Final Exam

Take Home Portion

Hayden Atchley

2022-12-03

1

The full model using all the second-order term is given in Table 1. Many of the *p*-values are not significant, so we use a stepwise function to choose the one with the best (lowest) C(p). The resulting model is given in Table 2.

Table 1: Strength Second-Order Model

term	estimate	std.error	statistic	p.value
(Intercept)	169020.28	13571.35	12.454	0.000
Force	-1119.35	72.02	-15.541	0.000
Power	42.88	220.82	0.194	0.847
Temp	-447.78	74.31	-6.026	0.000
Minutes	158.74	92.63	1.714	0.093
I(Force^2)	1.88	0.12	15.166	0.000
I(Power^2)	-2.11	1.51	-1.400	0.168
I(Temp^2)	0.65	0.16	3.942	0.000
I(Minutes^2)	-0.06	0.22	-0.247	0.806
Force:Power	0.13	0.64	0.207	0.837
Force:Temp	1.16	0.18	6.369	0.000
Force:Minutes	-0.14	0.27	-0.503	0.617
Power:Temp	0.80	0.93	0.856	0.396
Power:Minutes	-0.98	0.87	-1.125	0.266
Temp:Minutes	-0.67	0.31	-2.135	0.038

Table 2: Final Strength Model

term	estimate	std.error	statistic	p.value
(Intercept)	122950.054	12072.841	10.184	0.000
I(Force^2)	1.773	0.175	10.103	0.000
Force	-940.643	93.229	-10.090	0.000
Power	-2.169	19.740	-0.110	0.913
I(Minutes^2)	-0.670	0.246	-2.728	0.008
Temp:Minutes	-0.153	0.077	-1.981	0.052
Force:Minutes	0.417	0.131	3.189	0.002

Based on this model, we can predict the highest strength using inputs from our data. Table 3 gives the first few rows of the input data sorted by highest predicted strength. The first row of the table gives the approximate values of the predictors that will give the highest strength, along with the predicted highest strength.

Table 3: Highest Strength Predictions

Force	Power	Temp	Minutes	Strength	PredictStrength
311	31	136	63	6129	5994
301	19	129	74	3877	4537
215	23	137	71	3994	4102
300	33	155	72	4965	4050
300	25	155	59	5035	3890
217	22	147	71	3329	3705

2

We start with `NO3` as the response variable. We are interested in the effects of density after adjusting for the other demographic variables, so we use these in a regression model. The model is given in Table 4. From this table we can see that the NO3 is correlated with density after adjusting for the other factors, with a highly significant *p*-value.

Table 4: NO3 Model

term	estimate	std.error	statistic	p.value
(Intercept)	-24.823	37.59	-0.660	0.5134
RUNOFF	0.000	0.00	0.064	0.9495
DEPOS	-0.527	1.16	-0.455	0.6519
NPREC	0.203	0.28	0.734	0.4677
AREA	0.000	0.00	-0.044	0.9649
DISCHARG	-0.094	0.06	-1.603	0.1181
PREC	3.350	1.61	2.080	0.0451
DENSITY	0.641	0.11	6.099	0.0000

Now we look at `EXPORT` as the response variable (Table 5). Again, we see that density is correlated, with a small *p*-value. We can conclude that both `NO3` and `EXPORT` are significantly correlated for density after adjusting for the remaining demographic variables.

Table 5: EXPORT Model

term	estimate	std.error	statistic	p.value
(Intercept)	-150.882	347.93	-0.434	0.6673
RUNOFF	0.001	0.01	0.194	0.8471
DEPOS	17.848	10.71	1.667	0.1047
NPREC	-1.854	2.56	-0.723	0.4744
AREA	0.000	0.00	-0.314	0.7558
DISCHARG	0.623	0.55	1.144	0.2607
PREC	-0.972	14.91	-0.065	0.9484
DENSITY	6.330	0.97	6.507	0.0000

3

a

term	estimate	std.error	statistic	p.value
(Intercept)	-44.62	5.177	-8.62	2.15e-15
X	1.54	0.073	20.98	0.00e+00

$Y = -44.62 + 1.54X$

b

term	estimate	std.error	statistic	p.value
(Intercept)	-249.94	29.938	-8.35	1.000e-14
X	7.60	0.876	8.68	0.000e+00
I(X^2)	-0.04	0.006	-6.94	5.449e-11

$Y = -249.94 + 7.6X - 0.04X^2$

c

term	estimate	std.error	statistic	p.value
(Intercept)	-39.31	8.353	-4.71	0.0000
X	1.44	0.139	10.41	0.0000
GenderMale	2.64	3.260	0.81	0.4189

$Y = -39.31 + 1.44X + 2.64(?\text{Male})$

The difference in intercepts is not statistically significant.

d

term	estimate	std.error	statistic	p.value
(Intercept)	-116.64	10.448	-11.16	0.0e+00
X	2.74	0.174	15.71	0.0e+00
GenderMale	156.44	15.903	9.84	8.2e-19
X:GenderMale	-2.26	0.230	-9.81	9.7e-19

$Y = -116.64 + 2.74X + (156.44 - 2.26X)(?\text{Male})$

e

For females the slope is 2.74 and the intercept is -116.64. For males the slope is $2.74 - 2.26 = 0.48$ and the intercept is $-116.64 + 156.44 = 39.80$.

4

Model (a), with constant octane for all 4 regions:

term	estimate	std.error	statistic	p.value
(Intercept)	85.25	0.1	853	0

Model (b), with two groups (A/B and C/D):

term	estimate	std.error	statistic	p.value
(Intercept)	85.23	0.133	641.245	0.0000
GroupCD	0.05	0.204	0.247	0.8059

And model (c), with each region in its own group:

term	estimate	std.error	statistic	p.value
(Intercept)	85.568	0.167	512.656	0.0000
RegionRegion B	-0.786	0.253	-3.108	0.0028
RegionRegion C	-0.176	0.274	-0.642	0.5231
RegionRegion D	-0.387	0.257	-1.504	0.1374

Comparing these models using an extra sum of squares *F*-test gives:

term	df.residual	rss	df	sumsq	statistic	p.value
Octane ~ 1	67	45.5				
Octane ~ Group	66	45.5	1	0.04	0.068	0.7945
Octane ~ Region	64	39.2	2	6.24	5.092	0.0089

Based on these results, model (c) is preferable, as there is a significant difference between this and the other two models.

5

- (T/**F**) The coefficient of a term in a linear regression model will stay the same independent of the other terms in the model.
- A(n) _______ **(interaction)** term indicates how much the correlation of one explanatory variable to the response changes based on another explanatory variable.
- The equation $Y = \beta_0 + \beta_1 X + \beta_2 (X * Z)$ is interpreted graphically as follows:

  a. A single line
  b. Two parallel lines
  c. **Two lines with the same intercept but different slopes**
  d. Two lines with different intercepts and slopes