

۱-۱: داده‌های زیر از یک آزمایش به دست آمده‌اند. (در Excel فایل 4-SSC1)

Y	X_1	X_2	X_3
۱۱/۹۷	۴۷/۱	۳۳/۹	۷/۵
۸/۶۳	۷۲/۹	۳۳/۹	۷۵۰
۸/۸۰	۴۷/۱	۸/۱	۷۵۰
۱۰/۷۳	۶۰	۲۱	۷۵
۱۰/۶۹	۶۰	۲۱	۷۵
۱۳/۱۲	۷۲/۹۰	۸/۱	۷/۵
۱۲/۵۸	۴۷/۱۰	۸/۱	۷/۵
۱۲/۲۴	۷۲/۹۰	۳۳/۹	۷/۵
۱۰/۶۴	۶۰	۲۱	۷۵
۹/۰۹	۷۲/۹۰	۸/۱	۷۵۰
۸/۴۶	۴۷/۱۰	۳۳/۹	۷۵۰
۱۰/۶۵	۶۰	۲۱	۷۵
۷/۶۰	۶۰	۲۱	۳۰۰۰
۱۳/۰۶	۶۰	۲۱	۳
۱۰/۵۱	۳۹	۲۱	۷۵
۱۱/۲۲	۶۰	۰	۷۵
۱۰/۶۷	۶۰	۲۱	۷۵
۱۰/۲۴	۶۰	۴۲	۷۵
۱۰/۷۴	۸۱	۲۱	۷۵
۱۰/۶۹	۶۰	۲۱	۷۵

الف- برآوردهای کمترین توانهای دوم و خطای استاندارد پارامترهای مدل زیر را به دست آورید:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 Z_i + \varepsilon_i$$

که در آن $Z_i = \ln(X_{ri} + 1)$.

$b_0 = ..$	$b_1 =$
$b_r =$	$b_r =$
$S_{b_0} =$	$S_{b_1} =$
$S_{b_2} =$	$S_{b_3} =$

برای مدل بالا کدام یک از متغیرهای پیشگو به نظر میرسند مهم نباشند؟ توضیح دهید.
 ب- برای مدل قسمت (الف) مجموع توانهای دوم مانده ($SS_{residual}$) را به دو بخش افراز کرده و آماره F را برای آزمون نقص برازش (Lack of fit test) محاسبه کنید.

منبع تغییرات	d.f.	S.S.	M.S.	F
<hr/>				
نقص برازش				
خطای خالص (Pure error)				
<hr/>				
مانده (Residuals)				

آیا مدل قسمت (الف) مناسب به نظر میرسد؟ توضیح دهید

پ- نمودار مانده‌های استودنتیده (Studentized Residuals) حاصل از مدل قسمت (الف) را در مقابل X_1 ، X_2 ، \hat{Y}_i و ترتیبی که مشاهدات گزارش شده‌اند، رسم کنید. همچنین نمودار احتمال نرمال را برای مانده‌ها رسم کرده و آماره شاپیرو-ویلک را محاسبه کنید. با این بررسی‌ها در باره فرض‌های مدل رگرسیونی چه می‌توان گفت؟

پ- نمودار مانده‌های استودنتیده (Studentized Residuals) حاصل از مدل قسمت (الف) را در مقابل X_1 ، X_2 ، \hat{Y}_i و ترتیبی که مشاهدات گزارش شده‌اند، رسم کنید. همچنین نمودار احتمال نرمال را برای مانده‌ها رسم کرده و آماره شاپیرو-ویلک را محاسبه کنید. با این بررسی‌ها در باره فرض‌های مدل رگرسیونی چه می‌توان گفت؟

ت- برآوردهای کمترین توانهای دوم پارامترهای مدل زیر و خطای استاندارد آنها را به دست آورید.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 Z_i + \beta_4 X_{1i} X_{2i} + \beta_5 X_{1i} Z_i + \beta_6 X_{2i} Z_i + \beta_7 Z_i^2 + \beta_8 X_{1i} Z_i^2 + \beta_9 X_{2i} Z_i^2 + \varepsilon_i$$

که در آن $Z_i = \ln(X_{2i} + 1)$.

کدام یک از جمله‌های این مدل مهم به نظر نمی‌رسند. (چرا؟)

ح- قسمت (ب) را برای مدل قسمت (ت) تکرار کرده و گزارش کنید.

$$F_{lack-of-fit} = \dots \quad d.f. = \dots \quad P-value = \dots$$

استنباط خود را بیان کنید.

خ- مقادیر R^2 برای مدل‌های قسمت‌های (الف) و (ت) را با بزرگترین مقدار ممکن R^2 مقایسه کنید

$$R^2_{(الف)} = \dots \quad R^2_{(ت)} = \dots \quad R^2_{max} = \dots$$

ج- فرض صفر اینکه مدل قسمت (الف) مناسب است را در مقابل فرض اینکه مدل قسمت (ت) مناسب است، آزمون کنید.

$$F = \dots \quad d.f. = \dots \quad P-value = \dots$$

با توجه به نتایج قسمت‌های (خ) و (ج) به چه نتیجه‌ای می‌رسید؟

چ-قسمت(پ) را برای مدل قسمت (ت) انجام داده، نتیجه گیری نمائید؟

د-پارامتر β_1 را در مدل قسمت (ت) تفسیر کنید

۱-۲: یک پژوهشگر دو نمونه تصادفی، هر یک به حجم ۳۰، یکی از بین مردان سیگاری و دیگری از بین مردان غیرسیگاری انتخاب کرده، متغیرهای زیر را در مورد هر ۶۰ نفر اندازه گیری نمود.

۱-کلسترول H HDL

۲-کلسترول L LDL

۳-فشار خون بالا SYS

۴-فشار خون پایین DIS

۵-چربی FAT

۶-وزن WEI

این پژوهشگر سپس از هر یک از دو گروه فوق، ۱۵ نفر را به طور تصادفی انتخاب کرده، هر ۳۰ نفر را در طی سه ماه بعد به رعایت یک برنامه ورزشی وادار نموده و در پایان این دوره مقادیر متغیرهای فوق را، در مورد هر ۶۰ نفر، مجدداً اندازه گیری نمود. داده های این دو نوبت اندازه گیری در Excel بنام فایل 4-SSC2 داده شده است.

توجه:

۱-برای انجام موارد زیر از هر نرم افزار آماری می توانید استفاده کنید.

۲-کلیه خروجی نرم افزار (پس از ویرایش) باید به ممتحنین تحویل داده شود.

۳-در مورد هر متغیر پسوند ۱ نشان دهنده اندازه گیری در ابتدای دوره سه ماهه و

۴-در مورد هر متغیر پسوند ۲ نشان دهنده اندازه گیری در انتهای دوره سه ماهه است.

۵-در ستون cig، "yes" به معنی سیگاری بودن است.

۶-مرز معنی داری p-مقدار=۰/۰۵ و مرز p-مقدار برای خروج متغیر = ۰/۱۰.

۷-در مورد هر بخش کلیه فرض هائی را که لازم است قائل شوید بنویسید

الف) رابطه فشار خون بالا (SYS1) را با سایر متغیرهای کمی فوق در بین کلیه افراد براساس داده های قبل از ورزش به روش رگرسیون پسر و بررسی کرده، ضرایب معادله، ضریب تعیین و ضرایب همبستگی جزئی الگوی نهایی را برآورد نموده و نتایج را تفسیر نمایید. پس از رسم نمودار Q-Q برای باقیمانده های الگوی نهایی، نرمال بودن باقیمانده ها را آزمون کنید.

ب) بررسی اثر سیگار (cig) و ورزش (exe) را بر متغیر وابسته وزن (wei) با استفاده از یک الگوی دو عاملی با فرض نرمال بودن متغیر وابسته (و هر فرض دیگری که لازم است قائل شوید ذکر کنید).

ب-۱) وجود اثر متقابل را آزمون و نتیجه را تفسیر کنید.

ب-۲) با توجه به وجود یا نبود اثر متقابل معنی داری هر یک از دو عامل را بر متغیر وابسته معنی کنید.

پ) تکرار (ب) با متغیر وابسته "کلسترول H" بجای "wei"

ت) در صورتیکه نتوانیم توزیع داده ها را نرمال فرض کنیم، این فرضیه را که "در مورد افراد سیگاری ورزش بر کلسترول H بی تأثیر است" آزمون نمایید.



group	cig	hdl1	ldl1	dis1	sis1	wei1	fat1	hdl2	ldl2	dis2	sis2	wei2	fat2	
1	t	yes	37	149	80	110	75	20	49	127	80	110	70	18
2	t	yes	36	142	80	120	72	23	51	119	80	120	71	22
3	t	yes	46	149	100	130	76	24	50	148	90	130	72	23
4	t	yes	44	168	80	120	85	25	48	118	80	110	78	22
5	t	yes	40	214	80	130	72	21	43	192	80	120	72	21
6	t	yes	38	124	90	130	81	25	40	112	80	120	76	23
7	t	yes	37	246	80	120	70	21	38	217	80	120	69	21
8	t	yes	38	116	80	120	75	24	38	110	80	120	71	22
9	t	yes	42	84	60	90	68	22	49	83	60	90	67	21
10	t	yes	41	142	80	130	75	24	47	127	80	120	71	22
11	t	yes	37	156	90	130	74	22	39	120	80	120	72	21
12	t	yes	36	146	80	120	84	26	38	131	80	120	79	23
13	t	yes	40	142	90	130	80	24	55	137	80	130	75	22
14	t	yes	36	121	80	120	72	21	40	112	80	120	72	21
15	t	yes	37	200	90	130	75	22	39	164	80	120	71	21
16	t	no	37	136	90	140	70	27	34	128	85	120	70	27
17	t	no	30	125	90	150	76	28	30	126	80	130	74	26
18	t	no	26	100	90	120	90	28	30	99	80	120	84	26
19	t	no	27	75	80	130	81	29	29	76	80	130	75	28
20	t	no	34	150	90	130	82	24	31	141	85	120	82	24

group	cig	hdl1	ldl1	dis1	sis1	wei1	fat1	hdl2	ldl2	dis2	sis2	wei2	fat2
21 t	no	30	59	90	130	80	24	30	58	85	120	79	24
22 t	no	31	116	80	120	70	25	32	118	80	110	71	25
23 t	no	29	100	80	120	75	24	30	107	80	120	73	23
24 t	no	33	156	70	120	73	24	34	151	70	120	72	23
25 t	no	30	132	80	120	87	28	30	132	80	120	81	26
26 t	no	34	75	90	140	75	22	33	76	80	130	74	22
27 t	no	27	127	80	120	82	26	31	136	80	120	78	23
28 t	no	28	105	90	130	72	21	29	111	85	120	72	21
29 t	no	35	131	80	120	75	22	35	142	80	110	70	20
30 t	no	30	127	80	120	74	23	34	128	80	120	73	22
31 c	yes	39	173	90	120	80	24	39	176	90	120	81	24
32 c	yes	49	106	90	130	64	21	44	109	90	130	64	21
33 c	yes	40	112	80	120	67	25	36	122	80	120	67	25
34 c	yes	44	168	80	120	72	23	41	170	80	120	72	23
35 c	yes	41	160	80	120	70	25	42	150	80	120	74	25
36 c	yes	45	234	80	120	69	21	46	228	90	130	70	22
37 c	yes	37	116	80	120	75	24	38	126	80	120	70	23
38 c	yes	35	142	70	110	67	22	35	140	80	120	68	22
39 c	yes	43	79	90	130	54	18	42	82	80	130	54	18
40 c	yes	35	119	90	130	81	24	40	112	80	130	81	24

group	cig	hdl1	ldl1	dis1	sis1	wei1	fat1	hdl2	ldl2	dis2	sis2	wei2	fat2	
41	c	yes	37	174	80	130	72	22	37	180	80	130	72	22
42	c	yes	37	65	90	140	74	22	39	60	80	130	74	22
43	c	yes	36	144	80	120	85	28	40	140	80	120	85	28
44	c	yes	37	205	90	130	74	22	38	186	90	120	71	21
45	c	yes	35	142	80	120	73	21	36	160	80	120	75	22
46	c	no	31	128	90	160	74	26	31	127	90	160	73	26
47	c	no	31	120	80	130	72	27	32	122	80	130	72	27
48	c	no	32	127	80	120	86	26	30	128	80	130	88	26
49	c	no	28	92	82	140	75	26	29	102	85	140	75	26
50	c	no	29	100	80	120	80	25	28	107	80	120	78	25
51	c	no	31	87	90	120	76	25	30	82	85	120	75	25
52	c	no	30	118	90	120	73	26	31	109	80	120	74	26
53	c	no	32	102	80	110	75	24	32	116	80	120	75	24
54	c	no	30	149	85	140	81	24	31	147	90	140	80	24
55	c	no	35	125	85	120	77	25	34	126	70	110	78	25
56	c	no	33	71	80	110	72	23	35	85	80	110	72	23
57	c	no	31	116	70	110	70	23	29	112	80	110	71	23
58	c	no	28	135	85	150	82	22	28	142	85	150	81	23
59	c	no	36	147	90	130	78	26	37	134	85	120	76	27
60	c	no	32	120	80	110	76	24	30	119	90	110	77	23