

## ۱۰.۲ تمرین‌ها

۱. برای داده‌های table.b1 در تمرین ۲ فصل ۱، استنباط‌های زیر را انجام دهید (با فرض نرمال بودن)

- الف. آزمون معنی‌داری خط رگرسیون و بازه‌ی اطمینان ۹۵ درصد برای شیب خط رگرسیون؛  
 ب. آزمون بدون عرض از مبدأ بودن مدل و بازه‌ی اطمینان ۹۵ درصد برای عرض از مبدأ خط رگرسیون؛  
 ج. بازه‌ی اطمینان ۹۵ درصد برای خط رگرسیون.

۲. برای داده‌های وزن و فشار خون تمرین ۳ فصل ۱

- الف. بازه‌ی اطمینان ۹۵ درصد مجانبی برای ضریب همبستگی  $\rho$  به دست آورید.  
 ب. آیا در نظر گرفتن یک مدل رگرسیون خطی بدون عرض از مبدأ برای این داده‌ها منطقی است؟ چرا؟ آزمون فرض مربوط را انجام دهید.  
 ج. یک مدل رگرسیون خطی بدون عرض از مبدأ به داده‌ها برازش دهید و برای پارامتر شیب خط رگرسیونی در این مدل یک بازه‌ی اطمینان ۹۵ درصد به دست آورید.  
 د. یک بازه‌ی پیش‌بینی برای فشار خون فردی با وزن ۲۰۰ پوند را به دست آورید.  
 ۳. فرض کنید از یک جامعه، نمونه‌ای به اندازه  $n = ۳۸$  انتخاب کرده‌ایم و از هر فرد دو ویژگی  $(x, y)$  مشاهده شده است. فرض کنید  $(x, y)$  دارای توزیع نرمال دومتغیره باشد. پس از مشاهده‌ی نمونه، نتیجه‌های زیر به دست آمده‌اند

$$S_{xx} = ۱۰۰, S_{xy} = ۱۴۰, S_{yy} = ۲۹۶,$$

- آزمون معنی‌داری شیب خط رگرسیونی را در سطح  $\alpha = ۰/۰۵$  انجام دهید.  
 ۴. برای داده‌های مثال‌های ۱ و ۲ و ۳ فصل ۱، آزمون فرض صفر بودن ضرایب‌های رگرسیونی را در سطح ۵ درصد انجام دهید.  
 ۵. با به کارگیری داده‌های مثال‌های ۴ و ۵ و ۶ فصل ۱، برای ضرایب‌های رگرسیونی بازه‌ی اطمینان ۹۵ درصد به دست آورید.

## ۶. برای داده‌های تمرین ۲۱ فصل ۱

الف. آزمون فرض صفر بودن شیب خط رگرسیونی را در سطح ۵ درصد انجام دهید.

ب. بازه‌ی اطمینان ۹۵ درصد برای عرض از مبدأ خط رگرسیونی را به دست آورید.

ج. از پاسخ قسمت «ب» درباره‌ی صفر بودن یا نبودن عرض از مبدأ خط رگرسیونی اظهارنظر کنید. اگر پاسخ مثبت بود، یک مدل بدون عرض از مبدأ به داده‌ها برازش داده و آزمون قسمت «الف» را برای مدل جدید دوباره انجام دهید.

د. برای خط رگرسیونی برازش داده شده یک بازه‌ی اطمینان ۹۵ درصد به دست آورید.

۷. مدل رگرسیون خطی بدون عرض از مبدأ  $y_i = \beta_1 x_i + \epsilon_i$ ،  $i = 1, 2, \dots, n$  را در نظر بگیرید که در آن  $\epsilon_i$ ها متغیرهای تصادفی مستقل و دارای توزیع نرمال با واریانس  $\sigma^2$  هستند. فرض کنید  $x_0$  یک مقدار اختیاری از متغیر پیشگوی  $x$  باشد. نشان دهید یک بازه‌ی اطمینان  $100(1 - \alpha)\%$  برای  $E(y|x_0)$  از رابطه‌ی زیر به دست می‌آید

$$\hat{\beta}_1 x_0 \pm t_{(n-1; \frac{\alpha}{2})} S_e \sqrt{\frac{x_0^2}{\sum_{i=1}^n x_i^2}}$$

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

۸. داده‌های جدول ۲.۲ مربوط به ۲۷ درخت کاج است که ضریب گسیختگی ( $y$ ) با واحد کیلوپاسکال (KPa) و وزن مخصوص آن ( $x$ ) با واحد گرم بر سانتی متر مکعب ( $\text{gr/cm}^3$ ) اندازه‌گیری شده‌اند (مایرز، ۱۹۹۰).

الف. خط رگرسیونی کم‌ترین توان دوم را برازش دهید.

ب. برآورد واریانس جمله‌ی خطای مدل را به دست آورید و به کمک آن برآورد واریانس پارامترهای مدل رگرسیون خطی را محاسبه کنید.

ج. آزمون معنی‌داری مدل رگرسیون خطی را به دو روش انجام دهید.

د. جدول تحلیل واریانس را تشکیل داده و به کمک آن مشخص کنید آیا مدل رگرسیون خطی معنی‌دار است یا خیر؟

ه. بازه‌ی اطمینان و بازه‌ی پیش‌بینی را به ازای مقدار معین  $x_0 = 0.58$  برای متغیر پیشگو به دست آورید.

جدول ۲.۲: داده‌های درختان کاج

$y$	$x$	$y$	$x$	$y$	$x$
۲۹۱۸۶	۰/۴۱۴	۶۷۷۰۵	۰/۵۳۰	۷۳۴۶۶	۰/۵۳۱
۲۹۲۶۶	۰/۳۸۳	۶۶۰۸۸	۰/۵۶۹	۷۸۶۱۰	۰/۵۵۰
۲۶۲۱۵	۰/۳۹۹	۷۸۴۸۶	۰/۵۵۸	۶۷۶۵۷	۰/۵۵۶
۳۰۱۶۲	۰/۴۰۲	۸۹۸۶۹	۰/۵۷۷	۷۴۰۱۷	۰/۵۲۳
۳۸۸۶۷	۰/۴۴۲	۷۷۳۶۹	۰/۵۷۲	۸۷۲۹۱	۰/۶۰۲
۳۷۸۳۱	۰/۴۲۲	۶۷۰۹۵	۰/۵۴۸	۸۶۸۳۶	۰/۵۶۹
۴۴۵۷۶	۰/۴۶۶	۸۵۱۵۶	۰/۵۸۱	۸۲۵۴۰	۰/۵۴۴
۴۶۰۹۷	۰/۵۰۰	۶۹۵۷۱	۰/۵۵۷	۸۱۶۹۹	۰/۵۵۷
۵۹۶۹۸	۰/۵۱۴	۸۴۱۶۰	۰/۵۵۰	۸۲۰۹۶	۰/۵۳۰

جدول ۳.۲: داده‌های مربوط به فعالیت شدید بدنی ورزشکاران

$y$	$x$	$y$	$x$	$y$	$x$
۴۲/۳۳	۹۱۸	۳۶/۲۳	۱۰۴۵	۵۳/۲۹	۷۴۳
۵۳/۱۰	۸۰۵	۴۹/۶۶	۸۱۰	۴۷/۱۸	۸۰۳
۴۲/۰۸	۸۹۲	۴۱/۴۹	۹۲۷	۵۶/۹۱	۶۸۳
۵۰/۰۶	۹۶۲	۴۶/۱۷	۸۱۳	۴۷/۸۰	۸۴۴
۴۲/۴۵	۹۶۸	۴۸/۱۸	۸۵۸	۴۸/۶۵	۷۵۵
۴۲/۴۶	۹۰۷	۴۳/۲۱	۸۶۰	۵۳/۶۹	۷۰۰
۴۷/۸۲	۷۷۰	۵۱/۸۱	۷۶۰	۶۰/۶۲	۷۴۸
۴۹/۹۲	۷۴۳	۵۳/۲۸	۷۴۷	۵۶/۷۳	۷۷۵

و. موارد «الف» تا «ه» را با به‌کارگیری نرم‌افزار R انجام دهید.

۹. یکی از آزمون‌هایی که به‌طور معمول بر روی ورزشکاران حرفه‌ای انجام می‌شود، اندازه‌گیری میزان اکسیژن دریافتی در مدت معینی از فعالیت شدید بدنی است. برای این منظور از تعداد ۲۴ ورزشکار حرفه‌ای خواسته شده که مسافت ۲ مایل را بر روی تردمیل دویده و در این حالت دو متغیر مدت زمان طی مسافت بر حسب ثانیه ( $x$ ) و میزان اکسیژن دریافتی ( $y$ ) اندازه‌گیری شده که نتایج آن در جدول ۳.۲ آمده است (مایرز، ۱۹۹۰).

الف. پارامترهای مدل رگرسیون خطی ساده را برآورد کنید.

ب. آزمون کنید آیا مدل رگرسیون خطی معنی‌دار است یا خیر؟

ج. مشخص کنید به ازای هر ده ثانیه دویدن بیشتر، به طور متوسط چه میزان اکسیژن بیش‌تر

جدول ۴.۲: داده‌های نمره‌های دانشجویان

۸۲	۵۱	۸۹	۸۴	۶۶	۹۵	۷۳	۸۲	$x$
۸۹	۶۲	۷۳	۷۹	۷۶	۸۹	۸۳	۷۶	$y$
۸۷	۴۹	۳۴	۸۱	۶۰	۹۰	۹۰	۷۵	$x$
۷۴	۲۵	۵۱	۶۹	۴۸	۸۵	۸۵	۷۷	$y$

مصرف می‌شود؟

د. مقدارهای برازش یافته‌ی متغیر پاسخ و مانده‌ها را به دست آورید.

۱۰. داده‌های جدول ۴.۲ نمره‌های میان‌ترم درس آمار ( $x$ ) و نمره‌ی نهایی ( $y$ ) دانشجویان (از ۱۰۰ نمره) را نشان می‌دهد.

الف. یک مدل رگرسیون خطی ساده به داده‌ها برازش دهید که بر اساس آن بتوان نمره‌ی نهایی دانشجویان را از روی نمره‌ی میان‌ترم آن‌ها پیش‌بینی کرد.

ب. معنی‌داری مدل رگرسیون خطی ساده را به کمک تحلیل واریانس بررسی کنید.

ج. بازه‌ی اطمینان ۹۰ درصد را برای ضریب‌های رگرسیون خطی به دست آورید.

د. موارد بالا را با به‌کارگیری نرم‌افزار R انجام دهید.

۱۱. خلاصه‌ی اطلاعات زیر بر اساس ۲۰ داده به دست آمده است

$$\bar{x} = 15/53, \bar{y} = 61/85, S_{xx} = 64/51, S_{yy} = 3744/36, S_{xy} = 446/17$$

با توجه به اطلاعات داده شده به پرسش‌های زیر پاسخ دهید.

الف. معادله‌ی خط رگرسیونی را بنویسید.

ب. مقدار  $S_e^2$  را به دست آورید.

ج. آماره‌ی آزمون  $H_0: \beta_1 = 0$  در برابر  $H_1: \beta_1 \neq 0$  را به دست آورید.

د. ضریب همبستگی خطی نمونه‌ای پیرسون را به دست آورده و آماره‌ی آزمون  $H_0: \rho = 0$  را در برابر  $H_1: \rho \neq 0$  محاسبه کنید.

ه. یک بازه‌ی اطمینان ۹۹ درصد برای  $\beta$  به دست آورید.



و. جدول تحلیل واریانس مربوط به مدل بالا را تشکیل دهید.

۱۲. فرض کنید  $(x, y) \sim N(2, 2, 3, 1, 0/7)$  باشد، ضریب‌های معادله‌ی رگرسیون خطی

$$y = \beta_0 + \beta_1 x + \epsilon$$

و واریانس جمله‌ی خطا را به دست آورید.

۱۳. در برازش یک مدل رگرسیون خطی مقدارهای برازش یافته‌ی جدول زیر به دست آمده است.

$y_i$	۲	۲/۵	۳/۲	۴/۱	۵
$\hat{y}_i$	۱/۵	۲	۳	۵	۴/۵

الف. برآورد واریانس مدل را به دست آورید.

ب. جدول تحلیل واریانس را تشکیل داده و تکمیل کنید.

۱۴. با توجه به نامنفی بودن چندجمله‌ای

$$Q(\lambda) = \frac{1}{n} \sum_{i=1}^n (\lambda(x_i - \bar{x}) + (y_i - \bar{y}))^2,$$

ثابت کنید ضریب همبستگی نمونه‌ای  $(r)$  برای داده‌های  $(x_1, y_1), \dots, (x_n, y_n)$ ، بین ۱- و ۱ است. همچنین  $|r| = 1$ ، اگر برای عددهای حقیقی  $a, b, c$  داشته باشیم  $ax_i + by_i = c$ ،  $i = 1, \dots, n$ .