حال با به‌کارگیری روش کم‌ترین توان دوم، بایستی مجموع توان دوم خطاهای تبدیل یافته در زیر را کمینه کنیم،

$$Q^*(\beta_0, \beta_1) = \sum_{i=1}^{n} (\sqrt{w_i}\epsilon_i)^2 = \sum_{i=1}^{n} [\sqrt{w_i}(y_i - \beta_0 + \beta_1 x_i)]^2$$

با توجه به کاو بودن تابع بالا، β_0 و β_1 که Q^* را کمینه کند، در معادله‌های زیر صدق می‌کنند

$$\begin{cases} \frac{\partial Q^*(\beta_0,\beta_1)}{\partial \beta_0} = -2\sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i) = 0 \\ \frac{\partial Q^*(\beta_0,\beta_1)}{\partial \beta_1} = -2\sum_{i=1}^{n} w_i x_i(y_i - \beta_0 - \beta_1 x_i) = 0 \end{cases} \quad (11.3)$$

از حل معادله‌های (۱۱.۳) که به **معادله‌های نرمال وزنی** معروف هستند، داریم

$$\hat{\beta}_0 = \bar{y}_w - \hat{\beta}_1 \bar{x}_w \ , \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} w_i x_i(y_i - \bar{y}_w)}{\sum_{i=1}^{n} w_i x_i^2 - \sum_{i=1}^{n} w_i \bar{x}_w^2} \quad (12.3)$$

که در آن $\bar{x}_w = \frac{1}{\sum_{i=1}^{n} w_i}\sum_{i=1}^{n} w_i x_i$ و $\bar{y}_w = \frac{1}{\sum_{i=1}^{n} w_i}\sum_{i=1}^{n} w_i y_i$. براوردهای به‌دست آمده در (۱۲.۳) را براوردهای کم‌ترین توان دوم وزنی برای عرض از مبدأ و شیب مدل رگرسیون خطی ساده با واریانس ناهمگن می‌گویند.

۸.۳ تمرین‌ها

۱. در مدل‌های رگرسیونی زیر تعیین کنید که آیا مدل مورد نظر قابل تبدیل به یک مدل خطی هست یا خیر؟ در صورت مثبت بودن پاسخ، تبدیل‌های مورد نیاز برای خطی سازی و مدل خطی شده را

بنویسید.

الف. $\mathrm{E}(y|x) = e^{\beta_0 + \beta_1 \ln(x_1 + x_2)}$ ب. $\mathrm{E}(y|x) = \beta_0 + x_1 e^{\beta_2 x_2}$

ج. $\mathrm{E}(y|x) = \beta_0 + e^{\beta_1 x}$ د. $\mathrm{E}(y|x) = \beta_0 + \frac{\beta_1}{x_1} + \beta_2 x_2^2$

ه. $\mathrm{E}(y|x) = \frac{1}{1+e^{-(\beta_0 + \beta_1 x_1 \beta_2 x_2)}}$ و. $\mathrm{E}(y|x) = \beta_0 + \ln(\beta_1 x_1) + \beta_2 x_2$

ز. $\mathrm{E}(y|x) = \frac{\beta_0 \beta_1}{\beta_0 \sin^2 x + \beta_1 \cos^2 x}$ ح. $\mathrm{E}(y|x) = \beta_1 + (1 + \beta_1) e^{-\beta_2 (x-2)}$

۲. در تمرین ۱، تفسیر ضریب‌های رگرسیونی β_i، $i = 0, 1, 2$ را برای هر مورد بیان کنید.

۳. داده‌های جدول ۳.۳ مربوط به مطالعه‌ای درباره‌ی سانحه‌های رانندگی و ایمنی وسایل نقلیه است. متغیر x سرعت وسیله‌ی نقلیه و y فاصله‌ی طی شده تا توقف کامل پس از ترمز کردن را نشان می‌دهد. نتیجه‌ی ۱۹ آزمایش در جدول ۳.۳ آمده است.

جدول ۳.۳: داده‌های سانحه‌های رانندگی

x	۲۵	۳۵	۶۰	۴۵	۵۰	۳۷/۵	۳۰
y	۳۷/۴	۵۷/۷	۳۳۷/۶	۱۴۲/۵	۱۸۲/۴	۶۷/۵	۳۷/۵
x	۵۵	۶۰	۲۲/۵	۶۵	۵۰	۲۰	۱۵
y	۲۲۵/۰	۲۵۸/۱	۲۵/۰	۲۹۷/۴	۱۷۰/۰	۲۰/۰	۱۳/۵
x	۲۷/۵	۵۵	۴۰	۴۵	۱۷/۵		
y	۴۰/۸	۲۰۷/۸	۱۰۵/۵	۱۳۲/۶	۱۹/۱		

الف. نمودار پراکنش (x, y) را رسم کنید. آیا به نظر شما مدل رگرسیون خطی برای داده‌ها مناسب است؟

ب. مدل رگرسیون خطی ساده را به داده‌ها برازش دهید و مانده‌ها را حساب کنید. مانده‌های استیودنتیده را به دست آورید.

ج. نمودار مانده‌های استیودنتیده در برابر x را رسم کنید و در مورد آن اظهارنظر کنید.

د. تبدیل $y^* = \sqrt{y}$ را به‌کاربرید، مدل رگرسیون خطی y^* در برابر x را به دست آورده و نمودار مانده‌های استیودنتیده این مدل را رسم کنید.

ه. مقدار ضریب تعیین دو مدل قسمت «ب» و «د» را به دست آورده و در مورد نتیجه‌های به دست آمده اظهارنظر کنید.

۴. داده‌های جدول ۴.۳ (کارول و اسپیگلمن، ۱۹۸۶) به بررسی ارتباط میان فشار در یک منبع و حجم مایع درون منبع می‌پردازد.

جدول ۴.۳: داده‌های فشار و حجم مایع درون منبع

حجم	فشار	حجم	فشار	حجم	فشار
۲۰۸۴	۴۵۹۹	۲۸۴۲	۶۳۸۰	۳۷۸۹	۸۵۹۹
۲۰۸۴	۴۶۰۰	۳۰۳۰	۶۸۱۸	۳۷۸۹	۸۶۰۰
۲۲۷۳	۵۰۴۴	۳۰۳۱	۶۸۱۷	۳۹۷۹	۹۰۴۸
۲۲۷۳	۵۰۴۳	۳۰۳۱	۶۸۱۸	۳۹۷۹	۹۰۴۸
۲۲۷۳	۵۰۴۴	۳۲۲۱	۷۲۶۶	۴۱۶۷	۹۴۸۴
۲۴۶۳	۵۴۸۸	۳۲۲۱	۷۲۶۸	۴۱۶۸	۹۴۸۷
۲۴۶۳	۵۴۸۷	۳۴۰۹	۷۷۰۹	۴۱۶۸	۹۴۸۷
۲۶۵۱	۵۹۳۱	۳۴۱۰	۷۷۱۰	۴۳۵۸	۹۹۳۶
۲۶۵۲	۵۹۳۲	۳۶۰۰	۸۱۵۶	۴۳۵۸	۹۹۳۸
۲۶۵۲	۵۹۳۲	۳۶۰۰	۸۱۵۸	۴۵۴۶	۱۰۳۷۷
۲۸۴۲	۶۳۸۰	۳۷۸۸	۸۵۹۷	۴۵۴۷	۱۰۳۷۹

الف. مدل رگرسیون خطی را به این داده‌های برازش دهید و مانده‌ها را به دست آورید.

ب. نمودار Q-Q مانده‌ها را رسم کرده و توصیف کنید.

ج. نمودار مانده‌های استیودنتیده در برابر مقدارهای برازش داده شده را رسم و تفسیر کنید.

د. فرض کنید ترتیب زمانی جمع آوری داده‌ها به صورت جدول بالا باشد. نمودار مانده‌ها در برابر زمان را رسم کرده و تفسیر کنید.

۵. برای داده‌های جدول ۴.۳، آزمون نقص برازش مدل رگرسیون خطی را انجام دهید.

۶. داده‌های جدول ۵.۳ (مونت‌گمری، ۲۰۱۳) مدت زمان ارسال یک نوع کالای نوشیدنی (y) و تعداد آن (x) را برای ۲۵ داده نشان می‌دهد.

الف. مدل رگرسیون خطی ساده را به این داده‌ها برازش داده و مانده‌ها را به دست آورید.

ب. نمودار مانده‌های استیودنتیده در برابر x را رسم و تفسیر کنید.

ج. به نظر شما چه تبدیلی برای اصلاح ناهمگنی واریانس خطاهای مشاهده شده در قسمت «ب» مناسب است؟

۷. داده‌های جدول ۱.۹ تمرین ۲۲ فصل ۱ را درنظر بگیرید. هدف تعیین رابطه‌ی بین شدت بیماری و درجه‌ی حرارت بدن بیمار است. با رسم نمودار مانده‌ها در برابر مقدارهای برازش یافته، درباره‌ی درستی فرض‌های ف.۱ تا ف.۵ در این داده‌ها اظهارنظر کنید.

جدول ۵.۳: داده‌های تعداد و مدت زمان ارسال یک کالای نوشیدنی

x	y	x	y	x	y
۶	۱۵/۳۵	۳۰	۷۹/۲۴	۷	۱۶/۶۸
۷	۱۹/۰۰	۵	۲۱/۵۰	۳	۱۱/۵۰
۳	۹/۵۰	۱۶	۴۰/۳۳	۳	۱۲/۳۰
۱۷	۳۵/۱۰	۱۰	۲۱/۰۰	۴	۱۴/۸۸
۱۰	۱۷/۹۰	۴	۱۳/۵۰	۶	۱۳/۷۵
۲۶	۵۲/۳۲	۶	۱۹/۷۵	۷	۱۸/۱۱
۹	۱۸/۷۵	۹	۲۴/۰۰	۲	۸/۰۰
۸	۱۹/۸۳	۱۰	۲۹/۰۰	۷	۱۷/۸۳
				۴	۱۰/۷۵

۸. رابطه‌ی درآمد افراد به میلیون تومان (y) برحسب تحصیلات آن‌ها به سال (x) بر اساس یک نمونه تصادفی به صورت $\widehat{\ln(y)} = ۰/۱ + ۰/۰۴x$ برازش داده شده است. اگر فرد A تحصیلاتی دو سال بیش‌تر از فرد B داشته باشد، نسبت متوسط درآمد فرد A به فرد B را به دست آورید.

۹. داده‌های جدول ۶.۳ میزان ماده‌ی شیمیایی (y) را نشان می‌دهد که در ۱۰۰ گرم آب در درجه‌ی حرارت‌های گوناگون (x) حل می‌شود. آزمون نقص برازش خط رگرسیونی را انجام دهید.

جدول ۶.۳: داده‌های درجه‌ی حرارت و میزان ماده‌ی شیمیایی حل شده در آب

y			x
۶	۵	۴	۰
۱۰	۹	۱۱	۵
۲	۱۳	۱۷	۱۰
۱۸	۲۲	۲۰	۱۵

۱۰. داده‌های جدول۷.۳ (بینگهام و فرای، ۲۰۱۰) زمان‌های ثبت شده به دقیقه (y) و سن به سال (x) را برای نفرات اول دوی ماراتن و نیمه مارتن را نشان می‌دهد. مدل رگرسیون خطی مناسب برای داده‌ها را به‌دست آورید.

۱۱. بردار داده‌های $y = (۹/۸, ۱۱/۰, ۱۳/۲, ۱۵/۱, ۱۶/۰)$ شاخص قیمت یک محصول را در سال‌های $t = ۱, \ldots, ۵$ نشان می‌دهد.

الف. کدام یک از دو مدل $\mathrm{E}(y|t) = a + bt$ و $\mathrm{E}(y|t) = Ae^{Bt}$ بهتر به داده‌ها برازش داده می‌شود؟

جدول ۷.۳: داده‌های سن و زمان ثبت شده برای نفرات اول دوی ماراتون و نیمه ماراتون

ماراتن		نیمه ماراتن	
x	y	x	y
۴۶	۸۵/۶۲	۴۶/۵	۱۶۶/۸۷
۴۸	۸۴/۹۰	۴۷/۰	۱۷۳/۲۵
۴۹	۸۷/۸۸	۴۷/۵	۱۷۵/۱۷
۵۰	۸۷/۸۸	۴۹/۵	۱۷۸/۹۷
۵۱	۸۷/۵۷	۵۰/۵	۱۷۶/۶۳
۵۷	۹۰/۲۵	۵۴/۵	۱۷۵/۰۳
۵۹	۸۸/۴۰	۵۶/۰	۱۸۰/۳۲
۶۰	۸۹/۴۵	۵۸/۵	۱۸۳/۰۲
۶۱	۹۶/۳۸	۵۹/۵	۱۹۲/۳۳
۶۲	۹۴/۶۲	۶۰/۰	۱۹۱/۷۳

ب. آیا مدل درجه‌ی دوم $\mathrm{E}(y|t) = a + bt + ct^2$ سبب بهبود معنی‌داری نسبت به مدل خطی $\mathrm{E}(y|t) = a + bt$ می‌شود؟

۱۲. مدل رگرسیون خطی ساده $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$، $i = 1, \ldots, n$ را در نظر بگیرید، که در آن واریانس ϵ_i برابر $\sigma^2 x_i^2$ است. آیا با تبدیل $y^* = \frac{y}{x}$ و $x^* = \frac{1}{x}$ واریانس خطاها ثابت می‌شود؟

۱۳. داده‌های جدول ۸.۳ متوسط تعداد باکتری‌های باقی‌مانده در کنسروهای غذایی و دقایق حرارت‌دهی در دمای ۳۰۰ درجه فارنهایت را نشان می‌دهد (مونت‌گمری، ۲۰۱۳). نمودار پراکنش مشاهدات را رسم کرده و تبدیل مناسب برای خطی‌سازی مدل برازش داده شده به داده‌ها را تشخیص دهید.

جدول ۸.۳: داده‌های متوسط تعداد باکتری‌های باقی‌مانده در کنسروهای غذایی و دقایق حرارت‌دهی

تعداد باکتری	دقایق حرارت‌دهی	تعداد باکتری	دقایق حرارت‌دهی
۱۷۵	۱	۴۹	۷
۱۰۸	۲	۳۱	۸
۹۵	۳	۲۸	۹
۸۲	۴	۱۷	۱۰
۷۱	۵	۱۶	۱۱
۵۰	۶	۱۱	۱۲

۱۴. داده‌های جدول ۹.۳ شمارش تعداد گل‌های کوکنار را در مزرعه‌های جو در سطح‌های گوناگون یک عامل (تیمار) و بلوک‌های گوناگون نشان می‌دهد (بینگهام و فرای، ۲۰۱۰). متغیر پاسخ y

را برابر تعداد گل‌ها و متغیر پیشگو را برابر متغیر دو سطحی عامل‌ها در نظرگرفته و جدول تحلیل واریانس را تشکیل دهید. سپس تبدیل $y^* = \sqrt{y}$ را به‌کاربرده و دوباره تحلیل واریانس را انجام داده و نتیجه‌های به دست آمده از دو مدل را با هم مقایسه کنید.

جدول ۹.۳: داده‌های تعداد گل‌های کوکنار در مزرعه‌های جو

تیمار	A	B	C	D	E
بلوک ۱	۴۳۸	۵۳۸	۷۷	۱۷	۱۸
بلوک ۲	۴۴۲	۴۲۲	۶۱	۳۱	۲۶
بلوک ۳	۳۱۹	۳۷۷	۱۵۷	۸۷	۷۷
بلوک ۴	۳۸۰	۳۱۵	۵۲	۱۶	۲۰

۱۵. داده‌های جدول ۱۰.۳ نسبت خوشه‌های ذرت بدون استفاده را در سطح‌های گوناگون یک عامل (تیمار) و بلوک‌های گوناگون نشان می‌دهد (بینگهام و فرای، ۲۰۱۰). متغیر پاسخ y را برابر نسبت خوشه‌های ذرت بدون استفاده و متغیر پیشگو را برابر متغیر دو سطحی عامل‌ها در نظرگرفته و جدول تحلیل واریانس را تشکیل دهید. سپس تبدیل $y^* = \arcsin(\sqrt{y})$ را به‌کاربرده و دوباره تحلیل واریانس را انجام داده و نتیجه‌های به دست آمده از دو مدل را با هم مقایسه کنید.

جدول ۱۰.۳: داده‌های نسبت خوشه‌های ذرت بدون استفاده

تیمار	A	B	C	D
بلوک ۱	۰/۴۲۴	۰/۳۳۳	۰/۰۸۵	۰/۱۶۶
بلوک ۲	۰/۳۴۴	۰/۳۳۳	۰/۲۱۹	۰/۱۹۳
بلوک ۳	۰/۲۴۱	۰/۰۵۰	۰/۰۶۲	۰/۱۶۶
بلوک ۴	۰/۳۹۵	۰/۲۶۳	۰/۱۶۰	۰/۰۲۱
بلوک ۵	۰/۵۵۵	۰/۳۰۲	۰/۱۳۵	۰/۱۱۱
بلوک ۶	۰/۴۹۱	۰/۲۸۶	۰/۱۵۴	۰/۱۱۱