Modeling Writeup

Aidan Mullan

2/10/2019

Regression Models

3 models were fit to the Pulmonary Embolism dataset. First, a logistic regression model was fit to determine what factors affect the positivity rate of CT scans for PE. Then, two linear models were fit to the effective dose of radiation and amount of contrast given to patients for their scan.

Positivity Model

The full model for rate of positive CT scans considered Gender, Age Range, BMI Range, Type of Procedure, Location of Admittance, and Effective Dose as variables of interest. Within the Type of Procedure variable, the 43 "double rule out" conditions had to be removed from analysis since none of these patients had positive test results. In the end, only location and BMI range were found to be significant predictors for positivity. Coefficients and standard errors for this final model are given in Table 1. The default group for BMI is the under 25 group, and the default for location is ED.

Table 1: Positivity Model Coefficients

	Estimate	Std. Error	p
Intercept	-2.414	0.140	$<.001$ $***$
BMI 25-40	0.294	0.149	.048 $*$
BMI >40	0.107	0.204	.600
Loc: ICU	0.543	0.172	.002 $**$
Loc: IN	0.295	0.151	.051 .
Loc: OUT	-0.147	0.262	.574

Note: Significance: . $< .1$, $* <.05$, $** < .005$, $*** < .001$

It is important to note here that although the BMI and Location variables were significant predictors for positivity rate, not all pairwise differences within each factor are significant. Post-hoc analyses were performed using a Bonferroni correction to reduce the chance of spurious significance.

For the BMI groups, a significant difference was found between the under 25 and between 25-40 factors ($Z = 1.980$, $p = .024$). No other differences were significant (over 40 and under 25: $Z = 0.525$, $p = .300$; over 40 and between 25 and 40: $Z = 1.037$, $p = 0.15$)

For the location variable, significant differences were found between in-patients and ED patients ($Z = 1.951$, $p = .026$), ICU patients and out-patients ($Z = 2.396$, $p = .008$), and between ICU patients and ED patients ($Z = 3.150$, $p < .001$). No other pairwise differences were significant (OUT and ED: $Z = 0.94$, $p = .174$; OUT and IN: $Z = 0.36$, $p = .360$; ICU and IN: $Z = 1.288$, $p = .099$).

Model Interpretation

Since the positivity model utilized logistic regression, the coefficients are interpretated in terms of odds ratios and probability of a positive test result. The way of reading these coefficients is to add the terms of interest and then exponentiate. This provides the odds of a positive test result. For example, the odds that a person with a BMI over 40 who was admitted in the ICU has a positive test result is

$$\exp(-2.414 + 0.107 + 0.543) = 0.171$$

To convert this to a probability, you only need to compute $\frac{\text{Odds}}{1+\text{Odds}}$. In the above example, the probability of a positive test result is

$$\frac{0.171}{1 + 0.171} = .146$$

Therefore we get that the probability of a positive test result for an individual from the ICU with a BMI over 40 is 14.6%.

To compare two groups, such as in-patients and out-patiends, you simply need to divide the two odds. This gives us the odds ratio for a positive test result. The odds ratio for in-patients compared to out-patients is given by

$$\frac{\exp(.295)}{\exp(-0.147)} = 1.56$$

This tells us that the odds that an in-patient has a positive test result are 56% greater than the odds that an out-patient has a positive test result, assuming both patients fall in the same BMI range. To get the reverse odds ratio (out-patient compared to in-patient), you take the inverse of the original ratio ($\frac{1}{1.56} = 0.64$).

All coefficient interpretations are given below.

BMI Groups

- 25-40 compared to <25
  The odds that an individual with a BMI between 25-40 has a positive test result are 34.2% greater (OR = 1.342) than the odds for someone under 25 BMI.
- >40 compared to <25
  The odds that an individual with a BMI over 40 has a positive test result are 11.3% greater (OR = 1.113) than the odds for someone under 25 BMI.
- >40 compared to 25-40
  The odds that an individual with a BMI over 40 has a positive test result are 11.1% less (OR = 0.889) than the odds for someone between 25 and 40 BMI.

Location Groups

- OUT compared to ED
  The odds that an individual addmitted as an out-patient has a positive test result are 13.7% less (OR = 0.863) than the odds for someone addmitted from ED
- IN compared to ED
  The odds that an individual addmitted as an in-patient has a positive test result are 34.3% greater (OR = 1.343) than the odds for someone addmitted from ED
- ICU compared to ED
  The odds that an individual addmitted in the ICU has a positive test result are 72.1% greater (OR = 1.721) than the odds for someone addmitted from ED
- IN compared to OUT
  The odds that an individual addmitted as an in-patient has a positive test result are 55.6% greater (OR = 1.556) than the odds for someone addmitted as an out-patient
- ICU compared to OUT
  The odds that an individual addmitted to the ICU has a positive test result are 99.4% greater (OR = 1.994) than the odds for someone addmitted as an out-patient

- ICU compared to IN
  The odds that an individual addmitted to the ICU has a positive test result are 28.1% greater (OR = 1.281) than the odds for someone addmitted from as an in-patient

Effective Dosage Model

The full model for effective dose considered CT type, Age and BMI range, Gender, and Location of admittance. Effective dosage was log-transformed to normalize the distribution, and then a linear model was fit to the data. Variable selection found that every variable was significant, and coefficients are given in Table 2. The default factors were: BMI range - under 25, Location - ED, CT type - 128SSwIR, Age - under 18.

Table 2: Positivity Model Coefficients

	Estimate	Std. Error	p
Intercept	0.305	0.138	.027 $*$
BMI 25-40	0.491	0.029	$<$.001 $***$
BMI >40	1.131	0.040	$<$.001 $***$
Loc: ICU	0.297	0.039	$<$.001 $***$
Loc: IN	0.013	0.032	.678
Loc: OUT	0.032	0.051	.528
CT: 64SSwIR	0.109	0.064	.087 .
CT: 64SSWoIR	0.526	0.085	$<$.001 $***$
CT: DS Scanner	-0.118	0.027	$<$.001 $***$
Age: 18-35	0.751	0.134	$<$.001 $***$
Age: >35	0.795	0.134	$<$.001 $***$
Gender: Male	0.217	0.026	$<$.001 $***$

Note: Significance: . $< .1$, $* <.05$, $** < .01$, $*** < .001$

Model Interpretation

Since this is a linear model with a log-transformed response variable, the interpretation is similar to the logistic case. All we need to do is add up the coefficients for each factor we are considering and exponentiate. For example, suppose we have a male between the ages of 18 and 35, with a BMI over 40. He was admitted from the ICU and received a CT scan on the 64SSwIR machine. According to our model, his effective dose would be

$$\exp(0.305 + 0.217 + 0.751 + 0.297 + 0.109) = 5.36$$

If we are interested in comparing the change in effective dose between two levels of a single factor, we would do the same thing as the logistic case. Here, we would exponentiate the two coefficients and take their difference. For example, suppose we want to compare the 64SSwIR scanner with the 64SSWoIR scanner. Then we calculate

$$\frac{\exp(.109)}{\exp(.526)} = .659$$

which tells us that the 64SSwIR scanner is associated with 34.1% decreased effective dose.

Contrast Model

Similar to the effective dose model, the full model for contrast considered CT type, Age and BMI range, Gender, and Location of admittance. No transformation was necessary, so a standard linear model was fit to the data. Again, variable selection found that every variable was significant, and coefficients are given in Table 3. The default factors were: BMI range - under 25, Location - ED, CT type - 128SSwIR, Age - under 18.

Table 3: Positivity Model Coefficients

	Estimate	Std. Error	p
Intercept	66.364	3.184	<.001 $***$
BMI 25-40	3.343	0.805	<.001 $***$
BMI >40	17.264	1.099	<.001 $***$
Loc: ICU	5.918	1.085	<.001 $***$
Loc: IN	1.238	0.884	.161
Loc: OUT	2.696	1.420	.058 .
CT: 64SSwIR	7.507	1.767	<.001 $***$
CT: 64SSWoIR	6.791	2.347	.004 $**$
CT: DS Scanner	-9.073	0.740	<.001 $***$
Age: 18-35	21.716	3.849	<.001 $***$
Age: >35	21.373	3.725	<.001 $***$
Gender: Male	4.324	0.707	<.001 $***$

Note: Significance: . $< .1$, $* < .05$, $** < .01$, $*** < .001$

Model Interpretation

This contrast model is a standard linear regression, so the interpretation is pretty straightforward. Since all of our variables are factors, the coefficient for a given factor is the associated difference in contrast between the given factor level and the default. For example, patients with a BMI over 40 are associated with an increase in contrast of 17.26 units over patients with a BMI under 25.

To compare two factor levels that are not the default, we need to take the difference in their coefficients. Suppose we want to compare ICU patients with in-patients. A patient in the ICU receives on average of $5.918 - 1.238 = 4.68$ units more contrast than an in-patient.

Otherwise, to determine the amount of contrast associated with a given patient, all we need to do is add up the coefficients for each category that the given patient falls into. A 45 year old female with a BMI between 25 and 40 admitted as an out-patient and scanned on a DS scanner would be associated with $66.364 + 21.373 + 3.343 + 2.696 - 9.073 = 84.703$ units of contrast.