Assignment 8

Bikram Halder

2023-04-05

Problem 1

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	Y
1	17.75	60.2	5.83	69.0	1.49	77.9	2.42	74.4	34.0
2	14.76	57.5	3.83	75.0	2.72	77.2	3.30	72.6	32.0
3	27.99	62.3	5.17	72.0	3.12	75.8	7.10	72.2	43.0
4	16.76	60.5	1.64	77.8	3.45	76.1	3.01	70.5	40.0
5	11.36	69.5	3.49	77.2	3.85	79.7	2.84	73.4	23.0
6	22.71	55.0	7.00	65.9	3.35	79.4	2.42	73.6	38.4
7	17.91	66.2	2.85	70.1	0.51	83.4	3.48	79.2	20.0
8	23.31	61.8	3.80	69.0	2.63	75.9	3.99	77.8	44.6
9	18.53	59.5	4.67	69.2	4.24	76.5	3.82	75.7	46.3
10	18.56	66.4	5.32	71.4	3.15	76.2	4.72	70.7	52.2
11	12.45	58.4	3.56	71.3	4.57	76.7	6.44	70.7	52.3
12	16.05	66.0	6.20	70.0	2.24	75.1	1.94	75.1	51.0
13	27.10	59.3	5.93	69.7	4.89	74.3	3.17	72.2	59.9
14	19.05	57.5	6.16	71.6	4.56	75.4	5.07	74.0	54.7
15	20.79	64.6	5.88	71.7	3.73	72.6	5.88	71.8	52.0
16	21.88	55.1	4.70	64.1	2.96	72.1	3.43	72.5	43.5
17	20.02	56.5	6.41	69.8	2.45	73.8	3.56	68.9	56.7
18	23.17	55.6	10.39	66.3	1.72	72.8	1.49	80.6	30.5
19	19.15	59.2	3.42	68.6	4.14	75.0	2.54	73.9	60.5
20	18.28	63.5	5.51	72.4	3.47	76.2	2.34	73.0	46.1
21	18.45	59.8	5.70	68.4	4.65	69.7	2.39	67.7	48.2
22	22.00	62.2	6.11	65.2	4.45	72.1	6.21	70.5	43.1
23	19.05	59.6	5.40	74.2	3.84	74.7	4.78	70.0	62.2
24	15.67	60.0	5.31	73.2	3.28	74.6	2.33	73.2	52.9
25	15.92	55.6	6.36	72.9	1.79	77.4	7.10	72.1	53.9
26	16.75	63.6	3.07	67.2	3.29	79.8	1.79	77.2	48.4
27	12.34	62.4	2.56	74.7	4.51	72.7	4.42	73.0	52.8
28	15.82	59.0	4.84	68.9	3.54	77.9	3.76	72.9	62.1
29	15.24	62.5	3.80	66.4	7.55	70.5	2.55	73.0	66.0
30	21.72	62.8	4.11	71.5	2.29	72.3	4.92	76.3	64.2
31	25.08	59.7	4.43	67.4	2.76	72.6	5.36	73.2	63.2
32	17.79	57.4	3.36	69.4	5.51	72.6	3.04	72.4	75.4
33	26.61	66.6	3.12	69.1	6.27	71.6	4.31	72.5	76.0

The above data set provides information on the variables X_1 = Year, 1 denoting 1930, X_2 = Pre-season precipitation, X_3 = May temperature, X_4 = June rain, X_5 = June temperature, X_6 = July rain, X_7 = July temperature, X_8 = August rain, X_9 = August temperature, and Y = X_{10} = Corn Yield, for the state of Iowa.

Using the forward selection method of stepwise regression, construct a multiple regression model for Corn Yield based on the predictors available suchthat the model has at most 4 predictors and these predictors explain at least 70% of the variability in Corn Yield.

Solution

We have

$$
\begin{aligned}
\operatorname{Corr}(X_1, Y) &= 0.751769616309195 \\
\operatorname{Corr}(X_2, Y) &= 0.192254990500002 \\
\operatorname{Corr}(X_3, Y) &= -0.101795078871127 \\
\operatorname{Corr}(X_4, Y) &= -0.143852643399597 \\
\operatorname{Corr}(X_5, Y) &= -0.14887625581978 \\
\operatorname{Corr}(X_6, Y) &= 0.581004126431327 \\
\operatorname{Corr}(X_7, Y) &= -0.579630217192443 \\
\operatorname{Corr}(X_8, Y) &= 0.209020782245619 \\
\operatorname{Corr}(X_9, Y) &= -0.342884971142907
\end{aligned}
$$

Among which the absolute maximum is $\operatorname{Corr}(X_1, Y) = 0.7517696$. But R^2-score for the model $Y \sim X_1$ is $0.5651576 < 0.70$.

So we find

$$
\begin{aligned}
\operatorname{Corr}(X_2, Y|X_1) &= 0.241837860792189 \\
\operatorname{Corr}(X_3, Y|X_1) &= -0.108903917548665 \\
\operatorname{Corr}(X_4, Y|X_1) &= -0.13792989690489 \\
\operatorname{Corr}(X_5, Y|X_1) &= 0.078783791740625 \\
\operatorname{Corr}(X_6, Y|X_1) &= 0.461627365772735 \\
\operatorname{Corr}(X_7, Y|X_1) &= -0.307111438276225 \\
\operatorname{Corr}(X_8, Y|X_1) &= 0.248867305219476 \\
\operatorname{Corr}(X_9, Y|X_1) &= -0.426500342231511
\end{aligned}
$$

Among which the absolute maximum is $\operatorname{Corr}(X_6, Y|X_1) = 0.4616274$. But R^2-score for the model $Y \sim X_1 + X_6$ is $0.6578224 < 0.70$.

So we find

$$
\begin{aligned}
\operatorname{Corr}(X_2, Y|X_1, X_6) &= 0.300038611141932 \\
\operatorname{Corr}(X_3, Y|X_1, X_6) &= -0.179398775787862 \\
\operatorname{Corr}(X_4, Y|X_1, X_6) &= -0.0128078684982457 \\
\operatorname{Corr}(X_5, Y|X_1, X_6) &= 0.0759728637870637 \\
\operatorname{Corr}(X_7, Y|X_1, X_6) &= -0.158816380491146 \\
\operatorname{Corr}(X_8, Y|X_1, X_6) &= 0.284062595649432 \\
\operatorname{Corr}(X_9, Y|X_1, X_6) &= -0.287927197687353
\end{aligned}
$$

Among which the absolute maximum is $\operatorname{Corr}(X_2, Y|X_1, X_6) = 0.3000386$. But R^2-score for the model $Y \sim X_1 + X_2 + X_6$ is $0.6886263 < 0.70$.

So we find

$$\begin{aligned}
\mathrm{Corr}\,(X_3, Y|X_1, X_2, X_6) &= -0.152797944791066 \\
\mathrm{Corr}\,(X_4, Y|X_1, X_2, X_6) &= -0.12459244201669 \\
\mathrm{Corr}\,(X_5, Y|X_1, X_2, X_6) &= 0.251774564688212 \\
\mathrm{Corr}\,(X_7, Y|X_1, X_2, X_6) &= -0.0499070701252337 \\
\mathrm{Corr}\,(X_8, Y|X_1, X_2, X_6) &= 0.241092329779549 \\
\mathrm{Corr}\,(X_9, Y|X_1, X_2, X_6) &= -0.327243940162817
\end{aligned}$$

Among which the absolute maximum is $\mathrm{Corr}\,(X_9, Y|X_1, X_2, X_6) =$ -0.3272439. And R^2-score for the model $Y \sim X_1 + X_2 + X_6 + X_9$ is 0.7219709 $>$ 0.70.

Hence, the multiple regression model for the corn yield with 4 the predictors: Year (X_1), Pre-Season precipitation (X_2), July rain (X_6) and August temperature (X_9) explains 72.2% variability in corn yield.

Problem 2

An experiment was run to determine whether four specific temperatures affect the density of a certain type of brick. The experiment led to the following data.

Temperature (μ_i)	Density$_1(y_{i1})$	Density$_2(y_{i2})$	Density$_3(y_{i3})$	Density$_4(y_{i4})$
100	21.8	21.9	21.7	21.6
125	21.7	21.4	21.5	21.4
150	21.9	21.8	21.8	21.6
175	21.9	21.7	21.8	21.4

Does the firing temperature affect the density of the bricks?

Solution

We fit the linear model: $y_{ij} = \mu_i + \epsilon_{ij}; j = 1, 2, 3, 4, i = 1, 2, 3, 4$
and, test the hypothesis, H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
We have, $k = 4$, $n = 16$ Now, *RSS* $= 0.2975$, $RSS_{H_0} = 0.484375$. So we compute the $F-$ratio $= 2.51260504 < 2.60552492 = F_{k-1,n-k}$. Therefore, the firing temperature does not affect the density of the bricks.