MATH 204 - ASSIGNMENT 2

Please Hand in Assignment in the Lecture on Wednesday 7th March.

The study of lung expiratory pressure capacity in sufferers from cystic fibrosis is to be studied, with the objective of diagnosing whether any other measured variables significantly influence the variability in this response variable.

Reference: O'Neill S, Leahy F, Pasterkamp H, Tal A. The effects of chronic hyperinflation, nutritional status, and posture on respiratory muscle strength in cystic fibrosis. *American Review of Respiratory Disease*, 1983 Dec;128(6):10511054.

Experiment: 25 sufferers from cystic fibrosis were measured for their **maximal static expiratory pressure**; this is the continuous response variable, denoted **pemax**.

Nine other variables were also recorded. Eight are *continuous covariates*

- **age**: Age in years.
- **height**: Height (cm).
- **weight**: Weight (kg).
- **bmp**: Body mass (% of normal).
- **fev1**: Forced expiratory volume.
- **rv**: Residual volume.
- **frc**: Functional residual capacity.
- **tlc**: Total lung capacity.

The ninth is a *factor predictor*

- **sex**: Sex of patient - 0=Male, 1=Female.

Research question: Which of the variables are useful in explaining the variation in response variable **pemax** ?

(a) For each of the nine independent variables, fit a *simple linear regression model*, and identify those variables which appear to contribute significantly to the variation in the response.

Note that the factor predictor **must** not be treated in the same way as the continuous covariates when the model is fitted using the software.

10 Marks

(b) Using *multiple regression* fit a linear regression model of the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_9 x_9 + \epsilon$$

and identify which independent variables appear to influence the response significantly. Comment on the global fit of the model, and contrast the results with those from (a).

5 Marks

(c) Using the

$$\texttt{Analyze} \longrightarrow \texttt{Correlate} \longrightarrow \texttt{Bivariate}$$

pulldown menu selections, compute the correlations between the continuous covariates.

3 Marks

(d) The results in (c) help to explain the results in (a) and (b). Explain why.

2 Marks

CYSTIC FIBROSIS DATA SET

id	age	height	weight	bmp	fev1	rv	frc	tlc	sex	pemax
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y
1	7	109	13.1	68	32	258	183	137	0	95
2	7	112	12.9	65	19	449	245	134	1	85
3	8	124	14.1	64	22	441	268	147	0	100
4	8	125	16.2	67	41	234	146	124	1	85
5	8	127	21.5	93	52	202	131	104	0	95
6	9	130	17.5	68	44	308	155	118	0	80
7	11	139	30.7	89	28	305	179	119	1	65
8	12	150	28.4	69	18	369	198	103	1	110
9	12	146	25.1	67	24	312	194	128	0	70
10	13	155	31.5	68	23	413	225	136	1	95
11	13	156	39.9	89	39	206	142	95	0	110
12	14	153	42.1	90	26	253	191	121	1	90
13	14	160	45.6	93	45	174	139	108	0	100
14	15	158	51.2	93	45	158	124	90	1	80
15	16	160	35.9	66	31	302	133	101	1	134
16	17	153	34.8	70	29	204	118	120	1	134
17	17	174	44.7	70	49	187	104	103	0	165
18	17	176	60.1	92	29	188	129	130	1	120
19	17	171	42.6	69	38	172	130	103	0	130
20	19	156	37.2	72	21	216	119	81	1	85
21	19	174	54.6	86	37	184	118	101	0	85
22	20	178	64.0	86	34	225	148	135	0	160
23	23	180	73.8	97	57	171	108	98	0	165
24	23	175	51.1	71	33	224	131	113	0	95
25	23	179	71.5	95	52	225	127	101	0	195

The data may be downloaded

- in plain text format from

  http://www.math.mcgill.ca/∼dstephens/204/Data/CystFibr.txt

- in SPSS format from

  http://www.math.mcgill.ca/∼dstephens/204/Data/CystFibr.sav

You may use SPSS, and the output generated, provided that you write comments pointing out the key results.