Spring 2023 **Firstname M. Lastname**

University of Central Florida
College of Business

QMB 6911
Capstone Project in Business Analytics

Solutions: Problem Set #7

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 #6. In doing so, I will investigate any nonlinear relationships by incorporating a nonlinear but parametric specification for the value of the dimensions of the reels: the width, diameter, and density, which constitute the continuous variables in the dataset. This parametric analysis will be performed using the Box-Tidwell framework to investigate whether the value of these characteristics are best described with parametric nonlinear forms.

	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 #6.

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 by investigating the potential for nonlinear functional forms.

3 Nonlinear Specifications

4 The Box–Tidwell Transformation

The Box–Tidwell function tests for non-linear relationships to the mean of the dependent variable. The nonlinearity is in the form of an exponential transformation in the form of the Box-Cox transformation, except that the transformation is taken on the explanatory variables.

4.1 Transformation of Width

Performing the transformation on the width variable produces a modified form of the linear model. This specification allows a single exponential transformation on width, rather than a linear form.

```
 MLE of lambda Score Statistic (z) Pr(>|z|)
       1.1615              0.1587   0.8739

iterations =  5
```

The R output is the statistics for a test of nonlinearity: that the exponent λ in the Box–Tidwell transformation is zero. The ”`MLE of lambda`” statistic is the optimal exponent on horsepower. Similar to the Box-Cox transformation, with Box-Tidwell, the exponents are on the explanatory variables and are all called lambda, in contrast to the parameter τ in our class notes. The exponent is not significantly different from one. This supports a linear form for Width, confirming our result from the nonparametric analysis.

4.2 Transformation of Diameter

```
 MLE of lambda Score Statistic (z) Pr(>|z|)
     -0.35213             -1.2106    0.226

iterations =  4
```

This is very weak evidence for an inverse square root transformation for diameter but it is not estimated accurately. There is no evidence for a transformation with a coefficient different from 1, which suggests a purely linear relationship between `log_Price` and diameter. Next, I will consider the possibility of nonlinearity in the value of the density of a reel.

4.3 Transformation of Density

```
 MLE of lambda Score Statistic (z) Pr(>|z|)
      0.19315             -1.3448   0.1787

iterations =  10
```

Similar to Diameter, this is very weak evidence for a square root transformation for diameter but it is not estimated accurately. Conclude that the linear model is the best choice, when nonlinear options are restricted to this parametric form. It is possible, however, that a nonlinear relationship exists but takes on a form not captured by this specific framework.