۵۱	۱۴۶	۱۵۰
وزن	۲۱۰	۲۰۰	۱۴۹	۱۵۸	۱۶۹	۱۷۰	۱۷۲
فشارخون	۱۴۰	۱۴۸	۱۲۵	۱۳۳	۱۳۵	۱۵۰	۱۳۵
وزن	۱۵۹	۱۶۸	۱۷۴	۱۸۳	۲۱۵	۱۹۵	۱۸۰
فشارخون	۱۲۸	۱۳۲	۱۴۹	۱۵۸	۱۵۰	۱۶۳	۱۵۶
وزن	۱۴۳	۲۴۰	۲۳۵	۱۹۲	۱۸۷		
فشارخون	۱۲۴	۱۷۰	۱۶۵	۱۶۰	۱۵۹		

۴. فرض کنید $y_1 = \theta + \epsilon_1$، $y_2 = \theta - \phi + \epsilon_2$ و $y_3 = \theta + 2\phi + \epsilon_3$، که در آن $E(\epsilon_i) = 0$، $i = 1, 2, 3$. براورد کم‌ترین توان دوم θ و ϕ را بیابید.

۵. فرض کنید می‌خواهیم براورد کم‌ترین توان دوم مدل

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \; i = 1, \ldots, n$$

را با در نظر گرفتن قید $\beta_1 = 2\beta_0$ به‌دست آوریم. تابع لاگرانژ را تشکیل داده و براوردهای کم‌ترین توان دوم برای β_0 و β_1 را با در نظر گرفتن این قید به‌دست آورید. به شیوه‌ای دیگر، در مدل بالا قرار دهید $\beta_1 = 2\beta_0$ و براورد کم‌ترین توان دوم β_1 را به‌دست آورید و پاسخ را با حالت پیش مقایسه کنید.

۶. فرض کنید $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$، $i = 1, \ldots, n$، که در آن β_0 معلوم و β_1 نامعلوم است. اگر $\hat{\beta}_1$ براورد کم‌ترین توان دوم β_1 باشد، آن‌گاه

$$\hat{y}_i = \beta_0 + \hat{\beta}_1 x_i, \; e_i = y_i - \hat{y}_i,$$

درستی یا نادرستی عبارت‌های زیر را بررسی کنید.

الف. $\sum_{i=1}^{n} e_i = 0$، ب. $\sum_{i=1}^{n} e_i x_i = 0$، ج. $\sum_{i=1}^{n} e_i \hat{y}_i = 0$.

۷. در مدل رگرسیون خطی ساده $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$، $i = 1, \ldots, n$، با فرض $\epsilon_i \sim N(0, \sigma^2)$ مقدار $P(y_i > \beta_0 + \beta_1 x_i)$ را به‌دست آورید.

۸. نشان دهید در مدل رگرسیون خطی گذرا از مبدأ $y_i = \beta_1 x_i + \epsilon_i$، $i = 1, \ldots, n$،

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

یک براوردگر نااریب برای β_1 است و واریانس $\hat{\beta}_1$ را با واریانس براوردگر کم‌ترین توان دوم β_1 مقایسه کنید.

۹. برای مدل رگرسیون خطی ساده و مقدارهای برازش یافته به روش براورد کم‌ترین توان دوم نشان دهید

$$\sum_{i=1}^{n} \mathrm{Var}(\hat{y}_i) = 2\sigma^2$$

۱۰. در مدل رگرسیون خطی ساده تمرین ۹، مقدار $\mathrm{Cov}(\hat{y}_i, \bar{y})$ را به‌دست آورید.

۱۱. فرض کنید در برازش دو خط رگرسیونی

$$y_i = \alpha_0 + \alpha_1 x_i + \epsilon_i, x_i = \beta_0 + \beta_1 z_i + e_i,\ i = 1, \ldots, n$$

براوردهای α_0، α_1، β_0 و β_1 به روش کم‌ترین توان دوم به ترتیب به صورت $\hat{\alpha}_0$، $\hat{\alpha}_1$، $\hat{\beta}_0$ و $\hat{\beta}_1$ به‌دست آمده باشند. اگر بخواهیم در مدل $y_i = \theta_0 + \theta_1 z_i + \epsilon_i$، $i = 1, \ldots, n$، θ_0 و θ_1 را براورد کنیم، براوردهای θ_0 و θ_1 را بر حسب $\hat{\alpha}_0$، $\hat{\alpha}_1$، $\hat{\beta}_0$ و $\hat{\beta}_1$ به‌دست آورید.

۱۲. فرض کنید میان دو متغیر تصادفی x و y رابطه‌ی $\frac{x}{4y} + \frac{y}{x} = 1$ برقرار است. ضریب همبستگی $u = 3x + 5$ و $v = 5y + 3$ را به‌دست آورید.

۱۳. فرض کنید y_1 و y_2 دو متغیر تصادفی مستقل با میانگین‌های به ترتیب 2β و 3β باشند. براورد β را به روش کم‌ترین توان دوم به‌دست آورید.

۱۴. فرض کنید متغیرهای تصادفی $y_1, \ldots, y_n$ به ترتیب دارای میانگین‌های $\beta, 2\beta, \ldots, n\beta$ باشند. براورد β را به روش کم‌ترین توان دوم به‌دست آورید.

۱۵. فرض کنید $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$، $i = 1, \ldots, n$، با انجام تبدیل‌های زیر

$$x_i^* = \frac{x_i - \bar{x}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}},\quad y_i^* = \frac{y_i - \bar{y}}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}},\quad i = 1, \ldots, n$$

براورد $\hat{\beta}_1$ (به روش کم‌ترین توان دوم) را بر حسب x_i^*ها و y_i^*ها به‌دست آورید و در مورد پاسخ به‌دست آمده اظهارنظر کنید.

۱۶. در مدل رگرسیون خطی با مقدارهای برازش داده شده به شیوه‌ی کم‌ترین توان دوم به صورت زیر

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \ i = 1, \ldots, n,$$

مقدار واریانس نمونه‌ای مقدارهای برازش داده شده یعنی

$$\frac{1}{n-1}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

را بر حسب r و مجموع‌های توان دوم x یا y بیابید که در آن $\bar{y}$ میانگین $\hat{y}_i$ها است.

۱۷. اگر براوردهای کم‌ترین توان دوم β_0 و β_1 در معادله‌ی مدل رگرسیون خطی $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$، $i = 1, \ldots, n$، برابر $\hat{\beta}_0$ و $\hat{\beta}_1$ باشند و قرار دهیم $x_i^* = a x_i$ و $y_i^* = b y_i$ که در آن a و b مقدارهای حقیقی معلوم هستند، براوردهای کم‌ترین توان دوم α_0 و α_1 را در مدل

$$y_i^* = \alpha_0 + \alpha_1 x_i^* + e_i, \ \ i = 1, \ldots, n,$$

بر حسب $\hat{\beta}_0$ و $\hat{\beta}_1$ به‌دست آورید.

۱۸. فرض کنید $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$، $i = 1, \ldots, n$ و به اشتباه مدل بدون عرض از مبدأ

$$y_i = \beta_1 x_i + \epsilon_i, \ \ i = 1, \ldots, n,$$

را به داده‌ها برازش دهیم. میانگین توان دوم خطا $(\mathrm{E}(\hat{\beta}_1 - \beta_1)^2)$ در مدل گذرا از مبدأ را با $\mathrm{Var}(\hat{\beta}_1)$ در مدل درست با عرض از مبدأ مقایسه کنید.

۱۹. در مدل رگرسیون خطی ساده برازش داده شده به روش کم‌ترین توان دوم با مانده‌های $e_1, \ldots, e_n$، $\mathrm{Var}(e_i)$ها را به‌دست آورید.

۲۰. هسویی و همکاران (۱۹۹۵) اثر نسبت مولار نوعی اسید (x) را بر روی چسبندگی ذاتی نوعی پلی‌استر (y) بررسی کرده‌اند. داده‌های زیر از این مطالعه در جدول ۸.۱ آمده است (مونتگومری،

۲۰۱۳). این مجموعه داده در بسته‌ی نرم‌افزاری MPV با نام p۲۱۴. قابل دسترسی است.

جدول ۸.۱: داده‌های نسبت مولار و چسبندگی پلی استر

نسبت مولار	۱/۰	۰/۹	۰/۸	۰/۷	۰/۶	۰/۵	۰/۴	۰/۳
چسبندگی	۰/۴۵	۰/۲۰	۰/۳۴	۰/۵۸	۰/۷۰	۰/۵۷	۰/۵۵	۰/۴۴

الف. نمودار پراکنش داده‌ها را رسم کنید.

ب. مدل رگرسیون خطی ساده را به داده‌ها برازش داده و ضریب‌های مدل را براورد کنید.

ج. مقدارهای برازش یافته‌ی چسبندگی پلی استر را محاسبه کنید.

۲۱. داده‌های جدول ۹.۱ شدت یک بیماری خاص و درجه‌ی حرارت بدن بیمار را برای ۱۰ بیمار نشان می‌دهد (انجمن آسیب شناسی گیاهی آمریکا [1]). هدف تعیین رابطه‌ی بین شدت بیماری و درجه‌ی حرارت بدن بیمار است.

جدول ۹.۱: داده‌های شدت بیماری

شدت بیماری	۱/۹	۳/۱	۳/۳	۴/۸	۵/۳	۶/۱	۶/۴	۷/۶	۹/۸	۱۲/۴
درجه‌ی حرارت	۲	۱	۵	۵	۲۰	۲۰	۲۳	۱۰	۳۰	۲۵

الف. مدل رگرسیون خطی ساده را به داده‌ها برازش داده و ضریب‌های مدل را براورد کنید.

ب. مقدارهای برازش یافته‌ی شدت بیماری را محاسبه کنید.

۲۲. در مدل رگرسیون خطی ساده $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$، $i = 1, \ldots, n$، نشان دهید

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$$

۲۳. درستی رابطه‌ی زیر را نشان دهید (r ضریب همبستگی نمونه‌ای بین x و y است).

$$\text{SSE} = (1 - r^2)S_{yy}$$

[1] American Phytopathological Society: [http:\\www.apsnet.org]