University of Central Florida
College of Business

QMB 6911
Capstone Project in Business Analytics

Solutions: Problem Set #10

1 Data Description

By engaging an industry consultant to gather relevant and appropriate information, your firm has been able to put together data concerning 248 different fly-fishing reels, over one-half of which are produced in the United States, with the remainder being produced in Asia—either in China or Korea. These data are contained in the file `FlyReels.csv`, which is available in the `Data` folder. Each fly-fishing reel in the data set is a row, while the columns correspond to the variables whose names and definitions are the following:

Variable	Definition
Name	product name (a string)
Brand	brand name (a string)
Weight	weight of reel in ounces (a real number)
Diameter	diameter of reel in inches (a real number)
Width	width of reel in inches (a real number)
Price	price of reel in dollars (a real number)
Sealed	whether the reel is sealed; "Yes" versus "No" (a string)
Country	country of manufacture, (a string)
Machined	whether the reel is machined versus cast; machined="Yes", while cast="No" (a string)

I will revisit the recommended linear model from Problem Set #7, which was supported in Problem Sets #8 and #9 by considering other nonlinear specifications within a Generalized Additive Model.

Then I will further investigate this nonlinear relationship by considering the issue of sample selection: fly reel manufacturers may produce fly reels in each country with specific qualities based on their perceived value to typical American customers, in ways that are not represented by the variables in the dataset.

	Model 1
(Intercept)	2.00999***
	(0.26125)
Width	0.33575*
	(0.15622)
Diameter	0.39567***
	(0.05076)
Density	1.21296***
	(0.21948)
SealedYes	0.62731***
	(0.08622)
MachinedYes	0.64934***
	(0.08320)
made_in_USATRUE	0.74633***
	(0.09247)
SealedYes:made_in_USATRUE	-0.29519**
	(0.10092)
R^2	0.74893
Adj. R^2	0.74160
Num. obs.	248

$^{***}p < 0.001;$ $^{**}p < 0.01;$ $^{*}p < 0.05$

Tab. 1: Linear Model for Fly Reel Prices

2 Linear Regression Model

A natural staring point is the recommended linear model from Problem Set #7.

2.1 Linear Model with `Sealed*Made_in_USA` Interaction

Last week I investigated whether the functional form should include different specifications by country of manufacture. The model included the continuous variables width, diameter, and density, as well as categorical variables for country of manufacture, and whether or not the reels were sealed or machined. In addition to the indicator for the country of manufacture, the model included an indicator for an interaction between the the country of manufacture indicator and the indicator for whether the reels were sealed or unsealed. The dependent variable was chosen as the logarithm of the fly reel price, since the results were similar to those from the model with the optimal Box-Cox transformation, without the added complexity. The results of this regression specification are shown in Table 1.

Next, I will attempt to improve on this specification, using Tobit models for sample selection.

	Model 1	Model 2	Model 3	Model 4
(Intercept)	3.35***	3.48***	2.03***	2.41***
	(0.30)	(0.28)	(0.47)	(0.41)
Width	0.28		0.39	
	(0.22)		(0.23)	
Diameter	0.44***	0.49***	0.36***	0.41***
	(0.07)	(0.05)	(0.08)	(0.07)
Density	1.13***	1.06***	1.32**	1.06**
	(0.25)	(0.25)	(0.41)	(0.38)
SealedYes	0.32***	0.32***	0.65***	0.64***
	(0.06)	(0.06)	(0.10)	(0.10)
MachinedYes			0.65***	0.63***
			(0.09)	(0.09)
R^2	0.54	0.53	0.75	0.74
Adj. R^2	0.52	0.52	0.73	0.73
Num. obs.	135	135	113	113

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Tab. 2: Regression Models by Country of Manufacture

2.2 Comparison by Country of Manufacture

Table 2 shows the results of a series of regression models on different samples by country of manufacture. Model 1 shows the results for the full model using the sample of fly reels made in the USA and Model 3 shows the remaining fly reels made in China or Korea. Models 2 and 4 show the estimates from a reduced model on each sample, eliminating the variable `Width`, which was not significant in either of these smaller samples.

The width of the reel is insignificant in both samples and the coefficients are qualitatively similar across the samples, as well as matching in significance. This suggests that one model might be sufficient.

3 Sample Selection

3.1 Predicting the Selection into Samples

The specification in Table 1 assumes a linear functional form for the relationship between characteristics and prices of fly reels, without selecting into samples by brand. To investigate this relationship further, consider the set of variables that are related to whether or not a manufacturer decides to manufacture fly reels in America or overseas with the characteristics observed in the dataset.

3.2 Estimating a Sample Selection Model