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.

	Model 1	Model 2
(Intercept)	-3.46545	3.38980^{**}
	(159.27332)	(1.06142)
Weight	0.11330	0.13348^{*}
	(0.19412)	(0.05711)
Diameter	-0.08526	
	(0.86262)	
Width	-0.71463	-2.45931^{**}
	(1.68668)	(0.81177)
Volume	-0.03294	
	(0.09631)	
Density	-2.19723	-2.02613^{*}
	(2.46062)	(0.98455)
SealedYes	-1.23411^{***}	-0.68650^{***}
	(0.23449)	(0.18639)
MachinedYes	6.58401	
	(159.21535)	
AIC	267.46344	328.68383
BIC	295.57087	346.25098
Log Likelihood	-125.73172	-159.34192
Deviance	251.46344	318.68383
Num. obs.	248	248

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

Tab. 3: Probit Models for Country-of-Manufacture Selection of Fly Reels

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.

Table 3 shows the estimates for a probit model to predict the selection into samples by country of manufacture. Model 1 in Table 3 shows a preliminary probit model to predict the selection indicator, with all the other explanatory variables in the model. American fly reel manufacture seems only to be related to whether or not the reels are sealed. Model 2 shows the result of a variable-reduction exercise to eliminate variables that are not statistically significant. These estimates provide a concise but useful model to indicate the fly reel designs that manufacturers would prefer to manufacture on American soil.

This model is used to specify the selection equation of the sample selection estimates discussed next.

3.2 Estimating a Sample Selection Model

Table 4 shows the estimates from a model that accounts for sample selection. The models are estimates from the Tobit model of type 5, which is a model specification that allows for switching of the observations in the sample into two models: one for the value of fly reels made in the USA and the other for fly reels produced elsewhere.

Each column shows the estimates from a separate model and the series of models is the result of a downward selection procedure in which a statistically insignificant variable was removed from each of the previous models in the sequence. In each model, the estimates are grouped into three categories. The first block of coefficients describe the selection model to determine whether a fly reel design would be manufactured in the USA. These coefficients are denoted by the prefix"S:". Below these lies two blocks of coefficients for the observation equations. The notation "O: ¡name of variable¿ (i)" indicates the coefficient for the particular variable in the observation equation for sample i. In this specification, the first observation equation represents fly reels made overseas (`made_in_USA == 0`), while equation 2 represents the fly reels made in the USA (`made_in_USA == 1`).

Model 1 shows the estimates from the full model. Several of the coefficients in the model are statistically insignificant and the model has numerical issues. In particular, some of the standard errors are undefined, evidenced by the missing standard error for `sigma1`. This suggests that the likelihood function is flat in some areas of the parameter space. This model has its imperfections but is a good start. Several variables are statistically insignificant but these can be removed one by one to produce a refined model. The goal will be to obtain a final model that has well-defined standard errors for all variables and, ideally, all coefficients statistically significant.

Model 2 shows the estimates from a reduced model, eliminate `Diameter` from the other-country equation. There still exists coefficients that are statistically insignificant and the numerical issue remains for `sigma1`. One step further, Model 3 eliminates `Width` from the American observation equation. Again, this model is an improvement but a numerical issue remains.

Model 4 excludes `Weight` from the selection equation. The produces a set of estimates that are numerically stable, within a strictly concave region of the likelihood function. This model, however, still contains some variables that are statistically significant. Model 5 excludes `Density` from the selection equation. Again, this model is well-behaved numerically but one statistically insignificant variable remains.

The next step is to estimate what would be Model 6 by eliminating `Width` from the selection equation. This specification is problematic, however, as it produces an error message indicating severe multicollinearity. With these numerical problems, it is better to keep the additional variable in the selection equation, even though it may be statistically insignificant. Notice that the remaining selection variable `Sealed` appears in both observation equations and there is no variable in the selection equation that is excluded. The other extreme would offer better performance, that is, having some variables in the selection equation that are not included in the observation equations.

	Model 1	Model 2	Model 3	Model 4	Model 5
S: (Intercept)	3.38999^{*}	3.38995^{*}	3.39163^{**}	1.84476^{*}	1.06496^{*}
	(1.52501)	(1.39731)	(1.24301)	(0.88778)	(0.49241)
S: Width	-2.45920	-2.45922^{*}	-2.45815^{*}	-0.87031	-0.59098
	(1.26540)	(1.18220)	(1.06044)	(0.51987)	(0.45219)
S: Weight	0.13399	0.13392	0.13475		
	(0.08555)	(0.08136)	(0.07909)		
S: Density	-2.02599	-2.02603	-2.02510	-0.88282	
	(1.36507)	(1.18246)	(1.15198)	(0.83583)	
S: SealedYes	-0.68657^{**}	-0.68656^{**}	-0.68891^{***}	-0.52031^{**}	-0.47216^{**}
	(0.25772)	(0.21197)	(0.15923)	(0.17879)	(0.17222)
O: (Intercept) (1)	0.26764	0.23849	0.30737	1.89824^{**}	2.14641^{***}
	(1.08958)	(0.88866)	(0.78180)	(0.59989)	(0.50263)
O: Width (1)	1.73696^{***}	1.75076^{***}			
	(0.51414)	(0.50948)			
O: Diameter (1)	0.02868				
	(0.18074)				
O: Density (1)	1.70252	1.70122	1.73905^{*}	1.09840^{*}	0.79690^{*}
	(0.97121)	(0.88239)	(0.79321)	(0.48704)	(0.37891)
O: SealedYes (1)	1.22730^{***}	1.23870^{***}	1.24406^{***}	0.92288^{***}	0.90755^{***}
	(0.21538)	(0.22743)	(0.18759)	(0.11757)	(0.11456)
O: MachinedYes	0.61018^{**}	0.60781^{**}	0.63222^{***}	0.75933^{***}	0.76203^{***}
	(0.21795)	(0.20627)	(0.18592)	(0.09168)	(0.09128)
O: (Intercept) (2)	3.32465^{***}	3.32465^{***}	3.30003^{***}	3.49131^{***}	3.44177^{***}
	(0.30694)	(0.30651)	(0.30380)	(0.28482)	(0.27958)
O: Width (2)	0.14939	0.14939			
	(0.26002)	(0.25871)			
O: Diameter (2)	0.47103^{***}				
	(0.07550)				
O: Density (2)	1.09216^{***}	1.09216^{***}	1.04813^{***}	1.00088^{***}	1.06657^{***}
	(0.26518)	(0.26347)	(0.26497)	(0.26029)	(0.24389)
O: SealedYes (2)	0.26663^{**}	0.26662^{**}	0.21502^{***}	0.27774^{***}	0.28495^{***}
	(0.08286)	(0.08173)	(0.06041)	(0.07395)	(0.07554)
sigma1	0.83812	0.89629	0.80395	0.52930^{***}	0.53170^{***}
				(0.07186)	(0.06979)
sigma2	0.31216^{***}	0.31216^{***}	0.33612^{***}	0.31969^{***}	0.31675^{***}
	(0.04046)	(0.04029)	(0.02417)	(0.04534)	(0.04844)
rho1	-1.00000	-0.99511	-1.00000	-0.87806^{***}	-0.88135^{***}
	(3.66988)			(0.08675)	(0.08048)
rho2	0.39798	0.39798	0.75807^{***}	0.46352	0.43112
	(0.44205)	(0.43892)	(0.06565)	(0.40092)	(0.48226)
O: Diameter		0.47102^{***}	0.50888^{***}	0.47502^{***}	0.48018^{***}
		(0.07531)	(0.05880)	(0.05735)	(0.05664)
O: Width			1.81227^{***}	1.25522^{***}	1.17131^{***}
			(0.46808)	(0.30287)	(0.28044)
AIC	591.29253	617.14337	581.03621	513.34065	512.46369
BIC	661.56110	683.89852	644.27793	573.06894	568.67855
Log Likelihood	-275.64627	-289.57169	-272.51811	-239.67032	-240.23184
Num. obs.	248	248	248	248	248
Censored	113	113	113	113	113
Observed	135	135	135	135	135

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

Tab. 4: Selection Models for Fly Reel Prices

I revert back to Model 5 to analyze the differences between country of manufacture, in the rightmost column of Table 4. First of all, this confirms that machined reels are more valuable, with a coefficient of 0.76, instead of 0.63, as was found in the separate linear regression model. It also justifies the fact that only machined reels are produced in the USA. This could relate to some advantage in American production technology or, rather, the outdated casting techniques used for cheaper reels produced overseas.

In the model for American-made reels, the coefficients on all three variables are statistically the same as those in the separate linear regression model. For the fly reels made overseas, however, `Width` replaces `Diameter` as a proxy for the size of the reels. As a consequence, density is a relatively less valuable feature. The value of sealed reels, however, is even greater, after accounting for the selection of different production techniques jointly with manufacturing location. The coefficient on sealed reels jumps to 0.90, compared to 0.65 in the linear model. This suggests a higher premium than indicated earlier, once we also consider the choice of country of manufacture. Similarly, the value of machined reels rises from 0.65 to 0.76 in the selection model, suggesting an even higher for this design when produced overseas.

In conclusion, an American fly reel producer should not consider producing cast reels, unless the purpose is to explore the change in value. I suspect, however, that this outcome has been observed in the past, so the older production techniques have been abandoned for a reason. Machined reels are also more valuable when produced overseas, so a company has to compare the difference in labor costs with the relative cost of producing machined reels overseas.

Similarly, sealed reels produce a much higher premium overseas than was originally estimated with the linear model. Reels produced overseas should be sealed, unless the cost of changing the manufacturing process would outweigh this premium. Likewise for the American reels, except the premium is one third the size.

However it is measured, the size of a reel matters: bigger reels are more valuable. A manufacturer can compare the cost of materials with the premiums attached to those dimensions when producing a reel.

Perhaps the largest difference is in the intercept term: 1.07 overseas vs 2.14 in the USA, double the size. In the linear model, the intercepts were 2.41 vs 3.48, which measures a similar percentage difference. No matter how it is measured, there exists a substantial premium for fly reels made in the USA.