University of Central Florida
College of Business

QMB 6911
Capstone Project in Business Analytics

Solutions: Problem Set #6

1 Analyzing the Dependent Variable

1.1 Probability Density Function of Fly Reel Prices

Figure 1 is the kernel-smoothed probability density function of the natural
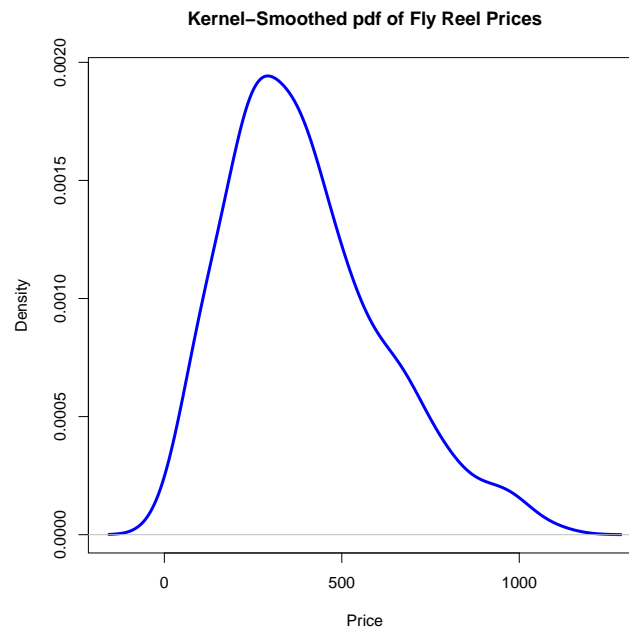logarithm of price:

Fig. 1: Probability Density Function of Fly Reel Prices

1.2 Probability Density Function of the Log. of Fly Reel Prices

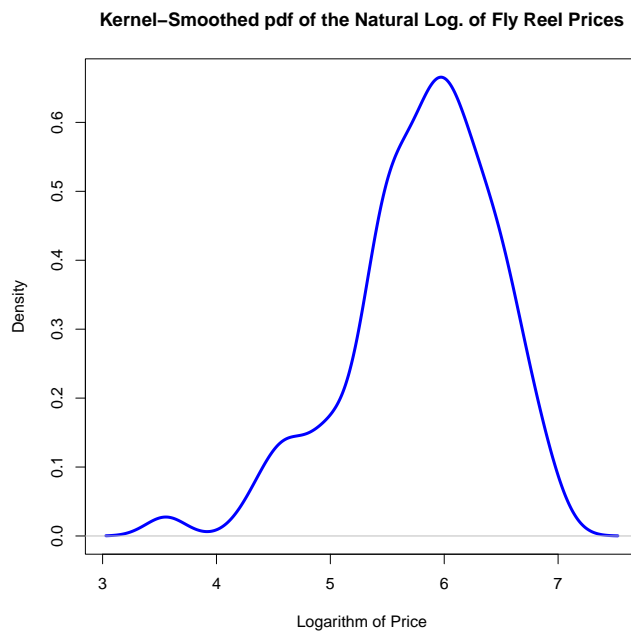Figure 2 is the kernel-smoothed probability density function of the natural logarithm of price:

Kernel−Smoothed pdf of the Natural Log. of Fly Reel Prices

Fig. 2: Probability Density Function of Fly Reel Prices

1.3 Probability Density Function of the Box-Cox Transformation of Fly Reel Prices

Figure 3 is the kernel-smoothed probability density function of the natural logarithm of price. Although the transformed model does look closest to normal, all three distributions are reasonably symmetric, indicating that the use of the transformation is subject to judgment. Whether it is worth using the transformation can be settled by analyzing the results of regression models with different forms of the dependent variable.
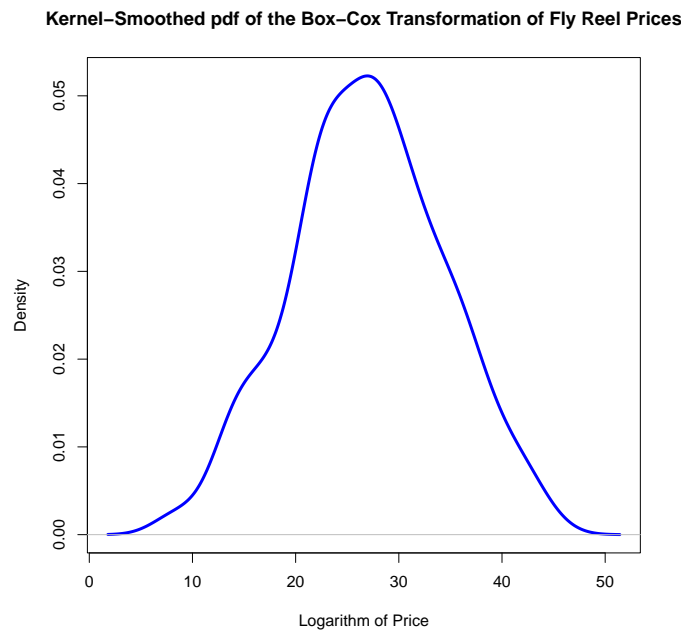
Kernel−Smoothed pdf of the Box−Cox Transformation of Fly Reel Prices

Fig. 3: Probability Density Function of Transformed Fly Reel Prices

1.4 Comparison by Transformation of Dependent Variable

Table 1 shows the results of a series of regression models with different definitions of the dependent variable. Model 1 uses the fly reel prices without transformation, Model 2 uses the logarithm of the fly reel prices, and Model 3 uses the Box-Cox transformation of fly reel prices, with the optimal value of the parameter in the exponent. Although the model built on the original price levels has statistically significant coefficients, the two transformed

	Model 1	Model 2	Model 3
(Intercept)	−62.63	4.05***	8.51**
	(83.91)	(0.21)	(2.57)
Width	−157.49*	−0.31	−4.58*
	(63.68)	(0.16)	(1.95)
Diameter	−4.97	0.06	0.32
	(27.25)	(0.07)	(0.84)
Weight	55.69***	0.11***	1.56***
	(7.06)	(0.02)	(0.22)
SealedYes	145.82***	0.42***	5.02***
	(19.77)	(0.05)	(0.61)
MachinedYes	98.97**	0.73***	6.11***
	(29.72)	(0.08)	(0.91)
made_in_USATRUE	201.07***	0.51***	6.47***
	(19.26)	(0.05)	(0.59)
R^2	0.66	0.75	0.72
Adj. R^2	0.66	0.74	0.72
Num. obs.	248	248	248

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

Tab. 1: Regression Models with Different Dependent Variables

models have a better fit, with higher values of $\bar{R}^2$. Since the difference between the latter two models is marginal, it is better to model the logarithm of fly reel prices, which has the added advantage of interpretability of the coefficients, which approximately represent percentage changes in fly reel prices.

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	3.94***	5.19***	5.24***	4.13***	3.91***
	(0.08)	(0.26)	(0.08)	(0.32)	(0.11)
Weight	0.10***	0.10***	0.11***	0.11***	0.09***
	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)
SealedYes	0.42***	0.32***	0.32***	0.65***	0.64***
	(0.05)	(0.06)	(0.05)	(0.09)	(0.09)
MachinedYes	0.76***			0.62***	0.65***
	(0.07)			(0.09)	(0.09)
made_in_USATRUE	0.52***				
	(0.05)				
Width		−0.32		−0.27	
		(0.23)		(0.23)	
Diameter		0.10		−0.01	
		(0.09)		(0.11)	
R^2	0.74	0.55	0.54	0.75	0.75
Adj. R^2	0.74	0.53	0.53	0.74	0.74
Num. obs.	248	135	135	113	113

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

Tab. 2: Regression Models by Country of Manufacture

1.5 Comparison by Country of Manufacture

In contrast to our analysis of tractor prices, we will take the more aggressive step of splitting the sample to estimate separate models before performing any variable reduction. Table 2 shows the results of a series of regression models on different samples by country of manufacture. Models 2 and 3 show the results for the sample of fly reels made in the USA and Models 4 and 5 show the results for the remaining fly reels made in China or Korea. Model 1 is the model built on the full sample, excluding the insignificant variables relating to the size of the reels.

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. A notable exception is the difference in value of sealed reels between the two subsamples. This suggests that one model might be sufficient.

To test this statistically, I conduct an F-test. This compares a single model with only an indicator for the country of manufacture (the restricted model) with a separate model for each country. In this case, the full, unre-

stricted model has $K = 2 \times 4 = 8$ parameters, one for each variable in two models. The test that all of the coefficients are the same has $M = 4 - 1 = 3$ restrictions. The one restriction fewer accounts for the made-in-USA indicator in the full model, which allows for two separate intercepts. The F-statistic has a value of

$$\frac{(RSS_M - RSS)/M}{RSS/(N-K)} = \frac{(26.08593 - 25.05317)/3}{25.05317/235} = 3.297812.$$

This value is greater than 1, so we can compare it to the critical value of the F-statistic at the specified degrees of freedom for a conventional level of significance. These critical values are 3.864, 2.642, and 2.107 at the 1%, 5%, and 10% levels of significance, respectively.

This places the F-statistic between the critical values for the 5 and 10 percent levels of significance. Conclude that fly reel prices may have some difference by country of manufacture but the difference is marginal. This suggests little justification for separate models by country of manufacture. We can investigate small differences between the models.

In the next section, I will consider a hybrid model with some interaction terms with separate coefficients by country of manufacture.

1.6   Models with Interaction Terms

Table 3 shows the results of a series of regression models with different specifications of interaction terms by country of manufacture. The model built on the full sample, without interaction terms, is shown in Model 1. The interaction with sealed and country of manufacture is significant in both specifications Model 2 and Model 4. In contrast, the interaction with weight was not significant. Since all variables are significant in this model and it has the highest $\bar{R}^2$, Model 2 is the recommended model.

As a sensitivity analysis, I estimated the models again after adding back the excluded variables of width and diameter using the recommended specification and found neither to be significant in the recommended specification. It is often worth revisiting past decisions whenever you make a change in model specification.

In terms of the production decision, there exists a substantial premium for fly reels made in the USA but this premium is half as large for sealed reels, however, this reduced premium is still smaller than the additional value for a sealed reel, since sealed reels already carry a high premium.

	Model 1	Model 2	Model 3	Model 4
(Intercept)	3.94***	3.87***	3.95***	3.91***
	(0.08)	(0.08)	(0.10)	(0.10)
Weight	0.10***	0.10***	0.10***	0.09***
	(0.01)	(0.01)	(0.01)	(0.01)
SealedYes	0.42***	0.63***	0.42***	0.64***
	(0.05)	(0.08)	(0.05)	(0.08)
MachinedYes	0.76***	0.65***	0.76***	0.65***
	(0.07)	(0.08)	(0.07)	(0.08)
made_in_USATRUE	0.52***	0.76***	0.50***	0.68***
	(0.05)	(0.09)	(0.14)	(0.14)
SealedYes:made_in_USATRUE		-0.31**		-0.32**
		(0.10)		(0.10)
Weight:made_in_USATRUE			0.00	0.01
			(0.02)	(0.02)
R^2	0.74	0.75	0.74	0.75
Adj. R^2	0.74	0.75	0.74	0.75
Num. obs.	248	248	248	248

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

Tab. 3: Regression Models with Interaction Terms by Country of Manufacture